ELDA: Using Edges to Have an Edge on Semantic Segmentation Based UDA Supplementary Material

Ting-Hsuan Liao*[1]
tingforun@gapp.nthu.edu.tw
Huang-Ru Liao*[1]
mimiliao2000@gapp.nthu.edu.tw
Shan-Ya Yang†[1]
sonya.tw2000@gapp.nthu.edu.tw
Jie-En Yao†[1]
matt1129yao@gapp.nthu.edu.tw
Li-Yuan Tsao†[1]
abcdabcd10211@gapp.nthu.edu.tw
Hsu-Shen Liu†[1]
tenhoshi9023@gapp.nthu.edu.tw
Bo-Wun Cheng‡[1]
bobcheng15@gapp.nthu.edu.tw
Chen-Hao Chao‡[1]
lance_chao@gapp.nthu.edu.tw
Chia-Che Chang‡[2]
chia-che.chang@mediatek.com
Yi-Chen Lo‡[2]
yichen.lo@mediatek.com
Chun-Yi Lee[1]
cylee@gapp.nthu.edu.tw

[1] Elsa Lab,
Department of Computer Science,
National Tsing Hua University,
No. 101, Section 2, Kuang-Fu Road,
Hsinchu City 300,
Taiwan

[2] MediaTek, Inc.,
No. 1, Dusing 1st Road,
Hsinchu Science Park,
Hsinchu City 300,
Taiwan

*, †, and ‡ indicate equal contributions.

1 Notation Table

The symbols used in this paper are summarized in Table 1.

2 Datasets

In this section, we describe the details of the datasets used in the experiments of this paper.

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

Symbol	Description
X_s	A set of training images from the source domain.
X_t	A set of training images from the target domain.
Y_s	The set of ground truth semantic segmentation labels corresponding to X_s.
x_s	An image from the source domain.
y_s	The ground truth semantic segmentation label corresponding to x_s.
$\mathrm{C}(x_s)$	The edges generated by the Canny algorithm from x_s.
x_t	An image from the target domain.
$y_t^{'}$	The pseudo semantic segmentation label corresponding to x_t.
$\mathrm{C}(x_t)$	The edges generated by the Canny algorithm from x_t.
$\mathrm{C}(\cdot)$	The Canny operator.
$\hat{y}_s^{\text{init}}$	The initial semantic segmentation prediction for x_s.
$\hat{y}_t^{\text{init}}$	The initial semantic segmentation prediction for x_t.
$\hat{y}_s^{\text{final}}$	The final semantic segmentation prediction for x_s.
$\hat{y}_t^{\text{final}}$	The final semantic segmentation prediction for x_t.
$\hat{e}_s^{\text{init}}$	The initial edge prediction for x_s.
$\hat{e}_t^{\text{init}}$	The initial edge prediction for x_t.
$\hat{e}_s^{\text{final}}$	The final edge prediction for x_s.
$\hat{e}_t^{\text{final}}$	The final edge prediction for x_t.
λ	A balancing factor for the segmentation and edge prediction tasks.

Table 1: A list of the symbols used in this paper.

2.1 Cityscapes Dataset

The Cityscapes dataset [2] is a real-world dataset containing semantic segmentation of a number of urban scenes. We use $2,975$ unlabeled images for training and 500 images for validation. The resolution is set to 2048×1024 pixels.

2.2 GTA5 Dataset

The GTA5 dataset [5] is a synthetic dataset. It contains $24,966$ images with the resolution set to 1914×1052 pixels. The images are labeled with Cityscapes-style annotations that include 19 classes. In our experiments, we use the entire dataset as our source domain samples for training ELDA as well as the baselines.

2.3 SYNTHIA Dataset

The SYNTHIA dataset [6] is a synthetic dataset that is generated from an autonomous driving simulator of a number of urban scenes. In our experiments, we use the SYNTHIA-RAND-CITYSCAPES subset, which contains 9,400 images with the resolution set to 1280×760 pixels. The images are labeled with Cityscapes-style annotations of 16 classes.

3 Implementation Details

In this section, we elaborate on the implementation details of ELDA. We adopt Deeplabv2[1] with ResNet-101[3] as the backbone of ELDA. Our model is trained using the SGD optimizer [7] with a learning rate 2.5×10^{-4} decayed by a factor of 0.9 for $250,000$ iterations.

The weight decay of the model parameters is set to 5×10^{-4}, and the momentum is set to 0.9. Images are randomly cropped to 512×512 pixels. The balancing factor λ is set as 0.01 and 1 for GTA5→Cityscapes and SYNTHIA→Cityscapes benchmarks, respectively. The hyper-parameter σ for the Canny operator is set to 2. The batch size is set to 4, where in each mini-batch there are two images from X_s and two images from X_t. The code is implemented using Python with the PyTorch library [4]. All the methods are trained and evaluated on an NVIDIA Tesla V100 GPU.

4 Visualization of Additional Qualitative Results

In this section, we present additional qualitative results, including the semantic segmentation predictions from *source only*, CorDA [8], and ELDA in Fig. 1, as well as the edge predictions from ELDA and the Canny algorithm in Fig. 2.

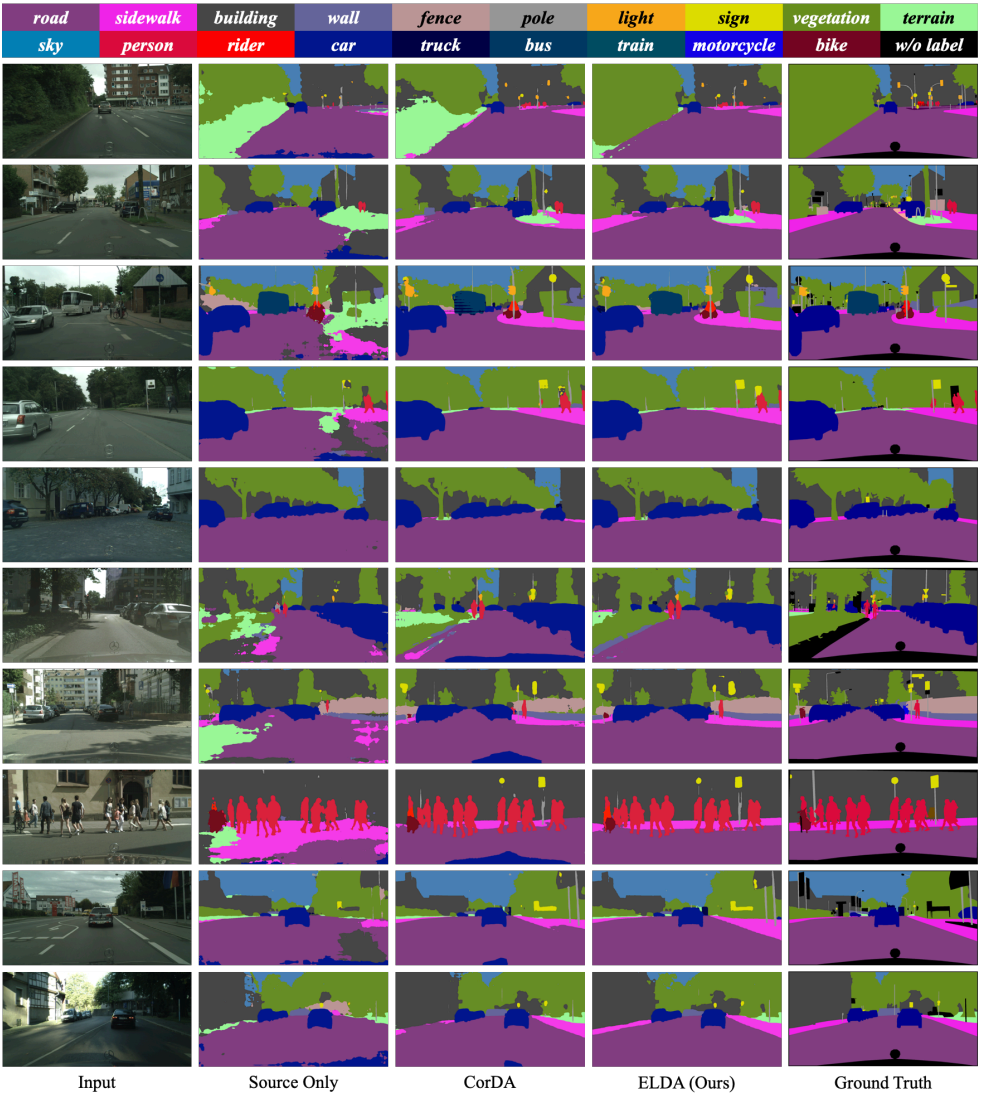

Figure 1: An illustration of the semantic segmentation predictions from *source only*, CorDA [8], and ELDA, along with their corresponding input images and the ground truths. It can be observed that the predictions from ELDA are less noisy, and are able to preserve the boundaries in the predicted segmentation maps better than the other two baseline approaches.

Figure 2: A comparison of the edge predictions from ELDA and the Canny algorithm. It can be observed that the edges predicted by ELDA are close to those generated by the Canny algorithm, indicating that ELDA is able to capture the essential edge features from its input images.

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, 2017.

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding, 2016.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[4] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[5] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games, 2016.

[6] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[7] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[8] Qin Wang, Dengxin Dai, Lukas Hoyer, Luc Van Gool, and Olga Fink. Domain adaptive semantic segmentation with self-supervised depth estimation, 2021.