

令和 6 年 9 月 30 日現在

機関番号：11301
研究種目：基盤研究(B)（一般）
研究期間：2019～2023
課題番号：19H04078
研究課題名（和文）雑音畳込みニューラルネットワークの研究開発

研究課題名（英文）On a noise convolutional neural network

研究代表者

中原 啓貴（Nakahara, Hiroki）

東北大学・未踏スケールデータアナリティクスセンター・教授

研究者番号：20624414

交付決定額（研究期間全体）：（直接経費） 13,200,000円

研究成果の概要（和文）：雑音CNNは既存CNNと等価であることを導出した。雑音CNN専用回路の設計を行い、FPGAプロトタイプ実装を行った。雑音生成回路と1×1サイズ畳込みを組み合わせた構成に適したノイズ生成回路を調査し、回路実装した。GPUと比較して提案手法の優位性を示した。雑音CNN回路に既存のパラメータ削減方法である低ビット化と枝刈り手法を適用して更に性能向上させた。雑音畳込みは既存畳込みと等価なので、既存手法と組合せが可能であることを示した。また、雑音CNNの学習高速化のためのFPGAクラスタの環境構築を行い、雑音CNNの学習方式を検討した。

研究成果の学術的意義や社会的意義

電力やデバイスの制約で実現できなかった高度な認識技術が組み込み機器に実現できた。また、設計のボトルネックであった学習時間が短縮された。研究期間後は、専用チップ化による更なる性能向上とコスト削減に取り組む予定である。

研究成果の概要（英文）：We derived that the noise CNN is equivalent to the existing CNN. We designed a dedicated circuit for noise CNN and implemented an FPGA prototype. We investigated a noise generation circuit suitable for a configuration that combines a noise generation circuit and 1×1 size convolution, and implemented the circuit. We demonstrated the superiority of the proposed method compared to GPU. We further improved the performance of the noisy CNN circuit by applying existing parameter reduction methods such as bit reduction and pruning. Since noise convolution is equivalent to existing convolution, we showed that it can be combined with existing methods. We also constructed an FPGA cluster environment to speed up the learning of noisy CNNs, and investigated the learning method for noisy CNNs.

研究分野：AI, Machine learning, FPGA

キーワード：AI Machine learning FPGA

様式 C-19

科学研究費助成事業（科学研究費補助金）研究成果報告書

様式 C-19、F-19-1（共通）

1. 研究開始当初の背景

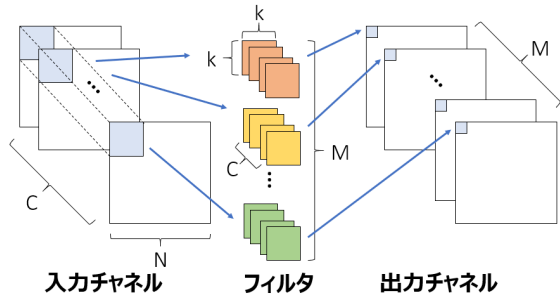


図 1: 畳み込み演算

表 1: 代表的な CNN のパラメータ数

CNN	層数	パラメータ数
AlexNet	8	60M
VGG	19	140M
GoogLeNet	22	6M
ResNet	152	60M

ニューラルネットワーク(NN: Neural Network)の一種である畳み込みニューラルネットワーク(CNN: Convolutional Neural Network) はパターン認識フィルタがカスケードされた構造を持ち、ビッグデータによる訓練を可能とするため、幅広いコンピュータビジョンアプリケーション（クラス分類、物体認識、意味的領域分割、姿勢推定など）において既存の機械学習ベース手法よりも優れた認識精度を達成している。CNN はネットワークを深くするほど認識精度が上がるため、大規模な CNN が求められる。すなわち、CNN は計算量や内部パラメータの数が膨大であり、高性能な GPU サーバーで実装される場合が多い。

図 1 に典型的な CNN における畳み込み演算を示す。C を入力チャンネル数、M を出力チャンネル数、N を入力画像サイズ、k をフィルタサイズとする。ここで、パラメータ(重み)は k^2CM 個であり、典型的な CNN では後段になるにつれてチャンネル数 MC が増える。従って、CNN のパラメータ量は後段部分が支配的である。表 2 に代表的な CNN のパラメータサイズを示す。画像処理の主なプラットフォームは組込み機器であり、組込み GPU や FPGA のオンチップメモリ(数 Mbit 程度)に対してパラメータは膨大であり、DDR メモリ等のオフチップメモリに格納しなければならないが、速度低下(主にオフチップメモリのアクセス時間)と電力のオーバーヘッドが生じてしまう。表 2 より、既存 CNN を直接実装した場合、大量のパラメータを処理するため、極端に処理速度が遅く電力消費が大きい状況である。 また、学習に多大な時間を要する問題もある。

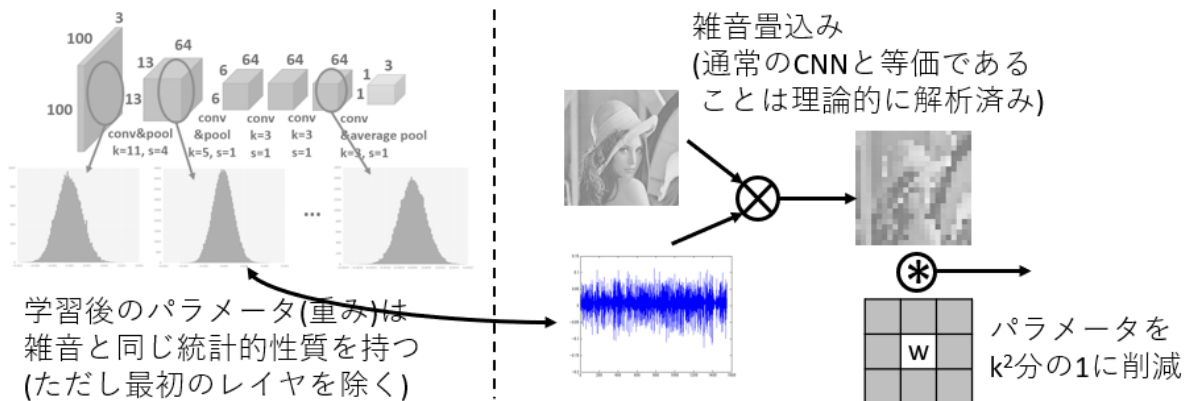


図 2 : CNN パラメータのヒストグラム(左, AlexNet)と提案する雑音畳み込み演算(右)

2. 研究の目的

図2(左)に手書き文字データセットである MNIST を用いて CNN の一種である AlexNet を学習した後のパラメータ (重み) のヒストグラムを示す。興味深いのは十分な学習が終わると、パラメータのヒストグラムは最初のレイヤを除いて 0 を中心とした対照的な分布に従うことである。すなわち、「大部分の重みは雑音でモデル化し、データに依存する分布の偏りのみ学習によって補完できるのではないか」という仮説である。もしこの仮説が成立すれば、大部分のパラメータが雑音によって置換できるため、最小の 1×1 サイズ畳み込みで十分である。このとき、パラメータを k^2 分の 1 (通常, $k=3$ が使われる) まで削減することで、計算量もほぼ 1 桁削減できる。また、パラメータを削減することでオンチップメモリに格納できれば、オフチップメモリへのアクセス時間と電力を削減できる。もし、既存の削減法と併用できれば強力なパラメータ削減法になり得る。予備実験において、特定の条件下で仮説、すなわち雑音畳み込み演算が畳み込み演算が等価であることを明らかにしている。

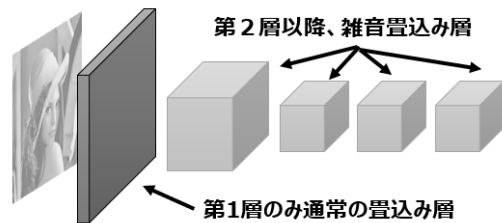


図 3: 提案する雑音畳み込み CNN 回路(右)

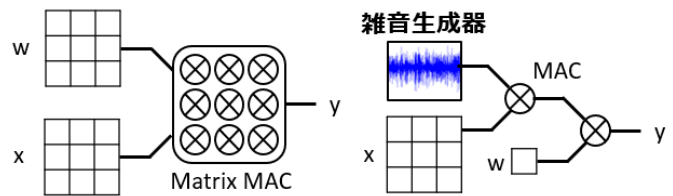


図 4: 既存畳み込み回路(左)と雑音畳み込み回路(右)

図 3 に提案する雑音畳み込み CNN を示す。仮定を満たすために、初段のレイヤは通常の畳み込みを行う。2 層目以降、提案する雑音畳み込み層で実装する。画像認識ベンチマーク (CIFAR10) と代表的な CNN(ResNet18)で認識率の比較を予備的に実験した。雑音層を加えても認識率の低下はごくわずかであり、理論的な解析を実験的に示せた。このとき、パラメータの削減率は 8.6 分の 1であった。また、パラメータ数の削減により、学習時間を 41%削減できた。パラメータ数に比べて学習時間を短縮できなかったのは、予備実験で用いた GPU が雑音生成の計算オーバーヘッドが大きいからであった。図 4 に既存の畳み込み回路と提案する雑音畳み込み回路の比較を示す。大部分のパラメータを雑音生成器で置換するため、ボトルネックであったパラメータを保持するメモリ量を削減できる。また、演算回数そのものを最大で k^2 分の 1 に削減するため、高速化を達成できる見通しが立った。GPU による雑音畳み込みは雑音生成器 (乱数生成器) がボトルネックであるため、本研究計画では雑音 CNN 専用回路を FPGA プロトタイプ実装を通して研究開発する。また、雑音 CNN の学習高速化クラスタの開発を行う。

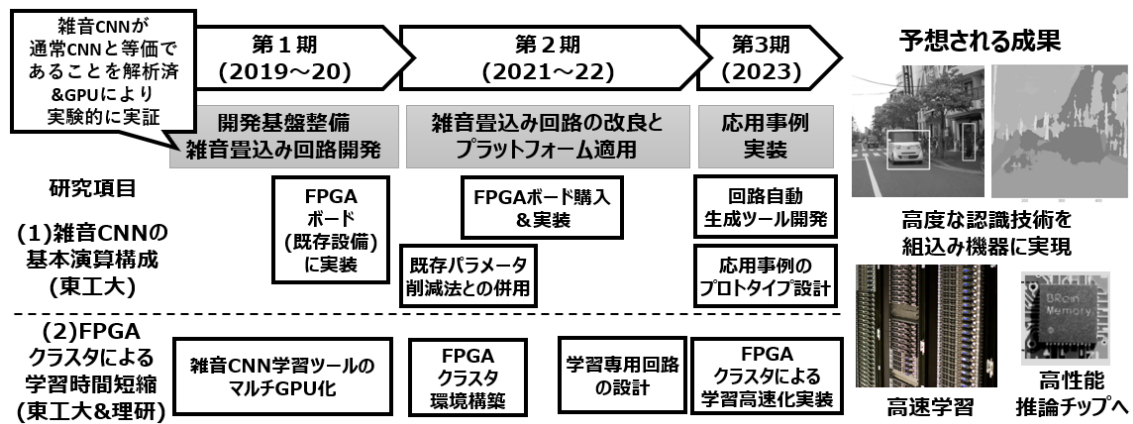


図5：研究スケジュール

3. 研究の方法

本研究開発のスケジュールを図5に示す。以下、研究方法の詳細を述べる。

第1期 (2019~20年度)：雑音 CNN の学習時間が回路開発に占める割合が大きいため、マルチ GPU 学習環境を整備する。雑音 CNN 専用回路の設計を行い、FPGA プロトタイプ実装を行う。雑音生成回路と 1×1 サイズ畳込みを組み合わせる構成に適したノイズ生成回路を調査し、回路実装する。GPU と比較して提案手法の優位性を示す。

第2期 (2021~22年度)：雑音 CNN 回路に既存のパラメータ削減方法である低ビット化と枝刈り手法を適用して更に性能向上させる。予備実験により、雑音畳込みは既存畳込みと等価なので、既存手法と組合せが可能である。これら組合せの FPGA 実装を行い、更に性能を向上させる。また、雑音 CNN の学習高速化のための FPGA クラスタの環境構築を行う。雑音 CNN の学習方式を検討する。

第3期 (2023年度)：前年度までに培った基礎技術を用いて FPGA プロトタイプを応用事例 (例えば、姿勢推定、シーン認識、物体認識など) に適用する。提案手法によりパラメータ自体を削減したので学習による更新量も k^2 分の 1 に削減できるはずである。しかし、汎用 GPU ではノイズ生成部分がボトルネックになることが予想できるため、専用回路によるノイズ生成器と 1×1 サイズ畳込み回路の組み合わせによる学習が必要である。分担者 (理研佐野) の開発中の FPGA クラスタとその設計手法を用いて学習専用回路を FPGA 上に実現する。最後に、本研究成果を元に雑音畳込み回路自動生成ツールの開発に取り組む。

4. 研究成果

雑音 CNN は既存 CNN と等価であることを導出した。雑音 CNN 専用回路の設計を行い、FPGA プロトタイプ実装を行った。雑音生成回路と 1×1 サイズ畳込みを組み合わせる構成に適したノイズ生成回路を調査し、回路実装した。GPU と比較して提案手法の優位性を示した。雑音 CNN 回路に既存のパラメータ削減方法である低ビット化と枝刈り手法を適用して更に性能向上させた。雑音畳込みは既存畳込みと等価なので、既存手法と組合せが可能であることを示した。また、雑音 CNN の学習高速化のための FPGA クラスタの環境構築を行い、雑音 CNN の学習方式を検討した。

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件／うち国際共著 1件／うちオープンアクセス 0件）

1. 著者名 Takeshi Senoo, Akira Jinguji, Ryosuke Kuramochi, Hiroki Nakahara	4. 巻 105-D(12)
2. 論文標題 Multilayer Perceptron Training Accelerator Using Systolic Array	5. 発行年 2022年
3. 雑誌名 EICE Trans. Inf. Syst.	6. 最初と最後の頁 2048-2056
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Zhiqiang Que, Hiroki Nakahara, Eriko Nurvitadhi, Andrew Boutros, Hongxiang Fan, Chenglong Zeng, Jiuxi Meng, Kuen Hung Tsoi, Xinyu Niu, Wayne Luk	4. 巻 30(2)
2. 論文標題 Recurrent Neural Networks With Column-Wise Matrix-Vector Multiplication on FPGAs	5. 発行年 2022年
3. 雑誌名 IEEE Trans. Very Large Scale Integr. Syst.	6. 最初と最後の頁 227-237
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Naoto Soga, Shimpei Sato, Hiroki Nakahara	4. 巻 104-D(8)
2. 論文標題 Energy-Efficient ECG Signals Outlier Detection Hardware Using a Sparse Robust Deep Autoencoder	5. 発行年 2021年
3. 雑誌名 EICE Trans. Inf. Syst.	6. 最初と最後の頁 1121-1129
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Akira Jinguji, Shimpei Sato, Hiroki Nakahara	4. 巻 104-D(12)
2. 論文標題 Weight Sparseness for a Feature-Map-Split-CNN Toward Low-Cost Embedded FPGAs	5. 発行年 2021年
3. 雑誌名 EICE Trans. Inf. Syst.	6. 最初と最後の頁 2040-2047
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ryosuke Kuramochi, Hiroki Nakahara	4. 巻 104-D(12)
2. 論文標題 A Low-Latency Inference of Randomly Wired Convolutional Neural Networks on an FPGA	5. 発行年 2021年
3. 雑誌名 IEICE Trans. Inf. Syst.	6. 最初と最後の頁 2068-2077
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masayuki Shimoda, Youki Sada, Hiroki Nakahara	4. 巻 93(5)
2. 論文標題 FPGA-Based Inter-layer Pipelined Accelerators for Filter-Wise Weight-Balanced Sparse Fully Convolutional Networks with Overlapped Tiling	5. 発行年 2021年
3. 雑誌名 J. Signal Process. Syst.	6. 最初と最後の頁 499-512
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masayuki Shimoda, Youki Sada, Ryosuke Kuramochi, Shimpei Sato, Hiroki Nakahara	4. 巻 103-D(12)
2. 論文標題 SENTEI: Filter-Wise Pruning with Distillation towards Efficient Sparse Convolutional Neural Network Accelerators	5. 発行年 2020年
3. 雑誌名 IEICE Trans. Inf. Syst.	6. 最初と最後の頁 2463-2470
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2020PAP0013	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hiroki Nakahara, Haruyoshi Yonekawa, Tomoya Fujii, Masayuki Shimoda, Shimpei Sato	4. 巻 102-D(5)
2. 論文標題 GUINNESS: A GUI Based Binarized Deep Neural Network Framework for Software Programmers	5. 発行年 2019年
3. 雑誌名 IEICE Trans. Inf. Syst.	6. 最初と最後の頁 1003-1011
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masayuki Shimoda, Shimpei Sato, Hiroki Nakahara	4. 巻 102-D(5)
2. 論文標題 Power Efficient Object Detector with an Event-Driven Camera for Moving Object Surveillance on an FPGA	5. 発行年 2019年
3. 雑誌名 IEICE Trans. Inf. Syst.	6. 最初と最後の頁 1020-1028
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計14件 (うち招待講演 1件 / うち国際学会 14件)

1. 発表者名 Takeshi Senoo, Akira Jinguji, Ryosuke Kuramochi, Hiroki Nakahara
2. 発表標題 A Multilayer Perceptron Training Accelerator using Systolic Array
3. 学会等名 IEEE APCCAS (国際学会)
4. 発表年 2021年

1. 発表者名 Kota Ando, Jaehoon Yu, Kazutoshi Hirose, Hiroki Nakahara, Kazushi Kawamura, Thiem Van Chu, Masato Motomura
2. 発表標題 Edge Inference Engine for Deep & Random Sparse Neural Networks with 4-bit Cartesian-Product MAC Array and Pipelined Activation Aligner
3. 学会等名 IEEE HCS (国際学会)
4. 発表年 2021年

1. 発表者名 Naoto Soga, Ryosuke Kuramochi, Hiroki Nakahara
2. 発表標題 A High-Throughput Detection Circuit based on $2q+1$ -Valued Deep Neural Networks
3. 学会等名 IEEE ISMVL (国際学会)
4. 発表年 2021年

1. 発表者名 Hiroki Nakahara, Zhiqiang Que, Wayne Luk
2. 発表標題 High-Throughput Convolutional Neural Network on an FPGA by Customized JPEG Compression
3. 学会等名 FCCM2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Zhiqiang Que, Hiroki Nakahara, Eriko Nurvitadhi, Hongxiang Fan, Chenglong Zeng, Jiuxi Meng, Xinyu Niu, Wayne Luk
2. 発表標題 Optimizing Reconfigurable Recurrent Neural Networks
3. 学会等名 FCCM2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Ryosuke Kuramochi, Hiroki Nakahara
2. 発表標題 An FPGA-Based Low-Latency Accelerator for Randomly Wired Neural Networks
3. 学会等名 FPL2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Hiroki Nakahara
2. 発表標題 2n+1-valued SSS-Net: Uniform Shift, Channel Sparseness, and Channel Shuffle
3. 学会等名 ISMVL2020 (国際学会)
4. 発表年 2020年

1 . 発表者名 Youki Sada, Naoto Soga, Masayuki Shimoda, Akira Jinguji, Shimpei Sato, Hiroki Nakahara
2 . 発表標題 Fast Monocular Depth Estimation on an FPGA
3 . 学会等名 IPDPS Workshops 2020 (国際学会)
4 . 発表年 2020年

1 . 発表者名 Masayuki Shimoda, Youki Sada, Hiroki Nakahara
2 . 発表標題 Filter-Wise Pruning Approach to FPGA Implementation of Fully Convolutional Network for Semantic Segmentation
3 . 学会等名 ARC2019 (国際学会)
4 . 発表年 2019年

1 . 発表者名 Hiroki Nakahara, Youki Sada, Masayuki Shimoda, Kouki Sayama, Akira Jinguji, Shimpei Sato
2 . 発表標題 FPGA-Based Training Accelerator Utilizing Sparseness of Convolutional Neural Network.
3 . 学会等名 FPL2019 (国際学会)
4 . 発表年 2019年

1 . 発表者名 Youki Sada, Masayuki Shimoda, Akira Jinguji, Hiroki Nakahara
2 . 発表標題 A Dataflow Pipelining Architecture for Tile Segmentation with a Sparse MobileNet on an FPGA
3 . 学会等名 FPT 2019 (国際学会)
4 . 発表年 2019年

1. 発表者名 Atsuki Munakata, Hiroki Nakahara, Shimpei Sato
2. 発表標題 Noise Convolutional Neural Networks and FPGA Implementation
3. 学会等名 ISMVL 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Ryosuke Kuramochi, Youki Sada, Masayuki Shimoda, Shimpei Sato
2. 発表標題 Many Universal Convolution Cores for Ensemble Sparse Convolutional Neural Networks
3. 学会等名 MCSoc 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Hiroki Nakahara
2. 発表標題 Deep Learning Accelerator for an Intelligent Camera
3. 学会等名 HEART 2019 (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	佐野 健太郎 (Sano Kentaro) (00323048)	国立研究開発法人理化学研究所・計算科学研究センター・ チームリーダー (82401)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	佐藤 真平 (Sato Simpei) (80782763)	信州大学・学術研究院工学系・助教 (13601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関