論 文 内 容 の 要 旨

博士論文題目

CGRA 型プログラマブルアクセラレータの画像認識応用に関する研究

氏　　名　一倉　孝弘

高精度な画像認識が実現可能なConvolutional Neural Networks (CNNs) を組み込み機器に搭載することが期待されている．ただし，CNNs では多くの計算が必要であり，従来の機器に搭載されているCPU やGPU では時間がかかり過ぎる．これに対し，CNNs の計算に特化したアクセラレータ(Domain Specific Accelerators (DSAs)) が多数報告されている．しかし，微細化によるコストダウンの恩恵がなくなった現状においてLSI 開発費の回収を図るには，様々なメモリ参照パターンに対応できるアクセラレータが望ましい．そこで，本論文では，様々な離散ステンシル計算を効率よく計算することを目的に開発したEMAXV を改良したEMAXVR を提案する．EMAXVR は，ローカルメモリを備えるCGRA(Coarse Grained Reconfigurable Architecture) を採用したプログラマブルアクセラレータのEMAXV に対して，CNNs の計算を効率化するために，1) 多重ループ制御機構，2) マルチディレクションブロードキャストバスと3) スクラッチパッドメモリを搭載した．その結果，CNNs のAlexNet とVGG16 の畳み込み層の計算を行った時に，ARM CoretexA9 に比べて60 倍以上，Vivante GC2000+に比べて約20 倍高速に計算できることを確認した．さらに，DSAs と比較して，計算性能指標である演算器利用率が18 %低いものの，消費電力指標である計算回数に対するDRAM アクセス量が，同等にまで迫れることを確認した．

氏　名	一　倉　　孝　宏

（論文審査結果の要旨）（Ａ４　１枚　１、２００字程度）

高速・高精度画像認識のために，組み込み機器にConvolutional Neural Networks (CNNs) を搭載することが求められている．しかしながら，CNNsの計算量は膨大であり，従来機器に搭載さるCPUやGPUでは実用的な時間で計算することができない．そこでCNNsの計算に特化したアクセラレータが多数提案されているが，微細化によるコストダウンの恩恵がなくなった現状においてLSI 開発費を回収するためには，様々なメモリ参照パタンに対応できるアクセラレータが望ましい．

本研究は，CGRA 型プログラマブルアクセラレータの画像認識応用に関するものである．離散ステンシル計算やLight-Fieldのレンダリングで高い性能を発揮しているCGRA型プログラマブルアクセラレータをCNNs向けに改良することで，CGRA型プログラマブルアクセラレータの適応範囲をCNNsにも拡大する．最終試験では，以下の主要な成果について報告があった．

【１】CNNs の計算時間の大部分(90 %以上) を占める多重ループの畳み込み演算を効率よく計算するために多重ループの計算を 1 回の起動で実行する仕組みを提案し，シングルループの計算と比較して畳み込み演算の計算速度を 20 倍高速化した．さらに，電力消費が大きな外部メモリへのアクセスを抑制するために，マルチディレクションブロードキャストによるデータ転送回数の低減と，スクラッチパッドメモリの追加により，外部メモリ通信量を従来の 15 %に低減した．

【２】カーネルサイズなどの各種パラメータが異なる畳み込み演算において，アクセラレータの演算器利用率を最大化する命令写像を示し，従来の命令マップより計算速度を1.5 倍高速化した．

【３】クロックサイクルアキュレートのソフトウェアシミュレータにより性能評価を行い，組み込みCPU やGPU より高い性能を実現することを示した．さらに，プログラマビリティを有したアクセラレータでありながら，CNNs の計算に特化したアクセラレータに対して，演算器利用率は少し劣るものの，外部メモリの通信量を同等に迫れることを示した．

本研究で，CGRA型プログラマブルアクセラレータをCNNsの計算に利用するための構成を確認したことで，複数の用途で使用できるアクセラレータでありながら，期待されている組み込み機器にCNNsを利用した高精度な画像認識機能の搭載を実現できると予想される．

以上，本論文は学術上，実際上寄与するところが少なくない．よって，本論文は博士(工学)の学位論文として価値あるものと認める．