

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7566969号

(P7566969)

(45)発行日 令和6年10月15日(2024.10.15)

(24)登録日 令和6年10月4日(2024.10.4)

(51)国際特許分類

F I

G 0 6 N	3/096(2023.01)	G 0 6 N	3/096	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00	3 5 0 B
G 0 6 V	10/82 (2022.01)	G 0 6 V	10/82	

請求項の数 17 (全22頁)

(21)出願番号	特願2023-70644(P2023-70644)	(73)特許権者	514322098
(22)出願日	令和5年4月24日(2023.4.24)		ベイジン バイドゥ ネットコム サイエ
(65)公開番号	特開2024-35052(P2024-35052A)		ンス テクノロジー カンパニー リミテ
(43)公開日	令和6年3月13日(2024.3.13)		ッド
審査請求日	令和5年4月24日(2023.4.24)		Beijing Baidu Netco
(31)優先権主張番号	202211059602.5		m Science Technolog
(32)優先日	令和4年8月30日(2022.8.30)		y Co., Ltd.
(33)優先権主張国・地域又は機関	中国(CN)		中華人民共和国 ベキン 100085,
			ハイディアン ディストリクト, シャン
			ディ テンス ストリート, 10番, バ
			イドゥ キャンパス 2階
			2/F Baidu Campus, N
			o.10, Shangdi 10th
			Street, Haidian Dis
			trict, Beijing 1000
			最終頁に続く

(54)【発明の名称】 軽量モデルトレーニング方法、画像処理方法、軽量モデルトレーニング装置、画像処理装置、電子デバイス、記憶媒体及びコンピュータプログラム

(57)【特許請求の範囲】

【請求項1】

コンピュータにより実行される、第1取得モジュールと、データ拡張モジュールと、予測モジュールと、第1決定モジュールと、第2決定モジュールと、トレーニングモジュールと、第3決定モジュールと、を備える、軽量モデルトレーニング装置において、

前記第1取得モジュールが、e回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得することであって、前記eは最大反復回数であるE以下の正の整数であり、前記Eは1より大きい正の整数である、取得することと、前記データ拡張モジュールが、それぞれ前記第1拡張確率と前記第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得ることと、

前記予測モジュールが、前記第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、前記第2データセットに基づいて前記生徒モデルの第3出力値と前記教師モデルの第4出力値とを得ることと、

前記第1決定モジュールが、前記第1出力値と前記第2出力値とに基づいて蒸留損失関数を決定し、前記第3出力値と前記第4出力値とに基づいて真値損失関数を決定することと、第2決定モジュールが、前記蒸留損失関数と前記真値損失関数とに基づいて目標損失関数を決定することと、

前記トレーニングモジュールが、前記目標損失関数に基づいて前記生徒モデルをトレーニングし、前記第3決定モジュールが、前記eが前記Eよりも小さい場合に、e+1回目の

10

20

反復に用いられるべき第 1 拡張確率または目標重みを決定することと、を含む、
軽量モデルトレーニング方法。

【請求項 2】

前記軽量モデルトレーニング装置は、第 2 取得モジュールと、第 4 決定モジュールと、をさらに備え、

前記軽量モデルトレーニング方法は、
前記第 2 取得モジュールが、最大拡張確率を取得することと、
前記第 4 決定モジュールが、前記最大拡張確率と、前記最大反復回数と、前記第 1 拡張確率とに基づいて、前記第 2 拡張確率を決定することと、をさらに含む、

請求項 1 に記載の軽量モデルトレーニング方法。

10

【請求項 3】

前記第 3 決定モジュールが、前記 e + 1 回目の反復に用いられるべき第 1 拡張確率を決定することは、

前記最大拡張確率と、前記最大反復回数と、前記 e 回目の前記第 1 拡張確率とに基づいて、前記 e + 1 回目の反復に用いられるべき第 1 拡張確率を決定すること、を含む、

請求項 2 に記載の軽量モデルトレーニング方法。

【請求項 4】

前記軽量モデルトレーニング装置は、第 3 取得モジュールをさらに備え、

前記軽量モデルトレーニング方法は、
前記第 3 取得モジュールが、最大目標重みを取得すること、を含み、
前記第 3 決定モジュールが、前記 e + 1 回目の反復に用いられるべき目標重みを決定することは、

前記最大目標重みと、前記最大反復回数と、前記 e 回目の前記目標重みとに基づいて、前記 e + 1 回目の反復に用いられるべき目標重みを決定すること、を含む、

請求項 1 に記載の軽量モデルトレーニング方法。

20

【請求項 5】

前記第 2 決定モジュールが、前記蒸留損失関数と前記真値損失関数とに基づいて目標損失関数を決定することは、

前記目標重みが最大目標重み以上である場合に、または前記蒸留損失関数が前記真値損失関数以上である場合に、前記蒸留損失関数を前記目標損失関数として決定する一方、そうでない場合に、前記真値損失関数を前記目標損失関数として決定すること、を含む、

請求項 1 に記載の軽量モデルトレーニング方法。

30

【請求項 6】

前記第 1 決定モジュールが、前記第 1 出力値と前記第 2 出力値とに基づいて蒸留損失関数を決定することは、

下記の式により前記蒸留損失関数を決定すること、を含み、

【数 1】

$$l1 = (a + a_{dft} * 2/E) * L_{dist}(o1s, o1t) + (1 - a - a_{dft} * 2/E) * L_{gt}(o1s, gt)$$

ここで、 $l1$ は前記蒸留損失関数を表し、 $L_{dist}(o1s, o1t)$ は前記第 1 出力値と前記第 2 出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o1s, gt)$ は前記第 1 出力値と真値とに基づいて決定された真値項損失値を表し、 a は前記目標重みを表し、 a_{dft} は最大目標重みを表し、 E は前記最大反復回数を表し、 gt は前記真値を表し、 $o1s$ は前記第 1 出力値を表し、 $o1t$ は前記第 2 出力値を表す、

請求項 1 に記載の軽量モデルトレーニング方法。

40

【請求項 7】

前記第 1 決定モジュールが、前記第 3 出力値と前記第 4 出力値とに基づいて真値損失関数を決定することは、

下記の式により前記真値損失関数を決定すること、を含み、

50

【数 2】

$$l2 = a * L_{dist}(o2s, o2t) + (1 - a) * L_{gt}(o2s, gt)$$

ここで、 $l2$ は前記真値損失関数を表し、 $L_{dist}(o2s, o2t)$ は前記第3出力値と前記第4出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o2s, gt)$ は前記第3出力値と真値とに基づいて決定された真値項損失値を表し、 a は前記目標重みを表し、 gt は前記真値を表し、 $o2s$ は前記第3出力値を表し、 $o2t$ は前記第4出力値を表す、

請求項1に記載の軽量モデルトレーニング方法。

10

【請求項 8】

e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得するための第1取得モジュールであって、前記 e は最大反復回数である E 以下の正の整数であり、前記 E は1より大きい正の整数である第1取得モジュールと、それぞれ前記第1拡張確率と前記第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得るためのデータ拡張モジュールと、

前記第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、前記第2データセットに基づいて前記生徒モデルの第3出力値と前記教師モデルの第4出力値とを得るための予測モジュールと、

20

前記第1出力値と前記第2出力値とに基づいて蒸留損失関数を決定し、前記第3出力値と前記第4出力値とに基づいて真値損失関数を決定するための第1決定モジュールと、

前記蒸留損失関数と前記真値損失関数とに基づいて目標損失関数を決定するための第2決定モジュールと、

前記目標損失関数に基づいて前記生徒モデルをトレーニングするためのトレーニングモジュールと、

前記 e が前記 E よりも小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重みを決定するための第3決定モジュールと、を備える、

軽量モデルトレーニング装置。

【請求項 9】

30

前記軽量モデルトレーニング装置は、

最大拡張確率を取得するための第2取得モジュールと、

前記最大拡張確率と、前記最大反復回数と、前記第1拡張確率とに基づいて、前記第2拡張確率を決定するための第4決定モジュールと、をさらに備える、

請求項8に記載の軽量モデルトレーニング装置。

【請求項 10】

前記第3決定モジュールは、最大拡張確率と、前記最大反復回数と、前記 e 回目の前記第1拡張確率とに基づいて、前記 $e + 1$ 回目の反復に用いられるべき第1拡張確率を決定するための第1決定サブユニット、を備える、

40

請求項8に記載の軽量モデルトレーニング装置。

【請求項 11】

前記軽量モデルトレーニング装置は、

最大目標重みを取得するための第3取得モジュール、をさらに備え、

前記第3決定モジュールは、

前記最大目標重みと、前記最大反復回数と、前記 e 回目の前記目標重みとに基づいて、前記 $e + 1$ 回目の反復に用いられるべき目標重みを決定するための第2決定サブユニット、を備える、

請求項8に記載の軽量モデルトレーニング装置。

【請求項 12】

50

前記第 2 決定モジュールは、

前記目標重みが最大目標重み以上である場合に、または前記蒸留損失関数が前記真値損失関数以上である場合に、前記蒸留損失関数を前記目標損失関数として決定する一方、そうでない場合に、前記真値損失関数を前記目標損失関数として決定するための第 3 決定サブユニット、を備える、

請求項 8 に記載の軽量モデルトレーニング装置。

【請求項 13】

前記第 1 決定モジュールは、

下記の式により前記蒸留損失関数を決定するための第 4 決定サブユニット、を備え、

【数 3】

$$l1 = (a + a_{aft} * 2/E) * L_{dist}(o1s, o1t) + (1 - a - a_{aft} * 2/E) * L_{gt}(o1s, gt)$$

ここで、 $l1$ は前記蒸留損失関数を表し、 $L_{dist}(o1s, o1t)$ は前記第 1 出力値と前記第 2 出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o1s, gt)$ は前記第 1 出力値と真値とに基づいて決定された真値項損失値を表し、 a は前記目標重みを表し、 a_{aft} は最大目標重みを表し、 E は前記最大反復回数を表し、 gt は前記真値を表し、 $o1s$ は前記第 1 出力値を表し、 $o1t$ は前記第 2 出力値を表す、

請求項 8 に記載の軽量モデルトレーニング装置。

【請求項 14】

前記第 1 決定モジュールは、

下記の式により前記真値損失関数を決定するための第 5 決定サブユニット、を備え、

【数 4】

$$l2 = a * L_{dist}(o2s, o2t) + (1 - a) * L_{gt}(o2s, gt)$$

ここで、 $l2$ は前記真値損失関数を表し、 $L_{dist}(o2s, o2t)$ は前記第 3 出力値と前記第 4 出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o2s, gt)$ は前記第 3 出力値と真値とに基づいて決定された真値項損失値を表し、 a は前記目標重みを表し、 gt は前記真値を表し、 $o2s$ は前記第 3 出力値を表し、 $o2t$ は前記第 4 出力値を表す、

請求項 8 に記載の軽量モデルトレーニング装置。

【請求項 15】

少なくとも 1 つのプロセッサと、

前記少なくとも 1 つのプロセッサと通信接続されるメモリと、を備え、

前記メモリには、前記少なくとも 1 つのプロセッサで実行可能な命令が記憶されており、前記命令は、前記少なくとも 1 つのプロセッサによって実行されると、前記少なくとも 1 つのプロセッサに、請求項 1 から請求項 7 のいずれか 1 項に記載の方法を実行させる、電子デバイス。

【請求項 16】

コンピュータに請求項 1 から請求項 7 のいずれか 1 項に記載の方法を実行させるコンピュータ命令を記憶した非一時的なコンピュータ可読記憶媒体。

【請求項 17】

プロセッサにより実行されると、請求項 1 から請求項 7 のいずれか 1 項に記載の方法を実現するコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【 0 0 0 1 】

本開示は、コンピュータ技術分野に関し、特に、機械学習、コンピュータビジョン、画像処理などの技術分野に関する。

【背景技術】

【 0 0 0 2 】

知識蒸留はモデル圧縮の一般的な手法であり、生徒モデルも教師モデルと同様の精度メトリクスを得られるように、より精度の高い教師モデルを用いて生徒モデルのトレーニングプロセスを指導することを指す。知識蒸留自体は正則化の効果を持ち、オーバーフィッティングのモデルに対して知識蒸留の手法を用いてトレーニングを行うことで、ある程度オーバーフィッティングを回避することができる。しかし、一部の軽量モデルについては、知識蒸留策略を使用してトレーニングを行うと、さらにアンダーフィッティングをもたらし、モデルの精度の低下やモデルの非収束につながる可能性がある。

10

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 3 】

本開示は、軽量モデルトレーニング方法、画像処理方法、軽量モデルトレーニング装置、画像処理装置、電子デバイス、記憶媒体及びコンピュータプログラムを提供する。

【課題を解決するための手段】

【 0 0 0 4 】

本開示の第1の態様では、軽量モデルトレーニング方法を提供し、当該方法は、 e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得することと、ここで、 e は最大反復回数である E 以下の正の整数であり、 E は1より大きい正の整数であり、それぞれ第1拡張確率と第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得ることと、第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値とを得ることと、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定することと、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定することと、目標損失関数に基づいて生徒モデルをトレーニングし、 e が E よりも小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重みを決定することと、を含む。

20

30

【 0 0 0 5 】

本開示の第2の様態では、画像処理方法を提供し、当該方法は、目標シーンにおける処理対象画像を受信することと、処理対象画像を生徒モデルに入力し、生徒モデルが出力した、処理対象画像に対する処理結果を取得することと、を含み、ここで、生徒モデルは、第1の態様により提供された軽量モデルトレーニング方法を用いて得られる。

【 0 0 0 6 】

本開示の第3の様態では、軽量モデルトレーニング装置を提供し、当該装置は、 e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得するための第1取得モジュールと、ここで、 e は最大反復回数である E 以下の正の整数であり、 E は1より大きい正の整数であり、それぞれ第1拡張確率と第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得るためのデータ拡張モジュールと、第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値とを得るための予測モジュールと、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定するための第1決定モジュールと、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定するための第2決定モジュールと、目標損失関数に基づいて生徒モデルをトレーニングするためのトレーニングモジュールと、 e が E よりも小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重み

40

50

を決定するための第3決定モジュールと、を備える。

【0007】

本開示の第4の様態では、画像処理装置を提供し、当該装置は、目標シーンにおける処理対象画像を受信するための受信モジュールと、処理対象画像を生徒モデルに入力し、生徒モデルが出力した、処理対象画像に対する処理結果を取得するための処理モジュールと、を備え、ここで、生徒モデルは、第1の様態により提供された軽量モデルトレーニング方法を用いて得られる。

【0008】

本開示の第5の様態では、電子デバイスを提供し、当該電子デバイスは、少なくとも1つのプロセッサと、該少なくとも1つのプロセッサと通信接続されるメモリと、を備え、該メモリには、該少なくとも1つのプロセッサで実行可能な命令が記憶されており、該命令は、該少なくとも1つのプロセッサによって実行されると、該少なくとも1つのプロセッサに、上述の第1の様態及び/又は第2の様態により提供された方法を実行させる。

10

【0009】

本開示の第6の様態では、コンピュータに、上述の第1の様態及び/又は第2の様態により提供された方法を実行させるコンピュータ命令を記憶した非一時的なコンピュータ可読記憶媒体を提供する。

【0010】

本開示の第7の様態では、プロセッサにより実行されると、上述の第1の様態及び/又は第2の様態により提供された方法を実現するコンピュータプログラムを提供する。

20

【0011】

本実施の技術方案によれば、軽量モデルのトレーニング精度を向上させることができる。

【0012】

ここに記載された内容は、本開示の実施例のキーポイント又は重要な特徴を記述することを意図せず、また、本開示の範囲を制限することにも用いられないことを理解すべきである。本開示の他の特徴については、下記の明細書を通して説明を促す。

【図面の簡単な説明】

【0013】

添付図面において、別段の定めがない限り、複数の図面の同一の符号は同一または類似の部品または要素を示す。これらの図面は必ずしも比例して描かれているとは限らない。これらの図面は、本願によって開示された一部の実施形態のみを図示しており、本願の範囲を限定するものとみなされるべきではないことを理解されたい。

30

【図1】本開示の実施例による軽量モデルトレーニング方法のフローを示す概略図である。

【図2】本開示の実施例による知識蒸留に基づく軽量モデルのトレーニングのフローを示す概略図である。

【図3】本開示の実施例による画像処理方法のフローを示す概略図である。

【図4】本開示の実施例による軽量モデルトレーニング装置の構成を示す概略図である。

【図5】本開示の実施例による画像処理装置の構成を示す概略図である。

【図6】本開示の実施例による軽量モデルのトレーニングシーンを示す概略図である。

【図7】本開示の実施例による画像処理シーンを示す概略図である。

40

【図8】本開示の実施例による軽量モデルトレーニング方法又は画像処理方法を実施するための電子デバイスのブロック図である。

【発明を実施するための形態】

【0014】

以下では、本開示の例示的な実施例を、理解を容易にするために本開示の実施例の様々な詳細を含む添付の図面に関連して説明するが、これらは単に例示的なものであると考えるべきである。したがって、当業者は、本開示の範囲及び精神を逸脱することなく、本明細書に記載された実施例に様々な変更及び修正を加えることができることを認識すべきである。同様に、以下の説明では、周知の機能及び構成については、明確化及び簡明化のために説明を省略する。

50

【 0 0 1 5 】

本開示の実施例において、「第1」、「第2」等の用語は、特定の順序又は前後関係を示すために用いられるものではなく、類似する対象を区別するためのものである。また、「含む」、「有する」及びこれらに準ずるその他の用語は、非排他的に含む場合も含むという意味で用いられる。例えば、一連のステップ又はユニットを含むプロセス、方法、システム、製品又は機器は、明記されているステップ又はユニットに限られるわけではなく、明記されていない又はこれらのプロセス、方法、製品又は機器に固有のその他のステップ又はユニットを含んでもよい。

【 0 0 1 6 】

関連技術の中で、知識蒸留法は主に、モデル出力に基づく蒸留であって、すなわち、教師モデルの出力を用いて生徒モデルの出力を直接監督すること、中間層特徴に基づく蒸留であって、すなわち、教師モデルの中間特徴図を用いて生徒モデルの中間特徴図の出力を監視すること、関係に基づく蒸留であって、すなわち、教師モデルの異なるサンプルに対する類似度などのメトリックと生徒モデルの差異を考慮し、これに基づいて生徒モデルのトレーニングを指導することのようなものがある。

10

【 0 0 1 7 】

上記の知識蒸留法について、蒸留の損失関数及び特徴選択プロセスに対する設計と選択を行うことが多いが、生徒モデルのフィッティング状態に対する研究は少ない。

【 0 0 1 8 】

知識蒸留過程において、損失関数は蒸留タスクに対する損失関数（知識蒸留損失関数と呼ぶことができる）と真値の損失関数（真値損失関数と呼ぶことができる）を含み、ここで、知識蒸留損失関数はモデルの汎化性能を増加することができ、正規化の効果を発揮させるが、真値損失関数は、モデルを迅速に収束させることができる。

20

【 0 0 1 9 】

ディープラーニングタスクでは、モデルの汎化性能を向上させるために、一般的に大量のデータ拡張過程が含まれるが、トレーニング過程において、一般的にデータ拡張の操作や確率などが変化しないことを保証するため、それによってモデルのトレーニング過程に不一致をもたらしやすい。また、トレーニング前期では、モデルトレーニングの目標はより早く収束することであり、トレーニング後期では、モデルトレーニングの目標はより良く汎化することであるため、データ拡張が変化しないことを維持すると、2つのトレーニング段階の目標を同時に満たすことができない。

30

【 0 0 2 0 】

現在の知識蒸留アルゴリズムは、生徒モデルのフィッティング状態を考慮しておらず、教師モデルを用いて生徒モデルの出力を監督すると、小規模モデルの知識蒸留過程にアンダーフィッティングの状態が生じやすく、それによってモデルの精度を低下させる。

【 0 0 2 1 】

本開示は、上記の問題及び他の潜在的な問題のうちの1つまたは複数を少なくとも部分的に解決するためになされたものであり、知識蒸留戦略を用いて軽量モデルをトレーニングする際に発生するアンダーフィッティング現象を効果的に回避し、軽量モデルのトレーニング精度を高め、軽量モデルの知識蒸留の精度を高めることができる軽量モデルトレーニング方法を提案する。

40

【 0 0 2 2 】

本開示の実施例は、軽量モデルトレーニング方法を提供し、図1は、本開示の実施例による、軽量モデルトレーニング装置に適用可能な軽量モデルトレーニング方法のフローを示す概略図である。該軽量モデルトレーニング装置は、固定デバイス及び/またはモバイルデバイスを含むがこれに限定されない電子デバイスに配置されてもよい。例えば、固定デバイスは、サーバーを含むがこれに限定されず、サーバーはクラウドサーバーであってもよいし、通常のサーバーであってもよい。例えば、モバイルデバイスは、携帯電話、タブレットのうちの1つまたは複数の端末を含むがこれらに限定されない。いくつかの可能な実現形態では、該軽量モデルトレーニング方法は、プロセッサがメモリに記憶されたコン

50

コンピュータ可読命令を呼び出すことによって実現されることもできる。図 1 に示すように、該軽量モデルトレーニング方法は以下のステップを含む。

【0023】

S101において、e回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得し、ここで、eは最大反復回数であるE以下の正の整数であり、Eは1より大きい正の整数である。

【0024】

S102において、それぞれ第1拡張確率と第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得る。

【0025】

S103において、第1データセットに基づいて生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値とを得、ここで、該生徒モデルは軽量モデルである。

【0026】

S104において、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定する。

【0027】

S105において、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定する。

【0028】

S106において、目標損失関数に基づいて生徒モデルをトレーニングし、eがEよりも小さい場合に、e+1回目の反復に用いられるべき第1拡張確率または目標重みを決定する。

【0029】

本開示の実施例では、生徒モデルは画像処理を行うために用いられる。本開示は、生徒モデルの機能をそれ自体のシーン及び用途に応じて限定しない。たとえば、生徒モデルは画像分類に用いられる。また、例えば、生徒モデルは画像認識に用いられる。別の例として、生徒モデルは目標検出に用いられる。以上は例示的な説明に過ぎず、生徒モデルの全ての可能な機能タイプを限定するものではなく、ここでは列挙しない。

【0030】

本開示の実施例では、教師モデルはすでにトレーニング済みのモデルである。なお、本開示は教師モデルの大きさを限定するものではなく、教師モデルは大規模モデルであっても小規模モデルであってもよい。

【0031】

本開示の実施例では、Eの値は、速度要求や精度要求などの要求に応じて設定されてもよい。

【0032】

本開示の実施例では、第2拡張確率は、第1拡張確率とは異なる値である。通常、第2拡張確率の値は、第1拡張確率の値よりも大きい。

【0033】

本開示の実施例では、総損失関数は、蒸留項損失と真値項損失とを含み、ここで、蒸留項損失の重みが α である場合に、真値項損失重みは $1 - \alpha$ である。

【0034】

本開示の実施例では、e回目の反復が開始する前に、e回目の反復に入力されるデータセットは、e-1回目の反復で生成された第1データセットと第2データセットとに基づく集合であってもよいし、e-1回目の反復で生成された第1データセットと第2データセットとに基づく部分的な集合であってもよいし、元のデータセットであってもよい。以上は例示的な説明に過ぎず、e回目の反復に入力させるデータセットに含まれるデータの全ての可能なタイプを限定するものではなく、ここでは列挙しない。

【0035】

本開示の実施例では、データ拡張は、データセット(トレーニングセットとも呼ばれる

10

20

30

40

50

)のデータ量を増やすために用いられる。データ拡張により、データセットの多様性を増やし、オーバーフィッティングを減らしてモデルの汎化能力を向上させることができる。

【0036】

本開示の実施例において、蒸留損失関数、真値損失関数、及び目標損失関数を如何に決定することについて、後続の実施例で詳細に説明するので、ここではこれ以上述べない。

【0037】

本開示の実施例では、生徒モデルのネットワーク構造は、畳み込みニューラルネットワーク(Convolutional Neural Networks, CNN)、transformerを含むことができる。以上は例示的な説明に過ぎず、生徒モデルに含まれるネットワークの全ての可能なタイプを限定するものではなく、ここでは列挙しない。

【0038】

本開示の実施例に記載された技術方案は、 e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び目標重みを取得し、それぞれ第1拡張確率と第2拡張確率とに基づいてデータセットに対してデータ拡張を行って、第1データセットと第2データセットとを得、第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値とを得、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定し、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定し、目標損失関数に基づいて生徒モデルをトレーニングし、 e が E より小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重みを決定する。このように、データ拡張により、データセットの多様性を増やし、オーバーフィッティングを減らして軽量モデルの汎化能力を向上させることができる。同時に、 e が E より小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率或いは目標重みを決定することは、トレーニング過程全体における第1拡張確率或いは目標重みを動的に調整することができ、知識蒸留策略を使用して軽量モデルをトレーニングする際に現れたアンダーフィッティング現象を有効に回避することができ、それによってモデルのトレーニング精度を高めることができる。

【0039】

いくつかの実施例では、該軽量モデルトレーニング方法は、以下のステップをさらに含むことができる。

【0040】

S107において、最大拡張確率を取得し、最大拡張確率と、最大反復回数と、第1拡張確率とに基づいて第2拡張確率を決定する。

【0041】

本開示の実施例では、第1拡張確率が p_i 、最大拡張確率が p_{aug} 、最大反復回数が E であると、第2拡張確率は、

【数1】

$$p_i + p_{aug} * 2/E$$

となる。

【0042】

ここで、最大拡張確率は、速度要求や精度要求、あるいは汎化能力要求などの要求に応じて設定または調整されてもよい。

【0043】

本開示の実施例では、S107はS101の前に実行される。

【0044】

本開示は、最大拡張確率の取得方式について限定しない。例えば、ユーザインタフェースを介してトレーニング関連パラメータを受信する。また、例えば、端末から送信されたトレーニング関連パラメータを受信する。該トレーニング関連パラメータは、最大拡張確率を含む。以上は例示的な説明に過ぎず、最大拡張確率の全ての可能なソースを限定する

10

20

30

40

50

ものではなく、ここでは列挙しない。

【0045】

なお、最大拡張確率の設定値は生徒モデルごとに異なっていてもよい。

【0046】

このようにして、決定された第2拡張確率を、第1拡張確率及び最大反復回数の両方に関連付けると共に、最大拡張確率にも関連付けることができ、第2データセットの汎化性と正確性を向上させることができ、決定された生徒モデルの出力した第3出力値と教師モデルの出力した第4出力値をより正確にすることができ、目標損失関数の選択の精度を向上させることに役立ち、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

10

【0047】

いくつかの実施例では、 $e + 1$ 回目の反復に用いられるべき第1拡張確率を決定することは、最大拡張確率と、最大反復回数と、 e 回目の第1拡張確率とに基づいて、 $e + 1$ 回目の反復に用いられるべき第1拡張確率を決定することを含む。

【0048】

本開示の実施例では、 $e + 1$ 回目の反復に用いられるべき第1拡張確率は

【数2】

$$p_i = p_i + p_{aug} * 2/E$$

である。ここで、第1拡張確率は p_i であり、最大拡張確率は p_{aug} であり、最大反復回数は E である。

20

【0049】

このようにして、決定された第1拡張確率を、前回の第1拡張確率及び最大反復回数の両方に関連付けると共に、最大拡張確率にも関連付けることができ、第1データセットの汎化性及び正確性を向上させることができ、決定された生徒モデルの出力した第1出力値及び教師モデルの出力した第2出力値をより正確にすることができ、目標損失関数の選択の精度を向上させることに役立ち、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

【0050】

いくつかの実施例では、該軽量モデルトレーニング方法は、S108において、最大目標重みを取得することをさらに含むことができる。

30

【0051】

本開示の実施例では、S108はS101の前に実行される。S108はS107の前に実行されてもよく、S108はS107の後に実行されてもよく、S108はS107と同時に実行されてもよい。

【0052】

本開示は、最大目標重みの取得方法について限定しない。例えば、ユーザインタフェースを介してトレーニング関連パラメータを受信する。また、例えば、端末から送信されたトレーニング関連パラメータを受信する。該トレーニング関連パラメータは、最大目標重みを含む。以上は例示的な説明に過ぎず、最大目標重みの全ての可能な取得方法を限定するものではなく、ここでは列挙しない。

40

【0053】

なお、最大目標重みの設定値は、生徒モデルごとに異なっていてもよい。

【0054】

いくつかの実施例では、 $e + 1$ 回目の反復に用いられるべき目標重みを決定することは、最大目標重みと、最大反復回数と、 e 回目の目標重みとに基づいて、 $e + 1$ 回目の反復に用いられるべき目標重みを決定することを含む。

【0055】

本開示の実施例では、 $e + 1$ 回目の反復に用いられるべき最大目標重みは

【数3】

50

$$a = a + a_{dft} * 2/E$$

である。ここで、 e 回目の目標重みは a であり、最大目標重みは a_{dft} であり、最大反復回数は E である。

【0056】

このようにして、決定された目標重みを、前回の目標重み及び最大反復回数の両方に関連付けると共に、最大目標重みにも関連付けることができ、目標損失関数の選択の精度を向上させることに役立ち、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

【0057】

本開示の実施例では、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定することは、目標重みが最大目標重み以上である場合に、または蒸留損失関数が真値損失関数以上である場合に、蒸留損失関数を目標損失関数として決定する一方、そうでない場合に、真値損失関数を目標損失関数として決定する。

【数4】

本開示の実施例では、 $a \geq a_{dft}$ 、または、 $I1 \geq I2$ の場合に、損失関数 $I2$ を用いてモデルの逆伝播と勾配更新を行い、 $e + 1$ 回目の反復に用いられるべき第1拡張確率を、

$$p_i = p_i + p_{aug} * 2/E$$

と設定する一方、そうでない場合に、損失関数 $I1$ を用いてモデルの逆伝播と勾配更新を行い、 $e + 1$ 回目の反復に用いられるべき目標重みを、

【数5】

$$a = a + a_{dft} * 2/E$$

と設定する。

【0058】

このように、トレーニング過程において、トレーニング状況に応じて適切な目標損失関数を選択することができ、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

【0059】

本開示の実施例では、第1出力値と第2出力値とに基づいて蒸留損失関数を決定することは、蒸留損失関数を次の式に基づいて決定することを含む。

【数6】

$$l1 = (a + a_{dft} * 2/E) * L_{dist}(o1s, o1t) + (1 - a - a_{dft} * 2/E) * L_{gt}(o1s, gt)$$

ここで、 $l1$ は蒸留損失関数を表し、 $L_{dist}(o1s, o1t)$ は第1出力値と第2出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o1s, gt)$ は第1出力値と真値とに基づいて決定された真値項損失値を表し、 a は目標重みを表し、 a_{dft} は最大目標重みを表し、 E は最大反復回数を表し、 gt は真値を表し、 $o1s$ は第1出力値を表し、 $o1t$ は第2出力値を表す。

【0060】

このようにして、蒸留損失関数を迅速に決定することができ、蒸留損失関数を、目標重み及び最大反復回数に関連付けることができ、蒸留損失関数の精度を向上させることに役立つ、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

【0061】

本開示の実施例では、第3出力値と第4出力値とに基づいて真値損失関数を決定することは、真値損失関数を次の式に基づいて決定することを含む。

【数7】

10

20

30

40

50

$$l2 = a * L_{dist}(o2s, o2t) + (1 - a) * L_{gt}(o2s, gt)$$

ここで、 $l2$ は真値損失関数を表し、 $L_{dist}(o2s, o2t)$ は第3出力値と第4出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o2s, gt)$ は第3出力値と真値とに基づいて決定された真値項損失値を表し、 a は目標重みを表し、 gt は真値を表し、 $o2s$ は第3出力値を表し、 $o2t$ は第4出力値を表す。

【0062】

これにより、真値損失関数を迅速に決定することができ、真値損失関数を、目標重み及び最大反復回数に関連付けることができ、真値損失関数の精度を向上させることに役立つ、ひいては軽量モデルのトレーニング精度を向上させることに役立つ。

10

【0063】

図2は、知識蒸留に基づく軽量モデルのレーニングのフローを示す概略図であり、図2に示すように、該フローは以下のものを含む。

【0064】

S201において、生徒モデル(F_s)と教師モデル(F_t)とを定義し、最大反復回数 E を定義し、初期反復回数 $e = 1$ と定義する。

【0065】

ここで、教師モデルはすでにトレーニング済みのモデルである。

【数8】

S202において、 e 回目の反復過程において、第1拡張確率 p_i と第2拡張確率

$$p_i = p_i + p_{aug} * 2/E$$

を用いて元のデータを拡張して前処理し、処理後のデータ b_1 と b_2 とを得る。

S203において、処理後のデータ b_1 を生徒モデルと教師モデルとに送り、それぞれ生徒モデルの出力する $o1s$ と教師モデルの出力する $o1t$ とを得る。

S204において、処理後のデータ b_2 を生徒モデルと教師モデルとに送り、それぞれ生徒モデルの出力する $o2s$ と教師モデルの出力する $o2t$ とを得る。

S205において、損失関数 $l1$ と損失関数 $l2$ とを計算する。

S206において、 $a \geq a_{dft}$ 、または、 $l1 \geq l2$ を満たすか否かを判断し、YESであればS207を実行する一方、NOであれば、S208を実行する。

30

S207において、 $l2$ を目標損失関数として決定し、 $e < E$ の場合に、 $e + 1$ 回目の反復で用いられるべき第1拡張確率を、

【数9】

$$p_i = p_i + p_{aug} * 2/E$$

と設定する。

S208において、 $l1$ を目標損失関数として決定し、 $e < E$ の場合に、 $e + 1$ 回目の反復に用いられるべき目標重みを、

40

【数10】

$$a = a + a_{dft} * 2/E$$

と設定する。

【0066】

S209において、目標損失関数に基づいて生徒モデルに対してモデルの逆伝播と勾配更新を行う。

50

【0067】

S210において、 $e < E$ を満たすか否かを判断し、YESであれば $e = e + 1$ と設定して、ステップS202に戻る一方、そうでなければ、S211を実行する。

【0068】

S211において、モデルを保存して、反復を終了する。

【0069】

このようにして、知識蒸留に基づく軽量モデルトレーニングの過程において、データ拡張確率、蒸留損失関数重みがトレーニング過程に与える影響を考慮して、トレーニング過程において、蒸留損失関数の重み、データ拡張の確率など、正則化に関連するモデルパラメータを動的に調整し、最終的に軽量モデル知識蒸留の精度を向上させることを提案した。

10

【0070】

以上の処理フローは全ての知識蒸留アルゴリズムと融合することができ、異なる視覚タスクに適用することができる。視覚タスクには、画像認識、画像分類、目標検出、画像分割などの処理を含むがこれに限定されない画像処理が含まれる。

【0071】

図2に示されたフローチャートは単に概略的なものであって限定的なものではなく、当業者は図2の例に基づいて様々な明白な変更及び/または置換を行うことができ、結果として得られた技術方案はなお本開示の実施例の開示範囲に属することが理解されるべきである。

【0072】

20

上述した軽量モデルトレーニング方法によってトレーニングされた生徒モデルに基づいて、本開示の実施例では、コンピュータ、携帯電話、タブレット等を含むがこれに限定されない電子デバイスに適用される画像処理方法を開示し、本開示は電子デバイスの種類について限定しない。図3に示すように、該画像処理方法は以下のステップを含む。

【0073】

S301において、目標シーンにおける処理対象画像を受信する。

【0074】

S302において、処理対象画像を生徒モデルに入力し、生徒モデルが出力した、処理対象画像に対する処理結果を取得する。

【0075】

30

ここで、該生徒モデルは、図1に示す軽量モデルトレーニング方法を用いて得られる。

【0076】

このようにして、トレーニング済みの生徒モデルを用いて画像処理を行うことで、画像処理の精度を向上させることができる。

【0077】

いくつかの実施例では、目標シーンにおける処理対象画像を受信することは、画像分類シーンにおける処理対象画像を取得することと、画像認識シーンにおける処理対象画像を取得することと、目標検出シーンにおける処理対象画像を取得することと、うちの少なくとも1つを含む。

【0078】

40

目標シーンは、以上に列挙したシーンを含むがこれに限定されるものではなく、画像処理に関連する全てのシーンを目標シーンとすることができ、ここでは列挙しないことを理解されたい。

【0079】

このようにして、トレーニング済みの生徒モデルを用いて、異なる目標シーンにおける画像を処理することができ、ひいては生徒モデルの画像処理能力を向上させることができる。

【0080】

本開示の実施例では、軽量モデルトレーニング装置を提供し、図4に示すように、該装置は、 e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重み

50

である目標重みを取得するための第1取得モジュール401であって、ここで、 e は最大反復回数である E 以下の正の整数であり、 E は1より大きい正の整数である第1取得モジュール401と、それぞれ第1拡張確率と第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得るためのデータ拡張モジュール402と、第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値とを得るための予測モジュール403と、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定するための第1決定モジュール404と、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定するための第2決定モジュール405と、目標損失関数に基づいて生徒モデルをトレーニングするためのトレーニングモジュール406と、 e が E よりも小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重みを決定するための第3決定モジュール407と、を備える。

10

【0081】

いくつかの実施例では、該軽量トレーニング装置は、最大拡張確率を取得するための第2取得モジュール408（図4に示されていない）と、最大拡張確率と、最大反復回数と、第1拡張確率とに基づいて、第2拡張確率を決定するための第4決定モジュール409（図4に図示せず）と、をさらに備えることができる。

【0082】

いくつかの実施例では、該第3決定モジュール407は、最大拡張確率と、最大反復回数と、 e 回目の第1拡張確率とに基づいて、 $e + 1$ 回目の反復に用いられるべき第1拡張確率を決定するための第1決定サブユニットを備える。

20

【0083】

いくつかの実施例では、該軽量トレーニング装置は、最大目標重みを取得するための第3取得モジュール410（図4に示されていない）をさらに備えることができ、ここで、該第3決定モジュール407は、最大目標重みと、最大反復回数と、 e 回目の目標重みとに基づいて、 $e + 1$ 回の反復に用いられるべき目標重みを決定するための第2決定サブユニットを備える。

【0084】

いくつかの実施例では、第2決定モジュール405は、目標重みが最大目標重み以上である場合に、または蒸留損失関数が真値損失関数以上である場合に、蒸留損失関数を目標損失関数として決定する一方、そうでない場合に、真値損失関数を目標損失関数として決定するための第3決定サブユニットを備える。

30

【0085】

いくつかの実施例では、該第1決定モジュール404は、下記の式により蒸留損失関数を決定するための第4決定サブユニットを備える。

【数11】

$$l1 = (a + a_{dft} * 2/E) * L_{dist}(o1s, o1t) + (1 - a - a_{dft} * 2/E) * L_{gt}(o1s, gt)$$

ここで、 $l1$ は蒸留損失関数を表し、 $L_{dist}(o1s, o1t)$ は第1出力値と第2出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o1s, gt)$ は第1出力値と真値とに基づいて決定された真値項損失値を表し、 a は目標重みを表し、 a_{dft} は最大目標重みを表し、 E は最大反復回数を表し、 gt は真値を表し、 $o1s$ は第1出力値を表し、 $o1t$ は第2出力値を表す。

40

【0086】

いくつかの実施例では、該第1決定モジュール404は、下記の式により真値損失関数を決定するための第5決定サブユニットを備える。

【数12】

50

$$l2 = a * L_{dist}(o2s, o2t) + (1 - a) * L_{gt}(o2s, gt)$$

ここで、 $l2$ は真値損失関数を表し、 $L_{dist}(o2s, o2t)$ は第3出力値と第4出力値とに基づいて決定された蒸留項損失値を表し、 $L_{gt}(o2s, gt)$ は第3出力値と真値とに基づいて決定された真値項損失値を表し、 a は目標重みを表し、 gt は真値を表し、 $o2s$ は第3出力値を表し、 $o2t$ は第4出力値を表す。

【0087】

本開示の実施例による軽量モデルトレーニング装置における各処理モジュールの機能は、前述の軽量モデルトレーニング方法の関連する説明を参照して当業者に理解されるものとし、本開示の実施例による軽量モデルトレーニング装置における各処理モジュールは、本開示の実施例に記載された機能を実現するアナログ回路によって実装されてもよいし、本開示の実施例に記載された機能を実行するソフトウェアが電子デバイス上で動作することによって実装されてもよいことは当業者に理解されるものとする。

10

【0088】

本開示の実施例による軽量モデルトレーニング装置は、軽量モデルのトレーニング精度を向上させることができる。

【0089】

本開示の実施例では、電子デバイスに適用される画像処理装置を提供し、図5に示すように、該装置は、目標シーンにおける処理対象画像を受信するための受信モジュール501と、処理対象画像を生徒モデルに入力し、生徒モデルが出力した、処理対象画像に対する処理結果を取得するための処理モジュール502と、を備え、ここで、該生徒モデルは、軽量モデルトレーニング方法を用いて得られる。

20

【0090】

いくつかの実施例では、該受信モジュール501が目標シーンにおける処理対象画像を受信することは、画像分類シーンにおける処理対象画像を取得することと、画像認識シーンにおける処理対象画像を取得することと、目標検出シーンにおける処理対象画像を取得することと、のうちの少なくとも1つを含む。

【0091】

本開示の実施例による画像処理装置における各処理モジュールの機能は、前述の画像処理方法の関連する説明を参照して当業者に理解されるものとし、本開示の実施例による画像処理装置における各処理モジュールは、本開示の実施例に記載された機能を実現するアナログ回路によって実装されてもよいし、本開示の実施例に記載された機能を実行するソフトウェアが電子デバイス上で動作することによって実装されてもよいことは、当業者に理解されるものとする。

30

【0092】

本開示の実施例による画像処理装置は、画像処理の精度を向上させることができる。

【0093】

本開示の実施例はまた、軽量モデルのトレーニングシーンを示す概略図を提供し、図6に示すように、クラウドサーバーのような電子デバイスは、端末から送信されたトレーニングタスクに基づいて、各トレーニングタスクに教師モデルと、トレーニングすべき生徒モデルと、を決定する。ここで、異なる端末から送信されるトレーニングタスクは、異なる使用シーンに対するトレーニングタスクであってもよい。電子デバイスは、生徒モデルをトレーニングし、以下のステップを含み、すなわち、 e 回目の反復に用いられる第1拡張確率、第2拡張確率、及び蒸留項損失値の重みである目標重みを取得することによって、ここで、 e は最大反復回数である E 以下の正の整数であり、 E は1より大きい正の整数であること、それぞれ第1拡張確率と第2拡張確率とに基づいて、データセットに対してデータ拡張を行って、第1データセットと第2データセットとを得ることと、第1データセットに基づいて軽量モデルである生徒モデルの第1出力値と教師モデルの第2出力値とを得、第2データセットに基づいて生徒モデルの第3出力値と教師モデルの第4出力値と

40

50

を得ることと、第1出力値と第2出力値とに基づいて蒸留損失関数を決定し、第3出力値と第4出力値とに基づいて真値損失関数を決定することと、蒸留損失関数と真値損失関数とに基づいて目標損失関数を決定することと、目標損失関数に基づいて生徒モデルをトレーニングし、 e が E よりも小さい場合に、 $e + 1$ 回目の反復に用いられるべき第1拡張確率または目標重みを決定することと、を含む。電子デバイスはトレーニング済みの生徒モデルを各端末に返す。

【0094】

また、本開示の実施例はまた、画像処理のシーンを示す概略図を提供し、図7に示すように、クラウドサーバー等の電子デバイスは、各端末から送信された処理対象画像と指定された目標シーンを受信し、各処理対象画像を、目標シーンに対応する生徒モデルに入力し、対応する生徒モデルが出力した画像処理結果を得る。

10

【0095】

本開示では、端末及び電子デバイスの数について限定せず、実用的には、複数の端末、複数の電子デバイスを含むことができる。

【0096】

図6及び図7に示すシーンの図は、単に例示的なものであり、限定的なものではなく、当業者は、図6及び図7の例に基づいて様々な明白な変更及び/または置換を行うことができ、結果として得られた技術方案はなお本開示の実施例の開示範囲に属することは理解されるべきである。

【0097】

本開示の技術方案に関わる利用者の個人情報の取得、保管及び利用等は、関係する法令の規定に準拠し、公序良俗に反しないものである。

20

【0098】

本開示の実施例によれば、本開示はまた、電子デバイス、可読記憶媒体、及びコンピュータプログラム製品を提供する。

【0099】

図8は、本開示の実施例を実装するための電子デバイス800のブロック図である。電子デバイスは、各形式のデジタルコンピュータを指し、例えば、ラップトップコンピュータ、デスクトップコンピュータ、ワークステーション、パーソナルデジタルアシスタント、サーバー、ブレードサーバー、大型コンピュータ、及びその他の適合するコンピュータが挙げられる。電子デバイスは、各形式の移動装置をさらに指し、例えば、パーソナルデジタルアシスタント、セルラー電話、インテリジェントフォン、ウェアラブルデバイス、及びその他の類似のコンピュータ装置が挙げられる。本開示に記載されているコンポーネント、それらの接続関係、及び機能は例示的なものに過ぎず、本開示に記載・特定されているものの実現を限定するわけではない。

30

【0100】

図8に示すように、デバイス800は、リードオンリーメモリ(ROM)802に記憶されたコンピュータプログラム命令、又は記憶ユニット808からランダムアクセスメモリ(RAM)803にローディングされたコンピュータプログラム命令に基づいて、各種の適切な動作と処理を実行できる計算ユニット801を含む。RAM803には、デバイス800の動作に必要な各種のプログラム及びデータをさらに記憶することができる。計算ユニット801と、ROM802と、RAM803とは、バス804を介して互いに接続されている。入力/出力(Input/Output, I/O)インターフェース805もバス804に接続されている。

40

【0101】

デバイス800における複数のコンポーネントは、I/Oインターフェース805に接続されており、その複数のコンポーネントは、キーボードやマウス等の入力ユニット806と、種々なディスプレイやスピーカ等の出力ユニット807と、磁気ディスクや光学ディスク等の記憶ユニット808と、ネットワークカード、モデム、無線通信トランシーバ一等の通信ユニット809と、を備える。通信ユニット809は、デバイス800がイン

50

ターネットのようなコンピュータネット及び/又は種々なキャリアネットワークを介して他の機器と情報/データを交換することを許可する。

【0102】

計算ユニット801は、処理及びコンピューティング能力を有する様々な汎用及び/又は専用の処理コンポーネントであってもよい。計算ユニット801のいくつかの例としては、中央処理装置(Central Processing Unit, CPU)、グラフィックス処理ユニット(Graphics Processing Unit, GPU)、様々な専用の人工知能(Artificial Intelligence, AI)コンピューティングチップ、様々な機械学習モデルアルゴリズムを実行する計算ユニット、デジタル信号プロセッサ(Digital Signal Processor, DSP)、及び任意の適切なプロセッサ、コントローラ、マイクロコントローラ等を備えるが、これらに限定されない。計算ユニット801は、上述で説明された各方法及び処理、例えば軽量モデルトレーニング方法及び/または画像処理方法を実行する。例えば、いくつかの実施例では、軽量モデルトレーニング方法及び/または画像処理方法を、記憶ユニット808のような機械読み取り可能な媒体に有形的に含まれるコンピュータソフトウェアプログラムとして実現することができる。一部の実施例では、コンピュータプログラムの一部又は全ては、ROM802及び/又は通信ユニット809を介して、デバイス800にロード及び/又はインストールすることができる。コンピュータプログラムがRAM803にロードされて計算ユニット801によって実行される場合に、前述した軽量モデルトレーニング方法及び/または画像処理方法の1つ又は複数のステップを実行することができる。追加可能に、他の実施例では、計算ユニット801は、他の任意の適当な方式(例えば、ファームウェア)により軽量モデルトレーニング方法及び/または画像処理方法を実行するように構成することができる。

【0103】

ここで記載されているシステム又は技術の各種の実施例は、デジタル電子回路システム、集積回路システム、フィールドプログラマブルゲートアレイ(Field Programmable Gate Array, FPGA)、特定用途向け集積回路(Application Specific Integrated Circuit, ASIC)、特定用途向け標準品(Application Specific Standard Products, ASSP)、システムオンチップ(System on Chip, SOC)、コンプレックスプログラマブルロジックデバイス(Complex Programmable Logic Device, CPLD)、コンピュータのハードウェア、ファームウェア、ソフトウェア、及び/又はこれらの組み合わせによって実現することができる。これらの各実施形態は、少なくとも1つのプログラマブルプロセッサを含むプログラマブルシステムにて実行及び/又は解釈される1つ又は複数のコンピュータプログラムにより実行することを含み得、該プログラマブルプロセッサは、ストレージシステム、少なくとも1つの入力デバイス、及び少なくとも1つの出力デバイスからデータ及び命令を受け取り、データ及び命令を該ストレージシステム、該少なくとも1つの入力デバイス、及び該少なくとも1つの出力デバイスに転送することができる専用又は汎用のプログラマブルプロセッサであってもよい。

【0104】

本開示の方法を実行するためのプログラムコードは、1つ又は複数のプログラミング言語の任意の組み合わせで作成することができる。これらのプログラムコードは、汎用コンピュータ、専用コンピュータ又は他のプログラミングデータ処理装置のプロセッサ又はコントローラに提供されることにより、プログラムコードがプロセッサ又はコントローラによって実行される場合に、フローチャート及び/又はブロック図に規定された機能/動作を実行することができる。プログラムコードは、完全にマシンで実行されてもよいし、部分的にマシンで実行されてもよいし、独立したソフトパッケージとして部分的にマシンで実行されるとともに部分的にリモートマシンで実行されてもよいし、又は完全にリモートマシン又はサーバーで実行されてもよい。

10

20

30

40

50

【0105】

本開示の説明において、機械読み取り可能な媒体は、有形な媒体であってもよく、命令実行システム、装置又は機器によって、又は命令実行システム、装置又は機器と合わせて用いられるプログラムを含み、又は記憶する。機械読み取り可能な媒体は、機械読み取り可能な信号媒体又は機械読み取り可能な記憶媒体であってもよい。機械読み取り可能な媒体は、電子、磁気、光学、電磁、赤外線、又は半導体システム、装置、又はデバイス、又は前述した内容の任意の適切な組み合わせを含むことができるがこれらに限定されない。機械読み取り可能な記憶媒体のさらなる具体例として、1つ又は複数の配線による電氣的接続、ポータブルコンピュータディスクカートリッジ、ハードディスク、ランダムアクセスメモリ、リードオンリーメモリ、消去可能なプログラマブルリードオンリーメモリ (Erasable Programmable Read-Only Memory, EPROM)、フラッシュメモリ、光ファイバー、ポータブルコンパクトディスクリードオンリーメモリ (Compact Disk Read Only Memory, CD-ROM)、光学記憶装置、磁気記憶装置、又は前述した内容の任意の組み合わせを含む。

10

【0106】

ユーザとのインタラクションを提供するために、コンピュータでここに記載されているシステム及び技術を実施することができ、該コンピュータは、ユーザに情報を表示するための表示装置 (例えば、陰極線管 (Cathode Ray Tube, CRT) 又は液晶ディスプレイ (Liquid Crystal Display, LCD) モニター)、ユーザが入力をコンピュータに提供するためのキーボード及びポインティングデバイス (例えば、マウス又はトラックボール等) を備えることができる。ユーザとのインタラクションを提供するために、他の種類の装置を使用することもでき、例えば、ユーザに提供するフィードバックは、いかなる形式のセンサーフィードバック (例えば、視覚フィードバック、聴覚フィードバック、又は触覚フィードバック等) であってもよく、また、いかなる形式 (例えば、音響入力、音声入力、触覚入力等) によって、ユーザからの入力を受付取るができる。

20

【0107】

ここに記載されているシステムと技術を、バックグラウンド部品に含まれるコンピューティングシステム (例えば、データサーバーとして)、又はミドルウェア部品を含むコンピューティングシステム (例えば、アプリケーションサーバー)、又はフロント部品を含むコンピューティングシステム (例えば、グラフィックスユーザーインターフェース又はネットワークブラウザを有するユーザコンピュータが挙げられ、ユーザがグラフィックスユーザーインターフェース又は該ネットワークブラウザによって、ここに記載されているシステムと技術の実施形態とインタラクションすることができる)、又はこのようなバックグラウンド部品、ミドルウェア部品、又はフロント部品のいかなる組合したコンピューティングシステムで実施することができる。如何なる形式又はメディアのデジタルデータ通信 (例えば、通信ネットワーク) を介して、システムの部品を互いに接続することができる。通信ネットワークの例は、ローカルエリアネットワーク (Local Area Network, LAN)、ワイドエリアネットワーク (Wide Area Network, WAN) 及びインターネットを含む。

30

40

【0108】

コンピュータシステムは、クライアントとサーバーを含み得る。通常、クライアントとサーバーは、互いに離れており、通信ネットワークを介してインタラクションを行うことが一般的である。対応するコンピュータで動作することで、クライアント-サーバーの関係を有するコンピュータプログラムによってクライアントとサーバーの関係を生み出す。サーバーは、クラウドサーバー、分散型システムのサーバー、ブロックチェーンを組み込んだサーバーなどであってもよい。上記に示された様々な態様のフローを使用して、ステップを新たにソート、追加、又は削除することが可能であることを理解すべきである。例えば、本開示で記載された各ステップは、並列に実行しても良いし、順次に実行しても良いし、異なる順序で実行しても良い。本開示で開示された技術方案が所望する結果を実現

50

することができる限り、本開示ではこれに限定されない。

【 0 1 0 9 】

上記具体的な実施例は、本開示の保護範囲に対する限定を構成するものではない。当業者は、設計事項やその他の要因によって、様々な修正、組み合わせ、サブ組み合わせ、及び代替が可能であることを理解するべきである。本開示の要旨及び原理原則内における変更、均等な置換及び改善等は、いずれも本開示の保護範囲に含まれるべきである。

10

20

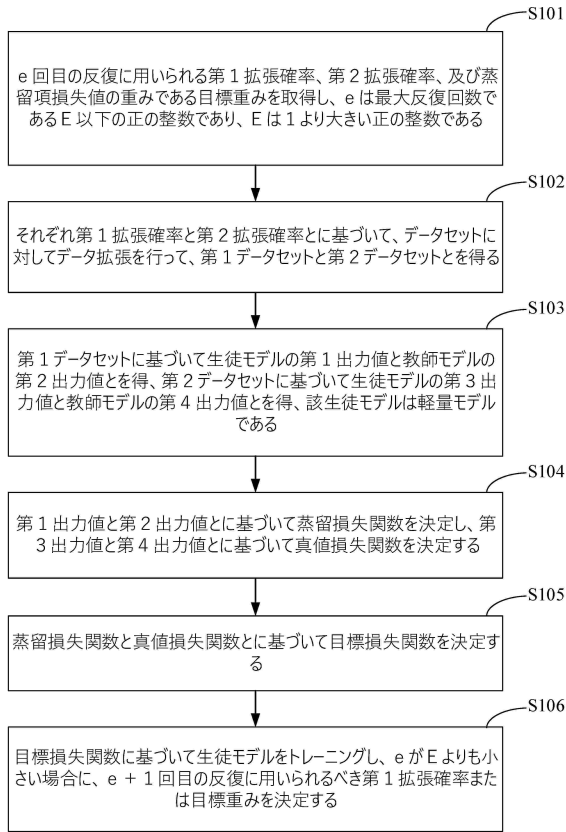
30

40

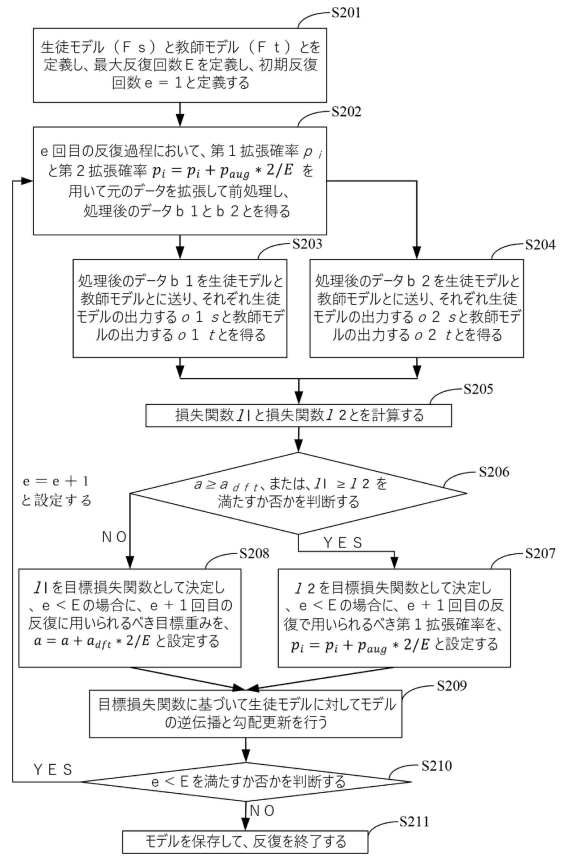
50

【 図 面 】

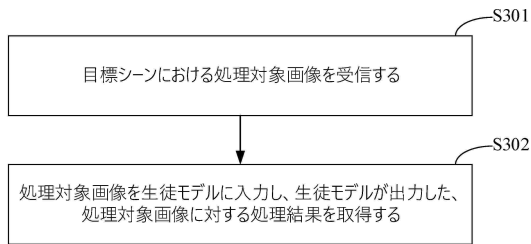
【 図 1 】



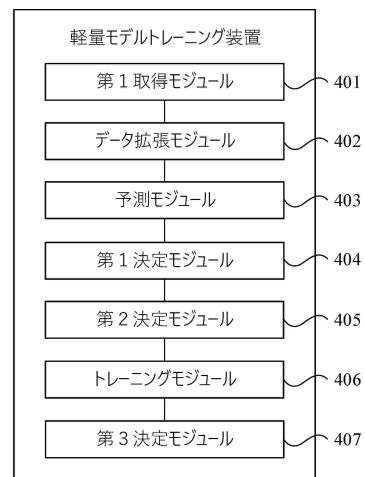
【 図 2 】



【 図 3 】



【 図 4 】



10

20

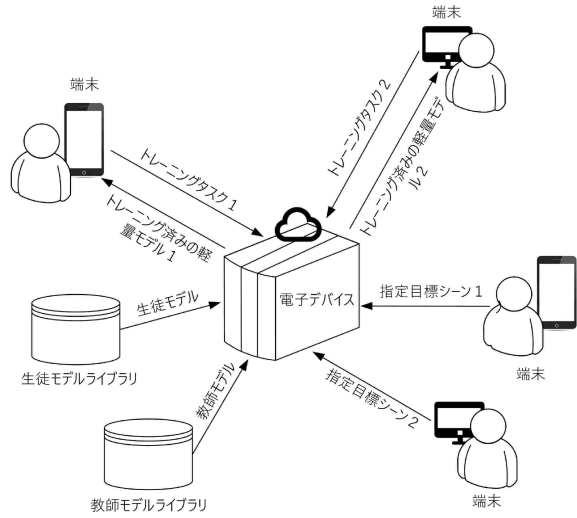
30

40

【図5】

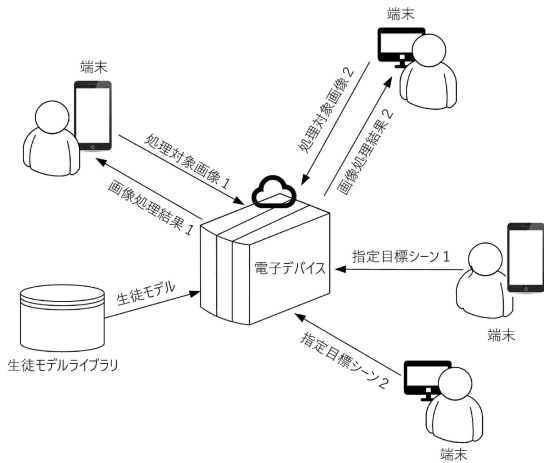


【図6】



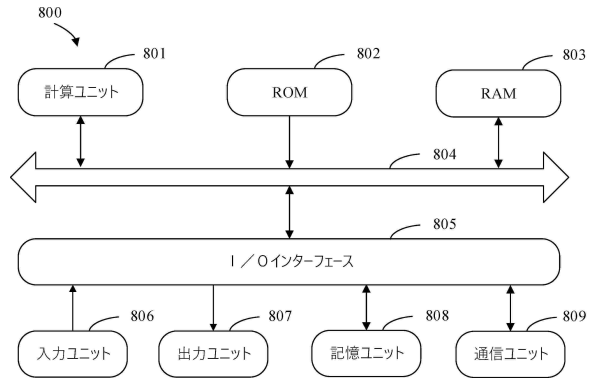
10

【図7】



30

【図8】



20

40

50

フロントページの続き

- 85, China
- (74)代理人 100079108
弁理士 稲葉 良幸
- (74)代理人 100109346
弁理士 大貫 敏史
- (74)代理人 100117189
弁理士 江口 昭彦
- (74)代理人 100134120
弁理士 内藤 和彦
- (74)代理人 100108213
弁理士 阿部 豊隆
- (72)発明者 グオ, ルオユ
中華人民共和国, ベイジン ハイディアן ディストリクト シャンディ テンス ストリート ナン
バー 10 バイドゥ キャンパス 2階
- (72)発明者 ドウ, ユニン
中華人民共和国, ベイジン ハイディアן ディストリクト シャンディ テンス ストリート ナン
バー 10 バイドゥ キャンパス 2階
- (72)発明者 リ, チェンシア
中華人民共和国, ベイジン ハイディアן ディストリクト シャンディ テンス ストリート ナン
バー 10 バイドゥ キャンパス 2階
- (72)発明者 ライ, バオファ
中華人民共和国, ベイジン ハイディアן ディストリクト シャンディ テンス ストリート ナン
バー 10 バイドゥ キャンパス 2階
- (72)発明者 マ, ヤンジュン
中華人民共和国, ベイジン ハイディアן ディストリクト シャンディ テンス ストリート ナン
バー 10 バイドゥ キャンパス 2階
- 審査官 北川 純次
- (56)参考文献 国際公開第2021/044591(WO, A1)
国際公開第2022/132967(WO, A1)
中国特許出願公開第112990342(CN, A)
中国特許出願公開第113592023(CN, A)
中国特許出願公開第114863248(CN, A)
- (58)調査した分野 (Int.Cl., DB名)
G06N 3/02 - 3/10
G06N 20/00
G06T 7/00
G06V 10/82