

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7611248号
(P7611248)

(45)発行日 令和7年1月9日(2025.1.9)

(24)登録日 令和6年12月25日(2024.12.25)

(51)国際特許分類		F I	
G 0 6 T	7/11 (2017.01)	G 0 6 T	7/11
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 B
G 0 6 T	7/174(2017.01)	G 0 6 T	7/174

請求項の数 13 (全39頁)

(21)出願番号	特願2022-528234(P2022-528234)	(73)特許権者	518156417 ズークス インコーポレイテッド 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
(86)(22)出願日	令和2年11月9日(2020.11.9)	(74)代理人	110001243 弁理士法人谷・阿部特許事務所
(65)公表番号	特表2023-501716(P2023-501716 A)	(72)発明者	ブラヴィーン スリニヴァサン 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 ズークス インコーポ レイテッド内
(43)公表日	令和5年1月18日(2023.1.18)	(72)発明者	クラタース ゴエル 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
(86)国際出願番号	PCT/US2020/059734		
(87)国際公開番号	WO2021/096817		
(87)国際公開日	令和3年5月20日(2021.5.20)		
審査請求日	令和5年9月21日(2023.9.21)		
(31)優先権主張番号	16/732,243		
(32)優先日	令和1年12月31日(2019.12.31)		
(33)優先権主張国・地域又は機関	米国(US)		
(31)優先権主張番号	62/935,636		
(32)優先日	令和1年11月15日(2019.11.15)		
	最終頁に続く		最終頁に続く

(54)【発明の名称】 セマンティックおよび/または深度認識インスタンスセグメンテーションのためのマルチタスク学習

(57)【特許請求の範囲】

【請求項1】

1つまたは複数のプロセッサと、
コンピュータ実行可能命令を格納したメモリと、
を含むシステムであって、
前記コンピュータ実行可能命令は、前記1つまたは複数のプロセッサによって実行されると、前記システムに、
画像データを受信することと、
前記画像データの少なくとも一部を機械学習(ML)モデルに入力することと、
前記MLモデルによって、前記画像データに示されるオブジェクトに関連付けられた
関心領域(ROI)を決定することと、
前記MLモデルによって、および前記ROIに少なくとも部分的に基づいて、追加出力を決定することであって、前記追加出力は、
前記オブジェクトに関連付けられたセマンティックセグメンテーションであって、
前記セマンティックセグメンテーションが前記オブジェクトの分類を示す、セマンティックセグメンテーションと、
前記画像の少なくとも前記一部に関連付けられた深度データと、
前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、
ことと、
前記セマンティックセグメンテーション、前記深度データ、および前記インスタンスセグ

10

20

メンテーションに少なくとも部分的に基づいて、3次元ROIを決定することと、前記深度データと前記3次元ROIの境界との間の差に少なくとも部分的に基づいて一貫性損失を決定することと、

トレーニングされたMLモデルとして、および前記一貫性損失に少なくとも部分的に基づいて、前記MLモデルの1つまたは複数のパラメータを変更することと、を含む動作を実行させる、システム。

【請求項2】

前記ROIを決定することが、

第1の解像度に関連付けられた第1の特徴のセットを決定することと、

第2の解像度に関連付けられた第2の特徴のセットを決定することと、

に少なくとも部分的に基づいており、

前記追加出力を決定することは、さらに、前記第1の特徴のセットおよび前記第2の特徴のセットに少なくとも部分的に基づいている、請求項1に記載のシステム。

【請求項3】

前記動作が、前記オブジェクトの中心を示す方向データを決定することをさらに含み、前記インスタンスセグメンテーションを決定することが、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づいている、請求項1または2に記載のシステム。

【請求項4】

前記動作が、

前記オブジェクトの中心を示す方向データを決定することを備え、

前記3次元ROIを決定することは、さらに、前記方向データに少なくとも部分的に基づく、請求項1乃至3のいずれか一項に記載のシステム。

【請求項5】

前記一貫性損失を前記決定することは、

前記セマンティックセグメンテーション、深度データまたはインスタンスセグメンテーションの1つまたは複数に少なくとも部分的に基づいて、2次元境界領域を決定することと、

前記ROIと前記2次元境界領域との間の差を決定することと、

を含む、請求項1乃至4のいずれか一項に記載のシステム。

【請求項6】

前記動作が、前記セマンティックセグメンテーション、前記深度データ、および前記インスタンスセグメンテーションの少なくとも1つに関連付けられた確実性を決定することをさらに含み、前記一貫性損失がさらに、不確実性に少なくとも部分的に基づいている、請求項1乃至5のいずれか一項に記載のシステム。

【請求項7】

コンピュータ実行可能命令を格納した非一時的コンピュータ可読媒体であって、前記コンピュータ実行可能命令は、1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに

画像データを受信することと、

前記画像データの少なくとも一部を機械学習(ML)モデルに入力することと、

前記MLモデルによって、前記画像データに示されるオブジェクトに関連付けられた関心領域(ROI)を決定することと、

前記MLモデルによって、および前記ROIに少なくとも部分的に基づいて、追加出力を決定することであって、前記追加出力は、

前記オブジェクトに関連付けられたセマンティックセグメンテーションであって、前記セマンティックセグメンテーションが前記オブジェクトの分類を示す、セマンティックセグメンテーションと、

前記画像の少なくとも前記一部に関連付けられた深度データと、

前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、

10

20

30

40

50

ことと、

前記セマンティックセグメンテーション、前記深度データ、および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、3次元ROIを決定することと、前記深度データと前記3次元ROIの境界との間の差に少なくとも部分的に基づいて一貫性損失を決定することと、

トレーニングされたMLモデルとして、および前記一貫性損失に少なくとも部分的に基づいて、前記MLモデルの1つまたは複数のパラメータを変更することと、を含む動作を実行させる、非一時的コンピュータ可読媒体。

【請求項8】

前記ROIを決定することが、

第1の解像度に関連付けられた第1の特徴のセットを決定することと、

第2の解像度に関連付けられた第2の特徴のセットを決定することと、

に少なくとも部分的に基づいており、

前記追加出力を決定することは、さらに、前記第1の特徴のセットおよび前記第2の特徴のセットに少なくとも部分的に基づいている、請求項7に記載の非一時的コンピュータ可読媒体。

【請求項9】

前記動作が、前記オブジェクトの中心を示す方向データを決定することをさらに含み、前記インスタンスセグメンテーションを決定することが、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づいている、請求項7または8に記載の非一時的コンピュータ可読媒体。

【請求項10】

前記動作が、

前記オブジェクトの中心を示す方向データを決定することを備え、

前記3次元ROIを決定することは、さらに、前記方向データに少なくとも部分的に基づき、請求項7乃至9のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項11】

前記一貫性損失を前記決定することは、

前記セマンティックセグメンテーション、深度データまたはインスタンスセグメンテーションの1つまたは複数に少なくとも部分的に基づいて、2次元境界領域を決定することと、

前記ROIと前記2次元境界領域との間の差を決定することと、

を含む、請求項7乃至10のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項12】

前記動作が、前記セマンティックセグメンテーション、前記深度データ、および前記インスタンスセグメンテーションの少なくとも1つに関連付けられた確実性を決定することをさらに含み、前記一貫性損失がさらに、不確実性に少なくとも部分的に基づいている、請求項7乃至11のいずれか一項に記載の非一時的コンピュータ可読媒体。

【請求項13】

前記深度データは、離散的な深度を示す深度ピンの出力と、前記深度ピンからのオフセットを示す深度残差とを含む、請求項7乃至12のいずれか一項に記載の非一時的コンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、セマンティックおよび/または深度認識インスタンスセグメンテーションのためのマルチタスク学習に関する。

【0002】

〔関連出願〕

本出願は、2019年11月15日に出願された米国仮出願第62/935,636号

10

20

30

40

50

、および2019年12月31日に出願された米国非仮特許出願第16/732,243号の利益を主張し、これらの全体が本明細書に組み込まれる。

【背景技術】

【0003】

コンピュータビジョンは、自律車両を動作させること、セキュリティ目的で個人を識別することなど、様々なアプリケーションにおいて使用される。コンピュータビジョン技術は、画像内に表される環境に関する情報を決定し、コンピュータがさらなる動作（例えば、検出されたオブジェクトを追跡する）を実行するために使用できる形態でその情報をコンピュータに提供するソフトウェアコンポーネントを構築することを含み得る。オブジェクト検出の精度を向上させるためにコンピュータビジョンの進歩が行われているが、多くのコンピュータビジョン技術は、リアルタイムアプリケーションに有用であるように画像を処理するのに時間がかかりすぎ、複数のニューラルネットワークの使用を必要とする場合があり、それらをレンダリングするメモリ空間を使い果たして、自動運転車両などの様々なアプリケーションに使用することを不可能にする。

10

【図面の簡単な説明】

【0004】

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、参照番号の左端の数字は、その参照番号が最初に現れる図を識別する。異なる図における同じ参照番号は、同様のまたは同等の項目を示す。

【0005】

【図1】図1は、自律車両が、本明細書で説明される機械学習（ML）アーキテクチャを使用して1つまたは複数の出力を決定し、出力を使用して軌道を生成する例示的なシナリオを示す。

20

【図2】図2は、本明細書で説明されるMLアーキテクチャおよびトレーニングコンポーネントを含む例示的なシステムのブロック図を示す。

【図3A】図3Aは、本明細書で説明されるMLアーキテクチャのバックボーンコンポーネントのブロック図を示す。バックボーンコンポーネントは、画像およびバックボーンコンポーネントの層のトレーニングに少なくとも部分的に基づいて特徴を生成し得る。

【図3B】図3Bは、バックボーンコンポーネントの層に関連付けられたMLアーキテクチャの関心領域（ROI）コンポーネントのブロック図を示す。ROIコンポーネントは、画像内で検出されたオブジェクトに関連付けられたROI、ROIに関連付けられた分類、および/または信頼度を生成し得る。

30

【図3C】図3Cは、例示的な画像内で検出されたオブジェクトに関連付けられたROIおよび分類の例を示す。

【図4A】図4Aは、MLアーキテクチャの追加のまたは代替のコンポーネント、すなわち、集約コンポーネント、セマンティックセグメンテーションコンポーネント、センターポージングコンポーネント、および/または深度コンポーネントのブロック図を示す。

【図4B】図4Bは、例示的な画像に少なくとも部分的に基づいて、MLアーキテクチャによってそれぞれ決定されるセマンティックセグメンテーション、方向データ、および深度データの例を示す。

40

【図4C】図4Cは、例示的な画像に少なくとも部分的に基づいて、MLアーキテクチャによってそれぞれ決定されるセマンティックセグメンテーション、方向データ、および深度データの例を示す。

【図4D】図4Dは、例示的な画像に少なくとも部分的に基づいて、MLアーキテクチャによってそれぞれ決定されるセマンティックセグメンテーション、方向データ、および深度データの例を示す。

【図5A】図5Aは、MLアーキテクチャの追加のまたは代替のコンポーネント、すなわち、トリミングおよび/またはプーリングコンポーネントおよび/またはインスタンスセグメンテーションコンポーネントのブロック図を示す。

【図5B】図5Bは、例示的な画像に少なくとも部分的に基づいて、MLアーキテクチャ

50

によって決定されるインスタンスセグメンテーションの例を示す。

【図 5 C】図 5 C は、ML アーキテクチャの追加のまたは代替のコンポーネント、すなわちトリミングおよび/またはプーリングコンポーネントおよび/または 3 次元 ROI コンポーネントのブロック図を図示する。

【図 5 D】図 5 D は、例示的な画像に少なくとも部分的に基づいて、ML アーキテクチャによって決定される 3 次元 ROI の例を示す。

【図 6】図 6 は、本明細書で説明される ML アーキテクチャを使用してオブジェクト検出を生成する、および/またはオブジェクト検出に少なくとも部分的に基づいて自律車両を制御するための例示的なプロセスのフロー図を示す。

【図 7】図 7 は、本明細書で説明される ML アーキテクチャをトレーニングするための例示的なプロセスのフロー図を示す。

【発明を実施するための形態】

【0006】

本明細書で説明される技術は、オブジェクト検出の正確さおよび/または精度を増加させること、単一の機械学習 (ML) モデルから利用可能なオブジェクト検出に関する情報の量を増加させること、様々なコンピュータビジョンアーチファクト (例えば、オブジェクト検出の境界でのトレイル) を減少させること、および/または技術がリアルタイムで実行され得るように処理時間を減少させることによって、コンピュータビジョンを改善し得る。いくつかの例では、本明細書で説明される ML モデルは、消費者グレードのハードウェア (例えば、消費者グレードのグラフィックス処理ユニット (GPU)) 上で每秒 30 以上の速度で、本明細書で説明される 4 つまたは複数の出力を含むオブジェクト検出を出力し得る。この動作速度は、自律車両制御、拡張現実などの多くのリアルタイムアプリケーションに十分である。

【0007】

本明細書で説明される ML アーキテクチャは、画像を受信し、トレーニングされて、4 つまたは複数の出力を出力し得るが、ML アーキテクチャは、より多くまたはより少ない出力を出力し得ることが企図される。いくつかの例では、ML アーキテクチャは、オブジェクトに関連付けられた 2 次元領域 (ROI)、分類、セマンティックセグメンテーション、方向ロジット、深度データ (例えば、深度ピンおよび/または深度残差)、および/またはインスタンスセグメンテーションを含むオブジェクト検出を決定し得る。追加的または代替的に、ML アーキテクチャは、オブジェクトに関連付けられた 3 次元関心領域を出力するためのコンポーネントを含み得る。いくつかの例では、ML アーキテクチャは、単一の順方向伝搬のパスにおいてこのデータのいずれかを出力し得る。

【0008】

本明細書で説明される技術は、ML アーキテクチャのコンポーネントを共同トレーニングすることを備え得、これは、一組のニューラルネットワーク層と、ROI (例えば、2 次元および/または 3 次元)、セマンティックセグメンテーション、方向ロジット、深度データ、および/またはインスタンスセグメンテーションを決定するためのそれぞれのコンポーネントとを含むバックボーン ML モデルを含み得る。簡単にするために、本明細書で説明される出力の各々は、まとめて「タスク」と呼ばれる。例えば、ML アーキテクチャは、オブジェクトに関連付けられた ROI および/または分類を決定するタスクに関連付けられた検出コンポーネント、セマンティックセグメンテーションを決定するタスクに関連付けられた別のコンポーネントなどを含む。

【0009】

いくつかの例では、ML モデルのコンポーネントを共同トレーニングすることは、トレーニングデータセットを ML モデルに提供することと、ML モデルから予測された出力を受信することとを含み得る。例えば、トレーニングデータは、少なくとも第 1 の画像を含み得、予測された出力は、第 1 の画像に関連付けられた本明細書に記載のタスクのそれぞれについて、それぞれの出力を含み得る。コンポーネントを共同トレーニングすることは、出力とトレーニングデータによって示されるそれぞれのグラントゥールス情報との間

10

20

30

40

50

のエラーに基づいてジョイント損失を決定することと、ジョイント損失に少なくとも部分的に基づいてコンポーネントを修正する（例えば、勾配降下を使用して）こととを含み得る。本明細書で説明される技術は、ジョイント損失を調整して、損失の一貫性を強制し得る。

【0010】

例えば、一貫性を強制することは、タスクに関連付けられた不確実性を決定することによって、不確実性は、それによって生成された出力がランドトゥルスデータに対して、正しい/適合するそれぞれのコンポーネントの信頼度を示す、ことと、出力およびランドトゥルスデータに少なくとも部分的に基づいて決定された損失を調整することとを含み得る。調整は、不確実性に少なくとも部分的に基づいて損失をスケールアップすることとを含み得る。一貫性を強制することは、追加的または代替的に、同様になるように信頼度を推進することとを含み得る。例えば、ROIコンポーネントは、2次元ROIおよびそれに関連付けられた信頼度を出力し得、セマンティックセグメンテーションコンポーネントは、同じ分類に関連付けられた画像のピクセルの集合と、各ピクセルに関連付けられたそれぞれの信頼度とを示すセマンティックセグメンテーションを出力し得る。本技術は、セマンティックセグメンテーションに関連付けられた平均信頼度または代表信頼度（例えば、セマンティックセグメンテーションに関連付けられた信頼度にわたって合計エリアテーブルを使用して決定された近似平均）を決定することと、セマンティックセグメンテーションに関連付けられた平均および/または代表信頼度と、2次元ROIに関連付けられた信頼度との間の差に少なくとも部分的に基づいて一貫性損失を決定することとを含み得る。当然、任意の数の一貫性損失は、そのようなネットワークをトレーニングするために使用され得る。追加の例は、ネットワークによって出力されたROIと、インスタンスセグメンテーション、セマンティックセグメンテーション、および/または方向データの1つまたは複数に基づいて決定された境界領域とを比較する（例えば、それらの間の差を決定することと、3次元ROIを画像フレーム内に投影し、得られた投影領域を2次元ROIと比較することと、MLアーキテクチャによって出力されたlidarデータと深度データとの間の差を決定することと、lidarデータ、深度データ、および/または3次元ROIに関連付けられた境界領域などの間の差を決定することと、を含むが、これらに限定されない。

【0011】

いくつかの例では、トレーニングデータ内に含まれるランドトゥルスは、教師ありランドトゥルスデータ（例えば、人間および/または機械にラベル付けされた）、半教師あり（例えば、データのサブセットのみがラベル付けされた）、および/または教師なし（例えば、ラベルが提供されていない場合）であり得る。いくつかの例では、本明細書で説明されるMLアーキテクチャの深度コンポーネントによって生成される深度データに関連付けられた損失を決定するために、lidarデータがランドトゥルスデータとして使用されるときなど、ランドトゥルスデータはまばら（sparse）であり得る。そのようなデータは、半教師あり学習の例であり得る。これらの技術はこれを矯正（remedy）し、それぞれのセンサ測定値をMLアーキテクチャによって生成された出力データのグループ（より濃密な）に関連付けることによって、センサ測定値をランドトゥルスデータの有用なソースとする。

【0012】

例えば、MLアーキテクチャは、画像の各ピクセルまでに関連付けられた深度データを出力し得るが、一方で、画像に関連付けられたlidarポイントの数は、ピクセルの数よりもはるかに少ない場合がある。とはいえ、本技術は、lidarポイントの数、ROI、セマンティックセグメンテーション、インスタンスセグメンテーション、および/または方向データ（例えば、オブジェクトの中心を指す方向ロジット）に少なくとも部分的に基づいて、lidarポイントをピクセルのグループ（または出力の他の離散部分）と関連付けることとを含み得る。ピクセルのグループに関連付けられたlidarポイントは、そのピクセルのグループのランドトゥルスデータとして機能する。

10

20

30

40

50

【 0 0 1 3 】

いくつかの例では、本明細書で説明されるMLアーキテクチャは、それぞれの特徴を生成する一組の層を含むバックボーンコンポーネントを含み得る。本明細書で説明される技術は、これらの特徴を特徴データ構造（例えば、高密度な特徴データマップ）に集約することを含み得る。例えば、特徴を特徴データ構造に集約することは、特徴を共通の解像度にアップサンプリングすることと、アップサンプリングされた特徴の要素ごとの合計および/または連結を決定することとを含み得る。いくつかの例では、特徴データ構造の集約/作成は、追加的または代替的に、合計された特徴を畳み込んでチャンネルの数を減少させること（例えば、チャンネルごとのプーリングを達成するような 1×1 フィルタを使用すること）、その上で1つまたは複数の atrous 畳み込みを実行すること（例えば、拡張率を増加させること）、および/またはもう一度畳み込んでチャンネルの数を復元すること（例えば、特徴を追加のチャンネルに投影するような 1×1 フィルタを使用すること）を含み得る。

10

【 0 0 1 4 】

いくつかの例では、2次元ROIは、バックボーン層によって決定された特徴から直接生成され得、一方で、セマンティックセグメンテーション、方向ロジット、および/または深度データは、特徴データ構造（合計、連結、および/または畳み込みデータ）に少なくとも部分的に基づいて決定され得る。技術は、2次元ROIに少なくとも部分的に基づいてセマンティックセグメンテーション、方向ロジット、および/または深度データをトリミング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ことと、トリミングを一緒に連結することと、トリミングされたおよび連結されたデータからインスタンスセグメンテーションを決定することとを含み得る。同じオブジェクトに関連付けられた3次元ROIを決定することは、オブジェクトのインスタンスセグメンテーションを生成するために使用されるのと同じトリミングされ連結されたデータを取得することと、オブジェクトに関連付けられた画像トリミングとそれに対するインスタンスセグメンテーションとを連結することとを含み得る。換言すれば、3次元ROIを決定することは、セマンティックセグメンテーション、方向ロジット、深度データ、元の画像、および/またはインスタンスセグメンテーションのトリミングに少なくとも部分的に基づいて生成され得る。

20

【 0 0 1 5 】

いくつかの既存のコンピュータビジョン技術とは逆に、本明細書で説明されるコンポーネントは、異なるタスク（例えば、ROI生成、セマンティックセグメンテーションなど）に専念するサブネットワークを有する1つのネットワークの一部であり得る。コンポーネントは共同トレーニングされ得、これは、ネットワークを通じて画像を順方向伝搬し、本明細書で説明されるコンポーネントのそれぞれを通じて本明細書で説明される損失を順方向伝搬することを含み得る。ことが理解される。

30

【 0 0 1 6 】

[例示的なシナリオ]

図1は、車両102を含む例示的なシナリオ100を示している。いくつかの例では、車両102は、米国運輸省道路交通安全局によって発行されたレベル5分類に従って動作するよう構成される自律車両であり得、これは、ドライバー（または乗員）の常時車両制御を期待することなく、全行程に対する全ての安全上重要な機能を実行することが可能な車両を説明する。しかし、他の例では、車両102は、任意の他のレベルまたは分類を有する完全なまたは部分的な自律車両であり得る。本明細書で説明される技術は、自律車両などのロボット制御以外に対して適用され得ることが企図される。例えば、本明細書で説明される技術は、マイニング、製造、拡張現実などおよび/またはコンピュータビジョンを組み込む任意の技術に適用され得る。さらに、車両102は陸上車両として描写されているが、車両102は、宇宙船、船舶、採掘車両などであり得る。いくつかの例では、車両102はシミュレーションされた車両としてシミュレーションにおいて表され得る。簡単にするために、本明細書における説明は、シミュレートされた車両と現実世界の車両と

40

50

を区別しない。したがって、「車両」への言及は、シミュレートされた車両および/または現実世界の車両を指し得る。本明細書で説明されるデータおよび/またはセンサは、現実世界および/またはシミュレートであり得る。

【0017】

本明細書で説明される技術によれば、車両102は、車両102のセンサ104からセンサデータを受信し得る。例えば、センサ104は、画像センサ（例えば、可視光カメラ、赤外線カメラ）位置センサ（例えば、全地球測位システム（GPS）センサ）慣性センサ（例えば、加速度計センサ、ジャイロ스코プセンサなど）磁場センサ（例えば、コンパス）位置/速度/加速度センサ（例えば、速度計、駆動システムセンサ）深度位置センサ（例えば、lidarセンサ、radarセンサ、sonarセンサ、飛行時間（ToF）カメラ、深度カメラ、超音波および/もしくはsonarセンサ、ならびに/または他の深度感知センサ）オーディオセンサ（例えば、マイクロフォン）および/または環境センサ（例えば、気圧計、湿度計など）、を含み得る。

10

【0018】

センサ104は、センサデータを生成し得、これは、車両102に関連付けられたコンピューティングデバイス106によって受信され得る。しかし、他の例では、センサ104および/またはコンピューティングデバイス106のいくつかまたは全ては、車両102から離れて別個であり、および/または遠隔に配置され得、データキャプチャ、処理、コマンド、および/または制御は、有線および/または無線ネットワークを介して1つまたは複数のリモートコンピューティングデバイスによって車両102との間で通信され得る。

20

【0019】

コンピューティングデバイス106は、知覚コンポーネント110、計画コンポーネント112、機械学習（ML）アーキテクチャ114、および/またはシステムコントローラ116を格納するメモリ108を備え得る。いくつかの例では、知覚コンポーネント110は、衝突回避コンポーネントの一部であり得る二次知覚コンポーネントなどの他の知覚コンポーネントの一次知覚コンポーネントであり得る。知覚コンポーネント110は、パイプラインの1つまたは複数のMLコンポーネントの1つであり得るMLアーキテクチャ114を含み得る。MLアーキテクチャ114は、様々なコンピュータビジョンタスクを達成するように、すなわち、画像データに少なくとも部分的に基づいて、車両を取り巻く環境内に何があるかを決定するように、構成され得る。いくつかの例では、知覚コンポーネント110、計画コンポーネント112、および/またはMLアーキテクチャ114は、1つまたは複数のGPU、MLモデル、カルマンフィルタ、コンピュータ実行可能命令などを含み得るハードウェアおよび/またはソフトウェアのパイプラインを含み得る。

30

【0020】

一般に、知覚コンポーネント110は、車両102を取り囲む環境に何があるかを決定し得、計画コンポーネント112は、知覚コンポーネント110から受信した情報に従って車両102をどのように動作させるかを決定し得る。

【0021】

いくつかの例では、知覚コンポーネント110はセンサ104からセンサデータを受信し、車両102の近傍のオブジェクトに関連するデータ（例えば、検出されたオブジェクトに関連付けられた分類、インスタンスセグメンテーション、セマンティックセグメンテーション、2および/または3次元境界ボックス、軌道）、車両の目的地を特定する経路データ、道路の特性（例えば、自律車両をローカライゼーションするのに有用な異なるセンサモダリティにおいて検出可能な特徴）を識別するグローバルマップデータ、車両に近接して検出された特性（例えば、建物、木、フェンス、消火栓、一時停止標識、および様々なセンサモダリティにおいて検出可能な任意の他の特徴の位置および/または寸法）などを識別するローカルマップデータ、を決定し得る。知覚コンポーネント110によって決定されるオブジェクト分類は、例えば、乗用車、歩行者、自転車乗り、配送トラック、セミトラック、交通標識などの異なるオブジェクトタイプを区別し得る。軌道は、過去、

40

50

現在、および/または予測されたオブジェクトの位置、速度、加速度、および/または方向を含み得る。知覚コンポーネント 110 によって生成されたデータは、知覚データと総称され得る。知覚コンポーネント 110 が知覚データを生成すると、知覚コンポーネント 110 は、知覚データを計画コンポーネント 112 に提供し得る。

【0022】

計画コンポーネント 112 は、知覚コンポーネント 110 から受信された知覚データおよび/またはローカライゼーションコンポーネント 226 から受信されたローカライゼーションデータを使用して、1つまたは複数の軌道を決定し、経路またはルートを横断するように車両 102 の動きを制御し、および/またはそうでなければ、車両 102 の動作を制御し得るが、任意のそのような動作は、様々な他のコンポーネントで実行され得る（例えば、ローカライゼーションは、ローカライゼーションコンポーネントによって実行され得、これは知覚データに少なくとも部分的に基づき得る）。いくつかの例では、計画コンポーネント 112 は、知覚データおよび/または他の情報、例えば、1つまたは複数のマップ、ローカライゼーションコンポーネントによって生成されたローカライゼーションデータ（この図では図示せず）など、に少なくとも部分的に基づいて軌道 118 を決定し得る。

10

【0023】

例えば、計画コンポーネント 112 は、第 1 の位置から第 2 の位置への車両 102 の経路を決定し、実質的に同時に、知覚データおよび/またはシミュレートされた知覚データ（そのようなデータにおいて検出されたオブジェクトに関する予測をさらに含み得る）に少なくとも部分的に基づいて、車両がルートを横断するように制御するために（例えば、検出されたオブジェクトのいずれかを回避するために）、receding horizon 技術（例えば、1 マイクロ秒、半秒）に従って車両 102 の運動を制御するための複数の潜在的軌道を生成し、潜在的軌道の 1 つを、車両 102 の駆動コンポーネントに送信され得る駆動制御信号を生成するために使用され得る車両 102 の軌道 118 として選択し得る。図 1 は、方向、速度、および/または加速度を示す矢印として表されるそのような軌道 118 の例を示すが、軌道自体は、コントローラ 116 のための命令を含み得、これは、次いで、車両 102 の駆動システムを作動させ得る。軌道 118 は、車両位置、車両速度、および/または車両加速度をもたらし得るステアリング角度および/またはステアリング速度を実現するように車両 102 の駆動コンポーネントを作動させるためのコントローラ 116 に対する命令を含み得る。例えば、軌道 118 は、コントローラ 116 がトラッキングするためのターゲット方向、ターゲットステアリング角度、ターゲットステアリングレート、ターゲット位置、ターゲット速度、および/またはターゲット加速度を含み得る。

20

30

【0024】

いくつかの例では、コントローラ 116 は、軌道 118 をトラッキングするのに十分な車両 102 の駆動コンポーネントを作動させるためのソフトウェアおよび/またはハードウェアを含み得る。例えば、コントローラ 116 は 1 つまたは複数の比例積分微分 (PID) コントローラを含み得る。

【0025】

いくつかの例では、ML アーキテクチャ 114 は、センサ 104 の 1 つまたは複数の画像センサから画像 120 などの 1 つまたは複数の画像を受信し得る。いくつかの例では、ML アーキテクチャ 114 は、画像センサから画像のストリームを受信し得る。画像センサは、画像を ML アーキテクチャ 114 および/または他のコンポーネントに、ML アーキテクチャ 114 の出力と同期し得るまたはし得ないレートで出力するように構成され得る。本明細書で説明される技術によれば、ML アーキテクチャ 114 は、消費者グレードのハードウェア上で毎秒 30 以上のレートで、本明細書にて説明される出力を生成し得るが、いくつかの例では、ML アーキテクチャ 114 は必要に応じてより低速であり得る。

40

【0026】

いくつかの例では、本明細書で説明される ML アーキテクチャ 114 は、バックボーンコンポーネントおよび様々なサブネットワークを有する単一のネットワークであってよく、

50

それらのすべてが本明細書の説明に従って共同でトレーニングされるが、追加または代替の例では、ネットワークの少なくともいくつかをフリーズ、または1つもしくは複数の他のコンポーネントとは別個にトレーニングし得る。本明細書で説明されるMLアーキテクチャ114は、画像を受信し、画像内のオブジェクトに関連付けられた2次元関心領域(ROI)、画像に関連付けられたセマンティックセグメンテーション、画像に関連付けられた方向データ(例えば、対応するオブジェクトの中心を指すピクセルごとのベクトルを備え得る)、画像に関連付けられた深度データ(深度ピンおよびオフセットの形態であり得る)、オブジェクトに関連付けられたインスタンスセグメンテーション、および/または3次元ROIを出力するように構成され得る。これらのそれぞれは、本明細書では異なるタスクと称され、異なるそれぞれのコンポーネントと関連付けられ得る。少なくとも1つの非限定的な例では、MLアーキテクチャ114は、単一の順方向伝搬における出力を生成し得る。

10

【0027】

ROIは、境界ボックス、いくつかの他の境界形状、および/またはマスクを含み得る。セマンティックセグメンテーションは、それに関連付けられた分類のピクセルごとの表示(例えば、「歩行者」、「車両」、「自転車乗り」、「特大車両」、「連結式車両」、「動物」などのセマンティックラベル)を含み得るが、セマンティックラベルは、画像および/または特徴マップの任意の他の離散部分(例えば、領域、ピクセルのクラスタ)に関連付けられ得る。方向データは、オブジェクトの最も近い中心の方向のピクセルごと(または他の離散部分ごと)の表示を含み得る。画像の離散部分に関連付けられた方向データの部分は、方向ロジットと呼ばれ得、オブジェクトの中心が方向ロジットによって示される離散部分に対する方向にある尤度の表示を含み得る。深度データは、画像センサから画像の一部に関連付けられた表面までの距離の表示を備え得、これは、いくつかの例では、深度「ピン」およびオフセットの表示を含み得る。

20

【0028】

例えば、図1は、画像120を使用してシングルパスにおいてMLアーキテクチャ114によって生成された出力のいくつかを表す出力122を示す。出力122は、画像120内で検出されたオブジェクトと関連付けられた3次元ROI124を含み、深度データのそれぞれの離散部分に重ねられた画像データを伴う深度データを表す。画像120において可視ではない環境の部分は、出力において可視ではなく、深度データは、車両102からの距離が増加するにつれてよりまばらになることに留意されたい。また、出力122の表現は、MLアーキテクチャ114によって生成された4つまたは複数の出力の2つの表現のみを含むことに留意されたい。計画コンポーネント112によって使用される出力122は、画像データ、深度データ、および/または3次元ROIに加えて、またはその代わりに、2次元ROI、方向データ、および/またはインスタンスセグメンテーションを含み得る。

30

【0029】

[例示的なシステム]

図2は、本明細書で説明される技術を実装する例示的なシステム200のブロック図を示す。いくつかの例では、例示的なシステム200は、図1の車両102を表し得る車両202を含み得る。いくつかの例では、車両202は、米国運輸省道路交通安全局によって発行されたレベル5分類に従って動作するよう構成される自律車両であり得、これは、ドライバー(または乗員)の常時車両制御を期待することなく、全行程に対する全ての安全上重要な機能を実行することが可能な車両を説明する。しかし、他の例では、車両202は、他のレベルまたは分類を有する完全にまたは部分的な自律車両であり得る。さらに、いくつかの例では、本明細書に記載の技術は、非自律車両によっても使用可能であり得る。

40

【0030】

車両202は、車両コンピューティングデバイス204、センサ206、エミッタ208、ネットワークインタフェース210、および/または駆動コンポーネント212を含

50

み得る。車両コンピューティングデバイス 204 は、コンピューティングデバイス 106 を表し得、センサ 206 はセンサ 104 を表し得る。システム 200 は、追加的または代替的にコンピューティングデバイス 214 を含み得る。

【0031】

いくつか例では、センサ 206 は、センサ 104 を表し得、lidar センサ、radar センサ、超音波トランスデューサ、sonar センサ、位置センサ（例えば、全地球測位システム（GPS）、コンパスなど）、慣性センサ（例えば、慣性測定ユニット（IMU）、加速度計、磁力計、ジャイロ스코ープなど）、画像センサ（例えば、赤緑青（RGB）、赤外線（IR）、強度、深度、飛行時間カメラなど）、マイクロフォン、ホイールエンコーダ、環境センサ（例えば、温度計、湿度計、光センサ、圧力センサなど）などを含み得る。センサ 206 はこれらまたは他のタイプのセンサのそれぞれの複数の例を含み得る。例えば、radar センサは、車両 202 の角部、前部、後部、側部、および/または上部に位置する個々の radar センサを含み得る。別の例として、カメラは、車両 202 の外部および/または内部に関する様々な場所に配置された複数のカメラを含み得る。センサ 206 は、車両コンピューティングデバイス 204 および/またはコンピューティングデバイス 214 に入力を提供し得る。

10

【0032】

車両 202 はまた、上記のように、光および/または音を放出するためのエミッタ 208 を含み得る。この例におけるエミッタ 208 は、車両 202 の乗客と通信するための内部オーディオおよびビジュアルエミッタを含み得る。限定ではなく例として、内部エミッタは、スピーカー、ライト、サイン、ディスプレイスクリーン、タッチスクリーン、触覚エミッタ（例えば、振動および/またはフォースフィードバック）、機械式アクチュエータ（例えば、シートベルトテンショナー、シートポジションナー、ヘッドレストポジションナーなど）などを含み得る。この例におけるエミッタ 208 はまた外部エミッタを含み得る。限定ではなく例として、この例における外部エミッタは、進行方向を信号で伝えるためのライトまたは車両動作の他のインジケータ（例えば、インジケータライト、サイン、ライトアレイなど）、および歩行者または他の近くの車両と聴覚的に通信するための 1 つまたは複数のオーディオエミッタ（例えば、スピーカー、スピーカーアレイ、ホーンなど）を含み、それらの 1 つまたは複数の音響ビームステアリング技術を含む。

20

【0033】

車両 202 はまた、車両 202 と 1 つまたは複数の他のローカルまたはリモートコンピューティングデバイスとの間の通信を可能にするネットワークインタフェース 210 を含み得る。例えば、ネットワークインタフェース 210 は、車両 202 および/または駆動コンポーネント 212 上の他のローカルコンピューティングデバイスとの通信を容易にし得る。また、ネットワークインタフェース 210 は、追加的または代替的に、車両が他の近くのコンピューティングデバイス（例えば、他の車両、交通信号など）と通信することを可能にし得る。ネットワークインタフェース 210 は、追加的または代替的に、車両 202 がコンピューティングデバイス 214 と通信することを可能にし得る。いくつかの例では、コンピューティングデバイス 214 は、分散コンピューティングシステム（例えば、クラウドコンピューティングアーキテクチャ）の 1 つまたは複数のノードを含み得る。

30

40

【0034】

ネットワークインタフェース 210 は、車両コンピューティングデバイス 204 を別のコンピューティングデバイスまたはネットワーク 216 などのネットワークに接続するための物理的および/または論理的インタフェースを含み得る。例えば、ネットワークインタフェース 210 は、IEEE 200.11 規格によって定義された周波数などを介する Wi-Fi ベースの通信、Bluetooth（登録商標）などの短距離無線周波数、セルラー通信（例えば、2G、3G、4G、4GLTE、5G など）、またはそれぞれのコンピューティングデバイスが他のコンピューティングデバイスとインタフェースすることを可能にする任意の適切な有線もしくは無線通信プロトコルを可能にし得る。いくつかの例では、車両コンピューティングデバイス 204 および/またはセンサ 206 は、ネット

50

ワーク 2 1 6 を介して、所定の期間が経過した後、ほぼリアルタイムでなど、特定の頻度でコンピューティングデバイス 2 1 4 にセンサデータを送信し得る。

【 0 0 3 5 】

いくつかの例では、車両 2 0 2 は、1 つまたは複数の駆動コンポーネント 2 1 2 を含み得る。いくつかの例では、車両 2 0 2 は、単一の駆動コンポーネント 2 1 2 を有し得る。いくつかの例では、駆動コンポーネント 2 1 2 は、駆動コンポーネント 2 1 2 のおよび / または車両 2 0 2 の周囲の状態を検出するための 1 つまたは複数のセンサを含み得る。限定ではなく例として、駆動コンポーネント 2 1 2 のセンサは、駆動コンポーネントのホイールの回転を感知するための 1 つまたは複数のホイールエンコーダ（例えば、ロータリエンコーダ）、駆動コンポーネントの向きおよび加速度を測定するための慣性センサ（例えば、IMU、加速度計、ジャイロスコープ、磁力計など）、カメラまたは他の画像センサ、駆動コンポーネントの周囲のオブジェクトを音響的に検出するための超音波センサ、lidar センサ、radar センサなどを含み得る。ホイールエンコーダなどのいくつかのセンサは、駆動コンポーネント 2 1 2 に固有であり得る。いくつかのケースでは、駆動コンポーネント 2 1 2 上のセンサは、車両 2 0 2 の対応するシステム（例えば、センサ 2 0 6）と重複または補足し得る。

【 0 0 3 6 】

駆動コンポーネント 2 1 2 は、高電圧バッテリー、車両を推進するためのモータ、バッテリーからの直流電流を、他の車両システムによる使用のための交流電流に変換するインバータ、ステアリングモータおよびステアリングラック（電動であり得る）を含むステアリングシステム、油圧または電気アクチュエータを含むブレーキシステム、油圧および / または空気圧コンポーネントを含むサスペンションシステム、トラクションの損失を軽減し、制御を維持するためにブレーキ力を分配するための安定性制御システム、HVAC システム、照明（例えば、車両の外部周囲を照明するためのヘッド / テールライトなどの照明）、および 1 つまたは複数の他のシステム（例えば、冷却システム、安全システム、車載充電システム、DC / DC コンバータなどの他の電気コンポーネント、高電圧接合部、高電圧ケーブル、充電システム、充電ポートなど）を含む、車両システムの多くを含み得る。さらに、駆動コンポーネント 2 1 2 は、センサからデータを受信して前処理を様々な車両システムの動作を制御し得る駆動コンポーネントコントローラを含み得る。いくつかの例では、駆動コンポーネントコントローラは、1 つまたは複数のプロセッサおよび 1 つまたは複数のプロセッサと通信可能に結合されたメモリを含み得る。メモリは、駆動コンポーネント 2 1 2 の様々な機能を実行する 1 つまたは複数のコンポーネントを格納し得る。さらに、駆動コンポーネント 2 1 2 はまた、それぞれの駆動コンポーネントによる、1 つまたは複数の他のローカルまたはリモートコンピューティングデバイスとの通信を可能にする 1 つまたは複数の通信接続部を含み得る。

【 0 0 3 7 】

車両コンピューティングデバイス 2 0 4 は、プロセッサ 2 1 8 と、1 つまたは複数のプロセッサ 2 1 8 と通信可能に結合されたメモリ 2 2 0 とを含み得る。メモリ 2 2 0 は、メモリ 1 0 8 を表し得る。コンピューティングデバイス 2 1 4 はまた、プロセッサ 2 2 2、および / またはメモリ 2 2 4 を含み得る。プロセッサ 2 1 8 および / または 2 2 2 は、データを処理し、本明細書に記載されるような動作を実行するための命令を実行することが可能な任意の適切なプロセッサであり得る。限定ではなく例として、プロセッサ 2 1 8 および / または 2 2 2 は、1 つまたは複数の中央処理ユニット（CPU）、グラフィックス処理ユニット（GPU）、集積回路（例えば、特定用途向け集積回路（ASIC））、ゲートアレイ（例えば、フィールドプログラマブルゲートアレイ（FPGA））、および / または電子データを処理してその電子データをレジスタおよび / またはメモリに格納され得る他の電子データに変換する任意の他のデバイスまたはデバイスの一部を含み得る。

【 0 0 3 8 】

メモリ 2 2 0 および / または 2 2 4 は、非一時的コンピュータ可読媒体の例であり得る。メモリ 2 2 0 および / または 2 2 4 は、オペレーティングシステム、および本明細書で

10

20

30

40

50

説明される方法および様々なシステムに起因する機能を実装するための1つまたは複数のソフトウェアアプリケーション、命令、プログラム、および/またはデータを格納し得る。様々な実施形態において、メモリは、スタティックRAM (SRAM)、シンクロナスDRAM (SDRAM)、不揮発性/フラッシュタイプメモリ、または情報を格納することが可能な任意の他のタイプのメモリのような任意の適切なメモリ技術を用いて実装され得る。本明細書で説明されるアーキテクチャ、システム、および個々の要素は、多くの他の論理的、プログラムの、および物理的なコンポーネントを含み得、それらの添付図面に図示されるものは、単に本明細書での説明に関連する例にすぎない。

【0039】

いくつかの例では、メモリ220および/またはメモリ224は、ローカライゼーションコンポーネント226、知覚コンポーネント228、計画コンポーネント230、MLアーキテクチャ232、マップ234、および/またはシステムコントローラ236を格納し得る。知覚コンポーネント228は、知覚コンポーネント110を表し得、計画コンポーネント230は計画コンポーネント112を表し得、および/またはMLアーキテクチャ232はMLアーキテクチャ114を表し得る。

10

【0040】

少なくとも1つの例において、ローカライゼーションコンポーネント226は、車両202の位置、速度および/または方向(例えば、x位置、y位置、z位置、ロール、ピッチ、またはヨーの1つまたは複数)を決定するためにセンサ206からのデータを受信するハードウェアおよび/またはソフトウェアを含み得る。例えば、ローカライゼーションコンポーネント226は、環境のマップ234を含み、および/または要求/受信し得、マップ234内の自律車両の位置、速度、および/または方向を継続的に決定できる。いくつかの例では、ローカライゼーションコンポーネント226は、SLAM(同時にローカライゼーションおよびマッピング)、CLAMS(同時に較正、ローカライゼーション、およびマッピング)、相対SLAM、バンドル調整、非線形最小二乗最適化などを利用し、画像データ、lidarデータ、radarデータ、IMUデータ、GPSデータ、ホイールエンコーダデータなどを受信し、自律車両の位置、姿勢、および/または速度を正確に決定し得る。いくつかの例では、本明細書で説明されるように、ローカライゼーションコンポーネント226は、車両202の様々なコンポーネントにデータを提供して、軌道を生成するためのおよび/または地図データを生成するための自律車両の初期位置を決定し得る。いくつかの例では、ローカライゼーションコンポーネント226は、マッピングコンポーネント234に、環境に対する車両202の姿勢(例えば、位置および/または方向)、および/またはそれに関連付けられたセンサデータを提供し得る(例えば、マップ234に対する位置および/または方向を介して)。

20

30

【0041】

いくつかの例では、知覚コンポーネント228は、ハードウェアおよび/またはソフトウェアに実装された予測システムを含み得る。知覚コンポーネント228は、車両202の周囲の環境内のオブジェクトを検出し(例えば、オブジェクトが存在することを識別し)、オブジェクトを分類し(例えば、検出されたオブジェクトに関連付けられたオブジェクトタイプを決定し)、センサデータおよび/または環境の他の表現をセグメント化し(例えば、センサデータおよび/または環境の表現の一部を、検出されたオブジェクトおよび/またはオブジェクトタイプに関連付けられていると識別し)、オブジェクトに関連付けられた特性(例えば、オブジェクトに関連付けられた現在の、予測された、および/または以前の位置、方向、速度、および/または加速度を識別する軌道)を決定するなどをし得る。知覚コンポーネント228によって決定されるデータは知覚データと呼ばれる。

40

【0042】

計画コンポーネント230は、ローカライゼーションコンポーネント226から車両202の位置データならびに/または方向データを、および/または知覚コンポーネント228から知覚データを受信し得、このデータのいずれかに少なくとも部分的に基づいて車両202の動作を制御する命令を決定し得る。いくつかの例では、命令を決定することは

50

、命令が関連付けられているシステムに関連付けられたフォーマットに少なくとも部分的に基づいて命令を決定することを含み得る（例えば、自律車両の動きを制御するための第1の命令は、システムコントローラ236および/または駆動コンポーネント212が解析/実行させ得るメッセージおよび/または信号（例えば、アナログ、デジタル、空気圧、運動学的）の第1のフォーマットでフォーマットされ得、エミッタ208のための第2の命令は、それに関連付けられた第2のフォーマットに従ってフォーマットされ得る）。

【0043】

メモリ220および/または224は、追加的または代替的に、衝突回避システム、ライドマネジメントシステムなどを格納し得る。ローカライゼーションコンポーネント226、知覚コンポーネント228、計画コンポーネント230、MLアーキテクチャ232、マップ234、および/またはシステムコントローラ236は、メモリ220に格納されているように図示されるが、これらのコンポーネントのいずれかは、プロセッサ実行可能命令、MLモデル（例えば、ニューラルネットワーク）、および/またはハードウェアを含み得、これらのコンポーネントのいずれかの全てまたは一部はメモリ224に格納、またはコンピューティングデバイス214の一部として構成され得る。いくつかの例では、車両202上で動作するマッピングコンポーネントは、コンピューティングデバイス214への送信のためのセンサデータ（例えば、生センサデータ、センサデータアライメント、知覚ラベル付きセンサデータ）、姿勢データ、および/または知覚データを収集しおよび/または符号化し得る。車両および/またはコンピューティングデバイス214上で動作するマッピングコンポーネントは、本明細書で説明される動作を実行して、リンク修正（a link modification）に少なくとも部分的に基づいてマップを生成し得る。

【0044】

いくつかの例では、コンピューティングデバイス214（および/または204）は、トレーニングコンポーネント238を含み得る。いくつかの例では、トレーニングコンポーネントは、1つまたは複数の自律車両から教師あり、半教師あり、および/または教師なしトレーニングデータを生成および/または収集し、本明細書で説明されるMLアーキテクチャ114をトレーニングするためのコンポーネントを含み得る。

【0045】

MLアーキテクチャ232は、車両202および/またはコンピューティングデバイス214上で動作し得る。いくつかの例では、MLアーキテクチャ232は、センサ206、ローカライゼーションコンポーネント226、パイプライン内の知覚コンポーネント228の他のコンポーネント、および/または計画コンポーネント230から下流（出力を受信する）であり得る。

【0046】

ローカライゼーションコンポーネント226、知覚コンポーネント228、計画コンポーネント230、MLアーキテクチャ232、トレーニングコンポーネント238、および/またはシステム200の他のコンポーネントは1つまたは複数のMLモデルを含み得る。例えば、ローカライゼーションコンポーネント226、知覚コンポーネント228、計画コンポーネント230、MLアーキテクチャ232、および/またはトレーニングコンポーネント238は、それぞれ異なるMLモデルパイプラインを含み得る。いくつかの例では、MLモデルは、ニューラルネットワークを含み得る。例示的なニューラルネットワークは、入力データを一連の接続された層を通過させて出力を生成する生物学的に着想されたアルゴリズムである。ニューラルネットワークにおけるそれぞれの層が別のニューラルネットワークを含むこともでき、または任意の数の層（畳み込み層であるか否か）を含むこともできる。本開示のコンテキストで理解できるように、ニューラルネットワークは機械学習を利用でき、これは、学習されたパラメータに基づいて出力が生成されるそのようなアルゴリズムの広範なクラスを指すことができる。

【0047】

ニューラルネットワークのコンテキストで説明されるが、任意のタイプの機械学習を本開示と一致して使用できる。例えば、機械学習アルゴリズムは、回帰アルゴリズム（例え

10

20

30

40

50

ば、通常最小二乗回帰（OLS）、線形回帰、ロジスティック回帰、段階的回帰、多変量適応回帰スプライン（MARS）、局所的に推定される散布図の平滑化（LOESS）、インスタンススペースのアルゴリズム（例えば、リッジ回帰、最小絶対値縮小選択演算子（LASSO）、弾性ネット、最小角回帰（LARS）、決定木アルゴリズム（例えば、分類回帰木（CART）、反復二分法3（ID3）、カイ二乗自動相互作用検出（CHAID）、決定切り株、条件付き決定木）、ベイジアンアルゴリズム（例えば、ナイーブベイズ、ガウスナイーブベイズ、多項式ナイーブベイズ、平均1依存性推定器（AODE）、ベイジアン信頼度ネットワーク（BNN）、ベイジアンネットワーク）、クラスタリングアルゴリズム（例えば、k平均法、kメジアン法、期待値最大化（EM）、階層的クラスタリング）、関連規則学習アルゴリズム（例えば、パーセプトロン、バックプロパゲーション、ホップフィールドネットワーク、動径基底関数ネットワーク（RBFN）、深層学習アルゴリズム（例えば、深層ボルツマンマシン（DBM）、深層信頼ネットワーク（DBN）、畳み込みニューラルネットワーク（CNN）、積層型オートエンコーダ）、次元削減アルゴリズム（例えば、主成分分析（PCA）、主成分回帰（PCR）、部分最小二乗回帰（PLSR）、サモンマッピング、多次元スケーリング（MDS）、射影追跡法、線形判別分析（LDA）、混合判別分析（MDA）、二次判別分析（QDA）、フレキシブル判別分析（FDA）、アンサンブルアルゴリズム（例えば、ブースティング、ブートストラップ集約（バギング）、エイダブースト、階層型一般化（ブレンディング）、勾配ブースティングマシン（GBM）、勾配ブースト回帰木（GBRT）、ランダムフォレスト）、SVM（サポートベクトルマシン）、教師あり学習、教師なし学習、半教師あり学習などを含むことができるが、これらに限定されない。アーキテクチャの追加の例は、ResNet50、ResNet101、VGG、DenseNet、PointNetなどのニューラルネットワークを含む。

10

20

【0048】

メモリ220は、追加的または代替的に、1つまたは複数のシステムコントローラ236を格納し得、これは、車両202のステアリング、推進、ブレーキ、安全、エミッタ、通信、および他のシステムを制御するように構成され得る。これらのシステムコントローラ236は、駆動コンポーネント212および/または車両202の他のコンポーネントの対応するシステムと通信し、および/または制御し得る。システムコントローラ236は、計画コンポーネント230から受信した命令に少なくとも部分的に基づいて、車両202の動作を制御し得る。

30

【0049】

図2は分散システムとして示されているが、代替の例では、車両202のコンポーネントは、コンピューティングデバイス214に関連付けられ得、および/またはコンピューティングデバイス214のコンポーネントは、車両202に関連付けられ得ることに留意されたい。すなわち、車両202は、コンピューティングデバイス214に関連付けられた機能の1つまたは複数を実行し得、逆もまた同様である。

【0050】

[MLアーキテクチャおよび関連付けられたタスク出力の例]

図3Aは、MLアーキテクチャ232および/またはMLアーキテクチャ114を表し得る、例示的なMLアーキテクチャ300の一部のブロック図を示す。MLアーキテクチャ300は、バックボーンコンポーネント302を含み得る。バックボーンコンポーネントは層304などの1つまたは複数の層を含み得、これは畳み込み層/フィルタ、ReLU関数、バッチ正規化、サブサンプリング関数（例えば、最大プール、平均プール、L2ノルム）、損失関数/フィードバック（少なくともトレーニング中）などを含み得る。いくつかの例では、例示的なMLモデル200は、例えば、畳み込みネットワークなどのニューラルネットワークを含み得る。ニューラルネットワークのコンテキストで説明したが、任意のタイプの機械学習を本開示と一致させて使用し得る。例えば、機械学習アルゴリズムは、回帰アルゴリズム、インスタンススペースのアルゴリズム、ベイジアンアルゴリズム、相関ルール学習アルゴリズム、深層学習アルゴリズムなどを含み得るが、これらに限

40

50

定されない。少なくとも1つの非限定的な例では、バックボーンコンポーネント302は、RetinaNet、VGG、ResNetネットワーク（例えば、ResNet50、ResNet101）などを含み得る。

【0051】

いくつかの例では、バックボーンコンポーネント302の各層は、特徴306乃至310などの特徴を出力し得る。3つの特徴が示されているが、特徴の数は、バックボーンコンポーネント302の層の数に、少なくとも部分的に依存し得ることが理解される。バックボーンコンポーネント302は、この例では3つの層を有するが、バックボーンコンポーネント302は、より少ないまたはより多くを有し得ることが理解される。いくつかの例では、特徴の1つ、例えば、特徴306は層によって出力される特徴マップを含み得る。層の関数が、それへの入力のコピュータおよび/またはニューラルネットワーク変換であり得る出力を含み得るので、特徴306は人的に意味のある用語で記載されない場合がある。したがって、関数は、バックボーンコンポーネント302のそれぞれのコンポーネントの層によって生成された値の高次元フィールド（例えば、ベクトルおよび/またはテンソルを生成した層の学習パラメータに基づいて決定されたデータの固有の特性を表す値のベクトルおよび/またはテンソル）を含み得る。

10

【0052】

いくつかの例では、バックボーンコンポーネント302は、画像120を受信し、画像120を決定された特徴306乃至310にバックボーンコンポーネント302の層の1つまたは複数を通して順方向伝搬し得る。いくつかの例では、特徴306乃至310は、バックボーンコンポーネント302の層の関数に応じて、異なる解像度および/またはサイズを有し得る。例えば、特徴306は、最小サイズを有し得、特徴310は、特徴306乃至310と比較して、最大サイズを有し得る。例えば、第1の層は、以前の層と比較して画像をダウンサンプリングし得る。いくつかの例では、バックボーンの層は、フィルタ/カーネルの寸法に応じて、1つまたは複数の重みまたはそれに関連付けられたバイアス値を有する、および/またはそれに関連付けられた1つまたは複数のハイパーパラメータを有する、フィルタ/カーネルを含み得る。例えば、ハイパーパラメータは、フィルタの寸法（例えば、フィルタに関連付けられた重みの数を決定し得る、例えば、3×3フィルタは、例えば、最大9重みを含み得る）、ストライド、パディング、パディング値（例えば、ゼロパディング、1つのパディング）、拡張率などを含み得る。

20

30

【0053】

図3Bは、MLアーキテクチャ300のROIコンポーネント312乃至316のブロック図を示す。いくつかの例では、ROIコンポーネント312乃至316はそれぞれ、バックボーンコンポーネント302の異なる層から特徴を受信し得る。例えば、ROIコンポーネント312は、バックボーンコンポーネントの層304から特徴306を受信し得る。

【0054】

ROIコンポーネント312乃至316は、それぞれ、オブジェクトに関連付けられたROIおよび/または分類を決定するようにトレーニングされ得る。ROIコンポーネント312乃至316は、YOLO構造などの同じMLモデル構造、および/または同じハイパーパラメータを含み得るが、追加または代替の例では、それらは異なる構造および/またはハイパーパラメータを含み得る。構造は、コンポーネントのサブコンポーネント間の順序、タイプ、および/または接続性を定義し得る（例えば、第1の畳み込み層は、生のセンサデータを受信し、そこから出力を生成し、第2の畳み込み層に出力を提供する第1のマックスプール関数に出力を提供するなど）。コンポーネントに関連付けられたハイパーパラメータは、例えば、畳み込み層内のフィルタの数および/もしくは次元、ならびに/またはコンポーネントに関連付けられた空間範囲、ストライド、パディングの量、パディング値（例えば、0パディング、フラクショナルパディング）、入力サイズ（例えば、次元 $W_1 \times H_1 \times D_1$ または任意の他の次元数を有するテンソル）および/もしくはタイプ（例えば、生センサデータ、例示的なMLモデル200の前のコンポーネントから受信さ

40

50

れたテンソル)、出力サイズおよび/もしくはタイプ(例えば、次元 $W_1 \times H_1 \times D_1$ または $W_2 \times H_2 \times D_2$ を有する次元を有するテンソル)などのような構造のプロパティを定義し得る。パラメータは、ハイパーパラメータとは対照的に、例えば、フィルタなどの層またはそのコンポーネントに関連付けられた重みおよび/またはバイアスなど、トレーニング中に修正される任意のパラメータ含み得る。異なるROIコンポーネント312乃至316によって生成された異なるROIは、特徴306乃至310の解像度の差に少なくとも部分的に基づいて異なるサイズであり得る。いくつかの例では、ROIコンポーネント312乃至316によって生成されたROIは収集され得、冗長ROIは破棄され得、および結果として生じるROIはMLアーキテクチャ300の次の部分に転送される。

【0055】

例えば、ROIコンポーネント312を取ると、ROIコンポーネント312は、特徴306に少なくとも部分的に基づいて、ROI318および/または分類320(図中では「クラス」と略記)を生成し得る。いくつかの例では、ROI318を生成することは境界形状の中心および/または範囲(例えば、寸法)を決定することを含み得、これは分類320に関連付けられたアンカーに少なくとも部分的に基づき得る。分類320は、ROI318に関連付けられた意味分類および/またはROI318に基づくアンカーを含み得る。いくつかの例では、各分類は、1つまたは複数のアンカーに関連付けられ得、ROIコンポーネント312によって出力されるROI318は、複数のROIおよび信頼度の中から最も高い信頼度に関連付けられたROI318であり得る。例えば、ROIコンポーネント312は、(例えば、非最大抑制(NMS)アルゴリズムを使用して)画像内に表されたオブジェクトとの関連付けのために、特徴306および/または特徴306自体に少なくとも部分的に基づいてROIコンポーネント312によって生成された第1の特徴マップ内の複数のROIの中から第1のROIを選択し、第1のROI画像を関連付けるかどうかを決定し得る。いくつかの例では、ROIコンポーネント312は、ROI318に関連付けられた信頼度を出力し得る。

【0056】

いくつかの例では、ROIコンポーネントは、各アンカーについての分類を決定するためのもの、および各アンカーに関してROIサイズを回帰させるためのものの2つのサブネットワークを含み得る。本明細書で使用される場合、2次元ROIは、境界ボックス(または他の形状)、分類、および/または信頼度を含み得る。

【0057】

図3Cは、画像120から検出された車両に関連付けられたROIおよび分類322の例を示す。ROIは、描写される例において境界矩形を含むが、ROIは、追加的または代替的に、アンカー形状に応じて、任意の他のタイプのマスクまたは境界形状であり得る。

【0058】

図4Aを参照すると、図4Aは、MLアーキテクチャ300の追加または代替のコンポーネントのブロック図を示す。例えば、MLアーキテクチャ300は、集約コンポーネント400、セマンティックセグメンテーションコンポーネント402、センターポーティングコンポーネント404、および/または深度コンポーネント406を含み得る。いくつかの例では、ROIコンポーネント312乃至316、セマンティックセグメンテーションコンポーネント402、センターポーティングコンポーネント404、および/または深度コンポーネント406は、本明細書に記載される共同学習技術に少なくとも部分的に基づいて共同学習され得る。バックボーンコンポーネント302によって生成された特徴(例えば、306乃至310)は、集約コンポーネント400にて受信され得る。

【0059】

集約コンポーネント400は、特徴が共通の解像度(例えば、画像120の8分の1スケール、または任意の他の共通のスケール)を有するようにアップサンプリングし、アップサンプリングされた特徴の要素ごとの合計を決定し得る。いくつかの例では、アップサンプリングステージは、畳み込み(例えば、他のフィルタサイズが企図されているが、学習されたパラメータを含み得る 3×3 フィルタを使用して)、バッチ正規化、ReLU、

10

20

30

40

50

および2×バイリニアアップサンプリングを含み得る。特徴のセットの解像度に応じて、特徴のセットは1つまたは複数のアップサンプリングステージを通過させられて、共通の解像度に達し得る。追加的または代替的な例では、特徴は学習されたパラメータを含み得る一連のa t r o u s 畳み込みを通過させられ得る。a t r o u s 畳み込みを含まない上述のアップサンプリングは、十分に意味論的に意味のある高い解像度の特徴マップを達成し得、a t r o u s 畳み込みを使用することと比較して、計算およびメモリ使用量を低減し得る。いくつかの例では、特徴が共通のスケールにアップサンプリングされると、特徴は密度の高い特徴マップとして合計され得る。

【0060】

本技術は、追加または代替として、アップサンプリングおよび合計された特徴のチャンネルの数を低減することによって（例えば、チャンネルワイスプールを実行するために1×1畳み込みを使用して）、受容野を増加させ、および/または高密度特徴マップ内のエッジをさらに分解し、1つまたは複数のa t r o u s 畳み込み（例えば、増加する拡張率で、例えば、2、4、および8の拡張率で3つの畳み込みであるが、任意の他の数の畳み込みまたは拡張率が使用され得る）を実行し、1×1畳み込みを適用することによってチャンネルの数を復元し得る高密度ピクセル単位エンコードを含み得、その任意の畳み込みは、異なる学習されたパラメータを含み得る。これらの動作の結果は、高密度な特徴マップであり得る特徴データ構造408である。この技術は、リアルタイムで使用され得、MLモデルの受容野を増加させながら特徴の解像度を維持する。

【0061】

いくつかの例では、特徴データ構造408は、セマンティックセグメンテーションコンポーネント402、センターポティングコンポーネント404、および/または深度コンポーネント406によって使用され得る。この共有データ使用は、計算および/またはメモリ使用を低減し得る。いくつかの例では、セマンティックセグメンテーションコンポーネント402、センターポティングコンポーネント404、および/または深度コンポーネント406は、それぞれ、特徴データ構造408を本明細書で説明されるタスク固有の出力に投影するためのフィルタを含み得る。

【0062】

セマンティックセグメンテーションコンポーネント402は、画像120のセマンティックセグメンテーション410および/またはそれに関連付けられた信頼度412を決定し得る。例えば、セマンティックセグメンテーションは、画像120の離散部分に関連付けられたセマンティックラベル（例えば、ピクセルごとの分類ラベル）および/または分類が正しい尤度を示す信頼度を含み得る。例えば、図4Bは、画像120の一部に関連付けられた例示的なセマンティックセグメンテーション414を示す。いくつかの例では、セマンティックセグメンテーションコンポーネント402は、セマンティックセグメンテーション410および/または信頼度412（例えば、特徴データ構造408をセマンティックセグメンテーションおよび/または信頼度空間に投影する）を生成するために、1×1畳み込み、4×バイリニアアップサンプリング、および/またはs o f t m a x層を含み得る。例示的なセマンティックセグメンテーション414は、分類「車両」に関連付けられた複数の離散部分（例えば、ピクセル）と、分類「地上」に関連付けられた複数の他の離散部分とを示す。いくつかの例では、信頼度はロジットによって示され得る。

【0063】

センターポティングコンポーネント404は、特徴データ構造408に少なくとも部分的に基づいて方向データ416を決定し得、方向データは、画像120の離散部分に関連付けられた方向および/または信頼度を含む。いくつかの例では、信頼度はロジットによって示され得るが、確率などの他の例が企図される。方向は、離散部分から最も近いオブジェクトの中心への方向を示し得る。図4Cは、画像120の一部のそれぞれの離散部分に関連付けられた非常に限られた数の方向ロジットを含む例示的な方向データ418を示す。灰色の線は、方向データには現れず、視覚的参照のためにのみ現れることに留意されたい。

10

20

30

40

50

【 0 0 6 4 】

深度コンポーネント 4 0 6 は、画像 1 2 0 の離散部分に関連付けられた深度ビン 4 2 0 および / または深度残差 4 2 2 を決定し得る。いくつかの例では、深度ビンは、画像センサおよび / またはビンの中心 (および / または任意の他の中間点) からの距離の範囲を含み得る。いくつかの例では、深度ビン 4 2 0 を決定することは分類タスクであり得、一方、深度残差を決定することは回帰タスクであり得る。いくつかの例では、深度残差は、深度ビンに少なくとも部分的に基づき得る。深度残差は、例えば、深度ビンの中心または深度ビンの端部など、深度ビンに関連付けられた基準点からのオフセットを含み得る。図 4 D は、画像 1 2 0 に関連付けられて深度コンポーネント 4 0 6 によって決定される例示的な深度データ 4 2 4 を示す。いくつかの例では、深度コンポーネント 4 0 6 は、残差 4 2 2 を最終深度として出力深度ビン 4 2 0 の中心を合計し得る。

10

【 0 0 6 5 】

さらに説明するための非限定的な例では、深度コンポーネント 4 0 6 は、深度間隔 $[\alpha, \beta]$ を K 個の重複しない離散深度ビンに量子化 (離散化) し得る。いくつかの例では、量子化は、ログ空間、逆深度空間、または線形空間において行われ得るが、ログ空間が本明細書で説明される。ログ空間に深度ビンを設定することは、大きな深度値を有する領域におけるダウンウェイトトレーニング損失の効果を有し得、これは、画像が一般に遠隔のオブジェクトに関するより少ない情報を含むため、より不正確であり得る (例えば、出力 1 2 2 を参照) 。少なくとも 1 つの非限定的な例では、ビンのエッジは、

20

【 0 0 6 6 】

【 数 1 】

$$t_i = \exp \left(\log \alpha + \frac{\log \left(\frac{\beta}{\alpha} \right) * i}{K} \right) \quad (1)$$

【 0 0 6 7 】

に従って定義され得、ここで、 $t_i \in \{t_1, \dots, t_{K-1}\}$ は K 個のエッジであり、 t_{i+1} はビンの他端の対応するエッジである。ピクセルについてのグラントゥールス深度が g_d である場合、 $\epsilon \in [t_i, t_{i+1})$ である場合、およびその場合に限り、それはビンインデックス $i \in \{0, \dots, K-1\}$ を割り当てられ得る。いくつかの例では、回帰は、少なくとも部分的に中点に基づき得る。ログ空間における i 番目のビンの中点は、

30

【 0 0 6 8 】

【 数 2 】

$$m_i = \frac{\log(t_{i+1}) + \log(t_i)}{2} \quad (2)$$

40

【 0 0 6 9 】

と定義され得る。

【 0 0 7 0 】

ビン i に入るグラントゥールス深度 g_d について、グラントゥールス残差 r_d は、

【 0 0 7 1 】

【 数 3 】

50

$$r_d = \frac{(\log(g_d) - m_i)}{\log(t_{i+1}) - \log(t_i)} \quad (3)$$

【 0 0 7 2 】

として計算され得る。

【 0 0 7 3 】

推論時に、深度コンポーネント406は、畳み込み（例えば、他のフィルタサイズが企図されるが、学習されたパラメータを備え得る 1×1 フィルタを使用する）を特徴データ構造408に適用して、そのピクセルにおける深度が対応する深度ビン（例えば、ロジット l_i について、 $[t_i, t_{i+1})$ ）に入る尤度に対応するピクセルごとの K 個のsoftmaxロジット l_0, \dots, l_{K-1} を生成し得る。いくつかの例では、深度コンポーネント406は、残差を生成するために並列計算を行い得る。例えば、深度コンポーネント406は、特徴データ構造408を畳み込み（例えば、他のフィルタサイズが企図されるが、学習されたパラメータを備え得る 1×1 フィルタを使用して）、ピクセルごとの K 個の予測された残差 r_0, \dots, r_{K-1} を生成し得る。いくつかの例では、深度コンポーネント406は、予測された残差値に制限を課し得ないが、他の例では、深度コンポーネント406は、予測された残差をビン幅の半分に制限するなどのように、予測された残差を制限し得る。

10

20

【 0 0 7 4 】

ログ空間を使用する例では、特定のピクセルおよび特定のビン i についての深度値を決定することは、以下の式

【 0 0 7 5 】

【数4】

$$d_i = \exp(m_i + r_i(\log(t_{i+1}) - \log(t_i))) \quad (4)$$

【 0 0 7 6 】

を評価すること、を含み得る。

【 0 0 7 7 】

いくつかの例では、ピクセルに関連付けるためにビンの中から深度 d_i を選択するために、深度コンポーネント406は、最大ロジット l_i を有する深度を選択し得る。深度コンポーネント406は、追加または代替として、局所近傍内のロジットに関連付けられた深度値に少なくとも部分的に基づいて、この深度を平滑化し得る。例えば、平滑化は、局所近傍内のロジットから決定された確率分布に少なくとも部分的に基づいて行うことができる。局所近傍は、例えば、 $N = (i-1, i, i+1)$ として定義され得るが、近傍をより広く定義することが可能であり、平滑化は、

30

40

$$\sum_{j \in N} P_j d_j$$

として定義され得、ここで、

$$P_j = \frac{\exp(l_j)}{\sum_{k \in N} \exp(l_k)}$$

である。

【 0 0 7 8 】

いくつかの例では、本明細書で説明される深度コンポーネント406の動作は、「トレ

50

イル」アーチファクトを減少させ得る。これは、最も高いロジットを有する深度ピンを選択することが、各ピクセルにおける潜在的にマルチモーダルな深度分布の単一モードの選択を可能にするためであり得る。これにより、ピクセルは、背景深度またはオブジェクト深度のいずれかに暗示的に割り当てられ得る。

【0079】

図5Aを参照すると、図5Aは、MLアーキテクチャ300の追加または代替のコンポーネント、すなわち、トリミングおよび/もしくはプーリングコンポーネント500ならびに/またはインスタンスセグメンテーションコンポーネント502、のブロック図を示す。いくつかの例では、トリミングおよび/またはプーリングコンポーネント500は、ROIを受信し(Dにおいて)、ROIに関連付けられたセマンティックセグメンテーション410、方向データ416、ならびに/または深度データ420および/もしくは422の(例えば、トリミングおよび/またはプーリング)部分を決定し得る。トリミングおよび/またはプーリングコンポーネント500は、共通の解像度でない結果として生じる部分のいずれかをアップサンプリングし、部分を一緒に連結し得る(504にて)。いくつかの例では、トリミングおよび/またはプーリングコンポーネント500は、合計されたエリアテーブルに少なくとも部分的に基づいて、セマンティックセグメンテーション410のトリミングに関連付けられた信頼度を決定し得る。いくつかの例では、セマンティックセグメンテーション410のトリミングに合計エリアテーブルを適用することは、セマンティックセグメンテーション410のトリミングに関連付けられた信頼度の平均信頼度の近似を示す代表信頼度を決定し得る。追加または代替の例では、トリミングおよび/またはプーリングコンポーネント500は、セマンティックセグメンテーション410のトリミングに関連付けられた平均信頼度を決定し得る。いくつかの例では、代表または平均信頼度は、トレーニング中および/または推論中に使用され得る。

【0080】

いくつかの例では、インスタンスセグメンテーションコンポーネント502は、セマンティックセグメンテーション410のトリミングされた部分、方向データ416、ならびに/または深度データ420および/もしくは422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506を生成し得る。いくつかの例では、インスタンスセグメンテーションコンポーネント502は、セマンティックセグメンテーション410、方向データ416、ならびに/または深度データ420および/もしくは422を(例えば、学習されたパラメータを備え得る1×1フィルタを使用して)畳み込んで、オブジェクトが検出されたか否かのバイナリ表示を決定し得る。例えば、図5Bは、例示的なインスタンスセグメンテーション508を示す。オブジェクトの分類またはオブジェクトの境界を定める形状を示すROIを区別するセマンティックセグメンテーションとは異なり、インスタンスセグメンテーション508は、オブジェクトが検出されるまたは検出されないというバイナリ表示を含み得る。

【0081】

第1の非限定的な例では、インスタンスセグメンテーションコンポーネント502は、セマンティックセグメンテーションデータ410および方向データ416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506を決定し得る。例えば、インスタンスセグメンテーションコンポーネント502は、セマンティックセグメンテーションから分類に関連付けられたチャンネル(例えば、歩行者チャンネル)を選択し、歩行者チャンネル内のROI出力に少なくとも部分的に基づいて領域をトリミングし得る。インスタンスセグメンテーションコンポーネント502は、方向チャンネルから領域の方向ロジットを集め(例えば、プーリングし)、プーリングされた方向ロジットにそってトリミングされたセマンティックセグメンテーションロジットを使用して、前景/背景セグメンテーションを行い得る。第2の非限定的な追加的または代替的な例では、インスタンスセグメンテーションコンポーネント502は、第1の非限定的な例で説明される動作におけるセマンティックセグメンテーション410のために深度データ420および/または422を置き換え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506を決

10

20

30

40

50

定し得る。第3の非限定的な例では、インスタンスセグメンテーションコンポーネント502は、セマンティックセグメンテーションデータ410、方向データ416、ならびに深度データ420および/または422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506を決定し得る。そのような例では、そのトリミングされた(および/またはプーリングされた)部分の各々は、(例えば、学習されたパラメータを含み得る1×1フィルタを使用して)連結され、畳み込まれ得る。

【0082】

図5Cは、MLアーキテクチャ300の追加のまたは代替のコンポーネント、すなわちトリミングおよび/またはプーリングコンポーネント510および/または3次元ROIコンポーネント512のブロック図を図示する。トリミングおよび/またはプーリングコンポーネント510は、トリミングおよび/またはプーリングコンポーネント500と同じまたは異なるコンポーネントであり得、それらのいずれかは、それぞれ、インスタンスセグメンテーション502および/または3次元ROIコンポーネント512の一部であり得る。いくつかの例では、インスタンスセグメンテーションのためのトリミングおよび/またはプーリングコンポーネント500によってトリミングおよび/またはプーリングされたデータは、インスタンスセグメンテーション502および画像120と共にトリミングおよび/またはプーリングコンポーネント510に提供され得る。トリミングおよび/またはプーリングコンポーネント500でデータをトリミングおよび/またはプーリングするために使用される同じROIは、画像120および/またはインスタンスセグメンテーション506をトリミングするために使用され得、それらの各々またはいずれかは、504においてアップサンプリングおよび連結された、セマンティックセグメンテーションデータ410、方向データ416、ならびに深度データ420および/または422の部分にアップサンプリングおよび/または連結され得る(514において)。

【0083】

3次元ROIコンポーネント512は、学習されたパラメータを含むフィルタを含み得る1つまたは複数の畳み込み層を含み得る。3次元ROIコンポーネント512は、トリミングされた、プーリングされた、アップサンプリングされた、および/または連結された画像、インスタンスセグメンテーション、セマンティックセグメンテーションデータ410、方向データ416、ならびに/または深度データ420および/もしくは422に少なくとも部分的に基づいて、3次元ROI516を生成し得る。

【0084】

図5Dは、3次元ROIコンポーネント512によって決定される3次元ROI518の例を示す。描写される例における3次元ROI518は、3次元境界ボックスである。いくつかの例では、3次元境界ボックスは、それによって識別されるオブジェクトに関連付けられた位置、方向、姿勢(例えば、方向)、および/またはサイズ(例えば、長さ、幅、高さなど)を含み得る。

【0085】

[例示的な処理]

図6は、本明細書で説明されるMLアーキテクチャを使用してオブジェクト検出を生成する、および/またはオブジェクト検出に少なくとも部分的に基づいて自律車両を制御するための例示的なプロセス600のフロー図を示す。いくつかの例では、例示的なプロセス600は、知覚コンポーネント228および/またはMLアーキテクチャ300によって実行され得る。

【0086】

動作602において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、画像データを受信することを含み得る。画像データは、本明細書で説明されるMLアーキテクチャに入力され得る。

【0087】

動作604において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、MLアーキテクチャによるオブジェクト検出を決定することを含み得る。い

10

20

30

40

50

くつかの例では、オブジェクト検出は、ROI、分類、セマンティックセグメンテーション、深度データ、インスタンスセグメンテーション、および/または3次元ROIを含み得る。オブジェクト検出を決定することは、MLアーキテクチャの異なる部分によって達成される、本明細書で説明される1つまたは複数の動作(例えば、動作606乃至622の少なくとも1つ)を含み得、これは、コンポーネントのパイプラインを含み得る、

【0088】

動作606において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、バックボーンコンポーネントによって、画像データに少なくとも部分的に基づいて、特徴のセットを決定することを含み得る。特徴のセットは、1つまたは複数の特徴マップ(例えば、異なる解像度で)であってよく、特徴マップの特徴は、画像データの一部に関連付けられた値を含み得る。例えば、バックボーンコンポーネントは、RetinaNet、VGG、ResNetネットワーク(例えば、ResNet50、ResNet101)などを含み得、特徴のセットは、1つまたは複数の特徴マップあり得、それらのそれぞれは、バックボーンコンポーネントの異なる層によって出力され得る。

【0089】

動作608において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、特徴のセットを特徴データ構造に集約することを含み得る。例えば、特徴のセットは、異なる解像度の1つまたは複数の特徴マップを含み得る。特徴のセットを集約することは、1つまたは複数の特徴マップを共通の解像度にスケーリングすることと、スケーリングされた特徴マップを特徴データ構造に要素ごとに合計することと、を含み得る。要素ごとに合計することに加えてまたは代替として、本技術は、要素ごとに合計された特徴マップをダウンサンプリングすること(例えば、チャンネルごとにプーリングするために1x1畳み込みを使用すること)、増大する拡張率を使用して1つまたは複数のatrous畳み込みを決定すること、および/または結果として生じる特徴マップをアップサンプリングすることを含む、高密度なピクセルごとの符号化を含み得る。いくつかの例では、結果として生じる特徴データ構造マップは、MLアーキテクチャの1つまたは複数のコンポーネントに提供され得る。例えば、特徴データ構造は、ROIコンポーネント、セマンティックセグメンテーションコンポーネント、センターポティングコンポーネント、インスタンスセグメンテーションコンポーネント、および/または3次元ROIコンポーネントへの入力として提供され得る。

【0090】

動作610において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、バックボーンコンポーネントによって決定される特徴のセットに少なくとも部分的に基づいてROIを決定することを含み得る。いくつかの例では、ROIコンポーネントは、検出されたオブジェクトによって占有されているとしてROIが示す画像の領域に関連付けられた2次元ROI、分類、および/または信頼スコアを生成し得る。いくつかの例では、ROIコンポーネントは、バックボーンコンポーネントの各層に関連付けられ得、異なるサイズ/解像度に関連付けられたROIを生成し得る。例えば、第1のROIコンポーネントは小さなオブジェクトを検出し得、第2のROIコンポーネントはより大きなオブジェクトを検出し得る。しかし、他の技術が企図される。

【0091】

動作612において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、特徴データ構造に少なくとも部分的に基づいてセマンティックセグメンテーションを決定することを含み得る。いくつかの例では、セマンティックセグメンテーションは、分類(例えば、自転車、歩行者、車両)に関連付けられているとして画像の領域を識別し得る。MLアーキテクチャのセマンティックセグメンテーション部分は、追加的または代替的に、セマンティックセグメンテーションの離散部分(例えば、ピクセル)と関連付けられた信頼度を決定することを含み得る動作612を達成し得る。いくつかの例では、セマンティックセグメンテーション部分は、元の画像の解像度でピクセルごとの分類を生成するために、1x1畳み込み、4xバイリニアアップサンプリング、およびソフト

10

20

30

40

50

マックス層を含む出力ヘッドを含み得るが、他の構成が企図される。1×1畳み込みは、本明細書で説明される技術に従ってトレーニングされる学習されたパラメータを含み得、1×1畳み込みは、代替的に、別のサイズのフィルタであり得ることに留意されたい。

【0092】

動作614において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、特徴データ構造に少なくとも部分的に基づいて方向データを決定することを含み得る。MLアーキテクチャのセンターポータリング部分は、特徴データ構造に少なくとも部分的に基づいて方向データを生成し得る。

【0093】

動作616において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、特徴データ構造に少なくとも部分的に基づいて深度データを決定することを含み得る。いくつかの例では、MLアーキテクチャの深度部分は、1×1畳み込みを特徴データ構造に適用して、そのピクセルにおける深度がそのロジットについての対応する深度ピンに入る尤度に対応するピクセルごとのK個のソフトマックスロジットを生成し得る。1×1畳み込みは、本明細書で説明される技術に従ってトレーニングされる学習されたパラメータを含み得、1×1畳み込みは、代替的に、別のサイズのフィルタであり得ることに留意されたい。深度部分は、追加または代替の1×1畳み込みを特徴データ構造に適用して、ピクセルごとの残差を予測し得る。1×1畳み込みは、本明細書で説明される技術に従ってトレーニングされる学習されたパラメータを含み得、1×1畳み込みは、代替的に、別のサイズのフィルタであり得ることに留意されたい。深度は、ログ空間推定のための上記の式(4)に従って予測され得る。いくつかの例では、最大尤度に関連付けられた深度ピンは、ピクセルとの関連付けのために選択され得、および/またはその深度ピンによって示される深度は、ピクセルを囲む領域内のピクセルによって示される深度に少なくとも部分的に基づいて平滑化され得る。

【0094】

動作618において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、ROI、セマンティックセグメンテーション、方向データ、および/または深度データ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を含み得る。いくつかの例では、ROIは、セマンティックセグメンテーション、方向データ、および/または深度データをトリミングするために使用され得る。実施形態に応じて、インスタンスセグメンテーションは、第1の例では(トリミングされた)セマンティックデータおよび方向データ、第2の例では(トリミングされた)深度データおよび方向データ、ならびに/または第3の例では(トリミングされた)深度データ、セマンティックデータ、および方向データに少なくとも部分的に基づいて決定され得るが、任意の他の組合せが企図される。第3の例によれば、ROIに関連付けられた予測クラスのセマンティックセグメンテーションロジット、方向ロジット、および深度ロジットは、インスタンスマスクを推定するために、1×1畳み込みを使用して連結され得る。1×1畳み込みは、本明細書で説明される技術に従ってトレーニングされる学習されたパラメータを含み得、1×1畳み込みは、代替的に、別のサイズのフィルタであり得ることに留意されたい。

【0095】

動作620において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、3次元ROIを決定することを含み得る。例えば、3次元ROIを決定することは、ROIに関連付けられたセマンティックセグメンテーション、深度データ、方向データ、および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き得る。

【0096】

動作622において、例示的なプロセス600は、本明細書で説明される技術のいずれかに従って、オブジェクト検出に少なくとも部分的に基づいて自律車両を制御することを含み得る。例えば、自律車両は、ROI、セマンティックセグメンテーション、深度データ、インスタンスセグメンテーション、および/または3次元ROIに少なくとも部分的に基づいて、自律車両の動きまたは他の動作を制御するための軌道または他のコマンドを

10

20

30

40

50

決定し得る。

【0097】

図7は、本明細書で説明されるMLアーキテクチャをトレーニングするための例示的なプロセス700のフロー図を示す。いくつかの例では、例示的なプロセス700は、知覚コンポーネント228、MLアーキテクチャ300、および/またはトレーニングコンポーネント238によって実行され得る。

【0098】

動作702において、例示的なプロセス700は、本明細書で説明される技術のいずれかに従って、トレーニングデータを受信することを含み得る。例えば、トレーニングデータは、画像704およびそれに関連付けられたグラントゥルース706を含み得る。いくつかの例では、グラントゥルースは、MLアーキテクチャによって達成されるタスクの各タイプに利用可能でないことがある。例えば、トレーニングデータとして使用するために利用可能な画像は、グラントゥルースインスタンスセグメンテーション、深度データ、方向データ、および/または3次元ROIではなく、グラントゥルースROIおよびグラントゥルース意味分類であらかじめラベル付けされ得る。

10

【0099】

そのような例では、トレーニングデータはバッチを含み得、各バッチは異なるグラントゥルースに関連付けられる。例えば、トレーニングデータの第1のバッチ708(1)は、ROIグラントゥルースデータに関連付けられた画像を含み得、第2のバッチ708(2)は、深度グラントゥルースデータ(例えば、lidarデータ)に関連付けられた画像を含み得、および/または第nのバッチ708(n)は、セマンティックセグメンテーショングラントゥルースデータに関連付けられた画像を含み得る。

20

【0100】

いくつかの例では、トレーニングデータ内に含まれるグラントゥルースは、教師ありグラントゥルースデータ(例えば、人間および/または機械にラベル付けされた)、半教師あり(例えば、データのサブセットのみがラベル付けされた)、および/または教師なし(例えば、ラベルが提供されていない場合)であり得る。いくつかの例では、本明細書で説明されるMLアーキテクチャの深度コンポーネントによって生成される深度データに関連付けられた損失を決定するために、lidarデータがグラントゥルースデータとして使用されるときなど、グラントゥルースデータはまばらであり得る。そのようなデータは、半教師あり学習の例であり得る。これらの技術はこれを矯正し、それぞれのセンサ測定値をMLアーキテクチャによって生成された出力データのグループ(より濃密な)に関連付けることによって、センサ測定値をグラントゥルースデータの有用なソースとする。その全体が本明細書に組み込まれる、2019年11月14日に出願された米国特許出願第16/684,554号、およびその全体が本明細書に組み込まれる、2019年11月14日に出願された米国特許出願第16/684,568号を参照されたい。

30

【0101】

動作710において、例示的なプロセス700は、本明細書で説明される技術のいずれかに従って、トレーニングデータに少なくとも部分的に基づいてMLアーキテクチャのコンポーネントを共同トレーニングすることを含み得る。

40

【0102】

動作712において、例示的なプロセス700は、本明細書で説明される技術のいずれかに従って、MLアーキテクチャを1つまたは複数の自律車両に送信することを含み得る。

【0103】

MLアーキテクチャのコンポーネントを共同トレーニングすること(動作710)は、本明細書で説明されるサブ動作をさらに含み得る。コンポーネントを共同トレーニングすることは、異なるコンポーネントのパラメータがジョイント損失を最小化するように変更されるように、コンポーネントの各々の出力に基づいているジョイント損失を決定することと、ジョイント損失を、MLアーキテクチャ全体を通してバックプロパゲートすることとを含み得る。追加的または代替的に、共同トレーニングは、ジョイント損失を構成する

50

損失の間の一貫性を強制することを含み得る。

【0104】

動作708において、MLアーキテクチャを共同トレーニングすることは、トレーニングデータに少なくとも部分的に基づいてMLアーキテクチャから出力を受信することを含み得る。MLアーキテクチャから出力を受信することは、MLアーキテクチャへの入力として画像を提供することに少なくとも部分的に基づき得、受信された出力は、動作604に少なくとも部分的に基づき得る。いくつかの例では、MLアーキテクチャから出力を受信することは、ROI、分類、セマンティックセグメンテーション、方向データ、深度データ、インスタンスセグメンテーション、および/または3次元ROIを受信することを含み得、これらのそれぞれは、MLアーキテクチャの異なる部分のそれぞれの出力と称され得る。そのような出力は、トレーニングデータの画像704のそれぞれについて受信され得る。例えば、MLアーキテクチャに画像704を提供することに応答してMLアーキテクチャから受信した出力714は、バッチに関連付けられた次元および/または他の部分を含む高次元データ構造であり得（例えば、部分716は、バッチ708(n)に関連付けられ得る）、特定のコンポーネントの出力は、そのデータ構造の別の部分に関連付けられ得る（例えば、部分718は、全てのバッチにわたるセマンティックセグメンテーションタスクに関連付けられた出力714の部分であり得る）。

10

【0105】

動作720において、例示的な動作710は、特定のタスクのために利用可能なグラントゥールスに対応する出力710のサブセットを決定することを含み得る。例えば、動作720は、セマンティックセグメンテーション損失を生成するのに適格な出力714のサブセット722を決定することを含み得る。例えば、これは、セマンティックセグメンテーショングラントゥールスが利用可能であった画像に少なくとも部分的に基づいて生成された出力714のサブセット（すなわち、図示された例では、部分716に対応するバッチ708(n)）を決定することと、セマンティックセグメンテーション（すなわち、部分718）を示す出力の次元を決定することとを含み得る。いくつかの例では、動作720は、画像ごとにすべてのタスクタイプについてグラントゥールスが利用可能でない場合に使用され得る。言い換えると、各画像は、出力を生成するMLアーキテクチャの各コンポーネントに関連付けられたグラントゥールスデータに関連付けられていない。

20

【0106】

とはいえ、動作724において、例示的な動作710は、異なるタスクに関連付けられた損失のセットを決定することを含み得る。いくつかの例では、損失を決定することは、タスク固有の損失を決定することと、損失の1つまたは複数にわたって一貫性を強制することとを含み得る。次いで、タスク固有の損失はジョイント損失に合計され得、これはMLアーキテクチャを通じて逆伝播する可能性がある。

30

【0107】

例えば、タスク固有の損失を決定することは、2次元および/または3次元ROIに関して決定され得る、分類損失、 \mathcal{L}_c 、および/または回帰損失、 \mathcal{L}_b （例えば、アンカーに関して）、各ピクセルをセマンティッククラスの1つ（セマンティックセグメンテーションデータ用）に分類するためのピクセル単位の損失、 \mathcal{L}_s 、分類損失（方向ロジットがビンニングされている場合）および/または方向ロジットの回帰損失、 \mathcal{L}_d 、深度ピンの分類損失、 \mathcal{L}_m および深度残差の回帰損失、 \mathcal{L}_r 、および/またはバイナリクロスエントロピー損失

40

であり得る、インスタンスセグメンテーション損失の後に適用され得るインスタンスマスク損失、 \mathcal{L}_i を決定することを含み得る。いくつかの例では、これらの損失の各々は、ジョイント損失を形成するように合計され得る。

【0108】

50

いくつかの例では、一貫性は、ジョイント損失が1つのタスクのトレーニング

【 0 1 0 9 】

【 数 5 】

$$\mathcal{L}_{total} = \sum_t^T \lambda_t \mathcal{L}_t \quad (5)$$

【 0 1 1 0 】

によって歪められないように、重みを使用してパラメータにわたって強制され得、ここで、 λ_t はt番目のタスクに関連付けられた重みであり、 \mathcal{L}_t はt番目のタスクについて決定された損失である。

10

上述した例では、 $t \in \{c, b, s, d, m, r, i\}$ である。

【 0 1 1 1 】

追加または代替の例では、MLアーキテクチャのコンポーネントは、追加パラメータ、各タスク $t \in [1, \dots, T]$ についての不確実性の尺度 (σ_t) をそれぞれ出力し得る。そのような例では、動作726において、例示的な動作710は、損失のセットの損失を重み付けすることを含み得る。いくつかの例では、安定のために、不確実性は、 $s_t = \log \sigma_t^2$ としてログ空間において定義され得る。トレーニングのための修正された(重み付けされた)総損失関数は、

20

【 0 1 1 2 】

【 数 6 】

$$\mathcal{L}_{total}^h = \sum_t \tau_t \exp(-s_t) \mathcal{L}_t + \frac{s_t}{2} \quad (6)$$

【 0 1 1 3 】

によって与えられ得、ここで、

30

【 0 1 1 4 】

【 数 7 】

$$\tau_t = \begin{cases} 1 & \text{if } t \text{ is a classification task} \\ 0.5 & \text{if } t \text{ is a regression task} \end{cases} \quad (7)$$

【 0 1 1 5 】

である。

【 0 1 1 6 】

追加的または代替的に、一貫性損失が損失に追加され得る。一貫性を強制することは、第1の出力と第2の出力との間の差を決定することと、差に少なくとも部分的に基づいて損失を決定することを含み得る。例えば、セマンティックセグメンテーションおよび深度データ、2次元ROIおよび3次元ROI、セマンティックセグメンテーションおよび分類、深度データおよび3次元ROI、および/または本明細書で説明される出力の他の組み合わせの間の差が決定され得る。追加または代替として、一貫性を強制することは、信頼度を類似させることを含み得る。例えば、ROIコンポーネントは、2次元ROIおよびそれに関連付けられた信頼度を出力し得、セマンティックセグメンテーションコンポーネントは、同じ分類に関連付けられた画像のピクセルの集合と、各ピクセルに関連付けられたそれぞれの信頼度とを示すセマンティックセグメンテーションを出力し得る。本技術は、セマンティックセグメンテーションに関連付けられた平均信頼度または代表信頼度(例えば、セマンティックセグメンテーションに関連付けられた信頼度にわたって合計工

40

50

リアテーブルを使用して決定された近似的平均)を決定することと、セマンティックセグメンテーションに関連付けられた平均および/または代表信頼度と、2次元ROIに関連付けられた信頼度との間の差に少なくとも部分的に基づいて一貫性損失を決定することとを含み得る。当然、任意の数の一貫性損失を使用し得る。

【0117】

動作728において、例示的な動作710は、MLアーキテクチャのコンポーネントを修正して、動作722および/または724で決定されたジョイント損失を最小化することを含み得る。ジョイント損失は、MLアーキテクチャ300を通して逆伝播され得、これは、本明細書で説明される各コンポーネントのゼロまたは複数のパラメータを調整し、ジョイント損失を低減することを含み得る。

10

【0118】

[例示的な発明内容]

A. 画像データを受信することと、前記画像データの少なくとも一部を機械学習(ML)モデルに入力することと、前記MLモデルによって、前記画像に示されるオブジェクトに関連付けられた関心領域(ROI)を決定することと、前記MLモデルによって、および前記ROIに少なくとも部分的に基づいて、追加出力を決定することとであって、前記追加出力は、

前記オブジェクトに関連付けられたセマンティックセグメンテーションであって、前記セマンティックは前記オブジェクトの分類を示す、セマンティックセグメンテーションと、前記オブジェクトの中心を示す方向データと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記方向データ、前記深度データ、または前記インスタンスセグメンテーションの2つまたは複数に少なくとも部分的に基づいて一貫性損失を決定することと、トレーニングされたMLモデルとして、および前記一貫性損失に少なくとも部分的に基づいて、前記MLモデルの1つまたは複数のパラメータを変更することと、トレーニングされたMLモデルを自律車両に送信すること、を含む方法。

20

【0119】

B. 前記ROIを決定することが、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することと、に少なくとも部分的に基づいており、前記追加出力を決定することはさらに、前記第1の特徴のセットおよび前記第2の特徴のセットに少なくとも部分的に基づいている、段落Aに記載の方法。

30

【0120】

C. 前記MLモデルによって、および、前記ROI、前記セマンティックセグメンテーション、前記方向データ、前記深度データ、または前記インスタンスセグメンテーションの2つまたは複数に少なくとも部分的に基づいて、前記オブジェクトに関連付けられた3次元ROIを決定すること、をさらに含む、段落Aまたは段落Bのいずれかに記載の方法。

【0121】

D. 前記一貫性損失を決定することは、前記セマンティックセグメンテーション、深度データ、インスタンスセグメンテーション、または前記3次元ROIの少なくとも1つに少なくとも部分的に基づいて、2次元境界領域を決定することと、前記ROIと前記2次元境界領域との間の差を決定することと、を含む、段落A乃至Cのいずれか1つに記載の方法。

40

【0122】

E. 前記深度データは、離散的な深度を示す深度ピン出力と、前記深度ピンからのオフセットを示す深度残差とを含む、段落A乃至Dのいずれか1つに記載の方法。

【0123】

F. 1つまたは複数のプロセッサと、コンピュータ実行可能命令を格納したメモリと、を含むシステムであって、前記コンピュータ実行可能命令は前記一つまたは複数のプロセ

50

ッサによって実行されると、前記システムに、画像データを受信することと、前記画像データの少なくとも一部を機械学習（ML）モデルに入力することと、前記MLモデルによって、前記画像に示されるオブジェクトに関連付けられた関心領域（ROI）を決定することと、前記MLモデルによって、および前記ROIに少なくとも部分的に基づいて、追加出力を決定することと、前記追加出力は、前記オブジェクトに関連付けられたセマンティックセグメンテーションであって、前記セマンティックが前記オブジェクトの分類を示す、セマンティックセグメンテーションと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記深度データ、または前記インスタンスセグメンテーションの2つまたは複数に少なくとも部分的に基づいて一貫性損失を決定することと、トレーニングされたMLモデルとして、および前記一貫性損失に少なくとも部分的に基づいて、前記MLモデルの1つまたは複数のパラメータを変更することと、を含む動作を実行させる、システム。

10

【0124】

G．前記ROIを決定することが、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することと、に少なくとも部分的に基づいており、前記追加出力を決定することはさらに、前記第1の特徴のセットおよび前記第2の特徴のセットに少なくとも部分的に基づいている、段落Fに記載のシステム。

20

【0125】

H．前記動作が、前記オブジェクトの中心を示す方向データを決定することをさらに含み、前記インスタンスセグメンテーションを決定することが、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づいている、段落FまたはGのいずれかに記載のシステム。

【0126】

I．前記動作が、前記オブジェクトの中心を示す方向データを決定することと、前記セマンティックセグメンテーション、前記深度データ、前記方向データ、および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、3次元ROIを決定することと、をさらに含み、段落F乃至Hのいずれか1つに記載のシステム。

【0127】

J．前記一貫性損失を前記決定することは、前記深度データと前記3次元ROIの境界との間の差を決定することを含む、段落F乃至Iのいずれか1つに記載のシステム。

30

【0128】

K．前記一貫性損失を前記決定することは、前記セマンティックセグメンテーション、深度データまたはインスタンスセグメンテーションの1つまたは複数に少なくとも部分的に基づいて、2次元境界領域を決定することと、前記ROIと前記2次元境界領域との間の差を決定することと、を含む、段落F乃至Jのいずれか1つに記載のシステム。

【0129】

L．前記動作が、前記セマンティックセグメンテーション、前記深度データ、および前記インスタンスセグメンテーションの少なくとも1つに関連付けられた確実性を決定することをさらに含み、前記一貫性損失がさらに、前記不確実性に少なくとも部分的に基づいている、段落F乃至Kのいずれか1つに記載のシステム。

40

【0130】

M．コンピュータ実行可能命令を格納した非一時的コンピュータ可読媒体であって、前記コンピュータ実行可能命令は、1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、画像データを受信することと、前記画像データの少なくとも一部を機械学習（ML）モデルに入力することと、前記MLモデルによって、前記画像に示されるオブジェクトに関連付けられた関心領域（ROI）を決定することと、前記MLモデルによって、および前記ROIに少なくとも部分的に基づいて、追加出力を決定することと、前記追加出力は、前記オブジェクトに関連付けられたセマンティック

50

クセグメンテーションであって、前記セマンティックが前記オブジェクトの分類を示す、セマンティックセグメンテーションと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記深度データ、または前記インスタンスセグメンテーションの2つまたは複数に少なくとも部分的に基づいて一貫性損失を決定することと、トレーニングされたMLモデルとして、および前記一貫性損失に少なくとも部分的に基づいて、前記MLモデルの1つまたは複数のパラメータを変更することと、を含む動作を実行させる、非一時的コンピュータ可読媒体。

【0131】

N．前記ROIを決定することが、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することと、に少なくとも部分的に基づいており、前記追加出力を決定することはさらに、前記第1の特徴のセットおよび前記第2の特徴のセットに少なくとも部分的に基づいている、段落Mに記載の非一時的コンピュータ可読媒体。

10

【0132】

O．前記動作が、前記オブジェクトの中心を示す方向データを決定することをさらに含み、前記インスタンスセグメンテーションを決定することが、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づいている、段落MまたはNのいずれかに記載の非一時的コンピュータ可読媒体。

【0133】

P．前記動作が、前記オブジェクトの中心を示す方向データを決定することと、前記セマンティックセグメンテーション、前記深度データ、前記方向データ、および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、3次元ROIを決定することと、さらに含み、段落M乃至Oのいずれか1つに記載の非一時的コンピュータ可読媒体。

20

【0134】

Q．前記一貫性損失を前記決定することは、前記深度データと前記3次元ROIの境界との間の差を決定することを含む、段落M乃至Pのいずれか1つに記載の非一時的コンピュータ可読媒体。

【0135】

R．一貫性損失を前記決定することは、前記セマンティックセグメンテーション、深度データまたはインスタンスセグメンテーションの1つまたは複数に少なくとも部分的に基づいて、2次元境界領域を決定することと、前記ROIと前記2次元境界領域との間の差を決定することと、を含む、段落M乃至Qのいずれか1つに記載の非一時的コンピュータ可読媒体。

30

【0136】

S．前記動作が、前記セマンティックセグメンテーション、前記深度データ、および前記インスタンスセグメンテーションの少なくとも1つに関連付けられた確実性を決定することをさらに含み、前記一貫性損失がさらに、前記不確実性に少なくとも部分的に基づいている、段落M乃至Rのいずれか1つに記載の非一時的コンピュータ可読媒体。

【0137】

T．前記深度データは、離散的な深度を示す深度ピン出力と、前記深度ピンからのオフセットを示す深度残差とを含む、段落M乃至Sのいずれか1つに記載の非一時的なコンピュータ可読媒体。

40

【0138】

U．1つまたは複数のプロセッサと、プロセッサ実行可能命令を格納するメモリと、を含むシステムであって、前記プロセッサ実行可能命令は前記一つまたは複数のプロセッサによって実行されると、前記システムに、自律車両に関連付けられた画像センサから画像を受信することと、前記画像の少なくとも一部を機械学習(ML)モデルに入力することと、前記MLモデルによって、出力のセットを決定することであって、前記出力のセットは、前記画像に示されるオブジェクトに関連付けられた関心領域(ROI)と、前記オブ

50

ジェクトに関連付けられたセマンティックセグメンテーションであって、前記オブジェクトの分類を示す、前記セマンティックセグメンテーションと、前記オブジェクトの中心を示す方向データと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記インスタンスセグメンテーション、または前記深度データの少なくとも1つに少なくとも部分的に基づいて前記自律車両を制御することと、を含む動作を実行させる、システム。

【0139】

V. 前記出力のセットを決定することは、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することと、前記第1の解像度は前記第2の解像度とは異なる、ことと、アップサンプリングされた特徴として、第1の解像度と同じ解像度を有する前記第2の特徴をアップサンプリングすることと、組み合わせられた特徴として、前記アップサンプリングされた特徴を前記第1の特徴と組み合わせることと、前記セマンティックセグメンテーション、深度データ、方向データ、またはインスタンスセグメンテーションの少なくとも1つが、前記組み合わせられた特徴に少なくとも部分的に基づいている、ことと、を含む、段落Uに記載のシステム。

10

【0140】

W. 前記出力のセットが、3次元ROIをさらに含む、段落UまたはVのいずれかに記載のシステム。

20

【0141】

X. 前記深度データを決定することは、深度ピンのセットの中から深度ピンを決定することと、前記深度ピンは、前記環境の離散部分に関連付けられている、ことと、前記深度ピンに関連付けられた深度残差を決定することと、前記深度残差は、前記深度ピンに関連付けられた位置からの前記離散部分に関連付けられた表面の偏差を示す、ことと、を含む、段落U乃至Wのいずれか1つに記載のシステム。

【0142】

Y. 前記深度ピンを決定することが、前記離散部分を囲む領域内の他の離散部分のロジットの平均または確率分布を決定することに少なくとも部分的に基づいて、平滑化されたロジットのセットを決定することと、前記深度ピンが、前記平滑化されたロジットのセットの中の最大平滑化されたロジット値に関連付けられていることを決定することに少なくとも部分的に基づいて、前記深度ピンのセットの中から前記深度ピンを選択することと、を含む、段落U乃至Xのいずれか1つに記載のシステム。

30

【0143】

Z. 自律車両に関連付けられた画像センサから画像を受信することと、前記画像の少なくとも一部を機械学習(ML)モデルに入力することと、前記MLモデルによって、出力のセットを決定することと、前記出力のセットは、前記オブジェクトに関連付けられたセマンティックセグメンテーションと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記インスタンスセグメンテーション、または前記深度データの少なくとも1つに少なくとも部分的に基づいて、前記自律車両を制御することと、を含む方法。

40

【0144】

AA. 前記出力のセットを決定することは、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することと、前記第1の解像度は前記第2の解像度とは異なる、ことと、アップサンプリングされた特徴として、第1の解像度と同じ解像度を有する前記第2の特徴をアップサンプリングすることと、組み合わせられた特徴として、前記アップサンプリングされた特徴を前記第1の特徴と組み合わせることと、前記セマンティックセグメンテーション

50

ン、深度データ、またはインスタンスセグメンテーションの少なくとも1つが、前記組み合わせられた特徴に少なくとも部分的に基づいている、ことと、を含む、段落Zに記載の方法。

【0145】

AB．前記出力のセットが、3次元ROIをさらに含む、段落ZまたはAAのいずれかに記載の方法。

【0146】

AC．前記出力のセットは、前記オブジェクトの中心を示す方向データをさらに含み、前記3次元セグメンテーションを決定することは、前記セマンティックセグメンテーション、前記深度データ、前記方向データ、および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いる、Z乃至ABのいずれか1つに記載の方法。

10

【0147】

AD．前記深度データを決定することは、深度ピンのセットの中から深度ピンを決定することであって、前記深度ピンは、前記環境の離散部分に関連付けられている、ことと、前記深度ピンに関連付けられた深度残差を決定することであって、前記深度残差は、前記深度ピンに関連付けられた位置からの前記離散部分に関連付けられた表面の偏差を示す、ことと、を含む、段落Z乃至ACにいずれか1つに記載の方法。

【0148】

AE．前記深度ピンを決定することは、前記離散部分を囲む領域内の他の離散部分のロジットの平均または確率分布を決定することに少なくとも部分的に基づいて、平滑化されたロジットのセットを決定することと、前記深度ピンが、前記平滑化されたロジットのセットの中の最大平滑化されたロジット値に関連付けられていることを決定することに少なくとも部分的に基づいて、前記深度ピンのセットの中から前記深度ピンを選択することと、を含む、段落Z乃至ADにいずれか1つに記載の方法。

20

【0149】

AF．前記出力のセットは、前記オブジェクトの中心を示す方向データをさらに含み、前記インスタンスセグメンテーションを決定することは、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づき、Z乃至AEのいずれか1つに記載の方法。

【0150】

AG．コンピュータ実行可能命令を格納した非一時的コンピュータ可読媒体であって、前記コンピュータ実行可能命令は、1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、自律車両に関連付けられた画像センサから画像を受信することと、前記画像の少なくとも一部を機械学習(ML)モデルに入力することと、前記MLモデルによって、出力のセットを決定することであって、前記出力のセットは、前記オブジェクトに関連付けられたセマンティックセグメンテーションと、前記画像の少なくとも前記一部に関連付けられた深度データと、前記オブジェクトに関連付けられたインスタンスセグメンテーションと、を含む、ことと、前記ROI、前記セマンティックセグメンテーション、前記インスタンスセグメンテーション、または前記深度データの少なくとも1つに少なくとも部分的に基づいて、前記自律車両を制御することと、を含む動作を実行させる、非一時的コンピュータ可読媒体。

30

40

【0151】

AH．前記出力のセットを決定することは、第1の解像度に関連付けられた第1の特徴のセットを決定することと、第2の解像度に関連付けられた第2の特徴のセットを決定することであって、前記第1の解像度は前記第2の解像度とは異なる、ことと、アップサンプリングされた特徴として、第1の解像度と同じ解像度を有する前記第2の特徴をアップサンプリングすることと、組み合わせられた特徴として、前記アップサンプリングされた特徴を前記第1の特徴と組み合わせることであって、前記セマンティックセグメンテーション、深度データ、またはインスタンスセグメンテーションの少なくとも1つが、前記組み合わせられた特徴に少なくとも部分的に基づいている、ことと、を含む、段落AGに記載の

50

非一時的コンピュータ可読媒体。

【0152】

A I . 前記出力のセットを決定することは、ダウンサンプリングされた特徴として、前記組み合わせられた特徴をダウンサンプリングして、前記組み合わせられた特徴に関連付けられたチャンネルの数を減少させることと、畳み込まれた特徴として、異なる拡張速度に従ってダウンサンプリングされた特徴を2回以上畳み込むことと、特徴データ構造として、前記畳み込み特徴をアップサンプリングすることであって、前記セマンティックセグメンテーション、深度データ、またはインスタンスセグメンテーションの少なくとも1つが、前記特徴データ構造に少なくとも部分的に基づいている、ことと、をさらに含む、段落A G またはA H に記載の非一時的コンピュータ可読媒体。

10

【0153】

A J . 前記出力のセットは、3次元ROIをさらに含む、段落A G 乃至A I のいずれか1つに記載の非一時的コンピュータ可読媒体。

【0154】

A K . 前記出力のセットは、前記オブジェクトの中心を示す方向データをさらに含み、前記3次元を決定することは、前記セマンティックセグメンテーション、前記深度データ、前記方向データ、および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いる、段落A J に記載の非一時的コンピュータ可読媒体。

【0155】

A L . 前記深度データを決定することは、深度ピンのセットの中から深度ピンを決定することであって、前記深度ピンは、前記環境の離散部分に関連付けられている、ことと、前記深度ピンに関連付けられた深度残差を決定することであって、前記深度残差は、前記深度ピンに関連付けられた位置からの前記離散部分に関連付けられた表面の偏差を示す、ことと、段落A G 乃至A K のいずれか1つに記載の非一時的コンピュータ可読媒体。

20

【0156】

A M . 前記深度ピンを決定することは、前記離散部分を囲む領域内の他の離散部分のロジットの平均または確率分布を決定することに少なくとも部分的に基づいて、平滑化されたロジットのセットを決定することと、前記深度ピンが、前記平滑化されたロジットのセットの中の最大平滑化されたロジット値に関連付けられていることを決定することに少なくとも部分的に基づいて、前記深度ピンのセットの中から前記深度ピンを選択することと、段落A G 乃至A L のいずれか1つに記載の非一時的コンピュータ可読媒体。

30

【0157】

A N . 前記出力のセットは、前記オブジェクトの中心を示す方向データをさらに含み、前記インスタンスセグメンテーションを決定することは、前記セマンティックセグメンテーション、前記深度データ、および前記方向データに少なくとも部分的に基づく、段落A G 乃至A M のいずれか1つに記載の非一時的コンピュータ可読媒体。

【0158】

A O . 1つまたは複数のプロセッサと、プロセッサ実行可能命令を格納したメモリと、を含むシステムであって、前記プロセッサ実行可能命令は前記一つまたは複数のプロセッサによって実行されると、前記システムに、請求項A乃至FまたはZ乃至AFのいずれか1つに記載の動作のいずれかを含む動作を実行させる、システム。

40

【0159】

A P . 1つまたは複数のプロセッサと、プロセッサ実行可能命令を格納するメモリと、を含む自律車両であって、前記プロセッサ実行可能命令は前記一つまたは複数のプロセッサによって実行されると、前記システムに、請求項A乃至FまたはZ乃至AFのいずれか1つに記載の動作のいずれかを含む動作を実行させる、自律車両。

【0160】

A P . 1つまたは複数のセンサをさらに備える、段落A P に記載の自律車両。

【0161】

A Q . 一つまたは複数のプロセッサによって実行されると、前記一つまたは複数のプロ

50

セッサに、請求項 A 乃至 F または Z 乃至 A F のいずれか 1 つに記載の動作のいずれかを含む動作を実行させるプロセッサ実行可能命令を格納する、非一時的コンピュータ可読媒体。

【0162】

[結論]

主題は、構造的特徴および/または方法論的行為に特有の言語で説明されてきたが、添付の特許請求の範囲において定義される主題は、説明される特定の特征または行為に必ずしも限定されないことを理解されたい。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形態として開示される。

【0163】

本明細書で説明されるコンポーネントは、任意のタイプのコンピュータ可読媒体に格納され得る、およびソフトウェアおよび/またはハードウェアで実装され得る命令を表す。上記で説明した方法およびプロセスのすべては、1つまたは複数のコンピュータもしくはプロセッサ、ハードウェア、またはそれらのいくつかの組合せによって実行されるソフトウェアコードコンポーネントおよび/またはコンピュータ実行可能命令において具現化され、それらを介して完全に自動化され得る。代替として、方法の一部または全部は、専用コンピュータハードウェアにおいて具現化され得る。

10

【0164】

本明細書で説明されるプロセスの少なくともいくつかは論理フローグラフとして示され、その各動作は、ハードウェア、ソフトウェア、またはそれらの組合せで実装され得る動作のシーケンスを表す。ソフトウェアのコンテキストにおいて、動作は、1つまたは複数の非一時的コンピュータ可読記憶媒体に格納されたコンピュータ実行可能命令を表し、コンピュータ実行可能命令は、1つまたは複数のプロセッサによって実行されると、コンピュータまたは自律車両に、記載された動作を実行させる。一般に、コンピュータ実行可能命令は、特定の機能を実行するまたは特定の抽象データタイプ実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。動作が説明される順序は、限定として解釈されることを意図されず、任意の数の説明される動作が、プロセスを実装するために、任意の順序でおよび/または並列に組み合わせることができる。

20

【0165】

特に「し得る(may)」、「できる(could)」、「し得る(may)」または「し得る(might)」などの条件付き言語は、特に明記しない限り、コンテキスト内で、特定の例が特定の特征、要素および/またはステップを含み、他の例が含まないことを示すと理解される。したがって、そのような条件付き言語は、一般に、特定の特征、要素、および/またはステップが1つまたは複数の例に任意の手段で必要とされること、または1つまたは複数の例が、ユーザ入力またはプロンプトの有無にかかわらず、特定の特征、要素、および/またはステップが任意の特定の例に含まれるまたは実行されるべきかどうかを決定するための論理を必ず含むことを暗示することを意図するものではない。

30

【0166】

「X、Y、またはZの少なくとも1つ」という句などの接続的な言語は、別段に具体的に述べられない限り、項目、用語などが、X、Y、もしくはZのいずれか、または複数の各要素を含むそれらの任意の組合せであり得ることを提示すると理解されるべきである。単数形として明示的に記載されていない限り、「a」は単数形および複数形を意味する。

40

【0167】

本明細書に記載され、および/または添付の図面に示されるフロー図における任意のルーチンの説明、要素、またはブロックは、ルーチンにおける特定の論理機能または要素を実装するための1つまたは複数のコンピュータ実行可能命令を含むコードのモジュール、セグメント、または部分を潜在的に表すものとして理解されるべきである。代替実施形態は、本明細書で説明される例の範囲内に含まれ、要素または機能は、当業者によって理解されるように、関与する機能に応じて、実質的に同期して、逆の順序で、追加の動作とともに、または動作を省略することを含めて、図示または説明する順序から削除または実行され得る。

50

【0168】

上述の例に対して多くの变形および修正を行い得、その要素は、他の許容可能な例の中にあるものとして理解されるべきである。そのようなすべての修正および变形は、本開示の範囲内で本明細書に含まれ、以下の特許請求の範囲によって保護されることが意図されている。

【図面】

【図1】

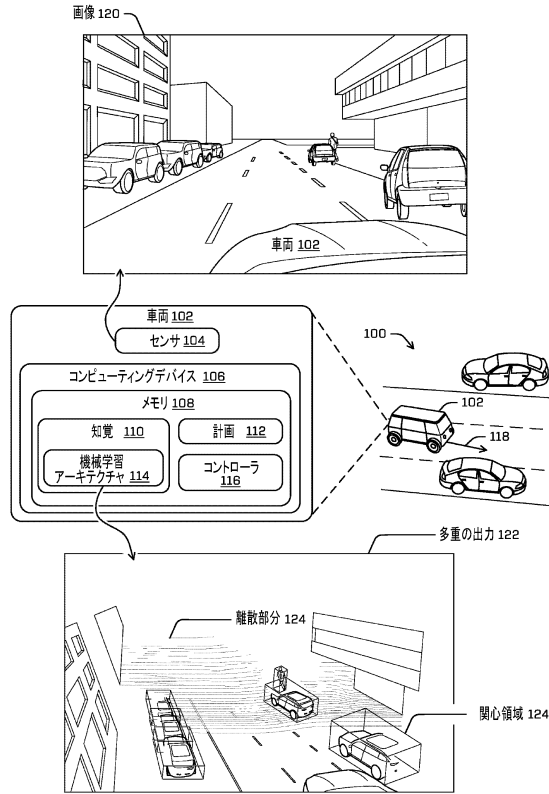


FIG. 1

【図2】

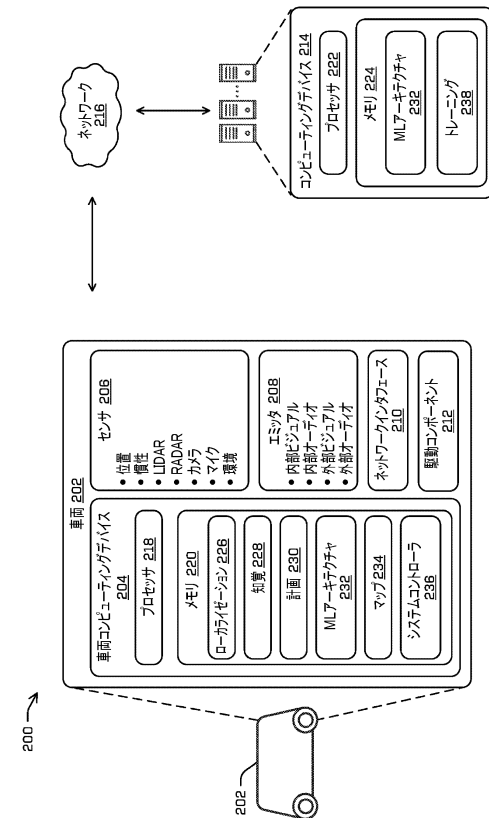


FIG. 2

【図3A】

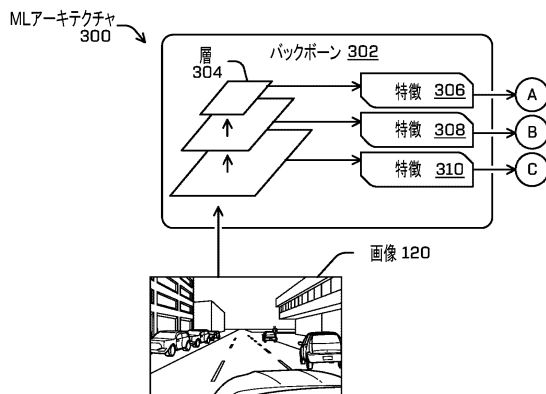


FIG. 3A

【図3B】

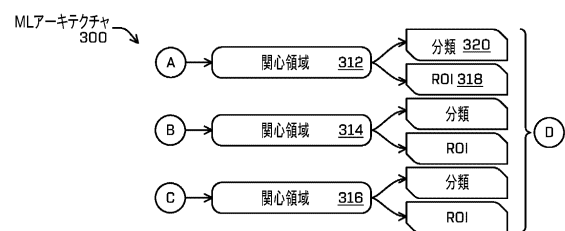


FIG. 3B

10

20

30

40

50

【 図 3 C 】

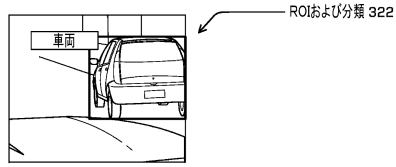


FIG. 3C

【 図 4 A 】

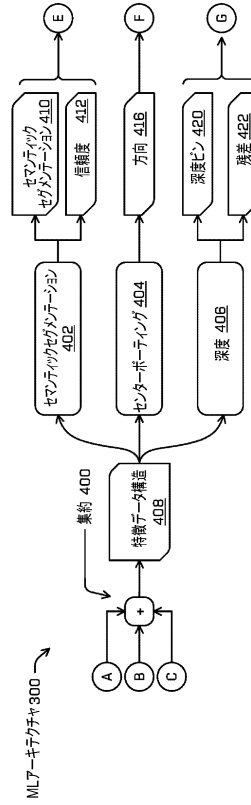


FIG. 4A

【 図 4 B 】

セマンティックセグメンテーション 414

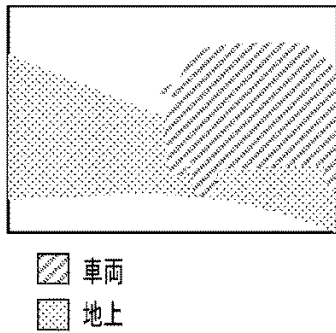


FIG. 4B

【 図 4 C 】

方向データ 418

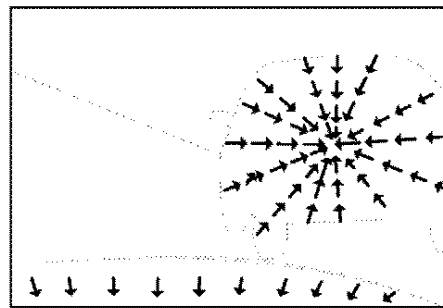


FIG. 4C

10

20

30

40

50

【図 4 D】

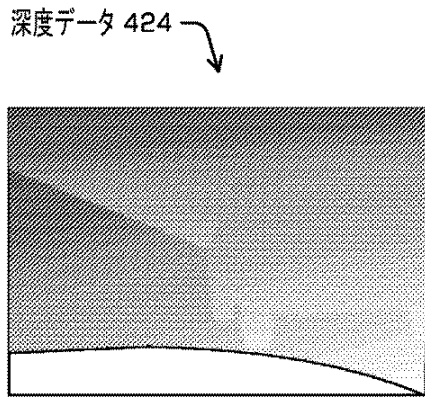


FIG. 4D

【図 5 A】

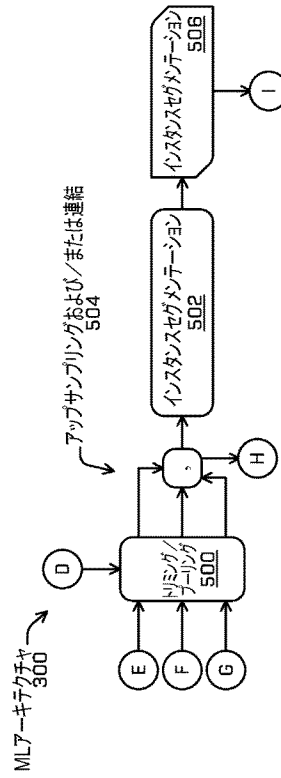


FIG. 5A

【図 5 B】

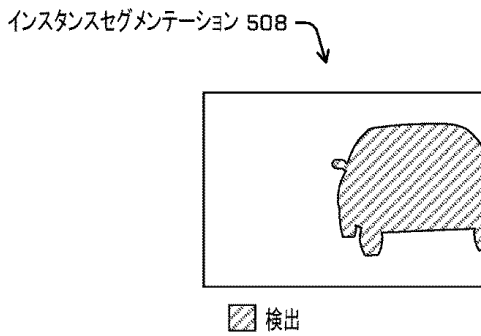


FIG. 5B

【図 5 C】

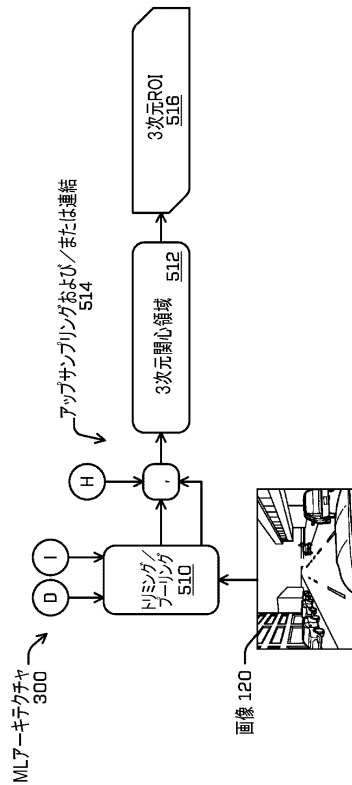


FIG. 5C

10

20

30

40

50

【 図 5 D 】

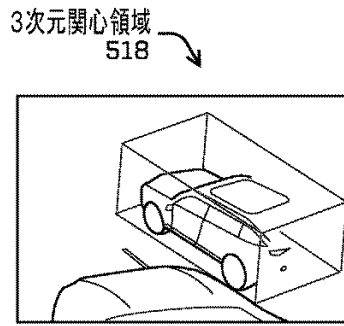
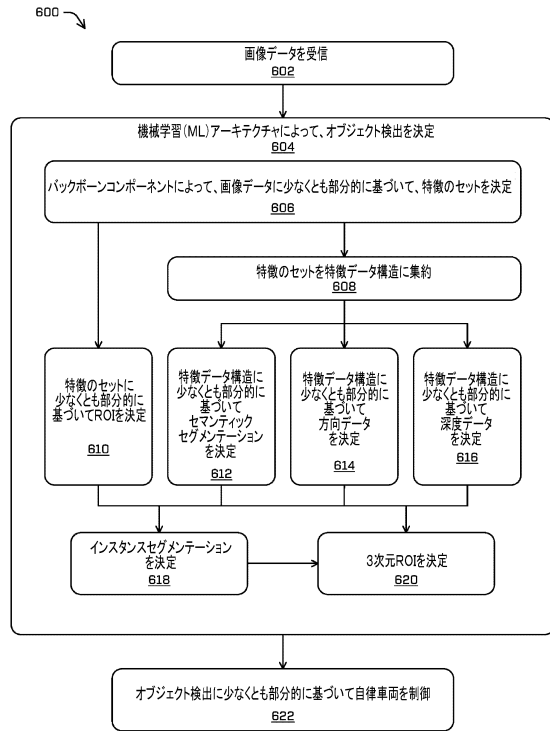


FIG. 5D

【 図 6 】

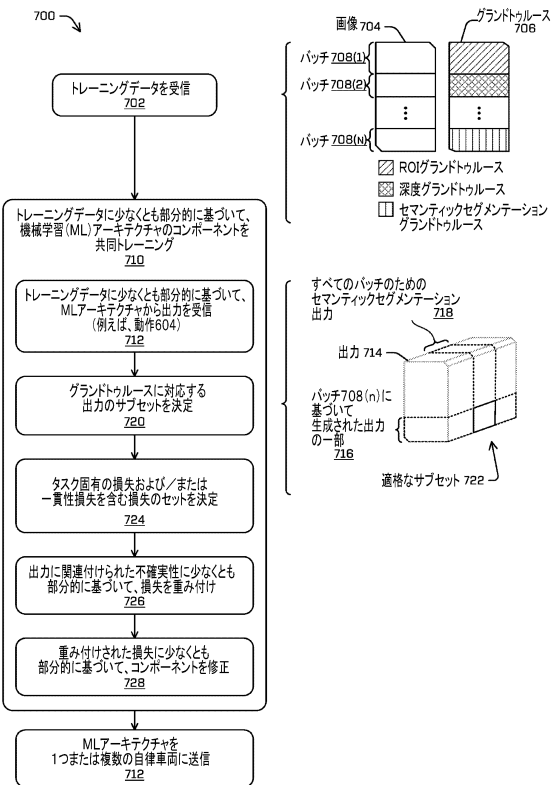


10

20

FIG. 6

【 図 7 】



30

40

FIG. 7

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)

ライブ 1149 ズークス インコーポレイテッド内

(72)発明者 サラ タリク

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

(72)発明者 ジェームズ ウィリアム ベイジー フィルビン

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 佐田 宏史

(56)参考文献

米国特許出願公開第2019/0147245 (US, A1)

米国特許出願公開第2019/0057507 (US, A1)

米国特許出願公開第2019/0130275 (US, A1)

中国特許出願公開第109191515 (CN, A)

Alex Kendall et al., "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 米国, IEEE, 2018年06月18日, pp.7482-7491

Danfei Xu et al., "PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation", arXiv, 米国, CORNELL UNIVERSITY, 2018年08月25日, pp.1-11, <https://arxiv.org/abs/1711.10871v2>

嶋田 彩人、外6名, "生体蛍光観察動画の深度を考慮した深層学習による細胞追跡精度の改善", 情報処理学会論文誌 数理モデル化と応用 (TOM), 日本, 情報処理学会, 2019年07月17日, Vol.12, No.2, pp.82-91

Liqin Huang et al., "Measuring the absolute distance of a front vehicle from an in-car camera based on monocular vision and instance segmentation", Journal of Electronic Imaging, 米国, SPIE and IS&T, 2018年07月24日, Vol.27, No.4, pp.1-10

Yingting Li et al., "Certainty-Driven Consistency Loss for Semi-supervised Learning", arXiv, 米国, CORNELL UNIVERSITY, 2019年01月17日, pp.1-10, <https://arxiv.org/abs/1901.05657v1>Jean Lahoud et al., "3D Instance Segmentation via Multi-Task Metric Learning", arXiv, 米国, CORNELL UNIVERSITY, 2019年11月01日, pp.1-11, <https://arxiv.org/abs/1906.08650>

(58)調査した分野 (Int.Cl., D B 名)

G 0 6 T 1 / 0 0 , 7 / 0 0 - 7 / 9 0

G 0 6 V 1 0 / 0 0 - 1 0 / 9 8