

トーン特徴量を損失関数に用いた cGAN でのトーン貼り

Tone Pasting Using cGANs with Tone Feature Loss

坪田 亘記 相澤 清晴
Koki Tsubota Kiyoharu Aizawa

東京大学
The University of Tokyo

Tone pasting is one of the processes of manga creation and there is a demand for automatic tone pasting. In this study, we tackle a task of automatic tone pasting of manga characters. Tone pasting is difficult because tones have characteristic patterns. It is hard to learn tone patterns for usual conditional generative adversarial networks (cGANs) which are combined with L_1 loss or perceptual loss. To train pasting tones in a tone pattern aware manner, we introduce *tone feature loss* to cGANs. *Tone feature loss* is the distance between tone features of target images and those of generated images. We performed experiments on two characters in Manga109 and showed our results are equal to or more visually appealing than those by a baseline method.

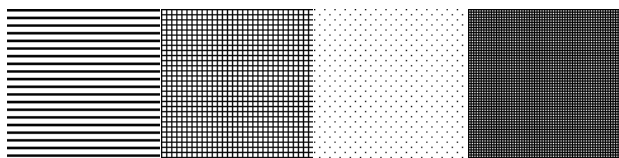


図 1: トーンの例. トーンには特有のパターンが存在する.

1. はじめに

漫画の制作において、トーン貼りは欠かせない工程である。漫画の制作の現場では、人手によってトーン貼られている。トーンの貼り方にはアナログの方法とデジタルの方法がある。前者は、トーンを貼りたい領域の形に切り貼りするために人手が必要となる。また後者についても、領域を手動で選択する必要がある。トーン貼りの自動化は、このような人手のコストを削減するため必要がある。

本稿では、キャラクターの線画に対してトーンを貼ることを目的とする。トーンは、自然画像やイラストとは異なり、特有のパターンが存在する (図 1)。自然画像の生成や線画の着色に用いられるような手法 [Isola 17, Ci 18] では、ベタに近いトーンは生成することはできるが、粗い縞模様といったパターンは生成することが難しい。これは既存手法がトーンのパターンを考慮していないためである。

我々は、トーン貼りの学習において、トーンのパターンを考慮した学習を行う手法を提案する。一般的な画像から画像への生成の学習に用いられる conditional generative adversarial network (cGAN) の枠組みを利用し、その学習時に、損失関数として目標の画像と生成画像のトーン特徴量間の距離を用いる。トーン特徴量としては、同じトーン領域内では同じ特徴量を示すものを利用する。実験により、我々の提案手法がトーンのパターンを学習しやすい手法になっていることを確認する。

2. 関連研究

2.1 線画・漫画の着色

線画の着色は、キャラクターの線画に対するトーン貼りとは、線画内の領域を選択して色を置くという点で近いタスクである。pix2pix [Isola 17] では、cGAN を用いて自然画像のエッジから自然画像を生成している。また、cGAN は漫画の着色 [Hensman 17, Furusawa 17] やイラストの着色 [Ci 18] でも用いられる。本研究でも既存の線画の着色手法に倣い、cGAN を用いてトーン貼りを行う。

2.2 cGAN を用いた画像生成

cGAN では、損失関数として生成画像を本物の画像に近づけるような学習をするための adversarial loss と、タスク依存の損失関数を組み合わせて用いる。タスク依存の損失関数として、 L_1 loss を用いる手法 [Isola 17] や、 L_2 loss を用いる手法 [Pathak 16]、perceptual loss [Ci 18] を用いる手法がある。

これらの手法は、トーンのパターンを考慮した損失関数とはなっていない。 L_1 loss や L_2 loss はトーンの貼り方が 1 ピクセルずれただけでも損失が発生してしまう。また perceptual loss は自然画像で学習した畳み込みニューラルネットワークから得られる特徴量間の距離を損失関数としており、特徴量がトーンのパターンを考慮していない。本研究ではトーンのパターンを考慮するために、タスク依存の損失関数として、トーン特徴量の距離に基づく損失関数を用いる。

3. 手法

トーンのパターンを考慮するために、cGAN で用いられるタスク依存の損失関数としてトーン特徴量間の距離とする手法を提案する。手法の概要を図 2 に示す。まず、cGAN を用いたベースライン手法と言える pix2pix [Isola 17] を説明した上で、提案手法の説明を行う。

以降で数式を用いて説明するために文字を定める。 x が入力画像 (線画) であり、 y が x に対応したトーン付きの画像である。また cGAN における G が生成器であり、 D が識別器である。

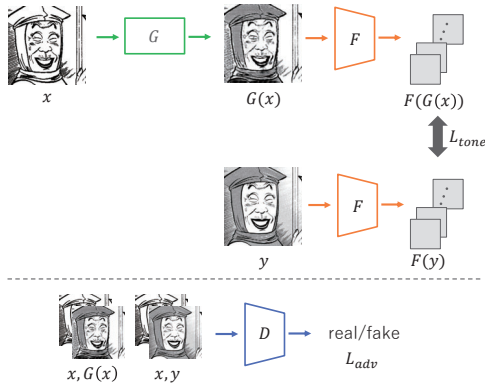


図 2: 手法概要. ©やまだ 浩一

3.1 pix2pix

pix2pix [Isola 17] は、ペアの学習データを用いて画像から画像への変換を行う手法である。生成画像を本物の画像に近づけるような学習を行うための adversarial loss

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

を導入している。また、タスク依存の損失関数としては L_1 loss を用いている。 L_1 loss は以下の式で定められる。

$$\mathcal{L}_1(G) = \mathbb{E}_{x,y} [\|y - G(x)\|_1]. \quad (2)$$

全体の損失関数としては、この両式を組み合わせた、

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \lambda \mathcal{L}_1(G), \quad (3)$$

で定められる。この損失関数を用いて、以下の式に基づき G の学習を行う。

$$G^* = \arg \min_G \max_D \mathcal{L}(G, D). \quad (4)$$

しかし、前述したように、タスク依存の損失関数として L_1 を用いた場合は、トーンのパターンの学習が難しい。

3.2 トーン特徴量を用いた損失関数

提案手法は、cGAN におけるタスク依存の損失関数として、トーンのパターンが学習しやすいようにトーン特徴量間の距離を用いる手法である。トーンの特徴抽出器を F と定めると、

$$\mathcal{L}_{tone}(G) = \mathbb{E}_{x,y} [\|F(y) - F(G(x))\|_1], \quad (5)$$

と損失関数は表される。本研究では F として、Gabor フィルタを用いた特徴量 [Qu 06] を利用する。この F は微分可能であることに注意が必要である。

タスク依存の損失関数として \mathcal{L}_{tone} を導入した場合、全体の損失関数は以下の式で表される。

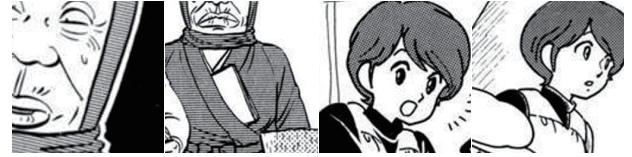
$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \lambda_{tone} \mathcal{L}_{tone}(G). \quad (6)$$

この損失関数を、式 4 を用いて最適化する。

4. 実験

4.1 実験設定

トーン付きの画像データセットとして、漫画画像データセットの Manga109 [Matsui 16, Ogawa 18] の漫画を用いた。デー

図 3: 学習データの例。左側が“留”，右側が“篠原かもき”。
©やまだ 浩一 ©ふくやま けいこ

タセットに付与されている，“body” のアノテーションのバウンディングボックスに従って画像の切り出しを行い、単一キャラクターについて実験を行った。対象としたキャラクターは，“平成爺メン”に登場する“留”と“タップ君の探偵室”に登場する“篠原かもき”である。画像の例を図 3 に示す。“留”はトーンが粗く，“篠原かもき”はトーンが細かいことが特徴である。

“留”の画像は 151 枚あり、また“篠原かもき”の画像は 239 枚あった。“留”の画像は、学習用データを 104 枚、検証用データを 14 枚、テストデータを 33 枚と無作為に分割した。“篠原かもき”の画像は、学習用データを 168 枚、検証用データを 20 枚、テストデータを 51 枚と無作為に分割した。

トーン付き画像と対応した線画データセットは存在しないため、トーン除去の手法を漫画に適用した後、2 値化することで、線画とトーン付き画像のペアを作成した。トーン除去の手法としては [Li 17] を用いた。

ネットワークは、Isola ら [Isola 17] と同様のものを利用した。すなわち、cGAN における生成器として U-net [Ronneberger 15] に基づいたネットワークを用い、識別器として PatchGAN を用いた。学習時の係数は、 $\lambda_{tone} = 100$ とした。

入力画像については、最短の辺の長さが 256 になるようにアスペクト比を変更せずにリサイズした後、学習時は 256x256 にランダムクロップを行い、テスト時はセンタークロップを行った。

4.2 実験結果

実験結果を図 4 及び図 5 に示す。ベースライン手法は pix2pix である。図 4 から、提案手法はベースライン手法と比較してトーンのパターンを生成できていることが分かる。また、pix2pix における λ が小さい場合は、トーンのパターンを生成できている部分もあるが、一方で \mathcal{L}_1 の重み λ が小さくなることで塗るべき領域の学習が難しくなっている。図 5 のようにトーンが細かい場合は、提案手法・ベースライン手法いずれの場合でも十分学習できている。

5. 結論

本稿では、キャラクターの線画に対するトーン貼りに取り組んだ。パターンを考慮するために、cGAN でのタスク依存の損失関数として、目標の画像と生成画像とのトーン特徴量間の距離を損失関数として用いた。実験により、提案手法がベースライン手法である pix2pix と比較して、有効な場合があることを示した。

今後の展望として、2 つ考えている。1 つは、学習データに存在しないキャラクターへのトーン貼りである。線画の着色の研究のように、複数キャラクターを用いて学習することで実現できると考えている。

もう 1 つは、テストデータとして純粋な線画を用いた場合の実験である。本稿では、テストデータとして漫画から抽出し

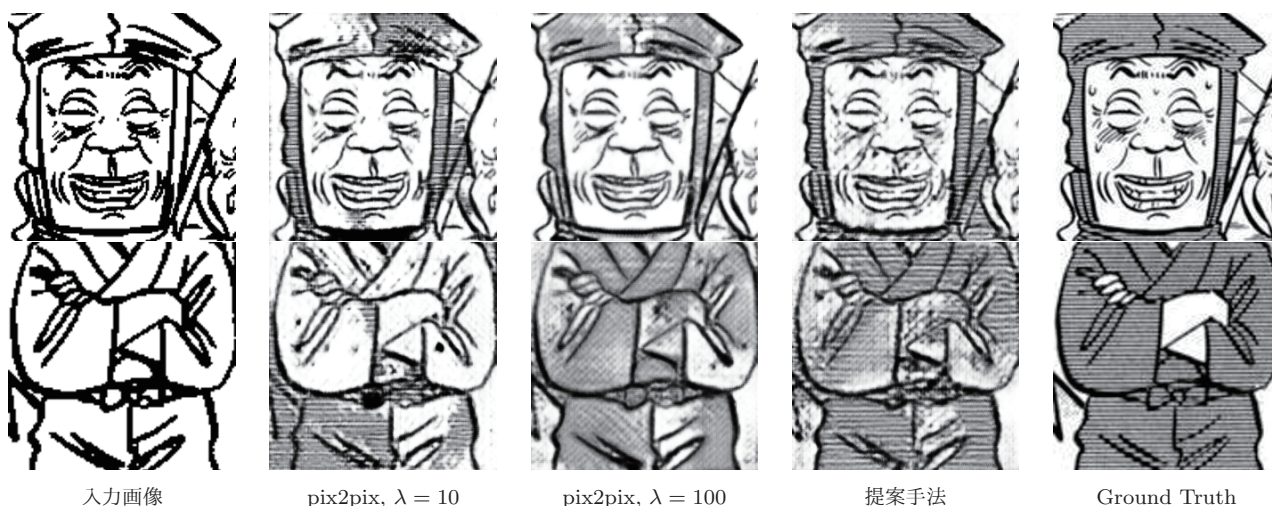


図 4: “留”での実験結果。©やまだ 浩一

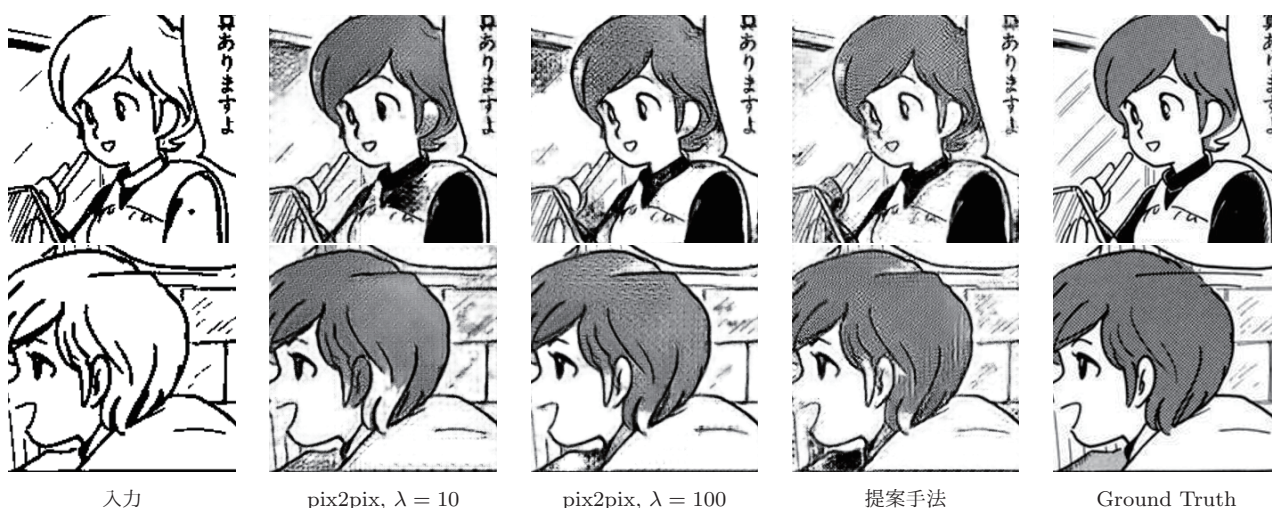


図 5: “篠原かもき”での実験結果。©ふくやま けいこ

た線画を用いて実験を行っている。そのため、得られた線画には元々貼ってあったトーンの情報が残っている可能性がある。

参考文献

- [Ci 18] Ci, Y., Ma, X., Wang, Z., Li, H., and Luo, Z.: User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks, in *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 1536–1544 (2018)
- [Furusawa 17] Furusawa, C., Hiroshiba, K., Ogaki, K., and Odagiri, Y.: Comicolorization: Semi-automatic Manga Colorization, in *SIGGRAPH Asia Technical Briefs* (2017)
- [Hensman 17] Hensman, P. and Aizawa, K.: cGAN-Based Manga Colorization Using a Single Training Image, in *2nd International Workshop on coMics Analysis, Processing, and Understanding, 14th IAPR International Conference on Document Analysis and Recognition*, pp. 72–77 (2017)
- [Isola 17] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in *Proceedings of the IEEE Conference on CVPR*, pp. 5967–5976 (2017)
- [Li 17] Li, C., Liu, X., and Wong, T.-T.: Deep Extraction of Manga Structural Lines, *ACM Transactions on Graphics*, Vol. 36, No. 4, pp. 117:1–117:12 (2017)
- [Matsui 16] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K.: Sketch-based manga retrieval using manga109 dataset, *Multimedia Tools and Applications* (2016)
- [Ogawa 18] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T., and Aizawa, K.: Object Detection for Comics using Manga109 Annotations, *arXiv preprint arXiv:1803.08670* (2018)
- [Pathak 16] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A.: Context encoders: Feature

learning by inpainting, in *Proceedings of the IEEE Conference on CVPR*, pp. 2536–2544 (2016)

[Qu 06] Qu, Y., Wong, T.-T., and Heng, P.-A.: Manga Colorization, *ACM Transactions on Graphics*, Vol. 25, No. 3, pp. 1214–1220 (2006)

[Ronneberger 15] Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241 (2015)