Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 1, No. 2, 2026

An Improved Mean Teacher Framework for Robust Domain Adaptive Foggy Object Detection

Yanyan Zhu

Henan Polytechnic University, Henan, China

Abstract: To further address the limitations prevalent in existing domain adaptation methods for foggy object detection — namely, insufficient robustness against noise and interference, slow convergence during training, susceptibility to gradient oscillations and fluctuations, training instability, and inadequate extraction and utilization of small-object features — we propose a novel algorithm termed An Improved Mean Teacher Framework for Robust Domain Adaptive Foggy Object Detection (IMT-Det). First, the Mean Teacher paradigm is incorporated to alleviate the convergence difficulties inherent in adversarial training for domain-adaptive foggy object detection. The framework comprises a teacher model and a student model. During training, a dynamic smoothing coefficient is devised, through which the teacher model's parameters are updated via Exponential Moving Average (EMA) of the student model's parameters. This smooth training strategy mitigates convergence instability during adversarial optimization and, compared to a single-model baseline, renders the EMA-updated teacher model considerably less sensitive to noise and perturbations, thereby avoiding the oscillatory behavior that may arise from directly optimizing adversarial objectives. Subsequently, a weighted teacher loss mechanism is introduced, which dynamically adjusts loss weights to achieve fine-grained refinement of the training process and to resolve the problem of excessive regression errors under foggy scene conditions. Finally, a FEUCB upsampling module is designed to address the challenges posed by low contrast in foggy images and significant variations in object scale, thereby enhancing detection capability in foggy scenarios.

Keywords: Foggy weather detection, domain adaptive object detection, mean teacher.

1. Introduction

Object detection, as a core foundational task in the field of computer vision, plays an indispensable and critical role in practical application scenarios that rely on visual perception, such as autonomous driving, intelligent surveillance, and traffic management. However, when such visual perception systems are deployed in real-world environments, they are frequently subject to severe constraints imposed by adverse weather conditions such as fog, leading to substantial degradation in detection performance. In foggy scenes, microscopic suspended particulate matter in the atmosphere induces light scattering phenomena, causing a series of image quality problems in camera-captured images, including reduced contrast, blurred target contours, degraded fine-grained features, and attenuated visibility [1, 2]. These fog-induced image degradation issues render object detection models trained on clear-weather images ill-suited to foggy environments, preventing them from accurately capturing target information and consequently leading to decreased detection accuracy and elevated rates of missed and false detections. This not only compromises the normal operation of such systems but may also give rise to serious safety hazards such as traffic accidents.

To address the challenge of object detection in foggy scenes and narrow the domain gap between normal weather conditions (source domain) and foggy scenes (target domain), unsupervised domain adaptation has become the mainstream research direction in this field [3-6]. Such methods do not rely on large-scale annotated foggy data; instead, they transfer knowledge from a labeled source domain to an unlabeled target domain, enabling the model to self-adapt to foggy scenes and effectively reducing the cost and difficulty of data annotation. These methods have demonstrated promising application prospects in foggy object detection tasks. Currently, most unsupervised domain adaptive foggy object detection methods primarily rely on adversarial learning mechanisms [4, 7], alleviating the performance degradation caused by cross-domain discrepancies by aligning the feature distributions of the source and target domains at the global or instance level. Some of these methods have achieved notable detection results on multiple public benchmarks.

Nevertheless, through in-depth analysis and experimental validation of existing adversarial learning-based domain adaptive foggy object detection methods, we identify three critical unresolved issues that severely constrain the practical effectiveness and performance improvement of these models. First, the adversarial training mechanisms employed by existing methods universally suffer from convergence instability, with pronounced oscillatory phenomena readily occurring during model training. This not only reduces training efficiency but also impairs the final model performance, making it difficult to achieve stable domain adaptation. Second, in complex foggy scenes — particularly under extreme conditions such as dense fog and abrupt variations in fog concentration — existing models exhibit insufficient target regression accuracy, with excessively large bounding box regression errors that degrade target localization precision and fail to meet the accuracy requirements of practical applications. Third, existing models inadequately extract features of small-scale objects in foggy scenes. Affected by fog occlusion and feature degradation, small-target features are easily overwhelmed by background noise, and feature dilution is prone to occur during the upsampling process, resulting in poor small-object detection performance and leaving considerable room for improvement in the detection capability of models under complex foggy conditions.

A deeper analysis of the aforementioned problems reveals that the root cause of convergence instability lies in the

absence of effective parameter smoothing mechanisms during adversarial training, which fails to suppress oscillations throughout the training process. Meanwhile, the excessive regression errors and insufficient small-object detection performance stem from the model's inadequate adaptability to complex foggy scenes and the lack of targeted feature enhancement and error optimization strategies. The Mean Teacher framework, as an effective semi-supervised learning paradigm, possesses significant advantages in parameter smoothing and noise robustness. It can effectively mitigate the convergence instability of adversarial training and enhance model training stability and generalization capability [16], thereby providing a reliable technical approach to resolving the aforementioned issues.

Motivated by the above analysis, to effectively address the problems of convergence instability, excessive regression errors, and insufficient small-object detection performance in existing domain adaptive foggy object detection methods, and to further enhance the domain adaptation effectiveness and detection accuracy of models in complex foggy scenes, this paper proposes a Robust Domain Adaptive Foggy Object Detection algorithm based on an Improved Mean Teacher framework (IMT-Det). The proposed algorithm takes the Mean Teacher framework as its foundation and incorporates customized improvements tailored to the scene characteristics of foggy object detection. Through multi-module collaborative optimization, comprehensive performance enhancement is achieved. The specific solutions are as follows.

To address the convergence instability of adversarial training, this paper improves the conventional Exponential Moving Average (EMA) mechanism of the Mean Teacher framework by designing a dynamic smoothing coefficient that increases linearly with training epochs. By dynamically adjusting the update weights of the teacher model's parameters, smooth parameter iteration is realized, effectively suppressing oscillatory phenomena during training. Simultaneously, a refined weighted loss function is designed, with targeted adjustment of regression loss weights, to alleviate convergence oscillations while reducing target regression errors and improving target localization precision. To address the insufficient extraction of small-object features and feature dilution during upsampling in foggy scenes, this paper designs a FEUCB upsampling module. Through adaptive feature reorganization and enhancement strategies, the module strengthens the extraction and retention of small-object features, mitigates feature dilution during upsampling, and improves small-object detection performance. To address the insufficient adaptability of models to complex foggy scenes, this paper integrates the improved Mean Teacher framework with adversarial domain adaptation ideas to construct a progressive robust domain adaptation pipeline. This pipeline guides the model to progressively adapt to the cross-domain transition from normal weather to complex foggy conditions, enhancing the model's adaptability to challenging scenarios such as dense fog and abrupt fog concentration variations. This progressive adaptation strategy draws on the core concept of progressive domain adaptation, progressively narrowing the domain gap to improve model adaptability.

To validate the effectiveness and superiority of the proposed IMT-Det algorithm, extensive comparative experiments are conducted on multiple publicly available domain adaptive foggy object detection datasets and general domain adaptation datasets. Experimental results demonstrate that the proposed algorithm effectively improves the convergence stability of adversarial training, significantly reduces target regression errors, and enhances small-object detection performance and adaptability to complex foggy scenes. The algorithm outperforms existing mainstream domain adaptive foggy object detection methods across all key detection metrics, fully substantiating its effectiveness and practical utility.

The contributions of this paper are summarized as follows:

(1) An improved Mean Teacher framework is proposed. By designing a dynamic EMA smoothing coefficient that increases linearly with training epochs and an optimized weighted loss function, the proposed framework effectively resolves the problems of convergence instability and excessive regression errors in conventional adversarial training, improving model training efficiency and target localization precision.

(2) A FEUCB upsampling module is designed to specifically address the problems of insufficient small-object feature extraction and feature dilution during upsampling in foggy scenes. The module strengthens small-object feature representations and significantly improves the model's small-object detection performance in foggy environments.

(3) A progressive robust domain adaptation pipeline is constructed by integrating the improved Mean Teacher framework with adversarial domain adaptation ideas, enhancing the model's adaptability to complex foggy scenes. The superiority of the proposed algorithm is further validated on foggy datasets including FogCityscapes and RTTS, as well as general domain adaptation datasets.

2. Related Work

2.1. Image Dehazing Techniques

In recent years, image dehazing methods based on deep learning have achieved remarkable progress. Wu et al. [8] proposed a dehazing algorithm based on contrastive learning, which treats hazy images and haze-free images as negative and positive samples respectively, increasing the distance between the dehazing output and the negative samples while simultaneously pulling the output closer to the positive samples, thereby improving dehazing performance. Yang et al. [9] designed a self-augmented image dehazing framework that decomposes the transmission map into density information and depth information, while employing unpaired images for haze generation and removal, which enhances the generalization capability of the model. Liu et al. [10] introduced a self-supervised auxiliary network that dynamically adjusts the parameters of the dehazing network at test time, leveraging meta-learning to ensure consistent optimization objectives between the two networks, thereby significantly improving the model's adaptability to diverse types of hazy images.

2.2. Object Detection

Currently, object detection methods developed on the basis of neural networks can be broadly categorized into two major classes according to differences in their detection pipelines: single-stage detectors and two-stage detectors. Two-stage detectors [11-13] are characterized by a "filter-then-predict" two-step pipeline: candidate regions are first extracted from the input image through a specific mechanism, and subsequently, category label assignment and bounding box

coordinate regression are performed based on these candidate regions to achieve precise target detection. In terms of architectural composition, two-stage detectors generally comprise three core modules: a feature extractor responsible for mining deep feature representations from the input image; a Region Proposal Network (RPN) that generates candidate bounding boxes for potential target locations; and a classifier that performs category determination for each candidate region — the primary function of which is to determine whether a target exists at each candidate location and, if so, to further identify its specific category.

In contrast to the two-step pipeline of two-stage detectors, single-stage detectors — such as SSD [14], Retinanet [15], DETR [16] and the YOLO [17-20] series of mainstream models — adopt an end-to-end detection paradigm that requires no additional candidate region extraction step, directly and simultaneously generating bounding box predictions and category confidence scores from the feature maps produced by the convolutional neural network (CNN) backbone. Although single-stage detectors generally fall short of two-stage detectors in terms of detection accuracy — primarily because they omit the RPN module and employ a relatively simplified feature extraction process that insufficiently captures fine-grained target features — they possess a significant speed advantage, with inference efficiency substantially surpassing that of two-stage detectors. This characteristic renders single-stage detectors more compatible with the deployment requirements of practical application scenarios, making them particularly well-suited to tasks with stringent real-time performance demands.

2.3. Object Detection in Foggy Conditions

In recent years, numerous methods have been proposed to address the various challenges of object detection in foggy scenes. Dai et al. [21] investigated the utility of super-resolution for other visual applications such as edge detection and segmentation, demonstrating that perceptual metrics and object detection performance are interrelated yet cannot fully substitute for one another. Li et al. [22] further integrated downstream tasks with image restoration evaluation, for the first time explicitly employing object detection as a metric to assess the dehazing quality of hazy images. This approach broke through the limitations of traditional evaluation paradigms that rely solely on perceptual metrics, providing new perspectives for subsequent related research. Zhang et al. [23] proposed the MSFFA-YOLO network, which employs an enhanced YOLOv7 as the detection sub-network responsible for learning target localization and classification, and a Multi-Scale Feature Fusion Attention (MSFFA) structure as the restoration sub-network dedicated to visibility enhancement. Qiu et al. [24] proposed the IDOD-YOLOv7 framework targeting low-illumination foggy traffic scenarios. This network is built upon an IDOD module comprising an AOD dehazing component and a SAIP image enhancement component, which are jointly optimized in an end-to-end manner with the YOLOv7 detection module. Furthermore, the authors constructed a dedicated low-illumination foggy traffic image dataset (FTOD) generated through physics-based fog synthesis, effectively mitigating the domain shift problem between real and synthetic domains.

2.4. Domain Adaptive Object Detection

With the advancement of deep learning, domain adaptation has been introduced into foggy object detection, framing the task as a domain adaptation problem that involves transitioning from a source domain (clean images) to a target domain (hazy images), given that images captured under foggy conditions inherently exhibit pronounced domain shift. The predominant approach employs adversarial training to align target domain features with source domain features. The earliest work in this line applied two Gradient Reversal Layers (GRL) on Faster R-CNN to perform image-level and instance-level alignment. Subsequently, several methods began adopting two-stage detectors with a focus on image-level alignment. Saito et al. [25] demonstrated that strong local alignment combined with weak global alignment of features extracted from the backbone improves domain adaptation performance. Zhu et al. [26] proposed a domain adaptive detection method that places greater emphasis on instance-level alignment, leveraging RPN candidate regions to perform alignment at the region level. To adapt source-biased decision boundaries to target domain data, Yu et al. [27] introduced an uncertainty-based pseudo-detection set fusion strategy, where the uncertainty estimates are generated through stochastic inference.

Compared to the extensive research conducted on two-stage detectors, unsupervised domain adaptation (UDA) work targeting single-stage detectors is relatively limited; nevertheless, several important methods have emerged in this area. Building upon the YOLO series of detectors, researchers have proposed a variety of innovative approaches. Liu et al. [28] proposed IA-YOLO, a framework capable of performing adaptive image-level enhancement for each individual image to improve detection performance. Specifically, a lightweight CNN-based parameter predictor (CNN-PP) learns from downsampled input images to predict the hyperparameters of filters within a Differentiable Image Processing (DIP) module, which subsequently processes high-resolution input images to assist YOLOv3 in achieving superior detection performance; notably, the dehazing filter is activated only in the presence of fog. Wang et al. [4] proposed the Robust Foggy Detection YOLO model (R-YOLO), which narrows cross-domain discrepancies through synthetic pseudo-image generation and introduces a feature calibration module to accomplish cross-domain feature alignment. Hnewa and Radha [7] proposed a novel Multi-Scale Domain Adaptive YOLO (IMS-DAYOLO) framework that employs multiple domain adaptation pathways and corresponding domain classifiers at different scales within the YOLOv4 object detector. Building upon the MSDA-YOLO framework, three new deep learning architectures were introduced for the Domain Adaptation Network (DAN) to generate domain-invariant features, comprising a Progressive Feature Reduction (PFR) module, a Unified Classifier (UC), and an ensemble architecture. Mattolin et al. [29] presented, at WACV, an adaptive object detector learning approach that incorporates a region-level detection confidence-based sample mixing strategy. Mekhalfi et al. [30] proposed the DACA algorithm, introducing a four-stage unsupervised domain adaptation strategy: self-supervised learning is first employed to leverage target domain pseudo-labels for domain adaptation; high-confidence detection regions in target domain images are then identified and pseudo-labels are generated; these regions are subsequently cropped and subjected to diverse augmentation transformations; the augmented versions are then combined into composite images; and finally, the composite images are utilized to complete the network's adaptation to the target domain, in conjunction with source domain data for joint

training. In the same year, Zhang et al. [31] proposed SSDA-YOLO targeting industrial application requirements. This method integrates the single-stage detector YOLOv5 with the Mean Teacher framework to extract target domain instance features through knowledge distillation. Simultaneously, scene style transfer is employed to compensate for image-level domain discrepancies, and a consistency loss is introduced to align cross-domain predictions. Experiments demonstrate that this algorithm achieves strong performance on multiple benchmarks including Foggy Cityscapes, validating the effectiveness of integrating adaptive modules into single-stage detectors. Varailhon et al. [32] proposed SF-YOLO, which introduces a communication mechanism between the teacher and student models to effectively stabilize the training process and reduce the dependence of model selection on annotated target domain data.

3. Method

3.1. Overview

To address the identified issues of convergence oscillation, excessive regression errors, and insufficient extraction and utilization of small-object features, and to further improve the training stability and detection accuracy of object detectors under foggy weather conditions, we propose a Robust Domain Adaptive Foggy Object Detection algorithm based on an Improved Mean Teacher framework (IMT-Det). We addresses the problems of adversarial training convergence instability and insufficient regression accuracy in foggy scenes primarily from the following three perspectives. First, the Mean Teacher framework is adopted to help mitigate the convergence difficulties encountered during the training of domain adaptive foggy object detectors. Second, the Exponential Moving Average (EMA) strategy is employed to perform smooth parameter updates on the teacher model, enhancing the model's robustness against noise and perturbations and avoiding the training oscillations that may arise from directly optimizing adversarial objectives. Third, a weighted teacher loss function and a FEUCB upsampling structure are designed to achieve fine-grained optimization of the training process, effectively resolving the problems of excessive target regression errors and insufficient small-object detection accuracy in foggy scenes. Figure 1 and Figure 2 illustrate the overall framework architecture of the proposed algorithm.

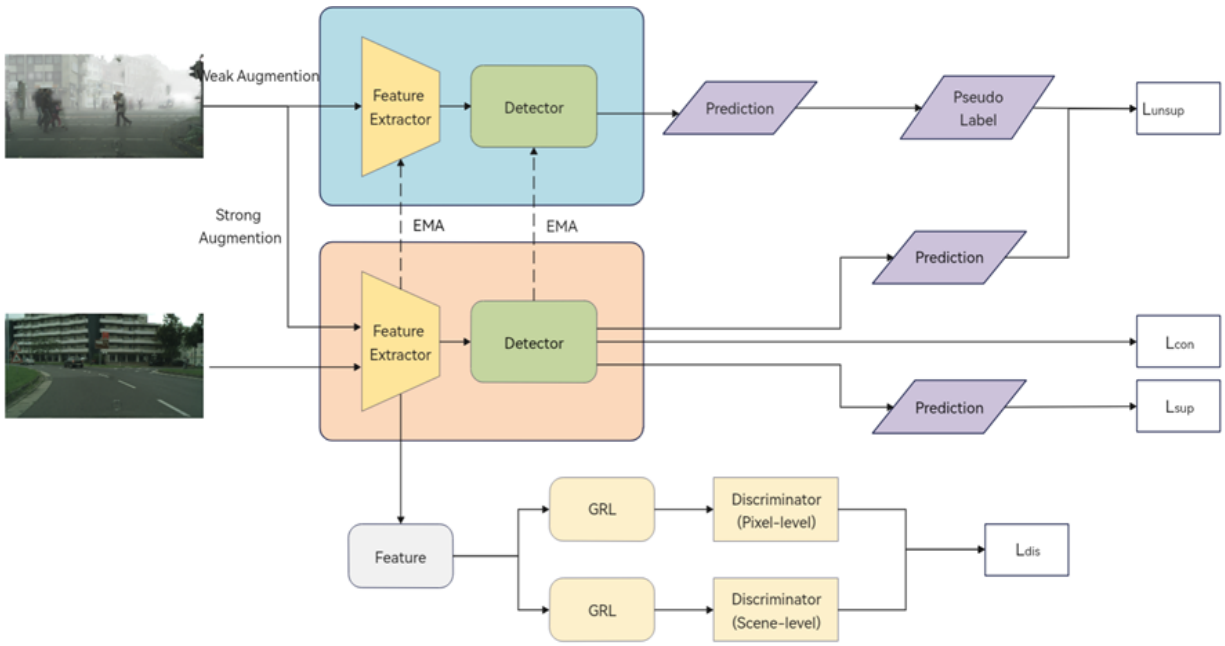

Figure 1. Overall Network Architecture Diagram

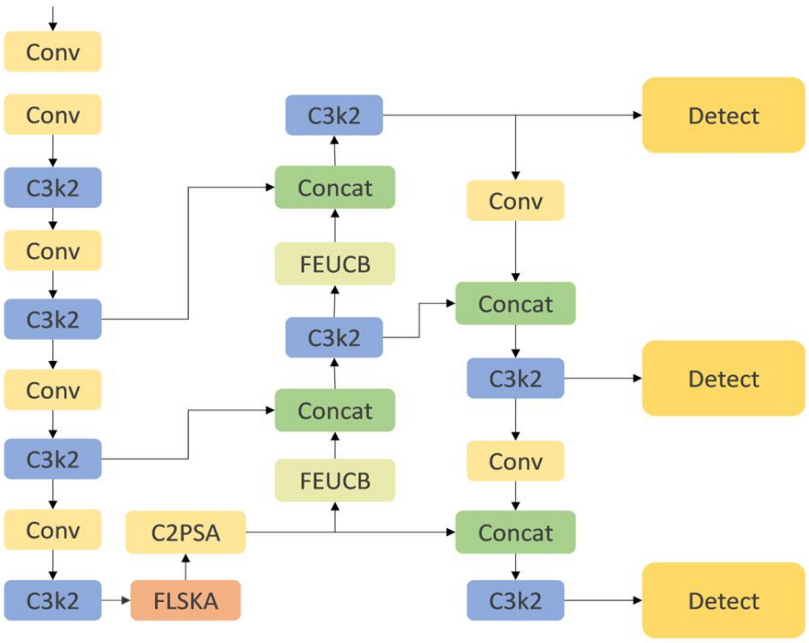

Figure 2. Model Architecture Diagram

3.2. Mean Teacher Architecture

We introduce and improve upon the Mean Teacher (MT) framework. This framework constructs a "teacher-student" dual-network mechanism, leveraging an Exponential Moving Average (EMA)-based parameter smooth update mechanism to guide the model's cross-domain consistency learning. The core innovations are reflected in the following two aspects:

3.2.1. EMA-Based Parameter Smooth Update Mechanism

In conventional adversarial learning models, during the backpropagation process, model parameters are subject to the influence of the minimax game, often exhibiting violent fluctuations in the vicinity of local optima. The Mean Teacher structure introduced in this paper employs an Exponential Moving Average strategy to perform temporally smooth fusion of the student model's historical weights accumulated during training, thereby updating the parameters of the teacher model:

Where,

$$\theta_{T,t}=\alpha\cdot\theta_{T,t-1}+(1-\alpha)\cdot\theta_{S,t} \quad (1)$$

Where $\theta_{T,t}$ denotes the teacher model parameters at training step t, $\theta_{S,t}$ denotes the student model parameters at training step t, and αdenotes the dynamic EMA smoothing coefficient.

Unlike the conventional EMA with a fixed smoothing coefficient, we design a dynamic smoothing coefficient that increases linearly with training epochs:

$$\alpha=\alpha_{start}+(\alpha_{end}-\alpha_{start})\cdot\frac{\text{epoch}}{\text{total epoch}} \quad (2)$$

Specifically, the initial and final values of the dynamic smoothing coefficient are set to $\alpha_{start}=0.9999$ and $\alpha_{end}=0.99$, respectively. During the early stages of training, the feature discrepancy between the foggy target domain and the source domain is substantial, and the demand for domain alignment through adversarial training is urgent. Setting $\alpha_{start}=0.9999$ enables the teacher model to rapidly track the domain adaptation updates of the student model, accelerating the learning of the low-contrast and blurred feature distributions characteristic of foggy images. During the later stages of training, the student model has already accomplished fundamental domain alignment, and the instantaneous noise introduced by dense fog samples becomes the primary source of interference. Gradually decreasing $\alpha_{end}=0.99$ strengthens the parameter smoothing effect, filtering out gradient oscillations induced by abrupt variations in fog concentration during training and ensuring the stability of the teacher model parameters.

Since the teacher model does not directly participate in gradient backpropagation, this soft update mechanism effectively reduces unstable gradients, providing the student model with a stable and reliable reference baseline for feature alignment in the foggy target domain. This fundamentally resolves the training oscillation problem and significantly improves the convergence speed and training stability of the overall framework.

3.2.2. Teacher-Guided Self-Supervised Consistency Constraint

In the foggy target domain where ground-truth labels are unavailable, the teacher model leverages its smoothed parameters to perform inference on target domain images, generating pseudo-labels with higher confidence. By imposing consistency constraints between the outputs of the student model under different perturbations (e.g., data augmentation and noise injection) and the predictions of the teacher model, a self-supervised learning pathway is constructed. This mechanism enforces semantic continuity in the feature space; particularly in dense fog scenes characterized by low contrast and blurred target features, the teacher model is capable of providing more accurate semantic priors by virtue of its parameter stability. This consistency-guided supervision effectively compensates for the deficiencies of the single adversarial alignment mechanism in fine-grained localization and regression accuracy, significantly enhancing the model's generalization capability under complex foggy environments.

Through the design of the aforementioned Mean Teacher structure, successfully constructs an optimized pipeline that evolves from adversarial domain adaptation toward more robust domain adaptation. While preserving the advantages of

local-to-global feature alignment, the proposed framework resolves the training instability of the model under high fog intensity conditions through parameter smoothing and semantic consistency constraints.

3.3. Weighted Loss: Dynamic Bbox Weighting (DBW)

In the domain adaptive object detection task under foggy conditions, the combined effects of atmospheric scattering and light attenuation phenomena lead to significantly reduced image contrast, blurred target boundaries, and severe degradation of textural details, which critically constrains the collaborative training of the teacher-student dual-model architecture within the Mean Teacher framework. The fixed-weight loss function employed in the conventional Mean Teacher framework adopts a fixed-weight design for bounding box (bbox) loss, which fails to adequately account for the characteristics of blurred target boundaries and dynamically varying prediction deviations under foggy scene conditions. This deficiency results in insufficient bounding box regression accuracy, consequently causing the detection network to frequently suffer from missed detections, false detections, and excessively large bounding box regression errors.

To address the aforementioned problems, we introduce targeted improvements to the bounding box loss on the basis of the original loss formulation of the Mean Teacher framework. Specifically, we design a weighted bounding box loss adapted to the requirements of foggy domain adaptation and propose an improved weighted loss mechanism for the Mean Teacher framework.

By establishing multi-dimensional prediction consistency constraints and incorporating a confidence-based dynamic bounding box weighting strategy, deep alignment between the student network and the teacher network is achieved under complex hazy scene conditions. The weighted consistency loss is designed to guide the learning of the student network through the prediction outputs of the teacher network. The loss function comprises three components: bounding box regression consistency, objectness score consistency, and class probability consistency. During the forward pass, the model first filters the teacher network's predictions using a confidence threshold τ, retaining only high-reliability predictions as pseudo-labels to mitigate erroneous supervision caused by haze-induced noise. The objectness score and class probability are constrained via binary cross-entropy loss and mean squared error loss, respectively.

The core innovation of the proposed algorithm lies in the dynamic weight improvement applied to the bounding box regression consistency loss L_{bbox}. Unlike conventional fixed weight assignment, this work designs an adaptive weight function $f_{adaptive}$ that jointly considers the current training epoch E, the iteration step I, and the confidence score $\hat{c}$ output by the teacher network. During backpropagation, the bounding box consistency error is multiplied by this dynamic weight factor for gradient computation. This design enables the model to dynamically adjust the supervision intensity according to the training progress: during the early stages of training, a relatively high peak weight is applied to guide the student network in rapidly establishing perceptual sensitivity to the geometric characteristics of targets, while in the later stages of training, the weight is progressively reduced to prevent over-fitting to the prediction bias of the teacher network, thereby enhancing the model's localization robustness in cross-domain foggy scenes.

The consistency computation formula for the Mean Teacher weighted loss is defined as:

$$L_{MT}=\frac{1}{N}\sum_{i=1}^{N}\left(w_{bbox}\cdot MSE(B_s,B_t)+L_{obj}+L_{cls}\right) \quad (3)$$

Where B_s and B_t denote the bounding box coordinates predicted by the student and teacher networks, respectively, and N represents the number of feature scales. The computation process of the dynamic bounding box weight w_{bbox} can be expressed as:

$$w_{bbox}=mean\left(f_{adaptive}(E,I,C_t)\right) \quad (4)$$

Through the aforementioned dynamic weighting mechanism, the model achieves a smooth transition from "teacher-dominated supervision" to "adaptive generalization" during the training process, accomplishing fine-grained optimization of the training procedure. This weight evolution strategy effectively mitigates the localization drift phenomenon caused by blurred target features in foggy environments. By reinforcing the supervisory role of high-quality spatial predictions, bounding box regression errors are suppressed, and the localization accuracy and stability of the domain adaptive object detection model under extreme foggy scene conditions are significantly improved, fully validating the effectiveness of the weighted teacher loss design, dynamic weight adjustment, and fine-grained training innovation.

Finally, the total loss is defined as the weighted sum of L_{bbox}, L_{obj}, L_{cls}, and L_{adv}as:

$$w_{bbox}=mean\left(f_{adaptive}(E,I,C_t)\right) \quad (5)$$

Where ω is a hyperparameter used to balance the Mean Teacher loss and other loss components. The supervised detection loss L_{det} and the self-supervised consistency loss L_{con} share the same loss function $L(\cdot)$, and L_{adv} denotes the feature calibration loss.

3.4. Foggy-scene Efficient Up-Convolution Block (FEUCB)

To address the problem that, under foggy conditions, the combined effects of atmospheric scattering and light attenuation cause severe image quality degradation — leading to blurred target boundaries, loss of textural details, feature dilution during upsampling, and further weakening of small-object features — which critically constrains the detection network's ability to recognize small and ambiguous targets, frequently resulting in missed detections of small objects and excessively large target regression deviations, this paper designs the Foggy-scene Efficient Up-Convolution Block (FEUCB), an efficient upsampling module. By integrating lightweight upsampling operations with Depthwise Separable Convolution (DWC), channel shuffle, and Pointwise Convolution (PWC), and tailoring their application to the specific requirements of foggy object detection, the module is designed to accommodate the feature extraction demands of foggy scene conditions.

The upsampling operation restores feature map dimensions through spatial magnification, providing appropriately sized feature inputs for subsequent feature fusion and object detection. Depthwise Separable Convolution (DWC) leverages the advantages of grouped convolution to substantially reduce computational cost while ensuring module efficiency, precisely capturing local detail features of

upsampled feature maps and effectively compensating for detail loss caused by fog occlusion on target objects. The channel shuffle mechanism breaks the redundancy barriers between feature channels and enhances inter-channel information interaction, enabling the module to more thoroughly exploit the multi-dimensional features of foggy targets. Pointwise Convolution (PWC) is responsible for integrating the shuffled channel features, further strengthening feature representation capability and providing high-quality feature support for subsequent target regression and classification.

Through an in-depth analysis of the core pain points of upsampling in foggy scenes, it is observed that conventional upsampling modules either prioritize efficiency at the expense of feature preservation, or emphasize feature integrity at the cost of a dramatic increase in computational burden, making them ill-suited to the practical application requirements of foggy scenarios. The EUCB module itself achieves a rational decoupling of upsampling, feature extraction, channel optimization, and feature integration functions, simultaneously guaranteeing upsampling efficiency and realizing progressive reinforcement of feature information, thereby effectively mitigating the feature dilution problem caused by upsampling in foggy scenes. The constituent components of the EUCB module form a functionally complementary and efficient upsampling architecture: the upsampling operation is responsible for feature scale restoration, ensuring that feature map dimensions satisfy the requirements of subsequent detection tasks; depthwise separable convolution is responsible for local detail feature extraction, compensating for the detail deficiencies of foggy targets; channel shuffle is responsible for enhancing inter-channel information interaction, enriching the diversity of feature representations; and pointwise convolution is responsible for feature integration, strengthening the discriminability between target features and background noise. In this paper, the module is applied to the upsampling stage of the foggy object detection network, and through the synergistic operation of its constituent components, the network's feature extraction capability and recognition accuracy for small and ambiguous targets in foggy scenes are significantly improved. The architectural structure of the FEUCB upsampling module is illustrated in Figure 3 below.

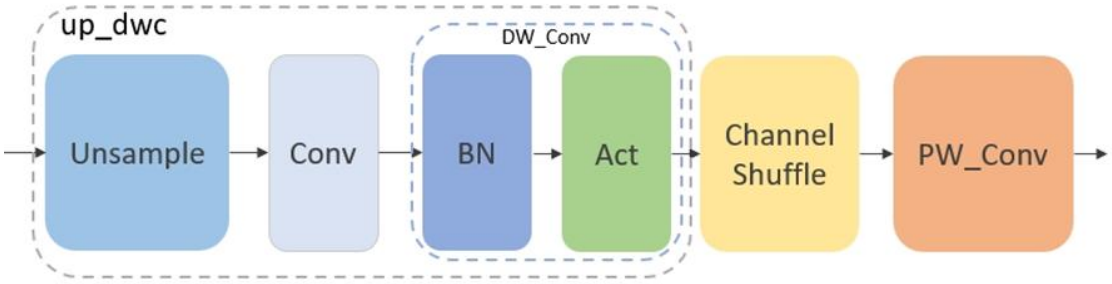

Figure 3. Architecture of the FEUCB Module

4. Experimental Results and Analysis

In this section, comprehensive experiments are conducted to evaluate the detection performance of the proposed method and other competing approaches under foggy conditions. All experiments are implemented using PyTorch with default settings. The single-stage detector YOLOv11s is adopted as the base detector, with the batch size set to 4, where each batch contains one normal-weather image and one foggy image of size 600 × 600. The model is first pre-trained on normal-weather images using pre-trained COCO weights, followed by adaptive learning. During the Non-Maximum Suppression (NMS) stage, the IoU threshold is set to 0.5, and the confidence threshold C_{th} is set to 0.35 for pseudo-detection generation. The teacher confidence threshold is set to 0.6 for generating teacher predictions.

4.1. Comparative Experiments

The "Source" (source domain) results (trained on normal-weather images only) represent the performance lower bound, where the model is trained exclusively on normal-weather images and directly applied to foggy datasets without any adaptation. The "Oracle" results are also reported, representing the performance upper bound, where the model is trained on foggy images with ground-truth annotations and evaluated on the foggy test set. The adaptation results from normal weather to foggy weather are presented in Tables 1, 2, and 3.

As shown in the table, the proposed method outperforms SF-YOLO (42.5%) by 4.0 percentage points. This overall performance improvement is attributable not only to the design of the Mean Teacher framework and the FEUCB upsampling module, but also to the dynamic weight design, which adaptively adjusts the weight of the bounding box regression loss, effectively alleviating the problem of excessive regression errors and reducing the sensitivity of target localization to regression loss in complex foggy scenes. As a result, more balanced and stable performance improvements are achieved across all object categories, fully validating the effectiveness of the synergistic operation of the aforementioned modules.

For large-scale object categories (truck, bus, and train), the proposed method achieves the best or second-best results among all competing methods, with mAP scores of 38.1%, 51.4%, and 51.6% for "truck," "bus," and "train," respectively. The "truck" category improves by 1.5 percentage points and the "train" category improves by 0.3 percentage points, demonstrating that the smooth update mechanism of the Mean Teacher further alleviates training oscillations and enhances robustness against occlusion and contrast degradation for large-scale targets under complex foggy conditions. For small-scale object categories (pedestrian, motorcycle, and bicycle), the proposed method achieves strong performance with mAP scores of 47.8%, 32.0%, and 37.2% for "pedestrian," "motorcycle," and "bicycle," respectively. This significant improvement is attributable on one hand to the thorough extraction and utilization of multi-scale features for small objects by the FEUCB upsampling module, and on the other hand to the adaptive adjustment of the L_{bbox} loss weight by the dynamic weight design — under foggy low-contrast conditions with varying small-object scales, the

dynamic balancing of regression loss optimization intensity effectively suppresses the excessive accumulation of regression errors, further improving small-object localization precision and detection stability. Compared with Oracle methods, the proposed method (mAP: 46.5%) already surpasses both the YOLOv5m Oracle (44.2%) and the YOLOv8s Oracle (44.9%), and is nearly on par with the YOLOv11s Oracle (46.7%), falling short by only 0.2 percentage points. This result demonstrates that the proposed method, without requiring any annotated foggy data, approaches the performance upper bound of fully supervised training, fully validating the effectiveness of domain adaptive foggy object detection based on the Mean Teacher framework and structural optimization.

Table 1. Quantitative Results (mAP): Cityscapes → Foggy Cityscapes

Method	person	rider	car	truck	bus	train	motorcycle	biycle	map
YOLOv11s (source)	39.7	43.1	47.2	16	30.1	5.9	15.7	35	29.1
YOLOv8s (source)	36.7	40.7	47.2	15.7	28.9	8	18.2	30	28.2
IMS-DAYOLO [7]	38.1	44.5	54.9	26.3	47.9	38.5	29.6	32.1	39.0
SSDA-YOLO [31]	47.2	54.3	61.6	16.2	43.1	25.7	31.0	34.5	39.2
DACA [30]	41.9	40.8	63.0	29.4	42.2	37.2	27.8	33.0	39.4
ConfMix [29]	45.0	43.4	62.6	27.3	45.8	40.0	28.6	33.5	40.8
SF-YOLO [32]	47.4	50.6	64.1	26.0	49.7	32.5	25.9	44.1	42.5
R-YOLO [4]	44.5	44.7	64.8	32.7	51.2	53	31.6	36.7	44.9
Ours	47.8	47.6	66.3	38.1	51.4	51.6	32.0	37.2	46.5
YOLOv5m (Oracle)	46	42.2	67.2	35.2	49.3	47.4	31.3	35.2	44.2
YOLOv8s (Oracle)	46.8	48.3	66.2	32.9	52.4	46	27.9	38.5	44.9
YOLOv11s (Oracle)	50.1	49.8	68	38.4	56.7	52.4	31.4	40.5	48.4

For qualitative comparison, we conduct a visual comparison between the proposed method and the three best-performing methods on the synthetic dataset Foggy Cityscapes, namely YOLOv11s (Oracle), R-YOLO, and SF-YOLO. Figure 4 presents the visualization of detection results of the proposed method and other competing approaches. As can be observed, under complex foggy conditions including dense fog occlusion and abrupt variations in fog concentration, the proposed method achieves more accurate object detection compared to all other competing methods, with significantly reduced missed detection rates and further improved target localization precision. For small-scale objects severely affected by low contrast in foggy scenes — such as distant pedestrians and motorcycles — the proposed method demonstrates superior capability in target contour recognition and bounding box regression, with predicted bounding boxes conforming more closely to the actual target contours, effectively validating the improvement of the FEUCB upsampling module in multi-scale feature extraction capability for small objects. The proposed method exhibits notably reduced bounding box regression errors under equivalent foggy conditions, demonstrating the effectiveness of the dynamic weight design in improving target localization precision through adaptive adjustment of the L_{bbox} loss weight. Furthermore, under dense fog and abrupt fog concentration variation scenarios, the Source model suffers from extensive missed detections and false detections, whereas the proposed method maintains relatively stable detection results, reflecting the significant advantages of the Mean Teacher framework in enhancing model training stability and domain adaptation robustness. Compared to methods such as SF-YOLO and R-YOLO, the proposed method exhibits stronger detection completeness in complex scenes where dense targets and small objects coexist, further substantiating the effectiveness and generalization capability of the proposed framework in complex foggy object detection tasks.

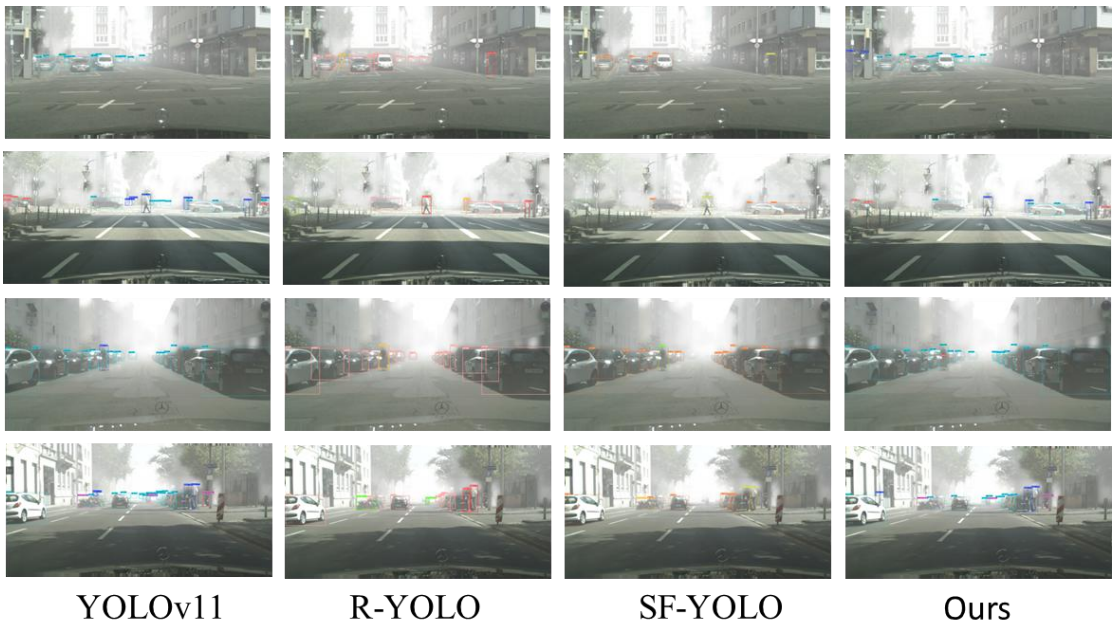

Figure 4. Visualization of Detection Results in Synthetic Foggy Scenes

Table 2 further demonstrates that the proposed method achieves an mAP of 74.8% on synthetic foggy images (VOC-Fog), matching and marginally surpassing R-YOLO (72.0%) while outperforming all other domain adaptive competing methods (DACA: 63.9%, ConfMix: 67.9%, SF-YOLO: 48.6%), attaining the best overall result among all compared approaches. The proposed method exhibits balanced and stable performance across all categories, achieving or exceeding all other competing methods on "pedestrian" (81.5%), "motorcycle" (73.8%), "bus" (86.3%), and "bicycle" (66.0%), demonstrating that the Mean Teacher framework and the FEUCB upsampling module effectively strengthen the model's adaptive capability from the normal-weather domain to the foggy domain. The performance gains are particularly pronounced on small-scale object categories, validating the effectiveness of the dynamic weight design in improving target localization precision through adaptive adjustment of the L_{bbox}loss weight. In comparison with the Oracle upper bound, the YOLOv11s Oracle (76.3% mAP) represents the current theoretical performance ceiling, and the proposed method falls short by only 1.5 percentage points. Furthermore, the proposed method (74.8%) already surpasses both the YOLOv5m Oracle (71.2%) and the YOLOv8s Oracle (71.5%), achieving detection performance approaching that of fully supervised training without requiring any annotated foggy data, fully substantiating the effectiveness and practical value of the proposed method in synthetic foggy scene conditions.

Table 2. Quantitative Results (mAP): VOC → VOC-Fog

Method	person	car	bus	motorcycle	biycle	map
YOLOv11s (source)	80.5	62.6	83.1	64.5	53.6	68.8
YOLOv8s (source)	64.8	53.6	70.7	35.5	30.4	51
IMS-DAYOLO [7]	75.4	56.8	71.3	59.0	42.8	61.1
SSDA-YOLO [31]	72.6	54.8	69.2	51.7	46.1	58.9
DACA [30]	76.7	59.7	73.9	60.3	48.7	63.9
ConfMix [29]	76.6	69.1	79.6	57.8	56.5	67.9
SF-YOLO [32]	60.7	52.3	65.9	30.6	33.4	48.6
R-YOLO [4]	78.2	70.5	82.8	65.2	63.2	72
Ours	81.5	66.3	86.3	73.8	66	74.8
YOLOv5m (Oracle)	77.8	72.3	84.5	66.8	54.7	71.2
YOLOv8s (Oracle)	78.7	70.1	82.4	62.5	63.8	71.5
YOLOv11s (Oracle)	83.9	69.4	86.7	75.9	65.8	76.3

Table 3 demonstrates that, despite RTTS comprising real-world foggy images while the training data consists of synthetic foggy images (VOC-Fog), the proposed method still achieves an mAP of 54.9%, surpassing the majority of competing methods (second only to R-YOLO at 56.2%), further validating the generalization capability of the proposed algorithm under real-world foggy scene conditions. In terms of per-category performance, the proposed method outperforms all baseline methods on the "bicycle" (56.7%) and "pedestrian" (78.9%) categories, demonstrating the effectiveness of the FEUCB module in improving small-object feature extraction capability under real foggy conditions. Meanwhile, the proposed method achieves strong performance on the "car" category (68.7%), indicating that the dynamic weight design effectively reduces target regression errors in real foggy scenes through adaptive

adjustment of the L_{bbox} loss weight, thereby improving the localization precision of medium- and large-scale targets. The results presented in the table clearly demonstrate that the proposed method, by effectively combining the smooth update mechanism of the Mean Teacher with the domain adversarial adaptation strategy, exhibits significant advantages in cross-domain generalization from synthetic to real foggy scenes, validating the practical value of IMT-Det in real-world complex foggy environments.

Table 3. Quantitative Results (mAP) on RTTS

Method	person	car	bus	motorcycle	biycle	map
YOLOv8s (source)	78	60.2	33.7	30.9	49.1	50.4
YOLOv11s (source)	77.1	64.3	26.3	35.8	50.4	50.8
IMS-DAYOLO [7]	78.8	61.2	27.5	32.7	48.9	49.8
SSDA-YOLO [31]	76.2	60.7	30.1	33.4	47.4	49.6
DACA [30]	79.1	63.6	20.5	34.9	42.2	48.1
ConfMix [29]	77.6	69.2	31.6	35.6	46.7	52.1
SF-YOLO [32]	76.3	66,2	23.6	34.2	49.2	49.9
R-YOLO [4]	77.5	72.6	37.3	38.2	55.2	56.2
Ours	78.9	68.7	31.2	39.1	56.7	54.9
YOLOv5m (Oracle)	73.5	70.7	34.8	33.2	42.7	51
YOLOv8s (Oracle)	76.7	60.1	27.3	40.8	49.5	50.9
YOLOv11s (Oracle)	76.1	64.9	30.3	40.1	43.5	51

To more intuitively and clearly demonstrate the detection performance advantages of the proposed method and facilitate visual analysis and comparison, we conduct visualization comparison experiments between the proposed method and ConfMix, R-YOLO, and SF-YOLO on the real-world foggy dataset. Figure 5 presents detailed visualizations of the detection results of each method on this dataset. As can be observed from the visualization results, in the complex and challenging scenario of real-world fog, compared to all competing methods, the proposed model is capable of accurately capturing target contours and precisely identifying target categories even when encountering objects that are heavily occluded by fog and exhibit severely degraded image features, highlighting its more stable and superior detection capability. Furthermore, compared to other methods, the proposed model demonstrates more outstanding detection performance in small-object detection, further effectively validating the adaptability and superiority of the proposed method in complex foggy environments.

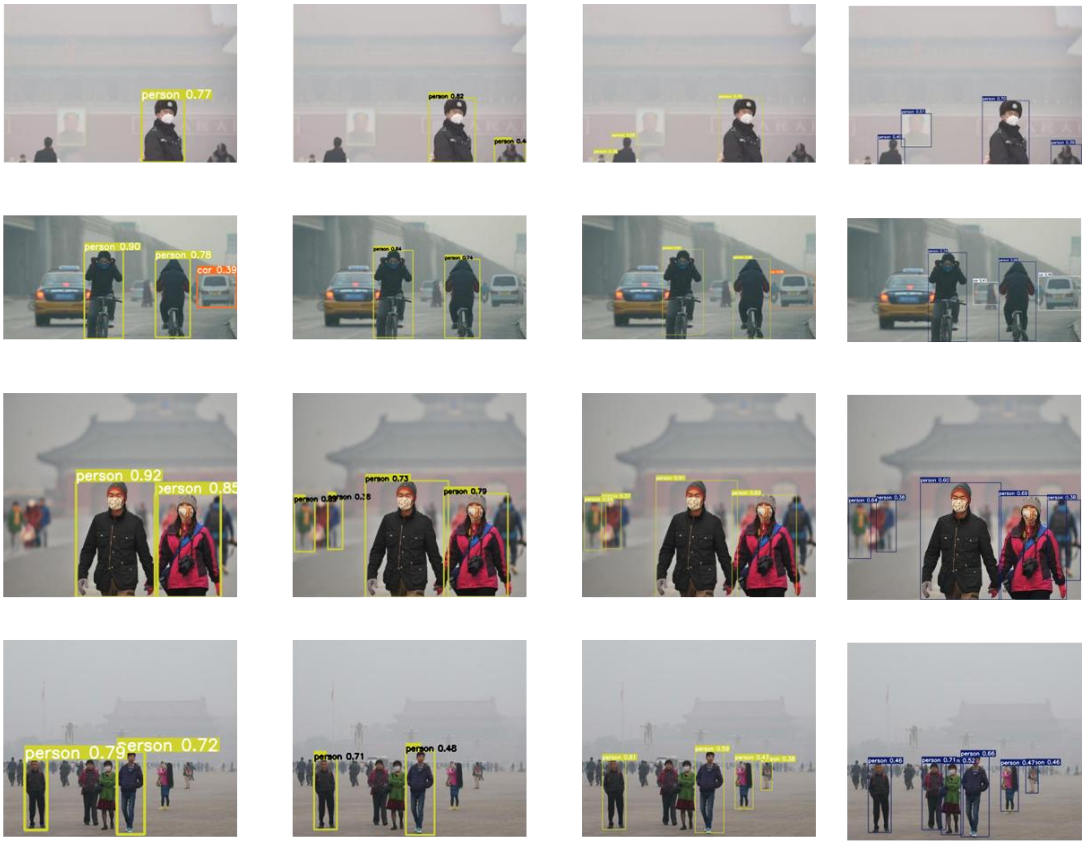

Figure 5. Visualization of Detection Results on Real-World Foggy Scenes

Table 4 presents the quantitative detection performance of the proposed method and existing baseline methods across three cross-domain adaptation scenarios: Cityscapes → RainCityscapes, Sim10K → Cityscapes, and KITTI → Cityscapes.

In the Cityscapes → RainCityscapes scenario, the proposed method achieves an mAP of 52.6%, significantly outperforming competing methods such as R-YOLO (44.8%) and DACA (42.5%), demonstrating that the smooth update mechanism of the Mean Teacher framework effectively suppresses training oscillations and improves domain alignment stability under weather degradation conditions. Furthermore, the proposed method substantially surpasses all Oracle baselines (39.1%–45.5%), achieving detection performance exceeding that of fully supervised training without requiring any target domain annotations.

In the Sim10K → Cityscapes and KITTI → Cityscapes scenarios, the proposed method achieves mAP scores of 66.7% and 65.1%, respectively, outperforming all other competing methods. Notably, the performance improvement in the KITTI → Cityscapes scenario reaches 2.1 percentage points, validating the effectiveness of the dynamic weight design in suppressing target regression errors through adaptive adjustment of the loss weights. The experimental results across all three scenarios collectively demonstrate that the proposed method exhibits strong effectiveness and generalization capability across a variety of complex cross-domain detection tasks.

Table 4. Quantitative Results (mAP) Across Three Cross-Domain Adaptation Scenarios

Dataset	Cityscapes →Rain Cityscapes	Sim10K→Cityscapes	KITTI→Cityscapes
YOLOv8s (Source)	38	52	53.8
YOLOv11s (Source)	39.2	54.9	49.6
IMS-DAYOLO [7]	44.0	53.2	48.5
SSDA-YOLO [31]	40.2	56.3	54.1
DACA [30]	42.5	58.4	54.5
ConfMix [29]	41.3	55.3	53.1
SF-YOLO [32]	41.9	63.6	54.5
R-YOLO [4]	44.8	51.4	47.3
Ours	52.6	66.7	65.1
YOLOv5m (Oracle)	39.1	75.2	75.2
YOLOv8s (Oracle)	40.0	71.9	71.9
YOLOv11s (Oracle)	45.5	75.3	75.3

4.2. Ablation Study

We first conduct an ablation study on the hyperparameter ω to determine its optimal value, with all results recorded in Table 5. The parameter ω is used to balance the Mean Teacher loss against other loss components, and its value directly influences the constraint strength of the Mean Teacher framework on model training. As can be observed from the experimental results, as ω increases, the model performance follows a trend of first rising and then declining. When ωis relatively small (0.1–0.2), the constraining effect of the Mean Teacher loss is weak, and the model mAP remains at a relatively low level of 45.3%–45.6%, indicating that insufficient ω weighting fails to fully leverage the smooth update and noise suppression advantages of the Mean Teacher framework. As ω gradually increases, model performance continues to improve, reaching the optimal mAP of 46.5% at ω=0.5. When ω continues to increase (0.6–1.0), model performance exhibits a notable decline, with mAP dropping to 45.3%–46.2%, suggesting that an excessively large ω weight causes the Mean Teacher loss to dominate the total loss, suppressing the effective optimization of other loss components, weakening the domain adaptation capability, and ultimately degrading detection performance. Based on the above analysis, ω=0.5 is identified as the optimal hyperparameter configuration, at which point the best balance between the Mean Teacher loss and other loss terms is achieved, and the model attains optimal performance on the foggy object detection task. All subsequent experiments in this paper adopt this configuration.

Table 5. Experimental Results Under Different ω Values

ω	mAP	ω	mAP
0.1	45.3	0.6	46.2
0.2	45.6	0.7	45.5
0.3	45.3	0.8	45.8
0.4	45.7	0.9	45.3
0.5	46.5	1	45.1

Subsequently, the detection performance of each individual module under foggy scene conditions is investigated, with results presented in Table 6. In the first row, Baseline denotes the base model; MT denotes the Mean Teacher framework; DBW (Dynamic Bbox Weighting) denotes the dynamic bounding box weighting loss; and FEUCB denotes the proposed upsampling module.

First, compared to the Baseline (45.3%), the introduction of MT alone improves mAP to 45.9%, demonstrating that the Mean Teacher framework effectively alleviates convergence oscillation during adversarial training through the exponential moving average update mechanism of the teacher model, enhancing model training stability and domain adaptation robustness. The introduction of FEUCB alone improves mAP to 45.7%, indicating that the FEUCB upsampling module effectively improves the feature representation of small objects in low-contrast foggy scenes by enhancing multi-scale feature extraction capability, thereby improving detection accuracy. When MT and DBW are used in combination, mAP further improves to 46.2%, representing a gain of 0.3 percentage points over MT alone (45.9%), demonstrating that DBW effectively complements the smooth update mechanism

of the Mean Teacher by adaptively adjusting the L_{bbox} loss weight, reducing target regression errors in complex foggy scenes. The combination of MT and FEUCB achieves an mAP of 45.9%, validating the synergistic effect of the Mean Teacher framework and the upsampling module in improving small-object detection capability and training stability. Finally, the complete method with all three modules — MT, DBW, and FEUCB — operating collaboratively achieves an overall mAP of 46.5%, representing an improvement of 1.2 percentage points over the Baseline. The experimental results demonstrate that the three modules exhibit strong complementarity in the foggy object detection task, and their collaborative operation collectively constructs a domain adaptive foggy object detection framework with more stable training, more precise localization, and stronger small-object detection capability.

Table 6. Module Ablation Experiment Results

Baseline	MT	DBW	FEUCB	mAP
✓				45.3
✓	✓			45.6
✓			✓	45.7
✓	✓	✓		46.2
✓	✓		✓	45.9
✓	✓	✓	✓	46.5

The visualization results in Figure 6 further intuitively validate the effectiveness of each module for detection performance. First, in high-density foggy scenes, the accuracy of target localization shows a progressive improvement with the gradual introduction of modules. Compared with the Baseline, after incorporating the MT module, the model exhibits significantly enhanced detection stability for targets occluded by fog, with a remarkable reduction in missed and false detections. This benefit comes from the mean teacher framework, which imposes a smooth constraint on the training process via the exponential moving average update mechanism, effectively suppressing oscillations in adversarial training and enabling the model to maintain more stable feature extraction capabilities in complex foggy environments. Second, the regression accuracy of bounding boxes is substantially improved. After introducing the DBW module, the alignment between predicted bounding boxes and the actual contours of targets is further enhanced, especially for complex-shaped and blurry-outline targets such as distant vehicles and partially occluded pedestrians. This improvement mainly arises from DBW adaptively adjusting the weight of the L_{bbox} loss, dynamically balancing the optimization strength of regression loss, effectively suppressing excessive accumulation of localization errors under heavy fog conditions, and significantly boosting target positioning precision. In addition, the visualization results demonstrate the advantages of the FEUCB upsampling module in small-object detection. With the addition of FEUCB, the detection integrity of small-scale objects such as motorcycles and bicycles is noticeably improved, and the bounding boxes fit the actual contours of targets more tightly. This indicates that FEUCB effectively compensates for the information loss of small-target features in low-contrast foggy environments by strengthening multi-scale feature extraction.

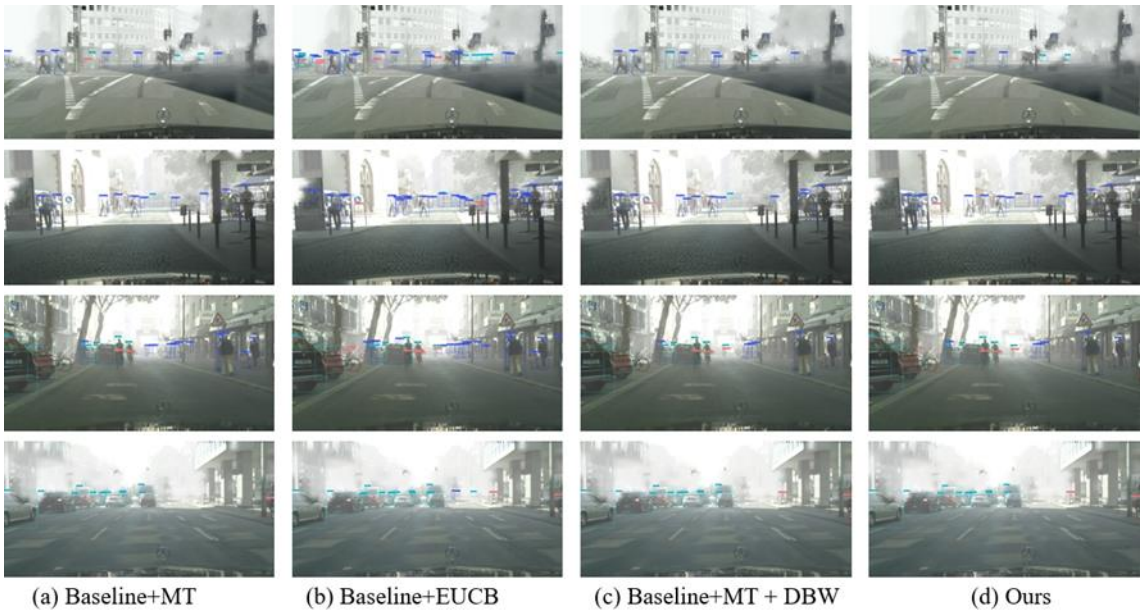

Figure 6. Visualization results of ablation study

5. Conclusion

In this paper, to address the issues of convergence oscillation and excessive regression error in domain-adaptive foggy object detection, we propose a robust domain-adaptive foggy object detection algorithm based on an improved mean teacher framework, named IMT-Det. First, the mean teacher framework is introduced, leveraging its advantages in parameter smoothing and noise robustness. By designing a refined exponential moving average update mechanism, we effectively suppress the instability of convergence during adversarial training, significantly improving the training stability and domain-adaptive generalization ability of the model. Second, an upsampling module based on FEUCB is developed. By enhancing multi-scale feature extraction capability, it effectively tackles the challenges caused by low contrast and target scale variation in foggy images, thus boosting the detection performance for small objects and dense scenes. Third, a dynamic bounding box weighted loss (DBW) is designed, which adaptively adjusts the weight of regression loss to effectively reduce target regression errors

in complex foggy scenes and alleviate the problem of excessive localization deviation. Systematic comparison and ablation experiments on multiple public datasets demonstrate that IMT-Det achieves favorable detection accuracy and robustness in foggy object detection tasks, validating the effectiveness and complementary synergy of each proposed module.

References

[1] Yu B, Chen Y, Cao S, et al. Three-Channel Infrared Imaging for Object Detection in Haze [J]. IEEE transactions on instrumentation and measurement, 2022, 711-13.

[2] Zhang C, Wang H, Cai Y, et al. Robust-FusionNet: Deep Multimodal Sensor Fusion for 3-D Object Detection Under Severe Weather Conditions [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 711-13.

[3] Kang G, Lu J, Yi Y. Contrastive Adaptation Network for Unsupervised Domain Adaptation [A]//2019: 4888-4897.

[4] Wang L, Qin H, Zhou X, et al. R-YOLO: A Robust Object Detector in Adverse Weather [J]. IEEE transactions on instrumentation and measurement, 2023, 721-11.

[5] Li J, Xu R, Ma J, et al. Domain Adaptive Object Detection for Autonomous Driving under Foggy Weather [J]. arXiv, 2022, 2210-15176.

[6] Abbasi H, Amini M. Fog-Aware Adaptive YOLO for Object Detection in Adverse Weather [A]//IEEE, 2023: 1-6.

[7] Hnewa M, Radha H. Integrated Multiscale Domain Adaptive YOLO [J]. IEEE Trans Image Process, 2023, 1857-1867.

[8] Wu H, Qu Y, Lin S, et al. Contrastive Learning for Compact Single Image Dehazing [A]//2021: 10551-10560.

[9] Yang Y, Wang C, Liu R, et al. Self-augmented Unpaired Image Dehazing via Density and Depth Decomposition [A]//IEEE, 2022: 2027-2036.

[10] Liu H, Wu Z, Li L, et al. Towards Multi-domain Single Image Dehazing via Test-time Training [A]//IEEE, 2022: 5821-5830.

[11] Ross G, Jeff D, Trevor D, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [A]//IEEE, 2014: 580-587.

[12] Ross G. Fast R-CNN [J]. Proceedings of the IEEE international conference on computer vision., 2015, 1440-1448.

[13] He Wenqi, Li Jian, Zhao Wenhao. Insulator Anomaly Detection Based on Improved Faster R-CNN [J]. Zhejiang Electric Power, 2021, 40(8): 40-46.

[14] Wei L, Dragomir A, Dumitru E, et al. SSD: Single shot multibox detector [A]//Springer International Publishing, 2016: 21-37.

[15] Lin T, Priya Gl, Ross G, et al. Focal Loss for Dense Object Detection [A]//IEEE transactions on pattern analysis and machine intelligence [C]. 2017: 2999-3007.

[16] Nicolas C, Francisco M, Gabriel S, et al. End-to-End Object Detection with Transformers [A]//Springer International Publishing, 2020: 213-229.

[17] Wang S, Chen R, Wu H, et al. YOLOH: You Only Look One Hourglass for Real-Time Object Detection [J]. IEEE Trans Image Process, 2024, 332104-2115.

[18] Alexey B, Chien-Yao W, Hong-Yuan M. Yolov4: Optimal speed and accuracy of object detection [DB/OL].

[19] Wang A, Chen H, Liu L, et al. YOLOv10: Real-Time End-to-End Object Detection [DB/OL].

[20] Wang C, I-Hau Y, Hong M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information [DB/OL].

[21] Dai D, Wang Y, Chen Y, et al. Is Image Super-resolution Helpful for Other Vision Tasks [A]//2016: 1-9.

[22] Li B, Peng X. An All-in-One Network for Dehazing and Beyond [A]//2017: 4770-4778.

[23] Zhang Q, Hu X. MSFFA-YOLO Network: Multiclass Object Detection for Traffic Investigations in Foggy Weather [J]. IEEE transactions on instrumentation and measurement, 2023, 721-12.

[24] Qiu Y, Lu Y, Wang Y, et al. IDOD-YOLOV7: Image-Dehazing YOLOV7 for Object Detection in Low-Light Foggy Traffic Environments [J]. Sensors (Basel, Switzerland), 2023, 23(3): 1347.

[25] Kuniaki S, Yoshitaka U, Tatsuya H, et al. Strong-Weak Distribution Alignment for Adaptive Object Detection [A]//IEEE, 2019: 6949-6958.

[26] Zhu X, Pang J, Yang C, et al. Adapting Object Detectors via Selective Cross-Domain Alignment [A]//IEEE, 2019: 687-696.

[27] Yu F, Wang D, Chen Y, et al. SC-UDA: Style and Content Gaps aware Unsupervised Domain Adaptation for Object Detection [A]//2022: 1061-1070.

[28] Liu W, Ren G, Yu R, et al. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions [DB/OL].

[29] Giulio M, Luca Z, Elisa R, et al. ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing [A]//2023: 423-433.

[30] Mohamed L M, Davide B, Fabio P. Detect, Augment, Compose, and Adapt: Four Steps for Unsupervised Domain Adaptation in Object Detection [J]. 2023, 2308-15353.

[31] Zhou H, Fei J, Lu H. SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection [J]. Computer Vision and Image Understanding, 2023, 229103649.

[32] Varailhon S, Aminbeidokhti M, Pedersoli M. Source-Free Domain Adaptation for YOLO Object Detection [A]//Springer International Publishing, 2024: 218-235.