

# 敵対的生成ネットワークを用いた画像生成に関する研究

2018SE084 鈴木一路 2018SE086 鈴木祐人

指導教員：野呂昌満

## 1 はじめに

敵対的生成ネットワーク (GAN: Generative Adversarial Networks)[1] の実問題を考えた場合、そのデータ量の確保が第一の課題となっている。

Data Augmentation(以下, DA)とは、その課題に対する解を与える研究である。DAは既存のデータに変換を加えてデータ量を増やす手法である。例えば医療における病理診断に用いられる。病理ガラススライドを少し回転させても病院医は同じ癌と捉えることができるが、AIは別のものと判断する。画像を回転させたり、位置を少しずらしたりすることで別の画像として学習データを水増しすることができる [2]。

DAは画像分類で広く使われ、精度向上に大きく貢献してきたが、GANの分野ではあまり研究されてこなかった。GANにDAを適用した例はあるが、どのDAがGANにとって効果的なのかは具体的に示されていない。

本研究の目的は、GANの訓練において、適切なDAの方法を提案することである。この際、GANは森野ら [3] の一般的な構成を用いる。GANの構成は変えずに、異なるDA方法によってGANの訓練に有効であるか実験する。

本研究の課題は以下の通りである。

- (1)DAを適用したデータセット作成方法の提案
- (2)提案方法によるGANへの適用と評価

これらを解決することで、GAN訓練の精度向上だけでなく、データ収集にかかる時間や労力の削減が期待される。

## 2 関連研究

### 2.1 GAN

敵対的生成ネットワークとは機械学習技術の一種で、同時に訓練される2つのモデルからなる。1つ目(生成器: Generator)は偽のデータを生成し、もう1つ(識別器: Discriminator)は本物のデータを偽のデータと見分けるように訓練される。

生成器の目標は、訓練データ内に含まれる本物データと見分けがつかないような偽のデータを作り出すことである。識別器の目標は、訓練データに含まれる本物のデータと、生成器が作り出した偽のデータとを区別することである。

図1にGAN構成と訓練の概略を示す。GANの訓練は識別器と生成器の訓練を交互に繰り返すことで、2つのネットワークが互いに競い合うように訓練される。

#### 1. 識別器の訓練

- a. 訓練データからランダムに本物のサンプル  $x$  を取り出す

- b. 新しい乱数ベクトル  $z$  を生成し、生成器ネットワークを用いて偽のサンプル  $x^*$  を生成
  - c. 識別器ネットワークを使って、 $x$  と  $x^*$  を分類
  - d. 分類誤差を計算し、全誤差を誤差逆伝播法により、識別器の訓練可能なパラメータを更新、分類誤差を最小化
- #### 2. 生成器の訓練
- a. 新しい乱数ベクトル  $z$  を生成し、生成器ネットワークを用いて偽のサンプル  $x^*$  を生成
  - b. 識別器ネットワークを使って、 $x^*$  が本物かどうかを判定
  - c. 分類誤差を計算し、全誤差を誤差逆伝播法により、生成器の訓練可能なパラメータを更新、識別器の誤差を最大化

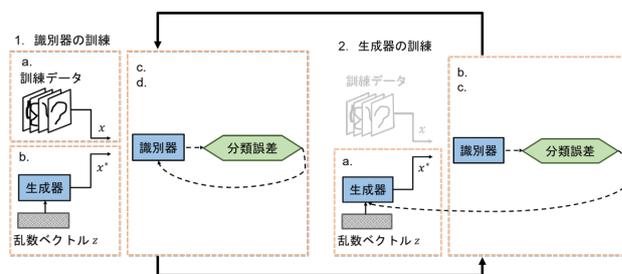


図1 GAN構成と訓練の概略

### 2.2 Data Augmentation(DA)

Data Augmentationとは既存のデータから新しく作成された合成データを追加することによって、データ数を水増しする技術である。DAには、トレーニングデータセットのサイズと品質を向上させる一連の手法が含まれており、それらを使用してより優れたディープラーニングモデルを構築できる。

### 2.3 Inception v3

Inception v3は深さが48層の畳み込みニューラルネットワークである。120万のRGB画像で構成された事前学習済みのネットワークをImageNetデータベースから読み込みできる。画像  $x$  が与えられた場合、確率のベクトル  $p(y|x) \in [0, 1]^{1000}$  の形式でクラスラベル  $y$  が出力される。これはネットワークが各クラスに割り当てる確率を示す。

### 2.4 Inception Score(IS)

Inception Score[4]とはGANにおける評価指標の1つである。ISはImageNetで事前トレーニングされたInception v3を使用する。ISは2つの分布の差異を表すKLダイバージェンスを用いて計算される。 $x_i$ を画像( $i$ は画像のインデックスで総数は $N$ とする)、 $y$ をラベルとするとISは式(1)のように表される。

$p(y|x_i)$  は個々の生成画像の認識結果の確率を、 $p(y)$  は生成画像全体の認識結果の確率の平均を表し、これらは分布に従うとすると図2のようになる。

$$IS = \exp\left(\frac{1}{N} \sum_i \text{KL}(p(y|x_i)||p(y))\right) \quad (1)$$

入力画像が特定のクラスの場合、2つの分布間の距離が小さいほど良い生成画像ができていると言えるので、ISが小さいほど性能が高いということになる。すなわち本スコアでは以下の2つの観点を満たすほど生成スコアが良いとされている。

- (1) 画像識別器が識別しやすいこと
- (2) 識別されるラベルが特定のクラスに集中していること

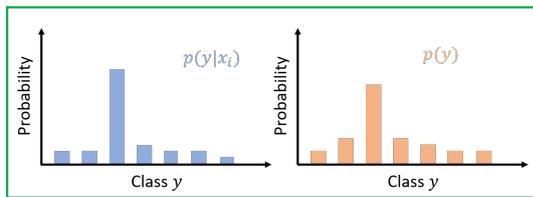


図2 理想的な2つの確率分布の例

本研究では、GANによって生成された画像を、ISで画像の良さを評価する。

### 2.5 関連研究を踏まえて

DAを用いたGANの研究はされてきたが、元のデータセットとDA後の最終的なデータセットの比率が、最適なパフォーマンスのモデルになるという保証はない[5]。

本研究では、関連研究を踏まえ、DAを適用したデータセットの提案と評価を課題とする。

### 3 本研究課題における問題解決のアプローチ

本研究では、Data Augmentationにコントラスト、ネガポジ反転、左右反転、上下反転、GANで生成した画像(以下、生成画像)をデータセットにした。DAにより、GANの訓練が十分に行えるかどうか予測できなかったため、図3のDA方策を実施した。



図3 5種類のData Augmentation

以下のステップを踏まえ、アプローチを行う。

条件：集まったデータ(例：CelebA)を1000枚とする。

1. CelebAに左右反転やネガポジ反転などのDAを適用
2. CelebA1000枚の画像をGANで生成
3. 1と2の中から1000枚の画像を抽出
4. CelebA 1000枚と3のデータを合わせた2000枚のデータセット作成

5. 4のデータセットを用いてGANで画像生成、それをISで評価

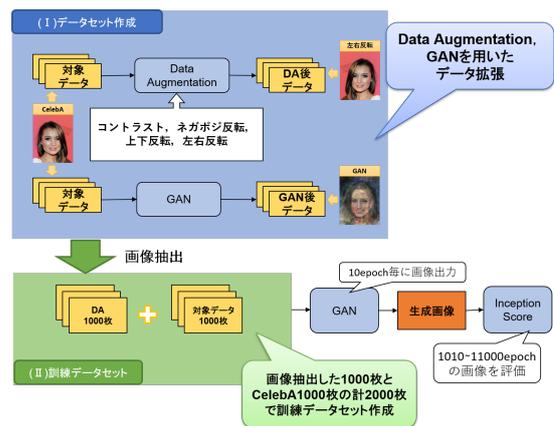


図4 アプローチ

## 4 実験内容と結果

### 4.1 データセット

CelebA[6]は有名人の顔画像を178×218ピクセルで202,599枚集めたデータセットである。その中から1000枚画像を抽出し、データセットに用いた。

ポメラニアンはThe Oxford-IIIT Pet Dataset[7]とGoogle画像検索から1,000枚集めたデータセットである。これらの取得した画像を128×128にリサイズして適用した。

### 4.2 対象データへの実装と評価

対象データはCelebAとポメラニアンとする。GANによって生成された画像はInception Scoreによって評価した。

本研究では、1種類のDAを施した1000枚のデータセットと2種類のDAを1000枚のうち1:1の比率で作成したデータセットで実験を行う。後者の条件から、DAの枚数を1000枚から2000枚に増やした場合の実験を行う。以下に実装と評価を示す。

(1) CelebA1000枚、1種類のDA1000枚の計2000枚

(1)の生成結果を図5に示す。GANによって生成された画像1000枚を抽出し、ISで計算した結果を図6に示す。他の生成された画像も同様に計算した。

ISは、ネガポジ反転が最も高い結果となった。ネガポジ反転の画像を見ると人間の顔にはふさわしくない色が見られた。上下反転はDAなしよりISは高いが、首が2つあるような画像が生成された。



図5 1種類のDAの生成結果

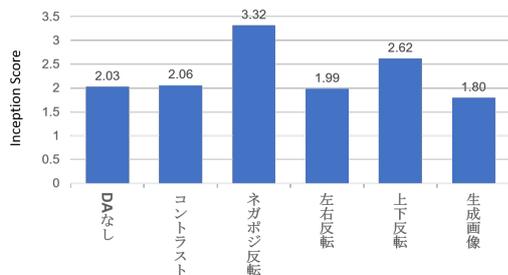


図6 1種類のDAをデータセットにしたISの比較

(2) CelebA1000枚, 1:1で合わせた2種類のDA1000枚の計2000枚

(2)の生成結果を図7(左)に示し, 図8(青)にISの結果を示す. ネガポジ反転や上下反転はデータセットとしてふさわしくない, それらを除いた2種類のDAの組み合わせを実験した. (2)のISはDAなしに比べ, どの場合も同等あるいはスコアが下がっている.

(3) CelebA1000枚, 1:1で合わせた2種類のDA1000枚の計2000枚

(3)の生成結果を図7(右)に示し, 図8(橙)にISの結果を示す. ISはDA1000枚に比べ, スコアが下がっている. データ量を増やすことで生成された画像の精度が上がったといえる.



図7 DA1000枚(左)とDA2000枚(右)の生成結果

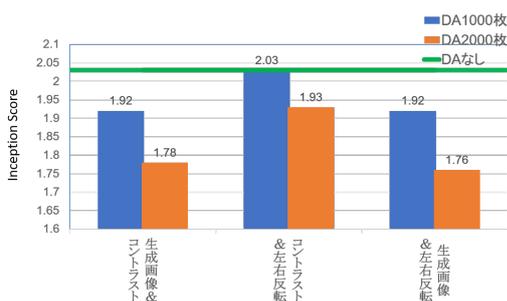


図8 2種類のDAをデータセットにしたISの比較

(4) ポメラニアン1000枚, 1:1で合わせた2種類のDA1000枚の計2000枚

(4)の生成結果を図9に示し, 図10にISの結果を示す. ISは生成画像が最も低くなった. DAなしはepochが進むに連れ輪郭がはっきりしなくなるが, DAの画像生成は輪郭を保っている, ISに影響出たのではないといえる.

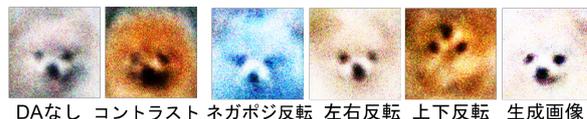


図9 1種類のDAの生成結果

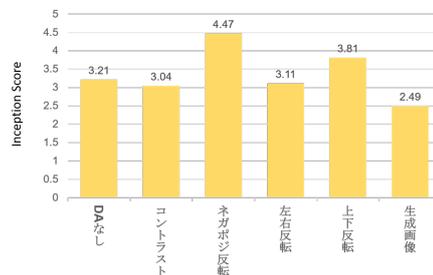


図10 2種類のDAをデータセットにしたISの比較

(5) ポメラニアン1000枚, 1:1で合わせた2種類のDA1000枚の計2000枚

(5)の結果を図11(左)に示し, 図12(黄緑)にISの結果を示す. ISはDAなしに比べ, スコアが下がっている.

(6) ポメラニアン1000枚, 1:1で合わせた2種類のDA1000枚の計2000枚

(6)の結果を図11(右)に示し, 図12(紫)にISの結果を示す. DAの量を増やしてもコントラストと左右反転の結果はスコアが上がった. CelebAの場合はスコアが下がっている, 対象データによってISに違いが出るのではないかといえる.



図11 DA1000枚(左)とDA2000枚(右)の生成結果



図12 2種類のDAをデータセットにしたISの比較

### 4.3 CelebA とポメラニアンのIS比較

図6と図10, 図8と図12を比較するとグラフの凹凸が似たような傾向になっている. 対象データに問わず, 適用

する DA の種類によって IS に影響が出たのではないかといえる。

## 5 考察

### 5.1 研究課題に対する考察

一般的に、AI の知的レベルを向上させるには、学習データの質と量が重要とされている。Chen ら [8] の研究によると、学習用データの量が多いほど精度が向上している。

データの質の観点からは、元のデータセットが 1000 枚と DA1000 枚の画像を 1:1 の比率で合わせて実験を行った。DA によって増やした学習データの量を 1000 枚に固定し、異なる DA 方策を実験することにより、DA の質を確かめることができる。

データの量の観点からは、DA 方策は変えずに、DA の枚数を 1000 枚から 2000 枚に増やして実験を行った。DA の枚数を変化させることにより、DA の質を確かめることができる。

これらを実験することにより、AI の知的レベルを向上させる実験を十分に行ったといえる。

### 5.2 関連研究の妥当性

以下の 2 点において妥当性を評価した。

#### (1) Data Augmentation

DA の種類によって IS にそれぞれ変化が見られた。これは学習データの質が変化したことにより IS に影響したためである。ネガポジ反転や上下反転など、目的に合わないパターンや実データに出てこないバリエーションをデータセットにすると、ふさわしくない顔画像が生成される。学習データの量によっても GAN 訓練の精度向上が期待できる。

データの質の観点からは、ネガポジ反転や上下反転など、目的に合わないパターンや実データに出てこないバリエーションをデータセットにすると、ふさわしくない顔画像が生成されることが確認できた。左右反転や生成画像は、IS が DA なしより低いので、DA にふさわしい方策といえる。

データの量の観点からは、DA の枚数を増やした場合、6 通りの実験から 5 通りに IS が小さくなることが確認できた。GAN の実問題を考えた場合、そのデータ量の確保が問題であるが、DA の枚数を増やす方策は必要であるといえる。

#### (2) Inception Score

我々の画像生成が成功した場合、入力画像 X に含まれる画像が特定のクラスに集中しており、かつ、高い確率で識別できている画像である。すなわち、入力画像 X が Inception v3 によって作りたいと思っていた画像に集中している分布である。本研究では、少ないデータセットの条件で実験を行っており、GAN によって生成される画像の多様性が低くなるので、上記の分布となるといえる。

本研究で実装まで至らなかったが、ResNet-50 という、画像分類における学習済みモデルを用いることで、生成された画像がどのクラスに分類されるか確率を算出できる。

入力画像 X が作りたいと思っていた画像の分布に集中しているかどうか確認できるので、ResNet-50 を用いて実験する必要がある。

## 6 おわりに

本研究では、大量のデータセットを集めることが困難な場合、DA を適用することで GAN 訓練の精度向上ができるか実験した。GAN によって生成された画像を、1 つの評価指標である IS で評価したことにより、DA の種類によって効果があることが示された。IS だけで画像生成の評価を行うのは懸念点がある。IS は画像 1 枚ごとの分類結果を結果として出力しない。生成された画像の質が悪くても、生成された画像すべての認識結果が同じような確率であれば IS が小さくなる。今後は実データにふさわしい DA 方法を考え、ResNet-50 を用いることで適切な評価を行い、GAN 訓練に相応しい DA データセットを作成する必要がある。

## 参考文献

- [1] Jakub Langr and Vladimir Bok: “GANs in Action: Deep learning with Generative Adversarial Networks”, Manning Publications, 2019 [大和田 茂 (監訳), 実践 GAN~敵対的生成ネットワークによる深層学習~, マイナビ出版, 2020].
- [2] 吉澤明彦: “人工知能 (AI) を用いた病理診断”, 京都府立医科大学雑誌 128(8), pp. 561-570, 2019.
- [3] 森野佑哉, 天野将志, 鈴木裕介, 綿絹礼菜: “フィーチャに基づく GAN の段階的学習方法の提案と評価”, 南山大学理工学部 2020 年度卒業論文, 2021.
- [4] Shane Barratt and Rishi Sharma: “A Note on the Inception Score”, Proc. ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Jun 2018.
- [5] Conor Shorten and Taghi M Khoshgoftaar: “A survey on Image Data Augmentation for Deep Learning”, Jurnal of Big Data, July 2019.
- [6] Ziwei Liu, Ping Luo, Xiaogang Wang and Xiaoou Tang: “Large-scale CelebFaces Attributes (CelebA) Dataset”, <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>, (参照 2022-02-02).
- [7] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman and C.V.Jawahar: “The Oxford-IIIT Pet Dataset”, <https://www.robots.ox.ac.uk/~vgg/data/pets/>, (参照 2022-02-02).
- [8] Chen Sun, Abhinav Shrivastava, Saurabh Singh and Abhinav Gupta: “Revisiting Unreasonable Effectiveness of Data in Deep Learning Era”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 843-852, 2017.