

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年11月3日(03.11.2022)



(10) 国際公開番号

WO 2022/230147 A1

- (51) 国際特許分類:
G06T 7/00 (2017.01)
- (21) 国際出願番号: PCT/JP2021/017091
- (22) 国際出願日: 2021年4月28日(28.04.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 伊藤 淳(ITO, Jun); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産

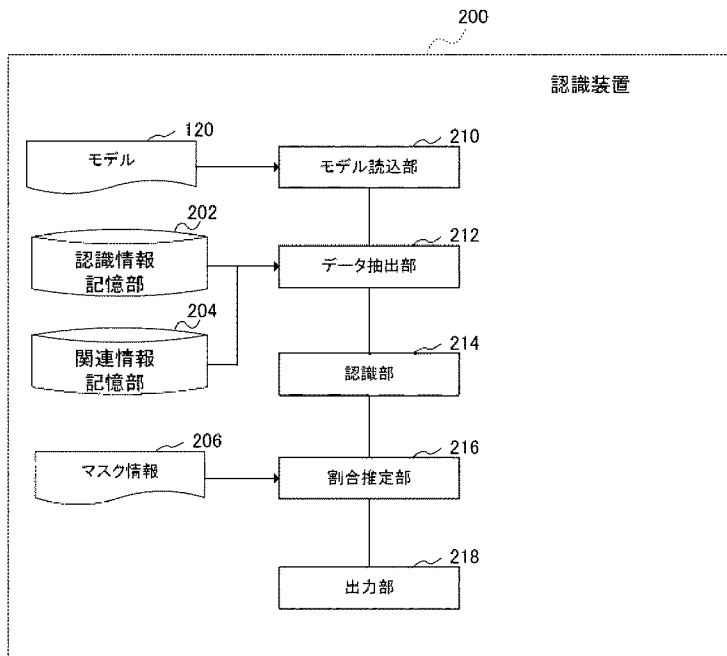
センタ内 Tokyo (JP). 横山 法子(YOKOYAMA, Noriko); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 山本 千尋(YAMAMOTO, Chihiro); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 尾花 和昭(OBANA, Kazuaki); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 特許業務法人太陽国際特許事務所 (TAIYO, NAKAJIMA & KATO); 〒1600022 東京都新宿区新宿4丁目3番17号 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ,

(54) Title: RECOGNITION DEVICE, RECOGNITION METHOD, RECOGNITION PROGRAM, MODEL TRAINING DEVICE, MODEL TRAINING METHOD, AND MODEL TRAINING PROGRAM

(54) 発明の名称: 認識装置、認識方法、認識プログラム、モデル学習装置、モデル学習方法、及びモデル学習プログラム



- 120 Model
200 Recognition device
202 Recognition information storage unit
204 Related information storage unit
206 Mask information
210 Model reading unit
212 Data extraction unit
214 Recognition unit
216 Ratio estimation unit
218 Output unit

(57) Abstract: This recognition device comprises a data extraction unit, a recognition unit, and a ratio estimation unit. In relation to a recognition target image, which is a post-treatment image from among pre-treatment and post-treatment images obtained by photographing a container containing an object before and after treatment of the container, the data extraction unit acquires related information related to the object, and extracts recognition target data that is a combination

BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

of the recognition target image and the related information. The recognition unit inputs the recognition target data to a trained model to output a recognition result recognizing regions in which at least the container, the object, and other objects are divided. The ratio estimation unit estimates a ratio of the object in the recognition target image on the basis of the recognition result and an area ratio in the pre-treatment image stored in advance. The model transforms the recognition target image into a feature map, and performs weighted calculation of the feature map using latent information obtained from the related information, to thereby recognize the regions.

(57) 要約: 認識装置は、データ抽出部と、認識部と、割合推定部とを含む。データ抽出部は、対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、対象物に関連する関連情報を取得し、認識対象画像及び関連情報の組み合わせである認識対象データを抽出する。認識部は、認識対象データを予め学習されたモデルへの入力としてモデルの出力により、容器と、対象物と、対象物以外と、が少なくとも区分される領域を認識した認識結果を出力する。割合推定部は、予め保存された前後を撮影した前の画像における面積割合と、認識結果とに基づいて、認識対象画像における対象物の割合を推定する。モデルは、認識対象画像を特徴量マップへと変換し、関連情報から得られる潜在情報により特徴量マップを重み付けて算出することにより、領域を認識する。

明 細 書

発明の名称：

認識装置、認識方法、認識プログラム、モデル学習装置、モデル学習方法、及びモデル学習プログラム

技術分野

[0001] 開示の技術は、認識装置、認識方法、認識プログラム、モデル学習装置、モデル学習方法、及びモデル学習プログラムに関する。

背景技術

[0002] 従来、画像に何が写っているかを認識する手法として、U-Netなどに代表されるセマンティックセグメンテーションが用いられてきた（非特許文献1参照）。何が写っているかを矩形ごとのラベルで表現する物体検出と異なり、セマンティックセグメンテーションは画素ごとのラベルで表現するため、被写体の領域を細かく捉えることができる。

先行技術文献

非特許文献

[0003] 非特許文献1：O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in MICCAI 2015.

発明の概要

発明が解決しようとする課題

[0004] しかし、認識の難易度が高い場合に、従来の画像情報のみを用いたセマンティックセグメンテーション技術をそのまま適用するだけでは認識精度が低い場合がある。

[0005] 開示の技術は、上記の点に鑑みてなされたものであり、認識の困難な画像の領域の認識を可能にする認識装置、認識方法、認識プログラム、モデル学習装置、モデル学習方法、及びモデル学習プログラムを提供することを目的とする。

課題を解決するための手段

[0006] 本開示の第1態様は、認識装置であって、対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出するデータ抽出部と、前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力する認識部と、予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定する割合推定部と、を含み、前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する。

[0007] 本開示の第2態様は、モデル学習装置であって、対象物を収めた容器に対する処置の前後を撮影した学習用の後の画像、前記後の画像に対応する学習用のマスク画像、及び前記対象物に関連する関連情報を含む学習用のデータを入力とし、モデルにより、前記画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識したマスク画像を認識結果として出力する認識部と、前記認識結果のマスク画像と、前記学習用のデータに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるように前記モデルのパラメータを更新するモデル更新部と、を含む。

発明の効果

[0008] 開示の技術によれば、認識の困難な画像の領域の認識を可能にすることができる。

図面の簡単な説明

[0009] [図1]モデル学習装置及び認識装置のハードウェア構成を示すブロック図であ

る。

[図2]本実施形態のモデル学習装置の構成を示すブロック図である。

[図3]学習用情報記憶部のデータ構造の例である。

[図4]関連情報記憶部のデータ構造の例である。

[図5]本実施形態の認識装置の構成を示すブロック図である。

[図6]マスク情報のデータ構造の例である。

[図7]出力部の出力データのデータ構造の例である。

[図8]モデルにおけるセマンティックセグメンテーションのネットワーク構成例である。

[図9]特微量マップのチャンネル成分に対して、重み付き特微量マップ算出処理を行う例である。

[図10]特微量マップの空間成分に対して、重み付き特微量マップ算出処理を行う例である。

[図11]モデル学習装置によるモデル学習処理の流れを示すフローチャートである。

[図12]認識装置による認識処理の流れを示すフローチャートである。

発明を実施するための形態

[0010] 以下、開示の技術の実施形態の一例を、図面を参照しつつ説明する。なお、各図面において同一又は等価な構成要素及び部分には同一の参照符号を付与している。また、図面の寸法比率は、説明の都合上誇張されており、実際の比率とは異なる場合がある。

[0011] まず、本開示の概要について説明する。

[0012] セマンティックセグメンテーションの技術は、例えば、食事後の食器を撮影した画像から食べ残しを認識するようなユースケースにおいても適用ができる。

[0013] しかし、上記課題において説明したように、食事後の食器を撮影した画像から食べ残しを認識するユースケースにおいて、従来の画像情報のみを用いたセマンティックセグメンテーション技術をそのまま適用するだけでは認識

精度が低い場合がある。皿に残ったものがすべて食べ残しとなるわけではなく、食べ残しに該当しない残留物も含まれるため、認識の難易度が高いことが原因である。例えば、お椀に液体が残っていた場合、メニューがスープであれば食べ残しとなるが、ラーメンであれば完飲しないことも考えられるため食べ残しに該当しない残留物となるべきである。食べ残しに該当しない残留物のその他の例として、エビのしっぽ、パセリ、ソースやドレッシングによる皿の汚れ、などがあげられる。

[0014] なお、以下の実施形態の説明においては食器の食べ残しを認識する態様を例に説明するが、容器の対象物の認識全般に適用可能である。

[0015] 本実施形態の構成について説明する。実施形態は、モデル学習装置と、認識装置とのそれぞれについて説明する。

[0016] 図1は、モデル学習装置100及び認識装置200のハードウェア構成を示すブロック図である。モデル学習装置100及び認識装置200は同様のハードウェア構成とすることができる。

[0017] 図1に示すように、モデル学習装置100は、CPU (Central Processing Unit) 11、ROM (Read Only Memory) 12、RAM (Random Access Memory) 13、ストレージ14、入力部15、表示部16及び通信インタフェース (I/F) 17を有する。各構成は、バス19を介して相互に通信可能に接続されている。

[0018] CPU 11は、中央演算処理ユニットであり、各種プログラムを実行したり、各部を制御したりする。すなわち、CPU 11は、ROM 12又はストレージ14からプログラムを読み出し、RAM 13を作業領域としてプログラムを実行する。CPU 11は、ROM 12又はストレージ14に記憶されているプログラムに従って、上記各構成の制御及び各種の演算処理を行う。本実施形態では、ROM 12又はストレージ14には、モデル学習プログラムが格納されている。

[0019] ROM 12は、各種プログラム及び各種データを格納する。RAM 13は

、作業領域として一時的にプログラム又はデータを記憶する。ストレージ14は、HDD (Hard Disk Drive) 又はSSD (Solid State Drive) 等の記憶装置により構成され、オペレーティングシステムを含む各種プログラム、及び各種データを格納する。

[0020] 入力部15は、マウス等のポインティングデバイス、及びキーボードを含み、各種の入力を行うために使用される。

[0021] 表示部16は、例えば、液晶ディスプレイであり、各種の情報を表示する。表示部16は、タッチパネル方式を採用して、入力部15として機能してもよい。

[0022] 通信インタフェース17は、端末等の他の機器と通信するためのインタフェースである。当該通信には、例えば、イーサネット（登録商標）若しくはFDDI等の有線通信の規格、又は、4G、5G、若しくはWi-Fi（登録商標）等の無線通信の規格が用いられる。

[0023] 認識装置200についても同様に、CPU21、ROM22、RAM23、ストレージ24、入力部25、表示部26及び通信I/F27を有する。各構成は、バス29を介して相互に通信可能に接続されている。ROM22又はストレージ24には、認識プログラムが格納されている。ハードウェア構成の各部についての説明は、モデル学習装置100と同様であるため省略する。

[0024] 次に、モデル学習装置100の各機能構成について説明する。

[0025] 図2は、本実施形態のモデル学習装置100の構成を示すブロック図である。各機能構成は、CPU11がROM12又はストレージ14に記憶されたモデル学習プログラムを読み出し、RAM13に展開して実行することにより実現される。

[0026] 図2に示すように、モデル学習装置100は、学習用情報記憶部102と、関連情報記憶部104と、データ分割部110と、認識部112と、モデル更新部114と、モデル書込部116と、モデル120とを含んで構成される。

[0027] 図3は学習用情報記憶部102のデータ構造の例である。学習用情報記憶部102には、少なくとも食事後画像、マスク画像、及びメニューIDが含まれていることとする。食事後画像は過去の食事後の画像を学習のために保存したものである。マスク画像は食事後画像中の背景、皿、食べ残し、及び食べ残しに該当しない残留物などの領域を人手によりマスキング（色分け）したものである。メニューIDは食事後画像に対応するメニューのIDであり、関連情報記憶部104のテーブルを参照する際のキーとして使用する。なお、皿が、本開示の容器の一例である。食べ残し（食べ物）が、本開示の対象物の一例である。食べ残しに該当しない残留物が、本開示の対象物以外の一例である。食事後画像が、本開示の対象物を収めた容器に対する処置の前後を撮影した後の画像の一例である。処置とは、本実施形態では食事に適用され、処置前が食事前（配膳時）、処置後が食事後である。

[0028] 図4は関連情報記憶部104のデータ構造の例である。関連情報記憶部104には、少なくともメニューID、メニュー関連情報、及び配膳時面積割合が含まれていることとする。また、メニュー関連情報はメニュー名、食材名、皿の種類など、メニューに関連する情報を少なくとも1つ含んでいることとする。配膳時面積割合は、配膳時における皿の面積に対するそれぞれの食べ物の面積割合である。配膳時面積割合は、配膳時の画像とメニュー関連情報を認識部112に入力し、出力されたマスク画像から計算する、又は配膳時の画像から目視で概算する、などの方法で数値を得た結果が保存されている。メニュー関連情報が、本開示の対象物に関連する関連情報の一例である。なお、対象物に関連する関連情報には、容器、及び対象物以外の情報も含まれる。配膳時面積割合が、本開示の処置の前後を撮影した前の画像における面積割合の一例である。

[0029] データ分割部110は、学習用情報記憶部102、及び関連情報記憶部104を入力とし、学習データとテストデータに学習用のデータを分割する。学習データ及びテストデータは、学習用情報記憶部102の食事後画像、マスク画像、及びメニューIDをキーにして参照した関連情報記憶部104の

メニュー関連情報からなる。学習データとテストデータに構造的な違いはなく、モデル更新部 114 での用途が異なる。学習データ及びテストデータが、本開示の学習用のデータの一例である。

[0030] 認識部 112 及びモデル更新部 114 の繰り返し処理によりモデルのパラメータが更新される。なお、モデルのネットワーク構成については後述する。

[0031] 認識部 112 は、学習データ及びテストデータを入力とし、モデルのセマンティックセグメンテーションにより背景、皿、食べ残し、及び食べ残しに該当しない残留物などが区分される領域を認識する。認識結果はマスク画像として出力される。従来の技術では、画像情報以外の入力を想定した構成となっていなかったが、本開示の手法では関連情報記憶部 104 のメニュー関連情報も入力とした構成になっている点が特徴である。なお、認識部 112 の詳細は後述する。

[0032] モデル更新部 114 は、学習データを入力とした時の認識結果のマスク画像と、学習データに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるようにモデルのパラメータを更新する。また、モデル更新部 114 は、テストデータを入力とした時の認識結果のマスク画像と、テストデータに含まれるマスク画像の一致度を正解率として数値化し、モデルの汎化性能を計測する。前回の学習時より汎化性能の低下が認められる場合は学習を終了してモデル書込部 116 の処理に移る。汎化性能の低下が認められない場合は学習を継続し、認識部 112 の処理に戻る。

[0033] モデル書込部 116 は、学習したモデル 120 を外部ファイルとして出力する。

[0034] モデル学習装置 100 において、画像情報以外の入力であるメニュー関連情報が存在すること、画像情報以外の入力を扱える認識部 112 になっていることが本開示の手法の特徴である。それ以外の構成は機械学習においてモデルを学習するための一般的な構成を適用すればよい。

[0035] 次に、認識装置 200 の各機能構成について説明する。

- [0036] 図5は、本実施形態の認識装置200の構成を示すブロック図である。各機能構成は、CPU21がROM22又はストレージ24に記憶された認識プログラムを読み出し、RAM23に展開して実行することにより実現される。
- [0037] 図5に示すように、認識装置200は、モデル120と、認識情報記憶部202と、関連情報記憶部204と、マスク情報206と、モデル読込部210と、データ抽出部212と、認識部214と、割合推定部216と、出力部218とを含んで構成される。
- [0038] モデル120は、モデル学習装置100で学習されたモデルである。
- [0039] 認識情報記憶部202は、データ構造として、少なくとも食事後画像、及びメニューIDが含まれていることとする。認識情報記憶部202のデータ構造は、図3に示した学習用情報記憶部102のデータ構造からマスク画像を除いた形式である。
- [0040] 関連情報記憶部204は、モデル学習装置100の関連情報記憶部104と同様のデータ構造を持ち、メニューID、メニュー関連情報、及び配膳時面積割合が含まれる。
- [0041] 図6は、マスク情報206のデータ構造の例である。マスクIDはマスク画像のカラーパレットのID、マスク名はそれに対応するマスクが何であることを示した名前である。例えば、カラーパレットの0番がRGB=(0, 0, 0)の黒色だったとすると、インデックスカラーのマスク画像において黒色となっている領域は背景であるということを意味する。
- [0042] モデル読込部210は、モデル120のファイルを読み込み、メモリ上に展開する。
- [0043] データ抽出部212は、認識情報記憶部202の食事後画像と、関連情報記憶部204のメニュー関連情報とを入力とし、認識対象データを抽出する。認識対象データは、食事後画像、及び食事後画像に対応するメニュー関連情報からなる。学習データ及びテストデータと異なり、マスク画像は認識対象データには含まれない。

[0044] 認識部 214 は、認識対象データをモデル 120 への入力として、モデル 120 の出力により、背景、皿、食べ残し、及び食べ残しに該当しない残留物などが区分される領域を認識した認識結果を出力する。入力を受けたモデル 120 ではセマンティックセグメンテーションを行う。認識結果はマスク画像として出力される。モデル学習時の認識部 112 と入力は異なるが同じ出力のマスク画像である。

[0045] 割合推定部 216 は、関連情報記憶部 204 に保存された配膳時面積割合と、認識部 214 で出力される認識結果であるマスク画像とに基づいて、食べ残し割合を推定する。マスク画像はインデックスカラーとなっているため、インデックスが何であるかは、図 6 のマスク情報 206 を読み込み、マスク ID（インデックス）に対応するマスク名を取得することで解決する。以下に食べ残し割合の計算方法を示す。

[0046] 関連情報記憶部 204 の配膳時面積割合でポテトの数値が 0.15 と保存されており、マスク画像に皿が 128 ピクセル、ポテトが 24 ピクセル、ブロッコリーが 18 ピクセル、ソース汚れが 30 ピクセル写っていたとする。食べ残し及び食べ残しに該当しない残留物は皿の上に位置していると考えられるため、マスク画像において皿の面積に対するポテトの面積の割合を求めると、以下のように計算できる。

$$24 \div (128 + 24 + 18 + 30) = 0.12$$

[0047] したがって、食べ残し割合は、以下の計算となり、配膳時のポテトに対して 80% が食べ残しになっている、と計算できる。

$$100 * 0.12 \div 0.15 = 80$$

[0048] ある食べ物 t における食べ残し割合を r_t 、配膳時面積割合を a_t 、マスク画像に写っている物を $m \in M$ 、そのピクセル数を p_m とすると、以下の (1) 式で一般化される。

[数1]

$$r_t = \min \left(100, \frac{100 p_{m=t}}{a_t \sum_{m \in M} p_m} \right)$$

．．．（１）

[0049] $\min$ は最小値を返す関数であり、食べ残し割合の上限を100%にするために使用する。なお、食べ残しに該当しない残留物は食べ残しではないため、食べ残し割合の計算は行われない。

[0050] 出力部218は、割合推定部216の食べ残し割合の計算結果を外部システムに向けて出力する。出力データ構造は出力にあたって外部システムの入力インタフェースに合わせて変更する。図7は出力部218の出力データのデータ構造の例である。食べ残し品目はマスク情報206のマスク名と同じである。食べ残し割合は割合推定部216によって推定された値である。

[0051] 次に、モデル120のネットワークを説明する。図8は、モデル120におけるセマンティックセグメンテーションのネットワーク構成例である。入力された食事後画像は、畳み込み（convolution）、最大プーリング（max pooling）、及びアップサンプリング（upsampling）などの処理により、異なるシェイプの特徴量マップへと変換され、最終的にマスク画像として出力される。破線部で囲まれた部分がネットワーク構成上の特徴である。このネットワーク構成では、食事後画像以外にメニュー関連情報が入力として受け付けられるようになっており、ある特徴量マップをメニュー関連情報によって重み付け、重み付き特徴量マップとして出力する機能を有している。図8では中段のスキップ（skip）部分に重み付き特徴量マップ算出処理を適用する例を示したが、本処理は入出力前後で特徴量マップのシェイプが変わらないため、任意の特徴量マップに対して適用することができる。また、図8では1箇所だけに本処理を適用する例を示したが、複数箇所に適用することができる。重み付き特徴量マップ算出処理はいくつかの処理方法が考えられ、次に代表例を2つ挙げて説明する。

[0052] 図9は、特徴量マップのチャンネル成分に対して、重み付き特徴量マップ算出処理を行う例である。入力の特徴量マップ f_{in} 、出力の重み付き特徴量マップ f_{out} はともにシェイプが（H，W，C）である。もう1つの入力であるメニュー関連情報 m は、例えばメニュー名などテキストデータである場合

は、形態素解析などにより単語分割処理を行った上で単語埋め込み (Word Embedding) によって単語ごとにベクトルに変換する。そして、メニュー名に含まれる全単語のベクトルの平均値をとるなどして用意する。皿の種類などカテゴリカルデータである場合は、One-hot エンコーディングによってベクトルに変換して用意する。メニュー関連情報 m は全結合層によって潜在情報 C' に変換される。メニュー関連情報の情報源が複数ある場合、例えばメニュー名と皿の種類を同時に利用したい場合は、複数の情報源のそれぞれで得たベクトルを連結したメニュー関連情報 m とすることで、複数の情報源に対応できる。そのため柔軟に追加情報の増減に対応できる。

[0053] 以下にチャンネル成分の重み w_c 、重み付き特徴量マップ f_{out} の算出式 (2-1)、(2-2) を示す。

[数2]

$$w_c = F_{c4} \left(F_{c3} \left(F_{concat} (F_{c1}(f_{in}), F_{c2}(m)) \right) \right) \quad \dots (2-1)$$

$$f_{out} = w_c \otimes f_{in} \quad \dots (2-2)$$

[0054] なお、 $F_{concat}(X, Y)$ は、テンソル X とテンソル Y を連結する操作を意味する関数である。

[0055] 図10は、特徴量マップの空間成分に対して、重み付き特徴量マップ算出処理を行う例である。空間成分に重みを適用するため、チャンネル成分に重みを適用した時と途中の特徴量マップや関数は異なっているが、入出力は同じである。

[0056] 以下に空間成分の重み w_s 、重み付き特徴量マップ f_{out} の算出式 (3-1)、(3-2) を示す。

[数3]

$$w_s = F_{s4} \left(F_{concat} \left(F_{s1}(f_{in}), F_{s3}(F_{s2}(m)) \right) \right) \dots (3-1)$$

$$f_{out} = w_s \otimes f_{in} \dots (3-2)$$

[0057] 次に、モデル学習装置100及び認識装置200の作用について説明する。

[0058] 図11は、モデル学習装置100によるモデル学習処理の流れを示すフローチャートである。CPU11がROM12又はストレージ14からモデル学習プログラムを読み出して、RAM13に展開して実行することにより、モデル学習処理が行なわれる。

[0059] ステップS100において、CPU11は、データ分割部110として、学習用情報記憶部102、及び関連情報記憶部104を入力とし、学習データとテストデータにデータを分割する。

[0060] ステップS102において、CPU11は、認識部112として、学習データ及びテストデータを入力とし、モデルのセマンティックセグメンテーションにより背景、皿、食べ残し、及び食べ残しに該当しない残留物などが区分される領域を認識する。

[0061] ステップS104において、CPU11は、モデル更新部114として、学習データを入力とした時の認識結果のマスク画像と、学習データに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるようにモデルのパラメータを更新する。

[0062] ステップS106において、CPU11は、モデル更新部114として、テストデータを入力とした時の認識結果のマスク画像と、テストデータに含まれるマスク画像の一致度を正解率として数値化し、モデルの汎化性能を計測する。

- [0063] ステップS 1 0 8において、CPU 1 1は、モデル更新部 1 1 4として、汎化性能の低下があるか否かを判定する。前回の学習時より汎化性能の低下がある場合はステップS 1 1 0へ移行し、汎化性能の低下がない場合はステップS 1 0 2に戻って処理を繰り返す。
- [0064] ステップS 1 1 0において、CPU 1 1は、モデル書込部 1 1 6として、学習したモデル 1 2 0を外部ファイルとして出力する。
- [0065] 以上説明したように本実施形態のモデル学習装置 1 0 0によれば、認識の困難な画像の領域の認識を可能にするモデルのパラメータを学習できる。
- [0066] 図 1 2は、認識装置 2 0 0による認識処理の流れを示すフローチャートである。CPU 2 1がROM 2 2又はストレージ 2 4から認識プログラムを読み出して、RAM 2 3に展開して実行することにより、認識処理が行なわれる。
- [0067] ステップS 2 0 0において、CPU 2 1は、モデル読込部 2 1 0として、モデル 1 2 0のファイルを読み込み、メモリ上に展開する。
- [0068] ステップS 2 0 2において、CPU 2 1は、データ抽出部 2 1 2として、認識情報記憶部 2 0 2の食事後画像と、関連情報記憶部 2 0 4のメニュー関連情報とを入力とし、認識対象データを抽出する。
- [0069] ステップS 2 0 4において、CPU 2 1は、認識部 2 1 4として、認識対象データをモデル 1 2 0への入力として、モデル 1 2 0の出力により、背景、皿、食べ残し、及び食べ残しに該当しない残留物などが区分される領域を認識した認識結果を出力する。入力を受けたモデル 1 2 0ではセマンティックセグメンテーションを行う。認識結果はマスク画像として出力される。
- [0070] ステップS 2 0 6において、CPU 2 1は、割合推定部 2 1 6として、関連情報記憶部 2 0 4に保存された配膳時面積割合と、出力された認識結果であるマスク画像とに基づいて、食べ残し割合を推定する。
- [0071] ステップS 2 0 8において、CPU 2 1は、出力部 2 1 8として、割合推定部 2 1 6の食べ残し割合の計算結果を外部システムに向けて出力する。
- [0072] 以上説明したように本実施形態の認識装置 2 0 0によれば、認識の困難な

画像の領域の認識を可能にする。

[0073] また、メニュー関連情報から得た潜在情報と画像情報の関係性を重みとして利用することで、メニュー関連情報と食べ残しの関係性、メニュー関連情報と食べ残しに該当しない残留物との関係性を学習することができ、認識精度が向上する。

[0074] また、特徴量マップのチャンネル成分、又は空間成分の重みが算出される。そのため、どんな画像情報に対してどのようなメニュー関連情報が入力されると、特徴量マップのどのチャンネルや空間が注目されるのかが重みの値の大小から明らかになり、認識根拠についての説明可能性が与えられる。

[0075] なお、上記実施形態でCPUがソフトウェア（プログラム）を読み込んで実行したモデル学習処理又は認識処理を、CPU以外の各種のプロセッサが実行してもよい。この場合のプロセッサとしては、FPGA（Field-Programmable Gate Array）等の製造後に回路構成を変更可能なPLD（Programmable Logic Device）、並びに、GPU（Graphics Processing Unit）、及びASIC（Application Specific Integrated Circuit）等の特定の処理を実行させるために専用に設計された回路構成を有するプロセッサである専用電気回路等が例示される。また、モデル学習処理又は認識処理を、これらの各種のプロセッサのうちの1つで実行してもよいし、同種又は異種の2つ以上のプロセッサの組み合わせ（例えば、複数のFPGA、複数のGPU、及びCPUとFPGAとの組み合わせ等）で実行してもよい。また、これらの各種のプロセッサのハードウェア的な構造は、より具体的には、半導体素子等の回路素子を組み合わせた電気回路である。

[0076] また、上記実施形態では、モデル学習プログラム又は認識プログラムがストレージに予め記憶（インストール）されている態様を説明したが、これに限定されない。プログラムは、CD-ROM（Compact Disk Read Only Memory）、DVD-ROM（Digital

Versatile Disk Read Only Memory)、及びUSB(Universal Serial Bus)メモリ等の非一時的(non-transitory)記憶媒体に記憶された形態で提供されてもよい。また、プログラムは、ネットワークを介して外部装置からダウンロードされる形態としてもよい。

[0077] 以上の実施形態に関し、更に以下の付記を開示する。

[0078] (付記項1)

メモリと、

前記メモリに接続された少なくとも1つのプロセッサと、

を含み、

前記プロセッサは、

対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出し、

前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力し、

予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定し、

前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する、

ように構成されている認識装置。

[0079] (付記項2)

認識処理を実行するようにコンピュータによって実行可能なプログラムを記憶した非一時的記憶媒体であって、

対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対

象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出し、

前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力し、

予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定し、

前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する、

非一時的記憶媒体。

符号の説明

- [0080] 1 0 0 モデル学習装置
- 1 0 2 学習用情報記憶部
- 1 0 4 関連情報記憶部
- 1 1 0 データ分割部
- 1 1 2 認識部
- 1 1 4 モデル更新部
- 1 1 6 モデル書込部
- 1 2 0 モデル
- 2 0 0 認識装置
- 2 0 2 認識情報記憶部
- 2 0 4 関連情報記憶部
- 2 0 6 マスク情報
- 2 1 0 モデル読込部
- 2 1 2 データ抽出部
- 2 1 4 認識部

2 1 6 割合推定部

2 1 8 出力部

請求の範囲

- [請求項1] 対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出するデータ抽出部と、
- 前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力する認識部と、
- 予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定する割合推定部と、を含み、
- 前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する、
- 認識装置。
- [請求項2] 前記モデルにおいて、前記関連情報は全結合層によって前記潜在情報に変換される構成とし、前記関連情報の情報源を単一又は複数とする請求項1に記載の認識装置。
- [請求項3] 前記モデルは、前記特徴量マップのチャンネル成分、又は空間成分に対して重み付き特徴量マップ算出処理を行う請求項1又は請求項2に記載の認識装置。
- [請求項4] 対象物を収めた容器に対する処置の前後を撮影した学習用の後の画像、前記後の画像に対応する学習用のマスク画像、及び前記対象物に関連する関連情報を含む学習用のデータを入力とし、モデルにより、前記画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される

領域を認識したマスク画像を認識結果として出力する認識部と、

前記認識結果のマスク画像と、前記学習用のデータに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるように前記モデルのパラメータを更新するモデル更新部と、

を含むモデル学習装置。

[請求項5]

対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出し、

前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力し、

予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定し、

前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する、

ことを含む処理をコンピュータに実行させる認識方法。

[請求項6]

対象物を収めた容器に対する処置の前後を撮影した学習用の後の画像、前記後の画像に対応する学習用のマスク画像、及び前記対象物に関連する関連情報を含む学習用のデータを入力とし、モデルにより、前記画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識したマスク画像を認識結果として出力し、

前記認識結果のマスク画像と、前記学習用のデータに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるように前記モ

デルのパラメータを更新する、

ことを含む処理をコンピュータに実行させるモデル学習方法。

[請求項7]

対象物を収めた容器に対する処置の前後を撮影した後の画像である認識対象画像について、前記対象物に関連する関連情報を取得し、前記認識対象画像及び前記関連情報の組み合わせである認識対象データを抽出し、

前記認識対象データを予め学習されたモデルへの入力として前記モデルの出力により、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識した認識結果を出力し、

予め保存された前記前後を撮影した前の画像における面積割合と、前記認識結果とに基づいて、前記認識対象画像における前記対象物の割合を推定し、

前記モデルは、前記認識対象画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記領域を認識する、

処理をコンピュータに実行させる認識プログラム。

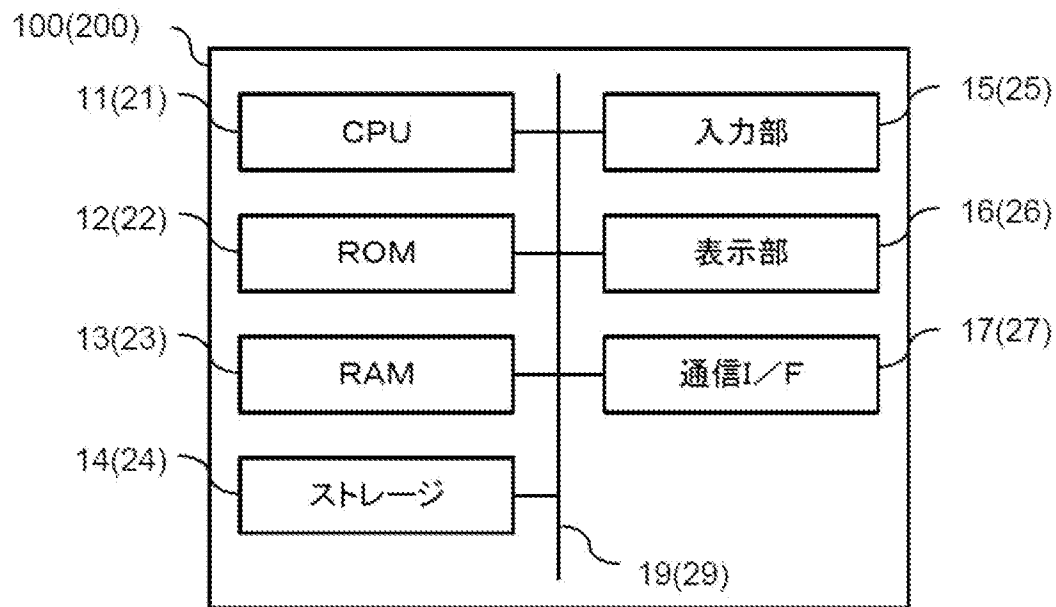
[請求項8]

対象物を収めた容器に対する処置の前後を撮影した学習用の後の画像、前記後の画像に対応する学習用のマスク画像、及び前記対象物に関連する関連情報を含む学習用のデータを入力とし、モデルにより、前記画像を特徴量マップへと変換し、前記関連情報から得られる潜在情報により前記特徴量マップを重み付けて算出することにより、前記容器と、前記対象物と、前記対象物以外と、が少なくとも区分される領域を認識したマスク画像を認識結果として出力し、

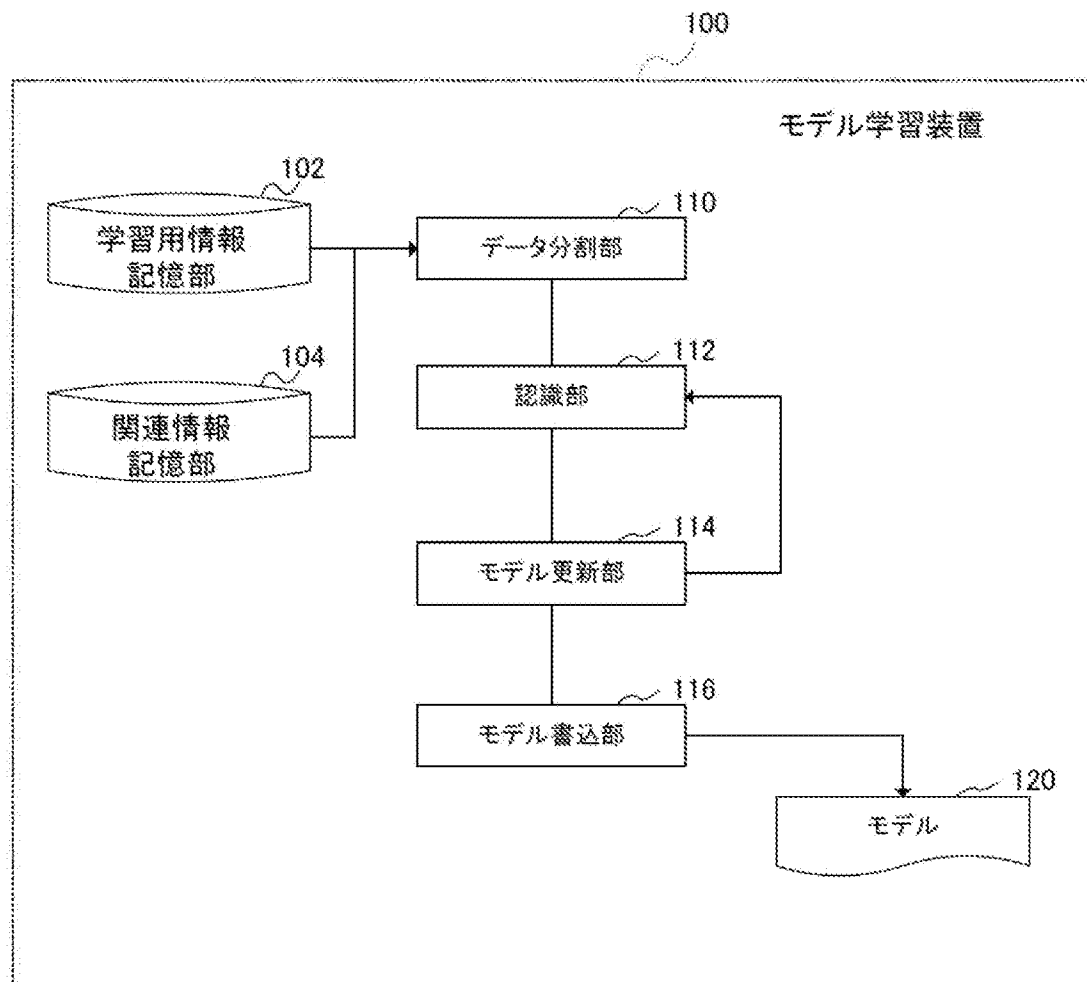
前記認識結果のマスク画像と、前記学習用のデータに含まれるマスク画像の差分を損失として数値化し、損失が少なくなるように前記モデルのパラメータを更新する、

処理をコンピュータに実行させるモデル学習プログラム。

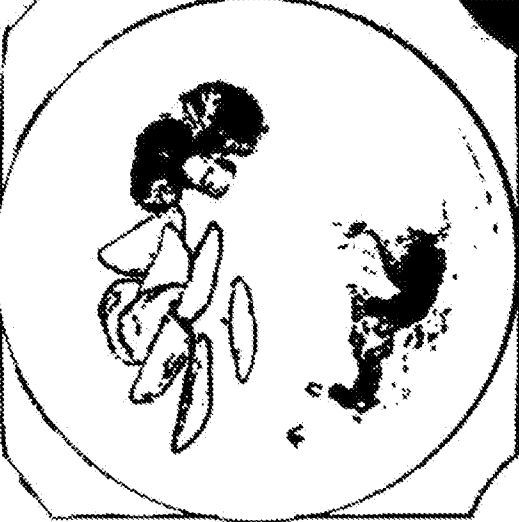
[図1]



[図2]



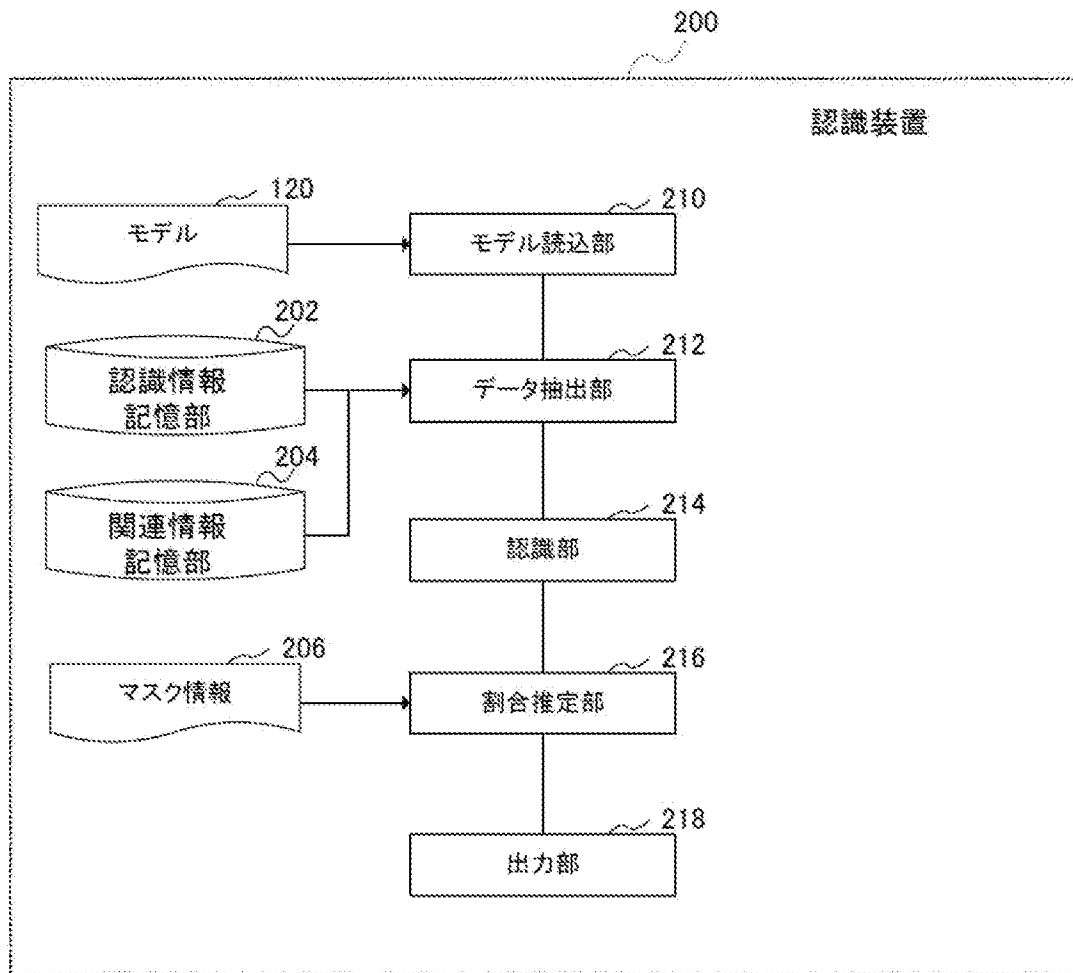
[図3]

...		...
メニューID	12345678	...
マスク画像		...
食事後画像		...

[図4]

メニュー関連情報					
メニューID	メニュー名	食材名	皿の種類	...	配膳時面積割合
12345678	ハンバーグ定食	牛豚ひき肉 じゃがいも ブロッコリー ...	白丸皿	...	ハンバーグ：0.4 ポテト：0.15 ブロッコリー：0.1 ...
...	...	...	...	...	...

[図5]



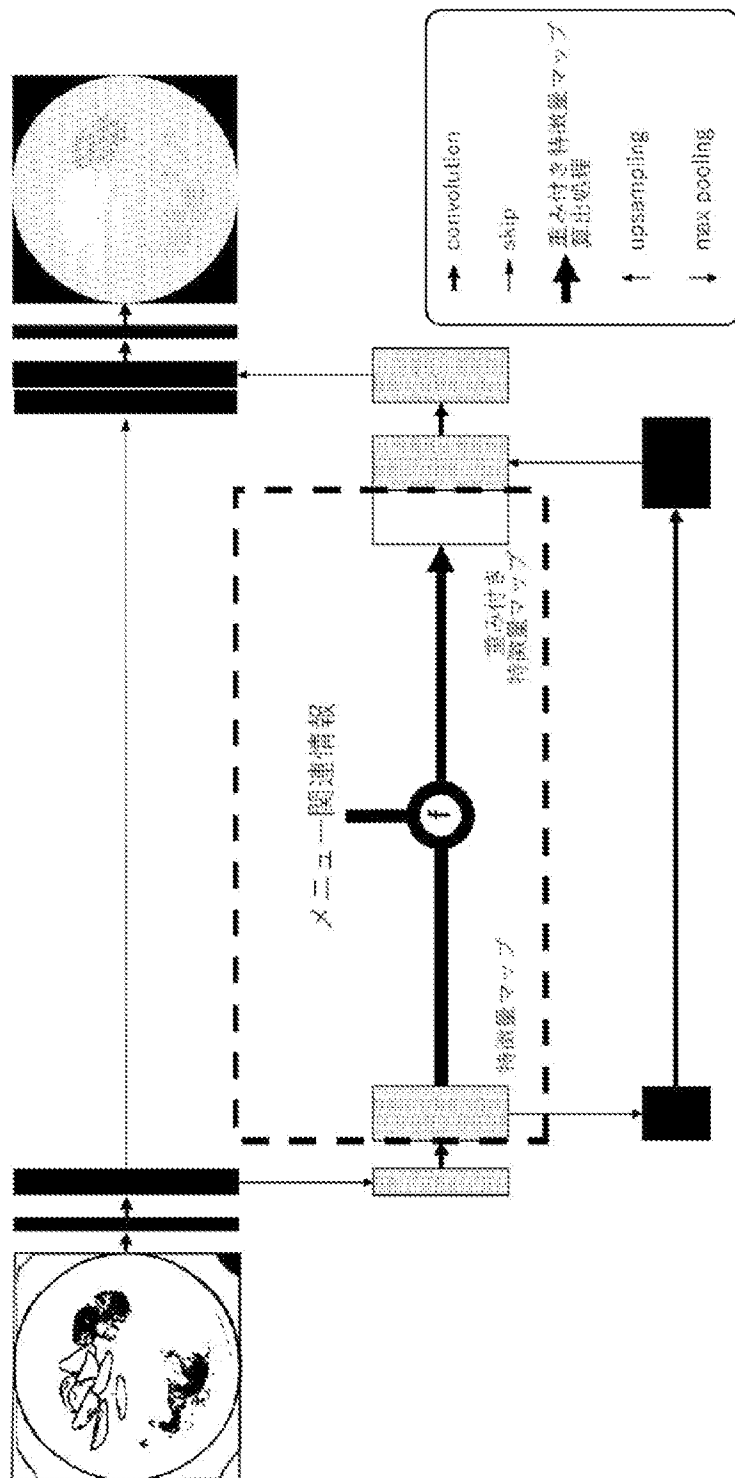
[図6]

マスクID	マスク名
0	背景
1	ポテト
2	ブロッコリー
...	...

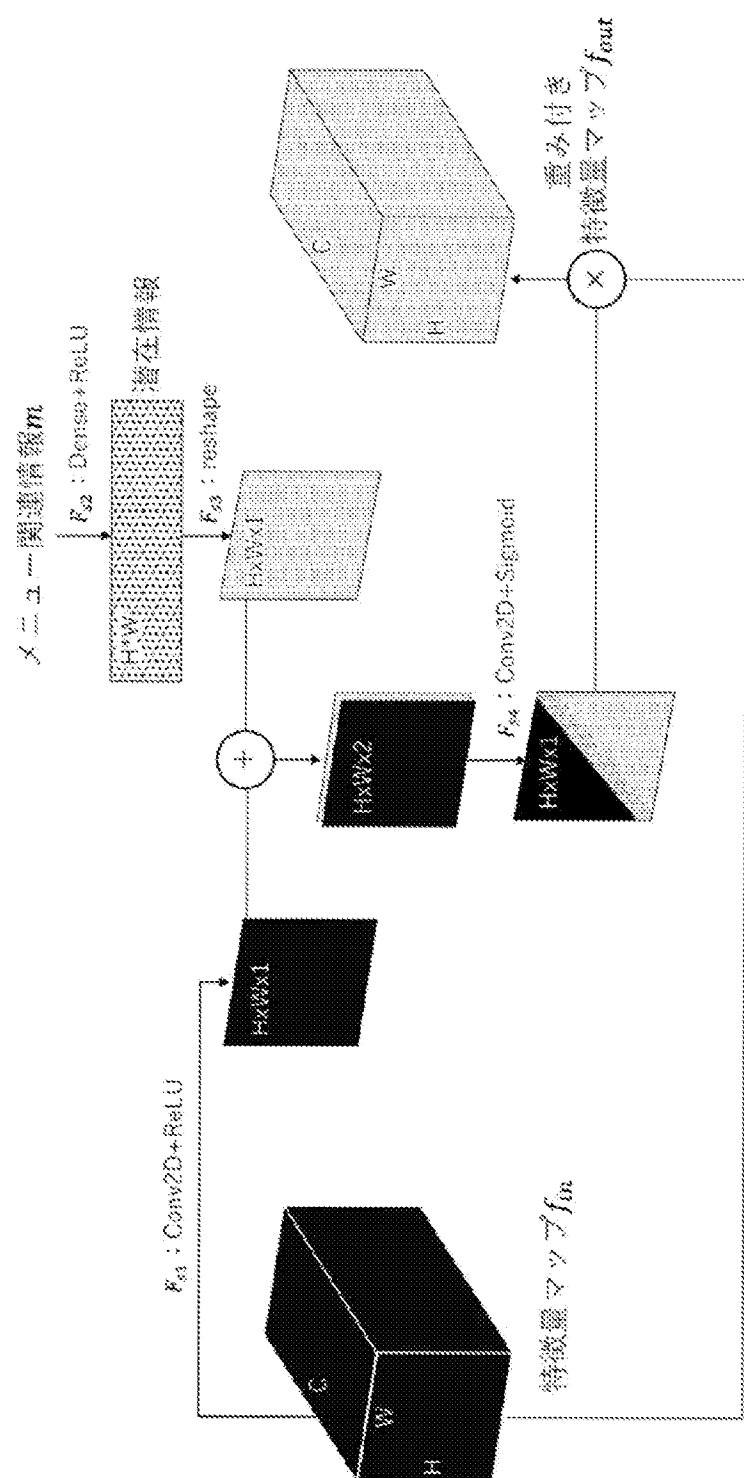
[図7]

食べ残し品目	食べ残し割合
ポテト	80%
ブロッコリー	100%
...	...

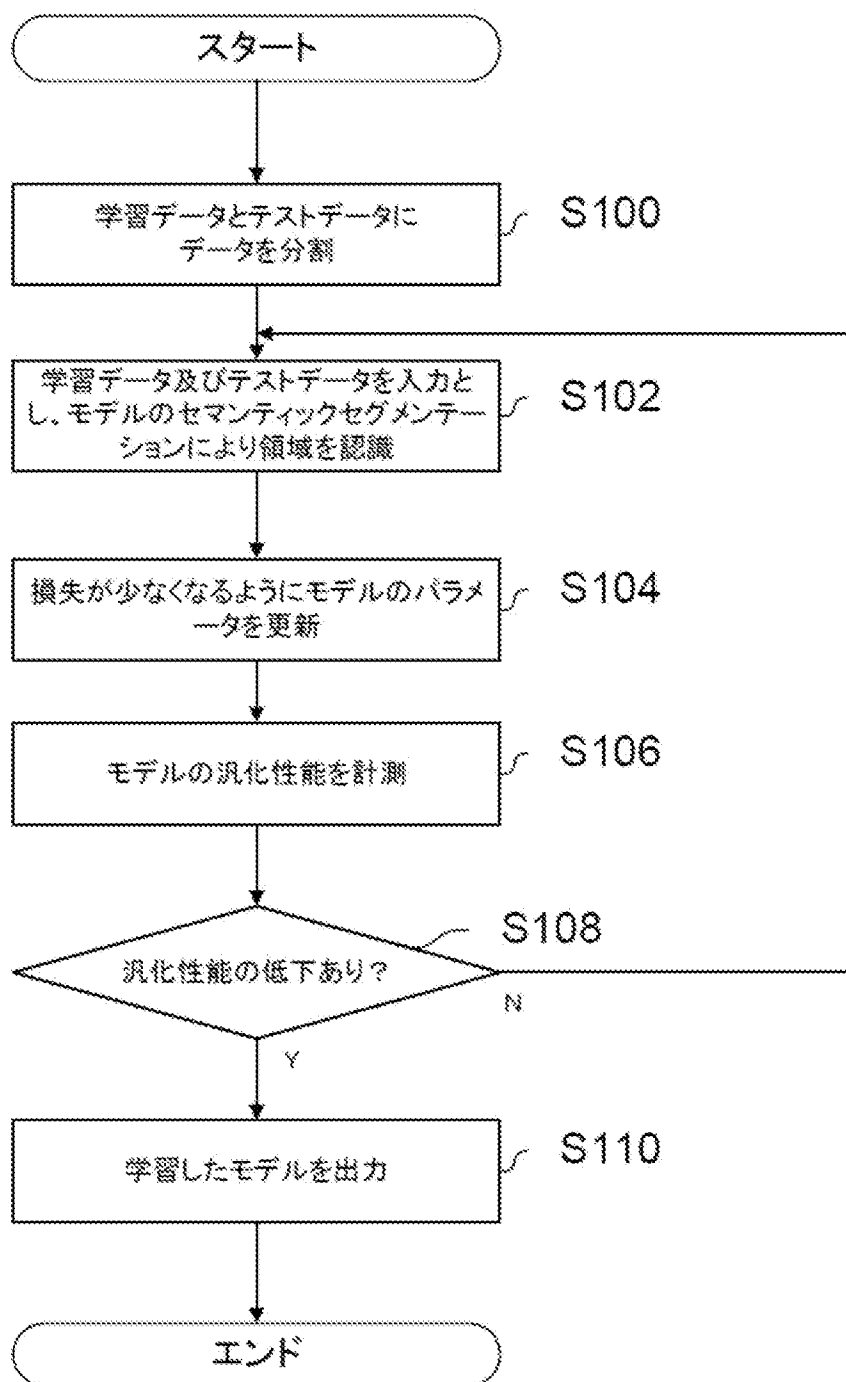
[図8]



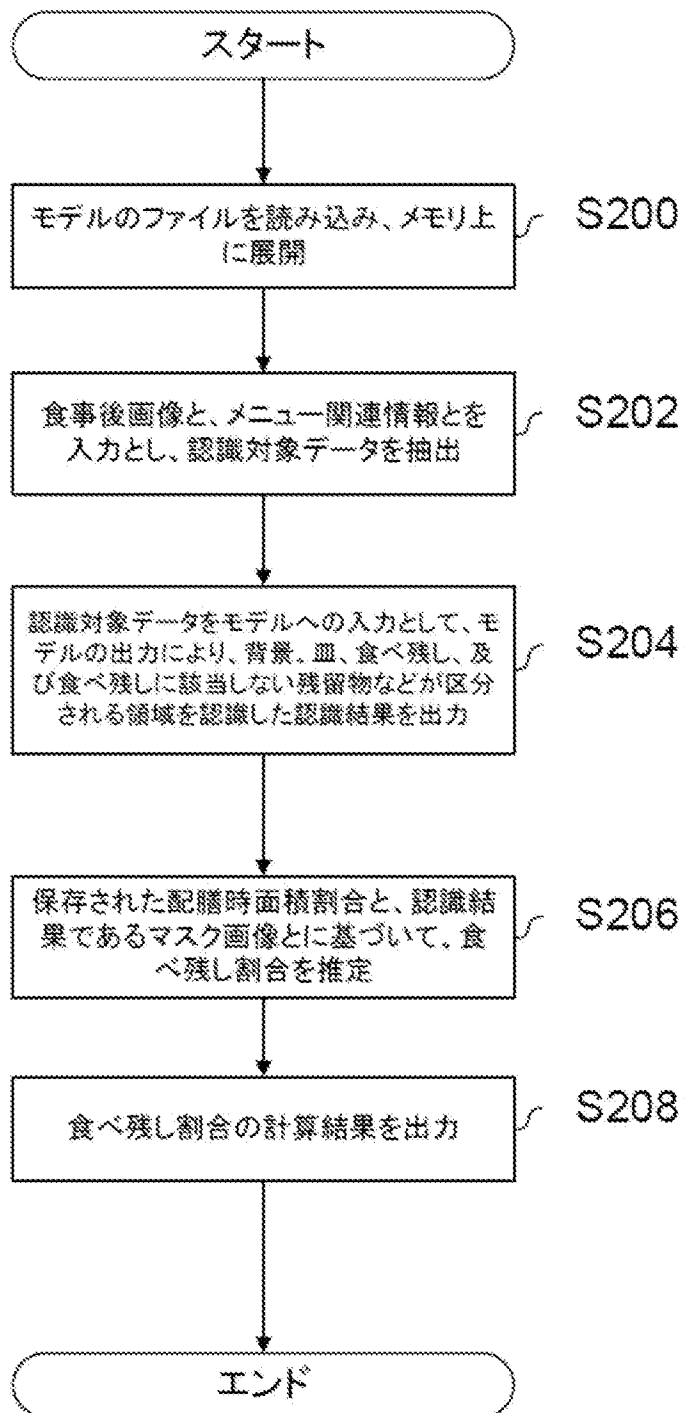
[図10]



[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/017091

A. CLASSIFICATION OF SUBJECT MATTER**G06T 7/00**(2017.01)i

FI: G06T7/00 350C

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T7/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2015-138452 A (TOSHIBA TEC KK) 30 July 2015 (2015-07-30) entire text, all drawings	1-8
A	横山 新, 宿泊施設のピュツフェを対象とした追加調理指示システムの構築, 電子情 通信学会技術研究報告, 12 September 2019, vol. 119, no. 206, pp. 13-18 entire text, all drawings, (YOKOYAMA, Arata. Construction of Additional Cooking Order System for Accommodation Buffet. IEICE Technical Report.)	1-8

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

24 June 2021

Date of mailing of the international search report

06 July 2021

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2021/017091

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2015-138452 A	30 July 2015	US 2015/0206259 A1	
entire text, all drawings			

A. 発明の属する分野の分類（国際特許分類（IPC）） G06T 7/00(2017.01)i FI: G06T7/00 350C														
B. 調査を行った分野														
調査を行った最小限資料（国際特許分類（IPC）） G06T7/00														
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2021年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2021年	日本国実用新案登録公報	1996 - 2021年	日本国登録実用新案公報	1994 - 2021年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2021年													
日本国実用新案登録公報	1996 - 2021年													
日本国登録実用新案公報	1994 - 2021年													
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）														
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
A	JP 2015-138452 A（東芝テック株式会社）30.07.2015（2015 - 07 - 30） 全文、全図	1-8												
A	横山 新, 宿泊施設のビュッフェを対象とした追加調理指示システムの構築, 電子情報通信学会技術研究報告, 2019.09.12, Vol.119 No.206, p.13-p.18 全文、全図	1-8												
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 24.06.2021	国際調査報告の発送日 06.07.2021													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 山田 辰美 5H 8383 電話番号 03-3581-1101 内線 3531													

国際出願番号
PCT/JP2021/017091

引用文献	公表日	パテントファミリー文献	公表日
JP2015-138452A	30.07.2015	US2015/0206259A1	
全文, 全図			