

(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1月6日(06.01.2022)



(10) 国際公開番号

WO 2022/004389 A1

- (51) 国際特許分類:
G06N 3/04 (2006.01) G06N 3/08 (2006.01)
- (21) 国際出願番号: PCT/JP2021/022886
- (22) 国際出願日: 2021年6月16日(16.06.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2020-115034 2020年7月2日(02.07.2020) JP
- (71) 出願人: 株式会社ニコン (NIKON CORPORATION) [JP/JP]; 〒1086290 東京都港区港南二丁目15番3号 Tokyo (JP).
- (72) 発明者: 藤嶋 敏之 (FUJISHIMA, Toshiyuki); 〒1086290 東京都港区港南2丁目15番3号 株式会社ニコン内 Tokyo (JP).
- (74) 代理人: 青木 篤, 外 (AOKI, Atsushi et al.); 〒1050001 東京都港区虎ノ門一丁目23

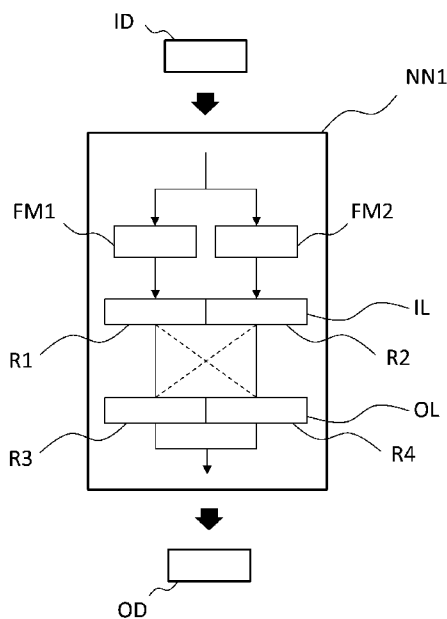
番1号 虎ノ門ヒルズ森タワー 青和特許法律事務所 Tokyo (JP).

- (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, 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,

(54) Title: CALCULATION DEVICE, LEARNING METHOD, AND NEURAL NETWORK CREATION SYSTEM

(54) 発明の名称: 演算装置、学習方法、およびニューラルネットワークの作成方法

図4



(57) Abstract: This calculation device comprises: a generation unit that generates, on the basis of input data, a first feature map and a second feature map in which a channel direction feature quantity is the same; and a calculation unit which causes a neural network to execute a calculation, said neural network comprising an input layer that has, in the channel direction, a first region to which the first feature map is inputted and a second region to which the second feature map is inputted, and an output layer which has, in the channel direction, a third region to which a calculation result corresponding to the first region is outputted and a fourth region to which a calculation result corresponding to the second region is outputted. The third region is linked to the first region and at least part of the second region, and furthermore, the calculation coefficient for the link from the second region to the third region is essentially set to zero. The fourth region is linked to the second region and at least part of the first region, and furthermore, the calculation coefficient for the link from the first region to the fourth region is essentially set to zero.

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：演算装置は、入力データに基づいて、チャネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する生成部と、第1特徴マップが入力される第1領域および第2特徴マップが入力される第2領域をチャネル方向に有する入力層と、第1領域に対応する演算結果が出力される第3領域および第2領域に対応する演算結果が出力される第4領域をチャネル方向に有する出力層とを備えるニューラルネットワークに演算を実行させる演算部と、を備え、第3領域は、第1領域および少なくとも第2領域の一部に結合されるとともに、第2領域から第3領域にいたる結合の演算係数は実質的にゼロに設定され、第4領域は、第2領域および少なくとも第1領域の一部に結合されるとともに、第1領域から第4領域にいたる結合の演算係数は実質的にゼロに設定される。

明 細 書

発明の名称：

演算装置、学習方法、およびニューラルネットワークの作成方法

技術分野

[0001] 本開示は、ニューラルネットワークを使用した演算装置、学習方法、およびニューラルネットワークの作成方法に関する。

背景技術

[0002] 従来、データの内容に基づく学習の結果に従って認識を行うニューラルネットワークの構成が提案されている（例えば非特許文献1を参照）。

先行技術文献

非特許文献

[0003] 非特許文献1：Kaiming He et. al., “Deep Residual Learning for Image Recognition,” arXiv:1512.03385, 2015

発明の概要

[0004] 本開示にかかる演算装置は、入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する生成部と、第1特徴マップが入力される第1領域および第2特徴マップが入力される第2領域をチャンネル方向に有する入力層と、第1領域に対応する演算結果が出力される第3領域および第2領域に対応する演算結果が出力される第4領域をチャンネル方向に有する出力層とを備えるニューラルネットワークに演算を実行させる演算部と、を備え、第3領域は、第1領域および少なくとも第2領域の一部に結合されるとともに、第2領域から第3領域にいたる結合の演算係数は実質的にゼロに設定され、第4領域は、第2領域および少なくとも第1領域の一部に結合されるとともに、第1領域から第4領域にいたる結合の演算係数は実質的にゼロに設定される。

[0005] 本開示にかかる学習方法は、入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する第1ステップ

と、第1特徴マップが入力される第1領域および第2特徴マップが入力される第2領域をチャンネル方向に有する入力層と、第1領域に対応する演算結果が出力される第3領域および第2領域に対応する演算結果が出力される第4領域をチャンネル方向に有する出力層とを備えるようニューラルネットワークを作成する第2ステップと、コンピュータが、入力データとして教師データをニューラルネットワークに入力し、入力層と出力層との結合の演算係数を更新する第3ステップと、を含み、第2ステップは、第3領域を第1領域および少なくとも第2領域の一部に結合するステップと、第4領域を第2領域および少なくとも第1領域の一部に結合するステップと、を含み、第3ステップは、第2領域から第3領域にいたる結合の演算係数および第1領域から第4領域にいたる結合の演算係数が実質的にゼロになるよう、演算係数を更新するステップを含む。

[0006] 本開示にかかる学習方法は、入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する第1ステップと、第1特徴マップが入力される第1領域および第2特徴マップが入力される第2領域をチャンネル方向に有する入力層と、第1領域に対応する演算結果が出力される第3領域および第2領域に対応する演算結果が出力される第4領域をチャンネル方向に有する出力層とを備えるようニューラルネットワークを作成する第2ステップと、コンピュータが、入力データとして教師データをニューラルネットワークに入力し、入力層と出力層との結合の演算係数を更新する第3ステップと、を含み、第2ステップは、第3領域を第1領域および少なくとも第2領域の一部に結合するステップと、第4領域を第2領域および少なくとも第1領域の一部に結合するステップと、を含み、第3ステップは、第2領域から第3領域にいたる結合の演算係数および第1領域から第4領域にいたる結合の演算係数を実質的にゼロに維持しながら、その他の結合の演算係数を更新するステップを含む。

[0007] 本開示にかかるニューラルネットワークの作成方法は、入力された特徴マップの特徴量に対し演算係数を用いて所定の演算を実行した結果を出力する

演算層をそれぞれ1以上備えるように第1演算経路および第2演算経路を作成するとともに、同一の演算を実行する演算層の位置が第1演算経路および第2演算経路において対応する位置となるように各演算層を配置するステップと、第1演算経路において第2演算経路に配置された演算層に対応する演算層が配置されていない位置、および、第2演算経路において第1演算経路に配置された演算層に対応する演算層が配置されていない位置に、対応する位置に配置された演算層と同一の演算を実行する補助演算層を追加するステップと、少なくともいずれか一方に補助演算層を追加された第1演算経路および第2演算経路に、それぞれ教師データを入力して演算層の演算係数を更新するステップと、演算係数が更新された第1演算経路および第2演算経路において対応するそれぞれの層をチャンネル方向に連結するステップと、を含む。

[0008] 本開示にかかるニューラルネットワークの作成方法は、入力された特徴マップの特徴量に対し演算係数を用いて所定の演算を実行した結果を出力する演算層をそれぞれ1以上備えるように第1演算経路および第2演算経路を作成するとともに、同一の演算を実行する演算層の位置が第1演算経路および第2演算経路において対応する位置となるように各演算層を配置するステップと、第1演算経路および第2演算経路にそれぞれ教師データを入力して演算層の演算係数を更新するステップと、第1演算経路において第2演算経路に配置された演算層に対応する演算層が配置されていない位置、および、第2演算経路において第1演算経路に配置された演算層に対応する演算層が配置されていない位置に、対応する位置に配置された演算層と同一の演算を実行する補助演算層を追加するステップと、少なくともいずれか一方に補助演算層が追加された第1演算経路および第2演算経路において対応するそれぞれの層をチャンネル方向に連結するステップと、を含む。

図面の簡単な説明

[0009] [図1]図1は、コンピュータの概略構成を示す模式図である。

[図2]図2は、プロセッサの機能ブロック図である。

[図3]図3は、生成部による特徴マップの生成を説明する図である。

[図4]図4は、ニューラルネットワークの構造を示す模式図である。

[図5]図5は、ニューラルネットワーク作成第1処理の第1フローチャートである。

[図6A]図6Aは、ニューラルネットワーク作成第1処理における第1の状態を示す図である。

[図6B]図6Bは、ニューラルネットワーク作成第1処理における第2の状態を示す図である。

[図6C]図6Cは、ニューラルネットワーク作成第1処理における第3の状態を示す図である。

[図7]図7は、演算経路における結合の第1の例を示す図である。

[図8]図8は、演算経路における結合の第2の例を示す図である。

[図9]図9は、補助演算層における活性化演算の概要を示す図である。

[図10]図10は、学習処理の動作フローチャートである。

[図11]図11は、ニューラルネットワーク作成第2処理のフローチャートである。

[図12]図12は、ニューラルネットワーク作成第1処理の第2フローチャートである。

[図13]図13は、ニューラルネットワーク作成第1処理の第3フローチャートである。

[図14]図14は、演算装置を搭載するカメラの構成を示す模式図である。

[図15]図15は、演算装置を搭載するカメラの処理フローチャートである。

発明を実施するための形態

[0010] 以下、図面を参照して演算装置、学習方法、およびニューラルネットワークの作成方法について詳細に説明する。ただし、本発明は図面または以下に記載される実施形態には限定されないことを理解されたい。

[0011] 図1はコンピュータの概略構成を示す模式図であり、図2はプロセッサの機能ブロック図である。

- [0012] 本実施形態において、演算装置の一例であるコンピュータ 1 は、所定の入力データに基づいて特徴マップを生成し、特徴マップをニューラルネットワークに入力して演算結果を出力する。そのために、コンピュータ 1 は、入出力インタフェース 11 と、メモリ 12 と、プロセッサ 13 とを備える。
- [0013] 入出力インタフェース 11 は、コンピュータ 1 が処理すべきデータを受け付け、または、コンピュータ 1 により処理されたデータを出力するためのインタフェース回路を有する。入出力インタフェース 11 は、例えばコンピュータ 1 を通信ネットワークに接続するための通信インタフェース回路、またはコンピュータ 1 をキーボード、ディスプレイといった各種周辺機器と接続するための周辺機器インタフェース回路を含む。入出力インタフェース 11 は、所定のデータの入力を受け付けて、処理結果を出力する。
- [0014] メモリ 12 は記憶部の一例であり、例えば、半導体メモリ、磁気ディスク装置および光ディスク装置のうちの少なくとも 1 つを有する。メモリ 12 は、プロセッサ 13 による処理に用いられるドライバプログラム、オペレーティングシステムプログラム、アプリケーションプログラム、データ等を記憶する。例えば、メモリ 12 はアプリケーションプログラムとして、本開示の演算処理を実現するためのコンピュータプログラムを記憶する。また、メモリ 12 はデータとして、ニューラルネットワークを規定するためのパラメータ群（層数、層構成、カーネル、重み係数等）を記憶する。各種プログラムは、コンピュータ読み取り可能な可搬型記録媒体から、公知のセットアッププログラム等を用いてメモリ 12 にインストールされてよい。コンピュータ読み取り可能な可搬型記録媒体とは、例えば C D - R O M (Compact Disc Read-Only Memory)、D V D - R O M (DVD Read-Only Memory) 等である。
- [0015] プロセッサ 13 は、1 以上のプロセッサおよびその周辺回路を備える。プロセッサ 13 は、コンピュータ 1 の全体的な動作を統括的に制御する処理回路であり、例えば、C P U (Central Processing Unit) である。プロセッサ 13 は、大量の定型的演算を並列に処理可能な G P U (Graphics Processing Unit) を有してもよい。プロセッサ 13 は、コンピュータ 1 の各種処理がメ

メモリ 12 に記憶されているプログラム等に基づいて適切な手段で実行されるように、入出力インタフェース 11 等の動作を制御する。プロセッサ 13 は、メモリ 12 に記憶されているプログラム（オペレーティングシステムプログラム、ドライバプログラム、アプリケーションプログラム等）に基づいて処理を実行する。また、プロセッサ 13 は、複数のプログラム（アプリケーションプログラム等）を並列に実行することができる。

[0016] 図 2 に示すように、プロセッサ 13 は、入力された入力データに基づいて特徴マップを生成するための生成部 131 と、ニューラルネットワーク NN1 による演算を実行するための演算処理部 132 を有する。演算処理部 132 は、プロセッサ 13 が有するプロセッサ上で実行されるプログラムによって実装される機能モジュールである。あるいは、演算処理部 132 は、独立した集積回路、マイクロプロセッサ、またはファームウェアとしてコンピュータ 1 に実装されてもよい。

[0017] 図 3 は、生成部 131 による特徴マップの生成を説明する図である。生成部 131 は、入力データ ID に基づいて、第 1 特徴マップ FM1 および第 2 特徴マップ FM2 を生成する。入力データ ID は、例えば画素ごとに RGB の対応値を有する画像データである。第 1 特徴マップ FM1 および第 2 特徴マップ FM2 において、チャンネル（C）方向の特徴量は共通している。すなわち、第 1 特徴マップ FM1 および第 2 特徴マップ FM2 において、対応するチャンネルにおける特徴量は同一である。

[0018] 生成部 131 は、入力データ ID に対し、畳み込み演算、転置畳み込み演算、プーリング演算、活性化演算といった所定の演算を行うことにより、第 1 特徴マップ FM1 および第 2 特徴マップ FM2 を生成する。生成部 131 は、例えば画像データである入力データ ID に対し、RGB のそれぞれについて画素ごとに畳み込み演算を実行することにより、第 1 特徴マップ FM1 および第 2 特徴マップ FM2 を生成する。また、生成部 131 は、入力データ ID と同一のデータを、第 1 特徴マップ FM1 および第 2 特徴マップ FM2 として生成してもよい。

- [0019] 図4は、ニューラルネットワークNN1の構造を示す模式図である。
- [0020] ニューラルネットワークNN1には、上述の生成部131により入力データIDに基づいて生成された第1特徴マップFM1および第2特徴マップFM2が入力される。そして、ニューラルネットワークNN1は、第1特徴マップFM1および第2特徴マップFM2に対し、畳み込み演算、転置畳み込み演算、プーリング演算、活性化演算のうち少なくとも1つを行い、出力データODを出力する。ニューラルネットワークNN1は、検出対象となる物体が表わされた複数の画像、および、各画像上における物体の位置、種類等を示す情報を教師データとして用いて、誤差逆伝搬法に従って学習される。ニューラルネットワークNN1は、例えば、畳み込みの演算経路とスキップ接続の演算経路とを有するResNetである。ニューラルネットワークNN1は、例えば画像データの入力に基づいて、画像データに表わされる物体を検出する。
- [0021] ニューラルネットワークNN1は、入力層ILと出力層OLとを備える。入力層ILは、チャンネル方向に、第1領域R1および第2領域R2を有する。第1領域R1には、入力データIDに基づいて生成された第1特徴マップFM1が入力される。第2領域R2には、入力データIDに基づいて生成された第2特徴マップFM2が入力される。
- [0022] 出力層OLは、チャンネル方向に、第3領域R3および第4領域R4を有する。第3領域R3は、第1領域R1と結合される。また、第3領域R3は、第2領域R2の少なくとも一部に結合される。このとき、第2領域R2から第3領域R3にいたる結合の演算係数は、実質的にゼロに設定される。
- [0023] 第4領域R4は、第2領域R2と結合される。また、第4領域R4は、第1領域R1の少なくとも一部に結合される。このとき、第1領域R1から第4領域R4にいたる結合の演算係数は、実質的にゼロに設定される。
- [0024] ニューラルネットワークNN1は、第3領域R3の出力および第4領域R4の出力に基づいて出力データODを生成し、出力する。
- [0025] 本明細書において、結合の演算係数が実質的にゼロに設定されるとは、前

の層の特徴量が後の層における演算の結果に影響を与えないように演算係数が設定されることをいい、詳細は後述する。

[0026] ニューラルネットワークNN1は、入力層ILと出力層OLとの間に、少なくとも1つ、好ましくは複数の中間層を有していてもよい。

[0027] 図5は、ニューラルネットワーク作成第1処理の第1フローチャートであり、図6(a)－(c)は、ニューラルネットワーク作成第1処理におけるニューラルネットワークの状態を示す図である。コンピュータは、図5に示す各ステップの処理を実行することにより、ニューラルネットワークNN1を作成することができる。

[0028] まず、コンピュータは、演算層をそれぞれ1以上備える第1演算経路P1および第2演算経路P2を作成する。そして、同一の演算を実行する演算層の位置が第1演算経路P1および第2演算経路P2において対応する位置となるように各演算層を配置する（ステップS11）。

[0029] 図6Aは、3つの演算層L11－L13を有する第1演算経路P1と、2つの演算層L21－L22を有する第2演算経路P2とが作成され、各演算層が配置された状態を示している。

[0030] 第1演算経路P1は、畳み込み演算を実行する演算層L11およびL12と、活性化演算を実行する演算層L13とを有する。第1演算経路P1には、 $W \times H \times C_i$ の次元を有する第1特徴マップFM1が入力される（Wは幅、Hは高さ、Cはチャンネルを表す）。演算層L11は、入力された第1特徴マップFM1に対して、例えば 3×3 のカーネルを用いた畳み込み演算を実行し、 $W \times H \times C_m$ の次元を有する特徴マップを出力する。演算層L12は、演算層L11から出力された特徴マップに対して、例えば 3×3 のカーネルを用いた畳み込み演算を実行し、 $W \times H \times C_o$ の次元を有する特徴マップを出力する。演算層L13は、演算層L12から出力された特徴マップに対して、例えばReLU関数による活性化演算を実行し、 $W \times H \times C_o$ の次元を有する特徴マップを出力する。

[0031] 第2演算経路P2は、畳み込み演算を実行する演算層L21およびL22

を有する。第2演算経路P2には、 $W \times H \times C_i$ の次元を有する第2特徴マップFM2が入力される。演算層L21は、入力された第2特徴マップFM2に対して、例えば 3×3 のカーネルを用いた畳み込み演算を実行し、 $W \times H \times C_r$ の次元を有する特徴マップを出力する。演算層L22は、演算層L21から出力された特徴マップに対して、例えば 1×1 のカーネルを用いた畳み込み演算を実行し、 $W \times H \times C_o$ の次元を有する特徴マップを出力する。

[0032] 第1演算経路P1の出力および第2演算経路の出力の和は、ニューラルネットワークNN1の出力として次の処理に渡される。

[0033] 次に、コンピュータは、第1演算経路P1において第2演算経路P2に配置された演算層に対応する演算層が配置されていない位置、および、第2演算経路において第1演算経路P1に配置された演算層に対応する演算層が配置されていない位置に、補助演算層を追加する（ステップS12）。図6Aの例では、第1演算経路P1において、第2演算経路P2に配置された演算層L22に対応する演算層は配置されていない。また、第2演算経路P2において、第1演算経路P1に配置された演算層L11および演算層L13に対応する演算層は配置されていない。コンピュータは、図6Bに示すように、補助演算層FL1-FL3を追加する。補助演算層FL1は、演算層L11と同様の畳み込み演算を実行する演算層である。補助演算層FL2は、演算層L13と同様の活性化演算を実行する演算層である。補助演算層FL3は、演算層L22と同様の畳み込み演算を実行する演算層である。また、補助演算層FL1-3では、入力される特徴量と実質的に等しい特徴量を出力するように演算係数が設定される。

[0034] そして、コンピュータは、第1演算経路P1および第2演算経路P2において対応するそれぞれの層をチャネル方向に連結し（ステップS14）、ニューラルネットワークNN1を生成する。本開示において、「演算層をチャネル方向に連結する」とは、複数の演算層をチャネル方向に連続させて1つの演算層を構成することをいう。図6Cに示すように、対応する位置に配置

された演算層 L_{11} および補助演算層 FL_1 、演算層 L_{12} および演算層 L_{21} 、演算層 L_{13} および補助演算層 FL_2 、補助演算層 FL_3 および演算層 L_{22} が、それぞれチャンネル方向に連結される。第1演算経路 P_1 および第2演算経路 P_2 の対応する位置で連結された各層は、入力された特徴マップに対してそれぞれ同一の演算を実行する。例えば、演算層 L_{11} と補助演算層 FL_1 とが連結された層は、第1特徴マップ FM_1 と第2特徴マップ FM_2 とが連結された特徴マップに対して 3×3 のカーネルを用いた畳み込み演算を実行し、 $W \times H \times (C_m + C_r)$ の次元を有する特徴マップを出力する。

[0035] チャンネル方向の連結において、第1演算経路 P_1 の演算層の少なくとも一部は、第2演算経路 P_2 の演算層の少なくとも一部に結合される。このとき、第1演算経路 P_1 から第2演算経路 P_2 にいたる結合の演算係数、および、第2演算経路 P_2 から第1演算経路 P_1 にいたる結合の演算係数は、実質的にゼロに設定される。

[0036] 図7は演算経路における結合の第1の例を示す図であり、図8は演算経路における結合の第2の例を示す図である。図7、図8に示す例では、演算経路における各ノードは、チャンネル方向に結合されている。

[0037] 図7は、第1特徴マップ FM_1 および第2特徴マップ FM_2 と、連結された演算層 L_{11} および補助演算層 FL_1 との結合の例を示している。演算層 L_{11} は第1演算経路 P_1 に含まれ、第1特徴マップ FM_1 が入力される。補助演算層 FL_1 は第2演算経路 P_2 に含まれ、第2特徴マップ FM_2 が入力される。第1特徴マップ FM_1 に含まれる一部のチャンネルは、第2特徴マップ FM_2 にも含まれている。演算層 L_{11} および補助演算層 FL_1 は、入力層 IL に対応する。

[0038] 演算層 L_{11} の各チャンネル $C_{m1} - C_{m4}$ には、第1特徴マップ FM_1 および第2特徴マップ FM_2 の各チャンネル $C_{i1} - C_{i16}$ が全結合される。また、補助演算層 FL_1 の各チャンネル $C_{r1} - C_{r8}$ には、第1特徴マップ FM_1 および第2特徴マップ FM_2 の各チャンネル $C_{i1} - C_{i16}$ が全結合

される。

[0039] 演算層L 1 1の各チャンネルC_{m 1}－C_{m 4}と第2特徴マップF M 2の各チャンネルのうち第1特徴マップF M 1に含まれないチャンネルC_{i 1 1}－C_{i 1 6}との結合の演算係数は、実質的にゼロに設定される。補助演算層F L 1の各チャンネルC_{r 1}－C_{r 8}と、対応するチャンネルC_{i 9}－C_{i 1 6}との結合以外の結合の演算係数は、実質的にゼロに設定される。このように演算係数を設定することにより、一部のチャンネル（C_{i 9}－C_{i 1 0}）が共通する第1特徴マップF M 1および第2特徴マップF M 2を、第1演算経路P 1および第2演算経路P 2に分岐させることができる。また、補助演算層F L 1は、第2演算経路の後続の演算層に対して、入力された第2特徴マップF M 2を変更せずに出力することができる。

[0040] 演算層で実行される畳み込み演算は、以下の式1により一般的に表現される。

[0041] [数1]

$$y_{ij}^{c_o} = \sum_{c_i, r, s} \{W \cdot x_{i+r, j+s}^{c_i}\} + b^{c_o}, \quad W = \{w_{rs}^{c_o c_i}\} \quad (\text{式1})$$

[0042] 式1において、演算層L 1 1および補助演算層F L 1で実行される畳み込み演算に用いられる演算係数であるウェイトWは、以下に示す式2のように設定される。

[0043] [数2]

$$W = \begin{pmatrix} W^{f1} \\ W^{h1} \end{pmatrix} \quad (\text{式2})$$

[0044] すなわち、演算層L 1 1における結合のウェイトはW^{f1}であり、補助演算層F L 1における結合のウェイトはW^{h1}である。W^{f1}のうち、第1特徴マップF M 1との結合のウェイトは学習により更新され、第2特徴マップF M 2のうち第1特徴マップF M 1に含まれないチャンネルとの結合のウェイトは実質的にゼロになるよう設定される。

[0045] また、W^{h1}のうち、第1特徴マップF M 1のうち第2特徴マップF M 2に含

まれないチャネルとの結合のウェイトは、実質的にゼロになるよう設定され、第2特徴マップFM2との結合のウェイトは、第2特徴マップFM2と補助演算層FL1との対応するチャネルどうし（例えばC_i21とC_r1）の結合のウェイトが所定値（例えば1）、他のチャネルとの結合のウェイトが実質的に0となるよう予め設定され、学習により更新されない。

[0046] W^{h1} を以下に示す式3のように設定することにより、補助演算層FL1における畳み込み演算において、第2特徴マップFM2と等しい値を補助演算層FL1から出力させることができる。なお、式3において、 δ はクロネッカーのデルタである。また、式3において、 r および s の値域は $-(k-1)/2 \sim (k-1)/2$ であり、 $c_r > c_o$ である。

[0047] [数3]

$$W^{h1} = \{\delta_{r0}\delta_{s0}\delta_{c_i c_r}\} \quad (\text{式3})$$

[0048] 図8は、連結された演算層L11および補助演算層FL1と、連結された演算層L12および演算層L21との結合の例を示している。演算層L11および演算層L12は第1演算経路P1に含まれる。補助演算層FL1および演算層L21は第2演算経路P2に含まれる。演算層L12および演算層L21は、入力層ILと出力層OLとの間の中間層に該当する。

[0049] 演算層L12の各チャネルC_o1、C_o2には、演算層L11の各チャネルC_m1－C_m4および補助演算層FL1の各チャネルC_r1－C_r8が全結合される。演算層L12の各チャネルC_o1、C_o2と補助演算層FL1の各チャネルC_r1－C_r8との結合の演算係数は、実質的にゼロに設定される。また、演算層L21の各チャネルC_r1－C_r8には、演算層L11の各チャネルC_m1－C_m4および補助演算層FL1の各チャネルC_r1－C_r8が全結合される。演算層L21の各チャネルC_r1－C_r8と演算層L11の各チャネルC_m1－C_m4との結合の演算係数は、実質的にゼロに設定される。

[0050] このとき、式1における演算係数であるウェイトWは、以下に示す式4のよ

うに設定される。

[0051] [数4]

$$W = \begin{pmatrix} W^{f2} & 0 \\ 0 & W^{h2} \end{pmatrix} \quad (\text{式4})$$

[0052] すなわち、演算層 L 1 1 と演算層 L 1 2 との結合のウェイトは W^{f2} であり、演算層 L 1 1 と演算層 L 2 1 との結合のウェイト、および、補助演算層 F L 1 と演算層 L 1 2 との結合のウェイトは 0 であり、補助演算層 F L 1 と演算層 L 2 1 との結合のウェイトは W^{h2} である。 W^{f2} は、演算層 L 1 1 と演算層 L 1 2 とがチャネル方向に全結合されるように設定され、学習により更新される。 W^{h2} は、予め設定され、学習により更新されない。

[0053] また、補助演算層 F L 3 と演算層 L 2 2 とが連結された演算層による畳み込み演算に用いられるウェイト W を以下に示す式 5 のように設定することにより、補助演算層 F L 3 の出力と演算層 L 2 2 の出力との和演算を実行することができる。すなわち、補助演算層 F L 3 の演算に用いられるウェイトを W^{h3} とし、演算層 L 2 2 の演算に用いられるウェイトを W^{f3} とする。式 5 において、 λ は、補助演算層 F L 3 の出力と演算層 L 2 2 の出力との和に要求されるウェイト比に応じて設定される係数テンソルである。 λ の値は 1 であってよい。

[0054] [数5]

$$W = (\lambda W^{f3} \quad W^{h3}) \quad (\text{式5})$$

[0055] 図 9 は、補助演算層 F L 2 における活性化演算の概要を示す図である。

[0056] 補助演算層 F L 2 は、演算層 L 1 3 と同様の活性化演算を実行し、演算係数が実質的にゼロに設定される。活性化演算について演算係数が実質的にゼロとするため、補助演算層 F L 2 は、入力側から順に、オフセット演算層 F L 2 1 と、活性化演算層 F L 2 2 と、逆オフセット演算層 F L 2 3 とを有する。

[0057] 補助演算層 F L 2 は、活性化演算層 F L 2 2 による活性化演算の直前にオ

フセット演算層FL21によるオフセット演算を実行し、活性化演算の直後に逆オフセット演算層FL23による逆オフセット演算を実行する。オフセット演算とは、入力値を活性化演算の前後で値が実質的に変化しない値域に変換する演算である。また、逆オフセット演算とは、オフセット演算により変換された値を元の入力値に変換する演算である。

[0058] 活性化演算層FL22がReLU関数による活性化演算を実行する場合、オフセット演算層FL21は、入力値に所定の正の値を加算することによりオフセット演算を実行することができる。このとき、逆オフセット演算層FL23は、オフセット演算により加算された所定の正の値を入力値から減算することにより逆オフセット演算を実行することができる。

[0059] また、オフセット演算層FL21は、例えば入力値に対するバッチ正規化を行うことにより、入力値の正規化を行ってもよい。ミニバッチにおける平均を μ 、標準偏差を σ とするとき、 σ を1、 μ を十分大きな値となるように正規化を行うことにより、ReLU関数による活性化演算に対するオフセット演算を実行することができる。また、 μ を0、 σ を十分大きな値となるように正規化を行うことにより、シグモイド関数による活性化演算に対するオフセット演算を実行することができる。逆オフセット演算層FL23は、オフセット演算層FL21における演算を打ち消す演算を実行する。

[0060] オフセット演算層FL21は、前の演算層において、畳み込み演算のバイアスを1以外の値に設定することにより実装されてもよい。同様に、逆オフセット演算層FL23は、後の演算層において、畳み込み演算のバイアスを1以外の値に設定することにより実装されてもよい。

[0061] また、活性化演算層FL22がシグモイド関数による活性化演算を実行する場合、オフセット演算層FL21は、入力値をそのまま出力してもよい。この場合、逆オフセット演算層FL23は、シグモイド関数の逆関数による演算を実行し、元の入力値を求める。

[0062] ニューラルネットワークNN1は、転置畳み込み演算を実行する演算層を有してもよい。転置畳み込み演算を実行する演算層に対応する補助演算層に

において、畳み込み演算の場合と同様に、対応するチャネルの特徴量に対するウェイト以外が0となるよう設定される。これにより、転置畳み込み演算を実行する演算層に対応する補助演算層は、入力された特徴マップをそのまま出力することができる。

[0063] ニューラルネットワークNN1は、プーリング演算を実行する演算層を有してもよい。プーリング演算を実行する演算層に対応する補助演算層では、プーリング演算における対象領域のサイズに応じたストライド（例えば2×2プーリングの場合、ストライド2）で1×1のカーネルを用いた畳み込み演算を実行する。

[0064] ニューラルネットワークNN1は、バッチ正規化演算を実行する演算層を有してもよい。バッチ正規化演算を実行する演算層に対応する補助演算層では、バッチ正規化演算において算出された μ 、 σ を、実際の値にかかわらずそれぞれ0、1とすることにより、入力された特徴マップをそのまま出力することができる。

[0065] ニューラルネットワークNN1では、第1演算経路P1に含まれる所定サイズ（例えば3×3）のカーネルを用いた畳み込み演算を実行する演算層と第2演算経路P2に含まれる異なるサイズ（例えば5×5）のカーネルを用いた畳み込み演算を実行する演算層が結合されてもよい。この場合、結合された演算層では、大きい方のカーネルを用いた畳み込み演算が実行され、小さい方のカーネルを用いる畳み込み演算が行われる演算経路では、カーネルの外側における演算係数がゼロに設定される。同様に、ダイレートの異なるカーネルを用いた畳み込み演算を実行する演算層を結合することも可能である。この場合、結合された演算層ではダイレートの大きいカーネルを使用し、ダイレートの小さいカーネルを用いる畳み込み演算では、不要な演算係数がゼロに設定される。

[0066] 以上の処理によって作成されたニューラルネットワークNN1において、演算層L11に対応するチャネル方向の領域は第1領域R1に対応し、補助演算層FL1に対応するチャネル方向の領域は第2領域R2に対応する。そ

して、チャンネル方向に連結された演算層 L 1 1 および補助演算層 F L 1 は、入力層 I L に対応する。また、ニューラルネットワーク N N 1 において、補助演算層 F L 3 に対応するチャンネル方向の領域は第 3 領域 R 3 に対応し、演算層 L 2 2 に対応するチャンネル方向の領域は第 4 領域 R 4 に対応する。そして、チャンネル方向に連結された補助演算層 F L 3 および演算層 L 2 2 は、出力層 O L に対応する。

[0067] 図 7、8 に示すように、補助演算層 F L 3 は、第 1 特徴マップ F M 1 および第 2 特徴マップ F M 2 に結合され、第 2 特徴マップ F M 2 との結合の演算係数は実質的にゼロに設定されている。同様に、演算層 L 2 2 は、第 1 特徴マップ F M 1 および第 2 特徴マップ F M 2 に結合され、第 1 特徴マップ F M 1 との結合の演算係数は実質的にゼロに設定されている。

[0068] 複数の演算経路を有するニューラルネットワークをコンピュータに実装するには、演算経路ごとに演算処理を実行する必要がある。そのため、汎用的な F P G A (Field-Programmable Gate Array) または並列計算エンジンにおけるこのようなニューラルネットワークの実装は、複雑なものとなる。本実施形態のニューラルネットワーク N N 1 は、複数の演算経路による演算を、連結された 1 つの演算経路によって実現することができるため、汎用的な F P G A または汎用計算エンジンに容易に実装することができる。また、本実施形態のニューラルネットワーク N N 1 は、保持が必要な演算の中間結果データのデータ量を、複数の演算経路のまま実装する場合よりも削減することができる。

[0069] 図 1 0 は、学習処理の動作フローチャートである。コンピュータは、図 1 0 に示す各ステップの処理を実行することにより、入力データに基づいてニューラルネットワークの演算係数を更新させ、入力データに基づく処理をニューラルネットワークに学習させることができる。

[0070] まず、コンピュータの生成部 1 3 1 は、教師データである入力データに基づいて、チャンネル方向の特徴量が共通する第 1 特徴マップ F M 1 および第 2 特徴マップ F M 2 を生成する (ステップ S 2 1)。

[0071] 次に、コンピュータは、入力層と出力層とを備えるニューラルネットワークを作成する（ステップS 2 2）。入力層は、第 1 特徴マップF M 1 が入力される第 1 領域R 1 および第 2 特徴マップF M 2 が入力される第 2 領域R 2 をチャンネル方向に有する。出力層は、第 1 領域R 1 に対応する演算結果が出力される第 3 領域R 3 および第 2 領域R 2 に対応する演算結果が出力される第 4 領域R 4 をチャンネル方向に有する。

[0072] 続いて、コンピュータは、第 1 特徴マップF M 1 および第 2 特徴マップF M 2 をニューラルネットワークN N 1 に入力する。そして、コンピュータは、第 3 領域R 3 および第 4 領域R 4 における出力値と教師データに示される正解との誤差が小さくなるように、例えば勾配降下法および誤差逆伝搬法といった手法を用いて、入力層と出力層との結合の演算係数を更新し（ステップS 2 3）、学習処理を終了する。

[0073] ステップS 2 3において、コンピュータは、第 2 領域R 2 から第 3 領域R 3 にいたる結合の演算係数および第 1 領域R 1 から第 4 領域R 4 にいたる結合の演算係数が実質的にゼロになるよう、演算係数を更新する。

[0074] また、ステップS 2 3において、コンピュータは、第 2 領域R 2 から第 3 領域R 3 にいたる結合の演算係数および第 1 領域R 1 から第 4 領域R 4 にいたる結合の演算係数が実質的にゼロに維持しながら、その他の結合の演算係数を更新してもよい。

[0075] 図 1 1 は、ニューラルネットワーク作成第 2 処理のフローチャートである。コンピュータは、図 1 0 に示す学習処理におけるステップS 2 2 が呼び出されるたびに図 1 1 に示されるフローチャートを実行する。

[0076] まず、コンピュータは、第 1 特徴マップF M 1 が入力される第 1 領域R 1 および第 2 特徴マップF M 2 が入力される第 2 領域R 2 をチャンネル方向に有する入力層を作成する（ステップS 3 1）。続いて、コンピュータは、第 3 領域R 3 および第 4 領域R 4 をチャンネル方向に有する出力層を作成する（ステップS 3 2）。

[0077] 次に、コンピュータは、第 3 領域R 3 を第 1 領域R 1 および少なくとも第

2領域R2の一部に結合する（ステップS33）。続いて、コンピュータは、第4領域R4を第2領域R2および少なくとも第1領域R1の一部に結合し（ステップS34）、ニューラルネットワーク作成第2処理を終了する。コンピュータは、ステップS31およびステップS32を逆の順序で実行してもよい。また、コンピュータは、ステップS33およびステップS34を逆の順序で実行してもよい。このように作成されたニューラルネットワークは、図4に表されるニューラルネットワークNN1に相当する。

[0078] 図12は、ニューラルネットワーク作成第1処理の第2フローチャートである。ニューラルネットワーク作成第1処理の第2フローチャートは、ニューラルネットワーク作成第1処理の第1フローチャートにおけるステップS12とステップS13との間にステップS43を追加したものに相当する。そのため、ニューラルネットワーク作成処理の第1フローチャートと共通するステップS41、S42、S44については説明を省略する。

[0079] コンピュータは、第1演算経路P1および第2演算経路P2にそれぞれ教師データを入力し、出力値と教師データに示される正解との誤差が小さくなるように、第1演算経路P1および第2演算経路P2における結合の演算係数を更新する（ステップS43）。このとき、第1演算経路P1に含まれる演算層および第2演算経路P2に含まれる演算層はチャンネル方向に連結されていないため、コンピュータは、演算係数の更新を第1演算経路P1および第2演算経路P2について独立に実行することができる。

[0080] 図13は、ニューラルネットワーク作成第1処理の第3フローチャートである。ニューラルネットワーク作成第1処理の第3フローチャートは、ニューラルネットワーク作成第1処理の第1フローチャートにおけるステップS11とステップS12との間にステップS52を追加したものに相当する。そのため、ニューラルネットワーク作成第1処理の第1フローチャートと共通するステップS51、S53、S54については説明を省略する。また、ステップS52の処理は、ニューラルネットワーク作成第1処理の第2フローチャートにおけるステップS43の処理と同様であるので、ステップS5

2についても説明を省略する。

[0081] 次に、演算装置をカメラに搭載した例について説明する。図14は、演算装置を搭載するカメラの構成を示す模式図である。

[0082] 本実施形態にかかる演算装置は、カメラ2に搭載されるマイクロコントローラ4である。カメラ2は、撮像装置の一例であり、イメージセンサ3、マイクロコントローラ4およびレンズ5を備える。イメージセンサ3は、撮像素子の一例であり、レンズ5で集光された被写体からの光を画像データに変換する。マイクロコントローラ4は、プロセッサ、メモリ、入出力インタフェースを集積した集積回路である。マイクロコントローラ4は、イメージセンサ3から入力された画像データに基づいて演算を実行し、オートフォーカスの対象とすべき画像データ中の領域を示す信号を出力する。また、マイクロコントローラ4は、合焦制御部として動作し、画像データに表される所定の領域に合焦するようにレンズ5を制御する合焦制御信号を出力する。レンズ5は、光学系の一例であり、被写体からの光をイメージセンサ3に集光させる。レンズ5は、オートフォーカス機構51を有し、オートフォーカス機構51は、マイクロコントローラ4から出力される合焦制御信号に基づいてレンズ5を被写体に合焦させる。

[0083] 本実施形態にかかる演算装置は、複数の演算経路による演算を行うResNetのようなニューラルネットワークを、少ないキャッシュメモリにより実装することができる。

[0084] 図15は、演算装置を備えたカメラの処理フローチャートである。カメラ2は、ユーザによる不図示のシャッターボタンの半押しなどのオートフォーカス実行を指示する動作の検出に応じて、図15の処理を開始する。カメラ2は、オートフォーカス実行を指示する動作が検出されている間、図15の処理を繰り返し実行してもよい。

[0085] まず、イメージセンサ3は、集光された光に応じた画像データをマイクロコントローラ4に送信する（ステップS101）。

[0086] 次に、マイクロコントローラ4は、受信した画像データに基づいて、チャ

ネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する（ステップS102）。続いて、マイクロコントローラ4は、第1特徴マップおよび第2特徴マップをニューラルネットワークに入力し、画像データに含まれる識別対象物に対応する領域を識別する演算を実行させる（ステップS103）。

[0087] 続いて、マイクロコントローラ4は、演算により識別された識別対象物に対応する領域に合焦させるための合焦制御信号を生成し、オートフォーカス機構51に送信する。オートフォーカス機構51は、マイクロコントローラ4より受信した合焦制御信号に基づいて、レンズ5に含まれる合焦レンズの位置を変更する（ステップS104）。画像の特定された領域に焦点が合っているか否かは、コントラストAF方式、位相差AF方式などの方式により判定される。

[0088] 当業者は、本開示の精神および範囲から外れることなく、種々の変更、置換および修正をこれに加えることが可能であることを理解されたい。

符号の説明

[0089] 1 コンピュータ
 131 生成部
 132 演算処理部
 NN1 ニューラルネットワーク

請求の範囲

- [請求項1] 入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する生成部と、
- 前記第1特徴マップが入力される第1領域および前記第2特徴マップが入力される第2領域を前記チャンネル方向に有する入力層と、前記第1領域に対応する演算結果が出力される第3領域および前記第2領域に対応する演算結果が出力される第4領域を前記チャンネル方向に有する出力層とを備えるニューラルネットワークに演算を実行させる演算部と、を備え、
- 前記第3領域は、前記第1領域および少なくとも前記第2領域の一部に結合されるとともに、前記第2領域から前記第3領域にいたる結合の演算係数は実質的にゼロに設定され、
- 前記第4領域は、前記第2領域および少なくとも前記第1領域の一部に結合されるとともに、前記第1領域から前記第4領域にいたる結合の演算係数は実質的にゼロに設定される、
- 演算装置。
- [請求項2] 前記ニューラルネットワークは、前記入力層と前記出力層との間に少なくとも1つの中間層を備える、請求項1に記載の演算装置。
- [請求項3] 前記生成部は、前記入力データと同一のデータを前記第1特徴マップおよび前記第2特徴マップとして生成する、請求項1または2に記載の演算装置。
- [請求項4] 前記ニューラルネットワークは、畳み込み演算、転置畳み込み演算、プーリング演算、活性化演算のうち少なくとも1つを行う、請求項1－3のいずれか一項に記載の演算装置。
- [請求項5] 入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する第1ステップと、
- 前記第1特徴マップが入力される第1領域および前記第2特徴マップが入力される第2領域を前記チャンネル方向に有する入力層と、前記

第1領域に対応する演算結果が出力される第3領域および前記第2領域に対応する演算結果が出力される第4領域を前記チャンネル方向に有する出力層とを備えるようニューラルネットワークを作成する第2ステップと、

コンピュータが、前記入力データとして教師データを前記ニューラルネットワークに入力し、前記入力層と前記出力層との結合の演算係数を更新する第3ステップと、を含み、

前記第2ステップは、

前記第3領域を前記第1領域および少なくとも前記第2領域の一部に結合するステップと、

前記第4領域を前記第2領域および少なくとも前記第1領域の一部に結合するステップと、を含み、

前記第3ステップは、前記第2領域から前記第3領域にいたる結合の演算係数および前記第1領域から前記第4領域にいたる結合の演算係数が実質的にゼロになるよう、演算係数を更新するステップを含む、

学習方法。

[請求項6]

入力データに基づいて、チャンネル方向の特徴量が共通する第1特徴マップおよび第2特徴マップを生成する第1ステップと、

前記第1特徴マップが入力される第1領域および前記第2特徴マップが入力される第2領域を前記チャンネル方向に有する入力層と、前記第1領域に対応する演算結果が出力される第3領域および前記第2領域に対応する演算結果が出力される第4領域を前記チャンネル方向に有する出力層とを備えるようニューラルネットワークを作成する第2ステップと、

コンピュータが、前記入力データとして教師データを前記ニューラルネットワークに入力し、前記入力層と出力層との結合の演算係数を更新する第3ステップと、を含み、

前記第2ステップは、

前記第3領域を前記第1領域および少なくとも前記第2領域の一部に結合するステップと、

前記第4領域を前記第2領域および少なくとも前記第1領域の一部に結合するステップと、を含み、

前記第3ステップは、前記第2領域から前記第3領域にいたる結合の演算係数および前記第1領域から前記第4領域にいたる結合の演算係数を実質的にゼロに維持しながら、その他の結合の演算係数を更新するステップを含む、

学習方法。

[請求項7]

入力された特徴マップの特徴量に対し演算係数を用いて所定の演算を実行した結果を出力する演算層をそれぞれ1以上備えるように第1演算経路および第2演算経路を作成するとともに、同一の演算を実行する前記演算層の位置が前記第1演算経路および前記第2演算経路において対応する位置となるように各演算層を配置するステップと、

前記第1演算経路において前記第2演算経路に配置された演算層に対応する演算層が配置されていない位置、および、前記第2演算経路において前記第1演算経路に配置された演算層に対応する演算層が配置されていない位置に、対応する位置に配置された演算層と同一の演算を実行する補助演算層を追加するステップと、

少なくともいずれか一方に前記補助演算層を追加された前記第1演算経路および前記第2演算経路に、それぞれ教師データを入力して前記演算層の前記演算係数を更新するステップと、

前記演算係数が更新された前記第1演算経路および前記第2演算経路において対応するそれぞれの層をチャネル方向に連結するステップと、

を含む、ニューラルネットワークの作成方法。

[請求項8]

入力された特徴マップの特徴量に対し演算係数を用いて所定の演算

を実行した結果を出力する演算層をそれぞれ1以上備えるように第1演算経路および第2演算経路を作成するとともに、同一の演算を実行する前記演算層の位置が前記第1演算経路および前記第2演算経路において対応する位置となるように各演算層を配置するステップと、

前記第1演算経路および前記第2演算経路にそれぞれ教師データを入力して前記演算層の前記演算係数を更新するステップと、

前記第1演算経路において前記第2演算経路に配置された演算層に対応する演算層が配置されていない位置、および、前記第2演算経路において前記第1演算経路に配置された演算層に対応する演算層が配置されていない位置に、対応する位置に配置された演算層と同一の演算を実行する補助演算層を追加するステップと、

少なくともいずれか一方に前記補助演算層が追加された前記第1演算経路および前記第2演算経路において対応するそれぞれの層をチャンネル方向に連結するステップと、

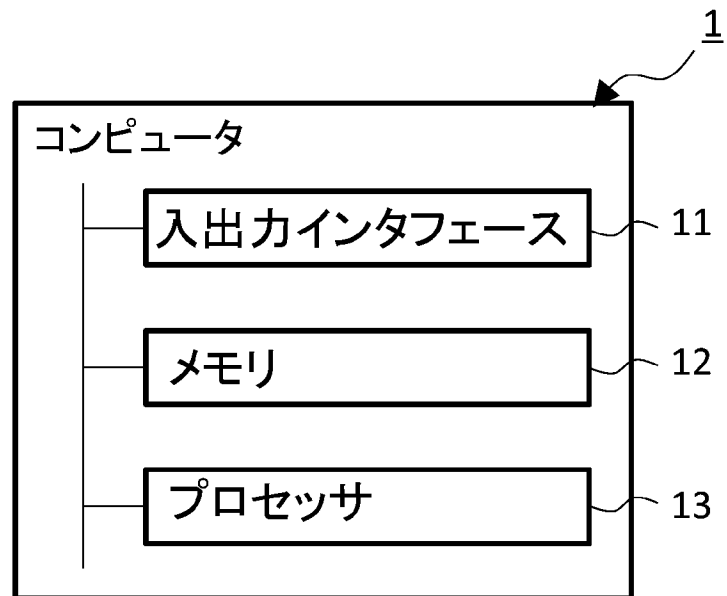
を含む、ニューラルネットワークの作成方法。

[請求項9]

前記補助演算層は、入力される特徴量と実質的に等しい値を出力するよう演算係数が設定された層である、請求項7または8に記載のニューラルネットワークの作成方法。

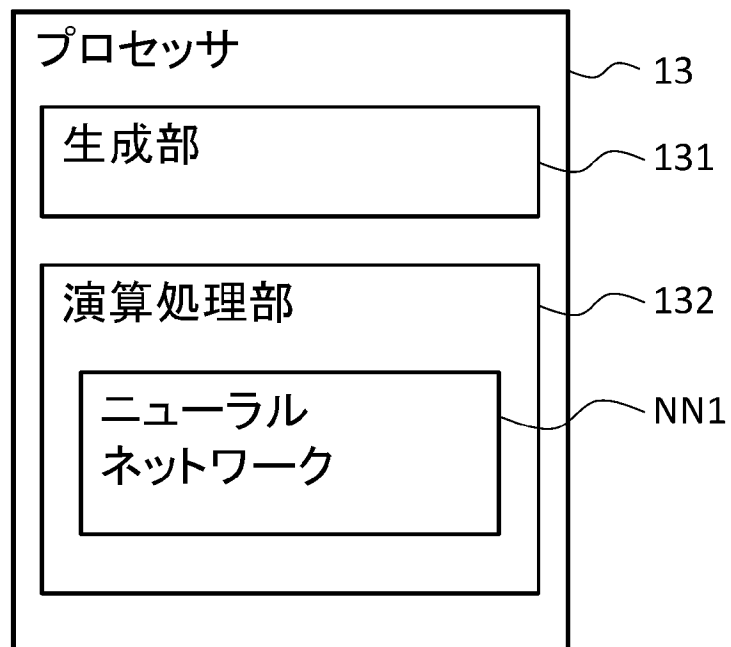
[図1]

図1



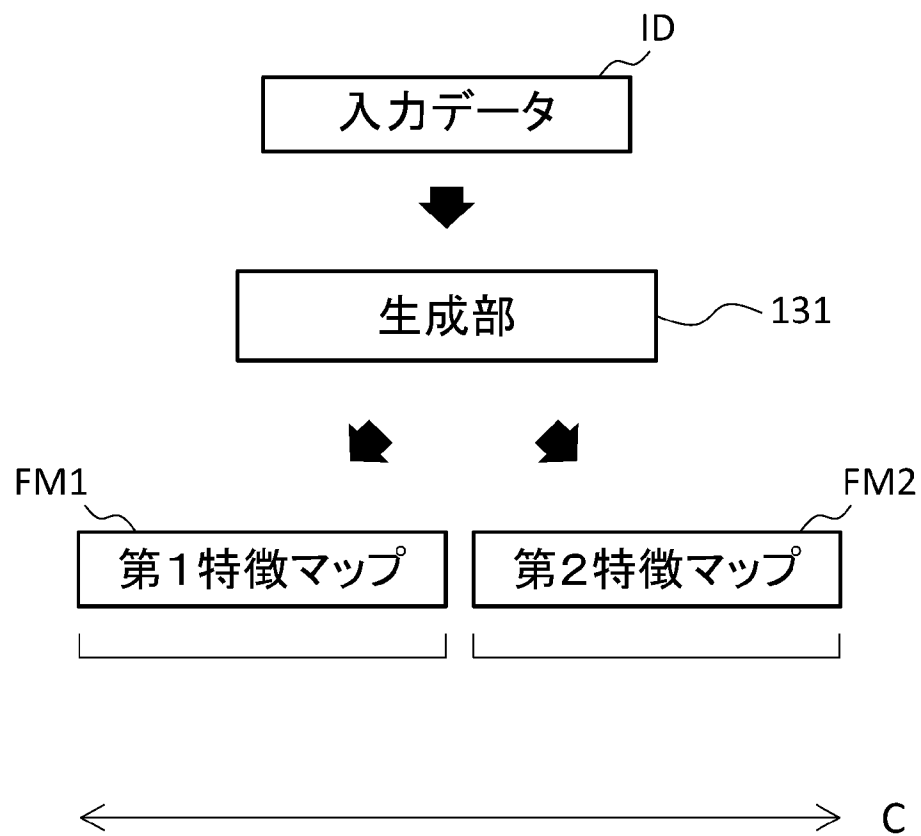
[図2]

図2



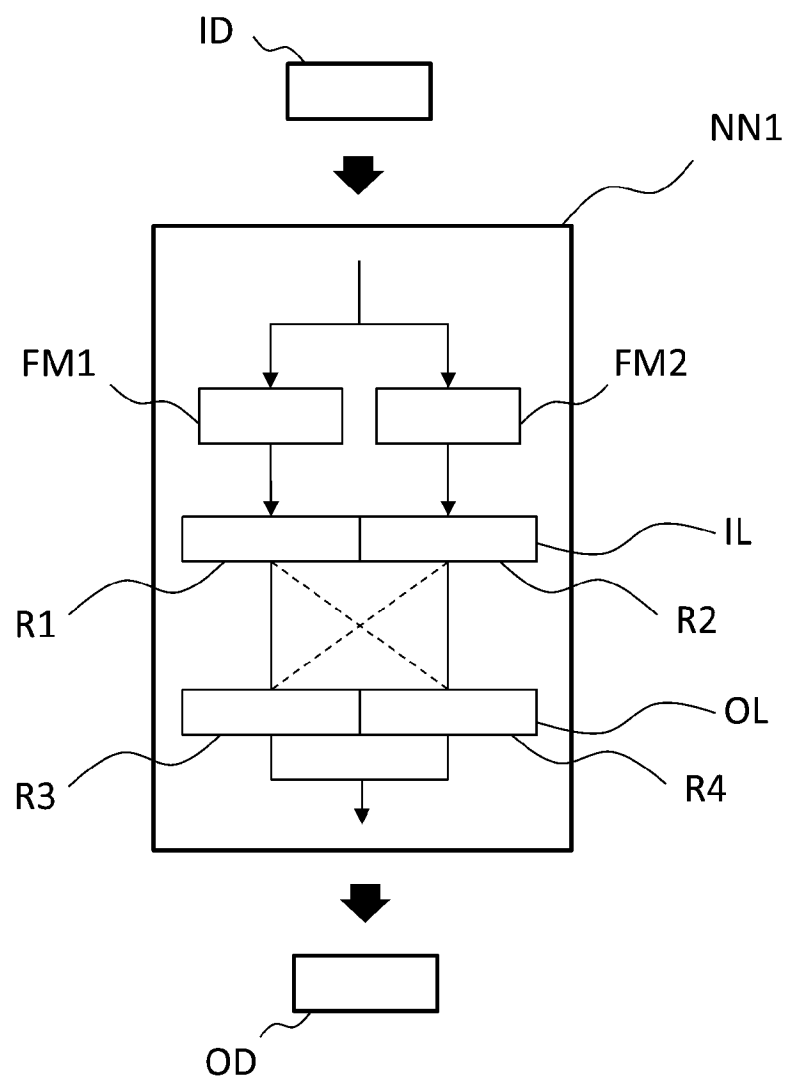
[図3]

図3



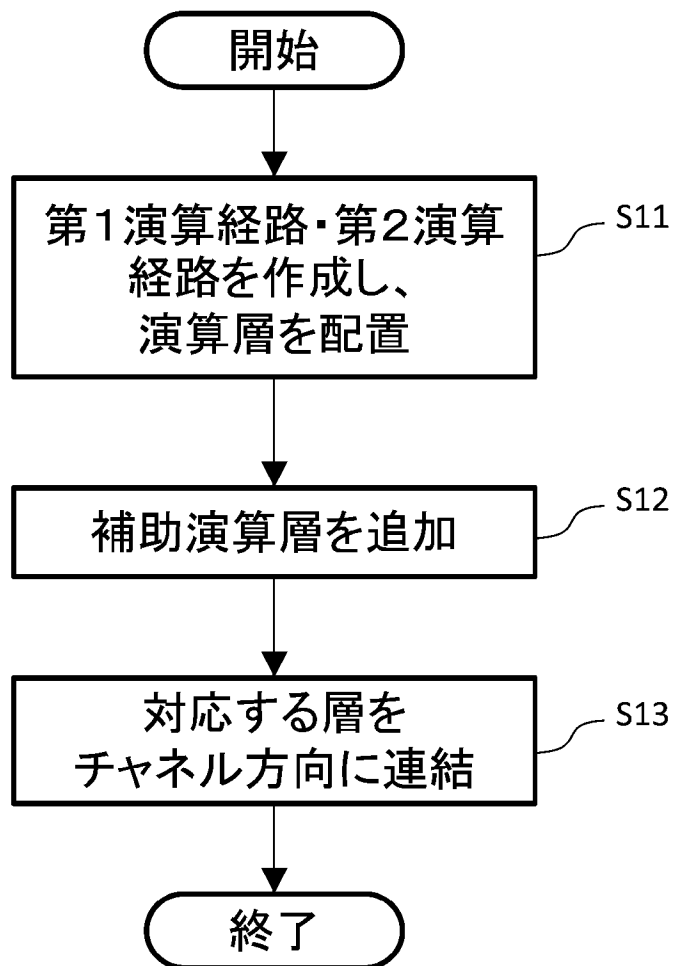
[図4]

図4



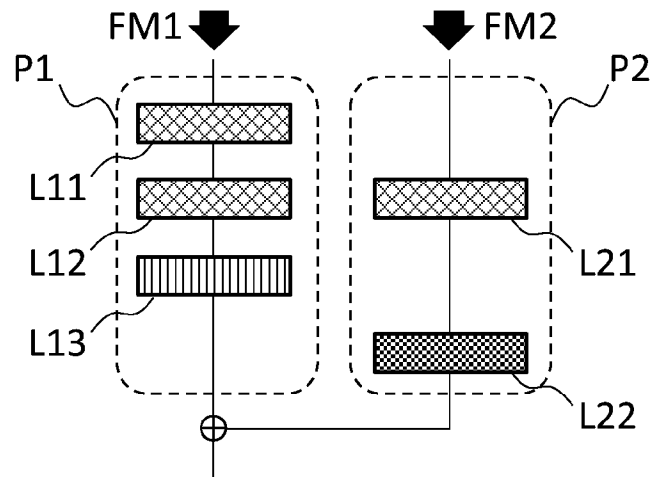
[図5]

図5



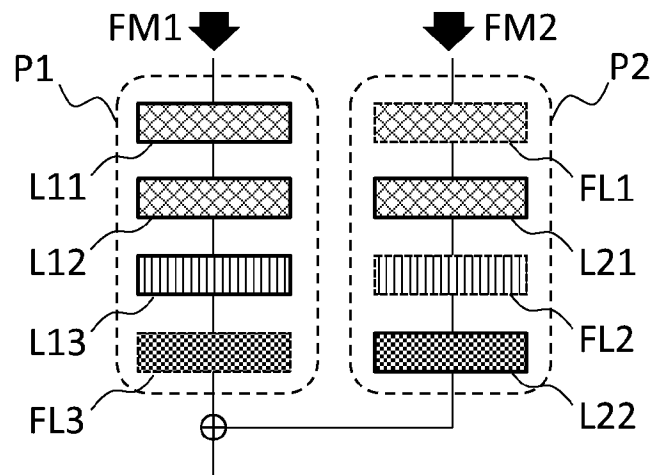
[図6A]

図6A



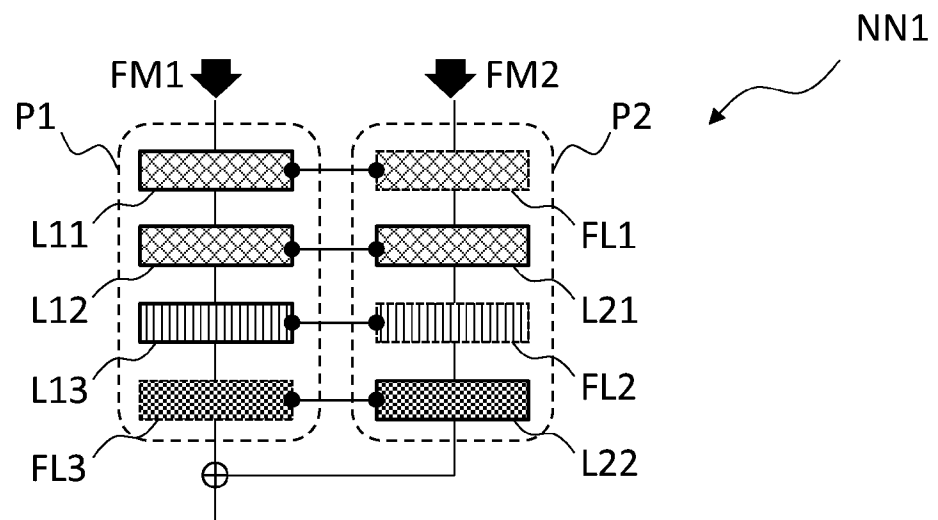
[図6B]

図6B



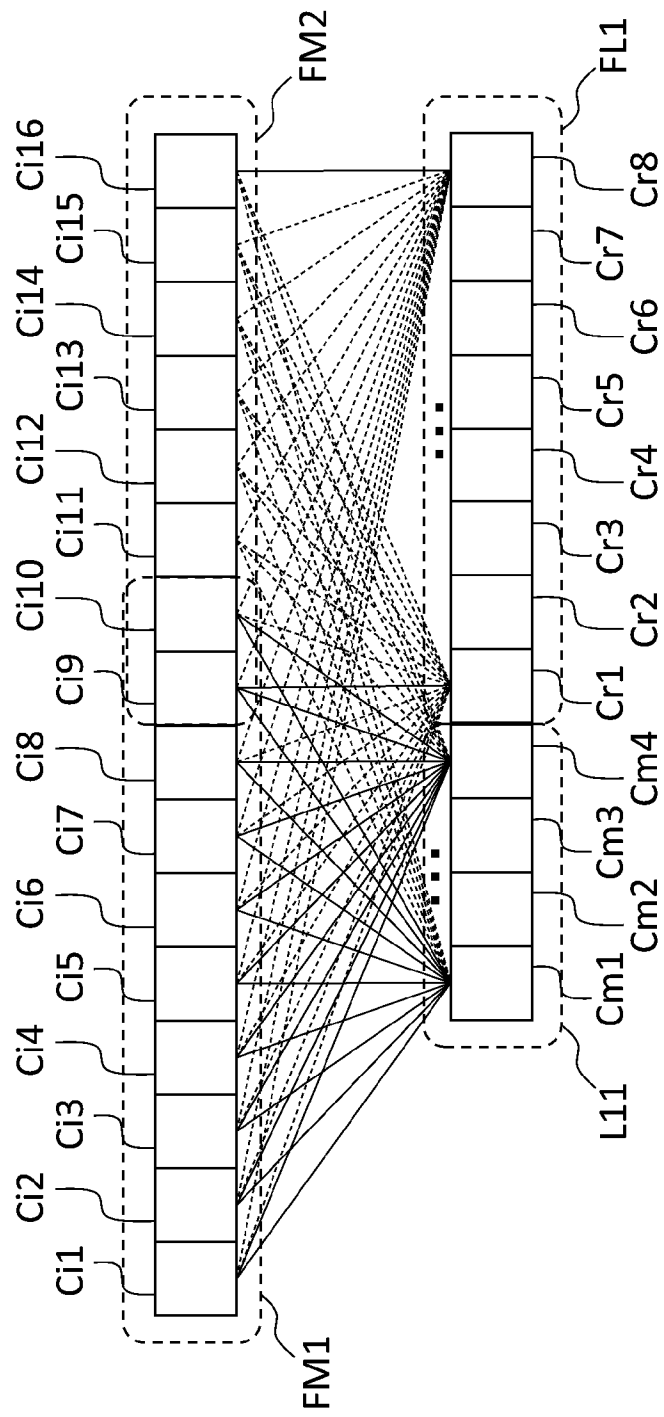
[図6C]

図6C



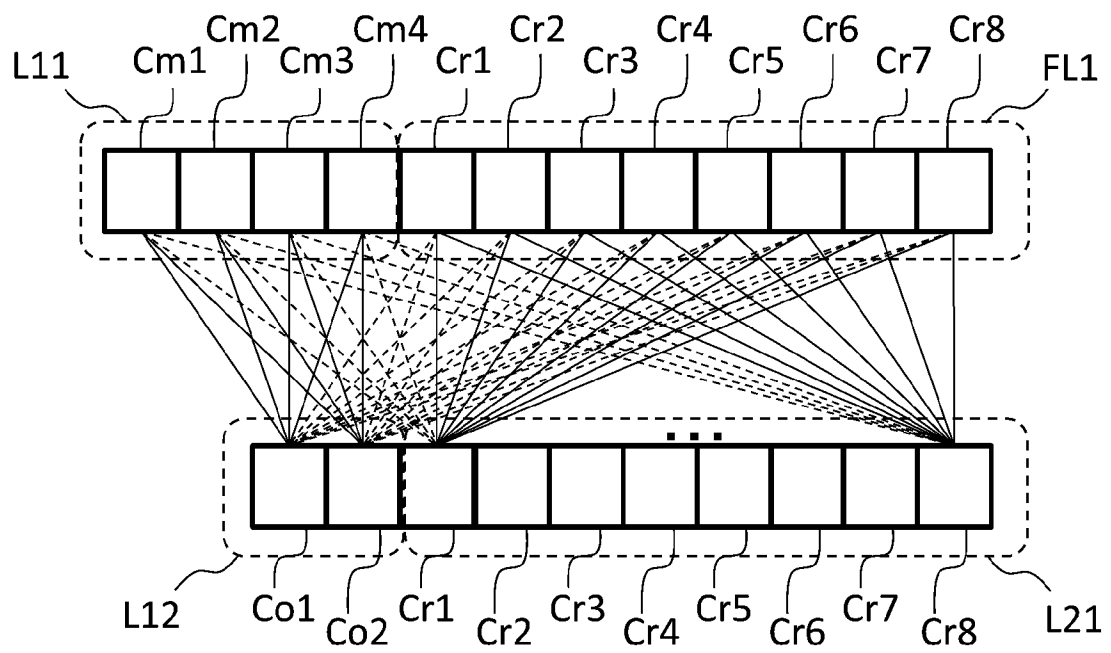
[図7]

図7



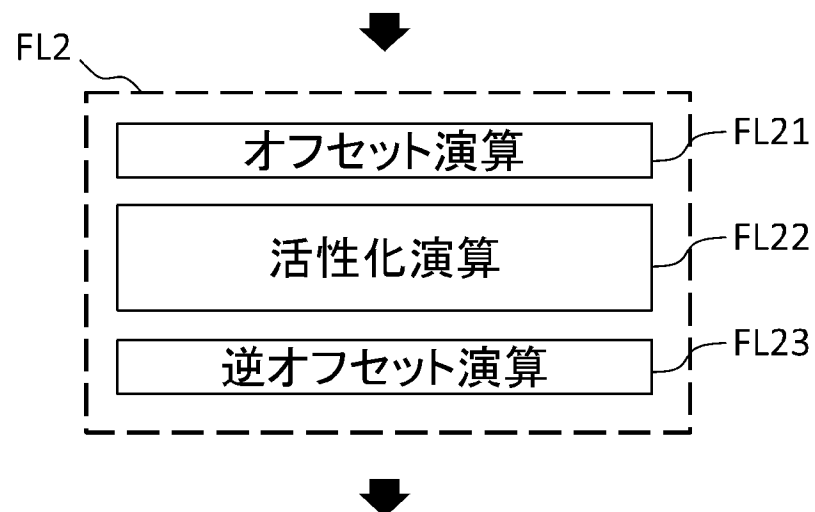
[図8]

図8



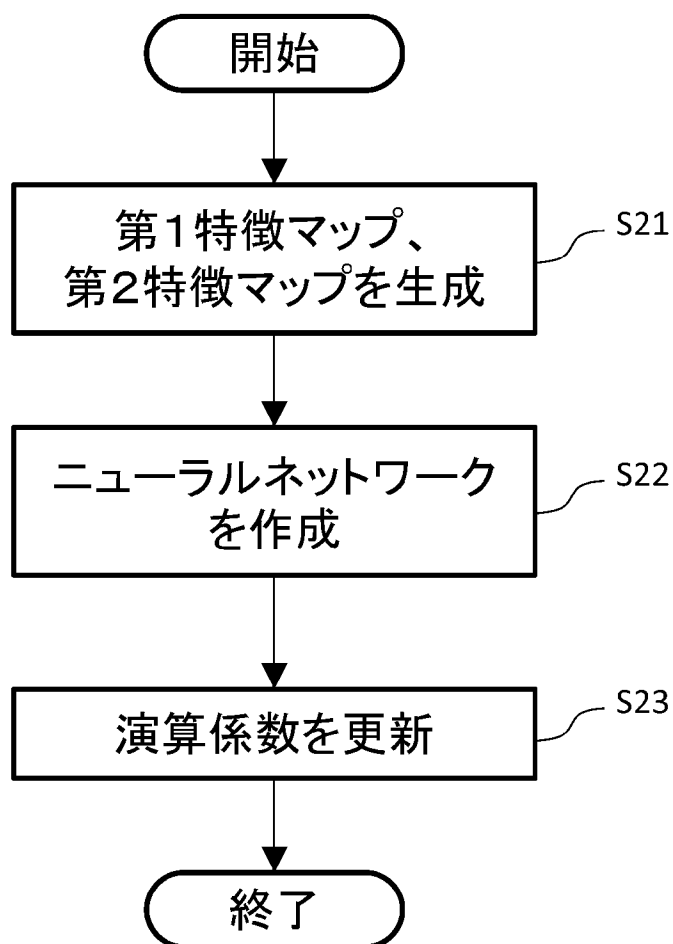
[図9]

図9



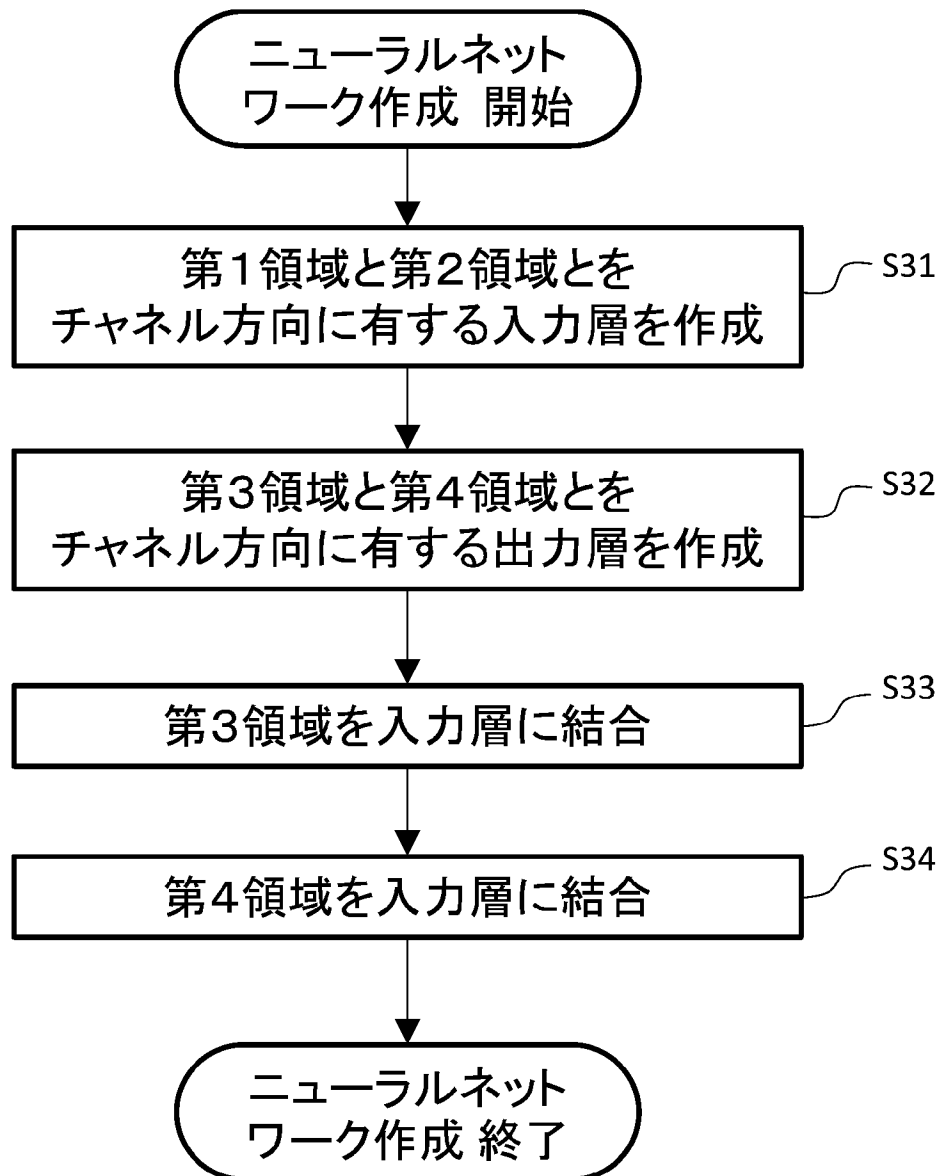
[図10]

図10



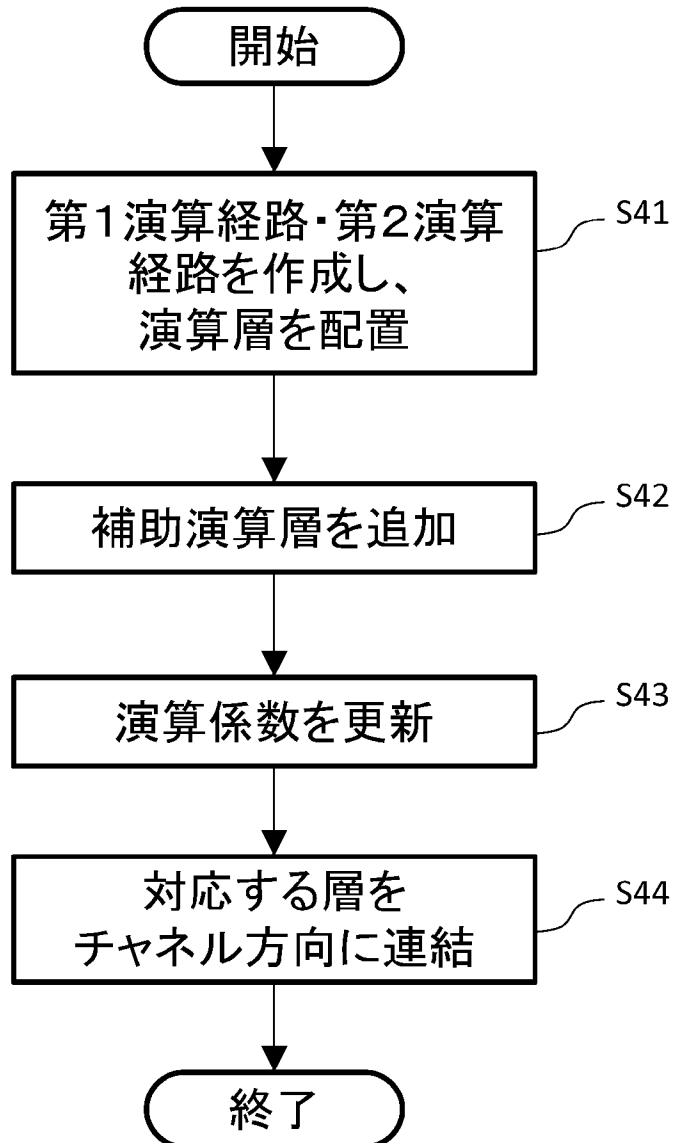
[図11]

図11



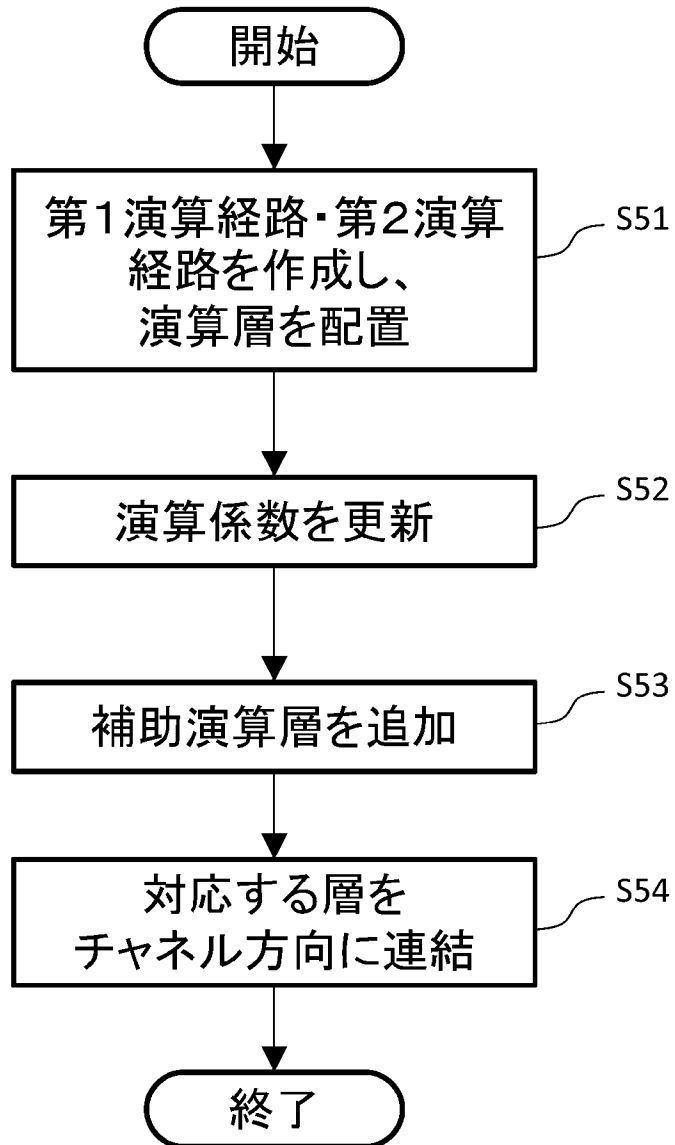
[図12]

図12



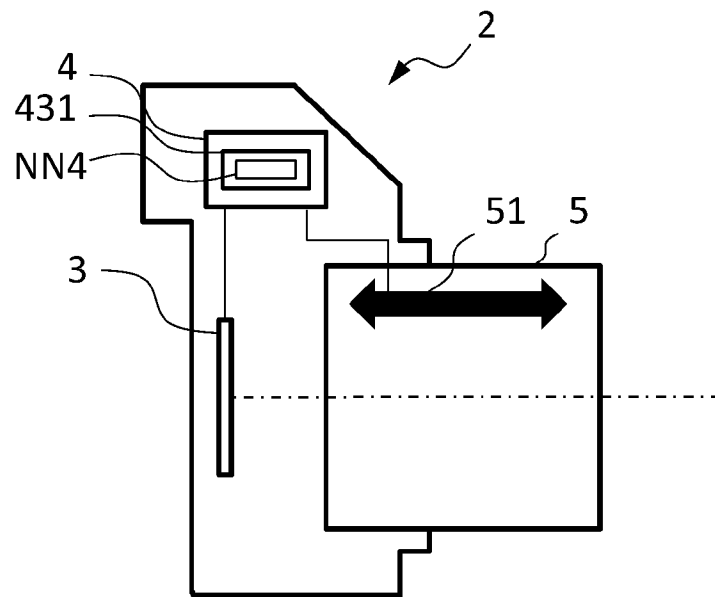
[図13]

図13



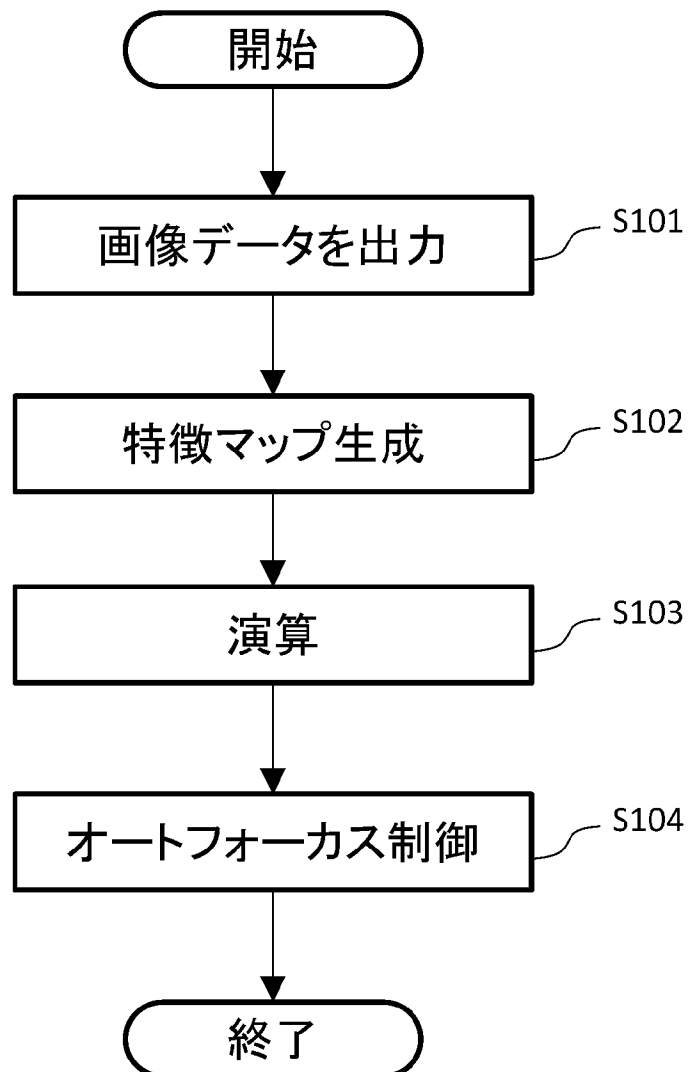
[図14]

図14



[図15]

図15



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/022886

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/04 (2006.01) i; G06N 3/08 (2006.01) i
FI: G06N3/04; 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)

C06N3/04; C06N3/08

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	KRIZHEVSKY, Alex et al., "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems 25 (NIPS 2012) [online], 2012 [retrieved on 30 August 2021], Retrieved from the Internet: <URL: https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf >, in particular, "3.5 Over all Architecture"	1-9
A	CHEN, Yunpeng et al., "Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution", arXiv[online], 18 August 2019 [retrieved on 30 August 2021], Retrieved from the Internet: <URL: https://export.arxiv.org/pdf/1904.05049v3.pdf >, in particular, "3. Method"	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

"I" 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
30 August 2021 (30.08.2021)

Date of mailing of the international search report
07 September 2021 (07.09.2021)

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

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/022886

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2019-159940 A (PREFERRED NETWORKS, INC.) 19 September 2019 (2019-09-19)	1-9

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2021/022886

Patent Documents
referred in the
Report

Publication
Date

Patent Family

Publication
Date

JP 2019-159940 A

19 Sep. 2019

(Family: none)

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 3/04(2006.01)i; G06N 3/08(2006.01)i FI: G06N3/04; 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）） G06N3/04; 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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2021年</td> </tr> </table> 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）			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				
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													
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	KRIZHEVSKY, Alex ほか, ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25(NIPS 2012)[online], 2012 [retrieved on 2021.08.30], Retrieved from the Internet: <URL: https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf> 特に”3.5 Overall Architecture”	1-9												
A	CHEN, Yunpeng ほか, Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution, arXiv[online], 2019.08.18 [retrieved on 2021.08.30], Retrieved from the Internet: <URL: https://export.arxiv.org/pdf/1904.05049v3.pdf> 特に”3. Method”	1-9												
A	JP 2019-159940 A (株式会社 Preferred Networks) 19.09.2019 (2019-09-19)	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。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“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</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの</td> <td>“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</td> </tr> <tr> <td>“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</td> <td>“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</td> </tr> <tr> <td>“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）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“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</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの	“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	“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	“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	“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）	“&” 同一パテントファミリー文献	“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	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの	“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													
“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	“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													
“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）	“&” 同一パテントファミリー文献													
“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														
国際調査を完了した日 30.08.2021	国際調査報告の発送日 07.09.2021													
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） 渡辺 順哉 5B 1157 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/JP2021/022886

引用文献	公表日	パテントファミリー文献	公表日
JP 2019-159940 A	19.09.2019	(ファミリーなし)	