No. A-12

会津大学

# マルチチャネル畳み込み演算の高電力効率・高計算効率アクセラレータ

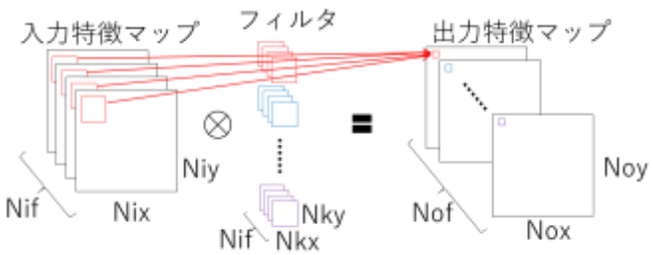
上級准教授　富岡　洋一

## 概要

〇画像中の特定の物体を検出したり、画像を各カテゴリの領域に分割するといった画像認識を高精度に実現する最先端技術のひとつとして、畳み込みニューラルネットワーク（CNN）が活用されています。

〇CNNの推論処理は大量の積和演算を実行する必要があるため、発熱量や消費電力に厳しい制約のあるエッジデバイスでリアルタイム推論処理を実現するために、電力効率の良い専用アクセラレータが必要です。

〇本研究では、CNNの推論計算のほとんどの割合を占めるマルチチャネル畳み込み計算を、高並列かつ高電力効率で実現する新しいアルゴリズムとその高速化のための回路を提案しました。




## 実用化の可能性

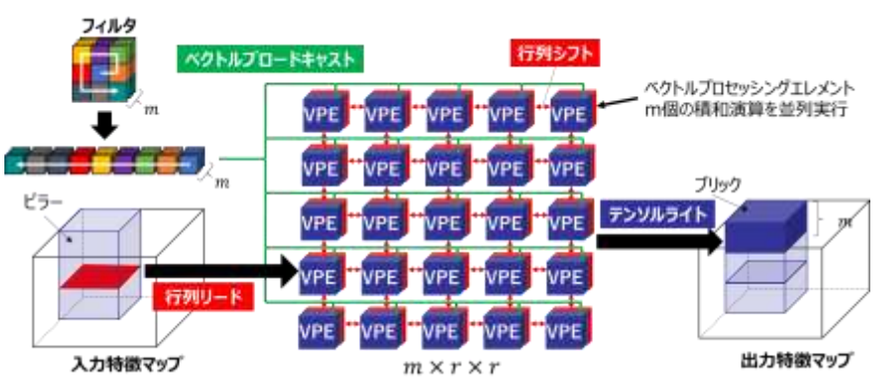
〇本技術は最先端のCNNを用いた画像認識を高電力効率かつ高速に実行するための技術であり、エッジ環境で画像認識を必要とする多様なアプリケーションへの応用が考えられます。例えば、応用先として、人工知能を搭載した自動運転車両やスマート監視カメラ等のシステムオンチップ(SoC) を検討しています。

## UBICからのメッセージ

第3次AIブームの火付け役となったCNNですが、その処理の重たさから従来はクラウドサーバ等での処理が主流でした。昨今ではセンサ技術の向上とともにエッジ側での処理が注目を浴びるようになり、処理の軽量化や省電力化が重要な鍵を握っています。本技術はCNNの中で最も計算負荷の高い部分を高度に並列化するとともに、デバイス内でのデータ転送を局所化することにより、処理の高速化と省電力化を図っています。ハード・ソフト両面からのこのような技術により、高度なAI処理もエッジ側で実現できるようになりつつあります。

## 研究概要図

Broadcast-Compute-Shift畳み込み演算方式




特徴

- 出力特徴マップの画素、チャネル、入力特徴マップのチャネルに関するマルチチャネル畳み込み演算の潜在的な並列性を引き出し、高並列度の計算を実現
- 低エネルギーコストの近傍データ転送を活用し、レジスタ上でデータを再利用することで高エネルギーコストのメモリアクセスを削減し、低消費電力化

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）「高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発／革新的AIエッジコンピューティング技術の開発／ソフトテンソルプロセッサによる超広範囲センシングAIエッジ技術の研究開発」の委託業務の結果得られたものです。

**高度なAI処理をエッジにおいて高効率かつ省電力で実現する**