

Attention を導入したアニメ線画の自動彩色手法の開発および検証

南谷 大輔† 米澤 弘毅‡

名城大学理工学研究科情報工学専攻† 名城大学情報工学部情報工学科‡

1 はじめに

近年のアニメ業界では若年層の減少によって労働力が不足し技術を伝承できないといった課題があり、このままではアニメ文化を維持並びに発展させることが困難な状況に陥る可能性があると言われてしている。しかし、業界では液晶タブレットの使用を代表としたコンピュータ技術を用いた制作支援技術を取り入れるなど労力の削減を目指す取り組みも行われている。

ニューラルネットワークを用いた自動彩色については実画像を対象としたものが多く、イラストなどの絵を対象とした研究はあまり行われていない。

本研究では参考画像とアニメ線画をニューラルネットワークへ入力し、参考画像の色使いをヒントに線画を彩色する手法の開発を行う。これにより希望する色で線画を彩色することが目標である。

2 先行研究

2.1 Tag2Pix[1]

この手法は、タグ情報をヒントにして線画を彩色する手法であり、GAN（敵対的生成ネットワーク）が基になっている。タグ情報には、「red hair」のように彩色したいパーツと対応する色の情報が紐づけられている。タグ情報は畳み込み処理などを行い色情報へ変換された後に Generator の Decoder 部分にて線画と結合され、線画の彩色を行う。

3 提案手法

3.1 提案手法の概要

今回構築したネットワークは以前提案したネットワーク[2]を改良したもので、GAN（敵対的生成ネットワーク）が基になっている。構築したネットワークの概要を図1に示す。

このネットワークはGANのGeneratorの部分であり、参考画像から得られる色情報をヒントとして線画を彩色する。参考画像から得られる色情報については3.2節にて説明する。

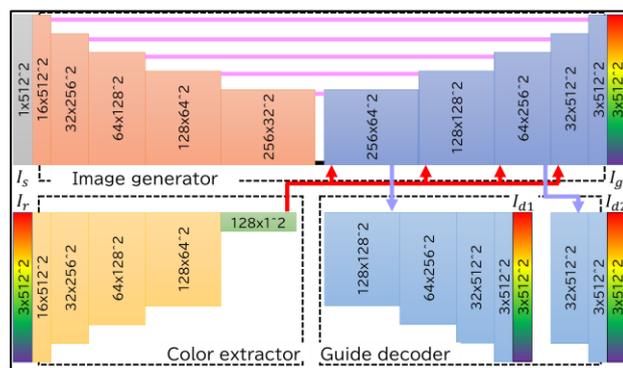


図1 ネットワークの概要図

ネットワークの概要（図1）について説明する。 I_s は入力画像、 I_r は参考画像、 I_g は生成画像である。 I_{d1} はガイドデコーダ1による生成画像、 I_{d2} はガイドデコーダ2による生成画像であり、ガイドデコーダと損失関数は以前提案したネットワーク[2]と同じものを使用した。

図1のブロック内に書かれている数字は対象のブロックが出力するテンソルのサイズを次元数×画像のピクセル数の2乗で表している。DiscriminatorについてはTag2Pix[1]と同じものを使用した。

3.2 色情報の生成

色情報はカラーの参考画像を用いて、図1のColor extractorの部分で生成される。Color extractorは4個のEncoderモジュールとAttentionモジュールから構成され、Encoderモジュールは以前提案したネットワーク[2]と同じものを使用した。

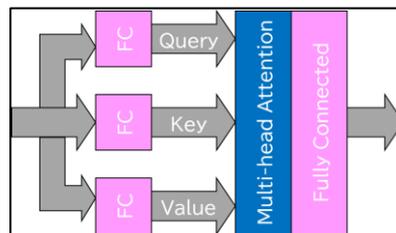


図2 Attentionモジュールのネットワーク図

Attentionモジュールのネットワーク図を図2に示す。AttentionモジュールはColor extractorの最後の部分にあり、Global Max Pooling層を経由し

Towards anime drawing colorization with Attention modules

†Daisuke NANYA · Meijo University

‡Kouki YONEZAWA · Meijo University

た後で処理を行う。以前提案したネットワーク[2]では単に全結合層のみであったが、今回は Multi-Head Attention の前後に全結合層を導入した。

参考画像はネットワークの学習時に変形処理を行った後に Color extractor へ入力される。これは、テスト時に線画と参考画像の構図が異なる場合でも彩色できるようにするためである。変形処理は、torchvision ライブラリの RandomPerspective を用いて行った。distortion_scale を 0.5 と設定し、空白になった場所には白色で塗りつぶされるようにした。ただし、テスト時は参考画像の変形処理は行わない。

3.3 入力画像と色情報の結合

入力画像と色情報の結合は Decoder の部分で行われる。色情報との結合には Tag2Pix [1]で提案されている手法を参考に構築した。

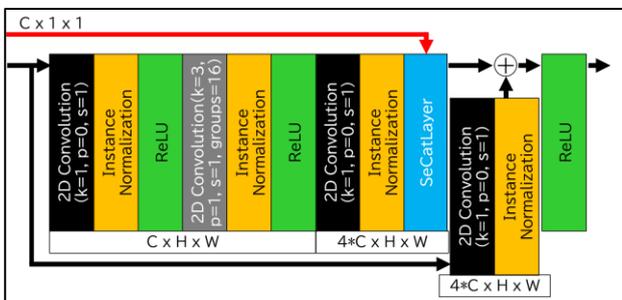


図 3 Decoder モジュールのネットワーク図

Decoder モジュールのネットワーク図を図 3 に示す。Tag2Pix では正規化にバッチ正規化を使用していたが、提案手法ではインスタンス正規化に変更しバッチサイズによる影響を排除した。Decoder モジュールの利用回数は Tag2Pi と同じく、256次元で12回、128次元で8回、64次元で5回、32次元で5回であり、次元を下げる際にはピクセルシャッフルによるアップサンプリングを行っている。

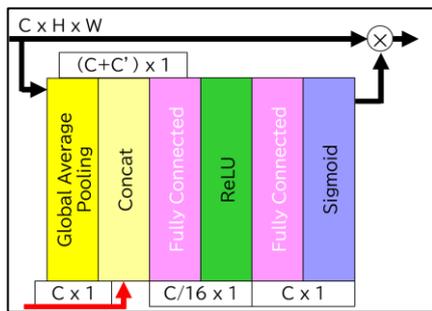


図 4 SeCatLayer のネットワーク図

SeCatLayer のネットワークを図 4 に示す。これは Tag2Pix でも使用されたものである。

4 検証結果

4.1 使用するデータセット

ネットワークの学習には、Anime Sketch

Colorization Pair データセット[3]を利用した。このデータセットには、カラー画像とそれに対する線画がセットになっており、全部で 17769 枚の画像がある。その中から 14424 枚をネットワークの学習用、3545 枚をテスト用として利用した。

4.2 出力画像の検証

出力画像の検証には学習で使用していないテスト用データの 3545 枚を用いる。学習におけるバッチサイズは4、エポック数は60である。検証手法には FID (フレッシュ開始距離) を用いる。各手法における FID の値を表 1 に示す。

表 1 各手法における FID の値

手法	FID ↓
pix2pix[4]	39.81
以前の提案手法[2]	13.38
提案手法(ヘッド数:1)	17.33
提案手法(ヘッド数:2)	18.54
提案手法(ヘッド数:4)	13.84
提案手法(ヘッド数:8)	13.75
提案手法(ヘッド数:16)	14.61

5 おわりに

本研究ではアニメ線画の彩色を目的としてネットワークの構築を行った。また Attention の利用は色情報の生成のみに留まったが、今後は Encoder や Decoder 部分に導入し、より高品質な画像を生成できるネットワークの構築に取り組みたい。

参考文献

- [1] H. Kim, H.Y. Jhoo, E. Park, S. Yoo, Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss, in: 2019: pp. 9056–9065. https://openaccess.thecvf.com/content_ICCV_2019/html/Kim_Tag2Pix_Line_Art_Colorization_Using_Text_Tag_With_SECat_and_ICCV_2019_paper.html.
- [2] D. Nanya, K. Yonezawa, Towards anime drawing colorization with reference-guided convolutional neural network and composition verification, in: Information Processing Society of Japan, 2023.
- [3] TAEBUM KIM, Anime Sketch Colorization Pair, Kaggle.Com. (2019). <https://www.kaggle.com/datasets/ktaebum/anime-sketch-colorization-pair>.
- [4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, (n.d.). <https://arxiv.org/pdf/1611.07004.pdf>.