

どんな研究？

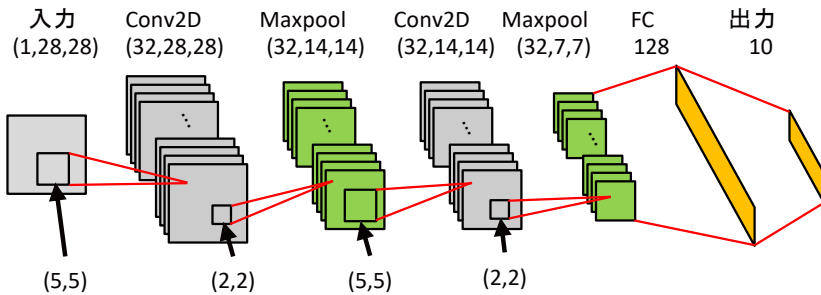
IoTデバイス等での簡易で高速な処理を目的として、機械学習のハードウェア化が注目されています。ニューラルネットワークの推論においては、2値データを用いた高速・省資源処理方式が提案されているが、ネットワーク自体の学習はほとんどの場合クラウド側で行われています。本研究では、オンサイトの学習を容易にするニューラルネットワークの学習手法に取り組んでいます。

何がわかる？

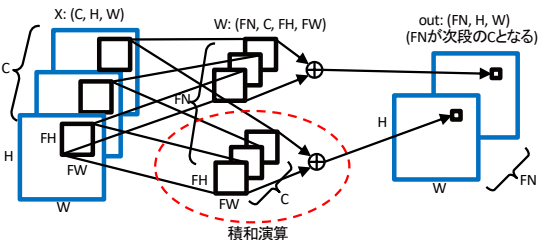
ニューラルネットワークとは？
 その基本的な仕組みとは？
 学習とは？
 ハードウェア実装することのメリットは？
 問題点は？
 解決のアプローチは？

ニューラルネットワークの推論と学習

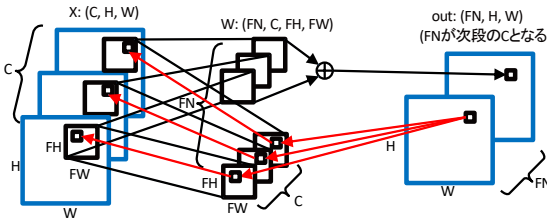
◆ニューラルネットワークとは(Convolutional Neural Networkの例)



◆ Forward演算 (Conv2D層)



◆ Backward演算 (Conv2D層)



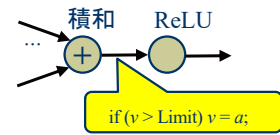
研究内容

◆ 解決のアプローチ

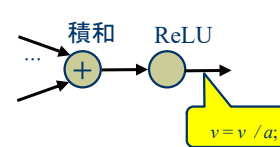
- 整数演算を前提とした学習方式を提案し、どれほど推論の正解率が下がるかを評価
- ハードウェア量(使用リソース)削減を評価

◆ Forward演算

- 入力・重みが整数: すべて整数演算で可能
- 値の圧縮手法
- クリップ処理



● スケールダウン処理



◆ 整数化(MNISTデータセットでの例)

- 入力: 閾値を200程度として2値化
- 重み初期値: $\text{int}(\text{整数} \cdot w_i(l))$ に平均0, 分散1の正規分布を持つ乱数を乗じた数
- 重み更新の粒度に影響

◆ Backward演算

- 出力層のノード値が正解に向かうための変位
- Rewardと呼ぶ: 誤差の負数に相当

