

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6842395号  
(P6842395)

(45) 発行日 令和3年3月17日 (2021.3.17)

(24) 登録日 令和3年2月24日 (2021.2.24)

(51) Int. Cl.

F I

G 0 6 F 16/50 (2019.01)

G 0 6 F 16/50

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 3 5 0 C

請求項の数 10 外国語出願 (全 14 頁)

(21) 出願番号 特願2017-182561 (P2017-182561)  
 (22) 出願日 平成29年9月22日 (2017.9.22)  
 (65) 公開番号 特開2018-101406 (P2018-101406A)  
 (43) 公開日 平成30年6月28日 (2018.6.28)  
 審査請求日 令和2年7月8日 (2020.7.8)  
 (31) 優先権主張番号 16192142.4  
 (32) 優先日 平成28年10月4日 (2016.10.4)  
 (33) 優先権主張国・地域又は機関  
 欧州特許庁 (EP)

早期審査対象出願

(73) 特許権者 502208205  
 アクシス アーバー  
 スウェーデン国 2 2 3 6 9 ルンド,  
 グレンデン 1  
 (74) 代理人 110002077  
 園田・小林特許業務法人  
 (72) 発明者 ダニエルソン, ニクラス  
 スウェーデン国 2 2 3 6 9 ルンド,  
 エンダラヴェーゲン 1 4, シー/オ  
 ー アクシス コミュニケーションズ ア  
 ーバー

最終頁に続く

(54) 【発明の名称】 ニューラルネットワークにトレーニングデータを提供するための画像解析アルゴリズムの使用

(57) 【特許請求の範囲】

【請求項 1】

エンコーディングプロセス中に、画像にビットレートバジェットを分配するときにビデオカメラで使用される、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像内のマクロブロックの関連度を決定するためのデータを前記特定の入力用デジタル画像の各々について出力するニューラルネットワークをトレーニングするためのコンピュータプログラムであって、

第1の複数のデジタルトレーニング画像について、前記マクロブロック内の画像特性に基づく関連度で前記マクロブロックを自動的にラベル付けするように構成された第1の画像解析アルゴリズムを使用することによって、前記複数のデジタルトレーニング画像の各特定

10

のデジタルトレーニング画像内の前記マクロブロックの関連度を決定することと (S 4 0 2)、  
 ユーザが特定した状況を表示する第2の複数のデジタルトレーニング画像について、マクロブロックの画像特性に基づく関連度で前記マクロブロックを自動的にラベル付けするように構成された第2の画像解析アルゴリズムを使用することにより、前記第2の複数のデジタルトレーニング画像の各特定

のデジタルトレーニング画像内の前記マクロブロックの関連度を決定すること (S 4 0 4)、  
 によってトレーニングデータセットを決定することであって、

前記第2の画像解析アルゴリズムは、前記ユーザが特定した状況を表示する画像を解析するように特別に適合されている点で、第1の画像解析アルゴリズムとは異なり、

20

比較的高い関連度のマクロブロックは、より低い関連度を有するマクロブロックと比較して、エンコーディングプロセス中にそのマクロブロックに前記ビットレートバジェットのより大きい部分が費やされることをビデオカメラに示す、前記トレーニングデータセットを決定すること、

前記第 1 および第 2 の複数のデジタルトレーニング画像および前記関連するラベルを前記トレーニングデータセットとして含めること、および

前記ニューラルネットワークのトレーニング ( S 4 0 8 ) に前記トレーニングデータセットを使用すること、  
を含み、

前記第 1 の複数のデジタルトレーニング画像および前記関連する関連度は、前記ニューラルネットワークを最初にトレーニングするのに使用され、その後、前記第 2 の複数のデジタルトレーニング画像および前記関連する関連度は、前記ニューラルネットワークの前記トレーニングを前記ユーザが特定した状況に微調整するのに使用される、コンピュータプログラム。

【請求項 2】

前記第 1 の画像解析アルゴリズムが、

前記マクロブロックの空間的統計測定値を計算することと、

前記複数のデジタルトレーニング画像の画像シーケンスに時間的フィルタリングを適用することによって、前記マクロブロックの時間的統計測定値を計算することであって、前記画像シーケンスは前記特定のデジタルトレーニング画像を含む、前記マクロブロックの時間的統計測定値を計算することと、

前記マクロブロックの前記時間的統計測定値および前記空間的統計測定値を重み付けすることによって重み付け統計測定値を計算して、前記重み付け統計測定値に基づいて前記マクロブロックに関連度をラベル付けすることと

を行うことにより、前記マクロブロックに関連度を自動的にラベル付けするように構成される、請求項 1 に記載のコンピュータプログラム。

【請求項 3】

前記第 2 の複数のデジタルトレーニング画像が、前記第 1 の複数のデジタルトレーニング画像の一部である少なくとも 1 つのデジタルトレーニング画像を含む、請求項 1 または 2 に記載のコンピュータプログラム。

【請求項 4】

前記第 2 の複数のデジタルトレーニング画像が、前記第 1 の複数のデジタルトレーニング画像の一部ではない少なくとも 1 つのデジタルトレーニング画像を含む、請求項 1 から 3 のいずれか一項に記載のコンピュータプログラム。

【請求項 5】

前記トレーニングデータセットが、前記第 1 の複数のデジタルトレーニング画像の中の少なくとも 1 つのデジタルトレーニング画像のマクロブロックに関連度を手動でラベル付けすること ( S 4 0 6 ) によってさらに決定される、請求項 1 から 4 のいずれか一項に記載のコンピュータプログラム。

【請求項 6】

前記トレーニングデータセットが、前記第 2 の複数のデジタルトレーニング画像の中の少なくとも 1 つのデジタルトレーニング画像のマクロブロックに関連度を手動でラベル付けすることによってさらに決定される、請求項 1 から 5 のいずれか一項に記載のコンピュータプログラム。

【請求項 7】

請求項 1 から 6 のいずれか一項に記載のコンピュータプログラムを収容した、コンピュータ可読記憶媒体。

【請求項 8】

エンコーディングプロセス中に、画像にビットレートバジェットを分配するときにビデオカメラで使用される、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像

10

20

30

40

50

内のマクロブロックの関連度を決定するためのデータを前記特定の入力用デジタル画像の各々について出力するニューラルネットワークをトレーニングするように構成されたプロセッサを備えるコンピュータであって、前記ニューラルネットワークの前記トレーニングが、

第1の複数のデジタルトレーニング画像について、前記マクロブロック内の画像特性に基づく関連度で前記マクロブロックを自動的にラベル付けするように構成された第1の画像解析アルゴリズムを使用することによって、前記複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像内の前記マクロブロックの関連度を決定すること、

ユーザが特定した状況を表示する第2の複数のデジタルトレーニング画像について、マクロブロックの画像特性に基づく関連度で前記マクロブロックを自動的にラベル付けするように構成された第2の画像解析アルゴリズムを使用することにより、前記第2の複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像内の前記マクロブロックの関連度を決定すること（S404）、

とによってトレーニングデータセットを決定することであって、

前記第2の画像解析アルゴリズムは、前記ユーザが特定した状況を表示する画像を解析するように特別に適合されている点で、第1の画像解析アルゴリズムとは異なり、

比較的高い関連度のマクロブロックは、より低い関連度を有するマクロブロックと比較して、エンコーディングプロセス中にそのマクロブロックに前記ビットレートバジェットのより大きい部分が費やされることをビデオカメラに示す、前記トレーニングデータセットを決定すること、

前記第1および第2の複数のデジタルトレーニング画像および前記関連するラベルを前記トレーニングデータセットとして含めること、および

前記ニューラルネットワークのトレーニングに前記トレーニングデータセットを使用すること、

を含み、

前記第1の複数のデジタルトレーニング画像および前記関連する関連度は、前記ニューラルネットワークを最初にトレーニングするのに使用され、その後、前記第2の複数のデジタルトレーニング画像および前記関連する関連度は、前記ニューラルネットワークの前記トレーニングを前記ユーザが特定した状況に微調整するのに使用される、コンピュータ。

#### 【請求項9】

ビデオカメラと請求項8に記載のコンピュータとを備えるシステムであって、

前記プロセッサが、結果として得られたトレーニングされたニューラルネットワークを前記カメラに提供する（S410）ようにさらに構成され、

前記カメラが、

前記トレーニングされたニューラルネットワークを受信し、

複数のデジタル画像を取り込み、前記複数のデジタル画像の画像データを前記トレーニングされたニューラルネットワークに供給し、前記複数のデジタル画像の各特定のデジタル画像について、前記特定のデジタル画像のマクロブロックに各々対応する複数の関連度を取得するように構成された、システム。

#### 【請求項10】

前記カメラが、

前記取得された関連度に従って前記特定のデジタル画像の圧縮度を制御することによって前記特定のデジタル画像をエンコードするようにさらに構成された、請求項9に記載のシステム。

#### 【発明の詳細な説明】

#### 【技術分野】

#### 【0001】

本発明は一般に、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像内のピクセル群の関連度を決定するためのデータを特定の入力用デジタル画像の各々について

10

20

30

40

50

出力するニューラルネットワークをトレーニングするための方法、コンピュータプログラム、コンピュータ、およびシステムに関する。

【背景技術】

【0002】

画像処理では、多くのアプリケーションは、さまざまな画像領域に対してリソースの割り当てを変えることが必要となる。たとえば、圧縮パラメータは、画像のある種の特性に基づいて選択することが可能であり、或いは、画像領域は、送信信頼性と効率等との間のトレードオフを最適なものにするために、さまざまな程度のエラー修正を受けることがあり得る。特定の画像領域に割り当てられるべきリソースの量を決定するための画像領域の関連度の自動識別は、重要な問題であり、そのようなアルゴリズムを実行することはまた、貴重なCPU時間を必要とする場合もある。これは、たとえば多数のプログラムが、監視カメラのような組み込みプラットフォーム上で限られたリソースをめぐって競合する場合には、問題を生じる可能性がある。

10

【0003】

したがって、このような状況において改善の必要がある。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】欧州特許第3021583号明細書

【発明の概要】

20

【発明が解決しようとする課題】

【0005】

上記のことを考慮して、本発明の目的は、上記で説明されている欠点の1つまたは複数、を、解決するかまたは少なくとも低減することである。概して、上記の目的は、添付の独立請求項によって達成される。

【課題を解決するための手段】

【0006】

第1の態様によれば、本発明は、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像内のピクセル群の関連度を決定するためのデータを特定の入力用デジタル画像の各々について出力するニューラルネットワークをトレーニングするための方法により実現される。方法は、複数のデジタルトレーニング画像について、ピクセル群に関連度を自動的にラベル付けするように構成された第1の画像解析アルゴリズムを使用することにより複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像におけるピクセル群の関連度を決定することによって、トレーニングデータセットを決定することと、複数のデジタルトレーニング画像およびラベルをトレーニングデータセットとして含めると、ニューラルネットワークのトレーニングのためにトレーニングデータセットを使用することを含む。

30

【0007】

「ニューラルネットワーク」という用語は、本明細書のコンテキストでは、脳内のニューロンの巨大なネットワークと類似した、相互接続されたノード群として理解されたい。ニューラルネットワークはまた、「人工ニューラルネットワーク」(ANN)と称されることもある。このコンテキストで使用され得るニューラルネットワークの特定のタイプは、畳み込みニューラルネットワーク(CNN)である。「ディープラーニング」という用語もまた一般に使用される。

40

【0008】

「関連度」という用語は、本明細書のコンテキストでは、画像内の対応する領域の画像コンテンツの(たとえばユーザにとっての)関心度の格付けとして理解されたい。言い換えるならば、関連度は、画像内の領域(ピクセル群)に対応し、関連度は、その領域の画像コンテンツが画像の閲覧者にとってどの程度関心を引くものであるかを規定する。これは有利なことに、画像のビットレートバジェットを分配するために使用されてもよい。たと

50

えば、関連度は、圧縮レベルの設定を通じて、対応する領域の圧縮度を制御するため、またはエンコーダがその領域のコーディングをどの程度積極的にスキップすべきかを決定するために使用されてもよい。関連度はさらに、たとえばエンコードされた画像を送信する際のエラー修正の度合いを決定するための、エラーコーディングの目的で使用されてもよい。画像の特定の領域の比較的高い関連度は、ビットレートバジェットのより大きい部分が、関連度のより低い別の画像領域と比較して、その特定の領域に費やされ得ることを意味する。

#### 【0009】

従来技術において関連度に代わって使用されることもあるその他の用語として、「関心領域」すなわち「ROI」を挙げることができる。

10

#### 【0010】

「ピクセル群に関連度を自動的にラベル付けするように構成された画像解析アルゴリズム」という表現は、本明細書のコンテキストでは、その部分の画像特性（つまり画像解析）に基づいて画像の対応する領域（たとえばマクロブロック）の画像コンテンツの（たとえばユーザにとっての）関心度を自動的に（たとえば、コンピュータを用いて、人間の補助なしに）格付けするために使用され得る任意の適切なアルゴリズムとして理解されたい。画像特性は、たとえばデジタル画像の一部に存在するエッジ、ノイズの存在、繰り返しテクスチャ加工された部分、顔認識、オブジェクト検出などのように、空間的であってもよい。画像特性はまた、たとえば動き検出のように、時間的なものであってもよい。適切な画像解析アルゴリズムの例は、Axis Communicationにより提供されるZipsreamアルゴリズムである。このアルゴリズムの一部は、たとえば欧州特許第3021583号明細書において説明されている。

20

#### 【0011】

本発明者らは、上記で説明されている既存の画像解析アルゴリズムを使用してトレーニングされるニューラルネットワークを採用することがいくつかの利点を有することに注目した。たとえば、ニューラルネットワークは極めて反復的ではあるが、極めて単純な膨大な量の計算によって機能するため、ニューラルネットワークの設計は、極めて単純で、ハードウェアの実装に十分に順応することができる。さらに、モデル圧縮の分野で現在進行中の研究は、監視カメラのような組み込みプラットフォーム上に置くことが徐々に実現可能になりつつあるニューラルネットワークを利用する。このように、監視カメラのような組み込みプラットフォームでニューラルネットワークを使用することは、カメラで実際の画像解析アルゴリズムを実行することよりも有利となり得る。

30

#### 【0012】

一部の実施形態によれば、第1の画像解析アルゴリズムは、ピクセル群の空間的統計測定値を計算することと、複数のデジタルトレーニング画像の画像シーケンスに時間フィルタリングを適用することによって、ピクセル群の時間的統計測定値を計算することによって、画像シーケンスは特定のデジタルトレーニング画像を含む、ピクセル群の時間的統計測定値を計算することと、を行うことにより、ピクセル群に関連度を自動的にラベル付けするように構成される。

#### 【0013】

次いで、2つの統計測定値は、ピクセル群の時間的統計測定値および空間的統計測定値を重み付けすることによって重み付け統計測定値を計算して、重み付け統計測定値に基づいてピクセル群に関連度をラベル付けするために使用され得る。使用される重みは、たとえば0.5、0.3、0.7のような静的な重みであってもよく、または関与する画像領域の画像特性に応じて異なってもよい。

40

#### 【0014】

一部の実施形態によれば、トレーニングデータセットを決定することは、第2の複数のデジタルトレーニング画像について、ピクセル群に関連度を自動的にラベル付けするように構成された第2の画像解析アルゴリズムを使用することにより、第2の複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像におけるピクセル群の関連度を決

50

定することを含み、第2の画像解析アルゴリズムは第1の画像解析アルゴリズムとは異なる。

【0015】

たとえば、(より少ない)複数の画像について、より複雑または特殊化された画像解析アルゴリズムが、たとえばデジタル画像の特定のコンテンツに関するトレーニングデータをニューラルネットワークに提供するために採用されてもよい。そのような特定のコンテンツの例は、交通量を示す画像コンテンツか、または多くの人々がカメラのそばを通り過ぎるエントランスを示す画像コンテンツなどであってもよい。

【0016】

本発明の実施形態の利点は、ニューラルネットワークが、正しくトレーニングされる場合、第1の画像解析アルゴリズムのような手作りのアルゴリズムと比較して、はるかに優れたパフォーマンスを有する可能性があることである。本発明の実施形態の利点は、膨大な量のトレーニングデータが、第1の画像解析アルゴリズムとの間で生成されて、そのアルゴリズムと同程度で実行するためのアルゴリズムを迅速に獲得できることである。その後、ニューラルネットワークは、第2の複数の画像およびそのレベル(第2の画像解析アルゴリズムによって決定される)のような、その他のより「高度な」トレーニングデータで微調整されてもよい。また、微調整が必要とするデータは通常、初期トレーニング(つまり、第1の複数の画像を使用すること)よりもはるかに少なくてすむ。言い換えるならば、第2のトレーニングを行う目的は、ユーザ固有の関連度を提供することである。画像のある区域は、一部のユーザにとっては重要であるが、その他のタイプの使用事例では関心を引くものではない。この第2のトレーニング(第2の複数の画像および第2の画像解析アルゴリズムを使用する)を追加することで、ビットレートバジェットはユーザの関連分野で容易に費やすことができる。

【0017】

第2の複数のデジタルトレーニング画像は、第1の複数のデジタルトレーニング画像の一部である少なくとも1つのデジタルトレーニング画像を備えることができる。第2の複数のデジタルトレーニング画像は、たとえば、第1の複数のデジタルトレーニング画像のサブセットで構成されてもよい。第2の複数のデジタルトレーニング画像はまた、第1の複数のデジタルトレーニング画像の一部ではない少なくとも1つのデジタルトレーニング画像を備えることができる、つまり第2の複数のデジタルトレーニング画像は第1の複数のデジタルトレーニング画像のサブセットではない。

【0018】

一部の実施形態によれば、トレーニングデータセットは、第1の複数のデジタルトレーニング画像の中の少なくとも1つのデジタルトレーニング画像のピクセル群に関連度を手動でラベル付けすることによってさらに決定される。その他の実施形態によれば、トレーニングデータセットは、第2の複数のデジタルトレーニング画像の中の少なくとも1つのデジタルトレーニング画像のピクセル群に関連度を手動でラベル付けすることによってさらに決定される。トレーニングデータセットの画像の中のピクセル群に手動でラベル付けすることは、トレーニングデータセットを微調整するために使用されてもよい。第2の複数の画像および第2の画像解析アルゴリズムの使用と併せて、上記で説明されているものと同じ論拠を使用すると、この実施形態では、ビットレートバジェットはユーザの関連分野で容易に費やすことができる。

【0019】

第2の態様において、本発明は、処理機能を有するデバイスによって実行されるとき、第1の態様の方法を遂行するように適合された命令を伴うコンピュータ可読記憶媒体を備えるコンピュータプログラム製品を提供する。

【0020】

第3の態様によれば、本発明は、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像内のピクセル群の関連度を決定するためのデータを特定の入力用デジタル画像の各々について出力するニューラルネットワークをトレーニングするように構成されたブ

ロセッサを備えるコンピュータを提供する。ニューラルネットワークのトレーニングは、複数のデジタルトレーニング画像について、ピクセル群に関連度を自動的にラベル付けするように構成された第1の画像解析アルゴリズムを使用することにより複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像におけるピクセル群の関連度を決定することによって、トレーニングデータセットを決定することと、複数のデジタルトレーニング画像およびラベルをトレーニングデータセットとして含めることと、ニューラルネットワークのトレーニングのためにトレーニングデータセットを使用することを含む。

【0021】

第4の態様において、本発明は、ビデオカメラおよびコンピュータを備えるシステムを提供する。コンピュータは、複数の入力用デジタル画像を受信し、特定の入力用デジタル画像内のピクセル群の関連度を決定するためのデータを特定の入力用デジタル画像の各々について出力するニューラルネットワークをトレーニングするように構成されたプロセッサを有し、ニューラルネットワークのトレーニングは、複数のデジタルトレーニング画像について、ピクセル群に関連度を自動的にラベル付けするように構成された第1の画像解析アルゴリズムを使用することにより複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像におけるピクセル群の関連度を決定することによって、トレーニングデータセットを決定することと、複数のデジタルトレーニング画像およびラベルをトレーニングデータセットとして含めることと、ニューラルネットワークのトレーニングのためにトレーニングデータセットを使用することを含む。プロセッサは、結果として得られたトレーニングされたニューラルネットワークをカメラに提供するようにさらに構成され、カメラは、トレーニングされたニューラルネットワークを受信するように構築される。カメラは、複数のデジタル画像を取り込み、複数のデジタル画像の画像データをトレーニングされたニューラルネットワークに提供し、複数のデジタル画像の各特定のデジタル画像について、特定のデジタル画像のピクセル群に各々対応する複数の関連度を取得するように、さらに構築される。

【0022】

「トレーニングされたニューラルネットワーク」という用語は、本明細書のコンテキストでは、たとえば、ニューラルネットワーク内のノードのセットアップ（ノードが相互接続される方法）、およびニューラルネットワークにおいて使用される重み（フィルタ、重みバンクなど）を指定するテキストファイルとして理解されたい。たとえばバイナリファイルを介するなど、コンピュータとカメラとの間でトレーニングされたニューラルネットワークを提供するその他の方法が、同様に可能である。

【0023】

一部の実施形態によれば、カメラは、取得された関連度に従って特定のデジタル画像の圧縮度を制御することによって特定のデジタル画像をエンコードするようにさらに構築される。

【0024】

圧縮度は、画像の特定の部分（領域、区域など）について圧縮比を設定することによって制御されてもよい。圧縮比は、たとえば0～51（H.264）、1～255（VP9）、または1～100の範囲の、量子化パラメータ値、QP値によって具現化されてもよい。この説明において、「圧縮度」、「圧縮比」、および「圧縮レベル」という用語は、広い意味で、元の、圧縮解除された、ピクセルの表現よりも少ないビットを使用するピクセルのエンコーディングを示すために同義的に使用されることに留意されたい。上記で説明されているように、エンコーディングプロセスのその他の部分は、たとえば関連度に基づいてエンコードされた画像を送信する際のエラー修正の度合いを決定するために、エラーコーディングのような、関連度に基づいてもよい。

【0025】

第2、第3、および第4の態様は概して、第1の態様と同じ特徴および利点を有することができる。

【0026】

本発明の上記、ならびに追加の目的、特徴、および利点は、同じ参照番号が類似する要素について使用される添付の図面を参照して、本発明の実施形態の以下の例示的かつ非限定的な詳細な説明を通じてさらに深く理解されるであろう。

【図面の簡単な説明】

【 0 0 2 7 】

【図 1】 畳み込みニューラルネットワークを概略的に示す図である。

【図 2】 実施形態によるニューラルネットワークがトレーニングされる方法を示す図である。

【図 3】 実施形態によるビデオカメラおよびコンピュータを備えるシステムを示す図である。

【図 4】 実施形態によるニューラルネットワークをトレーニングする方法を示す図である。

【発明を実施するための形態】

【 0 0 2 8 】

図 1 は、一例として畳み込みニューラルネットワーク (CNN) 100 を示す。この明細書のコンテキストでは、入力用デジタル画像におけるピクセル群の関連度を決定するための CNN の使用は単に一例に過ぎないことに留意されたい。たとえば別のタイプのフィードフォワードニューラルネットワーク (つまり CNN 以外)、再帰型ネットワークなど、任意の適切なニューラルネットワークアーキテクチャが、この目的で採用されてもよい。CNN 100 において、畳み込みレイヤ 102、および完全接続レイヤ 104 という 2 つのタイプのレイヤが存在することができる。完全接続レイヤは、2 つ以上の出力を有するレイヤで置き換えられてもよいことが想定される。CNN 100 において、任意の数の 2 つの異なるレイヤが存在することができる。図 1 において、2 つの畳み込みレイヤ 102 a、102 b が存在する。畳み込みレイヤ 102 には、2 つの回帰ヘッド 104 が追加される。回帰ヘッド 104 は、この例において、畳み込みレイヤ 102 から出力 106 を供給される完全接続レイヤ 104 a ~ b、104 c ~ d のセットである。このコンテキストにおける CNN の利点は、(画像解析の目的に適合されたフィルタおよびフィルタ重みで構成された) 畳み込みレイヤ 102 の同じセットが、関連度を決定する目的のために使用され得ることであってもよい。このセット 102 には、追加の回帰ヘッド 104 が追加されてもよい (畳み込みレイヤ 102 のセットはその他の画像解析の目的でトレーニングされたその他の回帰ヘッドにすでに接続されている)。次いで、追加の回帰ヘッドは、畳み込みレイヤ 102 からの出力データ 106 を使用して関連度を決定するためにトレーニングされてもよい。回帰ヘッド 104 の計算複雑性は、畳み込みレイヤ 102 のセットと比較すると低い場合が多く、そのためニューラルネットワーク 100 を実行するハードウェアに追加の要求を加えることはほとんどない。

【 0 0 2 9 】

図 2 は、ニューラルネットワーク 100 をトレーニングするように構築されたデバイス 200 を一例として示す。トレーニングの実施形態は、これ以降、図 2 および図 4 と併せて説明される。デバイス 200 は、処理機能を有する任意のデバイスであってもよく、これ以降、コンピュータ 200 と称される。コンピュータ 200 は、最初に複数のデジタルトレーニング画像 208 を受信することによって、トレーニングデータセット 212 を決定することができる。次いで、デジタルトレーニング画像 208 は、デジタルトレーニング画像 208 の各々におけるピクセル群に対する関連度を決定し S 402、自動的にラベル付けするように構成される第 1 の画像解析アルゴリズム 204 に供給されてもよい。第 1 の画像解析アルゴリズム 204 は、一部の実施形態に従ってデジタルトレーニング画像の空間的統計測定値に重点を置くことができる。この場合、各デジタルトレーニング画像は、第 1 の画像解析アルゴリズム 204 によって個別に処理されてもよい。通常の空間的統計測定値は、エッジ検出、色分布、顔検出、および空間ノイズ検出であってもよい。代替的に、または追加的に、第 1 の画像解析アルゴリズム 204 はまた、デジタルトレーニング画像のシーケンスのオブジェクトの動きのような、デジタルトレーニング画像の時間

10

20

30

40

50



的特徴を考慮に入れることもできる。この理由から、第1の画像解析アルゴリズムは、複数のデジタルトレーニング画像の画像シーケンスに時間的フィルタリングを適用することによって、各特定のデジタルトレーニング画像におけるピクセル群の時間的統計測定値を計算するように構成されてもよく、画像シーケンスは特定のデジタルトレーニング画像を含む。この場合、第1の画像解析アルゴリズム204は、アルゴリズムの1つの実行において、つまり特定のデジタルトレーニング画像における特定のピクセル群の時間的統計測定値を計算する際に、複数のデジタルトレーニング画像を考慮する必要がある。時間的統計測定値の例は、動き検出、時間ノイズ検出、画像間の照明条件の変化の検出などである。

【0030】

10

空間的統計測定値および時間的統計測定値の両方がピクセルの特定群について計算される場合、第1の画像解析アルゴリズム204は、ピクセル群の時間的統計測定値および空間的統計測定値を重み付けすることによって重み付け統計測定値を計算し、重み付け統計測定値に基づいてピクセル群に関連度をラベル付けするように構成されてもよい。適用される重みは、0から1の任意の値であってもよく（時間的統計測定値および空間的統計測定値が正規化される場合、それ以外の場合はその他の範囲の重みが適用する）、第1の画像解析アルゴリズム204の対象分野に応じて決まる。

【0031】

たとえば、第1の画像解析アルゴリズム204が、移動する人物を含む画像の部分にビットレートバジェットを費やすように適合される場合、アルゴリズムは、画像のそのような部分に高い関連度をラベル付けすべきである。第1の画像解析アルゴリズムは、この実施形態において、画像内のピクセル群が移動する人物に対応するかどうかを決定するために、対応する空間的および時間的統計測定値を計算する必要がある。もう1つの例は、第1の画像解析アルゴリズム204が街路を細部にわたり表示する保存区域に重点を置くように構築される（つまりビットレートバジェットがそのような区域に費やされる）ものであってもよく、この場合、計算される必要があるのは空間的統計測定値のみであってもよい。

20

【0032】

要約すると、第1の画像解析アルゴリズム204は、各デジタルトレーニング画像208の特定のピクセル群に関連度を自動的にラベル付けするように構成される。この目的で第1の画像解析アルゴリズム204のみが使用される実施形態について、複数のデジタルトレーニング画像208および計算されたラベルは、トレーニングデータセット212を形成する。

30

【0033】

一部の実施形態によれば、トレーニングデータセット212を決定することは、第2の複数のデジタルトレーニング画像を第2の画像解析アルゴリズム206を通じて実行することを含む。

【0034】

第2の複数のデジタルトレーニング画像は、第1の複数のデジタルトレーニング画像208のサブセットであってもよい。

40

【0035】

あるいは、第2の複数のデジタルトレーニング画像は、第1の複数のデジタルトレーニング画像208のサブセットではない。

【0036】

たとえば、第2の複数のデジタルトレーニング画像は、たとえば交通状況、カジノ、小売店、銀行、航空機監視など、ニューラルネットワークのアプリケーションの特定の範囲を対象としてもよい。この理由により、これらの特定の状況も対象とすることができる第2の画像解析アルゴリズム206を使用してそのような状況を表示する画像210を解析することは有利となり得る。すなわち、一部の実施形態によれば、第1の画像解析アルゴリズム204は、各種の画像コンテンツで良好に機能するより汎用のアルゴリズムであっ

50

てもよく、第2の画像解析アルゴリズム206は、特定の画像コンテンツを対象とする。一部の実施形態によれば、第1の画像解析アルゴリズム204は、計算複雑性がさほど高くはなく、そのため膨大な画像（たとえば>10000または>1000000）で実行することが可能であってもよく、第2の画像解析アルゴリズム206は、計算複雑性がより高くてもよい。

#### 【0037】

したがって、コンピュータ200は、第2の複数のデジタルトレーニング画像について、ピクセル群に関連度を自動的にラベル付けするように構成された第2の画像解析アルゴリズムを使用することにより、第2の複数のデジタルトレーニング画像の各特定のデジタルトレーニング画像におけるピクセル群の関連度を決定するS404ように構築されてもよく、第2の画像解析アルゴリズムは第1の画像解析アルゴリズムとは異なる。上記で説明されているように、第1の画像解析アルゴリズムは、あらゆるタイプの入力画像で等しく良好に機能する汎用アルゴリズムであってもよく、第2の画像解析アルゴリズムは、よりプロファイルベースの解析を提供する（つまり一部の特定のコンテンツを備えるかまたは特定の状況を表示する画像を対象とする）ことができる。

#### 【0038】

一部の実施形態によれば、トレーニングデータセット212は、第1の複数のデジタルトレーニング画像の中の少なくとも1つのデジタルトレーニング画像のピクセル群に関連度を手動で（人物214により）ラベル付けすることによってさらに決定されるS406。代替的に、または追加的に、トレーニングデータセット212は、第2の複数のデジタルトレーニング画像の中の少なくとも1つのデジタルトレーニング画像のピクセル群に関連度を手動でラベル付けすることによってさらに決定される。この実施形態は、第1および/または第2の画像解析アルゴリズム204、206が関連度を正しくラベル付けできなかったデジタルトレーニング画像の一部の特定の特徴に対して、トレーニングデータセットが微調整される必要がある場合に有利となりうる。トレーニングデータセット212は、一部の実施形態により、少なくとも1つのさらなるデジタルトレーニング画像（第1/第2の複数のデジタルトレーニング画像の一部ではない）のピクセル群に関連度を手動でラベル付けすることによってさらに決定されてもよく、少なくとも1つのさらなるデジタルトレーニング画像および対応するラベルをトレーニングデータセットに含む。この実施形態は、図2には含まれていない。

#### 【0039】

したがって、トレーニングデータセット212は、第1および/または第2の画像解析アルゴリズム204、206によって解析された複数のデジタルトレーニング画像、および場合によっては人物214によって微調整されたラベルを備える。

#### 【0040】

次いで、トレーニングデータセット212は、ニューラルネットワーク100をトレーニングするS408のために使用される。トレーニングデータセット212を使用してニューラルネットワーク100をトレーニングすることは、当業者に実施が任せられ、使用されるニューラルネットワーク100のアーキテクチャに従って行なわれる。一般に、ニューラルネットワークは、各入力用デジタル画像の関連度のマップを、トレーニングデジタル画像のピクセル群に対するグラウンドトゥールズとして損失関数およびラベルを使用して、その内部重みを調整することによって提供するようにトレーニングされる。損失関数は、たとえば、平均二乗誤差の最小化をもたらす、L2ノルムに基づいてもよい。損失関数は、たとえば、可能な関連度の数、入力データのサイズなどに基づいて規定されてもよい。損失関数が規定されている場合、重みは通常、たとえばオプティマイザの確率的勾配降下法アルゴリズムと共に標準化誤差逆伝播法を通じて更新される。当業者に知られているように、確率的勾配降下法は、トレーニング中の入力画像に基づいて望ましい出力を生成することに可能な限り近いモデルを取得するように、ニューラルネットワークの重みをいかにして更新するかを計算するための最もよく知られている方法である。

#### 【0041】

第2の複数のトレーニング画像がトレーニングに使用され、第1の複数のトレーニング画像のサブセットである場合、第2の複数のデジタルトレーニング画像およびその対応するラベルは通常、ニューラルネットワークが第1の複数のデジタルトレーニング画像でトレーニングされた後にニューラルネットワークのトレーニングに使用されることに留意されたい。

【0042】

たとえば、トレーニング中に何万もの画像がニューラルネットワーク100に入力され、トレーニングデジタル画像の一部ではないデジタル画像上で、ただしはるかに迅速かつ計算複雑性の低い方法で、第1および/または第2の画像解析アルゴリズム204、206に相当する(またはより良い)結果を生成することができるまで、ニューラルネットワーク100の内部重みが調整される。

10

【0043】

ニューラルネットワークの設計により、有利な使用領域は、特定用途向集積回路(ASIC)のような、ニューラルネットワークを実行するように特に設計されたハードウェアを備えるビデオカメラなどの低い使用可能計算リソース(たとえば、第1/第2の画像アルゴリズムを実行するのに十分ではない)を有し得るデバイスにある。そのため、デバイス200は、そのようなデバイス304に、トレーニングされたニューラルネットワークを、たとえば無線ネットワーク302を介して提供するように構築されてもよく、それによってデバイスはトレーニングされたニューラルネットワークで構成されるS410。これは、図3に示される。トレーニングされたニューラルネットワークが、内部フィルタおよびトレーニングされたニューラルネットワークの重みを備えるテキストファイルのような、任意の適切な方法で配布されてもよいことに留意されたい。また、コンピュータメモリ(たとえば、USB)または有線接続のような、トレーニングされたニューラルネットワークのカメラへの配布の任意のその他の手段が使用されてもよいことに留意されたい。次いで、カメラは、カメラ304によって取り込まれ、次いで適宜にエンコードされた画像の関連度を決定するために、トレーニングされたニューラルネットワークを使用するように構築される。そのため、カメラ304は、

20

複数のデジタル画像を取り込み、

複数のデジタル画像の画像データをトレーニングされたニューラルネットワークに提供し、複数のデジタル画像の各特定のデジタル画像について、特定のデジタル画像のピクセル群に各々対応する複数の関連度を取得するように構築されてもよい。

30

【0044】

一部の実施形態によれば、カメラは、取得された複数の関連度に従って特定のデジタル画像の圧縮度を制御することによって特定のデジタル画像をエンコードするように構築されてもよい。関連度と圧縮度との間のマッピングは線形であってもよい、つまり比較的高い関連度は常に、比較的低い関連度と比べて、対応するピクセル群のより低い圧縮度に対応してもよい。その他の実施形態によれば、マッピングは段階的なものである、つまり関連度の範囲は、同じ圧縮度にマップする。画像の多数の区域、または画像シーケンスが、関連すると見なされる画像コンテンツを含む場合には、使用可能な帯域幅を超えることがないように、マッピングはまた、使用可能な帯域幅を検討する必要がある。

40

【0045】

上記で説明されているように、一部の実施形態によれば、カメラ304は、ニューラルネットワークを実行するために専用のハードウェアを備えることができる。その他の実施形態によれば、ニューラルネットワークのソフトウェア実施は、使用されるべき関連度を取得するために使用される。いずれにせよ、上記で説明されているニューラルネットワークを使用することにより達成される結果は、関連度を取得するために第1/第2の画像解析アルゴリズム204、206を実際に行う場合と比較して、大幅に改善され得る。

【符号の説明】

【0046】

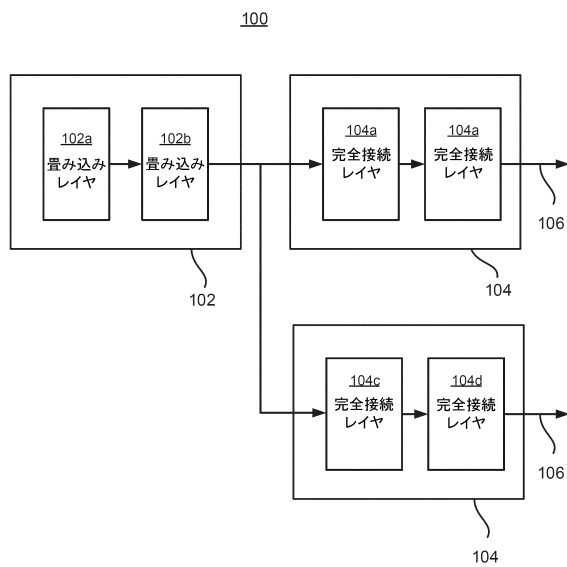
100 ニューラルネットワーク

50

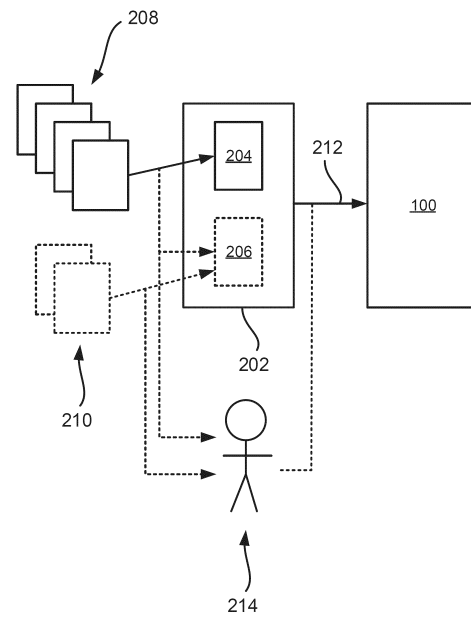
- 1 0 2 畳み込みレイヤ
- 1 0 2 a 畳み込みレイヤ
- 1 0 2 b 畳み込みレイヤ
- 1 0 4 回帰ヘッド
- 1 0 4 a 完全接続レイヤ
- 1 0 4 b 完全接続レイヤ
- 1 0 4 c 完全接続レイヤ
- 1 0 4 d 完全接続レイヤ
- 1 0 6 出力
- 2 0 0 デバイス、コンピュータ
- 2 0 8 デジタルトレーニング画像
- 2 1 2 トレーニングデータセット
- 2 0 4 第 1 の画像解析アルゴリズム
- 2 0 6 第 2 の画像解析アルゴリズム
- 2 1 4 人物
- 2 1 0 画像
- 3 0 2 無線ネットワーク
- 3 0 4 カメラ

10

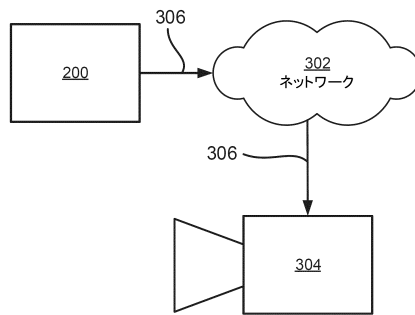
【図 1】



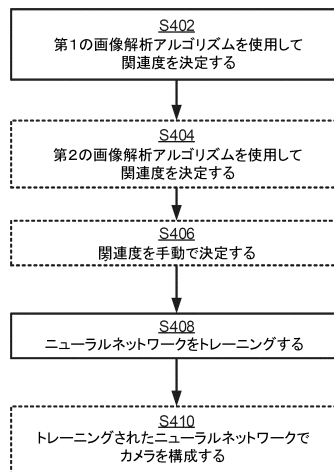
【図 2】



【図 3】



【図 4】



---

フロントページの続き

(72)発明者 ダニエルソン ファン, シン  
スウェーデン国 223 69 ルンド, エンダラヴェーゲン 14, シーノール アクシス  
コミュニケーションズ アーバー

審査官 吉田 誠

(56)参考文献 特表2010-537263(JP, A)  
特開2010-252276(JP, A)

(58)調査した分野(Int.Cl., DB名)  
G06F 16/00 - 16/958  
G06T 7/00 - 7/90