

VGG と U-Net を組み合わせたニューラルネットワークによる

グレースケール画像の自動着色

Image Colorization by Neural Networks Combining VGG and U-Net

原口 俊樹[†] 黒木 修隆[†] 廣瀬 哲也[‡] 沼 昌宏[†]
 Toshiki Haraguchi Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

1. はじめに

現存する過去の写真や戦時中の写真は、グレースケールのものが多く存在している。これらの画像に対して着色を行うことにより、当時の様子をより理解することや、戦争の惨劇をより身近なものとして感じることができるようになる。しかし、グレースケール画像はカラー情報の多くが失われているため色の復元は困難を極める。もし人間の手で行うと、着色者の経験と勘に大きく依存するという問題がある。

2. 関連研究

Kamayar ら[1]は Generative Adversarial Network(GAN)をベースとした着色手法を提案している。これはグレースケール画像を CNN に通して生成したカラー画像と、真のカラー画像とを競合させることにより学習を進める手法である。一方、Gustav ら[2]の手法は VGG-16 のいくつかの中間層から同じ空間座標に対応する出力を取り出し、それらを組み合わせることで画素の色を推定する手法である。これらの従来手法では、入力された画像にどのような物体が画像中に存在しているかの認識は行わず、各入力画像に対するテクスチャやエッジなどを用いて着色を行っていた。そのため、同様のテクスチャやエッジを持った異なる物体を、同一の色で着色するという問題点があった。

3. 提案手法

本稿では、物体認識用 VGG の特徴マップを U-Net に結合した自動着色ネットワークを提案する。

3.1 概要

提案手法において着色されるグレースケール画像は、U-Net および VGG-16 の二つに入力される。VGG-16 は、入力画像にどのような物体が存在するかを認識し、その結果を U-Net に伝達する。U-Net は VGG-16 から伝達された認識結果と、入力されたグレースケール画像を用いて着色を行う。これにより、画像内の物体の種類を考慮した自動着色を行う。

3.2 提案手法のネットワーク構成

図 2 に提案モデルである、VGG-16 と U-Net を組み合わせた構造を示す。VGG-16 は、1,000 クラス分類が可能な学習済みネットワークであり、入力された画像を畳み込んでいくにつれて、その画像がどのクラスに分類されるかを認識

[†] 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

[‡] 大阪大学大学院工学研究科

Graduate School of Engineering, Osaka University

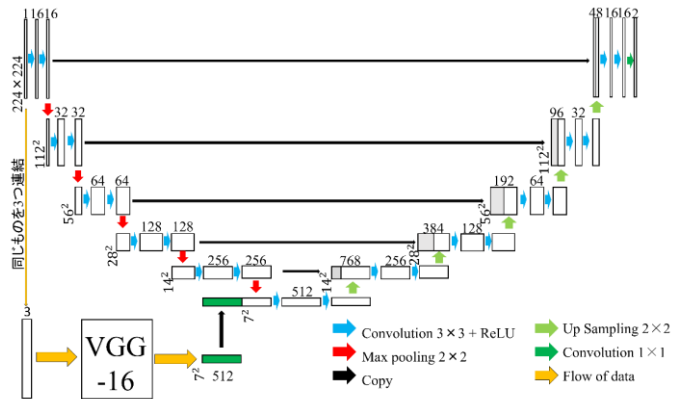


図 2 VGG-16 と U-Net を組み合わせた構造



図 1 VGG-16 に入力する偽画像

するための特徴マップが形成されていく。また、U-Net は Encoder-Decoder で構成されているため、U-Net における最深部では、入力画像の重要な特徴が抽出された特徴マップが形成されている。この両者の特徴マップを U-Net の最深部で結合し、Decoder への入力として用いる。これにより、物体のクラス・テクスチャの両方を用いて学習を進めることができるため、より精度の高い着色が期待される。

4. 評価実験

従来手法[1],[2]と提案手法の比較を行った。従来手法[1]と提案手法では学習画像データセットに Places365 の学習画像 1,803,460 枚、評価画像に Places365 の評価画像 328,500 枚を用いた。学習回数は 180 万回。評価指標として、平均絶対誤差(MAE)、平均平方二乗誤差(RMSE)、ピーク信号対雑音比(PSNR)、pixel 毎の着色精度、主観的評価を用いた。なお、従来手法[2]では学習画像データセットに ImageNet の画像を用いているが、参考値として文献中の値を参照した。

表 1 各種指標による評価結果

手法	MAE	RMSE	PSNR	ACC2%	ACC5%
Kamyar ら	7.5	N/A	N/A	18.3	47.3
Gustav ら	N/A	0.299	24.45	N/A	N/A
提案手法	6.9	0.251	31.58	38.1	74.4

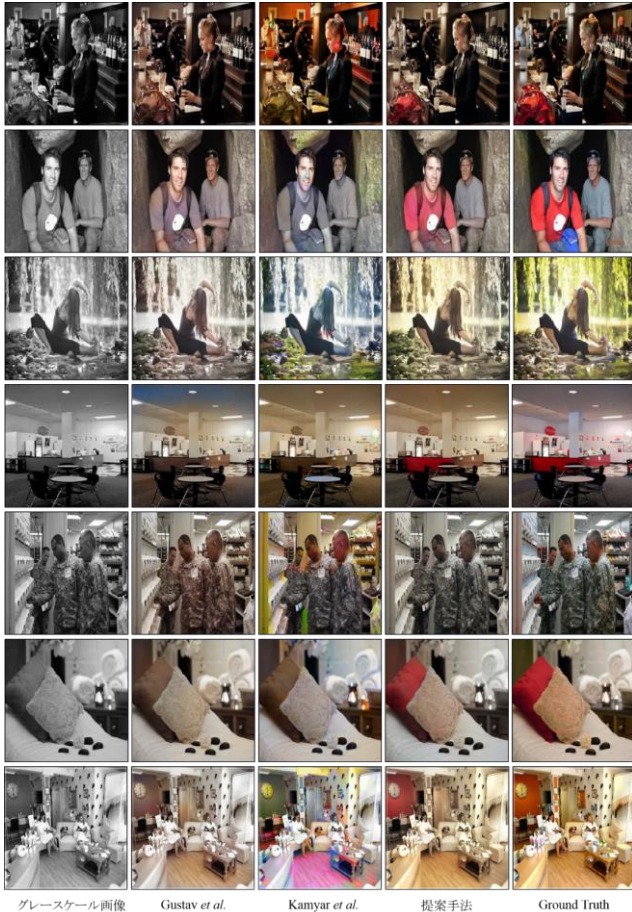


図 3 提案手法による着色結果

2 つ目の実験では VGG-16 の認識結果が着色に及ぼす影響を調べる. U-Net には実際に着色する画像を入力し, VGG-16 には U-Net に入力する画像とは異なる画像を入力し, その着色結果を比較する. 入力する画像は図 1 に示す 8 枚の画像を用いた.

4.1 評価実験

評価指標ごとの結果を表 1 に示す. 評価実験の結果, 提案手法では, 従来手法である Kamyar ら[1], Gustav ら[2]と比較して, MAE では 0.6 pt 減少, RMSE では 0.048 pt 減少, PSNR では 7.13 dB 上昇, pixel 毎の着色精度では, $\epsilon = 2\%$ としたときは 19.8 pt 上昇, $\epsilon = 5\%$ としたときは 27.1 pt 上昇した. このようにすべての指標について提案手法が良好な結果を示した.

4.2 主観的評価

図 3 に提案手法による着色結果を示した. 建物の壁や細かな物体に対しても着色できており, 使われている色の数も多く存在していた. また着色の際に様々な色が考えられ



図 4 物体認識が着色に及ぼす影響

るものに対しては, 赤褐色系の色を多く塗る傾向にあることが分かった. これは学習画像に出現する色の頻度と関係があると考えられる.

4.3 物体認識が着色に及ぼす影響の評価実験

U-Net への入力画像を固定し, VGG-16 への入力画像を変化させた場合の出力結果を図 4 に示した. 偽画像 ①, ② のような花の画像を入力すると全体的に緑・赤で着色を行った. これは花を入力したことにより, U-Net は着色の際にエッジ部分を草や葉っぱと認識したためであると考えられる. また, 偽画像 ③, ④, ⑤ では画像中に人間が写っているため評価画像中の人間部分に, 適切な着色がなされていた. しかし, 背景部分に関しては入力された偽画像によって異なる着色が行われていたため, U-Net は VGG-16 の出力を参考に行っていることが確認できた. さらに偽画像 ⑥, ⑦, ⑧ を入力した際は, それらが自然画像のため, 評価画像の細かなエッジを草・葉っぱと認識し, 緑に着色していることが確認できた. また, 自然画像中に多く存在する空に関してはうまく着色できていた. このような結果から, 提案手法のモデルは VGG-16 の認識結果を大きな参考にして着色を行っていると考えられる.

5. まとめ

本研究では, 高精度な自動着色手法の実現を目的とし, VGG-16 と U-Net を組み合わせたグレースケール画像の自動着色を提案した. 評価実験の結果, 従来の Gustav ら, Kamyar らの手法と比較して, 提案手法は, MAE が 0.6 pt 減少, RMSE が 0.048 pt 減少, PSNR が 7.13 dB 上昇, pixel 毎の着色精度が, $\epsilon = 2\%$ としたときは 19.8 pt 上昇, $\epsilon = 5\%$ としたときは 27.1 pt 上昇した. 今後は赤褐色系以外にも可能性の高い色が複数存在するような物体に対して, 色の選択方法を検討する必要がある.

参考文献

- [1] KamyarNazeri, EricNg, and MehranEbrahimi, "Image Colorization using Generative Adversarial Networks", IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2018.
- [2] GustavLarsson,MichaelMaire,GregoryShakhnarovich, "Learning Representations for Automatic Colorization", European Conference on Computer Vision (ECCV),2016