

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

## Supplementary Material

Houzhong Fang<sup>1†\*</sup>, Xiaolin Wang<sup>1†</sup>, Zengyang Li<sup>1</sup>, Lu Wang<sup>1</sup>, Qingshan Li<sup>1</sup>, Yi Chang<sup>2</sup>, Luxin Yan<sup>2</sup>

<sup>1</sup>Xidian University <sup>2</sup>Huazhong University of Science and Technology

houzhongfang@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

## 1. Overview

In this supplementary material, we provide additional details and more experimental results to further validate the proposed UniCD framework. The structure of this material is organized as follows.

**Details of the IRBFD Dataset.** In Section 2, we describe the construction of the IRBFD dataset, including the synthetic process for generating nonuniformity bias fields in IRBFD-syn and the characteristics of real-world scenes in IRBFD-real. These datasets aim to comprehensively benchmark nonuniformity correction (NUC) and UAV detection methods.

**Further Validation Results of the UniCD Framework.** In Section 3, we detail the training process of UniCD, including both separate and union training phases. We highlight the design of loss functions to ensure effective cooperation between the NUC and detection modules.

In Section 4, we present additional quantitative results that demonstrate the superior performance of UniCD compared to state-of-the-art methods. UniCD achieves high detection accuracy and robustness across synthetic and real-world datasets, even under challenging scenarios.

In Section 5, we provide more qualitative results to visually illustrate the advantages of UniCD. The visualizations emphasize its ability to correct severe nonuniformity while maintaining accurate target detection.

**Further Ablation Studies.** In Section 6, we conduct ablation studies to evaluate the impact of key components, such as the degree of polynomial used in the NUC module and the inclusion of the TEBS and BR losses. These studies confirm the adaptability and effectiveness of UniCD under various configurations.

**Details of Metric Computation.** In Section 7, we calculate the signal-to-clutter ratio gain (SCRG) to quantify the improvement in target detectability achieved by UniCD.

In Section 8, we compute the cosine similarity between feature maps to analyze the feature alignment introduced by different correction methods. This further supports the superiority of UniCD in enhancing feature representation for UAV detection.

This supplementary material highlights the robustness, efficiency, and effectiveness of the UniCD framework across a wide range of experimental scenarios, providing strong support for the claims made in the main paper.

## 2. More Details about the Dataset IRBFD

### 2.1. Generation Process of the Synthetic Dataset IRBFD-syn

**Selection of Infrared Clear Images.** To ensure the broad applicability and diversity of the dataset, we select the largest publicly available infrared UAV dataset [3] as the foundation. From this dataset, we uniformly and randomly sampled 30,000 images to serve as the infrared clear images. After sampling, we conduct a manual review to ensure that every image contains UAV targets. These images, along with their tracking annotations, are converted into the VOC dataset format for ease of use and compatibility with existing tools.

**Generation of Degraded Images with Nonuniformity Bias Fields.** We model nonuniformity bias field using the following bivariate polynomial:

$$B(x_i, y_j) = \sum_{t=0}^D \sum_{s=0}^{D-t} a_{t,s} x_i^t y_j^s, \quad (1)$$

where coefficients  $a_{t,s}$  are randomly setted to simulate varying bias levels.  $D$  denotes the degree of the polynomial. By varying the degree  $D$ , we can obtain bias fields with different basis surfaces. In our work, we set the degree  $D$  to 3. Then, by adding the bias field to a clear infrared image, we

\*Corresponding author. †Equal contribution.

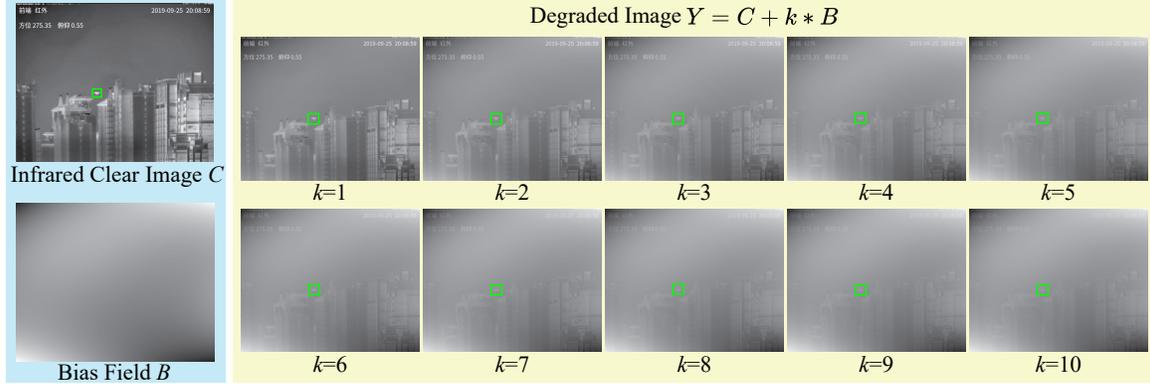


Figure 1. Visualization of degraded images with varying levels of nonuniformity degradation controlled by  $k$ .

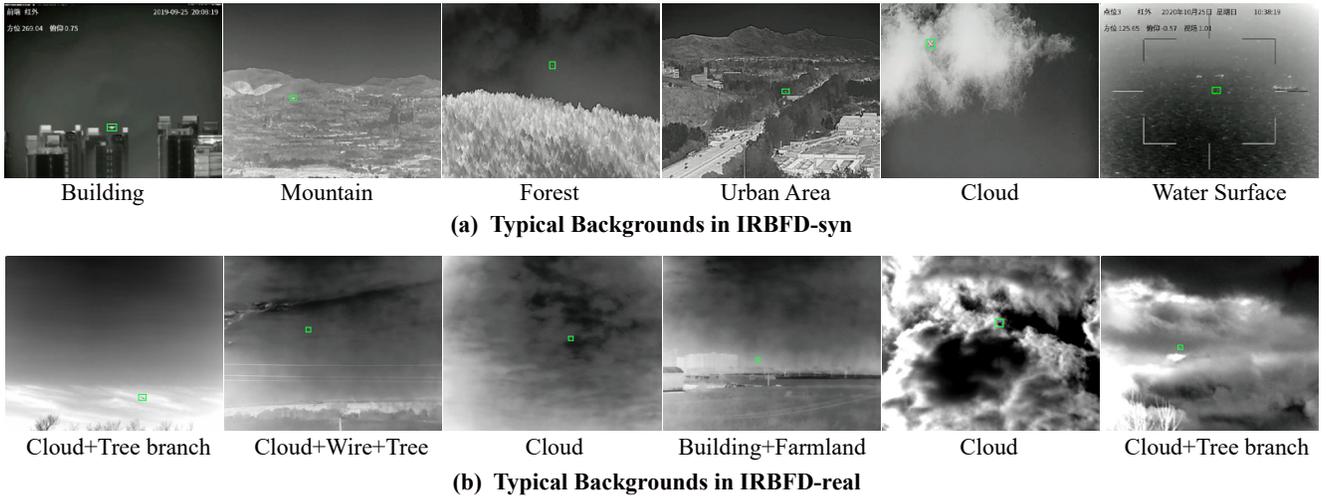


Figure 2. Typical backgrounds in UAV surveillance scenarios from the IRBFD dataset.

can obtain the degraded image with bias fields:

$$Y = C + k * B, \quad (2)$$

where  $Y$ ,  $C$ , and  $B$  represent the degraded image, the clean image, and the bias field, respectively.  $k$  is employed to control the severity of nonuniformity. The effects of different levels of degradation are shown in Fig. 1. When the degree of degradation is very low ( $k$  is much smaller than 10), existing target detection methods can detect UAV targets without requiring correction. However, as the degradation becomes more severe ( $k$  is much greater than 10), most existing correction methods fail and existing detection methods also struggle to detect the UAV targets. Therefore, we set  $k=10$  in the dataset to evaluate the effectiveness of UniCD under challenging nonuniformity conditions.

## 2.2. Statistics and Analysis of the IRBFD Dataset

**Diversity of Backgrounds.** In Fig. 2, the first row presents typical background images from the synthetic

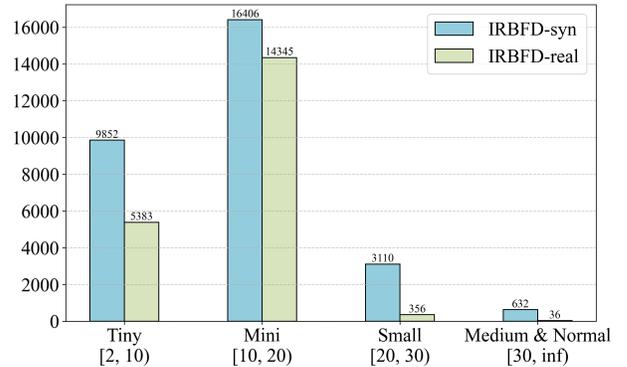


Figure 3. Statistics of target scales in the IRBFD dataset.

dataset IRBFD-syn. For clarity of visualization, we display the infrared clear images at the top of Fig. 2. These scenes encompass two different lighting conditions (day and night), two seasons (autumn and winter), and a vari-

ety of backgrounds, including buildings (30%), mountains (20%), forests (5%), urban areas (30%), clouds (10%), and water surfaces (3%) [3].

In the real dataset IRBFD-real, we collect UAV target data under various real-world bias field degradation scenarios. The dataset encompasses a diverse range of backgrounds, including dense clouds, trees, power lines, buildings, and farmlands, as shown at the bottom of Fig. 2. These real-world scenes were carefully selected to reflect practical environmental complexities, providing a comprehensive testbed for evaluating the robustness of the proposed method against real nonuniformity effects.

**Scale Variation of Infrared UAV Targets.** As shown in Fig. 3, in the IRBFD-syn dataset, tiny-scale UAV targets (Tiny, [2, 10]) account for approximately 37.8% (9,852 targets), mini-scale UAV targets (Mini, [10, 20]) account for approximately 63.1% (16,406 targets), small-scale UAV targets (Small, [20, 30]) account for approximately 11.9% (3,110 targets), and medium & normal-scale UAV targets (Medium & Normal, [30, inf]) account for less than 2.5% (632 targets).

In contrast, in the IRBFD-real dataset, tiny-scale UAV targets account for approximately 25.8% (5,383 targets), mini-scale UAV targets account for approximately 68.8% (14,345 targets), small-scale UAV targets account for only 1.7% (356 targets), and medium & normal-scale UAV targets account for less than 0.2% (36 targets).

The real-world dataset, IRBFD-real, closely reflects the challenges of real-world anti-UAV scenarios. The significant proportion of tiny and mini UAV targets highlights the difficulty of detecting smaller targets, making it more representative of real-world application needs.

**Comprehensiveness of UAV Target Position Distribution.** Figure 4 illustrates the position distribution of UAV targets within the IRBFD dataset, including both the synthetic dataset (IRBFD-syn) and the real-world dataset (IRBFD-real).

The UAV targets in the IRBFD dataset are comprehensively distributed across the entire image space, as shown in both subfigures. This design ensures that the datasets cover a wide range of spatial configurations, providing diverse scenarios for evaluating nonuniformity correction (NUC) and UAV detection methods. The distribution of targets reflects a balanced dataset that supports robust model training and testing across varying environmental conditions.

### 3. More Training Details about the UniCD

When training UniCD on the IRBFD-syn dataset, we initially train the NUC module and the infrared UAV target detection module separately. During the training of the NUC module, the loss function is defined as shown in Sec. 3, Eq. (5) of the main text. For the infrared UAV target detection module, the loss function is defined as shown in

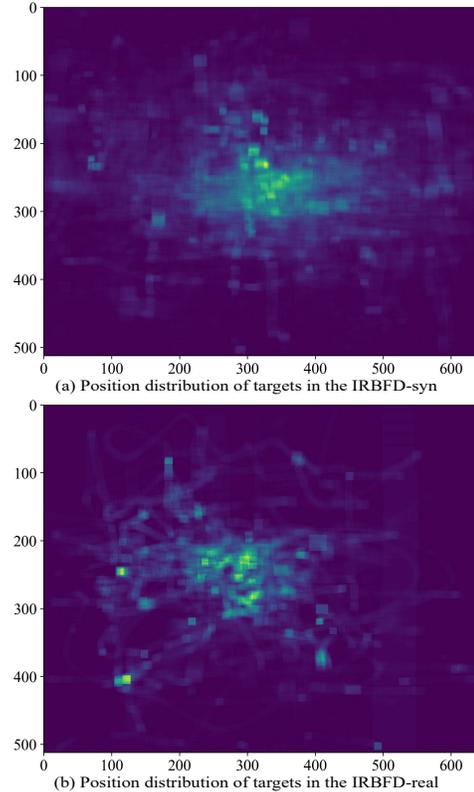


Figure 4. Position distribution of targets in the IRBFD dataset.

Eq. (8) of the main text. Once both sub-modules are sufficiently trained, we proceed to train the union framework. In this phase, the loss function is given by Eq. (11) of the main text, where the BR loss is designed to balance the conflict between the NUC module and the UAV target detection module.

On the IRBFD-real dataset, we pre-train only the UAV target detection module, while the NUC module retains the weights trained on the IRBFD-syn dataset without further updates. During union training, the NUC module is frozen, and the loss function defined in Eq. (11) of the main text is used for optimization.

### 4. More Quantitative Comparisons of UniCD with Existing Methods

We further expand the experiments on the synthetic dataset IRBFD-syn and the real-world dataset IRBFD-real, with the results presented in Tab. 1 and Tab. 2, to verify the superior performance of the proposed UniCD framework.

On the IRBFD-syn dataset, we evaluate additional target detection methods (DINO [10] and MSHNet [7]) under separate strategies. The experiments demonstrate that, among the separate strategies, DL-driven correction methods (e.g., DMRN [1]) significantly improve image qual-

Table 1. Quantitative comparison of the proposed method with SOTA methods on the synthetic dataset IRBFD-syn. For the separation strategy, each correction method corresponds to multiple different detection methods. **Bold** and underline indicate the best and the second best results, respectively.

Strategy	Module			Metrics				
	NUC	Detection	Pub'Year	PSNR $\uparrow$	SSIM $\uparrow$	P $\uparrow$	R $\uparrow$	FPS $\uparrow$
Direct	-	Deformable DETR	ICLR'21	-	-	0.614	0.630	24
		DINO	ICLR'23			0.904	0.640	26
		DAGNet	TII'23			<u>0.994</u>	0.635	<b>43</b>
		LESPS	CVPR'23			0.033	0.446	12
		MSHNet	CVPR'24			0.407	0.421	41
		YOLO11L	2024			0.963	0.602	<u>42</u>
Separate	Liu	DINO	IPT'16	16.800	0.8289	0.868	0.599	<1
		MSHNet				0.686	<u>0.663</u>	<1
		YOLO11L				0.898	0.574	<1
		DAGNet				0.978	0.578	<1
		DINO				0.841	0.585	23
	DMRN	MSHNet	GRSL'19	24.467	0.8600	0.669	0.663	35
		YOLO11L				0.923	0.550	35
		DAGNet				0.966	0.595	36
		DINO				0.813	0.557	<1
	Shi	MSHNet	AO'22	13.974	0.7783	0.584	0.618	<1
		YOLO11L				0.924	0.455	<1
		DAGNet				0.966	0.472	<1
		DINO				0.086	0.115	21
	TV-DIP	MSHNet	IPT'23	13.397	0.6374	0.078	0.077	29
		YOLO11L				0.131	0.020	29
		DAGNet				0.599	0.020	30
		DINO				0.417	0.294	<1
	AHBC	MSHNet	TGRS'24	13.954	0.6763	0.300	0.208	<1
YOLO11L		0.825				0.080	<1	
DAGNet		0.724				0.040	<1	
DINO		0.417				0.294	<1	
Union	UniCD	-	<b>31.961</b>	<b>0.9827</b>	<b>0.999</b>	<b>0.822</b>	32	

Table 2. Quantitative comparison of the proposed method with SOTA methods on the real dataset IRBFD-real. **Bold** and underline indicate the best and the second best results, respectively.

Strategy	NUC	Detection	SCRG $\uparrow$	P $\uparrow$	R $\uparrow$
Direct	-	DINO	-	0.971	0.660
		YOLO11L		0.966	0.843
		DAGNet		<u>0.992</u>	<u>0.871</u>
TV-DIP	-	DINO	0.412	0.094	0.083
		YOLO11L		0.521	0.024
		DAGNet		0.663	0.026
		DINO		0.687	0.480
Separate	DMRN	YOLO11L	0.997	0.918	0.296
		DAGNet	0.929	0.345	
		DINO	0.964	0.649	
AHBC	-	YOLO11L	<u>1.146</u>	0.940	0.633
		DAGNet	0.986	0.699	
		DINO	0.964	0.649	
Union	UniCD	-	<b>1.286</b>	<b>0.994</b>	<b>0.901</b>

ity and achieve notable PSNR and SSIM results. However, their detection performance is limited due to the independent handling of correction and detection modules, resulting in low recall. Traditional model-driven correction methods (e.g., Liu [6], Shi [8], and AHBC [9]) struggle to handle severe degradation scenarios, further impairing detection accuracy. In contrast, UniCD achieves a PSNR of 31.961 and an SSIM of 0.9827 while maintaining outstanding detection precision (P = 0.999) and recall (R = 0.822),

showcasing its ability to handle image correction and target detection simultaneously in a unified framework.

On the IRBFD-real dataset, we expand the analysis with more NUC methods and target detection combinations under a separate processing strategy. The experiments reveal that separate strategies, such as TV-DIP [5] and AHBC, face significant limitations on real-world data. For instance, TV-DIP often deteriorates the image content, resulting in extremely low detection precision and recall. While AHBC performs better in certain scenarios, its overall precision and recall still fell short compared to UniCD. In contrast, UniCD consistently achieves a SCRГ of 1.286 and surpassed all other methods with a precision of P = 0.994 and a recall of R = 0.901. These results highlight UniCD's adaptability and robustness in handling diverse real-world nonuniformity scenarios.

Furthermore, UniCD's real-time processing capability (32 FPS) ensures its practical deployment even under resource-constrained conditions. The experimental results further confirm that UniCD provides a comprehensive solution to address the challenges posed by nonuniformity in infrared UAV detection tasks.

## 5. More Qualitative Comparisons of UniCD with Existing Methods

We evaluate more correction-then-detection methods on both synthetic and real-world datasets across diverse scenarios, including buildings, hillsides, clouds, forests, and urban regions. These visualizations further validate the findings from the quantitative experiments presented in the main paper.

Figure 5 presents the qualitative results on the synthetic dataset. The analysis is supplemented by including DINO and MSHNet as detection methods applied to various NUC correction approaches. The results indicate that traditional NUC methods, such as Liu, TV-DIP, and AHBC, paired with DINO or MSHNet, often fail to effectively handle severe nonuniformity effects:

- Liu + DINO/MSHNet results in extremely low detection confidence in most scenarios (e.g., 0.02 confidence or undetected targets in hillside scenes) due to insufficient correction of the bias field.
- TV-DIP + DINO/MSHNet fails to restore sufficient target features, particularly in forested or cluttered environments, leading to missed detections.
- AHBC + DINO/MSHNet, while slightly more effective, still produces false positives or weak confidence scores.

In contrast, our UniCD consistently outperforms all these combinations, achieving clear corrected images and accurate UAV target detection with high confidence scores (1.00 across all scenarios). This superior performance is attributed to the robust integration of parametric modeling in the NUC module and the auxiliary loss in the detection

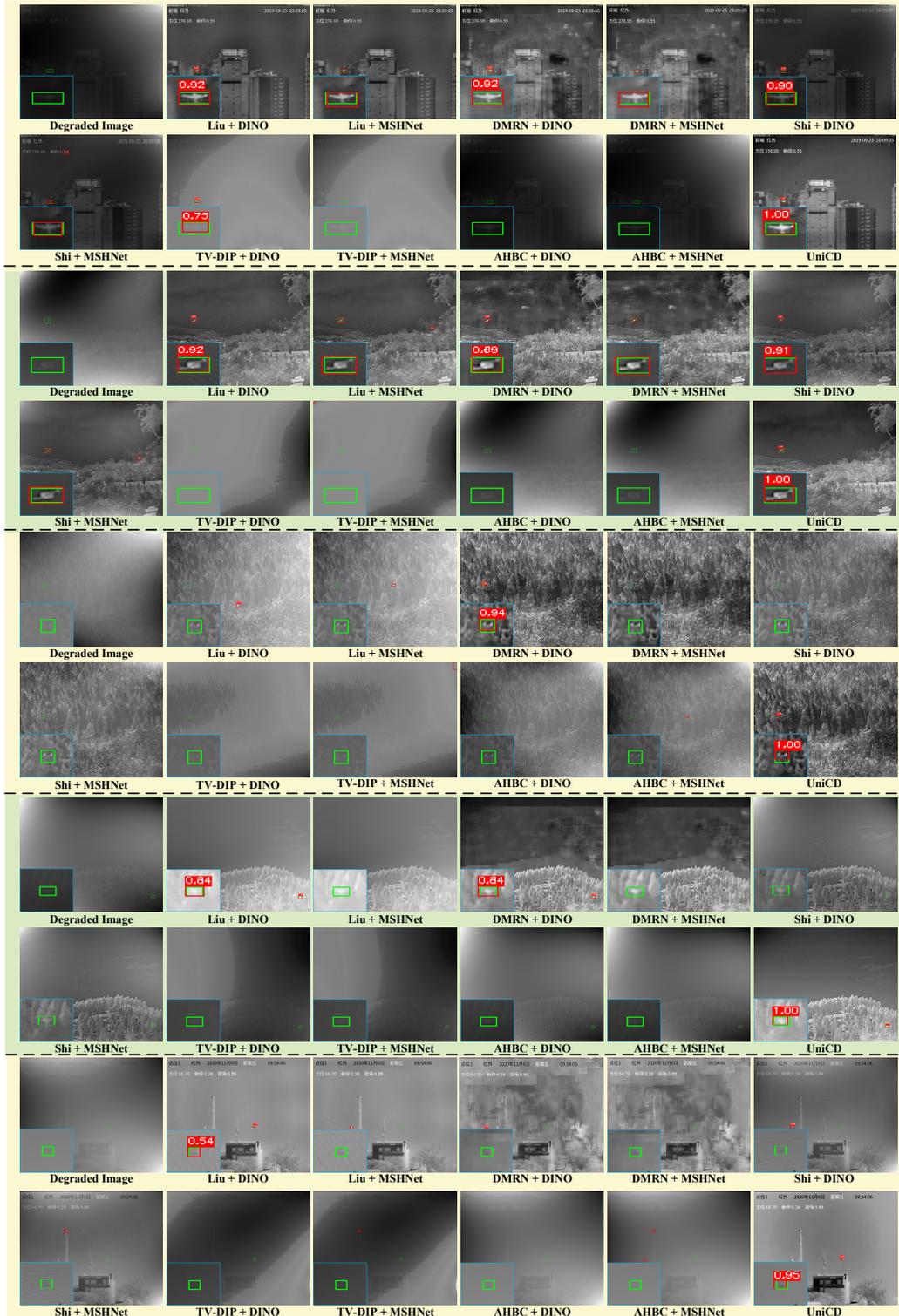


Figure 5. Visual comparison of results from separate correction-then-detection methods and our UniCD on the synthetic dataset IRBFD-syn . Closed-up views are shown in the left bottom corner. Boxes in green and red represent ground-truth and correctly detected targets, respectively.

backbone, enabling accurate parameter estimation and enhanced feature representation under severe nonuniformity conditions.

Figure 6 shows the qualitative results on the real-world dataset. We further add the following content: (1) DINO as the detection network applied to different NUC correction methods. (2) DMRN as a newly introduced NUC method, combined with three detection methods: DINO, YOLO11L, and DAGNet. The results demonstrate the limitations of conventional NUC methods when applied to real-world data:

- TV-DIP + DINO fails to effectively correct the bias field and instead degrades image content, resulting in extremely low detection confidence scores or entirely missed targets.
- DMRN + DINO, while improving bias field correction, introduces block artifacts that impair detection accuracy. DMRN + YOLO11L [4] and DMRN + DAGNet [2] show moderate improvements but struggle to maintain consistent detection in complex urban or cloudy scenes.
- AHBC + DINO performs inconsistently, with detection confidence dropping as low as 0.36 in dense cloud scenarios.

In contrast, our UniCD provides consistently superior results across all scenarios, achieving high detection confidence scores (1.00) even in challenging conditions, such as dense clouds and cluttered urban environments. These results highlight UniCD’s ability to simultaneously balance correction and detection tasks while effectively addressing complex real-world nonuniformity effects.

Overall, these supplementary visualization results demonstrate the superiority of UniCD compared to conventional approaches in both synthetic and real-world settings. This further reinforces UniCD’s capability to address challenging nonuniformity effects while delivering robust and reliable UAV detection performance.

## 6. Further Ablation Studies

### 6.1. Impact of Polynomial Degree on the NUC Module

In this section, we provide a detailed discussion on the rationale for selecting the third-degree polynomial model.

**Definition of the Polynomial Basis Surfaces.** As presented in Eq. (1), the polynomial model  $B(x_i, y_j)$  serves as an effective representation of a surface in a two-dimensional space, where each term  $x_i^t y_j^s$  represents a basis surface that contributes to the overall shape of the surface. The influence of each basis surface is modulated by its corresponding coefficient  $a_{t,s}$ , which determines the weight and impact of that component. By appropriately adjusting these coefficients, the polynomial model can capture complex spatial variations and transformations, making it well-suited

Table 3. Robustness analysis of UniCD across different levels of nonuniformity degradation.

$k$	PSNR	SSIM	P	R
3	29.119	0.9891	0.998	0.820
5	34.907	0.9950	0.998	0.820
12	38.361	0.9968	0.998	0.820

for modeling nonuniformity bias fields. This structure representation strikes a balance between expressiveness and computational efficiency, providing a flexible yet compact framework for accurately representing diverse patterns of bias field distortions.

**Why Choose the Third-degree Polynomial?** As shown in Fig. 7, the comparison of 2D basis surfaces and their corresponding 3D visualizations across different polynomial degrees highlights key considerations for selecting the optimal degree. Higher-degree polynomials inherently encompass the basis surfaces of lower-degree ones, which indicates that very low-degree polynomials lack sufficient expressive power to model complex spatial variations in the bias field effectively. However, as the degree increases beyond three, redundant basis surfaces emerge (highlighted in red boxes in the figure), introducing overlapping representations. This redundancy not only diminishes modeling efficiency but also increases the risk of overfitting and numerical instability.

In addition, higher-degree polynomials require estimating a larger number of coefficients, which complicates the optimization process and slows convergence during training. In contrast, the third-degree polynomial strikes an ideal balance by offering adequate expressive power while maintaining computational simplicity and stability, making it the optimal choice for our framework.

Furthermore, the quantitative results in Tab. 3 of the main text also demonstrate that the third-degree polynomial achieves the best performance among all configurations. A second-degree polynomial, with only 6 coefficients, struggles to model the bias field accurately, resulting in poor correction performance (PSNR = 13.5279, SSIM = 0.7590) and low recall (R = 0.433). Increasing the degree to three improves both correction quality (PSNR = 39.050, SSIM = 0.9970) and detection performance (R = 0.810), achieving the best balance between accuracy and complexity. Polynomials of higher degrees (e.g., 4 and 5) introduce more coefficients, leading to a slight decline in correction performance and recall (e.g., PSNR = 31.744, R = 0.788 for degree 4). This indicates diminishing returns and potential risks of overfitting.

### 6.2. Impact of Different Levels of Degradation on UniCD

To analyze the adaptability of our proposed UniCD framework, we conduct experiments on the IRBFD-syn dataset

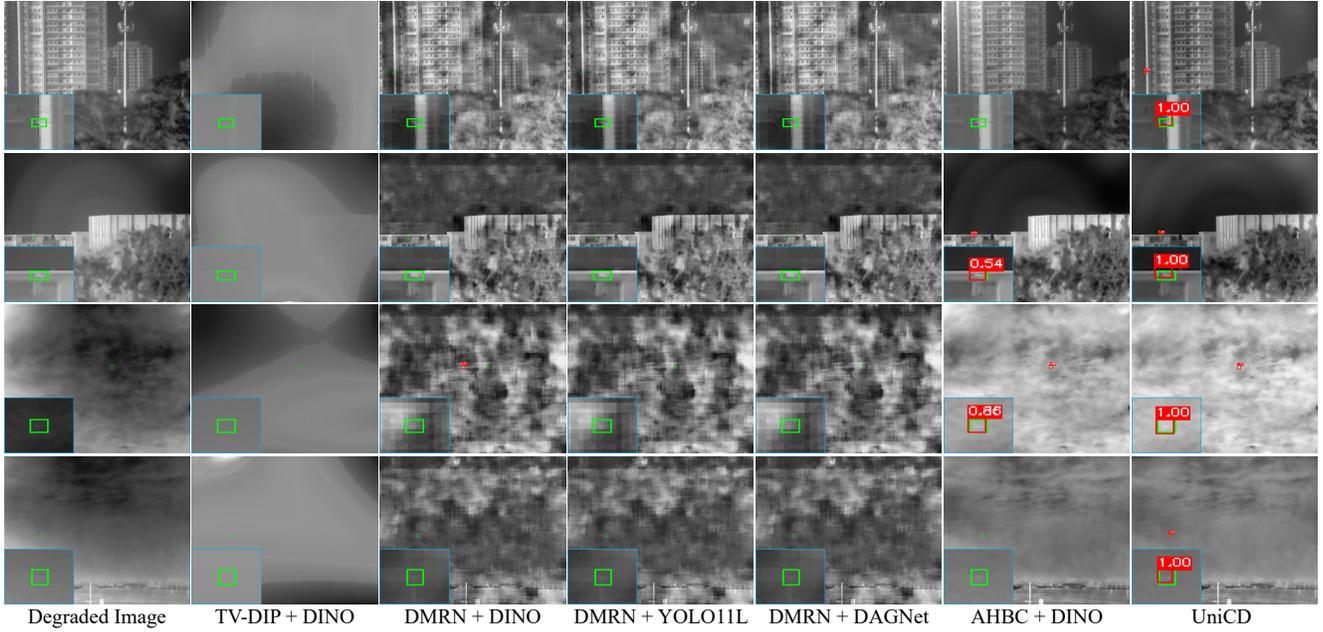


Figure 6. Visual comparison of results from separate correction-then-detection methods and our UniCD on the real dataset IRBFD-real.

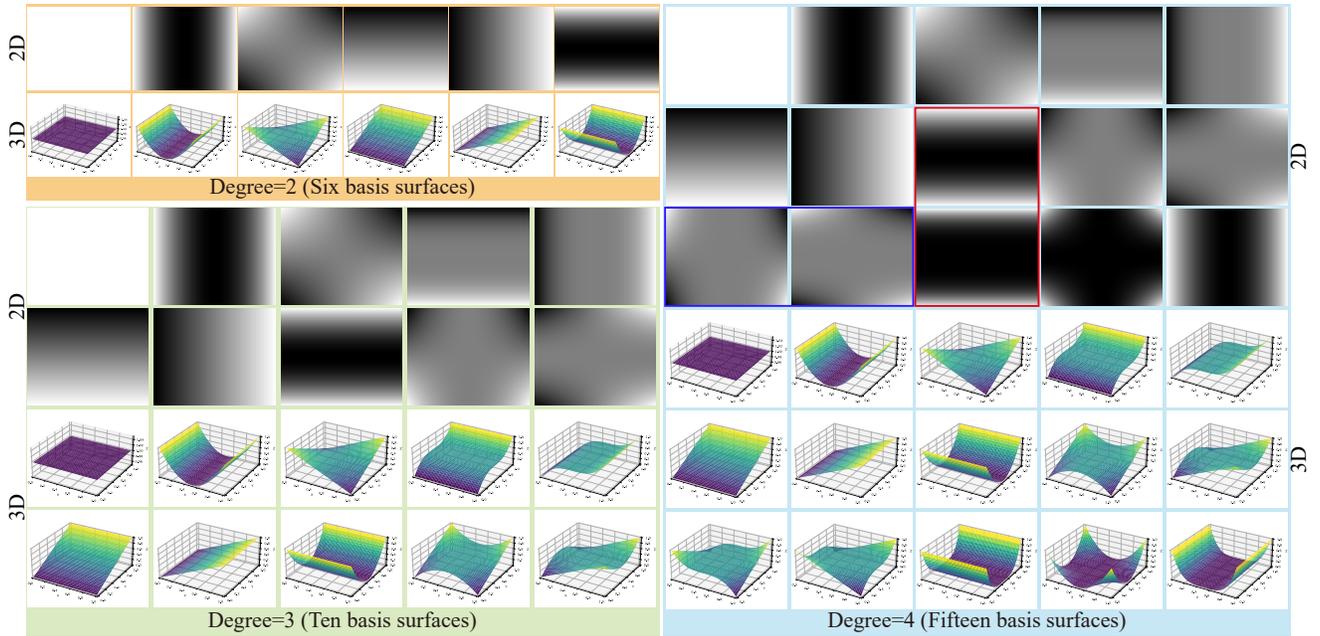


Figure 7. Visualization of 2D and 3D basis surfaces for polynomials of different degrees. The surface sets for the second-degree, third-degree, and fourth-degree models are highlighted with yellow, green, and blue background boxes, respectively. We employ blue and red boxes to represent similar redundant surfaces for a degree equal to 4.

by training the model on images with a fixed degradation level ( $k = 10$ ) and testing it across varying levels of degradation ( $k = 3, 5, 12$ ). The results, summarized in Tab. 3 demonstrate the robustness of our method under different nonuniformity conditions.

Specifically, the PSNR and SSIM values increase consistently with higher  $k$ -values, indicating that the bias field correction module effectively reconstructs clearer images as the degradation severity rises. This improvement reflects the model’s ability to handle more challenging conditions

Table 4. Comparison of number of parameters and computational complexity for different NUC methods.

Methods	Params (M)	FLOPs (G)
DMRN	0.7398	76.2524
TV-DIP	2.1817	75.3930
Our NUC	0.3966	1.6809

with greater nonuniformity intensity.

Despite the variation in image quality, the detection performance metrics (P and R) remain consistent across all tested  $k$ -values. Both precision (P = 0.998) and recall (R = 0.820) are maintained at high levels, showcasing UniCD’s ability to balance correction and detection tasks effectively. This result highlights the flexibility and generalization capability of the framework, ensuring reliable performance even when deployed under varying real-world degradation scenarios.

These findings confirm that UniCD is not only robust to different nonuniformity levels but also achieves stable detection results, reinforcing its practical applicability in diverse environmental conditions.

### 6.3. Number of Parameters and Computational Complexity of the NUC Module

Table 4 gives a comparison of number of parameters and computational complexity (measured in FLOPs) among different NUC methods, including DMRN, TV-DIP, and our proposed NUC module. They are all methods based on deep learning. The results demonstrate the significant advantages of our NUC method in terms of both model efficiency and computational complexity.

- **Number of Parameters.** Our NUC module requires only 0.3966 million parameters, which is approximately 46.4% smaller than DMRN (0.7398M) and 81.8% smaller than TV-DIP (2.1817M). This reduction in number of parameters directly translates to lower memory usage and faster runtime performance, making it more suitable for resource-constrained applications.
- **Computational Complexity.** The proposed NUC module achieves a significant reduction in Floating Point Operations (FLOPs), requiring only 1.6809G, which is 55 times smaller than DMRN (76.2524G) and 44 times smaller than TV-DIP (75.3930G). Such a significant reduction indicates that our module is highly optimized for real-time processing, further enhancing its practical deployment capabilities.

In summary, the results demonstrate that our NUC module not only achieves superior nonuniformity correction performance but also does so with minimal computational overhead, significantly outperforming other methods in terms of efficiency. These attributes make our method an ideal solution for real-world scenarios where computational

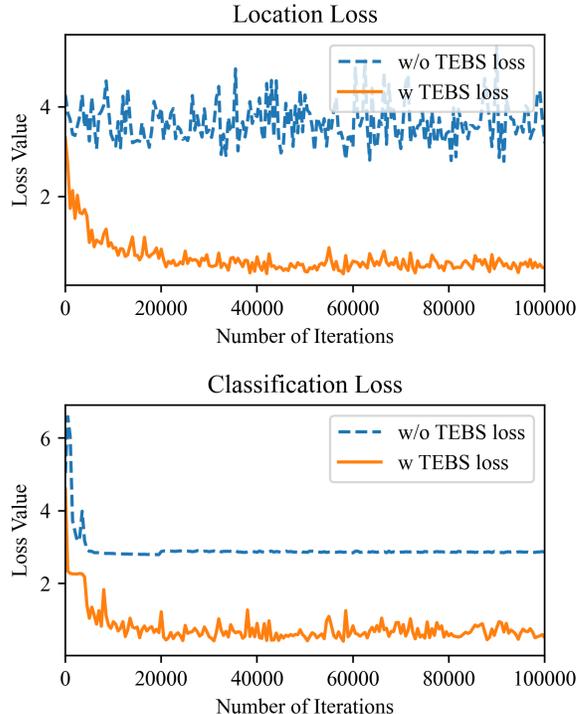


Figure 8. Comparison of the localization and classification loss curves during training without and with the TEBS loss.

resources are limited.

### 6.4. Impact of the TEBS Loss on the Detection Module

In this section, we further demonstrate the superiority of the TEBS loss, highlighting its effectiveness in improving training performance, as well as enhancing target features while suppressing background.

Figure 8 illustrates the comparison of the localization and classification loss curves during training without and with the TEBS loss. The upper plot shows the location loss, while the lower plot represents the classification loss. It is evident that with the TEBS loss (solid orange line), both the location and classification losses converge more smoothly and to significantly lower values compared to training without the TEBS loss (dashed blue line). Specifically, the TEBS loss effectively stabilizes the training process, reduces oscillations, and enhances convergence speed, demonstrating its clear advantage in improving both localization and classification performance.

Figure 9 shows the comparison of feature maps from different stages of the detection backbone without and with the TEBS loss. The TEBS loss imposes explicit supervision on the target and background masks during training, which significantly enhances the feature representation of UAV tar-

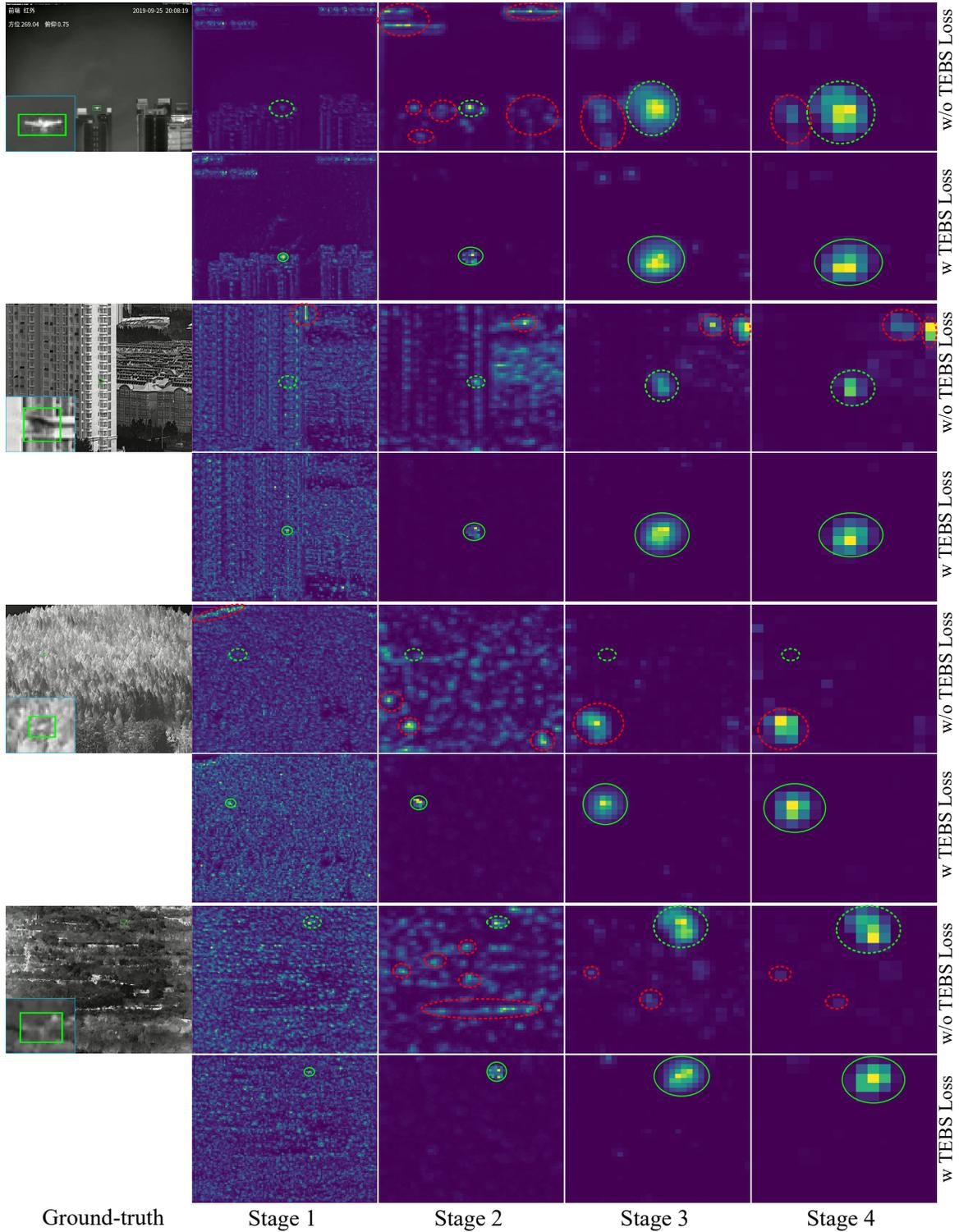


Figure 9. Comparison of feature maps from different stages of the detection backbone without and with the TEBS loss. The green dashed circles and solid circles represent the target enhancement effects without and with the TEBS loss, respectively. The red dashed circles represent the residual background.

Table 5. Impact of different union training methods on performance.

Row	Correction loss	Frozen NUC	Detection loss	BR loss	Metrics			
					PSNR	SSIM	P	R
1	✓	×	✓	×	33.024	0.9827	0.998	0.791
2	×	×	✓	×	17.940	0.8910	0.989	0.811
3	×	✓	✓	×	<b>37.722</b>	<b>0.9960</b>	0.998	0.804
4	×	×	✓	✓	31.961	0.9827	<b>0.999</b>	<b>0.822</b>

gets while suppressing irrelevant background information.

From Stage 1 to Stage 4, it can be observed that without the TEBS loss, the feature maps exhibit noticeable residual background interference (highlighted by red circles), especially in complex scenes such as urban areas or dense vegetation. This interference not only reduces the distinguishability of the UAV target but also increases the likelihood of false positives or missed detections. In contrast, with the TEBS loss, the feature maps progressively refine the representation of the UAV targets, as indicated by the high-intensity regions (green circles) becoming more focused and distinct. The background clutter is effectively suppressed, particularly in later stages of the backbone network.

This improvement highlights the contribution of the TEBS loss in enforcing spatial and semantic constraints on the backbone features, enabling the network to prioritize target regions while mitigating distractions from irrelevant features. As a result, the detection performance of the framework is substantially enhanced, as corroborated by the quantitative results and visualizations provided in the main text and supplementary material.

### 6.5. Impact of the BR Loss on the Union Framework

In this section, we focus on discussing the impact of various union training methods on the relationship between the NUC and target detection sub-tasks. Table 5 presents a comparison of four different union training strategies for combining the correction and detection modules, highlighting the impact of the BR loss in the UniCD framework.

(1) *Training with the correction and detection losses* (Row 1). In this approach, the correction and detection losses are directly added to jointly train the two modules. However, since the optimization objectives of the two losses differ, this method introduces conflicts that prevent either module from achieving optimal performance. As a result, both correction and detection metrics suffer from suboptimal outcomes.

(2) *Training with the detection loss only* (Row 2). This strategy uses only the detection loss to train both the correction and detection modules. Due to the lack of supervision on the correction module, its performance degrades significantly, leading to a sharp decline in PSNR and SSIM. Although the detection metrics improve slightly (e.g., Recall

= 0.811), the severe degradation in correction quality makes this method unsuitable for scenarios requiring high-quality visual outputs.

(3) *Training with the detection loss and the frozen correction module* (Row 3). In this method, the pre-trained correction module is frozen, and only the detection module is updated during training. While this approach achieves the best correction metrics (PSNR = 37.722, SSIM = 0.9960), the lack of interaction between the modules means that the correction module does not retain information beneficial to detection. Consequently, the detection metrics decline, with Recall dropping to 0.804.

(4) *Training UniCD with the detection loss and the BR loss* (Row 4). Our UniCD framework introduces the BR loss to enforce feature-level supervision from the detection backbone on the correction module. This self-supervised mechanism ensures that the correction module not only maintains visual quality (PSNR = 31.961, SSIM = 0.9827) but also aligns with the detection objectives. By jointly training both modules, the detection module can guide the correction module to produce feature maps that enhance UAV detection performance, achieving the highest detection metrics (P = 0.999, R = 0.822). This demonstrates the effectiveness of the BR loss in balancing the objectives of correction and detection while achieving superior overall performance.

### 6.6. Detection-Friendliness of the NUC Module under Union Training

In this section, we verify that the proposed union training method enables the NUC module to produce detection-friendly results, which are reflected in the feature maps of the detection backbone. As shown in Fig. 10, we present the feature maps across four stages of the detection backbone under separate and union training strategies: the separate training approach (first row and third row) and the union training approach (second row and fourth row).

In the separate training approach, the NUC module and the detection network are trained independently. As shown in the feature maps from the separate training approach, the extracted features exhibit scattered and noisy activations, with limited focus on the UAV target regions. This indicates a lack of alignment between the correction and detection tasks, leading to suboptimal feature representations for the detection network.

In contrast, the union training approach integrates the NUC module and detection network into a unified end-to-end framework. As observed in the second row and fourth row, the feature maps from union training display more concentrated and structured activations around the UAV target regions. This improvement highlights the detection-friendliness of the features produced by the NUC module under joint optimization, effectively suppressing irrelevant

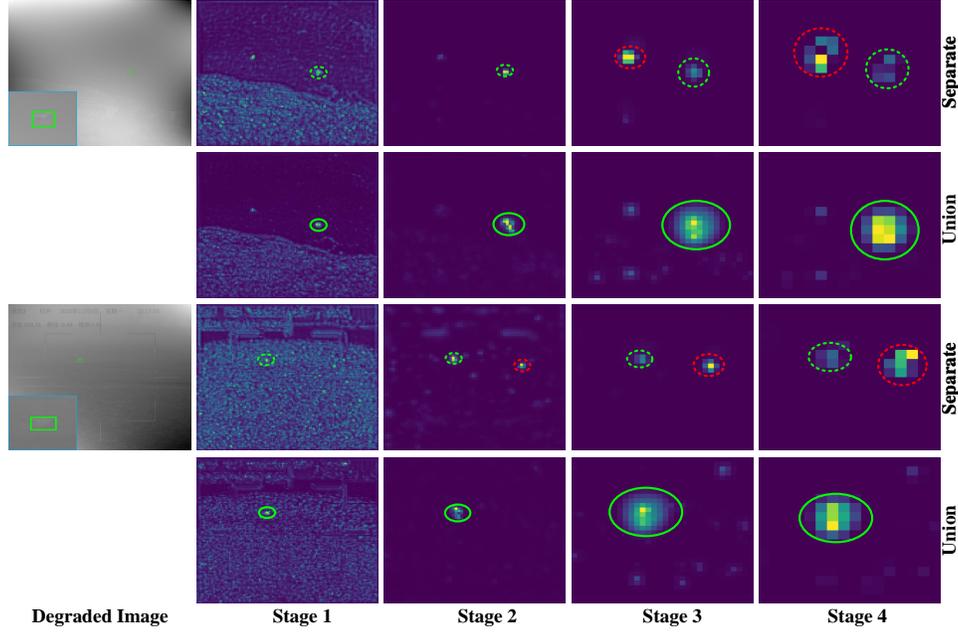


Figure 10. Visualization comparison of feature maps across four stages of the detection backbone under separate and union training. The green and red dashed circles represent the enhanced target region and the significant residual background, respectively.

background information and enhancing the features critical for accurate UAV detection.

This comparison demonstrates that union training not only improves the compatibility between the NUC and detection modules but also enables the NUC module to produce features that are more beneficial for downstream detection tasks, resulting in improved overall performance.

## 7. Calculation of Signal-to-Clutter Ratio Gain (SCRG)

The signal-to-clutter ratio (SCR) is utilized to measure the difficulty of target detection in a local region, can be calculated by:

$$\text{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b}, \quad (3)$$

where  $\mu_t$  and  $\mu_b$  represent the average pixel values of the target region and the surrounding neighboring region, respectively.  $\sigma_b$  is the standard deviation of the pixel values in the surrounding neighboring region of the target. As shown in Fig. 11, we assume the size of the small UAV target is  $a \times b$ , and then the size of its background region is  $(a + 2d) \times (b + 2d)$ , where  $d$  is the pixel width of neighboring area. We set  $d = 5$  pixels in our experiment.

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, which can be defined

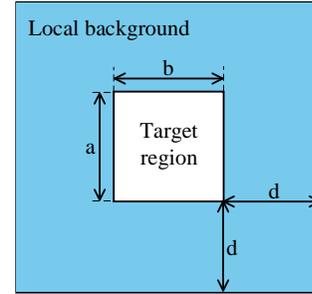


Figure 11. The bounding box of a small target and the adjacent background box.

as:

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}}. \quad (4)$$

## 8. Calculation of the Cosine Similarity Between Two Feature Maps

To compute the cosine similarity between two feature maps  $A$  and  $B$ , we assume they have the same shape,  $m \times n$  (i.e.,  $A$  and  $B$  each have  $m$  rows and  $n$  columns). The cosine similarity between feature maps can be interpreted as the cosine similarity between each pair of row vectors from  $A$  and  $B$ , resulting in an  $m \times m$  similarity matrix.

### 8.1. Definition of Cosine Similarity

The cosine similarity between two vectors  $\mathbf{a}$  and  $\mathbf{b}$  is defined as:

$$\text{Cos\_Sim}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}, \quad (5)$$

where:

- $\mathbf{a} \cdot \mathbf{b} = \sum_{k=1}^n a_k b_k$  is the dot product of  $\mathbf{a}$  and  $\mathbf{b}$ .
- $\|\mathbf{a}\|_2 = \sqrt{\sum_{k=1}^n a_k^2}$  and  $\|\mathbf{b}\|_2 = \sqrt{\sum_{k=1}^n b_k^2}$  are the  $L_2$  norms of  $\mathbf{a}$  and  $\mathbf{b}$ , respectively.

### 8.2. Computing the Dot Product Matrix

For two feature maps  $A$  and  $B$ , let  $A_i$  and  $B_j$  represent the  $i$ -th row of  $A$  and the  $j$ -th row of  $B$ , respectively. We can construct a dot product matrix  $D$ , where each element  $D_{ij}$  represents the dot product of row  $A_i$  with row  $B_j$ :

$$D_{ij} = A_i \cdot B_j = \sum_{k=1}^n A_{ik} B_{jk}. \quad (6)$$

Thus, the matrix  $D$  can be expressed as:

$$D = AB^T. \quad (7)$$

### 8.3. Calculating the $L_2$ Norms of Row Vectors

For each row  $A_i$  of feature map  $A$  and each row  $B_j$  of feature map  $B$ , we compute their  $L_2$  norms:

$$\|A_i\|_2 = \sqrt{\sum_{k=1}^n A_{ik}^2} \quad \text{and} \quad \|B_j\|_2 = \sqrt{\sum_{k=1}^n B_{jk}^2}. \quad (8)$$

The norms for all rows of  $A$  and  $B$  can be represented as column vectors:

$$\|A\|_2 = \begin{bmatrix} \|A_1\|_2 \\ \|A_2\|_2 \\ \vdots \\ \|A_m\|_2 \end{bmatrix}, \quad \|B\|_2 = \begin{bmatrix} \|B_1\|_2 \\ \|B_2\|_2 \\ \vdots \\ \|B_m\|_2 \end{bmatrix}. \quad (9)$$

### 8.4. Forming the Outer Product of Norms

Using the  $L_2$  norms  $\|A\|_2$  and  $\|B\|_2$ , we construct an  $m \times m$  matrix  $N$ , where each element  $N_{ij}$  represents the product of the norms:

$$N_{ij} = \|A_i\|_2 \|B_j\|_2. \quad (10)$$

Therefore, the matrix  $N$  can be represented as:

$$N = \|A\|_2 \|B\|_2^T. \quad (11)$$

### 8.5. Calculating the Cosine Similarity Matrix

Finally, we compute the cosine similarity matrix by dividing each element of the dot product matrix  $D$  by the corresponding element in the norm product matrix  $N$ :

$$\text{Cos\_Sim}_{ij} = \frac{D_{ij}}{N_{ij}} = \frac{A_i \cdot B_j}{\|A_i\|_2 \|B_j\|_2}. \quad (12)$$

In matrix form, the cosine similarity between feature maps  $A$  and  $B$  is given by:

$$\text{Cos\_Sim}(A, B) = \frac{AB^T}{\|A\|_2 \|B\|_2^T}. \quad (13)$$

### 8.6. Final Formula Summary

Thus, the cosine similarity matrix can be expressed as:

$$\text{Cos\_Sim}(A, B) = \frac{AB^T}{\sqrt{\sum_{k=1}^n A_{ik}^2} \cdot \sqrt{\sum_{k=1}^n B_{jk}^2}}. \quad (14)$$

This provides the cosine similarity for each pair of row vectors in  $A$  and  $B$ .

## References

- [1] 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. 3
- [2] 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. 6
- [3] 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. 1, 3
- [4] Glenn Jocher and Jing Qiu. Ultralytics YOLO11, 2024. 6
- [5] 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. 4
- [6] 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. 4
- [7] 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. 3
- [8] 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. 4

- [9] 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. [4](#)
- [10] 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. [3](#)