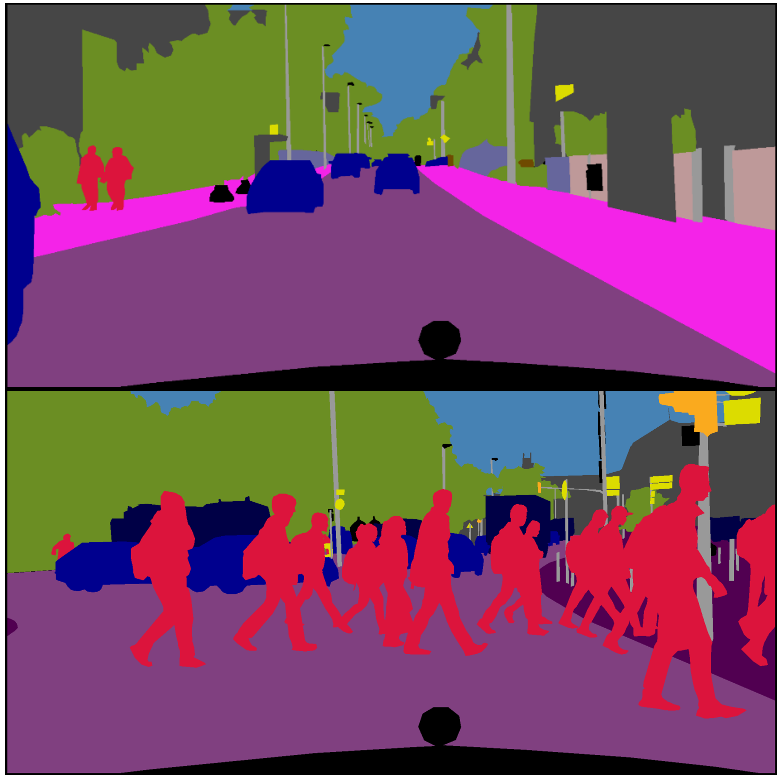
CVPR 2022

Domain Adaptation

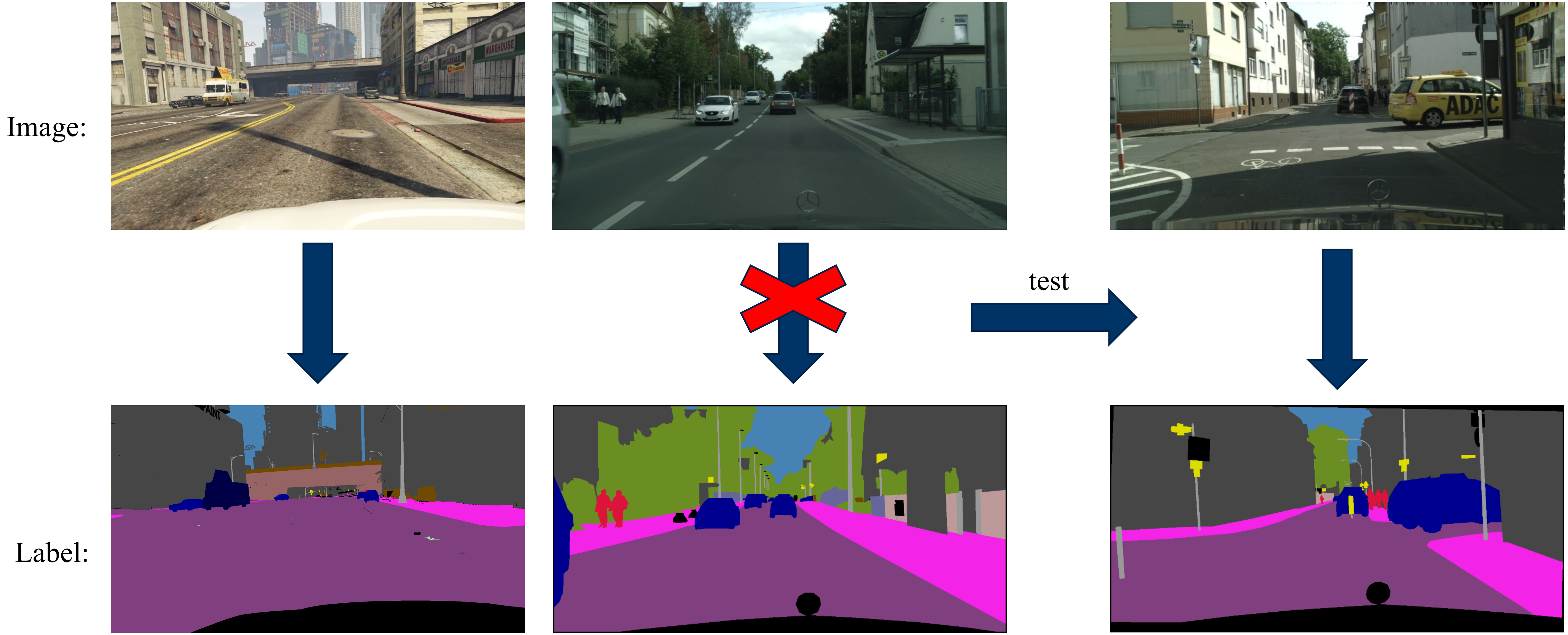
01 Background

Annotations are particularly costly as every pixel has to be labeled



01 What is UDA?

UDA: Unsupervised Domain Adaptation



Domain Adaptation

Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

Yahao Liu Jinhong Deng Jiale Tao Tong Chu Lixin Duan Wen Li*
School of Computer Science and Engineering & Shenzhen Institute for Advanced Study,
University of Electronic Science and Technology of China
{lyhaolive, jhdeng1997, jialetao.std, uestcchutong, lxduan, liwenbnu }@gmail.com

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **CVPR 2022**
- **Motivation**

Ignore the label shift problem, which commonly exists in CDSS tasks, since the label distribution is often different across domains.

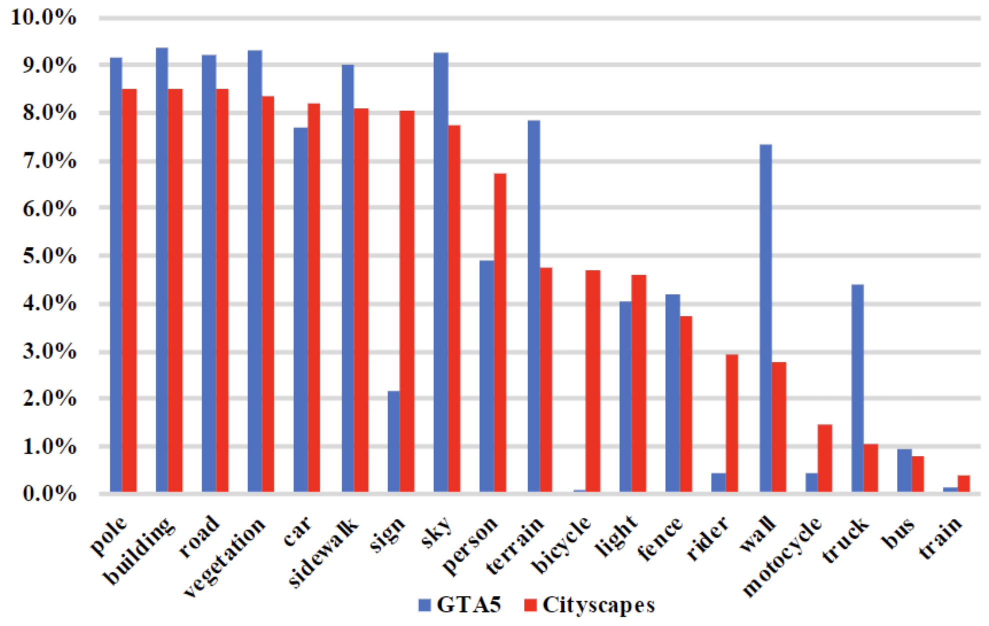


Figure 1. Label distribution in GTA5 and Cityscapes. There is an obvious label shift problem between the two datasets. For example, the frequency of "rider" in GTA5 is much less than that in Cityscapes while that of "wall" is opposite (Best viewed in color).

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

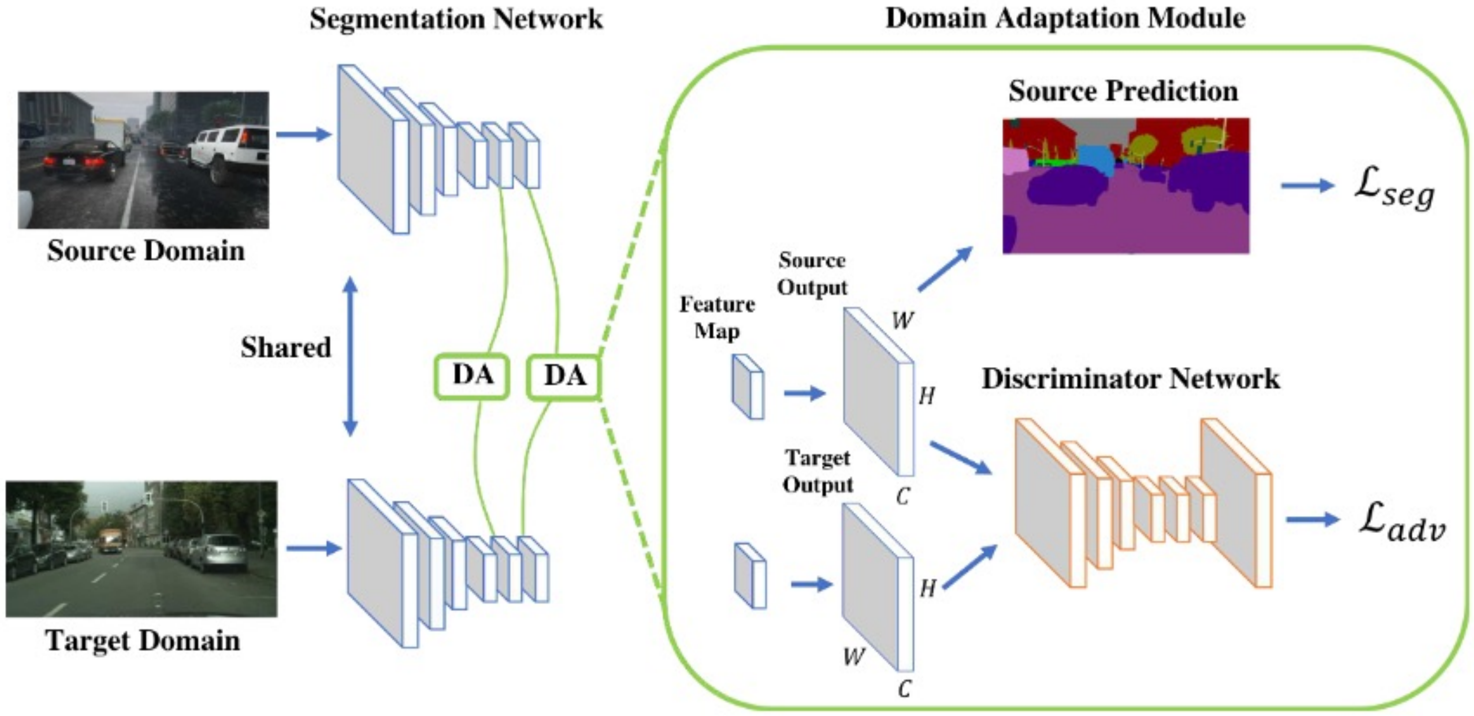
• Contribution

- propose to address the label shift issue for CDSS tasks in a more realistic scenario (i.e., the conditional distributions are different across domains) and reveal that the classifier bias is the critical factor leading to poor generalization on the target domain
- propose two simple yet effective methods to rectify the classifier bias from source to target by remolding the classifier predictions after explicitly aligning the conditional distribution

- **Motivation**
- $G = C \circ F$
- $G(x) = C\big(F(x)\big) = p(Y|F(x))$
- **Bayes' theorem** : $C\big(F(x)\big) = P(Y|F(x)) = \dfrac{P\big(F(x)\big|Y\big)P(Y)}{P(F(x))}$
- $C_s(F(x)) \propto P_s(F_S(x)|Y)P_s(Y)$
- $C_t(F(x)) \propto P_t(F_t(x)|Y)P_t(Y)$
- $C_t(F(x)) \propto \dfrac{C_s(F(x))P_t(Y)}{P_s(Y)}$

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Conditional Distribution Alignment**



$$\min_{F,C} \mathcal{L}_{seg} + \lambda_{adv}\mathcal{L}_{adv},$$

$$\min_{D} \mathcal{L}_{D},$$

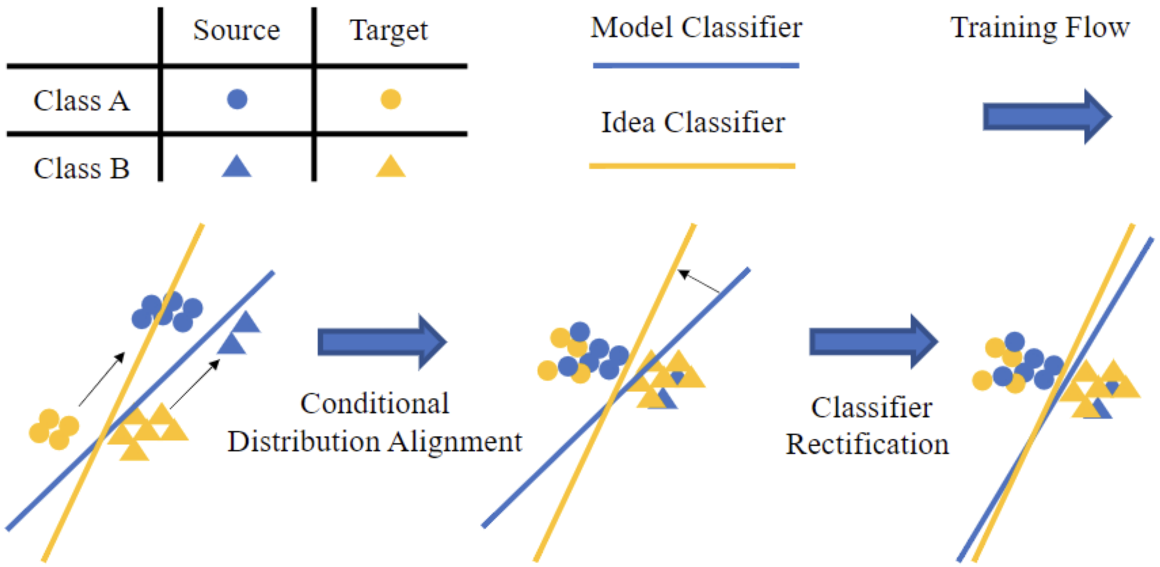
$$\mathcal{L}_{seg} = -\sum_{i=1}^{N_s}\sum_{k=1}^{K} y_s^{ik} \log\left(C\left(F\left(x_s^i\right)\right)\right),$$

$$\mathcal{L}_{adv} = -\sum_{j=1}^{N_t}\sum_{k=1}^{K} a_t^{jk} \log D\left(d=0, y=k \mid F(x_t^j)\right)$$

$$\mathcal{L}_{D} = -\sum_{i=1}^{N_s}\sum_{k=1}^{K} a_s^{ik} \log D(d=0, y=k \mid F(x_s^i)) - \sum_{j=1}^{N_t}\sum_{k=1}^{K} a_t^{jk} \log D(d=1, y=k \mid F(x_t^j)),$$

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Classifier Rectification(CR)**



$$C_s\left(F(x)\right) \propto \frac{C_t\left(F\left(x\right)\right) P_s\left(Y\right)}{P_t\left(Y\right)}.$$

$$\min_{C_{cda}} \mathcal{L}'_{\text{seg}},$$

$$\mathcal{L}'_{\text{seg}} = -\sum_{i=1}^{n_s}\sum_{k=1}^{K} y_s^{ik} \log(\hat{p}_s^{ik}),$$

$$\hat{p}_s^{ik} = \frac{p_s^{ik} \cdot \frac{P_s(Y=k)}{P_t(Y=k)}}{\sum_{k'=1}^{K}\left(p_s^{ik'} \cdot \frac{P_s(Y=k')}{P_t(Y=k')}\right)}.$$

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Inference Adjustment (IA)**

$$y^i_{IA} = \text{argmax}_k \left(p^i_t \cdot \frac{P_t(Y)}{P_s(Y)} \right).$$

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Label Distribution Estimation**
- **we denote the count of the image-level label of i-th source image as $I_s^i(k)$**

$$I_s^i(k) = \mathbb{1}\left[\left(\sum_{h=1}^{H}\sum_{w=1}^{W}\mathbb{1}\left[y_s^i(h,w) == k\right]\right) > n_s\right]$$

$$P_s(\hat{Y} = k) = \frac{\sum_{i=1}^{N_s} I_s^i(k)}{\sum_{k=1}^{K}\sum_{i=1}^{N_s} I_s^i(k)}.$$

$$p_i(k) = \log\left[\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\exp\left(G_{cda}\left(X_t^i\right)(h,w,k)\right)\right], \tag{17}$$

$$I_t^i(k) = \mathbb{1}\left[p_i(k) > p_{pix}(k)\right], \tag{18}$$

$$P_t(\hat{Y} = k) = \frac{\sum_{i=1}^{N_t} I_t^i(k)}{\sum_{k=1}^{K}\sum_{i=1}^{N_t} I_t^i(k)}. \tag{19}$$

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Experimental Results**
- **GTA5:**

Method	road	sidewalk	building	wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
Source	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
AdaptSegNet [42]	86.5	36.0	79.9	23.4	23.3	23.9	35.2	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.4
AdvEnt [43]	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
CLAN [29]	87.0	27.1	79.6	27.3	23.3	28.3	35.5	24.2	83.6	27.4	74.2	58.6	28.0	76.2	33.1	36.7	6.7	31.9	31.4	43.2
FADA [44]	87.0	37.6	83.3	36.9	25.3	30.9	35.3	21.0	82.7	36.8	83.1	58.3	34.1	83.3	31.5	35.0	24.4	34.3	32.0	46.9
Our IA	87.9	37.0	83.3	37.0	25.0	31.0	35.7	24.9	83.4	38.9	85.7	58.0	35.4	83.6	35.3	36.3	30.7	32.5	45.2	48.8
Our CR	89.1	34.3	83.6	38.3	27.5	28.9	34.7	17.6	84.2	41.0	85.1	57.8	33.7	85.1	38.5	41.3	30.7	31.1	48.0	49.0
FDA [47]	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
DACS [41]	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
CRST [55]	91.0	55.4	80.0	33.7	21.4	37.3	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	26.9	26.0	42.3	47.1
FADA+SD [44]	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
IAST [32]	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
CLS [25]+IAST	**94.7**	60.1	85.6	39.5	24.4	44.1	39.5	20.6	88.7	38.7	80.3	67.2	35.1	86.5	37.0	45.4	**39.0**	37.9	46.2	53.0
Ours+SD	91.2	45.1	85.5	41.0	30.8	36.0	41.1	19.3	87.4	45.7	88.7	64.4	37.8	87.5	41.8	51.2	11.2	41.6	54.9	52.7
Ours+IAST	94.1	**61.3**	86.5	39.3	33.5	38.3	48.9	38.5	87.2	44.2	89.3	63.4	38.3	86.2	30.5	43.0	33.6	43.1	54.8	55.5
R-MRNet [52]	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
ProDA [49]	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
Ours+ProDA	92.9	52.7	87.2	39.4	41.3	43.9	**55.0**	52.9	**89.3**	**48.2**	**91.2**	**71.4**	36.0	**90.2**	**67.9**	**59.8**	0.0	48.5	**59.3**	**59.3**

- **Experimental Results**
- **SYNTHIA :**

Method	road	sidewalk	building	wall*	fence*	pole*	light	sign	vege.	sky	person	rider	car	bus	motor	bike	mIoU*	mIoU
Source	55.6	23.8	74.6	9.2	0.2	24.4	6.1	12.1	74.8	79.0	55.3	19.1	39.6	23.3	13.7	25.0	38.6	33.5
AdaptSegNet [42]	81.7	39.1	78.4	11.1	0.3	25.8	6.8	9.0	79.1	80.8	54.8	21.0	66.8	34.7	13.8	29.9	45.8	39.6
AdvEnt [43]	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	84.1	57.9	23.8	73.3	36.4	14.2	33.0	48.0	41.2
CLAN [29]	81.3	37.0	80.1	-	-	-	16.1	13.7	78.2	81.5	53.4	21.2	73.0	32.9	22.6	30.7	47.8	-
FADA [44]	81.3	35.1	80.8	9.6	0.2	26.8	9.1	17.8	82.4	81.5	49.9	18.8	78.9	33.3	15.3	33.7	47.5	40.9
Our IA	82.2	35.6	80.8	9.0	0.2	27.1	12.4	21.3	82.3	80.7	54.4	21.2	80.0	36.6	14.0	42.2	49.5	42.5
Our CR	83.6	36.2	80.9	10.3	0.1	27.4	17.6	22.8	81.5	81.2	54.6	20.1	80.3	38.1	11.1	42.9	50.1	43.0
FDA [47]	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	-
DACS [41]	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	54.8	48.3
CRST [55]	67.7	32.2	73.9	10.7	1.6	37.4	22.2	31.2	80.8	80.5	60.8	29.1	82.8	25.0	19.4	45.3	50.1	43.8
FADA+SD [44]	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	52.5	45.2
IAST [32]	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	57.0	49.8
Ours+SD	86.9	42.9	83.3	9.9	0.0	35.3	17.2	26.0	85.4	83.0	62.0	18.5	86.7	51.4	12.8	50.0	54.3	47.0
Ours+IAST	84.6	43.0	84.1	**38.1**	0.5	36.7	32.9	36.2	83.1	81.9	65.6	33.4	80.5	34.5	38.2	53.1	57.8	51.6
R-MRNet [52]	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	54.9	47.9
ProDA [49]	**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	62.0	55.5
Ours+ProDA	82.5	37.2	81.1	23.8	0.0	**45.7**	**57.2**	**47.6**	87.7	85.8	74.1	28.6	**88.4**	**66.0**	**47.0**	**55.3**	**64.5**	**56.7**

01 Undoing the Damage of Label Shift for Cross-domain Semantic Segmentation

- **Experimental Results**

methods	mIoU	Δ
Baseline+SD [44]	49.2	
Ours+SD	52.7	3.5↑
IAST [32]	51.5	
Baseline+IAST	53.2	
Ours+IAST	55.5	2.3↑
ProDA [49] stage1	53.7	
Baseline+ProDA stage1	55.1	
Ours+ProDA stage1	57.6	2.5↑
ProDA	57.5	
Ours+ProDA	59.3	1.8↑

methods	classifier	mIoU	Δ
FADA [44]	Original ASPP	46.9	
Our IA	Original ASPP	48.8	1.9↑
Our CR	Original ASPP	49.0	2.1↑
FADA [44]	Modified ASPP	47.6	
Our IA	Modified ASPP	49.2	1.6↑
Our CR	Modified ASPP	48.7	1.1↑

Domain Adaptation

Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

Ruihuang Li[1], Shuai Li[1], Chenhang He[1], Yabin Zhang[1], Xu Jia[2], Lei Zhang[1*]
[1]The Hong Kong Polytechnic University, [2]Dalian University of Technology
{csrhli, csshuaili, csche, csybzhang, cslzhang}@comp.polyu.edu.hk, xjia@dlut.edu.cn

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **CVPR 2022**
- **Motivation**

pixel-wise cluster assignments could reveal the intrinsic distributions of pixels in target domain, and provide useful supervision for model training

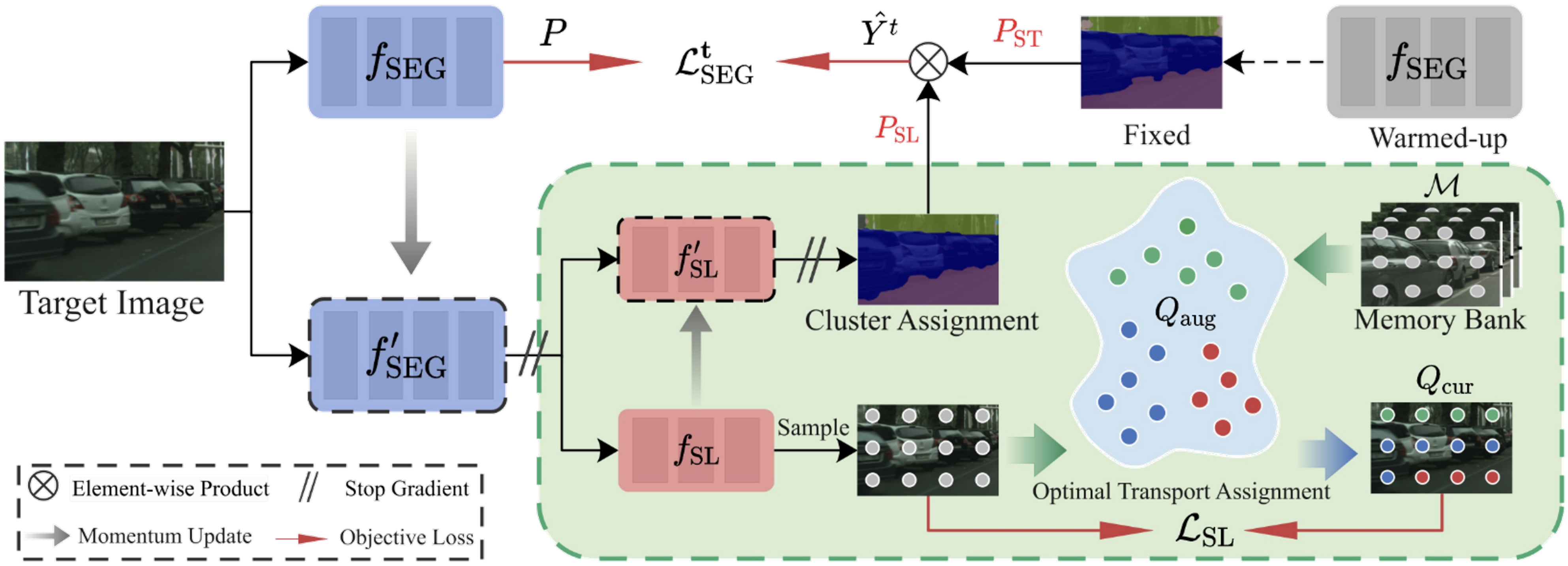
01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Contribution**

  - A pixel-level self-labeling module is developed for domain adaptive semantic segmentation. We cluster pixels in an online fashion and simultaneously rectify pseudo labels based on the resulting cluster assignments.
  - propose two simple yet effective methods to rectify the classifier bias from source to target by remolding the classifier predictions after explicitly aligning the conditional distribution

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Overview of the Proposed Model**

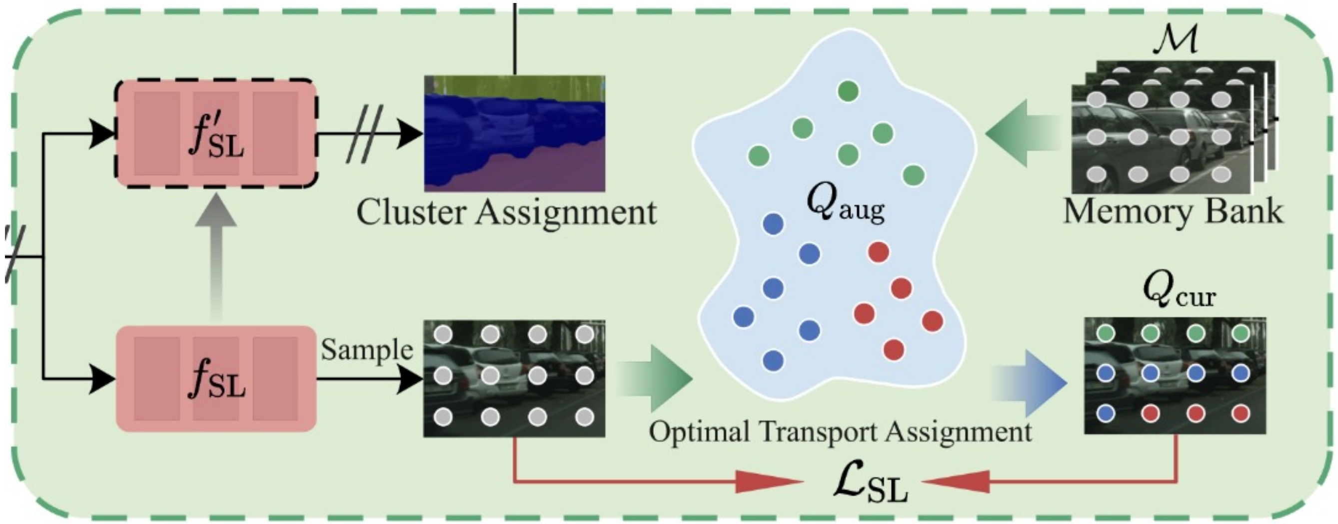


Class-balanced Pixel-level Self-Labeling Module

$$\hat{Y}^{t,(c)}_{n,i} = \begin{cases} 1, & \text{if } c = \underset{c*}{argmax}(P^{(c*)}_{\text{SL},n,i} \cdot P^{(c*)}_{\text{ST},n,i}) \\ 0, & \text{otherwise} \end{cases}, \quad \mathcal{L}^t_{\text{SEG}} = -\sum_{n=1}^{N_T}\sum_{i=1}^{H\times W}\sum_{c=1}^{C} \hat{Y}^{t,(c)}_{n,i} \log P^{(c)}_{n,i}. \quad \mathcal{L}^s_{\text{SEG}} = -\sum_{n=1}^{N_S}\sum_{i=1}^{H\times W}\sum_{c=1}^{C} Y^{s,(c)}_{n,i} \log P^{(c)}_{n,i}.$$

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Online Pixel-Level Self-Labeling**



Class-balanced Pixel-level Self-Labeling Module

$$p_m^{(c)} = \frac{\exp(\frac{1}{\tau} f_{\mathrm{SL}}^{(c)}(z_m))}{\sum_{c'} \exp(\frac{1}{\tau} f_{\mathrm{SL}}^{(c')}(z_m))}, \ c \in \{1, \cdots, C\},$$

$$\mathcal{L}_{\mathrm{SL}} = -\frac{1}{M} \sum_{m=1}^{M} \sum_{c=1}^{C} q_m^{(c)} \log p_m^{(c)} \quad s.t.\ Q \in \mathbb{Q},$$

$$\text{with} \quad \mathbb{Q} := \{Q \in \mathbb{R}_{+}^{C \times M} | Q\mathbf{1}_M = r, Q^T \mathbf{1}_C = h\}.$$

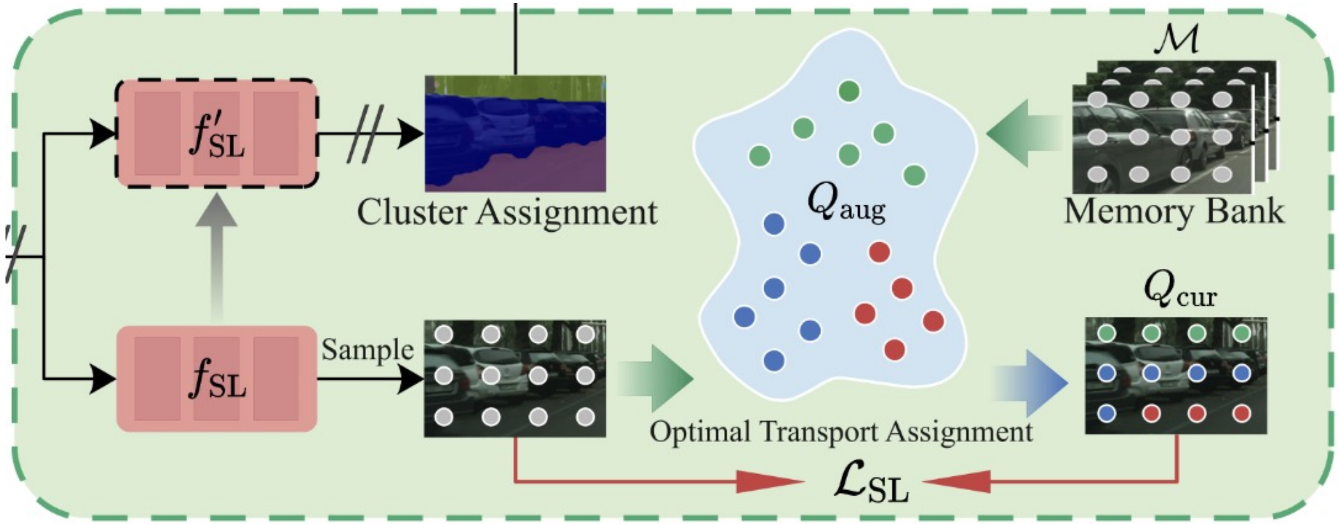
$$Q^* = \mathrm{diag}(\alpha) \exp(\frac{f_{\mathrm{SL}}(\hat{Z})}{\varepsilon}) \,\mathrm{diag}(\beta),$$

Weight Initialization

$$\bar{\mathbf{z}}_c = \frac{1}{|\Gamma_c|} \sum_{n=1}^{N_T} \sum_{i=1}^{H \times W} Y_{\mathrm{ST},n,i}^{(c)} \cdot z_{n,i},$$

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Class-Balanced Sampling.**



Class-balanced Pixel-level Self-Labeling Module

$$\delta_n^{(c)} = \frac{1}{H \times W} \sum_i^{H \times W} \hat{Y}_{n,i}^{t,(c)},$$

$$M_c = \lfloor M \times \delta_n^{(c)} \rfloor$$

Distribution Alignment: simultaneously optimizing Q and $\hat{P}$ in Eq. 5 may lead to degenerated results that all data points are trivially assigned to a single cluster.

$$\delta_{pseudo}^{(c)}|_0 = \frac{1}{N_T \times H \times W} \sum_n^{N_T} \sum_i^{H \times W} Y_{\mathrm{ST},n,i}^{t,(c)}.$$

$$\delta_{pseudo}^{(c)}|_k = \alpha \delta_{pseudo}^{(c)}|_{k-1} + (1-\alpha)\delta_n^{(c)}.$$

$$r = \delta_{pseudo}, \quad h = \frac{1}{M}\mathbf{1}_M.$$

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Experimental Results**
- **GTA5:**

Method	road	sideway	building	wall	fence	pole	light	sign	vege	terrace	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
AdaptSeg [39]	86.5	25.9	79.8	22.1	20.0	23.6	33.1	21.8	81.8	25.9	75.9	57.3	26.2	76.3	29.8	32.1	7.2	29.5	32.5	41.4
CyCADA [17]	86.7	35.6	80.1	19.8	17.5	38.0	39.9	41.5	82.7	27.9	73.6	64.9	19.0	65.0	12.0	28.6	4.5	31.1	42.0	42.7
ADVENT [41]	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
CBST [56]	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
FADA [42]	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_UDA [51]	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 [48]	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
PIT [30]	87.5	43.4	78.8	31.2	30.2	36.3	39.3	42.0	79.2	37.1	79.3	65.4	37.5	83.2	**46.0**	45.6	25.7	23.5	49.9	50.6
IAST [31]	**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
ProDA [50]	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
CPSL (ours)	91.7	52.9	83.6	**43.0**	32.3	**43.7**	**51.3**	42.8	85.4	37.6	81.1	**69.5**	30.0	**88.1**	44.1	**59.9**	24.9	**47.2**	48.4	**55.7**
ProDA+*distill*	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
CPSL+*distill*	**92.3**	**59.9**	**84.9**	45.7	29.7	**52.8**	**61.5**	**59.5**	87.9	41.5	**85.0**	**73.0**	35.5	**90.4**	**48.7**	**73.9**	**26.3**	**53.8**	53.9	**60.8**

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Experimental Results**
- **SYNTHIA:**

Method	road	sideway	building	wall	fence	pole	light	sign	vege	sky	person	rider	car	bus	motor	bike	$mIoU^{13}$	$mIoU^{16}$
AdaptSeg [39]	79.2	37.2	78.8	-	-	-	9.9	10.5	78.2	80.5	53.5	19.6	67.0	29.5	21.6	31.3	45.9	-
ADVENT [41]	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	84.1	57.9	23.8	73.3	36.4	14.2	33.0	48.0	41.2
CBST [56]	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	48.9	42.6
CAG_UDA [51]	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	51.5	44.5
PIT [30]	83.1	27.6	81.5	8.9	0.3	21.8	26.4	33.8	76.4	78.8	64.2	27.6	79.6	31.2	31.0	31.3	51.8	44.0
FADA [42]	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	52.5	45.2
FDA [48]	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	-
PyCDA [26]	75.5	30.9	83.3	20.8	0.7	32.7	27.3	33.5	84.7	85.0	64.1	25.4	85.0	45.2	21.2	32.0	53.3	46.7
IAST [31]	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	57.0	49.8
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	59.3	52.6
ProDA [50]	87.1	44.0	83.2	**26.9**	0.7	42.0	45.8	**34.2**	86.7	81.3	68.4	22.1	87.7	50.0	31.4	38.6	58.5	51.9
CPSL (ours)	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**	**61.7**	**54.4**
ProDA+*distill*	**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	62.0	55.5
CPSL+*distill*	87.2	43.9	**85.5**	33.6	0.3	**47.7**	**57.4**	**37.2**	87.8	**88.5**	**79.0**	**32.0**	**90.6**	49.4	**50.8**	**59.8**	**65.3**	**57.9**

01 Class-Balanced Pixel-Level Self-Labeling for Domain Adaptive Semantic Segmentation

- **Experimental Results**

Configuration	mIoU	Δ
w/o SL	47.8	-7.9
w/o CB	51.8	-3.9
w/o ST	39.4	-16.3
w/o Init	49.9	-5.8
w/o Aug	54.2	-1.5
w/o Mom	54.6	-1.1
CPSL	**55.7**	-

# samples	mIoU
64	54.9
128	55.3
256	55.5
512	**55.7**
1024	54.3
2048	53.4

Domain Adaptation

Towards Fewer Annotations: Active Learning via Region Impurity and Prediction Uncertainty for Domain Adaptive Semantic Segmentation

Binhui Xie[1] Longhui Yuan[1] Shuang Li[1] ✉ Chi Harold Liu[1] Xinjing Cheng[2,3]
[1]School of Computer Science and Technology, Beijing Institute of Technology
[2]School of Software, BNRist, Tsinghua University [3]Inceptio Technology
{binhuixie,longhuiyuan,shuangli,chiliu}@bit.edu.cn cnorbot@gmail.com

01 Towards Fewer Annotations: Active Learning via Region Impurity and Prediction Uncertainty for Domain Adaptive Semantic Segmentation

- **CVPR 2022**
- **Motivation**

(a) Target image　(b) Image-based selection (100%)

(c) Point-based selection (2.2%)　(d) Region-based selection (2.2%)

Figure 1. Illustration of different selection strategies. **Image-based selection** (e.g., MADA [41]) picks a few target samples and label the entire image, which is probably inefficient. **Point-based selection** (e.g., LabOR [54]) chooses scarce points about which the model is uncertain, while uncertainty estimation at point level is prone to lump pixels that come from particular categories. **Our region-based selection** asks for more annotations of regions with more categories as well as object boundaries in an effective way.

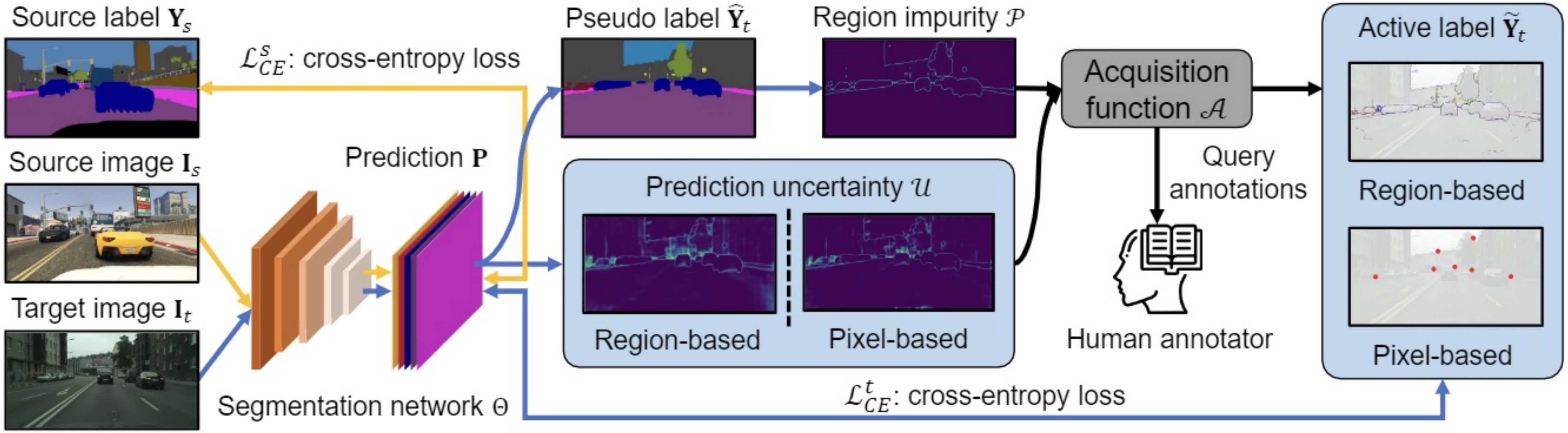
01 Towards Fewer Annotations: Active Learning via Region Impurity and Prediction Uncertainty for Domain Adaptive Semantic Segmentation

• Contribution

- benchmark the performance of prior methods for active domain adaptation regarding semantic segmentation and uncover that methods using image-based or point-based selection strategies are not effective
- propose a region-based acquisition strategy for domain adaptive semantic segmentation, termed RIPU, that utilizes region impurity and prediction uncertainty to identify image regions that are both diverse in spatial adjacency and uncertain in prediction output

01 Active Learning via Region Impurity and Prediction Uncertainty

- **Overview of the Proposed Model**

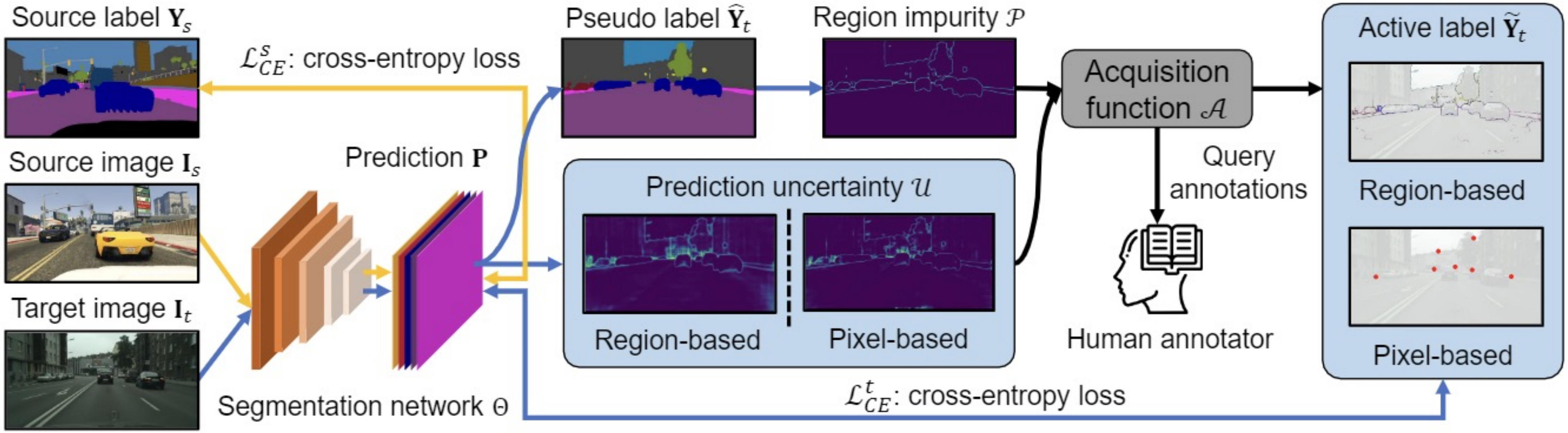


- **Region Generation: k-square-neighbors** $\mathcal{N}_k(i,j) = \{(u,v) \mid |u-i| \le k, |v-j| \le k\}$.
- **Region Impurity:** $\mathcal{N}_k^c(i,j) = \{(u,v) \in \mathcal{N}_k(i,j) \mid \widehat{\mathbf{Y}}_t^{(u,v)} = c\}$. $\mathcal{P}^{(i,j)} = -\sum_{c=1}^{C} \frac{|\mathcal{N}_k^c(i,j)|}{|\mathcal{N}_k(i,j)|} \log \frac{|\mathcal{N}_k^c(i,j)|}{|\mathcal{N}_k(i,j)|}$,
- **Prediction Uncertainty:** $\mathcal{U}^{(i,j)} = \frac{1}{|\mathcal{N}_k(i,j)|} \sum_{(u,v)\in\mathcal{N}_k(i,j)} \mathcal{H}^{(u,v)}$

$$\mathcal{A}(\mathbf{I}_t;\Theta^n) = \mathcal{U} \odot \mathcal{P} ,$$

01 Active Learning via Region Impurity and Prediction Uncertainty

- **Overview of the Proposed Model**



$$\mathcal{L}_{sup} = \mathcal{L}^s_{CE}(\mathbf{I}_s, \mathbf{Y}_s) + \mathcal{L}^t_{CE}(\mathbf{I}_t, \tilde{\mathbf{Y}}_t),$$

$$\mathcal{L}^s_{cr} = \frac{1}{|\mathbf{I}_s|} \sum_{(i,j)\in\mathbf{I}_s} \left\|\mathbf{P}^{(i,j)} - \overline{\mathbf{P}}^{(i,j)}\right\|_1$$

$$\pi(\mathbf{P}_t^{(i,j,c)}) = \begin{cases} 1 & \text{if } \mathbf{P}_t^{(i,j,c)} < \tau, \\ 0 & \text{otherwise}, \end{cases} \qquad \mathcal{L}^t_{nl} = \frac{-1}{\Lambda(\mathbf{I}_t)} \sum_{(i,j)\in\mathbf{I}_t} \sum_{c=1}^{C} \pi(\mathbf{P}_t^{(i,j,c)}) \log(1 - \mathbf{P}_t^{(i,j,c)}),$$

01 Active Learning via Region Impurity and Prediction Uncertainty

- **Experimental Results：GTAV----Cityscapes**

Method	Budget	road	side.	buil.	wall	fence	pole	light	sign	veg.	terr.	sky	pers.	rider	car	truck	bus	train	motor	bike	mIoU
Source Only	-	75.8	16.8	77.2	12.5	21.0	25.5	30.1	20.1	81.3	24.6	70.3	53.8	26.4	49.9	17.2	25.9	6.5	25.3	36.0	36.6
CBST [83]	-	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
MRKLD [84]	-	91.0	55.4	80.0	33.7	21.4	37.3	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	26.9	26.0	42.3	47.1
Seg-Uncertainty [82]	-	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
TPLD [54]	-	94.2	60.5	82.8	36.6	16.6	39.3	29.0	25.5	85.6	44.9	84.4	60.6	27.4	84.1	37.0	47.0	31.2	36.1	50.3	51.2
DPL-Dual [7]	-	92.8	54.4	86.2	41.6	32.7	36.4	49.0	34.0	85.8	41.3	86.0	63.2	34.2	87.2	39.3	44.5	18.7	42.6	43.1	53.3
ProDA [80]	-	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
LabOR [53]	40 pixels	**96.1**	**71.8**	**88.8**	47.0	46.5	**42.2**	**53.1**	**60.6**	**89.4**	55.1	**91.4**	**70.8**	44.7	90.6	56.7	47.9	39.1	47.3	62.7	63.5
Ours (PA)	40 pixels	95.5	69.2	88.2	**48.0**	**46.5**	36.9	45.2	55.7	88.5	**55.3**	90.2	69.2	**46.1**	**91.2**	**70.7**	**73.0**	**58.2**	**50.1**	**65.9**	**65.5**
LabOR [53]	2.2%	**96.6**	**77.0**	89.6	47.8	50.7	**48.0**	**56.6**	**63.5**	89.5	**57.8**	91.6	72.0	47.3	91.7	62.1	61.9	48.9	47.9	65.3	66.6
Ours (RA)	2.2%	96.5	74.1	**89.7**	**53.1**	**51.0**	43.8	53.4	62.2	**90.0**	57.6	**92.6**	**73.0**	**53.0**	**92.8**	**73.8**	**78.5**	**62.0**	**55.6**	**70.0**	**69.6**
Fully Supervised	100%	96.8	77.5	90.0	53.5	51.5	47.6	55.6	62.9	90.2	58.2	92.3	73.7	52.3	92.4	74.3	77.1	64.5	52.4	70.1	70.2
AADA$^\sharp$ [60]	5%	92.2	59.9	87.3	36.4	45.7	46.1	50.6	59.5	88.3	44.0	90.2	69.7	38.2	90.0	55.3	45.1	32.0	32.6	62.9	59.3
MADA$^\sharp$ [40]	5%	95.1	69.8	88.5	43.3	48.7	45.7	53.3	59.2	89.1	46.7	91.5	73.9	50.1	91.2	60.6	56.9	48.4	51.6	68.7	64.9
Ours (RA)$^\sharp$	5%	**97.0**	**77.3**	**90.4**	**54.6**	**53.2**	**47.7**	**55.9**	**64.1**	**90.2**	**59.2**	**93.2**	**75.0**	**54.8**	**92.7**	**73.0**	**79.7**	**68.9**	**55.5**	**70.3**	**71.2**
Fully Supervised$^\sharp$	100%	97.4	77.9	91.1	54.9	53.7	51.9	57.9	64.7	91.1	57.8	93.2	74.7	54.8	93.6	76.4	79.3	67.8	55.6	71.3	71.9

01 Active Learning via Region Impurity and Prediction Uncertainty

• Experimental Results：SYNTHIA----Cityscapes

Method	Budget	road	side.	buil.	wall*	fence*	pole*	light	sign	veg.	sky	pers.	rider	car	bus	motor	bike	mIoU	mIoU*
Source Only	-	64.3	21.3	73.1	2.4	1.1	31.4	7.0	27.7	63.1	67.6	42.2	19.9	73.1	15.3	10.5	38.9	34.9	40.3
CBST [83]	-	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	48.9
MRKLD [84]	-	67.7	32.2	73.9	10.7	1.6	37.4	22.2	31.2	80.8	80.5	60.8	29.1	82.8	25.0	19.4	45.3	43.8	50.1
DPL-Dual [7]	-	87.5	45.7	82.8	13.3	0.6	33.2	22.0	20.1	83.1	86.0	56.6	21.9	83.1	40.3	29.8	45.7	47.0	54.2
TPLD [54]	-	80.9	44.3	82.2	19.9	0.3	40.6	20.5	30.1	77.2	80.9	60.6	25.5	84.8	41.1	24.7	43.7	47.3	53.5
Seg-Uncertainty [82]	-	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	54.9
ProDA [22]	-	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 (PA)	40 pixels	95.8	71.9	87.8	39.9	41.5	38.3	47.1	54.2	89.2	90.8	69.9	48.5	91.4	71.5	52.2	67.2	66.1	72.1
Ours (RA)	2.2%	**96.8**	**76.6**	**89.6**	**45.0**	**47.7**	**45.0**	**53.0**	**62.5**	**90.6**	**92.7**	**73.0**	**52.9**	**93.1**	**80.5**	**52.4**	**70.1**	**70.1**	**75.7**
Fully Supervised	100%	96.7	77.8	90.2	40.1	49.8	52.2	58.5	67.6	91.7	93.8	74.9	52.0	92.6	70.5	50.6	70.2	70.6	75.9
AADA$^\sharp$ [60]	5%	91.3	57.6	86.9	37.6	48.3	45.0	50.4	58.5	88.2	90.3	69.4	37.9	89.9	44.5	32.8	62.5	61.9	66.2
MADA$^\sharp$ [40]	5%	96.5	74.6	88.8	45.9	43.8	46.7	52.4	60.5	89.7	92.2	74.1	51.2	90.9	60.3	52.4	69.4	68.1	73.3
Ours (RA)$^\sharp$	5%	**97.0**	**78.9**	**89.9**	**47.2**	**50.7**	**48.5**	**55.2**	**63.9**	**91.1**	**93.0**	**74.4**	**54.1**	**92.9**	**79.9**	**55.3**	**71.0**	**71.4**	**76.7**
Fully Supervised$^\sharp$	100%	97.5	81.4	90.9	48.5	51.3	53.6	59.4	68.1	91.7	93.4	75.6	51.9	93.2	75.6	52.0	71.2	72.2	77.1

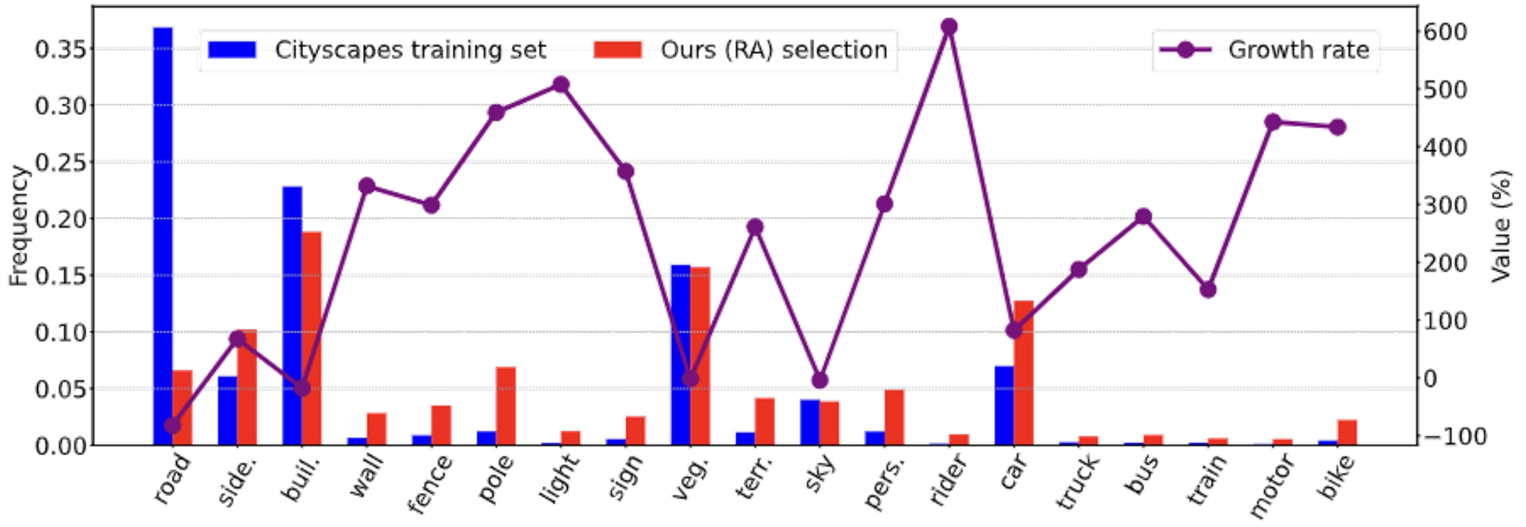
01 Active Learning via Region Impurity and Prediction Uncertainty

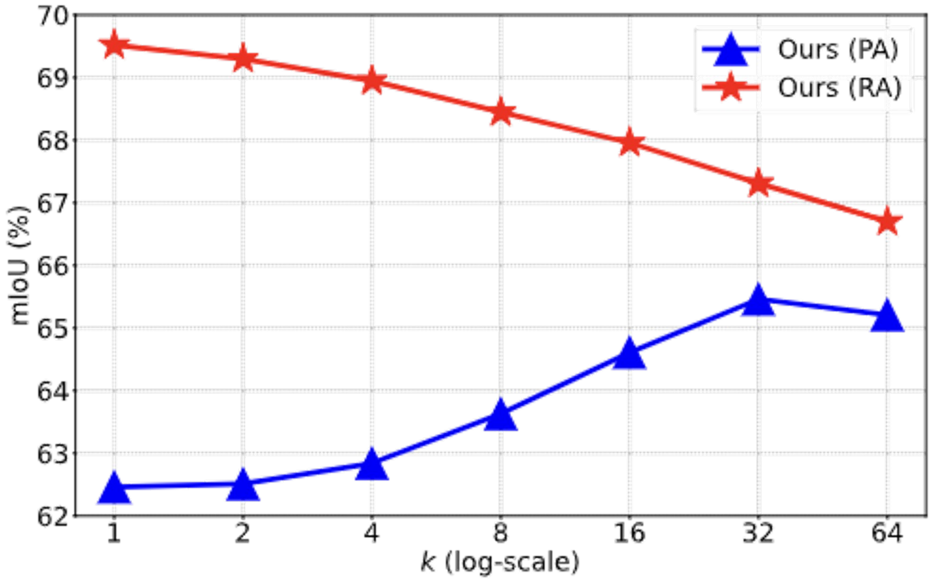
• Experimental Result

	Selection		Training		GTAV	SYNTHIA
Method	Impurity	Uncertainty	$\mathcal{L}^s_{cr}$	$\mathcal{L}^t_{nl}$	mIoU	mIoU
RAND: randomly selecting regions (2.2%)					63.8	64.7
Fully Supervised: all labeled source and target					70.2	70.6
(a)	✓				68.1	69.0
(b)		✓			66.2	67.9
(c)	✓	✓			68.5	69.2
(d)	✓	✓	✓		69.0	69.7
(e)	✓	✓		✓	69.2	69.8
(f)	✓	✓	✓	✓	**69.6**	**70.1**

Table 4. Experiments on different **active selection methods**.

Method	Budget	mIoU	Budget	mIoU
RAND	40 pixels	60.3	2.2%	63.8
ENT [52]	40 pixels	55.0	2.2%	66.2
SCONF [10]	40 pixels	59.1	2.2%	66.5
Ours	PA, 40 pixels	**64.9**	RA, 2.2%	**68.5**





Domain Adaptation

HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

Lukas Hoyer[1], Dengxin Dai[2], and Luc Van Gool[1,3]

[1] ETH Zurich, Switzerland {lhoyer,vangool}@vision.ee.ethz.ch
[2] MPI for Informatics, Germany ddai@mpi-inf.mpg.de
[3] KU Leuven, Belgium

01 DAformer

- **Previous Work：DAformer**

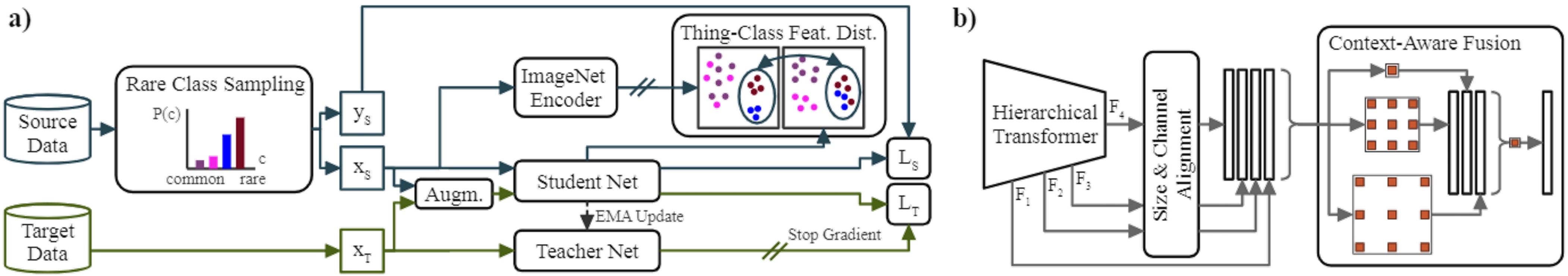


Figure 2. Overview of our UDA framework with Rare Class Sampling, Thing-Class Feature Distance, and DAFormer network.

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **arXiv**
- **Motivation**

UDA methods are usually more GPU memory intensive than regular supervised training as images from multiple domains, additional networks (e.g teacher model or domain discriminator), and additional losses are required for UDA training

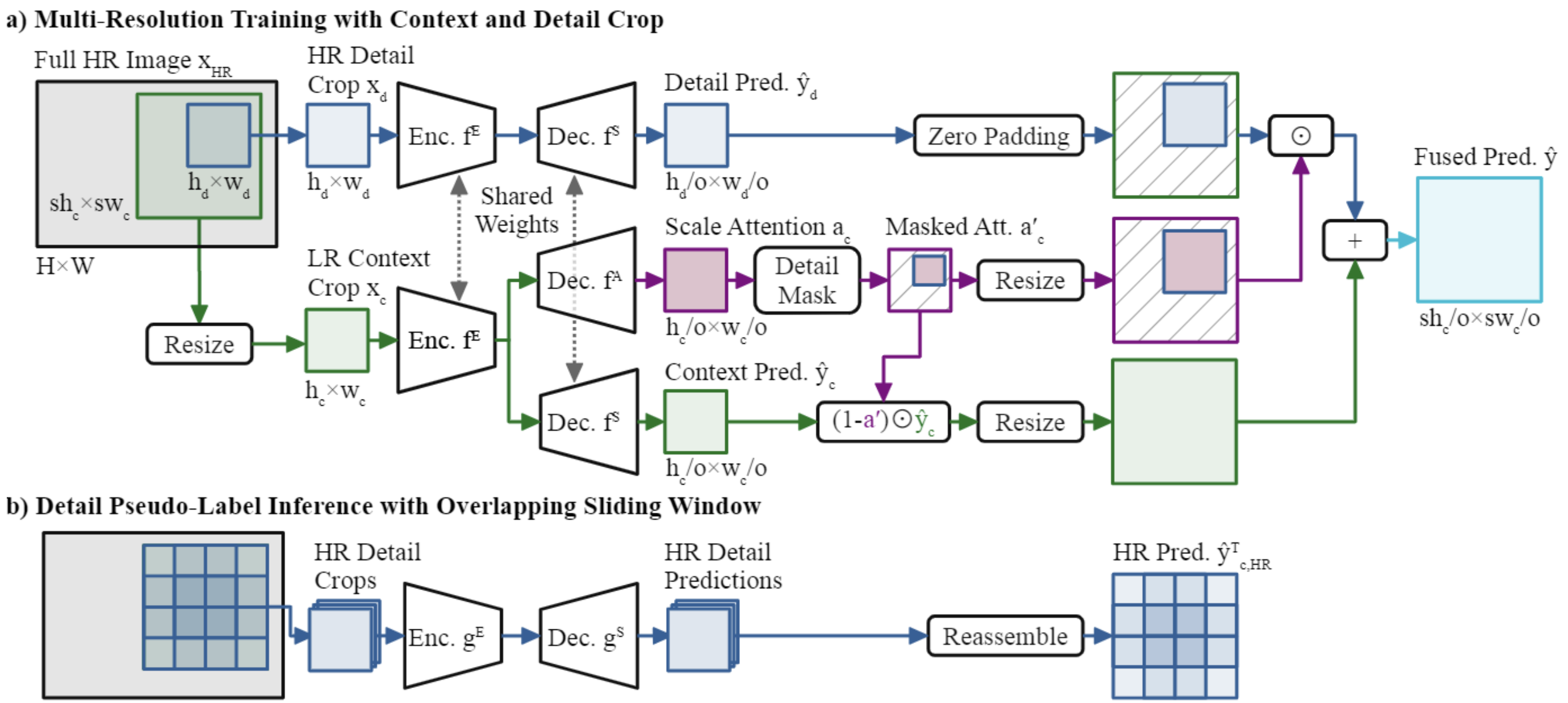
predictions from low-resolution (LR) inputs often fail to recognize small objects such as distant traffic lights and to preserve fine segmentation details such as limbs of distant pedestrians

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Contribution**
  - studying the influence of resolution and crop size
  - exploiting HR inputs for adapting small objects and fine segmentation details
  - applying multi-resolution training with a learned scale attention for object-scale-dependent adaptation
  - proposing a nested context and detail crop for memory-efficient training

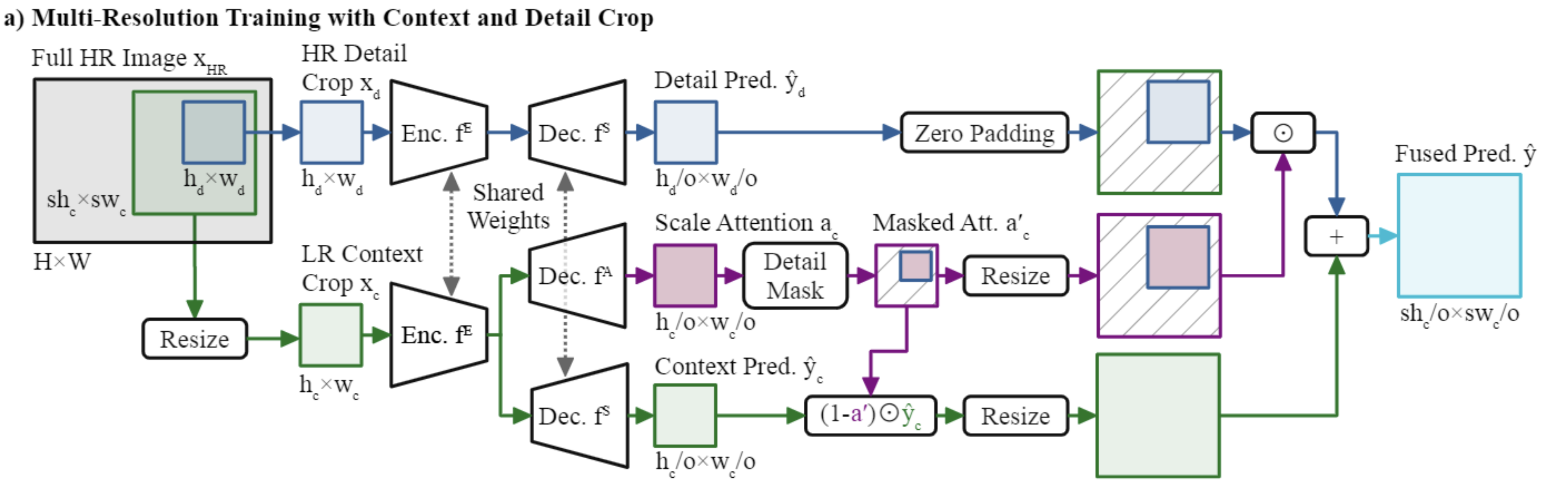
01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Overview of the Proposed Model**



01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Context and Detail Crop**



LR Context:

$$x_{c,HR} = x_{HR}[b_{c,1} : b_{c,2}, b_{c,3} : b_{c,4}] \,, \quad x_c = \zeta(x_{c,HR}, 1/s)$$

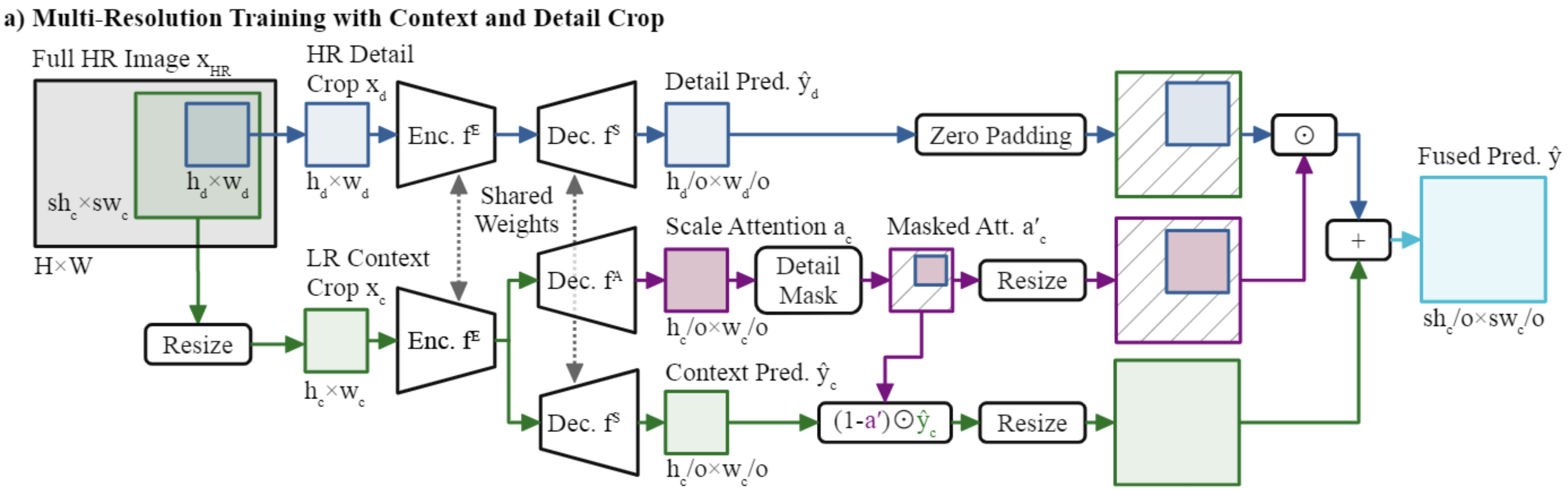
$$b_{c,1} \sim \mathcal{U}\{0, (H - sh_c)/k\} \cdot k \,, \quad b_{c,2} = b_{c,1} + sh_c \,,$$
$$b_{c,3} \sim \mathcal{U}\{0, (W - sw_c)/k\} \cdot k \,, \quad b_{c,4} = b_{c,3} + sw_c \,.$$

HR Detail:

$$x_d = x_{c,HR}[b_{d,1} : b_{d,2}, b_{d,3} : b_{d,4}] \,,$$
$$b_{d,1} \sim \mathcal{U}\{0, (sh_c - h_d)/k\} \cdot k \,, \quad b_{d,2} = b_{d,1} + h_d \,,$$
$$b_{d,3} \sim \mathcal{U}\{0, (sw_c - w_d)/k\} \cdot k \,, \quad b_{d,4} = b_{d,3} + w_d \,.$$

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Multi-Resolution Fusion**



scale attention

$$a_c = \sigma(f^S(f^A(x_c))) \quad a'_c \in \mathbb{R}^{\frac{h_c}{o} \times \frac{w_c}{o}}, \quad a'_c(i,j) = \begin{cases} a_c(i,j) & \text{if } \frac{b_{d,1}}{s \cdot o} \leq i < \frac{b_{d,2}}{s \cdot o} \wedge \frac{b_{d,3}}{s \cdot o} \leq j < \frac{b_{d,4}}{s \cdot o} \\ 0 & \text{otherwise} \end{cases}$$

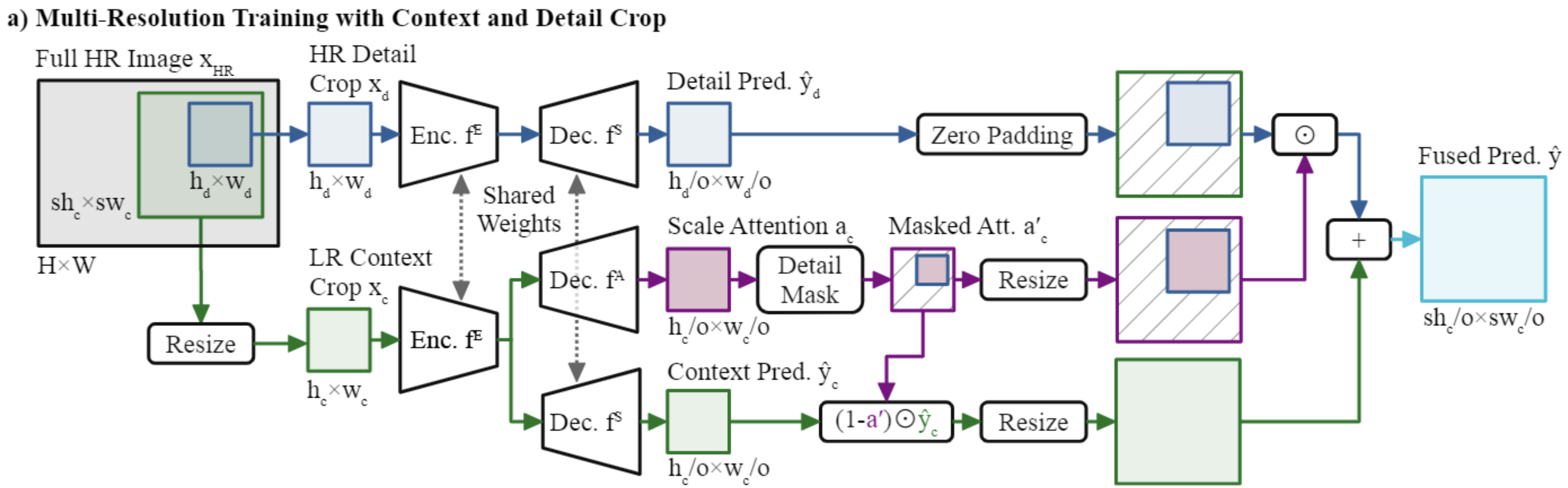
$$\hat{y}'_d(i,j) = \begin{cases} \hat{y}_d(i - \frac{b_{d,1}}{o}, j - \frac{b_{d,3}}{o}) & \text{if } \frac{b_{d,1}}{o} \leq i < \frac{b_{d,2}}{o} \wedge \frac{b_{d,3}}{o} \leq j < \frac{b_{d,4}}{o} \\ 0 & \text{otherwise} \end{cases}.$$

attention-weighted sum:

$$\hat{y}_{c,F} = \zeta((1 - a'_c) \odot \hat{y}_c, s) + \zeta(a'_c, s) \odot \hat{y}'_d\,.$$

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Multi-Resolution Fusion**



Source:

$$\mathcal{L}_{HRDA}^{S} = (1 - \lambda_d)\mathcal{L}_{ce}(\hat{y}_{c,F}^{S}, y_{c,HR}^{S}, 1) + \lambda_d \mathcal{L}_{ce}(\hat{y}_d^{S}, y_d^{S}, 1)$$

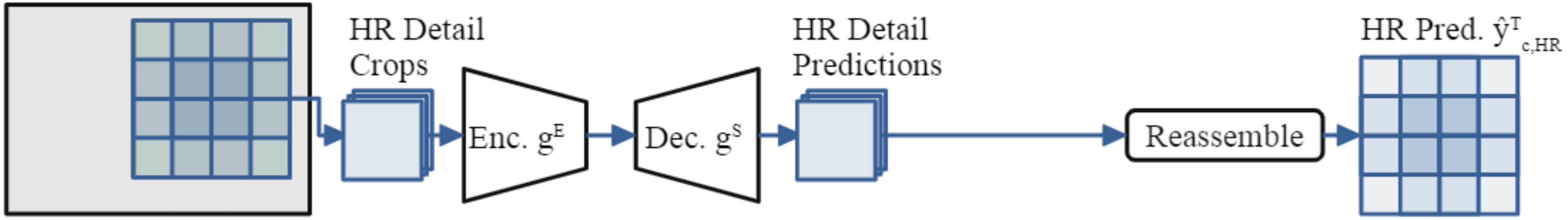
Target:

$$\mathcal{L}_{HRDA}^{T} = (1 - \lambda_d)\mathcal{L}_{ce}(\hat{y}_{c,F}^{T}, p_{c,F}^{T}, q_{c,F}^{T}) + \lambda_d \mathcal{L}_{ce}(\hat{y}_d^{T}, p_d^{T}, q_d^{T}) \, .$$

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Pseudo-Label Generation with Overlapping Sliding Window**

b) Detail Pseudo-Label Inference with Overlapping Sliding Window



The underlying HRDA prediction $\hat{y}^T_{c,F}$ is fused from the LR prediction $\hat{y}^T_c$ and HR prediction $\hat{y}^T_{c,HR}$ using the full scale attention a^T_c

$$\hat{y}^T_{c,F} = \zeta((1 - a^T_c) \odot \hat{y}^T_c, s) + \zeta(a^T_c, s) \odot \hat{y}^T_{c,HR} \, .$$

01 HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation

- **Experimental Results**
- **GTA5:**

	Road	S.walk	Build.	Wall	Fence	Pole	Tr.Light	Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike	mIoU
GTA5 → Cityscapes																				
CBST [97]	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
DACS [62]	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
CorDA [69]	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
BAPA [41]	94.4	61.0	88.0	26.8	39.9	38.3	46.1	55.3	87.8	46.1	89.4	68.8	40.0	90.2	60.4	59.0	0.0	45.1	54.2	57.4
ProDA [84]	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
DAFormer [29]	<u>95.7</u>	<u>70.2</u>	<u>89.4</u>	<u>53.5</u>	<u>48.1</u>	<u>49.6</u>	<u>55.8</u>	<u>59.4</u>	<u>89.9</u>	<u>47.9</u>	<u>92.5</u>	<u>72.2</u>	<u>44.7</u>	<u>92.3</u>	<u>74.5</u>	<u>78.2</u>	<u>65.1</u>	<u>55.9</u>	<u>61.8</u>	<u>68.3</u>
HRDA	**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**

01 ProCST: Boosting Semantic Segmentation using Progressive Cyclic Style-Transfer

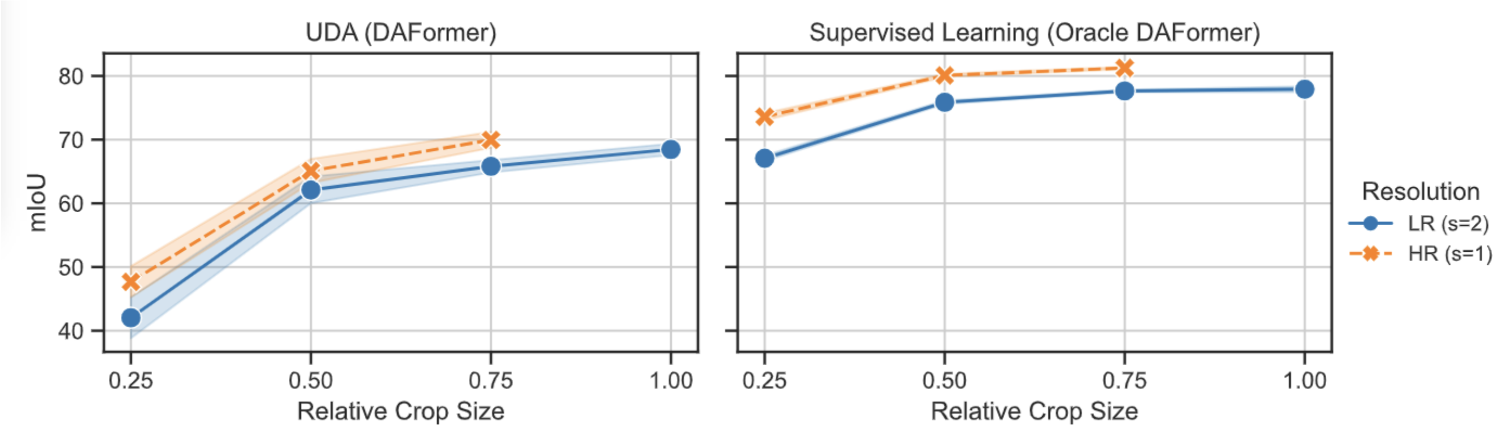
- **Experimental Results**
- **SYNTHIA :**

	Synthia $\rightarrow$ Cityscapes																			
CBST [97]	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
DACS [62]	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
BAPA [41]	91.7	53.8	83.9	22.4	0.8	34.9	30.5	42.8	86.6	–	88.2	66.0	34.1	86.6	–	51.3	–	29.4	50.5	53.3
CorDA [69]	**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
ProDA [84]	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
DAFormer [29]	84.5	40.7	88.4	41.5	**6.5**	50.0	55.0	54.6	86.0	–	89.8	73.2	48.2	87.2	–	53.2	–	53.9	61.7	60.9
HRDA	85.2	47.7	**88.8**	**49.5**	4.8	**57.2**	**65.7**	**60.9**	85.3	–	**92.9**	**79.4**	**52.8**	**89.0**	–	**64.7**	–	**63.9**	**64.9**	**65.8**

01 ProCST: Boosting Semantic Segmentation using Progressive Cyclic Style-Transfer

- **Experimental Results**

	UDA Method	Network	w/o HRDA	w/ HRDA	Improvement
1	Entropy Min. [65]	DeepLabV2 [4]	44.3 ±0.4	46.7 ±1.2	+2.4
2	Adversarial [63]	DeepLabV2 [4]	44.2 ±0.1	47.1 ±1.0	+2.9
3	DACS [62]	DeepLabV2 [4]	53.9 ±0.6	59.4 ±1.2	+5.5
4	DAFormer [29]	DAFormer [29]	68.3 ±0.5	73.8 ±0.3	+5.5



01 ProCST: Boosting Semantic Segmentation using Progressive Cyclic Style-Transfer

• Experimental Results

Table 3. HRDA context size. XR_a denotes crops with resolution XR (s_{LR}=2, s_{HR}=1) and relative crop size a=$h/\frac{H_T}{s_{XR}}$.

	Context Crop	Detail Crop	mIoU
1	$LR_{0.5}$	–	62.1 $\pm$2.1
2	–	$HR_{0.5}$	65.1 $\pm$1.9
3	$LR_{0.5}$	$HR_{0.5}$	68.5 $\pm$0.6
4	$LR_{0.75}$	$HR_{0.5}$	71.1 $\pm$1.7
5	$LR_{1.0}$	$HR_{0.5}$	73.8 $\pm$0.3

Table 4. HRDA detail size. XR_a denotes crops with resolution XR (s_{LR}=2, s_{HR}=1) and relative crop size a=$h/\frac{H_T}{s_{XR}}$.

	Context Crop	Detail Crop	mIoU
1	$LR_{1.0}$	–	68.5 $\pm$0.9
2	–	$HR_{0.25}$	47.7 $\pm$2.4
3	$LR_{1.0}$	$HR_{0.25}$	70.6 $\pm$0.7
4	$LR_{1.0}$	$HR_{0.375}$	71.7 $\pm$0.4
5	$LR_{1.0}$	$HR_{0.5}$	73.8 $\pm$0.3

Table 5. Comparison of HRDA with naive HR crops that have a comparable GPU memory footprint ($HR_{0.75}$).

	Context	Detail	Mem.	mIoU
1	–	$HR_{0.75}$	22.0 GB	70.0 $\pm$1.2
2	$LR_{0.75}$	$HR_{0.375}$	13.5 GB	71.3 $\pm$0.3
3	$LR_{1.0}$	$HR_{0.5}$	22.5 GB	73.8 $\pm$0.3

Table 6. HRDA detail crop variants. Up-LR: LR crop upsampled to HR resolution.

	Context Crop	Detail Crop	mIoU
1	$LR_{1.0}$	–	68.5 $\pm$0.9
2	$LR_{1.0}$	$LR_{0.5}$	69.1 $\pm$0.4
3	$LR_{1.0}$	Up-$LR_{0.5}$	71.9 $\pm$1.5
4	$LR_{1.0}$	$HR_{0.5}$	73.8 $\pm$0.3

01 ProCST: Boosting Semantic Segmentation using Progressive Cyclic Style-Transfer

- **Experimental Results**

Table 7. Component ablation of HRDA.

	Context	Detail	Scale Attention	Overlapping Detail	Detail Loss	mIoU
1	–	✓	–	–	–	65.1 ±1.9
2	✓	–	–	–	–	68.5 ±0.9
3	✓	✓	Average	–	–	67.5 ±0.8
4	✓	✓	Learned	–	–	71.5 ±0.5
5	✓	✓	Learned	✓	–	72.4 ±0.1
6	✓	✓	Learned	✓	✓	73.8 ±0.3