

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 6 日現在

機関番号：12605

研究種目：基盤研究(C)（一般）

研究期間：2014～2016

課題番号：26330060

研究課題名（和文）同期シフトデータ転送に基づくディープラーニング・ハードウェア構成法の研究

研究課題名（英文）Study on deep learning hardware construction method based on synchronous shift data transfer

研究代表者

北澤 仁志（KITAZAWA, HITOSHI）

東京農工大学・工学（系）研究科（研究院）・教授

研究者番号：60345329

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究ではディープラーニングの高速処理と高性能化のために、FPGAを用いた並列処理ハードウェアの実現、および動画処理への応用を目指した。まず、SIMDアレイ型回路を用いた“同期シフトデータ転送”を適用した誤差逆伝搬学習回路をFPGAで実現した。次に、スループット1クロックのパイプライン積和演算器を用いたディープラーニングの高速処理回路を実現した。更に、ディープラーニングの応用として1枚のFPGAボードでカメラ64台の映像を入力し、移動物体の抽出、追跡、識別を行うMulti-Stream Trackingハードウェアにディープラーニングを適用し、高い精度で物体識別ができることを示した。

研究成果の概要（英文）：In this research, we aimed to realize parallel processing hardware using FPGA and applying it to motion image processing for high speed processing and high identification accuracy of deep learning. First, using an SIMD array type circuit, a back propagation learning circuit applying "synchronous shift data transfer" was realized with an FPGA. Next, a high-speed processing circuit of deep learning using pipelined multiply-accumulate circuits with throughput 1 clock per pixel was realized. Furthermore, as an application of deep learning, we show that deep learning can identify objects with high accuracy in Multi-Stream Tracking hardware that inputs images from 64 cameras and extracts, tracks and identifies moving objects by using one FPGA board.

研究分野：動画処理ハードウェア

キーワード：ディープラーニング FPGA リンコンフィギャラブルシステム 物体識別 移動物体抽出

1. 研究開始当初の背景

(1)ディープラーニングが、画像や音声の認識などにおいて従来手法を超える高い性能を実現したことから広く注目を集めた。この手法は大きな可能性を秘めているが、適切な構造の決定や誤差逆伝搬による学習に極めて多大な処理時間を要する。一方、ディープラーニングには本質的に高い並列度があり、ハード化による高速化が試行され始めた。

(2) ディープラーニングの中心となる計算は、 $\sum W_{ji}X_i$  のように入力  $X_i$  の重み付き総和を求めるものである。これを高速に処理するために木構造積和演算器が用いられるが、複数の演算器を用いるとメモリアクセス競合が生じる。これを避けるため GPU ではメモリを複数のバンクに分割しているが、分割された他のメモリバンクへのアクセスによってバス競合が生じる。また、キャッシュメモリの書き戻しや再コピーのため演算効率は大きく低下してしまう。

(3)提案者は FPGA を用いた並列処理において、メモリアクセス競合やバス競合を避けることのできる“同期シフトデータ転送”を考案し、動画像中の移動物体のトラッキングと電磁界解析において、GPU を凌ぐ処理速度が得られることを示した。これをディープラーニングに応用すれば学習の高速処理が実現できると考えられる。

2. 研究の目的

(1)ディープラーニングを効率良く実行するため、SIMD (Single-Instruction Multi Data) アレイ型、パイプライン型、マルチコア型など種々並列処理手法について、適切なハードウェア構成法を探す。

(2)ディープラーニングの計算には、識別を行う前進計算部分と、識別誤差を逆伝搬させて枝の重みを学習する誤差逆伝搬部分があるが、これらの性質は大きく異なる。処理に応じて前述のどの並列処理手法を適用するのが良いかを明らかにする。

(3)ディープラーニングの性能を具体的応用により明らかにするため、複数カメラ映像の実時間解析に適用する。種々の構造や学習手法を試行し、動画像中の物体識別器として有効に動作させるための拡張・改良を行う。

3. 研究の方法

(1)まず、SIMD アレイ上で同期シフトデータ転送を用い、ディープラーニングの前進計算回路および誤差逆伝搬学習回路を作成する。前進計算では畳込み回路、プーリング回路、全結合型識別回路などを用いる。また、誤差逆伝搬学習用回路では重みの更新値の総和を求める専用回路を用いる。実装する回路のノード数、重み数と、FPGA のロジックエレ

メント、レジスタ、メモリ、DSP 演算器の使用量から実装可能な回路規模を予測する。また、CPU や GPU と比較した実行速度、および、段数やノード数と処理時間の関係を明らかにする。

(2)次に、パイプライン型回路を用いた場合の前進計算、および誤差逆伝搬学習回路について(1)と同様に性能を予測・評価する。

(3)動画像解析への応用におけるディープラーニングの性能を明らかにするため、1枚のFPGA ボードでカメラ 64 台の映像を入力し、移動物体の抽出、追跡、識別を行う Multi-Stream Tracking ハードウェアにおいて、物体識別部分にディープラーニングを適用したときの識別精度および動作速度を調べる。また、識別精度を改善するため種々の拡張改良を行う。

4. 研究成果

(1)SIMD アレイでの並列処理手法としてメモリアクセス競合やバス競合を生じない“同期シフトデータ転送”を適用したディープラーニングの前進計算回路、および誤差逆伝搬学習回路を FPGA 上に作成した。実験回路は CIFAR-10 の物体識別回路とし、FPGA には Stratix V 5SGSMD5 を用いた。図 1 に全体構成を、図 2 に PE (Processing Element) の構成を示す。図 3 に示すように同期シフトデータ転送により畳込み演算を実現した。また、

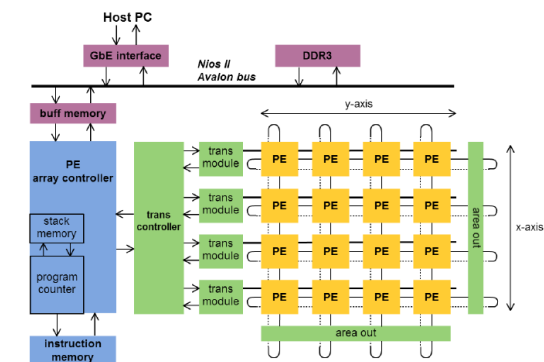


図 1. SIMD アレイ型ディープラーニング回路

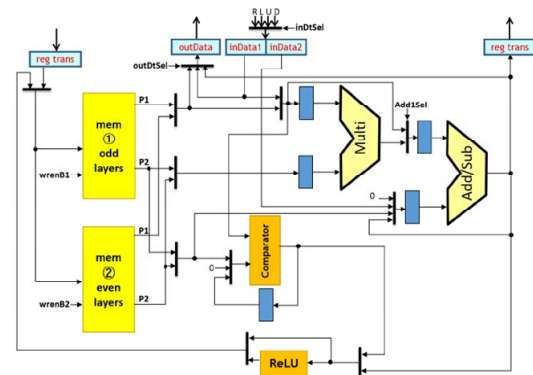


図 2. PE (Processing Element) の構成

誤差逆伝播の高速処理のためのリダクション回路を実装した。用いた FPGA の資源量とクロック速度を表 1 に示す。処理速度は 1 フレーム当たり前進計算で 5.1ms, 誤差逆伝播で 14.1ms であり, 処理速度では GPU を超えることはできなかった。主な原因は, 係数  $W_{ij}$  が極めて多量でメモリに入りきらず外部メモリとの通信待ちになること, 出力段に近づいてノード数が減少したところでは並列度が低下してしまうこと, SIMD アレイ型はディープラーニングの各層を逐次処理することとなるため並列度が上がり GPU とのクロック速度の差を取り戻せなかったためであることを明らかにした(学会発表)。

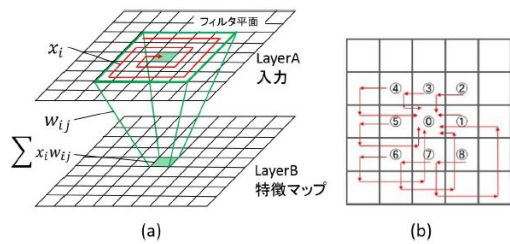


図 3. 同期シフトデータ転送を用いた畳込み

表 1. FPGA の実装結果

デバイス	Altera Stratix V 5SGSMD5K2F40C2N
ALMs	152,900 / 172,600 (89%)
レジスタ	113,227 / 690,400 (16%)
DSP	770 / 1,590 (48%)
M20K ブロックメモリ	1,571 / 2,014 (78%)
動作速度	61.33 MHz

(2)パイプライン形式では, 図 4 の畳込み演算を行うスループット 1 クロックのパイプライン式積和演算器を, 各レイヤに対応して実装する形式を用いた。また, 終段の密結合層の代わりに総和回路を用いることで, 性能

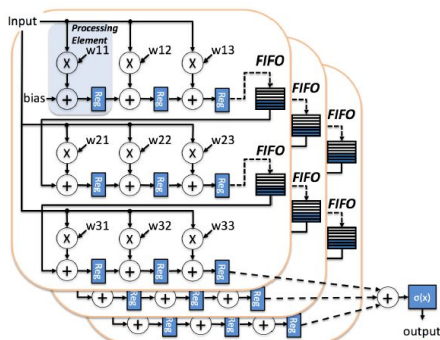


図 4. パイプライン型畳込み演算回路

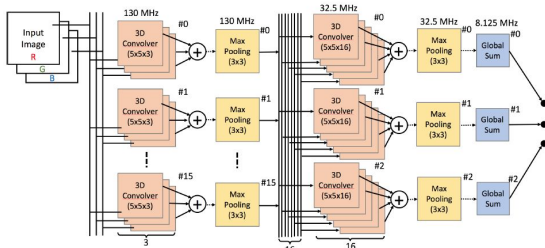


図 5. 移動物体識別用回路の全体構成

を落とさずに資源が大きく節減できることを示した。全体構成を図 5 に示す。この回路ではディープラーニングの全部の層が同時に動作し高い並列度が得られる。一方, 処理できるディープラーニング回路の大きさは実装できる積和演算器の数で制限される。なお, 必要な積和演算器の数は畳込みのフィルタの大きさで決まり, 入力画像サイズには依存しない。画像サイズに影響されるのは FIFO として用いるバッファメモリの長さのみであり, 大きな画像にも適用できる。回路全体のスループットは 1 ピクセル / 1 クロックであり, 320 × 240 の車載カメラ画像中の前走車両の接近識別において識別率 99.0%, 約 1600 フレーム / 秒の処理速度を実現した。演算速度は 409GOPS であり, 入手できる資料から推測したところでは, この時点での FPGA 1 チップ構成の最速であった(学会発表)。更に処理速度を上げるため, 最大値プーリング後の並列度低下による演算効率低下を抑えるように改良を行った(学会発表)。

(3) 1 枚の FPGA ボードで固定監視カメラ 64 台の映像を入力し, 移動物体の抽出, 追跡, 識別を行うことを目指した Multi-Stream Tracking ハードウェアを作成し, 物体識別部分にディープラーニングを適用した。図 6 に機能概要を示す。背景差分による移動物体抽出回路は SIMD アレイで実現されており, ディープラーニング用に位置スケール変換機能, および識別機能を作成した。これらの機能が, 移動物体抽出と同一の SIMD アレイ上で, 特別なハードウェアを付加することなく, 制御プログラムの拡張のみで実現できることを示した。なお, 実シーン中の移動物体を高い精度で識別するには, 抽出された物体をフルスケールに変換する必要がある。ディープラーニングではこれをプーリングにより実現できるとされているが, 多大な資源を要するため FPGA への実装は不可能であり, 図 6 のように PE 間でのピクセルデータの転送により実現した。SIMD アレイ型は本質的にプログラム制御方式のプロセッサであるため, ステップ数を増やせば上記のように任意の処理が

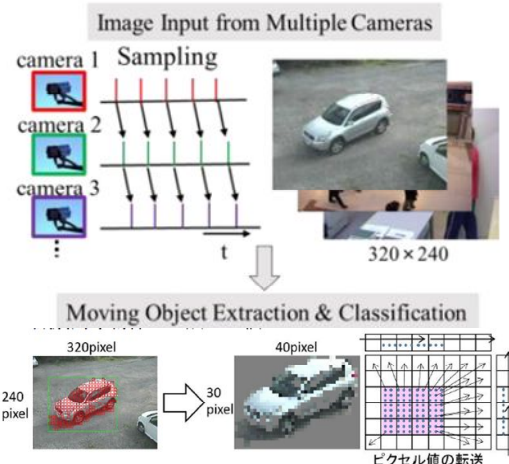


図 6. Multi-Stream Tracking

実現できるが処理時間は長くなる．一方，パイプライン型は畳込み演算のような定型処理は極めて高速に実行できるが，背景差分やフルスケール変換を組み込むことはできない．実験回路の処理速度は GPU には及ばなかったが，人，車両，背景の 3 物体の識別時間 3.7msec，識別率 94%を得た(学会発表， )．

(4)(3)の Multi-Stream Tracking において識別対象の種類に増やすと識別率が大きく低下した．原因は畳込みフィルタに用いるメモリの不足，および人手で作成した学習データの欠陥であった．これを改善するため，学習データからラベルミスやアウトライヤなどの欠陥を取り除く方法を考案した．これにより実データの識別率が 1.5%～3.0%改善できた．次に，SIMD アレイにおいて畳込みフィルタの重みを全 PE に持たせるのではなく，1つの代表 PE からブロードキャストする形式とした．これによりメモリ利用効率が向上し，フィルタの数，サイズを拡張でき，5 物体での識別率を 89.2%から 92.4%に改善した．ディープラーニングの応用により，カメラ 64 台の映像中の移動物体の抽出と識別を 2 回/秒実行する処理速度を実現した．用いた回路の構成を図 7 に示す(学会発表， )．

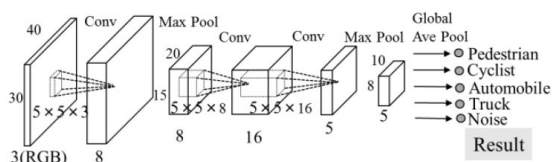


図 7. 移動物体識別回路の構成

(5)本研究で得られたディープラーニングのハードウェアの性能をまとめると表 2 のようになる．前進計算についてはパイプライン型が最も高速であった．SIMD アレイ型は前進計算のみでは低速であったが，背景差分など他の処理と組み合わせると実装するときは，付加回路が不要で同期シフトデータ転送が有効に利用できる．誤差逆伝搬に関しては GPU を超える構成は実現できなかった．GPU でも十分な処理速度では無く，今後は学習用ハードウェアの更なる改良の他，進化的アルゴリズムや粒子群最適化のように誤差逆伝搬を用いない学習手法が重要になると考えられる．

表 2. ディープラーニングのハードウェアの比較

手法	前進識別	誤差逆伝搬
SIMDアレイ型	△他処理との併用に有効	×極めて遅い
パイプライン型	◎高速，規模に制限	×メモリアクセス多大
GPU	○	△

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 7 件)

Shunpei Takaki, and Hitoshi Kitazawa,

Multi-stream Tracking Hardware with DNN Object Classifier using FPGA, International Workshop on Advanced Image Technology 2017 (IWAIT 2017), 2017 年 1 月 8 日, Penang (マレーシア)  
Yo Takeda, and Hitoshi Kitazawa, Filtering Training Data by Using Cross-Validation for Robust Machine Learning, International Workshop on Advanced Image Technology 2017 (IWAIT 2017), 2017 年 1 月 8 日, Penang (マレーシア)．

李寧, 高木俊平, 富岡洋一, 北澤仁志, データフロー構成による高演算効率 DCNN を用いた高速移動物体の識別, 情報処理学会第 78 回全国大会, 2016 年 3 月 10 日, 慶応義塾大学(神奈川県・横浜市)．

高木俊平, 李寧, 北澤仁志, SIMD アレイ型 FPGA を用いた Multi-Stream Tracking における DCNN による物体識別, 情報処理学会第 78 回全国大会, 2016 年 3 月 10 日, 慶応義塾大学(神奈川県・横浜市)．

Ning Li, Shunpei Takaki, Yoichi Tomioka, and Hitoshi Kitazawa, A Multistage Dataflow Implementation of a Deep Convolutional Neural Network Based on FPGA For High-Speed Object Recognition, 2016 IEEE Southwest Symposium on Image Analysis and Interpretation, 2016 年 3 月 6 日, Santa Fe (アメリカ)．

Ryota Takasu, Yoichi Tomioka, Takashi Aoki, and Hitoshi Kitazawa, An FPGA Implementation of Multi-stream Tracking Hardware using 2D SIMD Array, 23rd ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA2015), 2015 年 2 月 23 日, Monterey (アメリカ)．

李寧, 富岡洋一, 北澤仁志, 同期シフトデータ転送に基づく Deep Convolutional Neural Network の FPGA 実装, 電子情報通信学会 VLSI 設計技術研究会, 2015 年 1 月 30 日, 慶応義塾大学(神奈川県・横浜市)．

## 6. 研究組織

### (1) 研究代表者

北澤 仁志 (KITAZAWA, Hitoshi)

東京農工大学・工学院工学研究院・教授

研究者番号： 60345329

### (2) 研究分担者

富岡 洋一 (TOMIOKA, Yoichi)

会津大学・コンピュータ理工学部・准教授

研究者番号： 10574072