

GAN によるグレースケール画像のカラー化

研究者:阿多 健史郎 指導教員:濱川 恭央

1. 序論

カメラが発明されたばかりの写真は全て白黒であり、現代のカラー写真を見慣れている私達は、それを見ても情景がすぐさま思い浮かばない。カラー写真を一目見れば、その場所の景観だけではなく、雰囲気までも感じることができる。そこで本稿では、グレースケール画像を現代写真のようにカラー化することで情報量を増やし、過去の景観を可能な限り再現する。風景のみならず建築物や車などの人工物までカラー化することは、従来のネットワークでは困難である。また、従来手法と位置付けるベースラインネットワークは、一度の学習で、画像を生成するシステムである。しかし GAN は、一度の学習で、ベースラインのような画像生成に加え、その画像の精度を別のネットワークで判別するため、より高精度で画像を生成可能となる。このため本稿では Goodfellow 氏ら[1]の提案した敵対的生成ネットワーク (Generative Adversarial Networks, 以下 GAN)を用いて自動カラー化を行う。

2. Generative Adversarial Networks

GAN は、日本語では敵対的生成ネットワークといい、新しいタイプの生成モデルである。GAN は、Generator と Discriminator の 2 つのニューラルネットワークから成っており、それぞれ異なる役割を持つ。図 1 に GAN の概要図を示す。

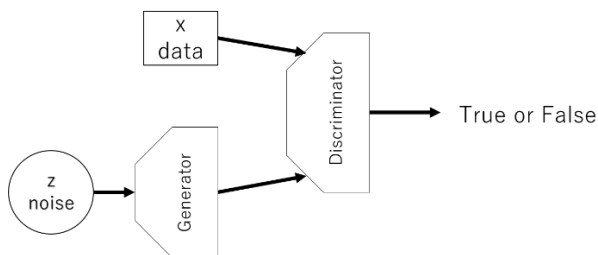


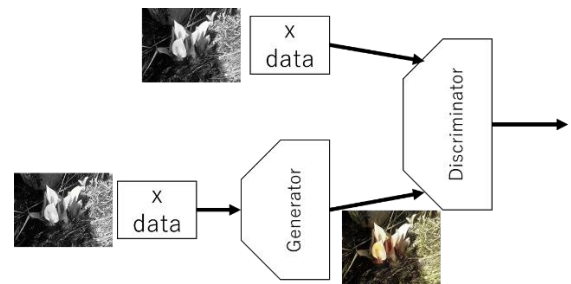
図 1: GAN の概要図

まず図 1 にて左側の六角形ブロックで表されている Generator の役割は、画像を生成することである。通常の GAN では 100 次元程度のノイズから画像を生成する。次に図 1 にて右側の六角形ブロックで表されている Discriminator の役割は、入力の画像が元のデータ分布から与えられているのか、それとも Generator により生成された画像なのか判別することである。最終的には、Generator は、Discriminator が本物と間違えるような画像を生成することを目指して学習し、Discriminator

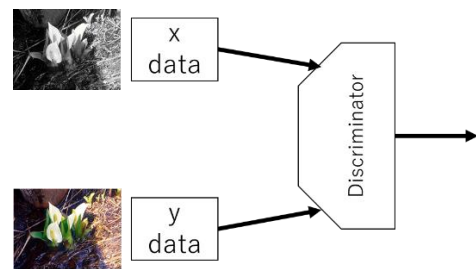
は Generator の画像を間違えることなく判別することを目指して学習する。

2.1 Conditional GAN

従来の GAN では、Generator の入力は無ランダムなノイズデータである。しかし本稿はカラー画像の生成を目的とするので、入力をグレースケール画像とする。このため、Conditional GAN と呼ばれるモデルを採用する。これによりカラー化などの画像変換タスクに対応可能となる。図 2 に Conditional GAN の概要図を示す。



(a) Generator の結果を判断する場合



(b) 正解画像を判断する場合

図 2: Conditional GAN の概要図

図 2 より Discriminator は、グレースケール画像を条件として、Generator の生成画像及びオリジナルデータの正解画像の 2 種類のカラー画像を入力とし、どのペアにオリジナルデータの正解画像が含まれているかを判別する。前者が図 2(a)であり、後者が図 2(b)である。

3. シミュレーション

3.1 ネットワークモデル

Generator 及び Discriminator で畳み込みニューラルネットワーク (Convolutional Neural Network, 以下 CNN) を使用する。ベースラインネットワークと GAN の Generator 及び Discriminator のアーキテクチャを表 1 に示す。GAN で用いられている手法は[1]で提唱され、GAN の学習精度を高める。なお、ベースラインには Discriminator がなく、生成ネットワークのみである。

表 1: ネットワークのアーキテクチャ

	活性化関数 (隠れ層/ 出力層)	バッチ 正規化	ドロップ アウト
ベースライン	ReLU/ Sigmoid	なし	なし
Generator	LeakyReLU/ Tanh	あり	あり
Discriminator	LeakyReLU/ Softmax	あり	なし

3.2 学習

学習には公開データセットの CIFAR-100 を使用する。訓練データ 5 万枚及びテストデータ 1 万枚を含むデータセットであり、100 クラスに分類される画像を含む。また、画像サイズは 32×32 であるので、ネットワークもそれに応じた形状となっている。学習エポック数は 100 である。

4. 評価

4.1 SSIM での評価

学習結果の画像の精度を評価するため、SSIM を用いる。

SSIM とは、Structural Similarity の略で、画像の類似度を示す画像評価式である。式(1)に、数式を示す。なお、SSIM は画像内の局所領域毎に算出され、 μ_x と μ_y は領域内の画素値の平均を表し、 σ_x と σ_y は画素値の標準偏差、 σ_{xy} は画素値の共分散を表す。また、 c_1 及び c_2 は分母の値が非常に小さくなった際に評価値が不定とならないための定数である。

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

SSIM は、画素値、コントラスト、構造の 3 点の変化を捉える。これにより、PSNR や MSE などの画像評価式よりも主観的で人間の感覚に近い値を算出する性質を持つ。値域は $[0, 1]$ である。また、結果の表にはテストデータでの測定値の平均を示す。

表 2 にベースラインネットワークとの比較結果を示し、生成画像を図 3 に示す。また、表 3 に過去研究である、CNN を用いたカラー化[2]及び HSV 空間におけるカラー化[3]との比較結果を示す。

表 2: ベースラインネットワークとの比較結果

	ベースライン	GAN
SSIM	0.7070	0.9084

表 3: 過去研究との比較結果

	CNN	HSV 空間	GAN
SSIM	0.8430	0.8500	0.9084

4.2 彩度での評価

生成画像の鮮やかさを比較するために、彩度の比較を行う。比較結果を表 4 に示す。

表 4: 彩度の比較結果

	オリジナル	ベースライン	GAN
SSIM	0.3216	0.1952	0.2561



(a) 結果 1



(b) 結果 2

図 3: カラー化結果

各画像の上段: オリジナル画像

各画像の中段: ベースラインの生成画像

各画像の下段: GAN の生成画像

5. 結論

GAN を用いたグレースケール画像の自動カラー化を高精度で行うことができた。SSIM 及び彩度ともに、GAN で生成したカラー画像がベースラインネットワークで生成したカラー画像を上回った。

今後の課題として、自然物以外のカラー化においては異なる色付けをしている場合も見られたので、大規模なデータセットで学習を行う、新たな手法を提案するなどが挙げられる。

参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. 2014
- [2] 池田伊織, CNN を用いた風景画像の自動カラー化, 鹿児島工業高等専門学校, 2017
- [3] 下吉賢信, HSV 空間における白黒画像のカラー化, 鹿児島工業高等専門学校, 2017