

Layer Fusion を導入した MobileNets

日大生産工 ○山浦 和佳 日大生産工 山内 ゆかり

1. まえがき

深層畳み込みニューラルネットワーク (deep convolutional neural network) 1) が画像認識の精度を競う ImageNet Large Scale Visual Recognition Challenge (ILSVRC) の2012年で他と大きな差をつけて優勝してから、ディープラーニングの主流となっている。それ以降、日々改良がなされ、精度が上がるとともに、それと比例してデータ量や計算コスト、計算時間などが増大している。精度が向上することは必ずしもネットワークのサイズと計算速度を効率的に利用しているとは限らない。近年スマートフォンやノートパソコンなどが普及している中でニューラルネットワークを動作させるには、計算資源が限られた中で精度とコストの兼ね合いの良くなるようにネットワークの構成を見直し、モバイル端末でも使えるように、遅延の少ない軽いネットワークのモデルが必要とされている。

Andrew G. Howard らは畳み込みニューラルネットワークのネットワークの分離 2) を提案し、約1/8~1/9の計算量の削減に成功している。また、James O' Neill らはニューラルネットワークの内部変数が多いことに着目し、層を融合 3) することにより、精度誤差2%以内で3倍の圧縮率を実現した。

本研究では、軽いニューラルネットワークである MobileNets に層融合を導入することによる計算量の削減を提案し、CIFAR-10 における計算機実験により提案手法と精度、計算量を比較し、報告する。

2. 従来研究

2.1 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク 4) は、主に畳み込み層、プーリング層、全結合層に分かれており、畳み込み層では画像と重みを掛け合わせ、画像の特徴抽出を行う。重みと画像の特徴が一致すると値が大きくなるため特徴が伝播されていく。プーリング層ではあまり情報を失わずにデータ量を減らし、位置のずれに関係なく学習する。また、重みが存在しないため計算が軽い。全結合層ではこれらの特徴を受け取り、クラス分類や画像識別を行う。

2.2 MobileNets

MobileNets では畳み込みニューラルネットワークの構成を変更し、計算量を減らす工夫が

なされている。通常、畳み込みニューラルネットワークはチャンネル方向と空間方向に対して同時に畳み込みを行う。入力チャンネル数 N 、カーネルサイズ K 、出力チャンネル数 M の場合通常の畳み込みニューラルネットワークの計算量は

$$K * K * N * M \quad (1)$$

となる。

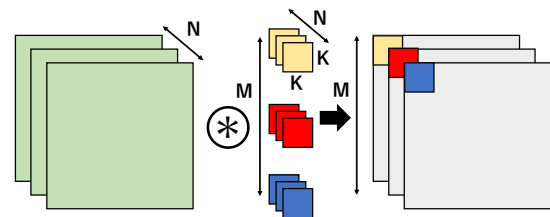


Figure 1 通常の畳み込みニューラルネットワーク

従来研究ではこの要素を空間方向とチャンネル方向に分解することによって大幅な計算量の削減を図っている。通常の畳み込みはすべての計算量が掛け合わせであるのに対し、分離することによって計算量が足し合わせになる。これによって精度を下げずに大幅な計算量の削減に成功している。Depthwise Convolution では入力チャンネル数と出力チャンネル数が等しくチャンネル方向の足し合わせは行わない。通常の畳み込みと違い入力チャンネル数 N が出力チャンネル数 M と等しくなるため計算量は

$$K * K * N \quad (2)$$

Pointwise Convolution ではカーネルサイズが1の畳み込みを行い、チャンネル方向の計算を行う。カーネルサイズ K が1になるため計算量は

$$N * M \quad (3)$$

チャンネル方向と空間方向に分離して畳み込みを行うことによって、計算量は

$$K * K * N + N * M \quad (4)$$

となり約1/8~1/9に演算量の削減になっている。

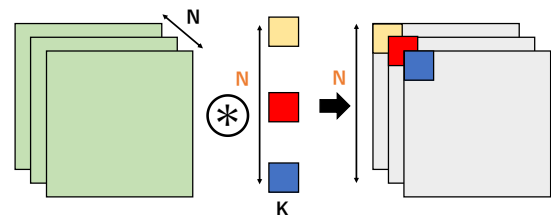


Figure 2 空間方向の畳み込み

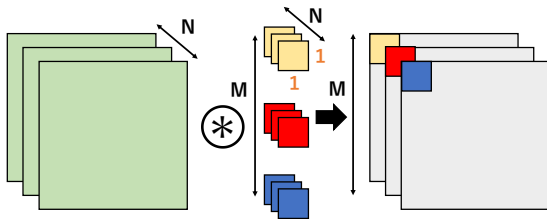


Figure 3 チャネル方向の畳み込み

3. 提案手法

本研究ではニューラルネットワーク、とくに畳み込みニューラルネットワークの内部変数の数が多いことに着目し、層融合を提案する。層融合とは複数の計算を一つにまとめ、内部変数の数を減らすことができる。入力が x 、 n 個の層を経て得る出力 $f(x)$ は

$$f(x) = W_n g(\dots g(W_2 g(W_1 x + b_1) + b_2) \dots) + b_n \quad (5)$$

となる。

従来研究のDepthwise ConvolutionからPointwise Convolutionまでの流れを層融合しその間の内部変数をなくすことによって精度を下げることなく、計算量を削減する。手書き文字データセットMNIST 5)の場合、Segmentとしてみなした内部で層融合を行う。

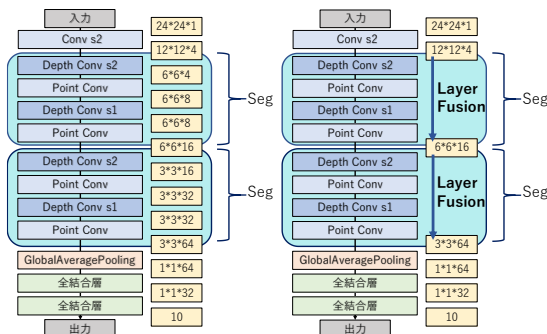


Figure 4 従来研究と提案手法

4. 実験および検討

手書き文字データセットMNISTおよびカラー画像で画像サイズが 32×32 のCIFAR-10 6)を用いて精度、処理時間をそれぞれ測定し、従来研究と提案手法、指標として通常の畳み込みニューラルネットワークで比較する。MNIST及びCIFAR-10で精度を下げることなく、どれだけ計算量を減らすことができるかデータを得る。



Figure 5 MNIST と CIFAR-10

5. まとめ

畳み込みニューラルネットワークがディープラーニングの主流となっている。日々精度は向上しているが、必ずしもネットワークのサイズと計算速度を効率的に利用できているとは限らない。畳み込みニューラルネットワークをスマートフォンやノートパソコンなどのモバイル端末で利用するために、限られた計算資源の中で、遅延の少ない軽いニューラルネットワークのモデルが必要となっている。

本研究では、軽いニューラルネットワークであるMobileNetsに層融合を導入することによる計算量の削減を提案し、CIFAR-10における計算機実験により提案手法と精度、計算量を比較した。

参考文献

- 1) Alex Krizhevsky , Ilya Sutskever , Geoffrey E. Hinton “ImageNet Classification with Deep Convolutional Neural Networks”(2012)
- 2) Andrew G.Howard , Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam” MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications ”(2017)
- 3) James O’Neill, Greg Ver Steeg & Aram Galstyan “Compressing Deep Neural Networks via Layer Fusion ”(2020)
- 4) Yann LeCun ,Leon Bottou, Yoshua Bengio, Patrick Haffner” GradientBased Learning Applied to Document Recognition”(1998)
- 5) “THE MNIST DATABASE of handwritten digits”, <http://yann.lecun.com/exdb/mnist/>
- 6) “CIFAR-10 dataset”, <https://www.cs.toronto.edu/~kriz/cifar.html>