

(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5月12日(12.05.2022)



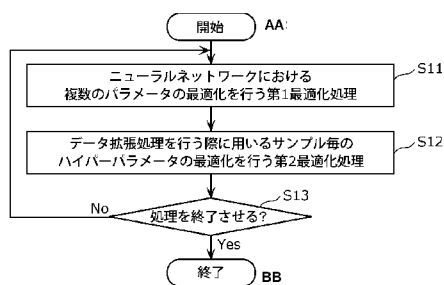
(10) 国際公開番号

WO 2022/097709 A1

- (51) 国際特許分類:
G06N 3/08 (2006.01)
- (21) 国際出願番号: PCT/JP2021/040736
- (22) 国際出願日: 2021年11月5日(05.11.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
63/110,570 2020年11月6日(06.11.2020) US
- (71) 出願人: パナソニック インテレクチュアル
プロパティ コーポレーション オブ アメ
リカ(PANASONIC INTELLECTUAL PROPER-
TY CORPORATION OF AMERICA) [US/US];
90504 カリフォルニア州, トーランス, ス
イト 4 5 0, ウェスト 1 9 0 ストリ
ート 2 0 5 0 California (US).
- (72) 発明者: 石坂 隼(ISHIZAKA, Shun); 〒5718501
大阪府門真市大字門真1006番地 パナソニッ
ク株式会社内 Osaka (JP). 小塚 和紀(KOZUKA,
Kazuki). 築澤 宗太郎(TSUKIZAWA, Sotaro).
グドフスキー デニス(GUDOVSKIY, Denis).
- (74) 代理人: 新居 広守, 外 (NII, Hiromori et al.);
〒5320011 大阪府大阪市淀川区西中島5丁目
3番10号タナカ・イトーピア新大阪ビル6
階新居国際特許事務所内 Osaka (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,
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.

(54) Title: DATA AUGMENTATION METHOD, LEARNING DEVICE, AND PROGRAM

(54) 発明の名称: データ拡張方法、学習装置およびプログラム



S11 First optimization processing to optimize plurality of parameters in neural network
S12 Second optimization processing to optimize hyperparameters for each sample used when performing data augmentation processing
S13 Terminate processing?
AA: Start
BB End

(57) Abstract: This data augmentation method includes alternately carrying out first optimization processing (S11) to optimize a plurality of parameters of a DNN, and second optimization processing (S12) to optimize a hyperparameter for each sample for data augmentation processing. In the first optimization processing: the DNN is caused to predict a first augmentation label from a first augmentation sample obtained by subjecting a first sample included in a training dataset to data augmentation processing; a first error function for the first augmentation label and a first correct answer label of the first sample is calculated, and a plurality of parameters are updated on the basis of the first error function. In the second optimization processing: a second sample is acquired from an evaluation dataset having a distribution similar to that of a test dataset; the DNN of which the plurality of parameters have been updated is caused to predict a second label from the second sample; a second error function of the second label and a second correct answer label of the second sample is calculated; and the hyperparameter is updated on the basis of a gradient obtained by differentiating the second error function with respect to the hyperparameter.

WO 2022/097709 A1

(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))

(57) 要約 : DNNの複数のパラメータの最適化を行う第1最適化処理 (S 1 1) と、データ拡張処理のためのサンプル毎のハイパーパラメータの最適化を行う第2最適化処理 (S 1 2) とを交互に行う。第1最適化処理では、学習用データセットに含まれる第1サンプルにデータ拡張処理を行った第1拡張サンプルから、第1拡張ラベルをDNNに予測させ、第1拡張ラベルと、第1サンプルの第1正解ラベルとについての第1誤差関数を算出し、第1誤差関数に基づき、複数のパラメータを更新する。第2最適化処理では、テスト用データセットの分布に類似する評価用データセットから、第2サンプルを取得し、複数のパラメータが更新されたDNNに、第2サンプルから第2ラベルを予測させ、第2ラベルと、第2サンプルの第2正解ラベルとの第2誤差関数を算出し、第2誤差関数をハイパーパラメータについて微分して得た勾配に基づいて、ハイパーパラメータを更新する。

明 細 書

発明の名称：データ拡張方法、学習装置およびプログラム

技術分野

[0001] 本開示は、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法、学習装置およびプログラムに関する。

背景技術

[0002] 画像認識を応用したシステムの開発において、高精度なAIモデルを獲得するために、学習用の画像データを多く集める必要がある。

[0003] 人物など汎用的な対象については画像データなどのサンプルを大規模に集めやすい一方で、ベビーカーなど汎用的でない対象についてサンプルを大規模には集められない。また、特定の製品、異常サンプルなどそもそも少量しか存在しない対象もサンプルを大規模に集められない。

[0004] これに対して、学習用のサンプル数が少ない場合でも、サンプル数を増加させることができるデータ拡張という手法がある。データ拡張は、画像データなどの学習用のサンプルに対して変換処理を行うことでサンプルを増やす手法であるが、種々のデータ拡張が提案されている（例えば非特許文献1）。

[0005] 非特許文献1には、画像の一部を別の値で塗りつぶしてデータ拡張する技術が開示されている。

先行技術文献

非特許文献

[0006] 非特許文献1：Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. AutoAugment: Learning augmentation policies from data. arXiv:1805.09501, 2018.

非特許文献2：Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast AutoAugment. In Advances in Neural Information Processing Systems, 2019.

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、非特許文献1で開示された技術では、学習用のサンプル数を増やすことができるものの、高精度なAIモデルを獲得できない場合がある。換言すると、学習用のサンプルに対して適切に変換処理を行わないと、学習による識別面（識別線）の生成が理想的な境界線から大きく外れてしまい、予測精度の向上に寄与しないばかりか予測精度を低下させてしまう場合も生じる。

[0008] 本開示は、上述の事情を鑑みてなされたもので、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができるデータ拡張方法等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

[0009] 上記課題を解決するために、本開示の一形態に係るデータ拡張方法は、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法であって、前記ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、前記第1最適化処理では、前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、前記第2最適化処理では、前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、前記複数のパラメータが更新された前記ニューラルネットワークに、前記第2サンプルから、第2ラ

ベルを予測させ、前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新する。

[0010] なお、これらの全般的または具体的な態様は、システム、方法、集積回路、コンピュータプログラムまたはコンピュータで読み取り可能なCD-ROMなどの記録媒体で実現されてもよく、システム、方法、集積回路、コンピュータプログラムおよび記録媒体の任意な組み合わせで実現されてもよい。

発明の効果

[0011] 本開示のデータ拡張方法等によれば、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

図面の簡単な説明

[0012] [図1]図1は、実施の形態に係る学習装置の構成の一例を示すブロック図である。

[図2]図2は、実施の形態に係る学習装置の機能をソフトウェアにより実現するコンピュータのハードウェア構成の一例を示す図である。

[図3]図3は、実施の形態に係る学習装置の動作概要を示すフローチャートである。

[図4]図4は、図3に示す第1最適化処理の詳細動作を示すフローチャートである。

[図5]図5は、図3に示す第2最適化処理の詳細動作を示すフローチャートである。

[図6]図6は、実施例1に係る第1最適化処理及び第2最適化処理を交互に行う処理例を概念的に説明するための図である。

[図7A]図7Aは、実施例1に係る評価用データセットの一例を示す図である。

[図7B]図7Bは、実施例1に係る学習用データセットの一例を示す図である。

。

[図8A]図8Aは、実施例1に係る第1最適化処理を行うための機能例を概念的に説明するための図である。

[図8B]図8Bは、図8Aに示すLOSS機能の具体的な機能を説明するための図である。

[図9]図9は、実施例1に係る第1最適化処理及び第2最適化処理を交互に行う最適化処理手順の一例を示すアルゴリズム1の擬似コードを示す図である

。

[図10]図10は、比較例に係るすべてのデータ間で共有するハイパーパラメータを用いてデータ拡張を行った場合の問題点を説明するための図である。

[図11]図11は、学習処理による識別面の生成を概念的に説明するための図である。

[図12A]図12Aは、データ拡張処理を概念的に説明するための図である。

[図12B]図12Bは、学習用データセットに含まれるデータ数が少ない場合における学習処理による識別面の生成の問題点を概念的に説明するための図である。

[図12C]図12Cは、データ拡張処理によりデータ拡張された学習用データセットを用いた学習処理による識別面の生成の一例を概念的に説明するための図である。

[図13]図13は、比較例に係るデータ拡張処理における問題点を説明するための図である。

[図14]図14は、本開示に係るデータ拡張処理によりデータ拡張された学習用データセットを用いて生成した識別面を概念的に説明するための図である

。

[図15]図15は、実験例に係る学習用データセットを用いて学習後に、テスト用データセットでエラー率を評価した検証結果を示す図である。

[図16]図16は、実験例に係る学習用データセットを用いて学習させたときの精度学習曲線を示す図である。

発明を実施するための形態

[0013] 本開示の一形態に係るデータ拡張方法は、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法であって、前記ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、前記第1最適化処理では、前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、前記第2最適化処理では、前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、前記複数のパラメータが更新された前記ニューラルネットワークに、前記第2サンプルから、第2ラベルを予測させ、前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新する。

[0014] これによれば、ハイパーパラメータを最適化できるので、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0015] より具体的には、ニューラルネットワークの複数のパラメータの最適化を行うニューラルネットワークの学習の過程で、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化することができる。

よって、データ拡張処理によりデータ拡張された学習用データを用いて、ニューラルネットワークを学習させても、学習による識別面の生成を理想的な識別面に近づけることができるように、学習用データセットのサンプル数を増やすことができる。このようにして、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0016] さらに、サンプル毎に、ニューラルネットワークの複数のパラメータとデータ拡張処理のハイパーパラメータを交互に最適化を行うことで、サンプル毎のハイパーパラメータの最適化をニューラルネットワークの複数のパラメータの学習とともに行える。よって、ニューラルネットワークの複数のパラメータの最適化とサンプル毎のハイパーパラメータの最適化に要する計算コストを抑制できる。

[0017] ここで、例えば、前記ハイパーパラメータを更新する際、前記学習用データセットに前記データ拡張処理を行うことで得られるデータ拡張後の学習用データセットの分布と、前記評価用データセットの分布とのずれを減らすよう、前記勾配に基づいて前記ハイパーパラメータを更新する。

[0018] これにより、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化するように、ハイパーパラメータを更新できる。よって、データ拡張処理によりデータ拡張された学習用データを用いて、ニューラルネットワークを学習させても、学習による識別面の生成を理想的な識別面に近づけることができる。

[0019] また、例えば、前記ハイパーパラメータは、前記データ拡張処理を示す関数において陰関数化され、前記ニューラルネットワークは、完全微分可能な関数であるとしてもよい。

[0020] これにより、陰関数の微分手法を用いることで、ハイパーパラメータについての偏微分を計算することができる。

[0021] また、例えば、前記第1誤差関数を算出する際、前記第1正解ラベルが示す正解値をソフトラベル化して得た第1正解ソフトラベルと、前記第1拡張ラベルとの誤差をカルバックライブラーダイバージェンスを用いて評価する

前記第1誤差関数を算出してもよい。

[0022] これにより、データ拡張後に含まれるサンプルに対する正解ラベルのノイズを抑制することができる。よって、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化するように、サンプル毎のハイパーパラメータを最適化できる。

[0023] また、例えば、前記第1誤差関数を算出する際、前記第1誤差関数に対して、さらに、サンプル毎に算出される重みづけを行ってもよい。

[0024] これにより、サンプルのバイアスを捉えて抑制することができる。よって、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化するように、サンプル毎のハイパーパラメータを最適化できる。

[0025] また、本開示の一形態に係る学習装置は、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法を行うための学習装置であって、プロセッサと、メモリとを備え、前記プロセッサは、前記メモリを用いて、前記ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、前記第1最適化処理では、前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、前記第2最適化処理では、前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、前記複数のパラメータが更新さ

れた前記ニューラルネットワークに、前記第2サンプルから、第2ラベルを予測させ、前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新する。

[0026] なお、これらの包括的または具体的な態様は、システム、方法、集積回路、コンピュータプログラムまたはコンピュータで読み取り可能なCD-ROMなどの記録媒体で実現されてもよく、システム、方法、集積回路、コンピュータプログラムおよび記録媒体の任意な組み合わせで実現されてもよい。

[0027] 以下で説明する実施の形態は、いずれも本開示の一具体例を示すものである。以下の実施の形態で示される数値、形状、構成要素、ステップ、ステップの順序などは、一例であり、本開示を限定する主旨ではない。また、以下の実施の形態における構成要素のうち、最上位概念を示す独立請求項に記載されていない構成要素については、任意の構成要素として説明される。また全ての実施の形態において、各々の内容を組み合わせることもできる。

[0028] (実施の形態)

以下では、図面を参照しながら、実施の形態に係る学習装置10の情報処理方法等の説明を行う。

[0029] [1.1 学習装置10の構成]

図1は、実施の形態に係る学習装置10の構成の一例を示すブロック図である。

[0030] 学習装置10は、ニューラルネットワークの学習を行うとともに、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法を行うための装置である。学習装置10は、ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行う。

[0031] 本実施の形態では、学習装置 10 は、図 1 に示すように取得部 11 と、データ拡張処理部 12 と、DNN 13 と、誤差算出部 14 と、最適化部 15 とで構成されている。

[0032] [1. 1. 1 取得部 11]

取得部 11 は、第 1 最適化処理において、学習用データセットから、第 1 サンプルを取得し、データ拡張処理部 12 に入力する。ここで、学習用データセットは、例えば画像データなどの複数のサンプルとそれらの正解ラベルとからなる。なお、サンプルは、画像データである場合に限らない。

[0033] また、取得部 11 は、第 1 最適化処理において、学習用データセットから、第 1 サンプルの正解を示す第 1 正解ラベルを取得し、誤差算出部 14 に入力する。

[0034] また、取得部 11 は、第 2 最適化処理において、評価用データセットから、第 2 サンプルを取得し、DNN 13 に入力する。ここで、評価用データセットは、学習済の DNN 13 の性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである。なお、評価用データセットと、テスト用データセットとは異なる。評価用データセットは、既存のデータセットのテスト用データセットを分割することで生成することができる。

[0035] また、取得部 11 は、第 2 最適化処理において、評価用データセットから、第 2 サンプルの正解を示す第 2 正解ラベルを取得し、誤差算出部 14 に入力する。

[0036] なお、第 2 最適化処理において、取得部 11 は、学習用データセットから、第 3 サンプルを取得し、データ拡張処理部 12 に入力してもよい。この場合、取得部 11 は、学習用データセットから、第 3 サンプルの正解を示す第 3 正解ラベルを取得し、誤差算出部 14 に入力すればよい。第 3 サンプルは、第 1 サンプルと異なってもよいし同じであってもよい。

[0037] [1. 1. 2 データ拡張処理部 12]

データ拡張処理部 12 は、第 1 最適化処理において、学習用データセットに含まれる第 1 サンプルに、データ拡張処理を行い、第 1 拡張サンプルを出

力する。より具体的には、第1最適化処理において、データ拡張処理部12は、サンプル毎のデータ変換処理を規定する変数であるハイパーパラメータを用いて、第1サンプルに、データ拡張処理を行う。ハイパーパラメータは、例えば回転処理、ズーム処理、並進処理、色変換処理といったデータの変換処理を規定する変数であり、サンプル毎に規定される。

[0038] 本実施の形態では、データ拡張処理部12は、データ拡張処理を示す関数を、学習用データセットから取得されたサンプルに作用させることで、データ拡張処理を行う。ここで、ハイパーパラメータは、データ拡張処理を示す関数において陰関数化されている。

[0039] ここで、データ拡張処理を示す関数を g とし、第1サンプルに対するハイパーパラメータを λ_1 とする。この場合、データ拡張処理部12は、第1最適化処理において、 $g(\lambda_1)$ を第1サンプルに作用させることで、第1サンプルのデータ拡張処理を行うことができ、第1拡張サンプルを出力することができる。したがって、データ拡張処理部12は、第1最適化処理において、第 i (i は自然数)サンプルに対するハイパーパラメータを λ_i とする場合、 $g(\lambda_i)$ を第 i サンプルに作用させることで、第 i サンプルのデータ拡張処理を行うことができる。

[0040] なお、第2最適化処理において、データ拡張処理部12は、取得部11により第3サンプルが入力される場合、第3サンプルに、データ拡張処理を行い、第3拡張サンプルを出力すればよい。この場合、データ拡張処理部12は、第2最適化処理において、第3サンプルに対するハイパーパラメータを λ_3 とする $g(\lambda_3)$ を第3サンプルに作用させることで、第3サンプルのデータ拡張処理を行うことができるので、第3拡張サンプルを出力することができる。

[0041] [1. 1. 3 DNN13]

DNN13は、CNNなど、複数のパラメータを有するニューラルネットワークにより構成される。DNN13は、第1最適化処理において、取得部11により入力された第1拡張サンプルから、第1拡張ラベルを予測する。

複数のパラメータは、例えば重みであるが、重みに限らない。

[0042] 本実施の形態では、DNN13は、第1最適化処理において、ニューラルネットワークによる予測処理を示す関数を、第1拡張サンプルに作用させることで、第1拡張サンプルの正解としての第1拡張ラベルを予測する予測処理を行う。ここで、ニューラルネットワークによる予測処理を示す関数は、完全微分可能な関数である。また、複数のパラメータは、ニューラルネットワークによる予測処理を示す関数において陰関数化されている。

[0043] ここで、ニューラルネットワークによる予測処理を示す関数を f とし、重みを含む複数のパラメータを θ とすると、DNN13は、第1最適化処理において、 $f(\theta)$ を第1拡張サンプルに作用させることで、第1拡張ラベルを予測することができる。

[0044] また、DNN13は、第1最適化処理により複数のパラメータが更新された後、取得部11により入力された、評価用データセットに含まれる第2サンプルから、第2ラベルを予測する。換言すると、本実施の形態では、第2最適化処理において、DNN13は、第1最適化処理により複数のパラメータが更新された $f(\theta)$ を第2サンプルに作用させることで、第2ラベルを予測する。

[0045] なお、DNN13は、第2最適化処理において、データ拡張処理部12により第3拡張サンプルが入力された場合、第3拡張サンプルから、第3拡張サンプルの正解としての第3拡張ラベルを予測すればよい。

[0046] [1.1.4 誤差算出部14]

誤差算出部14は、第1最適化処理において、第1拡張ラベルと、学習用データセットに含まれる第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出する。ここでの第1誤差関数は、カルバック・ライブラー・ダイバージェンス（KLダイバージェンス）というロス関数である。第1誤差関数は、クロスエントロピー誤差関数であってもよい。

[0047] なお、カルバック・ライブラー・ダイバージェンスは、確率論と情報理論

とにおける2つの確率分布の差異を計る尺度として用いることができる。つまり、カルバック・ライブラー・ダイバージェンスは、正解分布である確率分布と推定分布である確率分布とがどのくらい似ているかを評価できるロス関数である。

[0048] また、誤差算出部14は、第1最適化処理において、第1正解ラベルが示す正解値をソフトラベル化して得た第1正解ソフトラベルと、第1拡張ラベルとの誤差をKLダイバージェンスを用いて評価する第1誤差関数を算出してもよい。ここで、ソフトラベルは、第1拡張サンプルと正解ラベルの類似度を離散値（ハードラベル）ではなく、割合で表すことを意味する。例えば、ソフトラベルとしては、ガンベルソフトマックス（gumbel-softmax）またはソフトマックス（softmax）を用いることができる。これにより、データ拡張後に含まれるサンプルに対する正解ラベルのノイズを抑制することができる。

[0049] さらに、第1最適化処理では、誤差算出部14は、算出した第1誤差関数に対して、さらに、サンプル毎に算出される重みづけを行ってもよい。

[0050] また、誤差算出部14は、第2最適化処理において、第2ラベルと、評価用データセットに含まれる第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出する。ここで、第2誤差関数は、第1誤差関数と同様に、KLダイバージェンスというロス関数であってもよいし、クロスエントロピー誤差関数であってもよい。

[0051] なお、第2最適化処理において、DNN13が、第3拡張サンプルから第3拡張ラベルを予測する場合、第1誤差関数を用いて、第3拡張ラベルと、学習用データセットに含まれる第3サンプルの正解を示す第3正解ラベルとについての誤差を評価すればよい。

[0052] [1. 1. 5 最適化部15]

最適化部15は、第1最適化処理において、誤差算出部14により算出された第1誤差関数に基づき、複数のパラメータを更新する。これにより、最適化部15は、ニューラルネットワークにおける複数のパラメータの最適化

を行う第1最適化処理を行うことができる。

[0053] また、最適化部15は、第2最適化処理において、誤差算出部14により算出された第2誤差関数を、ハイパーパラメータについて偏微分して得た勾配に基づいて、ハイパーパラメータを更新する。なお、ハイパーパラメータは、陰関数化されているので、陰関数の微分手法を用いることで、ハイパーパラメータについての偏微分を計算することができる。ここで、最適化部15は、学習用データセットにデータ拡張処理を行うことで得られるデータ拡張後の学習用データセットの分布と、評価用データセットの分布とのずれを減らすよう、勾配に基づいてハイパーパラメータを更新する。これにより、最適化部15は、第2最適化処理において、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うことができる。

[0054] 以上のように構成された学習装置10は、ニューラルネットワークにおける複数のパラメータに対する第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のハイパーパラメータに対する第2最適化処理とを交互に行う。これにより、データ拡張処理がされた学習用データセットの分布を、テスト用データセットの分布に合わせることができるので、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0055] より具体的には、上記のようにしてデータ拡張処理がされたサンプル数が増えた学習用データを用いて、ニューラルネットワークを学習させても、学習による識別面の生成を理想的な識別面に近づけることができるように、学習用データセットのサンプル数を増やすことができる。このようにして、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0056] [1.2 学習装置10のハードウェア構成]

図2は、実施の形態に係る学習装置10の機能をソフトウェアにより実現するコンピュータ1000のハードウェア構成の一例を示す図である。

- [0057] コンピュータ1000は、図2に示すように、入力装置1001、出力装置1002、CPU1003、内蔵ストレージ1004、RAM1005、読取装置1007、送受信装置1008およびバス1009を備えるコンピュータである。入力装置1001、出力装置1002、CPU1003、内蔵ストレージ1004、RAM1005、読取装置1007および送受信装置1008は、バス1009により接続される。
- [0058] 入力装置1001は、入力ボタン、タッチパッド、タッチパネルディスプレイなどといったユーザインタフェースとなる装置であり、ユーザの操作を受け付ける。なお、入力装置1001は、ユーザの接触操作を受け付ける他、音声での操作、リモコン等での遠隔操作を受け付ける構成であってもよい。
- [0059] 出力装置1002は、入力装置1001と兼用されており、タッチパッドまたはタッチパネルディスプレイなどによって構成され、ユーザに知らせべき情報を通知する。
- [0060] 内蔵ストレージ1004は、フラッシュメモリなどである。また、内蔵ストレージ1004は、学習装置10の機能を実現するためのプログラム、複数のパラメータを有するニューラルネットワーク、ニューラルネットワークによる予測処理を示す関数 f 及びデータ拡張処理を示す関数 g などが予め記憶されていてもよい。
- [0061] RAM1005は、ランダムアクセスメモリ (Random Access Memory) であり、プログラム又はアプリケーションの実行に際してデータ等の記憶に利用される。
- [0062] 読取装置1007は、USB (Universal Serial Bus) メモリなどの記録媒体から情報を読み取る。読取装置1007は、上記のようなプログラムやアプリケーションが記録された記録媒体からそのプログラム、アプリケーションを読み取り、内蔵ストレージ1004に記憶させる。
- [0063] 送受信装置1008は、無線又は有線で通信を行うための通信回路である。送受信装置1008は、例えばネットワークに接続されたサーバ装置また

はクラウドと通信を行い、サーバ装置またはクラウドから上記のようなプログラム、アプリケーションをダウンロードして内蔵ストレージ1004に記憶させてもよい。

[0064] CPU1003は、中央演算処理装置 (Central Processing Unit) であり、内蔵ストレージ1004に記憶されたプログラム、アプリケーションをRAM1005にコピーし、そのプログラムやアプリケーションに含まれる命令をRAM1005から順次読み出して実行する。なお、内蔵ストレージ1004から直接実行しても良い。

[0065] [1.3 学習装置10の動作]

次に、上記のように構成された学習装置10の動作について説明する。

[0066] 図3は、実施の形態に係る学習装置10の動作概要を示すフローチャートである。

[0067] まず、学習装置10は、ニューラルネットワークにおける複数のパラメータの最適化を行う第1最適化処理を行う (S11)。次に、学習装置10は、データ拡張処理を行う際に用いるサンプル毎のハイパーパラメータの最適化を行う第2最適化処理を行う (S12)。次に、学習装置10は、処理を終了させるかを判定する (S13)。学習装置10は、例えば既定の回数 (エポック回数) だけ第1最適化処理及び第2最適化処理を行った場合、処理を終了させると判定し (S13でYes)、本動作を終了させる。一方、学習装置10は、例えば既定の回数 (エポック回数) まで第1最適化処理及び第2最適化処理を行っていない場合、処理を終了させないと判定し (S13でNo)、ステップS11の処理から繰り返す。

[0068] 図4は、図3に示す第1最適化処理の詳細動作を示すフローチャートである。

[0069] 第1最適化処理において、まず、学習装置10は、学習用データセットに含まれる第1サンプルに、データ拡張処理を行い、第1拡張サンプルを取得する (S111)。ここで、学習装置10は、サンプル毎のデータ変換処理を規定する変数であるハイパーパラメータを用いて、第1サンプルに、デー

タ拡張処理を行う。例えば、データ拡張処理を示す関数を g とし、第1サンプルに対するハイパーパラメータを λ_1 とすると、学習装置10は、 $g(\lambda_1)$ を第1サンプルに作用させることで、第1サンプルに対してデータ拡張処理を行った第1拡張サンプルを取得する。

[0070] 次に、学習装置10は、ニューラルネットワークに、ステップS111で取得した第1拡張サンプルから、第1拡張ラベルを予測させる(S112)。例えば、ニューラルネットワークによる予測処理を示す関数を f とし、重みを含む複数のパラメータを θ とすると、学習装置10は、 $f(\theta)$ を第1拡張サンプルに作用させることで、ニューラルネットワークに、第1拡張ラベルを予測させる。

[0071] 次に、学習装置10は、ステップS112で予測させた第1拡張ラベルと、学習用データセットに含まれる第1サンプルの第1正解ラベルとについての誤差を評価するための第1誤差関数を算出する(S113)。ここで、第1誤差関数は、例えばKLダイバージェンスというロス関数である。なお、上述したように、学習装置10は、第1正解ラベルが示す正解値をソフトラベル化して得た第1正解ソフトラベルと、第1拡張ラベルとの誤差をKLダイバージェンスを用いて評価する第1誤差関数を算出してもよい。学習装置10は、さらに、算出した第1誤差関数に対して、さらに、サンプル毎に算出される重みづけを行ってもよい。

[0072] 次に、学習装置10は、ステップS113で算出した第1誤差関数に基づき、複数のパラメータを更新する(S114)。

[0073] このようにして、学習装置10は、初期または第2最適処理後のデータ拡張処理により水増しされた学習用データセットを用いて、ニューラルネットワークにおける重みなどの複数のパラメータの最適化を行う第1最適化処理を行うことができる。

[0074] 図5は、図3に示す第2最適化処理の詳細動作を示すフローチャートである。

[0075] 第2最適化処理において、まず、学習装置10は、テスト用データセット

の分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得する（S 1 2 1）。

[0076] 次に、学習装置10は、第1最適化処理において複数のパラメータが更新されたニューラルネットワークに、第2サンプルから、第2ラベルを予測させる（S 1 2 2）。例えば、ニューラルネットワークによる予測処理を示す関数を f とし、重みを含む複数のパラメータを θ とすると、学習装置10は、 $f(\theta)$ を第2サンプルに作用させることで、ニューラルネットワークに、第2ラベルを予測させる。

[0077] 次に、学習装置10は、ステップS 1 2 2で予測させた第2ラベルと、評価用データセットに含まれる第2サンプルの第2正解ラベルとの誤差を評価するための第2誤差関数を算出する（S 1 2 3）。ここで、第2誤差関数は、例えばKLダイバージェンスというロス関数であるが、クロスエントロピー誤差関数であってもよい。

[0078] 次に、学習装置10は、ステップS 1 2 3で算出した第2誤差関数を、ハイパーパラメータについて偏微分して得た勾配に基づいて、ハイパーパラメータを更新する（S 1 2 4）。ここで、学習装置10は、学習用データセットにデータ拡張処理を行うことで得られるデータ拡張後の学習用データセットの分布と、評価用データセットの分布とのずれを減らすよう、勾配に基づいてハイパーパラメータを更新する。

[0079] このようにして、学習装置10は、評価用データセットを用いて、データ拡張処理を行う際に用いるサンプル毎のハイパーパラメータの最適化を行う第2最適化処理を行うことができる。

[0080] （実施例1）

以下、本実施の形態の実施例1について説明する。

[0081] 図6は、実施例1に係る第1最適化処理及び第2最適化処理を交互に行う処理例を概念的に説明するための図である。

[0082] Process 1 2 aは、データ拡張処理部12が行うデータ拡張処理機能を示し、 $g(\lambda_i)$ は、当該データ拡張処理を示す関数であってハイパーパラメータ

λ_i が陰関数化された関数を示している。図6に示す例では、第1最適化処理において、Process 12 aは、学習用データセットからサンプル (x_i, y_i) を取得し、サンプル (x_i) に対して、ハイパーパラメータ λ_i でデータ拡張処理を行う。そして、Process 12 aは、データ拡張処理により得た拡張サンプル $g(x_i, y_i, \lambda_i)$ をDNN 13 aに出力する。

[0083] DNN 13 aは、DNN 13の複数のパラメータ θ を有するニューラルネットワークを示し、 $f(\theta)$ は、当該ニューラルネットワークによる予測処理を示す関数を示している。図6に示す例では、第1最適化処理において、Process 12 aから出力された拡張サンプル $g(x_i, y_i, \lambda_i)$ を用いて、DNN 13 aのニューラルネットワークが学習されることが示されている。

[0084] また、図6に示す例では、第2最適化処理において、評価用データセットからサンプル (x_i^v, y_i^v) を取得し、サンプル (x_i^v) をDNN 13 aに入力する。DNN 13 aは、サンプル (x_i^v) に対して、関数 $f(\theta)$ を作用させて、予測処理を行う。そして、DNN 13 aは、予測処理により得たサンプル (x_i^v) に対する正解ラベルを出力する。

[0085] AutoD015 aは、最適化部15が行う第2最適化処理機能を示し、 $\partial L_v / \partial \lambda$ は、DNN 13 aの予測処理の結果に基づき算出した誤差関数 L_v のハイパーパラメータ λ について微分（偏微分）して得た勾配を示す。より具体的には、第2最適化処理では、予測処理により得たサンプル (x_i^v) に対する正解ラベルと評価用データセットから取得したサンプル (x_i^v) の正解ラベル (y_i^v) についての誤差関数 L_v を算出する。そして、AutoD015 aは、算出した誤差関数 L_v を、 λ_i について偏微分して得た勾配に基づいて、 λ_i を更新することで、ハイパーパラメータ λ を最適化する。

[0086] 図7 Aは、実施例1に係る評価用データセットの一例を示す図である。図7 Bは、実施例1に係る学習用データセットの一例を示す図である。

[0087] 図7 Aに示す評価用データセットのサンプルは、桁数1, 6, 7, 9の画像データである。図7 Aには、学習済のDNN 13 aの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットのサンプルが

例示されている。

- [0088] 一方、図7Bに示す学習用データセットのサンプルは、桁数0、0、1、2、6、9等の画像データである。図7Bには、バイアスが含まれたサンプルと、サンプルに対する正解ラベルにノイズ（ラベルノイズ）が含まれていることが例示されている。
- [0089] 図8Aは、実施例1に係る第1最適化処理を行うための機能例を概念的に説明するための図である。図8Bは、図8Aに示すLOSS14bの具体的な機能を説明するための図である。
- [0090] Data Sampler 11bは、第1最適化処理において取得部11が行う取得処理機能を示し、 $i \sim U(1, N)$ で示される学習用データセットから、サンプル x_i を取得する。
- [0091] Augment 12bは、第1最適化処理においてデータ拡張処理部12が行うデータ拡張処理機能を示し、 $g^A(\lambda^A)$ は、当該データ拡張処理を示す関数であってハイパーパラメータ λ^A が陰関数化された関数を示している。図8Aに示す例では、Augment 12bにおける $g^A(\lambda^A)$ は、第2最適化処理において勾配 $\partial L_v / \partial \lambda$ によりハイパーパラメータ λ^A が更新されている。
- [0092] Augment 12bは、Data Sampler 11bにより取得されたサンプル x_i に対して、 $g^A(\lambda^A)$ を作用させてデータ拡張処理を行う。そして、Data Sampler 11bは、データ拡張処理により得た拡張サンプル x_i^A をDNN 13bに出力する。
- [0093] DNN 13bは、DNN 13の複数のパラメータ θ を有するニューラルネットワークを示し、 $f(\theta)$ は、当該ニューラルネットワークによる予測処理を示す関数を示している。図8Aに示す例では、第1最適化処理において、DNN 13bは、Augment 12bから出力された拡張サンプル x_i^A に対して、 $f(\theta)$ を作用させてデータ予測処理を行う。そして、DNN 13bは、予測処理により得た拡張サンプル x_i^A に対する正解ラベル

[数1]

$$\hat{y}_i$$

を出力する。

- [0094] LOSS 1 4 b は、第 1 最適化処理における誤差関数（第 1 誤差関数）を算出する処理機能を示す。 $g_{s,w}(\lambda^{s,w})$ は、サンプルのバイアスを捉えるためのサンプル毎の重み（損失重み）と、ノイズの多い正解ラベルに対応するためのソフトラベルとを用いた第 1 最適化処理における誤差関数（第 1 誤差関数）を示す。 $\lambda^{s,w}$ は、損失重みとソフトラベルのハイパーパラメータを示し、陰関数化されている。
- [0095] なお、ハイパーパラメータ $\lambda^{s,w}$ は、第 2 最適化処理において勾配 $\partial L_v / \partial \lambda$ により更新されている。
- [0096] より詳細には、LOSS 1 4 b は、図 8 B に示すように、Soft-label 1 4 1 と、Weight KL Div 1 4 2 と、Reweight 1 4 3 との機能を有する。
- [0097] Soft-label 1 4 1 は、例えば、ガンベルソフトマックスまたはソフトマックスを用いて、入力された正解ラベルが示す正解値をソフトラベル化して、出力する。 $g_s(\lambda^s)$ は、ノイズの多い正解ラベルに対応するためのソフトラベル化処理する関数を示し、ハイパーパラメータ λ^s が陰関数化されている。
- [0098] 図 8 B に示す例では、Soft-label 1 4 1 は、 $i \sim U(1, N)$ で示される学習用データセットから、サンプル x_i に対する正解ラベル y_i が入力され、入力された示す正解ラベル y_i をソフトラベル化する。Soft-label 1 4 1 は、ソフトラベル化したソフトラベル y_i^s を Weight KL Div 1 4 2 に出力する。ここで、例えば、Soft-label 1 4 1 は、 $y_i^s = g_s(y_i, \lambda_i^s) = \text{softmax}(\lambda_i^s)$ の式を用いて正解ラベル y_i からサンプル毎のソフトラベル y_i^s を算出する。ここで、 λ_i^s は、サンプル毎のソフトラベルのハイパーパラメータであり、陰関数化されている。
- [0099] Weight KL Div 1 4 2 は、サンプル毎の重み（損失重み）と、ノイズの多い正解ラベルに対応するためのソフトラベルとを用いた第 1 最適化処理における誤差関数（第 1 誤差関数）を算出する。より具体的には、Weight KL Div 1 4 2 は、Soft-label 1 4 1 により出力されたソフトラベル y_i^s と DNN 1 3

bにより出力された正解ラベル

[数2]

$$\hat{y}_i$$

との誤差を評価するKLダイバージェンスである第1誤差関数Lを算出する。そして、Weight KL Div 1 4 2は、算出したReweight 1 4 3により算出された重み w_i （スカラ量）でかけることで、重み付けされた第1誤差関数

[数3]

$$w_i L(y_i^S, \hat{y}_i)$$

を算出する。

[0100] Reweight 1 4 3は、softplusなどの活性化関数を用いて、サンプルのバイアスを捉えるためのサンプル毎の重み w_i を算出する。 $g_w(\lambda^w)$ は、サンプルのバイアスを捉えるための誤差関数Lに重み付けされるサンプル毎の重み（損失重み）を算出する関数を示す。

[0101] 図8Bに示す例では、Reweight 1 4 3は、 $g_w(L_i, \lambda_i^w) = w_i L_i$ の式における重み w_i を算出し、Weight KL Div 1 4 2に出力する。例えばReweight 1 4 3は、 $w_i = 1.44 \times \text{softplus}(\lambda_i^w)$ の式を用いてサンプル毎の重み w_i を算出する。ここで、 λ_i^w は、サンプル毎の重み（損失重み）のハイパーパラメータであり、陰関数化されている。

[0102] 第1最適化処理において、上記のように算出された重み付けされた第1誤差関数

[数4]

$$w_i L(y_i^S, \hat{y}_i)$$

を用いて、DNN 1 3 bの複数のパラメータ θ を更新することができる。

[0103] 続いて、図9を用いて、実施例1に係る第1最適化処理及び第2最適化処理を交互に行う最適化処理手順について説明する。

- [0104] 図9は、実施例1に係る第1最適化処理及び第2最適化処理を交互に行う最適化処理手順の一例を示すアルゴリズム1の擬似コードを示す図である。なお、図9に示される λ などの変数、 $f(\theta)$ などの関数は、図8A及び図8Bで説明した通りである。図9では、上記における学習用データセットのサンプルをデータと称して表現されている。
- [0105] 図9に示すアルゴリズム1は、例えば学習装置10のプロセッサにより行われる。
- [0106] 図9に示すように、1行目の手順では、パラメータ θ とハイパーパラメータ λ との初期化が行われることが規定されている。
- [0107] 2行目では、3行目～19行目の手順すなわち第1最適化処理及び第2最適化処理がエポック回数まで繰り返されることが規定されている。
- [0108] 3行目では、4行目～9行目の手順すなわち第1最適化処理がバッチ個数まで繰り返されることが規定されている。
- [0109] 4行目では、学習用データセット D_{train} からバッチがサンプリングされることが規定されている。なおバッチとは、学習用データセットを幾つかのサブセットに分けたときのサブセットに含まれるデータ（上記でのサンプル）に該当する。
- [0110] 5行目では、サンプリングされたデータ x がデータ拡張処理されることが規定されている。
- [0111] 6行目では、5行目においてデータ拡張処理されたデータ x^A から正解ラベルが予測されることが規定されている。
- [0112] 7行目では、4行目で学習用データセット D_{train} からサンプリングされたデータ x に対する正解ラベルであって学習用データセット D_{train} から取得された正解ラベルのソフトラベルが生成されることが規定されている。
- [0113] 8行目では、重み付き誤差関数がKLダイバージェンスで計算され、パラメータ θ で微分(∇_{θ})されることが規定されている。
- [0114] 9行目では、ニューラルネットワークによる予測処理を示す関数 $f(\theta)$ のパラメータ θ が更新されることが規定されている。

- [0115] 10行目では、エポック回数がE回数を超えていたら、11行目以降の処理を行われることが規定されている。
- [0116] 11行目では、12行目～19行目の手順すなわち第2最適化処理がバッチ個数まで繰り返されることが規定されている。
- [0117] 12行目では、学習用データセット D_{train} からバッチがサンプリングされることが規定されている。
- [0118] 13行目では、評価用データセット D_{val} からバッチがサンプリングされることが規定されている。
- [0119] 14行目では、13行目でサンプリングされた評価用データセット D_{val} のデータ X^v から正解ラベルが予測されることが規定されている。
- [0120] 15行目では、12行目でサンプリングされた学習用データセット D_{train} のデータがデータ拡張処理されたデータ X^A から正解ラベルが予測されることが規定されている。
- [0121] 16行目では、12行目で学習用データセット D_{train} からサンプリングされたデータ x に対する正解ラベルであって学習用データセット D_{train} から取得された正解ラベルのソフトラベルが生成されることが規定されている。
- [0122] 17行目及び18行目では、18行目に示される誤差関数 L_v と誤差関数 L とが計算され、誤差関数 L_v をハイパーパラメータ λ で微分(∇_{λ})されることが規定されている。ここで、誤差関数 L_v は、評価用データセット D_{val} のデータと14行目で予測されたその正解ラベルとの誤差を評価する誤差関数である。誤差関数 L は、16行目において生成されたソフトラベルと15行目において予測されたデータ拡張処理されたデータ X^A の正解ラベルとの誤差を評価する誤差関数に重みが付されたものである。
- [0123] 18行目では、ハイパーパラメータ λ が更新されることが規定されている。なお、19行目に示されるHOは、hyperparameter optimizationの略語である。
- [0124] このように規定されるアルゴリズム1を実行することで、第1最適化処理及び第2最適化処理を交互に行うことができる。

[0125] [1. 4 効果等]

以上のように、本実施の形態に係る学習装置10及びデータ拡張方法によれば、ハイパーパラメータを最適化できるので、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0126] より具体的には、ニューラルネットワークの複数のパラメータの最適化を行うニューラルネットワークの学習の過程で、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化することができる。よって、データ拡張処理によりデータ拡張された学習用データを用いて、ニューラルネットワークを学習させても、学習による識別面の生成を理想的な識別面に近づけることができるように、学習用データセットのサンプル数を増やすことができる。このようにして、高精度なAIモデルを獲得できるように学習用データセットのサンプル数を増やすことができる。

[0127] ところで、従来、データ拡張処理を行うためには学習用データセットの変換方法を定める大規模なハイパーパラメータを調整する必要があった。また、データの分布の偏り及び正解ラベルの不正確さを含む不完全な学習用データセットを用いて、ニューラルネットワークの複数のパラメータを学習させても、高精度なAIモデルを獲得できないという問題もある。したがって、ニューラルネットワークの学習のための計算と、データ拡張処理を行うために学習用データセットの大規模なハイパーパラメータを最適化する計算といった2重の最適化計算が必要になる。このため、多大な計算コストを要しないとハイパーパラメータを最適化できなかった。

[0128] これに対して、本実施の形態に係る学習装置10及びデータ拡張方法によれば、サンプル毎に、ニューラルネットワークの複数のパラメータとデータ拡張処理のハイパーパラメータを交互に最適化を行うことで、サンプル毎のハイパーパラメータの最適化をニューラルネットワークの複数のパラメータの学習とともに行える。よって、ニューラルネットワークの複数のパラメータの最適化とサンプル毎のハイパーパラメータの最適化に要する計算コスト

を抑制できるという効果も奏する。

[0129] ここで、本実施の形態に係る学習装置10及びデータ拡張方法では、学習用データセットにデータ拡張処理を行うことで得られるデータ拡張後の学習用データセットの分布と、評価用データセットの分布とのずれを減らすよう、勾配に基づいてハイパーパラメータを更新する。

[0130] これにより、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化するように、ハイパーパラメータを更新できる。よって、データ拡張処理によりデータ拡張された学習用データを用いて、ニューラルネットワークを学習させても、学習による識別面の生成を理想的な識別面に近づけることができる。

[0131] また、本実施の形態に係る学習装置10及びデータ拡張方法では、ハイパーパラメータは、少なくともデータ拡張処理を示す関数において陰関数化され、ニューラルネットワークは、完全微分可能な関数である。これにより、陰関数の微分手法を用いることで、ハイパーパラメータについての偏微分を計算することができる。

[0132] また、本実施の形態に係る学習装置10及びデータ拡張方法では、ノイズの多い正解ラベルに対応するためのソフトラベルを用いた第1最適化処理における第1誤差関数を算出する。これにより、データ拡張後に含まれるサンプルに対する正解ラベルのノイズを抑制することができる。よって、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のずれを最小化するように、サンプル毎のハイパーパラメータを最適化できる。

[0133] また、本実施の形態に係る学習装置10及びデータ拡張方法では、算出した第1誤差関数に対して、サンプルのバイアスを捉えるためのサンプル毎の重みを用いて重み付けする。

[0134] これにより、サンプルのバイアスを捉えて抑制することができる。よって、サンプル毎のハイパーパラメータについて偏微分して得た勾配を用いて、データ拡張された学習用データセットとテスト用データセットとの分布のず

れを最小化するように、サンプル毎のハイパーパラメータを最適化できる。

[0135] (比較例)

以下、比較例に係るデータ拡張方法とその問題点について説明する。

[0136] 図10は、比較例に係るすべてのデータ間で共有するハイパーパラメータを用いてデータ拡張を行った場合の問題点を説明するための図である。

[0137] 図10では、すべてのデータ間で共有するハイパーパラメータ λ を用いたデータ拡張処理 $g(\lambda)$ でデータ拡張された学習用データセットを用いて、ニューラルネットワーク $f(\theta)$ を学習させたときの識別面(識別線)が示されている。また、図10では、中心の点線四角の識別面より上方領域及び下方領域での一番小さい実線の円は、学習用データの分布を概念的に示している。また、上方領域及び下方領域での点線の円は、データ拡張処理 $g(\lambda)$ でデータ拡張された学習用データセットの分布を概念的に示している。点線四角における大きい実線の円は、テスト用データセットの分布を概念的に示している。

[0138] 図10に示す比較例に係るデータ拡張処理 $g(\lambda)$ でデータ拡張された場合、学習用データセットのすべてのデータすなわち数字の分布が均等に増加される。このようにデータ拡張された学習用データセットを用いて学習されて得た識別面は、テスト用データセットの一部を横切っているため、テスト用データセットを正しく識別できないという問題がある。

[0139] つまり、比較例に係るデータ拡張方法では、学習による識別面の生成が理想的な境界線から大きく外れてしまうという問題がある。なお、この問題は、学習用データセットの分布に応じてデータ拡張されるための生じていると考えられる。

[0140] 以下、比較例に係るデータ拡張の問題について具体的に説明する。

[0141] 図11は、学習処理による識別面の生成を概念的に説明するための図である。図11に示すように、分類問題において、A1は、種々の犬または猫の画像とそれらの正解ラベルからなる学習用データセットを用いて学習処理が行われると、特徴抽出して識別面 d_1 を生成する。識別面 d_1 の生成は、図1

1からわかるように、概念的には犬と猫とを見分ける境界線を引くことに該当する。

[0142] 図12Aは、データ拡張処理を概念的に説明するための図である。

[0143] 図12Aに示すように、学習用データセットのデータ拡張処理は、学習用データセットのデータに対して回転処理、ズーム処理、並進処理、色変換処理などの変換処理を行うことで、学習用データセットを構成するデータを擬似的に増加、水増しさせる処理である。

[0144] 図12Bは、学習用データセットに含まれるデータ数が少ない場合における学習処理による識別面の生成の問題点を概念的に説明するための図である。図12Cは、データ拡張処理によりデータ拡張された学習用データセットを用いた学習処理による識別面の生成の一例を概念的に説明するための図である。

[0145] 図12Bに示すように、学習用データセットに含まれるデータ数が少ない場合には、犬と猫のデータ間の隙間が大きいので、A1は、識別面 d_1 、 d_2 、 d_3 で示されるように、犬と猫とを見分ける境界線をどこに引けばわからない。

[0146] 一方、図12Cに示すように、データ拡張処理により適切にデータ拡張された場合、図12Bに示される隙間をデータ拡張されたデータを擬似的に埋めることができるため、A1は、識別面 d_1 で示されるように、犬と猫とを見分ける境界線が引きやすくなる。

[0147] 続いて、データ拡張処理における課題について説明する。

[0148] 図13は、比較例に係るデータ拡張処理における問題点を説明するための図である。図13には、データ拡張をうまくしないと学習により生成される識別面の位置が理想的な識別面からずれてしまうことが示されている。

[0149] すなわち、図13では、猫のデータ51をデータ拡張することで、データ拡張後の猫のデータに関する学習用データセットの分布が分布 Q_1 となっている。同様に、犬のデータ52をデータ拡張することで、データ拡張後の犬のデータに関する学習用データセットの分布が分布 Q_2 となっている。このよう

な場合、A1は、データ拡張後の学習用データセットを学習することで、理想的な識別面 d_4 での位置からずれた位置の識別面 d_5 を生成してしまう。すると、テスト用データセットである犬の画像50がデータ拡張後の犬のデータに関する学習用データセットの分布 Q_2 から外れた位置にある場合、A1は、テスト用データセットである犬の画像50に対しては、猫と識別してしまうことになる。

[0150] このように、学習用データセットの分布に応じてデータ拡張する場合、学習により生成される識別面の位置が理想的な識別面からずれてしまう場合がある。

[0151] 図14は、本開示に係るデータ拡張処理によりデータ拡張された学習用データセットを用いて生成した識別面を概念的に説明するための図である。

[0152] 本開示では、テスト用データセットの分布に合わせて、学習用データセットをデータ拡張処理する。より具体的には、猫のデータ51を猫についてのテスト用データセットの分布 Q_{t1} に合わせてデータ拡張することで、データ拡張後の猫のデータに関する学習用データセットの分布を、分布 Q_3 となるようにすることができる。同様に、犬のデータ52を猫についてのテスト用データセットの分布 Q_{t2} に合わせてデータ拡張することで、データ拡張後の犬のデータに関する学習用データセットの分布を、分布 Q_4 となるようにすることができる。

[0153] これにより、A1は、理想的な識別面 d_4 での位置に近い位置の識別面 d_6 を生成することができる。

[0154] (実験例)

本開示のデータ拡張方法の効果について、The Street View House Numbers (SVHN) Datasetを用いて検証を行ったので、その検証結果を実験例として説明する。

[0155] 図15は、実験例に係る学習用データセットを用いて学習後に、テスト用データセットでエラー率を評価した検証結果を示す図である。

[0156] SVHNデータセットには、600,000桁を超える画像といった大量

のラベル付きデータが組み込まれている。SVHNデータセットは、73,257桁の学習用データセットと、26,032桁のテスト用データセットとで構成されている。また、SVHNデータセットには、追加の学習用データとして使用できる531,131のやや難易度の低いサンプルが追加されている。

- [0157] なお、本実験例では、SVHNデータセットの学習用データセットの32%を分割して評価用データセットとした。
- [0158] また、本実験例では、学習用データセットとして、データの偏り及びラベルノイズのない学習用データセットと、データの偏りまたはラベルノイズのある学習用データセットを用いて検証した。図15では、データの偏りを示すクラスインバランス比（IR）とラベルノイズ比（NR）とにより、データの偏りの有無、ラベルノイズの有無を示している。なお、SVHNデータセットの学習用データセットは、データの偏り及びラベルノイズのない学習用データセットに該当する。
- [0159] 例えば、IR-NRが「100-0.1」である場合、全画像データのうちのランダムな10%（NR=0.1）の画像データがランダムに反転され、ラベルノイズがあることを意味する。また、例えば{0~4}を示す画像データの数が{5~9}を示す画像データの数の100倍となっており、クラスのサンプル数が不均一すなわちデータに偏りがあることを意味する。
- [0160] したがって、図15において、IR-NRが「1-0.0」である学習用データセットは、データの偏り及びラベルノイズのない学習用データセットに該当する。IR-NRが「100-0.0」である学習用データセットは、データの偏りはあるがラベルノイズのない学習用データセットに該当する。同様に、IR-NRが「1-0.1」である学習用データセットは、データの偏りはないが、ラベルノイズがある学習用データセットに該当する。IR-NRが「100-0.1」である学習用データセットは、データの偏り及びラベルノイズのある学習用データセットに該当する。
- [0161] また、図15において、 $\lambda^A(\text{ours})$ 、 $\lambda^{A,W}(\text{ours})$ 、 $\lambda^{A,W,S}$

(ours) は、本開示のデータ拡張方法を示す。本開示のデータ拡張方法としては、図9に示すアルゴリズム1を用いて学習処理すなわち第1最適化処理及び第2最適化処理が行われた。第2最適化処理は、50回目のエポックの後に開始した。

[0162] また、図15において、 $\lambda^A(\text{ours})$ は、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータ λ^A のみに対して最適化されたことを示している。また、 $\lambda^{A,W,S}(\text{ours})$ は、データ拡張処理と損失重みとソフトラベルのハイパーパラメータに対して最適化されたことを示している。 $\lambda^{A,W}(\text{ours})$ は、データ拡張処理と損失重みのハイパーパラメータに対して最適化されたことを示している。これらの最適化処理は、図9に示すアルゴリズム1を用いて行った。

[0163] なお、図15では、比較例としての $\lambda^{\text{ASHA}}(\text{ours})$ の検証結果が示されている。図15において、 $\lambda^{\text{ASHA}}(\text{ours})$ は、データ毎ではなくすべてのデータ間で共有するハイパーパラメータ λ^{ASHA} に対して最適化されたことを示している。

[0164] さらに、図15では、比較例としてBaselineとFAAとの検証結果も示されている。より具体的には、Baselineは、非特許文献1に記載される方法すなわちランダムクロッピング、水平反転、消去などの標準的な変換処理のみを用いてデータ拡張処理をする場合の検証結果を示している。FAA (Fast AutoAugment) は、非特許文献2に記載される方法でデータ拡張処理をする場合の検証結果を示している。

[0165] 図15からわかるように、データの偏りまたはラベルノイズのある学習用データセットを用いる場合、 λ^{ASHA} 、Baseline、FAAの検証結果と比較すると、 $\lambda^A(\text{ours})$ 及び $\lambda^{A,W}(\text{ours})$ の検証結果は同等程度かそれ以下のエラー率（トップ1エラー率）を達成している。

[0166] さらに、 λ^{ASHA} 、Baseline、FAAの検証結果と比較すると、 $\lambda^{A,W,S}(\text{ours})$ の検証結果は、学習用データセットにデータの偏りの有無、ラベルノイズの有無に関わらず、改善されたエラー率（トップ1エラー率）を達成し

ている。

[0167] 以上から、ハイパーパラメータ $\lambda^{A, W, S}$ (ours) に対して最適化だけでなくハイパーパラメータ λ^A (ours) に対して最適化も有効であることがわかる。

[0168] 図16は、実験例に係る学習用データセットを用いて学習させたときの精度学習曲線を示す図である。図16では、 $1R-NR$ が「1-0.0」である学習用データセットを用いて学習させたときの精度学習曲線(c)が示されている。なお、図16において、上記の $\lambda^{A, W, S}$ (ours) に対する最適化処理がAutoD0に該当し、上記の $\lambda^{A, W, S}$ (ours) に対する最適化処理(本開示のデータ拡張方法)のうち、第2最適化処理の開始がAutoD0 startとして示されている。また、AutoD0 startは、50回目のエポックの後に開始されている。

[0169] なお、図16では、比較例として、上記のBaselineとFAAとを実験例に係る学習用データセットを用いて学習させたときの精度学習曲線(a)、(b)も示されている。

[0170] 図16からわかるように、本開示のデータ拡張方法は、AutoD0すなわち第2最適化処理を開始しないと性能向上しないのがわかる。また、本開示のデータ拡張方法は、AutoD0すなわち第2最適化処理を行うことにより、ハイパーパラメータ $\lambda^{A, W, S}$ を有効に最適化でき、性能が急激に向上することがわかる。

[0171] (他の実施態様の可能性)

以上、実施の形態において本開示のデータ拡張方法及び学習装置について説明したが、各処理が実施される主体や装置に関しては特に限定しない。ローカルに配置された特定の装置内に組み込まれたプロセッサなどによって処理されてもよい。またローカルの装置と異なる場所に配置されているクラウドサーバなどによって処理されてもよい。

[0172] なお、本開示は、上記実施の形態に限定されるものではない。例えば、本明細書において記載した構成要素を任意に組み合わせて、また、構成要素の

いくつかを除外して実現される別の実施の形態を本開示の実施の形態としてもよい。また、上記実施の形態に対して本開示の主旨、すなわち、請求の範囲に記載される文言が示す意味を逸脱しない範囲で当業者が思いつく各種変形を施して得られる変形例も本開示に含まれる。

[0173] また、本開示は、さらに、以下のような場合も含まれる。

[0174] (1) 上記の装置は、具体的には、マイクロプロセッサ、ROM、RAM、ハードディスクユニット、ディスプレイユニット、キーボード、マウスなどから構成されるコンピュータシステムである。前記RAMまたはハードディスクユニットには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムに従って動作することにより、各装置は、その機能を達成する。ここでコンピュータプログラムは、所定の機能を達成するために、コンピュータに対する指令を示す命令コードが複数個組み合わされて構成されたものである。

[0175] (2) 上記の装置を構成する構成要素の一部または全部は、1個のシステムLSI (Large Scale Integration: 大規模集積回路) から構成されているとしてもよい。システムLSIは、複数の構成部を1個のチップ上に集積して製造された超多機能LSIであり、具体的には、マイクロプロセッサ、ROM、RAMなどを含んで構成されるコンピュータシステムである。前記RAMには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムに従って動作することにより、システムLSIは、その機能を達成する。

[0176] (3) 上記の装置を構成する構成要素の一部または全部は、各装置に脱着可能なICカードまたは単体のモジュールから構成されているとしてもよい。前記ICカードまたは前記モジュールは、マイクロプロセッサ、ROM、RAMなどから構成されるコンピュータシステムである。前記ICカードまたは前記モジュールは、上記の超多機能LSIを含むとしてもよい。マイクロプロセッサが、コンピュータプログラムに従って動作することにより、前記ICカードまたは前記モジュールは、その機能を達成する。このICカー

ドまたはこのモジュールは、耐タンパ性を有するとしてもよい。

[0177] (4) また、本開示は、上記に示す方法であるとしてもよい。また、これらの方法をコンピュータにより実現するコンピュータプログラムであるとしてもよいし、前記コンピュータプログラムからなるデジタル信号であるとしてもよい。

[0178] (5) また、本開示は、前記コンピュータプログラムまたは前記デジタル信号をコンピュータで読み取り可能な記録媒体、例えば、フレキシブルディスク、ハードディスク、CD-ROM、MO、DVD、DVD-ROM、DVD-RAM、BD (Blu-ray(登録商標) Disc)、半導体メモリなどに記録したものとしてもよい。また、これらの記録媒体に記録されている前記デジタル信号であるとしてもよい。

[0179] また、本開示は、前記コンピュータプログラムまたは前記デジタル信号を、電気通信回線、無線または有線通信回線、インターネットを代表とするネットワーク、データ放送等を経由して伝送するものとしてもよい。

[0180] また、本開示は、マイクロプロセッサとメモリを備えたコンピュータシステムであって、前記メモリは、上記コンピュータプログラムを記憶しており、前記マイクロプロセッサは、前記コンピュータプログラムに従って動作するとしてもよい。

[0181] また、前記プログラムまたは前記デジタル信号を前記記録媒体に記録して移送することにより、または前記プログラムまたは前記デジタル信号を、前記ネットワーク等を経由して移送することにより、独立した他のコンピュータシステムにより実施するとしてもよい。

産業上の利用可能性

[0182] 本開示は、ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法、学習装置およびプログラムに利用でき、特に計算コストを抑制してニューラルネットワークの学習とサンプル毎のハイパーパラメータの最適化とを行うことができるデータ拡張方法、学習装置およびプログラムに利用できる。

符号の説明

[0183]	1 0	学習装置
	1 1	取得部
	1 1 b	Data Sampler
	1 2	データ拡張処理部
	1 2 a	Process
	1 2 b	Augment
	1 4	誤差算出部
	1 5	最適化部
	1 5 a	AutoD0
	1 4 1	Soft-label
	1 4 2	Weight KL Div
	1 4 3	Reweight
	1 0 0 0	コンピュータ
	1 0 0 1	入力装置
	1 0 0 2	出力装置
	1 0 0 4	内蔵ストレージ
	1 0 0 7	読取装置
	1 0 0 8	送受信装置
	1 0 0 9	バス

請求の範囲

[請求項1]

ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法であって、

前記ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、

前記第1最適化処理では、

前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、

前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、

前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、

算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、

前記第2最適化処理では、

前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、

前記複数のパラメータが更新された前記ニューラルネットワークに、前記第2サンプルから、第2ラベルを予測させ、

前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、

算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新する、

データ拡張方法。

[請求項2]

前記ハイパーパラメータを更新する際、

前記学習用データセットに前記データ拡張処理を行うことで得られるデータ拡張後の学習用データセットの分布と、前記評価用データセットの分布とのずれを減らすよう、前記勾配に基づいて前記ハイパーパラメータを更新する、

請求項1に記載のデータ拡張方法。

[請求項3]

前記ハイパーパラメータは、前記データ拡張処理を示す関数において陰関数化され、

前記ニューラルネットワークは、完全微分可能な関数である、

請求項1または2に記載のデータ拡張方法。

[請求項4]

前記第1誤差関数を算出する際、前記第1正解ラベルが示す正解値をソフトラベル化して得た第1正解ソフトラベルと、前記第1拡張ラベルとの誤差をカルバックライブライダイバージェンスを用いて評価する前記第1誤差関数を算出する、

請求項1～3のいずれか1項に記載のデータ拡張方法。

[請求項5]

前記第1誤差関数を算出する際、前記第1誤差関数に対して、さらに、サンプル毎に算出される重みづけを行う、

請求項1～4のいずれか1項に記載のデータ拡張方法。

[請求項6]

ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法を行うための学習装置であって、

プロセッサと、メモリとを備え、

前記プロセッサは、前記メモリを用いて、

前記ニューラルネットワークにおける、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、

前記第1最適化処理では、

前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、

前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、

前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、

算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、

前記第2最適化処理では、

前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、

前記複数のパラメータが更新された前記ニューラルネットワークに、前記第2サンプルから、第2ラベルを予測させ、

前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、

算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新する、学習装置。

[請求項7]

ニューラルネットワークの学習に用いる学習用データセットのデータ拡張方法をコンピュータに実行させるプログラムであって、

前記ニューラルネットワーク、重みを含む複数のパラメータの最適化を行うための第1最適化処理と、データ拡張処理を行う際に用いるサンプル毎のデータ変換処理を規定する変数であるハイパーパラメータの最適化を行うための第2最適化処理とを交互に行い、

前記第1最適化処理では、

前記学習用データセットに含まれる第1サンプルに、前記データ拡張処理を行わせることで、第1拡張サンプルを取得し、

前記ニューラルネットワークに、前記第1拡張サンプルから、第1拡張ラベルを予測させ、

前記第1拡張ラベルと、前記学習用データセットに含まれる前記第1サンプルの正解を示す第1正解ラベルとについての誤差を評価するための第1誤差関数を算出し、

算出した前記第1誤差関数に基づき、前記複数のパラメータを更新し、

前記第2最適化処理では、

前記ニューラルネットワークの性能を評価するためのテスト用データセットの分布に類似した分布のデータセットである評価用データセットから、第2サンプルを取得し、

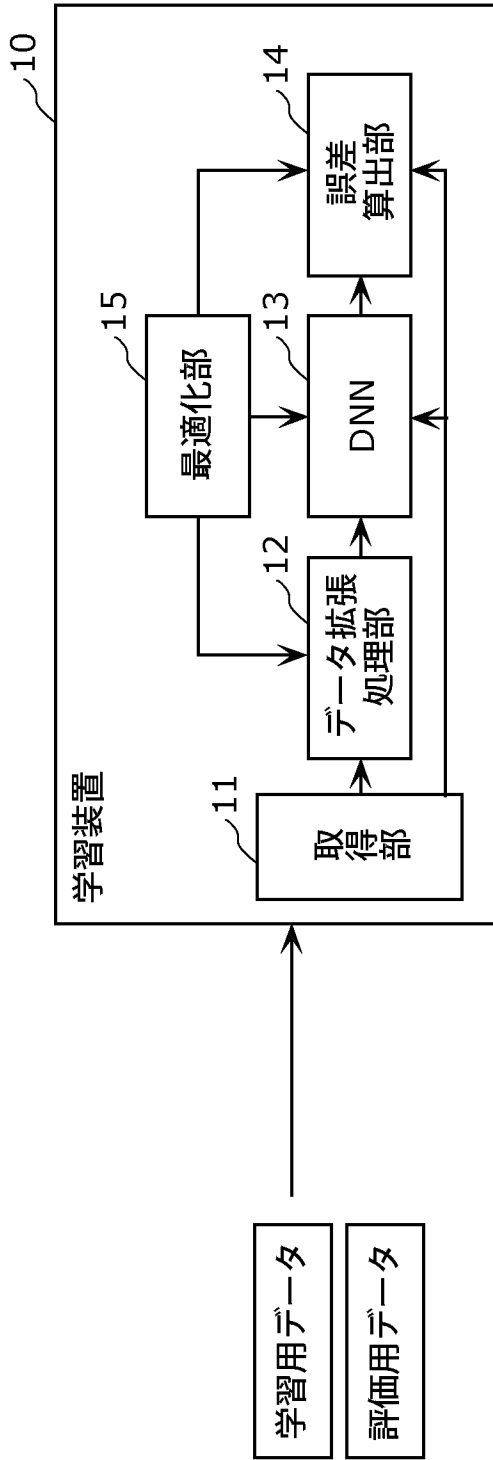
前記複数のパラメータが更新された前記ニューラルネットワークに、前記第2サンプルから、第2ラベルを予測させ、

前記第2ラベルと、前記評価用データセットに含まれる前記第2サンプルの正解を示す第2正解ラベルとの誤差を評価するための第2誤差関数を算出し、

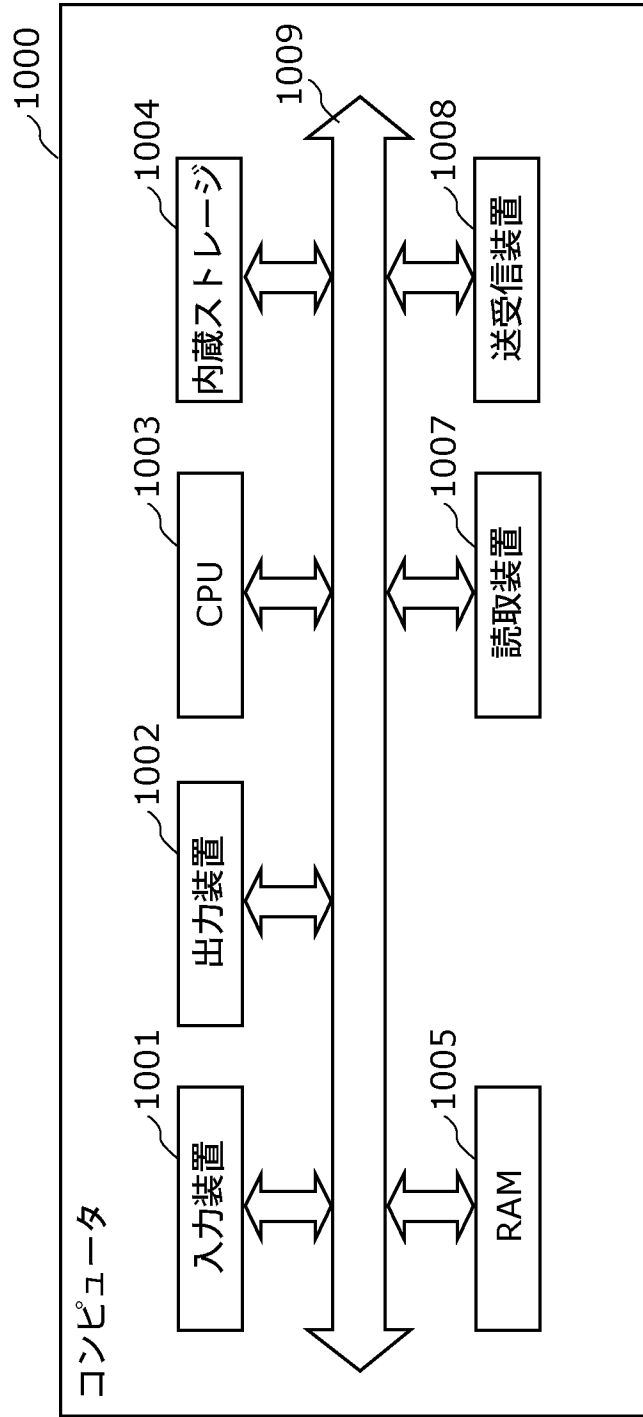
算出した前記第2誤差関数を、前記ハイパーパラメータについて偏微分して得た勾配に基づいて、前記ハイパーパラメータを更新することを、

コンピュータに実行させるプログラム。

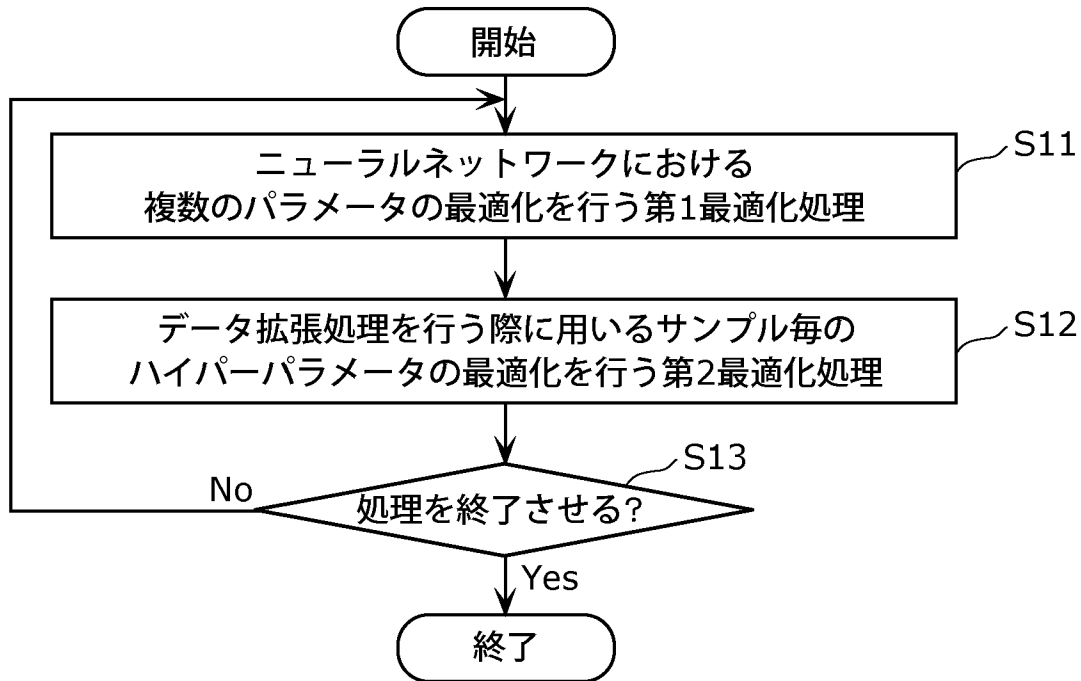
[図1]



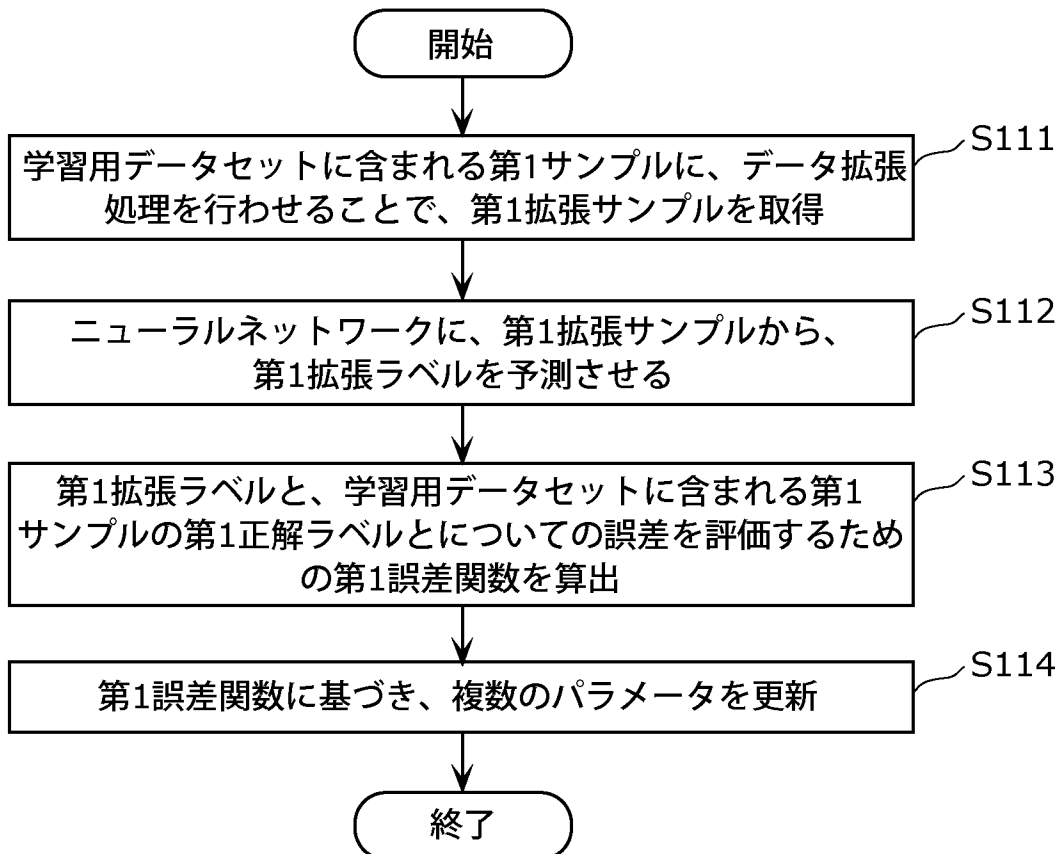
[図2]



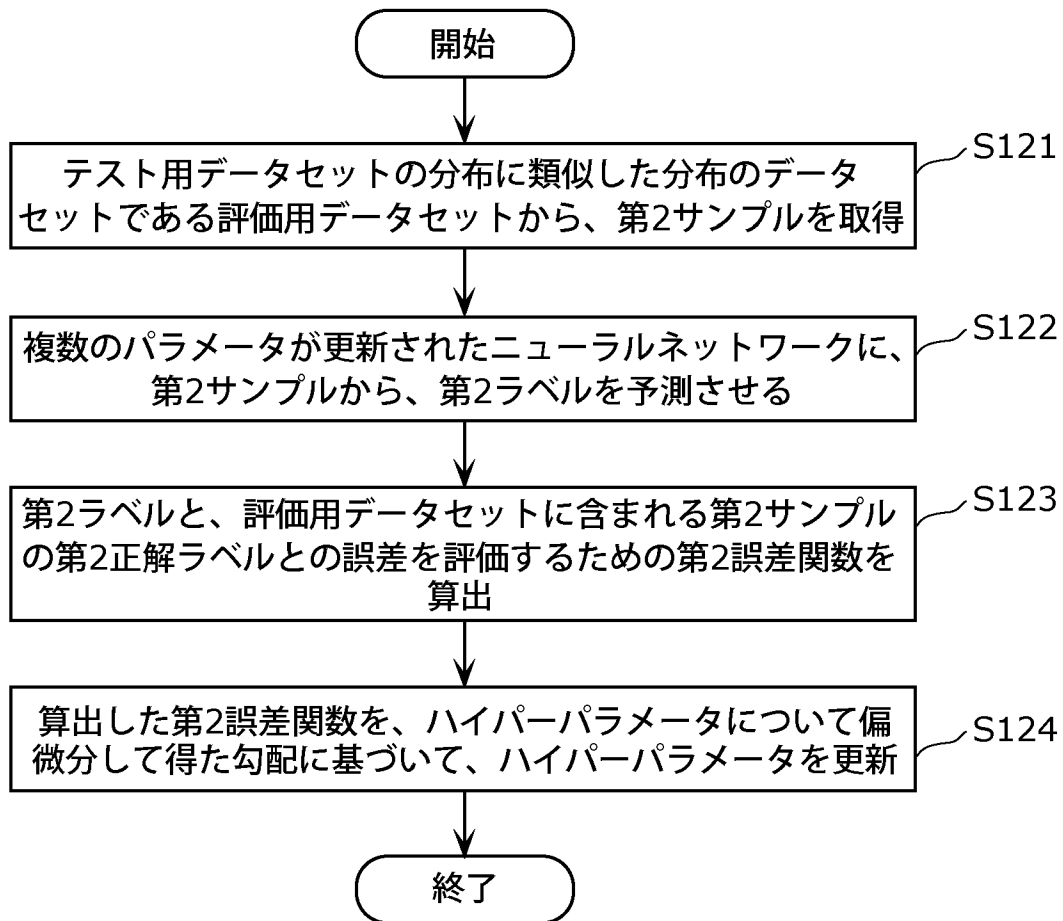
[図3]



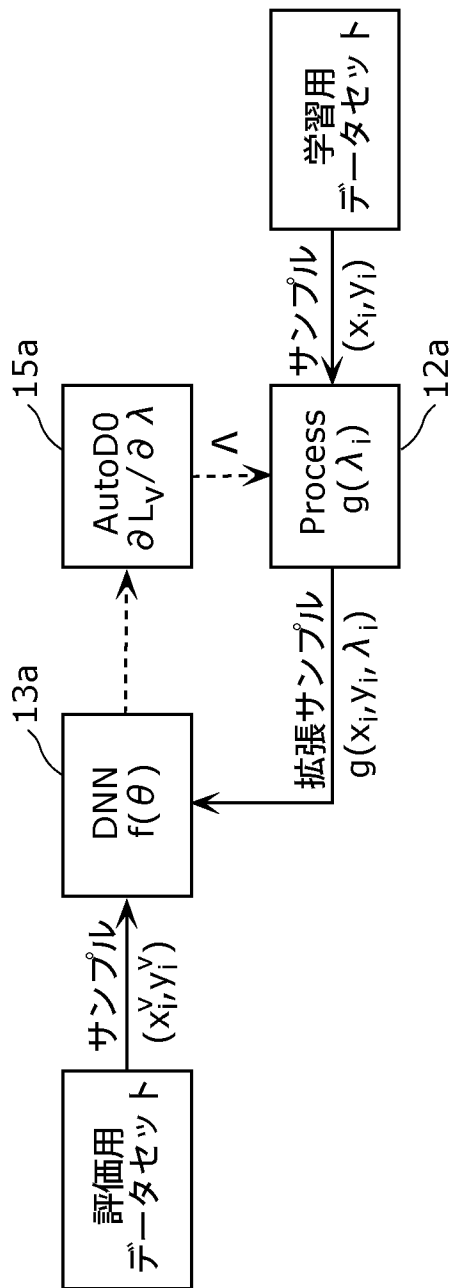
[図4]



[図5]



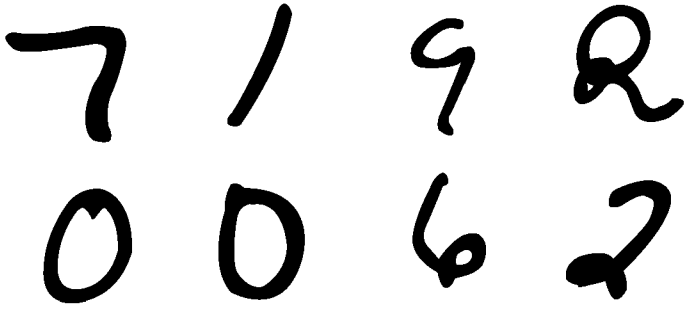
[図6]



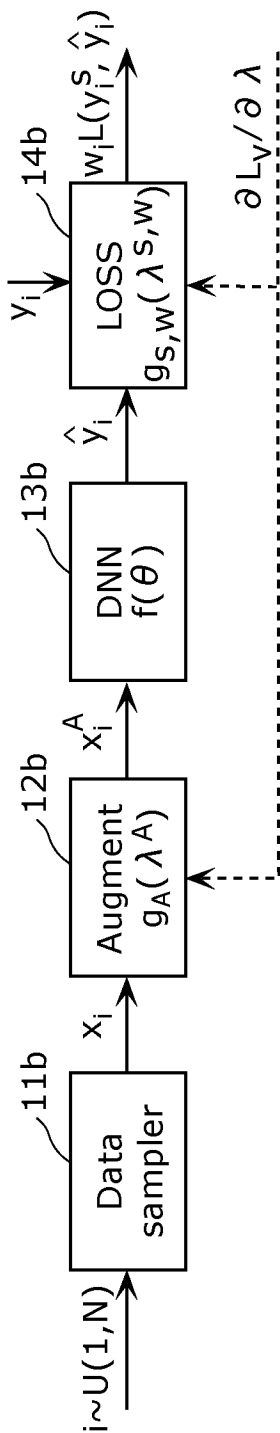
[図7A]

6-79

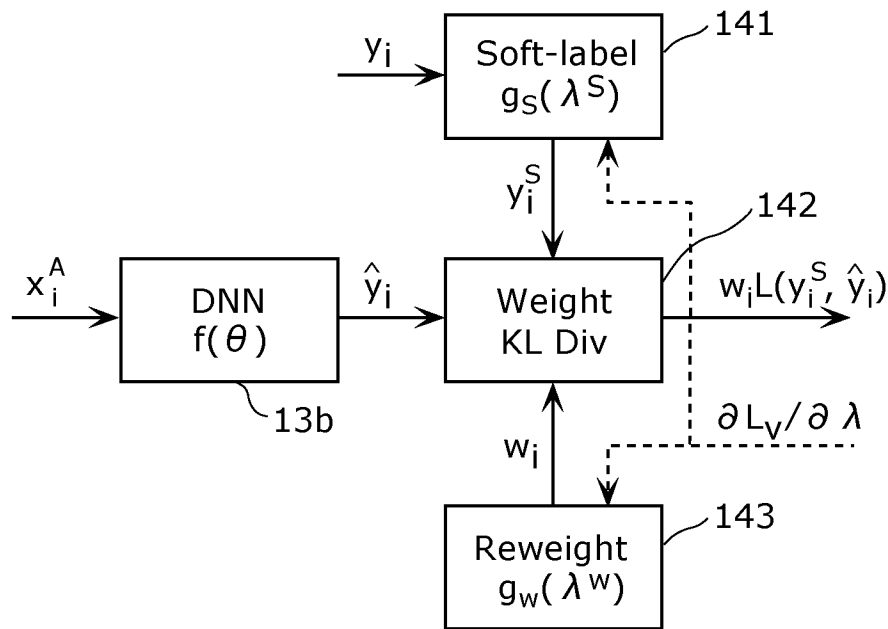
[7B]



[8A]



[圖8B]

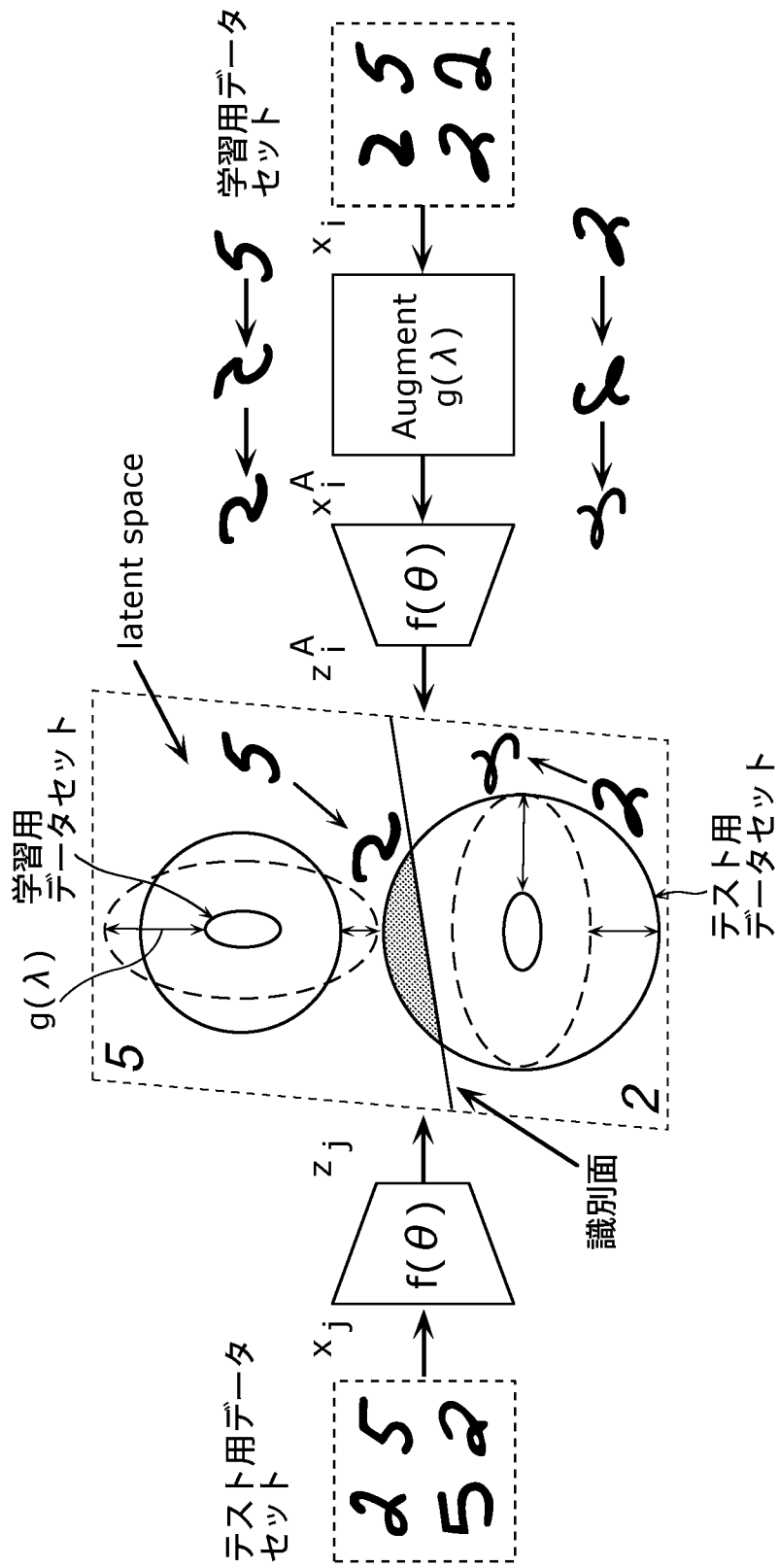


[圖9]

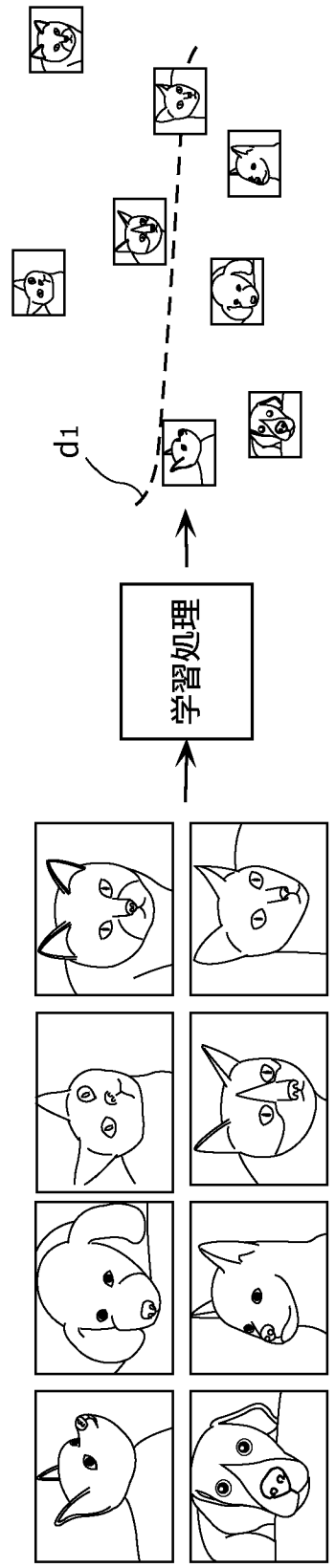
Algorithm 1 optimization of $f(\theta), g(\lambda)$

- 1: Initialize parameters θ and hyperparameters λ
- 2: for epoch = 1...epochs do
- 3: for batch = 1...batches do
- 4: sample batch $\{(x, y)\}_{b_i \in B}$ from D_{train}
- 5: augment data $x^A = g_A(x, \lambda^A)$
- 6: predict $\hat{y} = f(x^A, \theta)$
- 7: generate soft-labels $y^S = g_S(y, \lambda^S)$
- 8: calculate $\nabla_{\theta} (wL(y^S, \hat{y}))$
- 9: update θ using task optimizer
- 10: if epoch > E then
- 11: for batch = 1...batches do
- 12: sample $\{(x, y)\}_{b_i \in B}$ from D_{train}
- 13: sample $\{(x^V, y^V)\}_{b_j \in B}$ from D_{val}
- 14: predict $\hat{y}^V = f(x^V, \theta)$
- 15: predict $\hat{y} = f(x^A, \theta) = f(g_A(x, \lambda^A), \theta)$
- 16: generate soft-labels $y^S = g_S(y, \lambda^S)$
- 17: calculate $\nabla_{\lambda} L_V$ using, where
- 18: $L = wL(y^S, \hat{y})$ and $L_V = L_V(y^V, \hat{y}^V)$
- 19: update λ using HO optimizer

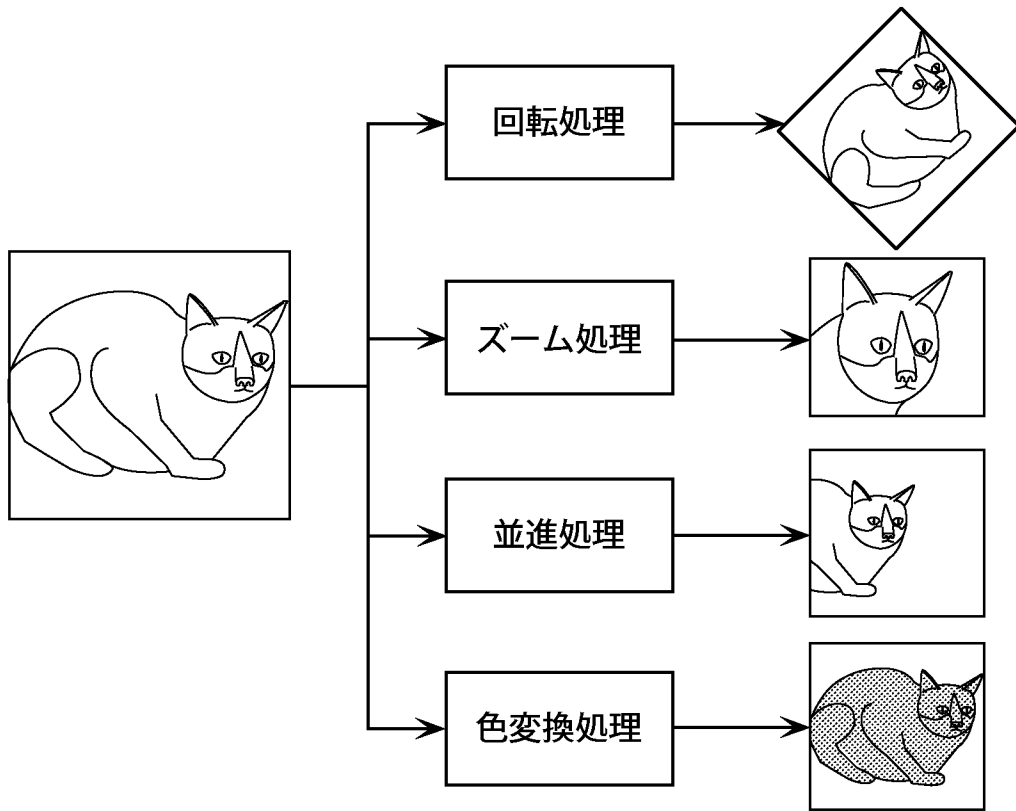
[図10]



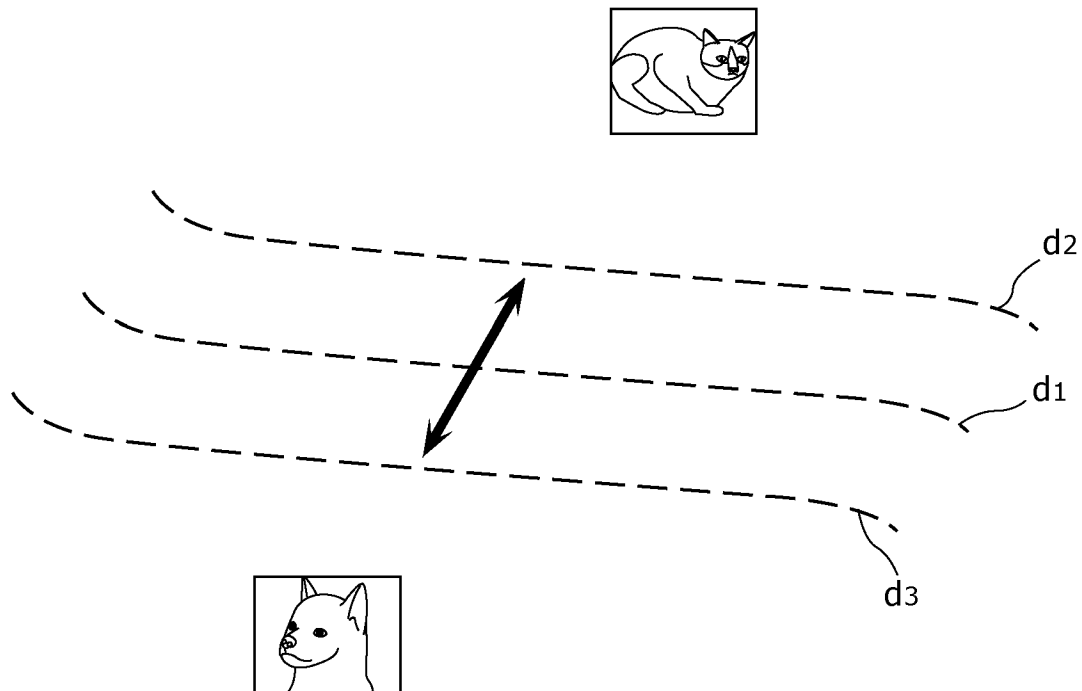
[图11]



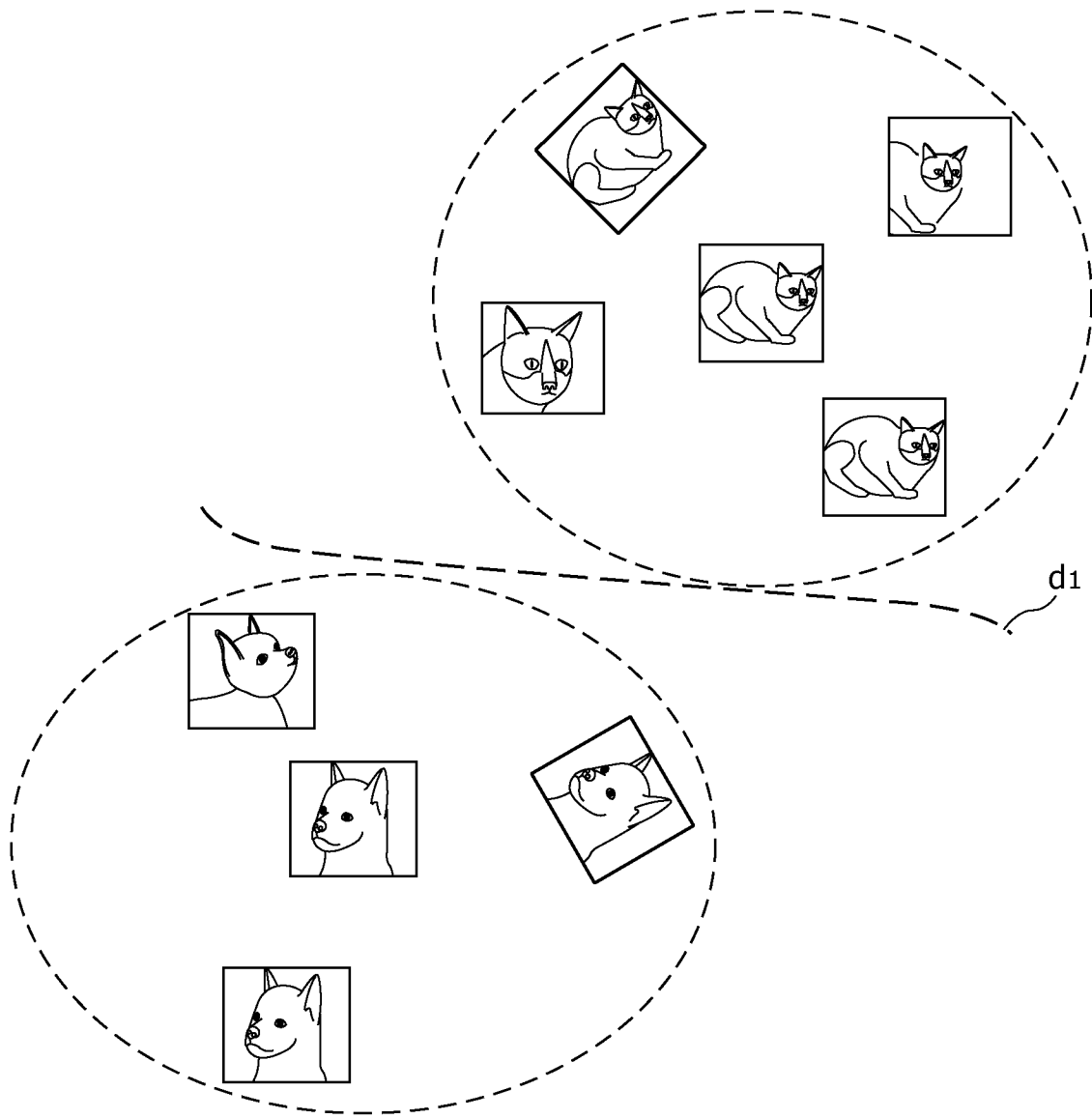
[図12A]



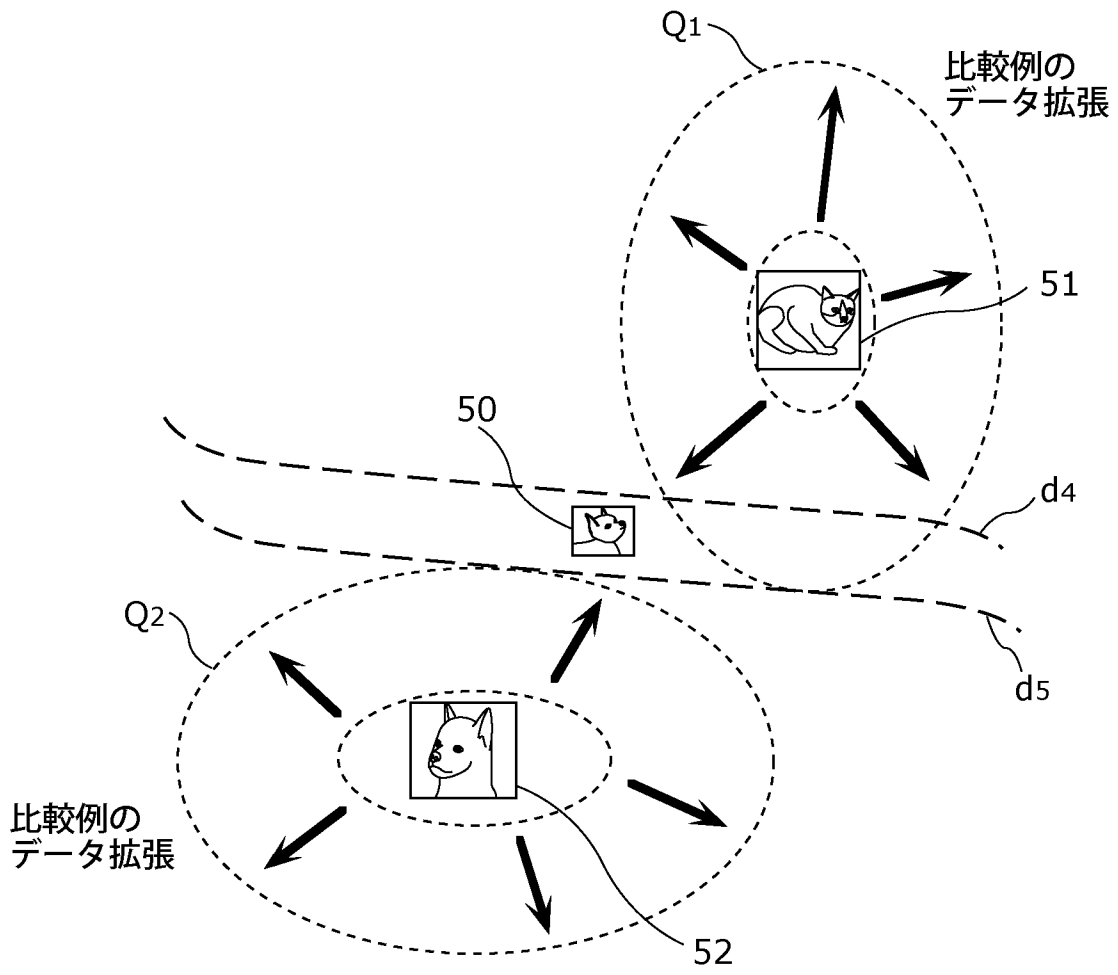
[図12B]



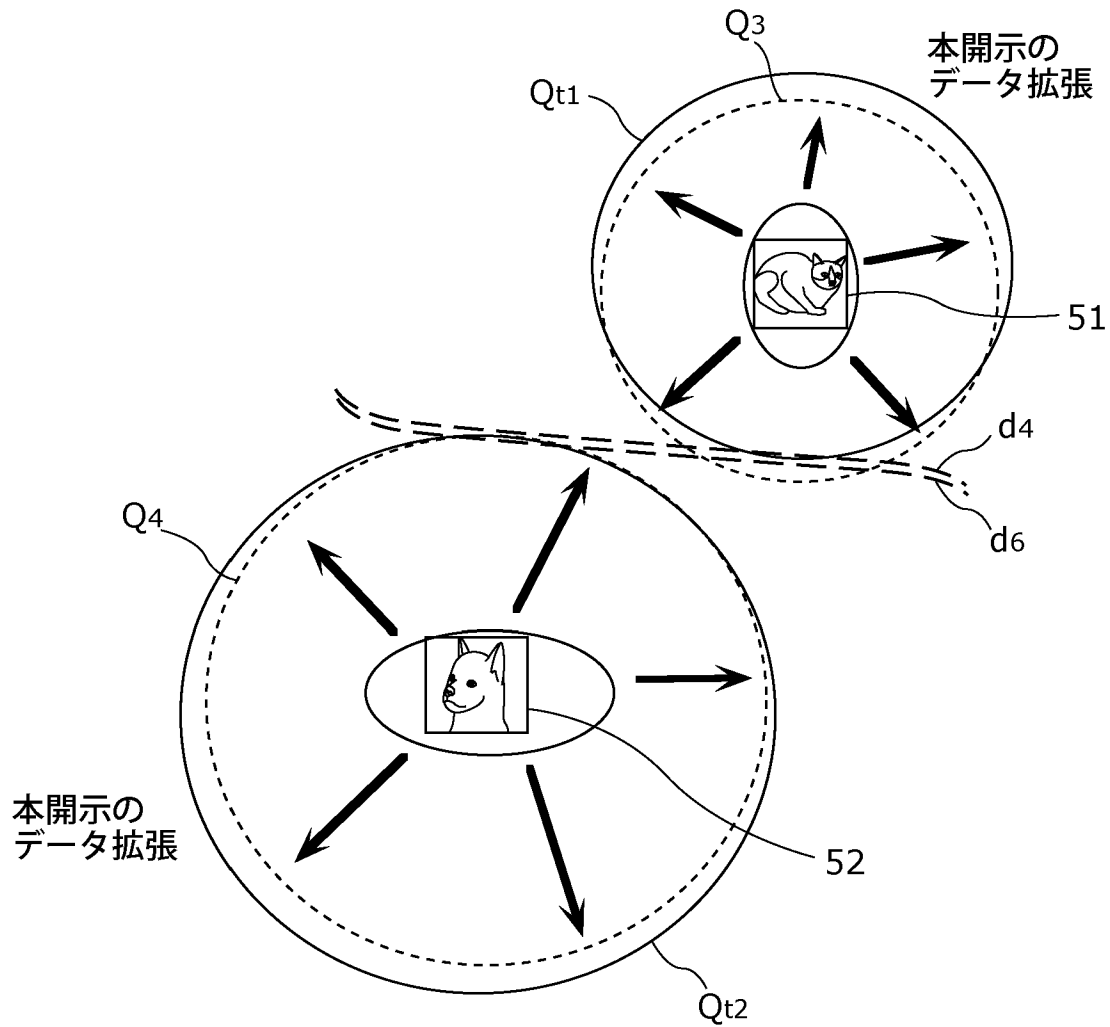
[図12C]



[図13]



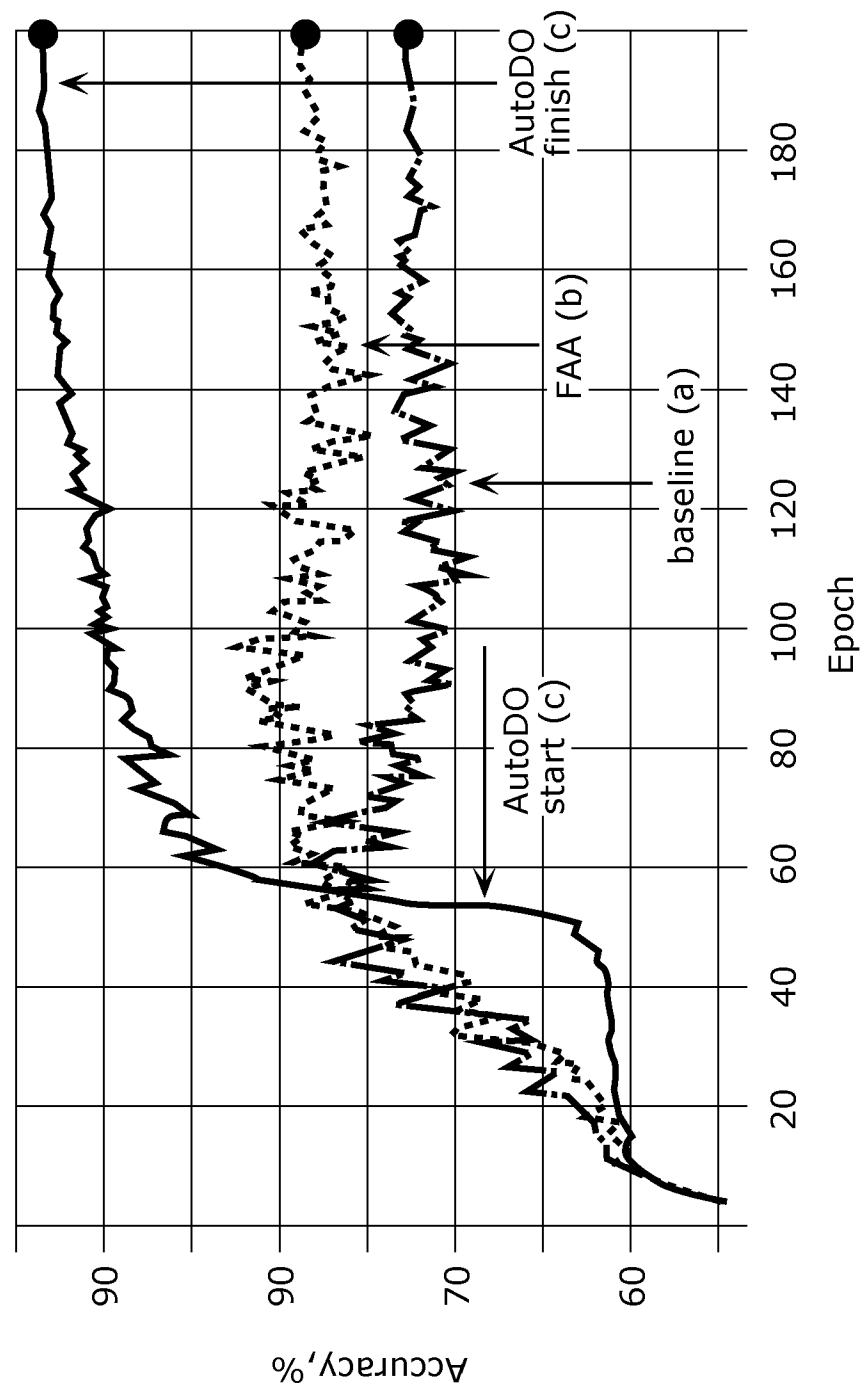
[図14]



[図15]

Alg./IR-NR	1-0.0	100-0.0	1-0.1	100-0.1
Baseline	3.6±0.10	13.6±0.69	5.3±0.27	20.0±1.92
FAA	2.8±0.02	11.5±0.32	3.7±0.08	15.3±1.07
$\lambda^{ASHA}(\text{ours})$	2.8±0.10	12.6±1.53	3.0±0.17	13.7±0.77
$\lambda^A(\text{ours})$	2.7±0.09	10.2±0.50	3.0±0.07	12.3±0.80
$\lambda^{A,W}(\text{ours})$	2.8±0.04	6.1±0.22	2.8±0.07	8.1±0.14
$\lambda^{A,W,S}(\text{ours})$	2.5±0.04	5.3±0.21	2.6±0.05	6.3±0.57

[図16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/040736

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06N 3/08</i> (2006.01)i FI: G06N3/08		
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) G06N3/08		
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 1971-2022 Registered utility model specifications of Japan 1996-2022 Published registered utility model applications of Japan 1994-2022		
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.
X	HATAYA, R., ET AL., Meta Approach to Data Augmentation Optimization, arXiv.org [online], 14 June 2020, [retrieved on 12 January 2022], Retrieved from the Internet: <https://arxiv.org/abs/2006.07965> sections 1, 3, algorithm 1	1-3, 6, 7
Y	sections 1, 3, algorithm 1	4, 5
Y	US 2019/0205748 A1 (INTERNATIONAL BUSINESS MACHINES CORP.) 04 July 2019 (2019-07-04) paragraph [0060]	4, 5
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 12 January 2022		Date of mailing of the international search report 25 January 2022
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/JP2021/040736

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2019/0205748 A1	04 July 2019	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 3/08(2006.01)i FI: G06N3/08		
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）） G06N3/08		
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		
日本国実用新案公報	1922 - 1996年	
日本国公開実用新案公報	1971 - 2022年	
日本国実用新案登録公報	1996 - 2022年	
日本国登録実用新案公報	1994 - 2022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
X	HATAYA, R., ET AL., Meta Approach to Data Augmentation Optimization, arXiv.org [online], 2020.06.14, [retrieved on 2022-01-12], Retrieved from the Internet: <https://arxiv.org/abs/2006.07965> Secs 1, 3, Algorithm 1	1-3, 6, 7
Y	Secs 1, 3, Algorithm 1	4, 5
Y	US 2019/0205748 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 04.07.2019 (2019 - 07 - 04) para. [0060]	4, 5
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 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 “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 “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） “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 “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 “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 “&” 同一パテントファミリー文献	
国際調査を完了した日	12.01.2022	国際調査報告の発送日 25.01.2022
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） 金木 陽一 5B 4876 電話番号 03-3581-1101 内線 3545	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2021/040736

引用文献	公表日	パテントファミリー文献	公表日
US 2019/0205748 A1	04.07.2019	(ファミリーなし)	