The Parables of the Mustard Seed and the Yeast: Extremely Low-Budget, High-Performance Nighttime Semantic Segmentation

Shiqin Wang[1,2], **Xin Xu**[1,2*], **Haoyang Chen**[3], **Kui Jiang**[4], **Zheng Wang**[3]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065 China
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System
[3]National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, China
[4]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001 China
wangshiqin@wust.edu.cn, xuxin@wust.edu.cn, haoyangchen@whu.edu.cn, jiangkui@hit.edu.cn, wangzwhu@whu.edu.cn

Abstract

Nighttime Semantic Segmentation (NSS) is essential to many cutting-edge vision applications. However, existing technologies overly rely on massive labeled data, whose annotation is time-consuming and laborious. In this paper, we pioneer a new task focusing on exploring the potential of training strategy and framework design with limited annotation to achieve high-performance NSS. Insufficient information at very low labeling budgets can easily lead to under-optimization or overfitting of the model. Our solution comprises two main components: i) a novel region-based active sampling strategy called Contextual-Aware Region Query (CARQ), which identifies highly informative target nighttime regions for labeling; and ii) an innovative Fragmentation Synergy Active Domain Adaptation framework (FS-ADA), which progressively broadcasts the limited annotation to the unlabeled regions, achieving high performance with a minimal annotation budget. Extensive experiments demonstrate that our method outperforms state-of-the-art UDA-NSS & ADA-SS methods across four day-to-nighttime benchmarks, and generalizes well to foggy, rainy, & snowy scenes. In particular only with 1% target nighttime data annotation, our method is on par with the mainstream fully-supervised methods on the BDD100K-Night val dataset.

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 (Wu et al. 2021b; Wang et al. 2024), visual surveillance (Wang et al. 2023, 2020), and robotic vision (DeSouza and Kak 2002; Xu et al. 2021). Recent fully supervised NSS methods have achieved great success by leveraging abundant nighttime labeled data, yet acquiring full-pixel annotations is labor-intensive. To avoid annotation efforts, some researchers multiplex daytime-trained models at night, yet encounter severe performance degradation due to the huge domain gap (Wu et al. 2021b). Subsequently, Unsupervised Domain Adaptation Nighttime Semantic Segmentation (UDA-NSS) attracts much attention to adapt models trained on well-labeled source daytime domain to the unlabeled nighttime domain. The performance of UDA-NSS can outperform pure daytime-trained models ("Source Only"), yet still lags behind fully supervised counterparts (Figure 1 (I)).

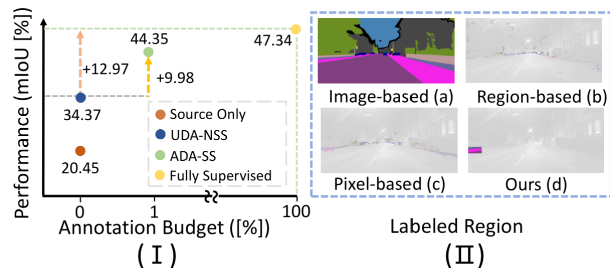

Figure 1: I) Compared with UDA-NSS, the Fully Supervised method requires 100% of the budget to increase the mIoU by 12.97%. while ADA only requires 1% of the budget to increase the mIoU by 9.98%. II) (a) treats an entire image as the individual query sample; (b) and (c) choose either 3×3 or 1×1 pixel regions as the query granularity. We choose the image patch as the query granularity (d), which makes a trade-off between sample diversity (compared to (a)) and reduces labeling costs (compared to (b) and (c)).

The main reason contributing to the large performance gap between UDA-NSS and fully supervised counterparts is the scarce annotated nighttime data. Since collecting generous paired annotations is cumbersome and costly, leveraging the Active Domain Adaptation (ADA) technique, *i.e.,* labels few target nighttime domain data for model training, provides an alternative and reasonable solution. Consequently, we pioneer the Active Domain Adaptive Nighttime Semantic Segmentation (ADA-NSS) task, which queries the most informative target nighttime samples for labeling and performs effective learning to maximize NSS performance.

In the previous studies of ADA, bare efforts have been made on Active Domain Adaptive Semantic Segmentation (ADA-SS). In ADA-SS, the design of the label query, including acquisition functions and query granularity, is vital to maximizing supervision within the given budget. In acquisition functions, current methods mainly utilize prediction uncertainty (Franco et al. 2024) or feature space diversity (Ning et al. 2023) to select valuable data. These strategies are less effective under the large domain gap, as in the ADA-NSS task (Xie et al. 2023a; Han et al. 2023; Wu

*Corresponding author.

et al. 2022). The primary reasons are: (i) uncertainty estimation on the target domain is usually miscalibrated under the large domain gap, leading to sampling outliers or redundant instances; and (ii) diversity selection may select samples already well-aligned with the source domain, resulting in redundant annotations. Regarding query granularity, early approaches treat an entire image (Ning et al. 2023) (Figure 1 (II) (a)) as the individual query sample. Recent methods chose a non-overlapped region (3×3) (Xie et al. 2022; Wu et al. 2022) (Figure 1 (II) (b)) or individual pixel (1×1) (Franco et al. 2024) (Figure 1 (II) (c)) as the individual query sample. These labeling units are sub-optimal as the former lacks sample diversity, and the latter incurs high labeling costs comparable to labeling the entire image. After annotating the queried target data, existing ADA-SS methods fine-tune the model either in a supervised manner using labeled target data (Xie et al. 2022; Wu et al. 2022; Franco et al. 2024), or in a semi-supervised learning manner with both labeled and unlabeled target data (Ning et al. 2023). However, optimizing the model blindly with limited labeled data inescapably suffers from model over-fitting or divergence of learning ability. These limitations render existing ADA-SS methods inapplicable to the ADA-NSS task.

To better promote the ADA-NSS task, two primary and urgent issues are as follows: *i) how to query the most valuable target nighttime data for annotation; ii) how to perform more effective training with the limited queried data?*

For the first issue, we argue that data maximizing model adaptability[1] holds the highest annotation value. Firstly, the significant domain discrepancy between the source and target domains hinders domain adaptation and affects the model's adaptability. Secondly, class imbalance in training samples induces confirmation bias toward common classes (Hoyer, Dai, and Van Gool 2022a), exacerbating the problem. Consequently, we would like to consider domain degradation and transferability of different semantic classes. Under adverse weather conditions, weather-specific distraction factors, such as nighttime illumination, induce domain degradation. We devised a Distraction Perception data selection strategy that analyses the severity of weather-specific distractions within regions while using a saliency metric to assess their importance comprehensively. Meanwhile, we propose a Semantic Complexity data selection strategy from the semantic classes perspective to identify more valuable regions by comprehensively evaluating the types, quantities, and proportions of semantic classes in each region's predictions. To leverage the complementary information of the above two value metrics, we introduce the Value Mixture of Experts to dynamically assign the importance of each strategy ("valuable metric expert"), thus constructing a plug-and-play data selection strategy, *Contextual-Aware Region Query (CARQ)*. In response to the second sub-task, augmenting limited labeled target data or improving its optimization capability towards unlabeled data during fine-tuning can improve the model's adaptability. To this end, we designed a Fragmentation Synergy Active Domain Adaptation framework (FS-ADA), achieving high performance with low annotation cost. Concretely, we first perform restoration and segmentation using spatially aligned target day-night fragments to facilitate domain adaptation as a warmup step. CARQ identifies the most valuable region, which is then manually annotated. Afterward, to maximize the information from limited annotated data, we design the *Focal Information Global Broadcast (FIGB)*, including the Label Propagation Module (LPM) and Label Generalization Module (LGM). LPM progressively propagates knowledge from labeled regions to neighboring regions, continuously incorporating confident data into the labeled dataset. LGM further utilizes labeled data to guide the learning of unlabeled data.

[1]Model adaptability refers to the model's ability to adapt in the target domain, demonstrated by its performance.

Our main contributions are summarized as follows:

- We present a novel plug-and-play region-based active sampling algorithm, CARQ, identifying the most informative regions. Extensive experiments demonstrate that our algorithm achieves a 0.2–7.4 mIoU [%] improvement over existing ADA-SS methods at night.
- We pioneer a new ADA-NSS task, aiming to facilitate further the development of the NSS field. Taking it further, we introduce FS-ADA, which achieves performance comparable to the fully supervised model using only 1% of the target nighttime annotation budget on the BDD100K-Night val dataset.
- Experiments on four widely used NSS benchmarks show that our method outperforms state-of-the-art ADA-SS methods, particularly achieving a 24.4% performance gain on BDD100K-Night val based on DeepLab-v2. Notably, our method also exhibits superior generalization in adverse conditions like fog, rain, and snow.

Related Work

Nighttime Semantic Segmentation

Existing Nighttime Semantic Segmentation (NSS) methods are categorized into fully-supervised and unsupervised domain adaptation. For the former, it directly trains a model on the labeled nighttime data. Early, Tan *et al.* proposed the first large-scale NSS dataset (Tan et al. 2021). Then, researchers focused on architecture optimization via implicitly learning the entangled representations (Xie et al. 2023b; Liu et al. 2023, 2024b) or explicitly estimating the effect of lighting on semantics (Wei et al. 2023). Fully-supervised methods, though successful, heavily rely on extensive pixel-wise annotated nighttime data, which is labor-intensive and costly.

The latter focuses on adapting semantic segmentation models from the labeled daytime source domain to the unlabeled nighttime target domain. Early UDA-NSS methods primarily narrow the illumination gap between source and target domains via utilizing an intermediate twilight domain as the bridge (Sakaridis, Dai, and Van Gool 2022) or image transferring networks to unify domain styles (Wu et al. 2021a,b; Yang et al. 2021). However, these methods overlook inherent dataset gaps caused by camera equipment and urban appearance, *i.e.,* treating daytime images from distinct datasets as identical, which hinders domain adaptation. Some works regarded cross-domain correlation as the concrete representation of domain shift for adaptation (Ding, Li, and Tian 2023; Dong, Kang, and Ming 2023). However, constructing source and target domain pairs with identical differences is hard, yielding limited performance gains. Recent studies performed cross-domain mixing to implicitly address dataset and illumination gaps (Huang, Yao, and Zhou 2023; Wang et al. 2023; Xie et al. 2024). However, these methods still struggle with accurate semantic segmentation in low-visibility and complex nighttime illuminations.

Active Domain Adaptation Semantic Segmentation

Existing Active Domain Adaptation Semantic Segmentation (ADA-SS) methods are classified into 1) image-based, 2) pixel-based, and 3) region-based methods based on the sampling unit. Image-based methods (Ning et al. 2023) actively selected a subset of images and annotated the entire image. Under the limited labeling budget, the diversity of samples is restricted, resulting in subpar final effectiveness. Pixel-based methods (Zhang and Zhang 2022; Franco et al. 2024) selected individual pixels for labeling, which is less budget-efficient. To gain cost efficiency, recent ADA-SS methods (Xie et al. 2022; Wu et al. 2022; Liu et al. 2024a) treated non-overlapped local image regions as individual samples. Xie *et al.* selected image regions that are diverse in spatial adjacency and uncertain in predictions (Xie et al. 2022). Wu *et al.* considered both uncertainty and class balance when querying labels (Wu et al. 2022). These methods are designed to reduce the domain gap between synthetic and real urban street scenes, losing their effectiveness at night for ignoring the nighttime illumination distraction.

Methodology

As a premise, a fully labeled source daytime domain $D_S = \left\{(x_s^i, y_s^i)\right\}_{i=1}^{N_s}$ and an unlabeled target nighttime domain $D_T = \left\{(x_t^j)\right\}_{j=1}^{N_t}$ jointly participate in the ADA-NSS task. Here N_s and N_t denote the number of images in the source and target domains, respectively. Afterward, our FS-ADA framework can be demonstrated sequentially: 1) Train an initial UDA method with $D_S \cup D_T$ as a warmup step; 2) Query a few target nighttime data $D_{TL} = \left\{((x_t^j)^{(n)}, (y_t^j)^{(n)})\right\}_{j=1}^{N_t}$ via CARQ strategy to be labeled, and the remaining unlabeled is denoted as $D_{TU} = \left\{(x_t^j)^{(m)}\right\}_{j=1}^{N_t}$, where n and m is the labeled and unlabeled regions, respectively; 3) Augment and propagate queried labels via our FIGB to perform more effective training. Our FS-ADA framework is illustrated in Figure 2.

Warmup step. The step performs image restoration and segmentation. Using provided paired target images (or only target night images), a fully (or un) supervised restoration network I_n is fine-tuned to fit new data by constraining semantic predictions and pixel consistency of aligned target fragments (or replacing training data with D_T). The segmentation network g_θ performs the collaborative learning of aligned target fragments and complementary classes (or adaptive enhanced domain adaptive classes).

Contextual-Aware Region Query

We present the CARQ strategy considering domain difference and semantic class attributes via the following *Distraction Perception* and *Semantic Complexity*.

Distraction Perception. The weather-specific distraction severity δ in D_T is used to indirectly measure the huge degradation from D_S to D_T. Inspired by (Xu et al. 2023), low-level Frequency Components (FCs), extracted from the image's frequency spectrum through a band-pass filter with indices [0,2)∪[32,64], can represent the domain-variant contents (*i.e,* distraction). $\mathbf{x}_t^j$ is decomposed into a set of n non-overlapping image patches of the same size $S^i_{\mathbf{x}_t^j} = \left\{(\mathbf{x}_t^j)^i\right\}_{i=1}^n$. The mean value of low-level FCs for the i-th image patch in $\mathbf{x}_t^j$ is calculated to assess the δ as follows:

$$\delta(S^i) = \frac{1}{N_t}\sum_{x=1}^{h}\sum_{y=1}^{w} iDCT(Cat(\alpha(u)\alpha(v)\sum_{x=1}^{h}\sum_{y=1}^{w} S^i_{\mathbf{x}_t^j} \cos\left[\frac{\pi(2x+1)u}{2h}\right]\cos\left[\frac{\pi(2y+1)v}{2w}\right])) \tag{1}$$

where N_t, w, and h denote the total pixel number, width, and height of the i-th patch, respectively. $iDCT$ is the inverse 2D-DCT function, transforming low-level FCs into the original image space. Cat denotes the concatenation operation. u and v are the horizontal and vertical FCs, respectively. $\alpha(u)$ and $\alpha(v)$ are normalized coefficients.

To identify the most valuable patches, we search regions where distraction severity (*e.g.,* night illumination) significantly affects model predictions. Specifically, its influence f on model predictions for i-th image patch in the restored target nighttime image $\mathbf{x}_{te}^j$ is calculated as:

$$f(S^i) = \frac{1}{N_t}\sum_{c=1}^{C}\mathrm{ReLU}\left(\sum_{k=1}^{K}\left(\frac{1}{N_t}\sum_{x=1}^{h_k}\sum_{y=1}^{w_k}\frac{\partial \underset{c\in\{1,\ldots,C\}}{\arg\max}\, g_\theta\left(S^i_{\mathbf{x}_{te}^j}\right)^c}{\partial A^k_{xy}}\right)A^k\right) \tag{2}$$

where N_t and $\mathcal{C}$ are the total number of pixel, and semantic classes of predictions in the i-th image patch, respectively. ReLU highlights areas that positively contribute to the decision for class c. $\{A^k\}_{k=1}^K$ are selected feature maps from the shallow layer[2]. K is the number of kernels. x, y is pixel location. h_k and w_k is the hight and width of each feature map A^k. g_θ is our segmentation network.

The most disturbed regions in D_T contain extremely limited information, and learning them well may hinder the model's generalization to all D_T scenes. Meanwhile, regions more significant to model predictions offer greater potential for overall performance improvement. To comprehensively evaluate each image patch's value in D_T, we propose a new Distraction Perception $\mathcal{D}(\cdot)$ strategy, calculated as follows:

$$\mathcal{D}(\cdot) = \frac{e^{f(S^i)} - e^{-f(S^i)}}{2ln\left(1 + e^{\delta(S^i)}\right)} \tag{3}$$

[2]the second convolutional layer of DeepLab-v2 (Huang et al. 2023), or the last Layer Norm of the second Transformer Block for DAFormer/HRDA (Chen et al. 2023), respectively.

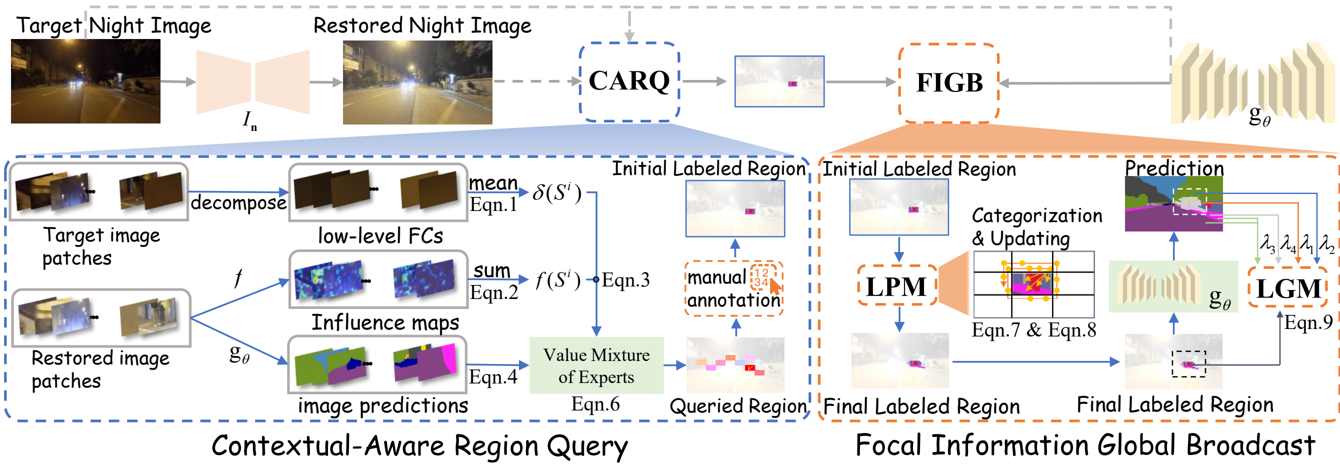

Figure 2: **The Overview of FS-ADA framework.** In detail, an image restoration network I_n is fine-tuned to fit the target data distribution, and then the target image is processed through I_n to obtain the restored image. In the target region query process, based on the knowledge of the target image, restored image, and semantic segmentation network g_θ, CARQ adaptively queries the most valuable region in the target domain image, which is provided to the annotator to annotate manually and gets the initial labeled region. After that, to maximize the information of the labeled regions, FIGB implements a global broadcast of labeling information. Specifically, FIGB consists of LPM and LGM. The former propagates the information of the initial labeled region to its neighbors and the latter propagates the information of all labels to the global.

where $\delta(S^i)$ and $f(S^i)$ are the severity of weather-specific distraction and its impact on model predictions, respectively, for the i-th image patch in any image from D_T.

Semantic Complexity. In semantic segmentation, a higher number of predicted classes in the region contributes more to performance gains. Prediction uncertainty also quantifies the contribution to performance gains. Each region's semantic attributes include the *number of classes*, *class type*, and *number of class pixels*, which are all derived from the prediction and together reflect the model's familiarity with it. Given a region, our newly proposed region Semantic complexity $\mathcal{S}(\cdot)$ is calculated as follows:

$$\mathcal{S}(\cdot) = log(\Omega) \times \sum_{c=1}^{C} \left[\omega_c \times \frac{exp(\frac{\xi}{h \cdot w})}{\sum\limits_{c=1}^{C} exp(\frac{\xi}{h \cdot w})} \right]$$
$$\Omega = \sum_{c=1}^{C} \mathbb{I}(\underset{c\in\{1,\ldots,C\}}{\arg\max}\, g_\theta \left((x_{te}^{j})^{(i)} \right) = c) \tag{4}$$
$$\xi = \left| \underset{c\in\{1,\ldots,C\}}{\arg\max}\, g_\theta \left((x_{te}^{j})^{(i)} \right)^c \right|$$

where Ω and ξ represent the *number of classes* and the *number of c-th class pixels* in the i-th region, respectively. h and w are the height and width of the region, respectively. $\mathbb{I}$ is an indicator function. x_{te}^{j} is the restored target night image. $\mathcal{C}$ denotes all semantic classes. w_c denotes the model's prediction uncertainty for *c-th class type*, calculated as follows:

$$\omega_c = \frac{1}{\varphi} \sum_{j=1}^{N_t} \sum_{i=1}^{H\times W} I_{x_{te}^{j}}^{(i)} * \mathbb{I} \left(\underset{c\in\{1,\ldots,C\}}{\arg\max}\, g_\theta \left((x_{te}^{j})^{(i)} \right)^c = c \right)$$
$$\varphi = \sum_{j=1}^{N_t} \sum_{i=1}^{H\times W} \left| \underset{c\in\{1,\ldots,C\}}{\arg\max}\, g_\theta \left(\left(x_{te}^{j} \right)^{(i)} \right)^c \right| \tag{5}$$
$$I_{x_{te}^{j}}^{(i)} = -\frac{1}{\log(C)} \sum_{c=1}^{C} p_{x_{te}^{j}}^{(i,c)} \log p_{x_{te}^{j}}^{(i,c)}$$

where φ is the number of pixels for the c-th class in the restored target nighttime domain $D_{TE} = \left\{ (x_{te}^{j}) \right\}_{j=1}^{N_t}$. H and W are the image height and width, respectively. $I_{x_{te}^{j}}^{(i)}$ is the normalized pixel-wised entropy map overall C classes. $|\cdot|$ denotes the number of pixels in the set. $p_{x_{te}^{j}}^{(i,c)}$ denotes the softmax probability of i-th pixel in x_{te}^{j}.

Value Mixture of Experts. To better measure the value of $S_{\mathbf{x}_t^j}$ via $\mathcal{D}(\cdot)$ and $\mathcal{S}(\cdot)$, a Value Mixture of Experts is proposed to leverage the complementary information from the above two value metrics (*i.e.,* "value metric experts"). Specifically, the final acquisition function is obtained via dynamically assigning importance to each expert as follows:

$$\mathcal{A}(S_{\mathbf{x}_t^j}^{(i)}) = \textstyle\sum_{i=1}^{n} G(x)_i E_i(x) = G(x)_1 \mathcal{D}(\cdot) + G(x)_2 \mathcal{S}(\cdot) \tag{6}$$

where $\sum_{i=1}^{n} G(x)_i = 1$ and $G(x)_i$ are the weight of expert $f_i(x)$. Initially, $G(x)_i$ is set to normalized value gains (*i.e.,* mIoU gains), then dynamically updated by the value gains ($\triangle E_i(x)$). When total $\mathcal{S}(\cdot)$ decreases/increases, $G(x)_2$ is increased/decreased by α. Otherwise, it remains unchanged. $G(x)_i$ is updated with the new normalized weight.

Focal Information Global Broadcast

To supplement and maximize information from limited labeled data, we design FIGB, comprising LPM and LGM. LPM propagates labels to neighboring regions, incorporating confident data into the labeled set $D_{TL} = \left\{((x_t^j)^{(n)}, (y_t^j)^{(n)})\right\}_{j=1}^{N_t}$. LGM utilizes these labels to guide the unlabeled data's learning $D_{TU} = \left\{(x_t^j)^{(m)}\right\}_{j=1}^{N_t}$, where n and m are labeled and unlabeled regions, respectively.

Label Propagation Module. Inspired by the neighboring regions having a high probability of containing the same semantic class in semantic segmentation, LPM progressively propagates pixel labels from labeled to neighboring regions in a growing manner. The key issues in this process are initial seed location and similarity criterion. For the initial seed location, we average the features of each class c in the labeled regions to estimate the c-th class prototype A^c as:

$$A^c = \frac{1}{|\Lambda_c|} \sum_{(x,y)\in D_{TL}} \mathbb{I}\{y = c\} \cdot f\left(x\right)|_c \tag{7}$$

where $|\Lambda_c|$ is the number of pixels belongs to category c and $\mathbb{I}$ is an indicator function. x and y are the image pixel and corresponding label in D_{TL}, respectively. $f(x)|_c$ denotes the feature output of our segmentation network g_θ for c.

For the similarity criterion, we first obtain the predicted class $\hat{c}$ and confidence $\mathcal{C}$ for each pixel in the neighboring regions. To ensure accurate label propagation, we propagate the label c^* in D_{TL} to pixel x in neighboring regions when the following condition is satisfied:

$$\begin{aligned}&(y_{te}^j)^{(m')} = \underset{c^*\in y'}{\arg\max}\ cos\left(f(x), A^{c^*}\right) \text{ only if } \mathcal{C} \geq 0.2 \text{ and}\\&cos\left(f(x), A^{c^*}\right) > \frac{\sum_{(x',y')\in D_{TL}} \mathbb{I}\{y' = c^*\} \cdot cos\left(f\left(x'\right)|_{c^*}, A^{c^*}\right)}{\sum_{(x',y')\in D_{TL}} \mathbb{I}\{y' = c^*\}}\end{aligned} \tag{8}$$

where x is the pixel in the neighboring regions. cos is the cosine similarity between feature vectors. x' and y' are the image pixel and corresponding label in D_{TL}, respectively. $\mathcal{C} < 0.2$ are false positives and discarded (Brüggemann et al. 2023). m' are all propagated pixels. As the labels progressively propagate, we online update A^c by adding the feature of pixel x with propagated label c when $\hat{c}$ is equal to c^*.

Label Generalization Module. To enhance the global impact of labeled data, a prediction consistency loss, $\mathcal{L}_{con}$, is designed to constrain the consistency between model predictions and labels in D_{TL}. In semantic segmentation, features of the same semantic class are closer in feature space than those of different classes. Meanwhile, the closer the spatial location relationship is considered, the greater the feature similarity of the same semantic class. We further enforce feature consistency between non-propagated neighboring regions and labeled regions of the same semantic class, as well as between non-neighbors and labeled regions in D_{TL}. In adverse weather conditions (*e.g.*, nighttime), semantic classes of the same category exhibit similar interference for their spatial adjacency. Finally, we propose a spatial interference consistency loss $\mathcal{L}_{sic}$ to constrain the interference consistency within semantic classes of the same category. After LPM, the segmentation network g_θ is optimized via our LGM with the training loss represented as follows:

$$\begin{aligned}\mathcal{L} = {} & \lambda_1 \mathcal{L}_{con}\left(g_\theta\left(I_n((x_t^j)^{(n)})\right), (y_t^j)^{(n)}\right)\\& + \lambda_2 \sum_{c=1}^{C}\left(1 - cos\left(\sum_{x\in D_{TU1}} f(x)|_c, \sum_{(x,y)\in D_{TL}} \mathbb{I}\{y = c\} \cdot f(x)|_c\right)\right)\\& + \lambda_3 \sum_{c=1}^{C}\left(1 - cos\left(\sum_{x\in D_{TU2}} f(x)|_c, \sum_{(x,y)\in D_{TL}} \mathbb{I}\{y = c\} \cdot f(x)|_c\right)\right)\\& + \lambda_4 \mathcal{L}_{sic}\left(\sum_{s}\left(1 - cos\left(\sum_{(x,y)\in D_T} \mathbb{I}\{y = c_1\} \cdot \phi(x), \sum_{(x,y)\in D_T} \mathbb{I}\{y = c_2\} \cdot \phi(x)\right)\right)\right)\end{aligned} \tag{9}$$

where $(x_t^j)^{(n)}$ and $(y_t^j)^{(n)}$ denote all image and labels in D_{TL} including all propagated pixels, respectively. $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ is the coefficient. cos is the cosine similarity between feature vectors. $\mathbb{I}$ is an indicator function, and f is the feature extractor of the g_θ. D_{TU1} and D_{TU2} denote the

	Method	mIoU↑
Backbone: DeepLab-v2 (Chen et al. 2017)		
Source Only	DeepLab-v2 (Chen et al. 2017)	24.6
UDA-NSS	DANNet (Wu et al. 2021a)	41.3
	Bi-Mix (Yang et al. 2021)	40.2
	Refign (Bruggemann et al. 2023)	49.8
ADA-SS	RIPU (Xie et al. 2022) (1%)	49.7
	D^2ADA (Wu et al. 2022) (1%)	52.0
	HALO (Franco et al. 2024) (1%)	50.3
ADA-NSS	FS-ADA (1%)	52.4
Fully Supervised	DeepLab-v2 (Chen et al. 2017)	**54.8**
	IA-Seg (Liu et al. 2023)	47.9
Backbone: DeepLabv3+ (Chen et al. 2018)		
Source Only	DeepLabv3+ (Chen et al. 2018)	33.2
UDA-NSS	DeepLabv3+ (Chen et al. 2018)	36.1
ADA-SS	RIPU (Xie et al. 2022) (1%)	50.0
	D^2ADA (Wu et al. 2022) (1%)	50.4
ADA-NSS	FS-ADA (1%)	52.4
Fully Supervised	DeepLabv3+ (Chen et al. 2018)	**57.0**
Backbone: DAFormer (Hoyer, Dai, and Van Gool 2022a)		
Source Only	DAFormer (Hoyer, Dai, and Van Gool 2022a)	39.8
UDA-NSS	DAFormer (Hoyer, Dai, and Van Gool 2022a)	45.8
	Refign (Bruggemann et al. 2023)	54.8
	DTPS (Huang, Yao, and Zhou 2023)	53.8
	InforMS (Wang et al. 2023)	56.9
	PIG (Xie et al. 2024)	51.0
ADA-NSS	FS-ADA (1%)	59.1
Fully Supervised	DAFormer (Hoyer, Dai, and Van Gool 2022a)	**62.0**
Backbone: HRDA (Hoyer, Dai, and Van Gool 2022b)		
Source Only	HRDA (Hoyer, Dai, and Van Gool 2022b)	47.7
UDA-NSS	HRDA (Hoyer, Dai, and Van Gool 2022b)	53.1
	Refign (Bruggemann et al. 2023)	63.5
	InforMS (Wang et al. 2023)	65.1
	PIG (Xie et al. 2024)	56.9
ADA-NSS	FS-ADA (1%)	66.3
Fully Supervised	HRDA (Hoyer, Dai, and Van Gool 2022b)	**67.4**

Table 1: Comparison on ACDC-night-test for Cityscapes→ACDC nighttime adaptation. The best results for each backbone are shown in bold. Our FS-ADA based on different backbones outperforms existing ADA-SS methods with the same 1% target annotation budget.

	Method	D1	D2	D3	D4
Backbone: DeepLab-v2 (Chen et al. 2017)					
Source Only	DeepLab-v2	16.3	17.9	18.4	12.3
UDA-NSS	DeepLab-v2	23.3	25.6	27.1	31.5
ADA-SS	RIPU (1%)	43.8	47.0	50.0	30.3
	D^2ADA (1%)	45.2	38.7	44.4	24.8
ADA-NSS	FS-ADA (1%)	45.6	47.7	50.2	**37.7**
Fully Supervised	DeepLab-v2	**48.1**	**48.7**	**51.2**	36.3
	IA-Seg	42.9	48.3	51.0	34.5
Backbone: DAFormer (Hoyer, Dai, and Van Gool 2022a)					
Source Only	DAFormer	29.1	27.2	30.2	22.3
UDA-NSS	DAFormer	36.8	32.4	38.9	37.8
ADA-NSS	FS-ADA (1%)	**52.1**	41.4	43.1	**40.5**
Fully Supervised	DAFormer	51.2	**42.5**	**45.3**	39.4
Backbone: HRDA (Hoyer, Dai, and Van Gool 2022b)					
Source Only	HRDA	34.9	31.7	34.9	29.6
UDA-NSS	HRDA	37.4	32.6	42.9	31.3
ADA-NSS	FS-ADA (1%)	**57.8**	43.5	50.8	**42.0**
Fully Supervised	HRDA	55.9	**45.8**	**52.7**	40.0

Table 2: Comparison with existing methods on D1 (ACDC-night-val), D2 (NightCity val), D3 (NightCity-fine val) and D4 (BDD100K-Night val). Best results per backbone are bolded.

set from non-propagated regions in the neighboring regions of the initial labeled data, and the non-neighboring region, respectively. ϕ is the interference estimator. c_1 and c_2 denotes the semantic class of the same category s.

Experiments

Experimental Datasets

We evaluate our approach using 1) four day-to-nighttime SS benchmarks: Cityscapes→ACDC adaptation for nighttime, Cityscapes→NightCity, Cityscapes→NightCity-fine, and BDD100K daytime→BDD100K nighttime; 2) one clear-to-adverse SS benchmark: Cityscapes→ACDC; and 3) Cityscapes→ACDC foggy/rainy/snowy subset.

Cityscapes (Cordts et al. 2016) contains 2,975 training, 500 validation, and 1,525 testing daytime images. ACDC (Sakaridis, Dai, and Van Gool 2021) includes 4,006 images captured under adverse conditions (fog, snow, rain, and nighttime). NightCity (Tan et al. 2021) and NightCity-fine (Wei et al. 2023) are nighttime datasets, with the latter being label-optimized. BDD100K (Yu et al. 2020) consists of 100,000 images, 320/34 night images in train/val sets are picked to the component nighttime subset (BDD100K-Night) and the others belong to the daytime subset.

Comparison with State-of-the-art Methods

Comparison on ACDC nighttime scene. We conduct comparative analyses of various methods, including UDA-NSS, ADA-SS, and Fully Supervised approaches, based on different backbones. The quantitative results of FS-ADA and existing methods on the ACDC-night-test and the ACDC-night-val datasets are reported in Table 1 and Table 2, respectively. Notably, FS-ADA outperforms existing UDA-SS and ADA-SS methods on the ACDC-night-test and ACDC-night-val datasets with the same 1% target annotation budget. To further validate the effectiveness, we compare it with the Fully Supervised. Concisely, our FS-ADA based on DAFormer and HRDA with only 1% target annotation budget can outperform the Fully Supervised models on the ACDC-night-val dataset, even by a significant margin (+0.9 mIoU [%] and +1.9 mIoU [%]). It confirmed that our method is effective and efficient.

We further visualize the segmentation results predicted by our proposed FS-ADA based on DeepLab-v2 and HRDA in comparison to existing UDA-NSS and ADA-SS methods in Figure 3. Comparatively, the segmentation results of our proposed method are closer to the ground truth. Besides, our method achieves more accurate predictions for dynamic and rare classes with a few pixels in the source & target domains.

Comparison on NightCity and NightCity-fine. Quantitative results on the NightCity val and NightCity-fine val datasets are reported in Table 2. FS-ADA outperforms current state-of-the-art ADA-SS methods with the 1% target annotation budget, verifying the effectiveness of our method. The corresponding visual comparison is shown in Figure 3.

Comparison on BDD100K-Night. We perform the BDD100K-day→BDD100K-night training set adaptation. The quantitative results on the BDD-100K-night val dataset are displayed in Table 2, with sample visualization results presented in Figure 3. FS-ADA achieves sufficient gain compared to the fully supervised competitor with only 1% target budget, verifying the effectiveness of our method.

Restoration	Segmentation		ACDC-night-val	
Fine-tune	Aligned fragments	Complementary classes	DAFormer	HRDA
			45.9	53.3
✓			47.0	54.7
✓	✓		48.0	55.5
✓	✓	✓	**49.2**	**55.7**

Table 3: Ablation study of our restoration and segmentation framework during the warmup step on the ACDC-night-val.

Selection	Training		ACDC-night-val	
CARQ	LPM	LGM	DAFormer	HRDA
			49.2	55.7
✓			51.0	56.4
✓	✓		51.2	56.9
✓		✓	51.9	57.0
✓	✓	✓	**52.1**	**57.8**

Table 4: Ablation Study of Different Components Combinations on the ACDC-night-val dataset.

Ablation Study

Restoration and Segmentation Evaluation. To validate the effectiveness of each strategy in our restoration and segmentation framework during the warmup step, we conduct several ablation studies on the ACDC-night-val, shown in Table 3. For the baseline implementation (Row 1), D_T are directly restored to D_{TE} by existing mainstream enhancement network, *e.g.,* PyDiff (Zhou, Yang, and Yang 2023),

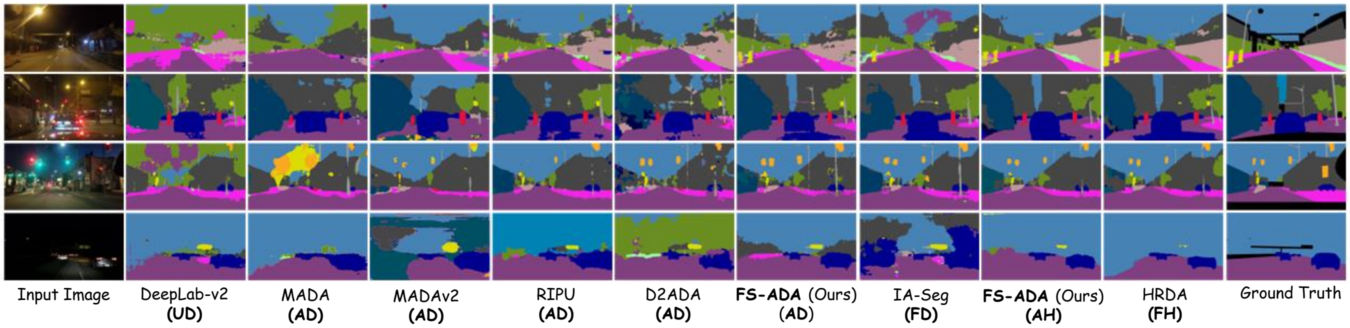

Figure 3: The qualitative comparison between our FS-ADA and existing state-of-the-art methods based on DeepLab-v2 and HRDA on the ACDC-night-val, NightCity val, NightCity-fine val, and BDD100K-Night val dataset, respectively. Comparatively, the segmentation results of our FS-ADA method are superior to existing UDA-SS and ADA-SS methods, approaching more fully supervised performance. In parentheses, the first letters U, A, and F stand for UDA, ADA, and Fully Supervised, respectively; the second letters D and H stand for DeepLab-v2 and HRDA, respectively.

		Clear-to-Adverse-Weather	Clear-to-Foggy	Clear-to-Rainy	Clear-to-Snowy
	Method	ACDC test	ACDC-fog-test	ACDC-rain-test	ACDC-snow-test
Backbone: DeepLab-v2 (Chen et al. 2017)					
UDA-SS	DeepLab-v2 (Chen et al. 2017)	38.0	33.5	44.5	40.2
	Refign (Bruggemann et al. 2023)	48.0	51.1	57.0	57.1
	VBLC (Li et al. 2023)	47.8	-	-	-
	CMA (Brüggemann et al. 2023)	50.4	-	-	-
	CISS (Sakaridis et al. 2023)	47.2	42.1	54.2	49.1
	FS-ADA (1%)	**58.0**	51.9	**59.2**	**57.7**
Fully Supervised	DeepLab-v2 (Chen et al. 2017)	55.3	**52.2**	57.6	56.8

Table 5: Comparisons with existing UDA-SS and Fully Supervised on adverse, foggy, rainy, and snowy across four benchmarks.

then performs domain adaptation from the D_S to the D_{TE} via the existing UDA-NSS method, *e.g.,* Refign (Bruggemann et al. 2023). Then, we fine-tune the restoration network to fit our data distribution (Row 2). By doing so, 1.1/1.4 mIoU [%] performance gain is achieved, respectively. Then, we constraint the prediction consistency of D_{TE} and aligned target reference domain when performing domain adaptation from D_S to D_{TE} (Row 3). Finally, we further optimize the complementary classes learning, which achieves the best performance on the ACDC-night-val. It validates that each design is effective.

Overall Component Evaluation. To assess the efficacy of each constituent within FS-ADA, we train several model variants, whose backbone is DAFormer or HRDA, and evaluate their performance on the ACDC-night-val dataset, as delineated in Table 4. We choose our designed restoration and segmentation framework as the baseline. Subsequently, we adopt CARQ to select the nighttime regions with the 1% budget, followed by fine-tuning the model in a supervised manner. This procedure yielded a notable 1.8/0.7 mIoU [%] improvement, respectively. Building upon this, we adopt LPM and LGM to optimize our model, respectively. Both improve performance, with the latter showing greater improvement. When two manners are integrated into a unified selection criterion, the performance is further boosted to 52.1/57.8 mIoU [%], respectively, verifying the overall effectiveness of our proposed methods.

Generalization to adverse, foggy, rainy, and snowy

To further verify the generalization of our FS-ADA across scenes, extensive experiments are conducted on the Cityscapes→ACDC and Cityscapes→ACDC weather-specific subsets (Clear-to-Adverse, Clear-to-Foggy, Clear-to-Rainy, and Clear-to-Snowy), respectively. Quantitative results on each test dataset are shown in Table 5. Our FS-ADA outperforms existing UDA-SS methods and matches the Fully Supervised method with only 1% target annotation budget, verifying its effectiveness and generalization.

Conclusion

We initiated the ADA-NSS task, pioneering a new direction in the nighttime semantic segmentation field. Our proposed CARQ data sampling strategy deeply considers information biases caused by domain degradation and semantic complexity, enabling it to select the most informative areas from the target domain adaptively. Additionally, we introduced a Label Propagation Module and a Label Generalization Module, which facilitate the propagation of limited area information to the global scale. Overall, using minimal annotated data, our FS-ADA framework has achieved significant advancements in semantic segmentation performance. Extensive experiments show that our method outperforms state-of-the-art UDA-NSS and ADA-SS methods on four public NSS benchmarks. With only 1% of the target nightly data annotated, our method performs similarly to mainstream fully supervised methods on the BDD100K-Night val dataset.

Acknowledgements

This work was supported by the Natural Science Foundation of China (62376201, 61602349, U1803262, and 61440016), and the Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System. This research was also financially supported by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), under Grant No.GML-KF-24-09.

References

Brüggemann, D.; Sakaridis, C.; Brödermann, T.; and Van Gool, L. 2023. Contrastive model adaptation for cross-condition robustness in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11378–11387. Paris, France: IEEE.

Bruggemann, D.; Sakaridis, C.; Truong, P.; and Van Gool, L. 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*, WACV '23, 3174–3184. Waikola, HI, USA: IEEE.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 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): 834–848.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision*, 801–818. Munich, Germany: Springer.

Chen, Y.; Liu, H.; Yin, H.; and Fan, B. 2023. Building Vision Transformers with Hierarchy Aware Feature Aggregation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5908–5918. Paris, France: IEEE.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3213–3223. Las Vegas, NV, USA: IEEE.

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

Ding, F.; Li, J.; and Tian, W. 2023. Dual-level Consistency Learning for Unsupervised Domain Adaptive Nighttime Semantic Segmentation. In *2023 IEEE International Conference on Multimedia and Expo*, ICME '23, 420–425. Brisbane, Australia: IEEE.

Dong, C.; Kang, X.; and Ming, A. 2023. ICDA: Illumination-Coupled Domain Adaptation Framework for Unsupervised Nighttime Semantic Segmentation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 672–680. Macao, SAR, China: IEEE.

Franco, L.; Mandica, P.; Kallidromitis, K.; Guillory, D.; Li, Y.-T.; and Galasso, F. 2024. Hyperbolic Active Learning for Semantic Segmentation under Domain Shift. In *International Conference on Machine Learning*, 1–12. Vienna, Austria.

Han, K.; Kim, Y.; Han, D.; Lee, H.; and Hong, S. 2023. TL-ADA: Transferable loss-based active domain adaptation. *Neural Networks*, 161: 670–681.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022a. 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*, 9924–9935. New Orleans, LA, USA: IEEE.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022b. HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation. In *European Conference on Computer Vision*, 372–391. Tel Aviv, Israel: Springer.

Huang, F.; Yao, Z.; and Zhou, W. 2023. DTBS: Dual-Teacher Bi-Directional Self-Training for Domain Adaptation in Nighttime Semantic Segmentation. In *European Conference on Artificial Intelligence*, 1084–1091. Kraków, Poland.

Huang, W.; Chen, C.; Li, Y.; Li, J.; Li, C.; Song, F.; Yan, Y.; and Xiong, Z. 2023. Style projected clustering for domain generalized semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3061–3071. Vancouver, BC, Canada: IEEE.

Li, M.; Xie, B.; Li, S.; Liu, C. H.; and Cheng, X. 2023. VBLC: visibility boosting and logit-constraint learning for domain adaptive semantic segmentation under adverse conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8605–8613. Washington, DC, USA.

Liu, S.; Jiang, Z.; Li, Y.; Peng, J.; Wang, Y.; and Lin, W. 2024a. Density Matters: Improved Core-Set for Active Domain Adaptive Segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '24, 13999–14007. Vancouver, Canada: AAAI Press.

Liu, W.; Cai, J.; Li, Q.; Liao, C.; Cao, J.; He, S.; and Yu, Y. 2024b. Learning Nighttime Semantic Segmentation the Hard Way. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(7): 1–17.

Liu, W.; Li, W.; Zhu, J.; Cui, M.; Xie, X.; and Zhang, L. 2023. Improving nighttime driving-scene segmentation via dual image-adaptive learnable filters. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10): 5855–5867.

Ning, M.; Lu, D.; Xie, Y.; Chen, D.; Wei, D.; Zheng, Y.; Tian, Y.; Yan, S.; and Yuan, L. 2023. MADAv2: Advanced Multi-Anchor Based Active Domain Adaptation Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 13553–13566.

Sakaridis, C.; Bruggemann, D.; Yu, F.; and Van Gool, L. 2023. Condition-invariant semantic segmentation. *arXiv preprint arXiv:2305.17349*.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding. In *Proceedings of the*

IEEE/CVF International Conference on Computer Vision, 10765–10775. Montreal, BC, Canada: IEEE.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2022. Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6): 3139–3153.

Tan, X.; Xu, K.; Cao, Y.; Zhang, Y.; Ma, L.; and Lau, R. W. 2021. Night-Time Scene Parsing With a Large Real Dataset. *IEEE Transactions on Image Processing*, 30: 9085–9098.

Wang, S.; Xu, X.; Chen, H.; Jiang, K.; Wang, Z.; and Tang, K. 2024. Low-Light Salient Object Detection Meets the Small Size. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 1–13.

Wang, S.; Xu, X.; Liu, L.; and Tian, J. 2020. Multi-level feature fusion model-based real-time person re-identification for forensics. *Journal of Real-Time Image Processing*, 17(1): 73–81.

Wang, S.; Xu, X.; Ma, X.; Jiang, K.; and Wang, Z. 2023. Informative Classes Matter: Towards Unsupervised Domain Adaptive Nighttime Semantic Segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 163–172. Ottawa, ON, Canada: Association for Computing Machinery.

Wei, Z.; Chen, L.; Tu, T.; Ling, P.; Chen, H.; and Jin, Y. 2023. Disentangle then Parse: Night-time Semantic Segmentation with Illumination Disentanglement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 21593–21603. Paris, France: IEEE.

Wu, T.-H.; Liou, Y.-S.; Yuan, S.-J.; Lee, H.-Y.; Chen, T.-I.; Huang, K.-C.; and Hsu, W. H. 2022. D^2ada: Dynamic density-aware active domain adaptation for semantic segmentation. In *European Conference on Computer Vision*, 449–467. Tel Aviv, Israel: Springer.

Wu, X.; Wu, Z.; Guo, H.; Ju, L.; and Wang, S. 2021a. 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*, 15769–15778. Virtual Event: IEEE.

Wu, X.; Wu, Z.; Ju, L.; and Wang, S. 2021b. A one-stage domain adaptation network with image alignment for unsupervised nighttime semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 58–72.

Xie, B.; Yuan, L.; Li, S.; Liu, C. H.; Cheng, X.; and Wang, G. 2022. Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8068–8078. New Orleans, LA, USA: IEEE.

Xie, M.; Li, S.; Zhang, R.; and Liu, C. H. 2023a. Dirichlet-based uncertainty calibration for active domain adaptation. In *Proceedings of the Eleventh International Conference on Learning Representations*, 1–12. Kigali, Rwanda: OpenReview.net.

Xie, Z.; Qiu, R.; Wang, S.; Tan, X.; Xie, Y.; and Ma, L. 2024. PIG: Prompt Images Guidance for Night-Time Scene Parsing. *IEEE Transactions on Image Processing*, 33: 3921–3921.

Xie, Z.; Wang, S.; Xu, K.; Zhang, Z.; Tan, X.; Xie, Y.; and Ma, L. 2023b. Boosting Night-time Scene Parsing with Learnable Frequency. *IEEE Transactions on Image Processing*, 32: 2386–2398.

Xu, M.; Qin, L.; Chen, W.; Pu, S.; and Zhang, L. 2023. Multi-view adversarial discriminator: Mine the non-causal factors for object detection in unseen domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8103–8112. Vancouver, BC, Canada: IEEE.

Xu, X.; Wang, S.; Wang, Z.; Zhang, X.; and Hu, R. 2021. Exploring image enhancement for salient object detection in low light images. *ACM transactions on multimedia computing, communications, and applications*, 17(1s): 1–19.

Yang, G.; Zhong, Z.; Tang, H.; Ding, M.; Sebe, N.; and Ricci, E. 2021. Bi-mix: Bidirectional mixing for domain adaptive nighttime semantic segmentation. arXiv:2111.10339.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2636–2645. Seattle, WA, USA: IEEE.

Zhang, H.; and Zhang, R. 2022. Active domain adaptation with multi-level contrastive units for semantic segmentation. In *Proceedings of the Asian Conference on Computer Vision*, ACCV '22, 1640–1657. Macau SAR, China: Springer.

Zhou, D.; Yang, Z.; and Yang, Y. 2023. Pyramid Diffusion Models For Low-light Image Enhancement. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 1795–1803. Macao, SAR, China: IEEE.