

ランドマーク情報に基づく漫画画像合成

OSA, Ryusei / 長, 瑠星

(出版者 / Publisher)

法政大学大学院理工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学研究科編

(巻 / Volume)

64

(開始ページ / Start Page)

1

(終了ページ / End Page)

3

(発行年 / Year)

2023-03-24

(URL)

<https://doi.org/10.15002/00026391>

ランドマーク情報に基づく漫画画像合成

IMAGE SYTHESIS FOR MANGA BASED ON LANDMARKS

長瑠星

Rysei OSA

指導教員 Jinjia Zhou

法政大学理工学研究科応用情報工学専攻修士課程

With the goal of creating a Manga drawing assistance system using automatic generation techniques based on Deep Learning, this paper describes the construction of a system that generates Manga characters. I propose a method to generate a cartoon image from random noise and landmark information and a method to generate a cartoon image from the original image, its landmark information and target landmark information. And, I created Manga image datasets for each method.

Key Words : Manga, Deep Learning, Image Synthesis

1. はじめに

近年、コロナ禍の巣ごもり需要もあり、漫画の売り上げが増加している。全国出版協会出版科学研究所は漫画が過去最大の市場規模に達したと発表した[1]。漫画の作画は通常一人の漫画家と複数人のアシスタントによって行われる。そのうち、メインとなるキャラクターの作画については、漫画家一人によって時間をかけて行われることがほとんどである。これは作画の技術が複雑であり作者によって様々な表現方法が存在し、他人による再現が非常に困難であるためであると考えられる。

キャラクターの顔はパーツの形や配置に多くの表現方法が存在し、表情や向きが変化しても元の表現を維持する必要があるため高度な作画技術を要する。

そのため、機械学習を用いたキャラクターの表情や向きに対応した自動作画システムの作成は作画に要する時間の節約に有効であり、漫画家が物語の構成等他の重要な作業に集中することができる。このシステムはアニメ等のカラーイラストのアニメーション作成にも転用可能である。

また、漫画やアニメのデータセット拡張にも利用でき、それらに関連した画像認識モデルの精度向上にも期待できる。本研究では、ランダムノイズからの画像生成が可能なランドマーク情報に基づく漫画画像合成手法 1 とポーズ転送モデル Pose Attention Transfer[2]を用いたランドマーク情報に基づく漫画画像合成手法 2 を提案する。

2. 既存研究

Cartoon GAN[3]、Anime GAN[4]では対になっていない学習データで写実的な画像のアニメ、漫画スタイルへの変換を行った。Anime GAN は Cartoon GAN の改良であり、256×256 ピクセルの 6,656 枚の実写画像と作画監督ごとに同じアニメーション作品内での約 2000 枚の画像を用い学習を行った。Anime GAN によるスタイル変換画像を図 1 に示す。一番左の列はスタイル変換前の写真であり、それ以外の各列はそれぞれ対応した作者のスタイルのア

ニメ画像である。

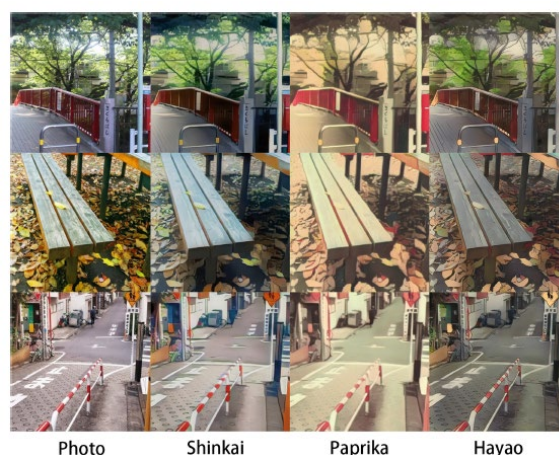


図 1 Anime GAN によるスタイル変換画像

Progressive Pose Attention Transfer for Person Image Generation で提案されたネットワーク

3. 既存手法の問題点

Cartoon GAN、Anime GAN では入力に写真が必要であり利用時にコストがかかる。また、絵柄ごとにネットワークを学習させる必要があり、大量の学習用データを準備する必要があるという問題がある。

4. 提案手法

3.1 データセットの作成

Manga109[5]の顔領域のアノテーションデータを基に抽出した顔画像と、ebook japan(<https://ebookjapan.yahoo.co.jp/>) © eBook Initiative Japan Co.,Ltd. © Yahoo Japan)の無料試し読み可能漫画 100 巻(少年・青年向け 50 巻、少女・女性向け 50 巻)においてスクレイピングを行い集めた画像に手で顔領域の判定を行い集めた 128×128 ピクセルの画像計 8100 枚の画像を漫画顔画像データとした。漫画顔画像データのうち 1500 枚を無作為に

選び手動で顔ランドマークのラベル付けを行った。そして、その 1500 枚のラベル付きデータを用いて HRNet[6]の訓練を行い、学習後の HRNet で残りの 6600 枚の画像の顔ランドマークの推定を行い、全 8100 枚のラベル付き漫画顔画像データを作成した。さらに、このデータセットから人物ごとに手動でグループ分けを行った。各グループの画像数は 2-10 枚になるようにし、訓練用が 575 グループ、計 2350 枚、テスト用が 50 ペア 100 枚のポーズ変換顔画像データセットを作成した。

3.2 漫画画像合成手法 1

8100 枚のラベル付き漫画顔画像データを用いて BicycleGAN を基にした図 2 の構成のネットワークを学習させた。上のネットワークでは Encoder を用いて元画像を正規分布のパラメータ(平均、分散)に変換する。そこからサンプリングした値 z と目標ランドマーク画像を Generator の入力として画像生成を行う。生成画像のランドマークの識別を行う学習済み Landmark Detector を追加し、識別結果が目標ランドマークから遠い場合にペナルティを付加した。ここで Landmark Detector は学習データ作成時に学習させた HRNet である。下のネットワークは生成画像から z を復元するためのネットワークであり、標準正規分布からサンプリングした z を Generator に入力する。その生成画像を Encoder に入力したとき、出力の一つである平均と z の差を縮めるように Generator を学習させる。

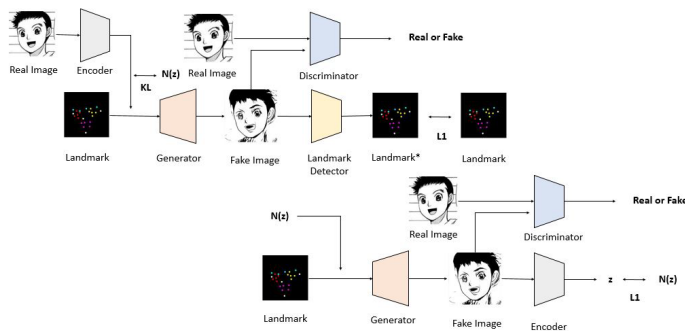


図 2 漫画画像合成手法 1 のネットワーク構成

3.3 漫画画像合成手法 2

ポーズ変換顔画像データセットを用いて図 3 のネットワークの学習を行った。このネットワークは Pose-Attentional Transfer Network を基にしたものである。学習時には全 2350 枚の中からランダムに 1 枚選び、その画像が属するグループの別の画像をポーズ転移先データとして学習を行うように設定した。

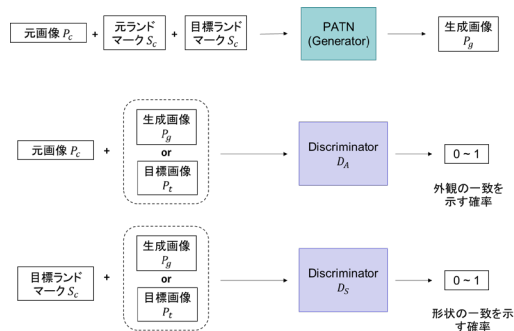


図 3 漫画画像合成手法 2 のネットワーク構成

3.4 評価指標

また、生成画像が学習に使用したデータとどのくらい類似しているかを示す定量的な評価指標として SSIM[7]を用いた。

5. 実験結果

5.1 手法 1 の結果

学習後の図 2 のネットワークの Generator による生成結果を図 4 に示す。Generator の入力としてランドマーク画像と標準正規分布からサンプリングしたランダムノイズを用いた。一番左の列は入力に使用したランドマークを表し、その他の列はランダムノイズを入力した際の出力画像である。図 4 から、髪の色についてはノイズによって多様性が見られたが、髪型の変化が入力ランドマークに依存してしまっていた。

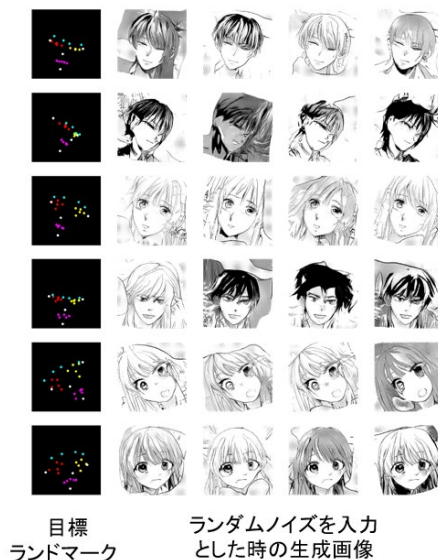


図 4 手法 1 の生成結果

5.2 手法 2 の結果

学習後の図 3 のネットワークの Generator による生成結果を図 5 に示す。右列から順に元画像、元画像のランドマーク、変換後の正解画像、変換後のランドマーク、変換後の生成画像を示す。背景や髪の輪郭は崩れているが、髪や目の質感をある程度維持しつつ、指定されたランドマー

クにポーズ変換が行えることが確認できた。

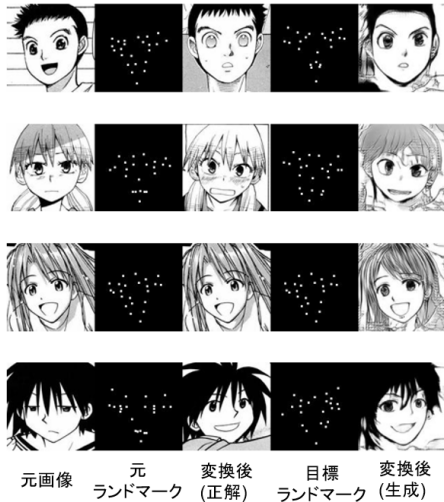


図5 手法2の生成結果

表1に手法2の生成画像におけるSSIM値とProgressive Pose Attention Transfer for Person Image Generationの実験におけるSSIM値の比較を示す。背景が単純であるDeepFashionにおけるSSIM値には及ばなかったが小規模なデータセットであるが263,632ペアのデータセットであるMarket-1501に近い値を得ることが出来た。

表1 本実験でのSSIM値とProgressive Pose Attention Transfer for Person Image Generationの実験におけるSSIM値の比較

データセット	SSIM
本実験におけるポーズ変換 顔画像データセット	0.262
Market-1501 (実写の全身画像)	0.311
DeepFashion (実写の全身画像 背景白)	0.773

6. 考察

手法1は1つのランドマーク情報と調整可能な潜在ベクトルを入力として画像の生成が可能であるため、作画の技術を持っていないユーザーでも容易に利用可能であると考えられる。一方で、特定の絵柄を再現することが困難であるため、自分や特定の漫画家の絵柄で画像を生成するという用途には向かないと考えられる。また、ランドマーク情報が変化すると、絵柄も変化してしまうためアニメーションへの利用にも不向きである。

手法2は変化元となる画像とそのランドマークの組を用意する必要があるが入力画像の絵柄を忠実に再現することができる。そのため、作画の技術を持っていないユーザーの利用は困難であるが、技術を持つ漫画家が効率化のために利用する際に有用であると考えられる。また、

ランドマーク情報が変化しても絵柄が変化しないためアニメーションへの利用が可能である。

生成画像の質の低下には、データセットの不足、多様な表現方法の存在が理由として考えられる。データセットの不足は研究で容易に利用可能で整ったデータセットがほとんど存在せず、収集が困難であることに起因する。多様な表現方法の存在はすべての画像について背景が異なることに加え、喜怒哀楽などの表現のために、同一漫画内の同一人物の画像でも顔のパーツの色が異なる場合があることが原因であると考えられる。

7. まとめ

漫画人物顔画像の自動生成システムの構築を行った。ランダムノイズとランドマーク情報から漫画画像を生成する手法1と元画像とそのランドマーク情報と目標ランドマーク情報から漫画画像を生成する手法2を提案した。また、それぞれの手法に対応した漫画画像データセットを作成した。手法1は作画技術を持たないユーザーが0から生成を行うことに適しており、手法2は作画技術を持つ漫画家の絵柄を再現し、時間の節約に有効であることがわかった。

参考文献

- [1] 全国出版協会出版科学研究所, 「コミック販売額」
<https://shuppankagaku.com/statistics/comic/>,
Copyright © 2023 AJPEA.
- [2] Zhu, Zhen and Huang, Teng and Shi, Baoguang and Yu, Miao and Wang, Bofei and Bai, Xiang, Progressive Pose Attention Transfer for Person Image Generation. In CVPR. pp. 2347-2356. 2019.
- [3] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. Cartoongan: Generative adversarial networks for photo cartoonization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9465-9474, 2018.
- [4] Jie Chen, Gang Liu, Xin Chen. AnimeGAN: A Novel Lightweight GAN for Photo Animation. In ISICA 2019: Artificial Intelligence Algorithms and Applications pp 242-256. 2019.
- [5] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K. Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications. 2017.
- [6] Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong. Deep High-Resolution Representation Learning for Human Pose Estimation. In CVPR. 2019.
- [7] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. In IEEE transactions on image processing. pp. 600-612. 2014.