

# 敵対的生成ネットワークを利用した高解像画像生成の研究

入江 一帆

指導教官 藤本 健司

## 1 はじめに

近年、深層学習と呼ばれる手法が、教師あり学習の画像認識や音声認識などの、分類、認識の分野において成果を上げている。さらには新たな分野として、敵対的生成ネットワークを用いた画像生成モデルも考案され、教師なし学習での画像生成に成功している。しかし、高解像度の画像を生成したいとき、現在の生成モデルでは、学習コストが増大してしまい、出力精度も低下するため、生成画像をそのまま利用するには有用なレベルであるといえない。そこで本研究では、高解像度の画像を生成するための手法を比較検討し、生成モデルの性能を向上させることで、有用な画像生成を実現し、創造活動の発展に寄与することを目的とした。

## 2 研究内容

まず、Deep Convolutional Generative Adversarial Networks[1](以下、DCGAN)と、多段階 GAN[2]を用いて画像を生成した。その後、Super Resolution Convolutional Neural Network[3](以下、SRCNN)を用いて超解像を行った。

### 2.1 DCGAN

DCGAN は、図 1 のように、Generator(以下、G)と Discriminator(以下、D)の 2 つのネットワークを組み合わせる。G は数次元ベクトルを入力として、畳み込みと Up Sampling を繰り返し画像  $Y_g$  を生成する。D は畳み込みを用いて入力画像が G で生成した画像  $Y_g$  であるのか、訓練画像  $Y_t$  であるのかを判別する。D は判定精度を向上させることを目指し、G は D が誤判定することを目指して学習を進める。これらを繰り返すことにより、G が質の高い画像を生成できるようになる。

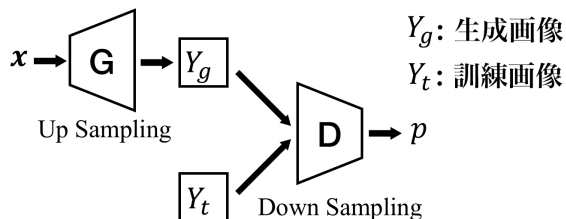


図 1 DCGAN の原理図

### 2.2 多段階 GAN

多段階 GAN は、複数の DCGAN を組み合わせて高解像度の画像を生成する手法である。高解像度の画像を生成しようとしたとき、1 段の DCGAN で画像を生成するよりも、多段階 GAN で生成するほうが、良い結果を得られることが知られている。図 2 に 2 段のときの多段階 GAN の原理図を示す。

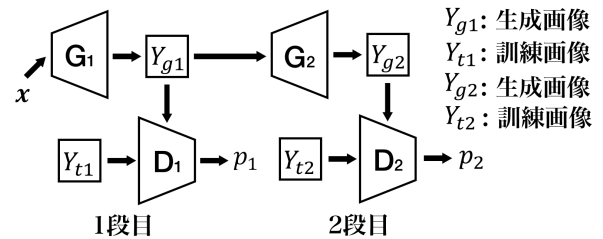


図 2 多段階 GAN の原理図

### 2.3 SRCNN

SRCNN は、低解像度の画像を拡大してから畳み込みを用いて補正を行い、高解像度の画像を推定することによって、超解像を実現するネットワークである。

## 3 研究結果

Labeled Faces in the Wild で配布されている顔画像データセットから無作為に 4096 枚抽出し、それらを出力画像サイズ  $h \times h$  [pixels] にリサイズして、訓練データとした。

### 3.1 DCGAN

出力画像サイズ  $h \times h$  を  $32 \times 32$ ,  $64 \times 64$ ,  $128 \times 128$  [pixels] として DCGAN モデルを訓練した。  $h \times h = 64 \times 64$  のときの生成画像の一例を図 3 に示す。



図 3 DCGAN 生成画像の一例

### 3.2 多段階 GAN

出力画像サイズ  $h \times h = 128 \times 128$ [pixels] として DCGAN モデルを訓練した. 生成画像の一例を図 4 に示す.



図 4 多段階 GAN 生成画像の一例

### 3.3 SRCNN

入力画像サイズ  $h \times h = 64 \times 64$ [pixels], 出力画像サイズ  $2h \times 2h = 128 \times 128$ [pixels] として, SRCNN モデルを訓練した. 超解像結果の一例を図 5 に示す.

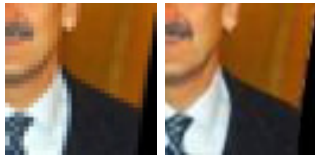


図 5 超解像結果の一例

### 3.4 性能比較

画像評価において定量的評価と人間の主観の間には相違があることが知られている. 本研究では利用目的に基づき、人間の主観による評価を得るために、10代から50代の男女50人に対してアンケート調査を行った.

GAN を高解像度に対応させる手法として以下の2つの結果について、学習反復数を共通としたときと、学習時間を共通としたときを比較した.

- 1) 多段階 GAN を用いて  $128 \times 128$ [pixels] の画像を作成する方法 (以下, 手法 A).
- 2) DCGAN により  $64 \times 64$ [pixels] の画像を作成し、それを SRCNN により  $128 \times 128$ [pixels] に拡大する方法 (以下, 手法 B).

手法 A の多段階 GAN と、手法 B の DCGAN の学習反復数を 6000[epochs] 共通としたときの結果を表 1 に示す. このとき手法 B の SRCNN の学習反復数は 2000[epochs] とした.

表 1 反復数共通のときの結果

方法	選択率 [%]	反復数 [epochs]	時間 [hrs]
手法 A	50.0	6000	113.8
手法 B	50.0	6000, 2000	75.2

手法 A と手法 B の学習時間を約 75[hours] 共通としたときの結果を表 2 に示す.

表 2 学習時間共通のときの結果

方法	選択率 [%]	反復数 [epochs]	時間 [hrs]
手法 A	22.7	4000	76.8
手法 B	77.3	6000, 2000	75.2

## 4 結論

DCGAN では出力画像の解像度が高くなるに伴い、学習時間も跳ね上がる. さらに、ある一定の解像度を超えると画質が向上しにくくなる. 従って、一段の DCGAN では出力精度の向上に限界があり、最も性能を発揮できるのは  $64 \times 64$ [pixels] 程度であると考えられる.

高解像度の画像を出力したいとき、多段階 GAN を用いると質の高い画像を得られることがわかった. しかし、学習コストが跳ね上がる問題は維持されたままであった.

DCGAN で出力した低解像度の画像を SRCNN で超解像することにより、質の高い高解像度の画像を出力できることがわかった. また、同じ反復数では多段階 GAN よりも短い学習時間で訓練が完了することも確認できた.

アンケート調査の結果より得た、人間の感性による評価では、多段階 GAN を用いて高解像度の画像を出力するよりも、DCGAN と SRCNN を組み合わせて出力する方が、優れた結果を得られることがわかった.

## 参考文献

- [1] Alec Radford et al. : "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" (2015).
- [2] Han Zhang et al. : "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks" (2016).
- [3] Chao Dong et al. : "Image Super-Resolution Using Deep Convolutional Networks" (2014).