(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7683723号
(P7683723)

(45)発行日　令和7年5月27日(2025.5.27)

(24)登録日　令和7年5月19日(2025.5.19)

(51)国際特許分類

Ｇ０６Ｔ	*7/00*	*(2017.01)*	Ｇ０６Ｔ	7/00
Ｇ０６Ｎ	*3/08*	*(2023.01)*	Ｇ０６Ｎ	3/08
Ｇ０６Ｎ	*20/00*	*(2019.01)*	Ｇ０６Ｎ	20/00

ＦＩ

Ｇ０６Ｔ　7/00　３５０Ｂ

請求項の数　10　（全23頁）

(21)出願番号　特願2023-555935(P2023-555935)
(86)(22)出願日　令和3年10月26日(2021.10.26)
(86)国際出願番号　PCT/JP2021/039520
(87)国際公開番号　WO2023/073813
(87)国際公開日　令和5年5月4日(2023.5.4)
審査請求日　令和6年4月5日(2024.4.5)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100124811
弁理士　馬場 資博
(74)代理人　100088959
弁理士　境 廣巳
(74)代理人　100097157
弁理士　桂木 雄二
(74)代理人　100187724
弁理士　唐鎌 睦
(72)発明者　朴 君
東京都港区芝五丁目７番１号 日本電気株式会社内
審査官　藤原 敬利

最終頁に続く

(54)【発明の名称】　画像処理システム

(57)【特許請求の範囲】

【請求項１】

画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する学習部を含み、

前記学習済みモデルは、

前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、

前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、

前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、

前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、

を含む画像処理システム。

【請求項２】

前記第３のコンポーネントは、前記推論タスク毎に抽出された第２の特徴量のうちの１つを基準特徴量とし、前記基準特徴量以外の前記第２の特徴量のサイズを前記基準特徴量のサイズに合わせて変更し、前記基準特徴量のサイズに合わせてサイズを変更した後の前記基準特徴量以外の前記第２の特徴量と前記基準特徴量とを結合して前記第３の特徴量を生成し、前記推論タスク毎に、前記第３の特徴量のサイズを前記第４のコンポーネントの

入力サイズに合わせて変更して前記第４のコンポーネントへ出力する、
請求項１に記載の画像処理システム。
【請求項３】
前記第３のコンポーネントは、前記推論タスクに対応するサブコンポーネントを含み、
前記サブコンポーネントは、対応する前記推論タスクの前記第２の特徴量を基準特徴量とし、対応する前記推論タスク以外の前記第２の特徴量のサイズを前記基準特徴量のサイズに合わせて変更し、前記基準特徴量のサイズに合わせてサイズを変更した後の前記対応する前記推論タスク以外の前記第２の特徴量と前記基準特徴量とを結合して前記第３の特徴量を生成し、前記第４のコンポーネントへ出力する、
請求項１に記載の画像処理システム。
【請求項４】
前記学習部は、複数の学習段階に分けて前記学習済みモデルの学習を行い、
前記複数の学習段階は、少なくとも、
前記複数の推論タスクのうちの何れか１つを学習対象タスクとし、前記学習対象タスク以外の推論タスクに係る前記第２のコンポーネントおよび前記第４のコンポーネントと前記第１のコンポーネントのパラメータを固定して、前記学習対象タスクに係る前記第２のコンポーネントおよび前記第４のコンポーネントのパラメータを学習する第１の学習段階と、
前記第１のコンポーネントのパラメータを固定して、前記複数の推論タスクのそれぞれに係る前記第２のコンポーネントおよび前記第４のコンポーネントのパラメータを学習する第２の学習段階と、を含む、
請求項１乃至３の何れかに記載の画像処理システム。
【請求項５】
前記推論タスクに対応して設けられた第４のコンポーネントは、前記第３の特徴量を構成する複数の前記第２の特徴量のうち、対応する前記推論タスクの第２の特徴量の優先度合いを定める重みをそれ以外の第２の特徴量の優先度合いを定める重みより大きくする、
請求項１乃至４の何れかに記載の画像処理システム。
【請求項６】
学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を出力する推論部を含み、
前記学習済みモデルは、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、
前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、
前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、
を含む画像処理システム。
【請求項７】
コンピュータによる画像処理方法であって、
前記コンピュータが、画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成し、
前記生成では、前記コンピュータが、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させ

る、
画像処理方法。
【請求項８】
コンピュータによる画像処理方法であって、
前記コンピュータが、学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力し、
前記推定では、前記コンピュータが、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
画像処理方法。
【請求項９】
コンピュータに、画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する処理を行わせるためのプログラムであって、
前記生成では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラム。
【請求項１０】
コンピュータに、学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力する処理を行わせるためのプログラムであって、
前記推定では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、画像処理システム、画像処理方法、および、記録媒体に関する。
【背景技術】
【０００２】
単一の多層ニューラルネットワークＤＮＮ（Ｄｅｅｐ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ）で複数のタスクを同時に学習および推定する手法がある。この手法はマルチタスク学習と呼ばれる。マルチタスク学習は、タスク数に比例して増加する学習および推定時間を削減することができる。これにより、複数のタスクから得られる情報が必要な人物画像解析などのアプリケーションにおいてマルチタスク学習は有効な手法の一つになっている。
【０００３】
マルチタスク学習の一例が特許文献１に記載されている。特許文献１に記載の技術（以下、本発明に関連する技術と記す）では、ＤＮＮは、人物の顔が写っている画像から複数

のタスクに共通な特徴量x^Lを抽出する。次に、ＤＮＮは、特徴量x^Lから顔の表情を識別するタスクに固有な特徴量を抽出して推定結果y^cを出力すると共に、それと並行して、特徴量x^Lから顔領域の目や鼻の位置を推定するタスクに固有な特徴量を抽出して推定結果y^rを出力する。

【先行技術文献】

【特許文献】

【０００４】

【文献】特開２０１８－５５３７７号公報

【発明の概要】

【発明が解決しようとする課題】

【０００５】

しかしながら、本発明に関連する技術では、画像から全てのタスクに共通な特徴量を抽出し、この共通な特徴量からタスクに固有な特徴量を抽出して各タスクの推定結果を推定するように構成されている。そのため、或るタスクに固有な特徴量を他のタスクの推定に利用できないという課題がある。

【０００６】

本発明は、上述した課題、すなわち、複数のタスク間でタスク固有な特徴量を相互に利用できない、という課題を解決する画像処理システムを提供することにある。

【課題を解決するための手段】

【０００７】

本発明の一形態に係る画像処理システムは、

画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する学習部を含み、

前記学習済みモデルは、

前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、

前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、

前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、

前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、

を含むように構成されている。

【０００８】

本発明の他の形態に係る画像処理システムは、

学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を出力する推論部を含み、

前記学習済みモデルは、

前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、

前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、

前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、

前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、

を含むように構成されている。

【０００９】

本発明の他の形態に係る画像処理方法は、

画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成し、

前記生成では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、ように構成されている。
【００１０】
本発明の他の形態に係る画像処理方法は、
学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力し、
前記推定では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、ように構成されている。
【００１１】
本発明の他の形態に係るコンピュータ読み取り可能な記録媒体は、
コンピュータに、画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する処理を行わせるためのプログラムであって、
前記生成では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラムを記録するように構成されている。
【００１２】
本発明の他の形態に係るコンピュータ読み取り可能な記録媒体は、
コンピュータに、学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力する処理を行わせるためのプログラムであって、
前記推定では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラムを記録するように構成されている。
【発明の効果】
【００１３】
本発明は、上述したような構成を有することにより、複数のタスク間でタスク固有な特徴量を相互に利用することができる。このため、複数のタスクのそれぞれにおいて、当該タスク固有の特徴量と他のタスク固有の特徴量とを考慮した学習および推定が可能となる。
【図面の簡単な説明】
【００１４】
【図１】本発明の第１の実施形態に係る画像処理装置のブロック図である。

【図２】本発明の第１の実施形態に係る画像処理装置における学習フェーズの動作の一例を示すフローチャートである。
【図３】本発明の第１の実施形態に係る画像処理装置における推定フェーズの動作の一例を示すフローチャートである。
【図４】本発明の第１の実施形態で使用するモデルの一例を示す構成図である。
【図５】本発明の第１の実施形態で使用するモデルのコンポーネントＣＭ３の一例を示す構成図である。
【図６】本発明の第１の実施形態で使用するモデルのコンポーネントＣＭ３の他の例を示す構成図である。
【図７】本発明の第１の実施形態で使用するモデルの機械学習に用いられる訓練データのリストの一例を示す図である。
【図８】本発明の第１の実施形態に係る画像処理装置における学習部の学習処理の一例を示すフローチャートである。
【図９】本発明の第２の実施形態に係る画像処理装置のブロック図である。
【図１０】本発明の第３の実施形態に係る画像処理装置のブロック図である。
【発明を実施するための形態】
【００１５】
次に、本発明の実施の形態について、図面を参照して詳細に説明する。
［第１の実施の形態］
図１は、本発明の第１の実施形態に係る画像処理装置１０のブロック図である。この画像処理装置１０は、画像から互いに相違する複数の推論タスクを行うように構成されている。図１を参照すると、画像処理装置１０は、カメラＩ／Ｆ（インターフェース）部１１と通信Ｉ／Ｆ部１２と操作入力部１３と画面表示部１４と記憶部１５と演算処理部１６とを含んで構成されている。
【００１６】
カメラＩ／Ｆ部１１は、有線または無線により画像サーバ１７に接続され、画像サーバ１７と演算処理部１６との間でデータの送受信を行うように構成されている。画像サーバ１７は、有線または無線によりカメラ１８に接続され、カメラ１８で撮影されたそれぞれ撮影時刻が異なる複数の画像を過去一定期間分蓄積するように構成されている。カメラ１８は、例えば、数百万画素程度の画素容量を有するＣＣＤ（Ｃｈａｒｇｅ－Ｃｏｕｐｌｅｄ　Ｄｅｖｉｃｅ）イメージセンサやＣＭＯＳ（Ｃｏｍｐｌｅｍｅｎｔａｒｙ　ＭＯＳ）イメージセンサを備えたカラーカメラや白黒カメラであってよい。カメラ１８は、防犯・監視の目的のために多くの人が行きかう街頭、屋内などに設置されたカメラであってよい。或いはカメラ１８は、車などの移動体に搭載されて移動しながら同一或いは異なる撮影領域を撮影するカメラであってよい。カメラ１８は１台に限定されず、異なる場所から異なる撮影領域を撮影する複数台のカメラであってよい。
【００１７】
通信Ｉ／Ｆ部１２は、データ通信回路から構成され、有線または無線によって図示しない外部装置との間でデータ通信を行うように構成されている。操作入力部１３は、キーボードやマウスなどの操作入力装置から構成され、オペレータの操作を検出して演算処理部１６に出力するように構成されている。画面表示部１４は、ＬＣＤ（Ｌｉｑｕｉｄ　Ｃｒｙｓｔａｌ　Ｄｉｓｐｌａｙ）などの画面表示装置から構成され、演算処理部１６からの指示に応じて、各種情報を画面表示するように構成されている。
【００１８】
記憶部１５は、ハードディスクやメモリなどの記憶装置から構成され、演算処理部１６における各種処理に必要な処理情報およびプログラム１５１を記憶するように構成されている。プログラム１５１は、演算処理部１６に読み込まれて実行されることにより各種処理部を実現するプログラムであり、通信Ｉ／Ｆ部１２などのデータ入出力機能を介して図示しない外部装置や記録媒体から予め読み込まれて記憶部１５に保存される。記憶部１５に記憶される主な処理情報には、画像情報１５２、モデル１５３、および推定結果情報１

５４がある。
【００１９】
画像情報１５２は、カメラＩ／Ｆ部１１を通じて画像サーバ１７から取得されたカメラ１８のフレーム画像である。
【００２０】
モデル１５３は、カメラ１８のフレーム画像から互いに相違する複数の推論タスクを同時に学習および推定する機械学習モデルである。モデル１５３は、例えば、ＤＣＮＮ（Ｄｅｅｐ Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋ）を用いて構成してよい。本実施形態では、モデル１５３は、物体検出、ポーズ推定、セマンティックセグメンテーション推定の３つの推論タスクを行うようにパラメータが学習される。パラメータが学習されたモデルを学習済みのモデルと呼び、学習前のモデルと区別する。
【００２１】
物体検出は、画像内のクラスと物体位置を検出する。物体検出の結果は、クラス名、クラスの推定信頼度、物体位置を表すバウンディングボックス（以下、矩形と記す）を含む。検出するクラスは、例えば人物であってよい。但し、検出するクラスは人物に限定されず、動物や物であってもよい。
【００２２】
ポーズ推定は、画像内の人物の骨格情報を推定する。人物の骨格情報は、人体を構成する関節の位置を表す情報を含んでいる。関節は、首や肩などの関節のみならず、目や鼻などの顔のパーツも関節に含めてよい。ポーズ推定の結果は、関節名（関節ＩＤ）、関節の位置、関節の信頼度を含む。
【００２３】
セマンティックセグメンテーション推定は、画像内の各ピクセルのクラスを推定する。セマンティックセグメンテーション推定の結果は、各ピクセルのクラスを含む。推定するクラスは、物体検出で検出するクラスと同じである。
【００２４】
推定結果情報１５４は、学習済みのモデル１５３を用いて画像から推定した結果を表す情報である。推定結果情報１５４は、物体検出結果、ポーズ推定結果、および、セマンティックセグメンテーション推定結果を含む。
【００２５】
演算処理部１６は、ＭＰＵなどの１以上のプロセッサとその周辺回路を有し、記憶部１５からプログラム１５１を読み込んで実行することにより、上記ハードウェアとプログラム１５１とを協働させて各種処理部を実現するように構成されている。演算処理部１６で実現される主な処理部には、取得部１６１、学習部１６２、および、推定部１６３がある。
【００２６】
取得部１６１は、カメラＩ／Ｆ部１１を通じて画像サーバ１７から、カメラ１８で撮影された動画を構成するフレーム画像あるいはそれをダウンサンプリングしたフレーム画像を取得し、記憶部１５に画像情報１５２として保存するように構成されている。取得されたフレーム画像には、カメラＩＤおよび撮影時刻が付加されている。フレーム画像の撮影時刻は、フレーム毎に相違する。
【００２７】
学習部１６２は、訓練データを用いて、モデル１５３に上記３つの推論タスクを同時に学習させるように構成されている。即ち、学習部１６２は、画像から上記３つの推論タスクを行う学習済みモデル１５３を生成する。学習部２１は、上記生成では、モデル１５３に、画像から上記３つの推論タスクに共通な第１の特徴量を抽出させ、次に、推論タスク毎に、第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、次に、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、次に、推論タスク毎に、第３の特徴量から対応する推論タスクの推論結果を出力させる。
【００２８】
推定部１６３は、学習済みモデル１５３を用いて、画像から上記３つの推論タスクの推

論結果を推定して出力するように構成されている。推定部３１は、上記推定では、学習済みモデル１５３に、先ず、画像から上記３つの推論タスクに共通な第１の特徴量を抽出させ、次に、推論タスク毎に、第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、次に、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、次に、推論タスク毎に、第３の特徴量から対応する推論タスクの推論結果を出力させる。

【００２９】

次に、画像処理装置１０の動作を説明する。画像処理装置１０のフェーズは、学習フェーズと推定フェーズとに大別される。学習フェーズは、モデル１５３を機械学習するフェーズである。推定フェーズは、学習済みのモデル１５３を用いて、画像から上記３つの推論タスクの推論結果を推定し、出力するフェーズである。

【００３０】

図２は学習フェーズの動作の一例を示すフローチャートである。図２を参照すると、先ず、取得部１６１は、カメラＩ／Ｆ部１１を通じて画像サーバ１７からカメラ１８で撮影されたフレーム画像を取得し、記憶部１５に画像情報１５２として保存する（ステップＳ１）。次に、学習部１６２は、モデル１５３の機械学習に用いる訓練データを作成する（ステップＳ２）。次に、学習部１６２は、訓練データを用い、入力を画像とし、出力を上記３つの推論タスクの推定結果とするモデル１５３を機械学習させ、学習済みのモデル１５３を生成する（ステップＳ３）。

【００３１】

図３は推定フェーズの動作の一例を示すフローチャートである。図３を参照すると、先ず、取得部１６１は、カメラＩ／Ｆ部１１０を通じて画像サーバ１７からカメラ１８で撮影されたフレーム画像を取得し、記憶部１５に画像情報１５２として保存する（ステップＳ１１）。

【００３２】

次に、推定部１６３は、学習済みのモデル１５３を用いて、画像情報１５２に含まれるフレーム画像から上記３つの推論タスクの推定結果を同時に推定する（ステップＳ１２）。次に、推定部１６３は、推定した３つの推論タスクの推定結果を画面表示部１４に表示し、または／および、通信Ｉ／Ｆ部１２を通じて外部装置へ送信する（ステップＳ１３）。

【００３３】

続いて、モデル１５３と学習部１６２を詳細に説明する。

【００３４】

先ず、モデル１５３の詳細を説明する。

【００３５】

図４は、モデル１５３として利用できるマルチタスクモデルの一例を示す構成図である。この例のモデル１５３は、８個のコンポーネントＣＭから構成され、全体が１つの多層ニューラルネットワークになっている。

【００３６】

コンポーネントＣＭ１は、多層ニューラルネットワークの低層側に設けられ、画像を入力し、全てのタスクに共通な低次な特徴量ＦＭ１を抽出するように構成されている。コンポーネントＣＭ１は、バックボーンとも呼ばれる。コンポーネントＣＭ１によって抽出された特徴量ＦＭ１は、低次な特徴マップとも称する。コンポーネントＣＭ１は、１以上の畳み込み層を含んで構成してよい。例えば、コンポーネントＣＭ１は、ＳＳＤ（Ｓｉｎｇｌｅ Ｓｈｏｔ ＭｕｌｔｉＢｏｘ Ｄｅｔｅｃｔｏｒ）の構成要素であるＶＧＧ－１６を使用してよい。或いは、コンポーネントＣＭ１は、例えば、ＯｐｅｎＰｏｓｅの構成要素であるＶＧＧ－１９を使用してよい。或いは、コンポーネントＣＭ１は、例えば、ＳｅｇＮｅｔの構成要素であるエンコーダを使用してよい。或いは、コンポーネントＣＭ１は、例えば、ＳＳＤやＯｐｅｎＰｏｓｅやＳｅｇＮｅｔ以外のモデルのバックボーンを使用してよい。

【００３７】

コンポーネントＣＭ２－１は、コンポーネントＣＭ１から特徴量ＦＭ１を入力し、物体検出タスクに固有な高次の特徴量ＦＭ２－１を抽出するように構成されている。コンポーネントＣＭ２－１は、１以上の畳み込み層を含んで構成してよい。例えば、コンポーネントＣＭ２－１は、ＳＳＤの構成要素である特別な畳み込み層（Ｅｘｔｒａ　Ｆｅａｔｕｒｅ　Ｌａｙｅｒｓ）を使用してよい。但し、コンポーネントＣＭ２－１は上記に限定されず、ＳＳＤ以外の物体検出モデルにおいて物体検出タスクに固有な高次の特徴量を抽出する畳込み層を使用してよい。

【００３８】

コンポーネントＣＭ２－２は、コンポーネントＣＭ１から特徴量ＦＭ１を入力し、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２を抽出するように構成されている。コンポーネントＣＭ２－２は、１以上の畳み込み層を含んで構成してよい。例えば、コンポーネントＣＭ２－２は、ＯｐｅｎＰｏｓｅの構成要素である、キーポイントの位置を表すＰａｒｔ　Ｃｏｎｆｉｄｅｎｃｅ　Ｍａｐを生成する畳み込み層と、キーポイント間の関連度を表すＰａｒｔ　Ａｆｆｉｎｉｔｙ　Ｆｉｅｌｄｓを生成する畳み込み層と、生成されたＰａｒｔ　Ｃｏｎｆｉｄｅｎｃｅ　ＭａｐとＰａｒｔ　Ａｆｆｉｎｉｔｙ　Ｆｉｅｌｄｓと抽出元の特徴量ＦＭ１とを結合（ｃｏｎｃａｔｅｎａｔｅ）する層（これによって結合して得られた特徴マップを以下、ＯｐｅｎＰｏｓｅ特徴マップと記す）とを使用してよい。但し、コンポーネントＣＭ２－２は上記に限定されず、ＯｐｅｎＰｏｓｅ以外のポーズ推定モデルにおいてポーズ推定タスクに固有な高次の特徴量を抽出する畳込み層を使用してよい。

【００３９】

コンポーネントＣＭ２－３は、コンポーネントＣＭ１から特徴量ＦＭ１を入力し、セマンティックセグメンテーション推定タスクに固有な高次の特徴量ＦＭ２－３を抽出するように構成されている。コンポーネントＣＭ２－３は、１以上の畳み込み層を含んで構成してよい。例えば、コンポーネントＣＭ２－３は、ＳｅｇＮｅｔの構成要素であるデコーダを使用してよい。但し、コンポーネントＣＭ２－３は上記に限定されず、ＳｅｇＮｅｔ以外のセマンティックセグメンテーション推定モデルにおいてセマンティックセグメンテーション推定タスクに固有な高次の特徴量を抽出する畳込み層を使用してよい。

【００４０】

コンポーネントＣＭ３は、コンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ－２－３から特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３を入力し、これら３つの特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３を結合（ｃｏｎｃａｔｅｎａｔｅ）して得られる特徴量ＦＭ３－１、ＦＭ３－２、ＦＭ３－３を生成するように構成されている。

【００４１】

図５は、コンポーネントＣＭ３の一例を示す構成図である。この例のコンポーネントＣＭ３は、リサイズ部ＣＭ３－１と結合部ＣＭ３－２とリサイズ部ＣＭ３－３とを含んで構成されている。

【００４２】

リサイズ部ＣＭ３－１は、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３を結合し得るように、それらのサイズを合わせるように構成されている。リサイズ部ＣＭ３－１は、３つの特徴量のうちの何れか１つの特徴量を基準特徴量とし、基準特徴量のサイズに合わせて、残り２つの特徴量のサイズを変更する。例えば、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３のサイズを、それぞれ、３８×３８、７０×７０、２４０×３２０とし、基準特徴量を特徴量ＦＭ２－１とする。この場合、リサイズ部ＣＭ３－１は、特徴量ＦＭ２－２のサイズを７０×７０から３８×３８に変更した特徴量ＦＭ２－２’を生成して出力する。また、リサイズ部ＣＭ３－１は、特徴量ＦＭ２－３のサイズを２４０×３２０から３８×３８に変更した特徴量ＦＭ２－３’を生成して出力する。また、リサイズ部ＣＭ３－１は、特徴量ＦＭ２－１についてはサイズを変更せず、特徴量ＦＭ２－１そのものを特徴量ＦＭ２－１’として出力する。

【００４３】

結合部ＣＭ３－２は、リサイズ部ＣＭ３－１から特徴量ＦＭ２－１’、ＦＭ２－２’、Ｆ

Ｍ２－３’を入力し、これらを結合して得られる特徴量ＦＭ３を生成して出力する。例えば、結合部ＣＭ３－２は、それぞれが３８×３８のサイズである特徴量ＦＭ２－１’、ＦＭ２－２’、ＦＭ２－３’を入力し、３８×３８×３のサイズの特徴量ＦＭ３を生成して出力する。このように、特徴量の結合により、チャネル数（次元数）が増加する。

【００４４】

リサイズ部ＣＭ３－３は、結合部ＣＭ３－２から特徴量ＦＭ３を入力し、各タスクに応じたサイズに変更した特徴量ＦＭ３－１、ＦＭ３－２、ＦＭ３－３を生成して出力する。例えば、コンポーネントＣＭ４－１、ＣＭ４－２、ＣＭ４－３の入力サイズをそれぞれ、３８×３８×３、７０×７０×３、２４０×３２０×３とする。この場合、リサイズ部ＣＭ３－３は、特徴量ＦＭ３のサイズを３８×３８×３から７０×７０×３に変更した特徴量ＦＭ３－２を生成して、コンポーネントＣＭ４－２に出力する。また、リサイズ部ＣＭ３－３は、特徴量ＦＭ３のサイズを３８×３８×３から２４０×３２０×３に変更した特徴量ＦＭ３－３を生成して、コンポーネントＣＭ４－３に出力する。また、リサイズ部ＣＭ３－３は、３８×３８×３のサイズの特徴量ＦＭ３そのものを特徴量ＦＭ３－１として、コンポーネントＣＭ４－１に出力する。

【００４５】

図６は、コンポーネントＣＭ３の他の一例を示す構成図である。この例のコンポーネントＣＭ３は、３つのサブコンポーネントＣＭ３Ａ、ＣＭ３Ｂ、ＣＭ３Ｃを含んで構成されている。

【００４６】

サブコンポーネントＣＭ３Ａは、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３から物体検出タスクのコンポーネントＣＭ４－１のための特徴量ＦＭ３－１を生成して出力するように構成されている。サブコンポーネントＣＭ３Ａは、特徴量ＦＭ２－１のサイズに合わせて特徴量ＦＭ２－２、ＦＭ２－３のサイズを変更して得られる特徴量ＦＭ２－２’、ＦＭ２－３’を生成して出力するリサイズ部ＣＭ３Ａ－１と、特徴量ＦＭ２－１、ＦＭ２－２’、ＦＭ２－３’の３つを結合して得られる特徴量ＦＭ３－１を生成して出力する結合部ＣＭ３Ａ－２とを含んで構成されている。例えば、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３のサイズを、それぞれ、３８×３８、７０×７０、２４０×３２０とし、コンポーネントＣＭ４－１の入力サイズを３８×３８×３とする。この場合、リサイズ部ＣＭ３Ａ－１は、特徴量ＦＭ２－２のサイズを７０×７０から３８×３８に変更した特徴量ＦＭ２－２’を生成して出力し、特徴量ＦＭ２－３のサイズを２４０×３２０から３８×３８に変更した特徴量ＦＭ２－３’を生成して出力する。結合部ＣＭ３Ａ－２は、同じ３８×３８のサイズの特徴量ＦＭ２－１、ＦＭ２－２’ ＦＭ２－３’を結合して、３８×３８×３のサイズの特徴量ＦＭ３－１を生成して出力する。これにより、結合に伴う特徴量ＦＭ２－１の劣化を抑えることができる。

【００４７】

サブコンポーネントＣＭ３Ｂは、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３からポーズ推定タスクのコンポーネントＣＭ４－２のための特徴量ＦＭ３－２を生成して出力するように構成されている。サブコンポーネントＣＭ３Ｂは、特徴量ＦＭ２－２のサイズに合わせて特徴量ＦＭ２－１、ＦＭ２－３のサイズを変更して得られる特徴量ＦＭ２－１’、ＦＭ２－３’を生成して出力するリサイズ部ＣＭ３Ｂ－１と、特徴量ＦＭ２－１’、ＦＭ２－２、ＦＭ２－３’の３つを結合して得られる特徴量ＦＭ３－２を生成して出力する結合部ＣＭ３Ｂ－２とを含んで構成されている。例えば、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３のサイズを、それぞれ、３８×３８、７０×７０、２４０×３２０とし、コンポーネントＣＭ４－２の入力サイズを７０×７０×３とする。この場合、リサイズ部ＣＭ３Ｂ－１は、特徴量ＦＭ２－１のサイズを３８×３８から７０×７０に変更した特徴量ＦＭ２－１’を生成して出力し、特徴量ＦＭ２－３のサイズを２４０×３２０から７０×７０に変更した特徴量ＦＭ２－３’を生成して出力する。結合部ＣＭ３Ｂ－２は、同じ７０×７０のサイズの特徴量ＦＭ２－１’、ＦＭ２－２ ＦＭ２－３’を結合して、７０×７０×３のサイズの特徴量ＦＭ３－２を生成して出力する。これにより、結合に伴う特徴量ＦＭ２－２の劣

化を抑えることができる。
【００４８】
サブコンポーネントＣＭ３Ｃは、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３からセマンティックセグメンテーション推定タスクのコンポーネントＣＭ４－３のための特徴量ＦＭ３－３を生成して出力するように構成されている。サブコンポーネントＣＭ３Ｃは、特徴量ＦＭ２－３のサイズに合わせて特徴量ＦＭ２－１、ＦＭ２－２のサイズを変更して得られる特徴量ＦＭ２－１’、ＦＭ２－２’を生成して出力するリサイズ部ＣＭ３Ｃ－１と、特徴量ＦＭ２－１’、ＦＭ２－２’、ＦＭ２－３の３つを結合して得られる特徴量ＦＭ３－３を生成して出力する結合部ＣＭ３Ｃ－２とを含んで構成されている。例えば、特徴量ＦＭ２－１、ＦＭ２－２、ＦＭ２－３のサイズを、それぞれ、３８×３８、７０×７０、２４０×３２０とし、コンポーネントＣＭ４－３の入力サイズを２４０×３２０×３とする。この場合、リサイズ部ＣＭ３Ｃ－１は、特徴量ＦＭ２－１のサイズを３８×３８から２４０×２４０に変更した特徴量ＦＭ２－１’を生成して出力し、特徴量ＦＭ２－２のサイズを７０×７０から２４０×３２０に変更した特徴量ＦＭ２－２’を生成して出力する。結合部ＣＭ３Ｃ－２は、同じ２４０×３２０のサイズの特徴量ＦＭ２－１’、ＦＭ２－２’ ＦＭ２－３を結合して、２４０×３２０×３のサイズの特徴量ＦＭ３－３を生成して出力する。これにより、結合に伴う特徴量ＦＭ２－３の劣化を抑えることができる。
【００４９】
再び図４を参照すると、コンポーネントＣＭ４－１は、コンポーネントＣＭ３から特徴量ＦＭ３－１を入力し、特徴量ＦＭ３－１から物体検出タスクの推定結果ＥＲ１を推定して出力するように構成されている。特徴量ＦＭ３－１は、物体検出タスクに固有な高次の特徴量ＦＭ２－１だけでなく、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２とセマンティックセグメンテーション推定に固有な高次の特徴量ＦＭ２－３とを含んでいる。そのため、コンポーネントＣＭ４－１は、それら３つの高次の特徴量を考慮した学習および推定が可能になる。コンポーネントＣＭ４－１は、例えば、ＳＳＤを構成する特別な畳み込み層につながる出力層（Ｄｅｔｅｃｔｉｏｎｓ：８７３２ ｐｅｒ Ｃｌａｓｓ、Ｎｏｎ－Ｍａｘｉｍｕｍ Ｓｕｐｐｒｅｓｓｉｏｎ）を使用してよい。
【００５０】
ここで、コンポーネントＣＭ４－１は、物体検出タスクに固有な高次の特徴量ＦＭ２－１の優先度合いを定める重みをそれ以外の第２の特徴量の優先度合いを定める重みより大きく設定してよい。例えば、コンポーネントＣＭ４－１は、物体検出タスクに固有な高次の特徴量ＦＭ２－１の優先度合いを定める重みを０．５、それ以外の第２の特徴量の優先度合いを定める重みを０．２５としてよい。このように特徴量ＦＭ２－１に相対的に大きな重みを与えることにより、３つの高次の特徴量を考慮した学習および推定を可能にしつつ、物体検出タスクに固有な高次の特徴量ＦＭ２－１の重要度を上げることができる。
【００５１】
また、コンポーネントＣＭ４－１は、入力された特徴量ＦＭ３－１に対して１×１の畳み込み（Ｃｈａｎｎｅｌ－Ｗｉｓｅ Ｃｏｎｖｏｌｕｔｉｏｎ）を行うことにより、高次の特徴量の次元数を、例えば、３８×３８×３から３８×３８×１に削減してよい。これにより、コンポーネントＣＭ４－１として、ＳＳＤなどの既存モデルにおける、高次の特徴量から推定結果を推定して出力するネットワーク部分をそのまま利用することができる。
【００５２】
コンポーネントＣＭ４－２は、コンポーネントＣＭ３から特徴量ＦＭ３－２を入力し、特徴量ＦＭ３－２からポーズ推定タスクの推定結果ＥＲ２を推定して出力するように構成されている。特徴量ＦＭ３－２は、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２だけでなく、物体検出タスクに固有な高次の特徴量ＦＭ２－１とセマンティックセグメンテーション推定に固有な高次の特徴量ＦＭ２－３とが含まれている。そのため、コンポーネントＣＭ４－２は、それら３つの高次の特徴量を考慮した学習および推定が可能になる。コンポーネントＣＭ４－２は、例えば、ＯｐｅｎＰｏｓｅの構成要素である、ＯｐｅｎＰｏｓｅ特徴マップからポーズ推定結果を推定するネットワーク部分を用いてよい。

【００５３】
ここで、コンポーネントＣＭ４－２は、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２の優先度合いを定める重みをそれ以外の第２の特徴量の優先度合いを定める重みより大きく設定してよい。例えば、コンポーネントＣＭ４－２は、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２の優先度合いを定める重みを０．５、それ以外の第２の特徴量の優先度合いを定める重みを０．２５としてよい。このように特徴量ＦＭ２－２に相対的に大きな重みを与えることにより、３つの高次の特徴量を考慮した学習および推定を可能にしつつ、ポーズ推定タスクに固有な高次の特徴量ＦＭ２－２の重要度を上げることができる。
【００５４】
また、コンポーネントＣＭ４－２は、入力された特徴量ＦＭ３－２に対して１×１の畳み込み（Ｃｈａｎｎｅｌ－Ｗｉｓｅ　Ｃｏｎｖｏｌｕｔｉｏｎ）を行うことにより、高次の特徴量の次元数を、例えば、７０×７０×３から７０×７０×１に削減してよい。これにより、コンポーネントＣＭ４－２として、ＯｐｅｎＰｏｓｅなどの既存モデルにおける、高次の特徴量から推定結果を推定して出力するネットワーク部分をそのまま利用することができる。
【００５５】
コンポーネントＣＭ４－３は、コンポーネントＣＭ３から特徴量ＦＭ３－３を入力し、特徴量ＦＭ３－３からセマンティックセグメンテーション推定タスクの推定結果ＥＲ３を推定して出力するように構成されている。特徴量ＦＭ３－３は、セマンティックセグメンテーション推定タスクに固有な高次の特徴量ＦＭ２－３だけでなく、物体検出タスクに固有な高次の特徴量ＦＭ２－１とポーズ推定に固有な高次の特徴量ＦＭ２－２とが含まれている。そのため、コンポーネントＣＭ４－３は、それら３つの高次の特徴量を考慮した学習および推定が可能になる。コンポーネントＣＭ４－３は、例えば、ＳｅｇＮｅｔの構成要素であるソフトマックス層などを用いてよい。
【００５６】
ここで、コンポーネントＣＭ４－３は、セマンティックセグメンテーション推定タスクに固有な高次の特徴量ＦＭ２－３の優先度合いを定める重みをそれ以外の第２の特徴量の優先度合いを定める重みより大きく設定してよい。例えば、コンポーネントＣＭ４－３は、セマンティックセグメンテーション推定タスクに固有な高次の特徴量ＦＭ２－３の優先度合いを定める重みを０．５、それ以外の第２の特徴量の優先度合いを定める重みを０．２５としてよい。このように特徴量ＦＭ２－３に相対的に大きな重みを与えることにより、３つの高次の特徴量を考慮した学習および推定を可能にしつつ、セマンティックセグメンテーション推定タスクに固有な高次の特徴量ＦＭ２－３の重要度を上げることができる。
【００５７】
また、コンポーネントＣＭ４－３は、入力された特徴量ＦＭ３－３に対して１×１の畳み込み（Ｃｈａｎｎｅｌ－Ｗｉｓｅ　Ｃｏｎｖｏｌｕｔｉｏｎ）を行うことにより、高次の特徴量の次元数を、例えば、２４０×３２０×３から２４０×３２０×１に削減してよい。これにより、コンポーネントＣＭ４－３として、ＳｅｇＮｅｔなどの既存モデルにおける、高次の特徴量から推定結果を推定して出力するネットワーク部分をそのまま利用することができる。
【００５８】
続いて、学習部１６２の詳細を説明する。
【００５９】
先ず、モデル１５３の機械学習に用いられる訓練データについて説明する。
【００６０】
図７は、モデル１５３の機械学習に用いられる訓練データのリストの一例を示す。図７を参照すると、このリストには、合計ｎ個の訓練データが登録されている。個々の訓練データは、訓練データを一意に識別するＩＤ、画像、物体検出ラベル、ポーズ推定ラベル、セマンティックセグメンテーション推定ラベルの各項目から構成されている。

【００６１】
画像の項目には、カメラ１８で撮影されたフレーム画像が設定される。物体検出ラベルの項目には、ラベルの有無と、ラベルが有る場合には、ラベル情報である画像中に存在する人物などのクラスとその位置情報（矩形情報）が設定される。ポーズ推定ラベルの項目には、ラベルの有無と、ラベルが有る場合には、画像中に存在する関節の関節名（関節ＩＤ）とその位置情報が設定される。セマンティックセグメンテーション推定ラベルの項目には、ラベルの有無と、ラベルが有る場合には、画像の各ピクセルのクラスが設定される。このように、訓練データ群の中には、３つのラベル（物体検出ラベル、ポーズ推定ラベル、セマンティックセグメンテーション推定ラベル）の項目の全てにラベル情報が設定されたもの以外に、一部のラベルの項目だけにラベル情報が設定されたものが含まれていてよい。
【００６２】
上述したような訓練データは、例えば、ユーザとの間の対話的処理によって作成されてよい。例えば、学習部１６２は、取得部１６１によって取得されたカメラ１８の画像を画面表示部１４に表示し、操作入力部１３を通じてユーザから当該画像のラベル情報を受け付ける。そして、学習部１６２は、表示した画像と受け付けたラベル情報との組を１つの訓練データとして作成する。学習部１６２は、同様の方法により、必要十分な数の訓練データを作成する。但し、訓練データの作成方法は上記に限定されない。
【００６３】
次に、訓練データを用いて、学習部１６２がモデル１５３を学習する方法について説明する。
【００６４】
図８は、学習部１６２の学習処理の一例を示すフローチャートである。この例の学習処理は、図４に示される構成のモデル１５３を学習対象モデルとする。また、この例の学習処理は、モデル１５３全体を一気に学習するのではなく、学習するネットワーク部分を徐々に拡大しながら学習を行う。これにより、安定した学習が行える。具体的には、以下の４つの学習段階を経る。
【００６５】
（１）学習段階１
学習段階１では、学習部１６２は、物体検出に係る深い層のネットワーク部分であるコンポーネントＣＭ２－１およびＣＭ４－１だけを学習する。このとき、バックボーンであるコンポーネントＣＭ－１、ポーズ推定に係る深い層のネットワーク部分であるコンポーネントＣＭ２－２およびＣＭ４－２、並びに、セマンティックセグメンテーション推定に係る深い層のネットワーク部分であるコンポーネントＣＭ２－３およびＣＭ４－３のパラメータは固定しておく。
（２）学習段階２
学習段階２では、学習部１６２は、物体検出およびポーズ推定に係る深い層のネットワーク部分であるコンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ４－１、および、ＣＭ４－２だけを学習する。このとき、バックボーンであるコンポーネントＣＭ－１、セマンティックセグメンテーション推定に係る深い層のネットワーク部分であるコンポーネントＣＭ２－３およびＣＭ４－３のパラメータは固定しておく。
（３）学習段階３
学習段階３では、学習部１６２は、全ての推論タスク、すなわち物体検出、ポーズ推定およびセマンティックセグメンテーション推定に係る深い層のネットワーク部分であるコンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ２－３、ＣＭ４－１、ＣＭ４－２、およびＣＭ４－３だけを学習する。このとき、バックボーンであるコンポーネントＣＭ－１のパラメータは固定しておく。
（４）学習段階４
学習段階４では、学習部１６２は、モデル全体、すなわちバックボーンであるコンポーネントＣＭ－１、物体検出、ポーズ推定およびセマンティックセグメンテーション推定に

係る深い層のネットワーク部分であるコンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ２－３、ＣＭ４－１、ＣＭ４－２、およびＣＭ４－３を学習する。

【００６６】

図８を参照すると、学習部１６２は、モデル１５３の機械学習に用いられる訓練データ群から、各学習段階で使用する訓練データ群を作成する（ステップＳ２１）。

【００６７】

例えば、学習部１６２は、ステップＳ２１において、図７で説明したような訓練データのリストから、学習段階３で使用する訓練データ群および学習段階４で使用する訓練データ群を、それぞれ必要な数だけ作成する。学習段階３および学習段階４では、３つのラベル（物体検出ラベル、ポーズ推定ラベル、セマンティックセグメンテーション推定ラベル）の項目の全てにラベル情報が設定された訓練データが必要である。そのため、学習部１６２は、そのような条件を満たす訓練データをリストから抽出することにより、学習段階３で使用する訓練データ群および学習段階４で使用する訓練データ群を作成する。

【００６８】

また、学習部１６２は、ステップＳ２１において、リストの残りの訓練データ群から、学習段階２で使用する訓練データ群を作成する。学習段階２では、物体検出ラベルおよびポーズ推定ラベルの項目にラベル情報が設定された訓練データ（セマンティックセグメンテーション推定ラベル情報の有無は不問である）が必要である。そのため、学習部１６２は、そのような条件を満たす訓練データをリストから抽出することにより、学習段階２で使用する訓練データ群を作成する。

【００６９】

また、学習部１６２は、ステップＳ２１において、リストの残りの訓練データ群から、学習段階１で使用する訓練データ群を作成する。学習段階１では、物体検出ラベルの項目にラベル情報が設定された訓練データ（ポーズ推定ラベル情報およびセマンティックセグメンテーション推定ラベル情報の有無は不問である）が必要である。そのため、学習部１６２は、そのような条件を満たす訓練データをリストから抽出することにより、学習段階１で使用する訓練データ群を作成する。

【００７０】

次に、学習部１６２は、学習段階１、学習段階２、学習段階３、学習段階４の順に、それぞれ所定の終了条件が成立するまで、各段階の学習を行う（ステップＳ２２～Ｓ２５）。各段階の学習では、訓練データに含まれる画像をモデル１５３に入力したときにモデル１５３の出力として得られる推論タスクの推論結果と訓練データに含まれるラベル情報との誤差を、予め与えられた損失関数を用いて算出する。損失関数は、物体検出タスク、ポーズ推定タスク、セマンティックセグメンテーション推定タスク毎に存在する。物体検出タスクの損失関数をＬ１、ポーズ推定タスクの損失関数をＬ２、セマンティックセグメンテーション推定タスクの損失関数をＬ３とそれぞれ表記する。

【００７１】

学習段階１では、損失関数Ｌ１で算出した損失を最小化するようにモデル１５３のコンポーネントＣＭ２－１、ＣＭ４－１のパラメータを学習する。学習段階２では、損失関数Ｌ１で算出した損失と損失関数Ｌ２で算出した損失の総和（例えば重み付き和）を最小化するようにモデル１５３のコンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ４－１、ＣＭ４－２のパラメータを学習する。学習段階３では、損失関数Ｌ１で算出した損失と損失関数Ｌ２で算出した損失と損失関数Ｌ３で算出した損失の総和（例えば重み付き和）を最小化するようにモデル１５３のコンポーネントＣＭ２－１、ＣＭ２－２、ＣＭ２－３、ＣＭ４－１、ＣＭ４－２、ＣＭ４－３のパラメータを学習する。学習段階４では、損失関数Ｌ１で算出した損失と損失関数Ｌ２で算出した損失と損失関数Ｌ３で算出した損失の総和（例えば重み付き和）を最小化するようにモデル１５３のコンポーネントＣＭ１、ＣＭ２－１、ＣＭ２－２、ＣＭ２－３、ＣＭ４－１、ＣＭ４－２、ＣＭ４－３のパラメータを学習する。各学習では、例えば、勾配降下法と誤差逆伝搬法を用いてよい。

【００７２】

以上、訓練データを用いてモデル１５３を学習する方法の例について説明した。しかし、本発明に適用可能な学習方法は以上の例に限定されない。例えば、次のような学習方法であってもよい。即ち、最初に、物体検出に係るコンポーネントＣＭ２－１、ＣＭ４－１だけを学習する（他のコンポーネントＣＭ１、ＣＭ２－２、ＣＭ２－３、ＣＭ４－２、ＣＭ４－３のパラメータは固定する）。次に、ポーズ推定に係るコンポーネントＣＭ２－２、ＣＭ４－２だけを学習する（他のコンポーネントＣＭ１、ＣＭ２－１、ＣＭ２－３、ＣＭ４－１、ＣＭ４－３のパラメータは固定する）。次に、セマンティックセグメンテーション推定に係るコンポーネントＣＭ２－３、ＣＭ４－３だけを学習する（他のコンポーネントＣＭ１、ＣＭ２－１、ＣＭ２－３、ＣＭ４－１、ＣＭ４－３のパラメータは固定する）。次に、全ての推論タスクに係るコンポーネントＣＭ２－１～ＣＭ２－３、ＣＭ３－１～ＣＭ３－３）だけを学習する（コンポーネントＣＭ１のパラメータは固定する。次に、モデル全体のコンポーネントＣＭ１、ＣＭ２－１～ＣＭ２－３、ＣＭ４－１～ＣＭ４－３を学習する。

【００７３】

以上説明したように、本実施形態に係る画像処理装置１０によれば、複数のタスク間でタスク固有な高次の特徴量を相互に利用することができる。このため、複数のタスクのそれぞれにおいて、当該タスク固有の高次な特徴量と他のタスク固有の高次な特徴量とを考慮した学習および推定が可能となる。

【００７４】

続いて、本実施形態の変形例について説明する。

【００７５】

＜変形例１＞

上記実施形態では、モデル１５３は、セマンティックセグメンテーション推定を行うように構成されていた。しかし、モデル１５３は、セマンティックセグメンテーション推定の代わりにインスタントセマンティックセグメンテーション推定を行うように構成されていてよい。この場合、例えば図４に示されるマルチタスクモデル１５３のコンポーネントＣＭ３とコンポーネントＣＭ４－３との間に、特徴量ＦＭ３－３から物体検出を行うコンポーネントを追加し、コンポーネントＣＭ４－３は物体検出された個々のクラスの矩形毎にピクセル単位でクラスを推定するように構成してよい。

【００７６】

＜変形例２＞

上記実施形態では、モデル１５３は、物体検出、ポーズ推定、セマンティックセグメンテーション推定の３つの推論タスクを行うように構成されていた。しかし、モデル１５３は、物体検出、ポーズ推定、および、セマンティックセグメンテーション推定のうちの何れか２つの推論タスクのみを行うように構成してよい。或いは、モデル１５３が行う推論タスクは、物体検出、ポーズ推定、セマンティックセグメンテーション推定に限定されず、それら以外のタスクであってもよい。

【００７７】

［第２の実施の形態］

図９は、本発明の第２の実施形態に係る画像処理システム２０のブロック図である。図９を参照すると、画像処理システム２０は、学習部２１と学習済みモデル２２とを備えている。

【００７８】

学習部２１は、画像から互いに相違する複数の推論タスクを行う学習済みモデル２２を生成するように構成されている。学習部２１は、例えば図１の学習部１６２と同様に構成することができるが、それに限定されない。

【００７９】

学習済みモデル２２は、上記画像から上記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、上記推論タスクに対応して設けられ、上記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、上記

推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、上記推論タスクに対応して設けられ、上記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、を含むように構成されている。

【００８０】

上述のように構成された画像処理システム２０は、以下のように動作する。即ち、学習部２１は、画像から互いに相違する複数の推論タスクを行う学習済みモデル２２を生成する。学習部２１は、上記生成では、学習済みモデル２２に、画像から複数の推論タスクに共通な第１の特徴量を抽出させ、次に、推論タスク毎に、第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、次に、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、次に、推論タスク毎に、第３の特徴量から対応する推論タスクの推論結果を出力させる。

【００８１】

以上のように構成され動作する画像処理システム２０によれば、複数の推論タスク間でタスク固有な特徴量を相互に利用することができる。その理由は、画像処理システム２０は、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成し、第３の特徴量から対応する推論タスクの推論結果を出力するように構成されているためである。このため、複数の推論タスクのそれぞれにおいて、当該タスク固有の特徴量と他のタスク固有の特徴量とを考慮した学習および推定が可能となる。

【００８２】

［第３の実施の形態］

図１０は、本発明の第３の実施形態に係る画像処理システム３０のブロック図である。図１０を参照すると、画像処理システム３０は、推定部３１と学習済みモデル３２とを備えている。

【００８３】

推定部３１は、学習済みモデル３２を用いて、画像から互いに相違する複数の推論タスクの推論結果を出力するように構成されている。推定部３１は、例えば図１の推定部１６３と同様に構成することができるが、それに限定されない。

【００８４】

学習済みモデル３２は、上記画像から上記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、上記推論タスクに対応して設けられ、上記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、上記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、上記推論タスクに対応して設けられ、上記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、を含むように構成されている。

【００８５】

上述のように構成された画像処理システム３０は、以下のように動作する。即ち、推定部３１は、学習済みモデル３２を用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力する。推定部３１は、上記推定では、学習済みモデル３２に、先ず、画像から複数の推論タスクに共通な第１の特徴量を抽出させ、次に、推論タスク毎に、第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、次に、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、次に、推論タスク毎に、第３の特徴量から対応する推論タスクの推論結果を出力させる。

【００８６】

以上のように構成され動作する画像処理システム３０によれば、複数の推論タスク間でタスク固有な特徴量を相互に利用することができる。その理由は、画像処理システム３０は、推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成し、第３の特徴量から対応する推論タスクの推論結果を出力するように構成されているためである。このため、複数の推論タスクのそれぞれにおいて、当該タスク固有の特徴量と他のタスク固有の特徴量とを考慮した学習および推定が可能となる。

【００８７】

以上、上記各実施形態を参照して本発明を説明したが、本発明は、上述した実施形態に限定されるものではない。本発明の構成や詳細には、本発明の範囲内で当業者が理解しうる様々な変更をすることができる。

【産業上の利用可能性】

【００８８】

本発明は、カメラ画像などの画像から、物体検出、ポーズ推定、セマンティックセグメンテーション推定などの複数の推論タスクを行う分野全般に利用できる。

【００８９】

上記の実施形態の一部又は全部は、以下の付記のようにも記載され得るが、以下には限られない。

［付記１］

画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する学習部を含み、

前記学習済みモデルは、

前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、

前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、

前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、

前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、

を含む画像処理システム。

［付記２］

前記第３のコンポーネントは、前記複数の第２の特徴量のうちの１つを基準特徴量とし、前記基準特徴量以外の前記第２の特徴量のサイズを前記基準特徴量のサイズに合わせて変更し、前記サイズ変更後の前記基準特徴量以外の前記第２の特徴量と前記基準特徴量とを結合して前記第３の特徴量を生成し、前記推論タスク毎に、前記第３の特徴量のサイズを前記第４のコンポーネントの入力サイズに合わせて変更して前記第４のコンポーネントへ出力する、

付記１に記載の画像処理システム。

［付記３］

前記第３のコンポーネントは、前記推論タスクに対応するサブコンポーネントを含み、前記サブコンポーネントは、対応する前記推論タスクの前記第２の特徴量を基準特徴量とし、対応する前記推論タスク以外の前記第２の特徴量のサイズを前記基準特徴量のサイズに合わせて変更し、前記サイズ変更後の前記対応する前記推論タスク以外の前記第２の特徴量と前記基準特徴量とを結合して前記第３の特徴量を生成し、前記第４のコンポーネントへ出力する、

付記１に記載の画像処理システム。

［付記４］

前記学習部は、複数の学習段階に分けて前記学習済みモデルの学習を行い、

前記複数の学習段階は、少なくとも、

前記複数の推論タスクのうちの何れか１つを学習対象タスクとし、前記学習対象タスク以外の推論タスクに係る前記第２のコンポーネントおよび前記第３のコンポーネントと前記第１のコンポーネントのパラメータを固定して、前記学習対象タスクに係る前記第２のコンポーネントおよび前記第３のコンポーネントのパラメータを学習する第１の学習段階と、

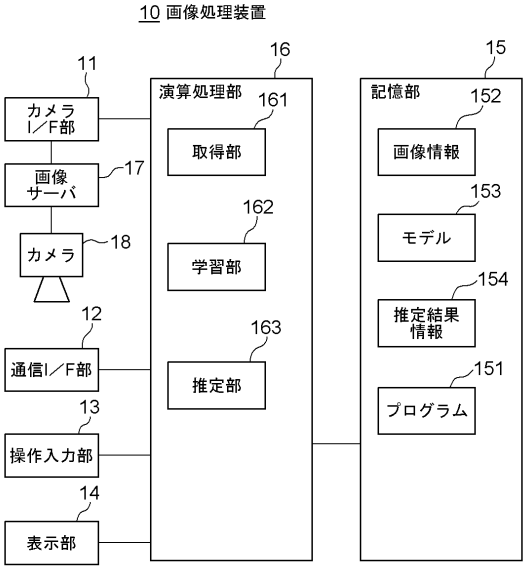
前記第１のコンポーネントのパラメータを固定して、前記全ての推論タスクに係る前記第２のコンポーネントおよび前記第３のコンポーネントのパラメータを学習する第２の学習段階と、を含む、

付記１乃至３の何れかに記載の画像処理システム。
［付記５］
前記推論タスクに対応して設けられた第４のコンポーネントは、前記第３の特徴量を構成する複数の前記第２の特徴量のうち、対応する前記推論タスクの第２の特徴量の優先度合いを定める重みをそれ以外の第２の特徴量の優先度合いを定める重みより大きくする、
付記１乃至４の何れかに記載の画像処理システム。
［付記６］
前記推論タスクに対応して設けられた第４のコンポーネントは、入力された前記第３の特徴量に対して１×１の畳み込みを行うことにより、前記第３の特徴量の次元数を削減する、
付記１乃至５の何れかに記載の画像処理システム。
［付記７］
前記複数の推論タスクは、物体検出タスク、ポーズ推定タスク、セマンティックセグメンテーション推定タスクを含む、
付記１乃至６の何れかに記載の画像処理システム。
［付記８］
前記学習済みモデルを用いて、画像から前記複数の推論タスクの推論結果を出力する推論部を、更に含む、
付記１乃至７の何れかに記載の画像処理システム。
［付記９］
学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を出力する推論部を含み、
前記学習済みモデルは、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出する第１のコンポーネントと、
前記推論タスクに対応して設けられ、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出する第２のコンポーネントと、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成する第３のコンポーネントと、
前記推論タスクに対応して設けられ、前記第３の特徴量から対応する推論タスクの推論結果を出力する第４のコンポーネントと、
を含む画像処理システム。
［付記１０］
画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成し、
前記生成では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
画像処理方法。
［付記１１］
学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力し、
前記推定では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、

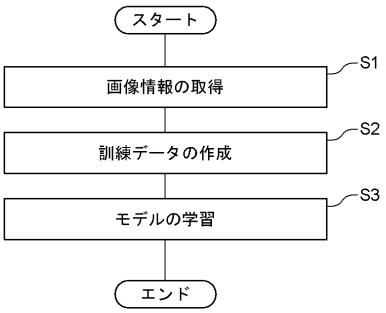
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
画像処理方法。
［付記１２］
コンピュータに、画像から互いに相違する複数の推論タスクを行う学習済みモデルを生成する処理を行わせるためのプログラムであって、
前記生成では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラムを記録したコンピュータ読み取り可能な記録媒体。
［付記１３］
コンピュータに、学習済みモデルを用いて、画像から互いに相違する複数の推論タスクの推論結果を推定して出力する処理を行わせるためのプログラムであって、
前記推定では、前記学習済みモデルに、
前記画像から前記複数の推論タスクに共通な第１の特徴量を抽出させ、
前記推論タスク毎に、前記第１の特徴量から対応する推論タスクに固有な第２の特徴量を抽出させ、
前記推論タスク毎に抽出された第２の特徴量を結合して第３の特徴量を生成させ、
前記推論タスク毎に、前記第３の特徴量から対応する推論タスクの推論結果を出力させる、
プログラムを記録したコンピュータ読み取り可能な記録媒体。
【符号の説明】
【００９０】
１０　画像処理装置
１１　カメラＩ／Ｆ部
１２　通信Ｉ／Ｆ部
１３　操作入力部
１４　画面表示部
１５　記憶部
１６　演算処理部
１７　画像サーバ
１８　カメラ
１５１　プログラム
１５２　画像情報
１５３　モデル
１５４　推定結果情報
１６１　取得部
１６２　学習部
１６３　推定部

【図面】

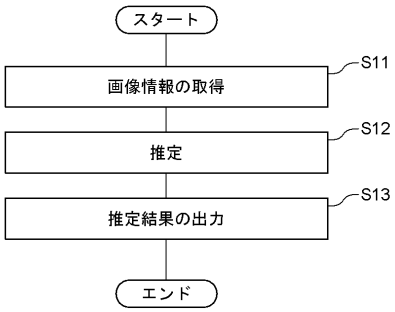
【図１】


10 画像処理装置
11
カメラ I／F部
17
画像 サーバ
18
カメラ
12
通信I／F部
13
操作入力部
14
表示部
16
演算処理部
161
取得部
162
学習部
163
推定部
15
記憶部
152
画像情報
153
モデル
154
推定結果 情報
151
プログラム

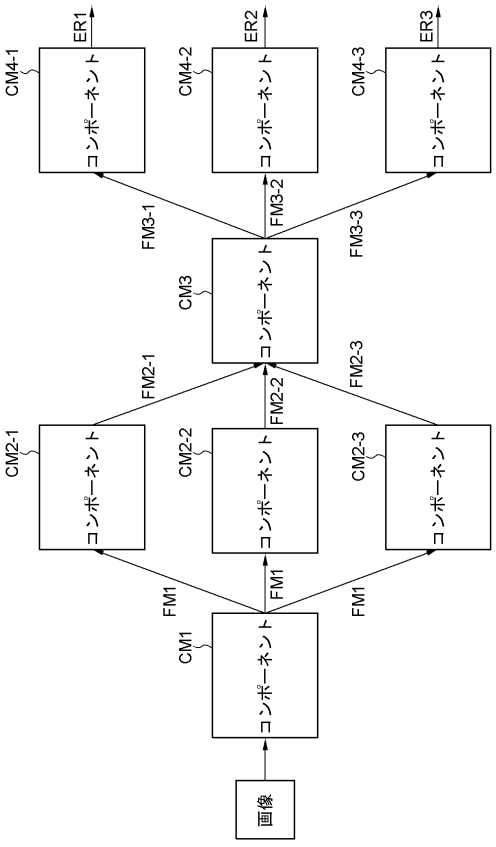
【図２】


スタート
S1
画像情報の取得
S2
訓練データの作成
S3
モデルの学習
エンド

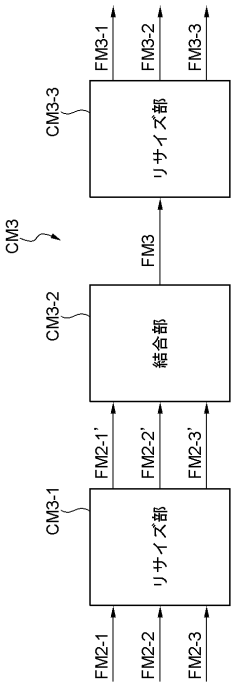
【図３】


スタート
S11
画像情報の取得
S12
推定
S13
推定結果の出力
エンド

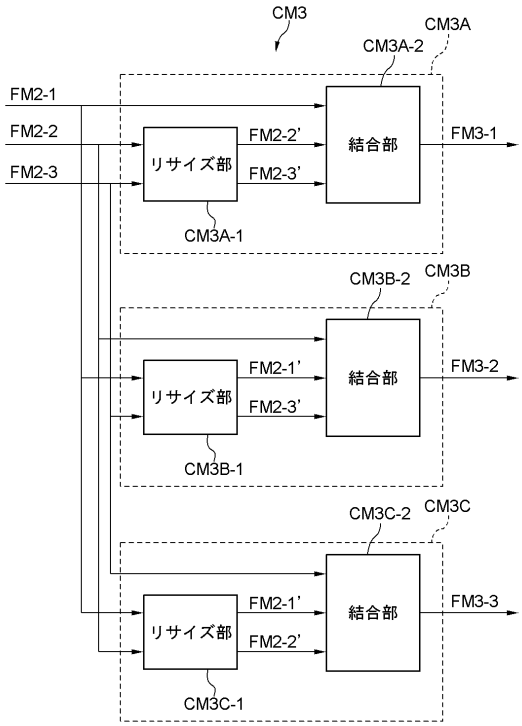
【図４】


ER1
ER2
ER3
CM4-1
CM4-2
CM4-3
コンポーネント
FM3-1
FM3-2
FM3-3
CM3
FM2-1
FM2-2
FM2-3
CM2-1
CM2-2
CM2-3
FM1
CM1
画像

【図５】



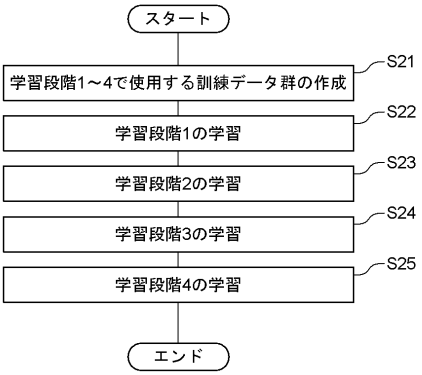
【図６】



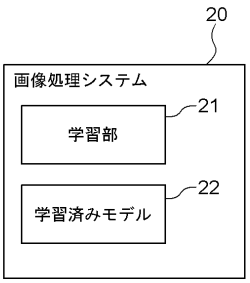
【図７】

ID	画像	物体検出ラベル	ポーズ推定ラベル	セグメンテーション推定ラベル
1	G1	有	有	有
2	G2	有	有	無
3	G3	有	無	無
・・・	・・・	・・・	・・・	・・・
n	Gn	有	有	有

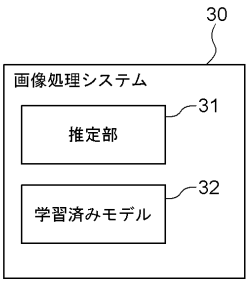
【図８】



【図９】



【図１０】



フロントページの続き

(56)参考文献 特開２０２１－０２１９７８（ＪＰ，Ａ）
特開２０１９－１９２００９（ＪＰ，Ａ）

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｔ ７／００－ ７／９０
Ｇ０６Ｖ １０／００－２０／９０
Ｇ０６Ｎ ３／００－９９／００