

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	コンピューティング・アーキテクチャ (中島 康彦 (教授))		
学籍番号	2011401	提出日	令和 4年 1月 13日
学生氏名	赤部 知也		
論文題目	確率的計算基盤のAIアプリケーションによる評価		
要旨			
情報化社会の進展はエネルギー消費に影響を及ぼしており、演算のエネルギー効率を向上させる新しい手法が必要とされている。ニューラルネットワークの分野でも電力問題は無視できない。画像認識AIなどで広く使われている深層畳み込みニューラルネットワーク(Deep Convolutional Neural Networks: 以下, DCNNs)は膨大なパラメータを高速に演算する必要がある。また、全結合層でのパラメータの多さは巨大なメモリバンド幅を必要とする。現在はDCNNsの演算はGPUなどの広いメモリバンド幅を持つアクセラレータでSIMDを使って実現している。しかし、近年AIが求められる領域はIoT機器や車載など多岐に及ぶ。その中にはバッテリー駆動であったり、スペースに限りがあったりするため、消費電力の増加や大型化しやすい広いメモリバンド幅を持つアクセラレータは使用が困難である。そこで小面積、低消費電力である粗粒度再構成型アレイ(Coarse Grained Reconfigurable Arrays: 以下, CGRA)型アクセラレータIMAX2を開発してきた。これは演算器の近くにローカルメモリを置くことで無駄なデータの移動を削減し、狭いメモリバンド幅を最大限再利用することで高い演算効率を実現している。 最近の研究によると、ニューラルネットワークでは高い演算の精度は必要ないとされる。このような状況の中で注目を集めているのが確率を使って演算を行う確率的コンピューティングである。確率を使う近似計算のため、厳密計算と比べると精度は低下するものの、乗算をANDゲート1個で実現できることからハードウェアリソースの削減と周波数性能の向上が可能である。そのため、精度の低下がある程度許容できるニューラルネットワークの推論とは相性が良い。 本研究では、確率的積和演算器をIMAX2に組み込み、より高効率なデバイスを目指した。導入準備として行った行列積演算フローの変形により、CGRAの起動回数は0.2%に削減し、画像認識の実行速度は46倍に向上した。また、確率的積和演算器は従来の積和演算器と比較して、高い周波数性能を示した。			