

(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
2019年3月28日(28.03.2019)



(10) 国際公開番号

WO 2019/059191 A1

(51) 国際特許分類:
G06N 3/063 (2006.01)

(21) 国際出願番号: PCT/JP2018/034522

(22) 国際出願日: 2018年9月18日(18.09.2018)

(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 2017-180457 2017年9月20日(20.09.2017) JP

(71) 出願人: 国立大学法人東京工業大学
(TOKYO INSTITUTE OF TECHNOLOGY) [JP/
JP]; 〒1528550 東京都目黒区大岡山 2 -
1 2 - 1 Tokyo (JP).

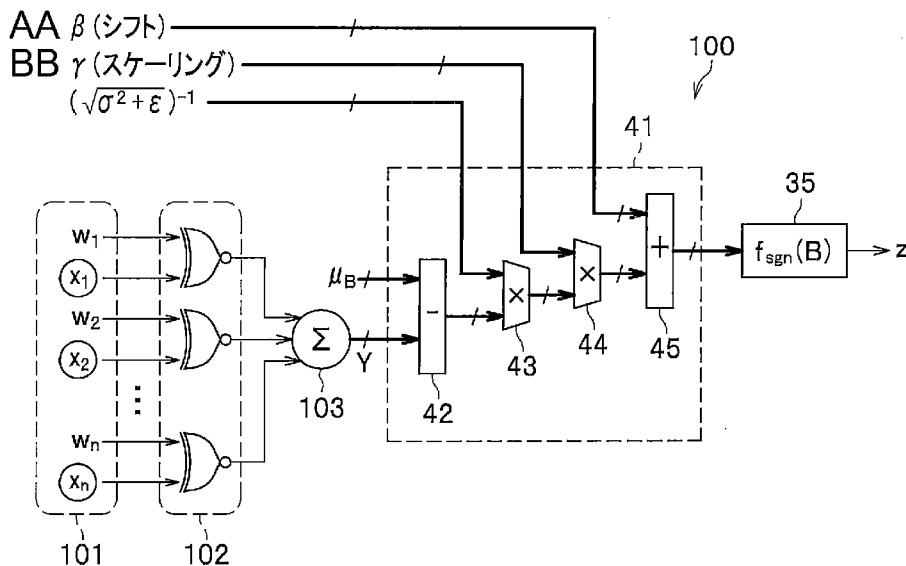
(72) 発明者: 中原 啓 貴 (NAKAHARA Hiroki);
〒1528550 東京都目黒区大岡山 2 - 1 2 - 1 国
立大学法人東京工業大学内 Tokyo (JP).

(74) 代理人: 特許業務法人磯野国際特許商
標事務所 (ISONO INTERNATIONAL PATENT
OFFICE, P.C.); 〒1050001 東京都港区虎ノ
門一丁目1番18号 ヒューリック
虎ノ門ビル 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,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

(54) Title: NEURAL NETWORK CIRCUIT DEVICE, NEURAL NETWORK, NEURAL NETWORK PROCESSING METHOD, AND PROGRAM FOR IMPLEMENTING NEURAL NETWORK

(54) 発明の名称: ニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラム



AA Shift
BB Scaling

(57) Abstract: Provided are a neural network circuit device, a neural network, a neural network processing method, and a program for implementing a neural network, with which bias is not required. A binarized neural network circuit (100) is provided with: input nodes for inputting input values x1 to xn (xi) (binary) and an input unit (101) for inputting weightings w1 to wn (wi) (binary); an XNOR gate circuit (102) which accepts the input values x1 to xn and the weightings w1 to wn and performs XNOR logic; a summing circuit (103) which performs a sum of the XNOR logical values; a batch normalizing circuit (41) which corrects variation biasing due to binarization by

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)）

means of a process to widen a normalization range and shift the center thereof; and an activation function circuit (35) which uses an activation function $f_{\text{sgn}}(B)$ to convert a signal B obtained by batch normalizing a signal Y obtained by said summing.

(57) 要約：バイアスが不要なニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラムを提供する。2値化ニューラルネットワーク回路（100）は、入力値 $x_1 \sim x_n$ （ x_i ）（2値）を入力する入力ノードおよび重み $w_1 \sim w_n$ （ w_i ）（2値）を入力する入力部（101）と、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR論理を取るXNORゲート回路（102）と、各XNOR論理値の総和を取る総和回路（103）と、2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路（41）と、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換する活性化関数回路（35）と、を備える。

明 細 書

発明の名称：

ニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラム

技術分野

[0001] 本発明は、ニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラムに関する。

背景技術

[0002] 古典的な順伝搬型ニューラルネットワーク（F F N N : Feedforward Neural Network）、R B F（Radial Basis Function）ネットワーク、正規化したR B F ネットワーク、自己組織化マップなどがある。R B F Nは、誤差逆伝搬法に用いる活性化関数に放射基底関数を用いる。しかし、中間層が多く取れず高精度認識判定が難しかったり、HW規模が大きく処理時間がかかる、などの問題があり手書き文字認識など応用分野が限定されていた。

近年、A D A S（advanced driver assistance system）用の画像認識や自動翻訳などで注目を集める新方式として畳み込みニューラルネットワーク（C N N : Convolutional Neural Network）（層間が全結合でないN N）や再帰型ニューラルネットワーク（双方向伝搬）が登場している。C N Nは、ディープニューラルネットワーク（D N N : Deep Neural Network）に畳込み演算を付加したものである。

[0003] 特許文献1には、誤り訂正符号の検査行列に基づいて、階層型ニューラルネットワークにおける疎結合のノード間で学習された重みの値と入力信号とを用いて、問題を解く処理部を備える処理装置が記載されている。

[0004] 既存のC N Nは、短精度（多ビット）による積和演算回路で構成されており、多数の乗算回路が必要である。このため、面積・消費電力が多くなる欠点があった。そこで、2値化した精度、すなわち+1と-1（または0と

1)のみ用いてCNNを構成する回路が提案されている（例えば、非特許文献1～4参照）。

[0005] 非特許文献1～4の技術では、精度を2値に落とすことでCNNの認識精度も落としてしまう。これを避けて2値化CNNの精度を維持するためには、バッチ正規化回路が必要である。

先行技術文献

特許文献

[0006] 特許文献1：特開2016-173843号公報

非特許文献

[0007] 非特許文献1：M. Courbariaux, I. Hubara, D. Soudry, R.E. Yaniv, Y. Bengio, “Binarized neural networks: Training deep neural networks with weights and activations constrained to +1 or -1,” Computer Research Repository (CoRR), 「2値化NNのアルゴリズム」、[online]、2016年3月、[平成28年10月5日検索]、<URL: <http://arxiv.org/pdf/1602.02830v3.pdf> >

非特許文献2：Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, Ali Farhadi, “XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks,” Computer Vision and Pattern recognition, 「2値化NNのアルゴリズム」、[online]、2016年3月、[平成28年10月5日検索]、<URL: <https://arxiv.org/pdf/1603.05279v4> >

非特許文献3：Hiroki Nakahara, Haruyoshi Yonekawa, Tsutomu Sasao, Hisashi Iwamoto and Masato Motomura, ” A Memory-Based Realization of a Binarized Deep Convolutional Neural Network,” Proc. of the 2016 International Conference on Field-Programmable Technology (FPT), Xi'an, China, Dec 2016 (To Appear).

非特許文献4：Eriko Nurvitadhi, David Sheffield, Jaewoong Sim, Asit Mishra, Ganesh Venkatesh, Debbie Marr, ” Accelerating Binarized Neural Networks: Comparison of FPGA, CPU, GPU, and ASIC,” Proc. of the 2016

International Conference on Field-Programmable Technology (FPT), Xi'an, China, Dec 2016 (To Appear).

発明の概要

発明が解決しようとする課題

[0008] CNNでは、学習を進めると重みが均等に分布する。しかし、学習データには偏りが存在するため、完全には均等に分布せず、その調整のためにバイアスによる補正が必要であった。学習データにもよるが、バイアスの精度は固定小数点精度で30～40ビットになり、浮動小数点精度を使ったとしても、加算器などの回路が必要であった。バイアスがあることで、面積・消費電力が増大するという課題があった。

[0009] 本発明は、このような事情に鑑みてなされたものであり、バイアスが不要なニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラムを提供することを課題とする。

課題を解決するための手段

[0010] 前記した課題を解決するため、本発明に係るニューラルネットワーク回路装置は、入力層、1以上の中間層、および、出力層を少なくとも含むニューラルネットワーク回路装置であって、前記中間層の中で、2値の入力値 x_i および重み w_i を受け取り、論理演算を行う論理回路部と、前記論理回路部の出力の総和を取る総和回路部と、2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路部と、総和を取った信号 Y のバッチ正規化した信号 B を活性化関数 $f_{\text{sgn}}(B)$ で変換する活性化関数回路部と、を備え、前記総和を取った信号 Y は、下記式で示される[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ただし、

γ : スケーリング係数

β : シフト値

μ'_B : バイアスを除く平均値。入力値 $\times 0$ のときのバイアス値 w_0 とし、ミニバッチの平均値 μ_B とするとき、 $w_0 - \mu_B$

σ^2_B : ミニバッチの分散値

ε : 定数

ことを特徴とする。

その他の手段については、発明を実施するための形態のなかで説明する。

発明の効果

[0011] 本発明によれば、バイアスが不要なニューラルネットワーク回路装置、ニューラルネットワーク、ニューラルネットワーク処理方法およびニューラルネットワークの実行プログラムを提供することができる。

図面の簡単な説明

[0012] [図1]ディープニューラルネットワーク (DNN) の構造の一例を説明する図である。

[図2]比較例のニューラルネットワークのニューラルネットワーク回路の構成の一例を示す図である。

[図3]図2に示すニューラルネットワーク回路における活性化関数 $f_{act}(Y)$ を示す図である。

[図4]図2に示すニューラルネットワーク回路の乗算回路をXNORゲート回路に置き換えた2値化ニューラルネットワーク回路の構成の一例を示す図である。

[図5]図4に示す2値化ニューラルネットワーク回路における活性化関数 $f_{sgn}(B)$ を示す図である。

[図6]比較例のバッチ正規化回路を備える2値化ニューラルネットワーク回路の構成の一例を示す図である。

[図7]ニューラルネットワークの2値化ニューラルネットワーク回路のスケーリング (γ) による正規化を示す図である。

[図8]ニューラルネットワークの2値化ニューラルネットワーク回路のシフト(β)による $-1 \sim +1$ の制限を示す図である。

[図9]本発明の実施形態に係るディープニューラルネットワークの2値化ニューラルネットワーク回路の構成を示す図である。

[図10A]本発明の実施形態に係るディープニューラルネットワークの「バッチ正規化なし」の認識精度を説明する図である。

[図10B]本発明の実施形態に係るディープニューラルネットワークの「バッチ正規化あり」(バイアス項あり)の認識精度を説明する図である。

[図10C]本発明の実施形態に係るディープニューラルネットワークの「バッチ正規化あり」(バイアス項なし)の認識精度を説明する図である。

[図11]本発明の実施形態に係るディープニューラルネットワークの2値化ニューラルネットワーク回路と既存の2値化ニューラルネットワーク回路との比較を行った結果を表にして示す図である。

[図12]本発明の実施形態に係るディープニューラルネットワークの2値化ニューラルネットワーク回路の実装例を説明する図である。

[図13]各FPGA実装のハードウェアの量を比較して示す図である。

[図14]変形例1のディープニューラルネットワークの2値化ニューラルネットワーク回路の構成を示す図である。

[図15]変形例2のディープニューラルネットワークの2値化ニューラルネットワーク回路の構成を示す図である。

発明を実施するための形態

[0013] 以下、図面を参照して本発明を実施するための形態(以下、「本実施形態」という)におけるディープニューラルネットワークについて説明する。

(背景説明)

図1は、ディープニューラルネットワーク(DNN)の構造の一例を説明する図である。

図1に示すように、ディープニューラルネットワーク(DNN)1は、入力層(input layer)11、任意の数の中間層である隠れ層(hidden layer)

1 2、出力層 (output layer) 1 3を有して構成される。

入力層 (input layer) 1 1は、複数個 (ここでは8) の入力ノード (ニューロン) を有する。隠れ層 1 2は、複数 (ここでは3層 (hidden layer1, hidden layer2, hidden layer3)) である。実際には、隠れ層 1 2の層数 n は、例えば20～100に達する。出力層 1 3は、識別対象の数 (ここでは4) の出力ノード (ニューロン) を有する。なお、層数およびノード数 (ニューロン数) は、一例である。

ディープニューラルネットワーク 1は、入力層 1 1と隠れ層 1 2のノード間が全て結合し、隠れ層 1 2と出力層 1 3のノード間が全て結合している。

[0014] 入力層 1 1、隠れ層 1 2および出力層 1 3には、任意の数のノード (図1の○印参照) が存在する。このノードは、入力を受け取り、値を出力する関数である。入力層 1 1には、入力ノードとは別に独立した値を入れるバイアス (bias) ノードがある。構成は、複数のノードを持つ層を重ねることで構築される。伝播は、受け取った入力に対して重み (weight) をかけ、受け取った入力を次層に活性化関数 (activation function) で変換して出力する。活性化関数は、sigmoid関数やtanh関数などの非線形関数、ReLU (Rectified Linear Unit function: 正規化線形関数) がある。ノード数を増やすことで、扱う変数を増やし、多数の要素を加味して値/境界を決定できる。層数を増やすことで、直線境界の組み合わせ、複雑な境界を表現できる。学習は、誤差を計算し、それを基に各層の重みを調整する。学習は、誤差を最小化する最適化問題を解くことであり、最適化問題の解法は誤差逆伝播法 (Backpropagation) を使うのが一般的である。誤差は、二乗和誤差を使うのが一般的である。汎化能力を高めるために、誤差に正則化項を加算する。誤差逆伝播法は、誤差を出力層 1 3から伝播させていき、各層の重みを調整する。

[0015] 図1のディープニューラルネットワーク 1の構成を2次元に展開することで画像処理に適したCNNを構築できる。また、ディープニューラルネットワーク 1にフィードバックを入れることで、双方向に信号が伝播するRNN (Recurrent Neural Network: 再帰型ニューラルネットワーク) を構成する

ことができる。

[0016] 図1の太破線三角部に示すように、ディープニューラルネットワーク1は、多層のニューラルネットワークを実現する回路（以下、ニューラルネットワーク回路という）2から構成されている。

本技術は、ニューラルネットワーク回路2を対象とする。ニューラルネットワーク回路2の適用箇所および適用数は限定されない。例えば、隠れ層12の層数 n ：20～30の場合、これらの層のどの位置に適用してもよく、またどのノードを入出力ノードとするものでもよい。さらに、ディープニューラルネットワーク1に限らず、どのようなニューラルネットワークでもよい。ただし、入力層11または出力層13のノード出力には、2値化出力ではなく多ビット出力が求められるので、ニューラルネットワーク回路2は、対象外である。ただし、出力層13のノードを構成する回路に、乗算回路が残ったとしても面積的には問題にはならない。

なお、入力データに対し学習済のものを評価していくことを前提としている。したがって、学習結果として重み w_i は既に得られている。

[0017] <ニューラルネットワーク回路>

図2は、比較例のニューラルネットワーク回路の構成の一例を示す図である。

比較例のニューラルネットワーク回路20は、図1のディープニューラルネットワーク1を構成するニューラルネットワーク回路2に適用できる。なお、以下の各図の表記において、値が多ビットである場合は太実線矢印とバンドルで、また値が2値である場合は細太実線矢印で示す。

ニューラルネットワーク回路20は、入力値（判別データ） $X_1 \sim X_n$ （多ビット）を入力する入力ノードおよび重み $W_1 \sim W_n$ （多ビット）を入力する入力部21と、バイアス W_0 （多ビット）を入力するバイアス W_0 入力部22と、入力値 $X_1 \sim X_n$ および重み $W_1 \sim W_n$ を受け取り、入力値 $X_1 \sim X_n$ に重み $W_1 \sim W_n$ をそれぞれ乗算する複数の乗算回路23と、各乗算値とバイアス W_0 との総和を取る総和回路24と、総和を取った信号 Y を活

活性化関数 $f_{act}(Y)$ で変換する活性化関数回路 25 と、を備えて構成される。

以上の構成において、ニューラルネットワーク回路 20 は、入力値 $X_1 \sim X_n$ (多ビット) を受け取り、重み $W_1 \sim W_n$ を乗算した後に、バイアス W_0 を含めて総和を取った信号 Y を活性化関数回路 25 を通すことで人間のニューロンに模した処理を実現している。

[0018] 図 3 は、前記図 2 に示すニューラルネットワーク回路 20 における活性化関数 $f_{act}(Y)$ を示す図である。図 3 は、横軸に総和を取った信号 Y 、縦軸に活性化関数 $f_{act}(Y)$ の値をとる。図 3 の符号 $\bigcirc$ 印は、 ± 1 の範囲の値をとる正側の活性化値 (状態値)、図 3 の符号 $\times$ 印は、 ± 1 の範囲の値をとる負側の活性化値である。

ニューラルネットワーク回路 20 (図 2 参照) は、多ビットで高い認識精度を実現している。このため、活性化関数回路 25 (図 2 参照) において、非線形な活性化関数 $f_{act}(Y)$ を用いることができる。すなわち、図 4 に示すように、非線形な活性化関数 $f_{act}(Y)$ は、傾きが非ゼロとなる部分 (図 4 の破線囲み部分参照) に ± 1 の範囲の値をとる活性化値を設定できる。このため、ニューラルネットワーク回路 20 は、多様な活性を実現でき、認識精度は実用的な値になっていた。しかし、ニューラルネットワーク回路 20 は、大量の乗算回路 23 が必要になる。加えて、ニューラルネットワーク回路 20 は、入出力・重みが多ビットであることにより、大量のメモリが必要であり、読み書きの速度 (メモリ容量・帯域) も問題である。

[0019] <単に 2 値化した 2 値化ニューラルネットワーク回路>

図 2 に示す比較例のニューラルネットワーク回路 20 は、短精度 (多ビット) による積和演算回路で構成されている。このため、多数の乗算回路 21 が必要であり、面積・消費電力が多くなる欠点があった。また、入出力・重みが多ビットであることで大量のメモリが必要であり、読み書きの速度 (メモリ容量・帯域) が問題となっていた。

そこで、2 値化した精度、すなわち $+1$ と -1 のみ用いてニューラルネットワーク回路 2 (図 1 参照) を構成する回路が提案された (非特許文献 1 ~

4)。具体的には、図2に示すニューラルネットワーク回路20の乗算回路21を、論理ゲート（例えばXNORゲート回路）に置き換えることが考えられる。

[0020] 図4は、比較例の図2に示すニューラルネットワーク回路20の乗算回路21をXNORゲート回路に置き換えた2値化ニューラルネットワーク回路の構成の一例を示す図である。

比較例の2値化ニューラルネットワーク回路30は、図1のニューラルネットワーク回路2に適用できる。

図4に示すように、比較例の2値化ニューラルネットワーク回路30は、入力値 $x_1 \sim x_n$ （2値）を入力する入力ノードおよび重み $w_1 \sim w_n$ （2値）を入力する入力部31と、バイアス w_0 （2値）を入力するバイアス w_0 入力部32と、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR（Exclusive NOR：否定排他的論理和）論理を取る複数のXNORゲート回路33と、XNORゲート回路33の各XNOR論理値とバイアス w_0 との総和を取る総和回路34と、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f \operatorname{sgn}(B)$ で変換する活性化関数回路35と、を備えて構成される。

2値化ニューラルネットワーク回路30は、乗算回路23（図2参照）がXNOR論理を実現するXNORゲート回路33に置き換えられている。このため、乗算回路23を構成する際に必要であった面積を削減することができる。また、入力値 $x_1 \sim x_n$ 、出力値 z 、および重み $w_1 \sim w_n$ は、いずれも2値（-1と+1）であるため、多値である場合と比較してメモリ量を大幅に削減でき、メモリ帯域を向上させることができる。

[0021] 図5は、比較例の前記図4に示す2値化ニューラルネットワーク回路30における活性化関数 $f \operatorname{sgn}(B)$ を示す図である。図5は、横軸に総和を取った信号Y、縦軸に活性化関数 $f \operatorname{sgn}(B)$ の値をとる。図5の符号○印は、±1の範囲の値をとる正側の活性化値、図5の符号×印は、±1の範囲の値をとる負側の活性化値である。

2値化ニューラルネットワーク回路30は、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を単に2値化している。このため、図5の符号aに示すように、 ± 1 のみ扱う活性化関数しか扱えないため、誤差が頻繁に生じてしまう。また、傾きが非ゼロとなる区間（図5の破線囲み部分参照）が不均等となり学習が上手く行われない。すなわち、図6の符号bに示すように、不均等な幅により微分が定義できない。その結果、単に2値化した2値化ニューラルネットワーク回路40は、認識精度が大幅に落ち込んでしまう。

そこで、非特許文献1～4には、既存の2値化ニューラルネットワークの精度を維持するためにバッチ正規化を行う技術が記載されている。

[0022] <バッチ正規化回路を備える2値化ニューラルネットワーク回路>

図6は、比較例の2値化した精度を是正して、CNNの認識精度を保つバッチ正規化回路（BN：Batch Normalization）を備える2値化ニューラルネットワーク回路40の構成の一例を示す図である。図4と同一構成部分には同一符号を付している。

図6に示すように、比較例の2値化ニューラルネットワーク回路40は、入力値 $x_1 \sim x_n$ （2値）を入力する入力ノード $x_1 \sim x_n$ 、重み $w_1 \sim w_n$ （2値）を入力する入力部31と、バイアスB（多ビット）を入力するバイアスB入力部32と、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR（Exclusive NOR：否定排他的論理和）論理を取る複数のXNORゲート回路33と、XNORゲート回路33の各XNOR論理値とバイアスBとの総和を取る総和回路34と、2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路41と、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換する活性化関数回路35と、を備えて構成される。

[0023] バッチ正規化回路41は、重み総和を取った信号Yと平均値（ μ_B ）との差分をとる減算器42と、減算器42出力とミニバッチの分散値（ σ^2_B ）および定数（ ϵ ）の和の根の逆数とを乗算する第1乗算回路43と、第1乗算回路43出力に、スケーリング（ γ ）値（多ビット）による正規化を行う第2

算回路 44 と、スケーリング係数 (γ) による正規化後、シフト値 (β) (多ビット) によりシフトして 2 分類を行う加算器 45 とを有する。スケーリング係数 (γ) およびシフト値 (β) の各パラメータは、事前に学習時に求めておく。

[0024] 学習データ毎に重みを更新していると、計算時間がかかり特定のデータに依存してしまう。そこで、バッチというまとまったデータ数単位で更新を行う手法が採られる。ミニバッチとは、バッチを更に小さくした単位であり、現在はこのミニバッチを用いる。

[0025] 2 値化ニューラルネットワーク回路 40 は、バッチ正規化回路 41 を備えることで、2 値化した精度を是正して、CNN の認識精度を保つようにする。

なお、入力値 $x_1 \sim x_n$ と重み $w_1 \sim w_n$ との XNOR 論理を取る論理回路であれば、XNOR ゲートに限らずどのような論理ゲートでもよい。例えば、XOR 回路を用いて総和をとり、活性化関数の否定を取ってもよい。

[0026] <バッチ正規化回路が必要となる理由>

比較例の 2 値化ニューラルネットワーク回路 40 のバッチ正規化回路 41 が必要となる理由について説明する。

図 7 および図 8 は、比較例の 2 値化ニューラルネットワーク回路 40 のバッチ正規化による効果を説明する図である。図 7 は、比較例のスケーリング係数 (γ) による正規化を示す図、図 8 は、比較例のシフト値 (β) による $-1 \sim +1$ の制限を示す図である。

バッチ正規化とは、2 値化によるバラツキの偏りを是正する回路であり、重み総和後、スケーリング係数 (γ) による正規化を行った後、シフト値 (β) による適切な活性化による 2 分類を行う。これらのパラメータは事前に学習時に求めておく。具体的には、下記の通りである。

[0027] 図 7 の白抜矢印および符号 c に示すように、バッチ正規化回路 41 の乗算回路 42 (図 6 参照) は、重み総和後の信号 (結果) Y を、スケーリング係数 (γ) により、幅「2」(図 7 の網掛け部参照) に正規化する。これによ

り、図5の幅（図5の網掛け部参照）と比較して分かるように、単に2値化した2値化ニューラルネットワーク回路30では、不均等な幅により微分が定義できなかった不具合が、スケーリング係数（ γ ）により幅「2」に正規化することで、不均等な幅が抑制される。

[0028] その上で、図8の白抜矢印および符号dに示すように、バッチ正規化回路41の加算器45（図6参照）は、スケーリング係数（ γ ）による正規化後の値を、シフト値（ β ）により $-1 \sim +1$ の範囲になるよう制限する。すなわち、図5の幅（図5の網掛け部参照）と比較して分かるように、図5の幅（図5の網掛け部参照）が、 $+1$ 側により多くシフトしている場合には、シフト値（ β ）により、スケーリング係数（ γ ）による正規化後の値を $-1 \sim +1$ に制限することで、この幅の中心を0とする。図5の例では、負側の活性化値（図5の破線囲み部の符号×印参照）が、本来あるべき負側に戻される。これにより、誤差の発生が減少し、認識精度を高めることができる。

このように、比較例の2値化ニューラルネットワーク回路40には、バッチ正規化回路41が必要である。

[0029] <バイアス項による補正の必要性>

上述したように、学習を進めると重みが均等に分布する。しかし、学習データには偏りが存在するため、完全には均等に分布せず、その調整のためにごくわずかなバイアス項による補正が必要であった。学習データにもよるがバイアス項の精度は、固定小数点精度で30～40ビットになり、浮動小数点精度を使ったとしても、加算器などの回路が必要であった。

[0030] すなわち、学習データの偏りが存在するので完全に均等分布は非常に困難であり、それを調整するためにバイアス（もしくはバイアスに相当する操作）が必要となる。このため、バイアスは多ビットである必要がある。また、バイアス値は、学習データや学習期間に応じて、刻々と変化する。

バイアスがない場合、実用に耐えない。例えば、後記する図10の例では、Classification Errorが90%程度になると想定される。

バイアスがあることの直接的な課題は、高精度による回路が必要とされる

ことである。このような高精度な回路は、面積・消費電力が大きいものとなる。

[0031] (本発明の原理説明)

本発明の着眼点は、バッチ正規化の操作を導入したNNに対して、これと等価なNNを解析的に求めると、バイアス項が不要なNNを得ることができることを発見したことである。

すなわち、重み積和後に2値化ニューラルネットワーク回路40のバッチ正規化回路41(図6参照)に入力される信号をYとすると、バッチ正規化回路41から出力される信号(Yと等価となる信号)Y'(中間値)は、次式(1)で示される。

[0032] [数1]

$$\begin{aligned} Y' &= \gamma \frac{Y - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \\ &= \frac{\gamma}{\sqrt{\sigma_B^2 + \epsilon}} \left(Y - \left(\mu_B - \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta \right) \right) \quad \dots (1) \end{aligned}$$

ただし、

γ : スケーリング係数

β : シフト値

μ_B : ミニバッチの平均値

σ_B^2 : ミニバッチの分散値

ϵ : 定数(0による除算を避けるための定数)

[0033] ここで、上記スケーリング係数(γ)、シフト値(β)、ミニバッチの平均値(μ_B)、ミニバッチの分散値(σ_B^2)、および定数(ϵ)は、学習時にバッチ正規化により得られる値である。

[0034] 2値化ニューラルネットワーク回路40の活性化関数を通した出力は、+1または-1(信号の割り当てによっては0または1)である。また、上記式(1)の係数 $\gamma / \sqrt{(\sigma_B^2 + \epsilon)^{-1}}$ により、2値化ニューラルネットワー

ク回路40の中間信号を変換しても活性化関数を通した値は変化しないので、無視できる。

したがって、上記式(1)は、次式(2)となる。

[0035] [数2]

$$Y' = (Y - (\mu_B - \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma})) \quad \dots(2)$$

[0036] ここで、2値化活性化関数の値は、中間値 Y' が正か負かで+1または-1(信号の割り当てによっては0または1)である。したがって、2値化活性化関数の値 $f'_{sgn}(Y)$ は、下記式(3)の条件で決まる。

[0037] [数3]

$$f'_{sgn}(Y') = \begin{cases} 1 & \text{if } Y' < -\mu_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \\ -1 & \text{上記以外} \end{cases} \quad \dots(3)$$

[0038] よって、これらの解析的な操作から重み積和演算は、下記式(4)のよう
に得られる。

[0039] [数4]

$$Y = \sum_{i=0}^n w_i x_i - \mu_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta \quad \dots(4)$$

[0040] さて、単に2値化した2値化ニューラルネットワーク回路20(図2参照)
より、入力値 $x_0 = 1$ である。このとき、 w_0 はバイアス値であったから、
上記式(4)は、下記式(5)となる。

[0041] [数5]

$$Y = \sum_{i=1}^n w_i x_i + (w_0 - \mu_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots(5)$$

[0042] ここで、上記式(5)の第1項の総和演算が $i = 1$ で始まることに注意されたい。
すなわち、上記式(5)の第1項は、バイアス値が含まれないニューラルネットワークを表している。
ちなみに、上記式(4)の第1項の総和

演算は $i = 0$ で始まっている。

ここで、上記式 (5) に $w_0 - \mu_B = \mu'_B$ を導入すると、上記式 (5) は、下記式 (6) となる。 μ'_B は、バイアスを除く平均値である。

[0043] [数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{Y} \beta) \quad \dots (6)$$

[0044] 上記式 (6) は、次のことを意味する。すなわち、上記式 (6) は、バイアスを除いたニューラルネットワークとバッチ正規化を用いて学習することで、正しく2値化ニューラルネットワークが実現できることを表している。このとき、上記式 (6) を実現する回路構成は、 μ'_B を学習する、すなわち、従来学習していたミニバッチの平均値とバイアス値を減算した値を学習するものといえる。

[0045] 従来のニューラルネットワークでは、中間層、特に後半部になるにつれて学習が収束する。後半部になるにつれて学習が収束することで、重みの分散が一定となる。このため、それを調整するバイアス値が極めて小さくなる問題があった。

上記式 (6) は、バイアス値を個別に学習しなくても等価な2値ニューラルネットワークが学習できることを示している。解析的な洞察がなければ成り立たない。

[0046] [実施形態の構成]

図9は、本発明の実施形態に係るニューラルネットワークの2値化ニューラルネットワーク回路の構成を示す図である。比較例の図6と同一構成部分には、同一符号を付している。

本実施形態の2値化ニューラルネットワーク回路は、ディープニューラルネットワークへの実装技術を提供する。

2値化ニューラルネットワーク回路100は、図1のニューラルネットワーク回路2に適用できる。

2値化ニューラルネットワーク回路100（ニューラルネットワーク回路

装置)は、バイアスが不要な2値化ニューラルネットワーク回路である。

図9に示すように、2値化ニューラルネットワーク回路100は、入力値 $x_1 \sim x_n$ (x_i) (2値)を入力する入力ノードおよび重み $w_1 \sim w_n$ (w_i) (2値)を入力する入力部101と、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR論理を取るXNORゲート回路102 (論理回路部)と、各XNOR論理値の総和を取る総和回路103 (総和回路部)と、2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路41と、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換する活性化関数回路35と、を備えて構成される。

2値化ニューラルネットワーク回路100は、前記式(6)で示されるバイアス項を不要とする2値化CNNである。

[0047] 2値化ニューラルネットワーク回路100は、ディープニューラルネットワーク1の隠れ層12 (図1参照)に適用される。ここでは、ディープニューラルネットワーク1において、入力値に対し学習済のものを評価していくことを前提としている。

なお、NNでは、重みが、クライアントの認識物体毎に全て異なる。また学習により毎回異なることがある。画像処理では係数は、全て同じであり、この点でNNと画像処理では、HWが大きく異なる。

[0048] XNORゲート回路102は、排他的論理和を含むどのような論理回路部でもよい。すなわち、入力値 $x_1 \sim x_n$ と重み $w_1 \sim w_n$ との論理を取る論理回路であれば、XNORゲートに限らずどのようなゲート回路でもよい。例えば、XORゲートにNOTゲートを組み合わせる、AND、ORゲートを組み合わせる、さらにはトランジスタスイッチを用いて作製するなど、論理的に等しいものであればどのようなものでもよい。

[0049] 2値化ニューラルネットワーク回路100は、バイアスが不要な2値化ニューラルネットワーク回路であり、比較例の2値化ニューラルネットワーク回路40 (図6参照)で必要であったバイアスb入力部32が削除されてい

る。

したがって、総和回路 103 は、各 XNOR 論理値の総和のみを取る。すなわち、総和回路 103 は、比較例の 2 値化ニューラルネットワーク回路 400 (図 6 参照) 総和回路 35 のように、各 XNOR 論理値とバイアス b との総和を取ることはない。

[0050] バッチ正規化回路 41 は、重み総和を取った信号 Y と平均値 (μ_B) との差分をとる減算器 42 と、減算器 42 出力とミニバッチの分散値 (σ^2_B) および定数 (ϵ) とを乗算する第 1 乗算回路 43 と、第 1 乗算回路 43 出力に、スケーリング (γ) 値 (多ビット) による正規化を行う第 2 乗算回路 44 と、スケーリング係数 (γ) による正規化後、シフト値 (β) (多ビット) によりシフトして 2 分類を行う加算器 45 と、からなる。

[0051] 活性化回路 35 は、総和を取った信号 Y に対して符号ビットのみを出力する活性化関数回路を模擬する回路となっている。符号ビットは、総和を取った多ビット信号 Y を活性化するかしないかで示す 2 値信号である。

[0052] このように、2 値化ニューラルネットワーク回路 100 は、総和回路 103 が、式 (6) で示されるように各 XNOR 論理値の総和のみを取る。このため、2 値化ニューラルネットワーク回路 100 は、バイアスが不要なニューラルネットワーク回路となっている。

[0053] 以下、上述のように構成された 2 値化ニューラルネットワーク回路 100 の動作について説明する。

2 値化ニューラルネットワーク回路 100 は、図 1 に示すディープニューラルネットワーク 1 のニューラルネットワーク回路 2 に用いられる。この場合、2 値化ニューラルネットワーク回路 100 の入力ノード $x_1 \sim x_n$ は、図 1 に示すディープニューラルネットワーク 1 の hidden layer1 の入力ノードである。入力部 101 には、隠れ層 12 の hidden layer1 の入力ノードの入力値 $x_1 \sim x_n$ (2 値) および重み $w_1 \sim w_n$ (2 値) が入力される。

乗算の代わりとなる XNOR ゲート回路 102 では、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR 論理により 2 値 ($-1/+1$) の

乗算を行う。XNORゲート回路102を通した後、総和回路103は、各XNOR論理値の総和を取る。そして、バッチ正規化回路41は、総和を取った信号Yをバッチ正規化する。活性化関数回路35は、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換する。

[0054] 2値化ニューラルネットワーク回路100は、比較例の多ビット構成の乗算回路21（図2参照）がXNOR論理を実現するXNORゲート回路102に置き換えられている。このため、乗算回路21を構成する際に必要であった面積を削減することができる。また、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ は、いずれも2値（-1 / +1）であるため、多ビット（多値）である場合と比較してメモリ容量を大幅に削減でき、メモリ帯域を向上させることができる。

[0055] [学習の実施例]

本実施形態の効果を確認するため、VGG11（隠れ層が11層）ベンチマークNNを実装し、学習が成功するか確認した。VGG11は、良く使われているベンチマークで再現性があるものである。

図10A-Cは、比較例および本実施形態の2値化VGG11モデルを用いて画像認識タスクCIFAR10を学習させた結果の認識精度を説明する図である。図10Aは「バッチ正規化なし」（バイアス項あり）で構成したニューラルネットワーク回路30（図4参照）の認識精度、図10Bは「バッチ正規化あり」（バイアス項あり）で構成した2値化ニューラルネットワーク回路40（図6参照）の認識精度、図10Cは「バッチ正規化あり」（バイアス項なし）で構成した本実施形態の2値化ニューラルネットワーク回路100（図9参照）の認識精度をそれぞれ示す。図10Cは、本実施形態をVGG11ベンチマークNNで実装し確認したものである。

図10A-Cの横軸は、利用した学習データに対して更新を終えたサイクルであるエポック（epoch）数（学習回数）、縦軸は誤認識率（Classification error）である。また、図10A-Cは、ディープニューラルネットワーク用のフレームワークソフトウェアChainer（登録商標）のfloat32 精度CNN

を用いている。

[0056] <バッチ正規化の効果>

図10Aの「バッチ正規化なし」に示すように、比較例の単に2値化した2値化ニューラルネットワーク回路30（図4参照）では、誤認識率が大きく（エポック数200以上；誤認識率約70%）認識精度は悪い。また、学習を続けても認識精度の改善は見られない（学習が成功していない）。

これに対して、比較例の図10Bの「バッチ正規化あり」（バイアス項あり）で構成した2値化ニューラルネットワーク回路40（図6参照）と本実施形態の図10Cの「バッチ正規化あり」（バイアス項なし）で構成した2値化ニューラルネットワーク回路100（図9参照）は、どちらも学習を続けるに従って誤認識率が落ちており、エポック数400以上で誤認識率は小さくなり（約20%）、学習が成功していることを示している。

このように、バッチ正規化回路41が無ければ学習が成功していない。2値化ニューラルネットワーク回路30では、バッチ正規化回路41が必要であることがあらためて確認された。

[0057] <バイアス項を不要とする2値化CNNバッチ正規化の効果>

比較例の図10Bの「バッチ正規化あり」（バイアス項あり）と本実施形態の図10Cの「バッチ正規化あり」（バイアス項なし）とを比較して分かるように、バッチ正規化回路41がある場合には、バイアス項の有無については、認識精度にほとんど影響を与えないことが確認された。

すなわち、図10Cの「バッチ正規化あり」（バイアス項なし）で構成した本実施形態のニューラルネットワーク回路100（図9参照）は、比較例の図10Bの「バッチ正規化あり」（バイアス項あり）で構成したニューラルネットワーク回路40と比較して、バイアス項をなくしても認識精度を落とさないことが確認された。

[0058] 図11は、本実施形態の2値化ニューラルネットワーク回路100をFPGA(Digilent 社 NetFPGA-1G-CML)上に実装し、既存の多ビット実装法との比較を行った結果を表にして示す図である。

図11の表は、表下欄外に表記した[Zhao et al.]～[FINN]の学会発表者（論文発表年）のニューラルネットワークと本実施形態のニューラルネットワークをFPGA（Digilent社 ZedBoard）上に実現した場合に、各項目を対比して示したものである。

図11の表の用語は下記の通りである。

Implementationは、実装した手法・研究グループ名である。

(Year)は、関連文献が発表された年である。

FPGA Boardは、FPGA（field-programmable gate array）搭載ボードの名称である。

(FPGA)は、FPGA搭載ボードに搭載されているFPGAの型番であり、比較条件を揃えるため同一のFPGAを使用する。

Clock [MHz] は、FPGAの動作周波数である。動作周波数が高いほど高速に動作する。

#LUTsは、FPGAのLUT（Look-Up Table）消費量であり、面積を意味する。

#18Kb BRAMsは、FPGAの内部メモリブロックの消費量であり、面積を意味する。

#DSP Blocksは、FPGAの内部積和演算ブロックの消費量であり、面積を意味する。

Test Errorは、テスト画像のエラー率であり、認識精度を意味する。

Time [msec] は、認識時間（単位はミリ秒）である。

(FPS)は、Frames Per Secondの略で単位時間あたりに認識できる画像数である。(FPS)の数値が高いほど高速である。

Power [W]は、消費電力（単位はワット）である。

FPS/Wattは、消費電力効率である。

FPS/LUTは、面積効率である。

FPS/BRAMは、メモリ量効率である。

[0059] 図11の表は、「Platform」（プラットフォーム）、「FPGA Board」（使用FPGAボード）、「Clock (MHz)」（同期化のための内部クロック）、「LUT, BRAM,

DSP Block」(メモリ・DSP数)、「Test Error」(誤認識率)、「Time (msec) (FPS)」(処理時間(処理速度))、「Power (W)」(消費電力)、「FPS/Watt, FPS/ LUT, FPS/BRAM」(データ転送待ち時間／外部にメモリを付けた場合の転送速度)、の各項目を対比して示した。この表において、特に注目すべき事項は下記の通りである。

[0060] <消費電力>

本実施形態の2値化ニューラルネットワーク回路100は、表の従来例と比較して、電力のバランスが取れていることが挙げられる。従来例では、「Power (W)」に示すように、4.7W, 2.5Wと消費電力が大きい。消費電力が大きいので、これを回避する制御方法が複雑である。「Power (W)」に示すように、本実施形態では、従来例[Zhao et al.]と比較して消費電力を2.3Wと約1／2に低減することができた。

[0061] <チップ面積>

本実施形態の2値化ニューラルネットワーク回路100は、バイアスが不要であること、乗算回路が2値論理ゲートであること、から、表の「LUTs, BRAM, DSP Block」に示すように、チップ面積が46900から14509と約1／3に減り、外付けメモリが不要となる、メモリコントローラが単純になることなどの効果がある。チップ面積は価格に比例するので、価格も2桁程度安くなることが期待できる。

[0062] <性能等価>

本実施形態の2値化ニューラルネットワーク回路100は、表の「FPS/Watt, FPS/ LUT, FPS/BRAM」に示すように、性能パワー効率、面積を見ずにパワー効率だけを見たものでも35.7と182.6と約5倍となっている。さらに、データ転送速度も168と120と約3倍となっている。

[0063] [実装例]

図12は、本発明の実施形態に係る2値化ニューラルネットワーク回路の実装例を説明する図である。

<STEP1>

まず、与えられたデータセット（今回はImageNet、画像認識タスク用にデータ）を既存のディープニューラルネットワーク用のフレームワークソフトウェアであるChainer（登録商標）を用いてGPU（Graphics Processing Unit）を有するコンピュータ201上で学習を行った。学習は、GPU上で実行する。このコンピュータ201は、ARMプロセッサなどのCPU（Central Processing Unit）と、メモリと、ハードディスクなどの記憶手段（記憶部）と、ネットワークインタフェースを含むI/Oポートとを有する。このコンピュータは、CPU201が、メモリ上に読み込んだプログラム（2値化したニューラルネットワークの実行プログラム）を実行することにより、後記する各処理部により構成される制御部（制御手段）を動作させる。

[0064] <STEP2>

次に、自動生成ツールを用いて、本実施形態の2値化ニューラルネットワーク回路100と等価なC++コードを自動生成し、C++コード202を得た。

[0065] <STEP3>

次に、FPGAベンダの高位合成ツール(Xilinx社SDSoC)（登録商標）を用いて、FPGA（field-programmable gate array）合成用にHDL（hardware description language）203を生成した。例えば、高位合成ツール(Xilinx社SDSoC)では、実現したい論理回路をハードウェア記述言語（Verilog HDL/VHDL）を用いて記述し、提供されたCADツールでビットストリームに合成する。そして、FPGAにこのビットストリームを送信するとFPGAに回路が実現する。

[0066] <STEP4>

次に、従来のFPGA合成ツールVivado（登録商標）を用いて、FPGA上に実現（FPGA合成204）して画像認識タスクの検証を行った。

[0067] <STEP5>

検証後、基板205を完成させた。基板205には、2値化ニューラルネットワーク回路100がハードウェア化されて実装されている。

[0068] 以上説明したように、本実施形態に係る2値化ニューラルネットワーク回路100(図9参照)は、入力値 $x_1 \sim x_n$ (x_i)(2値)を入力する入力ノードおよび重み $w_1 \sim w_n$ (w_i)(2値)を入力する入力部101と、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR論理を取るXNORゲート回路102と、各XNOR論理値の総和を取る総和回路103と、2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路41と、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換する活性化関数回路35と、を備える。総和を取った信号Yは、式(6)で示される。

[0069] また、ニューラルネットワーク処理方法では、入力値 $x_1 \sim x_n$ (x_i)および重み $w_1 \sim w_n$ (w_i)を入力するステップと、入力値 $x_1 \sim x_n$ および重み $w_1 \sim w_n$ を受け取り、XNOR論理を取るステップと、式(6)で示されるように、各XNOR論理値の総和のみを取るステップと、バッチ正規化を行うステップと、総和を取った信号Yのバッチ正規化した信号Bを活性化関数 $f_{\text{sgn}}(B)$ で変換するステップと、を実行する。

[0070] 学習時にバイアス項を不要とする2値化CNNで学習を行い、回路実現時もバイアス項を不要とする。

[0071] これにより、バイアス項を保持するメモリ・加算器回路が不要であり、CNNを2値の入力・重みとバッチ正規化回路41で実現できる。このため、認識精度をほとんど落とすことなく面積・消費電力・速度において優れるCNNを実現できる。

例えば、図11の表に示すように、本実施形態に係る2値化ニューラルネットワーク回路100は、消費電力(Power [W])を半分に削減でき、面積を約30分の1(FPS/Watt, FPS/LUT, FPS/BRAMの相乗効果)に削減できた。

[0072] 本実施形態では、2値化ニューラルネットワーク回路40(図6参照)で必須であったバイアス自体を不要とすることから面積が削減でき、メモリ量も削減できる。また、図10(b)(c)を比較してわかるように、本実施形態の2値化ニューラルネットワーク回路100は、認識精度について差が

ない。

[0073] 図13は、各FPGA実装のハードウェアの量を比較して示す図である。ザイリンクス社のFPGAを用いて固定小数点精度、2値化、本実施形態がそれぞれ要求するハードウェアの量を計算した。具体的には、TensorFlowチュートリアルCNNを実装した。また、Digilent社のNetFPGA-1G-CMLボード上に実装した。

各FPGA実装の比較は、本実施形態（2値化(バッチ正規化+バイアスなし)）、比較例の固定小数点(16ビット)、および比較例の2値化(バイアスのみ)である。FPGA実装のハードウェアの量は、FF(flip-flop)数、LUT数、18Kb BRAM数、およびDSP(digital signal processor) 48E数で示される。

[0074] 図13に示すように、本実施形態（2値化(バッチ正規化+バイアスなし)）および2値化(バイアスのみ)は、固定小数点精度に対して、FF数、LUT数、18Kb BRAM数、およびDSP48E数の、いずれのハードウェア量も減少していることが確かめられた。また、本実施形態（2値化(バッチ正規化+バイアスなし)）は、2値化(バイアスのみ)に対してもハードウェア量（面積）が1～2%程度の増加で済んでいることが分かる。

[0075] さらに、本実施形態の効果について述べる。

（1）バッチ正規化なしのCNNとの比較

本実施形態は、バッチ正規化なしのCNNと比較して、バイアス項回路が不要になる、かつ、学習時にバイアスが不要になるので学習が容易になる利点がある。ただし、バッチ正規化項用の回路が必要である。また、図13の本実施形態（2値化(バッチ正規化+バイアスなし)）と比較例の2値化(バイアスのみ)との比較から分かるように、ハードウェア量（面積）、電力が数%増加する。

[0076] （2）学習時の比較

前記図10B、Cに示すように、本実施形態の2値化ニューラルネットワーク回路100の認識の精度（図10C参照）と、比較例の「バッチ正規化

あり」（バイアス項あり）で構成した２値化ニューラルネットワーク回路４０の認識精度（図１０Ｂ参照）、および学習時間に差はなく、ほぼ同じとみてよいことが分かる。なお、回路実現時の差異は、図１３で示されている。

[0077] （３）設計容易性

バイアス値がある場合、その値が極めて小さい値（固定小数点３０～４０ビット）であるので回路設計に留意すべきである。

[0078] 本実施形態によれば、既存のバイアスを有する２値化ニューラルネットワーク回路と比較して、消費電力を半分に削減でき、面積を約３０分の１に削減（図１１参照）しつつ、図１０に示したように認識精度はほぼ等価なCNNを構成できることが判明した。ディープラーニングを用いたADAS（Advanced Driver Assistance System：先進運転支援システム）カメラ画像認識用のエッジ組み込み装置ハードウェア方式として実用化が期待される。特にADASでは、車載する上で高信頼性と低発熱が要求される。本実施形態に係る２値化ニューラルネットワーク回路１００は、図１１の表に示すように、消費電力（Power [W]）が格段に低減していることに加え、外付けメモリが不要であるので、メモリを冷却する冷却ファンや冷却フィンも不要である。ADASカメラに搭載して好適である。

[0079] [変形例]

図１４は、変形例１のディープニューラルネットワークの２値化ニューラルネットワーク回路の構成を示す図である。図９と同一構成部分には同一符号を付して重複箇所の説明を省略する。

変形例１の２値化ニューラルネットワーク回路１００Ａは、図１のニューラルネットワーク回路２に適用できる。

２値化ニューラルネットワーク回路１００Ａ（ニューラルネットワーク回路装置）は、バイアスが不要な２値化ニューラルネットワーク回路である。

図１４に示すように、２値化ニューラルネットワーク回路１００Ａは、図９の２値化ニューラルネットワーク回路１００にさらに、バイアス値を記憶するバイアスメモリ１１０（記憶部、バイアス値入力部）を備えて構成され

る。

[0080] 2値化ニューラルネットワーク回路100Aは、バイアスメモリ110に記憶されたバイアス値を読み出して総和回路103に出力する。この場合、2値化ニューラルネットワーク回路100Aは、バイアスを用いる2値化ニューラルネットワーク回路40（図6参照）と同様に、総和回路103にバイアス値が入力される構成となる。総和回路35は、XNORゲート回路34の各XNOR論理値とバイアス値との総和を取ることになり、2値化ニューラルネットワーク回路100Aは、バイアスを用いる2値化ニューラルネットワーク回路40（図6参照）と等価なニューラルネットワーク処理方法を実現できる。

一方、2値化ニューラルネットワーク回路100Aは、バイアスメモリ110に記憶されたバイアス値の読み出しに代えて、総和回路103に0を書き込むことで、式（6）の回路を実行する2値化ニューラルネットワーク回路100Aのニューラルネットワーク処理方法を実現できる。この場合、2値化ニューラルネットワーク回路100Aは、バイアスを不要とする2値化CNNである。

[0081] 変形例1によれば、2値化ニューラルネットワーク回路100Aを、バイアスを用いる既存の2値化ニューラルネットワーク回路40（図6参照）に置き換え（流用して）使用することができ、汎用的に適用可能である。特に、バイアスを用いる既存の2値化ニューラルネットワーク回路40そのもの、またバイアスを用いる既存の2値化ニューラルネットワーク回路40に接続されるメモリや読出し／書込み制御部の設計変更や検証を行うことなく適用できる。また、既存の2値化ニューラルネットワーク回路40で蓄積した資源を活用することができる。

[0082] 図15は、変形例2のディープニューラルネットワークの2値化ニューラルネットワーク回路の構成を示す図である。図9と同一構成部分には同一符号を付して重複箇所の説明を省略する。

変形例2の2値化ニューラルネットワーク回路100Bは、図1のニュー

ラルネットワーク回路 2 に適用できる。

2 値化ニューラルネットワーク回路 1 0 0 B (ニューラルネットワーク回路装置) は、バイアスが不要な 2 値化ニューラルネットワーク回路である。

図 1 5 に示すように、2 値化ニューラルネットワーク回路 1 0 0 B は、図 9 の 2 値化ニューラルネットワーク回路 1 0 0 にさらに、バイアス値 B を入力するバイアス値入力部 1 2 0 と、バイアス値 B の総和回路 1 0 3 への入力をオンオフするスイッチ 1 2 1 と、を備えて構成される。

[0083] 2 値化ニューラルネットワーク回路 1 0 0 B は、バイアスメモリ 1 1 0 に記憶されたバイアス値を読み出して総和回路 1 0 3 に出力する。この場合、2 値化ニューラルネットワーク回路 1 0 0 B は、バイアスを用いる 2 値化ニューラルネットワーク回路 4 0 (図 6 参照) と同様に、総和回路 1 0 3 にバイアス値 B が入力される構成となる。総和回路 3 5 は、X N O R ゲート回路 3 4 の各 X N O R 論理値とバイアス値 B との総和を取ることになり、2 値化ニューラルネットワーク回路 1 0 0 B は、バイアスを用いる 2 値化ニューラルネットワーク回路 4 0 (図 6 参照) と等価なニューラルネットワーク処理方法を実現できる。

一方、2 値化ニューラルネットワーク回路 1 0 0 B は、スイッチ 1 2 1 をオフにすることで、バイアス値 B の総和回路 1 0 3 への入力をオフし、式 (6) の回路を実行する 2 値化ニューラルネットワーク回路 1 0 0 A のニューラルネットワーク処理方法を実現できる。

[0084] 変形例 2 によれば、2 値化ニューラルネットワーク回路 1 0 0 B を、バイアスを用いる既存の 2 値化ニューラルネットワーク回路 4 0 (図 6 参照) に置き換え (流用して) 使用することができ、変形例 1 の場合と同様の効果を得ることができる。変形例 2 は、変形例 1 の 2 値化ニューラルネットワーク回路 1 0 0 A に比べ構成がより簡素であり、さらに汎用的に適用できる。

[0085] 本発明は上記の実施形態例に限定されるものではなく、特許請求の範囲に記載した本発明の要旨を逸脱しない限りにおいて、他の変形例、応用例を含む。

例えば、乗算回路としての論理ゲートに代えて、LUT (Look-Up Table) を用いてもよい。このLUTは、XNOR論理を行うXNORゲート回路102 (図9参照) に代えて、FPGAの基本構成要素であるルックアップテーブルを用いる。LUTは、2入力 ($x1$, $w1$) に対する2値 ($-1/+1$) のXNOR論理結果Yを格納する。LUTを用いることで、バッチ正規化回路の面積とパラメータを格納するメモリ面積・メモリ帯域を無くすることができ、かつ、性能的には等価な回路構成を実現することができる。LUTは、FPGAの基本構成要素であり、FPGA合成の際の親和性が高く、FPGAによる実装が容易である。

[0086] また、上記した実施形態例は本発明をわかりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。また、ある実施形態例の構成の一部を他の実施形態例の構成に置き換えることが可能であり、また、ある実施形態例の構成に他の実施形態例の構成を加えることも可能である。また、実施形態例は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形例は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

[0087] また、上記実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上述文書中や図面中に示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散

・統合して構成することができる。

[0088] また、上記の各構成、機能、処理部、処理手段等は、それらの一部または全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行するためのソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリや、ハードディスク、SSD (Solid State Drive) 等の記録装置、または、IC (Integrated Circuit) カード、SD (Secure Digital) カード、光ディスク等の記録媒体に保持することができる。

また、上記実施の形態では、装置は、ニューラルネットワーク回路装置という名称を用いたが、これは説明の便宜上であり、名称はディープニューラルネットワーク回路、ニューラルネットワーク装置、パーセプトロン等であってもよい。また、方法およびプログラムは、ニューラルネットワーク処理方法という名称を用いたが、ニューラルネットワーク演算方法、ニューラルネットプログラム等であってもよい。

符号の説明

- [0089]
- 1 ディープニューラルネットワーク
 - 2 ニューラルネットワーク回路
 - 1 1 入力層
 - 1 2 隠れ層 (中間層)
 - 1 3 出力層
 - 3 5 活性化回路 (活性化回路部, 活性化回路手段)
 - 4 1 バッチ正規化回路
 - 4 2 減算器
 - 4 3 第1乗算回路
 - 4 4 第2乗算回路
 - 4 5 加算器
 - 1 0 0, 1 0 0 A, 1 0 0 B 2値化ニューラルネットワーク回路 (ニュー

ーラルネットワーク回路装置)

1 0 1 入力部

1 0 2 XNORゲート回路(論理回路部, 論理回路手段)

1 0 3 総和回路(総和回路部, 総和回路手段)

1 1 0 バイアスメモリ(記憶部, バイアス値入力部)

1 2 0 バイアス値入力部

1 2 1 スイッチ

$x_1 \sim x_n (x_i)$ 入力値(2値)

$w_1 \sim w_n (w_i)$ 重み(2値)

請求の範囲

[請求項1] 入力層、1 以上の中間層、および、出力層を少なくとも含むニューラルネットワーク回路装置であって、

前記中間層の中で、2 値の入力値 x_i および重み w_i を受け取り、論理演算を行う論理回路部と、

前記論理回路部の出力の総和を取る総和回路部と、

2 値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路部と、

総和を取った信号 Y のバッチ正規化した信号 B を活性化関数 $f \operatorname{sgn}(B)$ で変換する活性化関数回路部と、を備え、

前記総和を取った信号 Y は、下記式で示される

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ただし、

γ : スケーリング係数

β : シフト値

μ'_B : バイアスを除く平均値。入力値 x_0 のときのバイアス値 w_0 とし、ミニバッチの平均値 μ_B とするとき、 $w_0 - \mu_B$

σ^2_B : ミニバッチの分散値

ϵ : 定数

ことを特徴とする記載のニューラルネットワーク回路装置。

[請求項2] 前記総和を取った信号 Y が、下記式 (4) で示される場合、

[数4]

$$Y = \sum_{i=0}^n w_i x_i - \mu_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta \quad \dots (4)$$

ただし、

μ_B : ミニバッチの平均値

入力値 $x_0 = 1$ 、このときのバイアス値 w_0 をもとに、前記式（４）を、下記式（５）に変形し、

[数5]

$$Y = \sum_{i=1}^n w_i x_i + (w_0 - \mu_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (5)$$

さらに、前記バイアス値 w_0 から前記ミニバッチの平均値 μ_B を減じた値を前記 μ'_B に置き換えて、下記式（６）とする

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ことを特徴とする請求項１に記載のニューラルネットワーク回路装置。

[請求項3]

バイアス値を入力するバイアス値入力部を備え、
前記総和回路部は、

前記論理回路部の出力と前記バイアス値との総和を取る

ことを特徴とする請求項１に記載のニューラルネットワーク回路装置。

[請求項4]

前記バイアス値を記憶する記憶部を備え、
前記バイアス値入力部は、

前記記憶部に記憶された前記バイアス値を読み出して前記総和回路部に出力するとともに、前記バイアス値の非使用時には前記総和回路部に０を書き込んで、下記式（６）の回路を実行する

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ことを特徴とする請求項３に記載のニューラルネットワーク回路装置。

[請求項5] 前記バイアス値の前記総和回路部への入力をオンオフするスイッチを備え、

前記バイアス値の非使用時には前記スイッチをオフして、下記式（6）の回路を実行する

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ことを特徴とする請求項3に記載のニューラルネットワーク回路装置。

[請求項6] 前記論理回路部は、否定排他的論理和または排他的論理和を含むことを特徴とする請求項1に記載のニューラルネットワーク回路装置。

[請求項7] 請求項1乃至6のいずれか1項に記載のニューラルネットワーク回路装置を備えるニューラルネットワーク。

[請求項8] 入力層、1以上の中間層、および、出力層を少なくとも含むニューラルネットワーク処理方法であって、

前記中間層の中で、2値の入力値 x_i および重み w_i を受け取り、論理演算を行うステップと、

論理回路部の出力の総和を取るステップと、

2値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するステップと、

総和を取った信号 Y のバッチ正規化した信号 B を活性化関数 $f \operatorname{sgn}(B)$ で変換するステップと、を有し、

前記総和を取った信号 Y は、下記式で示される

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \epsilon}}{\gamma} \beta) \quad \dots (6)$$

ただし、

γ : スケーリング係数

β : シフト値

μ'_B : バイアスを除く平均値。入力値 x_0 のときのバイアス値 w_0 とし、ミニバッチの平均値 μ_B とするとき、 $w_0 - \mu_B$

σ^2_B : ミニバッチの分散値

ε : 定数

ことを特徴とする記載のニューラルネットワーク処理方法。

[請求項9]

入力層、1 以上の中間層、および、出力層を少なくとも含むニューラルネットワーク回路装置としてのコンピュータを、

前記中間層の中で、2 値の入力値 x_i および重み w_i を受け取り、論理演算を行う論理回路手段、

論理回路部の出力の総和を取る総和回路手段、

2 値化によるバラツキの偏りを正規化範囲を広げ中心をシフトさせる処理で是正するバッチ正規化回路手段、

総和を取った信号 Y のバッチ正規化した信号 B を活性化関数 $f \operatorname{sgn}(B)$ で変換する活性化関数回路手段、

ただし、前記総和を取った信号 Y は、下記式で示される

[数6]

$$Y = \sum_{i=1}^n w_i x_i + (\mu'_B + \frac{\sqrt{\sigma_B^2 + \varepsilon}}{\gamma} \beta) \quad \dots (6)$$

γ : スケーリング係数

β : シフト値

μ'_B : バイアスを除く平均値。入力値 x_0 のときのバイアス値 w_0 とし、ミニバッチの平均値 μ_B とするとき、 $w_0 - \mu_B$

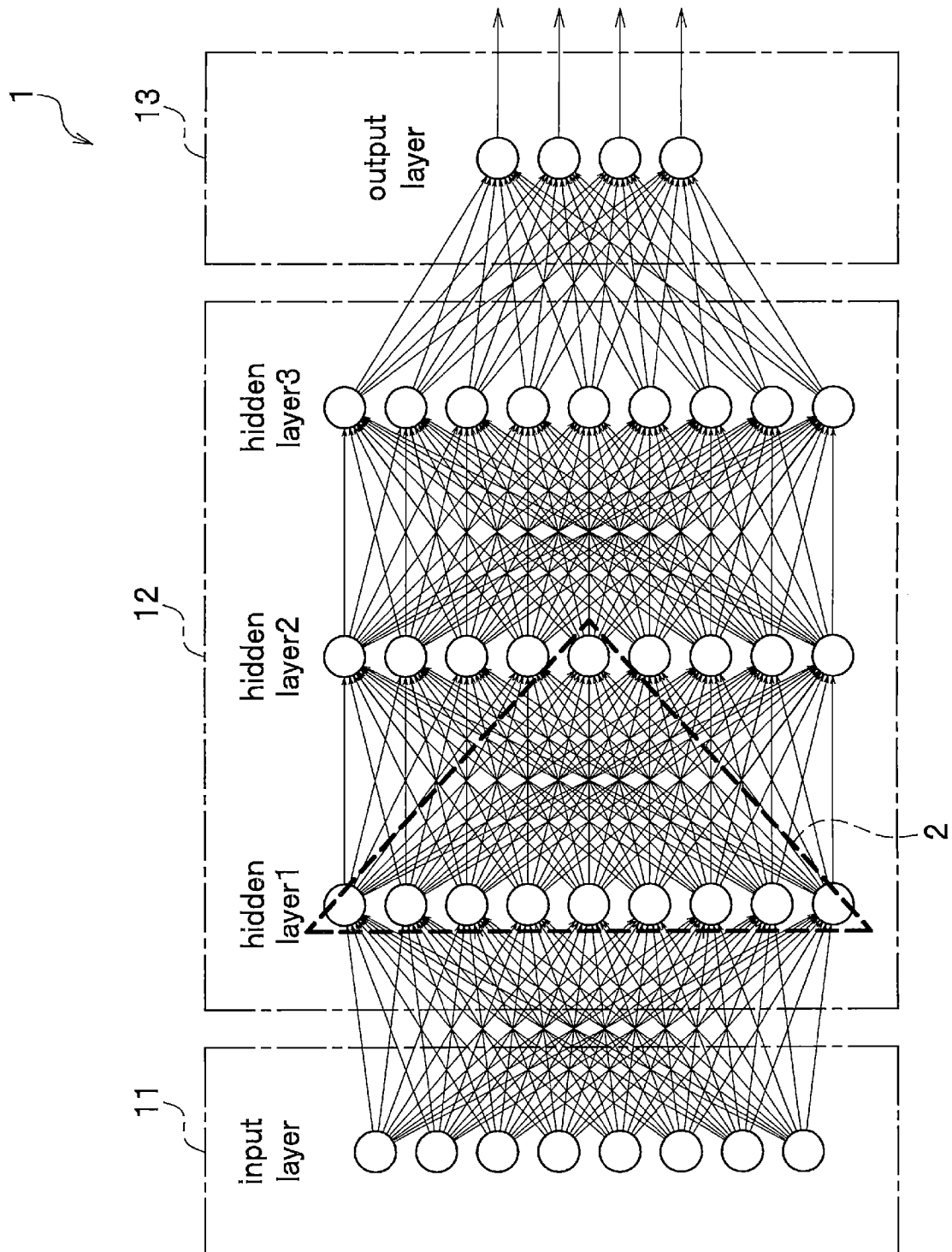
σ^2_B : ミニバッチの分散値

ε : 定数

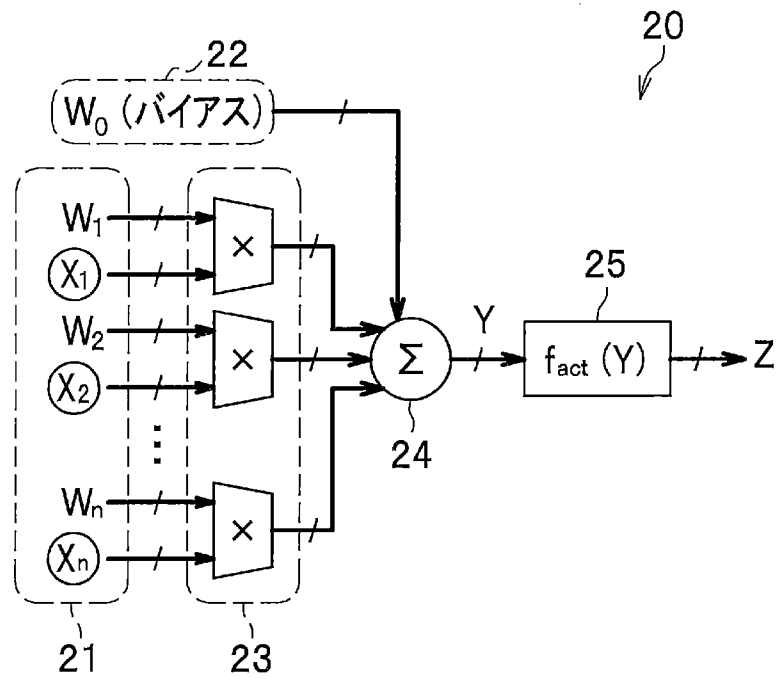
として機能させるためのニューラルネットワークの実行プログラム

。

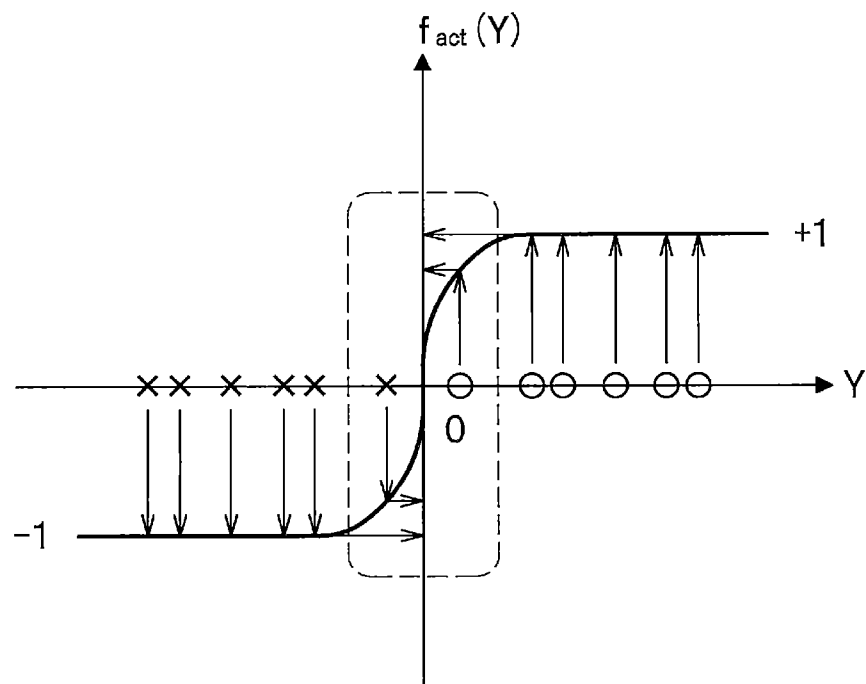
[図1]



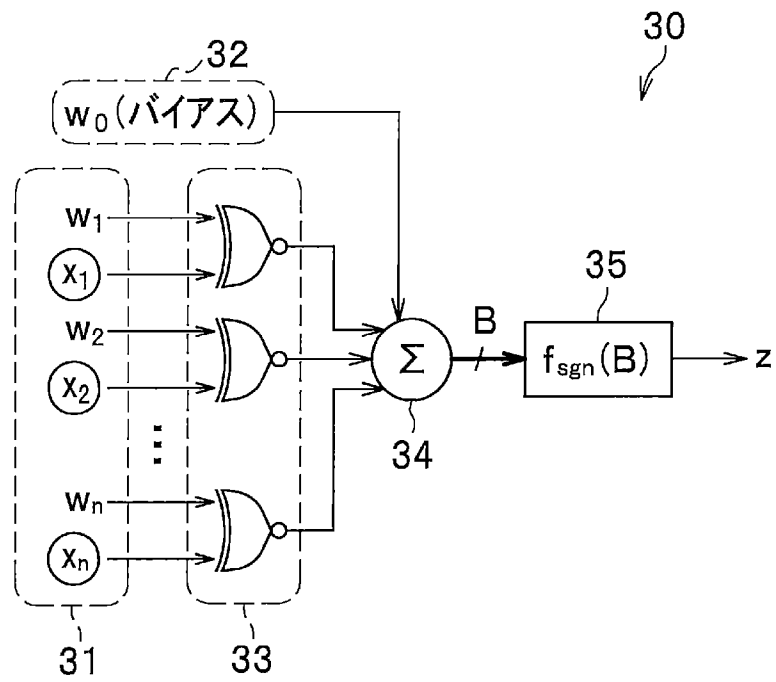
[図2]



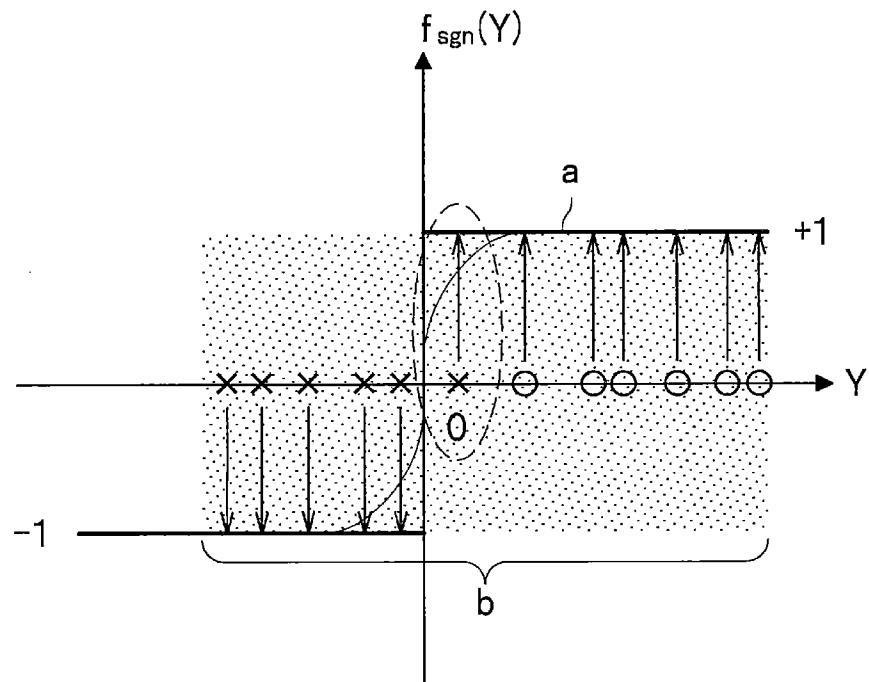
[図3]



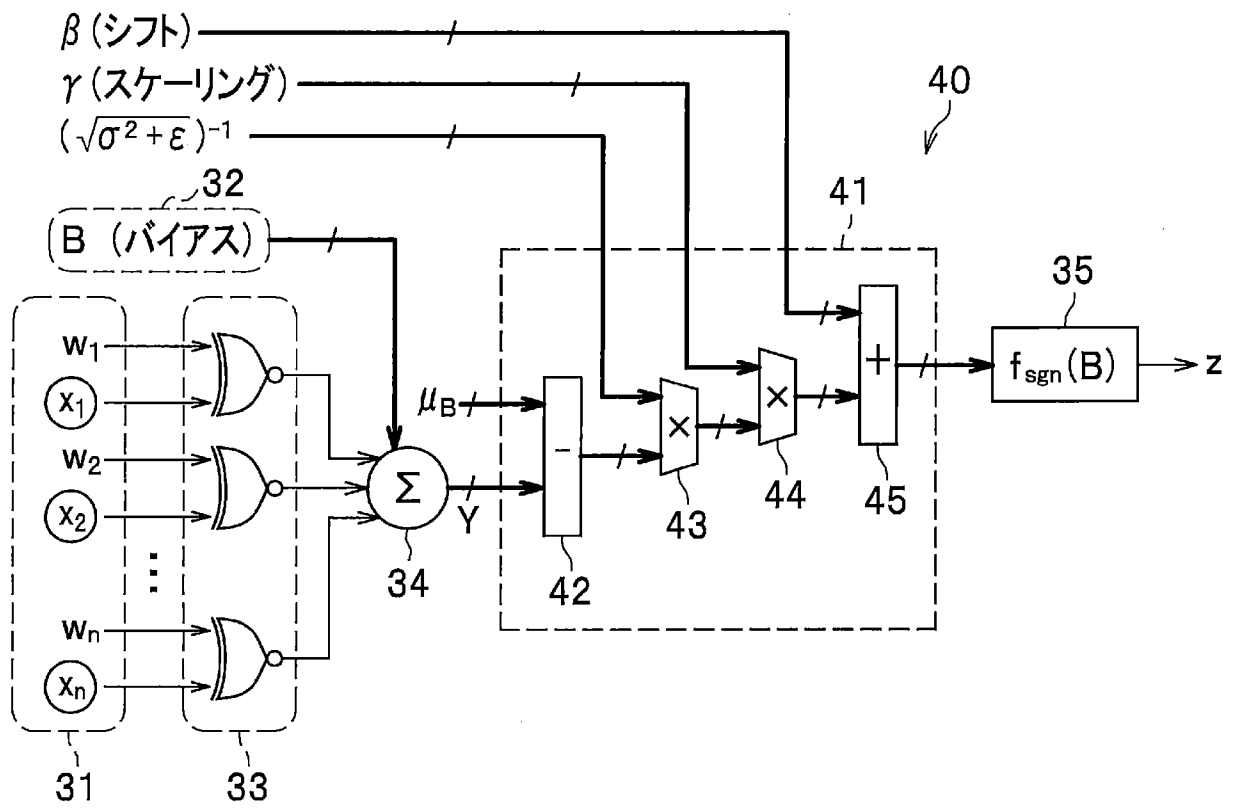
[図4]



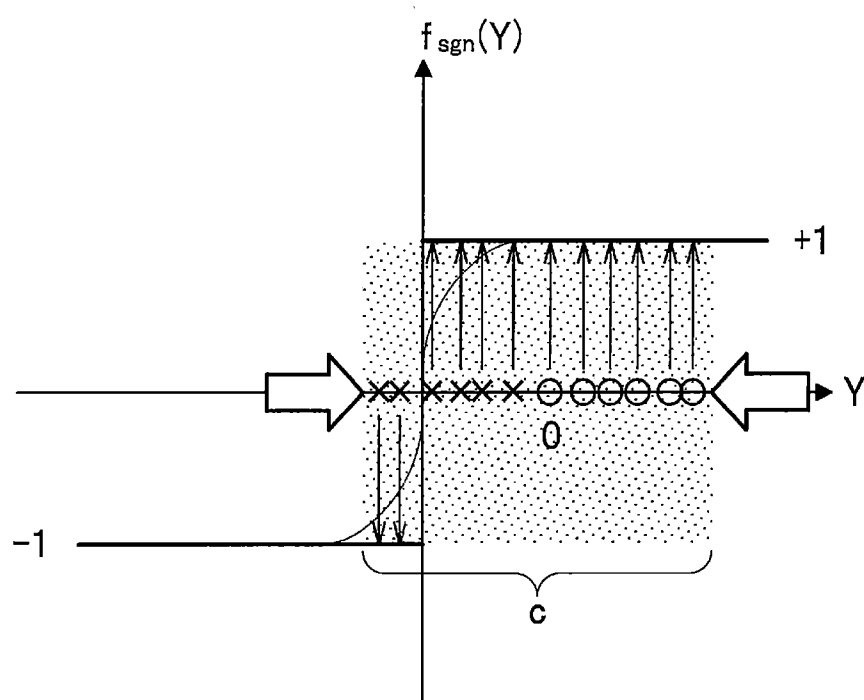
[図5]



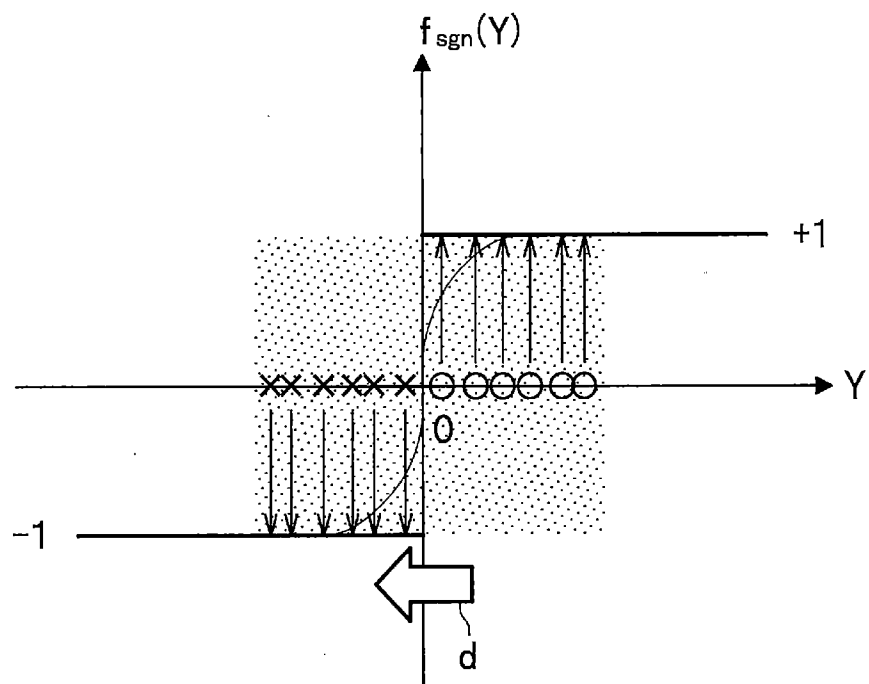
[図6]



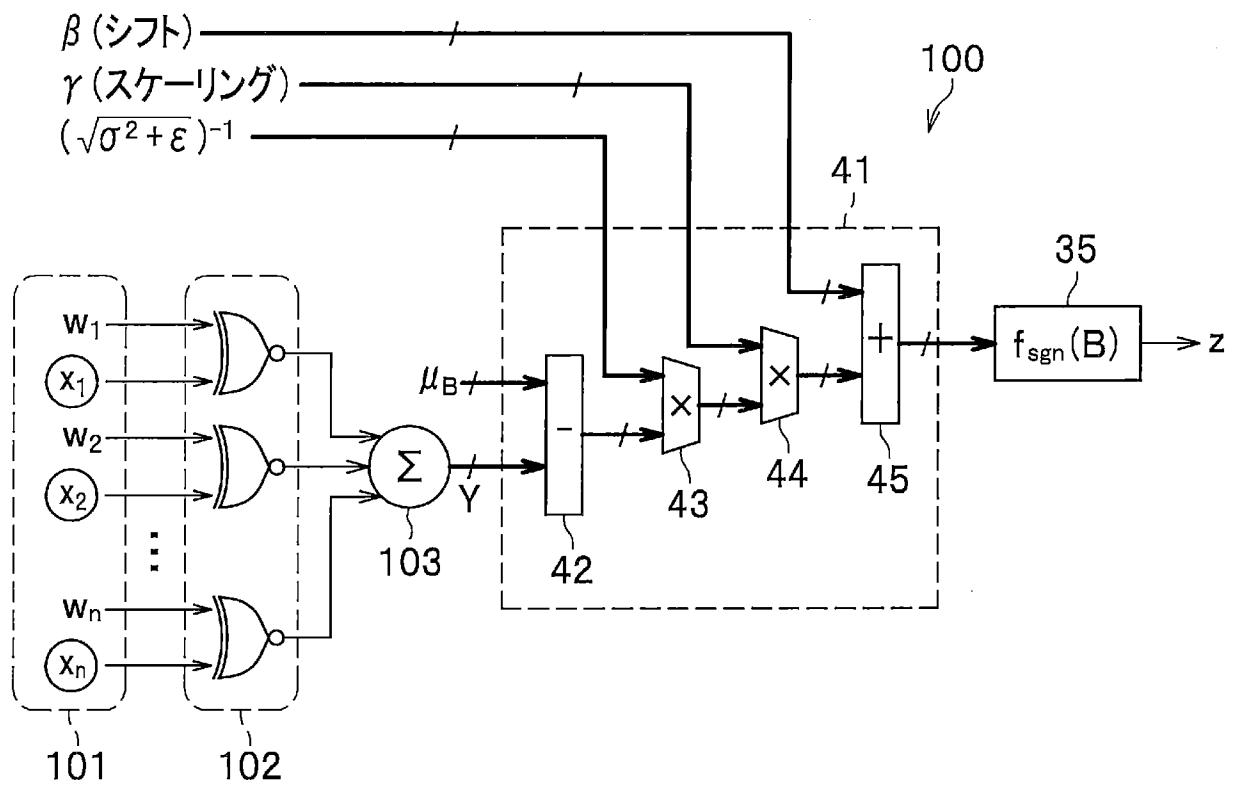
[図7]



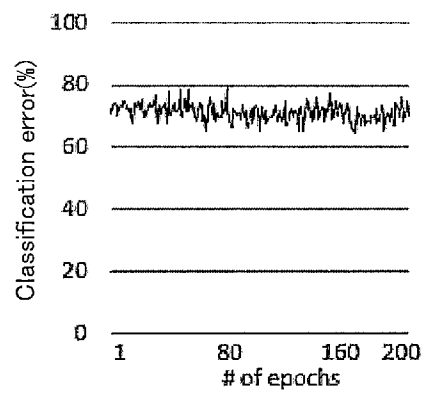
[図8]



[図9]

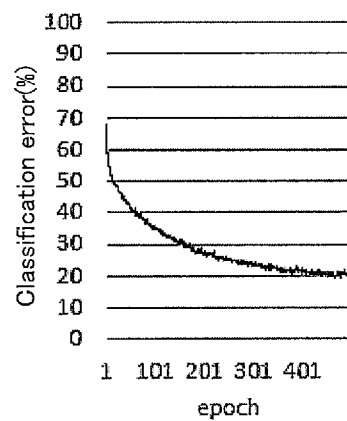


[図10A]



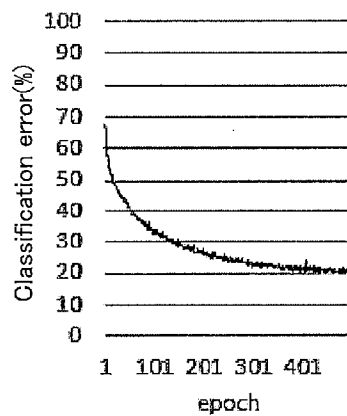
バッチ正規化なし
(バイアス項あり)

[図10B]



バッチ正規化あり
(バイアス項あり)

[図10C]



バッチ正規化あり
(バイアス項なし)

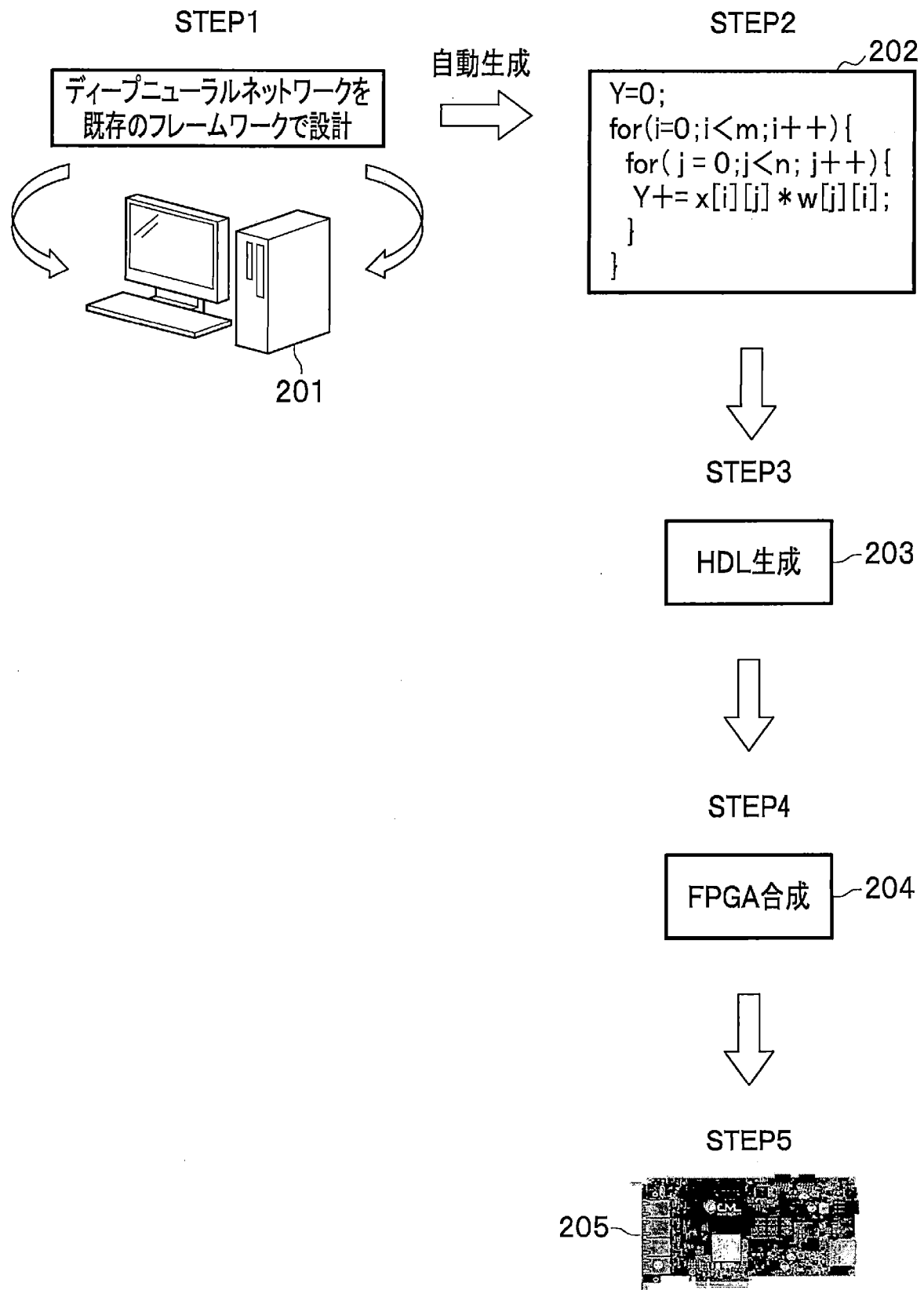
[図11]

Implementation (Year)	Zhao et al. (2017)	FINN (2017)	本実施形態
FPGA Board (FPGA)	Zedboard (XC7Z020)	PYNQ board (XC7Z020)	Zedboard (XC7Z020)
Clock (MHz)	143	166	143
# LUTs	46900	42823	14509
# 18Kb BRAMs	94	270	32
# DSP Blocks	3	32	1
Test Error	12.27%	19.9%	18.2%
Time [msec] (FPS)	5.94 (168)	2.24 (445)	2.37 (420)
Power [W]	4.7	2.5	2.3
FPS/Watt	35.7	178.0	182.6
FPS/LUT	35.8×10^{-4}	103.9×10^{-4}	289.4×10^{-4}
FPS/BRAM	1.8	1.6	13.1

[Zhao et al.] R. Zhao, W. Song, W. Zhang, T. Xing, J.-H. Lin, M. Srivastava, R. Gupta and Z. Zhang, "Accelerating Binarized Convolutional Neural Networks with Software-Programmable FPGAs," ISFPGA, 2017, pp.15-24.

[FINN] Y. Umuroglu, N. J. Fraser, G. Gambardella, M. Blott, P. Leong, M. Jahre, and K. Vissers, "FINN: A Framework for Fast, Scalable Binarized Neural Network Inference," ISFPGA, 2017.

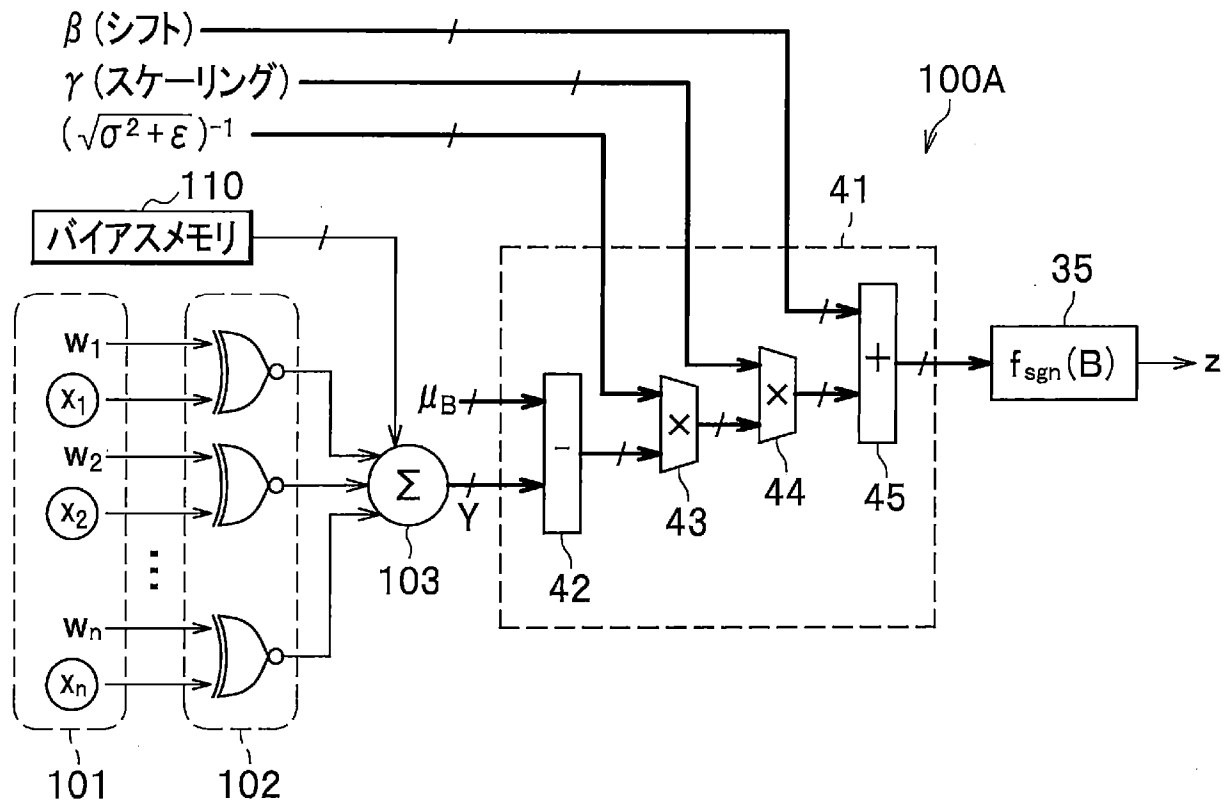
[図12]



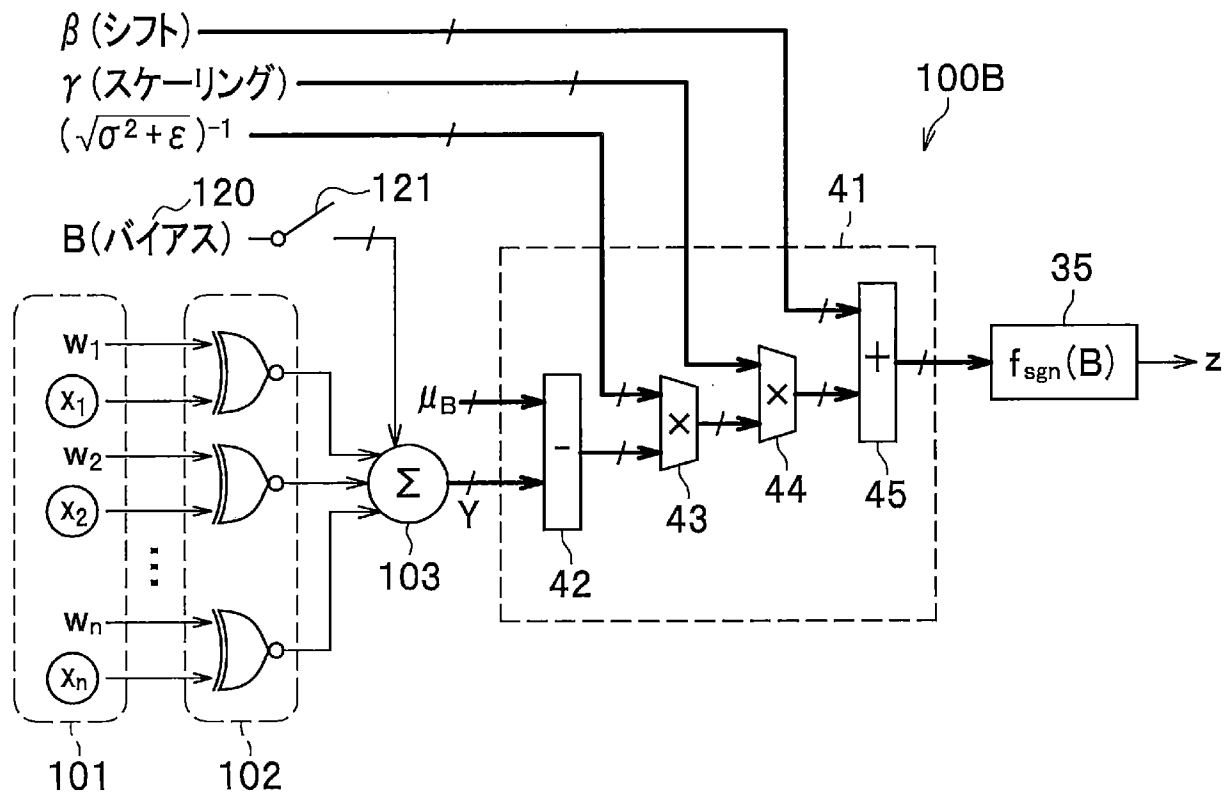
[図13]

	固定小数点 (16ビット)	本実施形態 2値化 (バッチ正規化+ バイアスなし)	2値化 (バイアスのみ)
FF 数	15,664	7,374	7,046
LUT 数	14,731	11,300	10,952
18kb BRAM 数	563	154	153
DSP 48E 数	52	30	26

[図14]



[図15]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/034522

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06N3/063 (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. G06N3/00-99/00

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 1971-2018

Registered utility model specifications of Japan 1996-2018

Published registered utility model applications of Japan 1994-2018

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

THE ACM DIGITAL LIBRARY

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	中原啓貴ほか, FPGA 向けディープラーニング開発環境 GUINNESS について, 電子情報通信学会技術研究報告, 18 September 2017, vol. 117, no. 221, pp. 51-56, (NAKAHARA, Hiroki et al., "GUINNESS : A GUI based Binarized Deep Neural Network Framework for an FPGA", IEICE technical report)	1-9
P, A	WO 2018/101275 A1 (TOKYO INSTITUTE OF TECHNOLOGY) 07 June 2018, entire text, all drawings & JP 2018-92377 A & JP 6183980 B1	1-9

☐ 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
14 November 2018 (14.11.2018)

Date of mailing of the international search report
27 November 2018 (27.11.2018)

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. G06N3/063(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. G06N3/00-99/00

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	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018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）

THE ACM DIGITAL LIBRARY

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	中原啓貴ほか, FPGA 向けディープラーニング開発環境 GUINNESS について, 電子情報通信学会技術研究報告, 2017.09.18, 第117巻, 第221号, pp.51-56	1-9
P, A	WO 2018/101275 A1 (国立大学法人東京工業大学) 2018.06.07, 全文, 全図 & JP 2018-92377 A & JP 6183980 B1	1-9

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

14.11.2018

国際調査報告の発送日

27.11.2018

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）

北元 健太

5 B

3 8 5 6

電話番号 03-3581-1101 内線 3545