Auxiliary Domain-guided Adaptive Object Detection in Adverse Weather Conditions

Zhuobin Fu[1], Kan Chang[1]✉, Mingyang Ling[2], Qingzhi Zhang[1], and Enze Qi[1]

[1] School of Computer and Electronic Information, Guangxi University, China
fzb@st.gxu.edu.cn, kanchang@gxu.edu.cn, {yizhi,enze7}@st.gxu.edu.cn
[2] School of Electrical Engineering, Guangxi University, China
lingmy@st.gxu.edu.cn

Abstract. To enhance detection accuracy in adverse weather conditions, domain adaptation methods that extract domain-invariant features from both the source and target domains have been present for one-stage detectors. However, the use of pseudo-labels in the instance-level domain adaptation inevitably introduces noise. To tackle this challenge, we propose an auxiliary domain-guided adaptive one-stage detection method. Firstly, a generative network is used to transform source domain images into the auxiliary domain. To form a suitable auxiliary domain that can provide reliable guidance for instance-level adaptation of the detector, the generated images are required to possess a similar style to that of the target domain, while also being restricted to maintaining the same object categories and location information as the source domain images. Secondly, for instance-level adaptation, we treat the same object from different domains as positive samples and different objects as negative samples, and utilize contrastive learning to ensure that only the domain shift, rather than other domain-irrelevant disparities, is reduced during the adaptation process. Experimental results demonstrate that our approach obtains a significant improvement over state-of-the-art (SOTA) algorithms on real datasets captured under adverse weather conditions.

Keywords: object detection · domain adaption · contrastive learning · convolutional neural network · adverse weather

1 Introduction

Recently, many convolutional neural networks (CNN)-based object detectors, such as Faster R-CNN [35] and YOLO [34], have achieved great success in object detection. However, these detection models are commonly trained under normal weather conditions. When applied in adverse weather conditions, such as low-light and foggy conditions, the performance of the normal object detectors could reduce significantly.

To improve the detection accuracy of detectors on images captured under adverse weather conditions, several methods utilize image enhancement techniques [41, 42, 12, 4, 28, 23, 38, 44] on the input images before performing detection. However, most image enhancement algorithms are designed for the human

visual system rather than machine vision, which might neglect the latent high-level features beneficial for high-level vision tasks [45, 33].

The second type of methods use the multi-task learning (MTL) strategy. For instance, Huang et al. [19] designed a framework to jointly learn visibility enhancement, object classification, and object localization in hazy conditions; besides object detection, the estimation of the low-light degradation parameters was additionally introduced by Cui et al. [7]. However, the MTL-based methods usually require high-quality images or precise degradation parameters for training, which are normally unavailable in real-world applications.

Some researchers jointly consider image enhancement and detection by cascading an image enhancement model with an object detector and training them in an end-to-end manner [29, 33, 24, 46, 48]. For instance, Liu et al. [29] proposed IA-YOLO, which uses a CNN-based parameter predictor to learn the hyper-parameters required by image processing operations; Qin et al. [33] used a Laplacian pyramid to decompose the image into different frequency components that are enhanced in a coarse-to-fine manner; Li et al. [24] proposed BAD-Net, where two parallel detector backbones are applied to extract features from both the input and enhanced images. Although the aforementioned methods have achieved high detection accuracy, obtaining a large amount of annotated data for adverse weather conditions remains a challenging task. Note that in the above references, synthetic datasets were employed for training. However, the simple degradation model employed in these datasets may significantly differ from the diverse and complex degradations encountered in real-world scenarios.

Unlike the above methods, the domain adaptation methods [24, 8, 30, 40] jointly use annotated data under normal weather conditions (known as source domain) and unannotated data under adverse weather conditions (known as target domain) to train domain-adaptive detectors, so that the backbone is able to extract domain-invariant features from the input. Recently, many domain-adaptive detection methods have been present [6, 36, 22, 26]. Thought domain adaptive detectors have achieved remarkable success, a majority of them rely on two-stage detectors, such as Faster R-CNN [35]. However, for real-world applications, these two-stage detectors may not be the optimal choice due to their relatively slow processing speed, which can hinder real-time performance.

Only a few domain adaptation methods have been proposed for one-stage detectors recently. Hnewa et al. [16] first proposed a cross-domain adaptation method for one-stage object detectors, where a gradient reversal layer (GRL) [10] followed by a domain discriminator is employed to multiple scales of features. As Hnewa et al. [16] only considered image-level adaptation, Zhang et al. [49] further introduced multi-scale instance alignment to reduce the domain shift at instance level. Nevertheless, in contrast to two-stage detectors, one-stage detectors, such as YOLO detectors [34], do not produce region proposals using a region proposal network (RPN) before predicting objects. To solve this problem, Zhang et al. [49] directly used the pseudo-labels generated by the YOLO detector and the ROI pooling operation to extract instance-level features, which inevitably introduces noise. During instance alignment, it also presents a challenge to minimize the

domain shift while simultaneously preserving other domain-irrelevant disparities, including variations in scales and appearance between two distinct objects.

To address the above limitations, this paper introduces a novel approach called Auxiliary Domain-guided Domain Adaptive object detection for the one-stage detector YOLO (AD-DAYOLO). To generate a high-quality auxiliary domain with similar adverse weather conditions to the target domain, we apply an unpaired image-to-image translation method [31] on the source domain images. The auxiliary domain shares the same labels as the source domain and overcomes the limitations associated with pseudo-labels. However, to well preserve the foreground information in source domain images, the ground-truth (GT) labels of source domain are used as an additional supervision for training the image-to-image translation model. After that, all the images from the source, target and auxiliary domains are fed to the AD-DAYOLO framework for training. In AD-DAYOLO, we perform image-level domain adaptation between the source and target domains, and also conduct instance-level domain adaptation between the source and auxiliary domains. During instance-level domain adaptation, we minimize the distance between corresponding objects in the source and auxiliary domains to facilitate the learning of domain-invariant features. On the contrary, when dealing with distinct objects observed in different domains, our objective is to maximize their distance from each other, so that domain-irrelevant disparities will be preserved and not diminished during the adaptation process. To this end, contrastive learning [13, 11] is introduced to the instance-level domain adaptation of AD-DAYOLO.

In summary, our contributions are threefold: 1) We generate an auxiliary domain by utilizing a regularized image-to-image translation model that incorporates GT labels from the source domain. This approach ensures that the auxiliary images not only capture similar weather conditions as the target domain but also possess identical labels as the source domain. 2) We introduce contrastive learning into the instance-level domain adaptation. This approach enables the instance alignment process to focus more on reducing domain shift, rather than domain-irrelevant disparities between instances. 3) By concurrently conducting image-level and instance-level adaptation, our AD-DAYOLO achieves a significant improvement compared to state-of-the-art (SOTA) methods on realistic datasets under various adverse weather conditions.

2 Related Work

2.1 Domain Translation Methods

Some methods have employed generative models (e.g., CycleGAN [50]) to synthesize target domain images for training. For instance, to produce pseudo-labels, Lang et al. [22] proposed training the detector on images obtained through image-to-image translation; Hsu et al. [18] introduced a progressive domain adaptive object detector, where an intermediate domain is formed by translating source domain images to the target domain; Inoue et al. [20] proposed

fine-tuning the detector on artificially generated images with instance-level annotations in the target domain. However, simply synthesizing the target domain via image-to-image translation may lead to suboptimal performance, as the foreground of a source domain image could be heavily obscured by the synthesized adverse weather conditions, such as dense fog, heavy darkness, etc.

In this paper, we also generate the auxiliary domain using image-to-image translation. However, we go a step further by utilizing the GT labels in the source domain to constrain the image-to-image translation process. This ensures that the generated images retain crucial foreground information from the source domain. Furthermore, unlike conventional domain translation methods, we facilitate instance-level domain adaptation between the source and auxiliary domains. This approach promotes the detection backbone to extract features that are invariant across domains, thereby enhancing the detector's ability to handle domain shift.

2.2 Domain Adaptation via Adversarial Feature Learning

To learn domain-invariant features, this type of methods employ one or many domain discriminators and train the detector and discriminators adversarially.

Chen et al. [6] first employed adversarial training for domain adaptative object detection. Following this work, many two-stage domain adaptive detectors have been proposed. Saito et al. [36] proposed to conduct strong and weak global alignment during adaptation; He et al. [15] proposed a hierarchical domain feature alignment module where multiple adversarial domain classifiers are designed; Sindagi et al. [37] used weather-specific prior knowledge to define a novel prior-adversarial loss; Li et al. [26] proposed a new adversarial GRL to perform adversarial mining for hard examples.

For one-stage detectors, Hnewa et al. [16] proposed the multi-scale domain adaptive YOLO (MS-DAYOLO) framework based on the YOLO detection framework. Latter on, to maintain consistency in domain classification results across different scales of domain discriminators, Hnewa et al. [17] proposed IMS-DAYOLO by merging the three domain discriminators into a single one. In addition to image-level adaptation, instance-level adaptation should also be incorporated to further improve the adaptation ability. However, as the region proposals are not available in one-stage detectors, conducting instance-level domain adaptation in one-stage object detectors is a more challenging task. To address this issue, Zhang et al. [49] took advantage of pseudo-labels and successfully incorporated instance-level domain adaptation to the YOLO detector.

Although the above approaches effectively alleviate the domain shift, it is worth noting that the during instance-level adaptation, domain-irrelevant shift, such as the appearance of distinct objects, might also be aligned, which is harmful to the detection task. In addition, the presence of noisy pseudo-labels can also potentially mislead the training process of a domain-adaptive one-stage detector.

2.3 Mean Teacher for Domain Adaptive Object Detectors

The mean teacher model [39] was first used in cross-domain object detection by [2], where pseudo-labels generated by the teacher model is utilized to supervise the training of the student model in the target domain. Latter on, Chen et al. [5] introduced probability predictions into the mean teacher framework, and improved pseudo-labeling by applying uncertainty-aware self-training. Li et al. [27] proposed an adaptive teacher model that utilizes weak-strong augmentation and adversarial learning.

Although the mean teacher model has been widely adopted for cross-domain object detection, the pseudo labels used for training the student model may contain substantial noise, which significantly limits its performance. To address this issue, in contrastive mean teacher (CMT) [3], object-level contrastive learning is incorporated to the framework to enhance the representation learning.

Unlike the mean teacher model which applies self-training strategy, our method leverages auxiliary domain data containing GT labels as the source domain. Consequently, our method does not suffer from noisy pseudo labels. Although our method and CMT [3] both utilize contrastive learning, their objectives differ significantly. In CMT, contrastive learning is applied between the instance-level features extracted by the teacher and student models, aiming to mitigate the impact of noisy pseudo labels. On the contrary, our method applies contrastive learning between the instance-level features extracted from the source and auxiliary domains, enabling the cross-domain adaptive detector to focus more on diminishing domain shift, rather than other domain-irrelevant distribution shift.

3 Proposed Method

3.1 Overview of the Framework

We propose the AD-DAYOLO detector to tackle the issue of domain adaptation from the source domain (normal weather conditions) to the target domain (adverse weather conditions), and the framework is shown in Fig. 1.

As can be seen, AD-DAYOLO detector consists of two stages of training. In the first training stage, we generate auxiliary domain images by using the contrastive-learning-based unpaired image-to-image translation (CUT) [31] model. However, in order to ensure that an auxiliary domain image shares the same labels as its corresponding source domain image, it is important to maintain semantic consistency between the foregrounds of these two images. To achieve this, we cascade a YOLO detector, which has been pre-trained on a source domain dataset, with the CUT model. This allows us to predict the object categories and localizations from the auxiliary domain image generated by the CUT model. During training, the pre-trained YOLO detector is kept frozen, and the CUT model is trained with the supervision of both the YOLO detection loss [21] and the original unpaired GAN loss and PatchNCE loss [31].

In the second training stage, the source domain, target domain and auxiliary domain images are all fed to the YOLO backbone for training a domain adaptive

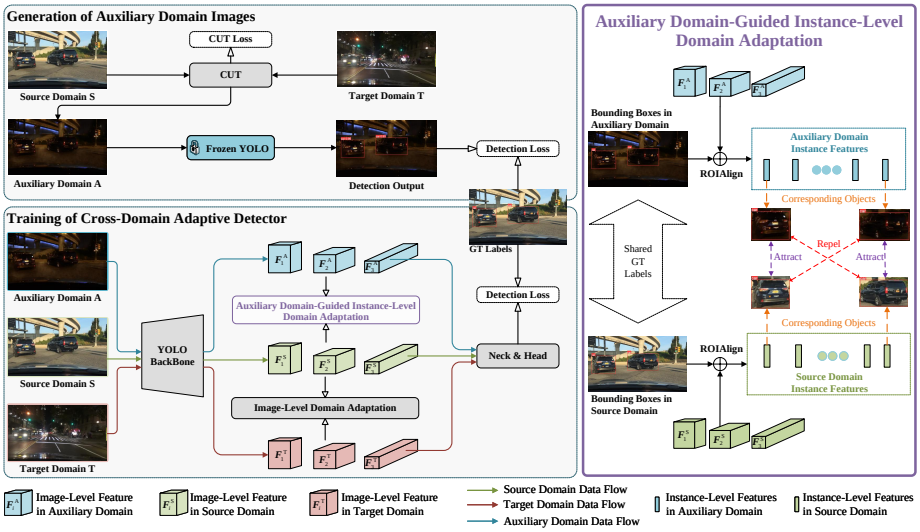


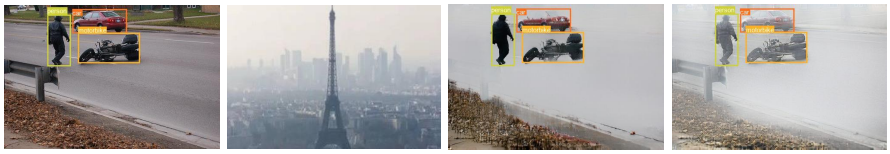
Fig. 1. Overview of the proposed AD-DAYOLO framework. It consists of two training stages. In the first stage, under the weak supervision provided by a frozen YOLO detector, we train a generative model to produce auxiliary domain images. In the second stage, all the source domain, target domain and auxiliary domain images are fed to the YOLO backbone. During this stage, image-level adaptation is performed between the source and target domains. Simultaneously, instance-level adaptation is conducted between the source and auxiliary domains. Moreover, contrastive learning is used to encourage the model to reduce domain shift between the instance-level features.

YOLO detector. Inspired by MS-DAYOLO [16], we extract features from both the source and target domain images to conduct image-level domain adaptation. However, as the target domain images do not contain annotations, we choose not to conduct instance-level adaptation using them. This decision helps us avoid using noisy pseudo labels for training. Instead, we perform instance-level domain adaptation between the source domain and auxiliary domain images, as they share the same GT labels. To achieve this, we utilize ROIAlign [14] on the extracted multi-scale features. Additionally, we employ contrastive learning [13, 11] to encourage the model to reduce domain shift between the extracted instance-level features.

3.2 Generation of Auxiliary Domain Images

In our approach, the auxiliary domain images are supposed to simulate adverse weather conditions similar to those found in the target domain images, and also supposed to share the same GT labels as the source domain images. With such an auxiliary domain, we can avoid the issues associated with noisy pseudo labels.

To this end, we leverage image-to-image translation model to synthesize the auxiliary domain data from the source domain images. Here we choose the un-



(a) Source domain image (b) Target domain image (c) Original CUT result (d) Our method result

Fig. 2. Comparison of images generated by the original CUT model and our method. Fig. 2 (d) preserves essential object details and more closely resembles the target domain than Fig. 2 (c).

paired image-to-image translation method CUT [31]. As mentioned in the study by Wu et al. [43], the CUT model has demonstrated superior performance in terms of structural similarity index (SSIM) scores for image-to-image translation tasks. This indicates that the synthesized images generated by the CUT model are capable of retaining the structural information of the source domain images. However, directly use the CUT model may not lead to satisfactory result, as the CUT model is not constrained to fully preserve the semantic information of objects in the source domain images. As an example, Fig. 2 (c) shows that parts of the objects in the image generated by the original CUT model are obscured by dense fog, which can lead to a loss of important visual details. Such results, where objects are significantly affected or wiped out by adverse weather conditions, cannot be considered as qualified auxiliary domain images.

To address this issue, we use annotations from the source domain images to guide the image generation model. As shown in Fig. 1, a YOLO detector is pre-trained by using annotated source domain images, and then it is cascaded with CUT. During training, the pre-trained YOLO detector is kept frozen, and we use the original CUT loss [31], including GAN loss and PatchNCE loss, together with the YOLO detection loss [21] to constrain the training of the CUT model. Therefore, the foreground-guided loss function for training the CUT model (FCUT) is represented by:

$$\mathcal{L}_{\text{FCUT}} = \mathcal{L}_{\text{CUT}} + \lambda_{\text{D}} \mathcal{L}_{\text{DET}}(Y(G(\mathbf{I}_i^{\text{S}})), \mathbf{B}_i^{\text{S}}, \mathbf{C}_i^{\text{S}}) \tag{1}$$

where $\mathcal{L}_{\text{CUT}}$ indicates the original CUT loss [31]; λ_{D} is a trade-off parameter; $\mathcal{L}_{\text{DET}}$ stands for the YOLO detection loss; $G(\cdot)$ and $Y(\cdot)$ represent the CUT model that converts a source domain image to an auxiliary domain image, and the frozen YOLO detector that predicts the object classes and the localizations, respectively. For the annotated source domain dataset $D_{\text{S}} = \{(\mathbf{I}_i^{\text{S}}, \mathbf{B}_i^{\text{S}}, \mathbf{C}_i^{\text{S}})\}_{i=1}^{N}$, $\mathbf{I}_i^{\text{S}}$, $\mathbf{C}_i^{\text{S}}$, $\mathbf{B}_i^{\text{S}}$, and N represent the i-th source domain image, the annotated object classes in the image, the annotated object localizations in the image, and the number of images in the source domain dataset, respectively.

Fig. 2 shows a comparison of images generated by the original CUT model and the CUT model trained by using Eq. (1). The results clearly demonstrate that our approach successfully maintains the original semantic information of objects in the source domain image, while accurately synthesizing an adverse weather condition that closely resembles the target domain image.

3.3 Auxiliary Domain-guided Instance-level Adaptation

We propose the auxiliary domain-guided instance-level adaptation method. On the one hand, since the auxiliary domain images share GT labels with the source domain, performing instance-level adaptation between the source and auxiliary domains mitigates the impact of noisy pseudo labels. Therefore, we directly use the GT labels to extract instance-level features from the multi-scale image-level features. On the other hand, to ensure the model focus more on minimizing domain shift rather than other domain-irrelevant distribution shift, we further introduce contrastive learning to the instance-level adaptation. These two main steps of our approach are detailed as follows.

Extraction of instance-level features. For instance-level domain adaptation, images from the source and auxiliary domains are paired. Each of these N pairs shares the same GT bounding boxes $\mathcal{B} = \{\mathbf{B}_1^{\mathrm{S}}, \mathbf{B}_2^{\mathrm{S}}, ..., \mathbf{B}_N^{\mathrm{S}}\}$. We denote the image-level features extracted from the j-th convolution layer of YOLO backbone as $\mathbf{F}_j^{\mathrm{S}}$ for the source domain, and $\mathbf{F}_j^{\mathrm{A}}$ for the auxiliary domain, respectively. With $\mathcal{B}$, we apply ROIAlign [14] followed by normalization [13] on both $\mathbf{F}_j^{\mathrm{S}}$ and $\mathbf{F}_j^{\mathrm{A}}$ to extract RoI features from the corresponding image-level features. Since the YOLO backbone produces multi-scale image-level features, ROIAlign [14] is applied, resulting in instance-level features with different channel dimensions.

Multi-scale instance-level contrastive learning. We introduce contrastive learning [13, 11] to the instance-level domain adaptation to enhance the representation ability of the model. Unlike CMT [3], which defines positive samples as the instances with the same category, our approach defines positive samples as the same object located in different domains. Due to the usage of the auxiliary domain, each object in the source domain image is related to its domain-shift version in the auxiliary domain. Consequently, minimizing the distances between positive samples facilitates the learning of domain-invariant features by the model. Additionally, for each object in the source domain, we define a distinct object from the auxiliary domain as a negative sample. By maximizing the distances between negative samples, we ensure that the domain-irrelevant data distribution shift, such as objects with different appearances and scales, is not diminished during domain adaptation. This approach allows the model to capture discriminant instance-level features, resulting in a more robust detection capability. Based on the above analysis, the contrastive loss for a specific scale of features is formulated as:

$$\ell_j = \frac{1}{N_{\mathrm{I}}} \sum_{i=1}^{N_{\mathrm{I}}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\mathbf{z}_{ij}^{\mathrm{S}} \cdot \mathbf{z}_{pj}^{\mathrm{A}} / \tau)}{\exp(\mathbf{z}_{ij}^{\mathrm{S}} \cdot \mathbf{z}_{pj}^{\mathrm{A}} / \tau) + \sum_{a \in \mathrm{Neg}(i)} \exp(\mathbf{z}_{ij}^{\mathrm{S}} \cdot \mathbf{z}_{aj}^{\mathrm{A}} / \tau)} \quad (2)$$

where $\mathbf{z}_{ij}^{\mathrm{S}}$ indicates the source domain instance-level features for the i-th object extracted from the j-th scale of image-level feature; $P(i) = \{p | \mathbf{B}_p^{\mathrm{I}} = \mathbf{B}_i^{\mathrm{I}}, p \in \{1, ..., N_{\mathrm{I}}\}\}$ indicates the set of positive samples which share the same bounding

Algorithm 1 Training of AD-DAYOLO framework.

Require: Annotated source domain data $\mathcal{D}_{\mathrm{S}}$, Unannotated target domain data $\mathcal{D}_{\mathrm{T}}$
Stage 1: Generation of Auxiliary Domain Images
Pre-train YOLO model on $\mathcal{D}_{\mathrm{S}}$;
Cascade the CUT model with the pre-trained YOLO model;
Train CUT model by using the FCUT loss shown in Eq. (1);
Synthesize auxiliary domain images by using the CUT model, i.e., $\mathcal{D}_{\mathrm{A}} = G(\mathcal{D}_{\mathrm{S}})$;
Stage 2: Training of Cross-domain Adaptive Detector
Feed $\mathcal{D}_{\mathrm{S}}$, $\mathcal{D}_{\mathrm{A}}$, and $\mathcal{D}_{\mathrm{T}}$ into the YOLO backbone;
Use the loss function shown in Eq. (4) to train the cross-domain adaptive detector.

box $\mathbf{B}_i^{\mathrm{I}}$ as the i-th source domain object, with N_{I} denoting the number of objects in an image; $\mathbf{z}_{pj}^{\mathrm{A}}$ stands for the auxiliary domain instance-level features for the p-th positive sample extracted from the j-th scale of image-level feature; $\mathrm{Neg}(i)$ is the negative set for the i-th object; $\tau > 0$ is a temperature parameter. By jointly considering multiple scales of instance-level features, the multi-scale instance-level contrastive learning loss is calculated by

$$\mathcal{L}_{\mathrm{INS}} = \sum_{j=1}^{3} \ell_j \tag{3}$$

3.4 Loss Function and the Training Procedure of AD-DAYOLO

To train our AD-ADYOLO, the following loss function is employed:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{DET}} + \lambda_1 \mathcal{L}_{\mathrm{DC}} + \lambda_2 \mathcal{L}_{\mathrm{INS}} \tag{4}$$

where $\mathcal{L}_{\mathrm{DET}}$ denotes the YOLO detection loss [21]; $\mathcal{L}_{\mathrm{INS}}$ is the multi-scale instance-level contrastive learning loss defined in Equ. 3; $\mathcal{L}_{\mathrm{DC}}$ stands for the image-level domain classification loss, which is calculated based on the source domain and target domain features and has the same definition as MS-DAYOLO [16]; λ_1 and λ_2 are two trade-off parameters.

For a better understanding, Algo. 1 summaries the training process of the proposed AD-DAYOLO framework. Note that when training the cross-domain adaptive detector, we conduct image-level adaptation between multi-scale features $\{\mathbf{F}_j^{\mathrm{S}}\}$ from the source domain and $\{\mathbf{F}_j^{\mathrm{T}}\}$ from the target domain, and also perform instance-level adaptation between multi-scale instance-level features $\{\mathbf{z}_{ij}^{\mathrm{S}}\}$ from the source domain and $\{\mathbf{z}_{ij}^{\mathrm{A}}\}$ from the auxiliary domain.

4 Experiments and Analysis

4.1 Implementation Details

All experiments are implemented using the Pytorch framework [32] and carried out on an NVIDIA RTX3090Ti GPU. The YOLOv5s is used as the detector of

Table 1. Overview of the datasets used in our experiments.

Dataset	Type	Images	Instances	Categories
BDD_day_s	daytime dataset	6647	98601	7
BDD_night_t	realistic nighttime	15090	164715	7
BDD_night_test	realistic nighttime	2133	23898	7
VOC_clear	fog-free dataset	8111	19461	5
URHI	realistic foggy dataset	4807	-	-
RTTS	realistic foggy dataset	4322	29598	5

our method. The training and testing images are fixed at a resolution of 640×640. During training, we apply default data augmentation techniques of YOLOv5s. The backbone of YOLOv5 is initialized with the weights pre-trained on the COCO dataset. Similar to [49, 25, 3], we train our model by using the stochastic gradient descent (SGD) method [1]. Our SGD optimizer employs a learning rate of 0.01, weight decay of 0.0005, and momentum of 0.937, consistent with [25]. We empirically determined the following hyperparameters: λ_{D} in Eq. (1) is set to 10, λ_1 and λ_2 in Eq. (4) are set to 0.5 and 0.01, respectively. Following [3], the temperature parameter τ in Eq. (2) is set as 0.07. For practical reasons, we use fixed hyperparameters across scenarios, although these may not be optimal due to limited exploration. Despite this, the experimental results will demonstrate the effectiveness of our method in the following sections.

4.2 Datasets

We perform domain adaptation on two types of adverse weather conditions, including adaptation from daytime to nighttime and from clear-weather to foggy scenes. The used datasets are summarized in Tab. 1. For daytime-to-nighttime adaptation, the images with clear weather conditions in BDD100K [47] are selected and divided to different subsets. "BDD_day_s" and "BDD_night_t" represent the daytime source domain and nighttime target domain subsets for training, respectively, and "BDD_night_test" is the nighttime subset selected from the validation set of BDD100K for testing. For the adaptation from clear-weather to foggy scenes, the widely used PASCAl VOC [9] is chosen as the source domain dataset, while the unannotated realistic hazy images (URHI) dataset [23] is considered as the target domain dataset for training. As the testing dataset, the real-world task-driven testing set (RTTS) with annotations is used. However, only 5 of 20 categories in VOC dataset match with the 5 categories in RTTS dataset. Therefore, we first filter out the unmatched categories in VOC, and obtain a sub-sets called "VOC_clear" for training.

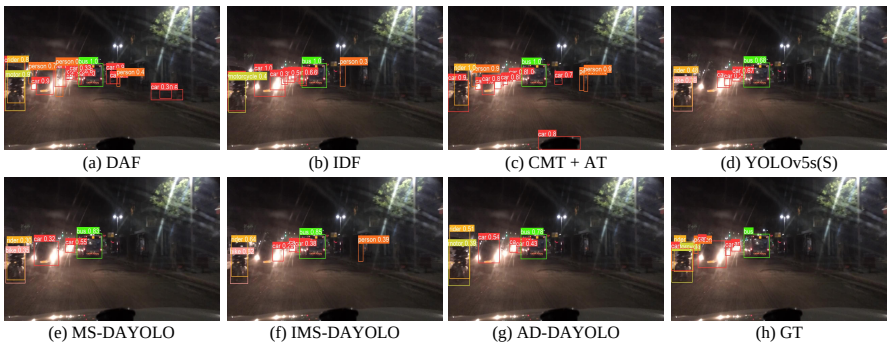
4.3 Adaptation from Daytime to Nighttime

For the purpose of daytime-to-nighttime adaptation, a comprehensive comparison is conducted, which involves various domain adaptive two-stage detectors such as DAF [26], IDF [22], PT [5], AT [27], and CMT [3], as well as domain

Table 2. Performance comparisons for daytime-to-nighttime domain adaptation (AP50 (%)). The results in bold and underlined denote the best and the second best results, respectively. YOLOv5s(S) and YOLOv5s(T) are considered as the performance lower bound and upper bound for the YOLOv5s detector, respectively.

Method	Detector	Car	Bike	Person	Rider	Motor	Bus	Truck	**mAP50**
CMT [3] + PT [5]	Faster RCNN	53.8	12.4	30.7	6.4	4.3	13.3	24.2	20.7
IDF [22]		49.9	21.9	25.6	8.4	5.3	23.3	24.8	22.7
DAF [26]		51.7	29.8	31.8	25.3	14.0	31.0	27.6	30.2
CMT [3] + AT [27]		**67.2**	<u>38.7</u>	42.1	<u>25.9</u>	21.8	<u>44.9</u>	39.1	<u>40.0</u>
YOLOv5s(S)	YOLOv5s	59.4	26.4	36.9	18.8	18.2	30.1	25.9	30.8
MS-DAYOLO [16]		61.8	25.1	39.9	19.3	9.0	42.5	40.1	34.0
S-DAYOLO* [25]		63.9	32.5	**44.8**	25.1	**27.5**	42.6	39.4	39.4
IMS-DAYOLO [17]		65.0	34.1	42.8	25.5	<u>24.2</u>	43.4	<u>42.4</u>	39.6
AD-DAYOLO (Ours)		<u>65.2</u>	**40.6**	<u>44.5</u>	**26.4**	21.5	**45.0**	**46.2**	**41.3**
YOLOv5s(T)		74.4	41.3	49.0	31.3	29.4	51.9	52.3	47.1

* indicates results obtained from [25].



(a) DAF (b) IDF (c) CMT + AT (d) YOLOv5s(S)

(e) MS-DAYOLO (f) IMS-DAYOLO (g) AD-DAYOLO (h) GT

Fig. 3. Detection results obtained on image from BDD100K.

adaptive one-stage detectors namely MS-DAYOLO [16], IMS-DAYOLO [17], and S-DAYOLO [25]. Among the compared adaptive two-stage detectors, DAF and IDF are adversarial feature learning methods, while PT, AT and CMT are mean teacher-based methods. Specifically, CMT combines contrastive learning with mean teacher. Among the compared adaptive one-stage detectors, MS-DAYOLO and IMS-DAYOLO only use image-level domain adaptation, while S-DAYOLO further incorporates instance-level domain adaptation by using pseudo-labels.

Tab. 2 shows the average precision (AP) achieved by various methods on the task of daytime-to-nighttime adaptation. In Tab. 2, the AP for each class is calculated at an intersection over union (IoU) threshold of 0.5, and the mean average precision (mAP) across all classes is also reported. The models denoted as YOLOv5s(S) and YOLOv5s(T) correspond to the models trained on the source domain dataset "BDD_day_s" and the target domain dataset "BDD_night_t" with annotations, respectively. Therefore, the results obtained by YOLOv5s(S) and YOLOv5s(T) are considered as the lower and upper bound performance. Ex-

cept YOLOv5s(S) and YOLOv5s(T), all the other methods are trained by using "BDD_day_s" with annotations and "BDD_night_t" without annotations.

From Tab. 2 we can conclude that: 1) IMS-DAYOLO addresses the inconsistency issue of multi-classifier strategy in MS-DAYOLO, thus achieving 5.6% higher mAP50 than MS-DAYOLO; 2) S-DAYOLO includes an instance-level domain adaptation module and uses an image generation model to augment the source domain dataset, and thus results in a 5.4% higher mAP50 than MS-DAYOLO; 3) among the two-stage adaptive detectors, "CMT + AT" achieves the best results, suggesting the effectiveness of the combination of contrastive learning and mean teacher model; 4) compared to S-DAYOLO and "CMT + AT", our method AD-DAYOLO shows 1.9% and 1.3% mAP50 improvement, respectively, which well demonstrates the effectiveness of our auxiliary domain-guided instance-level adaptation.

Fig. 3 gives qualitative comparisons among the detection results obtained by different adaptation methods on the test image from "BDD_night_test". As can be seen, our approach has the highest accuracy, and the detection results are close to the GT results.

4.4 Adaptation from Clear-weather to Foggy Scenes

Tab. 3 presents the detection results for the adaptation from clear-weather to foggy scenes. Similar to the situation of daytime-to-nighttime adaptation, YOLOv5s(S) denotes the model trained by using the source domain dataset "VOC_clear", and all the other methods are trained on both the annotated source domain dataset "VOC_clear" and the unannotated target domain dataset URHI. The results are obtained on the target domain test set RTTS.

Tab. 3 shows a trend similar to Tab. 2. Moreover, it can be observed that: 1) compared to the nighttime condition, the images under foggy conditions commonly show higher contrast more details, and thus all domain adaptation methods achieve higher AP values than the results in Tab. 2; 2) although multiple domain discriminators in MS-DAYOLO [16] may lead to inconsistent discrimination results, MS-DAYOLO achieves 1.8% higher mAP50 than IMS-DAYOLO [17], which may be attributed to the increased robustness brought by multiple discriminators; 3) our approach AD-DAYOLO achieves the highest mAP50, although the improvement over the other competing methods is less than that in Tab. 2.

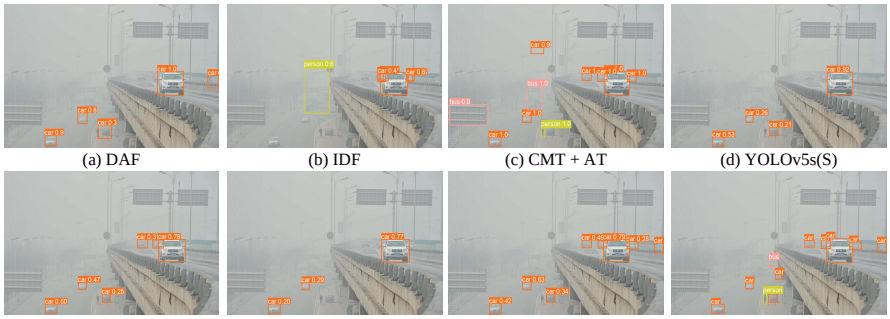
The qualitative comparisons among different methods on the image from RTTS are presented in Fig. 4. For the GT result, the GT labels are directly displayed on the original foggy image. As can be observed, our approach detects the most objects and does not have any false positive results.

4.5 Ablation Study and Efficiency Comparison

Tab. 4 provides ablation study of our AD-DAYOLO framework. The experiments are conducted for daytime-to-nighttime adaptation. In Tab. 4, five variants are compared. Compared to original YOLOv5s(S), variant $\mathbb{N}_1$ incorporates instance-level domain adaptation by using adversarial feature learning strategy

Table 3. Performance comparisons for the adaptation from clear-weather to foggy scenes (AP50 (%)). The results in bold and underlined denote the best and the second best results, respectively. Note that the results of YOLOv5s(T) is not available in this table as the URHI dataset does not contain annotations.

Method	Detector	Bicycle	Bus	Car	Motorbike	Person	**mAP50**
CMT [3] + PT [5]	Faster RCNN	31.1	2.5	39.7	11.8	65.8	30.2
IDF [22]		42.3	9.3	31.1	27.1	65.1	35.0
DAF [26]		52.3	26.4	50.4	46.6	70.9	49.3
CMT [3] + AT [27]		**67.4**	**40.4**	**72.7**	<u>56.1</u>	<u>77.2</u>	<u>62.8</u>
YOLOv5s(S)	YOLOv5s	59.8	33.8	71.0	53.5	80.5	59.7
IMS-DAYOLO [17]		61.2	33.0	65.1	51.9	80.5	58.3
MS-DAYOLO [16]		61.4	35.2	68.0	54.6	81.3	60.1
AD-DAYOLO (Ours)		<u>65.1</u>	<u>38.6</u>	<u>71.1</u>	**59.1**	**82.4**	**63.2**



(a) DAF (b) IDF (c) CMT + AT (d) YOLOv5s(S)

(e) MS-DAYOLO (f) IMS-DAYOLO (g) AD-DAYOLO (h) GT

Fig. 4. Detection results obtained on image from RTTS.

and pseudo labels; in variant $\mathbb{N}_2$, instead of using pseudo labels, the normal CUT model is used to generate the auxiliary domain for instance-level adaptation; different from $\mathbb{N}_2$, $\mathbb{N}_3$ uses the FCUT loss in Eq. (1) to train the CUT model; variant $\mathbb{N}_4$ applies multi-scale contrastive learning to replace the adversarial feature learning in variant $\mathbb{N}_3$; finally, variant $\mathbb{N}_5$ further combines the instance-level adaptation with the image-level adaptation module.

The following conclusions can be made based on Tab. 4: 1) variant $\mathbb{N}_2$ is superior to $\mathbb{N}_1$ by 0.9% mAP50, indicating that using the auxiliary domain to conduct instance-level adaptation is more effective than applying pseudo-labels; 2) compared with $\mathbb{N}_2$, variant $\mathbb{N}_3$ shows a 1.8% improvement in mAP50, demonstrating the effectiveness of the proposed FCUT loss; 3) variant $\mathbb{N}_4$ achieves a 0.8% improvement in mAP50 over variant $\mathbb{N}_3$, indicating that the multi-scale contrastive learning strategy is beneficial for making the model focus more on domain shift rather than domain-irrelevant data distribution shift; 4) variant $\mathbb{N}_5$ outperforms $\mathbb{N}_4$ by 1.3%, suggesting that the image-level adaptation and instance-level domain adaptation are complementary.

Tab. 5 shows the comparison of efficiency. As can be observed, although AD-DAYOLO training time is longer than MS-DAYOLO that employs only image-

Table 4. Ablation Study of AD-DAYOLO framework for the daytime-to-nighttime adaptation (mAP50 (%)).

Variants	Image-Level Adaptation	Instance-Level Adaptation				mAP50
		Pseudo-labels	CUT	FCUT	Contrast Learning	
YOLOv5s(S)	×	×	×	×	×	30.8
$\mathbb{N}_1$	×	✓	×	×	×	36.5
$\mathbb{N}_2$	×	×	✓	×	×	37.4
$\mathbb{N}_3$	×	×	×	✓	×	39.2
$\mathbb{N}_4$	×	×	×	✓	✓	40.0
$\mathbb{N}_5$	✓	×	×	✓	✓	41.3

Table 5. Comparison of efficiency. Runtime is tested on a RTX 3090Ti GPU with an image size of 640×640. The domain adaptation modules are removed during testing.

Method	Training Phase		Testing Phase	
	Param. (M)	Training Time (h)	Param. (M)	Runtime (ms)
MS-DAYOLO	10.15	22.57	7.02	6.6
IMS-DAYOLO	11.43	22.86	7.02	6.6
CMT + AT	184.61	27.67	92.31	42.7
AD-DAYOLO (Our)	31.87	33.66	7.02	6.6

level domain adaptation, its testing speed matches that of MS-DAYOLO and is significantly faster than the "CMT + AT" method.

5 Conclusion

To improve the detection accuracy in adverse weather conditions, this paper proposes the AD-DAYOLO framework, which leverages the auxiliary domain to guide the instance-level adaptation. Firstly, to produce a suitable auxiliary domain for training, we introduce the FCUT loss to train the CUT model, in which the foreground information of the source domain image is used as an additional constraint. Secondly, the GT bounding boxes in source domain are applied to extract instance-level features from the auxiliary domain and the source domain, respectively. By leveraging the contrastive learning strategy on the instance-level features, the model is forced to concentrate on minimizing the domain shift rather than other domain-irrelevant data distribution shift. Finally, in addition to the instance-level adaptation, our framework also conducts image-level adaptation between the source domain and the target domain images. Experimental results demonstrate that the proposed method effectively promotes the model to focus on extracting domain-invariant features, thereby enhancing the cross-domain adaptation capability of the AD-DAYOLO framework.

Acknowledgments. This work was supported by NSFC under Grant 62171145, Guangxi Key R&D Program under Grant AB23075106, and Guangxi Key Laboratory of Multimedia Communications and Network Technology.

References

1. Bottou, L.: Stochastic gradient descent tricks. In: Neural Networks: Tricks of the Trade: Second Edition, pp. 421–436. Springer (2012)
2. Cai, Q., Pan, Y., Ngo, C.W., Tian, X., Duan, L., Yao, T.: Exploring Object Relation in Mean Teacher for Cross-Domain Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11449–11458. Long Beach, CA, USA (2019)
3. Cao, S., Joshi, D., Gui, L.Y., Wang, Y.X.: Contrastive Mean Teacher for Domain Adaptive Object Detectors. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 23839–23848. Vancouver, BC, Canada (2023)
4. Chang, K., Li, H., Tan, Y., Ding, P.L.K., Li, B.: A two-stage convolutional neural network for joint demosaicking and super-resolution. IEEE Transactions on Circuits and Systems for Video Technology **32**(7), 4238–4254 (Jul 2022)
5. Chen, M., Chen, W., Yang, S., Song, J., Wang, X., Zhang, L., Yan, Y., Qi, D., Zhuang, Y., Xie, D., Pu, S.: Learning Domain Adaptive Object Detection with Probabilistic Teacher. In: International Conference on Machine Learning (ICML). vol. 162, pp. 3040–3055. Baltimore, Maryland, USA (2022)
6. Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L.: Domain Adaptive Faster R-CNN for Object Detection in the Wild. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3339–3348. Salt Lake City, UT, USA (2018)
7. Cui, Z., Qi, G.J., Gu, L., You, S., Zhang, Z., Harada, T.: Multitask AET with Orthogonal Tangent Regularity for Dark Object Detection. In: IEEE International Conference on Computer Vision (ICCV). pp. 2533–2542. Montreal, QC, Canada (2021)
8. Duan, L., Tsang, I.W., Xu, D.: Domain Transfer Multiple Kernel Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(3), 465–479 (2012)
9. Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J.M., Zisserman, A.: The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision **88**(2), 303–338 (2010)
10. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.S.: Domain-Adversarial Training of Neural Networks. Journal of Machine Learning Research **17**(59), 1–35 (2016)
11. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality Reduction by Learning an Invariant Mapping. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1735–1742. New York, NY, USA (2006)
12. Han, F., Chang, K., Li, G., Ling, M., Huang, M., Gao, Z.: Illumination-aware divide-and-conquer network for improperly-exposed image enhancement. Neural Networks **180**, 106733 (2024)
13. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum Contrast for Unsupervised Visual Representation Learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9726–9735. Seattle, WA, USA (2020)
14. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988. Venice, Italy (2017)
15. He, Z., Zhang, L.: Multi-Adversarial Faster-RCNN for Unrestricted Object Detection. In: IEEE International Conference on Computer Vision (ICCV). pp. 6667–6676. Seoul, Korea (South) (2019)
16. Hnewa, M., Radha, H.: Multiscale Domain Adaptive Yolo For Cross-Domain Object Detection. In: IEEE International Conference on Image Processing (ICIP). pp. 3323–3327. Anchorage, AK, USA (2021)

17. Hnewa, M., Radha, H.: Integrated Multiscale Domain Adaptive YOLO. IEEE Transactions on Image Processing **32**, 1857–1867 (2023)
18. Hsu, H.K., Yao, C.H., Tsai, Y.H., Hung, W.C., Tseng, H.Y., Singh, M., Yang, M.H.: Progressive Domain Adaptation for Object Detection. In: IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 738–746. Snowmass, CO, USA (2020)
19. Huang, S.C., Le, T.H., Jaw, D.W.: DSNet: Joint Semantic Learning for Object Detection in Inclement Weather Conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(8), 2623–2633 (2021)
20. Inoue, N., Furuta, R., Yamasaki, T., Aizawa, K.: Cross-Domain Weakly-Supervised Object Detection Through Progressive Domain Adaptation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5001–5009. Salt Lake City, UT, USA (2018)
21. Jocher, G.: Yolov5. https://github.com/ultralytics/yolov5 (2020)
22. Lang, Q., Zhang, L., Shi, W., Chen, W., Pu, S.: Exploring Implicit Domain-Invariant Features for Domain Adaptive Object Detection. IEEE Transactions on Circuits and Systems for Video Technology **33**(4), 1816–1826 (2023)
23. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking Single-Image Dehazing and Beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2019)
24. Li, C., Zhou, H., Liu, Y., Yang, C., Xie, Y., Li, Z., Zhu, L.: Detection-Friendly Dehazing: Object Detection in Real-World Hazy Scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(7), 8284–8295 (2023)
25. Li, G., Ji, Z., Qu, X., Zhou, R., Cao, D.: Cross-Domain Object Detection for Autonomous Driving: A Stepwise Domain Adaptative YOLO Approach. IEEE Transactions on Intelligent Vehicles **7**(3), 603–615 (2022)
26. Li, J., Xu, R., Ma, J., Zou, Q., Ma, J., Yu, H.: Domain Adaptive Object Detection for Autonomous Driving under Foggy Weather. In: IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 612–622. Waikoloa, HI, USA (2023)
27. Li, Y.J., Dai, X., Ma, C.Y., Liu, Y.C., Chen, K., Wu, B., He, Z., Kitani, K., Vajda, P.: Cross-Domain Adaptive Teacher for Object Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7571–7580. New Orleans, LA, USA (2022)
28. Ling, M., Chang, K., Huang, M., Li, H., Dang, S., Li, B.: PRNet: Pyramid restoration network for raw image super-resolution. IEEE Transactions on Computational Imaging **10**, 479–495 (2024)
29. Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., Zhang, L.: Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. In: AAAI Conference on Artificial Intelligence (AAAI). pp. 1792–1800. Virtual Event (2022)
30. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Unsupervised Domain Adaptation with Residual Transfer Networks. In: Conference on Neural Information Processing Systems (NeurIPS). pp. 136–144. Barcelona, Spain (2016)
31. Park, T., Efros, A.A., Zhang, R., Zhu, J.: Contrastive Learning for Unpaired Image-to-Image Translation. In: European Conference on Computer Vision (ECCV). vol. 12354, pp. 319–345. Glasgow,UK (2020)
32. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)

33. Qin, Q., Chang, K., Huang, M., Li, G.: DENet: Detection-driven Enhancement Network for Object Detection Under Adverse Weather Conditions. In: Asian Conference on Computer Vision (ACCV). vol. 13843, pp. 491–507. Virtual Event (2022)
34. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788. Las Vegas, NV, USA (2016)
35. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1137–1149 (2017)
36. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-Weak Distribution Alignment for Adaptive Object Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6949–6958. Long Beach, CA, USA (2019)
37. Sindagi, V.A., Oza, P., Yasarla, R., Patel, V.M.: Prior-Based Domain Adaptive Object Detection for Hazy and Rainy Conditions. In: European Conference on Computer Vision (ECCV). vol. 12359, pp. 763–780. Glasgow,UK (2020)
38. Song, Y., He, Z., Qian, H., Du, X.: Vision transformers for single image dehazing. IEEE Transactions on Image Processing **32**, 1927–1941 (2023)
39. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Conference on Neural Information Processing Systems (NeurIPS). pp. 1195–1204. Long Beach, CA, USA (2017)
40. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial Discriminative Domain Adaptation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2962–2971. Honolulu, HI, USA (2017)
41. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A General U-Shaped Transformer for Image Restoration. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17662–17672. New Orleans, LA, USA (2022)
42. Wei, X., Chang, K., Li, G., Huang, M., Qin, Q.: DLEN: Deep Laplacian Enhancement Networks for Low-Light Images. In: IEEE International Conference on Image Processing (ICIP). pp. 2120–2124. Kuala Lumpur, Malaysia (2023)
43. Wu, C.H., De La Torre, F.: A Latent Space of Stochastic Diffusion Models for Zero-Shot Image Editing and Guidance. In: IEEE International Conference on Computer Vision (ICCV). pp. 7344–7353. Paris, France (2023)
44. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10551–10560 (2021)
45. Yan, T., Li, H., Sun, B., Wang, Z., Luo, Z.: Discriminative Feature Mining and Enhancement Network for Low-Resolution Fine-Grained Image Recognition. IEEE Transactions on Circuits and Systems for Video Technology **32**(8), 5319–5330 (2022)
46. Yin, M., Ling, M., Chang, K., Yuan, Z., Qin, Q., Chen, B.: Joint Image and Feature Enhancement for Object Detection under Adverse Weather Conditions. In: International Joint Conference on Neural Networks (IJCNN). pp. 1–8. Yokohama, Japan (2024)
47. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2633–2642. Seattle, WA, USA (2020)

48. Yuan, Z., Ling, M., Chang, K., Yin, M., Li, M., Chen, B.: A Two-Stage Enhancement Method for Object Detection on Low-Resolution Images. In: International Joint Conference on Neural Networks (IJCNN). pp. 1–8. Yokohama, Japan (2024)
49. Zhang, S., Tuo, H., Hu, J., Jing, Z.: Domain Adaptive YOLO for One-Stage Cross-Domain Detection. In: Asian Conference on Machine Learning (ACML). vol. 157, pp. 785–797. Virtual Event (2021)
50. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: IEEE International Conference on Computer Vision (ICCV). pp. 2223–2232. Venice, Italy (2017)