

氏名	永井 晃大	指導教員	藤枝 直輝
題目	PYNQ-Z1 ボードにおける FINN-L を利用した文字認識システムの開発		

1 はじめに

ニューラルネットワーク技術は、近年めざましい進展を遂げている。人間の脳の神経細胞ネットワークを模倣したこの技術は、複雑なパターンの学習やデータの分類・予測に広く用いられている。その応用例は自動運転、医療画像解析、自然言語処理など、さまざまな分野に見られ、社会に大きな影響を与えている。

ニューラルネットワークの実用化には、高い計算能力と効率的な処理が不可欠である。特に、リアルタイム処理が求められる分野では、大量のメモリ使用や高い消費電力が課題となっている。これに対し、FPGA (Field-Programmable Gate Array) は、柔軟なプログラミング可能性と効率的なハードウェア設計により、ニューラルネットワークの推論処理において有効な解決策となりうる。

2 目的

本研究の主な目的は、ニューラルネットワークの一種である LSTM (Long-Short Term Memory) を用いた文字認識システムを、FPGA を使って開発することである。近年注目を集める Transformer アーキテクチャと比較して、LSTM はモデルサイズが小さく、FPGA への実装に適している。本研究では、FINN-L [1] という FINN [2] の拡張を基にして、文字認識システムの柔軟性を高め、実際のタスクに適応させるための改良を行う。

3 先行研究

FINN は、FPGA を活用したニューラルネットワークの量子化技術をサポートするフレームワークであり、特に入出力、特徴量、重みのデータを少ビット数で処理することに特化している。その拡張である FINN-L は、FPGA 上での可変精度 LSTM ネットワークを実装することに焦点を当てている。これらのフレームワークは、ハードウェアの容量や消費電力を削減しつつ、高速な計算を実現するために重要な技術である。

しかしながら FINN-L には、実際の応用例に関する詳細な記述が不足している。本研究ではこのギャップを埋めることを目的としている。

4 研究方法の概要

本研究で開発する文字認識システムは、複数行の OCR タスクを対象とし、入力として与えられた複数行の文章画像から各行に書かれた文字列を正確に出力することを目標とする。本研究で FINN-L に対して施した改良は以下の3つである。

第1の改良は、複数行の文章画像を行ごとに分割する、文字列検出処理の開発である。この処理は、Haar-Like 特徴に基づくアルゴリズムを使用して文書画像を縦方向に走査し、文字列と非文字列の部分を区別する。この段階では、ソフトウェアベースの実装を行い、抽出された文字列画像を後続の処理に送る。

第2の改良は、LSTM 推論器のインタフェースの変更である。先行研究で使用された FINN-L の評価システムでは、アクセラレータは AXI インタフェースを介して能動的にデータをやり取りしていたが、本研究ではデータのやり取りを受動的な AXI Stream インタフェースに変更する。また、制御方法も FPGA システムで広く用いられる Python による方法に変更し、将来的な拡張や改良に対する柔軟性を高める。

第3の改良は、タスクに適合するデータセットの作成である。英語のテキスト画像を含む独自のデータセットを「不思議の国のアリス」の文章を使用して生成し、500枚の画像に対して文字サイズのランダム化や空白の挿入などのオーグメンテーションを施す。学習には、FINN-L のトレーニングパイプラインである pytorch-ocr を使用する。

これらの手順を通じて、PYNQ-Z1 上で文字列認識システムを構築し、実際の文章画像に対して OCR を行った。その結果、97.54% の文字列認識率を達成した。

参考文献

- [1] V. Rybalkin et al., "FINN-L: Library Extensions and Design Trade-off Analysis for Variable Precision LSTM Networks on FPGAs, in FPL 2018, pp. 890-897, 2018.
- [2] Y. Umuroglu et al., "FINN: A Framework for Fast, Scalable Binarized NN Inference," in FPGA '17, pp. 65-74, 2017.