

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7367183号  
(P7367183)

(45)発行日 令和5年10月23日(2023.10.23)

(24)登録日 令和5年10月13日(2023.10.13)

(51)国際特許分類 F I  
 G 0 8 G 1/16 (2006.01) G 0 8 G 1/16 C  
 G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 6 5 0 A

請求項の数 19 (全21頁)

(21)出願番号	特願2022-510922(P2022-510922)	(73)特許権者	317015065 ウェイモ エルエルシー アメリカ合衆国 カリフォルニア州 9 4 0 4 3 マウンテン ビュー アンフィシ アター パークウェイ 1 6 0 0
(86)(22)出願日	令和2年7月21日(2020.7.21)	(74)代理人	100079108 弁理士 稲葉 良幸
(65)公表番号	特表2022-546283(P2022-546283 A)	(74)代理人	100126480 弁理士 佐藤 睦
(43)公表日	令和4年11月4日(2022.11.4)	(72)発明者	マーヴェット, ハイマン アメリカ合衆国, カリフォルニア州 9 4 0 4 3, マウンテン ビュー, アンフ イシアター パークウェイ 1 6 0 0
(86)国際出願番号	PCT/US2020/042841	(72)発明者	オガール, アビジット アメリカ合衆国, カリフォルニア州 9
(87)国際公開番号	WO2021/040910		最終頁に続く
(87)国際公開日	令和3年3月4日(2021.3.4)		
審査請求日	令和4年4月14日(2022.4.14)		
(31)優先権主張番号	16/557,246		
(32)優先日	令和1年8月30日(2019.8.30)		
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 占有予測ニューラルネットワーク

(57)【特許請求の範囲】

【請求項 1】

1つ以上のデータ処理装置によって実施される方法であって、  
 現時点での車両の近くの環境を特徴付ける、前記車両のセンサシステムによって生成されたセンサデータを受信することであって、前記センサデータが、それぞれ異なる時点でキャプチャされた環境を特徴付ける複数のセンササンプルを含むことと、

ニューラルネットワークを使用して前記センサデータを含むネットワーク入力を処理して、前記環境の領域に関する占有予測出力を生成することであって、

前記占有予測出力が、前記現時点の後の1つ以上の未来の時間間隔について、前記環境の前記領域が前記未来の時間間隔中に前記環境内のエージェントによって占有されるそれぞれの可能性を特徴付け、

前記ネットワーク入力が、前記ニューラルネットワークの入力レイヤに提供され、前記環境の前記領域に関する前記占有予測出力が、前記ニューラルネットワークの出力レイヤによって出力されることと、

前記占有予測出力を前記車両の計画システムに提供して、前記車両の未来の軌道を計画する計画決定を生成することと、を含む、方法。

【請求項 2】

前記センササンプルが、前記車両の1つ以上のカメラセンサによって生成された画像を含む、請求項 1 に記載の方法。

【請求項 3】

前記センササンプルが、前記車両の1つ以上のL I D A Rセンサによって生成された点群データ、前記車両の1つ以上のレーダーセンサによって生成されたセンサデータ、またはその両方、を含む、請求項1または2に記載の方法。

【請求項4】

前記現時点の後の前記未来の時間間隔が、前記現時点から始まる時間間隔を含む、請求項1～3のいずれか一項に記載の方法。

【請求項5】

前記現時点の後の複数の所与の未来の時間間隔のそれぞれについて、前記占有予測出力が、前記環境の前記領域が、前記所与の未来の時間間隔中に、前記環境内のエージェントによって占有されるそれぞれの可能性を特徴付ける、請求項1～4のいずれか一項に記載の方法。

10

【請求項6】

前記環境の複数の所与の領域のそれぞれについて、前記占有予測出力が、前記環境の前記所与の領域が、前記現時点の後の未来の各時間間隔中に、前記環境内のエージェントによって占有されるそれぞれの可能性を特徴付ける、請求項1～5のいずれか一項に記載の方法。

【請求項7】

前記環境の前記領域が、車道の交差点である、請求項1～6のいずれか一項に記載の方法。

【請求項8】

前記計画システムによって、前記車道の前記交差点に関する前記占有予測出力に基づいて、前記車両に前記車道の前記交差点を横断させる計画決定を生成すること、をさらに含む、請求項7に記載の方法。

20

【請求項9】

前記環境の前記領域が、静止したエージェントによって占有されている車道の一部分である、請求項1～6のいずれか一項に記載の方法。

【請求項10】

前記計画システムによって、前記静止したエージェントによって占有されている前記車道の前記一部分に関する前記占有予測出力に基づいて、前記車両を減速させる計画決定を生成すること、をさらに含む、請求項9に記載の方法。

30

【請求項11】

ニューラルネットワークが、複数の畳み込みニューラルネットワークレイヤを含む畳み込みニューラルネットワークである、請求項1～10のいずれか一項に記載の方法。

【請求項12】

前記ニューラルネットワークを使用して前記センサデータを処理して、前記環境の前記領域に関する前記占有予測出力を生成することが、

1つ以上のニューラルネットワークレイヤの第1のセットを使用して前記センサデータの第1のサブセットを処理して、第1の中間出力を生成することと、

1つ以上のニューラルネットワークレイヤの第2のセットを使用して前記センサデータの第2のサブセットを処理して、第2の中間出力を生成することと、

1つ以上のニューラルネットワークレイヤの第3のセットを使用して前記第1の中間出力および前記第2の中間出力を処理して、前記占有予測出力を生成することと、を含む、請求項1～11のいずれか一項に記載の方法。

40

【請求項13】

前記センサデータの前記第1のサブセットが、第1のモダリティのセンサデータを含み、前記センサデータの前記第2のサブセットが、第2の異なるモダリティのセンサデータを含む、請求項12に記載の方法。

【請求項14】

前記ニューラルネットワークが、前記車両の前記近くの前記環境内の他のエージェントの現在または未来の位置を明示的に位置特定することなく、前記占有予測出力を生成する

50

、請求項 1 ~ 1 3 のいずれか一項に記載の方法。

【請求項 1 5】

前記ネットワーク入力、前記車両の前記近くの前記環境の静的な特徴を特徴付ける道路グラフデータをさらに含む、請求項 1 ~ 1 4 のいずれか一項に記載の方法。

【請求項 1 6】

前記センサデータが、それぞれの時点でキャプチャされた画像を含み、

前記ニューラルネットワークを使用して前記画像を処理する前に、1つ以上の事前定義されたトリミング操作を各画像に適用すること、をさらに含む、請求項 1 ~ 1 5 のいずれか一項に記載の方法。

【請求項 1 7】

前記センサデータが、それぞれの時点でキャプチャされた画像を含み、

前記ニューラルネットワークが、前記画像を処理して画像トリミング操作を指定するデータを生成するように構成されているトリミングサブネットワークを含み、

前記ニューラルネットワークを使用して前記センサデータを処理することが、前記トリミングサブネットワークによって指定された前記画像トリミング操作を前記画像に適用すること、を含む、請求項 1 ~ 1 5 のいずれか一項に記載の方法。

【請求項 1 8】

システムであって、

1つ以上のコンピュータと、

前記1つ以上のコンピュータに通信可能に結合された1つ以上の記憶デバイスであって、前記1つ以上の記憶デバイスは、命令を記憶しており、前記命令は、前記1つ以上のコンピュータによって実行されたときに、前記1つ以上のコンピュータに請求項 1 ~ 1 7 のいずれか一項に記載のそれぞれの方法の動作を行わせる、1つ以上の記憶デバイスと、を備える、システム。

【請求項 1 9】

1つ以上のコンピュータによって実行されたときに、前記1つ以上のコンピュータに請求項 1 ~ 1 7 のいずれか一項に記載のそれぞれの方法の動作を行わせる命令を記憶している、1つ以上の非一時的コンピュータ記憶媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本明細書は、自律型車両に関する。

【背景技術】

【0 0 0 2】

自律型車両は、自動運転する車、船舶、および航空機を含む。自律型車両は、様々な車載センサを使用して環境を特徴付けるセンサデータを生成し、コンピュータシステムを使用してセンサデータを処理して、制御およびナビゲーションの決定を行う。

【0 0 0 3】

自律型車両は、機械学習モデルを使用して制御およびナビゲーションの決定を行うことができる。機械学習モデルは、入力を受信し、受信した入力に基づいて出力を生成する。一部の機械学習モデルは、パラメトリックモデルであり、受信した入力とモデルのパラメータの値とに基づいて出力を生成する。

【0 0 0 4】

一部の機械学習モデルは、モデルの複数のレイヤを使用して、受信した入力に対する出力を生成するディープモデルである。例えば、ディープニューラルネットワークは、出力レイヤと、それぞれが受信した入力に非線形変換を適用して出力を生成する1つ以上の隠れレイヤと、を含むディープマシンラーニングモデルである。

【発明の概要】

【0 0 0 5】

本明細書は、車両の車載システムが、その環境の領域（例えば、交差点の一部）が、未

10

20

30

40

50

来の時間間隔（例えば、現時点の後の次のT秒、例えば、現時点の後の時間T1～T2）で環境内のエージェント（例えば、別の車両）によって占有されるかどうかをどのように予測することができるかについて説明する。

【0006】

第1の態様によれば、現時点での車両の近くの環境を特徴付ける、車両のセンサシステムによって生成されたセンサデータを受信することを含む方法が提供される。センサデータには、それぞれ異なる時点でキャプチャされた環境を特徴付けるセンササンプルのセットが含まれる。センサデータを含むネットワーク入力は、ニューラルネットワークを使用して処理され、環境の領域の占有予測出力を生成する。占有予測出力は、現時点の後の1つ以上の未来の時間間隔について、未来の時間間隔中に環境の領域が環境内のエージェントによって占有される可能性をそれぞれ特徴付ける。ネットワーク入力は、ニューラルネットワークの入力レイヤに提供され、環境の領域に関する占有予測出力は、ニューラルネットワークの出力レイヤによって出力される。占有予測出力を車両の計画システムに提供して、車両の未来の軌道を計画する計画決定を生成する。

10

【0007】

いくつかの実装形態では、センササンプルは、車両の1つ以上のカメラセンサによって生成された画像を含む。

【0008】

いくつかの実装形態では、センササンプルは、車両の1つ以上のLIDARセンサによって生成された点群データ、車両の1つ以上のレーダーセンサによって生成されたセンサデータ、またはその両方、を含む。

20

【0009】

いくつかの実装形態では、現時点の後の未来の時間間隔は、現時点から始まる時間間隔を含む。

【0010】

いくつかの実装形態では、現時点の後の所与の未来の時間間隔のセットのそれぞれについて、占有予測出力は、環境の領域が、所与の未来の時間間隔中に、環境内のエージェントによって占有されるそれぞれの可能性を特徴付ける。

【0011】

いくつかの実装形態では、環境の所与の領域のセットのそれぞれについて、占有予測出力は、環境の所与の領域が、現時点の後の未来の各時間間隔中に、環境内のエージェントによって占有されるそれぞれの可能性を特徴付ける。

30

【0012】

いくつかの実装形態では、環境の領域は、車道の交差点である。

【0013】

いくつかの実装形態では、この方法は、計画システムによって、車道の交差点に関する占有予測出力に基づいて、車両に車道の交差点を横断させる計画決定を生成することをさらに含む。

【0014】

いくつかの実装形態では、環境の領域は、静止したエージェントによって占有されている車道の一部である。

40

【0015】

いくつかの実装形態では、この方法は、計画システムによって、静止したエージェントによって占有されている車道の一部に関する占有予測出力に基づいて、車両を減速させる計画決定を生成することをさらに含む。

【0016】

いくつかの実装形態では、ニューラルネットワークは、畳み込みニューラルネットワークレイヤのセットを含む畳み込みニューラルネットワークである。

【0017】

いくつかの実装形態では、ニューラルネットワークを使用してセンサデータを処理して

50

、環境の領域に関する占有予測出力を生成することは、1つ以上のニューラルネットワークレイヤの第1のセットを使用してセンサデータの第1のサブセットを処理して、第1の中間出力を生成することと、1つ以上のニューラルネットワークレイヤの第2のセットを使用してセンサデータの第2のサブセットを処理して、第2の中間出力を生成することと、1つ以上のニューラルネットワークレイヤの第3のセットを使用して第1の中間出力および第2の中間出力を処理して、占有予測出力を生成することと、を含む。

【0018】

いくつかの実装形態では、センサデータの第1のサブセットは、第1のモダリティのセンサデータを含み、センサデータの第2のサブセットは、第2の異なるモダリティのセンサデータを含む。

10

【0019】

いくつかの実装形態では、ニューラルネットワークは、車両の近くの環境内の他のエージェントの現在または未来の位置を明示的に位置特定することなく、占有予測出力を生成する。

【0020】

いくつかの実装形態では、ネットワーク入力は、車両の近くの環境の静的な特徴を特徴付ける道路グラフデータをさらに含む。

【0021】

いくつかの実装形態では、センサデータは、それぞれの時点でキャプチャされた画像を含み、この方法は、ニューラルネットワークを使用して画像を処理する前に、各画像に1つ以上の事前定義されたトリミング操作を適用することをさらに含む。

20

【0022】

いくつかの実装形態では、センサデータは、それぞれの時点でキャプチャされた画像を含み、ニューラルネットワークは、画像を処理して、画像トリミング操作を指定するデータを生成するように構成されたトリミングサブネットワークを含み、ニューラルネットワークを使用してセンサデータを処理することは、トリミングサブネットワークによって指定された画像トリミング操作を画像に適用することを含む。

【0023】

第2の態様によれば、1つ以上のコンピュータと、1つ以上のコンピュータに通信可能に結合された1つ以上の記憶デバイスであって、1つ以上の記憶デバイスは、命令を記憶しており、命令は、1つ以上のコンピュータによって実行されたときに、1つ以上のコンピュータに第1の態様の方法の動作を行わせる、1つ以上の記憶デバイスと、を含む、システムが提供される。

30

【0024】

第3の態様によれば、1つ以上のコンピュータによって実行されたときに、1つ以上のコンピュータに第1の態様の方法の動作を行わせる、命令を記憶している、1つ以上の非一時的なコンピュータ記憶媒体が提供される。

【0025】

本明細書に記載の主題の特定の実施形態は、以下の利点の1つ以上を実現するように実装することができる。

40

【0026】

本明細書で説明されている占有予測ニューラルネットワークは、センサデータ（例えば、カメラデータ、レーダーデータ、LIDARデータ、またはこれらの何らかの組み合わせ）および道路グラフデータを処理して、環境のターゲット領域が未来の時間間隔においてエージェントによって占有されるかどうかを直接予測することができる。これに対し、従来の占有予測システムは、環境内の各エージェントを位置特定および追跡する中間ステップを実行することによって、環境のターゲット領域が未来の時間間隔においてエージェントによって占有されるかどうかを予測し得る。従来のシステムは、その位置特定データおよび追跡データを使用して、他のエージェントのそれぞれの現在の位置および速度を決定し、次にこのデータを使用して、他のエージェントが未来においてターゲット領域を占

50

有するかどうか、およびいつ占有するかを予測し得る。しかしながら、エージェントの位置特定および追跡は、特にエージェントが、車両の測量センサ（例えば、L I D A Rおよびレーダーセンサ、または3D空間位置特定データを提供するその他のセンサ）の有効範囲を超えているときに、困難でエラーが発生しやすくなり得る。エージェントの位置特定および追跡における固有のエラーおよび不確実性は、従来の占有予測システムによって行われる計算を通じて伝播し、ターゲット領域に関する未来の占有予測を不正確にさせ得る。したがって、センサデータを処理して、環境内の各エージェントを明示的に位置特定および追跡することなく、ターゲット領域の未来の占有を直接予測することにより、本明細書で説明されている占有予測ニューラルネットワークは、一部の従来の占有予測システムよりも正確な占有予測出力を生成し得る。車両の未来の軌道を計画する計画システムは、占有予測ニューラルネットワークによって生成された占有予測出力を使用して、車両をより効率的、安全、かつ快適に動作させるアクションを選択することができる。以下でより詳細に説明するように、本明細書で説明する占有予測ニューラルネットワークは、従来の占有予測システムを補完し得、車両の計画システムは、車両の未来の軌道を計画する際に両方を使用し得る。

10

#### 【0027】

本明細書で説明されている占有予測ニューラルネットワークは、一部の従来の占有予測システムよりも消費する計算リソースが少なくなり得る。例えば、環境内の各エージェントを位置特定および追跡する従来の占有予測システムでは、システムによって行われる計算の量は、環境内のエージェントの数に少なくとも比例して増加し得る。これに対し、本明細書で説明されている占有予測ニューラルネットワークは、環境内の他のエージェントの数に関係なく、1つのフォワードパスを使用してセンサデータを処理することによって占有予測出力を生成する。すなわち、従来の占有予測システムは、環境内の各エージェントが未来の時間間隔においてターゲット領域を占有するかどうかを個別に予測し得るが、占有予測ニューラルネットワークは、任意のエージェントが未来の時間間隔においてターゲット領域を占有するかどうかを直接予測し得る。

20

#### 【0028】

本明細書で説明されている占有予測ニューラルネットワークは、手動のラベル付けを必要とせず、ログに記録されたセンサデータから自動的に生成されるトレーニングデータでトレーニングすることができる。これに対し、エージェントの位置特定および追跡を行う従来の占有予測システムのトレーニングには、例えば、カメラデータ、L I D A Rデータ、レーダーデータ内のオブジェクトの周囲の境界ボックスに人が手動で注釈を付ける場合に、手動でラベル付けされた大量のトレーニングデータが必要になり得る。

30

#### 【0029】

本明細書の主題の1つ以上の実施形態の詳細を、添付の図面および以下の説明に記載する。主題の他の特徴、態様、および利点は、明細書、図面、および特許請求の範囲から、明らかになるであろう。

#### 【図面の簡単な説明】

#### 【0030】

【図1】車両の車載システムが占有予測ニューラルネットワークを使用して、車両の未来の軌道を計画する環境の例を示す。

40

【図2】占有予測ニューラルネットワークを使用する車両の車載システムの例のブロック図である。

【図3】占有予測ニューラルネットワークの例示的なアーキテクチャを示す。

【図4】占有予測ニューラルネットワークをトレーニングするための例示的なトレーニングシステムを示す。

【図5】占有予測ニューラルネットワークを使用して車両の未来の軌道を計画する計画決定を生成するための例示的なプロセスの流れ図である。

#### 【0031】

様々な図面の中の同様の参照番号および名称は、同様の要素を示す。

50

**【発明を実施するための形態】****【0032】**

本明細書は、車両の車載システムが、その環境の領域（例えば、交差点の一部）が、1つ以上の未来の時間間隔（例えば、次のT秒間中）で環境内のエージェント（例えば、車両）によって占有されるかどうかをどのように予測することができるかを説明する。環境の領域が未来の時間間隔においてエージェントによって占有されるかどうかを予測するために、車載システムは、複数の時点（現時点まで）での環境を特徴付けるセンサデータ（例えば、カメラ画像）を「占有予測」ニューラルネットワークに提供することができる。占有予測ニューラルネットワークは、センサデータを処理して、環境の領域が未来の時間間隔においてエージェントによって占有されるかどうかを予測する占有予測出力を生成するように構成されている。

10

**【0033】**

車両の車載システムは、占有予測出力を使用して、車両を制御するアクションを計画および実行することができる。例として、図1に示される環境100内の車両102を取り上げる。この例では、車両102の車載システムは、車両102が安全に交差点104に進入し、ターゲット領域112を横断することによって対向車線（例えば、車両106、108、および110）に合流することを可能にするアクションを計画している。車両102の車載システムは、占有予測ニューラルネットワークを使用して車両102のカメラシステムによってキャプチャされた対向車線を描写する一連のカメラ画像を処理して、ターゲット領域112が次の数秒内で占有されるかどうかを予測することができる。次の数秒間は、ターゲット領域に他のエージェントがいまいであるかと決定することに対応して、車両102の車載システムは、車両102を安全に交差点に進入させ、対向車線に合流させるアクションを実行することができる。

20

**【0034】**

占有予測ニューラルネットワークを使用することにより、車両の車載システムが、環境内の他のエージェントを明示的に位置特定および追跡することなく、ターゲット領域が未来の時間間隔において占有されるかどうかを効果的に予測することを可能にし得る。エージェントを位置特定することとは、環境内のエージェントの空間的な場所を決定することを指し、エージェントを追跡することとは、例えば、エージェントの速度を推定するために、複数の時間ステップにわたってエージェントを繰り返し位置特定することを指す。エージェントの位置を特定することは、特にエージェントが車両の測量センサ（例えば、LIDARまたはレーダーセンサ）の制限された範囲を超えている場合に、困難な場合がある。占有予測ニューラルネットワークは、センサデータに存在する様々な手がかりを暗黙的に利用することを学習することによって、車両の測量センサの範囲を超えるエージェントが存在する場合でも、ターゲット領域の未来の占有を効果的に予測することができる。例えば、占有予測ニューラルネットワークは、他のエージェントの速度および近接性を、車両のカメラセンサによって複数の時間ステップにわたって生成された画像で、それらが拡大（または縮小）するように見える速度に基づいて、暗黙的に特徴付けることを学習し得る。

30

**【0035】**

これらの特徴および他の特徴については、以下でより詳細に説明する。

40

**【0036】**

図2は、車両200の未来の軌道を計画する計画決定を生成する一部として占有予測ニューラルネットワーク204を使用する、車両200の例示的な車載システム202のブロック図である。車載システム202は、ハードウェアコンポーネントおよびソフトウェアコンポーネントから構成され、これらのコンポーネントのうちの一部またはすべては、物理的に車両200に搭載されて位置する。図1の車両200は、自動車として描かれており、本明細書の例は、自動車を参照して説明されているが、占有予測ニューラルネットワークは、他の様々な設定のいずれかで使用することができる。例えば、倉庫または製造環境をナビゲートしているロボットエージェントもまた、計画決定の生成の一部として占

50

有予測ニューラルネットワークを使用してもよい。

【 0 0 3 7 】

車載システム 2 0 2 は、センサシステム 2 0 6、占有予測ニューラルネットワーク 2 0 4、計画システム 2 0 8、および制御システム 2 1 0 を含み、これらはそれぞれ、次に詳細に説明される。

【 0 0 3 8 】

センサシステム 2 0 6 は、車載システム 2 0 2 が車両 2 0 0 の近くの環境を「見る」ことを可能にする。より具体的には、センサシステム 2 0 6 は、1 つ以上のセンサ、例えば、カメラセンサ、L I D A R センサ、およびレーダーセンサを使用して、継続的に（すなわち、複数の時点のそれぞれで）、車両の近くの環境を特徴付けるセンサデータ 2 1 2 を生成する。例えば、センサシステム 2 0 6 は、測量センサ（例えば、L I D A R またはレーダーセンサ）を使用して、環境の幾何学的形状を特徴付ける点群を生成することができる。点群とは、環境内の表面上のそれぞれの点の位置をそれぞれが表すデータ点の集まりを指す（例えば、車、歩行者、道路標識、または建物の表面上の点）。各データ点は、例えば、標準化された座標フレーム内の  $x - y - z$  座標、例えば、車両 2 0 0 を中心とする座標フレームによって表され得る。別の例として、センサシステム 2 0 6 は、カメラセンサを使用して、環境の視覚的外観を特徴付ける画像（例えば、赤 - 緑 - 青（R G B）画像）を生成し得る。本明細書全体を通して使用されるように、「時点」でキャプチャされるものとして参照されるセンササンプルは、一定期間にわたってキャプチャされ得、例えば、L I D A R センサによって生成されるセンササンプルは、1 0 0 m s の期間にわたる L I D A R センサの 1 回転中にキャプチャされるデータからなり得る。この例では、L I D A R センサのサンプルがキャプチャされた時点は、例えば、L I D A R センサの回転の最終時点と見なされ得る。

【 0 0 3 9 】

占有予測ニューラルネットワーク 2 0 4 は、( i ) 一連の複数の時点で生成されたセンサデータ 2 1 2、および ( i i ) 環境のターゲット領域の占有予測出力 2 1 4 を生成するための道路グラフデータ 2 1 8 を処理するように構成されている。図 2 に示す例では、現時点は、 $t$  であり、センサデータ 2 1 2 は、時点  $t$ （すなわち、現時点）、 $t - 1$ （例えば、現時点の 1 秒前）、および  $t - 2$ （例えば、現時点の 2 秒前）で生成される。道路グラフデータ 2 1 8 は、車両の近くの環境、例えば、環境の様々な静的な特徴の位置およびタイプ、例えば、車線、道路標識、横断歩道、信号機などを特徴付ける。道路グラフデータ 2 1 8 は、様々な方法のいずれかで表すことができ、例えば、車両の近くの環境のトップダウンビューを表す 2 D 画像として表すことができる。センサデータ 2 1 2 および道路グラフデータ 2 1 8 に加えて、占有予測ニューラルネットワークは、例えば、環境内の信号機の状態（例えば、赤、黄、または緑）を指定する、または車両 2 0 0 の意図された未来の軌道（すなわち、他のエージェントの行動に影響を与え得る）を指定する、他の入力を追加的に処理し得る。占有予測出力 2 1 4 は、環境のターゲット領域が、現時点の後の未来の時間間隔、例えば、現時点から始まる未来の時間間隔において、エージェントによって占有される可能性を特徴付ける。より一般的には、占有予測出力 2 1 4 は、以下でより詳細に説明するように、環境の 1 つ以上のターゲット領域が、現時点の後の 1 つ以上の未来の時間間隔のそれぞれの間にエージェントによって占有されるそれぞれの可能性を特徴付け得る。

【 0 0 4 0 】

いくつかの実装形態では、占有予測出力 2 1 4 は、環境の複数のターゲット領域のそれぞれが未来の時間間隔においてエージェントによって占有されるそれぞれの可能性を特徴付け得る。例えば、環境のターゲット領域は、車両 2 0 0 の現在の場所を中心とする事前定義されたグリッドに配置された事前定義されたサイズ（例えば、1 0 フィート  $\times$  1 0 フィート）を有する領域であり得る。ヒートマップ 2 1 6 は、占有予測出力 2 1 4 の例を示している。ヒートマップ 2 1 6 のそれぞれは、セルのグリッドによって表され、各セルは、環境のそれぞれの領域に対応する（グリッド自体は、車両の場所を中心とし得る）。各

10

20

30

40

50

セルの陰影は、セルに対応する環境の領域が未来の時間間隔、例えば、時間間隔  $[t, t + 1]$ 、 $[t, t + 2]$ 、および  $[t, t + 3]$  において、エージェントによって占有される可能性を表し、例えば、暗い陰影は、その領域が未来の時間間隔において占有される可能性がより高いことに対応する。例示的なヒートマップ 216 は、デカルト座標系におけるセルのグリッドを描写しているが、より一般的には、セルは、任意の適切な座標系、例えば、極座標系で表され得る。

#### 【0041】

任意選択的に、環境のターゲット領域が事前定義されるのではなく、占有予測ニューラルネットワーク 204 は、占有予測出力によって特徴付けられるターゲット領域を指定する追加の入力を処理するように構成されてもよい。例えば、占有予測ニューラルネットワーク 204 は、環境のターゲット領域を指定するマスク画像（すなわち、センサデータ 212 に加えて）を処理するように構成されてもよい。マスク画像は、数値の 2 次元配列として表すことができ、配列の各成分は、環境のそれぞれの部分に対応する。ターゲット領域に対応するマスク画像の成分は、第 1 の値（例えば、値 1）を有し得、一方、マスク画像の他の成分は、異なる値（例えば、値 0）を有し得る。車載システム 202 は、環境内の車両の現在の状態に基づいて、環境のターゲット領域を動的に選択し得る。例えば、車両 200 を所与の車線に進入させる前に、車載システム 202 は、例えば、所与の車線を指定するマスク画像を含む占有予測ニューラルネットワーク 204 に入力を提供することによって、所与の車線に対応するターゲット領域を選択し得る。

#### 【0042】

任意選択的に、占有予測出力 214 は、複数の未来の時間間隔のそれぞれの間に各ターゲット領域が占有される可能性を特徴付けてもよい。一例では、未来の時間間隔のセットは、現時点から開始する連続的で重複しない未来の時間間隔であり得、例えば、未来の時間間隔は、以下によって与えられ得る。

#### 【数 1】

$$\{[t_0 + (n - 1) \cdot i, t_0 + n \cdot i]\}_{n=1}^N \quad (1)$$

式中、 $t_0$  は、現時点、 $i$  は、未来の各時間間隔の一定期間、 $N$  は、未来の時間間隔の総数である。別の例では、未来の時間間隔のセットは、それぞれが現時点から開始する異なる期間の未来の時間間隔であり得る。図 2 に示される例を参照すると、占有予測出力 214 は、未来の各時間間隔  $[t, t + 1]$ 、 $[t, t + 2]$ 、 $[t, t + 3]$ （式中、 $t$  は、現時点である）に対応するそれぞれの占有予測ヒートマップ 216 を指定し得る。

#### 【0043】

一般に、占有予測ニューラルネットワーク 204 は、それがその説明された機能を行うことを可能にする任意の適切なニューラルネットワークアーキテクチャを有し得る。例示的な占有予測ニューラルネットワークアーキテクチャは、図 3 を参照してより詳細に説明される。さらに、占有予測ネットワーク 204 は、前の時点からのセンサデータサンプル 212 に加えて、他の入力を処理するように構成され得る。例えば、占有予測ネットワーク 204 は、車両の近くの環境の静的な構造、例えば、車線の構造、ならびに環境内の静的なオブジェクトの位置およびタイプ（例えば、横断歩道、信号機、道路標識）を定義する道路グラフデータを追加的に処理してもよい。

#### 【0044】

計画システム 208 は、（例えば、環境の現在の状態および予測される未来の状態を特徴付ける他の様々なデータに加えて）占有予測出力を処理して、車両 200 の未来の軌道を計画する計画決定を生成するように構成されている。図 1 を参照して説明される例では、ターゲット領域 112 が次の 10 秒間占有されないままであることを指定する占有予測出力を受信することに対応して、計画システム 208 は、車両 200 にターゲット領域を横断させる計画決定を生成し得る。計画決定は、操縦および加速アクションなどの一連の

1つ以上のアクションによって指定され得る。

【0045】

制御システム210は、計画システム208によって生成された計画決定によって指定されたアクションを実行することによって、車両200の動作の一部またはすべてを制御するように構成されている。例えば、一連の操縦および加速アクションを実行するための計画決定を受信することに応答して、制御システムは、電子信号を車両の操縦および加速制御ユニットに送信して、それらに操縦および加速アクションを実行させ得る。

【0046】

任意選択的に、車載システム202は、占有予測ニューラルネットワーク204に加えて、従来の占有予測システムを実装してもよい。従来の占有予測システムは、各エージェントを位置特定および追跡して、エージェントの現在の位置および速度を明示的に推定し、次に、エージェントがその現在の位置および速度に基づいてターゲット領域を占有するかどうか、およびいつ占有するかを予測し得る。従来のシステムによって生成された占有予測出力は、占有予測ニューラルネットワーク204によって生成された占有予測出力を補完し得る。例えば、占有予測ニューラルネットワーク204は、環境内の1つ以上のエージェントが車両200の測量センサ（例えば、LIDARおよびレーダー）の有効範囲を超えているときに、従来の占有予測システムよりも正確な占有予測出力を生成し得る。他方、従来の占有予測システムは、車両200の測量センサの有効範囲内のエージェントに対してより正確な占有予測出力を生成することができる。計画システム208は、占有予測ニューラルネットワーク204および従来の占有予測システムの両方を使用して計画決定を生成することにより、それらの補完的な性質を利用し得る。

【0047】

様々な状況において、車載システム202は、特定のエージェントを正確に位置特定することができないにもかかわらず、占有予測ニューラルネットワーク204を使用して、車両200の安全で快適な未来の軌道を計画することができる。次にいくつかの実施例を示す。

【0048】

一例では、車載システム202は、占有予測ネットワークを使用して、車両の測量センサの有効範囲外にある複数の対向するエージェントを伴う高速道路に車両を方向転換させる軌道（例えば、対向車線を横切って左折するための）を効果的に計画することができる。

【0049】

別の例では、車載システム202は、占有予測ニューラルネットワークを使用して、車両の測量センサの有効範囲外にある、道路上の静止したエージェントの前の車両を減速させる軌道を効果的に計画することができる。この例では、静止したエージェントは、車両からかなり離れているため（例えば、100メートル以上）、または障害物（例えば、丘や障壁）が車両と静止したエージェントとの間の視線を部分的に遮っているため、測量センサの有効範囲外にあり得る。

【0050】

別の例では、トラック（例えば、セミトレーラートラック）の車載システムが、占有予測ニューラルネットワークを使用して、トラックの長引く停止時間に対応するための長期占有予測を行い得る。例えば、トラックの車載システムは、占有予測ニューラルネットワークを使用して、トラックの前の道路の車線が、例えば、停車中もしくは低速で移動する車両のために、または道路工事を示すコーンのために、ブロックされているかどうかについての長期予測を行い得る。

【0051】

別の例では、車両の車載システムは、占有予測ニューラルネットワークを使用して、環境の複数の異なる視点からの、例えば、車両の異なる側に位置するカメラセンサからのセンサデータを考慮に入れた占有予測出力を生成し得る。複数の視点からのセンサデータを使用することにより、例えば、車両の測量センサ（LIDARおよびレーダーセンサなど）が同時に環境の1つの視点しか特徴付けることができない状況でも、車載システムが正

10

20

30

40

50

確な占有予測出力を作成することを可能にし得る。

【 0 0 5 2 】

いくつかの実装形態では、占有予測ニューラルネットワークは、( i ) ターゲット領域が未来の時間間隔において占有されるかどうかと、( i i ) 特定のタイプのエージェントが未来の時間間隔においてターゲット領域を占有する可能性と、の両方を特徴付ける占有予測出力を生成するように構成され得る。エージェントの「タイプ」は、例えば、車両、自転車乗り、歩行者、または他の任意の適切なタイプであり得る。ターゲット領域ごとにおよび未来の時間間隔ごとに、占有予測ニューラルネットワークは、事前定義されたエージェントタイプのセット内のタイプごとに、そのタイプのエージェントが未来の時間間隔においてターゲット領域を占有する可能性を指定する出力を生成し得る。

10

【 0 0 5 3 】

場合によっては、車両 2 0 0 は、実世界の環境をナビゲートする実世界の車両ではなく、シミュレーションシステム内のシミュレートされた環境をナビゲートするシミュレートされた車両であり得る。シミュレートされた車両の場合（またはログに記録されたセンサデータを分析する文脈で）、「現」時点は、未来の占有予測が生成される任意の時点を指す（すなわち、ターゲット領域が現時点の後の未来の時間間隔において占有されるかどうかを特徴付ける）と理解され得る。

【 0 0 5 4 】

図 3 は、占有予測ニューラルネットワークの例示的なアーキテクチャ 3 0 0 を示している。占有予測ネットワークは、図 2 を参照して説明したように、複数の時点のそれぞれでの車両の近くの環境を特徴付けるセンササンプルのセットを処理して、占有予測出力 3 0 2 を生成するように構成されている。

20

【 0 0 5 5 】

例示的なアーキテクチャ 3 0 0 では、占有予測ニューラルネットワークは、カメラ画像のセット 3 0 4、L I D A R 画像のセット 3 0 6、および道路グラフ 3 1 4 を含む入力を処理する。一例では、占有予測ニューラルネットワークは、それぞれ 0 . 2 秒間隔でキャプチャされた 1 0 個のカメラ画像 3 0 4、すなわち、2 秒の期間にわたるカメラ画像を処理し得る。同様に、占有予測ニューラルネットワークは、それぞれ 0 . 2 秒間隔でキャプチャされ、かつカメラ画像 3 0 4 とほぼ同じ 2 秒の期間にわたる 1 0 個の L I D A R 画像 3 0 6 を処理し得る。L I D A R 「画像」とは、L I D A R 点群のトップダウン表現を指す。

30

【 0 0 5 6 】

占有予測ニューラルネットワークは、画像特徴ニューラルネットワークを使用してカメラ画像 3 0 4 の結合されたセットを処理して、カメラ画像 3 0 4 の代替表現を生成する。同様に、占有予測ニューラルネットワークは、L I D A R 特徴ニューラルネットワークを使用して L I D A R 画像 3 0 6 の結合されたセットを処理して、L I D A R 画像 3 0 6 の中間表現を生成する。画像特徴ニューラルネットワーク 3 0 8 および L I D A R 特徴ニューラルネットワーク 3 1 0 は、例えば、一連の 2 次元 ( 2 D ) または 3 次元 ( 3 D ) の畳み込みレイヤを含むそれぞれのアーキテクチャを有し得る。カメラ画像 3 0 4 および L I D A R 画像 3 0 6 の代替表現はそれぞれ、数値の順序付けられた集まり、例えば、2 つの「空間」次元および 1 つ以上の「チャンネル」次元を有する数値の配列として表され得る。

40

【 0 0 5 7 】

占有予測ニューラルネットワークは、カメラ画像 3 0 4 および L I D A R 画像 3 0 6 の代替表現を結合し、融合ニューラルネットワーク 3 1 2 を使用して結合表現を処理して、占有予測出力を生成する。融合ニューラルネットワークは、一連の畳み込みレイヤとそれに続く出力レイヤを含むアーキテクチャを有し得る。占有予測出力 3 0 2 は、例えば、図 2 を参照して説明されたように、例えば、それぞれが未来の時間間隔中のターゲット領域のグリッドの占有可能性をそれぞれ特徴付ける、例えば、一連のヒートマップとして表され得る。一例では、占有予測ニューラルネットワークは、複数の出力レイヤを有し得、各出力レイヤは、それぞれの未来の時間間隔に対応する占有予測出力を生成する。所与の未

50

来の時間間隔に対応する出力レイヤは、例えば、環境の各ターゲット領域に対応するそれぞれのニューロンを有するシグモイド出力レイヤであり得る。

【0058】

占有予測ニューラルネットワークは、車両の近くの環境を特徴付ける道路グラフの表現をさらに処理し得る。例えば、占有予測ニューラルネットワークは、道路グラフの表現を、(i)カメラ画像304およびLIDAR画像306、または(ii)カメラ画像304およびLIDAR画像306の代替表現に結合し得る。

【0059】

場合によっては、占有予測ニューラルネットワークによって処理されたカメラ画像304のそれぞれは、より広い視野を有する別のカメラ画像からトリミングされ得る。例えば、各カメラ画像304は、車両から離れた一定の距離（例えば、50メートルまたは150メートル）を中心とする環境の領域を描写する事前定義された画像トリミングであり得る。別の例として、カメラ画像304が事前定義された方法でトリミングされるのではなく、占有予測ネットワークは、例えば、残りの占有予測ニューラルネットワークと共同でトレーニングされる、学習済みアテンション機構を使用してカメラ画像304を動的にトリミングし得る。

【0060】

学習済みアテンション機構は、例えば、カメラ画像304を処理して、カメラ画像からトリミングされる領域、例えば、カメラ画像の304の基準フレーム内の境界ボックスの頂点、を定義するデータを生成する「トリミング」ニューラルネットワークによって実装され得る。この例では、占有予測ニューラルネットワークは、画像特徴ニューラルネットワーク308にそれらを提供する前に、トリミングニューラルネットワークの出力に従ってカメラ画像304をトリミングし得る。占有予測ニューラルネットワークは、同様の学習済み（または事前定義済み）アテンション機構を使用して、LIDAR機能ニューラルネットワーク310を使用してそれら进行处理する前に、LIDAR画像306をトリミングすることができる。トリミングニューラルネットワークは、任意の適切なニューラルネットワークアーキテクチャ、例えば、S. Ren, K. He, R. Girshick, J. Sunの“Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Neural Information Processing Systems (NeurIPS) Conference, (2015)を参照して説明されている領域提案ネットワークアーキテクチャを有することができる。アテンション機構を使用してセンササンプル（例えば、カメラ画像およびLIDAR画像）をトリミングすることにより、センササンプルの最も関連性の高い部分に焦点を当てることで、占有予測ニューラルネットワークはより正確な占有予測出力を生成できるようになり得る。

【0061】

占有予測ニューラルネットワークの他の様々なアーキテクチャが可能である。例えば、占有予測ニューラルネットワークは、それぞれが異なるカメラ画像のセットを処理するように構成されている複数の画像特徴ニューラルネットワークを有し得る。一例では、各画像特徴ニューラルネットワークは、異なる時点での車両から一定の距離を中心とする環境の領域を描写するカメラ画像のセットを処理し得る。別の例では、各画像特徴ニューラルネットワークは、それぞれの向きを有するそれぞれのカメラによってキャプチャされたカメラ画像のセットを処理し得、例えば、1つの画像特徴ニューラルネットワークは、正面カメラによってキャプチャされた画像を処理し得、別の画像特徴ニューラルネットワークは、側面カメラによってキャプチャされた画像を処理し得る。

【0062】

図4は、例示的なトレーニングシステム400を示している。トレーニングシステム400は、以下に説明するシステム、コンポーネント、および技法が実装される1つ以上の場所（例えば、データセンター402）にある1つ以上のコンピュータ上にコンピュータプログラムとして実装されるシステムの例である。

10

20

30

40

50

## 【 0 0 6 3 】

トレーニングシステム 4 0 0 は、車両 4 1 0 の車載システム 4 0 8 の占有予測ニューラルネットワークのモデル重み 4 0 6 のためのトレーニング済み値 4 0 4 を決定し、次に、例えば、有線または無線接続を使用して、トレーニング済み重み値 4 0 4 を車載システムに提供するように構成されている。占有予測ニューラルネットワークのモデル重み 4 0 6 は、例えば、占有予測ニューラルネットワークのニューラルネットワークレイヤのレイヤパラメータの値を定義する、占有予測ニューラルネットワークのトレーニング可能なパラメータを指す。

## 【 0 0 6 4 】

トレーニングシステム 4 0 0 は、トレーニングデータ 4 1 4 のセットで占有予測ニューラルネットワーク 4 1 2 をトレーニングすることによって、トレーニング済み重み値 4 0 4 を決定する。トレーニングデータ 4 1 4 は、トレーニング例のセットからなり、各トレーニング例は、( i ) 異なる時点でキャプチャされたセンサデータサンプル（例えば、カメラ画像および L I D A R データ）を含むトレーニング入力、および ( i i ) ターゲット占有出力を含む。ターゲット占有出力は、対応するトレーニング入力を処理することによって占有予測ニューラルネットワーク 4 1 2 によって生成されるべき出力を表す。

## 【 0 0 6 5 】

トレーニングシステム 4 0 0 は、ログに記録されたセンサデータ 4 1 6 のセット、すなわち、1 つ以上の車両のセンサによって以前に生成された格納されたセンサデータを処理することによって、自動化された方法でトレーニングデータ 4 1 4 を生成することができる。トレーニング例を生成するために、トレーニングシステム 4 0 0 は、トレーニングデータからの一連の時間ステップにわたってセンサデータサンプル（例えば、カメラ画像および L I D A R データ）を集約することによってトレーニング入力を識別することができる。トレーニングシステム 4 0 0 は、後続のタイムステップからのセンサデータサンプルを処理して、環境の 1 つ以上のターゲット領域が後続のタイムステップで占有されたかどうかを判断することによって、トレーニング入力に対応するターゲット占有出力を識別することができる。すなわち、トレーニングシステム 4 0 0 は、ログに記録されたセンサデータ 4 1 6 を処理して、第 1 の期間にわたって測定されたセンサデータを後続の時点で測定されたセンサデータ（例えば、数秒後のターゲット領域におけるエージェントの占有を特徴付ける）と確実に関連付けることによって、トレーニング例を生成し得る。後続の時点で、環境内の他のエージェントは、ログに記録されたセンサデータを生成した車両のセンサの近くにいた場合があり、したがって、後続の時点で測定されたセンサデータは、第 1 の期間中に測定されたセンサデータよりも正確にエージェントを特徴付け得る。このようにして、トレーニングシステム 4 0 0 は、人間の介入を最小限にするか、または全くなしに、トレーニングデータ 4 1 4 を生成することができる。

## 【 0 0 6 6 】

トレーニングシステム 4 0 0 は、複数のトレーニング反復にわたってトレーニングデータ 4 1 4 について占有予測ニューラルネットワーク 4 1 2 をトレーニングすることができる。各トレーニング反復で、トレーニングシステム 4 0 0 は、トレーニング例の「バッチ」（セット）を選択し、占有予測ニューラルネットワーク 4 1 2 の現在の重み 4 1 8 に従ってトレーニング例のトレーニング入力を処理して、対応するトレーニング出力 4 2 0 を生成し得る。トレーニングエンジン 4 2 2 は、( i ) トレーニング出力 4 2 0 と、( i i ) トレーニング例によって指定された対応するターゲット占有予測出力と、の間の類似性を測定する目的関数の勾配を計算することによって、モデル重み 4 0 6 への更新 4 2 4 を決定し得る。目的関数は、例えば、交差エントロピー目的関数であり得、トレーニングエンジン 4 2 2 は、バックプロパゲーション法を使用して目的関数の勾配を計算し得、トレーニングエンジン 4 2 2 は、任意の適切な勾配降下アルゴリズム、例えば、A d a m または R M S p r o p を使用して勾配から重み更新 4 2 4 を決定し得る。

## 【 0 0 6 7 】

トレーニングシステム 4 0 0 は、トレーニング終了基準が満たされたとき、例えば、所

定の回数のトレーニング反復が行われたときに、または目的関数の値が所定の閾値を満たしたときに、トレーニングが完了したと判断し得る。トレーニングが完了したと判断した後、トレーニングシステム 400 は、図 2 を参照して説明したように、車両 410 の未来の軌道を計画する計画決定を生成するために使用するために、占有予測ネットワーク 412 のトレーニング済み重み値 404 を車載システム 408 に提供することができる。

【0068】

図 5 は、占有予測ニューラルネットワークを使用して車両の未来の軌道を計画する計画決定を生成するための例示的なプロセス 500 の流れ図である。便宜上、プロセス 500 は、1 つ以上の場所に位置する 1 つ以上のコンピュータのシステムによって実行されるものとして説明する。例えば、車載システム、例えば、本明細書に従って適切にプログラムされた図 2 の車載システム 202 は、プロセス 500 を実行することができる。

10

【0069】

システムは、現時点での車両の近くの環境を特徴付ける車両のセンサシステムによって生成されたセンサデータを受信する(502)。センサデータには、それぞれ異なる時点でキャプチャされた環境を特徴付けるセンサデータサンプルが含まれる。センサデータサンプルとは、ある時点で車両のセンサによって生成されたセンサデータの要素(例えば、カメラ画像または L I D A R 点群)を指す。センサデータサンプルには、例えば、車両のカメラセンサによって生成された画像、車両の L I D A R もしくはレーダーセンサによって生成された点群データ、またはその両方が含まれ得る。

【0070】

20

システムは、占有予測ニューラルネットワークを使用してセンサデータを含むネットワーク入力を処理して、環境の領域の 1 つ以上の占有予測出力を生成する(504)。ネットワーク入力は、ニューラルネットワークの入力レイヤに提供され、占有予測出力は、ニューラルネットワークの出力レイヤによって出力される。ニューラルネットワークの「入力レイヤ」とは、外部入力を受信し(すなわち、別のニューラルネットワークレイヤからではなくニューラルネットワークの外部から)、ニューラルネットワークの他の 1 つ以上のレイヤにその外部入力を提供するニューラルネットワークレイヤを指す。占有予測出力は、現時点の後の 1 つ以上の未来の時間間隔について、現時点の後の未来の時間間隔中に環境の領域が環境内のエージェントによって占有されるそれぞれの可能性を特徴付ける。未来の時間間隔には、現時点から始まる時間間隔が含まれ得る。場合によっては、占有予測出力は、現時点の後の複数の未来の時間間隔のそれぞれにおいて環境の領域がエージェントによって占有されるそれぞれの可能性を特徴付け得る。場合によっては、未来の時間間隔ごとに、占有予測出力は、環境の複数の領域のそれぞれが未来の時間間隔においてエージェントによって占有されるそれぞれの可能性を特徴付ける。ネットワーク入力には、追加データ、例えば、道路グラフデータが含まれ得る。

30

【0071】

場合によっては、システムは、占有予測ニューラルネットワークを使用してセンサデータ画像を処理する前に、事前定義されたトリミング操作をセンサデータ画像に適用し得る。場合によっては、占有予測ニューラルネットワークは、ネットワーク入力に含まれる画像を処理して画像トリミング操作を指定するデータを生成するように構成されているトリミングサブネットワークを含み得る。占有予測ニューラルネットワークは、占有予測出力を生成する一部として、画像にトリミング操作を適用し得る。

40

【0072】

占有予測ニューラルネットワークは、複数の畳み込みニューラルネットワークレイヤを含むニューラルネットワークアーキテクチャを有し得る。一例では、占有予測出力を生成するために、占有予測ニューラルネットワークは、1 つ以上のニューラルネットワークレイヤの第 1 のセットを使用して第 1 のモダリティ(例えば、カメラ画像)を有するセンサデータの第 1 のサブセットを処理して、第 1 の中間出力を生成し得る。占有予測ニューラルネットワークは、1 つ以上のニューラルネットワークレイヤの第 2 のセットを使用して第 2 のモダリティ(例えば、L I D A R 画像)を有するセンサデータの第 2 のサブセット

50

を処理して、第2の中間出力を生成し得る。次に、占有予測ニューラルネットワークは、例えば、図3を参照してより詳細に説明したように、1つ以上のニューラルネットワークレイヤの第3のセットを使用して第1の中間出力および第2の中間出力を処理することによって、占有予測出力を生成し得る。一般に、占有予測ニューラルネットワークは、環境内の他のエージェントの位置を明示的に位置特定することなく、占有予測出力を生成する。

【0073】

システムは、車両の計画システムに占有予測出力を提供して、車両の未来の軌道を計画する計画決定を生成する(506)。計画決定は、車両によって実行される一連の1つ以上のアクション(例えば、操縦または加速アクション)を指定し得る。一例では、車両は、車道の交差点の近くにあり得、計画システムは、交差点に関する占有予測出力を処理して、車両が交差点を横断するようにする計画決定を生成し得る。別の例では、車両の近くにいるエージェントのうちの1つが、車道に静止した状態であり得、計画システムは、静止したエージェントが占める環境の部分の占有予測出力を処理して、車両を減速させる計画決定を生成し得る。

10

【0074】

本明細書では、システムおよびコンピュータプログラムコンポーネントに関連して「構成された」という用語を使用している。1つ以上のコンピュータのシステムが特定の操作または動作を実行するように構成されているとは、システムが、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組み合わせをインストールしており、それらが動作中に、システムにその操作または動作を実行させることを意味する。1つ以上のコンピュータプログラムが特定の操作または動作を実行するように構成されているとは、1つ以上のプログラムが、データ処理装置によって実行されると、装置に操作または動作を行わせる命令を含むことを意味する。

20

【0075】

本明細書に記載の主題および機能的動作の実施形態は、デジタル電子回路内に、有形的に具現化されたコンピュータソフトウェアもしくはファームウェア内に、本明細書に開示された構造体およびそれらの構造上の等価物を含むコンピュータハードウェア内に、またはそれらのうちの1つ以上を組み合わせ、実装することができる。本明細書に記載の主題の実施形態は、1つ以上のコンピュータプログラムとして、すなわち、データ処理装置によって実行するために、またはデータ処理装置の動作を制御するために有形の非一時的記憶媒体に符号化されたコンピュータプログラム命令の1つ以上のモジュールとして実装することができる。コンピュータ記憶媒体は、機械可読記憶デバイス、機械可読記憶基板、ランダムもしくはシリアルアクセスメモリデバイス、またはこれらのうちの1つ以上の組み合わせであってもよい。代替的に、またはさらに、プログラム命令は、人工的に生成された伝播信号、例えば、データ処理装置によって実行するために好適な受信装置に送信される情報を符号化するために生成される機械生成の電気、光、または電磁信号に符号化され得る。

30

【0076】

「データ処理装置」という用語は、データ処理ハードウェアを指し、データを処理するためのあらゆる種類の装置、デバイス、および機械を包含し、それらには、例として、プログラマブルプロセッサ、コンピュータ、または複数のプロセッサもしくはコンピュータが含まれる。装置はまた、例えば、FPGA(フィールドプログラマブルゲートアレイ)もしくはASIC(特定用途向け集積回路)などの専用論理回路であってもよく、またはそれをさらに含んでもよい。装置は、ハードウェアに加えて、コンピュータプログラムのための実行環境を作り出すコード、例えば、プロセッサファームウェア、プロトコルスタック、データベース管理システム、オペレーティングシステム、またはそれらのうちの1つ以上の組み合わせを構成するコードを任意選択的に含むことができる。

40

【0077】

プログラム、ソフトウェア、ソフトウェアアプリケーション、アプリ、モジュール、ソフトウェアモジュール、スクリプト、またはコードとも称され得るか、もしくは記載され

50

得る、コンピュータプログラムは、コンパイラ型もしくはインタプリタ型言語、または宣言型言語もしくは手続き型言語を含む、任意の形式のプログラミング言語で記述され得、かつスタンドアロンプログラムもしくはモジュール、コンポーネント、サブルーチン、またはコンピューティング環境で使用するのに適した他のユニットを含む、任意の形式で展開され得る。プログラムは、ファイルシステム内のファイルに対応する場合もあるが、必ずしもそうである必要はない。プログラムは、他のプログラムもしくはデータを保持するファイルの一部、例えば、マークアップ言語文書に格納された1つ以上のスクリプト、当該プログラム専用の単一ファイル、または複数の調整ファイル、例えば、1つ以上のモジュール、サブプログラム、もしくはコードの一部を格納するファイルに格納することができる。コンピュータプログラムは、1つのコンピュータもしくは1つの場所に配置された複数のコンピュータ上で実行されるように展開されるか、または複数の場所に分散されて、データ通信ネットワークで相互接続することができる。

10

## 【0078】

本明細書において、「エンジン」という用語は、1つ以上の特定の機能を遂行するようにプログラムされている、ソフトウェアベースのシステム、サブシステム、またはプロセスを指すために広く使用される。概して、エンジンは、1つ以上のソフトウェアモジュールまたはコンポーネントとして実装され、1つ以上の場所にある1つ以上のコンピュータにインストールされる。場合によっては、1つ以上のコンピュータが特定のエンジン専用になり、それ以外の場合は、複数のエンジンを1つ以上の同じコンピュータにインストールして実行することができる。

20

## 【0079】

本明細書に記載のプロセスおよび論理フローは、1つ以上のプログラマブルコンピュータが1つ以上のコンピュータプログラムを実行して、入力データを操作し、かつ出力を生成することで機能を果たすことによって実行することができる。プロセスおよび論理フローはまた、FPGAもしくはASICなどの専用論理回路によって、または専用論理回路と1つ以上のプログラムされたコンピュータとの組み合わせによって実行することができる。

## 【0080】

コンピュータプログラムの実行に好適なコンピュータは、汎用もしくは専用のマイクロプロセッサまたはその両方、あるいは他の種類の中央処理装置に基づくことができる。概して、中央処理装置は、読み取り専用メモリもしくはランダムアクセスメモリ、またはその両方から命令およびデータを受信することになる。コンピュータの本質的な要素は、命令を遂行または実行するための中央処理装置ならびに命令およびデータを格納するための1つ以上のメモリデバイスである。中央処理装置およびメモリは、専用論理回路によって補完されるか、またはその回路に組み込むことができる。概して、コンピュータはまた、例えば、磁気ディスク、光磁気ディスク、もしくは、光ディスクなど、データを格納するための1つ以上の大容量記憶デバイスを含むか、または、それらかとの間でデータを送受信するように動作可能に結合されることになる。しかしながら、コンピュータはそのようなデバイスを有する必要はない。さらに、コンピュータは別のデバイス、例えば、ほんの数例を挙げると、携帯電話、電子手帳(PDA)、モバイルオーディオもしくはビデオプレーヤ、ゲームコンソール、全地球測位システム(GPS)受信機、またはポータブル記憶デバイス、例えば、ユニバーサルシリアルバス(USB)フラッシュドライブに組み込むことができる。

30

40

## 【0081】

コンピュータプログラム命令およびデータを格納するのに好適なコンピュータ可読媒体は、不揮発性メモリ、媒体、およびメモリデバイスのあらゆる形態を含み、例として、EPROM、EEPROM、およびフラッシュメモリデバイスなどの半導体メモリデバイス、内蔵ハードディスクまたは取り外し可能なディスクなどの磁気ディスク、光磁気ディスク、ならびにCD-ROMおよびDVD-ROMディスクを含む。

## 【0082】

50

ユーザとの対話を提供するために、本明細書に記載の主題の実施形態は、コンピュータ上で実装することができ、コンピュータは、ユーザに情報を表示するための表示デバイス、例えば、CRT（陰極線管）もしくはLCD（液晶ディスプレイ）モニタ、ならびにユーザがコンピュータに入力を提供することができる、キーボードおよびマウス、トラックボールなどのポインティングデバイスを有する。他の種類のデバイスを使用して、ユーザとの対話を提供することもでき、例えば、ユーザに提供されるフィードバックは、視覚的フィードバック、聴覚的フィードバック、または触覚的フィードバックなどの任意の形式の感覚的フィードバックであり得、ユーザからの入力は、音響、音声、または触覚入力を含む任意の形式で受信することができる。さらに、コンピュータは、ユーザが使用するデバイスとの間で文書を送受信することによって、例えば、ウェブブラウザから受信した要求に回答して、ユーザのデバイス上のウェブブラウザにウェブページを送信することによって、ユーザと対話することができる。また、コンピュータは、テキストメッセージまたは他の形式のメッセージを、メッセージアプリケーションを実行するスマートフォンなどのパーソナルデバイスに送信し、代わりにユーザからの応答メッセージを受信することにより、ユーザと対話することができる。

10

**【0083】**

機械学習モデルを実装するためのデータ処理装置には、例えば、機械学習のトレーニングまたは本稼働の一般的で演算集約型の部分、すなわち、推論、ワークロードを処理するための専用ハードウェアアクセラレータユニットを含めることもできる。

**【0084】**

20

機械学習モデルは、TensorFlowフレームワーク、Microsoft Cognitive Toolkitフレームワーク、Apache Singaフレームワーク、またはApache MXNetフレームワークなどの機械学習フレームワークを使用して実装および展開することができる。

**【0085】**

本明細書に記載の主題の実施形態は、バックエンドコンポーネントを、例えば、データサーバとして含むか、またはミドルウェアコンポーネントを、例えば、アプリケーションサーバとして含むか、またはフロントエンドコンポーネント、例えば、それらを経由してユーザが本明細書に記載の主題の実装形態と対話することができる、グラフィカルユーザインターフェース、ウェブブラウザ、もしくはアプリを有する、クライアントコンピュータを含む、コンピューティングシステム、あるいは1つ以上のかかるバックエンド、ミドルウェア、もしくはフロントエンドのコンポーネントの任意の組み合わせに実装され得る。システムのコンポーネントは、デジタルデータ通信の任意の形式または媒体、例えば、通信ネットワークによって相互接続され得る。通信ネットワークの例には、ローカルエリアネットワーク（LAN）およびワイドエリアネットワーク（WAN）、例えば、インターネットが含まれる。

30

**【0086】**

コンピューティングシステムには、クライアントとサーバとを含むことができる。クライアントおよびサーバは概して、互いにリモートであり、典型的には通信ネットワークを介して相互作用する。クライアントとサーバとの関係は、それぞれのコンピュータで実行され、かつ互いにクライアントとサーバとの関係を有している、コンピュータプログラムによって生じる。いくつかの実施形態では、サーバは、データ、例えば、HTMLページを、ユーザデバイスに、例えば、クライアントとして動作するデバイスと対話するユーザにデータを表示し、ユーザからのユーザ入力を受信する目的で、送信する。ユーザデバイスで生成されたデータは、例えば、ユーザ対話の結果は、デバイスからサーバで受信することができる。

40

**【0087】**

本明細書は多くの特定の实装形態の詳細を含んでいるが、これらは、任意の発明の範囲、または特許請求され得る事項の範囲に対する限定として解釈されるべきではなく、特定の発明の特定の实装形態に特異的であり得る特徴の説明として解釈されるべきである。別

50

個の実施形態の文脈で本明細書に記載された特定の特徴を、単一の実施形態で組み合わせで実装することもできる。逆に、単一の実施形態の文脈で記載されている様々な特徴は、複数の実施形態で、別個に、または任意の好適な部分組み合わせで実装することもできる。さらに、特徴は、特定の組み合わせで作用するものとして上記に説明され、当初はそのように特許請求されることさえあるが、場合によっては、特許請求された組み合わせからの1つ以上の特徴が、その組み合わせから削除される可能性もあり、特許請求された組み合わせが、部分組み合わせまたは部分組み合わせの変形を対象とすることもある。

【0088】

同様に、動作が特定の順序で図面に描写され、請求の範囲に列挙されているが、これは、所望の結果を達成するために、かかる動作がその示された特定の順序、もしくは一連の順序で実行されるべきであること、または例証したすべての動作が実行されるべきであることを要求するものとして理解されるべきではない。特定の状況では、マルチタスクおよび並列処理が有利な場合がある。さらに、上記の実施形態における様々なシステムモジュールおよびコンポーネントの分離は、すべての実施形態においてかかる分離を必要とするものとして理解されるべきではなく、記載されたプログラムコンポーネントおよびシステムは、概して、単一のソフトウェア製品内に共に一体化され得るか、または複数のソフトウェア製品にパッケージ化され得ることを理解されたい。

10

【0089】

主題の特定の実施形態を説明してきた。他の実施形態は、以下の特許請求の範囲内に存在する。例えば、特許請求の範囲に列挙された動作は、異なる順序で実行することができ、望ましい結果を依然として達成することができる。一例として、添付の図に描写されたプロセスは、望ましい結果を達成するために、必ずしも示された特定の順序、または一連の順序を必要としない。場合によっては、マルチタスクおよび並列処理が有利となり得る。

20

【0090】

特許請求の範囲：

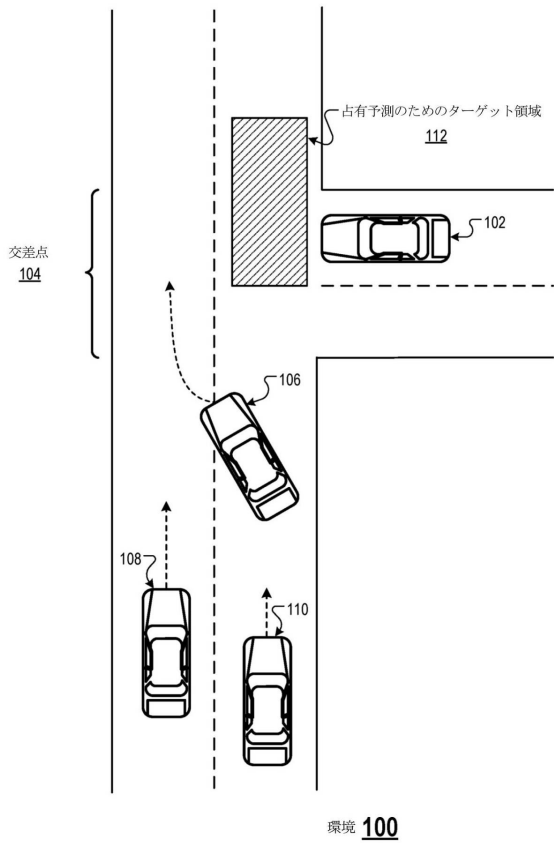
30

40

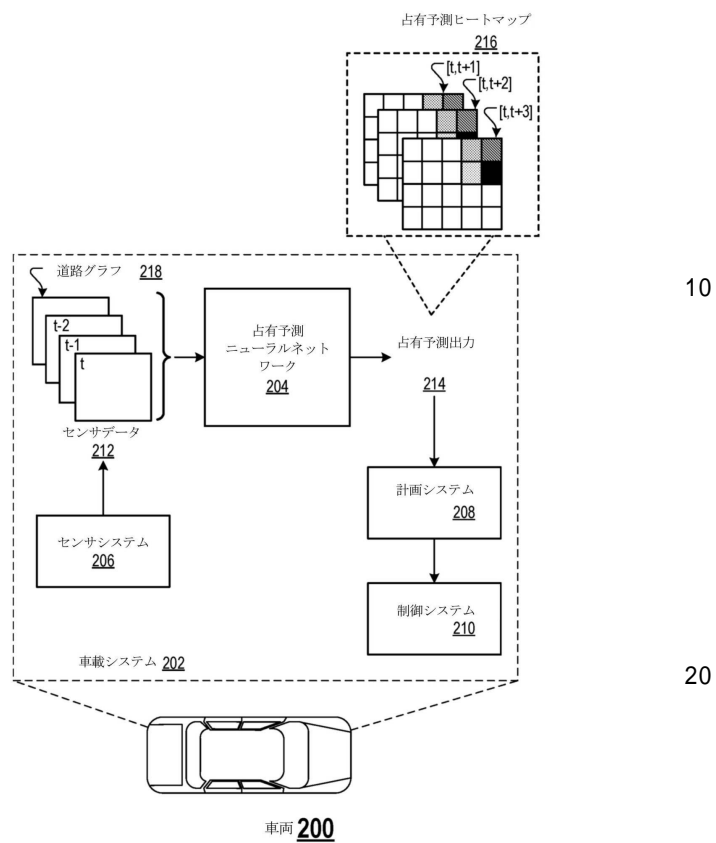
50

【 図面 】

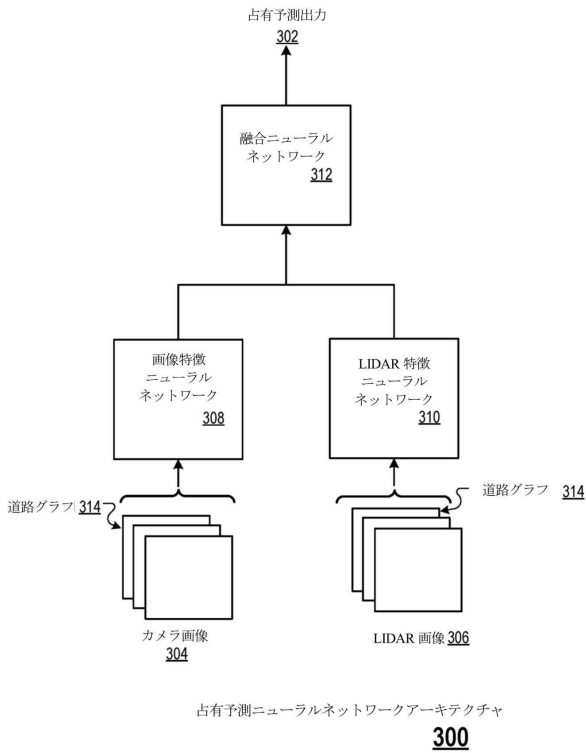
【 図 1 】



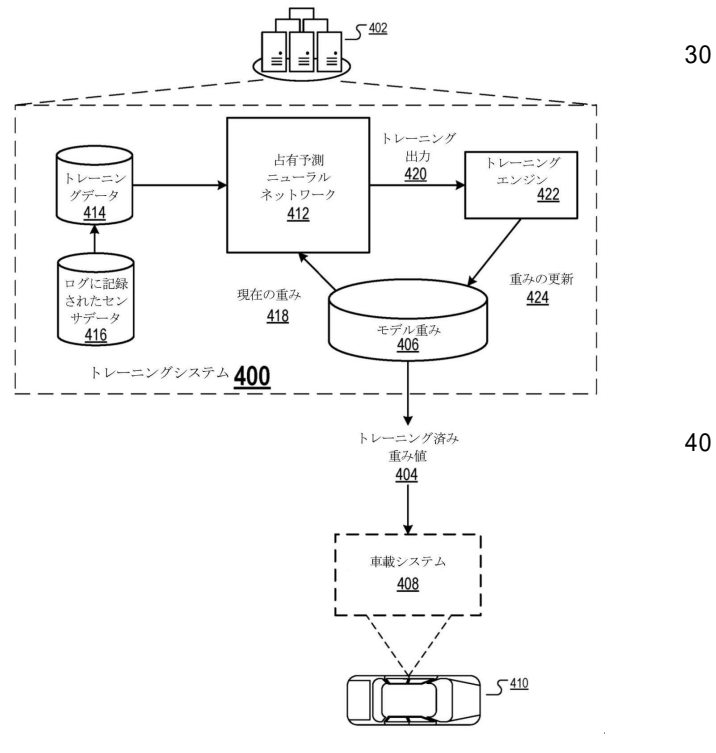
【 図 2 】



【 図 3 】



【 図 4 】



10

20

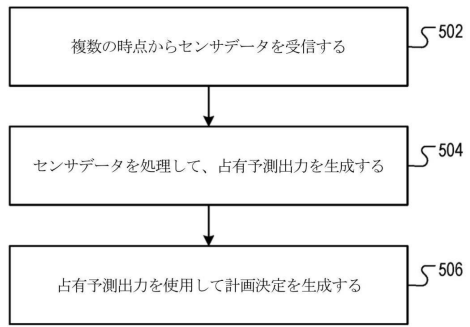
30

40

50

【 図 5 】

500 ↻



10

20

30

40

50

## フロントページの続き

4 0 8 6 サニーベール レドンド テラス 3 7 4

(72)発明者 チェン, ユ - ハン

アメリカ合衆国, カリフォルニア州 9 4 0 4 3, マウンテン ビュー, アンフィシアター パーク  
ウェイ 1 6 0 0

(72)発明者 アヤヴァチ, アルパー

アメリカ合衆国, カリフォルニア州 9 4 0 4 3, マウンテン ビュー, アンフィシアター パーク  
ウェイ 1 6 0 0

審査官 佐々木 佳祐

(56)参考文献 米国特許出願公開第 2 0 1 9 / 0 1 5 2 4 9 0 ( U S , A 1 )

特開 2 0 1 6 - 1 7 7 7 2 9 ( J P , A )

国際公開第 2 0 1 6 / 0 2 4 3 1 8 ( W O , A 1 )

(58)調査した分野 (Int.Cl., D B 名)

G 0 8 G 1 / 0 0 - 9 9 / 0 0

G 0 6 T 7 / 0 0