

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7648638号
(P7648638)

(45)発行日 令和7年3月18日(2025.3.18)

(24)登録日 令和7年3月10日(2025.3.10)

(51)国際特許分類 F I
G 0 6 N 3/084(2023.01) G 0 6 N 3/084
G 0 6 N 3/0464(2023.01) G 0 6 N 3/0464

請求項の数 15 (全25頁)

(21)出願番号	特願2022-545299(P2022-545299)	(73)特許権者	306037311 富士フイルム株式会社 東京都港区西麻布2丁目26番30号
(86)(22)出願日	令和3年4月20日(2021.4.20)	(74)代理人	100140992 弁理士 松浦 憲政
(86)国際出願番号	PCT/JP2021/016004	(74)代理人	100170069 弁理士 大原 一樹
(87)国際公開番号	WO2022/044425	(74)代理人	100128635 弁理士 松村 潔
(87)国際公開日	令和4年3月3日(2022.3.3)	(74)代理人	100153822 弁理士 増田 重之
審査請求日	令和5年4月5日(2023.4.5)	(74)代理人	100083116 弁理士 松浦 憲三
(31)優先権主張番号	特願2020-144568(P2020-144568)	(72)発明者	大酒 正明 東京都港区西麻布2丁目26番30号
(32)優先日	令和2年8月28日(2020.8.28)		最終頁に続く
(33)優先権主張国・地域又は機関	日本国(JP)		

(54)【発明の名称】 学習装置、学習方法、プログラム、学習済みモデル、及び内視鏡システム

(57)【特許請求の範囲】

【請求項1】

認識器の学習モデルと前記学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置において、

前記学習モデルは、

第1の条件で取得された複数のデータで構成される第1のデータ群から選択された第1のデータが入力されて、第1の特徴量を出力する第1の入力層と、

前記第1の入力層とは独立した第2の入力層であって、前記第1のデータ群を構成するデータと同一のカテゴリに属し前記第1の条件とは異なる第2の条件で取得された複数のデータで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量を出力する第2の入力層と、

前記第1の入力層及び前記第2の入力層に対して共通の中間層であって、前記第1の特徴量が入力された場合には第1の中間特徴量を出力し、前記第2の特徴量が入力された場合には第2の中間特徴量を出力する第1の中間層と、

前記第1の中間特徴量が入力され、前記第1の中間特徴量に基づく第1の正規化特徴量を出力する第1の正規化層と、

前記第2の中間特徴量が入力され、前記第2の中間特徴量に基づく第2の正規化特徴量を出力する第2の正規化層と、

前記第1の正規化層及び前記第2の正規化層に対して共通の中間層であって、前記第1の正規化特徴量が入力された場合には第3の中間特徴量を出力し、前記第2の正規化特徴

量が入力された場合には第 4 の中間特徴量を出力する第 2 の中間層と、

前記第 3 の中間特徴量または前記第 4 の中間特徴量が入力され、前記第 3 の中間特徴量が入力された場合には前記第 3 の中間特徴量に基づく第 1 の認識結果を出力し、前記第 4 の中間特徴量が入力された場合には前記第 4 の中間特徴量に基づく第 2 の認識結果を出力する出力層と、

を含む階層型ネットワークを備え、

前記学習制御部は、前記第 1 の認識結果と前記第 1 のデータの正解との第 1 の誤差に基づいて前記学習モデルを学習させる第 1 の学習、及び前記第 2 の認識結果と前記第 2 のデータの正解との第 2 の誤差に基づいて前記学習モデルを学習させる第 2 の学習を行わせる学習装置であって、

10

前記第 1 の入力層は前記第 1 の条件で取得された第 1 の画像データを前記第 1 のデータとして入力し、

前記第 2 の入力層は前記第 1 の条件とは異なる前記第 2 の条件で取得された第 2 の画像データを前記第 2 のデータとして入力し、

前記第 1 の入力層は、白色光で取得された前記第 1 の画像データを入力し、

前記第 2 の入力層は前記白色光よりも狭帯域である光で取得された前記第 2 の画像データを入力し、

前記第 1 の正規化層及び前記第 2 の正規化層では、色に関してそれぞれ異なる条件で正規化を行う、学習装置。

【請求項 2】

20

前記学習制御部は、少なくとも前記第 1 の学習を 2 回行わせ、

前記第 2 の中間層は、1 回目の前記第 1 の学習における前記第 3 の中間特徴量が出力された後であって、2 回目の前記第 1 の学習における前記第 3 の中間特徴量が出力される前の期間に、前記第 2 の学習における前記第 4 の中間特徴量を出力する請求項 1 に記載の学習装置。

【請求項 3】

前記学習制御部は、少なくとも前記第 1 の学習を 2 回行わせ、

前記第 2 の中間層は、1 回目の前記第 1 の学習における前記第 3 の中間特徴量の出力、及び 2 回目の前記第 1 の学習における前記第 3 の中間特徴量の出力が完了した後に、前記第 2 の学習における前記第 4 の中間特徴量を出力する請求項 1 に記載の学習装置。

30

【請求項 4】

前記階層型ネットワークは、畳み込みニューラルネットワークである請求項 1 から 3 のいずれか 1 項に記載の学習装置。

【請求項 5】

前記第 1 の正規化層はバッチノーマライゼーション処理により前記第 1 の正規化特徴量を算出し、及び前記第 2 の正規化層はバッチノーマライゼーション処理により前記第 2 の正規化特徴量を算出する請求項 1 から 4 のいずれか 1 項に記載の学習装置。

【請求項 6】

前記第 1 の入力層は、畳み込み演算、プーリング処理、バッチノーマライゼーション処理、活性化処理のいずれか一つを含む演算によって前記第 1 の特徴量を出力する請求項 1 から 5 のいずれか 1 項に記載の学習装置。

40

【請求項 7】

前記第 2 の入力層は、畳み込み演算、プーリング処理、バッチノーマライゼーション処理、活性化処理のいずれか一つを含む演算によって前記第 2 の特徴量を出力する請求項 1 から 6 のいずれか 1 項に記載の学習装置。

【請求項 8】

前記第 1 の中間層は、畳み込み演算、プーリング処理、及び活性化処理のいずれか一つを含む演算によって前記第 1 の中間特徴量または前記第 2 の中間特徴量を出力する請求項 1 から 7 のいずれか 1 項に記載の学習装置。

【請求項 9】

50

前記第 2 の中間層は、畳み込み演算、プーリング処理、及び活性化処理のいずれか一つを含む演算によって前記第 3 の中間特徴量または前記第 4 の中間特徴量を出力する請求項 1 から 8 のいずれか 1 項に記載の学習装置。

【請求項 1 0】

前記第 1 の入力層は、前記白色光よりも狭帯域である光である第 1 の光で取得された画像データを前記第 1 の画像データとして入力し、

前記第 2 の入力層は前記第 1 の光とは異なる第 2 の光で取得された画像データを前記第 2 の画像データとして入力する請求項 1 から 9 のいずれか 1 項に記載の学習装置。

【請求項 1 1】

認識器の学習モデルと前記学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置の学習方法であって、

10

前記学習モデルは、第 1 の条件で取得された複数のデータで構成される第 1 のデータ群から選択された第 1 のデータが入力されて、第 1 の特徴量を出力する第 1 の入力層と、前記第 1 の入力層とは独立した第 2 の入力層であって、前記第 1 のデータ群を構成するデータと同一のカテゴリに属し前記第 1 の条件とは異なる第 2 の条件で取得された複数のデータで構成される第 2 のデータ群から選択された第 2 のデータが入力されて、第 2 の特徴量を出力する第 2 の入力層と、前記第 1 の入力層及び前記第 2 の入力層に対して共通の中間層であって、前記第 1 の特徴量が入力された場合には第 1 の中間特徴量を出力し、前記第 2 の特徴量が入力された場合には第 2 の中間特徴量を出力する第 1 の中間層と、前記第 1 の中間特徴量が入力され、前記第 1 の中間特徴量に基づく第 1 の正規化特徴量を出力する第 1 の正規化層と、前記第 2 の中間特徴量が入力され、前記第 2 の中間特徴量に基づく第 2 の正規化特徴量を出力する第 2 の正規化層と、前記第 1 の正規化層及び前記第 2 の正規化層に対して共通の中間層であって、前記第 1 の正規化特徴量が入力された場合には第 3 の中間特徴量を出力し、前記第 2 の正規化特徴量が入力された場合には第 4 の中間特徴量を出力する第 2 の中間層と、前記第 3 の中間特徴量または前記第 4 の中間特徴量が入力され、前記第 3 の中間特徴量が入力された場合には前記第 3 の中間特徴量に基づく第 1 の認識結果を出力し、前記第 4 の中間特徴量が入力された場合には前記第 4 の中間特徴量に基づく第 2 の認識結果を出力する出力層と、を含む階層型ネットワークを備え、

20

前記学習制御部により、

前記第 1 の認識結果と前記第 1 のデータの正解との第 1 の誤差に基づいて前記学習モデルを学習させる第 1 の学習工程と、

30

前記第 2 の認識結果と前記第 2 のデータの正解との第 2 の誤差に基づいて前記学習モデルを学習させる第 2 の学習工程と、

を含む学習方法であって、

前記第 1 の入力層は前記第 1 の条件で取得された第 1 の画像データを前記第 1 のデータとして入力し、

前記第 2 の入力層は前記第 1 の条件とは異なる前記第 2 の条件で取得された第 2 の画像データを前記第 2 のデータとして入力し、

前記第 1 の入力層は、白色光で取得された前記第 1 の画像データを入力し、

前記第 2 の入力層は前記白色光よりも狭帯域である光で取得された前記第 2 の画像データを入力し、

40

前記第 1 の正規化層及び前記第 2 の正規化層では、色に関してそれぞれ異なる条件で正規化を行う、学習方法。

【請求項 1 2】

認識器の学習モデルと前記学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置の学習方法を実行するプログラムであって、

前記学習モデルは、第 1 の条件で取得された複数のデータで構成される第 1 のデータ群から選択された第 1 のデータが入力されて、第 1 の特徴量を出力する第 1 の入力層と、前記第 1 の入力層とは独立した第 2 の入力層であって、前記第 1 のデータ群を構成するデータと同一のカテゴリに属し前記第 1 の条件とは異なる第 2 の条件で取得された複数のデー

50

タで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量
 を出力する第2の入力層と、前記第1の入力層及び前記第2の入力層に対して共通の中間
 層であって、前記第1の特徴量が入力された場合には第1の中間特徴量を出力し、前記第
 2の特徴量が入力された場合には第2の中間特徴量を出力する第1の中間層と、前記第1
 の中間特徴量が入力され、前記第1の中間特徴量に基づく第1の正規化特徴量を出力する
 第1の正規化層と、前記第2の中間特徴量が入力され、前記第2の中間特徴量に基づく第
 2の正規化特徴量を出力する第2の正規化層と、前記第1の正規化層及び前記第2の正規
 化層に対して共通の中間層であって、前記第1の正規化特徴量が入力された場合には第3
 の中間特徴量を出力し、前記第2の正規化特徴量が入力された場合には第4の中間特徴量
 を出力する第2の中間層と、前記第3の中間特徴量または前記第4の中間特徴量が入力さ
 れ、前記第3の中間特徴量が入力された場合には前記第3の中間特徴量に基づく第1の認
 識結果を出力し、前記第4の中間特徴量が入力された場合には前記第4の中間特徴量に基
 づく第2の認識結果を出力する出力層と、を含む階層型ネットワークを備え、

10

前記学習制御部により、

前記第1の認識結果と前記第1のデータの正解との第1の誤差に基づいて前記学習モデ
 ルを学習させる第1の学習工程と、

前記第2の認識結果と前記第2のデータの正解との第2の誤差に基づいて前記学習モデ
 ルを学習させる第2の学習工程と、

を含む学習方法を実行させるプログラムであって、

前記第1の入力層は前記第1の条件で取得された第1の画像データを前記第1のデータ
 として入力し、

20

前記第2の入力層は前記第1の条件とは異なる前記第2の条件で取得された第2の画像
 データを前記第2のデータとして入力し、

前記第1の入力層は、白色光で取得された前記第1の画像データを入力し、

前記第2の入力層は前記白色光よりも狭帯域である光で取得された前記第2の画像デー
 タを入力し、

前記第1の正規化層及び前記第2の正規化層では、色に関してそれぞれ異なる条件で正
 規化を行う、プログラム。

【請求項13】

認識器の学習モデルと前記学習モデルを学習させる学習制御部とを構成するプロセッサ
 を備えた学習装置の学習方法を実行することによって得られるプログラムで構成される学
 習済みモデルであって、

30

前記学習モデルは、第1の条件で取得された複数のデータで構成される第1のデータ群
 から選択された第1のデータが入力されて、第1の特徴量を出力する第1の入力層と、前
 記第1の入力層とは独立した第2の入力層であって、前記第1のデータ群を構成するデー
 タと同一のカテゴリに属し前記第1の条件とは異なる第2の条件で取得された複数のデー
 タで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量
 を出力する第2の入力層と、前記第1の入力層及び前記第2の入力層に対して共通の中間
 層であって、前記第1の特徴量が入力された場合には第1の中間特徴量を出力し、前記第
 2の特徴量が入力された場合には第2の中間特徴量を出力する第1の中間層と、前記第1
 の中間特徴量が入力され、前記第1の中間特徴量に基づく第1の正規化特徴量を出力する
 第1の正規化層と、前記第2の中間特徴量が入力され、前記第2の中間特徴量に基づく第
 2の正規化特徴量を出力する第2の正規化層と、前記第1の正規化層及び前記第2の正規
 化層に対して共通の中間層であって、前記第1の正規化特徴量が入力された場合には第3
 の中間特徴量を出力し、前記第2の正規化特徴量が入力された場合には第4の中間特徴量
 を出力する第2の中間層と、前記第3の中間特徴量または前記第4の中間特徴量が入力さ
 れ、前記第3の中間特徴量が入力された場合には前記第3の中間特徴量に基づく第1の認
 識結果を出力し、前記第4の中間特徴量が入力された場合には前記第4の中間特徴量に基
 づく第2の認識結果を出力する出力層と、を含む階層型ネットワークとして構成され、
 前記学習済みモデルは、

40

50

前記学習制御部により行われる、

前記第1の認識結果と前記第1のデータの正解との第1の誤差に基づいて前記学習モデルを学習させる第1の学習工程と、

前記第2の認識結果と前記第2のデータの正解との第2の誤差に基づいて前記学習モデルを学習させる第2の学習工程と、によって、学習されたものであり、

前記第1の入力層は前記第1の条件で取得された第1の画像データを前記第1のデータとして入力し、

前記第2の入力層は前記第1の条件とは異なる前記第2の条件で取得された第2の画像データを前記第2のデータとして入力し、

前記第1の入力層は、白色光で取得された前記第1の画像データを入力し、

前記第2の入力層は前記白色光よりも狭帯域である光で取得された前記第2の画像データを入力し、

前記第1の正規化層及び前記第2の正規化層では、色に関してそれぞれ異なる条件で正規化を行うよう、コンピュータを機能させるための学習済みモデル。

【請求項14】

請求項13に記載の前記認識器の学習済みモデルを搭載する内視鏡システム。

【請求項15】

前記第1の条件と前記第2の条件とでは、撮像装置、観察光の波長バランス、解像度、及び画像に施す画像処理のうち少なくとも1つが異なる請求項14に記載の内視鏡システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、学習装置、学習方法、プログラム、学習済みモデル、及び内視鏡システムに関し、特に階層型ネットワークを用いて学習を行う学習装置、学習方法、プログラム、学習済みモデル、及び内視鏡システムに関する。

【背景技術】

【0002】

機械学習の分野では階層型ネットワークを用いて学習を行うことが知られている。階層型ネットワークは一般に特徴抽出や認識等を行う複数の層から構成されるが、具体的なネットワーク構成や学習方法には種々の態様が存在する。

【0003】

例えば、特許文献1には、互いに異なる条件で取得された第1のデータ群と第2のデータ群を適切に学習することを目的とした学習装置が記載されている。具体的には、互いに異なる条件で取得された第1のデータ群と第2のデータ群とを、互いに独立した第1の入力層及び第2の入力層にそれぞれ入力し、第1の入力層及び第2の入力層に対して共通の中間層を設けた階層型ネットワークが記載されている。

【0004】

さらに、機械学習を進める上で、非特許文献1に記載されているように、算出された特徴量を正規化することで、認識器の精度を向上させる技術が知られている。

【先行技術文献】

【特許文献】

【0005】

【文献】国際公開第2020/022027号公報

【非特許文献】

【0006】

【文献】Sergey Ioffe, Christian Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", [online], 2015年3月2日, Cornell University, arXiv:1502.03167v3[cs.LG], (2020年8月17日検索), インターネット URL : <https://arxiv.org/abs/1502.03167>

10

20

30

40

50

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、上述した特許文献1に記載された階層型ネットワークのように、互いに異なる条件で取得された第1のデータ群と第2のデータ群とで得られたそれぞれの特徴量に対して、非特許文献1に記載された技術のような正規化を適切に適用することができないという問題がある。特徴量の正規化を行う場合には、本来は同じ条件で取得されたデータ群毎に、異なる正規化を行うべきである。しかしながら、特許文献1に記載された階層型ネットワークでは、中間層を共通とする構成にしているため、中間層から出力された特徴量に対して入力されたデータ毎に異なる条件で正規化を行うことができず、効率的な学習が行えない場合がある。

10

【0008】

本発明はこのような事情に鑑みてなされたもので、その目的は、互いに異なる条件で取得されたデータを使用して学習を行う場合であっても、効率の良い学習を行うことができる学習装置、学習方法、プログラム、学習済みモデル、及び内視鏡システムを提供することである。

【課題を解決するための手段】

【0009】

上記目的を達成するための本発明の一の態様である学習装置は、認識器の学習モデルと学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置において、学習モデルは、第1の条件で取得された複数のデータで構成される第1のデータ群から選択された第1のデータが入力されて、第1の特徴量を出力する第1の入力層と、第1の入力層とは独立した第2の入力層であって、第1のデータ群を構成するデータと同一のカテゴリに属し第1の条件とは異なる第2の条件で取得された複数のデータで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量を出力する第2の入力層と、第1の入力層及び第2の入力層に対して共通の中間層であって、第1の特徴量が入力された場合には第1の中間特徴量を出力し、第2の特徴量が入力された場合には第2の中間特徴量を出力する第1の中間層と、第1の中間特徴量が入力され、第1の中間特徴量に基づく第1の正規化特徴量を出力する第1の正規化層と、第2の中間特徴量が入力され、第2の中間特徴量に基づく第2の正規化特徴量を出力する第2の正規化層と、第1の正規化層及び第2の正規化層に対して共通の中間層であって、第1の正規化特徴量が入力された場合には第3の中間特徴量を出力し、第2の正規化特徴量が入力された場合には第4の中間特徴量を出力する第2の中間層と、第3の中間特徴量または第4の中間特徴量が入力され、第3の中間特徴量が入力された場合には第3の中間特徴量に基づく第1の認識結果を出力し、第4の中間特徴量が入力された場合には第4の中間特徴量に基づく第2の認識結果を出力する出力層と、を含む階層型ネットワークを備える学習装置であって、学習制御部は、第1の認識結果と第1のデータの正解との第1の誤差に基づいて学習モデルを学習させる第1の学習、及び第2の認識結果と第2のデータの正解との第2の誤差に基づいて学習モデルを学習させる第2の学習を行わせる。

20

30

【0010】

本態様では、第1の中間層は、第1のデータに基づく第1の特徴量が入力された場合には第1の中間特徴量を出力し、第2のデータに基づく第2の特徴量が入力された場合には第2の中間特徴量を出力する。そして、第1の正規化層は第1の中間特徴量を入力し第1の正規化特徴量を出力し、第2の正規化層は第2の中間特徴量を入力し第2の正規化特徴量を出力する。第2の中間層は、第1の正規化特徴量及び第2の正規化特徴量を入力する。これにより、本態様は、第1のデータに由来する第1の中間特徴量と第2のデータに由来する第2の中間特徴量とを別々の条件で正規化することができるので、第1の中間特徴量と第2の中間特徴量とを適切に正規化することができ、効率的な学習を行うことができる。

40

【0011】

また、本態様では、独立した第1、第2の入力層に第1、第2のデータをそれぞれ入力

50

し、第1、第2の入力層でそれぞれ特徴量を算出することで、第1、第2の入力層の一方における特徴量算出が他方の入力層における特徴量算出の影響を受けないようにしている。また本態様では、入力層（第1の入力層及び第2の入力層）における特徴抽出に加えて、さらに第1、第2の入力層に共通な第1の中間層において第1の中間特徴量と第2の中間特徴量とが算出されるので、入力層で第1、第2のデータから算出した特徴量を第1の中間層における中間特徴量算出に反映することができる。第2の中間層も、第1の正規化層及び第2の正規化層に共通であるので、同様に、第1の正規特徴量及び第2の正規化特徴量を第2の中間層における中間特徴量算出に反映することができる。また、階層型ネットワークはパラメータが多いため過学習になりがちであるが、大量にデータを与えることで過学習を回避できる。本態様に係る学習装置では、中間層は第1、第2のデータを合わせた大量のデータで学習できるため過学習になりやすく、一方、入力層は第1、第2の入力層に独立しておりそれぞれの入力層のパラメータは少なくなるため、少量のデータでも過学習になりにくい。本態様によれば、このようにして同一のカテゴリに属し異なる条件で取得されたデータを適切に学習することができる。

10

【0012】

なお本態様及び以下の各態様において、「第1、第2の入力層から出力される特徴量に基づく第1、第2の特徴量」については、第1、第2の入力層から出力される特徴量をそのまま第1、第2の特徴量として入力してもよいし、第1、第2の入力層から出力される特徴量に何らかの処理を施した特徴量を第1、第2の特徴量として入力してもよい。また、「同一のカテゴリに属し」とは、画像と画像、テキストとテキスト、音声と音声のような組み合わせを意味し、「第1の条件と第2の条件が異なる」には「同じ条件で取得されたデータを2つに分ける」ことは含まれない。

20

【0013】

また、本態様及び以下の各態様において、第1、第2の入力層、及び中間層は1つの層で構成されていてもよいし、複数の層から構成されていてもよい。また、第1、第2の入力層を構成する層の数は同じでもよいし、違っていてもよい。階層型ネットワークは、第1、第2の入力層、中間層の他に出力層、認識層等を含んでいてもよい。

【0014】

また、本態様及び以下の各態様において、第1、第2の入力層から出力される特徴量が第1、第2のデータの特徴を適切に表現できるように、学習の結果（例えば、認識結果と正解データとの誤差、損失等）を考慮して第1、第2の入力層の層数や各層におけるパラメータを調整することが好ましい。また、中間層についても、同様に学習の結果を考慮して中間層の層数や各層におけるパラメータを調整することが好ましい。

30

【0015】

好ましくは、学習制御部は、少なくとも第1の学習を2回行わせ、第2の中間層は、1回目の第1の学習における第3の中間特徴量が出力された後であって、2回目の第1の学習における第3の中間特徴量が出力される前の期間に、第2の学習における第4の中間特徴量を出力する。

【0016】

第1の学習を多数回連続して行いその後第2の学習を行う場合、中間層で算出される特徴量が第1のデータの影響を強く受けてしまい第2のデータに対する学習（特徴量の算出）が適切に行われない可能性がある（逆の場合も同様である）。このため本態様では、第3の中間特徴量の算出が終了してから他の第3の中間特徴量の算出が始まるまでの期間に第4の中間特徴量の算出を実行しており、これにより第4の中間特徴量の算出の際に算出される特徴量が第1のデータの影響を過度に受けることを避け、第1、第2のデータに対して適切に学習を行うことができる。

40

【0017】

好ましくは、学習制御部は、少なくとも第1の学習を2回行わせ、第2の中間層は、1回目の第1の学習における第3の中間特徴量の出力、及び2回目の第1の学習における第3の中間特徴量の出力が完了した後に、第2の学習における第4の中間特徴量を出力する。

50

【 0 0 1 8 】

本態様では、上述したのと同様に第 3 の中間特徴量の算出の際に算出される特徴量が第 1 のデータの影響を過度に受けることを避け、第 1、第 2 のデータに対して適切に学習を行うことができる。

【 0 0 1 9 】

好ましくは、階層型ネットワークは、畳み込みニューラルネットワークである。

【 0 0 2 0 】

好ましくは、第 1 の正規化層はバッチノーマライゼーション処理により第 1 の正規化特徴量を算出し、及び第 2 の正規化層はバッチノーマライゼーション処理により第 2 の正規化特徴量を算出する。

10

【 0 0 2 1 】

好ましくは、第 1 の入力層は、畳み込み演算、プーリング処理、バッチノーマライゼーション処理、活性化処理のいずれか一つを含む演算によって第 1 の特徴量を出力する。

【 0 0 2 2 】

好ましくは、第 2 の入力層は、畳み込み演算、プーリング処理、バッチノーマライゼーション処理、活性化処理のいずれか一つを含む演算によって第 2 の特徴量を出力する。

【 0 0 2 3 】

好ましくは、第 1 の中間層は、畳み込み演算、プーリング処理、及び活性化処理のいずれか一つを含む演算によって第 1 の中間特徴量または第 2 の中間特徴量を出力する。

【 0 0 2 4 】

好ましくは、第 2 の中間層は、畳み込み演算、プーリング処理、及び活性化処理のいずれか一つを含む演算によって第 3 の中間特徴量または第 4 の中間特徴量を出力する。

20

【 0 0 2 5 】

好ましくは、第 1 の入力層は第 1 の条件で取得された第 1 の画像データを第 1 のデータとして入力し、第 2 の入力層は第 1 の条件とは異なる第 2 の条件で取得された第 2 の画像データを第 2 のデータとして入力する。

【 0 0 2 6 】

好ましくは、第 1 の条件と第 2 の条件とでは、撮像装置、観察光の波長バランス、解像度、及び画像に施す画像処理のうち少なくとも 1 つが異なる。

【 0 0 2 7 】

なお、本態様において「撮像装置が異なる」とは、「モダリティは同じで、機種、型番、性能等が異なる」ことを意味するものとする。例えば、内視鏡装置と CT 装置ではモダリティが異なる。また、「観察光の波長バランスが異なる」とは、観察光の波長帯域及び/または観察光における各波長帯域の強度の相対関係が異なることを意味する。また、「画像に施す画像処理が異なる」には、例えば特定の波長成分の影響を強調または低減する処理、あるいは特定の対象や領域を強調または目立たなくする処理が含まれるが、これらに限定されるものではない。

30

【 0 0 2 8 】

好ましくは、第 1 の入力層は第 1 の観察光により取得された第 1 の医用画像のデータを第 1 の画像データとして入力し、第 2 の入力層は第 1 の観察光と波長バランスが異なる第 2 の観察光により取得された第 2 の医用画像のデータを第 2 の画像データとして入力する。

40

【 0 0 2 9 】

「撮影された画像で被写体のどのような構造が明確に（あるいは不明確に）映るか」は撮影に用いる観察光の波長バランスに依存するため、診断や診察の場面では波長バランスが異なる複数の観察光により画像を取得する場合があるが、本態様ではそのような場合でも画像の学習を適切に行うことができる。なお本態様及び以下の各態様において、「医用画像」は「医療画像」ともいう。

【 0 0 3 0 】

好ましくは、第 1 の入力層は白色光を第 1 の観察光として取得された第 1 の医用画像のデータを第 1 の画像データとして入力し、第 2 の入力層は狭帯域光を第 2 の観察光として

50

取得された第2の医用画像のデータを第2の画像データとして入力する。

【0031】

医用画像を取得する場合、ユーザの目視による確認等のため白色光を観察光とする画像を取得することが多い。一方、狭帯域光の場合、波長により被検体の細部や深部等、白色光画像と異なる構造を観察できるが、目視観察には適していないため白色光画像と比べて取得される画像の数が少ない。本態様では、このような場合でも適切に学習を行うことができる。なお、本態様において「狭帯域光」は青色光、紫色光等短波長の観察光でもよいし、赤色光、赤外光等長波長の観察光でもよい。

【0032】

好ましくは、第1の入力層は第1の狭帯域光を第1の観察光として取得された第1の医用画像のデータを第1の画像データとして入力し、第2の入力層は第1の狭帯域光とは異なる第2の狭帯域光を第2の観察光として取得された第2の医用画像のデータを第2の画像データとして入力する。

10

【0033】

医用画像を取得する場合、画像の利用目的によっては観察光として複数の狭帯域光を用いて画像を取得することがあるが、本態様によればそのような場合でも適切に学習を行うことができる。なお「第1の狭帯域光とは異なる第2の狭帯域光」とは、第1の狭帯域光と第2の狭帯域光とで観察光の波長帯域及び/または観察光の強度が異なることを意味する。

【0034】

本発明の他の態様である学習方法は、認識器の学習モデルと学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置の学習方法であって、学習モデルは、第1の条件で取得された複数のデータで構成される第1のデータ群から選択された第1のデータが入力されて、第1の特徴量を出力する第1の入力層と、第1の入力層とは独立した第2の入力層であって、第1のデータ群を構成するデータと同一のカテゴリに属し第1の条件とは異なる第2の条件で取得された複数のデータで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量を出力する第2の入力層と、第1の入力層及び第2の入力層に対して共通の中間層であって、第1の特徴量が入力された場合には第1の中間特徴量を出力し、第2の特徴量が入力された場合には第2の中間特徴量を出力する第1の中間層と、第1の中間特徴量が入力され、第1の中間特徴量に基づく第1の正規化特徴量を出力する第1の正規化層と、第2の中間特徴量が入力され、第2の中間特徴量に基づく第2の正規化特徴量を出力する第2の正規化層と、第1の正規化層及び第2の正規化層に対して共通の中間層であって、第1の正規化特徴量が入力された場合には第3の中間特徴量を出力し、第2の正規化特徴量が入力された場合には第4の中間特徴量を出力する第2の中間層と、第3の中間特徴量または第4の中間特徴量が入力され、第3の中間特徴量が入力された場合には第3の中間特徴量に基づく第1の認識結果を出力し、第4の中間特徴量が入力された場合には第4の中間特徴量に基づく第2の認識結果を出力する出力層と、を含む階層型ネットワークを備える学習装置の学習方法であって、学習制御部により、第1の認識結果と第1のデータの正解との第1の誤差に基づいて学習モデルを学習させる第1の学習工程と、第2の認識結果と第2のデータの正解との第2の誤差に基づいて学習モデルを学習させる第2の学習工程と、を含む。

20

30

40

【0035】

本発明の他の態様であるプログラムは、認識器の学習モデルと学習モデルを学習させる学習制御部とを構成するプロセッサを備えた学習装置の学習方法を実行するプログラムであって、学習モデルは、第1の条件で取得された複数のデータで構成される第1のデータ群から選択された第1のデータが入力されて、第1の特徴量を出力する第1の入力層と、第1の入力層とは独立した第2の入力層であって、第1のデータ群を構成するデータと同一のカテゴリに属し第1の条件とは異なる第2の条件で取得された複数のデータで構成される第2のデータ群から選択された第2のデータが入力されて、第2の特徴量を出力する第2の入力層と、第1の入力層及び第2の入力層に対して共通の中間層であって、第1の

50

特徴量が入力された場合には第 1 の中間特徴量を出力し、第 2 の特徴量が入力された場合には第 2 の中間特徴量を出力する第 1 の中間層と、第 1 の中間特徴量が入力され、第 1 の中間特徴量に基づく第 1 の正規化特徴量を出力する第 1 の正規化層と、第 2 の中間特徴量が入力され、第 2 の中間特徴量に基づく第 2 の正規化特徴量を出力する第 2 の正規化層と、第 1 の正規化層及び第 2 の正規化層に対して共通の中間層であって、第 1 の正規化特徴量が入力された場合には第 3 の中間特徴量を出力し、第 2 の正規化特徴量が入力された場合には第 4 の中間特徴量を出力する第 2 の中間層と、第 3 の中間特徴量または第 4 の中間特徴量が入力され、第 3 の中間特徴量が入力された場合には第 3 の中間特徴量に基づく第 1 の認識結果を出力し、第 4 の中間特徴量が入力された場合には第 4 の中間特徴量に基づく第 2 の認識結果を出力する出力層と、を含む階層型ネットワークを備える学習装置の学習方法を実行するプログラムであって、学習制御部により、第 1 の認識結果と第 1 のデータの正解との第 1 の誤差に基づいて学習モデルを学習させる第 1 の学習工程と、第 2 の認識結果と第 2 のデータの正解との第 2 の誤差に基づいて学習モデルを学習させる第 2 の学習工程と、を含む学習方法を実行させる。

10

【0036】

本発明の他の態様である認識器の学習済みモデルは、上述の学習方法によって得られる。

【0037】

本発明の他の態様である内視鏡システムは、上述の認識器の学習済みモデルを搭載する。

【0038】

好ましくは、第 1 の条件と前記第 2 の条件とでは、撮像装置、観察光の波長バランス、解像度、及び画像に施す画像処理のうち少なくとも 1 つが異なる。

20

【発明の効果】

【0039】

本発明によれば、互いに異なる条件で取得されたデータを使用して学習を行う場合であっても、効率の良い学習を行うことができる。

【図面の簡単な説明】

【0040】

【図 1】図 1 は、学習装置の構成を示すブロック図である。

【図 2】図 2 は、CNN の層構成の例を示す図である。

【図 3】図 3 は、図 2 で示した CNN の各層における入力及び出力されるデータ及び特徴量等を示す図である。

30

【図 4】図 4 は、学習装置で実行される学習方法を示すフローチャートである。

【図 5】図 5 は、第 1 の学習を説明する図である。

【図 6】図 6 は、第 2 の学習を説明する図である。

【図 7】図 7 は、第 1 の中間層に入力する特徴量を切り替える様子を示す図である。

【図 8】図 8 は、第 1 の入力層及び第 2 の入力層から第 1 の中間層に特徴量を入力する際の畳み込み (Convolution) の様子を示す図である。

【図 9】図 9 は、第 1 の学習、第 2 の学習のパターンを示す図である。

【図 10】図 10 は、第 1 の学習、第 2 の学習の他のパターンを示す図である。

【発明を実施するための形態】

40

【0041】

以下、添付図面に従って本発明に係る学習装置、学習方法、プログラム、学習済みモデル、及び内視鏡システムの好ましい実施の形態について説明する。

【0042】

< 学習装置の構成 >

図 1 は、本実施形態に係る学習装置 10 の構成を示すブロック図である。学習装置 10 は、被検体内に挿入される内視鏡で撮像された画像に基づく認識処理を行う認識器 100 と、通常光 (白色光) を観察光として取得した複数の内視鏡画像を記録する第 1 の画像データベース 201 と、特殊光 (狭帯域光) を観察光として取得した複数の内視鏡画像を記録する第 2 の画像データベース 202 とを備える。なお、以下の説明では通常光 (白色光

50

)を観察光として得られた画像を「通常光画像」(または「白色光画像」)といい、特殊光(狭帯域光)を観察光として得られた画像を「特殊光画像」(または「狭帯域光画像」)という。第1の画像データベース201、第2の画像データベース202に記録される内視鏡画像は、医用画像の一例である。

【0043】

<第1、第2の画像データベース>

<通常光画像及び特殊光画像>

第1の画像データベース201及び第2の画像データベース202は、ハードディスク等の記録媒体により構成される。第1の画像データベース201には通常光を観察光(第1の観察光)として撮影された複数の通常光画像(第1のデータ群、第1のデータ、第1の画像データ、第1の医用画像)が記録され、第2の画像データベース202には特殊光を観察光(第2の観察光)として撮影された複数の特殊光画像(第2のデータ群、第2のデータ、第2の画像データ、第2の医用画像)が記録される。すなわち、第1の画像データベース201に記録された複数の通常光画像は本発明における「第1の条件で取得された複数のデータ」の一態様であり、第2の画像データベース202に記録された複数の特殊光画像は本発明における「第1の条件とは異なる第2の条件で取得された複数のデータ」の一態様である。特殊光画像を撮影する特殊光(狭帯域光)は例えば青色狭帯域光とすることができるが、赤色狭帯域光等他の波長でもよい。また、上述の例では第1、第2の観察光が白色光と狭帯域光である場合について説明しているが、波長帯域及び/または強度が異なる第1、第2の狭帯域光を観察光として取得された内視鏡画像等の医用画像を用いてもよい。

10

20

【0044】

このように、通常光画像の取得条件(第1の条件)と特殊光画像の取得条件(第2の条件)は観察光の波長バランスが異なるが、この他、通常光画像と特殊光画像とで撮像装置、解像度、及び画像に施す画像処理が異なってもよい。すなわち、第1の条件と第2の条件とで撮像装置、観察光の波長バランス、解像度、及び画像に施す画像処理のうち少なくとも1つが異なってもよい。「撮像装置が異なる」には光学系の特性やプロセッサの性能が異なる内視鏡を用いていることが含まれるが、これに限定されるものではない。また、「画像に施す画像処理が異なる」には、注目領域等特定の領域を強調または目立たなくする処理、特定の波長成分の影響を強調または低減する処理の有無及び/または程度が異なることが含まれるが、これに限定されるものではない。

30

【0045】

<データ取得条件によるデータ数の違い>

内視鏡を用いた観察や検査を行う場合、ユーザは通常光(白色光)を観察光として取得された画像をモニタに表示させて確認するケースが多い。観察や検査の目的、状況(例えば、通常光では病変の構造が観察しづらい)により狭帯域光等の特殊光を観察光として画像をケースもあるが、通常光と比較すると観察光としての使用頻度が低く、そのため特殊光画像は通常光画像よりも著しく数が少ないケースが多い。機械学習により画像の学習及び/または認識を行う場合、特殊光画像についても学習及び/または認識を行う必要があるが、データ数が少ないと通常光画像と比較して学習及び/または認識の精度が低下するおそれがある。このような状況に鑑み、本実施形態では後述する階層型ネットワークの構成を採用してデータ数に差がある状況でも適切に学習及び/または認識できるようにしている。

40

【0046】

<内視鏡画像の正解データ>

第1の画像データベース201及び第2の画像データベース202は、上述した内視鏡画像に加え、注目領域(ROI: Region of Interest)を識別するための「正解データ」を画像と対応させて記憶する。具体的には、第1の画像データベース201は複数の通常光画像にそれぞれ対応する複数の正解データを記憶し、第2の画像データベース202は複数の特殊光画像にそれぞれ対応する複数の正解データを記憶する。正解データは、内視

50

鏡画像に対して医師が指定した注目領域や鑑別結果であることが好ましい。

【 0 0 4 7 】

< 認識器の構成 >

認識器 1 0 0 は、画像取得部 1 1 0、操作部 1 2 0、制御部 1 3 0、表示部 1 4 0、記録部 1 5 0、及び処理部 1 6 0 から構成されている。

【 0 0 4 8 】

画像取得部 1 1 0 は、外部サーバ、データベース等とネットワークを介して通信する装置等により構成され、学習や認識に用いる内視鏡画像や正解データを第 1 の画像データベース 2 0 1、第 2 の画像データベース 2 0 2 から取得する。画像取得部 1 1 0 は、図示せぬネットワークで学習装置 1 0 と接続された内視鏡システム、病院内サーバ等からも内視鏡画像を取得することができる。操作部 1 2 0 は図示せぬキーボード、マウス等の入力デバイスを備え、ユーザはこれらデバイスを介して画像取得、学習や認識等の処理に必要な操作を行うことができる。制御部 1 3 0 は記録部 1 5 0 に記録された各種プログラムを読み込み、操作部 1 2 0 から入力される指令に従って、学習装置 1 0 全体の動作を制御する。また制御部 1 3 0 は、後述する誤差算出部 1 6 4 が算出した誤差（損失）を CNN 1 6 2（CNN：Convolutional Neural Network、畳み込みニューラルネットワーク）に逆伝搬することにより、CNN 1 6 2 の重みパラメータを更新する。すなわち、制御部 1 3 0 は、CNN 1 6 2 に学習を行わせる学習制御部としての機能を有する。また、CNN 1 6 2 は、認識器 1 0 0 の学習モデルである。CNN 1 6 2 において以下で説明する第 1 の学習及び第 2 の学習が行われると、CNN 1 6 2 は認識器 1 0 0 の学習済みモデルとなる。

【 0 0 4 9 】

表示部 1 4 0 はモニタ 1 4 2（表示装置）を備え、内視鏡画像、学習結果、認識結果、処理条件設定画面等を表示する。記録部 1 5 0 は図示せぬ ROM（Read Only Memory）、RAM（Random Access Memory）、ハードディスク等で構成され、画像取得部 1 1 0 が取得したデータ、処理部 1 6 0 での学習結果や認識結果等を記録する。また、記録部 1 5 0 は内視鏡画像（医用画像）の学習、認識を行うためのプログラム（本発明の学習方法を学習装置 1 0 に実行させるプログラムを含む）を記録する。処理部 1 6 0 は、階層型ネットワークである CNN 1 6 2、及び CNN 1 6 2 の出力（認識結果）と上述した「正解データ」とに基づいて損失（誤差）を算出する誤差算出部 1 6 4 を備える。

【 0 0 5 0 】

< 各種のプロセッサによる機能の実現 >

上述した画像取得部 1 1 0、制御部 1 3 0、処理部 1 6 0（CNN 1 6 2、誤差算出部 1 6 4）の機能は、各種のプロセッサ（processor）を用いて実現できる。各種のプロセッサには、例えばソフトウェア（プログラム）を実行して各種の機能を実現する汎用的なプロセッサである CPU（Central Processing Unit）が含まれる。また、上述した各種のプロセッサには、画像処理に特化したプロセッサである GPU（Graphics Processing Unit）、FPGA（Field Programmable Gate Array）などの製造後に回路構成を変更可能なプロセッサであるプログラマブルロジックデバイス（Programmable Logic Device：PLD）も含まれる。さらに、ASIC（Application Specific Integrated Circuit）など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なども上述した各種のプロセッサに含まれる。

【 0 0 5 1 】

各部の機能は 1 つのプロセッサにより実現されてもよいし、同種または異種の複数のプロセッサ（例えば、複数の FPGA、あるいは CPU と FPGA の組み合わせ、または CPU と GPU の組み合わせ）で実現されてもよい。また、複数の機能を 1 つのプロセッサで実現してもよい。複数の機能を 1 つのプロセッサで構成する例としては、第 1 に、コンピュータに代表されるように、1 つ以上の CPU とソフトウェアの組合せで 1 つのプロセッサを構成し、このプロセッサが複数の機能として実現する形態がある。第 2 に、システムオンチップ（System On Chip：SoC）などに代表されるように、システム全体の機

10

20

30

40

50

能を1つのIC (Integrated Circuit) チップで実現するプロセッサを使用する形態がある。このように、各種の機能は、ハードウェア的な構造として、上述した各種のプロセッサを1つ以上用いて構成される。

【0052】

さらに、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、半導体素子などの回路素子を組み合わせた電気回路 (circuitry) である。

【0053】

上述したプロセッサあるいは電気回路がソフトウェア (プログラム) を実行する際は、実行するソフトウェアのプロセッサ (コンピュータ) 読み取り可能なコードをROM (Read Only Memory) 等の非一時的記録媒体に記憶しておき、プロセッサがそのソフトウェアを参照する。非一時的記録媒体に記憶しておくソフトウェアは、本発明に係る学習方法を実行するためのプログラムを含む。ROMではなく各種光磁気記録装置、半導体メモリ等の非一時的記録媒体にコードを記録してもよい。ソフトウェアを用いた処理の際には例えばRAM (Random Access Memory) が一時的記憶領域として用いられ、また例えば不図示のEEPROM (Electrically Erasable and Programmable Read Only Memory) に記憶されたデータを参照することもできる。これらのROM、RAM、EEPROM等は、記録部150に備えられたものを用いることができる。

【0054】

<CNNの層構成>

次に、CNN162の層構成に関して、図2及び図3に沿って説明をする。

【0055】

図2はCNN162の層構成の例を示す図である。図3は、図2で示したCNN162の各層における入力及び出力されるデータ及び特徴量等を示す図である。図2及び図3に示す例において、CNN162は、第1の入力層301 (第1の入力層) と、第2の入力層302 (第2の入力層) と、第1の中間層303 (中間層) と、第1の正規化層311 (第1の正規化層) と、第2の正規化層312 (第2の正規化層) と、第2の中間層313 (第2の中間層) と、出力層304 (出力層) とを含む。

【0056】

第1の入力層301は第1の画像データベース201に記憶された通常光画像 (第1のデータ群) から選択された画像 (第1のデータ) を入力して特徴量 (第1の特徴量) を出力する。

【0057】

第2の入力層302は第1の入力層301とは独立した入力層であり、第2の画像データベース202に記憶された特殊光画像 (第2のデータ群) から選択された画像 (第2のデータ) を入力して特徴量 (第2の特徴量) を出力する。

【0058】

第1の中間層303は第1の入力層301及び第2の入力層302に対して共通の中間層である。第1の中間層303は、第1の入力層301が出力した第1の特徴量 (A1) が入力された場合には、第1の中間特徴量 (B1) を出力する。また、第1の中間層303は、第2の入力層302が出力した第2の特徴量 (A2) が入力された場合には、第2の中間特徴量 (B2) を出力する。なお、第1の中間層303及び第2の中間層313の出力する特徴量の切り替えに関しては後で説明する。

【0059】

第1の正規化層311は、第1の中間層303から出力される第1の中間特徴量 (B1) が入力され、第1の中間特徴量に基づく第1の正規化特徴量 (C1) を出力する。

【0060】

第2の正規化層312は、第1の中間層303から出力される第2の中間特徴量 (B2) が入力され、第2の中間特徴量に基づく第2の正規化特徴量 (C2) を出力する。

【0061】

第2の中間層313は、第1の正規化層311及び第2の正規化層312に対して共通

10

20

30

40

50

の中間層である。第2の中間層313は、第1の正規化層311から出力される第1の正規化特徴量(C1)が入力された場合には第3の中間特徴量(D1)を出力する。また、第2の中間層313は、第2の正規化層312から出力される第2の特徴量(C2)が入力された場合には第4の特徴量(D2)を出力する。

【0062】

出力層304は、第2の中間層313から特徴量が入力され、第1の入力層301または第2の入力層302に入力された画像における認識結果を出力する。具体的には、出力層304は、第2の中間層313から出力された第3の中間特徴量(D1)が入力された場合には、第3の特徴量(D1)に基づく第1の認識結果(E1)を出力する。また、出力層304は、第2の中間層313から出力された第4の中間特徴量(D2)が入力された場合には、第4の中間特徴量(D2)に基づく第2の認識結果(E2)を出力する。ここで、第1の認識結果(E1)は第1のデータの認識結果であり、第2の認識結果(E2)は第2のデータの認識結果である。

10

【0063】

なお、第1の入力層301と、第1の中間層303と、第1の正規化層311と、第2の中間層313、出力層304とは、複数の「ノード」が「エッジ」で結ばれた構造となっており、複数の重みパラメータを保持している。また、第2の入力層302と、第1の中間層303と、第2の正規化層312と、第2の中間層313と、出力層304とは、複数の「ノード」が「エッジ」で結ばれた構造となっており、複数の重みパラメータを保持している。そして、これらの重みパラメータの値は、学習が進むにつれて変化していく。

20

【0064】

次に、CNN162を構成する各層での処理に関して説明を行う。

【0065】

<入力層及び中間層における処理>

第1の入力層301及び第2の入力層302の各層は、畳み込み演算、プーリング処理、活性化処理、及びバッチノーマライゼーション処理のいずれか一つを含む演算によって特徴量を出力する。第1の中間層303及び第2の中間層313の各層は、畳み込み演算、プーリング処理、及び活性化処理のいずれか一つを含む演算によって特徴量を出力する。例えば、第1の入力層301及び第2の入力層302の各層は、畳み込み演算、プーリング処理、活性化処理、及びバッチノーマライゼーションの演算が層状に組み合わせられており、特徴量を出力する。例えば第1の中間層303及び第2の中間層313の各層は、畳み込み演算、プーリング処理、及び活性化処理の演算が層状に組み合わせられており、特徴量を出力する。

30

【0066】

畳み込み演算は、入力されたデータ(例えば画像)にフィルタを使用した畳み込み演算により特徴マップを取得する処理である。畳み込み演算は、画像からのエッジ抽出等の特徴抽出の役割を担う。このフィルタを用いた畳み込み演算により、1つのフィルタに対して1チャンネル(1枚)の特徴マップが生成される。特徴マップのサイズは、畳み込みによりダウンスケーリングされ、各層で畳み込みが行われるにつれて小さくなって行く。

【0067】

プーリング処理は、畳み込み演算により出力された特徴マップを縮小(または拡大)して新たな特徴マップとする処理である。プーリング処理は、抽出された特徴が、平行移動などによる影響を受けないようにロバスト性を与える役割を担う。

40

【0068】

活性化処理は、特徴マップに対して活性化関数を使用して演算を行う。活性化関数としては、ジグモイド関数やReLU(Rectified Linear Unit)が使用される。

【0069】

バッチノーマライゼーション処理は学習を行う際のミニバッチを単位としてデータの分布を正規化する処理であり、学習を速く進行させる、初期値への依存性を下げる、過学習を抑制する等の役割を担う。

50

【 0 0 7 0 】

第1の入力層301、第2の入力層302、第1の中間層303、及び第2の中間層313は、これらの処理を行う1または複数の層により構成することができる。なお、層の構成は畳み込み演算、プーリング処理、活性化処理、及びバッチノーマライゼーション処理を行う層を1つずつ含む場合に限らず、いずれかの層が複数含まれていてもよい。

【 0 0 7 1 】

これら第1の入力層301、第2の入力層302、第1の中間層303、及び第2の中間層313の層のうち、入力側に近い層では低次の特徴抽出（エッジの抽出等）が行われ、出力側に近づくにつれて高次の特徴抽出（対象物の形状、構造等に関する特徴の抽出）が行われる。

【 0 0 7 2 】

< 正規化層における処理 >

第1の正規化層311及び第2の正規化層312は、入力された特徴量を正規化する。具体的には、第1の正規化層311及び第2の正規化層312は、入力された特徴量分布を正規化し、正規化特徴量を出力する。ここで、第1の正規化層311は、第1のデータに基づく第1の中間特徴量（B1）を正規化し、第2の正規化層312は、第2のデータに基づく第2の中間特徴量（B2）を正規化する。このように、CNN162では、第1の中間特徴量（B1）の専用の第1の正規化層311と、第2の中間特徴量（B2）の専用の第2の正規化層312とを独立に設けている。これにより、第1の中間特徴量（B1）及び第2の中間特徴量（B2）は、それぞれ個別独立の適切な条件で正規化されることになる。ここで仮に、第1の中間特徴量（B1）及び第2の中間特徴量（B2）を共通の正規化層で同じ条件で正規化を行うと、正規化処理の効果が小さくなってしまったり、正規化処理を行うことによりかえって、CNN162の学習が効率良く進まなくなったりする。これは、ことなる条件で取得された第1のデータと第2のデータに由来する2つの特徴量を正規化すると、その中間の特徴量への正規化が行われるからである。従って、CNN162では、第1の中間層303と第2の中間層313との間に、第1の中間特徴量（B1）専用の第1の正規化層311と第2の中間特徴量（B2）専用の第2の正規化層312とを設けることにより、第1のデータ及び第2のデータのそれぞれに適した正規化処理が実現されている。また、第1の正規化層311及び第2の正規化層312は、第1の中間層303と第2の中間層313とに挟まれる位置に並列に設けられる。これにより、第1の中間層303で出力された第1の中間特徴量（B1）及び第2の中間特徴量（B2）の正規化をそれぞれ行い、正規化した特徴量（第1の正規化特徴量及び第2の正規化特徴量）をさらに第2の中間層313に出力することができる。なお、第1の正規化層311及び第2の正規化層312で行われる正規化処理は、例えばバッチノーマライゼーション処理である。例えば、バッチノーマライゼーション処理により、第1の中間特徴量（B1）の分布が平均0分散1となるように、第2の中間特徴量（B2）の分布が平均0分散1となるように正規化処理が行われる。具体例として、第1のデータとして通常光の医療画像、第2のデータとして特殊光の医療画像を用いた場合には、第1の正規化層311と第2の正規化層312とで、色に関してそれぞれ異なる条件で正規化が行われることがある。このように、第1の正規化層311及び第2の正規化層312を設けることにより、CNN162は、異なる条件で取得された第1のデータ及び第2のデータを使用して学習を行う場合であっても、それぞれ適切に正規化を行うことができ、効率の良い学習を行うことができる。なお、上述した第1の入力層301及び第2の入力層302においてもバッチノーマライゼーション処理が行われるが、第1の入力層301及び第2の入力層302は、それぞれ第1のデータまたは第2のデータのみしか入力されないため、第1のデータ専用または第2のデータ専用のバッチノーマライゼーション処理となる。一方、第1の中間層303は性質の異なる第1のデータと第2のデータに由来する特徴量が入力されるため、分岐した第1の正規化層311と第2の正規化層312とを設けて正規化を正しく行っている。

【 0 0 7 3 】

10

20

30

40

50

< 出力層における処理 >

出力層 304 は、第 2 の中間層 313 から出力された特徴量に基づき、入力された画像（通常光画像、特殊光画像）に映っている注目領域の位置検出を行ってその結果を出力する層である。出力層 304 は、第 2 の中間層 313 から得られる「特徴マップ」により、画像に写っている注目領域の位置を画素レベルで把握する。即ち、内視鏡画像の画素ごとに注目領域に属するか否かを検出し、その検出結果を出力することができる。

【0074】

出力層 304 は、病変に関する鑑別を実行して鑑別結果を出力するものでもよい。例えば、出力層 304 は、内視鏡画像を「腫瘍性」、「非腫瘍性」、「その他」の 3 つのカテゴリに分類し、鑑別結果として「腫瘍性」、「非腫瘍性」及び「その他」に対応する 3 つのスコア（3 つのスコアの合計は 100%）として出力してもよいし、3 つのスコアから明確に分類できる場合には分類結果を出力してもよい。なお鑑別結果を出力する場合、出力層 304 が最後の 1 層または複数の層として全結合層を有することが好ましい。

10

【0075】

< 学習方法 >

次に、上述した学習装置 10 で実行される学習方法に関して説明する。図 4 は、学習装置 10 で実行される学習方法を示すフローチャートである。

【0076】

まず学習装置 10 の制御部 130 により、第 1 の学習工程が行われ（ステップ S106 を参照）、その後第 2 の学習工程が行われる（ステップ S112 を参照）。

20

【0077】

まず、第 1 の学習について説明する。処理部 160 は、第 1 の入力層 301 で第 1 の特徴量の算出処理（ステップ S101）を行う。次に処理部 160 は、第 1 の中間層 303 で第 1 の中間特徴量算出処理（ステップ S102）を行う。次に処理部 160 は、第 1 の正規化層 311 で第 1 の正規化特徴量算出処理（ステップ S103）を行う。次に処理部 160 は、第 2 の中間層 313 で第 3 の中間特徴量算出処理（ステップ S104）を行う。次に処理部 160 は、出力層 304 で第 1 の認識結果出力処理（ステップ S105）を行う。その後、制御部 130 は、第 1 の学習を CNN 162 に行わせる（ステップ S106）。

【0078】

次に、第 2 の学習について説明する。上述した第 1 の学習の後に第 2 の学習が行われる。処理部 160 は、第 2 の入力層 302 で第 2 の特徴量算出処理（ステップ S107）を行う。次に処理部 160 は、第 1 の中間層 303 で第 2 の中間特徴量算出処理（ステップ S108）を行う。次に処理部 160 は、第 2 の正規化層 312 で第 2 の正規化特徴量算出処理（ステップ S109）を行う。次に処理部 160 は、第 2 の中間層 313 で第 4 の中間特徴量算出処理（ステップ S110）を行う。次に処理部 160 は、出力層 304 で第 2 の認識結果出力処理（ステップ S111）を行う。その後、制御部 130 は、第 2 の学習を CNN 162 に行わせる（ステップ S112）。

30

【0079】

次に、第 1 の学習及び第 2 の学習における各処理に関して、詳しく説明を行う。

40

【0080】

< 第 1 の学習 >

図 5 は、CNN 162 における第 1 の学習を説明する図である。なお、図 5 において下向き矢印は第 1 の入力層 301 から、第 1 の中間層 303、第 1 の正規化層 311、及び第 2 の中間層 313 を経て出力層 304 に至る方向に情報が伝達されること（学習方向）を意味し、学習方向とは逆の上向き矢印は出力層 304 から、第 2 の中間層 313、第 1 の正規化層 311、第 1 の中間層 303、第 1 の入力層 301 に情報が伝達されること（後述する誤差逆伝搬）を意味する。

【0081】

[第 1 の特徴量算出処理]

50

第1の学習では、第1の画像データベース201に記録されている複数の通常光画像から選択された複数の画像(第1のデータ)でミニバッチを構成し、第1の入力層301に入力する。そして、第1の入力層301において、第1の特徴量算出処理(ステップS101)が行われて、第1の特徴量が算出される。

【0082】

[第1の中間特徴量算出処理]

第1の中間層303には上述のように第1の入力層301及び第2の入力層302が接続されているので、学習の際には第1の入力層301の出力と第2の入力層302の出力とを切り替えて入力する。図5に示すように、第1の中間層303は、第1の入力層301から出力される第1の特徴量が第1の中間層303へ入力された場合には、第1の中間特徴量を算出する(ステップS102)。

10

【0083】

図7は第1の中間層303に入力する特徴量を切り替える様子を示す図である。図7(a)は第1の特徴量を第1の中間層303へ入力する状態(第1の入力層301に含まれる層を構成するノード301Aからの出力を、第1の中間層303を構成するノード303Aに入力している)を示している。入力の際、第1の入力層301から出力される特徴量を第1の特徴量としてそのまま第1の中間層303へ入力してもよいし、適宜重みを乗じた特徴量を第1の特徴量として第1の中間層303へ入力してもよい(図8を参照)。なお、図中の実線は上述した出力の切り替えによりノードからデータが出力または入力されている状態を示し、図中の点線はノードからデータが出力または入力されていない状態を示す。ノード301A、303Aは概念的に示したものであり、数は特に限定されない。これらの点については図8でも同様である。

20

【0084】

図8は第1の入力層301及び第2の入力層302から第1の中間層303に特徴量を入力する際の畳み込み(Convolution)の様子を示す図である。図8の(a)部分は、第1の入力層301のノードX11、X12、X13の出力に対しそれぞれ重みパラメータW11、W12、W13を乗じて第1の中間層303のノードY11に入力する様子を示している(同図に示す状態では、ノードX10からはノードY11に入力されていない)。同図ではノードX11、X12、X13とノードY11との入力関係を示しているが、第1の中間層303の他のノードY10、Y12、Y13についても同様の関係が成立する。

30

【0085】

[第1の正規化特徴量算出処理]

処理部160は、第1の正規化層311で第1の正規化特徴量算出処理を行う。具体的には、処理部160は、第1の中間層303から出力される第1の中間特徴量に基づいて、第1の正規化特徴量を算出する(ステップS103)。

【0086】

[第3の中間特徴量算出処理]

処理部160は、第2の中間層313で第3の中間特徴量算出処理を行う(ステップS104)。具体的には、処理部160は、第1の正規化層311から出力される第2の正規化特徴量に基づいて、第3の中間特徴量を算出する。なお、第2の中間層313では、上述した第1の中間特徴量算出処理と同様に、第1の正規化層311から出力される第1の正規化特徴量と第2の正規化層312から出力される第2の正規化特徴量とが切り替えて入力される。第3の中間特徴量算出処理の詳細な説明は、第1の中間特徴量算出処理と同様であるので省略する。

40

【0087】

[第1の認識結果出力処理]

出力層304は、第2の中間層313で算出した第3の中間特徴量を入力し、第1の認識結果出力処理を行って、第1の認識結果を出力する(ステップS105)。

【0088】

50

[第 1 の学習の処理 (誤差逆伝搬による重みパラメータの更新)]

誤差算出部 164 は出力層 304 が出力する第 1 の認識結果と第 1 のデータに対する正解とを比較して損失 (第 1 の誤差) を計算する。なお、後で説明する第 2 の学習では、誤差算出部 164 は出力層 304 が出力する第 2 の認識結果と第 2 のデータに対する正解とを比較して損失 (第 2 の誤差) を計算する。そして誤差算出部 164 は、算出した損失が小さくなるように、図 5 に示すように、第 1 の入力層 301、第 1 の中間層 303、第 1 の正規化層 311、第 2 の中間層 313、及び出力層 304 における重みパラメータを出力側の層から入力側の層に向けて更新していく (誤差逆伝搬)。これらのパラメータの更新は、第 1 の学習 (ステップ S 106) である。

【 0089 】

< 第 2 の学習 >

図 6 は、CNN 162 における第 2 の学習を説明する図である。なお、図 6 において下向き矢印は第 2 の入力層 302 から、第 1 の中間層 303、第 2 の正規化層 312、及び第 2 の中間層 313 を経て出力層 304 に至る方向に情報が伝達されること (学習方向) を意味し、学習方向とは逆の上向き矢印は出力層 304 から、第 2 の中間層 313、第 2 の正規化層 312、第 1 の中間層 303、第 2 の入力層 302 に情報が伝達されること (後述する誤差逆伝搬) を意味する。

【 0090 】

[第 2 の特徴量算出処理]

第 2 の学習では、第 2 の画像データベース 202 に記録されている複数の特殊光画像から選択された複数の画像 (第 2 のデータ) でミニバッチを構成し、第 2 の入力層 302 に入力する。そして、第 2 の入力層 302 において、第 2 の特徴量算出処理 (ステップ S 107) が行われて、第 2 の特徴量が算出される。

【 0091 】

[第 2 の中間特徴量算出処理]

第 1 の中間層 303 は、第 2 の特徴量が入力され、第 2 の中間特徴量算出処理を行うことにより、第 2 の中間特徴量を算出する (ステップ S 108)。第 1 の中間層 303 には上述のように第 1 の入力層 301 及び第 2 の入力層 302 が接続されているので、学習の際には第 1 の入力層 301 の出力と第 2 の入力層 302 の出力とを切り替えて入力する。

【 0092 】

第 2 の学習の際には、図 6 に示すように出力を切り替え、第 2 の入力層 302 からの出力を第 1 の中間層 303 へ入力する。図 7 (b) は、第 2 の特徴量を第 1 の中間層 303 へ入力する様子 (第 2 の入力層 302 を構成するノード 302 A からの出力を、第 1 の中間層 303 を構成するノード 303 A へ入力) を示す図である。図 6 に示す状態で、第 2 の入力層 302 から出力される特徴量に基づく第 2 の特徴量を第 1 の中間層 303 へ入力して、第 1 の中間層 303 で第 2 の中間特徴量を算出する。図 7 (b) は第 2 の特徴量を第 1 の中間層 303 へ入力する状態を示している。

【 0093 】

図 8 (b) は、同図の (a) 部分と同様に、第 2 の入力層 302 のノード X 21、X 22、X 23 の出力に対しそれぞれ重みパラメータ W 21、W 22、W 23 を乗じて第 1 の中間層 303 のノード Y 11 へ入力する様子を示している (同図に示す状態では、ノード X 20 からはノード Y 11 へ入力されていない)。同図ではノード X 21、X 22、X 23 とノード Y 11 との入力関係を示しているが、第 1 の中間層 303 の他のノード Y 10、Y 12、Y 13 についても同様の関係が成立する。

【 0094 】

なお、第 2 の学習における「第 2 の正規化特徴量算出処理 (ステップ S 109)」、「第 4 の中間特徴量算出処理 (ステップ S 110)」、「第 2 の認識結果出力処理 (ステップ S 111)」、「第 2 の学習 (ステップ S 112)」は、第 1 の学習における「第 1 の正規化特徴量算出処理 (ステップ S 103)」、「第 3 の中間特徴量算出処理 (ステップ S 104)」、「第 1 の認識結果出力処理 (ステップ S 105)」、「第 1 の学習 (ステップ S 10

10

20

30

40

50

6)」と同様の処理が行われるので、説明は省略する。

【0095】

<学習パターンの例>

上述した学習方法の説明では、第1の学習及び第2の学習をそれぞれ1回行う例について説明をしたが、学習装置10が行う学習方法はこれに限定されるものではない。第1の学習と第2の学習がそれぞれ少なくとも1回実行されていればよく、様々な態様が採用される。以下、処理の回数及び順番の例について説明する。

【0096】

(第1の例)

第1の例では、第2の中間層313は、1回目の第1の学習における第3の中間特徴量が出力された後であって、2回目の第1の学習における第3の中間特徴量が出力される前の期間に、第2の学習における第4の中間特徴量を出力する。

10

【0097】

例えば、図9(a)に示す順番で処理を繰り返す。図中「A」、「B」とあるのはそれぞれ「第2の中間層313における第3の中間特徴量の算出」、「第2の中間層313における第4の中間特徴量の算出」を意味し、ミニバッチ単位で1回、2回...と数える。

【0098】

(第2の例)

第2の例では、第2の中間層313は、1回目の第1の学習における第3の中間特徴量の出力、及び2回目の第1の学習における第3の中間特徴量の出力が完了した後に、第2の学習における第4の中間特徴量を出力する。例えば、図9(b)に示す順番で処理を繰り返す。図9(b)での、「A」、「B」とあるのは同図の(a)部分と同じ意味である。この場合、図9(c)に示すように「B」を2回続けて行ってもよい。

20

【0099】

(第3の例)

第3の例では、学習装置10は、第1の学習を複数回連続して行った後に、第2の学習を複数回連続して行う。例えば、学習装置10は、図10に示す順番により学習を行う。なお、図10での「第1」、「第2」とあるのはそれぞれ、「第1の学習」、「第2の学習」である。なお、図9、10に示すパターンは例示であり、他にも様々なパターンで学習を行うことができる。

30

【0100】

<効果>

学習装置10は、第1の中間層303は、第1のデータに基づく第1の特徴量が入力された場合には第1の中間特徴量を出力し、第2のデータに基づく第2の特徴量が入力された場合には第2の中間特徴量を出力する。そして、第1の正規化層311は第1の中間特徴量を入力し第1の正規化特徴量を出力し、第2の正規化層312は第2の中間特徴量を入力し第2の正規化量を出力する。第2の中間層313は、第1の正規化特徴量及び第2の正規化特徴量を入力する。これにより、本態様は、第1のデータに由来する第1の中間特徴量と第2のデータに由来する第2の中間特徴量とを別の条件で正規化することができるので、第1の中間特徴量と第2の中間特徴量とを適切に正規化することができ、効率的な学習を行うことができる。

40

【0101】

また、学習装置10では、独立した第1、第2の入力層に第1、第2のデータをそれぞれ入力し、第1、第2の入力層でそれぞれ特徴量を算出することで、第1、第2の入力層の一方における特徴量算出が他方の入力層における特徴量算出の影響を受けないようにしている。また学習装置10では、入力層(第1の入力層301及び第2の入力層302)における特徴抽出に加えて、さらに第1、第2の入力層に共通な第1の中間層303において第1の中間特徴量と第2の中間特徴量とが算出されるので、入力層で第1、第2のデータから算出した特徴量を第1の中間層303における中間特徴量算出に反映することができる。第2の中間層313も、第1の正規化層311及び第2の正規化層312に共通

50

であるので、同様に、第1の正規特徴量及び第2の正規化特徴量を第2の中間層313における中間特徴量算出に反映することができる。また、階層型ネットワークはパラメータが多いため過学習になりがちであるが、大量にデータを与えることで過学習を回避できる。また、学習装置10では、中間層は第1、第2のデータを合わせた大量のデータで学習できるため過学習になりやすく、一方、入力層は第1、第2の入力層に独立しておりそれぞれの入力層のパラメータは少なくなるため、少量のデータでも過学習になりにくい。

【0102】

学習装置10によれば、このようにして同一のカテゴリに属し異なる条件で取得されたデータを適切に学習することができる。

【0103】

<合成ミニバッチによる学習>

なお、上述した学習パターンでは第1、第2のデータについてミニバッチ単位で別個に特徴量算出を行っているが、第1の中間層303への入力直前に第1、第2のミニバッチを1つのミニバッチに合成してもよい。具体的には、第1の画像データベース201に記録されている複数の通常光画像から選択された複数の画像(第1のデータ)でミニバッチ(第1のミニバッチ)を構成し、第1の入力層301に入力して特徴量を算出する。また、第2の画像データベース202に記録されている複数の特殊光画像から選択された複数の画像(第2のデータ)でミニバッチ(第2のミニバッチ)を構成し、第2の入力層302に入力して特徴量を算出する。これらの特徴量について第1の中間層303への入力直前に第1、第2のミニバッチを1つのミニバッチに合成して第1の中間層303に入力してもよい。

【0104】

<認識処理>

上述した学習(第1の学習及び第2の学習)が進むと、認識器100のCNN162は学習済みモデルとなる。学習済みモデルとなったCNN162を使用した認識(推論)処理においては、第1の入力層301または第2の入力層302を取り外した構成にして認識を行ってもよい。例えば、図5に示したように第2の入力層302を取り外して第1の入力層301のみが接続されている状態で、第1のデータに対して認識を行うことができる。また、図6に示したように第1の入力層301を取り外して第2の入力層302のみが接続されている状態で、第2のデータに対して認識を行うことができる。

【0105】

<第1狭帯域光画像及び第2狭帯域光画像による学習>

上述した例では、通常光画像(白色光画像)及び特殊光画像(例えば、青色特殊光画像)を用いた学習について説明したが、観察光の波長バランスが異なる複数の狭帯域光画像を用いて学習を行ってもよい。第1の入力層301は第1の狭帯域光を第1の観察光として取得された第1の医用画像のデータを第1の画像データとして入力し、第2の入力層302は第1の狭帯域光とは異なる第2の狭帯域光を第2の観察光として取得された第2の医用画像のデータを第2の画像データとして入力してもよい。この場合、狭帯域光の組み合わせとしては複数の青色狭帯域光、青色狭帯域光と紫色狭帯域光、複数の赤色狭帯域光等を用いることができる。

【0106】

<その他のデータを用いた学習>

上述の実施形態では異なる観察光で取得された内視鏡画像を用いた学習について説明したが、本発明に係る学習装置及び学習方法ではCT装置(Computed Tomography)、MRI(Magnetic Resonance Imaging)装置等の内視鏡画像以外の他の医用画像を用いる場合でも同様に学習することができる。また、医用画像以外の画像(例えば、人物、動物、風景等他の画像)を用いる場合でも同様に学習することができる。さらに、入力するデータが画像でなく文章、音声等の場合も同様に学習することができる。

【0107】

以上で本発明の例に関して説明してきたが、本発明は上述した実施の形態に限定されず

10

20

30

40

50

、本発明の趣旨を逸脱しない範囲で種々の変形が可能であることは言うまでもない。

【符号の説明】

【 0 1 0 8 】

1 0	: 学習装置	
1 0 0	: 認識器	
1 1 0	: 画像取得部	
1 2 0	: 操作部	
1 3 0	: 制御部	
1 4 0	: 表示部	
1 4 2	: モニタ	10
1 5 0	: 記録部	
1 6 0	: 処理部	
1 6 4	: 誤差算出部	
2 0 1	: 第 1 の画像データベース	
2 0 2	: 第 2 の画像データベース	
3 0 1	: 第 1 の入力層	
3 0 2	: 第 2 の入力層	
3 0 3	: 第 1 の中間層	
3 0 4	: 出力層	
3 1 1	: 第 1 の正規化層	20
3 1 2	: 第 2 の正規化層	
3 1 3	: 第 2 の中間層	

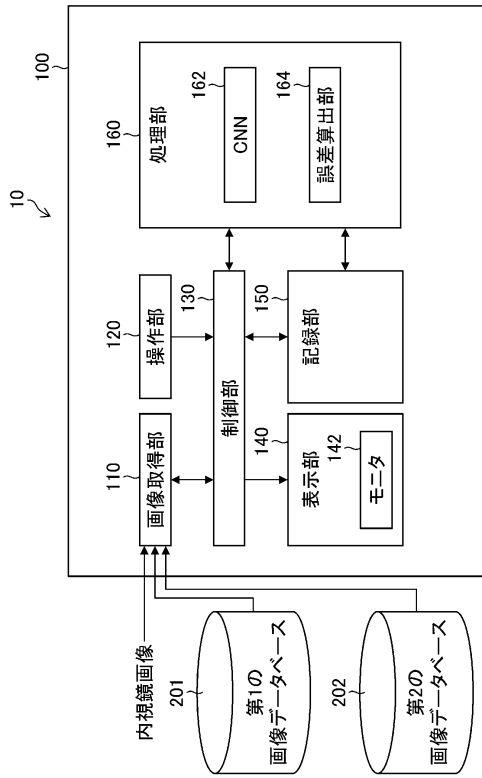
30

40

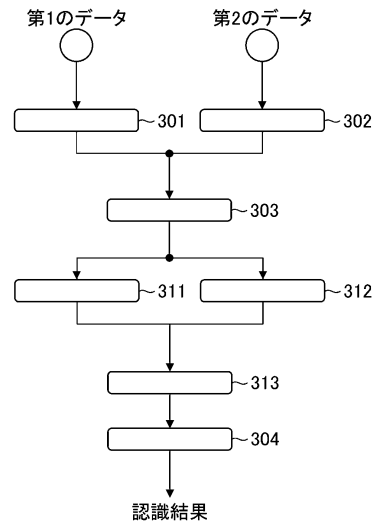
50

【図面】

【図 1】



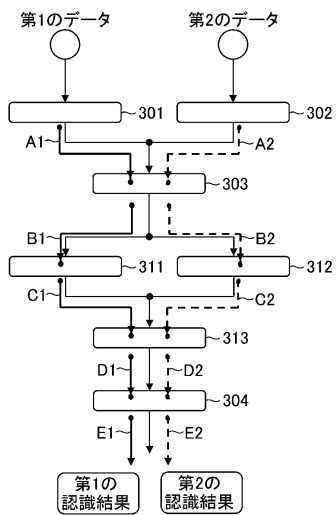
【図 2】



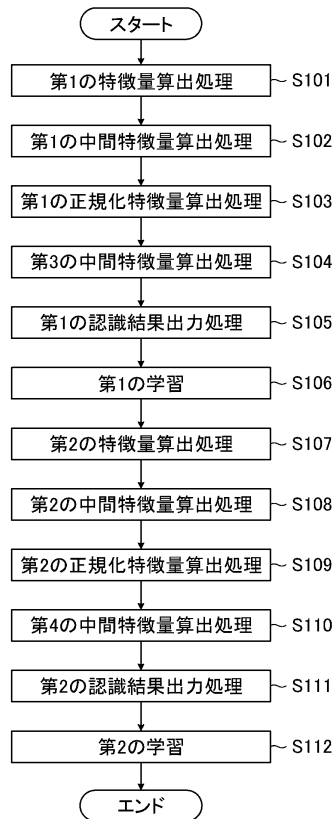
10

20

【図 3】



【図 4】

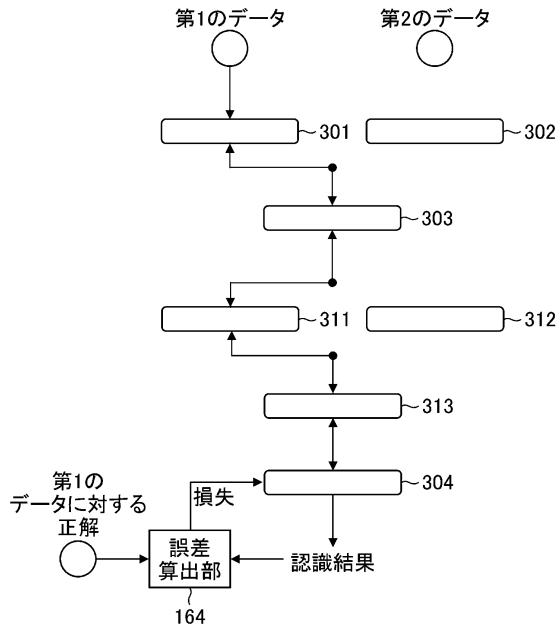


30

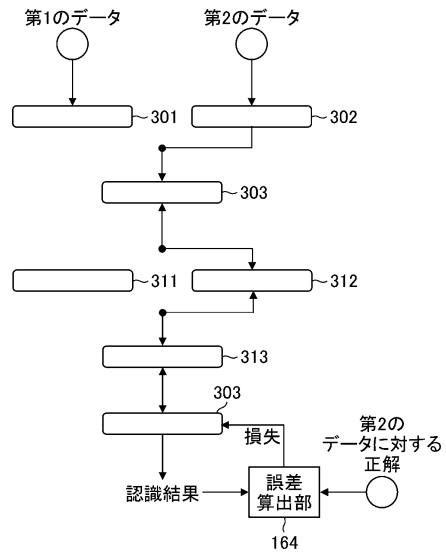
40

50

【 図 5 】



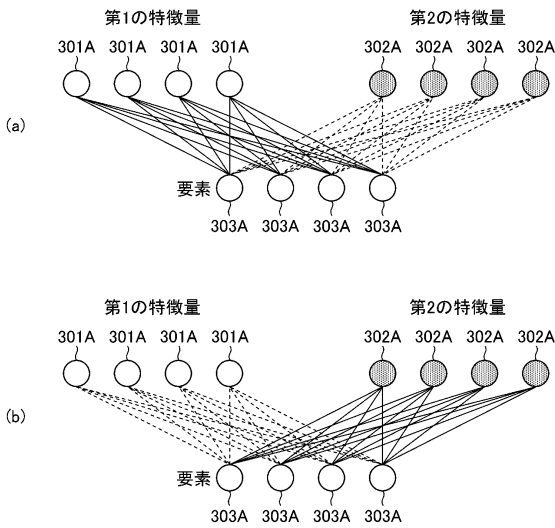
【 図 6 】



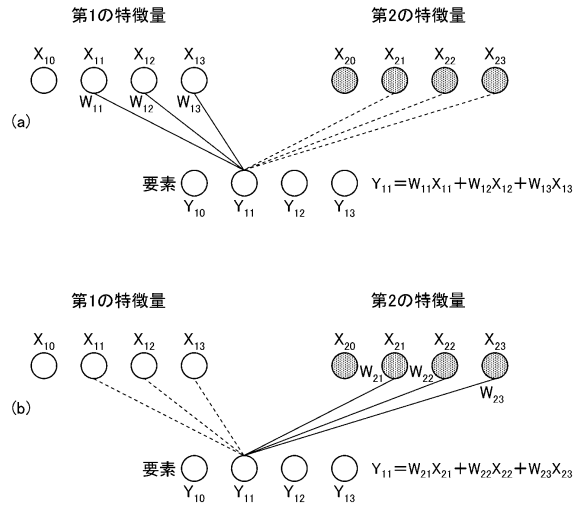
10

20

【 図 7 】



【 図 8 】

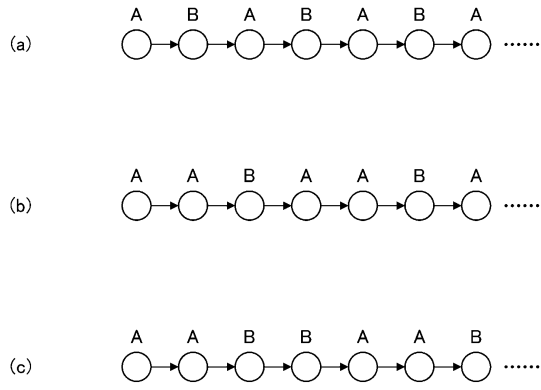


30

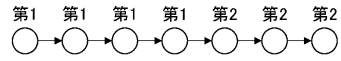
40

50

【 図 9 】



【 図 10 】



10

20

30

40

50

フロントページの続き

富士フイルム株式会社内

審査官 渡辺 順哉

(56)参考文献 国際公開第2020/022027(WO, A1)

特開2015-102806(JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06N 3/00-99/00