Informative Classes Matter: Towards Unsupervised Domain Adaptive Nighttime Semantic Segmentation

Shiqin Wang
Wuhan University of Science and Technology
wangshiqin@wust.edu.cn

Xin Xu*
Wuhan University of Science and Technology
xuxin@wust.edu.cn

Xianzheng Ma
Shanghai AI Laboratory
maxianzheng@pjlab.org.cn

Kui Jiang
Harbin Institute of Technology
jiangkui@hit.edu.cn

Zheng Wang
Wuhan University
wangzwhu@whu.edu.cn

ABSTRACT

Unsupervised Domain Adaptive Nighttime Semantic Segmentation (UDA-NSS) aims to adapt a robust model from a labeled daytime domain to an unlabeled nighttime domain. However, current advanced segmentation methods ignore the illumination effect and class discrepancies of different semantic classes during domain adaptation, showing an uneven prediction phenomenon. It is the completely ignored and underexplored issues of "hard-to-adapt" classes that some classes have a large performance gap between existing UDA-NSS methods and supervised learning counterparts while others have a very low performance gap. To realize "hard-to-adapt" classes' more sufficient learning and facilitate the UDA-NSS task, we present an Online Informative Class Sampling (OICS) strategy to adaptively mine informative classes from the target nighttime domain according to the corresponding spectrogram mean and the class frequency via our Informative Mixture of Experts. Furthermore, an Informativeness-based cross-domain Mixed Sampling (InforMS) framework is designed to focus on informative classes from the target nighttime domain by vesting their higher sampling probabilities when cross-domain mixing sampling and achieves better performance in UDA-NSS tasks. Consequently, our method outperforms state-of-the-art UDA-NSS methods by large margins on three widely-used benchmarks (*e.g.* , ACDC, Dark Zurich, and Nighttime Driving). Notably, our method achieves state-of-the-art performance with 65.1% mIoU on ACDC-night-test and 55.4% mIoU on ACDC-night-val.

CCS CONCEPTS

• **Information systems** → *Multimedia information systems*; • **Computing methodologies** → *Computer vision*; • **Applied computing** → *Computer forensics.*

*Corresponding author.

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org/10.1145/3581783.3611956

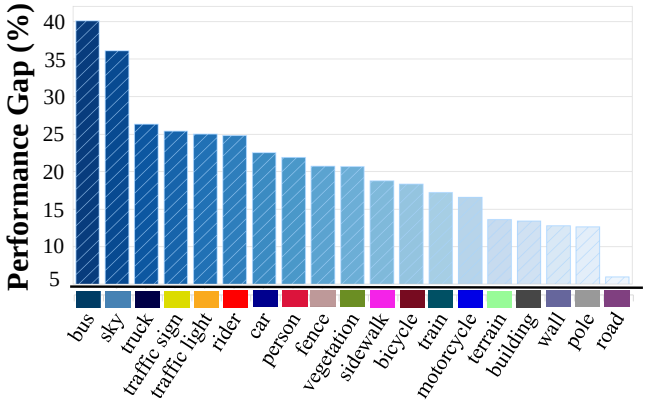

Figure 1: Uneven prediction phenomenon. The horizontal coordinate represents the semantic classes and the vertical coordinate demonstrates their Performance Gap (*i.e.* , average IoU [%] gap) between existing UDA-NSS methods and supervised supervision counterparts. We find the uneven prediction phenomenon that some classes (*e.g.* , bus and sky) have a larger performance gap while others (*e.g.* , pole and road) have a very low performance gap. In detail, we compute the average mIoU of 19 semantic classes on the nighttime dataset (*i.e.* , the ACDC-night-val and the ACDC-night-test from the ACDC dataset [41]) via existing UDA-NSS methods [3, 14, 15, 53, 66] and their corresponding fully supervision methods at the target nighttime domain, respectively.

KEYWORDS

unsupervised domain adaptive nighttime semantic segmentation; class sampling

ACM Reference Format:
Shiqin Wang, Xin Xu, Xianzheng Ma, Kui Jiang, and Zheng Wang. 2023. Informative Classes Matter: Towards Unsupervised Domain Adaptive Nighttime Semantic Segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3581783.3611956

1 INTRODUCTION

Nighttime Semantic Segmentation (NSS), aiming to label each pixel of a given nighttime image to an object category, has been widely applied to autonomous driving [24, 53, 54, 69], visual surveillance [45,

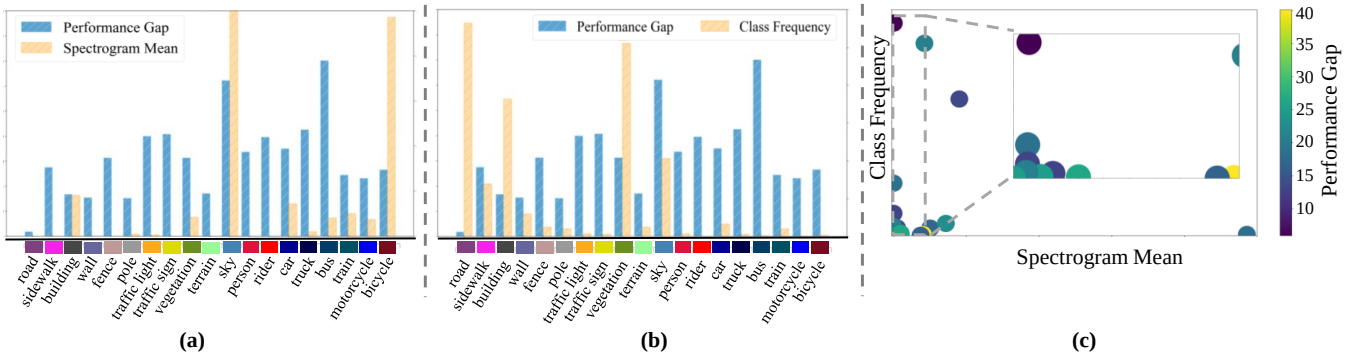

Figure 2: Correlation between Performance Gap and Spectrogram Mean (a) / Class Frequency (b) of each class in the target nighttime domain predictions when existing UDA-NSS methods [3, 14, 15, 53, 66] testing on the ACDC-night-val and the ACDC-night-test dataset. Then, we show the joint influence of Spectrogram Mean and Class Frequency on the Performance Gap (c). Each point symbolizes a class. Note that the top left corner, *i.e.* , classes with a low spectrogram mean and a high class frequency have a low performance gap.

50, 51, 63], and robotic vision [11]. It is easy to obtain large-scale real-world nighttime images, yet constructing corresponding nighttime datasets with well-labeled pixel-level annotations is labor-intensive and time-consuming, which makes traditional supervised learning methods [37, 61, 65] encounter bottlenecks in NSS tasks. Some researchers multiplex strategies of Daytime Semantic Segmentation (DSS), but encounter serious performance degradation at night due to the large domain gap between daytime and nighttime data [35, 54, 60]. To alleviate the data annotation requirement and bridge the domain gap, Unsupervised Domain Adaptive Nighttime Semantic Segmentation (UDA-NSS) [53, 54] attracts much attention to transferring the labeled daytime domain knowledge to the unlabeled nighttime domain through domain adaptation.

To facilitate the UDA-NSS task, on the one hand, previous UDA-NSS methods primarily narrow the illumination gap via utilizing an intermediate twilight domain as the bridge [9, 39, 40] or utilizing the image transferring network to transform different domains as the same style [12, 27, 38, 42, 44, 45, 49, 53, 54, 60, 66], **ignoring the illumination degradation of different semantic classes.** On the other hand, due to the class imbalance problem in the training samples, different classes are prone to have distinct transferability [52], i.e., some classes that occupy small portions of the pixel are inherently hard to transfer across domains. Existing methods usually introduce re-weighting strategy [27, 53, 54] to boost the performance for classes of smaller-size objects via assigning their higher weight or train the target nighttime samples in a class-balanced manner [60, 66] to optimize the model, **ignoring weights dynamic adjustment of different semantic classes during domain adaptation.**

The above two issues result in some "hard-to-adapt" classes severely insufficient learning. As shown in Figure 1, it is reported that there exists an uneven prediction phenomenon that some classes have a larger performance gap between existing UDA-NSS methods and supervised learning counterparts while others have a very low performance gap. Even though existing re-sampling [32, 57, 71, 72] and re-weighting [33, 34] strategies designed for UDA semantic segmentation are applied in UDA-NSS, the uneven prediction phenomenon is still not well addressed. This phenomenon indicates that existing UDA-NSS methods mainly tend to easy-to-adapt classes learning but neglect hard classes during domain adaptation. Another way to look at these hard classes, they have more considerable potential for improvement and contribute more to bridging the domain gap, dubbed as the "informative" classes in this study. In this way, we argue that the performance of the UDA-NSS model can be significantly improved by mining "informative" classes and then training them effectively during domain adaptation. However, labels of the target nighttime domain are unavailable during domain adaptation, so the performance gap of each class is unknown. That's why the classes that need to be focused on cannot be selected solely based on the performance gap of each class shown in Figure 1. Therefore, the dynamic mining and training of "informative" classes during domain adaptation is essential to encourage maximizing segmentation performance on the target nighttime domain.

Based on this observation, a new perspective from cross-domain mixed sampling exploiting informative class mining is pioneered in this study to facilitate the UDA-NSS task. For a better understanding, we systematically investigate this problem from two aspects: *1) How to mine informative classes adaptively during domain adaptation? 2) How to boost the UDA-NSS task with informative classes?*

Targeting the first question, we consider illumination degradation and transferability of different semantic classes. Inspired by [5, 58, 67], the low-level spectrum can scene illumination variations, and such image frequency distribution difference can represent the illumination gap. Furthermore, the class frequency can be dynamically obtained from the portion of every class in target nighttime domain predictions, reflecting the familiarity of the UDA-NSS model with the corresponding class. And such class frequency can effectively reflect the transferability of different semantic classes. We correspondingly compute the spectrogram mean and the class frequency of each class in the target nighttime domain predictions and corresponding input images, respectively. In detail, we utilize the Pearson Correlation Coefficient [1] (γ) [7] to measure the correlation between the performance gap and the above two characteristics

[1] A common metric that measures the strength and direction of the relationship between two variables. The range is from -1 to 1. A value closer to -1 or 1 indicates a stronger negative or positive correlation, respectively, and 0 implies no correlation.

(See Figure 2). It is observed the performance gap has a relatively strong positive/moderate negative correlation with the above two characteristics (γ is 0.64 and -0.41, respectively). To leverage the complementary information of the above two characteristics, we propose the Informative Mixture of Experts to dynamically assign the importance of each character ("informative class metric expert"), thus constructing an Online Informative Class Sampling (OICS) strategy. For the second aspect, during the domain adaptation, cross-domain mixed sampling can implicitly help the network to better deal with the domain gap [6, 28, 48] and focusing more on informative classes from the target domain contribute better to the domain adaptation. To this end, we design an Informativeness-based cross-domain Mixed Sampling (InforMS) framework, which focuses on informative classes via vesting their higher sampling probabilities when cross-domain mixing sampling and realizes some hard-to-adapt classes' sufficient learning. The main contributions are summarized as follows:

- We empirically find an uneven prediction phenomenon and pioneer a new perspective from cross-domain mixed sampling exploiting informative class mining to facilitate the UDA-NSS task, unexplored by previous methods.
- We present an Online Informative Class Sampling strategy, which can adaptively mine informative classes according to the spectrogram mean and the class Frequency of each class via our Informative Mixture of Experts. Besides, we designed an Informativeness-based cross-domain Mixed Sampling (InforMS) framework to realize some hard-to-adapt classes' more sufficient learning.
- Experiments on three widely-used benchmarks show that our proposed method outperforms state-of-the-art UDA-NSS methods by large margins and achieves a new state-of-the-art performance of nighttime semantic segmentation.

2 RELATED WORK

Previous UDA-NSS methods primarily narrow the illumination gap via the intermediate twilight domain [9, 39, 40] or the image transferring network [12, 27, 38, 42, 44, 45, 49, 53, 54, 60, 66]. Dai *et al.* [9] first leverages an intermediate twilight domain as the bridge to progressively transfer knowledge from the source daytime domain to the target nighttime domain. Sakaridis *et al.* [39, 40] further extend it by leveraging geometry information to refine the semantic predictions. However, introducing twilight domains for gradual domain adaptation brings a more significant computational burden. Motivated by the widely studied image style transfer [17, 21, 26], one can try translating different domains into the same style. Earlier studies [38, 42, 44, 45, 60] along this direction apply adversarial models, *e.g.,* CycleGAN [70], to achieve the style translation from daytime to nighttime or vice versa. Recently, some studies utilize the image relighting network [49, 53, 54, 66] or a light enhancement network [27] to transfer different domains as the same style. However, all the above UDA-NSS methods ignore the illumination degradation of different semantic classes, resulting in some "hard-to-adapt" classes severely insufficient learning.

The class imbalance issue also exacerbates the problem, as different classes exhibit distinct transferability [52]. Classes with small portions of the pixel are inherently hard to transfer across domains.

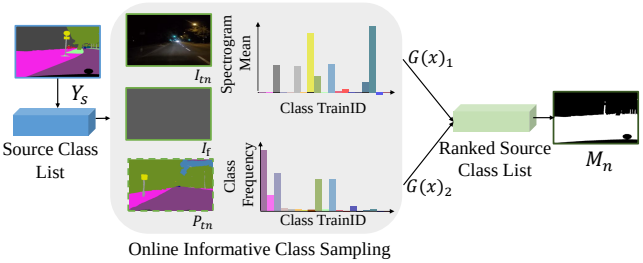

Figure 3: Online Informative Class Sampling. During training, the initial step is to derive the source class list from the ground truth (Y_s) of the source image (I_s). Following this, the spectrogram mean, and class frequency for each of the 19 semantic classes are computed, utilizing predictions from all target nighttime domain images and corresponding spectrogram segmentation maps generated from nighttime images (T_n). Subsequently, the source class list is ranked based on informativeness in descending order, dividing it into more informative classes (upper half) and less informative classes (lower half). Finally, a binary mask (M_n) is created, assigning a value of one to pixels in the ground truth (Y_s) of the source image from less informative classes and zero to all others.

Existing UDA-NSS approaches [27, 53, 54] utilize a re-weighting strategy to boost the prediction accuracy of smaller-size objects, which assigns the weight according to its proportion of pixels in the source domain, *i.e.,* the smaller the proportion and the higher the weight. Meanwhile, considering the class imbalance in the target nighttime domain, Yang *et al.* [66] train the nighttime samples in a class-balanced manner via mixing pixels of randomly selected classes from the source sample with the nighttime sample. Xu *et al.* [60] adopts an offline class-balanced process of generating pseudo labels. However, all the above UDA-NSS methods neglect weights dynamic adjustment of different semantic classes, leading to severe underlearning of "hard-to-adapt" classes.

3 METHODOLOGY

In this section, we investigate the UDA-NSS task from two aspects: 1) How to mine informative classes adaptively during domain adaptation? 2) How to boost the UDA-NSS task with informative classes?

3.1 Online Informative Class Sampling

Here we present the Online Informative Class Sampling strategy to adaptively mine informative classes according to the spectrogram mean and the class frequency metric via our Informative Mixture of Experts, as illustrated in Figure 3.

3.1.1 Spectrogram Mean. The principal disparity between the source daytime domain and the target nighttime domain stems from the illumination difference [23, 36, 45, 53, 54, 63, 68]. Challenges in domain adaptation arise from low-light conditions and non-uniform illumination at nighttime in the target domain [10, 18–20, 46]. The low-level spectrum represents scene illumination variations, while image frequency distributions can reflect the day-time/night-time discrepancies [5, 58, 67]. To illustrate this point, we conducted a

quantitative dataset-level study. Utilizing the Discrete Cosine Transform (DCT), we computed the spectrogram mean of images from the Cityscapes [8] (source daytime domain) and the ACDC-N [41] (target nighttime domain), respectively [59]. Our analysis revealed that the spectrogram mean of nighttime scenes (0.057566497) exceeded that of daytime scenes (-5.74667e-05), indicating illumination discrepancies. Crucially, higher spectrogram mean values are associated with more informative classes. Thus, the "Spectrogram Mean" can effectively evaluate the informativeness of each class, reflecting illumination discrepancies between the two domains.

To begin with, we obtain the predictions of all target nighttime domain images. We employ the two-dimensional DCT to transfer the spatial domain to the frequency domain as follows:

$$F(u,v) = c(u)c(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)B_{x,y}^{u,v}$$
$$B_{x,y}^{u,v} = \cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)v\pi}{2N} \tag{1}$$

where u and v denote the horizontal and vertical frequency components, respectively. (x, y) is the spatial locations of the image block and N represents the size of the image block. $F(u, v)$ is the 2D DCT frequency spectrum, $u, v \in \{0, 1, \cdots, n-1\}$, and $f(x, y)$ is a two-dimensional vector element of $N \times N$ in the spatial domain, $x, y \in \{0, 1, \cdots, N-1\}$. $c(u)$ and $c(v)$ are compensation factors as:

$$c(u), c(v) = \begin{cases} \sqrt{\frac{1}{N}}, & u, v = 0 \\ \sqrt{\frac{2}{N}}, & u, v \neq 0 \end{cases} \tag{2}$$

Then, we compute the spectrum mean of each class as follows:

$$I_c = \frac{1}{N_c}\sum_{i=1}^{N_c}\sum_{j=1}^{H\times W} F(u,v)^{(j,c)}; c \in C \tag{3}$$

where N_c denotes the number of pixels for the c-th class in the target nighttime domain predictions. H and W denote the height and width of the image, respectively.

3.1.2 Class Frequency. Motivated by that deep neural networks tend to predict their familiar classes and ignore unfamiliar classes. Hence, the "Class Frequency" can be employed to metric the informativeness level of each class, which can reflect the familiarity of the model with the class. That is if the class is rarely predicted by the model, *i.e.* , the class is unfamiliar to the model. The lower the class frequency, the more informative the class.

We first compute the c-th class frequency f_c of the whole target nighttime domain's predictions.

$$f_c = [\frac{\sum_{i=1}^{N}\sum_{j=1}^{H\times W}\left[c = \arg\max_{c'} g_\theta\left(x_T^{(i)}\right)^{(j,c')}\right]}{N \cdot H \cdot W}] \tag{4}$$

where N represents the number of images in the target nighttime domain, and $g_\theta\left(x_T^{(i)}\right)^{(j,c')}$ denotes the segmentation prediction of the target nighttime domain.

3.1.3 Informative Mixture of Experts. Rather than training a single network with one informative class metric, we propose an informative Mixture of Experts to leverage the complementary information

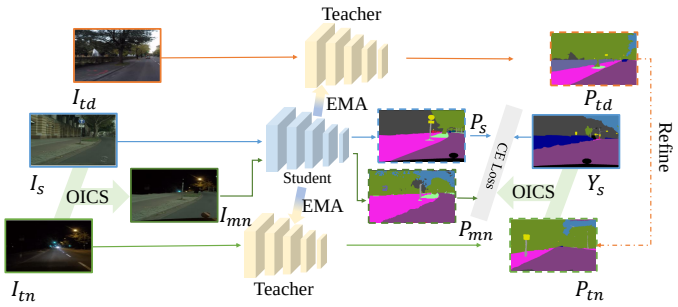

Figure 4: Informativeness-based cross-domain Mixed Sampling (InforMS) framework. Three input image I_s, I_{td} and I_{tn} are from the source domain S_d and two target domains T_d and T_n, respectively. They undergo the corresponding semantic segmentation network, yielding predictions P_s, P_{td}, and P_{tn}. Our proposed OICS strategy is employed to generate binary masks (M_n) by adaptively mining informative classes from T_n. Consequently, the mixed images (I_{mn}) and mixed labels (Y_{mn}) are generated. The student net subsequently processes the I_{mn} to ascertain segmentation predictions (P_{mn}). The final stage involves optimizing the image predictions and their associated labels using a categorical Cross-Entropy (CE) Loss.

from the above two informative class metrics (also termed "informative class metric experts"). Specifically, we dynamically assign the importance of each expert during domain adaptation as follows:

$$y = \sum_{i=1}^{n} G(x)_i E_i(x) = G(x)_1 I_c + G(x)_2 f_c \tag{5}$$

where $\sum_{i=1}^{n} G(x)_i = 1$ and $G(x)_i$ represents the weight assigned to the expert $f_i(x)$. And $G(x)_i$ is determined by the informativeness gains ($\triangle E_i(x)$) of informative classes. In detail, we first assign two experts the corresponding normalized informativeness gains (mIoU gains). Then, we calculate the spectrogram mean and the class frequency of informative classes at time t and t+1, respectively. When the spectrogram mean decreases or the class frequency increases, the weight of the corresponding expert is increased by 0.1. Otherwise, it remains the same. Finally, we obtain the normalized weight of each expert, which will be used in the next iteration.

3.2 Informativeness based cross-domain Mixed Sampling (InforMS) framework

Figure 4 depicts the overall architecture of our proposed method. It involves a source daytime domain S_d and two coarsely aligned target domains T_d and T_n, where T_d and T_n represent the target daytime and nighttime domain, respectively. Only the source domain S_d has ground-truth semantic segmentation during training. Three images I_{td}, I_d, and I_{tn}, sampled from T_d, S_d, and T_n, respectively, serve as input. We first feed the inputs into the corresponding semantic segmentation network to obtain their segmentation predictions (category-likelihood map) P_{td}, P_s, and $P_{tn}^{'}$ for the three domains, respectively. Considering there exists coarse alignment of semantic content between the target daytime image and the target nighttime image, we then utilize the refinement module [3] to refine $P_{tn}^{'}$ using the P_{td}. In online self-training, the weights of the teacher

Table 1: Comparison of state-of-the-art unsupervised domain adaptation on Cityscapes→ACDC adaptation for nighttime. The per-category results on ACDC-night-test are shown below. The best results are presented in bold

Method	Backbone	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIoU**
GCMA [39]	RefineNet	78.6	45.9	58.5	17.7	18.6	37.5	43.6	43.5	58.7	39.2	22.5	57.9	29.9	72.1	21.5	56.3	41.8	35.7	35.4	42.9
MGCDA [40]	RefineNet	74.5	52.5	69.4	7.7	10.8	38.4	40.2	43.3	61.5	36.3	37.6	55.3	25.6	71.2	10.9	46.4	32.6	27.3	33.8	40.8
DANNet [53]	DeepLab-v2	90.0	55.6	71.1	30.8	21.0	24.3	24.1	30.7	66.1	28.2	75.9	38.9	31.1	60.6	10.4	30.3	42.0	22.0	31.4	41.3
DANNet [53]	RefineNet	91.3	60.5	77.2	38.9	27.2	28.0	29.2	33.3	68.7	32.8	80.1	46.2	26.4	71.9	36.0	53.7	67.5	39.3	36.1	49.7
DANNet [53]	PSPNet	90.5	62.1	76.5	34.9	30.3	27.1	29.9	31.4	69.4	41.0	77.7	48.9	29.6	67.9	15.7	43.7	54.3	29.8	36.2	47.2
DANIA [54]	DeepLab-v2	89.4	55.1	75.2	37.5	27.1	33.1	37.0	37.2	70.5	34.2	78.9	47.8	32.4	64.9	6.1	34.8	41.7	27.3	36.2	45.6
DANIA [54]	RefineNet	92.1	63.0	79.0	39.3	30.7	33.1	39.3	40.2	70.7	37.5	80.6	49.1	28.9	71.5	48.3	56.0	59.0	43.3	36.9	52.6
DANIA [54]	PSPNet	91.0	60.9	77.7	40.3	30.7	34.3	37.9	34.5	70.0	37.2	79.6	45.7	32.6	66.4	11.1	37.0	60.7	32.6	37.9	48.3
Bi-Mix [66]	DeepLab-v2	88.7	58.1	71.1	34.2	22.3	27.3	26.9	31.9	63.2	26.4	64.5	39.2	25.3	62.6	3.4	31.9	32.9	25.2	29.6	40.2
Bi-Mix [66]	RefineNet	90.3	60.8	72.5	38.8	23.7	33.3	22.0	31.1	62.5	20.3	60.8	45.1	24.9	70.2	55.0	50.8	59.6	38.8	34.1	47.1
Bi-Mix [66]	PSPNet	90.0	60.3	67.2	37.7	24.3	31.9	31.9	32.6	67.7	32.6	52.6	44.1	27.8	62.9	32.4	39.0	53.9	33.0	34.1	45.1
GPS-GLASS [22]	PSPNet	91.8	65.0	76.4	38.1	30.0	35.8	38.5	37.6	69.2	41.4	79.8	45.8	31.2	69.6	38.0	59.9	45.7	24.9	37.2	50.3
Refign [3]	DeepLab-v2	92.7	67.5	82.0	40.2	21.0	39.2	36.5	44.1	60.1	35.4	53.9	54.5	28.0	75.1	18.4	49.0	62.5	42.0	43.7	49.8
InforMS (Ours)	DeepLab-v2	93.9	69.9	81.7	37.0	31.6	40.7	36.0	44.2	70.2	39.5	76.6	52.8	33.3	75.1	14.6	51.4	71.6	29.7	45.2	52.4
DAFormer [14]	DAFormer	92.3	64.6	70.1	28.7	18.5	45.8	11.3	41.5	42.7	41.9	0.0	55.4	30.0	74.3	40.3	45.8	81.4	39.4	47.0	45.8
Refign [3]	DAFormer	91.0	63.5	81.7	41.6	18.9	50.1	51.6	53.4	72.0	46.9	80.6	59.4	29.5	71.4	15.0	50.0	76.9	42.6	45.6	54.8
InforMS (Ours)	DAFormer	93.4	71.2	84.6	45.8	22.8	50.2	50.4	50.2	75.0	48.3	83.1	58.8	30.1	73.5	16.9	53.6	79.9	46.1	47.3	56.9
HRDA [15]	HRDA	87.2	46.9	79.1	46.2	18.0	51.4	41.0	48.5	41.8	46.7	0.0	63.2	36.9	81.0	**65.2**	**77.7**	83.6	46.0	49.0	53.1
Refign [3]	HRDA	94.8	76.7	86.5	**55.8**	41.7	47.7	55.3	55.3	**79.4**	**50.4**	**86.9**	65.2	40.6	81.2	46.7	75.6	83.6	34.0	48.7	63.5
InforMS (Ours)	HRDA	**95.3**	**77.9**	**87.8**	52.0	**42.5**	**51.6**	**56.2**	**63.0**	78.3	49.1	86.5	**67.0**	**41.8**	**82.9**	46.8	69.1	**88.8**	**48.1**	**52.1**	**65.1**

net h_ϕ are set as the Exponential Moving Average (EMA) of the weights of the student net g_θ after each training step t [1, 47].

During early training stages, the network's predictions are unreliable, and it tends to learn easy-to-adapt classes [2, 25, 29, 56, 62, 64]. Following DACS [48], we randomly select half of the source domain image's classes to constitute the final class list c. Later in training, the model converges gradually, and more attention should be paid to revising hard-to-adapt class predictions [3]. This progression can be captured through the model confidence, *i.e.* , the normalized entropy of the target nighttime probability maps. At that time, we obtain the informativeness level of each class via our proposed OICS strategy and rank the source domain image's class informativeness in ascending order. The front half of the ranked classes list is selected as informative classes to form the final class list c.

Corresponding binary mask M is generated via setting the pixels from the final class list c to value one in M and all others to value zero. By doing so, the mixed image (I_{mn}) can preserve as many informative classes as possible from the target nighttime domain, which can promote the segmentation of informative classes at nighttime. For labels, we mix the ground-truth label of the source domain image (Y_s) with the refined pseudo-labels of the target nighttime domain image (P_{tn}) in the same way as the mixed images to obtain the mixed pseudo labels (Y_{mn}). Finally, we apply CE Loss [43] to the images' predictions and corresponding labels.

4 EXPERIMENTS

4.1 Datasets and evaluation metrics

The mean of category-wise intersection-over-union (mIoU) is applied as the evaluation metric, where higher values indicate better performance. The following datasets are utilized for model training and performance evaluation:

Cityscapes [8]: The Cityscapes dataset contains 2,975 training, 500 validation, and 1,525 testing images. This paper uses the Cityscapes training set as the source domain.

ACDC [41]: The ACDC dataset includes four adverse weather conditions: fog, snow, rain, and nighttime. Each condition is divided into 400 training, 100 validation, and 500 test images, except for the nighttime set, which has 106 validation images. In this paper, we utilize the nighttime training images of ACDC (ACDC-N) and their corresponding normal-condition images (ACDC-D) as the target domain. ACDC-night-val and ACDC-night-test are adopted for evaluation. The test set is withheld for testing online[2].

Dark Zurich [39]: The Dark Zurich dataset contains 2,416 nighttime images, 2,920 twilight images, and 3,041 daytime images for training, all unlabeled. Following previous works [12, 53, 54, 66], we utilize 2,416 coarsely aligned night-day image pairs as the target dataset. The Dark Zurich dataset also includes 50 annotated for validation (Dark Zurich-val) and 151 for testing (Dark Zurich-test). Dark Zurich-test is withheld for testing online[3].

Nighttime Driving [9]: The Nighttime Driving test set contains 50 nighttime images. In our study, the test set is utilized for evaluation.

4.2 Experimental settings

Our proposed InforMS is implemented using PyTorch on a single RTX 3090 GPU. To showcase the flexibility of InforMS, we combine it with state-of-the-art UDA methods. We choose Refign [3] (using DeepLab-v2 [4], DAFormer [14], and HRDA [15]) as the baseline. In the training stage, following [14, 30, 55], we train the network with AdamW [31], a learning rate of $\eta_{\text{base}} = 6 \times 10^{-5}$ and a weight decay of 0.01. We incorporate linear learning rate warmup [13] with

[2]ACDC website: https://acdc.vision.ee.ethz.ch/

[3]https://codalab.lisn.upsaclay.fr/competitions/3783

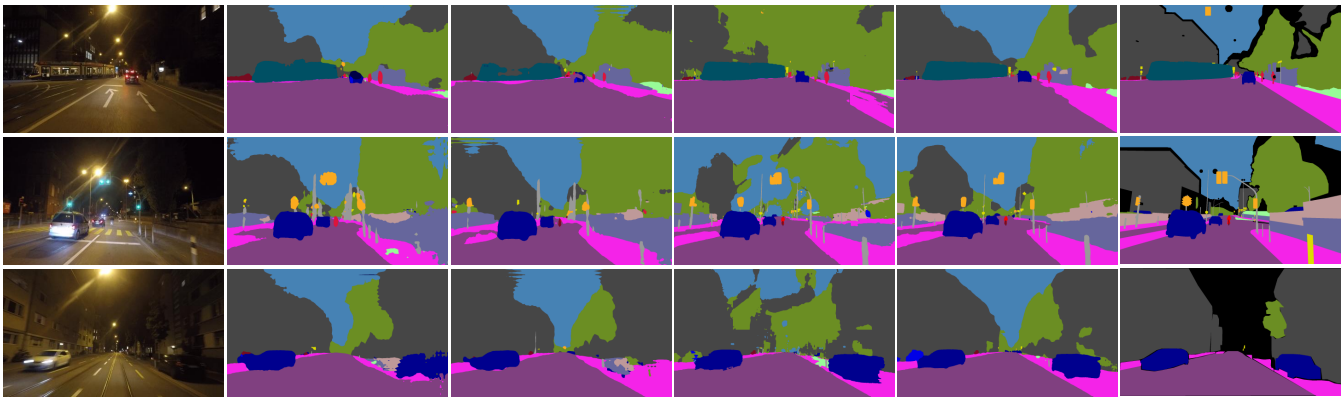

Figure 5: The qualitative comparison between our approach and existing state-of-the-art UDA-NSS methods on the ACDC-night-val set (Top row), the Dark Zurich-val set (Middle row), and the Night Driving test (Bottom row).

Table 2: Comparison of state-of-the-art unsupervised domain adaptation on Cityscapes→ACDC adaptation for nighttime. The per-category results on ACDC-night-val are shown below. The best results are presented in bold

Method	Backbone	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIoU**
DANNet [53]	DeepLab-v2	88.9	51.8	63.9	21.3	22.7	22.5	22.9	23.9	61.7	16.2	77.8	21.1	4.7	28.7	0.7	0.0	41.3	19.2	34.5	32.8
DANNet [53]	RefineNet	90.1	55.6	72.4	32.6	28.8	26.0	30.5	22.8	64.6	15.8	80.1	24.7	0.2	53.3	2.8	0.0	62.6	21.7	28.3	37.5
DANNet [53]	PSPNet	89.7	58.2	71.8	30.3	36.5	22.7	32.3	22.2	69.5	21.6	80.4	32.5	16.6	42.5	2.5	0.0	61.6	8.6	46.9	38.3
Bi-Mix [66]	DeepLab-v2	88.4	60.4	62.2	27.7	27.1	28.0	29.9	27.6	58.9	16.4	58.3	23.3	2.1	37.4	0.4	0.0	31.4	8.2	34.5	32.8
Bi-Mix [66]	RefineNet	90.1	58.7	61.4	27.0	27.4	36.1	25.0	25.3	57.6	12.1	59.6	23.5	5.4	50.3	7.3	0.0	46.2	17.2	26.6	34.6
Bi-Mix [66]	PSPNet	90.0	60.6	63.9	33.0	30.4	30.2	29.4	24.5	63.4	17.9	66.5	27.8	14.0	51.1	2.9	0.0	60.4	18.5	39.6	38.1
GPS-GLASS [22]	PSPNet	92.3	67.7	68.9	33.0	32.6	37.7	46.5	31.1	68.6	20.6	81.7	21.6	10.0	53.8	2.4	0.0	11.3	5.8	36.8	38.0
Refign [3]	DeepLab-v2	92.2	65.6	74.5	33.3	20.6	43.8	41.6	30.7	57.8	14.9	53.1	34.6	5.3	61.7	16.6	0.0	69.0	17.4	27.6	40.0
InforMS (Ours)	DeepLab-v2	93.5	67.7	72.6	37.0	22.7	40.8	33.8	32.6	63.7	10.3	65.0	44.2	16.8	65.6	21.2	0.0	79.7	23.8	44.1	44.0
DAFormer [14]	DAFormer	90.5	61.7	59.9	30.2	27.5	47.6	7.3	26.1	42.3	19.4	0.0	33.6	18.0	59.3	8.8	0.0	82.5	17.8	30.2	36.8
Refign [3]	DAFormer	87.2	55.8	76.2	39.2	20.8	52.4	54.5	41.1	68.6	22.0	78.8	38.5	18.2	60.7	13.2	0.0	67.5	12.4	41.9	44.7
InforMS (Ours)	DAFormer	92.7	69.2	78.7	41.9	32.4	49.6	51.4	33.6	71.9	26.5	82.7	33.2	**18.1**	64.3	14.2	0.0	74.2	19.2	52.1	47.7
HRDA [15]	HRDA	82.8	38.8	68.7	41.9	28.4	**52.7**	41.3	33.5	41.3	22.6	0.0	46.8	**5.3**	66.1	**18.1**	0.0	58.6	14.1	50.2	37.4
Refign [3]	HRDA	94.1	75.3	81.8	44.9	42.0	46.1	62.9	40.0	**75.7**	**30.3**	**85.3**	41.6	4.7	68.9	17.5	0.0	88.9	15.6	**56.1**	51.1
InforMS (Ours)	HRDA	**94.7**	**76.2**	**81.8**	**48.8**	**43.4**	51.2	**64.0**	**47.9**	73.5	27.9	83.0	**51.9**	4.2	**70.4**	11.9	0.0	**90.4**	**23.6**	52.8	**55.4**

warmup iterations set to 1.5k. It is trained on a batch of a 1024×1024 random crop on Cityscapes and ACDC dataset for 60k iterations. To generate mixed images, we employ our designed OICS strategy, followed by applying Color jittering and Gaussian blurring. And we further apply the rare class sampling [14] to address the long-tail distribution of the source domain and increase α to 0.999.

4.3 Comparison with State-of-the-art Methods

4.3.1 Comparison on ACDC. We present comparisons to several kinds of methods, including 1) *UDA-NSS methods*: GCMA [39], MGCDA [40], DANNet [53], DANIA [54], Bi-Mix [66], GPS-GLASS [22], and Refign [3]; 2) *backbones*: DAFormer [14] and HRDA [15]. The quantitative results of mIoU performances on the ACDC-night-test and the ACDC-night-val dataset are reported in Table 1 and Table 2, respectively. It is worth noting that our InforMS, using either DeepLab-v2, DAFormer, or HRDA, outperforms existing methods on both the ACDC-night-test and ACDC-night-val datasets. Among DeepLabv2-based methods, our proposed InforMS achieves the best performance. Besides, InforMS with DAFormer as the backbone boosts the performance of DAFormer by a substantial 24.2% and 29.6% on the ACDC-night-test and the ACDC-night-val, respectively. Applying InforMS on top of HRDA [15] results in a mIoU of 65.1% and 55.4% on the ACDC-night-test and ACDC-night-val, leading the new state-of-the-art in UDA-NSS from Cityscapes to ACDC-N. And we also observe that our approach on top of different backbones, on average, has achieved comparable effects in some hard-to-adapt classes, *e.g.* , sky, traffic light, traffic sign, and vegetation. Sample visually assessing results on ACDC-night-val in Figure 5 also verify such findings. Specifically, we calculated the average performance of our method and existing UDA-NSS methods [3, 14, 15, 53, 66] based on different backbones for each semantic class on the ACDC-night-val and the ACDC-night-test dataset. Our method demonstrated performance improvements in all classes, and notably, the performance improvement achieved by informative classes surpassed that of uninformative classes when compared with existing UDA-NSS methods. This proves that our method can realize "hard-to-adapt" classes' more sufficient learning and facilitate the UDA-NSS task.

Table 3: The per-category results on Dark Zurich-test by current state-of-the-art methods. The best results are presented in bold

Method	Backbone	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mIoU**
DMAda [9]	RefineNet	75.5	29.1	48.6	21.3	14.3	34.3	36.8	29.9	49.4	13.8	0.4	43.3	50.2	69.4	18.4	0.0	27.6	34.9	11.9	32.1
GCMA [39]	RefineNet	81.7	46.9	58.8	22.0	20.0	41.2	40.5	41.6	64.8	31.0	32.1	53.5	47.5	75.5	39.2	0.0	49.6	30.7	21.0	42.0
MGCDA [40]	RefineNet	80.3	49.3	66.2	7.8	11.0	41.4	38.9	39.0	64.1	18.0	55.8	52.1	53.5	74.7	66.0	0.0	37.5	29.1	22.7	42.5
CDAda [60]	RefineNet	90.5	60.6	67.9	37.0	19.3	42.9	36.4	35.3	66.9	24.4	79.8	45.4	42.9	70.8	51.7	0.0	29.7	27.7	26.2	45.0
DANNet [53]	DeepLab-v2	88.6	53.4	69.8	34.0	20.0	25.0	31.5	35.9	69.5	32.2	82.3	44.2	43.7	54.1	22.0	0.1	40.9	36.0	24.1	42.5
DANIA [54]	DeepLab-v2	89.4	60.6	72.3	34.5	23.7	37.3	32.8	40.0	72.1	33.0	84.1	44.7	48.9	59.0	9.8	0.1	40.1	38.4	30.5	44.8
Bi-Mix [66]	DeepLab-v2	89.9	59.7	68.5	32.6	21.9	33.5	26.4	33.5	64.1	31.3	74.4	38.4	29.2	63.2	6.8	0.1	31.8	37.9	21.3	40.2
CCDistill [12]	RefineNet	89.6	58.1	70.6	36.6	22.5	33.0	27.0	30.5	68.3	33.0	80.9	42.3	40.1	69.4	58.1	0.1	72.6	47.7	21.3	47.5
SFNET-N [49]	-	94.3	74.0	79.4	43.8	31.9	43.8	57.9	50.1	73.4	36.1	85.5	60.6	53.6	86.9	8.2	41.2	82.2	45.2	33.7	56.9
GPS-GLASS [22]	PSPNet	91.6	63.1	71.2	34.7	26.7	41.4	39.7	38.4	68.6	34.8	83.7	41.3	40.8	69.6	21.5	0.0	63.5	32.1	19.4	46.4
DIAL-Filters [27]	DeepLab-v2	88.7	55.8	69.8	34.7	17.1	31.7	26.6	34.4	69.0	25.9	80.1	45.1	43.3	67.6	10.9	1.1	66.1	37.6	20.5	43.5
LoopDA [44]	PSPNet	86.3	46.3	76.1	30.3	22.5	32.5	34.1	34.8	62.6	19.5	84.3	46.6	51.5	73.2	60.7	3.1	73.4	26.2	24.8	46.8
DeepLab-v2 [4]	DeepLab-v2	79.0	21.8	53.0	13.3	11.2	22.5	20.2	22.1	43.5	10.4	18.0	37.4	33.8	64.1	6.4	0.0	52.3	30.4	7.4	28.8
Refign [3]	DeepLab-v2	89.9	59.7	69.5	28.5	11.6	39.0	17.1	35.0	35.7	18.8	30.4	38.8	43.1	72.3	73.7	0.0	61.6	33.9	24.7	41.2
InforMS (Ours)	DeepLab-v2	94.1	69.5	77.6	40.7	18.5	47.5	41.9	40.5	68.7	33.9	86.0	57.0	49.7	78.4	63.1	**26.6**	79.4	34.4	37.1	55.0
DAFormer [14]	DAFormer	93.5	65.5	73.3	39.4	19.2	53.3	44.1	44.0	59.5	34.5	66.6	53.4	52.7	82.1	52.7	9.5	89.3	**50.5**	38.5	53.8
Refign [3]	DAFormer	91.8	65.0	80.9	37.9	25.8	56.2	45.2	51.0	78.7	31.0	88.9	58.8	52.9	77.8	51.8	6.1	90.8	40.2	37.1	56.2
InforMS (Ours)	DAFormer	92.1	66.6	76.1	29.5	25.4	**58.3**	48.5	52.9	74.6	**41.8**	87.8	57.8	53.5	82.5	76.7	1.0	87.2	44.2	40.7	57.7
HRDA [15]	HRDA	89.1	52.7	73.4	39.1	24.3	54.6	49.2	39.6	59.8	31.3	71.6	56.2	57.5	83.5	72.3	3.9	81.5	49.7	40.9	54.2
MIC [16]	HRDA	94.8	75.0	84.0	55.1	28.4	**62.0**	35.5	52.6	59.2	**46.8**	70.0	65.2	**61.7**	82.1	64.2	**18.5**	91.3	52.6	44.0	60.2
Refign [3]	HRDA	**95.4**	76.1	86.9	**53.7**	**37.1**	58.0	**57.2**	63.7	**79.4**	42.0	91.3	63.8	56.6	83.8	**85.6**	5.1	**91.7**	47.4	39.5	63.9
InforMS (Ours)	HRDA	94.0	72.8	**87.2**	52.7	34.4	60.3	54.1	**65.9**	78.2	40.6	**91.3**	**65.2**	56.2	**87.8**	82.1	14.0	91.6	**59.3**	**45.6**	**64.7**

Table 4: Comparison with state-of-the-art methods on Dark Zurich-val set and Nighttime Driving test set

Method	Backbone	Dark Zurich-val	Nighttime Driving
		mIoU	mIoU
DMAda [9]	RefineNet	-	36.1
GCMA [39]	RefineNet	26.7	45.6
MGCDA [40]	RefineNet	26.1	49.4
CDAda [60]	RefineNet	36.0	50.9
DANNet [53]	DeepLab-v2	32.1	44.1
DANIA [54]	DeepLab-v2	-	47.1
Bi-Mix [66]	DeepLab-v2	33.2	44.2
SFNET-N [49]	-	-	57.4
GPS-GLASS [22]	PSPNet	38.2	46.5
DIAL-Filters [27]	DeepLab-v2	-	44.6
LoopDA [44]	PSPNet	37.6	49.6
DeepLab-v2 [4]	DeepLab-v2	-	25.4
InforMS (Ours)	DeepLab-v2	40.0	53.7
DAFormer [14]	DAFormer	37.1	54.0
InforMS (Ours)	DAFormer	45.1	56.0
HRDA [15]	HRDA	42.1	54.1
InforMS (Ours)	HRDA	**52.5**	**58.5**

4.3.2 Comparison on Dark Zurich. To verify the effectiveness of our method, we further conduct comparative experiments on the Dark Zurich-test and the Dark Zurich-val, the results are shown in Table 3 and Table 4, respectively. Similarly, we compare our proposed method with several kinds of methods, including 1) *UDA-NSS methods*: DMAda [9], GCMA [39], MGCDA [40], CDAda [60], DANNet [53], DANIA [54], Bi-Mix [66], CCDistill [12], SFNET-N [49], GPS-GLASS [22], DIAL-Filters [27], LoopDA [44], and MIC [16]; 2) *backbones*: DeepLab-v2 [4], DAFormer [14] and HRDA [15]. Notably, our method with HRDA as the backbone achieves the best performance on the Dark Zurich-test and Dark Zurich-val datasets, with a remarkable 1.3% and 24.7% improvement in mIoU compared to the leading method, respectively. The visualization comparison of the Dark Zurich-val is shown in Figure 5.

4.3.3 Comparison on Nighttime Driving. To show the generalization ability of our method, we also evaluate our approach that is adapted to Dark Zurich on the Nighttime Driving test set in Table 4, with sample visualization results presented in Figure 5. Our method with HRDA as the backbone still achieves the best performance on this dataset, with a 0.9% improvement of the overall mIoU over the leading method. This result verifies the ability of our method to generalize well to the Nighttime Driving dataset.

4.4 Ablation Study

In this section, we conduct a series of ablation studies to validate the contributions of each component to the final performance.

4.4.1 Effectiveness of Online Informative Class Sampling. We first investigate the impact of our proposed Online Informative Class Sampling strategy. We train several model variants and evaluate their performance on the ACDC-night-test and the ACDC-night-val datasets, as reported in Table 5. We choose Refign [3] as the baseline, and the reproduced results are reported as the "Initialization" of Table 5. Then, we separately utilize "Spectrogram Mean" or "Class Frequency" as our informativeness metric to select informative classes. Both "Spectrogram Mean" and "Class Frequency" bring performance improvements, "Spectrogram Mean" showing greater improvement. After that, our Informative Mixture of Experts boosts the performance to 65.1% mIoU and 55.4% mIoU for the ACDC-night-test and the ACDC-night-val, respectively. This demonstrates the necessity and effectiveness of dynamic weight adjustment.

Table 5: Ablation Study on several model variants of our method on ACDC-night-test and ACDC-night-val

HRDA				
	Component		ACDC-night-test	ACDC-night-val
Initialization	Baseline		63.5 (+0.0)	51.1 (+0.0)
Informative Class Sampling Metric	Spectrogram Mean	Class Frequency	ACDC-night-test	ACDC-night-val
	✓		65.1 (+1.6)	54.8 (+3.7)
		✓	64.4 (+0.9)	54.6 (+3.5)
	✓	✓	65.1 (+1.6)	55.4 (+4.3)
DeepLab-v2				
Initialization	Baseline		49.8 (+0.0)	40.0 (+0.0)
Informative Class Sampling Metric	Spectrogram Mean	Class Frequency	ACDC-night-test	ACDC-night-val
	✓		51.1 (+1.3)	42.9 (+2.9)
		✓	50.4 (+0.6)	42.4 (+2.4)
	✓	✓	52.4 (+2.6)	44.0 (+4.0)

Table 6: Comparison with different data sampling methods in InforMS (HRDA) on the ACDC-night-val set. (RS: Random Sampling, CBS: Class-balanced Sampling, RCS: Rare Class Sampling, OICS: Online Informative Class Sampling)

Method	RS	CBS	RCS	OICS (Ours)
ACDC-night-val	53.1	50.7	51.1	**55.4**

Table 7: Comparison with different proportion choice of the Ranked Source Class List on the ACDC-night-val

Proportion	Backbone	1/4	1/3	1/2	2/3	3/4
InforMS (Ours)	DeepLab-v2	41.7	42.1	**44.0**	42.5	41.9
InforMS (Ours)	HRDA	54.9	51.9	**55.4**	55.0	52.5

Furthermore, we compare our method with existing re-sampling methods for UDA semantic segmentation in Table 6. We can observe that our proposed OICS strategy achieves the best performance on the ACDC-night-val. Meanwhile, the comparison reveals that CBS and RCS strategies are unsuitable for UDA-NSS compared to RS. This observation is because these methods focus on addressing the long-tail distribution of the source domain while overlooking informative class learning during domain adaptation.

4.4.2 Backbone. We also explore the effectiveness of different backbones in Table 5. HRDA can bring more obvious performance improvement compared with DeepLab-v2 for its strong robustness. Despite that, our method (DeepLab-v2) still achieves comparable performance compared with existing UDA-NSS methods based on DeepLab-v2 on the ACDC-night-test and ACDC-night-val dataset.

4.4.3 The Proportion Choice. We conduct an ablation study about the proportion choice of the ranked source class List on the ACDC-night-val, as shown in Table 7. We find that 1/2 is the best choice.

4.4.4 The Weight Adjustment of Experts. We conduct an ablation study about the weight adjustment of experts on the ACDC-night-val, as shown in Table 8. We find that 0.1 is the best choice.

4.4.5 The Class Selection in InforMS. We explore the class selection of the target nighttime in InforMS on the ACDC-night-test and the ACDC-night-val, as shown in Table 9. In the proposed InforMS, our proposed OICS strategy selects informative classes from the target nighttime domain to carry out the cross-domain mixed sampling. For comparison, we select classes randomly ("Random Selection") or less informative classes ("Less Informative") from the target nighttime domain to carry out the task. We find that performances with Informative (Ours) are better than those with "Random Selection" or "Less Informative" by large margins.

Table 8: Comparison with different weight adjustment of experts on the ACDC-night-val

Weight Adjustment	Backbone	0.0	0.05	0.1	0.15	0.2	0.5	1.0
InforMS (Ours)	DeepLab-v2	42.3	41.8	**44.0**	42.1	42.3	41.8	41.1

Table 9: Comparison with different selection methods in InforMS on the ACDC-night-test and the ACDC-night-val

Method	Backbone	ACDC-night-test	ACDC-night-val
Random Selection	DeepLab-v2	49.8	40.0
Less Informative	DeepLab-v2	51.4	42.7
Informative (Ours)	DeepLab-v2	**52.4**	**44.0**
Random Selection	HRDA	63.5	51.1
Less Informative	HRDA	64.6	54.6
Informative (Ours)	HRDA	**65.1**	**55.4**

Table 10: Ablation Study Results of target day-to-night mixture on the ACDC-night-test and the ACDC-night-val

Method	Backbone	ACDC-night-test	ACDC-night-val
InforMS (Ours)	DeepLab-v2	**52.4**	**44.0**
w day-to-night DA	DeepLab-v2	51.4	42.8
InforMS (Ours)	HRDA	**65.1**	**55.4**
w day-to-night DA	HRDA	63.5	51.9

4.4.6 InforMS with or without target day-to-night mixture. We investigate the effect of the mixture between target daytime and nighttime by utilizing pseudo labels of daytime and nighttime during domain adaptation. The results are reported on the ACDC-night-test and ACDC-night-val datasets, as presented in Table 10. Our method significantly outperforms our method with day-to-night DA. The reason for this result is that the day-to-night mixture accumulates the noise of the target daytime domain's pseudo-labels, hindering the knowledge transferring of the domain adaptation.

5 CONCLUSION

In this paper, we have reconsidered the UDA-NSS task from cross-domain mixed sampling exploiting informative classes mining perspective. We find that the informativeness of classes is closely related to the spectrogram mean and the class frequency of each class in the target nighttime domain predictions and corresponding input images. To boost the UDA-NSS task, we presented an Online Informative Class Sampling strategy to adaptively mine informative classes based on the corresponding spectrogram mean and class frequency. Furthermore, we designed an Informativeness-based cross-domain Mixed Sampling (InforMS) framework, prioritizing informative class learning and boosting the UDA-NSS task. Experimental results on three wide-used benchmarks demonstrated the superiority of our method over existing UDA-NSS approaches.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China (61602349, U1803262, and 61440016), and the Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System.

REFERENCES

[1] Nikita Araslanov and Stefan Roth. 2021. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, TN, USA, 15384–15394.

[2] Devansh Arpit, Stanisaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. 2017. A closer look at memorization in deep networks. In *International Conference on Machine Learning*. JMLR.org, Sydney, NSW, Australia, 233–242.

[3] David Bruggemann, Christos Sakaridis, Prune Truong, and Luc Van Gool. 2023. Refign: Align and Refine for Adaptation of Semantic Segmentation to Adverse Conditions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, Waikoloa, HI, USA, 1–17.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4 (2017), 834–848.

[5] Weilong Chen, Meng Joo Er, and Shiqian Wu. 2006. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Transactions on Systems, Man, and Cybernetics* 36, 2 (2006), 458–466.

[6] Yiting Cheng, Fangyun Wei, Jianmin Bao, Dong Chen, and Wenqiang Zhang. 2023. ADPL: Adaptive Dual Path Learning for Domain Adaptation of Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2023), 9339–9356. Issue 8.

[7] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. *Noise reduction in speech processing* (2009), 1–4.

[8] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, NV, USA, 3213–3223.

[9] Dengxin Dai and Luc Van Gool. 2018. Dark model adaptation: Semantic image segmentation from daytime to nighttime. In *2018 21st International Conference on Intelligent Transportation Systems*. IEEE, Maui, HI, USA, 3819–3824.

[10] Xueqing Deng, Peng Wang, Xiaochen Lian, and Shawn Newsam. 2022. NightLab: A Dual-level Architecture with Hardness Detection for Segmentation at Night. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New Orleans, LA, USA, 16938–16948.

[11] Guilherme N DeSouza and Avinash C Kak. 2002. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 2 (2002), 237–267.

[12] Huan Gao, Jichang Guo, Guoli Wang, and Qian Zhang. 2022. Cross-Domain Correlation Distillation for Unsupervised Domain Adaptation in Nighttime Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New Orleans, LA, USA, 9913–9923.

[13] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. 2017. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677* (2017).

[14] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. 2022. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New Orleans, LA, USA, 9924–9935.

[15] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. 2022. HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation. In *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 372–391.

[16] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. 2023. MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Vancouver, Canada, 11721–11732.

[17] Somi Jeong, Youngjung Kim, Eungbean Lee, and Kwanghoon Sohn. 2021. Memory-guided Unsupervised Image-to-image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, TN, USA, 6558–6567.

[18] Kui Jiang, Zhongyuan Wang, Zheng Wang, Chen Chen, Peng Yi, Tao Lu, and Chia-Wen Lin. 2022. Degrade is upgrade: Learning degradation for low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, Virtual Event, 1078–1086.

[19] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zhen Han, Tao Lu, Baojin Huang, and Junjun Jiang. 2020. Decomposition makes better rain removal: An improved attention-guided deraining network. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2020), 3981–3995.

[20] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zheng Wang, Xiao Wang, Junjun Jiang, and Chia-Wen Lin. 2021. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing* 30 (2021), 7404–7418.

[21] Liming Jiang, Changxu Zhang, Mingyang Huang, Chunxiao Liu, Jianping Shi, and Chen Change Loy. 2020. Tsit: A simple and versatile framework for image-to-image translation. In *European Conference on Computer Vision*. Springer, Glasgow, UK, 206–222.

[22] Hongjae Lee, Changwoo Han, and Seung-Won Jung. 2022. GPS-GLASS: Learning Nighttime Semantic Segmentation Using Daytime Video and GPS data. *arXiv preprint arXiv:2207.13297* (2022).

[23] Attila Lengyel, Sourav Garg, Michael Milford, and Jan C van Gemert. 2021. Zero-shot day-night domain adaptation with a physics prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, QC, Canada, 4399–4409.

[24] Liang Liao, Wenyi Chen, Jing Xiao, Zheng Wang, Chia-Wen Lin, and Shin'ichi Satoh. 2022. Unsupervised foggy scene understanding via self spatial-temporal label diffusion. *IEEE Transactions on Image Processing* 31 (2022), 3525–3540.

[25] Liang Liao, Wenyi Chen, Zhen Zhang, Jing Xiao, Yan Yang, Chia-Wen Lin, and Shin'ichi Satoh. 2023. Only a few classes confusing: Pixel-wise candidate labels disambiguation for foggy scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, Washington, DC, USA, 1558–1567.

[26] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates Inc., Long Beach, CA, USA, 700–708.

[27] Wenyu Liu, Wentong Li, Jianke Zhu, Miaomiao Cui, Xuansong Xie, and Lei Zhang. 2022. Improving Nighttime Driving-Scene Segmentation via Dual Image-adaptive Learnable Filters. *arXiv preprint arXiv:2207.01331* (2022).

[28] Yahao Liu, Jinhong Deng, Xinchen Gao, Wen Li, and Lixin Duan. 2021. BAPA-Net: Boundary Adaptation and Prototype Alignment for Cross-domain Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, QC, Canada, 8801–8811.

[29] Yuting Liu, Zheng Wang, Miaojing Shi, Shin'ichi Satoh, Qijun Zhao, and Hongyu Yang. 2020. Towards unsupervised crowd counting via regression-detection bi-knowledge transfer. In *Proceedings of the 28th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 129–137.

[30] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, QC, Canada, 10012–10022.

[31] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*. OpenReview.net, New Orleans, LA, USA, 1925–1934.

[32] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. 2020. Adversarial style mining for one-shot unsupervised domain adaptation. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates Inc., Virtual, 20612–20623.

[33] Yawei Luo, Ping Liu, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. 2021. Category-level adversarial adaptation for semantic segmentation using purified features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 8 (2021), 3940–3956.

[34] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. 2019. Taking a Closer Look at Domain Shift: Category-Level Adversaries for Semantics Consistent Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Long Beach, CA, USA, 2507–2516.

[35] Xianzheng Ma, Zhixiang Wang, Yacheng Zhan, Yinqiang Zheng, Zheng Wang, Dengxin Dai, and Chia-Wen Lin. 2022. Both style and fog matter: Cumulative domain adaptation for semantic foggy scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New Orleans, LA, USA, 18922–18931.

[36] Junhan Peng, Jia Su, Yongqing Sun, Zheng Wang, and Chia-wen Lin. 2021. Semantic nighttime image segmentation via illumination and position aware domain adaptation. In *2021 IEEE International Conference on Image Processing*. IEEE, Anchorage, AK, USA, 1034–1038.

[37] Jian-Jun Qiao, Zhi-Qi Cheng, Xiao Wu, Wei Li, and Ji Zhang. 2022. Real-time Semantic Segmentation with Parallel Multiple Views Feature Augmentation. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 6300–6308.

[38] Eduardo Romera, Luis M Bergasa, Kailun Yang, Jose M Alvarez, and Rafael Barea. 2019. Bridging the day and night domain gap for semantic segmentation. In *2019 IEEE Intelligent Vehicles Symposium*. IEEE, Paris, France, 1312–1318.

[39] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2019. Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Seoul, Korea (South), 7374–7383.

[40] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2020. Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 6 (2020), 3139–3153.

[41] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2021. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, QC, Canada, 10765–10775.

[42] Nag Sauradip, Adak Saptakatha, and Das Sukhendu. 2019. What's There in the Dark. In *2019 IEEE International Conference on Image Processing*. IEEE, Taipei, Taiwan, 2996–3000.

[43] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

[44] Fengyi Shen, Zador Pataki, Akhil Gurram, Ziyuan Liu, He Wang, and Alois Knoll. 2023. LoopDA: Constructing Self-loops to Adapt Nighttime Semantic Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, Waikoloa, HI, USA, 3256–3266.

[45] Lei Sun, Kaiwei Wang, Kailun Yang, and Xiang. Kaite. 2019. See Clearer at Night: Towards Robust Nighttime Semantic Segmentation through Day-Night Image Conversion. *Artificial Intelligence and Machine Learning in Defense Applications* 11169 (2019), 77–89.

[46] Xin Tan, Ke Xu, Ying Cao, Yiheng Zhang, Lizhuang Ma, and Rynson WH Lau. 2021. Night-Time Scene Parsing With a Large Real Dataset. *IEEE Transactions on Image Processing* 30 (2021), 9085–9098.

[47] Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*. Curran Associates Inc., Long Beach, CA, USA, 1195–9935.

[48] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. 2021. DACS: Domain Adaptation via Cross-domain Mixed Sampling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, Waikoloa, HI, USA, 1379–1389.

[49] Hai Wang, Yanyan Chen, Yingfeng Cai, Long Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. 2022. SFNet-N: An Improved SFNet Algorithm for Semantic Segmentation of Low-Light Autonomous Driving Road Scenes. *IEEE Transactions on Intelligent Transportation Systems* 23, 11 (2022), 1–13.

[50] Xiao Wang, Jun Chen, Zheng Wang, Wu Liu, Shin'ichi Satoh, Chao Liang, and Chia-Wen Lin. 2021. When pedestrian detection meets nighttime surveillance: a new benchmark. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. IJCAI Proceedings, Yokohama, Yokohama, Japan, 509–515.

[51] Xiao Wang, Zheng Wang, Wu Liu, Xin Xu, Jing Chen, and Chia-Wen Lin. 2021. Consistency-constancy bi-knowledge learning for pedestrian detection in night surveillance. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 4463–4471.

[52] Yuxi Wang, Junran Peng, and ZhaoXiang Zhang. 2021. Uncertainty-aware Pseudo Label Refinery for Domain Adaptive Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, QC, Canada, 9092–9101.

[53] Xinyi Wu, Zhenyao Wu, Hao Guo, Lili Ju, and Song Wang. 2021. DANNet: A One-Stage Domain Adaptation Network for Unsupervised Nighttime Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Nashville, TN, USA, 15769–15778.

[54] Xinyi Wu, Zhenyao Wu, Lili Ju, and Song Wang. 2021. A One-Stage Domain Adaptation Network with Image Alignment for Unsupervised Nighttime Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2021), 58–72. Issue 1.

[55] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. 2021. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Advances in Neural Information Processing Systems*. Curran Associates Inc., virtual, 12077–12090.

[56] Pengyu Xie, Xin Xu, Zheng Wang, and Toshihiko Yamasaki. 2021. Unsupervised video person re-identification via noise and hard frame aware clustering. In *2021 IEEE International Conference on Multimedia and Expo*. IEEE, Shenzhen, China, 1–6.

[57] Pengyu Xie, Xin Xu, Zheng Wang, and Toshihiko Yamasaki. 2022. Sampling and re-weighting: Towards diverse frame aware unsupervised video person re-identification. *IEEE Transactions on Multimedia* 24 (2022), 4250–4261.

[58] Zhifeng Xie, Sen Wang, Ke Xu, Zhizhong Zhang, Xin Tan, Yuan Xie, and Lizhuang Ma. 2023. Boosting Night-time Scene Parsing with Learnable Frequency. *IEEE Transactions on Image Processing* 32 (2023), 2386–2398.

[59] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. 2020. Learning in the Frequency Domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Seattle, WA, USA, 1740–1749.

[60] Qi Xu, Yinan Ma, Jing Wu, Chengnian Long, and Xiaolin Huang. 2021. CDAda: A Curriculum Domain Adaptation for Nighttime Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Montreal, BC, Canada, 2962–2971.

[61] Xin Xu, Lei Liu, Xiaolong Zhang, Weili Guan, and Ruimin Hu. 2021. Rethinking data collection for person re-identification: active redundancy reduction. *Pattern Recognition* 113 (2021), 107827.

[62] Xin Xu, Wei Liu, Zheng Wang, Ruimin Hu, and Qi Tian. 2022. Towards generalizable person re-identification with a bi-stream generative model. *Pattern Recognition* 132 (2022), 108954.

[63] Xin Xu, Shiqin Wang, Zheng Wang, Xiaolong Zhang, and Ruimin Hu. 2021. Exploring image enhancement for salient object detection in low light images. *ACM transactions on multimedia computing, communications, and applications* 17, 1s (2021), 1–19.

[64] Xin Xu, Xin Yuan, Zheng Wang, Kai Zhang, and Ruimin Hu. 2022. Rank-in-rank loss for person re-identification. *ACM Transactions on Multimedia Computing, Communications, and Applications* 18, 2s (2022), 1–21.

[65] Danna Xue, Fei Yang, Pei Wang, Luis Herranz, Jinqiu Sun, Yu Zhu, and Yanning Zhang. 2022. SlimSeg: Slimmable Semantic Segmentation with Boundary Supervision. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 6539–6548.

[66] Guanglei Yang, Zhun Zhong, Hao Tang, Mingli Ding, Nicu Sebe, and Elisa Ricci. 2021. Bi-Mix: Bidirectional Mixing for Domain Adaptive Nighttime Semantic Segmentation. *arXiv preprint arXiv:2111.10339* (2021).

[67] Yanchao Yang and Stefano Soatto. 2020. FDA: Fourier Domain Adaptation for Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Seattle, WA, USA, 4085–4095.

[68] Zelong Zeng, Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. 2020. Illumination-adaptive person re-identification. *IEEE Transactions on Multimedia* 22, 12 (2020), 3064–3074.

[69] Xian Zhong, Shidong Tu, Xianzheng Ma, Kui Jiang, Wenxin Huang, and Zheng Wang. 2022. Rainy WCity: A real rainfall dataset with diverse conditions for semantic driving scene understanding. In *International Joint Conference on Artificial Intelligence*. IJCAI Proceedings, Vienna, Austria, 1743–1749.

[70] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, Venice, Italy, 2223–2232.

[71] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 289–305.

[72] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. 2019. Confidence Regularized Self-Training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Seoul, Korea (South), 5982–5991.