

1. はじめに

自動運転システムでは、歩行者や車両をはじめとする周囲の物体を様々な環境下で正確に認識する必要がある。セマンティックセグメンテーションを用いることで、物体や走行可能領域等のセマンティックマップを作成可能となる。セマンティックセグメンテーションは、学習時とドメインが異なる地域のデータで評価すると、道路の形状や建物、看板などの見え方の違いにより精度が低下することがある。そのため、地域ごとのデータに特化して学習したモデルを構築する必要がある。学習した全てのモデルを車両に搭載すると、メモリコストが増加してしまう。

そこで本研究では、モデルマージを用いたマルチドメインセマンティックセグメンテーションを提案する。各地域のデータで学習した複数のモデルを1つのモデルにマージしてメモリコストを削減する。また、マージの際に2つのモデルのマージ割合をブラックボックス最適化により自動探索する。評価実験により、提案手法はパラメータ数を維持しつつ複数のデータセットに対応可能なモデルを実現できることを示す。

2. マルチドメインセマンティックセグメンテーション

マルチドメイン学習の代表的なセマンティックセグメンテーション手法として、マルチヘッドモデル [1] がある。マルチヘッドモデルは、複数ドメイン情報を共有する共有ネットワークとドメインごとに対応する出力ヘッドを持つエンコーダ・デコーダ構造のセマンティックセグメンテーション手法である。ドメインごとに対応する出力ヘッドを持つことで異なるドメインデータセットを同時に学習することができる。しかし、ドメインの数に応じて出力ヘッドが増加するためパラメータ数が増加する。

3. 提案手法

本研究では、パラメータ数を維持したまま複数のドメインのデータで高い精度を達成するために、モデルマージを用いたマルチドメインセマンティックセグメンテーション手法を提案する。セマンティックセグメンテーションのモデルとして SegNeXt [2] を使用する。SegNeXt は ViT をベースとし、Self-Attention 部分をマルチスケール畳み込みに変更したモデルである。マルチスケール畳み込みは、カーネルサイズを 7, 11, 21 のように複数のサイズにすることで、異なるスケールの特徴を同時に抽出する。これにより、画像の局所的な特徴から大域的な特徴まで網羅的に獲得できる。提案手法のモデルマージのプロセスを図 1 に示す。

3.1 ブロックごとの加重平均によるモデルマージ

モデルマージは、学習済みの複数モデルを単一のモデルに統合する手法である。従来は、学習済みの各モデルのパラメータを加重平均して新たなモデルのパラメータとする方法が一般的である。加重平均は各モデルのパラメータを均等に扱うため、個々のモデルが持つドメインの特徴が新たなモデルのパラメータに引き継がれず、精度低下するという問題がある。そこで、本手法はモデルを構成する各ブロックに対してマージ割合を個別に設定し、加重平均によりマージする。これにより、各モデルが持つドメイン固有の特徴を維持できる。モデルマージにおける加重平均の式を (1) に示す。

$$\hat{\theta}_j = \alpha_j \cdot \theta_j^A + (1 - \alpha_j) \cdot \theta_j^B \quad (1)$$

$\hat{\theta}_j$ はマージ後のモデルにおける j 番目のブロックのパラメータ、 α_j は j 番目のブロックのマージの割合、 θ_j^A はモデル A の j 番目のブロックのパラメータ、 θ_j^B はモデル B の j 番目のブロックのパラメータである。

3.2 マージ割合のブラックボックス最適化

モデルマージをする際に、ブロック毎に 0.0 から 1.0 までの範囲で適切なマージ割合 α_j を設定する必要がある。プ

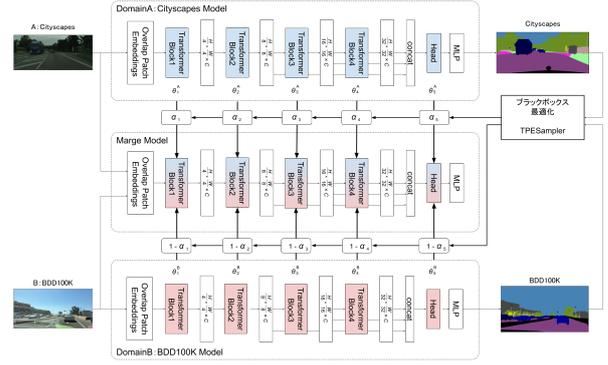


図 1: 提案手法

ロック数が多いと組み合わせ数が膨大となり、手動で探索することは困難である。そこで、ブラックボックス最適化手法である TPESampler [3] を使用し、ブロック毎のマージ割合 α_j を最適化する。本研究では、mIoU を TPESampler の目的関数としてブロック毎の最適なマージ割合を同時に探索する。TPESampler は、はじめにランダムで選ばれた 20 個の探索点に対して目的関数 mIoU で評価する。次に、mIoU の精度上位 25% とそれ以外の 2 つに分割し mIoU の精度上位 25% の探索点のカーネル密度推定 $l(x)$ とそれ以外の探索点のカーネル密度推定 $g(x)$ を式 (2), (3) より導出する。これらに基づいて、式 (4) の獲得関数が最大化するような探索点を決定する。

$$l(x) = \frac{1}{mh} \sum_{i=1}^m \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right) \quad (2)$$

$$g(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right) \quad (3)$$

$$EI(x) = \left(\gamma + \frac{g(x)}{l(x)}(1-\gamma)\right)^{-1} \quad (4)$$

ここで、 $EI(x)$ は獲得関数、 x は探索空間内の点、 x_i は過去に探索した点、 γ は探索範囲を調整する値、 $g(x)$ は低い時のマージ割合の確率分布、 $l(x)$ は mIoU が高い時のマージ割合の確率分布である。 m は mIoU の精度上位 25% の探索数、 n は mIoU の精度上位 25% 以外の探索数、 h はデータの範囲を決めるハイパーパラメータである。

4. 評価実験

提案手法の有効性を確認するために 2 種類のデータセットを用いて評価実験を行う。データ拡張として、水平方向のランダム反転、[0.5, 2.0] の範囲でのランダムスケール、1024×1024 ピクセルでランダムにクロップを用いる。そして、学習の最適化には AdamW を用いて、学習回数を 160000 イテレーション、初期学習率を 0.00006 に設定し、 $(1 - \frac{iter_{total}}{iter})^{0.9}$ を乗算して学習率を設定する。評価指標には mIoU を用いる。

4.1 データセット

本実験では、2つのデータセットを用いる。Cityscapes はドイツの 50 都市で撮影された車載データセットである。データセットは、合計で 5,000 枚の画像で構成されており、異なる時間帯、天候、交通状況などを網羅している。

Berkeley DeepDrive 100K (BDD100K) はアメリカで撮影された多様な走行シーンを含む大規模な車載動画データセットである。データセットは、合計で 10,000 枚で構築されている。

4.2 定量的評価

Cityscapes と BDD100K データセットによる定量的評価を表 1 に示す。Cityscapes で学習したモデルで Cityscapes

表 1: シングルドメイン学習と提案手法の比較 [%]

		マージ	最適化	パラメータ数	評価データセット		
					Cityscapes	BDD100K	平均
シングルドメイン学習	Cityscapes BDD100K			27.6M	82.84 66.17	57.66 66.56	70.25 66.36
マルチドメイン学習				27.6M	75.58	65.17	70.36
マルチヘッド [1]				28.8M	78.80	64.50	71.65
提案手法		✓ ✓	✓	27.6M	52.48 82.03	41.14 63.46	46.81 72.74

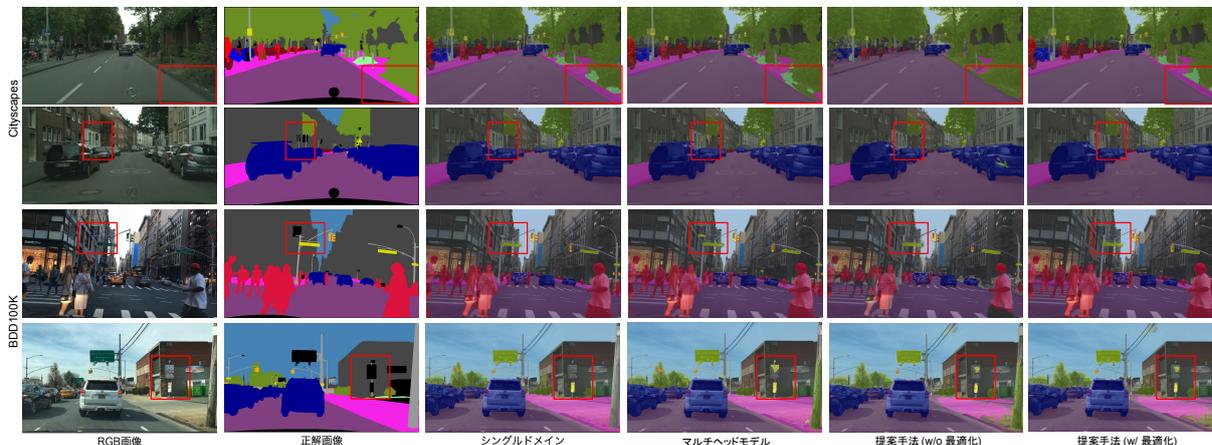


図 2: Cityscapes と BDD100K による定性的評価

に対して 82.84% と高精度であるが、BDD100K では 57.66% と精度が低下する。これより、学習するドメインと評価するドメインが異なる時、精度が大幅に低下していることが確認できる。提案手法と各シングルドメインの精度を比較すると、平均 1.98pt の精度低下が見られたものの、平均ではシングルドメインのモデルと比較して 2.46pt の精度向上した。

提案手法とマルチドメイン学習との比較では、Cityscapes において精度が向上し、BDD100K において精度が低下する結果となった。これは、学習データ数が Cityscapes が約 3,000 枚に対して BDD100K が 7,000 枚と異なるため、モデルが BDD100K データセットに適合した可能性が考えられる。

ブラックボックス最適化の有無による比較では、最適化の導入により両データセットにおいて精度向上した。このことから、マージ割合が重要であると言える。

4.3 ブロックごとのマージの割合

表 2 に最適化後のマージ割合 α 、パラメータの符号割合、パラメータの平均変化量を示す。 α が大きいほど Cityscapes のパラメータの割合が高く、 α が小さいほど BDD100K のパラメータの割合が高い。異符号の割合は、ブロックごとの Cityscapes モデルと BDD100K モデルのパラメータの符号が異なる割合である。パラメータの平均変化量は、ブロックごとのシングルドメインモデルと最適化を行った提案手法モデルでパラメータの平均変化量である。ブロック 1 から 4 では、 α の値が 0.5 前後であることから、二つのモデルのパラメータが均等に反映されたマージである。これは異符号の割合が 10% 前後と低く、パラメータの変化量が小さいため類似したブロックであると考えられる。一方 Head では、 α の値が 0.945 と Cityscapes モデルのパラメータの大部分を使用している。これは異符号の割合が 49% と高く、パラメータの平均変化量が大きいため異なる性能を持つ Head であると考えられる。

4.4 定性的評価

Cityscapes と BDD100K データセットによる定性的評価を図 2 に示す。赤枠で示す部分で精度の変化が見られる。マルチヘッドモデルでは狭い歩道の部分や看板などの詳細な部分でのクラスで誤認識している。最適化をしない場合

表 2: ブロックごとのマージの割合の比較

	マージ割合 α	異符号の 割合	パラメータの 平均変化量
ブロック 1	0.326	7.77	4.50×10^{-3}
ブロック 2	0.685	8.73	5.50×10^{-3}
ブロック 3	0.383	11.52	7.25×10^{-3}
ブロック 4	0.696	12.42	7.95×10^{-3}
Head	0.945	49.44	63.60×10^{-3}

は看板やポールのクラスを認識できていない箇所が見られる。最適化を行うと、マルチヘッドモデルで識別できなかった草木と歩道の境界線部分の識別の精度が向上し、シングルドメインと同等の結果となっている。

5. おわりに

本研究では、マージ割合を最適化したマージモデルによるマルチドメインセマンティックセグメンテーションを提案した。Cityscapes と BDD100K データセットの評価実験より、提案手法はマルチドメイン学習よりも高精度であることを確認した。今後は、データセットの数を増やした際の実験や、新たなマージの方法について行う。

参考文献

- [1] S. Masaki, *et al.*, “Multi-Domain Semantic-Segmentation using Multi-Head Model”, ITSC, 2021.
- [2] G. Guo, *et al.*, “SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation”, NeurIPS, 2023.
- [3] L. Munteanu, *et al.*, “Multiobjective Tree-Structured Parzen Estimator for Computationally Expensive Optimization Problems”, NeurIPS, 2020.

研究業績

- [1] 中村泰智 等, “マルチドメインに対応した半教師あり学習によるセマンティックセグメンテーション”, 画像センシングシンポジウム, 2023.
- [2] 中村泰智 等, “FixMatch を導入したマルチドメイン半教師あり学習によるセマンティックセグメンテーション”, 画像センシングシンポジウム, 2024.