

# 敵対的生成ネットワークでデカイ画像を生成してみた結果 www

入江 一帆

指導教官 藤本 健司

## 1 はじめに

最近、深層学習とかいう手法が教師あり学習の画像認識とか音声認識とかの分類認識の分野で成果上げてるんやわ。しかも流行りの敵対的生成ネットワークを使った画像生成モデルが考案されて教師無しでの画像生成に成功したんよ。でもな、でっかい画像を生成したいな思ったとき、今の生成モデルやったら学習コストがドカーンなって出力もグチャーなってもうて生成画像をそのまま使えるレベルではないねん。せやから、この研究ではでっかい画像を生成するための手法を比べたりしてめっちゃ強い生成モデルを作ってキレイな画像生成を実現したことで、おもしろいことが増えたらええなと思ってんねん。

## 2 やったこと

まず Deep Convolutional Generative Adversarial Networks[1](これからは DCGAN) と多段階 GAN[2] を使って画像を生成してん。ほんで Super Resolution Convolutional Neural Network[3](これからは SRCNN) を使って超解像をやったで。

### 2.1 DCGAN

DCGAN っていうのはな図 1 みたいに Generator(これからは G) と Discriminator(これからは D) の 2 つのネットワークをくっつけて作られるねん。G は数次元ベクトルをいれたったら畳み込みと Up Sampling をしまくって画像  $Y_g$  を作りよる。D は畳み込みを使って入力画像が G で作った画像  $Y_g$  なんか訓練画像  $Y_t$  なんかを判別しよる。D はちゃんと判定できるように G は D が間違ってしまうことを目指して学習するねん。これを何回もやることで G がええ感じの画像を生成するようになるんやわ。

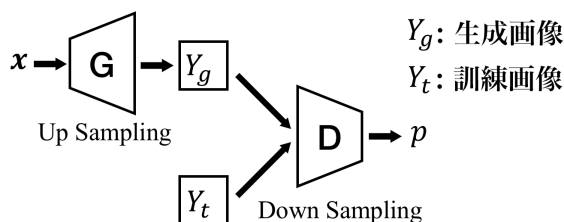


図 1 サルでもわかる DCGAN

### 2.2 多段階 GAN

多段階 GAN っていうのはな、DCGAN を何個も何個も使ってでっかい画像を生成する手法のことや。でっかい画像を作りたいなと思ったら 1 段の DCGAN でやるよりも多段階 GAN でやったほうが全然ええ結果になるって言われてるんよ。図 2 に 2 段のときの多段階 GAN の原理図を見したろ。

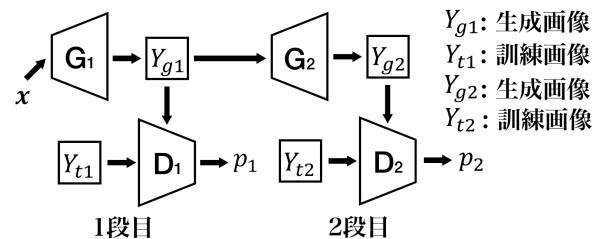


図 2 サルでもわかる多段階 GAN

### 2.3 SRCNN

SRCNN っていうのはな、ちっこい画像をでっかくしてから畳み込み使って手直してでっかい画像っぽく誤魔化すことで超解像をやりよるネットワークのことや。

## 3 やった結果

Labeled Faces in the Wild でばらまかれとる顔面の写真から適当に 4096 枚選んでそいつらを出力画像サイズ  $h \times h$ [pixels] に縮めて訓練データにしたわ。

### 3.1 DCGAN

出力画像サイズ  $h \times h$  を  $32 \times 32 \times 64 \times 128$ [pixels] にして DCGAN モデルを訓練したで。  $h \times h = 64 \times 64$  にしたときの生成画像をちよつとだけ図 3 に見せたら。



図 3 【マル秘】 DCGAN 生成画像

### 3.2 多段階 GAN

出力画像サイズ  $h \times h = 128 \times 128$ [pixels] にして DCGAN モデルを訓練した。生成画像をちょっとだけ図 4 に見せたら。



図 4 【閲覧注意】多段階 GAN 生成画像

### 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] にして SRCNN モデルを訓練した。超解像結果の一例をちょっとだけ図 5 に見せたら。

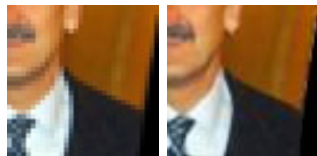


図 5 【衝撃】超解像結果

### 3.4 性能比較

生成した画像がええんか悪いんかわからんからアンケート調査してん。しっかり数字で評価するんと人間が評価するんとは全然ちやうから見た目による判定にしたわ。

GAN を高解像度もいけるようにする手法ってことで下の 2 つで学習反復数を一緒にしたときと学習時間を一緒にしたときのを比べた。

- 1) 多段階 GAN 使って  $128 \times 128$ [pixels] の画像を作る方法 (これからは手法 A)。
- 2) DCGAN で  $64 \times 64$ [pixels] の画像を作ってそれを SRCNN で  $128 \times 128$ [pixels] にでっかくする方法 (これからは手法 B)。

手法 A の多段階 GAN と手法 B の DCGAN の学習反復数を 6000[epochs] にしたときの結果を表 1 に書いとく。こんとき手法 B の SRCNN の学習反復数は 2000[epochs] やったわ。

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 75[hours] ぐらいにしたときの結果を表 2 に書いとく。

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 わかったこと

DCGAN は出力画像をでっかくしたら学習時間もバーンてなる。しかも画像の大きさが一線を超えるとグチャーてなってしまう。せやから一段の DCGAN ではキレイなの生成するのに限界があるねんよ。最高にやってくれるのは  $64 \times 64$ [pixels] ぐらいのときなんやろなって感じやわ。

でっかい画像欲しいなって思ったら多段階 GAN 使ったらええ画像出せるってわかったで。でもな学習コストがドッカーンてなる問題はそのままやってん。

DCGAN で作ったちっこい画像を SRCNN ででっかくすることでキレイな画像を作れることがわかったわ。ほんで同じ反復数やったら多段階 GAN よりもはよ訓練が終わるっていうこともわかったで。

アンケート調査の結果でわかった人間の独断と偏見による評価やと多段階 GAN を使ってでっかい画像を作るんやったら DCGAN と SRCNN をくっつけて作った方がええ結果になるねんて。

## ソース

- [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).