The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Residual Networks に対する確率的正則化の提案： Shake-ResDrop と Shake-SENet

Stochastic Regularization for Residual Networks: Shake-ResDrop and Shake-SENet

白濱　淳也　　川本　一彦
Junya Shirahama　　Kazuhiko Kamoto

千葉大学
Chiba University

Recently, residual networks (ResNets) and their improvements, such as stochastic regularization, have proven to be able to reduce overfitting during training processes. In this paper, we propose two stochastic models which combines stochastic regularization and attention mechanism. The two models are based on ShakeDrop, combining either SENet or Stochastic Depth with ShakeDrop itself. Both of our methods were able to improve existing ShakeDrop results on CIFAR-100.

1.　はじめに

深層学習を用いた画像認識タスクでは，畳み込みニューラルネットワーク（CNN）が主流となっている．Residual networks（ResNet）[He et al. 2016] は，ショートカット接続をもつ CNN で，2015 年に 150 層以上という大幅な深層化を実現した．ResNet は深層化に欠かせないため，現在でも広く利用されている．最近では，ResNet の汎化能力を向上させるために確率的な正則化が提案されている．Dropout[Srivastava et al 2014] のようにニューロンではなく，ResNet の層を確率的に落とす Stochastic Depth（ResDrop）[Haung et al. 2016] や，外乱を加えデータ拡張効果をもたらす Shake-Shake [Gastaldi et al. 2017] などの確率的手法が提案されている．さらに，ResDrop のように層を落とすのではなく Shake-Shake のように外乱を加えるか加えないか確率的に変化させ，ShakeDrop[Yamada et al. 2018] も提案されている．ShakeDrop は CIFAR-100 を用いた画像分類で当時の最高精度を記録した．さらに，ResNet にアテンション機構を導入した SENet[Hu et al. 2018] も提案されている．

本研究では，このようなアテンション機構や確率的正則化のモデルを組み合わせてより高性能な確率的正則化を提案する．

2.　提案手法

既存の確率的正則化手法である ShakeDrop を軸に ResDrop と SENet を組み合わせたアーキテクチャを提案する．

2.1　ShakeDrop と ResDrop

ShakeDrop と ResDrop を組み合わせる．ShakeDrop に層を落とす機構を加えるため，図 1 のように，residual モジュール，ショートカット接続 (出力に 0 をかけるため)，出力に乱数 $\alpha \in [-1, 1]$ をかける ShakeDrop モジュール，の 3 つの構造が学習中に現れることになる．その結果，ShakeDrop よりも強いアンサンブル学習の効果を得ることが期待できる．さらに，ショートカット接続が加わることにより，パラメータを減らすこともでき，学習時間を短縮することもできる．また，ResDrop や ShakeDrop と同様に層が深くなるにつれて，residual モジュールが現れる確率を低くし，残りの 2 つの確率が高くなるようにする．この確率の変化は，ネットワークが 5 層のとき図 2 に示すようになる．

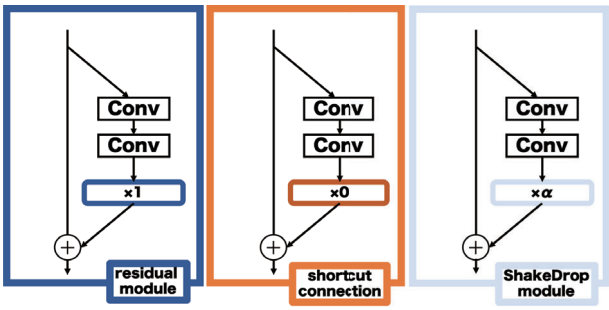

図 1: ShakeDrop と ResDrop 組み合わせて現れる 3 つの構造

2.2　ShakeDrop と SENet

この 2 つを組み合わせる方法としては図 3 に示すようになり，ShakeDrop させた後に SE Block を入れるか（図 3 左），前に入れるかとなる（図 3 右）．

ShakeDrop の後に加えた場合は ShakeDrop の特徴空間でのデータ拡張と SENet のアテンション機構を組み合わせることができる．この組み合わせでは，図 3 左に示すように，畳み込み層の出力に $\alpha \in [-1, 1]$ をかけるため符号反転も起こるうり，学習を乱しすぎることになると考えられる．乱数をかけない場合（residual モジュールになる）は SENet の本来のアテンション機構を生み出すことができるため，後に加えた場合も精度向上が見込まれる．

ShakeDrop の前に加えた場合は ShakeDrop のでのデータ拡張に SENet のアテンション機構を組み合わせることができる．ShakeDrop の乱数と同時に，畳み込み層からの出力を SE Block の重みでスケーリングするため，チャンネルのアテンションを含むデータ拡張を行うことができる．乱数が負を含むことによってチャンネルのアテンションの符号反転を入れることもできるため，負の値での強い摂動を更に与えれると考えられる [Yamada et al. 2018]．また前に加える場合後に加えた場合と同様に，乱数をかけないときは SENet を本来の形で与えられる．

連絡先: 川本一彦，千葉大学大学院工学研究院，千葉市稲毛区弥生町 1-33，kawa@faculty.chiba-u.jp

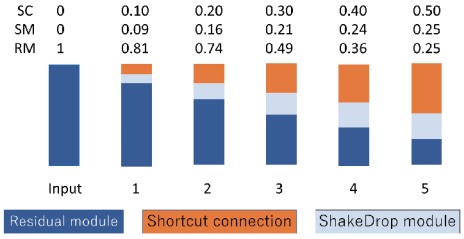

図 2: 5 層である場合の 3 種類の現れ方

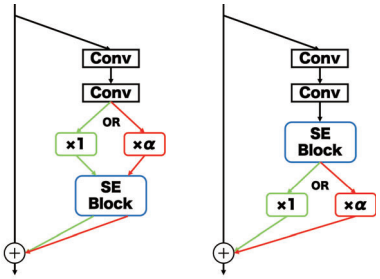

図 3: 左：ShakeDrop の後　右：ShakeDrop の前

3. 実験結果と考察

一般画像認識データセットの CIFAR-100 を用いて誤識別率で比較する．CIFAR-100 は 100 クラス各 600 枚，合計 60000 枚の画像を抽出したデータセットで，各クラス 500 枚を学習用データ，100 枚を評価用データとする．実験の設定を表 1 に記す．

表 2 に実験結果を示す．表 2 より，ShakeDrop 単独の手法よりも，3 つの提案手法で誤識別率は低くなっている．ResDrop と組み合わせでは，アンサンブル学習の効果を強めることができ性能向上できたと考えている．さらに，パラメータ数を減らすことができるため学習時間の減少も確認できる．ShakeDrop の前に SENet を組み込む手法が最も低い誤識別率を達成している．このことは，チャンネルのアテンションを含むデータ拡張をすることが性能向上に貢献することを示している．ShakeDrop の後に SENet を組み込む手法も，ShakeDrop 単独に比べて誤識別率は低くなっており，アテンション機構を単純に確率的正則化に組み合わせても有効であることを示している．

4. おわりに

本研究では，確率的正則化手法の ShakeDrop に，確率的正則化手法の ResDrop やアテンション機構の SENet を組み合わせることの有効性を従来手法と比較することにより実証した．今後は検証した確率的正則化方法とアテンション機構を同時に組み合わせることを検討している．

謝辞

本研究は JSPS 科研費 JP16K00231 の助成を受けたものです．

表 1: 実験設定

バッチサイズ	128
エポック数	1800
データ拡張	Random Erasing [Zoung et al. 2016]
学習率	0.2
重み減衰	0.0001
最適化関数	SGD with momentum
損失関数	クロスエントロピー誤差
学習計画	Cosine Learnig Scheduling [Loshchilov et al. 2017]

表 2: CIFAR-100 に対する実験結果

手法	誤識別率 (%)	学習時間 (h)
ShakeDrop	13.16	66.8
ShakeDrop and ResDrop	12.76	63.0
ShakeDrop after SENet	12.42	77.9
ShakeDrop before SENet	12.83	76.8

参考文献

[He et al. 2016] K.He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. CVPR, pp.770-778, 2016.

[Srivastava et al 2014] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” J. of Machine Learning, vol.15, pp.1929-1958, 2014.

[Haung et al. 2016] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Weinberger, “Deep networks with stochastic depth,” arXiv preprint arXiv:1603.09382v3, 2016.

[Gastaldi et al. 2017] X. Gastaldi, “Shake-shake regularization,” arXivpreprint arXiv:1705.07485v2, 2017.

[Yamada et al. 2018] Y. Yamada, M. Iwamura, and K Kise. “Shakedrop regularization,” arXiv preprint arXiv:1802.02375, 2018.

[Hu et al. 2018] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu. “Squeeze-and-excitation networks,” Proc. CVPR, pp.7132-7141, 2018.

[Zoung et al. 2016] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” arXiv preprint arXiv:1708.04896, 2017.

[Loshchilov et al. 2017] I. Loshchilov , F. Hutter. “stochastic gradient descent with restarts.” arXiv preprint arXiv:1608.03983, 2016.