

革新的コンピューティング技術の開拓
2019 年度採択研究者

2019 年度 実績報告書

陳 オリビア

横浜国立大学先端科学高等研究院
特任助教

アルゴリズム・ソフトウェア・ハードウェアの融合による超低電力ニューラルネットワーク
の構築

§ 1. 研究成果の概要

ここ数年で深層学習は沢山の注目を集め、様々なアプリケーションに応用されながら劇的な精度向上を果たしている。これらの多くは数百万から数十億の素子で構成され、複数の GPU と CPU の組み合わせによる超高速計算能力が重要な役割を果たしている。このような計算には大きな消費電力がともない、莫大なコストを必要とする。そこで本研究は、アルゴリズム、ソフトウェア、及びハードウェアを統合した超低電力ニューラルネットワーク

(Neural Network: NN) を提案する。初年度ではお主に、①アルゴリズムの面において、手書き数字画像データセット MNIST に向けた LeNET5 モデルの重み圧縮を行った。圧縮したモデルは、推論の精度を 99.03% に維持しながら従来より 600 倍以上軽量化することができた。さらに、②ソフトウェアについて、従来の遺伝的アルゴリズムを用いた配置配線の結果と比べ、最大 43% の面積削減かつ 100 倍以上の計算時間の短縮を達成した。

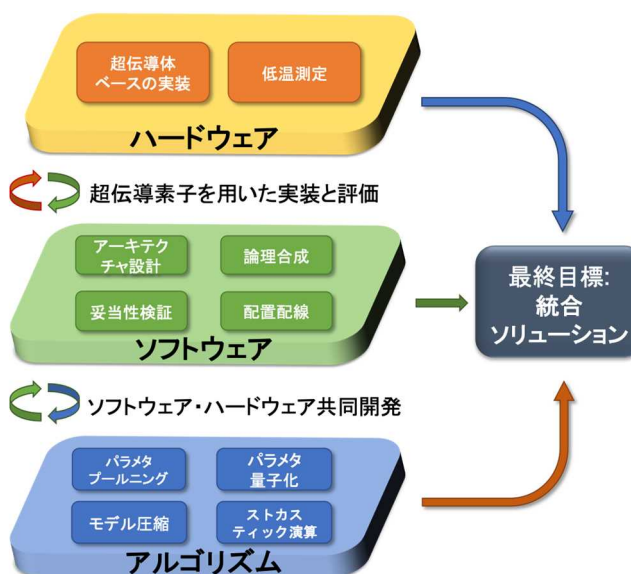


図1 提案するニューラルネットワークの全体像。