

1. はじめに

深層学習による高精度な物体識別は、大量の画像と教師データを必要とするが、そのアノテーションにかかる人的なコストは高い。そのため、CG 画像で代替する場合、実画像との差異があるため高精度な識別は期待できない。そこで、Generative Adversarial Networks(GAN) により、CG と実画像間のスタイルを変換することで差異を小さくすることが期待できる。しかし、スタイル変換が識別対象に特化しているとは言えない。そこで、本研究では、CycleGAN[1] に Attention 機構を導入することで識別対象に注視したスタイル変換手法を提案する。

2. Generative Adversarial Networks

GAN は、ノイズベクトルから画像を生成する Generator と、生成画像または実画像を入力し、どちらが入力されたかを識別する Discriminator を敵対的に学習することで、データセットに含まれない画像が生成可能である。CycleGAN は、異なる 2 つのドメインの画像を相互に変換することで、スタイル変換時に入力画像の特徴を保持した変換ができる。しかし、CycleGAN では色や形状が異なる複数の物体を単一ネットワークでスタイル変換することができないという問題がある。

3. 提案手法

本研究では、識別対象の注視領域を捉える Attention 機構を導入した CycleGAN による識別対象に注視した画像のスタイル変換手法を提案する。

3.1. ネットワーク構造

提案手法では、Cycle Consistent Adversarial Domain Adaptation(CyCADA)[2] の Semantic consistency loss を導入する。また、クラス分類時の注視領域をドメイン識別時に反映する Attention 機構への誤差として Attention consistency loss を導入する。提案手法のネットワーク構造を図 1 に示す。ここで、 x が CG 画像、 y が実画像、 \hat{x} が CG への変換画像、 \hat{y} が実画像への変換画像である。

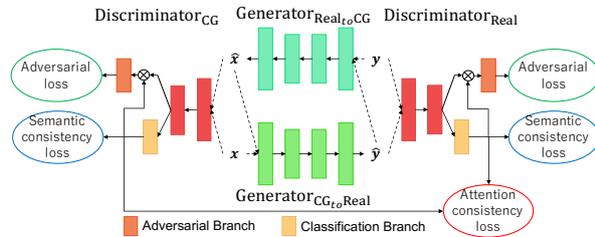


図 1: 提案手法のネットワーク構造

3.2. Attention consistency loss

Semantic consistency loss のみでは、クラス分類した際の判断根拠が異なる可能性がある。そこで、Attention consistency loss により注視領域をスタイル変換前後で拘束する。これにより、変換する領域以外の特徴を生成しないように学習することが可能である。Attention consistency loss (\mathcal{L}_{att}) を式 (1) に示す。ここで、 $D_{att}(\cdot)$ が Classification Branch における注視領域である。

$$\mathcal{L}_{att} = \frac{\|D_{att}(\hat{y}) - D_{att}(x)\|_1 + \|D_{att}(\hat{x}) - D_{att}(y)\|_1}{2} \quad (1)$$

4. 評価実験

提案手法で変換した画像の有効性を、識別学習により評価する。

4.1. 実験概要

本実験では、データセットに 10 種類の数字で構成された SVHN と SynthDigits を使用する。SVHN は実環境の画像 99,289 枚、SynthDigits は仮想環境の画像 488,953 枚

で構成されている。各データセットの一部を図 2 に示す。学習に使用するデータ数は、実画像を 26,032 枚、CG 画像を 479,400 枚とする。評価画像には学習に使用していない実画像すべてを用いる。評価方法は、CG 画像で事前学習した識別器へ実画像と変換画像を入力し、識別精度を比較する。



(a) SVHN (b) SynthDigits

図 2: データセットの一部

4.2. 実験結果

図 3 に実画像と提案手法で変換した画像の一部を示す。単一の Generator で Semantic な情報を欠損することなく変換できていることがわかる。



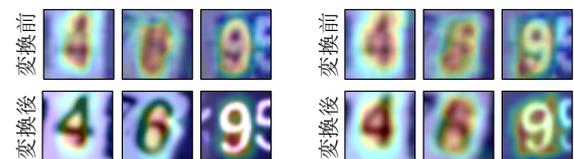
図 3: 各画像の比較

識別率を表 1 に示す。少量の実画像で学習した場合と比較して、CyCADA を用いて変換した場合は識別率が 1.80 ポイント低下した。この精度低下を解消するために、Attention 機構を導入することで識別率が 0.79 ポイント向上した。

表 1: 識別精度の比較

| 学習画像 | 評価画像 | 識別率 [%] |
|-------|------------------|---------|
| 実画像 | 実画像 | 80.95 |
| | 実画像 | 77.48 |
| CG 画像 | 変換画像 (CyCADA[2]) | 79.15 |
| | 変換画像 (提案手法) | 81.74 |

Attention 機構および Attention consistency loss の有無による注視領域の変化を図 4 に示す。Attention 機構の導入により、注視領域の縮小等の変化を抑制し識別対象に注視した変換することができた。以上より、提案手法を用いることで識別対象に効果的なスタイル変換が可能である。



(a) CyCADA (b) 提案手法

図 4: 提案手法による注視領域の変化

5. おわりに

Attention 機構を導入した CycleGAN による識別対象に注視した画像のスタイル変換法を提案し、精度比較にて 0.79%の精度向上を確認した。今後は、変換対象を特徴量ベースにすることで画像への復元を省略することによる識別精度の高精度化をする。

参考文献

- [1] J. Zhu, *et al.*, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, ICCV, 2017.
- [2] J. Hoffman, *et al.*, “CyCADA: Cycle-Consistent Adversarial Domain Adaptation”, ICML, 2018.