卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: 02/02/2021 (MM/DD/YYYY)

学科名 Department	情報理工	氏名 Name	高橋美帆	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	1W172186-9 CD		
研究題目 Title	ランドマーク情報を利用した WGAN-GP によるイラスト用顔画像の生成手法の研究 A Study on Generation Method of Face Image for Illustration by WGAN-GP Using Landmark Information				

1. まえがき

SNS におけるアイコンやゲームのアバター，ポスターなどで顔のイラストは利用される．Generative Adversarial Networks (GAN) [1]を用いることでイラスト用顔画像を生成できるが，それらは目がない画像や口がおかしい位置にあるなど，顔として認識できない画像も生成される．また，顔の形状や表情を指定して生成できないという問題がある．そこで，顔の形状や表情を指定するだけで簡単に顔のイラストを作成できるようになれば，イラスト用顔画像の生成における可能性が広がると考えられる．また，顔の位置情報であるランドマーク情報を利用することで画像ではなく顔画像として生成できると考え，ランドマーク情報を利用した，イラスト用顔画像の生成手法を提案する．

2. Wasserstein GAN-GP

WassersteinGAN-GP(WGAN-GP)[2]とは，GAN の安定化手法である Wasserstein GAN (WGAN)[3]を改良したものである．WGAN は Wasserstein 距離を用いることで GAN の不安定性を解消した．しかし，WGAN で利用される Weight Clipping はパラメータにより学習が上手くいかないことや，勾配消失をしてしまうなどの問題がある．そこで，WGAN-GP では Gradient Penalty を追加することで WGAN の安定化に成功した．

3. 提案手法

本研究で提案する手法の構造を図 1 に示す．生成画像が入力画像のランドマークを再現できるよう WGAN-GP の基本構造に Landmark Predictor を追加し，損失関数に新たな損失関数である Landmark Loss を追加する．Landmark Predictor は Convolutional Neural Networks (CNN)[4]で構成されている．画像を入力するとランドマークを出力するように事前に学習させておく．Landmark Loss は，入力画像(Input Image)におけるランドマークと Generator で生成された画像のランドマークの距離を表す損失関数である．

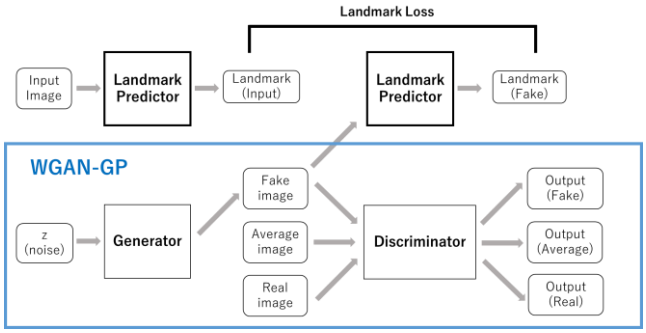

図 1　提案手法の構造

4. 実験

本研究では，kaggle で公開されている Bitmoji の顔画像データセット Bitmoji Faces [5]から 4,084 枚の画像を用いて WGAN-GP との比較実験を行った．検出するランドマークは目，鼻，口，顎の合計 8 箇所であり，指定方法を図 2 に示す．

提案手法と WGAN-GP における生成結果を図 2 に示し，それぞれの手法における品質の悪い画像の例を図 4 に示す．また，入力画像と生成画像のランドマークの距離と FID[6]を用いて評価を行った結果を表 1 に示す．

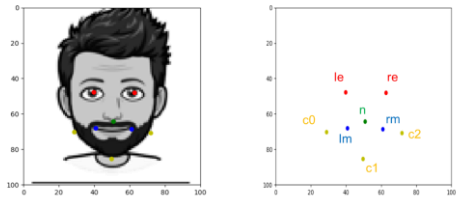

図 2　ランドマークの指定方法[5]

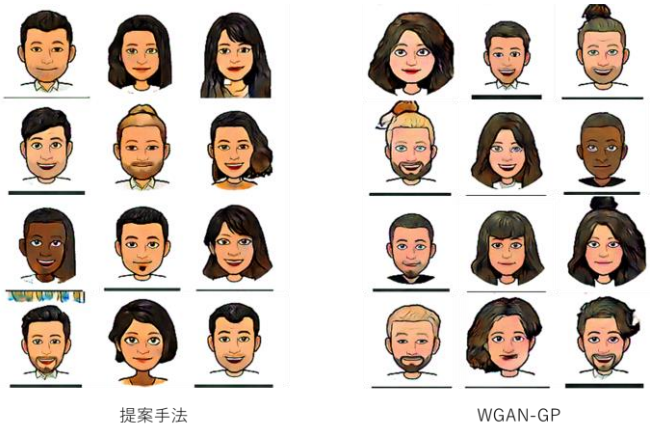

図 3 提案手法と WGAN-GP の生成結果

図 4　提案手法と WGAN-GP の品質の悪い画像

表 1　WGAN-GP との比較結果

モデル	ランドマーク距離	FID
WGAN-GP	0.2203	136.1
提案手法	0.2180	132.8

5. 考察

表 1 に示すように，提案手法におけるランドマーク距離は WGAN-GP より小さくなった. しかし，少し近づいただけで入力画像とランドマークが完全に一致する画像の生成はできなかった．ランドマーク情報は損失関数にのみ取り入れ，Generator に直接与えていないため，ランドマークの詳細な情報まで学習できなかったと考えられる．そのため，図 3 に示すように，提案手法と WGAN-GP の生成画像に大きな違いは見られない．しかし，図 4 に示すように品質の悪い画像において違いが生じた．WGAN-GP で生成した画像における品質の悪い画像は，目や鼻，口がぼやけており，顔として認識できない．一方，提案手法における品質の悪い画像は，画像全体が歪んでいるものはあるが，WGAN-GP に比べて目，鼻，口が崩れていない．また，目，鼻，口の形が明確になり，FID の値が小さくなった．よって，提案手法では，Landmark Loss を加えたことにより，ランドマークを認識できる画像を生成するように学習できたと考えられる．

6. まとめ

本研究では，ランドマーク情報を利用した WGAN-GP によるイラスト用顔画像の生成手法を提案し，WGAN-GP との比較を行った．ランドマーク情報を生成に利用するため，損失関数に Landmark Loss を追加した．実験した結果，入力画像のランドマークと一致する画像を生成できなかった．しかし，Landmark Loss の追加により，ランドマークを認識できるように学習でき，品質の高い画像が生成できた．よって，イラスト用顔画像において提案手法の有効性を確認できた．今後の課題として，より自由度の高い画像生成を行うために，Generator に直接ランドマーク情報を与える方法を検討する必要がある．

参考文献

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, “Generative Adversarial Networks”, Neural Information Processing Systems (NIPS), Dec. 2014.

[2] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, "Improved Training of Wasserstein GANs," arXiv preprint arXiv: 1704.00028, 2017.

[3] M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein GAN," arXiv preprint arXiv:1701.07875, 2017.

[4] [4]　斎藤康毅, ゼロから作る Deep Learning―Python で学ぶディープラーニングの理論と実装ー, オライリー・ジャパン, p.205, 2016

[5] M. Mozafari, “Bitmoji Faces”, kaggle, Aug 2020, https://www.kaggle.com/mostafamozafari/bitmoji-faces（2021 年 1 月現在）

[6] 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 preprint arXiv:1706.08500v6, 2018.

2020 年度　卒業論文

ランドマーク情報を利用した WGAN-GP による イラスト用顔画像の生成手法の研究

A Study on Generation Method of Face Image for Illustration by WGAN-GP Using Landmark Information

指導教員　渡辺　裕　教授

提出日　2021 年 2 月 2 日

早稲田大学　基幹理工学部　情報理工学科

1W172186-9

高橋　美帆

目次

第1章　序論

1.1　研究の背景

顔のイラストはTwitterやFacebookなどのソーシャルメディアで利用するアイコンやゲームのアバター，ポスターなど様々な場面で利用される．しかし顔のイラストを描くのは難しく，絵心やテクニックが必要とされる．そのため，顔の形状や表情を指定するだけで簡単に顔のイラストを作成できるようになれば，イラスト用顔画像の生成における可能性が広がると考えられる．また，顔の目や鼻，口などの位置情報であるランドマーク情報を利用し，画像生成を行うことで顔の特徴をとらえて学習できると考えられる．そこで，本研究ではランドマーク情報を利用した，イラスト用顔画像の生成手法を提案する．

1.2　本研究の目的

ディープラーニングの発展に伴い，Generative Adversarial Networks (GAN) [1]を用いた画像生成の研究が行われている．イラスト用顔画像もデータセットを用意することで簡単に生成ができる．しかし，それらで生成された画像の中には目がない画像や口がおかしい位置にあるなど，顔として認識できない画像も生成される．また，生成画像は学習に使用するデータセットに依存しており，顔の形状や表情を指定して生成することができない．そのため特定の形状や表情をもつ顔画像を生成するためには，時間をかけてデータセットに含まれる画像を選定しなくてはいけない．自分が生成したい形状や表情をもつデータがデータセットにない場合には，思い通りの顔画像は生成できない．そこで，データセットを選定することなく，顔の目や鼻，口などの位置情報を指定する．これにより顔画像の生成ができれば，自由度の高いイラスト用顔画像の生成が可能になる．したがって，本研究では顔のランドマーク情報を用いた画像生成により，指定した顔のランドマークをもつ顔画像の生成を目的とする．

1.3　関連研究

顔のランドマークを画像生成に取り入れた研究としては2019年にRuizheng Wu らによって提案されたLandmark Assisted CycleGAN for Cartoon Face Generation [2]が挙げられる．Ruizheng Wu らの研究では，CycleGAN [3]において顔のランドマーク情報を利用することで品質の高い画像の生成に成功している．しかし，CycleGANでは変換前と変換後の2種類の画像に関するデータがそれぞれ必要である．そのため，単一の画像で顔のランドマークを利用し，顔画像を生成している例はない．

1.4　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，目的及び関連研究について述べる.

第 2 章では本研究で用いるディープラーニングについて述べる.

第 3 章では本研究で提案する手法について述べる.

第 4 章では提案手法の実験，結果及び考察について述べる.

第 5 章では本論文の結論と今後の課題について述べる.

第2章　ディープラーニング

2.1　まえがき

本章では，本研究で用いるディープラーニングの技術である CNN，GAN について述べる．

2.2　ディープラーニングの概要

ディープラーニングとは機械学習の一つであり，階層の深いニューラルネットワークを用いて学習する．ニューラルネットワークとは，入力層，中間層，出力層が重なった構造になっており，適切な重みパラメータをデータから自動で学習することができるという特徴がある[4]．ディープラーニングではニューラルネットワークを階層化することにより、複雑な概念を単純な概念から構築して学習することを可能にしている[5].

2.3　CNN

Convolutional Neural Network(CNN)とは，ニューラルネットワークの一つであり，入力層，畳み込み層，プーリング層，全結合層，出力層を組み合わせることにより構成されている[4].CNN の基本構造を図 1 に示す. 全結合層だけでなく畳み込み層を追加することで，画像の 3 次元のデータをそのまま 3 次元のデータとして扱うことができ，形状を維持して学習ができる．そのため，CNN は画像認識や音声認識などさまざまな分野で利用されている．

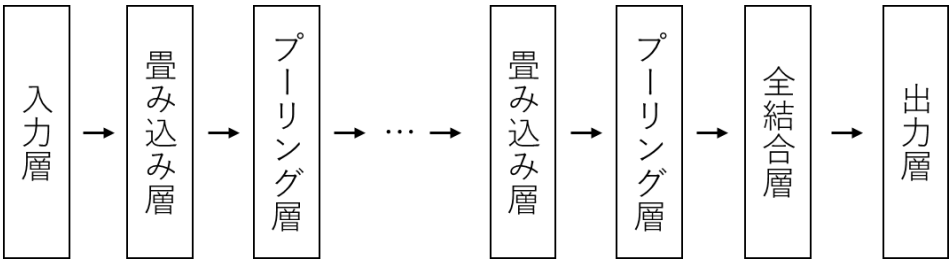

図 2.1 CNN の基本構造

2.4 GAN

2.4.1 GAN の概要

Generative Adversarial Networks (GAN)とは，2014 年に Ian J．Goodfellow らによって提案されたフレームワークである[1]．Generator と Discriminator の二つのモデルを同時に訓練し，敵対的プロセスを介することで生成モデルを推定する．GAN の基本構造を図 2.2 に示す．Generator では入力ノイズからデータセットに近い画像を生成し，Discriminator では入力されたデータがデータセットからきた本物の画像であるのか，それとも Generator が生成した偽物であるのかを識別する．Generator は Discriminator が間違った判定をするよう，より本物に近い画像の生成を学習し，Discriminator は本物と偽物を正しく識別できるように学習する．

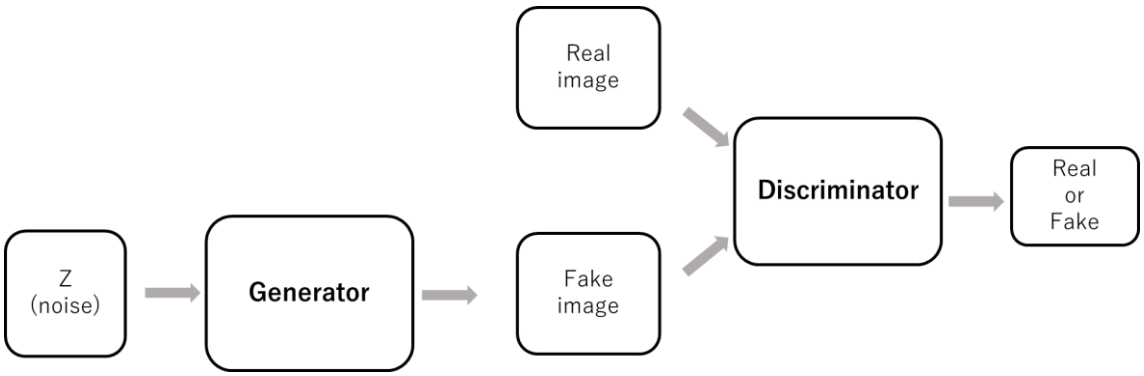

図 2.2　GAN の基本構造

2.4.2 WGAN-GP

GAN の安定化手法として 2017 年に Martin Arjovsky らによって Wasserstein GAN(WGAN) [6]が提案された．WGAN は，Generator の出力とデータセットの Wasserstein 距離を測り学習することで GAN の勾配消失やモード崩壊などの不安定性を改善した．しかし，WGAN で利用される Weight Clipping はパラメータにより学習が上手くいかないことや, 勾配消失をしてしまうなどの不安定性が生じる. そこで 2017 年に Martin Arjovsky らによって WGAN を改良した Wasserstein GAN-GP(WGAN-GP)[7]が提案された．WGAN-GP では Gradient Penalty を追加することで WGAN の安定化に成功した．WGAN-GP では以下の損失関数が最小化するように学習を行う．

$$\mathrm{L} = \mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[D(x)] + \lambda \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \tag{2.1}$$

2.5 むすび

本章では, 本研究で用いるディープラーニングの技術である CNN, GAN ついて述べた.

第3章　提案手法

3.1　まえがき

本章では，本研究で提案する手法の概要について述べる．本研究では，WGAN-GP の損失関数にランドマーク情報を追加した WGAN-GP を提案する．

3.2　提案手法

本研究で提案する手法の構造を図 3.1 に示す．

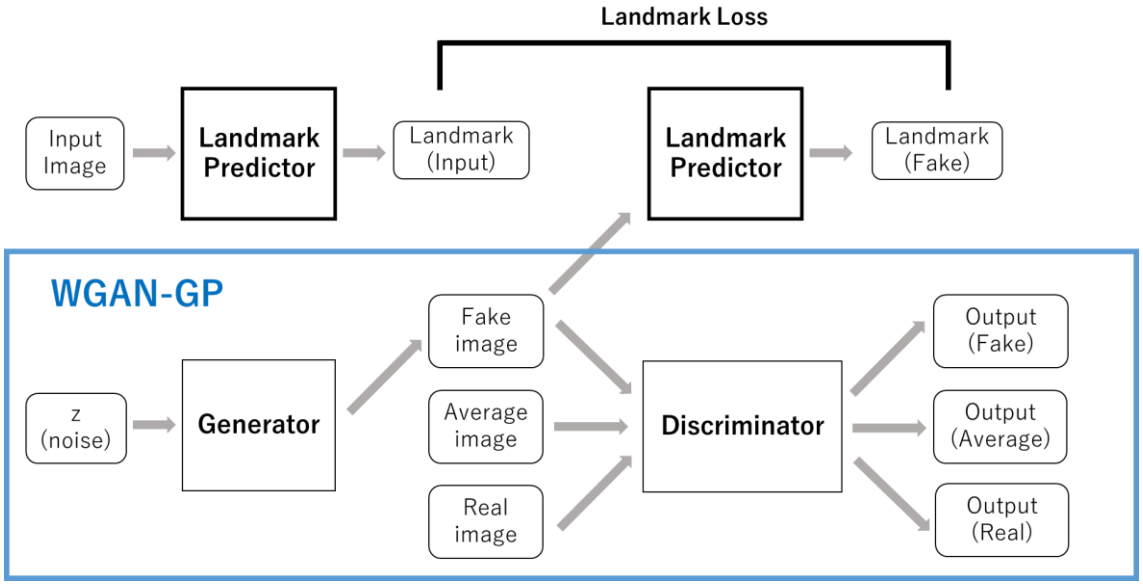

図 3.1　提案手法の構造

生成画像が入力画像のランドマークを再現できるよう WGAN-GP の基本構造に Landmark Predictor を追加し，損失関数に新たな損失関数である Landmark Loss を追加する．まず，Landmark Predictor について説明する．Landmark Predictor は CNN によって構成されている．使用するデータセットに手作業でランドマーク情報を加え，Landmark Predictor に画像を入力すると入力画像のランドマークを出力するように事前に学習させる．次に，Landmark Loss について説明する．Landmark Loss L_{Land} とは，以下に示すように，入力画像におけるランドマークと Generator で生成された画像のランドマークの距離を表す損失関数である．

$$L_{Land} = \lambda \| L(\tilde{x}) - l \|_2 \tag{3.1}$$

まず，生成したいランドマークをもつ画像を入力画像(input image)として Landmark Predictor に入力し，ランドマークlを出力する．また，Generator が生成した画像も同様に Landmark Predictor に入力し，ランドマーク$L(\tilde{x})$を出力する．それらの距離を測ることで Landmark Loss を求める．この損失関数は，生成画像が入力画像のランドマークと同じランドマークとなった場合に最小になる．

3.3 むすび

本章では，本研究で提案するイラストの顔画像生成のためのランドマーク情報を追加した WGAN-GP について述べた．

第4章　実験

4.1　まえがき

本章では，提案した手法の実験結果と評価および考察について述べる．

4.2　データセット

本研究では，kaggle で公開されている Bitmoji の顔画像データセット Bitmoji Faces [8] から 4,084 枚の画像を用いて実験を行った．本実験では，画像のサイズを 128×128[pixel] に正規化してから使用した．

4.3　ランドマークの指定方法

本実験では，入力画像と生成画像において検出するランドマークを左目 le，右目 re，鼻 n，口の右端 rm，口の左端 lm，顎の左側 c0，中央 c1，右側 c2 の合計 8 箇所に指定して実験を行った．ランドマークの指定方法を図 4.1 に示す．ランドマークを検出する際は画像を 100×100 [pixel]に正規化し，白黒画像に変換してから行う．

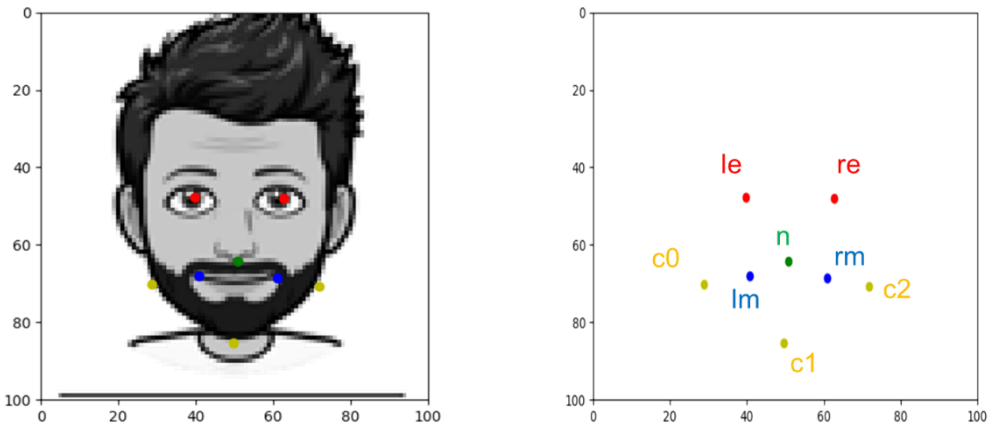

図 4.1　ランドマークの指定方法

4.4　提案手法と WGAN-GP の比較実験

学習回数を 25,000 回とし，提案手法と WGAN-GP で比較実験を行った．また，Discriminator (D) と Generator (G) の勾配更新回数比を D:G = 5:1，Gradient penalty の値を 1.0，Landmark Loss の値を 1.0 として実験を行った．ランドマークを指定するための入力画像は図 4.1 に示したものである．実験で使用したパラメータを表 4.1 に示し，提案手法と WGAN-GP により生成した画像を図 4.2，それぞれの品質の悪い画像の例を図 4.3 に示す．

表 4.1　実験で使用したパラメータ

パラメータ	値
学習回数 [回]	25,000
勾配更新回数比（D:G）	5:1
バッチサイズ	64
Gradient penalty の重み	1.0

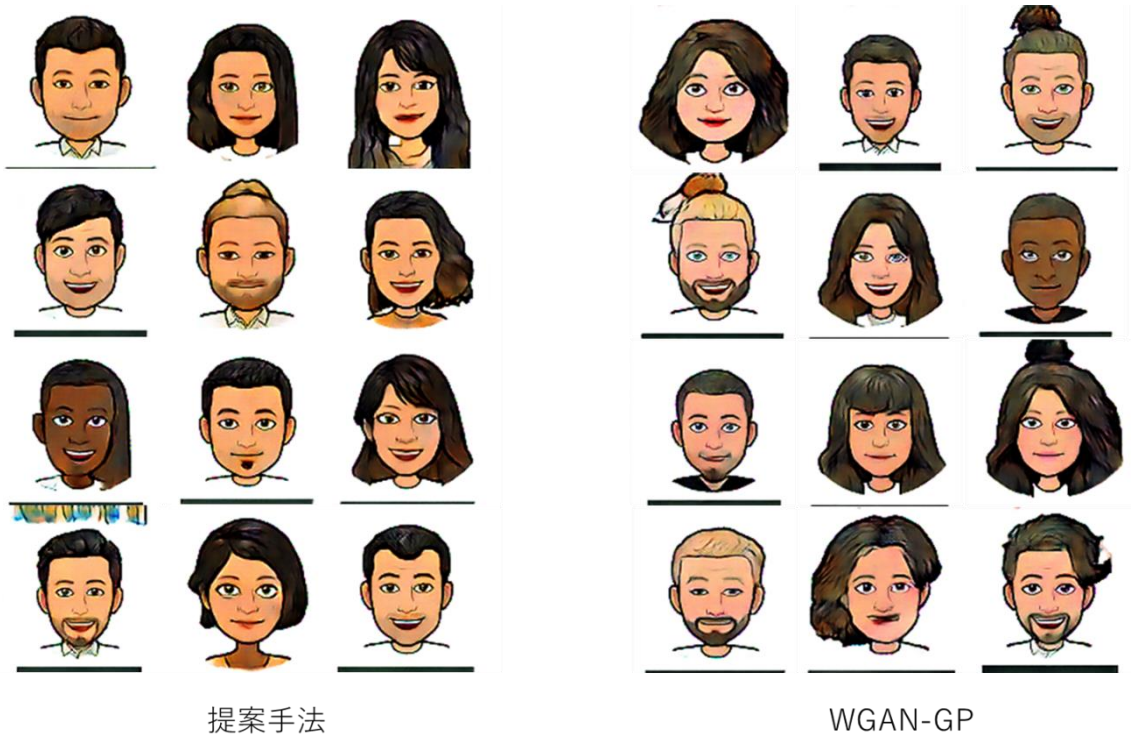

図 4.2　提案手法と WGAN-GP の生成結果

図 4.3　提案手法と WGAN-GP における品質の悪い画像

4.5　評価

4.5.1　入力画像と生成画像の Landmark の距離における評価

元のデータセットと WGAN-GP，および提案手法が生成した画像において，入力画像のランドマークとの距離を計測し，評価を行う．画像はそれぞれランダムに選択した 320 枚を用いた．比較画像はそれぞれ 100×100 [pixel]に正規化したものを用いる．入力画像とのランドマーク距離 $D_{Landmark}$ は以下の式で計算し，計算結果を表 4.2 に示す．

$$D_{Landmark} = \| L(\tilde{x}) - l \|_2 \tag{4.1}$$

表 4.2　入力画像とのランドマーク距離

画像	入力画像とのランドマーク距離
WGAN-GP の生成画像	0.2203
提案手法の生成画像	0.2180

4.5.2　FID による評価

WGAN-GP と提案手法の生成画像に対してそれぞれデータセットとの FID [9]を計算し，評価を行った．FID は二つのデータの距離を表すため，FID の値が小さいほどデータセットの画像に近い画像が生成できていると評価できる．画像はそれぞれ 320 枚用いた．FID の計算結果を表 4.3 に示す．

表 4.3　FID の評価

モデル	FID
WGAN-GP	136.1
提案手法	132.8

4.6　考察

表 4.2 に示すように, 提案手法は WGAN-GP よりもランドマーク距離が小さくなったが, 入力画像と同じランドマークをもつ画像を生成できなかった. その原因として, ランドマーク情報を Generator の生成条件に追加しなかったことが考えられる．ランドマーク情報を利用するために Landmark Loss を損失関数に追加したため，ランドマークが違うときに元の WGAN-GP の損失関数が小さくなったのか，Landmark Loss が小さくなったのかを判別することができない．そのため，提案手法では Generator がランドマークの詳細な情報まで学習できなかったと考えられる．

しかし, 表 4.3 に示すように提案手法の方は WGAN-GP に比べて FID の値が小さくなった. 図 4.3 に示すように, WGAN-GP で生成した画像における品質の悪い画像は, 目や鼻, 口がぼやけており顔として認識できない．一方で，提案手法においては，画像全体が歪んでいるものはあるが，ランドマーク情報として情報を与えた目，鼻，口が WGAN-GP に比べて崩れているものが少ない．また，目，鼻，口の形が明確になっている． Landmark Loss は Generator がランドマークを認識できない画像を生成したときに大きくなる. そのため, ランドマークを認識できない画像を生成したときに Landmark Loss が損失関数に大きく影響し，ランドマークが認識できる画像を生成するように学習できたと考えられる．

4.7　むすび

本章では，提案した手法の実験結果と評価，および考察について述べた．

第5章　結論と今後の課題

5.1　本論文のまとめと結論

これまでのイラスト生成では，画像として顔画像を生成するため顔として認識できない画像も生成されることがある，また，生成画像の顔の形状や表情を指定できないという問題点があった．そのため，顔画像においてランドマーク情報を利用した生成が有効であると考え，ランドマーク情報を利用した WGAN-GP によるイラスト用顔画像の生成手法を提案し，WGAN-GP との比較を行った．ランドマーク情報を利用して生成を行うために，WGAN-GP に新たに Landmark Loss という損失関数を追加した．実験した結果，提案手法は指定したランドマークと完全に一致する画像の生成はできなかった．しかし，Landmark Loss を追加したことにより顔として認識できない画像がされにくくなった．そのため，WGAN-GP よりも品質の高い画像の生成に成功した．これより，イラスト用顔画像において，ランドマーク情報を追加した学習が有効である．

5.2　今後の課題

イラスト用顔画像の生成において，ランドマーク情報を追加して学習させることで品質の高い画像が生成できた．しかし，入力画像と同じ形状，表情をもつ顔画像の生成はできなかった．自由度の高い顔画像生成を行うため，Generator の画像生成条件の見直しや他のネットワークにおいてもランドマーク情報を追加して実験するなど Generator に直接ランドマーク情報を与える方法を検討する必要がある．

謝辞

本研究の実験環境を整えてくださり，研究の方向性や丁寧なアドバイス，研究における問題点を提起してくださった渡辺裕教授に心より感謝申し上げます.

また，日頃から興味深い研究内容を共有してくださり，様々なアドバイスをくださった渡辺研究室の皆様に御礼申し上げます.

最後に，私をここまで育ててくださり，常に近くで支えてくれた家族に感謝いたします.

参考文献

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, “Generative Adversarial Networks”, Neural Information Processing Systems (NIPS), Dec. 2014.

[2] R. Wu, X. Gu, X. Tao, X. Shen, Y. W. Tai, and J. Jia, “Landmark Assisted CycleGAN for Cartoon Face Generation”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul 2019.

[3] J. Y. Zhu, T. Park, P .Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, IEEE International Conference on Computer Vision (ICCV), 2017.

[4] 斎藤康毅, ゼロから作る Deep Learning―Python で学ぶディープラーニングの理論と実装ー, オライリー・ジャパン, pp.39, 40, 205-207, 2016.

[5] Ian Goodfellow and Yoshua Bengio and Aaron Courville, “Deep Learning”, MIT Press, 2016, http://www.deeplearningbook.org（2021 年 1 月現在）.

[6] M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein GAN," arXiv preprint arXiv:1701.07875, 2017.

[7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, "Improved Training of Wasserstein GANs," arXiv preprint arXiv: 1704.00028, 2017.

[8] M . Mozafari , “Bitmoji Faces” , kaggle , Aug 2020 , https://www.kaggle.com/mostafamozafari/bitmoji-faces（2021 年 1 月現在）.

[9] 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 preprint arXiv:1706.08500v6, 2018.

図一覧

表一覧