

1. はじめに

セマンティックセグメンテーションは、ピクセルレベルでのクラス識別を行うことで、オブジェクトの種類だけでなく形状も認識することができる。学習と異なるシーンなどドメインの変化が生じると認識精度は著しく低下するため、対象となるドメインごとにモデルを学習する必要がある。そのため、様々な地域で運用される自動運転システムに用いる場合、地域ごとにそれぞれのデータで学習したモデルが必要となり、メモリコストが高くなる。そこで本研究では、単一のモデルで複数のドメインデータを学習するために、マルチヘッド構造を導入したセマンティックセグメンテーションを提案する。各ヘッドは、データセット固有のクラスを出力する。これにより、オブジェクトクラスが異なるデータセットを同時に学習することが可能となる。提案手法は、ベースモデルと比べて1.28倍のパラメータ増加で複数のドメインに対応可能なモデルを実現できることを確認した。

2. セマンティックセグメンテーション

セマンティックセグメンテーションの代表的な手法として、DeepLab v3+[1]がある。DeepLab v3+は、Encoder-Decoder構造と、Atrous Spatial Pyramid Pooling (ASPP)を採用している。ASPPは、異なるDilationの畳み込み処理を並列で行い、特徴マップを統合することで、異なるReceptive fieldを考慮する処理である。従来のセマンティックセグメンテーションは、未学習のドメインに対して認識精度が低下する問題がある。MSeg[2]は、データセットごとに付与されているラベルを共有ラベルに変換して学習することでマルチドメイン学習を実現している。しかし、共有ラベルの再定義にはコストがかかることや、一部のクラスは削除される問題がある。

3. 提案手法

本研究では、データセットを再構築することなく複数のドメインを同時に学習するために、ドメイン固有の出力が可能なマルチヘッド構造を導入したセグメンテーション手法を提案する。提案手法のネットワーク構造を図1に示す。

3.1. 共有ネットワーク

共有ネットワークは、単一のモデルで複数ドメイン情報を共有する部分である。この共有ネットワークには、ResNet101をBackboneにしたDeepLab v3+[1]を用いる。また、ResNet101のResidual BlockにDomain Attention (DA) module[3]を適用する。DA moduleは、ドメインに対するAttentionにより、共有したドメイン情報から入力に対して重要な特徴表現を獲得する機構である。DA moduleの構造を図2に示す。DA moduleは、SE AdapterとDomain Assignmentによって構成される。SE Adapterは、複数のSE moduleで構成されている。それぞれの出力を連結することで全ドメインの表現空間を形成できる。Domain Assignmentは、GAPと全結合層とソフトマックス層で構成されており、ドメインに適応した重みを獲得する。GAPの後の全結合層からの出力は、SE AdapterのSE moduleの数と同じになる。獲得した重みはSE Adapterの出力と乗算し、Sigmoid関数で算出する。これにより、SE Adapterからドメインに適応した重みベクトルが獲得できる。

3.2. ドメインに対応したマルチヘッド構造

一般的なセグメンテーションネットワークは、あらかじめ定義したクラスに対する出力を行うシングルヘッド構造であるため、クラス数が異なるデータセットを同時に学習できない。そこで、異なるオブジェクトクラスを持つデータセットも同時に学習するためマルチヘッド構造を提案する。共有ネットワークで獲得した特徴マップをデータセット固有の出力ヘッドに入力する。そのためデータセット固有のパラメータは、出力ヘッドのみになる。これにより、特徴抽出器は共有となるため、データセット増加によるパ

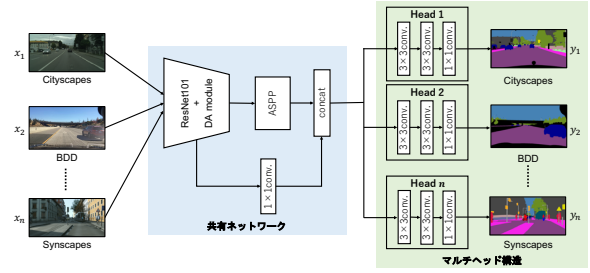


図 1: 提案手法のネットワーク構造

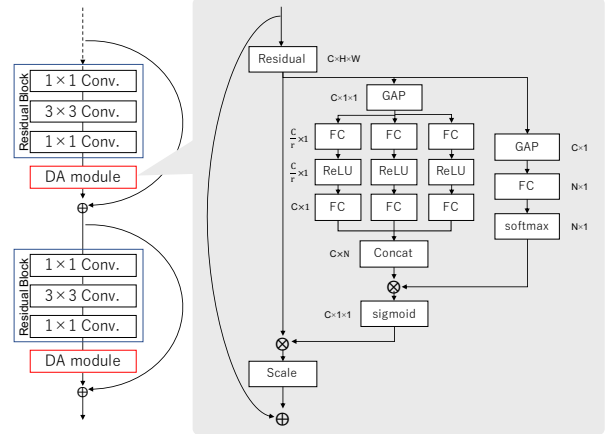


図 2: Domain Attention module の構造

ラメータ数の増加を抑えることが可能とする。

3.3. 損失関数

一般的なマルチドメイン学習 [3] では、異なるドメインを持つデータを順番に入力して、各ヘッドで損失を算出し、逐次逆伝播を行う。このとき、データセット毎にパラメータが更新されるため、逆伝播させる順番によって特定のデータセットにバイアスがかかる可能性がある。そこで、全てのドメインデータを入力して、各ヘッドで出力した損失を合計してから逆伝播する Mix Loss を使用する。 x を各ドメインの入力サンプル、 L_{CE} をクロスエントロピー損失関数として、 N 種類のドメインを学習する際の、損失関数 L は式 (1) のように求めることができる。

$$L = \sum_{n=1}^N L_{CE}(x_n) \quad (1)$$

全てのドメインの損失から逆伝播することで、ドメイン固有の偏りを防ぐことができる。

4. 評価実験

本章では、複数のデータセットを用いて提案手法の有効性を示すために、複数のデータセットを用いて評価実験を行う。

4.1. 実験概要

実験には、同一のオブジェクトクラスで構成されている Cityscapes, BDD, Synscapes の組み合わせと、異なるオブジェクトクラスで構成されている Cityscapes, A2D2, IDD の組み合わせを用いる。学習回数は 100 epoch、評価指標には mIoU を用いる。各データセットに含まれるデータ数は異なるため、1epoch あたりに用いるデータ数を学習枚数が最も多いデータセットに合わせることでデータ数を調整する。最適化には、モーメンタムを 0.9、重み減衰を 0.0001 に設定した SGD を用いる。学習時のミニバッチは、同じデータセットのデータのみで構成する。

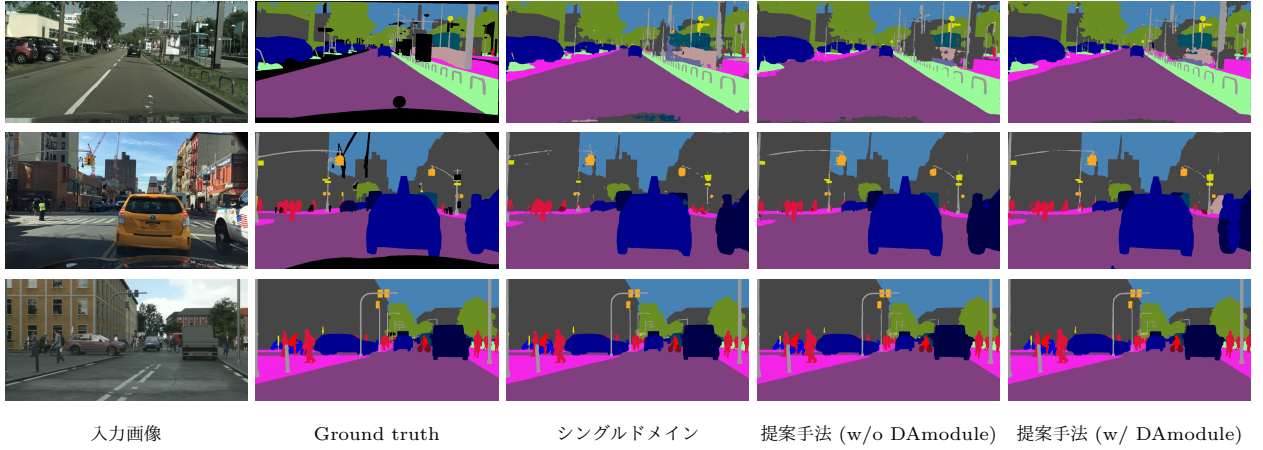


図 3: セマンティックセグメンテーション結果の比較

表 1: 同一クラス数のデータセットでの精度とパラメータ数 [%]

	Cityscapes	BDD	Synscapes	Mean	Params	削減率 [%]
シングルドメイン	77.57	61.55	91.55	76.89	178.02M	-
シングルヘッド	75.55	63.14	86.91	75.34	59.34M	66.67
提案手法 (w/o DA module)	77.92	62.51	90.14	76.86	61.94M	65.21
提案手法 (w/ DA module)	78.49	62.63	90.18	77.10	76.37M	57.10

表 2: Cityscapes でのクラス別精度 [%]

	road	sidewalk	building	wall	fence	pole	light	sign	veg.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
シングルドメイン	98.04	84.36	92.31	51.50	59.37	61.20	69.01	77.39	92.49	65.66	95.03	81.48	60.62	95.07	79.73	88.71	77.98	67.48	76.38	77.57
シングルヘッド	97.14	79.30	91.70	56.22	58.36	60.90	69.33	76.57	92.21	60.13	93.98	79.26	53.34	95.00	83.01	86.72	65.51	61.97	74.83	75.55
提案手法 (w/o DA module)	97.77	82.27	92.08	53.58	60.83	61.65	70.09	78.77	92.48	61.49	94.35	81.45	60.46	95.30	81.78	90.16	81.18	67.45	77.29	77.92
提案手法 (w/ DA module)	97.87	83.29	92.50	54.94	62.88	62.59	69.90	78.45	92.56	62.72	94.36	81.67	62.04	95.10	80.65	91.14	82.81	68.61	77.32	78.49

4.2. 同一クラスを持つデータセットでの学習

本実験では, Cityscapes, BDD, Synscapes を対象データセットとして用いる. 表 1 に認識精度とパラメータ数の比較, 表 2 に Cityscapes でのクラス別精度, 図 3 に各手法のセマンティックセグメンテーション結果を示す.

シングルヘッド構造との比較 マルチヘッド構造の有効性を確認するために, シングルヘッド構造との精度比較を行う. 表 1 より, マルチドメイン学習を行うとき, マルチヘッド構造でない場合, BDD では高い精度を達成したが, Cityscapes と Synscapes では精度が低下したことがわかる. これらのことから, 1 つのデータセットに偏ったモデルといえる. 一方, 提案手法では, 全てのデータセットに対して最も高い精度となり, バランスよく学習されていることがわかる.

DA module 適用による比較 表 1 より, マルチヘッド構造のネットワークに, DA module を適用することで, 全てのデータセットで DA module なしのマルチヘッド構造ネットワークよりも認識精度が向上した. また, Cityscapes と BDD では, シングルドメインで学習したモデルと比較して認識精度が向上した. 表 2 より, クラスごとの精度を比較すると, 19 クラス中 16 クラスの IoU が向上した. road や sky といった大きな物体だけでなく, pole や person といった小物体の精度も向上した. また, 図 3 の定性評価を比較すると, DA module を適用することで building と誤認識している pole クラスを正しく認識できている.

パラメータ数の比較 データセット単体で学習したシングルドメインのモデルは, パラメータ数がデータセットの数だけ増加する. 表 1 より, 提案手法では共有のネットワークを使用することで, シングルドメインのモデルを複数用意する場合よりもパラメータ数を 57.10% 削減した.

これらの結果より, 提案手法によりベースの精度以上, または同等の精度を僅かなパラメータ数の増加で達成することができた.

4.3. 異なるクラスを持つデータセットでの学習

次に, 異なるクラスを持つデータセットを同時に学習可能か実験する. 実験には, 19 クラスの Cityscapes, 18 ク

表 3: 異なるクラス数のデータセットでの精度 [%]

	Cityscapes	A2D2	IDD
シングルドメイン	77.57	78.08	67.80
提案手法	77.84	77.74	68.27

ラスの A2D2, 26 クラスの IDD を用いる. また, モデルは 4.2 節の実験で高精度であった, DA module を適用したモデルを使用する. シングルドメインとの精度比較を表 3 に示す. 表 3 より, シングルドメインの精度とほぼ同等の精度であることを確認した. この結果より, 異なるクラスを持つデータセットでも同時に学習可能であることを確認した.

5. おわりに

本研究では, 異なるドメインを同時に学習するマルチヘッド構造のセマンティックセグメンテーション手法を提案した. データセットごとに出力ヘッドを用意するマルチヘッド構造の導入により, 異なるクラスを持つデータセットに対しても単一のモデルで学習することを可能とした. 今後は, 他のベースネットワークへの適用により汎用性を確認する.

参考文献

- [1] L.-C.Chen, *et al.*, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, ECCV, 2018.
- [2] J.Lambert, *et al.*, “MSeg: A Composite Dataset for Multi-Domain Semantic Segmentation”, CVPR, 2020.
- [3] X.Wang, *et al.*, “Towards Universal Object Detection by Domain Attention”, CVPR, 2019.

研究業績

- [1] S.Masaki, *et al.*, “Multi-Domain Semantic-Segmentation using Multi-Head Model”, ITSC, 2021.

(他 3 件)