

Let There be Light: Improved Traffic Surveillance via Detail Preserving Night-to-Day Transfer

Lan Fu , Student Member, IEEE, Hongkai Yu , Member, IEEE, Felix Juefei-Xu , Member, IEEE, Jinlong Li , Qing Guo [†], Member, IEEE, and Song Wang , Senior Member, IEEE

Abstract—In recent years, image and video surveillance have made considerable progresses to the Intelligent Transportation Systems (ITS) with the help of deep Convolutional Neural Networks (CNNs). As one of the state-of-the-art perception approaches, detecting the interested objects in each frame of video surveillance is widely desired by ITS. Currently, object detection shows remarkable efficiency and reliability in standard scenarios such as daytime scenes with favorable illumination conditions. However, in face of adverse conditions such as the nighttime, object detection loses its accuracy significantly. One of the main causes of the problem is the lack of sufficient annotated detection datasets of nighttime scenes. In this paper, we propose a framework to alleviate the accuracy decline when object detection is taken to adverse conditions by using image translation method. We propose to utilize style translation based StyleMix method to acquire pairs of day time image and nighttime image as training data for following nighttime to daytime image translation. To alleviate the detail corruptions caused by Generative Adversarial Networks (GANs), we propose to utilize Kernel Prediction Network (KPN) based method to refine the nighttime to daytime image translation. The KPN network is trained with object detection task together to adapt the trained daytime model to nighttime vehicle detection directly. Experiments on vehicle detection verified the accuracy and effectiveness of the proposed approach.

Index Terms—nighttime object detection, image translation, kernel prediction network

I. INTRODUCTION

WITH the fast development of computer vision and deep Convolutional Neural Networks (CNNs), visual data understanding in image and video has attracted a lot of attention [1], [2], [3], [4], [5], [6], [7]. For example, in the Intelligent Transportation Systems (ITS), detecting the vehicles in each frame of the traffic surveillance video is important to extract the real-time traffic flow parameters [8] for the efficient traffic control and obtain the vehicle trajectories [9] for the calibrated traffic model simulation, *etc.*

Most of existing researches focus on daytime perception task through supervised learning, however, they generalize badly on adverse conditions such as nighttime scenarios [10]. The

Lan Fu and Song Wang are with the Department of Computer Science and Engineering, University of South Carolina, SC, USA, e-mail: lanf@email.sc.edu, songwang@cec.sc.edu.

Hongkai Yu and Jinlong Li are with the Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH, USA, e-mail: h.yu19@csuohio.edu and lijinlong1117@foxmail.com.

Felix Juefei-Xu is with the Alibaba Group, Sunnyvale, CA, USA, e-mail: juefei.xu@gmail.com.

Qing Guo is with the Nanyang Technological University, Singapore, e-mail: tsqingguo@ieee.org.

[†] indicates the co-corresponding authors: Qing Guo and Song Wang.

adversity of nighttime scenario poses two challenges for the success of perception task at nighttime: 1) nighttime data with a large amount of annotations is usually scarce compared to the large-scale daytime data, since accurate annotation of nighttime images is relatively hard to obtain. 2) The visual hazards, such as underexposure and noise, of nighttime images cause the extracted features corrupted.

One traditional way to solve this is to fine-tune the already-trained daytime perception model on the limited nighttime data, and hopefully it can perform well on nighttime scenarios, but it requires extra time and additional-labeled nighttime data for model fine-tuning. Another traditional way [10] might use Generative Adversarial Networks (GANs) based image to image translation methods in an unpaired way, such as CycleGAN [11] and UNIT [12], to transfer daytime images to fake nighttime images. Paired daytime and nighttime images are hard to obtain in the real-world applications, due to the dynamic traffic and environment changes. This kind of image translation considers this problem as domain adaptation for model fine-tuning on synthetic nighttime images without labeling the nighttime data. However, these methods also need extra time for model fine-tuning. In addition, GAN based image translation suffers from model collapse and does not preserve content details very well [13], [12], [11], [14], [15]. Bottleneck layers in a general deep generator hurt the learning ability of convolution kernels due to downsampling and upsampling operations, resulting in possible losing some structure details. Besides, unpaired training data of different domains limits the detail-preserving ability of generators due to the lack of pixel-wise correspondence.

In this paper, we would like to reuse the daytime perception model to nighttime scenarios. Our basic idea is to maximally use the pretrained daytime perception model, similar to works [16], [17], which could be easily extended to the nighttime tasks. Reversely to the traditional methods, we transfer the nighttime images back to the daytime style with the detail-preserving to reuse the trained daytime perception model, as shown in Figure 1. The strengths of this reverse way are obvious and promising: 1) there are no extra training efforts for the already trained daytime perception model and no needs to manually label the nighttime data; 2) image transfer could reduce the domain distribution discrepancy between daytime and nighttime data; 3) detail-preserving image transfer could better maintain the structure details than the GAN based image transfer.

Specifically, we propose a detail-preserving unpaired domain transfer method for this task, which mainly contains two components: 1) Style-transfer based StyleMix, 2) Kernel

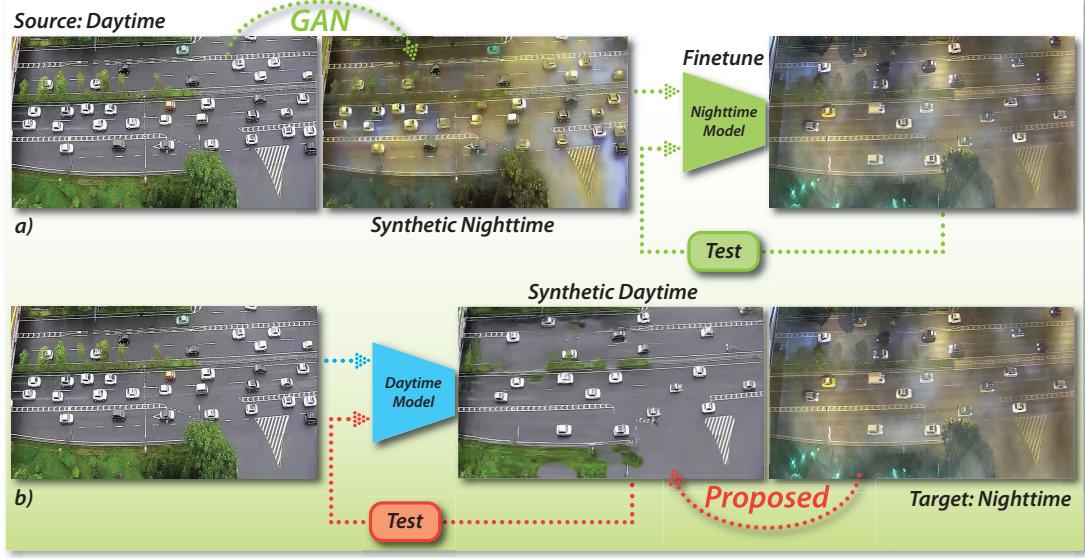


Figure 1: Illustration of the domain reuse problem: a) traditional method with style transfer and the nighttime model fine-tuned from the daytime model, b) the proposed detail-preserving Night-to-Day translation method without changing the daytime model.

Prediction Network (KPN) based nighttime to daytime image transfer. Without paired daytime-nighttime image pairs, we propose to utilize style translation based StyleMix method, inspired by AugMix [18], to acquire pairs of daytime and nighttime images as training data for the following nighttime to daytime image transfer. We can effectively alleviate the detail corruption caused by GAN: 1) The synthetic nighttime image and corresponding daytime image translation can provide pixel-wise correspondence for night-to-day translation. 2) Kernel prediction network based method can refine the nighttime to daytime image translation because the per-pixel kernel fusion can effectively utilize the neighboring region for each pixel and could learn more spatial context representing structure information. The proposed method can conduct daytime and nighttime vehicle detection with just one daytime model, which is more convenient in real-world applications.

In this paper, we choose the vehicle detection problem in the traffic surveillance video as a case study for the proposed approach. The KPN network is trained with object detection task together to adapt the trained daytime model to fit nighttime domain directly. Experimental results on a vehicle video dataset in daytime and nighttime verified the accuracy and effectiveness of the proposed approach. The contributions of this paper are summarized in the following:

- We propose a detail-preserving unpaired domain transfer method for nighttime vehicle detection to adapt the trained daytime model directly for nighttime vehicle detection.
- To solve the problem of lack of paired daytime-nighttime image pairs, we propose to utilize style translation based StyleMix method to obtain pairs of daytime image and nighttime image as training data. These training data are utilized by KPN network to perform nighttime to daytime image transfer.
- The comprehensive experimental results on a vehicle detec-

tion dataset from the video surveillance scenario in daytime and nighttime show that the proposed method achieved better vehicle detection performance in nighttime scenario.

In the following of this paper, Section II reviews the related work. Section III explains the proposed method. Experiment setting and results are described in Section IV, followed by a conclusion in Section V.

II. RELATED WORK

Object detection at nighttime: The state-of-the-art performance in object detection is rapidly improving in recent years. One-stage (SSD [19], YOLO [20], RetinaNet [21]) and two-stage (Faster R-CNN [22], Mask R-CNN [23]) detection frameworks have achieved promising performance in real-world applications. They generally require a large amount of manually labeled data for supervised learning. Nonetheless, most of them operate well at daytime, under favorable illumination conditions, and scale badly to nighttime scenarios with challenging lighting conditions. Further, manual annotation of nighttime images are hard and time-consuming, because even human cannot clearly discern objects in adverse nighttime scenario. Nighttime detection task has attracted a lot of attention recently. Domain specific works [24], [25], [26] explore the human detection at nighttime by considering the type of cameras. Other works [27], [28] pertain to vehicle detection in driving scenarios. Domain-invariant representations [29], [30] or fusion works [31] are designed to be robust to illumination changes. Image translation work [32] aims to improve retrieval-based localization at nighttime. In this work, we focus on vehicle detection at nighttime in traffic surveillance scenarios. We aim to adapt the daytime detection model to nighttime detection for reusing the daytime domain knowledge. It is also interesting to explore the robustness of daytime model from other perception tasks [33],

[34], *e.g.*, object tracking [35], [36], [37], [38], at nighttime scenario in the future.

Learning from synthetic data: In general, CNNs perform much worse when there is a domain shift between training and testing sets, which hurts the generalization ability of CNNs. Data augmentation techniques, like random cropping and affine transformation, are one way to improve the stability of networks in unfamiliar domains. Effective use of synthetic data [15], [39], [40] is another choice to achieve the same goal. It has been used in a lot of computer vision tasks such as crowd counting [15], semantic segmentation [39], person re-identification [40], *etc.* One common way of reducing dataset distribution bias is to make the synthetic data much more photo-realistic and minimize the domain shift at the same time. Some domain adaptation methods [41], [42], [43] can be used to learn the domain-invariant features to align the domains of synthetic and real data, so the model generalization ability can be improved. In this paper, we utilize pairs of synthetic nighttime and real daytime images based on style-transfer method for training a detail-preserving night-to-day network. Then, we adapt our trained translation network to transfer any nighttime image to daytime version so as to reuse the trained daytime detection model.

Style transfer: Many computer vision tasks need to translate an input image from one domain to another domain, which are viewed as the image translation problem. Generative adversarial networks (GANs) based methods are promising for image stylization, which aim to sample from a probability distribution to generate images. GANs include two models: a generative model and a discriminative model. The former captures the critical data distribution for image generation, while the latter aims to distinguish between real and generated samples. CycleGAN [11] extends the GAN-based image translation method to an unsupervised framework, where no paired data is required. It performs a full translation cycle both from source domain to target domain and from target domain back to source domain, and then it could regularize the high cycle consistency. Subsequent works encourage a shared latent feature space via a variational autoencoder (UNIT [12]). ComboGAN [44] and SMIT [45] extend to multiple domain translation. GcGAN [46] proposed that the translation network should keep geometry consistency.

Although unpaired image translation by GAN-based methods are popular for style transfer, the generated images might lack details due to the common existing downsampling and upsampling network operations. In this paper, we propose to train a detail-persevering network to achieve the unpaired domain transfer for nighttime object detection.

III. METHODOLOGY

In this section, we propose the *detail-preserving unpaired domain transfer* for performing high accuracy object detection in the nighttime without retraining the detectors on daytime dataset. We introduce the whole framework in Sec. III-A and reveal the challenges. Then, our two main contributions, *i.e.*, *scene-aware pixel-wise filtering* in Sec. III-B and *StyleMix* in Sec. III-C, help to address the challenges and achieve much better detection accuracy.

A. Detail-preserving Unpaired Domain Transfer for Nighttime Object Detection

We propose to perform nighttime object detection by transferring input nighttime images to the corresponding daytime versions for further object detection. This task could be simply formulated as

$$\hat{\mathbf{I}} = \phi(\mathbf{I}), \quad (1)$$

where the $\phi(\cdot)$ denotes a transfer function that can map the nighttime image \mathbf{I} to the corresponding daytime version. A straightforward way is to set $\phi(\cdot)$ as a popular generator that can be trained with the adversarial loss. Nevertheless, we argue that GAN-based transfer is hard to recover the details in the nighttime image, which is rather important for accurate object detection. As shown in Figure 2, the GAN-based method might destroy the detailed car structure, leading to missing detection. Actually, the night-to-day translation for object detection requires that the object-related details, *e.g.*, car's structure, should be preserved while different scene patterns in the night should be perceived and properly mapped to its daytime versions, posing two challenges for deep learning-based solutions: ① Popular deep generator based methods easily harm the object details due to the common existing bottleneck layers where the input image is transferred by downsampling and upsampling. ② It is hard to get paired dataset which is significantly important for training detail-persevering networks with pixel-wise correspondence.

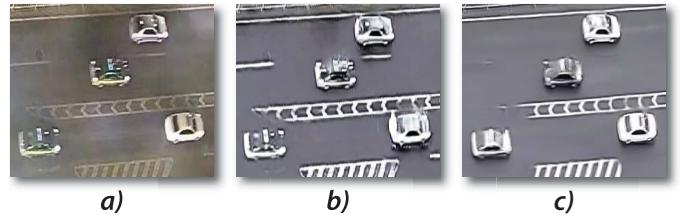


Figure 2: Image translation results of GAN-based method and the proposed method: a) target nighttime image, b) translated daytime image of a) by GAN-based method *CycleGAN* [11], c) translated daytime image of a) by the proposed method.

To address the first issue, we propose the *scene-aware pixel-wise filtering* in Sec. III-B for night-to-day transformation. Different to all existing works that employ a DNN as the transformer directly, our method maps the input image through a single-layer filtering whose kernels are predicted by an offline-trained DNN denoted as the *kernel prediction network*. Note that, the single-layer filtering (without any downsampling and upsampling operations) avoids the risk of missing important object-related details. Meanwhile, the DNN helps understand the scene and predict spatial-variant kernels for effective transformation. Specifically, kernel prediction network predicts a kernel for each pixel to capture local spatial context information to preserve more details, *e.g.*, structure information. Recent works [47], [48], [49], [50] have proved that per-pixel kernel prediction network can achieve image recovery with better details. To address the second challenge, we propose a style-transfer-based data augmentation method, *i.e.*, *StyleMix*

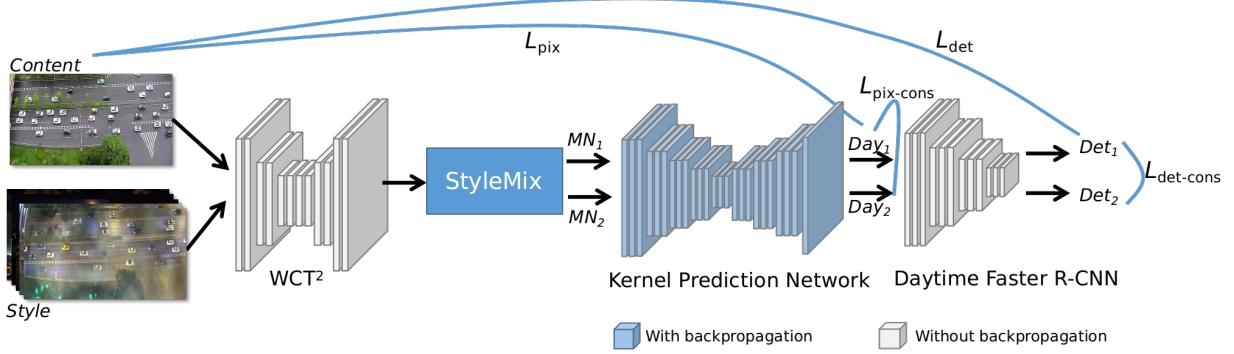


Figure 3: The proposed object detection pipeline at night with Night-to-Day image translation.

in Sec. III-C, to generate nighttime-daytime image pairs for training the kernel prediction network.

We show the whole framework in Figure 3. Intuitively, our method is a pre-process module transferring the input nighttime image to daytime version for further object detection, which is supported by a novel and simple data augmentation method, *i.e.*, StyleMix.

B. Scene-aware Pixel-wise Filtering

We propose the scene-aware pixel-wise filtering for the night-to-day transformation. Specifically, we reformulate Eq. (1) as

$$\hat{\mathbf{I}} = \mathbf{K} \circledast \mathbf{I}, \quad (2)$$

$$\text{with } \mathbf{K} = \phi(\mathbf{I}), \quad (3)$$

where \circledast denotes the pixel-wise filtering, \mathbf{K} is a pixel-wise filter $\in \mathbb{R}^{(k \times k) \times h \times w}$. Each vector in channel dimension $\mathbf{K}(i, j) \in \mathbb{R}^{(k \times k)}$ is a per-pixel kernel and can be applied to the $k \times k$ neighborhood region of each pixel in the input nighttime image \mathbf{I} by element-wise multiplication. The $\phi(\cdot)$ denotes the kernel prediction network and is used to perceive the input image and predict the suitable kernel for each pixel.

Then, we acquire the daytime version $\hat{\mathbf{I}}$ of the input image. Since it is pixel-wise filtering of the input nighttime image, it could largely preserve the image details without corruption. To fully leverage rich neighborhood information of every image pixel, a large kernel size k is desired, however, the computational and memory cost will increase as well. The kernel size k in our implementation is set to 5.

The framework of kernel prediction network is shown in Figure 4. In this work, the training input data for KPN is synthetic nighttime images from Sec. III-C. Specifically, two synthetic Mixed Nighttime images MN_1 and MN_2 with different style conditions are fed into KPN, respectively. KPN will output image-specific per-pixel filter for each image, respectively. Then element-wise multiplying the specific filter with the corresponding input image will generate the daytime version $\hat{\mathbf{I}}_i$, $i = 1, 2$.

The basic loss function $\mathcal{L}_{pix}(\hat{\mathbf{I}}_i, \hat{\mathbf{I}}^*)$ is the pixel-wise L_1 distance between the ground truth daytime image $\hat{\mathbf{I}}^*$ and the translated daytime image $\hat{\mathbf{I}}_i$. It is defined as

$$\mathcal{L}_{pix}(\hat{\mathbf{I}}_i, \hat{\mathbf{I}}^*) = \|\hat{\mathbf{I}}^* - \hat{\mathbf{I}}_i\|_1. \quad (4)$$

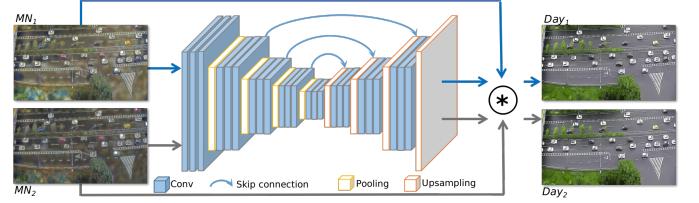


Figure 4: Illustration of the kernel prediction network based scene-aware pixel-wise filtering.

We also define a consistency loss $\mathcal{L}_{pix\text{-cons}}$ between $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ by measuring their L_1 distance. The equation is

$$\mathcal{L}_{pix\text{-cons}}(\hat{\mathbf{I}}_1, \hat{\mathbf{I}}_2) = \|\hat{\mathbf{I}}_1 - \hat{\mathbf{I}}_2\|_1. \quad (5)$$

C. StyleMix: Bridging the Gap to Nighttime Data

The style-transfer-based method is utilized to generate nighttime-daytime image pairs for KPN training. To bridge the shift of synthetic nighttime and target nighttime data, we propose the StyleMix strategy to embody the diversity of nighttime scenarios.

Specifically, style-transfer network can preserve the structure of input content image and stylize the content image according to the input style reference to implement image translation. As shown in Figure 3, we adopt a pretrained style transfer network, whitening and coloring transforms (WCT^2), to finish daytime to nighttime image translation. For the input of WCT^2 , daytime images are the content image and five real nighttime images act as the style reference. Five style reference images for following StyleMix are selected depending on the illumination condition of target nighttime scenarios. During daytime to nighttime image translation, StyleMix is involved to reduce the distribution shift of translated style and target nighttime style. It works in the way shown in Figure 5. Specifically, for each daytime input image, three style augmentation chains out of five style reference are randomly sampled, each of which consists of one to two randomly selected style transfer operations. Then the transferred images from these style augmentation chains are combined by pixel-level fusion to acquire mixed nighttime image MN_i in Sec. III-B. Pixel-wise fusion is implemented by pixel-wise convex operations between translated nighttime images and convex coefficients. We randomly sample from a Dirichlet (α, \dots, α) distribution

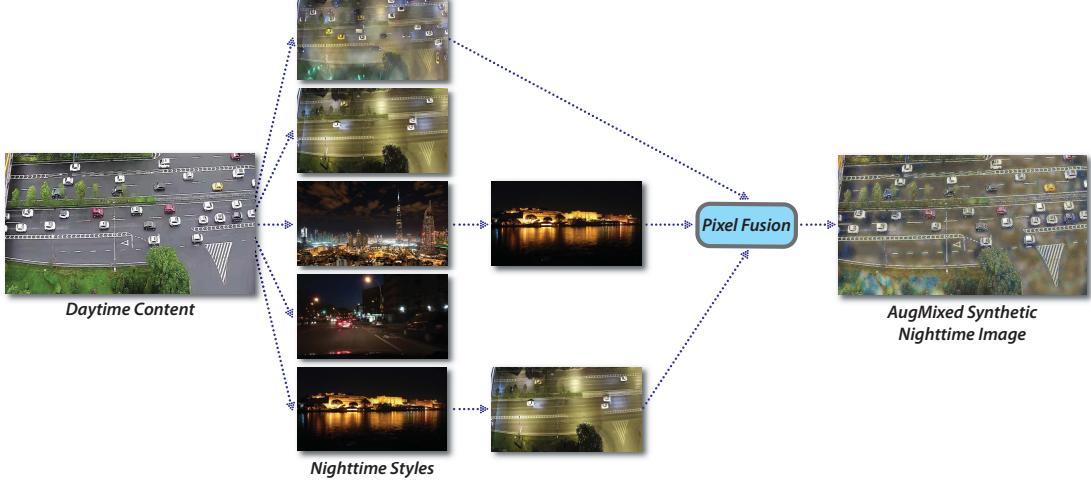


Figure 5: Illustration of the proposed StyleMix method to bridge the gap to nighttime data.

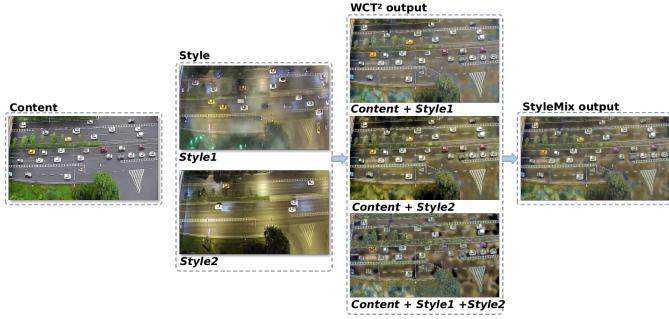


Figure 6: Sample visualization of the proposed StyleMix to generate synthetic nighttime images from real daytime images.

to construct the 3-dimensional vector of pixel-wise convex coefficients. Figure 6 shows one example of the pixel-wise fused synthetic nighttime image by StyleMix. The augmixed output of StyleMix are the pixel-wise fusion result of translated content image with different styles. It shows that StyleMix can effectively generate various kinds of synthetic nighttime images which are visually close to the real nighttime scene.

For detection task of the pipeline, we construct the detection loss $\mathcal{L}_{\text{det}}(Det_i, Det^*)$ and the detection consistency loss $\mathcal{L}_{\text{det-cons}}(Det_1, Det_2)$. We adopt Smooth L_1 loss [51] to calculate \mathcal{L}_{det} and $\mathcal{L}_{\text{det-cons}}$. The total loss \mathcal{L}_{N2D} of the pipeline is a weighted sum of \mathcal{L}_{pix} , $\mathcal{L}_{\text{pix-cons}}$, \mathcal{L}_{det} , and $\mathcal{L}_{\text{det-cons}}$. It is defined as

$$\mathcal{L}_{\text{N2D}} = \mathcal{L}_{\text{pix}} + \mathcal{L}_{\text{pix-cons}} + \lambda(\mathcal{L}_{\text{det}} + \mathcal{L}_{\text{det-cons}}), \quad (6)$$

where λ is set to 10 in our experiments.

IV. EXPERIMENTS

A. Datasets

In this paper, the public D&N-Car Benchmark [10] is utilized to verify the effectiveness of the proposed approach. It is a real traffic surveillance dataset in urban expressway scene recorded in the city Xi'an, China. This dataset includes 1,200 daytime images and 1,000 nighttime images with their ground truth

in the format of bounding boxes across different periods and dates, each of which is with resolution $1,280 \times 720$. There are total 57,059 vehicle instances in this dataset. The training set consists of 1,000 daytime traffic images with manual ground-truth labels, denoted as **Day-training**. The testing set includes 1,200 images, where 200 images are in daytime and 1,000 images are in nighttime. In the 200 daytime testing images, 100 images are in the normal traffic condition, denoted as **Day-normal**, and the other 100 images are in the congested traffic condition, denoted as **Day-congested**. The left 1,000 images of testing set constitute 4 subsets of nighttime traffic images (denoted as **Night1**, **Night2**, **Night3**, **Night4**). The details of the benchmark are shown in Table I. In the experiment, we denote the labeled daytime traffic images (Day-training) as the Source Domain **S**, and the unlabeled nighttime traffic images as the Target Domain **T**.

Table I: Details of the D&N-Car Benchmark [10].

	No. of images	No. of car instances	Time
Day-training	1000	32,456	19:10
Day-normal	100	3,173	19:00
Day-congested	100	4,539	14:30
Night1	250	7,322	21:30
Night2	250	5,554	21:30
Night3	250	1,738	23:50
Night4	250	2,277	00:20

B. Experimental Setting

We conduct experiments on two different scenarios: 1). Detect the vehicles during daytime by Faster R-CNN [22] model trained on **Day-training**; 2). Detect the vehicles during nighttime by trained Faster R-CNN model on **Day-training** after the proposed night-to-day image translation. The detailed experimental setting is as follows: 1) Scenario 1: We directly train a Faster R-CNN model on the dataset **Day-training** in a supervised way and test the images on **Day-normal** and **Day-congested**, respectively; 2) Scenario 2: For style-transfer-based StyleMix image translation aiming at acquiring pairs

of daytime and nighttime images, we utilize 1,000 images in **Day-training** set and 5 style reference images for nighttime images synthesis and augmixing styles. For KPN-based night-to-day training, there are 2,000 augmixed nighttime images for training in each epoch. Next, predicted daytime images are fed into detection task to further fit the translated daytime image for object detection. For inference, the trained KPN operates image translation for real nighttime images (**Night1**, **Night2**, **Night3**, and **Night4**), and then trained daytime detection model tests on translated nighttime images for performance evaluation.

We set the method that directly tests on nighttime images with trained daytime model Faster R-CNN [22] as a baseline. We also compare the proposed method with unpaired image translation methods UNIT [12], CycleGAN [11], and GcGAN [46] combining with Faster R-CNN in both day-to-night and night-to-day directions. To train the image translation models, the training dataset for daytime is the **Day-training** set and the training set for nighttime is a combination of **Night1**, **Night2**, **Night3**, and **Night4**.

We built our translation and detection pipeline in PyTorch. For object detection, we use ResNet50 as our backbone. For detection training, we utilize Stochastic Gradient Descent (SGD) to optimize our network and set the initial learning rate to 0.0001 and decay it after every 10 epochs. The experiments are conducted on a NVIDIA GTX 1080Ti GPU. For night-to-day image translation training, we train KPN with SGD by setting the learning rate to 0.002 for 200 epochs on two Tesla V100 GPUs. For a comprehensive performance evaluation, the widely-used object detection metric mAP (mean average precision) is used for evaluating the vehicle detection results. For all the experiments, the performance evaluation uses a uniform threshold of 0.5 for the Intersection Over Union (IoU) between the predicted bounding boxes and ground truth.

C. Results on Benchmark

We first report the detection results of one-stage detector SSD [19] and two-stage detector Faster R-CNN [22] for Scenario 1, shown in Table II. We can see that both of the mAP drop from about 99% to 88% when the traffic is congested. The congested situation increases the object detection difficulty, resulting in a lower detection performance compared to the uncrowded situation. Because there is not a clear difference in terms of mAP between SSD and Faster R-CNN, we choose Faster R-CNN as our baseline detector for following experiments.

Table II: Daytime vehicle detection results.

mAP(%)	SSD	Faster R-CNN
Day-normal	99.05	99.01
Day-congested	88.35	88.57

D. Results compared to day-to-night translation methods

We compare the detection results of nighttime vehicle to other image translation methods in a day-to-night direction. According to Scenario 2, the proposed method performs vehicle detection on translated daytime images obtained from KPN

by the daytime model obtained from Sec. IV-C. However, for comparison methods performing nighttime vehicle detection in a day-to-night direction, they require additionally training a nighttime model for nighttime vehicle detection, besides the daytime model. For example, taking CycleGAN as the day-to-night image translation method, we translate daytime images to fake/synthetic nighttime images in an unpaired way, and followed by training a Faster R-CNN_n detector on such fake/synthetic nighttime images with the same annotations of daytime images. Then we test the trained model on the nighttime images for vehicle detection. The comparison results are shown in Table III. We compare the detection results in the form of mAP for each subset of nighttime traffic images and the mean mAP for all of them. Day-to-night image translation methods UNIT [12], CycleGAN [11], and GcGAN [46] perform better than or comparable to the baseline Faster R-CNN which directly tests on nighttime images with daytime model of Sec. IV-C. Taking the dataset **Night4** as an example, the proposed method, based on night-to-day image translation without retraining one more model, achieves the highest 92.94% mAP, about 5.4% higher than Faster R-CNN_n + GcGAN_{d2n}, 3.3% higher than Faster R-CNN_n + CycleGAN_{d2n}, 4.8% higher than Faster R-CNN_n + UNIT_{d2n} and 5.9% higher than the baseline Faster R-CNN. The proposed method achieves the best mean mAP with 87.80% for all nighttime traffic images, despite that Faster R-CNN_n + GcGAN_{d2n} performs a little bit better on the **Night1** subset. We also provide a traditional method Mean-BGS [52] performing vehicle detection through background subtraction and the daytime model of SSD [19] performing vehicle detection directly on nighttime images. Both of them are worse than Faster R-CNN for nighttime vehicle detection. As the corresponding detection results are shown in Fig. 7, we can clearly see that the proposed method is robust to various light conditions. UNIT, CycleGAN and GcGAN based methods could not well detect vehicles under poor light conditions and missed many black vehicles compared to the proposed method, and Faster R-CNN without any image translation does not perform well due to the domain shift of daytime and nighttime scenarios.

Table III: Nighttime vehicle detection results based on day-to-night translation. Note that the Faster R-CNN_n model is trained on the fake/synthetic nighttime images.

Method \ mAP(%)	Night1	Night2	Night3	Night4	Mean
Mean-BGS [52]	54.03	49.09	52.16	55.56	52.71
SSD [19]	74.06	73.78	84.02	87.00	79.71
Faster R-CNN [22]	74.84	74.05	85.63	87.05	80.39
Faster R-CNN _n [22]+ UNIT _{d2n} [12]	70.56	77.13	82.87	88.19	79.68
Faster R-CNN _n [22]+ CycleGAN _{d2n} [11]	79.39	80.72	88.72	89.66	84.62
Faster R-CNN _n [22]+ GcGAN _{d2n} [46]	80.89	84.20	83.92	87.55	84.14
Proposed	80.25	84.81	93.20	92.94	87.80

E. Results compared to night-to-day translation methods

We also compare our method with these image translation methods in a night-to-day direction. Comparison translation methods, UNIT_{n2d} [12], CycleGAN_{n2d} [11], and GcGAN_{n2d} [46], first translate nighttime images to daytime-style images, and then these daytime-style images are fed into the daytime model, obtained from Sec. IV-C, for vehicle

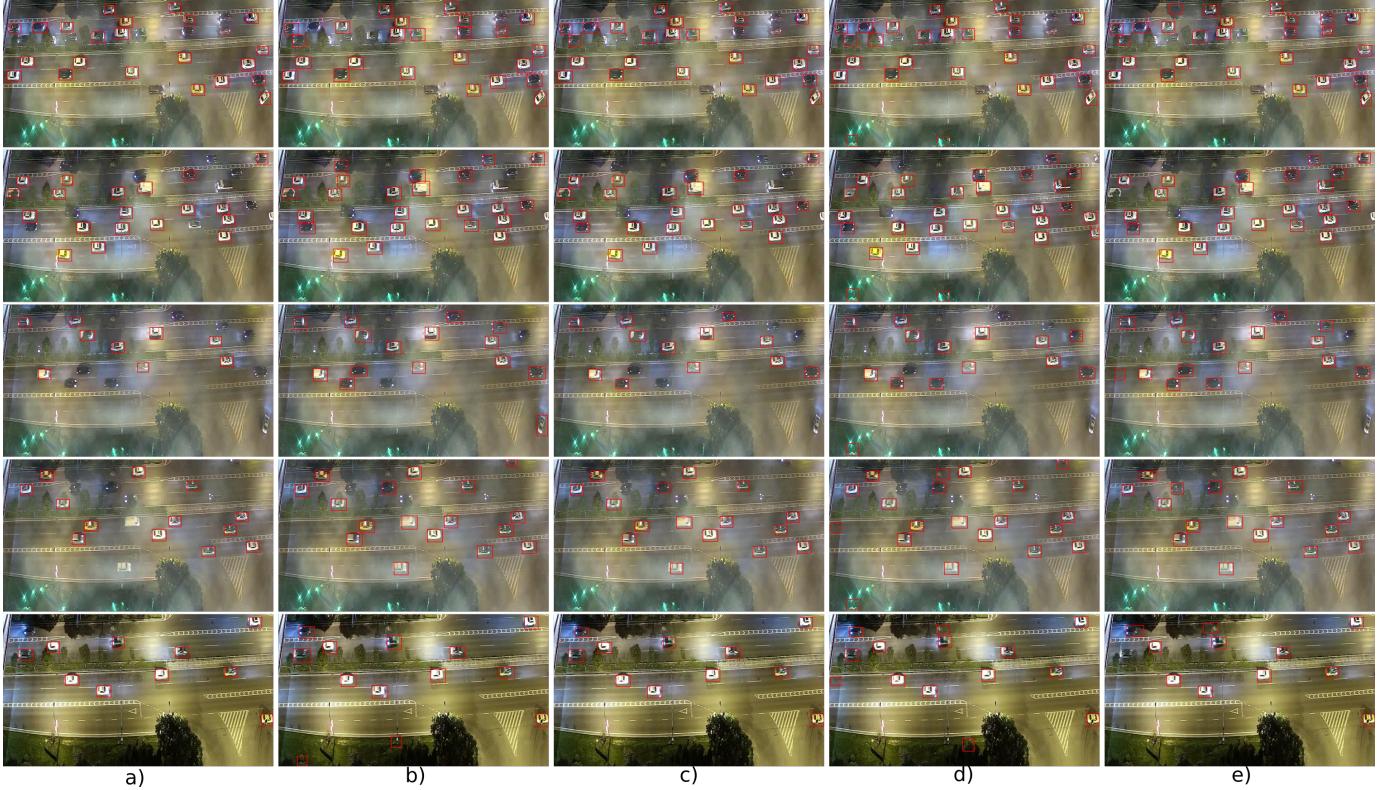


Figure 7: Visualization results of nighttime vehicle detection. a)-e) are the detection results from Faster R-CNN [22], Faster R-CNN_n + UNIT_{d2n} [12], Faster R-CNN_n + CycleGAN_{d2n} [11], Faster R-CNN_n + GcGAN_{d2n} [46], and the proposed method, respectively. Note: red bounding box indicates detection result.



Figure 8: Visualization results of image translation from nighttime to daytime. a) is target nighttime image, b) to e) are the image translation results of UNIT_{n2d} [12], CycleGAN_{n2d} [11], GcGAN_{n2d} [46], and proposed method, respectively.

detection. The results are shown in Table IV. It shows that the proposed method could achieve the best mean mAP performance than UNIT, CycleGAN, and GcGAN for nighttime vehicle detection via night-to-day translation, demonstrating the advances of the proposed method. This is because the proposed method considers the per-pixel kernel fusion of neighboring information for each pixel and the object detection task during image translation training, preserving more features, *e.g.*, structure details, which are critical for the detection task.

Table IV: Nighttime vehicle detection based on night-to-day translation.

Method \ mAP(%)	Night1	Night2	Night3	Night4	Mean
Mean-BGS [52]	54.03	49.09	52.16	55.56	52.71
SSD [19]	74.06	73.78	84.02	87.00	79.71
Faster R-CNN [22]	74.84	74.05	85.63	87.05	80.39
Faster R-CNN [22]+ UNIT _{nd} [12]	62.54	62.86	81.29	81.00	71.92
Faster R-CNN [22]+ CycleGAN _{nd} [11]	77.69	79.12	88.31	88.79	83.47
Faster R-CNN [22]+ GcGAN _{nd} [46]	83.57	83.80	79.26	83.94	82.64
Proposed	80.25	84.81	93.20	92.94	87.80

Visualization results by the above mentioned image translation methods from nighttime to daytime are shown in Fig. 8. It shows that the proposed method could recover the daytime scenario with details. UNIT method suffers from model collapse presenting poor local textures and details. Specifically, the translated images are blurred, especially for the shape and edge of vehicles inside the image. It is consist with the lower detection performance in the form of mAP in Table IV. CycleGAN could translate the texture of the vehicles from nighttime to daytime, but it is not robust to the intense road mirror reflections. It presents fake vehicles which do not exist in source nighttime images, resulting in more false positive detection samples. More black vehicles disappeared in the translated images by CycleGAN. GcGAN is also sensitive to such intense road mirror reflections, resulting in more fake vehicles in the translated images. Although the translated trees from the proposed method suffer from corruption, the goal of our work is to accurately detect vehicles at nighttime, we do not care much more about the tree corruption during the night-to-day translation. It is obvious that the translated cars are more natural with clear structures from the proposed method, resulting in better detection performance in Table IV. This is because the image translation training from nighttime to daytime makes full use of paired synthetic data with pixel-wise correspondence and per-pixel kernel fusion of neighboring information which provides rich spatial context information.

F. Ablation Study

Table V: Ablation study of the proposed method for the nighttime vehicle detection.

Method \ mAP(%)	Night1	Night2	Night3	Night4
Baseline	62.51	60.97	81.29	78.66
Baseline + StyleMix	75.90	76.20	88.20	86.90
Baseline + StyleMix + Zero	79.50	77.40	89.70	89.80
Baseline + StyleMix + Contrast	80.20	81.79	92.11	91.58

In this section, we evaluate the contribution of each step in the proposed method: 1) training KPN without the StyleMix,

instead, given each daytime image, we randomly select two style images from five style reference images to generate two synthetic nighttime images based on image translation, respectively. Two synthetic nighttime images are fed into KPN for training combined with detection task. We view this method as our *Baseline* in this section. 2) On the basis of step 1, we augment the style references in depth and width, denoted as *Baseline + StyleMix*. There are five style reference images for StyleMix, out of which two are from the nighttime traffic images of D&N-Car dataset. 3) For testing phase, we preprocess the target nighttime images by Zero-Reference network [53] to improve local contrast, then go through KPN for image translation to daytime scenario and followed by detection task via daytime detection model, denoted as *Baseline + StyleMix + Zero*. 4) Different with the preprocessing of step 4, we enhance the local contrast by improving the pixel value less than a threshold, denoted as *Baseline + StyleMix + Contrast*. The corresponding result of the proposed method in each step on the four night subsets in terms of mAP evaluation metric is shown in Table V.

We can clearly see the positive effect of each step with the increasing mAP performance. Taking dataset **Night1** for an example, the baseline method could achieve 62.51% in mAP. When augmenting and mixing the style reference images to embody the diversity of synthetic nighttime scenarios, mAP increases by about 13%. When prepossessing the target nighttime images with Zero-Reference network and Contrast, the mAP continues to increase to 79.50% and 80.20%, respectively.

We conduct ablation experiments to verify the effectiveness of the style reference setting for StyleMix. We construct a night-style image pool for style reference selection, which consists of 21 images, 7 from nighttime dataset of D&N-Car, *i.e.*, nighttime traffic images used in this paper, 7 from BDD dataset [54] with nighttime scene and the other 7 from WCT² publicized project website¹. We randomly choose 5 images from this night-style image pool as style reference for the whole image translation model. We conduct three experiments with different style reference setting for StyleMix. There are 1, 2, and 5 images from the nighttime images of D&N-Car dataset as different experiment settings: *StyleMix*₁, *StyleMix*₂ and *StyleMix*₅. The detection results are shown in Table VI. It turns out that the detection performance increases with more night-style images involved from the D&N-Car dataset. It is reasonable since we expect the StyleMix model to render the synthetic nighttime image closer to the corresponding nighttime style of nighttime images.

Table VI: Ablation study with different style reference setting.

Method \ mAP(%)	Night1	Night2	Night3	Night4	Mean
Baseline + <i>StyleMix</i> ₁ + Contrast	76.27	79.51	91.76	91.69	84.80
Baseline + <i>StyleMix</i> ₂ + Contrast	80.20	81.79	92.11	91.58	86.42
Baseline + <i>StyleMix</i>₅ + Contrast	80.25	84.81	93.20	92.94	87.80

V. CONCLUSIONS

In this paper, we proposed a detail-preserving method to implement the nighttime to daytime image translation and thus

¹<https://github.com/clovaai/WCT2>

adapt daytime trained detection model to nighttime detection. We firstly utilize style translation method to acquire paired images of daytime and nighttime, which are hard to obtain in real-world applications. We propose to stylemix the reference styles to embody the diversity of synthetic nighttime scenarios. The following nighttime to daytime translation is implemented based on kernel prediction network to avoid texture corruption and trained with detection task to make the translated daytime image not only visually photo-realistic to the daytime scenario but also fit the detection task to reuse the daytime domain knowledge. The proposed method can perform both daytime and nighttime vehicle detection with one model. Experimental results showed that the proposed method achieved effective and accurate nighttime detection results.

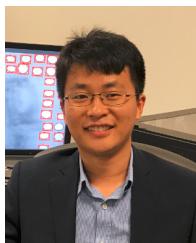
REFERENCES

- [1] Y. Zhang, X. Liang, D. Zhang, M. Tan, and E. P. Xing, “Unsupervised object-level video summarization with online motion auto-encoder,” *Pattern Recognition Letters*, vol. 130, pp. 376–385, 2020.
- [2] D. Zhang, J. Han, L. Yang, and D. Xu, “Spftm: a joint learning framework for localizing and segmenting objects in weakly labeled videos,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [3] L. Yang, J. Han, D. Zhang, N. Liu, and D. Zhang, “Segmentation in weakly labeled videos via a semantic ranking and optical warping network,” *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4025–4037, 2018.
- [4] D. Zhang, G. Guo, D. Huang, and J. Han, “Poseflow: A deep motion representation for understanding human behaviors in videos,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6762–6770.
- [5] J. Han, L. Yang, D. Zhang, X. Chang, and X. Liang, “Reinforcement cutting-agent learning for video object segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9080–9089.
- [6] H. Yu, D. Guo, Z. Yan, L. Fu, J. Simmons, C. P. Przybyla, and S. Wang, “Weakly supervised easy-to-hard learning for object detection in image sequences,” *Neurocomputing*, vol. 398, pp. 71–82, 2020.
- [7] S. Li, H. Yu, J. Zhang, K. Yang, and R. Bin, “Video-based traffic data collection system for multiple vehicle types,” *IET Intelligent Transport Systems*, vol. 8, no. 2, pp. 164–174, 2014.
- [8] R. Ke, Z. Li, J. Tang, Z. Pan, and Y. Wang, “Real-time traffic flow parameter estimation from uav video based on ensemble classifier and optical flow,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 54–64, 2018.
- [9] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, “High-resolution vehicle trajectory extraction and denoising from aerial videos,” *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [10] J. Li, Z. Xu, L. Fu, X. Zhou, and H. Yu, “Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework,” *Transportation Research Part C: Emerging Technologies*, vol. 124, p. 102946, 2021.
- [11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [12] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in *Advances in Neural Information Processing Systems*, 2017, pp. 700–708.
- [13] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, “Contrastive learning for unpaired image-to-image translation,” in *European Conference on Computer Vision*. Springer, 2020, pp. 319–345.
- [14] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in *European Conference on Computer Vision*, 2018, pp. 172–189.
- [15] Q. Wang, J. Gao, W. Lin, and Y. Yuan, “Learning from synthetic data for crowd counting in the wild,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8198–8207.
- [16] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, “Semantic annotation of high-resolution satellite images via weakly supervised learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660–3671, 2016.
- [17] J. Han, X. Yao, G. Cheng, X. Feng, and D. Xu, “P-cnn: Part-based convolutional neural networks for fine-grained visual categorization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [18] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, “Augmix: A simple data processing method to improve robustness and uncertainty,” in *International Conference on Learning Representations*, 2019.
- [19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *European Conference on Computer Vision*. Springer, 2016, pp. 21–37.
- [20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *IEEE International Conference on Computer Vision*, 2016, pp. 779–788.
- [21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- [22] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.
- [23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.
- [24] F. Xu, X. Liu, and K. Fujimura, “Pedestrian detection and tracking with night vision,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 63–71, 2005.
- [25] J. Ge, Y. Luo, and G. Tei, “Real-time pedestrian detection and tracking at nighttime for driver-assistance systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 283–298, 2009.
- [26] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, “Kaist multi-spectral day/night data set for autonomous and assisted driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.
- [27] H. Kuang, K.-F. Yang, L. Chen, Y.-J. Li, L. L. H. Chan, and H. Yan, “Bayes saliency-based object proposal generator for nighttime traffic images,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 814–825, 2017.
- [28] R. K. Satzoda and M. M. Trivedi, “Looking at vehicles in the night: Detection and dynamics of rear lights,” *IEEE Transactions on Intelligent Transportation Systems*, 2016.
- [29] J. M. Á. Alvarez and A. M. López, “Road detection based on illuminant invariance,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2010.
- [30] G. Ros and J. M. Alvarez, “Unsupervised image transformation for outdoor semantic labelling,” in *IEEE Intelligent Vehicles Symposium*. IEEE, 2015, pp. 537–542.
- [31] A. Valada, J. Vertens, A. Dhall, and W. Burgard, “Adapnet: Adaptive semantic segmentation in adverse environmental conditions,” in *IEEE International Conference on Robotics and Automation*. IEEE, 2017, pp. 4644–4651.
- [32] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, “Night-to-day image translation for retrieval-based localization,” in *IEEE International Conference on Robotics and Automation*. IEEE, 2019, pp. 5958–5964.
- [33] B. Tian, F. Juefei-Xu, Q. Guo*, W. Chan, Y. Cheng, X. Li, X. Xie, and S. Qin, “Bias field poses a threat to dnn-based x-ray recognition,” in *IEEE International Conference on Multimedia and Expo*, 2021.
- [34] Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, J. Wang, B. Yu, W. Feng, and Y. Liu, “Watch out! Motion is Blurring the Vision of Your Deep Neural Networks,” in *Advances in Neural Information Processing Systems*, 2020.
- [35] Q. Guo, W. Feng, R. Gao, Y. Liu, and S. Wang, “Exploring the effects of blur and deblurring to visual object tracking,” *IEEE Transactions on Image Processing*, vol. 30, pp. 1812–1824, 2021.
- [36] Q. Guo, R. Han, W. Feng, Z. Chen, and L. Wan, “Selective spatial regularization by reinforcement learned decision making for object tracking,” *IEEE Transactions on Image Processing*, vol. 29, pp. 2999–3013, 2020.
- [37] Q. Guo, X. Xie, F. Juefei-Xu, L. Ma, Z. Li, W. Xue, W. Feng, and Y. Liu, “Spark: Spatial-aware online incremental attack against visual tracking,” in *European Conference on Computer Vision*, vol. 2, 2020.
- [38] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, “Learning dynamic siamese network for visual object tracking,” in *IEEE International Conference on Computer Vision*, 2017, pp. 1763–1771.
- [39] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, “Fully convolutional adaptation networks for semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6810–6818.
- [40] S. Bak, P. Carr, and J.-F. Lalonde, “Domain adaptation through synthesis for unsupervised person re-identification,” in *European Conference on Computer Vision*, 2018, pp. 189–205.
- [41] P. Zhang, B. Zhang, D. Chen, L. Yuan, and F. Wen, “Cross-domain correspondence learning for exemplar-based image translation,” in *IEEE*

- Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5143–5153.
- [42] M. Kim and H. Byun, “Learning texture invariant representation for domain adaptation of semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 975–12 984.
- [43] Y. Zheng, D. Huang, S. Liu, and Y. Wang, “Cross-domain object detection through coarse-to-fine feature adaptation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 766–13 775.
- [44] A. Anoosheh, E. Agustsson, R. Timofte, and L. Van Gool, “ComboGAN: Unrestrained scalability for image domain translation,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 783–790.
- [45] A. Romero, P. Arbeláez, L. Van Gool, and R. Timofte, “Smit: Stochastic multi-label image-to-image translation,” in *IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [46] H. Fu, M. Gong, C. Wang, K. Batmanghelich, K. Zhang, and D. Tao, “Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2427–2436.
- [47] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, “Toward real-world single image super-resolution: A new benchmark and a new model,” in *IEEE International Conference on Computer Vision*, 2019, pp. 3086–3095.
- [48] B. Mildenhall, J. T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll, “Burst denoising with kernel prediction networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2502–2510.
- [49] Q. Guo, J. Sun, F. Juefei-Xu, L. Ma, X. Xie, W. Feng, and Y. Liu, “Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining,” in *AAAI Conference on Artificial Intelligence*, 2021.
- [50] L. Fu, C. Zhou, Q. Guo, F. Juefei-Xu, H. Yu, W. Feng, Y. Liu, and S. Wang, “Auto-exposure fusion for single-image shadow removal,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [51] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
- [52] S. Li, H. Yu, J. Zhang, K. Yang, and R. Bin, “Video-based traffic data collection system for multiple vehicle types,” *IET Intelligent Transport Systems*, vol. 8, no. 2, pp. 164–174, 2013.
- [53] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, “Zero-reference deep curve estimation for low-light image enhancement,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.
- [54] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “Bdd100k: A diverse driving dataset for heterogeneous multitask learning,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2636–2645.



Lan Fu is pursuing her Ph.D degree in the department of computer science and engineering at University of South Carolina, Columbia, SC, USA. Prior to that, she received the M.S. degree in Biomedical Engineering from Tianjin University, Tianjin, China. Her research interests include computer vision and deep learning and mainly focus on domain adaptation based object detection and image enhancement.



Hongkai Yu received the Ph.D. degree in computer science and engineering from University of South Carolina, Columbia, SC, USA, in 2018. He is currently an Assistant Professor in the Department of Electrical Engineering and Computer Science at Cleveland State University, Cleveland, OH, USA. His research interests include computer vision, machine learning, deep learning and intelligent transportation system.



Felix Juefei-Xu received the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University (CMU), Pittsburgh, PA, USA. Prior to that, he received the M.S. degree in Electrical and Computer Engineering and the M.S. degree in Machine Learning from CMU, and the B.S. degree in Electronic Engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China. Currently, he is a Research Scientist with the Alibaba Group, Sunnyvale, CA, USA, with research focus on a fuller understanding of deep learning where he is actively involved in deep learning that are statistically efficient and adversarially robust. He also has broader interests in pattern recognition, computer vision, machine learning, optimization, statistics, compressive sensing, and image processing. He is the recipient of multiple best/distinguished paper awards, including IJCB’11, BTAS’15-16, ASE’18, and ACCV’18.

exploring new methods

adversarially robust.

He also has broader interests in pattern recognition,

computer vision, machine learning, optimization,

statistics, compressive sensing,

and image processing.

He is the recipient of multiple best/distinguished paper

awards, including IJCB’11, BTAS’15-16, ASE’18, and ACCV’18.

Jinlong Li received the B.S. degree and M.S. degrees from Chang'an University, Xi'an, China in 2018 and 2021 respectively. His research interests include intelligent transportation system, computer vision, and deep learning.



Qing Guo received his B.S. degree in Electronic and Information Engineering from the North China Institute of Aerospace Engineering in 2011, M.E. degree in computer application technology from the College of Computer and Information Technology, China Three Gorges University in 2014, and the Ph.D. degree in computer application technology from the School of Computer Science and Technology, Tianjin University, China. He was a research fellow with the Nanyang Technology University, Singapore, from Dec. 2019 to Sep. 2020. He is currently a Wallenberg-NTU Presidential Postdoctoral Fellow with the Nanyang Technological University, Singapore. His research interests include computer vision, AI security, and image processing. He is a member of IEEE.

NTU Presidential Postdoctoral Fellow with the Nanyang Technological University, Singapore. His research interests include computer vision, AI security, and image processing. He is a member of IEEE.



Song Wang received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA, in 2002. He was a Research Assistant with the Image Formation and Processing Group, Beckman Institute, UIUC, from 1998 to 2002. In 2002, he joined the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA, where he is currently a Professor. His current research interests include computer vision, image processing, and machine learning. Dr. Wang is currently serving as the Publicity/Web Portal Chair of the Technical Committee of Pattern Analysis and Machine Intelligence of the IEEE Computer Society, an Associate Editor of IEEE Transaction on Pattern Analysis and Machine Intelligence, Pattern Recognition Letters, and Electronics Letters. He is a Senior Member of the IEEE and a member of the IEEE Computer Society.