厚德 求真 砺学 笃行

西安電子科技大學
XIDIAN UNIVERSITY

Connectivity-Driven Pseudo-Labeling Makes Stronger Cross-Domain Segmenters

Dong Zhao, Qi Zang, Shuang Wang*, Nicu Sebe, Zhun Zhong
School of Artificial Intelligence, Xidian University

*shwang@mail.xidian.edu.cn**

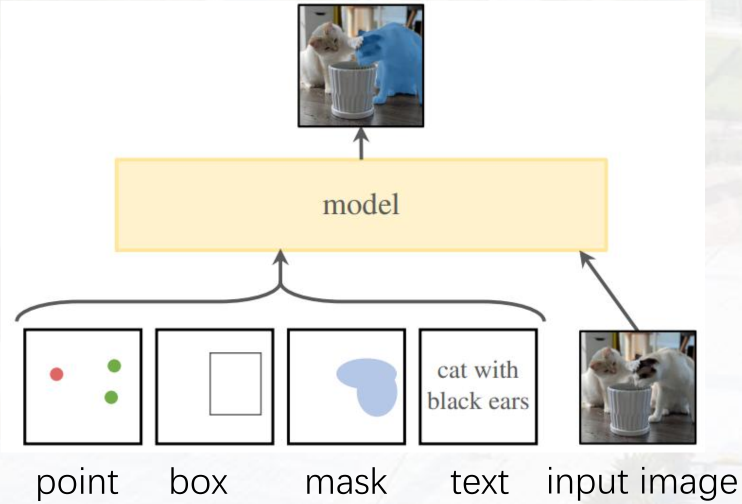
2024.11.13

Table of Contents

01

Research Motivations

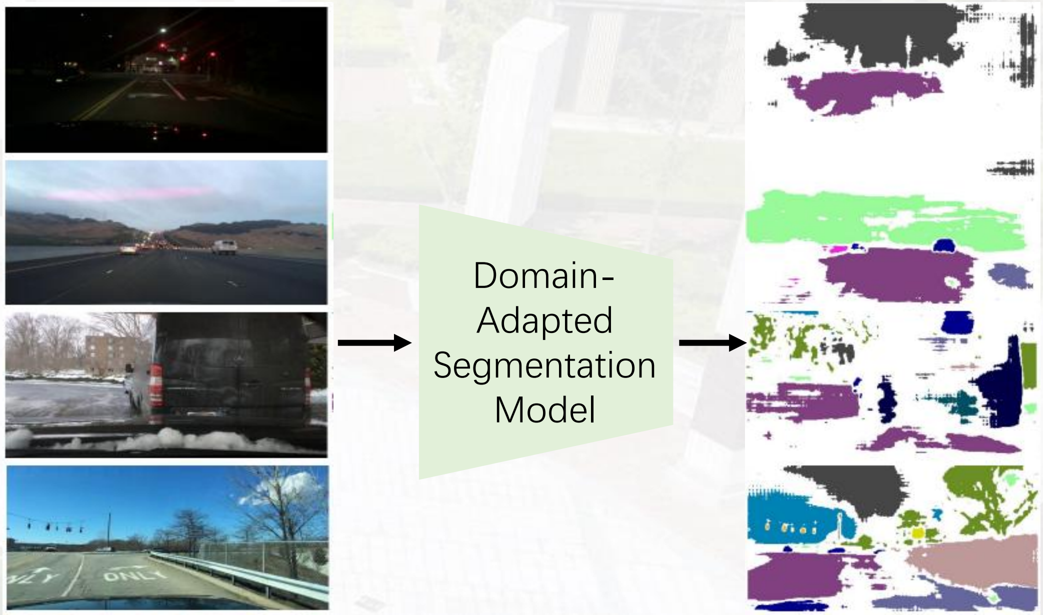
Our Motivations:



Segment anything model (SAM)

Complete mask

Target image+
Prompts built with
pseudo-labels



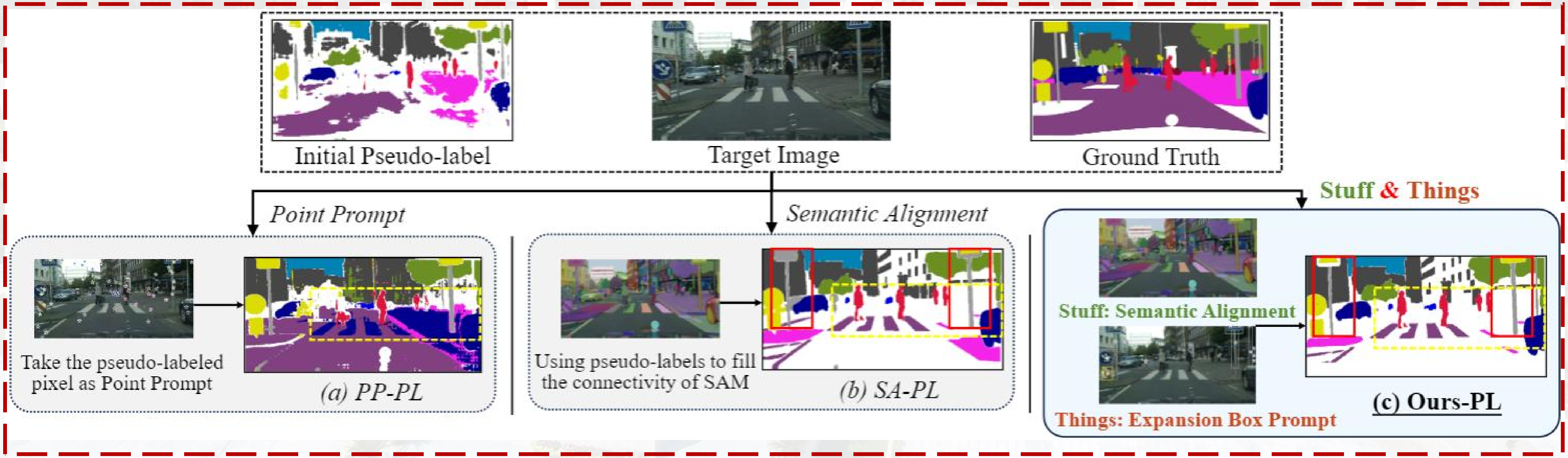
Target image

Domain adapted pseudo-labels

- Large-scale mask pre-training enables SAM' s powerful cross-domain capabilities
- Prompt interactive mode, support point, box, mask, text and other prompts

- There is speckle noise in the pseudo labels and has poor structure.
- Speckle noise is difficult to filter out

Our Motivations:



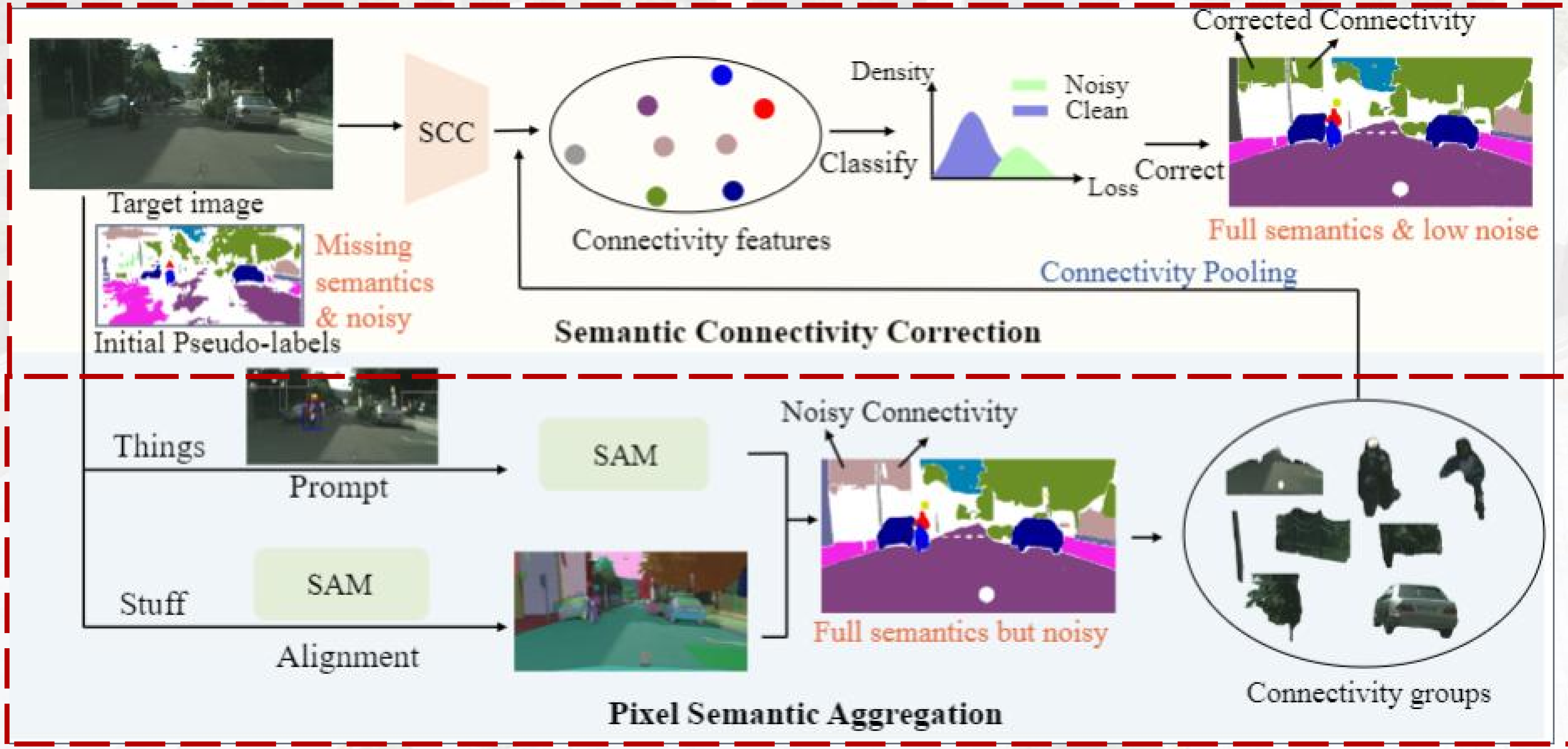
Experiments have shown that directly using the target domain image and its pseudo-label to prompt SAM has the following problems:

(1) Using points to prompt SAM is easily affected by pseudo-label noise, which will amplify the noise.

(2) Semantic alignment using SAM segmentation is prone to semantic confusion.

02

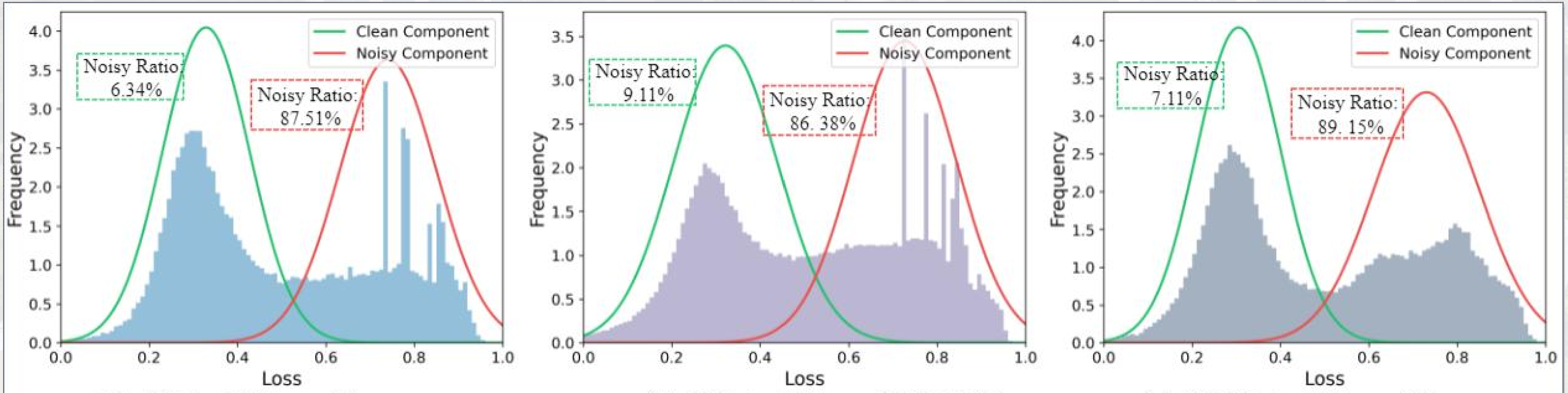
Method Design



- ➢ Proposes connectivity level denoising, using Gaussian distribution to model the category noise of the connected domain
- ✓ Simplifies noise discovery, making it easier to find open-set noise

- ➢ Designs a two-stream prompting method to semantically expand object and background categories respectively
- ✓ Optimizes the way pseudo-labels prompt SAM

03 Method Design



(a). UDA: GTA → Cityscapes (b). UDA: GTA → BDD100K (c). SFUDA: GTA → Cityscapes

- Experiments show that modeling noise distribution at the connectivity level is easier to discover category and open set noise than modeling noise distribution at the pixel level.
- As shown in the figure above, by modeling the loss of the connected domain with a two-component Gaussian mixture model, the low-noise connected domain and the high-noise connected domain can be clearly distinguished.

03

Experimental Results

Comparative Experiment:

- In the task settings of various domain adaptation, the methods in this chapter can improve the performance of existing methods.
- Our method can also be combined with the methods in the previous section to improve their performance.

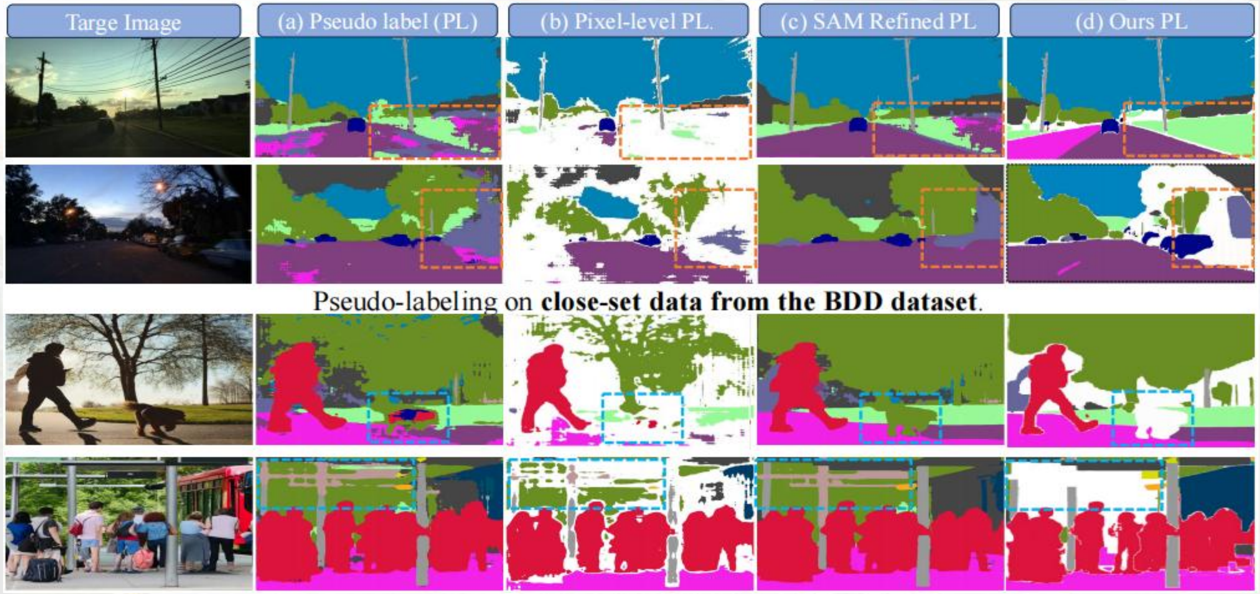
	Road	S.walk	Build.	Wall	Fence	Pole	Tr.Light	sign	Veget.	terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
Unsupervised domain adaptation: GTA → Cityscapes																				
AdvEnt [70] ICCV'19	89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
AdvEnt + Ours	92.0	61.0	87.0	51.0	49.4	48.9	44.5	44.3	86.7	50.0	87.9	63.3	46.0	89.7	57.6	54.6	5.6	47.7	51.6	**58.9** (+13.4)
ProDA [90] CVPR'21	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
ProDA + Ours	94.4	65.6	87.8	55.8	54.7	56.8	58.6	60.3	90.2	51.5	93.7	72.7	48.0	88.1	51.3	65.3	1.5	60.3	61.0	**64.1** (+9.4)
DAFormer [24] CVPR'22	95.7	70.2	89.4	53.5	48.1	49.6	55.8	59.4	89.9	47.9	92.5	72.2	44.7	92.3	74.5	78.2	65.1	55.9	61.8	68.2
DAFormer+Ours	96.2	74.4	90.9	56.7	49.7	60.5	62.7	69.4	92.4	54.9	93.9	77.1	53.1	96.6	83.1	82.2	72.5	62.6	65.6	**73.4** (+5.3)
HRDA [25] ECCV'22	96.4	74.4	91.0	61.6	51.5	57.1	63.9	69.3	91.3	48.4	94.2	79.0	52.9	93.9	84.1	85.7	75.9	63.9	67.5	73.8
HRDA+Ours	96.6	80.9	92.4	62.5	57.5	61.0	66.7	71.7	92.4	52.3	95.1	80.6	56.3	95.9	86.1	86.6	76.8	65.4	68.7	**76.1** (+2.3)
Source-free domain adaptation: GTA → Cityscapes																				
HCL [28] NIPS'21	92.0	55.0	80.4	33.5	24.6	37.1	35.1	28.8	83.0	37.6	82.3	59.4	27.6	83.6	32.3	36.6	14.1	28.7	43.0	48.1
HCL+Ours	94.6	62.5	88.6	48.4	41.6	45.2	43.5	32.9	84.0	45.3	91.6	66.0	47.5	89.0	42.6	58.8	31.5	47.2	56.2	**58.8** (+10.6)
DTST [93] CVPR'23	90.3	47.8	84.3	38.8	22.7	32.4	41.8	41.2	85.8	42.5	87.8	62.6	37.0	82.5	25.8	32.0	29.8	48.0	56.9	52.1
DTST+Ours	94.9	65.9	89.9	48.2	42.3	45.9	48.9	45.6	85.7	46.2	91.1	68.2	47.6	88.5	44.9	57.8	29.5	50.7	57.8	**60.5** (+8.4)
Black-box domain adaptation: GTA → Cityscapes																				
DINE [43] CVPR'22	88.2	44.2	83.5	14.1	32.4	23.5	24.6	36.8	85.4	38.3	85.3	59.8	27.4	84.7	30.1	42.2	0.0	42.7	45.3	46.7
DINE+Ours	89.6	60.8	84.1	46.3	38.4	44.0	41.6	32.2	82.1	41.7	86.6	63.4	44.9	83.9	41.5	58.6	0.0	40.5	54.1	**54.4** (+7.7)
BiMem [89] ICCV'23	94.2	59.5	81.7	35.2	22.9	21.6	10.0	34.3	85.2	42.4	85.0	56.8	26.4	85.6	37.2	47.4	0.2	39.9	50.9	48.2
BiMem+Ours	93.9	61.4	87.6	47.7	41.3	44.0	43.2	32.7	83.2	44.4	91.4	66.9	46.6	88.7	42.6	60.8	0.0	46.2	55.0	**56.7** (+8.5)

Comparative Experiment:

- Our method can provide a new approach to solving the domain generalization problem, which can alleviate the semantic and open set noise problems encountered by domain generalization methods when using unlabeled open data.

	Backbone	Using SAM	Cityscapes	BDD-100K	Mapillary	Average
SHADE [96] IJCV'23	ResNet-101	✗	46.6	43.7	45.5	45.3
TLDR [35] ICCV'23		✗	47.6	44.9	48.8	47.1
MoDify [32] ICCV'23		✗	48.8	44.2	47.5	46.8
+ CLOUDS [4] CVPR'24		✓	50.6	44.8	56.6	50.7
+ SeCo (Ours)		✓	**52.4**	**46.1**	**57.7**	**52.1**
HRDA [25] ECCV'22	MiT-B5	✗	57.4	49.1	61.1	55.9
+ CLOUDS [4] CVPR'24		✓	58.1	53.8	62.3	58.1
+ SeCo (Ours)		✓	**58.8**	**54.9**	**63.6**	**59.1**

SourceGTA5→	Compound Rainy	Compound Snowy	Compound Cloudy	Open Overcast	Average Compound + Open Overcast
Source Only	28.7	29.1	33.1	32.5	30.9
Unspervised domain adaptation: GTA5 → BDD-100k					
ML-BPM [56] ECCV'22	40.5	39.9	42.1	40.9	40.9
OSC [15] NIPS'23	-	-	-	-	44.0
PyCDA [46] CVPR'20	33.4	32.5	36.7	37.8	35.1
PyCDA + Ours	43.6	42.1	49.7	50.7	**46.5** (+11.4)
ProDA [90] CVPR'21	40.3	40.6	43.2	42.5	41.7
ProDA + Ours	47.6	45.7	51.9	52.6	**49.5** (+7.8)
Source-free domain adaptation: GTA5 → BDD-100k					
SFOCDA [95] TCSVT'22	35.4	33.4	41.4	41.2	37.9
SFOCDA + Ours	41.7	42.1	44.7	47.9	**44.1** (+6.2)



Comparative Experiment:

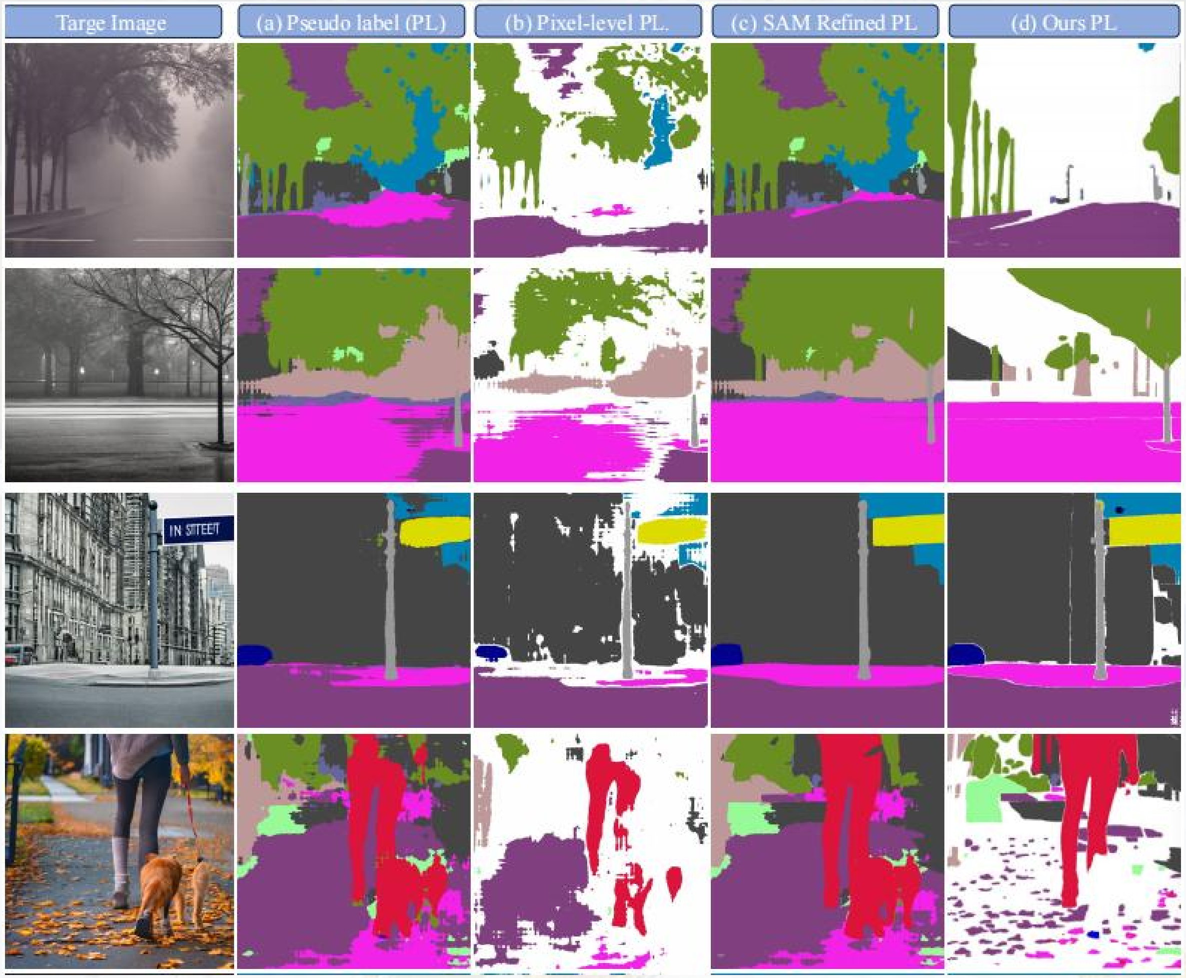
- Compared with various methods of using Segment Anything Model, including using SAM as pre-training and using SAM to optimize pseudo-labels, the method in this chapter shows advantages.

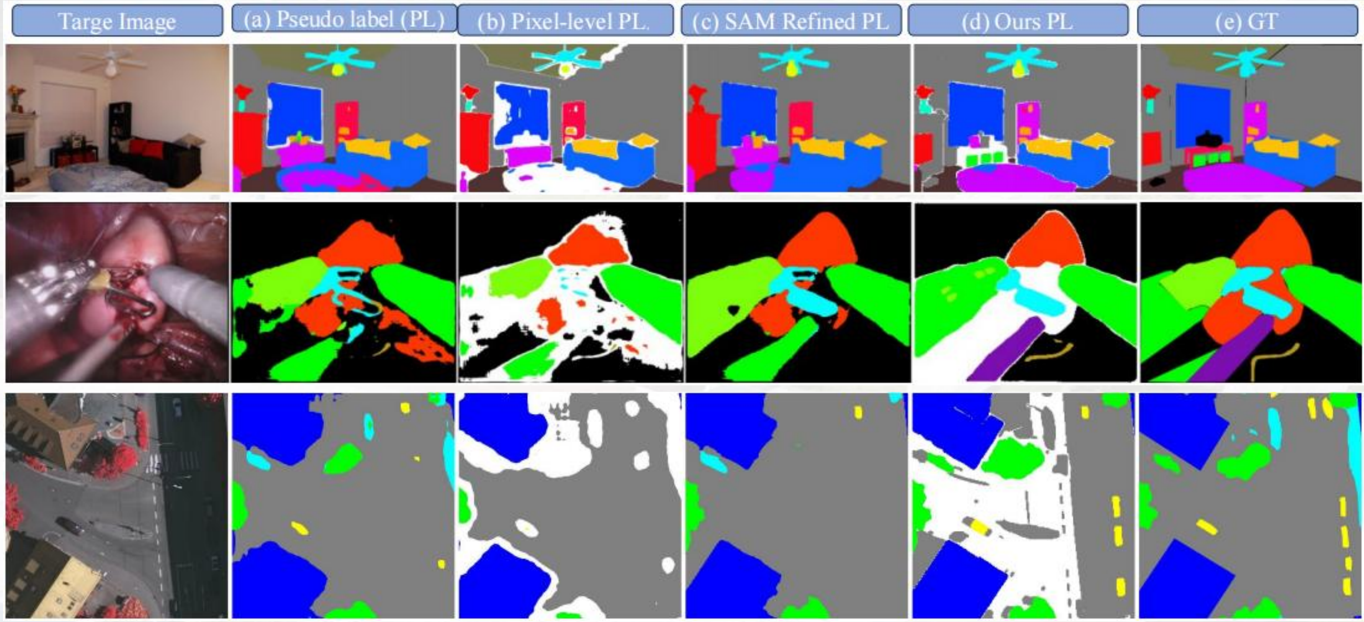
① Use the backbone of SAM to empower CDSS						
	DAFormer	+ (SAM ViT-B)	+ Ours	HRDA	+ (SAM ViT-B)	+ Ours
G2C	68.2	64.1 (-4.1)	**73.4** (+5.2)	73.8	69.1 (-4.7)	**76.1** (+2.3)
S2C	60.9	60.2 (-0.7)	**65.1** (+4.2)	65.8	63.7 (-2.1)	**68.5** (+2.3)

② Use CLIP + SAM (CSAM) to get the initial pseudo-label					
	CSAM	CSAM + DAFormer	Ours + DAFormer	CSAM + HRDA	Ours + HRDA
G2C	43.7	69.1	**73.4**	73.5	**76.1**
S2C	41.7	61.1	**65.1**	65.2	**68.5**

③ Use SAM on coarse pseudo-labels						
Method	Initial	Using Source-model's PL		Using UDA's PL		Upper Bound
		Vanilla SAM	Ours	Vanilla	Ours	
AdvEnt [70] ICCV'19	45.5	48.6 (+3.1)	53.9 (+8.4)	50.9 (+5.5)	**58.9 (+13.4)**	69.1
ProDA [90] CVPR'21	53.7	50.1 (-1.7)	58.2 (+4.5)	57.9 (+4.3)	**64.1 (+9.4)**	69.1
DAFormer [24] CVPR'22	68.2	67.7 (-0.5)	69.9 (+1.7)	69.7 (+1.5)	**73.4 (+5.3)**	76.4
HRDA [25] ECCV'22	73.8	72.7 (-1.1)	74.6 (+0.8)	74.6 (+0.8)	**76.1 (+2.3)**	77.1

Qualitative Analysis:





- In terms of the use of unlabeled open data, the method in this chapter shows that it can alleviate the semantic and open set noise problems.
- In indoor scenes, medical, and remote sensing scenes, the methods in this chapter have shown that they can improve the pseudo-label quality of existing domain adaptation methods.

厚德 求真 砺学 笃行

THANKS

Thanks!

QUESTIONS & ANSWERS

西安电子科技大学