Open-Set Domain Adaptation for Semantic Segmentation (OSDA-SS) with CRF and Unknown Class Reweighting

[1]Andy Wang, [+]Rahul Gomes, [*]Papia F. Rozario, [2]Daniel Chvat, [3]Landon Dierkes, [4]Grace Abraham, [5]Kristen North

[1]University of Wisconsin Madison, Mathematics and Computer Science, [+]University of Wisconsin Eau-Claire, Computer Science, [*]University of Wisconsin Eau-Claire Geography and Anthropology,
[2]University of California- Los Angeles, Computer Science, [3]Madison Area Technical College, Computer Science, [4]North Carolina State University, Computer Science, [5]El Camino College, Computer Science

Introduction

Unsupervised Domain Adaptation (UDA)

- Labeling data is costly and time consuming.
- UDA Aims to transfer knowledge from a labeled source domain to an unlabeled target domain.
- Most research on UDA for semantic segmentation are aimed towards closed-set UDA, where the classes between two domains are one to one.

OSDA for Semantic Segmentation

- Proposed recently by Choe et al. (2024)
- Two domains share the same C classes, but the target domain has additional classes.
- Aims to predict classes unknown from the source domain to a specialized unknown class.

The Dataset

- Synthetic Images (GTA5) and real urban street images (Cityscapes)
- GTA5 offering 24966 in-game images with resolution 1914 x 1050.
- Cityscapes offering 2975 training samples and 500 validation samples with resolution 2048 x 1024.
- Following Choe et al. (2024), 6 classes were set as the unknown class and was removed from the source label.
- The goal is to use OSDA to transfer knowledge from GTA5 to Cityscapes.

Known Classes
Road, Sidewalk, Building, Wall, Fence, Pole, Traffic Light, Vegetation, Terrain, Sky, Car, Bus, Motorcycle, Bicycle
Unknown Classes
Pole, Traffic Sign, Person, Rider, Truck, Train

Table 1: Chosen known/unknown classes for OSDA-SS.

Evaluation Metrics

- Besides using common metrics, we further evaluate by **H Score**, a harmonic mean of the mean Intersection over union (mIoU) for known and unknown classes. This can emphasize the importances of both classes.

$$H\ Score = \frac{2mIoU_{known}IoU_{unknown}}{mIoU_{known} + IoU_{unknown}}$$

Preliminary models and methods

DAFormer (Hoyer et al., 2022)

- Preliminary Closed-set UDA model for benchmarking.
- Employs self-training on target images to achieve effective UDA.
- Uses DACS Augmentation, RCS and Feature Distance to enhance accuracy.

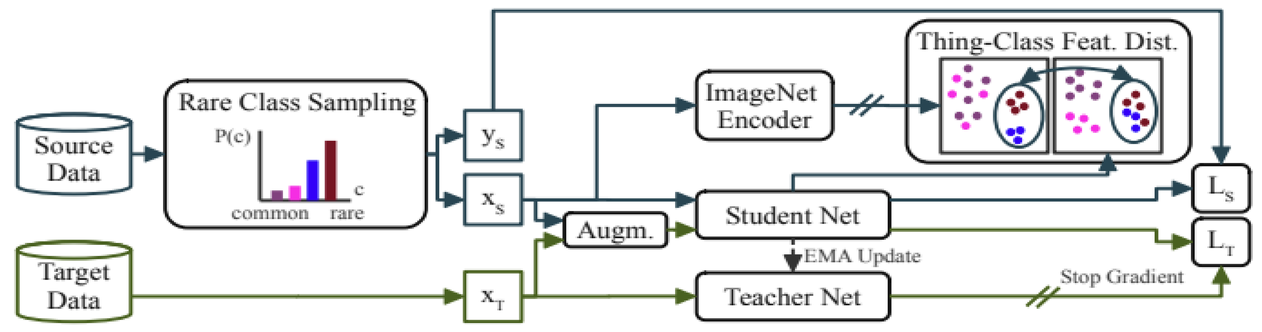

Fig. 1: DAFormer Model for Closed Set UDA.

Boundary and Unknown Shape-Aware Open-set Domain Adaptation (BUS) (Choe et al., 2024)

- Used Contrastive Learning methods to enforce boundary awareness.
- Proposed OpenReMix: image mixing method focused on unknown classes.
- Designed Unknown Head Expansion: Pixels in target pseudo labels with confidence below τ=0.5 are set as the unknown class for self-training.

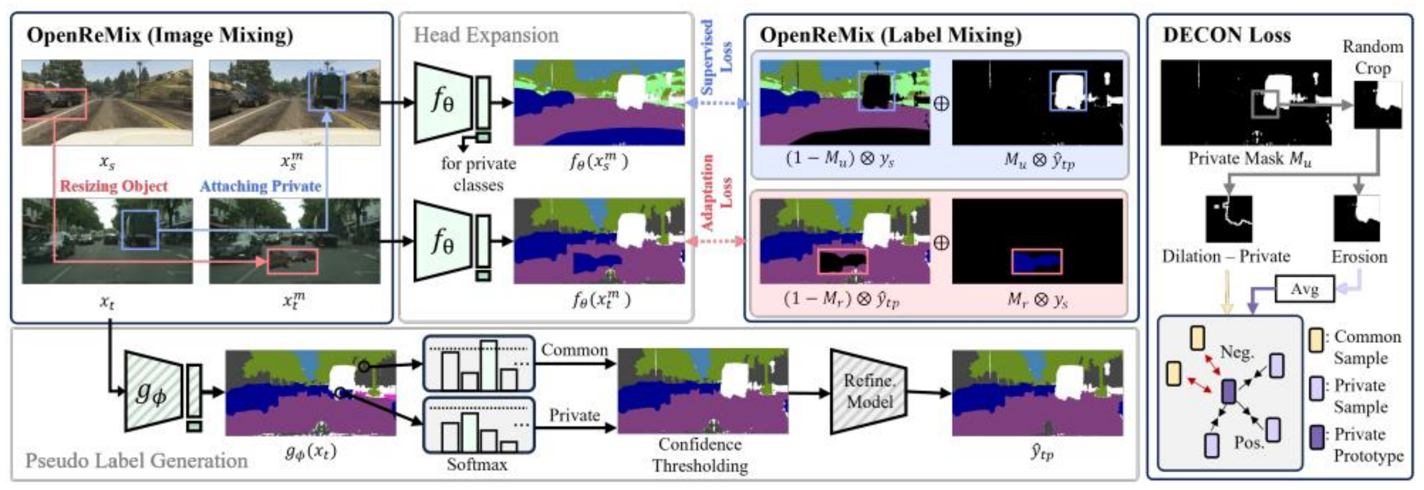

Fig. 2: BUS architecture for OSDA-SS.

Proposed methods

Pseudo Label Refinement with Conditional Random Fields (CRF)

- OSDA-SS heavily relies on the pseudo label's quality.
- CRFs are statistical models for structured predictions, ensuring spatial consistency and smooth boundaries between different regions in an image.
- We hypothesize by providing extra post-refinement, the model can produce high quality pseudo labels, facilitating better performance.

Image → UDA Model → Pseudo Label → CRF → Refined Label

Fig. 3: A flowchart of using CRF to refine pseudo labels.

Unknown Class Reweighting (ReW)

- The unknown class is only trained from pseudo labels, while known classes are trained from both pseudo labels and source labels, Making unknown classes get manifested by known classes easily.
- We hypothesize by weighting extra loss for unknown pixels with high confidence (0.7), the model can maintain focus on the unknown classes.
- We weight the unknown pixels with high confidence 1.5 times more.

Experiments and Results

- All models were train with 20000 iterations with a batch size of 4.
- Each Image has been processed into several 512 x 512 grids as input.
- All models use DAFormer as their base line model.

Method	$mIoU_{known}$	$IoU_{unknown}$	H Score	Diff
Head Exp	56.45	30.19	39.24	-
CRF	65.41	34.74	45.38	+6.14
CRF + ReW	**66.49**	35.2	46.03	+6.79
BUS	59.24	35.55	44.52	+5.28
BUS + CRF	61.29	37.27	46.36	+7.12
BUS + CRF + ReW	60.96	**38.99**	**47.56**	**+8.32**

Table 2: A comparison of different methods for the GTA->Cityscapes OSDA-SS task. Diff measures the difference of H Score compared to head expansion baseline.

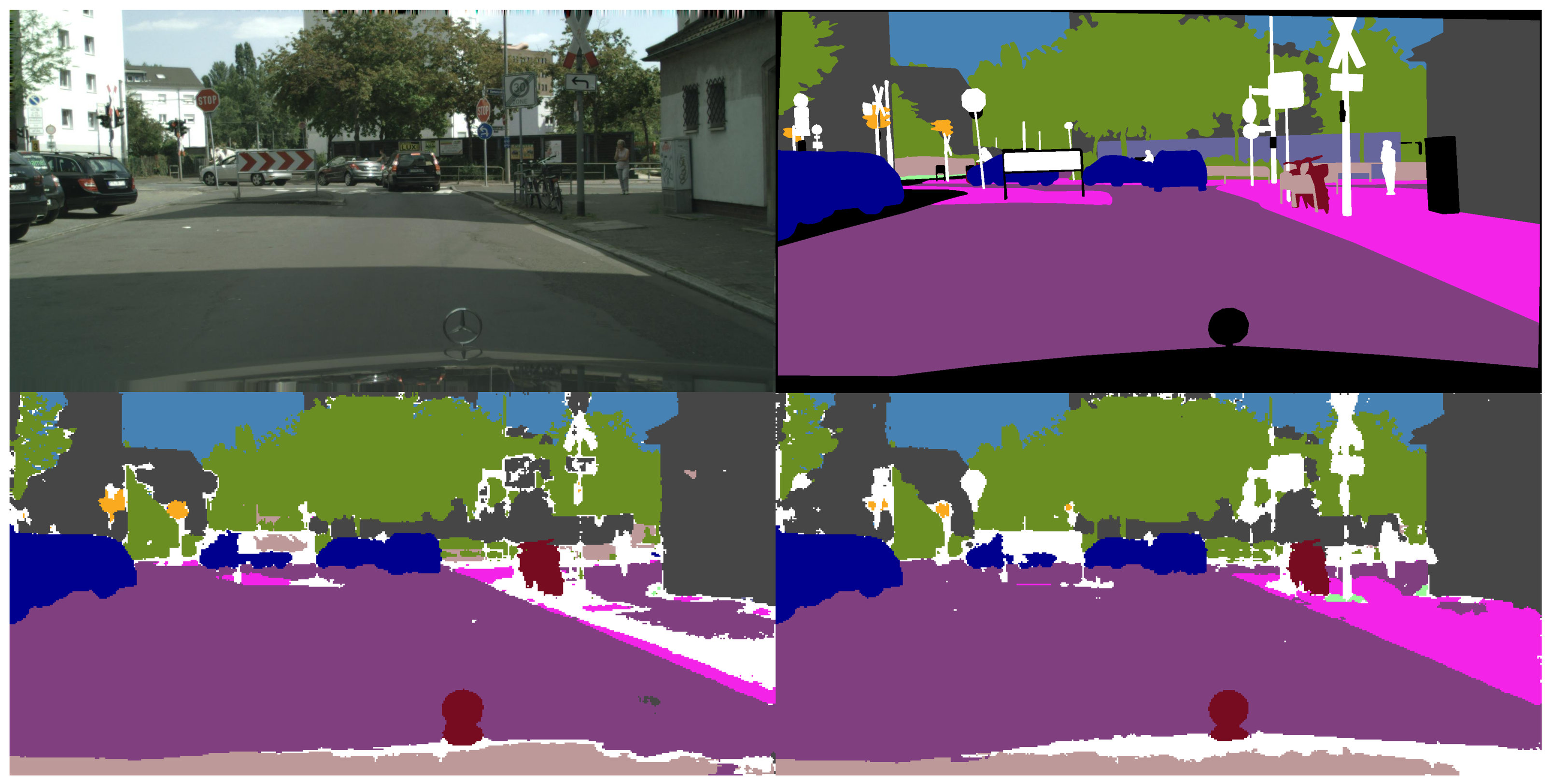

Fig. 4: Comparison Between head expansion baseline and BUS + CRF + ReW.
Top left: Original Image, Top Right: Ground Truth
Bottom left: head expansion, Bottom Right: BUS + CRF + ReW

Discussion & Future works

- CRF significantly enhanced the quality of pseudo labels, leading to a substantial increase in the H Score.
- ReW showed a slight improvement in the H Score. We suggest further exploration of hyperparameters to enhance the method's performance.
- Comparing different, newer UDA baseline models (HRDA, MIC) could provide valuable insights to the task.
- Incorporating generative models to blend domains could help the model learn target-specific features more effectively.

Acknowledgments

This research was funded by the National Science Foundation (NSF) Research Experience for Undergraduates (REU) grant OAC-2150191. We would also like to thank the Office of Research and Sponsored Programs (ORSP) at UW-Eau Claire for student support. The computational resources of this study was provided by the Blugold Center for High-Performance Computing under NSF grant CNS-1920220.

We thank the Office of Research and Sponsored Programs for supporting this research, and Learning & Technology Services for printing this poster.