

様式 C - 7 - 1

平成30年度科学研究費助成事業（科学研究費補助金）実績報告書（研究実績報告書）

所属研究機関名称		奈良先端科学技術大学院大学	機関番号	14603
研究代表者	部局	先端科学技術研究科		
	職	教授		
	氏名	中島 康彦		

1. 研究種目名 基盤研究(A)(一般) 2. 課題番号 17H00730

3. 研究課題名 エッジコンピューティングを牽引するニアメモリ高効率計算基盤

4. 研究期間 平成29年度～令和2年度 5. 領域番号・区分 -

6. 研究実績の概要

初年度、【1】メモリ主導型CGRAの探索と体系化および【3】CGRA仮想化技術の探索に関して、列方向マルチスレッディング技術、チップ内ローカルリカレントメモリ、メモリとして見せるインタフェースなどの成果を上げた。【2】プログラマビリティと性能チューニングについても開発済CGRAとの機能互換を達成し、既存ツールチェーンを積み上げて無駄の無い環境構築ができた。2年度は、LSI開発コスト削減のための技術開発を含む3つの発展的課題【4】エッジコンピューティング向け省面積CGRAのマルチチップ構成に関する検討と性能評価・分析；【5】エッジコンピューティング向けの共有CNN方式と分割推論モデルの構築と評価；【6】エッジコンピューティング向けのアナログアプロキシメイト演算方式の検討と性能評価；に取り組んだ。2年度の成果として、【4】マルチチップ構成のためのコンパイラ、詳細シミュレータ、マルチチップを構成する基本チップを模倣したFPGAプロトタイプを開発し評価した結果、既存組み込み向けGPUを上回る性能を確認；【5】GPUを用いた分割推論モデルの精度と通信量とのトレードオフ評価を完了；【6】近似計算の基本回路の精度とばらつきの評価を完了した。以上を踏まえ3年度は、【7】スケーラビリティのあるAXIスレーブカスケード型CGRAアクセラレータのFPGA実装および性能評価；【8】認識の確信度も表現可能な変分ベイズ推定手法の実装と評価；【9】よりハードウェアを小型化可能なストカスティック近似計算手法の実装と評価に取り組む。

7. キーワード

アクセラレータ エッジコンピューティング 近似計算

8. 現在までの進捗状況

区分 (1)当初の計画以上に進展している。
理由
初年度(1)列方向マルチスレッディング技術の考案；(2)チップ内ローカルリカレントメモリ(RMM)の導入；(3)ホストに対してメモリとして見せるインタフェースの考案；により、【1】エッジ高機能化に資するメモリ主導型CGRAの探索と体系化；【3】大規模専用ハードウェアを模倣できるCGRA仮想化技術の探索；に関して重要な成果を挙げることができた。また【2】プログラマビリティの飛躍的改善と性能チューニング手法の探索；についても、開発済CGRAとの機能互換を達成したことにより、既存ツールチェーンを積み上げて無駄の無い環境構築ができた。特に、ARM-SoC上に4列64行構成を実装し、実機動作の確認を完了した。また、計画を前倒しして28nmテクノロジーによる評価も完了した。以上、予定を上回る進捗であったため、2年度は、LSI開発コスト削減のための技術開発を含む3つの発展的課題【4】エッジコンピューティング向け省面積CGRAのマルチチップ構成に関する検討と性能評価・分析；【5】エッジコンピューティング向けの共有CNN方式と分割推論モデルの構築と評価；【6】エッジコンピューティング向けのアナログアプロキシメイト演算方式の検討と性能評価；に取り組んだ。2年度の成果として、【4】マルチチップ構成のためのコンパイラ、詳細シミュレータ、マルチチップを構成する基本チップを模倣したFPGAプロトタイプを開発し評価した結果、既存組み込み向けGPUを上回る性能を確認；【5】GPUを用いた分割推論モデルの精度と通信量とのトレードオフ評価を完了；【6】近似計算の基本回路の精度とばらつきの評価を完了した。

2 版

9. 今後の研究の推進方策

以上を踏まえ3年度は、【7】スケーラビリティのあるAXIスレーブカスケード型CGRAアクセラレータのFPGA実装および性能評価；【8】認識の確信度も表現可能な変分ベイズ推定手法の実装と評価；【9】よりハードウェアを小型化可能なストカスティック近似計算手法の実装と評価に取り組む。

10. 研究発表（平成30年度の研究成果）

〔雑誌論文〕 計6件（うち査読付論文 6件 / うち国際共著論文 1件 / うちオープンアクセス 0件）

1. 著者名 Ichikura Takahiro, Yamano Ryusuke, Kikutani Yuma, Zhang Renyuan, Nakashima Yasuhiko	4. 巻 -
2. 論文標題 EMAXVR: A programmable accelerator employing near ALU utilization to DSA	5. 発行年 2018年
3. 雑誌名 2018 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS)	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CoolChips.2018.8373078	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 Uetake Noriyuki, Zhang Renyuan, Nakada Takashi, Nakashima Yasuhiko	4. 巻 -
2. 論文標題 A programmable analog calculation unit for vector computations	5. 発行年 2018年
3. 雑誌名 2018 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS)	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CoolChips.2018.8373080	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 Zhang Renyuan, Uetake Noriyuki, Nakada Takashi, Nakashima Yasuhiko	4. 巻 38
2. 論文標題 Design of Programmable Analog Calculation Unit by Implementing Support Vector Regression for Approximate Computing	5. 発行年 2018年
3. 雑誌名 IEEE Micro	6. 最初と最後の頁 73~82
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/MM.2018.2873953	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 Tati Erlina, Yan Chen, Renyuan Zhang and Yasuhiko Nakashima	4. 巻 -
2. 論文標題 An Efficient Time-based Stochastic Computing Circuitry Employing Neuron-MOS	5. 発行年 2019年
3. 雑誌名 GLSVLSI2019	6. 最初と最後の頁 -
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 Yan CHEN, Jing ZHANG, Yuebing XU, Yingjie ZHANG, Renyuan ZHANG, and Yasuhiko Nakashima	4. 巻 -
2. 論文標題 A ReRAM-based Row-column-oriented Memory Architecture for Convolutional Neural Network	5. 発行年 2019年
3. 雑誌名 IEICE Trans. Electron	6. 最初と最後の頁 -
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 Takahiro ICHIKURA, Yuma KIKUTANI, and Yasuhiko NAKASHIMA	4. 巻 J102-D
2. 論文標題 DSA並みの効率を達成するCNNs拡張機能付きCGRAの提案と評価	5. 発行年 2019年
3. 雑誌名 IEICE Trans., Vol. J102-D, No.07	6. 最初と最後の頁 -
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6件 (うち招待講演 1件/うち国際学会 0件)

1. 発表者名 平賀由利亜, 福岡久和, 三谷剛正, 中田尚, 中島康彦
2. 発表標題 共有 CNN を用いた高効率な分割推論実行モデル
3. 学会等名 xSIG 2018: The 2nd. cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming
4. 発表年 2018年

2 版

1. 発表者名 Tati Erlina, Renyuan Zhang, Yasuhiko Nakashima
2. 発表標題 An Efficient Multiplier Employing Time-Encoded Stochastic Computing Circuit
3. 学会等名 信学技報, vol.118, no.339, CPSY2018-41, pp.47-52
4. 発表年 2018年

1. 発表者名 岩本淳, 菊谷雄真, 中島康彦
2. 発表標題 ユニット内フィードバックによるリニアアレイの多重ループ対応手法
3. 学会等名 信学技報, vol.118, no.339, CPSY2018-40, pp.33-38
4. 発表年 2018年

1. 発表者名 西本宏樹, 中田尚, 中島康彦
2. 発表標題 変分混合ガウスモデルアクセラレータ設計のための変分推論アルゴリズムの解析
3. 学会等名 信学技報, vol.118, no.334, VLD2018-62, pp.155-160
4. 発表年 2018年

1. 発表者名 中島康彦
2. 発表標題 AI専用ハードを横目に見ながらやるべきこと
3. 学会等名 信学技報, vol.118, no.339, CPSY2018-37, pp.3-8 (招待講演)
4. 発表年 2018年

1. 発表者名 Jun IWAMOTO, Yuma KIKUTANI, Renyuan ZHANG, and Yasuhiko NAKASHIM
2. 発表標題 CGRA Cascading for Narrow Memory Bandwidth and Low Cost
3. 学会等名 xSIG 2019: The 3rd. cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming
4. 発表年 2019年

〔図書〕 計0件

1 1. 研究成果による産業財産権の出願・取得状況

〔出願〕 計1件

産業財産権の名称 データ処理装置（高効率アクセラレータ構成方法）	発明者 中島康彦	権利者 同左
産業財産権の種類、番号 特許、PCT/JP2018/018169	出願年 2018年	国内・外国の別 外国

〔取得〕 計0件

1 2. 科研費を使用して開催した国際研究集会

計0件

1 3. 本研究に関連して実施した国際共同研究の実施状況

-

1 4. 備考

NAIST Computing ARchitecture Lab
<http://arch.naist.jp/>