

# GANを利用したアニメーション生成

## Video Cartoonization based on Generative Adversarial Networks

森山紘行\*<sup>1</sup>  
Hiroyuki Moriyama

李亜超\*<sup>1</sup>  
Yachao Li

下川原(佐藤)英理\*<sup>1</sup>  
Eri Sato-Shimokawara

山口亨\*<sup>1</sup>  
Toru Yamaguchi

\*<sup>1</sup> 首都大学東京大学院

Graduate School of System Design, Tokyo Metropolitan University

Animation has been playing an important role in the economy and culture. However, making animation is a hard work and costly. To solve the problem, we propose new solution for animation video generative model, which of baseline is Cartoon-GAN[1]. Cartoon-GAN is a kind of image to image style transfer network, and it has attained to generate cartoon style image from real-world scenes. In deep learning, there are usually two ways to process videos: 3d convolution or 2d convolution added with temporal processing. However, existing method doesn't achieve enough smoothness in the cartoon-style video making task. For our cartoon-style transfer task in video to video, our new solution is to use each two image frames and optical flow as an input for the generator. In this paper, we generated cartoon videos by adopting optical flow, which is effective to predict object motion.

### 1. はじめに

今や、アニメーションは社会の中で様々な価値を生み出してきている。多くのアニメーターによって作品が作られ、新しい経済効果や文化を生み出すことに貢献してきている。しかしこれまでのところ、アニメーション作成は人による作業が必要不可欠となっている。いくつかのアニメ制作会社のデータ[2]によれば、アニメの編集や作成には時間の消費量や、かかる費用が膨大であることがわかっている。そのため、もし私たちがアニメーション作成作業に関して、自動で行う手法を開発出来れば、その価値や意義は大きい。

画像から画像へのスタイル変換における最近の研究では、ほぼ良い結果を出してきている。この分野について、Cycle GAN[3]と呼ばれる手法を用いることで、機械学習時に整備された画像を用意しなくても、ある画僧から別の画像スタイルへの変換タスクにおいて一定の結果を出している。

従来の画像スタイル変換手法を応用し、ある画像を漫画風に変換する研究が行われるようになってきた。しかしこのタスクでは、a)高度なシンプル化、抽象化が必要 b)はっきりした輪郭、滑らかな色調変化、シンプルなテキストの実現が難しいという問題があった。この問題は本研究で目的としている、動画のアニメーションへのスタイル変換にも言えることである[4]。そこで我々は、2章で述べる Cartoon GAN[1]を活用することで、アニメーションらしいスタイルを出力している。

### 2. 関連研究

近年、画像のスタイル変換に関する研究では多くの場合、Ian Goodfellowらが2014年に提案したGANを基本として用いている[5]。本稿1節内のa), b)で説明したような、画

連絡先: 森山紘行, 首都大学東京大学院, moriyama-hiroyuki@ed.tmu.ac.jp

像の漫画風スタイル変換における問題点を解決するため、Cartoon GAN[1]では2つの新しい損失関数を活用している。一つは写真と漫画の間のスタイル変換に対応するため、VGGnetの高レベルな特徴マップの疎な正規化で定式化されているsemantic content lossである。またもう一つは、鮮明なエッジの維持が目的のedge-promoting adversarial lossである。これらの工夫により、現実世界の画像から漫画風画像へのスタイル変換が成功している。



図1. Cartoon GANを用いたスタイル変換の例

また、動画から動画へのスタイル変換タスク[4]にも焦点が当てられるようになってきている。このタスクにおいては、three stream(3次元畳み込み)かtwo stream(2次元畳み込みと時間的処理の組み合わせ)が主流となっている。Three stream[4]では、3次元の畳み込みを用いた、時空間変換器を提案し、この問題への1つの解決法であるとしている。しかし、3次元の畳み込みによる手法では動画内の動作情報を的確に捉えているのが明確には示されておらず、動画生成に適用するのは困難である。また、two streamとしてはLSTMを時間的処理として用いている[6][7]のものもある。しかし、LSTMは動画全体の特徴を取れない問題があり、これも動画生成への適用は難しい。

そこで本稿では、動画内全体に対して動きの特徴を取得するため、動画のスタイル変換における新規手法を次の3章のように提案する。

### 3. 提案手法

動画は、連続した画像フレームによって成り立っている。そこで本手法では、Cartoon GAN[1]を各画像フレームの生成用に用いる。動画と画像の違いは、動画は空間特徴だけでなく、時間的特性をも各フレーム間に持っていることである。我々の研究目標は滑らかなアニメーション生成であるため、次の 3.1 節で述べるようなオプティカルフローを時間的特性として用い、図 2 で示すシステム構成によるアニメーション生成を行った。

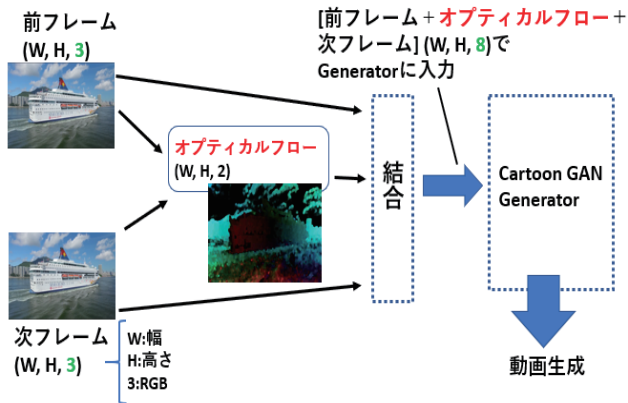


図 2. システム構成図

#### 3.1 オプティカルフロー

近年の研究により、オプティカルフローは動画生成において効果的な手法であることが分かっている [8]。そこで我々の手法では、2 つの画像フレーム間の差分ベクトルを計算するのにこれを用いた。

オプティカルフローには、疎と密、2 つのアルゴリズムがある。疎のタイプは画像内の特定の特徴点のみを分析するときに適用され、密のタイプは画像内の全てのピクセル移動に適用される。今回、2 つの画像フレーム間における、全ての差分ベクトルを取得して用いるため、密なオプティカルフローとして Gunnar-Farneback 法を利用した。

#### 3.2 Generator への入力

2 次元の畳み込みは、画像処理の特徴抽出における重要な技術の 1 つである。今回、Cartoon GAN に用いられている VGG や Resnet など、事前に学習済みの Imagenet を活用することで、空間特徴処理に向けた 2 次元畳み込みを利用出来る。また時間的特性に関しては、3.1 節で述べたオプティカルフローを適用している。

Cartoon GAN の generator に入力する前に、オプティカルフローによる処理を行う。動画に使われている画像フレームの中から、2 つを取り出す。画像は幅  $W$ 、高さ  $H$ 、色の RGB からなるため、 $(W, H, 3)$  の配列を持つ。前の画像と次の画像とでオプティカルフローを計算した結果、差分ベクトルとして 2 チャンネルの配列が返ってくる。従って、 $(W, H, 2)$  の配列となる。

前と後の画像フレーム、そしてオプティカルフローの配列を結合した  $(W, H, 8)$  の配列を Cartoon GAN の generator に入力し、動画生成を試みた。

### 4. 実験

実験に向け、我々は現実世界の動画とアニメーション動画のデータセットを用意した。これらは youtube と pixabay の中からそれぞれ人手で集めた。本稿 3 章で提案したシステムと用意したデータセットで実験を行った結果、図 3 のようなアニメーションが生成された。

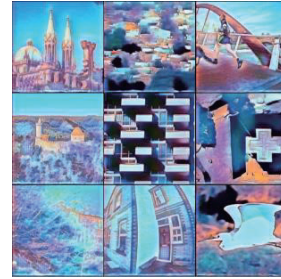


図 3. 生成されたアニメーション

### 5. まとめ

我々は、アニメーション生成において Cartoon GAN を利用し、generator への入力として、2 つの画像フレームとその間の差分ベクトルであるオプティカルフローを利用する手法を提案した。結果として、動画のアニメーションへのスタイル変換を行うことが出来た。

今後は、この生成結果を他のアニメーション生成手法とも比較し、動画の滑らかさについて評価を行う。

### 参考文献

- [1] Y. Chen, Y.-K. Lai, and Y.-J. Liu, "CartoonGAN": Generative Adversarial Networks for Photo Cartoonization," Cvpr2018, pp.9465-9474, 2017
- [2] G. Kucharska, "How much does it cost to produce animation – and why?" <https://www.studiopigeon.com/blog/how-much-does-it-cost-to-produce-animation-and-why/>, 05,01,2018.
- [3] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," Proceedings of the IEEE International Conference on Computer Vision, vol. 2017-October, pp.2242-2251, 2017.
- [4] D. Bashkurova, B. Usman, and K. Saenko, "Unsupervised Video-to-Video Translation", no. Nips, 2018.
- [5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in Advances in neural information processing systems, 2014, pp.2672-2680.
- [6] Mittal, Gaurav, Tanya Marwah, and Vineeth N. Balasubramanian. "Sync-Draw: Automatic Video Generation using Deep Recurrent Attentive Architectures." Proceedings of the 2017 ACM on Multimedia Conference. ACM, 2017.
- [7] Li, Yachao, and Toshihiro Komma. "Generating Videos Based on Convolutional Recurrent Generative Adversarial Networks." International Conference on Geometry and Graphics. Springer, Cham, 2018.
- [8] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, "Super slo-mo: High quality estimation of multiple intermediate frames for video interpolation," arXiv preprint arXiv:1712.00080,2017.