2022/05/06

Unsupervised Domain Adaptation for Semantic Segmentation

Dinh Viet Sang, Nguyen Thi Oanh, Muriel Visani, Kieu Dang Nam, Nguyen Viet Manh, Nguyen Minh Tu
Computer Vision Lab, BKAI

Contents

- Intro to unsupervised domain adaptation
- Project goal
- Distribution alignment and our proposed G2L method
- Experiments
- Conclusions and future works

Unsupervised domain adaptation (UDA)

UDA: transfer knowledge from a ***labeled source dataset*** to ***an unlabeled target dataset***.

- Source data: (x^s, y^s) → Training data
- Target data: (x^t) → Testing data

Training data / Testing data: mismatch

MNIST

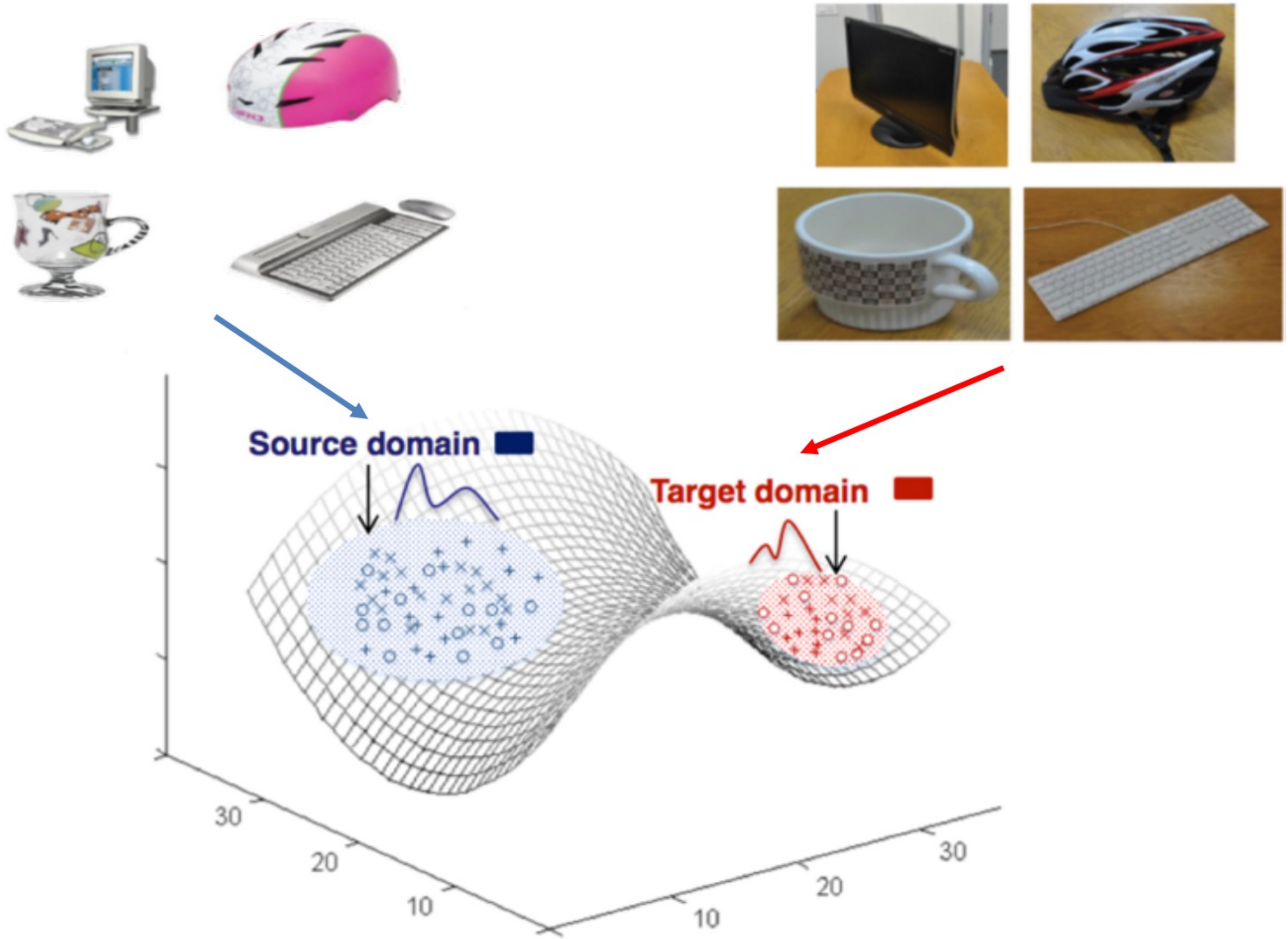
SOURCE — with label

TARGET — without label

MNIST-M

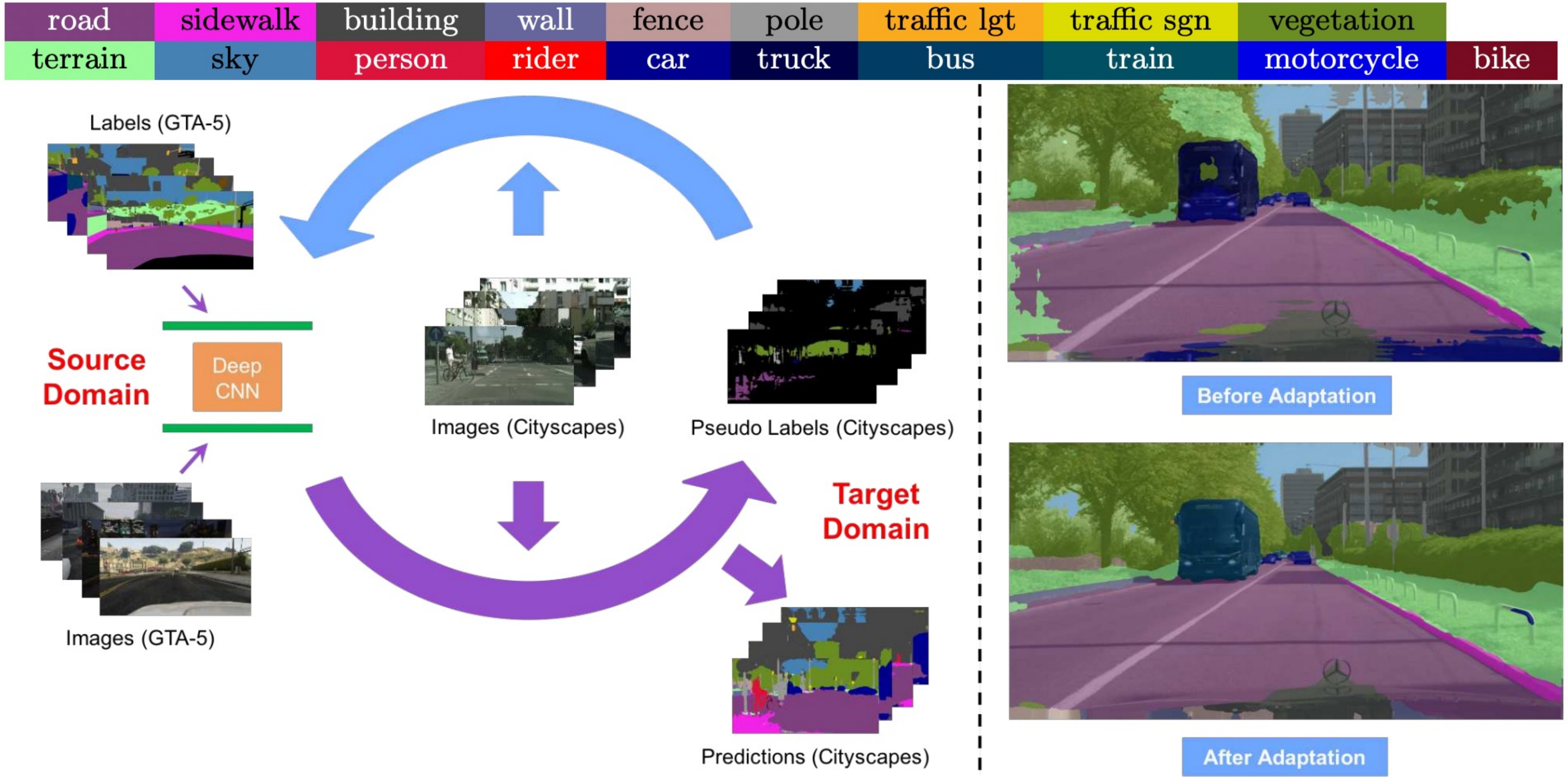
Domain shift

- Performance degradation on the target domain
- **Domain shift**: Discrepancy between the feature distributions of the two domains



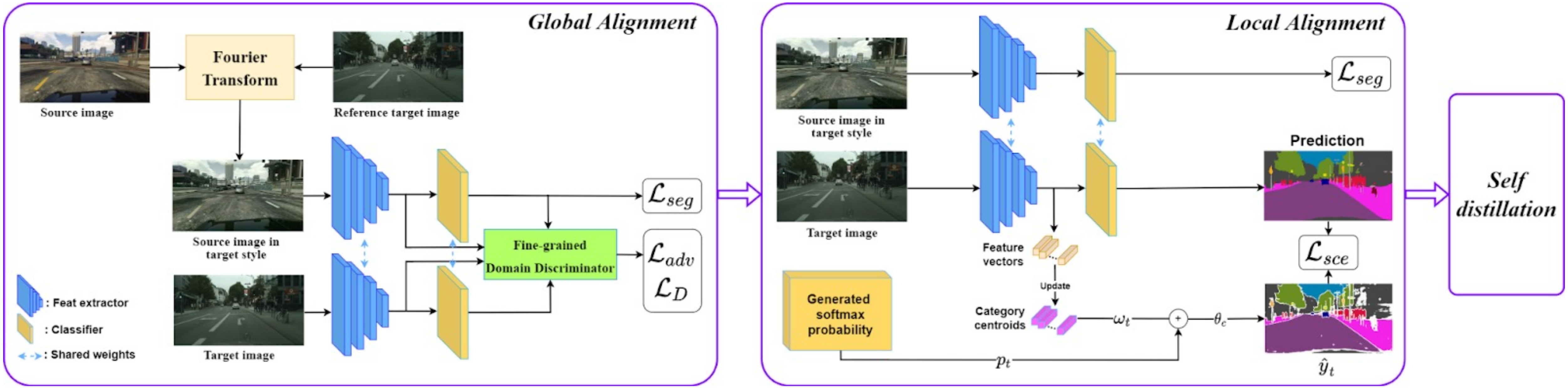
Project goal

- Scene Understanding for Autonomous Driving



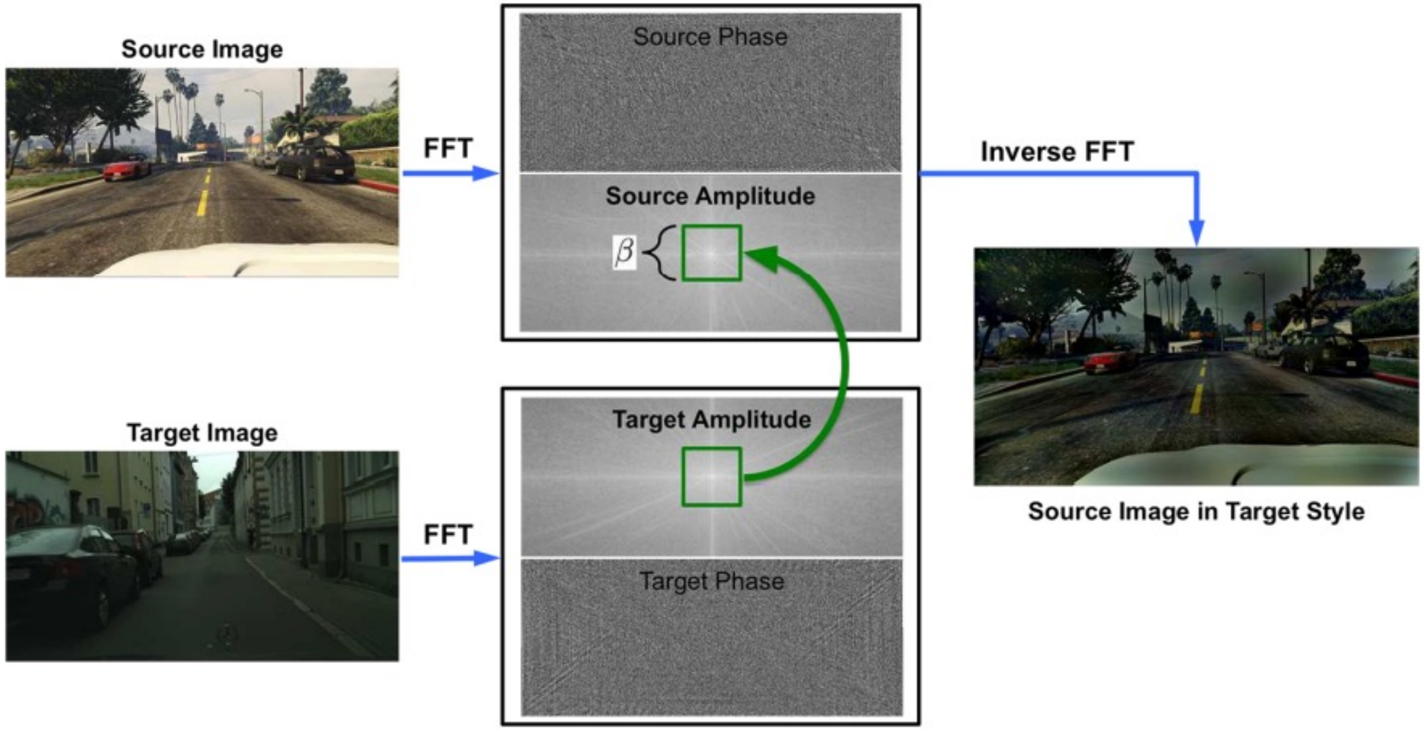
Distribution alignment

- Global alignment: photometric or feature-level
- Local alignment: class-based
- Our **G2L** method: leverages both global and local alignments



Photometric alignment

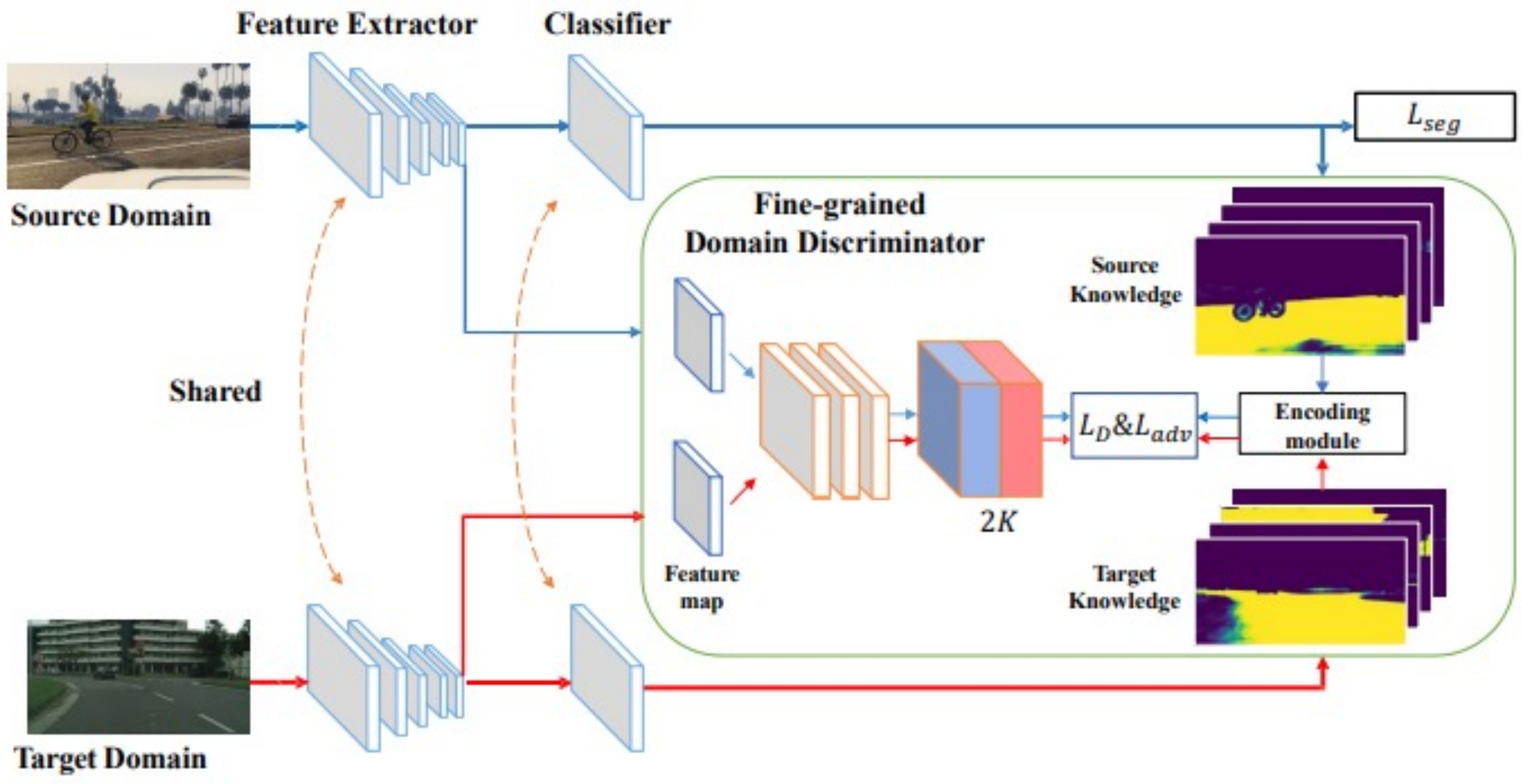
- Style transfer: CyCADA, BLD, DCAN, SA-I2I, FDA, …



FDA: Fourier Domain Adaptation for Semantic Segmentation

Feature alignment

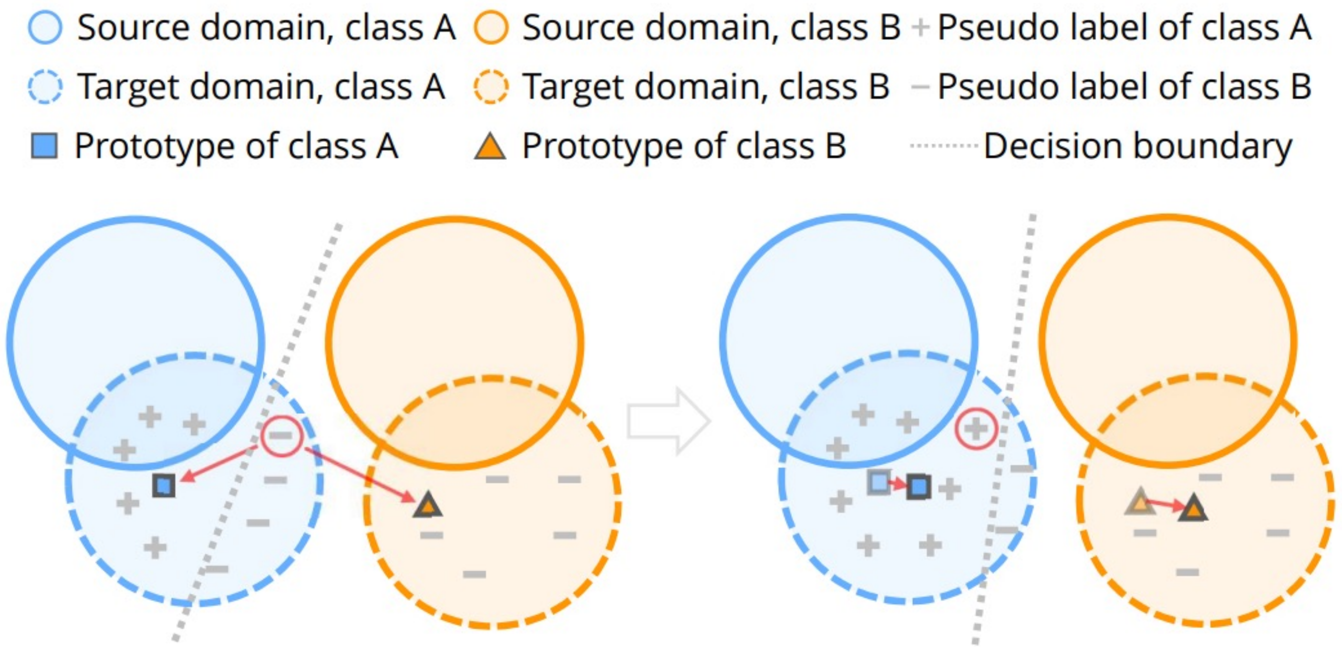
- Fine-grained adversarial learning



[Classes Matter: A Fine-grained Adversarial Approach to Cross-domain Semantic Segmentation]

Self-distillation with local alignment

- Self-distillation: the network teaches itself using generated pseudo labels
- Issue: Pseudo labels are often noisy
- Local alignment: class-wise information based pseudo-label denoising



Experiments

Table 1. Experimental results for GTA5 → Cityscapes adaptation. The numbers show the per-class Intersection over Union (IoU) for all of the 19 categories in Cityscapes. mIoU denotes the averaged scores over 19 categories.

Method	road	sw	build	wall	fen	pole	light	sign	veg	terr	sky	per	rider	car	truck	bus	train	moto	bike	mIoU
FADA [17]	92.5	47.5	85.1	37.6	32.8	33.4	33.8	18.4	85.3	37.7	83.5	63.2	**39.7**	87.5	32.9	47.8	1.6	34.9	39.5	49.2
CAG [24]	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
FDA [22]	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	53.1	16.9	27.7	46.4	50.5
IAST [12]	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 [15]	89.9	39.7	87.9	39.7	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
SAC [1]	90.4	53.9	86.6	42.4	27.3	45.1	48.5	42.7	87.4	40.1	86.1	67.5	29.7	88.5	49.1	54.6	9.8	26.6	45.3	53.8
DSP [5]	92.4	48.0	87.4	33.4	35.1	36.4	41.6	46.0	87.7	43.2	89.8	66.6	32.1	**89.9**	**57.0**	56.1	0.0	44.1	57.8	55.0
Coarse-to-Fine [11]	92.5	58.3	86.5	27.4	28.8	38.1	46.7	42.5	85.4	38.4	**91.8**	66.4	37.0	87.8	40.7	52.4	**44.6**	41.7	**59.0**	56.1
ProDA [23]	87.8	56.0	79.7	46.3	**44.8**	45.6	53.5	53.5	88.6	45.2	82.1	**70.7**	39.2	88.8	45.5	**59.4**	1.0	**48.9**	56.4	57.5
ProDA [23]*	91.5	48.8	85.3	43.0	37.7	45.7	54.8	54.6	89.2	49.8	80.2	69.9	18.0	87.5	29.8	49.8	0.3	42.6	52.2	54.3
ProDA [23]**	84.0	57.3	74.2	43.8	46.0	43.6	55.7	52.9	88.9	49.6	78.8	73.0	41.8	88.8	45.9	65.8	0.0	47.4	53.0	57.4
Ours G2L	**95.8**	**68.8**	**88.0**	**46,5**	37.5	**50.3**	**58.4**	**58.1**	**89.5**	**51.5**	83.1	69.0	33.6	89.6	41.3	**59.4**	36.9	46.9	29.9	**59.7**

Backbone: ResNet-101

**: Results reproduced by us with the similar training config as our G2L method.*

***: Results reproduced by us with the similar training config as our G2L method, using the pre-trained warm-up model, which is published on the page of ProDA.*

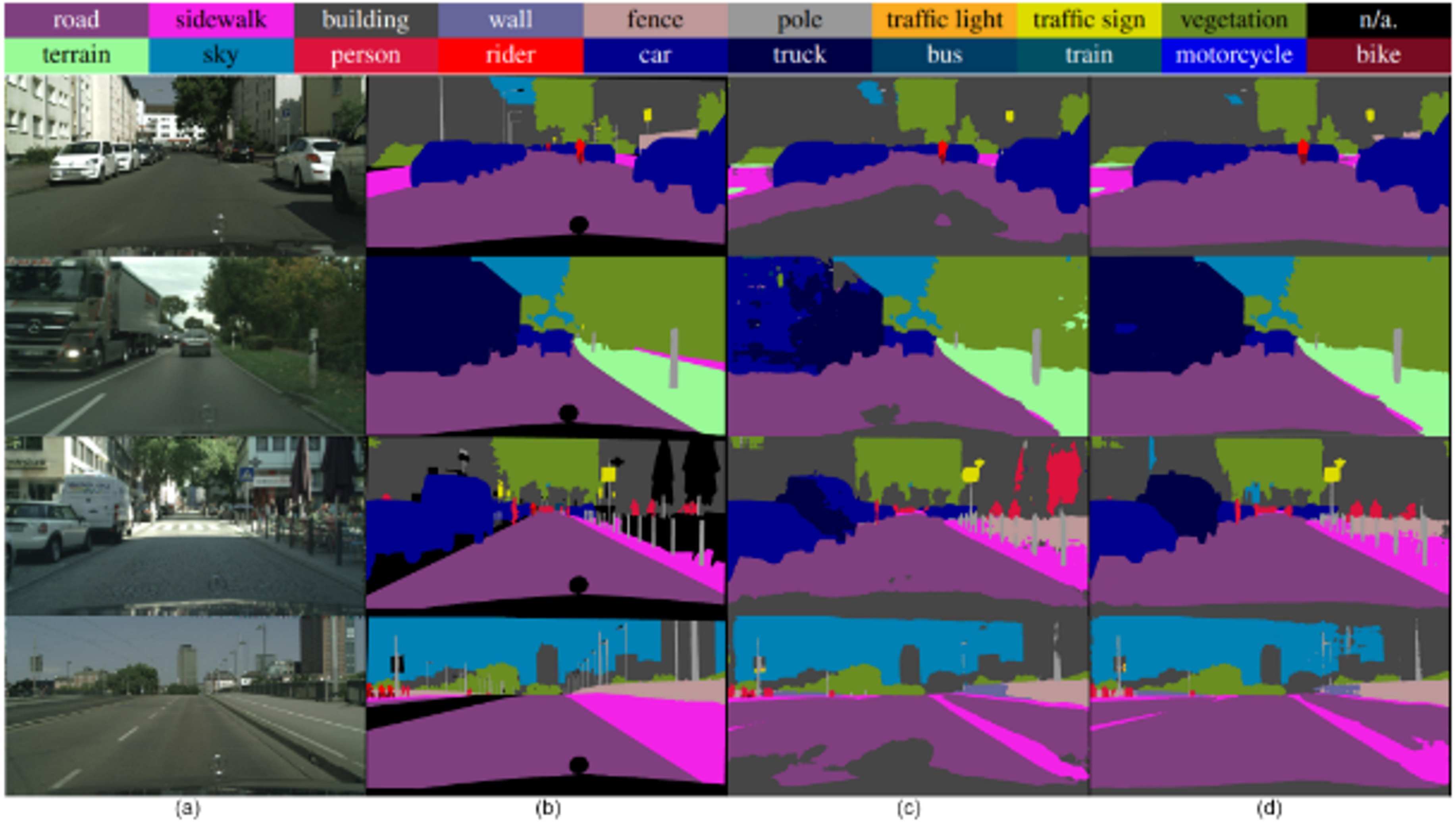
Experiments

Table 2. Experimental results for Synthia → Cityscapes adaptation. The numbers show the per-class Intersection over Union (IoU) of the 16 and 13 categories in Cityscapes. mIoU, and mIoU* denote the averaged scores over 16 and 13 categories respectively.

Method	road	sw	build	wall	fence	pole	light	sign	veg	sky	per	rider	car	bus	moto	bike	mIoU	mIoU*
FADA [17]	84.5	40.1	83.1	4.8	0.0	34.3	20.1	27.2	84.8	84.0	53.5	22.6	85.4	43.7	26.8	27.8	45.2	52.5
CAG (16 classes) [24]	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	—
CAG (13 classes) [24]	84.8	41.7	85.5	—	—	—	13.7	23.0	86.5	78.1	66.3	28.1	81.8	21.8	22.9	49.0	—	52.6
FDA [22]	79.3	35.0	73.2	—	—	—	19.9	24.0	61.7	82.6	61.4	31.1	83.9	40.8	38.4	51.1	—	52.5
IAST [12]	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	57.0
DACS [15]	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	54.8
SAC [1]	**89.3**	**47.2**	**85.5**	26.5	1.3	43.0	45.5	32.0	87.1	89.3	63.6	25.4	86.9	35.6	30.4	53.0	52.6	59.3
DSP [5]	86.4	42.0	82.0	2.1	1.8	34.0	31.6	33.2	87.2	88.5	64.1	31.9	83.8	**65.4**	28.8	54.0	51.0	59.9
Coarse-to-Fine [11]	75.7	30.0	81.9	11.5	2.5	35.3	18.0	32.7	86.2	90.1	65.1	33.2	83.3	36.5	35.3	**54.3**	48.2	55.5
ProDA [23]	87.8	45.7	84.6	**37.1**	0.6	44.0	**54.6**	37.0	**88.1**	84.4	**74.2**	24.3	88.2	51.1	**40.5**	45.6	55.5	62.0
Ours G2L	87.8	45.1	85.1	25.2	1.3	**46.9**	53.3	**46.4**	**88.1**	86.0	72.0	**40.6**	**91.4**	62.9	35.6	41.1	**56.8**	**64.4**

Backbone: ResNet-101

Experiments



*GTA5 → Cityscapes: **(a)** Input images, **(b)** Ground truth, **(c)** Results of ProDA, **(d)** Our G2L method*

Conclusion and future works

- Main achivements
  - Propose a new effective UDA method that leverages both glocal and local alignments
  - Achieve a significant improvement on popular benchmarks
- Other achievements
  - Generate a big synthesized dataset for scene understanding car simulator
  - Evaluate the model on the NAVER LABS dataset
- Future works
  - Online domain adaptation
  - Video-based/depth-guided domain adaptation


Thank You!