

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7566638号
(P7566638)

(45)発行日 令和6年10月15日(2024.10.15)

(24)登録日 令和6年10月4日(2024.10.4)

(51)国際特許分類		F I		
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00	
G 0 1 N	21/62 (2006.01)	G 0 1 N	21/62	Z
G 1 6 B	30/10 (2019.01)	G 1 6 B	30/10	
請求項の数 15 (全301頁)				
(21)出願番号	特願2020-572706(P2020-572706)	(73)特許権者	500358711	
(86)(22)出願日	令和2年3月22日(2020.3.22)		イルミナ インコーポレイテッド	
(65)公表番号	特表2022-535306(P2022-535306 A)		アメリカ合衆国 カリフォルニア州 9 2	
(43)公表日	令和4年8月8日(2022.8.8)		1 2 2 サンディエゴ イルミナ ウェイ	
(86)国際出願番号	PCT/US2020/024092	(74)代理人	5 2 0 0	
(87)国際公開番号	WO2020/191391		100108453	
(87)国際公開日	令和2年9月24日(2020.9.24)	(74)代理人	弁理士 村山 靖彦	
審査請求日	令和5年3月22日(2023.3.22)	(74)代理人	100110364	
(31)優先権主張番号	62/821,602		弁理士 実広 信哉	
(32)優先日	平成31年3月21日(2019.3.21)	(74)代理人	100133400	
(33)優先権主張国・地域又は機関	米国(US)	(72)発明者	弁理士 阿部 達彦	
(31)優先権主張番号	62/821,618		アニンディタ・ダッタ	
(32)優先日	平成31年3月21日(2019.3.21)		アメリカ合衆国・カリフォルニア・9 2	
最終頁に続く			1 2 2・サン・ディエゴ・イルミナ・ウ	
			エイ・5 2 0 0	
			最終頁に続く	

(54)【発明の名称】 人工知能ベースの配列決定

(57)【特許請求の範囲】

【請求項1】

テンプレート生成及びベースコールを含む、エンドツーエンド配列決定のコンピュータ実装の方法であって、

光学 픽セル解像度で 픽セルを含む第1の画像データ及び第2の画像データにアクセスすることであって、

前記第1の画像データが、配列決定動作の配列決定サイクルのうちの初期の1つで、配列決定システムによって捕捉される、クラスター及びクラスターの周囲の背景の画像を含み、

前記第2の画像データが、前記配列決定動作の前記配列決定サイクルで、前記配列決定システムによって捕捉される、前記クラスター及びクラスターの周囲の背景の画像を含む、アクセスすることと、

ニューラルネットワークベースのテンプレート生成器を介して前記第1の画像データを処理し、クラスターメタデータを識別するクラスターマップを生成することであって、

前記クラスターメタデータが、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及びクラスター境界のうちの少なくとも1つを含み、

前記ニューラルネットワークベースのテンプレート生成器が、前記クラスターメタデータへの前記クラスターの前記画像のマッピングのタスクで訓練される、生成することと、

アップサンプリングされたサブ 픽セル解像度でテンプレート画像内の前記クラスターメタデータを符号化することであって、

前記テンプレート画像のサブピクセル及び前記クラスターの前記画像の前記ピクセルが、同じ画像領域を表す、符号化することと、

前記テンプレート画像に基づいて、前記第2の画像データの前記ピクセルの強度値を修正し、前記クラスターメタデータを考慮する強度分布を有する前記第2の画像データの強度修正バージョンを生成することと、

ニューラルネットワークベースのベースコーラーを介して前記第2の画像データの前記強度修正バージョンを処理し、前記配列決定動作の1つ又はそれ以上の配列決定サイクルで、前記クラスターのうちの1つ又はそれ以上に対するベースコールを生成することであって、前記ニューラルネットワークベースのベースコーラーが、前記ベースコールへの前記クラスターの前記画像のマッピングのタスクで訓練される、生成することと、を含む、コンピュータ実装の方法。

10

【請求項2】

前記第2の画像データを前記テンプレート画像に補うことと、

前記ニューラルネットワークベースのベースコーラーを介して、前記テンプレート画像に補われる前記第2の画像データを処理し、前記配列決定動作の1つ又はそれ以上の配列決定サイクルで、前記クラスターのうちの1つ又はそれ以上に対するベースコールを生成することと、を更に含む、請求項1に記載のコンピュータ実装の方法。

【請求項3】

前記テンプレート画像内の各サブピクセルが、背景サブピクセル、クラスター中心サブピクセル、又はクラスター内部サブピクセルのいずれかとして識別される、請求項1に記載のコンピュータ実装の方法。

20

【請求項4】

前記第2の画像データの前記ピクセルの強度値を修正することが、前記第2の画像データの前記画像内のピクセルに対応する前記テンプレート画像内のどれくらい多くのサブピクセルが、前記クラスターのうちの1つ又はそれ以上の一部を含むかに基づいて、前記第2の画像データ内の1つ又はそれ以上のピクセルに対する領域重み付け係数を計算することと、前記領域重み付け係数に基づいて、前記ピクセルの強度を修正することと、を含み、

前記第2の画像データの前記ピクセルの強度値を修正することが、クラスター及びクラスターの周囲の背景の前記画像を前記アップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、前記テンプレート画像内の背景サブピクセルに対応する前記アップサンプリングされた画像内のサブピクセルに背景強度を割り当てることと、前記テンプレート画像内のクラスター中心サブピクセル及びクラスター内部サブピクセルに対応する前記アップサンプリングされた画像内のサブピクセルにクラスター強度を割り当てることと、を含み、前記背景強度が、ゼロ値を有し、

30

前記第2の画像データの前記ピクセルの強度値を修正することが、クラスター及びクラスターの周囲の背景の前記画像を前記アップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、前記テンプレート画像内の前記クラスター中心サブピクセル及び前記クラスター内部サブピクセルに対応する前記アップサンプリングされた画像内の前記ピクセルの構成サブピクセルのみの間で、光学ピクセル解像度内のピクセルの全体の強度を分散させることと、を含む、請求項1～3のいずれか一項に記載のコンピュータ実装の方法。

40

【請求項5】

前記クラスター強度が、前記光学ピクセル解像度で前記ピクセルの強度を補間することによって決定される、請求項4に記載のコンピュータ実装の方法。

【請求項6】

システムであって、

少なくとも1つのプロセッサと、

前記少なくとも1つのプロセッサによって実行されたときに、

光学ピクセル解像度でピクセルを含む第1の画像データ及び第2の画像データにアクセス

50

することであって、

前記第1の画像データが、配列決定動作の配列決定サイクルのうちの初期の1つで、配列決定システムによって捕捉される、クラスター及びクラスターの周囲の背景の画像を含み、前記第2の画像データが、前記配列決定動作の前記配列決定サイクルで、前記配列決定システムによって捕捉される、前記クラスター及びクラスターの周囲の背景の画像を含む、アクセスすることと、

ニューラルネットワークベースのテンプレート生成器を介して前記第1の画像データを処理し、クラスターメタデータを識別するクラスターマップを生成することであって、

前記クラスターメタデータが、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及びクラスター境界のうちの少なくとも1つを含み、

前記ニューラルネットワークベースのテンプレート生成器が、前記クラスターメタデータへの前記クラスターの前記画像のマッピングのタスクで訓練される、生成することと、

アップサンプリングされたサブピクセル解像度でテンプレート画像内の前記クラスターメタデータを符号化することであって、

前記テンプレート画像のサブピクセル及び前記クラスターの前記画像の前記ピクセルが、同じ画像領域を表す、符号化することと、

前記テンプレート画像に基づいて、前記第2の画像データの前記ピクセルの強度値を修正し、前記クラスターメタデータを考慮する強度分布を有する前記第2の画像データの強度修正バージョンを生成することと、

ニューラルネットワークベースのベースコーラーを介して前記第2の画像データの前記強度修正バージョンを処理し、前記配列決定動作の1つ又はそれ以上の配列決定サイクルで、前記クラスターのうちの1つ又はそれ以上に対するベースコールを生成することであって、前記ニューラルネットワークベースのベースコーラーが、前記ベースコールへの前記クラスターの前記画像のマッピングのタスクで訓練される、生成することと、

を前記システムに実行させる命令を含む非一時的なコンピュータ可読記憶媒体と、を備える、システム。

【請求項7】

前記第2の画像データを前記テンプレート画像に補うことと、

前記ニューラルネットワークベースのベースコーラーを介して、前記テンプレート画像に補われる前記第2の画像データを処理し、前記配列決定動作の1つ又はそれ以上の配列決定サイクルで、前記クラスターのうちの1つ又はそれ以上に対するベースコールを生成することと、を更に含む、請求項6に記載のシステム。

【請求項8】

前記テンプレート画像内の各サブピクセルが、背景サブピクセル、クラスター中心サブピクセル、又はクラスター内部サブピクセルのいずれかとして識別される、請求項6に記載のシステム。

【請求項9】

前記第2の画像データの前記ピクセルの強度値を修正することが、

前記第2の画像データの前記画像内のピクセルに対応する前記テンプレート画像内のどれくらい多くのサブピクセルが、前記クラスターのうちの1つ又はそれ以上の一部を含むかに基づいて、前記第2の画像データ内の1つ又はそれ以上のピクセルに対する領域重み付け係数を計算することと、

前記領域重み付け係数に基づいて、前記ピクセルの強度を修正することと、を含み、

前記第2の画像データの前記ピクセルの強度値を修正することが、クラスター及びクラスターの周囲の背景の前記画像を前記アップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、前記テンプレート画像内の背景サブピクセルに対応する前記アップサンプリングされた画像内のサブピクセルに背景強度を割り当てることと、前記テンプレート画像内のクラスター中心サブピクセル及びクラスター内部サブピクセルに対応する前記アップサンプリングされた画像内のサブピクセルにクラスター強度を割り当てることと、を含み、前記背景強度が、ゼロ値を

10

20

30

40

50

有し、

前記第2の画像データの前記ピクセルの強度値を修正することが、クラスター及びクラスターの周囲の背景の前記画像を前記アップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、前記テンプレート画像内の前記クラスター中心サブピクセル及び前記クラスター内部サブピクセルに対応する前記アップサンプリングされた画像内の前記ピクセルの構成サブピクセルのみの間で、光学ピクセル解像度内のピクセルの全体の強度を分散させることと、を含む、請求項6～8のいずれか一項に記載のシステム。

【請求項10】

前記クラスター強度が、前記光学ピクセル解像度で前記ピクセルの強度を補間することによって決定される、請求項9に記載のシステム。

【請求項11】

前記少なくとも1つのプロセッサによって実行されたときに、アクセスすることであって、

前記1つ又はそれ以上の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、

前記1つ又はそれ以上の配列決定サイクルのうちの前記現在の1つに先行する前記1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び

前記1つ又はそれ以上の配列決定サイクルのうちの前記現在の1つに続く前記1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の後続の画像セット、の各々で、前記光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、第2のニューラルネットワークの出力層を介して、前記1つ又はそれ以上の画像の修正バージョンから画像パッチの畳み込み表現を処理して、画像パッチの中心ピクセルに対して、A、C、T、及びGである、前記1つ又は複数の配列決定サイクルのうちの前記現在の1つで少なくとも1つの特定のクラスターに組み込まれる塩基の尤度を生成することと、前記尤度に基づいて、前記塩基をA、C、T、又はGと分類することと、によって、前記配列決定動作の前記1つ又はそれ以上の配列決定サイクルで、前記クラスターのうちの1つ又はそれ以上に対する前記ベースコールを生成すること、を前記システムに実行させる命令を更に含む、請求項6に記載のシステム。

【請求項12】

前記少なくとも1つのプロセッサによって実行されたときに、

それぞれの前記画像のピクセルに対し、それぞれのピクセルに対する前記テンプレート画像内の領域重み付け係数に基づいてピクセル強度値を修正することと、修正ピクセル強度値を有するピクセルを有する前記画像の各々の修正バージョンを生成することと、

を前記システムに実行させる命令を更に含む、請求項11に記載のシステム。

【請求項13】

前記少なくとも1つのプロセッサによって実行されたときに、

前記少なくとも1つの特定のクラスターに対して、各画像パッチがピクセルのアレイを有し、かつ、前記テンプレート画像で識別される特定のクラスターの中心ピクセルを含むように、各修正バージョンから画像パッチを抽出すること

を前記システムに実行させる命令をさらに含む、請求項12に記載のシステム。

【請求項14】

前記少なくとも1つのプロセッサによって実行されたときに、

前記画像の修正バージョンから抽出される前記画像パッチを前記第2のニューラルネットワークの畳み込みニューラルネットワークを介して畳み込んで、前記画像パッチの畳み込み表現を生成することと

を前記システムに実行させる命令をさらに含む、請求項13に記載のシステム。

【請求項15】

前記少なくとも1つのプロセッサによって実行されたときに、
前記修正ピクセル強度値を生成する前に、サイクル固有及び撮像チャネル固有の変換を使用
して、前記光学ピクセル解像度で捕捉される前記画像の各々を前記テンプレート画像と
位置合わせすること
を前記システムに実行させる命令をさらに含む、請求項12に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

(優先権出願)

本出願は、以下の出願の優先権又は利益を主張する。

【0002】

2019年3月21日に出願された「Training Data Generation for Artificial Intelligence-Based Sequencing」と題する米国仮特許出願第62/821,602号（代理人整理番号ILLM1008-1/IP-1693-PRV）、

【0003】

2019年3月21日に出願された「Artificial Intelligence-Based Generation of Sequencing Metadata」と題する米国仮特許出願第62/821,618号（代理人整理番号ILLM1008-3/IP-1741-PRV）、

【0004】

2019年3月21日に出願された「Artificial Intelligence-Based Base Calling」と題する米国仮特許出願第62/821,681号（代理人整理番号ILLM1008-4/IP-1744-PRV）、

【0005】

2019年3月21日に出願された「Artificial Intelligence-Based Quality Scoring」と題する米国仮特許出願第62/821,724号（代理人整理番号ILLM1008-7/IP-1747-PRV）、

【0006】

2019年3月21日に出願された「Artificial Intelligence-Based Sequencing」と題する米国仮特許出願第62/821,766号（代理人整理番号ILLM1008-9/IP-1752-PRV）、

【0007】

2019年6月14日に出願された「Training Data Generation for Artificial Intelligence-Based Sequencing」と題する蘭国特許出願第2023310号（代理人整理番号ILLM1008-11/IP-1693-NL）、

【0008】

2019年6月14日に出願された「Artificial Intelligence-Based Generation of Sequencing Metadata」と題する蘭国特許出願第2023311号（代理人整理番号ILLM1008-12/IP-1741-NL）、

【0009】

2019年6月14日に出願された「Artificial Intelligence-Based Base Calling」と題する蘭国特許出願第2023312号（代理人整理番号ILLM1008-13/IP-1744-NL）、

【0010】

2019年6月14日に出願された「Artificial Intelligence-Based Quality Scoring」と題する蘭国特許出願第2023314号（代理人整理番号ILLM1008-14/IP-1747-NL）、及び

10

20

30

40

50

【0011】

2019年6月14日に出願された「Artificial Intelligence-Based Sequencing」と題する蘭国特許出願第2023316号（代理人整理番号ILLM1008-15/IP-1752-NL）。

【0012】

2020年3月20日に出願された「Training Data Generation for Artificial Intelligence-Based Sequencing」と題する米国特許出願第16/825,987号（代理人整理番号ILLM1008-16/IP-1693-US）、

【0013】

2020年3月20日に出願された「Training Data Generation for Artificial Intelligence-Based Sequencing」と題する米国特許出願第16/825,991号（代理人整理番号ILLM1008-17/IP-1741-US）、

【0014】

2020年3月20日に出願された「Artificial Intelligence-Based Base Calling」と題する米国特許出願第16/826,126号（代理人整理番号ILLM1008-18/IP-1744-US）、

【0015】

2020年3月20日に出願された「Artificial Intelligence-Based Quality Scoring」と題する米国特許出願第16/826,134号（代理人整理番号ILLM1008-19/IP-1747-US）、

【0016】

2020年3月21日に出願された「Artificial Intelligence-Based Sequencing」と題する米国特許出願第16/826,168号（代理人整理番号ILLM1008-20/IP-1752-PRV）、

【0017】

本願と同時に出願され、その後PCT国際公開第WO _____号として公開されている「Training Data Generation for Artificial Intelligence-Based Sequencing」と題するPCT特許出願第PCT _____号（代理人整理番号ILLM1008-21/IP-1693-PCT）、

【0018】

本願と同時に出願され、その後、PCT国際公開第WO _____号として公開されている「Artificial Intelligence Based Generation of Sequencing Metadata」と題するPCT特許出願第PCT _____号（代理人整理番号ILLM1008-22/IP-1741-PCT）、

【0019】

本願と同時に出願され、その後PCT国際公開第WO _____号として公開されている「Artificial Intelligence-Based Base Calling」と題するPCT特許出願第PCT _____号（代理人整理番号ILLM1008-23/IP-1744-PCT）、及び

【0020】

本願と同時に出願され、その後PCT国際公開第WO _____号として公開されている「Artificial Intelligence-Based Quality Scoring」と題するPCT特許出願第PCT _____号（代理人整理番号ILLM1008-24/IP-1747-PCT）。

【0021】

優先権出願は、本明細書に完全に記載されているかのように、全ての目的のために参照

10

20

30

40

50

により本明細書に組み込まれる。

(組み込み)

【0022】

以下は、本明細書に完全に記載されているかのように、全ての目的のために参照により組み込まれる。

【0023】

2019年5月16日に出願された「Systems and Devices for Characterization and Performance Analysis of Pixel-Based Sequencing」と題する米国仮特許出願第62/849,091号（代理人整理番号ILLM1011-1/IP-1750-PRV）、

10

【0024】

2019年5月16日に出願された「Base Calling Using Convolutions」と題する米国特許仮出願第62/849,132号（代理人整理番号ILLM1011-2/IP-1750-PR2）、

【0025】

2019年5月16日に出願された「Base Calling Using Compact Convolutions」と題する米国仮特許仮出願第62/849,133号（代理人整理番号ILLM1011-3/IP-1750-PR3）、

【0026】

2020年2月20日に出願された「Artificial Intelligence-Based Base Calling of Index Sequences」と題する米国仮特許出願第62/979,384号（代理人整理番号ILLM1015-1/IP-1857-PRV）、

20

【0027】

2020年2月20日に出願された「Artificial Intelligence-Based Many-To-Many Base Calling」と題する米国仮特許出願第62/979,414号（代理人整理番号ILLM1016-1/IP-1858-PRV）、

【0028】

2020年2月20日に出願された「Knowledge Distillation-Based Compression of Artificial Intelligence-Based Base Caller」と題する米国仮特許出願第62/979,385号（代理人整理番号ILLM1017-1/IP-1859-PRV）、

30

【0029】

2020年2月20日に出願された「Multi-Cycle Cluster Based Real Time Analysis System」と題する米国特許仮出願第62/979,412号（代理人整理番号ILLM1020-1/IP-1866-PRV）、

【0030】

2020年2月20日に出願された「Data Compression for Artificial Intelligence-Based Base Calling」と題する米国仮特許出願第62/979,411号（代理人整理番号ILLM1029-1/IP-1964-PRV）、

40

【0031】

2020年2月20日に出願された「Squeezing Layer for Artificial Intelligence-Based Base Calling」と題する米国仮特許出願第62/979,399号（代理人整理番号ILLM1030-1/IP-1982-PRV）、

【0032】

Liu P, Hemani A, Paul K, Weis C, Jung M, Wehn

50

N. 3D-Stacked Many-Core Architecture for Biological Sequence Analysis Problems. Int J Parallel Prog. 2017, 45 (6) :1420-60,

[0033]

Z. Wu, K. Hammad, R. Mittmann, S. Magierowski, E. Ghafar-Zadeh, and X. Zhong, 「FPGA-Based DNA Basecalling Hardware Acceleration」, in Proc. IEEE 61st Int. Midwest Symp. Circuits Syst., Aug. 2018, pp. 1098-1101,

[0034]

Z. Wu, K. Hammad, E. Ghafar-Zadeh, and S. Magierowski, 「FPGA-Accelerated 3rd Generation DNA Sequencing」, in IEEE Transactions on Biomedical Circuits and Systems, Volume 14, Issue 1, Feb. 2020, pp. 65-74,

[0035]

Prabhakar et al., 「Plasticine: A Reconfigurable Architecture for Parallel Patterns」, ISCA'17, June 24-28, 2017, Toronto, ON, Canada,

[0036]

M. Lin, Q. Chen, and S. Yan, 「Network in Network」, in Proc. of ICLR, 2014,

[0037]

L. Sifre, 「Rigid-motion Scattering for Image Classification, Ph. D. thesis, 2014,

[0038]

L. Sifre and S. Mallat, 「Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination」, in Proc. of CVPR, 2013,

[0039]

F. Chollet, 「Xception: Deep Learning with Depthwise Separable Convolutions」, in Proc. of CVPR, 2017,

[0040]

X. Zhang, X. Zhou, M. Lin, and J. Sun, 「Shuffle Net: An Extremely Efficient Convolutional Neural Network for Mobile Devices」, in arXiv:1707.01083, 2017,

[0041]

K. He, X. Zhang, S. Ren, and J. Sun, 「Deep Residual Learning for Image Recognition」, in Proc. of CVPR, 2016,

[0042]

S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, 「Aggregated Residual Transformation For Deep NeuroNetworks」, Proc. of CVPR, 2017,

[0043]

A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko,

10

20

30

40

50

W. Wang, T. Weyand, M. Andreetto, and H. Adam, 「Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications」, in arXiv:1704.04861, 2017、

[0044]

M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, 「MobileNetV2: Inverted Residuals and Linear Bottlenecks」, in arXiv:1801.04381v3, 2018、

[0045]

Z. Qin, Z. Zhang, X. Chen and Y. Peng, 「FD-MobileNet: Improved MobileNet with a Fast Downsampling Strategy」, in arXiv:1802.03750, 2018、

[0046]

Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. CoRR, abs/1706.05588, 2017、

[0047]

J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. arXiv preprint arXiv:1611.10012, 2016、

[0048]

S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, 「WAVENET: A GENERATIVE MODEL FOR RAW AUDIO」, arXiv:1609.03499, 2016、

[0049]

S. O. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, S. Sengupta and M. Shoeybi, 「DEEP VOICE: REAL-TIME NEURAL TEXT-TO-SPEECH」, arXiv:1702.07825, 2017、

[0050]

F. Yu and V. Koltun, 「MULTI-SCALE CONTEXT AGGREGATION BY DILATED CONVOLUTIONS」, arXiv:1511.07122, 2016、

[0051]

K. He, X. Zhang, S. Ren, and J. Sun, 「DEEP RESIDUAL LEARNING FOR IMAGE RECOGNITION」, arXiv:1512.03385, 2015、

[0052]

R. K. Srivastava, K. Greff, and J. Schmidhuber, 「HIGHWAY NETWORKS」, arXiv:1505.00387, 2015、

[0053]

G. Huang, Z. Liu, L. van der Maaten and K. Q. W

10

20

30

40

50

einberger, 「DENTILY CONNECTED CONVOLUTIONAL NETWORKS」、arXiv:1608.06993, 2017、

[0054]

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, 「GOING DEEPER WITH CONVOLUTIONS」、arXiv:1409.4842, 2014、

[0055]

S. Ioffe and C. Szegedy, 「BATCH NORMALIZATION: ACCELERATING DEEP NETWORK TRAINING BY REDUCING INTERNAL COVARIATE SHIFT」、arXiv:1502.03167, 2015、

[0056]

J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, 「DILATED CONVOLUTIONAL NEURAL NETWORKS FOR CARDIOVASCULAR MR SEGMENTATION IN CONGENITAL HEART DISEASE」、arXiv:1704.03669, 2017、

[0057]

L. C. Piqueras, 「AUTOREGRESSIVE MODEL BASED ON A DEEP CONVOLUTIONAL NEURAL NETWORK FOR AUDIO GENERATION」、Tampere University of Technology, 2016、

[0058]

J. Wu, 「Introduction to Convolutional Neural Networks」、Nanjing University, 2017、

[0059]

「Illumina CMOS Chip and One-Channel SBS Chemistry」、Illumina, Inc. 2018, 2 pages、

[0060]

「skikit-image/peak.py at master」、GitHub, 5 pages, [2018-11-16に検索]。インターネット<URL:<https://github.com/scikit-image/scikit-image/blob/master/skimage/feature/peak.py#L25>>から検索、

[0061]

「3.3.9.11. Watershed and random walker for segmentation」、Scipy lecture notes, 2 pages, [2018-11-13に検索]。インターネット<URL:http://scipy-lectures.org/packages/scikit-image/uto_examples/plot_segmentations.html>から検索、

[0062]

Mordvintsev, Alexander and Revision, Abid K., 「Image Segmentation with Watershed Algorithm」、Revision 43532856, 2013, 6 pages [2018-11-13に検索]。インターネット<URL:https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_watershed/py_watershed.html>から検索、

10

20

30

40

50

【0063】

Mzur, 「Watershed.py」, 25 October 2017, 3 pages, [2018-11-13に検索]。インターネット<URL:<https://github.com/mzur/watershed/blob/master/Watershed.py>>から検索、

【0064】

Thakur, Pratibha, et. al. 「A Survey of Image Segmentation Techniques」、International Journal of Research in Computer Applications and Robotics, Vol. 2, Issue. 4, April 2014, Pg. :158-165、

10

【0065】

Long, Jonathan, et. al. , 「Fully Convolutional Networks for Semantic Segmentation」、:IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 39, Issue 4, 1 April 2017, 10 pages、

【0066】

Ronneberger, Olaf, et. al. , 「U-net:Convolutional networks for biomedical image segmentation」。In International Conference on Medical image computing and computer-assisted intervention, 18 May 2015, 8 pages、

20

【0067】

Xie, W. , et. al. , 「Microscopy cell counting and detection with fully convolutional regression networks」, Computer methods in biomechanics and biomedical engineering:Imaging&Visualization, 6 (3) , pp. 283-292, 2018、

【0068】

Xie, Yuanpu, et al. , 「Beyond classification :structured regression for robust cell detection using convolutional neural network」, International Conference on Medical Image Computing and Computer-Assisted Intervention. October 2015, 12 pages、

30

【0069】

Snuverink, I. A. F. , 「Deep Learning for Pixelwise Classification of Hyperspectral Images」、Master of Science Thesis, Delft University of Technology, 23 November 2017、19 pages、

40

【0070】

Shevchenko, A. , 「Keras weighted categorical_crossentropy」, 1 page、[2019-01-15に検索]。インターネット<URL:<https://gist.github.com/skeeeet/cad06d584548fb45eece1d4e28cfa98b>>から検索、

【0071】

van den Assem, D. C. F. , 「Predicting periodic And chaotic signals using Wavenets」、Mas

50

ter of Science Thesis, Delft University Of Technology, 18 August 2017, Pages 3-38、

【0072】

I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, 「CONVOLUTIONAL NETWORKS」、Deep Learning, MIT Press, 2016、及び

【0073】

J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahrourdy, B. Shuai, T. Liu, X. Wang, and G. Wang, 「RECENT ADVANCES IN CONVOLUTIONAL NEURAL NETWORKS」、arXiv:1512.07108, 2017。

【0074】

(発明の分野)

本開示の技術は、人工知能コンピュータ及びデジタルデータ処理システム、並びに、知能をエミュレーションするための対応するデータ処理方法及び製品（すなわち、知識ベースのシステム、推測システム及び知識取得システム）に関すると共に、不確実性（例えば、ファジー論理システム）、適応システム、機械学習システム、及び人工ニューラルネットワークを用いて推測するためのシステムを含む。具体的には、開示される技術は、データを分析するための深層畳み込みニューラルネットワークなどの深層ニューラルネットワークを使用することに関する。

【背景技術】

【0075】

このセクションで説明される主題は、単にこのセクションにおける言及の結果として、先行技術であると想定されるべきではない。同様に、このセクションで言及した問題、又は背景として提供された主題に関連する問題は、先行技術において以前に認識されていると想定されるべきではない。このセクションの主題は、単に、異なるアプローチを表し、それ自体はまた、特許請求される技術の実施に対応し得る。

【0076】

深層ニューラルネットワークは、高レベル機能を連続的にモデル化するために、複数の非線形及び複雑な変換層を使用する、人工ニューラルネットワークの類である。深層ニューラルネットワークは、観測された出力と予測出力との間の差を伝達してパラメータを調整する逆伝搬を介してフィードバックを提供する。深層ニューラルネットワークは、大きな訓練データセットの利用可能性、並列分散コンピューティングの能力、及び高度な訓練アルゴリズムと共に進化している。深層ニューラルネットワークは、コンピュータビジョン、音声認識、及び自然言語処理などの多数のドメインにおける主な進歩を促進している。

【0077】

畳み込みニューラルネットワーク (Convolutional neural network、CNN) 及び反復ニューラルネットワーク (Recurrent Neural Network、RNN) は、深層ニューラルネットワークの構成要素である。畳み込みニューラルネットワークは、特に、畳み込み層、非線形層、及びプーリング層を含む構造で画像認識に成功してきた。反復ニューラルネットワークは、パーセプトロン、長い短期メモリユニット、及びゲートされた反復単位のような構成単位間の周期的接続を有する入力データの連続的な情報を利用するように設計される。加えて、多くの他の出現深層ニューラルネットワークが、深層時空間ニューラルネットワーク、多次元反復ニューラルネットワーク、及び畳み込み自動エンコーダなどの限定された状況に関して提案されてきた。

【0078】

深層ニューラルネットワークの訓練の目標は、各層における重みパラメータの最適化であり、このことは、より好適な階層表現がデータから学習され得るように、より単純な特徴を複雑な特徴に徐々に組み合わせる。最適化プロセスの単一サイクルは、以下のように構成される。まず、訓練データセットと仮定すると、前方へのパスは、各層内の出力を順

10

20

30

40

50

次計算し、ネットワークを通して機能信号を順方向に伝搬する。最終出力層において、客観的な損失関数は、推測された出力と所与のラベルとの間の誤差を測定する。訓練誤差を最小化するために、後方へのパスは、連鎖ルールを使用して、誤差信号を逆伝搬し、ニューラルネットワーク全体の全ての重みに対する勾配を計算する。最後に、確率勾配降下に基づいて最適化アルゴリズムを使用して確率パラメータが更新される。バッチ勾配降下が完全データセットごとにパラメータ更新するのに対し、確率的勾配降下は、データ例の各々の小さいセットについて更新を実行することによって確率的近似値を提供する。いくつかの最適化アルゴリズムは確率的勾配降下に由来する。例えば、Adagrad and Adam訓練アルゴリズムは、それぞれ、各パラメータの更新頻度及び勾配のモーメントに基づいて学習率を適応的に修正しながら、確率的勾配降下を実行する。

10

【0079】

深層ニューラルネットワークの訓練における別のコア要素は規則化であり、規則化は、過剰適合を回避し、したがって良好な一般化性能を達成することを意図する戦略を指す。例えば、重み減衰は、重みパラメータがより小さい絶対値に収束するように、客観的損失関数にペナルティ項を追加する。ドロップアウトは、訓練中にニューラルネットワークから隠れたユニットをランダムに除去し、可能なサブネットワークの集合体と見なすことができる。ドロップアウトの能力を向上させるために、新たな起動関数、maxout、及びrnnDropと呼ばれる反復性ニューラルネットワークに対するドロップアウトのバリエーションが提案されている。更に、バッチ正規化は、ミニバッチ内の各アクティブ化に関するスカラー特徴の正規化を介した新たな規則化方法を提供し、各々の平均及び分散をパラメータとして学習する。

20

【0080】

配列データが多次元及び高次元であると仮定すると、深層ニューラルネットワークは、それらの広範な適用性及び強化された予測能力により、生物情報科学研究のためにかなり有望である。畳み込みニューラルネットワークは、モチーフ発見、病原性変異体識別、及び遺伝子発現推測などのゲノミクスにおける配列に基づく問題を解決するために採用されている。畳み込みニューラルネットワークは、DNAを研究するのに特に有用な重み共有戦略を使用するが、これは、短い配列モチーフを捕捉することができ、この配列モチーフは、有意な生物学的機能を有すると推定されるDNA中の局所的パターンを再現する。畳み込みニューラルネットワークの顕著な特徴は、畳み込みフィルタの使用である。

30

【0081】

精巧に設計され、手動で巧妙に作り上げられた特徴に基づく従来の分類アプローチとは異なり、畳み込みフィルタは、知識の情報表現に生入力データをマッピングするプロセスに類似した特徴の適応学習を実行する。この意味では、畳み込みフィルタは、そのようなフィルタのセットが入力内の関連するパターンを認識し、訓練手順中にそれ自体を更新することができるため、一連のモチーフスキャナーとして機能する。反復ニューラルネットワークは、タンパク質又はDNA配列などの様々な長さの連続的データにおける長距離依存性を捉えることができる。

【0082】

したがって、テンプレートの生成及びベースコールのための、理にかなった深層学習ベースの枠組みを使用する機会が生じる。

40

【0083】

ハイスループット技術の時代では、努力ごとに最も低いコストで解釈可能なデータの最高収率を蓄積することは、重大な課題を残している。クラスター形成のためにブリッジ増幅を利用するものなどの核酸配列決定のクラスターベースの方法は、核酸配列決定のスループットを増加させる目的に有益な貢献をしている。これらのクラスターベースの方法は、固体支持体上に固定化された核酸の密集した集団を配列決定することに依存し、典型的には、固体支持体上の別個の場所に位置する複数のクラスターを同時に配列決定する過程で生成される光信号を抑制するための画像解析ソフトウェアの使用を伴う。

【0084】

50

しかしながら、このような固相核酸クラスターベースの配列決定技術は、達成することができるスループットの量を制限する相当な障害に直面している。例えば、クラスターベースの配列決定方法では、空間的に分解されるには互いに物理的に近接し過ぎる、又は実際には、固体支持体上で物理的に重なり合う2つ又はそれ以上のクラスターの核酸配列を決定することは障害物をもたらす可能性がある。例えば、現在の画像解析ソフトウェアは、2つの重複クラスターのうちのどれから光信号が発せられたかを判定するための貴重な時間及び計算リソースを必要とする場合がある。結果として、得られ得る核酸配列情報の量及び／又は質に関して、様々な検出プラットフォームにとって妥協が不可避である。

【0085】

高密度核酸凝集体ベースのゲノミクス法は、ゲノム分析の他の領域にも同様に延在する。例えば、核酸クラスターベースのゲノミクスは、配列決定用途、診断及びスクリーニング、遺伝子発現分析、エピジェネティクス分析、多型の遺伝分析などに使用することができる。これらの核酸クラスターベースのゲノミクス技術のそれぞれは、厳密に近接して又は空間的に重複する核酸クラスターから生成されたデータを分解する能力がない場合に制限される。

【0086】

明らかに、ゲノミクス（例えば、任意の及び全ての動物、植物、微生物又は他の生物学的種又は集団のゲノム特性評価のための）、薬理ゲノミクス、トランスクリプトミクス、診断、予後、生物医学的リスク評価、臨床及び研究遺伝学、個人向け医療、薬物有効性及び薬物相互作用評価、獣医医学、農業、進化、及び生物学的研究、水性培養、林業、海洋調査、生態学的管理、及び環境管理、並びに他の目的を含む、様々な用途で迅速かつ費用効果の高い方法で取得できる核酸配列データの質と量を向上させる必要がある。

【0087】

開示される技術は、ハイスループット核酸配列決定技術におけるスループットのレベルを増加させることを含めて、これら及び同様のニーズに対処するニューラルネットワークベースの方法及びシステムを提供すると共に、他の関連する利点を提供する。

【0088】

特許又は出願ファイルは、カラーで創作された少なくとも1つの図面を含む。カラー図面（単数又は複数）を有するこの特許又は特許出願公開のコピーは、必要な料金の要求及び支払いの際に、庁によって提供される。カラー図面はまた、補足コンテンツタブを介してPAIR（patent application information retrieval：特許出願情報検索）で利用可能であってもよい。

【0089】

図面では、同様の参照文字は、概して、異なる図全体を通して同様の部分を指す。また、図面は必ずしも縮尺通りではなく、その代わりに、開示された技術の原理を例示することを強調している。以下の説明では、開示される技術の様々な実施態様が、以下の図面を参照して説明される。

【図面の簡単な説明】

【0090】

【図1】サブピクセルベースコールを使用してクラスターメタデータを決定する処理パイプラインの一実施態様を示す。

【図2】そのタイル内にクラスターを含むフローセルの一実施態様を示す。

【図3】8つのレーンを有するIllumina GA-EI x フローセルの一例を示す。

【図4】4チャネル化学の配列決定画像の画像セットを描写しており、すなわち、画像セットは、ピクセルドメイン内の4つの異なる波長帯域（画像／撮像チャネル）を使用して捕捉された4つの配列決定画像を有する。

【図5】配列決定画像をサブピクセル（又はサブピクセル領域）に分割する一実施態様である。

【図6】サブピクセルベースコール中に、ベースコーラーによって識別されたクラスターの予備中心座標を示す。

10

20

30

40

50

【図 7】クラスターメタデータを含むいわゆる「クラスターマップ」を生成するために、複数の配列決定サイクルにわたって生成されたサブピクセルベースコールをマージする一実施態様を示す。

【図 8 a】サブピクセルベースコールのマージによって生成されたクラスターマップの一例を示す。

【図 8 b】サブピクセルベースコールの一実施態様を示す。

【図 9】クラスターメタデータを識別するクラスターマップの別の例を示す。

【図 10】クラスターマップ内の不連続領域の質量中心 (Center Of Mass、COM) がどのように計算されるかを示す。

【図 11】不連続領域のサブピクセルから不連続領域の COM までのユークリッド距離に基づく加重減衰係数の計算の一実施態様を示す。

10

【図 12】サブピクセルベースコールによって生成された例示的なクラスターマップから導出された、例示的なグラウンドトゥルス減衰マップの一実施態様を示す。

【図 13】クラスターマップから三元マップを導出する一実施態様を示す。

【図 14】クラスターマップからバイナリマップを導出する一実施態様を示す。

【図 15】ニューラルネットワークベースのテンプレート生成器及びニューラルネットワークベースのベースコーラーを訓練するために使用される訓練データを生成する一実施態様を示すブロック図である。

【図 16】ニューラルネットワークベースのテンプレート生成器及びニューラルネットワークベースのベースコーラーを訓練するために使用される、開示された訓練例の特性を示す。

20

【図 17】開示されたニューラルネットワークベースのテンプレート生成器を介して入力画像データを処理し、アレイ内の各ユニットの出力値を生成する一実施態様を示す。一実施態様では、アレイは減衰マップである。別の実施態様では、アレイは三元マップである。更に別の実施態様では、アレイはバイナリマップである。

【図 18】クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を含むクラスターメタデータを導出するためにニューラルネットワークに基づくテンプレート生成器によって生成された、減衰マップ、三元マップ、又はバイナリマップに適用される後処理技術の一実施態様を示す。

【図 19】ピクセルドメイン内のクラスター強度を抽出する一実施態様を示す。

30

【図 20】サブピクセルドメイン内のクラスター強度を抽出する一実施態様を示す。

【図 21 a】ニューラルネットワークベースのテンプレート生成器の 3 つの異なる実施態様を示す。

【図 21 b】ニューラルネットワークベースのテンプレート生成器 1512 への入力として供給される入力画像データの一実施態様を示す。入力画像データは、配列決定動作の特定の数の初期配列サイクルの間に生成される配列決定画像を有する一連の画像セットを含む。

【図 22】入力画像データを形成する一連の「ダウンサイズの」画像セットを生成するために、図 21 b の一連の画像セットからパッチを抽出する一実施態様を示す。

【図 23】入力画像データを形成する一連の「アップサンプリングされた」画像セットを生成するために、図 21 b の一連の画像セットをアップサンプリングする一実施態様を示す。

40

【図 24】入力画像データを形成する一連の「アップサンプリング及びダウンサイズの」画像セットを生成するために、図 23 の一連のアップサンプリングされた画像セットからパッチを抽出する一実施態様を示す。

【図 25】ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥルスデータを生成する、全体的な例示のプロセスの一実施態様を示す。

【図 26】回帰モデルの一実施態様を示す。

【図 27】クラスターマップからグラウンドトゥルス減衰マップを生成する一実施態様を示す。グラウンドトゥルス減衰マップは、回帰モデルを訓練するためのグラウンドト

50

ウルスデータとして使用される。

【図 28】逆伝搬ベースの勾配更新技術を使用して回帰モデルを訓練する一実施態様である。

【図 29】推測中の回帰モデルによるテンプレート生成の一実施態様である。

【図 30】クラスターメタデータを識別するために、減衰マップを後処理に供する一実施態様を示す。

【図 31】クラスターを特徴付ける隣接するクラスター／クラスター内部サブピクセルの非重複グループを識別する、流域分割技術の一実施態様を示す。

【図 32】回帰モデルの例示的な U-Net 構造を示す表である。

【図 33】テンプレート画像内で識別されたクラスター形状情報を使用してクラスター強度を抽出する異なるアプローチを示す。

【図 34】回帰モデルの出力を使用したベースコールの異なるアプローチを示す。

【図 35】クラスター中心として非 COM 位置を使用することとは対照的に、RTA ベースコーラーがクラスター中心としてグラウンドトゥルス質量中心 (COM) 位置を使用するときのベースコール性能の差を示す。結果は、COM を使用することによりベースコールが改善されることを示す。

【図 36】左側に、回帰モデルを生成した減衰マップ例を示す。図 36 はまた、右側に、訓練中に回帰モデルが近位になる、例示的なグラウンドトゥルス減衰マップを示す。

【図 37】ピークを検出することによって減衰マップ内のクラスター中心を識別するピーククロケータの一実施態様を示す。

【図 38】回帰モデルによって生成された減衰マップ内のピーククロケータによって検出されたピークを、対応するグラウンドトゥルス減衰マップ内のピークと比較する。

【図 39】適合率と再現率の統計を使用して回帰モデルの性能を示す。

【図 40】20 pM のライブラリ濃度 (通常動作) について、RTA ベースコーラーと回帰モデルの性能とを比較する。

【図 41】30 pM のライブラリ濃度 (高密度動作) について、RTA ベースコーラーと回帰モデルの性能とを比較する。

【図 42】重複していない適切なリード対の数、すなわち、どちらのリードも回帰モデルによって検出された妥当な距離内で内側に位置合わせされていない対のリードの数を、RTA ベースコーラーによって検出されたものと比較する。

【図 43】回帰モデルによって生成された第 1 の減衰マップを右側に示す。左側では、図 43 は、回帰モデルによって生成された第 2 の減衰マップを示す。

【図 44】40 pM ライブラリ濃度 (高密度動作) について、RTA ベースコーラーと回帰モデルの性能とを比較する。

【図 45】回帰モデルによって生成された第 1 の減衰マップを左側に示す。右側では、図 45 は、第 1 の減衰マップに適用された閾値化、ピーク位置処理、及び流域分割技術の結果を示す。

【図 46】バイナリ分類モデルの一実施態様を示す。

【図 47】ソフトマックススコアを伴う逆伝搬ベースの勾配更新技術を使用してバイナリ分類モデルを訓練する一実施態様である。

【図 48】シグモイドスコアを伴う逆伝搬ベースの勾配更新技術を使用してバイナリ分類モデルを訓練する別の実施態様である。

【図 49】バイナリ分類モデルに供給された入力画像データ及びバイナリ分類モデルを訓練するために使用される対応するクラスラベルの別の実施態様を示す。

【図 50】推測中のバイナリ分類モデルによるテンプレート生成の一実施態様である。

【図 51】クラスター中心を識別するために、バイナリマップをピーク検出に供する一実施態様を示す。

【図 52 a】バイナリ分類モデルによって生成された例示的なバイナリマップを左側に示す。図 52 a はまた、右側に、訓練中にバイナリ分類モデルが近位になる、例示的なグラウンドトゥルスバイナリマップを示す。

10

20

30

40

50

【図 5 2 b】精度統計を使用してバイナリ分類モデルの性能を示す。

【図 5 3】バイナリ分類モデルの例示的な構造を示す表である。

【図 5 4】三元分類モデルの一実施態様を示す。

【図 5 5】逆伝搬ベースの勾配更新技術を使用して三元分類モデルを訓練する一実施態様である。

【図 5 6】三元分類モデルに供給された入力画像データ及び三元分類モデルを訓練するために使用される対応するクラスラベルの別の実施態様を示す。

【図 5 7】三元分類モデルの例示的な構造を示す表である。

【図 5 8】推測中の三元分類モデルによるテンプレート生成の一実施態様である。

【図 5 9】三元分類モデルによって生成された三元マップを示す。

10

【図 6 0】ユニットごとの出力値と共に三元分類モデル 5 4 0 0 によって生成されたユニット配列を示す。

【図 6 1】クラスター中心、クラスター背景、及びクラスター内部を識別するために、三元マップを後処理に供する一実施態様を示す。

【図 6 2 a】三元分類モデルの例示的予測を示す。

【図 6 2 b】三元分類モデルの他の例示的予測を示す。

【図 6 2 c】三元分類モデルの更に他の例示的予測を示す。

【図 6 3】図 6 2 a の三元分類モデルの出力からクラスター中心及びクラスター形状を導出する一実施態様を示す。

【図 6 4】バイナリ分類モデル、回帰モデル、及び R T A ベースコーラーのベースコール性能を比較する。

20

【図 6 5】3つの状況、5つの配列決定メトリック、及び2つの動作密度の下で、三元分類モデルの性能を R T A ベースコーラーの性能と比較する。

【図 6 6】図 6 5 で考察される3つの状況、5つの配列決定メトリック、及び2つの動作密度の下で、回帰モデルの性能を R T A ベースコーラーの性能と比較する。

【図 6 7】ニューラルネットワークベースのテンプレート生成器の最後から2番目の層に焦点を当てている。

【図 6 8】ニューラルネットワークベースのテンプレート生成器の最後から2番目の層が、逆伝搬ベースの勾配更新訓練の結果として学習したものを可視化する。図示された実施態様は、図 6 7 に示される最後から2番目の層の32個の訓練された畳み込みフィルタから24を可視化する。

30

【図 6 9】（青色での）バイナリ分類モデルのクラスター中心予測を、（ピンク色での）R T A ベースコーラーに重ね合わせる。

【図 7 0】バイナリ分類モデルの最後から2番目の層の訓練された畳み込みフィルタの可視化上に、（ピンク色で）R T A ベースのカラー（ピンク色で）によって作製されたクラスター中心予測を重ね合わせる。

【図 7 1】ニューラルネットワークベースのテンプレート生成器を訓練するために使用される訓練データの一実施態様を示す。

【図 7 2】ニューラルネットワークベースのテンプレート生成器のクラスター中心予測に基づいて画像位置合わせ用のビーズを使用する一実施態様である。

40

【図 7 3】ニューラルネットワークベースのテンプレート生成器によって識別されたクラスターのクラスター統計の一実施態様を示す。

【図 7 4】入力画像データが使用される初期配列決定サイクルの数が5から7に増加すると、ニューラルネットワークベースのテンプレート生成器が隣接するクラスター間を区別する能力がどのように改善されるかを示す。

【図 7 5】非 C O M 位置がクラスター中心として使用されるときとは対照的に、R T A ベースコーラーがクラスター中心としてグラウンドトゥールズ質量中心（C O M）位置を使用するときのベースコール性能の差を示す。

【図 7 6】追加で検出されたクラスターに関するニューラルネットワークベースのテンプレート生成器の性能を示す。

50

【図 7 7】ニューラルネットワークベースのテンプレート生成器を訓練するために使用される異なるデータセットを示す。

【図 7 8】一実施態様に係る、ベースコールのために R T A ベースコーラーによって使用される処理段階を示す。

【図 7 9】開示されたニューラルネットワークベースのベースコーラーを使用するベースコールの一実施態様を示す。

【図 8 0】サブピクセルドメインからピクセルドメインに、ニューラルネットワークベースのテンプレート生成器の出力から識別されるクラスター中心の場所／位置情報を変換する一実施態様である。

【図 8 1】参照クラスター中心からいわゆる「変換されたクラスター中心」を導出するために、サイクル固有及び画像チャンネル固有の変換を使用する一実施態様である。

【図 8 2】ニューラルネットワークベースのベースコーラーに供給される入力データの一部である画像パッチを示す。

【図 8 3】単一のターゲットクラスターが、ニューラルネットワークベースのベースコーラーによってベースコールされているときに、距離チャンネルの距離値を決定する一実施態様を示す。

【図 8 4】ピクセルとターゲットクラスターとの間で計算される距離値を符号化するピクセルごとの一実施態様を示す。

【図 8 5 a】複数のターゲットクラスターが、ニューラルネットワークベースのベースコーラーによって同時にベースコールされているときに、距離チャンネルの距離値を決定する一実施態様を示す。

【図 8 5 b】ターゲットクラスターの各々に対して、ピクセル中心に最も近いクラスター中心間距離に基づいて決定されるいくつかの最も近いピクセルを示す。

【図 8 6】ピクセルとクラスターのうちの最も近い 1 つとの間で計算される最小距離値を符号化するピクセルごとの一実施態様を示す。

【図 8 7】本明細書では「クラスター形状データ」と称される、ピクセルクラスター間の分類／属性／分類を使用する一実施態様を示す。

【図 8 8】クラスター形状データを使用して距離値を計算する一実施態様を示す。

【図 8 9】ピクセルと割り当てられたクラスターとの間で計算される距離値を符号化するピクセルごとの一実施態様を示す。

【図 9 0】異なる配列決定サイクルでデータの処理を分離するために使用されるニューラルネットワークベースのベースコーラーの専用構造の一実施態様を示す。

【図 9 1】分離された畳み込みの一実施態様を示す。

【図 9 2 a】組み合わせの畳み込みの一実施態様を示す。

【図 9 2 b】組み合わせの畳み込みの別の実施態様を示す。

【図 9 3】各畳み込み層が畳み込みフィルタのバンクを有する、ニューラルネットワークベースのベースコーラーの畳み込み層の一実施態様を示す。

【図 9 4】画像チャンネルを補うスケールリングチャンネルの 2 つの構成を示す。

【図 9 5 a】赤色画像及び緑色画像を生成する単一の配列決定サイクルの入力データの一実施態様を示す。

【図 9 5 b】画像チャンネルから生成される特徴マップに組み込まれる加法バイアスを供給する距離チャンネルの一実施態様を示す。

【図 9 6 a】単一のターゲットクラスターをベースコールする一実施態様を示す。

【図 9 6 b】単一のターゲットクラスターをベースコールする一実施態様を示す。

【図 9 6 c】単一のターゲットクラスターをベースコールする一実施態様を示す。

【図 9 7】複数のターゲットクラスターを同時にベースコールする一実施態様を示す。

【図 9 8】複数の連続する配列決定サイクルで複数のターゲットクラスターを同時にベースコールし、それによって、複数のターゲットクラスターの各々に対するベースコール配列を同時に生成する、一実施態様を示す。

【図 9 9】単一のクラスターベースコール実施態様のための次元性図を示す。

10

20

30

40

50

【図100】複数のクラスターにおける次元性図、単一の配列決定サイクルベースコール実施態様を示す。

【図101】複数のクラスターにおける次元性図、複数の配列決定サイクルベースコール実施態様を示す。

【図102a】多重サイクル入力データの例示的アレイ入力構成を示す。

【図102b】多重サイクル入力データの例示的スタック入力構成を示す。

【図103a】中心ピクセルでベースコールされているターゲットクラスターの中心を中心にするために画像パッチのピクセルを再構成する一実施態様を示す。

【図103b】(i) 中心ピクセルの中心がターゲットクラスターの中心と一致し、(ii) 非中心ピクセルがターゲットクラスターの中心から等距離である、別の例示的再構成／シフトされた画像パッチを示す。

10

【図104】標準的な畳み込みニューラルネットワーク及び再構成された入力を使用して、現在の配列決定サイクルで単一のターゲットクラスターをベースコールする一実施態様を示す。

【図105】標準的な畳み込みニューラルネットワーク及び位置合わせされた入力を使用して、現在の配列決定サイクルで複数のターゲットクラスターをベースコールする一実施態様を示す。

【図106】標準的な畳み込みニューラルネットワーク及び位置合わせされた入力を使用して、複数の配列決定サイクルで複数のターゲットクラスターをベースコールする一実施態様を示す。

20

【図107】ニューラルネットワークベースのベースコーラーを訓練する一実施態様を示す。

【図108a】ニューラルネットワークベースのベースコーラーとして使用されるハイブリッドニューラルネットワークの一実施態様を示す。

【図108b】現在の隠れ状態表現を生成するためにハイブリッドニューラルネットワークの反復モジュールによって使用される3D畳み込みの一実施態様を示す。

【図109】畳み込みモジュールの畳み込み層のカスケードを介して、ベースコールされる一連のt回の配列決定サイクル間の単一の配列決定サイクルのサイクルごとの入力データを処理する一実施態様を示す。

【図110】単一の配列決定サイクルのサイクルごとの入力データを、畳み込みモジュールの畳み込み層のカスケードによって生成される、その対応する畳み込み表現と混合する一実施態様を示す。

30

【図111】連続する配列決定サイクルの平坦化された混合表現をスタックとして配置する一実施態様を示す。

【図112a】図111のスタックに、前方及び後方の方向に3D畳み込みの反復適用を行い、一連のt回の配列決定サイクルの各々でクラスターの各々に対するベースコールを生成する一実施態様を示す。

【図112b】3D畳み込みを適用する長い短期メモリ(Long Short-Term Memory、LSTM)ネットワークの入力ゲート、活性化ゲート、忘却ゲート、及び出力ゲートを介して、平坦化された混合表現のグループを含む3D入力容積 $x(t)$ を処理する一実施態様を示す。LSTMネットワークは、ハイブリッドニューラルネットワークの反復モジュールの一部である。

40

【図113】ニューラルネットワークベースのベースコーラーを訓練するために使用される訓練データでのトリヌクレオチド(3量体)をバランスさせる一実施態様を示す。

【図114】ニューラルネットワークベースのベースコーラーに対して、RTAベースコーラーのベースコール精度を比較する。

【図115】RTAベースコーラーのタイル間の一般化を、同じタイル上のニューラルネットワークベースのベースコーラーのものと比較する。

【図116】RTAベースコーラーのタイル間の一般化を、同じタイル上及び異なるタイル上のニューラルネットワークベースのベースコーラーのものと比較する。

50

【図117】また、RTAベースコーラーのタイル間の一般化を異なるタイル上のニューラルネットワークベースのベースコーラーのものと比較する。

【図118】ニューラルネットワークベースのベースコーラーへの入力として供給される画像パッチの異なるサイズが、どのようにベースコール精度をもたらすかを示す。

【図119】A. baumannii及びE. coliからの訓練データにおけるニューラルネットワークベースのベースコーラーのレーン間の一般化を示す。

【図120】A. baumannii及びE. coliからの訓練データにおけるニューラルネットワークベースのベースコーラーのレーン間の一般化を示す。

【図121】A. baumannii及びE. coliからの訓練データにおけるニューラルネットワークベースのベースコーラーのレーン間の一般化を示す。

【図122】A. baumannii及びE. coliからの訓練データにおけるニューラルネットワークベースのベースコーラーのレーン間の一般化を示す。

【図123】図119、図120、図121、及び図122に関して上述したレーン間の一般化のための誤差プロファイルを示す。

【図124】図123の誤差プロファイルによって検出される誤差のソースを、緑色チャネル内の低クラスター強度に属させる。

【図125】2回の配列決定動作（リード1及びリード2）について、RTAベースコーラー及びニューラルネットワークベースのベースコーラーの誤差プロファイルを比較する。

【図126a】4つの異なる器具におけるニューラルネットワークベースのベースコーラーの動作間の一般化を示す。

【図126b】同じ器具上で実行される4つの異なる動作におけるニューラルネットワークベースのベースコーラーの動作間の一般化を示す。

【図127】ニューラルネットワークベースのベースコーラーを訓練するために使用される訓練データのゲノム統計を示す。

【図128】ニューラルネットワークベースのベースコーラーを訓練するために使用される訓練データのゲノムコンテキストを示す。

【図129】長いリード（例えば、 2×250 ）をベースコールする際のニューラルネットワークベースのベースコーラーのベースコール精度を示す。

【図130】ニューラルネットワークベースのベースコーラーが、画像パッチにわたる中心クラスターピクセル（複数可）及びその隣接するピクセルにどのように対処するかの一実施態様を示す。

【図131】一実施態様に係る、ニューラルネットワークベースのベースコーラーを訓練及び動作するために使用される様々なハードウェア構成要素及び構成を示す。他の実施態様では、異なるハードウェア構成要素及び構成が使用される。

【図132】ニューラルネットワークベースのベースコーラーを使用して実行され得る様々な配列決定タスクを示す。

【図133】t分布型確率的近傍埋め込み（t-Distributed Stochastic Neighbor Embedding、t-SNE）によって可視化される散乱プロットであり、ニューラルネットワークベースのベースコーラーのベースコール結果を示す。

【図134】品質スコアリングのためにニューラルネットワークベースのベースコーラーによって作成されるベースコール信頼性確率を選択する一実施態様を示す。

【図135】ニューラルネットワークベースの品質スコアリングの一実施態様を示す。

【図136a】ニューラルネットワークベースのベースコーラーによって作成される、品質スコアとベースコール信頼性予測との間の対応の一実施態様を示す。

【図136b】ニューラルネットワークベースのベースコーラーによって作成される、品質スコアとベースコール信頼性予測との間の対応の一実施態様を示す。

【図137】推測中にニューラルネットワークベースのベースコーラーによって作成されるベースコール信頼性予測から品質スコアを推測する一実施態様を示す。

【図138】ニューラルネットワークベースの品質スコアラーを訓練して、配列決定画像から導出される入力データを処理し、品質指標を直接生成する、一実施態様を示す。

10

20

30

40

50

【図139】推測中にニューラルネットワークベースの品質スコアラーの出力として品質指標を直接生成する一実施態様を示す。

【図140】ニューラルネットワークベースのテンプレート生成器、ニューラルネットワークベースのベースコーラー、及びニューラルネットワークベースの品質スコアラーへの入力として供給され得る変換されたデータを生成するために、無損失変換を使用する一実施態様を示す。

【図141】領域重み付け係数を使用して、ニューラルネットワークベースのテンプレート生成器をニューラルネットワークベースのベースコーラーと統合する一実施態様を示す。

【図142】アップサンプリング及び背景マスキングを使用して、ニューラルネットワークベースのテンプレート生成器をニューラルネットワークベースのベースコーラーと統合する別の実施態様を示す。

10

【図143】1ピクセル当たり単一のクラスターのみからの寄与に対する領域重み付け係数14300の一例を示す。

【図144】1ピクセル当たり複数のクラスターからの寄与に対する領域重み付け係数の一例を示す。

【図145】アップサンプリング及び背景マスキングのための補間を使用する一例を示す。

【図146】アップサンプリング及び背景マスキングのためのサブピクセルカウント重み付けを使用する一例を示す。

【図147A】配列決定システムの一実施態様を示す。配列決定システムは、構成可能なプロセッサを含む。

20

【図147B】配列決定システムの一実施態様を示す。配列決定システムは、構成可能なプロセッサを含む。

【図147C】ベースコールセンサー出力などの、配列決定システムからのセンサーデータの分析のためのシステムの簡略ブロック図である。

【図148A】ホストプロセッサによって実行される実行時プログラムの機能を含む、ベースコール動作の態様を示す簡略図である。

【図148B】図147Cに示されるものなどの構成可能プロセッサの構成の簡略図である。

【図149】本明細書に開示される技術を実施するために、図147Aの配列決定システムによって使用され得るコンピュータシステムである。

30

【図150】データ正規化及びデータ増強を含み得る、データ前処理の異なる実施態様を示す。

【図151】図150のデータ正規化技術（DeepRTA（norm））及びデータ増強技術（DeepRTA（augment））が、ニューラルネットワークベースのベースコーラーが細菌データで訓練され、ヒトデータで試験されるときにベースコール誤差割合を低減し、細菌データ及びヒトデータが、同じアッセイを共有する（例えば、両方ともイントロンデータを含む）ことを示す。

【図152】図151のデータ正規化技術（DeepRTA（norm））及びデータ増強技術（DeepRTA（augment））が、ニューラルネットワークベースのベースコーラーが非エキソンデータ（例えば、イントロンデータ）で訓練され、エキソンデータで試験されるときにベースコール誤差割合を低減することを示す。

40

【発明を実施するための形態】

【0091】

以下の説明は、開示された技術を当業者が作製及び使用することを可能にするために提示され、特定の用途及びその要件に関連して提供される。開示される実施態様に対する様々な修正は、当業者には容易に明らかとなり、本明細書で定義される一般原理は、開示される技術の趣旨及び範囲から逸脱することなく、他の実施態様及び用途に適用され得る。したがって、開示される技術は、示される実施態様に限定されることを意図するものではなく、本明細書に開示される原理及び特徴と一致する最も広い範囲を与えられるものである。

50

(導入)

【0092】

デジタル画像からのベースコールは、大規模に平行であり、計算的に集中的である。このことは、本発明者らの新規な技術を導入する前に識別する多数の技術的課題を提示する。

【0093】

評価されている画像セットからの信号は、塩基の分類が周期的に、特に塩基のますます長いストランドにわたって進行するにつれて次第に微弱である。塩基分類がストランドの長さにわたって延在するにつれて、信号対雑音比は減少し、信頼性が低下する。信頼性の更新された推定値は、塩基分類の変化の推定された信頼性として予想される。

【0094】

デジタル画像は、サンプルストランドの増幅されたクラスターから捕捉される。サンプルは、様々な物理的構造及び化学物質を使用して、ストランドを複製することにより増幅される。合成による配列決定中、タグは、サイクルで化学的に結合され、光るように刺激される。デジタルセンサーは、画像を生成するためにピクセルから読み出されるタグから光子を収集する。

【0095】

塩基を分類するためにデジタル画像を解釈するには、位置不確実性を解消することが必要であり、限られた画像解像度により障害がある。ベースコール中に収集される解像度よりも高い解像度では、撮像されたクラスターは、不規則な形状を有し、中心位置を不確定に有することが明らかである。クラスター位置は機械的に制御されず、そのため、クラスター中心はピクセル中心と位置合わせされない。ピクセル中心は、ピクセルに割り当てられた整数座標であり得る。他の実施態様では、ピクセルの左上角であってもよい。更に他の実施態様では、ピクセルの重心又は質量中心とすることができる。増幅は、均一なクラスター形状を生成しない。したがって、デジタル画像内のクラスター信号の分布は、規則的なパターンではなく統計的分布である。本発明者らは、この位置の不確実性を求める。

【0096】

信号クラスのうちの1つは、検出可能な信号を生成せず、「暗」信号に基づいて特定の位置に分類され得る。したがって、暗サイクル中に分類するためにテンプレートが必要である。テンプレートの生成は、暗信号の欠落を回避するために、複数の撮像サイクルを使用して初期位置不確実性を解消する。

【0097】

画像センサーのサイズ、倍率、及びステップデザインにおけるトレードオフは、センサーピクセル中心と一致するようにクラスター中心を処理するには大きすぎるピクセルサイズにつながる。本開示は、2つの感覚でピクセルを使用する。物理的センサーピクセルは、検出された光子を報告する光センサーの領域である。単にピクセルと呼ばれる論理ピクセルは、少なくとも1つの物理ピクセルに対応するデータであり、センサーピクセルから読み出されたデータである。ピクセルは、サブピクセル（例えば、4×4サブピクセル）に細分化されるか、又は「アップサンプリング」され得る。全ての光子が物理ピクセルの片側に当たって反対側ではない可能性を考慮するために、バイリニア補間又はエリア重み付けなどの補間によって、サブピクセルに値を割り当てることができる。ピクセルが物理ピクセルからデータにアフィン変換を適用することによって、ピクセルが再フレーミングされるときに、補間又は双線型補間も適用される。

【0098】

より大きい物理ピクセルは、より小さいピクセルよりも微弱な信号に対してより感度が高い。デジタルセンサーは時間と共に改善されるが、集光器表面積の物理的制限は避けられない。設計トレードオフを考慮すると、レガシーシステムは、センサーピクセルの3×3つのパッチから画像データを収集及び解析するように設計されており、そのクラスターの中心は、パッチの中心ピクセルのどこにあるかを収集及び分析するように設計されている。

【0099】

高解像度センサーは、一度に撮像された媒体の一部のみを捕捉する。センサーは、画像化された媒体の上にステップ付きで、全視野を覆う。1つの処理サイクル中に数千のデジタル画像を収集することができる。

【0100】

センサー及び照明設計は、塩基を分類するために使用される少なくとも4つの照明応答値を区別するために組み合わせられる。ベイヤーカラーフィルタアレイを有する従来のRGBカメラを使用した場合、4つのセンサーピクセルが単一のRGB値に組み合わせられる。これは、4倍の有効なセンサー分解能を低減するであろう。あるいは、画像化された媒体とセンサーとの間の位置に回転された異なる照明波長及び／又は異なるフィルタを使用し、単一の位置で収集することができる。4つの基本分類間を区別するために必要とされる画像の数は、システム間で異なる。いくつかのシステムは、異なるクラスの塩基に対して4つの強度レベルを有する1つの画像を使用する。他のシステムは、異なる照明波長（例えば、赤及び緑）を有する2つの画像、及び／又は塩基を分類するための一種の真理台を有するフィルタを使用する。システムはまた、特定の塩基クラスに調整された異なる照明波長及び／又はフィルタを有する4つの画像を使用することができる。

10

【0101】

デジタル画像の非常に平行な処理は、実際には、30～2000塩基対程度の比較的短いストランドを、長さのより長い、潜在的に数百万、又は更には長さが数十億である配列に位置合わせする必要がある。画像化された媒体上では冗長サンプルが望ましいため、配列の一部は、多数のサンプルリードによって被覆されてもよい。数千個のサンプルクラスターが単一の画像化された培地から撮像される。そのような多くのクラスターの大規模な処理は、コストを減少させる一方で、配列決定容量が増加している。

20

【0102】

配列決定の能力は、ムーアの法則を再現するペースで増加している。第1の配列決定コストは十億ドルであるが、Illumina（商標）などの2018年のサービスでは、（数）百ドルの結果を提供する。配列決定が主流に、かつ単価が降下するにつれて、分類のためにより少ないコンピューティング電力が利用可能であり、このことが、ほぼリアルタイム分類の課題を増加させる。これらの技術的課題を念頭に置いて、本発明者らは、開示された技術に転じる。

【0103】

開示された技術は、位置不確実性を解消するためのテンプレート生成中、及び分解された位置におけるクラスターの塩基分類中の両方の処理を改善する。開示される技術を適用することは、機械のコストを低減するために、より安価なハードウェアを使用することができる。ほぼリアルタイムの分析は費用効率が高くなり、画像収集と塩基分類との間の遅れを低減することができる。

30

【0104】

開示される技術は、センサーピクセルをサブピクセルに補間することによって生成されたアップサンプリングされた画像を使用し、次いで位置不確実性を解決するテンプレートを生成することができる。得られたサブピクセルは、そのサブピクセルがクラスターの中心にあるかのように、サブピクセルを処理する分類のためのベースコーラーに提出される。クラスターは、同じ塩基分類を繰り返し受信する隣接するサブピクセルのグループから特定される。この技術のこの態様は、既存のベースコール技術を活用して、クラスターの形状を特定し、クラスター中心をサブピクセル解像度で超検索することができる。

40

【0105】

開示される技術の別の態様は、信頼できる特定されたクラスター中心及び／又はクラスター形状を有する画像をペアリングする、グラウンドトゥールズを作成することである。深層学習システム及び他の機械学習アプローチは、実質的な訓練セットを必要とする。人間がキュレートしたデータは、コンパイルに費用がかかる。開示された技術を使用して、非標準的な動作モードで、人のキュレーターの介入又は費用を伴わずに、機密に分類された訓練データの大きなセットを生成することができる。訓練データは、CNNベースの深

50

層学習システムなどの非標準的な動作モードで、既存の分類子から入手可能なクラスター中心及び／又はクラスター形状を有する生画像を相関させる。1つの訓練画像を回転させ、反射させて、追加の等しく有効な実施例を生成することができる。訓練例は、全体画像内の所定のサイズの領域に焦点を合わせることができる。ベースコール中に評価されたコンテキストは、画像のサイズ又は画像化された媒体全体ではなく、例示的な訓練領域のサイズを決定する。

【0106】

開示される技術は、訓練データとして、又は塩基分類のためのテンプレートとして使用可能な、異なる種類のマップを生成することができ、このマップは、クラスター中心及び／又はクラスター形状をデジタル画像と相関させる。第1に、サブピクセルはクラスター中心として分類することができ、それによって、物理的センサーピクセル内のクラスター中心を局所化することができる。第2に、クラスター中心は、クラスター形状の重心として計算することができる。この位置は、選択された数値精度で報告することができる。第3に、クラスター中心は、サブピクセル又はピクセル解像度のいずれかで、減衰マップ内の周囲のサブピクセルで報告することができる。減衰マップは、クラスター中心からの領域の分離が増加するにつれて、領域内で検出された光子に与えられる重みを低減し、より遠い位置からの信号を減衰させる。第4に、隣接領域のクラスター内のサブピクセル又はピクセルに、バイナリ又は三元分類を適用することができる。バイナリ分類では、領域は、クラスター中心に属するか、又は背景として分類される。三元分類では、第3のクラスタイプは、クラスター内部を含むがクラスター中心ではない領域に割り当てられる。クラスター中心位置のサブピクセル分類は、より大きい光学ピクセル内の実数値クラスター中心座標に対して置換され得る。

【0107】

代替的なマップのスタイルは、最初に、グラウンドトゥールズデータセットとして生成することができ、又は訓練を行って、ニューラルネットワークを使用して生成することができる。例えば、クラスターは、適切な分類を有する隣接するサブピクセルの不連続領域として描写することができる。ニューラルネットワークからのマッピングされたクラスターの強度は、ピーク検出器フィルタによって後処理されて、中心が既に決定されていない場合、クラスター中心を計算することができる。いわゆる流域分析を適用することにより、隣接する領域を別個のクラスターに割り当てることができる。ニューラルネットワーク推測エンジンによって生成されるとき、マップは、デジタル画像の配列を評価し、ベースコールのサイクルにわたって塩基を分類するためのテンプレートとして使用することができる。

【0108】

塩基がデジタル画像の配列に分類されるとき、ニューラルネットワークは、過去及び将来のサイクルの画像チャンネルと共に、現在のサイクルで複数の画像チャンネルを処理する。クラスターでは、ストランドの一部は、合成の主要な過程の前又は後で動作し得、位相ずれのタグ付けは、プレフェイジング又はフェイジングとして知られている。プレフェイジング及びポストフェイジングの低い速度が経験的に観察されると仮定すると、プレフェイジング及びポストフェイジングから生じる信号内のノイズのほとんど全ては、現在、過去及び将来のサイクルで3つのサイクルだけでデジタル画像を処理するニューラルネットワークによって取り扱われ得る。

【0109】

現在のサイクル内のデジタル画像チャンネル間で、サイクル内の画像を位置合わせするための念入りの位置合わせは、正確な塩基分類に強く寄与する。誤差の他のソースの中でもとりわけ、波長及び不一致照明源の組み合わせは、測定されたクラスター中心場所で、小さい修正可能な差を生成する。並進、回転、及びスケールリングを伴う一般的なアフィン変換は、画像タイルにわたるクラスター中心を正確に位置合わせするために使用され得る。アフィン変換は、画像データを再構成し、かつクラスター中心に対するオフセットを解消するために使用され得る。

【0110】

画像データの再構成は、典型的には、アフィン変換を適用することによって、画像データを補間することを意味する。再構成は、ピクセルパッチの中心ピクセルの中央に対象のクラスター中心を置き得る。又は、それは、画像をテンプレートと位置合わせして、画像収集中にジッター及び他の不一致を克服し得る。再構成は、ピクセルパッチ内の全てのピクセルの強度値を調整することを含む。双1次及び双3次補間並びに重み付け領域調整は、代替戦略である。

【0111】

いくつかの実施態様では、クラスター中心座標は、追加の画像チャンネルとしてニューラルネットワークに供給され得る。

【0112】

距離信号はまた、塩基分類に寄与し得る。いくつかのタイプの距離信号は、クラスター中心からの領域の分離を反映する。最も強い光信号は、クラスター中心と一致すると見なされる。クラスター周囲に沿った光信号は、時には、近くのクラスターからのストレイ信号を含む。分類は、信号成分の寄与がクラスター中心からのその分離に従って減衰されるとき、より正確であることが観察されている。作用する距離信号は、単一のクラスター距離チャンネルと、マルチクラスター距離チャンネルと、マルチクラスター形状ベースの距離チャンネルと、を含む。単一のクラスター距離チャンネルは、中心ピクセル内のクラスター中心を有するパッチに適用する。次いで、パッチ内の全ての領域の距離は、中心ピクセル内のクラスター中心からの距離である。中心ピクセルと同じクラスターに属しないピクセルは、所与の計算された距離の代わりに、背景としてフラグ付けされ得る。マルチクラスター距離チャンネルは、最も近いクラスター中心に対する各領域の距離を事前計算する。これは、領域を誤ったクラスター中心に接続する可能性を有するが、その可能性は低い。マルチクラスター形状ベースの距離チャンネルは、隣接する領域を介して領域（サブピクセル又はピクセル）を、同じ塩基分類を生成するピクセル中心に関連付ける。いくつかの計算コストで、これは、誤ったピクセルに対する距離を測定する可能性を回避する。マルチクラスター及び距離信号に対するマルチクラスター形状ベースのアプローチは、画像内の複数のクラスターでの事前計算及び使用が行われるという利点を有する。

【0113】

信号対ノイズ比を改善するために、ノイズから信号を分離するように、ニューラルネットワークによって形状情報が使用され得る。上記の考察では、領域分類及び距離チャンネル情報の供給に対するいくつかのアプローチが識別された。いずれのアプローチにおいても、クラスターエッジを画定するために、クラスターの一部としてではなく背景として、領域がマークされ得る。ニューラルネットワークは、不規則なクラスター形状に関する得られた情報を利用するように訓練され得る。距離情報及び背景分類は、組み合わせられ得るか、又は別々に使用され得る。隣接するクラスターからの信号の分離は、クラスター密度が増加するにつれて、ますます重要になる。

【0114】

並行処理のスケールを増加させるための1つの方向は、画像化された媒体上のクラスター密度を増加させることである。密度の増加は、隣接する隣接部を有するクラスターを読み取るとき、背景ノイズを増加させるダウンサイドを有する。任意の（例えば、 3×3 のピクセルの）パッチの代わりに、形状データを使用することは、例えば、クラスター密度が増加するにつれて信号分離を維持するのに役立つ。

【0115】

開示される技術の一態様を適用して、塩基分類スコアはまた、品質を予測するために活用され得る。開示される技術は、直接又は予測モデルを介して、分類スコアを従来の S a n g e r 又は P h r e d クオリティ Q スコアと相関させることを含む。Q20、Q30、又はQ40などのスコアは、 $Q = -10 \log_{10} P$ によって塩基分類誤差確率に対数的に関連する。クラススコアのQスコアとの相関は、多出力ニューラルネットワーク又は多変量回帰分析を使用して実行され得る。品質スコアのリアルタイム計算の利点は、塩基

10

20

30

40

50

分類中、不備のある配列決定動作が早期に終了され得ることである。出願人は、動作を終了させるための時折の（まれな）決定が、分析配列を介して8分の1～4分の1の所で行われ得ることを見出した。終了させるための決定は、50サイクル後又は25～75サイクル後に行われ得る。そうではなく300～1000サイクルを動作する連続プロセスでは、早期の終了は、実質的なリソース節約をもたらす。

【0116】

専用の畳み込みニューラルネットワーク（CNN）構造は、複数サイクルにわたって塩基を分類するために使用され得る。1つの専門化は、処理の初期層中のデジタル画像チャネル間での分離を含む。畳み込みフィルタスタックは、サイクル間の処理を分離し、異なるサイクルからのデジタル画像セット間のクロストークを防止するように構造化され得る。サイクル間の処理の分離の動機は、異なるサイクルで撮られる画像が、残留位置合わせ誤差を有し、したがって位置合わせ不良であり、互いにランダムな並進オフセットを有することである。これは、センサーの運動段階の移動の有限精度に起因し、また、異なる周波数チャンネルで撮られる画像が異なる光路及び波長を有するために生じる。

【0117】

連続するサイクルから画像セットを使用する動機は、特定のサイクルでの信号へのプレフェイジング及びポストフェイジングの寄与が、2次の寄与であることである。画像収集サイクル間のデジタル画像セットの下位層の畳み込みを構造的に分離することが、畳み込みニューラルネットワークにとって有用であり得るということになる。

【0118】

畳み込みニューラルネットワーク構造はまた、クラスタリングに関する情報を取り扱う際に専用化され得る。クラスター中心及び／又は形状についてのテンプレートは、畳み込みニューラルネットワークがデジタル画像データと組み合わせる追加の情報を提供する。クラスター中心分類及び距離データは、サイクルにわたって繰り返し適用され得る。

【0119】

畳み込みニューラルネットワークは、画像フィールドで複数のクラスターを分類するように構造化され得る。複数のクラスターが分類されるとき、ピクセル又はサブピクセルについての距離チャンネルは、ピクセル又はサブピクセルが属する最も近いクラスター中心又は隣接するクラスター中心のいずれかに対する距離情報をよりコンパクトに含み得る。代替的に、各ピクセル若しくはサブピクセルに対して、又は少なくとも、クラスター中心を含む各々のものに対して、大きい距離ベクトルが供給され得、それは、所与のピクセルについてのコンテキストである、クラスター中心から全ての他のピクセルまでの完全な距離情報を与える。

【0120】

ベースコールでのテンプレート生成のいくつかの組み合わせは、距離チャンネルに取って代わるために領域重み付けにおける変動を使用し得る。ここでの考察は、距離チャンネルの代わりに、テンプレート生成器の出力がどのように直接使用され得るかについて説明する。

【0121】

我々は、テンプレート画像をピクセル値修正に直接適用することに影響を及ぼす3つの検討事項、すなわち、画像セットがピクセル又はサブピクセルドメイン内で処理されるかどうか、いずれのドメインにおいても、どのように領域重みが計算されるか、及びサブピクセルドメインで、補間強度値を修正するためにマスクとしてテンプレート画像を適用することを説明する。

【0122】

ピクセルドメインで塩基分類を実行することは、アップサンプリングから生じる、16倍などの計算の増加を要求しないという利点を有する。ピクセルドメインでは、畳み込みの上部層でさえ、必要とされない計算をキャンセルするためのロジックを追加する代わりに、収集されない計算の実行を正当化するのに十分なクラスター密度を有し得る。我々は、距離チャンネルなしでテンプレート画像データを直接使用する、ピクセルドメイン内の例から開始する。

【0123】

いくつかの実施態様では、分類は、特定のクラスターに焦点を当てる。これらの例では、クラスターの周囲におけるピクセルは、どの隣接するクラスターが分類の焦点であるかに応じて、異なる修正強度値を有し得る。サブピクセルドメイン内のテンプレート画像は、重複ピクセルが強度値を2つの異なるクラスターに与えることを示し得る。我々は、2つ又はそれ以上の隣接の又は隣接するクラスターが両方ともピクセルと重複するとき、両方とも光学ピクセルからの強度読み取り値に寄与するとき、「重複ピクセル」として光学ピクセルを指す。隆起線での異なる流域内へのレインフローの分離から名付けられる流域分析は、更なる隣接するクラスターを分離するために適用され得る。クラスターごとの分類のためにデータが受信されるとき、テンプレート画像は、クラスターの周囲に沿って重複ピクセルについての強度データを修正するために使用され得る。重複ピクセルは、どのクラスターが分類の焦点であるかに応じて、異なる修正強度を有し得る。

10

【0124】

ピクセルの修正強度は、アウェイクラスター（すなわち、ピクセルが強度放射を示す非ホームクラスター）とは対照的に、ホームクラスター（すなわち、ピクセルが属するクラスターか、又はピクセルが強度放射を主に示すクラスター）への、重複ピクセル内のサブピクセル寄与に基づいて低減され得る。5個のサブピクセルがホームクラスターの一部であり、2個のサブピクセルがアウェイクラスターの一部であると仮定する。次いで、7個のサブピクセルは、強度をホーム又はアウェイクラスターに与える。ホームクラスターに焦点を合わせている間、一実施態様では、16個のサブピクセルのうちの7個が強度をホーム又はアウェイクラスターに与えるため、重複ピクセルは、強度を $7/16$ 低減される。別の実施態様では、強度は、サブピクセルの総数によって分割されるホームクラスターに寄与するサブピクセルの領域に基づいて $5/16$ 低減される。第3の実施態様では、強度は、寄与するサブピクセルの総領域によって分割されるホームクラスターに寄与するサブピクセルの領域に基づいて、 $5/7$ 低減される。後者の2つの計算は、焦点がアウェイクラスターに変わるときに変化し、分子中に「2」を有する分数を生成する。

20

【0125】

当然のことながら、距離チャンネルがクラスター形状のサブピクセルマップと共に考慮されている場合、強度の更なる低減が適用され得る。

【0126】

分類の焦点であるクラスターについてのピクセル強度がテンプレート画像を使用して修正されると、修正ピクセル値は、ニューラルネットワークベースの分類子の層を介して畳み込まれて、修正画像を生成する。修正画像は、連続する配列決定サイクルで塩基を分類するために使用される。

30

【0127】

代替的に、ピクセルドメイン内の分類は、まとまった画像内の全てのピクセル又は全てのクラスターに対して並行して進行し得る。ピクセル値の1つの修正のみが、中間計算の再使用を確実にするために、このシナリオで適用され得る。上で与えられる分数のいずれもが、より小さい又はより大きい強度の減衰が望ましいかどうかに応じて、ピクセル強度を修正するために使用され得る。

40

【0128】

まとまった画像についてのピクセル強度がテンプレート画像を使用して修正されると、ピクセル及び周囲のコンテキストは、ニューラルネットワークベースの分類子の層を介して畳み込まれて、修正画像を生成し得る。まとまった画像において畳み込みを実行することにより、コンテキストを共有したピクセル間での中間計算の再使用が可能になる。修正画像は、連続する配列決定サイクルで塩基を分類するために使用される。

【0129】

この説明は、サブピクセルドメイン内の領域重みの適用のために並行にされ得る。並行は、重みが個々のサブピクセルについて計算され得ることである。重みは、光学ピクセルの異なるサブピクセル部分と同じであり得るが、そうする必要はない。ホーム及びアウェイ

50

イクラスターの上記のシナリオを、それぞれ、重複ピクセルの5個及び2個のサブピクセルで繰り返し、ホームクラスターに属するサブピクセルへの強度の割り当ては、ピクセル強度の $7/16$ 、 $5/16$ 、又は $5/7$ であり得る。再び、距離チャンネルがクラスター形状のサブピクセルマップと共に考慮されている場合、強度の更なる低減が適用され得る。

【0130】

まとまった画像についてのピクセル強度がテンプレート画像を使用して修正されると、サブピクセル及び周囲のコンテキストは、ニューラルネットワークベースの分類子の層を介して畳み込まれて、修正画像を生成し得る。まとまった画像において畳み込みを実行することにより、コンテキストを共有したサブピクセル間での中間計算の再使用が可能になる。修正画像は、連続する配列決定サイクルで塩基を分類するために使用される。

10

【0131】

別の代替案は、サブピクセルドメインで、バイナリマスクとしてテンプレート画像を、サブピクセルドメイン内に補間される画像データに適用することである。テンプレート画像は、クラスター間で背景ピクセルを必要とするように、又は異なるクラスターからのサブピクセルが隣接することを可能にするように配置され得る。テンプレート画像は、マスクとして適用され得る。マスクは、補間されたピクセルが補間によって割り当てられる値を維持するか、又はそれがテンプレート画像で背景として分類される場合、背景値（例えば、ゼロ）を受信するかどうかを決定する。

【0132】

再び、まとまった画像についてのピクセル強度がテンプレート画像を使用してマスクされると、サブピクセル及び周囲のコンテキストは、ニューラルネットワークベースの分類子の層を介して畳み込まれて、修正画像を生成し得る。まとまった画像において畳み込みを実行することにより、コンテキストを共有したサブピクセル間での中間計算の再使用が可能になる。修正画像は、連続する配列決定サイクルで塩基を分類するために使用される。

20

【0133】

開示される技術の特徴は、中間計算を再使用して、共有されたコンテキスト内で任意の数のクラスターを分類するように組み合わせ可能である。光学ピクセル解像度で、一実施態様では、ピクセルの約10パーセントは、分類されるクラスター中心を保持する。レガシーシステムでは、不規則な形状のクラスターの観察を仮定すると、 3×3 の光学ピクセルは、クラスター中心についての潜在的な信号寄与因子として分析するためにグループ化された。上部畳み込み層から離れた1つの 3×3 フィルタでさえ、クラスター密度は、光学ピクセルの半分よりも実質的に多くの光信号をクラスター中心でのピクセル内にロールアップする可能性が高い。超サンプリング解像度でのみ、上部畳み込み層についてのクラスター中心密度は、1パーセント未満に低下する。

30

【0134】

いくつかの実施態様では、共有コンテキストは実質的である。例えば、 15×15 の光学ピクセルコンテキストは、正確な塩基分類に寄与し得る。同等の $4 \times$ アップサンプリングされたコンテキストは、 60×60 のサブピクセルである。このコンテキストの範囲は、撮像中に、ニューラルネットワークが不均一な照明及び背景の影響を認識するのに役立つ。

40

【0135】

開示される技術は、デジタル画像入力で検出される境界を有するテンプレート入力内のクラスター境界を組み合わせるために、低畳み込み層で小さいフィルタを使用する。クラスター境界は、ニューラルネットワークが背景条件から信号を分離し、背景に対する画像処理を正規化するのに役立つ。

【0136】

開示される技術は、中間計算を実質的に再使用する。 $20 \sim 25$ 個のクラスター中心が、 15×15 の光学ピクセルのコンテキスト領域内に現れると仮定する。次いで、第1の層の畳み込みは、ブロックごとの畳み込みロールアップで $20 \sim 25$ 回再使用されることになる。再使用係数は、光学解像度での再使用係数が $1 \times$ 未満に低下する最初の時間であ

50

る最終層まで層ごとに低減される。

【0137】

複数の畳み込み層からのブロックごとのロールアップ訓練及び推測は、連続するロールアップを、ピクセル又はサブピクセルのブロックに適用する。ブロック周囲で、第1のデータブロックのロールアップ中に使用されるデータがロールアップの第2のブロックと重複し、再使用され得る重複ゾーンが存在する。ロールアップされ得、かつ再使用され得る、ピクセル値及び中間計算は、重複ゾーンによって囲まれる中心領域でブロック内にある。重複ゾーンで、 3×3 フィルタの適用によって、例えば、 15×15 から 13×13 に、コンテキストフィールドのサイズを漸進的に低減する畳み込み結果は、ブロック内の基礎計算の再使用を損なうことなくメモリを保存して、畳み込まれる値を保持する同じメモリブロック内に書き込まれ得る。より大きいブロックで、重複ゾーン内の中間計算の共有は、より少ないリソースを必要とする。より小さいブロックで、重複ゾーン内の中間計算を共有するために、複数のブロックを並行して計算することが可能であり得る。

10

【0138】

より大きいフィルタ及び拡張は、畳み込み層の数を低減し、それは、より低い畳み込み層がテンプレート及び／又はデジタル画像データでクラスター境界に反応した後の、分類を損なうことのない速度計算であり得る。

【0139】

テンプレートデータについての入力チャンネルは、デジタル画像フィールド内の複数のクラスター中心の分類と一致するテンプレート構造を作製するように選択され得る。上述の2つの代替案は、この一貫性基準、すなわち、全体のコンテキスト上の再構成及び距離マッピングを満たさない。再構成は、1つのクラスターだけの中心を光学ピクセルの中心に置く。クラスター中心を保持するとして分類されるピクセルに対する中心オフセットを供給することは、複数のクラスターを分類するためにより良好である。

20

【0140】

距離マッピングは、提供される場合、全てのピクセルが全コンテキスト上でそれ自体の距離マップを有しない限り、全コンテキスト領域にわたって実行するのが困難である。より単純な距離マップは、デジタル画像入力ブロックから複数のクラスターを分類するための有用な一貫性を提供する。

【0141】

ニューラルネットワークは、クラスターの境界でのピクセル又はサブピクセルのテンプレート内の分類から学習し得、そのため、距離チャンネルは、クラスター中心オフセットチャンネルを伴うバイナリ又は三元分類を供給するテンプレートによって取って代わり得る。使用されるとき、距離マップは、ピクセル（又はサブピクセル）が属するクラスター中心からのピクセルの距離を与え得る。又は、距離マップは、最も近いクラスター中心に対する距離を与え得る。距離マップは、背景ピクセルに割り当てられるフラグ値でバイナリ分類を符号化し得るか、又はそれは、ピクセル分類からの別個のチャンネルであり得る。クラスター中心オフセットと組み合わせられて、距離マップは、三元分類を符号化し得る。いくつかの実施態様では、特に1ビット又は2ビットでピクセル分類を符号化するものは、少なくとも開発中に、ピクセル分類及び距離についての別個のチャンネルを使用することが望ましい場合がある。

30

【0142】

開示される技術は、上部層内のいくつかの計算リソースを節約するための計算の低減を含み得る。クラスター中心オフセットチャンネル又は三元分類マップは、ピクセル中心の最終分類に寄与しないピクセル畳み込みの中心を識別するために使用され得る。多くのハードウェア／ソフトウェアの実施態様では、推測中にルックアップを実行し、畳み込みロールアップをスキップすることは、9個の乗算及び8個の加算さえ実行して 3×3 フィルタを適用するよりも、上部層（複数可）でより効率的であり得る。並列実行のための計算をパイプライン処理するカスタムハードウェアでは、全てのピクセルは、パイプライン内で分類され得る。次いで、クラスター中心と一致するピクセルのみに対する結果を収集する

40

50

ために、クラスター中心マップは、最終分類がそれらのピクセルに対してのみ所望されるため、最終畳み込み後に使用され得る。再び、光学ピクセルドメインでは、現在観察されているクラスター密度で、ピクセルの約10パーセントに対するロールアップ計算が収集される。4xアップサンプリングされたドメインでは、上部層内のサブピクセル分類の1パーセント未満が収集されるため、いくつかハードウェア上で、スキップされた畳み込みから、より多くの層が恩恵を受け得る。

(ニューラルネットワークに基づくテンプレート生成)

【0143】

テンプレート生成の第1の工程は、クラスターメタデータを特定することである。クラスターメタデータは、それらの中心、形状、サイズ、背景、及び／又は境界を含むクラスターの空間分布を識別する。

10

(クラスターメタデータの特定)

【0144】

図1は、サブピクセルベースコールを使用してクラスターメタデータを特定する処理パイプラインの一実施態様を示す。

【0145】

図2は、そのタイル内にクラスターを含むフローセルの一実施態様を示す。フローセルは、レーンに分割される。レーンは、「タイル」と呼ばれる非重複領域に更に分割される。配列決定手順中、タイル上の集団及びそれらの周囲の背景が画像化される。

【0146】

20

図3は、8つのレーンを有する例示的なIllumina GA-IIx (商標) フローセルを示す。図3はまた、1つのタイル及びそのクラスター及びそれらの周囲の背景上の拡大も示す。

【0147】

図4は、4チャンネル化学の配列決定画像の画像セットを描写しており、すなわち、画像セットは、ピクセルドメイン内の4つの異なる波長帯域 (画像／撮像チャンネル) を使用して捕捉された4つの配列決定画像を有する。画像セット内の各画像は、フローセルのタイルを覆い、タイル上のクラスターの強度放射を示し、フローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで特定の画像チャンネルのために捕捉された、それらの周囲の背景を示す。一実施態様では、各撮像チャンネルは、複数のフィルタ波長帯域のうちの1つに対応する。別の実施態様では、各撮像チャンネルは、配列決定サイクルで複数の撮像イベントのうちの1つに対応する。更に別の実施態様では、各撮像チャンネルは、特定のレーザーを用いた照明と特定の光学フィルタを通した撮像との組み合わせに対応する。クラスターの強度放射は、検体に関連する塩基を分類するために使用され得る検体から検出された信号を含む。例えば、強度放射は、タグが刺激され、1つ又はそれ以上のデジタルセンサーによって検出され得る、サイクル中に検体に化学的に取り付けられたタグによって放出される光子を示す信号であってもよい。

30

【0148】

図5は、配列決定画像をサブピクセル (又はサブピクセル領域) に分割する一実施態様である。図示の別の実施態様では、4分の1 (0.25) サブピクセルが使用され、これにより、配列決定画像内の各ピクセルが16個のサブピクセルに分割される。図示した配列決定画像が、20x20ピクセル、すなわち、400ピクセルの解像度を有することを前提として、分割は6400サブピクセルを生成する。サブピクセルのそれぞれは、サブピクセルベースコールのための領域中心として、ベースコーラーによって処理される。いくつかの実施態様では、このベースコーラーは、ニューラルネットワークベースの処理を使用しない。他の実施態様では、このベースコーラーは、ニューラルネットワークベースのベースコーラーである。

40

【0149】

所与の配列決定サイクル及び特定のサブピクセルに関して、ベースコーラーは、画像処理工程を実行し、配列決定サイクルの対応する画像セットからサブピクセルの強度データ

50

を抽出することによって、所与の配列決定サイクル特定のサブピクセルに対するベースコールを生成するように論理を用いて構成される。これは、サブピクセルのそれぞれ、及び複数の配列決定サイクルのそれぞれに対して行われる。また、Illumina MiSeq シーケンサの1800×1800ピクセル解像度タイル画像の1/4サブピクセル分割を用いて実験を行った。サブピクセルベースコールを、50回の配列決定サイクル及び10タイルのレーンについて行った。

【0150】

図6は、サブピクセルベースコール中に、ベースコーラーによって識別されたクラスターの予備中心座標を示す。図6はまた、予備中心座標を含む「原点サブピクセル」又は「中心サブピクセル」を示す。

【0151】

図7は、クラスターメタデータを含むいわゆる「クラスターマップ」を生成するために、複数の配列決定サイクルにわたって生成されたサブピクセルベースコールをマージする一例を示す。図示した実施態様では、サブピクセルベースコールは、第1の検索アプローチを使用してマージされる。

【0152】

図8aは、サブピクセルベースコールのマージによって生成されたクラスターマップの一例を示す。図8bは、サブピクセルベースコールの一例を示す。図8bはまた、サブピクセルベースから生成されたサブピクセルごとのベースコール配列を分析してクラスターマップを生成する一実施態様を示す。

(配列決定画像)

【0153】

クラスターメタデータ判定は、配列決定機器102（例えば、IlluminaのiSeq、HiSeqX、HiSeq3000、HiSeq4000、HiSeq2500、NovaSeq 6000、NextSeq、NextSeqDx、MiSeq及びMiSeqDx）によって生成された画像データを分析することを含む。以下の説明は、一実施態様に従って、画像データがどのように生成されるか、及びそれを描写するものを概説する。

【0154】

ベースコールは、配列決定機器102の生信号、すなわち、画像から抽出された強度データがDNA配列及び品質スコアにデコードされるプロセスである。一実施態様では、Illuminaプラットフォームは、ベースコールのための環状可逆終端（Cyclic Reversible Termination、CRT）化学を採用する。このプロセスは、新たに添加された各ヌクレオチドの放出信号を追跡しながら、改変されたヌクレオチドを有するテンプレートDNA鎖に相補的な成長した出現DNA鎖上に依存する。修飾されたヌクレオチドは、ヌクレオチド型のフルオロフォアシグナルをアンカーする3'の取り外し可能なブロックを有する。

【0155】

配列決定は繰り返しサイクルで行われ、それぞれは3つの工程、すなわち、（a）修飾されたヌクレオチドを追加することによって経鼻鎖を伸長する工程と、（b）光学系104の1つ又はそれ以上のレーザーを使用して蛍光団を励起し、光学系104の異なるフィルタを通して画像化して、配列決定画像108を生成する工程と、（c）蛍光団の開裂及び次の配列決定サイクルの準備における3'ブロックを除去する工程と、を含む。組み込み及び撮像サイクルを、指定された数の配列決定サイクルに繰り返し、全ての集団の読み取り長さを規定する。このアプローチを使用して、各サイクルはテンプレートストランドに沿って新しい位置を問い合わせる。

【0156】

Illuminaプラットフォームのトレメントパワーは、CRT反応を受ける数百万のクラスター又は更には数十億のクラスターを同時に実行及び感知する能力からシステムを形成する。配列決定プロセスは、フローセル202において、配列決定プロセス中に入力

10

20

30

40

50

DNA断片を保持する小さなスライドガラスである。フローセル202は、顕微鏡画像、励起レーザー、及び蛍光フィルタを含むハイスループット光学システム104に接続される。フローセル202は、レーン204と呼ばれる複数のチャンバを含む。レーン204は、互いに物理的に分離され、異なるタグ付けされた配列決定ライブラリを含んでもよく、試料交差汚染なしで区別可能である。撮像装置106（例えば、電荷結合素子（Charge-Coupled Device、CCD）又は相補的金属酸化物半導体（Complementary Metal-Oxide-Semiconductor、CMOS）センサーなどのソリッドステート撮像素子）は、タイル206と呼ばれる一連の非重複領域において、レーン204に沿った複数の場所でスナップショットを取る。

【0157】

例えば、Illumina Genome Analyzer IIのレーン当たり100タイル、及びIllumina HiSeq2000内のレーン当たり64個のタイルが存在する。タイル206は数十万〜数百万個のクラスターを保持する。明るいスポットとして示されるクラスターを有するタイルから生成された画像を208で示す。クラスター302は、テンプレート分子の約千個の同一のコピーを含むが、クラスターはサイズ及び形状が異なる。クラスターは、配列決定動作前に、入力ライブラリのブリッジ増幅によって、テンプレート分子から成長させる。増幅及びクラスター成長の目的は、撮像装置106が単一の蛍光団を確実に感知できないため、放出された信号の強度を増大させることである。しかしながら、クラスター302内のDNAフラグメントの物理的距離は小さいため、撮像装置106は、単一のスポット302として断片のクラスターを知覚する。

【0158】

配列決定動作の出力は、レーン、タイル、配列決定サイクル、及びフルオロフォア（208A、208C、208T、208G）の特定の組み合わせのための、ピクセルドメイン内のタイル上のクラスターの強度放射を示す配列決定画像108である。

【0159】

一実施態様では、バイオセンサーは、光センサーのアレイを備える。光センサーは、バイオセンサーの検出表面上の対応するピクセル領域（例えば、反応部位／ウェル／ナノセル）からの情報を感知するように構成されている。ピクセル領域内に配設された検体は、ピクセル領域、すなわち、関連する検体と関連付けられると言われる。配列決定サイクルでは、ピクセル領域に対応する光センサーは、関連する検体からの発光／光子を検出／捕捉／検知するように構成され、それに応じて、画像化されたチャンネルごとにピクセル信号を生成するように構成される。一実施態様では、各撮像チャンネルは、複数のフィルタ波長帯域のうちの1つに対応する。別の実施態様では、各撮像チャンネルは、配列決定サイクルで複数の撮像イベントのうちの1つに対応する。更に別の実施態様では、各撮像チャンネルは、特定のレーザーを用いた照明と特定の光学フィルタを通した撮像との組み合わせに対応する。

【0160】

光センサーからのピクセル信号は、（例えば、通信ポートを介して）バイオセンサーに結合された信号プロセッサに伝達される。各配列決定サイクル及び各画像化チャンネルについて、信号プロセッサは、ピクセルが対応する光センサーから得られるピクセル信号をそれぞれ描写／含有／示す／表す／特徴付ける画像を生成する。このようにして、画像内のピクセルは、（i）ピクセルによって表されるピクセル信号を生成したバイオセンサーの光センサーと、（i i）対応する光センサーによって放射が検出され、ピクセル信号に変換された関連検体と、（i i i）関連検体を保持するバイオセンサーの検出表面上のピクセル領域と、に対応する。

【0161】

例えば、配列決定動作が2つの異なる画像化チャンネル：赤色チャンネル及び緑色チャンネルを使用すると考える。次いで、各配列決定サイクルにおいて、信号プロセッサは、赤色画像及び緑色画像を生成する。このようにして、配列決定動作の一連のk配列決定サイクルについて、赤色画像及び緑色画像のk対を有する配列が出力として生成される。

【0162】

赤色画像及び緑色画像（すなわち、異なる画像化チャンネル）内のピクセルは、配列決定サイクル内で1対1の対応を有する。これは、一対の赤色画像及び緑色画像内の対応するピクセルが、異なる画像化チャンネル内にある同じ関連する検体の強度データを示すことを意味する。同様に、赤色画像及び緑色画像の対にわたるピクセルは、配列決定サイクル間に1対1の対応を有する。これは、赤色画像及び緑色画像の異なるペア内の対応するピクセルが、配列決定動作の異なる獲得イベント／時間ステップ（配列決定サイクル）に関して、同じ関連する検体の強度データを示すことを意味する。

【0163】

赤色画像及び緑色画像（すなわち、異なる画像化チャンネル）内の対応するピクセルは、第1の赤色チャンネル及び第2の緑チャンネル内の強度データを表す、「サイクルごとの画像」のピクセルと見なすことができる。ピクセルがピクセルエリアのサブセットのピクセル信号、すなわち、バイオセンサーの検出面の領域（タイル）を描写するサイクルごとの画像は、「サイクルごとのタイル画像」と呼ばれる。サイクルごとのタイル画像から抽出されたパッチは、「サイクルごとの画像パッチ」と呼ばれる。一実施態様では、パッチ抽出は、入力準備者によって実行される。

【0164】

画像データは、配列決定動作の一連のk配列決定サイクルのために生成された一連のサイクルごとの画像パッチを含む。サイクルごとの画像パッチ内のピクセルは、関連する検体のための強度データを含み、強度データは、関連付けられた検体からの排出を検出するように構成された対応する光センサーによって、1つ又はそれ以上の画像化チャンネル（例えば、赤色チャンネル及び緑色チャンネル）のために取得される。一実施態様では、単一のターゲットクラスターがベースコールされる場合、サイクルごとの画像パッチは、標的関連検体及び非中心ピクセルに関する強度データを含む中心ピクセルで中心に置かれ、サイクルごとの画像パッチ内の非中心ピクセルは、標的関連検体に隣接する関連する検体の強度データを含む。一実施態様では、画像データは、入力準備者によって調製される。

（サブピクセルベースコール）

【0165】

開示された技術は、配列決定動作中に生成された一連の画像セットにアクセスする。画像セットは、配列決定画像108を含む。配列決定動作のそれぞれの配列決定サイクル中にそれぞれ連続する画像セットが捕捉される。一連の画像（又は配列決定画像）は、フローセルのタイル及びそれらの周囲の背景上のクラスターを捕捉する。

【0166】

一実施態様では、配列決定動作は4つのチャンネル化学を利用し、各画像セットは4つの画像を有する。別の実施態様では、配列決定動作は2チャンネル化学を利用し、各画像セットは2つの画像を有する。更に別の実施態様では、配列決定動作は、1チャンネル化学を利用し、各画像セットは2つの画像を有する。更に他の実施態様では、各画像セットは1つの画像のみを有する。

【0167】

ピクセル領域の配列決定画像108は、まずサブピクセルアドレス指定器110によってサブピクセルドメインに変換され、サブピクセルドメイン内に配列決定画像112が生成される。一実施態様では、配列決定画像108内の各ピクセルは、16個のサブピクセル502に分割される。したがって、一実施態様では、サブピクセル502は、4分の1サブピクセルである。別の実施態様では、サブピクセル502は2分の1サブピクセルである。その結果、サブピクセルドメイン内の配列決定画像112のそれぞれは、複数のサブピクセル502を有する。

【0168】

次いで、サブピクセルは、ベースコーラー114への入力として別々に供給されて、サブピクセルの各々を4つの塩基（A、C、T、及びG）のうちの1つと分類するベースコールをベースコーラー114から取得する。これにより、配列決定動作の複数の配列決定

10

20

30

40

50

サイクルにわたって、サブピクセルのそれぞれについてのベースコール配列 1 1 6 を生成する。一実施態様では、サブピクセル 5 0 2 は、それらの整数又は非整数座標に基づいて、ベースコーラー 1 1 4 に識別される。複数の配列決定サイクル中に生成された画像セットにわたってサブピクセル 5 0 2 からの発光信号を追跡することにより、ベースコーラー 1 1 4 は、各サブピクセルの基礎 DNA 配列を回復する。この例を図 8 b に示す。

【0 1 6 9】

他の実施態様では、開示される技術は、ベースコーラー 1 1 4 から、5 つの塩基（A、C、T、G、及び N）のうちの 1 つとして、サブピクセルのそれぞれを分類する。このような実施態様では、N 個のベースコールは、通常、低いレベルの抽出された強度に起因する、決定されていないベースコールを示す。

【0 1 7 0】

ベースコーラー 1 1 4 のいくつかの例としては、非ニューラルネットワークベースの Illumina offerings、例えば、リアルタイム分析（Real Time Analysis、RTA）、Genome Analyzer Analysis Pipeline の Firecrest プログラム、統合 1 次分析報告（Integrated Primary Analysis and Reporting、IPAR）マシン、及びオフラインベースコーラー（Off-Line Basecaller、OLB）が挙げられる。例えば、ベースコーラー 1 1 4 は、最近傍強度抽出、ガウス系強度抽出、平均 2×2 サブピクセル領域に基づく強度抽出、 2×2 サブピクセル面積の最も明るい試験に基づく強度抽出、平均 3×3 サブピクセル面積、バイリニア強度抽出、双キュービク強度抽出、及び／又は加重面積被覆率に基づく強度抽出に基づく強度抽出のうちの少なくとも 1 つを含む、サブピクセルの強度を補間することによって、ベースコール配列を生成する。これらの技術は、「強度抽出方法」と題された付録に詳細に記載されている。

【0 1 7 1】

他の実施態様では、ベースコーラー 1 1 4 は、本明細書に開示されるニューラルネットワークベースのベースコーラー 1 5 1 4 などのニューラルネットワークベースのベースコーラーであり得る。

【0 1 7 2】

次いで、サブピクセルごとのベースコール配列 1 1 6 は、入力として検索器 1 1 8 に供給される。検索器 1 1 8 は、連続するサブピクセルの基本的なベースコール配列を検索する。連続するサブピクセルのベースコール配列は、ベースコールの所定の部分が、序数の位置ごとの基準（例えば、 ≥ 45 サイクルにおける 4 1 一致、 ≤ 45 サイクルにおける 4 不一致、 ≤ 50 サイクルにおける 4 不一致、又は ≤ 34 サイクルにおける 2 不一致）と一致するとき、連続するサブピクセルのベースコール配列は「実質的に一致する」。

【0 1 7 3】

次いで、検索器 1 1 8 は、実質的に一致するベースコール配列を共有する隣接するサブピクセルの、例えば 8 0 4 a ~ d などの、クラスターを識別するクラスターマップ 8 0 2 を生成する。本出願は、「不連続の」、「ばらばらな」、及び「非重複の」を互換的に使用する。検索は、クラスターの一部を含むサブピクセルをベースコールして、それらが実質的に一致するベースコール配列を共有する隣接するサブピクセルに、コールされたサブピクセルをリンクさせることを可能にすることを含む。いくつかの実施態様では、検索器 1 1 8 は、不連続領域の少なくとも一部が、クラスターとして処理される所定の最小数のサブピクセル（例えば、4、6、又は 10 サブピクセルを超える）を有することを必要とする。

【0 1 7 4】

いくつかの実施態様では、ベースコーラー 1 1 4 はまた、クラスターの予備中心座標を識別する。予備中心座標を含むサブピクセルは、原点サブピクセルと呼ばれる。ベースコーラー 1 1 4 及び対応する原点サブピクセル（6 0 6 a ~ c）によって識別されたいくつかの例示的な予備中心座標（6 0 4 a ~ c）が図 6 に示されている。しかしながら、以下に説明するように、原点サブピクセル（クラスターの予備中心座標）の識別は必要ではな

10

20

30

40

50

い。いくつかの実施態様では、検索器 118 は、原点サブピクセル 606 a ~ c から始まり連続的に連続する非原点サブピクセル 702 a ~ c を継続して、サブピクセルの実質的に一致するベースコール配列を識別するための、第 1 の検索を使用する。これは、以下に説明するように、任意選択的である。

(クラスターマップ)

【0175】

図 8 a は、サブピクセルベースコールのマージによって生成されたクラスターマップ 802 の一例を示す。クラスターマップは、複数の不連続領域（図 8 a において様々な色で示される）を識別する。各不連続領域は、タイル上のそれぞれのクラスターを表す連続するサブピクセルの非重複グループ（その配列決定画像から、かつクラスターマップがサブピクセルベースコールを介して生成される）の非重複グループを含む。不連続領域間の領域は、タイル上の背景を表す。背景領域内のサブピクセルは、「背景サブピクセル」と呼ばれる。不連続領域内のサブピクセルは、「クラスターサブピクセル」又は「クラスター内部サブピクセル」と呼ばれる。この説明では、原点サブピクセルは、RTA 又は別のベースコーラーによって決定される予備的な中心クラスター座標が位置するサブピクセルである。

【0176】

原点サブピクセルは、予備的な中心クラスター座標を含む。これは、原点サブピクセルによって覆われた領域が、予備的な中心クラスター座標位置と一致する座標位置を含むことを意味する。クラスターマップ 802 は論理サブピクセルの画像であるため、原点サブピクセルは、クラスターマップ内のサブピクセルの一部である。

【0177】

サブピクセルの基本的に一致するベースコール配列を有するクラスターを識別するための検索は、全てのサブピクセルについて検索を行うことができ、任意のサブピクセル（例えば、0, 0 サブピクセル又は任意のランダムサブピクセル）から開始することができるため、原点サブピクセル（クラスターの予備的な中心座標）の識別から始める必要はない。したがって、各サブピクセルは、実質的に一致するベースコール配列を別の連続サブピクセルと共有するかどうかを判定するために評価されるため、検索は、原点サブピクセルに依存しないため、検索は任意のサブピクセルで開始することができる。

【0178】

原点サブピクセルが使用されるか否かに関わらず、ベースコーラー 114 によって予測される原点サブピクセル（クラスターの初期中心座標）を含まない特定のクラスターが識別される。サブピクセルベースコールのマージによって識別され、原点サブピクセルを含まないクラスターのいくつかの例は、図 8 a のクラスター 812 a、812 b、812 c、812 d 及び 812 e である。したがって、開示される技術は、中心がベースコーラー 114 によって識別されていない場合がある追加の又は余分なクラスターを識別する。したがって、原点サブピクセル（クラスターの初期中心座標）を識別するためのベースコーラー 114 の使用は任意であり、連続するサブピクセルの実質的に一致するベースコール配列を検索するために必須ではない。

【0179】

一実施態様では、最初に、ベースコーラー 114 によって識別された原点サブピクセル（クラスターの初期中心座標）を使用して、（連続するサブピクセルの実質的に一致するベースコール配列を識別することによって）第 1 のクラスターセットを識別する。次いで、第 1 のクラスターセットの一部ではないサブピクセルは、（連続するサブピクセルの実質的に一致するベースコール配列を識別することによって）第 2 のクラスターセットを識別するために使用される。これにより、中心がベースコーラー 114 によって識別されない追加の又は余分なクラスターを識別するために開示された技術が可能になる。最後に、クラスターの第 1 及び第 2 のセットの一部ではないサブピクセルが背景サブピクセルとして識別される。

【0180】

図 8 b は、サブピクセルベースコールの一例を示す。図 8 b では、各配列決定サイクルは、4つの異なる波長帯域（画像／撮像チャネル）及び4つの異なる蛍光染料（各塩基について1つ）を使用して捕捉された4つの異なる画像（すなわち、A、C、T、G 画像）を有する画像セットを有する。

【0181】

この例では、画像内のピクセルを16個のサブピクセルに分割する。次いで、サブピクセルは、ベースコーラー114によって各配列決定サイクルで別々にベースコールされる。特定の配列決定サイクルで所与のサブピクセルをベースコールするために、ベースコーラー114は、4つのA、C、T、G 画像のそれぞれにおける所与のサブピクセルの強度を使用する。例えば、サイクル1の4つのA、C、T、G 画像の各々の各々においてサブピクセル1でカバーされた画像領域の強度を用いて、サイクル1でサブピクセル1をベースコールする。サブピクセル1については、これらの画像領域は、サイクル1の4つのA、C、T、G 画像のそれぞれにおける左上のピクセルの左上の1／16領域を含む。同様に、サイクルnの4つのA、C、T、G 画像の各々の各々におけるサブピクセルmでカバーされた画像領域の強度が、サイクルnでサブピクセルmをベースコールするために使用される。サブピクセルmについては、これらの画像領域は、サイクル1の4つのA、C、T、G 画像の各々におけるそれぞれの右下ピクセルの右下の1／16領域を含む。

【0182】

このプロセスは、複数の配列決定サイクルにわたってサブピクセルごとのベースコール配列116を生成する。次いで、検索器118は、連続するサブピクセルの対を評価して、それらが実質的に一致するベースコール配列を有するかどうかを判定する。はいの場合、一対のサブピクセルは、不連続領域内の同じクラスターに属するように、クラスターマップ802内に記憶される。いいえの場合、一対のサブピクセルは、同じ不連続領域に属しないように、クラスターマップ802内に記憶される。したがって、クラスターマップ802は、サブピクセルに対するベースコールが複数のサイクルにわたって実質的に一致するサブピクセルの連続セットを識別する。クラスターマップ802はしたがって、複数のクラスターからの情報を使用して、複数のクラスターを提供し、複数のクラスターの各クラスターは、単一のDNA鎖の配列データを提供することが高い信頼性を有する複数のクラスターを提供する。

【0183】

次いで、クラスターメタデータ生成器122は、クラスターマップ802を処理して、クラスターメタデータを決定することであって、それらの中心（810a）、形状、サイズ、背景、及び／又は境界を含むクラスターの空間分布を決定することを含む処理を実行する（図9）。

【0184】

いくつかの実施態様では、クラスターメタデータ生成器122は、クラスターマップ802内のサブピクセルを、非結合領域のいずれにも属さず、したがって、任意のクラスターに寄与しない、背景として識別する。このようなサブピクセルは、背景サブピクセル806a～cと呼ばれる。

【0185】

いくつかの実施態様では、クラスターマップ802は、ベースコール配列が実質的に一致しない2つの連続するサブピクセル間のクラスター境界部分808a～cを識別する。

【0186】

クラスターマップは、ニューラルネットワークベースのテンプレート生成器1512及びニューラルネットワークベースのベースコーラー1514などの分類子を訓練するためのグラウンドトゥールズとして使用するためのメモリ（例えば、クラスターマップデータストア120）に記憶される。クラスターメタデータはまた、メモリ（例えば、クラスターメタデータデータストア124）内に記憶され得る。

【0187】

図9は、クラスターの空間分布、クラスターの中心、クラスター形状、クラスターサイ

10

20

30

40

50

ズ、クラスター背景、及び／又はクラスター境界と共に、クラスターの空間分布を含むクラスターメタデータを識別するクラスターマップの別の例を示す。

(質量中心 (COM))

【0188】

図10は、クラスターマップ内の不連続領域の質量中心 (COM) がどのように計算されるかを示す。COMは、下流処理における対応するクラスターの「修正された」又は「改善された」中心として使用することができる。

【0189】

いくつかの実施態様では、クラスターごとに、質量中心計算器1004により、クラスターマップの不連続領域の質量中心を、不連続領域を形成するそれぞれの連続するサブピクセルの座標の平均として計算することによって、クラスターの超配置中心座標1006を決定する。次いで、分類子を訓練するためのグラウンドトゥールズとして使用するために、クラスター内のメモリ内のクラスターの超位置中心座標をクラスターごとに記憶する。

【0190】

いくつかの実施態様では、サブピクセル分類部がクラスターごとにクラスターマップ802の不連続領域804a~d内の質量中心サブピクセル1008をクラスターの超配置中心座標1006で識別する。

【0191】

他の別の実施態様では、クラスターマップは、補間を使用してアップサンプリングされる。アップサンプリングされたクラスターマップは、分類子を訓練するためのグラウンドトゥールズとして使用するためにメモリに記憶される。

(減衰係数及び減衰マップ)

【0192】

図11は、サブピクセルからサブピクセルが属する不連続領域の質量 (COM) の中心までのユークリッド距離に基づくサブピクセルに対する加重減衰係数の計算の一実施態様を示す。図示した別の実施態様では、加重減衰係数は、COMを含むサブピクセルに最も高い値を与え、COMから更に離れたサブピクセルについて減少する。加重減衰係数は、上述のサブピクセルベースコールから生成されたクラスターマップから、グラウンドトゥールズ減衰マップ1204を導出するために使用される。グラウンドトゥールズ減衰マップ1204は、ユニット配列を含み、配列内の各ユニットに少なくとも1つの出力値を割り当てる。いくつかの実施態様では、ユニットはサブピクセルであり、各サブピクセルは、加重減衰係数に基づいて出力値を割り当てられる。次いで、グラウンドトゥールズ減衰マップ1204は、開示されたニューラルネットワークベースのテンプレート生成器1512を訓練するためのグラウンドトゥールズとして使用される。いくつかの実施態様では、グラウンドトゥールズ減衰マップ1204からの情報もまた、開示されるニューラルネットワークベースのベースコーラー1514の入力を調製するためにも使用される。

【0193】

図12は、上述のようにサブピクセルベースコールによって生成された例示的なクラスターマップから導出された、例示的なグラウンドトゥールズ減衰マップ1204の一実施態様を示す。いくつかの実施態様では、クラスターごとにアップサンプリングされたクラスターマップにおいて、隣接するサブピクセルが属する不連続領域内の質量中心サブピクセル1104からの隣接するサブピクセルの距離1106に比例する減衰係数1102に基づいて、不連続領域内の隣接する各サブピクセルに値が割り当てられる。

【0194】

図12は、グラウンドトゥールズ減衰マップ1204を示す。一実施態様では、サブピクセル値は、ゼロと1との間で正規化された強度値である。別の実施態様では、アップサンプリングされたクラスターマップにおいて、背景として識別された全てのサブピクセルに同じ所定の値が割り当てられる。いくつかの実施態様では、所定の値はゼロ強度値である。

【0195】

いくつかの実施態様では、グラウンドトゥールース減衰マップ1204は、不連続領域内の連続するサブピクセル、及びそれらの割り当てられた値に基づいて背景として識別されるサブピクセルを表すアップサンプリングされたクラスターマップから、グラウンドトゥールース減衰マップ生成器1202によって生成される。グラウンドトゥールース減衰マップ1204は、分類子を訓練するためにグラウンドトゥールースとして使用するためにメモリに記憶される。一実施態様では、グラウンドトゥールース減衰マップ1204内の各サブピクセルは、ゼロと1との間で正規化された値を有する。

(三元(3クラス)マップ)

【0196】

図13は、クラスターマップからグラウンドトゥールース三元マップ1304を導出する一実施態様を示す。グラウンドトゥールース三元マップ1304は、ユニット配列を含み、アレイ内の各ユニットに少なくとも1つの出力値を割り当てる。名前によって、グラウンドトゥールース三元マップ1304の三元マップ実施態様は、各ユニットについて、第1の出力値が背景クラスの分類ラベル又はスコアに対応するように、アレイ内の各ユニットに3つの出力値を割り当て、第2の出力値は、クラスター中心クラスの分類ラベル又はスコアに対応し、第3の出力値は、クラスター／クラスター内部クラスの分類ラベル又はスコアに対応する。グラウンドトゥールース三元マップ1304は、ニューラルネットワークベースのテンプレート生成器1512を訓練するためのグラウンドトゥールースデータとして使用される。いくつかの実施態様では、グラウンドトゥールース三元マップ1304からの情報もまた、ニューラルネットワークベースのベースコーラー1514の入力を調製するために使用される。

【0197】

図13は、例示的なグラウンドトゥールース三元マップ1304を示す。別の実施態様では、アップサンプリングされたクラスターマップでは、不連続領域内の連続するサブピクセルは、同じクラスターに属するクラスター内部サブピクセルとして、質量中心サブピクセルをクラスター中心サブピクセルとして、及び背景サブピクセルをどのクラスターにも属しないサブピクセルとして、グラウンドトゥールース三元マップ生成器1302によってクラスターごとに分類される。いくつかの実施態様では、分類は、グラウンドトゥールース三元マップ1304に記憶される。これらの分類及びグラウンドトゥールース三元マップ1304は、分類子を訓練するためのグラウンドトゥールースとして使用するためにメモリに記憶される。

【0198】

他の別の実施態様では、クラスターごとに、クラスター内部サブピクセル、クラスター中心サブピクセル、及び背景サブピクセルの座標は、分類子を訓練するためのグラウンドトゥールースとして使用するためにメモリに記憶される。次いで、クラスターマップをアップサンプリングするために使用される因子によって座標をダウンスケールする。次いで、クラスターごとに、ダウンスケールされた座標は、分類子を訓練するためのグラウンドトゥールースとして使用するためにメモリに記憶される。

【0199】

更に他の実施態様では、グラウンドトゥールース三元マップ生成器1302は、クラスターマップを使用して、アップサンプリングされたクラスターマップから三元グラウンドトゥールースデータ1304を生成する。三元グラウンドトゥールースデータ1304は、背景クラスに属する背景サブピクセルをラベルし、クラスター中心クラスに属するクラスター中心サブピクセル、及びクラスター内部クラスに属するクラスター内部サブピクセルをラベル付けする。いくつかの可視化の実施態様では、色符号化を使用して、異なるクラスラベルを描写し、区別することができる。三元グラウンドトゥールースデータ1304は、分類子を訓練するためにグラウンドトゥールースとして使用するためにメモリに記憶される。

(バイナリ(2クラス)マップ)

【0200】

図14は、クラスターマップから、グラウンドトゥールースバイナリマップ1404を導

出する一実施態様を示す。バイナリマップ1404は、ユニット配列を含み、アレイ内の各ユニットに少なくとも1つの出力値を割り当てる。名前によって、バイナリマップは、各ユニットについて、第1の出力値がクラスター中心クラスの分類ラベル又はスコアに対応し、第2の出力値が非中心クラスの分類ラベル又はスコアに対応するように、2つの出力値をアレイ内の各ユニットに割り当てる。バイナリマップは、ニューラルネットワークベースのテンプレート生成器1512を訓練するためのグラウンドトゥールースデータとして使用される。いくつかの実施態様では、バイナリマップからの情報もまた、ニューラルネットワークベースのベースコーラー1514の入力を準備するために使用される。

【0201】

図14は、グラウンドトゥールースバイナリマップ1404を示す。グラウンドトゥールースバイナリマップ生成器1402は、クラスターマップ120を使用して、アップサンプリングされたクラスターマップからバイナリグラウンドトゥールースデータ1404を生成する。バイナリグラウンドトゥールースデータ1404は、クラスター中心サブピクセルをクラスター中心クラスに属するものとしてラベルし、他の全てのサブピクセルを非中心クラスに属するものとしてラベルする。バイナリグラウンドトゥールースデータ1404は、分類子を訓練するためにグラウンドトゥールースとして使用するためにメモリに記憶される。

【0202】

いくつかの実施態様では、開示される技術は、フローセルの複数のタイルのクラスターマップ120を生成し、クラスターマップをメモリに記憶し、それらの形状及びサイズを含むクラスターマップ120に基づいて、タイル内のクラスターの空間分布を決定する。次いで、開示された技術は、タイル内のクラスターのアップサンプリングされたクラスターマップ120において、クラスターごとにサブピクセルをクラスターごとに分類し、同じクラスターに属するクラスター内部サブピクセル、クラスター中心サブピクセル、及び背景サブピクセルに分類する。次いで、開示された技術は、分類子を訓練するためのグラウンドトゥールースとして使用するためのメモリに分類を記憶し、クラスター内のクラスターごとに、クラスター内部サブピクセルの座標、クラスター中心サブピクセル、及び分類子を訓練するためのグラウンドトゥールースとして使用するためにメモリ内に背景サブピクセルを記憶する。次いで、開示された技術は、クラスターマップをアップサンプリングするために使用される係数によって座標をダウンスケールし、クラスターごとに、分類子を訓練するためのグラウンドトゥールースとして使用するために、メモリ内のダウンスケールされた座標を記憶する。

【0203】

いくつかの実施態様では、フローセルは、クラスターを占有するウェルのアレイを有する、少なくとも1つのパターン化された表面を有する。そのような実施態様では、クラスターの決定された形状及びサイズに基づいて、開示される技術は、(1)ウェルのうちのどの1つが、少なくとも1つの群によって実質的に占有されているか、(2)ウェルのうちのどの1つが最小限に占有されているか、(3)ウェルのうちのどの1つが複数の集団によって共占有されているか、を特定する。これにより、同じウェル、すなわち、同じウェルを共有する2つ又はそれ以上のクラスターの中心、形状、及びサイズを共占する複数のクラスターのそれぞれのメタデータを決定することが可能になる。

【0204】

いくつかの実施態様では、サンプルがクラスターに増幅される固体支持体は、パターン化された表面を含む。「パターン化された表面」は、固体支持体の露出層内又はその上の異なる領域の配置を指す。例えば、1つ又はそれ以上の領域は、1つ又はそれ以上の増幅プライマーが存在する特徴であり得る。この特徴は、増幅プライマーが存在しない間質領域によって分離され得る。いくつかの実施態様では、パターンは、行及び列にある特徴のx-yフォーマットであり得る。いくつかの実施態様では、パターンは、特徴及び／又は間質領域の反復配列であり得る。いくつかの実施態様では、パターンは、特徴及び／又は間質領域のランダム配列であり得る。本明細書に記載される方法及び組成物において使用することができる例示的なパターン化表面は、米国特許第8,778,849号明細書、

10

20

30

40

50

米国特許第 9, 079, 148 号明細書、米国特許第 8, 778, 848 号明細書、及び米国特許出願公開第 2014/0243224 号明細書、に記載されており、それぞれ参照により本明細書に組み込まれる。

【0205】

いくつかの実施態様では、固体支持体は、表面にウェル又は窪みのアレイを含む。これは、フォトリソグラフィ、スタンピング技術、成形技術、及びマイクロエッチング技術を含むがこれらに限定されない様々な技術を使用して、技術分野において一般的に知られているように製造することができる。技術分野において理解されるように、使用される技術は、アレイ基板の組成及び形状に依存する。

【0206】

パターン付き表面内の特徴は、ガラス、シリコン、プラスチック、又はポリ（N-（5-アジドアセトアミルペンチル）アクリルアミド-co-アクリルアミド）（PAZAM、例えば、それぞれ、参照によりその全体が本明細書に組み込まれる米国特許出願公開第 2013/184796 号明細書、国際公開第 2016/066586 号及び同第 2015-002813 号を参照されたい）などのパターン化された共有結合ゲルを有する他の好適な固体支持体上のウェル（例えば、マイクロウェル又はナノウェル）配列におけるウェルであってもよい。このプロセスは、配列決定のために使用されるゲルパッドを作成し、これは、多数のサイクルで配列決定動作にわたって安定であり得る。ポリマーをウェルに共有結合することは、様々な用途の間に、構造化基材の寿命全体にわたってゲルを構造化特徴部に維持するのに有用である。しかしながら、多くの実施態様では、ゲルは、ウェルに共有結合される必要はない。例えば、いくつかの条件では、構造化基材の任意の部分に共有結合していない、シラン遊離アクリルアミド（Silane Free Acrylamide、SFA）（SFA、例えば、参照によりその全体が本明細書に組み込まれる米国特許第 8, 563, 477 号明細書を参照されたい）、をゲル材料として使用することができる。

【0207】

特定の別の実施態様では、構造化基材は、ウェル（例えば、マイクロウェル又はナノセル）を用いて固体支持材料をパターンニングし、パターン化された支持体をゲル材料（例えば、PAZAM、SFA、又はその化学修飾された変異体）でコーティングすることによって作製することができ、SFA（アジド-SFA）のアジド化バージョンなど、及びゲルコーティングされた支持体を、例えば化学研磨又は機械研磨によって研磨し、それによって、ウェル内にゲルを保持するが、ウェル間の構造化基材の表面上の間隙領域から実質的に全てのゲルを除去又は不活性化する。ゲル材料にプライマー核酸を付着させることができる。次いで、標的核酸（例えば、断片化されたヒトゲノム）の溶液を、個々の標的核酸が、ゲル材料に結合したプライマーとの相互作用を介して個々のウェルを種にするように、研磨された基質と接触させることができるが、標的核酸は、ゲル材料の非活性又は非活性に起因して、介在領域を占有しない。標的核酸の増幅は、介在領域内のゲルの非存在又は非活性が、増殖する核酸コロニーの外向きの移動を防止するため、ウェルに限定されるであろう。プロセスは、好都合に製造可能であり、スケール変更可能であり、マイクロ又はナノ製造方法を利用する。

【0208】

本明細書で使用するとき、用語「フローセル」は、1つ又はそれ以上の流体試薬を流通させることができる固体表面を含むチャンバを指す。本開示の方法において容易に使用することができるフローセル及び関連する流体システム及び検出プラットフォームの例は、例えば、それぞれ参照により本明細書に組み込まれる Bentley ら、Nature 456:53-59 (2008)、国際公開第 04/018497 号、米国特許第 7, 057, 026 号明細書、国際公開第 91/06678 号、同第 07/123744 号、米国特許第 7, 329, 492 号明細書、同第 7, 211, 414 号明細書、同第 7, 315, 019 号明細書、同第 7, 405, 281 号明細書、及び同第 2008/0108082 号明細書に記載されている。

【0209】

本開示全体を通して、増幅プライマーを参照するとき、用語「P5」及び「P7」が使用される。本明細書に提示される方法において、任意の好適な増幅プライマーを使用することができ、P5及びP7の使用は例示的な実施のみであることが理解されるであろう。フローセル上でのP5及びP7などの増幅プライマーの使用は、その全体が参照により本明細書に組み込まれる国際公開第2007/010251号、同第2006/064199号、同第2005/065814号、同第2015/106941号、同第1998/044151号及び同第2000/018957号の開示によって例示されるように、技術分野において既知である。例えば、任意の好適な順方向増幅プライマーは、固定化されているか又は溶液中にあるかに関わらず、相補的配列及び配列の増幅のために本明細書に提示される方法において有用であり得る。同様に、任意の好適な逆増幅プライマーは、固定化されているか又は溶液中にあるかに関わらず、相補的配列及び配列の増幅のために本明細書に提示される方法において有用であり得る。当業者であれば、本明細書に提示される核酸の捕捉及び増幅に好適なプライマー配列の設計及び使用方法を理解するであろう。

【0210】

いくつかの実施態様では、フローセルは、少なくとも1つのパターン化されていない表面を有し、クラスターは、非パターン化表面上で不均一に散乱される。

【0211】

いくつかの実施態様では、クラスターの密度は、約100,000クラスター/mm²～約1,000,000クラスター/mm²の範囲である。他の実施態様では、クラスターの密度は、約1,000,000クラスター/mm²～約10,000,000クラスター/mm²の範囲である。

【0212】

一実施態様では、ベースコーラーによって決定されたクラスターの予備中心座標は、タイルのテンプレート画像内に定義される。いくつかの実施態様では、画像座標系のピクセル解像度、画像座標系、及び測定スケールは、テンプレート画像及び画像と同じである。

【0213】

別の実施態様では、開示される技術は、フローセルのタイル上のクラスターに関するメタデータを決定することに関する。最初に、開示された技術は、(1)配列決定動作中に捕捉されたタイルの画像のセット、及び(2)ベースコーラーによって決定されたクラスターの予備中心座標にアクセスする。

【0214】

次いで、各画像セットについて、本開示の技術は、4つの塩基のうちの1つとして、(1)予備中心座標を含む原点サブピクセルと、(2)原点サブピクセルのそれぞれに連続的に連続している連続するサブピクセルの所定の近傍を取得する。これにより、原点サブピクセルのそれぞれ、及び連続するサブピクセルの所定の近傍のそれぞれに対して、ベースコール配列を生成する。連続するサブピクセルの所定の近傍は、原点サブピクセルを含むサブピクセルを中心とするm×nサブピクセルパッチとすることができる。一実施態様では、サブピクセルパッチは、3×3サブピクセルである。他の実施態様では、その画像パッチは、5×5、15×15、20×20などの任意のサイズであり得る。他の実施態様では、連続するサブピクセルの所定の近傍は、原点サブピクセルを含むサブピクセルを中心とするn個の接続されたサブピクセル近傍であり得る。

【0215】

一実施態様では、開示された技術は、非結合領域のいずれにも属しないクラスターマップ内のサブピクセルを背景として識別する。

【0216】

次に、開示される技術は、隣接するサブピクセルの不連続領域としてクラスターを識別するクラスターマップを生成し、そのクラスターマップは、(a)原点サブピクセルのうちの対応する1つの少なくとも一部に連続的に連続しており、(b)4つの塩基のうちの1つの実質的に一致するベースコール配列を、原点サブピクセルのうちの対応する1つの少なくとも一部と共有する。

【0217】

開示された技術は、次いで、クラスターマップをメモリに記憶し、クラスターマップ内の不連続領域に基づいてクラスターの形状及びサイズを決定する。他の実施態様では、クラスターの中心も決定される。

(テンプレート生成器のための訓練データの生成)

【0218】

図15は、ニューラルネットワークベースのテンプレート生成器1512及びニューラルネットワークベースのベースコーラー1514を訓練するために使用される訓練データを生成する一実施態様を示すブロック図である。

【0219】

図16は、ニューラルネットワークベースのテンプレート生成器1512及びニューラルネットワークベースのベースコーラー1514を訓練するために使用される開示された訓練例の特性を示す。各訓練例はタイルに対応し、対応するグラウンドトゥールースデータ表現でラベル付けされる。いくつかの実施態様では、グラウンドトゥールースデータ表現は、グラウンドトゥールース減衰マップ1204、グラウンドトゥールース三元マップ1304、又はグラウンドトゥールースバイナリマップ1404の形態のグラウンドトゥールースクラスターメタデータを識別するグラウンドトゥールースマスク又はグラウンドトゥールースマップである。いくつかの実施態様では、複数の訓練例は、同じタイルに対応する。

【0220】

一実施態様では、開示される技術は、ニューラルネットワークベースのテンプレート生成及びベースコーラーのための訓練データ1504を生成することに関する。最初に、開示された技術は、配列決定動作の複数のサイクルにわたって捕捉されたフローセル202の多数の画像108にアクセスする。フローセル202は、複数のタイルを有する。多数の画像108において、タイルのそれぞれは、複数のサイクルにわたって生成された一連の画像セットを有する。画像セット108の配列内の各画像は、特定の1つのサイクルにおける、タイルの特定の1つのクラスター302及びそれらの周囲の背景304の強度放射を示す。

【0221】

次いで、訓練セットコンストラクタ1502は、複数の訓練例を有する訓練セット1504を構築する。図16に示されるように、各訓練例は、タイルのうちの特定の1つに対応し、タイルのうちの特定の1つの画像セット1602の配列内の少なくとも一部の画像セットからの画像データを含む。一実施態様では、画像データは、タイルのうちの特定の1つの画像セット1602の配列内の少なくともいくつかの画像セット内の画像を含む。例えば、画像は、1800×1800の解像度を有し得る。他の実施態様形態では、100×100、3000×3000、10000×10000などの任意の解像度であり得る。更に他の実施態様では、画像データは、画像のそれぞれからの少なくとも1つの画像パッチを含む。一実施態様では、画像パッチは、タイルの特定の1つの部分を被覆する。一例では、画像パッチは、20×20の解像度を有し得る。他の実施態様形態では、画像パッチは、50×50、70×70、90×90、100×100、3000×3000、10000×10000などの任意の解像度を有することができる。

【0222】

いくつかの実施態様では、画像データは、画像パッチのアップサンプリングされた表現を含む。アップサンプリングされた表現は、例えば、80×80の解像度を有することができる。他の実施例では、アップサンプリングされた表現は、50×50、70×70、90×90、100×100、3000×3000、10000×10000などの任意の解像度を有することができる。

【0223】

いくつかの実施例では、複数の訓練例は、タイルのうちの同じ特定の1つに対応し、それぞれ、タイルのうちの同じ特定の1つの画像セット1602の配列内の少なくとも一部の画像セットのそれぞれの画像から異なる画像パッチをそれぞれ含む。このような実施態

10

20

30

40

50

様では、異なる画像パッチのうちの少なくとも一部は、互いに重複する。

【0224】

次いで、グラウンドトゥルス発生器1506は、訓練例のそれぞれに対して、少なくとも1つのグラウンドトゥルスデータ表現を生成する。グラウンドトゥルスデータ表現は、クラスターの空間分布、及びクラスターの形状、クラスターサイズ、及び／又はクラスター境界、及び／又はクラスターの中心のうちの少なくとも1つを含む、画像データによって表される、クラスターの空間分布及びそれらの周囲の背景のうちの少なくとも1つを識別する。

【0225】

一実施態様では、グラウンドトゥルスデータ表現は、隣接するサブピクセルの不連続領域としてクラスターを識別し、クラスターの中心は、不連続領域のうちの対応する領域内の質量サブピクセルの中心としてのクラスターの中心、及びそれらの周囲の背景として、そのクラスターを識別する。

【0226】

一実施態様では、グラウンドトゥルスデータ表現は、80×80のアップサンプリング解像度を有する。他の実施態様では、グラウンドトゥルスデータ表現は、50×50、70×70、90×90、100×100、3000×3000、10000×10000などの任意の解像度を有することができる。

【0227】

一実施態様では、グラウンドトゥルスデータ表現は、クラスター中心又は非中心であるかのいずれかとして、各サブピクセルを識別する。別の実施態様では、グラウンドトゥルスデータ表現は、クラスター内部、クラスター中心、又は周囲背景であるかのいずれかとして、各サブピクセルを識別する。

【0228】

いくつかの実施態様では、開示された技術は、ニューラルネットワークベースのテンプレート生成器1512及びニューラルネットワークベースのベースコーラー1514を訓練するための訓練データ1504として、訓練セット1504及び関連するグラウンドトゥルスデータ1508をメモリ内に記憶する。訓練は、訓練器1510によって操作される。

【0229】

いくつかの実施態様では、開示される技術は、様々なフローセル、配列決定機器、配列決定プロトコル、配列決定ケミストリー、配列決定試薬、及びクラスター密度に関する訓練データを生成する。

(ニューラルネットワークベースのテンプレート生成器)

【0230】

推測又は製造の実施態様において、開示される技術は、クラスターメタデータを決定するためにピーク検出及び分割を使用する。本開示の技術は、入力画像データ1702の代替表現1708を生成するために、ニューラルネットワーク1706を介して一連の画像セット1602から導出された入力画像データ1702を処理する。例えば、画像セットは特定の配列決定サイクル用であり、各画像チャネルA、C、T、及びGに1つずつ、合計4つの画像を含めることができる。したがって、50回の配列決定サイクルを有する配列決定動作の場合、そのような画像セットは50個、つまり合計200個の画像になる。時間的に配置されると、画像セット当たり4つの画像セットを有する画像セットが一連の画像セット1602を形成する。いくつかの実施態様では、特定のサイズの画像パッチが、50枚の画像セット内の各画像から抽出され、画像パッチセット当たり4つの画像パッチセットを形成し、一実施態様では、これは入力画像データ1702である。他の実施態様では、入力画像データ1702は、50回の配列決定サイクル未満、すなわち、1回、2回、3回、15回、20回の配列決定サイクルよりも少ない画像パッチセットに対して、画像パッチセットごとに4つの画像パッチを有する画像パッチセットを含む。

【0231】

図17は、ニューラルネットワークベースのテンプレート生成器1512を介して入力画像データ1702を処理し、アレイ内の各ユニットの出力値を生成する一実施態様を示す。一実施態様では、アレイは減衰マップ1716である。別の実施態様では、アレイは三元マップ1718である。更に別の実施態様では、アレイはバイナリマップ1720である。したがって、アレイは、入力画像データ1702内に表される複数の場所のそれぞれの1つ又はそれ以上の特性を表し得る。

【0232】

グラウンドトゥールース減衰マップ1204、グラウンドトゥールース三元マップ1304、及びグラウンドトゥールースバイナリマップ1404を含む、先の図の構造を使用してテンプレート生成器を訓練することとは異なり、減衰マップ1716、三元マップ1718及び／又はバイナリマップ1720は、訓練されたニューラルネットワークベースのテンプレート生成器1512の前方伝搬によって生成される。前方伝搬は、訓練中又は推測中であり得る。訓練中、逆方向伝搬ベースの勾配更新により、減衰マップ1716、三元マップ1718及びバイナリマップ1720（すなわち、累積的に出力1714）は、グラウンドトゥールース減衰マップ1204、グラウンドトゥールース三元マップ1304、及びグラウンドトゥールースバイナリマップ1404にそれぞれ漸進的に一致又は接近する。

【0233】

推測中に分析される画像アレイのサイズは、一実施態様によれば、入力画像データ1702のサイズに依存する（例えば、同じ又はアップスケールされた又はダウンスケールされたバージョンである）。各ユニットは、ピクセル、サブピクセル、又はスーパーピクセルを表すことができる。アレイの単位ごとの出力値は、減衰マップ1716、三元マップ1718、又はバイナリマップ1720を特徴付ける／表す／示すことができる。いくつかの実施態様では、入力画像データ1702はまた、ピクセル解像度、サブピクセル解像度、又はスーパーピクセル解像度のユニット配列である。そのような別の実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、入力アレイ内の各ユニットの出力値を生成するために、意味的セグメンテーション技術を使用する。入力画像データ1702に関する更なる詳細は、図21b、22、23、及び24及びそれらの考察において見出すことができる。

【0234】

いくつかの実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、参照により本明細書に組み込まれる、J. Long, E. Shelhamer, and T. Darrell, 「Fully convolutional networks for semantic segmentation」、CVPR、(2015)に記載されているものなどの完全な畳み込みネットワークである。他の実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、http://link.springer.com/chapter/10.1007/978-3-319-24574-4_28で入手可能であり、参照により本明細書に組み込まれる、Ronneberger O, Fischer P, Brox T., 「U-net: Convolutional networks for biomedical image segmentation」 Med. Image Comput. Comput. Assist. Interv. (2015)に記載されているものなど、デコーダとエンコーダとの間のデコーダとエンコーダとの間のスキップ接続を有するU-Netネットワークである。U-Net構造は、以下の2つの主要なサブ構造を有する自動エンコーダに似ている。1) 入力画像を取り込み、複数の畳み込み層を介してその空間解像度を低減して、符号化を生成するエンコーダと、を備える、システム。2) 出力として再構成された画像を生成するために、空間解像度を符号化し、増大させる表現をとるデコーダ。U-Netは、この構造に2つの革新を導入する：最初に、目的関数は、損失関数を使用して分割マスクを再構成するように設定され、第2に、エンコーダの畳み込み層は、スキップ接続を使用して、デコーダ内の同じ解像度の対応する層に接続される。更に更なる実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、エンコーダサブネットワーク及

10

20

30

40

50

び対応するデコーダネットワークを有する深層完全畳み込み分割ニューラルネットワークである。そのような別の実施態様では、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを完全入力解像度特徴マップにマッピングするデコーダの階層を含む。分割ネットワークに関する更なる詳細は、「Segmentation Networks」と題された付録に見出すことができる。

【0235】

一実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、畳み込みニューラルネットワークである。別の実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、反復ニューラルネットワークである。更に別の実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、残留ブロック及び残留接続を有する残留ニューラルネットワークである。更なる実施態様では、ニューラルネットワークベースのテンプレート生成器1512は、畳み込みニューラルネットワークと反復ニューラルネットワークとの組み合わせである。

【0236】

ニューラルネットワークベースのテンプレート生成器1512（すなわち、ニューラルネットワーク1706及び／又は出力層1710）は、様々なパディング及びストリディング構成を使用することができることを理解するであろう。それは、異なる出力機能（例えば、分類又は回帰）を使用することができる、1つ又はそれ以上の完全に接続された層を含んでも含まなくてもよい。それは、1D重畳、2D重畳、3D重畳、4D重畳、5D重畳、拡張又は無性重畳、転置重畳、奥行分離可能な重畳、1×1重畳、グループ重畳、扁平重畳、空間及びクロスチャネルの重畳、シャッフルされたグループ化重畳、空間的な分離可能な重畳、及び逆重畳を使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はゲート反復ユニット（Gated Recurrent Unit、GRU）など）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（Rectifying Linear Unit、ReLU）、漏れやすいReLU、指数関数的ライナーユニット（Exponential Liner Unit、ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

【0237】

いくつかの実施態様では、画像セット1602の配列内の各画像はタイルを覆い、タイル上のクラスターの強度放射を示し、フローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の撮像チャンネルのために捕捉された、それらの周囲の背景を示す。一実施態様では、入力画像データ1702は、画像セット1602の配列内の画像のそれぞれからの少なくとも1つの画像パッチを含む。このような別の実施態様では、画像パッチはタイルの一部を覆う。一例では、画像パッチは、20×20の解像度を有する。他の場合には、画像パッチの解像度は、20×20から10000×10000の範囲であり得る。別の実施態様では、入力画像データ1702は、画像セット1602の配列内の画像のそれぞれからの画像パッチのアップサンプリングされたサブピクセル解像度表現を含む。一実施例では、アップサンプリングされたサブピクセル表現は、80×80の解像度を有する。他の場合には、アップサンプリングされたサブピクセル表現の解像度は、80×80から10000×10000の範囲であり得る。

10

20

30

40

50

【0238】

入力画像データ1702は、クラスター及びそれらの周囲背景を描写する単位1704のアレイを有する。例えば、画像セットは特定の配列決定サイクル用であり、各画像チャネルA、C、T、及びGに1つずつ、合計4つの画像を含めることができる。したがって、50回の配列決定サイクルを有する配列決定動作の場合、そのような画像セットは50個、つまり合計200個の画像になる。時間的に配置されると、画像セット当たり4つの画像セットを有する画像セットが一連の画像セット1602を形成する。いくつかの実施態様では、特定のサイズの画像パッチが、50枚の画像セット内の各画像から抽出され、画像パッチセット当たり4つの画像パッチセットを形成し、一実施態様では、これは入力画像データ1702である。他の実施例では、入力画像データ1702は、50回の配列決定サイクル未満、すなわち、1回、2回、3回、15回、20回の配列決定サイクルよりも少ない画像パッチごと画像パッチセットを有する画像パッチセットを含む。代替表現は、特徴マップである。特徴マップは、ニューラルネットワークが畳み込みニューラルネットワークである場合、畳み込み特徴又は畳み込み表現であり得る。特徴マップは、ニューラルネットワークが反復ニューラルネットワークであるとき、隠れた状態特徴又は隠れた状態表現であり得る。

10

【0239】

次に、開示された技術は、出力層1710を介して代替表現1708を処理して、アレイ1704内の各ユニットに対する出力値1712を有する出力1714を生成する。出力層は、単位ごとの出力値を生成するソフトマックス又はシグモイドなどの分類層であり得る。一実施態様では、出力層は、単位ごとの出力値を生成するReLU層又は任意の他の起動機能層である。

20

【0240】

一実施態様では、入力画像データ1702内のユニットはピクセルであり、したがって、出力1714においてピクセルごとの出力値1712が生成される。別の実施態様では、入力画像データ1702内の単位はサブピクセルであり、したがって、サブピクセルごとの出力値1712が出力部1714において生成される。更に別の実施態様では、入力画像データ1702内のユニットはスーパーピクセルであり、したがってスーパーピクセルごとの出力値1712が出力部1714において生成される。

(減衰マップ、三元マップ及び／又はバイナリマップからのクラスターメタデータの導出)

30

【0241】

図18は、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を含むクラスターメタデータを導出するために、ニューラルネットワークベースのテンプレート生成器1512によって生成された減衰マップ1716、三元マップ1718、又はバイナリマップ1720に適用される後処理技術の一実施態様を示す。いくつかの実施態様では、後処理技術は、閾値保持器1802、ピークローケータ1806、及びセグメント化器1810を更に含むポストプロセッサ1814によって適用される。

【0242】

40

閾値化器1802への入力、開示されるニューラルネットワークベースのテンプレート生成器などのテンプレート生成器1512によって生成される、減衰マップ1716、三元マップ1718、又はバイナリマップ1720である。一実施態様では、閾値化器1802は、減衰マップ、三元マップ、又はバイナリマップ内の値に閾値を適用して、背景ユニット1804（すなわち、非クラスター背景を特徴付けるサブピクセル）、及び非背景ユニットを識別する。別の言い方をすれば、出力1714が生成されると、閾値化器1802は、ユニット1712の出力値に閾値を適用し、クラスターの周囲の背景を描写する「背景ユニット」1804及びクラスターに属する可能性のあるユニットを表す「非背景ユニット」としてユニット1712の第1のサブセットを分類するか、又は再分類することができる。閾値化器1802によって適用される閾値は、予め設定され得る。

50

【0243】

ピークロケータ1806への入力はまだ、ニューラルネットワークベースのテンプレート生成器1512によって生成される、減衰マップ1716、三元マップ1718、又はバイナリマップ1720である。一実施態様では、ピークロケータ1806は、減衰マップ1716内の値のピーク検出を、三元マップ1718、又はバイナリマップ1720に適用して、中心ユニット1808（すなわち、クラスター中心を特徴付ける中心サブピクセル）を識別する。言い換えれば、ピークロケータ1806は、出力1714内のユニット1712の出力値を処理し、クラスターの中心を含む「中心ユニット」1808としてユニット1712の第2のサブセットを分類する。いくつかの実施態様では、ピークロケータ1806によって検出されるクラスターの中心もまた、クラスターの質量中心である。次いで、中心ユニット1808は、セグメント化器1810に提供される。ピークロケータ1806に関する更なる詳細は、「ピーク検出」と題された付録に見出すことができる。

10

【0244】

閾値及びピーク検出は、並行して、又は他方の後に行うことができる。すなわち、それらは互いに依存しない。

【0245】

セグメント化器1810への入力はまだ、ニューラルネットワークベースのテンプレート生成器1512によって生成される、減衰マップ1716、三元マップ1718、又はバイナリマップ1720でもある。セグメント化器1810への追加の補足入力は、閾値化器1802によって識別された閾値化ユニット（背景、非背景）1804と、ピークロケータ1806によって識別された中心ユニット1808とを含む。セグメント化器1810は、背景、非背景1804、及び中心ユニット1808を使用して、不連続領域1812（すなわち、クラスターを特徴付ける隣接するクラスター／クラスター内部サブピクセルの非重複グループ）を識別する。言い換えれば、セグメント化器1810は、出力1714内のユニット1712の出力値を処理し、背景ユニット1804によって分離され、中心ユニット1808を中心とする連続ユニットの非重複領域としてクラスターの形状1812を決定するために、背景及び非背景ユニット1804、並びに中心ユニット1808を使用する。セグメント化器1810の出力は、クラスターメタデータ1812である。クラスターメタデータ1812は、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を識別する。

20

30

【0246】

一実施態様では、セグメント化器1810は、中心ユニット1808から始まり、各中心ユニットに関して、質量中心が中心ユニットに含まれる同じクラスターを示す連続的に連続するユニット群を決定する。一実施態様では、セグメント化器1810は、いわゆる「流域」セグメント化技術を使用して、連続するクラスターを、強度の谷で複数の隣接するクラスターに細分化する。流域分割技術及び他の分割技術に関する更なる詳細は、「Watershed Segmentation」と題された付録に見出すことができる。

【0247】

一実施態様では、出力1714内のユニット1712の出力値は、グラウンドトゥルス減衰マップ1204内で符号化されたものなどの連続値である。別の実施態様では、出力値は、グラウンドトゥルス三元マップ1304及びグラウンドトゥルスバイナリマップ1404にコードされているものなどのソフトマックススコアである。一実施態様に係るグラウンドトゥルス減衰マップ1204では、非重複領域のうちの対応する領域内の連続単位は、隣接するユニットが属する非重複領域内の中心ユニットからの連続ユニットの距離に従って重み付けされた出力値を有する。そのような実施態様では、中心ユニットは、非重複領域のうちのそれぞれの領域内で最も高い出力値を有する。上述したように、訓練中、後方伝搬ベースの勾配更新により、減衰マップ1716、三元マップ1718及びバイナリマップ1720（すなわち、累積的に出力1714）は、グラウンドトゥルス減衰マップ1204のグラウンドトゥルス三元マップ1304とグラウンドトゥル

40

50

ースバイナリマップ1404とをそれぞれ漸進的に一致又は接近させる。

(ピクセルドメインー規則的なクラスター形状からの強度抽出)

【0248】

ここで説明は、開示された技術によって決定されたクラスター形状を、クラスターの強度を抽出するために使用することができるかについて説明する。クラスターは典型的に不規則な形状及び輪郭を有するため、開示される技術は、どのサブピクセルがクラスター形状を表す不規則な形状の不連続領域に寄与するかを識別するために使用することができる。

【0249】

図19は、ピクセルドメイン内のクラスター強度を抽出する一実施態様を示す。「テンプレート画像」又は「テンプレート」は、減衰マップ1716、三元マップ1718及び／又はバイナリマップ1718に由来するクラスターメタデータ1812を含むか、又は識別するデータ構造を指すことができる。クラスターメタデータ1812は、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を識別する。

【0250】

いくつかの実施態様では、テンプレート画像は、アップサンプリングされたサブピクセルドメイン内にあり、クラスター境界を微細化レベルで区別する。しかしながら、クラスター及び背景強度データを含む配列決定画像108は、典型的には、ピクセルドメイン内にある。したがって、開示される技術は、アップサンプリングされたサブピクセル解像度内のテンプレート画像に符号化されたクラスター形状情報を使用して、光学的なピクセル解像度配列決定画像から不規則形状のクラスターの強度を抽出する2つの手法を提案している。図19に示される第1のアプローチでは、テンプレート画像内で識別された連続するサブピクセルの非重複グループは、ピクセル解像度配列決定画像内に位置し、それらの強度は補間によって抽出される。この強度抽出技術に関する更なる詳細は、図33及びその考察において見出すことができる。

【0251】

一実施態様では、非重複領域が不規則な輪郭を有し、ユニットがサブピクセルである場合、所与のクラスターのクラスター強度1912は、以下のように強度抽出器1902によって決定される。

【0252】

まず、サブピクセルロケータ1904は、所与のクラスターの形状を識別する隣接するサブピクセルの対応する非重複領域に基づいて、所与のクラスターのクラスター強度に寄与するサブピクセルを識別する。

【0253】

次に、サブピクセルロケータ1904は、現在の配列決定サイクルで1つ又はそれ以上の撮像チャンネルに対して生成された1つ又はそれ以上の光学ピクセル解像度画像1918内に識別されたサブピクセルを位置させる。一実施態様では、整数又は非整数座標（例えば、フローティングポイント）は、サブピクセルドメインを作成するために使用されるアップサンプリング係数に一致するダウンスケール係数に基づいてダウンスケールした後に、光学解像度画像、ピクセル解像度画像内に位置する。

【0254】

次いで、処理された画像内の識別されたサブピクセルの補間器及びサブピクセル強度結合器1906は、補間強度を組み合わせ、組み合わせられた補間強度を正規化して、画像のそれぞれにおける所与のクラスターのための画像ごとのクラスター強度を生成する。正規化は、正規化器1908によって実行され、正規化係数に基づく。一実施態様では、正規化係数は、識別されたサブピクセルの数である。これは、フローセル上のそれらの場所に応じてクラスターが受信する異なるクラスターサイズ及び不均一な照明を正規化／考慮するために行われる。

【0255】

最後に、クロスチャネルサブピクセル強度累算器1910は、画像のそれぞれに関する

10

20

30

40

50

画像ごとのクラスター強度を組み合わせ、現在の配列決定サイクルにおいて所与のクラスターのクラスター強度 1 9 1 2 を決定する。

【0 2 5 6】

次いで、所与のクラスターは、本出願で論じられたベースコールのうちのいずれか 1 つによって現在の配列決定サイクルでクラスター強度 1 9 1 2 に基づいてベースコールされ、ベースコール 1 9 1 6 を生成する。

【0 2 5 7】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー 1 5 1 4、すなわち、減衰マップ 1 7 1 6、三元マップ 1 7 1 8 及びバイナリマップ 1 7 2 0 の出力は、光学的ピクセルドメイン内にある。したがって、このような実施態様形態では、テンプレート画像はまた、光ピクセルドメイン内にある。

(サブピクセルドメインー規則的なクラスター形状からの強度抽出)

【0 2 5 8】

図 2 0 は、サブピクセルドメイン内のクラスター強度を抽出する第 2 のアプローチを示す。この第 2 のアプローチでは、光学的に配列決定画像、ピクセル解像度をサブピクセル解像度にアップサンプリングする。これにより、テンプレート画像内の「サブピクセルを描くクラスター形状」と、アップサンプリングされた配列決定画像における「サブピクセルを示すクラスター強度」との対応をもたらす。次いで、クラスター強度は、対応に基づいて抽出される。この強度抽出技術に関する更なる詳細は、図 3 3 及びその考察において見出すことができる。

【0 2 5 9】

一実施態様では、非重複領域が不規則な輪郭を有し、ユニットがサブピクセルである場合、所与のクラスターのクラスター強度 2 0 1 2 は、以下のように強度抽出器 2 0 0 2 によって決定される。

【0 2 6 0】

まず、サブピクセルロケータ 2 0 0 4 は、所与のクラスターの形状を識別する隣接するサブピクセルの対応する非重複領域に基づいて、所与のクラスターのクラスター強度に寄与するサブピクセルを識別する。

【0 2 6 1】

次いで、サブピクセルロケータ 2 0 0 4 は、現在の配列決定サイクルで 1 つ又はそれ以上の撮像チャネルのために生成された対応する光学的なピクセル解像度画像 1 9 1 8 からアップサンプリングされた 1 つ又はそれ以上のサブピクセル解像度画像 2 0 1 8 内に、識別されたサブピクセルを位置させる。アップサンプリングは、最近傍強度抽出、ガウス系強度抽出、平均 2×2 サブピクセル面積に基づく強度抽出、 2×2 サブピクセル面積の最も明るい試験に基づく強度抽出、平均 3×3 サブピクセル領域、バイリニア強度抽出、双次強度抽出、及び／又は加重領域被覆に基づく強度抽出により実行され得る。これらの技術は、「強度抽出方法」と題された付録に詳細に記載されている。テンプレート画像は、いくつかの実施態様では、強度抽出のためのマスクとして機能することができる。

【0 2 6 2】

次いで、アップサンプリングされた画像のそれぞれにおけるサブピクセル強度結合器 2 0 0 6 は、識別されたサブピクセルの強度を組み合わせ、組み合わせられた強度を正規化して、アップサンプリングされた画像のそれぞれにおける所与のクラスターのための画像ごとのクラスター強度を生成する。正規化は、正規化器 2 0 0 8 によって実行され、正規化係数に基づく。一実施態様では、正規化係数は、識別されたサブピクセルの数である。これは、フローセル上のそれらの場所に応じてクラスターが受信する異なるクラスターサイズ及び不均一な照明を正規化／考慮するために行われる。

【0 2 6 3】

最後に、クロスチャネルサブピクセル強度累算器 2 0 1 0 は、アップサンプリングされた画像のそれぞれについて、画像ごとのクラスター強度を組み合わせ、現在の配列決定

10

20

30

40

50

サイクルにおいて所与のクラスターのクラスター強度 2 0 1 2 を決定する。

【0 2 6 4】

次いで、所与のクラスターは、本出願で論じられたベースコールのうちのいずれか 1 つによって現在の配列決定サイクルでクラスター強度 2 0 1 2 に基づいてベースコールされ、ベースコール 2 0 1 6 を生成する。

(ニューラルネットワークベースのテンプレート生成器の種類)

【0 2 6 5】

ここでの考察は、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 の 3 つの異なる実施態様の詳細を説明する。図 2 1 a に示されており、(1) 減衰マップベースのテンプレート生成器 2 6 0 0 (回帰モデルとも呼ばれる)、(2) バイナリマップベースのテンプレート生成器 4 6 0 0 (バイナリ分類モデルとも呼ばれる)、及び (3) 三元マップベースのテンプレート生成器 5 4 0 0 (三元分類モデルとも呼ばれる) と、を含む。

10

【0 2 6 6】

一実施態様では、回帰モデル 2 6 0 0 は完全な畳み込みネットワークである。別の実施態様では、回帰モデル 2 6 0 0 は、デコーダとエンコーダとの間のスキップ接続を有する U-Net ネットワークである。一実施態様では、バイナリ分類モデル 4 6 0 0 は、完全な畳み込みネットワークである。別の実施態様では、バイナリ分類モデル 4 6 0 0 は、デコーダとエンコーダとの間のスキップ接続を有する U-Net ネットワークである。一実施態様では、三元分類モデル 5 4 0 0 は完全な畳み込みネットワークである。別の実施態様では、三元分類モデル 5 4 0 0 は、デコーダとエンコーダとの間のスキップ接続を有する U-Net ネットワークである。

20

(入力画像データ)

【0 2 6 7】

図 2 1 b は、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 への入力として供給される入力画像データ 1 7 0 2 の一実施態様を示す。入力画像データ 1 7 0 2 は、配列決定動作 (例えば、最初の 2 ~ 7 回の配列決定サイクル) の特定の数の初期配列決定サイクルの間に生成される配列決定画像 1 0 8 を有する一連の画像セット 2 1 0 0 を含む。

【0 2 6 8】

いくつかの実施態様では、配列決定画像 1 0 8 の強度は、背景について補正され、及び/又は親和性変換を用いて互いに整列される。一実施態様では、配列決定動作は 4 つのチャンネル化学を利用し、各画像セットは 4 つの画像を有する。別の実施態様では、配列決定動作は 2 チャンネル化学を利用し、各画像セットは 2 つの画像を有する。更に別の実施態様では、配列決定動作は、1 チャンネル化学を利用し、各画像セットは 2 つの画像を有する。更に他の実施態様では、各画像セットは 1 つの画像のみを有する。これら及び他の異なる実施態様は、付録 6 及び 9 に記載されている。

30

【0 2 6 9】

一連の画像セット 2 1 0 0 内の各画像 2 1 1 6 は、フローセル 2 1 0 2 のタイル 2 1 0 4 を覆い、タイル 2 1 0 4 上のクラスター 2 1 0 6 の強度放射、及び配列決定動作の複数の配列決定サイクルのうちの特定の 1 つで特定の画像チャンネルのために捕捉されたそれらの周囲の背景を示す。一実施例では、サイクル t 1 に関して、画像セットは、対応する蛍光染料で標識化され、対応する波長帯 (画像/撮像チャンネル) で撮像された各塩基 A、C、T、及び G 用の 1 つの画像を含む、4 つの画像 2 1 1 2 A、2 1 1 2 C、2 1 1 2 T、2 1 1 2 G を含む。

40

【0 2 7 0】

例示目的のために、画像 2 1 1 2 G では、図 2 1 b は、2 1 0 8 としてのクラスター強度放射及び 2 1 1 0 としての背景強度放射を示す。別の実施例では、サイクル t n に関して、画像セットはまた、対応する蛍光染料で標識化され、対応する波長帯 (画像/撮像チャンネル) で撮像された各塩基 A、C、T、及び G 用の 1 つの画像を含む、4 つの画像 2 1 1 4 A、2 1 1 4 C、2 1 1 4 T、2 1 1 4 G を含む。また、例示目的のために、画像 2

50

1 1 4 Aにおいて、図 2 1 bは、2 1 1 8としてクラスター強度放射を示し、画像 2 1 1 4 Tでは、背景強度放射を2 1 2 0として示す。

(非画像データ)

【0 2 7 1】

入力画像データ1 7 0 2は、強度チャンネル（撮像チャンネルとも呼ばれる）を使用して符号化される。特定の配列決定サイクルのためにシーケンサから取得されたc画像のそれぞれについて、別個の画像化チャンネルを使用して、その強度信号データを符号化する。例えば、配列決定動作が、各配列決定サイクルにおいて赤色画像及び緑色画像を生成する2チャンネル化学を使用すると考える。そのような場合、入力データ2 6 3 2は、(i)赤色画像内に捕捉された1つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示す、 $w \times h$ ピクセルを有する第1の赤色画像化チャンネルと、(ii)1つ又はそれ以上のクラスターの強度放射及び緑色画像内に捕捉されたそれらの周囲背景の強度放射を示す、 $w \times h$ ピクセルを有する第2の緑色画像化チャンネルと、を含む。

10

【0 2 7 2】

別の実施態様では、画像データは、ニューラルネットワークベースのテンプレート生成器1 5 1 2又はニューラルネットワークベースのベースコーラー1 5 1 4への入力として使用されない。その代わりに、ニューラルネットワークベースのテンプレート生成器1 5 1 2及びニューラルネットワークベースのベースコーラー1 5 1 4への入力は、分子延長中の水素イオンの放出によって誘発されるpH変化に基づく。pH変化は検出され、組み込まれた塩基の数に比例する電圧変化に変換される（例えば、Ion Torrentの場合）。

20

【0 2 7 3】

更に別の実施態様では、ニューラルネットワークベースのテンプレート生成器1 5 1 2及びニューラルネットワークベースのベースコーラー1 5 1 4への入力は、塩基の同一性を決定しながら、バイオセンサーを使用して、検体がナノ細孔を通過するとき、又はその開口部付近を通過する際に、電流の破壊を測定するためにバイオセンサーを使用するナノ細孔検知から構築される。例えば、Oxford Nanopore Technologies (ONT) 配列決定は、以下の概念に基づく：ナノ細孔を介して膜を介してDNA（又はRNA）の単鎖を通過させ、膜にわたって電位差を印加する。細孔内に存在するヌクレオチドは、細孔の電気抵抗に影響を及ぼし、そのため、経時的な電流測定は、細孔を通過するDNA塩基の配列を示すことができる。この電流信号（プロットされたときにその外観に起因する「押しつぶし」）は、ONTシーケンサによって収集された生データである。これらの測定値は、4 kHz周波数（例えば）で取られた16ビットの整数データ取得（Data Acquisition、DAC）値として記憶される。1秒当たり〜450塩基対のDNA鎖速度を用いて、これは、平均して、塩基当たり約9つの生観察を与える。次いで、この信号を処理して、個々の読み取りに対応する開孔信号の破断を識別する。これらの生信号の伸長は、ベースコールされ、DAC値をDNA塩基の配列に変換するプロセスである。いくつかの実施態様では、入力データ2 6 3 2は、正規化又はスケーリングされたDAC値を含む。

30

(パッチ抽出)

【0 2 7 4】

図 2 2は、入力画像データ1 7 0 2を形成する一連の「ダウンサイズの」画像セットを生成するために、図 2 1 bの一連の画像セット2 1 0 0からパッチを抽出する一実施態様を示す。図示の別の実施態様では、一連の画像セット2 1 0 0内の配列決定画像1 0 8は、サイズ $L \times L$ （例えば、2 0 0 0 $\times$ 2 0 0 0）である。他の実施態様では、 L は、1から10,000の範囲の任意の数である。

40

【0 2 7 5】

一実施態様では、パッチ抽出器2 2 0 2は、一連の画像セット2 1 0 0内の配列決定画像1 0 8からパッチを抽出し、一連のダウンサイズの画像セット2 2 0 6、2 2 0 8、2 2 1 0及び2 2 1 2を生成する。一連のダウンサイズ画像セット内の各画像は、一連の画

50

像セット 2100 内の対応する配列決定画像から抽出されるサイズ $M \times M$ (例えば、 20×20) のパッチである。パッチのサイズは予め設定することができる。他の別の実施態様では、 M は $1 \sim 1000$ の範囲の任意の数である。

【0276】

図 22 では、4 つの例示的な一連のダウンサイズ画像セットが示されている。第 1 の例示的な一連のダウンサイズ画像セット 2206 は、一連の画像セット 2100 内の配列決定画像 108 内の座標 $0, 0 \sim 20, 20$ から抽出される。第 2 の例示的な一連のダウンサイズ画像セット 2208 は、一連の画像セット 2100 内の配列決定画像 108 内の座標 $20, 20 \sim 40, 40$ から抽出される。第 3 の例示的な一連のダウンサイズ画像セット 2210 は、一連の画像セット 2100 内の配列決定画像 108 内の座標 $40, 40 \sim 60, 60$ から抽出される。第 4 の例示的な一連のダウンサイズ画像セット 2212 は、一連の画像セット 2100 内の配列決定画像 108 内の座標 $60, 60 \sim 80, 80$ から抽出される。

【0277】

いくつかの実施態様では、一連のダウンサイズの画像セットは、ニューラルネットワークベースのテンプレート生成器 1512 に入力として供給される入力画像データ 1702 を形成する。複数の一連のダウンサイズの画像セットを入力バッチとして同時に供給することができ、入力バッチ内の各シリーズに対して別個の出力を生成することができる。

(アップサンプリング)

【0278】

図 23 は、入力画像データ 1702 を形成する一連の「アップサンプリング」画像セット 2300 を生成するために、図 21b の一連の画像セット 2100 をアップサンプリングする一実施態様を示す。

【0279】

一実施態様では、アップサンプラー 2302 は、一連の画像セット 2100 内の配列決定画像 108 をアップサンプリング係数 (例えば、 $4 \times$) 及び一連のアップサンプリングされた画像セット 2300 によってアップサンプリングする。

【0280】

図示の別の実施態様では、一連の画像セット 2100 内の配列決定画像 108 は、サイズ $L \times L$ (例えば、 2000×2000) であり、アップサンプリング係数 4 によってアップサンプリングされて、一連のアップサンプリングされた画像セット 2300 内のサイズ $U \times U$ (例えば、 8000×8000) のアップサンプリングされた画像を生成する。

【0281】

一実施態様では、一連の画像セット 2100 内の配列決定画像 108 は、ニューラルネットワークベースのテンプレート生成器 1512 に直接供給され、アップサンプリングは、ニューラルネットワークベースのテンプレート生成器 1512 の初期層によって実行される。すなわち、アップサンプラー 2302 は、ニューラルネットワークベースのテンプレート生成器 1512 の一部であり、一連の画像セット 2100 内の配列決定画像 108 をアップサンプリングし、一連のアップサンプリングされた画像セット 2300 を生成する第 1 の層として動作する。

【0282】

いくつかの実施態様では、一連のアップサンプリングされた画像セット 2300 は、ニューラルネットワークベースのテンプレート生成器 1512 に入力として供給される入力画像データ 1702 を形成する。

【0283】

図 24 は、図 23 の一連のアップサンプリングされた画像セット 2300 からパッチを抽出して、入力画像データ 1702 を形成する一連の「アップサンプリング及びダウンサイズの」画像セット 2406、2408、2410 及び 2412 を生成する一実施態様を示す。

【0284】

一実施態様では、パッチ抽出器 2 2 0 2 は、一連のアップサンプリングされた画像セット 2 3 0 0 内のアップサンプリングされた画像からパッチを抽出し、一連のアップサンプリングされた画像セット 2 4 0 6、2 4 0 8、2 4 1 0 及びダウンサイズの画像セット 2 4 1 2 を生成する。一連のアップサンプリングされた画像セット及びダウンサイズの画像セット内のそれぞれのアップサンプリングされた画像は、一連のアップサンプリングされた画像セット 2 3 0 0 内の対応するアップサンプリングされた画像から抽出されるサイズ $M \times M$ (例えば、 80×80) のパッチである。パッチのサイズは予め設定することができる。他の別の実施態様では、 M は $1 \sim 1000$ の範囲の任意の数である。

【0285】

図 2 4 では、アップサンプリングされた及び小型化された画像セットの 4 つの例示的な一連が示されている。アップサンプリングされ、ダウンサイズの画像セット 2 4 0 6 の第 1 の例の一連は、一連のアップサンプリングされた画像セット 2 3 0 0 内のアップサンプリングされた画像内の座標 $0, 0 \sim 80, 80$ から抽出される。アップサンプリングされ、ダウンサイズの画像セット 2 4 0 8 の第 2 の例示的な一連は、一連のアップサンプリングされた画像セット 2 3 0 0 内のアップサンプリングされた画像内の座標 $80, 80 \sim 160, 160$ から抽出される。アップサンプリングされ、ダウンサイズされた画像セット 2 4 1 0 の第 3 の一連の例は、一連のアップサンプリングされた画像セット 2 3 0 0 内のアップサンプリングされた画像内の座標 $160, 160 \sim 240, 240$ から抽出される。アップサンプリングされ、ダウンサイズされた画像セット 2 4 1 2 の第 4 の一連の例は、一連のアップサンプリングされた画像セット 2 3 0 0 内のアップサンプリングされた画像内の座標 $240, 240 \sim 320, 320$ から抽出される。

【0286】

いくつかの実施態様では、一連のアップサンプリング及びダウンサイズの画像セットは、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 に入力として供給される入力画像データ 1 7 0 2 を形成する。複数の一連のアップサンプリングされた画像セット及びダウンサイズの画像セットは、入力バッチとして同時に供給され得、入力バッチ内の各シリーズに対して別個の出力を生成することができる。

(出力)

【0287】

3 つのモデルは、異なる出力を生成するように訓練される。これは、異なるタイプのグラウンドトゥールースデータ表現を訓練ラベルとして使用することによって達成される。回帰モデル 2 6 0 0 は、いわゆる「減衰マップ」1 7 1 6 を特徴付ける／表す出力を生成するように訓練される。バイナリ分類モデル 4 6 0 0 は、いわゆる「バイナリマップ」1 7 2 0 を特徴付ける／表す／表す出力を生成するよう訓練される。三元分類モデル 5 4 0 0 は、いわゆる「三元マップ」1 7 1 8 を特徴付ける／表す出力を生成するように訓練される。

【0288】

各タイプのモデルの出力 1 7 1 4 は、ユニット配列 1 7 1 2 を含む。ユニット 1 7 1 2 は、ピクセル、サブピクセル、又はスーパーピクセルであり得る。各タイプのモデルの出力は、ユニット配列の出力値が、回帰モデル 2 6 0 0 の場合の減衰マップ 1 7 1 6 と、バイナリ分類モデル 4 6 0 0 の場合のバイナリマップ 1 7 2 0 と、三元分類モデル 5 4 0 0 の場合の三元マップ 1 7 1 8 とを一緒に特徴付ける／表す／表すように、ユニットごとの出力値を含む。以下の詳細がある。

(グラウンドトゥールースデータ生成)

【0289】

図 2 5 は、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 を訓練するためのグラウンドトゥールースデータを生成する、全体的な例示的プロセスの一実施態様を示す。回帰モデル 2 6 0 0 に関して、グラウンドトゥールースデータは、減衰マップ 1 2 0 4 とすることができる。バイナリ分類モデル 4 6 0 0 では、グラウンドトゥールースデータは、バイナリマップ 1 4 0 4 であり得る。三元分類モデル 5 4 0 0 では、グラウンドトゥール

ースデータは三元マップ 1 3 0 4 とすることができる。グラウンドトゥールースデータは、クラスターメタデータから生成される。クラスターメタデータは、クラスターメタデータ生成器 1 2 2 によって生成される。グラウンドトゥールースデータは、グラウンドトゥールースデータ生成器 1 5 0 6 によって生成される。

【0 2 9 0】

図示の別の実施態様では、グラウンドトゥールースデータは、フローセル A のレーン A 上にあるタイル A のために生成される。グラウンドトゥールースデータは、配列決定動作 A 中に捕捉されたタイル A の配列決定画像 1 0 8 から生成される。タイル A の配列決定画像 1 0 8 は、ピクセル領域にある。配列決定サイクルごとに 4 つの配列決定画像を生成する 4 チャンネル化学を伴う一例では、5 0 個の配列決定サイクルのための 2 0 0 個の配列決定画像 1 0 8 がアクセスされる。2 0 0 個の配列決定画像 1 0 8 のそれぞれは、特定の配列決定サイクルで特定の画像チャンネル内に捕捉されたタイル A 及びそれらの周囲の背景上のクラスターの強度放射を示す。

【0 2 9 1】

サブピクセルアドレス指定器 1 1 0 は、配列決定画像 1 0 8 をサブピクセルドメインに変換し（例えば、各ピクセルを複数のサブピクセルに分割することによって）サブピクセルドメインに変換し、サブピクセルドメイン内に配列決定画像 1 1 2 を生成する。

【0 2 9 2】

次いで、ベースコーラー 1 1 4（例えば、R T A）は、サブピクセルドメイン内の配列決定画像 1 1 2 を処理し、各サブピクセル及び 5 0 個の配列決定サイクルのそれぞれについて、ベースコールを生成する。これは、本明細書では「サブピクセルベースコール」と称される。

【0 2 9 3】

次いで、サブピクセルベースコール 1 1 6 をマージして、各サブピクセルに対して、5 0 回の配列決定サイクルにわたってベースコール配列を生成する。各サブピクセルのベースコール配列は、5 0 個のベースコール、すなわち、5 0 個の配列決定サイクルのそれぞれに対する 1 つのベースコールを有する。

【0 2 9 4】

検索器 1 1 8 は、ペアワイズベースで連続するサブピクセルのベースコール配列を評価する。検索は、それぞれのサブピクセルを評価して、その連続するサブピクセルのうちのどのサブピクセルを、実質的に一致するベースコール配列を共有することを含む。ベースコールの所定の部分が、序数の位置ごとの基準（例えば、 ≥ 45 サイクルにおける 4 1 一致、 ≤ 45 サイクルにおける 4 不一致、 ≤ 50 サイクルにおける 4 不一致、又は ≤ 34 サイクルにおける 2 不一致）と一致するとき、連続するサブピクセルのベースコール配列は、「実質的に一致する」。

【0 2 9 5】

いくつかの実施態様では、ベースコーラー 1 1 4 はまた、クラスターの予備中心座標を識別する。予備中心座標を含むサブピクセルは、中心又は原点サブピクセルと呼ばれる。ベースコーラー 1 1 4 及び対応する原点サブピクセル（6 0 6 a ~ c）によって識別されたいくつかの例示的な予備中心座標（6 0 4 a ~ c）が図 6 に示されている。しかしながら、以下に説明するように、原点サブピクセル（クラスターの予備中心座標）の識別は必要ではない。いくつかの実施態様では、検索器 1 1 8 は、原点サブピクセル 6 0 6 a ~ c から始まり連続的に連続する非原点サブピクセル 7 0 2 a ~ c を継続して、サブピクセルの実質的に一致するベースコール配列を識別するための、第 1 の検索を使用する。これは、以下に説明するように、任意選択的である。

【0 2 9 6】

サブピクセルの基本的に一致するベースコール配列の検索は、全てのサブピクセルについて検索を行うことができ、その検索は、原点サブピクセルから開始する必要がなく、その代わりに、任意のサブピクセル（例えば、0, 0 サブピクセル又は任意のランダムサブピクセル）から開始することができないため、原点サブピクセル（クラスターの初期中心

10

20

30

40

50

座標)の識別を必要としない。したがって、各サブピクセルは、実質的に一致するベースコール配列を別の連続サブピクセルと共有するかどうかを判定するために評価されるため、検索は、原点サブピクセルを利用する必要はなく、任意のサブピクセルで開始することができる。

【0297】

原点サブピクセルが使用されるか否かに関わらず、ベースコーラー114によって予測される原点サブピクセル(クラスターの初期中心座標)を含まない特定のクラスターが識別される。サブピクセルベースコールのマージによって識別され、原点サブピクセルを含まないクラスターのいくつかの例は、図8aのクラスター812a、812b、812c、812d及び812eである。したがって、原点サブピクセル(クラスターの初期中心座標)を識別するためのベースコーラー114の使用は任意であり、サブピクセルの実質的に一致するベースコール配列の検索には必須ではない。

10

【0298】

検索器118:(1)いわゆる「不連続領域」として、実質的に一致するベースコール配列を有する連続するサブピクセルを識別し、(2)更に、(1)で既に識別されている非接合領域のいずれにも属しない、これらのサブピクセルのベースコール配列を更に評価し、追加の不連続領域を取得し、(3)次に、(1)及び(2)で既に識別されている不連続領域のいずれにも属しないサブピクセルとして背景サブピクセルを識別する。アクション(2)は、中心がベースコーラー114によって識別されない追加又は追加のクラスターを識別するために開示された技術を可能にする。

20

【0299】

検索器118の結果は、タイルAのいわゆる「クラスターマップ」で符号化され、クラスターマップデータストア120内に記憶される。クラスターマップでは、タイルA上のクラスターのそれぞれは、隣接するサブピクセルのそれぞれの不連続領域によって識別され、背景サブピクセルは、分離された領域を分離して、タイルA上の周囲の背景を識別する。

【0300】

質量中心(COM)計算機1004は、不連続領域を形成するそれぞれの連続するサブピクセルの座標の平均として、不連続領域のそれぞれのCOMを計算することによって、タイルA上のクラスターのそれぞれの中心を決定する。クラスターの質量中心は、COMデータ2502として記憶される。

30

【0301】

サブピクセル分類部2504は、クラスターマップ及びCOMデータ2502を使用してサブピクセル分類2506を生成する。サブピクセル分類2506は、(1)背景サブピクセル、(2)COMサブピクセル(それぞれの不連続領域のCOMを含む各不連続領域に関する1つのCOMサブピクセル)、及び(3)それぞれの不連続領域を形成するクラスター/クラスター内部サブピクセルと、を分類する。すなわち、クラスターマップ内の各サブピクセルには、3つのカテゴリのうちの1つが割り当てられる。

【0302】

一部の実施態様におけるサブピクセル分類2506に基づいて、(i)グラウンドトゥールース減衰マップ1204は、グラウンドトゥールース減衰マップ生成器1202によって生成され、(ii)グラウンドトゥールースバイナリマップ1304は、グラウンドトゥールースバイナリマップ生成器1302によって生成され、(iii)グラウンドトゥールース三元マップ1404は、グラウンドトゥールース三元マップ生成器1402によって生成される。

40

1.(回帰モデル)

【0303】

図26は、回帰モデル2600の一実施例を示す。図示の別の実施態様では、回帰モデル2600は、入力画像データ1702をエンコーダサブネットワーク及び対応するデコーダサブネットワークを介して処理する完全畳み込みネットワーク2602である。エン

50

コーダサブネットワークは、エンコーダの階層を含む。デコーダサブネットワークは、低解像度のエンコーダ機能マップを完全入力解像度減衰マップ1716にマッピングするデコーダの階層を含む。別の実施態様では、回帰モデル2600は、デコーダとエンコーダとの間のスキップ接続を有するU-Netネットワーク2604である。分割ネットワークに関する更なる詳細は、「Segmentation Networks」と題された付録に見出すことができる。

(減衰マップ)

【0304】

図27は、クラスターマップ2702からのグラウンドトゥールース減衰マップ1204を生成する一実施態様を示す。グラウンドトゥールース減衰マップ1204は、回帰モデル2600を訓練するためのグラウンドトゥールースデータとして使用される。グラウンドトゥールース減衰マップ1204では、グラウンドトゥールース減衰マップ生成器1202は、加重減衰係数に基づいて、各隣接するサブピクセルに加重減衰値を割り当てる。重み付け減衰値は、隣接するサブピクセルが属する不連続領域内の質量(COM)サブピクセルの中心からの隣接するサブピクセルのユークリッド距離に比例し、これにより、重み付き減衰値は、COMサブピクセルに対して最も高く(例えば、1又は100)、COMサブピクセルから更に離れたサブピクセルについて減少する。いくつかの実施態様では、加重減衰値は、100などの事前設定された係数で乗算される。

10

【0305】

更に、グラウンドトゥールース減衰マップ生成器1202は、全ての背景サブピクセルに同じ事前決定値(例えば、最小の背景値)を割り当てる。

20

【0306】

グラウンドトゥールース減衰マップ1204は、割り当てられた値に基づいて、不連続領域及び背景サブピクセル内の連続するサブピクセルを表している。グラウンドトゥールース減衰マップ1204はまた、割り当てられた値をユニット配列に記憶し、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。

(訓練)

【0307】

図28は、訓練2800中の訓練出力として回帰モデル2600によって生成された減衰マップ1716が地面のグラウンドトゥールース減衰マップ1204に漸進的に接近又は適合するまで回帰モデル2600のパラメータを修正する、逆伝搬ベースの勾配更新技術を使用した回帰モデル2600の訓練2800の一実施である。

30

【0308】

訓練2800は、減衰マップ1716とグラウンドトゥールース減衰マップ1204との間の誤差2806を最小化し、誤差2806に基づいて回帰モデル2600のパラメータを更新することを反復的に最適化することを含む。一実施態様では、損失関数は平均二乗誤差であり、減衰マップ1716及びグラウンドトゥールース減衰マップ1204における対応するサブピクセルの加重減衰値の間のサブピクセルごとに最小化される。

【0309】

訓練2800は、バッチングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬2808及び後方伝搬2810を含む。訓練データ1504は、入力画像データ1702として、一連のアップサンプリング及びダウンサイズの画像セットを含む。訓練データ1504は、アノテーター2806によって、グラウンドトゥールースラベルで注釈付けされる。訓練2800は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器1510によって操作可能である。

40

(推測)

【0310】

図29は、推測2900中の推測出力として回帰モデル2600によって減衰マップ1716が生成される推測2900の間の回帰モデル2600によるテンプレート生成の一実施態様である。減衰マップ1716の一例は、「Regression_Model_

50

Output」という名称の付録に開示されている。付録は、減衰マップ1716を共に表す単位加重減衰出力値2910を含む。

【0311】

推測2900は、バッチングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬2904を含む。推測2900は、入力画像データ1702として、一連のアップサンプリングされた画像セット及びダウンサイズの画像セットを含む推測データ2908に対して実行される。推測2900は、テスト2906によって動作可能である。

（流域分離）

【0312】

図30は、減衰マップ1716を（i）クラスター背景を特徴付ける背景サブピクセルを識別するように閾値化することと、（ii）クラスター中心を特徴付ける中心サブピクセルを識別するピーク検出と、を含む。閾値保持は、2値化された出力を生成するために、局所閾値バイナリを使用する閾値保持器1802によって実行される。ピーク検出は、クラスター中心を識別するためにピークロケータ1806によって実行される。ピークロケータに関する更なる詳細は、「ピーク検出」と題された付録に見出すことができる。

【0313】

図31は、背景サブピクセル及びそれぞれ閾値化器1802によって識別される中心サブピクセルを入力として取り、ピークロケータ1806が、隣接するクラスター間の強度の谷部を見つけ、クラスターを特徴付ける隣接するクラスター／クラスター内部サブピクセルの非重複グループを出力する、流域分割技術の一実施態様を示す。撥水分割技術に関する更なる詳細は、「Watershed Segmentation」と題された付録に見出すことができる。

【0314】

一実施態様では、流域分割3102は、入力（1）減衰マップ1716、（2）の無効化された出力値1802、及び（3）ピークロケータ1806によって識別されたクラスター中心を入力（1）マイナス出力値2910として取り込む。次いで、入力に基づいて、流域分割3102は出力部3104を生成する。出力3104では、各クラスター中心は、クラスター中心に属するサブピクセルの固有のセット／グループとして識別される（サブピクセルがバイナリ出力で「1」である限り、すなわち、背景サブピクセルではない）として識別される。更に、クラスターは、少なくとも4つのサブピクセルを含むことに基づいてフィルタリングされる。流域分割3102は、セグメント化器1810の一部であり得、セグメント化器1810は、次いで、ポストプロセッサ1814の一部である。

（ネットワーク構造）

【0315】

図32は、回帰モデル2600の例示的なU-Net構造を、回帰モデル2600の層の詳細、層の出力の次元性、モデルパラメータの大きさ、及び層間の相互接続の詳細を示す表である。同様の詳細は、本願に付録として提出された「Regression_Model_Example_Architecture」と題されたファイルに開示されている。

（クラスター強度抽出）

【0316】

図33は、テンプレート画像内で識別されたクラスター形状情報を使用してクラスター強度を抽出する異なるアプローチを示す。上述のように、テンプレート画像は、アップサンプリングされたサブピクセル解像度内のクラスター形状情報を識別する。しかしながら、クラスター強度情報は、典型的には光学解像度である配列決定画像108にある。

【0317】

第1の手法によれば、サブピクセルの座標は、配列決定画像108内に位置し、それらのそれぞれの強度は、双線形補間を使用して抽出され、クラスターに寄与するサブピクセルのカウントに基づいて正規化される。

10

20

30

40

50

【0318】

第2の手法は、ピクセルに寄与するサブピクセルの数に従ってピクセルの強度を変調するために、重み付けエリアカバー技術を使用する。ここでも、変調ピクセル強度は、サブピクセルカウントパラメータによって正規化される。

【0319】

第3の手法は、二次補間を使用して、配列決定画像をサブピクセルドメインにアップサンプルし、クラスターに属するアップサンプリングされたピクセルの強度を合計し、クラスターに属するアップサンプリングされたピクセルのカウントに基づいて合計された強度を正規化する。

(実験結果及び考察)

10

【0320】

図34は、回帰モデル2600の出力を使用したベースコールの異なるアプローチを示す。第1のアプローチでは、テンプレート画像内のニューラルネットワークベースのテンプレート生成器1512の出力から識別されたクラスター中心は、ベースコールのためのベースコール用のベースコーラー（例えば、本明細書では「RTAベースコール」と称されるIllumina's Time Analysisソフトウェア）に供給される。

【0321】

第2のアプローチでは、クラスターの中心の代わりに、テンプレート画像内のクラスター形状情報に基づいて、配列決定画像から抽出されたクラスター強度は、ベースコールのために、RTAベースコーラーに供給される。

20

【0322】

図35は、クラスター中心として非COM位置を使用することとは対照的に、RTAベースコールがクラスター中心として質量（COM）位置のグラウンドトゥルス中心を使用するときのベースコール性能の差を示す。結果は、COMを使用してベースコールを改善することを示す。

(モデル出力の実施例)

【0323】

図36は、回帰モデル2600によって生成された例示的な減衰マップ1716を左に示す。図36はまた、右側に、回帰モデル2600が訓練中に近似する、例示的なグラウンドトゥルス減衰マップ1204を示す。

30

【0324】

減衰マップ1716及びグラウンドトゥルス減衰マップ1204の両方は、隣接するサブピクセルの不連続領域としてクラスターを描写し、クラスターの中心は、不連続領域のうちの対応する領域の質量中心で中心サブピクセルとしてのクラスターの中心、及びそれらの周囲の背景としてのクラスターを示す。

【0325】

また、不連続領域のうちの対応する領域内の連続するサブピクセルは、隣接するサブピクセルが属する不連続領域内の中心サブピクセルからの連続サブピクセルの距離に応じて重み付けされた値を有する。一実施態様では、中心サブピクセルは、不連続領域のうちの対応する領域内で最も高い値を有する。一実施態様では、背景サブピクセルは全て、減衰マップ内で同じ最小の背景値を有する。

40

【0326】

図37は、ピーク3702を検出することによって、減衰マップ内のクラスター中心を識別するピークロケータ1806の一実施態様を示す。ピークロケータに関する更なる詳細は、「ピーク検出」と題された付録に見出すことができる。

【0327】

図38は、回帰モデル2600によって生成された減衰マップ1716内のピークロケータ1806によって検出されたピークを、対応するグラウンドトゥルス減衰マップ1204内のピークと比較する。赤色マーカーは、クラスター中心として回帰モデル2600によって予測されるピークであり、緑色マーカーは、クラスターの塊のグラウンドトゥ

50

ルース中心である。

(更なる実験結果及び考察)

【0328】

図39は、精度及び再較正統計を使用した回帰モデル2600の性能を示す。精度及び再較正統計は、回帰モデル2600が、全ての識別されたクラスター中心を回復するのに良好であることを実証する。

【0329】

図40は、20 pMのライブラリ濃度(通常動作)に対する、RTAベースコーラーを用いた回帰モデル2600の性能を比較する。RTAベースコーラーを実行することで、回帰モデル2600は、より高いクラスター密度環境(すなわち、988, 884クラスター)内の34, 323(4.46%)のクラスターを識別する。

10

【0330】

図40はまた、チェスチティフィルタ(「%PF」(パスフィルタ))を通過するクラスターの数、位置合わせされたリードの数(「配列された%」)、重複するリードの数(「%」)、「複製」、参照配列に位置合わせされた全てのリード(「%不一致」)、品質スコア30及び上記(「%Q30塩基」と呼ばれる塩基などについて、参照配列を一致させないリードの数などの他の配列決定メトリックの結果を示す。

【0331】

図41は、30 pMライブラリ濃度(高密度動作)に関する、RTAベースコーラーを用いた回帰モデル2600の性能を比較する。RTAベースコーラーを実行することで、回帰モデル2600は、34, 323(6.27%)より多くのクラスターを、遙かに高いクラスター密度環境(すなわち、1, 351, 588クラスター)で識別する。

20

【0332】

図41はまた、チェスチティフィルタ(「%PF」(パスフィルタ))を通過するクラスターの数、位置合わせされたリードの数(「配列された%」)、重複するリードの数(「%」)、「複製」、参照配列に位置合わせされた全てのリード(「%不一致」)、品質スコア30及び上記(「%Q30塩基」と呼ばれる塩基などについて、参照配列を一致させないリードの数などの他の配列決定メトリックの結果を示す。

【0333】

図42は、非重複(固有又は重複複製)の適切な読み取り対の数、すなわち、両方の読み取りが、回帰モデル2600によって検出された妥当な距離内で内側に位置合わせされた対のリードの数と、RTAベースコーラーによって検出されたものと比較したものである。比較は、20 pMの通常動作及び30 pMの高密度動作の両方で行われる。

30

【0334】

より重要なことに、図42は、開示されたニューラルネットワークベースのテンプレート生成器が、テンプレート生成に対する入力より少ない配列決定サイクルにおいて、より多くのクラスターを検出することができることを示す。ただ4回の配列決定サイクルでは、回帰モデル2600は、20 pMの通常動作中のRTAベースコーラーよりも、11%多く重複していない適切な読み取り対を識別し、30 pMの高密度動作中のRTAベースコーラーよりも33%多く重複していない適切な読み取り対を識別する。7回の配列決定サイクルでは、回帰モデル2600は、20 pMの通常動作中のRTAベースコーラーよりも、4.5%多く重複していない適切な読み取り対を識別し、30 pMの高密度動作中のRTAベースコーラーよりも6.3%多く重複していない適切な読み取り対を識別する。

40

【0335】

図43は、回帰モデル2600によって生成された第1の減衰マップを右側に示す。第1の減衰マップは、クラスターの形状、クラスターサイズ、及びクラスター中心を示すこれらの空間分布と共に、20 pMの通常動作中に撮像されたクラスター及びそれらの周囲の背景を識別する。

【0336】

50

左側では、図43は、回帰モデル2600によって生成された第2の減衰マップを示す。第2減衰マップは、30 pM高密度動作中に撮像されたクラスター及びそれらの周囲の背景を、クラスター形状、クラスターサイズ、及びクラスター中心を示すそれらの空間分布と共に識別する。

【0337】

図44は、40 pMのライブラリ濃度（高密度動作）について、RTAベースコーラーを用いた回帰モデル2600の性能を比較する。回帰モデル2600は、遥かに高いクラスター密度環境（すなわち、1,509,395クラスター）において、RTAベースコーラーよりも89,441,688のより整列した塩基を生成した。

【0338】

図44はまた、チェスチティフィルタ（「%PF」（パスフィルタ））を通過するクラスターの数、位置合わせされたリードの数（「配列された%」）、重複するリードの数（「%」）、「複製」、参照配列に位置合わせされた全てのリード（「%不一致」）、品質スコア30及び上記（「%Q30塩基」）と呼ばれる塩基などについて参照配列を不一致させるリードの数などの他の配列決定メトリックの結果を示す。

（モデル出力の更なる実施例）

【0339】

図45は、回帰モデル2600によって生成された第1の減衰マップを左に示す。第1の減衰マップは、40 pMの通常動作中に画像化されたクラスター及びそれらの周囲の背景を、クラスター形状、クラスターサイズ、及びクラスター中心を示すそれらの空間分布と共に識別する。

【0340】

右上では、図45は、閾値及び第1の減衰マップに適用されたピーク位置の結果を示して、それぞれのクラスターを互いから及び背景から区別し、それらのそれぞれのクラスター中心を識別する結果を示す。いくつかの実施態様では、それぞれのクラスターの強度が識別され、不整合率を低減するために適用されるシャースフィルタ（又は通過フィルタ）が特定される。

2.（バイナリ分類モデル）

【0341】

図46は、バイナリ分類モデル4600の一実施例を示す。図示の別の実施態様では、バイナリ分類モデル4600は、入力画像データ1702をエンコーダサブネットワーク及び対応するデコーダサブネットワークを介して処理する、ディープフル畳み込みセグメンテーションニューラルネットワークである。エンコーダサブネットワークは、エンコーダの階層を含む。デコーダサブネットワークは、低解像度のエンコーダ特徴部マップを完全入力解像度バイナリマップ1720にマッピングするデコーダの階層を含む。別の実施態様では、バイナリ分類モデル4600は、デコーダとエンコーダとの間のスキップ接続を有するU-Netネットワークである。分割ネットワークに関する更なる詳細は、「Segmentation Networks」と題された付録に見出すことができる。

（バイナリマップ）

【0342】

バイナリ分類モデル4600の最終出力層は、出力アレイ内の単位ごとに分類ラベルを生成する単位ごとの分類層である。いくつかの実施態様では、単位ごと区分層は、2つのクラスにわたってバイナリマップ1720内の各サブピクセルについて、ソフトマックス分類スコア分布を生成するサブピクセルごとの分類層であり、すなわち、クラスター中心クラス及び非クラスタークラス、及び所与のサブピクセルの分類ラベルは、対応するソフトマックス分類スコア分布から決定される。

【0343】

他の別の実施態様では、単位ごとの分類層は、単位の活性化が、単位が第1のクラスに属する確率として解釈されるように、バイナリマップ1720内の各サブピクセルについてのシグモイド分類スコアを生成するサブピクセルごとの分類層であり、逆に、1つから

10

20

30

40

50

の1つのマイナスは、第2のクラスに属する確率を与える。

【0344】

バイナリマップ1720は、予測される分類スコアに基づいて、各サブピクセルを表している。バイナリマップ1720はまた、ユニット配列内に予測値分類スコアを記憶し、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。

(訓練)

【0345】

図47は、バイナリ分類モデル4600のバイナリマップ1720が、グラウンドトゥールースバイナリマップ1404に漸進的に接近又は一致するまで、バイナリ分類モデル4600のパラメータを修正する逆伝搬ベースの勾配更新技術を使用したバイナリ分類モデル4600の訓練4700の一実施態様である。

10

【0346】

図示した実施態様では、バイナリ分類モデル4600の最終出力層は、ソフトマックスベースのサブピクセルごとの分類層である。ソフトマックス別の実施態様では、グラウンドトゥールースバイナリマップ生成器1402は、(i) クラスター中心値対(例えば、[1, 0])又は(ii) 非中心値対(例えば、[0, 1])のいずれかの各グラウンドトゥールースサブピクセルを割り当てる。

【0347】

クラスター中心値対[1, 0]において、第1の値[1]はクラスター中心クラスラベルを表し、第2の値[0]は、非中心クラスラベルを表す。非中心値対[0, 1]において、第1の値[0]はクラスター中心クラスラベルを表し、第2の値[1]は、非中心クラスラベルを表す。

20

【0348】

グラウンドトゥールースバイナリマップ1404は、割り当てられた値対／値に基づいて、各サブピクセルを表している。グラウンドトゥールースバイナリマップ1404はまた、割り当てられた値対／値をユニット配列に記憶し、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。

【0349】

訓練は、バイナリマップ1720とグラウンドトゥールースバイナリマップ1404との間の誤差4706(例えば、ソフトマックス誤差)を最小化する損失関数を反復的に最適化することと、誤差4706に基づいてバイナリ分類モデル4600のパラメータを更新することと、を含む。

30

【0350】

一実施態様では、損失関数は、カスタム重み付け二値クロスエントロピー損失であり、エラー4706は、図47に示されるように、予測される分類スコア(例えば、ソフトマックススコア)と標識されたクラススコア(例えば、ソフトマックススコア)との間のサブピクセルごとに最小化され、バイナリマップ1720及びグラウンドトゥールースバイナリマップ1404内の対応するサブピクセルの標識されたクラススコア(例えば、ソフトマックススコア)との間で最小化される。

【0351】

カスタム加重損失関数は、COMサブピクセルが誤分類されるたびに、褒賞(又はペナルティ)マトリックスで指定された対応する褒賞(又はペナルティ)重みを乗じて、COMサブピクセルに、より多くの重みを与える。カスタム加重損失関数に関する更なる詳細は、「Custom-Weighted Loss Function」と題された付録に見出すことができる。

40

【0352】

訓練4700は、バッティングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬4708及び後方伝搬4710を含む。訓練データ1504は、入力画像データ1702として、一連のアップサンプリング及びダウンサイズの画像セットを含む。訓練データ1504は、アノテーター2806によって、グラウンドトゥールースラベ

50

ルで注釈付けされる。訓練 2800 は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器 1510 によって操作可能である。

【0353】

図 48 は、バイナリ分類モデル 4600 の最終出力層がシグモイドベースのサブピクセルごとの分類層である、バイナリ分類モデル 4600 の訓練 4800 の別の実施態様である。

【0354】

シグモイド別の実施態様では、グラウンドトゥールースバイナリマップ生成器 1302 は、(i) クラスタ中心値（例えば、[1]）又は (ii) 非中心値（例えば、[0]）のいずれかの各グラウンドトゥールースサブピクセルを割り当てる。COM サブピクセルには、クラスタ中心値ペア／値が割り当てられ、他の全てのサブピクセルには、非中心値対／値が割り当てられる。

【0355】

クラスタ中心値では、0～1 の間の閾値中間値（例えば、0.5 を超える値）の値は、中心クラスラベルを表す。非中心値では、0～1 の閾値中間値（例えば、0.5 未満の値）を下回る値は、非中心クラスラベルを表す。

【0356】

グラウンドトゥールースバイナリマップ 1404 は、割り当てられた値対／値に基づいて、各サブピクセルを表している。グラウンドトゥールースバイナリマップ 1404 はまた、割り当てられた値対／値をユニット配列に記憶し、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。

【0357】

訓練は、バイナリマップ 1720 とグラウンドトゥールースバイナリマップ 1404 との間の誤差 4806（例えば、シグモイド誤差）を最小化する損失関数を反復的に最適化することと、誤差 4806 に基づいてバイナリ分類モデル 4600 のパラメータを更新することと、を含む。

【0358】

一実施態様では、損失関数は、カスタム重み付け二値クロスエントロピー損失であり、誤差 4806 は、図 48 に示されるように、バイナリマップ 1720 及びグラウンドトゥールースバイナリマップ 1404 内の対応するサブピクセルの予測スコア（例えば、シグモイドスコア）との間のサブピクセルごとに最小化され、図 48 に示されるように、二値マップ 1720 及びグラウンドトゥールースバイナリマップ 1404 における対応するサブピクセルの標識されたスコア（例えば、シグモイドスコア）で最小化される。

【0359】

カスタム加重損失関数は、COM サブピクセルが誤分類されるたびに、褒賞（又はペナルティ）マトリックスで指定された対応する褒賞（又はペナルティ）重みを乗じて、COM サブピクセルに、より多くの重みを与える。カスタム加重損失関数に関する更なる詳細は、「Custom-Weighted Loss Function」と題された付録に見出すことができる。

【0360】

訓練 4800 は、バッチングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬 4808 及び後方伝搬 4810 を含む。訓練データ 1504 は、入力画像データ 1702 として、一連のアップサンプリング及びダウンサイズの画像セットを含む。訓練データ 1504 は、アノテーター 2806 によって、グラウンドトゥールースラベルで注釈付けされる。訓練 2800 は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器 1510 によって操作可能である。

【0361】

図 49 は、バイナリ分類モデル 4600 に供給された入力画像データ 1702、及びバイナリ分類モデル 4600 を訓練するために使用される対応するクラスラベル 4904 の別の実施を示す。

【0362】

図示の別の実施態様では、入力画像データ1702は、一連でアップサンプリングされ、ダウンサイズの画像セット4902を含む。クラスラベル4904は、2つのクラスを含む。(1)「クラスター中心なし」及び(2)「クラスター中心」は、異なる出力値を使用して区別される。すなわち、(1)光緑色単位／サブピクセル4906は、クラスター中心を含まないバイナリ分類モデル4600によって予測されるサブピクセルを表し、(2)暗緑色サブピクセル4908は、クラスター中心を含むバイナリ分類モデル4600によって予測される単位／サブピクセルを表す。

(推測)

【0363】

図50は、推測5000中の推測出力としてバイナリマップ1720がバイナリ分類モデル4600によって生成される推測5000の間のバイナリ分類モデル4600によるテンプレート生成の一実施態様である。バイナリマップ1720の一例は、バイナリマップ1720と一緒に表す単位ごとのバイナリ分類スコア5010を含む。ソフトマックスアプリケーションでは、バイナリマップ1720は、非中心クラスの単位ごとの分類スコアの第1のアレイ5002aと、クラスター中心クラスの単位ごとの分類スコアの第2のアレイ5002bとを有する。

【0364】

推測5000は、バッティングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬5004を含む。推測5000は、入力画像データ1702として、一連のアップサンプリングされた画像セット及びダウンサイズの画像セットを含む推測データ2908に対して実行される。推測5000は、テスト2906によって動作可能である。

【0365】

いくつかの実施態様では、バイナリマップ1720は、クラスターメタデータを生成するために、閾値保持、ピーク検出、及び／又はウォータージェット分割などの、上述の後処理技術に供される。

(ピーク検出)

【0366】

図51は、クラスター中心を識別するために、バイナリマップ1720をピーク検出に供する一実施態様を示す。上述のように、バイナリマップ1720は、予測された分類スコアに基づいて各サブピクセルを分類するユニット配列であり、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。分類スコアは、ソフトマックススコア又はシグモイドスコアであり得る。

【0367】

ソフトマックス用途では、バイナリマップ1720は、2つのアレイを含む。(1)非中心クラスの単位ごとの分類スコアの第1のアレイ5002aと、(2)クラスター中心クラスの単位ごとの分類スコアの第2のアレイ5002bと、を含む。アレイの両方において、各ユニットは、入力内の対応するサブピクセルを表す。

【0368】

入力内のどのサブピクセルがクラスター中心を含み、クラスター中心を含まないかを判定するために、ピークロータ1806は、バイナリマップ1720内のユニット上にピーク検出を適用する。ピーク検出は、事前設定された閾値を上回る分類スコア(例えば、ソフトマックス／シグモイドスコア)を有する単位を識別する。識別されたユニットは、クラスター中心として推測され、入力内のそれらの対応するサブピクセルは、クラスター中心を含み、サブピクセル分類データストア5102内のクラスター中心サブピクセルとして記憶されると判定される。ピークロータ1806に関する更なる詳細は、「ピーク検出」と題された付録に見出すことができる。

【0369】

入力内の残りのユニット及びそれらの対応するサブピクセルは、クラスター中心を含ま

10

20

30

40

50

ず、サブピクセル分類データストア 5 1 0 2 内の非中心サブピクセルとして記憶される。

【0370】

いくつかの実施態様では、ピーク検出を適用する前に、特定の背景閾値（例えば、0.3）を下回る分類スコアを有するユニットは、ゼロに設定される。いくつかの実施態様では、入力内のそのようなユニット及びそれらの対応するサブピクセルは、クラスターを取り囲む背景を示すように推測され、サブピクセル分類データストア 5 1 0 2 内の背景サブピクセルとして記憶される。他の実施態様では、このようなユニットは、ノイズと見なされ、無視することができる。

（モデル出力の実施例）

【0371】

図 5 2 a は、バイナリ分類モデル 4 6 0 0 によって生成された例示的なバイナリマップを左側に示す。図 5 2 a はまた、右側に、バイナリ分類モデル 4 6 0 0 が訓練中に近似する、例示的なグラウンドトゥルスバイナリマップを示す。バイナリマップは、複数のサブピクセルを有し、クラスター中心又は非中心のいずれかとして各サブピクセルを分類する。同様に、グラウンドトゥルスバイナリマップは、複数のサブピクセルを有し、クラスター中心又は非中心のいずれかとして各サブピクセルを分類する。

（実験結果及び考察）

【0372】

図 5 2 b は、再較正及び精密統計を使用したバイナリ分類モデル 4 6 0 0 の性能を示す。これらの統計値を適用することにより、バイナリ分類モデル 4 6 0 0 は、RTA ベース

（ネットワーク構造）

【0373】

図 5 3 は、バイナリ分類モデル 4 6 0 0 の層の詳細、層の出力の次元性、モデルパラメータの大きさ、及び層間の相互接続の詳細と共に、バイナリ分類モデル 4 6 0 0 の例示的な構造を示す表である。同様の詳細は、「Binary Classification Model Example Architecture」という名称の付録に開示されている。

3. 三元（3クラス）分類モデル

【0374】

図 5 4 は、三元分類モデル 5 4 0 0 の一実施態様を示す。図示の別の実施態様では、三元分類モデル 5 4 0 0 は、入力画像データ 1 7 0 2 をエンコーダサブネットワーク及び対応するデコーダサブネットワークを介して処理する、深層完全畳み込みセグメンテーションニューラルネットワークである。エンコーダサブネットワークは、エンコーダの階層を含む。デコーダサブネットワークは、低解像度のエンコーダ特徴部マップを完全入力解像度の三元マップ 1 7 1 8 にマッピングするデコーダの階層を含む。別の実施態様では、三元分類モデル 5 4 0 0 は、デコーダとエンコーダとの間のスキップ接続を有する U-Net ネットワークである。分割ネットワークに関する更なる詳細は、「Segmentation Networks」と題された付録に見出すことができる。

（三元マップ）

【0375】

三元分類モデル 5 4 0 0 の最終出力層は、出力アレイ内の単位ごとに分類ラベルを生成する単位ごとの分類層である。いくつかの実施態様では、単位ごと区分層は、3つのクラスにわたって三元マップ 1 7 1 8 内の各サブピクセルについて、ソフトマックス分類スコア分布を生成するサブピクセルごとの分類層であり、すなわち、背景クラス、クラスター中心クラス、及びクラスター／クラスター内部クラス、及び所与のサブピクセルの分類ラベルは、対応するソフトマックス分類スコア分布から決定される。

【0376】

三元マップ 1 7 1 8 は、予測される分類スコアに基づいて、各サブピクセルを表している。三元マップ 1 7 1 8 はまた、ユニット配列内に予測値分類スコアを記憶し、アレイ内

10

20

30

40

50

の各ユニットは、入力内の対応するサブピクセルを表す。

(訓練)

【0377】

図55は、三元分類モデル5400の三元マップ1718が、訓練用グラウンドトゥール三元マップ1304に漸進的に接近又は適合するまで、三元分類モデル5400のパラメータを修正する、逆伝搬ベースの勾配更新技術を使用して、三元分類モデル5400を訓練5500する一実施態様である。

【0378】

図示した実施態様では、三元分類モデル5400の最終出力層は、ソフトマックススペースのサブピクセルごとの分類層である。ソフトマックス別の実施態様では、グラウンドトゥールの三元マップ生成器1402は、(i) 背景値トリプレット (例えば、[1、0、0])、(i i) クラスター中心値トリプレット (例えば、[0、1、0])、又は(i i i) クラスター／クラスター内部値トリプレット (例えば、[0、0、1]) のいずれかを割り当てる。

10

【0379】

背景サブピクセルに背景値トリプレットが割り当てられる。質量 (COM) サブピクセルの中心には、クラスター中心値トリプレットが割り当てられる。クラスター／クラスター内部サブピクセルには、クラスター／クラスター内部値トリプレットが割り当てられる。

【0380】

背景値トリプレット [1、0、0] において、第1の値 [1] は背景クラスラベルを表し、第2の値 [0] はクラスター中心ラベルを表し、第3の値 [0] はクラスター／クラスター内部クラスラベルを表す。

20

【0381】

クラスター中心値トリプレット [0、1、0] において、第1の値 [0] は、背景クラスラベルを表し、第2の値 [1] はクラスター中心ラベルを表し、第3の値 [0] はクラスター／クラスター内部クラスラベルを表す。

【0382】

クラスター／クラスター内部値トリプレット [0、0、1] において、第1の値 [0] は、背景クラスラベルを表し、第2の値 [0] はクラスター中心ラベルを表し、第3の値 [1] はクラスター／クラスター内部クラスラベルを表す。

30

【0383】

グラウンドトゥール三元マップ1304は、割り当てられた値トリプレットに基づいて、各サブピクセルを表している。グラウンドトゥール三元マップ1304はまた、割り当てられたトリプレットをユニット配列に記憶し、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。

【0384】

訓練は、三元マップ1718とグラウンドトゥール三元マップ1304との間の誤差5506 (例えば、ソフトマックス誤差) を最小化する損失関数を反復的に最適化することと、誤差5506に基づいて三元分類モデル5400のパラメータを更新することと、を含む。

40

【0385】

一実施態様では、損失関数は、カスタム重み付けカテゴリ化クロスエントロピー損失であり、エラー5506は、図54に示されるように、予測分類スコア (例えば、ソフトマックススコア) と標識されたクラススコア (例えば、ソフトマックススコア) との間のサブピクセルごとに最小化され、三元マップ1718及びグラウンドトゥール三元マップ1304内の対応するサブピクセルの標識されたクラススコア (例えば、ソフトマックススコア) との間で最小化される。

【0386】

カスタム加重損失関数は、COMサブピクセルが誤分類されるたびに、褒賞 (又はペナルティ) マトリックスで指定された対応する褒賞 (又はペナルティ) 重みを乗じて、CO

50

Mサブピクセルに、より多くの重みを与える。カスタム加重損失関数に関する更なる詳細は、「Custom-Weighted Loss Function」と題された付録に見出すことができる。

【0387】

訓練5500は、バッチングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬5508及び後方伝搬5510を含む。訓練データ1504は、入力画像データ1702として、一連のアップサンプリング及びダウンサイズの画像セットを含む。訓練データ1504は、アノテーター2806によって、グラウンドトゥールスラベルで注釈付けされる。訓練5500は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器1510によって操作可能である。

10

【0388】

図56は、三元分類モデル5400に供給された入力画像データ1702、及び三元分類モデル5400を訓練するために使用される対応するクラスラベルの一実施態様を示す。

【0389】

図示の別の実施態様では、入力画像データ1702は、一連でアップサンプリングされ、ダウンサイズの画像セット5602を含む。クラスラベル5604は、3つのクラスを含む。(1)「背景クラス」、(2)「クラスター中心クラス」、及び(3)「クラスター内部クラス」は、異なる出力値を使用して区別される。例えば、これらの異なる出力値の一部は、以下のように視覚的に表すことができる。(1)グレイユニット／サブピクセル5606は、背景であると三元分類モデル5400によって予測されるサブピクセルを表し、(2)暗緑色単位／サブピクセル5608は、クラスター中心を含有するために三元分類モデル5400によって予測されるサブピクセルを表し、及び(3)光緑色サブピクセル5610は、クラスターの内部を収容するために、三元分類モデル5400によって予測されるサブピクセルを表す。

20

(ネットワーク構造)

【0390】

図57は、三元分類モデル5400の層の詳細、層の出力の次元性、モデルパラメータの大きさ、及び層間の相互接続の詳細と共に、三元分類モデル5400の例示的な構造を示す表である。同様の詳細は、「Ternary_Classification_Model_Example_Architecture」という名称の付録に開示されている。

30

(推測)

【0391】

図58は、推測5800中の推測出力として三元マップ1718が三元分類モデル5400によって生成される推測5800中における三元分類モデル5400によるテンプレート生成の一実施態様である。三元マップ1718の一例は、「Ternary_Classification_Model_Output」と題された付録に開示されている。付録は、共に三元マップ1718を表す単位ごとのバイナリ分類スコア5810を含む。ソフトマックスアプリケーションでは、付録は、背景クラスの単位ごとの分類スコアの第1のアレイ5802aと、クラスター中心クラスの単位ごと分類スコアの第2のアレイ5802bと、クラスター／クラスター内部クラスの単位ごとの分類スコアの第3のアレイ5802cとを有する。

40

【0392】

推測5800は、バッチングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬5804を含む。推測5800は、入力画像データ1702として、一連のアップサンプリングされた画像セット及びダウンサイズの画像セットを含む推測データ2908に対して実行される。推測5000は、テスト2906によって動作可能である。

【0393】

いくつかの実施態様では、三元マップ1718は、閾値化、ピーク検出、及び／又はウ

50

オータージェット分割などの、上述の後処理技術を使用して、三元分類モデル 5 4 0 0 によって生成される。

【0 3 9 4】

図 5 9 は、3 つの対応するクラス、すなわち、背景クラス 5 9 0 6 のクラスター中心クラス 5 9 0 2 及びクラスター／クラスター内部クラス 5 9 0 4 の三元ソフトマックス分類スコア分布をそれぞれ有する三元分類モデル 5 4 0 0 によって生成された三元マップ 1 7 1 8 をグラフで示す。

【0 3 9 5】

図 6 0 は、単位ごとの出力値と共に三元分類モデル 5 4 0 0 によって生成されたユニット配列を示す。図示のように、各ユニットは、3 つの対応するクラス、すなわち、背景クラス 5 9 0 6 のクラスター中心クラス 5 9 0 2 及びクラスター／クラスター内部クラス 5 9 0 4 の 3 つの出力値を有する。各分類（列方向）について、各単位は、各単位の括弧内のクラスによって示されるように、最も高い出力値を有するクラスを割り当てられる。いくつかの実施態様では、出力値 6 0 0 2、6 0 0 4、及び 6 0 0 6 は、それぞれのクラス 5 9 0 6、5 9 0 2、及び 5 9 0 4（行ごと）のそれぞれについて分析される。

（ピーク検出及び流域分割）

【0 3 9 6】

図 6 1 は、クラスター中心、クラスター背景、及びクラスター内部を識別するために、三元マップ 1 7 1 8 を後処理に供する一実施態様を示す。上述のように、三元マップ 1 7 1 8 は、予測された分類スコアに基づいて各サブピクセルを分類するユニット配列であり、アレイ内の各ユニットは、入力内の対応するサブピクセルを表す。分類スコアは、ソフトマックススコアであり得る。

【0 3 9 7】

ソフトマックス用途では、三元マップ 1 7 1 8 は、3 つのアレイを含む。（1）背景クラスの単位ごとの分類スコアの第 1 のアレイ 5 8 0 2 a、（2）クラスター中心クラスの単位ごと分類スコアの第 2 のアレイ 5 8 0 2 b、及び（3）クラスター内部クラスに関する単位ごと分類スコアの第 3 のアレイ 5 8 0 2 c と、を含む。全ての 3 つのアレイにおいて、各ユニットは、入力内の対応するサブピクセルを表す。

【0 3 9 8】

入力内のどのサブピクセルがクラスターの内部を含有し、かつ背景を含むクラスター中心を含むかを判定するために、ピークロケータ 1 8 0 6 は、クラスター中心クラス 5 8 0 2 b の三元マップ 1 7 1 8 内のソフトマックス値にピーク検出を適用する。ピーク検出は、事前設定された閾値を上回る分類スコア（例えば、ソフトマックススコア）を有するユニットを識別する。識別されたユニットは、クラスター中心として推測され、入力内のそれらの対応するサブピクセルは、クラスター中心を含み、サブピクセル分類及びセグメント化データストア 6 1 0 2 内のクラスター中心サブピクセルとして記憶されると判定される。ピークロケータ 1 8 0 6 に関する更なる詳細は、「ピーク検出」と題された付録に見出すことができる。

【0 3 9 9】

いくつかの実施態様では、ピーク検出を適用する前に、特定のノイズ閾値（例えば、0.3）を下回る分類スコアを有するユニットは、ゼロに設定される。このようなユニットは、ノイズと見なすことができ、無視することができる。

【0 4 0 0】

また、特定の背景閾値（例えば、0.5 以上）を上回る背景クラス 5 8 0 2 a の分類スコアを有し、入力内のそれらの対応するサブピクセルは、クラスターを取り囲む背景を示すように推測され、サブピクセル分類及びセグメント化データストア 6 1 0 2 内の背景サブピクセルとして記憶される背景サブピクセルとして記憶される。

【0 4 0 1】

次いで、流域セグメント 3 1 0 2 によって操作される流域分割アルゴリズムが、クラスターの形状を決定するために使用される。いくつかの実施態様では、背景ユニット／サブ

10

20

30

40

50

ピクセルは、流域分割アルゴリズムによってマスクとして使用される。クラスター中心及びクラスター内部として推測される単位／サブピクセルの分類スコアは、いわゆる「クラスターラベル」を生成するために合計される。クラスター中心は、流域分割アルゴリズムによる強度谷部による分離のために、流域マーカーとして使用される。

【0402】

一実施態様では、負極化されたクラスターラベルは、セグメンテーションを実行し、背景サブピクセルによって分離された隣接するクラスター内部サブピクセルの不連続領域としてクラスター形状を生成する、流域セグメンター3102への入力画像として提供される。更に、各不連続領域は、対応するクラスター中心サブピクセルを含む。いくつかの実施態様では、対応するクラスター中心サブピクセルは、それが属する領域の中心である。他の実施態様では、不連続領域の質量（COM）の中心は、下にある位置座標に基づいて計算され、クラスターの新たな中心として記憶される。

10

【0403】

流域分割3102の出力は、サブピクセル分類及びセグメント化データストア6102に記憶される。流域分割アルゴリズム及び他の分割アルゴリズムに関する更なる詳細は、「Watershed Segmentation」と題された付録に見出すことができる。

【0404】

ピークローケータ1806及び流域分割3102の出力例が図62a、62b、63、及び64に示されている。

20

（モデル出力の実施例）

【0405】

図62aは、三元分類モデル5400の例示的予測を示す。図62aは、4つのマップを示し、各マップはユニット配列を有する。第1のマップ6202（左端の）は、クラスター中心クラス5802bの各ユニットの出力値を示す。第2のマップ6204は、クラスター／クラスター内部クラス5802cの各ユニットの出力値を示す。第3のマップ6206（右端）は、背景クラス5802aの各ユニットの出力値を示す。第4のマップ6208（底部）は、最も高い出力値を有するクラスラベルを各ユニットに割り当てる、グラウンドトゥールズ三元マップ6008のバイナリマスクである。

【0406】

30

図62bは、三元分類モデル5400の他の例示的予測を示す。図62bは、4つのマップを示し、各マップはユニット配列を有する。第1のマップ6212（最下部）は、クラスター／クラスター内部クラスの各ユニットの出力値を示す。第2のマップ6214は、クラスター中心クラスのそれぞれの単位の出力値を示す。第3のマップ6216（最も右の）は、背景クラスのそれぞれの単位の出力値を示す。第4のマップ（上部）6210は、最も高い出力値を有するクラスラベルを各ユニットに割り当てる、グラウンドトゥールズ三元マップである。

【0407】

図62cは、三元分類モデル5400の更に他の例示的予測を示す。図64は、4つのマップを示し、各マップはユニット配列を有する。第1のマップ6220（最下部）は、クラスター／クラスター内部クラスの各ユニットの出力値を示す。第2のマップ6222は、クラスター中心クラスのそれぞれの単位の出力値を示す。第3のマップ6224（最も右の）は、背景クラスのそれぞれの単位の出力値を示す。第4のマップ6218（上部）は、最も高い出力値を有するクラスラベルを各ユニットに割り当てる、グラウンドトゥールズ三元マップである。

40

【0408】

図63は、出力を後処理に供することによって、図62aの三元分類モデル5400の出力からクラスター中心及びクラスター形状を導出する一実施態様を示す。後処理（例えば、ピーク位置、ウォッシュセグメンテーション）は、クラスター形状データ及びクラスターマップ6310内で識別される他のメタデータを生成する。

50

(実験結果及び考察)

【0409】

図64は、バイナリ分類モデル4600、回帰モデル2600及びRTAベースコーラーの性能を比較する。性能は、様々な配列決定メトリックを使用して評価される。1つの指標は、検出されるクラスターの総数（「#クラスター」）であり、これは検出される固有のクラスター中心の数によって測定することができる。別のメトリックは、チェーティフィルタ（「%PF」（パスフィルタ））を通過する検出されたクラスターの数である。配列決定動作のサイクル1-25の間、チェーティフィルタは、画像抽出結果から少なくとも信頼性の高いクラスターを除去する。クラスターは、1つ以下のベースコールが、第1の25サイクルにおいて0.6未満のチェンティ値を有する場合には、「フィルタを通過する」。買い物客は、最も明るい塩基強度の比を、最も明るい試験と第2の最も明るい塩基強度との合計で割ったものとして定義される。この指標は、検出されたクラスターの量を超えており、また、その品質、すなわち、検出されたクラスターのうちのどれだけが、変異型コーリング及び変異型病原性アノテーションなどの正確なベースコール及び下流の二次及び三元分析のために使用され得る。

【0410】

検出されたクラスターが下流分析のためにどれくらい良好かを測定する他のメトリックとしては、検出されたクラスターから生成された整列されたリードの数（「配列された%」）、検出されたクラスターから生成された複製リードの数（「%Duplicate」）、検出されたクラスターから生成されたリードの数は、参照配列に位置合わせされた全てのリードについて参照配列を不一致させる（「不一致」）、検出されたクラスターから生成されたリードの数は、その部分がいずれかの側の基準配列に十分に一致しないため、位置合わせ（「軟クリップの%」）について無視され、検出されたクラスターについて呼ばれる塩基の数は、品質スコア30を有し、上にある（「%」）。Q30塩基）、検出されたクラスターから生成された対のリードの数は、妥当な距離（「全適切な読み取り対」）内で内側に位置合わせされたリードと、検出されたクラスターから生成されたユニークな又は重複した適切な読み取りペアの数（「非重複の正しい読み取り対」）。

【0411】

図64に示されるように、バイナリ分類モデル4600及び回帰モデル2600の両方は、測定基準の大部分でのテンプレート生成において、RTAベースコーラーを実行する。

【0412】

図65は、3つの状況、5つの配列決定メトリック、及び2つの動作密度下での、三元分類モデル5400の性能をRTAベースコーラーの性能と比較する。

【0413】

「RTA」と呼ばれる第1の状況では、クラスター中心は、RTAベースコーラーによって検出され、クラスターからの強度抽出は、RTAベースコーラーによって行われ、クラスターはまた、RTAベースコーラーを使用してベースコールされる。「RTA IE」と呼ばれる第2の状況では、クラスター中心は、三元分類モデル5400によって検出されるが、クラスターからの強度抽出は、RTAベースコーラーによって行われ、クラスターもまた、RTAベースコーラーを使用してベースコールされる。「Self IE」と呼ばれる第3の状況において、クラスター中心は、三元分類モデル5400によって検出され、クラスターからの強度抽出は、本明細書に開示されるクラスター形状ベースの強度抽出技術を使用して行われる（クラスター形状情報が三元分類モデル5400によって生成されることに留意されたい）。ただし、クラスターは、RTAベースコーラーを使用してベースコールされる。

【0414】

性能は、三元分類モデル5400と、以下の5つの測定基準に沿ったRTAベースコーラーとの間で比較される。（1）検出されたクラスターの総数（「#クラスター」）、（2）チェスチティフィルタ（「#PF」）を通過する検出されたクラスターの数、（3）検出されたクラスターから生成された固有の又は重複した適切なリード対の数（「#不重

複適切な読み出しペア」)、(4) 検出されたクラスターから生成された配列リードと、アライメント後の参照配列(「不一致率」)、及び(5) 品質スコア30を有する検出されたクラスターと(「%Q30」)との間の不一致の割合。

【0415】

3つの状況下の三元分類モデル5400とRTAベースコーラーとの間で性能を比較し、2種類の配列決定動作について5つのメトリックを比較する。(1) 20pMライブラリ濃度を有する通常動作、及び(2) 30pMライブラリ濃度を有する高密度動作。

【0416】

図65に示されるように、三元分類モデル5400は、全ての測定基準に対して、RTAベースコーラーを実行する。

【0417】

図66は、同じ3つの状況下で、5つの測定基準、及び2つの動作密度の下で、回帰モデル2600が全ての測定基準に対してRTAベースコーラーを実行することを示す。

【0418】

図67は、ニューラルネットワークベースのテンプレート生成器1512の最終層6702に焦点を当てる。

【0419】

図68は、ニューラルネットワークベースのテンプレート生成器1512の最終層6702が、逆伝搬ベースの勾配更新訓練の結果として学習したものを可視化する。図示された実施態様は、グラウンドトゥルースクラスター形状に重ねられた最終層6702の32つの畳み込みフィルタから24を可視化する。図68に示されるように、最終層6702は、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及びクラスター境界などのクラスターの空間分布を含むクラスターメタデータを学習している。

【0420】

図69は、(青色での)バイナリ分類モデル4600のクラスター中心予測を、(ピンク色で)RTAベースコーラーのものに重ね合わせる。予測は、Illumina NextSeqシーケンサからの画像データを配列決定するために行われる。

【0421】

図70は、バイナリ分類モデル4600の最終層の訓練された畳み込みフィルタの可視化上に、(ピンク色で)RTAベースコーラー(ピンク色で)によって作成されたクラスター中心予測を重ね合わせる。これらの畳み込みフィルタは、Illumina NextSeqシーケンサからの画像データの配列決定の結果として学習される。

【0422】

図71は、ニューラルネットワークベースのテンプレート生成器1512を訓練するために使用される訓練データの一実施態様を示す。この別の実施態様では、訓練データは、ストームプローブ画像を用いてデータを生成する高密度フローセルから取得される。別の実施態様では、訓練データは、より少ないブリッジ増幅サイクルでデータを生成する高密度フローセルから得られる。

【0423】

図72は、ニューラルネットワークベースのテンプレート生成器1512のクラスター中心予測に基づいて画像位置合わせ用のビーズを使用する一実施例である。

【0424】

図73は、ニューラルネットワークベースのテンプレート生成器1512によって識別されたクラスターのクラスター統計の一実施態様を示す。クラスター統計は、寄与するサブピクセル数及びGC含有量に基づくクラスターサイズを含む。

【0425】

図74は、入力画像データ1702が使用される初期配列決定サイクルの数が5~7増加すると、ニューラルネットワークベースのテンプレート生成器1512が隣接するクラスター間を区別する能力がどのように改善されるかを示す。5つの配列決定サイクルに関して、単一のクラスターは、連続するサブピクセルの単一の不連続領域によって識別され

10

20

30

40

50

る。7つの配列決定サイクルに関して、単一クラスターは、隣接するサブピクセルの独自の不連続領域をそれぞれ有する2つの隣接するクラスターに分割される。

【0426】

図75は、非COM位置がクラスター中心として使用されるときとは対照的に、RTAベースコーラーがクラスター中心として質量（COM）位置のグラウンドトゥールスを使用するときのベースコール性能の差を示す。

【0427】

図76は、追加の検出されたクラスター上のニューラルネットワークベースのテンプレート生成器1512の性能を示す。

【0428】

図77は、ニューラルネットワークベースのテンプレート生成器1512を訓練するために使用される異なるデータセットを示す。

【0429】

図78は、一実施態様に係る、ベースコールのためにRTAベースコーラーによって使用される処理段階を示す。図78はまた、2つの実施態様に係る、ベースコールのために、開示されたニューラルネットワークベースのベースコーラーによって使用される処理段階を示す。図78に示されるように、ニューラルネットワークベースのベースコーラー1514は、RTAベースコーラーによって使用される処理段階の多くを除去することによって、ベースコールプロセスを合理化し得る。合理化は、ベースコール精度及びスケールを改善する。ニューラルネットワークベースのベースコーラー1514の第1の実施態様では、それは、ニューラルネットワークベースのテンプレート生成器1512の出力から識別されるクラスター中心の場所／位置情報を使用してベースコールを実行する。第2の実施態様では、ニューラルネットワークベースのベースコーラー1514は、ベースコールのためにクラスター中心の場所／位置情報を使用しない。第2の実施態様は、パターン化されたフローセル設計がクラスター生成のために使用されるときに使用される。パターン化されたフローセルは、既知の基準場所に対して正確に配置されており、かつパターン化されたフローセル上に事前配置されたクラスター分布を提供する、ナノウェルを含む。他の実施態様では、ニューラルネットワークベースのベースコーラー1514は、ランダムフローセル上で生成されるクラスターをベースコールする。

（ニューラルネットワークベースのベースコール）

【0430】

ここでの考察は、ニューラルネットワークが訓練されて、配列決定画像をベースコールにマッピングするニューラルネットワークベースのベースコールについて説明する。考察は、以下のように構成される。最初に、ニューラルネットワークへの入力の説明される。次に、ニューラルネットワークの構造及び形態が説明される。最後に、ニューラルネットワークの出力が説明される。

入力

【0431】

図79は、ニューラルネットワーク7906を使用するベースコールの一実施態様を示す。

主入力：画像チャネル

【0432】

ニューラルネットワーク7906への主入力は、画像データ7902である。画像データ7902は、配列決定動作中にシーケンサ102によって生成される配列決定画像108から導出される。一実施態様では、画像データ7902は、配列決定画像102から抽出される $n \times n$ の画像パッチを含み、 n は、1～10、000の範囲の任意の数である。配列決定動作は、対応する m 個の画像チャネルに対して配列決定サイクルごとに m 個の画像（複数可）を生成し、画像パッチは、 m 個の画像（複数可）の各々から抽出されて、特定の配列決定サイクルで画像データを調製する。4-、2-、及び1-チャンネル化学などの異なる実施態様では、 m は、4又は2である。他の実施態様では、 m は、1、3、又は

10

20

30

40

50

4 よりも大きい。画像データ 7902 は、いくつかの実施態様では、光学ピクセルドメイン内にあり、他の実施態様では、アップサンプリングされたサブピクセルドメイン内にある。

【0433】

画像データ 7902 は、複数の配列決定サイクル（例えば、現在の配列決定サイクル、1 つ又はそれ以上の先行する配列決定サイクル、及び 1 つ又はそれ以上の連続する配列決定サイクル）のデータを含む。一実施態様では、画像データ 7902 は、3 回の配列決定サイクルのデータを含み、そのため、ベースコールされる現在の（時間 t ）配列決定サイクルのデータは、（ i ）左側のフランキング／コンテキスト／以前の／先行する／前の（時間 $t-1$ ）配列決定サイクルのデータ、及び（ i ）右側のフランキング／コンテキスト／次の／連続する／後続の（時間 $t+1$ ）配列決定サイクルのデータを伴う。他の実施態様では、画像データ 7902 は、単一の配列決定サイクルのデータを含む。

10

【0434】

画像データ 7902 は、1 つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示す。一実施態様では、単一のターゲットクラスターがベースコールされる時、画像パッチは、各画像パッチがその中心ピクセル内にターゲットクラスターの中心を含むような方法で、配列決定画像 108 から抽出され、概念は、本明細書で「ターゲットクラスター中心のパッチ抽出」と称される。

【0435】

画像データ 7902 は、強度チャンネル（画像チャンネルとも呼ばれる）を使用して入力データ 7904 で符号化される。特定の配列決定サイクルのためにシーケンサ 102 から取得された m 個の画像の各々に対して、別個の画像チャンネルを使用して、その強度データを符号化する。例えば、配列決定動作が、各配列決定サイクルで赤色画像及び緑色画像を生成する 2-チャンネル化学を使用すると考え、次いで、入力データ 7904 は、（ i ）赤色画像内に捕捉される 1 つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示す、 $n \times n$ のピクセルを有する第 1 の赤色画像チャンネルと、（ i ）緑色画像内に捕捉される 1 つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示す、 $n \times n$ のピクセルを有する第 2 の緑色画像チャンネルと、を含む。

20

補足入力：距離チャンネル

【0436】

画像データ 7902 は、補足距離データ（距離チャンネルとも呼ばれる）を伴う。距離チャンネルは、画像チャンネルから生成される特徴マップに組み込まれる加法バイアスを供給する。この加法バイアスは、距離チャンネルでピクセルごとに符号化されるピクセル中心クラスター中心（複数可）間距離に基づくため、ベースコール精度に寄与する。

30

【0437】

「単一のターゲットクラスター」ベースコール実施態様では、入力データ 7904 内の各画像チャンネル（画像パッチ）に対して、補足距離チャンネルは、その中心ピクセルを含むターゲットクラスターの中心からのそのピクセルの中心の距離を識別し、ベースコールされる。それによって、距離チャンネルは、画像パッチの中心ピクセルからの画像パッチのピクセルのそれぞれの距離を示す。

40

【0438】

「マルチクラスター」ベースコール実施態様では、入力データ 7904 内の各画像チャンネル（画像パッチ）に対して、補足距離チャンネルは、ピクセルとクラスターの各々との間の中心間距離に基づいて選択されるクラスターのうちの最も近い 1 つから、各ピクセルの中心間距離を識別する。

【0439】

「マルチクラスター形状ベースの」ベースコール実施態様では、入力データ 7904 内の各画像チャンネル（画像パッチ）に対して、補足距離チャンネルは、各クラスターピクセルを 1 つのクラスターのみで分類することに基づいて選択される、割り当てられたクラスターから、各クラスターピクセルの中心間距離を識別する。

50

補足入力：スケーリングチャネル

【0440】

画像データ7902は、異なるクラスターサイズ及び不均一な照明条件を考慮する、補足スケーリングデータ（スケーリングチャネルとも呼ばれる）を伴う。スケーリングチャネルはまた、画像チャネルから生成される特徴マップに組み込まれる加法バイアスを供給する。この加法バイアスは、スケーリングチャネルでピクセルごとに符号化される中心クラスターピクセル（複数可）の平均強度に基づくため、ベースコール精度に寄与する。

補足入力：クラスター中心座標

【0441】

いくつかの実施態様では、ニューラルネットワークベースのテンプレート生成器1512の出力から識別されるクラスター中心（複数可）の場所／位置情報7916（例えば、x-y座標）は、ニューラルネットワーク7906への補足入力として供給される。

補足入力：クラスター属性情報

【0442】

いくつかの実施態様では、ニューラルネットワーク7906は、どのピクセル又はサブピクセルが、同じクラスターを示し／同じクラスターに寄与し／同じクラスターに属する、背景ピクセル又はサブピクセル、クラスター中心ピクセル又はサブピクセル、及びクラスター／クラスター内部ピクセル又はサブピクセルであるかを分類する、クラスター属性情報を補足入力として受信する。他の実施態様では、減衰マップ、バイナリマップ、及び／若しくは三元マップ、又はそれらの変動は、ニューラルネットワーク7906への補足入力として供給される。

前処理：強度修正

【0443】

いくつかの実施態様では、入力データ7904は、距離チャネルを含まないが、代わりにニューラルネットワーク7906は、ニューラルネットワークベースのテンプレート生成器1512の出力、すなわち、減衰マップ、バイナリマップ、及び／又は三元マップの出力に基づいて修正される、修正画像データを入力として受信する。そのような実施態様では、画像データ7902の強度は、非存在距離チャネルを考慮するように修正される。

【0444】

他の実施態様では、画像データ7902は、1つ又はそれ以上の無損失変換動作（例えば、畳み込み、逆畳み込み、フーリエ変換）に供され、得られた修正画像データは、ニューラルネットワーク7906への入力として供給される。

ネットワーク構造及び形態

【0445】

ニューラルネットワーク7906はまた、本明細書で「ニューラルネットワークベースのベースコーラー」1514と称される。一実施態様では、ニューラルネットワークベースのベースコーラー1514は、多層パーセプトロン（Multilayer Perceptron、MLP）である。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、フィードフォワードニューラルネットワークである。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、完全に接続されたニューラルネットワークである。更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、完全畳み込みニューラルネットワークである。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、セマンティックセグメンテーションニューラルネットワークである。

【0446】

一実施態様では、ニューラルネットワークベースのベースコーラー1514は、複数の畳み込み層を有する畳み込みニューラルネットワーク（CNN）である。別の実施態様では、それは、長い短期メモリネットワーク（LSTM）、双方向LSTM（Bi-directional LSTM、Bi-LSTM）、又はゲートされた反復単位（GRU）などの反復ニューラルネットワーク（RNN）である。更に別の実施態様では、それは、CNN及びRNNの

10

20

30

40

50

両方を含む。

【0447】

更に他の実施態様では、ニューラルネットワークベースのベースコーラー1514は、1D畳み込み、2D畳み込み、3D畳み込み、4D畳み込み、5D畳み込み、拡張又は膨張畳み込み、転置畳み込み、奥行分離可能な畳み込み、pointwise畳み込み、1×1畳み込み、グループ畳み込み、扁平畳み込み、空間及びクロスチャネルの畳み込み、シャッフルされたグループ化畳み込み、空間的な分離可能な畳み込み、並びに逆畳み込みを使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はGRUなど）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（ReLU）、漏れやすいReLU、指数関数的ライナーユニット（ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

10

20

【0448】

ニューラルネットワークベースのベースコーラー1514は、入力データ7904を処理し、入力データ7904の代替表現7908を生成する。代替表現7908は、いくつかの実施態様では畳み込み表現であり、他の実施態様では隠れ表現である。次いで、代替表現7908は、出力層7910によって処理されて、出力7912を生成する。出力7912は、以下で説明するように、ベースコール（複数可）を生成するために使用される。

（出力）

【0449】

一実施態様では、ニューラルネットワークベースのベースコーラー1514は、特定の配列決定サイクルで単一のターゲットクラスターのためのベースコールを出力する。別の実施態様では、それは、特定の配列決定サイクルで複数のターゲットクラスター内の各ターゲットクラスターのためのベースコールを出力する。更に別の実施態様では、それは、複数の配列決定サイクル内の各配列決定サイクルで複数のターゲットクラスター内の各ターゲットクラスターのためのベースコールを出力し、それによって、各ターゲットクラスターのためのベースコール配列を生成する。

30

距離チャネル計算

【0450】

ここでの考察は、距離チャネルの距離値を計算する際に使用するために、クラスター中心（複数可）の適切な場所／位置情報（例えば、x-y座標）がどのように取得されるかについて説明する。

40

座標のダウンスケール

【0451】

図80は、サブピクセルドメインからピクセルドメインに、ニューラルネットワークベースのテンプレート生成器1512の出力から識別されるクラスター中心の場所／位置情報を変換する一実施態様である。

【0452】

クラスター中心場所／位置情報は、少なくとも、(i)中心ピクセルでベースコールされるターゲットクラスターの中心を含む配列決定画像108から画像パッチを抽出することによって入力データを構築するために、(ii)中心ピクセルが含まれるターゲットクラスターの中心からの画像パッチのピクセルの中心の距離を識別する距離チャネルを構築

50

するために、及び／又は (i i i) ニューラルネットワークベースのベースコーラー 1 5 1 4 への補足入力 7 9 1 6 として、ニューラルネットワークベースのベースコールのために使用される。

【 0 4 5 3 】

いくつかの実施態様では、クラスター中心場所／位置情報は、アップサンプリングされたサブピクセル解像度内のニューラルネットワークベースのテンプレート生成器 1 5 1 2 の出力から識別される。しかしながら、いくつかの実施態様では、ニューラルネットワークベースのベースコーラー 1 5 1 4 は、光学ピクセル解像度内である画像データ上で動作する。したがって、一実施態様では、クラスター中心場所／位置情報は、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 への入力として供給される画像データをアップサンプリングするために使用される同じアップサンプリング係数によって、クラスター中心の座標をダウンスケールすることによって、ピクセルドメインに変換される。

【 0 4 5 4 】

例えば、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 への入力として供給される画像パッチデータが、アップサンプリング係数 f によって、いくつかの初期配列決定サイクルから配列決定画像 1 0 8 をアップサンプリングすることによって導出されると考える。次いで、一実施態様では、ポストプロセッサ 1 8 1 4 によってニューラルネットワークベースのテンプレート生成器 1 5 1 2 によって生成され、テンプレート／テンプレート画像 8 0 0 4 内に記憶される、クラスター中心 8 0 0 2 の座標は、 f (除数) で除算される。これらのダウンスケールされたクラスター中心座標は、本明細書で「参照クラスター中心」8 0 0 8 と称され、テンプレート／テンプレート画像 8 0 0 4 内に記憶される。一実施態様では、ダウンスケールは、ダウンスケーラ 8 0 0 6 によって実行される。

座標の変換

【 0 4 5 5 】

図 8 1 は、参照クラスター中心 8 0 0 8 からいわゆる「変換されたクラスター中心」8 1 0 4 を導出するために、サイクル固有及び画像チャンネル固有の変換を使用する一実施態様である。そのようにする動機は、最初に説明される。

【 0 4 5 6 】

異なる配列決定サイクルで撮られる配列決定画像は、位置合わせ不良であり、互いにランダムな並進オフセットを有する。これは、センサーの運動段階の移動の有限精度に起因し、また、異なる画像／周波数チャンネルで撮られる画像が異なる光路及び波長を有するために生じる。その結果、参照クラスター中心と、配列決定画像内のクラスター中心の場所／位置との間にオフセットが存在する。このオフセットは、異なる配列決定サイクルで捕捉される画像間、及び異なる画像チャンネル内の同じ配列決定サイクルで捕捉される画像内で変化する。

【 0 4 5 7 】

このオフセットを考慮するために、サイクル固有及び画像チャンネル固有の変換が、参照クラスター中心に適用されて、各配列決定サイクルの画像パッチに対するそれぞれの変換されたクラスター中心を生成する。サイクル固有及び画像チャンネル固有の変換は、画像相関を使用して、完全な 6 パラメータのアフィン変換 (例えば、並進、回転、スケーリング、剪断、右鏡映、左鏡映) 又はプロクラスティクス変換 (例えば、並進、回転、スケーリング、任意選択的にアスペクト比への拡張) を決定する画像位置合わせプロセスによって導出され、その追加の詳細は、付録 1、2、3、及び 4 で見出され得る。

【 0 4 5 8 】

例えば、4 つのクラスター中心に対する参照クラスター中心は、 (x_1, y_1) ; (x_2, y_2) ; (x_3, y_3) ; (x_4, y_4) であり、配列決定動作は、赤色画像及び緑色画像が各配列決定サイクルで生成される 2-チャンネル化学を使用すると考える。次いで、例えば、配列決定サイクル 3、サイクル固有及び画像チャンネル固有の変換は、赤色画像に対して

【数 1】

$$\{\alpha_r^3, \beta_r^3, \chi_r^3, \delta_r^3, \varepsilon_r^3, \phi_r^3\}$$

であり、緑色画像に対して
【数 2】

$$\{\alpha_g^3, \beta_g^3, \chi_g^3, \delta_g^3, \varepsilon_g^3, \phi_g^3\}$$

10

である。
【0 4 5 9】

同様に、例えば、配列決定サイクル 9、サイクル固有及び画像チャネル固有の変換は、赤色画像に対して

【数 3】

$$\{\alpha_r^9, \beta_r^9, \chi_r^9, \delta_r^9, \varepsilon_r^9, \phi_r^9\}$$

20

であり、緑色画像に対して
【数 4】

$$\{\alpha_g^9, \beta_g^9, \chi_g^9, \delta_g^9, \varepsilon_g^9, \phi_g^9\}$$

である。
【0 4 6 0】

30

次いで、配列決定サイクル 3
【数 5】

$$(\hat{x}_1, \hat{y}_1); (\hat{x}_2, \hat{y}_2); (\hat{x}_3, \hat{y}_3); (\hat{x}_4, \hat{y}_4)$$

の赤色画像に対する変換されたクラスター中心は、変換
【数 6】

40

$$\{\alpha_r^3, \beta_r^3, \chi_r^3, \delta_r^3, \varepsilon_r^3, \phi_r^3\}$$

を参照クラスター中心 (x_1, y_1) ; (x_2, y_2) ; (x_3, y_3) ; (x_4, y_4)
に適用することによって導出され、配列決定サイクル 3
【数 7】

50

$$(\tilde{x}_1, \tilde{y}_1); (\tilde{x}_2, \tilde{y}_2); (\tilde{x}_3, \tilde{y}_3); (\tilde{x}_4, \tilde{y}_4)$$

の緑色画像に対する変換されたクラスター中心は、変換
【数 8】

$$\{\alpha_g^3, \beta_g^3, \chi_g^3, \delta_g^3, \varepsilon_g^3, \phi_g^3\}$$

10

を参照クラスター中心 $(x_1, y_1); (x_2, y_2); (x_3, y_3); (x_4, y_4)$
に適用することによって導出される。

【0 4 6 1】

同様に、配列決定サイクル 9

【数 9】

$$(\tilde{x}_1, \tilde{y}_1); (\tilde{x}_2, \tilde{y}_2); (\tilde{x}_3, \tilde{y}_3); (\tilde{x}_4, \tilde{y}_4)$$

20

の赤色画像に対する変換されたクラスター中心は、変換
【数 1 0】

$$\{\alpha_r^9, \beta_r^9, \chi_r^9, \delta_r^9, \varepsilon_r^9, \phi_r^9\}$$

30

を参照クラスター中心 $(x_1, y_1); (x_2, y_2); (x_3, y_3); (x_4, y_4)$
に適用することによって導出され、配列決定サイクル 9

【数 1 1】

$$(\tilde{x}_1, \tilde{y}_1); (\tilde{x}_2, \tilde{y}_2); (\tilde{x}_3, \tilde{y}_3); (\tilde{x}_4, \tilde{y}_4)$$

40

の緑色画像に対する変換されたクラスター中心は、変換
【数 1 2】

$$\{\alpha_g^9, \beta_g^9, \chi_g^9, \delta_g^9, \varepsilon_g^9, \phi_g^9\}$$

を参照クラスター中心 $(x_1, y_1); (x_2, y_2); (x_3, y_3); (x_4, y_4)$

50

)に適用することによって導出される。

【0462】

一実施態様では、変換は、変換器8102によって実行される。

【0463】

変換されたクラスター中心8104は、テンプレート／テンプレート画像8004内に記憶され、(i)(例えば、パッチ抽出器8106によって)対応する配列決定画像108からパッチ抽出を行うために、(ii)距離式

【数13】

$$(d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2})$$

10

で、対応する画像パッチに対する距離チャンネルを計算するために、かつ(iii)ベースコールされている対応する配列決定サイクルに対するニューラルネットワークベースのベースコーラー1514への補足入力として、それぞれ使用される。他の実施態様では、距離2乗、 $e^{-\text{距離}}$ 、及び $e^{-\text{距離}^2}$ などの異なる距離式が使用され得る。

画像パッチ

【0464】

図82は、ニューラルネットワークベースのベースコーラー1514に供給される入力データの一部である画像パッチ8202を示す。入力データは、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列を含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つに対する画像パッチを有する。

20

【0465】

例えば、配列決定動作は、各配列決定サイクルで赤色画像及び緑色画像を生成する2チャンネル化学を使用し、入力データは、配列決定動作の一連の3回の配列決定サイクル、すなわち、ベースコールされる現在の(時間t)配列決定サイクル、以前の(時間t-1)配列決定サイクル、及び次の(時間t+1)配列決定サイクルに及ぶデータを含むと考える。

30

【0466】

次いで、入力データは、サイクルごとの画像パッチセットの以下の配列、すなわち、現在の配列決定サイクルで捕捉される赤色及び緑色配列決定画像からそれぞれ抽出される、現在の赤色画像パッチ及び現在の緑色画像パッチを有する現在のサイクル画像パッチセット、以前の配列決定サイクルで捕捉される赤色及び緑色配列決定画像からそれぞれ抽出される、以前の赤色画像パッチ及び以前の緑色画像パッチを有する以前のサイクル画像パッチセット、並びに次の配列決定サイクルで捕捉される赤色及び緑色配列決定画像からそれぞれ抽出される、次の赤色画像パッチ及び次の緑色画像パッチを有する次のサイクル画像パッチセットを含む。

40

【0467】

各画像パッチのサイズは、 $n \times n$ であり得、 n は、1~10,000の範囲の任意の数であり得る。各画像パッチは、光学ピクセルドメイン、又はアップサンプリングされたサブピクセルドメイン内にあり得る。図82に示される実施態様では、抽出された画像ページ8202は、複数のクラスター1-m及びそれらの周囲の背景をカバーし／示すピクセルに対するピクセル強度データを有する。また、図示された実施態様では、画像パッチ8202は、その中心ピクセル内に含まれ、ターゲットクラスターの中心がベースコールされているような方法で抽出される。

【0468】

図82では、ピクセル中心は、黒い矩形によって示されており、整数場所／位置座標を

50

有し、クラスター中心は、紫色の円によって示されており、フローティングポイント場所／位置座標を有する。

(単一のターゲットクラスターに対する距離計算)

【0469】

図83は、単一のターゲットクラスターが、ニューラルネットワークベースのベースコーラー1514によってベースコールされているときに、距離チャンネルの距離値8302を決定する一実施態様を示す。ターゲットクラスターの中心は、ニューラルネットワークベースのベースコーラー1514への入力として供給される画像パッチの中心ピクセルに含まれる。距離値は、ピクセルごとに計算され、そのため、各ピクセルに対して、その中心とターゲットクラスターの中心との間の距離が決定される。したがって、入力データの一部である画像パッチの各々における各ピクセルに対して、距離値が計算される。

10

【0470】

図83は、特定の画像パッチに対する3つの距離値 d_1 、 d_c 、及び d_n を示す。一実施態様では、距離値8302は、以下の距離式を使用して計算される。すなわち、

【数14】

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

20

で、変換されたクラスター中心8104で動作する。他の実施態様では、距離2乗、 $e^{\wedge}$ -距離、及び $e^{\wedge}$ -距離2乗などの異なる距離式が使用され得る。

【0471】

他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、距離値8302は、サブピクセルドメイン内で計算される。

【0472】

したがって、単一のターゲットクラスターベースコール実施態様では、距離チャンネルは、ベースコールされているターゲットクラスターに関してのみ計算される。

【0473】

図84は、ピクセルとターゲットクラスターとの間で計算される距離値8302を符号化8402するピクセルごとの一実施態様を示す。一実施態様では、入力データで、距離チャンネルの一部としての距離値8302は、対応する各画像チャンネル(画像パッチ)を「ピクセル距離データ」として補う。配列決定サイクルごとに生成されている赤色画像及び緑色画像の例に戻って、入力データは、赤色画像チャンネル及び緑色画像チャンネルをピクセル距離データとしてそれぞれ補う、赤色距離チャンネル及び緑色距離チャンネルを含む。

30

【0474】

他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、距離チャンネルは、サブピクセルごとに符号化される。

(複数のターゲットクラスターに対する距離計算)

40

【0475】

図85aは、複数のターゲットクラスター1-mが、ニューラルネットワークベースのベースコーラー1514によって同時にベースコールされているときに、距離チャンネルの距離値8502を決定する一実施態様を示す。距離値は、ピクセルごとに計算され、そのため、各ピクセルに対して、その中心と複数のクラスター1-mの各々のそれぞれの中心との間の距離が決定され、(赤色での)最小距離値がピクセルに割り当てられる。

【0476】

したがって、距離チャンネルは、ピクセルとクラスターの各々との間の中心間距離に基づいて選択されるクラスターのうちの最も近い1つから、各ピクセルの中心間距離を識別する。図示された実施態様では、図85aは、2つのピクセル及び4つのクラスター中心に

50

対するピクセル中心クラスター中心間距離を示す。ピクセル 1 は、クラスター 1 に最も近く、ピクセル n は、クラスター 3 に最も近い。

【0477】

一実施態様では、距離値 8502 は、以下の距離式を使用して計算される。

【数15】

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

10

で、変換されたクラスター中心 8104 で動作する。他の実施態様では、距離 2 乗、 $e^{\wedge}$ - 距離、及び $e^{\wedge}$ - 距離 2 乗などの異なる距離式が使用され得る。

【0478】

他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、距離値 8502 は、サブピクセルドメイン内で計算される。

【0479】

したがって、マルチクラスターベースコール実施態様では、距離チャンネルは、複数のクラスターの中から最も近いクラスターに関して計算される。

【0480】

図 85b は、ターゲットクラスター 1-m の各々に対して、ピクセル中心最も近いクラスター中心間距離 8504 (d_1 、 d_2 、 d_{23} 、 d_{29} 、 d_{24} 、 d_{32} 、 d_n 、 d_{13} 、 d_{14} など) に基づいて決定される、いくつかの最も近いピクセルを示す。

【0481】

図 86 は、ピクセルとクラスターのうちの最も近い 1 つとの間で計算される最小距離値を符号化 8602 するピクセルごとの一実施態様を示す。他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、距離チャンネルは、サブピクセルごとに符号化される。

(クラスター形状に基づく複数のターゲットクラスターに対する距離計算)

【0482】

図 87 は、本明細書で「クラスター形状データ」又は「クラスター形状情報」と称される、ピクセルクラスター間の分類／属性／分類 8702 を使用して、複数のターゲットクラスター 1-m が、ニューラルネットワークベースのベースコーラー 1514 によって同時にベースコールされているときに、距離チャンネルのクラスター距離値 8802 を決定する、一実施態様を示す。最初に、以下のことは、クラスター形状データがどのように生成されるかについての簡単な検討である。

【0483】

上述のように、ニューラルネットワークベースのテンプレート生成器 1512 の出力は、同じクラスターを示し／同じクラスターに寄与し／同じクラスターに属する、背景ピクセル、中心ピクセル、及びクラスター／クラスター内部ピクセルとしてピクセルを分類するために使用される。このピクセルクラスター間分類情報は、ピクセル中心とクラスター中心との間の距離に関わらず、各ピクセルを 1 つのクラスターのみに属させるために使用され、クラスター形状データとして記憶される。

40

【0484】

図 87 に示される実施態様では、背景ピクセルは、グレーで着色され、クラスター 1 に属するピクセルは、黄色 (クラスター 1 ピクセル) で着色され、クラスター 2 に属するピクセルは、緑色 (クラスター 2 ピクセル) で着色され、クラスター 3 に属するピクセルは、赤色 (クラスター 3 ピクセル) で着色され、クラスター m に属するピクセルは、青色 (クラスター m ピクセル) で着色される。

【0485】

50

図 8 8 は、クラスター形状データを使用して距離値 8 8 0 2 を計算する一実施態様を示す。最初に、我々は、クラスター形状を考慮することなく計算される距離情報がなぜ誤差を生じやすいかを説明する。次いで、我々は、クラスター形状データがこの制限をどのように克服するかを説明する。

【0 4 8 6】

クラスター形状データを使用しない「マルチクラスター」ベースコール実施態様では（図 8 5 a - b 及び図 8 6）、ピクセルの中心間距離値は、複数のクラスターの中から最も近いクラスターに関して計算される。ここで、クラスター A に属するピクセルが、クラスター A の中心から更に離れているが、クラスター B の中心により近いときのシナリオを考える。そのような場合、クラスター形状データなしで、ピクセルは、（真に属する）クラスター A に向かい合って距離値を割り当てられる代わりに、（属していない）クラスター B に関して計算される距離値を割り当てられる。

【0 4 8 7】

「マルチクラスター形状ベースの」ベースコール実施態様は、生画像データ内で定義され、かつニューラルネットワークベースのテンプレート生成器 1 5 1 2 によって生成されるように、真のピクセルクラスター間マッピングを使用することによって、これを回避する。

【0 4 8 8】

2 つの実施態様間のコントラストは、ピクセル 3 4 及び 3 5 に関して見られ得る。図 8 5 b では、ピクセル 3 4 及び 3 5 の距離値は、クラスター形状データを考慮することなく、クラスター 3 の最も近い中心に関して計算される。しかしながら、図 8 8 では、クラスター形状データに基づいて、ピクセル 3 4 及び 3 5 の距離値 8 8 0 2 は、（それらが実際に属する）クラスター 2 に関して計算される。

【0 4 8 9】

図 8 8 では、クラスターピクセルは、クラスター強度を示し、背景ピクセルは、背景強度を示す。クラスター距離値は、クラスターのうちの 1 つのみに各クラスターピクセルを分類することに基づいて選択されるクラスターのうちの割り当てられた 1 つから、各クラスターピクセルの中心間距離を識別する。いくつかの実施態様では、背景ピクセルは、0 若しくは 0 . 1、又はいくつかの他の最小値などの所定の背景距離値を割り当てられる。

【0 4 9 0】

一実施態様では、上述のように、クラスター距離値 8 8 0 2 は、以下の距離式を使用して計算される。

【数 1 6】

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

で、変換されたクラスター中心 8 1 0 4 で動作する。他の実施態様では、距離 2 乗、 e^{-} 距離、及び e^{-} 距離 2 乗などの異なる距離式が使用され得る。

【0 4 9 1】

他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、クラスター距離値 8 8 0 2 は、サブピクセルドメイン内で計算され、クラスター及び背景属性 8 7 0 2 は、サブピクセルごとに生じる。

【0 4 9 2】

したがって、マルチクラスター形状ベースのベースコール実施態様では、距離チャンネルは、複数のクラスターの中から割り当てられたクラスターに関して計算される。割り当てられたクラスターは、生画像データ内で定義される真のピクセルクラスター間マッピングに従ってクラスターのうちの 1 つのみに各クラスターピクセルを分類することに基づいて

選択される。

【0493】

図89は、ピクセルと割り当てられたクラスターとの間で計算される距離値8702を符号化するピクセルごとの一実施態様を示す。他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、距離チャンネルは、サブピクセルごとに符号化される。

【0494】

深層学習は、多層ニューラルネットワークを使用する強力な機械学習技術である。コンピュータビジョン及び画像処理ドメイン内の1つの特に成功したネットワーク構造は、畳み込みニューラルネットワーク(CNN)であり、各層は、入力テンソル(画像状、多次元高密度アレイ)から異なる形状の出力テンソルへのフィードフォワード畳み込み変換を実行する。CNNは、最大3-又は4-Dのアレイ上で訓練を高速にする、画像の空間コヒーレンス及び汎用グラフィック処理ユニット(Graphics Processing Unit、GPU)の出現に起因する画像状入力に特に適する。これらの画像状特性を活用することは、サポートベクトルマシン(Support Vector Machine、SVM)又は多層パーセプトロン(MLP)などの他の学習方法と比較して、優れた経験的性能につながる。

【0495】

我々は、画像データ並びに補足距離及びスケールリングデータの両方を取り扱うために標準的なCNNを増強する専用構造を導入する。以下の詳細がある。

(専用構造)

【0496】

図90は、異なる配列決定サイクルでデータの処理を分離するために使用されるニューラルネットワークベースのベースコーラー1514の専用構造の一実施態様を示す。専用構造を使用する動機は、最初に説明される。

【0497】

上述のように、ニューラルネットワークベースのベースコーラー1514は、現在の配列決定サイクル、1つ又はそれ以上の先行する配列決定サイクル、及び1つ又はそれ以上の連続する配列決定サイクルでデータを処理する。追加の配列決定サイクルに対するデータは、配列固有のコンテキストを提供する。ニューラルネットワークベースのベースコーラー1514は、訓練中に配列固有のコンテキストを学習し、それらをベースコールする。更に、事前及び事後配列決定サイクルに対するデータは、プレフェイジング及びフェイジング信号の2次の寄与を現在の配列決定サイクルに提供する。

(空間畳み込み層)

【0498】

しかしながら、上述のように、異なる配列決定サイクルで、及び異なる画像チャンネル内に捕捉される画像は、位置合わせ不良であり、互いに残留位置合わせ誤差を有する。この位置合わせ不良を考慮するために、専用構造は、配列決定サイクル間で情報を混合せず、配列決定サイクル内で情報を混合するだけである、空間畳み込み層を含む。

【0499】

空間畳み込み層は、畳み込みの「専用の非共有」配列を介して複数の配列決定サイクルの各々に対して独立してデータを処理することによって分離を操作する、いわゆる「分離された畳み込み」を使用する。分離された畳み込みは、任意の他の配列決定サイクルのデータ及び得られた特徴マップ上で畳み込むことなく、所与の配列決定サイクル、すなわち、サイクル内のみのデータ及び得られた特徴マップ上で畳み込む。

【0500】

例えば、入力データが、(i)ベースコールされる現在の(時間t)配列決定サイクルに対する現在のデータと、(ii)以前の(時間t-1)配列決定サイクルに対する以前のデータと、(iii)次の(時間t+1)配列決定サイクルに対する次のデータと、を含むと考える。次いで、専用構造は、3つの別個のデータ処理パイプライン(又は畳み込みパイプライン)、すなわち、現在のデータ処理パイプライン、以前のデータ処理パイプ

10

20

30

40

50

ライン、及び次のデータ処理パイプラインを開始する。現在のデータ処理パイプラインは、現在の（時間 t ）配列決定サイクルに対する現在のデータを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「現在の空間畳み込み表現」を生成する。以前のデータ処理パイプラインは、以前の（時間 $t-1$ ）配列決定サイクルに対する以前のデータを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「以前の空間畳み込み表現」を生成する。次のデータ処理パイプラインは、次の（時間 $t+1$ ）配列決定サイクルに対する次のデータを入力として受信し、複数の空間畳み込み層を介して独立してそれを処理して、最終空間畳み込み層の出力としていわゆる「次の空間畳み込み表現」を生成する。

10

【0501】

いくつかの実施態様では、現在、以前、及び次の処理パイプラインは、並行して実行される。

【0502】

いくつかの実施態様では、空間畳み込み層は、専用構造内の空間畳み込みネットワーク（又はサブネットワーク）の一部である。

（時間的畳み込み層）

【0503】

ニューラルネットワークベースのベースコーラー1514は、配列決定サイクル間、すなわち、サイクル間で情報を混合する時間的畳み込み層を更に含む。時間的畳み込み層は、空間畳み込みネットワークからそれらの入力を受信し、それぞれのデータ処理パイプラインに対して最終空間畳み込み層によって生成される空間畳み込み表現で動作する。

20

【0504】

時間的畳み込み層のサイクル間動作性自由度は、空間畳み込みネットワークへの入力として供給される画像データ内に存在する位置合わせ不良特性が、空間畳み込み層の配列によって実行される、分離された畳み込みのカスケードによって空間畳み込み表現からパージされるという事実から生じる。

【0505】

時間的畳み込み層は、スライディングウィンドウベースでの連続する入力で入力チャンネル上でグループごとに畳み込む、いわゆる「組み合わせの畳み込み」を使用する。一実施態様では、連続する入力は、以前の空間畳み込み層又は以前の時間的畳み込み層によって生成される連続する出力である。

30

【0506】

いくつかの実施態様では、時間的畳み込み層は、専用構造内の時間的畳み込みネットワーク（又はサブネットワーク）の一部である。時間的畳み込みネットワークは、空間畳み込みネットワークからその入力を受信する。一実施態様では、時間的畳み込みネットワークの第1の時間的畳み込み層は、配列決定サイクル間の空間畳み込み表現をグループごとに組み合わせる。別の実施態様では、時間的畳み込みネットワークの後続の時間的畳み込み層は、以前の時間的畳み込み層の連続する出力を組み合わせる。

【0507】

最終時間的畳み込み層の出力は、出力を生成する出力層に供給される。出力は、1つ又はそれ以上の配列決定サイクルで1つ又はそれ以上のクラスターをベースコールするために使用される。

40

【0508】

以下のことは、分離された及び組み合わせの畳み込みのより詳細な考察である。

（分離された畳み込み）

【0509】

前方伝搬の間、専用構造は、2つの段階で複数の入力からの情報を処理する。第1の段階では、分離畳み込みは、入力間の情報の混合を防止するために使用される。第2の段階では、組み合わせの畳み込みは、入力間の情報を混合するために使用される。第2の段階

50

からの結果は、複数の入力に対して単一の推測を行うために使用される。

【0510】

これは、バッチモード技術とは異なり、畳み込み層は、バッチ内の複数の入力を同時に処理し、バッチ内の各入力に対して対応する推測を行う。対照的に、専用構造は、複数の入力を単一の推測にマッピングする。単一の推測は、4つの塩基（A、C、T、及びG）の各々に対する分類スコアなどの2つ又はそれ以上の予測を含み得る。

【0511】

一実施態様では、入力は、各入力異なる時間ステップで生成され、かつ複数の入力チャネルを有するように、時間的順序付けを有する。例えば、複数の入力は、以下の3つの入力、すなわち：時間ステップ（ t ）で現在の配列決定サイクルによって生成される現在の入力と、時間ステップ（ $t-1$ ）で以前の配列決定サイクルによって生成される以前の入力と、時間ステップ（ $t+1$ ）で次の配列決定サイクルによって生成される次の入力と、を含み得る。別の実施態様では、各入力は、1つ又はそれ以上の以前の畳み込み層によって現在の、以前の、及び次の入力からそれぞれ導出され、 k 個の特徴マップを含む。

【0512】

一実施態様では、各入力は、以下の5つの入力チャネル、すなわち、赤色画像チャネル（赤色）と、赤色距離チャネル（黄色）と、緑色画像チャネル（緑色）と、緑色距離チャネル（紫色）と、スケーリングチャネル（青色）と、を含み得る。別の実施態様では、各入力は、以前の畳み込み層によって生成される k 個の特徴マップを含み得、各特徴マップは、入力チャネルとして処理される。

【0513】

図91は、分離された畳み込みの一実施態様を示す。分離された畳み込みは、畳み込みフィルタを各入力に並行して適用することによって、複数の入力を一度に処理する。分離された畳み込みでは、畳み込みフィルタは、同じ入力内で入力チャネルを組み合わせ、異なる入力内で入力チャネルを組み合わせない。一実施態様では、同じ畳み込みフィルタは、各入力に並行して適用される。別の実施態様では、異なる畳み込みフィルタは、各入力に並行して適用される。いくつかの実施態様では、各空間畳み込み層は、 k 個の畳み込みフィルタのバンクを含み、その各々は、各入力に並行して適用される。

（組み合わせの畳み込み）

【0514】

組み合わせの畳み込みは、異なる入力の対応する入力チャネルをグループ化し、畳み込みフィルタを各グループに適用することによって、異なる入力間で情報を混合する。対応する入力チャネルのグループ化及び畳み込みフィルタの適用は、スライディングウィンドウベースで生じる。このコンテキストでは、ウィンドウは、例えば、2つの連続する配列決定サイクルに対する出力を表す、2つ又はそれ以上の連続する入力チャネルに及ぶ。ウィンドウがスライドウィンドウであるため、最も多くの入力チャネルは、2つ又はそれ以上のウィンドウで使用される。

【0515】

いくつかの実施態様では、異なる入力は、先行する空間又は時間的畳み込み層によって生成される出力配列から生じる。出力配列では、異なる入力は、連続する出力として配置され、したがって、連続する入力として次の時間的畳み込み層によって観察される。次いで、次の時間的畳み込み層では、組み合わせの畳み込みは、連続する入力内の対応する入力チャネルのグループに畳み込みフィルタを適用する。

【0516】

一実施態様では、連続する入力は、現在の入力が：時間ステップ（ t ）で現在の配列決定サイクルによって生成され、以前の入力が時間ステップ（ $t-1$ ）で以前の配列決定サイクルによって生成され、次の入力が時間ステップ（ $t+1$ ）で次の配列決定サイクルによって生成されるように、時間的順序付けを有する。別の実施態様では、各連続する入力は、1つ又はそれ以上の以前の畳み込み層によって現在の、以前の、及び次の入力からそれぞれ導出され、 k 個の特徴マップを含む。

【0517】

一実施態様では、各入力は、以下の5つの入力チャンネル、すなわち、赤色画像チャンネル（赤色）と、赤色距離チャンネル（黄色）と、緑色画像チャンネル（緑色）と、緑色距離チャンネル（紫色）と、スケーリングチャンネル（青色）と、を含み得る。別の実施態様では、各入力は、以前の畳み込み層によって生成される k 個の特徴マップを含み得、各特徴マップは、入力チャンネルとして処理される。

【0518】

畳み込みフィルタの深さ B は、対応する入力チャンネルがスライディングウィンドウベースで畳み込みフィルタによってグループごとに畳み込まれる、連続する入力の数に依存する。言い換えると、深さ B は、各スライディングウィンドウ及びグループサイズ内の連続する入力の数と等しい。

【0519】

図92aでは、2つの連続する入力からの対応する入力チャンネルは、各スライディングウィンドウ内で組み合わせられ、したがって $B=2$ である。図92bでは、3つの連続する入力からの対応する入力チャンネルは、各スライディングウィンドウ内で組み合わせられ、したがって $B=3$ である。

【0520】

一実施態様では、スライディングウィンドウは、同じ畳み込みフィルタを共有する。別の実施態様では、異なる畳み込みフィルタが、各スライディングウィンドウに対して使用される。いくつかの実施態様では、各時間的畳み込み層は、 k 個の畳み込みフィルタのバンクを含み、その各々は、スライディングウィンドウベースの連続する入力に適用される。（フィルタバンク）

【0521】

図93は、各畳み込み層が畳み込みフィルタのバンクを有する、ニューラルネットワークベースのベースコーラー1514の畳み込み層の一実施態様を示す。図93では、5個の畳み込み層が示されており、その各々は、64個の畳み込みフィルタのバンクを有する。いくつかの実施態様では、各空間畳み込み層は、 k 個の畳み込みフィルタのバンクを有し、 k は、1、2、8、64、128、256などの任意の数であり得る。いくつかの実施態様では、各時間的畳み込み層は、 k 個の畳み込みフィルタのバンクを有し、 k は、1、2、8、64、128、256などの任意の数であり得る。

【0522】

ここでの考察は、補足スケーリングチャンネル及びどのようにそれが計算されるかについて説明する。

（スケーリングチャンネル）

【0523】

図94は、画像チャンネルを補うスケーリングチャンネルの2つの構成を示す。スケーリングチャンネルは、ニューラルネットワークベースのベースコーラー1514に供給される入力データでピクセルごとに符号化される。異なるクラスターサイズ及び不均一な照明条件は、抽出されている広範囲のクラスター強度をもたらす。スケーリングチャンネルによって供給される加法バイアスは、クラスターにわたってクラスター強度を同等にする。他の実施態様では、画像パッチがアップサンプリングされたサブピクセル解像度内であるとき、スケーリングチャンネルは、サブピクセルごとに符号化される。

【0524】

単一のターゲットクラスターがベースコールされているとき、スケーリングチャンネルは、同じスケーリング値を全てのピクセルに割り当てる。複数のターゲットクラスターが同時にベースコールされているとき、スケーリングチャンネルは、クラスター形状データに基づいて、異なるスケーリング値をピクセルのグループに割り当てる。

【0525】

スケーリングチャンネル9410は、全てのピクセルに対して同じスケーリング値（ s_1 ）を有する。スケーリング値（ s_1 ）は、ターゲットクラスターの中心を含む中心ピクセ

10

20

30

40

50

ルの平均強度に基づく。一実施態様では、平均強度は、ターゲットクラスターに対してA及びTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に中心ピクセル観察の強度値を平均することによって計算される。

【0526】

スケーリングチャンネル9408は、クラスター形状データに基づいて、対応するクラスターに属するそれぞれのピクセルグループの異なるスケーリング値(s1、s2、s3、sm)を有する。各ピクセルグループは、対応するクラスターの中心を含む中心クラスターピクセルを含む。特定のピクセルグループのスケーリング値は、その中心クラスターピクセルの平均強度に基づく。一実施態様では、平均強度は、対応するクラスターに対してA及びTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に中心クラスターピクセル観察の強度値を平均することによって計算される。

10

【0527】

いくつかの実施態様では、背景ピクセルは、0若しくは0.1、又はいくつかの他の最小値であり得る、背景スケーリング値(sb)を割り当てられる。

【0528】

一実施態様では、スケーリングチャンネル9406及びそれらのスケーリング値は、強度スケーラ9404によって決定される。強度スケーラ9404は、平均強度を計算するために、先行する配列決定サイクルからクラスター強度データ9402を使用する。

【0529】

他の実施態様では、補足スケーリングチャンネルは、ニューラルネットワークベースのベースコーラー1514の最後の層の前又はそこへ、ニューラルネットワークベースのベースコーラー1514の1つ又はそれ以上の中間層の前又はそこへ、及び画像サイズに一致するようにピクセルごとにそれを符号化する代わりに単一の値としてなど、異なる方法での入力として提供され得る。

20

【0530】

ここでの考察は、ニューラルネットワークベースのベースコーラー1514に供給される入力データについて説明する。

入力データ：画像チャンネル、距離チャンネル、及びスケーリングチャンネル

【0531】

図95aは、赤色画像及び緑色画像を生成する単一の配列決定サイクルの入力データ9500の一実施態様を示す。入力データ9500は、以下を含む。

30

- ・赤色画像から抽出される画像パッチ内のピクセルに対する赤色強度データ9502(赤色)。赤色強度データ9502は、赤色画像チャンネルで符号化される。

- ・赤色強度データ9502をピクセルごとに補う赤色距離データ9504(黄色)。赤色距離データ9504は、赤色距離チャンネルで符号化される。

- ・緑色画像から抽出される画像パッチ内のピクセルに対する緑色強度データ9506(緑色)。緑色強度データ9506は、緑色画像チャンネルで符号化される。

- ・緑色強度データ9506をピクセルごとに補う緑色距離データ9508(紫色)。緑色距離データ9508は、緑色距離チャンネルで符号化される。

- ・赤色強度データ9502及び緑色強度データ9506をピクセルごとに補うスケーリングデータ9510(青色)。スケーリングデータ9510は、スケーリングチャンネルで符号化されている。

40

【0532】

他の実施態様では、入力データは、より少ない又はより多くの数の画像チャンネル及び補足距離チャンネルを含み得る。一例では、4チャンネル化学を使用する配列決定動作に対して、入力データは、各配列決定サイクル及び4つの補足距離チャンネルに対して4つの画像チャンネルを含む。

【0533】

ここでの考察は、距離チャンネル及びスケーリングチャンネルがベースコール精度にどのように寄与するかについて説明する。

50

(加法バイアス)

【0534】

図95bは、画像チャンネルから生成される特徴マップに組み込まれる加法バイアスを供給する距離チャンネルの一実施態様を示す。この加法バイアスは、距離チャンネルでピクセルごとに符号化されるピクセル中心クラスター中心（複数可）間距離に基づくため、ベースコール精度に寄与する。

【0535】

平均して、約 3×3 のピクセルは、1つのクラスターを含む。クラスターの中心での密度は、クラスターが実質的に中心場所から外向きに成長するため、周囲よりも高いと予想される。周囲クラスターピクセルは、近くのクラスターからの競合する信号を含み得る。したがって、中心クラスターピクセルは、最大強度領域と見なされ、クラスターを確実に識別するピーコンとして機能する。

【0536】

画像パッチのピクセルは、複数のクラスター（例えば、10～200クラスター）及びそれらの周囲の背景の強度放射を示す。追加のクラスターは、より広い半径からの情報を組み込み、強度放射が画像パッチで示される、基礎塩基を識別することによって、ベースコール予測に寄与する。言い換えると、クラスターのグループからの強度放射は、別個の塩基（A、C、T、又はG）に割り当てられ得る強度パターンを累積的に生成する。

【0537】

我々は、補足距離チャンネル内のクラスター中心（複数可）から各ピクセルの畳み込みフィルタ距離に明示的に通信することが、より高いベースコール精度をもたらすことを観察する。距離チャンネルは、どのピクセルがクラスター中心を含み、どのピクセルがクラスター中心からより離れているかを畳み込みフィルタに伝える。畳み込みフィルタは、（b）周囲クラスターピクセル、背景ピクセル、及びそれらから導出される特徴マップよりも多くの、（a）中心クラスターピクセル、それらの隣接するピクセル、及びそれらから導出される特徴マップに対処することによって、配列決定信号をその適切なソースクラスターに割り当てるために、この情報を使用する。対処の一例では、距離チャンネルは、（a）から生じる特徴マップに組み込まれる正の加法バイアスを供給するが、（b）から生じる特徴マップに組み込まれる負の加法バイアスを供給する。

【0538】

距離チャンネルは、画像チャンネルと同じ次元性を有する。これにより、畳み込みフィルタが、局所的な受容フィールド内の画像チャンネル及び距離チャンネルを別々に評価し、評価をコヒーレントに組み合わせることが可能になる。

【0539】

単一のターゲットクラスターがベースコールされているとき、距離チャンネルは、画像パッチの中心で1つの中心クラスターピクセルのみを識別する。複数のターゲットクラスターが同時にベースコールされているとき、距離チャンネルは、画像パッチにわたって分布する複数の中心クラスターピクセルを識別する。

【0540】

「単一のクラスター」距離チャンネルは、その中心ピクセルでベースコールされる単一のターゲットクラスターの中心を含む画像パッチに適用される。単一のクラスター距離チャンネルは、単一のターゲットクラスターに対する画像パッチ内の各ピクセルの中心間距離を含む。この実施態様では、画像パッチはまた、単一のターゲットクラスターに隣接する追加のクラスターを含むが、追加のクラスターはベースコールされない。

【0541】

「マルチクラスター」距離チャンネルは、そのそれぞれの中心クラスターピクセルでベースコールされる複数のターゲットクラスターの中心を含む画像パッチに適用される。マルチクラスター距離チャンネルは、複数のターゲットクラスターの中から最も近いクラスターに対する画像パッチ内の各ピクセルの中心間距離を含む。これは、誤ったクラスターに対する中心間距離を測定する可能性を有するが、その可能性は低い。

10

20

30

40

50

【0542】

「マルチクラスター形状ベースの」距離チャンネルは、そのそれぞれの中心クラスターピクセルでベースコールされる複数のターゲットクラスターの中心を含み、ピクセルクラスター間属性情報が知られている、画像パッチに適用される。マルチクラスター距離チャンネルは、それが属するか、又は複数のターゲットクラスターの中から属するクラスターに対する画像パッチ内の各クラスターピクセルの中心間距離を含む。背景ピクセルは、所与の計算された距離の代わりに、背景としてフラグ付けされ得る。

【0543】

図95bはまた、画像チャンネルから生成される特徴マップに組み込まれる加法バイアスを供給するスケーリングチャンネルの一実施態様を示す。この加法バイアスは、スケーリングチャンネルでピクセルごとに符号化される中心クラスターピクセル（複数可）の平均強度に基づくため、ベースコール精度に寄与する。距離チャンネルのコンテキスト内の加法バイアスに関する考察は、スケーリングチャンネルに類似して適用される。

（加法バイアスの例）

【0544】

図95bは、どのように、加法バイアスが距離及びスケーリングチャンネルから導出され、画像チャンネルから生成される特徴マップに組み込まれるかの例を更に示す。

【0545】

図95b、では、畳み込みフィルタ9514は、2つの画像チャンネル9502及び9506、2つの距離チャンネル9504及び9508、並びにスケーリングチャンネル9510にわたって局所的な受容フィールド9512（マゼンタ）を評価する。距離及びスケーリングチャンネルが別々に符号化されるため、加法バイアスは、チャンネル固有の畳み込みカーネル（又は特徴検出器）9516a～e（更にバイアス9516f）の各々の中間出力9516a～eが、局所的な受容フィールド9512に対する最終出力／特徴マップ要素9520としてチャンネルごとに蓄積9518されるときに生じる。この例では、2つの距離チャンネル9504及び9508によって供給される加法バイアスはそれぞれ、中間出力9516b及び9516dである。スケーリングチャンネル9510によって供給される加法バイアスは、中間出力9516eである。

【0546】

加法バイアスは、ベースコールのためにより重要で信頼性が高いと見なされる画像チャンネル内のそれらの特徴、すなわち、中心クラスターピクセル及びそれらの隣接するピクセルのピクセル強度をより重視することによって、特徴マップコンパイルプロセスをガイドする。訓練中に、グラウンドトゥルースベースコールとの比較から計算される勾配の逆伝搬は、畳み込みカーネルの重みを更新して、中心クラスターピクセル及びそれらの隣接するピクセルに対するより強い活性化を生成する。

【0547】

例えば、局所的な受容フィールド9512によってカバーされる隣接するピクセルのグループ内のピクセルが、クラスター中心を含み、次いで、距離チャンネル9504及び9508が、クラスター中心に対するピクセルの近接を反映すると考える。その結果、強度中間出力9516a及び9516cがチャンネルごとの蓄積9518で距離チャンネル加法バイアス9516b及び9516dとマージされるとき、その結果は、ピクセルの正にバイアスされた畳み込み表現9520である。

【0548】

対照的に、局所的な受容フィールド9512によってカバーされるピクセルがクラスター中心の近くでない場合、距離チャンネル9504及び9508は、クラスター中心からのそれらの分離を反映する。その結果、強度中間出力9516a及び9516cがチャンネルごとの蓄積9518で距離チャンネル加法バイアス9516b及び9516dとマージされるとき、その結果は、ピクセルの負にバイアスされた畳み込み表現9520である。

【0549】

同様に、スケーリングチャンネル9510から導出されるスケーリングチャンネル加法バイ

10

20

30

40

50

アス 9 5 1 6 e は、ピクセルの畳み込み表現 9 5 2 0 を正に又は負にバイアスし得る。

【0 5 5 0】

明確さのために、図 9 5 b は、単一の配列決定サイクルで入力データ 9 5 0 0 に対する単一の畳み込みフィルタ i 9 5 1 4 の適用を示す。当業者は、考察が、複数の畳み込みフィルタ（例えば、k フィルタのフィルタバンク、k は、8、1 6、3 2、6 4、1 2 8、2 5 6 などであり得る）まで、複数の畳み込み層（例えば、複数の空間及び時間的畳み込み層）、及び複数の配列決定サイクル（例えば、t、t + 1、t - 1）まで拡張され得ることを理解するであろう。

【0 5 5 1】

他の実施態様では、距離及びスケールリングチャンネル並びに画像チャンネルが同じ次元性を有するため、距離及びスケールリングチャンネルは、別々に符号化される代わりに、画像チャンネルに直接適用されて、変調ピクセル乗算）を生成する。更なる実施態様では、畳み込みカーネルの重みは、要素ごとの乗算中に画像チャンネルで最も重要な特徴を検出するように、距離及び画像チャンネルに基づいて決定される。更に他の実施態様では、第 1 の層に供給される代わりに、距離及びスケールリングチャンネルは、下流層及び／又はネットワーク（例えば、完全に接続されたネットワーク又は分類層）への補助入力として提供される。更に更なる実施態様では、距離及びスケールリングチャンネルは、第 1 の層に供給され、下流層及び／又はネットワークに（例えば、残留接続を介して）再供給される。

【0 5 5 2】

上記の考察は、k 個の入力チャンネルを有する 2 D 入力データに対するものである。3 D 入力までの拡張は、当業者によって理解されるであろう。簡潔には、容積入力は、大きさ $k \times l \times w \times h$ を有する 4 D テンソルであり、l は、追加の大きさ、長さである。各個々のカーネルは、4 D テンソルでスライスされた 4 D テンソルであり、3 D テンソル（チャンネル次元は、全体にわたってスライスされないためにくずれる）をもたらす。

【0 5 5 3】

他の実施態様では、入力データ 9 5 0 0 がアップサンプリングされたサブピクセル解像度内であるとき、距離及びスケールリングチャンネルは、サブピクセルごとに別々に符号化され、加法バイアスは、サブピクセルレベルで生じる。

（専用構造及び入力データを使用するベースコール）

【0 5 5 4】

ここでの考察は、専用構造及び入力データがニューラルネットワークベースのベースコールのためにどのように使用されるかについて説明する。

（単一のクラスターベースコール）

【0 5 5 5】

図 9 6 a、図 9 6 b、及び図 9 6 c は、単一のターゲットクラスターをベースコールする一実施態様を示す。専用構造は、3 回の配列決定サイクル、すなわち、ベースコールされる現在の（時間 t）配列決定サイクル、以前の（時間 t - 1）配列決定サイクル、及び次の（時間 t + 1）配列決定サイクルに対する入力データを処理し、現在の（時間 t）配列決定サイクルで単一のターゲットクラスターに対するベースコールを生成する。

【0 5 5 6】

図 9 6 a 及び図 9 6 b は、空間畳み込み層を示す。図 9 6 c は、いくつかの他の非畳み込み層と共に、時間的畳み込み層を示す。図 9 6 a 及び図 9 6 b では、垂直の点線は、特徴マップからの空間畳み込み層の境界を定め、水平の点鎖線は、3 回の配列決定サイクルに対応する 3 つの畳み込みパイプラインの境界を定める。

【0 5 5 7】

各配列決定サイクルに対して、入力データは、次元性のテンソル $n \times n \times m$ （例えば、図 9 5 a での入力テンソル 9 5 0 0）を含み、n は、正方形テンソルの幅及び高さを表し、m は、入力チャンネルの数を表し、3 回のサイクル $n \times n \times m \times t$ に対する入力データの次元性を作成する。

【0 5 5 8】

ここで、サイクルごとの各テンソルは、その画像チャネルの中心ピクセルで、単一のターゲットクラスターの中心を含む。それはまた、いくつかの隣接するクラスターの、及び特定の配列決定サイクルで画像チャネルの各々で捕捉されるそれらの周囲の背景の、単一のターゲットクラスターの強度放射を示す。図96aでは、2つの例示的画像チャネル、すなわち、赤色画像チャネル及び緑色画像チャネルが示されている。

【0559】

サイクルごとの各テンソルはまた、対応する画像チャネル（例えば、赤色距離チャネル及び緑色距離チャネル）を補う距離チャネルを含む。距離チャネルは、単一のターゲットクラスターに対する、対応する画像チャネル内の各ピクセルの中心間距離を識別する。サイクルごとの各テンソルは、画像チャネルの各々で強度値をピクセルごとにスケールするスケールリングチャネルを更に含む。

10

【0560】

専用構造は、5つの空間畳み込み層と、2つの時間的畳み込み層と、を有する。各空間畳み込み層は、次元性 $j \times j \times \vartheta$ の k 個の畳み込みフィルタのバンクを使用する分離された畳み込みを適用し、 j は、正方形フィルタの幅及び高さを表し、 ϑ はその深さを表す。各時間的畳み込み層は、次元性 $j \times j \times \alpha$ の k 個の畳み込みフィルタのバンクを使用する組み合わせの畳み込みを適用し、 j は、正方形フィルタの幅及び高さを表し、 α はその深さを表す。

【0561】

専用構造は、事前分類層（例えば、平坦化層及び高密度層）と、出力層（例えば、ソフトマックス分類層）と、を有する。事前分類層は、出力層に対する入力を調製する。出力層は、現在の（時間 t ）配列決定サイクルで、単一のターゲットクラスターに対するベースコールを生成する。

20

（空間次元性の一貫した低減）

【0562】

図96a、図96b、及び図96cはまた、畳み込みフィルタによって生成される、得られた特徴マップ（畳み込み表現若しくは中間畳み込み表現若しくは畳み込み特徴又は活性化マップ）を示す。サイクルごとのテンソルから開始して、得られた特徴マップの空間次元性は、1つの畳み込み層から次まで一定のステップサイズによって低減し、概念は、本明細書で「空間次元性の一貫した低減」と称される。図96a、図96b、及び図96cでは、空間次元性の一貫した低減のために、2の例示的一定のステップサイズが使用される。

30

【0563】

空間次元性の一貫した低減は、以下の式によって表現される。「現在の特征マップ空間次元性 = 以前の特征マップ空間次元性 - 畳み込みフィルタ空間次元性 + 1」空間次元性の一貫した低減により、畳み込みフィルタは、中心クラスターピクセル及びそれらの隣接するピクセルに対する注目の焦点を漸進的に狭め、中心クラスターピクセル及びそれらの隣接するピクセル間の局所的な依存性を捕捉する特徴を有する特徴マップを生成する。これは、次に、中心が中心クラスターピクセルに含まれるクラスターを正確にベースコールするのに役立つ。

40

【0564】

5つの空間畳み込み層の分離された畳み込みは、3つの配列決定サイクル間の情報の混合を防止し、3つの別個の畳み込みパイプラインを維持する。

【0565】

2つの時間的畳み込み層の組み合わせの畳み込みは、3つの配列決定サイクル間で情報を混合する。第1の時間的畳み込み層は、最終空間畳み込み層によって次及び現在の配列決定サイクルに対してそれぞれ生成される、次及び現在の空間畳み込み表現上で畳み込む。これは、第1の時間的出力をもたらす。第1の時間的畳み込み層はまた、最終空間畳み込み層によって現在及び以前の配列決定サイクルに対してそれぞれ生成される、現在及び以前の空間畳み込み表現上で畳み込む。これは、第2の時間的出力をもたらす。第2の時

50

間的畳み込み層は、第1及び第2の時間的出力上で畳み込み、最終時間的出力を生成する。

【0566】

いくつかの実施態様では、最終時間的出力は、平坦化層に供給されて、平坦化された出力を生成する。次いで、平坦化された出力は、高密度層に供給されて、高密度の出力を生成する。高密度出力は、出力層によって処理されて、現在の（時間 t ）配列決定サイクルで、単一のターゲットクラスターに対するベースコールを生成する。

【0567】

いくつかの実施態様では、出力層は、現在の配列決定サイクルで単一のターゲットクラスターに組み込まれる塩基がA、C、T、及びGである尤度（分類スコア）を生成し、尤度に基づいてA、C、T、又はGとして塩基を分類する（例えば、図96aでの塩基Aのような、最大尤度を有する塩基が選択される）。そのような実施態様では、尤度は、ソフトマックス分類層によって生成される、指数関数的に正規化されたスコアであり、1になる。

【0568】

いくつかの実施態様では、出力層は、単一のターゲットクラスターに対する出力ペアを導出する。出力ペアは、現在の配列決定サイクルで単一のターゲットクラスターに組み込まれる塩基がA、C、T、又はGであるクラスラベルを識別し、クラスラベルに基づいて、単一のターゲットクラスターをベースコールする。一実施態様では、1、0のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、1、1のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。別の実施態様では、1、1のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、0、5、0、5のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。更に別の実施態様では、1、0のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、0、5、0、5のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。更に更なる実施態様では、1、2のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、1、1のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。

【0569】

いくつかの実施態様では、出力層は、現在の配列決定サイクルで単一のターゲットクラスターに組み込まれる塩基がA、C、T、又はGであることを識別する、単一のターゲットクラスターに対するクラスラベルを導出し、クラスラベルに基づいて、単一のターゲットクラスターをベースコールする。一実施態様では、0、33のクラスラベルは、A塩基を識別し、0、66のクラスラベルは、C塩基を識別し、1のクラスラベルは、T塩基を識別し、0のクラスラベルは、G塩基を識別する。別の実施態様では、0、50のクラスラベルは、A塩基を識別し、0、75のクラスラベルは、C塩基を識別し、1のクラスラベルは、T塩基を識別し、0、25のクラスラベルは、G塩基を識別する。

【0570】

いくつかの実施態様では、出力層は、単一の出力値を導出し、塩基A、C、T、及びGに対応するクラス値範囲に対して単一の出力値を比較し、比較に基づいて、単一の出力値を特定のクラス値範囲に割り当て、割り当てに基づいて、単一のターゲットクラスターをベースコールする。一実施態様では、単一の出力値は、シグモイド関数を使用して導出され、単一の出力値は、0～1の範囲である。別の実施態様では、0～0、25のクラス値範囲は、A塩基を表し、0、25～0、50のクラス値範囲は、C塩基を表し、0、50～0、75のクラス値範囲は、T塩基を表し、0、75～1のクラス値範囲は、G塩基を表す。

【0571】

当業者は、他の実施態様では、専用構造が、より少ない又はより多くの数の配列決定サイクルに対する入力データを処理し得、より少ない又はより多くの数の空間及び時間的畳み込み層を含み得ることを理解するであろう。また、入力データの次元性、入力データ内のサイクルごとのテンソル、畳み込みフィルタ、得られた特徴マップ、及び出力は異なり

得る。また、畳み込み層内の畳み込みフィルタの数は異なり得る。それは、異なるパディング及びストリディング構成を使用し得る。それは、異なる分類機能（例えば、シグモイド又は回帰）を使用し得、完全に接続された層を含んでも含まなくてもよい。それは、1D重畳、2D重畳、3D重畳、4D重畳、5D重畳、拡張又は無性重畳、転置重畳、奥行分離可能な重畳、 1×1 重畳、グループ重畳、扁平重畳、空間及びクロスチャネルの重畳、シャッフルされたグループ化重畳、空間的な分離可能な重畳、及び逆重畳を使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はGRUなど）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（ReLU）、漏れやすいReLU、指数関数的ライナーユニット（ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

【0572】

単一のクラスターベースコールを説明してきたが、ここでの考察は、複数のクラスターベースコールについて説明する。

（複数のクラスターベースコール）

【0573】

入力データのサイズ及びフローセル上のクラスター密度に応じて、10～30万個のクラスターのいずれかは、入力ごとにニューラルネットワークベースのベースコーラー1514によって同時にベースコールされる。これを、並列プロセッサ上で実装されるデータ並列及び／又はモデル並列戦略に拡張し、バッチ又はサイズ10のミニバッチを使用することは、100～300万個のクラスターがバッチ又はミニバッチごとに同時にベースコールされることをもたらす。

【0574】

配列決定構成（例えば、クラスター密度、フローセル上のタイルの数）に応じて、タイルは、2万～30万個のクラスターを含む。別の実施態様では、IlluminaのNovaSeqシーケンサは、タイル当たり最大400万個のクラスターを有する。したがって、タイル（タイル画像）の配列決定画像は、2万～30万個のクラスター及びそれらの周囲の背景からの強度放射を示し得る。そのため、一実施態様では、全体のタイル画像を含む入力データを使用することは、30万個のクラスターが入力ごとに同時にベースコールされることをもたらす。別の実施態様では、入力データ内のサイズ 15×15 のピクセルの画像パッチを使用することは、100個未満のクラスターが入力ごとに同時にベースコールされることをもたらす。当業者は、これらの数が、配列決定構成、並列戦略、（例えば、最適構造ハイパーパラメータに基づく）構造の詳細、及び利用可能な計算に応じて変化し得ることを理解するであろう。

【0575】

図97は、複数のターゲットクラスターを同時にベースコールする一実施態様を示す。入力データは、上述の3つの配列決定サイクルに対する3つのテンソルを有する。サイクルごとの各テンソル（例えば、図95aでの入力テンソル9500）は、ベースコールされる複数のターゲットクラスター、及び特定の配列決定サイクルで画像チャネルの各々で捕捉されるそれらの周囲の背景の強度放射を示す。他の実施態様では、ベースコールされないいくつかの追加の隣接するクラスターはまた、コンテキストに含まれる。

【0576】

マルチクラスターベースコール実施態様では、サイクルごとの各テンソルは、対応する画像チャンネル（例えば、赤色距離チャンネル及び緑色距離チャンネル）を補う距離チャンネルを含む。距離チャンネルは、複数のターゲットクラスターの中から最も近いクラスターに対する、対応する画像チャンネル内の各ピクセルの中心間距離を識別する。

【0577】

マルチクラスター形状ベースのベースコール実施態様では、サイクルごとの各テンソルは、対応する画像チャンネル（例えば、赤色距離チャンネル及び緑色距離チャンネル）を補う距離チャンネルを含む。距離チャンネルは、それが属するか、又は複数のターゲットクラスターの中から属するクラスターに対する、対応する画像チャンネル内の各クラスターピクセルの中心間距離を識別する。

【0578】

サイクルごとの各テンソルは、画像チャンネルの各々で強度値をピクセルごとにスケールするスケーリングチャンネルを更に含む。

【0579】

図97では、サイクルごとの各テンソルの空間次元性は、図96aに示されるものよりも大きい。すなわち、図96aでの単一のターゲットクラスターベースコール実施態様では、サイクルごとの各テンソルの空間次元性は、 15×15 であるのに対し、図97での複数のクラスターベースコール実施態様では、サイクルごとの各テンソルの空間次元性は、 114×114 である。いくつかの実施態様によれば、追加のクラスターの強度放射を示すピクセル化データの量を多く有することにより、複数のクラスターに対して同時に予測されるベースコールの精度が改善する。

（冗長な畳み込みの回避）

【0580】

更に、サイクルごとの各テンソル内の画像チャンネルは、配列決定画像から抽出される画像パッチから取得される。いくつかの実施態様では、空間的に連続（例えば、左、右、上、及び底で連続）している、抽出された画像パッチ間に重複ピクセルが存在する。したがって、一実施態様では、重複ピクセルは、冗長な畳み込みを受けず、前の畳み込みからの結果は、重複ピクセルが後続の入力の一部であるとき、後続のインスタンスで再使用される。

【0581】

例えば、サイズ $n \times n$ のピクセルの第1の画像パッチは、配列決定画像から抽出され、サイズ $m \times m$ のピクセルの第2の画像パッチはまた、同じ配列決定画像から抽出され、そのため、第1及び第2の画像パッチは、空間的に連続しており、 $o \times o$ のピクセルの重複領域を共有すると考える。更に、 $o \times o$ のピクセルは、第1の画像パッチの一部として畳み込まれて、メモリ内に記憶される第1の畳み込み表現を生成すると考える。次いで、第2の画像パッチが畳み込まれるとき、 $o \times o$ のピクセルは、再び畳み込みされず、その代わりに、第1の畳み込み表現は、メモリから取得され、再使用される。いくつかの実施態様では、 $n = m$ である。他の実施態様では、それらは等しくない。

【0582】

次いで、入力データは、専用構造の空間及び時間的畳み込み層を介して処理されて、次元性 $w \times w \times k$ の最終時間的出力を生成する。ここでも、空間次元性事象の一貫した低減の下、空間次元性は、各畳み込み層で2の一定のステップサイズによって低減される。すなわち、入力データの $n \times n$ の空間次元性で開始して、最終時間的出力の $w \times w$ の空間次元性が導出される。

【0583】

次いで、空間次元性 $w \times w$ の最終時間的出力に基づいて、出力層は、ユニットの $w \times w$ のセット内の各ユニットに対するベースコールを生成する。一実施態様では、出力層は、ユニットごとに4つの塩基（A、C、T、及びG）に対する4つの分類スコアを生成するソフトマックス層である。すなわち、ユニットの $w \times w$ のセット内の各ユニットは、図97に示されるように、対応するソフトマックススコールドブルでの最大分類スコアに基づい

10

20

30

40

50

て、ベースコールを割り当てられる。いくつかの実施態様では、ユニットの $w \times w$ のセットは、平坦化層及び高密度層を介して最終時間的出力を処理する結果として導出されて、平坦化された出力及び高密度出力をそれぞれ生成する。そのような実施態様では、平坦化された出力は、 $w \times w \times k$ の要素を有し、高密度出力は、ユニットの $w \times w$ のセットを形成する $w \times w$ の要素を有する。

【0584】

複数のターゲットクラスターに対するベースコールは、ユニットの $w \times w$ のセット内のベースコールされたユニットのうちのどれが、中心クラスターピクセル、すなわち、複数のターゲットクラスターのそれぞれの中心を含む入力データ内のピクセルと一致するか、又はそれに対応するかを識別することによって取得される。所与のターゲットクラスターは、所与のターゲットクラスターの中心を含むピクセルと一致するか、又はそれに対応する、ユニットのベースコールを割り当てられる。言い換えると、中心クラスターピクセルと一致しないか、又はそれに対応しないユニットのベースコールは、フィルタリングされる。この機能は、いくつかの実施態様では専用構造の一部であるか、又は他の実施態様では後処理モジュールとして実装される、ベースコールフィルタリング層によって操作される。

【0585】

他の実施態様では、複数のターゲットクラスターに対するベースコールは、ユニットの $w \times w$ のセット内のベースコールされたユニットのどのグループが同じクラスターをカバーするかを識別することによって、すなわち、同じクラスターを示す入力データ内のピクセルグループを識別することによって取得される。次いで、各クラスター及びその対応するピクセルグループに対して、それぞれの4つの塩基クラス（A、C、T、及びG）の分類スコア（ソフトマックス確率）の平均は、ピクセルグループ内のピクセルにわたって計算され、最も高い平均分類スコアを有する塩基クラスは、クラスターをベースコールするために選択される。

【0586】

訓練中に、いくつかの実施態様では、グラウンドトゥールズ比較及び誤差計算は、中心クラスターピクセルと一致するか、又はそれに対応するそれらのユニットに対してのみ生じ、そのため、それらの予測されたベースコールは、グラウンドトゥールズラベルとして識別される正しいベースコールに対して評価される。

【0587】

複数のクラスターベースコールを説明してきたが、ここでの考察は、複数のクラスター及び複数のサイクルベースコールについて説明する。

（複数のクラスター及び複数のサイクルベースコール）

【0588】

図98は、複数の連続する配列決定サイクルで複数のターゲットクラスターを同時にベースコールし、それによって、複数のターゲットクラスターの各々に対するベースコール配列を同時に生成する、一実施態様を示す。

【0589】

上述の単一及び複数のベースコール実施態様では、1回の配列決定サイクル（現在の（時間 t ）配列決定サイクル）でのベースコールは、3回の配列決定サイクル（現在の（時間 t ）、以前の／左側のフランキング（時間 $t-1$ ）、及び次の／右側のフランキング（時間 $t+1$ ）配列決定サイクル）に対するデータを使用して予測され、右側及び左側のフランキング配列決定サイクルは、塩基トリプレットモチーフに対する配列固有のコンテキスト、並びにプレフェイジング及びフェイジング信号の2次の寄与を提供する。この関係は、以下の式によって表現される。「どのデータが入力データ（ t ）に含まれるかに対する配列決定サイクルの数＝ベースコールされている配列決定サイクルの数（ y ）＋右側及び左側のフランキング配列決定サイクルの数（ x ）」。

【0590】

図98では、入力データは、 t 回の配列決定サイクルに対する t 個のサイクルごとのデ

ンソルを含み、その次元性 $n \times n \times m \times t$ を作成し、 $n = 114$ であり、 $m = 5$ であり、 $t = 15$ である。他の実施態様では、これらの次元性は異なる。 t 回の配列決定サイクルのうち、第 t 番目の配列決定サイクル及び第 1 の配列決定サイクルは、右側及び左側のランキングコンテキスト x として機能し、それらの間の y 回の配列決定サイクルは、ベースコールされる。したがって、 $y = 13$ 、 $x = 2$ 、及び $t = y + x$ である。サイクルごとの各テンソルは、図 95 a での入力テンソル 9500 などの、画像チャンネル、対応する距離チャンネル、及びスケージングチャンネルを含む。

【0591】

次いで、 t 個のサイクルごとのテンソルを有する入力データは、専用構造の空間及び時間的畳み込み層を介して処理されて、 y 個の最終時間的出力を生成し、その各々は、ベースコールされている y 回の配列決定サイクルのうちのそれぞれ 1 つに対応する。 y 個の最終時間的出力の各々は、 $w \times w \times k$ の次元性を有する。ここでも、空間次元性事象の一貫した低減の下、空間次元性は、各畳み込み層で 2 の一定のステップサイズによって低減される。すなわち、入力データの $n \times n$ の空間次元性で開始して、 y 個の最終時間的出力の各々の $w \times w$ の空間次元性が導出される。

【0592】

次いで、 y 個の最終時間的出力の各々は、出力層によって並行して処理される。 y 個の最終時間的出力の各々に対して、出力層は、ユニットの $w \times w$ のセット内の各ユニットに対するベースコールを生成する。一実施態様では、出力層は、ユニットごとに 4 つの塩基 (A、C、T、及び G) に対する 4 つの分類スコアを生成するソフトマックス層である。すなわち、ユニットの $w \times w$ のセット内の各ユニットは、図 97 に示されるように、対応するソフトマックススコールドブルでの最大分類スコアに基づいて、ベースコールを割り当てられる。いくつかの実施態様では、ユニットの $w \times w$ のセットは、平坦化層及び高密度層を介して後にそれぞれ処理する結果として、 y 個の最終時間的出力の各々に対して導出されて、対応する平坦化された出力及び高密度出力を生成する。そのような実施態様では、平坦化された各出力は、 $w \times w \times k$ の要素を有し、各高密度出力は、ユニットの $w \times w$ のセットを形成する $w \times w$ の要素を有する。

【0593】

y 回の配列決定サイクルの各々に対して、複数のターゲットクラスターに対するベースコールは、ユニットの対応する $w \times w$ のセット内のベースコールされたユニットのうちのどれが、中心クラスターピクセル、すなわち、複数のターゲットクラスターのそれぞれの中心を含む入力データ内のピクセルと一致するか、又はそれに対応するかを識別することによって取得される。所与のターゲットクラスターは、所与のターゲットクラスターの中心を含むピクセルと一致するか、又はそれに対応する、ユニットのベースコールを割り当てられる。言い換えると、中心クラスターピクセルと一致しないか、又はそれに対応しないユニットのベースコールは、フィルタリングされる。この機能は、いくつかの実施態様では専用構造の一部であるか、又は他の実施態様では後処理モジュールとして実装される、ベースコールフィルタリング層によって操作される。

【0594】

訓練中に、いくつかの実施態様では、グラウンドトゥールズ比較及び誤差計算は、中心クラスターピクセルと一致するか、又はそれに対応するそれらのユニットに対してのみ生じ、そのため、それらの予測されたベースコールは、グラウンドトゥールズラベルとして識別される正しいベースコールに対して評価される。

【0595】

入力ごとに、その結果は、 y 回の配列決定サイクルの各々での複数のターゲットクラスターの各々に対するベースコール、すなわち、複数のターゲットクラスターの各々に対する長さ y のベースコール配列である。他の実施態様では、 y は、20、30、50、150、300 などである。当業者は、これらの数が、配列決定構成、並列戦略、(例えば、最適構造ハイパーパラメータに基づく) 構造の詳細、及び利用可能な計算に応じて変化し得ることを理解するであろう。

10

20

30

40

50

(エンドツーエンド次元性図)

【0596】

以下の説明は、上記データ次元性変化を実現するデータオペレータの次元性と共に、画像データからのベースコールの生成に關与する基礎データ次元性変化の異なる実施態様を示すために、次元性図を使用する。

【0597】

図99、図100、及び図101では、矩形は、空間及び時間的畳み込み層並びにソフトマックス分類層のようなデータオペレータを表し、角が丸みを帯びた矩形は、データオペレータによって生成されるデータ（例えば、特徴マップ）を表す。

【0598】

図99は、単一のクラスターベースコール実施態様に対する次元性図9900を示す。入力の「サイクル次元」は3であり、第1の時間的畳み込み層まで、得られた特徴マップに対するものであり続けることに留意されたい。3のサイクル次元は、3回の配列決定サイクルを表し、その連続性は、3回の配列決定サイクルに対する特徴マップが別々に生成され畳み込まれ、3回の配列決定サイクルの間に特徴が混合されないことを表す。分離された畳み込みパイプラインは、空間畳み込み層の深さごとの分離された畳み込みフィルタによって実現される。空間畳み込み層の深さごとの分離された畳み込みフィルタの「深さ次元性」は、1であることに留意されたい。これは、深さごとの分離された畳み込みフィルタが、所与の配列決定サイクル、すなわち、サイクル内のみのデータ及び得られた特徴マップ上で畳み込むことを可能にし、それらが任意の他の配列決定サイクルのデータ及び得られた特徴マップ上で畳み込むことを防止するものである。

【0599】

対照的に、時間的畳み込み層の深さごとの組み合わせの畳み込みフィルタの深さ次元性は、2であることに留意されたい。これは、深さごとの組み合わせの畳み込みフィルタが、複数の配列決定サイクルから得られた特徴マップ上でグループごとに畳み込み、かつ配列決定サイクル間の特徴を混合することを可能にするものである。

【0600】

また、2の一定のステップサイズによる「空間次元性」での一貫した低減に留意されたい。

【0601】

更に、4つの要素を有するベクトルは、ソフトマックス層によって指数関数的に正規化されて、4つの塩基（A、C、T、及びG）に対する分類スコア（すなわち、信頼スコア、確率、尤度、ソフトマックススコア）を生成する。最も高い（最大）ソフトマックススコアを有する塩基は、現在の配列決定サイクルでベースコールされている単一のターゲットクラスターを割り当てられる。

【0602】

当業者は、他の実施態様では、示された次元性が、配列決定構成、並列戦略、（例えば、最適構造ハイパーパラメータに基づく）構造の詳細、及び利用可能な計算に応じて変化する理解するであろう。

【0603】

図100は、複数のクラスターにおける次元性図10000、単一の配列決定サイクルベースコール実施態様を示す。単一のクラスターベースコールに関するサイクル、深さ、及び空間次元性に関する上記の考察は、この実施態様に適用される。

【0604】

ここで、ソフトマックス層は、10,000ユニットの各々に対して独立して動作し、10,000ユニットの各々に対するソフトマックススコアのそれぞれのクォドルプルを生成する。クォドルプルは、4つの塩基（A、C、T、及びG）に対応する。いくつかの実施態様では、10,000ユニットは、64,000平坦化ユニットの10,000高密度ユニットへの変換から導出される。

【0605】

次いで、10,000ユニットの各々のソフトマックススコアクォドルプルから、各クォドルプルでの最も高いソフトマックススコアを有する塩基は、10,000ユニットのうちのそれぞれの1つに割り当てられる。

【0606】

次いで、10,000ユニットのうち、現在の配列決定サイクルで同時にベースコールされている、2,500個のターゲットクラスターのそれぞれの中心を含む2,500個の中心クラスターピクセルに対応する2,500ユニットが選択される。選択された2,500ユニットに割り当てられる塩基は、次に、2,500個のターゲットクラスターのうちの対応するものに割り当てられる。

【0607】

当業者は、他の実施態様では、示された次元性が、配列決定構成、並列戦略、（例えば、最適構造ハイパーパラメータに基づく）構造の詳細、及び利用可能な計算に応じて変化し得ることを理解するであろう。

【0608】

図101は、複数のクラスターにおける次元性図10100、複数の配列決定サイクルベースコール実施態様を示す。単一のクラスターベースコールに関するサイクル、深さ、及び空間次元性に関する上記の考察は、この実施態様に適用される。

【0609】

更に、複数のクラスターベースコールに関する、ソフトマックスベースのベースコール分類に関する上記の考察は、ここでも適用される。しかしながら、ここで、2,500個のターゲットクラスターのソフトマックスベースのベースコール分類は、ベースコールされる13個の配列決定サイクルの各々に対して並行して生じ、それによって、2,500個のターゲットクラスターの各々に対する13個のベースコールを同時に生成する。

【0610】

当業者は、他の実施態様では、示された次元性が、配列決定構成、並列戦略、（例えば、最適構造ハイパーパラメータに基づく）構造の詳細、及び利用可能な計算に応じて変化し得ることを理解するであろう。

（アレイ入力v / s スタック入力）

【0611】

ここでの考察は、ニューラルネットワークベースのコラーへのマルチサイクル入力データが配置され得る2つの構成について説明する。第1の構成は、「アレイ入力」と呼ばれ、第2の構成は、「スタック入力」と呼ばれる。アレイ入力は、図102aに示され、図96a～図101に関して上述されている。アレイ入力は、サイクルごとの入力での画像パッチが、残留位置合わせ誤差に起因して互いに位置合わせ不良であるため、別個の列／ブロックで各配列決定サイクルの入力を符号化する。専用構造は、別個の列／ブロックの各々の処理を分離するために、配列された入力で使用される。また、距離チャンネルは、変換されたクラスター中心を使用して計算されて、サイクル内の画像パッチ間及びサイクルにわたる画像パッチ間の位置合わせ不良を考慮する。

【0612】

対照的に、図102bに示されるスタック入力は、単一の列／ブロックで異なる配列決定サイクルからの入力を符号化する。一実施態様では、これは、スタック入力内の画像パッチが、サイクル間及びサイクル内残留位置合わせ誤差を排除する、アフィン変換及び強度補間を介して互いに位置合わせされるため、専用構造を使用する必要性を除去する。いくつかの実施態様では、スタック入力は、全ての入力に対して共通のスケーリングチャンネルを有する。

【0613】

別の実施態様では、各画像パッチの中心ピクセルの中心が、ベースコールされている単一のターゲットクラスターの中心と一致するように、画像パッチを再構成又はシフトするために強度補間が使用される。これは、全ての非中心ピクセルが単一のターゲットクラスターの中心から等距離であるため、補足距離チャンネルを使用する必要性を除去する。距離

10

20

30

40

50

チャンネルなしのスタック入力、本明細書で「再構成された入力」と称され、図104に示される。

【0614】

しかしながら、そこで画像パッチが、ベースコールされる複数の中心クラスターピクセルを含むため、再構成は、複数のクラスターを伴うベースコール実施態様で実現可能でない場合がある。距離チャンネルなしかつ再構成なしのスタック入力は、本明細書で「位置合わせされた入力」と称され、図105及び図106に示される。距離チャンネルの計算が所望されないとき（例えば、計算制限に起因して）、位置合わせされた入力を使用され得、再構成は、実現可能でない。

【0615】

以下のセクションは、専用構造及び補足距離チャンネルを使用せず、その代わりに、標準的な畳み込み層及びフィルタを使用する、様々なベースコール実施態様を説明する。

再構成された入力：距離チャンネルなしの位置合わせされた画像パッチ

【0616】

図103aは、中心ピクセルでベースコールされているターゲットクラスターの中心を中心にするために画像パッチ10302のピクセルを再構成10300aする一実施態様を示す。図10300aに示されるように、ターゲットクラスターの中心（紫色）は、画像パッチ10302の中心ピクセル内にあるが、中心ピクセルの中心からのオフセット（赤色）にある。

【0617】

オフセットを排除するために、リフレーマ10304は、ピクセルの強度を補間することによって画像パッチ10302をシフトさせて、再構成を補償し、再構成／シフトされた画像パッチ10306を生成する。シフトされた画像パッチ10306では、中心ピクセルの中心は、ターゲットクラスターの中心と一致する。また、非中心ピクセルは、ターゲットクラスターの中心から等距離である。補間は、最近傍強度抽出、ガウス系強度抽出、平均2×2のサブピクセル領域に基づく強度抽出、最も明るい2×2のサブピクセル領域に基づく強度抽出、平均3×3のサブピクセル領域に基づく強度抽出、バイリニア強度抽出、双次強度抽出、及び／又は重み付け領域被覆に基づく強度抽出により実行され得る。これらの技術は、「強度抽出方法」と題された付録に詳細に記載されている。

【0618】

図103bは、(i) 中心ピクセルの中心がターゲットクラスターの中心と一致し、(ii) 非中心ピクセルがターゲットクラスターの中心から等距離である、別の例示的再構成／シフトされた画像パッチ10300bを示す。これらの2つの因子は、全ての非中心ピクセルがターゲットクラスターの中心に対して同じ程度の近接度を有するため、補足距離チャンネルを提供する必要性を除去する。

【0619】

図104は、標準的な畳み込みニューラルネットワーク及び再構成された入力を使用して、現在の配列決定サイクルで単一のターゲットクラスターをベースコールする一実施態様を示す。図示された実施態様では、再構成された入力は、ベースコールされている現在の(t)配列決定サイクルに対する現在の画像パッチセットと、以前の(t-1)配列決定サイクルに対する以前の画像パッチセットと、次の(t+1)配列決定サイクルに対する次の画像パッチセットと、を含む。各画像パッチセットは、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つに対する画像パッチを有する。図104は、2つの画像チャンネル、赤色チャンネル及び緑色チャンネルを示す。各画像パッチは、ベースコールされているターゲットクラスター、いくつかの隣接するクラスター、及びそれらの周囲の背景をカバーするピクセルに対するピクセル強度データを有する。再構成された入力はまた、共通のスケールリングチャンネルを含む。

【0620】

図103a～図103bに関して上で説明されるように、画像パッチが、中心でターゲットクラスターを中心にするように再構成又はシフトされるため、再構成された入力は、

10

20

30

40

50

任意の距離チャンネルを含まない。また、画像パッチは、サイクル間及びサイクル内残留位置合わせ誤差を除去するように互いに位置合わせされる。一実施態様では、これは、アフィン変換及び強度補間を使用して行われ、その追加の詳細は、付録 1、2、3、及び 4 で見出され得る。これらの因子は、専用構造を使用する必要性を除去し、その代わりに、標準的な畳み込みニューラルネットワークが、再構成された入力で使用される。

【0621】

図示された実施態様では、標準的な畳み込みニューラルネットワーク 10400 は、標準的な畳み込みフィルタを使用する 7 個の標準的な畳み込み層を含む。これは、（データが位置合わせされ、混合され得るため）配列決定サイクル間のデータの混合を防止するための分離された畳み込みパイプラインが存在しないことを意味する。いくつかの実施態様では、空間次元性事象の一貫した低減は、標準的な畳み込みフィルタを、中心クラスター中心及び他のピクセルよりも多くのその隣接するピクセルに対処するように教示するために使用される。

【0622】

次いで、再構成された入力は、標準的な畳み込み層を介して処理されて、最終畳み込み表現を生成する。最終畳み込み表現に基づいて、現在の配列決定サイクルでのターゲットクラスターに対するベースコールは、図 96c に関して上述したように、平坦化、高密度、及び分類層を使用して同様の様式で取得される。

【0623】

いくつかの実施態様では、プロセスは、複数の配列決定サイクルで反復されて、ターゲットクラスターに対するベースコールの配列を生成する。

【0624】

他の実施態様では、プロセスは、複数のターゲットクラスターに対する複数の配列決定サイクルで反復されて、複数のターゲットクラスター内の各ターゲットクラスターに対するベースコールの配列を生成する。

位置合わせされた入力：距離チャンネルなしの位置合わせされた画像パッチ及び再構成

【0625】

図 105 は、標準的な畳み込みニューラルネットワーク及び位置合わせされた入力を使用して、現在の配列決定サイクルで複数のターゲットクラスターをベースコールする一実施態様を示す。ここで、画像パッチが、ベースコールされている複数の中心クラスターピクセルを含むため、再構成は実現可能でない。その結果、位置合わせされた入力内の画像パッチは、再構成されない。更に、一実施態様によれば、補足距離チャンネルは、計算検討事項に起因して含まれない。

【0626】

次いで、位置合わせされた入力は、標準的な畳み込み層を介して処理されて、最終畳み込み表現を生成する。最終畳み込み表現に基づいて、ターゲットクラスターの各々に対するベースコールは、図 97 に関して上述したように、平坦化（任意選択）、高密度（任意選択）、分類、及びベースコールフィルタリング層を使用して、同様の様式で現在の配列決定サイクルで取得される。

【0627】

図 106 は、標準的な畳み込みニューラルネットワーク及び位置合わせされた入力を使用して、複数の配列決定サイクルで複数のターゲットクラスターをベースコールする一実施態様を示す。位置合わせされた入力は、標準的な畳み込み層を介して処理されて、ベースコールされている y 回の配列決定サイクルの各々に対して最終畳み込み表現を生成する。y 回の最終畳み込み表現に基づいて、ターゲットクラスターの各々に対するベースコールは、図 98 に関して上述したように、平坦化（任意選択）、高密度（任意選択）、分類、及びベースコールフィルタリング層を使用して、同様の様式でベースコールされている y 回の配列決定サイクルの各々に対して取得される。

【0628】

当業者は、他の実施態様では、標準的な畳み込みニューラルネットワークは、より少な

10

20

30

40

50

い又はより多くの数の配列決定サイクルに対する再構成された入力を処理し得、より少ない又はより多くの数の標準的な畳み込み層を含み得ることを理解するであろう。また、再構成された入力の次元性、再構成された入力内のサイクルごとのテンソル、畳み込みフィルタ、得られた特徴マップ、及び出力は異なり得る。また、畳み込み層内の畳み込みフィルタの数は異なり得る。それは、1D重畳、2D重畳、3D重畳、4D重畳、5D重畳、拡張又は無性重畳、転置重畳、奥行分離可能な重畳、 1×1 重畳、グループ重畳、扁平重畳、空間及びクロスチャネルの重畳、シャッフルされたグループ化重畳、空間的な分離可能な重畳、及び逆重畳を使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はGRUなど）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（ReLU）、漏れやすいReLU、指数関数的ライナーユニット（ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

（訓練）

【0629】

図107は、ニューラルネットワークベースのベースコーラー1514を訓練10700する一実施態様を示す。専用構造及び標準的な構造の両方で、ニューラルネットワークベースのベースコーラー1514は、正しいベースコール10708に対して、予測されたベースコール10704を比較し、比較に基づいて誤差10706を計算する、逆伝搬ベースの勾配更新技術を使用して訓練される。次いで、誤差10706が、後方伝搬10710中にニューラルネットワークベースのベースコーラー1514の重み及びパラメータに適用される勾配を計算するために使用される。訓練10700は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器1510によって操作可能である。

【0630】

訓練器1510は、（配列決定画像108から導出される）訓練データ10702を使用して、予測されたベースコール10704を生成する前方伝搬10712、及び誤差10706に基づいて重み及びパラメータを更新する後方伝搬10710の数千及び数百万回の反復でニューラルネットワークベースのベースコーラー1514を訓練する。訓練10700に関する追加の詳細は、「Deep Learning Tools」と題する付録で見出され得る。

CNN-RNNベースのベースコーラー

ハイブリッドニューラルネットワーク

【0631】

図108aは、ニューラルネットワークベースのベースコーラー1514として使用されるハイブリッドニューラルネットワーク10800aの一実施態様を示す。ハイブリッドニューラルネットワーク10800aは、少なくとも1つの畳み込みモジュール10804（又は畳み込みニューラルネットワーク（CNN））と、少なくとも1つの反復モジュール10808（又は反復ニューラルネットワーク（RNN））と、を含む。反復モジュール10808は、畳み込みモジュール10804から入力を使用及び／又は受信する。

【0632】

畳み込みモジュール10804は、1つ又はそれ以上の畳み込み層を介して入力データ10802を処理し、畳み込み出力10806を生成する。一実施態様では、入力データ10802は、「入力」と題するセクションで上述したように、主入力として画像チャネ

10

20

30

40

50

ル又は画像データのみを含む。ハイブリッドニューラルネットワーク10800aに供給される画像データは、上述の画像データ7902と同じであり得る。

【0633】

別の実施態様では、画像チャンネル又は画像データに加えて、入力データ10802はまた、「入力」と題するセクションで上述したように、距離チャンネル、スケーリングチャンネル、クラスター中心座標、及び／又はクラスター属性情報などの補足チャンネルを含む。

【0634】

画像データ（すなわち、入力データ10802）は、1つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を）示す。畳み込みモジュール10804は、畳み込み層を介して、配列決定動作の一連の配列決定サイクルに対する画像データを処理し、画像データの1つ又はそれ以上の畳み込み表現（すなわち、畳み込み出力10806）を生成する。

【0635】

一連の配列決定サイクルは、ベースコールされている t 回の配列決定サイクルに対する画像データを含み得、 t は、1～1000の任意の数である。 t が15～21であるとき、我々は、正確なベースコール結果を観察する。

【0636】

反復モジュール10810は、畳み込み出力10806を畳み込み、反復出力10810を生成する。特に、反復モジュール10810は、畳み込み表現及び以前の隠れ状態表現を畳み込むことに基づいて、現在の隠れ状態表現（すなわち、反復出力10810）を生成する。

【0637】

一実施態様では、反復モジュール10810は、3次元（3D）畳み込みを畳み込み表現及び以前の隠れ状態表現に適用し、以下のように数学的に定式化される、現在の隠れ状態表現を生成する。

$$h_t = W_{13D CONV} V_t + W_{23D CONV} h_{t-1}$$

であり、 h_t は、現在の時間ステップ t で生成される現在の隠れ状態表現を表し、

V_t は、現在の時間ステップ t で現在のスライディングウィンドウにおける入力容積を形成する、畳み込み表現のセット又はグループを表し、

$W_{13D CONV}$ は、 V_t に適用される第1の3D畳み込みフィルタの重みを表し、

h_{t-1} は、以前の時間ステップ $t-1$ で生成される以前の隠れ状態表現を表し、

$W_{23D CONV}$ は、 h_{t-1} に適用される第2の3D畳み込みフィルタの重みを表す。

【0638】

いくつかの実施態様では、 $W_{13D CONV}$ 及び $W_{23D CONV}$ は、重みが共有されるため同じである。

【0639】

次いで、出力モジュール10812は、反復出力10810に基づいて、ベースコール10814を生成する。いくつかの実施態様では、出力モジュール10812は、1つ又はそれ以上の完全に接続された層と、分類層（例えば、ソフトマックス）と、を含む。そのような実施態様では、現在の隠れ状態表現は、完全に接続された層を介して処理され、完全に接続された層の出力は、分類層を介して処理されて、ベースコール10814を生成する。

【0640】

ベースコール10814は、クラスターのうちの少なくとも1つ、及び配列決定サイクルのうちの少なくとも1つに対するベースコールを含む。いくつかの実施態様では、ベースコール10814は、クラスターの各々及び配列決定サイクルの各々に対するベースコールを含む。そのため、例えば、入力データ10802が25個のクラスター及び15回の配列決定サイクルに対する画像データを含むとき、ベースコール10802は、25個のクラスターの各々に対する15個のベースコールのベースコール配列を含む。

（3D畳み込み）

10

20

30

40

50

【0641】

図108bは、現在の隠れ状態表現を生成するためにハイブリッドニューラルネットワーク10800bの反復モジュール10810によって使用される3D畳み込み10800bの一実施態様を示す。

【0642】

3D畳み込みは、入力容積内に存在する各ボクセルが、畳み込みカーネルの同等の位置でのボクセルによって乗算される数学的動作である。最後に、結果の合計は、出力容積に追加される。図108bでは、3D畳み込み動作の表現を観察することが可能であり、入力10816で強調されるボクセル10816aは、カーネル10818内のそれらのそれぞれのボクセルで乗算される。これらの計算の後、それらの合計10820aは、出力10820に追加される。

10

【0643】

入力容積の座標が(x、y、z)によって与えられ、畳み込みカーネルが、サイズ(P、Q、R)を有するため、3D畳み込み動作は、以下のように数学的に定義され得る。

【数17】

$$O_{xyz} = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} K_{pqr} I_{(x+p)(y+q)(z+r)},$$

20

Oは、畳み込みの結果であり、

Iは、入力容積であり、

Kは、畳み込みカーネルであり、

(p、q、r)は、Kの座標である。

【0644】

バイアス項は、明確さを改善するために、上記の式から省略される。

【0645】

2D畳み込みのようなマトリックスから空間情報を抽出することに加えて、3D畳み込みは、連続するマトリックス間に存在する情報を抽出する。これにより、それらが3Dオブジェクトの空間情報及び配列決定画像のセットの時間的情報の両方をマッピングすることが可能になる。

30

(畳み込みモジュール)

【0646】

図109は、畳み込みモジュール10804の畳み込み層10900のカスケードを介して、ベースコールされる一連のt回の配列決定サイクル間の単一の配列決定サイクルのサイクルごとの入力データ10902を処理する一実施態様を示す。

【0647】

畳み込みモジュール10804は、畳み込み層10900のカスケードを介して、サイクルごとの入力データの配列内のサイクルごとの各入力データを別々に処理する。サイクルごとの入力データの配列は、ベースコールされている配列決定動作の一連の配列決定サイクルtに対して生成され、tは、1～1000の任意の数である。そのため、例えば、シリーズが15回の配列決定サイクルを含むとき、サイクルごとの入力データの配列は、15個の異なるサイクルごとの入力データを含む。

40

【0648】

一実施態様では、サイクルごとの各入力データは、画像チャネル(例えば、赤色チャネル及び緑色チャネル)又は画像データ(例えば、上述の画像データ7902)のみを含む。画像チャネル又は画像データは、一連のそれぞれの配列決定サイクルで捕捉される、1つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示す。別の実施態様では、画像チャネル又は画像データに加えて、サイクルごとの各入力データはまた、距離チ

50

チャンネル及びスケーリングチャンネル（例えば、上述の入力データ 9500）などの補足チャンネルを含む。

【0649】

図示された実施態様では、サイクルごとの入力データ 10902 は、ベースコールされる一連の t 回の配列決定サイクル間の単一の配列決定サイクルに対して、2つの画像チャンネル、すなわち、赤色チャンネル及び緑色チャンネルを含む。各画像チャンネルは、サイズ 15×15 の画像パッチで符号化される。畳み込みモジュール 10804 は、5 個の畳み込み層を含む。各畳み込み層は、サイズ 3×3 の 25 個の畳み込みフィルタのバンクを有する。更に、畳み込みフィルタは、入力画像又はテンソルの高さ及び幅を保存する、いわゆる同じパディングを使用する。同じパディングで、出力特徴マップが入力特徴と同じサイズを有するように、入力特徴にパディングが追加される。対照的に、いわゆる有効なパディングは、パディングがないことを意味する。

10

【0650】

第 1 の畳み込み層 10904 は、サイクルごとの入力データ 10902 を処理し、サイズ $15 \times 15 \times 25$ の第 1 の畳み込み表現 10906 を生成する。第 2 の畳み込み層 10908 は、第 1 の畳み込み表現 10906 を処理し、サイズ $15 \times 15 \times 25$ の第 2 の畳み込み表現 10910 を生成する。第 3 の畳み込み層 10912 は、第 2 の畳み込み表現 10910 を処理し、サイズ $15 \times 15 \times 25$ の第 3 の畳み込み表現 10914 を生成する。第 4 の畳み込み層 10916 は、第 3 の畳み込み表現 10914 を処理し、サイズ $15 \times 15 \times 25$ の第 4 の畳み込み表現 10918 を生成する。第 5 の畳み込み層 10920 は、第 4 の畳み込み表現 10918 を処理し、サイズ $15 \times 15 \times 25$ の第 5 の畳み込み表現 10922 を生成する。同じパディングは、得られた畳み込み表現（例えば、 15×15 ）の空間次元を保存することに留意されたい。いくつかの実施態様では、畳み込み層内の畳み込みフィルタの数は、2、4、16、32、64、128、256、512、及び 1024 などの 2 の累乗である。

20

【0651】

畳み込みがより深くなるにつれて、情報が失われ得る。これを考慮するために、いくつかの実施態様では、我々は、（1）最初のサイクルごとの入力データを再導入するために、及び（2）後の畳み込み層によって抽出される高レベルの空間特徴と、先の畳み込み層によって抽出される低レベルの空間特徴を組み合わせるために、スキップ接続を使用する。我々は、これがベースコール精度を改善することを観察している。

30

【0652】

図 110 は、畳み込みモジュール 10804 の畳み込み層 10900 のカスケードによって生成される、その対応する畳み込み表現 10906、10910、10914、10918、及び 10922 と、単一の配列決定サイクルのサイクルごとの入力データ 10902 を混合 11000 する一実施態様を示す。畳み込み表現 10906、10910、10914、10918、及び 10922 は、畳み込み表現 11004 の配列を形成するように連結されて、それは、次に、サイクルごとの入力データ 10902 と連結されて、混合表現 11006 を生成する。他の実施態様では、連結の代わりに積算が使用される。また、混合 11000 は、ミキサー 11002 によって動作可能である。

40

【0653】

次いで、平坦化器 11008 は、混合表現 11006 を平坦化し、サイクルごとの平坦化された混合表現 11010 を生成する。いくつかの実施態様では、平坦化された混合表現 11010 は、サイクルごとの入力データ 10902、並びに畳み込み表現 10906、10910、10914、10918、及び 10922（例えば、 15×1905 、すなわち、同じ行ごとの次元）と少なくとも 1 次元サイズを共有する、高次元ベクトル又は 2 次元（2D）アレイである。これは、下流 3D 畳み込みでの特徴抽出を容易にするデータでの対称性を誘発する。

【0654】

図 109 及び図 110 は、ベースコールされる一連の t 回の配列決定サイクル間の単一

50

の配列決定サイクルに対するサイクルごとの画像データ10902の処理を示す。畳み込みモジュール10804は、 t 回の配列決定サイクルの各々に対してそれぞれのサイクルごとの画像データを別々に処理し、 t 回の配列決定サイクルの各々に対して、それぞれのサイクルごとの平坦化された混合提示を生成する。

(スタック)

【0655】

図111は、連続する配列決定サイクルの平坦化された混合表現をスタック11100として配置する一実施態様を示す。図示された実施態様では、15回の配列決定サイクルに対する15個の平坦化された混合表現10904a~10904oは、スタック11100内にスタックされる。スタック11100は、3D畳み込みフィルタの同じ受容フィールドで、空間及び時間的次元（すなわち、複数の配列決定サイクル）の両方から利用可能な特徴を形成する3D入力容積である。スタックは、スタッカー11102によって動作可能である。他の実施態様では、スタック11100は、任意の次元性（例えば、1D、2D、4D、5Dなど）のテンソルであり得る。

(反復モジュール)

【0656】

我々は、配列決定データでの長期依存性を捕捉するために、特に、プレフェイジング及びフェイジングからのクロスサイクル配列決定画像での2次の寄与を考慮するために、反復処理を使用する。反復処理は、時間ステップの使用のために、連続データの分析に使用される。現在の時間ステップでの現在の隠れ状態表現は、(i)以前の時間ステップからの以前の隠れ状態表現、及び(ii)現在の時間ステップでの現在の入力の関数である。

【0657】

反復モジュール10808は、スタック11100に、前方及び後方の方向に3D畳み込みの反復適用（すなわち、反復処理11200）を行い、一連の t 回の配列決定サイクルの各々でクラスターの各々に対するベースコールを生成する。スライディングウィンドウベースでスタック11100内の平坦化された混合表現のサブセットから空間-時間的特徴を抽出するために、3D畳み込みが使用される。各スライディングウィンドウ(w)は、それぞれの配列決定サイクルに対応し、図112aでオレンジ色で強調される。いくつかの実施態様では、wは、同時にベースコールされている配列決定サイクルの総数に応じて、1、2、3、5、7、9、15、21などにパラメータ化される。一実施態様では、wは、同時にベースコールされている配列決定サイクルの総数の分数である。

【0658】

そのため、例えば、各スライディングウィンドウが、15個の平坦化された混合表現10904a~10904oを含むスタック11100からの3つの連続する平坦化された混合表現を含むと考える。次いで、第1のスライディングウィンドウ内の最初の3個の平坦化された混合表現10904a~10904cは、第1の配列決定サイクルに対応し、第2のスライディングウィンドウ内の次の3つの平坦化された混合表現10904b~10904dは、第2の配列決定サイクルに対応するなどである。いくつかの実施態様では、最終の平坦化された混合表現10904oで開始して、最終の配列決定サイクルに対応する最終のスライディングウィンドウ内の適切な数の平坦化された混合表現を符号化するために、パディングが使用される。

【0659】

各時間ステップで、反復モジュール10808は、(1)現在の入力 $x(t)$ 及び(2)以前の隠れ状態表現 $h(t-1)$ を受け入れ、現在の隠れ状態表現 $h(t)$ を計算する。現在の入力 $x(t)$ は、現在のスライディングウィンドウ(w)、オレンジ色)内にあるスタック11100からの平坦化された混合表現のサブセットのみを含む。したがって、各時間ステップで、現在の各入力 $x(t)$ は、複数の平坦化された混合表現（例えば、wに応じた1、2、3、5、7、9、15、又は21個の平坦化された混合表現）の3D容積である。例えば、(i)単一の平坦化された混合表現が大きさ 15×1905 を有する2次元(2D)であり、(ii)wが7であるとき、各時間ステップで、現在の各入

力 $x(t)$ は、大きさ $15 \times 1905 \times 7$ を有する 3D 容積である。

【0660】

反復モジュール 10808 は、第 1 の 3D 畳み込み ($W_{13DCONV}$) を現在の入力 $x(t)$ に、第 2 の 3D 畳み込み ($W_{23DCONV}$) を以前の隠れ状態表現 $h(t-1)$ に適用して、現在の隠れ状態表現 $h(t)$ を生成する。いくつかの実施態様では、 $W_{13DCONV}$ 及び $W_{23DCONV}$ は、重みが共有されるため同じである。

(ゲート処理)

【0661】

一実施態様では、反復モジュール 10808 は、長い短期メモリ (LSTM) ネットワーク又はゲートされた反復単位 (GRU) ネットワークなどのゲートされたネットワークを介して、現在の入力 $x(t)$ 及び以前の隠れ状態表現 $h(t-1)$ を処理する。例えば、LSTM 実施態様では、現在の入力 $x(t)$ は、以前の隠れ状態表現 $h(t-1)$ と共に、LSTM ユニットの 4 つのゲート、すなわち、入力ゲート、活性化ゲート、忘却ゲート、及び出力ゲートの各々を介して処理される。これは、3D 畳み込みを現在の入力 $x(t)$ 及び以前の隠れ状態表現 $h(t-1)$ に適用して、現在の隠れ状態表現 $h(t)$ を出力として生成する LSTM ユニットの介して、現在の入力 $x(t)$ 及び以前の隠れ状態表現 $h(t-1)$ を処理 11200b する、一実施態様を示す、図 112b に示される。そのような実施態様では、入力、活性化、忘却、及び出力ゲートの重みは、3D 畳み込みを適用する。

【0662】

いくつかの実施態様では、ゲートされたユニット (LSTM 又は GRU) は、双曲線正接及びシグモイドのような非線形／スカッシング関数を使用しない。

【0663】

一実施態様では、現在の入力 $x(t)$ 、以前の隠れ状態表現 $h(t-1)$ 、及び現在の隠れ状態表現 $h(t)$ は全て、同じ次元性を有する 3D 容積であり、3D 容積として入力、活性化、忘却、及び出力ゲートを介して処理されるか、又はそれによって生成される。

【0664】

一実施態様では、反復モジュール 10808 の 3D 畳み込みは、同じパディングと共に、サイズ 3×3 の 25 個の畳み込みフィルタのバンクを使用する。いくつかの実施態様では、畳み込みフィルタのサイズは、 5×5 である。いくつかの実施態様では、反復モジュール 10808 によって使用される畳み込みフィルタの数は、2、4、16、32、64、128、256、512、及び 1024 などの 2 の累乗によって因数分解される。

(双方向処理)

【0665】

反復モジュール 10808 は、最初に、スライディングウィンドウベースで始めから終わりまで (上から下に) スタック 11100 を処理し、前方横断

【数 18】

$$\vec{h}_t = 3DCONV(x_t + \vec{h}_{t-1})$$

に対する現在の隠れ状態表現 (ベクトル) の配列を生成する。

【0666】

次いで、反復モジュール 10808 は、スライディングウィンドウベースで終わりから始めまで (下から上に) スタック 11100 を処理し、後方／逆の横断

【数 19】

$$\overleftarrow{h}_t = 3D CONV(x_t + \overleftarrow{h}_{t-1})$$

に対する現在の隠れ状態表現（ベクトル）の配列を生成する。

【 0 6 6 7 】

いくつかの実施態様では、両方の方向に対して各時間ステップで、処理は、L S T M又はG R Uのゲートを使用する。例えば、各時間ステップで、前方の現在の入力 $x(t)$ は、L S T Mユニットの入力、活性化、忘却、及び出力ゲートを介して処理されて、前方の現在の隠れ状態表現

10

【数 2 0】

$$\overrightarrow{h}_t$$

を生成し、後方の現在の入力 $x(t)$ は、別のL S T Mユニットの入力、活性化、忘却、及び出力ゲートを介して処理されて、後方の現在の隠れ状態表現

20

【数 2 1】

$$\overleftarrow{h}_t$$

を生成する。

【 0 6 6 8 】

次いで、各時間ステップ／スライディングウィンドウ／配列決定サイクルに対して、反復モジュール 1 0 8 0 8 は、対応する前方及び後方の現在の隠れ状態表現を組み合わせ（連結若しくは合計又は平均し）、組み合わせられた隠れ状態表現

30

【数 2 2】

$$\overset{c}{h}_t = \begin{bmatrix} \overrightarrow{h}_t; \overleftarrow{h}_t \end{bmatrix}$$

を生成する。

【 0 6 6 9 】

次いで、組み合わせられた隠れ表現

40

【数 2 3】

$$\overset{c}{h}_t$$

50

は、1つ又はそれ以上の完全に接続されたネットワークを介して処理されて、高密度表現を生成する。次いで、高密度表現は、ソフトマックス層を介して処理されて、所与の配列決定サイクルでクラスターの各々に組み込まれる塩基がA、C、T、及びGである尤度を生成する。塩基は、尤度に基づいて、A、C、T、又はGとして分類される。これは、並行又は連続して、一連のt回の配列決定サイクルの各々（又は各時間ステップ／スライディングウィンドウ）に対して行われる。

【0670】

当業者は、他の実施態様では、ハイブリッド構造が、より少ない又はより多くの数の配列決定サイクルに対する入力データを処理し得、より少ない又はより多くの数の畳み込み及び反復層を含み得ることを理解するであろう。また、入力データの次元性、現在及び以前の隠れ状態表現、畳み込みフィルタ、得られた特徴マップ、並びに出力は異なり得る。また、畳み込み層内の畳み込みフィルタの数は異なり得る。それは、異なるパディング及びストリディング構成を使用し得る。それは、異なる分類機能（例えば、シグモイド又は回帰）を使用し得、完全に接続された層を含んでも含まなくてもよい。それは、1D重畳、2D重畳、3D重畳、4D重畳、5D重畳、拡張又は無性重畳、転置重畳、奥行分離可能な重畳、 1×1 重畳、グループ重畳、扁平重畳、空間及びクロスチャネルの重畳、シャッフルされたグループ化重畳、空間的な分離可能な重畳、及び逆重畳を使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はGRUなど）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（ReLU）、漏れやすいReLU、指数関数的ライナーユニット（ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

（実験結果及び考察）

【0671】

図113は、ニューラルネットワークベースのベースコーラー1514を訓練するために使用される訓練データでのトリヌクレオチド（3量体）をバランスさせる一実施態様を示す。バランスは、訓練データ内のゲノムに関する統計の非常に少ない学習をもたらし、次に、一般化を改善する。熱マップ11302は、「A. baumannii」と呼ばれる第1の生物に対する訓練データでのバランスされた3量体を示す。熱マップ11304は、「E. coli」と呼ばれる第2の生物に対する訓練データでのバランスされた3量体を示す。

【0672】

図114は、ニューラルネットワークベースのベースコーラー1514に対して、RTAベースコーラーのベースコール精度を比較する。図114に図示されているように、RTAベースコーラーは、2回の配列決定動作（リード1及びリード2）でより高い誤差割合を有する。すなわち、ニューラルネットワークベースのベースコーラー1514は、両方の配列決定動作で、RTAベースコーラーを実行する。

【0673】

図115は、RTAベースコーラーのタイル間の一般化を、同じタイル上のニューラルネットワークベースのベースコーラー1514のものと比較する。すなわち、ニューラルネットワークベースのベースコーラー1514で、データが訓練で使用される同じタイルに対するデータに対して推測（試験）が実行される。

【0674】

図116は、RTAベースコーラーのタイル間の一般化を、同じタイル上及び異なるタイル上のニューラルネットワークベースのベースコーラー1514のものと比較する。すなわち、ニューラルネットワークベースのベースコーラー1514は、第1のタイル上のクラスターに対するデータに対して訓練されるが、第2のタイル上のクラスターからのデータに対して推測を実行する。同じタイル実施態様では、ニューラルネットワークベースのベースコーラー1514は、タイル5上のクラスターからのデータに対して訓練され、タイル5上のクラスターからのデータに対して試験される。異なるタイル実施態様では、ニューラルネットワークベースのベースコーラー1514は、タイル10上のクラスターからのデータに対して訓練され、タイル5上のクラスターからのデータに対して試験される。

10

【0675】

図117はまた、RTAベースコーラーのタイル間の一般化を異なるタイル上のニューラルネットワークベースのベースコーラー1514のものと比較する。異なるタイル実施態様では、ニューラルネットワークベースのベースコーラー1514は、タイル10上のクラスターからのデータに対して訓練され、タイル5上のクラスターからのデータに対して試験されると、タイル20上のクラスターからのデータに対して訓練され、タイル5上のクラスターからのデータに対して試験される。

【0676】

図118は、ニューラルネットワークベースのベースコーラー1514への入力として供給される画像パッチの異なるサイズが、どのようにベースコール精度をもたらすかを示す。両方の配列決定動作（リード1及びリード2）では、パッチサイズが $3 \times 3 \sim 11 \times 11$ に増加するにつれて、誤差割合は減少する。すなわち、ニューラルネットワークベースのベースコーラー1514は、より大きい画像パッチを有するより正確なベースコールを生成する。いくつかの実施態様では、ベースコール精度は、 100×100 のピクセル以下である画像パッチを使用することによって、計算効率に対してバランスされる。他の実施態様では、 3000×3000 のピクセル（及びより大きいピクセル）の大きさの画像パッチが使用される。

20

【0677】

図119、図120、図121、及び図122は、A. baumannii及びE. coliからの訓練データにおけるニューラルネットワークベースのベースコーラー1514のレーン間の一般化を示す。

30

【0678】

図120に戻って、一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、フローセルの第1及び第2のレーンの両方のクラスターからのA. baumanniiデータに対して試験される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1及び第2のレーンの両方のクラスターからのA. baumanniiデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第1及び第2のレーンの両方のクラスターからのA. baumanniiデータに対して試験される。更に異なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1及び第2のレーンの両方のクラスターからのA. baumanniiデータに対して試験される。

40

【0679】

一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、フローセルの第1及び第2のレーンの両方のクラスターからのE. coliデータに対して試験

50

される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1及び第2のレーンの両方のクラスターからのE. coliデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのE. coliデータに対して試験される。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1及び第2のレーンの両方のクラスターからのE. coliデータに対して試験される。

【0680】

図120では、（誤差割合によって測定される）ベースコール精度は、2回の配列決定動作（例えば、リード1及びリード2）に対するこれらの実施態様の各々に対して示される。

【0681】

図121に戻って、一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのA. baumanniiデータに対して試験される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1のレーン上のクラスターからのA. baumanniiデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのA. baumanniiデータに対して試験される。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1のレーン上のクラスターからのA. baumanniiデータに対して試験される。

【0682】

一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのE. coliデータに対して試験される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1のレーン上のクラスターからのE. coliデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのE. coliデータに対して試験される。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第1のレーン上のクラスターからのE. coliデータに対して試験される。

【0683】

図121では、（誤差割合によって測定される）ベースコール精度は、2回の配列決定動作（例えば、リード1及びリード2）に対するこれらの実施態様の各々に対して示される。図120を図121と比較して、後のものによってカバーされる実施態様は、50～80パーセントの誤差低減をもたらすことが見られ得る。

【0684】

図122に戻って、一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、第2のレーン上のクラスターからのA. baumanniiデータに対して試験される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、

第2のレーン上のクラスターからのA. baumanniiデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第1のレーン上のクラスターからのA. baumanniiデータに対して試験される。第2の第1のレーンにおいて。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第2のレーン上のクラスターからのA. baumanniiデータに対して試験される。

【0685】

一実施態様では、ニューラルネットワークベースのベースコーラー1514は、フローセルの第1のレーン上のクラスターからのE. coliデータに対して訓練され、第2のレーン上のクラスターからのE. coliデータに対して試験される。別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第1のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第2のレーン上のクラスターからのE. coliデータに対して試験される。更に別の実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのE. coliデータに対して訓練され、第2のレーン上のクラスターからのE. coliデータに対して試験される。更に更なる実施態様では、ニューラルネットワークベースのベースコーラー1514は、第2のレーン上のクラスターからのA. baumanniiデータに対して訓練され、第2のレーン上のクラスターからのE. coliデータに対して試験される。

【0686】

図122では、(誤差割合によって測定される)ベースコール精度は、2回の配列決定動作(例えば、リード1及びリード2)に対するこれらの実施態様の各々に対して示される。図120を図122と比較して、後のものによってカバーされる実施態様は、50〜80パーセントの誤差低減をもたらすことが見られ得る。

【0687】

図123は、図119、図120、図121、及び図122に関して上述したレーン間の一般化のための誤差プロファイルを示す。一実施態様では、誤差プロファイルは、緑色チャンネル内のA及びT塩基をベースコールする際の誤差を検出する。

【0688】

図124は、図123の誤差プロファイルによって検出される誤差のソースを、緑色チャンネル内の低クラスター強度に属させる。

【0689】

図125は、2回の配列決定動作(リード1及びリード2)について、RTAベースコーラー及びニューラルネットワークベースのベースコーラー1514の誤差プロファイルと比較する。比較は、ニューラルネットワークベースのベースコーラー1514の優れたベースコール精度を確認する。

【0690】

図126aは、4つの異なる器具におけるニューラルネットワークベースのベースコーラー1514の動作間の一般化を示す。

【0691】

図126bは、同じ器具上で実行される4つの異なる動作におけるニューラルネットワークベースのベースコーラー1514の動作間の一般化を示す。

【0692】

図127は、ニューラルネットワークベースのベースコーラー1514を訓練するために使用される訓練データのゲノム統計を示す。

【0693】

図128は、ニューラルネットワークベースのベースコーラー1514を訓練するために使用される訓練データのゲノムコンテキストを示す。

10

20

30

40

50

【0694】

図129は、長いリード（例えば、 2×250 ）をベースコールする際のニューラルネットワークベースのベースコーラー1514のベースコール精度を示す。

【0695】

図130は、ニューラルネットワークベースのベースコーラー1514が、画像パッチにわたる中心クラスターピクセル（複数可）及びその隣接するピクセルにどのように対処するかの一実施態様を示す。

【0696】

図131は、一実施態様に係る、ニューラルネットワークベースのベースコーラー1514を訓練及び動作するために使用される様々なハードウェア構成要素及び構成を示す。他の実施態様では、異なるハードウェア構成要素及び構成が使用される。

10

【0697】

図132は、ニューラルネットワークベースのベースコーラー1514を使用して実行され得る様々な配列決定タスクを示す。いくつかの例は、品質スコアリング（Qスコアリング）と、変異体分類と、を含む。図132はまた、ニューラルネットワークベースのベースコーラー1514がベースコールを実行する、いくつかの例示的配列決定器具をリストする。

【0698】

図133は、t分布型確率的近傍埋め込み（t-SNE）によって可視化される散乱プロット13300であり、ニューラルネットワークベースのベースコーラー1514のベースコール結果を示す。散乱プロット13300は、ベースコール結果が $64(4^3)$ 個のグループにクラスタリングされ、各グループが、主として、特定の入力3量体（トリヌクレオチド繰り返しパターン）に対応することを示す。これは、ニューラルネットワークベースのベースコーラー1514が、少なくとも3つの配列決定サイクルに対する入力データを処理し、配列固有のモチーフを学習して、以前の連続するベースコールに基づいて現在のベースコールを生成するための場合である。

20

（品質スコアリング）

【0699】

品質スコアリングは、品質スコアを各ベースコールに割り当てるプロセスを指す。品質スコアは、品質表に基づいて、配列決定トレースの予測特徴の値を確率に変換するPhredフレームワークに従って定義される。品質表は、校正データセットに対して訓練することによって取得され、配列決定プラットフォームの特性が変化するとき更新される。品質スコアの確率的な解釈は、変異体コール及び配列アセンブリなどの下流分析での異なる配列決定リードの公平な統合を可能にする。したがって、品質スコアを定義するための有効なモデルは、任意のベースコーラーにとって不可欠である。

30

【0700】

我々は、品質スコアが何かを最初に説明する。品質スコアは、ベースコールでの配列決定誤差の確率の尺度である。高品質スコアは、ベースコールがより信頼性が高く、不正確である可能性がより低いことを意味する。例えば、塩基の品質スコアがQ30である場合、この塩基が不正確にコールされる確率は0.001である。これはまた、ベースコール精度が99.9%であることを示す。

40

【0701】

以下の表は、ベースコール品質スコア及びそれらの対応する誤差確率、ベースコール精度率、並びにベースコール誤差率間の関係を示す。

【表 1】

品質スコア	誤差確率	ベースコール誤差率	ベースコール精度率
Q10	0.1 (10個中1個)	10%	90%
Q20	0.01 (100個中1個)	1%	99%
Q30	0.001 (1,000個中1個)	0.1%	99.9%
Q40	0.0001 (10,000個中1個)	0.01%	99.99%
Q50	0.00001 (100,000個中1個)	0.001%	99.999%
Q60	0.000001 (1,000,000個中1個)	0.0001%	99.9999%

10

【0702】

ここで、我々は、品質スコアがどのように生成されるかを説明する。配列決定動作中、品質スコアは、全ての配列決定サイクルに対して、全てのタイルで、全てのクラスターに対する各ベースコールに割り当てられる。Illumina品質スコアは、2ステッププロセスで各ベースコールに対して計算される。各ベースコールに対して、品質予測値の数が計算される。品質予測値は、ベースコールが抽出されるクラスターの観察可能な特性である。これらは、強度プロファイル及び信号対ノイズ比などの特性を含み、ベースコール信頼性の様々な態様を測定する。それらは、ベースコールの品質と相関するように経験的に決定されている。

20

【0703】

品質表又はQ表としても知られる、品質モデルは、品質予測値の組み合わせをリストし、対応する品質スコアにそれらに関連付ける。この関係は、経験的データを使用して較正プロセスによって決定される。新しい品質スコアを推定するために、品質予測値は、新しいベースコールに対して計算され、事前較正された品質表内の値と比較される。

【0704】

ここで、我々は、品質表がどのように較正されるかを説明する。較正は、統計的品質表が、多数の器具上で配列決定される、様々な十分に特徴付けられたヒト及び非ヒトサンプルを含む経験的データから導出される、プロセスである。Phredアルゴリズムの修正バージョンを使用して、リードを適切な参照に位置合わせすることによって決定される生信号及び誤差率の特性を使用して品質表が開発及び改良される。

30

【0705】

ここで、我々は、品質表が時々変化する理由を説明する。品質表は、特定の器具構成及び化学のバージョンによって生成される動作に対して品質スコアを提供する。新しいハードウェア、ソフトウェア、又は化学バージョンなどの配列決定プラットフォームの有意な特性が変化するとき、品質モデルは、再較正を必要とする。例えば、配列決定化学の改善は、実質的な量の処理時間及び計算リソースを消費する新しいデータを正確にスコアリングするように品質表再較正を必要とする。

(ニューラルネットワークベースの品質スコアリング)

【0706】

我々は、品質予測値又は品質表を使用せず、その代わりに、十分に較正されたニューラルネットワークの予測における信頼性から品質スコアを推測する、品質スコアリングのためのニューラルネットワークベースの技術を開示する。ニューラルネットワークのコンテキストでは、「較正」は、主観的な予測と経験的な長い動作周波数との間の一貫性又は相関を指す。これは、確実性の頻度論者の概念であり、時間の90%で特定のラベルが正しいラベルであることをニューラルネットワークが主張する場合、評価中、確率90%が正しいものであると与えられる全てのラベルの90%が、正しいラベルであるべきである。較正は、精度に対する直交する懸念事項であることに留意されたい。ニューラルネットワークの予測は、正確であり得るが、誤って較正され得る。

40

【0707】

50

開示されたニューラルネットワークは、現実世界の配列決定動作のベースコールドメインを適切にモデル化する多様な配列決定特性を有する大規模な訓練セットで訓練されるため、十分に較正される。特に、様々な配列決定プラットフォーム、配列決定器具、配列決定プロトコル、配列決定化学、配列決定試薬、クラスター密度、及びフローセルから取得される配列決定画像は、ニューラルネットワークを訓練するための訓練例として使用される。他の実施態様では、異なるベースコール及び品質スコアリングモデルはそれぞれ、異なる配列決定プラットフォーム、配列決定器具、配列決定プロトコル、配列決定化学、配列決定試薬、クラスター密度、及び／又はフローセルのために使用される。

【0708】

4つのベースコールクラス（A、C、T、及びG）の各々に対して、多数の配列決定画像は、広範囲の配列決定条件の下で、それぞれのベースコールクラスを表す強度パターンを識別する訓練例として使用される。これは、次に、ニューラルネットワークの分類能力を訓練中に存在しない新しいクラスに拡張する必要性を除去する。更に、各訓練例は、適切な参照に対するリードの位置合わせに基づいて、対応するグラウンドトゥールズで正確にラベル付けされる。その結果は、予測における信頼性が、以下で数学的に表現される品質スコアリングに対する確実性尺度として解釈され得る、十分に較正されたニューラルネットワークである。

【0709】

$Y = \{A, C, T, G\}$ が、ベースコールクラスA、C、T、及びGに対するクラスラベルのセットを示し、 X が入力の空間を示すと仮定する。 $N_{\theta}(y|x)$ が、開示されたニューラルネットワークのうちの1つが入 $x \in X$ 力で予測する確率分布を示し、 θ が、ニューラルネットワークのパラメータを示すと仮定する。正しいラベル y_i を有する訓練例 x_i に対して、ニューラルネットワークは、ラベル

【数24】

$$\hat{y}_i = \operatorname{argmax}_{y \in Y} N_{\theta}(y|x_i)$$

を予測する。予測は、

【数25】

$$\hat{y}_i = y_i$$

及び0の場合、正確性スコア $c_i = 1$ を得て、そうでない場合、信頼性スコア

【数26】

$$r_i = N_{\theta}(\hat{y}_i|x_i)$$

を得る。

【0710】

ニューラルネットワーク $N_{\theta}(y|x)$ は、全体的に、 $(x_i, y_i) \in D$ 及び $r_i = \alpha$ で、 $c_i = 1$ は α である確率のために、データ分布 D において十分に較正される。例えば

、100の予測を仮定すると、Dからのサンプルから、信頼性0.8、80を有する各々は、ニューラルネットワーク $N_{\theta}(y|x)$ によって正確に分類される。より正式には、 $P_{\theta,D}(r,c)$ は、Dにおけるニューラルネットワーク $N_{\theta}(y|x)$ の予測の r 及び c 値における分布を表し、

【数27】

$$P_{\theta,D}(c=1|r=I_{\alpha})=\alpha \quad \forall_{\alpha} \in [0,1]$$

10

と表現され、 I_{α} は、 α 周囲の小さい非ゼロ間隔を表す。

【0711】

十分に較正されたニューラルネットワークが、品質予測値又は品質表とは異なり、多様な訓練セットで訓練されるため、それらは、器具構成及び化学バージョンに固有でない。これは、2つの利点を有する。最初に、異なるタイプの配列決定器具に対して、十分に較正されたニューラルネットワークは、別個の較正プロセスから異なる品質表を導出する必要性を除去する。次に、同じ配列決定器具に対して、それらは、配列決定器具の特性が変化するとき、再較正の必要性を除去する。以下の詳細がある。

(ソフトマックス信頼性確率からの品質スコアの推測)

20

【0712】

第1の十分に較正されたニューラルネットワークは、配列決定画像108から導出される入力データを処理し、塩基がA、C、T、及びGであるベースコール信頼性確率を生成する、ニューラルネットワークベースのベースコーラー1514である。ベースコール信頼性確率はまた、尤度又は分類スコアと見なされ得る。一実施態様では、ニューラルネットワークベースのベースコーラー1514は、ソフトマックス関数を使用して、ソフトマックススコアとしてベースコール信頼性確率を生成する。

【0713】

ソフトマックススコアが較正され(すなわち、それらがグラウンドトゥールズ正確性尤度を表し)、したがって、品質スコアに当然対応するため、品質スコアは、ニューラルネットワークベースのベースコーラー1514のソフトマックス関数によって生成されるベースコール信頼性確率から推測される。

30

【0714】

我々は、訓練中にニューラルネットワークベースのベースコーラー1514によって生成されるベースコール信頼性確率のセットを選択し、それらのベースコール誤差率(又はベースコール精度率)を決定することによって、ベースコール信頼性確率と品質スコアとの間の対応を実証する。

【0715】

そのため、例えば、我々は、ニューラルネットワークベースのベースコーラー1514によって生成されるベースコール信頼性確率「0.90」を選択する。我々は、ニューラルネットワークベースのベースコーラー1514が0.90のソフトマックススコアでベースコール予測を行ったとき、多数の(例えば、10000~10000000の範囲の)インスタンスを取る。多数のインスタンスは、検証セット又は試験セットのいずれかから取得され得る。次いで、我々は、多数のインスタンスのうちのそれぞれの1つに関連付けられる、対応するグラウンドトゥールズベースコールとの比較に基づいて、多数のインスタンスのうちのどれだけが、ベースコール予測が正しかったかを決定する。

40

【0716】

我々は、ベースコールが、10パーセントのミスコールで、多数のインスタンスの90パーセントで正確に予測されたことを観察している。これは、0.90のソフトマックススコアに対して、ベースコール誤差率が10%であり、ベースコール精度率が90%であ

50

り、それは次に、品質スコアQ10に対応する（上記表を参照）ことを意味する。同様に、0.99、0.999、0.9999、0.99999、及び0.999999のような他のソフトマックススコアに対して、我々は、それぞれ、品質スコアQ20、Q30、Q40、Q50、及びQ60との対応を観察している。これは、図136aに示される。他の実施態様では、我々は、Q9、Q11、Q12、Q23、Q25、Q29、Q37、及びQ39などのソフトマックススコアと品質スコアとの間の対応を観察している。

【0717】

我々はまた、ビニング品質スコアとの対応を観察している。例えば、0.80のソフトマックススコアは、ビニング品質スコアQ06に対応し、0.95のソフトマックススコアは、ビニング品質スコアQ15に対応し、0.993のソフトマックススコアは、ビニング品質スコアQ22に対応し、0.997のソフトマックススコアは、ビニング品質スコアQ27に対応し、0.9991のソフトマックススコアは、ビニング品質スコアQ33に対応し、0.9995のソフトマックススコアは、ビニング品質スコアQ37に対応し、0.9999のソフトマックススコアは、ビニング品質スコアQ40に対応する。これは、図136bに示される。

【0718】

本明細書で使用されるサンプルサイズは、小さいサンプル問題を回避するように大きく、例えば、10000～1000000の範囲であり得る。いくつかの実施態様では、ベースコール誤差率（又はベースコール精度率）を決定するために使用されるインスタンスのサンプルサイズは、評価されているソフトマックススコアに基づいて選択される。例えば、0.99のソフトマックススコアに対して、サンプルは、100個のインスタンスを含み、0.999のソフトマックススコアに対して、サンプルは、1000個のインスタンスを含み、0.9999のソフトマックススコアに対して、サンプルは、1万個のインスタンスを含み、0.99999のソフトマックススコアに対して、サンプルは、10万個のインスタンスを含み、0.999999のソフトマックススコアに対して、サンプルは、100万個のインスタンスを含む。

【0719】

ソフトマックスに関して、ソフトマックスは、マルチクラス分類のための出力活性化機能である。形式的には、いわゆるソフトマックス分類子を訓練することは、クラスではなくむしろ各クラスの尤度の信頼性予測を返すため、真の分類子よりもむしろクラス確率への回帰である。ソフトマックス関数は、値のクラスを取り、1になる確率にそれらを変換する。ソフトマックス関数は、任意の実数値のk次元ベクトルを、ゼロ～1の範囲内の実数値のk次元ベクトルにスカッシュする。したがって、ソフトマックス関数を使用することは、出力が有効で指数関数的に正規化された確率質量関数（非負及び1になる）であることを確実にする。

【0720】

【数28】

$$\tilde{y}_i$$

がベクトル

【数29】

$$\tilde{\mathbf{y}}=[\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n]$$

の第 i 番目の要素であるとする。

【数 3 0】

$$\hat{y}_i = (\text{softmax}(\tilde{z}))_i = \frac{\exp(\tilde{z}_i)}{\sum_{j=1}^{J-N} \exp(\tilde{z}_j)}$$

10

【0 7 2 1】

【数 3 1】

$$\tilde{y}$$

は、長さ n のベクトルであり、 n は、分類内のクラスの数である。これらの要素は、ゼロ～1 の値を有し、それらが有効な確率分布を表すように 1 になる。

20

【0 7 2 2】

例示的ソフトマックス活性化関数 1 3 4 0 6 は、図 1 3 4 に示される。ソフトマックス 1 3 4 0 6 は、

【数 3 2】

$$z \mapsto \text{softmax}\left(\left[z, \frac{z}{10}, -2z\right]\right)$$

30

として 3 つのクラスに適用される。3 つの出力は常に、1 になることに留意されたい。したがって、それらは、離散確率質量関数を定義する。

【0 7 2 3】

分類のために使用されるとき、

【数 3 3】

$$\tilde{y}_i$$

40

は、クラス i 内にある確率を与える。

【数 3 4】

$$P(Y=i|\tilde{z}) = (\text{softmax}(\tilde{z}))_i = \tilde{y}_i$$

【0 7 2 4】

「ソフトマックス」という名称は、幾分混乱し得る。関数は、マックス関数よりも、 a

50

r g m a x 関数に対してより密に関連付けられる。「ソフト」という用語は、ソフトマックス関数が連続的であり区別可能であるという事実に由来する。 a r g m a x 関数は、その結果がワンホットベクトルとして表され、連続的ではなく、又は区別可能でない。したがって、ソフトマックス関数は、 a r g m a x の「軟化した」バージョンを提供する。おそらく、ソフトマックス関数「 $\text{s o f t a r g m a x}$ 」をコールすることがより良好であるが、現在の名称は、定着した慣習である。

【0725】

図134は、品質スコアリングのためにニューラルネットワークベースのベースコーラー1514のベースコール信頼性確率10704を選択13400する一実施態様を示す。ニューラルネットワークベースのベースコーラー1514のベースコール信頼性確率10704は、分類スコア（例えば、ソフトマックススコア若しくはシグモイドスコア）又は回帰スコアであり得る。一実施態様では、ベースコール信頼性確率10704は、訓練10700中に生成される。

10

【0726】

いくつかの実施態様では、選択13400は、量子化に基づいて行われ、それは、ベースコール信頼性確率10704にアクセスし、かつ量子化分類スコア13404を生成する量子化器13402によって実行される。量子化分類スコア13404は、任意の実数であり得る。一実施態様では、量子化分類スコア13404は、

【数35】

20

$$0.9 \sum_{i=1}^n 0.1^{(i-1)}$$

と定義される選択式に基づいて選択される。別の実施態様では、量子化分類スコア13404は、

【数36】

30

$$\forall_{i=1}^{n-10} 0.1i$$

と定義される選択式に基づいて選択される。

【0727】

図135は、ニューラルネットワークベースの品質スコアリング13500の一実施態様を示す。量子化分類スコア13404の各々に対して、ベースコール誤差率13508及び／又はベースコール精度率13510は、対応するグラウンドトゥルースベースコール10708に対して（例えば、様々なサンプルサイズを有するバッチにおいて）そのベースコール予測10704を比較することによって決定される。比較は、比較器13502，によって実行され、それは次に、ベースコール誤差率判定器13504と、ベースコール精度率判定器13506と、を含む。

40

【0728】

次いで、量子化分類スコア13404と品質スコアとの間の対応を確立するために、適合判定器13512によって、量子化分類スコア13404とそれらのベースコール誤差率13508（及び／又はそれらのベースコール精度率13510）との間で適合が決定される。一実施態様では、適合判定器13512は、回帰モデルである。

【0729】

50

適合に基づいて、品質スコアは、相関器 1 3 5 1 4 よって、量子化分類スコア 1 3 4 0 4 と相関する。

【0 7 3 0】

図 1 3 6 a ~ 図 1 3 6 b は、ニューラルネットワークベースのベースコーラー 1 5 1 4 によって作成される、品質スコアとベースコール信頼性予測との間の対応 1 3 6 0 0 の一実施態様を示す。ニューラルネットワークベースのベースコーラー 1 5 1 4 のベースコール信頼性確率は、分類スコア（例えば、ソフトマックススコア若しくはシグモイドスコア）又は回帰スコアであり得る。図 1 3 6 a は、品質スコアに対する品質スコア対応スキーム 1 3 6 0 0 a である。図 1 3 6 b は、ピニング品質スコアに対する品質スコア対応スキーム 1 3 6 0 0 a である。

（推測）

【0 7 3 1】

図 1 3 7 は、推測 1 3 7 0 0 中にニューラルネットワークベースのベースコーラー 1 5 1 4 によって作成されるベースコール信頼性予測から品質スコアを推測する一実施態様を示す。ニューラルネットワークベースのベースコーラー 1 5 1 4 のベースコール信頼性確率は、分類スコア（例えば、ソフトマックススコア若しくはシグモイドスコア）又は回帰スコアであり得る。

【0 7 3 2】

推測 1 3 7 0 0 中、予測されたベースコール 1 3 7 0 6 は、そのベースコール信頼性確率（すなわち、最も高いソフトマックススコア（赤色））が最も対応する品質スコア 1 3 7 0 8 を割り当てられる。いくつかの実施態様では、品質スコア対応 1 3 6 0 0 は、品質スコア対応スキーム 1 3 6 0 0 a ~ 1 3 6 0 0 b を探すことによって作成され、品質スコア推測器 1 3 7 1 2 によって動作可能である。

【0 7 3 3】

いくつかの実施態様では、チャスティティフィルタ 1 3 7 1 0 は、そのコールされた塩基に割り当てられる品質スコア 1 3 7 0 8、又は連続するベースコールサイクルにおける平均品質スコアが予め設定された閾値を下回るとき、所与のクラスターのベースコールを終了する。

【0 7 3 4】

推測 1 3 7 0 0 は、バッティングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬 1 3 7 1 4 を含む。推測 1 3 7 0 0 は、（配列決定画像 1 0 8 及び／又は補足チャネル（例えば、距離チャネル、スケーリングチャネル）から導出される画像チャネルを有する）入力データを含む、推測データ 1 3 7 0 2 上で実行される。推測 1 3 7 0 0 は、テスト 1 3 7 0 4 によって動作可能である。

（ベースコール品質の直接の予測）

【0 7 3 5】

第 2 の十分に校正されたニューラルネットワークは、配列決定画像 1 0 8 から導出される入力データを処理し、品質指標を直接生成する、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 である。

【0 7 3 6】

一実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、多層パーセプトロン（MLP）である。別の実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、フィードフォワードニューラルネットワークである。更に別の実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、完全に接続されたニューラルネットワークである。更なる実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、完全畳み込みニューラルネットワークである。更に更なる実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、セマンティックセグメンテーションニューラルネットワークである。

【0 7 3 7】

一実施態様では、ニューラルネットワークベースの品質スコアラー 1 3 8 0 2 は、複数

10

20

30

40

50

の畳み込み層を有する畳み込みニューラルネットワーク（CNN）である。別の実施態様では、それは、長い短期メモリネットワーク（LSTM）、双方向LSTM（Bi-LSTM）、又はゲートされた反復単位（GRU）などの反復ニューラルネットワーク（RNN）である。更に別の実施態様では、それは、CNN及びRNNの両方を含む。

【0738】

更に他の実施態様では、ニューラルネットワークベースの品質スコアラー13802は、1D畳み込み、2D畳み込み、3D畳み込み、4D畳み込み、5D畳み込み、拡張又は膨張畳み込み、転置畳み込み、奥行分離可能な畳み込み、pointwise畳み込み、 1×1 畳み込み、グループ畳み込み、扁平畳み込み、空間及びクロスチャネルの畳み込み、シャッフルされたグループ化畳み込み、空間的な分離可能な畳み込み、並びに逆畳み込みを使用することができる。それは、ロジスティック回帰／対数損失、多クラスクロスエントロピー／ソフトマックス損失、二値クロスエントロピー損失、平均二乗誤差損失、L1損失、L2損失、平滑L1損失、及びHuber損失などの1つ又はそれ以上の損失機能を使用することができる。それは、TFRecord、圧縮符号化（例えば、PNG）、シャープ化、マップ変換に対する平行コール、バッチング、プリフェッチ、モデル並列、データ並列、及び同期／非同期SGDのような、任意の並列、効率、及び圧縮方式を使用することができる。これは、アップサンプリング層、ダウンサンプリング層、反復接続、ゲート及びゲートされたメモリユニット（LSTM又はGRUなど）、残留ブロック、残留接続、ハイウェイ接続、スキップ接続、ペエホル接続、アクティブ化機能（例えば、非線形変換関数は、整流線形ユニット（ReLU）、漏れやすいReLU、指数関数的ライナーユニット（ELU）、シグモイド及び双曲線正接（tanh））、バッチ正規化層、規則化層、ドロップアウト、プール層（例えば、最大又は平均プール）、グローバル平均プール層、及び注意機構のような非線形変換機能を含む。

【0739】

いくつかの実施態様では、ニューラルネットワークベースの品質スコアラー13802は、ニューラルネットワークベースのベースコーラー1514と同じ構造を有する。

【0740】

入力データは、配列決定画像108及び／又は補足チャネル（例えば、距離チャネル、スケーリングチャネル）から導出される画像チャネルを含み得る。ニューラルネットワークベースの品質スコアラー13802は、入力データを処理し、入力データの代替表現を生成する。代替表現は、いくつかの実施態様では畳み込み表現であり、他の実施態様では隠れ表現である。次いで、代替表現は、出力層によって処理されて、出力を生成する。出力は、品質指標を生成するために使用される。

【0741】

一実施態様では、同じ入力データが、ニューラルネットワークベースのベースコーラー1514及びニューラルネットワークベースの品質スコアラー13802に供給されて、（i）ニューラルネットワークベースのベースコーラー1514からベースコールを生成し、（ii）ニューラルネットワークベースの品質スコアラー13802から対応する品質指標を生成する。いくつかの実施態様では、ニューラルネットワークベースのベースコーラー1514及びニューラルネットワークベースの品質スコアラー13802は、エンドツーエンド逆伝搬と共同で訓練される。

【0742】

一実施態様では、ニューラルネットワークベースの品質スコアラー13802は、特定の配列決定サイクルで単一のターゲットクラスターのための品質指標を出力する。別の実施態様では、それは、特定の配列決定サイクルで複数のターゲットクラスター内の各ターゲットクラスターのための品質指標を出力する。更に別の実施態様では、それは、複数の配列決定サイクル内の各配列決定サイクルで複数のターゲットクラスター内の各ターゲットクラスターのための品質指標を出力し、それによって、各ターゲットクラスターのための品質指標配列を生成する。

【0743】

10

20

30

40

50

一実施態様では、ニューラルネットワークベースの品質スコアラ-13802は、配列決定画像108からのデータを含み、ベースコール品質グラウンドトゥールズでラベル付けされる訓練例で訓練される畳み込みニューラルネットワークである。ニューラルネットワークベースの品質スコアラ-13802は、畳み込みニューラルネットワーク13802のベースコール品質予測13804をベースコール品質グラウンドトゥールズ13808と漸進的に一致させる、逆伝搬ベースの勾配更新技術を使用して訓練される。いくつかの実施態様では、我々は、誤ったベースコールであった場合、塩基を0とラベル付けし、そうでない場合、塩基を1とラベル付けする。その結果、出力は、誤差の確率に対応する。一実施態様では、これは、入力特徴として配列コンテキストを使用する必要性を除去する。

10

【0744】

畳み込みニューラルネットワーク13802の入力モジュールは、1つ又はそれ以上のクラスターに対してコールされる1つ又はそれ以上の塩基の品質を決定するために、1つ又はそれ以上の配列決定サイクルで補足される配列決定画像108からのデータを畳み込みニューラルネットワーク13802に供給する。

【0745】

畳み込みニューラルネットワーク13802の出力モジュールは、畳み込みニューラルネットワーク13802による分析を、1つ又はそれ以上のクラスターに対してコールされる1つ又はそれ以上の塩基の品質を識別する出力13902に変換する。

【0746】

一実施態様では、出力モジュールは、高品質、中品質（任意選択的に、点線で示されるように）、及び低品質である品質状態に対する尤度を生成するソフトマックス分類層を更に備える。別の実施態様では、出力モジュールは、高品質及び低品質である品質状態に対する尤度を生成するソフトマックス分類層を更に備える。当業者は、異なって識別可能に品質スコアを汲み取る他のクラスが使用され得ることを理解するであろう。ソフトマックス分類層は、複数の品質スコアを割り当てられている品質に対する尤度を生成する。尤度に基づいて、品質は、複数の品質スコアのうちの1つから品質スコアを割り当てられる。品質スコアは、ベースコール誤差確率に対数的に基づく。複数の品質スコアは、Q6、Q10、Q15、Q20、Q22、Q27、Q30、Q33、Q37、Q40、及びQ50を含む。別の実施態様では、出力モジュールは、品質を識別する連続値を生成する回帰層を更に含む。

20

【0747】

いくつかの実施態様では、ニューラルネットワークベースの品質スコアラ-13802は、配列決定画像108からのデータを、コールされる塩基に対する品質予測値で補い、配列決定画像からのデータと共に、品質予測値を畳み込みニューラルネットワーク13802に供給する、補足入力モジュールを更に含む。

【0748】

いくつかの実施態様では、品質予測値は、オンライン重複、純度、フェイジング、start5、6量体スコア、モチーフ蓄積、endiness、近似ホモポリマー、強度減衰、最終チャスティティ、背景を有する信号重複（Signal Overlap With Background、SOWB）、及び／又はシフトされた純度G調整を含む。他の実施態様では、品質予測値は、ピーク高さ、ピーク幅、ピーク場所、相対的なピーク場所、ピーク高さ割り当て、ピーク間隔割り当て、及び／又はピーク対応を含む。品質予測値に関する追加の詳細は、参照により本明細書に完全に記載されているかのように組み込まれる、米国特許公開第2018/0274023号及び第2012/0020537号で見出され得る。

30

40

（訓練）

【0749】

図138は、ニューラルネットワークベースの品質スコアラ-13802を訓練13800として、配列決定画像108から導出される入力データを処理し、品質指標を直接生成する、一実施態様を示す。ニューラルネットワークベースの品質スコアラ-13802は

50

、正しい品質指標 13808 に対して、予測された品質指標 13804 を比較し、比較に基づいて誤差 13806 を計算する、逆伝搬ベースの勾配更新技術を使用して訓練される。次いで、誤差 13806 が、後方伝搬 13810 中にニューラルネットワークベースの品質スコアラー 13802 の重み及びパラメータに適用される勾配を計算するために使用される。訓練 13800 は、アダムなどの確率的勾配更新アルゴリズムを使用して、訓練器 1510 によって操作可能である。

【0750】

訓練器 1510 は、（配列決定画像 108 から導出される）訓練データ 13812 を使用して、予測された品質指標を生成する前方伝搬 13816、及び誤差 13806 に基づいて重み及びパラメータを更新する後方伝搬 13810 の数千及び数百万回の反復でニューラルネットワークベースの品質スコアラー 13802 を訓練する。いくつかの実施態様では、訓練データ 13812 は、品質予測値 13814 に補われる。訓練 13800 に関する追加の詳細は、「Deep Learning Tools」と題する付録で見出され得る。

（推測）

【0751】

図 139 は、推測 13900 中にニューラルネットワークベースの品質スコアラー 13802 の出力として品質指標を直接生成する一実施態様を示す。推測 13900 は、バッティングなどの平行四辺形技術を含む、数百、数千、及び／又は数百万回の前方伝搬 13908 を含む。推測 13900 は、（配列決定画像 108 及び／又は補足チャネル（例えば、距離チャネル、スケーリングチャネル）から導出される画像チャネルを有する）入力データを含む、推測データ 13904 上で実行される。いくつかの実施態様では、推測データ 13904 は、品質予測値 13906 に補われる。推測 13900 は、テスト 13910 によって動作可能である。

（無損失変換）

【0752】

図 140 は、ニューラルネットワークベースのテンプレート生成器 1512、ニューラルネットワークベースのベースコーラー 1514、及びニューラルネットワークベースの品質スコアラー 13802 への入力として供給され得る変換されたデータ 14004 を生成するために、無損失変換 1400 を使用する一実施態様を示す。無損失変換 1400 のいくつかの例は、畳み込み、逆畳み込み、及びフーリエ変換を含む。

【0753】

無損失変換 1400 は、複数のフィルタ $1-n$ （例えば、畳み込みカーネルを有する畳み込みフィルタ）を含む無損失変換器 14002 によって適用され得る。無損失変換 1400 は、入力データ 9500 及び／又は入力画像データ 1702 に適用されて、変換されたデータ 14004 を生成し得る。

【0754】

変換されたデータ 14004 は、ニューラルネットワークベースのテンプレート生成器 1512 への入力として供給されてクラスターメタデータを生成し、ニューラルネットワークベースのベースコーラー 1514 への入力として供給されてベースコールを生成し、及び／又はニューラルネットワークベースの品質スコアラー 13802 への入力として供給されて、品質スコアを生成し得る。

【0755】

いくつかの実施態様では、変換されたデータ 14004 は、逆畳み込み層 14006 によって逆畳み込みされて、入力データ 9500 及び／又は入力画像データ 1702 の必須特徴を再構成する。逆畳み込み層 14006 は、ニューラルネットワークベースのテンプレート生成器 1512、ニューラルネットワークベースのベースコーラー 1514、及び／又はニューラルネットワークベースの品質スコアラー 13802 の初期層であり得る。

（強度修正を伴うエンドツーエンド統合）

【0756】

10

20

30

40

50

ここでの考察は、ニューラルネットワークベースのテンプレート生成器 1512 が、強度修正技術を使用してニューラルネットワークベースのベースコーラー 1514 とどのように統合されるかについて説明する。

【0757】

上述のベースコール実施態様の多くでは、ニューラルネットワークベースのベースコーラー 1514 への入力として供給される入力データ 9500 は、(i) 配列決定画像 108 から導出される画像データ 7902 (画像チャンネル) と、(ii) ニューラルネットワークベースのテンプレート生成器 1512 の出力 1714 (例えば、減衰マップ 1716、三元マップ 1718、又はバイナリマップ 1720) から導出される補足距離データ (距離チャンネル) と、(iii) 補足スケーリングデータ (スケーリングチャンネル) と、を含む。これらの実施態様では、画像データ 7902 内の強度値は、修正されるのではなく、むしろ、画像データ 7902 内のどのピクセルがクラスター中心を含み、画像データ 7902 内のどのピクセルがクラスター中心からより離れているかを伝えることによって、クラスター形状情報を通信する距離値に補われる。

【0758】

ここで、我々は、クラスター形状情報を組み込むように画像データ 7902 を修正し、したがって、補足距離チャンネルを計算及び使用する必要性を除去する、ベースコール実施態様を開示する。画像データ 7902 は、減衰マップ 1716、三元マップ 1718、及びバイナリマップ 1720、に基づいて修正され、それは次に、ニューラルネットワークベースのテンプレート生成器 1512 の出力 1714 である。したがって、このコンテキストでは、「統合」は、ニューラルネットワークベースのテンプレート生成器 1512 (例えば、減衰マップ 1716、三元マップ 1718、及びバイナリマップ 1720) によって生成される情報に基づいて、ニューラルネットワークベースのベースコーラー 1514 によって処理されるデータを修正すること (前者を後者に補うのとは対照的) を指す。

【0759】

減衰及び三元マップの両方は、以下のようなサブピクセルを識別するクラスター形状情報を含む。(1) 背景サブピクセル、(2) クラスター中心サブピクセル、及び(3) 同じクラスターに属するクラスター又はクラスター内部サブピクセル。クラスター形状情報は、アップサンプリングされたサブピクセルドメイン内のテンプレート画像に含まれて、クラスター境界を微細化レベルで区別する。しかしながら、クラスター及び背景強度を含む画像データ 7902 は、典型的には、光学ピクセルドメイン内にある。

【0760】

テンプレート画像及び画像データ 7902 は、異なるドメイン内にあるが、それらは、同じ画像化領域を表す。テンプレート画像は、配列決定動作の特定の数の初期配列決定サイクルに対する入力画像データ 1702 の処理、及び減衰マップ 1716、三元マップ 1718、又はバイナリマップ 1720 の後処理から導出される。対照的に、クラスター形状情報組み込みに対する修正は、初期配列決定サイクルに対する画像データ 7902 に限定されないが、その代わりに、ベースコールされる各配列決定サイクルに対して画像データ 7902 に適用される。

【0761】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー 1514、すなわち、減衰マップ 1716、三元マップ 1718 及びバイナリマップ 1720 の出力は、光学的ピクセルドメイン内にある。したがって、このような実施態様形態では、テンプレート画像はまた、光ピクセルドメイン内にある。

【0762】

そのため、例えば、配列決定動作は、ベースコールされる 300 回の配列決定サイクルを含むと考える。次いで、テンプレート画像は、ニューラルネットワークベースのテンプレート生成器 1512 を介した第 1 の 2~7 回の配列決定サイクルに対する入力画像データ 1702 の処理、及び出力としてニューラルネットワークベースのテンプレート生成器

10

20

30

40

50

1512によって生成される、減衰マップ1716、三元マップ1718、又はバイナリマップ1720の後処理から導出される。一方、300回の配列決定サイクルの各々に対する画像データ7902は、クラスター形状情報のために修正され、次いで、ニューラルネットワークベースのベースコーラー1514を介して処理されて、300回の配列決定サイクルの各々に対してベースコールを生成する。

【0763】

我々は、ニューラルネットワークベースのベースコーラー1514によるベースコールのために、画像データ7902にクラスター形状情報を組み込む、強度修正技術を開示する。以下の詳細がある。

(領域重み付け係数 (Area Weighting Factoring, AWF))

10

【0764】

第1のタイプの強度修正技術は、強度修正が、光学ピクセルドメイン内の画像データ7902内のピクセルに適用される、領域重み付け係数技術である。

【0765】

テンプレート画像及び画像データ7902が同じ画像化領域を表すため、テンプレート画像内のサブピクセルブロックと画像データ7902内のそれぞれのピクセルとの間には、多対1の対応が存在する。例えば、テンプレート画像内の16個のサブピクセルの第1のブロックは、画像データ7902内の第1のピクセルに対応し、テンプレート画像内の16個のサブピクセルの第2のブロックは、画像データ7902内の第2のピクセルに対応するなどである。

20

【0766】

ベースコールされる所与のクラスターに対して、我々は、テンプレート画像からそのクラスター形状情報にアクセスし、画像データ7902内のどのピクセルが所与のクラスターの一部を含むか、すなわち、画像データ7902内のどのピクセルが所与のクラスターをカバーするか、又は所与のクラスターからの強度放射を示すかを識別する。

【0767】

次いで、画像データ7902内の識別された各ピクセルに対して、我々は、画像データ7902内の識別されたピクセルに対応するテンプレート画像内のどれくらい多くのサブピクセルが、所与のクラスターの一部を含むか、すなわち、画像データ7902内の識別されたピクセルに対応するテンプレート画像内のどれくらい多くのサブピクセルが所与のクラスターをカバーするか、又は所与のクラスターからの強度放射を示すかを決定する。次いで、決定されたサブピクセルカウントに基づいて、我々は、画像データ7902内の識別された各ピクセルに領域重み付け係数 (AWF) を割り当てる。

30

(1ピクセル当たり単一のクラスターの係数)

【0768】

単一のピクセル i に対する AWF は、以下のように計算される。

【数37】

$$\text{ピクセル}i\text{のAWF} = \frac{\text{ピクセル}i\text{に対応し、所与のクラスターを示す、テンプレート画像内のサブピクセルの数}}{\text{ピクセル}i\text{に対応するサブピクセルの総数}}$$

40

【0769】

上記 AWF 計算は、サブピクセルカウントから以下を除外する。(i) 背景サブピクセル、及び (ii) 任意の他のクラスターの一部を含むサブピクセル (すなわち、所与のクラスター以外のクラスターを示すサブピクセル)。この例は、図143に示される。

【0770】

次いで、我々は、その AWF に基づいて、識別された各ピクセルの強度値を修正する。これは、所与のクラスターをベースコールするためのニューラルネットワークベースのベ

50

ースコーラー 1 5 1 4 によって処理される画像データ 7 9 0 2 の修正バージョンをもたらす。

【0 7 7 1】

ピクセル i の修正強度値 (Modified Intensity Value、M I V) は、以下のように計算される。

ピクセル i の M I V = ピクセル i の A W F X (画像データ 7 9 0 2 内の) ピクセル i の光学強度値

【0 7 7 2】

図 1 4 3 は、1 ピクセル当たり単一のクラスターのみからの寄与に対する領域重み付け係数 1 4 3 0 0 の一例を示す。図 1 4 3 では、画像データ 7 9 0 2 の配列決定画像 1 4 3 0 4 内のピクセルの強度が修正される。配列決定画像 1 4 3 0 4 は、強度 1 0 0、1 4 0、1 6 0、及び 3 2 0 をそれぞれ有する 4 つのピクセルを含む。

【0 7 7 3】

テンプレート画像 1 4 3 0 2 は、配列決定画像 1 4 3 0 4 に対するクラスター形状情報を含む。テンプレート画像 1 4 3 0 2 は、配列決定画像 1 4 3 0 4 内の 4 つのピクセルにそれぞれ対応する 4 つのサブピクセルブロック (すなわち、配列決定画像 1 4 3 0 4 内の 1 ピクセル当たりのテンプレート画像 1 4 3 0 2 内の 1 6 個のサブピクセル) を含む。テンプレート画像 1 4 3 0 2 はまた、3 つのクラスター A、B、及び C に対する背景サブピクセル及びクラスターサブピクセルを識別する。

【0 7 7 4】

次いで、配列決定画像 1 4 3 0 4 内の 4 つのピクセルの各々に対する A W F は、1 ピクセル当たりのクラスター A のみを考慮するために計算され、テンプレート画像 1 4 3 0 2 内の A W F 1 4 3 0 6 として記憶される。第 2 及び第 3 のピクセルに対する A W F はそれぞれ、 $7/16$ 及び $8/16$ であることに留意されたい。第 2 のピクセルが 2 つのクラスター A 及び C からの寄与を受信したとしても、その A W F は、クラスター A (赤色) をカバーする 7 個のサブピクセルのみを考慮し、クラスター C (オレンジ色) をカバーする 4 個のサブピクセルを無視する。同様に、第 3 のピクセルが 2 つのクラスター A 及び B からの寄与を受信したとしても、その A W F は、クラスター A (赤色) をカバーする 8 個のサブピクセルのみを考慮し、クラスター B (緑色) をカバーする 4 個のサブピクセルを無視する。背景サブピクセルは、カウントされない。

【0 7 7 5】

A W F 1 4 3 0 6 は、4 つのピクセルの各々の強度を修正し、修正された配列決定画像 1 4 3 0 8 を生成するために更に使用される。修正された配列決定画像 1 4 3 0 8 は、ベースコールのためのニューラルネットワークベースのベースコーラー 1 5 1 4 によって処理される。

(1 ピクセル当たり複数のクラスターの係数)

【0 7 7 6】

いくつかの実施態様では、我々は、複数のクラスターから画像データ 7 9 0 2 内の単一のピクセルへの寄与を考慮する。複数のクラスターからの寄与を受信する単一のピクセル i に対する A W F は、以下のように計算される。

【数 3 8】

$$\text{ピクセル } i \text{ の A W F} = \frac{\text{ピクセル } i \text{ に対応し、少なくとも 1 つのクラスターを示す、テンプレート画像内のサブピクセルの数}}{\text{ピクセル } i \text{ に対応するサブピクセルの総数}}$$

【0 7 7 7】

上記 A W F 計算は、サブピクセルカウントから背景サブピクセルを除外するが、他のクラスターの一部を含むサブピクセル (すなわち、所与のクラスター以外のクラスターを示

すサブピクセル) をサブピクセルカウントに含む。この例は、図 1 4 4 に示される。

【0 7 7 8】

図 1 4 4 は、1 ピクセル当たり複数のクラスターからの寄与に対する領域重み付け係数 1 4 4 0 0 の一例を示す。図 1 4 4 では、画像データ 7 9 0 2 の配列決定画像 1 4 4 0 4 内のピクセルの強度が修正される。配列決定画像 1 4 4 0 4 は、強度 1 0 0、1 4 0、1 6 0、及び 3 2 0 をそれぞれ有する 4 つのピクセルを含む。

【0 7 7 9】

テンプレート画像 1 4 4 0 2 は、配列決定画像 1 4 4 0 4 に対するクラスター形状情報を含む。テンプレート画像 1 4 4 0 2 は、配列決定画像 1 4 4 0 4 内の 4 つのピクセルにそれぞれ対応する 4 つのサブピクセルブロック（すなわち、配列決定画像 1 4 4 0 4 内の 1 ピクセル当たりのテンプレート画像 1 4 3 0 2 内の 1 6 個のサブピクセル）を含む。テンプレート画像 1 4 4 0 2 はまた、3 つのクラスター A、B、及び C に対する背景サブピクセル及びクラスターサブピクセルを識別する。

【0 7 8 0】

次いで、配列決定画像 1 4 4 0 4 内の 4 つのピクセルの各々に対する AWF は、1 ピクセル当たり全ての 3 つのクラスター A、B、及び C を考慮するために計算され、テンプレート画像 1 4 4 0 2 内の AWF 1 4 4 0 6 として記憶される。第 2 及び第 3 のピクセルに対する AWF はそれぞれ、 $11/16$ 及び $12/16$ であることに留意されたい。第 2 のピクセルが 2 つのクラスター A 及び C からの寄与を受信するため、その AWF は、クラスター A（赤色）をカバーする 7 個のサブピクセルを考慮し、クラスター C（オレンジ色）をカバーする 4 個のサブピクセルも考慮する。同様に、第 3 のピクセルが 2 つのクラスター A 及び B からの寄与を受信するため、その AWF は、クラスター A（赤色）をカバーする 8 個のサブピクセルを考慮し、クラスター B（緑色）をカバーする 4 個のサブピクセルも考慮する。背景サブピクセルは、カウントされない。

【0 7 8 1】

AWF 1 4 4 0 6 は、4 つのピクセルの各々の強度を修正し、修正された配列決定画像 1 4 4 0 8 を生成するために更に使用される。修正された配列決定画像 1 4 4 0 8 は、ベースコールのためのニューラルネットワークベースのベースコーラー 1 5 1 4 によって処理される。

【0 7 8 2】

上述の領域重み付け係数技術は、単一のターゲットクラスターをベースコールするために使用され得、複数のターゲットクラスターを同時にベースコールするためにも使用され得る。

（アップサンプリング及び背景マスキング）

【0 7 8 3】

第 2 のタイプの強度修正技術は、画像データ 7 9 0 2 が、テンプレート画像と同じアップサンプリングされたサブピクセルドメイン内にあるように最初にアップサンプリングされて、次いで、強度修正が、画像データ 7 9 0 2 のアップサンプリングされたバージョンでのサブピクセルに適用される、アップサンプリング及び背景マスキングである。

【0 7 8 4】

テンプレート画像及び画像データ 7 9 0 2 が同じ画像化領域を表すため、テンプレート画像内のサブピクセルと画像データ 7 9 0 2 のアップサンプリングされたバージョンでのそれぞれのサブピクセルとの間には、1 対 1 の対応が存在する。例えば、テンプレート画像内の第 1 のサブピクセルは、画像データ 7 9 0 2 のアップサンプリングされたバージョンでの第 1 のサブピクセルに対応し、テンプレート画像内の第 2 のサブピクセルは、画像データ 7 9 0 2 のアップサンプリングされたバージョンでの第 2 のサブピクセルに対応するなどである。

【0 7 8 5】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー 1 5 1 4、すなわち、減衰マップ 1 7 1 6、

10

20

30

40

50

三元マップ 1 7 1 8 及びバイナリマップ 1 7 2 0 の出力は、光学的ピクセルドメイン内にある。したがって、このような実施態様形態では、テンプレート画像はまた、光ピクセルドメイン内にある。

(補間)

【0 7 8 6】

テンプレート画像内のクラスター形状情報を使用して、我々は、画像データ 7 9 0 2 内のピクセルにブロックごとに対応するテンプレート画像内のサブピクセル間で、テンプレート画像内のどのサブピクセルが、任意のクラスターからの／それをカバーする強度放射に寄与しない／それを示す背景サブピクセルであるか、及びどれが、少なくとも 1 つのクラスターからの／それをカバーする強度放射に寄与する／それを示すクラスターサブピクセルであるかを最初に識別する。

10

【0 7 8 7】

次いで、我々は、補間を使用して、アップサンプリングされたサブピクセルドメイン内の画像データ 7 9 0 2 をアップサンプリングし、画像データ 7 9 0 2 のアップサンプリングされたバージョンを生成して、(1) テンプレート画像内の識別された背景サブピクセルにそれぞれ対応する、画像データ 7 9 0 2 のアップサンプリングされたバージョンでのサブピクセルが、同じ背景強度(例えば、ゼロ値又はゼロに近い値)を割り当てられ、(2) テンプレート画像内の識別されたクラスターサブピクセルにそれぞれ対応する、画像データ 7 9 0 2 のアップサンプリングされたバージョンでのサブピクセルは、光学ピクセルドメイン内のピクセル強度から補間されるクラスター強度を割り当てられる。この例は、図 1 4 5 に示される。

20

【0 7 8 8】

図 1 4 5 は、アップサンプリング及び背景マスキング 1 4 5 0 0 のための補間を使用する一例を示す。図 1 4 5 では、画像データ 7 9 0 2 の配列決定画像 1 4 5 0 4 内のピクセルの強度が修正される。配列決定画像 1 4 5 0 4 は、強度 1 6 0、8 0、3 2 0、及び 2 0 0 をそれぞれ有する 4 つのピクセルを含む。

【0 7 8 9】

テンプレート画像 1 4 5 0 2 は、配列決定画像 1 4 5 0 4 に対するクラスター形状情報を含む。テンプレート画像 1 4 5 0 2 は、配列決定画像 1 4 5 0 4 内の 4 つのピクセルにそれぞれ対応する 4 つのサブピクセルブロック(すなわち、配列決定画像 1 4 5 0 4 内の 1 ピクセル当たりのテンプレート画像 1 4 5 0 2 内の 1 6 個のサブピクセル)を含む。テンプレート画像 1 4 5 0 2 はまた、3 つのクラスター A、B、及び C に対する背景サブピクセル及びクラスターサブピクセルを識別する。

30

【0 7 9 0】

補間は、配列決定画像 1 4 5 0 4 をアップサンプリングするために、及びサブピクセルを有するアップサンプリングされた配列決定画像 1 4 5 0 6 を生成するために使用される。補間は、背景強度を背景サブピクセルに割り当て、ピクセル強度から補間されるクラスター強度をクラスターサブピクセルに割り当てる。

(サブピクセルカウント重み付け)

【0 7 9 1】

ここで、クラスター強度は、異なって計算される。すなわち、ピクセル強度を補間する代わりに、光学ピクセルドメイン内の各ピクセルの強度は、対応するピクセルを構成する、画像データ 7 9 0 2 のアップサンプリングされたバージョンでのクラスターサブピクセル間で等しく分配される。各ピクセルに対して、その強度が等しく分散されるその構成クラスターサブピクセルのカウントは、複数のクラスターからの寄与を考慮する上述の領域重み付け係数(AWF)に基づいて決定され得る。背景サブピクセルは、上述のように、背景強度を割り当てられる。この例は、図 1 4 6 に示される。

40

【0 7 9 2】

図 1 4 6 は、アップサンプリング及び背景マスキング 1 4 6 0 0 のためのサブピクセルカウント重み付けを使用する一例を示す。図 1 4 6 では、画像データ 7 9 0 2 の配列決定

50

画像 1 4 6 0 4 内のピクセルの強度が修正される。配列決定画像 1 4 6 0 4 は、強度 1 6 0、8 0、3 2 0、及び 2 0 0 をそれぞれ有する 4 つのピクセルを含む。

【0 7 9 3】

テンプレート画像 1 4 6 0 2 は、配列決定画像 1 4 6 0 4 に対するクラスター形状情報を含む。テンプレート画像 1 4 6 0 2 は、配列決定画像 1 4 6 0 4 内の 4 つのピクセルにそれぞれ対応する 4 つのサブピクセルブロック（すなわち、配列決定画像 1 4 6 0 4 内の 1 ピクセル当たりのテンプレート画像 1 4 6 0 2 内の 1 6 個のサブピクセル）を含む。テンプレート画像 1 4 6 0 2 はまた、3 つのクラスター A、B、及び C に対する背景サブピクセル及びクラスターサブピクセルを識別する。

【0 7 9 4】

サブピクセルカウント重み付けは、配列決定画像 1 4 6 0 4 をアップサンプリングするために、及びサブピクセルを有するアップサンプリングされた配列決定画像 1 4 6 0 6 を生成するために使用される。サブピクセルカウント重み付けは、背景強度を背景サブピクセルに割り当て、各ピクセルの全体の強度をその構成クラスターサブピクセルに分配する。すなわち、ピクセルからその構成クラスターサブピクセルへの強度割り当ては、ピクセルを構成する背景サブピクセルへの割り当てがなく、又は最小の割り当てで、ピクセルの強度のいくつかを無駄にすることなく、ピクセルの強度の全てを利用する。いくつかの実施態様では、ピクセルの強度は、その構成クラスターサブピクセル間で等しく分配される。

【0 7 9 5】

他の実施態様では、アップサンプリングは、最近傍強度抽出、ガウス系強度抽出、平均 2×2 のサブピクセル領域に基づく強度抽出、最も明るい 2×2 のサブピクセル領域に基づく強度抽出、平均 3×3 のサブピクセル領域に基づく強度抽出、バイリニア強度抽出、双次強度抽出、及び／又は重み付け領域被覆に基づく強度抽出のうちの少なくとも 1 つを使用して実行される。

【0 7 9 6】

いくつかの実施態様では、アップサンプリングの前に、画像データ 7 9 0 2 は、サイクル固有及び撮像チャネル固有の変換を使用してテンプレート画像と位置合わせされる。

【0 7 9 7】

クラスター強度及び背景強度を含む画像データ 7 9 0 2 のアップサンプリングされたバージョンは、ベースコールのためにニューラルネットワークベースのベースコーラー 1 5 1 4 によって処理される。

【0 7 9 8】

他の実施態様では、減衰マップ、バイナリマップ、及び／又は三元マップでの値は、画像データ 7 9 0 2 内のピクセルの強度、又は画像データ 7 9 0 2 のアップサンプリングされたバージョンでのサブピクセルの強度を直接変調するために使用される。

（統合ワークフロー）

（領域重み付け係数）

【0 7 9 9】

図 1 4 1 は、領域重み付け係数を使用して、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 をニューラルネットワークベースのベースコーラー 1 5 1 4 と統合する一実施態様を示す。

【0 8 0 0】

最初に、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 は、配列決定動作のいくつかの初期配列決定サイクルに対する入力画像データ 1 7 0 2 を処理し、減衰マップ 1 7 1 6、三元マップ 1 7 1 8、又はバイナリマップ 1 7 2 0 を出力として生成する。入力画像データ 1 7 0 2 は、次に、図 2 1 b～図 2 4 を参照して上述したように、配列決定画像 1 0 8 から導出される。一実施態様では、入力画像データ 1 7 0 2 は、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 への入力として供給される前に、アップサンプリングされたサブピクセルドメイン／解像度内である。別の実施態様では、ニューラルネットワークベースのテンプレート生成器 1 5 1 2 のアップサンプリング層は、

10

20

30

40

50

アップサンプリングされたサブピクセルドメイン／解像度内にあるように、入力画像データ1702をアップサンプリングする。アップサンプリングは、2次補間などの補間技術によって達成され得る。

【0801】

ニューラルネットワークベースのテンプレート生成器1512の出力1714（減衰マップ1716、三元マップ1718、又はバイナリマップ1720）から、テンプレート画像14102は、上述のような後処理を介して導出される。テンプレート画像14202は、アップサンプリングされたサブピクセルドメイン／解像度内のクラスターメタデータを含む。クラスターメタデータ1812は、クラスター中心、クラスター形状、クラスター境界、及び／又はクラスター背景を識別する。「テンプレート画像」又は「テンプレート」は、減衰マップ1716、三元マップ1718及び／又はバイナリマップ1718に由来するクラスターメタデータ1812を含むか、又は識別するデータ構造を指すことができる。

10

【0802】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー1514、すなわち、減衰マップ1716、三元マップ1718及びバイナリマップ1720の出力は、光学的ピクセルドメイン内にある。したがって、そのような実施態様では、テンプレート画像14202はまた、光学ピクセルドメイン内にある。

【0803】

次いで、領域重み付け係数判定器14104は、上述のように、テンプレート画像14102を使用して、領域重み付け係数を決定し、テンプレート画像14102内に記憶する。

20

【0804】

次いで、配列決定動作の配列決定サイクルの各々に対して、画像データ7902は、テンプレート画像14102内に記憶される領域重み付け係数に基づいて、強度修正器14106によって修正される。他の実施態様では、領域重み付け係数は、他の箇所に記憶され得る。

【0805】

その結果は、ニューラルネットワークベースのベースコーラー1514によって処理されてベースコール14110を生成する、強度修正画像14108である。強度修正画像14108は、任意の補足距離チャンネルを含まないが、補足スケーリングチャンネルを含み得ることに留意されたい。

30

【0806】

他の実施態様では、強度修正は、配列決定動作の配列決定サイクルのサブセットに対してのみ実行される。

（アップサンプリング及び背景マスキング）

【0807】

図142は、アップサンプリング及び背景マスキングを使用して、ニューラルネットワークベースのテンプレート生成器1512をニューラルネットワークベースのベースコーラー1514と統合する別の実施態様を示す。

40

【0808】

最初に、ニューラルネットワークベースのテンプレート生成器1512は、配列決定動作のいくつかの初期配列決定サイクルに対する入力画像データ1702を処理し、減衰マップ1716、三元マップ1718、又はバイナリマップ1720を出力として生成する。入力画像データ1702は、次に、図21b～図24.を参照して上述したように、配列決定画像108から導出される。一実施態様では、入力画像データ1702は、ニューラルネットワークベースのテンプレート生成器1512への入力として供給される前に、アップサンプリングされたサブピクセルドメイン／解像度内である。別の実施態様では、ニューラルネットワークベースのテンプレート生成器1512のアップサンプリング層は

50

、アップサンプリングされたサブピクセルドメイン／解像度内にあるように、入力画像データ1702をアップサンプリングする。アップサンプリングは、2次補間などの補間技術によって達成され得る。

【0809】

ニューラルネットワークベースのテンプレート生成器1512の出力（減衰マップ1716、三元マップ1718、又はバイナリマップ1720）から、テンプレート画像14202は、上述のような後処理を介して導出される。テンプレート画像14202は、アップサンプリングされたサブピクセルドメイン／解像度内のクラスターメタデータを含む。クラスターメタデータ1812は、クラスター中心、クラスター形状、クラスター境界、及び／又はクラスター背景を識別する。「テンプレート画像」又は「テンプレート」は、減衰マップ1716、三元マップ1718及び／又はバイナリマップ1718に由来するクラスターメタデータ1812を含むか、又は識別するデータ構造を指すことができる。

10

【0810】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー1514、すなわち、減衰マップ1716、三元マップ1718及びバイナリマップ1720の出力は、光学的ピクセルドメイン内にある。したがって、そのような実施態様では、テンプレート画像14202はまた、光学ピクセルドメイン内にある。

【0811】

次いで、画像統合器14204は、上述のように、テンプレート画像14202を使用して、補間又はサブピクセルカウント重み付けを使用して配列決定動作の配列決定サイクルの各々に対して画像データ7902をアップサンプリングし、配列決定動作の配列決定サイクルの各々に対して画像データ7902のアップサンプリングされたバージョン14212を生成する。

20

【0812】

アップサンプリングは、画像アップサンプラー14208によって動作可能である。一実施態様では、画像データ7902のアップサンプリングされたバージョン14212は、ニューラルネットワークベースのベースコーラー1514への入力として供給される前に生成される。別の実施態様では、アップサンプリング層ニューラルネットワークベースのベースコーラー1514は、画像データ7902をアップサンプリングし、画像データ7902のアップサンプリングされたバージョン14212を生成する。アップサンプリングは、2次補間などの補間技術によって達成され得る。

30

【0813】

画像統合器14204はまた、上述のように、配列決定動作の配列決定サイクルの各々に対して、画像データ7902のアップサンプリングされたバージョン14212内の背景サブピクセルに背景マスクを適用する。背景マスキングは、背景マスカ14210によって適用される。

【0814】

いくつかの実施態様では、アップサンプリングの前に、画像統合器14204はまた、上述のように、配列決定動作の配列決定サイクルの各々に対して、画像データ7902をテンプレート画像14202と位置合わせする。位置合わせは、画像アライナー14206によって動作可能である。

40

【0815】

次いで、配列決定動作の配列決定サイクルの各々に対して、画像データ7902のアップサンプリングされたバージョン14212は、ニューラルネットワークベースのベースコーラー1514によって処理されて、ベースコール14214を生成する。画像データ7902のアップサンプリングされたバージョン14212は、任意の補足距離チャンネルを含まないが、補足スケーリングチャンネルを含み得ることに留意されたい。

【0816】

他の実施態様では、アップサンプリング及び背景マスキングは、配列決定動作の配列決

50

定サイクルのサブセットに対してのみ実行される。

(強度修正なしで、代わりに非距離補足チャンネルを使用するエンドツーエンド統合)

【0817】

ここでの考察は、ニューラルネットワークベースのテンプレート生成器1512が、配列決定画像の強度データを修正することなく、ニューラルネットワークベースのベースコーラー1514とどのように統合されるかについて説明する。以下で論じられる実施態様は、上述の補足距離チャンネルとは異なる新しい補足チャンネルを提供する。これらの新しい補足チャンネルはまた、クラスター形状情報を伝える。

1. 補足チャンネルとしての、減衰マップ、三元マップ、バイナリマップ

【0818】

ここで、我々は、ニューラルネットワークベースのテンプレート生成器1512の出力1714、すなわち、減衰マップ1716、三元マップ1718、及びバイナリマップ1720で、画像データ7902を補うベースコール実施態様を開示する。したがって、このコンテキストでは、「統合」は、ニューラルネットワークベースのテンプレート生成器1512(例えば、減衰マップ1716、3元マップ1718、及びバイナリマップ1720)によって生成される情報で、ニューラルネットワークベースのベースコーラー1514によって処理されるデータを補うことを指す。

【0819】

減衰マップ1716、三元マップ1718、及びバイナリマップ1720は、アップサンプリングされたサブピクセルドメイン内にあるが、クラスター及び背景強度を含む画像データ7902は、典型的には、光学ピクセルドメイン内にある。

【0820】

しかしながら、いくつかの実施態様では、クラスターサイズが十分に大きいとき、ニューラルネットワークベースのベースコーラー1514、すなわち、減衰マップ1716、三元マップ1718及びバイナリマップ1720の出力は、光学的ピクセルドメイン内にある。したがって、このような実施態様形態では、テンプレート画像はまた、光ピクセルドメイン内にある。

(入力画像データのアップサンプリング)

【0821】

減衰マップ1716、三元マップ1718、及びバイナリマップ1720がアップサンプリングされたサブピクセルドメイン内にあるとき、いくつかの実施態様では、入力画像データ1702は、アップサンプリングされたサブピクセルドメイン内にあるようにアップサンプリングされる。一実施態様では、アップサンプラー2302は、補間(例えば、2次補間)を使用して、一連の画像セット2100内の配列決定画像108をアップサンプリング係数(例えば、4x)及び一連のアップサンプリングされた画像セット2300によってアップサンプリングする。

【0822】

次いで、減衰マップ1716、三元マップ1718、又はバイナリマップ1720は、入力画像データ1702(アップサンプリングされたサブピクセルドメイン内でもある)でサブピクセルごとに補われ、入力画像データ1702(アップサンプリングされたサブピクセルドメイン内でもある)と共に補足チャンネルとしてニューラルネットワークベースのベースコーラー1514に供給される。

(減衰マップ、三元マップ、バイナリマップのダウンサンプリング)

【0823】

他の実施態様では、減衰マップ1716、三元マップ1718、及びバイナリマップ1720が、アップサンプリングされたサブピクセルドメイン内で生成されるとき、それらは、光学ピクセルドメイン内にあるようにダウンサンプリングされる。一実施態様では、ダウンサンプリングは、ダウンサンプリング係数に基づいてサブピクセルをグループ化することと、グループ化されたサブピクセルの出力値の平均を取り、それを光学ピクセルドメイン内の対応するピクセルに割り当てることと、を含み得る。出力値は、減衰マップ1

10

20

30

40

50

716の場合の重み付けされた減衰値である。出力値は、三元マップ1718の場合の三元分類スコアである。出力値は、バイナリマップ1720の場合の2方向分類スコアである。別の実施態様では、ダウンサンプリングは、同じクラスターに属することに基づいてサブピクセルをグループ化することと、グループ化されたサブピクセルの出力値の平均を取り、それを光学ピクセルドメイン内の対応するピクセル（複数可）に割り当てることと、を含み得る。

【0824】

次いで、光学ピクセルドメイン内の減衰マップ1716、三元マップ1718、又はバイナリマップ1720は、入力画像データ1702（光学ピクセルドメイン内でもある）でピクセルごとに補われ、入力画像データ1702（光学ピクセルドメイン内でもある）と共に補足チャネルとしてニューラルネットワークベースのベースコーラー1514に供給される。

2.（補足チャネルとしての領域重み付け係数）

【0825】

一実施態様では、テンプレート画像（例えば、14306及び14406）内に含まれる領域重み付け係数は、上述のように計算されるが、強度値を修正し、修正された配列決定画像（例えば、14308及び14408）を生成するために使用される代わりに、それら自体は、修正されていない元の配列決定画像（例えば、14304及び14404）に付随する補足チャネルとして提供される。すなわち、テンプレート画像（例えば、14306及び14406）内に含まれる領域重み付け係数が、光学ピクセルドメイン内にあるため、それらは、修正されていない入力画像データ1702（光学ピクセルドメイン内でもある）でピクセルごとに補われ、修正されていない入力画像データ1702（光学ピクセルドメイン内でもある）と共にニューラルネットワークベースのベースコーラー1514に補足チャネルとして供給される。

【0826】

したがって、このコンテキストでは、「統合」は、ニューラルネットワークベースのテンプレート生成器1512（例えば、減衰マップ1716、3元マップ1718、及びバイナリマップ1720）の出力から導出される情報（例えば、領域重み付け係数）で、ニューラルネットワークベースのベースコーラー1514によって処理されるデータを補うことを指す。

（データ前処理）

【0827】

いくつかの実施態様では、開示される技術は、画像データ202内のピクセルに適用され、かつ前処理された画像データ202pを生成する、前処理技術を使用する。そのような実施態様では、画像データ202の代わりに、前処理された画像データ202pは、ニューラルネットワークベースのベースコーラー1514への入力として提供される。データ前処理は、データプリプロセッサ15002によって動作可能であり、それは次に、データ正規化器15032及びデータ増強器15034を含み得る。

【0828】

図150は、データ正規化及びデータ増強を含み得る、データ前処理の異なる実施態様を示す。

（データ正規化）

【0829】

一実施態様では、データ正規化は、画像パッチごとに、画像データ202内のピクセルに適用される。これは、得られた正規化された画像パッチのピクセル強度ヒストグラムが、ゼロの第5のパーセンタイル及び1の第95のパーセンタイルを有するように、画像パッチ内のピクセルの強度値を正規化することを含む。すなわち、正規化された画像パッチでは、（i）5%のピクセルは、ゼロ未満の強度値を有し、（ii）別の5%のピクセルは、1よりも大きい強度値を有する。画像データ202のそれぞれの画像パッチは、別々に正規化され得るか、又は画像データ202は、全て一度に正規化され得る。その結果は

、前処理された画像データ202pの一例である、正規化された画像パッチ15016である。データ正規化は、データ正規化器15032によって動作可能である。

(データ増強)

【0830】

一実施態様では、データ増強は、画像データ202内のピクセルの強度値に適用される。これは、(i)画像データ202内の全てのピクセルの強度値に同じスケーリング係数を乗算することと、(ii)画像データ202内の全てのピクセルのスケーリングされた強度値に同じオフセット値を追加することと、を含む。単一のピクセルに対して、これは、以下の式によって表現され得る。

増強されたピクセル強度 (Augment Pixel Intensity、API) = $aX + b$

aは、スケーリング係数であり、Xは、元のピクセル強度であり、bは、オフセット値であり、aXは、スケーリングされたピクセル強度である。

【0831】

その結果は、前処理された画像データ202pの一例でもある、増強された画像パッチ15026である。データ増強は、データ増強器15034によって動作可能である。

【0832】

図151は、図150のデータ正規化技術 (DeepRTA (norm)) 及びデータ増強技術 (DeepRTA (augment)) が、ニューラルネットワークベースのベースコーラー1514が細菌データで訓練され、ヒトデータで試験されるときにベースコール誤差割合を低減し、細菌データ及びヒトデータが、同じアッセイを共有する (例えば、両方ともイントロンデータを含む) ことを示す。

【0833】

図152は、図150のデータ正規化技術 (DeepRTA (norm)) 及びデータ増強技術 (DeepRTA (augment)) が、ニューラルネットワークベースのベースコーラー1514が非エキソンデータ (例えば、イントロンデータ) で訓練され、エキソンデータで試験されるときにベースコール誤差割合を低減することを示す。

【0834】

言い換えると、図150のデータ正規化及びデータ増強技術は、ニューラルネットワークベースのベースコーラー1514が、訓練中に見られないデータをより良好に一般化し、したがって、過剰適合を低減することを可能にする。

【0835】

一実施態様では、データ増強は、訓練中及び推測中の両方で適用される。別の実施態様では、データ増強は、訓練中にのみ適用される。更に別の実施態様では、データ増強は、推測中にのみ適用される。

【0836】

図147A及び図147Bは、配列決定システムの一実施態様を示す。配列決定システムは、構成可能なプロセッサを含む。

【0837】

図147Cは、ベースコールセンサー出力などの、配列決定システムからのセンサーデータの分析のためのシステムの簡略ブロック図である。

【0838】

図148Aは、ホストプロセッサによって実行される実行時プログラムの機能を含む、ベースコール動作の態様を示す簡略図である。

【0839】

図148Bは、図147Cに示されるものなどの構成可能なプロセッサの構成の簡略図である。

【0840】

図149は、本明細書に開示される技術を実施するために、図147Aの配列決定システムによって使用され得るコンピュータシステムである。

(配列決定システム)

【0841】

図147A及び図147Bは、配列決定システム14700Aの一実施態様を示す。配列決定システム14700Aは、構成可能プロセッサ14746を含む。構成可能プロセッサ14746は、本明細書に開示されるベースコール技術を実施する。配列決定システムは、「シーケンサ」とも称される。

【0842】

配列決定システム14700Aは、生物学的物質又は化学物質のうちの少なくとも1つに関連する任意の情報又はデータを得ることができる。いくつかの実施態様では、配列決定システム14700Aは、ベンチトップデバイス又はデスクトップコンピュータと同様であり得るワークステーションである。例えば、所望の反応を実施するためのシステム及び構成要素の大部分（又は全て）は、共通のハウジング14702内にあってもよい。

【0843】

特定の実施態様では、配列決定システム14700Aは、*de novo sequencing*、全ゲノム又は標的ゲノム領域の再配列、及びメタゲノミクスを含むがこれらに限定されない、様々な用途のために構成された核酸配列決定システムである。シーケンサはまた、DNA又はRNA分析に使用されてもよい。いくつかの実施態様では、配列決定システム14700Aはまた、バイオセンサー内に反応部位を生成するように構成されてもよい。例えば、配列決定システム14700Aは、サンプルを受容し、サンプル由来のクロノウイルス増幅核酸の表面結合クラスターを生成するように構成され得る。各クラスターは、バイオセンサー内の反応部位を構成するか、又はその一部であってもよい。

【0844】

例示的な配列決定システム14700Aは、バイオセンサー14712と相互作用して、バイオセンサー14712内で所望の反応を行うように構成されたシステム容器又はインターフェース14710を含んでもよい。図147Aに関して以下の説明では、バイオセンサー14712はシステム受け部14710内に装填される。しかしながら、バイオセンサー14712を含むカートリッジは、システム受け部14710に挿入されてもよく、一部の状態では、カートリッジは一時的又は永久的に除去され得ることが理解される。上述のように、カートリッジは、とりわけ、流体制御及び流体貯蔵構成要素を含んでもよい。

【0845】

特定の実施態様では、配列決定システム14700Aは、バイオセンサー14712内で多数の平行反応を行うように構成されている。バイオセンサー14712は、所望の反応が生じ得る1つ又はそれ以上の反応部位を含む。反応部位は、例えば、バイオセンサーの固体表面に固定化されてもよく、又はバイオセンサーの対応する反応チャンバ内に位置するビーズ（又は他の可動基材）に固定化されてもよい。反応部位は、例えば、クロノウイルス増幅核酸のクラスターを含むことができる。バイオセンサー14712は、固体撮像装置（例えば、CCD又はCMOSイメージャ）及びそれに取り付けられたフローセルを含んでもよい。フローセルは、配列決定システム14700Aから溶液を受容し、溶液を反応部位に向けて方向付ける1つ又はそれ以上の流路を含んでもよい。任意選択的に、バイオセンサー14712は、熱エネルギーを流路の内外に伝達するための熱要素と係合するように構成することができる。

【0846】

配列決定システム14700Aは、相互に相互作用して、生物学的又は化学的分析のための所定の方法又はアッセイプロトコルを実行する、様々な構成要素、アセンブリ、及びシステム（又はサブシステム）を含んでもよい。例えば、配列決定システム14700Aは、配列決定システム14700Aの様々な構成要素、アセンブリ、及びサブシステムと通信してもよく、またバイオセンサー14712も含むシステムコントローラ14706を含む。例えば、システム容器14710に加えて、配列決定システム14700Aはまた、配列決定システム14700Aの流体ネットワーク及びバイオセンサー14712の流体の流れを制御する流体制御システム14708と、バイオアッセイシステムによって

使用され得る全ての流体（例えば、気体又は液体）を保持する流体貯蔵システム 1 4 7 1 4 と、流体ネットワーク、流体貯蔵システム 1 4 7 1 4 及び／又はバイオセンサー 1 4 7 1 2 内の流体の温度を調節し得る温度制御システム 1 4 7 0 4 と、バイオセンサー 1 4 7 1 2 を照明するように構成された照明システム 1 4 7 1 6 と、を備えていてもよい。上述のように、バイオセンサー 1 4 7 1 2 を有するカートリッジがシステム容器 1 4 7 1 0 内に装填される場合、カートリッジはまた、流体制御及び流体貯蔵構成要素を含んでもよい。

【0 8 4 7】

また、配列決定システム 1 4 7 0 0 A は、ユーザーと対話するユーザーインターフェース 1 4 7 1 8 を含んでもよい。例えば、ユーザーインターフェース 1 4 7 1 8 は、ユーザーから情報を表示又は要求するディスプレイ 1 4 7 2 0 と、ユーザー入力を受け取るためのユーザー入力デバイス 1 4 7 2 2 とを含むことができる。いくつかの実施態様では、ディスプレイ 1 4 7 2 0 及びユーザー入力デバイス 1 4 7 2 2 は、同じデバイスである。例えば、ユーザーインターフェース 1 4 7 1 8 は、個々のタッチの存在を検出し、またディスプレイ上のタッチの場所を識別するように構成されたタッチ感知ディスプレイを含んでもよい。しかしながら、マウス、タッチパッド、キーボード、キーパッド、ハンドヘルドスキャナー、音声認識システム、動き認識システムなどの他のユーザー入力デバイス 1 4 7 2 2 が使用されてもよい。以下でより詳細に説明するように、配列決定システム 1 4 7 0 0 A は、所望の反応を実施するために、バイオセンサー 1 4 7 1 2 （例えば、カートリッジの形態）を含む様々な構成要素と通信してもよい。配列決定システム 1 4 7 0 0 A はまた、バイオセンサーから得られたデータを分析して、ユーザーに所望の情報を提供する

10

20

【0 8 4 8】

システムコントローラ 1 4 7 0 6 は、マイクロコントローラ、低減命令セットコンピュータ（Reduced Instruction Set Computer、RISC）、特定用途向け集積回路（Application Specific Integrated Circuit、ASIC）、フィールドプログラマブルゲートアレイ（Field Programmable Gate Array、FPGA）、粗粒化再構成可能構造（Coarse-Grained Reconfigurable Architecture、CGRA）、論理回路、及び本明細書に記載される機能を実行することができる任意の他の回路又はプロセッサと、を備える。上記の実施例は、例示的なものに過ぎず、したがって、システムコントローラという用語の定義及び／又は意味を制限することを意図するものではない。例示的な実施態様では、システムコントローラ 1 4 7 0 6 は、検出データを取得し分析する少なくとも 1 つのために、1 つ又はそれ以上の記憶要素、メモリ、又はモジュール内に記憶された命令のセットを実行する。検出データは、ピクセル信号の複数の配列を含むことができ、それにより、数百万個のセンサー（又はピクセル）のそれぞれからのピクセル信号の配列を、多くのベースコールサイクルにわたって検出することができる。記憶要素は、配列決定システム 1 4 7 0 0 A 内の情報源又は物理メモリ要素の形態であってもよい。

30

【0 8 4 9】

命令セットは、本明細書に記載される様々な実施態様の方法及びプロセスなどの特定の動作を実行するように配列決定システム 1 4 7 0 0 A 又はバイオセンサー 1 4 7 1 2 に指示する様々なコマンドを含んでもよい。命令のセットは、有形の非一時的コンピュータ可読媒体又は媒体の一部を形成し得るソフトウェアプログラムの形態であってもよい。本明細書で使用する時、用語「ソフトウェア」及び「ファームウェア」は互換可能であり、ランダムアクセスメモリ（Random Access Memory、RAM）メモリ、リードオンリーメモリ（Read Only Memory、ROM）メモリ、EPROMメモリ、EEPROMメモリ、及び不揮発性RAM（Non-Volatile RAM、NVRAM）メモリを含むコンピュータによって実行されるメモリに記憶された任意のコンピュータプログラムを含む。上記メモリタイプは、例示的なものに過ぎず、したがって、コンピュータプログラムの記憶に使用可能なメモリの種類に限定されない。

40

【0 8 5 0】

ソフトウェアは、システムソフトウェア又はアプリケーションソフトウェアなどの様々

50

な形態であってもよい。更に、ソフトウェアは、別個のプログラムの集合、又はより大きいプログラム内のプログラムモジュール若しくはプログラムモジュールの一部の形態であってもよい。ソフトウェアはまた、オブジェクト指向プログラミングの形態のモジュール式プログラミングを含んでもよい。検出データを取得した後、検出データは、ユーザー入力に応じて処理された配列決定システム14700Aによって自動的に処理されてもよく、又は別の処理マシン（例えば、通信リンクを介したリモート要求）によって行われる要求に応じて処理されてもよい。図示の別の実施態様では、システムコントローラ14706は分析モジュール14744を含む。他の別の実施態様では、システムコントローラ14706は分析モジュール14744を含まず、代わりに分析モジュール14744へのアクセスを有する（例えば、分析モジュール14744は、クラウド上で別個にホスティングされ得る）。

10

【0851】

システムコントローラ14706は、通信リンクを介して、バイオセンサー14712及び配列決定システム14700Aの他の構成要素に接続されてもよい。システムコントローラ14706はまた、オフサイトシステム又はサーバに通信可能に接続されてもよい。通信リンクは、配線、コード、又は無線であってもよい。システムコントローラ14706は、ユーザーインターフェース14718及びユーザー入力デバイス14722からユーザー入力又はコマンドを受信してもよい。

【0852】

流体制御システム14708は、流体ネットワークを含み、流体ネットワークを通る1つ又はそれ以上の流体の流れを方向付けるように構成されている。流体ネットワークは、バイオセンサー14712及び流体貯蔵システム14714と流体連通していてもよい。例えば、流体貯蔵システム14714から流体を選択し、制御された方法でバイオセンサー14712に向けてもよく、又は流体は、バイオセンサー14712から引き出され、例えば、流体貯蔵システム14714内の廃棄物リザーバに向けられてもよい。図示されていないが、流体制御システム14708は、流体ネットワーク内の流体の流量又は圧力を検出する流量センサーを含んでもよい。センサーは、システムコントローラ14706と通信してもよい。

20

【0853】

温度制御システム14704は、流体ネットワーク、流体貯蔵システム14714及び／又はバイオセンサー14712の異なる領域における流体の温度を調節するように構成されている。例えば、温度制御システム14704は、バイオセンサー14712と相互作用し、バイオセンサー14712内の反応部位に沿って流れる流体の温度を制御する熱循環器を含んでもよい。温度制御システム14704はまた、配列決定システム14700A又はバイオセンサー14712の中実要素又は構成要素の温度を調節してもよい。図示されていないが、温度制御システム14704は、流体又は他の構成要素の温度を検出するためのセンサーを含んでもよい。センサーは、システムコントローラ14706と通信してもよい。

30

【0854】

流体貯蔵システム14714は、バイオセンサー14712と流体連通しており、所望の反応を行うために使用される様々な反応成分又は反応物質を貯蔵してもよい。流体貯蔵システム14714はまた、流体ネットワーク及びバイオセンサー14712を洗浄又は洗浄し、反応物質を希釈するための流体を貯蔵してもよい。例えば、流体貯蔵システム14714は、試料、試薬、酵素、他の生体分子、緩衝液、水性、及び非極性溶液などを保存するための様々なリザーバを含んでもよい。更に、流体貯蔵システム14714はまた、バイオセンサー14712から廃棄物を受容するための廃棄物リザーバを含んでもよい。カートリッジを含む実施態様形態では、カートリッジは、流体貯蔵システム、流体制御システム、又は温度制御システムのうちの1つ又はそれ以上を含み得る。したがって、これらのシステムに関する本明細書に記載される構成要素のうちの1つ又はそれ以上は、カートリッジハウジング内に収容され得る。例えば、カートリッジは、サンプル、試薬、酵

40

50

素、他の生体分子、緩衝液、水性、及び非極性溶液、廃棄物などを保存するための様々なリザーバを有し得る。したがって、流体貯蔵システム、流体制御システム、又は温度制御システムのうちの1つ又はそれ以上は、カートリッジ又は他のバイオセンサーを介してバイオアッセイシステムと取り外し可能に係合され得る。

【0855】

照明システム14716は、バイオセンサーを照明するための光源（例えば、1つ又はそれ以上の発光ダイオード（Light-Emitting Diode、LED））及び複数の光学構成要素を含んでもよい。光源の例としては、レーザー、アークランプ、LED、又はレーザーダイオードが挙げられる。光学部品は、例えば、反射器、偏光板、ビームスプリッタ、コリマ、レンズ、フィルタ、ウェッジ、プリズム、鏡、検出器などであってもよい。照明システムを使用する実施態様では、照明システム14716は、励起光を反応部位に向けるように構成されてもよい。一例として、蛍光団は、緑色の光の波長によって励起されてもよく、そのため、励起光の波長は約532nmであり得る。一実施態様では、照明システム14716は、バイオセンサー14712の表面の表面法線に平行な照明を生成するように構成されている。別の実施態様では、照明システム14716は、バイオセンサー14712の表面の表面法線に対してオフアングルである照明を生成するように構成されている。更に別の実施態様では、照明システム14716は、いくつかの平行照明及びある程度のオフアングル照明を含む複数の角度を有する照明を生成するように構成されている。

【0856】

システム容器又はインターフェース14710は、機械的、電氣的、及び流体的な方法のうちの少なくとも1つにおいてバイオセンサー14712と係合するように構成される。システム受け部14710は、バイオセンサー14712を所望の配向に保持して、バイオセンサー14712を通る流体の流れを容易にすることができる。システム受け部14710はまた、バイオセンサー14712と係合するように構成された電気接点を含んでもよく、それにより、配列決定システム14700Aは、バイオセンサー14712と通信してもよく、及び／又はバイオセンサー14712に電力を供給することができる。更に、システム容器14710は、バイオセンサー14712と係合するように構成された流体ポート（例えば、ノズル）を含んでもよい。いくつかの実施態様では、バイオセンサー14712は、電氣的に、また流体方式で、システム受け部14710に取り外し可能に連結される。

【0857】

加えて、配列決定システム14700Aは、他のシステム若しくはネットワークと遠隔で、又は他のバイオアッセイシステム14700Aと通信してもよい。バイオアッセイシステム14700Aによって得られた検出データは、リモートデータベースに記憶されてもよい。

【0858】

図147Bは、図147Aのシステムで使用することができるシステムコントローラ14706のブロック図である。一実施態様では、システムコントローラ14706は、互いに通信することができる1つ又はそれ以上のプロセッサ又はモジュールを含む。プロセッサ又はモジュールのそれぞれは、特定のプロセスを実行するためのアルゴリズム（例えば、有形及び／又は非一時的コンピュータ可読記憶媒体上に記憶された命令）又はサブアルゴリズムを含んでもよい。システムコントローラ14706は、モジュールの集合として概念的に例示されるが、専用ハードウェアボード、DSP、プロセッサなどの任意の組み合わせを利用して実装されてもよい。あるいは、システムコントローラ14706は、単一のプロセッサ又は複数のプロセッサを備えた既製のPCを使用して実装されてもよく、機能動作はプロセッサ間に分散される。更なる選択肢として、以下に記載されるモジュールは、特定のモジュール式機能が専用ハードウェアを利用して実施されるハイブリッド構成を利用して実装されてもよく、残りのモジュール式機能は、既製のPCなどを利用して実施される。モジュールはまた、処理ユニット内のソフトウェアモジュールとして実装されてもよい。

【0859】

動作中、通信ポート14750は、バイオセンサー14712（図147A）及び／又はサブシステム14708、14714、14704（図147A）から情報（例えば、データ）に情報（例えば、コマンド）を送信してもよい。実施態様形態では、通信ポート14750は、ピクセル信号の複数の配列を出力することができる。通信リンク14734は、ユーザーインターフェース14718からユーザー入力を受信し（図147A）、ユーザーインターフェース14718にデータ又は情報を送信してもよい。バイオセンサー14712又はサブシステム14708、14714、14704からのデータは、バイオアッセイセッション中に、システムコントローラ14706によってリアルタイムで処理されてもよい。追加的に又は代替的に、データは、バイオアッセイセッション中にシステムメモリ内に一時的に記憶され、リアルタイム又はオフライン操作よりも遅く処理されてもよい。

10

【0860】

図147Bに示すように、システムコントローラ14706は、中心処理装置（Central Processing Unit、CPU）14752と共に主制御モジュール14724と通信する複数のモジュール14726-14748を含んでもよい。主制御モジュール14724は、ユーザーインターフェース14718と通信してもよい（図147A）。モジュール14726-14748は、主制御モジュール14724と直接通信するものとして示されているが、モジュール14726-14748はまた、互いに、ユーザーインターフェース14718と、及びバイオセンサー14712と直接通信してもよい。また、モジュール14726-14748は、他のモジュールを介して主制御モジュール14724と通信してもよい。

20

【0861】

複数のモジュール14726-14748は、サブシステム14708、14714、14704及び14716とそれぞれ通信するシステムモジュール14728-14732、14726を含む。流体制御モジュール14728は、流体ネットワークを通る1つ又はそれ以上の流体の流れを制御するために、流体制御システム14708と通信して、流体ネットワークの弁及び流量センサーを制御してもよい。流体貯蔵モジュール14730は、流体が低い場合、又は廃棄物リザーバが容量又はその近くにあるときにユーザーに通知することができる。流体貯蔵モジュール14730はまた、流体が所望の温度で貯蔵され得るように、温度制御モジュール14732と通信してもよい。照明モジュール14726は、所望の反応（例えば、結合事象）が生じた後など、プロトコル中に指定された時間で反応部位を照明するために、照明システム14716と通信してもよい。いくつかの実施態様では、照明モジュール14726は、照明システム14716と通信して、指定された角度で反応部位を照明することができる。

30

【0862】

複数のモジュール14726-14748はまた、バイオセンサー14712と通信する装置モジュール14736と、バイオセンサー14712に関連する識別情報を判定する識別モジュール14738とを含んでもよい。装置モジュール14736は、例えば、システム容器14710と通信して、バイオセンサーが配列決定システム14700Aとの電氣的及び流体的接続を確立したことを確認することができる。識別モジュール14738は、バイオセンサー14712を識別する信号を受信してもよい。識別モジュール14738は、バイオセンサー14712の識別情報を使用して、他の情報をユーザーに提供してもよい。例えば、識別モジュール14738は、ロット番号、製造日、又はバイオセンサー14712で動作することが推奨されるプロトコルを決定し、その後表示してもよい。

40

【0863】

複数のモジュール14726-14748はまた、バイオセンサー14712から信号データ（例えば、画像データ）を受信及び分析する分析モジュール14744（信号処理モジュール又は信号プロセッサとも呼ばれる）も含む。分析モジュール14744は、検

50

出／画像データを記憶するためのメモリ（例えば、RAM又はフラッシュ）を含む。検出データは、ピクセル信号の複数の配列を含むことができ、それにより、数百万個のセンサー（又はピクセル）のそれぞれからのピクセル信号の配列を、多くのベースコールサイクルにわたって検出することができる。信号データは、その後の分析のために記憶されてもよく、又はユーザーインターフェース14718に送信されて、所望の情報をユーザーに表示することができる。いくつかの実施態様では、信号データは、分析モジュール14744が信号データを受信する前に、固体撮像素子（例えば、CMOS画像センサー）によって処理され得る。

【0864】

分析モジュール14744は、複数の配列決定サイクルのそれぞれにおいて、光検出器から画像データを取得するように構成される。画像データは、光検出器によって検出される発光信号から導出され、ニューラルネットワークベースの品質スコアラ－6102及び／又はニューラルネットワークベースのベースコーラ－218を介して、複数の配列決定サイクルの各々に対して画像データを処理し、複数の配列決定サイクルの各々で検体のうちの少なくともいくつかに対するベースコールを生成する。光検出器は、1つ又はそれ以上のオーバーヘッドカメラ（例えば、バイオセンサー14712上のクラスターの画像を上から撮影するIlluminaのGAIIxのCCDカメラ）の一部であってもよく、又はバイオセンサー14712自体の一部（例えば、バイオセンサー14712上のクラスターの下にあり、底部からのクラスターの画像を取るIlluminaのiSeqのCMOS画像センサー）であってもよい。

【0865】

光検出器の出力は、それぞれクラスターの強度放射及びそれらの周囲背景を示す配列決定画像である。配列決定画像は、配列決定中に配列にヌクレオチドを組み込む結果として生成される強度放射を示す。強度放射は、関連する検体及びそれらの周囲の背景からのものである。配列決定画像は、メモリ14748に記憶される。

【0866】

プロトコルモジュール14740及び14742は、メイン制御モジュール14724と通信して、所定のアッセイプロトコルを実施する際にサブシステム14708、14714及び14704の動作を制御する。プロトコルモジュール14740及び14742は、所定のプロトコルに従って特定の動作を実行するように配列決定システム14700Aに指示するための命令セットを含み得る。図示のように、プロトコルモジュールは、配列決定ごとの合成プロセスを実行するための様々なコマンドを発行するように構成された、配列合成（Sequencing-By-Synthesis、SBS）モジュール14740であってもよい。SBSにおいて、核酸テンプレートに沿った核酸プライマーの伸長を監視して、テンプレート中のヌクレオチド配列を決定する。下にある化学プロセスは、（例えば、ポリメラーゼ酵素により触媒される）又はライゲーション（例えば、リガーゼ酵素により触媒される）であり得る。特定のポリマー系SBSの実施態様では、プライマーに付加されるヌクレオチドの順序及び種類の検出を使用してテンプレートの配列を決定することができるように、蛍光標識ヌクレオチドをテンプレート依存様式でプライマー（それによってプライマーを伸長させる）に添加する。例えば、第1のSBSサイクルを開始するために、1つ又はそれ以上の標識されたヌクレオチド、DNAポリメラーゼなどを、核酸テンプレートのアレイを収容するフローセル内に／それを介して送達することができる。核酸テンプレートは、対応する反応部位に位置してもよい。プライマー伸長が、組み込まれる標識ヌクレオチドを、撮像事象を通して検出することができる、これらの反応部位が検出され得る。撮像イベントの間、照明システム14716は、反応部位に励起光を提供することができる。任意に、ヌクレオチドは、ヌクレオチドがプライマーに添加されると、更なるプライマー伸長を終結する可逆終端特性を更に含むことができる。例えば、可逆的ターミネーター部分を有するヌクレオチド類似体をプライマーに添加して、デブロッキング剤が部分を除去するためにデブロッキング剤が送達されるまで続く伸長が生じ得ない。したがって、可逆終端を使用する別の実施態様では、フローセル（検出前又は検出後）にデブロッ

10

20

30

40

50

キング試薬を送達するために、コマンドを与えることができる。1つ又はそれ以上のコマンドは、様々な送達工程間の洗浄（複数可）をもたらすために与えられ得る。次いで、サイクルをn回繰り返してプライマーをn個のヌクレオチドで伸長させることができ、それによって長さnの配列を検出する。例示的な配列決定技術は、例えば、Bentley et al., Nature 456:53-59 (200147)、国際公開第04/01147497号、米国特許第7,057,026号明細書、国際公開第91/0667147号、同第07/123744号、米国特許第7,329,492号明細書、同第7,211,414号明細書、同第7,315,019号明細書、米国特許第7,405,21471号明細書、及び同第200147/0147014701472号（それぞれ参照により本明細書に組み込まれる）に記載されている。

10

【0867】

SBSサイクルのヌクレオチド送達工程では、単一の種類のヌクレオチドのいずれかを一度に送達することができ、又は複数の異なるヌクレオチドタイプ（例えば、A、C、T、及びG）を送達することができる。一度に単一の種類のヌクレオチドのみが存在するヌクレオチド送達構成では、異なるヌクレオチドは、個別化された送達に固有の時間的分離に基づいて区別することができるため、異なるヌクレオチドは別個の標識を有する必要はない。したがって、配列決定方法又は装置は、単一の色検出を使用することができる。例えば、励起源は、単一の波長又は単一の波長範囲の励起のみを提供する必要がある。ある時点で、送達がフローセル内に存在する複数の異なるヌクレオチドをもたらすヌクレオチド送達構成では、異なるヌクレオチドタイプを組み込む部位は、混合物中のそれぞれのヌクレオチドタイプに結合された異なる蛍光標識に基づいて区別することができる。例えば、4つの異なる蛍光団のうちの1つをそれぞれ有する4つの異なるヌクレオチドを使用することができる。一実施態様では、4つの異なるフルオロフォアは、スペクトルの4つの異なる領域における励起を使用して区別することができる。例えば、4つの異なる励起放射線源を使用することができる。あるいは、4つ未満の異なる励起源を使用することができるが、単一源からの励起放射線の光学的濾過を使用して、フローセルにおいて異なる励起放射線の範囲を生成することができる。

20

【0868】

いくつかの実施態様では、4つ未満の異なる色を、4つの異なるヌクレオチドを有する混合物中で検出することができる。例えば、ヌクレオチドの対は、同じ波長で検出することができるが、対のうちの1つのメンバーに対する強度の差に基づいて、又は、対の他の部材について検出された信号と比較して明らかなシグナルを出現又は消失させる、対の1つのメンバーへの変化（例えば、化学修飾、光化学修飾、又は物理的改質を行うことを介して）に基づいて区別され得る。4個未満の色の検出を使用して4個の異なるヌクレオチドを区別するための例示的な装置及び方法が、例えば、米国特許出願第61/53147,294号明細書及び同第61/619,1477147号明細書に記載されており、それらの全体が参照により本明細書に組み込まれる。2012年9月21日出願された米国特許出願第13/624,200号は、その全体が参照により組み込まれる。

30

【0869】

複数のプロトコルモジュールはまた、バイオセンサー14712内の製品を増幅するための流体制御システム14708及び温度制御システム14704にコマンドを発行するように構成された試料調製（又は生成）モジュール14742を含んでもよい。例えば、バイオセンサー14712は、配列決定システム14700Aに係合されてもよい。増幅モジュール14742は、バイオセンサー14712内の反応チャンバに必要な増幅成分を送達するために、流体制御システム14708に命令を発行することができる。他の実施態様では、反応部位は、テンプレートDNA及び／又はプライマーなどの増幅のためのいくつかの成分を既に含有していてもよい。増幅成分を反応チャンバに送達した後、増幅モジュール14742は、既知の増幅プロトコルに従って異なる温度段階を通して温度制御システム14704にサイクルするように指示し得る。いくつかの実施態様では、増幅及び／又はヌクレオチドの取り込みは、等温的に実施される。

40

50

【0870】

SBSモジュール14740は、クローン性アンプリコンのクラスターがフローセルのチャンネル内の局所領域上に形成されるブリッジPCRを実行するコマンドを発行することができる。ブリッジPCRを介してアンプリコンを生成した後、アンプリコンを「線形化」して、一本鎖テンプレートDNAを作製してもよく、sstDNA及び配列決定プライマーは、関心領域に隣接する普遍配列にハイブリダイズされてもよい。例えば、合成方法による可逆的ターミネーター系配列決定を、上記のように又は以下のように使用することができる。

【0871】

各塩基性コーリング又は配列決定サイクルは、例えば、修飾DNAポリメラーゼ及び4種類のヌクレオチドの混合物を使用することによって達成することができる単一の塩基によってsstDNAを延長することができる。異なる種類のヌクレオチドは、固有の蛍光標識を有することができ、各ヌクレオチドは、各サイクルにおいて単一塩基の組み込みのみが生じることを可能にする可逆的ターミネーターを更に有し得る。SstDNA、励起光に単一の塩基を添加した後、反応部位に入射し、蛍光発光を検出することができる。検出後、蛍光標識及びターミネーターは、sstDNAから化学的に切断され得る。別の同様の基本コーリング又は配列決定サイクルは、以下の通りであってもよい。そのような配列決定プロトコルでは、SBSモジュール14740は、バイオセンサー14712を通る試薬及び酵素溶液の流れを方向付けるように流体制御システム14708に指示することができる。本明細書に記載される装置及び方法と共に利用することができる例示的な可逆性ターミネーターベースのSBS方法は、米国特許出願公開第2007/0166705(A1)号、米国特許出願公開第2006/01147147901(A1)号、米国特許第7,057,026号、米国特許出願公開第2006/0240439(A1)号明細書、米国特許出願公開第2006/0214714714709(A1)号明細書、国際公開第05/014914714号、米国特許出願公開第2005/014700900(A1)号明細書、国際公開第06/0147B199号及び国際公開第07/01470251号(それぞれ参照によりその全体が本明細書に組み込まれる)に記載されている。可逆性ターミネーター系SBSの例示的な試薬は、米国特許第7,541,444号、米国特許第7,057,026号明細書、同第7,414,14716号明細書、同第7,427,673号明細書、同第7,566,537号明細書、同第7,592,435号明細書、及び国際公開第07/141473536147号に記載されており、これらはそれぞれ参照によりその全体が本明細書に組み込まれる。

【0872】

いくつかの実施態様では、増幅及びSBSモジュールは、単一のアッセイプロトコルで動作してもよく、例えば、テンプレート核酸は増幅され、続いて同じカートリッジ内で配列される。

【0873】

配列決定システム14700Aはまた、ユーザーがアッセイプロトコルを再構成することを可能にし得る。例えば、決定システム14700Aは、決定されたプロトコルを修正するために、ユーザーインターフェース14718を通じてユーザーにオプションを提供することができる。例えば、バイオセンサー14712が増幅のために使用されると判定された場合、配列決定システム14700Aは、アニーリングサイクルの温度を要求し得る。更に、配列決定システム14700Aは、選択されたアッセイプロトコルに対して一般的に許容されないユーザー入力をユーザーが提供した場合に、ユーザーに警告を発行し得る。

【0874】

実施態様形態では、バイオセンサー14712は、センサー(又はピクセル)のミリオンを含み、それらのそれぞれは、連続するベースコールサイクルにわたって複数のピクセル信号の配列を生成する。分析モジュール14744は、センサーのアレイ上のセンサーの行方向及び/又は列方向の位置に従って、ピクセル信号の複数の配列を検出し、それら

を対応するセンサー（又はピクセル）に属させる。

【0875】

図147Cは、ベースコールセンサー出力などの、配列決定システム14700Aからのセンサーデータの分析のためのシステムの簡略ブロック図である。図147Cの例では、システムは、構成可能プロセッサ14746を含む。構成可能プロセッサ14746は、中心処理ユニット（CPU）14752（すなわち、ホストプロセッサ）によって実行される実行時プログラムと協調して、ベースコーラー（例えば、ニューラルネットワークベースの品質スコアラ－6102及び／又はニューラルネットワークベースのベースコーラー218）を実行することができる。配列決定システム14700Aは、バイオセンサー14712及びフローセルを含む。フローセルは、遺伝物質のクラスターが、遺伝物質中の塩基を識別するためにクラスター内の反応を引き起こすために使用される一連の検体フローに曝露される1つ又はそれ以上のタイルを含み得る。センサーは、タイルデータを提供するために、フローセルの各タイルにおける配列の各サイクルの反応を検知する。遺伝的配列決定はデータ集約的操作であり、このデータ集約的操作は、ベースコールセンサーデータを、ベースコール動作中に感知された各遺伝物質群のベースコールの配列に変換する。

【0876】

本実施例のシステムは、ベースコール動作を調整するための実行時プログラムを実行するCPU14752と、タイルデータのアレイの配列を記憶するメモリ14748Bと、ベースコール動作によって生成されたベースコールリードと、ベースコール動作で使

【0877】

配列決定システム14700Aは、バス14789によって構成可能プロセッサ14746に結合される。バス14789は、PCI-SIG規格（PCI Special Interest Group）によって現在維持及び開発されているPCIe規格（Peripheral Component Interconnect Express）と互換性のあるバス技術などの高スループット技術を使用して実装することができる。また、この例では、メモリ14748Aは、バス14793によって構成可能プロセッサ14746に結合される。メモリ14748Aは、構成可能プロセッサ14746を有する回路基板上に配置されたオンボードメモリであってもよい。メモリ14748Aは、ベースコール動作で使用される作業データの構成可能プロセッサ14746による高速アクセスに使用される。バス14793はまた、PCIe規格と互換性のあるバス技術などの高スループット技術を使用して実装することもできる。

【0878】

フィールドプログラマブルゲートアレイFPGA、粗いグレー構成可能な再構成可能アレイCGRAs、並びに他の構成可能かつ再構成可能なデバイスを含む構成可能なプロセッサは、コンピュータプログラムを実行する汎用プロセッサを使用して達成され得るよりも、より効率的に又はより高速に様々な機能を実装するように構成することができる。構成可能なプロセッサの構成は、時にはビットストリーム又はビットファイルと称される構成ファイルを生成するために機能的な説明を編集することと、構成ファイルをプロセッサ上の構成可能要素に配布することと、を含む。構成ファイルは、データフローパターンを設定するように回路を構成することにより、分散メモリ及び他のオンチップメモリリソースの使用、ルックアップテーブルコンテンツ、構成可能な論理ブロックの動作、及び構成可能な論理ブロックの動作、及び構成可能なアレイの構成可能な相互接続及び他の要素のような構成可能な実行ユニットとを含む。構成ファイルがフィールド内で変更され得る場

合、ロードされた構成ファイルを変更することによって構成ファイルを変更することができる場合に再構成可能である。例えば、構成ファイルは、揮発性SRAM要素内に、不揮発性読み書きメモリ素子内に記憶されてもよく、構成可能又は再構成可能なプロセッサ上の構成可能要素のアレイ間に分散されたものであってもよい。様々な市販の構成可能なプロセッサは、本明細書に記載されるようなベースコール動作において使用するのに好適である。例としては、Googleのテンソル処理ユニット(TPU)(商標)、GX4 Rackmount Series(商標)、GX9 Rackmount Series(商標)、NVIDIA DGX-1(商標)、Microsoft' Stratix V FPGA(商標)、GraphcoreのIntelligent Processor Unit(IPU)(商標)、QualcommのZeroth Platform(商標)(Snapdragon processors(商標)、NVIDIA Volta(商標)、NVIDIAのドライブPX(商標)、NVIDIAのJETSON TX1/TX2 MODULE(商標)、Intel's Nirvana™、Movidius VPU(商標)、Fujitsu DPI(商標)、アームDynamic IQ(商標)、IBM TrueNorth(商標)、Lambda GPU Server with Testa V100s(商標)、Xilinx Alveo(商標) U200、Xilinx Alveo(商標) U250、Xilinx Alveo(商標) U280、Intel/Altera Stratix(商標) GX2800、Intel/Altera Stratix(商標) GX2800、及びIntel Stratix(商標) GX10M、が含まれる。いくつかの実施例では、ホストCPUは、構成可能プロセッサと同じ集積回路上に実装することができる。

【0879】

本明細書に記載される実施形態は、構成可能プロセッサ14746を使用して、ニューラルネットワークベースの品質スコアラー6102及び/又はニューラルネットワークベースのベースコーラー218を実施する。構成可能プロセッサ14746の構成ファイルは、高レベルの記述言語HDL又はレジスタ転送レベルRTL言語仕様を使用して実行されるロジック機能を指定することによって実装することができる。本明細書は、選択された構成可能プロセッサが構成ファイルを生成するように設計されたリソースを使用してコンパイルすることができる。構成可能なプロセッサではない場合がある特定用途向け集積回路の設計を生成する目的で、同じ又は類似の仕様をコンパイルすることができる。

【0880】

したがって、本明細書に記載される全ての実施形態における構成可能プロセッサ構成可能プロセッサ14746の代替例は、特定用途向けASIC又は専用集積回路又は集積回路のセットを含む構成されたプロセッサを含み、又はシステムオンチップSOCデバイス、又は本明細書に記載されるようなニューラルネットワークベースのベースコール動作を実行するように構成された、システムオンチップSOCデバイス、又はグラフィック処理ユニット(GPU)プロセッサ若しくは粗粒化再構成可能構造(Coarse-Grained Reconfigurable Architecture、CGRA)プロセッサである。

【0881】

一般に、ニューラルネットワークの動作を実行するように構成された、本明細書に記載の構成可能なプロセッサ及び構成されたプロセッサは、本明細書ではニューラルネットワークプロセッサと称される。

【0882】

構成可能プロセッサ14746は、この例では、CPU14752によって実行されるプログラムを使用して、又は構成可能要素14791のアレイを構成する他のソースによってロードされた構成ファイルによって構成される(例えば、構成論理ブロック(Configuration Logic Block、CLB)、例えばルックアップテーブル(Look Up Table、LUT)、フリップフロップ、演算処理ユニット(PMU)、及び計算メモリユニット(Compute Memory Unit、CMU)、構成可能なI/Oブロック、プログラマブル相互接続)を計算して、ベースコール機能を実行する。この例では、構成は、バス14789及び1

4793に結合され、ベースコール動作で使用される要素間でデータ及び制御パラメータを分散する機能を実行するデータフローロジック14797を含む。

【0883】

また、構成可能プロセッサ14746は、ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218を実行するために、ベースコール実行ロジック14797を用いて構成されている。論理14797は、マルチサイクル実行クラスター（例えば、14779）を含み、この実施例では、実行クラスターXを介した実行クラスター1を含む。多重サイクル実行クラスターの数、動作の所望のスループットを伴うトレードオフ、及び構成可能プロセッサ14746上の利用可能なリソースに従って選択することができる。

10

【0884】

多重サイクル実行クラスターは、構成可能なプロセッサ14746上の構成可能な相互接続及びメモリリソースを使用して実装されるデータ流路14799によってデータフローロジック14797に結合される。また、マルチサイクル実行クラスターは、構成可能な相互接続及びメモリリソースを使用して、例えば構成可能プロセッサ14746上で実施される制御経路14795によってデータフローロジック14797に結合されている。それは、利用可能な実行クラスターを示す制御信号を提供し、ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の動作の実行のための入力ユニットを利用可能な実行クラスターに提供する準備ができており、ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の訓練されたパラメータを提供する準備ができており、ベースコール分類データの出力パッチ、並びにニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の実行に使用される他の制御データを提供する準備ができています。

20

【0885】

構成可能プロセッサ14746は、訓練されたパラメータを使用してニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の動作を実行して、ベースコール動作の検知サイクルに関する分類データを生成するように構成されている。ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の動作を実行して、ベースコール動作の被験者検知サイクルの分類データを生成する。ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の動作は、N個の検知サイクルのそれぞれの検知サイクルからのタイルデータのアレイの数Nを含む配列で動作し、N個の検知サイクルは、本明細書に記載される例での時間配列における動作ごとの1つの塩基位置に対する異なるベースコール動作のためのセンサーデータを提供する。任意選択的に、N個の検知サイクルのうちのいくつかは、実行される特定のニューラルネットワークモデルに従って必要に応じて、配列から出ることができる。数Nは、1を超える任意の数であり得る。本明細書に記載されるいくつかの実施例では、N個の検知サイクルの検知サイクルは、被験者の検知サイクルに先行する少なくとも1つの検知サイクル、及び被験者サイクルの後の少なくとも1回の検知サイクルについての検知サイクルのセットを表す。本明細書では、数Nが5以上の整数である、実施例が記載される。

30

40

【0886】

データフローロジック14797は、N個のアレイの空間的に位置合わせされたパッチのタイルデータを含む所与の動作のための入力ユニットを使用して、ニューラルネットワークベースの品質スコアラ6102及び／又はニューラルネットワークベースのベースコーラ218の動作のために、メモリ14748Aから構成可能プロセッサ14746に、タイルデータ及びモデルパラメータの少なくともいくつかの訓練されたパラメータを移動させるように構成されている。入力ユニットは、1回のDMA動作におけるダイレクトメモリアクセス動作によって、又は、配備されたニューラルネットワークの実行と協調

50

して、利用可能なタイムスロットの間に移動するより小さいユニット内で移動させることができる。

【0887】

本明細書に記載される感知サイクルのタイルデータは、1つ又はそれ以上の特徴を有するセンサーデータのアレイを含むことができる。例えば、センサーデータは、DNA、RNA、又は他の遺伝物質の遺伝的配列における塩基位置で4塩基のうちの1つを識別するために分析される2つの画像を含むことができる。タイルデータはまた、画像及びセンサーに関するメタデータを含むことができる。例えば、ベースコール動作の実施形態では、タイルデータは、タイル上の遺伝物質群の中心からのセンサーデータのアレイ内の各ピクセルの距離を示す中心情報からの距離などの、クラスターとの画像の位置合わせに関する情報を含むことができる。

10

【0888】

以下に記載されるように、ニューラルネットワークベースの品質スコアラ-6102及び／又はニューラルネットワークベースのベースコーラ-218の実行中に、タイルデータはまた、ニューラルネットワークベースの品質スコアラ-6102及び／又はニューラルネットワークベースのベースコーラ-218の実行中に生成されるデータも含み得る。それは、ニューラルネットワークベースの品質スコアラ-6102及び／又はニューラルネットワークベースのベースコーラ-218の動作中に再計算されるのではなく再使用することができる中間データと称される。例えば、ニューラルネットワークベースの品質スコアラ-6102及び／又はニューラルネットワークベースのベースコーラ-218の実行中に、データフローロジック14797は、タイルデータのアレイの所与のパッチのセンサーデータの代わりに、中間データをメモリ14748Aに書き込むことができる。このような実施形態は、以下により詳細に記載される。

20

【0889】

図示されているように、ベースコール動作の検知サイクルからタイルのセンサーデータを含むタイルデータを記憶する実行時プログラムによってアクセス可能なメモリ（例えば、14748A）を含む、ベースコールセンサー出力の分析のためのシステムが説明される。また、システムは、メモリへのアクセスを有する構成可能プロセッサ14746などのニューラルネットワークプロセッサを含む。ニューラルネットワークプロセッサは、訓練されたパラメータを使用してニューラルネットワークの動作を実行して、検知サイクルのための分類データを生成するように構成される。本明細書に記載されるように、ニューラルネットワークの動作は、被験者サイクルを含むN個の感知サイクルのそれぞれの感知サイクルからタイルデータのN個のアレイの配列で動作して、被験者サイクルの分類データを生成する。データフローロジック14797は、N個の検知サイクルのそれぞれの検知サイクルからのN個のアレイの空間的に位置合わせされたパッチのデータを含む入力ユニットを使用して、ニューラルネットワークの動作のために、メモリからニューラルネットワークプロセッサにタイルデータ及び訓練されたパラメータを移動させるために提供される。

30

【0890】

また、ニューラルネットワークプロセッサがメモリへのアクセスを有し、複数の実行クラスターを含み、ニューラルネットワークを実行するように構成された複数の実行クラスター内の実行クラスターを含むシステムも説明される。データフローロジック14797は、メモリへのアクセス、及び複数の実行クラスター内のクラスターを実行して、複数の実行クラスター内の利用可能な実行クラスターにタイルデータの入力ユニットを提供し、入力ユニットは、それぞれの感知サイクルからタイルデータのアレイの空間的に整列されたパッチの数Nを含む、入力ユニットと、被験者検知サイクルを含み、N個の空間的に整合されたパッチをニューラルネットワークに適用して、被験者検知サイクルの空間的に整合されたパッチの分類データの出力パッチを生成させるように、実行クラスターに、ニューラルネットワークにN個の空間的に位置合わせされたパッチを適用させることと、を含み、Nは1より大きい。

40

50

【0891】

図148Aは、ホストプロセッサによって実行される実行時プログラムの機能を含む、ベースコール動作の態様を示す簡略図である。この図では、フローセルからの画像センサーの出力は、ライン14800上で画像処理スレッド14801に提供され、画像処理スレッド8001は、個々のタイルのセンサーデータのアレイ内の位置合わせ及び配置、及び画像の再サンプリングなどの画像上のプロセスを実行することができ、フローセル内の各タイルのタイルクラスターマスクを計算するプロセスによって使用することができ、フローセルの対応するタイル上の遺伝子材料のクラスターに対応するセンサーデータのアレイ内のピクセルを識別するプロセスによって使用することができる。画像処理スレッド14801の出力は、CPU内のディスパッチロジック14810に、ライン14802上に提供され、それは、高速バス14803上又は高速バス14805上のデータキャッシュ14804（例えば、SSD記憶装置）に、ベースコール動作の状態に従って、図147Cの構成可能プロセッサ14746などのニューラルネットワークプロセッサハードウェア14820にタイルデータのアレイを転送する。処理され、変換された画像は、以前に使用されたサイクルを検知するために、データキャッシュ14804上に記憶され得る。ハードウェア14820は、ニューラルネットワークによって出力された分類データをディスパッチロジック148148に返し、ディスパッチロジック8080は、情報をデータキャッシュ14804に、又はライン14811上でスレッド14802に渡し、分類データを使用してベースコール及び品質スコア計算を実行し、ベースコール読み取りのための標準フォーマットでデータを配置することができる。ベースコール及び品質スコア計算を実行するスレッド14802の出力は、ベースコールリードを集約するスレッド14803に、ライン14812上に提供され、データ圧縮などの他の動作を実行し、結果として得られるベースコール出力を顧客による利用のために指定された宛先に書き込む。

【0892】

いくつかの実施形態では、ホストは、ニューラルネットワークを支持するハードウェア14820の出力の最終処理を実行する、スレッド（図示せず）を含むことができる。例えば、ハードウェア14820は、マルチクラスターニューラルネットワークの最終層から分類データの出力を提供することができる。ホストプロセッサは、ベースコール及び品質スコアスレッド14802によって使用されるデータを設定するために、分類データを超えて、ソフトマックス関数などの出力起動機能を実行することができる。また、ホストプロセッサは、ハードウェア14820への入力前のタイルデータのバッチ正規化などの入力動作（図示せず）を実行することができる。

【0893】

図148Bは、図147Cの構成などの構成可能プロセッサ14746の構成の簡略図である。図148Bでは、構成可能プロセッサ14746は、複数の高速PCIeインターフェースを有するFPGAを含む。FPGAは、図147Cを参照して説明されるデータフローロジック14797を含むラッパー14890を用いて構成されている。ラッパー14890は、CPU通信リンク14877を介してCPU内の実行時プログラムとのインターフェース及び調整を管理し、DRAM通信リンク14897を介してオンボードDRAM14899（例えば、メモリ14748A）との通信を管理する。ラッパー14890内のデータフローロジック14797は、数Nのサイクルのために、オンボードDRAM14899上のタイルデータのアレイをクラスター14885まで横断することによって取得されたパッチデータを提供し、クラスター14885からプロセスデータ14887を取得して、オンボードDRAM14899に配信する。ラッパー14890はまた、タイルデータの入力アレイ、及び分類データの出力パッチの両方について、オンボードDRAM14899とホストメモリとの間のデータの転送を管理する。ラッパーは、ライン14883上のパッチデータを割り当てられたクラスター14885に転送する。ラッパーは、オンボードDRAM14899から取得されたクラスター14885にライン14881の重みやバイアスなどの訓練されたパラメータを提供する。ラッパーは、CPU通信リンク14877を介してホスト上の実行時プログラムから提供されるか、又はそ

れに応答して生成されるクラスター14885に、ライン14879上の構成及び制御データを提供する。クラスターはまた、ホストからの制御信号と協働して使用されて、空間的に整列したパッチデータを提供し、クラスター14885のリソースを使用して、パッチデータを介して多重サイクルニューラルネットワークをパッチデータの上で実行するために、ホストからの制御信号と協働して使用されるラッパー14890に、ライン14889上の状態信号を提供することができる。

【0894】

上述のように、タイルデータの複数のパッチのうちの対応するパッチ上で実行するように構成されたラッパー14890によって管理される単一の構成可能なプロセッサ上に複数のクラスターが存在し得る。各クラスターは、本明細書に記載される複数の感知サイクルのタイルデータを使用して、被験者検知サイクルにおけるベースコールの分類データを提供するように構成することができる。

【0895】

システムの例では、フィルタ重み及びバイアスのようなカーネルデータを含むモデルデータをホストCPUから構成可能プロセッサに送信することができ、その結果、モデルは、サイクル数の関数として更新され得る。ベースコール動作は、代表的な例では、数百の感知サイクルの順序で含むことができる。ベースコール動作は、いくつかの実施形態では、ペアリングされた端部読み取りを含むことができる。例えば、モデル訓練されたパラメータは、20サイクルごと（又は他の数のサイクル）ごとに、又は特定のシステム及びニューラルネットワークモデルに実装される更新パターンに従って更新されてもよい。タイル上の遺伝的クラスター内の所与のストリングのための配列が、ストリングの第1の端部から下方に（又は上方に）延在する第1の部分と、ストリングの第2の端部から上方に（又は下方に）延在する第2の部分とを含む、ペアリングされた端部リードを含むいくつかの実施形態では、訓練されたパラメータは、第1の部分から第2の部分への遷移で更新され得る。

【0896】

いくつかの実施例では、タイルのための感知データの複数サイクルの画像データは、CPUから包装材14890に送信され得る。ラッパー14890は、任意選択的に、感知データの一部の前処理及び変換を行い、その情報をオンボードDRAM14899に書き込むことができる。各感知サイクルの入力タイルデータは、タイル当たり4000 x 3000ピクセル／タイル以上を含むセンサーデータのアレイを含むことができ、2つの特徴はタイルの2つの画像の色を表し、1ピクセル当たり1つ又は2つのバイトを含むセンサーデータのアレイを含むことができる。数Nが、多重サイクルニューラルネットワークの各動作において使用される3回の検知サイクルである実施形態では、多重サイクルニューラルネットワークの各動作のためのタイルデータのアレイは、数当たり数百メガバイトの数で消費することができる。システムのいくつかの実施形態では、タイルデータはまた、タイルごとに1回記憶されたDFCデータのアレイ、又はセンサーデータ及びタイルに関する他のタイプのメタデータも含む。

【0897】

動作中、多重サイクルクラスターが利用可能である場合、ラッパーは、パッチをクラスターに割り当てる。ラッパーはタイルの横断面にタイルデータの次のパッチをフェッチし、適切な制御及び構成情報と共に割り当てられたクラスターに送信する。クラスターは、構成可能プロセッサ上の十分なメモリを用いて構成されて、パッチを含むデータのパッチを、定位置に処理されているいくつかのシステム内で複数サイクルから保持するのに十分なメモリを有するように構成することができ、様々な実施形態では、ピンポンバッファ技術又はラスタ走査技術を使用して処理される。

【0898】

割り当てられたクラスターが、現在のパッチのニューラルネットワークのその動作を完了し、出力パッチを生成すると、それはラッパーに信号を送る。ラッパーは、割り当てられたクラスターから出力パッチを読み出すか、あるいは割り当てられたクラスターは、デ

10

20

30

40

50

ータをラッパーにプッシュする。次いで、ラッパーは、DRAM 14899内の処理されたタイルのための出力パッチを組み立てる。タイル全体の処理が完了し、データの出力パッチがDRAMに転送されると、ラッパーは、処理された出力アレイを、特定のフォーマットでホスト／CPUに返送する。いくつかの実施形態では、オンボードDRAM 14899は、ラッパー14890内のメモリ管理論理によって管理される。実行時プログラムは、リアルタイム分析を提供するために連続フローで動作する全てのサイクルについての全てのタイルデータのアレイの分析を完了するために、配列決定動作を制御することができる。

(コンピュータシステム)

【0899】

図149は、本明細書に開示される技術を実施するために配列決定システム800Aによって使用され得るコンピュータシステム14900である。コンピュータシステム14900は、バスサブシステム14955を介して多数の周囲デバイスと通信する、少なくとも1つの中心処理装置(CPU)14972を含む。これらの周囲デバイスは、例えば、メモリデバイス及びファイルストレージサブシステム14936、ユーザーインターフェース入力デバイス14938、ユーザーインターフェース出力デバイス14976及びネットワークインターフェースサブシステム14974を含む記憶サブシステム14910を含むことができる。入力及び出力デバイスは、コンピュータシステム14900とのユーザー対話を可能にする。ネットワークインターフェースサブシステム14974は、他のコンピュータシステム内の対応するインターフェースデバイスへのインターフェースを含む外部ネットワークへのインターフェースを提供する。

【0900】

一実施態様では、システムコントローラ7806は、記憶サブシステム14910及びユーザーインターフェース入力デバイス14938に通信可能にリンクされている。

【0901】

ユーザーインターフェース入力デバイス14938は、キーボードと、マウス、トラックボール、タッチパッド、又はグラフィックスタブレットなどのポインティングデバイスと、スキャナと、ディスプレイに組み込まれたタッチスクリーンと、音声認識システム及びマイクロフォンなどのオーディオ入力デバイスと、他の種類の入力デバイスと、を含むことができる。一般に、用語「入力デバイス」の使用は、コンピュータシステム14900に情報を入力するための全ての可能な種類のデバイス及び方法を含むことを意図する。

【0902】

ユーザーインターフェース出力デバイス14976は、ディスプレイサブシステム、プリンタ、ファックス装置、又はオーディオ出力デバイスなどの非視覚ディスプレイを含むことができる。ディスプレイサブシステムは、LEDディスプレイ、陰極線管(Cathode Ray Tube、CRT)、液晶ディスプレイ(Liquid Crystal Display、LCD)などのフラットパネルデバイス、投影デバイス、又は可視画像を作成するための何らかの他の機構を含むことができる。ディスプレイサブシステムはまた、音声出力デバイスなどの非視覚ディスプレイを提供することができる。一般に、用語「出力装置」の使用は、コンピュータシステム14900からユーザー又は別のマシン若しくはコンピュータシステムに情報を出力するための、全ての可能な種類のデバイス及び方法を含むことを意図する。

【0903】

記憶サブシステム14910は、本明細書に記載されるモジュール及び方法のうちのいくつか又は全ての機能を提供するプログラミング及びデータ構築物を記憶する。これらのソフトウェアモジュールは、概して、深層学習プロセッサ14978によって実行される。

【0904】

深層学習プロセッサ14978は、グラフィック処理ユニット(GPU)、フィールドプログラマブルゲートアレイ(FPGA)、特定用途向け集積回路(ASIC)、及び／又は粗粒化再構成可能構造(CGRA)であり得る。深層学習プロセッサ14978は、Google Cloud Platform(商標)、Xilinx(商標)及びCi

10

20

30

40

50

rrascale (商標) などの深層学習クラウドプラットフォームによってホスティングすることができる。深層学習プロセッサ14978の例は、GoogleのTensor Processing Unit (TPU) (商標)、GX4 Rackmount Series (商標)、GX149 Rackmount Series (商標) のようなラックマウントソリューション、NVIDIA DGX-1 (商標)、Microsoft's Stratix V FPGA (商標)、GraphcoreのIntelligent Processor Unit (IPU) (商標)、Snapdragon processors (商標) を有するQualcommのZeroth Platform (商標)、NVIDIAのVolta (商標)、NVIDIAのDRIVE PX (商標)、NVIDIAのJETSON TX1/TX2 MODULE (商標)、IntelのNirvana (商標)、Movidius VPU (商標)、Fujitsu DPI (商標)、ARMのDynamicIQ (商標)、IBM TrueNorth (商標)、Testa V100s (商標) を有するLambda GPU Server、及び他のものを含む。
【0905】

10

記憶サブシステム14910で使用されるメモリサブシステム14922は、プログラム実行中に命令及びデータを記憶するためのメインランダムアクセスメモリ (RAM) 14932と、固定命令が記憶された読み取り専用メモリ (ROM) 14934とを含む多数のメモリを含むことができる。ファイル記憶サブシステム14936は、プログラム及びデータファイルのための永続的な記憶装置を提供することができ、ハードディスクドライブ、関連する取り外し可能な媒体、ドライブ、光学ドライブ、又は取り外し可能な媒体カートリッジを含むことができる。特定の実施態様の機能を実装するモジュールは、ストレージサブシステム14910内のファイル記憶サブシステム14936によって、又はプロセッサによってアクセス可能な他のマシン内に記憶され得る。

20

【0906】

バスサブシステム14955は、コンピュータシステム14900の様々な構成要素及びサブシステムを、意図されるように互いに通信するための機構を提供する。バスサブシステム14955は、単一のバスとして概略的に示されているが、バスサブシステムの代替実施態様は、複数のバスを使用することができる。

【0907】

コンピュータシステム14900自体は、パーソナルコンピュータ、ポータブルコンピュータ、ワークステーション、コンピュータ端末、ネットワークコンピュータ、テレビ、メインフレーム、サーバファーム、緩く分散した一組の緩くネットワーク化されたコンピュータ、又は任意の他のデータ処理システム若しくはユーザーデバイスを含む様々なタイプのものであり得る。コンピュータ及びネットワークの変化の性質により、図149に示されるコンピュータシステム14900の説明は、本発明の好ましい実施態様を例示する目的のための特定の例としてのみ意図される。コンピュータシステム14900の多くの他の構成は、図149に示されるコンピュータシステムよりも多く又は少ない構成要素を有することができる。

30

(配列決定プロセス)

【0908】

本明細書に記載される実施態様は、配列変動を識別するために核酸配列を分析することに適用可能であり得る。実施態様は、遺伝子位置／遺伝子座の潜在的な変異体／対立遺伝子を分析し、遺伝子座の遺伝子型を決定するか、又は言い換えると、遺伝子座に対する遺伝子型コールを提供するために使用され得る。一例として、核酸配列は、米国特許出願公開第2016/0085910号及び米国特許出願公開第2013/0296175号に記載されている方法及びシステムに従って分析され得、その完全な主題は、それらの全体で参照により本明細書に明示的に組み込まれる。

40

【0909】

一実施態様では、配列決定プロセスは、DNAなどの核酸を含むか、又はそれを含むと思われるサンプルを受容することを含む。サンプルは、動物 (例えば、ヒト)、植物、細

50

菌、又は真菌などの既知の又は未知のソースからであり得る。サンプルは、ソースから直接取られ得る。例えば、血液又は唾液は、個体から直接取られ得る。代替的に、サンプルは、ソースから直接取得されない場合がある。次いで、1つ又はそれ以上のプロセッサは、配列決定のためにサンプルを調製するようにシステムに指示する。調製は、外来材料を除去すること、及び／又は特定の材料（例えば、DNA）を単離することを含み得る。生物サンプルは、特定のアッセイのための特徴を含むように調製され得る。例えば、生物サンプルは、配列合成（SBS）のために調製され得る。特定の実施態様では、調製は、ゲノムの特定の領域の増幅を含み得る。例えば、調製は、STR及び／又はSNPを含むことが知られている所定の遺伝子座を増幅することを含み得る。遺伝子座は、所定のプライマー配列を使用して増幅され得る。

10

【0910】

次に、1つ又はそれ以上のプロセッサは、サンプルを配列決定するようにシステムに指示する。配列決定は、様々な既知の配列決定プロトコルを介して実行され得る。特定の実施態様では、配列決定は、SBSを含む。SBSでは、光学基板の表面（例えば、フローセル内のチャンネルを少なくとも部分的に画定する表面）上に存在する増幅DNAの複数のクラスター（おそらく、数百万個のクラスター）を配列決定するために、複数の蛍光標識ヌクレオチドが使用される。フローセルは、フローセルが適切なフローセルホルダー内に配置される配列決定のための核酸サンプルを含み得る。

【0911】

核酸は、未知のターゲット配列に隣接する既知のプライマー配列を含むように調製され得る。第1のSBS配列決定サイクルを開始するために、1つ又はそれ以上の異なる標識ヌクレオチド及びDNAポリメラーゼなどが、流体フローサブシステムによってフローセル内に／それを介して流され得る。単一のタイプのヌクレオチドが一度に追加され得るか、又は配列決定手順で使用されるヌクレオチドが可逆終端特性を有するように特別に設計され得、したがって、配列決定反応の各サイクルが、いくつかのタイプの標識ヌクレオチド（例えば、A、C、T、G）の存在下で同時に生じることを可能にする。ヌクレオチドは、フルオロフォアなどの検出可能な標識部分を含み得る。4個のヌクレオチドと一緒に混合される場合、ポリメラーゼは、正しい塩基を選択して組み込むことができ、各配列は、単一の塩基によって伸長される。非組み込みヌクレオチドは、フローセルを介して洗浄溶液を流すことによって洗い流され得る。1つ又はそれ以上のレーザーは、核酸を励起し、蛍光を誘発し得る。核酸から放出される蛍光は、組み込み塩基のフルオロフォアに基づいており、異なるフルオロフォアは、異なる波長の発光を放出し得る。伸長及び検出されたDNA鎖から可逆終端基を除去するために、デブロッキング試薬がフローセルに追加され得る。次いで、デブロッキング試薬は、フローセルを介して洗浄溶液を流すことによって洗い流され得る。次いで、フローセルは、上記の標識ヌクレオチドの導入で開始して、配列決定の更なるサイクルの準備ができている。流体及び検出動作は、配列決定動作を完了させるために数回繰り返され得る。例示的配列決定方法は、例えば、Bentley et al., Nature 456: 53-59 (2008)、国際公開第WO04/018497号、米国特許第7,057,026号明細書、国際公開第WO91/06678号、国際公開第WO07/123744号、米国特許第7,329,492号明細書、米国特許第7,211,414号明細書、米国特許第7,315,019号明細書、米国特許第7,405,281号明細書、及び米国特許出願公開第2008/0108082号に記載されており、それぞれ参照により本明細書に組み込まれる。

20

30

40

【0912】

いくつかの実施態様では、核酸は、配列決定前又は配列決定中に、表面に結合され、増幅され得る。例えば、増幅は、ブリッジ増幅を使用して実行されて、表面上に核酸クラスターを形成し得る。有用なブリッジ増幅方法は、例えば、米国特許第5,641,658号明細書、米国特許出願公開第2002/0055100号、米国特許第7,115,400号明細書、米国特許出願公開第2004/0096853号、米国特許出願公開第2004/0002090号、米国特許出願公開第2007/0128624号、及び米国

50

特許出願公開第2008/0009420号に記載されており、これらはそれぞれ、その全体が参照により本明細書に組み込まれる。表面上の核酸を増幅するための別の有用な方法は、例えば、Lizardi et al., Nat. Genet. 19:225-232 (1998) 及び米国特許出願公開第2007/0099208 (A1) 号明細書に記載されているようなローリングサークル増幅 (Rolling Circle Amplification, RCA) であり、これらはそれぞれ、参照により本明細書に組み込まれる。

【0913】

一例示的SBSプロトコルは、例えば、国際公開第WO04/018497号、米国特許出願公開第2007/0166705号、及び米国特許第7,057,026号明細書に記載されているように、取り外し可能な3'ブロックを有する改変されたヌクレオチドを活用し、これらはそれぞれ、参照により本明細書に組み込まれる。例えば、SBS試薬の繰り返しサイクルは、例えば、ブリッジ増幅プロトコルの結果として、それに結合されたターゲット核酸を有するフローセルに送達され得る。核酸クラスターは、直鎖化溶液を使用して、単鎖形態に変換され得る。直鎖化溶液は、例えば、各クラスターの単鎖を切断することができる制限エンドヌクレアーゼを含み得る。開裂の他の方法は、制限酵素又はニックング酵素に対する代替として使用され得、とりわけ、化学的開裂 (例えば、過ヨウ素酸を用いたジオール結合の開裂)、熱又はアルカリに曝露することによる、エンドヌクレアーゼを用いた開裂による脱塩基部位の開裂 (例えば、NEB, Ipswich, Mass., USA 部品番号M5505Sによって供給されるような「USER」)、別の方法では、デオキシリボヌクレオチド、光化学的開裂、又はペプチドリンカーの開裂で構成された増幅産物に組み込まれるリボヌクレオチドの開裂を含む。直鎖化動作の後、配列決定プライマーは、配列決定されるターゲット核酸への配列決定プライマーのハイブリダイゼーションのための条件下でフローセルに送達され得る。

【0914】

次いで、フローセルは、単一のヌクレオチド付加によって各ターゲット核酸にハイブリダイズされるプライマーを伸長させるための条件下で、取り外し可能な3'ブロック及び蛍光標識を有する改変されたヌクレオチドを有するSBS伸長試薬と接触され得る。改変されたヌクレオチドが、配列決定されているテンプレートの領域に相補的な成長したポリヌクレオチド鎖に組み込まれると、更なる配列伸長を誘導するために利用可能な遊離の3'-OH基が存在せず、したがってポリメラーゼは、更なるヌクレオチドを付加することができないため、単一のヌクレオチドのみが、各プライマーに付加される。SBS伸長試薬は、除去され、放射線での励起下でサンプルを保護する構成要素を含む走査試薬と置き換えられ得る。走査試薬の例示的構成要素は、米国特許出願公開第2008/0280773 (A1) 号及び米国特許出願第13/018,255号に記載されており、それぞれ参照により本明細書に組み込まれる。次いで、伸長された核酸は、走査試薬の存在下で蛍光検出され得る。蛍光が検出されると、3'ブロックは、使用されるブロック基に適切なデブロック試薬を使用して除去され得る。それぞれのブロック基に有用な例示的デブロック試薬は、国際公開第004018497号、米国特許出願公開第2007/0166705 (A1) 号、及び米国特許第7,057,026号に記載されており、それぞれ参照により本明細書に組み込まれる。デブロック試薬は、更なるヌクレオチドの付加に現在適格である、3'-OH基を有する伸長されたプライマーにハイブリダイズされるターゲット核酸を残して洗い流され得る。したがって、伸長試薬、走査試薬、及びデブロック試薬の付加のサイクルは、1つ又はそれ以上の動作間での任意選択的な洗浄で、所望の配列が取得されるまで繰り返され得る。上記のサイクルは、改変されたヌクレオチドの各々が、特定の塩基に対応することが知られている、それに結合された異なるラベルを有するとき、サイクルごとの単一の伸長試薬送達動作を使用して実行され得る。異なる標識は、各組み込み動作中に付加されるヌクレオチド間の区別を容易にする。代替的に、各サイクルは、伸長試薬送達の別個の動作、続いて走査試薬送達及び検出の別個の動作を含み得、その場合、ヌクレオチドのうちの2つ又はそれ以上は、同じ標識を有し得、既知の送達の順序に基づい

10

20

30

40

50

て区別され得る。

【0915】

特定のSBSプロトコルに関して配列決定動作を上述したが、様々な他の分子分析のうちのいずれかを配列決定するための他のプロトコルが、望ましいように実行され得ることが理解されるであろう。

【0916】

次いで、システムの1つ又はそれ以上のプロセッサは、後続の分析に対する配列決定データを受信する。配列決定データは、BAMファイルなどの様々な方法でフォーマットされ得る。配列決定データは、例えば、多数のサンプルリードを含み得る。配列決定データは、ヌクレオチドの対応するサンプル配列を有する複数のサンプルリードを含み得る。1つのサンプルリードのみが論じられるが、配列決定データは、例えば、数百、数千、数十万、又は数百万個のサンプルリードを含み得ることを理解されたい。異なるサンプルリードは、異なる数のヌクレオチドを有し得る。例えば、サンプルリードは、10個のヌクレオチド〜約500個以上のヌクレオチドの範囲であり得る。サンプルリードは、ソース（複数可）の全体のゲノムに及び得る。一例として、サンプルリードは、疑わしいSTR又は疑わしいSNPを有する遺伝子座などの所定の遺伝子座に向けられる。

【0917】

各サンプルリードは、サンプル配列、サンプルフラグメント、又はターゲット配列と称され得るヌクレオチドの配列を含み得る。サンプル配列は、例えば、プライマー配列、ランキング配列、及びターゲット配列を含み得る。サンプル配列内のヌクレオチドの数は、30、40、50、60、70、80、90、100以上を含み得る。いくつかの実施態様では、1つ又はそれ以上のサンプルリード（又はサンプル配列）は、少なくとも150個のヌクレオチド、200個のヌクレオチド、300個のヌクレオチド、400個のヌクレオチド、500個以上のヌクレオチドを含む。いくつかの実施態様では、サンプルリードは、1000個よりも多くのヌクレオチド、2000個以上のヌクレオチドを含み得る。サンプルリード（又はサンプル配列）は、一方又は両方の末端にプライマー配列を含み得る。

【0918】

次に、1つ又はそれ以上のプロセッサは、配列決定データを分析して、潜在的な変異体コール（複数可）及びサンプル変異体コール（複数可）のサンプル変異体頻度を取得する。動作はまた、変異体コールアプリケーション又は変異体コーラーと称され得る。したがって、変異体コーラーは、変異体を識別又は検出し、変異体分類子は、検出された変異体を体細胞又は生殖細胞系として分類する。代替的な変異体コーラーは、本明細書での実施態様に従って利用され得、異なる変異体コーラーは、対象であるサンプルの特徴などに基づいて、実行されている配列決定動作のタイプに基づいて使用され得る。Illumina Inc. (San Diego, CA) によるPisces（商標）アプリケーションなどの、変異体コールアプリケーションの1つの非限定的な例は、<https://github.com/Illumina/Pisces>でホスティングされ、論文Dunn, Tamsen & Berry, Gwenn & Emig-Agius, Dorothea & Jiang, Yu & Iyer, Anita & Udar, Nitin & Stromberg, Michael. (2017). Pisces: An Accurate and Versatile Single Sample Somatic and Germline Variant Caller. 595-595. 10.1145/3107411.3108203に記載されており、その完全な主題は、その全体で参照により本明細書に明示的に組み込まれる。

【0919】

そのような変異体コールアプリケーションは、4つの連続的に実行されたモジュールを含み得る。

【0920】

(1) Pisces Read Stitcher: BAM中のペアのリード（同じ分子

10

20

30

40

50

のリード1及びリード2)をコンセンサスリードにステッチすることによって、ノイズを低減する。出力は、ステッチされたBAMである。

【0921】

(2) *Pisces Variant Caller*: 小さいSNV、挿入及び欠失をコールする。*Pisces*は、リード境界、基本フィルタリングアルゴリズム、及び単純ポアソンベースの変異体信頼性スコアリングアルゴリズムによって破壊される変異体を合体させるための変異体崩壊アルゴリズムを含む。出力は、VCFである。

【0922】

(3) *Pisces Variant Quality Recalibrator (VQR)*: 変異体コールが、熱損傷又はFFPE脱アミノ化に関連するパターンに過度に続く場合、VQR工程は、疑わしい変異体コールの変異体Qスコアを低下させる。出力は、調整されたVCFである。

【0923】

(4) *Pisces Variant Phaser (Scylla)*: 小変異体をクローンサブ集団からの複合対立遺伝子に組み立てるために、リードバックされたグリーディクスタリング方法を使用する。これは、下流ツールによる機能的結果のより正確な決定を可能にする。出力は、調整されたVCFである。

【0924】

追加的に又は代替的に、動作は、<https://github.com/Illumina/strelka>でホスティングされ、論文T Saunders, Christopher & Wong, Wendy & Swamy, Sajani & Becq, Jennifer & J Murray, Lisa & Cheetham, Keira. (2012)に記載されている、Illumina Inc.による変異体コールアプリケーション*Strelka* (商標) アプリケーションを利用し得る。*Strelka*: 配列決定された腫瘍-正常サンプルペアからの正確な体細胞小変異体コール。Bioinformatics (Oxford, England). 28. 1811-7. 10. 1093/bioinformatics/bts271、その完全な主題は、その全体で参照により本明細書に明示的に組み込まれる。更に、追加的に又は代替的に、動作は、<https://github.com/Illumina/strelka>でホスティングされ、論文Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Kallberg, M., Chen, X., Beyter, D., Krusche, P., and Saunders, C. T. (2017)に記載されている、Illumina Inc.による変異体コールアプリケーション*Strelka2* (商標) アプリケーションを利用し得る。*Strelka2*: 臨床配列決定アプリケーションのための高速で正確な変異体コール、その完全な主題は、その全体で参照により本明細書に明示的に組み込まれる。更に、追加的に又は代替的に、動作は、<https://github.com/Illumina/Nirvana/wiki>でホスティングされ、論文Stromberg, Michael & Roy, Rajat & Lajugie, Julien & Jiang, Yu & Li, Haochen & Margulies, Elliott. (2017)に記載されている、Illumina Inc.による*Nirvana* (商標) アプリケーションなどの変異体アノテーション/コールツールを利用し得る。*Nirvana: Clinical Grade Variant Annotator*. 596-596. 10. 1145/3107411. 3108204、その完全な主題は、その全体で参照により本明細書に明示的に組み込まれる。

【0925】

そのような変異体アノテーション/コールツールは、*Nirvana*に開示されているものなどの異なるアルゴリズム技術を適用し得る。

【0926】

a. 間隔アレイを有する全ての重複する転写物の識別: 機能的アノテーションに対して、我々は、変異体と重複する全ての転写物を識別し得、間隔ツリーが使用され得る。しか

10

20

30

40

50

しながら、間隔のセットが静的であり得るため、我々は、間隔アレイにそれを更に最適化することができた。間隔ツリーは、全ての重複する転写物を $O(\min(n, k \lg n))$ 時間で戻し、 n は、ツリーでの間隔の数であり、 k は、重複する間隔の数である。実際には、 k は、ほとんどの変異体について n と比較して非常に小さいため、間隔ツリー上の有効な実行時は、 $O(k \lg n)$ である。我々は、全ての間隔が、ソートされたアレイで記憶されている間隔アレイを生成することによって、 $O(\lg n + k)$ に改善し、その結果、我々は、第1の重複する間隔を見出す必要のみあり、次いで、残りの $(k - 1)$ を介して列挙する。

【0927】

b. CNVs / SVs (Yu) : コピー数変動及び構造変異体に対するアノテーションが提供され得る。小さい変異体のアノテーションと同様に、構造変異体 (Structural Variant, SV) と重複する転写物、及び以前に報告された構造変異体も、オンラインデータベースで注釈付けされ得る。小さい変異体とは異なり、あまりにも多くの転写物が大きいSVと重複するため、重複する転写物の全てが注釈付けされる必要があるというわけではない。その代わりに、部分的な重複する遺伝子に属する全ての重複する転写物は、注釈付けされ得る。具体的には、これらの転写物に対して、影響のあるイントロン、エキソン、及び構造変異体によって引き起こされる結果が報告され得る。全ての重複する転写物の出力を可能にするための選択肢が利用可能であるが、遺伝子シンボルなどのこれらの転写物に対する基本情報が報告され得、それが正規の重複であるか又は転写物との部分的な重複であるかどうかをフラグ付けし得る。各SV / CNVに対して、これらの変異体が研究されているかどうか、及び異なる集団でのそれらの頻度を知ることに対象である。したがって、我々は、1000個のゲノム、DGV及びClinGenなどの外部データベースで重複するSVを報告した。任意のカットオフを使用してどのSVが重複しているかを決定することを回避するために、代わりに、全ての重複する転写物が使用され得、相互の重複が計算され得、すなわち、重複する長さがこれら2つのSVの長さの最小値で除算される。

【0928】

c. 補足アノテーションの報告：補足アノテーションは、小さい及び構造的な変異体 (SV) の2つのタイプである。SVは、間隔としてモデル化され得、重複するSVを識別するために上述の間隔アレイを使用し得る。小さい変異体は、ポイントとしてモデル化され、位置及び (任意選択的に) 対立遺伝子によって一致する。したがって、それらは、バイナリ検索のようなアルゴリズムを使用して検索される。補足アノテーションデータベースが非常に大きくなり得るため、はるかに小さいインデックスが生成されて、染色体位置をマッピングして、補足アノテーションが存在する場所をファイルする。インデックスは、位置を使用してバイナリ検索され得る (染色体位置及びファイル場所で構成された) オブジェクトのソートされたアレイである。インデックスサイズを小さく維持するために、複数の位置 (最大、特定の最大カウント) は、第1の位置に対する値及び後続の位置に対するデルタのみを記憶する1つのオブジェクトに圧縮される。我々は、バイナリ検索を使用するため、実行時は、 $O(\lg n)$ であり、 n は、データベース内の項目の数である。

【0929】

d. VEPキャッシュファイル

【0930】

e. 転写物データベース：転写物キャッシュ (キャッシュ) 及び補足データベース (Supplementary database, SAdb) ファイルは、転写物及び補足アノテーションなどのデータオブジェクトのシリアル化されたダンプである。我々は、キャッシュに対する我々のデータソースとして、Ensembl VEPキャッシュを使用する。キャッシュを生成するために、全ての転写物が間隔アレイに挿入され、アレイの最終状態がキャッシュファイル内に記憶される。したがって、アノテーション中、我々は、事前に計算された間隔アレイをロードし、その上で検索を実行する必要だけある。キャッシュがメモリでロードされ、検索が非常に高速 (上述) であるため、重複する転写物を見出すことがNirvanaで非常に迅速である (総実行時の1%未満までプロファイルされる?)。

【0931】

f. 補足データベース：S A d bに対するデータソースは、補足材料下でリストされる。小さい変異体に対するS A d bは、（参照名及び位置によって識別される）データベース内の各オブジェクトが全ての関連する補足アノテーションを保持するように、全てのデータソースのk方向マージによって生成される。データソースファイルの解析中に遭遇する問題は、N i r v a n aのホームページで詳細に文書化されている。メモリ使用を制限するために、S A インデックスのみがメモリにロードされる。このインデックスは、補足アノテーションのためのファイル場所の迅速なルックアップを可能にする。しかしながら、データがディスクからフェッチされなければならないため、補足アノテーションの追加は、（総実行時の～30%でプロファイルされる）N i r v a n aの最大ボトルネックとして識別されている。

10

【0932】

g. 結果及び配列オントロジー：N i r v a n aの機能的アノテーション（提供される時）は、配列オントロジー（Sequence Ontology、SO）（<http://www.sequenceontology.org/>）ガイドラインに従う。時々、我々は、現在のSOの問題を識別し、SOチームと協働して、アノテーションの状態を改善するための機会を有した。

【0933】

そのような変異体アノテーションツールは、前処理を含み得る。例えば、N i r v a n aは、E x A C、E V S、1000 Genomes project、dbSNP、C l i n V a r、C o s m i c、D G V、及びC l i n G e nのような外部データソースからの多数のアノテーションを含んだ。これらのデータベースを完全に使用するために、我々は、それらからの情報をサニタイズする必要がある。我々は、異なるデータソースから存在する異なる競合に対処するための異なる戦略を実施した。例えば、同じ位置及び代替的な対立遺伝子に対する複数のdbSNPエントリーの場合、我々は、全てのIDをIDのカンマ区切りのリストに結合し、同じ対立遺伝子に対する異なるC A F値を有する複数のエントリーが存在する場合、我々は、第1のC A F値を使用する。競合するE x A C及びE V Sエントリーに対して、我々は、サンプルカウントの数を考え、より高いサンプルカウントを有するエントリーが使用される。1000 Genome Projectsでは、我々は、競合する対立遺伝子の対立遺伝子頻度を除去した。別の問題は、不正確な情報である。我々は、主に、1000 Genome Projectsから対立遺伝子頻度情報を抽出したが、我々は、G R C h 38に対して、情報フィールドで報告される対立遺伝子頻度は、利用できない遺伝子タイプを有するサンプルを除外せず、全てのサンプルに対して利用可能ではない変異体に対して収縮頻度をもたらすことに気づいた。我々のアノテーションの精度を保証するために、我々は、個々のレベルの遺伝子タイプの全てを使用して、真の対立遺伝子頻度を計算する。既知のように、同じ変異体は、異なる位置合わせに基づいて、異なる表現を有し得る。我々が、既に識別された変異体に対する情報を正確に報告できることを確認するために、我々は、異なるリソースから変異体を前処理して、それらに一貫した表現を有させる必要がある。全ての外部データソースに対して、我々は、参照対立遺伝子及び代替的な対立遺伝子の両方で、複製されたヌクレオチドを除去するために、対立遺伝子をトリミングした。C l i n V a rに対して、我々は、xmlファイルを直接解析し、多くの場合v c fファイルで使用される、全ての変異体に対して5個のプライム位置合わせを実行した。異なるデータベースは、同じセットの情報を含み得る。不必要な複製を回避するために、我々は、いくつかの複製された情報を除去した。例えば、我々は、より詳細な情報を有する1000人のゲノム内のこれらの変異体を既に報告したため、1000人ゲノムプロジェクトとしてデータソースを有するD G Vでの変異体を除去した。

20

30

40

【0934】

少なくともいくつかの実施態様によれば、変異体コールアプリケーションは、低周波数変異体、生殖細胞系コールなどに対するコールを提供する。非限定的な例として、変異体

50

コールアプリケーションは、腫瘍のみのサンプル及び／又は腫瘍－正常ペアのサンプル上で動作され得る。変異体コールアプリケーションは、単一ヌクレオチド変異（Single Nucleotide Variations、SNV）、複数ヌクレオチド変異（Multiple Nucleotide Variations、MNV）、インデルなどを検索し得る。変異体コールアプリケーションは、配列決定又はサンプル調製誤差に起因する不一致のためのフィルタリングを行っている間に、変異体を識別する。各変異体に対して、変異体コーラーは、参照配列、変異体の位置、及び潜在的な変異体配列（複数可）（例えば、A対C SNV、又はAG対A欠失）を識別する。変異体コールアプリケーションは、変異体が存在するという指標として、サンプル配列（又はサンプルフラグメント）、参照配列／フラグメント、及び変異体コールを識別する。変異体コールアプリケーションは、生フラグメントを識別し、生フラグメントの指定、潜在的な変異体コールを検証する生フラグメントの数のカウント、支持する変異体が生じた生フラグメント内の位置、及び他の関連する情報を出力し得る。生フラグメントの非限定的な例は、2重ステッチフラグメント、単純ステッチフラグメント、2重非ステッチフラグメント、及び単純非ステッチフラグメントを含む。

【0935】

変異体コールアプリケーションは、VCF又はGVCFFファイルなどの様々なフォーマットでコールを出力し得る。一例としてのみ、変異体コールアプリケーションは、MiSeq Reporterパイプライン（例えば、MiSeq（登録商標）シーケンサ器具上に実装されるとき）に含まれ得る。任意選択的に、アプリケーションは、様々なワークフローで実装され得る。分析は、所望の情報を取得するために指定された方法でサンプルリードを分析する、単一のプロトコル又はプロトコルの組み合わせを含み得る。

【0936】

次いで、1つ又はそれ以上のプロセッサは、潜在的な変異体コールに関連して検証動作を実行する。検証動作は、以下に説明するように、品質スコア及び／又は階層試験の階層に基づき得る。検証動作が、潜在的な変異体コールを認証又は検証するとき、検証動作は、（変異体コールアプリケーションからの）変異体コール情報をサンプル報告生成器に渡す。代替的に、検証動作が、潜在的な変異体コールを無効にするか、又は不適格にするとき、検証動作は、対応する指標（例えば、負のインジケータ、ノーコールインジケータ、無効なコールインジケータ）をサンプル報告生成器に渡す。検証動作はまた、変異体コールが正しいか、又は無効なコール指定が正しいという信頼度に関連する信頼性スコアを渡し得る。

【0937】

次に、1つ又はそれ以上のプロセッサは、サンプル報告を生成及び記憶する。サンプル報告は、例えば、サンプルに関して、複数の遺伝子座に関する情報を含み得る。例えば、所定のセットの遺伝子座の各遺伝子座に対して、サンプル報告は、遺伝子タイプコールを提供すること、遺伝子タイプコールが作製され得ないことを示すこと、遺伝子タイプコールの確実性に対する信頼性スコアを提供すること、又は1つ若しくはそれ以上の遺伝子座に関するアッセイで潜在的な問題を示すことのうちの少なくとも1つを行い得る。サンプル報告はまた、サンプルを提供した個体の性別を示し、及び／又はサンプルが複数のソースを含むことを示し得る。本明細書で使用するときに、「サンプル報告」は、遺伝子座若しくは遺伝子座の所定のセットのデジタルデータ（例えば、データファイル）、及び／又は遺伝子座若しくは遺伝子座のセットの印刷報告を含み得る。したがって、生成又は提供は、データファイルを生成すること、及び／若しくはサンプル報告を印刷すること、又はサンプル報告を表示することを含み得る。

【0938】

サンプル報告は、変異体コールが判定されたが、検証されなかったことを示し得る。変異体コールが無効であると判定されるとき、サンプル報告は、変異体コールを検証していない判定のための基準に関する追加の情報を示し得る。例えば、報告での追加の情報は、生フラグメントの説明と、生フラグメントが変異体コールを支持又は矛盾した範囲（例えば、カウント）と、を含み得る。追加的に又は代替的に、報告内の追加の情報は、本明細

書に記載される実施態様に従って取得される品質スコアを含み得る。

(変異体コールアプリケーション)

【0939】

本明細書に開示される実施態様は、配列決定データを分析して、潜在的な変異体コールを識別することを含む。変異体コールは、以前に実行された配列決定動作のために、記憶されたデータで実行され得る。追加的に又は代替的に、それは、配列決定動作が実行されている間に、リアルタイムで実行され得る。サンプルリードの各々は、対応する遺伝子座に割り当てられる。サンプルリードは、サンプルリードのヌクレオチドの配列、言い換えると、サンプルリード内のヌクレオチドの順序（例えば、A、C、G、T）に基づいて対応する遺伝子座に割り当てられ得る。この分析に基づいて、サンプルリードは、特定の遺伝子座の可能な変異体／対立遺伝子を含むものとして指定され得る。サンプルリードは、遺伝子座の可能な変異体／対立遺伝子を含むものとして指定された他のサンプルリードと共に収集（又は集約若しくはビンニング）され得る。割り当て動作はまた、サンプルリードが特定の遺伝子位置／座におそらく関連していると識別されるコール動作と称され得る。サンプルリードが分析されて、他のサンプルリードからサンプルリードを区別するヌクレオチドの1つ又はそれ以上の識別配列（例えば、プライマー配列）を配置し得る。より具体的には、識別配列（複数可）は、特定の遺伝子座に関連している他のサンプルリードからサンプルリードを識別し得る。

【0940】

割り当て動作は、識別配列の一連の n 個のヌクレオチドを分析して、識別配列の一連の n 個のヌクレオチドが選択配列のうちの1つ又はそれ以上と効果的に一致するかどうかを判定することを含み得る。特定の実施態様では、割り当て動作は、サンプル配列の第1の n 個のヌクレオチドを分析して、サンプル配列の第1の n 個のヌクレオチドが選択配列のうちの1つ又はそれ以上と効果的に一致するかどうかを判定することを含み得る。数 n は、様々な値を有し得、それは、プロトコルにプログラムされるか、又はユーザーによって入力され得る。例えば、数 n は、データベース内の最短選択配列のヌクレオチドの数として定義され得る。数 n は、所定の数であり得る。所定の数は、例えば、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、又は30個のヌクレオチドであり得る。しかしながら、他の実施態様では、より少ない又はより多くのヌクレオチドが使用され得る。数 n はまた、システムのユーザーなどの個人によって選択され得る。数 n は、1つ又はそれ以上の条件に基づき得る。例えば、数 n は、データベース内の最短プライマー配列のヌクレオチドの数、又は指定された数として定義され得、いずれの場合も、より小さい数である。いくつかの実施態様では、15個未満のヌクレオチドである任意のプライマー配列が例外として指定され得るように、15などの n の最小値が使用され得る。

【0941】

いくつかの場合では、識別配列の一連の n 個のヌクレオチドは、選択配列のヌクレオチドと正確に一致しない場合がある。それでもなお、識別配列は、識別配列が選択配列とほぼ同一である場合、選択配列と効果的に一致し得る。例えば、識別配列の一連の n 個のヌクレオチド（例えば、第1の n 個のヌクレオチド）が、不一致の指定された数（例えば、3）及び／又はシフトの指定された数（例えば、2）以下の選択配列と一致する場合、サンプルリードは、遺伝子座に対してコールされ得る。各不一致又はシフトが、サンプルリードとプライマー配列との間の差としてカウントし得るように、ルールが確立され得る。差の数が指定された数未満である場合、サンプルリードは、対応する遺伝子座に対してコールされ得る（すなわち、対応する遺伝子座に割り当てられ得る）。いくつかの実施態様では、サンプルリードの識別配列と遺伝子座に関連する選択配列との間の差の数に基づく一致スコアが決定され得る。一致スコアが、指定された一致閾値を超える場合、選択配列に対応する遺伝子座は、サンプルリードの潜在的な遺伝子座として指定され得る。いくつかの実施態様では、後続の分析が実行されて、サンプルリードが遺伝子座に対してコールされるかどうかを判定し得る。

【0942】

サンプルリードがデータベース内の選択配列のうちの1つと効果的に一致する（すなわち、上述のように正確に一致するか、又はほぼ一致する）場合、サンプルリードは、選択配列と相関する遺伝子座に割り当てられるか、又は指定される。これは、遺伝子座コール又は暫定遺伝子座コールと称され得、サンプルリードは、選択配列と相関する遺伝子座に対してコールされる。しかしながら、上述のように、サンプルリードは、1つを超える遺伝子座に対してコールされ得る。そのような実施態様では、更なる分析は、潜在的な遺伝子座のうちの1つのみに対してサンプルリードをコールするか、又は割り当てるために実行され得る。いくつかの実施態様では、参照配列のデータベースと比較されるサンプルリードは、ペアの末端配列決定からの第1のリードである。ペアの末端配列決定を実行するとき、サンプルリードと相関する（生フラグメントを表す）第2のリードが取得される。割り当ての後、割り当てられたリードで実行される後続の分析は、割り当てられたリードに対してコールされた遺伝子座のタイプに基づき得る。

10

【0943】

次に、サンプルリードが分析されて、潜在的な変異体コールを識別する。とりわけ、分析の結果は、潜在的な変異体コール、サンプル変異体頻度、参照配列、及び変異体が生じた、対象のゲノム配列内の位置を識別する。例えば、遺伝子座がSNPを含むことが知られている場合、遺伝子座に対してコールされた、割り当てられたリードは、割り当てられたリードのSNPを識別するために分析を受け得る。遺伝子座が多型反復DNA要素を含むことが知られている場合、割り当てられたリードは、サンプルリード内の多型反復DNA要素を識別又は特徴付けるために分析され得る。いくつかの実施態様では、割り当てられたリードがSTR遺伝子座及びSNP遺伝子座と効果的に一致する場合、警告又はフラグがサンプルリードに割り当てられ得る。サンプルリードは、STR遺伝子座及びSNP遺伝子座の両方として指定され得る。分析は、位置合わせプロトコルに従って、割り当てられたリードを位置合わせして、割り当てられたリードの配列及び／又は長さを決定することを含み得る。位置合わせプロトコルは、2013年3月15日に出版された国際特許出願第PCT/US2013/030867号（国際公開第WO2014/142831号）に記載される方法を含み得、その全体が参照により本明細書に組み込まれる。

20

【0944】

次いで、1つ又はそれ以上のプロセッサは、生フラグメントを分析して、支持変異体が生フラグメント内の対応する位置に存在するかどうかを判定する。様々なタイプの生フラグメントが識別され得る。例えば、変異体コーラーは、元の変異体コールを検証する変異体を示す、生フラグメントのタイプを識別し得る。例えば、生フラグメントのタイプは、2重ステッチフラグメント、単純ステッチフラグメント、2重非ステッチフラグメント、又は単純非ステッチフラグメントを表し得る。任意選択的に、他の生フラグメントは、前述の例の代わりに、又はそれに加えて識別され得る。各タイプの生フラグメントを識別することに関連して、変異体コーラーはまた、支持変異体が生じた生フラグメント内の位置、及び支持変異体を呈した生フラグメントの数のカウントを識別する。例えば、変異体コーラーは、生フラグメントの10個のリードが特定の位置Xで支持変異体を有する2重ステッチフラグメントを表すように識別されたという指標を出力し得る。変異体コーラーはまた、生フラグメントの5個のリードが特定の位置Yで支持変異体を有する単純非ステッチフラグメントを表すように識別されたという指標を出力し得る。変異体コーラーはまた、参照配列に対応する多数の生フラグメントを出力し得、したがって、対象のゲノム配列での潜在的な変異体コールを検証する証拠を別の方法で提供する支持変異体を含まなかった。

30

40

【0945】

次に、支持変異体、及び支持変異体が生じた位置を含む、生フラグメントのカウントが維持される。追加的に又は代替的に、（サンプルリード又はサンプルフラグメントでの潜在的な変異体コールの位置に対して）対象の位置で支持変異体を含まなかった生フラグメントのカウントが維持され得る。追加的に又は代替的に、参照配列に対応する生フラグメ

50

ントのカウントが維持され得、潜在的な変異体コールを認証又は確認しない。決定される情報は、変異体コール検証アプリケーションに出力され、潜在的な変異体コールを支持する生フラグメントのカウント及びタイプ、生フラグメントでの支持分散の位置、潜在的な変異体コールを支持しない生フラグメントのカウントなどを含む。

【0946】

潜在的な変異体コールが識別されるとき、プロセスは、潜在的な変異体コール、変異体配列、変異体位置、及びそれに関連する参照配列の指標を出力する。変異体コールは、誤差がコールプロセスに偽変異体を識別させ得るように、「潜在的な」変異体を表すように指定される。本明細書の実施態様によれば、潜在的な変異体コールが分析されて、偽変異体又は偽陽性を低減及び排除する。追加的に又は代替的に、プロセスは、サンプルリードに関連する1つ又はそれ以上の生フラグメントを分析し、生フラグメントに関連する、対応する変異体コールを出力する。

(技術的な改善及び用語)

【0947】

ベースコールは、蛍光標識されたタグを検体と共に組み込む又は取り付けを含む。検体は、ヌクレオチド又はオリゴヌクレオチドであってよく、タグは、特定のヌクレオチド型(A、C、T、又はG)であってもよい。励起光は、タグを有する検体に向けられ、タグは検出可能な蛍光シグナル又は強度発光を発する。強度発光は、検体に化学的に結合された励起タグによって放出される光子を示す。

【0948】

特許請求の範囲を含む本出願全体を通して、「画像、画像データ、又は検体及びそれらの周囲背景の強度放射を示す画像領域が使用されるとき、それらは、検体に取り付けられたタグの強度放射を指す。当業者であれば、取り付けられたタグの強度放射は、タグが取り付けられている検体の強度放射を表すか、又はそれに相当し、したがって互換的に使用されることを理解するであろう。同様に、検体の特性は、検体に取り付けられたタグ、又は取り付けられたタグからの強度放射の特性を指す。例えば、検体の中心とは、検体に取り付けられたタグによって放出される強度放射の中心を指す。別の実施例では、検体の周囲の背景とは、検体に取り付けられたタグによって放出される強度放射の周囲の背景を指す。

【0949】

特許、特許出願、論文、書籍、木、及びウェブページが挙げられるがこれらに限定されない、本出願において引用された文献及び類似材料は、その全体が参照により明示的に組み込まれる。組み込まれた文献及び類似の材料のうちの1つ又はそれ以上が、定義された用語、用語使用、記載された技術などを含むがこれらに限定されない、本出願とは異なる、又は矛盾する場合には、この出願は制御する。

【0950】

開示される技術は、核酸テンプレート又はその相補体、例えば、DNA若しくはRNAポリヌクレオチド又は他の核酸サンプルなどの核酸サンプルから得ることができる核酸配列情報の品質及び量を改善するためにニューラルネットワークを使用する。したがって、開示される技術の特定の実施は、以前に利用可能な方法と比較して、より高いスループットのポリヌクレオチド配列決定、例えば、より高いDNA又はRNA配列データの収集速度、配列データ収集におけるより高い効率、及び／又はそのような配列データを得る低コストを提供する。

【0951】

開示される技術は、ニューラルネットワークを使用して、固相核酸クラスターの中心を識別し、そのようなクラスターの配列決定中に生成される光信号を解析して、隣接する、隣接する、又は重複するクラスター間で曖昧さなく区別して、単一の離散したソースクラスターに配列決定シグナルを割り当てる。したがって、これら及び関連する実施態様は、高密度クラスターアレイの領域からの、配列データなどの有意義な情報の回収を可能にし、有用な情報は、重複する又は非常に近接して離間配置された隣接クラスターの影響を混

10

20

30

40

50

乱させることに起因して、そのような領域から以前に得られなかった場合がある。重複するシグナルの効果（例えば、核酸配列決定において使用されるような）の効果を含む。

【0952】

以下により詳細に記載されるように、特定の実施態様では、本明細書で提供されるように、1つ又は複数の核酸クラスターに固定化された固体支持体を含む組成物が提供される。各クラスターは、同じ配列の複数の固定化された核酸を含み、本明細書で提供されるような検出可能な中心標識を有する識別可能な中心を有し、識別可能な中心は、クラスター内の周囲領域において固定化された核酸と区別可能である。また、識別可能な中心を有するこのようなクラスターを作製及び使用するための方法も本明細書に記載される。

【0953】

本開示の実施態様は、多数の状況での使用が見出され、その利点は、クラスター内の実質的に中心の位置の位置を識別、決定、アノテーション、記録、ないしは別の方法で割り当てる能力から得られる、多くの状況において使用が見出されるであろう。ハイスループット核酸配列決定、光学的又は他のシグナルを個別のソースクラスターに割り当てるための画像解析アルゴリズムの開発、及び固定化された核酸クラスターの中心の認識が望ましい及び有益である他の用途が望ましい。

【0954】

特定の実施態様では、本発明は、核酸配列決定（例えば、「配列決定」）などのハイスループット核酸分析に関連する方法を企図する。例示的なハイスループット核酸解析としては、非限定的に、デノボ配列決定、再配列決定、全ゲノム配列決定、遺伝子発現解析、遺伝子発現モニタリング、エピジェネティクス分析、ゲノムメチル化分析、対立遺伝子特異的プライマー伸長（Allele Specific Primer Extension、APSE）、遺伝的多様性プロファイリング、全ゲノム多型発見及び解析、単一ヌクレオチド多型解析、ハイブリダイゼーション系配列決定法などが挙げられる。当業者は、様々な異なる核酸が、本発明の方法及び組成物を使用して分析され得ることを理解するであろう。

【0955】

本発明の実施は核酸配列決定に関連して記載されているが、それらは、異なる時点で取得された画像データ、空間位置、又は他の時間的若しくは物理的観点で取得された画像データが分析される任意の分野において適用可能である。例えば、本明細書に記載される方法及びシステムは、マイクロアレイ、生物学的検体、細胞、生物などからの画像データが取得され、異なる時点又は視点で取得され、分析される、分子生物学及び細胞生物学の分野において有用である。画像は、蛍光顕微鏡法、光学顕微鏡法、共焦点顕微鏡法、光学画像化法、磁気共鳴画像化法、トモグラフィー走査などが挙げられるが、これらに限定されない、技術分野において既知の任意の数の技術を使用して得ることができる。別の例として、本明細書に記載される方法及びシステムは、監視、空中、又は衛星撮像技術などによって取得された画像データが、異なる時点又は視点で取得され、分析される場合に適用することができる。本方法及びシステムは、視野内で取得された画像を分析するのに特に有用であり、この視野内で、観察される検体は、視野内の互いに対して同じ場所に留まる。しかしながら、検体は、別個の画像で異なる特性を有してもよく、例えば、検体は、視野の別々の画像において異なるように見える場合がある。例えば、検体は、異なる画像で検出された所与の検体の色とは異なるように見える場合があり、異なる画像内の所与の検体のために検出された信号の強度の変化、又は更には、1つの画像中の所与の検体の信号の外観、及び別の画像内の検体の信号の消失を示し得る。

【0956】

本明細書に記載される例は、学術分析又は商業的分析のための様々な生物学的又は化学的プロセス及びシステムにおいて使用されてもよい。より具体的には、本明細書に記載される例は、指定された反応を示すイベント、特性、品質、又は特性を検出することが望ましい様々なプロセス及びシステムにおいて使用されてもよい。例えば、本明細書に記載される例としては、光検出デバイス、バイオセンサー、及びそれらの構成要素、並びにバイオセンサーと共に動作するバイオアッセイシステムが挙げられる。いくつかの実施例では

10

20

30

40

50

、装置、バイオセンサー、及びシステムは、フローセルと、実質的に一体型構造と一緒に（取り外し可能に又は固定的に）結合された1つ又はそれ以上の光センサーと、を含み得る。

【0957】

装置、バイオセンサー、及びバイオアッセイシステムは、個別に又は集合的に検出される複数の指定された反応を実施するように構成されてもよい。装置、バイオセンサー、及びバイオアッセイシステムは、複数の指定された反応が並行して生じる多数のサイクルを実行するように構成されてもよい。例えば、装置、バイオセンサー、及びバイオアッセイシステムを使用して、酵素操作及び光又は画像検出／捕捉の反復サイクルを通して、DNA特徴の高密度配列を配列することができる。したがって、デバイス、バイオセンサー、及びバイオアッセイシステム（例えば、1つ又はそれ以上のカートリッジを介した）は、試薬又は他の反応成分を反応溶液中に送達する1つ又はそれ以上のマイクロ流体チャネル、バイオセンサー、及びバイオアッセイシステムを含んでもよい。いくつかの実施例では、反応溶液は、約5以下、又は約4以下、又は約3以下のpHを含むなど、実質的に酸性であってもよい。いくつかの他の実施例では、反応溶液は、約8以上、又は約9以上、又は約10以上のpHを含むなど、実質的にアルカリ性／塩基性であってもよい。本明細書で使用するとき、用語「酸性」及びその文法的変異体は、約7未満のpH値を指し、用語「塩基性」、「アルカリ性」及びその文法的変異型は、約7を超えるpH値を指す。

【0958】

いくつかの実施例では、反応部位は、均一又は反復パターンなどの所定の方法で提供又は離間される。いくつかの他の実施例では、反応部位はランダムに分布している。反応部位のそれぞれは、関連する反応部位からの光を検出する1つ又はそれ以上の光ガイド及び1つ又はそれ以上の光センサーと関連付けることができる。いくつかの実施例では、反応部位は、指定された反応を少なくとも部分的に区画化し得る反応凹部又はチャンバ内に位置する。

【0959】

本明細書で使用するとき、「指定された反応」は、対象となる検体などの対象となる化学物質又は生物学的物質の化学的、電気的、物理的、又は光学的特性（又は品質）のうちの少なくとも1つの変化を含む。特定の実施例では、指定された反応は、例えば、蛍光標識生体分子を対象とする蛍光標識生体分子の組み込みなどの正の結合事象である。より一般的には、指定された反応は、化学変換、化学変化、又は化学的相互作用であってもよい。指定された反応はまた、電気特性の変化であってもよい。特定の実施例では、指定された反応は、検体と蛍光標識された分子を組み込むことを含む。検体はオリゴヌクレオチドであってもよく、蛍光標識分子はヌクレオチドであってもよい。指定された反応は、励起光が標識ヌクレオチドを有するオリゴヌクレオチドに向けられ、蛍光団が検出可能な蛍光シグナルを発するときに、指定された反応が検出され得る。代替例では、検出された蛍光は、化学発光又は生物発光の結果である。指定された反応はまた、例えば、ドナーフルオロフォアをアクセプタ蛍光団に近接させることによって蛍光（又はF o r s t e r）共鳴エネルギー移動（Fluorescence Resonance Energy Transfer、F R E T）を増加させることができ、ドナーとアクセプタ蛍光団とを分離することによってF R E Tを減少させ、消光剤をフルオロフォアから分離することによって蛍光を増加させるか、又は消光剤及び蛍光団を共局在させることによって蛍光を減少させることができる。

【0960】

本明細書で使用するとき、「反応溶液」、「反応成分」又は「反応物質」は、少なくとも1つの指定された反応を得るために使用され得る任意の物質を含む。例えば、潜在的な反応成分としては、例えば、試薬、酵素、サンプル、他の生体分子、及び緩衝液が挙げられる。反応成分は、溶液中の反応部位に送達されてもよく、及び／又は反応部位で固定されてもよい。反応成分は、反応部位で固定化された対象検体などの別の物質と直接又は間接的に相互作用し得る。上記のように、反応溶液は、実質的に酸性であってもよい（すなわち、比較的高い酸性度を含む）（例えば、約5以下のpH、約4以下のpHを含む）、

又は約 3 以下の pH、又は実質的にアルカリ性／塩基性（すなわち、比較的高いアルカリ性／塩基性を含む）（例えば、約 8 以上の pH、約 9 以上の pH、又は約 10 以上の pH を含む）。

【0961】

本明細書で使用する時、用語「反応部位」は、少なくとも 1 つの指定された反応が生じ得る局所的領域である。反応部位は、物質がその上に固定され得る反応構造又は基材の支持表面を含んでもよい。例えば、反応部位は、その上に反応成分、例えば、その上に核酸のコロニーを有する反応構造（フローセルのチャンネル内に配置され得る）の表面を含んでもよい。いくつかのこのような実施例では、コロニー中の核酸は同じ配列を有し、例えば、一本鎖又は二本鎖テンプレートのクローンコピーである。しかしながら、いくつかの実施例では、反応部位は、例えば、一本鎖又は二本鎖形態で、単一の核酸分子のみを含有してもよい。

10

【0962】

複数の反応部位は、反応構造に沿ってランダムに分布してもよく、又は所定の様式で配置されてもよい（例えば、マイクロアレイなどのマトリックス内の並列）。反応部位はまた、指定された反応を区画化するように構成された空間領域又は容積を少なくとも部分的に画定する反応室又は凹部を含むことができる。本明細書で使用する時、用語「反応チャンバ」又は「反応凹部」は、支持構造体の画定された空間領域（多くの場合、流路と流体連通している）を含む。反応凹部は、周囲環境又は空間領域から少なくとも部分的に分離されてもよい。例えば、複数の反応凹部は、検出表面などの共有された壁によって互いに分離されてもよい。より具体的な例として、反応凹部は、検出表面の内面によって画定された窪み、ウェル、溝、空洞、又は窪みを含むナノセルであってもよく、ナノセルが流路と流体連通することができるように、開口部又は開口部（すなわち、開側面である）を有することができる。

20

【0963】

いくつかの実施例では、反応構造の反応凹部は、固体がその中に完全に又は部分的に挿入され得るように、固体（半固体を含む）に対してサイズ及び形状を定められる。例えば、反応凹部は、捕捉ビーズを収容するような大きさ及び形状であってもよい。捕捉ビーズは、クロノウイルス増幅 DNA 又はその上の他の物質を有してもよい。あるいは、反応凹部は、およそその数のビーズ又は固体基材を受容するような大きさ及び形状であってもよい。別の例として、反応凹部は、拡散又はフィルタ流体又は反応凹部に流入し得る溶液を制御するように構成された多孔質ゲル又は物質で充填されてもよい。

30

【0964】

いくつかの実施例では、光センサー（例えば、フォトダイオード）は、対応する反応部位と関連付けられる。反応部位に関連する光センサーは、関連する反応部位において指定された反応が生じたときに、少なくとも 1 つの光ガイドを介して、関連する反応部位からの光放射を検出するように構成されている。いくつかの場合では、複数の光センサー（例えば、光検出又はカメラデバイスのいくつかのピクセル）は、単一の反応部位に関連付けられてもよい。他の場合では、単一の光センサー（例えば、単一のピクセル）は、単一の反応部位に、又は反応部位の群と関連付けられてもよい。バイオセンサーの光センサー、反応部位、及び他の特徴は、光の少なくとも一部が反射されることなく光センサーによって直接検出されるように構成されてもよい。

40

【0965】

本明細書で使用する時、「生物学的又は化学物質」は、生体分子、対象試料、対象検体、及び他の化学化合物を含む。生物学的物質又は化学物質を使用して、他の化学化合物を検出、識別、若しくは分析するか、又は他の化学化合物を研究又は分析するための仲介として機能してもよい。特定の実施例では、生物学的物質又は化学物質は、生体分子を含む。本明細書で使用する時、「生体分子」は、バイオポリマー、ヌクレオチド、核酸、ポリヌクレオチド、オリゴヌクレオチド、タンパク質、酵素、ポリペプチド、抗体、抗原、リガンド、受容体、多糖類、炭水化物、ポリリン酸、細胞、組織、生物、若しくはそれ

50

らの断片、又は前述の種の類似体若しくは模倣体などの任意の他の生物学的に活性な化学化合物のうちの少なくとも1つを含む。更なる例では、生物学的若しくは化学物質又は生体分子は、酵素又は試薬などの別の反応の生成物、例えば、ピロ配列決定反応においてピロリン酸を検出するために使用される酵素又は試薬などの酵素又は試薬の生成物を検出する。ピロホスフェート検出に有用な酵素及び試薬は、例えば、参照によりその全体が組み込まれる米国特許公開第2005/0244870号に記載されている。

【0966】

生体分子、試料、及び生物学的物質又は化学物質は、天然に存在しても合成であってもよく、反応凹部又は領域内の溶液又は混合物中に懸濁されてもよい。生体分子、試料、及び生物学的物質又は化学物質もまた、固体相又はゲル材料に結合されてもよい。生体分子、試料、及び生物学的物質又は化学物質はまた、医薬組成物を含んでもよい。場合によっては、対象とする生体分子、試料、及び生物学的物質又は化学物質は、標的、プローブ、又は検体と呼ばれる場合がある。

【0967】

本明細書で使用する時、「バイオセンサー」は、反応部位又は反応部位に近接して生じる指定された反応を検出するように構成された複数の反応部位を有する反応構造を含む装置を含む。バイオセンサーは、固体光検出装置又は「撮像」デバイス（例えば、CCD又はCMOS光検出デバイス）、及び任意選択的に、それに取り付けられたフローセルを含んでもよい。フローセルは、反応部位と流体連通する少なくとも1つの流路を含み得る。1つの特定の例として、バイオセンサーは、生物学的アッセイシステムに流体的かつ電気的に結合するように構成される。バイオアッセイシステムは、所定のプロトコル（例えば、配列番号合成）に従って反応部位に反応溶液を送達し、複数の撮像事象を実施してもよい。例えば、バイオアッセイシステムは、反応溶液を反応部位に沿って流すことができる。反応溶液のうちの少なくとも1つは、同じ又は異なる蛍光標識を有する4種類のヌクレオチドを含んでもよい。ヌクレオチドは、反応部位の対応するオリゴヌクレオチドなどに結合してもよい。次いで、バイオアッセイシステムは、励起光源（例えば、発光ダイオード（LED）などの固体光源）を使用して反応部位を照明することができる。励起光は、波長の範囲を含む所定の波長又は波長を有してもよい。入射励起光によって励起された蛍光標識は、光センサーによって検出され得る発光信号（例えば、励起光とは異なる波長又は波長の光、及び潜在的に互いに異なる）を提供することができる。

【0968】

本明細書で使用する時、用語「固定化された」は、生体分子又は生物学的物質又は化学物質に関して使用されるとき、生体分子又は生物学的物質又は化学物質を、光検出デバイス又は反応構造の検出表面などの表面に実質的に付着させることを含む。例えば、生体分子又は生物学的物質又は化学物質は、非共有結合（例えば、静電力、ファンデルワールス、及び疎水性界面の脱水）を含む吸着技術、並びに官能基又はリンカーが生体分子の表面への結合を促進する共有結合技術を含む吸着技術を用いて、反応構造の表面に固定化されてもよい。生体分子又は生物学的物質若しくは化学物質を表面に固定化することは、表面の特性、生体分子又は生物学的物質若しくは化学物質を担持する液体媒体、並びに生体分子又は生物学的物質若しくは化学物質自体の特性に基づいてもよい。いくつかの場合において、表面は、生体分子（又は生物学的物質又は化学物質）を表面に固定化するのを容易にするために、表面を官能化（例えば、化学的又は物理的に修飾）してもよい。

【0969】

いくつかの実施例では、核酸は、その反応凹部の表面などの反応構造に固定化することができる。特定の実施例では、本明細書に記載される装置、バイオセンサー、バイオアッセイシステム及び方法は、天然ヌクレオチド、及び天然ヌクレオチドと相互作用するように構成された酵素の使用を含んでもよい。天然ヌクレオチドとしては、例えば、リボヌクレオチド又はデオキシリボヌクレオチドが挙げられる。天然ヌクレオチドは、一リン酸、二リン酸、又は三リン酸形態であってもよく、アデニン（A）、チミン（T）、ウラシル（U）、グアニン（G）、又はシトシン（C）から選択される塩基を有することができる。

しかしながら、上記ヌクレオチドの非天然ヌクレオチド、修飾ヌクレオチド、又は類似体を使用することができることが理解されるであろう。

【0970】

上記のように、生体分子又は生物学的物質又は化学物質は、反応構造の反応凹部内の反応部位に固定されてもよい。このような生体分子又は生物学的物質は、干渉嵌め、接着、共有結合、又は捕捉によって、反応凹部に物理的に保持又は固定化されてもよい。反応凹部に配置され得る物品又は固体の例としては、ポリマービーズ、ペレット、アガロースゲル、粉末、量子ドット、又は反応チャンバ内で圧縮及び／又は保持され得る他の固体が挙げられる。特定の実施態様では、反応凹部は、DNAオリゴヌクレオチドに共有結合することができるヒドロゲル層でコーティング又は充填されてもよい。特定の実施例では、DNAボールなどの核酸超構造は、例えば、反応凹部の内面に取り付けることによって、又は反応凹部に液体中に滞留することによって、反応凹部内又は反応凹部に配置することができる。DNAボール又は他の核酸超構造を実施することができ、次いで、反応凹部内又は反応凹部に配置することができる。あるいは、DNAボールは、反応凹部においてその場で合成することができる。反応凹部内に固定された物質は、固体、液体、又は気体状態であり得る。

10

【0971】

本明細書で使用する時、用語「検体」は、相対位置に従って他の点又は領域と区別することができるパターンの点又は領域を意味することを意図する。個々の検体は、特定の種類の1つ又はそれ以上の分子を含むことができる。例えば、検体は、特定の配列を有する単一の標的核酸分子を含むことができ、又は検体は、同じ配列（及び／又はその相補的配列）を有するいくつかの核酸分子を含むことができる。パターンの異なる検体である異なる分子は、パターン内の検体の場所に従って互いに分化させることができる。例示的な検体としては、基材中のウェル、基材中又は基材上のビーズ（又は他の粒子）、基材からの突出部、基材上の隆起部、基材上のゲル材料のパッド、又は基材内のチャンネルが挙げられる。

20

【0972】

検出、特徴付け、又は識別される様々な標的検体のいずれも、本明細書に記載される装置、システム、又は方法でを使用することができる。例示的な検体としては、限定するものではないが、核酸（例えば、DNA、RNA又はそれらの類似体）、タンパク質、多糖類、細胞、抗体、エピトープ、受容体、リガンド、酵素（例えば、キナーゼ、ホスファターゼ又はポリメラーゼ）、小分子薬物候補、細胞、ウイルス、生物などが挙げられるが、これらに限定されない。

30

【0973】

用語「検体」、「核酸」、「核酸分子」、及び「ポリヌクレオチド」という用語は、本明細書において互換的に使用される。様々な実施態様では、核酸は、特定の種類の核酸分析のために、本明細書で提供されるようなテンプレート（例えば、核酸テンプレート、又は核酸テンプレートに相補的な核酸相補体）として使用されてもよく、核酸増幅、核酸発現解析、及び／又は核酸配列決定、又はこれらの好適な組み合わせが挙げられるが、これらに限定されない。特定の実施における核酸としては、例えば、3' - 5' ホスホジエステル中のデオキシリボヌクレオチドの直鎖ポリマー、又はデオキシリボ核酸（Deoxyribonucleic Acid、DNA）、例えば、一本鎖及び二本鎖DNA、ゲノムDNA、コピーDNA若しくは相補的DNA（complementary DNA、cDNA）、組み換えDNA、又は任意の形態の合成DNA若しくは修飾DNAが挙げられる。他の実施態様では、核酸としては、例えば、3' - 5' ホスホジエステル中のリボヌクレオチドの直鎖ポリマー、又はリボ核酸（Ribonucleic Acid、RNA）などの他の結合、例えば、一本鎖及び二本鎖RNA、メッセンジャー（mRNA）、コピーRNA又は相補的RNA（complementary RNA、cRNA）、あるいはスプライシングされたmRNA、リボソームRNA、小核RNA（small nucleolar RNA、snRNA）、マイクロRNA（microRNA、miRNA）、低干渉RNA（small interfering RNA、siRNA）、ピウイRNA（piwi RNA、piRNA）

40

50

RNA)、又は任意の形態の合成若しくは修飾RNA。本発明の組成物及び方法において使用される核酸は、長さが変化してもよく、無傷又は完全長の分子若しくは断片、又はより大きい核酸分子のより小さい部分であってもよい。特定の実施態様では、核酸は、本明細書の他の箇所に記載されるように、1つ又はそれ以上の検出可能な標識を有してもよい。

【0974】

用語「検体」、「クラスター」、「核酸クラスター」、「核酸コロニー」、及び「DNAクラスター」は互換的に使用され、固体支持体に結合された核酸テンプレート及び／又はその相補体の複数のコピーを指す。典型的には、特定の好ましい実施態様では、核酸クラスターは、それらの5'末端を介して固体支持体に結合されたテンプレート核酸及び／又はその相補体の複数のコピーを含む。核酸クラスターを構成する核酸鎖のコピーは、一本鎖又は二本鎖形態であってよい。クラスター内に存在する核酸テンプレートのコピーは、例えば、標識部分の存在に起因して、互いに異なる対応する位置にヌクレオチドを有することができる。対応する位置はまた、異なる化学構造を有するが、ウラシル及びチミンの場合など、類似のWatson-Crick塩基対形成特性を有するアナログ構造を含むことができる。

【0975】

核酸のコロニーはまた、「核酸クラスター」と呼ばれることもある。核酸コロニーは、本明細書の他の箇所で更に詳細に記載されるように、クラスター増幅又はブリッジ増幅技術によって任意に作成することができる。標的配列の複数の反復は、ローリングサークル増幅手順を使用して作成された混乱剤などの単一の核酸分子中に存在し得る。

【0976】

本発明の核酸クラスターは、使用される条件に応じて、異なる形状、サイズ、及び密度を有することができる。例えば、クラスターは、実質的に円形、多面、ドーナツ形、又はリング形状の形状を有することができる。核酸クラスターの直径は、約0.2 μm ～約6 μm 、約0.3 μm ～約4 μm 、約0.4 μm ～約3 μm 、約0.5 μm ～約2 μm 、約0.75 μm ～約1.5 μm 、又は任意の介在直径であるように設計することができる。特定の実施態様において、核酸クラスターの直径は、約0.5 μm 、約1 μm 、約1.5 μm 、約2 μm 、約2.5 μm 、約3 μm 、約4 μm 、約5 μm 、又は約6 μm である。核酸クラスターの直径は、クラスターの産生において実施される増幅サイクルの数、核酸テンプレートの長さ、又はクラスターが形成される表面に付着したプライマーの密度を含むが、これらに限定されない多数のパラメータによって影響され得る。核酸クラスターの密度は、典型的には、0.1/mm²、1/mm²、10/mm²、100/mm²、1,000/mm²、10,000/mm²～100,000/mm²の範囲であるように設計することができる。本発明は、一部では、より高密度の核酸クラスター、例えば、100,000/mm²～1,000,000/mm²、及び1,000,000/mm²～10,000,000/mm²を更に企図する。

【0977】

本明細書で使用する時、「検体」は、検体又は視野内の対象領域である。マイクロアレイデバイス又は他の分子分析デバイスに関連して使用される場合、検体は、類似又は同一の分子によって占有される領域を指す。例えば、検体は、増幅オリゴヌクレオチド、又は同じ又は類似の配列を有するポリヌクレオチド又はポリペプチドの任意の他の群であり得る。他の実施態様では、検体は、試料上の物理的領域を占有する任意の要素又は要素群であり得る。例えば、検体は、ランドのパセル、水の本体などであってもよい。検体が撮像されると、各検体は、一部の領域を有する。したがって、多くの実施態様では、検体は、単に1つのピクセルではない。

【0978】

検体間の距離は、任意の数の方法で説明することができる。いくつかの実施態様では、検体間の距離は、1つの検体の中心から別の検体の中心まで説明することができる。他の実施態様では、距離は、1つの検体の縁部から別の検体の縁部まで、又は各検体の最も外側の識別可能な点間に記載することができる。検体の縁部は、チップ上の理論的若しくは

10

20

30

40

50

実際の物理的境界、又は検体の境界内のいくつかの点として説明することができる。他の実施態様では、距離は、試料上の固定点、又は試料の画像に関して説明することができる。

【0979】

一般に、分析方法に関して、いくつかの実施態様が本明細書に記載される。自動又は半自動化方法で方法を実行するためのシステムも提供されることが理解されるであろう。したがって、本開示は、ニューラルネットワークベースのテンプレート生成及びベースコールシステムを提供し、システムは、プロセッサと、記憶装置と、画像解析用のプログラムと、を含むことができ、プログラムは、本明細書に記載される方法のうちの1つ又はそれ以上を実行するための命令を含む。したがって、本明細書に記載される方法は、例えば、本明細書に記載されるか又は技術分野において既知の構成要素を有するコンピュータ上で実行することができる。

10

【0980】

本明細書に記載される方法及びシステムは、様々なオブジェクトのうちのいずれかを分析するのに有用である。特に有用な物体は、固体担体又は付着した検体を有する固相表面である。本明細書に記載される方法及びシステムは、 x y 平面における検体の繰り返しパターンを有する物体と共に使用される場合、利点を提供する。一例は、細胞、ウイルス、核酸、タンパク質、抗体、炭水化物、小分子（薬物候補など）、生物学的活性分子、又は他の対象検体の集合を有するマイクロアレイである。

【0981】

核酸及びポリペプチドなどの生物学的分子を有する検体を有するアレイの用途の数が増えてきた。このようなマイクロアレイは、典型的には、デオキシリボ核酸（DNA）又はリボ核酸（RNA）プローブが挙げられる。これらは、ヒト及び他の生物に存在するヌクレオチド配列に特異的である。特定の用途では、例えば、個々のDNA又はRNAプローブをアレイの個々の検体に取り付けることができる。既知のヒト又は生物からのものなどの試験サンプルは、標的核酸（例えば、遺伝子断片、mRNA、又はアンプリコン）が配列中のそれぞれの検体で相補的プローブにハイブリダイズするように、アレイに曝露することができる。プローブは、標的特異的プロセス（例えば、標的核酸上に存在する標識に起因して、又は検体においてハイブリダイズした形態で存在するプローブ又は標的の酵素標識に起因して）標識することができる。次いで、検体の上の特定の光の周波数を走査して、どの標的核酸が試料中に存在するかを識別することによって検査することができる。

20

30

【0982】

生物学的マイクロアレイは、遺伝子配列決定及び類似の用途に使用され得る。一般に、遺伝子配列決定は、DNA又はRNAの断片などの標的核酸の長さのヌクレオチドの順序を決定することを含む。比較的短い配列は、典型的には、各検体において配列決定され、得られた配列情報は、様々な生物情報科学法において使用されて、フラグメントが由来する多くの広範な長さの遺伝物質の配列を確実に決定するために、様々な生物情報科学法において使用されてもよい。特徴的断片の自動化されたコンピュータベースのアルゴリズムが開発されており、ゲノムマッピング、遺伝子の識別、及びそれらの機能などにおいて、より最近使用されてきた。マイクロアレイは、多数の変異体が存在するため、ゲノム含有量を特徴付けるのに特に有用であり、このことは、個々のプローブ及び標的に対して多くの実験を実施する代替物であるため、ゲノム含有量を特徴付けるのに特に有用である。マイクロアレイは、実用的な方法でこのような調査を実施するための理想的なフォーマットである。

40

【0983】

技術分野において既知の様々な検体アレイ（「マイクロアレイ」とも呼ばれる）のいずれも、本明細書に記載される方法又はシステムにおいて使用することができる。典型的なアレイは、それぞれが個々のプローブ又はプローブの集団を有する検体を含む。後者の場合、各検体におけるプローブの集団は、典型的には、単一種のプローブを有する均質である。例えば、核酸配列の場合、各検体は、それぞれ共通の配列を有する複数の核酸分子を有することができる。しかしながら、いくつかの実施態様では、アレイの各検体における

50

集団は、不均質であり得る。同様に、タンパク質配列は、単一のタンパク質又はタンパク質の集団を有する検体を有することができ、典型的には、同じアミノ酸配列を有するが、必ずしもそうではない。プローブは、例えば、プローブを表面に共有結合することによって、又はプローブと表面との非共有相互作用（複数可）を介して、アレイの表面に取り付けることができる。いくつかの実施態様では、核酸分子などのプローブは、例えば、米国特許出願第13／784,368号、及び米国特許出願公開第2011／0059865（A1）号明細書に記載されるようなゲル層を介して表面に取り付けられ得、それぞれ参照により本明細書に組み込まれる。

【0984】

例示的なアレイとしては、限定するものではないが、Illumina, Incから入手可能なBeadChipアレイ（San Diego, Calif.）又は他のもの、例えば、プローブが、表面上に存在するビーズ（例えば、表面上のウェル内のビーズ）に取り付けられる以下に記載されたものなどの他のものが挙げられる。米国特許第6,266,459号明細書、米国特許第6,355,431号明細書、米国特許第6,770,441号明細書、米国特許第6,859,570号明細書、又は米国特許第7,622,294号明細書、又はPCT国際公開第00／63437号。これらは、それぞれ参照により本明細書に組み込まれる。使用することができる市販のマイクロアレイの更なる例としては、例えば、VLSIPS（商標）（Very Large Scale Immobilized Polymer Synthesis）技術と呼ばれることがある技術に従って合成されたAffymetrix（登録商標）GeneChip（登録商標）マイクロアレイ又は他のマイクロアレイが挙げられる。スポットされたマイクロアレイはまた、本開示のいくつかの実施態様による方法又はシステムにおいて使用することができる。例示的なスポットティングされたマイクロアレイは、Amersham Biosciencesから入手可能なCodeLink（商標）Arrayである。有用な別のマイクロアレイは、Agilent Technologiesから入手可能なSurePrint™ Technologyなどのインクジェット印刷法を使用して製造されるものである。

【0985】

他の有用な配列としては、核酸配列決定用途で使用されるものが挙げられる。例えば、ゲノムフラグメント（多くの場合、クラスターと称される）のアンプリコンを有するアレイは、Bentley et al., Nature 456:53-59（2008）、国際公開第04／018497号、国際公開第91／06678号、国際公開第07／123744号、米国特許第7,329,492号明細書、米国特許第7,211,414号明細書、米国特許第7,315,019号明細書、米国特許第7,405,281号明細書、又は米国特許第7,057,026号明細書、又は米国特許出願公開第2008／0108082（A1）号明細書に記載されているように特に有用であり、それぞれ参照により本明細書に組み込まれる。核酸配列決定に有用な別の種類の配列は、エマルジョンPCR技術から生成される粒子の配列である。実施例は、Dressman et al., Proc. Natl. Acad. Sci. USA 100:8817-8822（2003）、国際公開第05／010145号、米国特許出願公開第2005／0130173号明細書又は米国特許出願公開第2005／0064460号明細書に記載されており、これらはそれぞれその全体が参照により本明細書に組み込まれる。

【0986】

核酸配列に使用される配列は、多くの場合、核酸検体のランダムな空間パターンを有する。例えば、Illumina Incから入手可能なHiSeq又はMiSeq配列決定プラットフォーム（San Diego, Calif.）は、核酸配列がランダム播種、続いてブリッジ増幅によって形成されるフローセルを利用する。しかしながら、パターン化された配列は、核酸配列又は他の分析用途にも使用することができる。パターン化されたアレイの例、それらの使用方法及びその使用法は、米国特許出願第13／787,396号明細書、米国特許出願第13／783,043号明細書、米国特許出願第13／

784, 368号明細書、米国特許出願公開第2013/0116153号明細書、及び米国特許出願公開第2012/0316086号明細書に記載されており、これらはそれぞれ参照により本明細書に組み込まれる。このようなパターン化された配列の検体を使用して、単一の核酸テンプレート分子を捕捉して、例えば、ブリッジ増幅を介して、均質なコロニーの後続の形成を行うことができる。このようなパターン化された配列は、核酸配列決定用途に特に有用である。

【0987】

アレイ（又は本明細書の方法又はシステムで使用される他の物体）上の検体のサイズは、特定の用途に適するように選択することができる。例えば、いくつかの実施態様では、アレイの検体は、単一の核酸分子のみを収容するサイズを有することができる。このサイズ範囲の複数の検体を有する表面は、単一分子分解能で検出するための分子の配列を構築するのに有用である。このサイズ範囲の検体もまた、核酸分子のコロニーをそれぞれ含む検体を有するアレイでの使用にも有用である。したがって、アレイの検体はそれぞれ、約 1mm^2 以下、約 $500\mu\text{m}^2$ 以下、約 $100\mu\text{m}^2$ 以下、約 $10\mu\text{m}^2$ 以下、約 $1\mu\text{m}^2$ 以下、約 500nm^2 以下、又は約 100nm^2 以下、約 10nm^2 以下、約 5nm^2 以下、又は約 1nm^2 以下の面積を有することができる。代替的に又は追加的に、アレイの検体は、約 1mm^2 以上、約 $500\mu\text{m}^2$ 以上、約 $100\mu\text{m}^2$ 以上、約 $10\mu\text{m}^2$ 以上、約 $1\mu\text{m}^2$ 以上、約 500nm^2 以上、約 100nm^2 以上、約 10nm^2 以上、約 5nm^2 以上、又は約 1nm^2 以上である。実際に、検体は、上記に例示したものから選択される上限と下限との間の範囲内の大きさを有することができる。表面の検体のいくつかのサイズ範囲が核酸及び核酸のスケールに関して例示されてきたが、これらのサイズ範囲の検体は、核酸を含まない用途に使用できることが理解されるであろう。検体のサイズは、核酸用途に使用されるスケールに必ずしも限定される必要はないことが更に理解されるであろう。

【0988】

検体のアレイなどの複数の検体を有する物体を含む実施例では、検体は、互いの間の空間で分離されている、別個のものとするすることができる。本発明において有用なアレイは、最大で $100\mu\text{m}$ 、 $50\mu\text{m}$ 、 $10\mu\text{m}$ 、 $5\mu\text{m}$ 、 $1\mu\text{m}$ 、 $0.5\mu\text{m}$ 以下の縁部から縁部までの距離によって分離される検体を有することができる。代替的に又は追加的に、アレイは、少なくとも $0.5\mu\text{m}$ 、 $1\mu\text{m}$ 、 $5\mu\text{m}$ 、 $10\mu\text{m}$ 、 $50\mu\text{m}$ 、 $100\mu\text{m}$ 、又はそれ以上の縁部から縁部までの距離によって分離される検体を有することができる。これらの範囲は、検体の平均縁部間隔及び縁部間隔、並びに最小又は最大間隔に適用することができる。

【0989】

いくつかの実施態様では、アレイの検体は、別個である必要はなく、代わりに、隣接する検体は互いに当接することができる。検体が別個であるか否かに関わらず、検体及び／又は検体のピッチの大きさは、アレイが所望の密度を有することができるように変化し得る。例えば、規則的なパターンにおける平均検体ピッチは、最大で $100\mu\text{m}$ 、 $50\mu\text{m}$ 、 $10\mu\text{m}$ 、 $5\mu\text{m}$ 、 $1\mu\text{m}$ 、 $0.5\mu\text{m}$ 以下であり得る。代替的に又は追加的に、規則的なパターンにおける平均検体ピッチは、少なくとも $0.5\mu\text{m}$ 、 $1\mu\text{m}$ 、 $5\mu\text{m}$ 、 $10\mu\text{m}$ 、 $50\mu\text{m}$ 、 $100\mu\text{m}$ 、又はそれ以上であり得る。これらの範囲は、規則的なパターンの最大ピッチ又は最小ピッチにも適用することができる。例えば、規則的なパターンの最大検体ピッチは、 $100\mu\text{m}$ 以下、 $50\mu\text{m}$ 以下、 $10\mu\text{m}$ 以下、 $5\mu\text{m}$ 以下、 $1\mu\text{m}$ 以下、 $0.5\mu\text{m}$ 以下とすることができ、かつ／又は規則的なパターンにおける最小検体ピッチは、少なくとも $0.5\mu\text{m}$ 、 $1\mu\text{m}$ 、 $5\mu\text{m}$ 、 $10\mu\text{m}$ 、 $50\mu\text{m}$ 、 $100\mu\text{m}$ 、又はそれ以上であり得る。

【0990】

アレイ内の検体の密度は、単位面積あたりに存在する検体の数に関しても理解され得る。例えば、アレイに関する検体の平均密度は、少なくとも約 1×10^3 検体/ mm^2 、 1×10^4 検体/ mm^2 、 1×10^5 検体/ mm^2 、 1×10^6 検体/ mm^2 、 1×10^6 検

体/mm²、 1×10^7 検体/mm²、 1×10^8 検体/mm²、又は 1×10^9 検体/mm²以上であり得る。あるいは、又はそれに加えて、アレイに関する検体の平均密度は、最大で約 1×10^9 検体/mm²、 1×10^8 検体/mm²、 1×10^7 検体/mm²、 1×10^6 検体/mm²、 1×10^5 検体/mm²、 1×10^4 検体/mm²、又は 1×10^3 検体/mm²以下であり得る。

【0991】

上記の範囲は、例えば、検体のアレイの全て又は一部を含む規則的なパターンの全て又は一部に適用することができる。

【0992】

パターン内の検体は、様々な形状のいずれかを有することができる。例えば、アレイの表面上などの2次元平面で観察される場合、検体は、丸みを帯びた、円形、楕円形、矩形、正方形、対称、非対称、三角形、多角形などに見える場合がある。検体は、例えば六角形又は直線パターンを含む規則的な繰り返しパターンで配置することができる。パターンは、所望のレベルのパッキングを達成するように選択され得る。例えば、円形検体は、六角形の配置で最適に充填される。当然のことながら、他の包装構成もまた、円形検体のために使用することができ、逆もまた同様である。

【0993】

パターンは、パターンの最小幾何学的単位を形成するサブセット内に存在する検体の数に関して特徴付けることができる。サブセットは、例えば、少なくとも約2、3、4、5、6、10以上の検体を含み得る。検体のサイズ及び密度に応じて、幾何学的単位は、 1 mm^2 、 $500 \mu\text{m}^2$ 、 $100 \mu\text{m}^2$ 、 $50 \mu\text{m}^2$ 、 $10 \mu\text{m}^2$ 、 $1 \mu\text{m}^2$ 、 500 nm^2 、 100 nm^2 、 50 nm^2 、 10 nm^2 以下の面積を占めることができる。代替的に又は追加的に、幾何学的単位は、 10 nm^2 、 50 nm^2 、 100 nm^2 、 500 nm^2 、 $1 \mu\text{m}^2$ 、 $10 \mu\text{m}^2$ 、 $50 \mu\text{m}^2$ 、 $100 \mu\text{m}^2$ 、 $500 \mu\text{m}^2$ 、 1 mm^2 以上の面積を占めることができる。形状、サイズ、ピッチなどの幾何学的単位における検体の特性は、アレイ又はパターンの検体に関して、より一般的に本明細書に記載されるものから選択することができる。

【0994】

検体の規則的なパターンを有するアレイは、検体の相対的な場所に対して順序付けられるが、各検体の1つ又はそれ以上の他の特性に対してランダムであってもよい。例えば、核酸配列の場合、核酸検体は、それらの相対的な位置に関して規則的であるが、任意の特定の検体に存在する核酸種に関する配列の知識に関してランダムであってもよい。より具体的な例として、テンプレート核酸を有する検体の反復パターンを播種し、各検体でテンプレートを増幅して、検体においてテンプレートのコピーを形成することによって形成される核酸配列（例えば、クラスター増幅又はブリッジ増幅を介して、核酸検体の規則的なパターンを有するが、配列にわたる核酸の配列の分布に関してランダムであろう。したがって、アレイ上の核酸材料の存在の検出は、検体の繰り返しパターンをもたらすことができるのに対し、配列特異的検出は、アレイにわたる信号の非反復分布をもたらすことができる。

【0995】

本明細書におけるパターン、順序、ランダム性などの説明は、アレイ上の検体などの物体上の検体にも関するだけでなく、画像中の検体にも関連することが理解されるであろう。したがって、パターン、順序、ランダム性などは、限定するものではないが、グラフィカルユーザーインターフェース又は他の出力デバイスなどのコンピュータ可読媒体又はコンピュータ構成要素を含むがこれらに限定されない、画像データを記憶、操作、又は通信するために使用される様々なフォーマットのうちのいずれかに存在することができる。

【0996】

本明細書で使用するとき、用語「画像」は、オブジェクトの全て又は一部の表現を意味することを意図する。表現は、光学的に検出された再現であり得る。例えば、蛍光、発光、散乱、又は吸収信号から画像を得ることができる。画像内に存在するオブジェクトの部

10

20

30

40

50

分は、物体の表面又は他の x y 面であり得る。典型的には、画像は 2 次元表現であるが、場合によっては、画像内の情報は、3 つ又はそれ以上の次元から導出することができる。画像は、光学的に検出された信号を含む必要はない。非光信号を代わりに存在させることができる。画像は、本明細書の他の箇所に記載されるもののうちの 1 つ又はそれ以上などの、コンピュータ可読フォーマット又は媒体に提供することができる。

【0997】

本明細書で使用する時、「画像」は、試料又は他の物体の少なくとも一部分の再現又は表現を指す。いくつかの実施態様では、再現は、例えばカメラ又は他の光学検出器によって生成される光再現である。再現は、非光学的再現、例えば、ナノ細孔検体のアレイから得られる電気信号の表現、又はイオン感応性 CMOS 検出器から得られた電気信号の表現であり得る。特定の実施態様では、非光学的再現性は、本明細書に記載される方法又は装置から除外され得る。画像は、例えば、100 μm 、50 μm 、10 μm 、5 μm 、1 μm 、又は 0.5 μm 未満離れたものを含む、様々な間隔のいずれかで存在する検体の検体を区別することができる解像度を有することができる。

【0998】

本明細書で使用する時、「取得」、「取得」、「及び同様の用語は、画像ファイルを取得するプロセスの任意の部分の指す。いくつかの実施態様では、データ取得は、標本の画像を生成することと、標本内の信号を探すことと、信号の画像を探すか又は生成するように検出デバイスに指示することと、画像ファイルの更なる分析又は変換のための命令、及び画像ファイルの任意の数の変換又は操作のための命令を与えることと、を含むことができる。

【0999】

本明細書で使用する時、用語「テンプレート」は、信号又は検体間の場所又は関係の表現を指す。したがって、いくつかの実施態様では、テンプレートは、検体中の検体に対応する信号の表現を有する物理的グリッドである。いくつかの実施態様では、テンプレートは、チャート、テーブル、テキストファイル、又は検体に対応する場所を示す他のコンピュータファイルであり得る。本明細書に提示される実施態様では、異なる基準点で捕捉された試料の画像のセットにわたって検体の場所を追跡するためにテンプレートが生成される。例えば、テンプレートは、別の検体に対する 1 つの検体の方向及び／又は距離を記述する x 、 y 座標、又は一連の値であり得る。

【1000】

本明細書で使用する時、用語「標本」は、画像が取り込まれる物体又は物体の領域を指すことができる。例えば、画像が土壌の表面から撮影される実施例では、ランドのパセルは、標本であり得る。生体分子の分析がフローセル内で行われる他の実施態様では、フローセルは、任意の数のサブディビジョンに分割されてもよく、これらのそれぞれは検体であってもよい。例えば、フローセルは、様々な流路又はレーンに分割されてもよく、各レーンは、画像化される 2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、140、160、180、200、400、600、800、1000 個以上の別個の領域に更に分割され得る。フローセルの一例は 8 つのレーンを有し、各レーンは 120 個の標本又はタイルに分割されている。別の実施態様では、試料は、複数のタイル、又は更にはフローセル全体で作製されてもよい。したがって、各検体の画像は、撮像されるより大きい表面の領域を表すことができる。

【1001】

本明細書に記載される範囲及び連続数リストへの言及は、列挙された数だけではなく、列挙された数字間の全ての実数を含むことが理解されるであろう。

【1002】

本明細書で使用する時、「参照点」は、画像間の任意の時間的又は物理的区別を指す。好ましい別の実施態様では、基準点は時間点である。より好ましい実施態様では、参照点は、配列決定反応中の時点又はサイクルである。しかしながら、用語「基準点」は、画

像を区別又は分離することができる、角度、回転、時間、又は他の態様などの画像を区別又は分離する他の態様を含むことができる。

【1003】

本明細書で使用する時、「画像のサブセット」は、セット内の画像のグループを指す。例えば、サブセットは、画像のセットから選択される1、2、3、4、6、8、10、12、14、16、18、20、30、40、50、60又は任意の数の画像を含んでもよい。特定の別の実施態様では、サブセットは、1、2、3、4、6、8、10、12、14、16、18、20、30、40、50、60以下、又は画像のセットから選択される任意の数の画像を含んでもよい。好ましい別の実施態様では、画像は、各サイクルに相關する4つの画像を有する1つ又はそれ以上の配列決定サイクルから得られる。したがって、例えば、サブセットは、4サイクルにわたって取得された16画像のグループであり得る。

10

【1004】

塩基は、ヌクレオチド塩基又はヌクレオチド、(アデニン)、C(シトシン)、T(チミン)、又はG(グアニン)を指す。本出願は、「塩基(複数可)」及び「ヌクレオチド(複数可)」を互換的に使用する。

【1005】

用語「染色体」は、DNA及びタンパク質成分(特にヒストン)を含むクロマチンストランドに由来する、生きている細胞の本発明の有効性を有する遺伝子キャリアを指す。本明細書では、従来の国際的に認識されている個々のヒトゲノム染色体番号付けシステムが本明細書で使用される。

20

【1006】

「部位」という用語は、参照ゲノム上の固有の位置(例えば、染色体ID、染色体位置及び配向)を指す。いくつかの実施態様では、部位は、残基、配列タグ、又は配列上のセグメントの位置であってもよい。用語「遺伝子座」は、参照染色体上の核酸配列又は多型の特定の位置を指すために使用されてもよい。

【1007】

本明細書における用語「試料」は、典型的には、配列決定及び／又はフェーズドされる核酸を含有する生物液、細胞、組織、器官、又は生物に由来するサンプル、又は配列決定及び／又はフェーズドされる核酸配列を少なくとも1つ含有する核酸の混合物に由来するサンプルを指す。このような試料としては、痰／口腔流体、羊水、血液、血液画分、細針生検試料(例えば、外科生検、針生検など)、尿、腹膜流体、胸膜流体、組織外植片、臓器培養物、及びこれらの任意の他の組織若しくは細胞調製物、又はこれらの画分若しくは誘導体が挙げられるが、これらに限定されない。サンプルは、多くの場合、ヒト被験者(例えば、患者)から採取されるが、試料は、イヌ、ネコ、ウマ、ヤギ、ヒツジ、ウシ、ブタなどを含むがこれらに限定されない、染色体を有する任意の生物から採取することができる。試料は、生物学的源から得られるように、又は試料の特性を修正する前処理後に、直接使用することができる。例えば、このような前処理は、血漿を血液から調製すること、粘性流体を希釈することなどを含んでもよい。前処理の方法には、濾過、沈殿、希釈、蒸留、混合、遠心分離、凍結、凍結乾燥、濃縮、増幅、核酸断片化、干渉成分の不活性化、試薬の添加、溶解などを含んでもよいが、これらに限定されない。

30

40

【1008】

用語「配列」は、互いに結合されたヌクレオチドの鎖を含むか、又はそれを表す。ヌクレオチドは、DNA又はRNAに基づくことができる。1つの配列は、複数のサブ配列を含んでもよいことを理解されたい。例えば、単一の配列(例えば、PCRアンプリコン)は、350ヌクレオチドを有してもよい。サンプル読み取りは、これらの350ヌクレオチド内に複数のサブ配列を含んでもよい。例えば、サンプル読み取りは、例えば、20-50ヌクレオチドを有する第1及び第2のフランキングサブ配列を含んでもよい。第1及び第2の隣接するサブ配列は、対応するサブ配列(例えば、40-100ヌクレオチド)を有する反復セグメントの両側に位置してもよい。隣接するサブ配列のそれぞれは、プラ

50

イマーサブ配列（例えば、10-30ヌクレオチド）を含んでもよい（又はその一部を含み得る）。読書を容易にするために、用語「サブ配列」は「配列」と称されるが、2つの配列は、共通のストランド上で互いに別個である必要はないことが理解される。本明細書に記載の様々な配列を区別するために、配列は、異なる標識（例えば、標的配列、プライマー配列、隣接配列、参照配列など）を与えられてもよい。「対立遺伝子」などの他の用語は、同様の物体を区別するために異なるラベルを与えられてもよい。アプリケーションは、「読み取り（単数又は複数）」及び「配列読み取り（単数又は複数）」を互換的に使用する。

【1009】

用語「paired end sequencing」は、標的フラグメントの両端を配列する配列決定方法を指す。ペレッドエンド配列決定は、ゲノム再構成及び反復セグメントの検出、並びに遺伝子融合及び新規転写物の検出を容易にし得る。パイレッドエンド配列決定の方法は、国際公開第07010252号、PCT出願第PCTGB2007/003798号、及び米国特許出願公開第2009/0088327号に記載されており、これらはそれぞれ参照により本明細書に組み込まれる。一実施例では、一連の操作は、以下のように実施されてもよく、（a）核酸のクラスターを生成する、（b）核酸を直鎖化すること、（c）第1の配列決定プライマーをハイブリダイズし、伸長の繰り返しサイクル、走査及び脱ブロッキングを行う。（d）相補的なコピーを合成することによって、フロー細胞表面上の標的核酸を「反転」し、（e）再合成された鎖を直鎖化し、（f）第2配列決定プライマーをハイブリダイズし、伸長の繰り返しサイクル、走査及び脱ブロッキングを行う。反転操作は、ブリッジ増幅の単一サイクルについて上述した試薬を送達することができる。

【1010】

用語「参照ゲノム」又は「参照配列」は、対象からの識別された配列を参照するために使用され得る任意の生物の部分的又は完全ないずれかの特定の既知のゲノム配列を指す。例えば、ヒト被験者に使用される参照ゲノム、並びに多くの他の生物が、National Center for Biotechnology Information at ncbi.nlm.nih.govで見出される。「ゲノム」は、核酸配列で発現される、生物又はウイルスの完全な遺伝子情報を指す。ゲノムは、DNAの遺伝子及び非コード配列の両方を含む。参照配列は、それに位置合わせされたリードよりも大きくてもよい。例えば、それは、少なくとも約100倍大きい、又は少なくとも約1000倍大きい、又は少なくとも約10,000倍大きい、又は少なくとも約105倍大きい、又は少なくとも約106倍大きい、又は少なくとも約107倍大きい場合がある。一実施例では、参照ゲノム配列は、完全長ヒトゲノムのものである。別の例では、参照ゲノム配列は、13番染色体などの特定のヒト染色体に限定される。いくつかの実施態様では、参照染色体は、ヒトゲノムバージョンhg19からの染色体配列である。このような配列は、染色体参照配列と呼ばれる場合があるが、用語参照ゲノムは、そのような配列を網羅することを意図している。参照配列の他の例としては、他の種のゲノム、並びに任意の種の染色体、サブ染色体領域（ストランドなど）などが挙げられる。様々な実施態様では、参照ゲノムは、複数の個体に由来するコンセンサス配列又は他の組み合わせである。しかしながら、特定の用途では、参照配列は、特定の個体から採取されてもよい。他の実施態様では、「ゲノム」はまた、ゲノム配列の特定の記憶形式及び表現を使用するいわゆる「グラフゲノム」も被覆する。一実施態様では、グラフゲノムは、線形ファイルにデータを記憶する。別の実施態様では、グラフゲノムは、代替的な配列（例えば、小さな差異を有する染色体の異なるコピー）がグラフ内の異なる経路として記憶されている表現を指す。グラフゲノムの実施に関する更なる情報は、<https://www.biorxiv.org/content/biorxiv/early/2018/03/20/194530.full.pdf>において見出すことができ、その内容は参照によりその全体が本明細書に組み込まれる。

【1011】

用語「読み取られる」は、ヌクレオチドサンプル又は参照のフラグメントを記述する配列データの集合を指す。用語「読み取られる」は、サンプル読み取り及び／又は参照読み取りを指し得る。典型的には、必ずしもそうではないが、読み取りは、サンプル又は参照における連続的な塩基対の短い配列を表す。読み取りは、サンプル又は参照フラグメントのベース対配列（A T C G）によって記号的に表されてもよい。読み取りが参照配列と一致するか、又は他の基準を満たすかを判定するために、メモリデバイスに記憶され、適切に処理されてもよい。読み取りは、配列決定機器から直接、又はサンプルに関する記憶された配列情報から間接的に得られてもよい。場合によっては、例えば、染色体又はゲノム領域又は遺伝子に位置合わせされ、特異的に割り当てられ得る、より大きな配列又は領域を識別するために使用することができる十分な長さ（例えば、少なくとも約25bp）のDNA配列である。

10

【1012】

次世代配列決定法としては、例えば、合成技術（Illumina）、ピロ配列決定（454）、イオン半導体技術（Ion Torrent sequencing）、一分子リアルタイム配列決定（Pacific Biosciences）及びライゲーションによる配列決定（SOLiD配列決定）が挙げられる。配列決定法に応じて、各読み取りの長さは、約30bp～10,000bpを超えて変化し得る。例えば、SOLiDシーケンサを用いたDNA配列決定法は、約50bpの核酸リードを生成する。別の例では、Ion Torrent Sequencingは、最大400bpの核酸リードを生成し、454のピロ配列は、約700bpの核酸リードを生成する。更に別の例では、単分子リアルタイム配列決定法は、10,000bp～15,000bpのリードを生成し得る。したがって、特定の実施態様では、核酸配列のリードは、30～100bp、50～200bp、又は50～400bpの長さを有する。

20

【1013】

用語「サンプル読み取り」、「サンプル配列」又は「サンプル断片」は、サンプルから対象とするゲノム配列に関する配列データを指す。例えば、サンプル読み取りは、順方向及び逆方向プライマー配列を有するPCRアンプリコンからの配列データを含む。配列データは、任意の選択配列手順から得ることができる。サンプルの読み取りは、例えば、配列単位合成（SBS）反応、配列決定・ライゲーション反応、又は反復要素の長さ及び／若しくは同一性を決定することが望ましい任意の他の好適な配列決定方法であり得る。サンプル読み取りは、複数のサンプル読み取りに由来するコンセンサス（例えば、平均又は加重）配列であり得る。特定の実施態様では、参照配列を提供することは、PCRアンプリコンのプライマー配列に基づいて目的の遺伝子座を識別することを含む。

30

【1014】

用語「生フラグメント」は、サンプル読み取り又はサンプル断片内の対象とする指定位置又は二次位置に少なくとも部分的に重なり合う、対象とするゲノム配列の一部の配列データを指す。生産物断片の非限定的な例としては、二重ステッチされた断片、単純なステッチされたフラグメント、及び単純な非ステッチの断片が挙げられる。用語「生」は、生のフラグメントがサンプル読み取りにおける配列データとのいくつかの関係を有する配列データを含むことを示すために使用され、生のフラグメントが、サンプル読み取りにおける潜在の変異体に対応し、かつそれを認証又は確認する支持変異体を示すかどうかに関わらず、使用される。用語「生フラグメント」は、フラグメントが、サンプル読み取りにおける変異型コールを検証する支持変異体を必ずしも含むことを示すものではない。例えば、サンプル読み取りが、第1の変異体を呈するために、変異型コールアプリケーションによって判定されるとき、この変異型コールアプリケーションは、1つ又はそれ以上の生のフラグメントが、サンプル読み取りにおける変異体を考慮して、そうでなければ発生することが予想され得る、対応する種類の「支持」変異体を欠くと判定することができる。

40

【1015】

用語「マッピング」、「整列された」、「整列している」、又は「整列する」という用語は、読み取り又はタグを参照配列と比較し、それによって、参照配列が読み取り配列を

50

含むかどうかを判定するプロセスを指す。参照配列が読み取られた場合、読み取りは参照配列にマップされてもよく、又は特定の別の実施態様では、参照配列内の特定の位置にマッピングされてもよい。いくつかの場合において、整列は、読み取りが特定の参照配列のメンバーであるか否か（すなわち、読み取りが参照配列中に存在するか又は存在していないか）かを単に伝える。例えば、ヒト染色体13についての参照配列に対する読み取りの位置合わせは、13番染色体の参照配列中に読み取りが存在するかどうかを伝える。この情報を提供するツールは、設定メンバシップ試験機と呼ばれることがある。場合によっては、位置合わせは、読み取り又はタグマップがある参照配列内の位置を更に示す。例えば、参照配列がヒトゲノム配列全体である場合、アライメントは、染色体13、上にリードが存在することを示してもよく、更に、読み取られたものが染色体13の特定の鎖及び／又は部位にあることを更に示してもよい。

10

【1016】

用語「インデル」は、生物のDNA中の塩基の挿入及び／又は欠失を指す。マイクロインデルは、1～50ヌクレオチドの正味変化をもたらすインデルを表す。インデルの長さが3の倍数でない限り、ゲノムの領域をコードする際に、フレームシフト変異が生じる。インデルは、点突然変異と対比することができる。インデル挿入物は、配列からヌクレオチドを欠失させるが、点変異は、DNA中の全体的な数を変えことなくヌクレオチドのうちの1つを置き換える置換の形態である。インデルはまた、隣接するヌクレオチドにおける置換として定義され得るTandem Base変異（Tandem Base Mutation、TBM）と対比することもできる（主に2つの隣接するヌクレオチドで置換されるが、隣接する3つのヌクレオチドでの置換が観察された）。

20

【1017】

用語「変異体」は、核酸参照とは異なる核酸配列を指す。典型的な核酸配列変異体としては、限定するものではないが、単一のヌクレオチド多型（Single Nucleotide Polymorphism、SNP）、短い欠失及び挿入多型（Indel）、コピー数変動（Copy Number Variation、CNV）、マイクロ衛星マーカー、又は短いタンDEM反復及び構造変異が挙げられる。体細胞変異体コールは、DNAサンプル中に低頻度で存在する変異体を識別するための努力である。体細胞変異体のコーリングは、癌治療の文脈において対象となる。癌は、DNAにおける変異の蓄積によって引き起こされる。腫瘍由来のDNAサンプルは、一般的に不均質であり、いくつかの正常細胞、癌進行の早期段階（より少ない突然変異を伴う）、及び一部の後期細胞（より多くの変異を有する）を含む。この不均一性のため、腫瘍を配列決定するとき（例えば、FFPEサンプルから）、体細胞変異は、多くの場合、低頻度で現れる。例えば、SNVは、所与の塩基を被覆するリードの10%のみに見られ得る。変異体分類子によって体細胞又は生殖細胞系として分類される変異体は、本明細書では「試験中の変異体」とも称される。

30

【1018】

用語「ノイズ」は、配列決定プロセス及び／又は変異型コールアプリケーションにおける1つ又はそれ以上の誤差から生じる、誤りのある変異型コールを指す。

【1019】

用語「変異体頻度」は、集団内の特定の遺伝子座における対立遺伝子（遺伝子の変異体）の相対頻度を表し、分画又は割合として表される。例えば、分画又は割合は、その対立遺伝子を保有する集団中の全ての染色体の割合であってもよい。一例として、サンプル変異体頻度は、個体から対象となるゲノム配列について得られたリード及び／又はサンプルの数に対応する「集団」にわたって、対象とするゲノム配列に沿った特定の遺伝子座／位置における対立遺伝子／変異体の相対頻度を表す。別の例として、ベースライン変異体頻度は、1つ又はそれ以上のベースラインゲノム配列に沿った特定の遺伝子座／位置における対立遺伝子／変異体の相対頻度を表し、ここで、1つ又はそれ以上のベースラインゲノム配列について得られた、1つ又はそれ以上のベースラインゲノム配列に沿った特定の遺伝子座／位置における対立遺伝子／変異体の相対頻度を表す。

40

【1020】

50

用語「変異型対立遺伝子頻度 (Variant Allele Frequency、VAF)」は、変異体を標的位置での全体的な被覆率で割った、配列決定されたリードの割合を指す。VAFは、変異体を保有する配列決定されたリードの割合の尺度である。

【1021】

用語「位置」、「指定位置」、及び「遺伝子座」は、ヌクレオチド配列内の1つ又はそれ以上のヌクレオチドの位置又は座標を指す。用語「位置」、「指定位置」、及び「遺伝子座」はまた、ヌクレオチドの配列における1つ又はそれ以上の塩基対の位置又は座標を指す。

【1022】

用語「ハプロタイプ」は、互いに遺伝する染色体上の隣接する部位における対立遺伝子の組み合わせを指す。ハプロタイプは、存在する場合、所与の座の組の間に生じた組み換え事象の数に応じて、1つの遺伝子座、いくつかの座、又は染色体全体であってもよい。

【1023】

本明細書における「閾値」という用語は、サンプル、核酸、又はその一部（例えば、読み取り）を特徴付けるためのカットオフとして使用される数値又は数値を指す。閾値は、経験的分析に基づいて変化してもよい。閾値は、そのような値を生じさせる源が特定の方法で分類されるべきかどうかを判定するために、測定値又は計算値と比較することができる。閾値は、経験的又は分析的に識別することができる。閾値の選択は、ユーザーが分類を行う必要があることを望む信頼度に依存する。閾値は、特定の目的（例えば、感度及び選択性のバランスのために）選択されてもよい。本明細書で使用する時、用語「閾値」は、分析の過程が変化し得る点、及び／又はアクションがトリガされ得る点を示す。閾値は、所定の数である必要はない。その代わりに、閾値は、例えば、複数の因子に基づく関数であってもよい。閾値は、状況に適応し得る。更に、閾値は、上限、下限、又は限界間の範囲を示し得る。

【1024】

いくつかの実施態様では、配列決定データに基づく指標又はスコアは、閾値と比較され得る。本明細書で使用する時、用語「メトリック」又は「スコア」は、配列決定データから決定された値又は結果を含んでもよく、又は配列決定データから決定された値又は結果に基づく関数を含んでもよい。閾値と同様に、指標又はスコアは、状況に適応し得る。例えば、指標又はスコアは、正規化された値であってもよい。スコア又はメトリックの例として、1つ又はそれ以上の実施態様は、データを分析する際にカウントスコアを使用することができる。計数スコアは、サンプル読み取りの数に基づいてもよい。サンプル読み取りは、サンプル読み取りが少なくとも1つの共通の特性又は品質を有するように、1つ又はそれ以上のフィルタリング段階を経ていてもよい。例えば、計数スコアを決定するために使用されるサンプル読み取りのそれぞれは、参照配列と整列されていてもよく、又は潜在的な対立遺伝子として割り当てられてもよい。共通の特性を有するサンプル読み取りの数をカウントして、読み取りカウントを決定することができる。カウントスコアは、読み取りカウントに基づいてもよい。いくつかの実施態様では、計数スコアは、読み取りカウントと等しい値であってもよい。他の実施例では、計数スコアは、読み取りカウント及び他の情報に基づいてもよい。例えば、計数スコアは、遺伝子座の特定の対立遺伝子の読み取りカウント、及び遺伝子座の合計リード数に基づいてもよい。いくつかの実施態様では、計数スコアは、遺伝子座の読み出しカウント及び以前に得られたデータに基づいてもよい。いくつかの実施態様では、計数スコアは、所定の値間の正規化スコアであってもよい。計数スコアはまた、サンプルの他の遺伝子座からのリードカウントの関数、又は対象となるサンプルと同時に動作された他のサンプルからのリードカウントの関数であってもよい。例えば、計数スコアは、特定の対立遺伝子の読み取りカウント及びサンプル中の他の遺伝子座の読み取りカウント、及び／又は他のサンプルからのリードカウントの関数であってもよい。一例として、他の遺伝子座からのリードカウント及び／又は他のサンプルからのリードカウントを使用して、特定の対立遺伝子についての計数スコアを正規化してもよい。

10

20

30

40

50

【1025】

用語「被覆率」又は「断片被覆率」は、配列の同じフラグメントに対する多数のサンプルリードの計数又は他の尺度を指す。読み取りカウントは、対応するフラグメントをカバーするリードの数のカウントを表し得る。あるいは、被覆率は、履歴知識、サンプルの知識、遺伝子座の知識などに基づく指定された因子を、読み取りカウントに掛けることによって決定されてもよい。

【1026】

用語「読み取り深さ」（従来、「x」が続く数）は、標的位置における重複アラインメントを伴う配列決定されたリードの数を指す。これは、多くの場合、1組の間隔（エキソン、遺伝子、又はパネルなど）のカットオフを超える平均又は百分率として表現される。例えば、臨床報告は、パネル平均被覆率が、標的化されたベースカバー>100×の98%を有する1,105×であると言うことができる。

10

【1027】

用語「ベースコール品質スコア」又は「Qスコア」は、単一の配列決定された塩基が正しい確率に反比例する0-50からの範囲のPHREDスケールされた確率を指す。例えば、20のQを有するTベースコールは、99.99%の確率で正しいと考えられる。Q<20での任意のベースコールは、低品質であると見なされるべきであり、変異体を支持する配列決定されたリードのかなりの割合が低い場合に識別される任意の変異体は、潜在的に偽陽性であると思なされるべきである。

【1028】

用語「変異体リード」又は「変異体リード番号」は、変異体の存在を支持する配列決定されたリードの数を指す。

20

【1029】

「ストリンディティ」(又はDNAストランド)に関して、DNA中の遺伝的メッセージは、文字A、G、C、及びTの文字、例えば、5'-AGGACA-3'として表すことができる。多くの場合、配列は、本明細書に示される方向、すなわち、5'端を左に、3'端を右に書き込む。DNAは、(特定のウイルスのように)一本鎖分子として生じる場合があるが、通常、二本鎖単位としてDNAを見つける。これは、2つの抗平行ストランドを有する二重螺旋構造を有する。この場合、「逆平行」という語は、2つのストランドが平行に動作するが、反対の極性を有することを意味する。二本鎖DNAは、塩基とペアリングによって一緒に保持され、ペアリングは、アデニン(A)対がチミン(T)及びシトシン(C)対とグアニン(G)との対となるように、常に保持される。このペアリングは相補性と呼ばれ、1本のDNA鎖は、他方の相補体であると言われる。したがって、二本鎖DNAは、これと同様に、2つのストランドとして表され得る。5'-AGGACA-3'及び3'-TCCGTG-5'。2つのストランドは、反対の極性を有することに留意されたい。したがって、2つのDNA鎖のストランド性は、基準ストランド及びその補体、順方向及び逆方向ストランド、トップ及びボトムストランド、センス及びアンチセンスストランド、又はWatson及びCrickストランドと呼ぶことができる。

30

【1030】

リードアライメント(リードマッピングとも呼ばれる)は、ゲノム中の配列が由来する場合に、参照するプロセスである。整列が行われると、所与の読み取りの「マッピング品質」又は「マッピング品質スコア(MAPQ)」は、ゲノム上のその位置が正しい確率を定量化する。マッピング品質は、位相スケールで符号化され、Pはアライメントが正しくない確率である。確率は、以下のように計算される。 $P = 10^{(-MAPQ/10)}$ であり、式中、MAPQは、マッピング品質である。例えば、-4の電力に対する $40 = 10$ のマッピング品質は、読み取りが不正確に位置合わせされた0.01%の機会が存在することを意味する。したがって、マッピング品質は、読み取りの基本品質、参照ゲノムの複雑性、及びパレドエンド情報などのいくつかの位置合わせ因子と関連付けられる。最初に、読み取りの基本品質が低い場合、観察された配列が誤っている可能性があり、したがってそのアライメントが誤っていることを意味する。第2に、マッピング能力はゲノムの複雑

40

50

さを指す。反復領域は、これらの領域に含まれるマップ及びリードをマッピングすることがより困難であり、通常、マッピング品質が低くなる。この文脈では、MAPQは、リードが一意的に整列されておらず、それらの実際の原点を決定することができないという事実を反映する。第3に、パリッドエンド配列決定データの場合、コンコダント対は、より良好に整列される可能性が高い。マッピング品質が高いほど、アライメントがより良好である。良好なマッピング品質と整合された読み取りは、通常、読み出し配列が良好であり、高いマッピング可能領域内ではわずかな不一致と位置合わせされたことを意味する。MAPQ値は、アライメント結果の品質管理として使用することができる。20よりも高いMAPQと位置合わせされたリードの割合は、通常、下流分析のためである。

【1031】

本明細書で使用する時、「信号」は、例えば画像内の発光、好ましくは発光などの検出可能な事象を指す。したがって、好ましい別の実施態様では、信号は、画像内に捕捉された任意の検出可能な発光（すなわち、「スポット」）を表すことができる。したがって、本明細書で使用する時、「信号」は、検体の検体からの実際の放出の両方を指すことができ、実際の検体と相関しない擬似発光を指すことができる。したがって、信号はノイズから生じ得、試験片の実際の検体を代表しないように後に廃棄することができる。

【1032】

本明細書で使用する時、用語「塊」は、一群の信号を指す。特定の実施態様では、信号は、異なる検体に由来する。好ましい別の実施態様では、信号塊は、一緒にクラスター化する信号群である。より好ましい実施態様では、シグナル凝集は、1つの増幅オリゴヌクレオチドによって覆われた物理的領域を表す。各信号塊は、理想的には、いくつかの信号（テンプレートサイクル当たり1つ、恐らくはクロストークによってより多く）として観察されるべきである。したがって、2つ（又はそれ以上）の信号が同じ信号の塊からテンプレートに含まれる、重複する信号が検出される。

【1033】

本明細書で使用する時、「最小」、「最大」、「最小化」、「最大化」、及びその文法的変異形などの用語は、絶対最大値又は最小値ではない値を含むことができる。いくつかの実施態様では、値は、最大値及び最小値付近を含む。他の実施例では、値は、局所的な最大値及び／又は局所最小値を含むことができる。いくつかの実施態様では、値は、絶対最大値又は最小値のみを含む。

【1034】

本明細書で使用する時、「クロストーク」は、別個の画像においても検出される1つの画像内の信号の検出を指す。好ましい別の実施態様では、クロストークは、放出された信号が2つの別個の検出チャンネルで検出されるときに発生し得る。例えば、放射された信号が1つの色で発生する場合、その信号の放射スペクトルは、別の色で別の放射された信号と重なってもよい。好ましい実施態様では、ヌクレオチド塩基A、C、G、及びTの存在を示すために使用される蛍光分子は、別個のチャンネルで検出される。しかし、A及びCの発光スペクトルは重複するため、色チャンネルを使用した検出中に、C色信号の一部が検出され得る。したがって、A信号とC信号との間のクロストークにより、1つのカラー画像からの信号が他のカラー画像に現れることを可能にする。いくつかの実施態様では、G及びTクロストークがある。いくつかの実施態様では、チャンネル間のクロストークの量は非対称である。チャンネル間のクロストークの量は、とりわけ、適切な放射スペクトルを有する信号分子の選択、並びに検出チャンネルのサイズ及び波長範囲の選択によって制御され得ることが理解されるであろう。

【1035】

本明細書で使用する時、「レジスタ」、「登録」、「登録」、及び同様の用語は、画像又はデータセット内の信号を、別の時点又は視点からの画像又はデータセット内の信号と相関させるための任意のプロセスを指す。例えば、位置合わせを使用して、画像のセットからの信号を整列させてテンプレートを形成することができる。別の例では、位置合わせを使用して、他の画像からの信号をテンプレートに位置合わせすることができる。1つ

10

20

30

40

50

の信号は、別の信号に直接又は間接的に登録されてもよい。例えば、画像「S」からの信号は、画像「G」に直接登録されてもよい。別の例として、画像「N」からの信号は、画像「G」に直接登録されてもよく、あるいは、画像「N」からの信号は、以前に画像「G」に登録された画像「S」に登録されてもよい。したがって、画像「N」からの信号は、画像「G」に間接的に登録される。

【1036】

本明細書で使用する時、用語「基準」は、物体内又は物体上の区別可能な基準点を意味することを意図する。基準点は、例えば、マーク、第2の物体、形状、縁部、領域、不規則性、チャンネル、ピット、ポストなどであり得る。基準点は、オブジェクトの画像内に、又はオブジェクトを検出することに由来する別のデータセット内に存在することができる。基準点は、物体の平面内のx及び／又はy座標によって指定することができる。代替的に又は追加的に、基準点は、例えば、物体と検出器との相対位置によって定義される、xy平面に直交するz座標によって指定することができる。基準点に対する1つ又はそれ以上の座標は、オブジェクトの1つ又はそれ以上の他の検体、又はオブジェクトに由来する画像又は他のデータセットに対して指定することができる。

【1037】

本明細書で使用する時、用語「光信号」は、例えば、蛍光、発光、散乱、又は吸収信号を含むことを意図する。光信号は、紫外線(Ultraviolet、UV)範囲(約200～390nm)、可視(Visible、VIS)範囲(約391～770nm)、赤外線(Infrared、IR)範囲(約0.771～25マイクロメートル)、又は電磁スペクトルの他の範囲で検出することができる。これらの範囲のうちの1つ又はそれ以上の全て又は一部を除外する方法で、光信号を検出することができる。

【1038】

本明細書で使用する時、用語「信号レベル」は、所望又は所定の特性を有する検出されたエネルギー又は符号化された情報の量又は量を意味することを意図する。例えば、光信号は、強度、波長、エネルギー、周波数、電力、輝度などのうちの1つ又はそれ以上によって定量化することができる。他の信号は、電圧、電流、電界強度、磁場強度、周波数、電力、温度などの特性に従って定量化することができる。信号の不在は、ゼロの信号レベル、又はノイズとは有意に区別されない信号レベルであると理解される。

【1039】

本明細書で使用する時、用語「シミュレートする」は、物理又は行動の特性を予測する物理的又は行動の表現又はモデルを作成することを意味することを意図する。表現又はモデルは、多くの場合、そのもの又は行動と区別可能であり得る。例えば、表現又はモデルは、色、被加工、サイズ、又は形状の全て又は一部から検出される信号の強度などの1つ又はそれ以上の特性に対するものと区別することができる。特定の実施態様では、表現又はモデルは、何か又は行為と比較して、理想化、誇張、ミュート、又は不完全であり得る。したがって、いくつかの実施態様では、モデルの表現は、例えば、上記の特性のうちの少なくとも1つに関して表すものであることを表すものであることができる。表現又はモデルは、本明細書の他の箇所に記載されるもののうちの1つ又はそれ以上などのコンピュータ可読フォーマット又は媒体に提供され得る。

【1040】

本明細書で使用する時、用語「特定の信号」は、背景エネルギー又は情報などの他のエネルギー又は情報にわたって選択的に観察される、検出されたエネルギー又は符号化情報を意味することを意図する。例えば、特定の信号は、特定の強度、波長、若しくは色で検出される光信号、特定の周波数、電力若しくは場強度で検出される電気信号、又は分光及び分析検出に関する技術分野において既知の他の信号であり得る。

【1041】

本明細書で使用する時、用語「スイング」は、物体の矩形部分を意味することを意図する。スイングは、ストリップの最長寸法に平行な方向に、物体と検出器との間の相対移動によって走査される細長いストリップであり得る。一般に、矩形部分又はストリップの

幅は、その全長に沿って一定である。物体の複数のスエージは、互いに平行であってもよい。物体の複数のスエージは、互いに重なり合い、互いに隣接するか、又は間質領域によって互いに分離され得る。

【1042】

本明細書で使用する時、用語「分散」は、予想される差、及び観察される差、又は2つ又はそれ以上の観測結果間の差を意味することを意図する。例えば、分散は、期待値と測定値との間の不一致であり得る。標準偏差、標準偏差の二乗、変動係数などの統計関数を使用して、分散を表すことができる。

【1043】

本明細書で使用する時、用語「x y 座標」は、x y 平面内の位置、サイズ、形状、及び／又は向きを指定する情報を意味することを意図する。情報は、例えば、デカルトシステムにおける数値座標であり得る。座標は、x 軸及び y 軸の一方又は両方に対して提供することができ、又は x y 平面内の別の場所に対して提供することができる。例えば、物体の検体の座標は、対象物の基準又は他の検体の位置に対する検体の場所を指定することができる。

【1044】

本明細書で使用する時、用語「x y 平面」は、直線軸 x 及び y によって画定される2次元領域を意味することを意図する。検出器及び検出器によって観察される物体を参照して使用される場合、検出器と検出されている物体との間の観測方向に直交するように更に指定することができる。

【1045】

本明細書で使用する時、用語「z 座標」は、x y 平面に直交する軸に沿った点、線、又は領域の位置を指定する情報を意味することを意図する。特定の別の実施態様では、z 軸は、検出器によって観察される物体の領域に直交する。例えば、光学系の焦点の方向は、z 軸に沿って指定されてもよい。

【1046】

いくつかの実施態様では、獲得された信号データは、アフィン変換を用いて変換される。いくつかのそのような実施態様では、テンプレートの生成は、色チャンネル間のアフィン変換が動作間に一貫しているという事実を使用する。この一貫性のため、標本中の検体の座標を決定する際に、デフォルトオフセットのセットを使用することができる。例えば、デフォルトオフセットファイルは、A チャンネルなどの1つのチャンネルに対する異なるチャンネルに対する相対変換（シフト、スケール、スキュー）を含むことができる。しかしながら、他の実施態様では、動作中及び／又は動作間の色チャンネルドリフト間のオフセットは、オフセット駆動型テンプレート生成を困難にする。このような実施例では、本明細書で提供される方法及びシステムは、オフセットしたテンプレート生成を利用することができ、これについては以下で更に説明する。

【1047】

上記の実施態様のいくつかの態様では、システムはフローセルを含み得る。一部の態様では、フローセルは、レーン、又は他の構成のタイルを含み、タイルの少なくとも一部は、1つ又はそれ以上の検体群を含む。一部の態様では、検体は、核酸などの複数の分子を含む。特定の態様では、フローセルは、標識されたヌクレオチド塩基を核酸の配列に送達し、それによって、核酸を含む検体に対応するシグナルを生成するように、検体内の核酸にハイブリダイズするプライマーを伸長させるように構成される。好ましい実施態様では、検体内の核酸は、互いに同一又は実質的に同一である。

【1048】

本明細書に記載される画像解析システムのいくつかにおいて、画像のセット内の各画像は、色信号を含み、異なる色は、異なるヌクレオチド塩基に対応する。一部の態様では、画像のセットの各画像は、少なくとも4つの異なる色から選択される単一の色を有する信号を含む。一部の態様では、画像のセット内の各画像は、4つの異なる色から選択される単一の色を有する信号を含む。本明細書に記載されるシステムのいくつかにおいて、核酸

10

20

30

40

50

は、4つの異なる画像を生成するように、4つの異なる標識ヌクレオチド塩基を分子の配列に提供することにより、核酸を配列決定することができ、各画像は単一の色を有するシグナルを含み、信号色が、4つの異なる画像のそれぞれに対して異なることにより、核酸内の特定の位置に存在する4つの可能なヌクレオチドに対応する4つのカラー画像のサイクルを生成する、方法。特定の態様では、システムは、追加の標識ヌクレオチド塩基を分子の配列に送達するように構成されたフローセルを含み、それによって複数のカラー画像のサイクルを生成する。

【1049】

好ましい実施態様形態では、本明細書で提供される方法は、プロセッサがアクティブにデータを取得しているか、又はプロセッサが低活動状態にあるかどうかを判定することを含み得る。多数の高品質画像を取得及び記憶することは、典型的には、大量の記憶容量を必要とする。更に、取得され記憶されると、画像データの分析はリソース集約的になり得、追加の画像データの取得及び記憶などの他の機能の処理能力を妨げる可能性がある。したがって、本明細書で使用するとき、用語「低活動状態」は、所与の時間におけるプロセッサの処理能力を指す。いくつかの実施態様では、低活動状態は、プロセッサがデータを取得及び／又は記憶していないときに生じる。いくつかの実施態様では、一部のデータ取得及び／又はストレージが行われる場合には、低いアクティビティ状態が生じるが、他の機能に干渉することなく画像解析が同時に生じ得るように、追加の処理能力が残る。

【1050】

本明細書で使用するとき、「競合を識別する」とは、複数のプロセスがリソースに対して競合する状況を識別することを指す。いくつかのそのような実施態様では、1つのプロセスは、別のプロセスに対して優先度を与えられる。いくつかの実施態様では、競合は、時間、処理能力、記憶能力、又は優先度が与えられる任意の他のリソースの割り当てに対する優先度を与える必要性に関連し得る。したがって、いくつかの実施態様では、処理時間又は容量が、データセットを分析し、データセットを取得及び／又は記憶するかのいずれかなどの2つのプロセス間に分散される場合、2つのプロセス間の不一致が存在し、プロセスのうちの1つに優先度を与えることによって解決することができる。

【1051】

本明細書では、画像解析を実行するためのシステムも提供される。システムは、プロセッサと、記憶容量と、画像解析用のプログラムと、を含むことができ、プログラムは、記憶のための第1のデータセット及び分析のための第2のデータセットを処理するための命令を含み、処理は、記憶装置上の第1のデータセットを取得及び／又は記憶することと、プロセッサが第1のデータセットを取得していないときに第2のデータセットを解析することと、を含む。特定の態様では、プログラムは、第1のデータセットを取得及び／又は記憶することと、第2のデータセットを解析することとの間の競合の少なくとも1つのインスタンスを識別するための命令を含み、第1のデータセットを取得及び／又は記憶することが優先度与えられるように、画像データを取得及び／又は記憶することが優先される。特定の態様では、第1のデータセットは、光学撮像装置から取得された画像ファイルを含む。特定の態様では、システムは、光学撮像装置を更に備える。一部の態様では、光学撮像装置は、光源と検出デバイスとを備える。

【1052】

本明細書で使用するとき、用語「プログラム」は、タスク又はプロセスを実行するための命令又はコマンドを指す。用語「プログラム」は、用語「モジュール」と互換的に使用され得る。特定の態様では、プログラムは、同じコマンドセットの下で実行される様々な命令のコンパイルであり得る。他の実施態様では、プログラムは、別個のバッチ又はファイルを参照することができる。

【1053】

以下に記載されるのは、本明細書に記載される画像解析を実行するための方法及びシステムを利用する驚くべき効果の一部である。いくつかの配列決定の実現例では、配列決定システムの有用性の重要な尺度は、その全体的な効率である。例えば、1日当たりに生成

10

20

30

40

50

されるマッピング可能なデータの量、並びに器具の設置及び動作の総コストは、経済的な配列決定ソリューションの重要な態様である。マッピング可能なデータを生成し、システムの効率を高めるための時間を短縮するために、リアルタイムのベースコールを機器コンピュータ上で有効にすることができ、配列決定ケミストリー及び画像化と並行して動作することができる。これにより、配列決定化学仕上げの前に、データ処理及び分析が完了することを可能にする。更に、中間データに必要な記憶を低減し、ネットワークを横切って移動する必要があるデータの量を制限することができる。

【1054】

配列出力が増加している間、本明細書で提供されるシステムからネットワークに転送された動作ごとのデータ、及び二次分析処理ハードウェアは、実質的に減少している。機器コンピュータ（取得コンピュータ）上でデータを変換することにより、ネットワークロードが劇的に低減される。これらのオン機器、オフネットワークデータ低減技術を伴わずに、DNA配列決定機器のフレッツの画像出力は、ほとんどのネットワークをクリップするであろう。

【1055】

ハイスループットDNA配列決定機器の広範な採用は、使用の容易さ、用途の範囲に対する支持、及び実質的に任意のlab環境に対する適合性によって、部分的に駆動されてきた。本明細書に提示される高度に効率的なアルゴリズムは、配列決定インスツルメントを制御することができる単純なワークステーションに、有意な分析機能を加えることを可能にする。計算ハードウェアの必要条件のこの低減は、配列決定出力レベルが増加し続けるにつれて、更に重要となる、いくつかの実用的な利点を有する。例えば、単純なタワー、熱生成、実験室設置面積、及び電力消費を最小限に抑えるために、画像解析及びベースコールを行うことによって、最小に保たれる。対照的に、他の商業的な配列決定技術は、1次分析のために、最大5回の処理電力で、そのコンピューティングインフラストラクチャを最近ランプアップして、熱出力及び電力消費の増加を開始する。したがって、いくつかの実施態様では、本明細書で提供される方法及びシステムの計算効率は、サーバハードウェアを最小限に抑えながら、それらの配列決定スループットを増加させることを可能にする。

【1056】

したがって、いくつかの実施態様では、本明細書に提示される方法及び／又はシステムは、状態マシンとして機能し、各試料の個々の状態の追跡を保ち、試料が次の状態に進む準備ができていることを検出すると、適切な処理を行い、試料をその状態に前進させる。状態マシンがファイルシステムを監視して、好ましい実施例に従って試料が次の状態に進む準備ができているかを判定する方法のより詳細な例が、以下の実施例1に記載されている。

【1057】

好ましい実施態様では、本明細書で提供される方法及びシステムは、マルチスレッドであり、構成可能な数のスレッドと協働することができる。したがって、例えば、核酸配列決定の文脈において、本明細書で提供される方法及びシステムは、リアルタイム分析のためのライブ配列決定動作中に背景において作用することができ、又はオフライン分析のために既存の画像データセットを使用して動作することができる。特定の好ましい実施態様では、方法及びシステムは、それぞれのスレッドを、それが関与する検体のそれ自体のサブセットを与えることによって、マルチスレッドを取り扱う。これにより、スレッド保持の可能性が最小限に抑えられる。

【1058】

本開示の方法は、検出装置を使用して物体の標的画像を取得する工程を含むことができ、この画像は、オブジェクト上の検体の繰り返しパターンを含む。表面の高解像度撮像が可能な検出装置が特に有用である。特定の実施態様では、検出装置は、本明細書に記載される密度、ピッチ、及び／又は検体サイズにおける検体を区別するのに十分な分解能を有するであろう。表面から画像又は画像データを得ることができる検出装置が特に有用であ

10

20

30

40

50

る。例示的な検出器は、物体と検出器とを静的関係に維持しつつ、面積画像を取得するように構成されたものである。走査装置も使用することができる。例えば、連続領域画像を取得する装置（例えば、「ステップ及びショット」検出器と呼ばれる）を使用することができる。また、物体の表面上の点又は線を連続的に走査して、表面の画像を構築するためにデータを蓄積するデバイスも有用である。点走査検出器は、表面の $x-y$ 平面内のラスト運動を介してオブジェクトの表面上の点（すなわち、小さい検出領域）を走査するように構成することができる。線走査検出器は、物体の表面の y 次元に沿った線を走査するように構成ことができ、この線の最長寸法は、 x 次元に沿って生じる。検出デバイス、物体、又はその両方を移動させて、走査検出を達成できることが理解されるであろう。例えば核酸配列決定用途において特に有用な検出装置は、米国特許出願公開第2012/0270305号明細書、米国特許出願公開第2013/0023422号明細書、及び米国特許出願公開第2013/0260372号明細書、及び米国特許第5,528,050号明細書、米国特許第5,719,391号明細書、米国特許第8,158,926号明細書及び米国特許第8,241,573号明細書に記載されており、これらはそれぞれ、参照により本明細書に組み込まれる。

【1059】

本明細書に開示される実施態様は、ソフトウェア、ファームウェア、ハードウェア、又はそれらの任意の組み合わせを生成するためのプログラミング技術又は工学技術を使用して、製造方法、装置、システム、又は物品として実装されてもよい。本明細書で使用する
とき、用語「製造物品」は、光学記憶デバイスなどのハードウェア又はコンピュータ可読
媒体、並びに揮発性又は不揮発性メモリデバイス内に実施態様されるコード又は論理を指
す。そのようなハードウェアとしては、フィールドプログラマブルゲートアレイ（FPGA）、粗粒度再構成可能構造（CGRA）、特定用途向け集積回路（ASIC）、複合プ
ログラマブル論理デバイス（Complex Programmable Logic Device、CPLD）、プ
ログラマブルロジックアレイ（Programmable Logic Array、PLA）、マイクロプロ
セッサ、又は他の同様の処理装置が挙げられるが、これらに限定されない。特定の実施態
様では、本明細書に記載される情報又はアルゴリズムは、非一過性記憶媒体中に存在する。

【1060】

特定の実施態様形態では、本明細書に記載されるコンピュータ実装の方法は、物体の複
数の画像が取得されている間、リアルタイムで発生することができる。このようなリアル
タイム分析は、核酸配列が流体及び検出工程の繰り返しサイクルに供される核酸配列決定
用途に特に有用である。配列決定データの分析は、多くの場合、本明細書に記載される方
法をリアルタイム又は背景で実行するのに有益であり得る一方で、他のデータ獲得又は分
析アルゴリズムがプロセス中である間に、本明細書に記載される方法を実行することが有
益であり得る。本方法で実行することができるリアルタイム分析法の例は、Illumina, Inc (San Diego, Calif.) から市販されており、及び／又は、
参照により本明細書に組み込まれる米国特許出願公開第2012/0020537号明細
書に記載されているMiSeq及びHiSeq配列決定機器に使用されるものである。

【1061】

1つ又はそれ以上のプログラムされたコンピュータによって形成され、本明細書に記載
される方法の1つ又はそれ以上の工程を実行するために実行されるコードを有するプログ
ラミングが、1つ又はそれ以上の機械可読媒体上に記憶されている、例示的データ分析シ
ステム。一実施態様では、例えば、システムは、標的オブジェクトからデータを取得する
ように構成された1つ又はそれ以上の検出システム（例えば、光学撮像システム）へのシ
ステムのネットワークを可能にするように設計されたインターフェースを含む。インター
フェースは、適切な場合には、データを受信及び条件することができる。特定の実施
態様では、検出システムは、例えば、アレイ又は他の物体の画像と一緒に形成する個々の
画像要素又はピクセルを表す画像データを出力する。プロセッサは、処理コードによって
定義された1つ又はそれ以上のルーチンに従って、受信した検出データを処理する。処理
コードは、様々な種類のメモリ回路に記憶されてもよい。

【1062】

現時点で企図される実施態様によれば、検出データ上で実行される処理コードは、検出データを分析して、データ内で可視又は符号化された個々の検体の場所、及び検体が検出されない場所（すなわち、検体が存在しないか、又は、既存の検体から有意な信号が検出されない場所）及びメタデータを判定するように設計されたデータ分析ルーチンを含む。特定の実施態様では、アレイ内の検体位置は、典型的には、撮像された検体に付着した蛍光染料の存在に起因して、非検体位置よりも明るく見える。検体は、例えば、検体におけるプローブの標的が検出されているアレイ内に存在しない場合、検体は、それらの周囲領域よりも明るく見える必要はないことが理解されるであろう。個々の検体が現れる色は、使用される染料、並びに撮像目的のために撮像システムによって使用される光の波長の関数であり得る。標的が結合されていない、又は特定のラベルを有さない検体は、マイクロアレイ内の予想される場所などの他の特性に従って識別することができる。

10

【1063】

データ分析ルーチンがデータ中に個々の検体を配置すると、値割り当てが実行され得る。一般に、値割り当ては、対応する場所における検出器構成要素（例えば、ピクセル）によって表されるデータの特性に基づいて、各検体にデジタル値を割り当てる。すなわち、例えば、画像化データが処理されるとき、値割り当てルーチンは、特定の場所で特定の色又は波長の光が特定の場所で検出されたことを認識するように設計されてもよい。典型的なDNA画像化用途では、例えば、4つの共通ヌクレオチドは、4つの別個の区別可能な色によって表される。次いで、各色は、そのヌクレオチドに対応する値を割り当てられてもよい。

20

【1064】

本明細書で使用するとき、用語「モジュール」、「システム」、又は「システムコントローラ」は、1つ又はそれ以上の機能を実行するように動作するハードウェア及び／又はソフトウェアシステム及び回路を含み得る。例えば、モジュール、システム、又はシステムコントローラは、コンピュータメモリなどの有形及び非一時的コンピュータ可読記憶媒体上に記憶された命令に基づいて動作を実行する、コンピュータプロセッサ、コントローラ、又は他のログベースのデバイスを含んでもよい。あるいは、モジュール、システム、又はシステムコントローラは、有線論理及び回路に基づいて動作を実行する、有線デバイスを含んでもよい。添付の図面に示されるモジュール、システム、又はシステムコントローラは、ソフトウェア又は結線命令に基づいて動作するハードウェア及び回路、ハードウェアを動作させるように命令するソフトウェア、又はそれらの組み合わせを表し得る。モジュール、システム、又はシステムコントローラは、1つ又はコンピュータマイクロプロセッサなどの1つ又はそれ以上のプロセッサを含む、及び／又はそれと接続されるハードウェア回路又は回路を含むか、又は表すことができる。

30

【1065】

本明細書で使用するとき、用語「ソフトウェア」及び「ファームウェア」は互換可能であり、RAMメモリ、ROMメモリ、EPROMメモリ、EEPROMメモリ、及び不揮発性RAM（NVRAM）メモリを含むコンピュータによって実行されるメモリに記憶された任意のコンピュータプログラムを含む。上記メモリタイプは単なる例であり、コンピュータプログラムの記憶に使用可能なメモリの種類に限定されるものではない。

40

【1066】

分子生物学分野では、使用中の核酸配列決定のためのプロセスの1つは、配列番号合成である。この技術は、非常に平行な配列決定プロジェクトに適用することができる。例えば、自動プラットフォームを使用することにより、数百万の配列決定反応を同時に行うことが可能である。したがって、本発明の実施態様の1つは、核酸配列決定中に生成された画像データを取得、記憶、及び分析するための器具及び方法に関する。

【1067】

取得及び記憶することができるデータ量の莫大なゲインは、合理化された画像解析方法を更により有益にする。例えば、本明細書に記載される画像解析方法は、設計者及びエン

50

ドユーザーの両方が、既存のコンピュータハードウェアの効率的な使用を行うことを可能にする。したがって、本明細書では、迅速に増加するデータ出力の面における処理データの計算量を低減する方法及びシステムが本明細書に提示される。例えば、DNA配列決定の分野では、収率は最近の過程で15倍に拡大され、DNA配列決定デバイスの単一の動作において何百ギガーゼに達する可能性がある。計算インフラストラクチャの要件が比例的に増加した場合、大規模なゲノム規模の実験は、ほとんどの研究者に到達していない。したがって、より多くの生配列データの生成は、二次分析及びデータ記憶の必要性を増加させ、データ輸送及び記憶の最適化を非常に有益にする。本明細書に提示される方法及びシステムのいくつかの実施態様は、使用可能な配列データを生成するために必要な時間、ハードウェア、ネットワーキング、及び実験室インフラストラクチャ要件を低減することができる。

10

【1068】

本開示は、方法を実行するための様々な方法及びシステムを説明する。方法のいくつかの例は、一連の工程として記載される。しかしながら、実施態様は、本明細書に記載される特定の工程及び／又は工程の順序に限定されないことを理解されたい。工程は省略されてもよく、工程は修正されてもよく、及び／又は他の工程が追加されてもよい。更に、本明細書に記載される工程を組み合わせることができ、工程は同時に実行されてもよく、工程は同時に実行されてもよく、工程は複数のサブ工程に分割されてもよく、工程は、異なる順序で実行されてもよく、又は工程（又は一連の工程）は、反復的に再実行されてもよい。加えて、本明細書には異なる方法が記載されているが、他の実施態様では、異なる方法（又は異なる方法の工程）を組み合わせてもよいことを理解されたい。

20

【1069】

いくつかの実施態様では、タスク又は動作を実行するように「構成された」処理ユニット、プロセッサ、モジュール、又はコンピューティングシステムは、タスク又は動作を実行するように特に構造化されていると理解され得る（例えば、タスク又は動作を実行するように調整又は意図された、及び／又はタスク若しくは動作を実行するように調整若しくは意図された1つ又はそれ以上のプログラム又は命令を有すること、及び／又はタスク若しくは動作を実行するように調整又は意図された処理回路の配置を有する）。明確さ及び疑義の回避のために、汎用コンピュータ（適切にプログラムされた場合にタスク又は動作を実行するように構成された）となり得る）汎用コンピュータは、タスク又は動作を実行するために具体的にプログラム又は構造的に変更されない限り、タスク又は動作を実行するように「構成されている」ように構成されていない）。

30

【1070】

更に、本明細書に記載される方法の操作は、操作が、商業的に妥当な時間期間内に、当業者には、平均的なヒト又は当業者によって実施されることができないように、十分に複雑であり得る。例えば、本方法は、そのような人が商業的に妥当な時間内で方法を完了できないように、比較的複雑な計算に依存し得る。

【1071】

本出願全体を通して、様々な刊行物、特許、又は特許出願が参照されている。これらの出版物の全体の開示は、本発明が属する技術分野の状態をより完全に説明するために、本出願において参照により本明細書に組み込まれる。

40

【1072】

用語「含む (comprising)」は、本明細書では、列挙された要素のみならず、任意の追加の要素を更に包含する、オープンエンドであることが意図される。

【1073】

本明細書で使用するとき、用語「それぞれ」は、項目の集合を参照して使用されるとき、集合内の個々の項目を識別することを意図しているが、必ずしも集合内の全ての項目を指すものではない。明示的な開示又は文脈がそうでないことを明確に指示する場合、例外が生じ得る。

【1074】

50

上記の実施例を参照して本発明を説明したが、本発明から逸脱することなく様々な修正を行うことができることを理解されたい。

【1075】

本出願のモジュールは、ハードウェア又はソフトウェアで実装することができ、図に示されるように、正確に同じブロックで分割される必要はない。いくつかは、異なるプロセッサ若しくはコンピュータ上に実装されてもよく、又は多数の異なるプロセッサ若しくはコンピュータの中で広がることもできる。加えて、モジュールの一部は、達成される機能に影響を及ぼすことなく、図に示されるものとは並行して、又は異なる順序で操作され得ることが理解されるであろう。また、本明細書で使用するとき、用語「モジュール」は、モジュールを構成するために本明細書で考慮することができる、「サブモジュール」を含むことができる。モジュールとして指定された図のブロックはまた、方法におけるフローチャート工程と考えることができる。

10

【1076】

本明細書で使用するとき、情報項目の「識別」は、その情報の項目の直接仕様を必ずしも必要としない。情報は、単に、一方向の1つ又はそれ以上の層を通じた実際の情報を単に参照することによって、又は情報の実際の項目を決定するのに十分である異なる情報の1つ又はそれ以上のアイテムを識別することによって、フィールド内で「識別され得る」ことができる。加えて、用語「指定する」は、本明細書では、「識別する」と同じであることを意味する。

【1077】

本明細書で使用するとき、所与の信号、イベント又は値は、「前デセサー信号、イベント又は前デセサー信号の値、所与の信号、イベント、又は値によって影響されるイベント又は値に依存する。介在処理要素、工程又は期間が存在する場合、所与の信号、イベント、又は値は、「前デセサー信号、イベント又は値」に依存して「存在する」ことができる。介在処理要素又は工程が2つ又はそれ以上の信号、イベント、又は値を組み合わせる場合、処理要素又は工程の信号出力は、「信号、イベント、又は値入力のそれぞれ」に依存していると見なされる。所与の信号、イベント又は値が前デセサー信号、イベント又は値と同じである場合、これは単に、所与の信号、イベント、又は値が「前デセサー信号、イベント又は値」に依存して「依存して」又は「依存して」又は「ベースデセサー信号、イベント又は値」に基づいて、「依存して」又は「依存する」と見なされる。別の信号、イベント又は値に対する所与の信号、イベント、又は値の「応答性」は、同様に定義される。

20

30

【1078】

本明細書で使用するとき、「並行して」又は「並行して」は、正確な同時性を必要としない。個人の1人の評価が、個人の別の評価が完了する前に開始する場合に十分である。

(特定の改善)

【1079】

我々は、ニューラルネットワークベースのテンプレート生成及びニューラルネットワークベースのベースコールの様々な実施態様を説明している。実施態様の1つ又はそれ以上の特徴を、塩基実施態様と組み合わせることができる。相互に排他的でない実施態様は、組み合わせ可能であると教示されている。実施態様の1つ又はそれ以上の特徴を他の実施態様と組み合わせることができる。本開示は、これらのオプションのユーザーを定期的に通知する。これらの選択肢を繰り返す列挙のいくつかの実施態様からの省略は、前述のセクションで教示されている組み合わせを制限するものとして解釈されるべきではない。これらの記載は、以下の実施のそれぞれに参照することにより本明細書に組み込まれる。

40

(サブピクセルベースコール)

【1080】

我々は、フローセルのタイル上の検体についてメタデータを決定するコンピュータ実装の方法を開示している。本方法は、配列決定動作中に生成された一連の画像セットにアクセスすることを含み、各画像セットは、配列決定動作のそれぞれの配列決定サイクル中に生成されたシリーズ内に設定された一連の画像セットにアクセスすることを含み、シリー

50

ズ中の各画像は複数のサブピクセルを有する。本方法は、ベースコーラーから、4つの塩基（A、C、T、及びG）のうちの1つとしてサブピクセルのそれぞれを分類するベースコールを取得することを含み、それによって、配列決定動作の複数の配列決定サイクルにわたってサブピクセルのそれぞれに対してベースコール配列を生成することを含む。本方法は、実質的に一致するベースコール配列を共有する連続するサブピクセルの不連続領域として検体を識別する検体マップを生成することを含む。本方法は、不連続領域に基づいて、それらの形状及びサイズを含む検体の空間分布を決定することと、分類子を訓練するためのグラウンドトゥールズとして使用するために、メモリに検体マップを記憶することと、を含む。

【1081】

開示されるこのセクション及び技術の他のセクションに記載される方法は、開示される追加の方法に関連して説明される以下の特徴及び／又は特徴のうちの1つ又はそれ以上を含むことができる。簡潔性の目的で、本出願に開示される特徴の組み合わせは、個別に列挙されず、特徴の各ベースセットで繰り返されない。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1082】

一実施態様では、本方法は、検体マップ内のサブピクセルを、非接合領域のいずれにも属しないサブピクセルを識別することを含む。一実施態様では、本方法は、ベースコーラーから、5つの塩基（A、C、T、G、及びN）のうちの1つとして、サブピクセルのそれぞれを分類することを含む。一実施態様では、検体マップは、ベースコール配列が実質的に一致しない2つの連続するサブピクセル間の検体境界部分を識別する。

【1083】

一実施態様では、本方法は、ベースコーラーによって判定された検体の予備中心座標における原点サブピクセルを識別することと、原点サブピクセルから始まり連続的に連続する非原点サブピクセルを継続することによって、ベースコール配列を実質的に一致させるための、第1の検索と、を含む。一実施態様では、方法は、検体ベースで検体の中心座標を決定することと、検体マップの不連続領域の質量中心を、不連続領域を形成するそれぞれの連続するサブピクセルの座標の平均として計算することと、検体上の検体の超位置中心座標を検体によって検体ベースで記憶して、分類子を訓練するためのグラウンドトゥールズとして使用することと、を含む。

【1084】

一実施態様では、方法は、検体ベースで検体マップの不連続領域内の質量サブピクセルの中心を識別することと、検体マップを補間を用いてアップサンプリングし、分類部を訓練するためにグラウンドトゥールズとして使用するために、メモリ内にアップサンプリングされた検体マップを記憶することと、を含み、検体による検体ベースのアップサンプリングされた検体マップにおいて、検体ごとに、隣接するサブピクセルが属する不連続領域内の質量サブピクセルの中心からの隣接するサブピクセルの距離に比例する減衰係数に基づいて、不連続領域内の各連続サブピクセルに値を割り当てることと、を含む。一実施態様では、値は、ゼロと1との間で正規化された強度値である。一実施態様では、方法は、アップサンプリングされた検体マップにおいて、背景として識別された全てのサブピクセルに同じ所定の値を割り当てることと、を含む。一実施態様では、所定の値はゼロ強度値である。

【1085】

一実施態様では、本方法は、分離された領域内の連続するサブピクセルを発現するアップサンプリングされた検体マップから減衰マップを生成することと、その割り当てられた値に基づいて背景として識別されたサブピクセルとから減衰マップを生成することと、分類部を訓練するためのグラウンドトゥールズとして使用するために、メモリ内に減衰マップを記憶することと、を含む。一実施態様では、減衰マップ内の各サブピクセルは、ゼロと1との間で正規化された値を有する。一実施態様では、本方法は、アップサンプリング

10

20

30

40

50

された検体マップにおいて、検体によって検体ベースで、同じ検体に属する検体内部サブピクセルとして、不連続領域内の連続するサブピクセルを検体中心サブピクセルとして分類することと、検体境界部分を境界サブピクセルとして含み、サブピクセルが背景サブピクセルとして背景として識別され、分類部を訓練するためのグラウンドトゥールズとして使用するために、メモリ内に分類を記憶することと、を含む。

【1086】

一実施態様では、方法は、検体による検体ベース、検体内部サブピクセルの座標、検体中心サブピクセル、境界サブピクセル、及び背景サブピクセルを、検体に基づいて検体に基づいて記憶することと、検体マップをアップサンプリングするために使用される因子によって座標をダウンスケールすることと、分類子を訓練するためのグラウンドトゥールズとして使用するために、メモリにダウンスケールされた座標を記憶することと、を含む。

10

【1087】

一実施態様では、方法は、アップサンプリングされた検体マップから生成されたバイナリグラウンドトゥールズデータにおいて、検体中心クラスに属する検体中心サブピクセルをラベル化するために色符号化を使用して、検体中心サブピクセルをラベル化することと、分類部を訓練するためのグラウンドトゥールズとして使用するために、メモリ内にバイナリグラウンドトゥールズデータを記憶することと、を含む。一実施態様では、方法は、背景クラスに属する背景サブピクセルをラベル化するためにカラーコーディングを使用して、アップサンプリングされた検体マップから生成された三元グラウンドトゥールズデータにおいて、背景クラスに属する背景サブピクセルをラベル化することと、検体中心サブピクセルは、検体中心クラスに属するものであり、検体内部サブピクセルは、検体内部クラスに属するものとして、及び分類子を訓練するためのグラウンドトゥールズとして使用するために、メモリ内に三元グラウンドトゥールズデータを記憶することと、を含む。

20

【1088】

一実施態様では、本方法は、フローセルの複数のタイルの検体マップを生成することと、検体マップをメモリに保存することと、それらの形状及びサイズを含む検体マップに基づいて、タイル内の検体の空間分布を決定することと、検体のアップサンプリングされた検体マップにおいて、検体によって検体ベースで、同じ検体、検体中心サブピクセル、境界サブピクセル、及び背景サブピクセルに属する検体内部サブピクセルとして分類することと、分類子を訓練するためのグラウンドトゥールズとして使用するためにメモリに分類を記憶することとあって、検体によって検体ベースで、検体内部サブピクセルの座標、検体中心サブピクセル、境界サブピクセルを記憶することと、分類子を訓練し、検体マップをアップサンプリングするために使用される係数によって座標をダウンスケールするために、メモリ内の背景サブピクセルが、検体マップをアップサンプリングするために使用される因子によって座標をダウンスケールすることと、分類子を訓練するためのグラウンドトゥールズとして使用するために、メモリにダウンスケールされた座標を記憶することと、を含む。

30

【1089】

一実施態様では、ベースコール配列は、ベースコールの所定の部分が、順序の位置ごとに一致するときに実質的に一致する。一実施態様では、ベースコーラーは、最近傍強度抽出、ガウス系強度抽出、平均 2×2 サブピクセル領域に基づく強度抽出のうちの少なくとも1つを含む、サブピクセルの強度を補間することによって、ベースコール配列を生成し、 2×2 サブピクセル面積の最も明るい試験に基づく強度抽出、平均 3×3 サブピクセル面積、バイリニア強度抽出、双キュービック強度抽出、及び／又は加重面積被覆率に基づく強度抽出に基づく強度抽出。一実施態様では、サブピクセルは、それらの整数又は非整数座標に基づいて、ベースコーラーに識別される。

40

【1090】

一実施態様では、本方法は、不連続領域の少なくとも一部が所定の最小数のサブピクセルを有することを必要とすることを含む。一実施態様では、フローセルは、検体を占有するウェルのアレイを有する、少なくとも1つのパターン化された表面を有する。そのよう

50

な実施態様では、本方法は、検体の決定された形状及びサイズに基づいて、ウェルのうちのどれが、ウェルのうちの1つが最小限に占有されている少なくとも1つの検体によって実質的に占有され、ウェルのうちの1つが、複数の検体によって共占有される。

【1091】

一実施態様では、フローセルは、少なくとも1つのパターン化されていない表面を有し、検体は、非パターン化表面上で不均一に散乱される。一実施態様では、検体の密度は、約100,000検体/mm²～約1,000,000検体/mm²の範囲である。一実施態様では、検体の密度は、約1,000,000検体/mm²～約10,000,000検体/mm²の範囲である。一実施態様では、サブピクセルは4分の1サブピクセルである。別の実施態様では、サブピクセルは半サブピクセルである。一実施態様では、ベースコーラーによって決定される検体の予備中心座標は、タイルのテンプレート画像内に定義され、画像座標系のピクセル解像度、画像座標系、及び測定スケールは、テンプレート画像及び画像と同じである。一実施態様では、各画像セットは、4つの画像を有する。別の実施態様では、各画像セットは2つの画像を有する。更に別の実施態様では、各画像セットは1つの画像を有する。一実施態様では、配列決定動作は、4チャンネル化学を利用する。別の実施態様では、配列決定動作は、2チャンネル化学を利用する。更に別の実施態様では、配列決定動作は、1チャンネル化学を利用する。

【1092】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1093】

我々は、フローセルのタイル上の検体についてメタデータを決定するコンピュータ実装の方法を開示している。本方法は、配列決定動作中に捕捉されたタイルの画像セット、及びベースコーラーによって決定された検体の予備中心座標にアクセスすることを含む。本方法は、各画像セットについて、基本的な中心座標を含む4つの基本原点サブピクセルのうちの1つとして、原点サブピクセルのそれぞれに連続的に連続している連続するサブピクセルの所定の近傍を含む4つの基本原点サブピクセルのうちの1つとして取得することと、それによって、ソースサブピクセルのそれぞれ、及び連続するサブピクセルの所定の近傍のそれぞれに対して、ベースコール配列を生成する。本方法は、原点サブピクセルのうちの対応する1つの少なくとも一部に連続的に隣接し、かつ、4つの塩基のうちの1つの実質的に一致するベースコール配列を、原点サブピクセルのうちの対応する1つの少なくとも一部と共有する、連続するサブピクセルの不連続領域として、検体を識別する検体マップを生成することを含む。本方法は、検体マップをメモリに保存することと、検体マップ内の不連続領域に基づいて、検体の形状及びサイズを決定することと、を含む。

【1094】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1095】

一実施態様では、連続するサブピクセルの所定の近傍は、原点サブピクセルを含むピクセルを中心とするm×nサブピクセルパッチであり、サブピクセルパッチは3×3ピクセルである。一実施態様では、連続するサブピクセルの所定の近傍は、原点サブピクセルを含むピクセルを中心とするn個の接続されたサブピクセル近傍である。一実施態様では、方法は、検体マップ内の、不連続領域のいずれにも属しないサブピクセルを背景として識別することを含む。

【1096】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(訓練データ生成)

【1097】

私たちは、ニューラルネットワークベースのテンプレート生成及びベースコールのための訓練データを生成するコンピュータ実装の方法を開示している。方法は、配列決定動作の複数のサイクルにわたって捕捉されたフローセルの多数の画像にアクセスすることを含み、フローセルは複数のタイルを有し、多数の画像において、各タイルは、複数のサイクルにわたって生成された一連の画像セットを有し、画像セットの配列内の各画像は、特定の1回のサイクルにおいて、特定のタイルのうちの特定の1つの上の検体及びそれらの周囲の背景の強度放射を示す。本方法は、複数の訓練例を有する訓練セットを構築することを含み、各訓練例は、タイルのうちの特定の1つに対応し、タイルのうちの特定の1つの画像セットの配列内の少なくとも一部の画像セットからの画像データを含む。本方法は、訓練例のそれぞれについて、少なくとも1つのグラウンドトゥルスデータ表現を生成することを含み、地面真理値データ表現は、その強度放射が画像データによって描かれているタイルのうちの特定の1つの上の、検体及びそれらの周囲の背景の空間分布のうちの少なくとも1つを識別することと、検体の形状、検体サイズ、及び／又は検体境界、及び／又は検体の中心のうちの少なくとも1つを含む。

【1098】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1099】

一実施態様では、画像データは、タイルのうちの特定の1つの画像セットの配列内の少なくとも一部の画像セットのそれぞれの画像を含み、画像は1800×1800の解像度を有する。一実施態様では、画像データは、画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチは、タイルのうちの特定の1つの一部分を覆い、解像度が20×20である。一実施態様では、画像データは、画像パッチのアップサンプリングされた表現を含み、アップサンプリングされた表現は、80×80の解像度を有する。一実施態様では、グラウンドトゥルスデータ表現は、80×80のアップサンプリング解像度を有する。

【1100】

一実施態様では、複数の訓練例は、タイルの同じ特定の1つに対応し、それぞれ、タイルのうちの同じ特定の画像セットの画像セットの配列内の各画像から異なる画像パッチを画像データとして含み、異なる画像パッチのうちの少なくとも一部は互いに重複する。一実施態様では、グラウンドトゥルスデータ表現は、隣接するサブピクセルの不連続領域として検体を識別し、検体の中心は、不連続領域のそれぞれの1つの内部の質量サブピクセルの中心として検体の中心、及びそれらの周囲の背景として、検体を識別する。一実施態様では、グラウンドトゥルスデータ表現は、色符号化を使用して、検体中心又は非中心のいずれかとして各サブピクセルを識別する。一実施態様では、グラウンドトゥルスデータ表現は、色符号化を使用して、検体内部、検体中心、又は周囲の背景のいずれかとして各サブピクセルを識別する。

【1101】

一実施態様では、本方法は、ニューラルネットワークベースのテンプレート生成及びベ

ースコールの訓練データとして、訓練セット及び関連するグラウンドトゥールースデータ表現内の訓練例を、メモリに記憶することを含む。一実施態様では、本方法は、様々なフローセル、配列決定機器、配列決定プロトコル、配列決定ケミストリー、配列決定試薬、及び検体密度の訓練データを生成することを含む。

【1102】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

10

(メタデータ及びベースコール生成)

【1103】

一実施態様では、方法は、シーケンサによって生成された検体の配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、ニューラルネットワークを訓練して、検体に関するメタデータを生成するための訓練データを使用することと、を含む。他の実施態様のための特定の实施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

20

【1104】

一実施態様では、方法は、シーケンサによって生成された検体の配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、ニューラルネットワークを訓練して検体をベースコールのための訓練データを使用することと、を含む。他の実施態様のための特定の实施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

(回帰モデル)

【1105】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、入力画像データを、ニューラルネットワークを介して画像セットの配列から処理し、入力画像データの代替表現を生成することを含む。画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャネルのために捕捉された、それらの周囲の背景とを示す。本方法は、出力層を介して代替表現を処理することと、検体を識別する出力を生成することと、を含み、その強度放射が、隣接するサブピクセルの不連続領域、不連続領域のうちのそれぞれの1つの質量中心での中心サブピクセルとしての検体の中心、及び不連続領域のいずれにも属しない背景サブピクセルとしてのそれらの周囲の背景として、入力画像データによって

40

50

表される。

【1106】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1107】

一実施態様では、不連続領域のうちの対応する領域内の隣接するサブピクセルは、隣接するサブピクセルが属する不連続領域内の中心サブピクセルからの隣接するサブピクセルの距離に従って重み付けされた強度値を有する。一実施態様では、中心サブピクセルは、不連続領域のうちの対応する領域内で最も高い強度値を有する。一実施態様では、背景サブピクセルは全て、出力において同じ最低強度値を有する。一実施態様では、出力層は、ゼロと1との間の強度値を正規化する。

【1108】

一実施態様では、方法は、ピークローケータを出力に適用して、出力におけるピーク強度を見つけ、ピーク強度に基づいて、検体の中心の位置座標を決定することと、入力画像データを作成するために使用されるアップサンプリング係数によって場所座標をダウンスケールすることと、検体をベースコールする際に使用するために、メモリにダウンスケールされた場所座標を記憶することと、を含む。一実施態様では、本方法は、同じ検体に属する検体内部サブピクセルとして、隣接するサブピクセルを、同じ検体に属する検体内部サブピクセルとして分類することと、検体をベースコールする際に使用するために、検体による検体ベースの検体内部サブピクセルの分類及びダウンスケールされた場所座標を記憶することと、を含む。一実施態様では、本方法は、検体ベースで、検体の中心のうちの対応する検体の内部サブピクセルの距離を判定することと、検体をベースコールする際に使用するために、検体による検体ベースでのメモリ内の距離を記憶することと、を含む。

【1109】

一実施態様では、本方法は、最近傍強度抽出、ガウス系強度抽出、平均 2×2 サブピクセル領域に基づく強度抽出のうちの少なくとも1つを使用することを含む、不連続領域のうちの対応する領域内の検体内部サブピクセルから強度を抽出することを含み、 2×2 個のサブピクセル領域の最も明るい試験に基づく強度抽出、平均 3×3 サブピクセル面積、双線形強度抽出、二次強度抽出、及び／又は強度抽出に基づく強度抽出、及び／又は加重領域被覆率に基づく強度抽出、及び／又は強度抽出に基づいて、強度抽出、及び／又は強度抽出に基づいて強度抽出することと、を含む。

【1110】

一実施態様では、本方法は、不連続領域に基づいて、検体形状、検体サイズ、及び／又は検体境界のうちの少なくとも1つを含む検体の空間分布を決定することと、検体をベースコールする際に使用するために、検体による検体ベースのメモリ内に関連する検体メタデータを保存することと、を含む。

【1111】

一実施態様では、入力画像データは、画像セットの配列内の画像を含み、画像は 3000×3000 の解像度を有する。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部分を覆い、解像度は 20×20 である。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、アップサンプリングされた表現は、 80×80 の解像度を有する。一実施態様では、出力は、 80×80 のアップサンプリング解像度を有する。

【1112】

一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳み込みセグメンテーションニューラルネ

10

20

30

40

50

ットワークであり、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを完全入力解像度特徴マップにマッピングするデコーダの階層を含む。一実施態様では、検体の密度は、約100,000検体/mm²～約1,000,000検体/mm²の範囲である。別の実施態様では、検体の密度は、約1,000,000検体/mm²～約10,000,000検体/mm²の範囲である。

【1113】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(訓練回帰モデル)

【1114】

本発明者らは、検体及び関連する検体メタデータを識別するためにニューラルネットワークを訓練するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを訓練するための訓練データを取得することを含む。訓練データは、訓練例を処理することによってニューラルネットワークによって生成されるべき、複数の訓練例及び対応するグラウンドトゥールースデータを含む。各訓練例は、画像セットの配列からの画像データを含む。画像セットの配列内の各画像は、フローセルのタイルを覆い、タイル上の検体の強度放射及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャネルのために捕捉された、それらの周囲の背景を示す。各グラウンドトゥールースデータは、対応する訓練例の画像データによって示され、隣接するサブピクセルの不連続領域として、検体の中心は、不連続領域のそれぞれの1つの質量中心における中心サブピクセルとしての検体の中心、及びそれらの周囲の背景として、対応する訓練例の画像データによって示される検体を識別する。本方法は、ニューラルネットワークを訓練し、出力とグラウンドトゥールースデータとの間の誤差を最小化する損失関数を反復的に最適化することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む。ニューラルネットワークを訓練し、訓練例の出力を生成することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む。

【1115】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1116】

一実施態様では、本方法は、最後の反復後の誤差収束の際に、メモリ内のニューラルネットワークの更新されたパラメータを記憶して、更なるニューラルネットワークベースのテンプレート生成及びベースコールに適用することを含む。一実施態様では、グラウンドトゥールースデータでは、不連続領域のうちの対応する領域内の隣接するサブピクセルは、隣接するサブピクセルが属する接合領域内の中心サブピクセルからの隣接するサブピクセルの距離に従って重み付けされた強度値を有する。一実施態様では、グラウンドトゥールースデータでは、中心サブピクセルは、それぞれの不連続領域内の最も高い強度値を有する。一実施態様では、グラウンドトゥールースデータでは、背景サブピクセルは全て、出力において同じ最低強度値を有する。一実施態様では、グラウンドトゥールースデータでは、強度値は、ゼロと1との間で正規化される。

【1117】

一実施態様では、損失関数は平均二乗誤差であり、出力及びグラウンドトゥールースにおける対応するサブピクセルの正規化された強度値とグラウンドトゥールースとの間のサブピ

クセル基準で最小化される。一実施態様では、グラウンドトゥルスデータは、関連する検体メタデータの一部として、検体形状、検体サイズ、及び／又は検体境界のうちの少なくとも1つを含む検体の空間分布を識別する。一実施態様では、画像データは画像セットの配列内の画像を含み、画像は 1800×1800 の解像度を有する。一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部を覆い、解像度は 20×20 である。一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、画像パッチのアップサンプリングされた表現は、 80×80 の解像度を有する。

【1118】

一実施態様では、訓練データにおいて、複数の訓練例は、それぞれ、同じタイルの画像セットの配列内の各画像からの画像データの異なる画像パッチとして、及び異なる画像パッチの少なくとも一部が互いに重複する。一実施態様では、グラウンドトゥルスデータは、 80×80 のアップサンプリング解像度を有する。一実施態様では、訓練データは、フローセルの複数のタイルの訓練例を含む。一実施態様では、訓練データは、様々なフローセル、配列決定インストール、配列決定プロトコル、配列決定ケミストリー、配列決定試薬、及び検体密度の訓練例を含む。一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有する深層完全畳み込みニューラルネットワークであり、エンコーダサブネットワークは、エンコーダの階層を含み、デコーダサブネットワークは、低解像度のエンコーダ特徴部マップを、最終分類層によるサブピクセルごとの分類のための完全入力解像度特徴マップにマッピングするデコーダの階層を含む。

【1119】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ニューラルネットワークベースのテンプレート生成器)

【1120】

我々は、フローセル上の検体に関するメタデータを決定するコンピュータ実装の方法を開示している。本方法は、検体の強度放射を描写する画像データにアクセスすることと、ニューラルネットワークの1つ又はそれ以上の層を介して画像データを処理することと、画像データの代替表現を生成することと、出力層を介して代替表現を処理し、検体及び／又は検体の中心の形状及びサイズのうちの少なくとも1つを識別する出力を生成することと、を含む、方法。

【1121】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1122】

一実施態様では、画像データは、検体の周囲の背景の強度放射を更に示す。そのような実施態様では、本方法は、検体間の周囲の背景及び境界を含む、フローセル上の検体の空間分布を識別する出力を含む。一実施態様では、方法は、出力に基づいて、フローセル上の検体の中心位置座標を決定することを含む。一実施態様では、ニューラルネットワークは、畳み込みニューラルネットワークである。一実施態様では、ニューラルネットワークは、反復ニューラルネットワークである。一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳

10

20

30

40

50

み込みセグメンテーションニューラルネットワークであり、続いて出力層はエンコーダサブネットワークがエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを完全入力解像度特徴マップにマッピングするデコーダの階層を含む。

【1123】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(バイナリ分類モデル)

【1124】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを介して画像セットの配列から入力画像データを処理することと、画像データの代替表現を生成することと、を含む。一実施態様では、画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉された、それらの周囲の背景を示す。本方法は、分類層を通して代替表現を処理することと、その強度放射が入力画像データによって示される検体の中心を識別する出力を生成することと、を含む。出力は複数のサブピクセルを有し、複数のサブピクセル内の各サブピクセルは、検体中心又は非中心のいずれかとして分類される。

【1125】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1126】

一実施態様では、分類層は、出力内の各サブピクセルを、検体中心である第1の尤度スコア、及び非中心である第2の尤度スコアを割り当てる。一実施態様では、第1及び第2の尤度スコアは、ソフトマックス関数に基づいて決定され、ゼロと1との間で指数関数的に正規化される。一実施態様では、第1及び第2の尤度スコアは、シグモイド関数に基づいて決定され、ゼロと1との間で正規化される。一実施態様では、出力における各サブピクセルは、第1及び第2の尤度スコアのうちの1つが他方よりも高いかに基づいて、検体中心又は非中心のいずれかとして分類される。一実施態様では、出力における各サブピクセルは、第1及び第2の尤度スコアが所定の閾値尤度スコアを上回るかどうかに基づいて、検体中心又は非中心のいずれかとして分類される。一実施態様では、出力は、検体のうちの対応する検体の質量中心の中心を識別する。一実施態様では、出力において、検体中心として分類されるサブピクセルには、同じ第1の所定の値が割り当てられ、非中心として分類されるサブピクセルは全て、同じ第2の所定の値を割り当てられる。一実施態様では、第1及び第2の所定の値は強度値である。一実施態様では、第1及び第2の所定の値は、連続値である。

【1127】

一実施態様では、本方法は、検体中心として分類されたサブピクセルの場所座標を決定することと、入力画像データを調製するために使用されるアップサンプリング係数によって場所座標をダウンスケールすることと、検体をベースコールする際に使用するために、メモリ内にダウンスケールされた場所座標を記憶することと、を含む。一実施態様では、入力画像データは、画像セットの配列内の画像を含み、画像は3000×3000の解像度を有する。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部を覆い、解像度

10

20

30

40

50

は 20×20 である。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、アップサンプリングされた表現は、 80×80 の解像度を有する。一実施態様では、出力は、 80×80 のアップサンプリング解像度を有する。

【1128】

一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダサブネットワークを有するディープフル畳み込みセグメンテーションニューラルネットワークであり、続いて分類層が、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを、分類層によるサブピクセルごとの分類のための完全入力解像度特徴マップにマッピングするデコーダの階層を含む。一実施態様では、検体の密度は、約 $100,000$ 検体/ $\text{mm}^2 \sim$ 約 $1,000,000$ 検体/ mm^2 の範囲である。別の実施態様では、検体の密度は、約 $1,000,000$ 検体/ $\text{mm}^2 \sim$ 約 $10,000,000$ 検体/ mm^2 の範囲である。

【1129】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(バイナリ分類モデルの訓練)

【1130】

本発明者らは、検体及び関連する検体メタデータを識別するためにニューラルネットワークを訓練するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを訓練するための訓練データを取得することを含む。訓練データは、訓練例を処理することによってニューラルネットワークによって生成されるべき、複数の訓練例及び対応するグラウンドトゥールースデータを含む。各訓練例は、画像セットの配列からの画像データを含む。画像セットの配列内の各画像は、フローセルのタイルを覆い、タイル上の検体の強度放射及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャネルのために捕捉された、それらの周囲の背景を示す。各グラウンドトゥールースデータは、対応する訓練例の画像データによって強度放射が示される、検体の中心を識別する。グラウンドトゥールースデータは複数のサブピクセルを有し、複数のサブピクセル内の各サブピクセルは、検体中心又は非中心のいずれかとして分類される。本方法は、ニューラルネットワークを訓練し、出力とグラウンドトゥールースデータとの間の誤差を最小化する損失関数を反復的に最適化することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む、ニューラルネットワークを訓練し、訓練例の出力を生成することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む。

【1131】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1132】

一実施態様では、本方法は、最後の反復後の誤差収束の際に、メモリ内のニューラルネットワークの更新されたパラメータを記憶して、更なるニューラルネットワークベースのテンプレート生成及びベースコールに適用することを含む。一実施態様では、グラウンドトゥールースデータにおいて、検体中心として分類されるサブピクセルは全て、同じ第1の所定のクラススコアを割り当てられ、非中心として分類されるサブピクセルは全て、同じ第2の所定のクラススコアが割り当てられる。一実施態様では、各出力において、各サブ

ピクセルは、検体中心である第1の予測スコアと、非中心である第2の予測スコアとを有する。一実施態様では、損失関数は、カスタム加重バイナリクロスエントロピー損失であり、出力及びグラウンドトゥールースにおける対応するサブピクセルの予測スコアとクラススコアとの間のサブピクセルベースで最小化される。一実施態様では、グラウンドトゥールースデータは、検体のうちの対応する検体の重心における中心を識別する。一実施態様では、グラウンドトゥールースでは、検体中心として分類されるサブピクセルは全て、同じ第1の所定の値を割り当てられ、非中心として分類されるサブピクセルは全て、同じ第2の所定の値を割り当てられる。一実施態様では、第1及び第2の所定の値は強度値である。別の実施態様では、第1及び第2の所定の値は、連続値である。

【1133】

一実施態様では、グラウンドトゥールースデータは、関連する検体メタデータの一部として、検体形状、検体サイズ、及び／又は検体境界のうちの少なくとも1つを含む検体の空間分布を識別する。一実施態様では、画像データは画像セットの配列内の画像を含み、画像は 1800×1800 の解像度を有する。一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部分を覆い、解像度は 20×20 である。一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、画像パッチのアップサンプリングされた表現は、 80×80 の解像度を有する。一実施態様では、訓練データにおいて、複数の訓練例は、それぞれ、同じタイルの画像セットの配列内の各画像からの画像データの異なる画像パッチとして、及び異なる画像パッチの少なくとも一部が互いに重複する。一実施態様では、グラウンドトゥールースデータは、 80×80 のアップサンプリング解像度を有する。一実施態様では、訓練データは、フローセルの複数のタイルの訓練例を含む。一実施態様では、訓練データは、様々なフローセル、配列決定インストール、配列決定プロトコル、配列決定ケミストリー、配列決定試薬、及び検体密度の訓練例を含む。一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳み込みセグメンテーションニューラルネットワークであり、続いて分類層が、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを、分類層によるサブピクセルごとの分類のための完全入力解像度特徴マップにマッピングするデコーダの階層を含む。

【1134】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(三元分類モデル)

【1135】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを介して画像セットの配列から入力画像データを処理することと、画像データの代替表現を生成することと、を含む。画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉された、それらの周囲の背景を示す。本方法は、分類層を通して代替表現を処理することと、検体の中心、検体形状、検体サイズ、及び／又は検体境界のうちの少なくとも1つを含む、入力画像データによって表される検体及びそれらの周囲の背景の空間分布を識別する出力を生成することと、を含む。出力は複数のサブピクセルを有し、複数のサブピクセル内の各サブピクセルは、背景、検体中心、又は検体内部のいずれかとして分類される。

【1136】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1137】

一実施態様では、分類層は、出力中の各サブピクセルを、背景である第1の尤度スコア、検体中心である第2の尤度スコア、及び検体内部である第3の尤度スコアを割り当てる。一実施態様では、第1、第2、及び第3の尤度スコアは、ソフトマックス関数に基づいて決定され、ゼロと1との間で指数関数的に正規化される。一実施態様では、出力における各サブピクセルは、第1、第2、及び第3の尤度スコアのうちの1つが最も高いかに基づいて、背景、検体中心、又は検体内部のいずれかとして分類される。一実施態様では、出力における各サブピクセルは、第1、第2、及び第3の尤度スコアが所定の閾値尤度スコアを上回るかどうかに基づいて、背景、検体中心、又は検体内部のいずれかとして分類される。一実施態様では、出力は、検体のうちの対応する検体の質量中心で検体中心を識別する。一実施態様では、出力において、背景として分類されるサブピクセルは全て、同じ第1の所定の値を割り当てられ、検体中心として分類されるサブピクセルは全て同じ第2の所定の値を割り当てられ、検体内部として分類されるサブピクセルは全て、同じ第3の所定の値が割り当てられる。一実施態様では、第1、第2、及び第3の所定の値は、強度値である。一実施態様では、第1、第2、及び第3の所定の値は、連続値である。

【1138】

一実施態様では、本方法は、検体基準で検体中心として分類されるサブピクセルの場所座標を決定することと、入力画像データを調製するために使用されるアップサンプリング係数によって場所座標をダウンスケールすることと、検体をベースコールするために、検体による検体ベースのメモリにダウンスケールされた場所座標を記憶することと、を含む。一実施態様では、本方法は、検体ベースの検体内部として分類されたサブピクセルの場所座標を決定することと、入力画像データを調製するために使用されるアップサンプリング係数によって場所座標をダウンスケールすることと、検体をベースコールするために、検体による検体ベースのメモリにダウンスケールされた場所座標を記憶することと、を含む。一実施態様では、本方法は、検体基準に基づいて、検体中心として分類されたサブピクセルのうちの対応する1つから検体内部として分類されるサブピクセルの距離を決定することと、検体をベースコールするために使用するために、検体による検体ベースでのメモリ内の距離を記憶することと、を含む。一実施態様では、本方法は、検体基準で、検体内部として分類されるサブピクセルから強度を抽出することを含み、これは、最近傍強度抽出、ガウス系強度抽出、平均 2×2 サブピクセル領域に基づく強度抽出のうちの少なくとも1つを使用することを含み、 2×2 個のサブピクセル領域の最も明るい試験に基づく強度抽出、平均 3×3 サブピクセル面積、双線形強度抽出、二次強度抽出、及び／又は強度抽出に基づく強度抽出、及び／又は加重領域被覆率に基づく強度抽出、及び／又は強度抽出に基づいて、強度抽出、及び／又は強度抽出に基づいて強度抽出することと、を含む。

【1139】

一実施態様では、入力画像データは、画像セットの配列内の画像を含み、画像は 3000×3000 の解像度を有する。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部分を覆い、解像度は 20×20 である。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、アップサンプリングされた表現は、 80×80 の解像度を有する。一実施態様では、出力は、 80×80 のアップサンプリング解像度を有する。一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳み込みセグメンテーションニューラルネットワークであり、続いて分類層が

、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを、分類層によるサブピクセルごとの分類のための完全入力解像度特徴マップにマッピングするデコーダの階層を含む。一実施態様では、検体の密度は、約100,000検体/mm²～約1,000,000検体/mm²の範囲である。別の実施態様では、検体の密度は、約1,000,000検体/mm²～約10,000,000検体/mm²の範囲である。

【1140】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ターナルクラス分類モデルの訓練)

【1141】

本発明者らは、検体及び関連する検体メタデータを識別するためにニューラルネットワークを訓練するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを訓練するための訓練データを取得することを含む。訓練データは、訓練例を処理することによってニューラルネットワークによって生成されるべき、複数の訓練例及び対応するグラウンドトゥールースデータを含む。各訓練例は、画像セットの配列からの画像データを含む。画像セットの配列内の各画像は、フローセルのタイルを覆い、タイル上の検体の強度放射及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャネルのために捕捉された、それらの周囲の背景とを示す。各グラウンドトゥールースデータは、検体中心、検体形状、検体サイズ、及び検体境界を含む、入力画像データによって表される、検体及びそれらの周囲の背景の空間分布を識別する。グラウンドトゥールースデータは複数のサブピクセルを有し、複数のサブピクセル内の各サブピクセルは、背景、検体中心、又は検体内部のいずれかとして分類される。本方法は、ニューラルネットワークを訓練し、出力とグラウンドトゥールースデータとの間の誤差を最小化する損失関数を反復的に最適化することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む、ニューラルネットワークを訓練し、訓練例の出力を生成することと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を含む。

【1142】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1143】

一実施態様では、本方法は、最後の反復後の誤差収束の際に、メモリ内のニューラルネットワークの更新されたパラメータを記憶して、更なるニューラルネットワークベースのテンプレート生成及びベースコールに適用することを含む。一実施態様では、グラウンドトゥールースデータにおいて、背景として分類されるサブピクセルは全て、同じ第1の所定のクラススコアを割り当てられ、検体中心として分類されるサブピクセルは全て、同じ第2の所定のクラススコアを割り当てられ、検体内部として分類されるサブピクセルは全て、同じ第3の所定のクラススコアが割り当てられる。

【1144】

一実施態様では、各出力において、各サブピクセルは、背景である第1の予測スコア、検体中心である第2の予測スコア、及び検体内部である第3の予測スコアを有する。一実施態様では、損失関数は、カスタム重み付け三元クロスエントロピー損失であり、出力及びグラウンドトゥールースにおける対応するサブピクセルの予測スコアとクラススコアとの

間のサブピクセルベースで最小化される。一実施態様では、グラウンドトゥルスデータは、検体のうちの対応する検体の質量中心で検体中心を識別する。一実施態様では、グラウンドトゥルスにおいて、背景として分類されるサブピクセルは全て、同じ第1の所定の値を割り当てられ、検体中心として分類されるサブピクセルは全て同じ第2の所定の値を割り当てられ、検体内部として分類されるサブピクセルは全て、同じ第3の所定の値が割り当てられる。一実施態様では、第1、第2、及び第3の所定の値は、強度値である。一実施態様では、第1、第2、及び第3の所定の値は、連続値である。一実施態様では、画像データは画像セットの配列内の画像を含み、画像は1800×1800の解像度を有する。一実施態様では、画像データは画像セットの配列内の画像を含み、画像は1800×1800の解像度を有する。

10

【1145】

一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部分を覆い、解像度は20×20である。一実施態様では、画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされた表現を含み、画像パッチのアップサンプリングされた表現は、80×80の解像度を有する。一実施態様では、訓練データにおいて、複数の訓練例は、それぞれ、同じタイルの画像セットの配列内の各画像からの画像データの異なる画像パッチとして、及び異なる画像パッチの少なくとも一部が互いに重複する。一実施態様では、グラウンドトゥルスデータは、80×80のアップサンプリング解像度を有する。一実施態様では、訓練データは、フローセルの複数のタイルの訓練例を含む。一実施態様では、訓練データは、様々なフローセル、配列決定インストール、配列決定プロトコル、配列決定ケミストリー、配列決定試薬、及び検体密度の訓練例を含む。一実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳み込みセグメンテーションニューラルネットワークであり、続いて分類層が、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを、分類層によるサブピクセルごとの分類のための完全入力解像度特徴マップにマッピングするデコーダの階層を含む。

20

【1146】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

(セグメント化)

【1147】

本発明者らは、検体メタデータを決定するコンピュータ実装の方法を開示する。本方法は、ニューラルネットワークを通じて一連の画像セットから導出された入力画像データを処理することと、入力画像データの代替表現を生成することと、を含む。入力画像データは、検体及びそれらの周囲背景を描写するユニットのアレイを有する。本方法は、出力層を介して代替表現を処理し、アレイ内の各ユニットの出力値を生成することを含む。本方法は、ユニットの出力値を閾値化することと、周囲の背景を描写する背景ユニットとして、ユニットの第1のサブセットを分類することと、を含む。本方法は、ユニットの出力値内にピークを配置することと、検体の中心を含む中心ユニットとして、ユニットの第2のサブセットを分類することと、を含む。本方法は、単位の出力値にセグメント化器を適用することと、背景ユニットによって分離され、中心ユニットを中心とした連続ユニットの非重複領域として、検体の形状を決定することと、を含む。セグメントは中心ユニットから始まり、各中心ユニットに関して、中心が中心ユニットに含まれる同じ検体を示す連続的に連続するユニット群を決定する。

40

【1148】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この

50

実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1149】

一実施態様では、ユニットはピクセルである。別の実施態様では、ユニットはサブピクセルである。更に別の実施態様では、ユニットはスーパーピクセルである。一実施態様では、出力値は連続値である。別の実施態様では、出力値は、ソフトマックススコアである。一実施態様では、非重複領域のうちの対応する領域内の連続単位は、隣接するユニットが属する非重複領域内の中心ユニットからの連続ユニットの距離に従って重み付けされた出力値を有する。一実施態様では、中心ユニットは、非重複領域のうちのそれぞれの領域内で最も高い出力値を有する。

【1150】

一実施態様では、非重複領域は不規則な輪郭を有し、ユニットはサブピクセルである。そのような実施態様において、方法は、所与の検体の形状を識別する連続するサブピクセルの対応する非重複領域に基づいて、所与の検体の検体強度に寄与するサブピクセルを識別することによって、所与の検体の検体強度を判定することと、現在の配列決定サイクルで1つ又はそれ以上の画像チャネルに対して生成された1つ又はそれ以上の光学ピクセル解像度画像内に識別されたサブピクセルを配置することと、画像のそれぞれにおいて、識別されたサブピクセルの強度を補間することと、補間強度を組み合わせ、組み合わせられた補間強度を正規化して、画像のそれぞれにおける所与の検体のための画像ごとの検体強度を生成することと、画像のそれぞれについて画像ごとの検体強度を合わせて、現在の配列決定サイクルにおいて、所与の検体の検体強度を決定することと、を含む。一実施態様では、正規化は正規化係数に基づいており、正規化係数は、識別されたサブピクセルの数である。一実施態様では、本方法は、現在の配列決定サイクルにおける検体強度に基づいて、所与の検体をベースコールすることを含む。

【1151】

一実施態様では、非重複領域は不規則な輪郭を有し、ユニットはサブピクセルである。そのような実施態様において、方法は、所与の検体の形状を識別する連続するサブピクセルの対応する非重複領域に基づいて、所与の検体の検体強度に寄与するサブピクセルを識別することによって、所与の検体の検体強度を判定することと、識別されたサブピクセルを、対応する光学からアップサンプリングされた1つ又はそれ以上のサブピクセル解像度画像内に配置することと、現在の配列決定サイクルで1つ又はそれ以上の画像チャネルに対して生成されたピクセル解像度画像であって、アップサンプリングされた画像のそれぞれにおいて、識別されたサブピクセルの強度を組み合わせ、組み合わせた強度を正規化して、アップサンプリングされた画像のそれぞれにおける所与の検体のための画像ごとの検体強度を生成することと、アップサンプリングされた画像のそれぞれに対する画像ごとの検体強度を合わせて、現在の配列決定サイクルでの所与の検体の検体強度を決定することと、を含む。一実施態様では、正規化は正規化係数に基づいており、正規化係数は、識別されたサブピクセルの数である。一実施態様では、本方法は、現在の配列決定サイクルにおける検体強度に基づいて、所与の検体をベースコールすることを含む。

【1152】

一実施態様では、画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャネルのために捕捉された、それらの周囲の背景とを示す。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの少なくとも1つの画像パッチを含み、画像パッチはタイルの一部分を覆い、解像度は 20×20 である。一実施態様では、入力画像データは、画像セットの配列内の画像のそれぞれからの画像パッチのアップサンプリングされたサブピクセル解像度表現を含み、アップサンプリングされたサブピクセル表現は、 80×80 の解像度を有する。

【1153】

一実施態様では、ニューラルネットワークは、畳み込みニューラルネットワークである。別の実施態様では、ニューラルネットワークは、反復ニューラルネットワークである。更に別の実施態様では、ニューラルネットワークは、残留ブロック及び残留接続を有する残留ニューラルネットワークである。更に更なる別の実施態様では、ニューラルネットワークは、エンコーダサブネットワーク及び対応するデコーダネットワークを有するディープフル畳み込みセグメンテーションニューラルネットワークであり、エンコーダサブネットワークはエンコーダの階層を含み、デコーダサブネットワークは、低解像度エンコーダ特徴マップを完全入力解像度特徴マップにマッピングするデコーダの階層を含む。

【1154】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ピーク検出)

【1155】

本発明者らは、検体メタデータを決定するコンピュータ実装の方法を開示する。本方法は、ニューラルネットワークを通じて一連の画像セットから導出された入力画像データを処理することと、入力画像データの代替表現を生成することと、を含む。入力画像データは、検体及びそれらの周囲背景を描写するユニットのアレイを有する。本方法は、出力層を介して代替表現を処理し、アレイ内の各ユニットの出力値を生成することを含む。本方法は、ユニットの出力値を閾値化することと、周囲の背景を描写する背景ユニットとして、ユニットの第1のサブセットを分類することと、を含む。本方法は、ユニットの出力値内にピークを配置することと、検体の中心を含む中心ユニットとして、ユニットの第2のサブセットを分類することと、を含む。

【1156】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1157】

一実施態様では、本方法は、単位の出力値にセグメント化器を適用することと、背景ユニットによって分離され、中心ユニットで中心に置かれた連続ユニットの非重複領域として、検体の形状を決定することと、を含む。セグメントは中心ユニットから始まり、各中心ユニットに関して、中心が中心ユニットに含まれる同じ検体を示す連続的に連続するユニット群を決定する。

【1158】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

ニューラルネットワークに基づく分析データ生成器

【1159】

一実施態様では、方法は、ニューラルネットワークを介して画像データを処理することと、画像データの代替表現を生成することと、を含む。画像データは、検体の強度放射を示す。本方法は、出力層を介して代替表現を処理することと、検体の空間分布、検体の形状、検体の中心、及び／又は検体間の境界、すなわち、検体境界／境界（複数）のうちの

10

20

30

40

50

少なくとも1つを含む、検体に関するメタデータを識別する出力を生成することと、を含む。他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

10

(ユニットベースの回帰モデル)

【1160】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、入力画像データを、ニューラルネットワークを介して画像セットの配列から処理し、入力画像データの代替表現を生成することを含む。画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉された、それらの周囲の背景を示す。本方法は、出力層を介して代替表現を処理することと、検体を識別する出力を生成することと、を含み、その強度放射が、隣接するユニットの不連続領域、不連続領域のうちのそれぞれの1つの質量中心での中心ユニットとしての検体の中心、及び不連続領域のいずれにも属しない背景ユニットとしてのそれらの周囲の背景として、入力画像データによって示される。

20

【1161】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1162】

一実施態様では、ユニットはピクセルである。別の実施態様では、ユニットはサブピクセルである。更に別の実施態様では、ユニットはスーパーピクセルである。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

(ユニットベースの結合分類モデル)

【1163】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを介して画像セットの配列から入力画像データを処理することと、画像データの代替表現を生成することと、を含む。画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉された、それらの周囲の背景とを示す。本方法は、分類層を通して代替表現を処理することと、その強度放射が入力画像データによって示される検体の中心を識別する出力を生成することと、を含む。出力は複数のユニットを有し、複数のユニット内の各ユニットは、検体中心又は非中心のいずれかとして分類される。

40

【1164】

50

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1165】

一実施態様では、ユニットはピクセルである。別の実施態様では、ユニットはサブピクセルである。更に別の実施態様では、ユニットはスーパーピクセルである。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ユニットベースの三元分類モデル)

【1166】

本発明者らは、フローセル及び関連する検体メタデータのタイル上の検体を識別するコンピュータ実装の方法を開示している。本方法は、ニューラルネットワークを介して画像セットの配列から入力画像データを処理することと、画像データの代替表現を生成することと、を含む。画像セットの配列内の各画像はタイルを覆い、タイル上の検体の強度放射、及びフローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉された、それらの周囲の背景を示す。本方法は、分類層を通して代替表現を処理することと、検体の中心、検体形状、検体サイズ、及び／又は検体境界のうちの少なくとも1つを含む、入力画像データによって表される検体及びそれらの周囲の背景の空間分布を識別する出力を生成することと、を含む。出力は複数のユニットを有し、複数のユニット内の各ユニットは、背景、検体中心、又は検体内部のいずれかとして分類される。

【1167】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1168】

一実施態様では、ユニットはピクセルである。別の実施態様では、ユニットはサブピクセルである。更に別の実施態様では、ユニットはスーパーピクセルである。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ベースコールー単一検体距離チャンネル)

【1169】

我々は、配列決定動作中にフローセルのタイル上で合成される検体をベースコールするニューラルネットワーク実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、テンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、テンプレート画像内の検体の参照中心を識別することを含む。本方法は、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイク

ルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で1つ又はそれ以上の画像にアクセスすることを含む。本方法は、テンプレート画像を有する、現在の、先行する、及び後続の画像セット内の画像の各々を位置合わせして、サイクル固有及び画像チャンネル固有の変換を決定することを含む。本方法は、検体の参照中心に変換を適用して、画像の各々で検体の変換された中心を識別することを含む。本方法は、ベースコールされている検体のうちの特定の1つに対して、現在の、先行する、後続の画像セット内の画像の各々から画像パッチを抽出することを含んで、各画像パッチは、画像のうちのそれぞれの1つで識別される検体のうちの特定の1つの変換された中心をその中心ピクセル内に含み、検体のうちの特定の1つ、検体のうちのいくつかの隣接するもの、及び画像チャンネルの対応する1つでのそれらの周囲の背景の強度放射を示す。本方法は、各画像パッチに対して、その中心ピクセルが含まれる検体のうちの特定の1つの変換された中心から、そのピクセルの中心の距離を識別する距離情報を生成することを含む。本方法は、距離情報を各画像パッチにピクセルごとに符号化することによって、入力データを構築することを含む。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで検体のうちの特定の1つに組み込まれる塩基の尤度を生成することを含む。本方法は、尤度に基づいて、塩基をA、C、T、又はGと分類することを含む。

10

20

【1170】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1171】

一実施態様では、本方法は、各画像パッチに対して、そのピクセルのうちのどれが検体のうちの特定の1つをカバーするか、及びそのピクセルのうちのどれがカバーしないかを識別する検体属性情報を生成することと、検体属性情報を各画像パッチにピクセルごとに符号化することによって入力データを構築することと、を含む。一実施態様では、検体のうちの特定の1つをカバーするピクセルは、検体属性情報で非ゼロ値を割り当てられる。一実施態様では、検体のうちの特定の1つをカバーしないピクセルは、検体属性情報でゼロ値を割り当てられる。一実施態様では、本方法は、検体の変換された中心の畳み込みニューラルネットワーク位置座標への入力として提供することを含む。そのような一実施態様では、入力は、畳み込みニューラルネットワークの第1の層に供給される。別のそのような実施態様では、入力は、畳み込みニューラルネットワークの1つ又はそれ以上の中間層に供給される。更に別のそのような実施態様では、入力は、畳み込みニューラルネットワークの最終層に供給される。一実施態様では、本方法は、画像パッチのピクセルに対応するスケール値を有する強度スケールチャンネルを畳み込みニューラルネットワークへの入力として提供することを含む。そのような実施態様では、スケール値は、検体のうちの特定の1つの中心を含む画像パッチの中心ピクセルの平均強度に基づく。一実施態様では、強度スケールチャンネルは、画像パッチの全てのピクセルの同じスケール値をピクセルごとに含む。一実施態様では、中心ピクセルの平均強度は、画像チャンネルのうちの対応する1つの各々に対して決定される。

30

40

【1172】

一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してA及びTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第1の画像チャンネルに対して決定される

50

。一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してA及びCベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第2の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してAベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第1の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してGベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第2の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第3の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちの特定の1つに対してCベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第3の画像チャンネルに対して決定される。

【1173】

一実施態様では、配列決定動作は、第1のリードプライマー及び第2のリードプライマーを使用して、検体内のフラグメントの両方の末端を順方向及び逆方向に配列決定する、ペアの末端配列決定を実施し、それによって、各フラグメントに対してリードペアを生成し、リードペアは、順方向リード及び逆方向リードを有する。一実施態様では、フラグメントの両方の末端は、連続して配列決定されて、順方向及び逆方向リードを次々に生成する。一実施態様では、フラグメントの両方の末端は、同時に配列決定されて、順方向及び逆方向リードを同時に生成する。一実施態様では、順方向及び逆方向リードは各々、1つ又はそれ以上のフラグメントを含む。一実施態様では、フラグメントのうちの1つ又はそれ以上は、連続して配列決定される。一実施態様では、フラグメントのうちの1つ又はそれ以上は、同時に配列決定される。一実施態様では、配列決定動作は、単一のリードプライマーを使用して1つの方向にフラグメントを配列決定する単一のリード配列決定を実施する。一実施態様では、配列決定動作は、ループ内のフラグメントの2本鎖コピーを配列決定する円形配列決定を実施し、ループは、所与のフラグメントの2本鎖コピーにおいて複数回反復する。一実施態様では、配列決定動作は、フラグメントのスタックコピーを配列決定するスタック配列決定を実施し、所与のフラグメントのスタックコピーは、垂直又は水平にスタックされる。一実施態様では、画像パッチのサイズは、 3×3 のピクセル～ 10000×10000 のピクセルの範囲である。

【1174】

一実施態様では、変換された中心は、浮動小数点座標値である。そのような実施態様では、本方法は、変換された中心の整数座標値を生成するための丸め動作を使用して浮動小数点座標値を丸めることと、その整数座標と変換された中心に対して生成される整数座標値との間の重複に基づいて中心ピクセルを識別することと、を含む。一実施態様では、丸め動作は、床関数、天井関数、及び／又はラウンド関数のうちの少なくとも1つである。一実施態様では、丸め動作は、整数関数及び／又は整数＋符号関数のうちの少なくとも1つである。一実施態様では、テンプレート生成器は、ニューラルネットワークベースのテンプレート生成器である。一実施態様では、出力層は、ソフトマックス層であり、尤度は、A、C、T、及びGである複数の配列決定サイクルのうちの現在の1つで、検体のうちの特定の1つに組み込まれる塩基の指数関数的に正規化されたスコア分布である。

【1175】

一実施態様では、画像チャンネルの各々1つは、複数のフィルタ波長帯域のうちの1つである。別の実施態様では、画像チャンネルの各々1つは、複数の画像イベントのうちの1つである。一実施態様では、フローセルは、検体を占有するウェルのアレイを有する、少なくとも1つのパターン化された表面を有する。別の実施態様では、フローセルは、少なくとも1つの非パターン化表面を有し、検体は、非パターン化表面上で不均一に散乱される

。一実施態様では、画像セットは、4個の画像を有する。別の実施態様では、画像セットは、2個の画像を有する。更に別の実施態様では、画像セットは、1個の画像を有する。一実施態様では、配列決定動作は、4チャンネル化学を利用する。別の実施態様では、配列決定動作は、2チャンネル化学を利用する。更に別の実施態様では、配列決定動作は、1チャンネル化学を利用する。

【1176】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1177】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。入力データは、配列決定動作の現在の配列決定サイクルで生成される現在の画像セット、現在の配列決定サイクルに先行する配列決定動作の1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び現在の配列決定サイクルに続く配列決定動作の1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で1つ又はそれ以上の画像から抽出される画像パッチを含む。画像パッチの各々は、ベースコールされているターゲット検体、いくつかの隣接する検体、及び対応する画像チャンネル内のそれらの周囲の背景の強度放射を示す。入力データは、画像パッチの中心ピクセルに位置するターゲット検体の中心からの画像パッチのピクセルの中心の距離を識別するために画像パッチの各々でピクセルごとに符号化される距離情報を更に含む。本方法は、出力層を介して畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルでターゲット検体をベースコールすることを含む。

【1178】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1179】

一実施態様では、本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである、現在の配列決定サイクルでターゲット検体に組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、を含む。一実施態様では、尤度は、ソフトマックス層によって生成される、指数関数的に正規化されたスコアである。

【1180】

一実施態様では、本方法は、出力から、現在の配列決定サイクルでターゲット検体に組み込まれる塩基がA、C、T、又はGであることのクラスラベルを識別するターゲット検体に対する出力ペアを導出することと、クラスラベルに基づいて、ターゲット検体をベースコールすることと、を含む。一実施態様では、1、0のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、1、1のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。別の実施態様では、1、1のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、0、0のクラスラベルは、G塩基を識別する。更に別の実施態様では、1、0のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、0、0のクラスラベルは、T塩基を識別し、0、0のクラスラ

ベルは、G塩基を識別する。更に更なる実施態様では、1、2のクラスラベルは、A塩基を識別し、0、1のクラスラベルは、C塩基を識別し、1、1のクラスラベルは、T塩基を識別し、0、0のクラスラベルは、G塩基を識別する。一実施態様では、本方法は、出力から、現在の配列決定サイクルでターゲット検体に組み込まれる塩基がA、C、T、又はGであることを識別するターゲット検体に対するクラスラベルを導出することと、クラスラベルに基づいて、ターゲット検体をベースコールすることと、を含む。一実施態様では、0.33のクラスラベルは、A塩基を識別し、0.66のクラスラベルは、C塩基を識別し、1のクラスラベルは、T塩基を識別し、0のクラスラベルは、G塩基を識別する。別の実施態様では、0.50のクラスラベルは、A塩基を識別し、0.75のクラスラベルは、C塩基を識別し、1のクラスラベルは、T塩基を識別し、0.25のクラスラベルは、G塩基を識別する。一実施態様では、本方法は、出力から単一の出力値を導出することと、塩基A、C、T、及びGに対応するクラス値範囲に対して単一の出力値を比較することと、比較に基づいて、単一の出力値を特定のクラス値範囲に割り当てることと、割り当てに基づいて、ターゲット検体をベースコールすることと、を含む。一実施態様では、単一の出力値は、シグモイド関数を使用して導出され、単一の出力値は、0～1の範囲である。別の実施態様では、0～0.25のクラス値範囲は、A塩基を表し、0.25～0.50のクラス値範囲は、C塩基を表し、0.50～0.75のクラス値範囲は、T塩基を表し、0.75～1のクラス値範囲は、G塩基を表す。

【1181】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1182】

我々は、配列決定動作中にフローセルのタイル上で合成される検体をベースコールするニューラルネットワーク実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、1つ又はそれ以上の画像チャネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、テンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、テンプレート画像内の検体の参照中心を識別することを含む。本方法は、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で1つ又はそれ以上の画像にアクセスすることを含む。本方法は、テンプレート画像を有する、現在の、先行する、及び後続の画像セット内の画像の各々を位置合わせして、サイクル固有及び画像チャネル固有の変換を決定することを含む。本方法は、検体の参照中心に変換を適用して、画像の各々で検体の変換された中心を識別することを含む。本方法は、ベースコールされている検体のうちの特定の1つに対して、現在の、先行する、後続の画像セット内の画像の各々から画像パッチを抽出することを含んで、各画像パッチは、画像のうちのそれぞれの1つで識別される検体のうちの特定の1つの変換された中心をその中心ピクセル内に含み、検体のうちの特定の1つ、検体のうちのいくつかの隣接するもの、及び画像チャネルの対応する1つでのそれらの周囲の背景の強度放射を示す。本方法は、各画像パッチに対して、その中心ピクセルが含まれる検体のうちの特定の1つの変換された中心から、そのピクセルの中心の距離を識別する距離情報を生成することを含む。本方法は、距離情報を各画像パッチにピクセルごとに符号化することによって、入力データを構築することを含む。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳

10

20

30

40

50

み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、複数の配列決定サイクルのうちの現在の1つで、検体のうちの特定の1つをベースコールすることを含む。

【1183】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1184】

一実施態様では、本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで検体のうちの特定の1つに組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、を含む。

【1185】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1186】

一実施態様では、コンピュータ実装の方法は、ニューラルネットワークを介して入力データを処理することと、入力データの代替表現を生成することと、を含む。入力データは、配列決定動作の1つ又はそれ以上の配列決定サイクルの各々に対するサイクルごとの画像データを含む。サイクルごとの画像データは、それぞれの配列決定サイクルで捕捉される、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す。本方法は、出力層を介して代替表現を処理することと、出力を生成することと、を含む。本方法は、出力に基づいて、配列決定サイクルのうちの1つ又はそれ以上で、検体のうちの1つ又はそれ以上をベースコールすることを含む。

【1187】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1188】

一実施態様では、本方法は、サイクルごとの画像データのピクセルと、検体のうちの1つ又はそれ以上との間の距離を識別する補足距離情報に、サイクルごとの画像データを伴わせることを含む。そのような実施態様では、距離は、ニューラルネットワーク及び出力層による処理で検体のうちの1つ又はそれ以上の中心、形状、及び／又は境界に関するコンテキストを組み込む。一実施態様では、本方法は、サイクルごとの画像データのピクセルにスケール値を割り当てる補足スケール情報に、サイクルごとの画像データを伴わせることを含む。そのような実施態様では、スケール値は、検体のうちの1つ又はそれ以上の強度の分散を考慮する。

【1189】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

10

20

30

40

50

(ベースコール—マルチ検体距離チャンネル)

【1190】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列を含む入力データにアクセスすることを含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つに対する画像パッチを有する。各画像パッチは、複数の検体及びそれらの周囲の背景をカバーするピクセルに対するピクセル強度データと、ピクセルと検体の各々との間の中心間距離に基づいて選択される検体のうちの最も近い1つから、各ピクセルの中心間距離を識別するピクセル距離データと、を有する。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである現在の配列決定サイクルで検体のうちのそれぞれの1つに組み込まれる塩基の尤度を識別する検体の各々に対するスコア分布を生成することを含む。本方法は、尤度に基づいて、検体の各々をベースコールすることを含む。

10

【1191】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

20

【1192】

一実施態様では、ピクセル距離データは、各画像パッチにピクセルごとに符号化される。一実施態様では、中心間距離は、検体の変換された中心の位置座標及びピクセル中心の位置座標を使用する距離式から導出される。一実施態様では、本方法は、各画像パッチのピクセルに対応するスケーリング値を有する畳み込みニューラルネットワーク強度スケーリングチャンネルへの入力として提供することを含み、スケーリング値は、検体の変換された中心を含む各画像パッチ内の中心ピクセルの平均強度の組み合わせに基づく。一実施態様では、強度スケーリングチャンネルは、画像パッチの全てのピクセルのピクセル強度データに、同じスケーリング値をピクセルごとに適用する。一実施態様では、強度スケーリングチャンネルは、ピクセル近傍ベースで、画像パッチのピクセルのピクセル強度データに異なるスケーリング値をピクセルごとに適用して、第1の中心ピクセルの平均強度から導出される第1のスケーリング値は、第1の中心ピクセルに連続的に連続する隣接するピクセルの第1のピクセル近傍に適用され、別の中心ピクセルの平均強度から導出される別のスケーリング値は、別の中心ピクセルに連続的に連続する隣接するピクセルの別のピクセル近傍に適用される。一実施態様では、ピクセル近傍は、中心ピクセルで中心にされる $m \times n$ のピクセルパッチであり、ピクセルパッチは、 3×3 のピクセルである。一実施態様では、ピクセル近傍は、中心ピクセルで中心にされる n 個の接続されたピクセル近傍である。一実施態様では、中心ピクセルの平均強度は、画像チャンネルの対応する1つの各々に対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちのそれぞれの1つに対してA及びTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第1の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちのそれぞれの1つに対してA及びCベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第2の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちのそれぞれの1つに対してAベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第1の画像チャンネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちのそれぞれの1つに対してGベースコールを生成した2つ又はそれ以上の先行する配列決定サイ

30

40

50

クル中に観察される中心ピクセルの強度値を平均することによって、第2の画像チャネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちそれぞれの1つに対してTベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第3の画像チャネルに対して決定される。一実施態様では、中心ピクセルの平均強度は、検体のうちのそれぞれの1つに対してCベースコールを生成した2つ又はそれ以上の先行する配列決定サイクル中に観察される中心ピクセルの強度値を平均することによって、第3の画像チャネルに対して決定される。一実施態様では、本方法は、各画像パッチに対して、そのピクセルのうちのどれが検体をカバーするか、及びそのピクセルのうちのどれがカバーしないかを識別する検体属性情報を生成することと、検体属性情報を各画像パッチにピクセルごとに符号化することによって入力データを構築することと、を含む。一実施態様では、検体をカバーするピクセルは、検体属性情報で非ゼロ値を割り当てられる。一実施態様では、検体をカバーしないピクセルは、検体属性情報でゼロ値を割り当てられる。一実施態様では、各画像パッチのサイズは、 3×3 のピクセル $\sim 10000 \times 10000$ のピクセルの範囲である。一実施態様では、出力層は、ソフトマックス層であり、スコア分布は、指数関数的に正規化されたスコア分布である。

【1193】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1194】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列を含む入力データにアクセスすることを含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャネルのうちのそれぞれの1つに対する画像パッチを有する。各画像パッチは、複数の検体及びそれらの周囲の背景をカバーするピクセルに対するピクセル強度データと、ピクセルと検体の各々との間の中心間距離に基づいて選択される検体のうちの最も近い1つから、各ピクセルの中心間距離を識別するピクセル距離データと、を有する。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。

【1195】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1196】

一実施態様では、本方法は、出力から、A、C、T、及びGである現在の配列決定サイクルで検体のうちのそれぞれの1つに組み込まれる塩基の尤度を識別する検体の各々に対するスコア分布を導出することと、尤度に基づいて、検体の各々をベースコールすることと、を含む。一実施態様では、出力層は、ソフトマックス層であり、スコア分布は、指数関数的に正規化されたスコア分布である。一実施態様では、本方法は、出力から、現在の配列決定サイクルで検体のうちのそれぞれの1つに組み込まれる塩基がA、C、T、及びGであることのクラスラベルを識別する検体の各々に対する出力ペアを導出することと、クラスラベルに基づいて、検体の各々をベースコールすることと、を含む。一実施態様で

は、本方法は、出力から単一の出力値を導出することと、塩基 A、C、T、及び G に対応するクラス値範囲に対して単一の出力値を比較することと、比較に基づいて、単一の出力値を特定のクラス値範囲に割り当てることと、割り当てに基づいて、検体の各々をベースコールすることと、を含む。一実施態様では、単一の出力値は、シグモイド関数を使用して導出され、単一の出力値は、0～1の範囲である。

【1197】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(ベースコールマルチ検体形状ベースの距離チャンネル)

【1198】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列を含む入力データにアクセスすることを含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つに対する画像パッチを有する。各画像パッチは、検体強度を示す検体ピクセル及び背景強度を示す背景ピクセルを使用して、複数の検体及びそれらの周囲の背景の強度放射を示す。各画像パッチは、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される検体のうちの割り当てられた1つから、各検体ピクセルの中心間距離を識別する検体距離データで符号化される。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである現在の配列決定サイクルで検体のうちのそれぞれの1つに組み込まれる塩基の尤度を識別する検体の各々に対するスコア分布を生成することを含む。本方法は、尤度に基づいて、検体の各々をベースコールすることを含む。

【1199】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1200】

一実施態様では、検体は、複数の検体ピクセルに及ぶ不規則な形状を有し、ピクセル対検体分類は、不規則な形状に基づく。一実施態様では、全ての背景ピクセルは、検体距離データ内の同じ最小中心間距離を割り当てられる。一実施態様では、全ての背景ピクセルは、同じ最小強度を割り当てられる。一実施態様では、各検体ピクセルは、ニューラルネットワークベースのテンプレート生成器によって生成される減衰マップに基づいて、検体のうちの1つのみに分類される。そのような実施態様では、減衰マップは、隣接するピクセルの不連続領域として検体を識別し、不連続領域のうちのそれぞれの1つの質量中心で中心ピクセルとして検体の中心を識別し、不連続領域のいずれにも属しない背景ピクセルとしてそれらの周囲の背景を識別する。一実施態様では、不連続領域のうちのそれぞれの1つでの隣接するピクセルは、隣接するピクセルが属する不連続領域内の中心ピクセルからの隣接するピクセルの距離に従って重み付けされる強度値を有する。一実施態様では、不連続領域のうちのそれぞれの1つでの隣接するピクセルは、同じ検体に属し、それを共に示す検体内部ピクセルとして分類され、検体ごとにメモリ内に記憶される。一実施態様では、中心ピクセルは、不連続領域のうちのそれぞれの1つで最も高い強度値を有する。一実施態様では、背景ピクセルは全て、減衰マップで同じ最低強度値を有する。一実施態様では、検体距離データは、各画像パッチにピクセルごとに符号化している。一実施態様

では、中心間距離は、検体の変換された中心の位置座標及びピクセル中心の位置座標を使用する距離式から導出される。一実施態様では、検体の変換された中心は、減衰マップによって識別される検体の中心に、サイクル固有及び画像チャンネル固有の変換を適用することによって導出される。

【1201】

一実施態様では、本方法は、各画像パッチのピクセルに対応するスケーリング値を有する強度スケーリングチャンネルを畳み込みニューラルネットワークへの入力として提供することを含む。そのような実施態様では、スケーリング値は、検体の変換された中心を含む各画像パッチ内の中心ピクセルの平均強度の組み合わせに基づく。一実施態様では、強度スケーリングチャンネルは、ピクセルグループベースで、画像パッチのピクセルのピクセル強度データに異なるスケーリング値をピクセルごとに適用して、第1の中心ピクセルの平均強度から導出される第1のスケーリング値は、第1の検体に属し、かつそれを共に示す、隣接するピクセルの第1のピクセルグループに適用され、別の検体の中心を含む別の中心ピクセルの平均強度から導出される別のスケーリング値は、別の検体に属し、かつそれを共に示す、隣接するピクセルの別のピクセルグループに適用される。一実施態様では、中心ピクセルの平均強度は、画像チャンネルの対応する1つの各々に対して決定される。一実施態様では、本方法は、各画像パッチに対して、そのピクセルのうちのどれが検体をカバーするか、及びそのピクセルのうちのどれがカバーしないかを識別する検体属性情報を生成することと、検体属性情報を各画像パッチにピクセルごとに符号化することによって入力データを構築することと、を含む。一実施態様では、検体をカバーするピクセルは、検体属性情報で非ゼロ値を割り当てられる。別の実施態様では、検体をカバーしないピクセルは、検体属性情報でゼロ値を割り当てられる。

【1202】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1203】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列を含む入力データにアクセスすることを含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つに対する画像パッチを有する。各画像パッチは、検体強度を示す検体ピクセル及び背景強度を示す背景ピクセルを使用して、複数の検体及びそれらの周囲の背景の強度放射を示す。各画像パッチは、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される検体のうちの割り当てられた1つから、各検体ピクセルの中心間距離を識別する検体距離データで符号化される。本方法は、畳み込みニューラルネットワークを介して、入力データを畳み込んで、入力データの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。

【1204】

他の実施態様のための特定の実施態様セクションで説明される特徴の各々は、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様

は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(専用構造)

【1205】

我々は、互いに位置合わせ誤差を有する配列決定画像を使用して検体をベースコールするネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列にアクセスすることを含む。配列は、サイクルごとの画像パッチセットにわたって、サイクルごとの画像パッチセット内で、画像パッチ間の位置合わせ誤差を有する。配列内の各画像パッチは、ベースコールされているターゲット検体、いくつかの隣接する検体、及び一連の対応する配列決定サイクルでの対応する画像チャンネル内のそれらの周囲の背景の強度情報を示す。配列内の各画像パッチは、その中心ピクセルに位置するターゲット検体の中心からのそのピクセルの中心の距離を識別する距離情報でピクセルごとに符号化される。本方法は、第1の畳み込みサブネットワークを介して、サイクルごとの各画像パッチセットを別々に処理して、各配列決定サイクルに対する中間畳み込み表現を生成することを含み、配列決定サイクル間ではなく配列決定サイクル内でのみ、強度及び距離情報を組み合わせ、かつ得られた畳み込み表現を組み合わせる畳み込みを適用することを含む。本方法は、第2の畳み込みサブネットワークを介して、一連の連続する配列決定サイクルに対する中間畳み込み表現をグループごとに処理して、一連の最終畳み込み表現を生成することを含み、配列決定サイクル間で、中間畳み込み表現を組み合わせ、かつ得られた畳み込み表現を組み合わせる畳み込みを適用することを含む。本方法は、出力層を介して最終畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルでターゲット検体をベースコールすることを含む。

【1206】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1207】

一実施態様では、配列内の各画像パッチは、複数の検体及びそれらの周囲の背景をカバーするピクセルに対するピクセル強度データと、ピクセルと検体の各々との間の中心間距離に基づいて選択される検体のうちの最も近い1つから、各ピクセルの中心間距離を識別するピクセル距離データと、を有する。そのような実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、配列内の各画像パッチは、検体強度を示す検体ピクセル及び背景強度を示す背景ピクセルを使用して、複数の検体及びそれらの周囲の背景の強度放射を示し、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される検体のうちの割り当てられた1つから、各検体ピクセルの中心間距離を識別する検体距離データで符号化される。そのような実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の第1の畳み込みサブネットワーク位置座標への入力として提供することを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の第2の畳み込みサブネットワーク位置座標への入力として提供することを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の出力層位置座標への入力として提供することを含む。

【1208】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体

を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1209】

我々は、位置合わせ誤差を有する画像データを使用して検体をベースコールするネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために入力データにアクセスすることを含む。入力データは、各配列決定サイクルに対する画像テンソルを有する。各画像テンソルは、1つ又はそれ以上の画像チャネルに対するデータを有し、各画像チャネルに対して、ベースコールされているターゲット検体、いくつかの隣接する検体、及び周囲の背景をカバーするピクセルに対するピクセル強度データ、並びにターゲット検体の中心からピクセルの中心までの距離に対するピクセル距離データを含む。入力データは、画像テンソルにわたるピクセル間のクロスサイクル位置合わせ誤差、及び画像テンソル内のピクセル間のクロス画像チャネル位置合わせ誤差を有する。本方法は、空間畳み込み層の配列を有する空間畳み込みネットワークを介して各入力テンソルを別々に処理して、各配列決定サイクルに対する空間畳み込み表現を生成することを含み、配列決定サイクル間ではなく配列決定サイクル内でのみ、ピクセル強度及び距離を組み合わせる第1の空間畳み込み層で開始することと、配列決定サイクル間ではなく一連の配列決定サイクル内の各配列決定サイクル内でのみ、先行する空間畳み込み層の出力を組み合わせる連続する空間畳み込み層を継続することと、を含む。本方法は、時間的畳み込み層の配列を有する時間的畳み込みネットワークを介して、連続する配列決定サイクルに対する空間畳み込み表現をグループごとに処理して、一連の時間的畳み込み表現を生成することを含み、一連の配列決定サイクル内の配列決定サイクル間で空間畳み込み表現を組み合わせる第1の時間的畳み込み層で開始することと、先行する時間的畳み込み層の連続する出力を組み合わせる、連続する時間的畳み込み層を継続することと、を含む。本方法は、出力層を介して時間的畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルでターゲット検体をベースコールすることを含む。

【1210】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1211】

一実施態様では、グループごとに処理することは、重複するスライディングウィンドウ内の連続する中間畳み込み表現上で畳み込むことを更に含む。一実施態様では、連続する時間的畳み込み層は、重複するスライディングウィンドウ内で連続する出力を組み合わせる。一実施態様では、ピクセル距離データは、各画像テンソルにピクセルごとに符号化している。一実施態様では、配列内の各画像テンソルは、複数の検体及びそれらの周囲の背景をカバーするピクセルに対するピクセル強度データと、ピクセルと検体の各々との間の中心間距離に基づいて選択される検体のうちの最も近い1つから、各ピクセルの中心間距離を識別するピクセル距離データと、を有する。一実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、配列内の各画像テンソルは、検体強度を示す検体ピクセル及び背景強度を示す背景ピクセルを使用して、複数の検体及びそれらの周囲の背景の強度放射を示し、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される検体のうちの割り当てられた1つから、各検体ピクセルの中心間距離を識別する検体距離データで符号化される。一実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の第1の畳み込みサブネットワーク位置座標への入力として提供することを含

む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の第2の畳み込みサブネットワーク位置座標への入力として提供することを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の出力層位置座標への入力として提供することを含む。

【1212】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

10

(再構成)

【1213】

我々は、配列決定動作中に合成される検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、配列決定動作の一連の配列決定サイクルのために生成されるサイクルごとの画像パッチセットの配列にアクセスすることを含む。配列内のサイクルごとの各画像パッチセットは、1つ又はそれ以上の画像チャネルのうちのそれぞれの1つに対する画像パッチを有する。各画像パッチは、ベースコールされているターゲット検体、いくつかの隣接する検体、及び周囲の背景をカバーするピクセルに対するピクセル強度データを有する。本方法は、各画像パッチのピクセルを再構成して、中心ピクセル内のターゲット検体の中心を中心にするを含む。本方法は、畳み込みニューラルネットワークを介して、再構成された画像パッチを畳み込んで、再構成された画像パッチの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して出力を生成することを含む。本方法は、出力に基づいて、現在の配列決定サイクルでターゲット検体をベースコールすることを含む。

20

【1214】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

30

【1215】

一実施態様では、再構成は、再構成を補償するための各画像パッチのピクセルの強度補間を更に含む。一実施態様では、強度補間は、最近傍強度抽出、ガウス系強度抽出、平均 2×2 のサブピクセル領域に基づく強度抽出、最も明るい 2×2 のサブピクセル領域に基づく強度抽出、平均 3×3 のサブピクセル領域に基づく強度抽出、バイリニア強度抽出、双次強度抽出、及び／又は重み付け領域被覆に基づく強度抽出のうちの少なくとも1つを更に含む。一実施態様では、再構成の前に、ターゲット検体の中心は、中心ピクセルの中心からオフセットで各画像パッチの中心ピクセルに位置する。一実施態様では、再構成は、各画像パッチの非中心ピクセルが、ターゲット検体のそれぞれの中心から等距離であることを必要とすることを更に含む。一実施態様では、配列内の各画像パッチは、複数の検体及びそれらの周囲の背景を示すピクセルに対するピクセル強度データと、ピクセルと検体の各々との間の中心間距離に基づいて選択される検体のうちの最も近い1つから、各ピクセルの中心間距離を識別するピクセル距離データと、を有する。一実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、配列内の各画像パッチは、検体強度を示す検体ピクセル及び背景強度を示す背景ピクセルを使用して、複数の検体及びそれらの周囲の背景の強度放射を示し、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される検体のうちの割り当てられた1つから、各検体ピクセルの中心間距離を識別する検体距離データで符号化される。一実施態様では、本方法は、出力に基づいて、現在の配列決定サイクルで検体の各々をベースコールすることを含む。一実施態様では、本方法は、ターゲット検

40

50

体及び／又は隣接する検体の第1の畳み込みサブネットワーク位置座標への入力として提供することを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の第2の畳み込みサブネットワーク位置座標への入力として提供することを含む。一実施態様では、本方法は、ターゲット検体及び／又は隣接する検体の出力層位置座標への入力として提供することを含む。

【1216】

我々は、フローセル上で検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、フローセル上で検体を合成する配列決定動作の複数の配列決定サイクル上で生成される画像セットの配列にアクセスすることを含む。画像セットの配列内の各画像は、フローセルの非重複領域をカバーし、複数の配列決定サイクルのうちのそれぞれの1つでの対応する画像チャンネル内に捕捉される、非重複領域上の検体のサブセット及びそれらの周囲の背景の強度放射を示す。本方法は、画像セットの配列から、複数の配列決定サイクルのうちの特定の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの特定の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの特定の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットを選択することによって、検体のサブセットのうちの特定の1つで複数の配列決定サイクルのうちの特定の1つで組み込まれるヌクレオチド塩基（A、C、T、又はG）を決定することを含む。本方法は、選択された画像セットの各々での画像から画像パッチを抽出することを含む。画像パッチは、検体のサブセットのうちの特定の1つで中心にされ、検体のサブセットから追加の隣接する検体を含む。本方法は、畳み込みニューラルネットワークの1つ又はそれ以上の層を介して、画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することを含む。本方法は、出力層を介して畳み込み表現を処理して、A、C、T、及びGである、ヌクレオチド塩基に対する尤度を生成することを含む。本方法は、尤度に基づいて、ヌクレオチド塩基をA、C、T、又はGと分類することを含む。

【1217】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1218】

一実施態様では、本方法は、複数の配列決定サイクルの各々に対して、選択、抽出、畳み込み、処理、及び分類を反復することによって、複数の配列決定サイクルにおいて検体のサブセットのうちの特定の1つに対してベースコールの配列を生成することを含む。一実施態様では、本方法は、サブセット内の複数の検体の各々に対して複数の配列決定サイクルの各々に対して、選択、抽出、畳み込み、処理、及び分類を反復することによって、複数の配列決定サイクルにおいてサブセット内の複数の検体に対してベースコールの配列を生成することを含む。一実施態様では、フローセルの非重複領域は、タイルである。一実施態様では、対応する画像チャンネルは、複数のフィルタ波長帯域のうちの1つである。一実施態様では、対応する画像チャンネルは、複数の画像イベントのうちの1つである。

【1219】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

（複数のサイクルでの複数のクラスターの同時のベースコール）

【1220】

我々は、フローセル上で検体をベースコールするニューラルネットワーク実装の方法を開示する。本方法は、画像セットの配列から入力画像データを取得することを含む。画像セットの配列は、フローセル上で検体を合成する配列決定動作の複数の配列決定サイクル上で生成される。画像セットの配列内の各画像は、フローセルの非重複領域をカバーし、複数の配列決定サイクルのうちのそれぞれの1つでの対応する画像チャンネル内に捕捉される、非重複領域上の検体のサブセット及びそれらの周囲の背景の強度放射を示す。本方法は、ニューラルネットワークの1つ又はそれ以上の層を介して、入力画像データを処理して、入力画像データの代替表現を生成することを含む。本方法は、出力層を介して代替表現を処理して、複数の配列決定サイクルの各々の各々でサブセット内の検体のうちの少なくともいくつかに組み込まれるヌクレオチド塩基（A、C、T、又はG）を識別する出力を生成し、それによって、複数の配列決定サイクルにおいて、サブセット内の検体のうちの少なくともいくつかの各々に対して、ベースコールの配列を生成することを含む。

10

【1 2 2 1】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1 2 2 2】

一実施態様では、出力層は、ソフトマックス層であり、出力は、A、C、T、及びGである、サブセット内の検体のうちの少なくともいくつかの各々で複数の配列決定サイクルの各々で組み込まれるヌクレオチド塩基の指数関数的に正規化されたスコア分布である。一実施態様では、入力画像データは、画像セットの配列内の画像を含む。一実施態様では、入力画像データは、画像セットの配列内の画像の各々からの少なくとも1つの画像パッチを含む。一実施態様では、ニューラルネットワークは、畳み込みニューラルネットワークである。別の実施態様では、ニューラルネットワークは、残留ニューラルネットワークである。更に別の実施態様では、ニューラルネットワークは、反復ニューラルネットワークである。

20

【1 2 2 3】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

（反復畳み込みベースのベースコール）

【1 2 2 4】

我々は、ベースコールに対するニューラルネットワークベースのシステムを開示する。システムは、反復モジュール及び畳み込みモジュールを有するハイブリッドニューラルネットワークを含む。反復モジュールは、畳み込みモジュールからの入力を使用する。畳み込みモジュールは、1つ又はそれ以上の畳み込み層を介して、配列決定動作の一連の配列決定サイクルに対する画像データを処理し、画像データの1つ又はそれ以上の畳み込み表現を生成する。画像データは、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す。反復モジュールは、畳み込み表現及び以前の隠れ状態表現を畳み込むことに基づいて、現在の隠れ状態表現を生成する。出力モジュールは、現在の隠れ状態表現に基づいて、検体のうちの少なくとも1つ、及び配列決定サイクルのうちの少なくとも1つに対してベースコールを生成する。

40

【1 2 2 5】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴

50

が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1226】

我々は、ベースコールのニューラルネットワーク実装の方法を開示する。本方法は、畳み込みニューラルネットワークの畳み込み層のカスケードを介して、サイクルごとの入力データの配列内のサイクルごとの各入力データを別々に処理することを含む。サイクルごとの入力データの配列は、配列決定動作の一連の配列決定サイクルに対して生成され、サイクルごとの各入力データは、それぞれの配列決定サイクルで捕捉される、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す画像チャネルを含む。本方法は、各配列決定サイクルに対して、別個の処理に基づいて、畳み込み層の各々で畳み込み表現を生成し、それによって、畳み込み表現の配列を生成することと、そのサイクルごとの入力データを、畳み込み表現のその対応する配列と混合し、混合表現を生成することと、その混合表現を平坦化し、平坦化された混合表現を生成することと、を含む。本方法は、連続する配列決定サイクルの平坦化された混合表現をスタックとして配置することを含む。本方法は、スライディングウィンドウベースでスタック内の平坦化された混合表現のサブセットにおいて畳み込む反復ニューラルネットワークを介して前方及び後方の方向にスタックを処理することを含み、各スライディングウィンドウがそれぞれの配列決定サイクルに対応し、(i)スタックにおける現在のスライディングウィンドウでの平坦化された混合表現のサブセット、及び(ii)以前の隠れ状態表現に基づいて、各配列決定サイクルに対して各時間ステップで現在の隠れ状態表現を連続的に生成する。本方法は、前方及び後方の方向でのスタックの処理の結果に基づいて、配列決定サイクルの各々で検体の各々をベースコールすることを含む。反復ニューラルネットワークは、LSTM及びGRUなどのゲートされた反復ニューラルネットワークであり得る。

【1227】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1228】

本方法は、時間ステップベースで所与の配列決定サイクルの前方及び後方の現在の隠れ状態表現を組み合わせることによって、所与の配列決定サイクルで検体の各々をベースコールし、組み合わせられた隠れ状態表現を生成することと、1つ又はそれ以上の完全に接続されたネットワークを介して、組み合わせられた隠れ状態表現を処理し、高密度表現を生成することと、ソフトマックス層を介して、高密度表現を処理して、A、C、T、及びGである、所与の配列決定サイクルで検体の各々に組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、を含む。一実施態様では、組み合わせることは、連結を含む。別の実施態様では、組み合わせることは、積算を含む。更に別の実施態様では、組み合わせることは、平均を含む。

【1229】

一実施態様では、サイクルごとの各入力データは、画像チャネルを補い、かつ対応する画像チャネル内のピクセルと1つ又はそれ以上の検体中心との間の中心間距離を含む、距離チャネルを含む。一実施態様では、サイクルごとの各入力データは、画像チャネル内の1つ又はそれ以上のピクセルの平均強度に基づいて、画像チャネルを補い、かつスケールリング値を含む、スケールリングチャネルを含む。一実施態様では、混合は、畳み込み表現及びサイクルごとの入力データを連結することを更に含む。一実施態様では、混合は、畳み込み表現及びサイクルごとの入力データを積算することを更に含む。一実施態様では、平坦化された混合表現は、2次元アレイである。一実施態様では、平坦化された混合表現のサブセットは、3次元容積である。一実施態様では、反復ニューラルネットワークは、3次元の畳み込みを3次元容積に適用する。一実施態様では、3次元畳み込みは、同じパデ

イングを使用する。一実施態様では、畳み込み層は、同じパディングを使用する。一実施態様では、反復ニューラルネットワークは、入力ゲート、活性化ゲート、忘却ゲート、及び出力ゲートを含む、長い短期メモリ（LSTM）ネットワークである。そのような実施態様では、本方法は、（i）スタックにおける現在のスライディングウィンドウ内の平坦化された混合表現のサブセット、及び（ii）入力ゲート、活性化ゲート、忘却ゲート、及び出力ゲートを介した以前の隠れ状態表現を処理することと、各配列決定サイクルに対して各時間ステップで現在の隠れ状態表現を生成することと、を含む。入力ゲート、活性化ゲート、忘却ゲート、及び出力ゲートは、（i）スタックにおける現在のスライディングウィンドウ内の平坦化された混合表現のサブセット、及び（ii）以前の隠れ状態表現に対して畳み込みを適用する。

10

【1230】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1231】

一実施態様では、ベースコールのニューラルネットワーク実装の方法は、畳み込みモジュールの1つ又はそれ以上の畳み込み層を介して、配列決定動作の一連の配列決定サイクルに対する画像データを畳み込むことと、画像データの1つ又はそれ以上の畳み込み表現を生成することと、を含む。画像データは、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す。本方法は、反復モジュールを介して、畳み込み表現及び以前の隠れ状態表現を畳み込むことと、現在の隠れ状態表現を生成することと、を含む。本方法は、出力モジュールを介して、現在の隠れ状態表現を処理することと、検体のうちの少なくとも1つ、及び配列決定サイクルのうちの少なくとも1つに対してベースコールを生成することと、を含む。

20

【1232】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

（品質スコアの推測）

【1233】

我々は、ニューラルネットワークベースのベースコーラーによってコールされる塩基に品質スコアを割り当てる、コンピュータ実装の方法を開示する。本方法は、訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される、予測されたベースコールの分類スコアを量子化することを含む。本方法は、量子化分類スコアとそれらのベースコール誤差率との間の適合を決定することを含む。すなわち、各量子化分類スコアに対して、量子化分類スコアを割り当てられる訓練データ内の訓練例のセットが決定される。訓練例の決定されたセット内の各訓練例に対して、訓練例に対する予測されたベースコールは、訓練例に対するグラウンドトゥルースベースコールと比較され、誤差率は、特定の量子化分類スコアに対する誤差率を提供するために、訓練例の決定されたセットにわたる比較から決定される。本方法は、適合に基づいて、品質スコアを量子化分類スコアに相関させることを含む。

40

50

【1234】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1235】

一実施態様では、量子化分類スコアのセットは、訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される予測されたベースコールの分類スコアのサブセットを含み、分類スコアは、実数である。一実施態様では、量子化分類スコアのセットは、訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される予測されたベースコールの全ての分類スコアを含み、分類スコアは、実数である。一実施態様では、分類スコアは、1になる指数関数的に正規化されたソフトマックススコアであり、ニューラルネットワークベースのベースコーラーのソフトマックス出力層によって生成される。一実施態様では、量子化分類スコアのセットは、

【数39】

$$0.9 \sum_{i=1}^n 0.1^{(i-1)}$$

と定義される選択式に基づいて選択され、ソフトマックススコアに適用される。一実施態様では、量子化分類スコアのセットは、

【数40】

$$\forall_{i=1}^{n=10} 0.1^i$$

と定義される選択式に基づいて選択され、ソフトマックススコアに適用される。一実施態様では、本方法は、相関に基づいて、推測中にニューラルネットワークベースのベースコーラーによってコールされる塩基に品質スコアを割り当てることを含む。一実施態様では、本方法は、推測中にニューラルネットワークベースのベースコーラーによってコールされる塩基に品質スコア対応スキームを適用することに基づいて、品質スコアを割り当てることを含む。そのような実施態様では、スキームは、推測中に、推測データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される、分類スコアの範囲をセット内の対応する量子化分類スコアにマッピングする。一実施態様では、本方法は、推測中に、品質スコアが現在のベースコールサイクルに対する設定された閾値を下回る検体をベースコールすることを停止させることを含む。一実施態様では、本方法は、推測中に、平均品質スコアが連続するベースコールサイクルの後に設定された閾値を下回る検体をベースコールすることを停止させることを含む。一実施態様では、予測されたベースコールを対応するグラウンドトゥールスペースコールと比較するために使用されるサンプルサイズは、各量子化分類スコアに固有である。一実施態様では、予測されたベースコールを対応するグラウンドトゥールスペースコールと比較するために使用されるサンプルサイズは、各量子化分類スコアに固有である。一実施態様では、適合は、回帰モデルを使用して決定される。一実施態様では、本方法は、各量子化分類スコアに対して、その予測されたベースコールを対応するグラウンドトゥールスペースコールと比較することによってベースコール精度率を決定することと、量子化分類スコアとそれらのベースコール精度率との間の適合を決定することと、を含む。一実施態様では、対応するグラウンドト

ールスペースコールは、多数の配列決定器具、配列決定化学、及び配列決定プロトコルで配列決定される、十分に特徴付けられたヒト及び非ヒトサンプルから導出される。

【1236】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

(品質スコアの予測)

【1237】

我々は、並行して動作し、かつメモリに結合された多数のプロセッサ上で動作するニューラルネットワークベースの品質スコアラを開示する。システムは、多数のプロセッサ上で動作する畳み込みニューラルネットワークを含む。畳み込みニューラルネットワークは、畳み込みニューラルネットワークのベースコール品質予測をベースコール品質グラウンドトゥールズと漸進的に一致させる逆伝搬ベースの勾配更新技術を使用して、配列決定画像からのデータを含む訓練例で訓練され、ベースコール品質グラウンドトゥールズでラベル付けされる。システムは、多数のプロセッサのうちの少なくとも1つで動作し、かつ1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質状態を決定するために、1つ又はそれ以上の配列決定サイクルで捕捉される配列決定画像からのデータを畳み込みニューラルネットワークに供給する、畳み込みニューラルネットワークの入力モジュールを含む。システムは、多数のプロセッサのうちの少なくとも1つで動作し、かつ畳み込みニューラルネットワークによる分析を、1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質状態を識別する出力に変換する、畳み込みニューラルネットワークの出力モジュールを含む。

【1238】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1239】

一実施態様では、出力モジュールは、高品質、中品質、及び低品質である品質状態に対する尤度を生成するソフトマックス分類層を更に備える。そのような実施態様では、尤度に基づいて、品質状態は、高品質、中質、又は低品質と分類される。一実施態様では、ソフトマックス分類層は、複数の品質スコアを割り当てられている品質状態に対する尤度を生成する。そのような実施態様では、尤度に基づいて、品質状態は、複数の品質スコアのうちの1つから品質スコアを割り当てられる。一実施態様では、品質スコアは、ベースコール誤差確率に対数的に基づき、複数の品質スコアは、Q6、Q10、Q43、Q20、Q22、Q27、Q30、Q33、Q37、Q40、及びQ50を含む。一実施態様では、出力モジュールは、品質状態を識別する連続値を生成する回帰層を更に含む。一実施態様では、システムは、配列決定画像からのデータを、コールされる塩基に対する品質予測値で補い、配列決定画像からのデータと共に、品質予測値を畳み込みニューラルネットワークに供給する、補足入力モジュールを含む。一実施態様では、品質予測値は、オンライン重複、純度、フェイジング、start5、6量体スコア、モチーフ蓄積、endiness、近似ホモポリマー、強度減衰、最終チャスティティ、背景を有する信号重複(SOWB)、及び／又はシフトされた純度G調整を含む。一実施態様では、品質予測値は、ピーク高さ、ピーク幅、ピーク場所、相対的なピーク場所、ピーク高さ比、ピーク間隔比、及び／又はピーク対応を含む。

【1240】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行する

ためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1241】

我々はまた、品質スコアリングのニューラルネットワーク実装の方法を開示する。本方法は、1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質状態を決定するために、1つ又はそれ以上の配列決定サイクルで補足される配列決定画像からのデータを畳み込みニューラルネットワークに供給することを含む。畳み込みニューラルネットワークは、配列決定画像からのデータを含み、かつベースコール品質グラウンドトゥールズでラベル付けされる訓練例で訓練される。訓練は、畳み込みニューラルネットワークのベースコール品質予測をベースコール品質グラウンドトゥールズと漸進的に一致させる、逆伝搬ベースの勾配更新技術を使用することを含む。本方法は、畳み込みニューラルネットワークによる分析を、1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質を識別する出力に変換することを含む。

【1242】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1243】

一実施態様では、コンピュータ実装の方法は、ニューラルネットワークを介して1つ又はそれ以上の検体に対して入力データを処理し、入力データの代替表現を生成することと、出力層を介して代替表現を処理して出力を生成することとであって、出力が、A、C、T、及びGである、検体のうちの特定の1つに組み込まれる塩基の尤度を識別する、生成することと、出力に基づいて、1つ又はそれ以上の検体に対して塩基をコールすることと、出力によって識別される尤度に基づいて、コールされた塩基の品質を決定することと、を含む。

【1244】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1245】

我々は、並行して動作し、かつメモリに結合された多数のプロセッサ上で動作するニューラルネットワークベースの品質スコアラーを開示する。システムは、ニューラルネットワークのベースコール品質予測をベースコール品質グラウンドトゥールズと漸進的に一致させる逆伝搬ベースの勾配更新技術を使用して、配列決定画像からのデータを含む訓練例で訓練され、かつベースコール品質グラウンドトゥールズでラベル付けされる、多数のプロセッサ上で動作するニューラルネットワークを含む。システムは、多数のプロセッサのうちの少なくとも1つで動作し、かつ1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質状態を決定するために、1つ又はそれ以上の配列決定サイクルで捕捉される配列決定画像からのデータをニューラルネットワークに供給する、ニューラ

10

20

30

40

50

ルネットワークの入力モジュールを含む。システムは、多数のプロセッサのうちの少なくとも1つで動作し、かつニューラルネットワークによる分析を、1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質状態を識別する出力に変換する、ニューラルネットワークの出力モジュールを含む。

【1246】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

(エンドツーエンド統合)

【1247】

コンピュータ実装の方法が提供され、本方法は、ニューラルネットワークを介して配列決定動作の1つ又はそれ以上の配列決定サイクルに対して配列決定システムによって捕捉される検体及びそれらの周囲の背景の画像を含む第1の画像データを処理することと、配列決定動作の1つ又はそれ以上の配列決定サイクルの検体のうちの1つ又はそれ以上に対してベースコールを生成することと、を含む。本方法は、1つ又はそれ以上の配列決定サイクルを実行して、検体の画像及びそれらの周囲の背景を捕捉することを含み得る。いくつかの実施態様では、本方法は、複数の配列決定サイクルを実行することを含み、複数の配列決定サイクルの各々は、画像データを生成する。コンピュータ実装の方法は、第1のニューラルネットワークを介して第1の入力を処理することと、第1の出力を生成することと、を含み得る。第1の入力は、配列決定動作のための配列決定システムによって捕捉される、検体及びそれらの周囲の背景の画像から導出される第1の画像データを含む。本方法は、ポストプロセッサを介して第1の出力を処理することと、第1の画像データのそれぞれの部分の1つ又はそれ以上の特性を示す、すなわち、検体及びそれらの周囲の背景に関するテンプレートデータを生成することと、を含み得る。本方法は、第2のニューラルネットワークを介して第2の入力を処理することと、第2の出力を生成することと、を含み得る。第2の入力は、テンプレートデータを使用して修正される第1の画像データ、テンプレートデータを使用して修正される第2の画像データ、及び／若しくは第1及び／若しくは第2の画像データ、並びに補足データを含み得る。補足データは、テンプレートデータを含み得る。第2の画像データは、検体及びそれらの周囲の背景の画像から導出される。第2の出力は、配列決定動作の1つ又はそれ以上の配列決定サイクルで、検体のうちの1つ又はそれ以上に対するベースコールを識別する。

【1248】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1249】

一実施態様では、メタデータは、アップサンプリングされたサブピクセル解像度でのテンプレート画像を含み、メタデータに基づいて、テンプレート画像内の各サブピクセルは、背景サブピクセル、検体中心サブピクセル、又は検体内部サブピクセルのいずれかとして識別される。一実施態様では、検体及びそれらの周囲の背景の画像は、光学ピクセル解像度で捕捉される。一実施態様では、本方法は、画像内のピクセルに対応する、テンプレート画像内のどれくらい多くのサブピクセルが検体のうちの1つ又はそれ以上の一部を含むかに基づいて、画像内のピクセルに対して領域重み付け係数を決定することと、領域重み付け係数に基づいて、ピクセルの強度を修正することと、第2のニューラルネットワークによるベースコールに対する第3の画像データとして第2の入力での修正強度を有するピクセルを含むことと、を含む。一実施態様では、本方法は、アップサンプリングされた

10

20

30

40

50

サブピクセル解像度に画像をアップサンプリングすることと、アップサンプリングされた画像を生成することと、を含む。アップサンプリングは、テンプレート画像内の背景サブピクセルに対応するアップサンプリングされた画像内のサブピクセルに背景強度を割り当てることと、テンプレート画像内の検体中心サブピクセル及び検体内部サブピクセルに対応するアップサンプリングされた画像内のサブピクセルに検体強度を割り当てることと、第2の入力内のアップサンプリングされた画像を、第2のニューラルネットワークによるベースコールに対する第3の画像データとして含むことと、を含む。一実施態様では、背景強度は、ゼロ又は最小値を有する。一実施態様では、検体強度は、光学ピクセル解像度内のピクセルの強度を補間することによって決定される。一実施態様では、本方法は、アップサンプリングされたサブピクセル解像度に画像をアップサンプリングすることと、アップサンプリングされた画像を生成することと、を含む。アップサンプリングは、テンプレート画像内の検体中心サブピクセル及び検体内部サブピクセルに対応するアップサンプリングされた画像内のピクセルの構成サブピクセルのみの間で、光学ピクセルドメイン内のピクセルの全体の強度を分散させることと、第2の入力内のアップサンプリングされた画像を、第2のニューラルネットワークによるベースコールに対する第3の画像データとして含むことと、を含む。一実施態様では、メタデータは、検体の中心を識別する。別の実施態様では、メタデータは、検体の形状を識別する。更に別の実施態様では、メタデータは、検体間の境界を識別する。一実施態様では、本方法は、第2の出力に基づいてベースコールの品質を決定することを含む。

【1250】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1251】

我々は、第1のニューラルネットワークを使用して、検体に関するメタデータを決定することであって、メタデータが、検体の中心、検体の形状、及び／又は検体間の境界を識別する、決定することと、第2のニューラルネットワークを使用して、決定されたメタデータに基づいて検体をベースコールすることと、を含む、コンピュータ実装の方法を開示する。

【1252】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1253】

一実施態様では、本方法は、決定されたメタデータに基づいて、第2のニューラルネットワークによる処理に対する入力を構築することを含む。入力は、第2のニューラルネットワークによる処理で検体の中心、形状、及び／又は境界に関するコンテキストを組み込む、修正強度値を含む。一実施態様では、本方法は、第2のニューラルネットワークを介して修正強度値を処理して、検体をベースコールすることを含む。一実施態様では、本方法は、決定されたメタデータに基づいて導出される補足データで処理するために、第2のニューラルネットワークに供給される入力を伴わせることを含む。補足データは、第2のニューラルネットワークによる処理で検体の中心、形状、及び／又は境界に関するコンテキストを組み込む。一実施態様では、本方法は、第2のニューラルネットワークを介して入力及び補足データを処理して、検体をベースコールすることを含む。

【1254】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行する

ためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1255】

我々は、検体に対して配列決定手順を実行することを含む、コンピュータ実装の方法を開示する。配列決定手順は、複数の配列決定サイクルを含み、複数の配列決定サイクルの各々は、画像データを生成する。一実施態様では、本方法は、ニューラルネットワークを介して、複数の配列決定サイクルの各々に対して画像データを処理することと、複数の配列決定サイクルの各々で、検体のうちの少なくともいくつかに対してベースコールを生成することと、を含む。

10

【1256】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1257】

一実施態様では、本方法は、ニューラルネットワークを介して複数の配列決定サイクルの各々に対して画像データを処理する前に、別のニューラルネットワークを介して複数の配列決定サイクルのうちのいくつかに対して画像データを処理することと、検体に関するメタデータを決定することと、を含む。メタデータは、検体の中心及び／又は形状を識別する。一実施態様では、本方法は、決定されたメタデータに基づいて、ニューラルネットワークを使用して、複数の配列決定サイクルの各々で、検体のうちの少なくともいくつかをベースコールすることを含む。

20

【1258】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

【1259】

我々は、バイオセンサーシステムに結合された受け部と、照明システムと、受け部に結合されており、かつ分析モジュールを有する、システムコントローラと、を備える、配列決定システムを開示する。バイオセンサーシステムは、光検出器のアレイを有するように構成されており、バイオセンサーシステムは、バイオセンサーを有し、バイオセンサーは、検体を含むように構成された反応部位を有する。照明システムは、励起光をバイオセンサーに向け、かつ反応部位内の検体を照明するように構成されている。検体のうちの少なくともいくつかは、照明されるときに発光信号を提供する。システムコントローラは、受け部に結合されており、分析モジュールを有する。分析モジュールは、複数の配列決定サイクルの各々で、光検出器から画像データを取得するように構成されている。画像データは、光検出器によって検出される発光信号から導出され、ニューラルネットワークを介して、複数の配列決定サイクルの各々に対して画像データを処理し、複数の配列決定サイクルの各々で検体のうちの少なくともいくつかに対するベースコールを生成する。

40

【1260】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

50

【1261】

我々は、アップサンプリングされたサブピクセル解像度でテンプレート画像で識別される検体中心、形状、及び境界を使用して、光学ピクセル解像度で捕捉される画像を解釈して、配列決定動作中にフローセルのタイル上で合成される検体をベースコールする、コンピュータ実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、光学ピクセル解像度で捕捉される1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、領域重み付け係数を有するテンプレート画像を生成することと、ニューラルネットワークベースのテンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、アップサンプリングされたサブピクセル解像度で検体の検体中心、形状、及び境界を識別することを含むことと、特定の検体の検体形状及び境界を評価して、特定の検体の一部を含む少なくとも1つのピクセルを識別して、識別されたピクセル内のどれくらい多くのサブピクセルが特定の検体の一部を含むかに基づいて、領域重み付け係数を設定して、テンプレート画像内の領域重み付け係数を記憶することと、特定の検体の一部も含むピクセルに対して、光学ピクセル解像度で捕捉される画像の各々でのピクセルに対して、評価を実行して、識別、設定、及び記憶することと、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、ニューラルネットワークベースのベースコーラーを介して画像の修正バージョンを処理して、修正バージョンの代替表現を生成することと、代替表現を使用して、特定の検体をベースコールすることと、を含む。

10

20

【1262】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1263】

一実施態様では、ベースコールは、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、画像の各々でのピクセルに対して、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、特定の検体に対して、各画像パッチが、ピクセルのアレイを有し、テンプレート画像で識別される特定の検体の中心をその中心ピクセル内に含むように、各修正バージョンから画像パッチを抽出することと、畳み込みニューラルネットワークを介して、画像の修正バージョンから抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、出力層を介して畳み込み表現を処理して、中心ピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで特定のクラスターに組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、を更に含む。一実施態様では、本方法は、ピクセル強度値を修正する前に、サイクル固有及び画像チャンネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを含む。

30

40

【1264】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行する

50

ためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1265】

我々は、アップサンプリングされたサブピクセル解像度でテンプレート画像で識別される検体中心、形状、及び境界を使用して、光学ピクセル解像度で捕捉される画像を解釈して、配列決定動作中にフローセルのタイル上で合成される検体をベースコールする、コンピュータ実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、光学ピクセル解像度で捕捉される1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、領域重み付け係数を有するテンプレート画像を生成することと、ニューラルネットワークベースのテンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、ピクセルが1次検体の一部を含む少なくとも1つの1次検体を決定して、ピクセル内のどれくらい多くのサブピクセルが1次検体の一部を含むかに基づいて、領域重み付け係数を設定することを含むことと、多数の検体及び多数のピクセルに対して、光学ピクセル解像度で捕捉される画像の各々でのピクセルに対して、評価を実行して、決定及び設定することと、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、ニューラルネットワークベースのベースコーラーの前方へのパスへの入力として、修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、ニューラルネットワークベースのベースコーラーを介して画像の修正バージョンを処理して、修正バージョンの代替表現を生成することと、前方へのパスの出力として、代替表現を使用して多数の検体の各々1つを同時にベースコールすることと、を含む。

【1266】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1267】

一実施態様では、ベースコールは、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、画像の各々でのピクセルに対して、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、各画像パッチが、ピクセルのアレイを有するように、各修正バージョンから画像パッチを抽出することと、畳み込みニューラルネットワークを介して、画像の修正バージョンから抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、出力層を介して畳み込み表現を処理して、アレイ内の各ピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、対応する検体の中心を含むそれぞれのピクセルに割り当てられる塩基分類に基づいて、多数の検体のうちの各々1つをベースコールすることと、を更に含む。一実施態様では、本方法は、ピクセル強度値を修正する前に、サイクル固有及び画像チャンネル固有の変換を使用して、光学ピクセル解像度

で捕捉される画像の各々をテンプレート画像と位置合わせすることを含む。

【1268】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1269】

我々は、アップサンプリングされたサブピクセル解像度でテンプレート画像で識別される検体中心、形状、及び境界を使用して、光学ピクセル解像度で捕捉される画像を解釈して、配列決定動作中にフローセルのタイル上で合成される検体をベースコールする、コンピュータ実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、光学ピクセル解像度で捕捉される1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、ニューラルネットワークベースのテンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、アップサンプリングされたサブピクセル解像度でテンプレート画像を生成することを含む。「複数の配列決定サイクルのうちの初期の1つ」によって、これは、1つ又はそれ以上の初期配列決定サイクル、例えば、配列決定サイクル1～10、2～10、2～8、又は2～7のうちの1つ又はそれ以上を指すことが理解されるであろう。テンプレート画像は、検体中心、背景を含み、光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングする検体に属するクラスにサブピクセルを分類し、アップサンプリングは、いかなる検体にも寄与しないとして、テンプレート画像で識別されるサブピクセルに背景強度を割り当てることと、ニューラルネットワークベースのベースコーラーの前方へのパスへの入力として、ニューラルネットワークベースのベースコーラーを介してアップサンプリングされた画像を処理して、アップサンプリングされた画像の代替表現を生成することと、前方へのパスの出力として、代替表現を使用して複数の検体を同時にベースコールすることと、を含む。

【1270】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

【1271】

一実施態様では、ベースコールは、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングすることとあって、いかなる検体にも寄与しないとして、テンプレート画像で識別されるサブピクセルに背景強度を割り当てることを含む、アップサンプリングすることと、各画像パッチがサブピクセルのアレイを有するように、各アップサンプリングされた画像から画像パッチを抽出することと、畳み込みニューラルネットワークを介して、アップサンプリングされた画像から抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、出力層を介して畳み込み表現を処理して、アレイ内の各サブピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで組

み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、対応する検体の中心を含むそれぞれのサブピクセルに割り当てられる塩基分類に基づいて、複数の検体のうちの各々1つをベースコールすることと、を更に含む。

【1272】

一実施態様では、本方法は、アップサンプリングの前に、サイクル固有及び画像チャンネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを含む。一実施態様では、アップサンプリングは、最近傍強度抽出、ガウス系強度抽出、平均 2×2 のサブピクセル領域に基づく強度抽出、最も明るい 2×2 のサブピクセル領域に基づく強度抽出、平均 3×3 のサブピクセル領域に基づく強度抽出、バイリニア強度抽出、双次強度抽出、及び／又は重み付け領域被覆に基づく強度抽出のうちの少なくとも1つを使用して実行される。一実施態様では、背景強度は、ゼロ値を有する。一実施態様では、背景強度は、ほぼゼロ値を有する。

10

【1273】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1274】

我々は、アップサンプリングされたサブピクセル解像度でテンプレート画像で識別される検体中心、形状、及び境界を使用して、光学ピクセル解像度で捕捉される画像を解釈して、配列決定動作中にフローセルのタイル上で合成される検体をベースコールする、コンピュータ実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、光学ピクセル解像度で捕捉される1つ又はそれ以上の画像チャンネルのうちのそれぞれの1つで検体及びそれらの周囲の背景の強度放射を示す。本方法は、サブピクセルごとの領域重み付け係数を有するテンプレート画像を生成することと、ニューラルネットワークベースのテンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、アップサンプリングされたサブピクセル解像度で検体の検体中心、形状、及び境界を識別することを含むことと、検体の検体形状及び境界を評価して、それぞれのピクセル内のどれくらい多くのサブピクセルが任意の検体の一部を含むかを決定し、それぞれのピクセル内のサブピクセルに対してサブピクセルごとの領域重み付け係数を設定し、テンプレート画像内のサブピクセルごとの領域重み付け係数を記憶することと、光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングすることと、サブピクセルごとの領域重み付け係数を適用することによって、いかなる検体にも寄与しないとして、テンプレート画像で識別されるそれぞれのピクセルの第1のサブピクセル間でそれぞれのピクセルの強度を分散させることを含む、アップサンプリングすることと、いかなる検体にも寄与しないとして、テンプレート画像で識別されるそれぞれのピクセル内の第2のサブピクセルに背景強度を割り当てることと、ニューラルネットワークベースのベースコーラーの前方へのパスへの入力として、ニューラルネットワークベースのベースコーラーを介してアップサンプリングされた画像を処理して、アップサンプリングされた画像の代替表現を生成することと、前方へのパスの出力として、代替表現を使用して複数の検体を同時にベースコールすることと、を含む。

20

30

40

【1275】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。

50

【1276】

一実施態様では、ベースコールは、複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングすることと、サブピクセルごとの領域重み付け係数を適用することによって、いかなる検体にも寄与しないとして、テンプレート画像で識別されるそれぞれのピクセルの第1のサブピクセル間でそれぞれのピクセルの強度を分散させることを含む、アップサンプリングすることと、いかなる検体にも寄与しないとして、テンプレート画像で識別されるそれぞれのピクセル内の第2のサブピクセルに背景強度を割り当てることと、各画像パッチがサブピクセルのアレイを有するように、各アップサンプリングされた画像から画像パッチを抽出することと、畳み込みニューラルネットワークを介して、アップサンプリングされた画像から抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、出力層を介して畳み込み表現を処理して、アレイ内の各サブピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで組み込まれる塩基の尤度を生成することと、尤度に基づいて、塩基をA、C、T、又はGと分類することと、対応する検体の中心を含むそれぞれのサブピクセルに割り当てられる塩基分類に基づいて、複数の検体のうちの各々1つをベースコールすることと、を更に含む。

10

20

【1277】

一実施態様では、本方法は、アップサンプリングの前に、サイクル固有及び画像チャネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを含む。一実施態様では、背景強度は、ゼロ値を有する。別の実施態様では、背景強度は、ほぼゼロ値を有する。

【1278】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

【1279】

一実施態様では、コンピュータ実装の方法は、特定の検体に対してアップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価して、特定の検体の一部を含む少なくとも1つのピクセルを識別して、識別されたピクセル内のどれくらい多くのサブピクセルが特定の検体の一部を含むかに基づいて、領域重み付け係数を設定することと、特定の検体の一部を含む識別されたピクセルに隣接するピクセルに対して、評価を実行して、決定及び設定することと、それぞれのピクセルに対する領域重み付け係数に基づく処理のために、識別されたピクセル及び隣接するピクセルのピクセル強度値を修正することと、を含む。

40

【1280】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ

50

以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

【1281】

一実施態様では、コンピュータ実装の方法は、アップサンプリングされたサブピクセルドメイン内のテンプレート画像内のピクセルを評価して、ピクセルが1次検体の一部を含む少なくとも1次検体を決定して、識別されたピクセル内のどれくらい多くのサブピクセルが1次検体の一部を含むかに基づいて、領域重み付け係数を設定することと、光学画像のフィールド内の多数のピクセルに対して、評価を実行して、決定及び設定することと、それぞれのピクセルに対する領域重み付け係数に基づく処理のために、識別されたピクセル及び隣接するピクセルのピクセル強度値を修正することと、を含む。

10

【1282】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

20

【1283】

一実施態様では、コンピュータ実装の方法は、アップサンプリングされたサブピクセルドメイン内のテンプレート画像にアクセスすることであって、テンプレート画像が、任意の検体の一部を含むサブピクセルを識別する、アクセスすることと、光学画像のフィールドのサブピクセルドメインへのアップサンプリング中に、いかなる検体にも寄与しないとして、テンプレート画像で識別されるサブピクセルに背景強度を割り当てることと、を含む。

【1284】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

30

【1285】

一実施態様では、コンピュータ実装の方法は、アップサンプリングされたサブピクセルドメイン内のテンプレート画像内の識別されたピクセルを評価して、識別されたピクセル内のどれくらい多くのサブピクセルが任意の検体の一部を含むかを決定して、識別されたピクセル内のサブピクセルに対してサブピクセルごとの領域重み付け係数を設定することと、光学画像のフィールド内の多数のピクセルに対して、評価を実行して、決定及び設定することと、テンプレート画像内の多数のピクセルに対してサブピクセルごとの領域重み付け係数を記憶することと、光学画像のフィールドのサブピクセルドメインへのアップサンプリング中に、サブピクセルごとの領域重み付け係数を適用することによって、任意の検体に寄与するとしてテンプレート画像で識別される特定のピクセルの第1のサブピクセル間の特定のピクセルの強度を分散させることと、任意の検体に寄与しないとしてテンプレ

40

50

レートで識別される特定のピクセルの第2のサブピクセルに背景強度を割り当てることと、を含む。

【1286】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

10

【1287】

我々は、アップサンプリングされたサブピクセル解像度でテンプレート画像で識別されるクラスター中心、形状、及び境界を使用して、光学ピクセル解像度で捕捉される画像を解釈して、配列決定動作中にフローセルのタイル上で合成されるデオキシリボ核酸（DNA）クラスターをベースコールする、コンピュータ実装の方法を開示し、配列決定動作は、複数の配列決定サイクルを有し、複数の配列決定サイクルの各々は、1つ又はそれ以上の画像を有する画像セットを生成し、画像の各々は、光学ピクセル解像度で捕捉される1つ又はそれ以上の撮像チャンネルのうちのそれぞれの1つでDNAクラスター及びそれらの周囲の背景の強度放射を示す。本方法は、領域重み付け係数を有するテンプレート画像を生成することと、ニューラルネットワークベースのテンプレート生成器を介して複数の配列決定サイクルのうちの初期の1つでそれぞれ生成される初期画像セットを処理して、ピクセルが1次DNAクラスターの一部を含む少なくとも1つの1次DNAクラスターを決定して、ピクセル内のどれくらい多くのサブピクセルが1次DNAクラスターの一部を含むかに基づいて、領域重み付け係数を設定することを含むことと、多数のDNAクラスター及び多数のピクセルに対して、評価を実行して、決定及び設定することと、画像内のピクセルで領域重み付け係数をピクセルごとに符号化することによって、光学ピクセル解像度で捕捉される画像の各々を、領域重み付け係数を有するテンプレート画像で補うことと、ニューラルネットワークベースのベースコーラーの前方へのパスへの入力として、ニューラルネットワークベースのベースコーラーを介して画像及び補足テンプレート画像を処理して、入力の代替表現を生成することと、前方へのパスの出力として、代替表現を使用して多数のDNAクラスターの各々1つを同時にベースコールすることと、を含む。

20

30

【1288】

他の実施態様のための特定の実施態様セクションで説明される特徴のそれぞれは、この実施態様に等しく適用される。上記のように、全ての他の特徴はここでは繰り返されず、参照により繰り返されるべきである。読者は、これらの実施態様において識別された特徴が、他の実施態様で識別された塩基特徴のセットと容易に組み合わせることができるかを理解するであろう。このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行するように動作可能な1つ又はそれ以上のプロセッサとを含むシステムを含むことができ、上記の方法のいずれかを実行することができる。

40

項目

【1289】

本開示はまた、以下の項目を含む。

項目セット

1. ニューラルネットワークを介して配列決定動作の1つ又はそれ以上の配列決定サイ

50

クルに対して配列決定システムによって捕捉される検体及びそれらの周囲の背景の画像を含む第1の画像データを処理することと、配列決定動作の1つ又はそれ以上の配列決定サイクルの検体のうちの1つ又はそれ以上に対してベースコールを生成することと、を含む、コンピュータ実装の方法。

2. 第1の画像データを処理することが、

第1のニューラルネットワークを介して第1の入力を処理して、第1の出力を生成することであって、第1の入力が、第1の画像データを含む、生成することと、

ポストプロセッサを介して第1の出力を処理して、第1の画像データのそれぞれの部分の1つ又はそれ以上の特性を示すテンプレートデータを生成することと、

第2のニューラルネットワークを介して第2の入力を処理して、第2の出力を生成することであって、第2の入力が、第1の画像データ及び補足データを含み、補足データが、テンプレートデータを含み、第2の出力が、配列決定動作の1つ又はそれ以上の配列決定サイクルで、検体のうちの1つ又はそれ以上に対するベースコールを識別する、生成することと、を含む、項目1に記載のコンピュータ実装の方法。

10

3. 第1の画像データを処理することが、

第1のニューラルネットワークを介して第1の入力を処理して、第1の出力を生成することであって、第1の入力が、第1の画像データを含む、生成することと、

ポストプロセッサを介して第1の出力を処理して、第1の画像データのそれぞれの部分の1つ又はそれ以上の特性を示すテンプレートデータを生成することと、

第2のニューラルネットワークを介して第2の入力を処理して、第2の出力を生成することであって、第2の入力が、テンプレートデータを使用して修正される第1の画像データを含み、第2の出力が、配列決定動作の1つ又はそれ以上の配列決定サイクルで、検体のうちの1つ又はそれ以上に対するベースコールを識別する、生成することと、を含む、項目1に記載のコンピュータ実装の方法。

20

4. 第2の入力が、テンプレートデータを使用して修正される第2の画像データを更に含み、第2の画像データが、配列決定動作の1つ又はそれ以上の追加の配列決定サイクルに対して、配列決定システムによって補足される検体及びそれらの周囲の背景の画像を含む、項目3に記載のコンピュータ実装の方法。

5. テンプレートデータが、テンプレート画像を含み、テンプレート画像が、アップサンプリングされたサブピクセル解像度である、項目2～5のいずれか一項に記載のコンピュータ実装の方法。

30

6. テンプレート画像内の各サブピクセルが、背景サブピクセル、検体中心サブピクセル、又は検体内部サブピクセルのいずれかとして識別される、項目5に記載のコンピュータ実装の方法。

7. 検体及びそれらの周囲の背景の画像が、光学ピクセル解像度で補足される、項目1～6のいずれか一項に記載のコンピュータ実装の方法。

8. テンプレートデータを使用する修正が、

第1及び／又は第2の画像データの画像内のピクセルに対応するテンプレートデータ内のどれくらい多くのサブピクセルが、検体のうちの1つ又はそれ以上の一部を含むかに基づいて、第1及び／又は第2の画像データ内の1つ又はそれ以上のピクセルに対する領域重み付け係数を計算することと、

40

領域重み付け係数に基づいて、ピクセルの強度を修正することと、を含む、項目3～7のいずれか一項に記載のコンピュータ実装の方法。

9. テンプレートデータを使用する修正が、

検体及びそれらの周囲の背景の画像をアップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、テンプレート画像内の背景サブピクセルに対応するアップサンプリングされた画像内のサブピクセルに背景強度を割り当てることと、テンプレート画像内の検体中心サブピクセル及び検体内部サブピクセルに対応するアップサンプリングされた画像内のサブピクセルに検体強度を割り当てることと、を含む、項目6に記載の、又は項目6に従属するときの項目7若しく

50

は 8 に記載のコンピュータ実装の方法。

10．背景強度が、ゼロ値を有する、項目 9 に記載のコンピュータ実装の方法。

11．検体強度が、光学ピクセル解像度でピクセルの強度を補間することによって決定される、項目 9 又は 10 に記載のコンピュータ実装の方法。

12．テンプレート画像を使用する修正が、

検体及びそれらの周囲の背景の画像をアップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、テンプレート画像内の検体中心サブピクセル及び検体内部サブピクセルに対応するアップサンプリングされた画像内のピクセルの構成サブピクセルのみの間で、光学ピクセルドメイン内のピクセルの全体の強度を分散させることと、を含む、項目 6 に記載の、又は項目 6 に従属する
10
ときの項目 7 ～ 11 のいずれか一項に記載のコンピュータ実装の方法。

13．テンプレートデータが、検体の空間分布、検体形状、検体の中心、及び検体境界からなる群から選択される特性のうちの少なくとも 1 つを識別する、項目 2 ～ 12 のいずれか一項に記載のコンピュータ実装の方法。

14．第 2 の出力に基づいて、ベースコールの品質を計算することを更に含む、項目 2 ～ 13 のいずれか一項に記載のコンピュータ実装の方法。

15．1 つ又はそれ以上の配列決定サイクルを実行して、検体及びそれらの周囲の背景の画像を捕捉することを更に含む、項目 1 ～ 14 のいずれか一項に記載のコンピュータ実装の方法。

16．複数の配列決定サイクルを実行することを更に含み、複数の配列決定サイクルの各々が、画像データを生成する、項目 1 ～ 15 のいずれか一項に記載のコンピュータ実装の方法。
20

17．第 1 のニューラルネットワークを使用して、検体に関するテンプレートデータを決定することであって、テンプレートデータが、検体の空間分布、検体形状、検体の中心、及び検体境界からなる群から選択される特性のうちの少なくとも 1 つを識別する、決定することと、

第 2 のニューラルネットワークを使用して、テンプレートデータに基づいて検体をベースコールすることと、を含む、コンピュータ実装の方法。

18．テンプレートデータが、検体の空間分布、検体形状、検体の中心、及び検体境界からなる群から選択される特性のうちの少なくとも 1 つを識別するための修正強度値と、
30

第 2 のニューラルネットワークを介して修正強度値を処理して、検体をベースコールすることと、を含む、項目 17 に記載のコンピュータ実装の方法。

19．テンプレートデータが、テンプレート画像を含む、項目 17 又は 18 に記載のコンピュータ実装の方法。

20．少なくとも 1 つの特定の検体に対してアップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価して、少なくとも 1 つの特定の検体の一部を含むピクセル、及び少なくとも 1 つの特定の検体の一部も含む、ピクセルに隣接するピクセルを識別することと、

識別されたピクセルの各々でのどれくらい多くのサブピクセルが、少なくとも 1 つの特定の検体の一部を含むかに基づいて、各ピクセルに対する領域重み付け係数を計算することと、
40

それぞれのピクセルに対する領域重み付け係数に基づく処理のために、識別されたピクセル及び隣接するピクセルのピクセル強度値を修正することと、を更に含む、項目 19 に記載のコンピュータ実装の方法。

21．テンプレート画像を評価することが、

第 1 のニューラルネットワークを介して、複数の配列決定サイクルのうちの 1 つ又はそれ以上の初期配列決定サイクルでそれぞれ生成される 1 つ又はそれ以上の初期画像セットを処理して、テンプレート画像を生成して、アップサンプリングされたサブピクセル解像度で検体の中心、形状、及び境界を識別することであって、各画像セットが、1 つ又はそれ以上の画像を含み、画像の各々が、光学ピクセル解像度で捕捉される 1 つ又はそれ以上
50

の撮像チャネルのうちのそれぞれの1つでの検体及びそれらの周囲の背景の強度放射を示す、識別することを更に含む、項目20に記載のコンピュータ実装の方法。

22. テンプレート画像を評価することが、

少なくとも1つの特定の検体の検体形状及び境界を評価して、少なくとも1つの特定の検体の一部を含む少なくとも1つのピクセル、及び少なくとも1つの特定の検体の一部も含む、ピクセルに隣接するピクセルを識別することを更に含む、方法が、

テンプレート画像内に領域重み付け係数を記憶することと、

修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、

第2のニューラルネットワークを介して画像の修正バージョンを処理して、修正バージョンの代替表現を生成することと、

代替表現を使用して、少なくとも1つの特定の検体をベースコールすることと、を更に含む、項目20又は21に記載のコンピュータ実装の方法。

23. ベースコールが、

複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び

複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、

画像の各々でのピクセルに対して、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、

修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、

少なくとも1つの特定の検体に対して、各画像パッチが、

ピクセルのアレイを有し、かつ

その中心ピクセルで、テンプレート画像で識別される特定の検体の中心を含むように、各修正バージョンから画像パッチを抽出することと、

第2のニューラルネットワークの畳み込みニューラルネットワークを介して、画像の修正バージョンから抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、

出力層を介して畳み込み表現を処理して、中心ピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで少なくとも1つの特定の検体に組み込まれる塩基の尤度を生成することと、

尤度に基づいて、塩基をA、C、T、又はGと分類することと、を更に含む、項目22に記載のコンピュータ実装の方法。

24. ピクセル強度値を修正する前に、サイクル固有及び撮像チャネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを更に含む、項目22又は23に記載のコンピュータ実装の方法。

25. アップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価して、任意の検体の一部を含むサブピクセルを識別することと、

任意の検体に寄与しないとしてテンプレート画像で識別されるサブピクセルに背景強度を割り当てることと、を更に含む、項目19に記載のコンピュータ実装の方法。

26. アップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価することが、

少なくとも1つのピクセル内のどれくらい多くのサブピクセルが、任意の検体の一部を含むかを計算することと、少なくとも1つのピクセル内のサブピクセルに対するサブピクセルごとの領域重み付け係数を計算することと、を更に含む、項目25に記載のコンピュータ実装の方法。

27. 方法が、

10

20

30

40

50

第1のニューラルネットワークを介して、複数の配列決定サイクルのうちの1つ又はそれ以上の初期配列決定サイクルでそれぞれ生成される1つ又はそれ以上の初期画像セットを処理して、アップサンプリングされたサブピクセル解像度でテンプレート画像を生成することであって、各画像セットが、1つ又はそれ以上の画像を含み、画像の各々が、光学ピクセル解像度で捕捉される1つ又はそれ以上の撮像チャンネルのうちのそれぞれの1つでの検体及びそれらの周囲の背景の強度放射を示し、テンプレート画像が、検体中心、背景、及び検体内部を含むクラスにサブピクセルを分類する、生成することと、

光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングすることと、任意の検体に寄与しないとしてテンプレート画像で識別される画像の各々のサブピクセルに背景強度を割り当てることと、

10

第2のニューラルネットワークを介してアップサンプリングされた画像を処理して、アップサンプリングされた画像の代替表現を生成することと、

代替表現を使用して、複数の検体をベースコールすることと、を含む、項目25又は26に記載のコンピュータ実装の方法。

28. 画像の各々をアップサンプリングすることが、

サブピクセルごとの領域重み付け係数を適用することによって、任意の検体に寄与するとしてテンプレート画像で識別される特定のピクセルの第1のサブピクセル間の特定のピクセルの強度を分散させることと、任意の検体に寄与しないとしてテンプレートで識別される特定のピクセルの第2のサブピクセルに背景強度を割り当てることと、を更に含む、項目27に記載のコンピュータ実装の方法。

20

29. アップサンプリングの前に、方法が、

複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び

複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることを含み、アップサンプリングした後に、方法が、

各画像パッチが、サブピクセルのアレイを有するように、各アップサンプリングされた画像から画像パッチを抽出することと、

30

第2のニューラルネットワークの畳み込みニューラルネットワークを介して、アップサンプリングされた画像から抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、

出力層を介して畳み込み表現を処理して、アレイ内の各サブピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで組み込まれる塩基の尤度を生成することと、

尤度に基づいて、塩基をA、C、T、又はGと分類することと、

対応する検体の中心を含むそれぞれのサブピクセルに割り当てられる塩基分類に基づいて、複数の検体のうちの各々1つをベースコールすることと、を含む、項目28に記載のコンピュータ実装の方法。

40

30. アップサンプリングの前に、サイクル固有及び撮像チャンネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを含み、項目28又は29に記載のコンピュータ実装の方法。

31. アップサンプリングが、最近傍強度抽出、ガウス系強度抽出、平均 2×2 のサブピクセル領域に基づく強度抽出、最も明るい 2×2 のサブピクセル領域に基づく強度抽出、平均 3×3 のサブピクセル領域に基づく強度抽出、バイリニア強度抽出、双次強度抽出、及び／又は重み付け領域被覆に基づく強度抽出のうちの少なくとも1つを使用して実行される、項目29又は30に記載のコンピュータ実装の方法。

32. バイオセンサーシステムに結合された受け部であって、バイオセンサーシステムが、光検出器のアレイを備えるように構成されており、バイオセンサーシステムが、バイ

50

オセンサーを備え、バイオセンサーが、検体を含むように構成された反応部位を含む、受け部と、

励起光をバイオセンサーに向け、かつ反応部位内の検体を照明するように構成された、照明システムであって、検体のうちの少なくともいくつかが、照明されるときに発光信号を提供する、照明システムと、

受け部に結合されており、かつ分析モジュールを備える、システムコントローラであって、分析モジュールが、

複数の配列決定サイクルの各々で、光検出器から、光検出器によって検出される発光信号から導出される画像データを取得し、

ニューラルネットワークを介して、複数の配列決定サイクルの各々に対して画像データを処理し、かつ複数の配列決定サイクルの各々で、検体のうちの少なくともいくつかに対してベースコールを生成するように構成されている、システムコントローラと、を備える、配列決定システム。

項目セット 2

1. ニューラルネットワークベースのテンプレート生成をニューラルネットワークベースのベースコールと統合することを含む、エンドツーエンド配列決定のコンピュータ実装の方法であって、

光学ピクセル解像度でピクセルを含む第 1 の画像データ及び第 2 の画像データにアクセスすることであって、

第 1 の画像データが、配列決定動作の配列決定サイクルのうちの初期の 1 つで、配列決定システムによって捕捉される、クラスター及びクラスターの周囲の背景の画像を含み、

第 2 の画像データが、配列決定動作の初期及び追加の配列決定サイクルで、配列決定システムによって捕捉される、クラスター及びそれらの周囲の背景の画像を含む、アクセスすることと、

ニューラルネットワークベースのテンプレート生成器を介して第 1 の画像データを処理し、クラスターメタデータを識別するクラスターマップを生成することであって

クラスターメタデータが、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界に基づいて、クラスターの空間分布情報を識別し、

ニューラルネットワークベースのテンプレート生成器が、クラスターメタデータへのクラスターの画像のマッピングのタスクで訓練される、生成することと、

アップサンプリングされたサブピクセル解像度でテンプレート画像内のクラスターの空間分布情報を符号化することであって、テンプレート画像のサブピクセル及びクラスターの画像のピクセルが、同じ画像化領域を表す、符号化することと、

テンプレート画像に基づいて、第 2 の画像データのピクセルの強度値を修正し、クラスターの空間分布情報を考慮する強度分布を有する第 2 の画像データの強度修正バージョンを生成することと、

ニューラルネットワークベースのベースコーラーを介して第 2 の画像データの強度修正バージョンを処理し、配列決定動作の 1 つ又はそれ以上の配列決定サイクルで、クラスターのうちの 1 つ又はそれ以上に対するベースコールを生成することであって、ニューラルネットワークベースのベースコーラーが、ベースコールへのクラスターの画像のマッピングのタスクで訓練される、生成することと、を含む、コンピュータ実装の方法。

2. 第 2 の画像データのピクセルの強度値を修正する代わりに、第 2 の画像データをテンプレート画像に補うことと、

ニューラルネットワークベースのベースコーラーを介して、テンプレート画像に補われる第 2 の画像データを処理し、配列決定動作の 1 つ又はそれ以上の配列決定サイクルで、クラスターのうちの 1 つ又はそれ以上に対するベースコールを生成することと、を更に含む、請求項 1 に記載のコンピュータ実装の方法。

3. テンプレート画像内の各サブピクセルが、背景サブピクセル、クラスター中心サブピクセル、又はクラスター内部サブピクセルのいずれかとして識別される、項目 1 に記載のコンピュータ実装の方法。

10

20

30

40

50

4. 第2の画像データのピクセルの強度値を修正することが、

第2の画像データの画像内のピクセルに対応するテンプレート画像内のどれくらい多くのサブピクセルが、クラスターのうちの1つ又はそれ以上の一部を含むかに基づいて、第2の画像データ内の1つ又はそれ以上のピクセルに対する領域重み付け係数を計算することと、

領域重み付け係数に基づいて、ピクセルの強度を修正することと、を含む、項目1〜3のいずれか一項に記載のコンピュータ実装の方法。

5. 第2の画像データのピクセルの強度値を修正することが、

クラスター及びそれらの周囲の背景の画像をアップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、テンプレート画像内の背景サブピクセルに対応するアップサンプリングされた画像内のサブピクセルに背景強度を割り当てることと、テンプレート画像内のクラスター中心サブピクセル及びクラスター内部サブピクセルに対応するアップサンプリングされた画像内のサブピクセルにクラスター強度を割り当てることと、を含む、項目1〜4のいずれか一項に記載のコンピュータ実装の方法。

6. 背景強度が、ゼロ値を有する、項目5に記載のコンピュータ実装の方法。

7. クラスター強度が、光学ピクセル解像度でピクセルの強度を補間することによって決定される、項目1〜6のいずれか一項に記載のコンピュータ実装の方法。

8. 第2の画像データのピクセルの強度値を修正することが、

クラスター及びそれらの周囲の背景の画像をアップサンプリングされたサブピクセル解像度にアップサンプリングして、アップサンプリングされた画像を生成することと、テンプレート画像内のクラスター中心サブピクセル及びクラスター内部サブピクセルに対応するアップサンプリングされた画像内のピクセルの構成サブピクセルのみの間で、光学ピクセルドメイン内のピクセルの全体の強度を分散させることと、を含む、項目1〜7のいずれか一項に記載のコンピュータ実装の方法。

9. 第1のニューラルネットワークを使用して、クラスターに関するテンプレート画像を決定することであって、テンプレート画像が、クラスターの空間分布、クラスター形状、クラスターの中心、及びクラスター境界からなる群から選択される特性のうちの少なくとも1つを識別する、決定することと、

第2のニューラルネットワークを使用して、テンプレート画像に基づいてクラスターをベースコールすることと、を含む、コンピュータ実装の方法。

10. テンプレート画像が、クラスターの空間分布、クラスター形状、クラスターの中心、及びクラスター境界からなる群から選択される特性のうちの少なくとも1つを識別するための修正強度値と、

第2のニューラルネットワークを介して修正強度値を処理して、クラスターをベースコールすることと、を含む、項目10に記載のコンピュータ実装の方法。

11. テンプレート画像が、テンプレート画像を含む、項目9又は10に記載のコンピュータ実装の方法。

12. 少なくとも1つの特定のクラスターに対してアップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価して、少なくとも1つの特定のクラスターの一部を含むピクセル、及び少なくとも1つの特定のクラスターの一部も含む、ピクセルに隣接するピクセルを識別することと、

識別されたピクセルの各々でのどれくらい多くのサブピクセルが、少なくとも1つの特定のクラスターの一部を含むかに基づいて、各ピクセルに対する領域重み付け係数を計算することと、

それぞれのピクセルに対する領域重み付け係数に基づく処理のために、識別されたピクセル及び隣接するピクセルのピクセル強度値を修正することと、を更に含む、項目11に記載のコンピュータ実装の方法。

13. テンプレート画像を評価することが、

第1のニューラルネットワークを介して、複数の配列決定サイクルのうちの1つ又はそ

10

20

30

40

50

れ以上の初期配列決定サイクルでそれぞれ生成される1つ又はそれ以上の初期画像セットを処理して、テンプレート画像を生成して、アップサンプリングされたサブピクセル解像度でクラスターの中心、形状、及び境界を識別することであって、各画像セットが、1つ又はそれ以上の画像を含み、画像の各々が、光学ピクセル解像度で捕捉される1つ又はそれ以上の撮像チャンネルのうちのそれぞれの1つでのクラスター及びそれらの周囲の背景の強度放射を示す、識別することを更に含む、項目12に記載のコンピュータ実装の方法。

14. テンプレート画像を評価することが、

少なくとも1つの特定のクラスターのクラスター形状及び境界を評価して、少なくとも1つの特定のクラスターの一部を含む少なくとも1つのピクセル、及び少なくとも1つの特定のクラスターの一部も含む、ピクセルに隣接するピクセルを識別することを更に含む、方法が、テンプレート画像内に領域重み付け係数を記憶することと、

10

修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、

第2のニューラルネットワークを介して画像の修正バージョンを処理して、修正バージョンの代替表現を生成することと、

代替表現を使用して、少なくとも1つの特定のクラスターをベースコールすることと、を更に含む、項目12又は13に記載のコンピュータ実装の方法。

15. ベースコールが、

複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、

複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることと、

20

画像の各々でのピクセルに対して、それぞれのピクセルに対するテンプレート画像内の領域重み付け係数に基づいて、ピクセル強度値を修正することと、

修正ピクセル強度値を有するピクセルを有する画像の各々の修正バージョンを生成することと、

少なくとも1つの特定のクラスターに対して、各画像パッチが、

ピクセルのアレイを有し、かつ

30

その中心ピクセルで、テンプレート画像で識別される特定のクラスターの中心を含むように、各修正バージョンから画像パッチを抽出することと、

第2のニューラルネットワークの畳み込みニューラルネットワークを介して、画像の修正バージョンから抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、

出力層を介して畳み込み表現を処理して、中心ピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで少なくとも1つの特定のクラスターに組み込まれる塩基の尤度を生成することと、

尤度に基づいて、塩基をA、C、T、又はGと分類することと、を更に含む、項目14に記載のコンピュータ実装の方法。

40

16. ピクセル強度値を修正する前に、サイクル固有及び撮像チャンネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることを更に含む、項目14又は15に記載のコンピュータ実装の方法。

17. アップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価して、任意のクラスターの一部を含むサブピクセルを識別することと、

いかなるクラスターにも寄与しないとしてテンプレート画像で識別されるサブピクセルに背景強度を割り当てることと、を更に含む、項目9に記載のコンピュータ実装の方法。

18. アップサンプリングされたサブピクセルドメイン内のテンプレート画像を評価することが、

少なくとも1つのピクセル内のどれくらい多くのサブピクセルが、任意のクラスターの

50

一部を含むかを計算することと、少なくとも1つのピクセル内のサブピクセルに対するサブピクセルごとの領域重み付け係数を計算することと、を更に含む、項目17に記載のコンピュータ実装の方法。

19. 方法が、

第1のニューラルネットワークを介して、複数の配列決定サイクルのうちの1つ又はそれ以上の初期配列決定サイクルでそれぞれ生成される1つ又はそれ以上の初期画像セットを処理して、アップサンプリングされたサブピクセル解像度でテンプレート画像を生成することであって、各画像セットが、1つ又はそれ以上の画像を含み、画像の各々が、光学ピクセル解像度で捕捉される1つ又はそれ以上の撮像チャネルのうちのそれぞれの1つでのクラスター及びそれらの周囲の背景の強度放射を示し、テンプレート画像が、クラスター中心、背景、及びクラスター内部を含むクラスにサブピクセルを分類する、生成することと、

10

光学ピクセル解像度で捕捉される画像の各々をサブピクセルドメインにアップサンプリングすることと、いかなるクラスターにも寄与しないとしてテンプレート画像で識別される画像の各々のサブピクセルに背景強度を割り当てることと、

第2のニューラルネットワークを介してアップサンプリングされた画像を処理して、アップサンプリングされた画像の代替表現を生成することと、

代替表現を使用して、複数のクラスターをベースコールすることと、を含む、項目17又は18に記載のコンピュータ実装の方法。

20

20. 画像の各々をアップサンプリングすることが、

サブピクセルごとの領域重み付け係数を適用することによって、任意のクラスターに寄与するとしてテンプレート画像で識別される特定のピクセルの第1のサブピクセル間の特定のピクセルの強度を分散させることと、いかなるクラスターにも寄与しないとしてテンプレートで識別される特定のピクセルの第2のサブピクセルに背景強度を割り当てることと、を更に含む、項目19に記載のコンピュータ実装の方法。

21. アップサンプリングの前に、方法が、

複数の配列決定サイクルのうちの現在の1つで生成される現在の画像セット、

複数の配列決定サイクルのうちの現在の1つに先行する複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び

30

複数の配列決定サイクルのうちの現在の1つに続く複数の配列決定サイクルのうちの1つ又はそれ以上でそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で光学ピクセル解像度で1つ又はそれ以上の画像にアクセスすることを含み、アップサンプリングした後に、方法が、

各画像パッチが、サブピクセルのアレイを有するように、各アップサンプリングされた画像から画像パッチを抽出することと、

第2のニューラルネットワークの畳み込みニューラルネットワークを介して、アップサンプリングされた画像から抽出される画像パッチを畳み込んで、画像パッチの畳み込み表現を生成することと、

出力層を介して畳み込み表現を処理して、アレイ内の各サブピクセルに対して、A、C、T、及びGである、複数の配列決定サイクルのうちの現在の1つで組み込まれる塩基の尤度を生成することと、

40

尤度に基づいて、塩基をA、C、T、又はGと分類することと、

対応するクラスターの中心を含むそれぞれのサブピクセルに割り当てられる塩基分類に基づいて、複数のクラスターのうちの各々1つをベースコールすることと、を含む、項目20に記載のコンピュータ実装の方法。

22. アップサンプリングの前に、サイクル固有及び撮像チャネル固有の変換を使用して、光学ピクセル解像度で捕捉される画像の各々をテンプレート画像と位置合わせすることとを更に含む、項目20又は21に記載のコンピュータ実装の方法。

23. バイオセンサーシステムに結合された受け部であって、バイオセンサーシステムが、光検出器のアレイを備えるように構成されており、バイオセンサーシステムが、バイ

50

オセンサーを備え、バイオセンサーが、クラスターを含むように構成された反応部位を含む、受け部と、

励起光をバイオセンサーに向け、かつ反応部位内のクラスターを照明するように構成された、照明システムであって、クラスターのうちの少なくともいくつかは、照明されるときに発光信号を提供する、照明システムと、

受け部に結合されており、かつ分析モジュールを備える、システムコントローラであって、分析モジュールが、

複数の配列決定サイクルの各々で、光検出器から、光検出器によって検出される発光信号から導出される画像データを取得し、

ニューラルネットワークを介して、複数の配列決定サイクルの各々に対して画像データを処理し、かつ複数の配列決定サイクルの各々で、クラスターのうちの少なくともいくつかに対してベースコールを生成するように構成されている、システムコントローラと、を備える、配列決定システム。

10

項目セット 3

1. ニューラルネットワークを介して入力データを処理し、入力データの代替表現を生成することであって、入力データが、配列決定動作の1つ又はそれ以上の配列決定サイクルの各々に対するサイクルごとのデータを含み、サイクルごとのデータが、それぞれの配列決定サイクルで1つ又はそれ以上の検体を示す、生成することと、

出力層を介して代替表現を処理することと、出力を生成することと、

出力に基づいて、配列決定サイクルのうちの1つ又はそれ以上で、検体のうちの1つ又はそれ以上をベースコールすることと、を含む、コンピュータ実装の方法。

20

2. サイクルごとのデータが、それぞれの配列決定サイクルで周囲の背景を示す、項目1に記載のニューラルネットワーク実装の方法。

3. 入力データが、画像データであり、サイクルごとのデータが、それぞれの配列決定サイクルで補足される、1つ又はそれ以上の検体及び周囲の背景を示す強度放射を含む、項目1又は2に記載のニューラルネットワーク実装の方法。

4. サイクルごとのデータのピクセルと、検体のうちの1つ又はそれ以上を示す強度放射を示すピクセルとの間の距離を識別する補足距離情報に、サイクルごとのデータを伴わせることを更に含む、項目3に記載のコンピュータ実装の方法。

5. サイクルごとのデータのピクセルにスケーリング値を割り当てる補足スケーリング情報に、サイクルごとのデータを伴わせることを更に含む、項目3に記載のコンピュータ実装の方法。

30

6. サイクルごとのデータが、それぞれの配列決定サイクルで検出される電圧変化を示す、項目1に記載のニューラルネットワーク実装の方法。

7. サイクルごとのデータが、それぞれの配列決定サイクルで測定される電流信号を示す、項目1に記載のニューラルネットワーク実装の方法。

8. 複数の配列決定サイクルを含む配列決定動作中に合成される検体をベースコールする、ニューラルネットワーク実装の方法であって、方法が、

畳み込みニューラルネットワークを介して入力データを畳み込んで、入力データの畳み込み表現を生成することであって、入力データが、配列決定動作の現在の配列決定サイクルで生成される現在の画像セット、現在の配列決定サイクルに先行する配列決定動作の1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の先行する画像セット、及び現在の配列決定サイクルに続く配列決定動作の1つ又はそれ以上の配列決定サイクルでそれぞれ生成される1つ又はそれ以上の後続の画像セットの各々で1つ又はそれ以上の画像から抽出される画像パッチを含み、画像パッチの各々が、ベースコールされているターゲット検体の強度放射を示し、

40

入力データが、画像パッチの中心ピクセルからの画像パッチのピクセルのそれぞれの距離を示す距離情報を更に含む、生成することと、

出力層を介して畳み込み表現を処理して、出力を生成することと、

出力に基づいて、現在の配列決定サイクルでターゲット検体をベースコールすることと

50

、を含む、ニューラルネットワーク実装の方法。

9. それぞれの検体を表す画像領域の中心の畳み込みニューラルネットワーク位置座標への入力として提供することを更に含み、

入力が、畳み込みニューラルネットワークの第1の層に提供され、入力が、畳み込みニューラルネットワークの1つ又はそれ以上の中間層に提供され、

入力が、畳み込みニューラルネットワークの最終層に提供される、項目8に記載のニューラルネットワーク実装の方法。

10. 画像パッチのピクセルに対応するスケーリング値を有する強度スケーリングチャネルを畳み込みニューラルネットワークへの入力として提供することを更に含み、

スケーリング値が、特定のターゲット検体を各々含む画像パッチの中心ピクセルの平均強度に基づく、項目8又は9に記載のニューラルネットワーク実装の方法。

11. 強度スケーリングチャネルが、画像パッチの全てのピクセルに対して同じスケーリング値をピクセルごとに含む、項目8～10のいずれか一項に記載のニューラルネットワーク実装の方法。

12. 各画像パッチが、それぞれのピクセルと、複数の検体のうちの最も近い1つとの間の距離を示すピクセル距離データを更に含み、複数の検体のうちの最も近い1つが、ピクセルと検体の各々との間の中心間距離に基づいて選択される、項目8に記載のニューラルネットワーク実装の方法。

13. 各画像パッチが、各検体ピクセルを検体のうちの1つのみに分類することに基づいて選択される、複数の検体のうちの割り当てられた1つからの各検体ピクセルの距離を識別する検体距離データを更に含む、項目8に記載のニューラルネットワーク実装の方法。

14. 畳み込みニューラルネットワークを介して入力データを畳み込んで、入力データの畳み込み表現を生成することが、

畳み込みニューラルネットワークの第1の畳み込みサブネットワークを介して、サイクルごとの各画像パッチセットを別々に処理して、各配列決定サイクルに対する中間畳み込み表現を生成し、配列決定サイクル間ではなく配列決定サイクル内でのみ、強度及び距離情報を組み合わせ、かつ得られた畳み込み表現を組み合わせる畳み込みを適用することを含むことと、

畳み込みニューラルネットワークの第2の畳み込みサブネットワークを介して、一連の連続する配列決定サイクルに対する中間畳み込み表現をグループごとに処理して、一連の最終畳み込み表現を生成し、配列決定サイクル間で、中間畳み込み表現を組み合わせ、かつ得られた畳み込み表現を組み合わせる畳み込みを適用することを含むことと、を含む、

出力層を介して畳み込み表現を処理して、出力を生成することが、出力層を介して最終畳み込み表現を処理することを含む、項目8～13のいずれか一項に記載のニューラルネットワーク実装の方法。

15. 各画像パッチのピクセルを再構成して、中心ピクセル内のターゲット検体の中心を中心にして、再構成された画像パッチを生成することを更に含み、

畳み込みニューラルネットワークを介して入力データを畳み込んで、入力データの畳み込み表現を生成することが、畳み込みニューラルネットワークを介して再構成された画像パッチを畳み込んで、畳み込み表現を生成することを含む、項目8～14のいずれか一項に記載のニューラルネットワーク実装の方法。

16. 再構成が、再構成を補償するための各画像パッチのピクセルの強度補間を更に含む、項目15に記載のニューラルネットワーク実装の方法。

17. 畳み込みニューラルネットワークの畳み込み層のカスケードを介して、サイクルごとの入力データの配列内のサイクルごとの各入力データを別々に処理することであって、サイクルごとの入力データの配列が、配列決定動作の一連の配列決定サイクルに対して生成され、

サイクルごとの各入力データが、それぞれの配列決定サイクルで捕捉される、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す、画像チャネルを含む、処理することと、

10

20

30

40

50

各配列決定サイクルに対して、別個の処理に基づいて、畳み込み層の各々で畳み込み表現を生成し、それによって、畳み込み表現の配列を生成することと、そのサイクルごとの入力データを、畳み込み表現のその対応する配列と混合し、混合表現を生成することと、その混合表現を平坦化することと、平坦化された混合表現を生成することと、

連続する配列決定サイクルの平坦化された混合表現をスタックとして配置することと、スライディングウィンドウベースでスタック内の平坦化された混合表現のサブセットにおいて畳み込む反復ニューラルネットワークを介して前方及び後方の方向にスタックを処理することであって、各スライディングウィンドウがそれぞれの配列決定サイクルに対応し、

(i) スタックにおける現在のスライディングウィンドウでの平坦化された混合表現のサブセット、及び (i i) 以前の隠れ状態表現に基づいて、各配列決定サイクルに対して各時間ステップで現在の隠れ状態表現を連続的に生成する、処理することと、

前方及び後方の方向でのスタックの処理の結果に基づいて、配列決定サイクルの各々で検体の各々をベースコールすることと、を含む、ベースコールのニューラルネットワーク実装の方法。

18. 時間ステップごとに所与の配列決定サイクルの前方及び後方の現在の隠れ状態表現を組み合わせ、組み合わせられた隠れ状態表現を生成することであって、組み合わせることが、連結若しくは積算又は平均を含む、生成すること、

1つ又はそれ以上の完全に接続されたネットワークを介して組み合わせられた隠れ状態表現を処理すること、及び高密度表現を生成すること、

ソフトマックス層を介して、高密度表現を処理して、A、C、T、及びGである、所与の配列決定サイクルで検体の各々に組み込まれる塩基の尤度を生成すること、並びに

尤度に基づいて、塩基をA、C、T、又はGと分類することによって、所与の配列決定サイクルで検体の各々をベースコールすることを更に含む、項目17に記載のニューラルネットワーク実装の方法。

19. 反復モジュール及び畳み込みモジュールを有するハイブリッドニューラルネットワークであって、反復モジュールが、畳み込みモジュールからの入力を使用する、ハイブリッドニューラルネットワークと、

1つ又はそれ以上の畳み込み層を介して配列決定動作の一連の配列決定サイクルに対する画像データを処理し、画像データの1つ又はそれ以上の畳み込み表現を生成する、畳み込みモジュールであって、画像データが、1つ又はそれ以上の検体及びそれらの周囲の背景の強度放射を示す、畳み込みモジュールと、

畳み込み表現及び以前の隠れ状態表現を畳み込むことに基づいて、現在の隠れ状態表現を生成する、反復モジュールと、

現在の隠れ状態表現に基づいて、検体のうちの少なくとも1つ、及び配列決定サイクルのうちの少なくとも1つに対してベースコールを生成する、出力モジュールと、を備える、ベースコールのためのニューラルネットワークベースのシステム。

20. ニューラルネットワークを介して入力データを処理して、入力データの代替表現を生成することであって、

入力データが、(i) 配列決定動作の1つ又はそれ以上の配列決定サイクルの各々に対するサイクルごとのデータと、(i i) 補足距離情報と、を含み、サイクルごとのデータが、1つ又はそれ以上のクラスター及び配列決定サイクルのうちのそれぞれの1つで捕捉される周囲の背景を示す強度放射を示すピクセルを含み、サイクルごとのデータが、サイクルごとのデータのピクセル間の距離を識別する補足距離情報を伴い、

ニューラルネットワークによるサイクルごとのデータのピクセルの処理中に、補足距離情報が、サイクルごとのデータのピクセルのうちのどれがクラスターの中心を含むか、及びサイクルごとのデータのピクセルのうちのどれがクラスターの中心からより離れているかをニューラルネットワークに伝える加法バイアスを供給する、生成することと、

出力層を介して代替表現を処理することと、出力を生成することと、

出力に基づいて、配列決定サイクルのうちの1つ又はそれ以上で、クラスターのうちの

10

20

30

40

50

1つ又はそれ以上をベースコールすることと、を含む、クラスターをベースコールする、コンピュータ実装の方法。

21. 加法バイアスが、ベースコールの精度を改善する、項目20に記載のコンピュータ実装の方法。

22. ニューラルネットワークが、補足距離情報を使用して、中心クラスターピクセル、それらの隣接するピクセル、並びに周囲クラスターピクセル、背景ピクセル、及びそれらから導出される代替表現よりも多くのそれらから導出される代替表現に対処することによって、配列決定信号をその適切なソースクラスターに割り当てる、項目21に記載のコンピュータ実装の方法。

23. ニューラルネットワークを介して入力データを処理して、入力データの代替表現を生成することであって、入力データが、

(i) 配列決定サイクルのうちのそれぞれの1つで1つ又はそれ以上のクラスターを示す強度放射を示すピクセルを含む配列決定動作の1つ又はそれ以上の配列決定サイクルの各々に対するサイクルごとのデータと、

(ii) サイクルごとのデータのピクセル間の距離を識別する補足距離情報と、を含み、

ニューラルネットワークによるサイクルごとのデータのピクセルの処理中に、補足距離情報が、サイクルごとのデータに伴い、サイクルごとのデータのピクセルのうちのどれがクラスターの中心を含むか、及びサイクルごとのデータのピクセルのうちのどれがクラスターの中心からより離れているかをニューラルネットワークに伝える、生成することと、

出力層を介して代替表現を処理することと、出力を生成することと、

出力に基づいて、配列決定サイクルのうちの1つ又はそれ以上で、クラスターのうちの1つ又はそれ以上をベースコールすることと、を含む、クラスターをベースコールする、コンピュータ実装の方法。

24. 補足距離情報が、ベースコールの精度を改善する、項目1に記載のコンピュータ実装の方法。

25. ニューラルネットワークが、補足距離情報を使用して、中心クラスターピクセル、それらの隣接するピクセル、並びに周囲クラスターピクセル、背景ピクセル、及びそれらから導出される代替表現よりも多くのそれらから導出される代替表現に対処することによって、配列決定信号をその適切なソースクラスターに割り当てる、請求項24に記載のコンピュータ実装の方法。

項目セット4

1. ニューラルネットワークベースのベースコーラーを介して1つ又はそれ以上の検体に対して入力データを処理し、入力データの代替表現を生成することと、

出力層を介して代替表現を処理して、出力を生成することであって、出力が、A、C、T、及びGである、検体のうちの特定の1つに組み込まれる塩基の尤度を識別する、生成することと、

出力に基づいて、検体のうちの1つ又はそれ以上に対して塩基をコールすることと、

出力によって識別される尤度に基づいて、コールされた塩基の品質スコアを決定することと、を含む、コンピュータ実装の方法。

2. 尤度に基づいて、コールされた塩基の品質スコアを決定することが、

訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される、ベースコールの分類スコアを量子化することと、

量子化分類スコアのセットを選択することと、

セット内の各量子化分類スコアに対して、その予測されたベースコールを対応するグラウンドトゥールスペースコールと比較することによって、ベースコール誤差率を決定することと、

量子化分類スコアとそれらのベースコール誤差率との間の適合を決定することと、

適合に基づいて、品質スコアを量子化分類スコアに相関させることと、を含む、項目1に記載のコンピュータ実装の方法。

3. 量子化分類スコアのセットが、訓練中の訓練データの処理に応答して、ニューラル

ネットワークベースのベースコーラーによって生成される予測されたベースコールの分類スコアのサブセットを含み、

分類スコアが、実数である、項目 1 又は 2 に記載のコンピュータ実装の方法。

4. 量子化分類スコアのセットが、訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される予測されたベースコールの全ての分類スコアを含み、

分類スコアが、実数である、項目 1 ～ 3 のいずれか一項に記載のコンピュータ実装の方法。

5. 分類スコアが、1 になる指数関数的に正規化されたソフトマックススコアであり、ニューラルネットワークベースのベースコーラーのソフトマックス出力層によって生成される、項目 1 ～ 4 のいずれか一項に記載のコンピュータ実装の方法。

6. 量子化分類スコアのセットが、

【数 4 1】

$$0.9 \sum_{i=1}^n 0.1^{(i-1)}$$

と定義される選択式に基づいて選択され、ソフトマックススコアに適用される、項目 1 ～ 5 のいずれか一項に記載のコンピュータ実装の方法。

7. 量子化分類スコアのセットが、

【数 4 2】

$$\forall_{i=1}^{n=10} 0.1^i$$

と定義される選択式に基づいて選択され、ソフトマックススコアに適用される、項目 1 ～ 6 のいずれか一項に記載のコンピュータ実装の方法。

8. 相関に基づいて、推測中にニューラルネットワークベースのベースコーラーによってコールされる塩基に品質スコアを割り当てることを更に含む、項目 1 ～ 7 のいずれか一項に記載のコンピュータ実装の方法。

9. 推測中にニューラルネットワークベースのベースコーラーによってコールされる塩基に品質スコア対応スキームを適用することに基づいて、品質スコアを割り当てることを更に含む、

スキームが、推測中に、推測データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される、分類スコアの範囲をセット内の対応する量子化分類スコアにマッピングする、項目 8 に記載のコンピュータ実装の方法。

10. 推測中に、品質スコアが現在のベースコールサイクルに対する設定された閾値を下回る検体をベースコールすることを停止させることを更に含む、項目 8 又は 9 に記載のコンピュータ実装の方法。

11. 推測中に、平均品質スコアが連続するベースコールサイクルの後に設定された閾値を下回る検体をベースコールすることを停止させることを更に含む、項目 8 ～ 10 のいずれか一項に記載のコンピュータ実装の方法。

12. 予測されたベースコールを対応するグラウンドトゥールベースコールと比較するために使用されるサンプルサイズが、各量子化分類スコアに固有である、項目 8 ～ 11

のいずれか一項に記載のコンピュータ実装の方法。

13. 適合が、回帰モデルを使用して決定される、項目8～12のいずれか一項に記載のコンピュータ実装の方法。

14. 各量子化分類スコアに対して、その予測されたベースコールを対応するグラウンドトゥールスペースコールと比較することによって、ベースコール精度率を決定することと、

量子化分類スコアとそれらのベースコール精度率との間の適合を決定することと、を更に含む、項目8～13のいずれか一項に記載のコンピュータ実装の方法。

15. 対応するグラウンドトゥールスペースコールが、多数の配列決定器具、配列決定化学、及び配列決定プロトコルで配列決定される、十分に特徴付けられたヒト及び非ヒトサンプルから導出される、項目8～14のいずれか一項に記載のコンピュータ実装の方法。

16. 並行して動作し、かつメモリに結合された多数のプロセッサと、

ニューラルネットワークのベースコール品質予測を、既知の正しいベースコールを識別するベースコール品質グラウンドトゥールスペースと漸進的に一致させる逆伝搬ベースの勾配更新技術を使用して、配列決定画像からのデータを含む訓練例で訓練され、かつベースコール品質グラウンドトゥールスペースでラベル付けされる、多数のプロセッサ上で動作するニューラルネットワークと、

多数のプロセッサのうちの少なくとも1つで動作し、かつ1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質を決定するために、1つ又はそれ以上の配列決定サイクルで捕捉される配列決定画像からのデータをニューラルネットワークに供給する、ニューラルネットワークの入力モジュールと、

多数のプロセッサのうちの少なくとも1つで動作し、かつニューラルネットワークによる分析を、1つ又はそれ以上の検体に対してコールされる1つ又はそれ以上の塩基の品質を識別する出力に変換する、ニューラルネットワークの出力モジュールと、を備える、ニューラルネットワークベースの品質スコアラー。

17. ニューラルネットワークが、畳み込みニューラルネットワークである、項目16に記載のニューラルネットワークベースの品質スコアラー。

18. 出力モジュールが、高品質、中品質、及び低品質である品質に対する尤度を生成するソフトマックス分類層を更に備え、

尤度に基づいて、品質を高品質、中品質、又は低品質と分類することを更に含む、項目16に記載のニューラルネットワークベースの品質スコアラー。

19. ソフトマックス分類層が、複数の品質スコアを割り当てられている品質に対する尤度を生成し、

尤度に基づいて、複数の品質スコアのうちの1つから品質に品質スコアを割り当てることを更に含む、項目16に記載のニューラルネットワークベースの品質スコアラー。

20. 品質スコアが、ベースコール誤差確率に対数的に基づき、

複数の品質スコアが、Q6、Q10、Q15、Q20、Q22、Q27、Q30、Q33、Q37、Q40、及びQ50を含む、項目16～19のいずれか一項に記載のニューラルネットワークベースの品質スコアラー。

21. 出力モジュールが、品質を識別する連続値を生成する回帰層を更に含む、項目16～20のいずれか一項に記載のニューラルネットワークベースの品質スコアラー。

22. コールされる塩基の品質予測値で配列決定画像からデータを補い、

配列決定画像からのデータと共に、品質予測値を畳み込みニューラルネットワークに供給する、補足入力モジュールを更に備える、項目16～21のいずれか一項に記載のニューラルネットワークベースの品質スコアラー。

23. 品質予測値が、オンライン重複、純度、フェイジング、start 5、6量体スコア、モチーフ蓄積、endiness、近似ホモポリマー、強度減衰、最終チャスティティ、背景を有する信号重複(SOWB)、及び／又はシフトされた純度G調整を含む、項目22に記載のニューラルネットワークベースの品質スコアラー。

24. 品質予測値が、ピーク高さ、ピーク幅、ピーク場所、相対的なピーク場所、ピー

10

20

30

40

50

ク高さ割り当て、ピーク間隔割り当て、及び／又はピーク対応を含む、項目 22 に記載のニューラルネットワークベースの品質スコアラー。

25. ベースコールの品質スコアを決定するコンピュータ実装の方法であって、

ニューラルネットワークベースのベースコーラーを介して1つ又はそれ以上のクラスターに対して入力データを処理し、入力データの代替表現を生成することと、

出力層を介して代替表現を処理して、出力を生成することであって、出力が、A、C、T、及びGである、クラスターのうちの特定の1つに組み込まれる塩基の尤度を識別する、生成することと、

出力に基づいて、クラスターのうちの1つ又はそれ以上に対して塩基をコールすることと、

訓練中の訓練データの処理に応答して、ニューラルネットワークベースのベースコーラーによって生成される、ベースコールの分類スコアを量子化すること、

量子化分類スコアのセットを選択すること、

セット内の各量子化分類スコアに対して、その予測されたベースコールを対応するグラウンドトゥールスペースコールと比較することによって、ベースコール誤差率を決定すること、

量子化分類スコアとそれらのベースコール誤差率との間の適合を決定すること、及び

適合に基づいて、品質スコアを量子化分類スコアに相関させることによって、出力によって識別される尤度に基づいて、コールされた塩基の品質スコアを決定することと、を含む、コンピュータ実装の方法。

26. 適合が、量子化分類スコアと品質スコアとの間の対応を示す、請求項1に記載のコンピュータ実装の方法。

27. ベースコールの品質スコアを決定するコンピュータ実装の方法であって、

ニューラルネットワークベースのベースコーラーを介して1つ又はそれ以上のクラスターに対して入力データを処理し、入力データの代替表現を生成することと、

出力層を介して代替表現を処理して、出力を生成することであって、出力が、A、C、T、及びGである、クラスターのうちの特定の1つに組み込まれる塩基の尤度を識別する、生成することと、

出力に基づいて、クラスターのうちの1つ又はそれ以上に対して塩基をコールすることと、

ニューラルネットワークベースのベースコーラーの訓練に対して較正される量子化スキームに基づいて、出力によって識別される尤度に基づいて、コールされた塩基の品質スコアを決定することであって、量子化スキームが、訓練データの処理に応答して、訓練中にニューラルネットワークベースのベースコーラーによって生成されるコールされた塩基の分類スコアを量子化すること、

量子化分類スコアのセットを選択すること、

セット内の各量子化分類スコアに対して、その予測されたベースコールを対応するグラウンドトゥールスペースコールと比較することによって、ベースコール誤差率を決定すること、

量子化分類スコアとそれらのベースコール誤差率との間の適合を決定すること、及び

適合に基づいて、品質スコアを量子化分類スコアに相関させることを含む、決定することと、を含む、コンピュータ実装の方法。

項目セット5

1. フローセルのタイル上の検体を示す画像領域を決定するコンピュータ実装の方法であって、

配列決定動作中に生成された一連の画像セットにアクセスすることであって、各画像セットが、配列決定動作のそれぞれの配列決定サイクル中に生成され、一連の各画像が、検体及びそれらの周囲の背景を示し、一連の各画像が、複数のサブピクセルを有するアクセスすることと、

サブピクセルの各々を分類するベースコールをベースコールから取得し、それによって

10

20

30

40

50

、配列決定動作の複数の配列決定サイクルにわたって、サブピクセルの各々に対してベースコール配列を生成することと、

実質的に一致するベースコール配列を共有する連続するサブピクセルの複数の不連続領域を決定することと、

決定された不連続領域を識別する検体マップを生成することと、を含む、コンピュータ実装の方法。

2. 連続するサブピクセルの決定された複数の不連続領域に基づいて分類子を訓練することを更に含み、分類子が、ニューラルネットワークベースのベースコーラーによるベースコールのための、入力画像データに表される複数の検体の各々の1つ又はそれ以上の特性を表す、減衰マップ、三元マップ、又はバイナリマップを生成するために、好ましくは、ハイスループット核酸配列決定技術におけるスループットのレベルを増加させるための、画像データを処理するためのニューラルネットワークベースのテンプレート生成器である、項目1に記載のコンピュータ実装の方法。

10

3. 不連続領域のいずれにも属しないサブピクセルを背景として識別することによって、検体マップを生成することと、を含む、項目1又は2に記載のコンピュータ実装の方法。

4. 検体マップが、ベースコール配列が実質的に一致しない2つの連続するサブピクセル間の検体境界部分を識別する、項目1-3のいずれか一項に記載のコンピュータ実装の方法。

5. 連続するサブピクセルの複数の不連続領域を決定することが、

ベースコーラーによって決定された検体の予備中心座標における原点サブピクセルを識別することと、

20

原点サブピクセルから開始し、連続的に連続する非原点サブピクセルを継続することによって、実質的に一致するベースコール配列を幅優先で検索することと、を更に含む、項目1-4のいずれか一項に記載のコンピュータ実装の方法。

6. 検体マップの不連続領域の質量中心を、不連続領域を形成するそれぞれの連続するサブピクセルの座標の平均として計算することによって、検体の超位置中心座標を決定することと、

分類子を訓練するためのグラウンドトゥールースとして使用するために、メモリ内の検体の超位置中心座標を記憶することと、を更に含む、項目1-5のいずれか一項に記載のコンピュータ実装の方法。

30

7. 検体の超位置中心座標における検体マップの不連続領域内の質量サブピクセルの中心を識別することと、

補間を使用して検体マップをアップサンプリングし、分類子を訓練するためのグラウンドトゥールースとして使用するために、メモリにアップサンプリングされた検体マップを記憶することと、

アップサンプリングされた検体マップで、連続するサブピクセルが属する不連続領域内の質量サブピクセルの中心からの連続するサブピクセルの距離に比例する減衰係数に基づいて、不連続領域内の各連続するサブピクセルに値を割り当てることと、を更に含む、項目6に記載のコンピュータ実装の方法。

8. 方法が、更に好ましくは、

40

それらの割り当てられた値に基づいて、不連続領域内の連続するサブピクセル、及び背景として識別されたサブピクセルを表す、アップサンプリングされた検体マップから減衰マップを生成することと、

分類子を訓練するためにグラウンドトゥールースとして使用するために、メモリに減衰マップを記憶することと、を含む、項目7に記載のコンピュータ実装の方法。

9. 方法が、更に好ましくは、

アップサンプリングされた検体マップにおいて、検体ベースで、不連続領域内の連続するサブピクセルを、同じ検体に属する検体内部サブピクセルとして分類することと、検体中心サブピクセルとしての質量サブピクセルの中心と、検体境界部分を境界サブピクセルとして含むサブピクセルと、背景サブピクセルとして背景として識別されたサブピクセル

50

とを分類することと、

分類子を訓練するためにグラウンドトゥールースとして使用するために、メモリに分類を記憶することと、を含む、項目 8 に記載のコンピュータ実装の方法。

10. 検体ベースで、検体内部サブピクセル、検体中心サブピクセル、境界サブピクセル、及び背景サブピクセルの座標を、分類子を訓練するためのグラウンドトゥールースとして使用するために、メモリ内に記憶することと、

検体マップをアップサンプリングするために使用される因子によって座標をダウンスケールすることと、

分類子を訓練するためのグラウンドトゥールースとして使用するために、検体ベースでメモリにダウンスケールされた座標を記憶することと、を含む、項目 1 ～ 9 のいずれか一項に記載のコンピュータ実装の方法。

10

11. アップサンプリングされた検体マップから生成されたバイナリグラウンドトゥールースデータにおいて、色符号化を使用して、検体中心クラスに属するように検体中心サブピクセルをラベル付けし、他の全てのサブピクセルが非中心クラスに属するものとしてラベル付けすることと、

分類子を訓練するためにグラウンドトゥールースとして使用するために、メモリにバイナリグラウンドトゥールースデータを記憶することと、を更に含む、項目 1 ～ 10 のいずれか一項に記載のコンピュータ実装の方法。

12. アップサンプリングされた検体マップから生成された三元グラウンドトゥールースデータにおいて、色符号化を使用して、背景クラスに属するとして背景サブピクセルをラベル付けし、検体中心クラスに属するとして検体中心サブピクセルをラベル付けし、検体内部クラスに属するとして検体内部サブピクセルをラベル付けすることと、

20

分類子を訓練するためのグラウンドトゥールースとして使用するために、メモリに三元グラウンドトゥールースデータを記憶することと、を更に含む、項目 1 ～ 11 のいずれか一項に記載の方法。

13. フローセルの複数のタイルの検体マップを生成することと、

検体マップをメモリに記憶し、それらの形状及びサイズを含む、検体マップに基づいて、タイル内の検体の空間分布を決定することと、

タイル中の検体のアップサンプリングされた検体マップにおいて、検体ベースで、同じ検体、検体中心サブピクセル、境界サブピクセル、及び背景サブピクセルに属する検体内部サブピクセルとして分類することと、

30

分類子を訓練するためにグラウンドトゥールースとして使用するために、メモリに分類を記憶することと、

タイルのわたる検体ベースで、分類子を訓練するためにグラウンドトゥールースとして使用するために、検体内部サブピクセル、検体中心サブピクセル、境界サブピクセル、及び背景サブピクセルの座標をメモリに記憶することと、

検体マップをアップサンプリングするために使用される係数によって座標をダウンスケールすることと、

分類子を訓練するためにグラウンドトゥールースとして使用するために、タイルにわたる検体ベースでダウンスケールされた座標をメモリに記憶することと、を更に含む、項目 1 ～ 12 のいずれか一項に記載のコンピュータ実装の方法。

40

14. ベースコール配列が、ベースコールの所定の部分が順序位置ごとに一致するときに実質的に一致する、項目 1 ～ 13 のいずれか一項に記載のコンピュータ実装の方法。

15. 実質的に一致するベースコール配列を共有する連続するサブピクセルの複数の不連続領域を決定することが、不連続領域のための所定の最小数のサブピクセルに基づく、項目 1 ～ 14 のいずれか一項に記載のコンピュータ実装の方法。

16. フローセルが、検体を占有するウェルのアレイを有する少なくとも 1 つのパターン化表面を有し、更に、

検体の決定された形状及びサイズに基づいて、

ウェルのうちのどれが、少なくとも 1 つの検体によって実質的に占有されているか、

50

ウェルのうちのどれが最小限に占有されているか、及び

ウェルのうちのどれが、複数の検体によって共占有されているかを決定することを更に含む、項目 1 ～ 15 のいずれか一項に記載のコンピュータ実装の方法。

17. フローセルのタイル上の検体に関するメタデータを決定するコンピュータ実装の方法であって、

配列決定動作中に捕捉されたタイルの画像のセットにアクセスすることと、ベースコーラーによって決定された検体の予備中心座標にアクセスすることと、

各画像セットに対して、ベースコーラーから、4つの塩基のうちの1つとしての、予備中心座標を含む原点サブピクセルのベースコール分類を取得することと、

原点サブピクセルのうちのそれぞれの1つに連続的に連続する、連続するサブピクセルの所定の近傍であって、それによって、原点サブピクセルの各々に対して、及び連続するサブピクセルの所定の近傍の各々に対して、ベースコール配列を生成することと、

原点サブピクセルのうちのそれぞれの1つの少なくともいくつかに連続的に連続する、連続するサブピクセルの不連続領域として検体を識別する検体マップを生成することと、

4つの塩基のうちの1つの実質的に一致するベースコール配列を原点サブピクセルのうちのそれぞれの1つの少なくともいくつかと共有することと、

検体マップをメモリに記憶し、検体マップ内の不連続領域に基づいて、検体の形状及びサイズを決定することと、を含む、コンピュータ実装の方法。

18. ニューラルネットワークベースのテンプレート生成及びベースコールのための訓練データを生成するコンピュータ実装の方法であって、

配列決定動作の複数のサイクルにわたって捕捉されたフローセルの多数の画像にアクセスすることであって、フローセルが複数のタイルを有し、多数の画像において、タイルの各々が、複数のサイクルにわたって生成された画像セットの配列を有し、画像セットの配列内の各画像が、特定の1回のサイクルにおける、特定のタイルのうちの特定の1つの検体及びそれらの周囲の背景の強度放射を示す、アクセスすることと、

複数の訓練例を有する訓練セットを構築することであって、各訓練例が、タイルのうちの特定の1つに対応し、タイルのうちの特定の1つの画像セットの配列内の少なくともいくつかの画像セットからの画像データを含む、構築することと、

訓練例の各々について少なくとも1つのグラウンドトゥールズデータ表現を生成することであって、グラウンドトゥールズ表現が、タイルのうちの特定の1つで検体の少なくとも1つの特性を識別し、その強度放射が、画像データによって示され、少なくとも部分的に、項目 1 ～ 17 のいずれか一項に記載の方法を使用して決定される、生成することと、を含む、コンピュータ実装の方法。

19. 検体の少なくとも1つの特性が、タイル上の検体の空間分布、検体形状、検体サイズ、検体境界、及び単一の検体を含む連続する領域の中心からなる群から選択される、項目 18 に記載のコンピュータ実装の方法。

20. 画像データが、タイルのうちの特定の1つの画像セットの配列内の少なくともいくつかの画像セットの各々の画像を含む、項目 18 又は 19 に記載のコンピュータ実装の方法。

21. 画像データが、画像の各々から少なくとも1つの画像パッチを含む、項目 18 ～ 20 のいずれか一項に記載のコンピュータ実装の方法。

22. 画像データが、画像パッチのアップサンプリングされた表現を含む、項目 18 ～ 21 のいずれか一項に記載のコンピュータ実装の方法。

23. 複数の訓練例が、タイルのうちの同じ特定の1つに対応し、タイルのうちの同じ特定の1つの画像セットの配列内の少なくともいくつかの画像セットの各々の各画像から異なる画像パッチを画像データとしてそれぞれ含み、

異なる画像パッチのうちの少なくともいくつかがい互いに重複する、項目 18 ～ 22 のいずれか一項に記載のコンピュータ実装の方法。

24. グラウンドトゥールズデータ表現が、隣接するサブピクセルの不連続領域として検体を識別し、不連続領域のうちのそれぞれの1つの内部の質量サブピクセルの中心とし

10

20

30

40

50

て検体の中心を識別し、不連続領域のうちのいずれにも属しないサブピクセルとしてそれらの周囲の背景を識別する、項目 18～23 のいずれか一項に記載のコンピュータ実装の方法。

25. 訓練セット及び関連するグラウンドトゥールースデータ表現内の訓練例を、ニューラルネットワークベースのテンプレート生成及びベースコールのための訓練データとして記憶することを更に含む、項目 18～24 のいずれか一項に記載のコンピュータ実装の方法。

26. シーケンサによって生成された検体の配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、

ニューラルネットワークを訓練するための訓練データを使用して、検体に関するメタデータを生成することと、を含む、コンピュータ実装の方法。

27. シーケンサによって生成された検体の配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、

ニューラルネットワークを訓練するための訓練データを使用して、検体をベースコールすることと、を含むコンピュータ実装の方法。

28. フローセルのタイル上の検体を示す画像領域を決定するコンピュータ実装の方法であって、

配列決定動作中に生成された一連の画像セットにアクセスすることであって、一連の各画像セットが、配列決定動作のそれぞれの配列決定サイクル中に生成され、一連の各画像が、検体及びそれらの周囲の背景を示し、一連の各画像が、複数のサブピクセルを有するアクセスすることと、

サブピクセルの各々を分類するベースコールをベースコーラーから取得し、それによって、配列決定動作の複数の配列決定サイクルにわたって、サブピクセルの各々についてベースコール配列を生成することと、

実質的に一致するベースコール配列を共有する連続するサブピクセルの複数の不連続領域を決定することと、を含む、コンピュータ実装の方法。

項目セット 6

1. クラスターメタデータ決定タスクのためのニューラルネットワークベースのテンプレート生成器を訓練するために、グラウンドトゥールース訓練データを生成するコンピュータ実装の方法であって、

配列決定動作中に生成された一連の画像セットにアクセスすることであって、一連の各画像セットが、配列決定動作のそれぞれの配列決定サイクル中に生成され、一連の画像が、クラスター及びそれらの周囲の背景を示し、一連の各画像が、ピクセルドメイン内のピクセルを有し、ピクセルの各々が、サブピクセルドメイン内の複数のサブピクセルに分割される、アクセスすることと、

サブピクセルの各々を 4 つの塩基（A、C、T、及び G）のうちの 1 つと分類するベースコールをベースコーラーから取得することであって、それによって、配列決定動作の複数の配列決定サイクルにわたって、サブピクセルの各々についてベースコール配列を生成することと、

実質的に一致するベースコール配列を共有する連続するサブピクセルの不連続領域としてクラスターを識別するクラスターマップを生成することと、

クラスターマップ内の不連続領域に基づいてクラスターメタデータを決定することであって、クラスターメタデータが、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を含む、決定することと、

クラスターメタデータを使用して、クラスターメタデータ決定タスクのためのニューラルネットワークベースのテンプレート生成器を訓練するためにグラウンドトゥールース訓練データを生成することと、を含み、グラウンドトゥールース訓練データが、減衰マップ、三元マップ、又はバイナリマップを含み、ニューラルネットワークベースのテンプレート生成器が、グラウンドトゥールース訓練データに基づいて、出力として減衰マップ、三元マップ、又はバイナリマップを生成するように訓練され、

10

20

30

40

50

推測中のクラスターメタデータ決定タスクの実行時に、クラスターメタデータが、次に、訓練されたニューラルネットワークベースのテンプレート生成器によって出力として生成される減衰マップ、三元マップ、又はバイナリマップから決定される、コンピュータ実装の方法。

2. ハイスループット核酸配列決定技術におけるスループットを増加させるために、ニューラルネットワークベースのベースコーラーによってベースコールするためのニューラルネットワークベースのテンプレート生成器による出力として生成された、減衰マップ、三元マップ、又はバイナリマップから導出されたクラスターメタデータを使用することを更に含む、項目1に記載のコンピュータ実装の方法。

3. 不連続領域のいずれにも属しないサブピクセルを背景として識別することによって、クラスターマップを生成することを更に含む、項目1に記載のコンピュータ実装の方法。

4. クラスターマップが、ベースコール配列が実質的に一致しない2つの連続するサブピクセル間のクラスター境界部分を識別する、項目1に記載のコンピュータ実装の方法。

5. クラスターマップが、ベースコーラーによって決定されるクラスターの予備中心座標における原点サブピクセルを識別すること、及び

原点サブピクセルから開始し、連続的に連続する非原点サブピクセルを継続することによって、実質的に一致するベースコール配列を幅優先で検索することに基づいて生成される、項目1に記載のコンピュータ実装の方法。

6. クラスターマップの不連続領域の質量中心を、不連続領域を形成するそれぞれの連続するサブピクセルの座標の平均として計算することによって、クラスターの超位置中心座標を決定することと、

ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールズ訓練データとして使用するために、メモリ内のクラスターの超位置中心座標を記憶することと、を更に含む、項目1に記載のコンピュータ実装の方法。

7. クラスターの超位置中心座標におけるクラスターマップの非接合領域内の質量サブピクセルの中心を識別することと、

補間を使用してクラスターマップをアップサンプリングし、ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールズ訓練データとして使用するために、メモリ内にアップサンプリングされたクラスターマップを記憶することと、

アップサンプリングされたクラスターマップで、連続するサブピクセルが属する不連続領域内の質量サブピクセルの中心からの連続するサブピクセルの距離に比例する減衰係数に基づいて、不連続領域内の各連続するサブピクセルに値を割り当てることと、を更に含む、項目6に記載のコンピュータ実装の方法。

8. それらの割り当てられた値に基づいて、不連続領域内の連続するサブピクセル、及び背景として識別されるサブピクセルを表す、アップサンプリングされたクラスターマップから減衰マップを生成することと、

ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールズ訓練データとして使用するために、メモリに減衰マップを記憶することと、を更に含む、項目7に記載のコンピュータ実装の方法。

9. アップサンプリングされたクラスターマップにおいて、クラスターごとに、不連続領域内の連続するサブピクセルを、同じクラスターに属するクラスター内部サブピクセルとして分類し、質量サブピクセルの中心をクラスター中心サブピクセルとして分類し、クラスター境界部分を含むサブピクセルを境界サブピクセルとして分類し、背景として識別されたサブピクセルを背景サブピクセルとして分類することと、

ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールズ訓練データとして使用するために、メモリに分類を記憶することと、を更に含む、項目8に記載のコンピュータ実装の方法。

10. クラスターごとに、クラスター内部サブピクセル、クラスター中心サブピクセル、境界サブピクセル、及び背景サブピクセルの座標を、ニューラルネットワークベースの

10

20

30

40

50

テンプレート生成器を訓練するためのグラウンドトゥールース訓練データとして使用するために、メモリ内に記憶することと、

クラスターマップをアップサンプリングするために使用される係数によって座標をダウンスケールすることと、

クラスターごとに、ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールース訓練データとして使用するために、メモリにダウンスケールされた座標を記憶することと、を更に含む、項目 9 に記載のコンピュータ実装の方法。

11. フローセルの複数のタイルのクラスターマップを生成することと、

クラスターマップをメモリに記憶し、クラスター中心、クラスター形状、クラスターサイズ、クラスター背景、及び／又はクラスター境界を含む、クラスターマップに基づいて、タイル内のクラスターのクラスターメタデータを決定することと、

タイル内のクラスターのアップサンプリングされたクラスターマップにおいて、クラスターごとに、サブピクセルを同じクラスター、クラスター中心サブピクセル、境界サブピクセル、及び背景サブピクセルに属するクラスター内部サブピクセルとして分類することと、

ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールース訓練データとして使用するために、メモリに分類を記憶することと、

タイルにわたるクラスターごとに、クラスター内部サブピクセル、クラスター中心サブピクセル、境界サブピクセル、及び背景サブピクセルの座標を、ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールース訓練データとして使用するために、メモリ内に記憶することと、

クラスターマップをアップサンプリングするために使用される係数によって座標をダウンスケールすることと、

タイルにわたるクラスターごとに、ニューラルネットワークベースのテンプレート生成器を訓練するためのグラウンドトゥールース訓練データとして使用するために、メモリ内のダウンスケールされた座標を記憶することと、を更に含む、項目 10 に記載のコンピュータ実装の方法。

12. ベースコール配列が、ベースコールの所定の部分が、順序位置ごとに一致するときに実質的に一致する、項目 11 に記載のコンピュータ実装の方法。

13. クラスターマップが、不連続領域のための所定の最小数のサブピクセルに基づいて生成される、項目 1 に記載のコンピュータ実装の方法。

14. フローセルが、クラスターを占有するウェルのアレイを有する少なくとも 1 つのパターン化表面を有し、更に、

クラスターの決定された形状及びサイズに基づいて、ウェルのうちのどれが、少なくとも 1 つクラスターによって実質的に占有されているか、ウェルのうちのどれが、最小限に占有されているか、及び

ウェルのうちのどれが、複数のクラスターによって共占有されているかを決定することを含む、項目 1 に記載のコンピュータ実装の方法。

15. フローセルのタイル上のクラスターに関するメタデータを決定するコンピュータ実装の方法であって、

配列決定動作中に捕捉されたタイルの画像のセットにアクセスすることと、ベースコーラーによって決定されたクラスターの予備中心座標にアクセスすることと、

各画像セットに対して、ベースコーラーから、4 つの塩基のうちの 1 つとしての、予備中心座標を含む原点サブピクセルのベースコール分類を取得することと、

原点サブピクセルのうちのそれぞれの 1 つに連続的に連続する、連続するサブピクセルの所定の近傍であって、それによって、原点サブピクセルの各々に対して、及び連続するサブピクセルの所定の近傍の各々に対して、ベースコール配列を生成することと、

原点サブピクセルのうちのそれぞれの 1 つの少なくともいくつかに連続的に連続し、かつ、4 つの塩基のうちの 1 つの実質的に一致するベースコール配列を、原点サブピクセルのうちのそれぞれの 1 つの少なくともいくつかと共有する、連続するサブピクセルの不連

10

20

30

40

50

続領域としてクラスターを識別する、クラスターマップを生成することと、

クラスターマップをメモリに記憶し、クラスターマップ内の不連続領域に基づいて、クラスターの形状及びサイズを決定することと、を含む、コンピュータ実装の方法。

16. ニューラルネットワークベースのテンプレート生成及びベースコールのための訓練データを生成するコンピュータ実装の方法であって、

配列決定動作の複数のサイクルにわたって捕捉されたフローセルの多数の画像にアクセスすることであって、フローセルが、複数のタイルを有し、多数の画像において、タイルの各々が、複数のサイクルにわたって生成された画像セットの配列を有し、画像セットの配列内の各画像が、特定の1つのサイクルでのタイルのうちの特定の1つでクラスター及びそれらの周囲の背景の強度放射を示す、アクセスすることと、

10

複数の訓練例を有する訓練セットを構築することであって、各訓練例が、タイルのうちの特定の1つに対応し、タイルのうちの特定の1つの画像セットの配列内の少なくともいくつかの画像セットからの画像データを含む、構築することと、

訓練例の各々について、少なくとも1つのグラウンドトゥルスデータ表現を生成することであって、グラウンドトゥルスデータ表現が、タイルのうちの特定の1つの検体の少なくとも1つの特性を識別し、その強度放射が、画像データによって示される、生成することと、を含む、コンピュータ実装の方法。

17. クラスターの少なくとも1つの特性が、タイル上のクラスターの空間分布、クラスター形状、クラスターサイズ、クラスター境界、及び単一のクラスターを含む連続する領域の中心からなる群から選択される、項目16に記載のコンピュータ実装の方法。

20

18. 画像データが、タイルのうちの特定の1つの画像セットの配列内の少なくともいくつかの画像セットの各々での画像を含む、項目16に記載のコンピュータ実装の方法。

19. 画像データが、画像の各々からの少なくとも1つの画像パッチを含む、項目18に記載のコンピュータ実装の方法。

20. 画像データが、画像パッチのアップサンプリングされた表現を含む、項目19に記載のコンピュータ実装の方法。

21. 複数の訓練例が、タイルのうちの同じ特定の1つに対応し、タイルのうちの同じ特定の1つの画像セットの配列内の少なくともいくつかの画像セットの各々での各画像からの異なる画像パッチを画像データとしてそれぞれ含み、

異なる画像パッチのうちの少なくともいくつかは互いに重複する、項目16に記載のコンピュータ実装の方法。

30

22. グラウンドトゥルスデータ表現が、隣接するサブピクセルの不連続領域としてクラスターを識別し、不連続領域のうちのそれぞれの1つの内部の質量サブピクセルの中心としてクラスターの中心を識別し、不連続領域のうちのいずれにも属しないサブピクセルとしてそれらの周囲の背景を識別する、項目16に記載のコンピュータ実装の方法。

23. 訓練セット及び関連するグラウンドトゥルスデータ表現内の訓練例を、ニューラルネットワークベースのテンプレート生成及びベースコールのための訓練データとしてメモリ内に記憶することを更に含む、項目16に記載のコンピュータ実装の方法。

24. シーケンサによって生成されたクラスターの配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、

40

ニューラルネットワークを訓練するための訓練データを使用して、クラスターに関するメタデータを生成することと、を含む、コンピュータ実装の方法。

25. シーケンサによって生成されたクラスターの配列決定画像にアクセスすることと、配列決定画像から訓練データを生成することと、

ニューラルネットワークを訓練するための訓練データを使用して、クラスターをベースコールすることと、を含む、コンピュータ実装の方法。

26. フローセルのタイル上の検体を示す画像領域を決定するコンピュータ実装の方法であって、

配列決定動作中に生成された一連の画像セットにアクセスすることであって、一連の各画像セットが、配列決定動作のそれぞれの配列決定サイクル中に生成され、一連の各画像

50

が、検体及びそれらの周囲の背景を示し、一連の各画像が、複数のサブピクセルを有するアクセスすることと、

サブピクセルの各々を分類するベースコールをベースコーラーから取得し、それによって、配列決定動作の複数の配列決定サイクルにわたって、サブピクセルの各々についてベースコール配列を生成することと、

実質的に一致するベースコール配列を共有する連続するサブピクセルの複数の不連続領域を決定することと、

決定された不連続領域を識別するクラスターマップを生成することと、を含む、コンピュータ実装の方法。

項目セット 7

1. 1つ又はそれ以上の検体に基づいて生成される画像データから検体データを決定する、ニューラルネットワーク実装の方法であって、

画像の配列から導出される入力画像データを受信することであって、画像の配列内の各画像が、画像化領域を表し、配列決定動作の複数の配列決定サイクルのうちのそれぞれの1つで強度放射の1つ又はそれ以上の検体及び周囲の背景を示す、強度放射を示し、

入力画像データが、画像の配列内の各画像から抽出される画像パッチを含む、受信することと、

ニューラルネットワークを介して入力画像データを処理して、入力画像データの代替表現を生成することと、

出力層を介して代替表現を処理して、画像化領域のそれぞれの部分の特性を示す出力を生成することと、を含む、ニューラルネットワーク実装の方法。

2. 特性が、一部が背景又は検体を表すかどうか、及び

一部が同じ検体を各々表す複数の連続する画像部分の中心を表すかどうかを含む、項目1に記載のニューラルネットワーク実装の方法。

3. 出力が、1つ又はそれ以上の検体を識別し、その強度放射が、隣接するユニットの不連続領域としての入力画像データ、不連続領域のうちのそれぞれの1つの質量中心での中心ユニットとしての1つ又はそれ以上の検体の中心、及び

不連続領域のいずれにも属しない背景ユニットとしての強度放射の周囲の背景によって示される、項目1に記載のニューラルネットワーク実装の方法。

4. 不連続領域のうちのそれぞれの1つでの隣接するユニットが、隣接するユニットが属する不連続領域内の中心ユニットからの隣接するユニットの距離に従って重み付けされる強度値を有する、項目3に記載のニューラルネットワーク実装の方法。

5. 出力が、検体又は背景として各部分を分類するバイナリマップである、項目1～4のいずれか一項に記載のニューラルネットワーク実装の方法。

6. 出力が、検体、背景、又は中心として各部分を分類する三元マップである、項目1～5のいずれか一項に記載のニューラルネットワーク実装の方法。

7. ピークロケータを出力に適用して、出力でピーク強度を見つけることと、

ピーク強度に基づいて、検体の中心の場所座標を決定することと、

入力画像データを準備するために使用されるアップサンプリング係数によって、場所座標をダウンスケールすることと、

検体のベースコールでの使用のために、ダウンスケールされた場所座標をメモリ内に記憶することと、を更に含む、項目1～6のいずれか一項に記載のニューラルネットワーク実装の方法。

8. 不連続領域のうちのそれぞれの1つでの隣接するユニットを、同じ検体に属する検体内部ユニットとして分類することと、

検体のベースコールでの使用のために、検体ごとにメモリ内の検体内部ユニットの分類及びダウンスケールされた場所座標を記憶することと、を更に含む、項目1～7のいずれか一項に記載のニューラルネットワーク実装の方法。

9. ニューラルネットワークを訓練するための訓練データを取得することであって、訓練データが、複数の訓練例及び対応するグラウンドトゥルスデータを含み、各訓練例が

10

20

30

40

50

、画像セットの配列からの画像データを含み、画像セットの配列内の各画像がフローセルのタイルを表し、フローセル上で実行される配列決定動作の複数の配列決定サイクルのうちの特定の1つで、特定の画像チャンネルのために捕捉される、タイル上の検体及びそれらの周囲の背景の強度放射を示し、

各グラウンドトゥールースデータが、訓練例のそれぞれの部分の特性を識別する、取得することと、

勾配降下訓練技術を使用して、ニューラルネットワークを訓練することと、グラウンドトゥールースデータと漸進的に一致する訓練例に対する出力を生成することと、出力とグラウンドトゥールースデータとの間の誤差を最小化する損失関数を反復的に最適化することを含むことと、誤差に基づいてニューラルネットワークのパラメータを更新することと、を更に含む、項目1～8のいずれか一項に記載のニューラルネットワーク実装の方法。

10

10．特性が、ユニットが中心又は非中心であるかどうかを識別することを含む、項目1～9のいずれか一項に記載のニューラルネットワーク実装の方法。

11．最後の反復後の誤差収束の際に、メモリ内のニューラルネットワークの更新されたパラメータを記憶して、更なるニューラルネットワークベースのテンプレート生成及びベースコールに適用することを更に含む、項目9に記載のニューラルネットワーク実装の方法。

12．グラウンドトゥールースデータで、不連続領域のうちのそれぞれの1つでの隣接するユニットが、隣接するユニットが属する不連続領域内の中心ユニットからの隣接するユニットの距離に従って重み付けされる強度値を有する、項目9～11のいずれか一項に記載のニューラルネットワーク実装の方法。

20

13．グラウンドトゥールースデータで、中心ユニットが、不連続領域のうちのそれぞれの1つの内部で最も高い強度値を有する、項目9～11のいずれか一項に記載のニューラルネットワーク実装の方法。

14．損失関数が、平均2乗誤差であり、誤差が、出力及びグラウンドトゥールースデータにおける対応するユニットの正規化された強度値間でユニットベースで最小化される、項目9～13のいずれか一項に記載のニューラルネットワーク実装の方法。

15．訓練データで、複数の訓練例がそれぞれ、同じタイルの画像セットの配列内の各画像からの異なる画像パッチを画像データとして含み、

異なる画像パッチのうちの少なくともいくつかが互いに重複する、項目9～14のいずれか一項に記載のニューラルネットワーク実装の方法。

30

16．グラウンドトゥールースデータで

検体中心として分類されるユニットは全て、同じ第1の所定のクラススコアを割り当てられ、

非中心として分類されるユニットは全て、同じ第2の所定のクラススコアを割り当てられる、項目9～15のいずれか一項に記載のニューラルネットワーク実装の方法。

17．損失関数が、カスタム重み付けバイナリクロスエントロピー損失であり、誤差が、出力及びグラウンドトゥールースデータにおける対応するユニットの予測スコアとクラススコアとの間でユニットベースで最小化される、項目9～16のいずれか一項に記載のニューラルネットワーク実装の方法。

40

18．グラウンドトゥールースデータで、背景として分類されるユニットが全て、同じ第1の所定のクラススコアを割り当てられ、検体中心として分類されるユニットが全て、同じ第2の所定のクラススコアを割り当てられ、

検体内部として分類されるユニットが全て、同じ第3の所定のクラススコアを割り当てられる、項目9～17のいずれか一項に記載のニューラルネットワーク実装の方法。

19．ユニットの出力値を閾値化することと、周囲の背景を示す背景ユニットとして、ユニットの第1のサブセットを分類することと、

ユニットの出力値内にピークを配置することと、検体の中心を含む中心ユニットとして、ユニットの第2のサブセットを分類することと、

ユニットの出力値にセグメント化器を適用することと、背景ユニットによって分離され

50

、かつ中心ユニットで中心にされる、連続するユニットの非重複領域として、検体の形状を決定することと、を更に含み、セグメント化器が、中心ユニットから始まり、各中心ユニットに対して、中心が中心ユニット内に含まれる同じ検体を示す連続的に連続するユニットの群を決定する、項目 1～18 のいずれか一項に記載のニューラルネットワーク実装の方法。

20．非重複領域が、不規則な輪郭を有し、ユニットが、複数のユニットであり、

所与の検体の形状を識別する、連続するユニットの対応する非重複領域に基づいて、所与の検体の検体強度に寄与するユニットを識別すること、

現在の配列決定サイクルで1つ又はそれ以上の画像チャネルに対して生成される1つ又はそれ以上の光学ピクセル解像度画像内に識別されたユニットを配置すること、

画像の各々で、識別されたユニットの強度を補間すること、補間強度を組み合わせること、及び組み合わせられた補間強度を正規化して、画像の各々で所与の検体に対する画像ごとの検体強度を生成すること、並びに

画像の各々に対して画像ごとの検体強度を組み合わせ、現在の配列決定サイクルで所与の検体の検体強度を決定することによって、所与の検体の検体強度を決定することを更に含む、項目 1～19 のいずれか一項に記載のニューラルネットワーク実装の方法。

21．非重複領域が、不規則な輪郭を有し、ユニットが、複数のユニットであり、

所与の検体の形状を識別する、連続するユニットの対応する非重複領域に基づいて、所与の検体の検体強度に寄与するユニットを識別すること、

現在の配列決定サイクルで1つ又はそれ以上の画像チャネルに対して生成される、対応する光学ピクセル解像度画像からアップサンプリングされる1つ又はそれ以上のユニット解像度画像内に識別されたユニットを配置すること、

アップサンプリングされた画像の各々で、識別されたユニットの強度を組み合わせること、及び組み合わせられた強度を正規化して、アップサンプリングされた画像の各々で所与の検体に対する画像ごとの検体強度を生成すること、並びに

アップサンプリングされた画像の各々に対して画像ごとの検体強度を組み合わせ、現在の配列決定サイクルで所与の検体の検体強度を決定することによって、所与の検体の検体強度を決定することを更に含む、項目 1～20 のいずれか一項に記載のニューラルネットワーク実装の方法。

22．正規化が、正規化係数に基づいており、

正規化係数が、識別されたユニットの数である、項目 1～21 のいずれか一項に記載のニューラルネットワーク実装の方法。

23．現在の配列決定サイクルにおける検体強度に基づいて、所与の検体をベースコールすることを更に含む、項目 1～22 のいずれか一項に記載のニューラルネットワーク実装の方法。

24．フローセル上の検体に関するメタデータを決定する、ニューラルネットワーク実装の方法であって、

検体の強度放射を示す画像データにアクセスすることと、

ニューラルネットワークの1つ又はそれ以上の層を介して画像データを処理することと、画像データの代替表現を生成することと、

出力層を介して代替表現を処理することと、検体の形状及びサイズ並びに／又は検体の中心のうちの少なくとも1つを識別する出力を生成することと、を含む、ニューラルネットワーク実装の方法。

25．画像データが、検体の周囲の背景の強度放射を更に示し、

検体間の周囲の背景及び境界を含む、フローセル上の検体の空間分布を識別する出力を更に含む、項目 24 に記載のニューラルネットワーク実装の方法。

26．ニューラルネットワークを介して画像データを処理し、画像データの代替表現を生成することであって、画像データが、検体の強度放射を示す、生成することと、

出力層を介して代替表現を処理することと、検体の空間分布、検体の形状、検体の中心、及び／又は検体間の境界のうちの少なくとも1つを含む、検体に関するメタデータを識

10

20

30

40

50

別する出力を生成することと、を含む、コンピュータ実装の方法。

27. 1つ又はそれ以上のクラスターに基づいて生成される画像データからクラスターメタデータを決定する、ニューラルネットワーク実装の方法であって、

画像の配列から導出される入力画像データを受信することであって、画像の配列内の各画像が、画像化領域を表し、配列決定動作の複数の配列決定サイクルのうちのそれぞれの1つで1つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示し、

入力画像データが、画像の配列内の各画像から抽出される画像パッチを含む、受信することと、

ニューラルネットワークを介して入力画像データを処理して、入力画像データの代替表現を生成することであって、ニューラルネットワークが、クラスター背景、クラスター中心、及びクラスター形状を決定することを含む、クラスターメタデータ決定タスクに対して訓練される、生成することと、

10

出力層を介して代替表現を処理して、画像化領域のそれぞれの部分の特性を示す出力を生成することと、

出力の出力値を閾値化することと、周囲の背景を示す背景部分として画像化領域のそれぞれの部分の第1のサブセットを分類することと、

出力の出力値内にピークを配置することと、クラスターの中心を含む中心部分として画像化領域のそれぞれの部分の第2のサブセットを分類することと、

出力の出力値にセグメント化器を適用することと、背景部分によって分離され、中心部分で中心にされる画像化領域の連続する部分の非重複領域として、クラスターの形状を決定することと、を含む、ニューラルネットワーク実装の方法。

20

28. 1つ又はそれ以上のクラスターに基づいて生成される画像データから、クラスター背景、クラスター中心、及びクラスター形状を含むクラスターメタデータを決定する、ニューラルネットワーク実装の方法であって、

画像の配列から導出される入力画像データを受信することであって、画像の配列内の各画像が、画像化領域を表し、配列決定動作の複数の配列決定サイクルのうちのそれぞれの1つで1つ又はそれ以上のクラスター及びそれらの周囲の背景の強度放射を示し、

入力画像データが、画像の配列内の各画像から抽出される画像パッチを含む、受信することと、

ニューラルネットワークを介して入力画像データを処理して、入力画像データの代替表現を生成することであって、ニューラルネットワークが、クラスター背景、クラスター中心、及びクラスター形状を決定することを含む、クラスターメタデータ決定タスクに対して訓練される、生成することと、

30

出力層を介して代替表現を処理して、画像化領域のそれぞれの部分の特性を示す出力を生成することと、

出力の出力値を閾値化することと、周囲の背景を示す背景部分として画像化領域のそれぞれの部分の第1のサブセットを分類することと、

出力の出力値内にピークを配置することと、クラスターの中心を含む中心部分として画像化領域のそれぞれの部分の第2のサブセットを分類することと、

出力の出力値にセグメント化器を適用することと、画像化領域の連続する部分の非重複領域として、クラスターの形状を決定することと、を含む、ニューラルネットワーク実装の方法。

40

【符号の説明】

【1290】

102 配列決定機器

104 光学系

106 撮像装置

108 配列決定画像

110 サブピクセルアドレス指定器

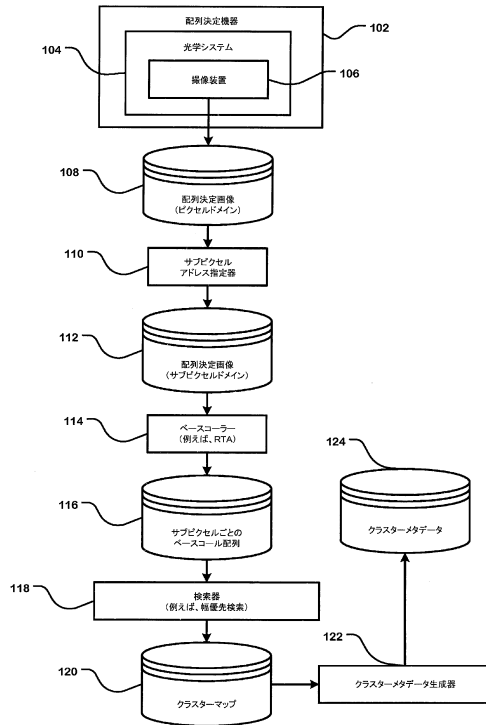
112 配列決定画像

50

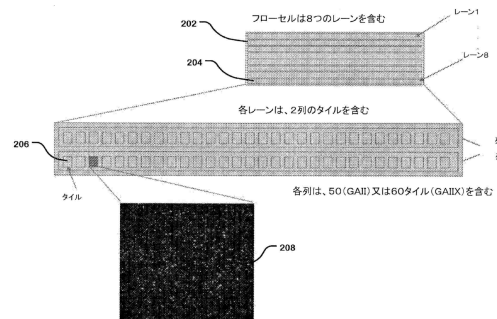
- 1 1 4 ベースコーラー
- 1 1 6 サブピクセルごとのベースコール配列
- 1 1 8 検索器
- 1 2 0 クラスタマップデータストア
- 1 2 2 クラスタメタデータ生成器
- 1 2 4 クラスタメタデータデータストア

【図面】

【図 1】



【図 2】



10

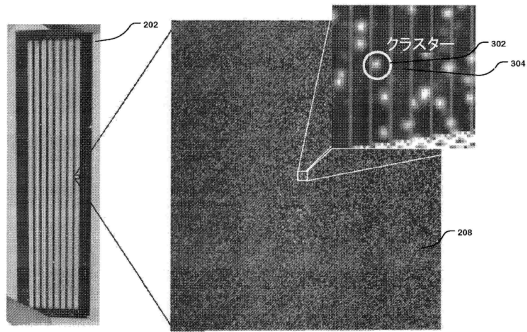
20

30

40

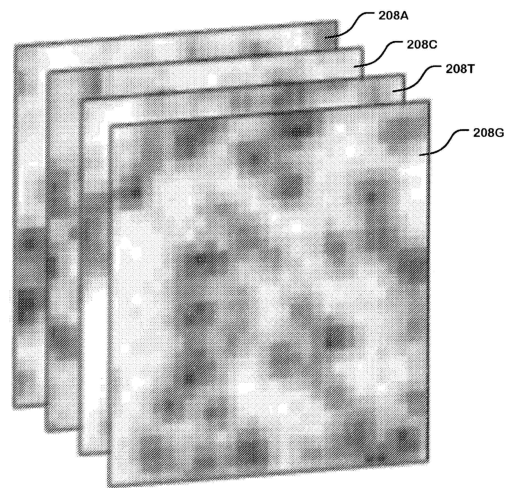
50

【図 3】



【図 4】

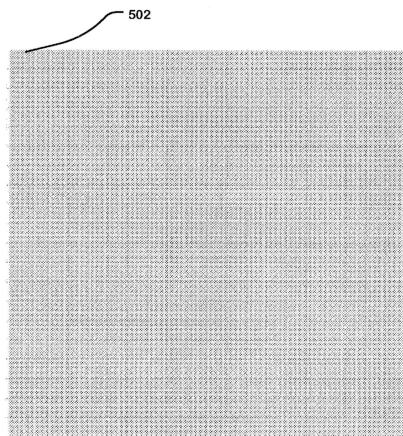
ピクセルドメイン108中の配列決定画像



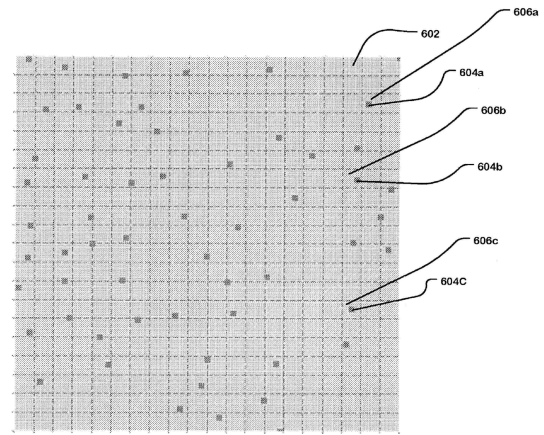
10

【図 5】

ベースコールされるサブピクセル



【図 6】



クラスター中心(RTA)及び原点サブピクセル

20

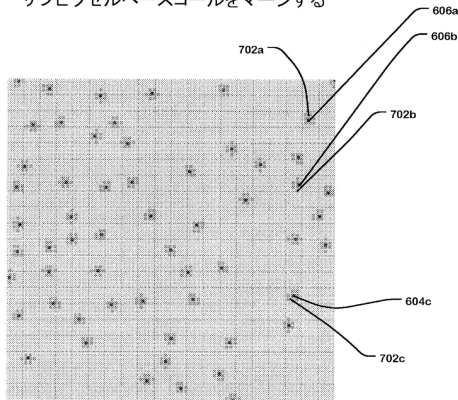
30

40

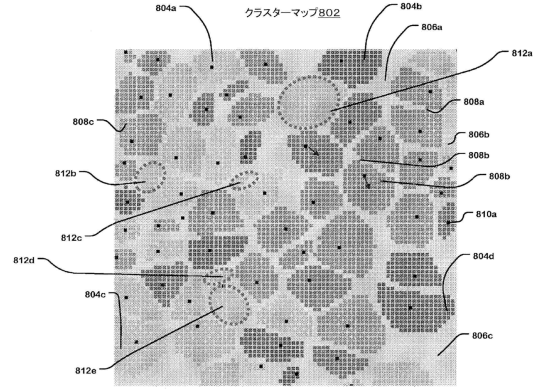
50

【図 7】

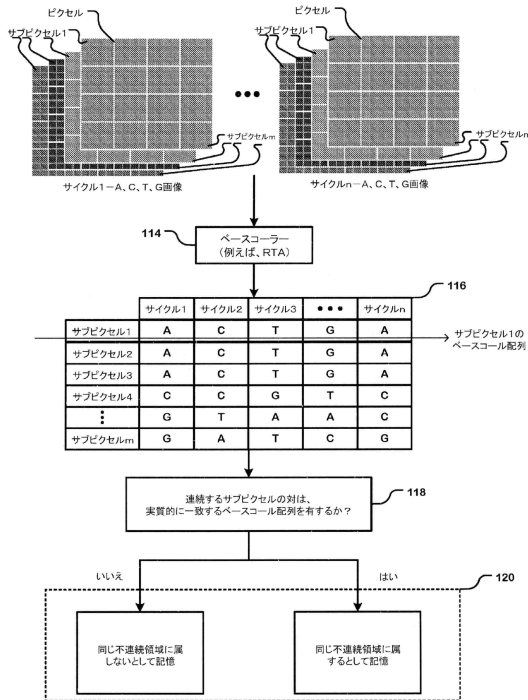
サブピクセルベースコルをマージする



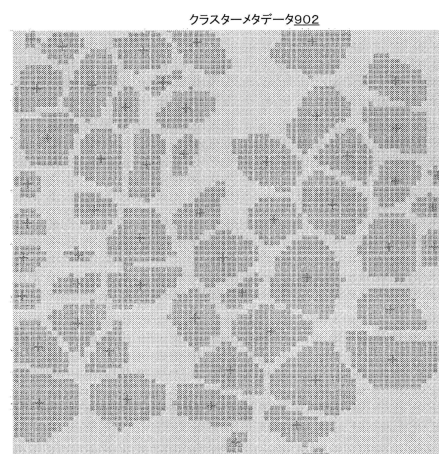
【図 8 a】



【図 8 b】



【図 9】



クラスターの空間分布—クラスター中心、形状、サイズ、背景、及び／又は境界

10

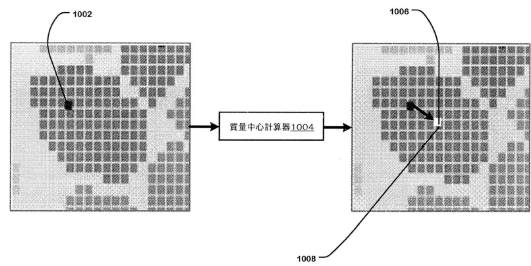
20

30

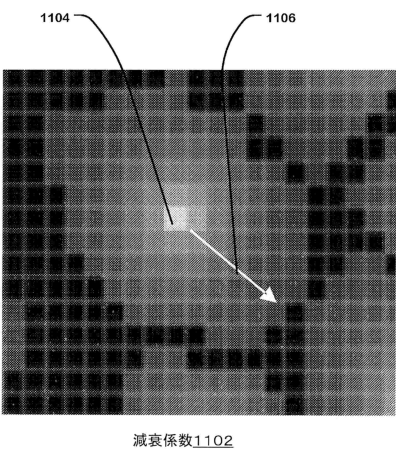
40

50

【図 10】

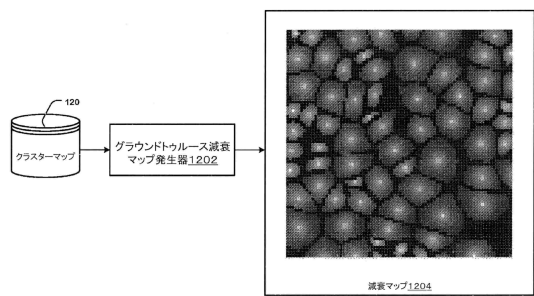


【図 11】

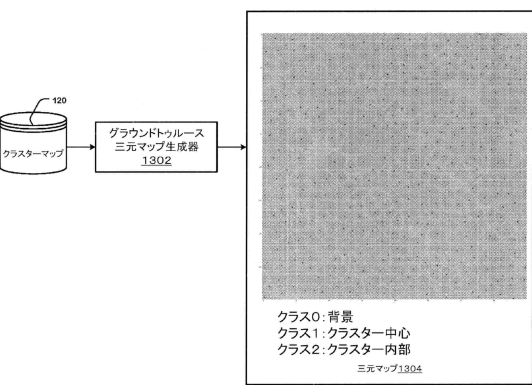


10

【図 12】



【図 13】



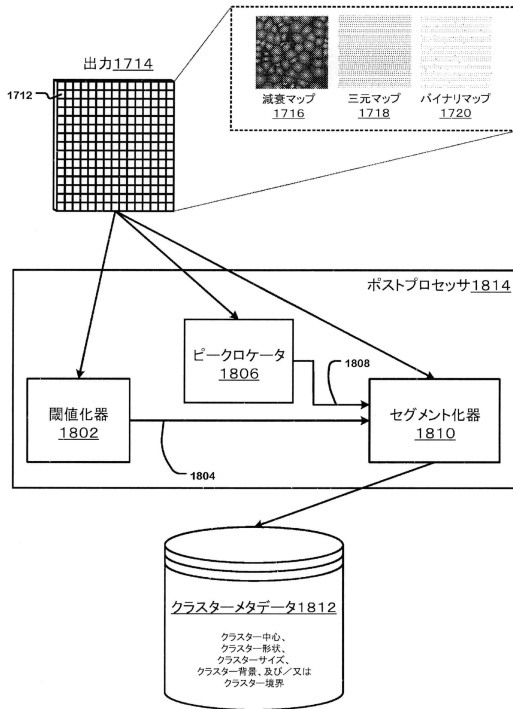
20

30

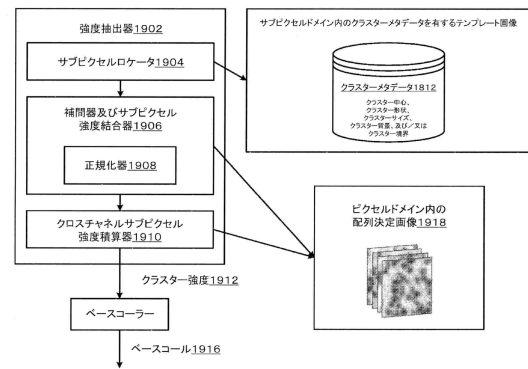
40

50

【図 18】



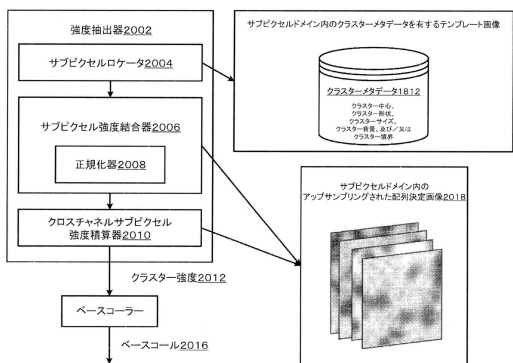
【図 19】



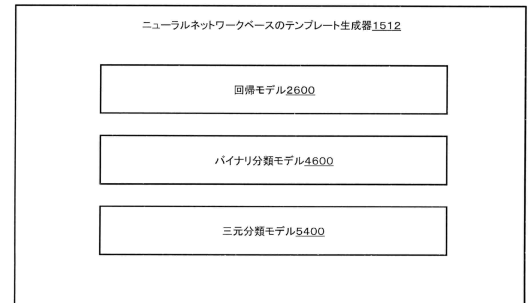
10

20

【図 20】



【図 21 a】

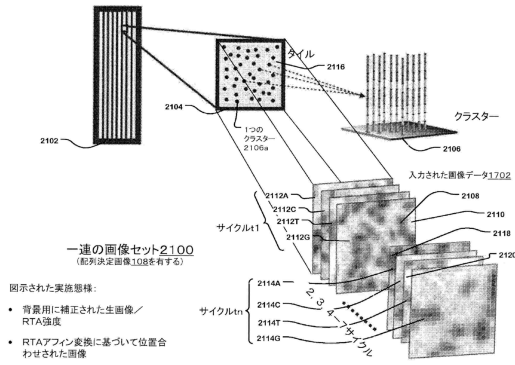


30

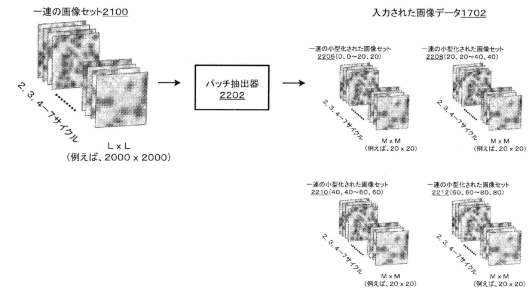
40

50

【図 2 1 b】

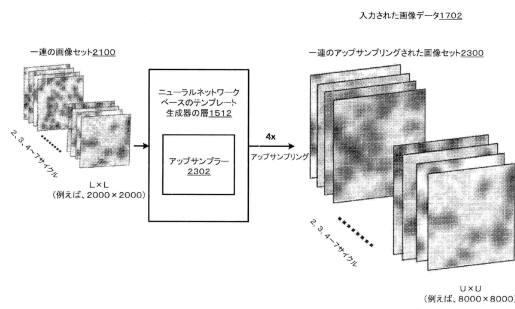


【図 2 2】

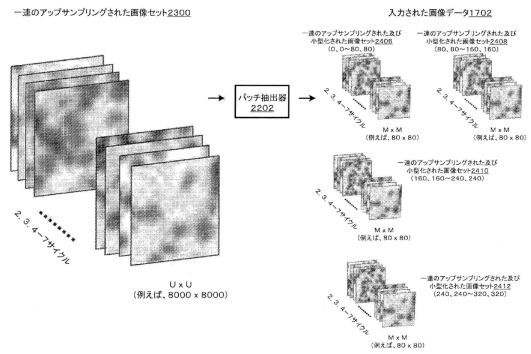


10

【図 2 3】

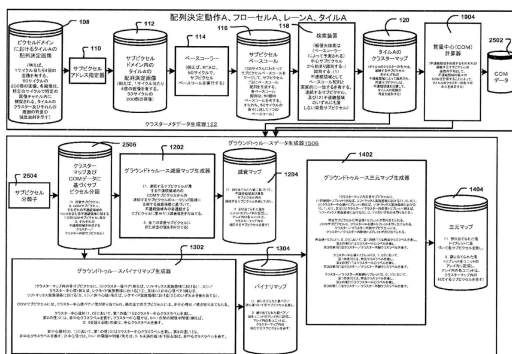


【図 2 4】

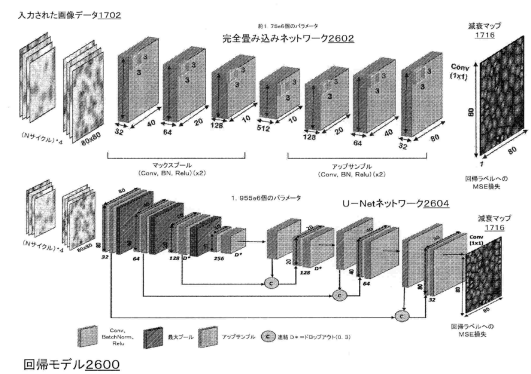


20

【図 2 5】



【図 2 6】

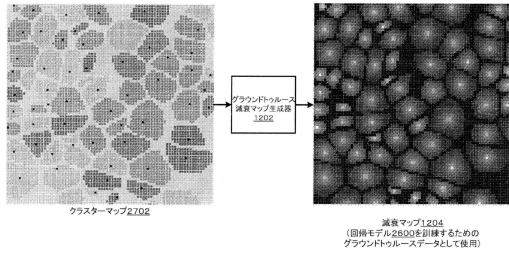


30

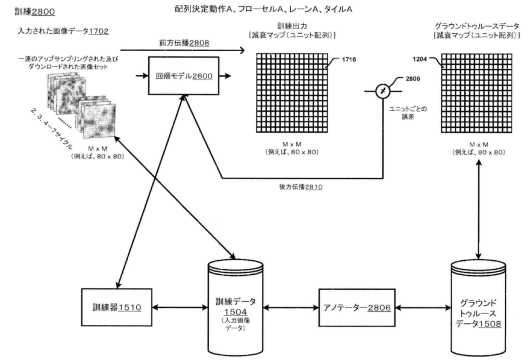
40

50

【図 27】

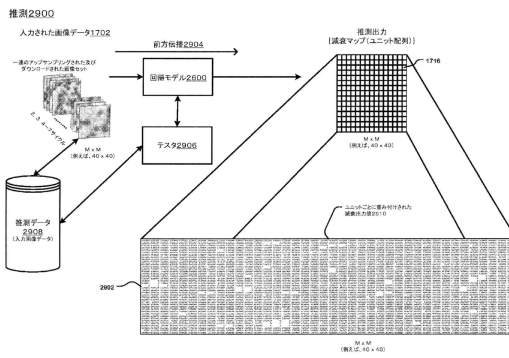


【図 28】

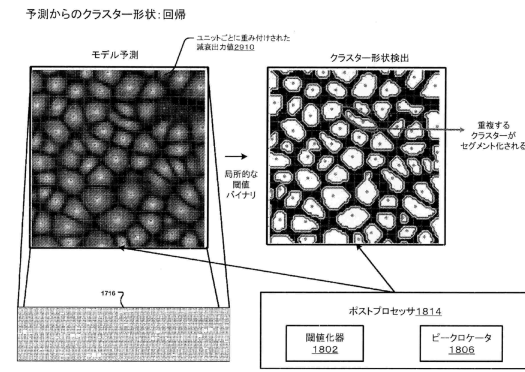


10

【図 29】



【図 30】



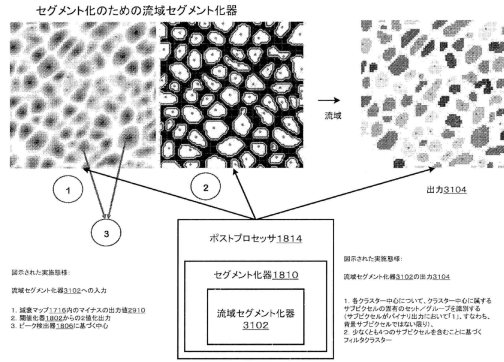
20

30

40

50

【図 3 1】



【図 3 2】

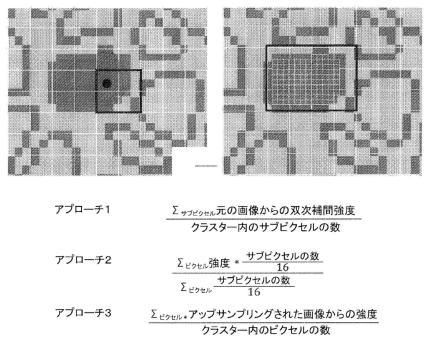
層(レイヤー)	出力形状	パラメータ	接続先
Input 1 (InputLayer)	(None, None, None, 28)	0	Input 1 [0] [0]
batch normalization 1 (BatchNormal)	(None, None, None, 28)	112	batch normalization 1 [0] [0]
conv2d 1 (Conv2D)	(None, None, None, 32)	8964	conv2d 1 [0] [0]
batch normalization 2 (BatchNormal)	(None, None, None, 32)	128	conv2d 1 [0] [0]
conv2d 2 (Conv2D)	(None, None, None, 32)	9216	batch normalization 2 [0] [0]
batch normalization 3 (BatchNormal)	(None, None, None, 32)	128	conv2d 2 [0] [0]
max pooling2d 1 (MaxPooling2D)	(None, None, None, 32)	0	batch normalization 3 [0] [0]
conv2d 3 (Conv2D)	(None, None, None, 64)	18432	max pooling2d 1 [0] [0]
batch normalization 4 (BatchNormal)	(None, None, None, 64)	256	conv2d 3 [0] [0]
conv2d 4 (Conv2D)	(None, None, None, 64)	36864	batch normalization 4 [0] [0]
batch normalization 5 (BatchNormal)	(None, None, None, 64)	256	conv2d 4 [0] [0]
max pooling2d 2 (MaxPooling2D)	(None, None, None, 64)	0	batch normalization 5 [0] [0]
conv2d 5 (Conv2D)	(None, None, None, 128)	137792	max pooling2d 2 [0] [0]
batch normalization 6 (BatchNormal)	(None, None, None, 128)	512	conv2d 5 [0] [0]
conv2d 6 (Conv2D)	(None, None, None, 128)	147456	batch normalization 6 [0] [0]
batch normalization 7 (BatchNormal)	(None, None, None, 128)	512	conv2d 6 [0] [0]
dropout 1 (Dropout)	(None, None, None, 128)	0	batch normalization 7 [0] [0]
max pooling2d 3 (MaxPooling2D)	(None, None, None, 128)	0	dropout 1 [0] [0]
conv2d 7 (Conv2D)	(None, None, None, 256)	294912	max pooling2d 3 [0] [0]
batch normalization 8 (BatchNormal)	(None, None, None, 256)	1024	conv2d 7 [0] [0]
conv2d 8 (Conv2D)	(None, None, None, 256)	318720	batch normalization 8 [0] [0]
batch normalization 9 (BatchNormal)	(None, None, None, 256)	1024	conv2d 8 [0] [0]
dropout 2 (Dropout)	(None, None, None, 256)	0	batch normalization 9 [0] [0]
up sampling2d 1 (UpSampling2D)	(None, None, None, 256)	0	dropout 2 [0] [0]
concatenate 1 (Concatenate)	(None, None, None, 512)	0	up sampling2d 1 [0] [0]
conv2d 9 (Conv2D)	(None, None, None, 512)	442368	concatenate 1 [0] [0]
batch normalization 10 (BatchNormal)	(None, None, None, 512)	512	conv2d 9 [0] [0]
conv2d 10 (Conv2D)	(None, None, None, 512)	474560	batch normalization 10 [0] [0]
batch normalization 11 (BatchNormal)	(None, None, None, 512)	512	conv2d 10 [0] [0]
dropout 3 (Dropout)	(None, None, None, 512)	0	batch normalization 11 [0] [0]
up sampling2d 2 (UpSampling2D)	(None, None, None, 512)	0	dropout 3 [0] [0]
concatenate 2 (Concatenate)	(None, None, None, 1024)	0	up sampling2d 2 [0] [0]
conv2d 11 (Conv2D)	(None, None, None, 64)	110784	concatenate 2 [0] [0]
batch normalization 12 (BatchNormal)	(None, None, None, 64)	256	conv2d 11 [0] [0]
conv2d 12 (Conv2D)	(None, None, None, 64)	106448	batch normalization 12 [0] [0]
batch normalization 13 (BatchNormal)	(None, None, None, 64)	256	conv2d 12 [0] [0]
up sampling2d 3 (UpSampling2D)	(None, None, None, 64)	0	batch normalization 13 [0] [0]
concatenate 3 (Concatenate)	(None, None, None, 128)	0	up sampling2d 3 [0] [0]
conv2d 13 (Conv2D)	(None, None, None, 32)	27648	concatenate 3 [0] [0]
batch normalization 14 (BatchNormal)	(None, None, None, 32)	128	conv2d 13 [0] [0]
conv2d 14 (Conv2D)	(None, None, None, 32)	6216	batch normalization 14 [0] [0]
batch normalization 15 (BatchNormal)	(None, None, None, 32)	128	conv2d 14 [0] [0]
prod (Conv2D)	(None, None, None, 1)	32	batch normalization 15 [0] [0]

10

20

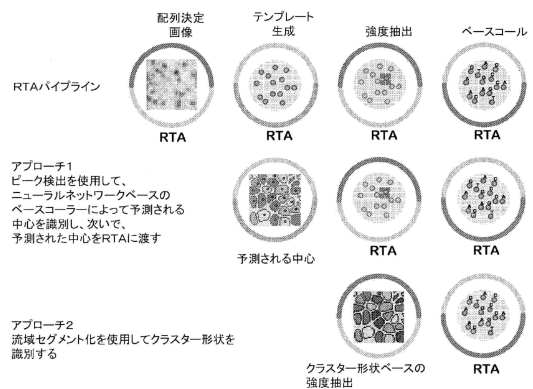
【図 3 3】

クラスター形状を使用した強度抽出



【図 3 4】

予測されるクラスターがどれくらい良好か？



30

40

50

【図 3 5】

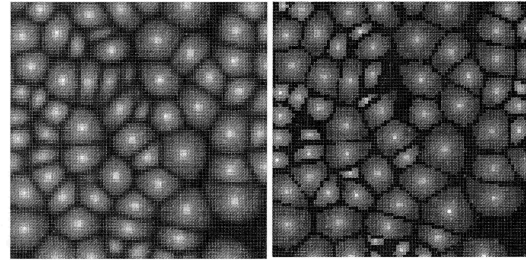
質量中心対RTA中心

質量中心	RTA中心
合計956048+0(QC通過リード+QC失敗リード) 158+0 二次 0+0 補足 841+0複製 936609+0 マッピング(97.97% N/A) 955890+0 配列決定中に対を成した 477945+0 リード1 477945+0 リード2 928926+0 適切に対を成した (97.18% N/A) 930732+0 それ自体とメートがマッピングされる 5719+0 島の個片(0.60% N/A) 0+0 異なるchrにマッピングされたメートを有する 0+0 異なるchrにマッピングされたメートを 有する(mapQ>=5)	合計956054+0(QC通過リード+QC失敗リード) 164+0 二次 0+0 補足 1045+0複製 932910+0 マッピング(97.68% N/A) 955890+0 配列決定中に対を成した 477945+0 リード1 477945+0 リード2 923322+0 適切にペアリングされた (96.59% N/A) 925276+0 それ自体とメートがマッピングされる 7470+0島の個片(0.75% N/A) 0+0 異なるchrにマッピングされたメートを有する 0+0 異なるchrにマッピングされたメートを 有する(mapQ>=5)
マッピングされた塩基141404101#無視クリップ マッピングされた塩基(cigar):140673703 塩基をトリミングする:0 塩基が複製される:126991 不一致:498799# NMフィールドから 誤差率:3.545787e-03 #不一致/ 塩基マッピング(cigar)	マッピングされた塩基140844646#無視クリップ マッピングされた塩基(cigar):139683478 塩基をトリミングする:0 塩基が複製される:157795 不一致:821045# NMフィールドから 誤差率:5.877897e-03 #不一致/ 塩基マッピング(cigar)

【図 3 6】

モデル予測

グラウンドトゥールース

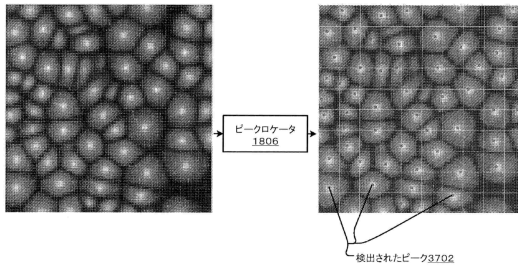


10

【図 3 7】

モデル予測

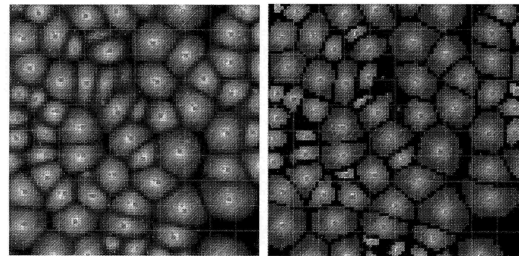
ピーク検出



【図 3 8】

モデル予測

グラウンドトゥールース



20

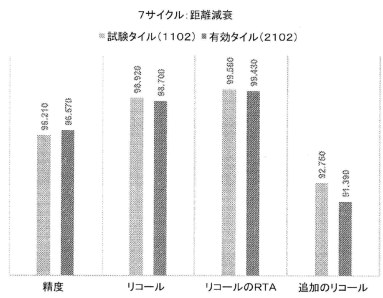
30

40

50

【図 3 9】

回帰モデル:最適化後



□ 識別された全ての中心を回収するのに良好

【図 4 0】

通常動作 20pMタイル:1102(7サイクル)

	RTA	境界 +/- 10(Unet)内
# クラスター	903,040	988,884
% PF	88.55% (799,684)	84.54% (836,197)
% 配列	97.50%	97.47%
% 重複	0.05%	0.12%
% 不一致	0.37%	34%
% ソフトクリップ	2.07%	2.05%
% Q30塩基	90.59%	91.42%
非重複の 適切なリード対	769,150	803,473 (4.46% 34,323)

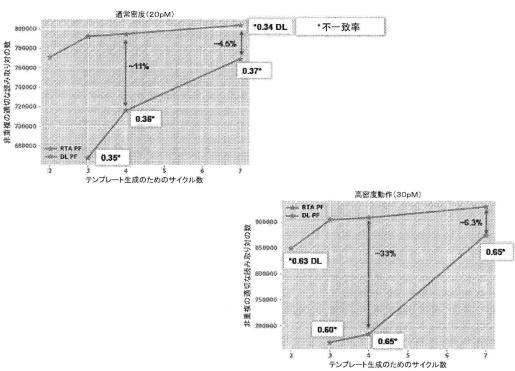
10

【図 4 1】

高密度動作30 pMタイル:2103(7サイクル)

	RTA	境界 +/- 10(Unet)内
# クラスター	1,157,285	1,351,588
% PF	78.58% (909,340)	71.62% (968,066)
% 配列	97.69%	97.62%
% 重複	0.05%	0.07%
% 不一致	0.65%	0.63%
% ソフトクリップ	2.60%	2.57%
% Q30塩基	83.45%	83.88%
非重複の 適切なリード対	874,000	928,787 (6.27%)

【図 4 2】



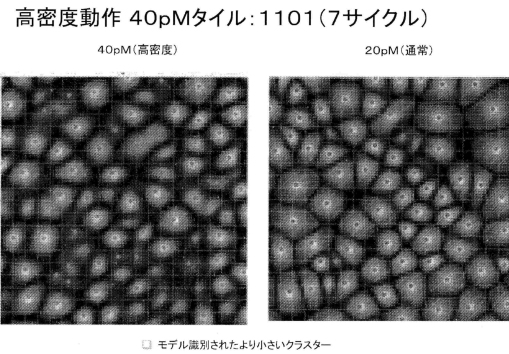
20

30

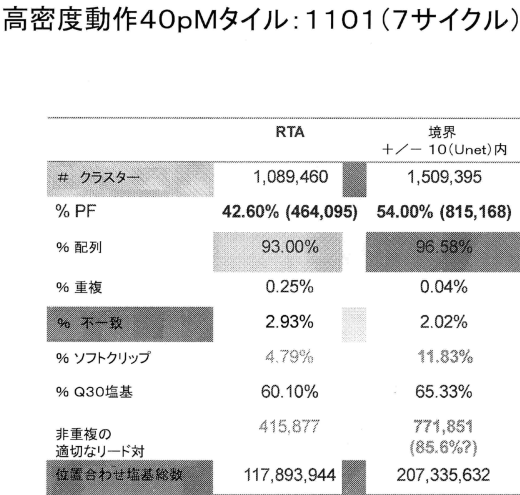
40

50

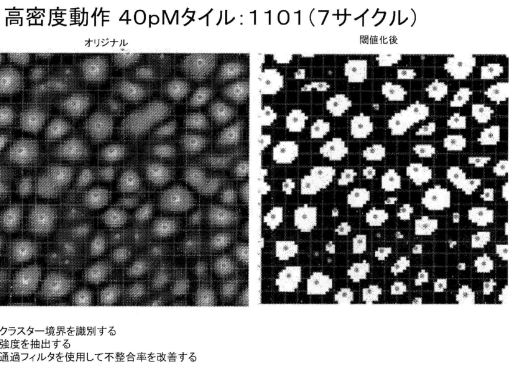
【図 4 3】



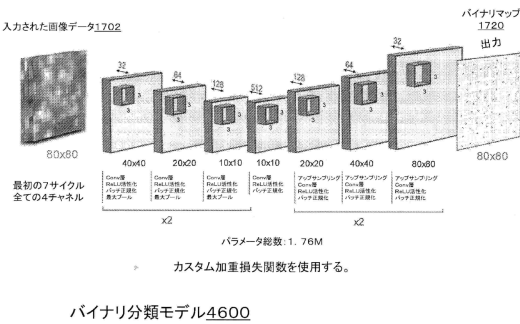
【図 4 4】



【図 4 5】



【図 4 6】



10

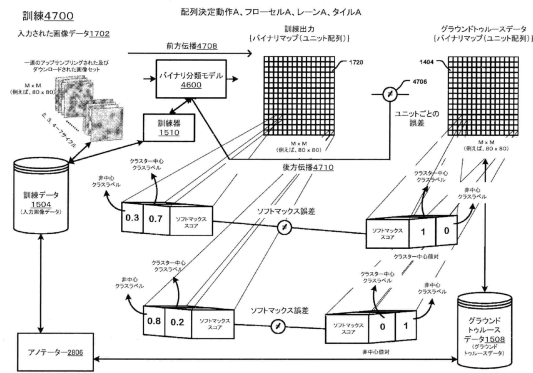
20

30

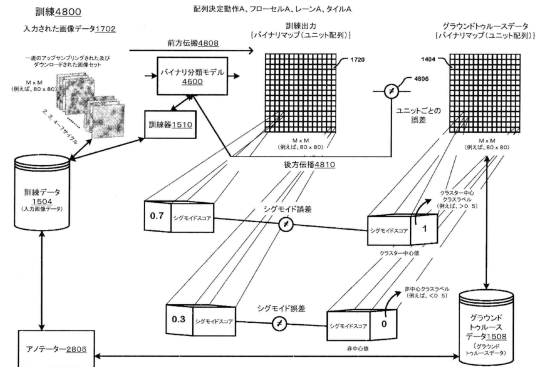
40

50

【図 47】



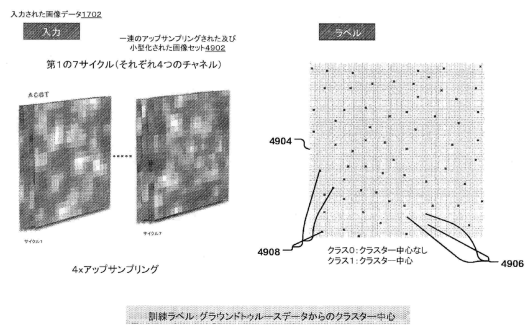
【圖 48】



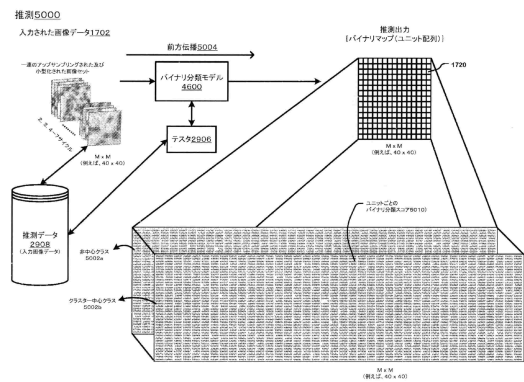
10

【図 49】

バイナリ分類: 訓練ラベル



【図 50】



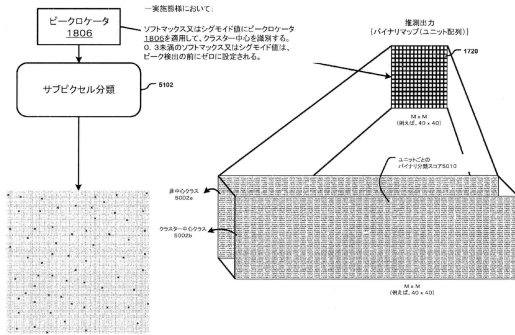
20

30

40

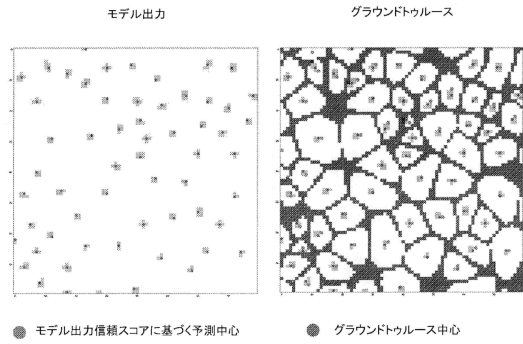
50

【図 5 1】



【図 5 2 a】

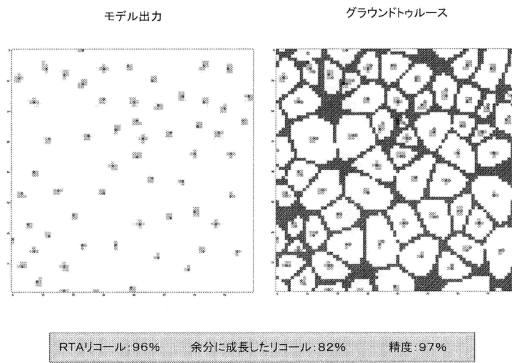
バイナリ分類: 予測



10

【図 5 2 b】

バイナリ分類: 予測



【図 5 3】

層(タイプ)	出力形状	パラメータ数
入力_1 (入力層)	(None, 80, 80, 24)	0
conv2d_1 (Conv2D)	(None, 80, 80, 24)	6084
プット_正規化_1 (バッチ)	(None, 80, 80, 24)	128
conv2d_2 (Conv2D)	(None, 80, 80, 24)	6216
プット_正規化_2 (バッチ)	(None, 80, 80, 24)	128
max_pooling2d_1 (MaxPooling2D)	(None, 40, 40, 24)	0
conv2d_3 (Conv2D)	(None, 40, 40, 64)	18432
プット_正規化_3 (バッチ)	(None, 40, 40, 64)	256
conv2d_4 (Conv2D)	(None, 40, 40, 64)	18464
プット_正規化_4 (バッチ)	(None, 40, 40, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 20, 20, 64)	0
conv2d_5 (Conv2D)	(None, 20, 20, 128)	73728
プット_正規化_5 (バッチ)	(None, 20, 20, 128)	512
conv2d_6 (Conv2D)	(None, 20, 20, 128)	147456
プット_正規化_6 (バッチ)	(None, 20, 20, 128)	512
ドロップアウト_1 (Dropout)	(None, 20, 20, 128)	0
max_pooling2d_3 (MaxPooling2D)	(None, 10, 10, 128)	0
conv2d_7 (Conv2D)	(None, 10, 10, 512)	589824
プット_正規化_7 (バッチ)	(None, 10, 10, 512)	2048
ドロップアウト_2 (Dropout)	(None, 10, 10, 512)	0
up_sampling2d_1 (UpSampling2D)	(None, 20, 20, 512)	0
conv2d_8 (Conv2D)	(None, 20, 20, 128)	88324
プット_正規化_8 (バッチ)	(None, 20, 20, 128)	512
conv2d_9 (Conv2D)	(None, 20, 20, 128)	147456
プット_正規化_9 (バッチ)	(None, 20, 20, 128)	512
up_sampling2d_2 (UpSampling2D)	(None, 40, 40, 128)	0
conv2d_10 (Conv2D)	(None, 40, 40, 64)	73728
プット_正規化_10 (バッチ)	(None, 40, 40, 64)	256
conv2d_11 (Conv2D)	(None, 40, 40, 64)	18464
プット_正規化_11 (バッチ)	(None, 40, 40, 64)	256
up_sampling2d_3 (UpSampling2D)	(None, 80, 80, 64)	0
conv2d_12 (Conv2D)	(None, 80, 80, 32)	18432
プット_正規化_12 (バッチ)	(None, 80, 80, 32)	128
conv2d_13 (Conv2D)	(None, 80, 80, 32)	9216
プット_正規化_13 (バッチ)	(None, 80, 80, 32)	128
全層に conv2d_13	(None, 80, 80, 21)	84
全層に conv2d_13	(None, 80, 80, 21)	84
訓練精度(バッチサイズ: 768)		0.84
訓練精度(バッチサイズ: 8192)		0.81

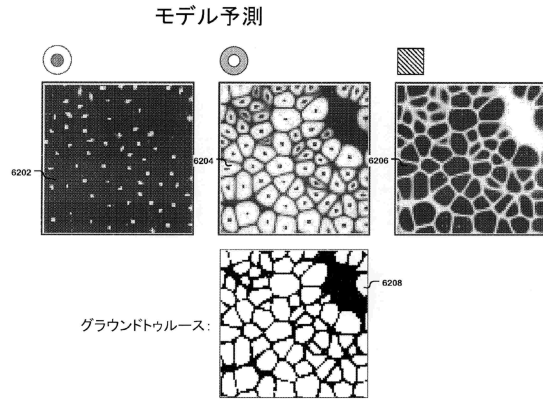
20

30

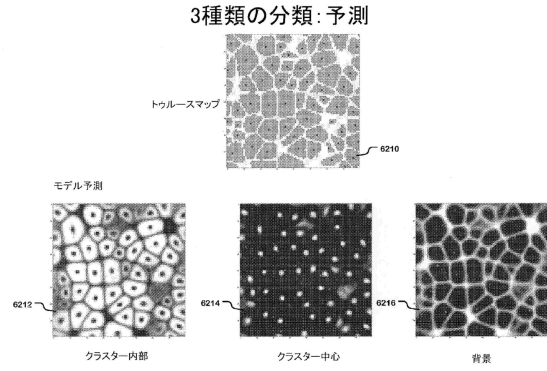
40

50

【図 6 2 a】

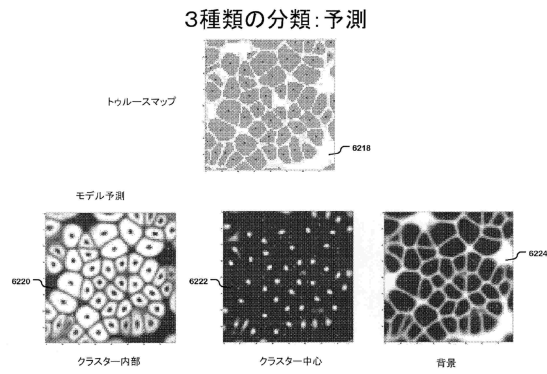


【図 6 2 b】

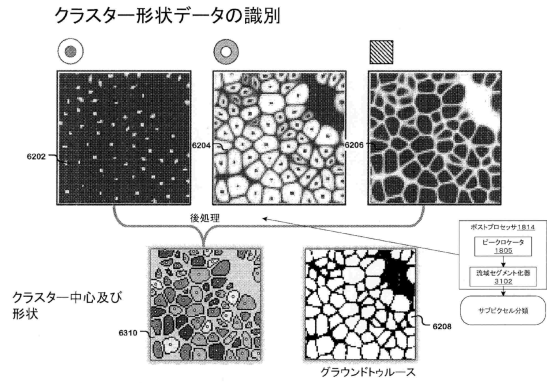


10

【図 6 2 c】



【図 6 3】



20

30

40

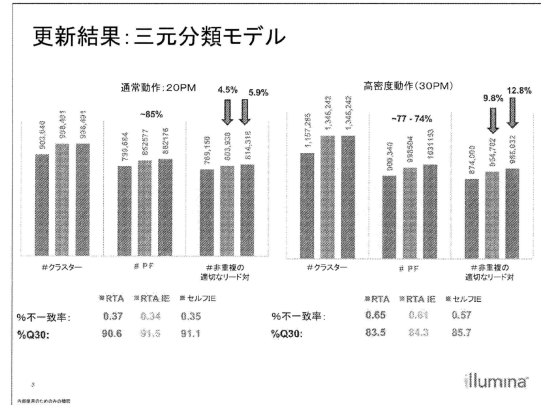
50

【図 6 4】

ベースコールの比較(RTA1. 18を使用)

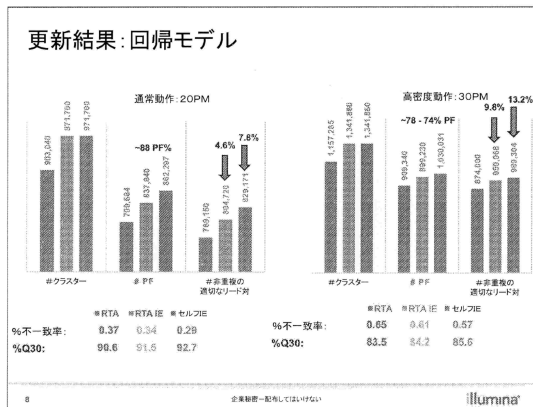
	RTA	DL バイナリ分類	DL 回帰
# クラスター	493,944	519,100	543,348
% PF	88.8%	86.99%	85.86%
% 配列	97.63%	97.68%	97.67%
% 重複	0.02%	0.2%	0.14%
% 不一致	0.34%	0.29%	0.29%
% ソフトクリップ	2.02%	1.97%	1.97%
% Q30塩基	91.94%	92.76%	92.81%
適切なリード対の合計	422,907	435,737	449,993
非重複の 適切なリード対	424,665	434,869	449,344

【図 6 5】



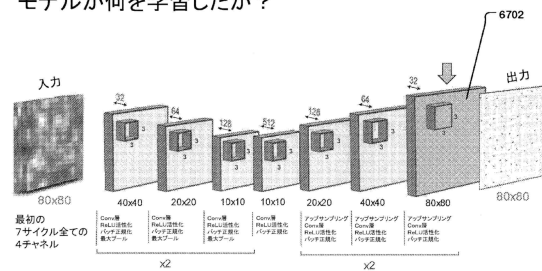
10

【図 6 6】



【図 6 7】

モデルが何を学習したか？



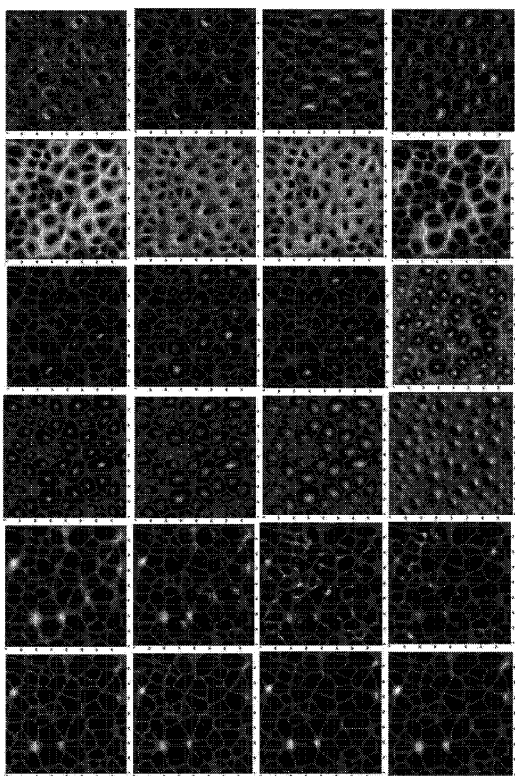
20

30

40

50

【圖 68】



【圖 6 9】

NextSeq

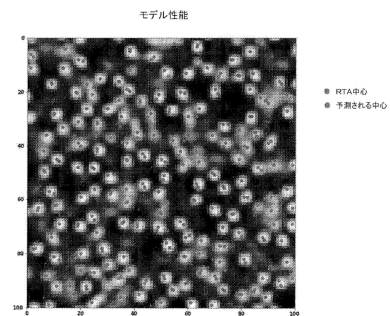


Figure 68

10

20

【図 70】

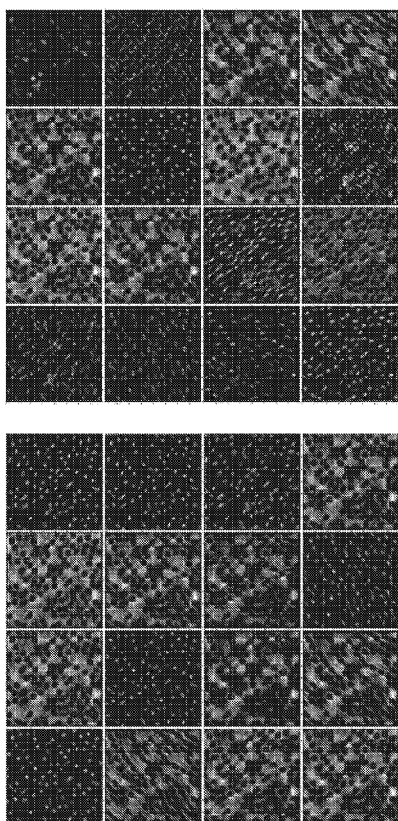
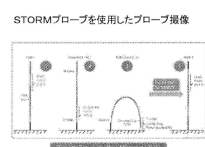


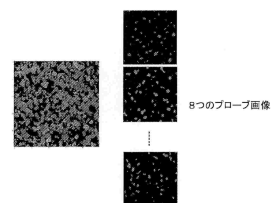
Figure 70

【圖 7 1】

プローブ撮像データ



STORMプローブ-17インデックス領域に混成するように設計された Alexa-Fluor647で標識された8オリゴ



データセット



リード1の101サイクル+17インデックスサイクル

Miseq, 10pM
Miseq, 100pM
Miseq, E. Coli 40pM

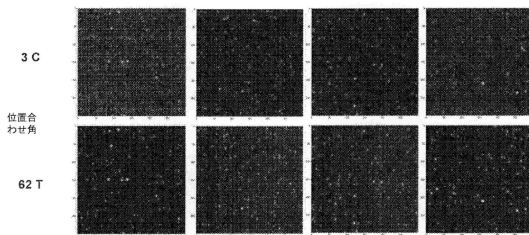
30

40

50

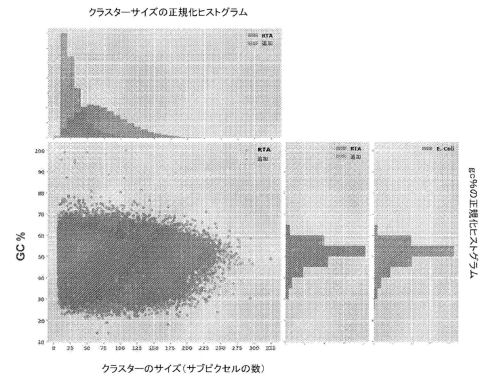
【図 7 2】

画像位置合わせのためのビーズを使用



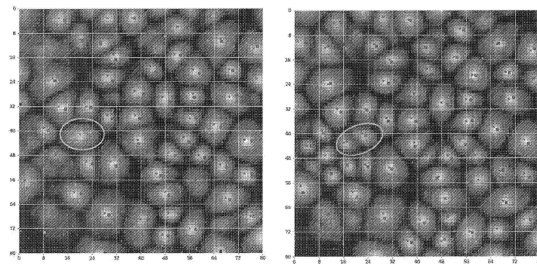
【図 7 3】

クラスター統計



【図 7 4】

モデル予測



【図 7 5】

RTA中心対質量中心

質量中心	RTA中心
合計956048+0(QC通過リード+QC失敗リード) 158+0 二次 0+0 補足 841+0 複製 956609+0 マッピング(97.97% N/A) 955890+0 配列決定中に対を成した 477945+0 リード1 477945+0 リード2 928925+0 適切にペアリングされた (97.18% N/A) 930732+0 それ自体とメートがマッピングされる 5719+0 彼の断片(0.60% N/A) 0+0 異なるchrにマッピングされたメートを有する 0+0 異なるchrにマッピングされたメートを有する(mapQ>=5)	合計956054+0(QC通過リード+QC失敗リード) 164+0 二次 0+0 補足 1045+0 複製 932910+0 マッピング(97.58% N/A) 955890+0 配列決定中に対を成した 477945+0 リード1 477945+0 リード2 923322+0 適切にペアリングされた (96.69% N/A) 925276+0 それ自体とメートがマッピングされる 7470+0 彼の断片(0.78% N/A) 0+0 異なるchrにマッピングされたメートを有する 0+0 異なるchrにマッピングされたメートを有する(mapQ>=5)
マッピングされた塩基141404101 # 無視クリップ マッピングされた塩基(cigar): 140673703 塩基をトリミングする: 0 塩基が複製される: 126991 不一致: 496799 # NMフィールドから 誤差率: 3.545787e-03 # 不一致/ 塩基マッピング(cigar)	マッピングされた塩基140844646 # 無視クリップ マッピングされた塩基(cigar): 139683478 塩基をトリミングする: 0 塩基が複製される: 157795 不一致: 821045 # NMフィールドから 誤差率: 5.877897e-03 # 不一致/ 塩基マッピング(cigar)

10

20

30

40

50

【圖 7 6】

追加成長クラスター

```

会計102777+0(QC通過リード+QC失敗リード)
27+0 二次
0+0 補足
129+0 複製
99134+0 マッピング(96.46% N/A)
102750+0 配列決定中に対を成した
51375+リード1
51375+リード2
96742+0 特性ベアリングされた(94.15% N/A)
97296+0 それ自体とメイトがマッピングされる
1811+0 個の個片(1.76% N/A)
0+0 異なるchrにマッピングされたメイトを有する
0+0 異なるchrにマッピングされたメイトを有する(mapQ>=5)

```

【圖 7 7】

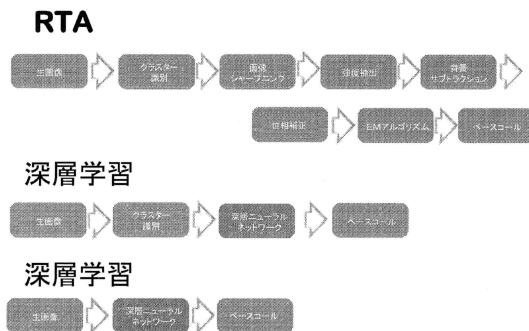
データセット

Miseq			Nextseq	
ライブラリ	密度	RTA PF%	ライブラリ	RTA PF%
20pM	1254 K/mm2	85%	1.3pM	61%
30pM	1524 K/mm2	72%	1.2pM	42%
40pM	1529 K/mm2	31%	3pM	14%

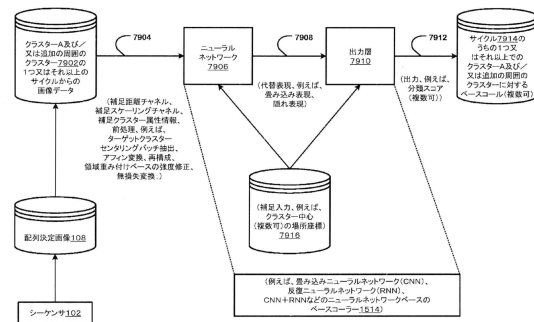
10

【圖 7 8】

ベースコール



【圖 7 9】



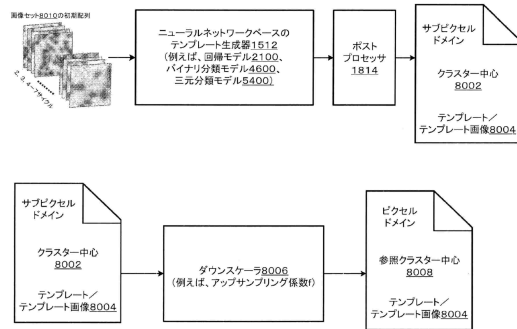
20

30

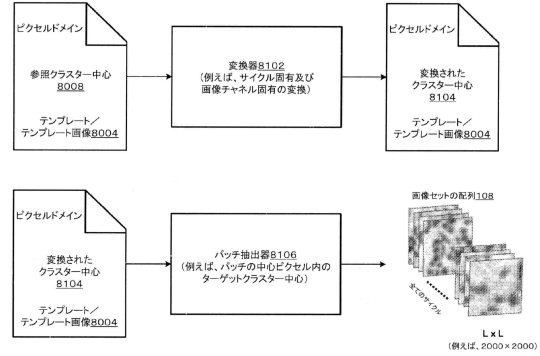
40

50

【図 80】

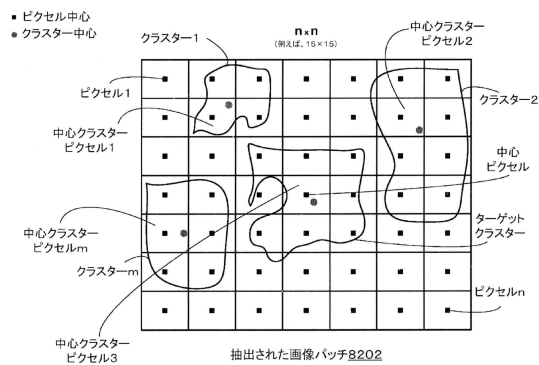


【図 81】

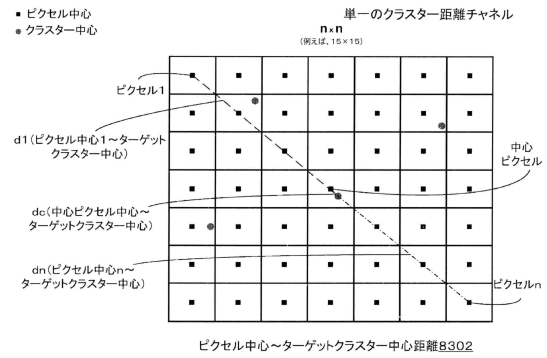


10

【図 82】

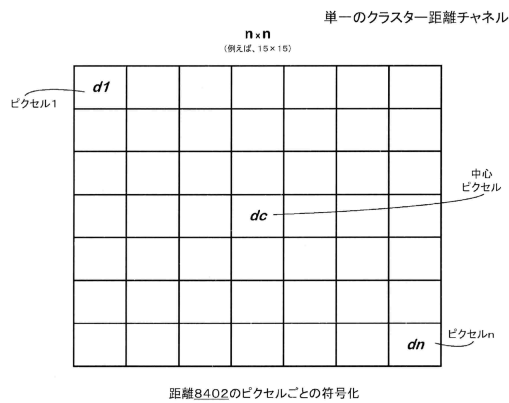


【図 83】



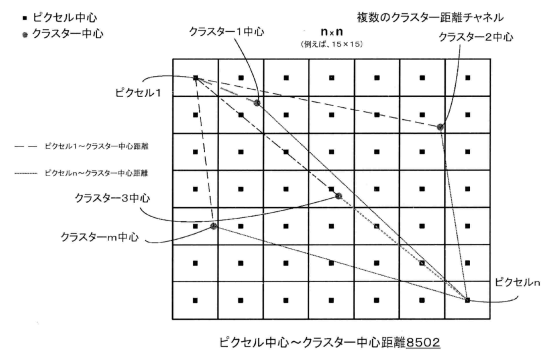
20

【図 84】



距離8402のピクセルごとの符号化

【図 85 a】



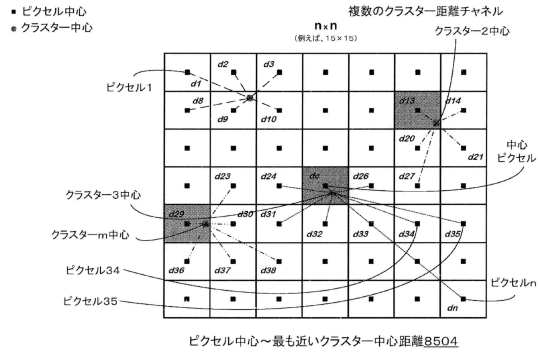
ピクセル中心〜クラスター中心距離8502

30

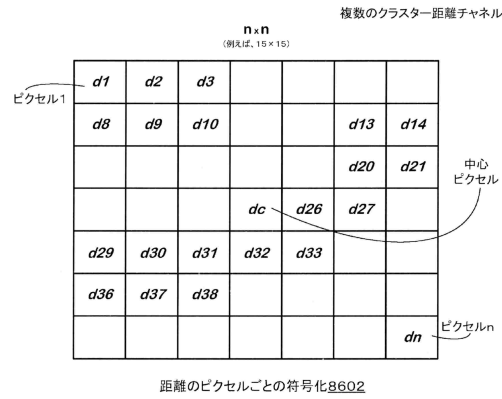
40

50

【図 85b】

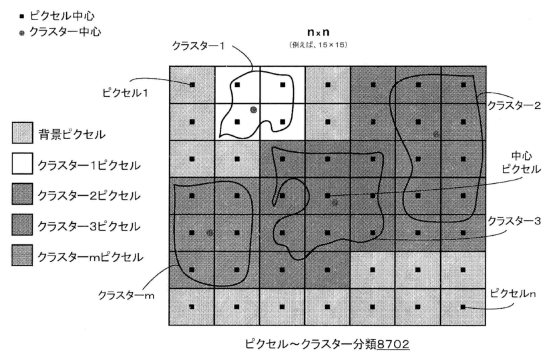


【図 86】

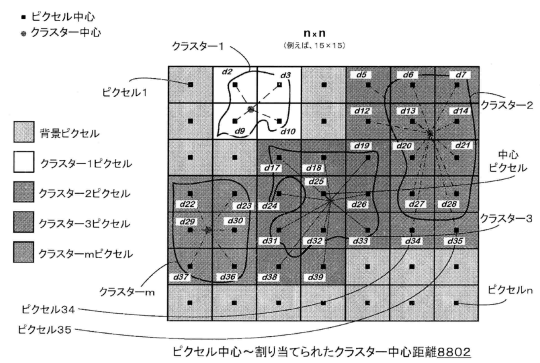


10

【図 87】



【図 88】



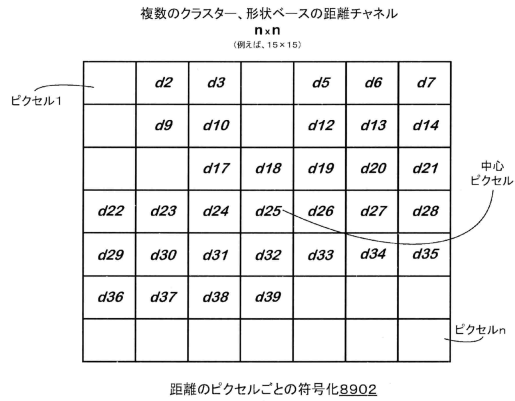
20

30

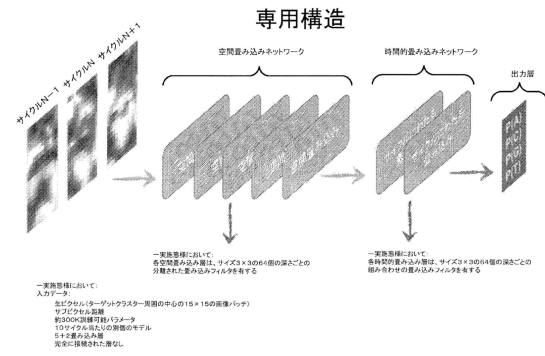
40

50

【図 89】

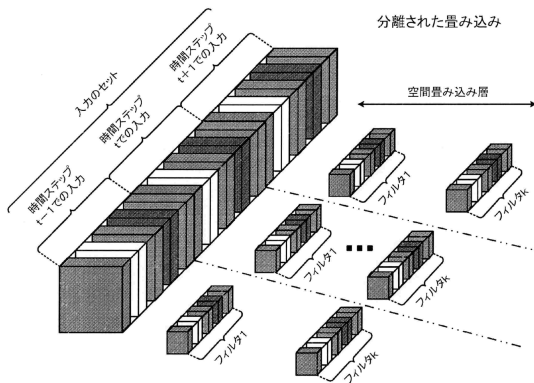


【図 90】

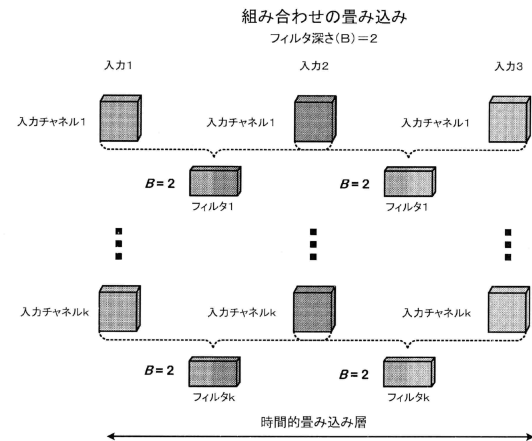


10

【図 91】



【図 92 a】



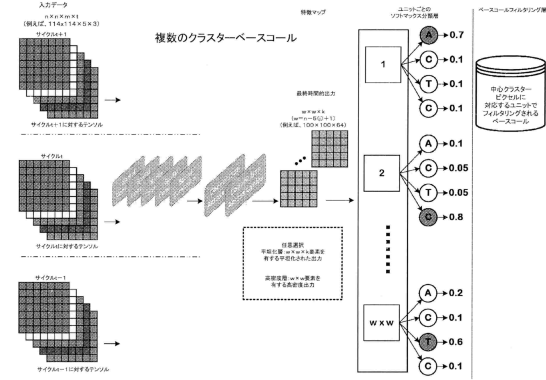
20

30

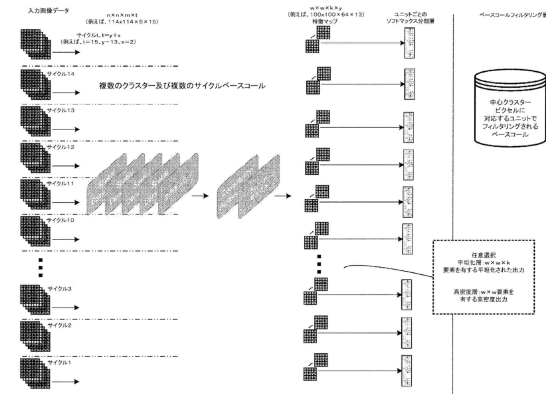
40

50

【図 9 7】



【図 9 8】



【図 9 9】



【図 1 0 0】



10

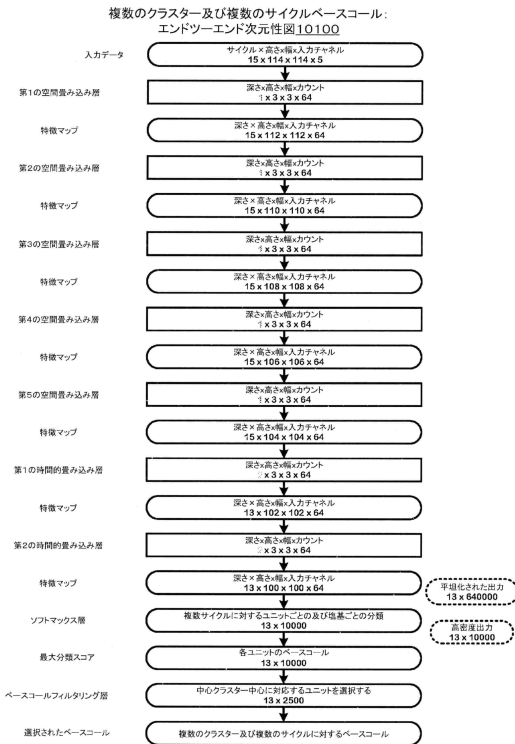
20

30

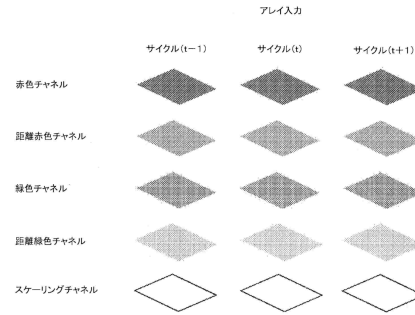
40

50

【図 101】



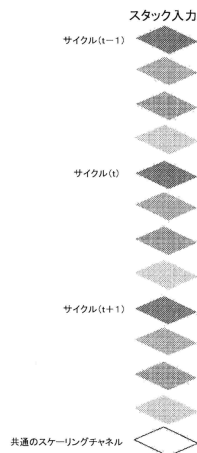
【図 102 a】



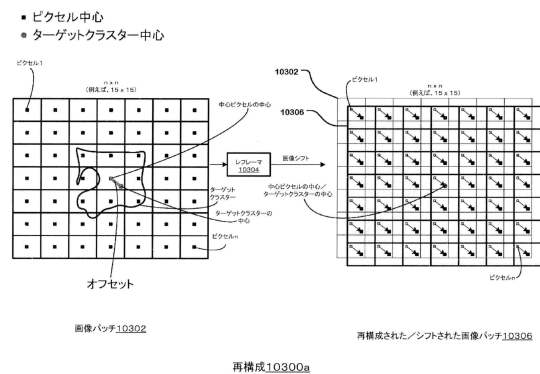
10

20

【図 102 b】



【図 103 a】

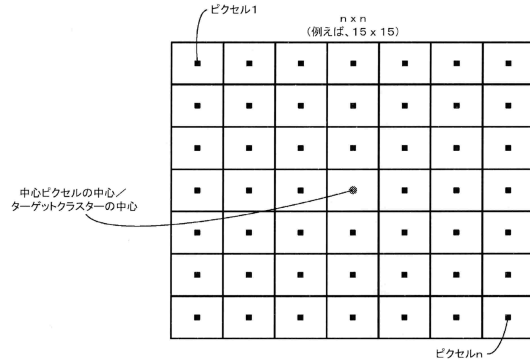


30

40

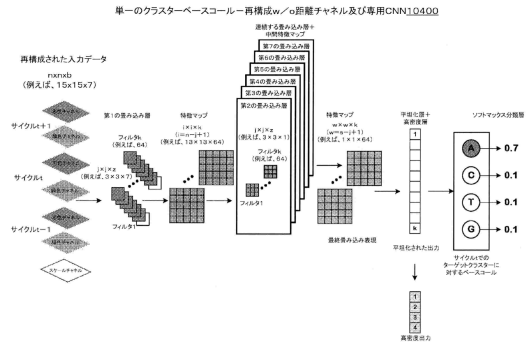
50

【図 103b】



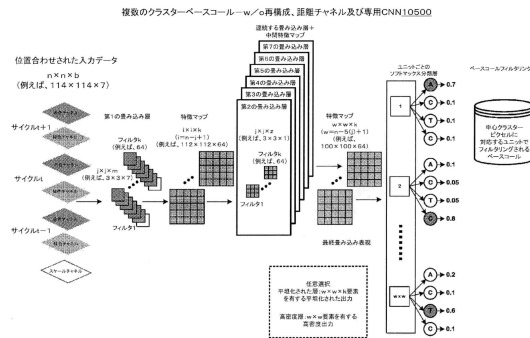
再構成された／シフトされた画像パッチ10300b

【図 104】

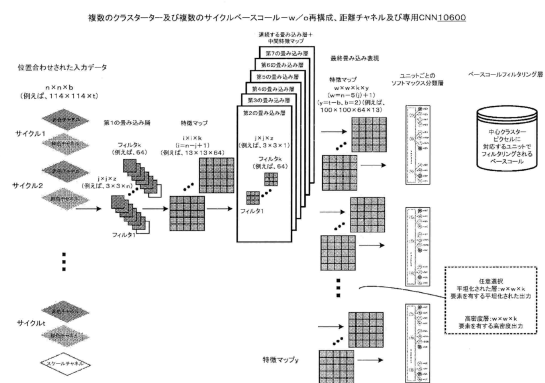


10

【図 105】



【図 106】



20

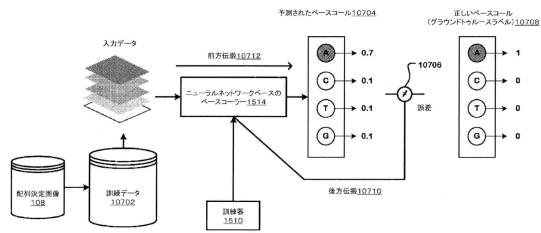
30

40

50

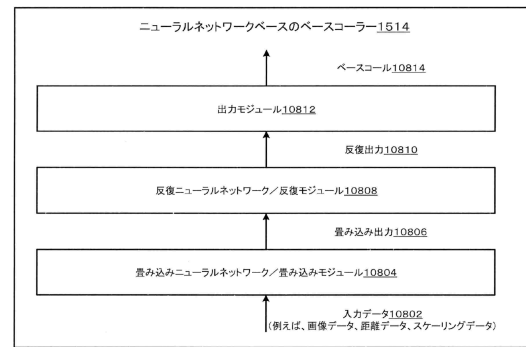
【図107】

ベースコーラーの訓練10700



【図108a】

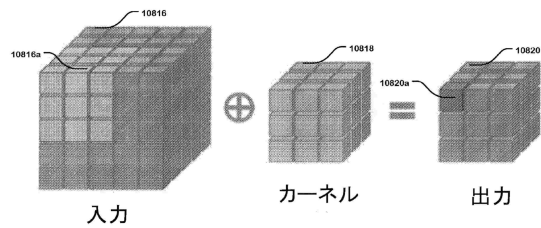
ベースコードに対するCNN-RNNベースのハイブリッドニューラルネットワーク10800a



10

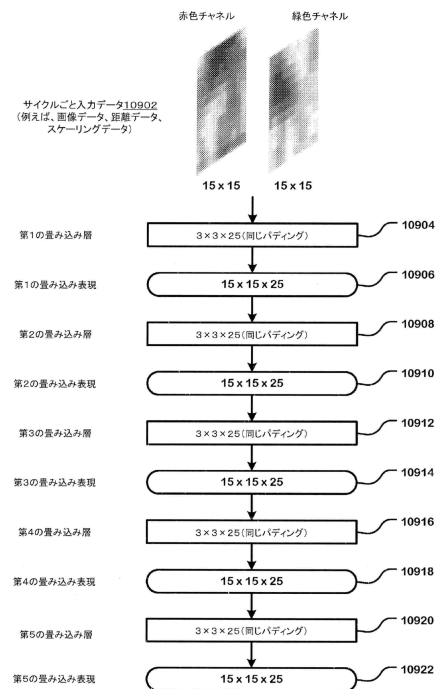
【図108b】

3D畳み込み10800b



【図109】

畳み込み層のカスケード10900



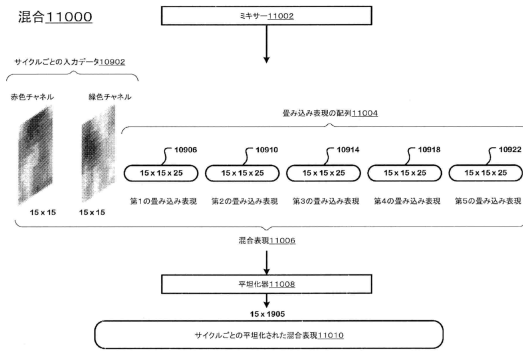
20

30

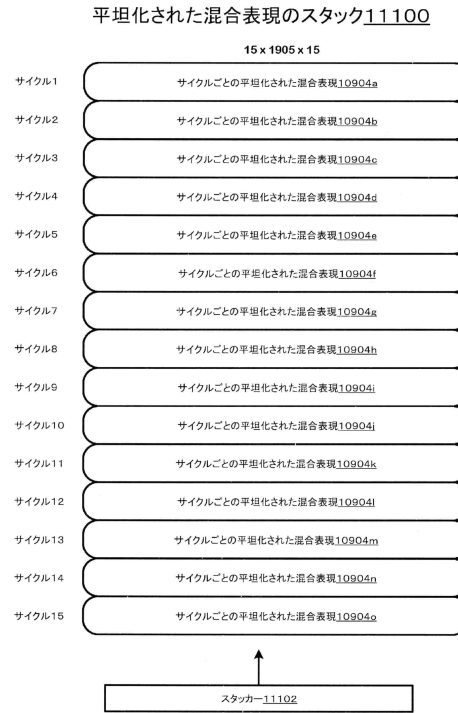
40

50

【図 1 1 0】



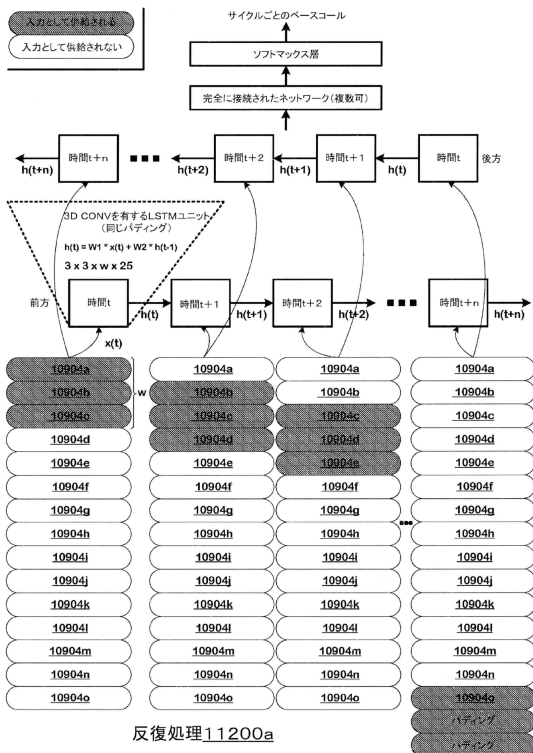
【図 1 1 1】



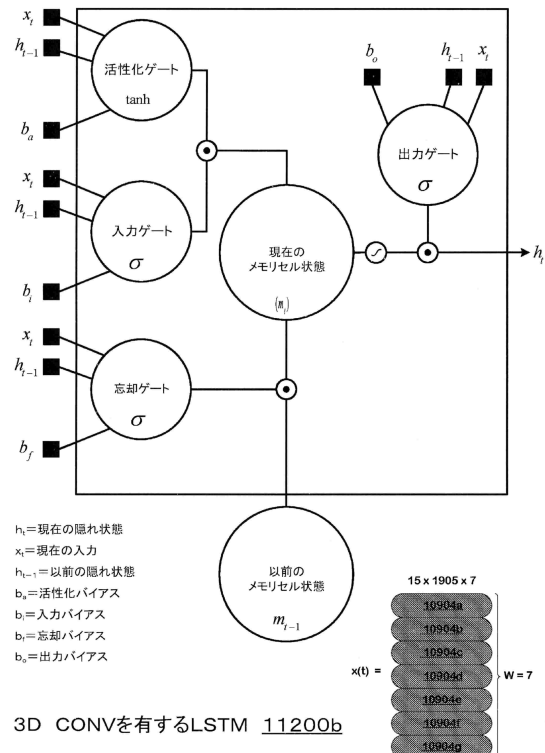
10

20

【図 1 1 2 a】



【図 1 1 2 b】

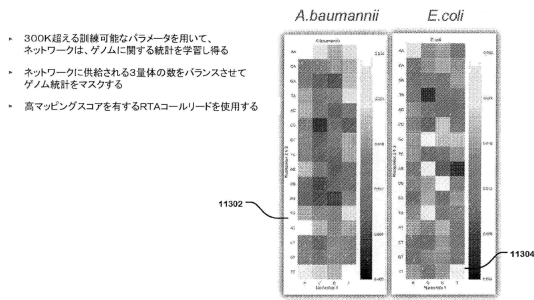


30

40

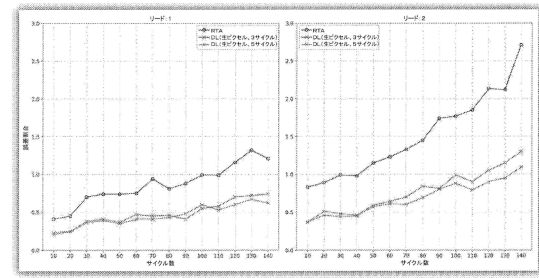
【図 1 1 3】

訓練データ内の3量体バランス



【図 1 1 4】

深層学習を使用するベースコール

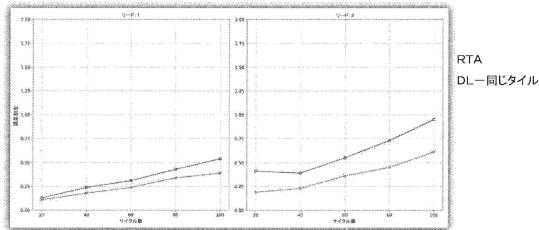


10

【図 1 1 5】

深層学習を使用するベースコール

タイル対タイル一般化

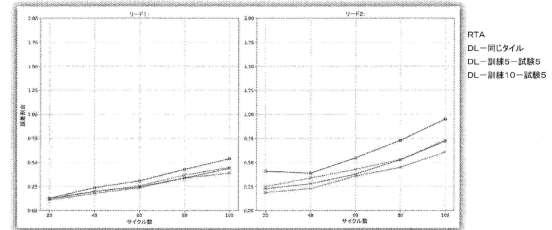


- NextSeq、2チャネル、ランダムフローセル
- 単一のモデル

【図 1 1 6】

深層学習を使用するベースコール

タイル対タイル一般化



- NextSeq、2チャネル、ランダムフローセル
- 単一のモデル

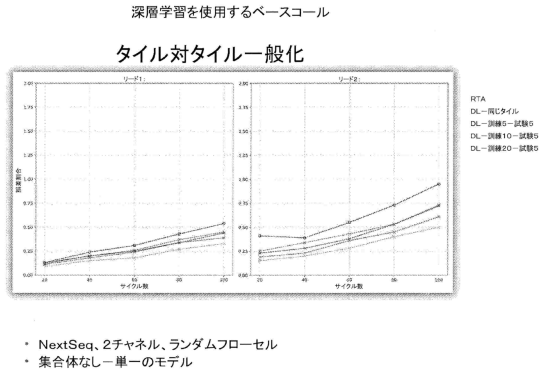
20

30

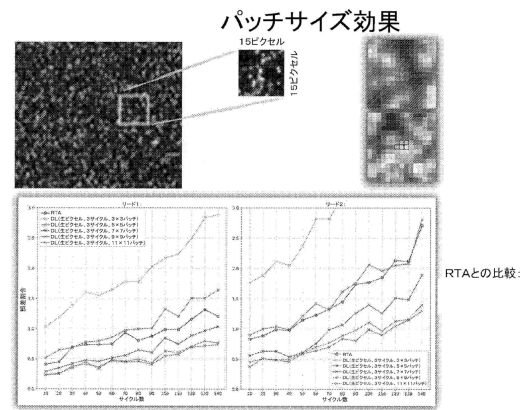
40

50

【図 1 1 7】

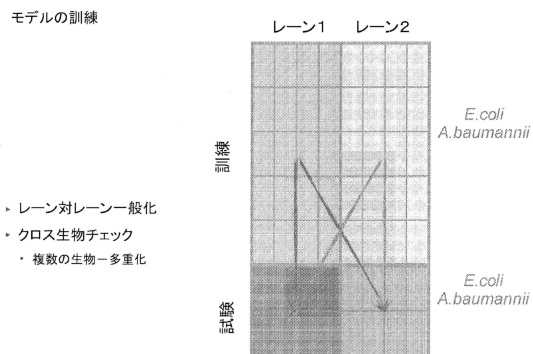


【図 1 1 8】

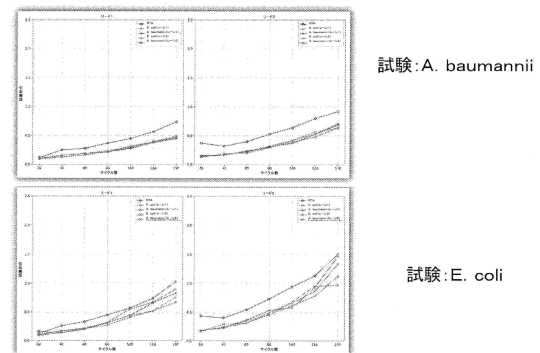


10

【図 1 1 9】



【図 1 2 0】



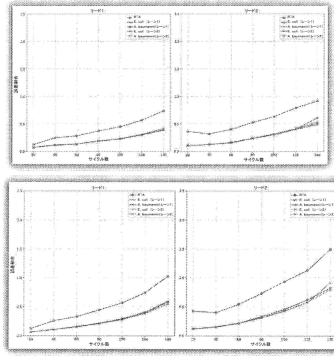
20

30

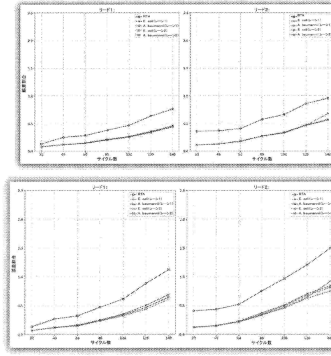
40

50

【図 1 2 1】

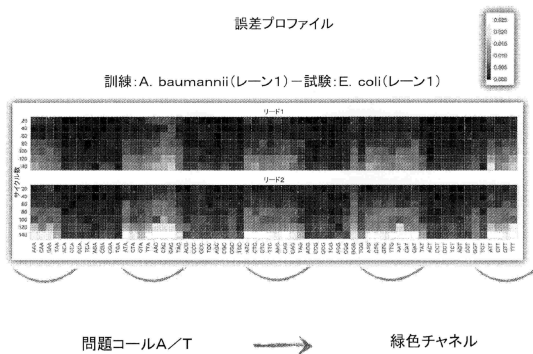
試験: A. baumannii
(レーン1)試験: E. coli
(レーン1)50~80%の
誤差低減

【図 1 2 2】

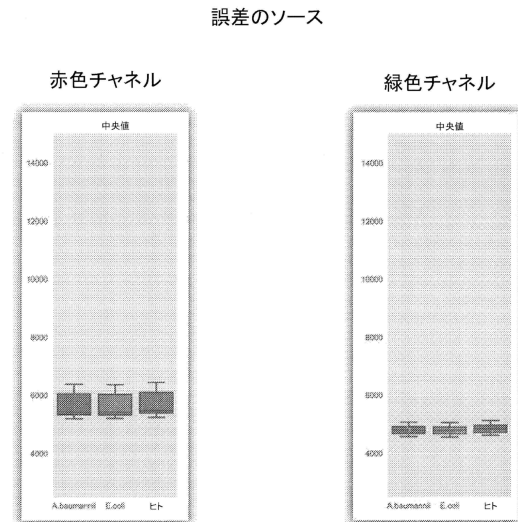
試験: A. baumannii
(レーン2)試験: E. coli
(レーン2)50~80%の
誤差低減

10

【図 1 2 3】



【図 1 2 4】



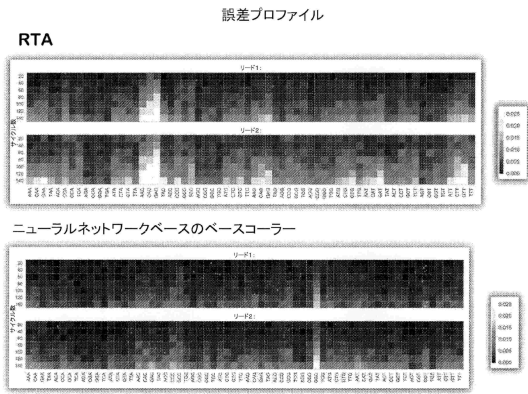
20

30

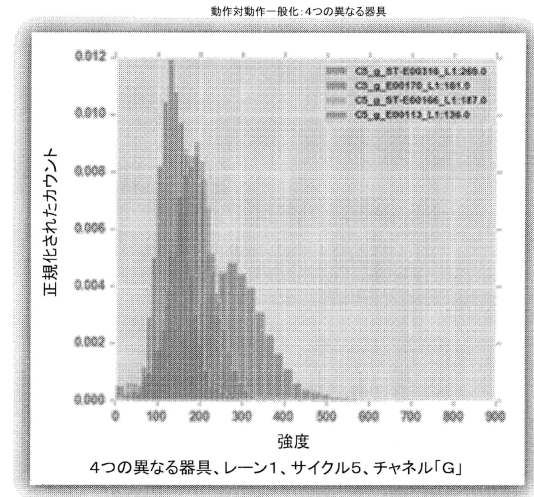
40

50

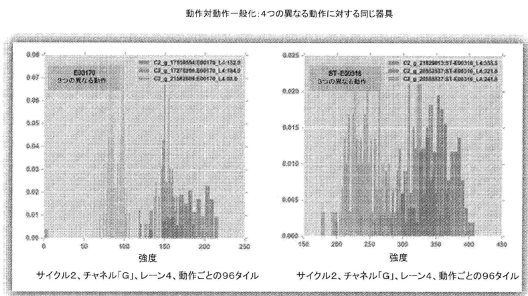
【図 1 2 5】



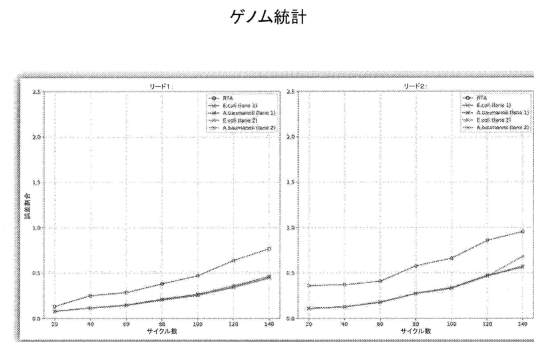
【図 1 2 6 a】



【図 1 2 6 b】



【図 1 2 7】



10

20

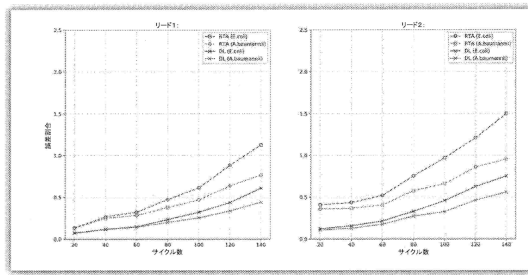
30

40

50

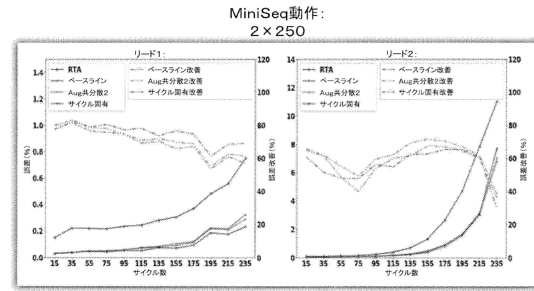
【図 1 2 8】

ゲノムコンテキスト



【図 1 2 9】

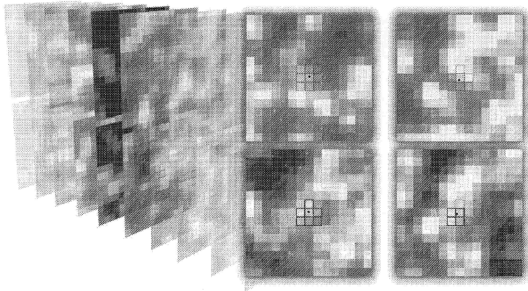
長いリード



10

【図 1 3 0】

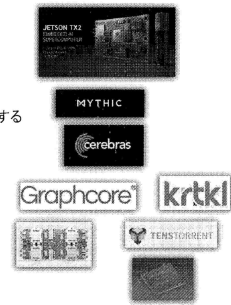
中心クラスターピクセル(複数可)及びその隣接するピクセルに注目する



【図 1 3 1】

ハードウェア

- 以下のような推測エンジン:
 - Jetson (Nvidia)
 - Mythic
- ソフトウェアベースの向上
 - FPGAを含む全てのプラットフォームと機能する
 - UCSDとの協働
- 器具上での利用可能なFPGA容量
- クラウドベースのソリューション
 - Firefly / Dragonfly
- メモリ



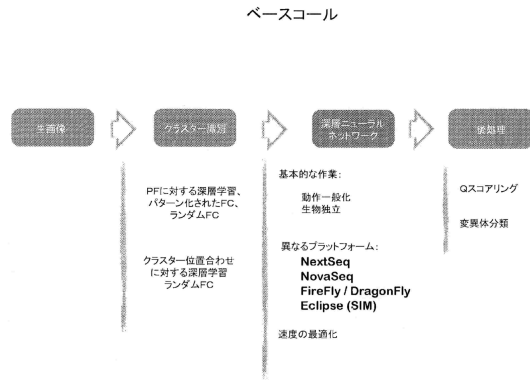
20

30

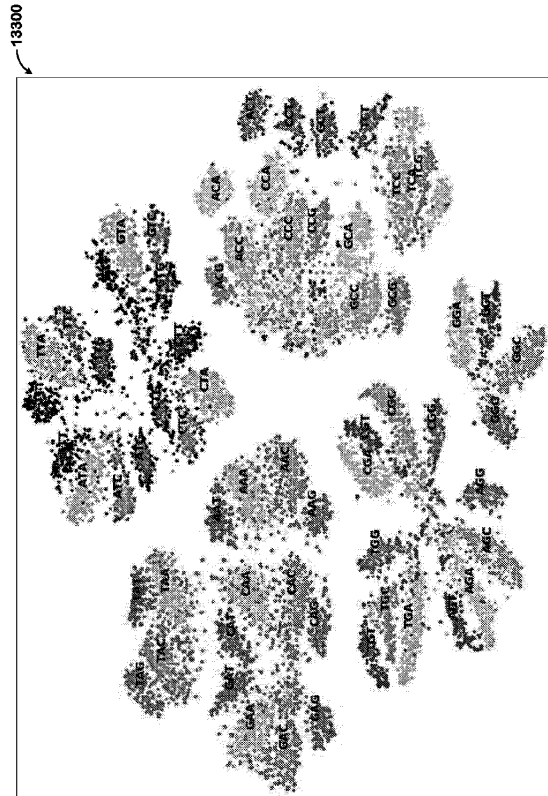
40

50

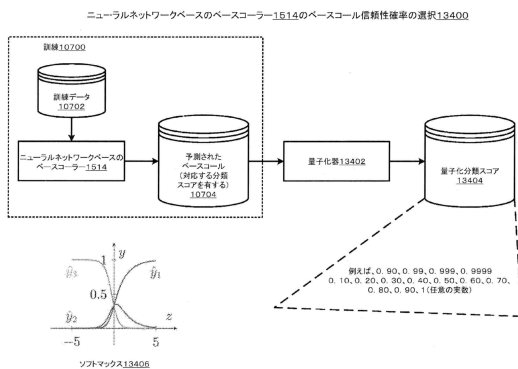
【図 1 3 2】



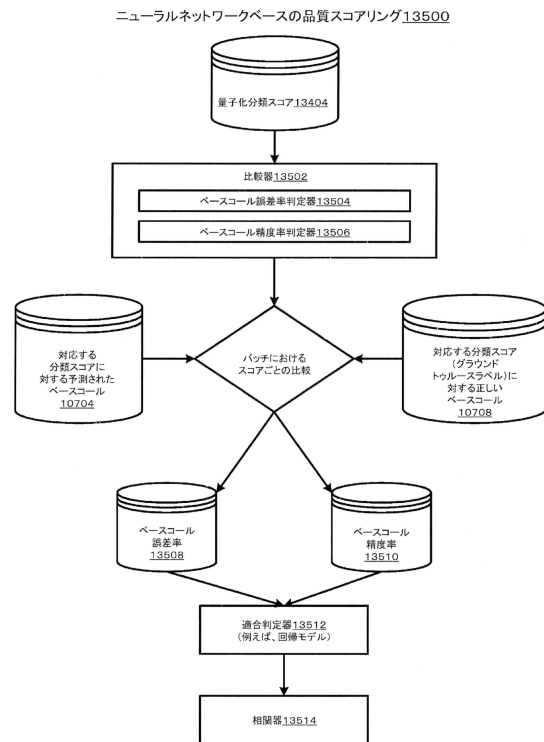
【図 1 3 3】



【図 1 3 4】



【図 1 3 5】



10

20

30

40

50

【図 1 3 6 a】

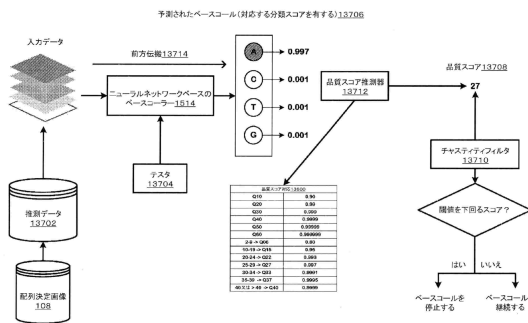
品質スコア対応スキーム13600a	
品質スコア	ソフトマックススコア
Q10	0.90
Q20	0.99
Q30	0.999
Q40	0.9999
Q50	0.99999
Q60	0.999999

【図 1 3 6 b】

品質スコア対応スキーム「J3600b」	
ピング品質スコア	ソフトマックススコア
Q2-Q9 → Q06	0.80
Q10-Q19 → Q15	0.95
Q20-Q24 → Q22	0.993
Q25-Q29 → Q27	0.997
Q30-Q34 → Q33	0.9991
Q35-Q39 → Q37	0.9995
Q40 又は Q40 → Q40	0.9999

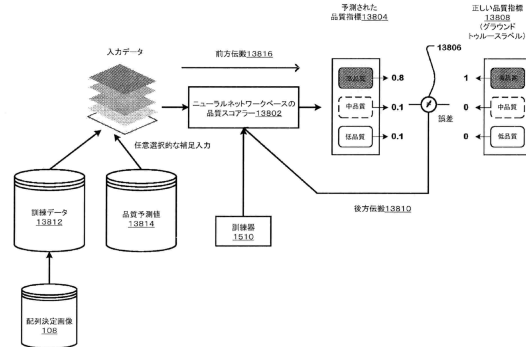
【図 137】

推測中の品質スコアリング13700



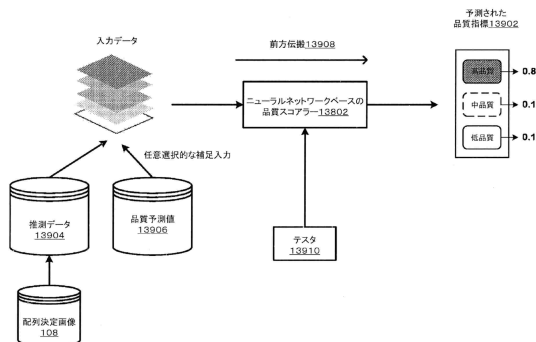
【圖 138】

ニューラルネットワークベースの品質スコアラー13802の訓練13800



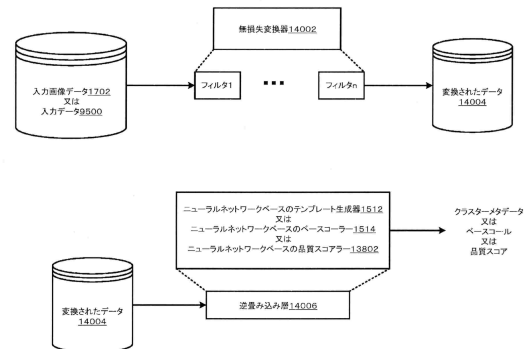
【図 139】

ニューラルネットワークベースの品質スコアラー13802を有する推測13900

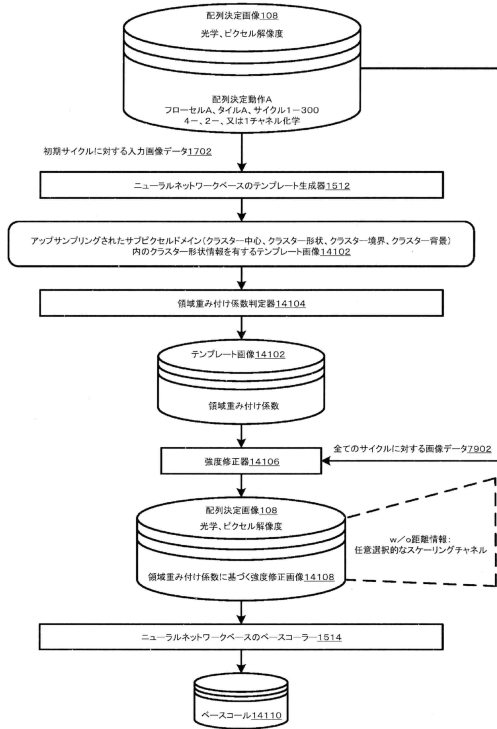


【図 140】

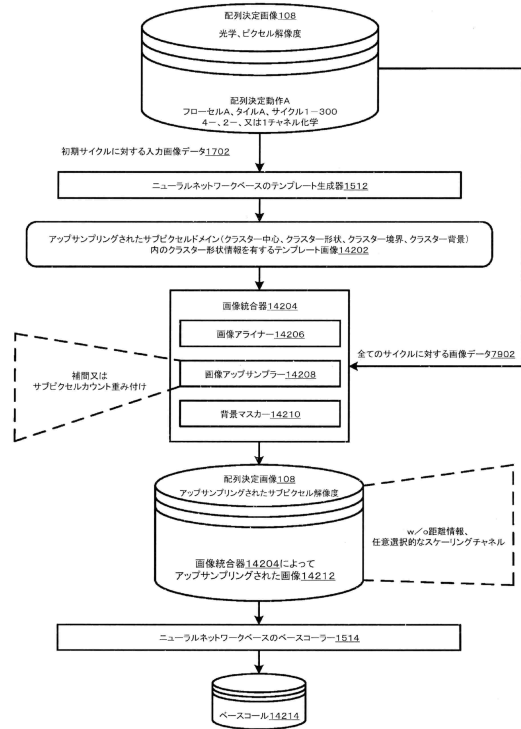
無損失変換14000



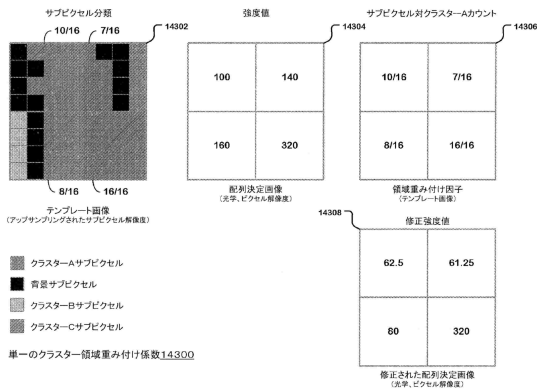
【図 1 4 1】



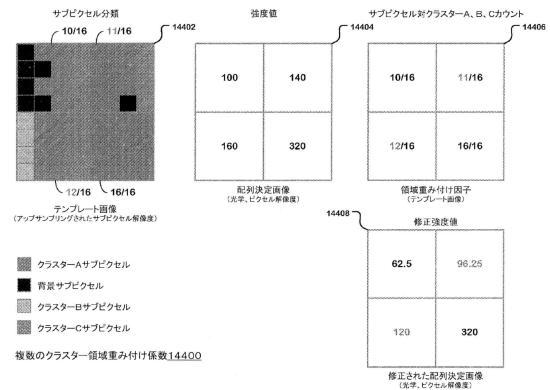
【図 1 4 2】



【図 1 4 3】



【図 1 4 4】



10

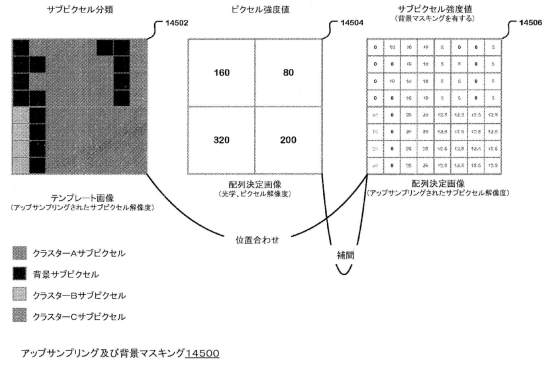
20

30

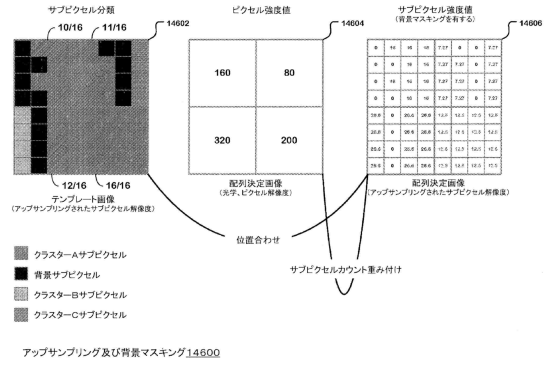
40

50

【図 145】

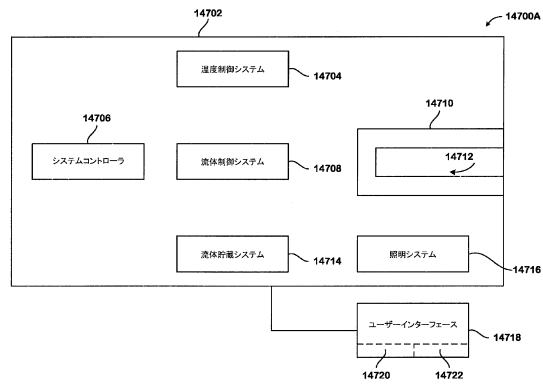


【図 146】

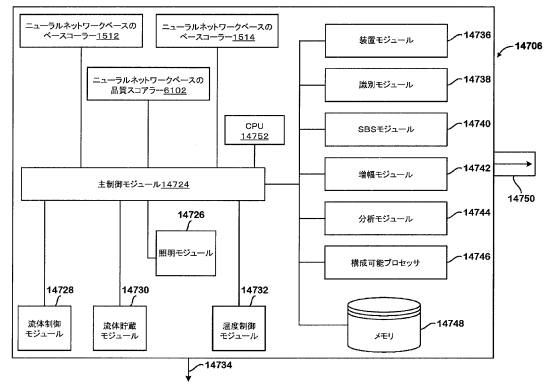


10

【図 147 A】



【図 147 B】



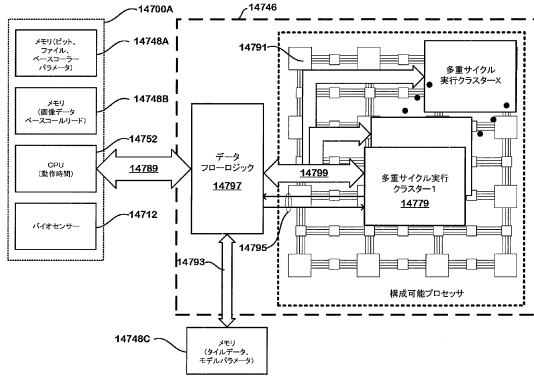
20

30

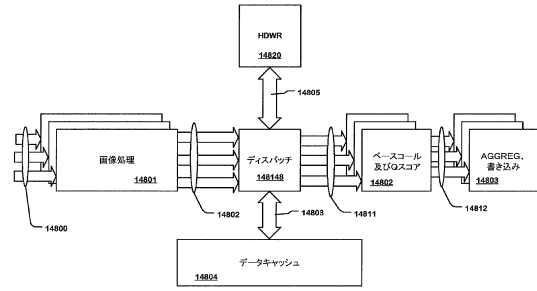
40

50

【図 147 C】

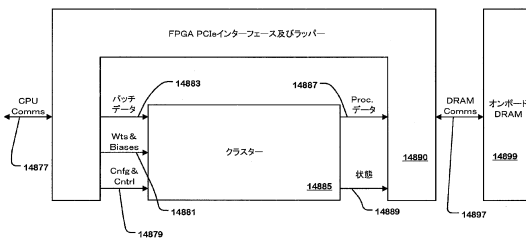


【図 148 A】

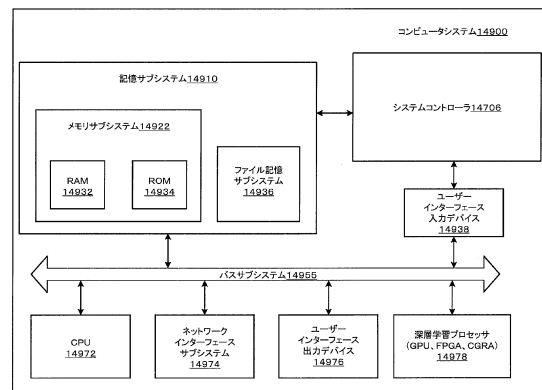


10

【図 148 B】



【図 149】



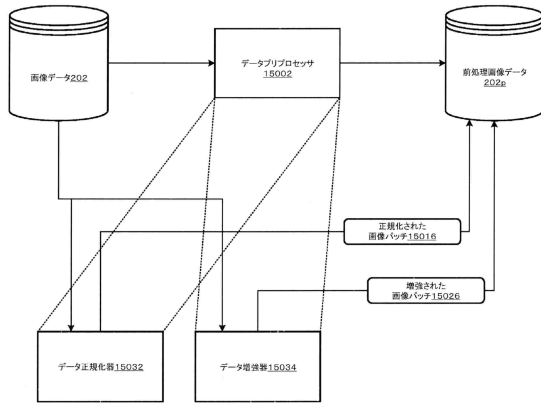
20

30

40

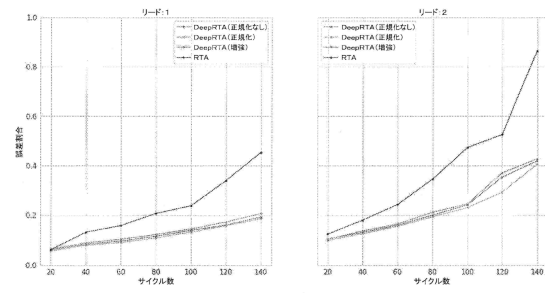
50

【図 150】



【図 151】

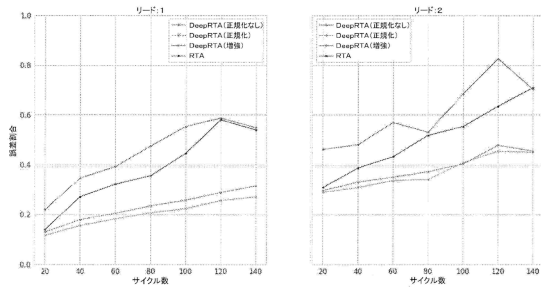
細菌で訓練、ヒトで試験(同じアッセイ)



10

【図 152】

アッセイ一般化—エクソーム配列決定



20

30

40

50

フロントページの続き

- (33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 62/821,681
(32)優先日 平成31年3月21日(2019.3.21)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 62/821,724
(32)優先日 平成31年3月21日(2019.3.21)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 62/821,766
(32)優先日 平成31年3月21日(2019.3.21)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 2023310
(32)優先日 令和1年6月14日(2019.6.14)
(33)優先権主張国・地域又は機関
オランダ(NL)
- (31)優先権主張番号 2023311
(32)優先日 令和1年6月14日(2019.6.14)
(33)優先権主張国・地域又は機関
オランダ(NL)
- (31)優先権主張番号 2023312
(32)優先日 令和1年6月14日(2019.6.14)
(33)優先権主張国・地域又は機関
オランダ(NL)
- (31)優先権主張番号 2023314
(32)優先日 令和1年6月14日(2019.6.14)
(33)優先権主張国・地域又は機関
オランダ(NL)
- (31)優先権主張番号 2023316
(32)優先日 令和1年6月14日(2019.6.14)
(33)優先権主張国・地域又は機関
オランダ(NL)
- (31)優先権主張番号 16/825,987
(32)優先日 令和2年3月20日(2020.3.20)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 16/825,991
(32)優先日 令和2年3月20日(2020.3.20)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 16/826,126
(32)優先日 令和2年3月20日(2020.3.20)
(33)優先権主張国・地域又は機関
米国(US)
- (31)優先権主張番号 16/826,134
(32)優先日 令和2年3月20日(2020.3.20)
(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 16/826,168

(32)優先日 令和2年3月21日(2020.3.21)

(33)優先権主張国・地域又は機関

米国(US)

(72)発明者 ドルナ・カシフハギギ

アメリカ合衆国・カリフォルニア・92122・サン・ディエゴ・イルミナ・ウェイ・5200

(72)発明者 アミラリ・キア

アメリカ合衆国・カリフォルニア・92122・サン・ディエゴ・イルミナ・ウェイ・5200

(72)発明者 キショール・ジャガナサン

アメリカ合衆国・カリフォルニア・92122・サン・ディエゴ・イルミナ・ウェイ・5200

(72)発明者 ジョン・ランドール・ゴベル

アメリカ合衆国・カリフォルニア・92122・サン・ディエゴ・イルミナ・ウェイ・5200

審査官 松田 芳子

(56)参考文献 米国特許出願公開第2012/0020537 (US, A1)

国際公開第2018/129314 (WO, A1)

(58)調査した分野 (Int.Cl., DB名)

C12Q 1/68

JSTPlus/JMEDPlus/JST7580 (JDreamIII)