

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7469237号
(P7469237)

(45)発行日 令和6年4月16日(2024.4.16)

(24)登録日 令和6年4月8日(2024.4.8)

(51)国際特許分類

G 0 6 T 7/00 (2017.01)

F I G 0 6 T 7/00 3 5 0 C

請求項の数 14 (全43頁)

(21)出願番号	特願2020-570843(P2020-570843)	(73)特許権者	518156417
(86)(22)出願日	令和1年6月19日(2019.6.19)		ズークス インコーポレイテッド
(65)公表番号	特表2021-528757(P2021-528757 A)		アメリカ合衆国 9 4 4 0 4 カリフォルニア州 フォスター シティー チェス ドライブ 1 1 4 9
(43)公表日	令和3年10月21日(2021.10.21)	(74)代理人	110001243
(86)国際出願番号	PCT/US2019/037967		弁理士法人谷・阿部特許事務所
(87)国際公開番号	WO2019/246250	(72)発明者	サラ タリク
(87)国際公開日	令和1年12月26日(2019.12.26)		アメリカ合衆国 9 4 4 0 4 カリフォルニア州 フォスター シティー チェス ドライブ 1 1 4 9
審査請求日	令和4年6月17日(2022.6.17)		ズークス インコーポレイテッド内
(31)優先権主張番号	16/013,729	(72)発明者	ジェームズ ウィリアム ベイジー フィルビン
(32)優先日	平成30年6月20日(2018.6.20)		アメリカ合衆国 9 4 4 0 4 カリフォル
(33)優先権主張国・地域又は機関	米国(US)		最終頁に続く
(31)優先権主張番号	16/013,748		
(32)優先日	平成30年6月20日(2018.6.20)		
	最終頁に続く		

(54)【発明の名称】 機械学習モデルの出力から推論されるインスタンスセグメンテーション

(57)【特許請求の範囲】

【請求項1】

機械学習（ML）モデルへの入力として、画像を提供するステップと、
前記MLモデルからの出力として、複数の特徴を含む特徴マップを受信するステップであって、前記複数の特徴は、複数の関心のリージョン（ROI）と、前記画像の一部分に関連付けられ、信頼スコアおよび第1の関心のリージョン（ROI）を含む前記複数の特徴のうちの個々の特徴とを含む、ステップと、
非極大抑制（NMS）の技法に少なくとも部分的に基づいて、出力されるROIを決定するステップであって、前記NMSの技法は、前記複数のROIのサブセットを決定することを含む、ステップと、
前記サブセットに関連付けられた前記画像の一部分を決定することに少なくとも部分的に基づいてインスタンスセグメンテーションを決定するステップであって、前記インスタンスセグメンテーションは、オブジェクトに関連付けられているとして前記画像の一部分を識別する、ステップと
を備えることを特徴とする方法。

【請求項2】

前記インスタンスセグメンテーションを決定するステップは、ROIの前記サブセットが、前記NMSの技法を介して抑制されているとして示されることを決定することに少なくとも部分的に基づくことを特徴とする請求項1に記載の方法。

【請求項3】

前記出力される R O I を出力すると決定するステップは、

前記出力される R O I が、前記複数の R O I の前記サブセットに関連付けられた前記信頼スコアの中から最大の信頼スコアに関連付けられると決定するステップと、

前記オブジェクトを表すとしての前記画像のリージョンと関連の前記出力される R O I を出力するステップと

を含むことを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記オブジェクトを表すとしての前記画像のピクセルを特定する前記インスタンスセグメンテーションに対応するマスクを出力するステップをさらに備えることを特徴とする請求項 1 に記載の方法。

10

【請求項 5】

前記 M L モデルへの入力として、第 1 の予め定義された数の、第 1 のオブジェクト分類に関連付けられる画像、および第 2 の予め定義された数の、第 2 のオブジェクト分類に関連付けられる画像を含む画像のバッチを提供することに少なくとも部分的に基づいて、前記 M L モデルをトレーニングするステップをさらに備え、

前記第 1 の予め定義された数および前記第 2 の予め定義された数が、前記複数の特徴の 1 つまたは複数の信頼レベルに少なくとも部分的に基づくことを特徴とする請求項 1 に記載の方法。

【請求項 6】

複数の分類に対応する各特徴に対して、複数の R O I を実質的に同時に出力する前記 M L モデルをトレーニングするステップであって、前記複数の分類が、車両、歩行者、または自転車のうちの 1 つまたは複数を含む、ステップをさらに備えることを特徴とする請求項 1 に記載の方法。

20

【請求項 7】

前記 M L モデルの正確度を示す前記 M L モデルに関連付けられた応答曲線を決定するステップであって、

前記画像における複数のオブジェクトを前記 M L モデルに提供すること、および

前記 M L モデルが異なるサイズのオブジェクトをどれくらい適切に分類しているかをトラッキングすること

を含む、ステップ

30

をさらに備えることを特徴とする請求項 1 に記載の方法。

【請求項 8】

前記出力される R O I、または前記インスタンスセグメンテーションのうちの 1 つに少なくとも部分的に基づいて、自律車両の動きを制御するための軌道を決定するステップをさらに備えることを特徴とする請求項 1、2、3、4、5、6、または 7 のいずれか一項に記載の方法。

【請求項 9】

コンピューターにおいて実行されると、請求項 1、2、3、4、5、6、または 7 のいずれか一項に記載の方法を実装する、コード化された命令を含むことを特徴とするコンピュータープログラム。

40

【請求項 10】

システムであって、

1 つまたは複数のプロセッサと、

前記 1 つまたは複数のプロセッサによって実行可能な命令を格納する 1 つまたは複数のコンピューター読み取り可能な媒体であって、前記命令は、実行されると、前記システムに、

機械学習 (M L) モデルへの入力として、画像を提供することと、

前記 M L モデルから、複数の特徴を含む特徴マップ、第 1 の関心のリージョン (R O I) を含む前記複数の特徴のうちの特徴、および第 1 の信頼スコアを受信することと、

非極大抑制 (N M S) の技法に少なくとも部分的に基づいて、第 2 の R O I を出力す

50

ることおよび前記第1のROIを抑制することを決定することと、

前記NMSの技法により抑制された1つまたは複数のROIに少なくとも部分的に基づいてインスタンスセグメンテーションを決定することであって、前記インスタンスセグメンテーションは、前記NMSにより抑制された前記1つまたは複数のROIに関連付けられた前記画像の一部分と、前記第2のROIに関連付けられた前記画像の一部分を含む、ことと

を含む動作を行わせる、コンピューター読み取り可能な媒体と
を備えたことを特徴とするシステム。

【請求項11】

前記動作は、オブジェクト分類に関連付けられたオブジェクトを表すとしての前記画像のピクセルを特定するマスクとして、前記インスタンスセグメンテーションを出力することをさらに含むことを特徴とする請求項10に記載のシステム。

【請求項12】

前記動作は、

前記インスタンスセグメンテーションに少なくとも部分的に基づいて、自律車両に環境の一部分を横切らせるように構成された軌道を決定することと、

前記自律車両に前記軌道を提供することと

をさらに含むことを特徴とする請求項11に記載のシステム。

【請求項13】

前記動作は、

前記MLモデルへの入力としておよび実質的に同時に、第1の予め定義された数の、第1のオブジェクト分類に関連付けられる画像、および第2の予め定義された数の、第2のオブジェクト分類に関連付けられる画像を含む画像のバッチを提供することに少なくとも部分的に基づいて、前記MLモデルをトレーニングすることをさらに含み、

前記第1の予め定義された数および前記第2の予め定義された数が、前記複数の特徴に関連付けられた1つまたは複数の信頼スコアに少なくとも部分的に基づいて決定されることを特徴とする請求項10に記載のシステム。

【請求項14】

前記動作は、

テスト画像、および関連付けられたグランドトゥールズに少なくとも部分的に基づいて、前記MLモデルをトレーニングすることであって、前記グランドトゥールズが、リファレンスリージョンおよびマスクを特定する、ことをさらに含み、

前記特徴マップが、前記画像の寸法の小部分である寸法を含むことを特徴とする請求項10に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、インスタンスセグメンテーションに関し、より詳細には、機械学習モデルの出力から推論されるインスタンスセグメンテーションに関する。

【背景技術】

【0002】

本特許出願は、2018年6月20日出願、シリアル番号16/013,729の「Machine Learning Techniques」という名称の米国実用特許出願の優先権を主張し、2018年6月20日出願、シリアル番号16/013,748の「Restricted Multi-Scale Inference for Machine Learning」という名称の米国実用特許出願の優先権を主張し、および2018年6月20日出願、シリアル番号16/013,764の「Instance Segmentation Inferred from Machine-Learning Model Output」という名称の米国実用特許出願の優先権を主張する。出願シリアル番号16/013,729、16/013,748、16/013,764は、参照によって本明細書にて完全に組み入れられる。

【0003】

コンピュータービジョンは、例えば、自律車両を動かすことなどのいくつかのアプリケーションに対してミッションクリティカルである。コンピューターに、人間の視覚を模倣する機能性を装備することは、画像を得て、画像から顕著な部分を特定し、コンピューターがさらなる動作を遂行するのに使用することが可能であるかたちにおいて、コンピューターに対して画像の顕著な部分を表すソフトウェアコンポーネントを構築することを含むことがある。コンピューターに上記の機能を装備するのに構築することが可能であるソフトウェアの1つのかたちは、ML（機械学習）モデルである。

【0004】

画像の顕著な部分を特定するMLモデルをトレーニングする先の試みは、欠陥のあるMLモデルに帰着した。例えば、MLモデルトレーニングのいくつかのかたちは、互いに近接するオブジェクト（たとえば、カメラのビューにおける別の歩行者の前を通り過ぎる歩行者）を正確に区別することが不可能であるMLモデルに帰着し、1つまたは両方のオブジェクトの無関係なおよび／または不正確なアイデンティフィケーションに帰着する。

【0005】

さらにその上、いくつかのMLモデルは、上記の欠陥のあるMLモデルよりもより正確なオブジェクトのアイデンティフィケーションを提供するが、リアルタイムのアプリケーションに役立つには、あまりにも多くの計算を必要とする、および／または特定の使用に都合がよいことがある高価な、専門化したコンピューティング装置を必要とすることがある。例えば、コンピュータービジョン用のMLモデルを使用する自律車両は、ビデオストリームを受信することがあり、50ミリ秒以下ごとに自律車両を動作させる決定をすることが必要とされることがある。オブジェクトを特定する今までに述べたMLモデルのいくつかに必要なとされる計算時間のために、検出されたオブジェクトは、オブジェクトが検出される時間によって著しく十分に移動していることがあり、従って、オブジェクト検出は、意思決定に対してははや当てにならない。例えば、今までに述べたMLモデルのいくつかは、100ミリ秒を超える計算時間を必要とすることがある。

【0006】

詳細な説明は、添付の図面を参照して説明される。図面において、参照符号の最も左の数字（複数可）は、参照符号が最初に現れる図面を特定する。異なる図面における同一の参照符号は、同様のまたは同じアイテムを示す。

【図面の簡単な説明】

【0007】

【図1】例示的なシナリオの例示的な空からの眺めを含む例示的なシナリオの例示的な画像を例示する。

【図2A】例示的な画像と、上記の画像がグリッドとしてどのように離散化されるかと、を例示する。

【図2B】例示的なROI（region of interest）と、ROIが生成された画像の部分とを例示する。

【図3】2つの領域を示す例示的なグランドトゥールースと、第2の、より小さい領域を示す例示的なシュリンクしたグランドトゥールースとを例示する。

【図4A】例示的なROIと、ROIに関連付けられた例示的な信頼スコアと、ROIが生成された画像の部分とを例示する。

【図4B】例示的なROIと、ROIに関連付けられた例示的な信頼スコアと、ROIが生成された画像の部分とを例示する。

【図4C】例示的なROIと、ROIに関連付けられた例示的な信頼スコアと、ROIが生成された画像の部分とを例示する。

【図4D】例示的なROIと、ROIに関連付けられた例示的な信頼スコアと、ROIが生成された画像の部分とを例示する。

【図5】グランドトゥールースによって示された領域に関し、グランドトゥールースによって示された領域に対する例示的なROIのアライメント度に関連付けられた図4A～4Dの例示的なROIを例示する。

10

20

30

40

50

【図 6 A】本明細書にて述べられるノンマキシマム・サプレッション・リプレースメント (non-maximum suppression replacement) の技法に係る ML モデルをトレーニングするための例示的な処理のフロー図を例示する。

【図 6 B】本明細書にて述べられるノンマキシマム・サプレッション・リプレースメントの技法に係る ML モデルをトレーニングするための例示的な処理のフロー図を例示する。

【図 7 A】画像（複数可）内の種々のサイズを有するオブジェクトに対する第 1 の ML モデルの例示的な第 1 の応答曲線を例示する。

【図 7 B】画像（複数可）内の種々のサイズを有するオブジェクトに対する第 2 の ML モデルの例示的な第 2 の応答曲線を例示する。

【図 8 A】本明細書にて述べられる技法に係る、入力として第 1 の ML モデルに提供される例示的な画像と、第 1 のスケーリングされた画像を受信することに応答する第 1 の ML モデルによって生成される ROI（複数可）とを例示する。

10

【図 8 B】本明細書にて述べられる技法に係る、第 2 のスケーリングされた画像としてリスケーリングされ、入力として第 1 の ML モデルに提供される図 8 A の例示的な画像と、第 2 のスケーリングされた画像を受信することに応答する第 1 の ML モデルによって生成される ROI（複数可）とを例示する。

【図 8 C】本明細書にて述べられる技法に係る、入力として第 2 の ML モデルに提供される図 8 A の例示的な画像と、例示的な画像を受信することに応答する第 2 の ML モデルによって生成される ROI（複数可）とを例示する。

【図 9 A】例示的な画像の一部に対して生成された例示的な ROI を例示する。

20

【図 9 B】例示的な画像における同一のオブジェクトに関連付けられた例示的な複数の ROI を例示する。

【図 9 C】本明細書にて述べられる技法による例示的なオブジェクトセグメンテーションを例示する。

【図 10】本明細書にて述べられる技法を実装する例示的なシステムを例示する。

【発明を実施するための形態】

【0008】

本明細書にて述べられる技法は、オブジェクトが、例えば、自律車両の制御などのアプリケーションにおける使用に対してリアルタイムに検出され得るように、オブジェクト検出の正確度を増すことと、正確なオブジェクトのアイデンティフィケーションを取得するための計算時間を減らすこととによって、コンピュータビジョンを改善する。本明細書にて述べられる技法の他のアプリケーションは、例えば、ビデオゲーム、拡張現実感などが予想される。

30

【0009】

本明細書にて述べられる技法は、画像を ML モデルに提供することと、ML モデルから、画像の異なる部分に対する複数の ROI (region of interest) を受信することを含むことがある。今までに述べた ROI は、ML モデルが画像においてオブジェクトの存在であると思うものを特定するどんなかたちでもあり得る。例えば、ROI は、検出されたオブジェクトに関連付けられているとして特定されるピクセルを示すボックス（「バウンディングボックス」）、検出されたオブジェクトに対応するピクセルを含むマスクなどを含むことがある。

40

【0010】

場合によっては、追加として、または代替えとして、ML モデルは、複数の ROI の各々に関連付けられた信頼スコア（または信頼情報）を出力することがある。例えば、ML モデルは、オブジェクトがあることを画像の一部から検出することがあり、そのオブジェクトが画像の中のどこにあるかを示す ROI を生成することがある。追加として、または代替えとして、ML モデルは、ML モデルが画像において顕著なオブジェクトを実際に特定したことを ML モデルがどれぐらい確信しているか、および／または ROI がオブジェクトにどれぐらい上手くフィットするかを本質的に示す信頼スコアを生成することがある。例えば、信頼スコアは、0 と 1 の間の値であることがあり、ここで、0 は、ML モデル

50

が、オブジェクトがROIに現れていることに全く確信がないという、MLモデルによるインディケーションを表し、1は、MLモデルが、オブジェクトがROIに現れていることを強く確信していることを示すが、他の順列は想定される。換言すれば、MLモデルは、オブジェクトがあるかもしれないと考えるのはどこかのインディケーションと、MLモデルが、オブジェクトを正しく特定したことをどれぐらい確信しているか、および／またはROIが、オブジェクトがイメージの中のどこにあるかをどれぐらい上手く指すかを示すスコアとを出力する。

【0011】

本明細書にて述べられる技法のいくつかは、より良いROI（オブジェクトが画像の中のどこにあるかをより正確に特定するROI）、および／または（たとえば、オブジェクトを含まないROIに対して0により近いスコアおよび／または顕著なオブジェクトを示すROIに対して1により近いスコアを生成する）より正確な信頼スコアを生成するために、および上記の正確度のROIを達成する計算時間を削減するために、MLモデルをトレーニングすることに向けられる。

【0012】

MLモデルは、ニューラルネットワーク、例えば、ランダムフォレストおよび／または向上された決定木のアンサンブル、有向非巡回グラフ（DAG）（たとえば、ノードがベイジアンネットワークとして組織される場合）、深層学習アルゴリズム（複数可）など、例えば、ANN（人工ニューラルネットワーク）（たとえば、RNN（リカレントニューラルネットワーク）、ResNet（residual neural network））など、DBN（deep belief network）、DSN（deep stacking network）などを含むことがある。MLモデルをトレーニングするための損失関数は、例えば、オブジェクトを表すとしてグラントゥールスによって示された画像の領域に対するROIのアライメント度（degree of alignment）を決定することを含むことがある。場合によっては、グラントゥールスによって示された領域に対するROIのアライメント度を決定することは、ROIがグラントゥールスにどれぐらい上手く「フィット」するかを決定するための1つのメトリックであるROIのIoU（intersection over union）を決定することを含むことがある。グラントゥールスによって示された領域に対するROIのフィットに関する他のインディケーションが、使用されることがある。場合によっては、グラントゥールスは、リファレンスリージョン（reference region）と呼ばれることが可能である。

【0013】

MLモデルをトレーニングすることが、数万、さらに数十万以上の画像をMLモデルに提供することと、重みに対する上記の訂正をすることを含むことがあるので、MLモデルが最も間違っているROIを特定することは、有利である。今述べたことは、MLモデルをトレーニングするのに費やす時間を大いに減らし、「非常に間違った」ROIおよび／または信頼スコアを考慮に入れるためにMLモデルになされる訂正が、「正しい」ROIおよび／または信頼スコアのラーニングを強化することによって洗い流されないもので、MLモデルの正確度を増すことがある。

【0014】

場合によっては、本明細書にて述べられる技法は、MLモデルをトレーニングするための特定のエグザンプルを選択することを含むことがある。今述べたエグザンプルを選択することは、例えば、信頼スコア（たとえば、最大から最小までの信頼スコア）、および／または信頼スコアにおける誤差（たとえば、NMSに従って抑制されたROIに対して、ROIに関連付けられた信頼スコアの誤差）によって複数のROIをソートすることと、上位n個のROIを選ぶこととを含むことがあるHEM（hard example mining；ハード・エグザンプル・マイニング）を含むことがある。場合によっては、ハード・エグザンプル・マイニングによってエグザンプルを選択することは、最大の信頼スコア（複数可）に関連付けられたROIを除外することがある。追加として、または代替えとして、技法は、n個のランダムなROIを選ぶことを含むことがある。場合によっては、数nは、画像におけるポジティブエグザンプル（positive example）（たとえば、画像に表された

10

20

30

40

50

【 0 0 1 5 】

20

30

40

【 0 0 1 8 】

50

によっては、第1のバッチは、数万または数十万の画像を含むことがある。

【0019】

場合によっては、MLモデルをトレーニングすることは、第1のステージに続く第2のステージを含むことがある。第2のステージの間、MLモデルをトレーニングすることは、ハードエグザンプルを含むトレーニングサブセットを使用することを含むことがある。場合によっては、追加として、第2のステージは、Focal Loss関数を使用して、MLモデルをトレーニングすることを含むことがある。場合によっては、Focal Loss関数は、上手く分類された(well-classified)ROIに対して計算された損失(誤差)が重みを減らされ、下手に分類された(poorly-classified)ROIに対して計算された損失がより少なく重み付けされるように、交差エントロピー損失関数(またはどんな他の損失関数でも)によって計算された損失を修正することがある。

10

【0020】

場合によっては、MLモデルの受容野(receptive field)は、MLモデルに、下手な信頼スコアおよび／または誤差値に関連付けられた受容野に対して、広すぎるオブジェクトに対するROIを生成させることがある。例えば、人間が、鼻が絵画に対して触れている状態で立っていたならば、絵画を確認すること、ましてや絵画における特定のオブジェクトを確認することは、とても困難であろう。今述べたことは、絵画の顕著な部分が、人間の受容野の外にあるからである。

【0021】

同様に、MLモデルは、MLモデルの受容野の外にあるオブジェクト(たとえば、オブジェクトを「見て」、MLモデルが「見る」ものを理解するためにMLモデルに対する受容野に入る十分なオブジェクトに対して、大きすぎるオブジェクト)を下手に検出することがある。

20

【0022】

今述べた問題を矯正するためのいくつかの技法は、MLモデルへの入力として画像を提供することと、サイズ範囲内のオブジェクトを得るためのMLモデルを使用することと、続いて画像をダウンスケーリングすることと、以前はサイズ範囲外であったオブジェクトが、(スケーリングされた後に)今やサイズ範囲内に入るようにMLモデルを通じて再実行することとを含んでいた。今述べたことは、反復的に繰り返されることがある。画像を小さくスケーリングすることは、スケールダウンされた画像において、より小さく見える大きなオブジェクトを作り、MLモデルの受容野内に入ることを可能にすることに注意する。入力画像のサイズに実質的に同様である受容野によってMLモデルをトレーニングすることが可能であることがある一方、そのようなMLモデルは、例えば、自動運転における使用に対して、遅い、または十分に応答しないことがある。

30

【0023】

追加のまたは代替の場合にて、本明細書にて説明される技法は、画像における見かけのオブジェクトサイズを介してオブジェクトを分類する(たとえば、ROIおよび／または信頼スコアを作り出す)ために、第1のMLモデルの正確度を示す第1のMLモデルに関連付けられた応答曲線を決定することを含むことがある。場合によっては、今述べた応答曲線を作り出すことは、画像(複数可)における複数のオブジェクトを第1のMLモデルに提供することと、第1のMLモデルが異なるサイズのオブジェクトをどれくらい上手く分類するかをトラッキングすることを含むことがある。技法は、ROIの正確度を決定し、ROIの正確度およびサイズ(たとえば、高さ、幅、面積)に関連付けることを含むことがある。少なくとも場合によっては、MLモデルの出力サイズは、入力スケールサイズに基づいた最適領域に基づいて制約されることがある。非限定の例として、入力として640×480の画像があるMLモデルは、約100×100～200×200のピクセルサイズを有するROIを出力するように強いられることがあるが、入力として940×940の画像があるMLモデルは、約250×250～300×300ピクセルのピクセル寸法を有するROIを出力するように強いられることがある。

40

【0024】

50

これらの正確度／ROIサイズのペアは、第1のMLモデルに対して決定された各ROIに対して、1つにされることがある。技法は、しきい値の正確度を満たす、または超える正確度に関連付けられたROIの範囲を特定することを含むことがある。今述べたサイズの範囲は、第1のMLモデルが、サイズの範囲のオブジェクトに対して「良い」ROIを決定することを示すことがある。場合によっては、第1のMLモデルは、サイズ範囲内のサイズを有するROIを出力することがあり、MLモデルによって決定される他のROIを抑制することがある。

【0025】

場合によっては、第1のバッチの画像は、スケールファクター（たとえば、0.75、0.5）を使用してスケールダウンされることがあり、スケーリングされた画像のバッチは、入力として、第2のMLモデルに提供されることがある。同様に、第2の応答曲線と、第2の範囲のサイズのとは、スケーリングされた画像のバッチに少なくとも部分的に基づいて、第2のMLモデルに対して決定されることがある。

【0026】

場合によっては、実行時に、画像は、入力として第1のMLモデルに提供されることがあり、画像は、スケーリングされ、入力として第2のMLモデルに提供されることがあり、および第1のMLモデルによって出力されるROIは、第1の範囲内のROIに制限されることがあり、第2のMLモデルの出力は、第2の範囲内のROIに制限されることがある。MLモデルは、異なるMLモデルに関連付けられた異なる範囲のROIがオーバーラップする、および／または互いに接するようにトレーニングされることがあって、オブジェクトサイズのより大きな範囲は、複数のMLモデルによって検出されることがある。

【0027】

簡単に言えば、技法は、少なくとも2つのMLモデルをトレーニングし、各MLモデルが「良い」ROIを生成するROIのサイズの範囲を決定することを含むことがある。そうすれば、画像をMLモデルに繰り返し提供すること、出力を受信すること、画像をスケーリングすること、それをMLモデルに再提供すること（re-providing）、第2の出力を受信することなどの代わりに、処理は、各MLモデルが、サイズの小さい範囲に対して最も良く応答するようにトレーニングされ、ROIをその小さい範囲の中に出力するように強いられる場合に、同一の画像のスケーリングされたバージョンを複数のMLモデルに提供すること（たとえば、スケーリングされた画像が、同一のアンスケーリングされた画像（unscaled image）のスケーリングされたバージョンである場合に、第1のスケーリングされた画像を第1のMLモデルに、第2の画像を第2のMLモデルになど提供すること）によって、短くなり、正確度が上げられることがある。一般に、上記の技法は、メモリー要件を減らす（たとえば、GPUのメモリー要件を減らした）一方、画像におけるオブジェクトを検出する速さおよび性能を改善することがある。

【0028】

さらに、本明細書にて説明される技法は、MLモデルへの入力として画像を提供し、MLモデルからの出力として、画像において検出された物体に関連付けられたROIと特定のピクセル（または画像の他の部分）を特定するオブジェクトセグメンテーション（たとえば、マスク、輪郭のセット）との両方を受信するための技法を含むことがある。画像を区分する（たとえば、マスク、または直線および／もしくは曲線のセットを出力することによって、たとえば、どのピクセルが、検出されたオブジェクトに対応するかを特定する）ために、第2のMLモデルおよび／または第2の動作を必要とする代わりに、本明細書にて述べられる技法は、実質的に同時に（たとえば、技術的なトレランスの範囲内で）、オブジェクトを検出し（たとえば、ROIを決定し）、画像を区分する。

【0029】

場合によっては、インスタンスセグメンテーション（たとえば、ピクセルベースによりピクセル上の画像を区分すること）のための技法は、MLモデルが、画像の複数の部分に対して、複数のROIと、関連付けられた複数の信頼スコアとを決定するまで、MLモデルにおいて画像を受信し、画像の一部（たとえば、ピクセル、ピクセルのクラスター）に

10

20

30

40

50

対するROIと、関連付けられた信頼スコアとを決定することを含むことがある。場合によっては、MLモデルは、出力のために、最大の信頼スコアに関連付けられたROI（「出力ROI」）を選択することによって画像を検出することがある。出力ROIが特定するオブジェクトを表す画像のリージョンを区分するために、技法は、信頼スコアに関連付けられた複数のROIについてのサブセットが、信頼スコアのしきい値を満たすまたは超える、最大の信頼スコアに関連付けられたROIと実質的にオーバーラップする、および／または最大の信頼スコアに関するしきい値信頼の範囲内にあることを決定し、（たとえば、ピクセルをマスクに加えることによって、ピクセルを含むように曲線を修正することによって）ROIが決定された部分から画像のセグメンテーションまでつなげることを含むことがある。

10

【0030】

違う説明では、（マスクであり得る）インスタンスセグメンテーションにおける画像のデータ（たとえば、ピクセル）を含むかどうかを決めるために、技法は、その特定のピクセル（または特徴マップ（feature map）の部分）に関連付けられたROIが、代表的なROIと、それに対応する信頼値とに実質的に類似するかどうかを決定することを含むことがある。少なくとも一例にて、インスタンスは、非最大抑制（非極大抑制とも呼ばれる）の下に抑制されるだろう単一のROIに対応するすべてのピクセルを含むことがある。

【0031】

場合によっては、オブジェクトに対してインスタンスセグメンテーションを生成するためにMLモデルをトレーニングすることは、グランドトゥルースのマスクと、オブジェクトの関連付けられたROIとに基づいてトレーニングすることを含むことがある。上記の例にて、損失関数は、予め定義された値を超える信頼スコアを有するROIを生成するために、マスクの外側にある画像の部分にペナルティーを与えるように構成されることがある。換言すれば、画像のピクセル（またはリージョン）の上記MLモデルは、上記のピクセルが、ピクセル（またはリージョン）に関連付けられたオブジェクトのマスクの外側に入るならば、たとえROIが適切であっても、ROIを出力すべきではない。本明細書が、ピクセルのマスクを出力することに帰着するインスタンスセグメンテーションを主に述べるが、画像のサブセットを特定するためのどんな適した方法でも（たとえば、点、線、曲線、および／または他のエッジインジケータのセットを出力すること）、使用されることがあると想定される。

20

【0032】

例示的なシナリオ

図1Aは、例示的な画像100と、例示的な画像100が生成されることがある例示的なシナリオ102の例示的な鳥瞰図とを例示する。例示的なシナリオ102は、自律車両104と、自律車両104と同一の環境（たとえば、例示的なシナリオ102の鳥瞰図に描かれている運転中の地面を含むことがある例示的な画像100に描かれる街区）において動いている3台の他の車両106、108、および110を含むことがある。自律車両104は、センサーデータを、自律車両104の1つまたは複数のセンサーから受信することがある。自律車両104は、自律車両の動きを制御するための軌道を決定するセンサーデータを使用することがある。センサーデータは、例えば、例示的な画像100などの画像を含むことがある。

30

40

【0033】

場合によっては、自律車両104は、運転手（または乗員）がいかなる時点においても車両を制御することが期待されない、全行程にすべてのセーフティクリティカルな機能を実行する性能がある車両を記述する、米国連邦道路交通安全局によって発行されるレベル5分類に従って、動作するように構成された自律車両であり得る。しかしながら、他の例において、自律車両104は、現在存在するまたは将来開発される他のどんなレベルまたは分類でも有する完全なまたは部分的な自律車両であり得る。さらにその上、場合によっては、説明されるコンピュータビジョンの技法は、自律的ではない車両によっても同様に使用可能であることがある。

50

【0034】

本明細書にて述べられる技法に従って、自律車両104は、自律車両104のセンサー112から画像を受信することがある、例えば、他の車両106、108、および110など、画像から検出されるオブジェクト（複数可）を検出することがある、および／または画像において検出されたオブジェクトを区分することがある。例えば、自律車両104は、他の車両106、108、110、運転可能な地面、交通信号などを特定し、区分することがある。場合によっては、今述べた検出することおよび／または区分することは、自律車両104のコンピューティングデバイス（複数可）118上において実行している認識エンジン116のMLモデル114に、入力としての画像を提供することによって達成されることがある。

10

【0035】

場合によっては、認識エンジン116は、例えば、MLモデル114など、自律車両104の環境から収集されたセンサーデータから物体を検出する、特定する、区分する、分類する、および／もしくは追跡するための1つまたは複数のMLモデル、ならびに／または他のコンピューター実行可能な命令を含むことがある。場合によっては、自律車両104の認識エンジン116は、車両104の1つまたは複数のセンサー112からセンサーデータを受信し、センサーデータから認識データ（perception data）を決定し、ならびにグローバルマップ上の自律車両104の位置をローカライズする、1つまたは複数の軌道を決定する、および／またはパスもしくはルートを横切る自律車両104の動きを制御するプランナー120による使用のために認識データをプランナー120に送信する認識エンジンを含むことがあるが、どんな上記の動作でも、種々の他のコンポーネントにおいて行われることがある（たとえば、ローカライゼーションは、示されないが、ローカライゼーションエンジンによって行われることがある）。

20

【0036】

例えば、認識エンジン116は、環境中の物体を検出し、物体（たとえば、乗用車、セミトラック（semi-truck）、ピックアップトラック、人、子供、犬、ボール）を分類することがあり得る。さらに、認識エンジン116は、物体のトラック（track）（たとえば、物体に関する履歴の、現在の、および／もしくは予測される進路、位置、速度、ならびに／または加速度）を決定することもある。場合によっては、追加としてまたは代替えとして、認識エンジン116は、（車両104をローカライズするためにどんなセンサーデータでも使用することがある、絵で示されない）ローカライゼーションエンジンによって決定される自律車両104の位置、自律車両104の近くにある物体に関係したデータ、車両の目的地を指定するルートデータ、車道の特性（たとえば、自律車両をローカライズするのに有用な様々なセンサーモダリティ（sensor modality）において検出可能な特徴）を特定するグローバルマップデータ、車両の近くに検出された特性（たとえば、建物、木、壁、消火栓、一時停止の標識、および種々のセンサーモダリティにおいて検出可能な他の特徴に関するロケーションならびに／または寸法）を特定するローカルマップデータなどを決定することがある。認識エンジン116によって生成されるデータは、「認識データ」と総称されていることがある。認識エンジン116が、画像からオブジェクトを特定したらおよび／または区分したら、および／または他の認識データを決定したらすぐに、認識は、オブジェクトの検出および／またはセグメンテーションを含む認識データを、プランナー120に提供することがある。

30

40

【0037】

場合によっては、プランナー120は、自律車両104の動きを制御するための軌道を生成するために、特定された、および／または区分されたオブジェクトを含む認識データを使用することがある。例えば、プランナー120は、自律車両104に対して第1のロケーションから第2のロケーションまでのルートを決

50

ブコントロール信号を生成するのに使用され得る自律車両 104 の軌道として、可能性がある軌道のうちの 1 つを選択することがある。

【0038】

ML モデルによる例示的な ROI (REGION OF INTEREST) 生成

図 2 A は、出力グリッド (output grid) 200 の各セルが画像 100 の一部を特定する例示的な画像 100 および例示的な出力グリッド 200 を例示する。画像 202 の一例の部分は、画像 100 の中心の近くに強調 (ボールド) される。本明細書にて述べられるような、「画像の一部」は、画像の単一のピクセル、および／または画像のピクセルの集まりを含むことがあることが想定される。場合によっては、機械学習モデルの出力は、特徴マップであり、個々のセルは、特徴マップの一部を表すことがある。上記の特徴マップは、複数のチャンネル、モデルによって決定される種々の要素 (複数可) (たとえば、信頼スコア、関心のリージョンなど) に関連付けられた各チャンネルを有することがある。図 2 A は、「画像の一部」、たとえば、例示的な部分 202 が画像 100 のピクセルの集まりを含む例を例示する。特徴マップの一部 (たとえば、セル) は、画像の一部に関連付けられることがある。それゆえ、例示的な部分 202 は、例示的な出力グリッド 200 の例示的なセル 202 と呼ばれることがある。

10

【0039】

場合によっては、例示的な出力グリッド 200 は、ML モデルによる出力として、例示的な画像 100 を離散化するやり方であり得る。例えば、ML モデルは、画像を受信し、出力グリッド 200 のセルごとに、1 つまたは複数の ROI および関連する信頼レベルを出力するように構成されることがある。少なくとも場合によっては、上記の出力グリッド 200 は、 $m \times n$ のピクセルの画像に従って、 $m/4 \times n/4$ のセルに離散化されることがある。場合によっては、セルは、4 ピクセル $\times$ 4 ピクセルであり得るが、どんな他の離散化でも想定される。場合によっては、上に述べたように、例示的な画像 100 は、特徴マップの複数の部分に離散化されることが可能である。すなわち、例は、セルのグリッドに限定されず、種々の実装が本明細書において想定される。

20

【0040】

場合によっては、ML モデルは、画像の各部分に対して (たとえば、図 2 A の各セルに対して)、ROI および／または信頼スコアを生成することがある。例えば、ML モデルは、例示的な部分 202 に対して、ROI および／または信頼スコアを決定することがある。場合によっては、ML モデルは、画像の一部の境界を越えて延びる ROI の境界を有する ROI を生成することがある。少なくとも場合によっては、各出力セルは、どんな 1 つまたは複数の分類に対してでも、中心の位置 (たとえば、 $\langle u, v \rangle$ 画像座標の位置)、エクステント (たとえば、幅および／または高さ)、および／または信頼レベルによって示される ROI と対応することがある。非限定の例として、各セルは、自動車、歩行者、自転車運転者、トラック／バス、交通信号、および／または停止標識の分類の各々に対して、中心、エクステント、および信頼に関連付けられることがある。

30

【0041】

図 2 B は、「自動車」の分類に関して、画像の例示的な部分 (すなわち、画像におけるセル)、および関連付けられた ROI (region of interest) を例示する。図 2 B は、ROI をバウンディングボックスとして例示するが、オブジェクトを表す画像のリージョンが、そうでなければ (たとえば、マスクによって) 示されることがあると理解される。本明細書にて述べられるように、ML モデルは、部分 204 (たとえば、セル 204) に対する ROI 204'、部分 206 に対する ROI 206'、および部分 208 に対する ROI 208' を決定することがある。場合によっては、ML モデルは、ROI 204' に関連する第 1 の信頼スコア、ROI 206' に関連する第 2 の信頼スコア、および ROI 208' に関連する第 3 の信頼スコアを決定することがある。信頼スコアは、関連する ROI が、オブジェクト (ここでは、自動車) を表す画像のリージョンを正確に表す確率を示すことがある。ROI 204'、206'、および 208' の各々は、それぞれ、別個のオブジェクト、すなわち、車両 106、108、および 110 を表す画像の別個のリージョンを特定

40

50

する。場合によっては、ROI204'は、MLモデルによって、画像100における車両106の「検出」と考えられることがある。

【0042】

例示的なML（機械学習）モデルトレーニング技法

図3は、例示的な画像100の部分300、第1のオブジェクト（すなわち、車両108）を表す画像の第1の領域を特定するグラントゥールス302、第2のオブジェクト（すなわち、車両110）を表す画像の第2の領域を特定する例示的なグラントゥールス304、ならびに2つのスケーリングされたグラントゥールス302'および304'を例示する。例示的なスケーリングされたグラントゥールス302'は、グラントゥールス302の中央の30%（「中央の30%のボックス302」）を定義する領域であり、例示的なスケーリングされたグラントゥールス304'は、グラントゥールス304の中央の30%を定義する領域である。もちろん、例示的なスケーリングされたグラントゥールス302'が、グラントゥールス302の中央の30%に対応するとして説明されるが、どんな百分率でも本明細書にて想定される。

【0043】

次の議論は、MLモデルをトレーニングするための技法と、MLモデルが、ROIを、および／または、より正確である信頼スコアを生成するために、MLモデルをトレーニングするやり方とに関するものである。場合によっては、一般に、図3～6は、トレーニングの間、MLモデルの種々の出力にペナルティーを与えるやり方を決定するための技法を例示する。

【0044】

図3に例示されるように、MLモデルは、画像100においてオーバーラップするとき（すなわち、オクルージョン）に現れるオブジェクト間の、例えば、車両110と112との間の違いを明らかにするのを可能にするような、今述べたスケーリングされた30%のリージョンを使用するだけでトレーニングされることがある。部分的に、グラントゥールス302がグラントゥールス304のオーバーラップをオクルージョンするので、上に述べた技法に従ってMLモデルをトレーニングすることは、画像の一部に関連されるべきオブジェクト（すなわち、出力するROI）を決めるときのあいまいさを避けることがあることに注意する。

【0045】

図4A～4Dは、例示的なROIと、ROIに関連付けられた例示的な信頼スコアと、ROIが生成された画像の部分とを例示する。

【0046】

図4Aは、例えば、「自動車」などの特定の分類について、MLモデルが部分400に対して決定した例示的なROI400'と、MLモデルがROI400'に対して決定した信頼スコア402とを例示する。場合によっては、信頼スコアは、MLモデルによって決定される、信頼スコアが関連付けられているROIが画像におけるオブジェクトを正確に特定する確率を示す。例えば、信頼スコアは、ROIがオブジェクトを完全に表すピクセルを含む領域にフィットする確率を示すことがある（たとえば、検出されたオブジェクト以外のものを表す、ROIによって特定されるピクセルがなく、ROIのエッジは、オブジェクトの境界と交わることなしにオブジェクトのエッジを囲み、オブジェクトの境界とROIの境界との間の最小量の空間にする）。

【0047】

図4Aにおいて描かれる例にて、信頼スコア402は、「0.97」である。場合によっては、MLモデルは、0.00と1.00との間の信頼スコアを決定することがあるが、他の値は想定される。今述べた例にて、0.00は、MLモデルが、画像の一部に対して生成したROIがオブジェクトを特定するとは思わないことを示すことがあり、および／または1.00は、MLモデルが、完ぺきなアライメント度によって、ROIがオブジェクトを特定するのを完全に確信していることを示すことがある。

【0048】

図4Bは、MLモデルが部分404に対して決定した例示的なROI404'と、MLモデルがROI404'に対して決定した信頼スコア406とを例示する。ROI404'は、それがわずかに小さ過ぎるとして、ROI400'と同様に上手くオブジェクト（すなわち、車両110）を表すリージョンに、フィットしていないことに注意する。今述べた例にて、信頼スコア406は、ROI404'が、完ぺきに、またはROI400'と同様に上手く、オブジェクト（すなわち、車両108）を表すリージョンにフィットしていないので、少なくとも最大の信頼スコア、信頼スコア402よりも正しく低い。さらに、部分400は、図4B～4Dに対して参照点として示され、これらの図のROIが、何とかして部分400に関連付けられることを意図していないのを継続することにも注意する。

【0049】

図4Cは、MLモデルが部分408に対して決定した例示的なROI408'と、MLモデルがROI408'に対して決定した信頼スコア410とを例示する。ROI408'は、それが大き過ぎるとして、上手く車両108を表すリージョンにフィットせず、単なる車両108よりも多くのオブジェクトを囲んでいることに注意する。平凡なフィットのために適切に、より低かった信頼スコア406とは異なり、信頼スコア410は、車両108を表す画像のリージョンと比べて、ROI408'の下手なフィットに対して高すぎると考えられることがある0.94の値を反映する。

【0050】

図4Dは、MLモデルが部分412に対して決定した例示的なROI412'と、MLモデルがROI412'に対して決定した信頼スコア414とを例示する。ROI412'は、それが非常に小さ過ぎるとして、上手くオブジェクト（すなわち、車両108）を表すリージョンに、フィットしていないことに注意する。信頼スコア414は、車両108を表す画像のリージョンと比べて、ROI412'の下手なフィットを適切に考えている0.13の値を反映する。

【0051】

図5は、NMS（non-maximum suppression；非最大抑制）の再割り当てに対してリージョンを選択する例を例示する。一般に、画像においてポジティブエグザンプルの疎のために、MLモデルは、より良い推定を出力するために、ポジティブエグザンプルとネガティブエグザンプルとのバランスを使用してトレーニングされることがある。さらに、トレーニングは、「ハード」エグザンプル（すなわち、正しくはないが、正しいことの高信頼を提供するエグザンプル）を組み入れることもある。場合によっては、同じ個数のポジティブエグザンプル、ネガティブエグザンプル、およびハードエグザンプルが、トレーニングのためのネットワークに提供される。少なくとも場合によっては、すべてのセル（画像の一部）の信頼レベルは、ソートされることがあり、誤ったROIを生成する最高信頼は、ハードエグザンプルとして選択されることがある。図3に関して例示された技法を使用するとき、最高信頼に関連付けられたROI内に入るが、中央の30%の部分ではなかった画像の一部分（たとえば、セル）は、一般に、例えば、「ハードエグザンプル」のようなセルをナイーブに示すことによってペナルティーが与えられることがある（すなわち、ネットワークは、それらが高い信頼レベルを有するが中央の30%に入らないので、今述べたセルを有するROIを出力しないように明示的にトレーニングされることがある）。図3に関して説明された技法が違いを明らかにする（より良いROIを提供する）のを助けることがあるが、これらの部分にナイーブにペナルティーを与えることによってモデルをトレーニングすることは、下手な性能（たとえば、悪い推定）を導くことがある。以下に詳細に説明するように、モデルをインテリジェントにトレーニングすることは、上記のナイーブな仮定を避けることがある。

【0052】

場合によっては、MLモデルは、ハードエグザンプルとして示された画像の一部分（たとえば、セル）が、NMSの下に抑制されるROIに関連付けられると決定することがある。換言すれば、モデルは、ハードエグザンプルとして選択されたセルが選択されるべきではなく、別のハードエグザンプルが、真のハードエグザンプルにペナルティーを与える

10

20

30

40

50

このために、損失をバックプロパゲーションするために選ばれるべきであると決定することがある。少なくとも場合によっては、良いボックスを出力したが、ハードエグザンプルとして選択されたセルは、代わりに良い（すなわち、損失が、セルに対応する出力を改良するためにバックプロパゲーションされるような）エグザンプルとして再配置されることがある。一般に、上記の処理は、NMS再割り当て（NMS reassignment）と呼ばれることがある。NMS再割り当ての支配下にあるべきである画像の部分、およびあるべきではない画像の部分に関する種々の例が、図5にて例示される。

【0053】

図5の至る所に、点線によって表された参照ROI（グラントゥールスのROI、および／またはNMS（non-maximum suppression）によって決定されたROI）（たとえば、参照500）と、例示的なROIのアライメント度とが、参照500に関して与えられている。例示的なROIが生成された画像の一部分は、目に見える前後関係に対して各々に再び現れる部分400を有して、前後関係に対して図5にて再び生成され続ける。場合によっては、生成された複数のROIのうち、画像におけるオブジェクトに対応することがある出力するROIを選択するために、MLモデルは、複数のROIをNMSすることがある、すなわち、MLモデルは、最大の信頼スコアに関連付けられたROIを決定することがあり、オブジェクトに関連付けられた出力するROIとして、そのROIを特定することがある。場合によっては、あるオブジェクト分類の複数のオブジェクトが画像において表される一方、MLモデルは、オーバーラップする（たとえば、しきい値のアライメント度を満たすまたは超える互いに関してアライメント度を有する）ROIのクラスターを決定することがあり、それによってオブジェクトが、ROIのクラスターの周りの画像において画像のどこかに表される尤度を示し、およびROIのクラスターをNMSしてオブジェクトに対する出力ROIを決定することがある。換言すれば、単一のオブジェクトに関連するROIを出力することは、オーバーラップする複数のROIのうち、最大の信頼スコアに関連付けられたROIを決定することに基づくことがある。MLモデルは、すべてのオブジェクトが特定され、NMSの支配下になるまで、クラスターを特定することと、クラスターに関してNMSを行うこととを繰り返すことがある。今述べたことは、ROIが検出される各オブジェクト分類に対してなされることがある。場合によっては、クラスターを特定することは、オーバーラップするROIを特定することに加えて、信頼スコアのしきい値を超える信頼スコアに関連付けられたROIを特定することを含むことがある。今述べたことは、ROIをしきい値化すること（thresholding）と呼ばれることがある。

【0054】

場合によっては、MLモデルをトレーニングするために、MLモデルは、ROIと参照ROI500との間のアライメント度を決定することがある。場合によっては、アライメント度を決定することは、グラントゥールスの領域500に対するROIのオーバーラップ量および／またはフィット量を反映するメトリックを決定することを含むことがある。場合によっては、今述べたことは、和集合上の積集合を決定することを含むことがあるが、MLモデルによって生成されるROIのオーバーラップ量／誤差を決定するための他の技法が使用されることがある。

【0055】

今述べた例にて、MLモデルは、例示的なROI400'がグラントゥールスの領域500に上手くフィットする、たとえば、例示的なROI400'に関して98%のアライメント／フィット度（degree of alignment/fit）を示すアライメント度502によって意味されることがある。例示的な信頼スコア402は、高いアライメント度の観点から、適切に高い。場合によっては、「良い」フィットは、0.70以上のアライメント度であるように定義されることがある。今述べた値（0.70）は、しきい値のアライメント度と考えられることがあるが、他の値が、例えば、0.80、0.50、および／または0.60など、使用されることがあると想定される。ここで、エグザンプル400'は、NMS再割り当ての支配下にあるだろう。換言すれば、たとえ400が中央の30%の部分では

10

20

30

40

50

なくても、それは、「良い」ROIを生成し、例えば、（ハードエグザンプルとして選択されるならば）ペナルティーが与えられるべきではない。

【0056】

同様に、MLモデルは、ROI 404'が0.90のアライメント度504に関連付けられると決定することによって、例示的なROI 404'が、例示的なROI 400'よりも何か悪いとはいえ、参照ROI 500にフィットすると決定することがある。それにもかかわらず、エグザンプル404は、（ハードエグザンプルとして選択されるならば）同様にNMS再割り当ての支配下にあるだろう。

【0057】

さらに、MLモデルは、例示的なROI 408'が、しきい値のアライメント度（たとえば、0.70）を満たさない、または超えないアライメント度506（すなわち、この例では、0.62）に関連付けられると決定することによって、ROI 408'が参照ROI 500に下手にフィットすると決定することもある。上記のように、たとえ例示的な部分的に部分408が参照500内に入っても、有効なハードエグザンプルを構成する。

【0058】

同様に、さらに、MLモデルは、例示的なROI 412'が、しきい値のアライメント度（たとえば、0.70）を満たさない、または超えないアライメント度508（すなわち、この例では、0.47）に関連付けられると決定することによって、ROI 412'が参照ROI 500に下手にフィットすると決定することもある。信頼スコアが低い（すなわち、今述べた例では、0.13）であるので、MLモデルは、例408／408'を有効なハードエグザンプルとして決定することがある。

【0059】

図6Aおよび6Bは、NMS再割り当てを介して選択されるROIのサブセットに少なくとも部分的に基づいてMLモデルをトレーニングするための例示的な処理600の挿絵入りのフロー図を例示する。動作602において、例示的な処理600は、ここで述べられる技法のうちのどれかに従って、画像を受信することを含むことがある。

【0060】

動作604において、例示的な処理600は、本明細書にて述べられる技法のうちのどれかに従って、複数のROI（たとえば、例612にて例示されるようなROI 400'、404'、408'、412'）および関連付けられた複数の信頼スコアを決定することを含むことがある。例えば、MLモデルは、出力セルを含む（または、より詳細には、複数の特徴を含む）画像に対して出力グリッドを生成することがあり（または、より詳細には、特徴マップを出力することがあり）、出力セル（たとえば、各出力セル）に対して分類ごとにROI（および関連付けられた信頼スコア）を決定することがある。非限定の例として、例えば、出力グリッドの各セルは、車両分類、自転車分類、歩行者分類などの各々に対してROIおよび信頼を有することがあるが、分類のどんな組み合わせでも想定される。今述べたことは、例示的な分類「自動車」に対して決定される、図6Aに描かれるような複数のROI（400'、404'、408'、412'）を生成することを含むことがある。

【0061】

動作606において、例示的な処理600は、本明細書にて述べられる技法のうちのどれかに従って、参照ROI（たとえば、エグザンプル614にて例示されるような参照ROI 500）を受信することを含むことがある。参照ROI 500（たとえば、リファレンスリジョン）は、画像においてオブジェクトを表す画像の領域に対して、人によるラベリング、またはグラントゥールズを確立するどんな他の適切な方法でも介して、受信されるグラントゥールズであり得る。場合によっては、参照ROIは、ROIが生成された分類に関連付けられた画像の領域を示すことがある。例えば、ROI 500は、上に述べたような分類「自動車」を表す画像の領域を示すことがある。動作608において、例示的な処理600は、本明細書にて述べられる技法のうちのどれかに従って、MLモデルをトレーニングするためのエグザンプルのサブセットを選択することを含むことがある。

。今述べたことは、画像の全部分のうちの1つまたは複数の中から画像の一部（たとえば、出力セル）を決定することを含んで、MLモデルをトレーニングするためのエグザンプルのサブセットに含むことがある。例えば、例示的な処理600は、NMS再割り当てを使用して、サブセットに含むポジティブエグザンプルを決定すること（608（a））、サブセットに含むネガティブエグザンプルを決定すること（608（b））、および／またはサブセットに含むハードエグザンプルを決定すること（608（c））を含むことがある。今述べたことは、以下に詳細に述べられる。場合によっては、追加としてまたは代替えとして、動作608は、MLモデルをトレーニングするためのエグザンプルのサブセットにおける包含のためにエグザンプルをランダムに選択することを含むことがある。

【0062】

動作608（a）は、ROIに対する信頼スコアが、最大の信頼スコアである、および／または信頼スコアのしきい値を満たすまたは超えることと、参照ROIに対するROIのアライメント度が、しきい値のアライメント度を超えることとを決定することを含むことがある（たとえば、ROIが、画像において表されるオブジェクトを正しく特定することを示すことがある）。動作608（a）は、ポジティブエグザンプルとして今述べた決定に基づいて、出力セル、および／またはMLモデルをトレーニングするためのサブセットにおける包含のために関連付けられたMLモデルによって生成されるデータのうちのどれかを選択することを含むことがある。

【0063】

動作608（b）は、ROIに対する信頼スコアが、最小の信頼スコアである、および／または信頼スコアのしきい値を満たさないことと、参照ROIに対するROIのアライメント度が、しきい値のアライメント度を満たさないこととを決定することを含むことがある。動作608（a）は、ネガティブエグザンプルとして今述べた決定に基づいて、出力セル、および／またはMLモデルをトレーニングするためのサブセットにおける包含のために関連付けられたMLモデルによって生成されるデータのうちのどれかを選択することを含むことがある。

【0064】

図6Bに移ると、動作608（c）は、一般に、ハードエグザンプルを選択すること（608（c）（1））を含むことがある。ハードエグザンプルは、機械学習モデルが最も間違っているエグザンプルと呼ばれることがある。たとえば、上記のハードエグザンプルは、対応しているROIの非常に高い信頼を有するセルに対応するが、上記のROIを誤って特定することがある、そうでなければ対応する出力に基づいてペナルティーが与えられるべきである。一般に、ハードエグザンプルは、ネガティブエグザンプルであり得る。ハードエグザンプルを選択することは、誤ったROIを生成した（そうでなければペナルティーが与えられるべきである）が、高い信頼スコアに関連付けられる画像の今述べた部分（たとえば、1つまたは複数のセル）を特定することと、信頼スコアによって部分をソートすることと、ハードエグザンプルとして、ソートされた残りのROIから、上位n個の関連付けられた部分を選択することとを含むことがある。しかしながら、どんな適切なハードエグザンプル選択方法でも想定されることが理解される。

【0065】

ハードエグザンプルが選択されたらすぐに、動作608（c）は、本明細書にて述べられる技法のうちのどれかに従って、選択されたハードエグザンプルを抑制するかどうかを決定すること（608（c）（2））を含むことがある。場合によっては、たとえば、オブジェクトに関連付けられたセルの部分のみが、関連付けられたデータ（たとえば、およびROI）を出力するようにトレーニングされる場合、実際に、上記のハードエグザンプルは（ナイーブに選択されるとき）、トレーニングのための良いエグザンプルを表すことがある（たとえば、セルは、オブジェクトに関連付けられた、正しく特定されたROIを有するが、たとえば、オブジェクトの中央の30％に属さないとしてペナルティーが与えられる）。上記の場合では、動作608（c）（2）は、もし選択されたハードエグザンプルが、ROIのリージョンの境界内にある画像の一部に対応する（そうでなければ、

10

20

30

40

50

有効な出力、たとえば、画像のその部分に関連付けられたオブジェクトに対応するROI、に関連付けられた) ならば、選択されたハードエグザンプルを抑制すると決定することがある。選択されたハードエグザンプルを抑制すると決定することに基づいて、動作608(c)(2)は、選択されたハードエグザンプルを捨てて、新しいハードエグザンプル(たとえば、上に述べたソートすることに従う最も間違っている次のエグザンプル)を選択するために動作608(c)(1)に戻ることを含むことがある。例えば、選択されたハードエグザンプルを捨てることは、選択されたハードエグザンプルを、MLモデルをトレーニングするためのサブセットから除外し、後続の選択において選択されることが不可能であるように、選択されたハードエグザンプルを、ソートされたROIから除外することを含むことがある。場合によっては、追加としてまたは代替えとして、選択されたハードエグザンプルを抑制するという決定は、参照ROIに対してハードエグザンプルに関連付けられたROIのアライメント度が、アライメント度のしきい値を満たすまたは超えると決定することを含むことがある。場合によっては、上記の決定は、ROIが、NMSの下に抑制されることがあるだろうかと決定することによってなされることがある。すなわち、NMSの動作は、高い信頼を有し、参照ROIに対応するROIに関連付けられるエグザンプルを決定するために、およびそうでなければハードエグザンプルであるだろうものからそれらを除外するために使用されることがある。

【0066】

場合によっては、動作608(c)(2)は、参照ROI内の画像の部分に対する、および「良い」アライメント度を有するROIを生成する、ハードエグザンプルの選択を抑制することがある。例えば、ROI400'および404'は、ハードエグザンプルとして選択されるならば、それらが参照ROI500内の画像の部分から生成され、それらが、参照ROI500に対するROIの「良い」フィットを示すしきい値のアライメント度を超えるアライメント度に関連付けられた各々であるので、動作608(c)(2)において抑制されることがあるエグザンプルである。

【0067】

少なくとも場合によっては、上記のエグザンプルを抑制すること(たとえば、捨てること)の反対として、上記のエグザンプルは、ポジティブエグザンプルとして再割り当てされることがある。場合によっては、ハードエグザンプルを抑制すると決定することは、ハードエグザンプルに関連付けられた目標の信頼スコアを変えることを含むことがある。例えば、参照ROIの中央の30%内における画像の部分に対して生成されたROIは、グラントゥールスの代表であるという決定に基づいて、最大の目標の信頼スコア(たとえば、1.0)に関連付けられることがある。しかしながら、ハードエグザンプルが抑制するとき、そのハードエグザンプルに関連付けられる、およびハードエグザンプルを使用してトレーニングするために使用される目標の信頼スコアが、修正されることがある。例えば、抑制されると決定される、ハードエグザンプルが選択されたならば、ハードエグザンプルに対する目標の信頼スコアは、1.0に変えられることがある。次に、トレーニングは、抑制されたハードエグザンプル、および新しい目標の信頼スコア、1.0を使用して、遂行されることがある。換言すれば、選択されたハード・ネガティブ・エグザンプルは、代わりに、ポジティブエグザンプルとして再割り当てされ、それによって、損失がバックプロパゲーションされることがある。選択されたハードエグザンプルが抑制されない(すなわち、エグザンプルが真のハードネガティブである)ならば、信頼損失は、エグザンプルが、どんなボックスでも出力するようにトレーニングされるが、対応する低い信頼値を有するような、0に設定されることがある。

【0068】

30%の中央のリージョンをグラントゥールスとして実装する少なくともいくつかの例にて、今までに述べた技法は、ただランキングだけに基づいて、ナイーブに選択される参照ROIの残りの70%(たとえば、中央の30%の部分ではないオブジェクトの部分)から生成されるエグザンプルを単に捨てた技法によるMLモデルの正確度を改善する。さらに、今までに述べた技法は、中央の30%の外側で参照ROIの内側にある画像の部

10

20

30

40

50

分が、適切なアライメント度のROIを生成することがあるので、中央の30%の外側に生成されたエグザンプルにペナルティーを与えることによってMLモデルをトレーニングする技法によるMLモデルの正確度も改善する、換言すれば、適切なROIを生成するために上記の部分にペナルティーを与えることは、実際には存在しない区分を学習するMLモデルを教えることがある（たとえば、適切なROIが生成された部分が、画像におけるオブジェクトの表現に実際には相関し、これにペナルティーを与えることは、実際には存在しない区分を作るMLモデルを教える）。

【0069】

動作610において、例示的な処理600は、本明細書にて述べられる技法のうちのどれかに従って、選択されたエグザンプルのサブセットを使用してMLモデルをトレーニングすることを含むことがある。

10

【0070】

少なくとも場合によっては、MLモデルをトレーニングすることは、第1のステージが、上記のNMS再割り当てを省略することがあり、第1のステージに続く第2のステージが、NMS再割り当てを含むことがある、少なくとも2つのステージを含むことがある。上記の例にて、第2のステージの間、画像の種々の部分（セル）のROIは、抑制されていないROIが、同一のまたは後続のステージにおいてMLモデルをトレーニングするために参照ROI500として使用されることがあるように、NMSの下に抑制されることがある。

【0071】

20

場合によっては、第1のステージは、第1のバッチの画像を、トレーニング用のMLモデルに提供することを含むことがある、第2のステージは、第2のバッチの画像を、トレーニング用のMLモデルに提供することを含むことがある。すなわち、異なるバッチのデータは、後続としてモデルをトレーニングするのに使用されることがある。

【0072】

場合によっては、第2のステージにおいてMLモデルをトレーニングすることは、Focal Loss関数を組み入れることがある。場合によっては、Focal Loss関数は、ファクター、 $(1 - p_t)^{\gamma}$ 、ただし、 $\gamma > 0$ （たとえば、 $\gamma = 1$ 、 $\gamma = 2$ 、 $\gamma = 5$ ）を加える交差エントロピー損失関数を含んで、ハード、誤分類されたエグザンプルに、上手く分類されたエグザンプル（たとえば、 $p_t > 0.5$ または $p_t > 0.7$ のエグザンプル）よりも、より重くペナルティーを与えることがある。例えば、Focal Loss関数は、次の式を使用して、ROIに対して評価されることがある。

30

【0073】

【数1】

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t)$$

【0074】

ただし、

【0075】

【数2】

40

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases}$$

【0076】

であり、

【0077】

【数3】

$$y \in \{\pm 1\}$$

【0078】

50

がグラントゥールス分類を示し、

【0079】

【数4】

$p \in [0,1]$

【0080】

が、ラベル $y = 1$ による分類に対するMLモデルの信頼スコアである。機能性として、今述べたことは、低い誤差に関連付けられるROIに対する交差エントロピー損失関数によって計算される誤差の重みを減らし（たとえば、 $p_t > 0.5$ に対して、信頼スコアが最適なスコアに近く、ROIがグラントゥールスに上手くフィットする）、より大きな誤差に徐々にペナルティーを与える。単に、MLモデルが、それが生成した実際に間違ったROIについて信頼すればするほど、ますます交差エントロピー損失は、Focal Loss関数によって変わらないままである。一方、Focal Loss関数は、上手く分類されたエグザンプルに対して計算された誤差の重みを減らし、たとえ多くの上手く分類されたエグザンプルがあっても、各々が生成する極小の誤差の効果は、効果的なトレーニングに関して、より少なくなるだろう。場合によっては、動作608の損失関数は、交差エントロピー損失関数、Huber関数、平均二乗誤差の計算などを含むことがある。

【0081】

場合によっては、3つのステージが想定され、第1のステージが、どんなFocal LossおよびNMS再割り当てでも省略するが、第2および第3のステージは、Focal Lossおよび／またはNMS再割り当てを組み入れる。少なくとも場合によっては、画像は、1つまたは複数のステージにおいて、別個にスケールリングされることがある。非限定の例として、クロップされたサイズ 240×240 の画像は、第1のステージにてトレーニングするときに第1のバッチにおいて使用されるが、サイズ 960×600 を有する画像クロップ (image crop) は、第3のステージにてモデルをトレーニングするのに使用される第3のバッチにおいて使用されることがある。一般に、画像クロップのサイズ、繰り返し数、学習率、バッチサイズ、および／またはバランシングは、ステージからステージへと変わることがある。

【0082】

場合によっては、第1のバッチの画像、および／または第2のバッチの画像は、MLモデルが、異なるタイプのオブジェクトを正確に検出するようにトレーニングされていることを確実にするために、データバランシングされることがある。1つのオブジェクト分類に対して連続的にさらされるMLモデルは、異なるタイプのオブジェクトに対してあまり頻繁にさらされないようにゆっくりと「忘れる」（たとえば、検出においてあまり正確にならない）ことがある。その上さらに、ある特定の分類のデータセットにエグザンプルがほとんど存在しない場合（たとえば、自動車が、自転車に関してデータセットを大量に支配するならば）、MLモデルは、画像に自転車がないと単に述べることによって、損失を最適化することがある。それゆえ、場合によっては、第1のバッチの画像、および／または第2のバッチの画像は、第1のオブジェクト分類に関連付けられる第1の予め定義された数の画像（たとえば、第1の予め定義された数の画像の各々が、少なくとも1人の歩行者を含む）、および第2のオブジェクト分類に関連付けられる第2の予め定義された数の画像（たとえば、第2の予め定義された数の画像の各々が、少なくとも1人の自転車に乗る人を含む）を含むことがあるが、分類および相対的な重みのいくつでも想定される。例えば、入力バッチは、MLモデルが検出するようにトレーニングされる各タイプのオブジェクトのうちの4つで、たとえば、自転車4台、自動車4台、歩行者4人、夜間画像4枚、日間画像4枚などで含むことがある32枚の画像を含むことがある。場合によっては、MLモデルをトレーニングすることは、画像を受信することと、画像に関連付けられたオブジェクト分類を決定することと、画像を、オブジェクト分類に関連付けられるトレーニングキュー (training queue) と関連付けることとを含むことがある。トレーニングの時間において、バッチは、各キューが、別個のオブジェクト分類および／または画像状況（

10

20

30

40

50

たとえば、夜／昼、障害物）に関連付けられる複数のキューの出力から構成されることがある。MLモデルが、いろいろなオブジェクト分類において連続的にトレーニングされることを確実にするによって、MLモデルは、オブジェクト分類を「忘れ」そうにない。

【0083】

例示的な応答曲線および制限された多重スケール推論

図7Aおよび7Bは、スケールに基づいてROIの出力サイズを抑制する例を例示する。一般に、ニューラルネットワークは、受容野と関連付けられることがある。上記の受容野は、入力信号のうちのどれぐらいのデータが、推論をするときに使用されるかを示すことがある。例えば、もしモデルの受容野が十分に大きくないならば、より大きいサイズのオブジェクトは、モデルを使用して適切に決定されないことがある。上のように、より大きなモデルは、すべてのオブジェクトを検出するのに用いられることがあるが、上記のモデルは、かなりのリソースおよび処理を必要とするだろうそしてそれは、アプリケーションを限定する（たとえば、上記のモデルは、処理時間およびメモリー要件のために自律車両のアプリケーションにおいて使用されることができないだろう）。今述べた限定を解消するために、1つまたは複数のより小さいネットワーク（すなわち、より小さいメモリーフットプリントおよび／または処理要件を有するネットワーク）が使用されることがある。上記の各ネットワークは、種々のROIサイズに対する特定の正確度と関連付けられ、上記の範囲内にROIを出力することのみに制限されることがある。サイズとの関連において述べたが、どんなデータ階層化（または他のメトリック）でも、本明細書にて想定される。非限定の例として、さらに、図7～8のうちのどれにおいても、例示される上記の技法は、どんな1つまたは複数の関心のリージョンのサイズ、局所コントラストなどに対してでも最適なネットワークを選択することがある。

【0084】

図7Aに例示されるように、第1のグラフ700は、第1のスケールに関する画像（複数可）内の種々のサイズを有するオブジェクトに対する第1のMLモデルの例示的な応答曲線702を含む。図7Bは、第2のスケールに関する画像（複数可）内の種々のサイズを有するオブジェクトに対する第2のMLモデルについての第2の例示的な応答曲線706を含む第2のグラフ704を例示する。第1のグラフ700および第2のグラフ704のy軸は、画像におけるオブジェクトを特定するROIおよび／または信頼スコアに関してモデルの正確度を示すことがあり、第1のグラフ700および第2のグラフ704のx軸は、画像におけるオブジェクトのサイズ、ROIのサイズ、および／または画像のサイズを示すことがある。一例にて、x軸は、ROIのサイズを示すことがある。

【0085】

場合によっては、技法は、画像における見かけのオブジェクトサイズを介してオブジェクトを分類する（たとえば、ROIおよび／または信頼スコアを作り出す）ために、第1のMLモデルの正確度を示す第1のMLモデルに関連付けられた応答曲線を決定することを含むことがある。今述べた応答曲線を生成するために、技法は、第1のMLモデルへの入力として、（1つまたは複数の画像に表される）複数のオブジェクトを提供することと、第1のMLモデルからの出力として、複数のオブジェクトに関連付けられた複数のROIおよび／または信頼スコアを受信することと、ROIおよび／または信頼スコアの各々に対して（たとえば、グラントゥールスと比較することによって、それぞれのグラントゥールスの領域に対する各ROIに対してアライメント度を計算することによって）、正確度を決定することを含むことがある。応答曲線を生成するために、技法は、ROIのサイズ（たとえば、高さおよび／または幅、領域）をROIの正確度（たとえば、ROIに関連付けられたアライメント度、アライメント度と信頼スコアとの間の不釣り合い）と関連付けることを含むことがある。これらの正確度／ROIサイズのペアは、画像のバッチから第1のMLモデルに対して決定されたROIに対して、1つにされることがある。

【0086】

場合によっては、数千の画像のバッチが、または数十万の画像のバッチでさえ、今述べた応答曲線が決定され得るように、入力として、第1のMLモデルに提供されることがあ

る。第1のMLモデルがバッチを処理したらすぐに、技法は、しきい値の正確度を満たすまたは超える正確度に関連付けられたROIサイズの範囲を特定することと、サイズの範囲を第1のMLモデルと関連付けることとを含むことがある。今述べたサイズの範囲は、第1のMLモデルが、サイズの範囲のオブジェクトに対して「良い」ROIを決定することを示すことがある。場合によっては、第1のMLモデルは、サイズの範囲内のサイズを有するROIを出力することがあり、第1のMLモデルによって決定される他のROIを抑制することがある。場合によっては、追加としてまたは代替えとして、ROIのサイズを介して正確度をプロットすることに対して、ROIの正確度は、グラントゥールースのサイズ、画像の一部分のサイズ、および／または画像のサイズを介してプロットされることがある。

10

【0087】

例えば、上に述べた技法に従って同様に生成された、しきい値の正確度は708によって示されることがあり、第1のMLモデルは応答曲線702と関連付けられることがあり、および第2のMLモデルは応答曲線706と関連付けられることがある。少なくとも場合によっては、第1および第2のMLモデルは、入力スケールにおいて異なるだけの、同一のモデルであり得る。描かれた例に従って、第1のMLモデルに関連付けられた第1の範囲は、サイズ900×900の入力画像に基づいて、100および200ピクセル間のピクセルの高さを有するROIのみを出力することがあるが、第2のMLモデルに関連付けられた第2の範囲は、サイズ300×300の入力画像に基づいて、50および100ピクセル間の高さを有するROIのみを出力することがある。場合によっては、第1の応答曲線702、および第2の応答曲線は、異なるしきい値の正確度レベルと関連付けられることが可能である。

20

【0088】

もちろん、モデルおよび／またはスケールのいくつでも、入力および出力のサイズに関する種々の範囲に対して決定されることがある。種々のスケールの（元の画像サイズに対して、より大きい、等しい、またはより小さい）画像は、画像におけるオブジェクトを検出する、分類する、および／または区分するために、種々の出力モデルに（連続してか実質的に同時にいずれかで）入力されることがある。

【0089】

少なくとも一例にて、選択されるモデル、範囲、または階層化の指定は、第3のモデルに少なくとも部分的に基づいて決定されることがある。非限定の例として、モデルは、望ましい出力を決定するために使用する、何というスケールおよび／または何というモデル間で区別するようにトレーニングされることがある。

30

【0090】

図8Aは、本明細書にて述べられる技法に従って、入力として第1のMLモデル802に提供される第1のスケールされた画像800と、第1のスケールされた画像を受信することに対応する第1のMLモデルによって生成される第1のスケールされた画像および2つのROI（804、806）とを例示する。場合によっては、第1のMLモデル802は、第1のサイズの範囲に関連付けられることがあり、車両108に対してROIを生成していることがあるが、ROIは、第1のサイズの範囲外に入っていて、第1のMLモデルによって抑制されている。一方、車両106に対応するROI804、および車両110に対応するROI806は、第1のサイズの範囲内に入ることがあり、それゆえ、第1のMLモデルによって出力されることがある。場合によっては、第1のスケールされた画像は、1のファクターによってスケールされることがある（たとえば、第1のスケールされた画像がアンスケールされる）。

40

【0091】

図8Bは、本明細書にて述べられる技法に従って、入力として第1のMLモデルに提供される例示的な第2のスケールされた画像808と、第2のスケールされた画像を受信することに対応する第1のMLモデルによって生成されるROI810とを例示する。場合によっては、第2のスケールされた画像は、1より小さいスケール係数によ

50

ってスケーリングされ、例示的な画像 800 のよりも小さいスケールの画像に帰着する。描かれた例にて、画像がスケールダウンされているので、今、車両 108 に対応する ROI は、第 1 の ML モデルと関連付けられた第 1 のサイズの範囲内に入り、車両 106 および 110 と関連付けられた ROI は、第 1 のサイズの範囲外に入っており、それゆえ、図 8B において、第 1 の ML モデル 802 によって抑制されている。場合によっては、第 2 のスケールファクターは（および／またはどんな追加のスケールファクターでも）、同一の画像が 2 つの異なるスケールにおいて第 1 の ML モデルに提供されるときに、ROI が同一のオブジェクトに対して冗長に生成されないように選択されることがある。

【0092】

場合によっては、第 1 のサイズの範囲は、第 1 の ML モデルの受容野に少なくとも部分的に基づくことがある。例えば、例示的な画像 800 における車両 108 の表現は、あまりにも大き過ぎて受容野にフィットしないことがあるが、例示的な画像 808 における車両 108 の表現は、スケールダウンされるとすぐに、第 1 の ML モデルの受容野にフィットすることがある。

【0093】

場合によっては、第 1 のサイズの範囲、および／または第 2 のサイズの範囲は、種々のサイズ（または、以下に述べられる種々のメトリックおよび／またはデータ階層化）を正確に区別するのに使用される種々の ML モデルの精度、再呼び出し、および／または正確度を最適化する回帰モデル（または他の機械学習されたモデル）に部分的に基づくことが可能である。

【0094】

図 8C は、画像においてすべてのオブジェクトを検出するための追加のまたは代替の例を例示する。反復的に（図 8B に描かれるような）、画像を ML モデルに提供することと、ROI を決定し、第 1 のサイズの範囲内のサイズ（複数可）を有する ROI（複数可）を出力することと、画像をスケーリングすることと、スケーリングされた画像を ML モデルに提供することなどの代わりに、例示的な画像 800 は、第 2 の ML モデル 812 が、第 1 の ML モデルによって検出されないオブジェクト（たとえば、今述べた例における車両 108）に対して、ROI 814 を決定することがあるように、第 2 のサイズの範囲と関連付けられた第 2 の ML モデル 812 に提供されることがある。場合によっては、画像は、各ディテクターが、別個のサイズの範囲と関連付けられ、サイズの範囲内の ROI を出力する場合、複数のディテクターに提供されることがある、および／または今述べた技法は、第 1 のスケーリングされた画像を第 1 のディテクターに提供することと、第 1 の出力を第 1 のディテクターから受信することと、第 1 のスケーリングされた画像を第 2 のスケーリングされた画像としてスケーリングすることと、第 2 のスケーリングされた画像を第 1 のディテクターに提供することと、第 2 の出力を第 2 のディテクターから受信することと、によって組み合わせられることがある。

【0095】

図 8A、8B、および 8C が、入力画像のサイズを変えることとの関連において、および／または予期される出力サイズに基づいて ML モデルを選択することとの関連において述べられるが、どんなメトリックおよび／またはデータ階層化が、適切な ML モデルを選択する、および／またはスケールを入力するのに使用されることが可能である。例えば、種々のメトリックは、限定されないが、コントラスト（たとえば、画像のリージョンおよび／またはピクセルの局所コントラスト）、輝度、チャネル数、色（たとえば、RGB、白黒、グレースケールなど）、サイズ（たとえば、幅および高さ）、面積、グリッドのサイズ、画像において特定されるオブジェクト数、時刻（たとえば、昼／夜）、天気（たとえば、雨、晴れ、雪、霧など）、分類などのうちの 1 つまたは複数を含むことが可能である。

【0096】

今述べたやり方にて、同一の ML モデル、および／または異なる ML モデルは、例えば、自律車両を制御するように、高度の正確度において、あまりにも計算時間がかかり過ぎ

10

20

30

40

50

てリアルタイムのアプリケーションに対して利用されないだろう大きなニューラルネットワークをトレーニングすることなく、異なるサイズのオブジェクトに対してROIを生成することがある。

【0097】

例示的なインスタンスセグメンテーション

図9Aは、例示的な画像904の一部分902に対してMLモデルによって決定された例示的なROI900を例示する。上に（たとえば、図5に関して）例示したエグザンプルを使ってのように、MLモデルはROIを出力するようにトレーニングされることがある。しかしながら、上のエグザンプルにおいてトレーニングされたモデルと違って、図9Aに例示されるMLモデルは、中央の30%と関連付けられたROIの代わりに、オブジェクトのマスクにおけるピクセルに関連付けられたそれらのROIを使用して、トレーニングされることがある。具体的には、画像におけるオブジェクトに対する画像のマスクおよび対応するROIは、MLモデルをトレーニングするときにグラントゥールスとして使用されることがある。例えば、上記の関心のリージョンをトレーニングすることは、マスクと関連付けられるROI（たとえば、マスクのピクセルに少なくとも部分的に基づくバウンディングボックス）に基づいて決定されることがある。図9Aに例示されるように、画像904の（ピクセルであり得る）一部分902は、オブジェクトのマスクの部分を形成し、関連するROI900を有することがある。

【0098】

図9Bは、例示的な画像904における同一のオブジェクト（すなわち、今述べた例における車両110）と関連付けられた例示的な複数のROI906を例示する。場合によっては、MLモデルは、例示的な画像904を受信することがあり、画像904の一部分および／またはピクセルごとに関連付けられた例示的な複数のROI906および／または信頼スコアを含む複数のROIを生成することがある。

【0099】

場合によっては、上のように、MLモデルは、複数の決定されたROIの外のオブジェクト（たとえば、オブジェクト110）に関連付けられる単一のROIを決定することがある。少なくとも一例にて、上記の検出は、NMSを使用して達成されることがある。本明細書にて述べられる例と同様に、すべてのピクセルまたは部分は、ROIを出力することがあり、複数のROIは、NMSによって抑制されることがある（たとえば、最大の信頼レベルに関連付けられないROIが抑制されることがある）。実質的に同時に、MLモデルは、車両110と関連付けられるROIと、出力用のROIの決定に少なくとも部分的に基づいて、オブジェクトに関連付けられる画像の特定の部分（たとえば、オブジェクトを表すピクセル）を特定するマスクとを決定することがある。

【0100】

少なくとも場合によっては、上記のMLモデルは、NMSの間に抑制されたピクセルおよび／またはリージョンのインディケーションを持続させることがある。画像のすべての上記ピクセルおよび／または部分は、オブジェクトのマスク（インスタンスセグメンテーション）と関連付けられることがある。場合によっては、さらに、NMSによって抑制されなかった極大のROIを生成した画像の部分は、マスクと関連付けられることもある。

【0101】

図9Cは、本明細書にて述べられる技法に従って、オブジェクトを特定する例示的なインスタンスセグメンテーション（すなわち、描かれた例におけるマスク）の2つの表現（910、912）を例示する。表現910は、例示的なマスクがオブジェクト（すなわち、車両110）を表すとして特定する画像の暗い部分を含む。明確のために、表現912は、オブジェクトをポジティブに特定するとしてマスクに含まれる画像の部分が車両110に対応することを示す図9Cに含まれる。場合によっては、表現910の画像の暗い部分の各々は、MLモデルが、NMSの間に抑制されたROIを生成した、またはそうでなければ上記のオブジェクトのROI（たとえば、最大の信頼スコアに関連付けられたROI）と関連付けられた、画像の部分に対応することがある。場合によっては、自律車両は

10

20

30

40

50

、自律車両の動きを制御するための軌道を生成するROIおよび／またはマスクを使用することがある。

【0102】

今述べた技法は、インスタンスセグメントを行うために別々の動作を必要とするかつての技法を越えた改良である。代わりに、本明細書にて述べられる技法は、従来の技法を使用してオブジェクト検出を行うのにかかるのと同じ計算時間において、オブジェクト検出と画像セグメンテーションとの両方を行う。場合によっては、本明細書にて説明されるMLモデルは、20ミリ秒で実行される。

【0103】

例示的なアーキテクチャー

図10は、本明細書において述べられる技法のどれかに従って、少なくとも1つの車両、例えば、自律車両などの動作を制御するための例示的な車両システム1002を含む例示的なアーキテクチャー1000のブロック図である。場合によっては、車両システム1002は、自律車両104の少なくとも一部を表すことがある。場合によっては、今述べたアーキテクチャーは、本明細書にて述べられるROI、画像セグメント、および／または信頼スコアに少なくとも部分的に基づいて、自律車両を制御するのに使用されることがある。

【0104】

場合によっては、車両システム1002は、（まとめるとコンピューティングデバイス（複数可）118を表すことがある）プロセッサ（複数可）1004および／またはメモリー1006を含むことがある。今までに述べたエレメントは、図10において組み合わせにより例示されるが、これらが、車両システム1002の別々のエレメントであり得ることと、場合によっては、システムのコンポーネントが、ハードウェアおよび／またはソフトウェアとして実装され得ることとが理解される。

【0105】

プロセッサ（複数可）1004は、1つのプロセッサを含むユニプロセッサシステム、またはいくつか（たとえば、2つ、4つ、8つ、または別の適切な個数）のプロセッサを含むマルチプロセッサシステムを含み得る。プロセッサ（複数可）1004は、命令を実行する能力があるどんな適切なプロセッサでもあり得る。例えば、種々の実装において、プロセッサ（複数可）は、いろいろなISA（命令セットアーキテクチャー）のどれでも、例えば、x86、PowerPC、SPARC、もしくはMIPSのISA、または他の適切などんなISAでも実装する汎用または組み込みプロセッサであり得る。マルチプロセッサシステムにおいて、各プロセッサ1004は、一般的に、必ずしもではないが、同一のISAを実装し得る。場合によっては、プロセッサ（複数可）1004は、CPU（中央処理装置）、GPU（グラフィックスプロセッサ）、FPGA（Field Programmable Gate Array）、ASIC（Application Specific Integrated Circuit）、またはその組み合わせを含み得る。

【0106】

例示的な車両システム1002は、メモリー1006を含み得る。場合によっては、メモリー1006は、実行可能な命令／モジュール、データ、および／またはプロセッサ（複数可）1004によりアクセス可能なデータアイテムを格納するように構成された非一時的なコンピューター読み取り可能な媒体を含み得る。種々の実装において、非一時的なコンピューター読み取り可能な媒体は、適切などんなメモリー技術でも、例えば、SRAM（スタティックRAM）、SDRAM（シンクロナスDRAM）、不揮発性／フラッシュ型メモリー、またはメモリーの他のどんなタイプなどでも使用して、実装され得る。例示される例において、例えば、上に説明されたような、望ましい動作を実装するプログラム命令およびデータは、非一時的なコンピューター読み取り可能なメモリー内に格納されて示される。他の実装において、プログラム命令、および／またはデータは、受信され、送られ、または例えば、非一時的なコンピューター読み取り可能な媒体、もしくは非一時的なコンピューター読み取り可能な媒体から分離した同様な媒体など、様々なタイプの

10

20

30

40

50

コンピュータアクセス可能な媒体に格納され得る。一般的に言って、非一時的な、コンピュータ読み取り可能なメモリは、ストレージ媒体またはメモリ媒体、例えば、例示的な車両システム1002に「I/O」（入力／出力）インターフェイス1008を介して結合されたフラッシュメモリ（たとえば、ソリッドステートメモリ）、磁気または光媒体（たとえば、ディスク）などを含み得る。非一時的なコンピュータ読み取り可能な媒体を介して格納されたプログラム命令およびデータは、例えば、ネットワークインターフェイス1010を介して実装され得るような通信媒体、例えば、ネットワークおよび／または無線リンクなどを介して運ばれ得る伝送媒体または信号、例えば、電氣的な、電磁氣的な、またはデジタルの信号などにより送信されることがあり得る。

【0107】

さらにその上、図10においてシングルユニットとして例示されるが、プロセッサ（複数可）1004およびメモリ1006は、車両の複数のコンピューティングデバイスにおよび／または複数の車両、データセンター、遠隔操作センターなどに分散され得る。

【0108】

場合によっては、「I/O」（入力／出力）インターフェイス1008は、プロセッサ（複数可）1004、メモリ1006、ネットワークインターフェイス1010、センサー（複数可）1012、I/Oデバイス1014、ドライブシステム1016、および／または車両システム1002の他のどのハードウェアの間にも、I/Oトラフィックを調整するように構成され得る。場合によっては、I/Oデバイス1014は、外部および／または内部のスピーカー（複数可）、ディスプレイ（複数可）、乗客の入力デバイス（複数可）などを含み得る。場合によっては、I/Oインターフェイス1008は、プロトコル、タイミング、または他のデータの変換を行い、あるコンポーネント（たとえば、非一時的なコンピュータ読み取り可能な媒体）からのデータ信号を、別のコンポーネント（たとえば、プロセッサ（複数可））による使用に適したフォーマットへとコンバートすることがあり得る。場合によっては、I/Oインターフェイス1008は、例えば、P C I（Peripheral Component Interconnect）バス規格、U S B（ユニバーサルシリアルバス）規格、またはそのバリエーションのような種々のタイプのペリフェラルバスを通じて取り付けられたデバイスに対するサポートを含み得る。いくつかの実装において、I/Oインターフェイス1008の機能は、例えば、ノースブリッジおよびサウスブリッジなどのような、2つ以上の別々のコンポーネントに分裂されることがあり得る。さらに、場合によっては、I/Oインターフェイス1008の機能性の一部またはすべて、例えば、メモリ1006へのインターフェイスなどは、プロセッサ（複数可）1004、および／または車両システム1002の1つまたは複数の他のコンポーネントに直に組み入れられることがあり得る。

【0109】

例示的な車両システム1002は、車両システム1002と1つまたは複数の他のデバイスとの間に通信リンク（すなわち、「ネットワーク」）を確立するように構成されたネットワークインターフェイス1010を含み得る。例えば、ネットワークインターフェイス1010は、車両システム1002と別の車両1018との間において第1のネットワーク1020を介して、および／または車両システム1002とリモートコンピューティングシステム1022との間において第2のネットワーク1024を介して、データが交換されることを可能にするように構成され得る。例えば、ネットワークインターフェイス1010は、別の車両1018および／またはリモートコンピューティングデバイス1022間の無線通信を可能にすることがあり得る。種々の実装において、ネットワークインターフェイス1010は、例えば、W i - F i ネットワークのような無線の一般的なデータネットワーク、および／または例えば、セルラー通信ネットワーク、衛星ネットワークなどのようなテレコミュニケーションネットワークを介した通信をサポートすることがあり得る。

【0110】

例示的な車両システム1002は、例えば、車両システム1002を環境中にローカラ

10

20

30

40

50

イズするように、環境中の1つまたは複数の物体を検出するように、環境を通じて例示的な車両システム1002の動きを検知し、環境データ（たとえば、周囲の温度、気圧、および湿度）を検知し、および／または例示的な車両システム1002の内部の状態（たとえば、乗客総数、内部温度、ノイズレベル）を検知するように構成されたセンサー（複数可）1012を含み得る。例えば、センサー（複数可）1012は、1つまたは複数のLIDARセンサー、1つまたは複数のカメラ（たとえば、RGBカメラ、強度（グレースケール）カメラ、赤外線カメラ、深度カメラ、ステレオカメラ）、1つまたは複数の磁気センサー、1つまたは複数のレーダーセンサー、1つまたは複数のソナーセンサー、1つまたは複数の音を検知するためのマイクロホン、1つまたは複数の（たとえば、加速度計およびジャイロ스코ープを含む）IMUセンサー、1つまたは複数のGPSセンサー、1つまたは複数のガイガーカウンターセンサー、1つまたは複数のホイールエンコーダー、1つまたは複数のドライブシステムセンサー、スピードセンサー、および／または例示的な車両システム1002の動作に関係した他のセンサーを含み得る。場合によっては、本明細書にて述べられるセンサーデータは、第1の車両により受信され、第2の車両に送信され得る。場合によっては、本明細書にて述べられる画像は、センサー（複数可）1012のうちのセンサーにおいて受信され、認識エンジン1026に提供されることがある。

10

【0111】

例示的な車両システム1002は、認識エンジン1026（たとえば、認識エンジン116）MLモデル1028（たとえば、MLモデル114）、およびプランナー1030（たとえば、プランナー120）を含むことがある。

20

【0112】

認識エンジン1026は、プロセッサ（複数可）1004により実行されると、センサーデータをセンサー（複数可）1012から入力として受信し、例えば、例示的な車両システム1002を取り巻く環境中の物体に関する1つまたは複数の姿勢（たとえば、位置および向き）、物体に関連付けられた物体追跡（object track）（たとえば、時間周期（たとえば、5秒）にわたる物体に関する位置履歴、速度履歴、加速度履歴、および／もしくは進路履歴）、ならびに／または物体（たとえば、歩行者、車両、自転車運転者など）に関連付けられたオブジェクト分類の典型的なデータを出力するプロセッサ（複数可）1004を構成する、メモリー1006に格納された命令を含み得る。場合によっては、認識エンジン1026は、1つまたは複数の物体に関する、より多くの物体の軌道を予測するように構成され得る。例えば、認識エンジン1026は、例えば、オブジェクトに関連付けられた予測される位置、軌道、および／もしくは速度の確率的決定またはマルチモーダル分布に基づいて、複数のオブジェクトの軌道を予測するように構成され得る。

30

【0113】

認識エンジン1026は、プロセッサ（複数可）1004によって実行されると、入力、例えば、画像としてセンサー（複数可）1012からセンサーデータを受信し、センサーデータをMLモデル1028に提供し、ROI、画像セグメント、および／または信頼スコアをMLモデル1028から受信し、認識エンジンがオブジェクトをセンサーデータから検出し、1つまたは複数のROI、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータを出力し得ることのインディケーションを出力するようにプロセッサ（複数可）1004を構成する、メモリー1006に格納された命令を含むことがある。さらに、ROI（複数可）、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータは、メモリー1006に格納されることもある。

40

【0114】

MLモデル1028は、プロセッサ（複数可）1004によって実行されると、画像を受信し、画像に少なくとも部分的に基づいて、画像におけるオブジェクトの表現を特定するROI（複数可）、および／またはインスタンスセグメンテーション（複数可）を出力することを決定するようにプロセッサ（複数可）1004を構成する、メモリー1006に格納された命令を含むことがある。MLモデル1028は、例えば、ランダムフォ

50

レストおよび／または向上された決定木のアンサンブルなどの決定木、またはそのどんな配置でも、DAG（有向非巡回グラフ）（たとえば、ノードがベイジアンネットワークとして組織される場合）、例えば、ANN（人工ニューラルネットワーク）、DBN（deep belief network）、DSN（deep stacking network）、またはRNN（リカレントニューラルネットワーク）などの深層学習アルゴリズム（複数可）などを含み得る。場合によっては、MLモデル1028は、特徴値が、出力を決定するためにプッシュされ得るノードを有するRNNを含み得る。例えば、MLモデル1028は、ノードの入力層、ノードの1つまたは複数の隠れ層、およびノードの出力層を含むことがあり得る。場合によっては、ノードの入力層は、画像および／またはその一部を受信し、1つまたは複数の隠れ層のノードを活性化するように構成されることがあり得る。出力層は、1つまたは複数の隠れ層のノードから刺激を受信し、最も活性化された出力層のノードに基づいて、および／または出力ノードの活性化量に対応し得る出力ノードの出力に基づいて、ROI、画像セグメント、および／または信頼スコアを出力するように構成されることがあり得る。場合によっては、個々の出力ノードは、画像の一部と対応することがある。

【0115】

認識エンジン1026は、ROI、画像セグメント、信頼スコア、および／または関係する他のデータを、プランナー1030が軌道生成するのに使用し得る他の追加情報（たとえば、オブジェクト分類、物体追跡、車両姿勢）とともに、プランナー1030に送信することがあり得る。場合によっては、追加としてまたは代替として、認識エンジン1026および／またはプランナー1030は、少なくとも部分的に、認識エンジン1026によって決定された信頼スコア（複数可）に基づいて、ネットワークインターフェイス1010を介してROI（複数可）、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータを、ネットワーク1024を介してリモートコンピューティングデバイス1022に、および／またはネットワーク1020を介して別の車両1018に送信することがあり得る。

【0116】

場合によっては、認識エンジン1026および／またはMLモデル1028は、別の車両1018および／またはリモートコンピューティングデバイス1022に配置されることがあり得る。場合によっては、別の車両1018および／またはリモートコンピューティングデバイス1022に配置された認識エンジンは、認識エンジン1026と決定を調整し得る。MLモデル1028が他の車両1018および／またはリモートコンピューティングデバイス1022に配置される場合において、他の車両1018および／またはリモートコンピューティングデバイス1022は、それぞれ、ネットワーク1020および／または1024を介して、画像を車両システム1002から受信することがあり、ROI（複数可）、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータを決定することがある。次に、他の車両1018および／またはリモートコンピューティングデバイス1022は、ROI（複数可）、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータを、車両システム1002のプランナー1030に戻す送信をすることがある。

【0117】

プランナー1030は、プロセッサ（複数可）1004により実行されると、例えば、環境中の例示的な車両システム1002のロケーションを表すデータ、および他のデータ、例えば、ローカルポーズデータ（local pose data）などと、ROI（複数可）、インスタンスセグメンテーション（複数可）、信頼スコア（複数可）、分類（複数可）、および／または関係する他のデータとを使用して、例示的な車両システム1002の軌道の典型的なデータを生成するプロセッサ（複数可）1004を構成する、メモリ1006に格納された命令を含み得る。場合によっては、プランナー1030は、実質的には連続的に（たとえば、後退ホライズン（receding horizon）の時間が想定されるが、1または2ミリ秒ごとに）、例示的な車両システム1002を制御するための複数の可能性が

10

20

30

40

50

ある軌道を生成し、車両を制御するための軌道のうちの1つを選択することがあり得る。選択は、現在のルート、静止車両がブロッキング車両である確率、現在の車両の軌道、および／または検出された物体の軌道データに少なくとも部分的に基づくことがあり得る。軌道を選択するとすぐに、プランナー1030は、軌道をドライブシステム1016に送信して、選択された軌道に従って例示的な車両システム1002を制御することがあり得る。

【0118】

場合によっては、さらに、認識エンジン1026、MLモデル1028、および／またはプランナー1030は、例えば、認識エンジンを実行するのに適したプロセッサ（たとえば、グラフィックスプロセッサ、FPGA）のような特殊化されたハードウェアを含み得る。

【0119】

例示的な箇条

A．システムは、1つまたは複数のプロセッサと、1つまたは複数のプロセッサによって実行可能な命令を格納する1つまたは複数のコンピューター読み取り可能な媒体とを含み、命令が、実行されると、システムに、ML（機械学習）モデルに入力として画像を提供することと、MLモデルから出力として特徴マップを受信し、特徴マップの一部分は、バウンディングボックスおよび信頼スコアに関連付けられることと、エグザンプルのサブセットに少なくとも部分的に基づいてMLモデルをトレーニングし、トレーニングは、オブジェクトを表す画像の領域に関連付けられたリファレンスリージョンを特定すること、第1のバウンディングボックスに関連付けられた第1の信頼スコアに少なくとも部分的に基づいて、エグザンプルのサブセットにおける包含のために第1のハードエグザンプルを選択し、第1の信頼スコアおよび第1のバウンディングボックスは、特徴マップの第1の一部分に関連付けられること、リファレンスリージョンに対する第1のバウンディングボックスの第1のアライメント度が、しきい値のアライメント度を満たすまたは超えることと決定すること、ならびにしきい値のアライメント度を満たすまたは超えることに少なくとも部分的に基づいて、第2のハードエグザンプルと第1のハードエグザンプルを取り替えることを含むこととを含む動作を行わせる。

【0120】

B．段落Aとしてのシステムは、動作が、リファレンスリージョンに対する第2のハードエグザンプルに関連付けられた第2のバウンディングボックスの第2のアライメント度がしきい値のアライメント度より小さいと決定することに少なくとも部分的に基づいてエグザンプルのサブセットにおける第2のハードエグザンプルを含むことをさらに含むと詳述する。

【0121】

C．段落AまたはBとしてのシステムは、リファレンスリージョンの第1の一部分が、ポジティブエグザンプルに関連付けられるとして示され、特徴マップの第1の一部分が、リファレンスリージョンの第1の一部分の外側においてリファレンスリージョンの第2の一部分に関連付けられると詳述する。

【0122】

D．段落Cとしてのシステムは、リファレンスリージョンの第1の一部分に関連付けられた第1の目標の信頼スコアが、最大の信頼スコアであり、第2の一部分に関連付けられた第2の目標の信頼スコアが、最小の信頼スコアであり、MLモデルをトレーニングすることが、第1のアライメント度がしきい値のアライメント度を満たすまたは超えることに少なくとも部分的に基づいて、第2の目標の信頼スコアの代わりに第1の目標の信頼スコアに第1のハードエグザンプルに関連付けることをさらに含むと詳述する。

【0123】

E．段落A～Dのうちのいずれかとしてのシステムは、第1のアライメント度が、第1のバウンディングボックスと領域との和集合によって分割されたリファレンスリージョンと第1のバウンディングボックスの積集合であると詳述する。

10

20

30

40

50

【0124】

F. コンピューター実装方法は、オブジェクトを表す画像を受信することと、画像においてオブジェクトを表す領域を示すリファレンスリージョンを特定することと、画像の一部分のサブセットを選択することと、サブセットに少なくとも部分的に基づいて、特徴マップ、分類情報に関連付けられた特徴マップの一部分、ROI (region of interest; 関心のリージョン) 情報、および信頼情報を出力するML (機械学習) モデルをトレーニングすることとを含み、画像の一部分のサブセットを選択することは、しきい値の信頼を満たすまたは超える第1の信頼情報を有する第1の一部分に少なくとも部分的に基づいて、特徴マップの第1の一部分を特定することと、比較として、第1の一部分に関連付けられたROI情報をリファレンスリージョンと比較することと、画像の一部分のサブセットとしておよび比較に少なくとも部分的に基づいて、特徴マップの第2の一部分を決定することとを含む。

10

【0125】

G. 段落Fとしてのコンピューター実装方法は、ROIを比較することが、リファレンスリージョンに関して第1の一部分のROI情報のNMS (non-maximal suppression; 非極大抑制) を行うことを含み、特徴マップの第2の一部分が、非極大抑制の下に抑制されたROIに関連付けられると詳述する。

【0126】

H. 段落Gとしてのコンピューター実装方法は、第1の一部分に、最高の信頼スコアに関連付けることをさらに含むと詳述する。

20

【0127】

I. 段落F~Hのうちのいずれかとしてのコンピューター実装方法は、画像に少なくとも部分的に基づいて、分類に対応する複数の出力ROIを実質的に同時に出力するMLモデルをトレーニングすることをさらに含み、分類が、1つまたは複数の車両、歩行者、または自転車運転者を含むと詳述する。

【0128】

J. 段落F~Iのうちのいずれかとしてのコンピューター実装方法は、リファレンスリージョンが、第1のリージョンおよび第2のリージョンを含み、さらに特徴マップの第2の一部分が、第1のリージョンの外的特徴を含むと詳述する。

【0129】

30

K. 段落Jとしてのコンピューター実装方法は、第1のリージョンが、リファレンスリージョンの中央部分と、分類情報のポジティブなアイデンティフィケーションとに関連付けられると詳述する。

【0130】

L. 段落F~Kのうちのいずれかとしてのコンピューター実装方法は、比較することが、ROI情報とリファレンスリージョンとの和集合によって分割されたリファレンスリージョンによって示される領域とROI情報の積集合としてのアライメント量を決定することを含むと詳述する。

【0131】

M. 段落Lとしてのコンピューター実装方法は、比較することが、しきい値のアライメント度を満たすまたは超えるアライメント量を決定することをさらに含むと詳述する。

40

【0132】

N. 段落F~Mのうちのいずれかとしてのコンピューター実装方法は、MLモデルをトレーニングすることが、少なくとも2つステージを含み、少なくとも2つのステージのうちの第1のステージが、MLモデルに出力として、画像の第1のバッチを提供することと、MLモデルからおよび画像の第1のバッチに少なくとも部分的に基づいて、画像の第1のバッチに関する第1の画像に関連付けられた中間ROIを受信することと、リファレンスリージョンとして中間ROIを特定することとを含み、少なくとも2つのステージのうちの第2のステージが、MLモデルに第2の入力として画像の第1のバッチを提供することと、MLモデルからおよび画像の第1のバッチに少なくとも部分的に基づいて、第2の

50

複数のROI、および関連付けられた第2の複数の信頼スコアを受信することと、MLモデルをトレーニングするためのエグザンプルの第2のサブセットにおける包含のためにおよびNMS（非極大抑制）の技法に少なくとも部分的に基づいて、第2の複数のROIに関連付けられた画像の一部分の中からMLモデルをトレーニングするためのハードエグザンプルを選択することと、第2のサブセットに少なくとも部分的に基づいてMLモデルをトレーニングすることと、を含むと詳述する。

【0133】

O. 段落Nとしてのコンピューター実装方法は、第2のステージが、Focal Lossを使用してMLモデルをトレーニングすることをさらに含むと詳述する。

【0134】

P. 段落F～Mのうちのいずれかとしてのコンピューター実装方法は、ROI情報がバウンディングボックスを含むと詳述する。

【0135】

Q. 段落F～Pのうちのいずれかとしてのコンピューター実装方法は、第1の予め定義された数の、第1のオブジェクト分類に関連付けられる画像、および第2の予め定義された数の、第2のオブジェクト分類に関連付けられる画像を含む画像のバッチを受信することと、MLモデルへの入力としての画像のバッチを提供することに少なくとも部分的に基づいて、MLモデルをトレーニングすることをさらに含み、第1の予め定義された数の画像、および第2の予め定義された数の画像は、特徴マップのうちの1つまたは複数の特徴に関連付けられた信頼情報に少なくとも部分的に基づくと詳述する。

【0136】

R. 非一時的なコンピューター読み取り可能な媒体は、実行されると、1つまたは複数のプロセッサに、ML（機械学習）モデルへの入力として画像を提供することと、MLモデルからの出力として、特徴マップ、画像におけるオブジェクトの示す第1のROI（関心のリージョン）を含む特徴マップの第1の一部分、および第1の信頼スコアを受信することと、テスト画像の複数の一部分に関連付けられたポジティブエグザンプルのサブセットおよびネガティブエグザンプルのサブセットに少なくとも部分的に基づいて、MLモデルをトレーニングすることと、を含む動作を行わせる命令のセットを有し、MLモデルをトレーニングすることが、MLモデルにテスト画像を提供し、テスト画像がグラントゥールスのリファレンスリージョンに対応するオブジェクトを表すことと、ハード・ネガティブ・エグザンプルとして、しきい値の信頼スコアを満たすまたは超える第2の信頼スコアを有する特徴マップの第2の一部分を決定することと、ハード・ネガティブ・エグザンプルが非極大抑制の技法の下に抑制される第2のROIに関連付けられると決定することと、ネガティブエグザンプルのサブセットからハード・ネガティブ・エグザンプルを除外することと、を含む。

【0137】

S. 段落Rとしての非一時的なコンピューター読み取り可能な媒体は、第1の一部分の第1のROIが、バウンディングボックスを含み、第1の一部分の第1の信頼スコアが、第1のROIがオブジェクトを表す画像のリージョンを正確に表す確率を示すと詳述する。

【0138】

T. 段落RまたはSとしての非一時的なコンピューター読み取り可能な媒体は、特徴マップの第1の一部分が、複数のROIに関連付けられ、複数のROIのうちのROIが、複数の分類に対応し、分類が、車両、自転車運転者、または歩行者を含むと詳述する。

【0139】

U. システムは、1つまたは複数のプロセッサと、1つまたは複数のプロセッサによって実行可能な命令を格納する1つまたは複数のコンピューター読み取り可能な媒体と含み、命令が、実行されると、システムに、第1のスケールを有する画像を受信することと、第1のML（機械学習）モデルに画像を提供することと、第1のMLモデルから、第1のオブジェクトの少なくとも一部分を表す画像の第1のリージョンの示す第1のバウンディングボックスを受信し、第1のバウンディングボックスがサイズの第1の範囲内の第

10

20

30

40

50

1のサイズを有することと、第2のスケールに少なくとも部分的に基づいて、スケーリングされた画像として画像をスケーリングすることと、第2のMLモデルにスケーリングされた画像を提供することと、第2のMLモデルから、第2のオブジェクトの少なくとも一部分を表す画像の第2のリージョンの示す第2のバウンディングボックスを受信し、第2のバウンディングボックスがサイズの第2の範囲内の第2のサイズを有することを含む動作を行わせる。

【0140】

V. 段落Uとしてのシステムは、実質的に同時に、画像が第1のMLモデルに提供され、スケーリングされた画像が第2のMLモデルに提供されると詳述する。

【0141】

W. 段落UまたはVとしてのシステムは、第1のMLモデルおよび第2のMLモデルが同一のMLモデルであると詳述する。

【0142】

X. 段落U～Wのうちのいずれかとしてのシステムは、動作が、MLモデルへの入力として、異なるサイズのオブジェクトを含むトレーニング画像の第1のバッチを提供することと、オブジェクトに関連付けられたグラントゥールズに少なくとも部分的に基づいて、ROIサイズの関数としてMLモデルの正確度を決定することと、しきい値の正確度を満たすまたは超える正確度に対応するROIサイズの範囲を特定することに少なくとも部分的に基づいて、サイズの第1の範囲を決定することとに少なくとも部分的に基づいて、MLモデルに対してサイズの第1の範囲を決定することをさらに含むと詳述する。

【0143】

Y. コンピューター実装方法は、第1のスケーリングされた画像および第2のスケーリングされた画像を受信し、第1のスケーリングされた画像および第2のスケーリングされた画像が画像に対応することと、第1のML（機械学習）モデルへの第1の入力として、第1のスケーリングされた画像を提供することと、第1のMLモデルからの第1の出力として、サイズの第1の範囲内のサイズを有する1つまたは複数の第1のROI（関心のリージョン）を受信することと、第2のMLモデルへの第2の入力として、第2のスケーリングされた画像を提供することと、第2のMLモデルからの第2の出力として、サイズの第2の範囲内のサイズを有する1つまたは複数の第2のROIを受信することを含む。

【0144】

Z. 段落Yとしてのコンピューター実装方法は、画像が、第1のオブジェクトおよび第2のオブジェクトの表現を含み、コンピューター実装方法が、第1のMLモデルによっておよび第1のスケーリングされた画像に少なくとも部分的に基づいて、第1のオブジェクトを表す第1のROIおよび第2のオブジェクトを表す第2のROIを決定することと、第2のMLモデルによっておよび第2のスケーリングされた画像に少なくとも部分的に基づいて、第1のオブジェクトを表す第3のROIおよび第2のオブジェクトを表す第4のROIを決定することと、をさらに含むと詳述する。

【0145】

AA. 段落Zとしてのコンピューター実装方法は、第1のROIが、サイズの第1の範囲外の第1のサイズを有し、第2のROIが、サイズの第1の範囲内の第2のサイズを有し、第3のROIが、サイズの第2の範囲外の第3のサイズを有し、第4のROIが、サイズの第2の範囲内の第4のサイズを有し、第1のMLモデルが第1のROIを捨て、第2のROIを出力し、第2のMLモデルが、第4のROIを捨て、第3のROIを出力すると詳述する。

【0146】

AB. 段落Y～AAのうちのいずれかとしてのコンピューター実装方法は、第1のスケールに少なくとも部分的に基づいて、画像をスケーリングして第1のスケーリングされた画像を生成することと、第2のスケールに少なくとも部分的に基づいて、画像をスケーリングして第2のスケーリングされた画像を生成することとをさらに含むと詳述する。

【0147】

10

20

30

40

50

A C．段落 Y～A B のうちのいずれかとしてのコンピューター実装方法は、実質的に同時に、第 1 のスケーリングされた画像が第 1 の M L モデルに提供され、第 2 のスケーリングされた画像が第 2 の M L モデルに提供されると詳述する。

【0148】

A D．段落 Y～A C のうちのいずれかとしてのコンピューター実装方法は、1 つまたは複数の第 1 の R O I のうちの第 1 の R O I に関連付けられた正確度に少なくとも部分的に基づいて、サイズの第 1 の範囲を決定することをさらに含み、決定することが、オブジェクトを表す画像の領域に関連付けられたグラントゥールズに少なくとも部分的に基づいて、グラントゥールズによって定義される領域に対して第 1 の R O I の正確度を決定することと、しきい値のアライメント度を満たすまたは超える 1 つまたは複数の第 1 の R O I に対して決定されたアライメント度に対応する R O I サイズの範囲を特定することに少なくとも部分的に基づいて、サイズの第 1 の範囲を決定することと、を含むと詳述する。

10

【0149】

A E．段落 Y～A D のうちのいずれかとしてのコンピューター実装方法は、画像のバッチを受信し、画像のバッチが、第 1 の予め定義された数の、第 1 のオブジェクト分類に関連付けられる画像、および第 2 の予め定義された数の、第 2 のオブジェクト分類に関連付けられる画像を含むことと、第 1 の M L モデルへの入力として画像のバッチを提供することに少なくとも部分的に基づいて、第 1 の M L モデルをトレーニングし、信頼スコアに少なくとも部分的に基づくとしての第 1 の予め定義された数の画像および第 2 の予め定義された数の画像が、第 1 の M L モデルまたは第 2 の M L モデルに関連付けられることと、をさらに含むと詳述する。

20

【0150】

A F．非一時的なコンピューター読み取り可能な媒体は、実行されると、1 つまたは複数のプロセッサに、第 1 のサイズ範囲に関連付けられた第 1 の M L (機械学習) モデルへの入力として、画像を提供することと、第 1 の M L モデルによっておよび画像に少なくとも部分的に基づいて、第 1 の複数の R O I (関心のリージョン) を決定することと、第 1 の複数の R O I のうちの第 1 の R O I のサブセットに関する第 1 の出力を抑制し、第 1 の R O I のサブセットに関する第 1 の出力を抑制することは、第 1 の R O I のサブセットが、第 1 のサイズ範囲外である第 1 のサイズに関連付けられると決定すること含むことと、第 1 の M L モデルからの第 2 の出力として、オブジェクトに関連付けられた 1 つまたは複数の第 1 の R O I、またはオブジェクトの寸法が第 1 のサイズ範囲外であるという第 1 のインディケーションを受信することとを含む動作を行わせる命令のセットを有する。

30

【0151】

A G．段落 A F としての非一時的なコンピューター読み取り可能な媒体は、オブジェクトの寸法が第 1 のサイズ範囲外であるという第 1 のインディケーションを受信することが、第 1 の R O I のサブセットが第 1 の複数の R O I のすべてを含むと決定することに少なくとも部分的に基づくと詳述する。

【0152】

A H．段落 A F または A G としての非一時的なコンピューター読み取り可能な媒体は、命令が、第 2 のサイズ範囲に関連付けられた第 2 の M L モデルへの入力として、画像を提供し、第 2 の M L モデルに画像を提供することが、第 1 の M L モデルに画像を提供することと実質的に同時に起こることと、第 2 の M L モデルによって、第 2 の複数の R O I を決定することと、第 2 の複数の R O I のうちの第 2 の R O I のサブセットに関する第 3 の出力を抑制し、第 2 の R O I のサブセットを抑制することが、第 2 の R O I のサブセットのうちの少なくとも R O I の一部分が、第 2 のサイズ範囲外である第 2 のサイズに関連付けられると決定すること含むことと、第 2 の M L モデルから、オブジェクトに関連付けられた 1 つまたは複数の第 2 の R O I、またはオブジェクトの寸法が第 2 のサイズ範囲外であるという第 2 のインディケーションを受信することとをさらに含むと詳述する。

40

【0153】

A I．段落 A H としての非一時的なコンピューター読み取り可能な媒体は、オブジェク

50

トに対応するROIが、画像におけるオブジェクトの寸法、第1のサイズ範囲、および第2のサイズ範囲に少なくとも部分的に基づいて、第1のMLモデルまたは第2のMLモデルから受信されると詳述する。

【0154】

AJ. 段落AHまたはAIとしての非一時的なコンピューター読み取り可能な媒体は、第1のMLモデルが、第1のサイズ範囲に関連付けられた第1のMLモデルに関する第1の正確度に少なくとも部分的に基づいて、第1のサイズ範囲内のサイズを有する1つまたは複数の第1のROIを出力し、第2のMLモデルが、第2のサイズ範囲に関連付けられた第2のMLモデルに関する第2の正確度に少なくとも部分的に基づいて、第2のサイズ範囲内のサイズを有する1つまたは複数の第2のROIを出力すると詳述する。

10

【0155】

AK. 段落AH～AJのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、動作が、1つまたは複数の第1のROI、または1つまたは複数の第2のROIに少なくとも部分的に基づいて、自律車両の動きを制御するための軌道を生成することをさらに含むと詳述する。

【0156】

AL. 段落AH～AKのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、機械学習モデルに少なくとも部分的に基づいて、第1のサイズ範囲および第2のサイズ範囲を選択することをさらに含むと詳述する。

【0157】

20

AM. 段落AF～ALのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、動作が、第1のMLモデルをトレーニングし、トレーニングすることが、第1のMLモデルへの入力として、リファレンスリージョンによって定義される領域に関連付けられたテストオブジェクトを含むテスト画像を提供すること、および第1のMLモデルによっておよびテスト画像に少なくとも部分的に基づいて、複数のROIを決定することを含むことと、リファレンスリージョンによって定義される領域のうちの領域に対して複数のROIのアライメント度を決定することと、しきい値のアライメント度を満たすまたは超えるアライメント度の一部に対応するオブジェクトサイズの全長を特定することによって少なくとも部分的に基づいて、第1のサイズ範囲を決定することによって少なくとも部分的に基づいて、第1のMLモデルに対して第1のサイズ範囲を決定することをさらに含むと詳述する。

30

【0158】

AN. 段落AF～AMのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、動作が、画像のバッチを受信し、画像のバッチが、第1の予め定義された数の、第1のオブジェクト分類に関連付けられる画像、および第2の予め定義された数の、第2のオブジェクト分類に関連付けられる画像を含むことと、第1のMLモデルへの入力として画像のバッチを提供することに少なくとも部分的に基づいて、第1のMLモデルをトレーニングし、第1の予め定義された数および第2の予め定義された数が、第1のMLモデルまたは第2のMLモデルに関連付けられる信頼スコアに少なくとも部分的に基づくことと、をさらに含むと詳述する。

40

【0159】

AO. コンピューター実装方法は、ML（機械学習）モデルへの入力として、画像を提供することと、MLモデルからの出力として、複数の特徴を含む特徴マップ、信頼スコアを含む複数の特徴のうちの特徴、分類情報、およびNMS（非極大抑制）の技法にしたがって決定されたROI（関心のリージョン）を受信することと、インスタンスセグメンテーションを使用して、同様のROIを有する特徴のサブセットに関連させることとを含む。

【0160】

AP. 段落AOとしてのコンピューター実装方法は、同様のROIを有する特徴のサブセットが、NMSの技法を介して抑制されているとして示される特徴を含むと詳述する。

【0161】

50

AQ. 段落APとしてのコンピューター実装方法は、第1の特徴に関連付けられた第1のROIを出力すると決定することが、第1のROIが、特徴のサブセットに関連付けられた信頼スコアのサブセットのうちの最大の信頼スコアに関連付けられると決定することと、オブジェクトを表すとして画像のリージョンに関連させる第1のROIを出力することを含むと詳述する。

【0162】

AR. 段落AO～AQのうちのいずれかとしてのコンピューター実装方法は、オブジェクトを表すとして画像のピクセルを特定するインスタンスセグメンテーションに対応するマスクを出力することをさらに含むと詳述する。

【0163】

AS. 段落ARとしてのコンピューター実装方法は、MLモデルによって出力されるROIまたはインスタンスセグメンテーションのうちの1つに少なくとも部分的に基づいて、自律車両の動きを制御するための軌道を生成することをさらに含むと詳述する。

【0164】

AT. 段落AO～ASのうちのいずれかとしてのコンピューター実装方法は、テスト画像、および関連付けられたグランドトゥールズに少なくとも部分的に基づいて、MLモデルをトレーニングすることをさらに含み、グランドトゥールズが、マスクを特定すると詳述する。

【0165】

AU. 段落AO～ATのうちのいずれかとしてのコンピューター実装方法は、複数の分類に対応する各特徴に対して、複数のROIを実質的に同時に出力するMLモデルをトレーニングすることをさらに含み、複数の分類が、1つまたは複数の車両、歩行者、または自転車を含むと詳述する。

【0166】

AV. 段落AO～AUのうちのいずれかとしてのコンピューター実装方法は、ROIがバウンディングボックスであると詳述する。

【0167】

AW. システムは、1つまたは複数のプロセッサと、1つまたは複数のプロセッサによって実行可能な命令を格納する1つまたは複数のコンピューター読み取り可能な媒体とを含み、命令が、実行されると、システムに、ML（機械学習）モデルへの入力として、画像を提供することと、MLモデルによって、複数の特徴を含む特徴マップ、ROI（関心のリージョン）を含む複数の特徴のうちの特徴、アイデンティフィケーション、および信頼スコアを決定することと、インスタンスセグメンテーションとして、複数の特徴に関連付けられたアイデンティフィケーションに少なくとも部分的に基づく出力画像を出力することと、を含む動作を行わせる。

【0168】

AX. 段落AWとしてのシステムは、特徴のROIが、NMS（非極大抑制）の技法にしたがって決定され、NMSの技法が、特徴に関連付けられた信頼スコアがしきい値の信頼スコアを満たすまたは超えると決定することと、ROIがしきい値のアライメント度を満たすまたは超える対応する特徴の対応するROIに対するアライメント度に関連付けられると決定することと、ROIとして、対応するROIを決定することと、アイデンティフィケーションとして、対応する特徴の対応するアイデンティフィケーションを決定することを含むと詳述する。

【0169】

AY. 段落AWまたはAXとしてのシステムは、動作が、オブジェクト分類に関連付けられたオブジェクトを表すとして画像のピクセルを特定するマスクとして、インスタンスセグメンテーションを出力することをさらに含むと詳述する。

【0170】

AZ. 段落AYとしてのシステムは、動作が、マスクに少なくとも部分的に基づいて、自律車両に環境の一部分を横切らせるように構成された軌道を決定することと、自律車両

10

20

30

40

50

に軌道を提供することとをさらに含むと詳述する。

【0171】

BA．段落AW～AZのうちのいずれかとしてのシステムは、動作が、MLモデルへの入力としておよび実質的に同時に、第1の予め定義された数の、第1のオブジェクト分類に関連付けられる画像、および第2の予め定義された数の、第2のオブジェクト分類に関連付けられる画像を含む画像のバッチを提供することに少なくとも部分的に基づいて、MLモデルをトレーニングすることをさらに含み、第1の予め定義された数および第2の予め定義された数が、複数の特徴に関連付けられた1つまたは複数の信頼スコアに少なくとも部分的に基づいて、決定されると詳述する。

【0172】

BB．段落AW～BAのうちのいずれかとしてのシステムは、動作が、テスト画像、および関連付けられたグラントゥールスに少なくとも部分的に基づいて、MLモデルをトレーニングすることをさらに含み、グラントゥールスが、リファレンスリージョンおよびマスクを特定すると詳述する。

【0173】

BC．非一時的なコンピューター読み取り可能な媒体は、実行されると、1つまたは複数のプロセッサに、ML（機械学習）モデルにおいて画像を受信することと、MLモデルによって、複数の特徴を含む特徴マップ、ROI（関心のリージョン）およびアイデンティフィケーションを含む複数の特徴のうちの特徴、画像における1つまたは複数のオブジェクトを示すROIを決定することと、インスタンスセグメンテーションとして、複数の特徴に関連付けられたアイデンティフィケーションを出力することを含む動作を行わせる命令のセットを有し、アイデンティフィケーションのうちの一アイデンティフィケーションが、特徴のROI、およびNMS（非極大抑制）の技法に少なくとも部分的に基づいて決定される。

【0174】

BD．段落BCとしての非一時的なコンピューター読み取り可能な媒体は、特徴のROIが、対応する特徴の対応するROI、特徴、および同一のアイデンティフィケーションを有する対応する特徴に基づいて、NMSの技法の下に抑制されるかどうかに基づいて、アイデンティフィケーションがさらに決定されると詳述する。

【0175】

BE．段落BCまたはBDとしての非一時的なコンピューター読み取り可能な媒体は、ROIが、エクステント、および中心の位置を有するバウンディングボックスであると詳述する。

【0176】

BF．段落BC～BEのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、特徴マップが、画像の寸法の小部分である寸法を有すると詳述する。

【0177】

BG．段落BC～BFのうちのいずれかとしての非一時的なコンピューター読み取り可能な媒体は、動作が、テスト画像、および関連付けられたグラントゥールスに少なくとも部分的に基づいて、MLモデルをトレーニングすることをさらに含み、グラントゥールスが、リファレンスリージョンおよびマスクを特定すると詳述する。

【0178】

BH．段落BGとしての非一時的なコンピューター読み取り可能な媒体は、MLモデルをトレーニングすることが、MLモデルへの入力として、第1の予め定義された数の、第1のオブジェクト分類に関連付けられた画像、および第2の予め定義された数の、第2のオブジェクト分類に関連付けられた画像を含む画像のバッチを提供することを含み、第1の予め定義された数および第2の予め定義された数が、複数の特徴に関する1つまたは複数の信頼レベルに少なくとも部分的に基づくとして詳述する。

【0179】

上に説明される例示的な箇条が、ある特定の実装に関して説明されるが、本文書の関連

10

20

30

40

50

において、さらに、例示的な箇条の内容は、方法、デバイス、システム、コンピューター読み取り可能な媒体、および／または別の実装を介して実装されることが可能であることが理解されるべきである。

【0180】

主題が構造的な特徴および／または方法論的な行為に特有の言葉において説明されたが、添付の特許請求の範囲において定義される主題が、説明される特定の特征または行為に必ずしも制限されないということは、理解されることである。むしろ、特定の特征および行為は、特許請求の範囲を実装する例示的なかたちとして開示される。

【0181】

本明細書において説明されるモジュールは、コンピューター読み取り可能な媒体のどんな種類にでも格納されることが可能であり、ソフトウェアおよび／またはハードウェアに実装されることが可能である命令を表す。上に説明される方法および処理のすべては、1つまたは複数のコンピューターもしくはプロセッサか、ハードウェアか、組み合わせかにより実行されるソフトウェアコードモジュールおよび／またはコンピューター実行可能な命令に具現化され完全に自動化されることが可能である。代替として、いくつかまたはすべての方法は、特殊化されたコンピューターハードウェアに具現化されることが可能である。

【0182】

特に他に述べられていない限り、条件付きの語、例えば、とりわけ、「可能である」、「可能であろう」、「あり得る」、または「あり得るであろう」などは、ある例が、他の例は含まないが、ある特徴、エレメントおよび／またはステップを含むことがある文脈内において理解される。したがって、一般に、上記の条件付きの語は、ある特徴、エレメントおよび／もしくはステップが、1つもしくは複数の例に何らかの点において必要とされることを、または必然的に1つまたは複数の例が、ユーザーの入力もしくはプロンプティングの有無に関わらず、ある特徴、エレメントおよび／またはステップが含まれるかどうか、もしくはどんな特定の例にでも行われるべきであるかどうかを決めるためのロジックを含むことを、暗示することが意図されない。

【0183】

特に他に述べられていない限り、結合的な語、例えば、語句「X、Y、またはZのうちの少なくとも1つ」などは、項目、用語などが、各要素の複数の含む、X、Y、もしくはZの、またはどんな組み合わせでも、いずれかであることが可能であることを示すことが理解されるべきである。単数として明示的に説明されない限り、「a」は、単数および複数を意味する。

【0184】

本明細書において説明されるフロー図における、および／または添付の図面に描かれるどんなルーチンの記述、エレメント、またはブロックでも、ルーチン中の特定の論理関数またはエレメントを実装するために、1つまたは複数のコンピューター実行可能な命令を含むモジュール、セグメント、またはコードの部分を潜在的に表すとして理解されるべきである。代替の実装は、エレメントまたはファンクションが、当業者に理解されるように含まれる機能に応じて、実質的に同時に、逆の順に、追加の操作とともに、または操作を省略することを含む、示されたまたは述べられたことから削除される、または順不同で実行されることが可能である、本明細書において説明される例の範囲内に含まれる。

【0185】

多くの変形および修正が、上述した例に対してなされることが可能であり、その要素が、他の許容可能な例の中にあるとして理解されるべきであることを強調すべきである。すべての上記の修正および変形は、本明細書において本開示の範囲内に含まれることが意図され、以下の特許請求の範囲によって保護される。

10

20

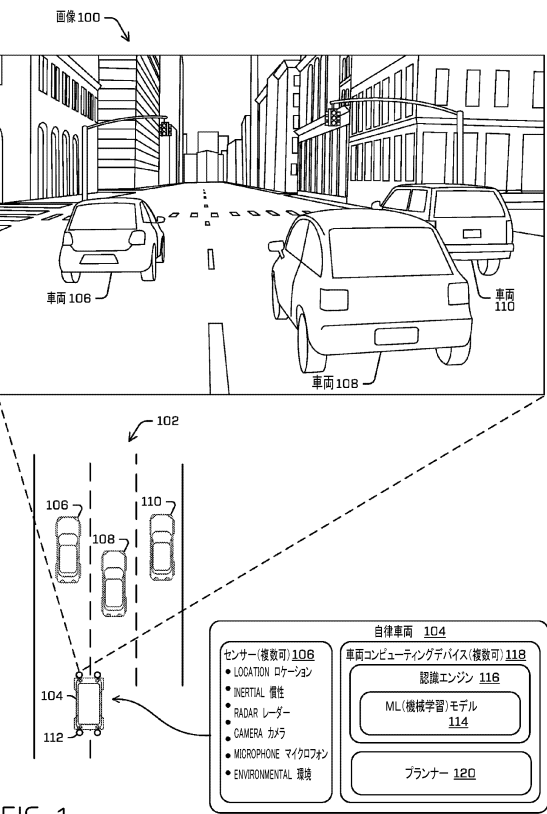
30

40

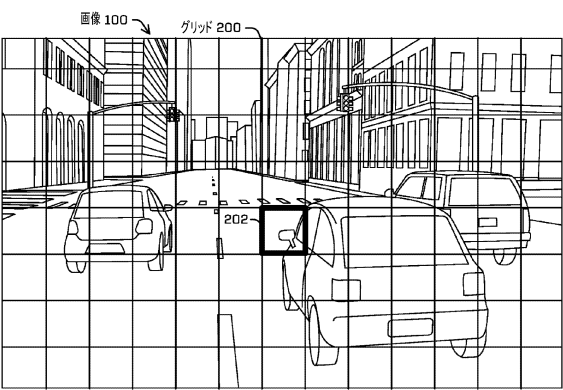
50

【図面】

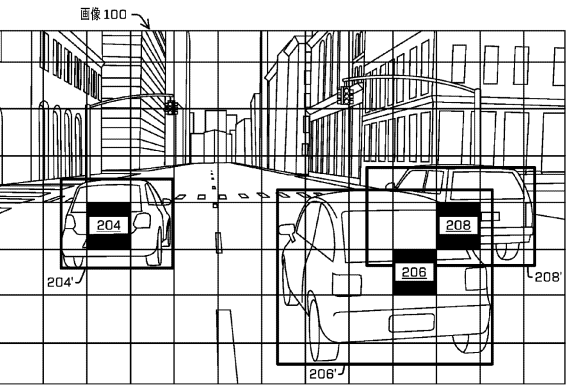
【図 1】



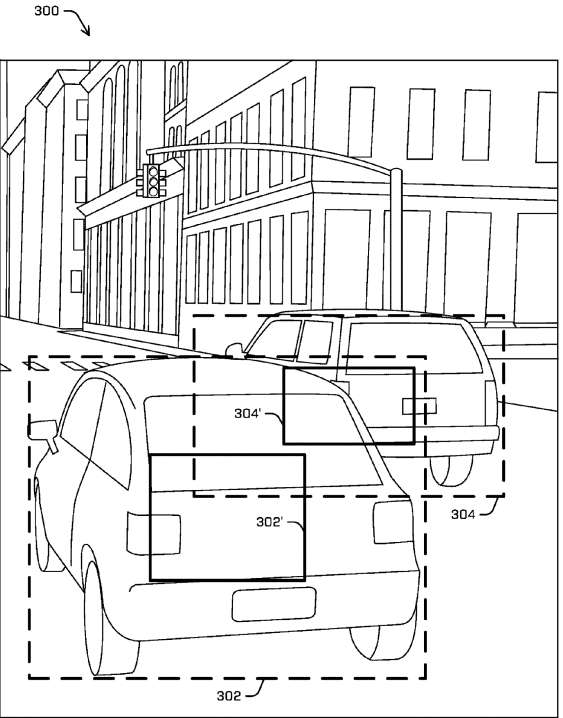
【図 2 A】



【図 2 B】



【図 3】



10

20

30

40

50

【図 4 A】

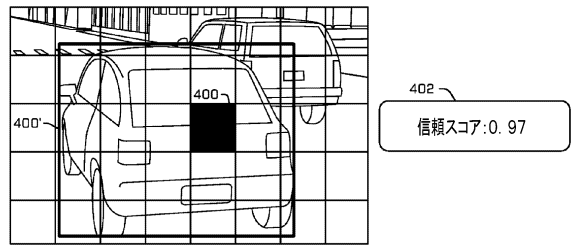


FIG. 4A

【図 4 B】

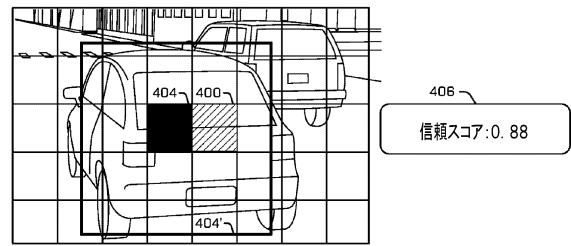


FIG. 4B

10

【図 4 C】

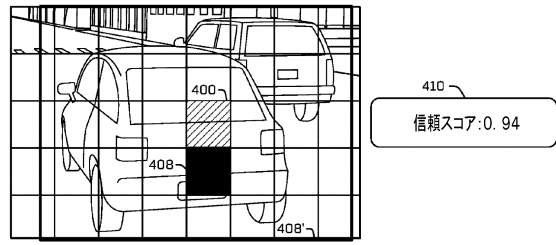


FIG. 4C

【図 4 D】

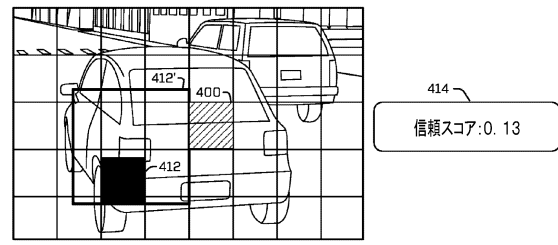


FIG. 4D

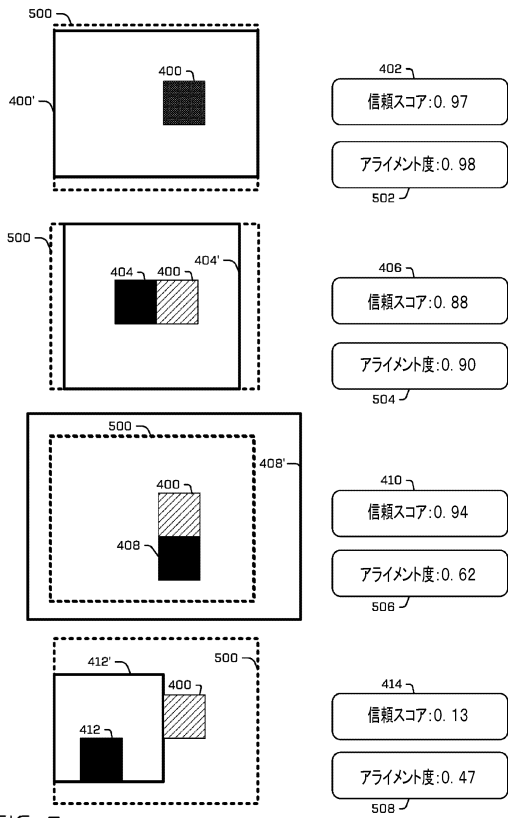
20

30

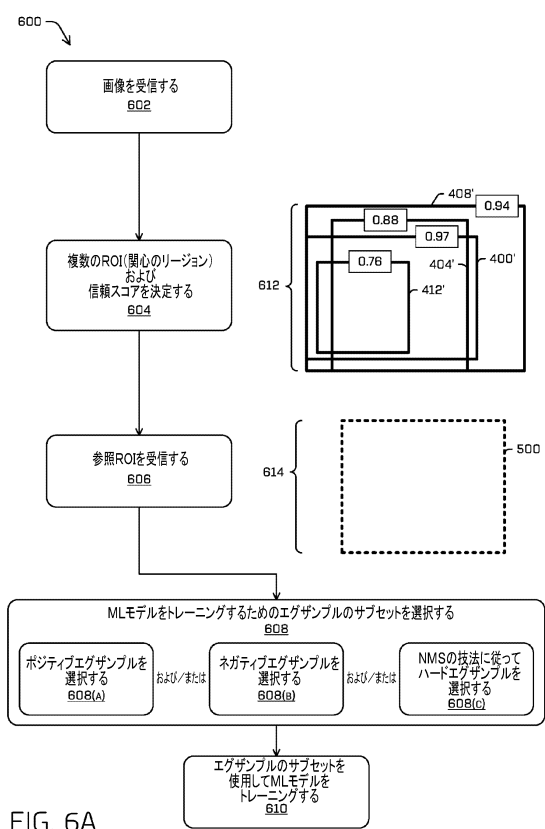
40

50

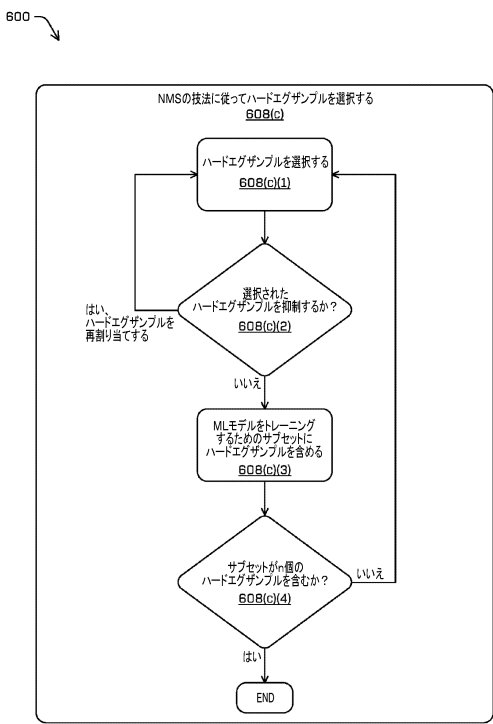
【図 5】



【図 6 A】



【図 6 B】



【図 7 A】

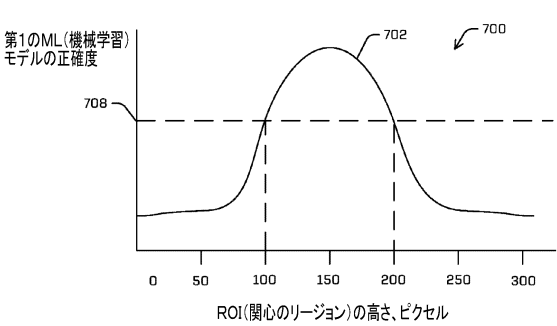


FIG. 6B

10

20

30

40

50

【図 7 B】

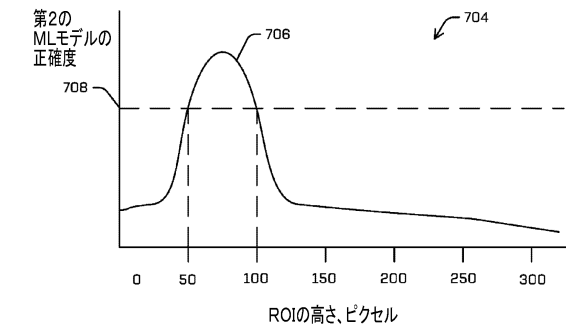


FIG. 7B

【図 8 A】

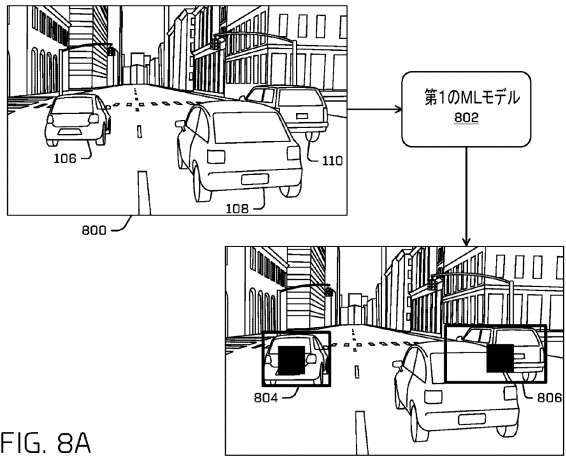


FIG. 8A

【図 8 B】

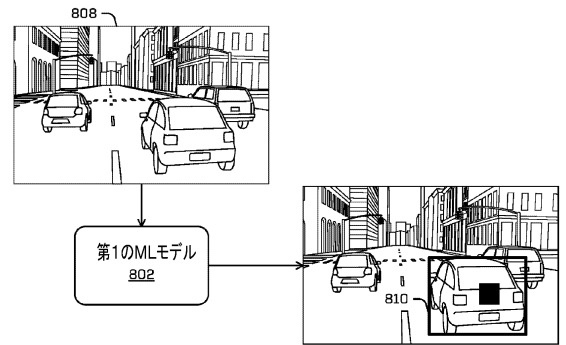


FIG. 8B

【図 8 C】

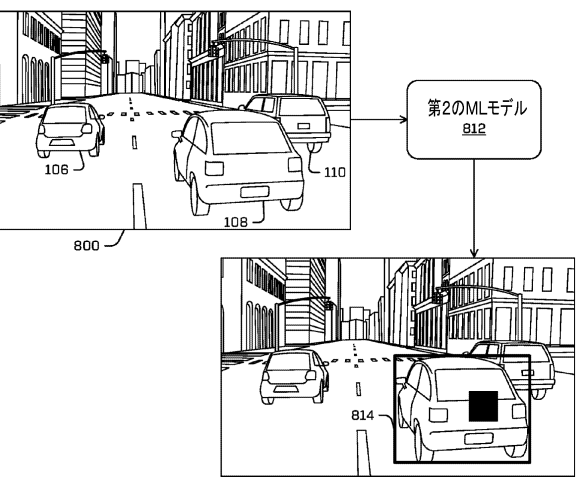


FIG. 8C

【図 9 A】

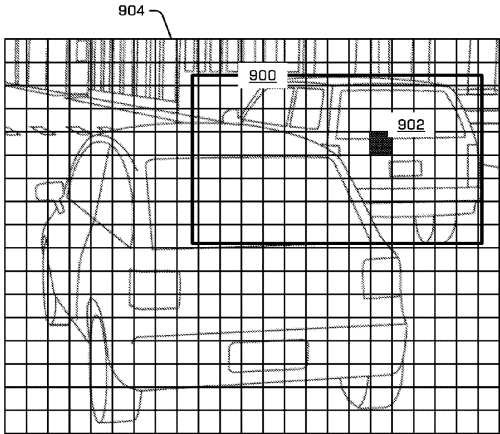


FIG. 9A

【図 9 B】

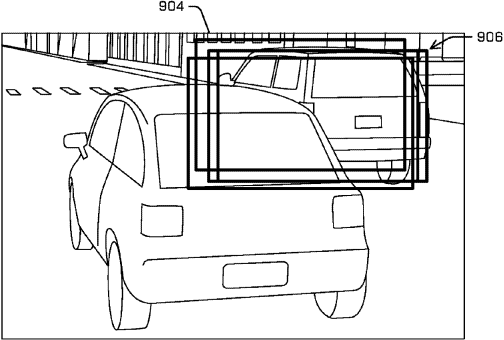


FIG. 9B

【図 9 C】

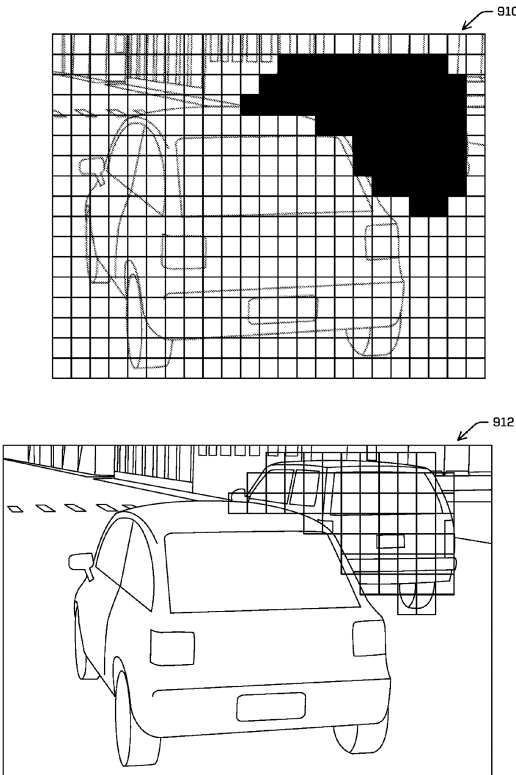


FIG. 9C

【図 10】

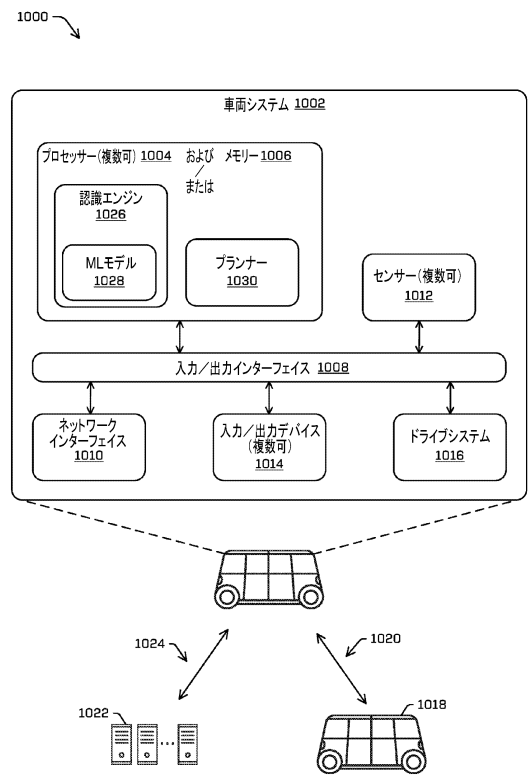


FIG. 10

10

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 16/013,764

(32)優先日 平成30年6月20日(2018.6.20)

(33)優先権主張国・地域又は機関

米国(US)

ニア州 フォスター シティー チェス ドライブ 1149 ズークス インコーポレイテッド内

(72)発明者 クラタース ゴエル

アメリカ合衆国 94404 カリフォルニア州 フォスター シティー チェス ドライブ 1149

ズークス インコーポレイテッド内

審査官 藤原 敬利

(56)参考文献 特開2017-224184 (JP, A)

特開2018-180945 (JP, A)

Navaneeth Bodla et al, Soft-NMS - Improving Object Detection with One Line of Code, 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017年12月25日, h

<https://ieeexplore.ieee.org/document/8237855>

(58)調査した分野 (Int.Cl., DB名)

G06T 7/00-7/90

G06V 10/00-20/90