

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7263216号
(P7263216)

(45)発行日 令和5年4月24日(2023.4.24)

(24)登録日 令和5年4月14日(2023.4.14)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 3 5 0 C

請求項の数 20 (全16頁)

(21)出願番号	特願2019-212083(P2019-212083)	(73)特許権者	504407000
(22)出願日	令和1年11月25日(2019.11.25)		パロ アルト リサーチ センター インコ
(65)公開番号	特開2020-98587(P2020-98587A)		ーポレイテッド
(43)公開日	令和2年6月25日(2020.6.25)		アメリカ合衆国 カリフォルニア州 9 4
審査請求日	令和4年11月18日(2022.11.18)		3 0 4 パロ アルト カイオーテ ヒル
(31)優先権主張番号	16/222,062	(74)代理人	100094569
(32)優先日	平成30年12月17日(2018.12.17)		弁理士 田中 伸一郎
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100109070
早期審査対象出願			弁理士 須田 洋之
		(74)代理人	100067013
			弁理士 大塚 文昭
		(74)代理人	100086771
			弁理士 西島 孝喜
		(74)代理人	100109335

最終頁に続く

(54)【発明の名称】 ワッサースタイン距離を使用する物体形状回帰

(57)【特許請求の範囲】

【請求項 1】

画像内の物体の外形を検出するための方法であって、
少なくとも1つの物体に関連付けられた2次元(2D)情報を含む2D画像を受信することと、
ランダムノイズ信号を生成することと、
前記受信した2D画像及び前記ランダムノイズ信号を形状回帰器モジュールに提供することであって、前記形状回帰器モジュールは、前記物体に関連付けられた前記2D情報に基づいて、形状回帰モデルを適用して、前記受信した画像内の物体の形状外形を予測し、前記形状回帰モデルは、アノテーションされた訓練画像のセットによって提供される2D情報を使用して訓練されたワッサースタイン敵対的生成ネットワーク(W-GAN)を含み、それぞれの訓練画像は、前記訓練画像内で少なくとも1つの物体の2Dの真の形状外形を定義することによってアノテーションされる、ことと、を含む、方法。

【請求項 2】

前記形状回帰モデルを訓練することを更に含み、前記形状回帰モデルを訓練することが、
前記アノテーションされた訓練画像及びランダムノイズ信号を前記形状回帰モデルに入力して、推定された形状外形を取得することと、
前記推定された形状外形及び真の形状外形を、前記推定された形状外形及び真の形状外形の確率分布間のワッサースタイン距離を計算する識別器モジュールに提供することと、
前記識別器モジュールによって計算された前記ワッサースタイン距離に基づいて前記形

状回帰モデルのパラメータを最適化することと、を含む、請求項 1 に記載の方法。

【請求項 3】

前記形状回帰モデルを最適化することが、前記計算されたワッサーズタイン距離が低減されるように前記形状回帰モデルのパラメータを更新することを含む、請求項 2 に記載の方法。

【請求項 4】

前記形状回帰器モジュールが第 1 のニューラルネットワークとして実装され、前記識別器モジュールが第 2 のニューラルネットワークとして実装される、請求項 2 に記載の方法。

【請求項 5】

前記第 1 及び第 2 のニューラルネットワークが共に、敵対的生成ネットワーク (G A N) を形成する、請求項 4 に記載の方法。

【請求項 6】

前記予測された形状外形が、前記物体の境界に沿ったデカルト $x - y$ 座標に関して定義される、請求項 1 に記載の方法。

【請求項 7】

前記予測された形状外形が、前記物体の境界に属する画像ピクセルの確率に関して定義される、請求項 1 に記載の方法。

【請求項 8】

画像内の物体の外形を検出するための装置であって、
プロセッサと、
前記プロセッサに結合され、命令を記憶するメモリであって、前記命令が、前記プロセッサによって実行されると、前記プロセッサに、方法を実行させる、メモリと、
を備え、
前記方法は、
少なくとも 1 つの物体に関連付けられた 2 D 情報を含む 2 次元画像を受信することと、
ランダムノイズ信号を生成することと、
前記物体に関連付けられた前記 2 D 情報に基づいて、形状回帰モデルを適用して、前記受信した画像内の物体の形状外形を予測するように構成された形状回帰器モジュールに、前記受信した 2 D 画像及び前記ランダムノイズ信号を提供することであって、前記形状回帰モデルは、アノテーションされた訓練画像のセットによって提供される 2 D 情報を使用して訓練されたワッサーズタイン敵対的生成ネットワーク (W - G A N) を含み、それぞれの訓練画像は、前記訓練画像内で少なくとも 1 つの物体の 2 D の真の形状外形を定義することによってアノテーションされる、ことと、を含む、装置。

【請求項 9】

前記形状回帰モデルの訓練を容易にする識別器モジュールを更に備え、
前記形状回帰モデルを訓練することが、
前記アノテーションされた訓練画像及びランダムノイズ信号を前記形状回帰モデルに入力して、推定された形状外形を取得することと、
前記推定された形状外形及び真の形状外形を、前記推定された形状外形及び真の形状外形の確率分布間のワッサーズタイン距離を計算する前記識別器モジュールに提供することと、

前記識別器モジュールによって計算された前記ワッサーズタイン距離に基づいて前記形状回帰モデルのパラメータを最適化することと、を含む、請求項 8 に記載の装置。

【請求項 10】

前記形状回帰モデルを最適化することが、前記計算されたワッサーズタイン距離が低減されるように前記形状回帰モデルのパラメータを更新することを含む、請求項 9 に記載の装置。

【請求項 11】

前記形状回帰器モジュールが第 1 のニューラルネットワークとして実装され、前記識別器モジュールが第 2 のニューラルネットワークとして実装される、請求項 9 に記載

10

20

30

40

50

の装置。

【請求項 1 2】

前記第 1 及び第 2 のニューラルネットワークが共に、敵対的生成ネットワーク (G A N) を形成する、請求項 1 1 に記載の装置。

【請求項 1 3】

前記予測された形状外形が、前記物体の境界に沿ったデカルト $x - y$ 座標に関して定義される、請求項 8 に記載の装置。

【請求項 1 4】

前記予測された形状外形が、前記物体の境界に属する画像ピクセルの確率に関して定義される、請求項 8 に記載の装置。

【請求項 1 5】

命令を記憶する非一時的コンピュータ可読記憶媒体であって、前記命令が、コンピュータによって実行されると、前記コンピュータに、画像内の物体の外形を検出するための方法を実行させ、前記方法が、

少なくとも 1 つの物体に関連付けられた 2 次元 (2 D) 情報を含む 2 D 画像を受信することと、

ランダムノイズ信号を生成することと、

前記受信した 2 D 画像及び前記ランダムノイズ信号を形状回帰器モジュールに提供することであって、前記形状回帰器モジュールは、前記物体に関連付けられた前記 2 D 情報に基づいて、形状回帰モデルを適用して、前記受信した画像内の物体の形状外形を予測し、前記形状回帰モデルは、アノテーションされた訓練画像のセットによって提供される 2 D 情報を使用して訓練されたワッサースタイン敵対的生成ネットワーク (W - G A N) を含み、それぞれの訓練画像は、前記訓練画像内で少なくとも 1 つの物体の 2 D の真の形状外形を定義することによってアノテーションされる、ことと、を含む、非一時的コンピュータ可読記憶媒体。

【請求項 1 6】

前記方法が、前記形状回帰モデルを訓練することを更に含み、前記形状回帰モデルを訓練することが、

前記アノテーションされた訓練画像及びランダムノイズ信号を前記形状回帰モデルに入力して、推定された形状外形を取得することと、

前記推定された形状外形及び真の形状外形を、前記推定された形状外形及び真の形状外形の確率分布間のワッサースタイン距離を計算する識別器モジュールに提供することと、

前記識別器モジュールによって計算された前記ワッサースタイン距離に基づいて前記形状回帰モデルのパラメータを最適化することと、を含む、請求項 1 5 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 1 7】

前記形状回帰モデルを最適化することが、前記計算されたワッサースタイン距離が低減されるように前記形状回帰モデルのパラメータを更新することを含む、請求項 1 6 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 1 8】

前記形状回帰器モジュールが第 1 のニューラルネットワークとして実装され、前記識別器モジュールが第 2 のニューラルネットワークとして実装され、前記第 1 及び第 2 のニューラルネットワークが共に、敵対的生成ネットワーク (G A N) を形成する、請求項 1 6 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 1 9】

前記予測された形状外形が、前記物体の境界に沿ったデカルト $x - y$ 座標に関して定義される、請求項 1 5 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 2 0】

前記予測された形状外形が、前記物体の境界に属する画像ピクセルの確率に関して定義される、請求項 1 5 に記載の非一時的コンピュータ可読記憶媒体。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、概して、コンピュータビジョンシステムに関する。より具体的には、本開示は、ニューラルネットワークフレームワークを使用して物体形状を予測するシステム及び方法に関する。

【0002】

物体検出（例えば、画像内の顔、建造物、車などの検出）は、コンピュータビジョンにおける基本的な問題である。物体検出における現在のアプローチは、多くの場合、物体の粗い軸平行境界ボックス（axis-aligned bounding boxes、AABB）を生成することを伴う。例えば、多くの顔検出アプリケーションは、検出された顔の周囲に矩形のボックスを作成する。そのようなボックスは、典型的には、物体の形状に適合せず、したがって、細粒シーン分析では限定された有用性を有する。一方、物体輪郭検出（例えば、物体の詳細な外形を検出すること）は、物体に関連付けられたより正確な形状情報を提供することができる。実際に、物体境界を正確に見つける能力は、分割、認識、及び場面理解を含む多くの視覚タスクに対して、広範囲に及ぶ影響を有し得る。

【0003】

Polygon-RNNなどの現在のアプローチは、再帰型ニューラルネットワーク（recurrent neural network、RNN）を使用して、物体の外形又は輪郭に沿う点を（多角形の一部として）予測することより、物体の境界又は輪郭を検出及び追跡する。より具体的には、物体の外形上の現在の予測点の画像パッチが与えられると、Polygon-RNNは、次の予測点の2D位置を回帰することができる。しかしながら、Polygon-RNNアプローチは、それ自身の欠点を有する。Polygon-RNNは、多くの場合、外形上の点の数に対応する複数の精選されたハイパーパラメータを必要とし、また第1の点の特別な処理が必要である。

【0004】

一実施形態は、画像内の物体の外形を検出するためのシステムを提供することができる。動作中、システムは、少なくとも1つの物体を含む画像を受信し、ランダムノイズ信号を生成し、受信した画像及びランダムノイズ信号を形状回帰器モジュールに提供し、形状回帰器モジュールは、形状回帰モデルを適用して、受信した画像内の物体の形状外形を予測する。

【0005】

この実施形態の変形例では、システムは、形状回帰モデルを更に訓練する。訓練中、システムは、訓練画像のセットを取得する。それぞれの訓練画像はアノテーションされ、かつ少なくとも1つの物体を含んでおり、アノテーションされた訓練画像は、真の形状外形のセットに関連付けられている。システムは、訓練画像及びランダムノイズ信号を形状回帰モデルに入力して、推定された形状外形を取得し、推定された形状外形及び真の形状外形を、推定された形状外形及び真の形状外形の確率分布間のワッサーズタイン距離を計算する識別器モジュールに提供し、識別器モジュールによって計算されたワッサーズタイン距離に基づいて形状回帰モデルのパラメータを最適化する。

【0006】

更なる変形例では、形状回帰モデルを最適化することは、計算されたワッサーズタイン距離が低減されるように形状回帰モデルのパラメータを更新することを含むことができる。

【0007】

更なる変形例では、形状回帰器モジュールは第1のニューラルネットワークとして実装され、識別器モジュールは第2のニューラルネットワークとして実装される。

【0008】

更なる変形例では、第1及び第2のニューラルネットワークは共に、敵対的生成ネットワーク（generative adversarial network、GAN）を形成する。

10

20

30

40

50

【 0 0 0 9 】

この実施形態の変形例では、形状外形は、物体の境界に沿ったデカルト $x - y$ 座標に関して定義される。

【 0 0 1 0 】

この実施形態の変形例では、形状外形は、物体の境界に属する画像ピクセルの確率に関して定義される。

【 0 0 1 1 】

一実施形態は、画像内の物体の外形を検出するための装置を提供することができる。装置は、少なくとも1つの物体を含む画像を受信するように構成された受信モジュールと、ランダムノイズ信号を生成するように構成されたノイズ生成器と、画像及びランダムノイズ信号を受信し、形状回帰モデルを適用して、受信した画像内の物体の形状外形を予測するように構成された形状回帰器モジュールと、を含むことができる。

【図面の簡単な説明】

【 0 0 1 2 】

【図1】例示的な敵対的生成ネットワーク (G A N) アーキテクチャ (従来技術) を示す。

【図2】本発明の一実施形態による、G A N 生成器の例示的な早期のイテレーション出力を示す。

【図3】本発明の一実施形態による、例示的な生成器のアーキテクチャを示す。

【図4】本発明の一実施形態による、例示的な批評器のアーキテクチャを示す。

【図5】本発明の一実施形態による、例示的な輪郭検出ワッサースタイン G A N (W a s s e r s t e i n G A N 、 W - G A N) のアーキテクチャを示す。

【図6】本発明の一実施形態による、複数回のイテレーション後の輪郭検出ワッサースタイン G A N (W - G A N) 生成器の出力を示す。

【図7A】本発明の一実施形態による、輪郭検出ワッサースタイン G A N (W - G A N) の例示的な訓練プロセスを示すフロー図を提示する。

【図7B】本発明の一実施形態による、例示的な形状検出プロセスを示すフロー図を提示する。

【図8】本発明の一実施形態による、物体形状検出システムを容易にする例示的なコンピュータシステムを示す。

【図9】本発明の一実施形態による、物体輪郭検出システムを容易にする例示的な装置を示す。

【発明を実施するための形態】

【 0 0 1 3 】

概要

本明細書に記載される実施形態は、画像内の物体の境界を自動的に検出する問題を解決する。より具体的には、システムは、物体境界を検出するように修正されている W - G A N (ワッサースタイン敵対的生成ネットワーク) に基づく。W - G A N は、画像の推測された境界と、画像の境界のグラウンドトゥールズ (すなわち、ヒューマンアノテーション) との間のワッサースタイン距離を最小化することによって訓練され得る。W - G A N フレームワークは、形状回帰器 (すなわち、生成器) 及び W スコア計算器 (すなわち、識別器) を含むことができる。形状回帰器は、画像、及びランダムノイズのセットを入力に取り、予測された境界点を出力することができる。識別器は、予測された境界点とグラウンドトゥールズ境界点との間のワッサースタイン距離を計算することができる。形状回帰器は、ワッサースタイン距離を最小化することによって訓練され得る。

【 0 0 1 4 】

ワッサースタイン G A N

敵対的生成ネットワーク (G A N) は、ゼロサムゲームフレームワークにおいて互いに競合する2つのニューラルネットワークからなるシステムによって実装される、教師なし機械学習で使用される人工知能アルゴリズムのークラスである。近年、G A N は、合成だがりアルな画像など、合成データを生成する手段として人気が高まってきている。これを

行うために、G A N は、典型的には、生成器ニューラルネットワーク（生成器と呼ばれる）と、識別器ニューラルネットワーク（識別器と呼ばれる）を含む。

【 0 0 1 5 】

動作中、生成器は、潜在空間から特定の対象となるデータ分布にマッピングすることを学習し、一方、識別器は、真のデータ分布からのインスタンスと生成器によって生成された候補との間の違いを識別する。生成器の訓練目的は、真のデータ分布から生じたかのように見える新規な合成されたインスタンスを生成することによって、識別器の誤り率を増加させる（すなわち、識別器ネットワークを「騙す」）ことである。一例として画像生成を使用すると、生成器は、合成画像サンプルを出力として生成し、これらの画像が実在画像であると識別器に「確信させる」ことによって合成画像サンプルの品質を改善しようとし得る。識別器は、実在画像サンプルと生成された合成画像サンプルとを見分けることを課せられる。識別器は、画像が全体として実在であるか否かを判定する。結果として、複数回のイテレーションにより、生成器は、実在画像の統計的特性を組み入れた合成画像を生成することを学習する。

10

【 0 0 1 6 】

図 1 は、例示的な敵対的生成ネットワーク（G A N）アーキテクチャ（従来技術）を示す。G A N 1 0 0 は、生成器 1 0 2 及び識別器 1 0 4 を含むことができる。生成器 1 0 2 は、正規分布又は一様分布を使用するノイズをサンプリングし、ディープネットワークを使用して出力（例えば、画像）を作成することができる。生成器 1 0 2 の出力は、識別器 1 0 4 に送信され得、識別器 1 0 4 は、その入力（すなわち、生成器 1 0 2 の出力）を訓練サンプルと比較することによってそれが実在するものであるか、又は生成されたものであるかを判定する。識別器 1 0 4 は、その入力を実在するものである可能性を推定するための値（例えば、 $D(x)$ ）を出力することができる。

20

【 0 0 1 7 】

ワッサースタイン距離（アースムーバーの距離（earth mover's distance）、又は E M D と呼ばれる）は、確率分布間に定義される距離関数である。ワッサースタイン G A N（W - G A N）は、ワッサースタイン距離をメトリック（例えば、損失関数）として使用する G A N を指す。より具体的には、識別器は、ワッサースタイン距離を使用して、生成器によって出力された画像の品質を判断し、合成画像の高品質化を促す。生成器の訓練目的は、生成された候補画像と真の画像（グラウンドトゥールースと呼ばれることもある）との間のワッサースタイン距離を最小化することである。分布 P_r と分布 P_g との間のワッサースタイン距離は、

30

【 0 0 1 8 】

【数 1】

$$W(P_r, P_g) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|]$$

【 0 0 1 9 】

として定義することができ、式中、

【 0 0 2 0 】

40

【数 2】

$$\prod(P_r, P_g)$$

【 0 0 2 1 】

は、周辺がそれぞれ P_r 及び P_g である全ての同時分布 (x, y) のセットを示す。直感的には、 (x, y) は、分布 P_r を分布 P_g に変換するために、どのくらいの「量」が x から y に移送されなければならないかを示す。確率分布の単純なシーケンスがワッサースタイン距離下で収束することは証明され得る。更に、ワッサースタイン距離が連続的かつ微分可能であり、したがって、訓練中に連続的に滑らかな勾配を提供可能であるこ

50

とは証明され得る。そのような特性は、識別器を訓練する際に有益であり得る。より具体的には、W - G A Nにおいて、識別器は、最適になるまで訓練され得る。

【 0 0 2 2 】

物体形状検出のためのW - G A N

コンピュータビジョンでは、画像外観を物体の存在及び位置に直接マッピングする回帰関数を学習することによって、回帰ベースの方法を物体検出のために使用することができる。具体的な例として、画像内の物体の外形点（すなわち、外形上の点）を生成するために、形状回帰器モジュールを使用することができる。より詳細には、形状回帰器モジュールの出力は、外形点の2D位置（例えば、デカルトx - y座標）であり得る。例えば、形状回帰器の入力が画像であれば、形状回帰器の出力は、画像の外観を与えられた座標のセット（例えば、 (x_1, y_1) 、 (x_2, y_2) 、 \dots 、 (x_n, y_n) ）又はこれらの座標の条件付き確率であり得る。

10

【 0 0 2 3 】

形状回帰器モジュールを実現するために、様々な技術を使用することができる。いくつかの実施形態では、形状回帰器モジュールは、フィードフォワードニューラルネットワークを実装する。更なる実施形態では、ニューラルネットワークは、多層パーセプトロンネットワークであり得る。

【 0 0 2 4 】

いくつかの実施形態では、形状回帰器モジュールは、G A Nの一部であり得る。より具体的には、形状回帰器モジュールは、G A Nの生成器であり得、候補の生成に関与し得る。合成画像が必要とされる用途では、G A Nの生成器は候補画像を生成するが、本発明の実施形態では、G A N生成器（すなわち、形状回帰器モジュール）は外形点を生成する。G A Nを訓練するために、複数のアノテーションされたサンプル画像を使用することができる。より具体的には、サンプル画像内の1つ以上の物体の外形が手動でマーキングされている。動作中、対象となる画像が特定の量のノイズ（元のサンプル画像の変形形態又は摂動画像を生成する）と共に生成器に送信され得、生成器は、物体の候補外形を生成することができる。図2は、一実施形態による、G A N生成器の例示的な早期のイテレーション出力を示す。図2中、曲線202は、ターゲット形状（すなわち、サンプル）であり得、曲線204は生成器出力であり得る。図2は、生成器の開始出力とターゲット画像との間の有意差を明確に示している。いくつかの実施形態では、画像の表現（例えば、多次元ベクトル）が、G A N生成器への入力として使用され得、ノイズは乱数生成器によって生成され得る。更に、ノイズは、正規分布又は一様分布を有し得る。

20

30

【 0 0 2 5 】

図3は、一実施形態による、例示的な生成器のアーキテクチャを示す。生成器300は、画像を受信するための画像受信モジュール302と、受信した画像を多次元ベクトルに変換するための画像表現モジュール304と、ランダムノイズを生成するための乱数生成器306と、合成器308と、複数の層（例えば、入力層、出力層、及び複数の隠れ層）を含むことができる、ニューラルネットワーク310と、を含むことができる。いくつかの実施形態では、ニューラルネットワーク310は、形状回帰器であり得る。生成器300はまた、ニューラルネットワーク310の出力を分析し、同出力を、外形点又は形状分布の2D座標など、所望の形式に変換するための出力モジュール312を含むこともできる。例えば、画像の各ピクセルについて、生成器300は、ピクセルが物体の外形上にある可能性を示す出力を生成することができる。いくつかの実施形態では、生成器300は、外形点のセット（例えば、 (x_1, y_1) 、 (x_2, y_2) 、 \dots 、 (x_n, y_n) ）又は外形点の確率分布を出力することができる。生成器300によって生成される外形点の数は、外形の形状に応じて、任意であり得ることに留意されたい。より複雑な外形は、それを記述するためにより多くの点を必要とし得るが、より単純な外形は、より少ない点を必要とし得る。一般に、生成器300によって生成又は出力される外形点の数は、識別器に提供されるグラウンドトゥルス外形点の数と一致しなくてもよい。このことは、ニューラルネットワークによって生成される外形点の数が訓練サンプルの数と同じである

40

50

ことを必要とする Polygon - RNN アプローチと比較して有利である。

【 0 0 2 6 】

ニューラルネットワーク 3 1 0 は画像を入力として受信するため、(x , y) 座標の形状分布は、ニューラルネットワーク 3 1 0 の内部に暗黙的に埋め込まれることに留意されたい。入力画像の形状を生成するために、ニューラルネットワーク 3 1 0 のフォワードパスを行うことができ、これにより、暗黙的な形状分布から (x , y) 座標がサンプリングされる。形状座標系は、入力画像の中心を中心とすることができる。例えば、x - y 平面の原点は、入力画像の中心にあることができる。

【 0 0 2 7 】

任意の種類の GAN と同様に、生成器 3 0 0 の出力は、比較のために識別器モジュール 10 に送信され得る。いくつかの実施形態では、識別器は、生成器 3 0 0 の出力とグラウンドトゥールスとの間のワッサーズタイン距離を計算する。ワッサーズタイン GAN の設定では、識別器は批評器とも呼ばれ得、批評器は、スカラーワッサーズタイン距離を出力する。図 4 は、本発明の一実施形態による、例示的な批評器のアーキテクチャを示す。

【 0 0 2 8 】

識別器 4 0 0 は、アノテーションされたサンプル画像を受信するための画像受信モジュール 4 0 2 を含むことができる。より具体的には、サンプル画像は、特定の種類の物体（例えば、顔、車、建造物など）の輪郭又は外形を示すようにアノテーションされている。識別器 4 0 0 はまた、生成器出力との比較に使用され得るグラウンドトゥールスを生成するためのグラウンドトゥールス生成モジュール 4 0 4 を含むこともできる。グラウンドトゥールスは、アノテーションされたサンプル画像から抽出され得る。いくつかの実施形態では、グラウンドトゥールスは、形状分布（例えば、外形点 (x * , y *) の確率分布であり得る。アノテーションされた画像内の外形点の数は、生成器 3 0 0 によって生成される外形点の数と同じでなくてもよい。識別器 4 0 0 はまた、生成器の出力を受信するための生成器出力受信モジュール 4 0 6 を含むこともできる。生成器の出力及びグラウンドトゥールスは、ワッサーズタイン距離計算器 4 0 8 に送信され得、ワッサーズタイン距離計算器 4 0 8 は、ワッサーズタイン距離を計算する。なお、ニューラルネットワークに、ワッサーズタイン距離 W を計算する上で必須である、1 - リプシッツ関数 f を学習させることが必要であることに留意されたい。いくつかの実施形態では、ワッサーズタイン距離は、以下のように計算され得る。

【 0 0 2 9 】

【 数 3 】

$$W(P_s, P_*) = \max_{\|f\|_L \leq 1} E_{(x,y) \sim P_s} [f(x,y)] - E_{(x,y) \sim P_*} [f(x,y)] , \quad (1)$$

【 0 0 3 0 】

式中、f は 1 - リプシッツ関数であり、P_s 及び P_{*} はそれぞれ、生成器出力及びグラウンドトゥールス分布である。f における最大重み値を制限するためにクリッピングが必要であることに留意されたい。より具体的には、識別器の重み係数は、ハイパーパラメータのセットによって制御される特定の範囲内である必要がある。

【 0 0 3 1 】

いくつかの実施形態では、x 座標及び y 座標の分布は、独立してモデル化され、生成され得る。換言すれば、ワッサーズタイン距離（又は W スコア）は、x 座標及び y 座標に関して独立して計算され得る。代替実施形態では、x 座標及び y 座標は、複雑な形状表現を介して一緒にモデル化され得る。ワッサーズタイン距離計算器 4 0 8 の出力は、W スコアと表される。

【 0 0 3 2 】

図 5 は、本発明の一実施形態による、例示的な輪郭検出ワッサーズタイン GAN (W - GAN) のアーキテクチャを示す。輪郭検出 W - GAN 5 0 0 は、生成器入力モジュール 5 0 2、形状回帰器ベースの生成器 5 0 4、識別器訓練入力モジュール 5 0 6、W スコア

10

20

30

40

50

ベースの識別器 508、及び最適化モジュール 510 を含むことができる。

【0033】

生成器入力モジュール 502 は、形状回帰器ベースの生成器 504 に入力を提供することに関与し得る。入力は、対象となる画像及びランダムノイズを含み得る。より具体的には、生成器入力モジュール 502 は、対象となる画像を、形状回帰器ベースの生成器 504 に送信され得る多次元ベクトルに変換することができる。ランダムノイズは、イテレーション間で更新され得る。いくつかの実施形態では、形状回帰器ベースの生成器 504 は、形状回帰器として機能するニューラルネットワーク（例えば、フィードフォワードニューラルネットワーク）を含むことができる。より具体的には、形状回帰器ベースの生成器 504 は、候補形状分布（例えば、外形点のセット）を生成するように構成され得る。画像には、形状分布に関連する全ての情報が暗黙的に含まれていることに留意されたい。したがって、形状回帰器ベースの生成器 504 は、回帰モデルを使用して、画像内の外形点（例えば、ピクセル）を特定することができる。形状回帰器ベースの生成器 504 によって生成される外形点の数は任意であり得、外形の形状によって決定され得る。外形点が 1 つずつ生成される Polygon-RNN アプローチと比較して、形状回帰器ベースの生成器 504 は、1 つのパスを使用して外形全体を生成することができる。形状回帰器ベースの生成器 504 の出力は、予測された外形点の座標（例えば、 (x_1, y_1) 、 (x_2, y_2) 、 \dots 、 (x_n, y_n) ）又はそれらの確率分布 $P(x, y)$ を含むことができる。形状回帰器ベースの生成器 504 の出力は、Wスコアベースの識別器 508 に送信され得る。

【0034】

識別器訓練入力モジュール 506 は、訓練サンプル（すなわち、アノテーションされたサンプル画像）を Wスコアベースの識別器 508 に送信することができる。いくつかの実施形態では、識別器訓練入力モジュール 506 は、グラウンドトゥルス外形点の座標（例えば、 (x_1^*, y_1^*) 、 (x_2^*, y_2^*) 、 \dots 、 (x_m^*, y_m^*) ）又はそれらの確率分布 $P(x^*, y^*)$ を Wスコアベースの識別器 508 に送信することができる。

【0035】

Wスコアベースの識別器 508 は、形状回帰器ベースの生成器 504 によって生成された外形点と、アノテーションされたサンプル画像内のグラウンドトゥルス外形点との間のワッサースタイン距離（又は Wスコア）を計算することができる。より具体的には、Wスコアは、式（1）を使用して計算され得る。いくつかの実施形態では、Wスコアベースの識別器 508 は、2 つの確率分布に基づいて 1 - リプシッツ関数を学習することに関与するニューラルネットワークを含むことができ、このニューラルネットワークは、Wスコアを計算する上で必須である。いくつかの実施形態では、Wスコアベースの識別器 508 は、外形点の x 座標及び y 座標に関して独立して Wスコアを計算し、次いで、それらを組み合わせる合計スコアを得ることができる。これは、より単純な 1 - リプシッツ関数（確率分布は 1 つの変数のみを有する）を可能にする。代替実施形態では、Wスコアベースの識別器 508 は、 x 座標及び y 座標の同時確率分布に基づいて Wスコアを計算することができる。

【0036】

Wスコアベースの識別器 508 は、形状回帰器ベースの生成器 504 のパラメータが固定されると、訓練され得る。所定の回数のイテレーション後、Wスコアベースの識別器 508 は、計算された Wスコアを最適化モジュール 510 に送信することができ、最適化モジュール 510 は、最適化条件が満たされたかどうかを判定することができる。最適化条件は、生成器イテレーションの回数が所定の閾値以上であること又は Wスコアが所定の値未満であることを含み得る。最適化条件が満たされると、最適化モジュール 510 は、形状回帰器ベースの生成器 504 を、その現在の予測を予測された外形点として出力するように構成することができる。そうでなければ、最適化モジュール 510 は、形状回帰器ベースの生成器 504 が自身のパラメータを調整し、外形点の新たな予測を生成することができるように、形状回帰器ベースの生成器 504 に制御信号を送信することができる。こ

のプロセスは、所定の最適化条件（例えば、生成器イテレーションの回数又はWスコアに基づく）が達成されるまで、それ自体を繰り返すことができる。一般に、イテレーション回数が多ければ、多くの場合、外形又は輪郭の予測は良好になり得る。形状回帰器ベースの生成器504の各イテレーションに対して、Wスコアベースの識別器508は、多くの場合、多数回（例えば、約1000回）の訓練イテレーションを実行することに留意されたい。換言すれば、形状回帰器ベースの生成器504のパラメータの各更新に対して、Wスコアベースの識別器508のパラメータは多数回更新されている。

【0037】

図6は、本発明の一実施形態による、複数回のイテレーション後の輪郭検出ワッサースタインGAN(W-GAN)生成器の出力を示す。上の図面は、149回のイテレーション後の結果を示し、下の図面は、949回のイテレーション後の結果を示す。この例では、訓練サンプルは、曲線602及び606として示される、ドメイン[0,]内の1次元正弦関数からランダムに選択された100個の点を含み得る。予測された形状（例えば、形状回帰器ベースの生成器の出力）は、曲線604及び608として示されている。形状回帰器ベースの生成器は、例えば、それぞれ50個のニューロンを有する2層のパーセプトロンを含み得、Wスコアベースの識別器は、形状回帰器ベースの生成器と同様のネットワーク構造を有し得る。入力ランダムノイズは、100次元ノイズであり得る。各イテレーションでは、Wスコアベースの識別器は1000ステップで訓練され、形状回帰器ベースの生成器は1ステップで訓練される（すなわち、1ステップのフォワードパスを行う）。図6からわかるように、イテレーション回数が増加すると、予測された形状は真のターゲット形状に近づく。

【0038】

図7Aは、本発明の一実施形態による、輪郭検出ワッサースタインGAN(W-GAN)の例示的な訓練プロセスを示すフロー図を提示する。動作中、システムは、複数の訓練画像を取得する（動作702）。訓練画像はアノテーションされていることに留意されたい。より具体的には、物体の輪郭又は外形は、それらの訓練画像内でラベル付けされており、形状分布のグラウンドトゥルースを表している。システムは、ランダムノイズを生成する（動作704）。訓練画像及びランダムノイズは組み合わせられ、形状回帰器に送信され得る（動作706）。いくつかの実施形態では、形状回帰器は、画像内の物体の輪郭又は外形を検出するように訓練され得るフィードフォワードニューラルネットワークを含むことができる。形状回帰器は、訓練画像の画像形状分布（例えば、少なくとも1つの物体の外形又は輪郭上の点の位置）を予測する（動作708）。いくつかの実施形態では、予測された画像形状分布は、複数の外形点及びそれらの確率分布を含むことができる。外形点の数は、アノテーションされたサンプル画像内でラベル付けされている外形点の数とは異なり得る。アノテーションされたサンプル画像（すなわち、グラウンドトゥルース）及び予測された形状分布は、Wスコア計算器に送信され得る（動作710）。

【0039】

Wスコア計算器は、予測された形状分布とグラウンドトゥルースとの間のワッサースタイン距離(Wスコア)を計算する（動作712）。より具体的には、Wスコア計算器内のニューラルネットワークは、計算されたWスコアを出力する前に、多数回のイテレーションにわたって訓練されてもよい。次いで、システムは、所定の最適化条件が満たされているかどうかを判定する（動作714）。所定の最適化条件は、形状回帰器のパラメータ更新イテレーションの回数が所定の閾値に達しているかどうか、又はWスコアが閾値未満であるかどうかを含み得る。最適化条件が満たされた場合、システムは形状回帰モデルを出力する（動作716）。そうでなければ、システムは形状回帰器のパラメータを更新し（動作718）、新たな予測が行われる（動作708）。より具体的には、形状回帰器のパラメータは、後に生成されるワッサースタイン距離が低減され得るように、すなわち、予測された形状分布とグラウンドトゥルース形状分布との間の類似性が増すように、更新される。

【0040】

図 7 B は、本発明の一実施形態による、例示的な形状検出プロセスを示すフロー図を提示する。動作中、システムは、少なくとも 1 つの物体を含む検出対象画像を受信し得る（動作 7 2 2）。システム内のランダムノイズ生成器は、ランダムノイズを生成し得る（動作 7 2 4）。いくつかの実施形態では、ランダムノイズは、正規分布又は一様分布を有し得る。次いで、システムは、前に訓練した形状回帰モデルを適用して、検出対象画像内の 1 つ以上の物体の形状外形を予測し得る（動作 7 2 6）。いくつかの実施形態では、画像及びノイズは形状回帰器に送信され得、形状回帰器は、訓練した形状回帰モデルを適用する。

【 0 0 4 1 】

一般に、本発明の実施形態は、画像内の物体境界又は外形の迅速な推測のための解決策を提供する。物体の周囲の矩形領域のみを検出する従来のアプローチと比較して、境界検出は、はるかに高い精度を提供することができる。一方、ピクセルレベルのインスタンス分割に基づくアプローチと比較して、現在の W - G A N アプローチは、はるかに高速であり得、したがって、リアルタイムアプリケーションを可能にする。例えば、カメラシステムが道路をスキャンすると、画像内の車が、輪郭検出 W - G A N を使用してリアルタイムで検出され、その車の外形でその車の位置がマーキングされる。同様に、様々な顔検出アプリケーション（例えば、ユーザ認証又は監視）では、矩形ボックスを使用して顔をマーキングする代わりに、各顔の実際の外形が、輪郭検出 W - G A N を使用して検出及びマーキングされ得る。更に、形状検出能力はまた、様々な物体検出アプリケーションの性能を向上させることもできる。

【 0 0 4 2 】

例示的なコンピュータシステム及び装置

図 8 は、本発明の一実施形態による、物体形状検出システムを容易にする例示的なコンピュータシステムを示す。コンピュータシステム 8 0 0 は、プロセッサ 8 0 2、メモリ 8 0 4、及び記憶デバイス 8 0 6 を含む。コンピュータシステム 8 0 0 は、ディスプレイデバイス 8 1 0、キーボード 8 1 2、及びポインティングデバイス 8 1 4 に結合され得、また、1 つ以上のネットワークインタフェースを介してネットワーク 8 0 8 に結合され得る。記憶デバイス 8 0 6 は、オペレーティングシステム 8 1 8 及び物体形状検出システム 8 2 0 を記憶することができる。

【 0 0 4 3 】

物体形状検出システム 8 2 0 は、コンピュータシステム 8 0 0 によって実行されると、コンピュータシステム 8 0 0 に、本開示で説明される方法及び / 又はプロセスを実行させることができる命令を含むことができる。物体形状検出システム 8 2 0 は、対象となる画像及びアノテーションされた訓練サンプルの両方を含めて、画像を受信するための命令（画像受信モジュール 8 2 2）と、画像から画像表現を抽出するための命令（画像表現モジュール 8 2 4）と、ノイズを生成するための命令（ノイズ生成器 8 2 6）と、を含むことができる。更に、物体形状検出システム 8 2 0 は、形状回帰モデルを適用するための命令（形状回帰器モジュール 8 2 8）と、予測された形状分布とグラウンドトゥルースとの間のワッサースタイン距離を推定するための命令（ワッサースタイン距離推定モジュール 8 3 0）と、推定されたワッサースタイン距離に基づいて形状回帰器モジュールのパラメータを最適化するための命令（最適化モジュール 8 3 2）と、を含むことができる。

【 0 0 4 4 】

図 9 は、本発明の一実施形態による、物体輪郭検出システムを容易にする例示的な装置を示す。装置 9 0 0 は、有線、無線、量子光、又は電気通信チャネルを介して互いに通信し得る複数のユニット又は装置を備えることができる。装置 9 0 0 は、1 つ以上の集積回路を使用して実現され得、図 9 に示されているものよりも少ない又は多いユニット又は装置を含み得る。更に、装置 9 0 0 は、コンピュータシステムに統合され得るか、又は他のコンピュータシステム及び / 若しくはデバイスと通信することができる別個のデバイスとして実現され得る。具体的には、装置 9 0 0 は、図 8 のコンピュータシステム 8 0 0 のモジュール 8 2 2 ~ 8 3 2 と同様の機能又は動作を実行するユニット 9 0 2 ~ 9 1 2 を備え

ることができ、これには、画像受信ユニット 902、画像表現ユニット 904、ノイズ生成ユニット 906、形状回帰器ユニット 908 と、ワッサーズタイン距離推定ユニット 910、及び最適化ユニット 912 が含まれる。装置 900 は、通信ユニット 914 を更に含むことができる。

【0045】

「発明を実施するための形態」セクションに記載される方法及び処理は、上記のようにコンピュータ可読記憶媒体に記憶され得るコード及び/又はデータとして具体化することができる。コンピュータシステムが、コンピュータ可読記憶媒体上に記憶されたコード及び/又はデータを読み取って実行すると、コンピュータシステムは、データ構造及びコードとして具体化され、コンピュータ可読記憶媒体内に記憶された方法及び処理を実行する。

10

【0046】

更に、上述の方法及び処理は、ハードウェアモジュール又は装置に含まれてもよい。ハードウェアモジュール又は装置としては、特定用途向け集積回路(application-specific integrated circuit、ASIC)チップ、フィールドプログラム可能ゲートアレイ(field-programmable gate array、FPGA)、特定の時刻に特定のソフトウェアモジュール又はコードを実行する専用又は共有プロセッサ、及び、既知の又は後に開発される他のプログラム可能論理デバイスを含むことができるが、これらに限定されない。ハードウェアモジュール又は装置が起動されると、それらの内部に含まれる方法及び処理が実行される。

20

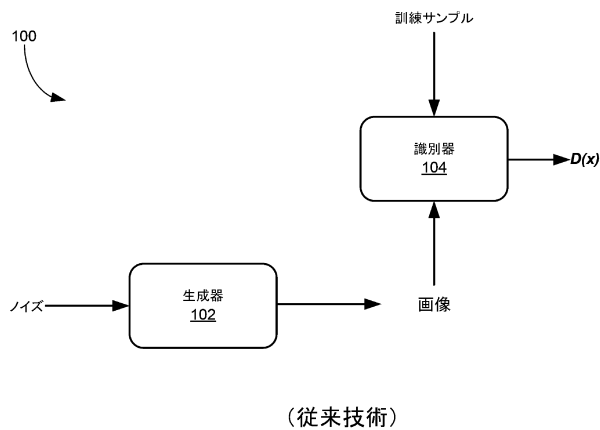
30

40

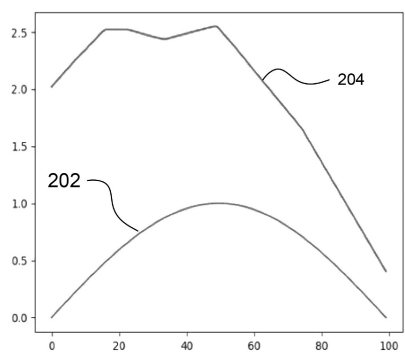
50

【図面】

【図 1】

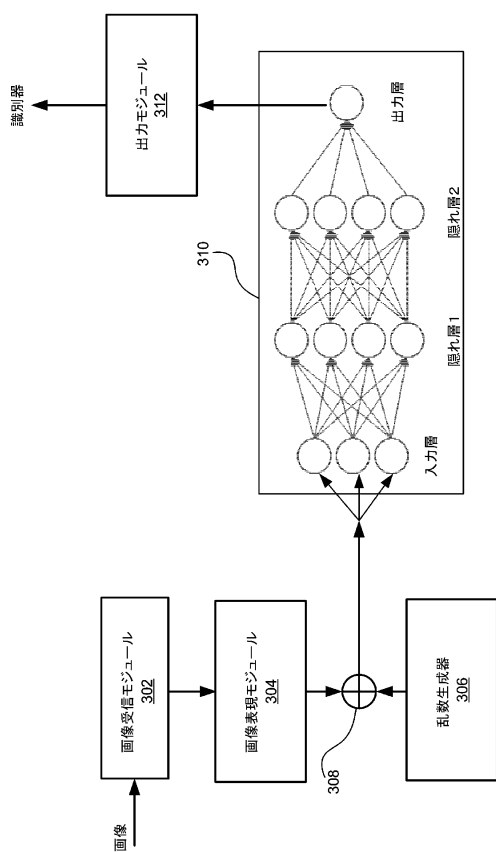


【図 2】

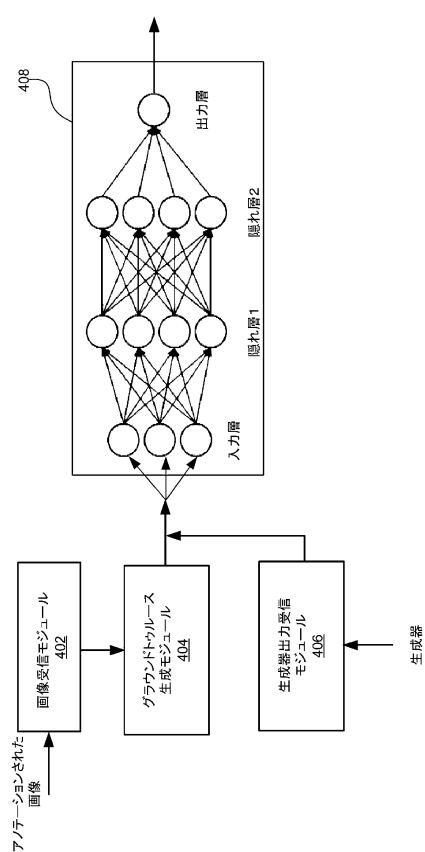


10

【図 3】



【図 4】



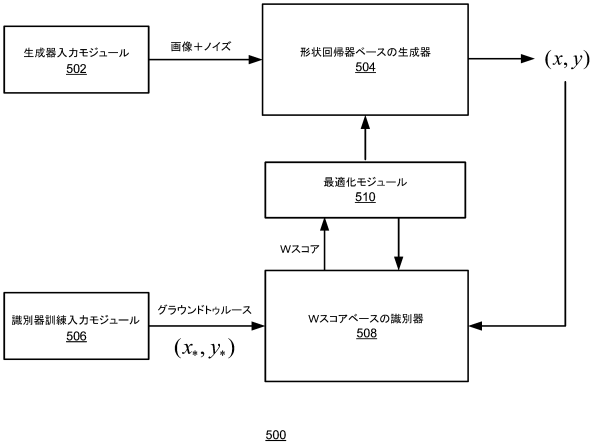
20

30

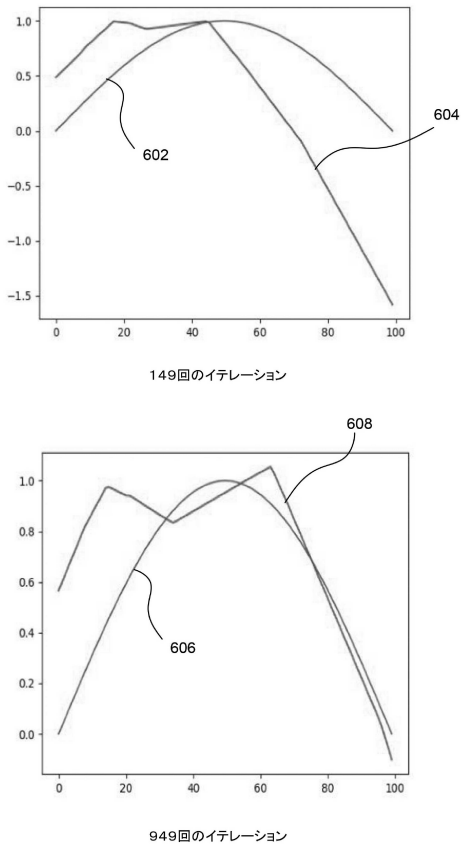
40

50

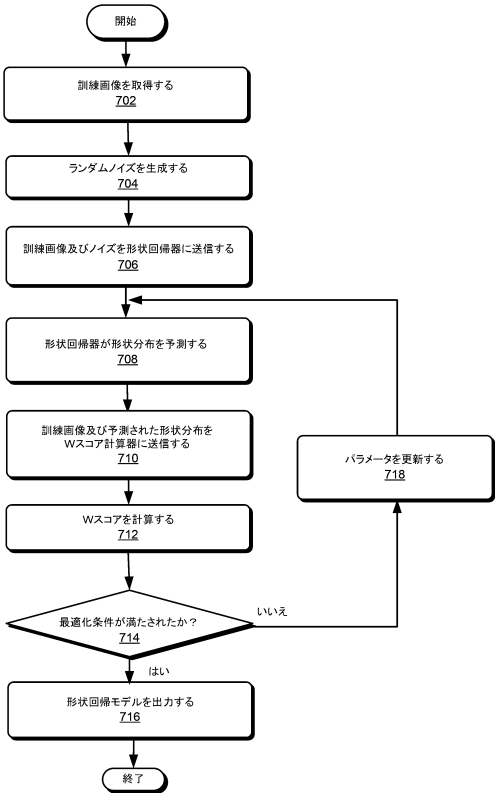
【図 5】



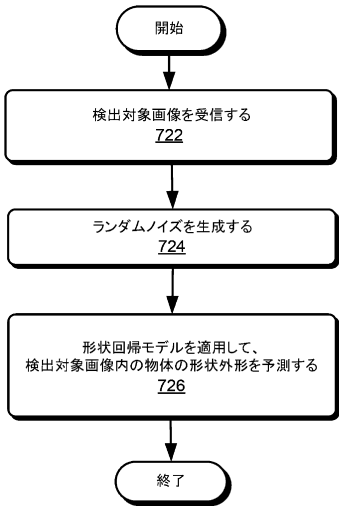
【図 6】



【図 7 A】



【図 7 B】



10

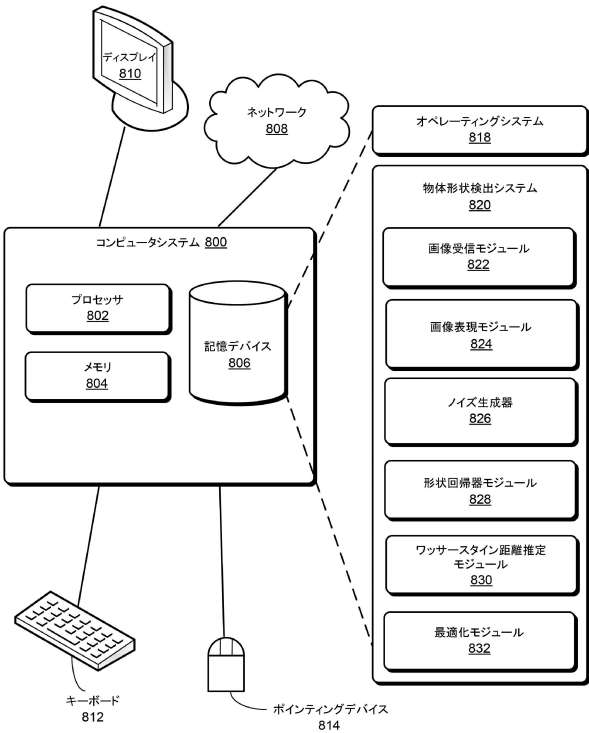
20

30

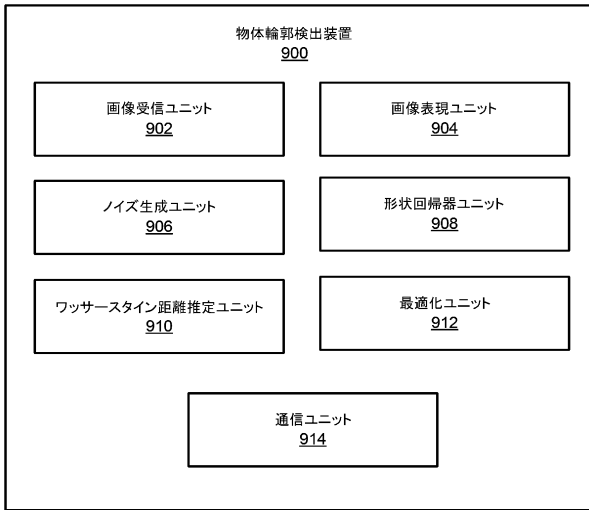
40

50

【図 8】



【図 9】



10

20

30

40

50

フロントページの続き

- 弁理士 上杉 浩
(74)代理人 100120525
弁理士 近藤 直樹
(74)代理人 100139712
弁理士 那須 威夫
(74)代理人 100158551
弁理士 山崎 貴明
(72)発明者 ジン・サン
アメリカ合衆国 カリフォルニア州 9 4 0 6 5 レッドウッド・シティ シャノン・ウェイ 5 3 0
ナンバー 4 1 0 2
(72)発明者 スリチャラン・カルーア・パリ・クマール
アメリカ合衆国 カリフォルニア州 9 4 0 4 1 マウンテン・ビュー ハイドン・アベニュー 1 8
0 2 アpartment 4
(72)発明者 ラジャ・バーラ
アメリカ合衆国 ニューヨーク州 1 4 5 3 4 ピッツフォード ウッドグリーン・ドライブ 6 5
審査官 千葉 久博
(56)参考文献 特開 2 0 1 8 - 7 3 3 9 3 (J P , A)
上西和樹, 外2名, “ 敵対的生成ネットワークを用いた, 3次元点群形状特徴量の教師なし学
習 ”, 研究報告コンピュータグラフィックスとビジュアル情報学 (CG), 日本, 情報処理学
会, 2018年06月14日, p.1-7
久保田涼介, 外1名, “ 全層畳み込みニューラルネットワークを用いた透明物体の輪郭抽出 ”,
電子情報通信学会技術研究報告, 日本, 一般社団法人電子情報通信学会, 2018年05月10日
, 第118巻, 第35号, p.41-46
榎木谷侑生, 外2名, “ Adversarial U-Net for Liver Segmentation ”, 電子情報通信学会技術研
究報告, 日本, 一般社団法人電子情報通信学会, 2018年03月12日, 第117巻, 第518号, p.
67-68
Christian F. Baumgartner, 外4名, "Visual Feature Attribution Using Wasserstein GANs", 2
018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 米国, IEEE, 201
8年06月23日, p.8309-8319
ZIQIANG ZHENG, 外4名, "Instance Map Based Image Synthesis With a Denoising Generativ
e Adversarial Network", IEEE Access, 第6巻, 米国, IEEE, 2018年06月20日, p.33654-3
3665
WANCHAO SU, 外4名, "Interactive Sketch-Based Normal Map Generation with Deep Neural
Networks", Proceedings of the ACM on Computer Graphics and Interactive Techniques,
第1巻, 第1号, 2018年07月25日, p.1-17
(58)調査した分野 (Int.Cl., D B 名)
G 0 6 T 7 / 0 0