

StyleGAN2 に基づくゲームキャラクターの 表情差分付き顔画像自動生成システム

陳 晴¹ 阿倍 博信¹

概要：近年、様々な実写顔やアニメ顔画像自動生成サービスが出てきた。しかし、これで得た画像は、ユーザの SNS アイコンとしては十分使えるが、ゲーム開発の素材としては表情差分生成機能がないため、不十分だと考えられる。そこで、本論文では、GAN の一方式である StyleGAN2 の画像生成と Style Mixing を組み合わせることで表情差分付き顔画像生成を自動化する方式について提案する。さらに、提案方式に基づき表情差分生成処理において誤りが発生した場合でも発生した誤りを手動でパラメータの値を修正する UI を持ったシステムを開発し、システム評価を行った。その結果、5 段階評価で最後まで表情差分付き顔画像生成を行うことと表情差分付き顔画像生成システムの使いやすさの平均値で 4.125 という結果が得られた。

An automatic facial image of game character generation system with facial expression difference based on StyleGAN2

Qing Chen¹ Hironobu Abe¹

1. はじめに

キャラクターを動かすゲームの開発においては、表情や顔向きの異なる表情差分付き顔画像が大量に必要な。通常は、ネットで探す、デザイナーに依頼する、自分で作る、などの方法が考えられるが、オリジナル素材を使用したい場合、デザイナーに依頼するか自ら制作する必要がある。しかし、この場合は時間と手間がかかってしまうという問題があった。

近年、様々な実写顔やアニメ顔画像自動生成サービス（例：This Waifu Does Not Exist[1], MakeGirlsMoe[2], Waifu Labs[3]）が出てきた。それらは敵対的生成ネットワーク(GAN[4]:Generative Adversarial Network)を用いて、指定のデータセットを学習して構築したモデルで画像を生成する。一般的に、パラメータの修正か、用意した各タイプ画像を選んで、最後にユーザが欲しい画像を 1 枚保存する。

また、この生成した画像を基に、さらに表情や顔向きの異なる表情差分を生成機能はついていない。もし表情や顔向きの異なる画像が欲しい場合、ユーザがパラメータを修正することが必要で、時間と手間が大変かかる。現時点では、指定のパラメータで編集可能だが、このパラメータを編集するとき、画像の他の部分も編集されてしまう場合もある。顔や髪なども変わってしまい、完全に別人になる可能性が大きい。これで得た画像は、ユーザの SNS アイコンとしては十分使えるが、ゲーム開発の素材としては不

十分と考えられる。

このような背景のもと、本研究ではゲーム作成支援のための顔画像生成と編集を自動化することで、オリジナル素材の利用と作業効率の向上を図る。画像生成を目的とした GAN の一方式である StyleGAN2[5]を用いて顔画像の表情差分生成方式を提案し、提案方式に基づく表情差分付き顔画像自動生成システムの開発及び有効性の確認を行う。

(図 1)



©2014 CloverLab.,Inc.

図 1 本研究の目標

以下、本論文では、2 章にて関連研究、3 章にて表情差分生成方式の提案、4 章にて表情差分付き顔画像自動生成システムの開発、5 章にて評価、6 章にて考察について述べ、最後にまとめを行う。

2. 関連研究

潜在空間のコントロールで顔画像の編集が可能である。関連研究として、StyleGAN2 に対応した潜在空間の研究が存在する。潜在空間の分析手法は、主に主成分分析に基づく GANSpace[6]、潜在セマンティック解析に基づく SeFa[7]がある。

GANSpace は、潜在空間または特徴空間に適用される主成分分析 (PCA) で、重要な潜在方向を特定する手法を提案した。PCA とは、多次元データのもつ情報をできるだけ

¹ 東京電機大学
Tokyo Denki University

損わずに、低次元空間に情報を縮約する方法である。普段の潜在空間分析は、ラベル付けされた生成画像や事前学習済み分類器などが必要となる。GANSpace では、教師なし方法で画像を分析し、潜在空間におけるより豊かな制御技術を発見することが可能である。

SeFa は、事前に学習済み潜在変数を直接分解することにより、潜在セマンティック解析(Latent Semantic Analysis, LSA)のための閉形式の因数分解アルゴリズムを提案した。潜在空間の無視できる方向を除外し、画像合成にとって重要な潜在方向を強調していると主張する。学習済みモデルから潜在空間を分析して、教師なしの方法で根本的な画像の変動要因を明らかにするのが可能である。

以上の2つの関連研究は教師なし学習であり、潜在方向を探索することが可能である。しかし、探索した潜在方向は画像のどの部分をコントロールできるかについて、手作業で1個ずつ確認しなければならない。

提案方式では、実際に必要な画像を大量に集め、重心を求める方法で潜在方向を探索し、コントロールしたい部分に関する潜在方向を見つける方式とする[8]。

3. 表情差分生成方式の提案

3.1 基本方針

顔画像の表情差分生成の際には、複数のパラメータが存在している。他のパラメータからの影響を減らすため、同じパラメータの画像を大量に集め、潜在変数の重心を算出する手法で探索を行い、構築した潜在空間からパラメータの潜在方向を求める。顔向きや表情などの表情差分付き顔画像の自動生成を可能にする。

3.2 概要

本論文にて提案する StyleGAN2 を用いた表情差分生成方式の概要を図2に示す。

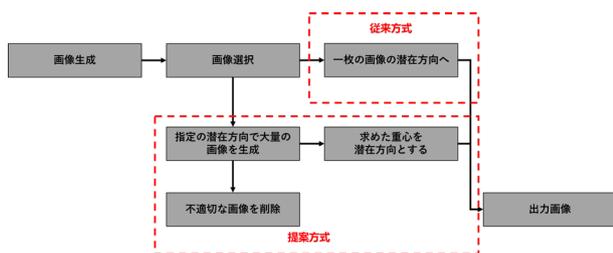


図2 提案方式の概要

まず、StyleGAN2 で転移学習したモデルを用いて顔画像を生成し、その中から表情差分生成処理用の画像を選択する。従来方式では、1枚の画像から求めた潜在方向で処理を行い、これで得た精度が低く、また他のパラメータも編集されてしまう問題も存在するため、表情差分生成処理に使用することは難しかった。そこで、1枚の画像の代わりに、従来方式により同じパラメータの画像を複数枚生成し、目視にて不適切な画像を除外した上で潜在空間における画像間の重心を算出、重心で求めた潜在方向として表情差分

生成処理を行うことで、出力画像（表情差分付き顔画像）を得る方式を提案する。

4. 表情差分付き顔画像自動生成システム

4.1 開発方針

本章では、3章で提案した表情差分生成方式を適用したシステムの開発について述べる。システムの開発にあたり3.1の基本方針に従い、表情差分生成処理で発生した誤りについては、ユーザが手動でパラメータの値を修正する方針とする。また、今回修正の対象とする誤りは、指定のパラメータで対応の画像にならなかった場合のみとする。対応の画像にならなかった場合、手動でパラメータの値を修正する。表情差分生成処理に対応するパラメータはSmile1, Smile2, Left, Rightの4種類とし、パラメータの説明を表1に示す。

表1 パラメータの説明

パラメータ	画像例	説明
Smile1		口があいている笑顔
Smile2		口が閉まっている笑顔
Left		顔が左に向いている
Right		顔が右に向いている

©2014 CloverLab.,Inc.

4.2 システム構成

本システムで使用するPCのスペックについて表2に示す。

表2 PCのスペック

プロセッサ	Intel Core i7-8700
システムクロック	3.20 GHz
RAM	32 GB
ビデオカード	NVIDIA GeForce GTX 1080Ti
OS	Ubuntu 16.04 LTS

本システムはpywebio[9]を用いてWebアプリケーションを構築した。pywebioとは、HTMLやJavaScriptファイルを作らずにWebアプリケーションを開発できるPythonのフレームワークである。顔画像生成、表情差分生成処理とUIは、Pythonを用いて、StyleGAN2のモデルで生成した画像を自動的に表情差分付き顔画像に変換する処理を実装した。

4.3 システム概要

提案方式を用いて表情差分付きのゲームキャラクタの顔画像を自動生成するシステムを構築した。本システムを使用することで、1枚の顔画像を生成し、この画像を基いてさらに表情や顔向きが異なる画像を自動生成する。ユーザーがUIを用いて表情差分生成処理のためのパラメータの値を修正し、異常な画像と指定されたパラメータになっていない顔画像の修正が可能である。本システムのアクティビティ図を図3に示す。

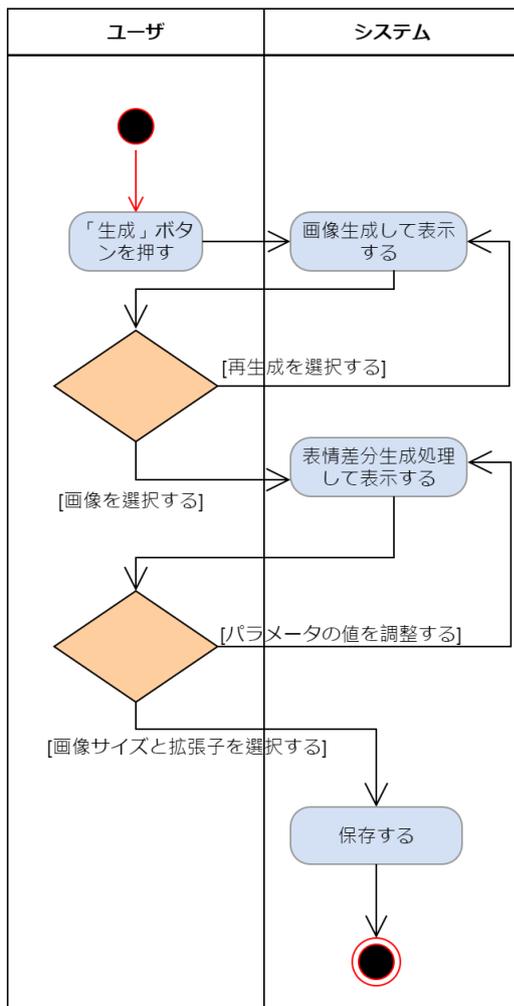


図3 本システムのアクティビティ図

4.4 画像生成機能

転移学習により構築した StyleGAN2 モデルを用いて、表情差分付き顔画像を4枚生成する。以下に、学習データの準備、転移学習によるモデルの構築と StyleGAN2 による画像生成について説明する。

4.4.1 学習データの準備

ガイドラインに従う形での二次利用が許諾されているスマホ RPG ゲームのゲームキャラクタの顔画像データ [10] を収集した。人型の顔画像が多いが、モンスターなどの顔画像も入っているため、モンスターなどの人物以外の顔画像データを削除し、筆者が事前生成した顔画像データ (130枚) を追加したデータとして372枚準備した。

次に収集した顔画像について、表情と顔向きに着目して、調査を行った。各パラメータについて、口が開いている笑顔 Smile1 に、口が閉まっている笑顔 Smile2 とし、表情に関するパラメータを表3に、顔向きに関するパラメータを表4に示す。

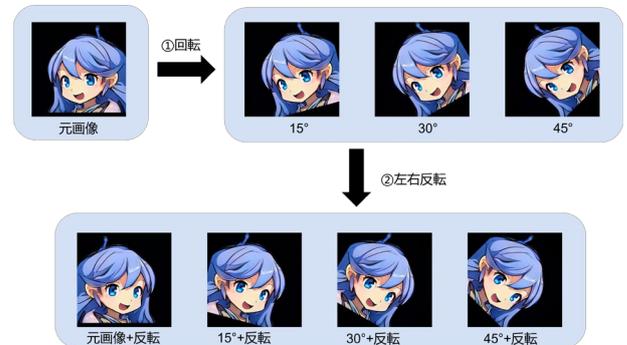
表3 データセットにおける表情パラメータ

パラメータ	Smile1	Smile2	その他	合計
枚数	137	114	121	372
%	37%	31%	32%	100%

表4 データセットにおける顔向きパラメータ

パラメータ	Left	Right	その他	合計
枚数	299	53	20	372
%	80%	14%	6%	100%

データ拡張として図4の通りで回転(15°, 30°, 45°), 左右反転処理を行った結果2,976枚となった。また、転移学習用の学習済みモデルは512px×512pxの画像を生成するため、学習データは同じサイズとした。



©2014 CloverLab.,Inc.

図4 学習データの前処理の流れ

4.4.2 転移学習によるモデルの構築

アニメの顔画像を学習済みの StyleGAN2 モデル [11] に対して、データ準備したデータセットを学習データとして転移学習を行う形で、ゲームキャラクタの顔画像を生成する StyleGAN2 のモデルを構築した。学習時間は23時間58分37秒となった。

4.4.3 StyleGAN2 による画像生成

作成した StyleGAN2 モデルで画像を4枚生成する場合、平均9秒かかる。例として画像を21枚生成した結果を図5に示す。図5のように、生成した画像の中には、髪に謎な飾りものが入っている場合や、顔になっていない画像なども含まれる場合がある。



©2014 CloverLab.,Inc.

図5 画像生成例

4.5 表情差分生成機能

StyleGAN2 では、2 枚の画像を Style Mixing により、画像編集を行うことができる。Style Mixing とは、画像の指定のレイヤーの潜在変数を入れ替える機能である。Style Mixing の概要を図 6 に示す。

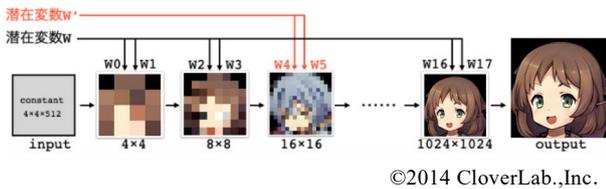


図 6 Style Mixing の概要

提案手法の通りで、代表となる 4 つのパラメータの画像を選んで、指定パラメータの画像を大量生成して潜在変数の重心を算出する。画像の指定のレイヤーの潜在変数を、算出した重心との差（求めた潜在方向）を比率(-1~1)で足し、この潜在変数を用いて表情差分付き顔画像を生成する。

4.6 UI

表情差分付き顔画像自動生成システムの UI について述べる。

4.6.1 ページ 1

システムの目的、手順、「生成」ボタンを表示する。「生成」ボタンを押すと、顔画像を画像生成してページ 2 に切り替える。



図 7 ページ 1 の画面

4.6.2 ページ 2

生成した 4 枚の画像の中から、ユーザが使いたい画像を選択する。「再生成」ボタンを押すと、新たな 4 枚画像の生成が可能である。「次へ」ボタンを押すと、表情差分生成処理を行ってページ 3 に切り替える。



図 8 ページ 2 の画面

4.6.3 ページ 3

(1)結果表示

選択した画像（左）と表情差分生成処理の結果（右）が表示される。

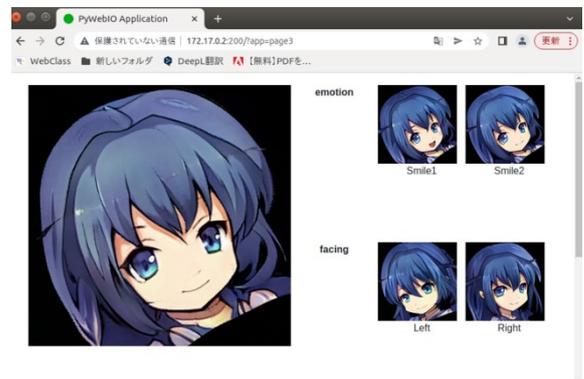


図 9 ページ 3 の結果表示

(2)パラメータ修正

修正したい画像を押すと、パラメータ修正ウインドウが表示される。4 つのパラメータの値を-1~1 修正した上で、「change」を押して表情差分生成処理を再び行う。「Recover」ボタンで選んだ画像を元に戻せる。



図 10 ページ 3 の修正画面

(3)指定のサイズ、拡張子で出力

表情差分生成処理の修正が終わった後、結果表示部分の下のサイズと拡張子のボタンを選択して、「保存」ボタンを押すことで、表示している画像結果を指定のサイズと拡張子を変更して zip ファイルに圧縮してダウンロードできる。



図 11 ページ 3 の結果表示

5. 評価

5.1 提案方式の評価

評価方式として、構築した StyleGAN2 のモデルを使用して生成した画像に基づき、指定の潜在方向へ編集された画像の精度を評価する。編集された画像の中に、顔になっていない画像と指定のパラメータになっていない画像を失敗画像にして成功画像と分ける。即ち、新たに 200 枚の顔画像を生成する。生成した顔画像に対して、表情差分生成処理を行う。目視で画像の確認を行い、成功画像と失敗画像分類する。最後に、成功画像の比率で精度を評価する。

提案方式の評価にあたり、Step1:従来方式、Step2:Step1 から算出した重心で求めた潜在方向での提案方式、Step3:Step2 から算出した重心で求めた潜在方向での提案方式とし、手作業で画像を分類して精度評価を行った。表 5 にその結果について示す。

表 5 提案方式の評価結果

パラメータ	Smile1	Smile2	Left	Right	平均値
Step1	37%	34%	25%	30.5%	31.6%
Step2	30.5%	36%	30.5%	30.5%	31.8%
Step3	49%	38%	48%	36.5%	42.9%

算出した重心で求めた潜在方向の差異を分かりやすくするため、例として 3 枚画像 (Seed1641, Seed0000, Seed0664) をあげて重心と Style Mixing した結果を図 13, 図 14, 図 15 に示す。

5.2 システムの評価

開発した表情差分付き顔画像自動生成システムの有効性の確認を目的としたシステム評価を実施した。具体的には、大学生 8 人に対して、本システムを用いて実際に表情差分付き顔画像自動生成作業を行う実験を実施し、アンケートを用いた評価を行った。実験者には 4 つのパラメータについて、画像のサンプルなどの説明資料を提示し、事前にパラメータを理解した状態で、アンケートによる主観評価を実施した。評価の流れのイメージを図 12 に示す。

評価方法として、まず画像生成を行い、実験者が選択した画像を処理する。次に、実験者は選択した画像の変換結果に対して、自由に 4 つのパラメータの値で修正 (-1~1) してもらい、指定の画像サイズと拡張子で結果をダウンロードする。これらの結果を表示したのち、アンケート評価を行う。

アンケート評価の項目を表 6 に、5 段階評価で評価結果



図 12 システム評価の流れ



©2014 CloverLab.,Inc.

図 13 Step1 の Style Mixing 結果



©2014 CloverLab.,Inc.

図 14 Step2 の Style Mixing 結果



©2014 CloverLab.,Inc.

図 15 Step3 の Style Mixing 結果

のグラフを図 16 に示す。8 名の実験者の成果物について、実験で表情差分生成処理の結果を奇数行で、実験者が修正した結果を偶数行で、8 回の実験の成果物の画像を図 17 に示す。

©2014 CloverLab.,Inc.

表 6 評価項目

アンケート内容	
Q1	最後まで表情差分付きを行うことができましたか？
Q2	自動で表情差分付いた結果について満足ですか？
Q3	システムは使いやすいですか？
Q4	システムの操作に迷うことがありましたか？
Q5	ボタンなど UI の位置は適切でしたか？
Q6	実際にあったらゲーム素材として使いたいですか？



図 16 評価結果

6. 考察

6.1 提案方式の評価に対する考察

6.1.1 評価結果の考察

表 5 から、複数枚画像から求めた潜在方向を用いた Step2 の Smile1 の精度が少し下がったが、前のステップと比較し、精度が穏やかに上がったという結果が得られた。すなわち、1 枚の画像の潜在方向と複数枚画像から求めた潜在方向と比べて、1 枚の画像より、複数枚画像から求めた潜在方向の方が指定の潜在方向へ編集する精度が高くなる。大量のデータから最も適切な潜在方向を特定、指定の潜在方向へ編集させる精度を上げるのが可能だと考察できた。また、精度の平均値が 50%以下となった原因は、図 5 の画像生成例のように顔になっていない画像も混ざっている可能性が大きいと考えられる。

そのため、事前に使えない顔画像を削除するか、モデルを再構築すること自動生成された顔画像の品質を上げる必要があると考えられる。

6.1.2 Style Mixing 結果の考察

潜在方向に関しては、画像編集用の潜在空間の重心を算出する手法で特定できていることが確認できた。図 13, 図 14, 図 15 の Style Mixing 結果を確認すると、1 行目の画像は各ステップで算出した重心画像である。この 12 枚で画像の中で、表情の方は、口が閉まっている Smile2 の口部の変化が極め



©2014 CloverLab.,Inc.
図 17 実験成果物

て小さくなった。元々口が開いている Smile1 の口が逆に小さくなった。顔向きはよく似ているが、髪の毛のところに謎な飾りが付いている結果となった。

表情について、指定のパラメータの通りで編集していない結果となった。例えば、元画像の口が閉まっている Seed1641, Seed0664 は、口が閉まっている Smile2 に編集した場合は、逆に口が少し開く状態になる。同じ方法で、元画像の口が開いている Seed0000 は、口が開いている Smile1 に編集した場合は、口がなくなる状態になる。顔向きについて、指定のパラメータの通りで編集したが、髪の毛も少し編集されて、別の髪型になった。髪の毛の上に変な飾りも付いている結果となった。

以上のように、元画像の状態を確認しないと、予想通りにならない場合や、おかしな画像が得られる可能性がある。

6.2 システムの評価に対する考察

6.2.1 アンケートに対する考察

図 16 において、5 段階評価で全問平均値が 3 以上となった。システム操作に関する Q4 は、2 人が 2 点をつけ、平均値が 3.125 となった。また、実験時は実験者がシステムの操作に迷ったため、時間かかった場合もあった。そのため、UI 上にボタンの説明や作業のフローなどを提示するなどの改善が必要であると考えられる。

Q1, Q2, Q3, Q4, Q5 に、中間値 3 を選んだ実験者が数名いた。現在の操作説明、UI の位置、画像結果などに満足できていないと考えられる。一方、実験者がパラメータの値を修正した際、編集前後画像の差異が小さい問題で数回修正が必要になった場合や、評価実験の際、想像通りにならなかったことで戸惑った場合や、パラメータ値を大きくし過ぎた場合に、画像が顔にならないこともあった。

6.2.2 成果物に対する考察

成果物に対して評価した結果、実験者により様々な状況が存在することが確認できた。確認できた状況は 1)修正後逆に誤りが発生した、2)ユーザ修正前後の差異が小さくなった、3)顔になっていないなど画像が入っている、などがあり、各項目の発生原因について考察する。

表情差分生成処理の結果について、実験者 1, 2, 4, 7 の画像は修正の必要のない画像が得られた。また、ユーザ修正を行った後、実験者 4, 6, 7, 8 の画像は各パラメータに対応した画像になり、逆に実験者 1, 2 の画像は違った結果になった。

実験者 3 が選んだ画像は、修正前後問わず、4 つのパラメータに対応した画像を得られなかった。実験者が適当に値を修正することも可能だが、元画像の髪に飾りものの部分が大きいため、潜在変数で再現しても顔画像にならない状況だったと考えられる。

実験者 4, 5 が選んだ画像は、修正前後の画像差異が小さいため、パラメータの値を修正せずそのまま保存する可

能性が大きい、実験者 4 が自動生成の結果のみで満足できると考えられるが、実験者 5 はパラメータを 0.1 ごとに修正しても画像が想像通りに編集できなかったため、戸惑った様子だった。

7. おわりに

本論文では、ゲーム制作における素材準備の一環である顔画像の表情差分生成処理を対象として、StyleGAN2 の生成画像に対して同じベース画像の処理を行うことで、重心を潜在方向として使う表情差分生成処理を自動化する方式について提案した。各パラメータの潜在変数の重心を算出する手法を用いて、StyleGAN2 で生成した画像の表情差分生成処理の自動化を試みた。提案方式について精度評価を行った結果、2 回で平均 31.8%、42.9%の精度を得られた。次に、提案方式を用いた表情差分生成処理において誤りが発生した場合、発生した誤りを手動でパラメータの値を修正する UI を持ったシステムを開発し、システム評価を行った。

その結果、従来方式と比較して、1 枚の画像より、複数枚画像から求めた潜在方向の方が指定の潜在方向へ編集させた精度が高くなった。大量のデータから最も適切な潜在方向を特定、指定の潜在方向へ編集させる精度を上げることが可能である。また、システムの評価結果により、本システムの有効性について確認することができた。システムの評価結果は 5 段階評価で平均 4.125 という結果が得られた。

今後の課題として、提案方式については、他の表情差分付き顔画像を増やすため、学習データに対応した表情差分付きの顔画像の追加が必要である。また、精度の向上のため、他の表情差分に必要な潜在方向を発見する手法の検討、モデル構造の見直しと後処理による精度の向上、などがある。また、システムについては、システム操作に迷った実験者がいたことから、システム上に操作の手順をわかりやすく提示することや、ユーザが認識できなかった領域を強調するなどの UI の改良についても検討していく予定である。

参考文献

- [1] Gwern Branwen: "This Waifu Does Not Exist", <https://www.thiswaifudoesnotexist.net>.
- [2] Yanghua Jin: "MakeGirlsMoe", <https://make.girls.moe>.
- [3] sizigi studios: "Waifu Labs" <https://waifulabs.com/>
- [4] Tero Karras, Samuli Laine, Timo Aila: A style-based generator architecture for generative adversarial Networks, <https://arxiv.org/abs/1812.04948> (2018).
- [5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila: Analyzing and improving the image quality of stylegan, <https://arxiv.org/abs/1912.04958> (2019).
- [6] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, Sylvain

- Paris, "GANSpace: Discovering Interpretable GAN Controls",
<https://arxiv.org/abs/2004.02546> (2020).
- [7] Yujun Shen, Bolei Zhou, "Closed-Form Factorization of Latent Semantics in GANs", <https://arxiv.org/abs/2007.06600> (2021).
- [8] 陳晴, 阿倍博信: "StyleGAN2に基づく表情差分付きゲームキャラクターの顔画像の自動生成", 情報処理学会 第84回全国大会 4N-01 (2022).
- [9] Weimin Wang: "pywebio",
<https://pywebio.readthedocs.io/en/latest/>.
- [10] クローバーラボ: "ゆるドラシル RPG ツクール素材提供サイト", <http://yurudora.com/tkool/>.
- [11] "2020-01-11-skylion-stylegan2-animeportraits-networksnapshot-024664.pkl". <https://www.gwern.net/Faces>.