

# フーリエ変換を用いた GAN による人物画像鮮明化

2016SS061 佐伯宥紀

指導教員：小市俊悟

## 1 はじめに

近年、カメラの解像度が急速に上がっている。そのため、購入時期やカメラの種類の違いによって同じ景色・被写体であっても解像度の差が大きい。ハードウェア的に解像度を上げることは今後、ますます難しくなることが予想される。したがって、ソフトウェア的に撮影後の写真の解像度を上げることができるようになることが期待される。このような技術は例えば医療現場のレントゲン写真において今までは設備に起因する解像度の限界により見ることができなかったものまで見ることを可能にする。また、衛星写真に適応すれば地形をより正確に把握ができるようになり、ハザードマップの精度を向上させることにもつながるであろう。

本研究では、このようなソフトウェア的な解像度の向上を機械学習、特に GAN と呼ばれる手法を用いて実現する。その際に、既存方法の問題点を克服するために、フーリエ変換を利用することを試みる。なお、機械学習によるソフトウェア的な画像の鮮明化とは、入力画像では不明瞭な境界をはっきりさせるといこと、強調させるということであり、超解像とも呼ばれる。その際の境界は物体と物体であったり、顔を構成する部位である目と瞼であったりする。注意すべきは、そのような境界は学習に基づいて予測されるものであるから実際には必ずしも正しいとは限らない。つまり、機械学習で鮮明化された画像は、実際の被写体とは異なり得る。図 1 は一番左が元の高解像度画像であり、真ん中がそれをダウンサンプリングし低解像度にしたものである。一番右は低解像度画像を元に GAN が鮮明化を行った結果である。



図 1 GAN による画像鮮明化の例

## 2 GAN と SRGAN

### 2.1 GAN について

本研究で用いる機械学習モデルは敵対的生成ネットワーク、英語の略称で GAN と呼ばれるものである。GAN とは図 2 のように 2 つのニューラルネットワークを競わせながら学習させるモデルである。2 つのニューラルネットワークのうち 1 つは Generator と呼ばれ、ランダムノイ

ズ入りのデータから所望のデータを生成する。Generator は生成したいデータの特徴の種にランダムノイズをあえて加え、このノイズを所望のデータに近づけるようにマッピングする。もう 1 つのネットワークは Discriminator と呼ばれ、Generator が生成した偽物のデータまたは本物のデータを入力として受け取り、その真偽を判定する。Generator が Discriminator に真偽を見破られない程度に学習できたとき、Generator は実在しないが本物のようなデータを生成することができることになる。

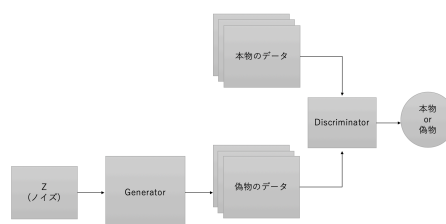


図 2 GAN の構造

### 2.2 SRGAN について

本研究では、GAN の中でも SRGAN[1] と呼ばれる手法を用いる。SRGAN の特徴の一つは、Generator に ResNet[2] と呼ばれる構造を持ったニューラルネットワークを利用することである。ResNet とは、2015 年に He らによって提案されたものであるが、従来のネットワークと比較して、層が非常に深いこと (100 層以上) がその特徴である。層が深いネットワークは、学習がうまくいけば表現力が高いため画像処理に有効であることが経験的に知られている。一方で層が深いことでいわゆる勾配消失が起き、学習に支障をきたしやすいため知られている。ResNet ではこの懸念点を、shortcut connection と呼ばれる構造を導入することで回避している。

## 3 SRGAN の問題点とフーリエ変換

### 3.1 SRGAN の問題点

SRGAN を用いて画像の鮮明化を実際に行なったところ、周期的なパターンを持つようなノイズが入りやすいことに気が付いた。図 3 は、そのような例であり、左が Generator の入力となる画像で、右が生成された画像である。このようなパターンを持つノイズは、不自然さを与えるので、取り除くことによって、より鮮明な画像を得ることができるのではないかと考えた。本研究では、SRGAN を利用することを前提に、このようなノイズの除去を目指す。



図3 元になった低解像度の画像(左)とSRGANによって生成された画像(右)

### 3.2 フーリエ変換

画像データは適当な変換, 例えば, グレースケールに変換することにより, 2次元の数値データとみなすことができるが, それに2次元(離散)フーリエ変換を適用すれば, そのデータをいくつかの平面波に分解することができる. その際, 元の画像に周期的な特徴がある場合, 分解後のデータには, 特定の波長の平面波が相対的に多く含まれることが知られている. フーリエ変換のこのような性質を利用すれば, SRGANが生成する画像の周期性を低減できるのではないかと考えた.

## 4 フーリエ変換を利用した生成方法

RGBの3チャンネルからなる低解像度画像 $I$ を, チャンネルごとにフーリエ変換した画像を $I_F$ とする.  $I_F$ の各要素は複素数からなるので, 実部 $I_{Fr}$ と虚部 $I_{Fi}$ に分離し,  $I_F = (I_{Fr}, I_{Fi})$ と表す. まず, 層構造がSRGANで用いられるものと同一であるような新たなGeneratorとして,  $F_r, F_i$ の二つを用意し, 各低解像度画像 $I$ を元に, フーリエ変換した $I_F$ を構成したのち, Generator  $F_r$ は $I_{Fr}$ を入力として $F_r(I_{Fr})$ を, Generator  $F_i$ は $I_{Fi}$ を入力として $F_i(I_{Fi})$ を出力するようにSRGANを改めた. Generator  $F_r$ と $F_i$ によって, 出力された $F_r(I_{Fr})$ と $F_i(I_{Fi})$ からフーリエ逆変換を用いて, 通常の画像として, あとはSRGANと同じ方法でGeneratorの損失を計算した. このような設定で学習を行ったところ, Generator  $F_r$ と $F_i$ によって生成される画像は, 1/4に縮小された顔が $4 \times 4$ に並んだ画像となった. そこで, SRGANで用いたGeneratorから, Sub-pixel Convolutionを取り除いたGenerator  $F'_r, F'_i$ を用意し, 必要な大きさの画像を得るためにGenerator  $F'_r, F'_i$ に入力する画像を入力前にバイリニア補間により縦横それぞれ4倍に拡大してから入力するようにした. これにより, Generatorのネットワークの規模が大きくなったが, 学習が進むに連れて, 通常の顔画像が生成されるようになった.

## 5 数値指標について

フーリエ変換を利用した生成方法を評価するために画質の優劣の評価にも用いられる3つの指標MSE, PSNR, SSIMを用いることにした. それらを用いた評価結果を表1と2に示す. SSIMに着目すれば元画像によっては, よ

り良い画像が生成できた.

表1 画像データA

データ種	MSE	PSNR	SSIM
SRGAN	103.96	27.96	0.83
SRGANF	783.64	19.19	0.71
SRGANGF	300.11	23.36	0.81

表2 画像データB

データ種	MSE	PSNR	SSIM
SRGAN	157.13	26.17	0.88
SRGANF	784.77	19.19	0.81
SRGANGF	245.75	24.23	0.89

## 6 おわりに

本研究を通して既存のSRGANの問題点と思われた周期的なノイズは図4に示す対数パワースペクトル密度からもわかるように除去できたが, 数値指標に大きな変化がないこともわかった. 主観的になるが, 人の目にはより良いと感じられるものもあった. 学習回数を増やすこと以外に画像の数値指標を向上させる方法として二つ考えられる. そのうちの一つは生成画像全体の色相を変化させることである. これをすることで生成画像が白みがかかっていることを抑えることができより良い結果を得ることができるかもしれない.

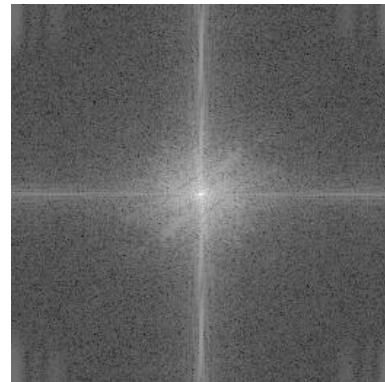


図4 SRGANGFにより生成された画像をフーリエ変換し, その対数パワースペクトル密度をグレースケール表示したもの

## 参考文献

- [1] Christian Ledig, et al.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv:1609.04802v5[cs.CV], 2017.
- [2] Kaiming He, et al.: Deep residual learning for image recognition. arXiv:1512.03385[cs.CV], 2015.