

特徴識別器を組み込んだ GAN による 複数の特徴を持つ顔画像の生成

2018SS068 内田翔太郎

指導教員：小市俊悟

1 はじめに

本研究では、敵対的生成ネットワーク (GAN) [2] の手法を用いて、指定された複数の特徴を持つ顔画像の生成を行う。人の顔は複数の特徴を持つことが普通であるが、所望の特徴が多いほど、学習データが少なくなるので、それらの特徴すべてを持つ顔画像の生成は容易ではない。本研究では、同様の問題に取り組んだ卒業研究 [1] を引き継ぎ、新たな試みとして、指定された特徴を正確に識別できる識別器 (特徴識別器) を作成し、それを学習の際に利用する方法を提案する。また、昨今では一定の学習が済んだ深層ネットワークに層を追加してより能力が高い深層ネットワークを作成する方法が注目されている。本研究でも層を追加することで、複数の特徴を持つ顔画像を効率的に生成できるかどうかを検証する。

2 手法の説明

提案手法の効果などを評価するために、はじめに、指定された一つの特徴を備える顔画像の生成を行う。その結果も踏まえて、指定された複数の特徴を備える顔画像の生成を行う。ただし、本研究では、指定された二つの特徴のみを扱うこととする。GAN とは、Discriminator と Generator と呼ばれる二つの深層ネットワークを用意し、それらが競合的な目的関数を持って学習することで、最終的に Generator に所望の機能を持たせるものである。下記では、 $x_i (i = 1, 2, \dots, n)$ は本物のデータを、 $z_i (i = 1, 2, \dots, m)$ はノイズを表す。

2.1 一つの特徴を持つ顔画像の生成

事前準備 指定された一つの特徴を備える顔画像を正確に識別できる識別器 D_2 を作成する。

pattern 1 Conditional GAN と同じ構造の深層ネットワークを用いて、Discriminator D_1 、Generator G_1 の学習を進め、 G_1 が指定された一つの特徴を備える顔画像を生成できるようにする。

pattern 2 pattern 1 の学習に、 D_2 を組み込み、 G_1 が指定された一つの特徴を備える顔画像を生成できるようにする。

pattern 3 pattern1 で得られた G_1 を元に、新たな層を追加した Generator G_4 を用意し、 G_4 の学習を進め、顔画像を生成できるようにする。

pattern 4 pattern 2 と pattern 3 を組み合わせて、 G_4 が指定された一つの特徴を備える顔画像を生成できるようにする。

pattern 1 における D_1 の目的関数は、

$$\sum_{i=1}^m [(D_1(x_i) - 1)^2 + (D_1(G_1(z_i)) - 0)^2]$$

である。真偽を正しく識別できるほど、この目的関数は小さくなる。一方、pattern 1 における G_1 の目的関数は、

$$\sum_{i=1}^m (D_1(G_1(z_i)) - 1)^2$$

である。 D_1 を騙すほど、この目的関数は小さくなる。

pattern 2 における D_1 の目的関数は、

$$\frac{1}{\alpha_1 + \alpha_2 + \alpha_3} \times \left\{ \alpha_1 \sum_{i=1}^m (D_1(x_i) - 1)^2 + \alpha_2 \sum_{i=1}^m (D_1(G_1(z_i)) - 0)^2 + \alpha_3 \sum_{i=1}^m (D_1(G_1(z_i)) - D_2(G_1(z_i)))^2 \right\}$$

である。真偽を正しく識別できるほど、この目的関数は小さくなる。分子の第 3 項として、 D_1 には G_1 が生成する画像について D_2 と同じ判定することが求められている。 D_2 は G_1 が生成する画像に特徴が表れているか否かを判定するため、この項の導入により、 D_1 も G_1 が生成する画像に特徴が表れているか否かを重視するようになることが期待できる。一方、pattern 2 における G_1 の目的関数は、

$$\frac{1}{\beta_1 + \beta_2} \times \left\{ \beta_1 \sum_{i=1}^m (D_1(G_1(z_i)) - 1)^2 + \beta_2 \sum_{i=1}^m (D_2(G_1(z_i)) - 1)^2 \right\}$$

である。 D_1 を騙すほど、この目的関数は小さくなる。分子の第 2 項として、 D_2 の識別結果が組み込まれている。 D_1 を騙そうとする点は通常の GAN と同じであるが、 D_2 が 1 と出力することは入力画像が特徴を備えることを意味するので、同時に D_2 に G_1 が生成する画像が特徴を備えた画像であると識別させることが目的となっている。また、pattern 3, 4 における D_1 、 G_1 の目的関数は、pattern 1, 2 における D_1 、 G_1 の目的関数にある G_1 を G_4 に変更したものである。

2.2 複数の特徴を持つ顔画像の生成

二つの特徴を持つ画像を生成できる Generator G_4 を作成する手順は以下の step1 から 4 である。

step 1 特徴 1 について、正確に識別できる識別器 D_2 を作成する。

step 2 特徴 2 について、正確に識別できる識別器 D_3 を作成する。

step 3 GAN を用いて、特徴 1 を持つ顔画像を生成できる Generator G_1 を作成する。 G_1 が生成した顔画像を step 1 で作成した D_2 を用いて識別し、その判定結果も G_1 の学習に利用する。

step 4 step 3 において、作成した G_1 に新たに層を追加して、特徴 1, 2 を持つ顔画像を生成できる Generator G_4 を GAN を用いて作成する。その際に、 G_4 が生成した顔画像を step 1 で作成した D_2 と step 2 で作成した D_3 を用いて識別し、その判定結果も G_4 の学習に利用する。

step 3 で用いる D_1 と G_1 の目的関数は、一つの特徴を持つ顔画像の生成における pattern 2 と同じものを利用する。

step 4 における D_4 の目的関数は、

$$\frac{1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \times \left\{ \alpha_1 \sum_{i=1}^m (D_4(x_i) - 1)^2 + \alpha_2 \sum_{i=1}^m (D_4(G_4(z_i)) - D_2(G_4(z_i)))^2 + \alpha_3 \sum_{i=1}^m (D_4(G_4(z_i)) - D_3(G_4(z_i)))^2 + \alpha_4 \sum_{i=1}^m (D_4(G_4(z_i)) - 0)^2 \right\}$$

である。真偽を正しく識別できるほど、この目的関数は小さくなる。step 4 の D_4 の目的関数には、分子の第 2 項、第 3 項として、識別器 D_2 , D_3 の識別結果が組み込まれている。 D_2 , D_3 が特徴の有無を識別できることから、これらの項の導入により、 D_4 も特徴の有無を重視ようになることが期待される。一方、step 4 における G_4 の目的関数は、

$$\frac{1}{\beta_1 + \beta_2 + \beta_3} \times \left\{ \beta_1 \sum_{i=1}^m (D_4(G_4(z_i)) - 1)^2 + \beta_2 \sum_{i=1}^m (D_2(G_4(z_i)) - 1)^2 + \beta_3 \sum_{i=1}^m (D_3(G_4(z_i)) - 1)^2 \right\}$$

である。 D_4 を騙すほど、この目的関数は小さくなる。分子の第 2 項、第 3 項として、識別器 D_2 , D_3 の識別結果が組み込まれている。 D_4 を騙そうとする点は通常の GAN と同じであるが、 D_2 と D_3 が 1 と出力することは入力画像が特徴を備えることを意味するので、同時に D_2 , D_3 に G_4 が生成する画像が各々の特徴を備えた画像であると識別させることが目的となっている。

2.3 目的関数で用いる重み

2.2 節で紹介した目的関数で用いた重みは具体的には次のようにした。下記では、学習の繰り返し回数（エポック

数）を n として、その限度を N で表す。重みは n に応じて変化させている。複数の特徴を持つ顔画像の生成における step 3 の D_1 の学習で用いた重みは $N = 400$ として、次である。

$$\begin{aligned} \alpha_1 &= \alpha_1(n) = 1 & (n \geq 0) \\ \alpha_2 &= \alpha_2(n) = 1 & (n \geq 0) \\ \alpha_3 &= \alpha_3(n) = \begin{cases} 0 & (n < 50) \\ \alpha_3(n-1) & (n \geq 50, \text{奇数}) \\ 0.0001 & \\ \times \left(1.0 - \frac{n-49}{N-49}\right) & \\ + 0.5 \times \frac{n-49}{N-49} & (n \geq 50, \text{偶数}) \end{cases} \end{aligned}$$

同じく step 3 の G_1 の学習で用いた重みは $N = 400$ として、次である。

$$\begin{aligned} \beta_1 &= \beta_1(n) = 1 & (n \geq 0) \\ \beta_2 &= \beta_2(n) = \begin{cases} 0.0001 & (n < 50) \\ \beta_2(n-1) & (n \geq 50, \text{奇数}) \\ 0.0001 \times \left(1.0 - \frac{n-49}{N-49}\right) & \\ + 0.5 \times \frac{n-49}{N-49} & (n \geq 50, \text{偶数}) \end{cases} \end{aligned}$$

step 4 の重みについては省略するが、step 3 と同様に n に応じて変化させている。

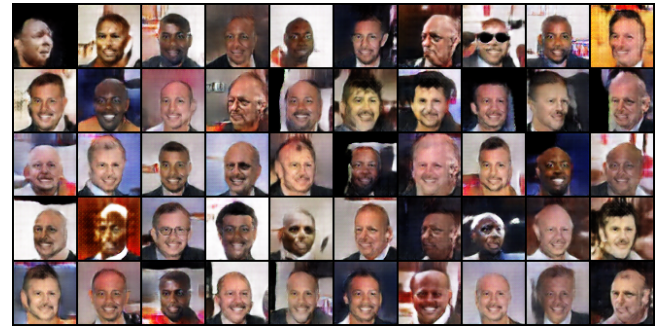


図 1 髪とスキンヘッドという特徴を持つ顔画像の生成結果

3 おわりに

本研究で提案した手法により得られた生成画像が図 1 である。この結果からも二つの特徴を持つ顔画像の生成におおむね成功したと言えるのではないかと考える。学習に特徴識別器を組み込むことや、一つの特徴を学習した Generator に層を追加することは、画像の質を向上させることがわかった。

参考文献

- [1] 杉崎智仁：『GAN を用いた複数の特徴を持つ顔画像の生成』。2019 年度南山大学理工学部システム数理学科卒業研究，2020
- [2] 宮本圭一郎，毛利拓也，大川洋平：『PyTorch ニューラルネットワーク実装ハンドブック』。秀和システム，東京，2019。