厚德 求真 砺学 笃行

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

西安電子科技大學
XIDIAN UNIVERSITY

Towards Better Stability and Adaptability: Improve Online Self-Training for Model Adaptation in Semantic Segmentation

Dong Zhao, Shuang Wang*, Qi Zang, Dou Quan, Xiutiao Ye, Licheng Jiao

School of Artificial Intelligence, Xidian University

shwang@mail.xidian.edu.cn*

2023.05.28

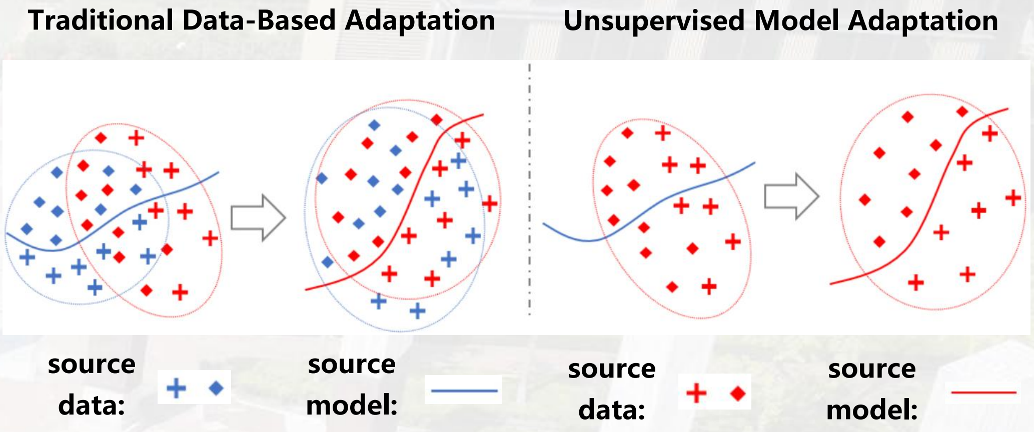
Table of Contents

01

Summary of the Paper

Our Motivations:

- **Existing Unsupervised Model Adaptation (UMA) methods adopt offline self-training (OST) but it requires expert intervention.**
- **Online self-training (ONST) avoids the drawback of OST by online co-evolving pseudo-labels, showing potential in unsupervised domain adaptation (UDA) with accessing source data.**



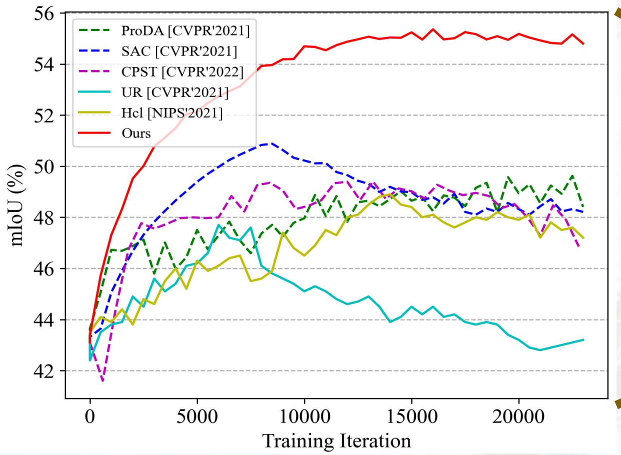


Figure. The dashed line is the ONST methods of UDA, and the solid line is the UMA methods.

- ONST is more competitive than existing UMA methods;
- ! But ONST methods suffer from impaired ***stability*** and ***adaptability*** in UMA.

- ***How to apply ONST to UMA without accessing source data?***

Contributions:

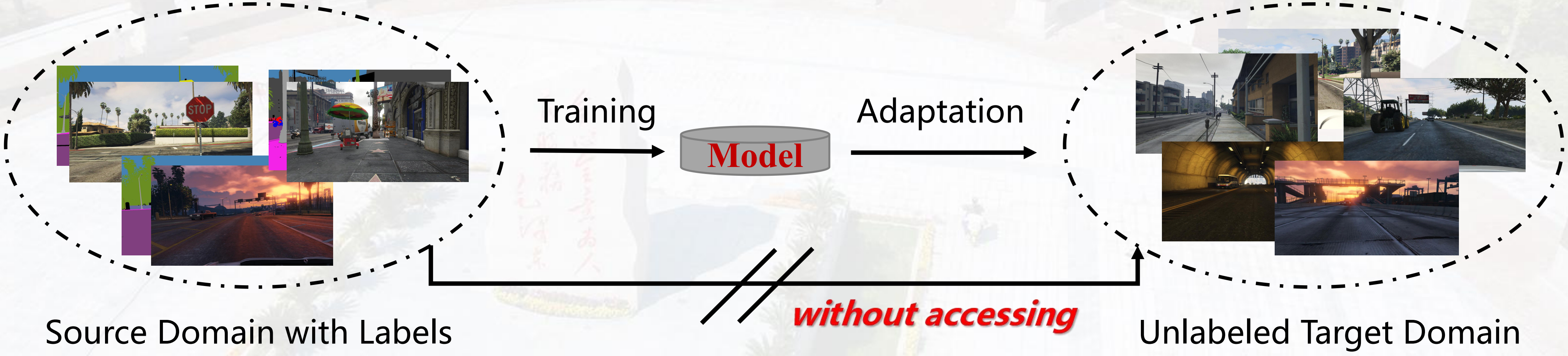
- We explore ***two reasons*** for the poor stability and adaptability of ONST in UMA: (1) the inopportune update of the teacher model; (2) the bias towards minority classes .
- For (1), we propose a Dynamic Teacher Update mechanism, which dynamically ***controls the update interval*** of the teacher model.
- For (2), we propose a Training-Consistency based Resampling strategy, which adaptively ***estimates the biased classes*** and resampling.

02

Research Motivations

Task Setting:

- **Unsupervised domain adaptation (UDA) in semantic segmentation transfers the knowledge of the source domain to the target one to improve the adaptability of the segmentation model in the target domain.**
- **However, the need to access labeled source data makes UDA unable to handle adaptation scenarios involving privacy, property rights protection, and confidentiality.**



- **The setting of Unsupervised Model Adaptation (UMA) is proposed, aiming to adapt the source-trained model to the unlabeled target domain without using source domain data.**

Problem:

- **Existing UMA methods adopt offline self-training (OST) that iteratively updates the pseudo-labels to retrain the models but it requires expert intervention.**
- **Online self-training (ONST) avoids the drawback of OST by online co-evolving pseudo-labels, showing potential in unsupervised domain adaptation (UDA) with accessing source data.**

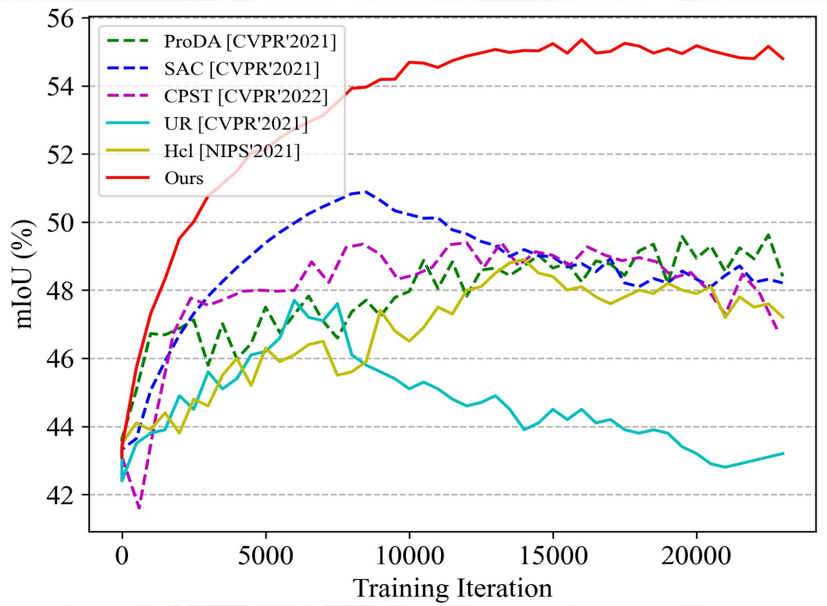


Figure 1. The dashed line is the ONST methods of UDA, and the solid line is the UMA methods.

Observation

- **ONST is more competitive than existing UMA methods;**
- **! But ONST methods suffer from impaired stability and adaptability in UMA.**

Question

- ***How to apply ONST to UMA without accessing source data?***

03

Method Design

Analyze:

- To apply ONST to UMA, we explore ***two reasons*** for the poor stability and adaptability of ONST in UMA:
- (1) the **inopportune update** of the teacher model;
- (2) the **bias towards minority classes** in the source-trained model.

Method Proposed:

- For (1), we propose a **Dynamic Teacher Update** mechanism, which dynamically controls the update interval of the teacher model.
- For (2), we propose a **Training-Consistency based Resampling** strategy, which adaptively estimates the biased classes and resampling.

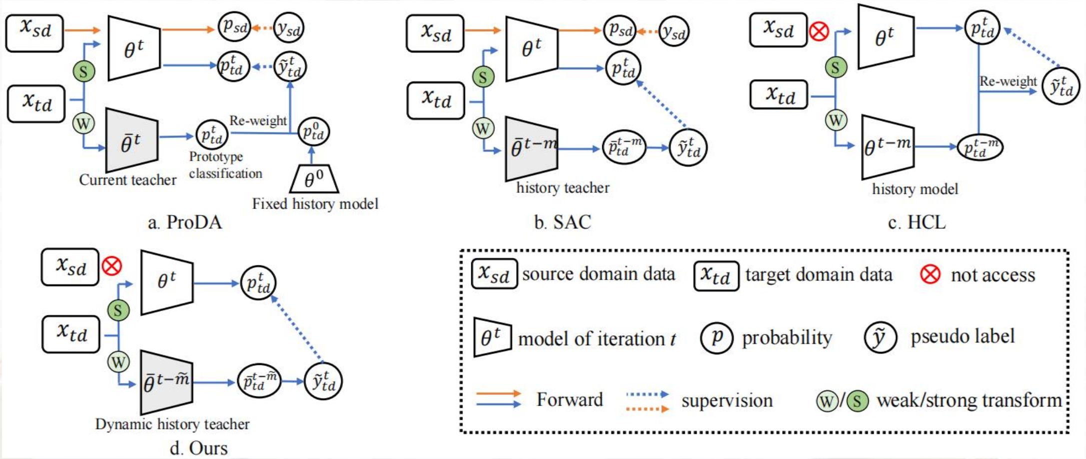


Figure 2. Schematics of different online self-training methods.

- In Fig.2 b and Fig.2 c, m denotes a **fixed update interval**;
- In Fig.2 d, $\tilde{m}$ denotes a **dynamic update interval**.

Task Background:

- Symbol definition:

➢ Let $D_{sd} = \{(x_{sd}, y_{sd})\}$ be the labeled source domain data, $D_{td} = \{x_{td}^n\}_{n=1}^N$ be the unlabeled target domain data;

➢ The D_{sd} and D_{td} share K categories;

➢ Let G and θ be the source-trained segmentation model and its parameters.

- Online Self-training (ONST) in UDA:

➢ Generally, in ONST, given a student model G_{stu} and an online updated teacher model G_{tea}. For G_{stu}, two losses are used for supervision:

$$\text{on source domain} \rightarrow L_{sd} = H[x_{sd}, y_{sd}],$$

$$\text{on target domain} \rightarrow L_{sd} = H\left[G_{stu}^{\theta_t}(W(x_{td})), G_{tea}^{\bar{\theta}_{t-m}}(S(x_{td}))\right],$$

G^{θ_t} is the model with parameters θ_t, m is the update interval of G_{tea}, t is the number of iterations.

➢ For G_{tea}, its parameters are the G_{stu}' s momentum-updated version:

$$\bar{\theta}_t = (1-\gamma)\theta_t + \gamma\bar{\theta}_{t-m},$$

γ is the update weights.

Dynamic Teacher Update (DTU):

- The **gain rate** of the student model over historical samples is **positively correlated** with stable **co-evolution**.

➢ gain rate (GR): $\frac{1}{T}\sum_t^T \delta\left(V\left[G_{stu}^{\theta_{t-m}}(D_{his})\right], V\left[G_{stu}^{\theta_t}(D_{his})\right]\right)$, $\delta(\cdot)$ is a comparison function, $V[\cdot]$ is a evaluation function.

➢ IEGR: $V[\cdot]$ is ***information entropy***; SNDGR: $V[\cdot]$ is ***soft neighborhood density***.

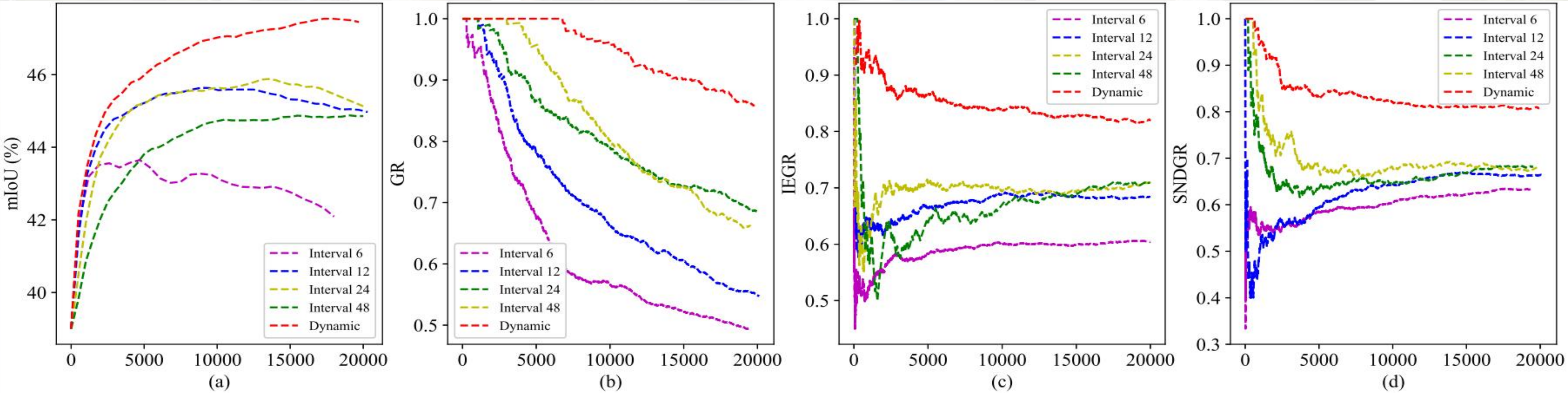


Figure 3. Training curves using different teacher update intervals.

- So $V[\cdot]$ can measure **whether the student model is evolved** to dynamically control update interval m.

Training-Consistency based Resampling (TCR):

- Online average class score is calculated from the output probability of the teacher model:

$$\mathrm{ACS}_k^t = \frac{1}{hw}\sum_{i,j} p_{i,j,k}^t,$$

p^t is the probability map generated by the $G_{tea}^{\bar{\theta}_t}$. The low-confidence classes in ACS are considered to be minorities.

- We then use ACS to determine the sampling rate for k-th class:

$$\mathrm{SR}_k = Normalize(1 - \mathrm{ACS}_k).$$

SR of the biased classes is larger!

- We regard the prediction consistency of as reliability to build reliable candidates for copy-paste resampling:

$$\mathrm{ReL}^n = \mathrm{IoU}\left[\aleph\left(G_{tea}^{\bar{\theta}_{t-1}}(x_{td}^n)\right), \aleph\left(G_{tea}^{\bar{\theta}_t}(x_{td}^n)\right)\right].$$

➢ For each class k, we sort target images according to ReL_k, and take the samples from the top C as the reliable candidates.

04

Experimental Results

Comparative Experiment:

- **We use pre-trained models from two simulation datasets GTA5 and SYNTHIA, and adapt them to real scenes Cityscapes and BDD-100k datasets.**

	SF	road	sidewalk	Building	Wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
IAST (ECCV'2020) [40]	✗	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
MetaCorr (CVPR'2021) [14]	✗	92.8	58.1	86.2	39.7	33.1	36.3	42.0	38.6	85.5	37.8	87.'6	62.8	31.7	84.8	35.7	50.3	2.0	36.8	48.0	52.1
ProDA (CVPR'2021) [59]	✗	91.5	52.4	82.9	42	35.7	40	44.4	43.3	87	43.8	79.5	66.5	31.4	86.7	41.1	52.5	0	45.4	53.8	53.7
SAC (CVPR'2021) [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
CPST(CVPR'2022) [31]	✗	91.7	52.9	83.6	43	32.3	43.7	51.3	42.8	85.4	37.6	81.1	69.5	30	88.1	44.1	59.9	24.9	47.2	48.4	55.7
Source model		65.0	16.1	68.7	18.6	16.8	21.3	31.4	11.2	83.0	22.0	78.0	54.4	33.8	73.9	12.7	30.7	13.7	28.1	19.7	36.8
URMDA (CVPR'2021) [9]	✓	92.3	55.2	81.6	30.8	18.8	37.1	17.7	12.1	84.2	35.9	83.8	57.7	24.1	81.7	27.5	44.3	6.9	24.1	40.4	45.1
SFDA (CVPR'2021) [36]	✓	91.7	52.7	82.2	28.7	20.3	36.5	30.6	23.6	81.7	35.6	84.8	59.5	22.6	83.4	29.6	32.4	11.8	23.8	39.6	45.8
SDF (MM'2021) [57]	✓	95.2	40.6	**85.2**	30.6	26.1	35.8	34.7	32.8	85.3	41.7	79.5	61.0	28.2	86.5	41.2	45.3	15.6	33.1	40.0	49.4
HCL (NIPS'2021) [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
DT-ST (Ours)	✓	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
Source model + DG [32]		80.2	30.2	79.6	30.7	20.3	31.9	36.1	18.6	80.6	23.9	75.2	63.0	36.2	84.8	31.2	36.1	4.4	31.2	28.0	43.3
ProDA† (CVPR'2021) [59]	✓	85.6	45.4	76.5	40.1	31.9	38.9	36.4	**47.4**	85.8	45.7	80.1	63.6	0	85.6	33.7	**51.2**	0	37.6	52.3	49.4
SAC† (CVPR'2021) [1]	✓	89.1	52.7	82.1	**40.3**	26.7	40.7	44.1	40.1	81.6	40.1	81.6	**67.4**	26.1	85.1	44.5	48.8	3.8	26.4	43.1	50.8
CPST† (CVPR'2022) [31]	✓	86.7	38.6	82.2	39.8	**32.1**	**40.8**	41.5	43.2	85.6	42.7	73.6	65.5	22.1	87.3	27.1	41.1	0	37.6	49.5	49.3
HCL (NIPS'2021) [21]	✓	92.6	54.6	82.8	33.2	26.2	39.8	38.1	31.9	84.5	38.6	85.3	61.3	30.2	85.4	33.1	41.6	14.4	27.3	44.0	49.7
DT-ST (Ours)	✓	**93.5**	**57.6**	84.7	36.5	25.2	33.4	**44.7**	36.7	**86.8**	**42.8**	81.3	62.3	**37.2**	**88.1**	**48.7**	50.6	**35.5**	**48.3**	**59.1**	**55.4**

Table 1. Experimental results for GTA5 → Cityscapes.

	SF	road	sidewalk	Building	Wall	fence	pole	light	sign	vege.	sky	person	rider	car	bus	mbike	bike	mIoU	mIoU*
IAST (ECCV'20) [40]	✗	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
MetaCorr (CVPR'21) [14]	✗	92.6	52.7	81.3	8.9	2.4	28.1	13.0	7.3	83.5	85.0	60.1	19.7	84.8	37.2	21.5	43.9	45.1	52.5
ProDA (CVPR'2021) [59]	✗	87.1	44	83.2	26.9	0.7	42	45.8	34.2	86.7	81.3	68.4	22.1	87.7	50	31.4	38.6	51.9	58.5
SAC (CVPR'2021) [1]	✗	89.3	47.2	85.5	26.5	1.3	43	45.5	32	87.1	89.3	63.6	25.4	86.9	35.6	30.4	53	52.6	59.3
CPST (CVPR'2022) [31]	✗	87.3	44.4	83.8	25.0	0.4	42.9	47.5	32.4	86.5	83.3	69.6	29.1	89.4	52.1	42.6	54.1	54.4	61.7
Source model		52.2	23.6	62.2	6.0	0.2	28.3	7.3	12.7	79.7	75.7	52.5	10.2	75.0	24.6	8.9	10.3	33.1	38.1
URMDA (CVPR'2021) [9]	✓	59.3	24.6	77	14	**1.8**	31.5	18.3	32	83.1	80.4	46.3	17.8	76.7	17	18.5	34.6	39.6	45
SFDA (CVPR'2021) [57]	✓	67.8	31.9	77.1	8.3	1.1	35.9	21.2	26.7	79.8	79.4	58.8	27.3	80.4	25.3	19.5	37.4	42.4	48.7
SDF (MM'2021) [57]	✓	**90.9**	45.5	80.8	3.6	0.5	28.6	8.5	26.1	83.4	83.6	55.2	25	79.5	32.8	20.2	43.9	44.2	51.9
HCL (NIPS'2021) [21]	✓	80.9	34.9	76.7	6.6	0.2	36.1	20.1	28.2	79.1	83.1	55.6	25.6	78.8	32.7	24.1	32.7	43.5	50.2
DT-ST (Ours)	✓	79.4	41.4	73.9	5.9	1.5	30.6	35.3	19.8	**86.0**	**86.0**	63.8	**28.6**	86.3	36.6	35.2	53.2	47.7	55.8
Source model + DG [32]		76.8	29.8	67.9	10.7	0.3	29.5	9.5	16.8	79.8	78.3	52.5	13.8	78.5	28.5	12.8	19.9	37.8	43.5
ProDA† (CVPR'2021) [59]	✓	79.9	35.7	75.5	**20.7**	0	**39.6**	36.5	**31.5**	84.2	80.6	64.2	9.6	85.3	40.9	24.9	35.8	46.6	52.7
SAC† (CVPR'2021) [1]	✓	84.7	39.6	80.9	16.3	0.2	38.4	**40.9**	27.4	82.5	84.7	59.1	16.6	82.3	31	20.8	36.1	46.3	52.8
CPST† (CVPR'2022) [31]	✓	80.9	28.7	81	20.4	1.2	38.6	36.3	31.4	85.3	74.4	64.2	12.6	87.2	31.9	16.3	42.8	45.8	51.8
HCL (NIPS'2021) [21]	✓	86.7	38.1	82.7	10.0	0.6	30.3	25.4	29.7	82.8	85.9	61.9	24.8	84.5	38.9	22.6	37.9	46.4	54.0
DT-ST (Ours)	✓	88.9	**45.8**	**83.3**	13.7	0.8	32.7	31.6	20.8	85.7	82.5	**64.4**	27.8	**88.1**	**50.9**	**37.6**	**57.3**	**50.7**	**58.8**

Table 2. Experimental results for SYNTHIA → Cityscapes.

Source GTA→	SF	Compound			Open	Avg	
		Rainy	Snowy	Cloudy	Overcast	C	C+O
Source Only		19.7	18.4	20.5	22.5	19.7	21.0
CBST [63]	✗	21.3	20.6	23.9	24.7	22.2	22.6
IBN-Net [42]	✗	20.6	21.9	26.1	25.5	22.8	23.5
PyCDA [34]	✗	21.7	22.3	25.9	25.4	23.3	23.8
OCDA [37]	✗	22.0	22.9	27.0	27.9	24.5	25.0
MOCDA [12]	✗	24.4	27.5	30.1	31.4	27.7	29.4
HCL [21]	✓	22.8	25.8	28.6	27.7	25.9	26.2
DT-ST (Ours)	✓	**26.7**	**28.1**	**32.1**	**32.5**	**30.1**	**31.3**

Table 3. Experimental results for GTA5 → BDD-100k.

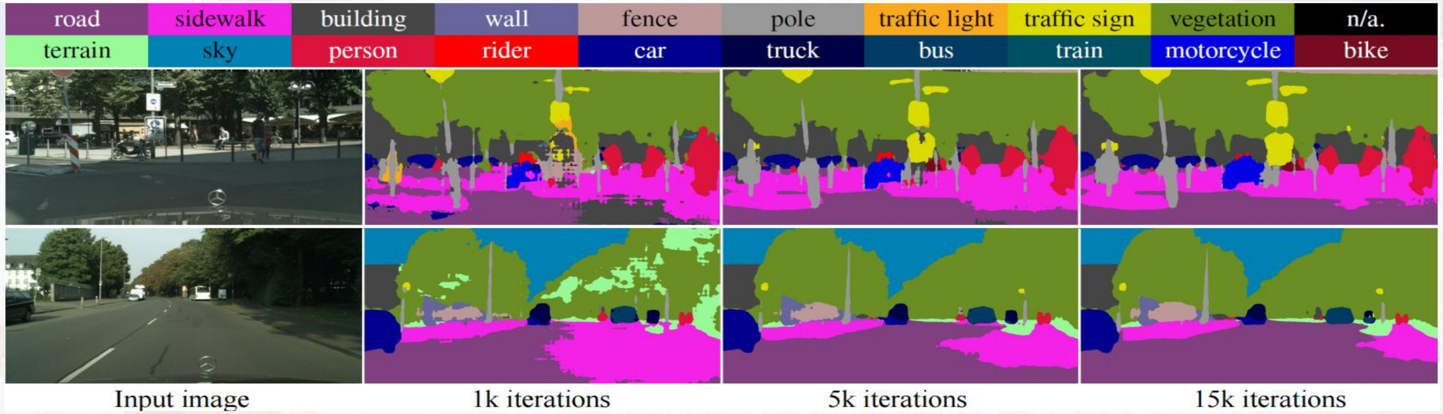


Figure 4. Visualization of output from the teacher model for different iterations.

Ablation Studies and Sensitivity Analysis:

source-trained	Base ST	DTU-E	DTU-SND	TCR-*Prob*	TCR	mIoU	gain
original						36.8	
	✓					46.2	+9.4
	✓	✓				47.2	+10.4
	✓		✓			47.8	+11.0
	✓			✓		49.3	+12.5
	✓				✓	50.6	+13.8
	✓		✓	✓		50.2	+11.0
	✓		✓		✓	**52.3**	**+15.5**
DG [32]						43.3	
	✓					50.7	+7.4
	✓	✓				51.3	+8.0
	✓		✓			52.7	+9.4
	✓			✓		52.8	+9.5
	✓				✓	54.4	+11.1
	✓		✓	✓		53.8	+10.5
	✓		✓		✓	**55.4**	**+12.1**

Table 4. We report mIoU scores (%) (val) using two source-trained models on GTA5 → Cityscapes task Under UMA setting. TCR-*Prob* denotes a strong copy-paste baseline using softmax entropy as reliability metrics.

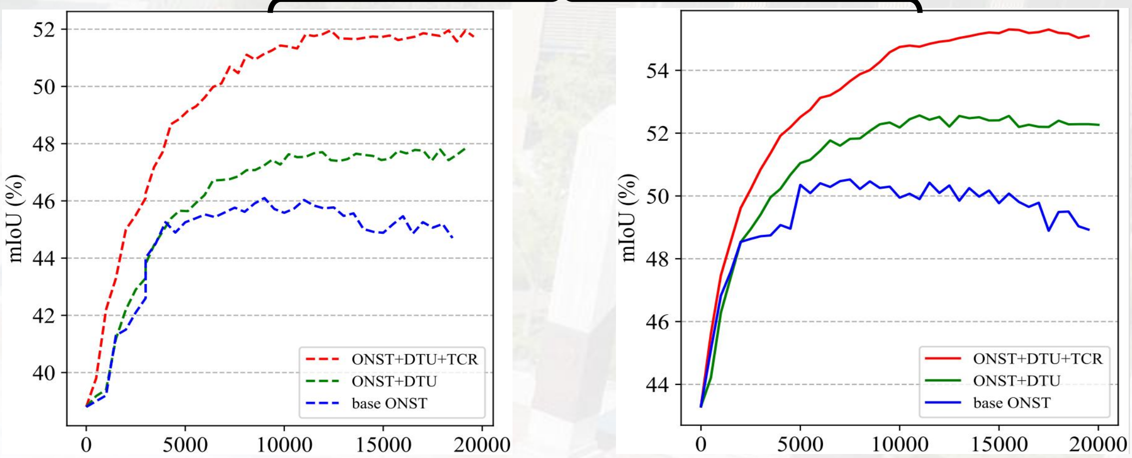
$C\downarrow$ $I\rightarrow$	2000	3000	4000	5000
30	54.1	54.2	54.3	54.1
40	54.3	54.5	54.6	54.8
50	54.8	55.2	55.5	55.1
60	54.8	55.1	55.3	55.2

Table 5. The mIoU (%) score on GTA5 → Cityscapes (val) with varying C% and I using domain generalization model.

M	30	50	70	90	110
GTA5 → Cityscapes	55.0	55.4	55.3	55.1	54.9
SYNTHIA → Cityscapes	50.4	50.7	50.7	50.5	50.1

Table 6. The mIoU scores (%) with varying M on different tasks.

Different pre-trained models



Original source-trained model

Domain generalization model

Figure 5. The mIoU score curve (val) of adding DTU and TCR.

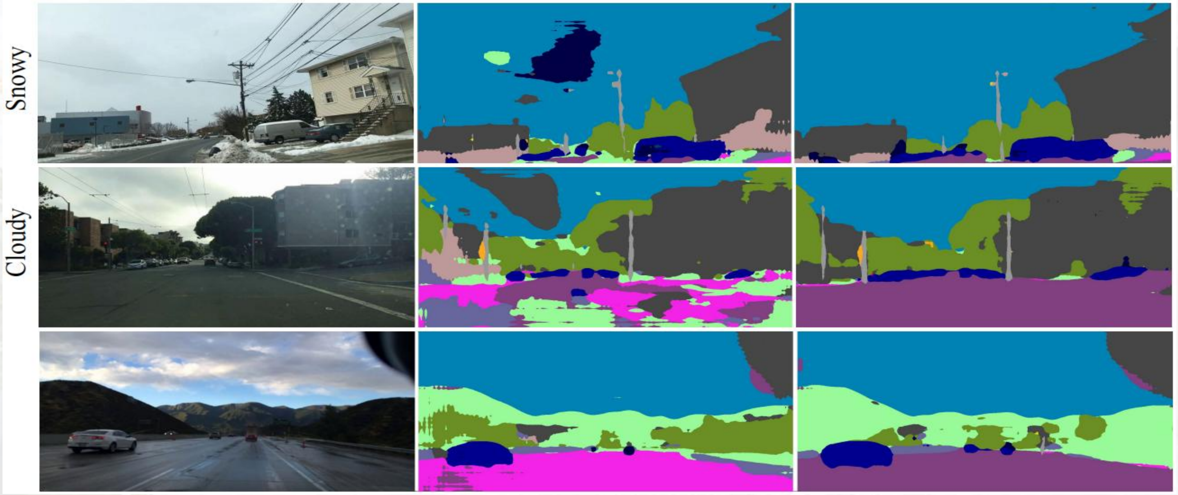


Figure 6. Visualization of the adaptation results from single-domain to mixed-domain, including rainy, snowy and cloudy scenes (i.e. GTA5 → BDD-100K).

厚德 求真 砺学 笃行

THANKS

Thanks!

QUESTIONS & ANSWERS

西安電子科技大學