

1. はじめに

セマンティックセグメンテーションは、入力画像の画素単位でクラス分類するタスクである。モデルの学習には、画素ごとにアノテーションした正解データが大量に必要である。また、学習と評価するドメインが異なると精度が著しく低下する問題もある。そこで本研究では、各ドメインに対応した出力を行うマルチヘッド構造を導入することでマルチドメインの問題を解決する。さらに、半教師あり学習を導入することで、アノテーションコストの削減を目指す。

2. マルチドメイン学習

マルチドメイン学習は、複数のドメインのデータセットを同時に学習する手法である。一般的なセグメンテーションネットワークでは、クラス数が異なるデータセットを単一のモデルで学習することは難しい。そこで、文献 [1] は全てのドメインで共通するバックボーンとドメインごとの出力を行うマルチヘッド構造を導入した。しかし、文献 [1] は教師あり学習のため、すべてのドメインに対する大量の正解付き学習データが必要である。

3. 提案手法

本研究ではアノテーションコストを削減するために、半教師あり学習である CAC[2] をマルチドメイン構造を持つセマンティックセグメンテーション手法に導入する。

3.1. ネットワーク構造

ベースネットワーク構造に、ResNet-50 をバックボーンにした DeepLabv3+ を用いる。ネットワーク構造を図 1 に示す。学習時に 1 枚の正解なし学習データから重複領域が存在するランダムなパッチを 2 つ抽出する。2 つのパッチからピクセル単位で対照学習を行うことで正解なし学習データの整合性を保つことができ、環境の変化に対するロバスト性が向上する。

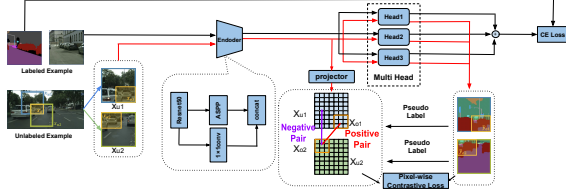


図 1: 提案手法の CAC の構成

3.2. 学習方法

半教師あり学習は、教師あり学習と教師なし学習を組み合わせる学習方法であり、CAC では擬似ラベリングを使用する。擬似ラベリングは、正解付き学習データを用いた教師あり学習を行ったモデルを用いて正解なし学習データに対して、擬似ラベルを付与する方法である。

教師あり学習

教師あり学習の損失関数は式 (1) の Cross-Entropy Loss (CELoss) を用いる。このとき p は教師データ、 q 正解付き学習データ、 i クラス数である。

$$L_{ce} = \sum_{n=1}^N p(i) \log(q(i)) \quad (1)$$

教師なし学習

教師なし学習の損失関数は式 (2) の Directional Contrastive Loss (DCLoss) を用いる。まず、正解なし学習データから重複領域を含む 2 つのパッチをランダムに切り出して、モデルに入力し、各画素に対するクラス分類を行う。そして、重複領域に着目し、クラス分類の信頼度が低い特徴を信頼度が高い特徴に近づくように学習を行う。

$$L_{dc}(\hat{y}) = -\frac{1}{N} \sum_{h,w} M_{d,p}^{h,w} \log \frac{r(\hat{y})}{r(\hat{y}) \sum M_{n,1}^{h,w} r(x_{o1}, x_{u2})} \quad (2)$$

ここで、 $\hat{y}$ は 2 つのパッチの重複領域、 x_{o1} は 1 枚目の重複領域のピクセル、 x_{u2} は 2 枚目のパッチのピクセル、 r

は類似度の指数関数、 N は重複空間のピクセル数、 $M_{n,1}^{h,w}$ は正解なしデータに擬似ラベルを付与する関数、 $M_{d,p}^{h,w}$ は閾値より信頼度の低い特徴を計算から除外する関数である。

マルチヘッド構造

マルチヘッド構造において各データセットごとに逆伝播を行う場合、学習を行うデータセットの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、Mix Loss により全てのドメインデータを入力し、各ヘッドで出力した Loss を合計してから逆伝播を行う。Mix Loss を式 (3) に示す。このとき λ は、教師なし Loss の重み、 L_{ce} は CELoss、 L_{dc} は DCLoss、 x_n は各ドメインの入力サンプル、 N はデータセット数である。

$$L = \sum_{n=1}^N (L_{ce} + \lambda L_{dc}) x_n \quad (3)$$

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評価する。

4.1. 実験概要

実験では Cityscapes, Synscapes, Berkley DeepDrive 100K (BDD100K) データセットを使用する。最適化手法に SGD を使用し、重み減衰を 0.0001、モーメントを 0.9、初期学習率を 0.01 とする。また、全てのモデルを各データセットにおいて 150 エポック学習する。さらに、半教師あり学習では各データセットで学習する総枚数の $\frac{1}{8}$ をラベルあり画像、 $\frac{7}{8}$ をラベルなし画像で学習する。教師あり学習は各データセットで学習する総枚数の $\frac{1}{8}$ をラベルあり画像で学習する。評価指標には mIoU を用いる。

4.2. 実験結果

定量的評価を表 1、結果画像を図 2 に示す。表 1 より、提案手法は従来手法と比較し、全てのデータセットで精度の向上が見られた。また、マルチドメインはシングルドメインと比較し、BDD100K データセットの認識精度が 9.22pt 向上した。定性的評価より、図 2 の赤丸で示すように rider と bicycle が正解であるが、シングルドメインでは people と認識している、一方、提案手法は rider と bicycle を正しく認識していることが確認できる。

表 1: シングルドメインと提案手法の比較 [%]

学習方法	Cityscapes	Synscapes	BDD100k	平均
シングルドメイン				
従来手法	61.83	80.22	41.99	61.35
提案手法	63.85	81.45	43.85	63.05
マルチドメイン				
従来手法	63.24	73.57	50.18	62.33
提案手法	65.52	77.60	53.07	65.40

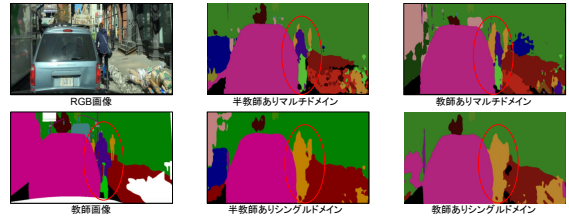


図 2: BDD100K データセットの結果画像

5. おわりに

本研究では、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手法の有効性を示した。今後は、データセットの数や新たな学習方法を考案し、より高い精度を得られる条件を検討する。

参考文献

- [1] M.Shota, et al., “Multi-Domain Semantic-Segmentation using Multi-Head Model”, ITSC, 2021.
- [2] X.Lai, et al., “Semi-supervised Semantic Segmentation with Directional Context-aware Consistency”, CVPR, 2021.