2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) | 979-8-3315-4364-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CVPR52734.2025.01111

Detection-Friendly Nonuniformity Correction: A Union Framework for Infrared UAV Target Detection

Houzhang Fang[1†*], Xiaolin Wang[1†], Zengyang Li[1], Lu Wang[1], Qingshan Li[1], Yi Chang[2], Luxin Yan[2]
[1]Xidian University [2]Huazhong University of Science and Technology
houzhangfang@xidian.edu.cn, wxl@stu.xidian.edu.cn, 22009200694@stu.xidian.edu.cn, wanglu@xidian.edu.cn, qshli@mail.xidian.edu.cn, yichang@hust.edu.cn, yanluxin@hust.edu.cn

Abstract

Infrared unmanned aerial vehicle (UAV) images captured using thermal detectors are often affected by temperature-dependent low-frequency nonuniformity, which significantly reduces the contrast of the images. Detecting UAV targets under nonuniform conditions is crucial in UAV surveillance applications. Existing methods typically treat infrared nonuniformity correction (NUC) as a preprocessing step for detection, which leads to suboptimal performance. Balancing the two tasks while enhancing detection-beneficial information remains challenging. In this paper, we present a detection-friendly union framework, termed UniCD, that simultaneously addresses both infrared NUC and UAV target detection tasks in an end-to-end manner. We first model NUC as a small number of parameter estimation problem jointly driven by priors and data to generate detection-conducive images. Then, we incorporate a new auxiliary loss with target mask supervision into the backbone of the infrared UAV target detection network to strengthen target features while suppressing the background. To better balance correction and detection, we introduce a detection-guided self-supervised loss to reduce feature discrepancies between the two tasks, thereby enhancing detection robustness to varying nonuniformity levels. Additionally, we construct a new benchmark composed of 50,000 infrared images in various nonuniformity types, multi-scale UAV targets and rich backgrounds with target annotations, called IRBFD. Extensive experiments on IRBFD demonstrate that our UniCD is a robust union framework for NUC and UAV target detection while achieving real-time processing capabilities. Dataset can be available at https://github.com/IVPLaboratory/UniCD.

1. Introduction

Unmanned aerial vehicle (UAV) detection based on infrared imaging is an important perception technology for monitoring UAV in both day and night scenarios. However, the thermal radiation from the optical lens and the camera housing causes the acquired infrared UAV images to often suffer from the temperature-dependent low-frequency nonuniformity effects [15, 23, 26, 33] (See the left column of Fig. 1). The optics-caused nonuniformity effect is also referred to as the bias field, which reduces the image contrast. Moreover, the infrared UAV targets typically have weak features and complex backgrounds [4, 5]. Nonuniformity bias field further exacerbates the difficulty of UAV target detection. Infrared nonuniformity correction (NUC) and target detection have achieved significant advancements in recent years [2, 4, 5, 17, 23, 28]. Previous methods focus on one aspect of the tasks and address the two tasks independently. As far as we know, no work considers the practical problem: infrared UAV target detection under the nonuniformity conditions.

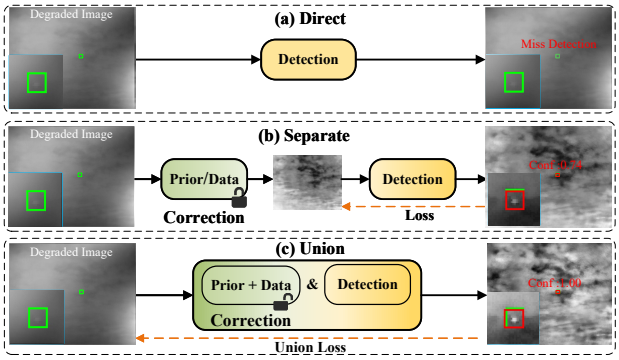


Figure 1. Three main categories of methods for UAV target detection in nonuniformity conditions. (a) Direct: detection models [10] are directly applied to nonuniformity degraded images. (b) Separate: correction model [15] serves as a pre-processing step, correcting images before passing them to detectors [10]. (c) Union: correction and detection are processed simultaneously in a unified framework. Previous methods solely concentrates on optimizing one task. Our UniCD concurrently emphasizes the joint enhancement of correction quality and detection accuracy.

To solve this problem, a simple strategy is to detect [10]

*Corresponding author. †Equal contribution.

2575-7075/25/$31.00 ©2025 IEEE
DOI 10.1109/CVPR52734.2025.01111

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

UAV targets directly on the degraded bias field images (See Fig. 1(a)), which easily leads to miss detection due to the weakening of target information. Another typical approach is the correction-then-detection paradigm (See Fig. 1(b)). NUC methods [15, 22, 26] are first adopted to remove the nonuniformity bias field, and the corrected images are then passed to the target detectors [4, 5, 10]. However, the existing NUC methods have limitations. Specifically, model-based NUC methods rely heavily on handcrafted features to model the images and bias fields, making them prone to overfitting the image content and thus struggle to handle complex or severely degraded bias fields [26]. Deep learning (DL)-based NUC methods depend on complex architectures and a large number of real input-output image pairs, which limits their practicality [2, 23]. Additionally, the NUC lacks supervision from the detection module to enhance detection-conducive information. Recently, joint methods for processing low-level images and high-level vision have already been proposed [11, 12, 18]. However, they are primarily designed for object detection under adverse weather conditions.

To overcome the above limitations, in this paper, we propose a detection-friedly union framework, termed UniCD, that simultaneously tackles both infrared bias field correction and UAV target detection. On the one hand, because of the spatially smooth nature of the bias field, we model it by the high-order bivariate polynomial [16, 22, 26], which can effectively fit the nonuniform bias field with different scales. As a result, accurately estimating the optimal polynomial coefficients is essential for ensuring the performance of bias field correction. In this work, we formulate bias field correction as a problem of predicting a small number of polynomial coefficients jointly driven by priors and data, which can be easily learned by a very lightweight network. The bias field has spatially continuous low-frequency characteristics, making transformer-based architecture well-suited for modeling this component. Additionally, we also integrate convolutional neural networks (CNNs) to capture the local details of the bias field. By incorporating parametric prior modeling and low-dimensional data-driven prediction, our approach avoids dependence on handcrafted features and real input-output data pairs, significantly improving correction performance. On the other hand, existing DL-based detection methods mainly focus on designing complex model architectures for extracting features [4, 5, 17, 19, 28, 31, 32]. We further introduce auxiliary loss with target mask supervision at different stages of the backbone in the infrared UAV target detection network without increasing computational complexity. Integrating this loss enhances the discriminative features of UAV targets while suppressing the background, thereby improving detection performance.

To balance correction and detection, we introduce a detection-guided self-supervised loss to reduce feature discrepancies between the correction and detection tasks. This loss enforces feature similarity between the corrected image and the reference image, both extracted by the detection backbone, thus ensuring high correction quality while enhancing features that are beneficial for detection. Our NUC model can also be flexibly integrated as a scalable module with existing detectors for infrared image bias field correction. Furthermore, we construct the bias field benchmark, IRBFD, consisting of 50,000 infrared images with varying nonuniformity types, multi-scale UAV targets, and rich backgrounds with target annotations, called IRBFD. Experimental results on the IRBFD demonstrate that UniCD outperforms state-of-the-art (SOTA) combined correction and detection methods in terms of precision and recall.

The contributions of this work can be summarized as:

- We propose a novel detection-friedly union framework, termed UniCD, that can simultaneously deal with NUC and infrared UAV target detection in an end-to-end manner. To the best of our knowledge, this is the first work to address both issues in a unified framework.
- We for the first time model nonuniformity bias field correction as a problem of predicting a small number of hyperparameters jointly driven by priors and data, which can be easily performed with a very lightweight network.
- We establish the first large benchmark called IRBFD to facilitate the research in the area of nonuniformity correction and infrared UAV target detection, which consists of 50,000 manually labeled infrared images with various nonuniformity levels, multi-scale UAV targets and rich backgrounds with target annotations.

2. Related Work

2.1. Nonuniformity Correction in Infrared Images

NUC methods for removing the bias field are broadly divided into two main categories: model-driven methods and data-driven methods. Model-driven methods typically leverage the prior constraints of the bias field and the images to remove the nonuniformity effects [1, 15, 16, 22, 25, 26]. However, these methods rely on handcrafted features with carefully tuned hyper-parameters, limiting their practicality in real-world applications. Data-driven DL methods [2, 14, 23] have gained increasing attention. However, DL-based correction models are often complex and rely on large amounts of training data with real correction labels, which restricts their widespread application in practical scenarios [6, 24, 27]. In contrast, we formulate a novel lightweight correction model driven by both parametric priors and data, which converges more easily.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

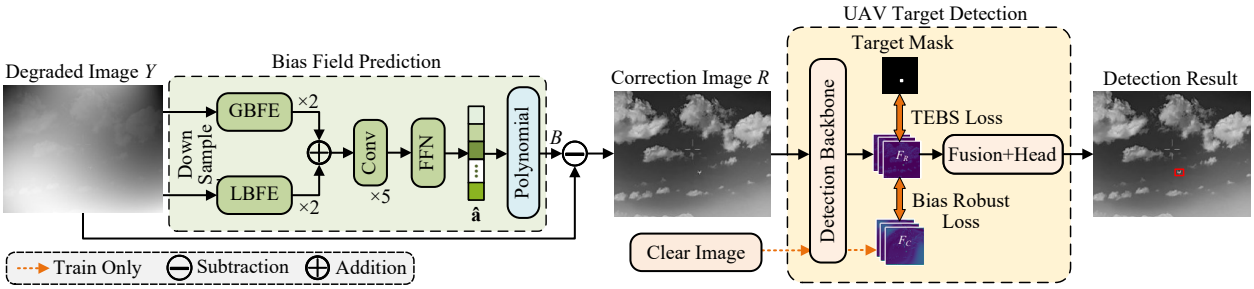


Figure 2. Overview of the proposed UniCD. Our UniCD integrates a bias field prediction network with an infrared UAV target detection network. These two components are fused into a unified pipeline and trained end-to-end. The target enhancement and background suppression (TEBS) loss is introduced to enhance UAV target features while suppressing the background. The bias robust loss is employed to balance correction and detection.

2.2. Infrared UAV Target Detection

In recent years, many methods have been developed to detect infrared UAV targets [3–5, 19, 21]. Infrared UAV targets are challenging due to their weak imaging features and complex backgrounds [7, 8, 29]. TAD [19] leveraged the inconsistent motion cues between UAV targets and the background to detect potential targets. Fang et al. [3] formulated the UAV detection task as a residual image prediction by learning the mapping from input images to residual images. DAGNet [5] introduced attention mechanisms to adaptively enhance the network's ability to discriminate between UAVs and the background. DANet [4] constructed a dynamic attention network for UAVs to enhance feature extraction capabilities. However, the above methods primarily focus on designing complex network structures for improving detection performance, while rarely exploring how to better enhance feature representation without increasing computational complexity. We focus on introducing new auxiliary losses into the backbone to boost the model's ability to represent UAV target features while suppressing the background.

2.3. Joint Low-Level Image Processing and High-Level Vision Tasks

Recently, joint approaches for image enhancement and object detection have emerged to further improve detection performance on low-quality images. One category of these methods utilized encoder-decoder architectures for image enhancement [12], but these modules are complex and hinder real-time performance. Another approach uses classical mathematical models for enhancement, replacing manually designed parameters with predictions from deep networks. IA-YOLO [18] integrated an adaptive enhancement strategy with YOLOv3 [20] via a differentiable image processing module, improving detection in foggy conditions. BAD-Net [11] introduced a dual-branch structure to minimize the impact of poor dehazing performance on the detection module. However, the above methods are developed for visible light images degraded by adverse weather conditions. To the best of our knowledge, no work has explored handling both NUC and UAV target detection for infrared imaging within a single framework. We propose an end-to-end framework to simultaneously improve both correction and detection performance.

3. Detection-Friendly Union Method

3.1. Overall Architecture

In this section, we propose a novel network architecture, UniCD, as shown in Fig. 2. The UniCD leverages a lightweight prediction network to estimate the bias field and then passes the corrected image to a UAV detection network tailored for UAV targets. Additionally, during the joint training of image correction and detection tasks, we employ a detection-guided self-supervised loss to minimize feature discrepancies between the two tasks. Finally, we construct a new dataset, IRBFD, to validate our approach.

3.2. Prior- and Data-Driven Nonuniformity Correction

Infrared UAV images captured by thermal detectors often suffer from low-frequency nonuniformity, which significantly impacts target detection performance. Existing model-driven correction methods often struggle to handle complex non-uniformities, while deep learning-based methods suffer from high network complexity. To address these issues, we propose a lightweight correction network that combines parametric prior knowledge with the strong learning capabilities of deep neural networks. The network architecture is shown in Fig. 2.

Generally, the degraded infrared image can be represented as follows [2, 16]:

$$Y = C + B, \tag{1}$$

where Y, C, and B represent the degraded image, the clear image, and the bias field, respectively. Thus, once B is available, the corrected image R can be obtained as $Y - B$. Under ideal conditions, R is theoretically equivalent to the clear image C. The bias field B possesses a

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

spatially smooth property, allowing us to model it using the following bivariate polynomial:

$$B\left(x_i, y_j\right) = \sum_{t=0}^{D} \sum_{s=0}^{D-t} a_{t,s} x_i^t y_j^s = \mathbf{m}^\top \mathbf{a}, \tag{2}$$

where (x_i, y_j), D, and $\mathbf{m}$ denote the image coordinates, the degree of the polynomial, and the column vector holding the monomial terms, respectively. The column vector $\mathbf{a}$ represents the coefficients of the polynomial formed by concatenating $\{a_{t,s}\}$. To reduce the redundancy of the basic plane and the computational complexity of higher-order models, we set the degree D to 3.

Accurate estimation of the polynomial coefficients $\mathbf{a}$ is crucial for improving bias field correction performance. In this work, we design a lightweight bias field prediction network that can estimate the model parameters accurately and efficiently.

In our NUC module, we first downsample the degraded image Y by a factor of two, resulting in the downsampled image Y_{down}. Then, we utilize the global bias field encoder (GBFE) and the local bias field encoder (LBFE) to extract features at different granularities. The GBFE, inspired by the RSTB [13] module, includes two Swin Transformer layers with the hidden layer channel dimension reduced to 16. The LBFE consists of two spatial attention modules in series, allowing it to capture localized features more effectively. The global and local features, F_{global} and F_{local}, are fused to form the final feature representation F_{fused}:

$$F_{\text{fused}} = \text{GBFE}(Y_{\text{down}}) + \text{LBFE}(Y_{\text{down}}). \tag{3}$$

Finally, the fused features F_{fused} pass through five 3×3 convolutional layers (Conv_5) and a fully connected (FC) layer to predict the final coefficient vector $\hat{\mathbf{a}}$:

$$\hat{\mathbf{a}} = \text{FC}(\text{Conv}_5(F_{\text{fused}})). \tag{4}$$

This approach enables the network to adaptively adjust to inputs with varying levels of degradation, effectively predicting the coefficient vector $\hat{\mathbf{a}}$ for bias field correction.

Analysis. Compared to existing model-driven methods, our approach avoids reliance on hand-crafted features. Compared to existing data-driven methods, our approach transforms the high-dimensional image space prediction into a low-dimensional data-driven problem with a few hyperparameters, effectively reducing computational complexity and improving correction performance.

Loss Function. We calculate the mean absolute error (MAE) loss between the predicted coefficients $\hat{\mathbf{a}}$ and the predefined coefficients $\mathbf{a}$ to minimize the discrepancy between them. The MAE loss is defined as:

$$L_{cor} = \frac{1}{N} \|\hat{\mathbf{a}} - \mathbf{a}\|_2^2, \tag{5}$$

where N is the total number of elements in the vector $\mathbf{a}$. By minimizing this loss, we improve the accuracy of the corrected image, making it closer to the ideal clear image. This loss function is used to pre-train the NUC module separately, serving as the initial weights for the joint training of correction and detection.

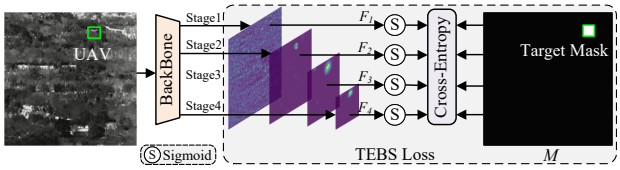


Figure 3. Calculation of the proposed feature enhancement and background suppression (TEBS) loss.

3.3. Mask-Supervised Infrared UAV Detector

We select DANet [4] as our baseline detector, which constructs a multi-scale dynamic perception network to address the challenges of multi-scale variations in UAV targets. Infrared UAV targets typically have weak features and complex backgrounds, and existing methods often enhance feature extraction through complex model architectures. In this work, we design an auxiliary loss function for further target enhancement and background suppression (TEBS) without increasing network complexity, as illustrated in Figs. 2 and 3.

Specifically, we first convert the bounding boxes of the ground-truth into a binary mask M, assigning a value of 1 to the target region and 0 to the background region:

$$M(x, y) = \begin{cases} 1, & \text{if } (x, y) \in \text{target region}, \\ 0, & \text{if } (x, y) \in \text{background region}. \end{cases} \tag{6}$$

This mask is then used to compute a binary cross-entropy loss with the feature maps F_i from the i-th stage of the backbone network. The TEBS loss L_{TEBS} is obtained by summing the losses from four stages of the backbone and averaging:

$$L_{TEBS} = \frac{1}{4} \sum_{i=1}^{4} L_{CE}(M, F_i), \tag{7}$$

where $L_{CE}(\cdot, \cdot)$ denotes the cross-entropy loss.

The TEBS loss offers three key benefits: (1) Supervision on the target regions helps the backbone network to quickly focus on the feature learning of infrared UAV targets and enhance localization accuracy; (2) Supervision on background regions effectively suppresses non-target features, thereby reducing clutter and noise interference in the features; (3) The loss enhances training efficiency by guiding the network to learn target features more accurately, leading to faster convergence.

The classification and regression losses are kept consistent with the baseline method. The final detector loss is written as:

$$L_{det} = L_{cls} + L_{reg} + \lambda L_{TEBS}, \tag{8}$$

where λ is set to 1 for the first 20 epochs to accelerate convergence and improve localization performance in the early

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

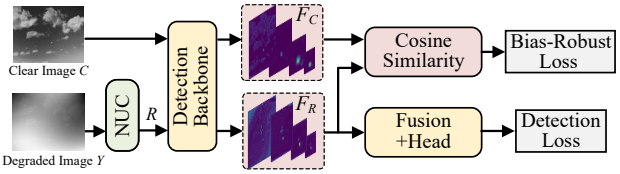


Figure 4. Construction of the bias-robust loss in the unified NUC and detection framework.

stages. After 20 epochs, λ is reduced to 0.01 to balance with the gradually decreasing loss and avoid impacting classification accuracy.

3.4. Balance Correction and Detection with Bias-Robust Loss

Existing research [11, 18] shows that incorporating supervision losses from low-level vision tasks in joint training can hinder high-level vision task performance. This is because low-level tasks focus on preserving fine image details, while high-level tasks aim to extract target-specific features and ignore irrelevant background information. This conflict can hinder convergence during joint training, causing the high-level vision task to settle into a local optimum.

To address this issue, we design a self-supervised loss, named bias robust (BR) loss, to achieve detection-friendly NUC, as illustrated in Fig. 4. Specifically, during joint training, the clear image C and the corrected image R obtained via the NUC module are simultaneously fed into the backbone of the detection model. Let $F_C^{(i)}$ and $F_R^{(i)}$ represent the feature maps from the i-th stage of the detection backbone, where $i = 1, 2, 3, 4$ denotes the four different stages. Here, clear images are employed only for the joint training of correction and detection on the synthetic dataset and are not applied to NUC training on real infrared images.

To evaluate the alignment of the corrected image with the clear image in the feature space, we compute the cosine similarity between the feature maps at each stage i as follows:

$$\text{Cos_Sim}(F_C^{(i)}, F_R^{(i)}) = \frac{F_C^{(i)} \cdot F_R^{(i)}}{\left\|F_C^{(i)}\right\| \left\|F_R^{(i)}\right\|}. \tag{9}$$

The details of the cosine similarity calculation are placed in the supplementary materials. The BR loss is obtained by summing the cosine similarities across all stages and averaging them:

$$L_{BR} = \frac{1}{4} \sum_{i=1}^{4} \left(1 - \text{Cos_Sim}(F_C^{(i)}, F_R^{(i)})\right). \tag{10}$$

The BR loss function is designed to maximize the consistency between the feature representations of the clear and corrected images, thereby enhancing the fidelity of the corrected image within the feature space and ensuring that it retains essential characteristics beneficial for detection. The final union loss function for the joint training process L_{uni} is written as:

$$L_{uni} = L_{det} + L_{BR}. \tag{11}$$

As illustrated in Fig. 1, the union loss L_{uni} ensures effective bias field correction while maximizing detection performance by backpropagating through each part of the network. Compared to separate training of NUC and detection networks, this integrated approach enables a more balanced optimization, leading to enhanced results in both correction and detection tasks.

3.5. IRBFD Dataset

We construct a new benchmark called IRBFD, comprising 30,000 synthetic nonuniformity infrared UAV images (IRBFD-syn) and 20,000 real-world infrared UAV images with nonuniformity field (IRBFD-real). The IRBFD-syn subset provides paired degraded and clear images, with the synthesis process based on Eqs. (1) and (2). The characteristics of the dataset are summarized as follows. (1) Multiple background types. The IRBFD includes multiple complex scenes, such as dense clouds, buildings, forests, urban areas, and sea. (2) Multi-scale variations. The distance between the UAVs and the sensor ranges from 50 meters to 2 kilometers, resulting in multi-scale variations of the targets. (3) Multiple UAV types. Such as the DJI Inspire, Matrice, Phantom, Mavic, and Mini series. All UAV positions are manually annotated. IRBFD serves as a comprehensive resource for evaluating the impact of non-uniformities on UAV target detection in real-world environments. All images have a size of 640×512. Additional details can be found in the supplementary material.

4. Experiment

4.1. Datasets and Evaluation Metrics

Datasets. We use two parts of the IRBFD dataset: IRBFD-syn and IRBFD-real. IRBFD-syn consists of simulated nonuniform infrared images, allowing us to train the model with controlled background images and varying nonuniformities. Training, validation, and testing sets are split in an 8:1:1 ratio. We train on the simulated dataset and directly validate on real-world dataset to demonstrate the generalizability of our method.

Evaluation Metrics. For NUC, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as objective evaluation metrics. For target detection, we evaluate the detection performance using precision (P) and recall (R). Lastly, for real-time performance, we use frames per second (FPS) as the evaluation metric. The signal-to-clutter ratio gain (SCRG) is the ratio of SCR in the corrected image to that in the original image, used to evaluate the improvement in target detectability achieved by the correction method.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

Table 1. Quantitative comparison of the proposed method with SOTA methods on the synthetic dataset IRBFD-syn. **Bold** and underline indicate the best and the second best results, respectively.

Strategy	Module			Metrics				
	NUC	Detection	Pub'Year	PSNR	SSIM	P	R	FPS
Direct	-	Deformable DETR	ICLR'21	-	-	0.614	0.630	24
		DINO	ICLR'23			0.904	0.640	26
		DAGNet	TII'23			0.994	0.635	**43**
		LESPS	CVPR'23			0.033	0.446	12
		MSHNet	CVPR'24			0.407	0.421	41
		YOLO11L	2024			0.963	0.602	42
Separate	Liu	YOLO11L	IPT'16	16.800	0.8289	0.898	0.574	<1
		DAGNet				0.978	0.578	<1
	DMRN	YOLO11L	GRSL'19	24.467	0.8600	0.923	0.550	35
		DAGNet				0.966	0.595	36
	Shi	YOLO11L	AO'22	13.974	0.7783	0.924	0.455	<1
		DAGNet				0.966	0.472	<1
	TV-DIP	YOLO11L	IPT'23	13.397	0.6374	0.131	0.020	29
		DAGNet				0.599	0.020	30
	AHBC	YOLO11L	TGRS'24	13.954	0.6763	0.825	0.080	<1
		DAGNet				0.724	0.040	<1
Union	UniCD		-	**31.961**	**0.9827**	**0.999**	**0.822**	32

Table 2. Quantitative comparison of the proposed method with SOTA methods on the real dataset IRBFD-real.

Strategy	NUC	Detection	SCRG	P	R
Direct	-	DINO	-	0.971	0.660
		YOLO11L		0.966	0.843
		DAGNet		0.992	0.871
Separate	TV-DIP	YOLO11L	0.412	0.521	0.024
		DAGNet		0.663	0.026
	AHBC	YOLO11L	1.146	0.940	0.633
		DAGNet		0.986	0.699
Union	UniCD		**1.286**	**0.994**	**0.901**

4.2. Implementation Details

We use Adam as the optimizer, with a learning rate of 0.001. The training lasts for 50 epochs, with a weight decay of 10^{-4} and a batch size of 4. During training, we only apply random horizontal flipping for data augmentation. Our experiments are conducted on an NVIDIA RTX 4090 with CUDA 12.4 and PyTorch 1.7. For NUC, we select Liu [15], Shi [22], and AHBC [26] as the traditional bias field correction methods; DMRN [2] and TV-DIP [14] as the DL-based correction methods. They are all designed for infrared images. For target detection, we select YOLO11L [9] (large version), DAGNet [5], LESPS [28], and MSHNet [17] as the convolutional neural network (CNN)-based target detection methods. DAGNet, LESPS, and MSHNet are designed for infrared target detection. Deformable DETR [34] and DINO [30] are the representative Transformer-based detection methods.

4.3. Quantitative Results

As shown in Tab. 1, the existing detection methods obtain low P and R values when detecting directly on the degraded bias field images. DAGNet has a high P value of 0.994 but a low recall. This indicates that the bias field has an adverse effect on the UAV target detection. For the separate strategy, the DL-based method DMRN enhances image quality, achieving notable improvements in PSNR and SSIM. Other correction methods, such as Liu, Shi, and AHBC, are less effective for severely degraded images, which further impacts detection accuracy. Table 2 shows that the three detection methods achieve high P and R when directly detecting on real images with low degradation levels. YOLO11L and DAGNet have lower P and R values for images corrected by TV-DIP, as TV-DIP not only lacks corrective effects but also deteriorates the image content. In contrast, the proposed UniCD achieves the best performance in terms of all the evaluation metrics compared with SOTA methods in Tab. 1 for the synthetic dataset and Tab. 2 for the real dataset. Specially, our UniCD achieves a real-time processing speed of 32 FPS. We also plot the R-R curves for our UniCD and other correction-then-detection methods on the IRBFD dataset shown in Fig. 5. Higher values of the area under the curve indicate better performance. It can be seen that our UniCD achieves the largest area under the curve among all correction-then-detection methods.

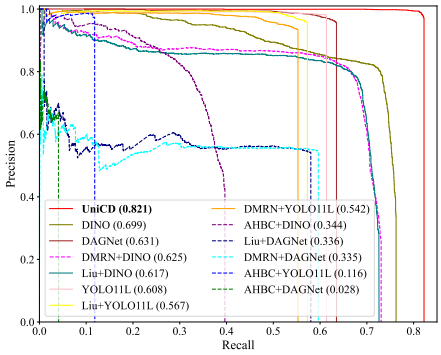


Figure 5. P-R curves of our UniCD and other correction-then-detection methods on the IRBFD-syn. The area values under the curves are placed after the method names.

4.4. Qualitative Results

As demonstrated in Fig. 6, we show qualitative results from various separate correction-then-detection methods and our UniCD on the IRBFD-syn dataset across three distinct scenarios: buildings, hillside, and clouds. As can be seen, even in severe degraded bias field situations, our UniCD can still perform high-quality image correction while accurately detecting the UAV targets. This is because the proposed NUC correction module integrates parametric modeling and a small number of model parameter predictions, enabling more accurate parameter estimation. Meanwhile, our detection network introduces auxiliary loss with target mask supervision into the backbone to enhance the features of UAV targets while suppressing the background, thereby improving detection performance. Conventional correction methods, such as Liu and AHBC, have limited modeling capabilities and are prone to producing bias field residuals.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

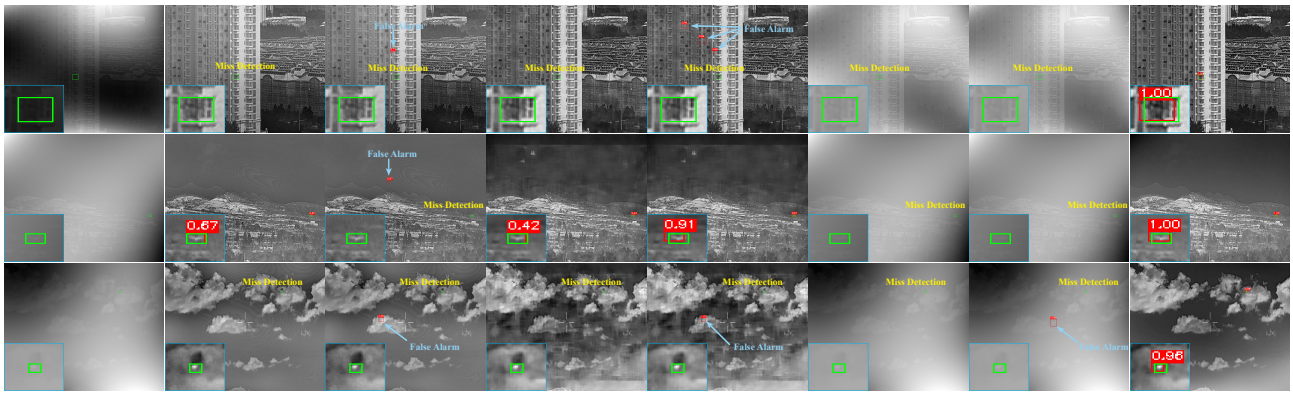


Figure 6. Visual comparison of results from separate correction followed by detection methods and our UniCD on the IRBFD-syn dataset. Closed-up views are shown in the left bottom corner. Boxes in green and red represent ground-truth and correctly detected targets, respectively.

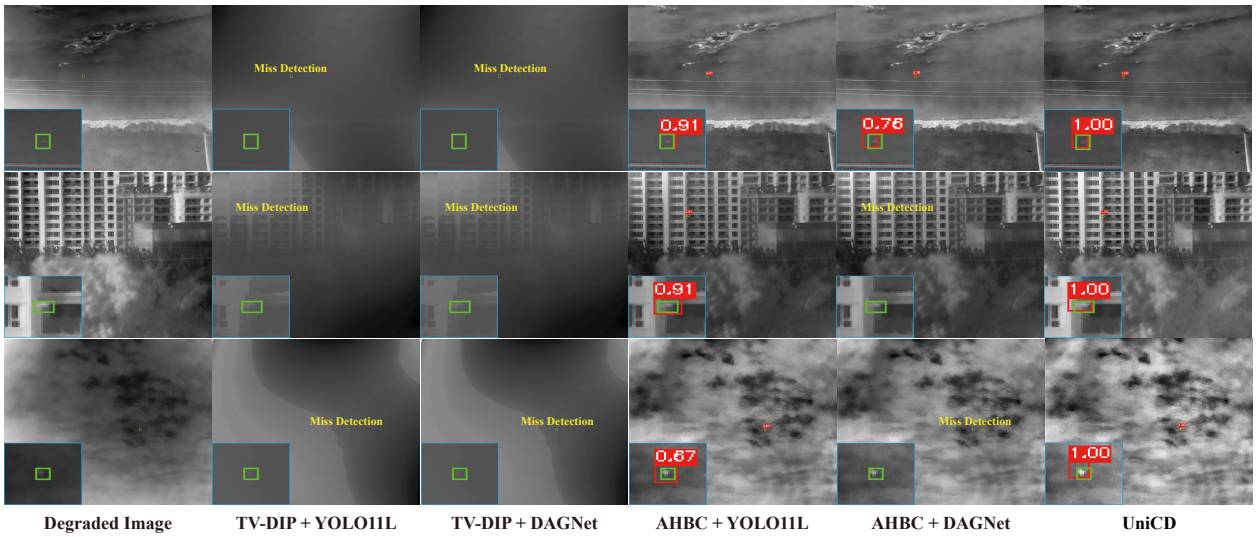


Figure 7. Visual comparison of results from separate correction-then-detection methods and our UniCD on the IRBFD-real dataset.

Table 3. Ablation study of polynomial degrees.

Degree	Number of coefficients	PSNR	SSIM	P	R
2	6	13.5279	0.7590	0.991	0.433
3	10	**39.050**	**0.9970**	**0.997**	**0.810**
4	15	31.744	0.9890	**0.997**	0.788
5	21	29.070	0.9830	**0.997**	0.787

Table 4. Ablation study of the LBFE and GBFE modules.

LBFE	GBFE	Params (M)	FLOPs (G)	PSNR	SSIM	FPS
×	×	0.3786	0.1192	27.218	0.9808	555
√	×	0.3812	0.2747	30.242	0.9877	370
×	√	0.3940	1.5159	37.888	0.9961	151
√	√	0.3966	1.6809	**39.050**	**0.9970**	116

The corrected image from DMRN exhibits block artifacts. Existing detection methods produce false alarms or miss detections on the above-corrected images; similar results are observed in Fig. 7 for scenes degraded by real bias fields. See the supplementary material for more visual results.

4.5. Ablation Study

In this section, we report ablation study results.

Impact of polynomial degrees. We conduct experiments to determine the optimal polynomial degree, as described in Tab. 8. The results indicate that the highest values for all correction and detection metrics are achieved when the degree is set to 3. Compared to higher-order polynomials, a third-order polynomial has lower complexity and less redundance. A lower degree implies weaker modeling capability, resulting in poor correction performance.

Impact of the LBFE and GBFE components. As presented in Tab. 4, when using only LBFE or GBFE, we can see that the PSNR of the correction model already achieves 30.242 and 37.888, respectively, surpassing the 27.218 of the baseline. When combined with LBFE and GBFE, further improvements can be achieved, reaching 39.050 with only 1.6809G FLOPs and 0.3966M parameters on an image size of 640×512. Additionally, the SSIM value has also shown some slight improvements. This suggests that the Transformer architecture combined with the CNN structure

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

Table 5. Ablation study of the auxiliary TEBS loss.

TEBS Loss	P	R
w/o	0.993	0.762
w	**0.997**	**0.810**

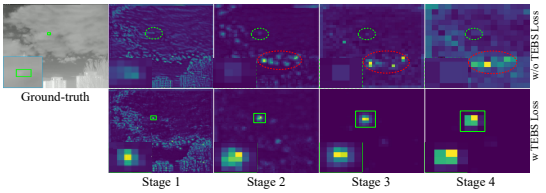


Figure 8. Comparison of feature maps from the detection backbone at different stages with and without TEBS loss.

Table 6. Ablation study of BR loss on the synthetic dataset IRBFD-syn. Here, the direct and separate strategies utilize the correction and detection modules proposed in UniCD.

Strategy	Correction loss	Detection loss	BR loss	Metrics			
				PSNR	SSIM	P	R
Direct	×	√	×	-	-	0.998	0.694
Separate	√	√	×	**37.722**	**0.9960**	**0.999**	0.793
Union	√	√	×	33.024	0.9827	0.998	0.791
Union	×	√	×	17.940	0.8910	0.989	0.811
UniCD	×	√	√	31.961	0.9827	**0.999**	**0.822**

can effectively model bias fields, significantly improving performance while maintaining high real-time efficiency.

Impact of the auxiliary TEBS loss. As shown in Tab. 5, the auxiliary TEBS loss added to the backbone of the UAV detection network leads to improvements in both P and R performance. The TEBS loss imposes strong constraints on the target and background masks for the backbone features, enhancing target features while suppressing the background, thus improving detection performance, as depicted in Fig. 8. In the absence of TEBS loss, the targets are generally weak and there is considerable residual background.

Impact of BR loss. As shown in Tab. 6, the direct detection results in low P and R. The correction-then-detection separate method without BR loss in the second row obtains high PSNR, SSIM and P, but with low R because of the independent processing of the two tasks. The union of correction and detection without BR loss in the third row leads to a decrease in PSNR, SSIM, P, and R due to the conflict between the two tasks. The union of correction and detection without correction and BR losses in the fourth row significantly reduces PSNR and SSIM owing to the detection module's sole constraint. Our UniCD achieves the highest R value and P value exceeding 0.99, demonstrating that our union framework effectively balances correction and detection through the self-supervised BR loss.

Effectiveness of UniCD on real dataset. As shown in Tab. 7, we use the NUC module weights trained on the IRBFD-syn and directly test them on the IRBFD-real. This configuration is referred to as UniCD*. The direct and separate methods with the modules in UniCD achieve a high P but a low R. Our UniCD obtains the highest R value and P value exceeding 0.99, thereby validating the effectiveness and generalization of our union framework on the real dataset.

Table 7. Ablation study of our UniCD on the IRBFD-real.

Strategy	SCRG	P	R
Direct	-	0.992	0.887
Separate	**1.286**	**0.998**	0.812
UniCD*	**1.286**	0.994	**0.901**

Table 8. Ablation study of varying levels of nonuniformity.

K-value	PSNR	SSIM
3	29.119	0.9891
5	34.907	0.9950
12	38.361	0.9968

Table 9. Ablation study of the union of our scalable NUC module with existing detection methods.

Our NUC	Detection	P	R	FPS
×	YOLO11L	0.835	0.075	42
	DAGNet	0.711	0.036	43
	LESPS	0.007	0.122	12
	MSHNet	0.284	0.221	42
√	YOLO11L	$0.977_{(+0.142)}$	$0.657_{(+0.582)}$	31
	DAGNet	$0.997_{(+0.286)}$	$0.722_{(+0.686)}$	31
	LESPS	$0.007_{(+0.000)}$	$0.459_{(+0.337)}$	11
	MSHNet	$0.776_{(+0.492)}$	$0.701_{(+0.480)}$	31

Generalization of the NUC Module. In Tab. 8, we test the UniCD on images with various degrees of nonuniformity degradation without retraining the NUC module. We control the severity of degradation using the formula $Y = C + k * B$, where the k-value determines the level of degradation. The correction results at different k-values indicate that the NUC module generalizes well to various degradation levels.

Scalability of the NUC Module. From Tab. 9, we observe that, except for the P value of LESPS, introducing our correction module significantly boosts the detection performance of several recent general and infrared target detection methods. This indicates that our NUC module can be flexibly integrated as a scalable component into existing detectors to enhance infrared images for detection purposes.

5. Conclusion

In this paper, we propose UniCD, an end-to-end framework that simultaneously addresses bias field correction and infrared UAV target detection. We develop a NUC module that removes bias fields and restores clear images with parameters adaptively predicted by a lightweight network. Additionally, we introduce auxiliary losses with mask supervision to enhance UAV target features and suppress the background. We also present a self-supervised feature loss to improve the robustness of detection to varying bias levels. Moreover, we construct a new dataset IRBFD to facilitate future research. Experimental results show that our UniCD outperforms previous approaches in both synthetic and real-world scenarios. Furthermore, our method shows great potential for deployment on resource-constrained edge devices.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

Acknowledgments. This work was supported by the Open Research Fund of the National Key Laboratory of Multispectral Information Intelligent Processing Technology under Grant 61421132301, and the Natural Science Foundation of Jiangsu Province under Grant BK202302028.

References

[1] Yanpeng Cao and Christel-Loic Tisse. Single-image-based solution for optics temperaturedependent nonuniformity correction in an uncooled long-wave infrared camera. *Optics Letters*, 39(3):646–648, 2014. 2

[2] Yi Chang, Luxin Yan, Li Liu, Houzhang Fang, and Sheng Zhong. Infrared aerothermal nonuniform correction via deep multiscale residual network. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1120 – 1124, 2019. 1, 2, 3, 6

[3] Houzhang Fang, Lan Ding, Liming Wang, Yi Chang, Luxin Yan, and Jinhui Han. Infrared small UAV target detection based on depthwise separable residual dense network and multiscale feature fusion. *IEEE Transactions on Instrumentation and Measurement*, 71:1–20, 2022. 3

[4] Houzhang Fang, Zikai Liao, Lu Wang, Qingshan Li, Yi Chang, Luxin Yan, and Xuhua Wang. DANet: Multi-scale UAV target detection with dynamic feature perception and scale-aware knowledge distillation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2121–2130, 2023. 1, 2, 3, 4

[5] Houzhang Fang, Zikai Liao, Xuhua Wang, Yi Chang, and Luxin Yan. Differentiated attention guided network over hierarchical and aggregated features for intelligent UAV surveillance. *IEEE Transactions on Industrial Informatics*, 19(9):9909–9920, 2023. 1, 2, 3, 6

[6] Blake Gella, Howard Zhang, Rishi Upadhyay, Tiffany Chang, Matthew Waliman, Yunhao Ba, Alex Wong, and Achuta Kadambi. WeatherProof: A paired-dataset approach to semantic segmentation in adverse weather. *arXiv preprint arXiv:2312.09534*, 2023. 2

[7] Bo Huang, Jianan Li, Junjie Chen, Gang Wang, Jian Zhao, and Tingfa Xu. Anti-UAV410: A thermal infrared benchmark and customized scheme for tracking drones in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):2852–2865, 2024. 3

[8] Nan Jiang, Kuiran Wang, Xiaoke Peng, Xuehui Yu, Qiang Wang, Junliang Xing, Guorong Li, Guodong Guo, Qixiang Ye, Jianbin Jiao, Jian Zhao, and Zhenjun Han. Anti-UAV: A large-scale benchmark for vision-based UAV tracking. *IEEE Transactions on Multimedia*, 25:486–500, 2023. 3

[9] Glenn Jocher and Jing Qiu. Ultralytics YOLO11, 2024. 6

[10] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8, 2023. 1, 2

[11] Chengyang Li, Heng Zhou, Yang Liu, Caidong Yang, Yongqiang Xie, Zhongbo Li, and Liping Zhu. Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8284–8295, 2023. 2, 3, 5

[12] Yi Li, Yi Chang, Changfeng Yu, and Luxin Yan. Close the loop: A unified bottom-up and top-down paradigm for joint image deraining and segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1438–1446, 2022. 2, 3

[13] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844, 2021. 4

[14] Kang Liu, Honglei Chen, Wenzhong Bao, and Jianlu Wang. Thermal imaging spatial noise removal via deep image prior and step-variable total variation regularization. *Infrared Physics & Technology*, 134:104888, 2023. 2, 6

[15] Li Liu, Luxin Yan, Hui Zhao, Xiaobing Dai, and Tianxu Zhang. Correction of aeroheating-induced intensity nonuniformity in infrared images. *Infrared Physics and Technology*, 76:235–241, 2016. 1, 2, 6

[16] Li Liu, Luping Xu, and Houzhang Fang. Simultaneous intensity bias estimation and stripe noise removal in infrared images using the global and local sparsity constraints. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3): 1777 – 1789, 2020. 2, 3

[17] Qiankun Liu, Rui Liu, Bolun Zheng, Hongkui Wang, and Ying Fu. Infrared small target detection with scale and location sensitivity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17490–17499, 2024. 1, 2, 6

[18] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive YOLO for object detection in adverse weather conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1792–1800, 2022. 2, 3, 5

[19] Yanyi Lyu, Zhunga Liu, Huandong Li, Dongxiu Guo, and Yimin Fu. A real-time and lightweight method for tiny airborne object detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3016–3025, 2023. 2, 3

[20] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement, 2018. 3

[21] Artem Rozantsev, Vincent Lepetit, and Pascal Fua. Detecting flying objects using a single moving camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5): 879–892, 2017. 3

[22] Yu Shi, Jisong Chen, Hanyu Hong, Yaozong Zhang, Nong Sang, and Tianxu Zhang. Multi-scale thermal radiation effects correction via a fast surface fitting with Chebyshev polynomials. *Applied Optics*, 61(25):7498 – 7507, 2022. 2, 6

[23] Yu Shi, Yixin Zhou, Lei Ma, Lei Wang, and Hanyu Hong. WDTSNet: Wavelet decomposition two-stage network for infrared thermal radiation effect correction. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1 – 17, 2024. 1, 2

[24] Rosaura G. VidalMata, Sreya Banerjee, Brandon Richard-Webster, Michael Albright, Pedro Davalos, Scott McCloskey, Ben Miller, Asong Tambo, Sushobhan Ghosh, Sudarshan Nagesh, Ye Yuan, Yueyu Hu, Junru Wu, Wenhan Yang, Xiaoshuai Zhang, Jiaying Liu, Zhangyang Wang,

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.

Hwann-Tzong Chen, Tzu-Wei Huang, Wen-Chi Chin, Yi-Chun Li, Mahmoud Lababidi, Charles Otto, and Walter J. Scheirer. Bridging the gap between computational photography and visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12):4272–4290, 2021. 2

[25] Yu Wang, Xiubao Sui, Yihong Wang, Tong Liu, Chuncheng Zhang, and Qian Chen. Contrast enhancement method in aero thermal radiation images based on cyclic multi-scale illumination self-similarity and gradient perception regularization. *Optics Express*, 32(2):1650–1668, 2024. 2

[26] Jun Xie, Lingfei Song, and Hua Huang. Thermal radiation bias correction for infrared images using Huber function-based loss. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. 1, 2, 6

[27] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J. Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, Yuqiang Zheng, Yanyun Qu, Yuhong Xie, Liang Chen, Zhonghao Li, Chen Hong, Hao Jiang, Siyuan Yang, Yan Liu, Xiaochao Qu, Pengfei Wan, Shuai Zheng, Minhui Zhong, Taiyi Su, Lingzhi He, Yandong Guo, Yao Zhao, Zhenfeng Zhu, Jinxiu Liang, Jingwen Wang, Tianyi Chen, Yuhui Quan, Yong Xu, Bo Liu, Xin Liu, Qi Sun, Tingyu Lin, Xiaochuan Li, Feng Lu, Lin Gu, Shengdi Zhou, Cong Cao, Shifeng Zhang, Cheng Chi, Chubing Zhuang, Zhen Lei, Stan Z. Li, Shizheng Wang, Ruizhe Liu, Dong Yi, Zheming Zuo, Jianning Chi, Huan Wang, Kai Wang, Yixiu Liu, Xingyu Gao, Zhenyu Chen, Chang Guo, Yongzhou Li, Huicai Zhong, Jing Huang, Heng Guo, Jianfei Yang, Wenjuan Liao, Jiangang Yang, Liguo Zhou, Mingyue Feng, and Likun Qin. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29: 5737–5752, 2020. 2

[28] Xinyi Ying, Li Liu, Yingqian Wang, Ruojing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, and Shilin Zhou. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15528–15538, 2023. 1, 2, 6

[29] Shenghai Yuan, Yizhuo Yang, Thien Hoang Nguyen, Thien-Minh Nguyen, Jianfei Yang, Fen Liu, Jianping Li, Han Wang, and Lihua Xie. MMAUD: A comprehensive multi-modal Anti-UAV dataset for modern miniature drone threats. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2745–2751, 2024. 3

[30] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel Ni, and Heung-Yeung Shum. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2023. 6

[31] Mingjin Zhang, Rui Zhang, Yuxiang Yang, Haichen Bai, Jing Zhang, and Jie Guo. ISNet: shape matters for infrared small target detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 867–876, 2022. 2

[32] Mingjin Zhang, Handi Yang, Jie Guo, Yunsong Li, Xinbo Gao, and Jing Zhang. IRPruneDet: efficient infrared small target detection via wavelet structure-regularized soft channel pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7224–7232, 2024. 2

[33] Yuanjie Zheng and James C. Gee. Estimation of image bias field with sparsity constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 255–262, 2010. 1

[34] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable DETR: Deformable transformers for end-to-end object detection. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2021. 6

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on August 15,2025 at 08:57:11 UTC from IEEE Xplore. Restrictions apply.