

# VAE を用いた 2D キャラクタの特徴抽出の検討

大道昇<sup>†1</sup> 大井翔<sup>†2</sup> 佐野睦夫<sup>†1</sup>

**概要:** 近年アニメ業界は成長し続けており、それに伴いアニメ画像を用いた画像生成や分類の研究も盛んにおこなわれている。なかでも GAN を利用した新たなアニメキャラクターの生成は大きな注目を集めている。他にも、視覚障害者が楽しむためのアニメキャラクターの分類や色構造のグラフを用いてアニメキャラクターの類似度を測定する論文がある。しかし、これらの研究では、分類精度があまりよくない。原因として、既存の現実の人に対する研究と違い、骨格や顔の部位の大きさが実際の人とはかけ離れていたり、絵師によって描き方が変わってくることに起因している。よって本研究では、VAE(Variational Auto Encoder)を用いることによって 2D キャラクターイラストの潜在的特徴抽出を行うことでキャラクターの分類などのプログラムの基礎的な部分の貢献ができないか検討する。結果として 2D キャラクターの確率的な分布から、一部の特徴を抽出することに成功した。

## 1. はじめに

一般社団法人日本動画協会のアニメ産業レポート 2018 サマリー版によると[1], 日本のアニメ市場は最高売上値が 5 年連続更新されており、ユーザーが支払った金額を推定したアニメ市場の推移によると、2015 年から衰退傾向にあるものの商品化が市場割合の 2 割を占めている。それに伴い、アニメに関する論文も徐々に増えてきている。また、イベントや地域おこし、名産品の紹介などイラストや着ぐるみなどのマスコットキャラクターとしてたびたび使用されている”ゆるいマスコットキャラクター(ゆるキャラ)”や、動画サイト YouTube でブームとなっているバーチャル YouTuber(VTuber)などがあり、マスコットキャラクターや 2D イラストキャラクターが人とのつながりやインタラクティブ性を生み出している。

近年、機械学習を用いた精度の高い画像を生成するモデルが盛んに研究されており、そのなかでも代表的な理論が GAN (Generative Adversarial Network) である[2]。また、GAN の理論を用いた研究は多く、用途に合った多様な GAN が考案されている。GAN の訓練画像はイラストにも適用することができ、2D キャラクターの顔画像を生成することができる GAN も考案されている。

このようにアニメや 2D のキャラクターを用いた研究は徐々に増えてきており、分類の分野では視覚障害者がアニメを楽しむためのアニメキャラクターの分類や[3], 色構造のグラフを用いてアニメキャラクターの類似度を測定する研究[4], 顔のパーツ座標からアニメキャラクターの顔の特徴量抽出などがある[5]。しかし、これらイラストのキャラクターを用いた研究では、写真などの現実の人画像を用いた研究より精度は劣っていると言える。原因として、まだまだ 2D キャラクターイラストを用いた研究が少なく、データがそろっていない点や既存の人物検出プログラムは学習データが人であ

り、キャラクターを検出することを想定していない。また、人を検出する時と違い、実際の人より目や顔が大きかったり、骨格が異なっていたり、色がべた塗りや水彩画のような色合いなど 2D のキャラクターイラストは絵師によって独特な表現がされているためだと思われる。そのため、2D イラストの画風ごとに学習しなおさなければならない。そのため、各 2D イラストの潜在的特徴を抽出することは将来的にさらに発展していくであろうアニメの分野の研究における基礎的な部分に必要なと思われる。

本研究では訓練画像から連続的に変化する画像を生成と、訓練画像の潜在的な分布が得ることができる VAE(Variational Auto Encoder)を用いることによって[6], 2D キャラクターイラストの潜在的特徴の抽出を検討する。VAE を用いることによって得られる潜在変数の分布から、2D キャ

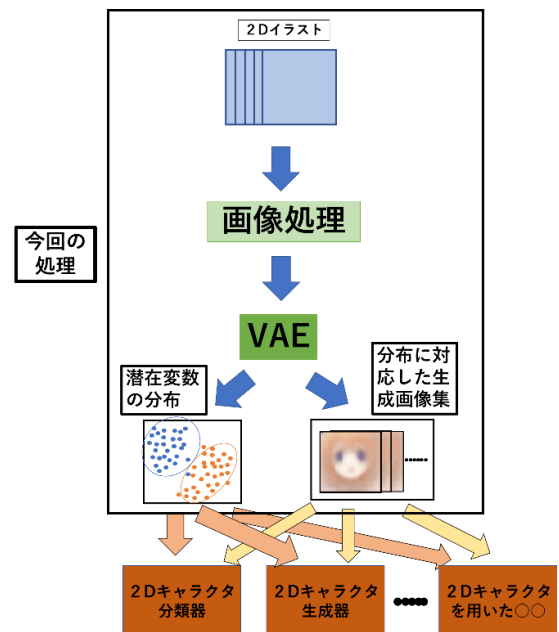


図1 提案処理の流れ

<sup>†1</sup> 大阪工業大学,

<sup>†2</sup> 立命館大学

ラクタのそれぞれの特徴だけでなく、各イラストの画風なども特徴として得られるのではないかと考える。

## 2. 関連研究

### 2.1 視覚障害者がアニメを楽しむための 機械学習によるキャラクター識別システム

この研究では視覚障害者がアニメ映像等のコンテンツを楽しむ際に生じる無言シーンでのキャラクタ把握問題に焦点を当てた情報保障やアクセス支援を目的としている。視覚による情報取得に制約があると、アニメ画像を視聴する際に「行動や状況を理解することの難しさ」、「キャラクタ識別の難しさ」、「似ているキャラクタの混同」といった問題があり、そこに焦点を当てている。この問題を解決するために機械学習によるキャラクタ識別のための支援機器の製作を検討している。

### 2.2 色構造のグラフを用いたアニメキャラクター画像間の類似度測定

この研究ではアニメ画像のキャラクタ間の類似度を測定する方法を提案し、その性質を実験結果によって示している。各キャラクタが顔の造りよりむしろ服装や髪の色と形によって主に描き分けられているということに注目し、キャラクタの特徴を効率的に取り出す手法を提案している。具体的にはキャラクタ領域を類似色セグメントに分割したものをグラフで表し、各セグメントの平均色の  $L^*a^*b^*$  表現に基づいて、部分グラフの Tree-walk に対する畳み込みツリーカーネル関数を利用している。色合いとその配置が良く似ているが異なるキャラクタの画像を含む実験の結果、良好な類似度を求めることが可能となった。

## 3. 提案システム

本研究では、アニメ関連の研究の基礎的な部分である 2D キャラクタの特徴を抽出することにより、今後のアニメなどの 2D イラストを用いた研究の貢献することを目指す。2D キャラクタの特徴を抽出するにあたって VAE を使用した。その際、2D キャラクタの顔画像のみを利用することとした。

### 3.1 画像処理

VAE で 2D キャラクタの特徴を抽出するにあたって、複数枚の 2D キャラクタの顔画像のイラストが必要となってくる。そのため、2D キャラクタのイラストの画像をサイトからダウンロードし、その後自動で 2D キャラクタの顔画像のみを切り抜くプログラムを作成した。2D キャラクタのイラスト画像は一枚一枚画像を検索してダウンロードするのは手間がかかるために、Google 画像のダウンロードツールである”google-images-download”を利用した[7]。この際に、後述するプログラムで利用できるように、2D キャラク



図2 “lbpcascade\_animeface”を用いたアニメキャラクターの顔検出[8]

画像の出典: 「アイドルマスター ミリオンライブ!」

©窪岡俊之 ©BANDAI NAMCO Entertainment Inc.

©BNEI/PROJECT iM@S

タのラベルごとに仕分けるようにしておく。前述した方法によってダウンロードした画像に対して OpenCV を用いた作成プログラムによって顔画像のみを切り抜く。その際に “lbpcascade\_animeface.xml” というカスケードファイルを使用した[8]。図2に顔検出の結果を示す。このカスケードファイルはアニメキャラクターの顔画像を検出できるように作成されている。本研究で使用する 2D キャラクタでもある程度検出されるようなので使用した。

### 3.2 VAE を用いた特徴抽出

画像処理を行った 2D キャラクタの顔画像イラストから VAE を用いて特徴抽出を行う。VAE は入力画像を Encoder で畳み込みを行い、潜在変数にまで次元削減をし、Decoder で同じ画像を出力する。この時、潜在変数を確率分布という構造に押し込めることを可能としたのが VAE である。そのため、学習時の入力データは訓練データのみで教師データを必要とせず、入力画像を表現する特徴を潜在変数の確率分布で確認することができる。また、確率分布に押し込められた潜在変数から、分布の位置に対応する画像を復元することができる。これらの VAE の特徴を利用することによって、2D キャラクタの顔画像イラストから潜在変数の分布によって特徴抽出を行うことができるか検討する。

VAE を用いて特徴抽出を行う際に、ラベル分けされたファイルを複数選択することによって、各ラベル分けされたイラストの特徴をラベルごとに指定した色によって潜在変数の分布として表現することができる。そのため、出力された確率分布にマッピングされた 2D キャラクタの顔画像イラストの潜在変数の位置から、特徴が抽出で生きているか確認する。また、VAE の入力画像のピクセル数、学習回数、潜在変数の次元数を変更できるようにし、変更したことによる変化を見られるようにした。既存の VAE は、確率分布を画像として出力する際に潜在変数が 2 次元までしか見られなかったため、潜在変数の次元数を増やすにあたって、次元数が 3 以上の場合は出力するそれぞれの次元での総当たりの確率分布を 1 つの画像に出力できるように改良した。

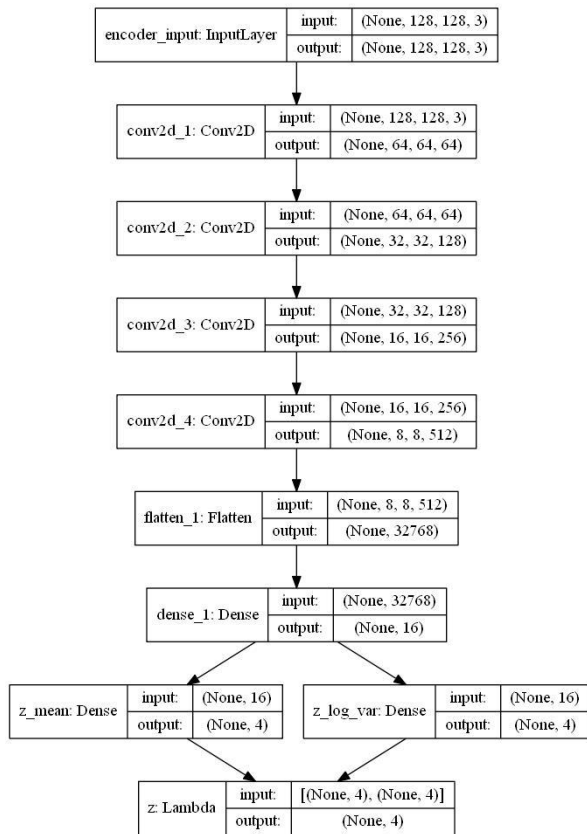


図3 使用する VAE の Encoder

使用する VAE の Encoder の中身の例を図3に示す. 提案する VAE は入力画像を  $128 \times 128$  の RGB 画像として入力できるようにしてあり, 畳み込み層は4層としている. そして, 図3では潜在変数の次元数を4としているが, この次元数は変更できるようにしてあり, 次元数の変化によってどのような影響が出るか比較できるようにしている. 学習回数は100回に統一している. また, 入力画像をテストデータと学習データをランダムに分けているので, 出力結果が毎回異なるようになっている.

## 4. 実験

2D キャラクタの顔画像を多数収集し, VAE の畳み込み層や潜在変数の次元数を変えることによって2D キャラクタの特徴抽出に違いが出るか検討する. 以下に, アニメキャラクターの特徴とゲームキャラクターの画風特徴の抽出実験を行う. ここで, アニメとゲームに着目した理由として, 我々の知る限り, 2D イラストが豊富であると考えたからである. また, キャラクタごとの特徴抽出にアニメキャラクターを用いる理由として, アニメキャラクターはゲームなどと違いキャラ数は比較的少ないが, 公式と2次創作を含めた各キャラクターのイラストが豊富であると考えたためである. そして, 画風の特徴抽出においてゲームキャラクターを選んだ理由として, キャラクタ数が多いため画風が統一された公式イラスト数が多いと考えたためである.

### 4.1 アニメキャラクターの特徴抽出

アニメキャラクターの特徴を抽出するにあたって, ”ご注文はうさぎですか?”と”がっこうぐらし!”のキャラクターイラストをそれぞれ使用した[9][10]. 使用したキャラクターは, ”ご注文はうさぎですか?”から”チノ”・”千夜”・”ココア”・”リゼ”, ”がっこうぐらし!”から”くるみ”である. 図4に使用するキャラクターの画像例を示す. それぞれのキャラクターに対して各500枚前後の顔画像イラストを収集した. この時, 収集した顔画像イラストは公式のイラスト以外にも2次創作のイラストも含めている. 特徴抽出する組み合わせは, ”チノ”・”千夜”・”ココア”と”リゼ”・”くるみ”である. ”チノ”・”千夜”・”ココア”の組み合わせを実験する理由として, VAE は2Dのキャラクターに対しても特徴を抽出できるか確かめるためである. ”リゼ”・”くるみ”の組み合わせを実験する理由として, ネット上でこの2キャラクターはとても似ていると言われているからである. そのため, VAEによって公式のイラスト以外にも2次創作のイラストから, 人が潜在的に2キャラクターを描き分けている特徴を抽出できないか確かめる. ”チノ”・”千夜”・”ココア”の組み合わせはVAEの潜在変数を4次元として実験した. 一方, ”リゼ”・”くるみ”の組み合わせは抽出が難しいと思われるので, 潜在変数の次元数を  $4 \cdot 6 \cdot 8$  の3パターン行った.

### 4.2 ゲームキャラクターの画風特徴の抽出

ゲームキャラクターの画風を特徴抽出するにあたって”グランブルーファンタジー(以下グラブル)”と”ぷよぷよ!!クエスト(以下ぷよクエ)”のキャラクターイラストを使用した[11][12]. キャラクターの顔画像イラストはそれぞれ1000枚ずつ集めた. 図5に使用するゲームキャラクターの画像例を示す. ここで”グラブル”と”ぷよクエ”を選んだ理由として, どちらも画風が統一されており, キャラクタごとに絵柄が変わらないためである. また図5を見てもわかるように, ”グラブル”と”ぷよクエ”は線の太さや顔のパーツの大きさ, 色合いが違っており, 2つの作品は画風が異なっていると言える. そのため, 2つの作品のキャラクターの顔画像イラストから特徴を抽出し比較することができると, それぞれの画風の特徴を抽出できると考える. ここでも画風の特徴抽出は難しいだろうと考えたため, 潜在変数の次元数を  $4 \cdot 6 \cdot 8$  の3パターン行った.

## 5. 結果と考察

アニメキャラクターのキャラクターごとの特徴抽出とゲームキャラクターの画風の特徴抽出の結果と考察をそれぞれ以下に示す.

### 5.1 アニメキャラクターの特徴抽出の結果と考察

”ココア”・”チノ”・”千夜”の特徴抽出結果を図6に示す. 赤色の点が”ココア”, 緑色の点が”チノ”, 青色の点が”千夜”となっている. 図6の上から2行の5つあるグラフからわ



図4 使用するアニメキャラクタの画像例

左から”ココア”, ”チノ”, ”千夜”, ”リゼ”, ”くるみ”

画像の出典: 左から4つの画像 「ご注文はうさぎですか??」 ©Koi・芳文社/ご注文は製作委員会ですか??  
右から1つの画像 「がっこうぐらし!」 ©Nitroplus/海法紀光・千葉サドル・芳文社/がっこうぐらし!製作委員会



図5 使用するゲームキャラクタの画像例

左二つが”グラブル”, 右二つが”ぷよクエ”

画像の出典: 左2つの画像 「グランブルーファンタジー」 ©Cygames, Inc.

右2つの画像 「ぷよぷよ!!クエスト」 ©SEGA

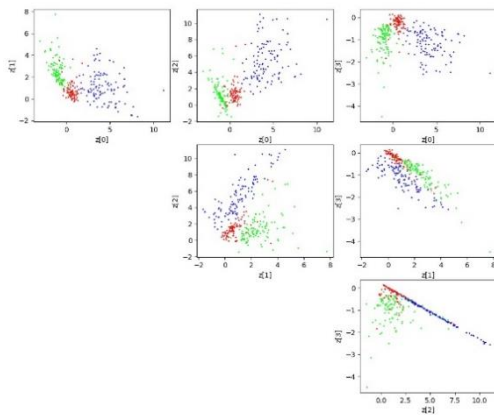


図6 ココア, チノ, 千夜の特徴抽出

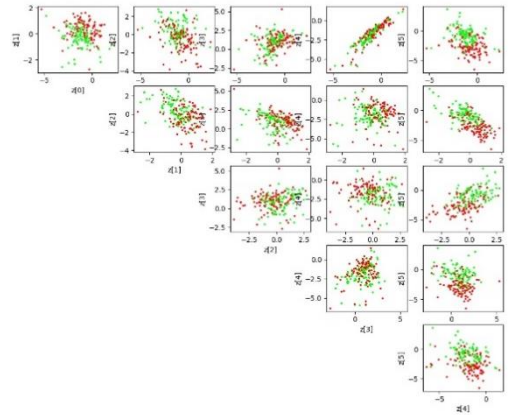


図8 リゼとくるみの特徴抽出(潜在変数: 6次元)

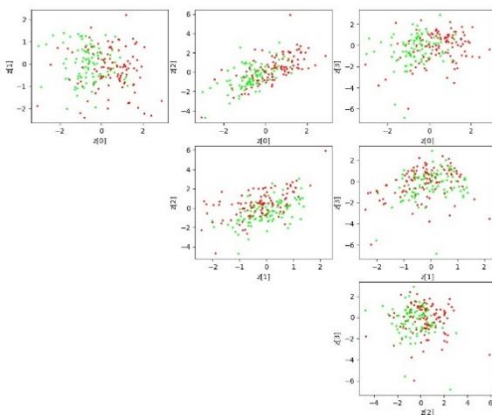


図7 リゼとくるみの特徴抽出(潜在変数: 4次元)

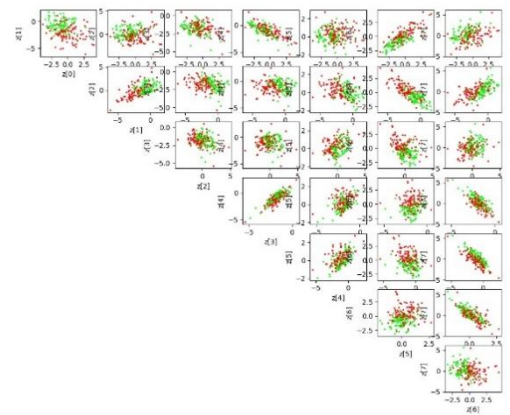


図9 リゼとくるみの特徴抽出(潜在変数: 8次元)



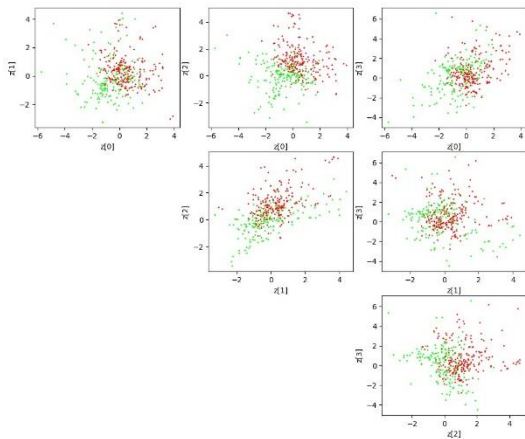


図 10 グラブルとぷよクエの特徴抽出(潜在変数：4次元)

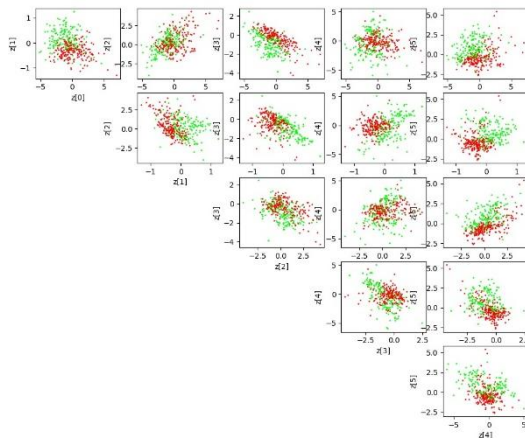


図 11 グラブルとぷよクエの特徴抽出(潜在変数：6次元)

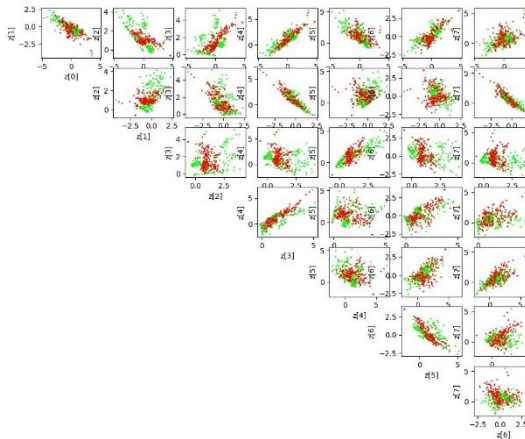


図 12 グラブルとぷよクエの特徴抽出(潜在変数：8次元)

かるように、赤緑青の点の分布に偏りが見られ、それぞれの特徴を抽出することができていると言える。“ココア”・“チノ”・“千夜”はそれぞれ髪の色が全く異なっていたためそれぞれの特徴を抽出することが容易だったのではないかと

考える。ただ、上から3列目のグラフは学習する過程で過学習を起こしてしまったために点群が直線的な形になり、分布の偏りがあまり見られなかったのだと思われる。

“リゼ”・“くるみ”の特徴抽出結果を図 7,8,9 に示す。図 7 の分布はばらつきが多いが、図 8,9 にかけて中央によって見える。しかし、完全に偏りが見られるものはない。“リゼ”・“と”くるみ”はネットで言われているように色合いも画風も似ているので、それぞれの特徴を抽出することは難しかったと考えられる。しかし、潜在変数の分布は完全に混ざり合っていないグラフもあるので、それぞれの潜在変数から重心をとるとある程度の分類精度を持つ分類器ができるのではないかと考える。

## 5.2 ゲームキャラクタの画風特徴の抽出結果と考察

グラブルとぷよクエの特徴抽出結果を図 10,11,12 に示す。図 10 は潜在変数の分布が広がりすぎており、赤と緑の点が重なっている部分も多いため、偏りが見られず特徴抽出できているとは言えない。図 12 もまとまっているものはいくつかあるが、偏りが見られないので特徴抽出はできていない。一方、図 11 の潜在変数が6次元のものは、赤と緑の点が重なっているグラフもあるが、上から1行目、左から3列目のグラフや、上から1行目、左から5列目のグラフなどはそれぞれの色の点がある程度まとまっており、重なっている部分も少ないので偏りが見られると言え、一部の画風特徴を抽出することができていると言える。

このような結果となった理由としては、“グラブル”と“ぷよクエ”それぞれのゲームには多数のキャラクタが存在し、キャラクタの色も多種多様である。そのため、色の割合がそれぞれで似通ったキャラクタが存在すると、潜在変数の分布の位置が近くなってしまったのではないかと考える。他にも、利用した 2D キャラクタの顔画像イラストは画像枚数を多く必要とするために顔の向きが真正面を向いていないキャラクタも存在した。そのため、顔のパーツの位置からわかるような特徴も、イラストごとに位置がばらついてしまったため、特徴抽出ができなかったと考えられる。一方、一部の偏りが見られた理由としては、学習用データとテスト用データの配分が、偶然に求めたい特徴抽出にあった分け方となったためだと思われる。

## 6. おわりに

本研究では、VAE を用いることによって潜在的な特徴の抽出を検討した。結果として、髪色が違ったキャラクタであればそれぞれの特徴を抽出することができた。しかし、髪色が似た場合や、画風の違いを得たい場合は特徴抽出がうまくいかなかった。今後の展望として、今回利用した 2D キャラクタの顔画像イラストは背景がそのまま入ってしまっていたので、背景の情報もそのキャラクタの特徴として反映されてしまった。そのため、前処理に背景除去をして

より良い特徴を検出したいと考える。また、2D キャラクタの顔の向きをできる限り真正面のイラストを使用すると、特徴抽出がよくなると考えるので、できる限り真正面のイラストを集め、物体検出プログラムの学習でも利用されている画像の増幅手法を用いて特徴抽出の向上を目指そうと考える。

今回得られた特徴は support vector machine(SVM)に活用することによって、VAE を用いた特徴抽出が 2D キャラクタを用いた認識分野の基礎的な部分への貢献ができるのかを確かめたいと考えている。

## 参考文献

- [1] 一般社団法人・日本動画協会：“アニメ産業レポート 2018 サマリー（日本語版）” 2018\_Anime\_ind\_rpt\_summary\_jp.pdf, [https://aja.gr.jp/jigyouchousa/sangyo\\_toukei/](https://aja.gr.jp/jigyouchousa/sangyo_toukei/) (2019-11-26)
- [2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: Generative Adversarial Networks, Advances in neural information processing systems, pp.2672-2680, 2014.
- [3] 吉野優, 中田一紀, 小林真, 関田巖, 巽久. 視覚障害者がアニメを楽しむための機械学習によるキャラクター識別システム. 一般社団法人 電子情報通信学会 信学技報, 2019-09
- [4] 中川雄貴, 坂本博康. 色構造のグラフを用いたアニメキャラクター画像間の類似度測定. 火の国情報シンポジウム 2013 プログラム
- [5] 山田 太雅, 棟方渚, 小野哲雄. アニメキャラクターの顔における特徴量抽出手法の提案と実装. 情報処理学会研究報告. Vol.2014-EC-32 No.6, 2014-05-30
- [6] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes., in Proc. 2nd International Conference on Learning Representations, 2014.
- [7] “Google Images Download”. <https://google-images-download.readthedocs.io/en/latest/>, (参照 2019-12-16).
- [8] “lbpcascade\_animeface”. [https://github.com/nagadomi/lbpcascade\\_animeface](https://github.com/nagadomi/lbpcascade_animeface), (参照 2019-12-16).
- [9] A WEB PAGE, TV アニメ「ご注文はうさぎですか?」公式サイト, <https://gochiusa.com/01/> (参照 2019-12-19).
- [10] A WEB PAGE, TV アニメ「がっこうぐらし!」公式サイト, <https://gakkougurashi.com/> (参照 2019-12-19).
- [11] A WEB PAGE, グランブルーファンタジー, <https://granbluefantasy.jp/> (参照 2019-12-19).
- [12] A WEB PAGE, ぷよぷよ!!クエスト(ぷよクエ)公式サイト | ぷよっと楽しいパズル RPG, <https://puyopuyoquest.sega-net.com/> (参照 2019-12-19).