

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年11月17日(17.11.2022)



(10) 国際公開番号

WO 2022/239240 A1

(51) 国際特許分類:
G06N 3/08 (2006.01)

(21) 国際出願番号: PCT/JP2021/018449

(22) 国際出願日: 2021年5月14日(14.05.2021)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 伊神 大貴 (**IKAMI Daiki**); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 入江 豪 (**IRIE**

Go); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 柴田 剛志 (**SHIBATA Takashi**); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP).

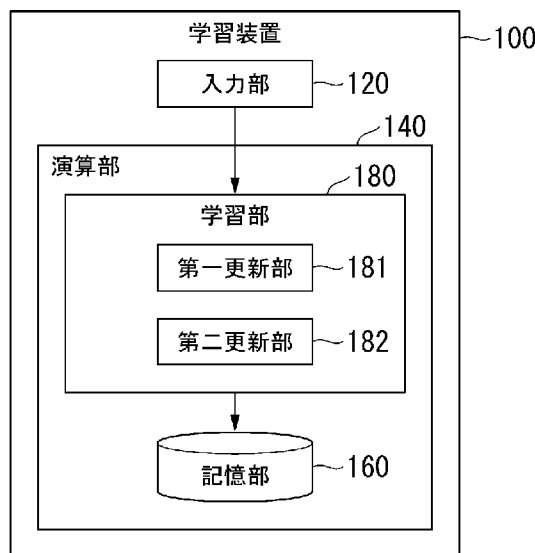
(74) 代理人: 特許業務法人 志賀国際特許事務所 (**SHIGA INTERNATIONAL PATENT OFFICE**); 〒1006620 東京都千代田区丸の内一丁目9番2号 Tokyo (JP).

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

(54) **Title:** TRAINING DEVICE, TRAINING METHOD, AND PROGRAM

(54) 発明の名称: 学習装置、学習方法及びプログラム

[図1]



100 Training device
120 Input unit
140 Calculation unit
160 Storage unit
180 Training unit
181 First updating unit
182 Second updating unit

(57) **Abstract:** This training device comprises a training unit that updates the weight of a neural network on the basis of a gradient of a loss function that calculates the error between output data obtained by inputting input data to the neural network, and correct data corresponding to the input data. The training unit comprises a first updating unit and a second updating unit. The first updating unit calculates gradients used in updating the weight of each channel of each layer of the neural network such that the average value of the gradients becomes zero, and updates the weight of each channel of each layer using the calculated gradients. The second updating unit updates the weights such that the norm of the weights updated by the first updating unit becomes a prescribed value.

KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

添付公開書類：

- 一 国際調査報告 (条約第21条(3))
-

(57) 要約：学習装置は、入力データをニューラルネットワークに入力して得られた出力データと、入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、ニューラルネットワークの重みを更新する学習部を備える。学習部は、第一更新部及び第二更新部を備える。第一更新部は、ニューラルネットワークの各層の各チャンネルの重みそれぞれの更新に用いる勾配の平均値が0となるように勾配を算出し、算出された勾配を用いて各層の各チャンネルの重みを更新する。第二更新部は、第一更新部により更新された重みのノルムが所定の値となるように重みを更新する。

明 細 書

発明の名称：学習装置、学習方法及びプログラム

技術分野

[0001] 本発明は、学習装置、学習方法及びプログラムに関する。

背景技術

[0002] AI（人工知能）技術の進展が目覚ましい。画像認識技術は、いまや数千クラスの物体を識別可能になった。音声認識技術は、自然雑音環境下でも安定した認識が可能な技術が多く実用化され、スマートフォンのような小型端末からでも気軽に利用できるようになっている。機械翻訳も誰もが利用可能になり、国際コミュニケーションの障壁が飛躍的に低下してきている。

[0003] これら近年のAIイノベーションを支えているのは、深層学習と呼ばれる深いニューラルネットワーク（深層ニューラルネットワーク）を学習する技術である。その台頭以来、画像・音・言語といった各種メディアの認識技術は、人間の能力を超えるほどに極めて高い精度レベルを達成するに至っている。

[0004] これら成功と普及の背後にある技術的要因は多岐にわたる。その要因の一つには、深層学習技術そのもの、すなわち、従来は困難であった深層ニューラルネットワークを、効果的に学習可能にする種々の学習技術が発明、あるいは再発明されてきたことが上げられよう。深層ニューラルネットワークの学習は、一般的には何らかの損失関数 L を最小化するようにネットワークパラメータ（重み）を最適化することによって行われる。しかし、単純にこれを実行しようとする、学習が安定しない、計算に時間がかかる、精度が低いなど、様々な問題に直面する。これに対し、例えば、効率的な学習を可能とするために、並列更新を可能にするmini-batch stochastic gradient descent（SGD）や、適切な更新ステップサイズを自動的に調整するAdamといった、種々の勾配法ベースの最適化アルゴリズムが発明されてきた。また、ニューラルネットワークの学習は初期値への依存性が非常に高い。そのため、

初期値の設定がしばしば勾配消失などの学習の失敗を引き起こす原因となっていた。この問題に対して、ネットワーク各層の出力分布に偏りを無くするような初期値設定手法が整備されたために、安定した学習が可能になった。

[0005] さらに、これらと並び重要な基本的技術要素の一つとなっているのが正規化である。正規化とは、変数の取りうる値が特定の範囲に収まるように、そのスケールや平均を変換して扱いやすくする操作である。正規化は、学習を安定化したり、収束を早めたり、精度を改善したりする効果がある。

[0006] ニューラルネットワークの学習における正規化の代表的な手法は、バッチ正規化 (batch normalization: BN) である (例えば、非特許文献 1 参照)。BN では、ネットワークの各層において、バッチ毎に各チャネルを平均 0 かつ分散 1 となるように標準正規化する。BN は学習安定化の効果が非常に高く、ネットワークの精度を増加させることができる。その一方で、BN には、いくつかの大きな弱点が存在する。まず、BN は入力を正規化するため、毎入力ごとに正規化処理を実行する必要がある。よって、計算コストが増加する。また、BN は学習時にはバッチ内の平均・分散を用いて正規化する。そのため、バッチサイズ (バッチ内のサンプル数) が小さい場合には、平均・分散が信頼性のある値にならず、性能が劣化することが知られている。また、学習時とテスト時でデータの統計量が大きく変わるような場合にも、やはり同様に精度が劣化するという問題がある。

[0007] 以上のような BN の問題点を鑑み、ネットワークの入力を直接正規化するのを避け、ネットワークの重みを正規化する重み正規化によって、BN の弱点を回避する試みがなされている。重み正規化が単体で用いられる場合には、一度重みを正規化しておけば、以降どのような入力になされた場合であっても、そのたびに正規化の必要がないため、計算量の面でメリットがある。また、バッチサイズにも依存しないため、統計量の信頼性やずれによる影響も受けにくい。

[0008] これまで、いくつかの重み正規化手法が開示されてきている。例えば、式 (1) に示すように、正規化前の各重み (ベクトル) v を、長さ g と方向 v

／ $||\mathbf{v}||$ に分解して個別に更新し、正規化後の重み $\mathbf{w}$ を得る正規化法がある（例えば、非特許文献2参照）。

[0009] [数1]

$$\mathbf{w} = \frac{g}{||\mathbf{v}||} \mathbf{v} \quad \dots (1)$$

[0010] また、式（2）に示すように、正規化前の重み $\mathbf{v}$ の k 番目の要素 v_k を、重み $\mathbf{w}$ の平均 μ と標準偏差 σ によって標準正規化し、対応する正規化後の重みの要素 w_k を得る方法がある（例えば、非特許文献3参照）。

[0011] [数2]

$$w_k = \frac{v_k - \mu}{\sigma} \quad \dots (2)$$

先行技術文献

非特許文献

[0012] 非特許文献1：Sergey Ioffe and Christian Szegedy, "Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift", In Proceedings of the 32nd International Conference on Machine Learning, Volume 37, p.448-456, 2015.

非特許文献2：Tim Salimans and Diederik P. Kingma, "Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks", In Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS), 2016.

非特許文献3：Siyuan Qiao, Huiyu Wang, Chenxi Liu, Wei Shen, Alan Yuille, "Weight standardization", arXiv preprint, arXiv:1903.10520v1, 2019.

発明の概要

発明が解決しようとする課題

[0013] 大局的にみれば、既存の発明はいずれも、正規化前の重みに対して、平均

を差し引いたり、ノルムや標準偏差で除算することによって別のパラメータに変換 (reparametrization) し、正規化後の重みを得ている。しかしながら、この方式には、学習を崩壊させる「勾配爆発」と呼ばれる問題が起こる可能性があるという根本的な弱点が存在する。このようなreparametrizationに基づく重み正規化法を採用し、正規化前の重み v を勾配ベースの最適化により求めようとした場合、その損失関数 L の重み v についての勾配 $\nabla_v L$ は式 (3) となる。

[0014] [数3]

$$\nabla_v L = \frac{1}{\sigma_v} \left(\nabla L - \frac{1}{m} \langle \mathbf{w}, \nabla_{\mathbf{w}} L \rangle \mathbf{w} \right) \quad \dots (3)$$

[0015] ここで、 $\nabla_{\mathbf{w}} L$ は、 L の正規化後の重み $\mathbf{w}$ についての勾配であり、 $\langle \rangle$ は内積演算であり、 σ_v は正規化前の重み v の標準偏差であり、 m は正規化後の重み $\mathbf{w}$ の次元を表す。ここで問題となるのは、勾配に標準偏差 σ_v による除算が含まれる点である。 σ_v は正規化前の重み v の標準偏差であるから、任意の実数値を取りうる。もし仮に、学習の過程で σ_v が非常に小さい値になるとすると、この勾配は無限大まで発散しうる。すなわち、学習の過程で勾配が発散し、学習自体が失敗する勾配爆発が生じうる。

[0016] 結果として、非特許文献2や非特許文献3に記載の既存の重み正規化法は、それら単体で利用することはできないため、必ずBNなどの活性値正規化法と組み合わせて用いられている。しかしながら、BNと組み合わせることによって、本来重み正規化法を単体で用いることができる場合に享受できるメリット、すなわち、BNの弱点を回避する効果が失われてしまう。そのため、本来のメリットを享受できるような重み正規化法が求められている。

[0017] 上記事情に鑑み、本発明は、計算コストの増加及び学習精度の低下を抑えながらニューラルネットワークを学習することができる学習装置、学習方法及びプログラムを提供することを目的としている。

課題を解決するための手段

[0018] 本発明の一態様の学習装置は、入力データをニューラルネットワークに入力して得られた出力データと、前記入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、前記ニューラルネットワークの重みを更新する学習部を備え、前記学習部は、前記ニューラルネットワークの各層の各チャンネルの前記重みそれぞれの更新に用いる勾配の平均値が0となるように前記勾配を算出し、算出された前記勾配を用いて各層の各チャンネルの前記重みを更新する第一更新部と、前記第一更新部により更新された前記重みのノルムが所定の値となるように前記重みを更新する第二更新部とを備える。

[0019] 本発明の一態様の学習方法は、入力データをニューラルネットワークに入力して得られた出力データと、前記入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、前記ニューラルネットワークの重みを更新する学習ステップを有し、前記学習ステップは、前記ニューラルネットワークの各層の各チャンネルの前記重みそれぞれの更新に用いる勾配の平均値が0となるように前記勾配を算出し、算出された前記勾配を用いて各層の各チャンネルの前記重みを更新する第一更新ステップと、前記第一更新ステップにおいて更新された前記重みのノルムが所定の値となるように前記重みを更新する第二更新ステップとを有する。

[0020] 本発明の一態様のプログラムは、コンピュータを、上述の学習装置として機能させる。

発明の効果

[0021] 本発明により、計算コストの増加及び学習精度の低下を抑えながらニューラルネットワークを学習することが可能となる。

図面の簡単な説明

[0022] [図1]本発明の第1の実施形態による学習装置の構成を示すブロック図である。

[図2]同実施形態による学習装置のハードウェア構成を示す図である。

[図3]同実施形態による学習装置が実行する学習処理を示すフロー図である。

[図4]同実施形態による学習装置の重み算出を説明するための図である。

[図5]同実施形態による学習装置を用いた評価実験の結果を示す図である。

[図6]同実施形態による学習装置を用いた評価実験の結果を示す図である。

[図7]第2の実施形態による学習装置を用いた評価実験の結果を示す図である。

発明を実施するための形態

[0023] 以下、図面を参照しながら本発明の実施形態を詳細に説明する。

[0024] <第1の実施形態>

第1の実施形態は、画像認識に適用した場合の実施形態の一例である。

[0025] <<装置構成>>

図1は、本実施形態による学習装置100の構成を示すブロック図である。学習装置100は、入力された学習用データに基づいて深層ニューラルネットワークを学習する。学習装置100は、入力部120と、演算部140とを備える。演算部140は、記憶部160と、学習部180とを備える。なお、記憶部160は、学習装置100の内部にあっても外部にあっても構わない。本実施形態においては、記憶部160は、学習装置100の内部に備えられる。

[0026] 入力部120は、学習用データを受け付ける。学習用データは、入力データと正解データとを対応付けた情報である。

[0027] 記憶部160は、深層ニューラルネットワークのネットワークパラメータである重みを記憶する。学習部180は、学習用データを用いて、深層ニューラルネットワークの学習処理を実行する。学習部180は、学習処理の結果得られた深層ニューラルネットワークの重みを記憶部160に出力する。学習部180は、第一更新部181と、第二更新部182とを備える。

[0028] 第一更新部181は、入力データを深層ニューラルネットワークに入力して出力データを得る。第一更新部181は、出力データと、入力データに対応した正解データとの誤差を算出する損失関数の勾配を算出する。このとき、第一更新部181は、深層ニューラルネットワークの各層の各要素それぞれ

れの重みの更新に用いられる勾配の平均値が0となるように勾配を算出する。第一更新部181は、算出された勾配を用いて重みを更新する。第二更新部182は、第一更新部181により更新された重みのノルムが所定値となるように重みを更新する。

[0029] 図2は、学習装置100のハードウェア構成を示す図である。学習装置100は、一例として、図2に示すコンピュータ200によって実現される。コンピュータ200は、CPU210と、メモリ220と、記憶部230と、入力部240と、表示部250とを含む。CPU210、メモリ220、記憶部230、入力部240及び表示部250は、バス260を介して互いに接続されている。

[0030] 記憶部230は、例えば、HDD (Hard disk drive)、SSD (Solid State Drive)、フラッシュメモリ等によって実現される。記憶部230には、コンピュータ200を学習装置100として機能させるためのプログラム232が記憶されている。CPU210は、プログラム232を記憶部230から読み出してメモリ220に展開し、プログラム232を実行する。なお、プログラム232をコンピュータ可読媒体に格納して提供してもよい。入力部240は、キーボードやマウスを含む。表示部250は、モニタを含む。

[0031] <<処理概要>>

本実施形態では、学習装置100が画像認識タスクを対象として、深層ニューラルネットワークの一種である畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を学習する場合について説明する。学習装置100の学習部180は、学習用のデータ群を入力として受け付けると、CNNを学習して出力する。

[0032] <<<CNNの構成>>>

CNNは、画像を入力として受け付けると、その画像の認識結果を出力する。CNNの構成は、このような入出力関係を実現できるものであれば任意の構成を採ることができる。例えば、畳み込み層Conv、平均プーリング

層 Avg Pool 、大域的平均プーリング層 GAP (Global Average Pooling)、全結合層 Linear を、次のように接続した9層の CNN を用いることができる。

[0033] $\text{Conv}(3, 64) - \text{Conv}(3, 64) - \text{Conv}(3, 64) - \text{Avg Pool}(2) - \text{Conv}(3, 128) - \text{Conv}(3, 128) - \text{Conv}(3, 128) - \text{Avg Pool}(2) - \text{Conv}(3, 256) - \text{Conv}(3, 256) - \text{Conv}(3, 256) - \text{GAP} - \text{Linear}$

[0034] $\text{Conv}(k, c)$ は、 c チャンネル、 $k \times k$ の大きさの畳み込みカーネルを持つ畳み込み層を表す。 $\text{Avg Pool}(s)$ は、ストライド s の平均プーリング層を表す。

[0035] 学習装置100は、 CNN として、参考文献1に記載のResNetや参考文献2に記載のWideResNetなどの CNN を用いても構わない。

[0036] [参考文献1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", in Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[0037] [参考文献2] Sergey Zagoruyko, Nikos Komodakis, "Wide Residual Networks", In Proceedings of British Machine Vision Conference, 2016.

[0038] <<<学習処理>>>

次に、学習装置100が CNN を学習する学習処理について説明する。学習処理は、 CNN のパラメータである重みを、学習用データに基づいて適切に決定するための処理である。

[0039] 学習装置100が学習処理を実行するためには、予め学習用データ D を準備する必要がある。学習用データは、画像 x と当該画像のクラスを表すクラスラベル y の組 (x, y) の集合 $D = \{(x, y)\}$ によって構成される。クラスラベル y は、該当するクラスのインデクスに対応する要素のみ1を取り、その他の要素は0を取る one hot vector である。この条件を満たす限り、

学習用データを構成するための手段は本実施形態の要点とは無関係であり、どのような手段で準備しても構わない。学習用データを、人手によって準備しても構わないし、自動的に準備しても構わない。

[0040] 図3は、本実施形態における学習装置100が実行する学習処理の概要を示すフロー図である。学習装置100は、上記のような学習用データDを用いて、次の工程により学習処理を実行する。

[0041] まず、ステップS101において、学習装置100の学習部180は、CNNのすべての重み $W = \{w\}$ を初期化する。便宜上、CNNの各層における各チャンネルの重みを w と表す。重み w は、 d 次元ベクトルである。初期化の方法は様々存在するが、好ましくは、Xavierの初期化（Glorotの初期化）、Heの初期化など、重みの平均値が0となるような公知の方法が用いられる。

[0042] 続いてステップS102において、学習部180の第一更新部181は、学習用データDから画像 x とクラスラベル y との組を取得する。第一更新部181は、画像 x をCNNに入力し、CNNの出力 z を求める。第一更新部181は、この出力 z と、画像 x に対応するクラスラベル y とに基づいて、損失関数 L のCNN各層及び各チャンネルの重み w に対する勾配 $\nabla_w L$ を求める。

[0043] 損失関数 L は、重み w に対していたるところで微分可能なものであればどのようなものを用いても構わない。本実施形態においては、損失関数 L として、Softmax Cross Entropyを用いる。Softmax Cross Entropyは、画像認識において用いられることが多い。損失関数 L は、式（4）のように示される。

[0044] [数4]

$$\mathcal{L} = -\frac{1}{n} \sum_{i,k} y_{ik} \log z_{ik} \quad \cdots (4)$$

[0045] y_{ik} は、 i 番目の画像 x に対応するクラスラベル y の k 番目の要素である

。 z_{ik} は、 i 番目の画像 x に対応する CNN の出力 z の k 番目の要素である。
 n は、学習用データの総数、すなわち、 D の集合の大きさである。

[0046] 通常、CNN（あるいは深層ニューラルネットワーク全般）の学習には、学習用データ D のうち一定数のサンプルごとに損失関数の値を計算し、その勾配を求めるミニバッチ学習が導入されることが多い。本ステップにおいても、第一更新部 181 は、ミニバッチ学習を実行してもよい。

[0047] 続いてステップ S103 において、第一更新部 181 は、CNN の各層及び各チャネルの重み w を、以下の式（5）に従い更新する。

[0048] [数5]

$$\mathbf{w} \leftarrow \mathbf{w} - \alpha \left(\nabla_{\mathbf{w}} \mathcal{L} - \frac{1}{d} (\mathbf{1}^T \nabla_{\mathbf{w}} \mathcal{L}) \mathbf{1} \right) \quad \dots (5)$$

[0049] ここで、太字のベクトル $\mathbf{1}$ は全要素が 1 のベクトルであり、 α は学習率である。 α には、任意の正の実数を与えてよい。 α は、通常は、学習用データやネットワークの構造等、諸条件に合わせて調整される。また、第一更新部 181 は、 α を学習の進行とともに変化させても構わない。例えば、第一更新部 181 は、初期値を $\alpha = 0.2$ とし、参考文献 3 の cosine annealing により α の値を変化させる。これにより、高い性能を達成できる。 $\nabla_{\mathbf{w}} \mathcal{L}$ は、損失関数 $\mathcal{L}$ の重み w についての勾配である。 d は、バッチサイズである。すなわち、 d は、1 イテレーションの中で同時に学習処理されるサンプル数である。 d は、任意の正の整数である。

[0050] [参考文献 3] Ilya Loshchilov, Frank Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts", In Proceedings of International Conference on Learning Representations, 2017.

[0051] 続いてステップ S104 において、第二更新部 182 は、CNN の各層及び各チャネルの重み w を、式（6）に従い更新する。

[0052] [数6]

$$\mathbf{w} \leftarrow \sqrt{c} \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad \dots (6)$$

[0053] ここで、 c は定数である。第二更新部182は、定数 c に任意の正の実数を与えてよいが、好ましくは $c = \sqrt{2}$ とする。この時、各層及び各チャンネルの重み w が相互に独立であり、入力 x が独立同分布サンプルとして与えられ、さらに、重み w と画像 x とが相互に独立であるような場合、全ての層の重みの分散を1にすることができる。

[0054] 学習部180は、以上のステップS102～ステップS104の処理を、所定の回数繰り返す。学習部180は、所定の回数の繰り返が終了した場合、ステップS105において、終了条件を満たしたと判定する。ステップS106において、第二更新部182は、最終的に学習した重み $W = \{w\}$ を記憶部160に出力し、図3の学習処理を終了する。記憶部160は、第二更新部182から出力された重み W を記憶する。

[0055] 数学的には、上記の学習処理は、各層及び各チャンネルの重み w の平均が0、かつ、ノルムが c となるような制約の下で、損失関数 L を最小化する手続きである。この手続きでは、重みを正規化するのではなく、最初から正規化された重みを求めることに相当する。具体的には、以下の式(7)の制約付き問題を解いている。

[0056] [数7]

$$\begin{aligned} & \min_{\{\mathbf{W}^{(l)}\}} \mathcal{L}(\{\mathbf{W}^{(l)}\}), \forall l \\ & \text{s.t. } \|\mathbf{w}_i^{(l)}\|_2^2 = c^{(l)}, \mathbf{1}^\top \mathbf{w}_i^{(l)} = 0, \forall i, l, \quad \dots (7) \end{aligned}$$

[0057] l はCNNの層のインデックス、 i は層内のチャンネルのインデックスである。 $W^{(l)}$ は l 層目の重み、 $w_i^{(l)}$ は l 層目の i 番目のチャンネルの重みである。 $c^{(l)}$ は定数である。 $c^{(l)}$ は、0より大きい任意の実数であるが、好ましくは2である。特定の条件を満たした場合に、分散を1とすることができる。式(7)における $\|\mathbf{w}_i^{(l)}\|_2^2 = c^{(l)}$ は、重みの分散を一定にする制約である。式(7)における $\mathbf{1}^\top \mathbf{w}_i^{(l)} = 0$ は、重みの平均を0にする制約である。

[0058] 式（５）における勾配 $\nabla_w L$ は従来技術と同様に求められる。式（５）の右辺において $(1/d)(1^T \nabla_w L)$ の項（以下、項Ａと記載）がなければ従来の勾配と同様であるが、項Ａがあることによって、勾配の平均が０になるように重みが更新される。そして、式（６）により、重みのノルムが c になるように正規化される。

[0059] 図４は、本実施形態における学習装置１００による重みの算出を説明するための図である。 $w^{(t)}$ は、 t 回目の学習処理で算出された重み w である。 $w^{(t)}$ のノルムは c である。すなわち、 $w^{(t)}$ は、 $c = ||w||_2$ で表される球 Q 上の点である。 $w^{(t)}$ から $-\alpha \nabla_w L$ を減算した値が w' である。つまり、 $w^{(t)}$ から w' への方向が、本来の勾配 $\nabla_w L$ の方向である。これは、式（５）の右辺に項Ａがない場合の値に相当する。項Ａを含んだ式（５）により、 $w^{(t)}$ における接平面 P に w' からおろした垂線の足である点 w^- が算出される。点 w^- は、式（５）の左辺の更新後の重み w に相当する。そして、式（６）により、 w^- から球面 Q におろした垂線が球面 Q と交わる点 $w^{(t+1)}$ が求められる。点 $w^{(t+1)}$ は、更新後の重み w である。 $w^{(t+1)}$ は球面 Q 上の点であるため、 $w^{(t+1)}$ のノルムは c である。

[0060] 上述のように、各層及び各チャネルの重み w の平均が０、かつ、ノルムが c となるような制約の下で、損失関数 L を最小化している。然るに、最終的な重みは、平均が０、分散が一定に正規化されたものとなっており、重み正規化と同様の効果を得ることができる。非特許文献２、非特許文献３に記載の従来の重み正規化法と異なり、本実施形態では、勾配に σ_v による除算が発生しない。よって、勾配爆発を起こさず安定した学習が可能である。この利点により、本実施形態は、ＢＮ等と組み合わせることなく学習可能であり、また、ＢＮと同等の精度を実現可能である。

[0061] <<<線形変換処理>>>

重み正規化は、全ての層における重みの平均値及び分散を画一化するものである。ゆえに、ＣＮＮ全体としての表現能力を低下させる可能性がある。この問題を回避するため、各層、各チャネルごとにスケールパラメータ γ と

バイアスパラメータ β を導入し、以下の式 (8) に示すように、各層、各チャネルの出力 p に線形変換を与えた新たな出力 q を出力するように変更してもよい。すなわち、第一更新部 181 は、入力データ x を深層ニューラルネットワークに入力した際に算出される各層、各チャネルの出力に、式 (8) の演算を行って次の層に入力する。学習部 180 は、重み w に加えて、これらのパラメータ γ 、 β を、重み w と同様に損失関数の勾配に基づいて学習してもよい。

[0062] [数8]

$$q = \gamma p + \beta \quad \dots (8)$$

[0063] これにより、各層、各チャネルの出力 q は、元の出力 p よりもダイナミックレンジの調整が容易になり、性能を高めることが可能になる。

[0064] <<<乱数乗算処理>>>

一般に、深層ニューラルネットワークの学習時に、適当なランダム性を導入することによって、最終的な性能を改善できる場合があることが知られている。代表的な例としては、ネットワークの出力や結合を一部ランダムに落とす dropout や、学習用データにランダムな変換を与えるデータ拡張がある。さらに、より最近は、参考文献 4 に記載の NoisyStudent などが挙げられる。また、活性値正規化法である BN においても、ミニバッチ統計量の推定バイアスがもたらすランダム性が、最終的な性能を改善する効果を与えていることを示唆する分析結果が報告されている（例えば、参考文献 5 参照）。

[0065] [参考文献 4] Jiacheng Sun, Xiangyong Cao, Hanwen Liang, Weiran Huang, Zewei Chen, Zhenguo Li, "New Interpretations of Normalization Methods in Deep Learning", In Proceedings of The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), 2020.

[0066] [参考文献 5] Cecilia Summers, Michael J. Dinneen, "Four things everyone should know to improve batch normalization", In Proceedings of the International Conference on Learning Representations, 2020.

[0067] これらの知見に着想を得て、本実施形態においても、学習時に、各層、各

チャネルの出力にランダム性を導入してもよい。具体的には、学習時においてのみ、式（８）の代わりに、式（９）によって各層、各チャネルの出力が定められる。式（９）により、線形変換が施されたニューラルネットワークの各層及び各要素の出力に、確率乱数が乗算される。第一更新部１８１は、入力データ x を深層ニューラルネットワークに入力した際に算出される各層、各チャネルの出力に式（９）の演算を行い、次の層に入力する。

[0068] [数9]

$$q = (1 + \varepsilon)\gamma p + \beta \quad \dots (9)$$

[0069] ここで、 ε は、平均０、分散 σ_l^2 を持つ正規乱数である。分散 σ_l^2 の値は、正の実数を取る限り任意の値を定めることができる。例えば、 $\sigma_l^2 = 0.4$ などと定めてもよい。一方、ResNetやWideResNetなど、層数の多いCNNを学習する場合には、層ごとにその値を変化させる方が、より大きな性能改善効果を得られる場合がある。CNNの全層数を L としたとき、本実施形態においては、次の式（１０）ように分散 σ_l を定める。

[0070] [数10]

$$\sigma_l = \frac{l}{L} s \quad \dots (10)$$

[0071] ここで、 s は正の実数を持つ定数であり、例えば $s = 0.4$ などとする。

[0072] 以上、本実施形態の学習方法の処理の詳細を説明した。

[0073] <<評価実験>>

以上説明した本実施形態に関する性能評価実験結果を示す。本実験では、評価用データとして、画像認識性能評価に広く用いられているCIFAR-10、CIFAR-100と呼ばれるデータを用いる。学習するCNNとして、先に示した９層のCNN、ResNet、及び、WideResNetを用いる。CNNを学習する際にはデータ拡張を用いることが多いが、本実験では、ごく一般的な水平反転及びランダムクロップを利用する標準的なデータ拡張（Standard）と、最適化したデータ拡張であるFast AutoAugment（FastAA）を用いる。

[0074] 以上の条件において、本実施形態による重み正規化の性能を、既存の活性化値正規化法／重み正規化法である、BN（非特許文献1）、GN（参考文献6）、WS+BN（非特許文献2）、WS+GN（非特許文献2）の性能と比較する。WS+BNは、非特許文献2に記載の重み正規化法であるWSをBNと併用したものを表し、WS+GNは同じくWSをGNと併用したものを表している。なお、WSは単体では学習に失敗し、十分な性能を得ることができないため、これらの活性化値正規化法と併用した場合の結果を示している。

[0075] [参考文献6] Yuxin Wu, Kaiming He, "Group normalization", In Proceedings of European Conference on Computer Vision, 2018.

[0076] 図5及び図6は、学習装置100を用いた評価実験の結果を示す図である。図5及び図6に示す表中の数値は、誤差率である。誤差率は、識別を誤った割合である。誤差率が、小さいほど性能がよいことを示す。また、同条件下のうち、最もよい性能を示した手法の数値に下線が付されている。図5及び図6に示す結果から、次の事実が観測できる。

[0077] (1) BNとWS+BNとを比較、あるいは、GNとWS+GNとを比較すると、その性能差はほとんどなく、条件によっては劣化する場合がある。すなわち、BN、GNを単体で利用しても十分な性能を得ることが可能であり、WSを組み合わせるメリットは限定的であると言える。

(2) 本実施形態は、既存手法の中では最も性能の高いBNとほぼ同色の性能を発揮しており、特にCIFAR-100においてはほとんどの場合にこれを上回る性能を達成している。

[0078] 以上のように、本実施形態は、重み正規化法でありながら、それ単体で活性化値正規化法であるBNと同程度の性能を達成できることが確認できる。すなわち、本実施形態によれば、重み正規化法本来のメリットを享受することができる高精度な学習を実現可能である。

[0079] <第2の実施形態>

第2の実施形態は、点群物体認識に適用した場合の実施形態の一例である

。本実施形態では、点群計測された物体を識別する点群物体認識タスクを対象として、点群を対象とした深層ニューラルネットワークの一種であるグラフ畳み込みニューラルネットワーク（Graph Convolutional Neural Network: GCNN）を学習する場合について説明する。本実施形態を、画像認識に適用した場合の実施形態との差分を中心に説明する。

[0080] <<装置構成>>

本実施形態の学習装置として、図1に示す第1の実施形態の学習装置100を用いることができる。学習装置100は、学習用のデータ群を入力として受け付けると、GCNNを学習し、学習したGCNNを出力する。

[0081] <<処理概要>>

<<<GCNNの構成>>>

GCNNは、点群データを入力として受け付けると、その点群データの認識結果を出力する。GCNNの構成は、このような入出力関係を実現できるものであれば任意の構成を採ることができる。例えば、本実施形態においては、参考文献7に記載のGCNNを用いる。

[0082] [参考文献7] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, Justin M. Solomon, "Dynamic Graph CNN for Learning on Point Clouds", ACM Transactions on Graphics (TOG), 38(5):1-12, 2019.

[0083] <<<学習処理>>>

学習処理については、図3に示す第1の実施形態の画像認識の場合における処理と同一である。入力データ形式には、第1の実施形態が画像、本実施形態は点群であるとの違いがある。しかし、本実施形態では、CNNではなくGCNNを用いるという利用する深層ニューラルネットワークの変更により、この違いが吸収される。そのため、以降の処理は、線形変換処理、乱数乗算処理も含め、全て同様の処理を適用することが可能である。

[0084] <<評価実験>>

本実施形態の一例に関する性能評価実験結果を示す。本実験では、評価用

データとして、点群物体認識性能評価に広く用いられているModelNet40と呼ばれるデータを用いる。先の画像認識性能に対する評価実験同様、本実施形態による重み正規化の性能を、BN、GN、WS+BN、WS+GNの性能と比較する。

[0085] 図7は、本実施形態の学習装置100を用いた性能評価実験の結果を示す図である。図6に示す表中の数値は認識正解率であり、大きいほど性能がよいことを示す。画像認識の場合の結果とは異なり、GNを除いてどの手法も大きな性能差は見られない。しかしながら、既存手法はいずれも活性化値正規化法を用いていることを鑑みると認識に要する計算コストは本実施形態よりも大きく、その点において本実施形態に利点がある。

[0086] 以上より、本実施形態の一例においても、重み正規化法でありながら、単体で活性化値正規化法であるBNと同程度の性能を達成できることが示された。

[0087] 以上、本発明の実施形態について詳細に説明した。

[0088] 本発明の実施形態における学習方法を、汎用演算処理装置、記憶装置等を備えたコンピュータやサーバ等により構成して、各処理がプログラムによって実行されるものとしてもよい。このプログラムは記憶装置に記憶されており、磁気ディスク、光ディスク、半導体メモリ等の記録媒体に記録することも、ネットワークを通して提供することも可能である。もちろん、その他いかなる構成要素についても、単一のコンピュータやサーバによって実現しなければならないものではなく、ネットワークによって接続された複数のコンピュータに分散して実現しても構わない。

[0089] 以上説明した実施形態によれば、学習装置は、学習部を備える。学習部は、入力データをニューラルネットワークに入力して得られた出力データと、入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、ニューラルネットワークの重みを更新する。学習部は、第一更新部と、第二更新部とを備える。第一更新部は、ニューラルネットワークの各層の各チャンネルの重みそれぞれの更新に用いる勾配の平均値が0となるように勾

配を算出し、算出した勾配を用いて各層の各チャンネルの重みを更新する。第二更新部は、第一更新部により更新された重みのノルムが所定の値となるように重みを更新する。

[0090] 第二更新部は、第一更新部により更新された重みを、当該重みのノルムにより除算して重みを更新する。また、第二更新部は、第一更新部により更新された重みを、当該重みのノルムにより除算し、さらに、定数を乗算することにより重みを更新する。

[0091] 第一更新部は、入力データをニューラルネットワークに入力することにより得られる各層及び各要素の出力に線形変換を施す処理と、線形変換のパラメータを損失関数の勾配に基づいて更新する処理とを行ってもよい。さらに、第一更新部は、線形変換が施されたニューラルネットワークの各層及び各要素の出力に確率乱数を乗算してもよい。

[0092] 本実施形態の学習装置、学習方法、及びプログラムによれば、単体で用いた場合であっても勾配爆発を起こさず、また、然るに重み正規化本来のメリットである、計算コストが小さい、バッチ統計量に依存しないといった性質を享受できる正規化法を提供できる。結果として、高精度な学習を実現することができる。

[0093] 以上、この発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこれら実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

符号の説明

[0094] 1 0 0 学習装置
 1 2 0 入力部
 1 4 0 演算部
 1 6 0 記憶部
 1 8 0 学習部
 1 8 1 第一更新部
 1 8 2 第二更新部

200 コンピュータ

210 CPU

220 メモリ

230 記憶部

240 入力部

250 表示部

請求の範囲

- [請求項1] 入力データをニューラルネットワークに入力して得られた出力データと、前記入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、前記ニューラルネットワークの重みを更新する学習部を備え、
- 前記学習部は、前記ニューラルネットワークの各層の各チャンネルの前記重みそれぞれの更新に用いる勾配の平均値が0となるように前記勾配を算出し、算出された前記勾配を用いて各層の各チャンネルの前記重みを更新する第一更新部と、
- 前記第一更新部により更新された前記重みのノルムが所定の値となるように前記重みを更新する第二更新部とを備える、
- 学習装置。
- [請求項2] 前記第二更新部は、前記第一更新部により更新された前記重みを、当該重みのノルムにより除算して前記重みを更新する、
- 請求項1に記載の学習装置。
- [請求項3] 前記第二更新部は、前記第一更新部により更新された前記重みを、当該重みのノルムにより除算し、さらに、定数を乗算することにより前記重みを更新する、
- 請求項1に記載の学習装置。
- [請求項4] 前記第一更新部は、前記入力データを前記ニューラルネットワークに入力することにより得られる各層及び各要素の出力に線形変換を施す処理と、前記線形変換のパラメータを前記損失関数の勾配に基づいて更新する処理とを行う、
- 請求項1に記載の学習装置。
- [請求項5] 前記第一更新部は、線形変換が施された前記ニューラルネットワークの各層及び各要素の出力に確率乱数を乗算する、
- 請求項4に記載の学習装置。
- [請求項6] 入力データをニューラルネットワークに入力して得られた出力デー

タと、前記入力データに対応した正解データとの誤差を算出する損失関数の勾配に基づいて、前記ニューラルネットワークの重みを更新する学習ステップを有し、

前記学習ステップは、

前記ニューラルネットワークの各層の各チャンネルの前記重みそれぞれの更新に用いる勾配の平均値が0となるように前記勾配を算出し、算出された前記勾配を用いて各層の各チャンネルの前記重みを更新する第一更新ステップと、

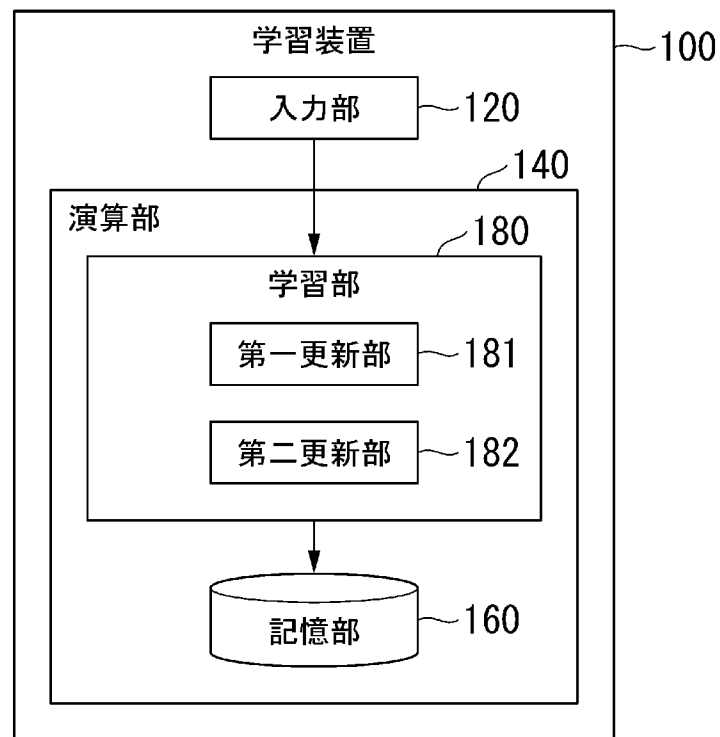
前記第一更新ステップにおいて更新された前記重みのノルムが所定の値となるように前記重みを更新する第二更新ステップとを有する、学習方法。

[請求項7]

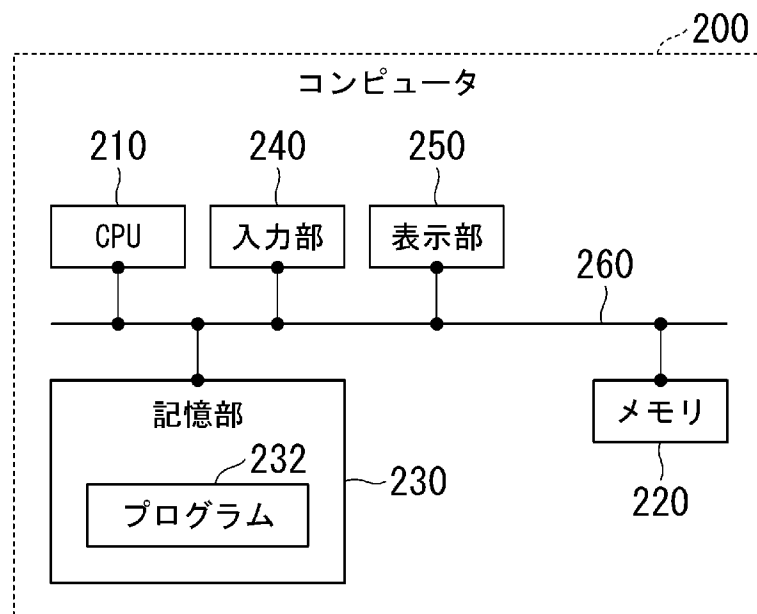
コンピュータを、

請求項1から請求項5のいずれか一項に記載の学習装置として機能させるためのプログラム。

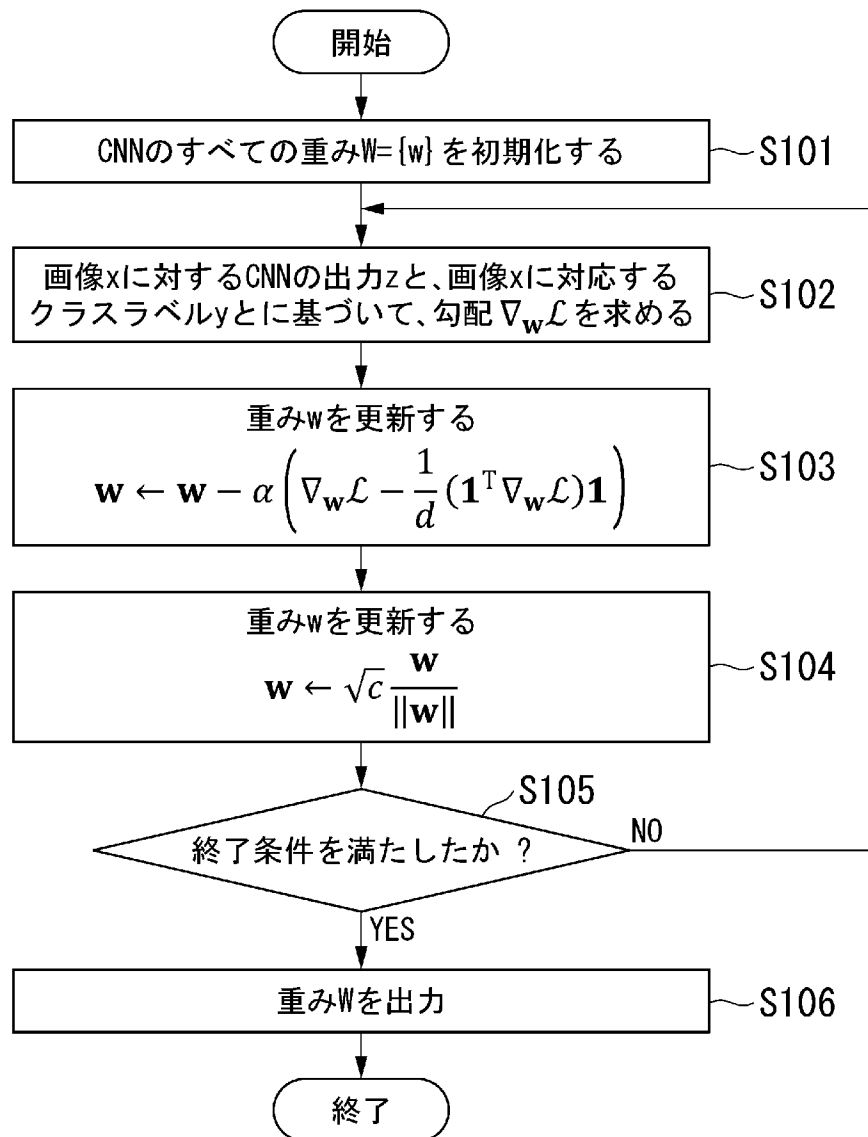
[図1]



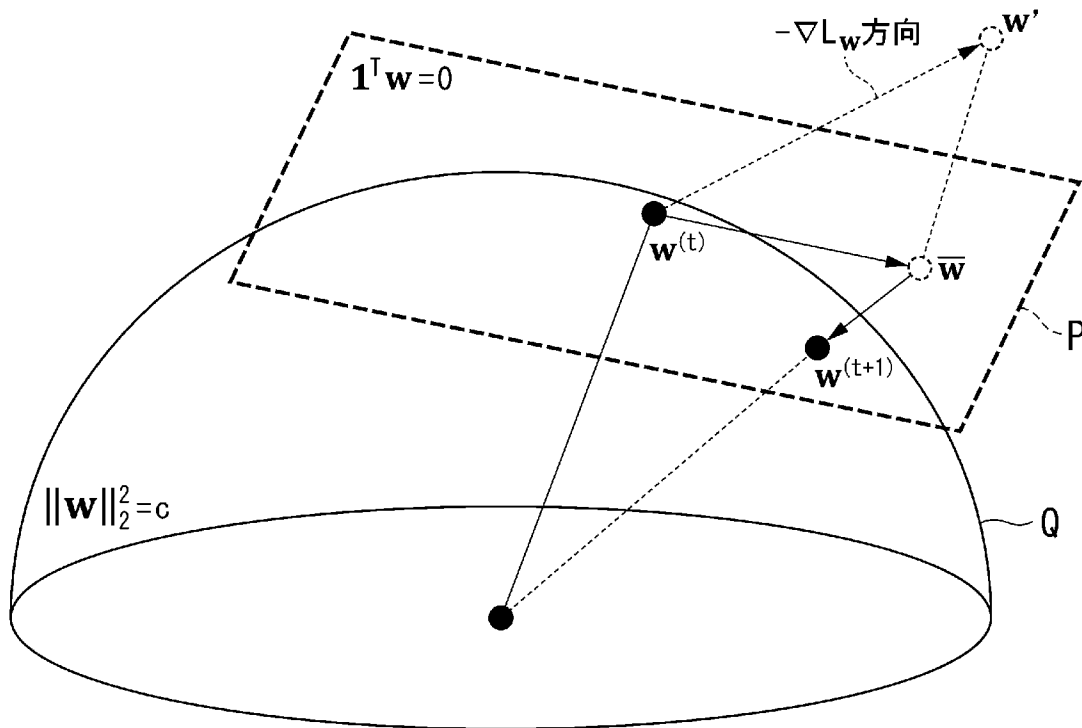
[図2]



[図3]



[図4]



[図5]

CIFAR-10の結果						
データ拡張	CNN	本実施形態	BN	GN	WS+BN	WS+GN
Standard	9層CNN	5.50	<u>5.36</u>	6.45	5.43	5.71
	ResNet	5.65	5.68	6.34	<u>5.52</u>	6.69
	WideResNet	4.19	<u>3.70</u>	5.01	4.78	4.67
FastAA	9層CNN	<u>4.18</u>	4.40	4.93	4.65	4.37
	ResNet	4.26	4.11	6.76	<u>3.92</u>	5.85
	WideResNet	<u>2.59</u>	2.64	3.68	3.42	3.29

[図6]

CIFAR-100の結果						
データ拡張	CNN	本実施形態	BN	GN	WS+BN	WS+GN
Standard	9層CNN	<u>24.26</u>	25.11	29.28	25.51	26.27
	ResNet	<u>25.16</u>	26.28	29.75	25.55	28.51
	WideResNet	<u>18.31</u>	18.48	22.63	19.82	22.62
FastAA	9層CNN	<u>21.03</u>	22.23	25.34	21.87	22.31
	ResNet	21.87	<u>21.84</u>	28.90	22.10	27.83
	WideResNet	<u>16.40</u>	16.57	19.96	19.18	19.41

[図7]

本実施形態	BN	GN	WS+BN	WS+GN
91.4	91.7	89.7	92.0	91.2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/018449

A. CLASSIFICATION OF SUBJECT MATTER**G06N 3/08**(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 1971-2021
 Registered utility model specifications of Japan 1996-2021
 Published registered utility model applications of Japan 1994-2021

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	得居 誠也, 最適化から見たディープラーニングの考え方, 経営の科学 オペレーションズ・リサーチ, 07 April 2015 (accession date), vol. 60, no. 4, pp. 191-197 chapter 6, (TOKUI, Seiya. Communications of the Operations Research Society of Japan.), non-official translation (Thinking way of deep learning from the perspective of optimization)	1-7
A	WO 2020/144820 A1 (NEC CORP) 16 July 2020 (2020-07-16) paragraphs [0047]-[0057]	1-7
A	OYMAK, Samet. Learning Compact Neural Networks with Regularization. arxiv.org. June 2018, https://arxiv.org/abs/1802.01223 chapters 2-3	1-7
T	IKAMI, Daiki et al. Constrained Weight Optimization for Learning Without Activation Normalization. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). June 2021 (accession date), https://ieeexplore.ieee.org/document/9423444 chapters 1-3	1-7

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

25 June 2021

Date of mailing of the international search report

06 July 2021

Name and mailing address of the ISA/JP

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

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2021/018449

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2020/144820 A1	16 July 2020	(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 1971-2021年 日本国実用新案登録公報 1996-2021年 日本国登録実用新案公報 1994-2021年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	得居 誠也, 最適化から見たディープラーニングの考え方, 経営の科学 オペレーションズ・リサーチ, 2015.04.07 (受入日), 第60巻 第4号, 191~197頁 第6章	1-7
A	WO 2020/144820 A1 (日本電気株式会社) 16.07.2020 (2020-07-16) 0047-0057段落	1-7
A	Samet Oymak, Learning Compact Neural Networks with Regularization, arxiv.org, 2018.06, https://arxiv.org/abs/1802.01223 第2-3章	1-7
T	Daiki Ikami等, Constrained Weight Optimization for Learning Without Activation Normalization, 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021.06 (受入日), https://ieeexplore.ieee.org/document/9423444 第1-3章	1-7
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的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 “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 “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 25.06.2021		国際調査報告の発送日 06.07.2021
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 牛丸 太希 5B 6297 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/JP2021/018449

引用文献			公表日	パテントファミリー文献	公表日
WO	2020/144820	A1	16.07.2020	(ファミリーなし)	