

1. はじめに

CycleGAN[1] は、ペア画像を必要としない GAN によるスタイル変換手法である。CycleGAN は 2 つのドメイン間における背景や物体の数が異なると、スタイル変換の際に背景の形状変化や物体の消失/出現が発生する問題がある。本研究では、この問題を解決するために、物体の大きさに合わせた複数の特徴マップサイズで損失計算が可能な Multi-Scale CycleGAN (MS-CycleGAN) を提案する。

2. CycleGAN

CycleGAN は、再変換画像と元画像の一貫性を保証する Cycle consistency loss の導入により、ペア画像なしでスタイル変換が可能な手法である。CycleGAN のネットワーク構造を図 1 に示す。CycleGAN は 2 つのドメイン間でスタイルを入れ替えるため、各ドメインに対する Generator, Discriminator が必要となる。Discriminator は画像を入力し、入力画像が本物である確率を出力する。Generator は一方のドメインの画像を入力し、他方のドメインにスタイルを変換する。CycleGAN は、背景の形状変化や物体の消失/出現が発生するという問題がある。また、同じドメイン内の画像でも物体サイズや物体数などがシーンにより異なるため、変換難易度が変化するという問題がある。

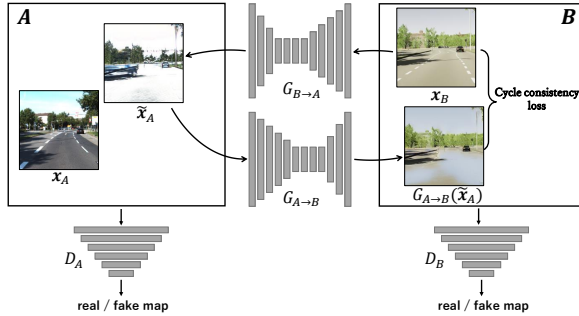


図 1: CycleGAN のネットワーク構造

3. 提案手法

従来の Discriminator は任意の特徴マップサイズのみで真贋判定するため、物体の大きさによる特徴の違いが考慮できていない。そこで、本研究では物体の大きさに合わせた複数の特徴マップサイズで損失計算が可能な Multi-Scale CycleGAN (MS-CycleGAN) を提案する。

3.1. MS-CycleGAN

MS-CycleGAN のネットワーク構造を図 2 に示す。Discriminator は Decoder を設けることでマルチスケールの真贋判定が可能となる。一方、Generator は CycleGAN と同様の構造を使用する。提案手法は、Discriminator の各出力層で得た特徴を学習することで、物体サイズの違いなど変換難易度を考慮した学習が期待できる。

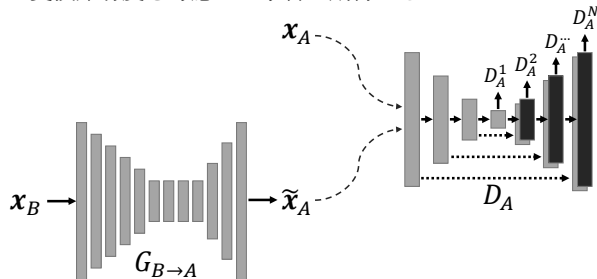


図 2: 提案手法のネットワーク構造

3.2. 敵対的損失の設計

CycleGAN は学習初期にテクスチャの変化が発生し、学習後期に形状変化が発生する傾向から、本研究では Generator と Discriminator それぞれの収束速度に着目した 2 つの敵対的損失を設計をする。式 (1) は Generator の学習

を早く収束させるため、Generator が物体の形状変化に鈍感となることを期待する。式 (2) は Discriminator の学習を早く収束させるため、Generator がテクスチャの変化のみで Discriminator を騙せるようになることを期待する。

$$\begin{aligned}\mathcal{L}_{G_{adv}} &= \mathbb{E}_{\mathbf{x}_B} \left[D_n^A(\tilde{\mathbf{x}}_A)^2 \right] \\ \mathcal{L}_{D_{adv}} &= \mathbb{E}_{\mathbf{x}_A} \left[(D_n^A(\mathbf{x}_A) - 1)^2 \right] + \mathbb{E}_{\mathbf{x}_B} \left[D_n^A(\tilde{\mathbf{x}}_A)^2 \right] \\ \mathcal{L}_{adv} &= \arg \min_{n \in N} \mathcal{L}_{G_{adv}} + \arg \max_{n \in N} \mathcal{L}_{D_{adv}}\end{aligned}\quad (1)$$

$$\mathcal{L}_{adv} = \arg \max_{n \in N} \mathcal{L}_{G_{adv}} + \arg \min_{n \in N} \mathcal{L}_{D_{adv}} \quad (2)$$

ここで、 N は Discriminator の出力層数、 $\mathbf{x}_A, \mathbf{x}_B$ は各ドメインの実画像、 $\tilde{\mathbf{x}}_A := G_{B \to A}(\mathbf{x}_B)$, $\tilde{\mathbf{x}}_B := G_{A \to B}(\mathbf{x}_A)$ である。

4. 評価実験

本実験では、提案手法の優位性を示すために、スタイル変換画像の定性的評価と定量的評価を行う。

4.1. 実験概要

データセットは、実世界画像として KITTI, CG 画像として Virtual KITTI 2 を使用する。実世界画像と CG 画像はそれぞれ 1,893 枚を使用し、バッチサイズを 1 として 200 エポック学習する。

4.2. 実験結果

スタイル変換結果を図 3 に示す。提案手法+式 (1) は、提案手法が従来の CycleGAN に比べてマルチスケール化により、物体の形状変化を抑制できた。提案手法+式 (2) は、変換に失敗した。Generator は Discriminator の学習に依存しているため、Discriminator の学習が早く収束する損失設計は提案手法の学習が不可能になることが確認できた。

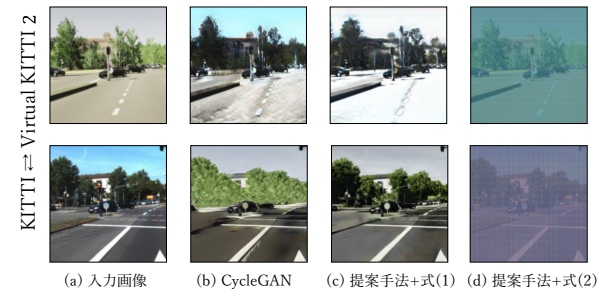


図 3: KITTI と Virtual KITTI 2 のスタイル変換結果

2 つの画像間の類似度を示す指標として LPIPS を用いた定量的評価を表 1 に示す。変換無しの場合と比較すると、CycleGAN が最も良く、次に提案手法+式 (1) が良いことが確認できた。提案手法+式 (2) は変換に失敗したため、変換無しよりも悪くなった。

表 1: LPIPS

変換無し	CycleGAN	提案手法	
		式 (1)	式 (2)
0.685	0.618	0.628	0.916

5. おわりに

本研究では、Multi-Scale CycleGAN (MS-CycleGAN) および損失関数を提案した。提案手法は、ある程度の物体の形状変化を抑制することが可能だが、ドメインが大幅に異なるデータには効果が低いことが分かった。今後は、ドメインの差異を埋める学習データの入力方法を検討する。

参考文献

- [1] J. Zhu, et al., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, ICCV, 2017.