

# GELU と TrivialAugment を導入した ResNet-RS の提案

日大生産工 ○佐藤 陸真 日大生産工 山内 ゆかり

## 1. まえがき

近年の機械学習において大きな影響を与えているモデルにResNet[1]がある。このモデルには、今まで問題とされていた勾配消失問題を解決するskip connectionが提案された。結果として今までより深層の学習で精度を上げることに成功した。Belloらは既存のResNetのアーキテクチャ、学習手法、スケールアップを改善しより精度を高め、同程度の精度を出すEfficientNet[2]よりも2.1~3.3倍計算速度が速いResNet-RS[3]を提案した。しかし、RandAugment[4]を用いるため、精度を出すためには事前学習に90エポック以上が必要とされているためパラメータの調整は簡単ではない。また、90エポック、350エポックでは350エポックのほうが正規化の時精度が高く、収束が遅いことが考えられる。

本研究では、ResNet-RSのデータ拡張手法をパラメータがないTrivialAugment[5]、活性化関数をより収束の早いGELU[6]に変更することで計算量の削減を目指す。

## 2. 従来研究

### 2.1 ResNet

ResNet[1]はCNNの問題としてあげられていた勾配消失問題を回避するために誤差をうまく伝搬するskip connectionを追加した。

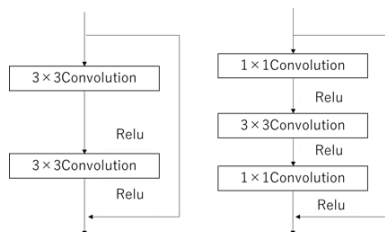


Figure 1 skip connection

skip connectionには、二つの方法があり左側が building block、右側が bottleneck building blockと呼ばれている。Figure 1のようにいくつかの畳み込み層をスキップし入力が入力と合わされるルートが追加された。残差ブロックでの式を以下に示す。

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

$F(x)$ が通常の畳み込みで渡される値で $x$ がスキップ接続で渡される値である。このように恒

等写像で $x$ を渡すことで勾配の消失を防いでいる。Figure 2はダウンスケール時の skip connectionである。

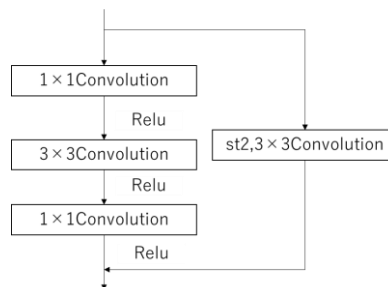


Figure 2 ResNetのダウンスケール方法

この方法ではskip connectionで3/4もの特徴量が無視してしまうという問題がある。

### 2.2 ResNet-RS

ResNet-RS[3]は従来のResNetのアーキテクチャ、学習方法、正規化手法、スケールアップを改善したモデルである。このモデルの特徴は既存のResNetのアーキテクチャに改善を加えただけで大きな変更がないにもかかわらず、同程度の精度を出すモデルより計算が速いことである。

アーキテクチャの改善として、ResNet-D[7]のダウンスケール時のスキップ接続のストライド-2 convolution(Figure 2)をストライド-2の2x2 average poolとストライドなしのconvolution(Figure 3)に変更する。

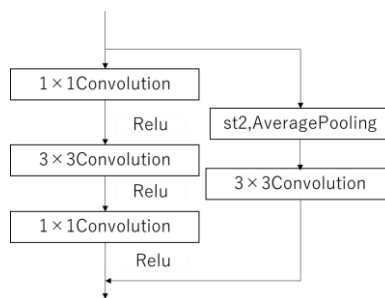


Figure 3 ResNet-Dのダウンスケール方法

この手法を用いることで畳み込み時の特徴が無視されなくなる。また、非ダウンスケール時はSqueeze-and-Excitation Networks[8]の

Squeeze-and-Excitationモジュールを採用しチャンネルごとのセルフアテンションを行っている。

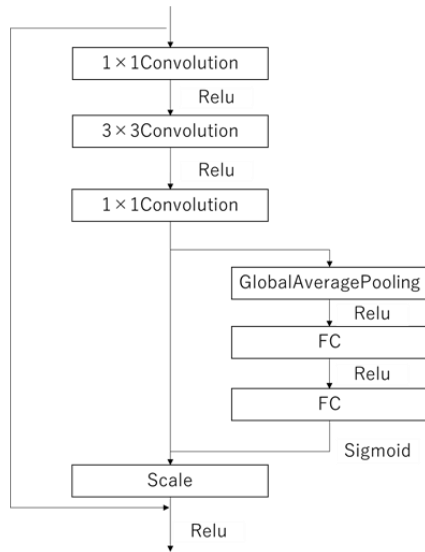


Figure 4 Squeeze-and-Excitation モジュール

学習方法はトレーニングのエポック数を増やす、学習率をcosine LR decayで決める方法で改善した。正規化手法は、RandAugment[4]、momentum optimizer、weight decay、label smoothing、dropout、stochastic depthを適用している。スケールアップ方法は、エポック数が多い学習では、層を深くし、少ないときはチャンネル数を増やす。

### 1.3 GELU

GELU[6]とは活性化関数の一つであり以下のように定義される。

$$GELU(x) = x\Phi(x) \quad (2)$$

誤差関数  $\text{erf}(\cdot)$ を用いると

$$GELU(x) = x \cdot \frac{1}{2} [1 + \text{erf}(x/\sqrt{2})] \quad (3)$$

近似を行い

$$GELU(x) \approx x\sigma(1.702x) \quad (4)$$

とする。ReLUと比較すると学習が早いという特徴がある。

### 1.4 TrivialAugment

TrivialAugment(TA)[5]はデータ拡張をランダムに選び、ランダムな強さで適応するデータ拡張手法である。AutoAugmentやRandAugmentと違いパラメータが不要なのにかかわらず同程度の精度を実現した。

## 3. 提案手法

解決したい問題点は収束速度の改善、計算コストの削減であり、活性化関数をReLUから

GELUに変更、データ拡張をRandAugmentからTrivialAugmentに変更することで解決できると考える。

## 4. 実験および検討

従来手法と今回提案したGELU、TrivialAugmentに変更したものをそれぞれ低エポック、高エポックでの実行速度、精度での比較を行う。

## 5. まとめ

GELUはDropOutと併用するときReLUよりも早く収束する特徴により改善されると考える。また、TrivialAugmentの事前学習が不要であることから計算量の削減ができると考える。

## 参考文献

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv preprint arXiv:1512.03385, (2015)
- [2] Mingxing Tan and Quoc V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, arXiv:1905.11946, (2019)
- [3] Irwan Bello, William Fedus, Xianzhi Du, Ekin D. Cubuk, Aravind Srinivas, Tsung-Yi Lin, Jonathon Shlens and Barret Zoph “Revisiting ResNets: Improved Training and Scaling Strategies”, arXiv:2103.07579, (2021)
- [4] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens and Quoc V. Le, “RandAugment: Practical automated data augmentation with a reduced search space”, arXiv preprint arXiv:1909.13719, (2019)
- [5] Samuel G. Müller and Frank Hutter, “TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation”, arXiv:2103.10158, (2021)
- [6] Dan Hendrycks and Kevin Gimpel, “GAUSSIAN ERROR LINEAR UNITS (GELUS)”, arXiv:1606.08415, (2016)
- [7] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, Mu Li, “Bag of Tricks for Image Classification with Convolutional Neural Networks”, arXiv:1812.01187, (2018)
- [8] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu, “Squeeze-and-Excitation Networks”, arXiv:1709.01507, (2019)