Semantic Segmentation and Adversarial Domain Adaptation

Student: Yifan Bao Rui Feng
Professor: Min Chi

Introduction

Image Segmentation: Deep learning-based image segmentation models, like deeplab family, often achieving the highest accuracy, have received a lot of attention.

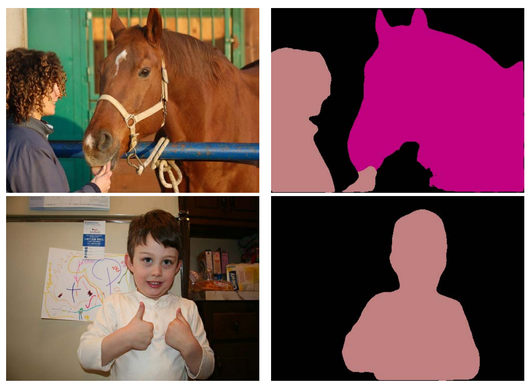

Figure1: Examples of Image Segmentation

Domain Adaptation: To reduce the labeling cost, unsupervised domain adaptation (UDA) approaches are proposed to transfer knowledge from labeled synthesized datasets to unlabeled real-world datasets.

Datasets

Cityscapes: It includes 5,000 annotated images with 2048 × 1024 resolution, grouped into 8 categories.

SYNTHIA: We use the SYNTHIA-RAND-CITYSCAPES subset consisting of 9,400 1280×760 synthetic images.

Metrics

Intersection over Union (IoU) is defined as

$$IoU = \frac{Intersection}{Union} = \frac{TP}{TP + FP + FN}$$

where TP means true positive. FP means false positive. FN means false negative.

Mean-IoU (mIoU) is defined as the average IoU over all classes.

$$mIoU == \frac{\sum_i IoU_i}{n}$$

where n is the number of classes and IoU_i is the IoU metric for $i-$th class.

References

[1] Chen L C; Zhu Y; Papandreou G, et al.: *Encoder-decoder with atrous separable convolution for semanticimage segmentation*, Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

[2] Tsai, Y.H.; Hung, W.C.; Schulter, S.; Sohn, K.; Yang, M.H.; Chandraker, M.: *Learning to adapt structuredoutput space for semantic segmentation.*, Proc. of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2018, pp. 7472–7481.

Acknowledgements

We thank Prof. Chi for insightful dicussions and ideas. We also thank North Carolina State University's GEARS program .

Model Framework

AdaptSegNet: According to Figure3, AdaptSegNet consists of two parts: one is the **segmentation network**, the other is the **domain adaptation module**.

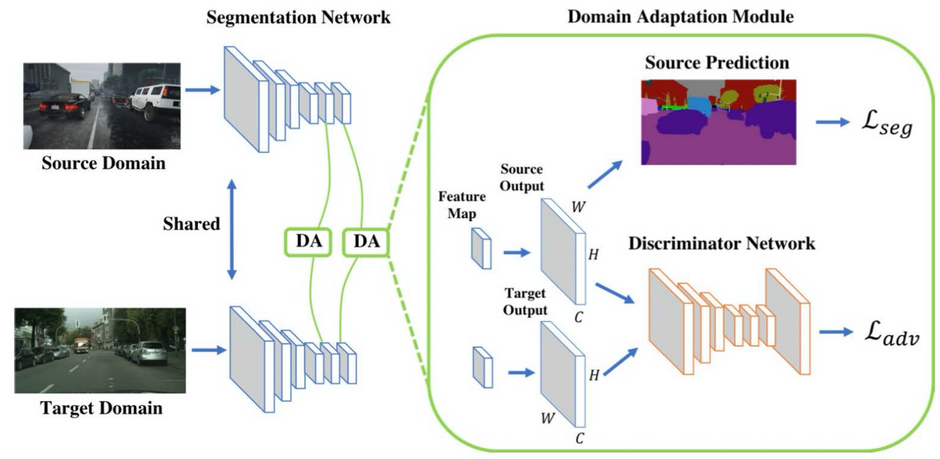

Figure2: AdaptSegnet Architecture (From [2])

In Figure2, we use DeepLabV2 and DeepLabV3+ as the segmentation network separately.

Qualitativie and Quantitative Results

In Figure3, we train DeeplabV2 model on SYNTHIA dataset. The first column is the target image, the second column is the groundtruth and the last is the segmentation result.

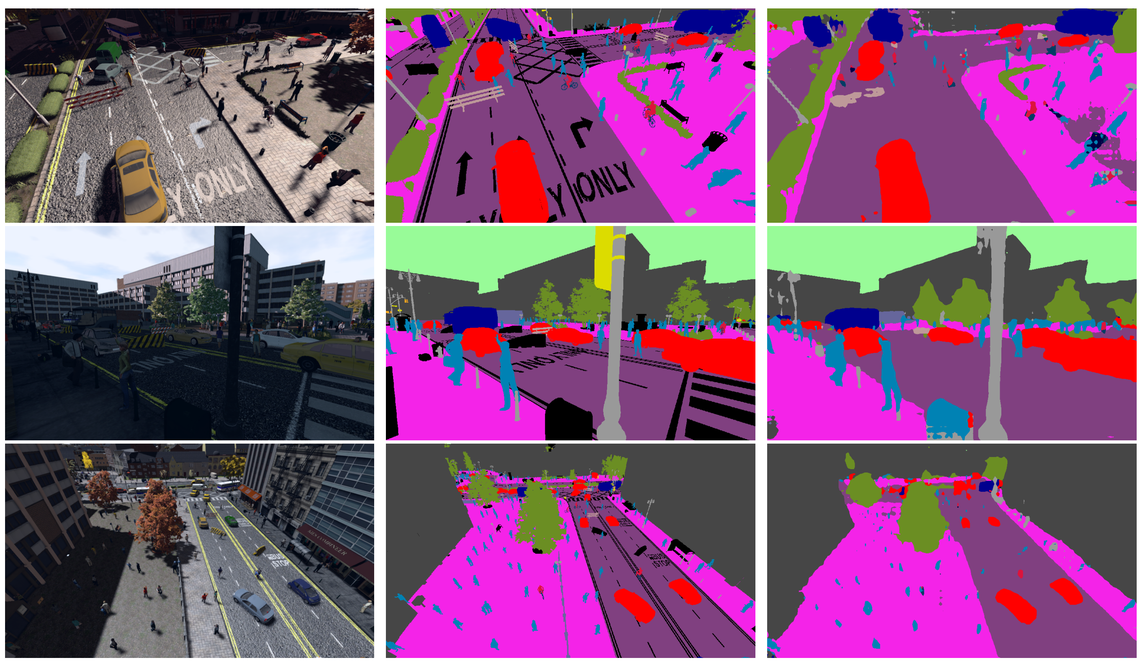

Figure3: Segmentation Result of DeeplabV2 on SYNTHIA

In Figure4, the first column is the target image in Cityscapes dataset. The second column is the groundtruth label. The next four columns are the segmentation results by with and without domain adapation. Specifically, BA(V2) means DeeplabV2 model without domain adapation; DA(V2) means DeeplabV2 model with domain adapation; BA(V3+) means DeeplabV3+ model without domain adapation; DA(V3+) means DeeplabV3+ model with domain adapation.

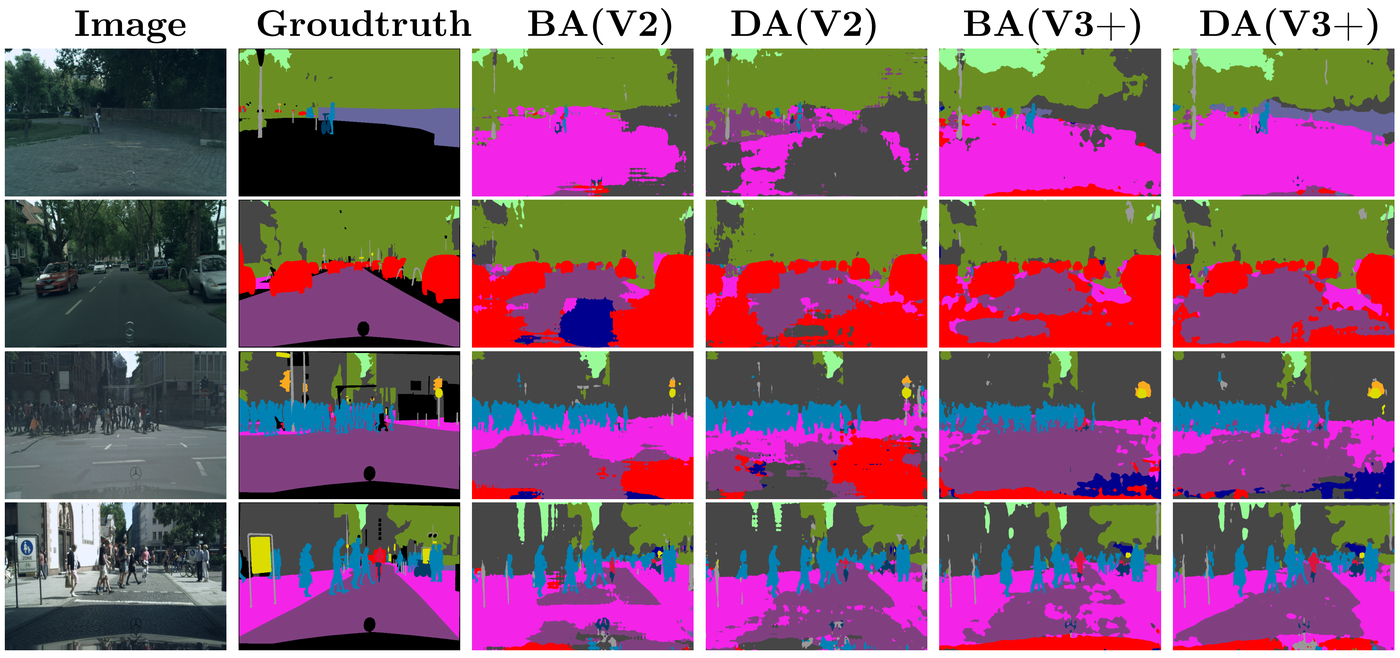

Figure4: Qualitative Image Segmentation Results on the Cityscapes Dataset

Table1: Quantitative Results for SYNTHIA-to-Cityscapes experiments

Method	Seg Model	building	walls*	veg	sky	person	motor	bike	mIoU (%)	mIoU* (%)
Source only	DeeplabV2	75.8	6.4	**79.0**	**81.1**	**57.6**	**20.8**	28.2	34.1	39.4
Source only	DeeplabV3+	74.8	5.3	74.0	71.9	52.0	10.0	22.5	33.2	38.6
AdaptSegNet	DeeplabV2	64.9	4.7	76.9	76.3	52.8	16.0	**34.7**	35.3	41.1
AdaptSegNet	DeeplabV3+	**76.8**	**10.1**	75.7	73.4	53.1	19.1	29.9	**37.0**	**42.8**