

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 02/01/2022 (MM/DD/YYYY)

学科名 Department	情報通信	氏名 Name	梶雄介	指導 教員 Advisor	渡辺 裕 印
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	CD 1w173035-3		
研究題目 Title	敵対的生成ネットワークによる二値画像としての顔イラストの画像生成 Image Generation of Face Illustration as Binary Image by Adversarial Generative Network				

1. まえがき

機械学習による画像処理による画像生成の質は近年著しく進化しており、その応用範囲は年々増えている。機械学習による画像処理の応用例の一つとしてキャラクターデザインの分野が存在する。一度に多くのキャラクターを量産できる機械学習を導入することができれば、大幅にキャラクターデザインにかかる時間を減らすことが可能となる。本研究では、キャラクターをデザインする上で重要である二値画像としての顔イラストの生成において、敵対的生成ネットワーク (Generative Adversarial Networks (GAN)) [1]による生成画像の改善策について検討することを目的とする。

2. 関連技術

2.1 CycleGAN

CycleGAN[2]は2017年にJ. Y. Zunらによって提案されたGANの一手法である。ソースドメインXからターゲットドメインYへの変換及び逆変換できるという特徴を持つ。通常、画像変換では学習データがペア画像であることが条件となる。しかし、CycleGANは特徴的な損失関数を複数用いることでペア画像なしで学習を行うことができる。

2.2 StyleGAN

StyleGAN[3]とは2018年にT. Karrasらが提案したGANの一手法である。最大の特徴はスタイル技術をGANに組み込んだ点にある。学習時に生成器及び識別器の畳み込み層の解像度を段階的に変換するProgressive-Growing GAN[4]を基本としている。StyleGANは潜在空間の操作を行うことで画風変換を行うことができる。本研究ではStyleGANの改良版であるStyleGAN2[5]を使用する。

2.3 SeFa

StyleGANの複雑な潜在空間を操作するためには、画像空間の特徴の変化に対応する潜在空間内の軸を探索することが一般的である。従来の技術

では探索した軸が複数の特徴を内包している点及び時間のかかる工程になっていることが問題となる。その改善策として提案されている手法にSeFA[6]がある。StyleGANの生成過程において直接軸を算出できるアルゴリズムを加える技術である。計従来の軸探索技術に比べより細かくかつ速く軸探索が行える。

3. 検討手法

CycleGANにおいては、既存研究[7]より学習するドメインX, Yの両方を二値画像に設定すると学習過程でモード崩壊が起こり、生成画像の質が落ちる。そのため本実験では二値画像とカラー画像両方を用いた学習によってCycleGANの生成画像の質の向上がもたらされるかどうかの検討を行う。StyleGANにおいては潜在空間の軸を探す技術としてSeFaを導入し、二値画像を用いた潜在空間の操作が可能かどうかの実験を行う。

データセットにはManga109[8]より新沢基栄氏の画像とカラー画像であるAnimeFace Character Dataset[9]を使用した。それぞれ正面を向いている画像を抽出し、同一解像度となるように調整した(図1)。



図1 実験に使用したデータセット

4. 実験

CycleGANによる画像生成結果を図2に示す。



図2 CycleGANによる生成画像

StyleGAN では学習データを拡張する Adaptive Discriminator Augmentation (ADA)[10]を採用した. StyleGAN2 に ADA 及び SeFa を組み込んだ学習モデルを使用した場合の画像生成結果を図3に示す.



図3 StyleGANによる生成画像

評価手法は GAN の評価手法として一般的である FID[11]を採用した. CycleGAN では変換先であるカラー画像と生成画像との FID を測定し, StyleGAN では学習データと生成画像間で FID の測定を行なった. 表1に FID の結果を示す.

表1 2方式の FID の値

	CycleGAN Manga2Anime	StyleGAN2 SeFa
FID	110.58	19.52

5. 考察

実験の結果, 潜在軸探索手法として SeFa を組み込んだ StyleGAN2 の生成画像の品質が優れることが分かった, 図2より CycleGAN では髪や目などパーツの形, 色などがはっきりしている部分に対する変換は妥当である. しかし肌と背景など形や色だけでは分別し難い部分に関しては形が崩れる傾向にある. CycleGAN がテクスチャの変換を行なっているためテクスチャとして判別しにくいものの質が落ちていることが分かる. 学習状況を調べるため Epoch ごとの結果を調べたところ, Epoch が進むにつれ変換されるパーツの種類が増えていることが分かった. 加えて損失関数を調べると二値画像のみの時に比べ正しく収束している様子が分かった. このことから

CycleGAN としては正しく学習しており, 数千~数万 Epoch をかければ質の良い生成画像ができると考えられる. 一方で処理時間の観点からは, 学習時間が数週間必要となり非現実的な処理量になると考えられる.

図3より StyleGAN に SeFa を加えたものは生成画像の質をあまり落とさずに細かい特徴の変換が可能になっていることが確認できた. さらに軸を算出しているためより細かく軸の分別が可能になっており複数の特徴を持つ軸の出現が少なかった. 加えて一つの軸を探す時間が数分と短く実用的である.

6. まとめ

本研究では, 二値画像の漫画を使用した CycleGAN 及び StyleGAN2 の改善策を検討し生成画像の比較を行なった. 実験より, 二値画像の漫画を学習させた CycleGAN の変換先をカラー画像と設定しても画像の質を上げることはできなかった. StyleGAN2 に SeFa を加えたモデルが二値画像の顔画像の生成に最も適していることを示した. しかし, StyleGAN2 に SeFa を加えたものでは任意に特徴を調整することが難しい. したがって潜在空間の中で任意の特徴を持つ軸の探索を行えるような手法が必要であると考えられる.

参考文献

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, WardeFarley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Networks," arXiv: 1406.2661, Jun. 2014.
- [2] J. Zhu, T. Park, P. Isola, A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," arXiv: 1703.10593, Mar. 2017.
- [3] T. Karras, S. Laine, T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," arXiv:1812.04948, Dec. 2018.
- [4] T. Karras, T. Aila, S. Laine, J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," CoRR, abs/1710.10196, 2017.
- [5] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," arXiv:1912.04958, Dec. 2019.
- [6] Y. Shen, B. Zhou, "Closed-Form Factorization of Latent Semantics in GANs," arXiv:2007.06600, Jul. 2020.
- [7] 蓮沼亮永: "敵対的生成ネットワークによる線画形状のスタイル変換" 早稲田大学卒業論文 Feb. 2021
- [8] K. Aizawa, A. Fujimoto, A. Otsubo, T. Ogawa, Y. Matsui, K. Tsubota, H. Ikuta, "Building a Manga Dataset "Manga109" with Annotations for Multimedia Applications," arXiv:2005.04425, May 2020.
- [9] "AnimeFace Character Dataset", <http://www.nurs.or.jp/~nagadomi/animeface-character-dataset/>, (Retrieved on Sep 17, 2021)
- [10] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, T. Aila, "Training Generative Adversarial Networks with Limited Data", arXiv:2006.06676v2, Oct. 2020.
- [11] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," arXiv:1706.08500, June 2017.

2021 年度卒業論文

敵対的生成ネットワークによる
二値画像としての顔イラストの画像生成
Image Generation of Face Illustration as
Binary Image by Generative Adversarial Networks

提出日 2022 年 2 月 1 日
指導教員 渡辺 裕 教授

早稲田大学 基幹理工学部
情報通信学科
1W173035-3 梶雄介

目次

第 1 章序論.....	1
1.1 研究背景.....	1
1.1.2 研究目的.....	1
1.1.3 本論文の構成.....	2
第 2 章関連研究.....	3
2.1 まえがき.....	3
2.2 敵対的生成ネットワーク Generative Adversarial Networks	3
2.2.1 概要.....	3
2.2.2 CycleGAN	4
2.2.3 StyleGAN	5
2.3 StyleGAN の関連技術.....	6
2.3.1 Adaptive Discriminator Augmentation	6
2.3.2 GANSpace.....	6
2.3.3 Closed-form-factorization	7
2.4 むすび.....	8
第 3 章検討手法.....	9
3.1 まえがき.....	9
3.2 データセット.....	9
3.3 CycleGAN による画像生成.....	10
3.4 StyleGAN による画像生成.....	10
3.4.1 StyleGAN の学習について.....	10
3.4.2 StyleGAN における生成画像の操作.....	11
3.5 評価手法.....	12
3.6 むすび.....	12
第 4 章実験.....	13
4.1 まえがき.....	13
4.2 実験概要.....	13
4.3 CycleGAN の実験.....	14
4.3.1 CycleGAN の実験結果.....	14
4.3.2 CycleGAN の結果考察.....	14

4.4 StyleGAN の実験.....	17
4.4.1 StyleGAN の実験結果.....	17
4.4.2 StyleGAN の結果考察.....	18
4.5 むすび.....	18
第 5 章結論と今後の課題.....	19
5.1 結論.....	19
5.2 今後の課題.....	19
謝辞.....	;20
参考文献.....	21
図一覧.....	22
表一覧.....	23

第1章 序論

1.1 研究背景

機械学習による画像処理による画像生成の質は近年著しく進化しており、その応用範囲は技術的な分野だけにとどまらず芸術の分野においても使用するケースが年々増えている。人工知能によって描かれる芸術作品も多く見るようになり、肖像画やイラストなど細かい芸術も機械学習によって作ることも可能となった。機械学習による画像処理の応用例の一つとしてキャラクターデザイン分野がある。現在ではデザインをする際にアナログやデジタルに関係なく一枚ずつキャラクターのパターンを描くことが必要であり、そのため膨大な時間を要する。機械学習を用いた画像生成の強みの一つとして画像を量産できる点があるため、ゲームや漫画におけるキャラクターデザインにおいて、一度に多くのキャラクターを量産できる機械学習を導入することができれば、大幅にキャラクターデザインにかかる時間を減らすことが可能となる。人が手書きで行っていた部分を機械学習に置き換えることで、日本が誇るアニメや漫画などといった文化の発展につながる。

1.2 研究目的

本研究では、敵対的生成ネットワークがキャラクターデザイン上で重要である顔イラストの生成において、十分な画像品質が得られるかどうかを示すことを目的とする。

機械学習の中で代表的な敵対的生成ネットワーク(Generative Adversarial Networks (GAN)) [1]での画像生成は従来の技術に比べかなりクオリティの高い顔画像やイラストなどが生成可能であることは認知されている。例として敵対的生成ネットワークを利用し、顔写真から肖像画への変換や様々な画風を再現するアートツールへの応用例などが存在している。しかし、GANによって生成される画像は限定的で万能ではなく改善の余地のある技術であることが知られている。さらに、写真など現実的な画像を用いた研究に比べイラストや漫画など2次元的なものを用いた画像生成の研究例が少ないことに加え、既存研究[2]において代表的なGANを使用した画像生成においていくつかの問題点が指摘されており、発展の余地があることを示している。本研究ではこうした問題点における敵対的生成ネットワーク(GAN)の改善策の検討を行う。

現在、敵対的生成ネットワーク(GAN)には様々な手法が存在し、同じ敵対的生成ネットワーク(GAN)といえどもそれぞれ画像の生成方法に違いが存在している。そちらの中で本研究ではいくつかの代表的なGANに二値の漫画の画像データ、カラーのアニメ画像を学習させ、顔の画像生成の評価および考察を行う。顔イラストにおけるGANが十分な画像品質を得ることができることを示すことでキャラクターデザインへの応用に結ぶことが可能であると考えられる。

1.3 本論文の構成

本論文の構成を以下に示す.

第1章は本章であり, 本研究の背景および目的について述べる.

第2章では, 本研究にて用いる GAN および GAN と並行して扱う技術について述べる.

第3章では, GAN を用いた画像生成の検討手法を紹介する.

第4章では, 検討手法を用いた実験, およびその結果と考察について述べる.

第5章では, 本研究の結論および今後の課題について述べる

第2章 関連研究

2.1 まえがき

本章では、本研究で取り扱う技術について述べる。はじめに GAN [1] を用いた画像生成手法について述べ、その GAN の一手法であり本実験で採用した CycleGAN [3], StyleGAN [4] について述べる。加えて、GAN をより深く説明するために、GAN におけるデータセットの拡張、GAN の潜在空間について記載する。

2.2 敵対性生成ネットワーク Generative Adversarial Networks

2.2.1 概要

GAN [1] は、2014 年に I.J. Goodfellow らによって提案された生成モデルである。その主な特徴は学習データから特徴を学習することで実在しないデータを学習した特徴に沿って生成することができることにある。GAN の構造を図 2.1 に示す。

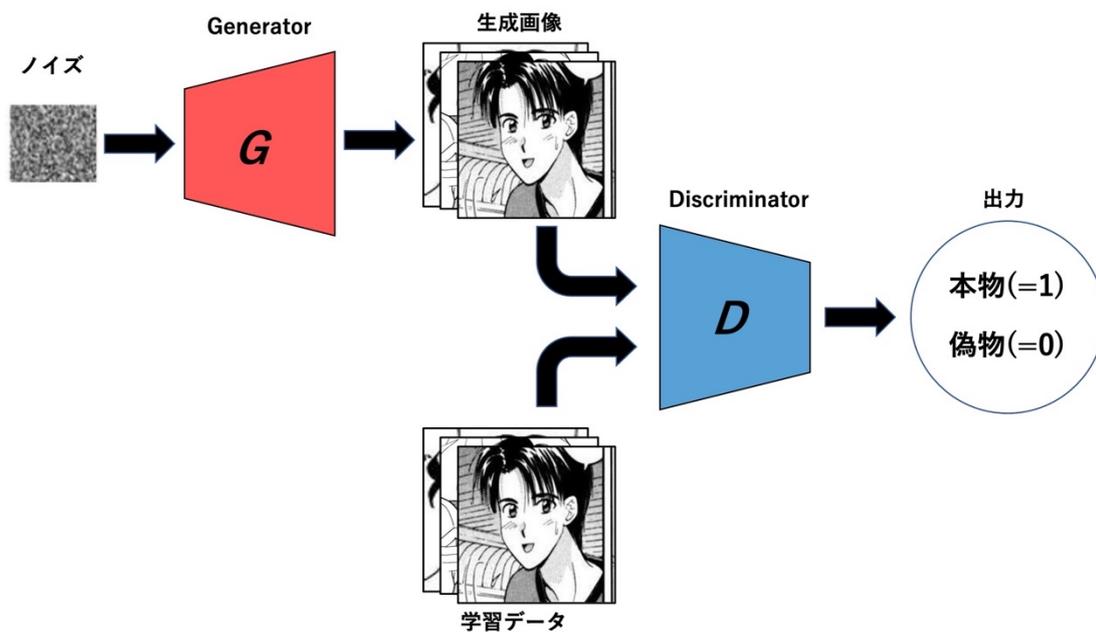


図 2.1 GAN の構造

GAN は大きく生成ネットワーク (Generator) という画像を生成する部分と識別ネットワーク (Discriminator) という生成画像の真偽を見極める部分の二つが存在する。Generator は学習データの特徴にあった画像を学習しながら生成していき、Discriminator はその生成した画像が学習データと比べ本物かどうかの見極めを行う。GAN はこの二つの部分が敵対的、つまりいたちごっこのような競争を行うような学習によってより学習データに近い、精度の高い画像を作ることができる。

2.2.2 CycleGAN

CycleGAN は 2017 年に J.Y.Zun らによって 2017 に提案された GAN の一手法である。訓練データなしに、ソースドメイン X からターゲットドメイン Y への変換およびその逆変換ができるという特徴を持つ。通常、画像変換には学習データが同じ姿勢、同じ位置であるペア画像であることが条件である。しかし、CycleGAN はペア画像なしで学習を行うことができる。CycleGAN の学習フレームワークを図 2.2 に示す。

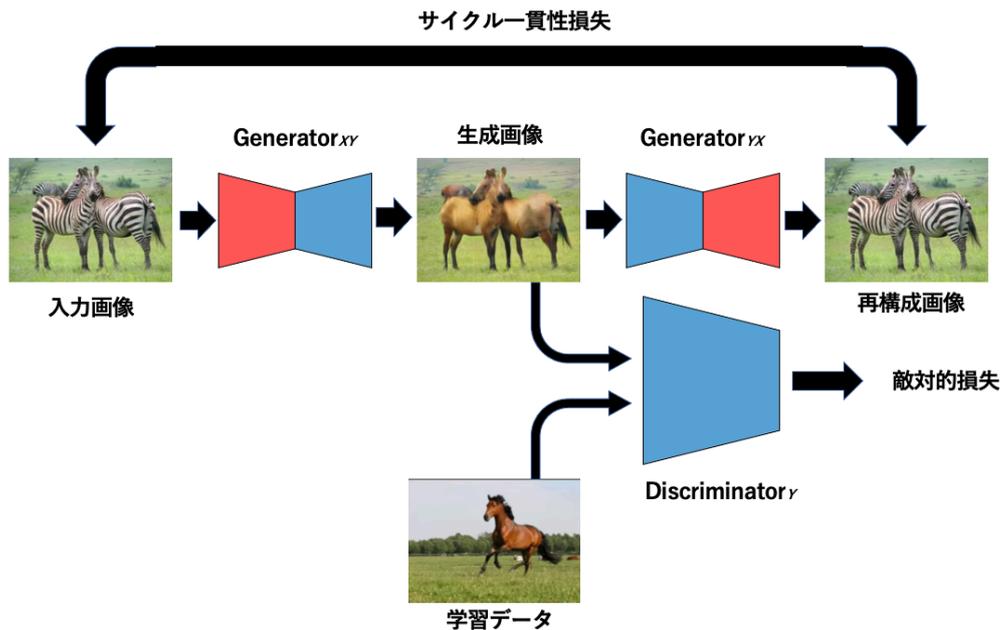


図 2.2 CycleGAN の構造

GAN はモデルを学習する中で妥当な損失関数をそれぞれ設定している。この設定した損失関数を元に出力における重みを調整している。CycleGAN にも特徴的な損失関数が存在している。その損失関数とは敵対的損失とサイクルー貫性損失である。CycleGAN はこの損失関数を扱うことでペア画像なしでの学習を実現している。CycleGAN では写像 G , F によって変換を行っている。写像 G はドメイン $G(X)$ (生成画像) がターゲットドメイン Y (学習データ) と遜色ないように敵対的損失を利用して学習を行う。対照的に逆写像 F においても $F(G(X))$ (再構成画像) がソースドメイン X に近づくためにサイクルー貫性損失を利用して学習を進める。最終的に Generator および Discriminator における敵対的損失をそれぞれ $LGAN(G, D_Y, X, Y)$, $LGAN(F, D_X, Y, X)$ と設定し、サイクルー貫性損失を $L_{cyc}(G, F)$ とすれば、相対的な重みの制御を行う λ を用いることで目的関数を式 2.1 のように表すことができる。

$$L(G, F, D_X, D_Y) = LGAN(G, D_Y, X, Y) + LGAN(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (2.1)$$

それに用いて式 2.2 を求めることで学習が行われる。

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \quad (2.2)$$

2.2.3 StyleGAN

StyleGAN とは 2018 年に T.Karras らが発表した GAN の一手法である。最大の特徴はスタイル技術を GAN に組み込んだ点にある。学習時に生成器および識別機の畳み込み層の解像度を学習時に段階的に変換する Progressive-Growing GAN[5] を基本としている。徐々に解像度を上げる学習を行うことによって高解像度の画像を安定して生成することができる。

2019 年には StyleGAN の一部を改良した StyleGAN2[6] が提案された。StyleGAN の出力画像において時より水滴のようなノイズができてしまうという問題点が存在していたため、その特徴的なノイズがなくなるように主に Generator に改良を加えたのが StyleGAN2 である。StyleGAN および StyleGAN2 の構造は基本的に同じであり、StyleGAN2 の構造を図 2.3 に示す。

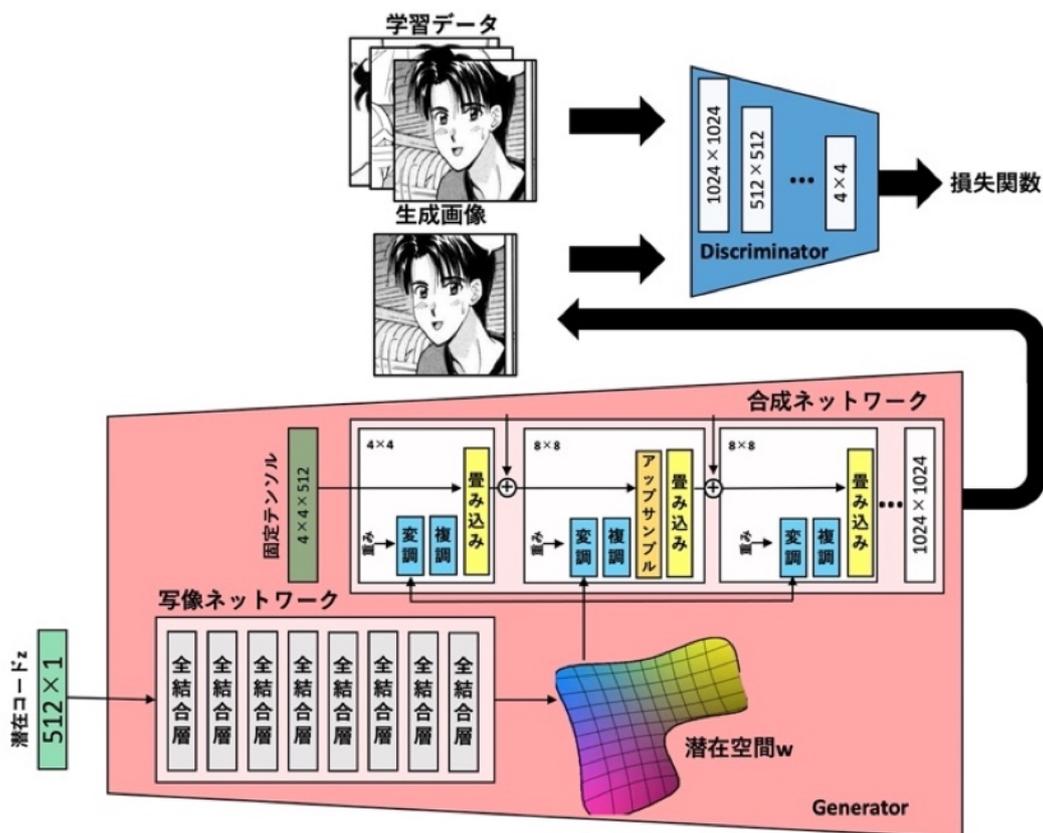


図 2.3 StyleGAN の構造

StyleGAN2 のネットワーク構造は主に二つに分けることができる。1つ目の写像ネットワーク (Mapping Network) ではランダムな潜在変数 z を 8 つの全結合層を用いて新たに潜在空間 w として形成し、潜在空間 w より抽出した特徴量を 2 つ目の合成ネットワーク (Synthesis Network) の各解像度にたたみ込むことで画像を生成する。この構造によってよ

り高解像度の画像を生成できる反面学習時に大量のデータがコストとして必要な面がある。しかし、この学習コストを削減することが可能になるなどといった関連技術もあり、それらを次の 2.3 で述べる。

2.3 StyleGAN の関連技術

2.3.1 Adaptive Discriminator Augmentation (ADA)

上述の学習時のコストを解決する方法として GAN では Adaptive Discriminator Augmentation (ADA) という技術が存在する。この Adaptive Discriminator Augmentation (ADA) は 2018 年に T.Karras[7] らが発表した。その特徴は GAN の学習を行う際の学習データを拡張するための技術にある。

この技術におけるデータ拡張は具体的には学習時のデータに上下左右の回転、反転や色の変換などといった加工を行うことによるデータの拡張を行っている。さらに、この ADA はデータ拡張を行う上で適切な強度に調整を行なった上で拡張を行うことができる。そのため単純なデータ拡張での学習によって上下反転した画像など必要のない加工が生成画像に施されることを阻止することが可能である。データ拡張を行うことで学習データが不十分な量にある状態においても Discriminator の分類に障害が起こる過学習という状態を阻止することができる。そのため十分な学習データがない状態での学習においてこの ADA は必須の技術と言える。

2.3.2 GANSpace

StyleGAN 単体では任意の画像を生成することは難しい。しかし、E.Arkonen[8] らが 2020 年に発表した GANSpace と併用することで生成画像の操作がある程度行うことができる。GAN の Generator が生成する画像はそれぞれシード値と対応しているため、生成された画像の特徴を細かく調整することが難儀である。しかし、StyleGAN においては写像ネットワークの存在によって 512 次元の潜在空間上の潜在変数と生成画像が対応している。そのことから、潜在空間を解析し、潜在変数の操作を行えば生成画像の髪型や髪色などといった細かい特徴の調整が可能である。GANSpace は潜在空間を PCA 分析することでこの 512 次元の潜在空間に存在する意味のあるベクトルを探す技術である。意味のあるベクトルというのは生成画像のセマンティックを変化するベクトルである。これを検出し、操作することで生成画像に特徴を加えられる。潜在空間およびベクトルのイメージを 2 次元に簡略した図を図 2.4 に示す

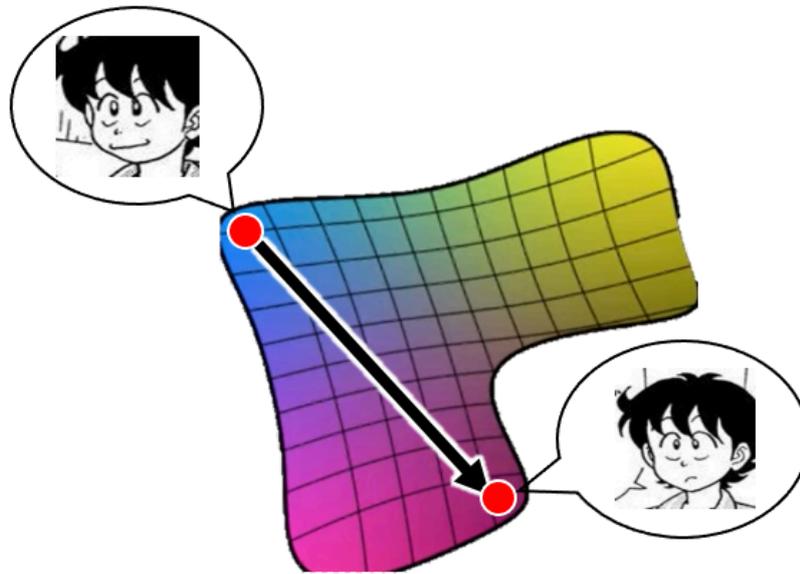


図 2.4 2次元で表現した潜在空間

また, GANSpace の強みの一つとして使用時に教師データがいらぬ点である.

従来の潜在空間の分析を行う技術において事前の教師データが必要だったことに対し, この GANSpace は教師データなしに変化の大きい意味のあるベクトルを見つけ出すことができる.しかし, GANSpace は大きく変化を持った軸の探索は可能であるが, 細かい変化を持つ軸を探索することには向いていない.つまり複数の特徴を保持する軸を検出してしまふことが多くなってしまう.その上ベクトルを一つずつ検出し, 狙った特徴を持っているか手探りで探す必要性があり, GANSpace では探索に時間が大きくかかってしまう問題がある.その改良版として提案されたのが 2.3.3 節にて述べる SeFa である.

2.3.3 Semantic Factorization(SeFa)

GANSpace に変わって 2020 年に Yujun Shen ら[9]によって新たに潜在空間における意味のあるベクトルを探索するために提案され得た技術が Semantic Factorization(SeFa)である. SeFa は直接 GAN の生成過程において探索のアルゴリズムを組み込むことでベクトルの探索を行うという手法である. GANSpace と違い意味のあるベクトルを計算によって算出することができるためより早く, 正確に, かつ細かく軸の探索が可能となる.加えて, GANSpace と同様に潜在空間の分析するための事前のサンプリングが必要ない.本手法では生成画像のセマンティックが潜在変数自体ではなくベクトル n が直接変化を加えていることに着目し, 線型結合層を利用することでベクトル n の探索を行っている線型結合層において潜在変数 z の変化によって得られる y の変化が線型結合層の重みとベクトル n の積によって表現できること.式 2.3 に示す.

$$\Delta y = FC(z') - FC(z) = (A(z + \alpha n) + b) - (Az + b) = \alpha \quad (2.3)$$

上述の式によって行列 A を含む方向 An を分解することができれば生成画像を意味ごとに分解できることがわかる。そのため、この An 最適化問題を解く必要がある。解いた結果、最終的に式 2.4 となる。

$$2A^T A n_i - 2\lambda_i n_i = 0 \quad (2.4)$$

式 2.4 が ATA の固有値問題への帰着が見て取れるため ATA の固有値を大きい順番に計算すれば良いこととなる。この計算を生成過程に組み込むことで意味のあるベクトルの探索を行っている。結果、GANSpace のように PCA 分析が必要なく、計算での軸探索が可能となった。

2.4 むすび

本章では、本研究において使用する GAN を初め、実際に取り扱う GAN の種類である CycleGAN と StyleGAN を紹介した。加えて StyleGAN を補助する技術である ADA と SeFa および関連した技術である GANSpace の紹介を行った。

第3章 検討手法

3.1 まえがき

本章では二値画像を学習させた各 GAN の生成画像が十分な画像品質を得られるかどうかを調べるための検討手法について述べる.初めに使用するデータセットの説明を行い,その後使用する CycleGAN, StyleGAN における画像生成の手法および評価手法について述べる.

3.2 データセット

本研究においては漫画やゲームにおけるキャラクタデザインに応用するため各 GAN の生成画像が十分な画像品質を得られるかどうかを目的としている.そのためデータセットとしては二値画像とカラー画像を用意した.二値画像のデータセットとしては日本の漫画を 109 種類内包している Manga109[10]を用いた.Manga109 は名前の通り 1980 年代から 2010 年代までの 109 作品もの漫画作品を収録しているデータセットである.本研究においてはキャラクタをデザインする上で重要である顔イラストを中心に評価を行う.そのため, Manga109 の収録より漫画の主人公の顔のパーツが全て画像内に収まっているイラストのみを抽出し本研究に用いた.主な使用画像を図 3.1 に示す.



図 3.1 Manga109 の画像

加えて本研究では二値の漫画画像だけでなくカラーのアニメ画像である AnimeFace Character Dataset [11]というデータセットも使用している.こちらは主にアニメ女性の顔イラストを生成するウェブサイトによって生成された画像を一つにまとめたデータセットで

ある.実験において Manga109 と同様に顔のパーツが全て画像内に収まっている画像を抽出した. 加えて Manga109 と同じ解像度になるように調整を行った. 主な使用画像を図 3.2 に示す.



図 3.2 AnimeFace Character Dataset の画像

3.3 CycleGAN による画像生成

CycleGAN は第 1 章に示した二値画像の画風変換についての既存研究において漫画のような二値の画像ドメイン同士の変換を行うことに適してないことがわかっている.その原因として記載されている点が二値の画像は形のみで表現された画像であることから, テクスチャの変換を主に扱う CycleGAN では十分に変換することができないことが挙げられている.本研究では二値画像とカラー画像のドメイン同士の変換を行うことで, 片方のドメインをカラー画像に設定することが CycleGAN の二値画像に対する改善策になるかを検証する.

CycleGAN では 2.2.2 節で記述したとおり学習用のデータが二つ必要である. 二値のドメイン同士の変換に比べ二値のドメインとカラー画像のドメインの変換による生成画像が十分な画像品質を得られるかどうかを試すため, 二値のデータセットとして Manga109 を使用し, カラー画像として AnimeFace Character Dataset より抽出した画像をデータセットとした. CycleGAN の学習時間は三日ほど要しており, Manga109 および AnimeFace Character Dataset よりそれぞれ 500 枚程度の画像を学習させた.

3.4 StyleGAN による画像生成

3.4.1 StyleGAN2+ADA の学習について

StyleGAN は 2.2.3 節に記述したとおり CycleGAN と違い二つのドメイン間の変換ではなく学習データより構築した潜在空間を参照して画像生成を行う.そのため学習におけるド

メイン数の制限はない。しかし潜在空間が 512 次元という高次元であるという特徴から学習データのスタイルを多く詰め込んだ場合、潜在空間が複雑化するため生成画像に多くの特徴が混ざり合うこととなる。その結果画像品質が著しく下がる。よって学習においては最大二つの画風を用いて学習を行った。

Manga109 の中から正面を向いている画像を抽出し、かつ画風を絞ったことにより用意できる学習データ数百枚と数千の学習データが必要である StyleGAN の学習としては不十分な量である。そのため 2.3.1 節で記載した ADA による StyleGAN の学習データの拡張を行った。つまり、本研究では ADA によって上下左右の反転などといった加工を加えてデータセットの増幅が行われる。

3.4.2 StyleGAN における生成画像の操作

CycleGAN においては、学習するデータのうち片方を変換先のドメインとして設定することでベースラインのドメイン内でありながら任意の特徴を持つ画像を生成できる。対照的に StyleGAN では生成画像に任意の特徴を加えるためには 2.3.2 節および 2.3.3 節で記載した潜在空間の操作が必要である。

通常の GAN では入力時のシード値と画像空間との対応関係がわかりづらく、生成画像に対して任意に細やかな加工を行うことが難しい。しかし、StyleGAN ではシード値である潜在コードをもとに写像ネットワークを通して潜在空間をマッピングしており、潜在空間によって画像空間を安易に捉えることができる。これは写像ネットワークによってそれぞれも画像データからスタイルが潜在空間に再現される際に画像データが持つ特徴を保持する画像空間と対応するように潜在空間が構築されるからである。例として学習データが顔画像であれば性別や髪型などといった画像空間と対応した潜在空間が存在している。そのため潜在空間の探索を通して画像空間との対応関係の分析を行うことで任意の特徴を持った潜在空間の次元とそれに対応する画像空間を把握することができる。生成画像の操作を行うためには画像空間と潜在空間の対応関係を応用し、ベースラインとなる潜在空間と任意の特徴を持つ潜在空間を行き来するベクトルの探索を行うことが重要である。つまり様々な特徴を持つ潜在空間の判別、探索を行うことで他の GAN には難しい細やかな生成画像の操作が可能である。

従来の研究において StyleGAN の潜在空間の探索は GANSpace が利用されてきた。しかし、既存研究[2]において GANSpace は教師データなしで潜在空間における探索ができることがわかっている。しかし、二値画像を学習させた StyleGAN において複数の特徴を持つ軸を検知することが多い点および軸の探索に長時間が必要であるなどといった問題点が挙げられていた。そのため本研究においては SeFa を検討手法として GANSpace の代わりに利用する。SeFa は GANSpace と違い生成過程の時点で探索を行うため、PCA 分析や教師データなどを必要とせず GANSpace に比べより早くかつ簡易に潜在空間の操作が可能となる。このような特徴を持つ SeFa がこうした課題への改善策になり得るかを試すととも

に二値画像との相性を評価する.そのために, SeFa を用いて 512 次元の潜在空間の中から生成画像に意味のある特徴を保持する次元を探し出し, 変化前と変化後の次元を行き来できるベクトルの探索を行うことで様々な特徴を持つ画像の生成を行う. 本実験では加えて StyleGAN を改良した StyleGAN2 を用いているため, 用意したモデルはまとめると StyleGAN2+ADA ということになる.

3.5 評価手法

本実験における評価手法は GAN の評価において一般的である Fréchet Inception Distance(FID)[12]を用いて行う.FID は 2017 年に H,Martin らが提案した評価手法である.GAN における評価は生成画像と学習データとの比較を通じて行われる.学習データをもとにした画像空間の分布と生成画像をもとにした画像空間の分布との特徴距離を図ることで評価を行う.この距離は直接的な距離ではなくモデルを多角的に評価するため Inception Network を用いた分布間の Fréchet 距離である. この Fréchet 距離を計算するのが Fréchet Inception Distance(FID)である.Inception Network を利用することで生成画像の品質だけでなくその多様性をも確率を通して評価することが可能である.画像がどのような物質であるかの確率を比較することで生成された画像がよりリアルで適切であるかの評価が行える.FID における距離が近い(出力の値が小さい)ほど生成画像と学習データが似ているという評価となる.

他の手法として Inception Score(IS)[13]や KID[14]などといった手法が存在するが今回は FID を使用する.IS は FID と同様に Inception Network を利用した評価である.しかし, 生成画像のみにしか Inception Network を適応しない.このため学習データにも Inception Network を適応できる FIDの方が評価手法として有用であると判断した.KID においては CycleGAN や StyleGAN において一般的な評価手法でないこと, また既存研究の多くが FID を評価手法として採用していることが多く, 既存研究との比較が安易であることから今回は FID を採用した.

3.6 むすび

第 3 章においては CycleGAN および StyleGAN において本研究での扱い方を明記した.初めに使用するデータセットである Manga109 および AnimeFace Character Dataset の説明を行った.その後使用する CycleGAN, StyleGAN における画像生成の手法および StyleGAN における SeFa を利用した潜在空間のベクトルの探索について述べた. また, 評価手法として FID を紹介した

第4章 実験

4.1 まえがき

第4章では、第3章にて検討手法として紹介した手法をもとに各GANにおける生成画像の品質の評価を行う。初めに実験に使用したデータセットについての詳しい記述を行いその後 CycleGAN, StyleGAN における二値画像を用いたそれぞれの実験結果およびその考察を行なった。

4.2 実験概要

CycleGAN ではデータセットとして二値の画像とカラー画像が必要であるため 3.2 節で記述した通り Manga109 および AnimeFace Character Dataset を使用した。しかし、Manga109 には 109 作品という莫大な量を収録しているためその中から漫画家を一人選択し、その方の画風を用いて学習を行った。選択したのは「ハイスクール!奇面組」と「ボクはしたたか君」を描かれている新沢基栄氏のものである。選択した理由としては Manga109 より正面を向いている画像を抽出した結果十分なデータ量を持つ画風であることが分かったためである。同時に髪型や目元などがはっきり描かれている顔イラストが多く、絵柄における認識のしやすさという点で選んだ。新沢基栄氏の主な画像を図 4.1 に示す。



図 4.1; 新沢基栄氏の画風

CycleGAN に新沢基栄氏の画像と AnimeFace Character Dataset を学習させその結果の評価を行った。加えて画質を統一するためにそれぞれの画質が同じになるように調整を行った。二値の画像同士を学習した場合との比較を行うことで、顔イラストを生成するための改善点を考察する。

StyleGAN においても CycleGAN と同様に新沢基栄氏の画像を学習させて実験を行った。二値画像およびカラーの顔イラスト画像を利用した SeFa の探索実験が行えるのかの実験を行い、二値画像における StyleGAN2+ADA+SeFa を利用した生成画像の品質の評価を行った。

4.3 CycleGAN の実験

4.3.1 CycleGAN の結果

CycleGAN に二値画像とカラー画像を学習させた結果を図 4.2 に示す。



図 4.2 CycleGAN の結果

結果として CycleGAN の結果として妥当ではなかった。髪色や一部のキャラクターの目はカラー画像の学習データのように変換できているが全体的に形が崩れる傾向にあった。CycleGAN では変換先であるカラー画像と生成画像との FID を測定した。FID の結果は 110.48 と品質として妥当な結果にならなかった。既存研究[2]にて二値画像同士の変換で測定された値が 60 前後だったことを考慮すると妥当でない判断できる。

4.3.2 CycleGAN の結果考察

CycleGAN の結果を踏まえると二値画像のみを設定した学習データを扱うことには適していないことが分かった。生成された画像を見ると髪など色の濃淡によって判別がはっきりしている部分目など特徴的な形を持っている部分の変換は妥当である。一方肌や背景など形、色などでは分別し難い部分や眉毛と細い目など形、色が似ている部分に関しては形が崩れる傾向があった。このことから CycleGAN がテクスチャの変換が行われるため二値の画像のみを設定した学習データはテクスチャの差異がないことから形のみでパーツや背景との分別がし難く、全体的に妥当ではない結果が得られないことが分かった。

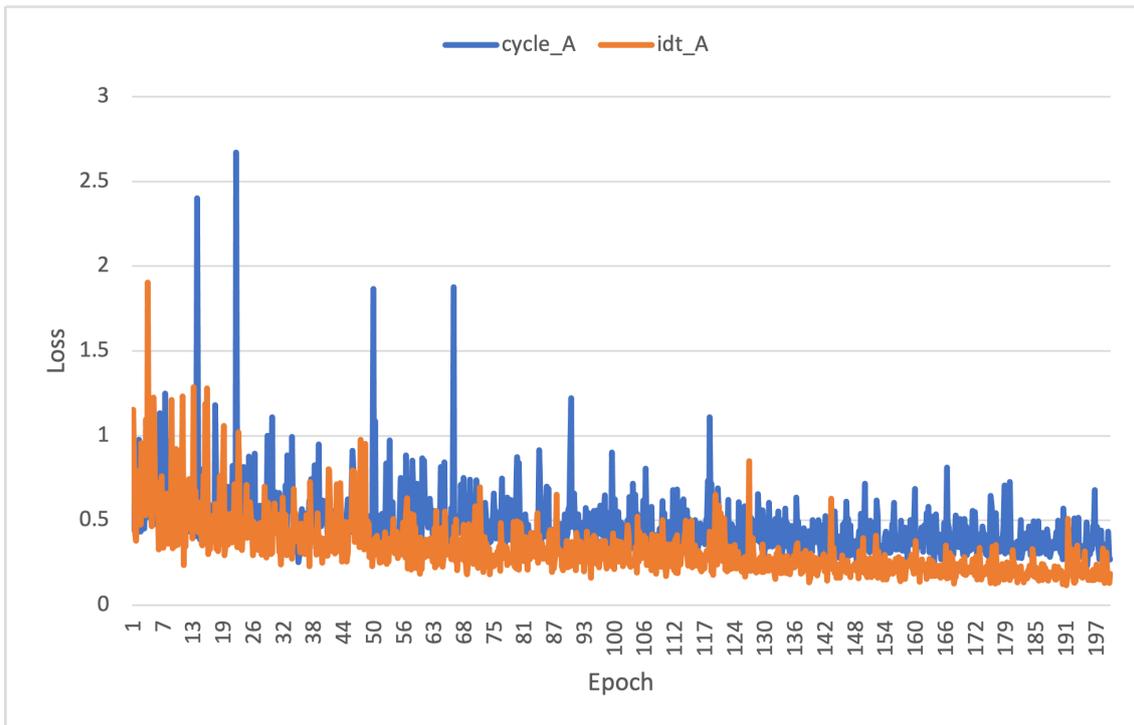
続いて学習面での評価を行う.最初に Epoch ごとの出力画像を図 4.3 に示す.



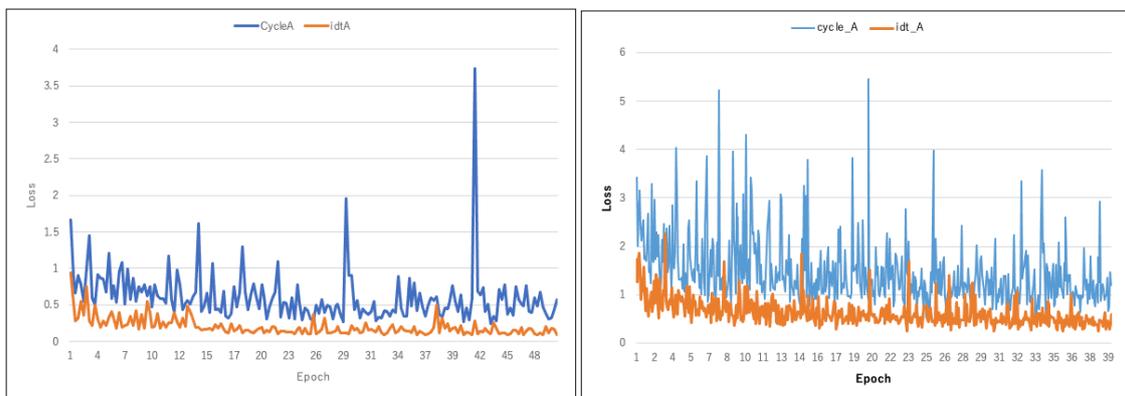
図 4.3 Epoch ごとにおける CycleGAN の生成画像

図 4.3 を見ると Epoch100 の時点では髪色のみの変換が行われている.一方で Epoch200 では目や襟などが変換されている.このことから学習が途中で止まることなく学習に進捗があることがわかる.

続いて損失関数の面から学習状況を調べた.ここで学習状況の参考とするために CycleGAN の変換で代表的な馬とシマウマの変換時の CycleGAN の損失関数の様子および二値の画像同士を学習させた場合の損失関数の計算も行った.この時の二値の画像は新沢基栄氏の画像と同じく Manga109 に収録されている佐佐木あつし氏の画像を使用した.損失関数の結果を図 4.4 に示す.



(a)



(b)

(c)

図 4.4 損失関数の比較

(a)manga2anime の損失関数

(b)manga2manga の損失関数(c)horse2zebra の損失関数

損失関数を調べると二値画像とアニメ画像の学習は馬としまうまと同じように緩やかに収束していることがわかる.このことから学習状況が正しく進んでいることが見える.対照的に二値画像間の CycleGAN では反対に学習過程でモード崩壊が起こっている.

以上の Epoch ごとの出力結果および損失関数の結果より CycleGAN の学習としては正しく進行していることが分かった。学習自体は正しく進んでいるが図 4.2 のような結果になることから数千～数万回 Epoch をかければ質の良い生成画像ができると考えられる。

しかし、処理時間の観点からは、学習時間が数週間必要となり非現実的な処理量になると考えられる。このことから、CycleGAN をキャラクタデザインに応用するには難しいと言える。

4.4 StyleGAN2 の実験

4.4.1 StyleGAN2 の結果

新沢基栄氏の画像と AnimeFace Character Dataset を別々に学習したそれぞれの StyleGAN 2 に SeFa を利用して潜在空間内にある意味のあるベクトルの探索を行った。画像生成は探索した軸を中心に行った。その軸上の変化の様子を図 4.5 に示す。

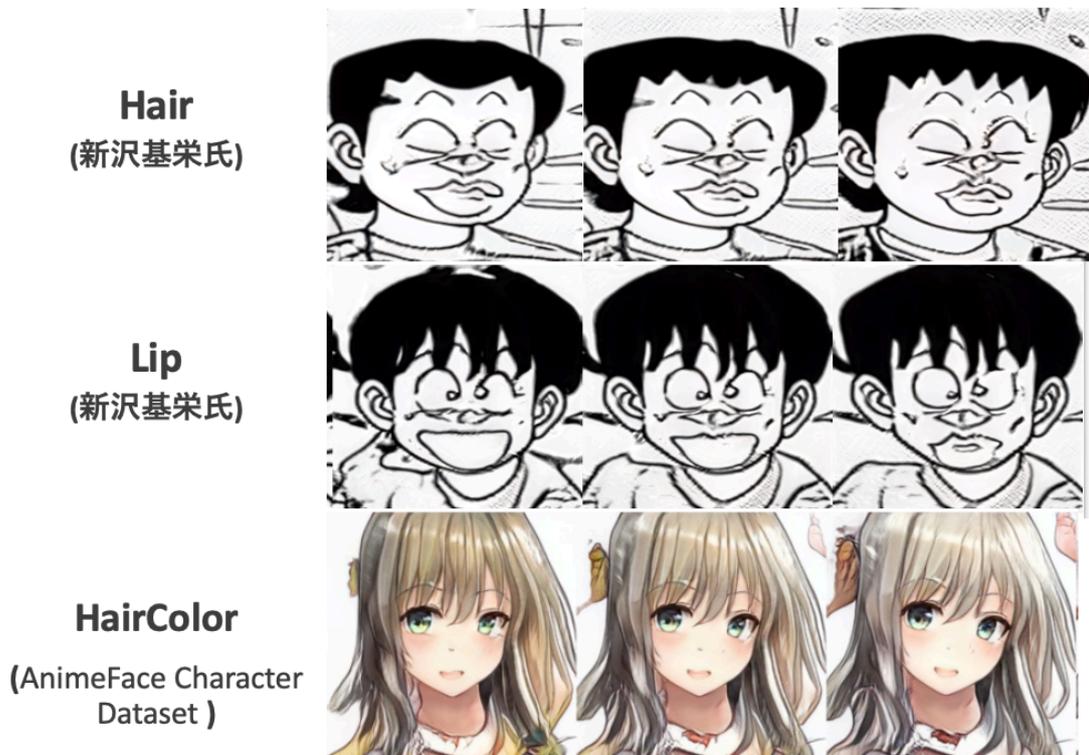


図 4.5 SeFa で探索した軸による画像生成

以上の図の通り SeFa を利用した潜在空間内の軸の探索を用いることで二値画像、カラー画像関係なくクオリティの高い画像の生成が行えることが分かった。FID は新沢基栄氏の画像を学習させたものを測定した結果が 19.52 となった。

4.4.2 StyleGAN2 の結果考察

図 4.2 の通り髪型や唇の変化などといった細かい特徴の変化を持つ画像の生成が行えることが分かった. SeFa を利用した軸の探索では一つの軸を探すのに要した時間は 3 分程度であった. 既存研究[2]で GANSpace では軸の一つ探すのに 1 時間程度かかっていたことが記述されていることと比較すると SeFa は効率よくかつ簡易に軸が探せる技術であることが確認できる. また, 既存研究[2]では GANSpace を用いると複数の特徴を持った軸が見つかる場合が多かったものの, SeFa での結果は複数の特徴を持つ軸の出力が少なかった. これは軸が直接算出されるため特徴が細かく判別することが可能になったためだと考えられる.

図 4.5 の通り AnimeFace Character Dataset を学習させた結果も Manga109 と同様に特徴の変化を見ることができ. このことより StyleGAN2 においては CycleGAN と違いテキストの差異がない場合においても生成画像のクオリティに差異がないことがわかる. AnimeFace Character Dataset は画風としてのバリエーションが新沢基栄氏の画風より多いため生成画像がより多くの特徴を跨ぐ可能性があった. しかし, 結果として複雑化することなく細かい特徴の変化を生成することができた.

簡単に特徴軸を見つけることが可能であることが示せたため, こうした特徴軸を別の画風を学習させたモデルに合成することができれば StyleGAN2 はよりキャラクターデザインに貢献できる技術になると考えられる. そのために 512 次元という複雑な潜在空間を細かく分析し, 似た特徴を持つ軸のラベリングを細かく整理する技術が必要であると考えられる.

FID に関しては 19.52 となり, CycleGAN の 110.58 と比べるでもとても良い結果になったことがわかる.

4.5 むすび

第 4 章では初めに Manga109 より抽出した新沢基栄氏の画風について述べた. その後 CycleGAN の結果を示し, 学習状況の様子を用いて考察を行なった. StyleGAN2 においてもその結果および考察を行い, 最後に FID を用いて各 GAN の評価を行なった.

第5章 結論および今後の課題

5.1 結論

実験結果より得た CycleGAN と StyleGAN2 の生成画像の FID と Manga109 を使用した既存研究[2]との比較の結果を表 5.1 にまとめた.

表 5.1 4 方式の FID の値

	CycleGAN Manga2Manga	CycleGAN Manga2Anime	StyleGAN2 +ADA+GANSpace	StyleGAN2 +ADA+SeFa
FID	61.01	110.58	20.17	19.52

表の結果より StyleGAN2 の結果が CycleGAN の結果より小さい値を示しており、このことから StyleGAN2 で生成したものが生成画像としてのクオリティが学習データに近いことが分かった.

二値画像である漫画からカラー画像に変換した本実験での CycleGAN においては、学習状況の比較より数千~数万 Epoch の学習時間をかけることで FID の値を抑えることができると考えられるが数千 Epoch の学習時間非現実的な処理量になると考えられる.

StyleGAN2 に SeFa を加えた結果、生成画像の質を落とすことなく細かい特徴の変換が可能になることを確認できた.さらに軸を算出しているためより細かく軸の分別が可能になっていたため複数の特徴を持つ軸の出現が少なかった.加えて一つの軸を探す時間が数分と短く GANSpace と同じく手探りでの探索ではあるが、より実用的な探索技術であると言える.

以上の実験結果より、二値の顔イラストにおいては StyleGAN の画像生成の方が有用であると判断できる.

5.2 今後の課題

キャラクタデザインにおいて StyleGAN2 を利用した場合においても ADA でデータ拡張を行なった上に SeFa で軸を探索する必要があり、工程が多い.GANSpace に代わる SeFa によって軸探索が簡易になるが全体として作業量が多く、機械学習をキャラクタデザインに導入するためには学習や工程を簡略化できる方法が必要であると考えられる.加えて StyleGAN は 512 次元という大きな潜在空間上にマッピングを行う関係上一度に多くの画風を学習してしまうと潜在空間が複雑化しすぎてしまい画風が混ざり合うこととなる.そのため生成画像の形の崩壊につながる.このため StyleGAN が生成できる範囲は限られており、多くの画風を学習するのには向いていない.解決方法として別々に画風を学習させたモデルを多数用意する方法が存在するが、非現実的なモデルだと言える.StyleGAN の生成画像の品質に問題はないため、キャラクタデザインに応用するためには生成できる幅が広が

るような技術，例えば複数のモデルの潜在空間を行き来する手法や潜在空間同士を合成し，分析する手法などが今後必要となる．

謝辞

本研究において長い時間ご指導し，研究を見守っていただいた渡辺教授には心から感謝申し上げます．

また，研究室でご意見，アドバイスをくださった同期の皆様並びに先輩方にも御礼申し上げます．

参考文献

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Networks”, arXiv:1406.2661, Jun, 2014,
- [2] 蓮沼亮永: “敵対的生成ネットワークによる線画形状のスタイル変換”早稲田大学卒業論文 Feb. 2021
- [3] J. Zhu, T. Park, P. Isola, A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, arXiv:1703.10593, Mar 30, 2017
- [4] T. Karras, S. Laine, T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks”, arXiv:1812.04948, Dec 12, 2018
- [5] T. Karras, T. Aila, S. Laine, and J. Lehtinen. “Progressive growing of GANs for improved quality, stability, and variation”. CoRR, abs/1710.10196, 2017
- [6] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, “Analyzing and Improving the Image Quality of StyleGAN”, arXiv:1912.04958, Dec 3, 2019
- [7] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, T. Aila, “Training Generative Adversarial Networks with Limited Data”, arXiv:2006.06676v2, Oct. 2020.β
- [8] E. Härkönen, A. Hertzmann, J. Lehtinen, S. Paris, “GANSpace Discovering Interpretable GAN Controls”, arXiv:2004.02546, 6 Apr 2020
- [9] Y. Shen, B. Zhou, “Closed-Form Factorization of Latent Semantics in GANs”, arXiv: 2007.06600, Jul 13, 2020
- [10] K. Aizawa, A. Fujimoto, A. Otsubo, T. Ogawa, Y. Matsui, K. Tsubota, H. Ikuta, “Building a Manga Dataset "Manga109" with Annotations for Multimedia Applications”, arXiv:2005.04425, May 9, 2020
- [11] “AnimeFace Character Dataset”, <http://www.nurs.or.jp/~nagadomi/animeface-character-dataset/>, (Retrieved on Sep 17, 2021)
- [12] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium”, arXiv:1706.08500, Jun 26, 2017
- [13] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. “Improved techniques for training GANs.” In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems* 29, pages 2234–2242, 2016
- [14] M. Bińkowski, D. J. Sutherland, M. Arbel, A. Gretton, “Demystifying MMD GANs”, arXiv:1801.01401, Jan 4, 2018

図一覧

図 2.1 GAN の構造.....	3
図 2.2 CycleGAN の構造.....	4
図 2.3 StyleGAN の構造	5
図 2.4 2次元で表現した潜在空間.....	7
図 3.1 Manga109 の画像.....	9
図 3.2 AnimeFace Character Dataset の画像.....	10
図 4.1 新沢基栄氏の画風.....	14
図 4.2 CycleGAN の結果.....	15
図 4.3 Epoch ごとにおける CycleGAN の生成画像.....	16
図 4.4 損失関数の比較.....	17
図 4.5 SeFa で探索した軸による画像生成.....	18

表一覽

表 5.1 各 FID	20
-------------------	----