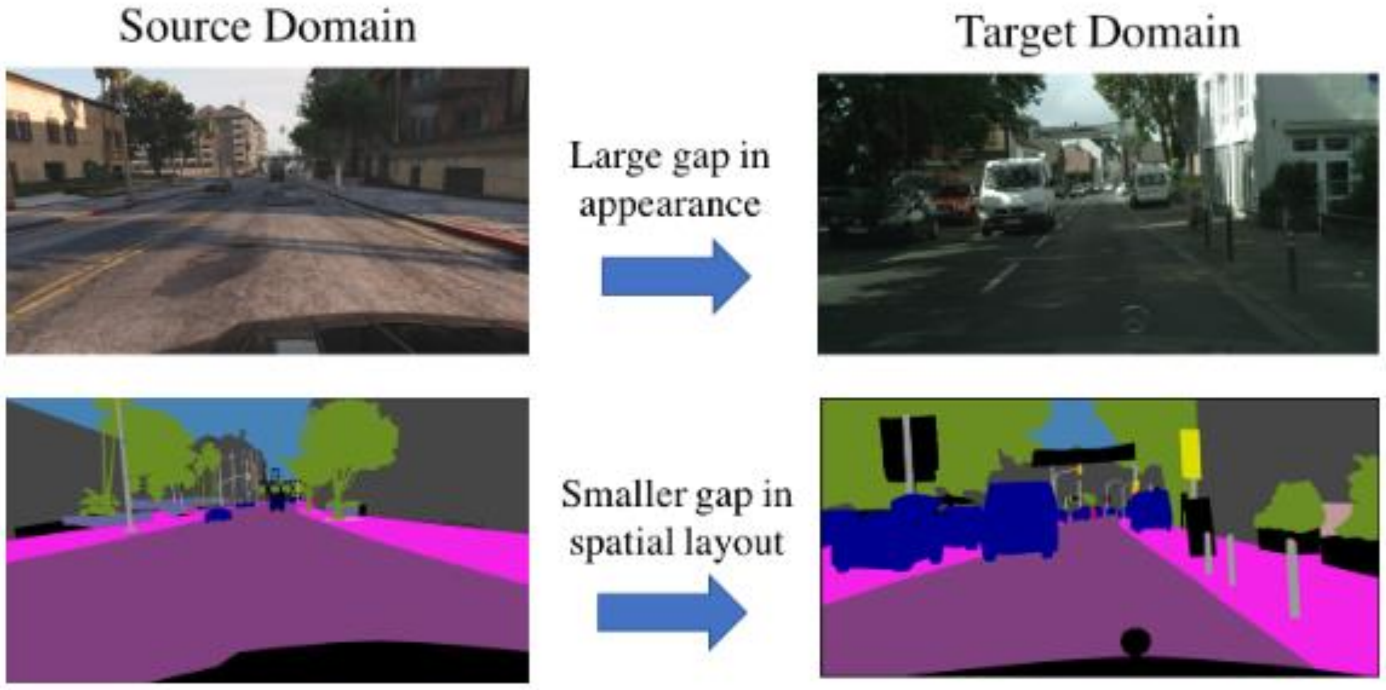
Domain Adaptive

Uncertainty-aware Pseudo Label Refinery for Domain Adaptive Semantic Segmentation

Yuxi Wang[1,2,5] Junran Peng[6] Zhaoxiang Zhang[1,2,3,4,5*]
[1] Institute of Automation, Chinese Academy of Sciences (CASIA)
[2] University of Chinese Academy of Sciences (UCAS)
[3] Centre for Artificial Intelligence and Robotics, HKISI_CAS
[4] Center for Excellence in Brain Science and Intelligence Technology, CAS
[5] National Laboratory of Pattern Recognition (NLPR) [6] Huawei Cloud&AI.
wangyuxi2016@ia.ac.cn, jrpeng4ever@gmail.com, zhaoxiang.zhang@ia.ac.cn

01 Foreground-aware Semantic Representations for Image Harmonization

- **Domain Adaptive**



01 Foreground-aware Semantic Representations for Image Harmonization

- **ICCV (2021)**

- **Motivation**

First, due to the class-imbalance, different categories are prone to have distinct transferability
Second, typical manually designed threshold methods [41, 40] generate pseudo labels according to the confidence scores, which is substantially hindered by the inevitable label noise

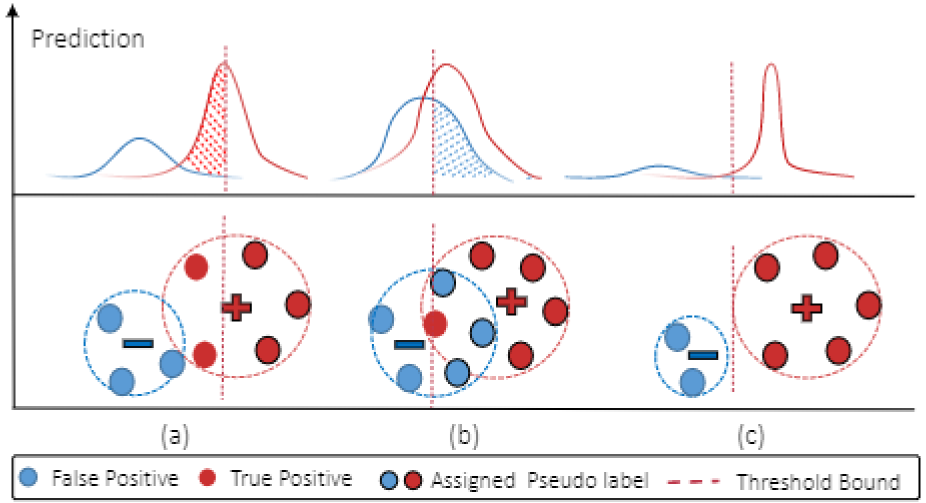


Figure 1. Problems for existing threshold-based pseudo labels generation. (a) Ignoring true positive predictions; (b) Assigning false positive predictions as labels (c) Ideal pseudo labels generation.

01 Foreground-aware Semantic Representations for Image Harmonization

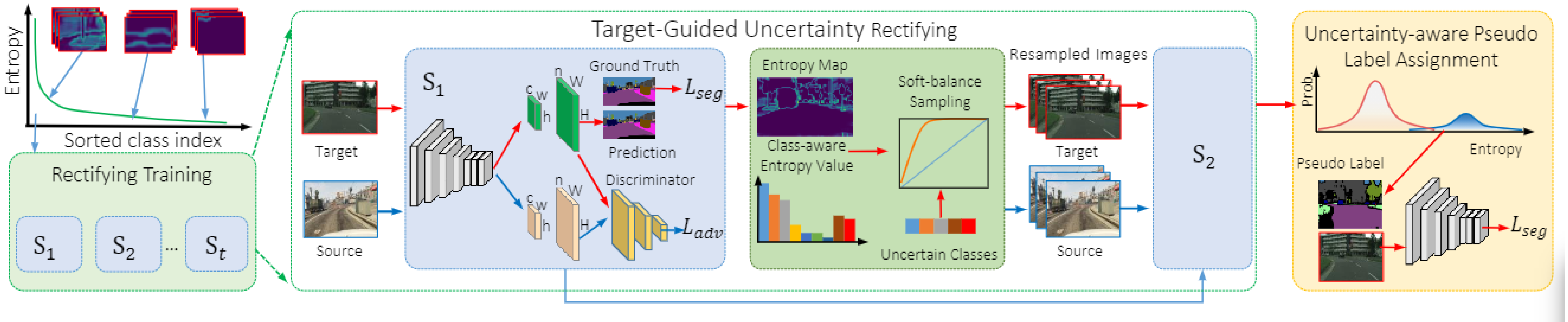
- **Contribution**
  - Propose to enhance the distribution alignment by resampling the training source images, whereas the resampling classes are designed according to the uncertainty statistics of the target domain.
  - Propose to select reliable pseudo labels by fitting the predictions to certainty and uncertainty modes using GMM. Pixels belonging to the certainty mode are assigned as pseudo labels.

01 Foreground-aware Semantic Representations for Image Harmonization

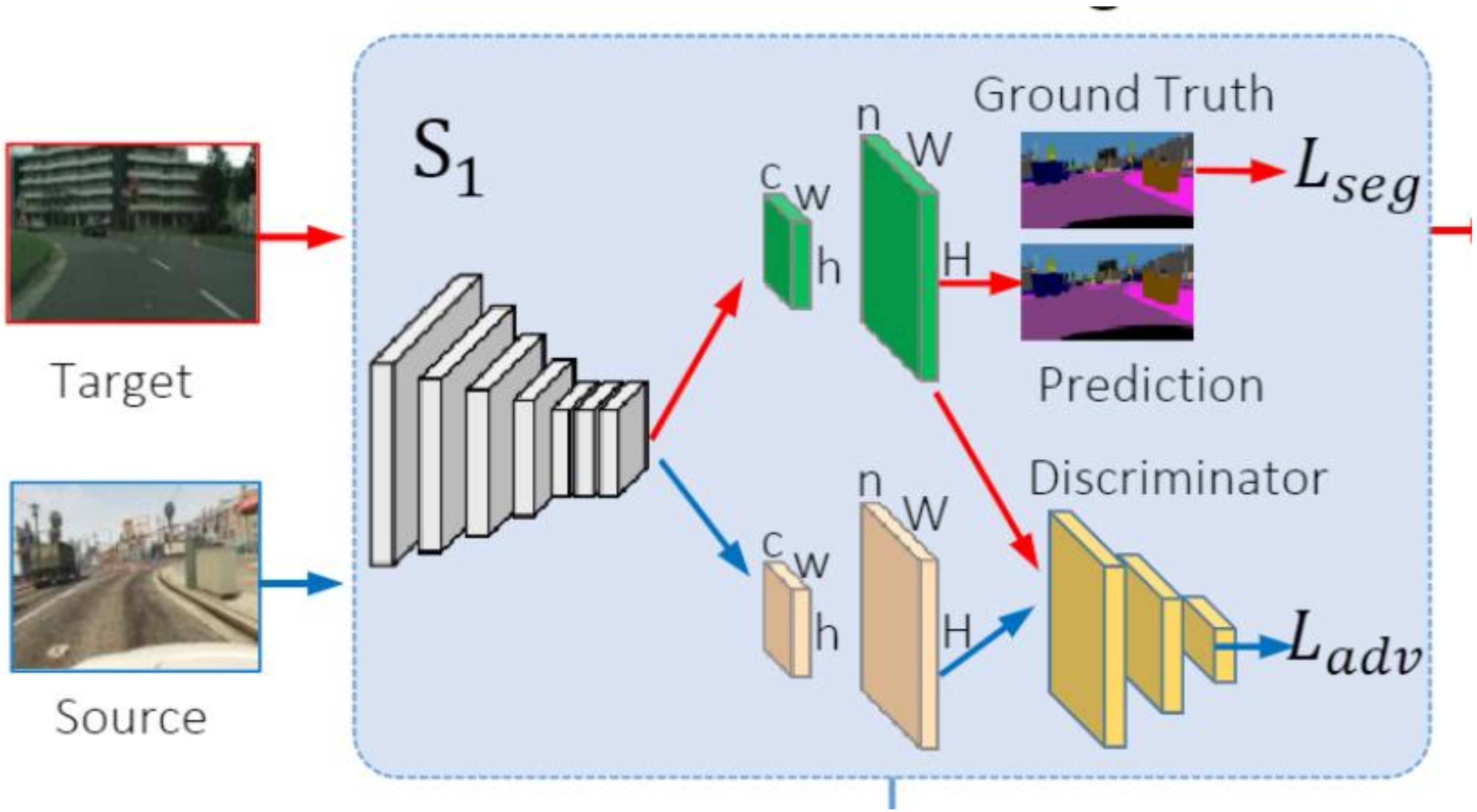
- **dataset**
  - GTA5 and SYNTHIA dataset
  - Cross-City dataset

01 Foreground-aware Semantic Representations for Image Harmonization

- **Overview of the Proposed Model**

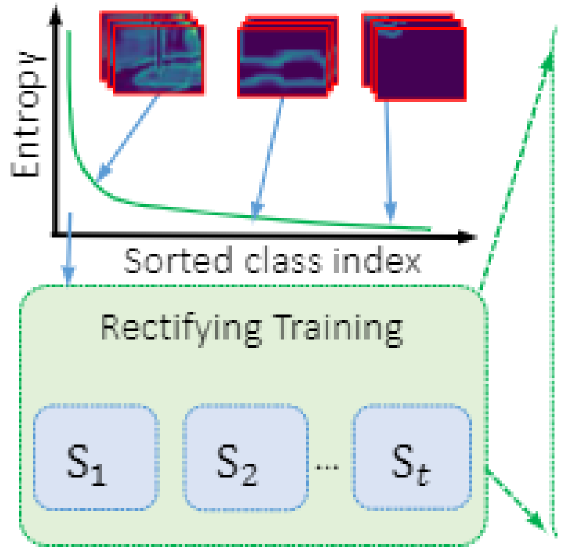


01 Foreground-aware Semantic Representations for Image Harmonization



01 Foreground-aware Semantic Representations for Image Harmonization

- **Target-guided uncertainty rectifying**



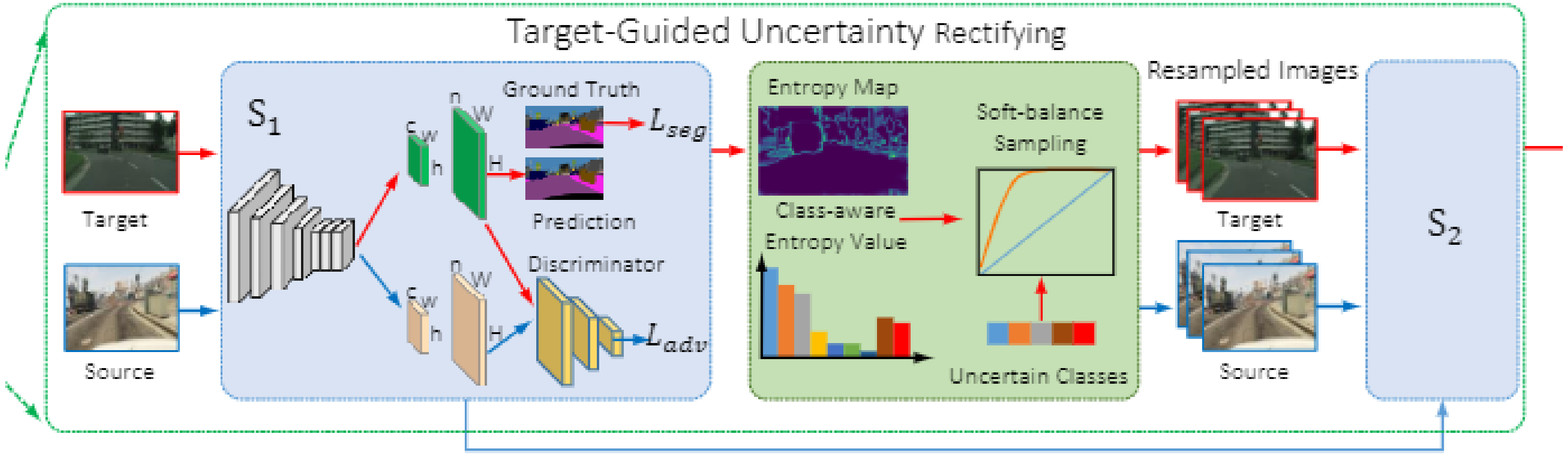
- To locate uncertainty-aware classes, we first calculate average category-level entropy $I_{\chi_t}^{c}$ on the whole target domain:

$$I_{\mathcal{X}_t}^{c} = \frac{1}{N_c} \sum_{x_t \in \mathcal{X}_t} \sum_{i} I_{x_t}^{(i)} * \mathbb{1}(\hat{y}_{x_t}^{(i,c)} = 1),$$

$$I_{x_t}^{(i)} = -\frac{1}{\log(C)} \sum_{c=1}^{C} p_{x_t}^{(i,c)} \log p_{x_t}^{(i,c)}.$$

- Then, we rank $I_{\chi_t}^{c}$ and obtain a subset S_k with top-k high-uncertain classes.

- **Soft-balance sampling for uncertainty-ware classes**



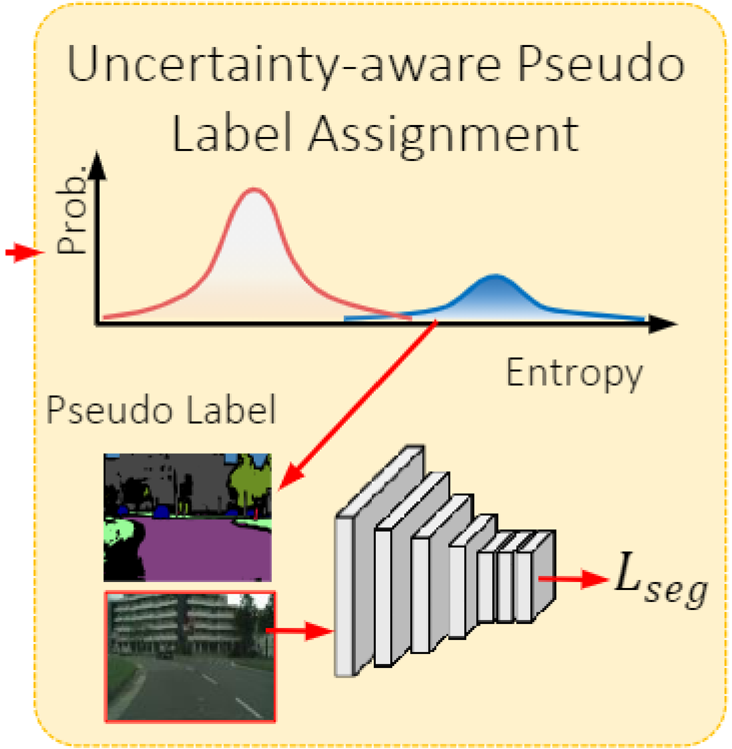
$$p_c(\mathcal{X}_s) = \frac{N_c(\mathcal{X}_s)^\lambda}{\sum_{c=1}^{C} N_c(\mathcal{X}_s)^\lambda} \frac{1}{N_c(\mathcal{X}_s)},$$

$$p_i(x_s) = \sum_{\hat{c}} \frac{N_{\hat{c}}(x_s)^\lambda \mathbb{1}(y_s^{\hat{c}} = 1)}{\sum_{\hat{c}} N_{\hat{c}}(x_s)^\lambda \mathbb{1}(y_s^{\hat{c}} = 1)} p_{\hat{c}}(\mathcal{X}_s),$$

$$\hat{p}_i(x_s) = 0.1 + \frac{1}{1 + \exp(-\alpha \times (p_i(x_s) - \mu))}.$$

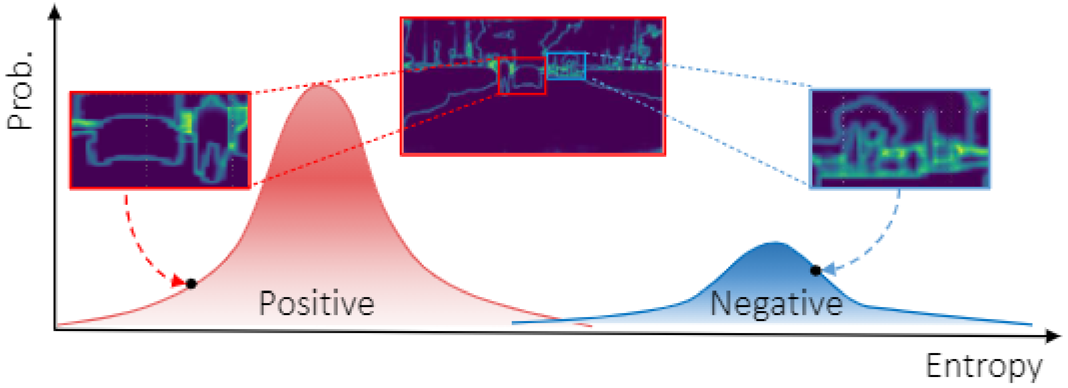
01 Foreground-aware Semantic Representations for Image Harmonization

- **Uncertainty-aware pseudo label assignment**



01 Foreground-aware Semantic Representations for Image Harmonization

- **Uncertainty-aware pseudo label assignment**



$$P_c(I^c_{x_t}) = w_{neg}\mathcal{N}_{neg}(I^c_{x_t};\mu_{neg},\sigma_{neg}) + w_{pos}\mathcal{N}_{pos}(I^c_{x_t};\mu_{pos},\sigma_{pos}),$$

- we use the Expectation-Maximization (EM) algorithm to optimize the distributions and weights (w_{neg},w_{pos}) following a uniform distribution.

01 Foreground-aware Semantic Representations for Image Harmonization

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

Table 1. Comparison to state-of-the-art methods of adaptation from GTA5 to Cityscapes based on ResNet-101 backbone. The top group is for adversarial adaptation ("AA"), while the bottom represents performance using self-training learning ("ST").

Method	road	side.	build.	wall	fence	pole	light	sign	vege.	terr.	sky	person	rider	car	truck	bus	train	motor.	bike	**mIoU**
AdaptSeg [31]	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
SIBAN [19]	88.5	35.4	79.5	26.3	24.3	28.5	32.5	18.3	81.2	**40.0**	76.5	58.1	25.8	82.6	30.3	34.4	3.4	21.6	21.5	42.6
CyCADA [11]	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
CLAN [20]	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
DISE [1]	91.5	47.5	82.5	31.3	25.6	33.0	33.7	25.8	82.7	28.8	**82.7**	62.4	30.8	85.2	27.7	34.5	6.4	25.2	24.4	45.4
ADVENT [33]	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
PatchAlign [32]	**92.3**	**51.9**	82.1	29.2	25.1	24.5	33.8	33.0	82.4	32.8	82.2	58.6	27.2	84.3	33.4	**46.3**	2.2	29.5	32.3	46.5
MRNet [38]	89.1	23.9	82.2	19.5	20.1	33.5	42.2	39.1	85.3	33.7	76.4	60.2	33.7	**86.0**	36.1	43.3	5.9	22.8	30.8	45.5
Ours (AA)	88.7	31.2	**83.7**	**34.1**	24.1	37.6	**42.9**	33.0	**85.8**	38.9	80.3	63.7	**34.2**	85.9	**41.2**	42.5	3.4	**33.8**	**42.5**	**48.8**
LSE [30]	90.2	40.0	83.5	31.9	26.4	32.6	38.7	37.5	81.0	34.2	84.6	61.6	33.4	82.5	32.8	45.9	6.7	29.1	30.6	47.5
PLCA [13]	84.0	30.4	82.4	35.3	24.8	32.2	36.8	24.5	85.5	37.2	78.6	**66.9**	32.8	85.5	40.4	48.0	8.8	29.8	41.8	47.7
BDL [15]	91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	**43.6**	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5
SIM [34]	90.6	44.7	84.8	34.3	28.7	31.6	35.0	37.6	84.7	43.3	85.3	57.0	31.5	83.8	**42.6**	48.5	1.9	30.4	39.0	49.2
TextDA [14]	**92.9**	**55.0**	85.3	34.2	31.1	34.9	40.7	34.0	85.2	40.1	**87.1**	61.0	31.1	82.5	32.3	42.9	0.3	36.4	46.1	50.2
FDA [36]	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
Zhe *et al.* [39]	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
Ours (ST)	90.5	38.7	**86.5**	**41.1**	**32.9**	**40.5**	48.2	42.1	**86.5**	36.8	84.2	64.5	**38.1**	**87.2**	34.8	50.4	0.2	**41.8**	**54.6**	**52.6**

01 Foreground-aware Semantic Representations for Image Harmonization

- **Ablation Experimental**
- **GTA5:**

Table 3. Ablation study on GTA5→Cityscapes. AA + IT acts as the baseline model with adversarial adaptation and image translation techniques; SR indicates the proposed soft-balance resampling strategy on source domain; TGAA is target-guided uncertainty rectifying adversarial adaptation; UPST stands for the proposed uncertainty-aware pseudo labels self-training process.

Method	AA	IT	SR	TGAA	UPST	mIoU
Source Only						36.6
+AA [31]	✓					42.9
+IT [15]	✓	✓				45.3
+SR	✓	✓	✓			46.1
+TGAA	✓	✓	✓	✓		48.8
+UPST	✓	✓	✓	✓	✓	**52.6**

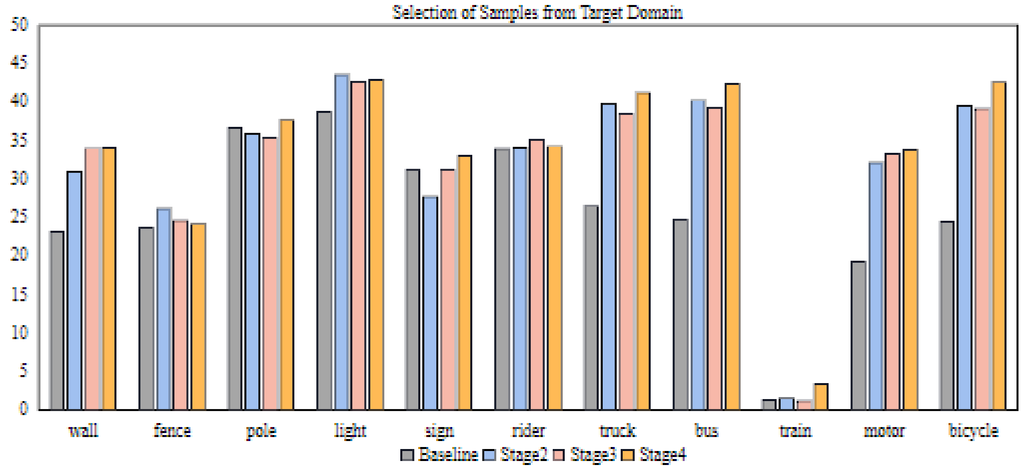


Figure 6. The influence of uncertainty-aware rectifying for selected infrequent categories on different sampling stages.

01 Foreground-aware Semantic Representations for Image Harmonization

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

Table 2. Comparison to state-of-the-art methods of adaptation from SYNTHIA to Cityscapes based on ResNet-101 backbone. The top group is for adversarial adaptation ("AA"), while the bottom group represents performance using self-training learning ("ST"). mIoU and mIoU* are averaged over 16 and 13 categories.

Method	road	side.	build.	wall	fence	pole	light	sign	vege.	sky	person	rider	car	bus	motor.	bike	**mIoU**	**mIoU***
AdaptSeg [31]	79.2	37.2	78.8	**10.5**	0.3	25.1	9.9	10.5	78.2	80.5	53.5	19.6	67.0	29.5	21.6	31.3	39.5	45.9
PatchAlign [32]	82.4	38.0	78.6	8.7	0.6	26.0	3.9	11.1	75.5	**84.6**	53.5	21.6	71.4	32.6	19.3	31.7	40.0	46.5
CLAN [20]	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
ADVENT [33]	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	41.2	48.0
DISE [1]	**91.7**	**53.5**	77.1	2.5	0.2	27.1	6.2	7.6	78.4	81.2	55.8	19.2	82.3	30.3	17.1	34.3	41.5	48.8
Ours (AA)	81.2	35.6	**81.5**	9.9	**0.8**	**35.9**	**29.6**	**19.9**	78.9	78.1	**62.8**	**27.1**	**83.7**	27.9	16.8	**53.1**	**45.2**	**52.0**
TextDA [14]	**92.6**	**53.2**	79.2	-	-	-	1.6	7.5	78.6	**84.4**	52.6	20.0	82.1	34.8	14.6	39.4	-	49.3
LSE [30]	82.9	43.1	78.1	9.3	0.6	28.2	9.1	14.4	77.0	83.5	58.1	25.9	71.9	38.0	29.4	31.2	42.6	49.4
BDL [15]	86.0	46.7	80.3	-	-	-	14.1	11.6	79.2	81.3	54.1	27.9	73.7	**42.2**	25.7	45.3	-	51.4
SIM [34]	83.0	44.0	80.3	-	-	-	17.1	15.8	**80.5**	81.8	59.9	**33.1**	70.2	37.3	28.5	45.8	-	52.1
FDA [36]	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
Ours (ST)	79.4	34.6	**83.5**	**19.3**	**2.8**	**35.3**	**32.1**	**26.9**	78.8	79.6	**66.6**	30.3	**86.1**	36.6	19.5	**56.9**	**48.0**	**54.6**

01 Foreground-aware Semantic Representations for Image Harmonization

- **Experimental Results**
-

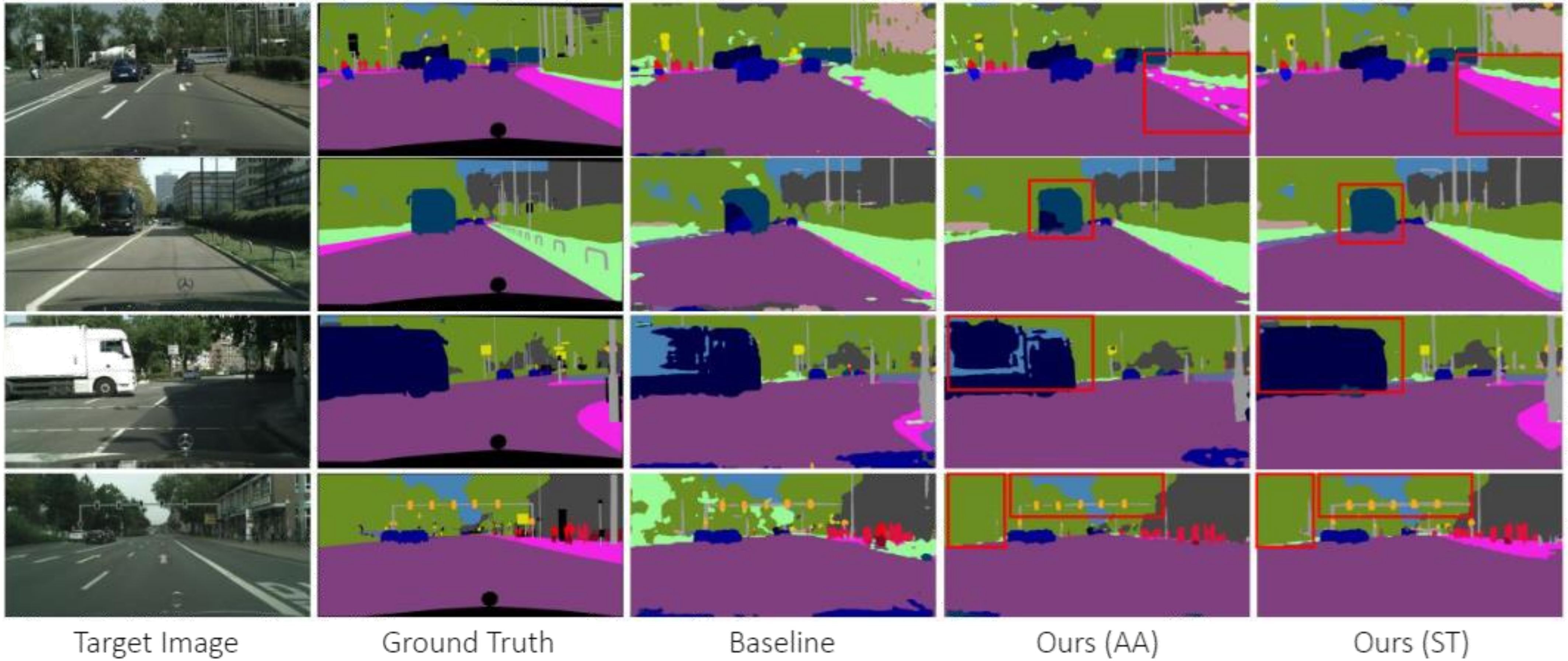


Figure 5. Visualization of the segmentation results. We perform our results from adversarial adaptation ("AA") and self-training learning ("ST"), respectively. The "baseline" model is achieved with adversarial learning and image transfer.