

## セグメンテーションを組み合わせた DCGAN による画像の自動生成 Automatic Image Generation by DCGAN with Semantic Segmentation

浦添 和哉<sup>†</sup> 黒木 修隆<sup>†</sup> 廣瀬 哲也<sup>†</sup> 沼 昌宏<sup>†</sup>  
Kazuya Urazoe Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

### 1. はじめに

画像認識分野において、既存の小規模データセットから大規模データセットを自動生成する手段が検討されている。従来の自動生成手法として、Deep Convolutional Generative Adversarial Networks (DCGAN) が挙げられる。しかし、従来手法では、生成物体と背景の境界線が曖昧になる問題点がある。

そこで、本稿では画像の品質向上を目的とし、セグメンテーションを組み合わせた DCGAN による画像の自動生成手法を提案する。

### 2. DCGAN の概要

まず、一般的な教師なし学習 Generative Adversarial Networks (GAN) [1] について説明する。図 1 に GAN のモデルを示す。GAN は生成器 (Generator) と識別器 (Discriminator) の 2 つのネットワークを競合させて学習を行う。生成器は入力されたランダムノイズから学習画像に類似したデータを生成する。対して、識別器は入力されたデータが学習データであるか、生成器から生成されたデータであるかを判定する。

ここで、生成器は識別器を誤判定させることを目標として学習を行う。対して識別器は入力されたデータの由来を正しく判定することを目標として学習を行う。このように生成器と識別器のネットワークを互いに競合させ、学習を行うことで、学習データに特徴量的に類似した画像を生成する生成器を得ることができる。

GAN を画像データに応用するために畳み込みニューラルネットワーク (CNN) を導入したものととして DCGAN [2] が挙げられる。しかし、従来の DCGAN の生成器は全体的にぼやけた、品質の悪い画像を生成することがある。これは生成対象物と背景を切り分けて学習ができないためである。

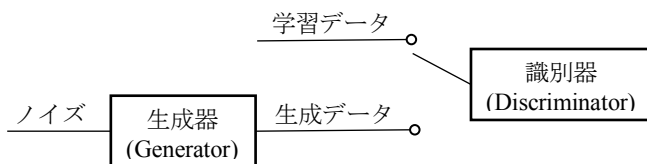


図 1 GAN のモデル

### 3. 提案手法

セグメンテーションを組み合わせた DCGAN による画像の自動生成手法を提案する。

#### 3.1 セグメンテーション

セグメンテーションとは、画像のピクセル単位でクラス分類を行う物体認識手法である。画像分類において、CNN は分類クラス数と同数の出力ユニットを持つ。セグメンテーションでは出力ユニット数が画像サイズとクラス数の積となり、各ピクセルに対し、それぞれのクラスの確率を出力する構造となっている。

セグメンテーションの一つに Fully Convolutional Networks for Semantic Segmentation (FCN) [3] と呼ばれる手法が挙げられる。FCN は一般物体認識用の CNN の後段にある全結合層を畳み込み層に置き換えることで、空間ボリュームを維持したまま出力を行うことができる。

本手法では VGG-16 に FCN を適用した voc-fcn8s によって、カラー画像 (RGB 3ch) からセグメンテーション画像 (1ch) を取得する。

#### 3.2 提案手法のネットワーク構造

図 2 に提案手法の生成器のネットワーク構造、図 3 に提案手法の識別器のネットワーク構造を示す。提案手法においては、生成器の出力層と識別器の入力層のチャンネル数を 4ch とする。生成器では 100 個のランダムノイズから自動生成された RGB 画像、および自動生成されたセグメンテーション画像を得る。一方、識別器には学習用画像とそのセグメンテーション画像を連結させた 4ch のデータ、または生成器で生成された 4ch のデータのどちらかを入力する。これにより、領域情報を考慮した学習が可能となるため、生成画像の品質向上が期待できる。

### 4. 評価実験

#### 4.1 実験内容

従来の DCGAN と、提案手法との比較実験を行った。学習用画像は人物画像データセット Labeled faces in the wild [4] から 1 万枚、画像サイズは 64×64 pixel とした。学習時のミニバッチを 50、エポック数を 1000 とした。評価項目は Inception Score とする。

<sup>†</sup> 神戸大学大学院工学研究科,  
Graduate School of Engineering, Kobe University

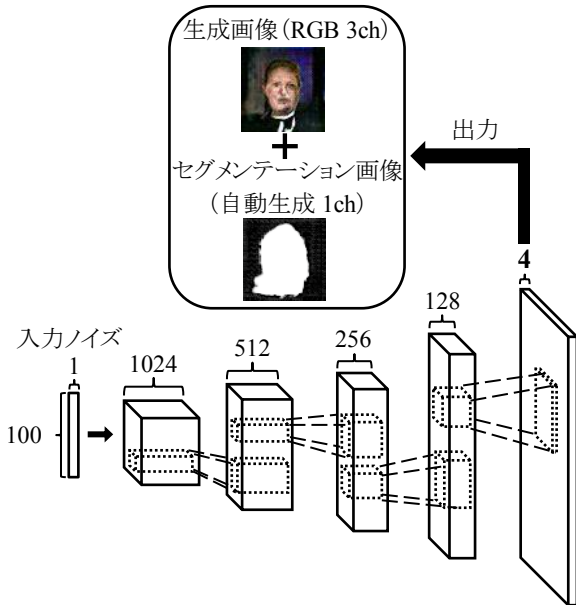


図2 提案手法の生成器のネットワーク構造

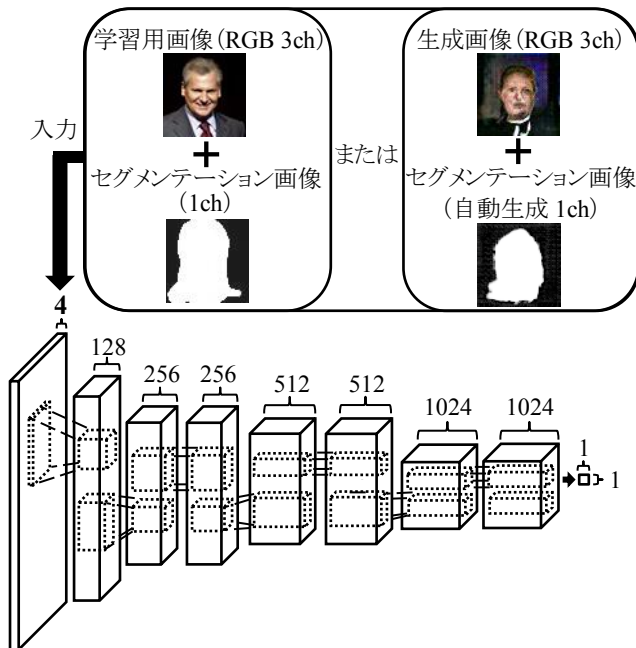


図3 提案手法の識別器のネットワーク構造

## 4.2 評価指標 (Inception Score)

Inception Score は GAN による生成画像の評価指標として用いられており、人間の主観的な品質評価と高い相関関係がある。生成画像  $x$  がクラス  $y$  と判定される確率を  $p(y|x)$ 、画像群全体に対する判定結果の確率分布の平均を  $p(y)$  とすると、Inception Score は

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

で表される。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る。

すなわち、 $p(y|x)$  は 1 つのクラスに突出し、 $p(y)$  はすべてのクラスに均等となるのが理想的な条件である。このとき、Inception Score の値が相対的に大きくなり、生成画像の品質が良好であるといえる。

表1 Inception Score の測定結果

評価指標	Inception Score
(1) オリジナルの人物画像 (10000 枚)	3.42 ± 0.04
(2) 従来手法による生成画像 (50000 枚)	2.80 ± 0.02
(3) 提案手法による生成画像 (50000 枚)	<b>2.86 ± 0.03</b>

## 4.3 実験結果と考察

表1に測定結果を示す。Inception Score において、(1) オリジナルの人物画像、(3) 提案手法、(2) 従来手法の順に高い値を得た。したがって、従来手法と比較し、提案手法による生成画像の品質がより優れていることが確認できた。

各手法において Inception Score が高かった生成画像の例を図4と図5に示す。従来手法では人物領域に背景画像が混在している生成画像があるが、提案手法では顔の輪郭が明瞭に表現されており、背景との混在が低減されていることが確認できる。このように主観的にも自然な画像が生成されていることがわかる。以上より、セグメンテーション情報を考慮した提案手法の有効性が確認できた。



図4 従来手法による生成画像



図5 提案手法による生成画像

## 5. まとめ

本稿では、自動生成画像の品質向上を目的とし、セグメンテーションを組み合わせた DCGAN による画像の生成手法を提案した。評価実験の結果、従来手法と比較して Inception Score が 0.06pt 向上し、また主観的にも自然な画像が生成されることを確認した。今後は多クラスのセグメンテーションを用いて複数の対象物が写った画像へ対応する予定である。

## 参考文献

- [1] J. 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)*, 2014.
- [2] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks", *International Conference on Learning Representations (ICLR)*, 2016.
- [3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [4] G. Huang, M. Ramesh, T. Berg, and E. Learned Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments", *Technical Report 07-49, University of Massachusetts, Amherst*, 2007.