

106. 深層学習の適用方法 (スタイル変換)

秋葉洋哉

2024 年 7 月 15 日

1 Pix2pix

1.1 概要

Pix2pix は、画像変換のためのモデルであり、画像のペアを入力として、画像の変換を行う。画像のペアとは、入力画像と出力画像の組み合わせであり、例えば、セグメンテーションマップと写真、白黒画像とカラー画像などがある。Pix2pix は、CGAN(Conditional Generative Adversarial Networks) を用いて学習を行う。

CGAN とは、あるクラスを与えた際に、そのクラスに対応する画像を生成する GAN である。Pix2pix は、その条件の代わりに画像を与えることで、何らかの変換を施した画像を出力するモデルである。Pix2pix ではこの画像の変換方法を学習していく。

各プレイヤーの役割は以下の通りである。

- Generator : 条件画像 x を基にある画像 $G(x,y)$ を生成する
- Discriminator : 条件画像 x を $G(x,y)$ に変換したものと、条件画像 x に真の変換が施された画像 y を見分ける

例えば、条件画像 x に白黒の画像に対して、真の変換後画像 y に着色した画像を与えたり、条件画像 x に航空写真の画像を与えて、真の変換後画像 y に道だけを抽出した画像を与えることで、学習を行う。

1.2 CGAN からの工夫 1: U-Net

Pix2pix は、Generator に U-Net を用いている。U-Net は、セマンティックセグメンテーションにおいて、ピクセル単位でクラス分類するネットワークである。U-Net は、Encoder-Decoder 構造を持ち、Encoder では画像の特徴を抽出し、Decoder では同じ階層の Encoder の特徴マップを元に画像を生成する。この構造により、画像の特徴を保持しつつ、画像を生成することができる。

Pix2pix は、U-Net の画像の位置関係を保持する構造を用いることで、元画像の特徴を保持しつつ、変換を行うことができる。

1.3 CGAN からの工夫 2: L1 正則化項の追加

Pix2pix は、Generator の損失関数に L1 正則化項を追加している。通常の GAN と異なり、Pix2pix では画像の変換方法を学習するため、条件画像と生成画像に視覚的一致性があることが重要である。このため、画像

の高周波成分 (色の変化が顕著な部分) を学習し、Generator が生成した画像がぼやけることを防ぐため、L1 正則化項を追加している。

1.4 CGAN からの工夫 3: PatchGAN

Pix2pix は、Discriminator に PatchGAN を用いている。PatchGAN は、画像全体ではなく、画像の一部 (パッチ) に対して識別を行う GAN である。この手法を用いて、各パッチ毎に Pix2pix を適用させることで、高周波成分を残す効果を高めることができる。