

アニメの静止画像を対象とした CycleGAN による画風変換手法の検討

西山 昂志¹ 當間 愛晃² 赤嶺 有平² 山田 孝治² 遠藤 聡²

概要: 日本においてアニメの歴史は長くそれに伴い様々な変化を経ている。例えば 1980 年代はセルアニメーションの作品が多かったが、現在ではコンピュータの発展に伴い、ほとんどがコンピュータアニメーションとなっている。製作方法の変化や技術の発展に伴い、アニメ作品の画風 (絵タッチ, 背景, 色合い等) も同様に変化していると考えられる。画風を変換する研究分野においては、芸術絵画や写真の画風を別の画風に変換する方法が提案されている。そこで、対象画像をアニメの静止画像とし、セルアニメーションが主流であった年代のアニメ画像を、現代のデジタル作画のアニメの画風に変換するタスクを考えた。本研究では、Image-to-image の手法の 1 つである CycleGAN をベースにアニメ画像の画風変換結果を報告する。

1. はじめに

日本ではじめて TV アニメが放送されたのは 1958 年であるとされている [7]。初の TV アニメ放送から現代に至るまでの過程で、TV アニメの作画や色合い、コントラストなどは絶えず変化し続けている。こういったアニメの画風の変化には、コンピュータ技術の進歩も大きく関わっており、変遷の 1 つとしてセルアニメーションからデジタル作画のアニメーションへの移行が考えられる。セルアニメーションとは、透明なセル数枚に登場人物や背景を直接書き込み、最後は重ね合わせて撮影するアニメーション制作方法で、おおよそ 1990 年代まではこの方法が主流であった。以降は徐々にコンピュータによるデジタル作画が用いられるようになり、現代ではセルアニメーションは消滅している。セルアニメーションからデジタル作画への移行の主目的は、制作工程の効率化によるコストカットであるが、この制作方法の変化がアニメの画風にも影響を与えている 1 つの理由と考えられる。これまでに制作されたアニメ作品は膨大な数であるが、新規タイトル制作だけではなく、セルアニメーションによって制作されている人気作品を、現代の技術 (デジタル作画, 3DCG など) でゼロから作り直すリメイク作品も登場している。Image-to-Image の研究分野では、写真を絵画の仕上がりに変換する画風変換の研

究も提案されていることから、セルアニメーションで制作されたアニメ画像をリメイク作品のようなデジタル作画で制作されている画風に変換するタスクを考えた。

本研究では、Image-to-Image の手法の 1 つである CycleGAN を用いて、作中の 1 キャラクターに焦点を絞り、セルアニメーションの年代のアニメ静止画像をデジタル作画の年代のアニメ画像に変換する。

2. 関連研究

Image-to-Image の研究は多岐にわたるが、Neural network を用いた変換手法として、Gatys らは [2] 画像を絵画のような仕上がりに変換する研究を発表している。これはコンテンツ画像とスタイル画像の 2 枚の画像を用い、VGG によって得られた特徴マップを合成することで、コンテンツ画像の物体の配置などを維持したまま、スタイル画像のテクスチャ情報を反映させることができる。また、Fujun ら [5] は入力にセグメント画像を追加することで、コンテンツ画像の構造厳格に維持したまま変換する手法を提案している。これらの手法はスタイル画像の画風を反映させた画像を生成するが、スタイル画像とコンテンツ画像の組み合わせに大きく依存する。

GAN (Generative Adversarial Network) [3] を用いた画像変換手法として、Isola らは [4] 変換前と変換後の画像のペアを用意することで、GAN を用いてある画像から別の画像を生成する手法を提案している。また、ペア画像の用意は非常に困難である問題点に対して、Zhu ら [8] はペア画像ではなく 2 つの画像群からドメインを学習する CycleGAN を提案している。

¹ 琉球大学理工学研究科情報工学専攻
Information Engineering Course, Graduate School of Engineering and Science, University of the Ryukyus

² 琉球大学工学部工学科知能情報コース
Computer Science and Intelligent Systems Program, Faculty of Eng., Univ. of the Ryukyus

3. 提案手法

本研究ではアニメ画像を対象に、セルアニメーションが主流であった年代のアニメ画像から、デジタル制作の年代のアニメ画風への変換を試みる。対象となる画像のペアデータ作成が非常に困難であることから、ペア画像を必要としない CycleGAN の枠組みを採用する。加えて、アニメ中の 1 登場キャラクターの構造を保持するための損失関数として、登場キャラクターの存在有無について学習した VGG16 のネットワーク構造の識別器から得られた特徴マップの差の L1 ノルムを追加する。

3.1 CycleGAN

CycleGAN[8] のゴールは与えられたドメイン X(以下 X) とドメイン Y(以下 Y) の両方のマッピング関数を学習することである。そのため、X を Y に変換する G と Y を X に変換する F の 2 つの生成器が必要となる。加えて、 $F(y)$ と X, $G(x)$ と Y を識別する D_X と D_Y の 2 つの識別器を用いる。

CycleGAN では GAN と同様に、adversarial loss が用いられる。これは、式 (1) によって定義される。式 (1) は G と D_Y についてのものであるが、同様に F と D_X についての adversarial loss も定義される。

$$\mathcal{L}_{adv}(G, D_Y, X, Y) = \mathbb{E}_y[\log D_Y(y)] + \mathbb{E}_x[\log(1 - D_Y(G(x)))] \quad (1)$$

加えて、cycle consistency loss と呼ばれる損失関数が用いられる。これは、式 (2) によって定義され、 $x \rightarrow G(x) \rightarrow F(G(x)) \rightarrow x$ となるようする損失関数である。

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_x\|F(G(x)) - x\|_1 + \mathbb{E}_y\|G(F(y)) - y\|_1 \quad (2)$$

式 (1)、式 (2) から、CycleGAN の損失関数 $\mathcal{L}_{cycleGAN}$ は式 (3) で表される。

$$\mathcal{L}_{cycleGAN}(G, F, D_X, D_Y) = \mathcal{L}_{adv}(G, D_Y, X, Y) + \mathcal{L}_{adv}(F, D_X, Y, X) + \mathcal{L}_{cyc}(G, F) \quad (3)$$

3.2 Content loss

CycleGAN に加えて、登場する 1 キャラクターの細かい構造を保持するため、content loss[1] を導入する。アニメ作中の 1 キャラクターを対象として、画像中の存在有無を判定する識別器を訓練する。今回は VGG16[6] のネットワーク構造を用い、式 (4) のように入力画像と生成画像の

表 1 CycleGAN で用いたデータ数

	ドメイン X	ドメイン Y
総数	1947	1604
訓練データ	1000	1000
テストデータ	100	100

特徴マップの差に対する L1 ノルムの正規化項を追加する。

$$\mathcal{L}_{con}(G, F) = \mathbb{E}_x\|VGG(G(x)) - VGG(x)\|_1 + \mathbb{E}_y\|VGG(F(y)) - VGG(y)\|_1 \quad (4)$$

また、変換前と変換後の画像間における色の構成を保持する正規化項として、identity loss を用いる。

$$\mathcal{L}_{idn}(G, F) = \mathbb{E}_y\|G(y) - y\|_1 + \mathbb{E}_x\|F(x) - x\|_1 \quad (5)$$

最終的な目的関数は式 (6) となる。

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{cycleGAN} + \lambda_{con}\mathcal{L}_{con} + \lambda_{idn}\mathcal{L}_{idn} \quad (6)$$

λ_{con} と λ_{idn} はそれぞれの損失関数のバランスを制御する重みである。

4. 実験

4.1 データセット

セルアニメーションの静止画像とデジタル作画アニメの静止画像として、「ドラえもん」の映画作品のキーフレームを用いた。同作品の映画タイトルを用意し、それぞれセルアニメーションの静止画像をオリジナル作品 [10]、デジタル作画の静止画像をリメイク作品 [9] とし、1 秒ごとにキーフレームを取り出し、それぞれの分割した画像群から登場キャラクター「ドラえもん」の存在有無の選別を行い、セルアニメーションで描かれたドラえもんであるドメイン X、デジタル作画で描かれたドラえもんドメイン Y とした。データ数はそれぞれ、1947、1604 枚となった (表 1)。

また、登場キャラクターの存在有無を判別する識別器訓練のため、映画ドラえもんの別作品を同様にキーフレーム分解し、存在しているデータと、存在していないデータを収集した。

4.2 パラメータ

パラメータの設定として、 λ_{con} と λ_{idn} の値を 10 に設定した。学習済みの特徴抽出部は VGG16 の conv4.3 とし、収集したドメイン X とドメイン Y のデータセットから訓練データとしてそれぞれ 1000 枚を用い、バッチサイズを 1 とした上で、80 エポックまで学習を行った。

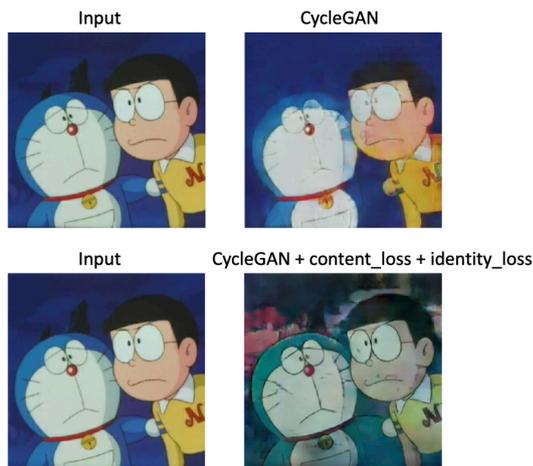


図 1 content loss の有無



図 2 人工物が多い結果

4.3 実験結果

得られた結果から content loss を用いた結果と、CycleGAN の結果を図 1 に示す。テスト画像 100 枚を目視によって確認したところ、content loss の導入により、図 1 の content loss 導入前よりキャラクターの構造が維持できているケースが確認できた。一方で、content loss の導入により図 2 のような、変換によって画面全体に人工物が多く現れる結果も確認できた。

5. まとめ

本研究では、CycleGAN を用いてセルアニメーションのアニメ画像から、デジタル作画のアニメ画像への変換を行った。作中の 1 キャラクターの焦点を絞り、アニメ画像中の対象キャラクターの存在有無を判別する VGG16 構造の識別器を用意し、その特徴産出部による差分を正則化項として追加した。これにより、目視による主観的な判断であるが、対象キャラクターの構造を維持するパターンを確認することができた。しかし、content loss の導入前の結果と比べて画像の崩壊が悪化するパターンも確認できた。

今後は出力画像の崩壊が悪化するパターンに対する改善策を検討するとともに、複数のキャラクターが映り込んだ画像に対してもキャラクターの構造を維持を可能にする手法に関して調査を行う。

参考文献

- [1] Chen, Y., Lai, Y. and Liu, Y.: CartoonGAN: Generative Adversarial Networks for Photo Cartoonization, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9465–9474 (online), DOI: 10.1109/CVPR.2018.00986 (2018).
- [2] Gatys, L. A., Ecker, A. S. and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423 (online), DOI: 10.1109/CVPR.2016.265 (2016).
- [3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Nets, *Advances in Neural Information Processing Systems 27* (Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. and Weinberger, K. Q., eds.), Curran Associates, Inc., pp. 2672–2680 (online), available from (<http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>) (2014).
- [4] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks (2016).
- [5] Luan, F., Paris, S., Shechtman, E. and Bala, K.: Deep Photo Style Transfer (2017).
- [6] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2014).
- [7] 津堅信之: 新版 アニメーション学入門, 平凡社新書 (2017).
- [8] Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (2017).
- [9] 寺本幸代, 真保裕一, 藤子・F・不二雄: ドラえもん のび太の新魔界大冒険 七人の魔法使い, 映画ドラえもん制作委員会 (2007).
- [10] 柴山努, 藤子不二雄: ドラえもん のび太の魔界大冒険, シンエイ動画, テレビ朝日, 小学館 (1984).