

1. はじめに

深層学習における学習画像の増幅手段として、幾何変換などを行うデータオーグメンテーションがある。また、多様なデータを作成するために、他のドメインデータを増幅手段として利用するアプローチも考えられる。様々なドメインデータを用いて学習することで、認識モデルの汎化性能の向上が期待されるが、認識対象との相関が低いドメインデータを学習すると、対象ドメインの認識に特化したモデルを作ることが困難となる。そこで、ターゲットドメインデータを作成するために、スタイル変換を用いる方法が考えられる。スタイル変換は、認識対象外のドメインを認識対象ドメインの画像様式へ変換する手法である。これにより、認識対象のドメインデータを増幅できる。しかし、スタイル変換は、ドメイン間の差が大きい程、変換品質が低下する傾向にある。

本研究では、中間ドメインを作成し、step-wise にスタイル変換を行うことで、ターゲットドメインの認識に貢献するデータ増幅方法である Domain-mixup を提案する。提案手法は、ドメイン間で画像を合成する Domain-mixup により、中間ドメインを経由しながらスタイル変換を行う。これにより、ターゲットドメインデータが増幅され、高精度な認識を補助することができる。

2. スタイル変換

スタイル変換は、ソースドメインデータをターゲットドメインデータへスタイルを変換するアプローチである。中でも CycleGAN [1] は、注釈データを使わずに画像中に含まれるクラスや属性の崩壊を起こさずにターゲットドメインにスタイル変換できる。これにより、ターゲットドメインへスタイル変換を行い、認識性能を向上させるためのデータ拡張に利用できる。しかし、変換画像はターゲットドメインのスタイルを獲得しても、ターゲットドメインと相関の低い画像へ変換されるがあるため、必ずしも認識性能の向上に繋がらないという問題がある。

3. 提案手法

本研究は、中間ドメインを経由してターゲットドメインへスタイル変換を行う。中間ドメインは、データセットに存在しないため、擬似的に中間ドメインを作成する必要がある。中間ドメインを作成するために、シーン間でジオメトリが一致するシーンレジストレーションデータセットと、画像を合成するために考案した Domain-mixup により中間ドメインを作成する。その後、ソースドメインから中間ドメイン画像へ、中間ドメインからターゲットドメイン画像へと step-wise にスタイル変換を行う。本手法によりターゲットドメインのデータを生成することで、高品質なターゲットドメインデータへ増幅し、認識性能の向上を図る。

3.1 Domain-mixup

Domain-mixup により、ソースドメインとターゲットドメインの中間ドメインを容易に作成できる。Domain-mixup は、mixup [2] と同様の手法を用いて画像を合成する。mixup の式を以下に示す。

$$X = (1 - \lambda)X_1 + \lambda X_2 \quad (1)$$

X は、合成画像を表す。 X_1, X_2 がそれぞれソースドメイン画像とターゲットドメイン画像に相当する。 λ は、合成するドメインの割合を決定する係数を示し、 λ は $[0.0 - 1.0]$ の範囲とする。また、Domain-mixup による中間ドメインの生成例について画像の関係性を図 1 に示す。

3.2 step-wise スタイル変換

段階的に画像の変換を行うために、中間ドメインを経由してスタイル変換を行う。ターゲットドメインを拡張するために用いるデータセットを Source only, Target only, Mixup only とする。Source only 及び Target only は、用意したそれぞれのドメインに属した RGB 画像を表し、



図 1：Domain-mixup による中間ドメイン画像の作成例

Mixup only は、Domain-mixup によって合成された RGB 画像を示す。中間ドメインは、ソースドメインから Mixup only へスタイル変換を行った画像と、Mixup only を統合したドメインとする。ターゲットドメインは、中間ドメインを Target only へスタイル変換を行った画像と Target only を統合ドメインとしている。これらを用いた学習の手順について、図 2 に示す。

Step1. Source only と Mixup only を用いて、ソースドメインを Mixup only ドメインへ変換し、中間ドメインを作成する。

Step2. Step1. で作成した中間ドメインをターゲットドメインへスタイル変換を行い、ターゲットドメインを拡張する。

Step3. Step2. で拡張したターゲットドメインを用いて、タスクネットワークを学習する。

3.3 データセット

提案手法の有効性を確認するために、タスク別にデータセットを構築する。

シーンレジストレーションデータセット 本データセットは CG を用いて作成され、2 種類のドメインから構成される。1 つ目は、Epic Games 社によって開発されたゲームエンジン Unreal Engine 4(UE4) で生成したリアルタイムレンダリング画像である。もう一方は、NVIDIA 社が提供する物理ベースのレンダリングテクノロジー V-Ray によって生成したプリレンダリング画像である。UE4 ドメイン及び V-Ray ドメインは、フォトリアリスティックな画像を生成し、実画像環境を模している。V-Ray は、レイトレーシングによるプリレンダリングを行うため、UE4 より実世界に近いシャドウや環境光の表現を実現できている。そして、UE4 画像と V-Ray 画像は、シーン内のテクスチャやライト条件は異なっているが、シーンの位置が一致している。

ARC データセット クラス分類タスクを解くために、Amazon Robotics Challenge データセット (ARC データセット) を用いる。ARC データセットは、カメラで撮影された実画像を模倣して CG データへ再現した単体アイテムデータセットである。CG は 2 種類のレンダラで作成され、1 つは UE4、もう一方は、プリレンダリングエンジンの I-Ray で作成された画像を用意する。そして、アイテム領域からランダムに切り出し、パッチ画像を作成する。

4. 評価実験

提案手法の有効性について、セマンティックセグメンテーションおよびクラス分類を行い評価実験を行う。

4.1 セマンティックセグメンテーション

本実験では、ターゲットドメインに対する認識性能の評価を行う。

実験概要 Source only 及び Target only は、それぞれ UE4 と V-Ray とする。スタイル変換手法は、CyCADA [3] の Pixel-level Adaptation を用いる。Step1. 及び Step2. のスタイル変換の学習回数はそれぞれ 60epoch とする。Step3. におけるセマンティックセグメンテーションの損失関数には Softmax Cross Entropy を用いて 100epoch を学習回数とする。セグメンテーションモデルは、Dilated Residual

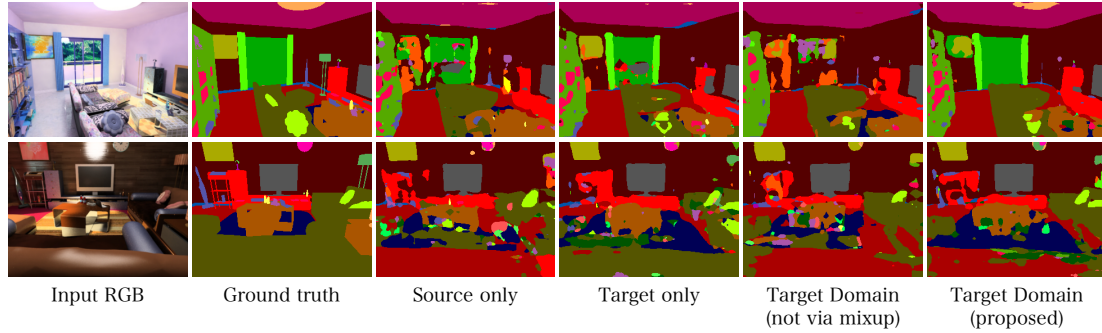
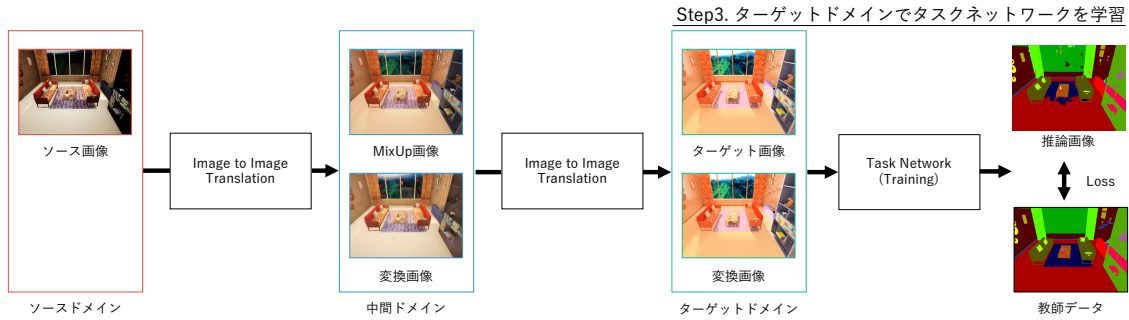


表 1: セマンティックセグメンテーションの評価結果

学習ドメイン	Pixel acc. [%]	Mean class acc. [%]	Mean IoU [%]
Source only(lower)	63.17	42.12	23.60
Target only	69.14	47.10	28.22
Target Domain(not via mixup)	66.62	48.00	28.81
Target Domain(proposed)	69.47	49.90	30.20

表 2: クラス分類の評価結果

学習ドメイン	Accuracy [%]
Source only(lower)	61.75
Target only	72.50
Real only	84.12
Target Domain	87.00

Networks を特徴抽出器とした DRN-C-26 [4] を用いて学習を行う。テストデータとして 500 枚用いて推論および評価を行う。評価指標には Pixel Accuracy, Mean Class Accuracy, Mean IoU を用いる。Pixel Accuracy はピクセルごとに推論値と教師ラベルの正誤を判定する。Mean Class Accuracy は、推論値をクラスごとに正誤判定を行い、正解領域に含まれる正推定数の割合の平均を求める。Mean IoU は、対象クラスの正解領域と誤推定した領域を考慮して正推定数の割合平均を求める。

実験結果 テストデータによる評価結果を表 1 に示す。提案手法によって生成した学習ドメインが Target Domain (proposed) である。提案手法との比較のために、中間ドメインを経由せず、ソースドメインからターゲットドメインへスタイル変換を行った学習ドメインを Target Domain (not via mixup) とする。表 1 より、提案手法が最も高精度に推論していることが分かる。図 3 の推論結果より、窓の領域やシーリングライトの領域など、他の学習ドメインでは正しく推論出来ていないクラスを改善出来ていることから評価指標の精度差へ影響を与えていると考えられる。

4.2 クラス分類

本手法は、異なる認識タスクにも適用できるため、ARC データセットによるクラス分類実験を行う。本実験では、テストデータを実画像パッチデータへ変更し、実画像に対する認識性能を評価する。

実験概要 Source only 及び Target only は、それぞれ UE4 と Iray とする。スタイル変換手法は、CycleGAN [1] を用いる。Step1. 及び Step2. のスタイル変換の学習回数はそれぞれ 200epoch とする。Step3. における損失関数には Softmax Cross Entropy を用いて 100epoch を学習回数とする。タスクモデルは、DRN-C-26 [4] を用いる。テストデータとして実画像 800 枚用いて推論および評価を行う。

実験結果 テストデータによる評価結果を表 2 に示す。提案手法によって生成した学習ドメインが Target Domain である。表 2 より、提案手法が最も高精度に推論していることが分かる。この結果から、単独の CG ドメインより複数の CG ドメインを用いて学習することで、実画像に対する認識性能の向上が見られる。

5. おわりに

本研究では、domain-mixup によって作成した中間ドメインを経由する step-wise なスタイル変換について提案した。実験結果より、提案手法によるデータ増幅の有効性について、セマンティックセグメンテーションおよびクラス分類タスクにて確認した。今後は、未知の実世界ドメインへ対応するスタイル変換方法や、より最適な中間ドメインの生成方法、実世界認識のためのドメイン適応について検討する。

参考文献

- [1] J.Y. Zhu, *et al.*, “Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks”, ICCV, 2017.
- [2] H. Zhang, *et al.*, “mixup: Beyond Empirical Risk Minimization”, ICLR, 2018.
- [3] J. Hoffman, *et al.*, “CyCADA: Cycle-Consistent Adversarial Domain Adaptation”, ICML, 2018.
- [4] F. Yu, *et al.*, “Dilated Residual Networks”, CVPR, 2017.

研究業績

- [1] 大角耕介 等, “Instance Weighting を用いた Gradient Reversal Layer によるドメイン適応の高精度化”, 第 21 回 画像の認識・理解シンポジウム, 2018.
- [2] K. Osumi *et al.*, “Domain Adaptation using a Gradient Reversal Layer with Instance Weighting”, MVA, 2019.