

パラメトリックなソフトサイン関数を用いた GAN による画像生成

日大生産工(院) ○滝浪 省吾 日大生産工 角田和彦
日大生産工 三井和男

1. まえがき

ディープラーニング技術の台頭を皮切りに、機械学習分野は大きく発展している。とりわけ強力なデータ生成モデルとして知られている Generative Adversarial Network (GAN) は発表されて以来大きな話題を集め、フルHDの写真画質画像を生成できる Progressive GAN¹⁾や風景写真を筆で描かれた絵画のようにできる CycleGAN²⁾など様々な派生が生まれている。しかしGANにはハイパーパラメータのチューニングが難しいことや、モード崩壊と呼ばれる類似したデータパターンを繰り返し生成する現象³⁾が起きてしまうといった課題があり学習が非常に不安定である。

そこで本研究では、GANの学習の安定性向上を目的として Convolutional Neural Network (CNN) の画像認識精度を改善させた Parametric Softsign Function⁴⁾ (PSF) を導入し、既存手法との生成画像の比較を行うことでGANにおけるPSFの有用性を検証した。

2. 活性化関数

2.1 Rectified Linear Unit

ディープラーニングで広く用いられている Rectified Linear Unit⁵⁾ (ReLU) は、学習の逆伝播時における勾配消失問題を抑制することのできる活性化関数である。しかしその性質上、入力値 x が0以下の場合に学習が滞ってしまう可能性があるという欠点を持つ。ReLUは次式のように示される。

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

これを補う形でReLUを発展させたものとして LeakyReLU⁶⁾ (LReLU) や Parametric ReLU⁷⁾ (PReLU) がある。LReLUは入力値 x

が0未満の場合にも学習が滞ることのないように、負の入力値における出力に僅かな勾配 α ($\alpha = 0.01 \sim 0.2$ 程度) を持たせている。また PReLUでは α の値を学習によって動的に決まるパラメータとして扱うことで、関数そのものを最適化させることができる。GANの学習においてはこれらを用いることでモデルのパフォーマンスが向上することが知られている。LReLU及びPReLUは次式のように示される。

$$f(x) = \begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases} \quad (2)$$

その他にも、ReLUの負の領域が非線形となる Exponential Linear Unit⁸⁾ (ELU) も提案されている。ELUは次式のように示される。

$$f(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases} \quad (3)$$

2.2 Parametric Softsign Function

本研究で取り入れたPSFはパラメータを持つPReLUと非線形の特徴を持つELUの双方の特徴を引き継いだものである。PSFは次式のように示される。

$$f(x) = \begin{cases} x & (x \geq 0) \\ \frac{e^{\alpha x}}{e^{\alpha} + |x|} & (x < 0) \end{cases} \quad (4)$$

ただし α はネットワークと共に学習するパラメータとする。

この活性化関数は、 $\alpha \rightarrow -\infty$ の極限でReLU関数に収束し、 $\alpha \rightarrow +\infty$ の極限で傾き1の線形関数に収束する (Fig.1参照)。PSFは負の領域において層ごとに関数が最適化される非線形な関数であるため、ネットワークの表現力向

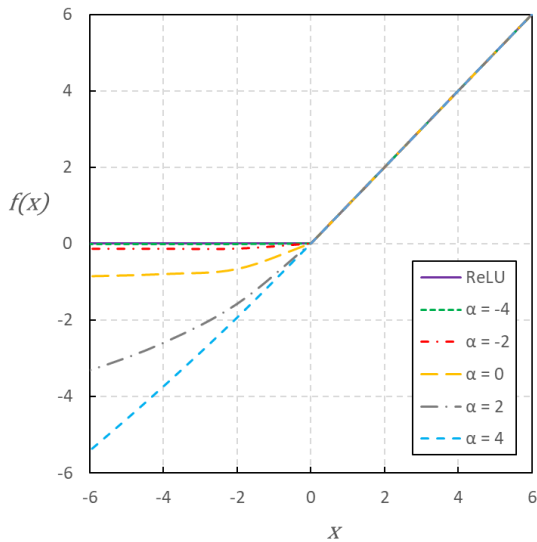


Fig.1 ReLU & Parametric Softsign Function

上に大きく貢献している。これにより、GANの生成モデルが多様な表現力を獲得することが期待できる。

3. モデルの構築

本研究ではDeep Convolutional GAN⁹⁾ (DCGAN) をpythonとkerasを利用して実装し、比較検証のため、既存の活性化関数としてLReLUを用いたモデルとその大部分をPSFに差し替えたモデルを用意して学習を行った。DCGANはRadfordらによって2016年に提案されたもので、GANの識別器と生成器の両方をCNNで表現することで生成画像の品質を格段に向上させる手法である。DCGANの生成器は下のFig.2¹⁾で示すように入力として受け取ったランダムノイズベクトルを転置畳み込みによって画像サイズに拡張していく。

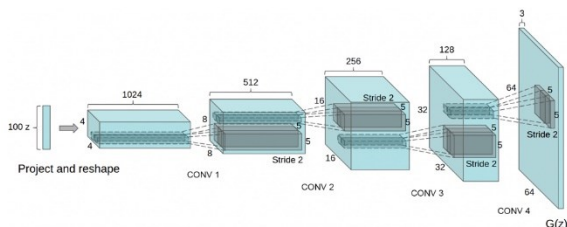


Fig.2 Generator's model at DCGAN

識別器はその逆で画像を入力として受け取り、畳み込み層を適用したのち最終層にてシグモイド関数を通して、入力画像が偽物ではなく本物である確率を出力する。訓練データには著名なデータセットであるMNISTとCIFAR-10の二つを使用する。CIFAR-10からは自動車の画像5000枚を訓練データとして抽出した。MNISTでは20000イテレーションの学習を行い、CIFAR-10ではバッチサイズを32として181エポックの学習を行った。また一定の回数分学習を終える度に画像を生成させて、学習過程の様子がわかるようにした。

4. 生成結果と損失の比較

4.1 MNIST

実装したPSFモデルとLReLUモデルによって生成された画像を観察し比較を行う。生成された画像をFig.3及びFig.4に示す。さらに、識別器と生成器それぞれにおいて両モデルの学習時の損失を比較したものをFig.5及びFig.6に示す。両モデル共に数字の形状を捉えて正しく学習できたことがわかる。学習過程においても0から9までの数字がまんべんなく表れていた。Fig.5より識別器の損失を比較するとLReLUモデルの損失は学習の終盤に差し掛かって不安定なのに対して、PSFモデルの損失は緩やかに下降していることが確認できる。Fig.6でも同様にPSFモデルの損失が終盤にかけて緩やかに推移している。

MNISTデータセットにおいては、PSFを導入することによって安定性の向上は見込めたが大きな効果を得るには至らなかった。

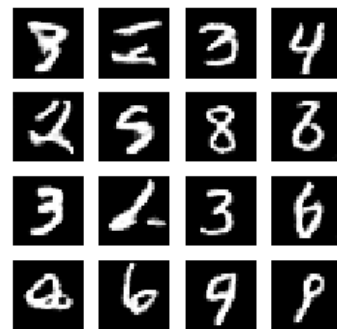


Fig.3 Generated MNIST images at PSF

¹⁾ 出典: Alec Radford, Luke Metz, Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Network. ICLR2016, pp.4.

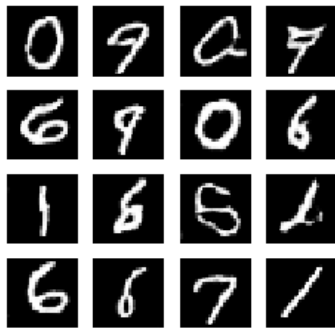


Fig.4 Generated MNIST images at LReLU



Fig.5 Discriminator's loss at MNIST

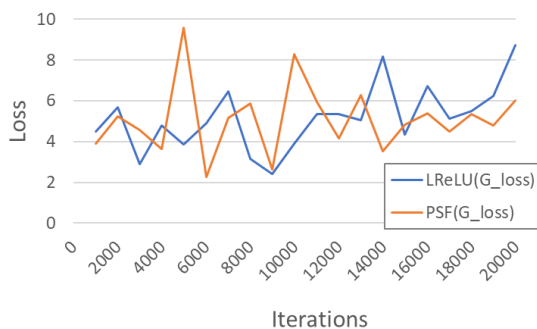


Fig.6 Generator's loss at MNIST

4.2 CIFAR-10

同様に生成された画像の比較を行う. PSFモデルとLReLUモデルによって生成された画像をそれぞれFig.7及びFig.8に示す. さらにFig.9及びFig.10では識別器の損失と生成器の損失のエポック毎の推移を示す. こちらでもLReLUモデルとPSFモデルで生成された画像の品質という面では大きな差異は見受けられなかった. PSFモデルでは車体は同一でありながら色が異なるものや背景が異なるものなど様々なパターンの画像生成に成功していた. 最終エポック後に生成された画像では類似した画像ばかりが生成されてしまい, モード崩壊の傾向が見られたがそれまでは安定して生

成することができていた. 対してLReLUモデルにおいては, 学習が100エポックを超えた時点ですでにFig.8のような類似した画像ばかりを生成してしまうことが確認でき, その後も頻繁にモード崩壊が発生していた. Fig.9の識別器の損失を見ると, PSFモデルが学習の初めから終わりにかけてジグザグかつ緩やかに下降しており, 非常に安定して学習できていたことが確認できる. 一方で, LReLUモデルでは160エポックを超えた頃から損失の増減が急激に不安定になっている. 最後にFig.10より生成器の損失を比較する. 全体的に見てPSFモデルの方がLReLUモデルの損失を上回ってしまっているのが確認できるが, これはLReLUモデルの生成器が識別器を騙せる生成パターンを見つけてしまいその生成パターンから抜け出せなくなるモード崩壊に依るところが大きいと考えられる.

CIFAR-10データセットにおいては, PSFの導入によってモード崩壊の発生を抑制し, 多様な画像を生成することができた. また, 識別器及び生成器の損失からも安定して学習できていたことが確認できた.



Fig.7 Generated automobile images at PSF



Fig.8 Generated automobile images at LReLU

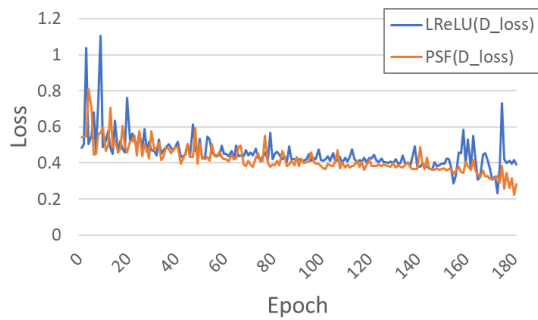


Fig.9 Discriminator's loss at CIFAR-10

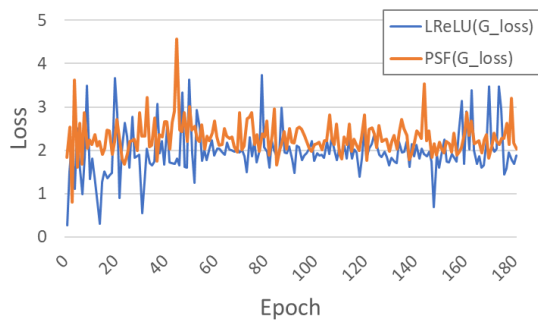


Fig.10 Generator's loss at Cifar-10

5. まとめ

本研究では、負の入力値を効果的に取り入れることのできる活性化関数をDCGANに導入し、既存手法との差異を検証した。

生成結果から、PSFを取り入れることで学習が安定し、生成結果に多様性が生まれたと同時にモード崩壊現象への抑制効果があることが確認できた。一方で、生成された画像の質について大きな差異はなかった。このことから、形状が単純な画像を生成させる際には、既存手法で十分に精度の高い画像が生成できるためあえてPSFを用いる必要性はないといえる。反対に、複雑な形状を持つ画像や統一性のない画像を生成させる際にはPSFが高い効果を発揮することが確認できた。

PSFは実装面において他の派生GANと組み合わせることが容易なため更なる性能向上が期待できる。ただしネットワークの層が複雑になるほどパラメータを最適化させる計算量が膨大になるため、相応の演算性能が要求される。今後の課題としてより高解像度、高品質な画像生成に挑戦することのほか計算コストの削減が必要だと考えられる。

参考文献

- 1) Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen., “Progressive Growing of GANs for Improved Quality, Stability, and Variation”, arXiv:1710.10196v3 [cs.NE] (2018)
- 2) Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, arXiv:1703.10593v7 [cs.CV] (2020)
- 3) Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Redford, Xi Chen., “Improved Techniques for Training GANs” arXiv:1606.03498v1 [cs.LG] (2016)
- 4) 榎本知幸, 角田和彦, 三浦慎一郎, “ニューラルネットワークにおける非線形活性化関数の提案及びその評価”, 第52回日本大学生産工学部学術講演会, (2019)
- 5) V. Nair and G.E. Hinton., “Rectified Linear Units Improverestricted Boltsmann Machines”, In ICML, (2010), pp807-814
- 6) Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li., “Empirical Evaluation of Rectified Activations in Convolutional Network”, arXiv:1505.00853, (2014), pp2
- 7) Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun., “Delving Deep into Rectifiers”, zrXiv:1502.01852v1 [cs.CV] (2015)
- 8) Djork-Arné Clevert, Thomas Unterthiner, Sepp Hochreiter., “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”, arXiv:1511.07289v5 [cs.LG] (2016)
- 9) Alec Radford, Luke Metz, Soumith Chintala., “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, arXiv:1511.06434v2 [cs.LG] (2016)