令和元年度 情報工学コース卒業研究報告要旨

枝廣・本田 研究室	氏　　名	下 平　健 太
卒 業 研 究 題 目	小規模組込みシステム向けDNNフレームワークの比較評価	

近年，多層のニューラルネットワーク (以下DNN) に代表される深層学習は，技術の発展が進み画像認識や音声認識および自然言語処理などの幅広い分野において使われ注目を集めている．組込みシステムにおいても機器の故障検知などで深層学習の推論の利用が求められている．現状の深層学習を実装するうえで用いられるフレームワーク（以下DNNフレームワーク）や開発環境は汎用PC向けに用意されていることが多く，学習と推論の両方にGPUやFPGAを用いることが一般的である．しかし，小規模な組込みシステムでは消費電力やメモリ使用量などのリソースの制約があるため，多くのリソースが必要なこれらの演算処理装置は使用できず，小規模組込みシステム向けのCPUであるマイコンを使用することになる．マイコンはメモリサイズや性能が低く，例えば最新のマイコンであるSTM社のSTM32F7のメモリサイズは最大2MBで，クロック周波数は216MHzである．マイコンにおける深層学習の推論を実現するためにはC言語での実装が必要であるため，前述した汎用PC向けDNNフレームワークを直接用いて実装することができない．さらに，性能の低いマイコンを用いるためには，メモリ使用量や計算量の削減が可能なパラメータの量子化やSIMD命令を用いた計算が必要である．以上のような観点から，小規模組込みシステムのDNNの実装は工数がかかるという問題がある．

この問題に対して，複数のマイコンベンダーから小規模組込みシステム向けDNNフレームワークがリリースされている．しかしながら，どのフレームワークもリリースから日が浅いため，機能や性能の比較評価が十分なされていない．

本研究では，複数の小規模組込みシステム向けDNNフレームワークの機能を調査し比較評価した．具体的には，STM社が提供しているX-CUBE-AIと，NXP社が提供しているeIQの2つについて調査し比較評価した．まず，小規模組込みシステムにおけるDNNの実装に必要な観点について整理した．次にそれぞれのフレームワークの特徴や設計フローについて調べ，最後に同一のネットワークに対する性能を比較し考察した．まず機能としてどちらのフレームワークも代表的な汎用PC向けDNNフレームワークで作成したネットワークモデルを読み込むことができる．さらにeIQでは，マイコンの持つSIMD命令に最適化されたDNNライブラリであるCMSIS-NNを用いてネットワーク構造を記述する実装にも対応している．X-CUBE-AIは，量子化を自動で行うことや，推論器のソースコードが自動生成されることなどから実装までの工数が少ないことに対し，eIQは量子化においてユーザーが別途ツールを用意する必要があることや，最終的なソースコードをユーザーが直接記述しなければならない．

同一のネットワークに対する推論処理の性能評価では，ネットワークモデルを読み込ませて実装した場合X-CUBE-AIは216MHzのマイコンで283ms，eIQは600MHzのマイコンで72msであり，eIQのCMSIS-NNで実装した場合96msだった．また，既存研究では216MHzのマイコンで同一のネットワークに対する推論をCMSIS-NNで実装した場合は99msと報告されており，eIQの結果はマイコンの性能に対し非常に低速であると分かった．この問題の原因を特定するために，CMSIS-NNとマイコンのアーキテクチャを調査した．eIQで用いたマイコンはCPUがROMに対して高速であるため，マイコンとしては大規模な32Kbytesのキャッシュを搭載している．一方、CMSIS-NNは既存のマイコンの多くがキャッシュを持たないことからSIMD演算の効率化のみを考慮しており，キャッシュの利用効率の悪いアルゴリズムとなっていたことが原因であった．