

クラス分類を用いた GAN における画像生成の多様化

鈴木 祐太[†] 坂上 文彦[†] 佐藤 淳[†]名古屋工業大学[†]

1.はじめに

近年、深層学習によりリアルな画像生成を行う Generative Adversarial Network (GAN) [1] の研究が進んでいる。GAN 研究の多くでは、1つの入力から1つの最適な解(リアルな画像)を出力することが重要視されており、生成画像の多様性は重要視されていなかった [2]。しかし、実際には多くの画像生成問題においては、1つの入力に対して最適な生成画像が多数存在しており、1つの入力から多様な画像が生成されることが求められる。MSGAN [3]では、この問題に対して GAN の多様性を評価する Loss を組み込むことで生成画像の多様化に成功した。しかし、これらの先行研究においても、訓練画像と比べると生成画像の多様性はまだ十分に高いとはいえない。

そこで本研究では、訓練画像と同等またはそれ以上の多様な画像を生成する手法を提案する。従来手法では、学習に用いる訓練画像の分布を知ることなく学習を行っていたが、本研究では事前に訓練画像の分布を推定し、訓練画像の多様性を疑似するよう学習を行うことで、より真の分布に近いバリエーション豊かな画像を生成する GAN を提案する。

2.提案法

2.1.訓練画像の分布推定

提案法では、まず GAN の学習前に k-means クラスタリングを用いて訓練画像の分布を推定する。最初に、事前学習済みの識別器を Encoder として用い、画像の特徴ベクトルを抽出する。次に、抽出した特徴ベクトルを用いて訓練画像の k-means クラスタリングを行い、訓練画像の分布を推定する。k-means クラスタリングにより、全ての訓練画像はそれぞれ一つのクラスに属することになる。以下では、推定した訓練画像の分布を用いて画像生成を行う GAN のネットワーク構造と学習方法について説明する。

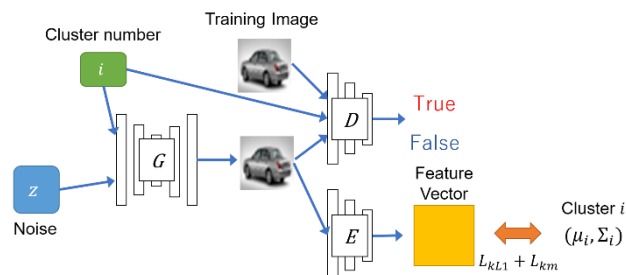


図 1: 提案法のネットワーク

2.2. ネットワーク構造

本研究では、画像を生成する際にクラス番号を入力し、入力したクラス番号に応じて異なる特徴の画像を生成することで多様性のある画像生成を行う。提案手法のネットワークでは、図 1 に示すように従来の GAN に対してクラス番号 i を入力する。クラス番号は Discriminator にも入力され、Discriminator は画像がリアルであるかだけでなく、クラス番号に合った画像を生成できているかどうかを判定する。また、図 1 の E は事前学習済みの特徴抽出器である。この特徴抽出器を通して得られた特徴ベクトルがクラス i の画像の特徴ベクトルに近づくよう学習を行う。特徴抽出器には EfficientNet-b0 を使用した。

提案法は image2image や text2image のような変換タスクに対しても適用可能であり、その場合はクラス番号とノイズだけでなく変換前の image や text も Generator と Discriminator の入力に加える。

2.3. 学習

訓練時には、学習に用いる訓練画像が属するクラス番号を Generator に入力して画像を生成し、生成された画像がそのクラスの特徴を持つよう学習を行う。先に述べたように、Generator が入力クラス番号に合った画像を生成できているかどうかの判定は Discriminator が行う。しかし、クラス分けされた画像は犬カテゴリや車カテゴリのように見た目ではっきり分かるようなカテゴリ分けされたものとはならない。そのため、Discriminator だけで、クラスらしさを判定するのは困難であると考えられる。

そこで、生成画像が入力クラスタの特徴を含んだ画像であるかどうかを判定するため、以下の2つの Loss を Generator の学習に加える。

$$L_{kl1} = |\mu_i - E(G(z, i))|$$

$$L_{km} = \sqrt{((E(G(z, i)) - \mu_i)^T \Sigma_i^{-1} (E(G(z, i)) - \mu_i))}$$

ここで、 μ_i はクラスタ i の平均、 Σ_i はクラスタ i の共分散行列であり、 L_{kl1} は生成画像とクラスタ i の中心との L1 距離、 L_{km} は生成画像とクラスタ i のマハラノビス距離である。マハラノビス距離を用いることにより、分散を考慮し、よりクラスタ固有の特徴を含んだ画像を生成するように学習を行う。ただし、マハラノビス距離のみでは学習が安定しないと考えられるため、L1 距離も学習に加えることで学習の安定化を行う。

本研究では、以上述べた2つの Loss に対して、さらに MSGAN [3] で導入された MS Loss を加え、全体の Loss を以下のように定義する。

$$L_{new} = L_{ori} + \lambda_{MS} L_{MS} + \lambda_{kl1} L_{kl1} + \lambda_{km} L_{km}$$

MS Loss も同時に用いることにより、クラスタ間での多様性に加えて同一クラスタ内においても多様性を出すように学習が行われ、MS Loss を加えない場合よりも、さらに多様な画像を生成するよう学習が行われると考えられる。

3. 実験

3.1. 画像生成

本研究では pix2pix [2] の facades データセットを用いて、提案法と従来法の比較実験を行った。図2は提案法、従来法それぞれで画像生成を行った結果である。提案法については入力したクラス番号も記載した。この図より、提案法では従来法 [2,3] と比べて多様な画像生成を行えていることが確認できる。特に、MSGAN [3] では濃い色の建物があまり生成されていないが、提案法では濃いオレンジや黄色の建物画像も作ることができており、非常にバリエーション豊かな画像が生成できていることが分かる。

3.2. 定量評価

次に、複数の指標を用いた定量評価実験について説明する。本研究では、生成画像のリアリティを評価する FID、訓練画像の分布と生成画像の分布の近さを評価する NDB と JSD、生成画像の多様性を評価する LPIPS の4つの指標を用いて定量評価実験を行った。その結果を表1に

示す。この表より、提案法では FID 以外の3つの指標において最も良い数値を示していることが確認できる。このことから提案法では、真の分布に近く多様な画像生成を行うことができていることが分かる。また、FID に関しては、低下しているもののその低下度合いは小さく、生成画像のクオリティを高く保った状態で多様化に成功したといえる。



図2：画像生成結果

表1：定量評価結果

	Pix2pix [2]	MSGAN [3]	Ours
FID ↓	98.23	88.84	92.27
NDB ↓	11	11	9
JSD ↓	0.0812	0.0559	0.0300
LPIPS ↑	0.0621	0.3752	0.4444

4. 結論

本研究では、GAN においてより多様な画像を生成するための手法について検討した。具体的には、訓練画像の分布を予め推定し学習時に利用する手法を提案した。さらに従来法との比較実験を行うことで提案法の有効性を示した。今後は、異なるデータセットでの実験、クラス数やパラメータの調整、よりよい特徴ベクトルの抽出方法などについて研究する予定である。

参考文献

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Networks, *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014).
- [2] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, *CVPR* (2017).
- [3] Mao, Q., Lee, H.-Y., Tseng, H.-Y., Ma, S. and Yang, M.H.: Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis, *IEEE Conference on Computer Vision and Pattern Recognition* (2019).