

# Anime-style Image Generation using GAN

Zhen Cui, Yasuaki Ito, Koji Nakano

Graduate School of Advanced Science and Engineering, Hiroshima University  
Kagamiyama 1-4-1, Higashi-hiroshima, 739-8527, Japan

Akihiko Kasagi

Fujitsu Ltd.

Kamikodanaka 4-1-1, Nakahara-ku, Kawasaki, Kanagawa 211-8588, Japan

**Abstract**—With the popularity of social networking services and smartphones, photo processing applications have become widely used, and there is a growing interest in more advanced photo processing techniques. One of these processing techniques is to convert photos to other styles. Many studies have been researched to automatically generate images using machine learning to deal with this problem in recent years. In the field of image generation, methods based on adversarial generative networks (GANs) have shown particularly good results. There are many studies of style transformation from photo to anime-style using the technique. However, most of these studies are limited to style transformation of the face part and background. The goal of this study is to convert a full-body photograph of a person into an image with style similar to that of an animated character. We prepared a dataset of full-body human photos and a dataset of full-body animated characters and used the unsupervised model CycleGAN to train the model. In addition, latent variables were added to obtain the diversity of the generated images. As a result, images with different animation styles were obtained with different latent variables.

**Index Terms**—machine learning, style transfer, GANs, anime-style

## I. はじめに

アニメや漫画内に登場する風景や人物は、しばしば実物を参考にして描かれることがある。この時、作者は対象物の外見や細かな特徴を捉えながら、描いている作品の雰囲気合った画像に変換している。通常この変換を可能にするには、知識や経験、画力などといった多くの土台となる技術が要求される。そのため、経験の浅いものにとってこの変換は非常に困難な作業となる。

近年、素人でも玄人と同様な完成度の画像を生成できる技術として、画像変換の分野における機械学習の技術に注目が集まっている。VAE [1]は入力画像をエンコードし、正規分布の特徴マップを得たあと、特徴マップをデコードし、生成した画像を元の画像に可能な限り一致させるように学習する。この特徴マップを別の正規分布の特徴マップに置き換えることで、同じ特徴を持つ別の画像として得ることができる。しかし、特徴マップの値がどのような特徴を表しているのか判別することができないため、特徴マップの値から任意の出力を与えることは困難である。Gatysらの研究 [2]では入力画像とスタイル画像の特徴を抽出し、それらを生成された画像と比較することで、入力画像の内容を保ちつつスタイルの変換をおこなった。

敵対的生成ネットワーク (GAN) [3]は2014年に提案されたネットワークで、近年の画像生成分野で大きな貢献を残

している技術である。GANは生成器と鑑別器が敵対的に学習することで、生成器が、ネットワークに教師として与えたデータに近いデータを生成することができる。2015年に提案されたDCGAN [4]は、先程のGANに畳み込みニューラルネットワーク (CNN) 追加することで、画像の変換により効果的な学習を可能とした。Karrasらの研究 [5]では訓練中にネットワークに畳み込み層をレイヤーごとに追加することで、生成画像を本物の画像のように生成することを示した。さらに画像から別の画像に変換する image-to-image 分野では、DCGANをベースにした pix2pix [6]が大きな成果を上げた。このモデルは生成器と鑑別器の損失を計算することで、教師あり学習を用いてスタイル変換を実現した。

しかし、実写画像からゴッホの作風を持った画像への変換や、顔のアニメ風変換など、変換前後のペアになるデータセットを持たない画像に対するスタイル変換などのタスクに対して、教師あり学習ができない場合も多い。そこで、ペアではないデータセットでも画像変換を学習できる教師なしのGANのモデルが提案された。StyleGAN [7]は、非線形マッピングネットワークを追加することで、任意の顔画像を別の特徴を持った顔画像への変換をおこなった。Pinkneyらの研究 [8]ではStyleGANをベースに、2つのスタイルの画像生成ネットワークを組み合わせて、人物の顔画像からその浮世絵スタイルの画像を生成した。CycleGAN [9]は二つの生成器と二つの鑑別器を利用することで、対象物の輪郭や細かな特徴を保持したまま、スタイルのみ変換することで質の高い画像を生成可能であることを示した。

CycleGANをベースとした研究も盛んに行われており、Wangらの研究 [10]ではスタイル変換のタスクを色、エッジ、構造の三つに分けて、それぞれを学習することで、風景写真からアニメ風画像への変換をおこなった。Landmark assisted CycleGAN [11]はCycleGANに基づいて、目、鼻、口の位置を示すkeypointを補助情報として追加することで、顔のスタイルの変換効果を向上させることに成功した。U-GAT-IT [12]では、学習する際のヒートマップを補助情報として加えることで、モデルが入力画像の注目領域をより判別しやすくなり、パーツや領域毎の重点的な画像変換を可能にした。また、Augmented CycleGAN [13]は、入力に潜在変数を付加し、その値を変化させることで、多様なスタイルをもつ画像を生成できることを示した。

これらの研究の大半は、風景または顔のスタイル変換に注目している。その一方、人物の全身のスタイル変換を対象とする研究はほとんど行われていない。そこで本研究は人物の全身画像をアニメスタイルの画像に変換する手法を

提案する。

人物写真からアニメ風に変換するタスクを pix2pix の様なモデルで学習させるためには、教師データとして、写真とそのアニメ画像のペアから成るデータセットが必要となる。しかし、実写画像とその被写体と同じポーズや髪型などを持っているアニメ画像のペアを大量に収集するの困難である。そのため、pix2pix のような教師データを必要とする手法は本研究にはあまり適していない。そこで、本研究はペアではないデータセットでも学習できる教師なしモデル CycleGAN をベースモデルとして採用した。しかし、CycleGAN は近い物体を同じような特徴に変換するため、局所的な特徴を多く持つ画像の変換にはしばしば向かない傾向にある。例えば馬からシマウマに変換するタスクでは、馬に人が乗っているとき、その人もシマウマ柄に変換してしまうことがある。そのため、本研究のデータセットでは、生成画像の品質を向上するために、人物とキャラクター以外の背景をすべて白にしている。さらに、実写人物とアニメキャラクターを全身が写る画像をデータセットとして収集し実験を行った。これは、人物が座っていたり、体の一部が隠されたりする画像をデータセットに含んだ状態で訓練したモデルでは、効果的な画像変換ができないためである。また、ネットワークを学習する際に、データセットの数が少ないため、訓練中にランダム cropping を用いて学習データを増やすことで、品質が向上させた。さらに、CycleGAN は一つの画像に対して一つの画像しか生成できない制限がある。そこで異なる服の種類や髪型をもつ画像を得るために、Augmented CycleGAN を利用した。結果として、潜在変数を変えることにより、一つの画像から Fig. 1 に示すような多様な生成結果を得られるモデルを作成できた。



Fig. 1. 提案手法によるアニメ画像の生成結果

本論文の残りの部分は以下のように構成されています。第 II 節では提案手法を紹介し、第 III 節では実験結果を示す。最後に、第 IV 節では本論文のまとめを行う。

## II. 提案手法

本節では、人物画像からアニメ風画像への変換を行う手法を提案する。教師なしモデル CycleGAN [9] は画像内の風景や人物をラベル付けし、区別させて学習することが容易でない。画像内に物体が多く存在する場合、主となる体以外の情報は画像変換の障害となることがある。そのため、さまざまなオブジェクトが含まれる画像を学習データとしてモデルを学習するのは難しい。そこで本研究は人物写真からアニメ風画像への変換をするために、新たなデータセットを作成し、CycleGAN を用いて学習を行う。

また、CycleGAN のモデルのみでは一つの画像に対してそれに対応する一つの画像しか生成できない。そのため、生

成した画像がユーザーの好みと一致しない場合がある。この問題を解決するために、本研究では、潜在変数を追加し、多様な結果を出す Augmented CycleGAN [13] を導入する。Augmented CycleGAN を使用することで、同じ入力画像に対して異なるスタイルを持つ複数の結果が得られるようになる。

### A. データセットの作成

先述の通り、CycleGAN は画像内の情報が多い場合は変換は難しいため、複雑な背景が含まれる画像を変換できないことが多い。本研究の目的は人物だけの変換であるため、Fig. 1 のように人物以外を削除したデータセットを用意する。また画像に複数の人物を写っていたり、人物の上半身だけ写っていたりするデータが混ざっているデータセットで CycleGAN を訓練する結果は良くないため、一人の全身が含まれる画像を学習データとして使用する。

### B. CycleGAN

CycleGAN は教師なしの画像変換によく使われるモデルである。そのモデルは Fig. 2 のように、二つの生成器と鑑別器で構成され、ドメイン  $X$  からドメイン  $Y$  への変換とドメイン  $Y$  からドメイン  $X$  への変換を同時に訓練する。ドメイン  $X$  の画像  $x$  はドメイン  $Y$  へ変換され、さらにドメイン  $X$  に逆変換される。この時、変換前後の画像をなるべく一致させることで、画像の内容を保つことができる。また、生成画像をドメイン  $Y$  の鑑別器で本物画像と判断されるように学習することで、よりドメイン  $Y$  に近い画像を生成することができる。

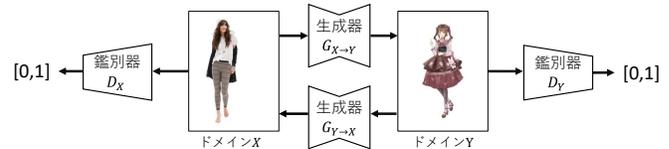


Fig. 2. CycleGAN の構造

CycleGAN は、訓練データ  $x \in X$ ,  $y \in Y$  が与えられたときに、ドメイン  $X$  と  $Y$  の間の特徴の差を学習する。生成器  $G_{X \rightarrow Y}$  と鑑別器  $D_Y$  に対して、敵対的損失は以下のように定義される。

$$L_{GAN}(G_{X \rightarrow Y}, D_Y) = \mathbb{E}_y[\log D_Y(y)] + \mathbb{E}_x[\log(1 - D_Y(G_{X \rightarrow Y}(x)))] \quad (1)$$

また、サイクル一貫性損失は以下のように定義されている。

$$L_{cyc} = \|(G_{Y \rightarrow X}(G_{X \rightarrow Y}(x)) - x)\|_1 + \|(G_{X \rightarrow X}(G_{Y \rightarrow X}(y)) - y)\|_1 \quad (2)$$

CycleGAN 全体の損失関数は以下のように定義されている。 $\lambda_1$  と  $\lambda_2$  は敵対的損失とサイクル一貫性損失のバランスを取るためのハイパーパラメータである。

$$L(G_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) = \lambda_1 L_{GAN}(G_{X \rightarrow Y}, D_Y) + \lambda_1 L_{GAN}(G_{Y \rightarrow X}, D_X) + \lambda_2 L_{cyc}(G_{X \rightarrow Y}, G_{Y \rightarrow X}) \quad (3)$$

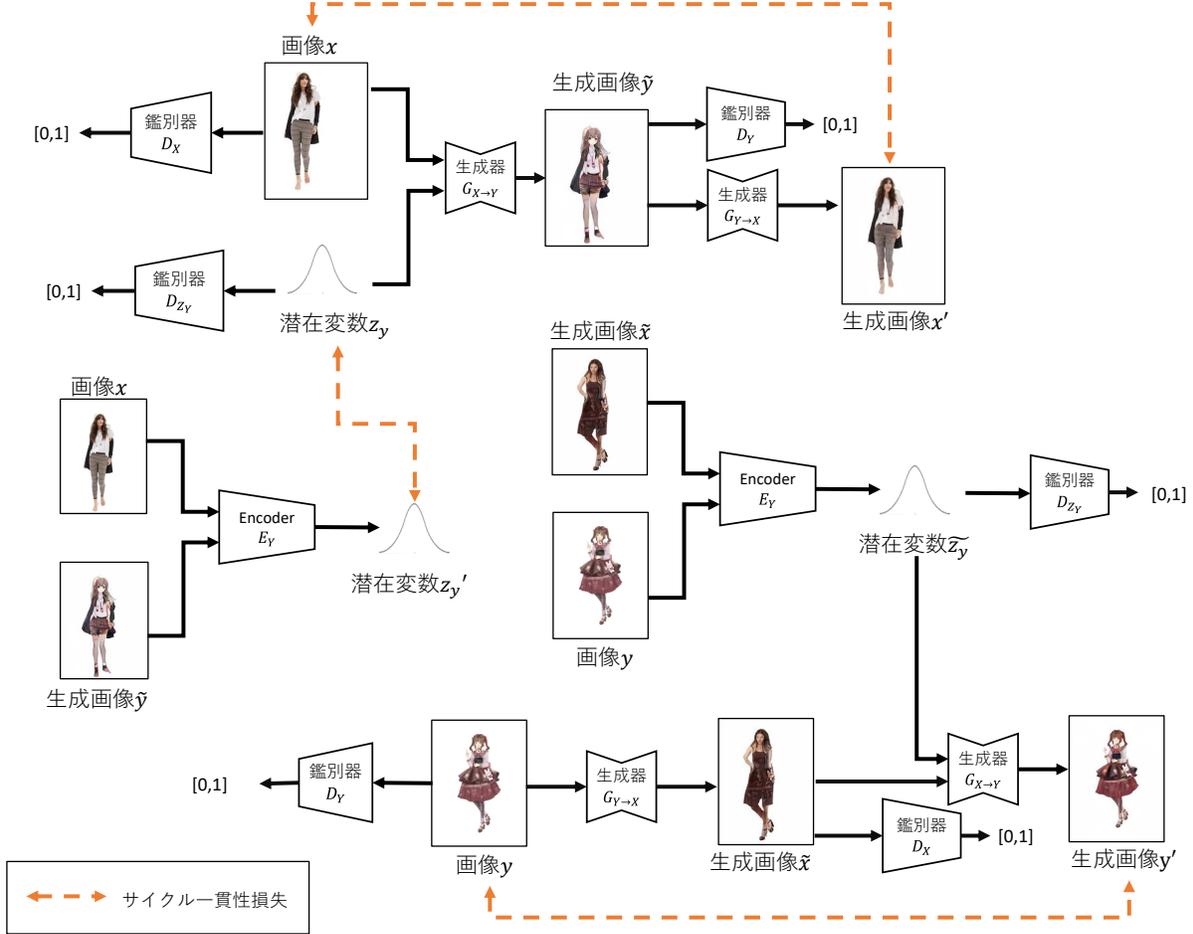


Fig. 3. Augmented CycleGAN モデル構造

### C. Augmented CycleGAN を用いた出力結果の多様化

CycleGAN では一つの人物写真から一つのアニメ画像しか生成できないため、ユーザーの好みの画像が出力されるとは限らない。そのため、複数の画像を生成できる Augmented CycleGAN [13] を導入する。提案手法で使用する Augmented CycleGAN のモデルの概要を Fig. 3 に示す。Augmented CycleGAN では CycleGAN のモデルに出力の多様性を与えるパラメータとして潜在変数  $G_{X \rightarrow Y}$  を入力として加える。さらに、潜在変数を生成するための Encoder ネットワーク  $E_Y$  とその鑑別器  $D_{Z_Y}$  を使用する。

前述した CycleGAN は一対一のマッピングに対して高い効果を発揮するが、その一方、一対多のマッピングに対する学習ができないという欠点を持つ。Augmented CycleGAN は一対多のマッピングを目的として提案されたモデルである。具体的には、各ドメインに補助的なパラメータとして潜在変数を追加することで、特徴空間を拡張してモデルを学習する。以下では Fig. 3 に示す Augmented CycleGAN のモデルにおける学習の詳細を示す。

入力画像  $x$  は、補助潜在変数  $z_y$  といっしょに CycleGAN の生成器  $G_{X \rightarrow Y}$  に入力され、生成された画像は鑑別器  $D_Y$

で判別される。敵対的損失関数は以下のように定義される。

$$L_{GAN}^B(G_{X \rightarrow Y}, D_Y) = \mathbb{E}_y[\log D_Y(y)] + \mathbb{E}_x[\log(1 - D_Y(G_{X \rightarrow Y}(x, z_y)))] \quad (4)$$

単に潜在変数を入力に加えるだけでは、潜在変数を無視して画像を生成するようにモデルを学習してしまう。それを避けるため、Augmented CycleGAN では、2つの Encoder ネットワーク  $E_X$  と  $E_Y$  と潜在変数の鑑別器  $D_{Z_X}$  と  $D_{Z_Y}$  を導入する。 $E_X$  と  $E_Y$  は元画像と生成画像を同時に入力し、潜在変数  $\tilde{z}_x$  と  $\tilde{z}_y$  を次の式を用いて生成する。

$$\begin{aligned} \tilde{z}_x &= E_X(x, G_{X \rightarrow Y}(x)) \\ \tilde{z}_y &= E_Y(y, G_{Y \rightarrow X}(y)) \end{aligned} \quad (5)$$

潜在変数  $\tilde{z}_x$  と与えた潜在変数  $z_x$  を一致させるように、潜在変数の敵対的損失は以下のように定義される。

$$L_{GAN}^{Z_X}(E_X, G_{X \rightarrow Y}, D_{Z_X}) = \mathbb{E}_{z_x}[\log D_{Z_X}(z_x)] + \mathbb{E}_{y}[\log(1 - D_{Z_X}(\tilde{z}_x))] \quad (6)$$

また、画像のサイクル一貫性損失は以下のように定義されている。

$$L_{cyc}^A(G_{X \rightarrow Y}, G_{Y \rightarrow X}, E_X) = \|x' - x\|_1, \quad (7)$$

$$\tilde{y} = G_{X \rightarrow Y}(x, z_y), \tilde{z}_x = E_X(x, \tilde{y}), x' = G_{Y \rightarrow X}(\tilde{y}, \tilde{z}_x)$$

生成された画像  $\tilde{y}$  を与えられた潜在変数  $z_y$  の依存性を強制するために、潜在変数のサイクル一貫性の損失は以下のように定義されている。

$$L_{cyc}^{Z_Y}(G_{X \rightarrow Y}, E_Y) = \|z'_y - z_y\|_1, \quad (8)$$

$$z'_y = E_Y(x, \tilde{y}), \tilde{y} = G_{X \rightarrow Y}(x, z_y)$$

ドメイン  $X$  からドメイン  $Y$  への損失関数は以下のように定義されている。

$$L(G_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y, E_X, E_Y) =$$

$$L_{GAN}^Y(G_{X \rightarrow Y}, D_Y) + L_{GAN}^{Z_X}(E_X, G_{X \rightarrow Y}, D_{Z_X})$$

$$+ \lambda_1 L_{cyc}^A(G_{X \rightarrow Y}, G_{Y \rightarrow X}, E_X) + \lambda_2 L_{cyc}^{Z_Y}(G_{X \rightarrow Y}, E_Y) \quad (9)$$

以上より、生成器が潜在変数に依存した画像を生成するようになる。

正規分布に従う潜在変数をいくつ追加し、一つの人物写真から複数のアニメ画像を生成することを目指す。

### III. 実験

実験では CycleGAN と Augmented CycleGAN を使用してアニメ風スタイル画像の生成を行った。各モデルの構造の詳細を Table I–VII に示す。

TABLE I  
CycleGAN の生成器  $G_{X \rightarrow Y}$  と生成器  $G_{Y \rightarrow X}$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	256,192,64	(7,1)
ConvLayer2	128,96,128	(3,2)
ConvLayer3	64,48,256	(3,2)
ResBlock1	64,48,256	(3,1)
ResBlock2	64,48,256	(3,1)
ResBlock3	64,48,256	(3,1)
ResBlock4	64,48,256	(3,1)
ResBlock5	64,48,256	(3,1)
ResBlock6	64,48,256	(3,1)
ResBlock7	64,48,256	(3,1)
ResBlock8	64,48,256	(3,1)
ResBlock9	64,48,256	(3,1)
DeConvLayer1	128,96,128	(3,2)
DeConvLayer2	256,192,64	(3,2)
ConvLayer4	256,192,3	(7,1)

TABLE II  
CycleGAN の鑑別器  $D_X$  と鑑別器  $D_Y$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	128,96,64	(4,2)
ConvLayer2	64,48,128	(4,2)
ConvLayer3	32,24,256	(4,2)
ConvLayer4	31,23,512	(4,1)
ConvLayer5	30,22,1	(4,1)

TABLE III  
Augmented CycleGAN の生成器  $G_{X \rightarrow Y}$  と生成器  $G_{Y \rightarrow X}$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	256,192,32	(7,1)
ConvLayer2	256,192,64	(3,1)
ConvLayer3	128,96,128	(3,2)
ResBlock1	128,96,128	(3,1)
ResBlock2	128,96,128	(3,1)
ResBlock3	128,96,128	(3,1)
DeConvLayer1	256,192,64	(3,2)
DeConvLayer2	256,192,32	(3,1)
ConvLayer4	256,192,3	(7,1)

TABLE IV  
Augmented CycleGAN の鑑別器  $D_X$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	128,96,32	(3,2)
ConvLayer2	64,48,64	(3,2)
ConvLayer3	32,24,128	(3,2)
ConvLayer4	16,12,128	(3,2)
ConvLayer5	13,9,1	(4,1)

TABLE V  
Augmented CycleGAN の鑑別器  $D_Y$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	128,96,64	(4,2)
ConvLayer2	64,48,128	(4,2)
ConvLayer3	63,47,256	(4,1)
ConvLayer4	62,46,256	(4,1)
ConvLayer5	61,45,1	(4,1)

TABLE VI  
Augmented CycleGAN の Encoder ネットワーク  $E_Y$  の構造

layer	output size	kernel, stride
Inputs	256,192,3	(-, -)
ConvLayer1	128,96,32	(3,2)
ConvLayer2	64,48,64	(3,2)
ConvLayer3	32,24,128	(3,2)
ConvLayer4	16,12,128	(3,2)
ConvLayer5	13,9,256	(4,1)
AvgPool1	1,1,256	(-, -)
ConvLayer-mu	1,1,16	(1,1)
ConvLayer-logvar	1,1,16	(1,1)

TABLE VII  
潜在変数の鑑別器  $D_{Z_Y}$  の構造

layer	output size
Inputs	1,1,16
Linear1	1,1,64
Linear2	1,1,64
Linear3	1,1,64
Linear4	1,1,1

モデルを学習するデータセットについて、教師なしモデル CycleGAN は変換対象のオブジェクト以外の情報を多く持つ画像の変換には、変換効果が良くない。そこでインター

ネットから収集した人物画像とアニメ画像で事前に実験を行った。データセットには、複数の人物が写っていたり、背景が複雑であったり、文字があったりする画像が含まれており、画像の変換がしばしばうまくいかないことが多く見られた。本研究は人物の全身を違和感なく変換することを目的としているため、データセット内の画像から、可能な限り人物以外の情報の削除をおこなった。具体的には、人物以外が写っている箇所を白色に塗りつぶした画像をデータセットの画像として使用した。

本研究で使われたデータセットは実写人物画像とアニメキャラクター画像から構成されている。

実写人物のデータセット：実写人物データセットは二つのデータセット、Fashion データセット [14] と COCO データセット [15] を使用した。次にそれぞれのデータセットの詳細について説明する。

Fashion データセットは DeepFashion Database [14] にある服装予測用のデータセットであり、10 万枚以上のデータが含まれている。本研究では、写真内の被写体のみが必要なため、DeepFashion Database で適切な画像を選び、人物検出器 PaddlePaddle [16] で人物の部分を検出し、背景を白にする。これにより、訓練データ 10,000 枚、テストデータ 252 枚の実写人物白背景の一人全身画像のデータセットを作成した。Fig. 4 はデータセットの一部を示している。



Fig. 4. Fashion データセットの一部

COCO データセットは COCO [15] にある 80 class の合計 33 万枚の画像が含まれているデータセットである。その中に人物の object segmentation や keypoint の情報が含まれている。それを用いて、白背景一人に全身画像のデータセット 8475 枚を取得した。Fig. 5 はデータセットの一部を示している。



Fig. 5. COCO データセットの一部

アニメキャラクターのデータセット：インターネットからアニメキャラクターの立ち絵を収集した。そして、手作業や segmentation モデル [17] でアニメキャラクター白背景の一人全身画像のデータセットを作成した。1578 枚を取得し、そのうち、1554 枚を訓練データ、24 枚をテストデータに設定した。Fig. 6 はデータセットの一部を示している。



Fig. 6. アニメキャラクターのデータセットの一部

COCO データセットとアニメキャラクターのデータセットで学習した結果と、Fashion データセットとアニメキャラクターのデータセットで学習した結果を Fig. 7 に示す。COCO データセットには人物が座っていたり、体の一部が隠されたり画像が多い。一方で、Fashion データセットのデータはプロの撮影写真データセットであり、人物の姿勢は正面で立つものが多いため、COCO データセットより Fashion データセットで学習したモデルの生成結果が向上している。本研究の残る実験は全部 Fashion データセットとアニメキャラクターのデータセットで行う。



Fig. 7. COCO データセットと Fashion データセットの実験結果

CycleGAN で訓練するとき、画像をランダム cropping してモデルに入力することで、データセットを豊富にすることができる。また、頭の位置は cropping することで変わるため、モデルの柔軟性が向上する。画像の端に人の部分が含まれている場合には、cropping した画像の人物は不完全の恐れがあるため、本研究は画像内の人物の部分小さくする。ここでは cropping しなかったデータセットと  $286 \times 215$  に resize し、訓練中に  $256 \times 192$  にランダム cropping したデータセットの二つの実験結果を比較する。結果を Fig. 8 に示す。cropping しないの訓練結果は、顔を幾つの固定パターンに崩壊することがある。訓練中にランダム cropping する方が効果が良いことが分かった。Augmented CycleGAN の実験は訓練中にランダム cropping する方法で行う。

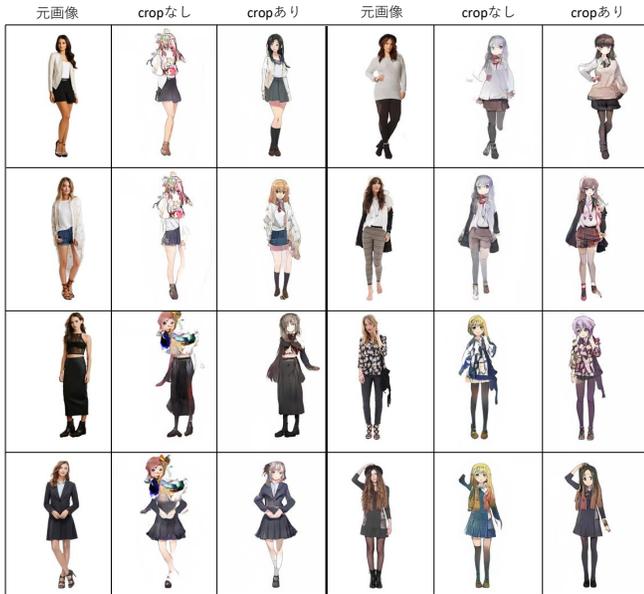


Fig. 8. データ cropping の有無の違い

#### A. 潜在変数の実験

Fashion データセットとアニメキャラクターのデータセットで Augmented CycleGAN を訓練した。入力の潜在変数の数を 1,4,8,16 に変換させ実験を行った。潜在変数の長さを 16 にしたときの交換結果を Fig. 9 に示す。同じ入力画像として、異なる潜在変数を与えることで、色だけでなく、スタイルも変換されていることがわかる。また、異なる画像に同じ潜在変数を追加すると、生成された画像の色とスタイルが類似した画像として変換されていることがわかる。この手法を用いることで、一枚の人物写真で多様なアニメ風画像を得ることができる。

#### IV. まとめ

本研究は実写人物の全身画像をアニメ風に変換することを目的と実験を行った。背景が複雑であったり、複数の人物や上半身と全身が混ざるデータセットには、CycleGAN は人物の変換は難しいと考え、本研究は白背景一人のデータセットを収集した。そのデータセットを CycleGAN で実験し、データセットの人物は単純すればするほど、生成結果が向上していることが検証した。また、訓練中のデータのランダム cropping はデータを豊富することで、訓練中にランダム cropping する学習が良い画像を生成することが分かった。

さらに、CycleGAN は一つの人物画像から一つのアニメ画像しか生成できないため、Augmented CycleGAN を導入し、生成結果がモデルに入力する潜在変数に大きく依存する。そこで、異なる潜在変数を入力の際に、より多様な生成結果が得ることができた。

現時点の Augmented CycleGAN のネットワークは特徴を捉え難くと考え、今後は、Augmented CycleGAN のネットワーク構造を改良し、より良い結果を得ることを目指している。また、入力の潜在変数はランダムで生成するものではなく、ユーザーが自らコントロールできるように改善する。その上で、人物写真から多様なアニメ画像を生成可能なア

プリを開発し、よりユーザーが楽しめるものを作るようにする。更に CycleGAN は人物の関節情報などの位置情報がないため、人物を崩れて生成する恐れがある。そのため、今後は人物の姿勢の keypoint を補助情報として CycleGAN をより良く人物写真からアニメ画像への変換に適していると考える。

#### REFERENCES

- [1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [2] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," *Advances in neural information processing systems*, vol. 28, pp. 262–270, 2015.
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [5] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.
- [6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2017.
- [7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
- [8] J. N. Pinkney and D. Adler, "Resolution dependent gan interpolation for controllable image synthesis between domains," *arXiv preprint arXiv:2010.05334*, 2020.
- [9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017.
- [10] X. Wang and J. Yu, "Learning to cartoonize using white-box cartoon representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8090–8099.
- [11] R. Wu, X. Gu, X. Tao, X. Shen, Y.-W. Tai *et al.*, "Landmark assisted cyclegan for cartoon face generation," *arXiv preprint arXiv:1907.01424*, 2019.
- [12] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *arXiv preprint arXiv:1907.10830*, 2019.
- [13] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," in *International Conference on Machine Learning*. PMLR, 2018, pp. 195–204.
- [14] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [15] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [16] Baidu China Co., Ltd., "PaddlePaddle," <https://github.com/PaddlePaddle/Paddle>.
- [17] Gaoding (Xiamen) Technology Co., Ltd. GaoDingKouTu. <https://koutu.gaoding.com/>.

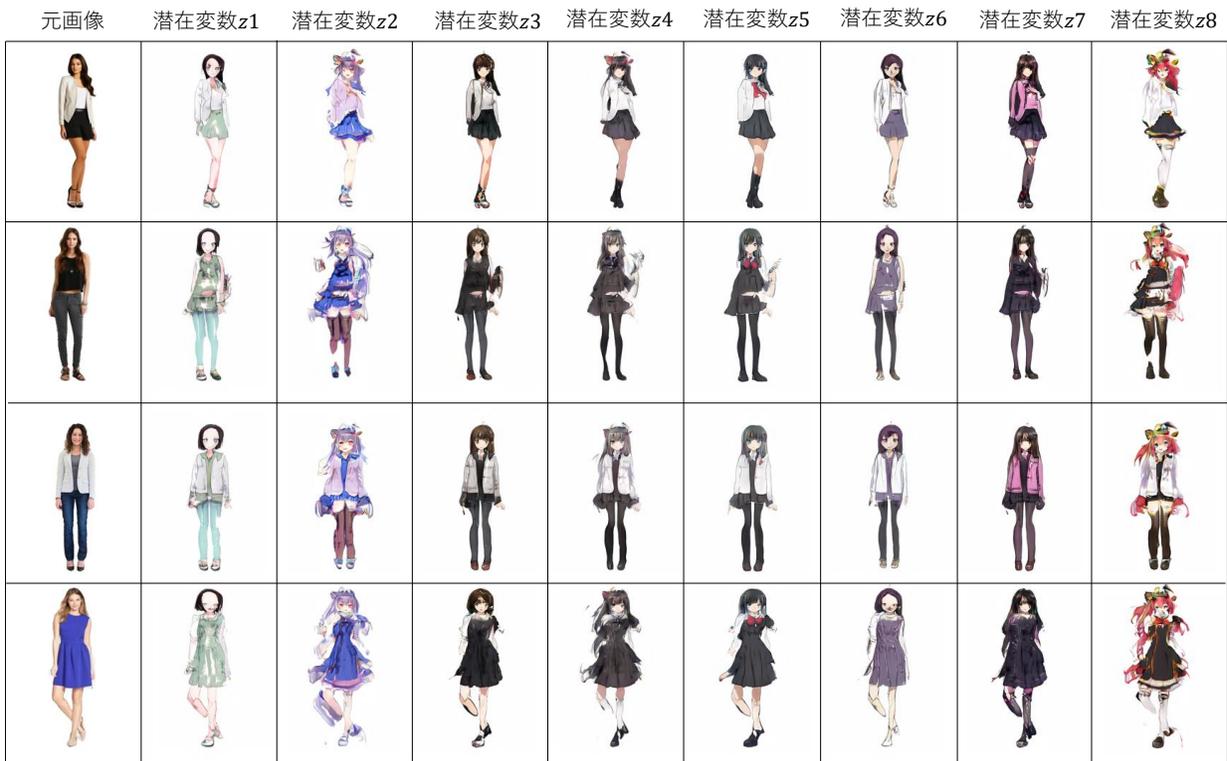


Fig. 9. Augmented CycleGAN を用いた異なる 8 通りの潜在変数に対するアニメ風画像生成結果