

(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
2020年3月19日(19.03.2020)



(10) 国際公開番号

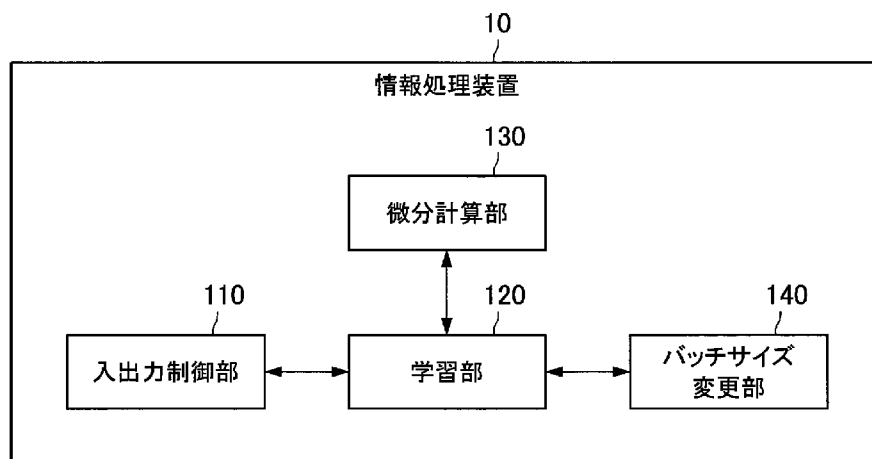
WO 2020/054083 A1

- (51) 国際特許分類:
G06N 3/08 (2006.01)
- (21) 国際出願番号: PCT/JP2018/042950
- (22) 国際出願日: 2018年11月21日(21.11.2018)
- (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:
特願 2018-169979 2018年9月11日(11.09.2018) JP
- (71) 出願人: ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 三上 裕明 (MIKAMI, Hiroaki); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング 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, JO, KE, KG, KH,

(54) Title: INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称: 情報処理装置および情報処理方法

[図3]



- 10 Information processing device
110 Input and output control unit
120 Learning unit
130 Differential calculation unit
140 Batch size changing unit

(57) Abstract: [Problem] To effectively achieve high-speed learning by DNN without depending on a learning method. [Solution] Provided is an information processing device provided with a learning unit for performing learning with the use of a neural network, wherein the learning unit dynamically changes a value of a batch size during learning on the basis of a gap value with an ideal state relating to learning output from the neural network. In addition, provided is an information processing method including learning with the use of a neural network by a processor, wherein the learning

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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, 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)）

further includes dynamically changing a value of a batch size during learning on the basis of a gap value with an ideal state relating to learning output from the neural network.

(57) 要約：【課題】DNNによる学習を学習手法に依らず効果的に高速化する。【解決手段】ニューラルネットワークを用いた学習を行う学習部、を備え、前記学習部は、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更する、情報処理装置が提供される。また、プロセッサが、ニューラルネットワークを用いた学習を行うこと、を含み、前記学習を行うことは、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更すること、をさらに含む、情報処理方法が提供される。

明 細 書

発明の名称： 情報処理装置および情報処理方法

技術分野

[0001] 本開示は、情報処理装置および情報処理方法に関する。

背景技術

[0002] 近年、脳神経系の仕組みを模した数学モデルであるニューラルネットワークが注目されている。また、ニューラルネットワークによる学習を高速化するための技術も多く提案されている。例えば、非特許文献1には、学習中にバッチサイズを変更する技術が開示されている。

先行技術文献

非特許文献

[0003] 非特許文献1：Samuel L. Smith、外3名、「Don't Decay the Learning Rate , Increase the Batch Size」、2017年11月1日、[Online]、[平成30年9月7日検索]、インターネット<<https://arxiv.org/pdf/1711.00489.pdf>>

発明の概要

発明が解決しようとする課題

[0004] しかし、非特許文献1に記載の技術は、特定の学習手法に依存しており、当該手法を採用しない学習には適用することが困難である。

課題を解決するための手段

[0005] 本開示によれば、ニューラルネットワークを用いた学習を行う学習部、を備え、前記学習部は、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更する、情報処理装置が提供される。

[0006] また、本開示によれば、プロセッサが、ニューラルネットワークを用いた学習を行うこと、を含み、前記学習を行うことは、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバ

ッチサイズの値を動的に変更すること、をさらに含む、情報処理方法が提供される。

図面の簡単な説明

[0007] [図1] `Step Learning rate decay`を適用した場合の損失の推移の一例を示す図である。

[図2] 本開示の一実施形態に係るバッチサイズ変更の概要について説明するための図である。

[図3] 同実施形態に係る情報処理装置の機能構成例を示すブロック図である。

[図4] 同実施形態に係る損失の傾きに基づくバッチサイズ変更を `ImageNet / ResNet-50` に適用した際の検証結果を示す図である。

[図5] 同実施形態に係るトレーニングの値に基づくバッチサイズ変更を `ImageNet / ResNet-50` に適用した際の検証結果を示す図である。

[図6] 同実施形態に係る損失に基づくバッチサイズ変更を `MNIST` を用いた学習に適用した際の検証結果を示す図である。

[図7] 同実施形態に係る損失に基づくバッチサイズ変更を `cifar10` を用いた学習に適用した際の検証結果を示す図である。

[図8] 同実施形態に係る損失に基づくバッチサイズ変更を `cifar10` を用いた学習に適用した際の検証結果を示す図である。

[図9] 同実施形態に係る損失の1回微分値に基づくバッチサイズの変更を実現する訓練スクリプトおよび損失傾き計算モジュールの一例を示す図である。

[図10] 同実施形態に係るエポックごとのバッチサイズ増加を `MNIST` を用いた学習に適用した場合の検証結果を示す図である。

[図11] 同実施形態に係る損失およびエポックに基づくバッチサイズ変化を `cifar10` を用いた学習に適用した場合の検証結果を示す図である。

[図12] 同実施形態に係る損失とエポックに基づくバッチサイズの増減を実現する訓練スクリプトの一例を示す図である。

[図13] 同実施形態に係るバッチサイズ変更部によるGPU中モデルの作り直しについて説明するための図である。

[図14]同実施形態に係るバッチサイズ変更部による計算ループ数の増減制御について説明するための図である。

[図15]同実施形態に係るバッチサイズ変更部による利用GPUの増減制御について説明するための図である。

[図16]同実施形態に係るバッチサイズ変更部による制御の流れを示すフローチャートである。

[図17]本開示の一実施形態に係るハードウェア構成例を示す図である。

発明を実施するための形態

[0008] 以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

[0009] なお、説明は以下の順序で行うものとする。

1. 実施形態

1. 1. 概要

1. 2. 情報処理装置10の機能構成例

1. 3. 検証結果

1. 4. バッチサイズ増減の実現手法

2. ハードウェア構成例

3. まとめ

[0010] <1. 実施形態>

<<1. 1. 概要>>

まず、本開示の一実施形態の概要について説明する。上述したように、近年、ニューラルネットワークによる学習を高速化するための技術が多く提案されている。一般に、DNN (Deep Neural Network) の学習に要する時間は、パラメータの更新回数と比例関係にあることから、当該更新回数を減らすことが学習の高速化に対し有効な手段となり得る。

[0011] パラメータの更新回数は、例えば、バッチサイズを増加させることで減ら

することが可能である。また、学習の後半においては、バッチサイズを増加させても学習が収束することが知られていることから、例えば、非特許文献１に開示されるように学習中にバッチサイズを変更し、できるだけ大きなバッチサイズを設定することで、パラメータの更新回数を削減することができ、ひいては学習を高速化させる効果が期待される。

[0012] しかし、非特許文献１に記載されるバッチサイズの変更手法は、特定の学習手法にのみ適用可能な技術である。ここで、上記の特定の学習手法とは、`Step Learning rate decay`と呼ばれる手法を指す。

[0013] 図１は、`Step Learning rate decay`を適用した場合の損失（`loss`）の推移の一例を示す図である。`Step Learning rate decay`とは、図１に示すように、学習率を階段状に下げること、損失を階段状に下げていく手法である。図１に示す一例を参照すると、エポック３０および６０付近において損失が大きく低下し、グラフが階段形状をなしていることがわかる。

[0014] 非特許文献１に記載の技術によれば、損失が大きく低下するエポック３０や６０のようなタイミングでバッチサイズを変更することが可能であるが、損失の推移が上記のような階段形状を示さない学習手法には適用することができない。

[0015] 本開示に係る技術思想は上記の点に着目して発想されたものであり、ＤＮＮによる学習を学習手法に依らず効果的に高速化することを可能とする。このために、本開示の一実施形態に係る情報処理装置１０は、ニューラルネットワークを用いた学習を行う学習部１２０を備え、学習部１２０は、ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更すること、を特徴の一つとする。

[0016] ここで、上記の理想状態とのギャップ値とは、期待される出力と実際の出力との差を定量的に表した指標であってよい。本実施形態に係る理想状態とのギャップ値には、例えば、損失が含まれる。また、本実施形態に係る理想

状態とのギャップ値は、トレーニングエラーやバリデーションエラーを含むうる。

[0017] なお、上記ギャップ値として用いられるトレーニングエラーおよびバリデーションエラーの一例としては、例えば、損失として用いられることもある平均二乗誤差（MSE : Mean Square Error）や平均絶対誤差（MAE : Mean Absolute Error）、画像分類において用いられるTop-k-error（特に、top-1-errorやtop-5-errorなど）、また物体検出において用いられるmAP（mean Average Precision）などが挙げられる。

[0018] ここで、図2を参照して、本実施形態に係るバッチサイズ変更の概要について説明する。図2には、エポックの経過に伴う損失の推移を表すグラフが示されている。なお、図2以降に示す各グラフにおいては、実線がバッチサイズ変更を伴わない損失の推移（Reference）を、破線が本実施形態に係るバッチサイズ変更を適用した損失の推移（Approach）をそれぞれ示している。

[0019] 本実施形態に係る学習部120は、例えば、損失に基づいて学習の収束が推定される場合、学習中にバッチサイズの値を増加させてよい。

[0020] 損失の値が小さくなることは、DNNが解に近づいていること、すなわち学習が収束に向かっていること（学習が安定していること）を示す。このことから、本実施形態に係る学習部120は、損失のn回微分値に基づいて、学習中にバッチサイズの値を増加させてもよい。

[0021] 例えば、本実施形態に係る学習部120は、損失の1回微分値、すなわち傾きが所定の閾値を下回る場合、バッチサイズの値を増加させることができる。図2に示す一例の場合、学習部120は、損失の傾きが落ち着いたタイミングT1（エポック30）でバッチサイズの値を32Kから64Kに増加させている。

[0022] また、例えば、本実施形態に係る学習部120は、損失の0回微分値、すなわち損失の値そのものが所定の閾値を下回る場合、バッチサイズの値を増

加させることもできる。ここで、上記の閾値が0.3である場合、学習部120は、損失の値が0.3を下回ったタイミングT2（エポック60）でバッチサイズの値を増加させてよい。なお、学習部120は、 $n > 2$ 以上の n 回微分値に基づいてバッチサイズの値を増加させてもよい。

[0023] ここで、ApproachとReferenceを比較すると、本実施形態に係るバッチサイズ変更手法を適用した場合であっても、学習が発散せずに性能を保っていることがわかる。すなわち、本実施形態に係る情報処理装置10が実現するバッチ変更手法によれば、学習性能の確保と、パラメータ更新回数の削減すなわち学習時間の短縮と、を両立することが可能となる。

[0024] また、本実施形態に係るバッチ変更手法によれば、図2に示すような、損失の推移が階段形状を示さない学習手法であっても、バッチサイズを増加させ、学習時間を短縮することが可能となる。このように、本実施形態に係る情報処理装置10によれば、DNNによる学習を学習手法に依らず効果的に高速化することが可能となる。

[0025] <<1. 2. 情報処理装置10の機能構成例>>

次に、本実施形態に係る情報処理装置10の機能構成例について説明する。図3は、本実施形態に係る情報処理装置10の機能構成例を示すブロック図である。図3を参照すると、本実施形態に係る情報処理装置10は、入出力制御部110、学習部120、微分計算部130、およびバッチサイズ変更部140を備える。

[0026] （入出力制御部110）

本実施形態に係る入出力制御部110は、DNNの学習に係るユーザインタフェースを制御する。例えば、本実施形態に係る入出力制御部110は、入力装置を介して入力された各種のデータを学習部120に引き渡す。また、例えば、入出力制御部110は、学習部120が出力する値を出力装置に引き渡す。

[0027] （学習部120）

本実施形態に係る学習部120は、DNNを用いた学習を行う。上述した

ように、本実施形態に係る学習部120は、DNNが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズを動的に変更すること、を特徴の一つとする。本実施形態に係る理想状態とのギャップ値は、例えば、損失、トレーニングエラー、バリデーションエラーなどを含む。

[0028] (微分計算部130)

本実施形態に係る微分計算部130は、学習部120から入力される損失にn回微分処理を行うことでn回微分値を算出し、当該n回微分値を学習部120に出力する。

[0029] (バッチサイズ変更部140)

本実施形態に係るバッチサイズ変更部140は、学習部120が設定したバッチサイズの値に基づいて、バッチサイズの増減を制御する機能を有する。本実施形態に係るバッチサイズ変更部140が有する機能の詳細については、別途後述する。

[0030] 以上、本実施形態に係る情報処理装置10の機能構成例について説明した。なお、図3を用いて説明した上記の構成はあくまで一例であり、本実施形態に係る情報処理装置10の機能構成は係る例に限定されない。本実施形態に係る情報処理装置10の機能構成は、仕様や運用に応じて柔軟に変形可能である。

[0031] <<1. 3. 検証結果>>

次に、本実施形態に係る情報処理装置10により実現されるバッチサイズ変更手法の検証結果について述べる。

[0032] まず、データセットにImageNetを、DNNにResNet-50を用いた場合の検証結果について説明する。図4は、本実施形態に係る損失の傾きに基づくバッチサイズ変更をImageNet/ResNet-50に適用した際の検証結果を示す図である。

[0033] ここでは、Referenceにおけるバッチサイズを32Kで固定し、学習を行った。一方、Approachにおいては、損失の1回微分値、す

なわち傾きが閾値を下回ったタイミングT3（エポック30）において、バッチサイズを32Kから68Kに増加させ学習を継続させた。

[0034] ReferenceとApproachを比較すると、本実施形態に係るバッチ変更手法によりバッチサイズを増加させた場合であっても、損失の収束に大きな影響を与えないことがわかる。

[0035] また、図5は、本実施形態に係るトレーニングの値に基づくバッチサイズ変更をImageNet/ResNet-50に適用した際の検証結果を示す図である。

[0036] ここでは、トレーニングエラーの0回微分値が閾値1.8を下回ったタイミングT4（エポック30）において、バッチサイズを2Kから20Kに増加させ学習を継続させた。

[0037] 図5を参照すると、トレーニングエラーの0回微分値に基づいてバッチサイズを増加させた場合であっても、影響なく学習が収束に向かっていることがわかる。

[0038] 次に、データセットにMNISTを用いた場合の検証結果について説明する。図6は、本実施形態に係る損失に基づくバッチサイズ変更をMNISTを用いた学習に適用した際の検証結果を示す図である。

[0039] ここでは、Referenceにおけるバッチサイズを128で固定し、学習を行った。一方、Approachにおいては、損失の1回微分値が閾値を下回り、かつ損失の0回微分値が閾値0.03を下回ったタイミングT5（エポック1）において、バッチサイズを128から3072に増加させ学習を継続させた。

[0040] 上記の制御の結果、パラメータの更新回数を2000回から560回に削減することができ、学習時間を大幅に短縮することができた。

[0041] 次に、データセットにcifar10を用いた場合の検証結果について説明する。図7および図8は、本実施形態に係る損失に基づくバッチサイズ変更をcifar10を用いた学習に適用した際の検証結果を示す図である。

[0042] 図7に係る検証では、Referenceにおけるバッチサイズを64で

固定し、学習を行った。一方、*Approach*においては、損失の1回微分値が閾値を下回り、かつ損失の0回微分値が閾値0.35を下回ったタイミングT6（エポック5）において、バッチサイズを64から1024に増加させ学習を継続させた。

[0043] 上記の制御の結果、パラメータの更新回数を20000回から5000回に削減することができ、学習時間を大幅に短縮することができた。

[0044] また、図8に係る検証では、図7に係る検証と同様に、*Reference*におけるバッチサイズを64で固定し、学習を行った。一方、*Approach*においては、損失の0回微分値が閾値0.35を下回ったタイミングT7（エポック8）において、バッチサイズを64から1024に増加させ学習を継続させた。

[0045] 上記の制御の結果、パラメータの更新回数を20000回から7250回に削減することができ、学習時間を大幅に短縮することができた。

[0046] 以上、本実施形態に係るバッチサイズ変更手法の検証結果について述べた。上述した検証結果では、本実施形態に係るバッチサイズ変更手法を適用した場合、性能にほぼ影響を与えずに、パラメータ更新回数を1/3～1/4程度削減できることが示されている。このように、本実施形態に係る情報処理装置10によれば、DNNによる学習を学習手法に依らず効果的に高速化することが可能となる。

[0047] なお、損失の1回微分値に基づくバッチサイズの変更は、例えば、図9に示すような訓練スクリプトTS1および損失傾き計算モジュールCMにより実現することができる。なお、図9においては、コードを擬似的に示している。

[0048] 図9に示す一例の場合、訓練スクリプトTS1では、まず、損失傾き取得APIすなわち損失傾き計算モジュールCMの呼び出し処理を実行し、返り値として現在の`loss_grad`の値が取得される。

[0049] 次に、取得された`loss_grad`の値と閾値との比較処理が実行され、ここで、`loss_grad`の値が閾値を下回る場合、バッチサイズの増

加処理が実行される。

[0050] 訓練スクリプトTS1では、学習が収束するまで、上記の各処理が繰り返し実行される。

[0051] また、損失傾き計算モジュールCMは、訓練スクリプトTS1により呼び出されると、保持するlossの値をloss_prevに退避し、新たに取得したlossとloss_prevとの差を求めることで、loss_gradを算出する。この際、損失傾き計算モジュールCMは、図示するように、損失の移動平均をとりノイズを除去する処理を行なってもよい。

[0052] なお、上記では、バッチサイズを増加させる場合を主な例に説明したが、本実施形態に係る学習部120は、損失に基づいて学習の発散が推定される場合には、学習中バッチサイズの値を減少させることも可能である。

[0053] 例えば、図5に示した一例の場合、期間D1においては学習がまだ不安定であることから、学習部120は、初期値として与えられた小さいバッチサイズの値を維持する。一方、期間D2において学習が安定した場合、学習部120はバッチサイズの値を増加させてよい。

[0054] しかし、Step learning rate decayでは、一般に学習率を下げた直後に大きく損失が下がることから、学習率を下げる前と比較して学習が発散しやすくなることが想定される。このため、本実施形態に係る学習部120は、期間D2において一度増加させたバッチサイズの値を期間D3においては減少させることで学習の収束を図ることができる。この際、学習部120は、例えば期間D1および期間D2の間となるようなバッチサイズの値を設定してよい。

[0055] 次に、本実施形態に係るエポックに基づくバッチサイズの変更について説明する。DNNによる学習では、学習率の減衰がない場合、学習が進むにつれ、すなわちエポックが重なるにつれ、学習が容易となる傾向が強く見られる。このため、本実施形態に係る学習部120は、エポックの経過に伴いバッチサイズの値を増加させることができる。例えば、本実施形態に係る学習部120は、エポックごとにバッチサイズの値を増加させてもよい。

- [0056] 図10は、本実施形態に係るエポックごとのバッチサイズ増加をMNISTを用いた学習に適用した場合の検証結果を示す図である。ここでは、バッチサイズの初期値として128を設定し、エポック1（タイミングT8）でバッチサイズを256に、エポック2（タイミングT9）でバッチサイズを512に、エポック3（タイミングT10）でバッチサイズを1024に、それぞれ倍増させる制御を行った。
- [0057] 上記の制御の結果、パラメータの更新回数を2000回から938に削減することができた。係る検証結果によれば、エポックごとにバッチサイズを増加させた場合であっても、損失の収束に大きな影響を与えず、学習時間を大幅に削減できることがわかる。
- [0058] また、本実施形態に係る学習部120は、損失やエポックに基づいてバッチサイズの値を増加させた結果、学習の発散が推定された場合には、発散前、すなわち直前のエポックにおけるネットワークモデルを再読み込みすることで、学習の収束を図ることも可能である。
- [0059] 図11は、本実施形態に係る損失およびエポックに基づくバッチサイズ変化をcifar10を用いた学習に適用した場合の検証結果を示す図である。
- [0060] ここでは、バッチサイズの初期値として64を設定し、損失の0回微分値の閾値を0.35に設定した。図11に示す一例の場合、エポック8（タイミングT11）において損失が閾値0.35を下回ったことに基づいてバッチサイズの増加処理を開始し、その後エポックごとにバッチサイズを増加させた。
- [0061] その後、エポック14（タイミングT12）においてバッチサイズを4Kに増加させたところ、学習の発散が推定された。このため、本実施形態に係る学習部120は、エポック15の開始時点でバッチサイズの値の増加を停止し、エポック14開始時点におけるモデルを再読み込みしたうえで、バッチサイズの値を2Kに固定し学習を継続した。
- [0062] このように、本実施形態に係る学習部120は、過去のエポックにおける

ネットワークモデルの再読み込みを行った場合、当該過去のエポックで設定した値よりも小さいバッチサイズの値を設定してよい。

[0063] 本実施形態に係る学習部120が有する上記の機能によれば、損失やエポックに基づいて自動でバッチサイズの値を増減することができ、学習の発散を回避しながらパラメータの更新回数を効果的に削減することが可能となる。

[0064] なお、上記のような損失とエポックに基づくバッチサイズの増減は、例えば、図12に示すような訓練スクリプトTS2より実現することができる。なお、図12においては、コードを擬似的に示している。

[0065] 図12に示す一例の場合、訓練スクリプトTS2は、まず、図9に示した損失傾き計算モジュールCMを呼び出し戻り値として取得した`loss_grad`を、閾値と比較している。ここで、`loss_grad`が閾値を下回る場合、訓練スクリプトTS2は、バッチサイズの自動増加を開始する。

[0066] その後、訓練スクリプトTS2は、損失が前エポックよりも閾値以上大きくなったか否かを判定する。ここで、損失の増大が認められる場合、訓練スクリプトTS2は、バッチサイズの自動増加を停止する。

[0067] また、この際、訓練スクリプトTS2は、前エポックにおけるDNNのネットワークモデルを再読み込みする。

[0068] <<1. 4. バッチサイズ増減の実現手法>>

続いて、本実施形態に係るバッチサイズ増減の実現手法について詳細に説明する。本実施形態に係るバッチサイズ変更部140は、学習部120が設定したバッチサイズの値を取得し、当該値に基づいてGPU (Graphics Processing Unit) を制御することで、バッチサイズの増減を実現する。

[0069] 例えば、本実施形態に係るバッチサイズ変更部140は、GPU中のモデルを作り直すことによりバッチサイズの増減を制御してもよい。

[0070] 図13は、本実施形態に係るバッチサイズ変更部140によるGPU中モデルの作り直しについて説明するための図である。この場合、まず、学習部

120が損失の値を微分計算部130に入力し、当該値のn回部分値を取得する。また、学習部120は、取得したn回部分値に基づいて変更後のバッチサイズの値を決定し、当該バッチサイズの値をバッチサイズ変更部140に入力する。

[0071] 次に、本実施形態に係るバッチサイズ変更部140は、入力されたバッチサイズの値に基づいて、現在学習に利用されているGPUにモデルの再作成指示を行う。

[0072] なお、図13では、GPU__0およびGPU__1のうち、GPU__0が現在学習に用いられており、GPU__0中のモデルのバッチサイズが32である場合において、バッチサイズ変更部140がGPU__0にモデルの再作成を指示し、当該モデルのバッチサイズを64に変更させる場合の一例が示されている。

[0073] 本実施形態に係るバッチサイズ変更部140による上記の制御によれば、情報処理装置10が有するGPUの数に影響されずグローバルにバッチサイズを上げることができ、またGPUの並列演算能力が活きることでさらなる高速化に繋がる効果も記載される。

[0074] また、例えば、本実施形態に係るバッチサイズ変更部140は、学習に係る計算のループ数を増減させることによりバッチサイズの増減を制御してもよい。上記のようなテクニックは、`accum-grad`とも称される。

[0075] 図14は、本実施形態に係るバッチサイズ変更部140による計算ループ数の増減制御について説明するための図である。この場合、本実施形態に係るバッチサイズ変更部140は、学習部120から入力されたバッチサイズの値に基づいて、現在学習に利用されているGPUに計算ループ数を変更するよう指示する。

[0076] なお、図14では、GPU__0およびGPU__1のうち、GPU__0が現在学習に用いられており、GPU__0中のモデルのバッチサイズが32である場合において、バッチサイズ変更部140がGPU__0に2回`accum-grad`を行うよう指示し、バッチサイズ32による学習が2回行われる

場合の一例が示されている。

[0077] 本実施形態に係るバッチサイズ変更部140による上記の制御によれば、GPUの数やメモリ容量に制限されることなくバッチサイズを上げることができ、また同期処理の回数が減るため、減った同期処理の回数分だけ学習を高速化することができる。

[0078] また、例えば、本実施形態に係るバッチサイズ変更部140は、学習に用いられるGPUの数を増減させることによりバッチサイズの増減を制御してもよい。

[0079] 図15は、本実施形態に係るバッチサイズ変更部140による利用GPUの増減制御について説明するための図である。この場合、本実施形態に係るバッチサイズ変更部140は、学習部120から入力されたバッチサイズの値に基づいて、現在学習に利用されていないGPUに稼働を指示する。

[0080] なお、図15では、GPU__0およびGPU__1のうち、GPU__0のみが現在学習に用いられている場合において、バッチサイズ変更部140がGPU__1に稼働を指示する場合の一例が示されている。

[0081] 本実施形態に係るバッチサイズ変更部140による上記の制御によれば、計算資源を増やすことで、その分学習を高速化することができる。

[0082] 以上、本実施形態に係るバッチサイズ変更部140によるバッチサイズの変更制御手法について説明した。なお、本実施形態に係るバッチサイズ変更部140は、例えば、図16に示すような優先度に基づいて、バッチサイズの変更制御手法を選択することで、バッチサイズ増加による高速化効果をより大きくすることができる。

[0083] 図16は、本実施形態に係るバッチサイズ変更部140による制御の流れを示すフローチャートである。

[0084] 図16を参照すると、バッチサイズ変更部140は、まず、追加で利用可能なGPUが存在するか否かを判定する(S1101)。

[0085] ここで、追加で利用可能なGPUが存在する場合(S1101: Yes)、バッチサイズ変更部140は、当該利用可能なGPUを学習に割り当てる

ことで、バッチサイズの増加を制御する（S 1 1 0 2）。

[0086] 続いて、バッチサイズ変更部 1 4 0 は、ステップ S 1 1 0 2 における処理により目的のバッチサイズを達成したかを判定する（S 1 1 0 3）。

[0087] ここで、目的のバッチサイズが達成されている場合（S 1 1 0 3 : Y e s）、バッチサイズ変更部 1 4 0 は、バッチサイズ変更に係る処理を終了する。

[0088] 一方、目的のバッチサイズが達成されていない場合（S 1 1 0 3 : N o）、または追加で利用可能な G P U が存在しない場合（S 1 1 0 1 : N o）、バッチサイズ変更部 1 4 0 は、現在利用中の G P U のメモリに空き容量が存在するか否かを判定する（S 1 1 0 4）。

[0089] ここで、現在利用中の G P U のメモリに空き容量が存在する場合（S 1 1 0 4 : Y e s）、バッチサイズ変更部 1 4 0 は、現在利用中の G P U 中のモデルを作り直すことで、バッチサイズの増加を制御する（S 1 1 0 5）。

[0090] 続いて、バッチサイズ変更部 1 4 0 は、ステップ S 1 1 0 5 における処理により目的のバッチサイズを達成したかを判定する（S 1 1 0 6）。

[0091] ここで、目的のバッチサイズが達成されている場合（S 1 1 0 6 : Y e s）、バッチサイズ変更部 1 4 0 は、バッチサイズ変更に係る処理を終了する。

[0092] 一方、目的のバッチサイズが達成されていない場合（S 1 1 0 6 : N o）、または現在利用中の G P U のメモリに空き容量が存在しない場合（S 1 1 0 4 : N o）、バッチサイズ変更部 1 4 0 は、学習に係る計算のループ数を増加させることで、バッチサイズの増加を制御し（S 1 1 0 7）、バッチサイズ変更に係る処理を終了する。

[0093] < 2. ハードウェア構成例 >

次に、本開示の一実施形態に係る情報処理装置 1 0 のハードウェア構成例について説明する。図 1 7 は、本開示の一実施形態に係る情報処理装置 1 0 のハードウェア構成例を示すブロック図である。図 1 7 を参照すると、情報処理装置 1 0 は、例えば、プロセッサ 8 7 1 と、ROM 8 7 2 と、RAM 8

７３と、ホストバス８７４と、ブリッジ８７５と、外部バス８７６と、インターフェース８７７と、入力装置８７８と、出力装置８７９と、ストレージ８８０と、ドライブ８８１と、接続ポート８８２と、通信装置８８３と、を有する。なお、ここで示すハードウェア構成は一例であり、構成要素の一部が省略されてもよい。また、ここで示される構成要素以外の構成要素をさらに含んでもよい。

[0094] （プロセッサ８７１）

プロセッサ８７１は、例えば、演算処理装置又は制御装置として機能し、ROM８７２、RAM８７３、ストレージ８８０、又はリムーバブル記録媒体９０１に記録された各種プログラムに基づいて各構成要素の動作全般又はその一部を制御する。プロセッサ８７１は、例えば、GPUやCPUを含む。なお、本開示の一実施形態に係る情報処理装置１０は、少なくとも２つのGPUを備える。

[0095] （ROM８７２、RAM８７３）

ROM８７２は、プロセッサ８７１に読み込まれるプログラムや演算に用いるデータ等を格納する手段である。RAM８７３には、例えば、プロセッサ８７１に読み込まれるプログラムや、そのプログラムを実行する際に適宜変化する各種パラメータ等が一時的又は永続的に格納される。

[0096] （ホストバス８７４、ブリッジ８７５、外部バス８７６、インターフェース８７７）

プロセッサ８７１、ROM８７２、RAM８７３は、例えば、高速なデータ伝送が可能なホストバス８７４を介して相互に接続される。一方、ホストバス８７４は、例えば、ブリッジ８７５を介して比較的データ伝送速度が低速な外部バス８７６に接続される。また、外部バス８７６は、インターフェース８７７を介して種々の構成要素と接続される。

[0097] （入力装置８７８）

入力装置８７８には、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ、及びレバー等が用いられる。さらに、入力装置８７８として

は、赤外線やその他の電波を利用して制御信号を送信することが可能なリモートコントローラ（以下、リモコン）が用いられることもある。また、入力装置 878 には、マイクロフォンなどの音声入力装置が含まれる。

[0098] （出力装置 879）

出力装置 879 は、例えば、CRT（Cathode Ray Tube）、LCD、又は有機 EL 等のディスプレイ装置、スピーカ、ヘッドホン等のオーディオ出力装置、プリンタ、携帯電話、又はファクシミリ等、取得した情報を利用者に対して視覚的又は聴覚的に通知することが可能な装置である。また、本開示に係る出力装置 879 は、触覚刺激を出力することが可能な種々の振動デバイスを含む。

[0099] （ストレージ 880）

ストレージ 880 は、各種のデータを格納するための装置である。ストレージ 880 としては、例えば、ハードディスクドライブ（HDD）等の磁気記憶デバイス、半導体記憶デバイス、光記憶デバイス、又は光磁気記憶デバイス等が用いられる。

[0100] （ドライブ 881）

ドライブ 881 は、例えば、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリ等のリムーバブル記録媒体 901 に記録された情報を読み出し、又はリムーバブル記録媒体 901 に情報を書き込む装置である。

[0101] （リムーバブル記録媒体 901）

リムーバブル記録媒体 901 は、例えば、DVD メディア、Blu-ray（登録商標）メディア、HD DVD メディア、各種の半導体記憶メディア等である。もちろん、リムーバブル記録媒体 901 は、例えば、非接触型 IC チップを搭載した IC カード、又は電子機器等であってもよい。

[0102] （接続ポート 882）

接続ポート 882 は、例えば、USB（Universal Serial Bus）ポート、IEEE1394 ポート、SCSI（Small Computer System Interface）、RS-232C ポ

ート、又は光オーディオ端子等のような外部接続機器 902 を接続するためのポートである。

[0103] (外部接続機器 902)

外部接続機器 902 は、例えば、プリンタ、携帯音楽プレーヤ、デジタルカメラ、デジタルビデオカメラ、又は IC レコーダ等である。

[0104] (通信装置 883)

通信装置 883 は、ネットワークに接続するための通信デバイスであり、例えば、有線又は無線 LAN、Bluetooth (登録商標)、又は WUSB (Wireless USB) 用の通信カード、光通信用のルータ、ADSL (Asymmetric Digital Subscriber Line) 用のルータ、又は各種通信用のモデム等である。

[0105] <3. まとめ>

以上説明したように、本開示の一実施形態に係る情報処理装置 10 は、ニューラルネットワークを用いた学習を行う学習部 120 を備え、学習部 120 は、ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズ of 値を動的に変更すること、を特徴の一つとする。係る構成によれば、DNN による学習を学習手法に依らず効果的に高速化することが可能となる。

[0106] 以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

[0107] また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

[0108] また、コンピュータに内蔵されるCPU、ROMおよびRAMなどのハードウェアに、情報処理装置10が有する構成と同等の機能を発揮させるためのプログラムも作成可能であり、当該プログラムを記録した、コンピュータに読み取り可能な非一過性の記録媒体も提供され得る。

[0109] また、本明細書の情報処理装置10の処理に係る各ステップは、必ずしもフローチャートに記載された順序に沿って時系列に処理される必要はない。例えば、情報処理装置10の処理に係る各ステップは、フローチャートに記載された順序と異なる順序で処理されても、並列的に処理されてもよい。

[0110] なお、以下のような構成も本開示の技術的範囲に属する。

(1)

ニューラルネットワークを用いた学習を行う学習部、
を備え、

前記学習部は、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更する、
情報処理装置。

(2)

前記理想状態とのギャップ値は、少なくとも損失を含む、
前記(1)に記載の情報処理装置。

(3)

前記学習部は、前記損失に基づいて学習の収束が推定される場合、学習中にバッチサイズの値を増加させる、
前記(2)に記載の情報処理装置。

(4)

前記学習部は、前記損失の n 回微分値に基づいて、学習中にバッチサイズの値を増加させる、
前記(3)に記載の情報処理装置。

(5)

前記学習部は、前記損失の値または前記損失の傾きの少なくともいずれかが閾値を下回ることに基づいて、学習中にバッチサイズの値を増加させる、前記（４）に記載の情報処理装置。

（６）

前記学習部は、前記損失に基づいて学習の発散が推定される場合、学習中にバッチサイズの値を減少させる、前記（２）～（５）のいずれかに記載の情報処理装置。

（７）

前記学習部は、エポックに基づいて、バッチサイズの値を動的に変更する、前記（１）～（６）のいずれかに記載の情報処理装置。

（８）

前記学習部は、エポックの経過に伴いバッチサイズの値を増加させる、前記（７）に記載の情報処理装置。

（９）

前記学習部は、前記理想状態とのギャップ値から学習の発散が推定される場合、過去のエポックにおけるネットワークモデルを再読み込みする、前記（８）に記載の情報処理装置。

（１０）

前記学習部は、前記過去のエポックにおけるネットワークモデルの再読み込みを行った場合、前記過去のエポックで設定した値よりも小さいバッチサイズの値を設定する、前記（９）に記載の情報処理装置。

（１１）

前記学習部は、エポックごとにバッチサイズの値を増加させる、前記（７）に記載の情報処理装置。

（１２）

前記学習部が設定した値に基づいて、バッチサイズの増減を制御するバツ

チサイズ変更部、

をさらに備える、

前記（１）～（１１）のいずれかに記載の情報処理装置。

（１３）

前記バッチサイズ変更部は、GPU中のモデルを作り直すことによりバッチサイズの増減を制御する、

前記（１２）に記載の情報処理装置。

（１４）

前記バッチサイズ変更部は、学習に係る計算のループ数を増減させることによりバッチサイズの増減を制御する、

前記（１２）または（１３）に記載の情報処理装置。

（１５）

前記バッチサイズ変更部は、学習に用いられるGPUの数を増減させることによりバッチサイズの増減を制御する、

前記（１２）～（１４）のいずれかに記載の情報処理装置。

（１６）

前記バッチサイズ変更部は、追加で利用可能なGPUが存在する場合、当該GPUを学習に割り当てることで、バッチサイズの増加を制御する、

前記（１２）～（１５）のいずれかに記載の情報処理装置。

（１７）

前記バッチサイズ変更部は、追加で利用可能なGPUが存在せず、かつ現在利用中のGPUのメモリに空き容量が存在する場合、現在利用中のGPU中のモデルを作り直すことで、バッチサイズを増加を制御する、

前記（１２）～（１６）のいずれかに記載の情報処理装置。

（１８）

前記バッチサイズ変更部は、現在利用中のGPUのメモリに空き容量が存在しない場合、学習に係る計算のループ数を増加させることで、バッチサイズの増加を制御する、

前記（１２）～（１７）のいずれかに記載の情報処理装置。

（１９）

前記理想状態とのギャップ値は、トレーニングエラーまたはバリデーションエラーのうち少なくともいずれかを含む、

前記（１）に記載の情報処理装置。

（２０）

プロセッサが、ニューラルネットワークを用いた学習を行うこと、
を含み、

前記学習を行うことは、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズを動的に変更すること、

をさらに含む、

情報処理方法。

符号の説明

[0111]	１０	情報処理装置
	１１０	入出力制御部
	１２０	学習部
	１３０	微分計算部
	１４０	バッチサイズ変更部

請求の範囲

- [請求項1] ニューラルネットワークを用いた学習を行う学習部、
を備え、
前記学習部は、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更する、
情報処理装置。
- [請求項2] 前記理想状態とのギャップ値は、少なくとも損失を含む、
請求項1に記載の情報処理装置。
- [請求項3] 前記学習部は、前記損失に基づいて学習の収束が推定される場合、
学習中にバッチサイズの値を増加させる、
請求項2に記載の情報処理装置。
- [請求項4] 前記学習部は、前記損失の n 回微分値に基づいて、学習中にバッチサイズの値を増加させる、
請求項3に記載の情報処理装置。
- [請求項5] 前記学習部は、前記損失の値または前記損失の傾きの少なくともいずれかが閾値を下回ることに基づいて、学習中にバッチサイズの値を増加させる、
請求項4に記載の情報処理装置。
- [請求項6] 前記学習部は、前記損失に基づいて学習の発散が推定される場合、
学習中にバッチサイズの値を減少させる、
請求項2に記載の情報処理装置。
- [請求項7] 前記学習部は、エポックに基づいて、バッチサイズの値を動的に変更する、
請求項1に記載の情報処理装置。
- [請求項8] 前記学習部は、エポックの経過に伴いバッチサイズの値を増加させる、
請求項7に記載の情報処理装置。

- [請求項9] 前記学習部は、前記理想状態とのギャップ値から学習の発散が推定される場合、過去のエポックにおけるネットワークモデルを再読み込みする、
請求項8に記載の情報処理装置。
- [請求項10] 前記学習部は、前記過去のエポックにおけるネットワークモデルの再読み込みを行った場合、前記過去のエポックで設定した値よりも小さいバッチサイズの値を設定する、
請求項9に記載の情報処理装置。
- [請求項11] 前記学習部は、エポックごとにバッチサイズの値を増加させる、
請求項7に記載の情報処理装置。
- [請求項12] 前記学習部が設定した値に基づいて、バッチサイズの増減を制御するバッチサイズ変更部、
をさらに備える、
請求項1に記載の情報処理装置。
- [請求項13] 前記バッチサイズ変更部は、GPU中のモデルを作り直すことによりバッチサイズの増減を制御する、
請求項12に記載の情報処理装置。
- [請求項14] 前記バッチサイズ変更部は、学習に係る計算のループ数を増減させることによりバッチサイズの増減を制御する、
請求項12に記載の情報処理装置。
- [請求項15] 前記バッチサイズ変更部は、学習に用いられるGPUの数を増減させることによりバッチサイズの増減を制御する、
請求項12に記載の情報処理装置。
- [請求項16] 前記バッチサイズ変更部は、追加で利用可能なGPUが存在する場合、当該GPUを学習に割り当てることで、バッチサイズの増加を制御する、
請求項12に記載の情報処理装置。
- [請求項17] 前記バッチサイズ変更部は、追加で利用可能なGPUが存在せず、

かつ現在利用中のGPUのメモリに空き容量が存在する場合、現在利用中のGPU中のモデルを作り直すことで、バッチサイズを増加を制御する、

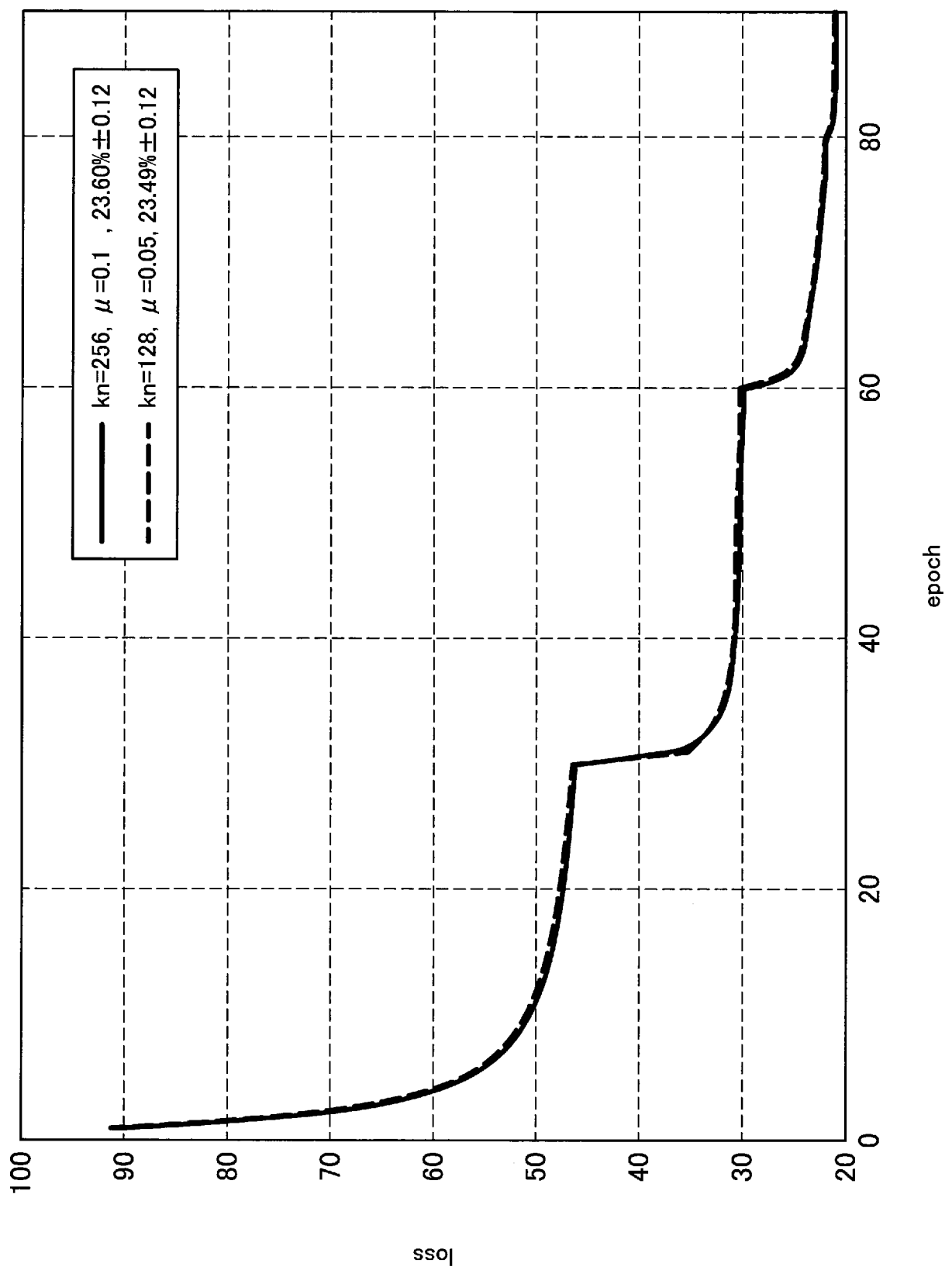
請求項12に記載の情報処理装置。

[請求項18] 前記バッチサイズ変更部は、現在利用中のGPUのメモリに空き容量が存在しない場合、学習に係る計算のループ数を増加させることで、バッチサイズの増加を制御する、
請求項12に記載の情報処理装置。

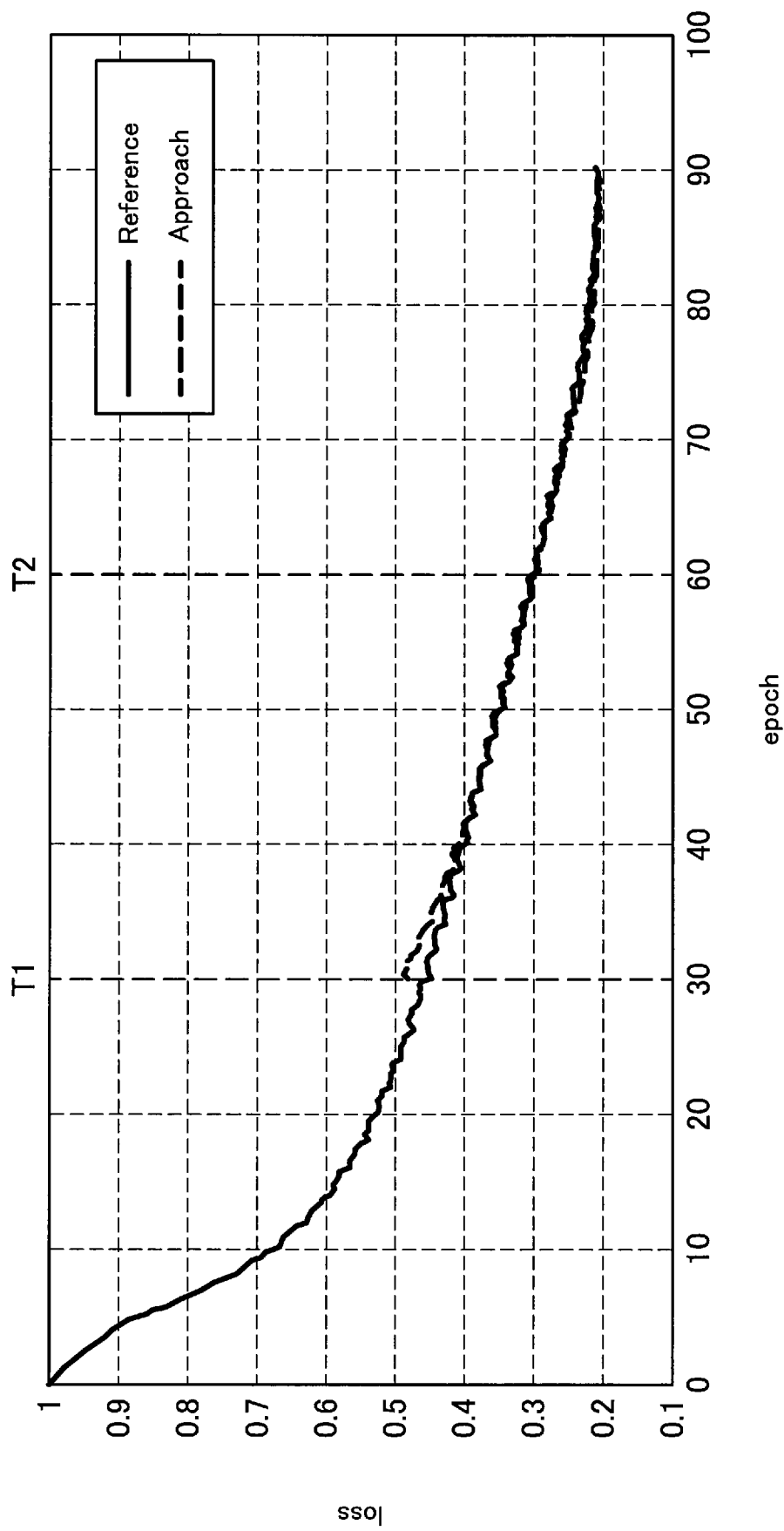
[請求項19] 前記理想状態とのギャップ値は、トレーニングエラーまたはバリデーションエラーのうち少なくともいずれかを含む、
請求項1に記載の情報処理装置。

[請求項20] プロセッサが、ニューラルネットワークを用いた学習を行うこと、
を含み、
前記学習を行うことは、前記ニューラルネットワークが出力する学習に係る理想状態とのギャップ値に基づいて、学習中にバッチサイズの値を動的に変更すること、
をさらに含む、
情報処理方法。

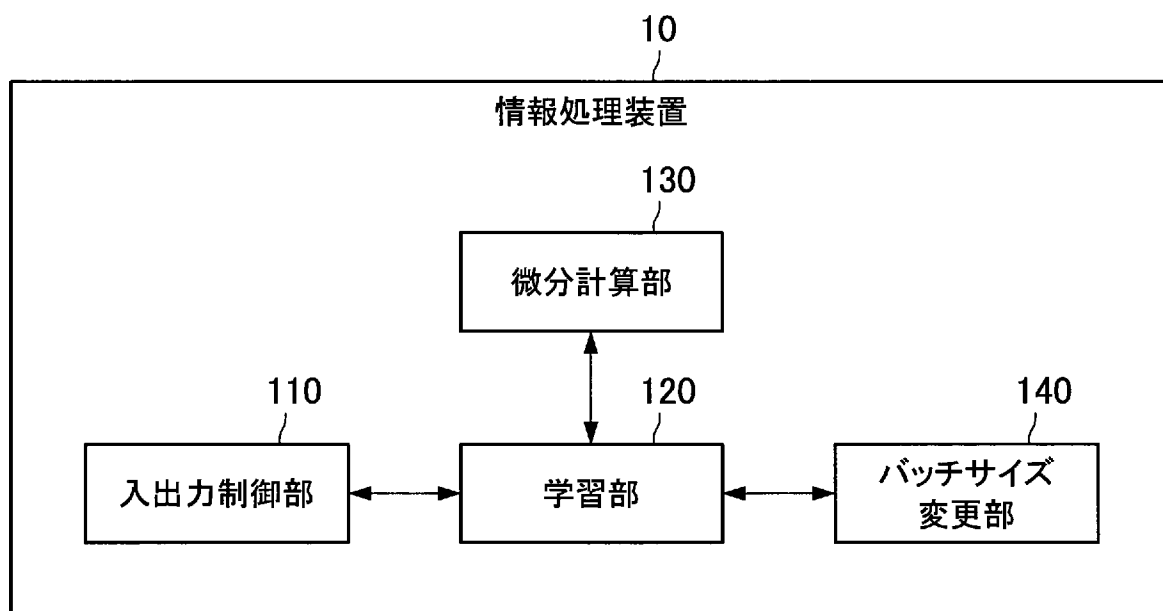
[図1]



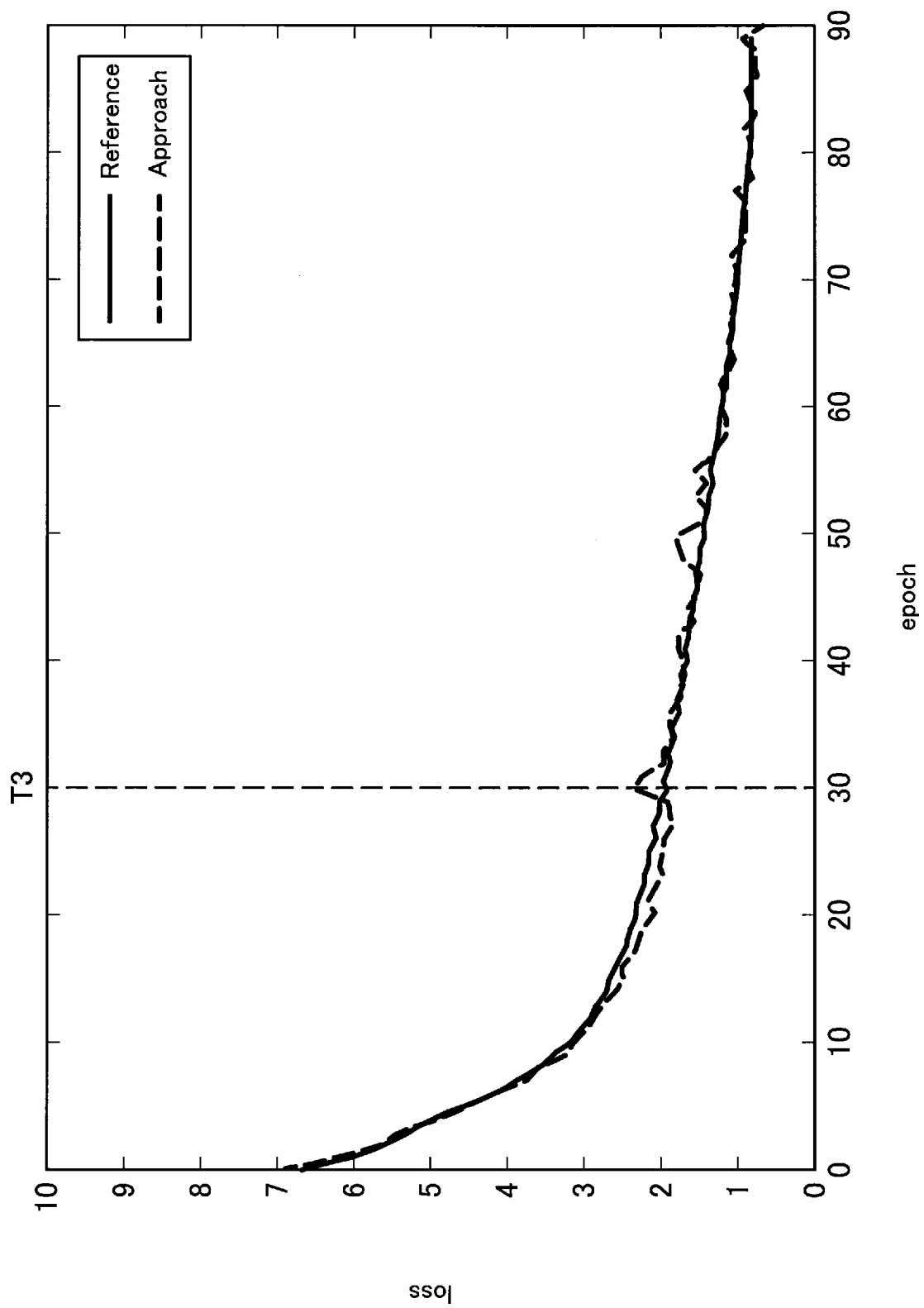
[図2]



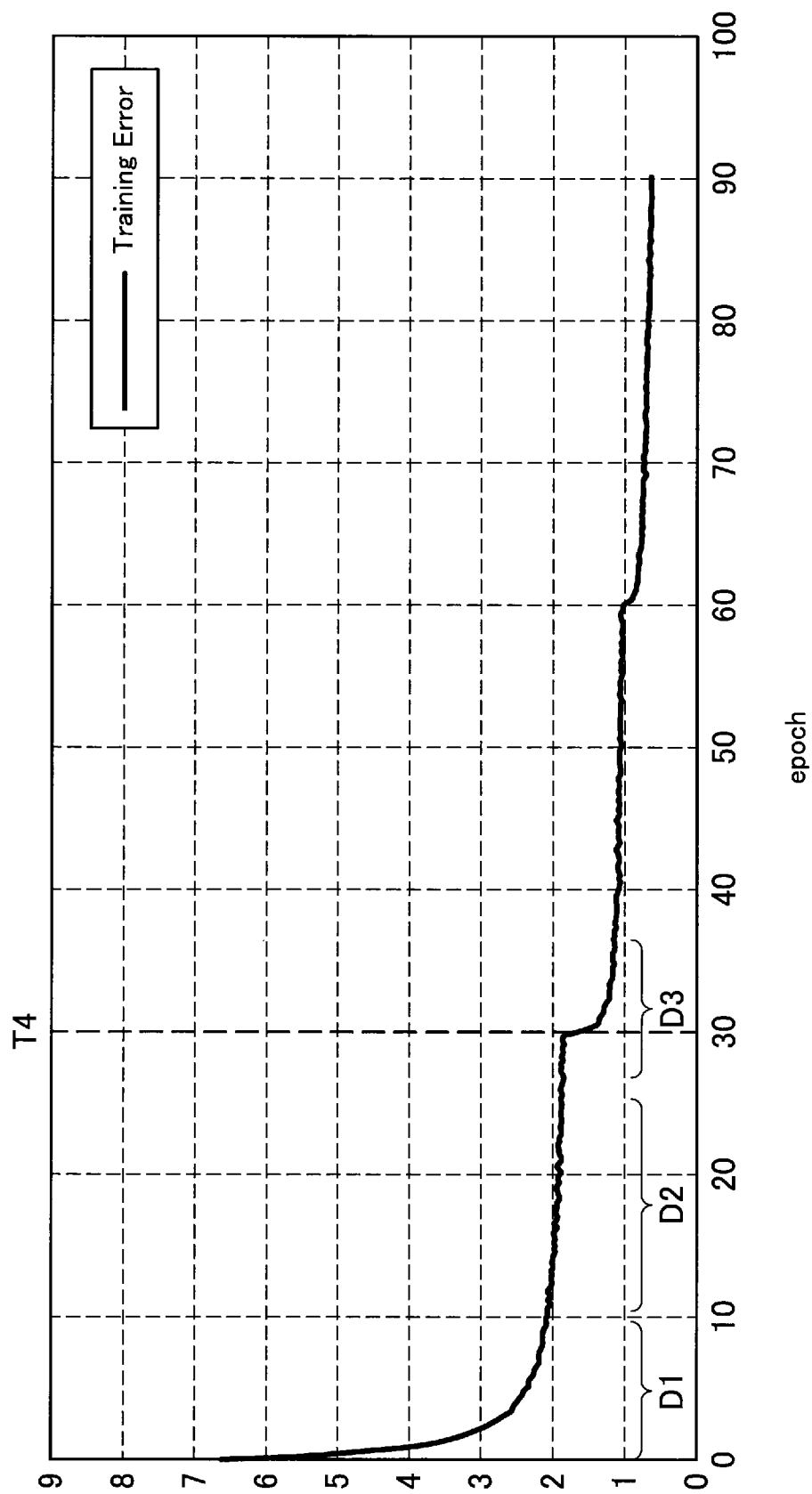
[図3]



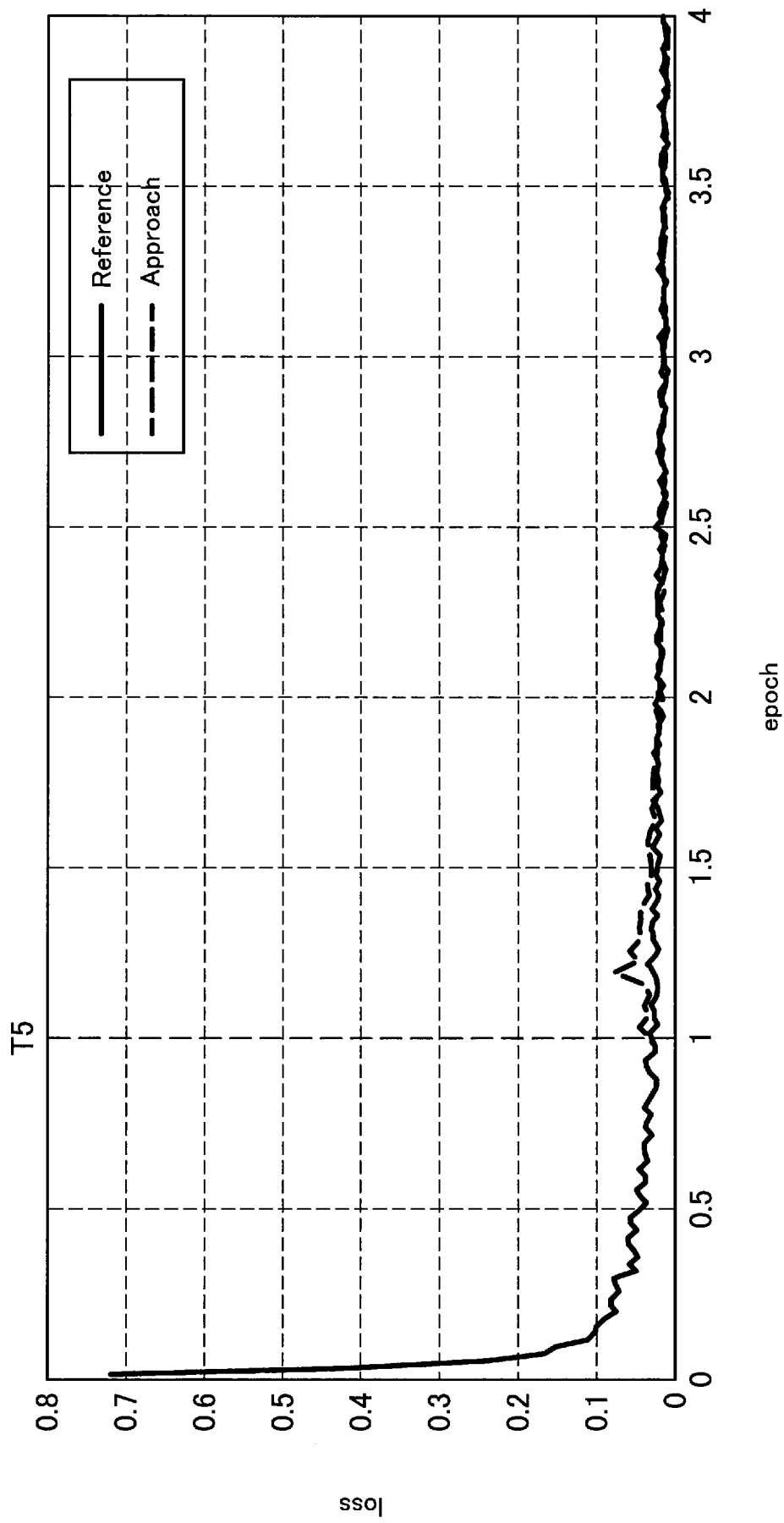
[図4]



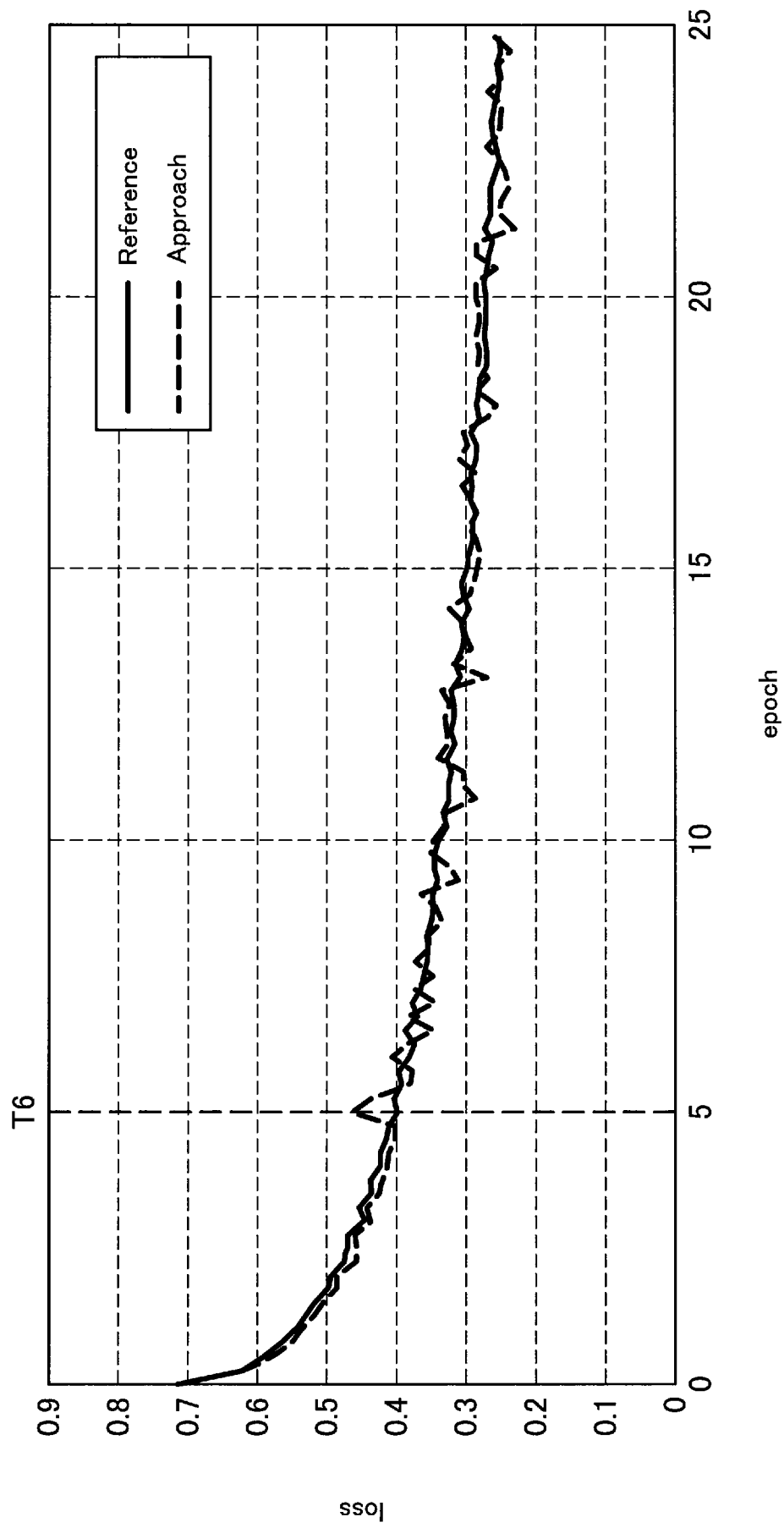
[図5]



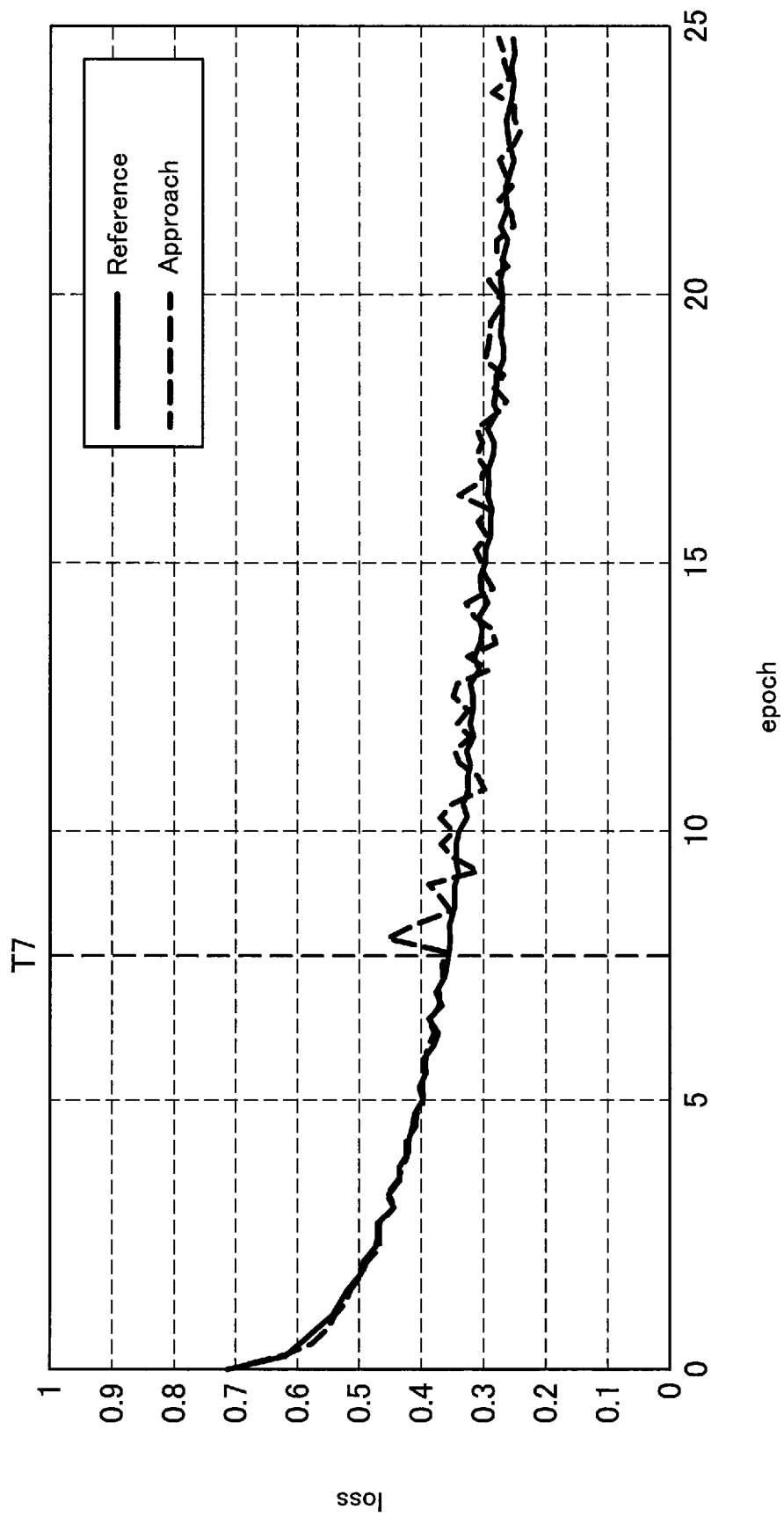
[図6]



[図7]



[図8]



[図9]

訓練スクリプト

```
while 収束するまで
  // 1iterの学習処理
  ...
  if 損失傾き取得API() < threshold:
    バッチサイズの増加
  ...
```

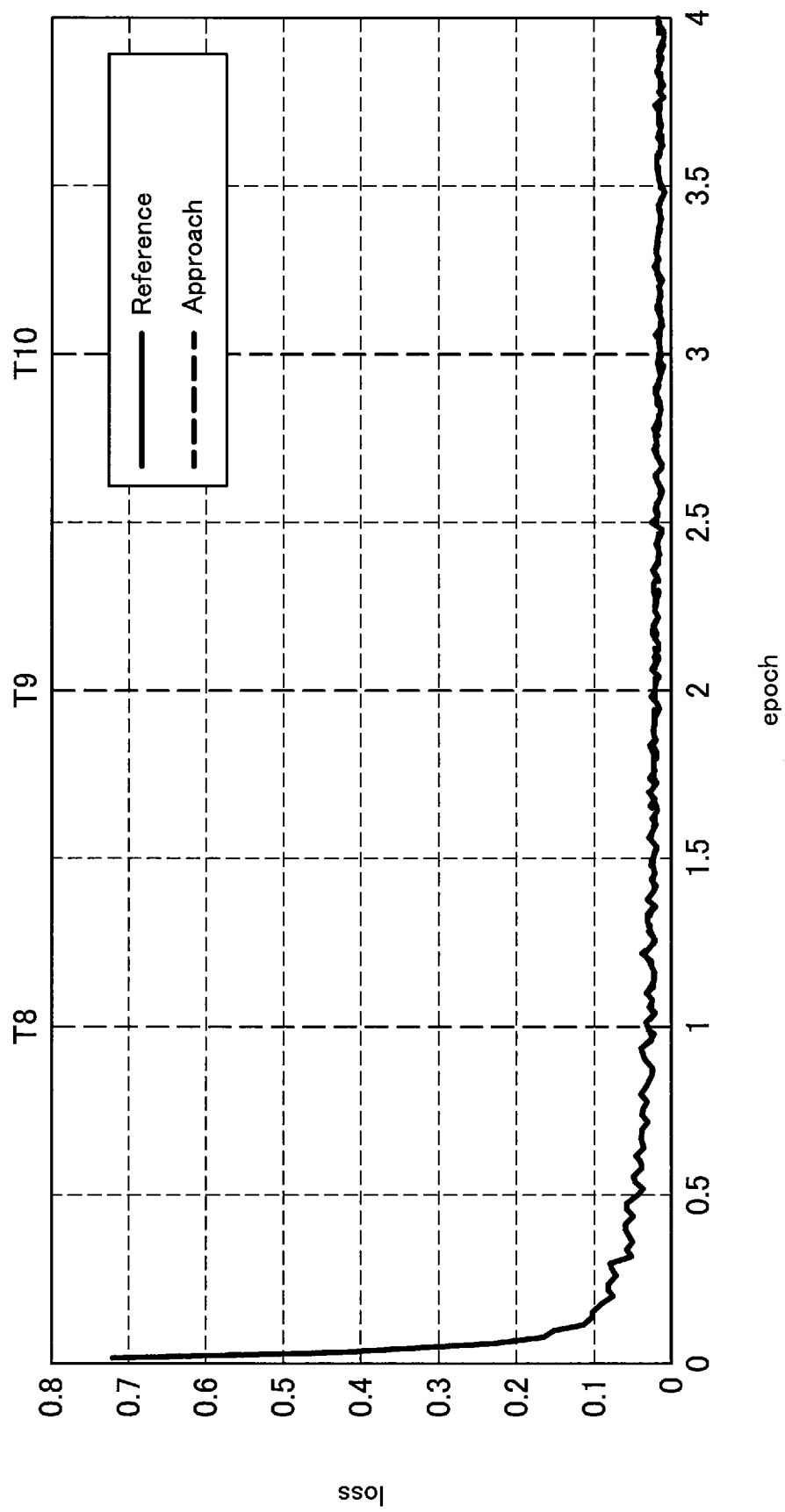
TS1

損失の傾きの計算モジュール

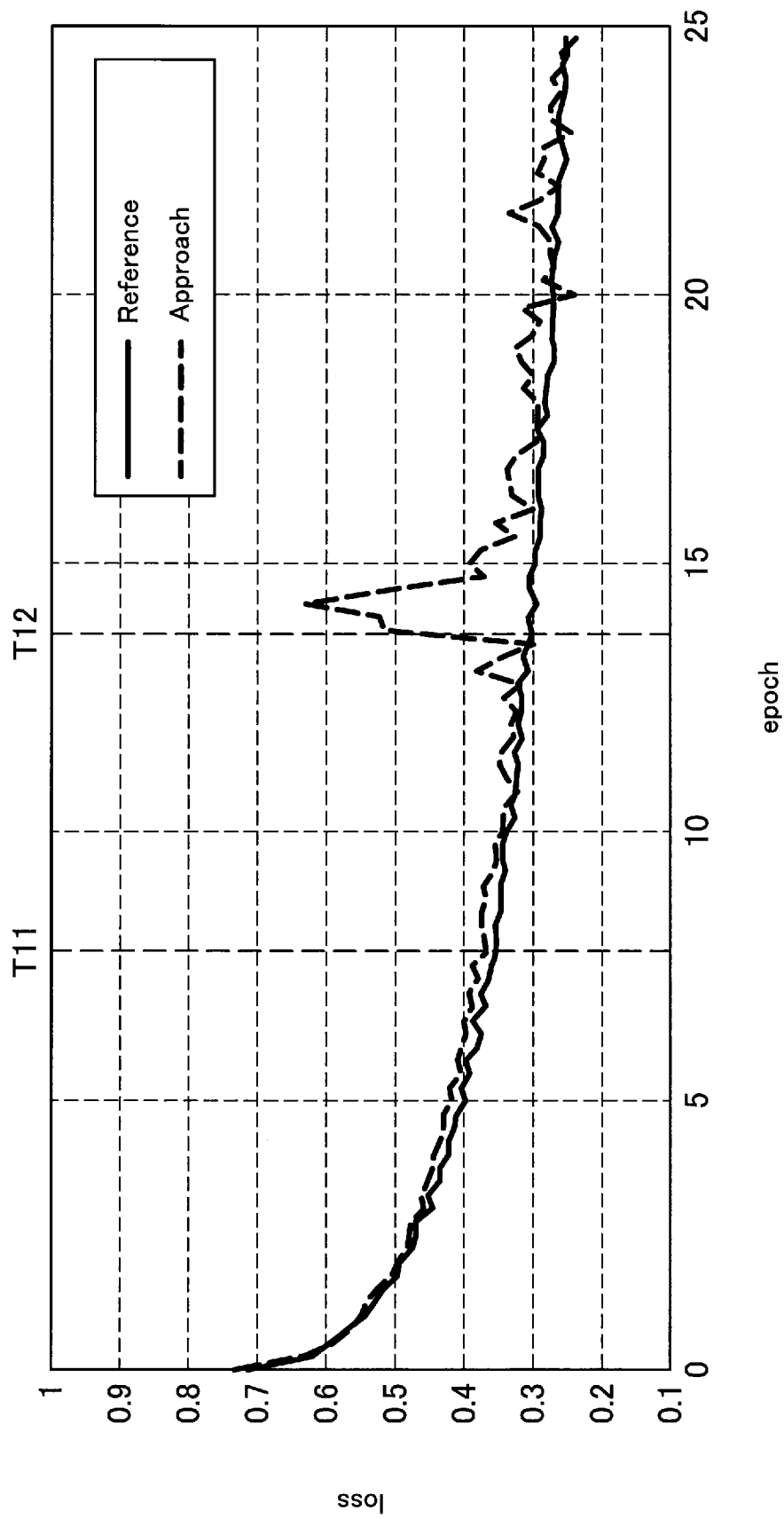
```
loss_prev = loss
loss = 0.9 * loss_prev + 0.1 * <現在のlossの値>
  //損失の移動平均をとりノイズを除去(必須ではない)
loss_grad = loss - loss_prev // 傾きを計算
```

CM

[図10]



[図11]

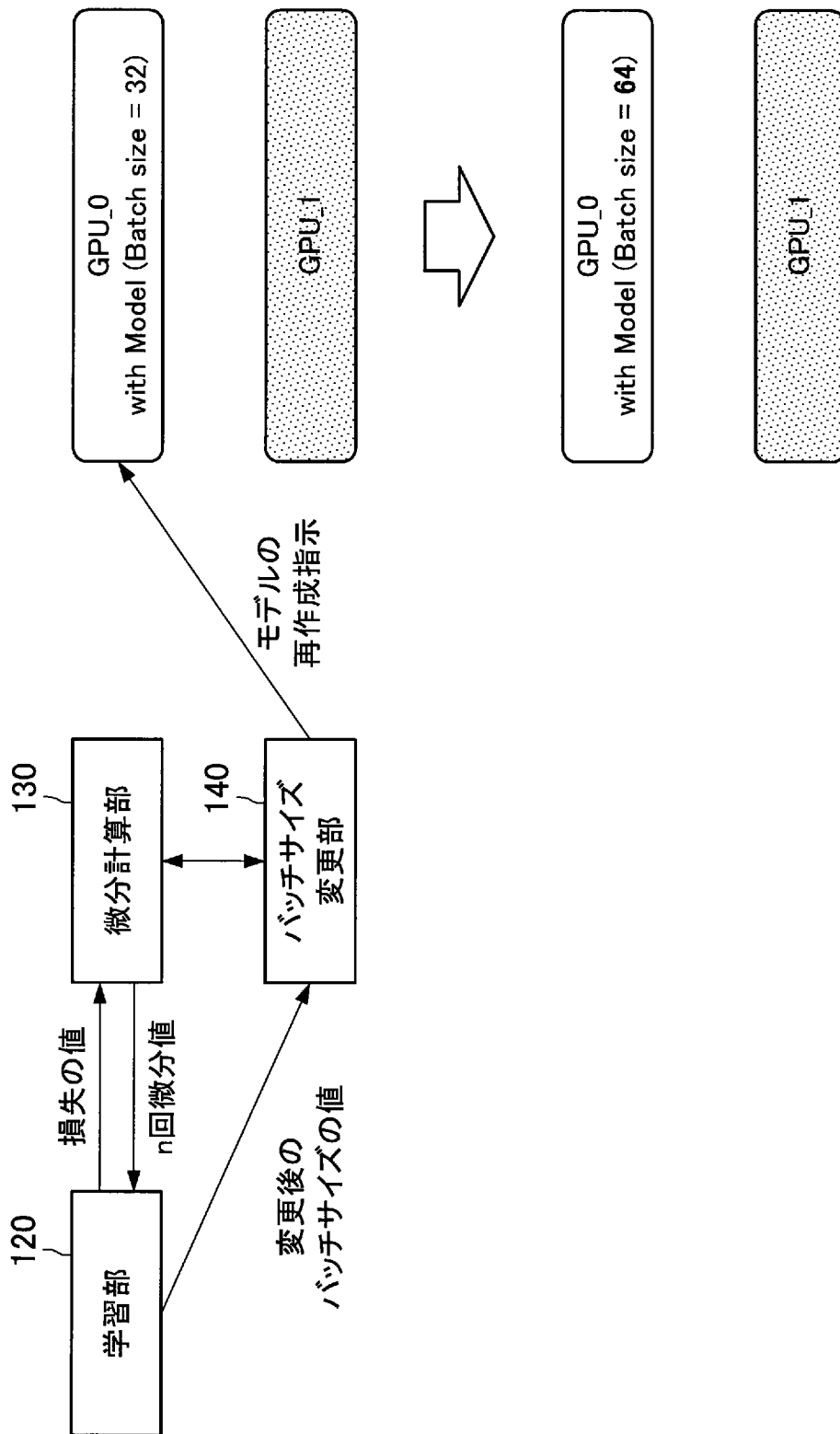


[図12]

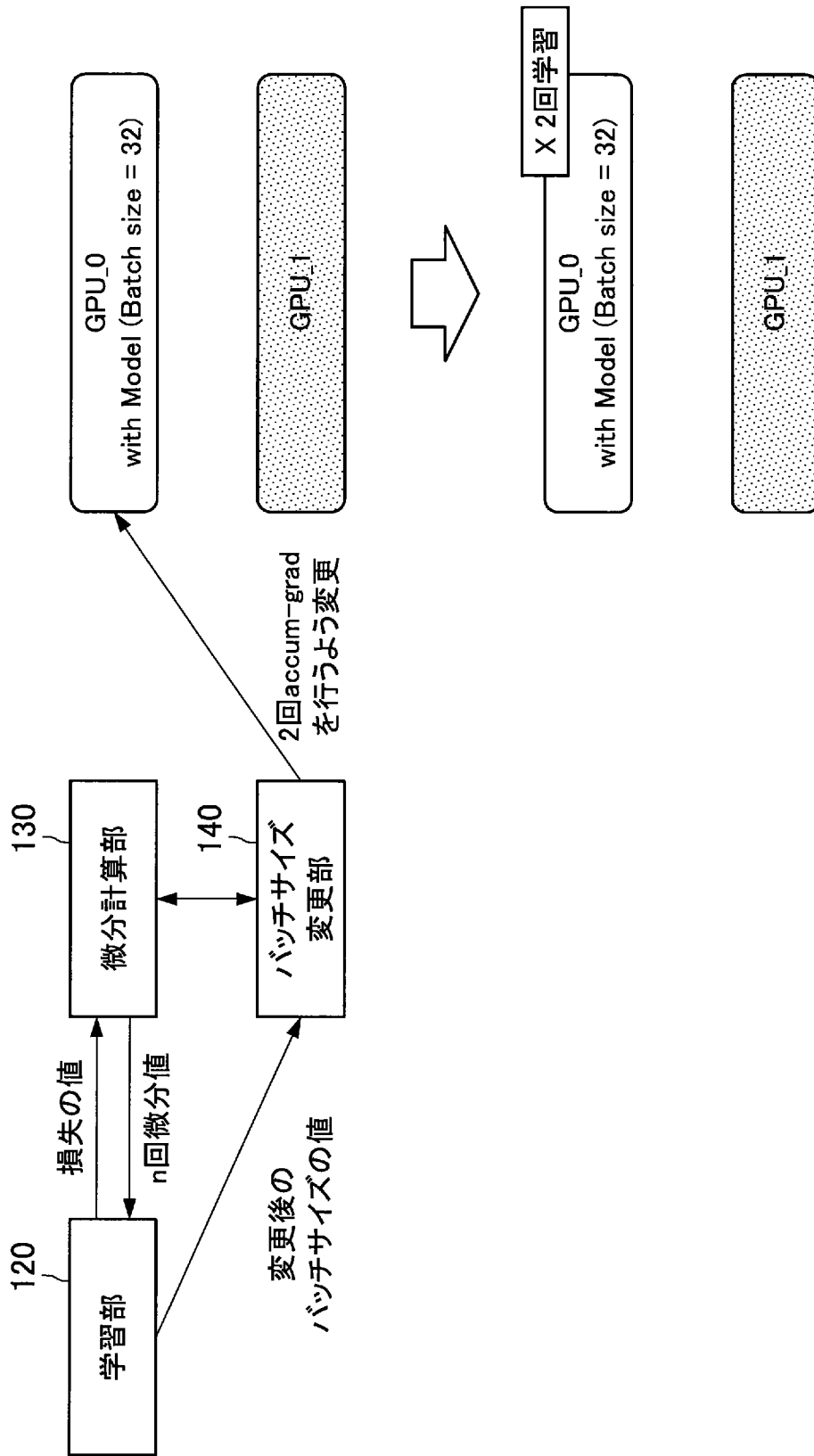
```
訓練スクリプト
while 収束するまで
  // 1エポックの学習処理
  ...
  if 損失傾き取得API() < threshold1: // 傾きが一定以下ならバッチサイズの自動増加を開始
    if ロス取得API() - loss_prev < threshold2: // 損失が大きくなった場合は自動増加を停止
      バッチサイズ /= 2
      前エポックにおけるDNNのモデルを再読み込み
    バッチサイズ *= 2
  loss_prev = 現在のlossの値
```

—TS2

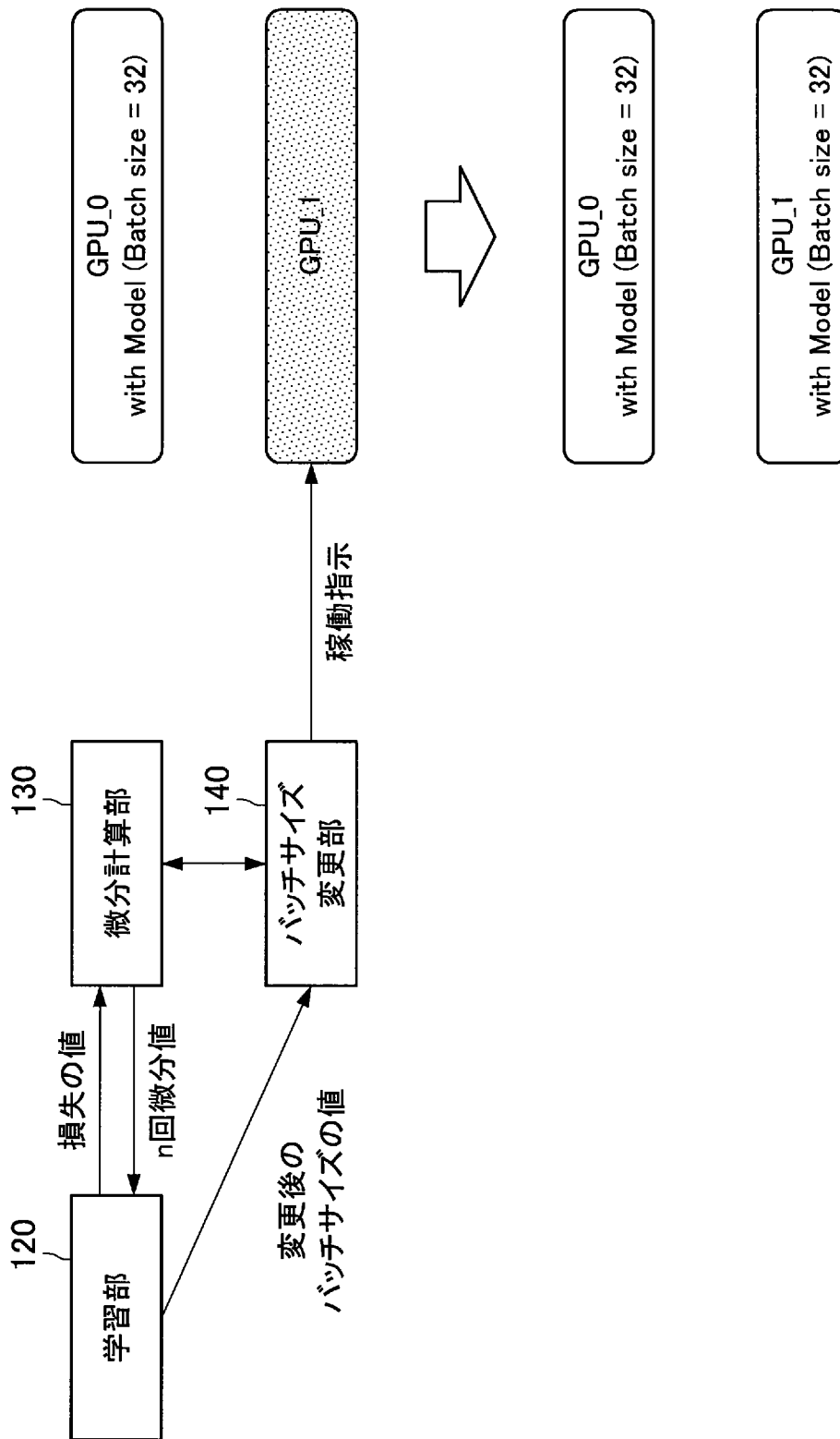
[図13]



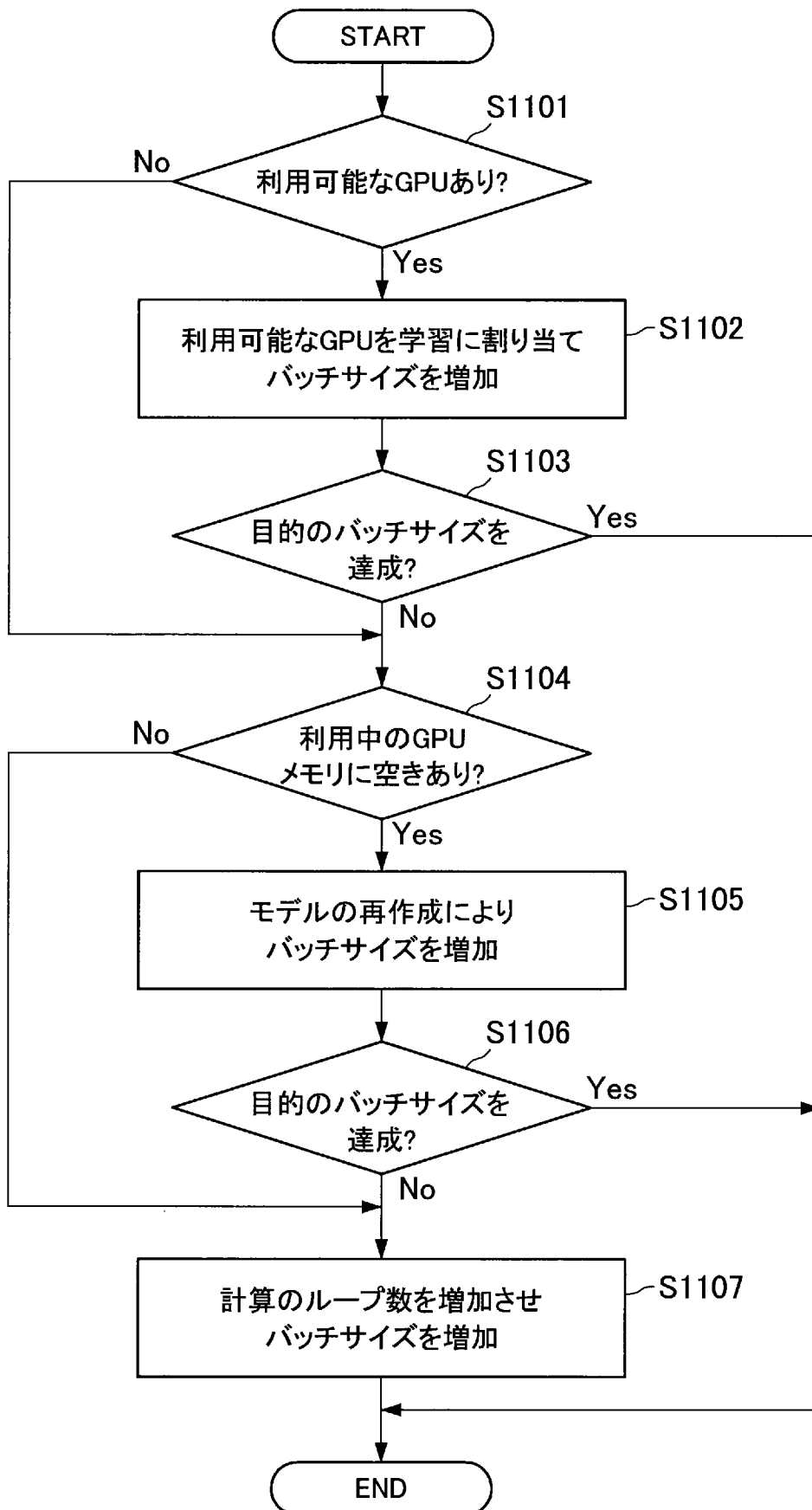
[図14]



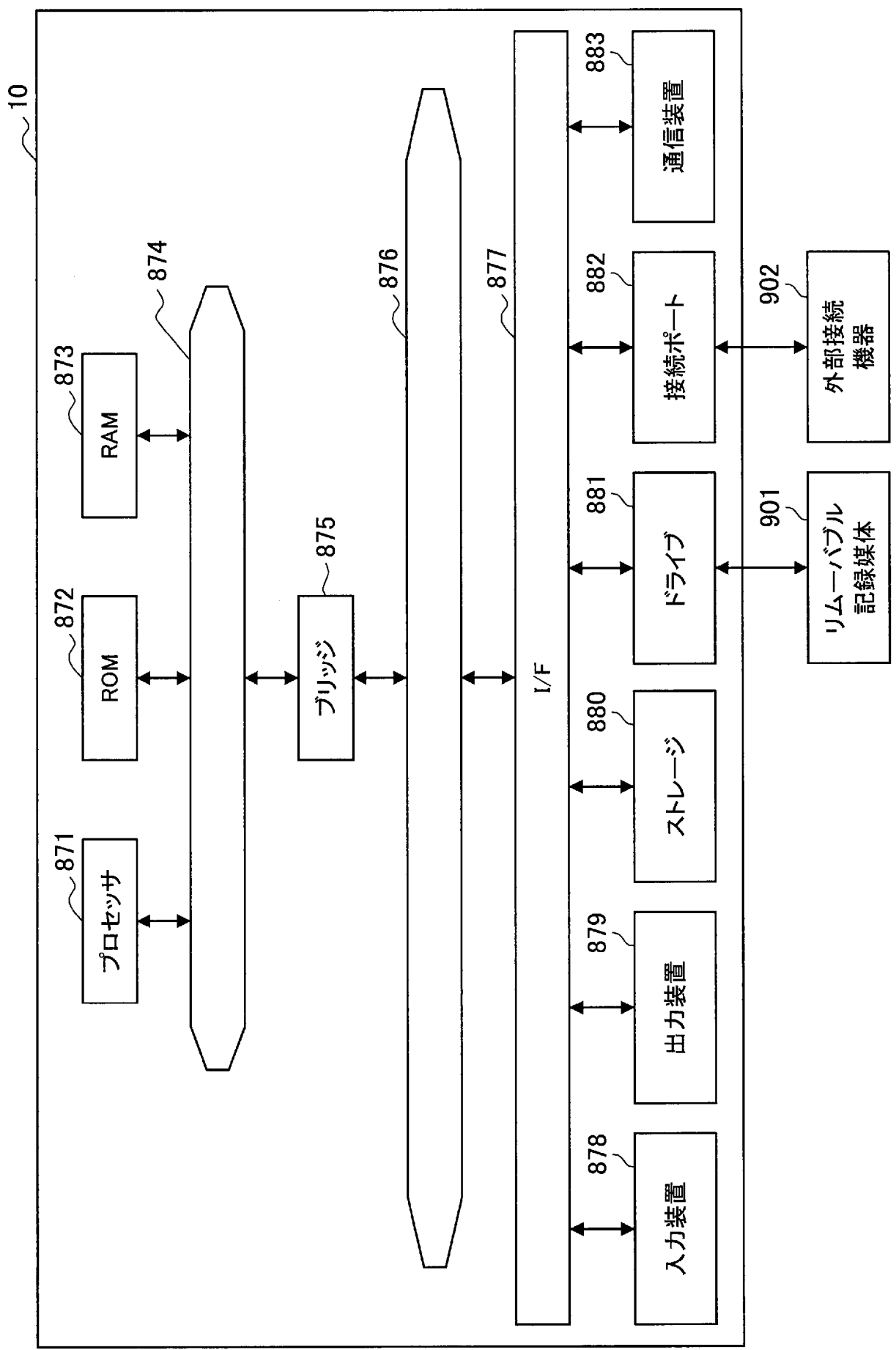
[図15]



[図16]



[図17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/042950

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06N3/08 (2006.01) i

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)

Int. Cl. 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 1971-2019

Registered utility model specifications of Japan 1996-2019

Published registered utility model applications of Japan 1994-2019

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2001-056802 A (OKI ELECTRIC INDUSTRY CO., LTD.) 27 February 2001, entire text (Family: none)	1-20
A	JP 03-223933 A (HITACHI, LTD.) 02 October 1991, entire text & US 5701394 A	1-20
A	US 2018/0144261 A1 (NANTOMICS, LLC.) 24 May 2018, entire text & WO 2018/094360 A2	1-20
A	二宮洋, 改良型 online 準ニュートン法によるニューラルネットワ ークの学習, 電子情報通信学会技術研究報告 NLP2009-81- NLP2009-125, non-linear problem, 04 November 2009, vol. 109, no. 269, pp. 187-192 (NINOMIYA, Hiroshi. Improved online quasi-Newton Method for Training of Neural Networks. IEICE Technical Report.)	1-20

☐ 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
04.02.2019

Date of mailing of the international search report
12.02.2019

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06N3/08 (2006.01) i

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））

Int.Cl. 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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2001-056802 A（沖電気工業株式会社）2001.02.27, 全文（ファ ミリーなし）	1-20
A	JP 03-223933 A（株式会社日立製作所）1991.10.02, 全文 & US 5701394 A	1-20
A	US 2018/0144261 A1（NANTOMICS, LLC.）2018.05.24, 全文 & WO 2018/094360 A2	1-20

☒ 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矛盾するものではなく、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

04.02.2019

国際調査報告の発送日

12.02.2019

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

多胡 滋

5 B

3562

電話番号 03-3581-1101 内線 3545

C (続き) . 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	二宮洋, 改良型 online 準ニュートン法によるニューラルネットワークの学習, 電子情報通信学会技術研究報告 NLP 2 0 0 9 - 8 1 - NLP 2 0 0 9 - 1 2 5 非線形問題, 2009.11.04, 第 109 巻, 第 269 号, pp.187-192	1 - 2 0