

動的グループ化畳み込みニューラルネットワークにおける 特徴割り当て最適化

日大生産工 ○本橋 卓也 日大生産工 山内 ゆかり

1. まえがき

近年ハードウェアが発展していくにつれて、画像データが身近なものになった。そして、世界中で画像が蓄積されていく中で、蓄積されたデータを有効活用しようという手段の一つとして、機械学習による画像認識がある。

その画像認識の手法の一つとして、畳み込みニューラルネットワーク(Convolutinal Neural networks:CNN) [1]という手法がある。

従来の畳み込みニューラルネットワーク(CNN)は多くの冗長性が有している。その冗長性を削ることで、精度を落とさずに計算量の削減を行うことができると考えられている。特に従来のCNNでは、畳み込み層の計算量が多くを占めており、CNNの計算時間削減のためにはこの畳み込み層の計算量の削減が重要であると考えられている。この畳み込み層の計算量削減のために、グループ化畳み込みニューラルネットワーク[2]が提案された。そして、そのグループ化畳み込みニューラルネットワークの精度向上を目的として、学習時にグループ分けを学習することで動的なグループ化を行うことができる機能[3](Fully Learnable Group Convolution : FLGC)が提案された。しかし、動的なグループ分けを行っても、グループ数が増えると精度が落ちるというグループ化畳み込みの問題点がある。

本研究では、動的なグループ分けを行う際に重要度の高い特徴マップを複数グループに振り分け、特徴マップを有効活用することでグループ数が増えても、画像の認識率が維持されることを目指す。

2. 従来手法

2-1. ResNet

CNNの構造上の制約になっていた、層を深くしすぎるとモデルの劣化が生じるという点を大幅に改善したネットワーク構造[4]である。基本的な構造は入力層、畳み込み層、プーリング層、Residual Block(以下残差ブロック)、Global Average Pooling、全結合層によってネットワークが構成されている。

2-1-1. 残差ブロック層

複数の畳み込み層、バッチ正規化と活性化関数、ショートカット接続(shortcut connection)を組み合わせて出力を算出する層である。標準的なブロック層Fig.1(a)と、ボトルネック構造になっているブロック層Fig.1(b)の二種類のブロック層が存在する。標準的なブロック層ではバッチ正規化とセットになった2層の畳み込み層を持ち、活性化関数はReLUを用いる。ボトルネック構造になっているブロック層では、バッチ正規化とセットになった3層の畳み込み層を持ち、活性化関数はReLUを用いる。また、使用されるブロック層の種類はネットワークの深さ次第で決定され、50層以降のネットワークではボトルネック構造のブロック層を使用する。

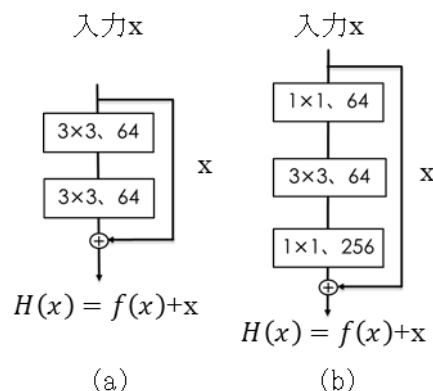


Fig.1 残差ブロック構造

また、一般的なCNNとは異なり、残差ブロック内で行われる学習には、1ブロックごとに算出されたResidual(残差)を学習する。

(1)の式は入力 x に対し、畳み込み処理 $F(x)$ を用いて、出力 $H(x)$ を表している。

$$H(x) = F(x) + x \quad (1)$$

2-1-2. Global Average Pooling層

残差ブロック層で算出されたそれぞれの特徴マップごとに、Global Average Poolingを行い、各特徴マップごとの画素の平均を算出する層である。

2-2. ResNeXt

ResNetの改良手法の1つであり、ResNetの残差ブロック層 $F(x)$ において、入力 x を複数グループ C にグループ分けを行い、それぞれのグループを同一の構造を持つネットワーク T_i で処理させた後に、その総和を取るグループ化畳み込みを導入した手法である。

(2)の式は入力 x を C グループに分け、それぞれをネットワーク T_i 上で処理を行い、その総和を出力 $F(x)$ として表している。

$$F(x) = \sum_{i=1}^C T_i(x) \quad (2)$$

2-3. Fully Learnable Group Convolution

従来のグループ化畳み込み処理では、入力、フィルターのグループ分けをFig.2(a)のように、ネットワークの構築時に一意に設定していた。

FLGCは、入力、フィルターのグループ分けを学習するパラメータを取り入れ、Fig.2(b)のように学習時にグループ分けを学習することで、推論時に無駄の少ないグループ分けを動的に行うことができるモジュールである。

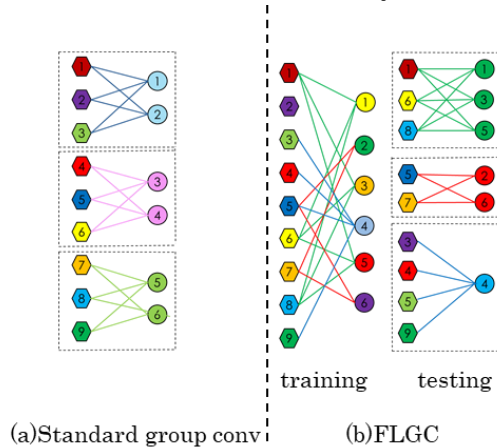


Fig.2 グループ畳み込み処理の概要

3. 提案手法

本研究では画像の認識率の向上を目的として重要度の高い特徴マップを複製して学習させるFLGCを提案する。

FLGCはグループ分けを決定するとき、学習されたパラメータによって決定されている。しかしこの方法では学習の結果次第で、複数のグループにマッチしやすい重要度の高い特徴マップが生じると考えられる。

そこで一度グループに振り分けた特徴マップの中でも、特に重要度の高い特徴マップを複製し他のグループにも振り分けることで、効率よく特徴量を抽出することで、画像の認識率の向上を目指す。

手法として、一度全ての特徴マップをパラメータに従って各グループに振り分ける。そして、

各グループ内でグループとのマッチングしている特徴マップの上から三位までを算出し、それらが持つ現グループとのパラメータの平均を算出する。そして、各グループ内に存在していない特徴マップの中で、グループとのパラメータが各グループが持つ上位のパラメータの平均よりも高ければ、その特徴マップを複製し該当するグループに追加する。そうすることで、重要度の高い特徴マップから有効な特徴の抽出を行うことができ、今まで活用されなかった有効な特徴を活用できるようになると考えられる。

4. 実験

ResNet50にFLGCを埋め込んだネットワークと、ResNet50にFLGCと提案手法を埋め込んだネットワークに対して、CIFAR-10を用いて画像認識の精度を測定し、提案手法の効果を確認する。CIFAR-10は32*32ピクセルの解像度を持ち、10のクラスと60000の画像から構成されたデータセットである。また、通常ResNet50は224*224の解像度を持つ画像に使用されているため、ネットワーク構造の調整として、7*7畳み込み層を3*3畳み込み層に置き換えて実験を行う。

5. まとめ

動的にグループ分けを行うグループ化畳み込みNNが持つ、グループ数が増えると精度が落ちるといった問題点に対し、特徴マップの有効活用を行うグループ化畳み込みNNを提案し、グループ数が増えても、画像の認識率を維持することを試みた。

【参考文献】

- [1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel. "Backpropagation applied to handwritten zip code recognition". *Neural Computation* 1 :541-551 1989.
- [2] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, Kaiming He, UC San Diego, Facebook AI Research, "Aggregated Residual Transformations for Deep Neural Networks", Accepted to CVPR 2017.
- [3] Xijun Wang, Meina Kan, Shiguang Shan, Xilin Chen, "Fully Learnable Group Convolution for Acceleration of Deep Neural Networks", Accepted by CVPR 2019
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research, "Deep Residual Learning for Image Recognition"