

1. はじめに

深層学習によるスタイル変換は、入力画像が持つスタイルを異なるスタイルへ変換するアプローチである。スタイル変換を用いて異なるドメインの画像に変換することで、物体認識のドメイン適応が可能となる。スタイル変換を用いたドメイン適応の代表的な手法である CyCADA[3] は、画像のクラスが一貫した変換は可能であるが、入力画像に対して生成できる変換画像は1枚である。

本研究では、スタイル変換の際にクラス情報を考慮して複数の変換画像を生成できるスタイル変換手法を提案する。さらに、画像の認識時にデータ拡張を行うテストタイム拡張に提案手法を用いることで高精度化が期待できる。提案手法は、クラス情報を考慮しながら様々なスタイル変換が可能となり、高いアンサンブル効果を実現できる。

2. 従来研究

2.1 スタイル変換

スタイル変換は、異なるスタイルを持つ画像間において、一方もしくは双方向に画像変換するアプローチである。スタイル変換の代表的な手法として、CycleGAN[1]がある。CycleGANは、変換先のペア画像を必要としない変換手法であり、2つのスタイル間で双方向に画像変換できる。また、3つ以上のスタイルを変換する手法として、StarGAN v2[2]がある。StarGAN v2は入力画像とは異なる画像、もしくはガウス分布に基づく潜在変数からスタイルコードを生成し、スタイルコードに埋め込まれたスタイルになるように入力画像を変換する。

2.2 ドメイン適応

ドメイン適応は、教師ラベルを持たないターゲットドメインと、教師ラベルを持つソースドメインの異なるドメイン間のギャップを軽減するように学習することで、ターゲットドメインでの識別精度を向上させる技術である。スタイル変換によるドメイン適応として Cycle-Consistent Adversarial Domain Adaptation(CyCADA)[3]が提案されている。CyCADAは、CycleGANにクラス情報を考慮する Semantic consistency を導入している。これにより、ソースドメインで事前学習した識別器を用いて、変換前後の画像におけるクラス情報が一致するように学習できる。

3. 提案手法

本研究では、StarGAN v2での画像変換時にクラス識別を行う補助タスクを追加することにより、スタイル変換時にクラス情報を考慮する。補助タスクには、CyCADAの Semantic consistency を用いる。これにより、本手法はクラス情報を考慮しながら複数の画像へ変換ができる。本手法を評価時にデータ拡張するテストタイム拡張に適用することで、高精度な識別が可能となる。

3.1 クラス情報を考慮した StarGAN v2

提案手法の構造を図1に示す。提案手法は、ソース画像 x_s とターゲット画像 x_t に対して以下の処理を順に行う。

Step1 潜在変数 z より生成されるスタイルコード \hat{s}_t を用いたスタイル変換。

Step2 実画像 x_s と変換画像 \tilde{x}_t の真贋判定。

Step3 実画像のスタイルコード \hat{s}_s を用いた画像 \tilde{x}_t の再構成。

Step4 実画像 x_s と変換画像 \tilde{x}_t のクラス識別。

ここで、GeneratorをEncoder-Decoderモデルとしたとき、入力画像が持つコンテキストをEncoderで抽出、Decoderでスタイル情報の付与を行うため、Generatorの中間特徴はドメイン不変の特徴量となる。そのため、Generatorの中間にクラス識別器を追加することにより、両方のドメインにおいて十分な識別性能が期待できる。また、追加したクラス分類器は画像変換時には用いない。

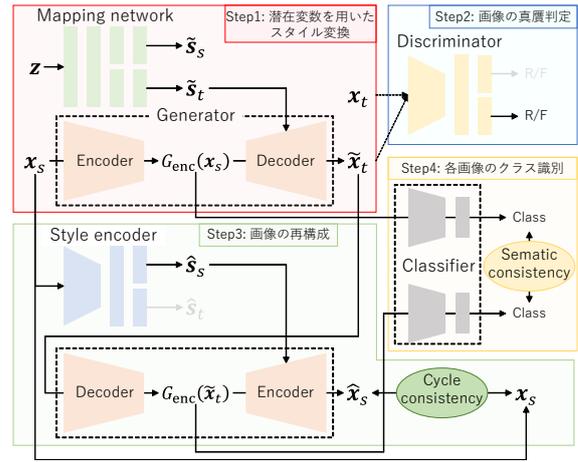


図1：提案手法の構造

3.2 誤差関数

誤差関数は、クラス識別器を学習する \mathcal{L}_{task} と、スタイル変換におけるクラス情報の一貫性を考慮する \mathcal{L}_{sem} を用いる。各誤差関数は式(1)、式(2)と定義する。

$$\mathcal{L}_{task} = CE(C(G_{enc}(x_s)), t_s) \quad (1)$$

$$\mathcal{L}_{sem} = CE(C(G_{enc}(G(x_s, s_t))), C(G_{enc}(x_s))) + CE(C(G_{enc}(G(x_t, s_s))), C(G_{enc}(x_t))) \quad (2)$$

ここで、 x が入力画像、 t_s がソースドメインの教師ラベル、 C がクラス分類器、 G が Generator、 s がスタイルコード、 CE がクロスエントロピー誤差である。また、添え字は各ドメインを表しており、 s がソースドメイン、 t がターゲットドメインを表している。 \mathcal{L}_{task} は、教師ラベルを持つソースドメインの画像のみで誤差を計算する。 \mathcal{L}_{sem} は、クラス分類器より出力される変換前のクラス確率を仮ラベルとし、変換後のクラス確率とのクロスエントロピー誤差を計算する。これにより、画像の変換前後でクラス情報の一貫性を考慮した学習が可能となる。

3.3 テストタイム拡張

テストタイム拡張は、識別時テストサンプルに対してデータ拡張を行い、各画像に対するクラス確率を平均して最終的な予測結果を決定する。本手法では、ソースドメインで事前学習した識別器へターゲットドメインの画像を入力する際に、スタイル変換によるドメイン適応を行う。同時に、潜在変数を複数回サンプリングし、様々なスタイルの画像へ変換する。そして、従来と同様に式(3)で定義するアンサンブル推論にて最終的な予測結果を決定する。

$$p_i = \frac{1}{N} \sum_{n=1}^N \sigma_i(f(G(x, F(z_n)))) \quad (3)$$

ここで、 i がクラス番号、 p が予測結果、 N が増幅数、 σ が Softmax 関数、 f が事前学習された識別器、 F が潜在変数よりスタイルコードを生成する Mapping network、 z が潜在変数である。

4. 評価実験

提案手法の有効性を示すために、変換画像の傾向調査とテストタイム拡張における識別性能の比較、および変換画像の分布調査を行う。評価実験に使用するデータセットは、数字認識用のデータセットである SVHN と SynthDigits を用いる。ここで、ソースドメインには、CG 画像である SynthDigits、ターゲットドメインは実画像である SVHN とし、クラス数は10である。また、テストタイム拡張における増幅数は10とする。

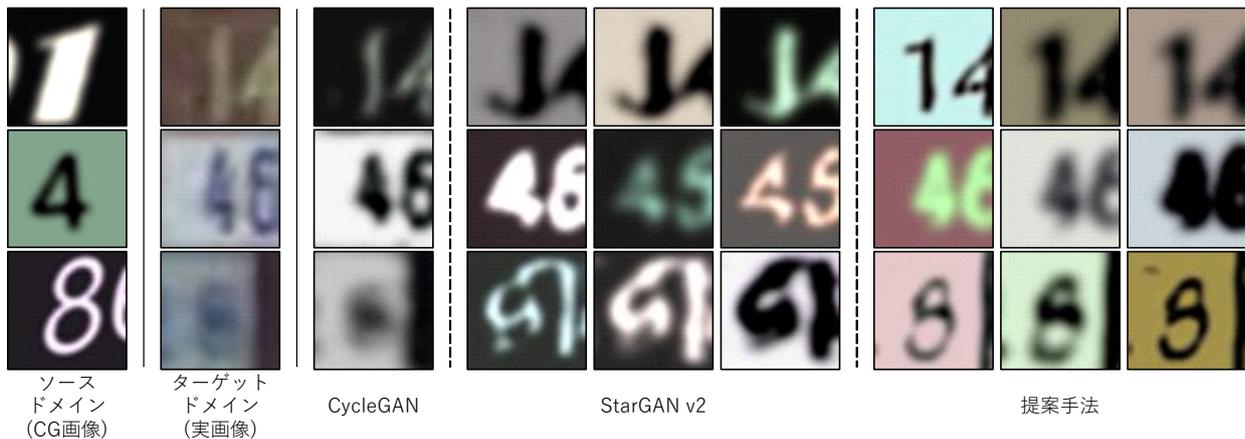


図 2 : 各手法による変換結果

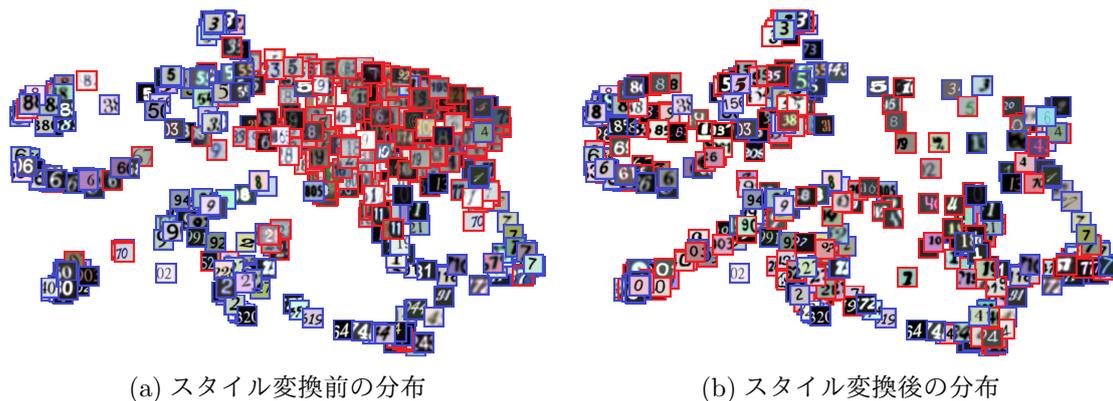


図 3 : UMAP による分布の可視化

表 1 : テストタイム拡張における識別精度

変換手法	増幅数	識別率 [%]	
		ResNet-20 (701.4k)	CNN (24.3k)
変換無し	-	74.88	66.54
ベースライン (StarGAN v2)	1	77.48	76.44
	10	79.85	79.50
提案手法	1	87.73	85.77
	10	89.21	87.91

4.1 変換画像の定性的評価

提案手法および従来手法 (CycleGAN, StarGAN v2) を用いて変換した結果を図 2 に示す。これより、従来手法ではクラス情報の維持が困難であり、変換元であるターゲットドメインの画像が低画質であった場合はその傾向が顕著である。一方、提案手法による変換は、クラス情報を維持した状態で異なるスタイルの画像へ変換できていることが分かる。

4.2 テストタイム拡張を用いた定量的評価

テストタイム拡張に使用する事前学習済み識別器には、ResNet-20 と、ドメイン間の差異を吸収することが困難な 5 層の CNN を用いて評価実験を行う。テストタイム拡張による識別精度を表 1 に示す。ここで、各ネットワークのパラメータ数はモデル名の下に示す。表 1 より、ResNet-20 の結果において、提案手法は変換無しと比較して識別精度の向上が確認できる。また、増幅数 1 では、提案手法による識別精度はベースラインの StarGAN v2 と比較し、10.25pt の精度向上を確認した。さらに、増幅数 10 では、増幅数 1 と比較して 1.48pt の精度向上を確認した。この傾向は CNN の識別結果においても同様である。

4.3 UMAP による変換画像の分布の調査

各ドメインの画像と提案手法で変換した画像を UMAP で次元圧縮し、可視化した結果を図 3 に示す。ここで、青枠がソースドメイン、赤枠がターゲットドメインを表している。次元圧縮に使用する画像枚数は、各クラスからランダムに 25 枚選択し、各ドメイン 250 枚とする。また、圧縮する特徴には定量的評価で使用した CNN の中間特徴を用いる。図 3(a) はスタイル変換を行っていないため、ドメインごとの分布が異なっている。一方、スタイル変換を行った図 3(b) は各クラスの分布がほぼ一致している。

5. おわりに

本研究では、クラス情報を考慮したスタイル変換手法および、テストタイム拡張におけるスタイル変換を用いたデータ拡張を提案した。評価実験において、提案手法によるスタイル変換は、クラス情報を考慮した変換ができていることを確認した。また、テストタイム拡張による定量的評価では、変換無しと比較して、14.33pt の精度向上を確認した。今後は、セマンティックセグメンテーションなどの異なるタスクへの適応を検討する。

参考文献

- [1] J.Y.Zhu, *et al.*, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks” In ICCV, 2017.
- [2] Y.Choi, *et al.*, “StarGAN v2: Diverse Image Synthesis for Multiple Domains”, In CVPR, 2020.
- [3] J.Hoffman, *et al.*, “CyCADA: Cycle Consistent Adversarial Domain Adaptation”, In ICML, 2018.

研究業績

- [1] 今枝航 等, “Attention 機構を導入した CycleGAN による識別に有効なスタイル変換”, 画像の認識・理解シンポジウム, 2019.
- [2] 今枝航 等, “Generative Adversarial Networks を用いたからあげピッキングにおける前処理としての画像変換”, ビジョン技術の実利用ワークショップ, 2020.