

H-038

生成画像品質を考慮した DCGAN を用いたデータオーギュメンテーション Data Augmentation using DCGAN considering generated image quality

渡部 宏樹[†]
Hiroki Watabe

渡辺 裕[†]
Hiroshi Watanabe

1. はじめに

スマートフォンの普及によりブログや SNS が身近になり、そこから多くの写真がインターネット上に流れるようになった。そのようなデータを解析し、様々なサービスに活用するためには画像のタグ情報が必要である。動物の品種情報は、画像のタグ情報として有用な情報の一つである。品種情報を得るために、畳み込みニューラルネットワーク (CNN) 等の機械学習手法を用いて品種識別が行われることがある。しかし、特定の品種について十分な訓練画像枚数を用意できないことがある。

そこで我々は、動物の顔画像に類似した画像を作成し、それを CNN の訓練データセットに加えることで学習画像枚数を補う手法を提案してきた。実験の結果、追加する類似画像の品質が CNN の性能に影響を及ぼすことが確認できた。

本稿では、生成画像の品質を評価し、品質の良い画像を CNN の訓練画像に追加する手法を提案する。実験により、追加する画像の品質を考慮しない場合に比べて、識別性能が向上することを確認した。

2. DCGAN

画像の生成器と判別器を共進化させることで、性能の良い生成器を得るシステムを Generative Adversarial Networks (GAN) という [1]。生成器は、一様分布などからサンプルされた乱数を入力として受け取り、これを種として画像を生成する。判別器は、入力画像が学習データセット由来か生成器由来かを判別する。GAN の学習の際、生成器は判別器がデータセット由来だと勘違いするような画像を生成するように学習する。判別器は入力画像がデータセット由来か生成器由来かを正しく判別できるように学習する。GAN は、生成器と判別器が競い合うように学習を進めることで、本物と見分けがつかない精度の画像を生成することを目的としている。

この GAN を特別な CNN で実装したものを Deep Convolutional Generative Adversarial Networks (DCGAN) という [2]。DCGAN は GAN よりも精密な画像を生成することが出来る。

3. 提案手法

3.1 評価値

GAN の生成画像の評価値の一つに Inception score がある [3]。これは二つの確率分布の KL ダイバージェンスを求めたもので、式(1)のように表される。

$p(y|x)$ は個々の生成画像の認識結果の確率分布を、 $p(y)$ は生成画像全体の認識結果の確率分布の平均を表す。

[†] 早稲田大学大学院 基幹理工学研究科

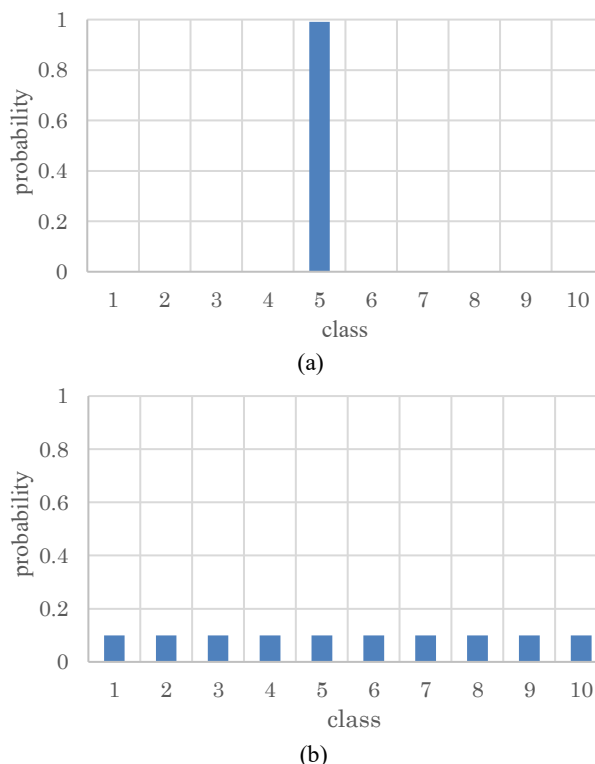


図 1: 理想的な確率分布の例

$$I = \exp(\mathbb{E}_x KL(p(y|x)||p(y))) \quad (1)$$

様々な種類の画像を学習した場合、GAN はそれぞれの画像が意味を持つ多様な画像群を生成する。そのため、理想的な $p(y|x)$ の分布は図 1(a) のように一つのクラスに突出したものとなる。一方 $p(y)$ の分布は図 1(b) のように全てのクラスに均等なものとなる。KL ダイバージェンスは二つの確率分布の差異を計る尺度であり、式(2)のように表される。 $p(y|x)$ および $p(y)$ が理想的な分布の場合、その差異は大きいため Inception score は大きな値を取る。Inception score は人の主観的な品質評価と相関があるとされている。

$$KL(P(x)||Q(x)) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (2)$$

本研究では、特定の品種のみを学習した DCGAN を使用する。そのため、理想的な $p(y)$ の分布も図 1(a) のようにある品種に突出したものとなる。よって、理想的な分布の場合、KL ダイバージェンスは小さな値を取る。そこで本研究では、Inception score の逆数を画像の評価値として用いる。

3.2 データオーギュメンテーション

本研究では、DCGAN を用いた CNN の訓練データセットのデータオーギュメンテーションを提案する。提案手法の流れは以下のとおりである。

1. DCGAN で学習を行い、類似画像の生成を行う。
2. 生成された類似画像から評価値の高いものを抽出する。
3. 抽出した類似画像を元のデータセットに加える。
4. 3 のデータセットに、通常のデータオーギュメンテーション(画像の反転、輝度変化)を施す。
5. 作成したデータセットで CNN の学習を行う。

DCGAN の生成画像は判別器の CNN が本物かどうかの区別がつかない画像である。そのため、基の画像に類似した CNN 特徴量を持ち、類似画像として機能すると推測される。CNN は訓練に使う画像枚数が多いほど高い性能を発揮する。よって、訓練画像枚数が十分でないデータセットにおいて、画像枚数を増やすのに有用なデータオーギュメンテーションとなると考えられる。

4. 評価実験

提案手法について、”The Oxford-IIIT-Pet dataset” [4]内の猫・犬の顔画像各 12 品種を用いて実験を行った。訓練・テストにはそれぞれ一品種あたり 100 枚の画像を使用した。提案手法では、さらに一品種あたり 100 枚の生成画像を訓練画像に追加した。評価指標として、式(3)で示される正解率を使用した。

$$\text{正解率(\%)} = \frac{\text{正解数}}{\text{テストデータ数}} \times 100 \quad (3)$$

生成された画像の例を図 2 に示す。評価値による品質評価の結果、品質の良いとされたものには比較的猫や犬に見える画像が多く見られた。品質の悪いとされたものには、顔の輪郭が崩れていたり、猫の形を保っていなかったりする画像が多く見られた。

実験の結果を表 1 に示す。ここで従来手法は、画像の反転や回転、輝度変化といった通常のデータオーギュメンテーション手法のみを用いた場合を示す。実験結果より、提案手法が従来手法よりも高い性能を示すことが確認できた。さらに、評価値が高い画像のみを追加した場合、生成画像の品質を考慮しない場合に比べて高い性能を示した。

また、猫に比べて犬の場合は提案手法による精度の向上が小さかった。これは、猫に比べて犬の生成画像の方が品質の悪い画像が多かったためと考えられる。

5. おわりに

本稿では、DCGAN により動物の顔の類似画像を生成し、評価値を用いて生成画像の評価を行った。そしてそれを CNN の訓練データセットに加えるデータオーギュメンテーションを提案した。実験の結果、品質の良い画像を用いることで猫の場合で 2.5%、犬の場合で 0.5%の性能向上を確認した。



(a: 品質の良い例)



(b: 品質の悪い例)

図 2: 生成画像例

表 1: 実験結果(正解率(%))

	猫	犬
従来手法	81.9	80.8
提案手法(品質ランダム)	82.3	81.2
提案手法(品質良い)	84.4	81.3

今後は DCGAN を改良して生成画像の品質を上げることで、さらなる性能向上を目指していく。

参考文献

- [1] A.Radford, L.Metz, and S.Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks”, In 4th Int. Conf. on Learning Representations (ICLR), No 67, pp.1-16, (2016)
- [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets”, Advances in Neural Information Processing Systems (NIPS), pp. 2672-2680, (2014)
- [3] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training GANs”, Advances in Neural Information Processing Systems (NIPS), pp.2234-2242, (2016).
- [4] O.M. Parkhi, A. Vedaldi, A. Zisserman, and C.V. Jawahar, “Cats and Dogs”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3498-3305, (2012)