(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5月19日(19.05.2022)

WIPO | PCT

(10) 国際公開番号
WO 2022/102255 A1

(51) 国際特許分類:
G06N 3/08 (2006.01)

(21) 国際出願番号： PCT/JP2021/034898

(22) 国際出願日： 2021年9月22日(22.09.2021)

(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 2020-188356 2020年11月11日(11.11.2020) JP

(71) 出願人: 日立Ａｓｔｅｍｏ株式会社(**HITACHI ASTEMO, LTD.**) [JP/JP]; 〒3128503 茨城県ひたちなか市高場２５２０番地 Ibaraki (JP).

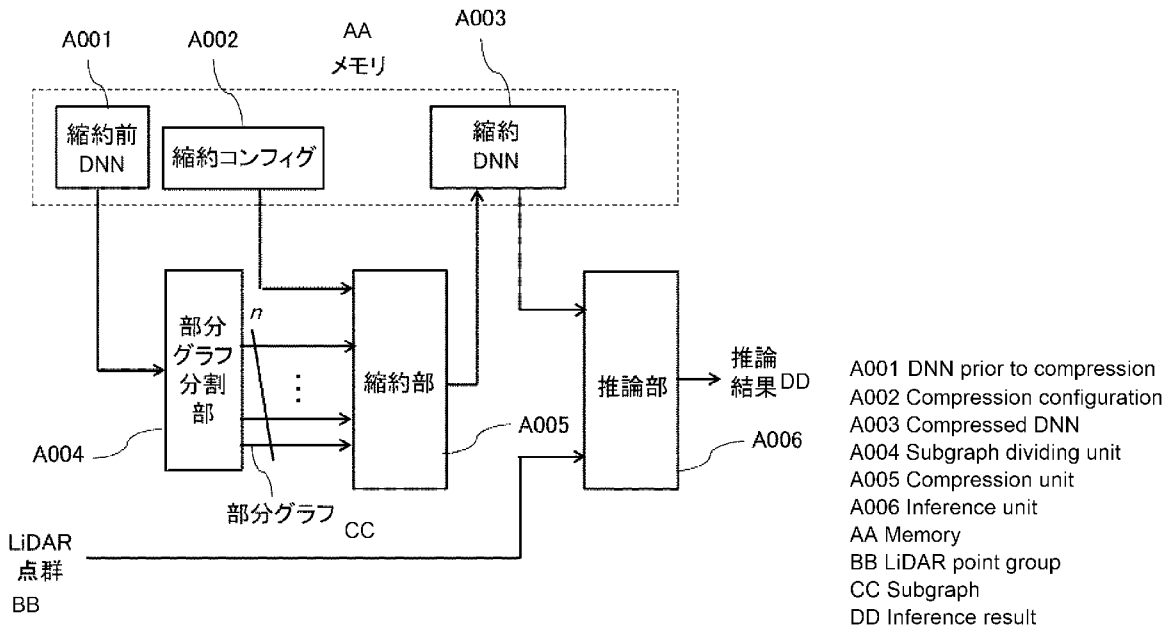
(72) 発明者: 村田 大智(**MURATA Daichi**); 〒1008280 東京都千代田区丸の内一丁目６番６号 株式会社日立製作所内 Tokyo (JP). 北山 晃(**KITAYAMA Akira**); 〒1008280 東京都千代田区丸の内一丁目６番６号 株式会社日立製作所内 Tokyo (JP). 伊藤 浩朗(**ITO Hiroaki**); 〒3128503 茨城県ひたちなか市高場２５２０番地 日立Ａｓｔｅｍｏ株式会社内 Ibaraki (JP). 黒田 昌芳(**KURODA Masayoshi**); 〒3128503 茨城県ひたちなか市高場２５２０番地 日立Ａｓｔｅｍｏ株式会社内 Ibaraki (JP).

(74) 代理人: 青稜特許業務法人(**SEIRYO I.P.C.**); 〒2200004 神奈川県横浜市西区北幸二丁目１０番２７号 Kanagawa (JP).

(54) **Title:** INFORMATION PROCESSING DEVICE AND NEURAL NETWORK COMPRESSION METHOD

(54) 発明の名称： 情報処理装置およびニューラルネットワーク縮約方法

図7



(57) **Abstract:** The present invention optimizes a compression algorithm to be applied to, as a unit, each subgraph of a neural network. A preferred aspect of the present invention is an information processing device for selecting an algorithm for compressing a neural network, said information processing device being provided with a subgraph dividing unit for dividing a neural network into subgraphs, and an optimization unit for outputting a compression configuration which associates each subgraph with one compression technique selected from a plurality of compression techniques.

WO 2022/102255 A1

［続葉有］

(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, 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, 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：ニューラルネットワークの部分グラフ単位で，適用する縮約アルゴリズムの最適化を実現する。 本発明の好ましい一側面は，ニューラルネットワークを縮約するアルゴリズムを選定する情報処理装置であって，ニューラルネットワークを部分グラフに分割する部分グラフ分割部と，前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けた縮約コンフィグを出力する最適化部と，を備える情報処理装置である。

明　細　書

発明の名称 ：

情報処理装置およびニューラルネットワーク縮約方法

技術分野

[0001] 本発明は，ニューラルネットワークの縮約に関する。

背景技術

[0002] 近年，車載ECU（Electronic Control Unit）にDNN（Deep Neural Network）を実装することで，高精度な物体認識や行動予測を用いたハイレベル自動運転の実現を目指す動きが加速している。

[0003] 図１に示すように，DNNは多数のニューロンから構成される演算レイヤを1層以上、多段に接続することで実現される。N層目の演算レイヤでは，N-1層目から出力された値を入力とし，この入力値を重み係数で重みづけすることで得られた結果をN+1層目の入力へ出力する。このとき，重み係数をアプリケーションに応じて適切な値に設定すること（学習）で，高い汎化性能を得ることができる。

[0004] 自動運転向けDNNの一例として，LiDAR（Light Detection and Ranging、Laser Imaging Detection and Ranging、「光検出と測距」ないし「レーザー画像検出と測距」）点群を用いて物体認識やセグメンテーションなどを実現する点群AI（Artificial Intelligence）がある。

[0005] 特許分文献１には、DNNに代表されるニューラルネットワークの演算量を削減（縮約）するための方法について開示がある。

先行技術文献

特許文献

[0006] 特許文献1：US 2018/046919 A1

発明の概要

発明が解決しようとする課題

[0007] 図２に点群AIの構成例を示す。LiDAR点群にVoxel（体積要素）を割り当て

，DNN等により疑似画像を生成する。疑似画像中の所定の物体の検出をCNN（Convolutional Neural Network）で行う。

[0008] 点群AIを自動運転に適用するためには，高解像度なLiDARを用いることで点群AIを高性能化する必要がある。よって自動運転向け点群AIの実装には，車載プロセッサの演算性能と比較して膨大な演算量が必要である。例えば，点群AIの一種であるPointPIllarsを用いて自動運転における物体認識を実現するためには，65GOPs（Giga Operations）以上の演算量が必要である。一方で，車載ECUへ搭載可能なプロセッサの演算性能は数10TOPS（Tera Operation per Second）程度であり，PointPillarsを含む自動運転システム全体をリアルタイム実行するためには不十分である。

[0009] そこで，自動運転向け点群AIの演算量削減と高精度な推論精度維持を両立できる演算量削減（縮約）アルゴリズムが必要である。

[0010] 特許文献1記載の技術では、DNNに代表されるニューラルネットワークの演算量を削減するため、Multi-iteration compression methodによりDNNの縮約を実現する。

[0011] 図３は，発明者らが創作した図であり，DNN縮約フローにおいて，任意の単一の縮約アルゴリズムを縮約対象DNN全体に一律適用する例を示している。

[0012] 学習データセット301と縮約前DNNモデル302に対して，例えば予め設定してある縮約条件から選択する等して初期縮約条件を決定する（S303）。縮約前DNNモデル302は，学習データセット301で学習済みとする。

[0013] 次に，初期縮約条件に基づいたDNN中の縮約箇所および縮約率に基づいて，縮約前DNNモデル302を縮約する（S304）。

[0014] 次に，縮約後のDNNモデルを学習データセット301で再学習し（S305），縮約前後のDNNモデルの出力の誤差を評価するなどして，縮約を終了するか判定する（S306）。縮約が終了した場合には，縮約後のDNNモデルを縮約DNNモデル307として記録する。

[0015] 縮約を終了しない場合には，縮約箇所や縮約率の最適条件を探索し（S308），再度縮約を行う（S304）。

[0016] しかし、点群AIなどの複雑な構造を有するDNNでは，DNNを構成する部位（部分グラフ）ごとに最適な縮約アルゴリズムが異なる。そこで，DNNの演算量削減と推論精度維持を両立する高精度縮約を実現するためには，部分グラフ単位で最適な縮約アルゴリズムを選定し適用する必要がある。

[0017] よって，上記の構成では，ニューラルネットワークの部分グラフ単位で，適用する縮約アルゴリズムの最適化を実現できないという課題があった。

課題を解決するための手段

[0018] 本発明の好ましい一側面は，ニューラルネットワークを縮約するアルゴリズムを選定する情報処理装置であって，ニューラルネットワークを部分グラフに分割する部分グラフ分割部と，前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けた縮約コンフィグを出力する最適化部と，を備える情報処理装置である。

[0019] 本発明の好ましい他の一側面は，ニューラルネットワークを部分グラフに分割する第１のステップ，前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けて仮縮約する第２のステップ，を実行するニューラルネットワーク縮約方法である。

発明の効果

[0020] ニューラルネットワークの部分グラフ単位で，適用する縮約アルゴリズムの最適化を実現できる。

図面の簡単な説明

[0021] [図1]Deep Neural Network(DNN)の構造を示す概念図。

[図2]点群AIの構成例を示す概念図。

[図3]DNN縮約フロー図（比較例）。

[図4]DNN縮約フロー図（実施例）。

[図5]縮約アルゴリズム最適化部の処理模式図。

[図6]実施例の効果を示すグラフ図。

[図7]実施例１/基本構成1のブロック図。

[図8A]実施例２/基本構成2（全体構成）のブロック図。

[図8B]縮約コンフィグの例を示す表図。

[図9]実施例２/基本構成2（縮約アルゴリズム最適化部）処理フロー図。

[図10]実施例２/基本構成2（縮約アルゴリズム最適化部）構成図。

[図11A]実施例２/基本構成2（部分グラフ分割部）処理フロー図。

[図11B]部分グラフ分割の概念図。

[図12]実施例２/基本構成2（摂動計算部1）構成図。

[図13]実施例２/基本構成2（摂動計算部2）構成図。

[図14]実施例２/基本構成2（縮約部）構成図。

[図15]実施例３/応用構成1の全体構成図。

[図16]実施例４/応用構成2の全体構成図。

[図17]実施例５/応用構成3の全体構成図。

発明を実施するための形態

[0022] 以下に実施例を図面を用いて説明する。ただし、本発明は以下に示す実施の形態の記載内容に限定して解釈されるものではない。本発明の思想ないし趣旨から逸脱しない範囲で、その具体的構成を変更し得ることは当業者であれば容易に理解される。

[0023] 以下に説明する実施例の構成において、同一部分又は同様な機能を有する部分には同一の符号を異なる図面間で共通して用い、重複する説明は省略することがある。

[0024] 同一あるいは同様な機能を有する要素が複数ある場合には、同一の符号に異なる添字を付して説明する場合がある。ただし、複数の要素を区別する必要がない場合には、添字を省略して説明する場合がある。

[0025] 本明細書等における「第１」、「第２」、「第３」などの表記は、構成要素を識別するために付するものであり、必ずしも、数、順序、もしくはその内容を限定するものではない。また、構成要素の識別のための番号は文脈毎に用いられ、一つの文脈で用いた番号が、他の文脈で必ずしも同一の構成を示すとは限らない。また、ある番号で識別された構成要素が、他の番号で識別された構成要素の機能を兼ねることを妨げるものではない。

[0026] 図面等において示す各構成の位置、大きさ、形状、範囲などは、発明の理解を容易にするため、実際の位置、大きさ、形状、範囲などを表していない場合がある。このため、本発明は、必ずしも、図面等に開示された位置、大きさ、形状、範囲などに限定されない。

[0027] 本明細書で引用した刊行物、特許および特許出願は、そのまま本明細書の説明の一部を構成する。

[0028] 本明細書において単数形で表される構成要素は、特段文脈で明らかに示されない限り、複数形を含むものとする。

[0029] 図４に実施例の一例のフローを示す。図３の縮約フローに対して縮約アルゴリズム最適化の処理（S409）を備えている。

[0030] 縮約アルゴリズム最適化（S409）では，まず縮約前DNNモデル402を複数の部分グラフに分割し，さらに部分グラフ単位で最適な縮約アルゴリズムを高効率に探索する。

[0031] 学習データセット401と縮約前DNNモデル402に対して，例えば予め設定してある縮約条件から選択する等して，部分グラフごとに初期縮約条件を決定する（S403）。縮約前DNNモデル402は，学習データセット401で学習済みとする。

[0032] 次に，初期縮約条件に基づいたDNN中の縮約箇所および縮約率に基づいて，縮約前DNNモデル402を部分グラフ毎に縮約する（S404）。

[0033] 次に，縮約後のDNNモデルを学習データセット401で再学習し（S405），縮約前後のDNNモデルの出力の誤差を評価するなどして，縮約を終了するか判定する（S406）。縮約が終了した場合には，縮約後のDNNモデルを縮約DNNモデル407として記録する。

[0034] ここで，縮約により検出精度が低下するため，検出精度を回復させるために再学習を実行している。再学習を実行しない場合は，縮約AIの検出率が大幅に低下する可能性がある。再学習には，基本的には縮約前DNNモデルの学習に用いたものと同じデータセットを利用してよいが，適用シーンなどに応じてデータセットは柔軟に変更することが可能である。

[0035] 縮約を終了しない場合には，縮約箇所や縮約率の最適条件を探索し（S408），再度縮約を行う（S404）。

[0036] 図５に縮約アルゴリズム最適化(S409)の処理を模式的に示す。まず入力された最大部分グラフ数および縮約前DNNモデル402をもとに，部分グラフ分割箇所の初期値を決定する。最大部分グラフ数は，予め設定あるいは，ユーザがその都度指定してよい。分割箇所は予め定めたルールにより、例えばニューロン数ができるだけ均等になるよう分割する。

[0037] 次に，分割された部分グラフごとに縮約アルゴリズムを適用し，縮約適用前後における各部分グラフの計算結果の誤差絶対値を摂動の大きさとして出力する。縮約アルゴリズムは，例えば予め準備された複数の中から選択する。部分グラフごとに異なる縮約アルゴリズムを選択可能である。

[0038] 摂動計算結果はマージして総合的に評価する。次に，任意の最適化アルゴリズムを用いて，評価した摂動の大きさが小さくなるように各部分グラフへ適用する縮約アルゴリズムの組み合わせを変更する。この時，縮約アルゴリズムの組み合わせを任意回数変更したにもかかわらず，摂動の大きさが一定閾値未満に収束しない場合，縮約前DNNモデル402の部分グラフ分割箇所を変更し，再度縮約アルゴリズの組み合わせ最適化を行う。

[0039] 図６に物体認識向け点群AIであるPointPillarsに本実施例を適用した場合の効果を示す。縮約前点群AI６０１のDNNモデル全体に，単一の縮約手法1b（Voxel数削減）を適用した場合，NG1のように検出率が大幅に低下した。また，DNNモデル全体に，単一の縮約手法2b（Low rank近似）を適用した場合，NG2のように処理時間は短縮されるものの検出率が低下した。

[0040] 本実施例では，縮約前点群AI６０１のDNNモデルのうち部分グラフ１に縮約手法1a（Voxelあたり点群数削減）を適用し，部分グラフ２に縮約手法2a（プルーニング）を適用した。この結果，最適縮約点群AI６０２が得られた。

[0041] なお、矢印６０３は，部分グラフ1に対する縮約手法1aで得られる検出率および処理時間の限界値を示しており，矢印６０４は，部分グラフ２に対する縮約手法2aで得られる検出率および処理時間の限界値を示している。斜点線

６０５はこれらの境界を示す。このように，部分グラフごとに適した縮約手法があることがわかる。

[0042] 本実施例の，最適縮約点群AI６０２は，PointPillarsの縮約に起因した検出率低下を抑制しつつ，推論時間を50%削減することが可能である。なお本実施例は，点群AIのみならず画像処理CNN等他のAIにも適用可能である。

実施例 1

[0043] 図７に実施例１の構成図を示す。まずメモリより読み出された縮約前DNN（A001）を入力として受け付け，部分グラフ分割部（A004）において縮約前DNNをn個の部分グラフへ分割する。次にn個の部分グラフおよびメモリから読み出された縮約コンフィグ（A002）を入力として受け付け，縮約部（A005）において部分グラフごとに縮約をそれぞれ適用する。縮約が適用された部分グラフは再結合され，縮約DNN（A003）としてメモリに格納される。最後にメモリから読み出された縮約DNN（A003）およびLiDAR点群を入力として受け付け，推論部（A006）においてDNNの推論を実行し，推論結果を出力する。

実施例 2

[0044] 図８Ａに実施例２の構成図を示す。まずメモリから読み出された縮約前DNN（B001）およびデータセット（B004）ならびに適用可能な縮約アルゴリズムが記載された縮約テーブル（B002）を縮約アルゴリズム最適化部（B006）に入力し，縮約前DNNに適用する縮約アルゴリズムを部分グラフ単位で最適化する。ここで各部分グラフと最適縮約アルゴリズムの対応関係は，縮約コンフィグ（B003）に書き出されメモリで保持される。なお，メモリに保持されたデータセット（B004）はセンサから取得されたLiDAR点群データにより適宜更新されるものとする。

[0045] 次に，縮約コンフィグ（B003）およびデータセット（B004）ならびにn個の部分グラフを縮約部（B007）に入力することで，縮約DNN（B005）を生成しメモリへ格納する。最後に，縮約DNN（B005）およびLiDAR点群を推論部（B008）に入力し，推論結果を得る。本実施例により，走行シーンの変化などに応じたDNN縮約手法の動的な調整が可能である。

[0046] 図８Ａのシステム構成は，サーバのような一般的な情報処理装置で構成可能である。情報記憶装置は，処理装置，記憶装置，入力装置，出力装置を備える。本実施例では計算や制御等の機能は，記憶装置に格納されたプログラムが処理装置によって実行されることで、定められた処理を他のハードウェアと協働して実現される。計算機などが実行するプログラム、その機能、あるいはその機能を実現する手段を、「機能」、「手段」、「部」、「ユニット」、「モジュール」等と呼ぶ場合がある。

[0047] 図８Ａにおいて，説明のため上記サーバの当然持つ構成は省略し，データや機能ブロックの形で表示している。LiDAR点群データは，例えば入力装置から入力される。縮約前DNN（B001），データセット（B004），縮約テーブル（B002）は，例えば記憶装置（メモリ）の一部であるハードディスク装置等にデータベースとして格納されており、同じく記憶装置の一部である半導体メモリなどに適宜呼び出されて使用される。縮約アルゴリズム最適化部（B006），縮約部（B007）は，記憶装置にプログラムとして実装されており，処理装置によって実行され所定の機能を実現する。縮約コンフィグ（B003），縮約DNN（B005）には，処理結果が格納される。

[0048] 推論部（B008）は，AIの処理を実行するためのデバイスであり，ニューラルネットワークを実装可能なモジュールとして市販されている。モジュールとして，例えばVersal（商標），Xavier（商標），NCS2（商品名）などが知られている。

[0049] 以上のサーバの構成は、単体の装置で構成してもよいし、あるいは、入力装置、出力装置、処理装置、記憶装置の任意の部分が、ネットワークで接続された他のコンピュータで構成されてもよい。

[0050] 本実施例中、ソフトウエアで構成した機能と同等の機能は、FPGA（Field Programmable Gate Array）、ASIC（Application Specific Integrated Circuit）などのハードウェアでも実現できる。

[0051] 図８Ｂは，縮約テーブル（B002）の例を示す表図である。推論部（B008）に適用可能なデバイスとして登録されている，デバイス＃１，＃２，＃３，

＃４に対して，各縮約手法との親和性が記録されている。親和性の値は１～０の範囲であり，１に近いものの親和性が高い。これらの情報は，公知情報あるいはユーザの知見により，予めデータベースとして登録しておく。

[0052] 以下で図８Ａの縮約アルゴリズム最適化部（B006）の構成および処理フローについて詳述する。図９に処理フローを，図１0に装置構成を示す。

[0053] 図９のフローを順次説明する。

S000：縮約前DNN，摂動計算部2（C006）から出力された部分グラフ摂動，部分グラフへの分割数上限である最大部分グラフ数，および最適化部（C007）から出力されたグラフ分割信号を部分グラフ分割部（C003）に入力し，縮約前DNNをn個の部分グラフに分割する。最大部分グラフ数は，予め設定するか，ユーザがその都度入力する。

[0054] S001：部分グラフ，縮約テーブル（B002）および最適化部（C007）より出力された部分グラフごとの縮約アルゴリズムを仮縮約部（C004）に入力し，各部分グラフに対応する縮約アルゴリズムを適用することで，仮縮約部分グラフを生成する。図８Ｂに示したように，縮約テーブル（B002）には，公知の縮約アルゴリズム、例えばプーリング，Row lank，Weight Sharing，量子化などが登録されているものとする。縮約アルゴリズムの選択は，例えば縮約テーブル（B002）において，実装用デバイスと親和性が高いものを優先的に選択する。

[0055] S002：縮約前DNN，データセット，仮縮約部分グラフを摂動計算部1（C005）および摂動計算部2（C006）に入力し，縮約前後における計算結果の摂動をそれぞれNN摂動および部分グラフ摂動として生成する。ここで部分グラフ摂動はログファイル(C008)としてメモリ上に記録される。

[0056] S003：NN摂動を最適化部（C007）に入力することで，NN摂動の値が小さくなるように縮約アルゴリズムの組み合わせを補正する。組み合わせの探索は例えば、縮約テーブルB002にて，親和性が高いアルゴリズムの組み合わせを網羅的に試行し，縮約アルゴリズム最適化のループ処理を行う。

[0057] S004：最適化部（C007）の試行において，所定のｋ回連続でNN摂動が任意

閾値以上の場合，グラフ分割信号をenableとし，S000の処理に戻る。それ以外の場合は，S005に進む。

[0058] S005：所定の終了条件により最適化の終了を判定する。終了条件としては，NN摂動が所定の条件を満たす，処理の制限時間あるいは制限回数を超える，などがある。縮約アルゴリズムの最適化を継続する場合はS001に戻る。それ以外は処理を完了する。

[0059] 次に，部分グラフ分割部（C003）について述べる。

図１１Ａに部分グラフ分割部の処理フローを示す。図９と合わせて説明する。

図１１Ｂに分割される部分グラフの概念を示す。

[0060] 部分グラフ分割部（C003）は，通常は待ち状態になっており，グラフ分割信号がenableになったとき起動する（S004）。AIの縮約処理開始直後は，部分グラフ摂動やグラフ分割信号が適用されないため，AIは分割されておらず（すなわち部分グラフ数ｎ＝１），AI全体に同じ縮約手法を適用する（S001）。適用する縮約手法は仮縮約部（C004）が，縮約テーブル（B002）を参照して最適なものを選択している。摂動計算部１（C005）は，AI全体に対して，縮約に起因した推論結果の摂動値を計算する（S002）。この時点では部分グラフ＝縮約前DNNなので，摂動計算部２（C006）は，摂動計算部１（C005）と同じ摂動値を出力する。縮約アルゴリズムの組み合わせを補正し（S003），終了条件を満足する場合には，AIを分割する必要がないので縮約処理は終了する（S005）。縮約アルゴリズムの組み合わせを補正しても，NN摂動が所定回連続で閾値以上となる場合には，グラフ分割信号をenableにしてS012（S000）へ進む（S004）。

[0061] S012：縮約前DNN（図１１Ｂのグラフ1101）をｍ分木の根に保存し，ｍ分木の根を分割対象ノードに設定する。ｍの値は予め設定あるいはユーザが都度入力すればよい。この例ではｍ＝２の固定値として説明する。図１１Ｂの縮約前DNNはこの例では８層とする。

[0062] S013：分割対象ノードに格納されたグラフをｍ分割し，分割対象ノードの

配下にあるｍ個の子ノードに保存する。図１１Ｂの例では、グラフ1101がｍ＝２分割され，グラフ1102とグラフ1103が子ノードとなる。分割の方法は、層単位でなるべく均等になるように分割しており，この例では，各４層のグラフが部分グラフとなる。この状態では，ｍ分木の葉ノードｍ個（深さ1の葉ノード:ｍ個）にそれぞれ分割後の部分グラフ1102と1103が格納される。

[0063] S014：仮縮約部（C004）は，分割後の部分グラフごとに縮約手法を適用し（S001），摂動計算部１（C005）がAI全体に対して，縮約に起因した推論結果の摂動値を計算する（S002）。縮約アルゴリズムを補正して（S003）も尚，縮約によるNN摂動が閾値より大きい場合は、グラフ分割信号をenableにしてS015へ進む（S004）。NN摂動が閾値より小さい場合は、グラフ分割の効果が得られたとして処理を終了する（S005）。

[0064] S015：ｍ個の子ノードに保存された部分グラフそれぞれに対して，摂動計算部２（C006）が部分グラフ摂動の値を取得する（S002）。図１１Ｂの例では部分グラフ1102と部分グラフ1103の摂動が計算される。

[0065] S016：ｍ個の子ノードのうち，部分グラフ摂動の絶対値が最大の子ノードを新しい分割対象ノードとして設定する。図１１Ｂの例では部分グラフ1102の摂動が大きいとし，部分グラフ1102が分割対象ノードとなる。

[0066] S017：生成されたｍ分木の葉の数が最大部分グラフ数以下であれば，S013に戻る。それ以外は，処理を終了する。図１１Ｂの例で，部分グラフ1102と部分グラフ1103まで分割されている場合は，これらが葉になるので，葉の数２が最大部分グラフ数を超えていれば分割は終了し、そうでなければS013に戻る。

[0067] S013：分割対象ノードのグラフ（この例では部分グラフ1102）が分割対象となり，ｍ＝２分割される。分割した部分グラフ1104と部分グラフ1105が，分割対象ノードのｍ個の子にそれぞれ格納される。この状態では、ｍ分木の葉ノード２ｍ－１個（深さ1の葉ノード:ｍ-１個＋深さ２の葉ノードｍ個）に格納された部分グラフがそれぞれ分割後の部分グラフとなる。この例では，深さ1の葉ノードは部分グラフ1103であり，深さ２の葉ノードは部分グラフ11

04と1105である。これら葉ノードの数（図１１Ｂの例では３）が部分グラフの数ｎに等しい。

[0068] S014：仮縮約部（C004）は，２ｍ－１個の部分グラフごとに縮約手法を適用し，摂動計算部１（C005）がAI全体に対して，縮約に起因した推論結果の摂動値を計算する。縮約アルゴリズムを補正し（S003），縮約によるNN摂動の値が閾値より大きい場合，部分グラフ分割信号をenableにしてさらに分割を繰り返す（S004）。次の分割対象は，部分グラフ1103と1104と1105の中で，最も部分グラフ摂動が大きいグラフとなる。

[0069] ループの終了条件は，部分グラフの分割数ｎ（ｍ分木の葉ノードの数）が最大値を超えた場合（S017）である。

[0070] 次に，摂動計算部1（C005）について述べる。摂動計算部1（C005）は，ネットワークモデル全体のNN摂動を計算する。NN摂動は，縮約手法の最適組み合わせ探索のために用いる。

[0071] 図１２に構成図を示す。まず0番目の仮縮約部分グラフおよびそれに対応する縮約前DNNの部分グラフを，順伝搬部B0および順伝搬部A0にそれぞれ入力し，同一のデータセットを順伝搬させる。同様にk番目の仮縮約部分グラフおよびそれに対応する縮約前DNNの部分グラフを，順伝搬部Bkおよび順伝搬部Akにそれぞれ入力し，ｋ－1番目の順伝搬部Bk－1の演算結果および順伝搬部Ak－1の演算結果をそれぞれ順伝搬させる。以上の操作をn番目の部分グラフまで繰り返し，順伝搬部AnおよびBnの演算結果の差の絶対値を摂動計算部により算出し，NN摂動を生成する。

[0072] 次に，摂動計算部2（C006）について述べる。摂動計算部2（C006）は，部分グラフごとの部分グラフ摂動を計算する。部分グラフ摂動は，ネットワークの部分グラフ分割の要否判定に用いる。

[0073] 図１３に構成図を示す。まず0番目の仮縮約部分グラフおよびそれに対応する縮約前DNNの部分グラフを，順伝搬部B0および順伝搬部A0にそれぞれ入力し，同一のデータセットを順伝搬させる。同様にk番目の仮縮約部分グラフおよびそれに対応する縮約前DNNの部分グラフを，順伝搬部Bkおよび順伝搬部Akに

それぞれ入力し，k－1番目の順伝搬部Ak－1の演算結果を双方に順伝搬させる。以上の操作をn番目の部分グラフまで繰り返す。このとき，順伝搬部AkおよびBkの演算結果の差の絶対値に正規化を適用した値を，部分グラフ摂動kとして生成する。正規化は，例えば（摂動p_i／縮約前NNの出力y_i）の演算で行うことができる。　最後に，縮約部（B007）について述べる。

[0074] 図１４に構成図を示す。まずn個の部分グラフおよび各部分グラフに適用する縮約アルゴリズムを記載した縮約コンフィグを入力として受け付け，演算量削減部（D001）において縮約を実行する。次に，各部分グラフに対する縮約結果を部分グラフ結合部（D002）において，一つのDNNに再結合する。再結合されたDNNおよびデータセットを用いて再学習部（D003）において再学習を実行し，縮約NNモデルを生成する。

実施例 3

[0075] 図１５に実施例３の構成を示す。実施例３は，実施例１に対し縮約コンフィグをデータセンタ側で生成するシステムを追加した構成である。具体的には，データセンタ側で縮約前DNN（E012）およびデータセット（E010）ならびに縮約テーブル（E013）を用いて縮約アルゴリズム最適化部（E009）により縮約コンフィグ（E011）を生成し，OTA送信機（E008）およびOTA受信機（E007）を介してAD-ECU側の縮約コンフィグ（E002）を更新する。

実施例 4

[0076] 図１6に実施例４の構成を示す。実施例４は，実施例２に対し縮約前DNNをOTA受信器（F009）により外部から受信可能とした構成である。

実施例 5

[0077] 図１７に実施例５の構成を示す。実施例５は，実施例４に対し縮約アルゴリズム最適化部で取得した摂動ログ（G010）を，メモリに記憶し，OTA送信器（G011）を通じて外部へ送信可能な構成である。

符号の説明

[0078] ・縮約前DNN・・・A001，B001，E001，E012，F001，G001

・縮約コンフィグ・・・A002, B003, C002, E002, E011, F003, G003

・縮約DNN・・・A003, B005, E003, F005, G005

・部分グラフ分割部・・・A004, C003, E004

・縮約部・・・A005, B007, E005, F007, G007

・推論部・・・A006, B008, E006, F008, G008

・縮約テーブル・・・B002, C001, E013, F002, G002

・データセット・・・B004, E010, F004, G004

・縮約アルゴリズム最適化部・・・B006, E009, F006, G006

・仮縮約部・・・C004

・摂動計算部1・・・C005

・摂動計算部2・・・C006

・最適化部・・・C007

・ログファイル・・・C008, G010

・演算量削減部・・・D001

・部分グラフ結合部・・・D002

・再学習部・・・D003

・OTA受信器・・・E007, F009, G009

・OTA送信器・・・E008, G011

請求の範囲

[請求項1] ニューラルネットワークを縮約するアルゴリズムを選定する情報処理装置であって，

ニューラルネットワークを部分グラフに分割する部分グラフ分割部と，

前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けた縮約コンフィグを出力する最適化部と，

を備える情報処理装置。

[請求項2] 前記部分グラフ分割部は，階層構造を有する前記ニューラルネットワークを，層単位で分割する，

請求項１記載の情報処理装置。

[請求項3] さらに仮縮約部と第１の摂動計算部を備え，

前記仮縮約部は，前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けて縮約して仮縮約部分グラフを出力し，

前記第１の摂動計算部は，前記ニューラルネットワークの順伝搬値および複数の前記仮縮約部分グラフを直列結合した順伝搬値の差に基づく，ニューラルネットワークの摂動を出力し，

前記最適化部は，前記ニューラルネットワークの摂動が，所定値を満たす縮約コンフィグを出力する，

請求項２記載の情報処理装置。

[請求項4] 前記仮縮約部は，前記ニューラルネットワークが実装されるデバイスの種類と適用すべき縮約手法の優先度が対応付けられた縮約テーブルを参照して，複数の縮約手法から一つを選択する，

請求項３記載の情報処理装置。

[請求項5] さらに第２の摂動計算部を備え，

前記第２の摂動計算部は，それぞれの前記部分グラフの順伝搬値および前記仮縮約部分グラフの順伝搬値の差に基づく，それぞれの部分

グラフの摂動を出力し，

前記部分グラフ分割部は，前記部分グラフの摂動に基づいて，さらに分割する部分グラフを選定する，

請求項３記載の情報処理装置。

[請求項6] 前記部分グラフ分割部は，現状において摂動が最大の部分グラフを，さらに分割する部分グラフとして選定する，

請求項５記載の情報処理装置。

[請求項7] さらに縮約部を備え，

前記縮約部は，演算量削減部と部分グラフ結合部を備え，

前記演算量削減部は，前記部分グラフと前記縮約コンフィグを入力とし，前記部分グラフのそれぞれに対して，前記縮約コンフィグに記載された縮約手法を適用して縮約後部分グラフを出力し，

前記部分グラフ結合部は，前記演算量削減部の出力を入力とし，前記縮約後部分グラフを結合して，縮約後ニューラルネットワークを出力する，

請求項１記載の情報処理装置。

[請求項8] さらに再学習部を備え、

前記再学習部は，前記縮約後ニューラルネットワークと学習用データセットを受付け，前記縮約後ニューラルネットワークを学習する，

請求項７記載の情報処理装置。

[請求項9] 前記部分グラフの摂動を摂動ログとしてメモリに保持することを特徴とする、

請求項５記載の情報処理装置。

[請求項10] 前記摂動ログを演算装置外に送信することを特徴とする、

請求項９記載の情報処理装置。

[請求項11] ニューラルネットワークを部分グラフに分割する第１のステップ，

前記部分グラフのそれぞれに対して，複数の縮約手法から選択された一つを対応付けて仮縮約する第２のステップ，

を実行するニューラルネットワーク縮約方法。

[請求項12] 仮縮約したニューラルネットワークの摂動を計算する第３のステップ，

前記ニューラルネットワークの摂動に基づいて，前記部分グラフに対する縮約手法の対応を変更する第４のステップ，

を実行する請求項１１記載のニューラルネットワーク縮約方法。

[請求項13] 前記部分グラフのそれぞれに対して，仮縮約した部分グラフ単位で部分グラフの摂動を計算する第５のステップ，

前記部分グラフの摂動に基づいて、さらに分割する部分グラフを選択する第６のステップ，

を実行する請求項１１記載のニューラルネットワーク縮約方法。

[請求項14] 前記第１のステップにおいて，前記部分グラフへの分割方法をｍ分木によって管理し，

前記ニューラルネットワークをｍ分木の根に保管し，前記ｍ分木の根に保管された前記ニューラルネットワークをｍ分割し，前記根の子である深さ１のノードにｍ個の部分グラフを保管し，前記ｍ分木の葉に保管された部分グラフを出力する，

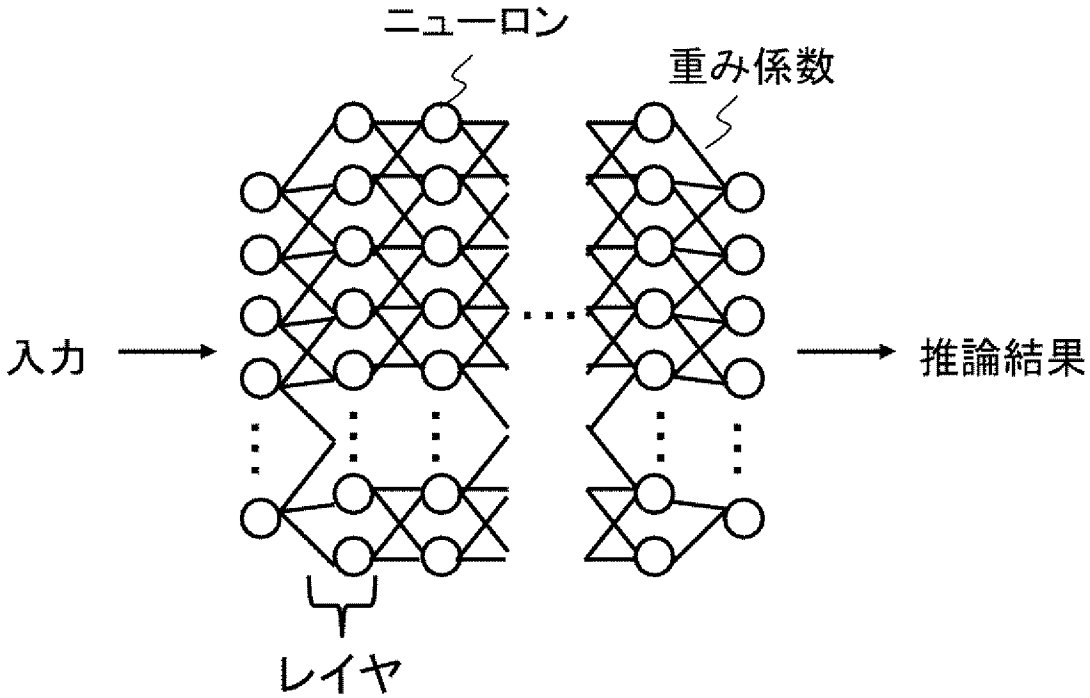
請求項１２記載のニューラルネットワーク縮約方法。

[請求項15] 前記根の子である深さ１のノードに保管されたｍ個の部分グラフのうち，部分グラフの摂動が最大のノードに格納された部分グラフをｍ分割し，前記部分グラフの摂動が最大のノードの子に保管する，

請求項１４記載のニューラルネットワーク縮約方法。

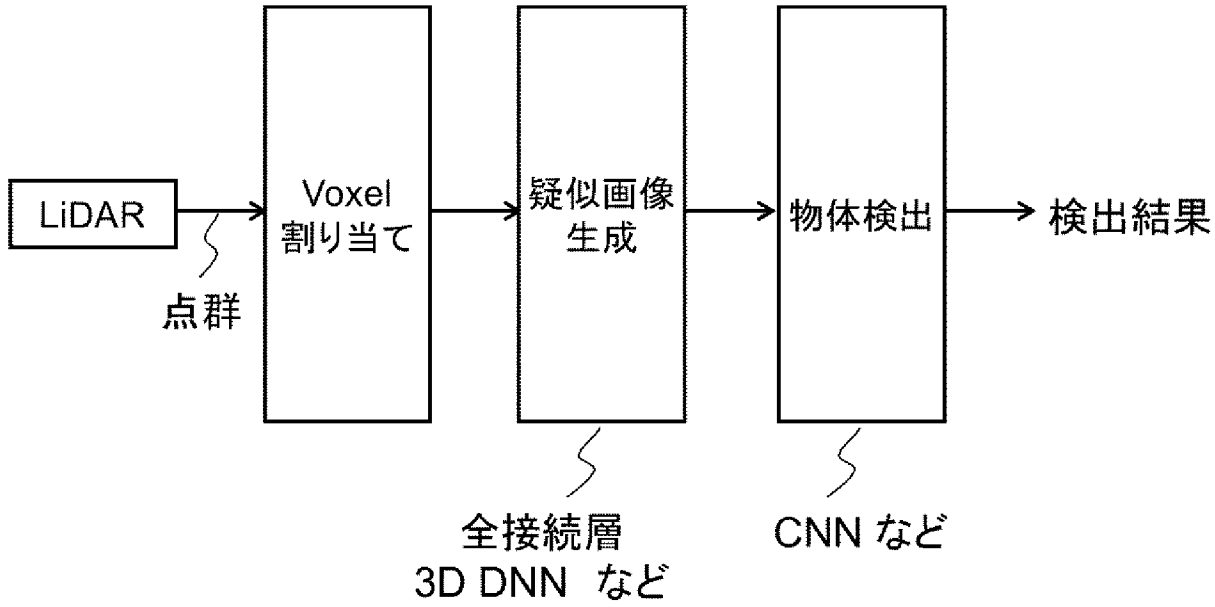
[図1]

図1



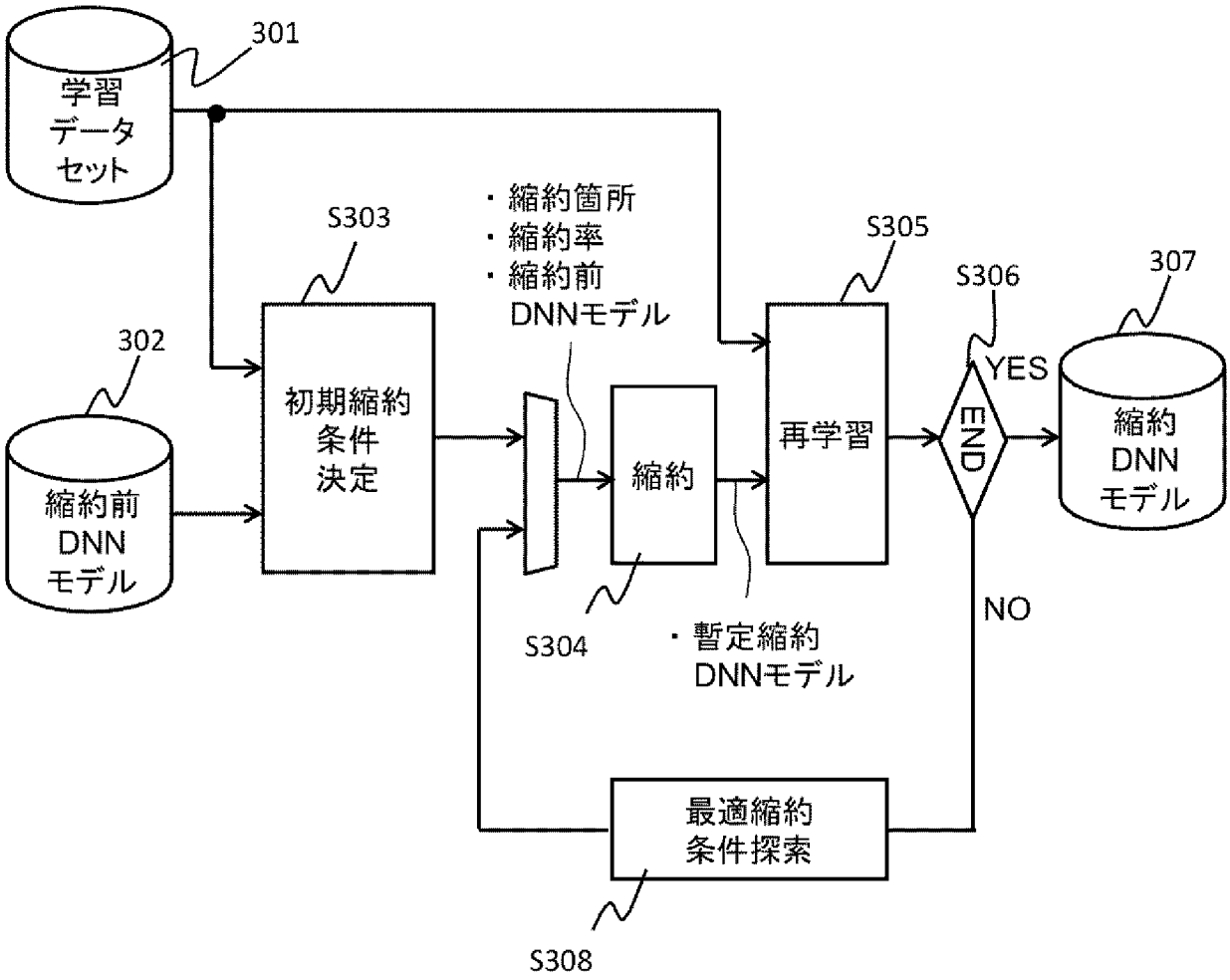
[図2]

図2



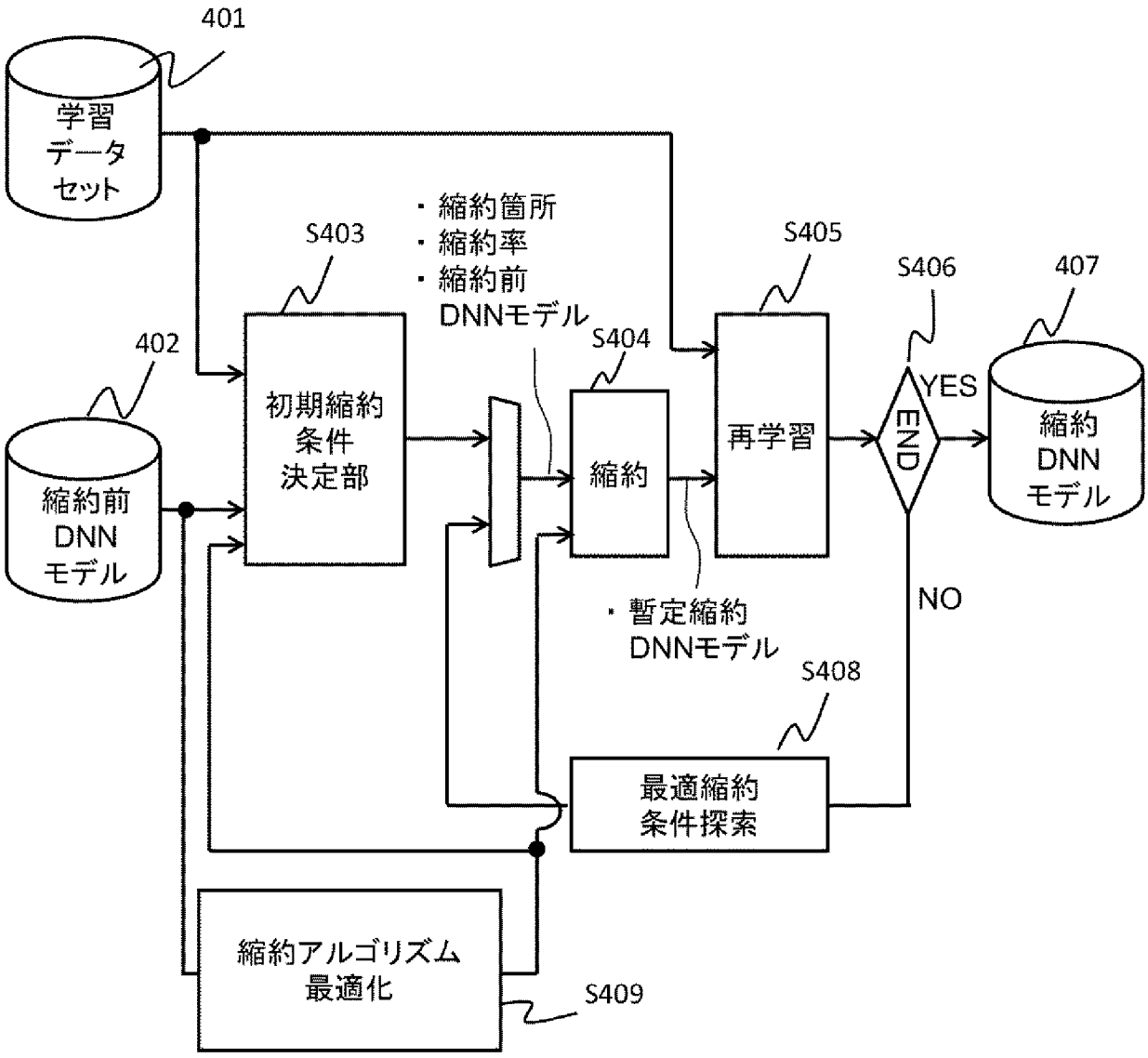
[図3]

図3


学習
データ
セット
301
302
縮約前
DNN
モデル
S303
初期縮約
条件
決定
・縮約箇所
・縮約率
・縮約前
DNNモデル
縮約
S304
・暫定縮約
DNNモデル
再学習
S305
S306
END
YES
NO
307
縮約
DNN
モデル
最適縮約
条件探索
S308

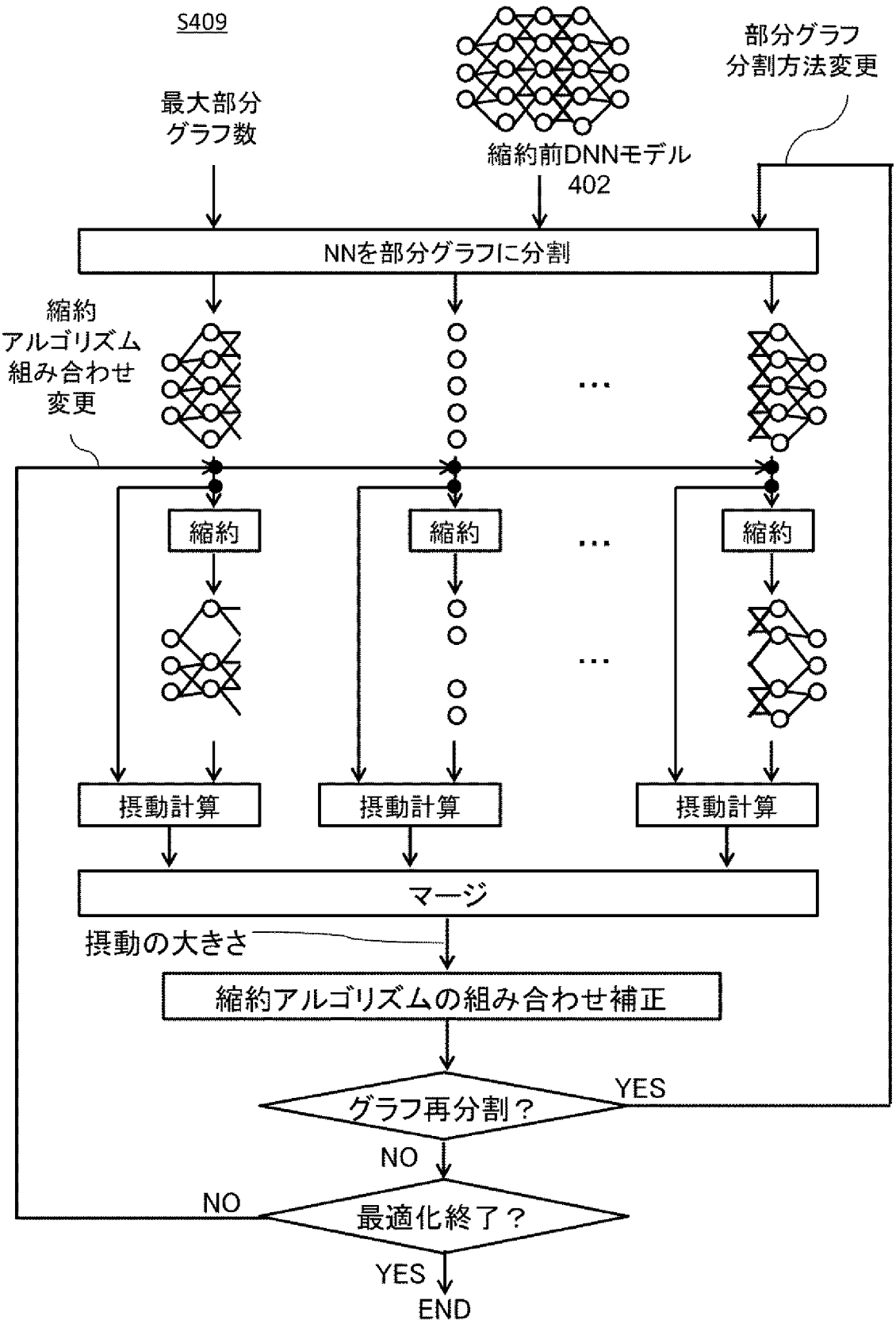
[図4]

図4


401
学習
データ
セット
・縮約箇所
・縮約率
・縮約前
DNNモデル
S403
S405
S406
407
402
S404
初期縮約
条件
決定部
縮約
再学習
END
YES
縮約
DNN
モデル
縮約前
DNN
モデル
・暫定縮約
DNNモデル
NO
S408
最適縮約
条件探索
縮約アルゴリズム
最適化
S409

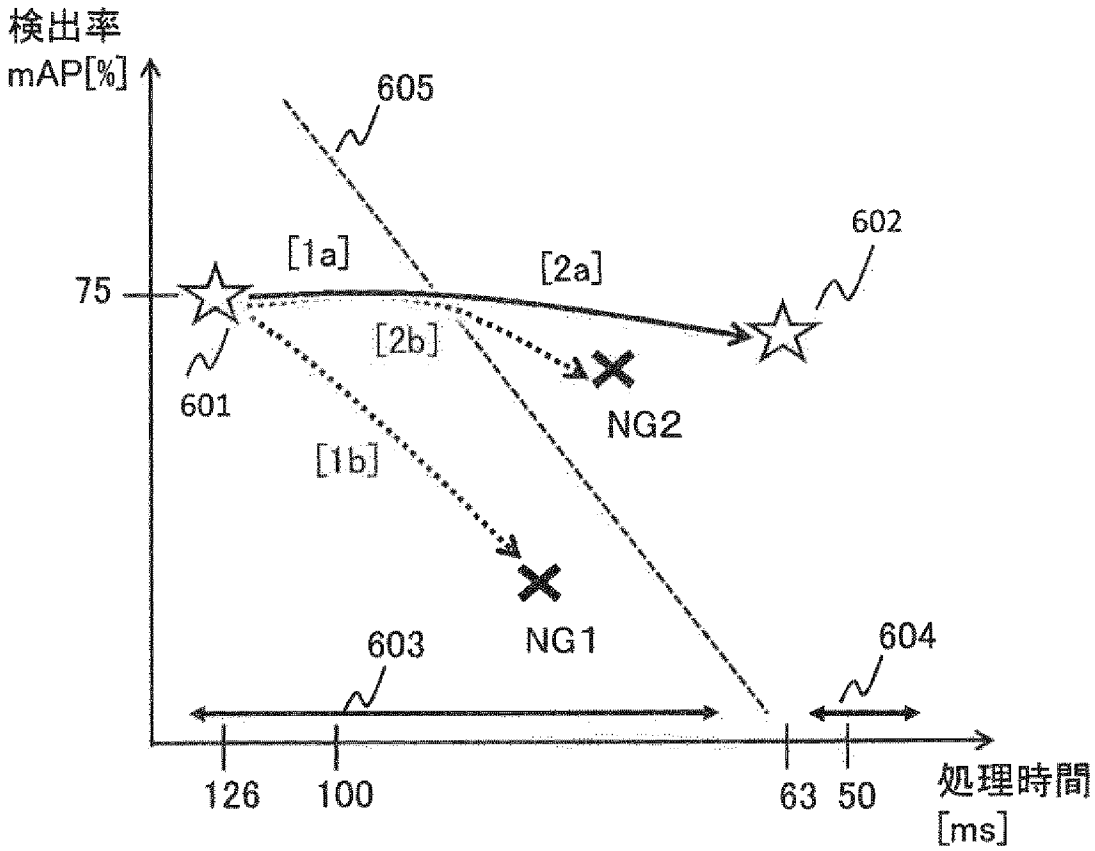
[図5]

図5


S409
最大部分
グラフ数
縮約前DNNモデル
402
部分グラフ
分割方法変更
NNを部分グラフに分割
縮約
アルゴリズム
組み合わせ
変更
縮約
縮約
縮約
摂動計算
摂動計算
摂動計算
マージ
摂動の大きさ
縮約アルゴリズムの組み合わせ補正
グラフ再分割？
YES
NO
NO
最適化終了？
YES
END

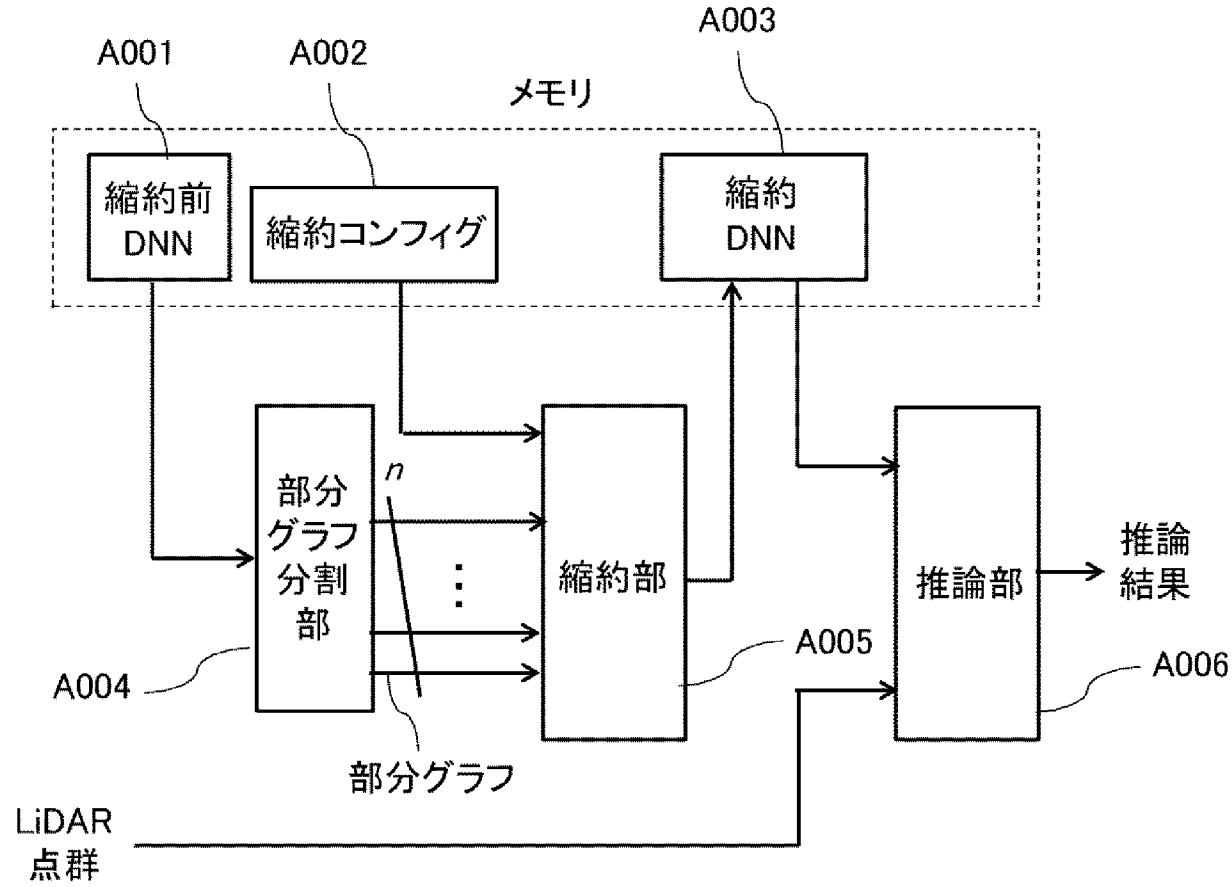
[図6]

図6


検出率
mAP[%]
605
602
[1a]
[2a]
75
[2b]
601
NG2
[1b]
NG1
603
604
126
100
63
50
処理時間
[ms]

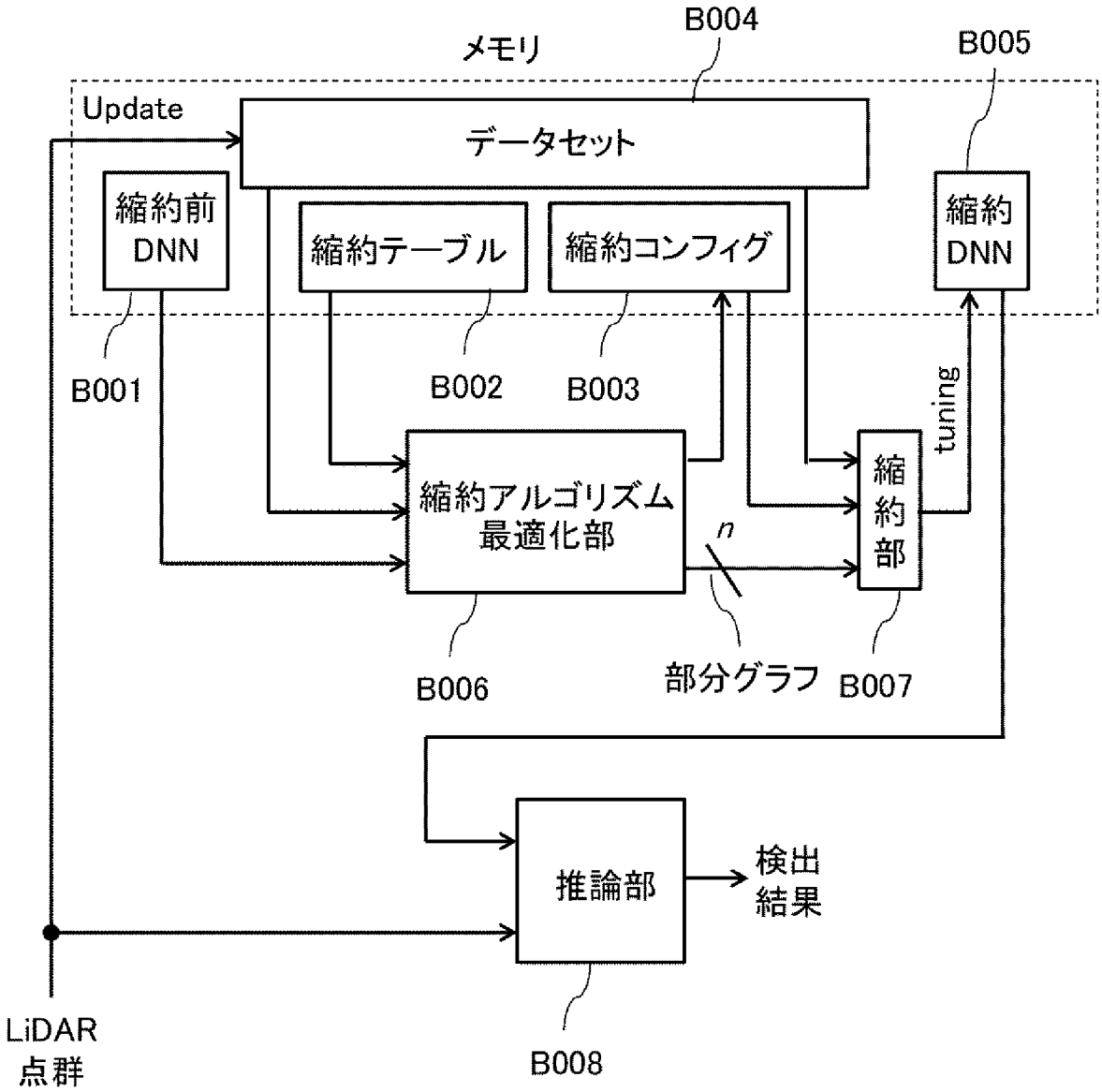
[図7]

図7


A001
A002
A003
メモリ
縮約前
DNN
縮約コンフィグ
縮約
DNN
部分
グラフ
分割
部
n
縮約部
推論部
推論
結果
A004
A005
A006
部分グラフ
LiDAR
点群

[図8A]

図8A


メモリ
B004
B005
Update
データセット
縮約前
DNN
縮約テーブル
縮約コンフィグ
縮約
DNN
B001
B002
B003
tuning
縮約アルゴリズム
最適化部
n
縮約
部
B006
部分グラフ
B007
推論部
検出
結果
LiDAR
点群
B008

[図8B]

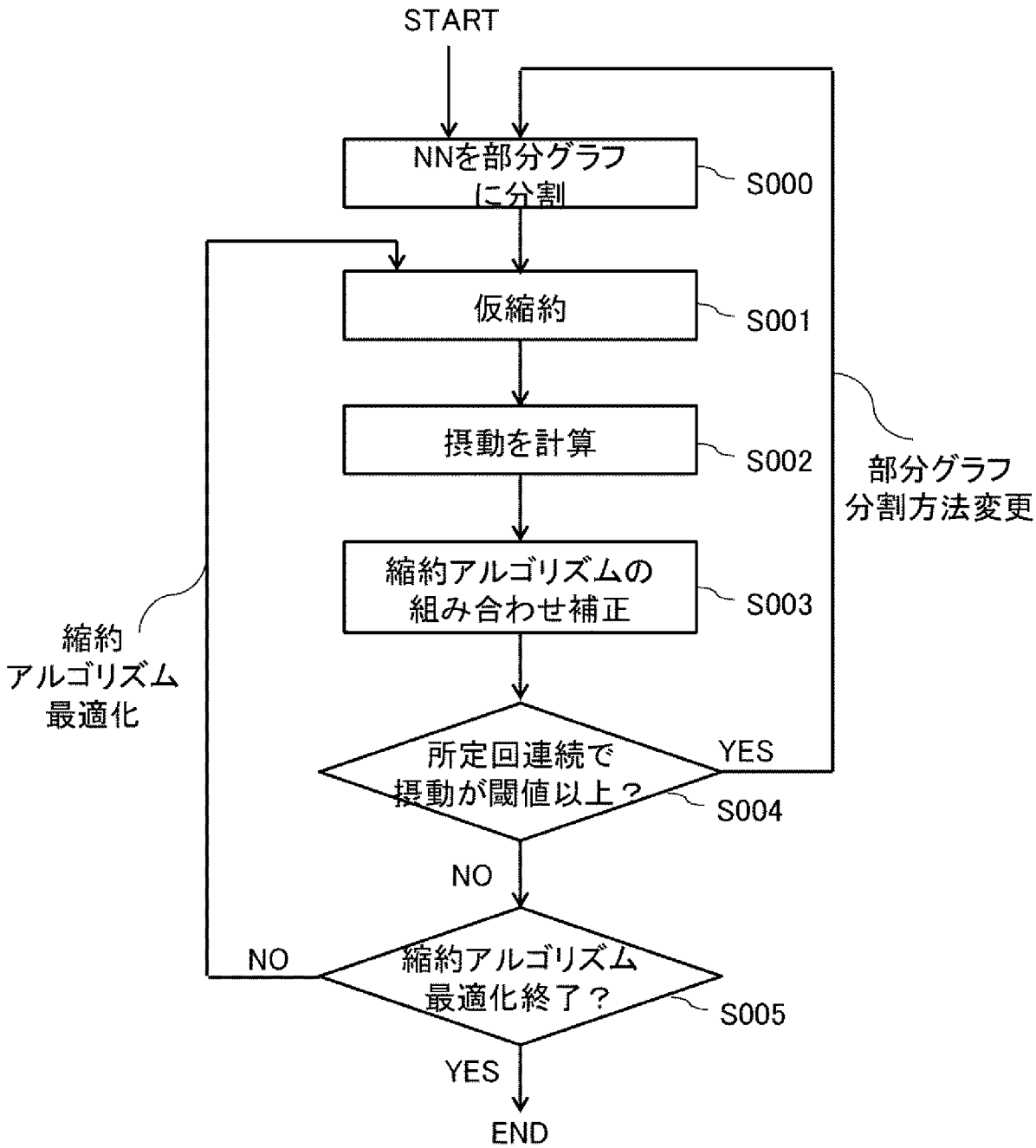
図8B

B002

縮約手法	#1	#2	#3	#4
プーリング	1.0	1.0	0.1	0.0
Row lank	0.5	0.0	1.0	0.0
Weight Sharing	0.0	0.0	0.1	0.0
・・				
・・				
量子化fp16	0.5	1.1	0.1	1.0
量子化INT8	0.2	1.1	0.1	1.0

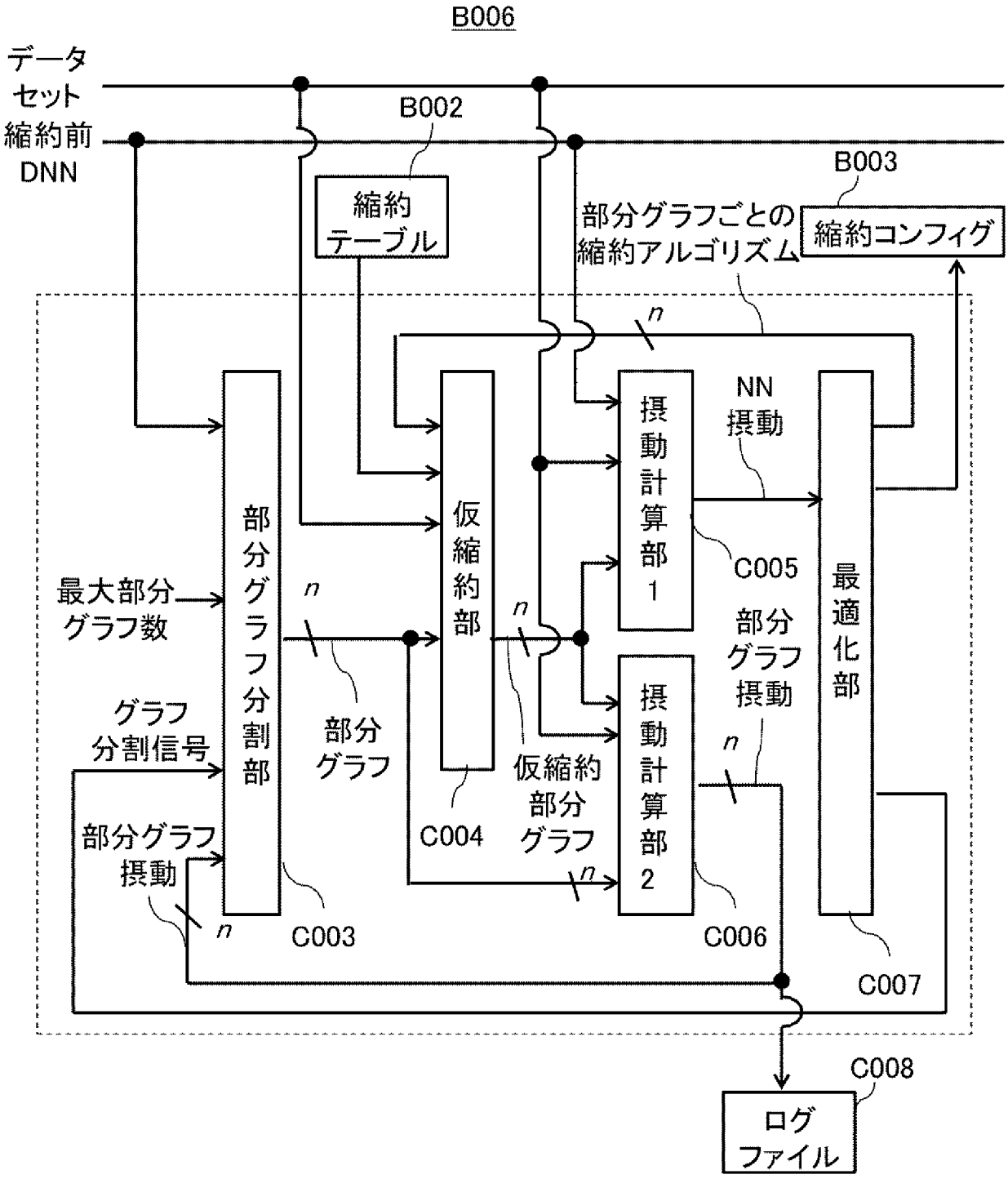
[図9]

図9


START
NNを部分グラフに分割
S000
仮縮約
S001
摂動を計算
S002
縮約アルゴリズムの組み合わせ補正
S003
所定回連続で摂動が閾値以上？
YES
S004
NO
縮約アルゴリズム最適化終了？
NO
S005
YES
END
縮約アルゴリズム最適化
部分グラフ分割方法変更

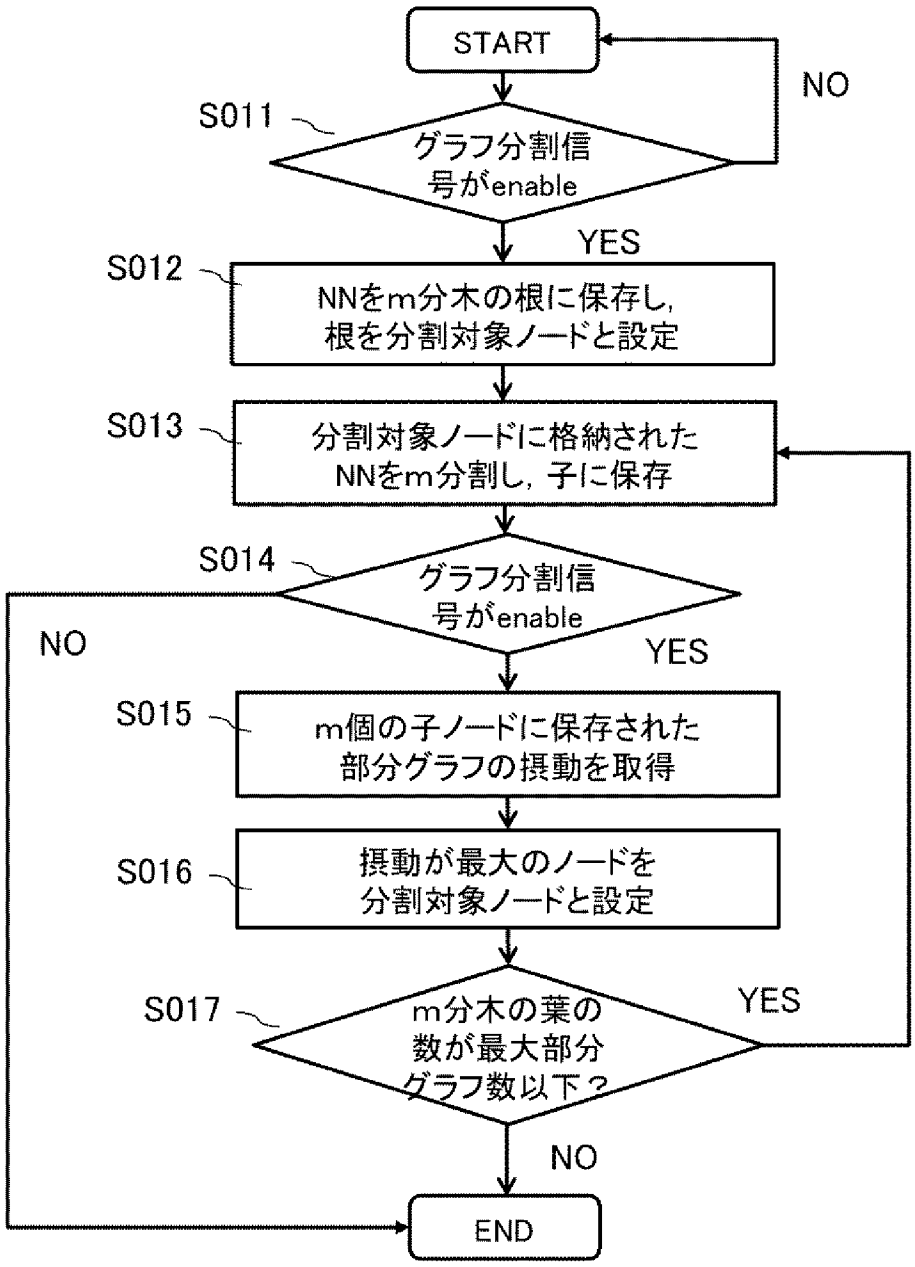
[図10]

図10


B006
データセット
縮約前DNN
B002
縮約テーブル
B003
部分グラフごとの縮約アルゴリズム
縮約コンフィグ
n
摂動計算部1
NN摂動
C005
最適化部
最大部分グラフ数
部分グラフ分割部
n
部分グラフ
仮縮約部
n
仮縮約部分グラフ
C004
摂動計算部2
部分グラフ摂動
n
C006
グラフ分割信号
部分グラフ摂動
n
C003
C007
C008
ログファイル

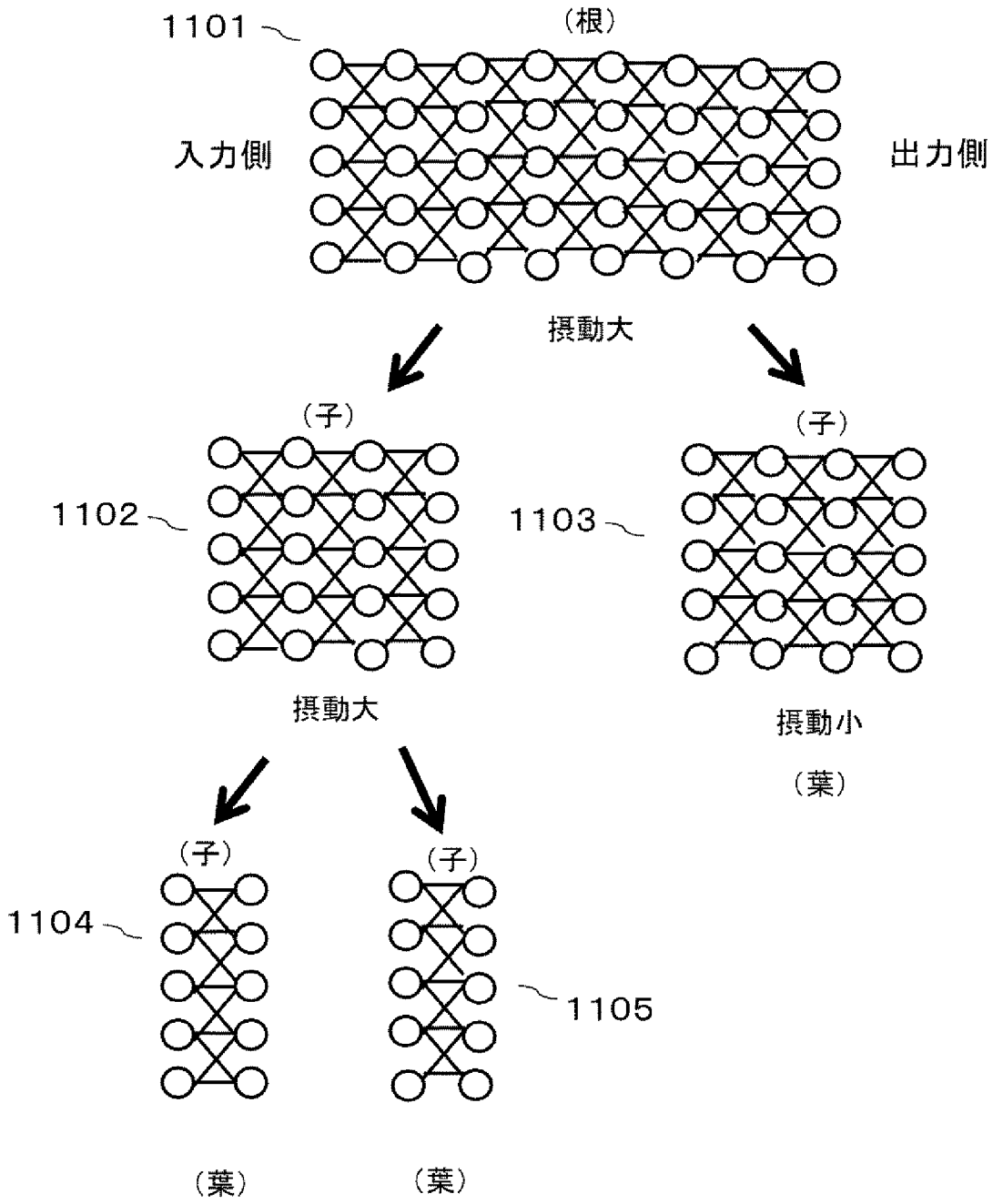
[図11A]

図11A


START
NO
S011
グラフ分割信号がenable
YES
S012
NNをm分木の根に保存し，根を分割対象ノードと設定
S013
分割対象ノードに格納されたNNをm分割し，子に保存
S014
グラフ分割信号がenable
NO
YES
S015
m個の子ノードに保存された部分グラフの摂動を取得
S016
摂動が最大のノードを分割対象ノードと設定
S017
m分木の葉の数が最大部分グラフ数以下？
YES
NO
END

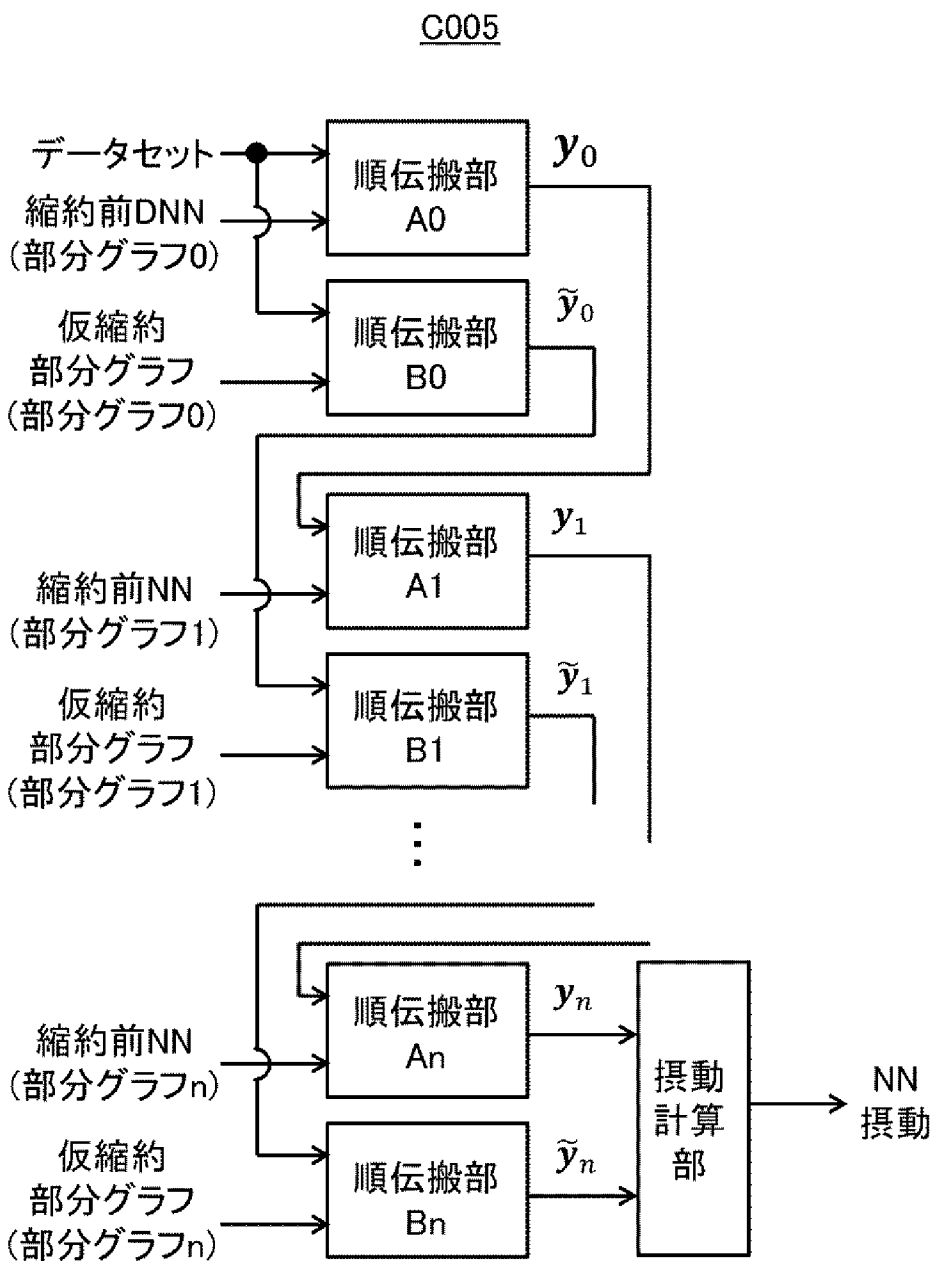
[図11B]

図11B



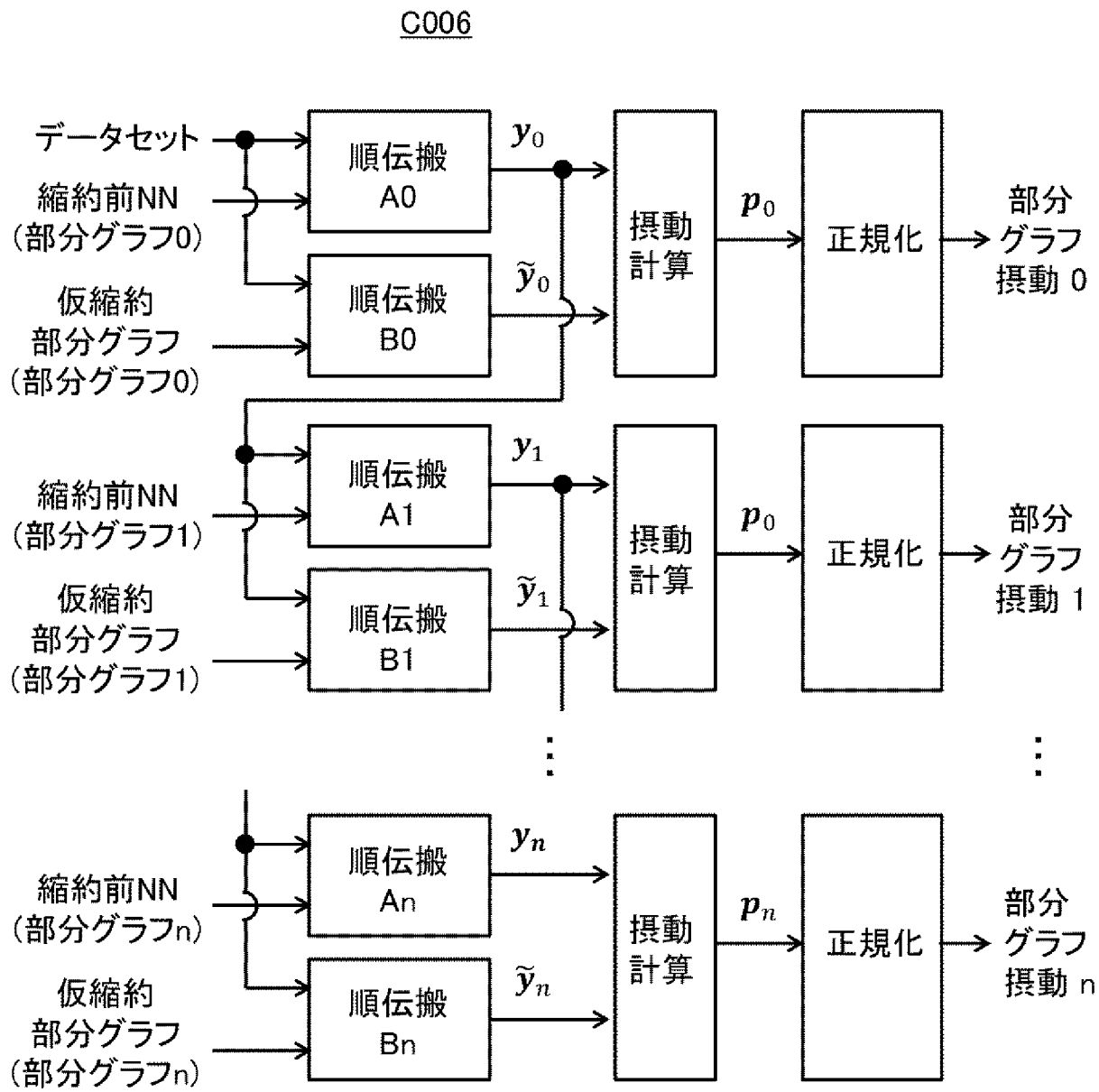
[図12]

図12


C005
データセット
縮約前DNN
(部分グラフ0)
仮縮約
部分グラフ
(部分グラフ0)
順伝搬部
A0
順伝搬部
B0
y_0
$\tilde{y}_0$
縮約前NN
(部分グラフ1)
仮縮約
部分グラフ
(部分グラフ1)
順伝搬部
A1
順伝搬部
B1
y_1
$\tilde{y}_1$
縮約前NN
(部分グラフn)
仮縮約
部分グラフ
(部分グラフn)
順伝搬部
An
順伝搬部
Bn
y_n
$\tilde{y}_n$
摂動
計算
部
NN
摂動

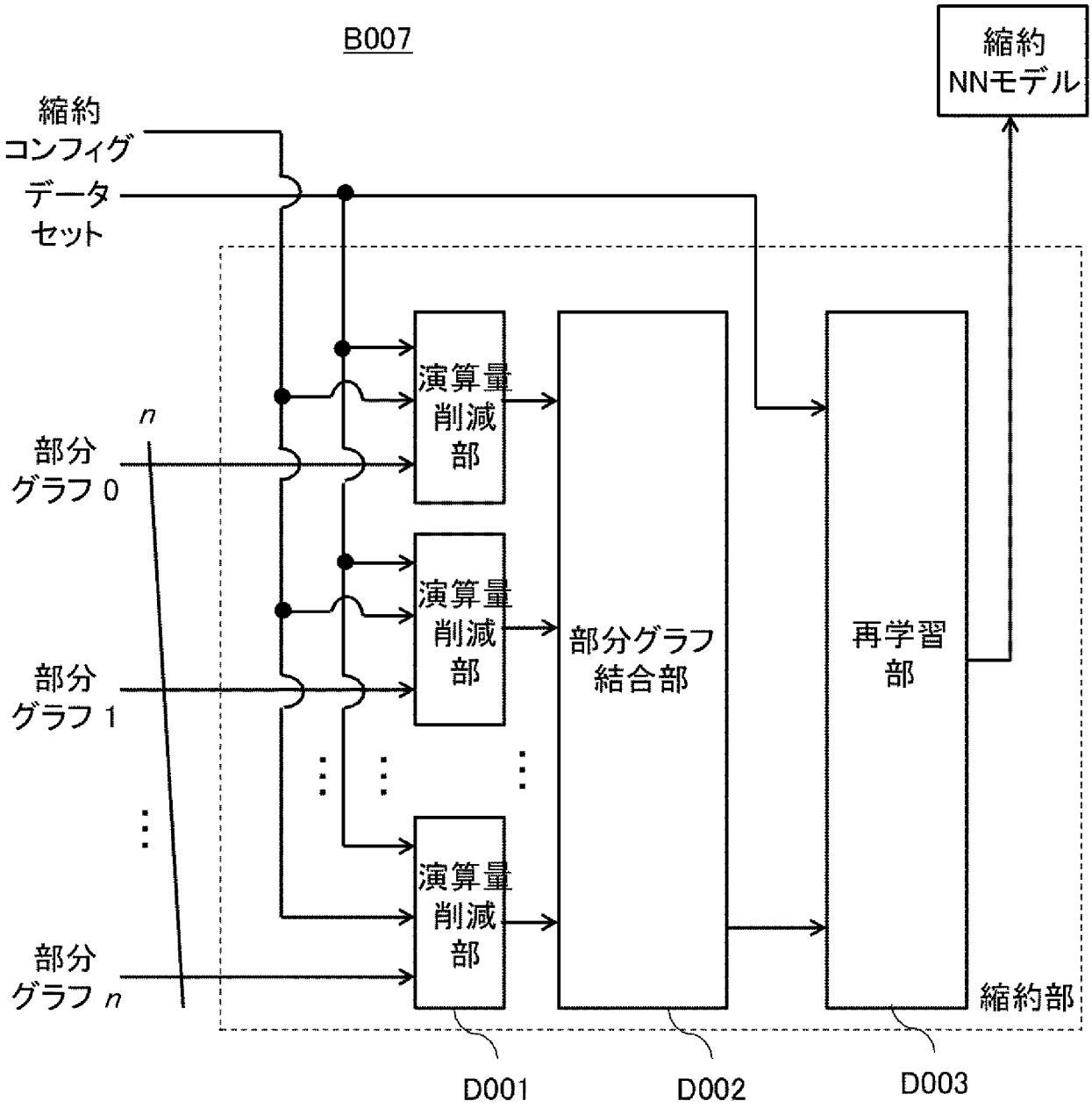
[図13]

図13


C006
データセット
縮約前NN
(部分グラフ0)
仮縮約
部分グラフ
(部分グラフ0)
順伝搬
A0
順伝搬
B0
y_0
$\tilde{y}_0$
摂動
計算
p_0
正規化
部分
グラフ
摂動 0
縮約前NN
(部分グラフ1)
仮縮約
部分グラフ
(部分グラフ1)
順伝搬
A1
順伝搬
B1
y_1
$\tilde{y}_1$
摂動
計算
p_0
正規化
部分
グラフ
摂動 1
縮約前NN
(部分グラフn)
仮縮約
部分グラフ
(部分グラフn)
順伝搬
An
順伝搬
Bn
y_n
$\tilde{y}_n$
摂動
計算
p_n
正規化
部分
グラフ
摂動 n

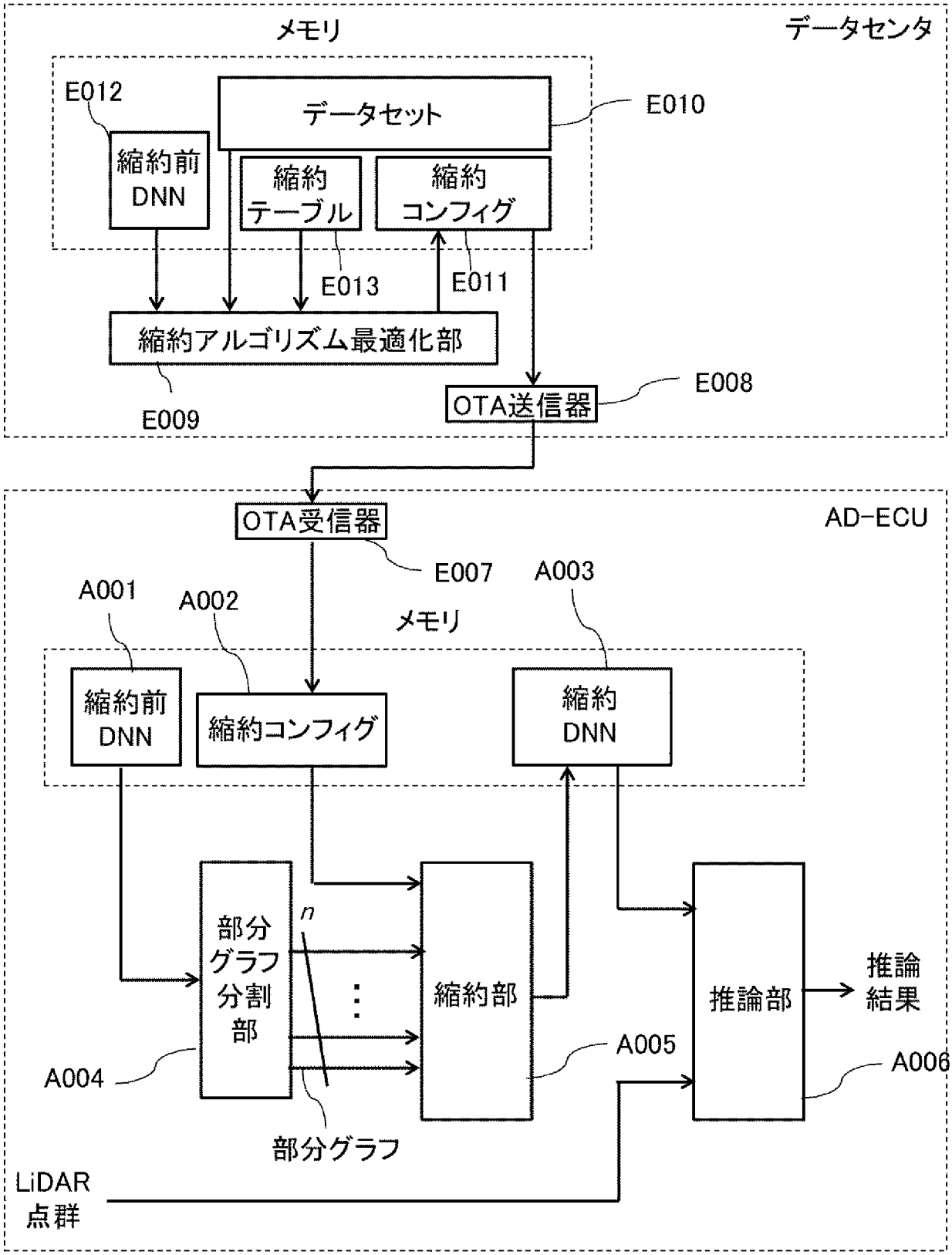
[図14]

図14


B007
縮約
NNモデル
縮約
コンフィグ
データ
セット
n
部分
グラフ 0
部分
グラフ 1
部分
グラフ n
演算量
削減
部
演算量
削減
部
演算量
削減
部
部分グラフ
結合部
再学習
部
縮約部
D001
D002
D003

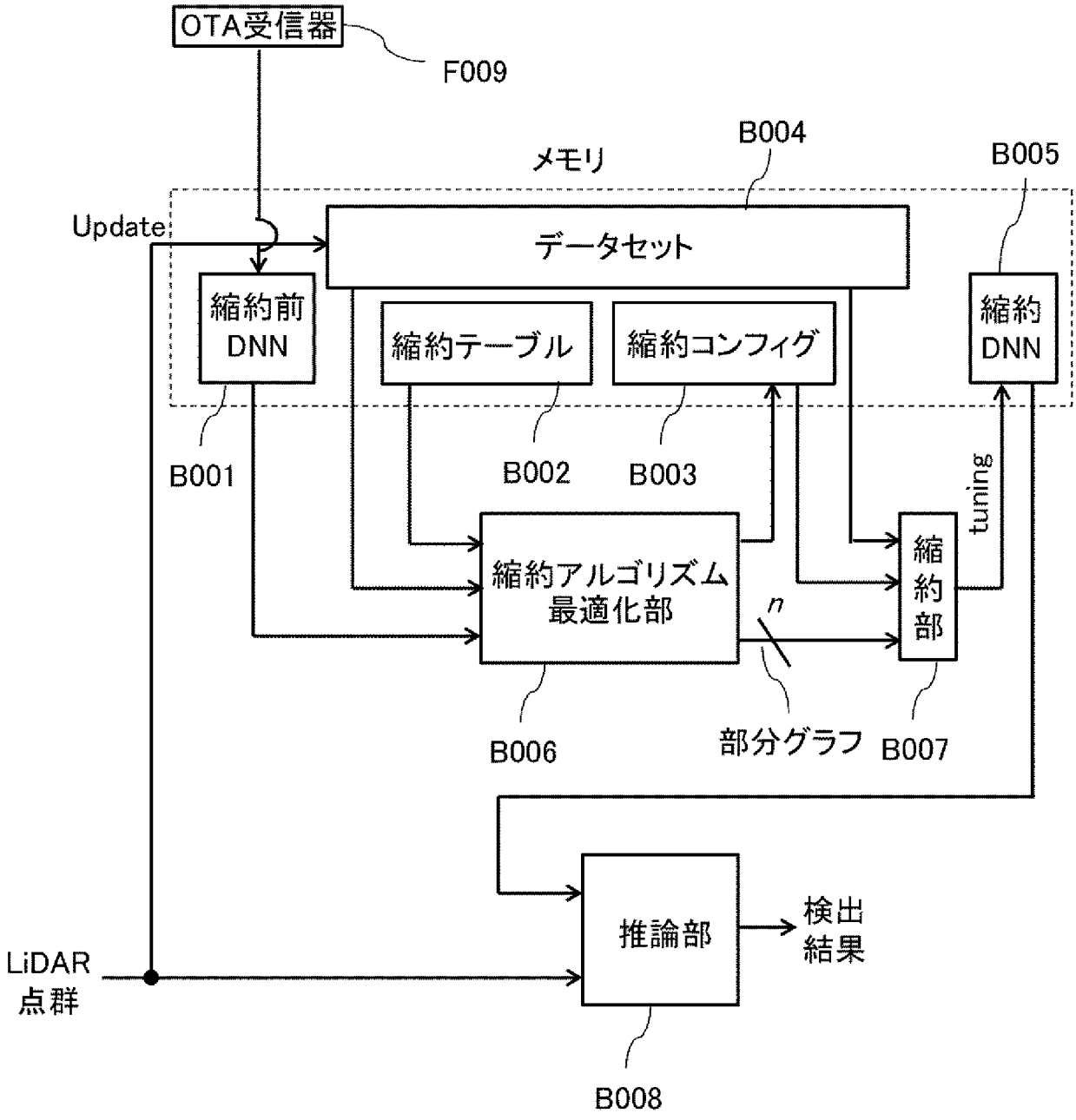
[図15]

図15


メモリ
データセンタ
E012
データセット
E010
縮約前
DNN
縮約
テーブル
縮約
コンフィグ
E013
E011
縮約アルゴリズム最適化部
E009
OTA送信器
E008
OTA受信器
AD-ECU
E007
A001
A002
A003
メモリ
縮約前
DNN
縮約コンフィグ
縮約
DNN
部分
グラフ
分割
部
n
縮約部
推論部
推論
結果
A004
A005
A006
部分グラフ
LiDAR
点群

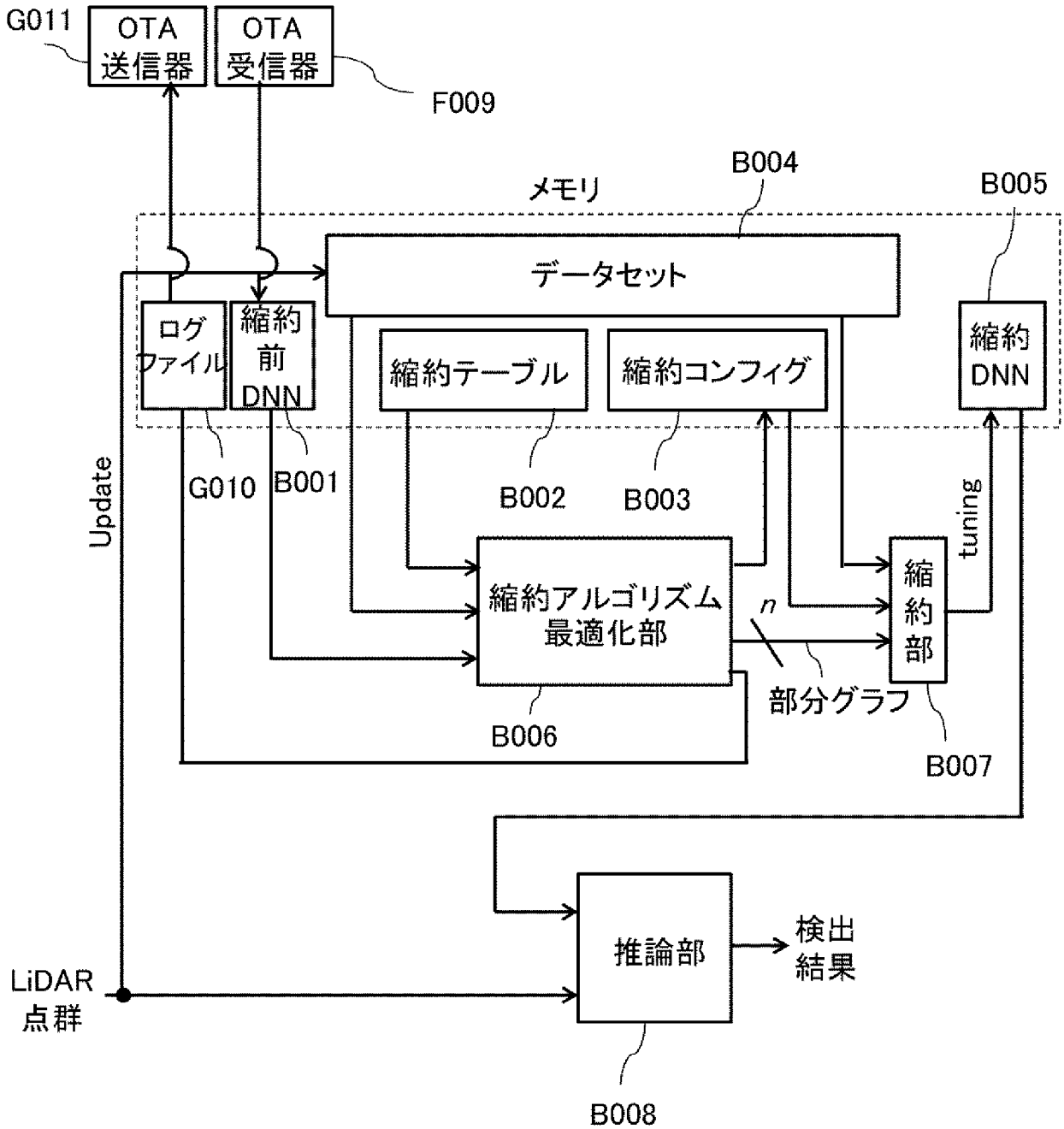
[図16]

図16


OTA受信器
F009
メモリ
B004
B005
Update
データセット
縮約前
DNN
縮約テーブル
縮約コンフィグ
縮約
DNN
B001
B002
B003
tuning
縮約アルゴリズム
最適化部
n
縮約
部
B006
部分グラフ
B007
推論部
検出
結果
LiDAR
点群
B008

[図17]

図17


G011
OTA
送信器
OTA
受信器
F009
メモリ
B004
B005
データセット
ログ
ファイル
縮約
前
DNN
縮約テーブル
縮約コンフィグ
縮約
DNN
G010
B001
B002
B003
Update
tuning
縮約アルゴリズム
最適化部
n
縮
約
部
部分グラフ
B006
B007
推論部
検出
結果
LiDAR
点群
B008

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2021/034898

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/08(2006.01)i
FI: G06N3/08 120

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N3/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2021
Registered utility model specifications of Japan 1996-2021
Published registered utility model applications of Japan 1994-2021

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/158058 A1 (HITACHI AUTOMOTIVE SYSTEMS, LTD.) 06 August 2020 (2020-08-06) paragraphs [0014]-[0032], [0080]-[0089]	1-3, 7
Y	paragraphs [0014]-[0032], [0080]-[0089]	4-6, 8-10
Y	WO 2019/146651 A1 (MORPHO, INC.) 01 August 2019 (2019-08-01) paragraphs [0035]-[0038], [0123], [0124]	4
Y	XU, Y., ET AL. Deep Neural Network Compression With Single and Multiple Level Quantization. Proceedings of the AAAI Conference on Artificial Intelligence. 29 April 2018, vol. 32, no. 1, pp. 4335-4342, [retrieved on 17 November 2021] Internet: <URL: https://ojs.aaai.org/index.php/AAAI/article/view/11663> "Overview" section	5, 6, 9, 10
Y	WO 2019/050771 A1 (PANASONIC INTELLECTUAL PROPERTY CORPORATION OF AMERICA) 14 March 2019 (2019-03-14) pp. 10, ll. 22-24	8

☐ 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	Date of mailing of the international search report
17 November 2021	**30 November 2021**
Name and mailing address of the ISA/JP **Japan Patent Office (ISA/JP)** **3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915** **Japan**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2021/034898

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☑ Claims Nos.: **11-15**
because they relate to subject matter not required to be searched by this Authority, namely:

Claims 11-15 are methods related to mathematical theory or the execution of purely mental acts, and thus pertain to subject matter for which an international search is not required.

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/034898

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2020/158058 A1	06 August 2020	JP 2020-123269 A	
		CN 113366509 A	
WO 2019/146651 A1	01 August 2019	US 2021/0158136 A1 paragraphs [0048]-[0051], [0150], [0151]	
		EP 3614312 A1	
		CN 110709862 A	
		JP 2019-128831 A	
WO 2019/050771 A1	14 March 2019	JP 2020-532777 A	
		US 2020/0097802 A1	
		EP 3679524 A1	
		CN 110663048 A	

国際調査報告

国際出願番号
PCT/JP2021/034898

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））
G06N 3/08(2006.01)i
FI: G06N3/08 120

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））
G06N3/08

最小限資料以外の資料で調査を行った分野に含まれるもの
日本国実用新案公報　１９２２－１９９６年
日本国公開実用新案公報　１９７１－２０２１年
日本国実用新案登録公報　１９９６－２０２１年
日本国登録実用新案公報　１９９４－２０２１年

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

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	WO 2020/158058 A1 (日立オートモティブシステムズ株式会社) 06.08.2020 (2020-08-06) 段落[0014]-[0032], [0080]-[0089]	1-3, 7
Y	段落[0014]-[0032], [0080]-[0089]	4-6, 8-10
Y	WO 2019/146651 A1 (株式会社モルフォ) 01.08.2019 (2019-08-01) 段落[0035]-[0038], [0123], [0124]	4
Y	XU, Y., ET AL., Deep Neural Network Compression With Single and Multiple Level Quantization, Proceedings of the AAAI Conference on Artificial Intelligence, 2018.04.29, Vol. 32, No. 1, pp. 4335-4342, [retrieved on 2021-11-17] Retrieved from the Internet: <URL: https://ojs.aaai.org/index.php/AAAI/article/view/11663> "Overview" section	5, 6, 9, 10
Y	WO 2019/050771 A1 (PANASONIC INTELLECTUAL PROPERTY CORPORATION OF AMERICA) 14.03.2019 (2019-03-14) p. 10, ll. 22-24	8

☐ C欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
"A" 特に関連のある文献ではなく、一般的技術水準を示すもの
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
"O" 口頭による開示、使用、展示等に言及する文献
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
"Y" 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
"&" 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
17.11.2021	30.11.2021

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）
金木　陽一　5B 4876

電話番号 03-3581-1101 内線 3545

国際調査報告

国際出願番号
PCT/JP2021/034898

第Ⅱ欄　請求の範囲の一部の調査ができないときの意見（第１ページの２の続き）

法第8条第3項（ＰＣＴ17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☑ 請求項 11-15
は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、

請求項11-15は、数学の理論、又は、純粋に精神的な行為の遂行に関する方法であり、国際調査をすることを要しない対象に係るものである。

2. ☐ 請求項
は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、

3. ☐ 請求項
は、従属請求の範囲であってＰＣＴ規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

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/JP2021/034898

引用文献	公表日	パテントファミリー文献	公表日
WO 2020/158058 A1	06.08.2020	JP 2020-123269 A	
		CN 113366509 A	
WO 2019/146651 A1	01.08.2019	US 2021/0158136 A1 paras. [0048]-[0051], [0150], [0151]	
		EP 3614312 A1	
		CN 110709862 A	
		JP 2019-128831 A	
WO 2019/050771 A1	14.03.2019	JP 2020-532777 A	
		US 2020/0097802 A1	
		EP 3679524 A1	
		CN 110663048 A	