

Article

Degradation Type-Aware Image Restoration for Effective Object Detection in Adverse Weather

Xiaochen Huang^{1,2} , Xiaofeng Wang¹, Qizhi Teng^{1,*} , Xiaohai He¹ and Honggang Chen^{1,3} ¹ College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China² Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin 300384, China³ Yunnan Key Laboratory of Software Engineering, Yunnan University, Kunming 650600, China

* Correspondence: qzteng@scu.edu.cn

Abstract: Despite significant advancements in CNN-based object detection technology, adverse weather conditions can disrupt imaging sensors' ability to capture clear images, thereby adversely impacting detection accuracy. Mainstream algorithms for adverse weather object detection enhance detection performance through image restoration methods. Nevertheless, the majority of these approaches are designed for a specific degradation scenario, making it difficult to adapt to diverse weather conditions. To cope with this issue, we put forward a degradation type-aware restoration-assisted object detection network, dubbed DTRDNet. It contains an object detection network with a shared feature encoder (SFE) and object detection decoder, a degradation discrimination image restoration decoder (DDIR), and a degradation category predictor (DCP). In the training phase, we jointly optimize the whole framework on a mixed weather dataset, including degraded images and clean images. Specifically, the degradation type information is incorporated in our DDIR to avoid the interaction between clean images and the restoration module. Furthermore, the DCP makes the SFE possess degradation category awareness ability, enhancing the detector's adaptability to diverse weather conditions and enabling it to furnish requisite environmental information as required. Both the DCP and the DDIR can be removed according to requirement in the inference stage to retain the real-time performance of the detection algorithm. Extensive experiments on clear, hazy, rainy, and snowy images demonstrate that our DTRDNet outperforms advanced object detection algorithms, achieving an average mAP of 79.38% across the four weather test sets.

Keywords: object detection in various weather scenes; restoration-assisted object detection; degradation type awareness; multi-task joint learning



Citation: Huang, X.; Wang, X.; Teng, Q.; He, X.; Chen, H. Degradation Type-Aware Image Restoration for Effective Object Detection in Adverse Weather. *Sensors* **2024**, *24*, 6330. <https://doi.org/10.3390/s24196330>

Academic Editors: Saeid (Adam) Pirasteh and Stefania Perri

Received: 8 August 2024

Revised: 18 September 2024

Accepted: 26 September 2024

Published: 30 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Object detection technology, which is dedicated to extracting target information such as spatial coordinates, semantic category, and physical dimensions from images or videos, constitutes one of the foundational tasks in the field of computer vision. In recent years, the rapid evolution of deep learning has catalyzed significant advancements in learning-based object detection technologies, showcasing exceptional accuracy and real-time performance across diverse datasets sourced from various sensors, including cameras, surveillance systems, vehicles, drones, and satellites [1–5]. Consequently, this technology is increasingly being utilized in various domains, including autonomous driving [6], environmental monitoring [7], military reconnaissance [8], and ship fire detection [9]. Nevertheless, mainstream detection algorithms are generally optimized and assessed using high-quality datasets such as MS COCO [10] and PASCAL VOC [11]. The data captured by sensors under adverse weather conditions frequently lead to degraded images characterized by poor visibility and blurred target details, posing significant challenges for the detection model in extracting target features. This constrains the model's generalization and robustness, resulting in a serious decline in detection performance under adverse weather conditions such as haze,

rain, and snow. Therefore, enhancing the stability of object detection algorithms in these severe environments has become an urgent issue that necessitates resolution.

To address this issue, the most intuitive solution is to employ image enhancement techniques [12–20] as a preprocessing step to mitigate the effects of adverse weather factors. While these techniques generally improve the visual quality of images, the resulting restored outputs may not be optimally tailored for machine perception. Consequently, the performance of individually trained two-stage restoration-detection frameworks is suboptimal. Furthermore, some researchers [21–23] have concurrently optimized restoration and detection networks in a cascading manner to ensure that the restored outputs are more conducive to the perception of the detection network. Although these methods do enhance detection accuracy under inclement weather conditions, they inevitably introduce additional model parameters, which compromises real-time object detection.

Recently, several studies [23–27] have focused on enhancing object detection in adverse weather conditions through transfer learning, particularly employing domain adaptation methods. These approaches primarily offset the domain gap [28,29] between the source (clear images) and target domains (degraded images), facilitating improved learning from clear to degraded images. While these methods can extract intrinsic features of input images rather than relying solely on restored features, they often overlook the fact that degraded factors complicate target feature extraction and lead to an over-reliance on source domain information, which ultimately limits their performance. In contrast to the aforementioned approaches, TogetherNet [6] integrates image restoration and object detection within a multi-task joint learning framework. It employs a restoration decoder that shares features encoded by the backbone of the detection network, along with a regression loss to optimize this shared backbone. As a result, clear image features can be extracted even when inputting degraded images, significantly enhancing object detection performance for such cases. However, this method is specifically designed for certain adverse weather conditions and has several limitations that affect the stability of the detection model: (1) the network lacks awareness of varying weather conditions when processing different types of images; (2) the entire joint learning framework is trained exclusively on degraded images, neglecting more common normal weather scenarios. As presented in Figure 1, although the adverse weather image object detector TogetherNet [6] significantly improves detection accuracy for hazy images, its performance on clear images deteriorates compared to YOLOXs [30]. Therefore, further investigation is needed to design a detection model that can adapt to diverse degraded images while maintaining performance on clear images.

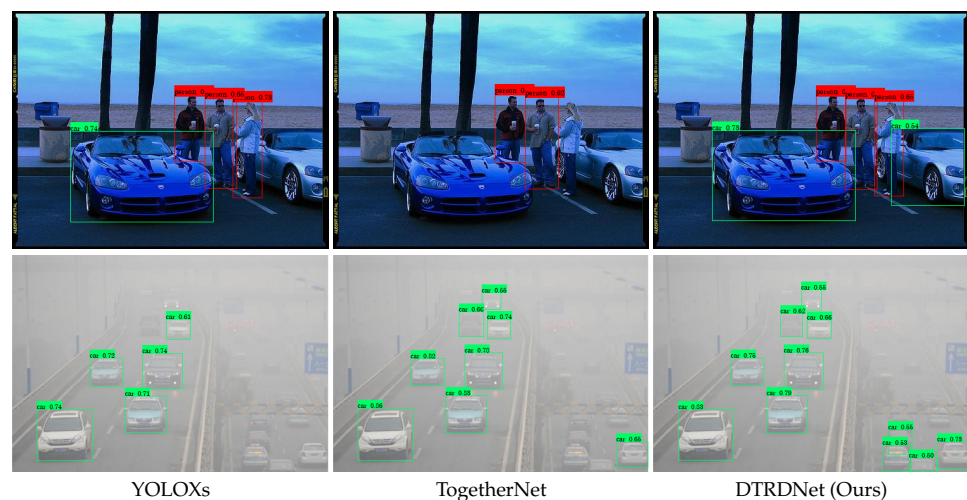


Figure 1. Detection results on clean and hazy images of YOLOXs [30], TogetherNet [6] and our DTRDNet. The three models are optimized for adverse weather image. The clean image is from VOC [11] and the hazy image is from RTTS [31].

In this study, we introduce DTRDNet by improving the restoration module and bridging image classification [32] based on TogetherNet [6], which is designed to improve precision and robustness in object detection for images captured under diverse weather conditions, including both clear and various degraded scenarios. Firstly, to avoid excessive model complexity and the limitation of adapting to only one specific degradation type, we construct a framework based on the joint learning paradigm similar to TogetherNet [6]. This involves feeding various degraded images along with clear images into the detection model during the training stage to better accommodate diverse weather conditions, allowing the model to learn more recognizable target features from clear images. However, introducing clear images during the training phase can interfere with the restoration module, and extracting restored features from clear images may hinder their detection. To eliminate the mutual interference between clear images and the restoration network, we embed a degradation discriminator (DD) between the shared feature encoder and the restoration decoder. This discriminator distinguishes between clear and degraded images using degradation type labels, ensuring the restoration decoder only processes degraded images. Additionally, removing unnecessary restoration training of clear images offers the added benefit of reducing training computational costs. Building upon this approach, we further optimize the restoration decoder to leverage more low-level features that enhance the extraction of detailed image features, thereby improving the model's ability to accurately position objects for detection. The combination of a DD and our optimized restoration decoder forms our degradation discrimination image restoration decoder (DDIR). Finally, we construct a degradation category predictor (DCP) to imbue the encoded features with awareness of degradation types by generating predicted degradation categories for each input during training. This enhancement improves the model's detection stability under diverse weather conditions and can provide additional environmental information as needed. As shown in Figure 1, our method outperforms both the current detector YOLOXs [30] and the adverse weather detector TogetherNet [6] in detecting both degraded and clear images.

Previous studies have predominantly focused on addressing specific adverse weather conditions, which is insufficient for effectively handling multiple weather scenarios with a single model. As a result, practical applications often require the deployment of multiple models to accommodate diverse weather conditions, incurring substantial costs. Our research aims to tackle the challenge of object detection under various complex weather conditions through a unified network. The proposed approach significantly enhances the accuracy and stability of the detector across different weather scenarios without incurring additional inference costs.

On the whole, our study makes the following contributions:

- On the basis of TogetherNet, we further develop the multi-task joint learning paradigm by bridging the tasks of detection, restoration, and classification to enhance the performance and stability of the detector in different environments.
- The proposed DDIR effectively mitigates the influence of clear data on the restoration model and improves the restoration decoder to bolster the model's capacity to extract detailed features. This facilitates the model to more effectively encode clear features for both clean and degraded images.
- The introduced DCP endows the feature extraction network with degradation category awareness, thereby enhancing the model's stability under varying weather conditions.

2. Related Work

2.1. Object Detection

Object detection is one of the most important tasks in the computer vision field, widely employed in domains such as ship detection [9,33], autonomous driving [34], and military surveillance [8]. Recently, with the rapid advancements in deep learning, learning-based detectors have emerged as the mainstream approach for both application and research. These can be broadly categorized into region proposal-based methods and regression-based methods. The regional proposal-based approach first generates the candidate proposal

ARES by methods such as selective search and edge boxes, and then performs adjustments using the subsequent detection model. The most representative proposal-based detector, R-CNN [35], first generates proposal areas through a convolutional neural network and then uses a support vector machine (SVM) to classify them. Inspired by the success of R-CNN in object detection, a large number of variants based on this framework have been developed, each aiming to enhance the original model's performance and address its limitations. These variants include Fast R-CNN [36], Faster R-CNN [37], Libra R-CNN [38], Dynamic R-CNN [39], and Mask R-CNN [40], among others. While the regional proposal-based approaches generally exhibit certain advantages in accuracy, they often come with higher computational costs and perform worse in terms of detection speed because the requirements for localization and classification are handled in two stages.

In contrast to the regional proposal-based approach, regression-based approaches represented by the YOLO series [5,8,30,41–45], SSD [46], RetinaNet [47], and CenterNet [48], among others, complete the prediction of target categories and localization in a single stage. This type of detector usually utilizes a feature encoder to extract image features and then decodes the features to predict the category and location information of targets. It should be noted that regression-based detection algorithms typically preset anchor boxes with different aspect ratios, which saves time and computational resources required to generate proposal boxes compared to regional proposal-based detectors, thus achieving faster detection speeds. In addition, the two optimizations of Focal Loss (FL) [47], used to mitigate the problem of positive and negative sample imbalance during the training stage, and Feature Pyramid Networks (FPN) [49], employed to address the issue of variation in target sizes, greatly improve the detection performance of regression-based single-stage detectors, making them better suited to satisfy the trade-off between performance and speed. In summary, although the performance of regression-based single-stage object detection algorithms is not as accurate as that of two-stage algorithms, they do not require generating proposal boxes and can directly output information about object categories, positions, and confidence in an end-to-end regression method. This provides these methods with a significant advantage in real-time performance.

2.2. Object Detection in Adverse Weather Conditions

The advancement of learning-based object detection technology has led to its widespread application across various scenarios. Nevertheless, conventional detection algorithms are susceptible to weather degradation, which poses challenges when encountering adverse environments. Although an increasing number of studies have concentrated on object detection in low-quality images, research specifically addressing this area remains limited.

Early algorithms for object detection in low-quality images predominantly employ image restoration techniques, including dehazing [12–14,50], deraining [14,16,21,51], and desnowing [17,18], among others. These methods function as preprocessing steps designed to restore degraded images before the execution of object detection. Intuitively, clearer images obtained from restoration algorithms should enhance detection performance. However, empirical evidence suggests that even when the clarity of restored images closely resembles that of real clear images, their detection performance may not improve. This discrepancy could be attributed to a domain gap between the restored images and actual clear images. Furthermore, some researchers have sought to integrate image restoration and object detection networks to alleviate the adverse effects of degraded information.

Sindagi et al. [23] improve the object in rainy and foggy weather by an unsupervised domain-adaptive framework based on priors. Liu et al. [24] develop an image-adaptive network for object detection in adverse weather conditions, which integrates restoration and detection models in an end-to-end framework that can learn appropriate parameters to enhance images for detection in a weakly supervised manner. Wang et al. [6] employ a unified joint learning paradigm to integrate image restoration and object detection, training a shared feature encoder for the two tasks and jointly optimizing them through restoration and detection losses, thereby enhancing detection performance under adverse weather

3.1. Overall Architecture

As illustrated in Figure 2, our proposed DTRDNet integrates three tasks: object detection, restoration, and classification. It primarily consists of three components: (1) an object detection network that incorporates a shared feature encoder (SFE) and an object detection decoder, (2) our proposed restoration decoder (DDIR), and (3) a degradation predictor (DCP). For convenience, we first introduce the overall workflow of our algorithm.

Differing with current optimization algorithms for adverse weather image object detection, we are devoted to utilizing restoration and degradation discrimination to improve detection performance for degraded images while retaining performance on clear images. For this reason, we mix the clear and degraded images together to optimize the network parameter during the training stage. The input is first processed by the SFE to extract multi-scale image features and learn degradation category information. Subsequently, the encoded features are fed into the object detection decoder, the DDIR, and the DCP to generate the detection results, restored clear images, and predicted degradation types. Finally, we formulate detection, restoration and degradation classification loss functions to jointly optimize the whole framework. This enables the SFE to distill knowledge from restoration and degradation classification tasks to assist with the detection task. Specifically, the DDIR enables the SFE to extract clear multi-scale features regardless of whether the input is a clear or degraded image, thereby improving detection performance under both clean and degraded weather conditions. The DIP endows the SFE with the capability to discern degradation type information within images, thereby enhancing stability when encountering various weather-induced degradations. Notably, during the inference phase, it is feasible to selectively omit the DDIR and the DCP to mitigate computational complexity without compromising object detection accuracy.

3.2. Object Detection Network

Although the aforementioned deficiencies require further attention, it is unequivocal that TogetherNet [6] represents a significant advancement in the detection of adverse weather conditions. In particular, its baseline network, YOLOXs, serves as a prominent regression-based detector, and the enhancements associated with DTFE and SC Conv have demonstrated considerable efficacy in object detection. Consequently, we retain its detection architecture, which comprises a backbone network and a detection head. The backbone operates as an SFE module to encode multi-scale image features for subsequent decoding modules, while the detection head functions as the object detection decoder within our framework.

3.2.1. Shared Feature Encoder

Given a clear or degraded input image $X \in \mathbb{R}^{3 \times 640 \times 640}$, the focus first divides the image into four patches and concatenates them to diminish spatial dimensions, thereby contributing to reduced computational costs. Then, a series of cascaded convolutions and CSP (Cross Stage Partial Network) are utilized to extract multi-scale features. In this process, SPP and DTFE are employed to enhance the feature representation capability of the entire encoder. There are four scale features extracted by the SFE, which can be mathematically expressed as follows:

$$f_{\times 4}, f_{\times 8}, f_{\times 16}, f_{\times 32} = E_{SFE}(X), \quad (1)$$

where $E_{SFE}(\cdot)$ denotes the encoding process of the SFE, and $f_{\times 4}, f_{\times 8}, f_{\times 16}$, and $f_{\times 32}$ represent encoded image features at different scales, with subscript $\times n$ indicating the size difference relative to the input image. Specifically, $f_{\times 4} \in \mathbb{R}^{64 \times 160 \times 160}$, $f_{\times 8} \in \mathbb{R}^{128 \times 80 \times 80}$, $f_{\times 16} \in \mathbb{R}^{256 \times 40 \times 40}$, and $f_{\times 32} \in \mathbb{R}^{512 \times 20 \times 20}$. These multi-scale features will be utilized by the object detection decoder (DDIR) and the DCP to generate results for object detection, image restoration, and degradation category prediction. Consequently, for object detection tasks, the optimized SFE following backpropagation ensures that these features contain informa-

tion from clear images and awareness of degradation categories. This significantly enhances feature extraction capabilities and improves generalization across diverse environments.

3.2.2. Object Detection Decoder

As previously mentioned, we employ the head of TogetherNet [6] as our object detection decoder. It includes a neck and detect head that have been enhanced by self-calibrated convolutions (SC Conv) [57]. The neck consists of a widely used feature pyramid network (FPN) [49] and a path aggregation network (PANet) [58] structure, which integrate high-level semantic features with low-level texture features to generate spatially scaled features for targets of varying sizes. Specifically, the FPN upsamples the high-level feature after adjusting its channel number using a 1×1 convolution, and then concatenates it with the corresponding scale's feature from the SFE. This operation is repeated twice to produce two fused features with different scales. The formulary description is as follows:

$$f_{\times 32}^{FPN} = Conv_{1,1}(f_{\times 32}), \quad (2)$$

$$f_{\times 16}^{FPN} = Conv_{1,1}(CSP(Cat[Up(f_{\times 32}^{FPN}), f_{\times 16}]_c)), \quad (3)$$

$$f_{\times 8}^{FPN} = CSP(Cat[SC(Up(f_{\times 16}^{FPN})), f_{\times 8}]_c), \quad (4)$$

where $Conv_{k,s}(\cdot)$ denotes convolution with a kernel size of $k \times k$ and a stride of s . The operation $Up(\cdot)$ signifies double upsampling. The notation $Cat[\cdot, \cdot]_c$ represents the concatenation operation along the channel dimension. $CSP(\cdot)$ and $SC(\cdot)$ denote the functions of CSP and SC modules, respectively. Followed by the PANet, it convolves the feature map $f_{\times 8}^{FPN}$ to generate two smaller-sized features, $f_{\times 16}^{PAN}$ and $f_{\times 32}^{PAN}$, as formulated below:

$$f_{\times 16}^{PAN} = CSP(Cat[SC(Conv_{3,2}(f_{\times 8}^{FPN})), f_{\times 16}^{FPN}]_c); \quad (5)$$

$$f_{\times 32}^{PAN} = CSP(Cat[SC(Conv_{3,2}(f_{\times 16}^{PAN})), f_{\times 32}^{FPN}]_c). \quad (6)$$

In the aforementioned FPN and PANet, self-calibrated convolutions and CSP are employed to expand the receptive field and further enhance features, which are subsequently decoupled for improved object detection. Ultimately, we utilize three detection heads to separate the features $f_{\times 8}^{FPN}$, $f_{\times 16}^{PAN}$, and $f_{\times 32}^{PAN}$ in order to output the detection results.

3.3. Restoration Network

The most current restoration-based adverse weather object detection algorithms focus solely on restoring clear images from single degraded inputs, which inadequately addresses the detection requirements across various scenarios when employing a single model. Furthermore, the utilization of multiple detection models tailored for various environments considerably complicates the deployment process. To enhance detection performance across diverse degraded images while ensuring effectiveness for clear images, we propose incorporating degradation type information to effectively differentiate between clear and degraded images. Subsequently, we optimize the image restoration decoder to establish the DDIR. It is designed to exclusively restore degraded images, thereby eliminating interference from clear images in the restoration process.

Figure 3 depicts the comprehensive restoration process of our DTRDNet. We integrate our proposed DD between the SFE and the restoration decoder to determine whether an input is a degraded image. For a batch of inputs, regardless of their classification as degraded or clear images, the SFE first extracts image features, while the DD utilizes degradation type labels to assess the degradation status of each image. If an image is identified as a degraded image, its multi-scale features extracted by the SFE are forwarded to the restoration decoder to generate a restored output. Otherwise, if the input is a clear image, we do not process the corresponding features with the restoration decoder; instead,

we utilize the clear input image as the output result. All restored degraded images and clear images are concatenated to generate the final restoration output. It should be noted that, to maintain the original order of corresponding images within a batch for calculating restoration loss, each image in the same batch is processed one by one during this procedure, which also helps alleviate the training burden. Specifically, as depicted in Figure 2, for a degraded image, our DDIR utilizes its multi-scale features $f_{\times 4}$, $f_{\times 8}$, $f_{\times 16}$, and $f_{\times 32}$ extracted by the SFE to generate restored images. The high-level feature $f_{\times 32}$ is restored to match input size through five cascading upsampling and convolution operations. In this process, lower-level features $f_{\times 16}$, $f_{\times 8}$, and $f_{\times 4}$ are progressively integrated. The process can be expressed as follows:

$$f_{\times 16}^R = \text{Conv}_{3,2}(\text{Up}(f_{\times 32}^D)), \quad (7)$$

$$f_{\times 8}^R = \text{Conv}_{3,2}(\text{Up}(\text{Cat}[f_{\times 16}^R, f_{\times 16}^D])), \quad (8)$$

$$f_{\times 4}^R = \text{Conv}_{3,2}(\text{Up}(\text{Cat}[f_{\times 8}^R, f_{\times 8}^D])), \quad (9)$$

$$R^D = \text{Conv}_{3,2}(\text{Up}(\text{Conv}_{3,2}(\text{Up}(f_{\times 4}^R)))), \quad (10)$$

where $f_{\times 32}^D$ refers to the $f_{\times 32}$ of degraded images, f^R denotes the feature produced by the DDIR, and $R^D \in \mathbb{R}^{3 \times H \times W}$ stands for the output of restoration decoder, with H and W denoting the spatial dimensions of input images. For inputs in a batch without degradation, we concatenate it with restored output R^D of the degraded images to form the final restoration results as follows:

$$R = \text{Cat}[R^D, X^C]_b, \quad (11)$$

where X^C refers to the clear image contained in inputs, R represents the final restoration result, and $\text{Cat}[\cdot, \cdot]_b$ denotes the operation of concatenation in the order of input batch.

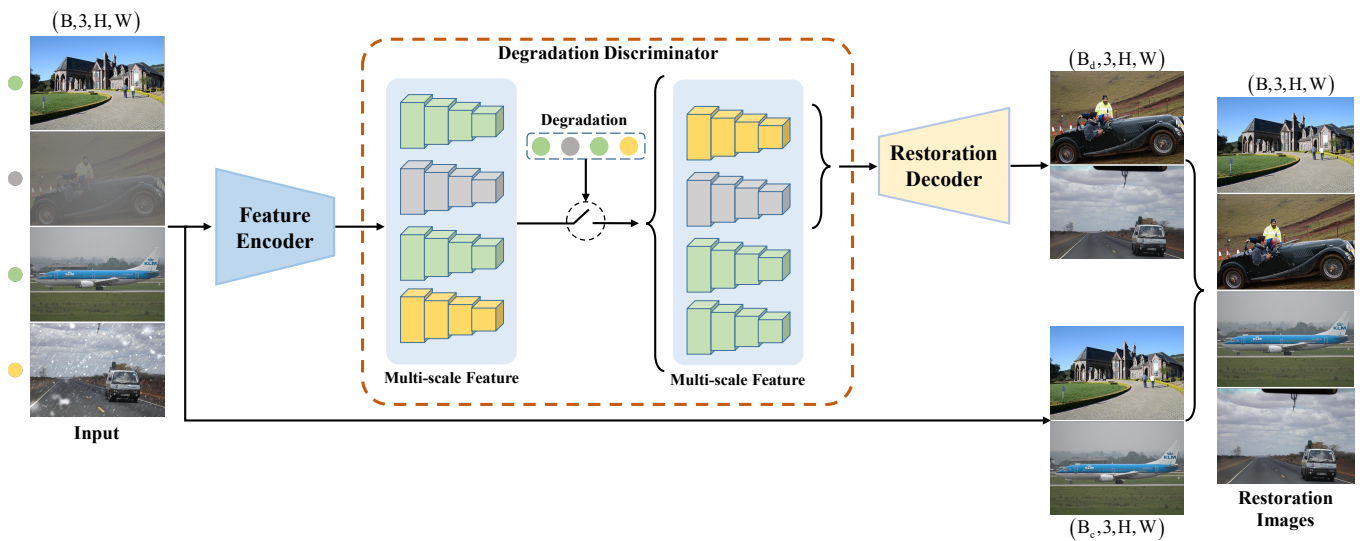


Figure 3. The restoration process of our DTRDNet. It can be seen as an encode–decode structure with the proposed degradation discriminator (DD) between encoder and decoder. The feature encoder adopts the SFE. Restoration decoder and DD make up our DDIR. B_d and B_c refer to the number of degraded and clear images of a batch of input.

3.4. Degradation Category Predictor

We incorporate degradation type information into the DDIR to bypass the process of restoring clear images, thereby mitigating its impact on model parameter updating. The primary objective remains extracting clear image features from degraded images. However, in reality, the features of a specific target vary under different weather conditions. This insight motivates us to recognize that for training a detection model adapting to various weather conditions, the model must possess weather awareness capabilities. Consequently,

we employ degradation type information as labels to train a degradation category classifier capable of predicting the weather in input images, referred to as DCP.

As exhibited in Figure 2, the DCP leverages the features of the SFE to predict the degradation category, aiming to avoid increased model complexity and learn degradation category awareness of the SFE to facilitate object detection in diverse weather conditions. The DCP incorporates a simple classifier that consists of a convolution layer with a 1×1 kernel size for adjusting the number of feature channels, a spatial pooling layer (Pool) for extracting global features for classification, and a fully connected layer (FC) for generating prediction results. The predicted degradation category (O^D) of the DCP can be formulated as follows:

$$O^D = FC(Pool(Conv_{1,1}(f_{\times 32}))). \quad (12)$$

3.5. Loss Function

To jointly train the entire framework for object detection, image restoration, and degradation classification, we introduce three distinct loss functions corresponding to the output of each decoder. Specific details are provided below.

Object Detection Loss: We evaluate the disparity between the object detection results and their corresponding labels to compute the detection loss, which consists of three components: (1) \mathcal{L}_{box} for the positional variance between predicted and actual bounding boxes, (2) \mathcal{L}_{cls} for target category prediction, and (3) \mathcal{L}_{obj} for bounding box confidence. This can be expressed as follows:

$$\mathcal{L}_{Det} = \lambda \mathcal{L}_{box} + \mathcal{L}_{cls} + \mathcal{L}_{obj}, \quad (13)$$

where the weight λ for \mathcal{L}_{box} is set to 5. The losses \mathcal{L}_{box} and \mathcal{L}_{cls} are computed using IOU loss and cross-entropy function, respectively. The loss function \mathcal{L}_{obj} adopts focus loss [47] to address the issue of imbalanced positive and negative samples in object detection tasks.

Image Restoration Loss: To acquire information for image restoration, we compute the mean square error (MSE) between the input and restoration result as our restoration loss \mathcal{L}_{Res} . It is important to note that, since both clear images and restored degraded images may be present in the final restoration result, the MSE of the corresponding clear image in the input X and restoration result R is zero. Therefore, the restoration loss \mathcal{L}_{Res} can be represented as follows:

$$\mathcal{L}_{Res} = \frac{1}{N} \sum_{i=1}^N \|X - R\|_2 = \frac{1}{N_D} \sum_{i=1}^{N_D} \|X_C^D - R^D\|_2, \quad (14)$$

where N is the batch size of input. N_D refers to the number of degraded images in a batch. R^D and X_C^D represent the restored degraded image and its corresponding clear image target, respectively.

Degradation Classification Loss: To make the network possess degradation category awareness, we introduce the cross-entropy function as the degradation classification loss \mathcal{L}_{dc} to assess whether the predicted results of DIP are consistent with the actual label, expressed as

$$\mathcal{L}_{dc} = - \sum_{i=1}^K y_i \log(O_i^D), y_i = \begin{cases} 1 & i = t \\ 0 & otherwise, \end{cases} \quad (15)$$

where K is set to 4, representing the degradation conditions of images, including clean, hazy, rainy, and snowy. t refers to the true label of degradation type.

Based on the above definition, the complete loss of our DTRDNet is formulated as

$$\mathcal{L} = \alpha \mathcal{L}_{Det} + \beta \mathcal{L}_{Res} + \gamma \mathcal{L}_{dc}, \quad (16)$$

where α , β , and γ are the balance coefficients for object detection loss, restoration loss, and degradation classification loss, respectively. We follow the set of TogetherNet [6] with $\alpha = 0.2$ and $\beta = 0.8$. The γ is set to 0.01. Based on the principle of back propagation, the parameters of the three decoders will be updated according to their respective loss functions to produce the desired output. The SFE will be optimized using three simultaneous loss functions, allowing it to extract and encode information for object detection, image restoration, and degradation classification. The optimization for image restoration ensures that the SFE can extract clear image features, even from degraded input images. Similarly, the regression for predicting degradation categories endows the shared feature with awareness of different degradation types. Both clear image features and degradation category awareness contribute to the performance and robustness of the detector under various weather conditions.

4. Experiment and Results

This section presents a comprehensive analysis of our experiments. We commence by detailing the experimental setup, which includes datasets and implementation specifics. Subsequently, we provide both quantitative and qualitative evaluations of our proposed DTRDNet, comparing its performance against other state-of-the-art object detection algorithms. Finally, we conduct an ablation experiment to validate the efficacy of our improvements.

4.1. Datasets

Train and Test Dataset. (1) VOC-FOG. This is a synthesized hazy dataset derived from the VOC dataset [11] proposed by TogetherNet [6], comprising a total of 11,707 hazy images across five object categories (namely, car, bus, motorbike, and person), with 9578 images designated for training (VOC-FOG-train) and 2129 for testing (VOC-FOG-test).

Train and Test Dataset. (2) VOC-Clean. We select the images utilized to synthesize the VOC-FOG dataset to constitute a clear image dataset, ensuring that it encompasses the same object categories and number of images as VOC-FOG. The training set, comprising 9578 images (VOC-Clean-train), is integrated with the degraded image training set for model training. The performance of the detector on clear images is assessed using its testing set (VOC-Clean-test), which consists of 2129 images.

Train and Test Dataset. (3) VOC-Rain. Similar to VOC-FOG, we employed the rain streak synthesis method of RainDS [59] to synthesize rainy images using the images from VOC-Clean. Furthermore, in order to more accurately simulate the effects of rainy days on imaging results, we synthesized various types of rain streaks with different orientations and intensities. The synthesized rainy images are divided into a training set (VOC-Rain-train) comprising 9578 images and a testing set (VOC-Rain-test) comprising 2129 images.

Train and Test Dataset. (4) VOC-Snow. We generate snowy images by combining high-quality images from the VOC-Clean dataset with snow masks from the CSD dataset [60]. The snow masks are applied to the high-quality images using a random weight ranging from 0.5 to 1.0, simulating varying intensities of snow. Subsequently, we create a snowy dataset (VOC-Snow) comprising 9578 training images (VOC-Snow-train) and 2129 testing images (VOC-Snow-test). Similarly, both the training and testing sets consist of five classes: car, bus, motorbike, bicycle, and person.

Real Scenario Test Dataset. (1) RTTS [31]. Due to the difficulty of capturing paired hazy-clean images in real-world scenarios, RTTS was proposed to evaluate the performance of dehazing algorithms from a task-driven perspective. This is a comprehensive dataset for object detection in real-world hazy scenes, containing 4322 hazy images annotated with five object categories (car, bus, motorbike, and person) corresponding to those in the training set.

Real Scenario Test Datasets. (2) Foggy Driving dataset [61]. This dataset provides a practical repository of fog-affected scene images tailor-made for applications in object detection and semantic segmentation, and contains a total of 101 authentic fog-affected

scene images annotated with 466 vehicle instances (including car, bus, train, truck, bicycle, and motorcycle) and 269 human instances. It is noteworthy that the dataset includes eight different categories of annotated instances, with only five of these categories selected for testing to maintain consistency with the training set.

4.2. Implementation Details

The implementation of this paper is conducted on the Ubuntu operating system using an NVIDIA 3090 GPU and the PyTorch framework. During training, we employ stochastic gradient descent (SGD) as the optimizer to update the model parameters. The initial learning rate is set to 1×10^{-2} , with a momentum of 0.937 and a weight decay of 5×10^{-4} . Additionally, we utilize a cosine annealing strategy [62] to reduce the learning rate for improved convergence. It is important to emphasize that the inclusion of various categories of degraded images will lead to a multiplicative increase in data volume. The utilization of a wider range of degradation types for training will result in an increase in the times of parameter updates. Consequently, to ensure fairness, different quantities of training datasets will be associated with diverse epoch settings in subsequent experiments. In terms of input, the spatial dimensions of all images are resized to 640×640 during both training and testing stages, with a fixed batch size of eight. Additionally, since mosaic data augmentation can disrupt the degradation type of the input and complicate the training of the restoration network, we close mosaic augmentation during the training phase. In contrast to universal detection algorithms, our proposed method employs clear images to supervise the training of the restoration branch. Consequently, we apply identical data enhancement to the input image and its corresponding clear image, including hue saturation value (HSV), random scaling, translation, and flipping. The augmented clear images are used as targets for restoration loss.

To comprehensively assess the performance of our algorithm across various weather conditions, including clear and adverse scenarios such as haze, rain, and snow, we conduct separate implementations for single degradation weather and diverse weather conditions. We then compare these results with those obtained from other advanced adverse weather image detection models.

4.3. Implementation Results

We assess the detection performance of our DTRDNet using the widely adopted mean average precision (mAP) metric across multiple target categories, with an intersection over union (IOU) set to 0.5. The comprehensive investigation are conducted under both single and multiple degradation conditions, with the results summarized in the table below. In the tables presented in this paper, the best results are highlighted in bold, while the second-best results are indicated by underlining. The following section presents comparative experiments of our proposed method in these scenarios, with our ablation experiments provided at the end.

4.3.1. Hazy Scene Object Detection

To assess the performance of the proposed DTRDNet in hazy conditions, we conducted a comparative analysis with other state-of-the-art object detection algorithms designed for adverse weather scenarios. Generally, these comparative methods can be categorized into three parts. (i) The first part includes the common “object detection (OD)” YOLOXs models as baselines, which are trained on degraded images (VOC-FOG-train) and clear images (VOC-Clean-train), respectively, denoted as YOLOXs and YOLOXs*. (ii) The second part is the “IR + OD” approach, which preprocesses the hazy images by the dehazing algorithms to produce clear images, and then detects the clear images using the YOLOXs [30] model trained on VOC-Clean-train. Five mainstream image dehazing algorithms are selected for this purpose, incorporating DCP [12], AODNet [63], FFANet [50], AECRNet [13], and RestorNet [14]. We combine these methods with YOLOXs* [30] to form “IR + OD” type algorithms, named DCP-YOLOXs*, AODNet-YOLOXs*, FFANet-YOLOXs*, AECRNet-

YOLOXs*, and RestorNet-YOLOXs*, respectively. It is worth noting that all the dehazing algorithms are trained on the OTS dataset from the RESIDE [31] dataset to ensure data consistency with RTTS [31]. (iii) The third part refers to the “Object Detection in Adverse Weather (ODAW)” algorithms that do not require the use of restoration algorithms or other preprocessing steps, including DS-Net [64], IA-YOLO [24], and TogetherNet [6]. To ensure a fair comparison, we retrained the aforementioned “ODAW” algorithms using the VOC-FOG-training dataset. As our framework requires information on different degradation types, we combine the VOC-Clean-train and VOC-FOG-train datasets to form our training set for hazy image object detection. We train the model for 50 epochs to mitigate overfitting that may arise from doubling our training images, while other methods are trained for 100 epochs. Subsequent experiments also adjust the number of training epochs based on the combination of datasets.

Object Detection Results on Synthetic Hazy Image Dataset. We first employ the synthetic dataset VOC-FOG-test to validate the performance of the proposed DTRDNet and the comparative algorithms. The results are presented in Table 1. It is evident that our proposed DTRDNet achieves a mAP of 79.86%, which is 1.72% higher than the recently advanced TogetherNet [6] dedicated to foggy object detection, with all categories achieving either the highest or second-highest average accuracy. Overall, our algorithm demonstrates superior detection performance and narrows the accuracy gap across different object classes.

Table 1. Object detection results on the VOC-FOG-test dataset [6] of synthetic hazy images. The type of algorithms are denoted by OD (“Object Detection”), IR + OD (“Image Restoration + Object Detection”), and ODAW (“Object Detection in Adverse Weather”). “*” means the model is trained on the degradation-free images, while the methods without “*” mean they are trained on degraded images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Class (AP (%))					mAP (%)
		Person	Bicycle	Car	Motorbike	Bus	
YOLOXs [30]	OD	67.67	83.28	77.75	68.91	81.70	75.86
YOLOXs* [30]	OD	73.09	57.22	69.55	59.83	77.34	67.41
DCP-YOLOXs* [12]	IR + OD	81.84	70.38	78.63	<u>73.48</u>	84.68	77.80
AODNet-YOLOXs* [63]	IR + OD	67.40	49.19	60.51	55.59	62.07	58.95
AECRNet-YOLOXs* [13]	IR + OD	80.47	67.82	76.97	72.46	82.73	76.09
RestorNet-YOLOXs* [14]	IR + OD	78.71	67.15	72.56	71.68	82.36	74.49
DS-Net [64]	ODAW	72.44	60.47	81.27	53.85	61.43	65.89
IA-YOLO [24]	ODAW	70.98	61.98	70.98	57.93	61.98	64.77
TogetherNet [6]	ODAW	<u>84.11</u>	69.26	79.59	72.12	85.62	<u>78.14</u>
DTRDNet (Ours)	ODAW	84.34	<u>72.96</u>	<u>80.87</u>	74.83	<u>84.34</u>	79.86

Object Detection Results on Real Scene Hazy Image Dataset. To demonstrate the effectiveness of our approach in real-world hazy conditions, we further conduct comparative experiments on the Foggy Driving dataset [61] and the RTTS dataset [31]. As shown in Tables 2 and 3, our method achieved the highest accuracy on both datasets. Specifically, for the Foggy Driving dataset [61], we observed an increase of 2.77% in mAP compared to YOLOXs and a 5.75% improvement over TogetherNet [6]. For the RTTS dataset, our algorithms also demonstrated significant improvements. Compared to YOLOXs [30] and TogetherNet [6], there are increases of 5.64% and 6.06% in mAP metrics, respectively. This significant improvement on the test set of real scenarios proves that our method can effectively mitigate overfitting.

Table 2. Object detection results on the Foggy Driving dataset [61] of real-world hazy images. “*” means the model is trained on the degradation-free images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Class (AP (%))					mAP (%)
		Person	Bicycle	Car	Motorbike	Bus	
YOLOXs [30]	OD	24.37	<u>22.33</u>	55.57	14.29	37.34	<u>30.78</u>
YOLOXs* [30]	OD	21.48	18.84	54.67	1.59	30.33	25.38
DCP-YOLOXs* [12]	IR + OD	21.57	17.85	55.30	3.57	39.92	27.64
AODNet-YOLOXs* [63]	IR + OD	20.66	20.92	54.03	4.76	24.78	25.03
AECRNet-YOLOXs* [13]	IR + OD	23.00	19.96	54.83	2.38	35.81	27.20
RestorNet-YOLOXs* [14]	IR + OD	23.24	18.81	53.71	2.38	35.70	26.77
DS-Net [64]	ODAW	26.74	20.54	<u>58.16</u>	7.14	36.11	29.74
IA-YOLO [24]	ODAW	16.20	11.76	41.43	4.76	17.55	18.34
TogetherNet [6]	ODAW	25.70	18.79	57.72	7.14	29.65	27.80
DTRDNet (Ours)	ODAW	<u>26.21</u>	30.28	58.58	14.29	<u>37.37</u>	33.55

Table 3. Object detection results on the RTTS [31] of real-world hazy images. “*” means the model is trained on the degradation-free images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Class (AP (%))					mAP (%)
		Person	Bicycle	Car	Motorbike	Bus	
YOLOXs [30]	OD	<u>81.78</u>	56.70	70.23	<u>49.48</u>	31.57	<u>57.95</u>
YOLOXs* [30]	OD	80.28	50.75	68.23	41.89	28.89	54.01
DCP-YOLOXs* [12]	IR + OD	81.16	51.34	71.13	47.20	31.09	56.38
AODNet-YOLOXs* [63]	IR + OD	78.16	44.49	65.54	38.88	25.57	50.53
AECRNet-YOLOXs* [13]	IR + OD	80.85	52.75	68.60	46.29	30.43	55.78
RestorNet-YOLOXs* [14]	IR + OD	77.48	51.43	60.92	43.12	29.16	52.42
DS-Net [64]	ODAW	68.81	18.02	46.13	15.15	15.44	32.71
IA-YOLO [24]	ODAW	67.25	35.28	41.14	20.97	13.64	35.66
TogetherNet [6]	ODAW	81.14	<u>54.12</u>	<u>72.39</u>	47.02	<u>32.99</u>	57.53
DTRDNet (Ours)	ODAW	83.22	57.86	76.71	61.17	38.97	63.59

Subsequently, we performed a joint analysis based on the findings presented in Tables 1–3. It is evident from Table 1 that models trained on clear images exhibit limited adaptability to degraded images when comparing YOLOXs and YOLOXs*. Furthermore, the “IR + OD” algorithms demonstrate some improvement in detection performance for YOLOXs* under general haze conditions. However, the results depicted in Tables 2 and 3 indicate that the efficacy of “IR + OD” algorithms is notably constrained when applied to real-world hazy images. This limitation may be attributed to a domain gap between synthetic hazy datasets used for dehazing model training and real-world hazy images, resulting in limited or adverse effects of the “IR + OD” method on actual hazy weather detection performance. In contrast, our algorithm outperforms the “OD”, “IR + OD”, and “ODAW” methods under single haze degradation conditions, demonstrating superior detection performance across both synthetic and real-world hazy scenes. Notably, our approach exhibits particularly obvious improvements in authentic foggy images, indicating our method has excellent generalization in real scenarios.

In addition to the above quantitative analysis, we also present the visual detection results of comparative algorithms on the VOC-FOG-test [11], Foggy Driving [61], and

RTTS [31] datasets in Figure 4. The detection results show that our method can correctly detect more targets with higher confidence, further testifying to the superiority of the proposed algorithm in both synthetic hazy images and real scene hazy images compared to the comparative algorithms.



Figure 4. Visualization of object detection results of hazy images on both synthetic and real-world datasets. We show the images from the VOC-FOG-test [6], RTTS [31], and Foggy Driving datasets [61], respectively, from first to third row.

4.3.2. Rainy Scene Object Detection

To evaluate the performance of the proposed DTRDNet in rainy image object detection, we conduct a comparative experiment using synthetic rainy images. Similarly to hazy scene object detection, we combine VOC-Clean-train and VOC-Rain-train as the training set to optimize our proposed model over 50 epochs for rainy object detection, while the comparative algorithms are trained on VOC-Rain-train for 100 epochs. We select YOLOXs trained on hazy images and YOLOXs* trained on clear images as the object detection (OD) methods. For IR + OD, we employ the restoration algorithms AirNet [51] and RestorNet [14] to preprocess the rainy images before object detection, titled the AirNet-YOLOXs* and RestorNet-YOLOXs*. For ODAW comparison, TogetherNet [6] is utilized for comparison. The results of these aforementioned object detection algorithms on rainy images are presented in Table 4.

As shown in Table 4, our approach demonstrates superior performance across all object categories, with the exception of “Bus”, where it also achieves near-optimal accuracy. Our method outperforms the advanced TogetherNet [6], increasing mAP by 1.38%, indicating a comprehensive enhancement in model accuracy. Notably, for “Bicycle” and “Motorbike”, which exhibited lower accuracy compared to other object classes, our algorithm improves their AP by 2.69% and 3.4%, respectively, resulting in more balanced detection performance across different categories. Additionally, Figure 5 presents visual detection results in rainy conditions, demonstrating that our proposed method can accurately detect more targets with higher confidence. These results confirm the effectiveness and superiority of our proposed approach for object detection in rainy conditions.

Table 4. Object detection results on the VOC-Rain-test of synthetic rainy images. “*” means the model is trained on the degradation-free images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Class (AP (%))					mAP (%)
		Person	Bicycle	Car	Motorbike	Bus	
YOLOXs [30]	OD	80.98	65.88	74.54	70.83	84.35	75.32
YOLOXs* [30]	OD	75.94	63.00	66.92	65.32	73.10	68.86
AirNet-YOLOXs* [51]	IR + OD	80.06	67.81	70.75	69.87	82.77	74.25
RestorNet-YOLOXs* [14]	IR + OD	80.44	68.29	71.68	70.16	82.54	74.62
TogetherNet [6]	ODAW	<u>83.53</u>	<u>69.83</u>	<u>77.88</u>	<u>74.26</u>	86.41	<u>78.38</u>
DTRDNet (Ours)	ODAW	83.87	72.52	78.59	77.66	<u>86.15</u>	79.76

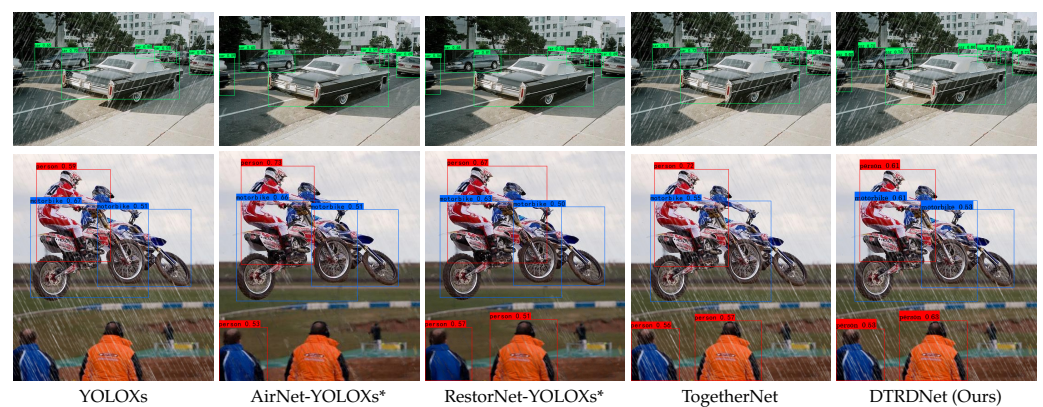


Figure 5. Visualization of object detection results of rainy images on VOC-Rain-test.

4.3.3. Snowy Scenes Object Detection

For snowy image object detection, we adopt the approach used for hazy and rainy images by leveraging the VOC-Snow-train and VOC-Clean-train datasets to train the proposed algorithms for 50 epochs, and VOC-Snow-train to train comparative algorithms for 100 epochs, followed by testing on VOC-Snow-test. The baseline method (OD) comprises YOLOXs and YOLOXs*, trained separately with snowy and clear images. In the IR + OD method, TKL [17] and LMQFormer [18] are chosen as image restoration methods for preprocessing snowy images, resulting in TKL-YOLOXs* and LMQFormer-YOLOXs*. For the ODAW method, TogetherNet is selected for comparison [6]. The specific outcomes are detailed in Table 5 and Figure 6.

According to Table 5, our DTRDNet demonstrates superior performance in the mAP metric and AP for each category compared to other methods. Specifically, the mAP shows a 2.36% increase compared to TogetherNet [6], with notable improvements in the “Motorbike” category, which exhibits a 4.56% increase in AP. Moreover, Figure 6 presents visual results of object detection on snowy images using comparative methods, further validating the effectiveness of our proposed approach. Our method excels at detecting multiple targets within complex scenes containing mixed categories and achieves higher accuracy in snow images.

4.3.4. Various Weather Scenes Object Detection

Herein above, extensive experiments have been conducted to assess the performance of the proposed DTRDNet under three adverse weather conditions: haze, rain, and snow. To further evaluate the object detection capabilities of our method across various weather conditions, we perform the experiments using a mixed dataset comprising VOC-Clean, VOC-FOG, VOC-Rain, and VOC-Snow. Specifically, the universal models are trained on the training set of the mixed dataset using our proposed method and comparison methods. The models

are then tested on the testing dataset with different degradation to evaluate their performance across multiple weather conditions. Given that the “IR + OD” type is not suitable for images depicting multiple weather scenes, we exclusively employed the “OD” and “ODAW” methods for comparison. Similarly to single degradation weather image object detection, YOLOXs* was trained on clear images for 100 epochs, while YOLOXs was trained on a mixed degradation dataset. As for “ODAW”, TogetherNet [6] is chosen for comparison. It should be noted that due to an expanded dataset in this part of the experiment, we set the training epochs to 25 for models trained on a mixed dataset. Table 6 and Figure 7 present quantitative and visual detection results across four different weather conditions.

Table 5. Object detection results on the VOC-Snow-test of synthetic snowy images. “*” means the model is trained on the degradation-free images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Class (AP (%))					mAP (%)
		Person	Bicycle	Car	Motorbike	Bus	
YOLOXs [30]	OD	81.16	66.64	75.43	70.87	83.28	75.48
YOLOXs* [30]	OD	78.40	64.88	70.80	56.90	81.15	70.43
TKL-YOLOXs* [51]	IR + OD	81.02	70.14	75.53	70.33	84.42	76.29
LMQFormer-YOLOXs* [14]	IR + OD	81.36	70.44	77.57	71.61	84.35	77.07
TogetherNet [6]	ODAW	<u>82.20</u>	<u>69.70</u>	<u>77.72</u>	<u>71.39</u>	<u>85.30</u>	<u>77.26</u>
DTRDNet (Ours)	ODAW	83.97	72.15	79.57	75.95	86.46	79.62



Figure 6. Visualization of object detection results of snow images on VOC-Snow-test dataset.

Table 6. Object detection results under multiple weather conditions. We show the detection results of clean, hazy, rainy, and snowy images tested on VOC-Clean-test, VOC-FOG-test, VOC-Rain-test, and VOC-Snow-test, respectively. “+” refers to the model trained on VOC-FOG-train, VOC-Rain-train, and VOC-Snow-train. “*” means the model is trained on the degradation-free images. The best results are bolded, and the sub-optimal results are indicated by underlining.

Method	Type	Datasets (mAP (%))				Average
		Clean	FOG	Rain	Snow	
YOLOXs [30]	OD	75.77	74.50	75.50	74.24	75.00
YOLOXs* [30]	OD	79.09	68.86	67.41	70.43	71.45
TogetherNet [6]	ODAW	78.86	76.74	76.78	77.03	77.35
DTRDNet(Ours) †	ODAW	<u>80.16</u>	<u>79.34</u>	<u>78.83</u>	78.81	<u>79.28</u>
DTRDNet (Ours)	ODAW	80.38	79.37	78.96	<u>78.80</u>	79.38

It is evident from Table 6 that the proposed method achieves superior results in object detection across various weather conditions. In comparison to TogetherNet, our method demonstrates an increase of 1.52%, 2.63%, 2.18%, and 1.77% in mAP for clean, foggy, rainy, and snowy images, respectively, resulting in an average mAP improvement of 2.03% across all four weather scenarios. Furthermore, our approach outperforms YOLOXs* in clear image object detection as well, highlighting its effectiveness in optimizing adverse weather image object detection while enhancing performance on clear images. On this basis, we further analyze the impact of introducing clear images on our method by training our model in VOC-FOG-train, VOC-Rain-train, and VOC-Snow-train, with the training epoch set to 33. The two experimental results of our DTRDNet show that the introduction of clear images has little effect on hazy and snowy images, which increases mAP by 0.03% in VOC-FOG-test and decreases by 0.01% in VOC-Snow-test. For VOC-Clean-test and VOC-Rain-test, there is an mAP boost of 0.22% and 0.13%, respectively, indicating that incorporating clear images into the training of our DTRDNet enhances detection performance for both clear and degraded images. This improvement arises because our method can better learn precise target features from clear images and eliminate interference between the clear image and the restoration module.

The visual results presented in Figure 7 illustrate the capability of our algorithm to detect multiple categories of targets under different weather conditions. In summary, our DTRDNet effectively accomplishes object detection across diverse weather conditions with a single trained model, thereby proposing a novel solution to tackle the challenges inherent in real-world object detection across diverse weather conditions.

4.3.5. Ablation Studies

The above experimental results demonstrate the superiority of our method in single and multiple degraded scene object detection. Furthermore, we conducted extensive ablation experiments to evaluate the effectiveness of our optimizations over DTRDNet, analyzing each of the improvements under real degradation conditions and synthetic multiple degraded scenes.

Effect of different components in real degraded scene. To assess the impact of each improvements in practical application scenarios, we selected the real hazy dataset RTTS [31] as a representative. As depicted in Table 7, we first reproduce the adverse weather image object detector TogetherNet [6], which yielded a mAP of 57.53%. Subsequently, by mixing the training set with clean images from VOC-Clean-train, we optimized TogetherNet and observed an increase in mAP by 1.34%, indicating that incorporating clear images into the training set contributes to object detection under real adverse weather conditions. This may be attributed to the model can better learn target features from clear images while adapting to degraded environment within mixed data. Furthermore, introducing a DD effectively mitigates interference from clear images on the restoration branch in TogetherNet, resulting in a notable 3.11% increase in mAP. Building upon this improvement, we further improve the restoration decoder to form our DDIR, which boosts mAP by 0.71%. Lastly, integrating the DCP gives an additional 0.9% boost in mAP. The number of epochs is set to 100 for training bare TogetherNet and to 50 for subsequent stages due to mixing clean and hazy images within the training set.

Effect of different components in synthetic multiple degradation condition. To assess the effectiveness of our enhancements in various weather image object detection, we also perform ablation using the multiple weather dataset comprising VOC-Clean, VOC-FOG, VOC-Rain, and VOC-Snow. As shown in Table 8, compared to the baseline YOLOXs, TogetherNet [6] demonstrates an obvious improvement of 2.37% in the mAP metric. Building upon this progress, we first integrate the DD to ensure that the restoration decoder exclusively processes degraded images. This approach eliminates mutual interference between clear images and the restoration decoder, resulting in an additional 1.55% increase in mAP performance. Furthermore, the improved restoration decoder to construct the DDIR facilitates clearer image feature extraction from degraded images, leading to a further

0.28% increase in the mAP metric. Finally, incorporating a DCP to learn degradation category awareness boosts mAP by 0.19%, indicating that models equipped with degradation type awareness can achieve higher precision for object detection under various weather conditions. Moreover, as our model is capable of extracting degradation type information from images, it can also output results from the DCP to assess environmental information from images if needed, which achieves a top1 precision of 99.27% for predicting the degradation category among the four weather conditions. In terms of model complexity, all our improvements can be removed during the inference stage without imposing any additional burden when conducting only object detection tasks. If environmental information is desired alongside object detection, this would only require an additional computational cost of 0.27 G.

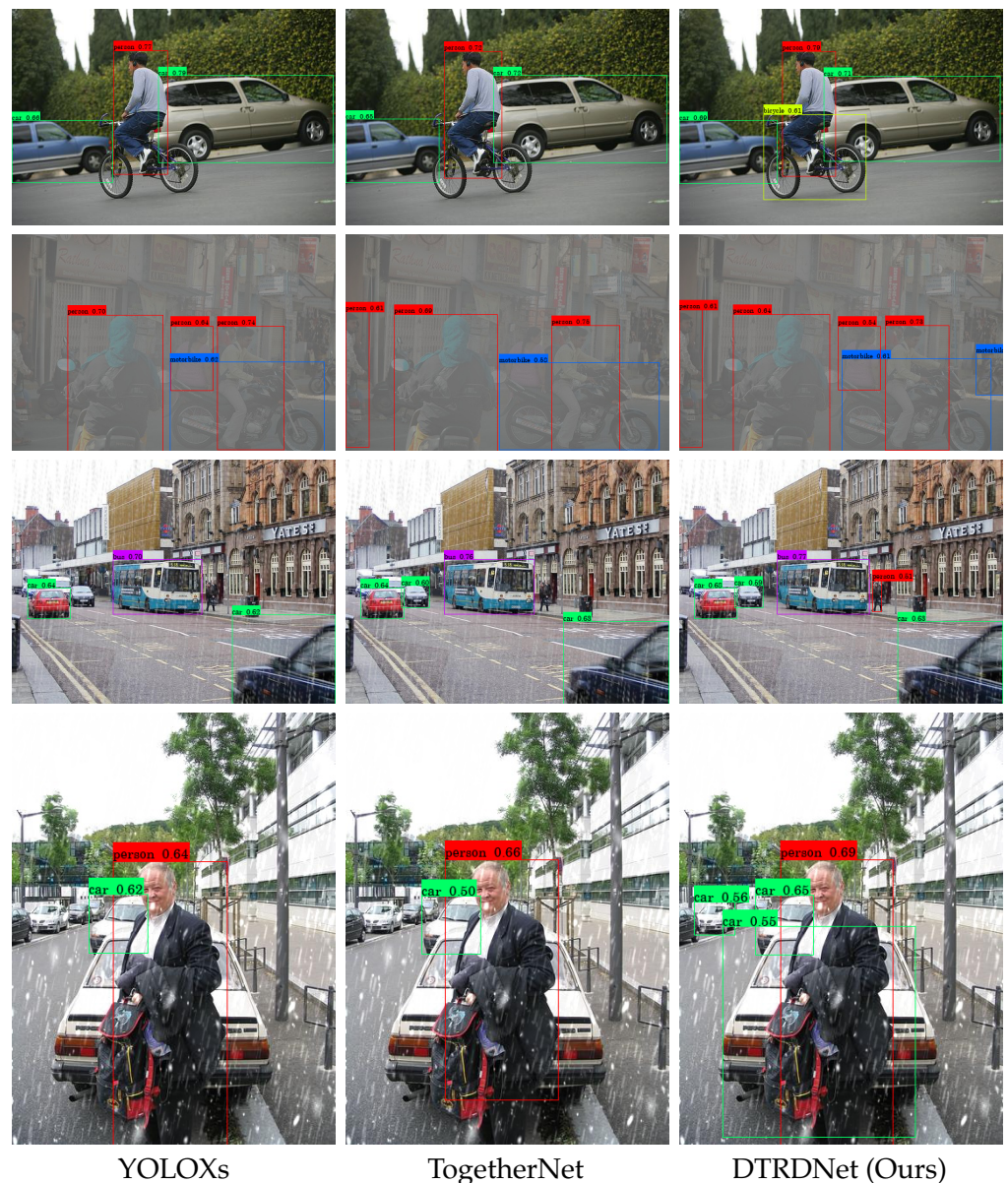


Figure 7. Visualization results of object detection in various weather conditions, including clean, hazy, rainy, and snowy scenes from top to bottom.

Table 7. The ablation experiments on real hazy dataset RTTS. “Mix” refers to the model trained on the mixed dataset of VOC-FOG-train and VOC-Clean-train. “DD” and “Res” express degradation discriminator and our improved image restoration decoder, respectively. “DD” and “Res” combine to form DDIR. “DCP” stands for degradation category predictor.

Model	Mix	DD	Res	DCP	mAP (%)
TogetherNet [6]	✗	✗	✗	✗	57.53
TogetherNet + Mix	✓	✗	✗	✗	58.87
TogetherNet + Mix + DD	✓	✓	✗	✗	61.98
TogetherNet + Mix + DDIR	✓	✓	✓	✗	62.69
TogetherNet + Mix + DDIR + DCP	✓	✓	✓	✓	63.59

Table 8. The ablation experiments on the multiple degradation dataset mixed by VOC-Clean, VOC-FOG, VOC-Rain, and VOC-Snow. “Top1” refers to the precision of degradation classifying. “✗” indicates the FLOPs for conducting detection and degradation type classification synchronously.

Model	DD	Res	DCP	mAP (%)	Top1	FLOPs (G)
YOLOXs [30]	✗	✗	✗	74.85	–	13.38
TogetherNet [6]	✗	✗	✗	77.22	–	17.29
TogetherNet + DD	✓	✗	✗	78.77	–	17.29
TogetherNet + DDIR	✓	✓	✗	79.05	–	17.29
TogetherNet + DDIR + DCP	✓	✓	✓	79.24	99.27	17.29/17.56 *

5. Limitations

Although the proposed methodology leverages degradation type information to enhance detection performance across diverse weather conditions, there are still certain limitations that warrant further improvement. First, the scarcity of clear-degraded image pairs necessitates the use of synthetically degraded images for model training. While this strategy is practical, it may potentially reduce the model’s ability to adapt to real-world adverse weather scenarios. Specifically, as illustrated in Figure 8, certain targets become virtually imperceptible under extreme weather conditions, significantly complicating detection and leading to missed detections. Secondly, our approach incorporates degradation information specific to different weather types, but does not account for the varying intensities of degradation within identical weather conditions. Consequently, the proposed method falls short in assessing the severity of these conditions. This limitation is exemplified by the fluctuating detection performance illustrated in Figure 9.

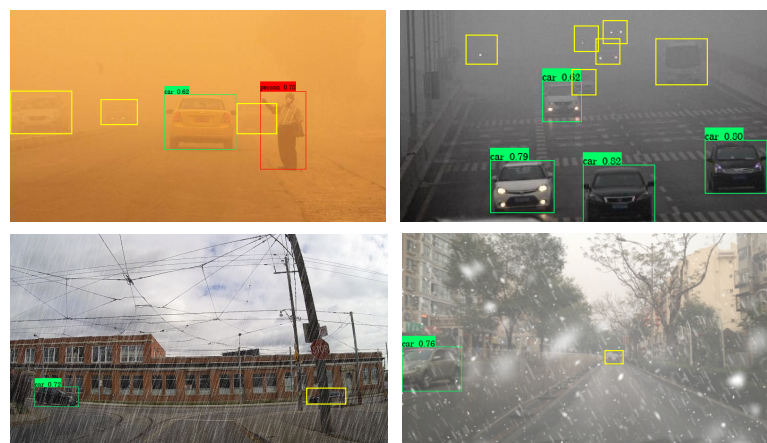


Figure 8. The failed case in extreme weather. The yellow boxes indicate targets that are undetected.

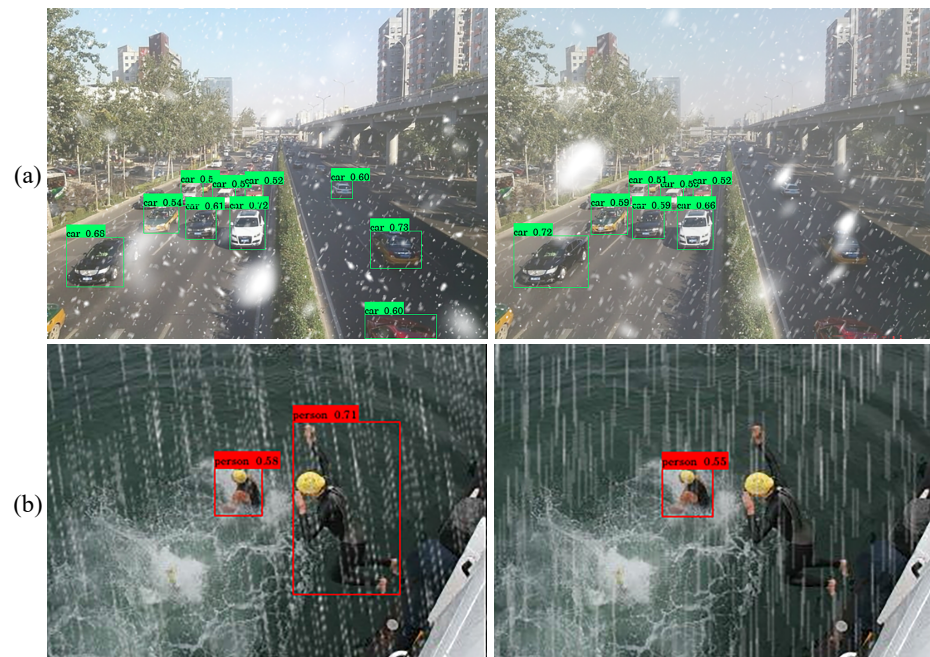


Figure 9. The detection performance difference caused by the varying of a certain degradation in snow and rain weather conditions. (a) and (b) refer to snow and rain images, respectively.

6. Conclusions

TogetherNet [6] integrates restoration and detection tasks to enhance the detection performance of single degraded image object detection. Building upon this, our research introduces DTRDNet, which aims to address the object detection challenge under multiple complex weather conditions using a unified model. We train the model on a mixed dataset encompassing multiple weather conditions to enhance its adaptability to various weather conditions. Under these circumstances, the image restoration module designed for degraded images does not contribute to detecting clear images, and the clean images also disrupt the restoration decoder during training. Consequently, we propose a degradation discrimination image restoration decoder (DDIR) to improve the restoration branch of TogetherNet [6], introducing a degradation discriminator (DD) that leverages degradation type labels to mitigate interference between undegraded images and the restoration decoder, while further optimizing it to extract more detailed image features from degraded images. However, direct application of degradation information is insufficient; hence, we introduce a degradation category predictor (DCP) and develop a joint learning framework by integrating degradation classification with image restoration and object detection to make the feature encoder learn degradation type awareness. This approach obviously enhances object detection performance under various weather conditions. The experiment results indicate that when employing a single model for the object detection across various degraded images, our approach significantly surpasses TogetherNet [6]. In the VOC-Clean, VOC-FOG, VOC-Rain, and VOC-Snow datasets, we observe improvements of 1.52%, 2.63%, 2.18%, and 1.77% in the mAP metric, respectively. In real hazy weather conditions, mAP increases by 6.06% and 5.75% on the RTTS [31] and Foggy Driving datasets [61], respectively, indicating that our method demonstrates greater applicability in actual adverse weather scenarios. Additionally, our model is capable of providing precise predictions of image degradation types at a minimal computational cost, if needed.

Our future work will focus on addressing the challenge of paired image training in unsupervised domain adaptation and evaluating the degree of degradation by incorporating text–image encoders. Additionally, we aim to enhance the performance of high-level visual tasks in real-world scenarios to better meet practical needs.

Author Contributions: Conceptualization, X.H. (Xiaochen Huang); methodology, X.H. (Xiaochen Huang); software, X.H. (Xiaochen Huang) and X.W.; validation, X.H. (Xiaochen Huang) and X.W.; formal analysis, X.H. (Xiaochen Huang) and H.C.; investigation, X.H. (Xiaochen Huang) and X.W.; resources, Q.T., H.C. and X.H. (Xiaohai He); data curation, X.W. and X.H. (Xiaochen Huang); writing—original draft preparation, X.H. (Xiaochen Huang); writing—review and editing, X.H. (Xiaochen Huang), H.C., and Q.T.; visualization, X.H. (Xiaochen Huang) and X.W.; supervision, Q.T.; project administration, H.C.; funding acquisition, H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan Science and Technology Program (2024YFHZ0212), the Open Foundation of Yunnan Key Laboratory of Software Engineering (2023SE206), Opening Foundation of Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology (TJUT-CVS20220001), China, the Fundamental Research Funds for the Central Universities (SCU2023D062 and 2022CDSN-15-SCU).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: Thanks to the editors and reviewers for their contributions to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Melnychenko, O.; Scislo, L.; Savenko, O.; Sachenko, A.; Radiuk, P. Intelligent Integrated System for Fruit Detection Using Multi-UAV Imaging and Deep Learning. *Sensors* **2024**, *24*, 1913. [[CrossRef](#)]
2. Qiu, Y.; Lu, Y.; Wang, Y.; Jiang, H. IDOD-YOLOV7: Image-Dehazing YOLOV7 for Object Detection in Low-Light Foggy Traffic Environments. *Sensors* **2023**, *23*, 1347. [[CrossRef](#)]
3. Bu, Y.; Ye, H.; Tie, Z.; Chen, Y.; Zhang, D. OD-YOLO: Robust SmObject Detection Model in Remote Sensing Image with a Novel Multi-Scale Feature Fusion. *Sensors* **2024**, *24*, 3596. [[CrossRef](#)]
4. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* **2023**, *23*, 7190. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, J.; Lei, J.; Xie, W.; Fang, Z.; Li, Y.; Du, Q. SuperYOLO: Super Resolution Assisted Object Detection in Multimodal Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5605415 [[CrossRef](#)]
6. Wang, Y.; Yan, X.; Zhang, K.; Gong, L.; Xie, H.; Wang, F.L.; Wei, M. Togethernet: Bridging image restoration and object detection together via dynamic enhancement learning. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2022; Volume 41, pp. 465–476.
7. Tang, W.; He, F.; Bashir, A.K.; Shao, X.; Cheng, Y.; Yu, K. A remote sensing image rotation object detection approach for real-time environmental monitoring. *Sustain. Energy Technol. Assess.* **2023**, *57*, 103270. [[CrossRef](#)]
8. Li, R.; Shen, Y. YOLOSRT: A deep learning method for small target detection in infrared remote sensing images based on super-resolution and YOLO. *Signal Process.* **2023**, *208*, 108962. [[CrossRef](#)]
9. Ergasheva, A.; Akhmedov, F.; Abdusalomov, A.; Kim, W. Advancing Maritime Safety: Early Detection of Ship Fires through Computer Vision, Deep Learning Approaches, and Histogram Equalization Techniques. *Fire* **2024**, *7*, 84. [[CrossRef](#)]
10. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
11. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
12. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
13. Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021*; pp. 10551–10560.
14. Wang, X.; Chen, H.; Gou, H.; He, J.; Wang, Z.; He, X.; Qing, L.; Sheriff, R.E. RestorNet: An efficient network for multiple degradation image restoration. *Knowl.-Based Syst.* **2023**, *282*, 111116. [[CrossRef](#)]
15. Yang, W.; Tan, R.T.; Feng, J.; Guo, Z.; Yan, S.; Liu, J. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1377–1393. [[CrossRef](#)] [[PubMed](#)]
16. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019*; pp. 3937–3946.

17. Chen, W.T.; Huang, Z.K.; Tsai, C.C.; Yang, H.H.; Ding, J.J.; Kuo, S.Y. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–22 June 2022; pp. 17653–17662.
18. Lin, J.; Jiang, N.; Zhang, Z.; Chen, W.; Zhao, T. LMQFormer: A laplace-prior-guided mask query transformer for lightweight snow removal. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 6225–6235. [[CrossRef](#)]
19. Wang, W.; Yan, D.; Wu, X.; He, W.; Chen, Z.; Yuan, X.; Li, L. Low-light image enhancement based on virtual exposure. *Signal Process. Image Commun.* **2023**, *118*, 117016. [[CrossRef](#)]
20. Liu, J.; Xu, D.; Yang, W.; Fan, M.; Huang, H. Benchmarking low-light image enhancement and beyond. *Int. J. Comput. Vis.* **2021**, *129*, 1153–1184. [[CrossRef](#)]
21. Wu, Y.; Guo, H.; Chakraborty, C.; Khosravi, M.R.; Berretti, S.; Wan, S. Edge computing driven low-light image dynamic enhancement for object detection. *IEEE Trans. Netw. Sci. Eng.* **2022**, *10*, 3086–3098. [[CrossRef](#)]
22. Wang, B.; Lu, T.; Zhang, Y. Feature-driven super-resolution for object detection. In Proceedings of the 2020 5th International Conference on Control, Robotics and Cybernetics (CRC), Penang, Malaysia, 21–23 November 2024; IEEE: Piscataway, NJ, USA, 2020; pp. 211–215.
23. Sindagi, V.A.; Oza, P.; Yasarla, R.; Patel, V.M. Prior-based domain adaptive object detection for hazy and rainy conditions. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 763–780.
24. Liu, W.; Ren, G.; Yu, R.; Guo, S.; Zhu, J.; Zhang, L. Image-adaptive YOLO for object detection in adverse weather conditions. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; Volume 36, pp. 1792–1800.
25. Hnewa, M.; Radha, H. Multiscale domain adaptive yolo for cross-domain object detection. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 3323–3327.
26. Rezaeianaran, F.; Shetty, R.; Aljundi, R.; Reino, D.O.; Zhang, S.; Schiele, B. Seeking similarities over differences: Similarity-based domain alignment for adaptive object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9204–9213.
27. Yunusov, N.; Islam, B.M.S.; Abdusalomov, A.; Kim, W. Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches. *Processes* **2024**, *12*, 1039. [[CrossRef](#)]
28. Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; Van Gool, L. Domain adaptive faster r-cnn for object detection in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3339–3348.
29. Gopalan, R.; Li, R.; Chellappa, R. Domain adaptation for object recognition: An unsupervised approach. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 999–1006.
30. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
31. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [[CrossRef](#)] [[PubMed](#)]
32. Kuldashboy, A.; Umirzakova, S.; Allaberdiev, S.; Nasimov, R.; Abdusalomov, A.; Im Cho, Y. Efficient image classification through collaborative knowledge distillation: A novel AlexNet modification approach. *Heliyon* **2024**, *10*, e34376. [[CrossRef](#)]
33. Ren, Z.; Tang, Y.; He, Z.; Tian, L.; Yang, Y.; Zhang, W. Ship detection in high-resolution optical remote sensing images aided by saliency information. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5623616. [[CrossRef](#)]
34. Li, J.; Xu, R.; Ma, J.; Zou, Q.; Ma, J.; Yu, H. Domain adaptive object detection for autonomous driving under foggy weather. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 612–622.
35. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
36. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
37. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
38. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra r-cnn: Towards balanced learning for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 821–830.
39. Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards high quality object detection via dynamic training. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 260–275.
40. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
41. Chen, S.; Lai, W.; Ye, J.; Ma, Y. A Fast and Low-Power Detection System for the Missing Pin Chip Based on YOLOv4-Tiny Algorithm. *Sensors* **2023**, *23*, 3918. [[CrossRef](#)]
42. Fan, X.; Xiao, D.; Li, Q.; Gong, R. Snow-CLOCs: Camera-LiDAR Object Candidate Fusion for 3D Object Detection in Snowy Conditions. *Sensors* **2024**, *24*, 4158. [[CrossRef](#)] [[PubMed](#)]

43. Huang, X.; Zhang, Y. ScanGuard-YOLO: Enhancing X-ray Prohibited Item Detection with Significant Performance Gains. *Sensors* **2024**, *24*, 102. [[CrossRef](#)] [[PubMed](#)]
44. Yin, X.; Yu, Z.; Fei, Z.; Lv, W.; Gao, X. Pe-yolo: Pyramid enhancement network for dark object detection. In Proceedings of the International Conference on Artificial Neural Networks, Heraklion, Greece, 26–29 September 2023; pp. 163–174.
45. Ding, Q.; Li, P.; Yan, X.; Shi, D.; Liang, L.; Wang, W.; Xie, H.; Li, J.; Wei, M. CF-YOLO: Cross fusion yolo for object detection in adverse weather with a high-quality real snow dataset. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 10749–10759. [[CrossRef](#)]
46. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
47. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
48. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
49. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
50. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
51. Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; Peng, X. All-in-one image restoration for unknown corruption. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17452–17462.
52. Wang, X.; Liu, X.; Yang, H.; Wang, Z.; Wen, X.; He, X.; Qing, L.; Chen, H. Degradation Modeling for Restoration-enhanced Object Detection in Adverse Weather Scenes. *IEEE Trans. Intell. Veh.* **2024**, 1–17. [[CrossRef](#)]
53. Shan, Y.; Lu, W.F.; Chew, C.M. Pixel and feature level based domain adaptation for object detection in autonomous driving. *Neurocomputing* **2019**, *367*, 31–38. [[CrossRef](#)]
54. Zhu, X.; Pang, J.; Yang, C.; Shi, J.; Lin, D. Adapting object detectors via selective cross-domain alignment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 687–696.
55. RoyChowdhury, A.; Chakrabarty, P.; Singh, A.; Jin, S.; Jiang, H.; Cao, L.; Learned-Miller, E. Automatic adaptation of object detectors to new domains using self-training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 780–790.
56. Zhang, S.; Tuo, H.; Hu, J.; Jing, Z. Domain adaptive yolo for one-stage cross-domain detection. In Proceedings of the Asian Conference on Machine Learning, Virtual, 17–19 November 2021; pp. 785–797.
57. Liu, J.J.; Hou, Q.; Cheng, M.M.; Wang, C.; Feng, J. Improving convolutional networks with self-calibrated convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 19 June 2020; pp. 10096–10105.
58. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
59. Quan, R.; Yu, X.; Liang, Y.; Yang, Y. Removing raindrops and rain streaks in one go. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9147–9156.
60. Chen, W.-T.; Fang, H.-Y.; Hsieh, C.-L.; Tsai, C.-C.; Chen, I.-H.; Ding, J.J.; Kuo, S.-Y. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 4196–4205.
61. Sakaridis, C.; Dai, D.; Van Gool, L. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vis.* **2018**, *126*, 973–992. [[CrossRef](#)]
62. Loshchilov, I.; Hutter, F.S. Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
63. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. Aod-net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4770–4778.
64. Huang, S.C.; Le, T.H.; Jaw, D.W. DSNet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2623–2633. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.