

(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1月2日(02.01.2020)



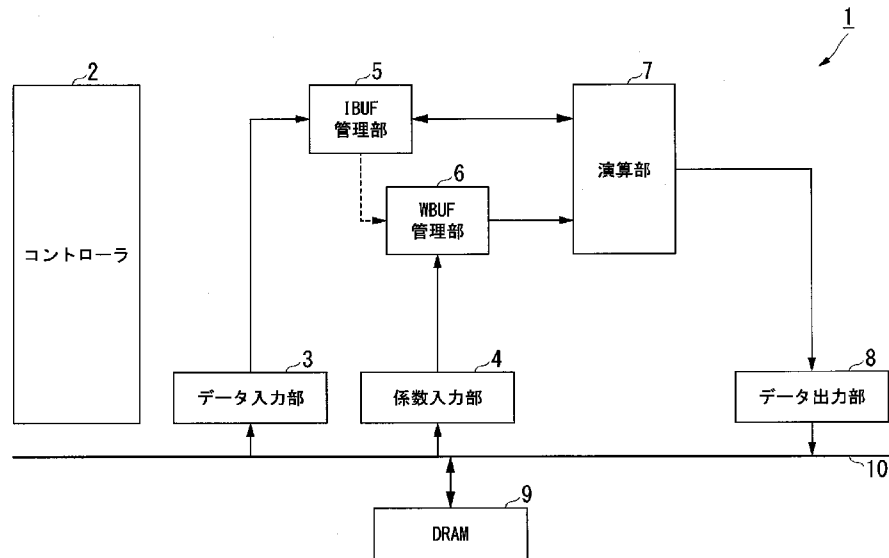
(10) 国際公開番号

WO 2020/003345 A1

- (51) 国際特許分類:
G06N 3/063 (2006.01) *G06F 17/16* (2006.01)
- (21) 国際出願番号: PCT/JP2018/023974
- (22) 国際出願日: 2018年6月25日(25.06.2018)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: オリンパス株式会社 (OLYMPUS CORPORATION) [JP/JP]; 〒1928507 東京都八王子市石川町2951番地 Tokyo (JP).
- (72) 発明者: 古川 英明 (FURUKAWA Hideaki); 〒1928507 東京都八王子市石川町2951番地 オリンパス株式会社内 Tokyo (JP).
- (74) 代理人: 棚井 澄雄, 外 (TANAI Sumio et al.); 〒1006620 東京都千代田区丸の内一丁目9番2号 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: ARITHMETIC PROCESSING DEVICE

(54) 発明の名称: 演算処理装置



- 2 Controller
- 3 Data input unit
- 4 Coefficient input unit
- 5 IBUF management unit
- 6 WBUF management unit
- 7 Computation unit
- 8 Data output unit

(57) Abstract: A SRAM write control unit of an arithmetic processing device for deep learning, which performs a Convolution process and a FullConnect process, virtually divides each SRAM constituting a data storage memory into a plurality of regions, switches write regions according to ID, and performs control so that different input feature value map data of the same coordinates is stored in the same SRAM.

[続葉有]

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

添付公開書類:

一 国際調査報告 (条約第21条(3))

(57) 要約: Convolution処理とFullConnect処理を行うディープラーニング用の演算処理装置のSRAM書き込み制御部は、データ格納メモリを構成するSRAMの各々を仮想的に複数の領域に分割し、IDによって書き込む領域を切り替えるとともに、同じSRAMには同一座標の異なる入力特徴量マップデータが格納されるように制御する。

明 細 書

発明の名称： 演算処理装置

技術分野

[0001] 本発明は、演算処理装置、より詳しくは、畳み込みニューラルネットワークを用いたディープラーニングを行う演算処理装置の回路構成に関する。

背景技術

[0002] 従来、複数の処理層が階層的に接続されたニューラルネットワークを用いて演算を実行する演算処理装置がある。特に画像認識を行う演算処理装置では、畳み込みニューラルネットワーク (Convolutional Neural Network、以下CNNという) を用いたディープラーニングが広く行われている。

[0003] CNNを用いたディープラーニングによる画像認識では、入力される画像データ (ピクセルデータ) に対して、CNNの複数の処理層における処理が順次施されることにより、画像に含まれる対象物が認識された最終的な演算結果データが得られる。CNNの処理層は、畳み込み演算処理、非線形処理、縮小処理 (プーリング処理) 等を含むConvolution処理を行うConvolution層 (畳み込み層) と、全ての入力 (ピクセルデータ) に係数を乗じて累積加算するFullConnect処理を行うFullConnect層 (全結合層) とに大きく分類される。

[0004] CNNを用いたディープラーニングによる画像認識は、高い認識率を実現できる。しかし、検出する被写体の種類を増やしたり、被写体検出精度を上げるためには、ネットワークを大きくする必要がある。そうするとデータ格納バッファや係数格納バッファが必然的に大容量になるが、LSI (Large-Scale Integrated circuit) にはあまり大容量のメモリを搭載できない。

[0005] また、画像認識処理におけるディープラーニングでは、(K-1) 層目とK層目におけるFM (Feature Map) サイズとFM数 (FMの面

数)の関係は次式のような関係になる場合が多く、回路としてメモリサイズを決定する際には最適化が困難である。

[0006] FMサイズ $[K] = 1 / 4 \times \text{FMサイズ} [K - 1]$

FM数 $[K] = 2 \times \text{FM数} [K - 1]$

[0007] 例えば、CNNのバリエーションの1つであるYolo_v2に対応可能な回路のメモリのサイズを考える場合、FMサイズとFM数の最大値だけで決定しようとする1GB程度必要となる。実際には、FM数とFMサイズは反比例的関係があるため、計算上は3MB程度あれば十分ではあるが、層別にメモリ管理を緻密に計算して設計しなければならない。

[0008] このような問題があることから、CNNは一般的には高性能PCやGPU (Graphics Processing Unit) を用いたソフトウェア処理で実装される。しかし、高速処理を実現するためには処理の重い部分をハードウェアで構成する必要がある。このようなハードウェア実装の例としては、特許文献1に記載がある。

[0009] 特許文献1は、複数の演算ブロックと複数のメモリを搭載することにより演算処理の効率化を図るようにした演算処理装置を開示している。複数の演算処理ユニットを持ち、それぞれに演算ブロックと、それと対になるバッファがあり、中継部を介して畳み込み演算処理を並列に実行すると共に、演算部間で累積加算データを授受することで、入力ネットワークが大きくても一度に活性化処理への入力を作成できるようになっている。

先行技術文献

特許文献

[0010] 特許文献1：特開2017-151604号公報

発明の概要

発明が解決しようとする課題

[0011] 特許文献1は、演算ブロックとデータ保持部が対をなす構成で、演算ブロック数を超えるネットワークに対しては冗長データ保持部を介して繰り返し累積

加算実行する事で対応できるようになっている。しかし、具体的な制御、特にデータ保持部では上述した様なFM数とFMサイズの関係に対応するためのバッファと制御の最適化が必要となるが、それについては考慮されていない。

[0012] 上述の事情を踏まえ、本発明は、ディープラーニングに適用でき、かつLSIに無理なく搭載できるメモリサイズの演算処理装置を提供することを目的とする。

課題を解決するための手段

[0013] 本発明の第一の態様は、Convolution処理とFullConnect処理を行うディープラーニング用の演算処理装置であって、3個以上のSRAMで構成されたデータ格納メモリと、入力特徴量マップデータから入力有効データ数をカウントして座標に変換するカウンタと、前記カウンタが変換した座標を前記データ格納メモリにおけるアドレスに変換して、前記入力特徴量マップデータを前記データ格納メモリに格納するSRAM書き込み制御部と、外部からのリクエストを受けて、前記データ格納メモリから所定の方法で格納されている前記入力特徴量マップデータを取り出すSRAM読み出し制御部と、を有するデータ格納メモリ管理部と；係数を格納する係数格納メモリを有し、前記データ格納メモリ管理部のステータスを参照して、前記データ格納メモリ管理部から取り出すデータに対応する係数を前記係数格納メモリから取り出す係数格納メモリ管理部と；格納用外部メモリから、前記入力特徴量マップデータを取得するデータ入力部と；前記格納用外部メモリから、前記係数を取得する係数入力部と；入力N並列、出力M並列の構成（N、 $M \geq 1$ の正数）で、前記データ格納メモリから前記入力特徴量マップデータを取得し、前記係数格納メモリから前記係数を取得して、フィルタ処理、累積加算処理、非線形演算処理およびプーリング処理を行う演算部と；前記演算部から出力されるM並列のデータを連結して、出力特徴量マップデータとして前記格納用外部メモリに出力するデータ出力部と；前記データ格納メモリ管理部、前記係数格納メモリ管理部、前記データ入力部、前記

データ出力部、および前記演算部を制御するコントローラと；を有し、前記カウンタは、入力された前記入力特徴量マップデータのIDを生成し、前記SRAM書き込み制御部は、前記データ格納メモリを構成する前記SRAMの各々を仮想的に複数の領域に分割し、前記IDによって書き込む領域を切り替えるとともに、同じSRAMには同一座標の異なる入力特徴量マップデータが格納されるように制御し、前記SRAM読み出し制御部は、取り出したい前記入力特徴量マップデータの前記IDに対応する前記領域にアクセスするように前記アドレスを切り替え、前記演算部は、N並列でフィルタ演算を実行するフィルタ演算部と、前記フィルタ演算部の演算結果を全て累積加算する第1加算器と、前記第1加算器の累積加算結果を後段で累積加算する第2加算器と、前記第2加算器の累積加算結果を保持するフリップフロップと、1つのデータ格納メモリに格納されている複数の入力特徴量マップデータを取り出すために、入力特徴量マップデータのIDを生成して、前記SRAM読み出し制御部から入力される入力特徴量マップデータと同じIDの前記係数を前記係数格納メモリ管理部から取得するよう制御し、全ての入力特徴量マップデータを取り終わるまでは、前記第2加算器が前記第1加算器の累積加算結果を累積加算するよう制御する演算制御部と、を有する。

[0014] 前記演算部は、同一座標の入力特徴量マップデータごとに演算処理を実行し、前記SRAM書き込み制御部は、前記同一座標の入力特徴量マップデータの演算処理が終わるごとに、前記同一座標の入力特徴量マップデータが格納されていた前記領域を開放して、前記入力特徴量マップデータの続きのデータを格納してもよい。

[0015] 前記SRAM読み出し制御部は、1つの入力特徴量マップデータから、出力するm枚分の出力特徴量マップデータの演算処理を実行するように前記データ格納メモリから前記入力特徴量マップデータを読み出し、前記データ出力部は、有効データをカウントしてどの出力特徴量マップデータであるかを識別し、適切に出力特徴量マップデータにおけるアドレスを生成して、前記格納用外部メモリに出力してもよい。

[0016] 前記S R A M読み出し制御部は、前記データ格納メモリから前記入力特徴量マップデータの一部データを読み出し、前記演算部は、前記S R A M読み出し制御部から入力された前記入力特徴量マップデータの一部データの演算処理を行い、前記出力特徴量マップデータの一部データを作成し、前記入力特徴量マップデータの前の一部データの処理が終わったら、前記入力特徴量マップデータの次の一部データの処理へと進んでもよい。

発明の効果

[0017] 本発明の各態様に係る演算処理装置によれば、i F M（入力特徴量マップ）サイズやi F M数（i F Mの面数）が大きくてデータ格納メモリ（I B U F）に入りきらないサイズでも、フィルタサイズ分のオーバーラップをせずに演算処理とデータ更新をシームレスに実施することができる。

図面の簡単な説明

- [0018] [図1]本発明の実施形態に係る演算処理装置の構成を示すブロック図である。
- [図2]本発明の実施形態に係る演算処理装置の演算部の構成を示す図である。
- [図3]本発明の実施形態に係る演算処理装置のI B U F（データ格納メモリ）管理部の構成を示す図である。
- [図4]本発明の実施形態に係る演算処理装置のI B U F（データ格納メモリ）の構成を示す図である。
- [図5]本発明の実施形態に係る演算処理装置において、1個の入力c hに入力されたi F M（入力特徴量マップ）が分割されてI B U Fに格納される様子を示す図である。
- [図6]本発明の実施形態に係る演算処理装置において、分割されてI B U Fに格納されたi F Mが処理される様子を示す図である。
- [図7]本発明の実施形態に係る演算処理装置において、 $n = 2$ の時の、i F M I Dによってアドレスオフセットを変化させるイメージを示す図である。
- [図8]本発明の実施形態に係る演算処理装置における、I B U Fへのデータ書き込み制御フローを示す図である。
- [図9]本発明の実施形態に係る演算処理装置における、i F Mのデータ配置と

係数の一例を示す図である。

[図10]本発明の実施形態に係る演算処理装置における、Convolution処理時のタイミングチャートを示す図である。

[図11]本発明の実施形態に係る演算処理装置における、SRAM読み出し制御部の制御フローの一例を示す図である。

[図12]本発明の実施形態に係る演算処理装置における、CNNの処理イメージを示す図である。

[図13]本発明の実施形態に係る演算処理装置における、SRAM読み出し制御部の制御フローの一例を示す図である。

[図14]本発明の実施形態に係る演算処理装置における、演算時のタイミングチャートを示す図である。

[図15]本発明の実施形態に係る演算処理装置において、データが書き出されるDRAMの出力イメージを示す図である。

発明を実施するための形態

[0019] (実施形態1)

本発明の実施形態について、図面を用いて説明する。図1は、本発明の実施形態に係る演算処理装置の構成を示すブロック図である。

[0020] 演算処理装置1は、コントローラ2と、データ入力部3と、係数入力部4と、IBUF（データ格納メモリ）管理部5と、WBUF（係数格納メモリ）管理部6と、演算部7と、データ出力部8を備える。データ入力部3と、係数入力部4と、データ出力部8は、バス10を介して、DRAM（格納用外部メモリ）9と接続されている。演算処理装置1は、入力特徴量マップ（iFM）から出力特徴量マップ（oFM）を生成する。

[0021] IBUF管理部5は、入力特徴量マップ（iFM）データ格納用のメモリ（データ格納メモリ、IBUF）と、データ格納メモリの管理・制御回路（データ格納メモリ制御回路）と、後述するカウンタ51と、SRAM書き込み制御部52と、N個のIBUF（データ格納メモリ）（0～N-1）と、SRAM読み出し制御部53とを備える。IBUFは、それぞれが3個以上

のSRAMから構成される。

[0022] IBUF管理部5は、入力データ（iFMデータ）中の有効データ数をカウントして座標に変換し、さらにそれをIBUFアドレス（IBUFにおけるアドレス）に変換し、データをデータ格納メモリに格納するとともに、所定の方法でiFMデータをデータ格納メモリから取り出す。IBUF管理部5は、複数のFMが入力されるときは、同一座標のデータが同じIBUFに入るように制御する。

[0023] WBUF管理部6は、係数格納用のメモリ（係数格納メモリ、WBUF）と、係数格納メモリの管理・制御回路（係数格納メモリ制御回路）を有する。WBUF管理部6は、IBUF管理部5のステータスを参照して、IBUF管理部5から取り出すデータに対応する係数を係数格納メモリから取り出す。

[0024] DRAM9は、iFMデータ、oFMデータおよび係数を格納する。データ入力部3は、DRAM9から所定の方法で、入力特徴量マップ（iFM）を取得し、IBUF（データ格納メモリ）管理部5に渡す。データ出力部8は、DRAM9に所定の方法で、出力特徴量マップ（oFM）データを書き出す。具体的には、データ出力部8は、演算部7から出力されるM並列のデータを連結してDRAM9に出力する。係数入力部4は、DRAM9から所定の方法で、係数を取得し、WBUF（係数格納メモリ）管理部6に渡す。

[0025] 演算部7は、IBUF（データ格納メモリ）管理部5からデータ、WBUF（係数格納メモリ）管理部6から係数を取得して、フィルタ処理・累積加算・非線形演算・プーリング処理等のデータ処理を行う。コントローラ2は、回路全体の制御を行う。

[0026] CNNでは、複数の処理層において、何層も類似の処理を実施する。本発明の実施形態に係る演算処理装置1を用いる場合、演算部7は、1層分の処理に必要なデータおよび係数だけを有して1層分の処理を行い、データおよび係数はDRAM9に逐一置かれる。そして、必要な層数分の処理が繰り返して実行される。演算処理装置1は最終出力データとして被写体推定結果を出

かし、この最終出力データを、プロセッサ（回路でもよい）を用いて処理することにより被写体推定結果を得る。

[0027] 図2は、本発明の実施形態に係る演算処理装置の演算部7の構成を示す図である。演算部7の入力チャネル数は N （ N は1以上の正数）、すなわち入力データ（ i FMデータ）は N 次元であり、 N 次元の入力データが並列して処理される（入力 N 並列）。

[0028] 演算部7の出力チャネル数は M （ M は1以上の正数）、すなわち出力データは M 次元であり、 M 次元の入力データが並列して出力される（出力 M 並列）。図2に示すように、1つの層において、各チャネル（ i $c h_0 \sim i$ $c h_N - 1$ ）ごとに、 i FMデータ（ $d_0 \sim d_N - 1$ ）と係数（ $k_0 \sim k_N - 1$ ）が入力され、1個のデータを出力する。この処理が M 層、並行して行われ、 M 個のデータ $o c h_0 \sim o c h_M - 1$ が出力される。

[0029] このように、演算部7は、入力チャネル数を N 、出力チャネル数を M として、並列度が $N \times M$ となる構成を取る。入力チャネル数 N および出力チャネル数 M の大きさは、 CNN の大きさに応じて設定（変更）することができるので、処理性能や回路規模を勘案して適切に設定する。

[0030] 演算部7は、演算部内各部の制御を行う演算制御部71を備える。また、演算部7は、各層ごとに、フィルタ演算部72と、第1加算器73と、第2加算器74と、FF（フリップフロップ）75と、非線形変換部76と、プーリング処理部77を備える。そして、このような層が M 個ある。

[0031] 演算制御部71が、演算部7の前段に対してリクエストを発行することにより、所定のデータがフィルタ演算部72に入力される。フィルタ演算部72は、内部で乗算器と加算器が N 並列で同時に実行できるように構成されており、入力データのフィルタ処理を行い、フィルタ処理の結果を N 並列で出力する。

[0032] 第1加算器73は、 N 並列で実行され出力されたフィルタ演算部72におけるフィルタ処理の結果を全て加算する。第2加算器74は、時分割で入力される第1加算器73の演算結果を累積加算する。第2加算器74は、全て

の i F M データを取り終わるまでは第 1 加算器 7 3 の演算結果を累積加算する。F F 7 5 は、累積加算の結果を保持するために設けられている。非線形変換部 7 6 は、第 2 加算器 7 4 および F F 7 5 での累積加算の結果に対して、A c t i v a t e 関数などによる非線形演算処理を行う。具体的な実装は特に規定しないが、例えば折れ線近似により非線形演算処理を行う。

[0033] プーリング処理部 7 7 は、非線形変換部 7 6 から入力された複数のデータの中から最大値を選択出力 (M a x P o o l i n g) する、平均値を算出 (A v e r a g e P o o l i n g) する、などのプーリング処理を行う。なお、非線形変換部 7 6 とプーリング処理部 7 7 は、演算制御部 7 1 によりスルーさせることができる。

[0034] 演算制御部 7 1 は、演算部 7 内の各部において行われる上述の処理の制御を行う。このような構成により、演算部 7 において、入力チャネル数 N および出力チャネル数 M の大きさは、C N N の大きさに応じて設定 (変更) することができるので、処理性能や回路規模を勘案して適切に設定する。

[0035] 図 3 は、本発明の実施形態に係る演算処理装置の I B U F (データ格納メモリ) 管理部 5 の構成を示す図である。I B U F 管理部 5 は、カウンタ 5 1 と、S R A M 書き込み制御部 5 2 と、N 個の I B U F (データ格納メモリ) (0 ~ N - 1) と、S R A M 読み出し制御部 5 3 とを備える。

[0036] 入力される i F M (入力特徴量マップ) データの数が入力の並列度 N を越える場合、i F M データが時分割入力されることがあるので、どの i F M データを入力しているかを識別するための I D (以後、i F M I D という) が必要となる。i F M I D は、データ入力部 3 が D R A M 9 からバースト転送の単位で取得したデータに対して 1 つ定まっている代表値であり、例えば、最初の 1 ~ N 面の i F M I D を 0、N + 1 ~ 2 N 面の i F M I D を 1、…のように、順次インクリメンタルに変化させる。

[0037] カウンタ 5 1 は、入力される i F M データの I D (i F M I D) を生成するとともに、i F M I D 毎に入力データ中の有効データ数 (入力有効データ数) をカウントして座標 (i F M 座標) に変換して出力する。

[0038] S R A M書き込み制御部52は、カウンタ51から出力されたi F M座標をI B U Fアドレス（I B U Fにおけるアドレス）に変換し、データ（i F Mデータ）をI B U Fに格納する。また、S R A M書き込み制御部52は、I B U Fを構成する各S R A Mを仮想的に複数の領域に分割し、i F M I Dによって書き込む領域を切り変える。また、S R A M書き込み制御部52は、同じS R A Mには同一座標の異なるi F Mデータが格納されるように制御する。

[0039] また、S R A M書き込み制御部52は、同一座標のi F Mデータの演算処理が終わるごとに、この同一座標のi F Mデータが格納されていた領域を開放して、i F Mデータの続きのデータを格納する。

[0040] S R A M読み出し制御部53は、外部から要求（r e q）を受けてI B U Fから所定の方法でデータを取り出す。また、S R A M読み出し制御部53は、取り出したいi F M I Dに対応するI B U F内の領域にアクセスするように、アドレスを切り替える。

[0041] 演算部7の演算制御部71は、1個のI B U Fに格納されている複数のi F Mを取り出すために、i F M I Dを生成して、S R A M読み出し制御部53から入力されるi F Mデータと同じi F M I Dの係数をW B U Fから取得するように制御する。また、演算制御部71は、全てのi F Mデータを取り終わるまでは、第2加算器74を用いて、第1加算器73の演算結果を累積加算する。

[0042] 図4は、本実施形態に係る演算処理装置のI B U F（データ格納メモリ）の構成を示す図である。I B U Fはそれぞれが3個以上のS R A Mから構成される。図4の例では、1個のI B U F（I B U F__i）は4つのS R A Mから構成されており、うち3個のS R A M（S R A M0～S R A M2）を使ってフィルタ演算に必要なデータを取り出す。残りの1個のS R A M（S R A M3）は、D R A Mから取得したデータを書き込むために用いる。

[0043] 演算部7における処理を終えて必要のなくなったデータは、S R A Mの1個ごとの単位で順次開放し、I B U F全体としてはサーキュレートバッファ

として動作する。なお、本例では I B U F は 4 個の S R A M から構成されているが、S R A M の数は 3 個以上であればいくつであっても構わない。S R A M の数を 3 個以上にすることで、例えば読み出し位置が S R A M 1 と S R A M 2 でオーバーラップする（またがる）場合でも、データを順次取り出している最中に S R A M 3 への書き込みを並行して行うことができる。

[0044] S R A M 1 の使用が終わったら w r i t e 用に開放しておく。処理実行中にいずれ S R A M 3 への書き込みが完了するので w r i t e 処理はすでに開放されている S R A M 1 に移る。その間、リード側は S R A M 2 ~ S R A M 3 の領域を使ってフィルタ処理を続ける。このように、使わないエリアを順次開放してデータ更新することでシームレスな処理（スライディングウィンドウ処理）をすることができる。

[0045] 次に、i F M 数（i F M の面数）が回路の入力並列度を超え、1 個の I B U F に複数の F M が入力されるケースを説明する。いま i F M 数 = $N \times n$ （ N ：入力並列度）であったとする。各入力 c h には n 面のデータが入力されることになる。そこで、I B U F 内を入力面数 n で分割し、分割されたそれぞれの領域内でスライディングウィンドウ処理を行う。各入力 c h に入力された n 面のデータを図 5 に示すように分割して I B U F に格納する。

[0046] 図 5 は、1 個の入力 c h に入力された i F M が分割されて I B U F に格納される様子を示す図である。図 6 は、分割されて I B U F に格納された i F M が処理される様子を示す図である。図は、I B U F が 4 つの S R A M（S R A M 1、…、S R A M 4）から構成されている例である。すなわち、I B U F は物理的に 4 つの S R A M に分かれている。

[0047] 各 S R A M（S R A M 1、…、S R A M 4）は仮想的に n 個に分割される。S R A M の各分割領域には、i F M の同じ領域（座標）のデータが格納されるようにする。具体的には、S R A M 1 の分割領域には、i F M の領域 A に対応するデータ i F M__0 ~ i F M__ $n-1$ が格納される。S R A M 2 の分割領域には、i F M の領域 B に対応する i F M__ n ~ i F M__ $2n-1$ が格納される。S R A M 3 の分割領域には、i F M の領域 C に対応する i F M

$iFM_2n \sim iFM_3n-1$ が格納される。SRAM4 の分割領域には、 iFM の領域Dに対応する $iFM_3n \sim iFM_4n-1$ が格納される。

[0048] すなわち、1個のSRAM内には iFM の同じ座標位置のデータが格納されるようにする。これは、 iFM の同じ座標位置のデータを同時に取り出して演算処理を行い、スライディングウィンドウ処理を行うためである。なお、SRAMの数を3個以上にすることで、 iFM の2領域（座標）にまたがるデータの処理（SRAM1 およびSRAM2 からのデータ読み込み）をしている場合でも、第3のSRAMへの書き込みを並行して行うことができる。

[0049] 図6の例では、まず、第1面の iFM_0 、第2面の iFM_1 、…、第 n 面の iFM_n-1 は同じ座標位置（領域A）のデータであり、これらはSRAM1の分割領域に格納される。第1面の iFM_n 、第2面の iFM_n+1 、…、第 n 面の iFM_2n-1 は同じ座標位置（領域B）のデータであり、これらはSRAM2の分割領域に格納される。第1面の iFM_2n 、第2面の iFM_2n+1 、…、第 n 面の iFM_3n-1 は同じ座標位置（領域C）のデータであり、これらはSRAM3の分割領域に格納される。第1面の iFM_3n 、第2面の iFM_3n+1 、…、第 n 面の iFM_4n-1 は同じ座標位置（領域D）のデータであり、これらはSRAM4の分割領域に格納される。

[0050] SRAM1に格納される iFM_0 、 iFM_1 、…、 iFM_n-1 は同じ座標位置のデータであり、同じタイミングで演算処理され、同じタイミングで不要になる。不要になったデータが格納されたSRAMは順次開放され、 iFM の続きのデータである、第1面の iFM_4n 、第2面の iFM_4n+1 、…、第 n 面の iFM_5n-1 を格納する。このように、スライディングウィンドウ処理が行われる。

[0051] 1個のSRAMの領域において、仮想的に分割された各領域サイズを「分割領域サイズ」と呼ぶことにする。分割領域サイズは、各領域において格納できる iFM のデータ量に対応する。使用済データは分割領域サイズ単位で

捨てることができるので、分割領域サイズ単位でスライディングウィンドウ処理を行い、SRAMの使用済領域を開放してiFMの続きのデータを入力することができる。

[0052] 具体的には、図6において、SRAM1のiFM₀、SRAM2のiFM_n、SRAM3のiFM_{2n}、SRAM4のiFM_{3n}の格納されている分割領域サイズ（第1面のデータが格納されている分割領域サイズ）において、スライディングウィンドウ処理が行われる。同時に、SRAM1のiFM₁、SRAM2のiFM_{n+1}、SRAM3のiFM_{2n+1}、SRAM4のiFM_{3n+1}の格納されている分割領域サイズ（第2面のデータが格納されている分割領域サイズ）において、スライディングウィンドウ処理が行われる。同時に、SRAM1のiFM_{n-1}、SRAM2のiFM_{2n-1}、SRAM3のiFM_{3n-1}、SRAM4のiFM_{4n-1}の格納されている分割領域サイズ（第n面のデータが格納されている分割領域サイズ）において、スライディングウィンドウ処理が行われる。

[0053] このようにして、本発明では、複数の演算部のそれぞれにおいて時分割で複数のFMを処理するようになっており、これにより大きなネットワークに対応することができる。

[0054] なお、iFMIDは面毎に0、1、2、…、n-1と増えていくので、iFMIDによってアドレスオフセットを変化させることで図6のようにiFMをIBUFに配置することができる。図7は、n=2の時の、iFMIDによってアドレスオフセットを変化させるイメージを示す図である。なお、n=1の場合はiFMIDを参照する必要が無いのでiFMIDをゼロ固定でよい。

[0055] 図8は、本実施形態に係る演算処理装置における、IBUFへのデータ書き込み制御フローを示す図である。ステップS1からiFM分割ループが始まる。ステップS2において、IBUFが書き込み可能か否かを判定する。IBUFが書き込み不可能の場合は、ステップS3に進み、待機（wait

) し、ステップ S 2 に戻る。

[0056] I B U F が書き込み可能の場合は、ステップ S 4 に進む。ステップ S 4 から分割領域サイズループが始まり、ステップ S 5 から i F M 数 (n) ループが始まる。ステップ S 6 においてアドレス計算を行い、ステップ S 7 において、データ入力があるか否かを判定する。データ入力がない場合は、ステップ S 8 に進み、待機 (w a i t) し、ステップ S 7 に戻る。

[0057] データ入力がある場合は、ステップ S 9 に進み、I B U F へデータ書き込みを行う。ステップ S 10 では、i F M 数 (n) 分の処理を行った否かを判定し、行っていない場合はステップ S 5 に戻り、行った場合はステップ S 11 に進む。ステップ S 11 では、分割領域サイズ分の処理を行った否かを判定し、行っていない場合はステップ S 4 に戻り、行った場合はステップ S 12 に進む。ステップ S 12 では、i F M 分割が完了したか否かを判定し、完了していない場合はステップ S 1 に戻り、完了した場合は、処理 (I B U F へのデータ書き込み処理) を終了する。

[0058] 次に、上述のように I B U F に入力されたデータを読み出すときの処理について説明する。各入力 c h には複数の i F M が入っているが、仮に 2 面の i F M (i F M 1、i F M 2) が入力されているとする。そして、i F M のデータ配置と、それに対して施すフィルタの係数が図 9 のようであったとする。図 9 は、i F M のデータ配置と係数の一例を示す図である。

[0059] i F M 1 において、データ配置は図 9 (a) のようになる。a 1、b 1、…、p 1 はデータを示し、4 × 4 のデータが配置されている。そのうち、3 × 3 のデータ (a 1、b 1、c 1、d 1、e 1、f 1、g 1、i 1、j 1、k 1) の各々に対して施すフィルタの係数は図 9 (b) のようになる。w 1、w 2、…、w 9 は各データ係数を示す。

[0060] 同様に、i F M 2 において、データ配置は図 9 (c) のようになる。a 2、b 2、…、p 2 はデータを示し、4 × 4 のデータが配置されている。そのうち、3 × 3 のデータ (a 2、b 2、c 2、d 2、e 2、f 2、g 2、i 2、j 2、k 2) の各々に対して施すフィルタの係数は図 9 (d) のようにな

る。 x_1 、 x_2 、 $\dots$ 、 x_9 は各データ係数を示す。

[0061] タイミングチャートは図10に示すようになる。図10は、本実施形態に係る演算処理装置における、Convolution処理時、iFM数=2Nの時のタイミングチャートを示す図である。3×3の9個のデータ $a_1 \sim k_1$ と係数 $w_1 \sim w_9$ のフィルタ演算結果がAであり、3×3の9個のデータ $a_2 \sim k_2$ と係数 $x_1 \sim x_9$ のフィルタ演算結果がBである。フィルタ演算結果は、第2加算器において時間方向で累積加算され、その結果がFFに保持される。第2加算器における最終的な累積加算結果が Σ である。

[0062] すなわち、iFM1とiFM2が時分割で入力され、第2加算器で累積加算され、その結果がFFに保持される。第2加算器の結果がAやA+Bの時は、まだ累積加算の途中であるので、非線形変換部には出力されない。第2加算器における累積加算が完了すると、その累積加算結果 Σ が非線形変換部に出力され、FFは初期値に戻る。非線形変換部は非線形変換 $f(\Sigma)$ を行う。

[0063] 図11は、本実施形態に係る演算処理装置における、SRAM読み出し制御部の制御フローの一例を示す図である。まず、ステップS21において、oFM__Vループが始まる。oFM__Vループは、oFMのデータをV方向（横方向）に処理するループである。そして、ステップS22において、oFM__Hループが始まる。oFM__Hループは、oFMのデータをH方向（縦方向）に処理するループである。ステップS23でWBUFアクセスアドレスを生成し、並列してステップS26でIBUFアクセスアドレスを生成する。

[0064] ステップS23でWBUFアクセスアドレスを生成すると、ステップS24でデータ格納判定を行う。データ格納判定がOKの場合は、ステップS29のプーリングループに進む。データ格納判定がNGの場合は、ステップS25で待機（wait）し、ステップS24に戻る。

[0065] また、ステップS26でIBUFアクセスアドレスを生成すると、ステップS27でデータ格納判定を行う。データ格納判定がOKの場合は、ステッ

プS 2 9のプーリンググループに進む。データ格納判定がNGの場合は、ステップS 2 8で待機（wait）し、ステップS 2 7に戻る。

[0066] ステップS 2 9のプーリンググループの後は、ステップS 3 0のi FM数グループ、ステップS 3 1のフィルタサイズグループと進み、ステップS 3 2において演算部の処理が行われる。ステップS 3 3では各グループを終えてよいかを判定し、終わっていない場合は各グループに戻り、全てのグループが終わった場合、制御フローは終了する。

[0067] I B U Fのデータ格納判定は、プーリング処理も含めてフィルタ処理に必要な全i FMの座標データがI B U Fに格納済みであるかの判定であり、最内側グループは判定後となっている。なお、判定方法は種々変更できるものとする。また、o FMのHサイズ単位で判定するのであればo FM_Hグループもデータ格納判定後に置く。

[0068] なお、本実施形態で説明したメモリ制御およびメモリ格納方法については、複数フレームを入力として出力データを生成するアルゴリズムにも応用できる。例えば、フレーム間相関を利用するM P E G等の圧縮処理やノイズリダクション処理にも適用できる。

[0069] （実施形態2）

実施形態1では、1つのC N Nの層を処理する回路とそこへ入力するi FMをどのようにI B U Fに展開し、処理するかについて説明した。実施形態2では、特に、o FM数（o FMの面数）が回路的な並列度Mを越えるケースに対応する。

[0070] 図1 2は、本実施形態のC N Nの処理イメージを示す図である。図1 2は、o FM数=2 Mの例を示し、1つの入力c hのみを表示している。o FM数が出力並列度Mを超え、o FM数=M×m（m>1）であるとする。C N Nは、全てのi FMの同一座標データからo FMの1画素を計算する処理である。

[0071] この時、o FMを1面ずつ作成すると、i FMはI B U F内でスライディングウィンドウ処理されているので、最初の部分のデータが上書きされてI

B U F内に存在しなくなる。このため、新しい○ F Mを作成する処理毎に、D R A Mから i F Mを読み直さなければならなくなってしまう。

[0072] そこで、本実施形態では、 I B U Fに格納されているデータで処理できる○ F Mを全て作成してしまうように制御する。具体的には、 I B U Fに入力されている i F M（一部しか格納されていない）から、1個の○ F Mについて可能なだけの処理回数を実行したら、 I B U Fを更新せずに次の○ F Mに対応する処理を行う（係数は適宜切り替える）。

[0073] このときの「可能なだけの処理回数」は、 I B U F上で単位としている分割領域サイズに設定すると効率的である。なぜなら I B U Fを構成する1個の S R A Mに入る i F M 1面分のデータが分割領域サイズであり、この単位で処理すると、出力したい○ F M数分の処理を終えた時点で S R A M 1個分のデータが使用済みとなるので開放できるからである。なお、分割領域サイズは i F Mにおけるデータ量であって、○ F M上のデータ量とは必ずしも一致しない。これは、プーリング処理による縮小が発生するためである。

[0074] ピクセル単位で○ F Mを変えない理由は、ある程度のデータ量を D R A M上に転送する方が、効率がよいためである。バッファのサイズや転送効率に問題がなければ、ピクセル単位で切り替えてもよい。

[0075] このように、本実施形態では、スライディングウィンドウ処理を行う前に、入っているデータでできる処理の全てを行う。なお、出力されるデータは、○ F Mが1個ずつ完成して出力されるのではなく、複数の○ F Mの一部分が少しずつ出力されて、最終的に複数の○ F Mが完成して出力される。

[0076] 図 1 3 は、本実施形態に係る演算処理装置における、 S R A M読み出し制御部の制御フローの一例を示す図である。「d Mカウンタ」は、分割領域サイズ分のループカウンタに相当する。まず、ステップ S 4 1で、d Mカウンタ=0とする。そして、ステップ S 4 2から○ F M__Vループが始まり、ステップ S 4 3から○ F M__Hループが始まる。○ F M__Vループは、○ F MのデータをV方向（横方向）に処理するループであり、○ F M__Hループは、○ F MのデータをH方向（縦方向）に処理するループである。

- [0077] ステップS44でWBUFアクセスアドレスを生成し、並行してステップS47でIBUFアクセスアドレスを生成する。
- [0078] ステップS44でWBUFアクセスアドレスを生成すると、ステップS45でデータ格納判定を行う。データ格納判定がOKの場合は、ステップS50のプーリンググループに進む。データ格納判定がNGの場合は、ステップS46で待機（wait）し、ステップS45に戻る。
- [0079] また、ステップS47でIBUFアクセスアドレスを生成すると、ステップS48でデータ格納判定を行う。データ格納判定がOKの場合は、ステップS50のプーリンググループに進む。データ格納判定がNGの場合は、ステップS49で待機（wait）し、ステップS48に戻る。
- [0080] ステップS50のプーリンググループの後は、ステップS51のiFM数グループ、ステップS52のフィルタサイズグループと進み、ステップS53において演算部の処理が行われる。ステップS54およびステップS55では各グループを終えてよいかを判定し、終わっていない場合は各グループに戻り、全てのグループが終わった場合は、ステップS56に進む。
- [0081] ステップS56では、dMカウンタが分割領域サイズ以上か否かを判定する。dMカウンタが分割領域サイズ以上である場合は、ステップS57に進み、dMカウンタ=0とする。そして、ステップS58において、係数を次のoFMへ送り、ステップS60へ進む。
- [0082] dMカウンタが分割領域サイズ以上でない場合（dMカウンタが分割領域サイズ未満である場合）は、ステップS59に進み、dMカウンタの値を増やす。そして、ステップS60へ進む。
- [0083] ステップS60では、oFM__VループおよびoFM__Hループを終えてよいかを判定し、終わっていない場合は各グループに戻り、全てのグループが終わった場合、制御フローは終了する。
- [0084] 図14は、本実施形態に係る演算処理装置における、iFM数=2N、oFM数=3Mの場合のタイミングチャートの一例を示す。分割領域サイズを4Dとすると、プーリングがあるためoFM上のデータ数はoFM数×D個

分となる。

[0085] データ出力部ではプーリング後の有効データ数をカウントして、現在どの○FMを受け取っているかを識別できるので、識別結果に従ってDRAM上に適切に展開する。図15は、データが書き出されるDRAMの出力イメージを示す図である。図中の数字は出力順を示す。

[0086] このように、本実施形態では、IBUF管理部内のSRAM読み出し制御部が、同じ入力データから、出力するm枚分の○FMのフィルタ処理を実行するようにIBUFからデータを取り出す。そして、データ出力部が、有効データをカウントしてどの○FMデータであるかを識別し、適切にアドレスを生成してDRAMに出力する。これにより、iFMを再読み込みすることなく出力並列度を越える○FM数を出力できる。

[0087] すなわち、SRAM読み出し制御部は、IBUFからiFMデータの一部データを読み出す。演算部は、SRAM読み出し制御部から入力されたiFMの一部データの演算処理を行い（読み込んだ所のデータを処理し）、○FMデータの一部データを作成する。そして、終わったら、iFMデータの次の一部データの処理へと進む。

[0088] 以上、本発明の一実施形態について説明したが、本発明の技術範囲は上記実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲において構成要素の組み合わせを変えたり、各構成要素に種々の変更を加えたり、削除したりすることができる。

[0089] 本明細書において「前、後ろ、上、下、右、左、縦、横、行、列」などの方向を示す言葉は、本発明の装置におけるこれらの方向を説明するために使用している。従って、本発明の明細書を説明するために使用されたこれらの言葉は、本発明の装置において相対的に解釈されるべきである。

産業上の利用可能性

[0090] 本発明は、畳み込みニューラルネットワークを用いたディープラーニングを行う演算処理装置に広く適用でき、iFMサイズやiFM数が大きくてデータ格納メモリ（IBUF）に入りきらないサイズでも、フィルタサイズ分

のオーバーラップをせずに演算処理とデータ更新をシームレスに実施することができる。

符号の説明

- [0091]
- 1 演算処理装置
 - 2 コントローラ
 - 3 データ入力部
 - 4 係数入力部
 - 5 I B U F (データ格納メモリ) 管理部
 - 6 W B U F (係数格納メモリ) 管理部
 - 7 演算部
 - 8 データ出力部
 - 9 D R A M (格納用外部メモリ)
 - 10 バス
 - 51 カウンタ
 - 52 S R A M書き込み制御部
 - 53 S R A M読み出し制御部
 - 71 演算制御部
 - 72 フィルタ演算部
 - 73 第1加算器
 - 74 第2加算器
 - 75 F F (フリップフロップ)
 - 76 非線形変換部
 - 77 プーリング処理部

請求の範囲

- [請求項1] Convolution処理とFullConnect処理を行うディープラーニング用の演算処理装置であって、
- 3個以上のSRAMで構成されたデータ格納メモリと、
- 入力特徴量マップデータから入力有効データ数をカウントして座標に変換するカウンタと、
- 前記カウンタが変換した座標を前記データ格納メモリにおけるアドレスに変換して、前記入力特徴量マップデータを前記データ格納メモリに格納するSRAM書き込み制御部と、
- 外部からのリクエストを受けて、前記データ格納メモリから所定の方法で格納されている前記入力特徴量マップデータを取り出すSRAM読み出し制御部と、
- を有するデータ格納メモリ管理部と；
- 係数を格納する係数格納メモリを有し、前記データ格納メモリ管理部のステータスを参照して、前記データ格納メモリ管理部から取り出すデータに対応する係数を前記係数格納メモリから取り出す係数格納メモリ管理部と；
- 格納用外部メモリから、前記入力特徴量マップデータを取得するデータ入力部と；
- 前記格納用外部メモリから、前記係数を取得する係数入力部と；
- 入力N並列、出力M並列の構成（ $N, M \geq 1$ の正数）で、前記データ格納メモリから前記入力特徴量マップデータを取得し、前記係数格納メモリから前記係数を取得して、フィルタ処理、累積加算処理、非線形演算処理およびプーリング処理を行う演算部と；
- 前記演算部から出力されるM並列のデータを連結して、出力特徴量マップデータとして前記格納用外部メモリに出力するデータ出力部と；
- ；
- 前記データ格納メモリ管理部、前記係数格納メモリ管理部、前記デ

ータ入力部、前記データ出力部、および前記演算部を制御するコントローラと；

を有し、

前記カウンタは、入力された前記入力特徴量マップデータのIDを生成し、

前記SRAM書き込み制御部は、前記データ格納メモリを構成する前記SRAMの各々を仮想的に複数の領域に分割し、前記IDによって書き込む領域を切り替えるとともに、同じSRAMには同一座標の異なる入力特徴量マップデータが格納されるように制御し、

前記SRAM読み出し制御部は、取り出したい前記入力特徴量マップデータのIDに対応する前記領域にアクセスするように前記アドレスを切り替え、

前記演算部は、

N並列でフィルタ演算を実行するフィルタ演算部と、

前記フィルタ演算部の演算結果を全て累積加算する第1加算器と、

前記第1加算器の累積加算結果を後段で累積加算する第2加算器と、

前記第2加算器の累積加算結果を保持するフリップフロップと、

1つのデータ格納メモリに格納されている複数の入力特徴量マップデータを取り出すために、入力特徴量マップデータのIDを生成して、前記SRAM読み出し制御部から入力される入力特徴量マップデータと同じIDの前記係数を前記係数格納メモリ管理部から取得するよう制御し、全ての入力特徴量マップデータを取り終わるまでは、前記第2加算器が前記第1加算器の累積加算結果を累積加算するよう制御する演算制御部と、

を有する演算処理装置。

[請求項2]

前記演算部は、同一座標の入力特徴量マップデータごとに演算処理

を実行し、

前記SRAM書き込み制御部は、前記同一座標の入力特徴量マップデータの演算処理が終わるごとに、前記同一座標の入力特徴量マップデータが格納されていた前記領域を開放して、前記入力特徴量マップデータの続きのデータを格納する

請求項1に記載の演算処理装置。

[請求項3]

前記SRAM読み出し制御部は、1つの入力特徴量マップデータから、出力するm枚分の出力特徴量マップデータの演算処理を実行するように前記データ格納メモリから前記入力特徴量マップデータを読み出し、

前記データ出力部は、有効データをカウントしてどの出力特徴量マップデータであるかを識別し、適切に出力特徴量マップデータにおけるアドレスを生成して、前記格納用外部メモリに出力する、

請求項1または2に記載の演算処理装置。

[請求項4]

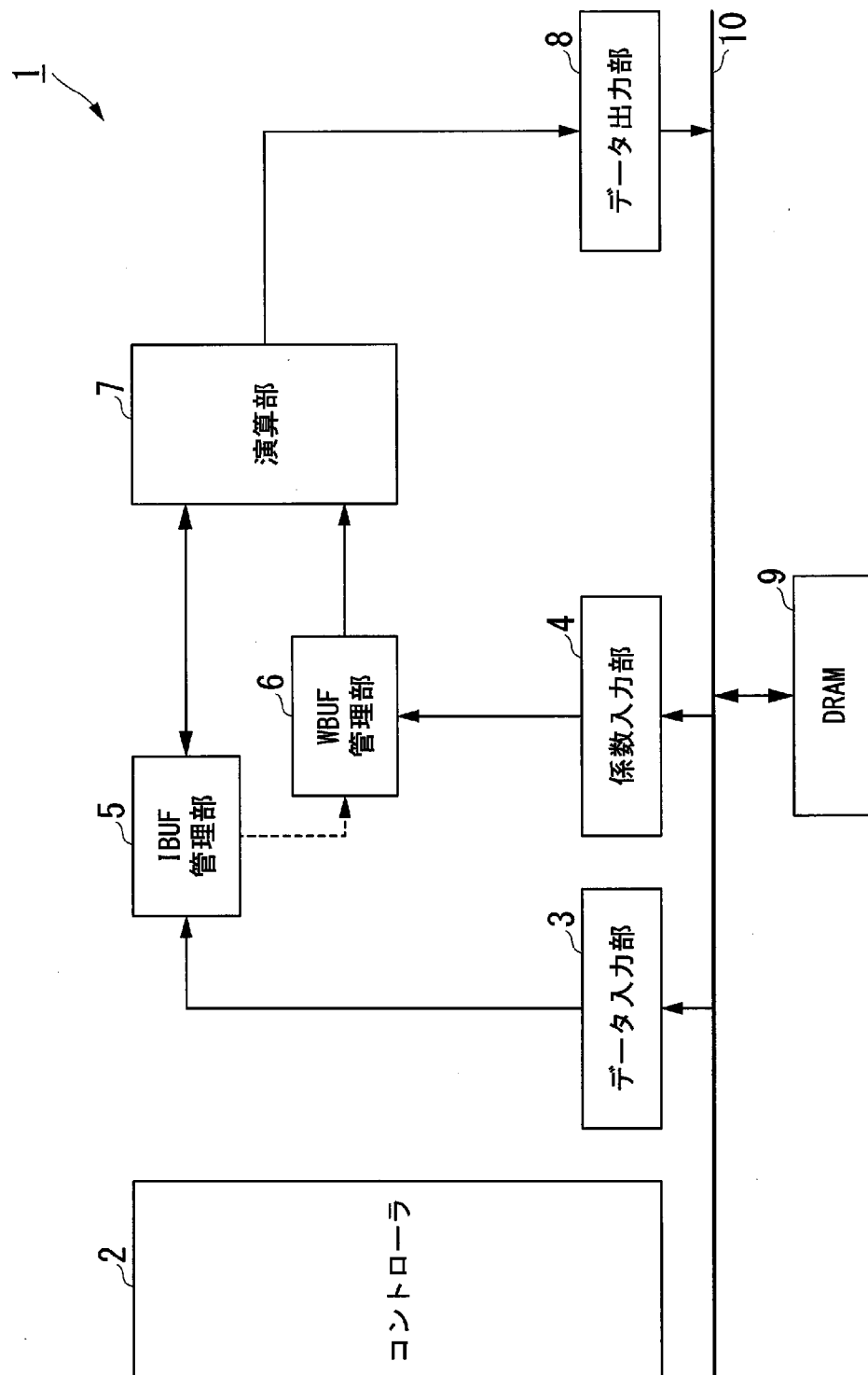
前記SRAM読み出し制御部は、前記データ格納メモリから前記入力特徴量マップデータの一部データを読み出し、

前記演算部は、前記SRAM読み出し制御部から入力された前記入力特徴量マップデータの一部データの演算処理を行い、前記出力特徴量マップデータの一部データを作成し、

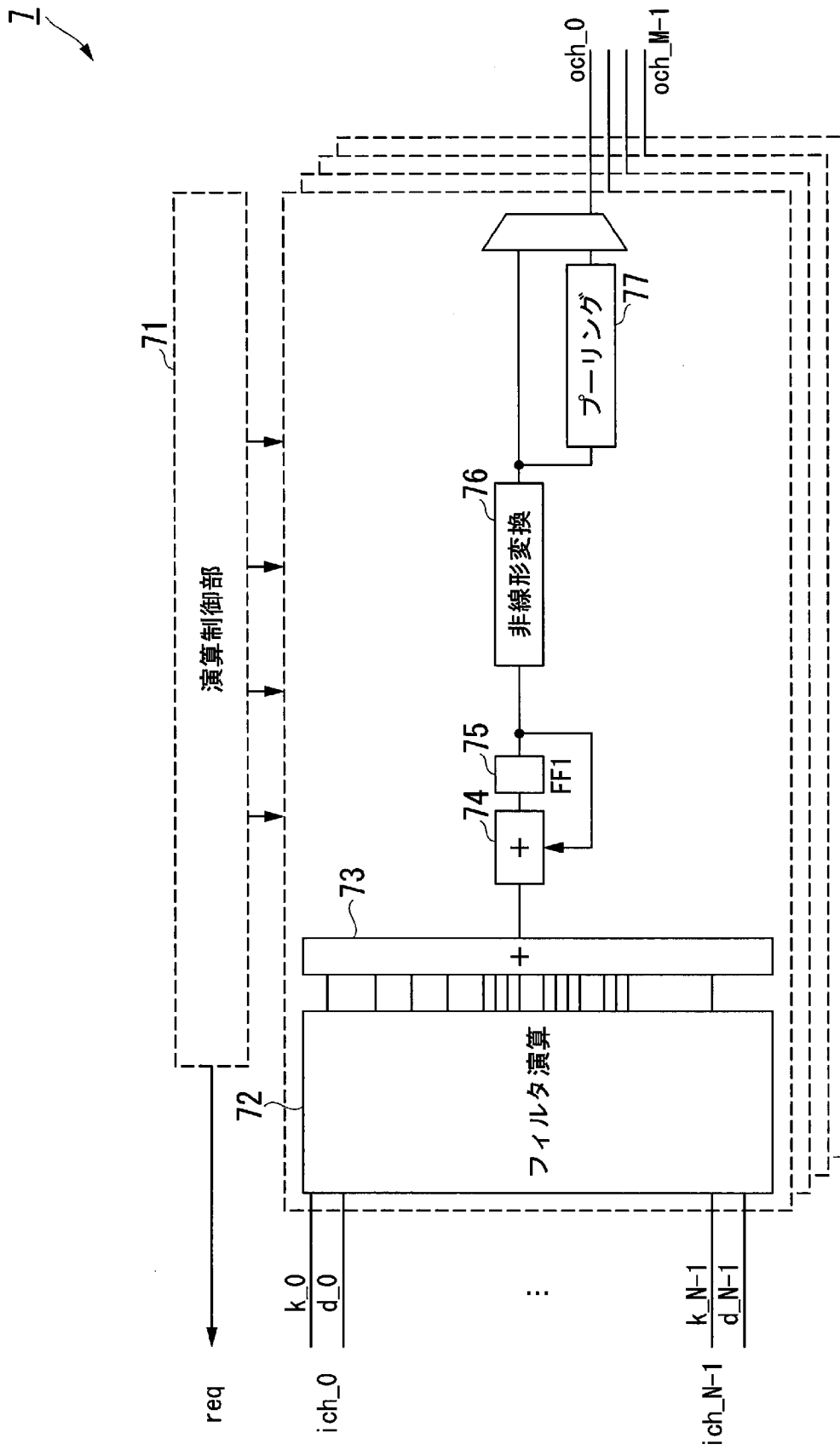
前記入力特徴量マップデータの前記一部データの処理が終わったら、前記入力特徴量マップデータの次の一部データの処理へと進む、

請求項1から3のいずれか1項に記載の演算処理装置。

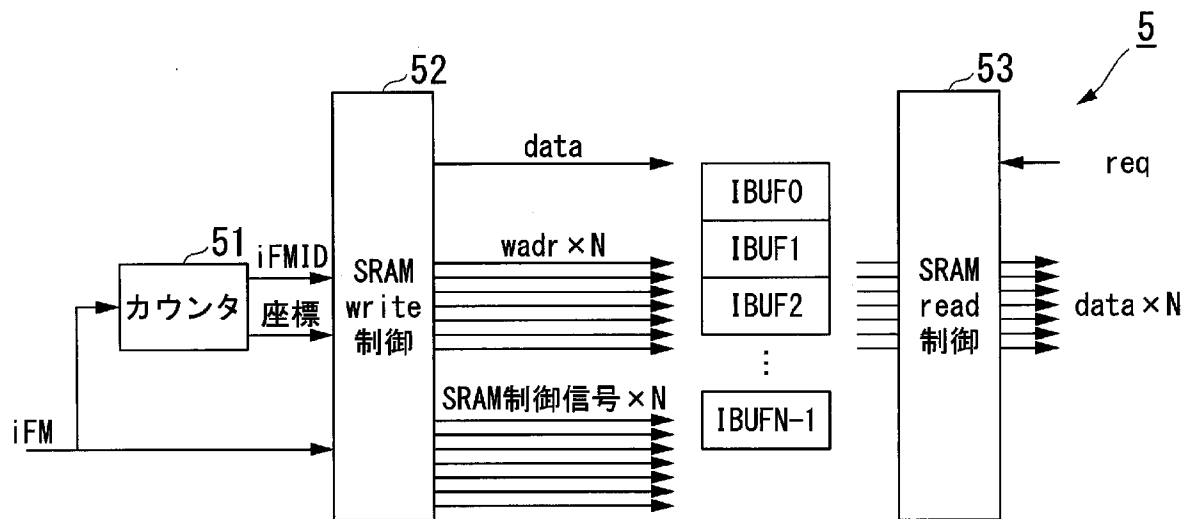
[図1]



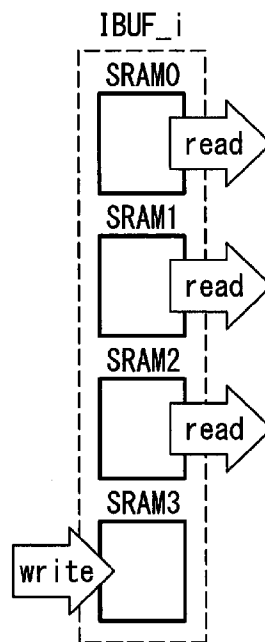
[図2]



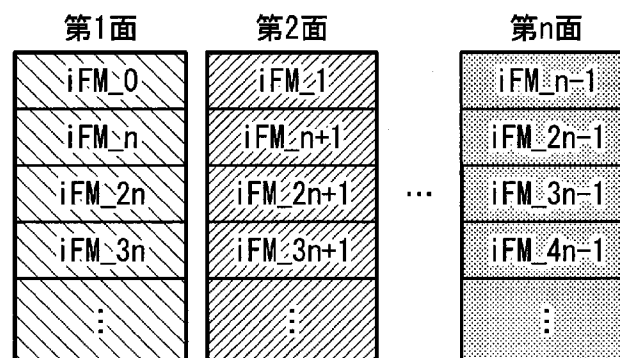
[図3]



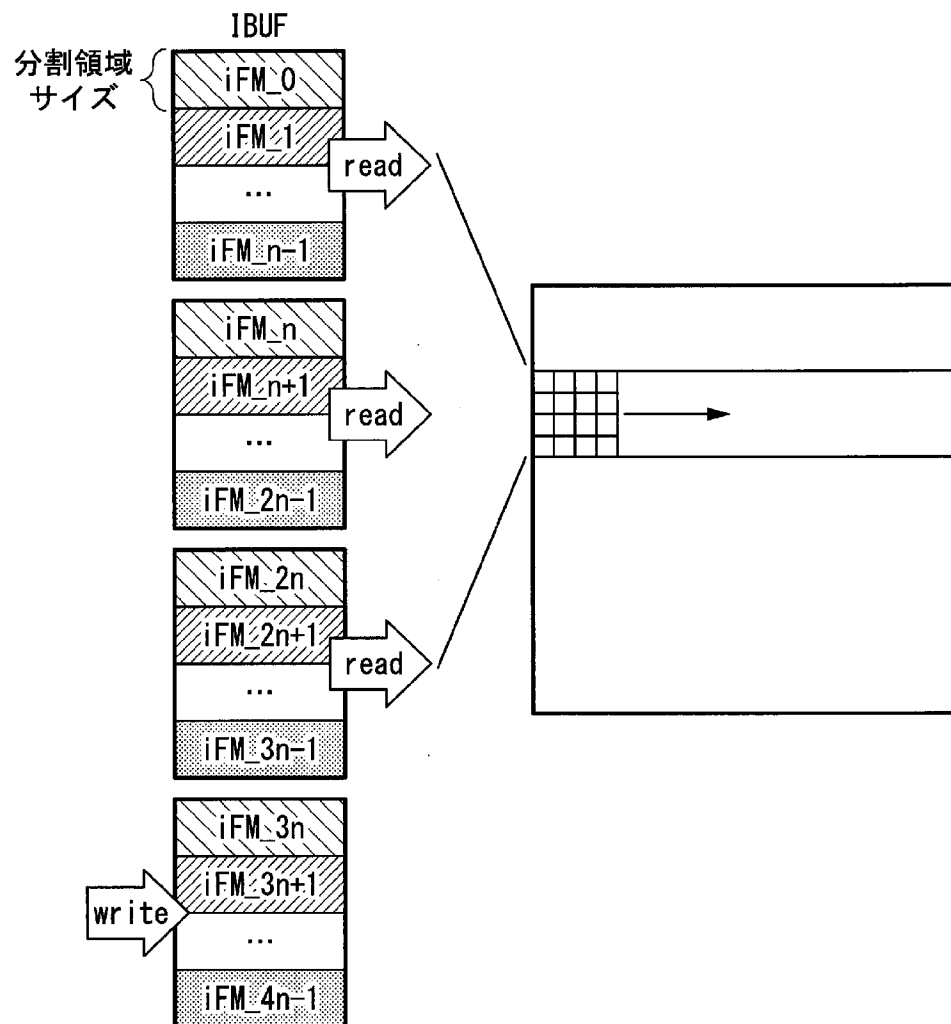
[図4]



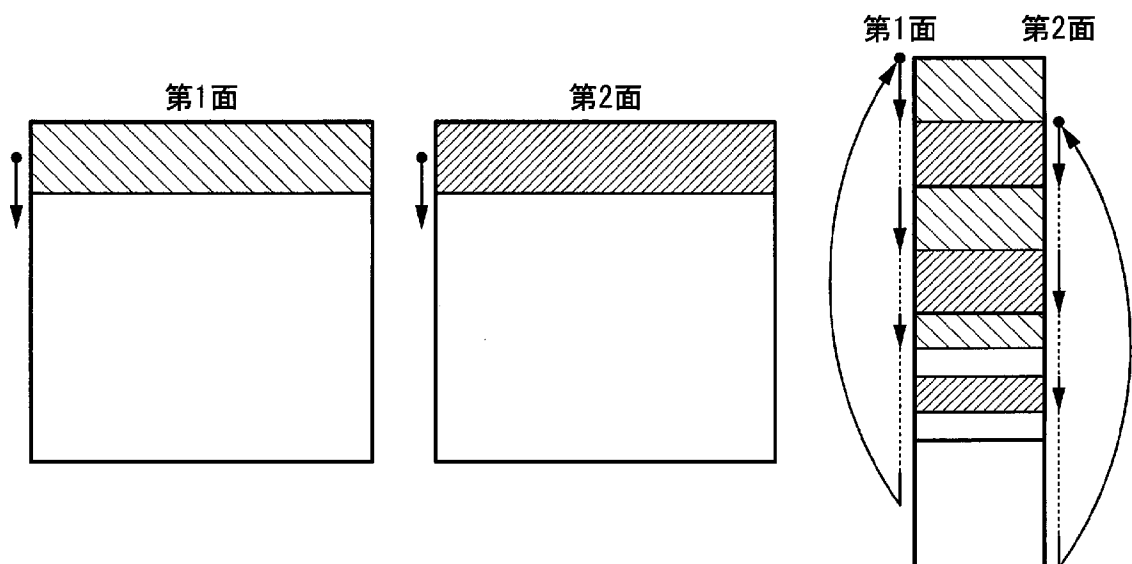
[図5]



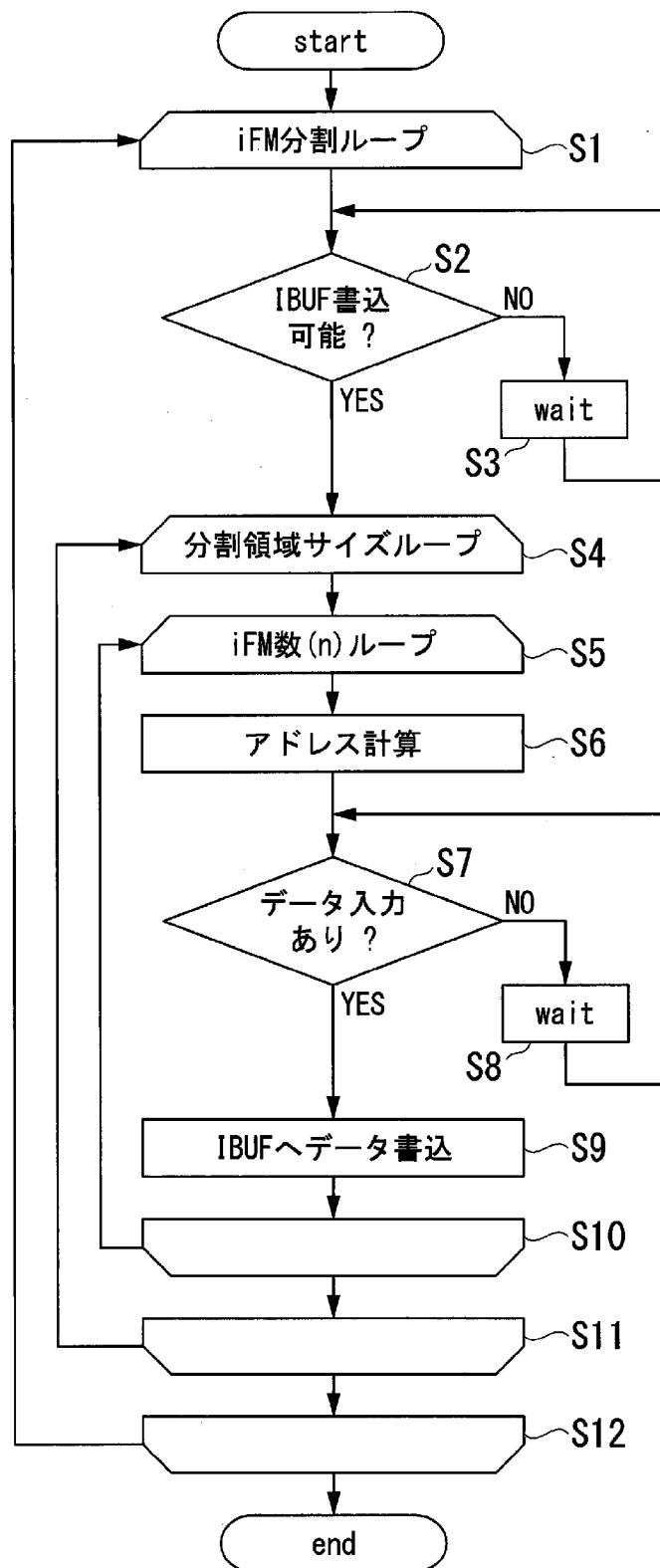
[図6]



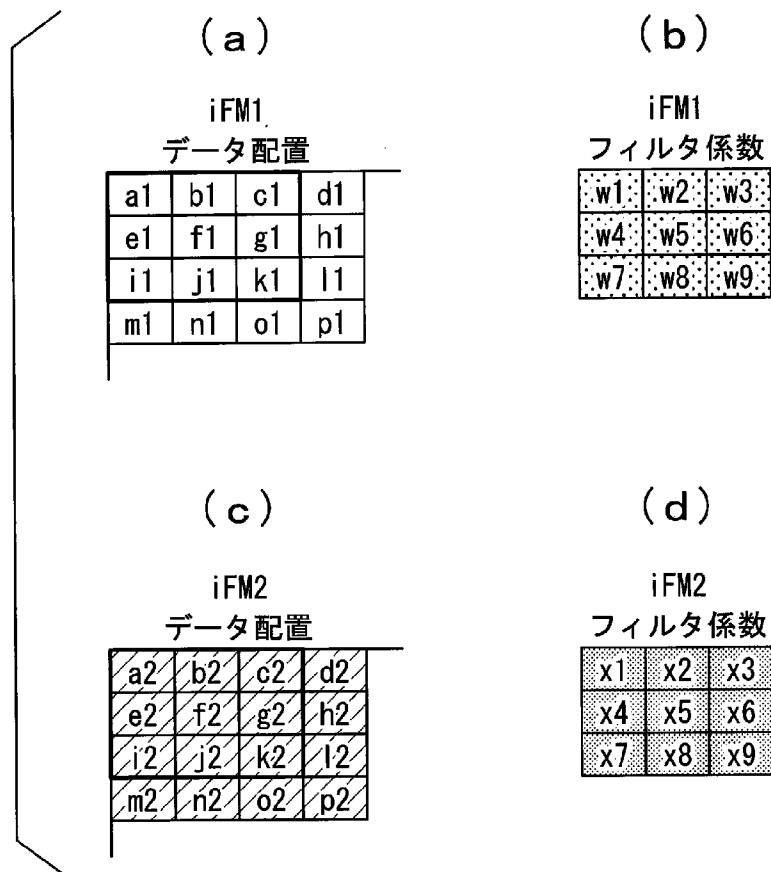
[図7]



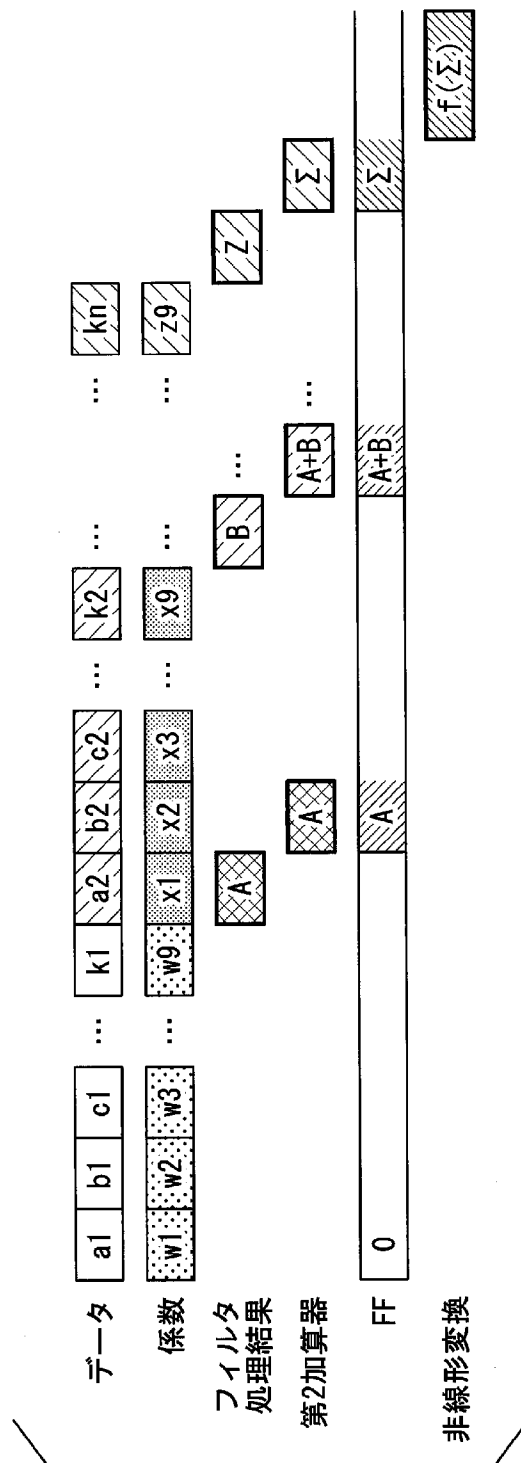
[図8]



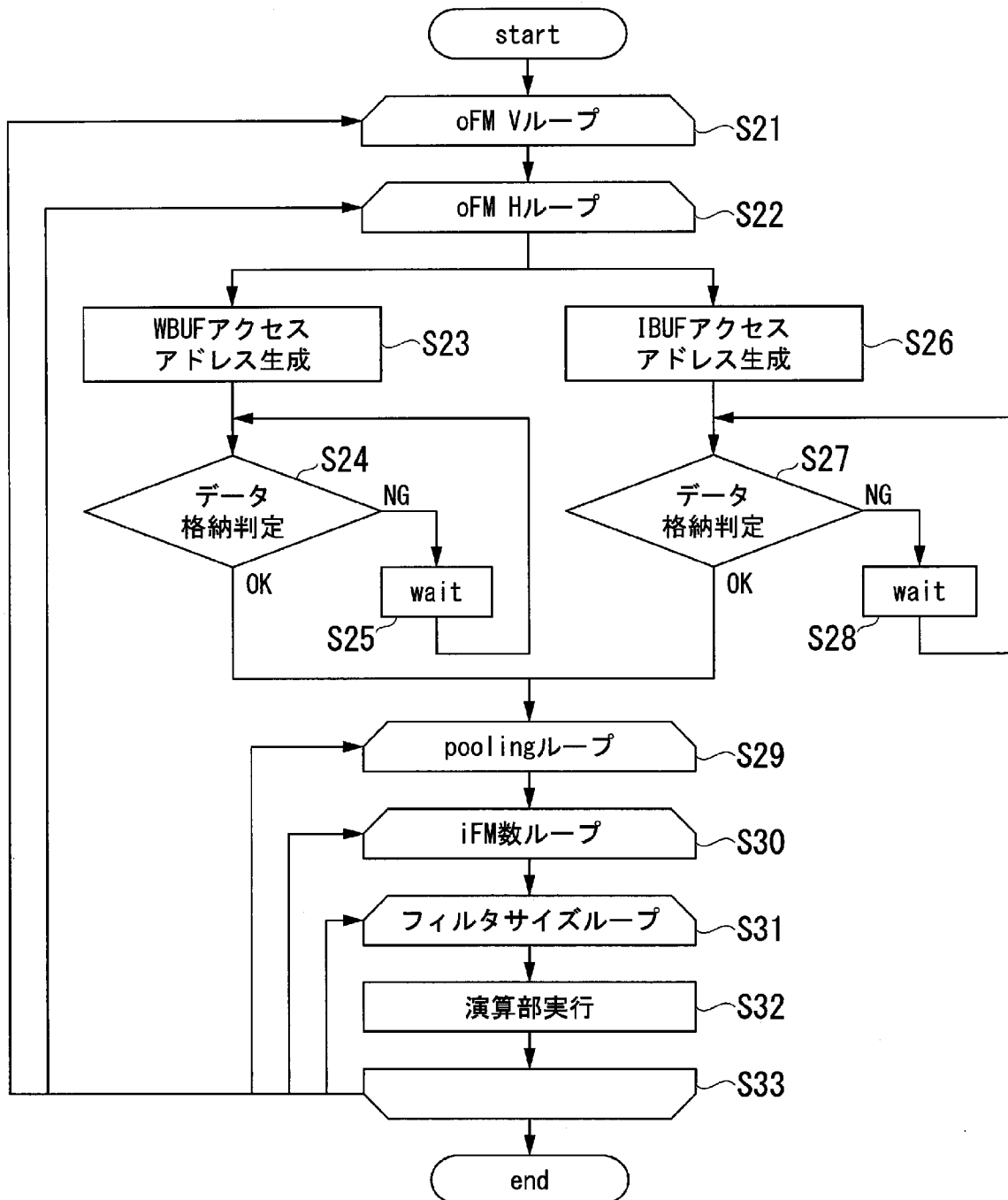
[図9]



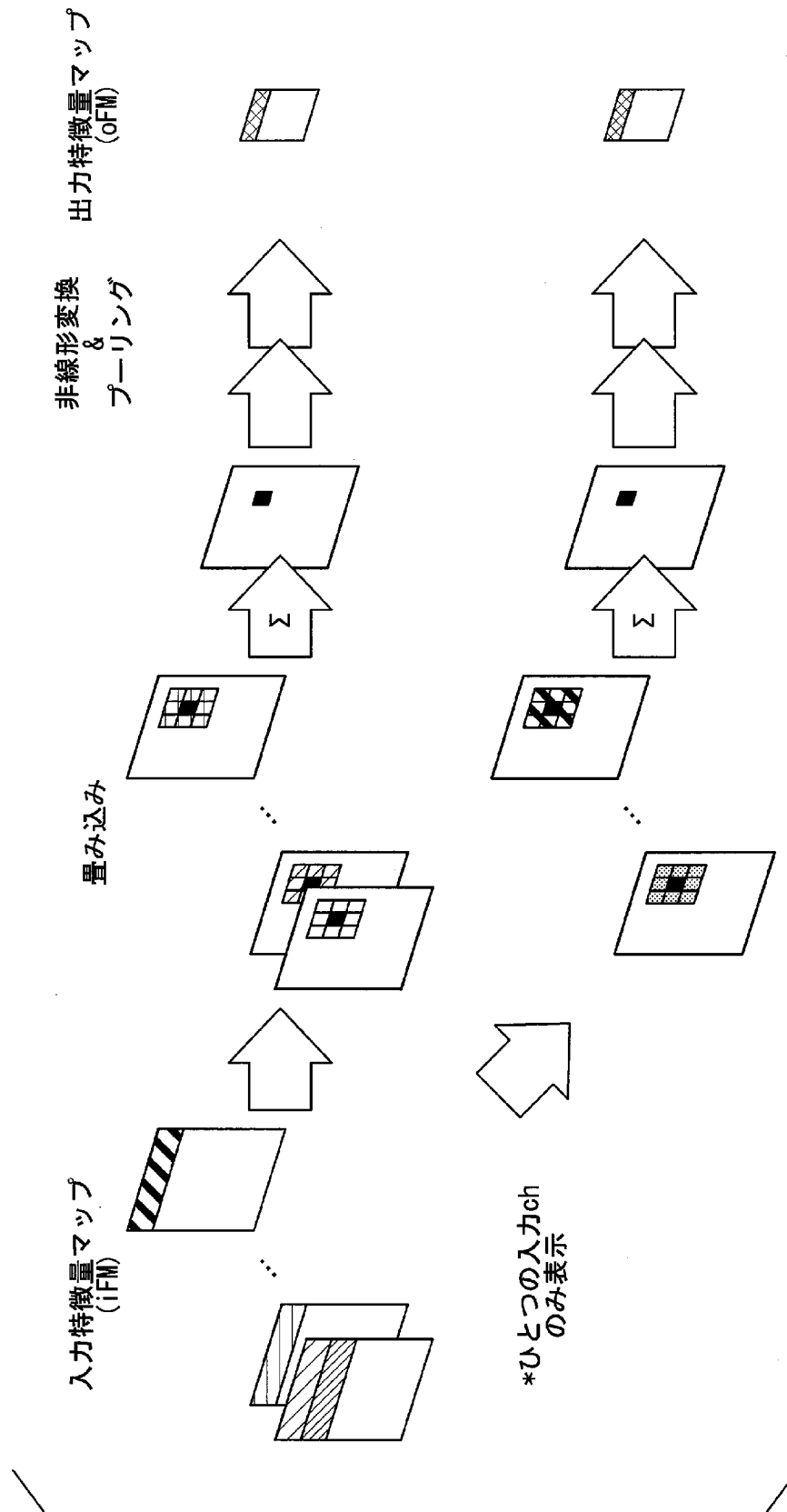
[図10]



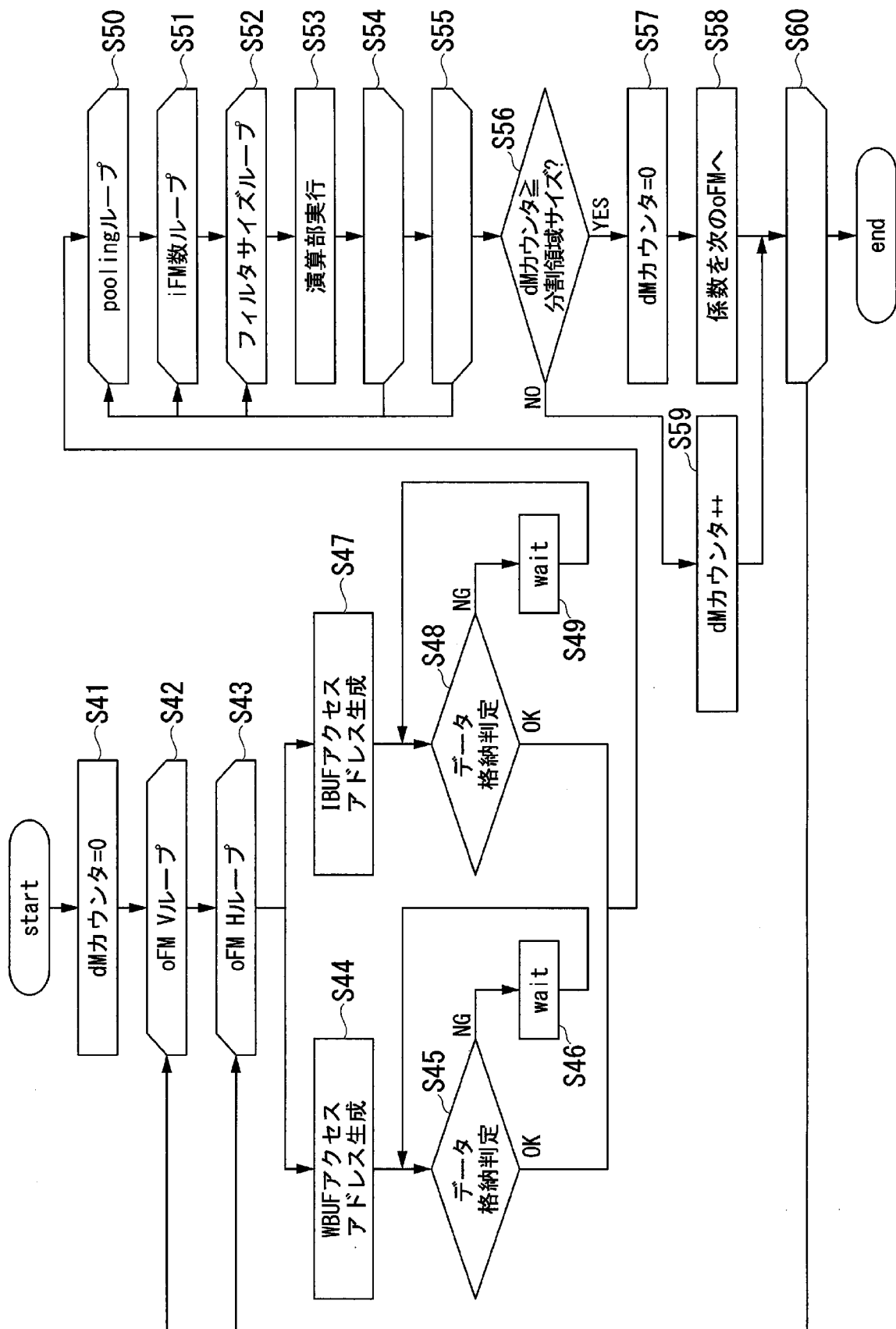
[図11]



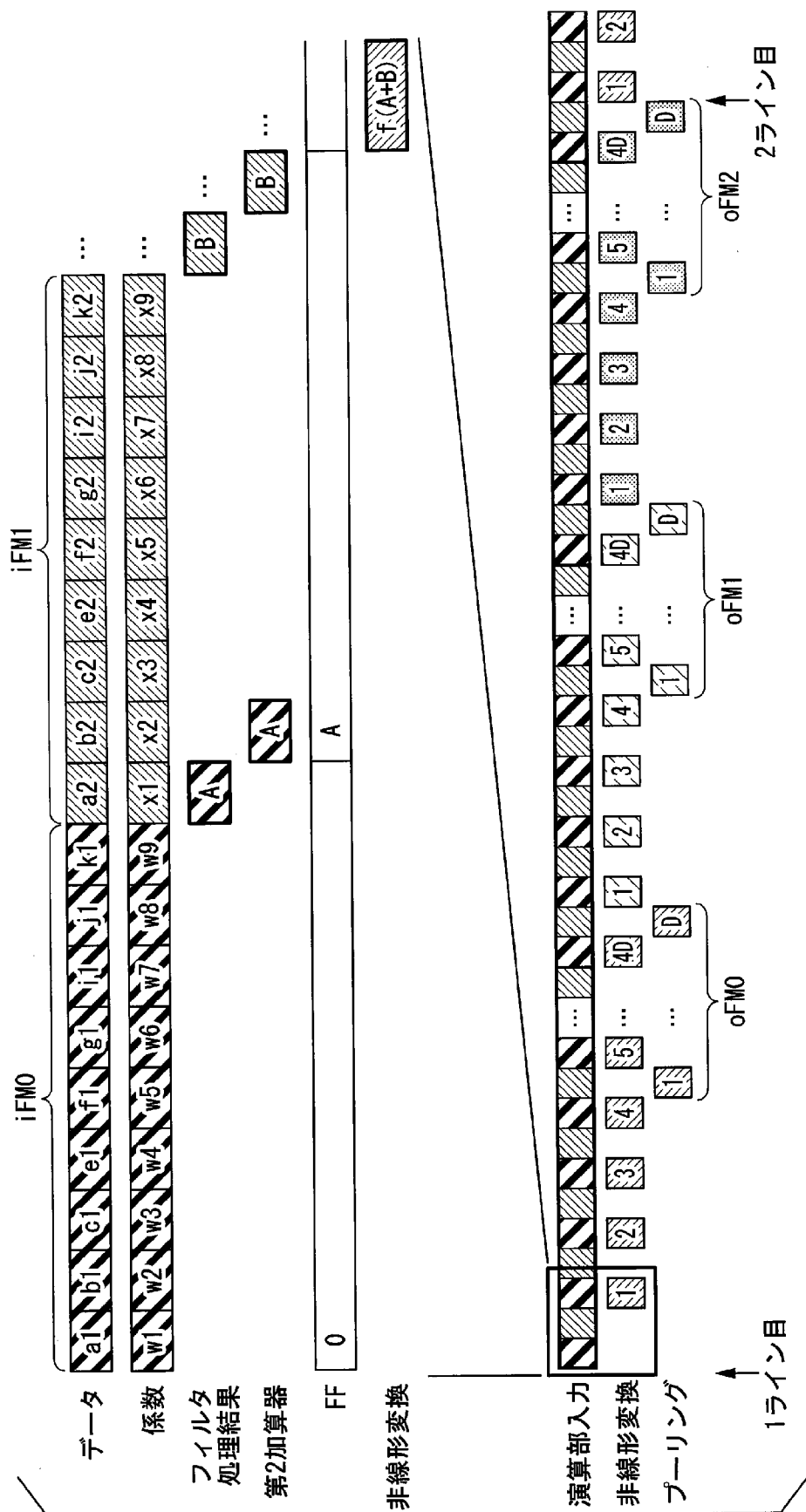
[図12]



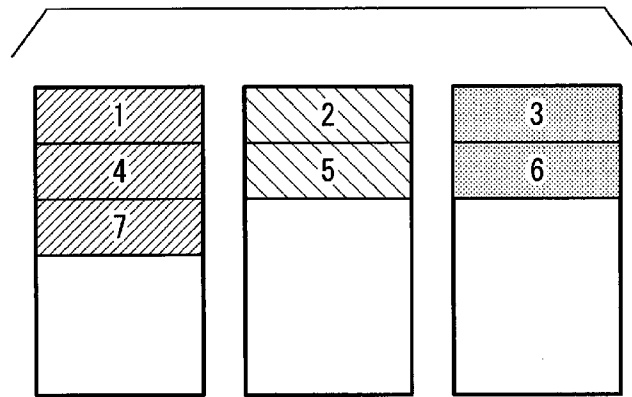
[図13]



[図14]



[図15]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/023974

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06N3/063 (2006.01) i, G06F17/16 (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. G06N3/063, G06F17/16

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)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2017-151604 A (DENSO CORP.) 31 August 2017 (Family: none)	1-4
A	US 2017/0277658 A1 (INTEL CORPORATION) 28 September 2017 & WO 2016/099779 A1 & EP 3035249 A1 & CN 107003989 A & KR 10-2017-0096105 A & TW 201640422 A	1-4
A	JP 2018-73102 A (CANON INC.) 10 May 2018 & US 2018/0121795 A1	1-4

☐ 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
21.08.2018

Date of mailing of the international search report
04.09.2018

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, G06F17/16(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. G06N3/063, G06F17/16

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）

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 2017-151604 A（株式会社デンソー）2017.08.31 （ファミリーなし）	1-4
A	US 2017/0277658 A1（INTEL CORPORATION）2017.09.28 & WO 2016/099779 A1 & EP 3035249 A1 & CN 107003989 A & KR 10-2017-0096105 A & TW 201640422 A	1-4
A	JP 2018-73102 A（キヤノン株式会社）2018.05.10 & US 2018/0121795 A1	1-4

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	「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
「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	「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
「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）	「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
「O」 口頭による開示、使用、展示等に言及する文献	「&」 同一パテントファミリー文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日

21.08.2018

国際調査報告の発送日

04.09.2018

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）

大桃 由紀雄

5 B

5583

電話番号 03-3581-1101 内線 3545