Vol.2020-ICS-197 No.11
2020/2/17

# CAMとGANを用いた 人間とキャラクターの顔画像変換

川名　晴也[†1,a)]　清 雄一[†1,b)]　田原 康之[†1,c)]　大須賀 昭彦[†1,d)]

**概要**：近年様々なメディアの発展により，自身を表すアイコンを使用する機会が多くなってきている．アイコンにアニメキャラクターを使用する者も多いが，既存のイラストを使用することは著作権等の問題があり，自分でオリジナルのイラストを用意するのもハードルが高い．そこで，人間の顔を自動でキャラクター風に変換するシステムがあれば，アイコンの作成をもっと簡易にできると考えられる．画像の変換には，GAN(Generative Adversarial Nets) と呼ばれる手法が一定の成果をあげている．しかしながら，人間とキャラクターの顔では特徴に違いが多いため，顔のパーツや雰囲気を残したまま変換することが難しいという課題がある．そこで CAM(ClassActivationMapping) を用いて特徴を抽出することで，きれいな変換をおこなえるのではないかと考えた．本研究では，CAM を用いることで人間の顔とキャラクターの顔の特徴を抽出し，CycleGAN を用いて人間の顔をキャラクターの画像に変換することを提案する．また，CAM を用いて特徴が強く出ている画像のみを選定することで，顔の特徴を残したまま変換することを目的とした．評価方法としては，変換前と変換後の画像を比較し，どの程度特徴を残し変換できているかをアンケートに回答してもらうことで評価した．従来手法と提案手法を比較した結果，提案手法のほうがより変換前の画像の特徴を残すことができた．

**キーワード**：CAM，GAN，画像変換，キャラクター

## 1. はじめに

### 1.1 背景

近年，様々なメディアの発展により自身を表すアイコンを使用する機会が多くなってきている [11]．しかし自分の写真をアイコンに使う者は少ない．プライバシーを気にする人が多いことや，メディア上では普段の自分を出したくないと考える人が多く存在するためである．そこでアイコンにアニメキャラクターを使用する者も多いが，既存のイラストを使用するのはいくつかの問題がある．ネットワーク上で公開するアイコンは著作権や肖像権の問題が発生してしまうためである．しかしながら，イラストを描いたりモデルを作成できる者は少なく，個人レベルでオリジナルのキャラクターを用意するのはハードルが高い．そこで，人間の顔を自動でキャラクターのように変換するシステムがあれば，アイコンの作成をもっと簡易に行うことが出来ると考えられる．

画像から画像の変換には，GAN(Generative Adversarial Nets) と呼ばれる手法が一定の成果をあげている [3]．GAN には様々な手法があり，画像の分野で多く成果をあげている手法の一つに pix2pix[4] がある．対となる画像を用意し学習させることで画像の変換を行う手法である．しかしながら，変換前後の一対一のペア画像を用意することが，人間の顔とキャラクターのイラストのデータセットでは難しい．それに対しペア画像を必要としない CycleGAN[7] という手法を使用した変換手法が存在する．CycleGAN は，変換したい二種類の属性の画像を用意し，属性 x から属性 y への変換と属性 y から属性 x の変換を交互に行うことで精度を高めていく手法である．

しかしながら，人間の顔とキャラクターのイラストでは特徴に違いが多く，顔のパーツや雰囲気を残したまま変換することが難しいという課題がある．そこで CAM(ClassActivationMapping)[8] と呼ばれるネットワークが画像の特徴を学習する際，どこを特徴として認識しているかをヒートマップのように表示する技術を用いて，特徴を可視化しバイアスをかけることで、変換すべき部位を明確にし自然な画像を生成することを提案する．

†1 現在，電気通信大学 〒 182-8585 東京都調布市調布ヶ丘 1-5-1
Presently with The University of Electro-Communications Chohu, Tokyo 182-8585, Japan
a) kawana.haruya@ohsuga.lab.uec.ac.jp
b) seiuny@uec.ac.jp
c) tahara@uec.ac.jp
d) ohsuga@uec.ac.jp

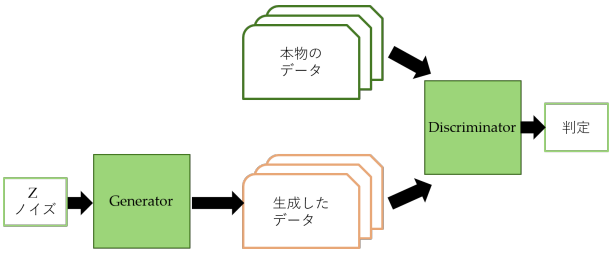


図 1　GAN のアーキテクチャ

### 1.2　研究内容

本研究では，CAM を用いることで人間の顔とキャラクターの顔の特徴を抽出し，CycleGAN を用いて人間の顔をキャラクターの画像に変換することを提案する．また，CAM を用いて特徴がより強く出ている画像のみをあらかじめ選定し学習をおこなうことで，顔の特徴を残したまま変換することを目的とした．また評価方法として，CycleGAN のみを用いて変換した画像と，CAM も用いた変換画像の変換前後の特徴を比較し，どれくらいの精度が出ているかをアンケートで回答してもらうことで，画像生成精度が向上することを示す．

## 2.　関連研究

本章では，本研究に用いた手法について述べる．

### 2.1　GAN

近年，ディープラーニングは様々な問題に対して使用されており，それら多くの問題に対して優秀な成果を収めてきた [1][2]．GAN(Generative Adversarial Networks) とはディープラーニングを応用した学習ネットワークで，日本語では敵対性生成ネットワークと呼ばれるものである．

ここで GAN のアーキテクチャのイメージ図を**図 1** に示す．特徴として，Generator と Discriminator という 2 つのネットワークを用いた低次元ベクトルを入力とする生成モデルであることがあげられる．Generator は，生成データの特徴にランダムノイズを入力することで，このノイズを所望のデータに近づくように学習していく．もう 1 つは Discriminator であり，Generator が生成した偽物のデータと用意された本物のデータの判別を行い，正答率が上がるように学習していく．この学習には minimax 法が用いられている．

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_z(z)[\log(1 - D(G(z)))] \quad (1)$$

minimax 法の式は式 (1) のように表される．右辺前半部分では値が大きくなるように，後半部分では値が小さくなるように数値を調整し，学習していく．このように 2 つのネットワークを交互に競合させながら学習することで，より本物に近い偽物のデータを生成できるようになる．

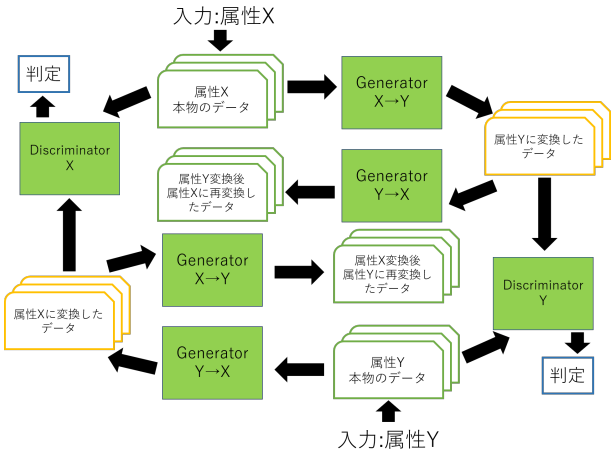


図 2　CycleGAN のアーキテクチャ

近年では，特に画像の分野において優秀な成績を収めている，CNN(Convolutional Neural Network)[5] を用いた学習モデルを応用した，DCGAN(Deep Convolutional GAN)[6] が使われていることが多い．

### 2.2　CycleGAN

GAN の学習による画像変換には様々な手法があるが，その多くで変換元の画像と，対となる変換先の画像のペアが学習用データとして必要となる．そこでペア画像を必要とせず画像変換が可能な手法として CycleGAN が提案された [7]．

ここで CycleGAN のアーキテクチャのイメージ図を**図 2** に示す．CycleGAN では，Generator 及び Discriminator を 2 組用いた学習モデルである．属性 x から属性 y への変換と属性 y から属性 x の変換を交互に行うことで精度を高めていく手法で，正解画像を必要としない．また，一度属性 y に変換したのち属性 x に再変換された画像は、元の属性 x と等しい必要があるため，その評価がなされる．

CycleGAN では 2 つの目的変数，AdversarialLoss と CycleConsistencyLoss を用いて学習を行い次のような数式となる．

$$\mathcal{L}_{full} = \mathcal{L}_{adv}(G_{X\to Y}, D_Y) + \mathcal{L}_{adv}(G_{Y\to X}, D_X) + \lambda_{cyc}\mathcal{L}_{cyc}(G_{X\to Y}, G_{Y\to X}) \quad (2)$$

CycleGAN の学習は式 (2) のように表される．

$$\mathcal{L}_{adv}(G_{X\to Y}, D_Y) = E_{y\sim p_{data}(y)}[\log D_Y(y)] + E_{x\sim p_{data}(x)}[\log(1 - D_Y(G_{X\to Y}(x)))] \quad (3)$$

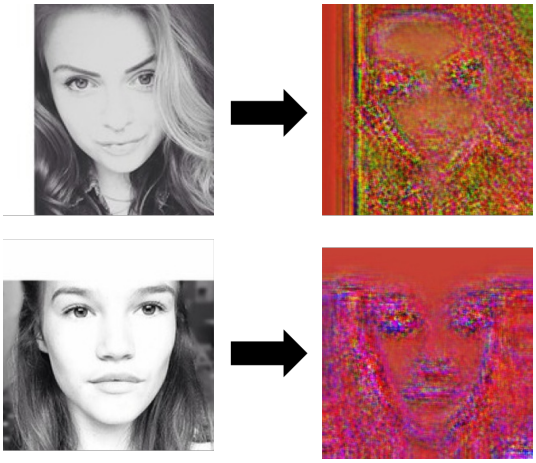

図 3　CAM を用いた画像

$$\mathcal{L}_{adv}(G_{Y\to X}, D_X) = E_{x\sim p_{data}(x)}[\log D_X(x)] + E_{y\sim p_{data}(y)}[\log(1 - D_X(G_{Y\to X}(y)))] \quad (4)$$

$$\mathcal{L}_{cyc} = E_{x\sim p_{data}}[|\, G_{Y\to X}(G_{X\to Y}(x)) - x\,|_1] + E_{y\sim p_{data}(y)}[|\, G_{X\to Y}(G_{Y\to X}(y)) - y\,|_1] \quad (5)$$

AdversarialLoss は式 (3)，式 (4) のように，CycleConsistencyLoss は式 (5) のように表される．

CycleGAN の応用例として，馬とシマウマの相互変換や，緑の葉を紅葉させる変換等が実現されており，いずれの変換でも高い精度を示している．

その反面，特徴の異なる画像同士 (例:犬と猫) の変換では，きれいな画像が生成されず精度が出なくなってしまうという現状がある．

### 2.3　CAM

CAM(ClassActivationMapping) は CNN が画像の分類をするときに，どこを判断基準にしているかを可視化する技術である．Grad-CAM という手法では，予測クラスの loss 値に大きく寄与している部分が，分類を行う上で重要な部分なのではないかといった予想を，画像上にヒートマップのように表示する [8].

これを応用した GuidedGrad-CAM は、Grad-CAM と GuidedBackPropagation という技術を組み合わせた技術である．GuidedBackPropagation は，CNN における最後の convolutional 層への入力の勾配が大きい部分を，分類する属性ごとに計算を行い，際立たせた上で Grad-CAM を行う方法である．

実際に CAM を顔画像に用いた例を**図 3** に表示する．

これらを行うことにより，人間の目からブラックボックスだった学習モデルの判断基準が鮮明にわかるようになる．

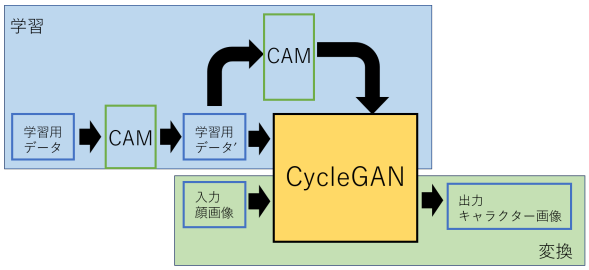


図 4　システムの概要図

## 3.　問題定義

本研究では，人間の顔の画像を入力として用い，出力としてキャラクター風のイラスト画像に変換することを想定する．その際，変換前の画像の特徴や雰囲気を可能な限り残して変換すること，およびキャラクターとして完成度の高い画像を生成することを目的とする．

## 4.　提案手法

### 4.1　概要

本章では，人間の顔画像からキャラクター風の顔画像への変換を，CycleGAN を用いて行う．その際，CAM を用い特徴の強く出ている画像を学習データとして用いることで，特徴を残したまま変換する方法を提案する．さらに，学習の際にも CAM を用いることで，顔の特徴を残したままで精度の高い画像を生成することを目的とする．本研究で提案するシステムの概要図を**図 4** に示す．

### 4.2　ネットワークのアーキテクチャ

CAM を使用したネットワークの，Generator のアーキテクチャのイメージ図を**図 5** に，Discriminator のアーキテクチャのイメージ図を**図 6** に示す．学習前に CAM を用いて入力画像の特徴量を抽出し正規化する．このデータを Generator の内部で掛け合わせることで，特徴の強い部分にバイアスをかけ学習させることを実現した．CAM のデータとサイズを合わせるため，CAM を掛け合わせる部分は Generator では Residual 層内，Discriminator では最終層の手前とした．

### 4.3　CAM による画像の選定

特徴を残した変換を行うため，前処理として学習用のデータの選定を行う．選定の方法は，CAM を用いて画像の特徴の抽出を行った後，顔だと判別できる画像かつ，目や口などの顔のパーツ部分の特徴が強く出ているものを手動で選定した．選定の例として，図 3 の場合，顔のパーツに特徴の集中している下の画像を選んだ．また CAM を用いて選定する画像は，人間の顔の画像のみとした．これは，キャラクターの顔はデフォルメされており，人間の顔ほど

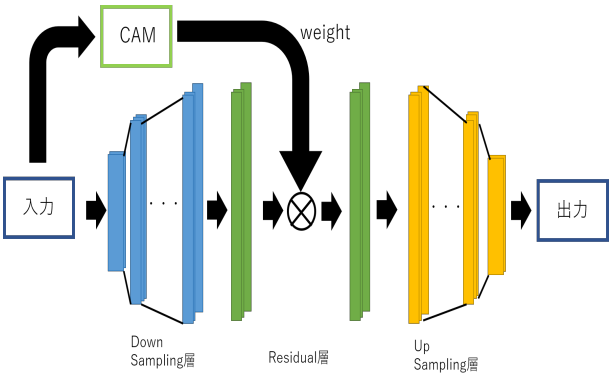


図 5　Generator のアーキテクチャ

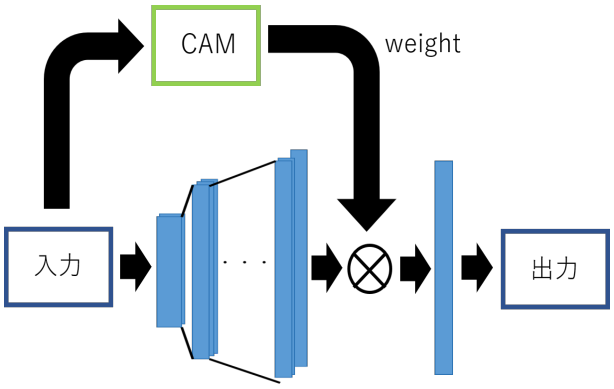


図 6　Discriminator のアーキテクチャ

特徴が多くないと考えられるため，人間の顔の特徴を強調することで特徴同士の相関がうまくできるのでははないかと考えたためである．

## 5. 画像生成実験

本章では，人間の顔画像からキャラクターの顔画像へ変換する，画像生成実験について述べる．

### 5.1 実験内容

提案手法として，CAM を用いて事前に人力で画像選定したものを，CAM と CycleGAN を用いて変換した．また比較対象として，CycleGAN のみを用いて変換する実験も行った．画像選定の基準としては，CAM を用いた結果の画像の目や口の部分に，特徴が強く表示されているものとした．

本実験では学習用データとして 500 枚を，CycleGAN のみの実験ではランダムに，提案手法では CAM を用いて選定した．テストデータは 50 枚をランダムに選定し，学習に用いた．また，学習回数は 100epoch とした．

### 5.2 データセット

本実験のデータセットとして使用した画像を**表 1** に示す．人間の顔画像として，Flickr-Faces-HQ Dataset (FFHQ) というデータを，キャラクターの顔画像として，AnimeFace Character Dataset を使用した [12][13].

使用した画像はすべて png ファイルであり，前処理として画像サイズを 256 × 256 にリサイズしたうえで，学習に用いることとした．

**表 1**　データセット

| データセット名 | サイズ | 詳細 |
|---|---|---|
| Flickr-Faces-HQ (FFHQ) | 1024 × 1024 | 様々な人種，性別の顔画像 データセット |
| AnimeFace Character | 160 × 160 | アニメキャラクターの顔画像 データセット |

### 5.3 ネットワーク構造

本実験で用いた，CycleGAN の Generator ネットワーク構造を**表 2** に，Discriminator のネットワーク構造を**表 3** に示す．

**表 2**　Generator のネットワーク構造

| name | in˙channels | out˙channels | filter | stride | pad |
|---|---|---|---|---|---|
| conv0 | 3 | 64 | 7 | 1 | 3 |
| conv1 | 64 | 128 | 3 | 2 | 1 |
| conv2 | 128 | 256 | 3 | 2 | 1 |
| res0 | 256 | 256 | 3 | 1 | 1 |
| res1 | 256 | 256 | 3 | 1 | 1 |
| res2 | 256 | 256 | 3 | 1 | 1 |
| res3 | 256 | 256 | 3 | 1 | 1 |
| conv3 | 256 | 128 | 3 | 1 | 1 |
| conv4 | 128 | 64 | 3 | 1 | 1 |
| conv5 | 64 | 3 | 7 | 1 | 3 |

**表 3**　Discriminator のネットワーク構造

| name | in˙channels | out˙channels | filter | stride | pad |
|---|---|---|---|---|---|
| conv0 | 3 | 64 | 4 | 2 | 1 |
| conv1 | 64 | 128 | 4 | 2 | 1 |
| conv2 | 128 | 256 | 4 | 2 | 1 |
| conv3 | 256 | 512 | 4 | 1 | 1 |
| conv4 | 256 | 1 | 3 | 1 | 1 |

### 5.4 実装環境

本実験では，CycleGAN の実装を以下の環境で行い，学習を行った．

- Python version : python3.6.0
- tensorflow version : 1.14
- CUDA version : cuda-8.0

### 5.5 実験結果

CycleGAN のみを用いて変換した画像の例を**図 7** に，提案手法を用いて変換した画像の例を**図 8** に示す．

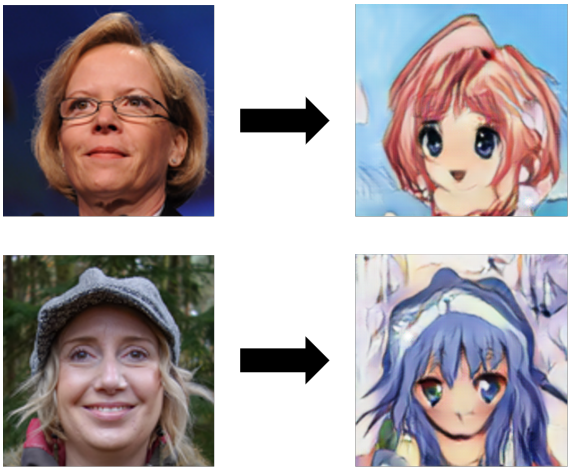


図 7　CycleGAN のみを用いた変換結果

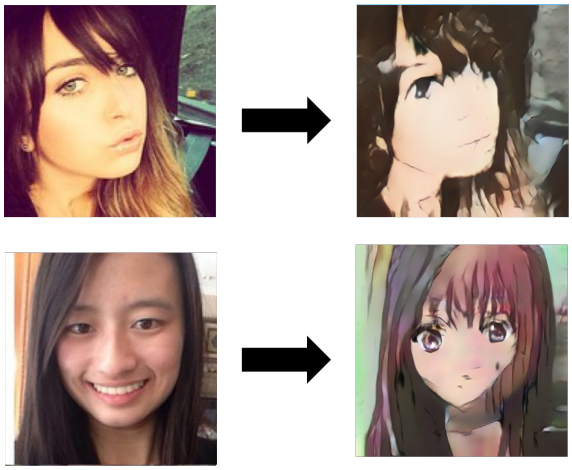


図 8　提案手法での実験結果

## 6. 評価実験

本章では，生成した画像についての評価実験について述べる．生成した画像が，きれいに変換できているかどうかを測るための指標として，画像識別ネットワークを用いた客観評価実験と，画像を実際に見てもらい，受けた印象をアンケート形式で答えてもらう主観評価実験の 2 種類を行った．

### 6.1 客観評価実験

客観評価として，それぞれの手法について生成された画像が，人間の顔画像か，キャラクターの画像かの分類を行った．変換したテスト画像 50 枚のうち，キャラクターと判別された画像の割合を表 4 に示す．

表 4　客観評価実験結果

| 学習方法 | キャラクターと判別された割合 |
|---|---|
| CycleGAN のみ | 0.72 |
| CycleGAN+CAM | 0.76 |

### 6.2 主観評価実験

主観評価として，それぞれの手法で生成された画像について，変換前の画像と変換後の画像の二枚を見比べてもらい，生成画像の評価を行っていた先行研究をもとに [9][10]，次の 3 つの観点で評価を行ってもらった．

(1) 顔として自然な画像であるか
(2) 変換前と髪型や顔のパーツが一致しているか
(3) 変換前の雰囲気を残しているか

なおアンケートは，グーグルフォームを用いたインターネット媒体で行い，20〜56 歳の男女 28 名に回答してもらった．内容として，2 手法それぞれの変換画像 5 枚ずつ計 10 枚について，どちらの手法で変換したかを伏せたうえで，5 段階 (1 が最も評価が低く，5 が最も評価が高い) で評価してもらった．実験結果を**図 9** に示す．

なお結果の数値は，5 段階の評価をそのまま得点とした場合の，各画像の平均獲得得点である．

## 7. 考察

客観評価実験では，若干ながら数値が上昇したものの，従来手法との大きな差は見られなかった．これは生成された画像が，そもそも人間とキャラクターの中間のようなものであるため，単純に二値分類としての精度がでなかったためだと推測した．

一方，主観評価アンケートでは結果に差がみられた．変換前と髪型や顔のパーツが一致しているか，変換前の雰囲気を残しているかの項目では，提案手法のほうが高い評価を得た．このことから，目標としていた変換前の特徴を残した画像の生成は達成できたといえる．

しかしながら，顔として自然な画像かの項目では従来手法のほうが高い評価を得ている．これは，事前の画像選定の際に，目や口に特徴が強く出ている画像のみを残したため，顔の輪郭部分やそのほかのパーツなどの学習がおろそかになってしまい，全体的な顔の完成度が落ちてしまったのではないかと考えられる．

## 8. おわりに

### 8.1 本論文のまとめ

本論文では，人間の顔をキャラクターの顔に変換することを想定し，変換前の画像の特徴を残したまま変換することを目的とした．変換手法として，CAM を用いることで特徴を明確にしたうえで，CycleGAN を用いて画像を生成することを提案した．また，生成した画像を評価するため，

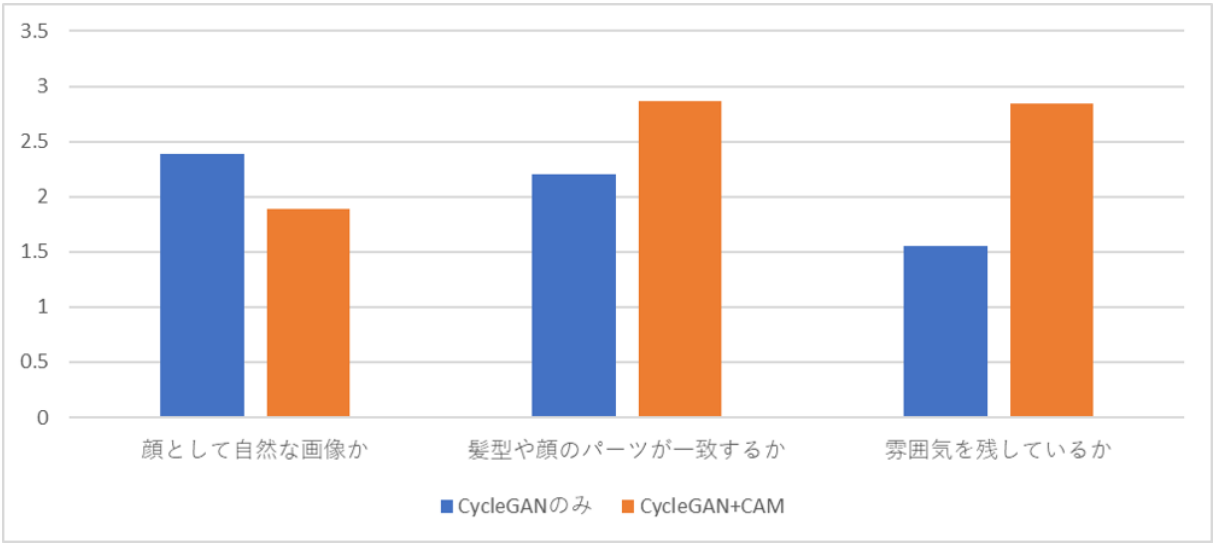


**図 9** アンケート結果

画像分類ネットワークを用いた客観評価と，アンケートによる主観評価の二種類を用いた．結果として，提案手法のほうが，より変換前の特徴を残すことに成功した．しかしながら，顔としての完成度が低くなってしまうという結果となった．

### 8.2 今後の課題

今後の課題として，画像枚数を増やしての実験があげられる．時間の制約上大規模な学習が行えなかったため，少数データから精度を上げる方法を検討したが，単純に学習枚数を増やすだけで，顔としての完成度が高くなることが期待できる．CAM で選別した学習用画像枚数を増加させることで，さらに精度が上がることが期待されるため，実験を行う必要があると考えられる．また，評価方法も増やす必要があると考えられる．今回客観評価方法が，画像がキャラクターに変換できているかという判別手法だけだったため，変換前と変換後の特徴を比較できる客観評価方法を検討する必要があると思われる．

**謝辞**

本研究は JSPS 科研費 JP17H04705，JP18H03229，JP18H03340，JP18K19835，JP19H04113，JP19K12107 の助成を受けたものです．

### 参考文献

[1] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, : Going Deeper with Convolutions, Proceeding of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, : Generative Adversarial Networks, 10 Jun 2014.

[3] Pengyuan Lyu, Xiang Bai, Cong Yao, Zhen Zhu, Tengteng Huang, Wenyu Liu, : Auto-Encoder Guided GAN for Chinese Calligraphy Synthesis, IAPR International Conference on Document Analysis and Recognition.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, arXiv:1611.07004v1 [cs.CV] 21 Nov 2016.

[5] LeCun, Y. , Boser, B. , Denker, J. S. , Henderson, D. , Howard, R. E. , Hubbard, W. , Jackel, L. D. : Backpropagation applied to handwritten zip code recognation, Neural computation, 1(4), pp.541-551, 1989

[6] Alec Radford, Luke Metz, Soumith Chintala, : Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434v2 [cs.LG] 7 Jan 2016.

[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, : Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 30 Mar 2017.

[8] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, : Grad-CAM:Visual Explanations from Deep Networks via Gradient-based Localization arXiv:1610.02391v4 [cs.CV] 3 Dec 2019.

[9] 佐川 友里香, 萩原 将文, : 属性を付与した DCGAN による顔画像生成システム, 日本感性工学会論文誌 Vol.17 No.3 pp.337-345 2018.

[10] 宮本 龍, 河合 紀彦, 山澤 一誠, 佐藤 智和, 横矢 直和, テクスチャの幾何学的変換と類似パターン位置を考慮したエネルギー最小化による画像修復. : 画像の認識・理解シンポジウム 2011.

[11] 総務省：ソーシャルメディアの利用状況, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd142210.html（参照 2020-01-13）.

[12] GitHub : NVlabs/ffhq-dataset, https://github.com/NVlabs/ffhq-dataset（参照 2019-06-29）.

[13] animeface-character-dataset, www.nurs.or.jp/~nagadomi/animeface-character-dataset/data/animeface-character-dataset.zip（参照 2019-06-29）.