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

Paper ID: 108

Ting-Hsuan Liao*, Huang-Ru Liao*, Shan-Ya Yang, Jie-En Yao, Li-Yuan Tsao, Hsu-Shen Liu, Chen-Hao Chao, Bo-Wun Cheng, Yi-Chen Lo, Chia-Che Chang, and Chun-Yi Lee

Elsa Lab, Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

BMVC 2022

Introduction

In this paper, we introduce Edge Learning based Domain Adaptation (ELDA), a novel unsupervised domain adaptation (UDA) framework which incorporates edge information into its training process to serve as a type of domain invariant information. In our experiments, we quantitatively and qualitatively demonstrate that the incorporation of edge information is indeed beneficial and effective, as it enables ELDA to outperform the contemporary state-of-the-art methods on two commonly adopted benchmarks for semantic segmentation based UDA tasks. We further provide ablation analysis to justify the decisions of ELDA.

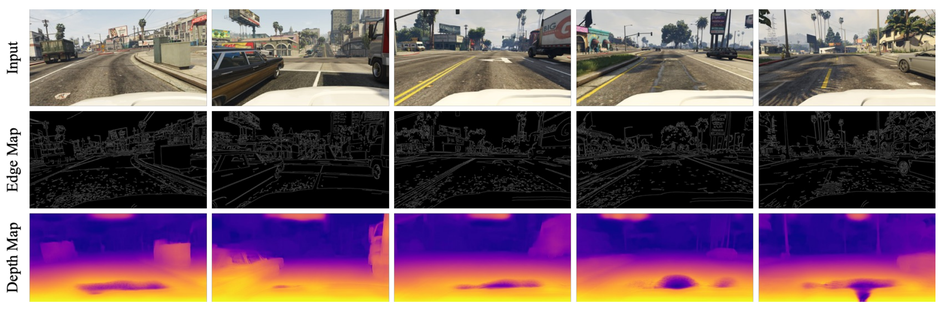

Figure 1. An example showing the differences of the depth maps extracted by CorDA and the edges of the images from GTA5 without the use of any ground truth labels.

Several methods have been proposed to leverage self-supervised learning (SSL) techniques to retrieve depth information to assist semantic-based UDA. Unfortunately, the methods that utilize SSL to retrieve depth information has two crucial constraints:

1. The computational cost of training an accurate auxiliary SSL-based depth estimation model is often expensive.
2. The performance is not comparable to physical sensors or supervised models in terms of accuracy, as depicted in Fig 1.

We propose to replace depth with edge. The benefits are twofold:

1. the computational cost of extracting edges from an input image is substantially lower than extracting depth map using SSL.
2. the quality of edges is typically much more consistent than that of depth.

The experimental results show that our method can achieve state-of-the-art performance on two commonly adopted benchmarks.

Methodology

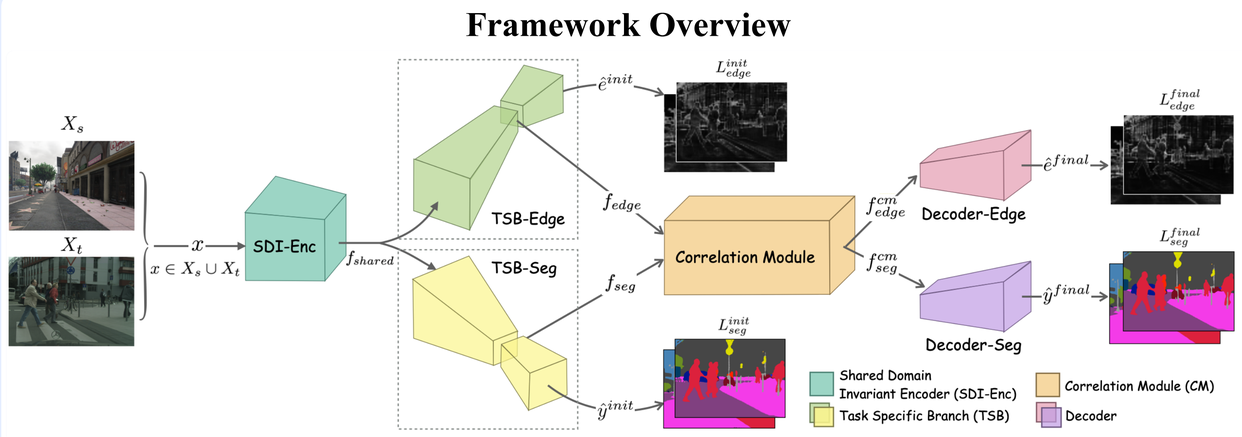

Figure 2. An overview of the proposed ELDA framework.

X_s: source dataset. X_t: target dataset. f: feature representation.
ê: edge prediction. $\hat{y}$: segmentation prediction. $*^{init}$: initial prediction. $*^{final}$: final prediction.

Model Components

Shared Domain Invariant Encoder (SDI-Enc)

ELDA employs the shared encoder technique for capturing both edge and segmentation features. An input image from either the source domain or the target domain is fed into the shared encoder to extract a shared feature f_{shared}.

Task Specific Branch (TSB)

To enable f_{shared} to be further interpreted into specific feature embeddings that bear edge and segmentation meanings, two separate branches of TSBs are utilized to generate initial edge and segmentation predictions. The encoders are in charge of encrypting f_{shared} into task specific features f_{edge} and f_{seg}, which are later fed to CM. On the other hand, the decoders are responsible for decoding f_{edge} and f_{seg} into $\hat{e}_s^{init}$ or $\hat{e}_e^{init}$ and $\hat{y}_s^{init}$ or $\hat{y}_e^{init}$, respectively, depending on the original domains of the input images, for updating SDI-Enc and the TSBs.

Correlation Module (CM)

With the goal of communicating information between the task specific latent embeddings f_{edge} and f_{seg}, we use a correlation module in the ELDA architecture. This operation helps the model to preserve the essential features from the two TSBs. CM can be formulated as the following equations, and illustrated as Fig. 3.

$$f_{seg}^{mid} = Conv(f_{seg}), f_{edge}^{mid} = Conv(f_{egde})$$

$$f_{seg}^{cm} = f_{seg} + f_{edge}^{mid} * Sigmoid(Conv(f_{edge}))$$

$$f_{edge}^{cm} = f_{edge} + f_{seg}^{mid} * Sigmoid(Conv(f_{seg}))$$

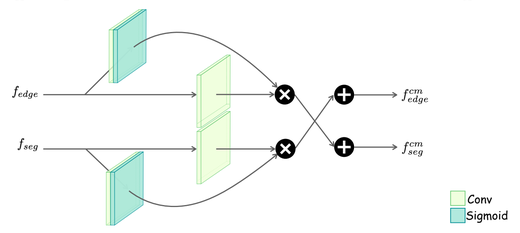

Figure 3. An overview of CM module.

Experimental Results

Quantitative Results

SYNTHIA → Cityscapes

Method	Aux.	Road	SideW	Build	Wall	Fence	Pole	Light	Sign	Veg	Sky	Person	Rider	Car	Bus	Motor	Bike	mIoU
Source only		51.8	17.0	73.0	7.1	0.2	25.4	9.4	10.2	70.7	84.0	55.6	13.7	68.0	2.9	8.5	16.1	32.1
CBST		68.0	29.9	76.3	10.8	1.4	33.9	22.8	29.5	77.6	78.3	60.6	28.3	81.6	23.5	18.8	39.8	42.6
CAG-UDA		84.7	40.8	81.7	7.8	0.0	35.1	13.3	22.7	84.5	77.6	64.2	27.8	80.9	19.7	22.7	48.3	44.5
Uncertainty		87.6	41.9	83.1	14.7	1.7	36.2	31.3	19.9	81.6	80.6	63.0	21.8	86.2	40.7	23.6	53.1	47.9
IAST		81.9	41.5	83.3	17.7	4.6	32.3	30.9	28.8	83.4	85.0	65.5	30.8	86.5	38.2	**33.1**	52.7	49.8
DACS		80.6	25.1	81.9	21.5	2.9	37.2	22.7	24.0	83.7	90.8	67.6	38.3	82.9	38.9	28.5	47.6	48.3
SPIGAN	✓	71.1	29.8	71.4	3.7	0.3	33.2	6.4	15.6	81.2	78.9	52.7	13.1	75.9	25.5	10.0	20.5	36.8
GIO-Ada	✓	78.3	29.2	76.9	11.4	0.3	26.5	10.8	17.2	81.7	81.9	45.8	15.4	68.0	15.9	7.5	30.4	37.3
DADA	✓	89.2	44.8	81.4	6.8	0.3	26.2	8.6	11.1	81.8	84.0	54.7	19.3	79.7	40.7	14.0	38.8	42.6
GUDA	✓	88.1	53.0	84.0	22.0	1.4	**39.6**	28.2	24.8	82.7	81.5	65.5	22.7	**89.3**	**50.5**	25.1	**57.5**	51.0
CorDA	✓	**93.3**	**61.6**	85.3	19.6	**5.1**	37.8	36.6	42.8	84.9	90.4	69.7	41.8	85.6	38.4	32.6	53.9	55.0
ELDA (Ours)	✓	92.6	56.6	**85.5**	**24.2**	2.1	37.6	**38.1**	**43.1**	**85.7**	**91.5**	**69.8**	**42.0**	87.2	47.6	20.0	50.1	**55.2**

GTA5 → Cityscapes

Method	Aux.	Road	SideW	Build	Wall	Fence	Pole	Light	Sign	Veg	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
Source only		70.1	18.4	66.1	12.8	17.4	22.1	30.8	16.1	79.1	14.4	71.3	57.1	23.7	77.5	29.5	37.0	4.9	29.6	31.5	37.3
CBST		91.8	53.5	80.5	32.7	21.0	34.0	28.9	20.4	83.9	34.2	80.9	53.1	24.0	82.7	30.3	35.9	16.0	25.9	42.8	45.9
CAG-UDA		90.4	51.6	83.8	34.2	27.8	38.4	25.3	48.4	85.4	38.2	78.1	58.6	34.6	84.7	21.9	42.7	**41.1**	29.3	37.2	50.2
Uncertainty		90.4	31.2	85.1	36.9	25.6	37.5	**48.8**	48.5	85.3	34.8	81.1	64.4	**36.8**	86.3	34.9	52.2	1.7	29.0	44.6	50.3
IAST		93.8	57.8	85.1	39.5	26.7	26.2	43.1	34.7	84.9	32.9	88.0	62.6	29.0	87.3	39.2	49.6	23.2	34.7	39.6	51.5
DACS		89.9	39.7	87.9	30.7	39.5	38.5	46.4	52.8	88.0	44.0	88.8	**67.2**	35.8	84.5	45.7	50.2	0.0	27.3	34.0	52.1
ProDA*		91.5	52.4	82.9	**42.0**	35.7	40.0	44.4	43.3	87.0	43.8	79.5	66.5	31.4	86.7	41.1	52.5	0.0	**45.4**	53.8	53.7
CorDA	✓	94.7	63.1	87.6	30.7	40.6	40.2	47.8	51.6	87.6	47.0	**89.7**	66.7	35.9	**90.2**	48.9	**57.5**	0.0	39.8	**56.0**	56.6
ELDA (Ours)	✓	**94.9**	**64.1**	**88.2**	35.0	**44.7**	**40.3**	47.0	**54.6**	**88.7**	**47.4**	88.9	67.0	31.1	**90.2**	**53.7**	56.0	0.0	41.7	55.5	**57.3**

These tables report the quantitative results evaluated on the GTA5 → Cityscapes and SYNTHIA → Cityscapes benchmarks. *Source Only* corresponds to the the model only trained on the images from the source domain. The distillation stage of ProDA is removed for fair comparison. The results show that ELDA reaches the state-of-the-art performance on both benchmarks.

Qualitative Results

Figure 4. The semantic segmentation results on GTA5→Cityscapes. It is observed that the predictions from ELDA are less fragmented and have more explicit boundaries.

Acknowledgement

MEDIATEK NSTC 國家科學及技術委員會 National Science and Technology Council TWCC TAIWAN COMPUTING CLOUD NVIDIA

Contact us

cylee@cs.nthu.edu.tw
sonya.tw2000@gapp.nthu.edu.tw

ELSA LAB

Our Lab Website