

複数の正則化手法による PyramidNet の精度向上

日大生産工(院) ○野口 尚馬

日大生産工 山内 ゆかり

1. まえがき

深層学習の一手法であるResNet[1]は、スキップ接続により勾配消失を防ぎ、100層を超える深い畳み込みニューラルネットワーク(CNN)の実現に成功した。ResNetが深層学習に与えた影響が大きく、現在でもResNetをベースにした改良モデルが多数提案されている。

本研究では、ResNet改良モデルの一つであるPyramidNet[2]に対し、複数の正則化手法を導入することで精度向上を試みる。また、PyramidNetの確率的正則化手法であるShakeDrop[3]の改良についても検討する。CIFAR-10とCIFAR-100を用いた実験により条件を変えて精度を比較し、提案手法の有効性を検証する。

2. 従来研究

2-1. PyramidNet

ResNetは特徴マップのサイズを半分にするダウンサンプリング層で一気にチャンネル数を増やすため、ネットワーク全体がダウンサンプリング層に依存する形となり、精度向上の妨げとなる問題があった。

PyramidNetはその問題の解決策として提案されたモデルであり、全てのブロック内で線形にチャンネル数を増やしていくのが特徴である。 k 番目のresidualユニット内のチャンネル数 D_k は式(1)で表される。PyramidNetでは各ユニットの入出力のチャンネル数が異なるため、ショートカットの際には足りない分のチャンネルをゼロで埋める方法が用いられる。

$$D_k = \begin{cases} 16 & \text{if } k = 1 \\ \lfloor D_{k-1} + \alpha/N \rfloor & \text{if } 2 \leq k \leq N+1 \end{cases} \quad (1)$$

また、PyramidNetではユニット内の構成についても変更が加えられている。ResNetでは、ユニットの構成は図1のようになっているが、PyramidNetではこれを図2のように変更している。この変更により勾配伝達の効率が上がり、精度が向上することが確かめられている。

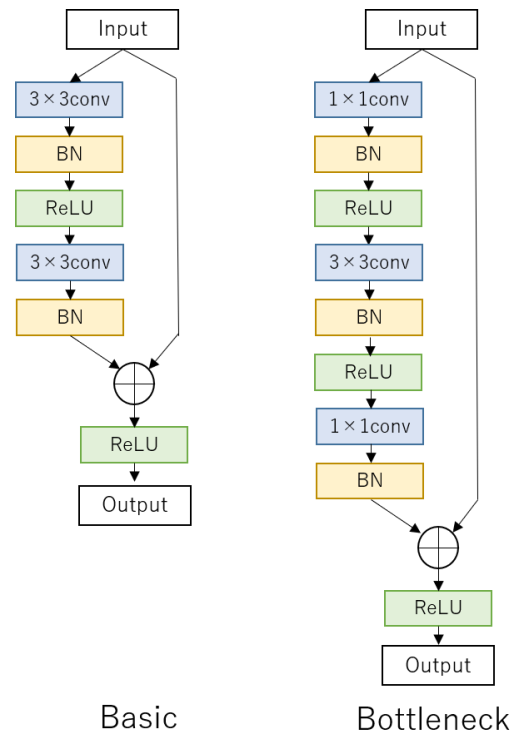


図1. ResNetのユニット構成

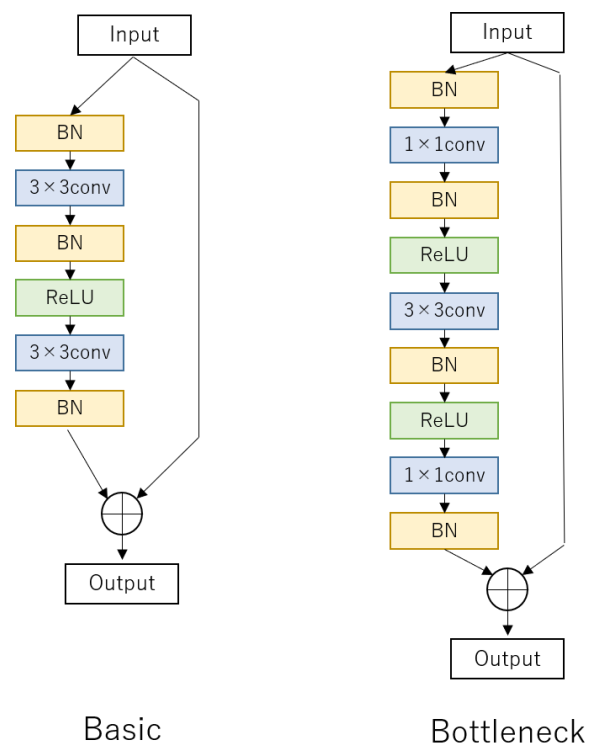


図2. PyramidNetのユニット構成

2-2. TrivialAugment

TrivialAugment[4]は、表1に示す複数の基本的なデータ拡張手法の中から1つをランダムに選び、ランダムな強さで適用する手法である。

表1. データ拡張手法の一覧

名称	説明
Identity	入力をそのまま出力
ShearX	x方向へのせん断
ShearY	y方向へのせん断
TranslateX	x方向への平行移動
TranslateY	y方向への平行移動
Rotate	回転
AutoContrast	RGB毎にヒストグラムが0~255になるように正規化
Invert	画素値反転
Equalize	RGBの各チャンネルの累積ヒストグラムが直線になるように補正
Solarize	指定された閾値よりも大きい画素値を反転
Posterize	画素値を8bitより少ないビット数で表現
Contrast	コントラストの調整
Color	カラーバランスの調整
Brightness	明るさの調整
Sharpness	鮮鋭化の調整

TrivialAugmentはパラメータを持たないためチューニングをする必要がないことに加え、類似手法であるAutoAugment[5]やRandAugment[6]に匹敵する精度を実現している。

TrivialAugmentの適用例を図3に示す。

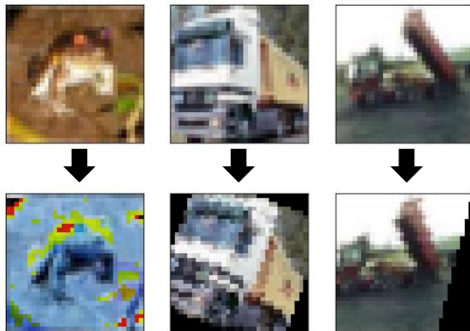


図3. TrivialAugmentの適用例

2-3. CutMix

CutMix[7]は、2枚の入力画像 x_a, x_b を切り貼りして新たな1枚の画像 $\tilde{x}$ を生成するデータ拡張手法である。また、それに合わせてラベル y_a, y_b も混ぜ合わせるのが特徴である。

まず、 $\lambda \in [0,1]$ をベータ分布Beta(α, α)からのサンプリングにより取得する。ここで、 α はハイ

パーパラメータである。次に、切り出す起点の座標 (r_x, r_y) をそれぞれ一様分布Uniform(0, H)とUniform(0, W)からサンプリングする。ここで、H, Wは画像の縦と横のサイズである。切り出す際の幅と高さは $\sqrt{1-\lambda}$ とし、それを元に切り出す終端の座標 (r_w, r_h) を求める。最後に、求めた $(r_x, r_y), (r_w, r_h)$ を使って一方の画像から一部を切り出し、もう一方の画像に貼り付けて新たな画像とする。新たに生成された画像のラベル $\tilde{y}$ は $\tilde{y} = \lambda y_a + (1 - \lambda) y_b$ とする。

CutMixの適用例を図4に示す。



図4. CutMixの適用例

2-4. SEblock

SEblock[8]は、図5のような構造のアーキテクチャであり、特徴マップをチャンネル毎に適応的に重み付けすることを目的としている。最初にGlobal Average Poolingでチャンネル毎の画素値平均を取り、次に全結合層でチャンネル数を C/r に圧縮してReLUを適用する。次にチャンネル数を元に戻し、Sigmoidを適用することでチャンネル毎の重みが得られる。最後にこの重みをスケールして元の入力と掛け合わせ、出力する。

SEblockはシンプルな構造でありながら、基本的にどのモデルにも導入でき、精度向上が期待できる。また、導入に伴う計算量の増加も少ない。

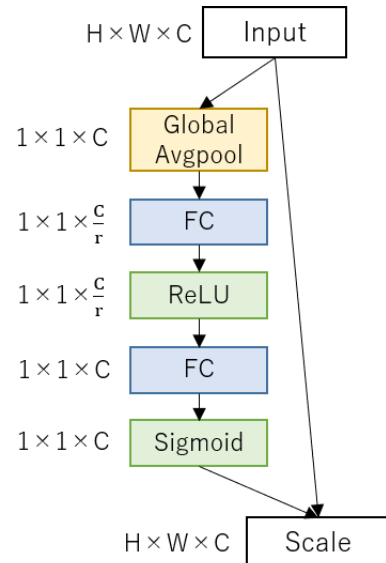


図5. SEblock

2-5. ShakeDrop

ShakeDropは、図6のように各ユニットの最後の加算の直前で順伝播時と逆伝播時にそれぞれ別々の一様乱数 α_l と β_l を掛ける正則化手法である。ただし、図6の b_l は確率 p_l で1を、確率 $1 - p_l$ で0を取るベルヌーイ変数である。テスト時には、順伝播時の期待値 $E[b_l + (1 - b_l)\alpha_l]$ を掛ける。確率 p_l は出力層に近いほど低くなるように設定され、式(2)で表される。ただし、 L はユニット数である。

$$p_l = 1 - \frac{l}{2L} \quad (2)$$

α_l と β_l の範囲は $\alpha_l \in [-1,1]$, $\beta_l \in [0,1]$ とした場合に精度が最も良くなることが実験的に確かめられている。ShakeDropを適用することで、過学習の抑制や精度向上の効果が得られることが確認されている。その理由として、ネットワーク全体としては正しい重み更新が行われるが、一部では正しくない重みが更新され、そのバランスによってネットワークの学習が安定しながら強い摂動が加わるためとされている。

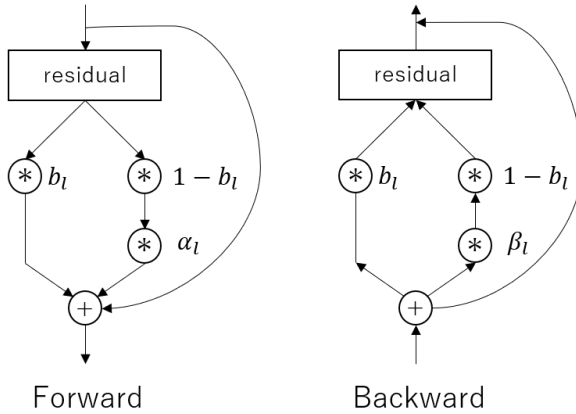


図6. ShakeDrop

3. 提案手法

本研究では、PyramidNetの精度向上を目的として、TrivialAugment、CutMix、SEblock、ShakeDropを導入し、どのような条件で最も画像認識の精度が向上するか検証する。図7に、PyramidNetにSEblockとShakeDropを組み込んだ際のresidualユニットを示す。

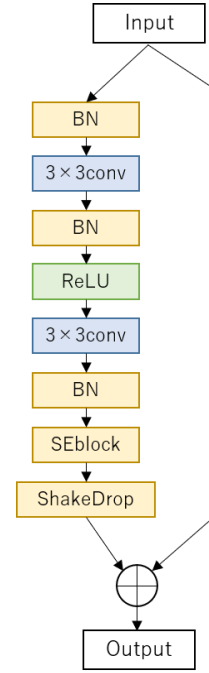


図7. 提案手法のresidualユニット

また、本研究ではShakeDropの改良についても検討する。ShakeDropの問題点として、ネットワークに摂動が加わるために収束が遅くなり、学習に通常の数倍のエポック数を要する点が挙げられるが、本研究では、その原因は適用確率の高さにあると仮定し、式(2)に係数 $\lambda \in [0,1]$ を導入することで p_l を式(3)のように変更することを提案する。

$$p_l = 1 - \frac{\lambda l}{2L} \quad (3)$$

$\lambda = 1$ のときは通常のShakeDropと同じになり、 $\lambda = 0$ のときはShakeDropを適用しない場合と同じになる。提案手法により、エポック数を増やさずに収束させることが可能か検証する。

4. 実験および検討

110層のPyramidNet($\alpha = 84$)に対し、表2に示すA~Fまでの条件でCIFAR-10とCIFAR-100による認識精度の測定を行う。

表2 条件

	A	B	C	D	E	F
CutMix			○	○	○	○
Trivial Aug			○	○	○	○
SEblock		○	○	○	○	○
ShakeDrop				$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$

エポック数は300とし、学習率は0.1を初期値として0.001までコサイン減衰させる。最適化手法にはSGDを使用し、Momentumは0.9に設定する。weight decayは0.0001とする。バッチサ

イズは128とする。また、基本のデータ拡張として4ピクセルの平行移動と確率0.5の左右反転を行う。CutMixを使用する場合は適用確率を1、 $\alpha = 1$ とする。また、SEblockを使用する場合はチャンネルの圧縮比を $r = 4$ とする。

実験結果を表3に示す。

表3 実験結果

	CIFAR-10 Error(%)	CIFAR-100 Error(%)
A	4.34	22.45
B	4.49	22.06
C	2.43	17.04
D	2.37	15.81
E	2.47	16.48
F	2.93	19.16

正則化を行っていないAやSEblockのみを追加したBではCIFAR-100で過学習の傾向が見られ、精度が伸びなかった。一方、FではShakeDropの効果で過学習が抑えられていたものの、300エポックでは収束せず、ShakeDropを使用しないCより精度が悪化した。今回の実験では、 $\lambda = 0.25$ としたDがCIFAR-10、CIFAR-100ともに最も良い精度となった。これは、ShakeDropの適用確率を下げるにより、精度向上と収束の速さを両立できたためだと考えられる。ただし、Dの場合でもCIFAR-100では完全に収束しなかった。

5. まとめ

本研究では、PyramidNetに複数の正則化手法を導入して精度向上を試みた。また、ShakeDropについては係数 λ を導入して確率 p_l の式を変更することでエポック数を増やさずに収束させられないか検証した。

結果、提案手法は $\lambda = 0.25$ とした際に、正則化をしない場合と比較してCIFAR-10で約2%、CIFAR-100で約6.6%の精度向上に成功した。今後の課題として、収束を更に速めるために最適化手法の変更なども検討したい。また今回の実験では λ を定数としたが、学習の進行に応じて変化させた場合についても検証したい。

参考文献

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun., “Deep Residual Learning for Image Recognition” (2015)
- [2] Dongyoon Han, Jiwhan Kim, Junmo Kim., “Deep Pyramidal Residual Networks” (2016)
- [3] Yoshihiro Yamada, Masakazu Iwamura, Takuya Akiba, Koichi Kise., “ShakeDrop

Regularization for Deep Residual Learning” (2018)

[4] Samuel G. Müller, Frank Hutter., “TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation” (2021)

[5] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, Quoc V. Le., “AutoAugment: Learning Augmentation Policies from Data” (2018)

[6] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, Quoc V. Le., “RandAugment: Practical automated data augmentation with a reduced search space” (2019)

[7] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo., “CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features” (2019)

[8] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu., “Squeeze-and-Excitation Networks” (2017)