Blur-Robust Detection via Feature Restoration: An End-to-End Framework for Prior-Guided Infrared UAV Target Detection ——Supplementary Material——

Xiaolin Wang[1], Houzhang Fang[1*], Qingshan Li[1], Lu Wang[1], Yi Chang[2], Luxin Yan[2]

[1]School of Computer Science and Technology, Xidian University, China

[2]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China

wxl@stu.xidian.edu.cn,houzhangfang@xidian.edu.cn,qshli@mail.xidian.edu.cn,wanglu@xidian.edu.cn,{yichang,yanluxin}@hust.edu.cn

1 Overview

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

- **Details of the IRBlurUAV dataset.** In Section 2,we elaborate on the construction of the IRBlurUAV dataset, including the generation process of the synthetic blurred images (IRBlurUAV-syn), the selection criteria for real-world blurred images (IRBlurUAV-real), and a thorough statistical analysis covering target scales, background diversity, and spatial distribution.
- **Further quantitative experimental results.** In Section 3, we report more extensive comparisons between our approach and state-of-the-art (SOTA) methods. Additionally, we include deblurring quality evaluation (PSNR/SSIM) and generalization tests on the external HIT-UAV dataset to verify generalization across scenarios.
- **Further qualitative comparison results.** In Section 4, we present more visual comparisons on both synthetic and real blurred infrared unmanned aerial vehicle (UAV) datasets, highlighting the differences in detection confidence and accuracy across methods under motion blur.
- **Further ablation studies.** In Section 5, we provide in-depth analyses on the influence of core components, such as feature-domain deblurring (FDD) and frequency structure guided module (FSGM). We also explore how different module configurations affect performance, and analyze feature-level quality improvements at various backbone stages.
- **Further details on metric computation.** In Section 6, we detail the calculation of the Signal-to-Clutter Ratio (SCR), including the mathematical formulation, parameter settings, and region definitions used to assess local detection difficulty.

*Corresponding author.

2 Further Details of the IRBlurUAV Dataset

2.1 Generation Process of the Synthetic Dataset IRBlurUAV-syn

Selection of infrared clear UAV images. To ensure the broad applicability and diversity of the dataset, we select the largest publicly available infrared UAV dataset Anti-UAV410 (Huang et al. 2024) 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 COCO dataset format for ease of use and compatibility with existing tools.

Generation of blurred UAV images. We adopt the blur kernel generation process described by Sayed and Brostow (2021), where camera motion is simulated via a stochastic trajectory governed by an acceleration model. Unlike their approach, which explores three different levels of motion randomness by setting the parameter $P \in \{P_1, P_2, P_3\}$, we fix the motion randomness parameter to a very small value, $P = 0.00005$, to approximate linear motion and better match the motion characteristics found in real-world anti-drone imaging scenarios.

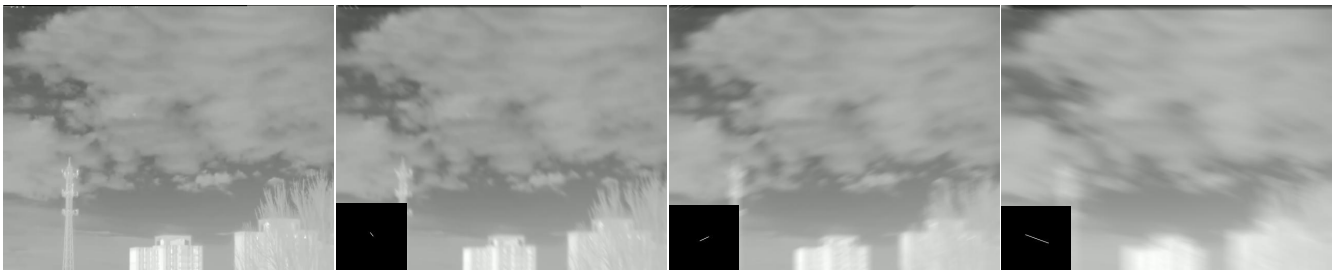
To simulate varying degrees of motion blur, we control the trajectory length via an exposure factor $E \in \left\{\frac{1}{10}, \frac{1}{5}, \frac{1}{2}\right\}$, which correspond to mild, moderate, and severe blur levels, respectively. The corresponding visualization is shown in Figure 1. The complete implementation details and the code used for kernel generation can be found in our open-source repository.

Formally, given a clear infrared UAV image $I(x, y)$, the blurred image $I_b(x, y)$ is generated as:

$$I_b(x, y) = (I * K)(x, y), \tag{1}$$

where $*$ denotes 2D convolution and $K(x, y)$ is the point spread function (PSF) derived from a simulated motion trajectory. The trajectory $\mathcal{T} = \{z_t\}_{t=1}^{T}$ is a complex-valued sequence with $z_t = x_t + iy_t$, generated via an acceleration-based model:

$$v_{t+1} = \frac{v_t + \Delta v_t}{\|v_t + \Delta v_t\|} \cdot \frac{L}{T}, \quad z_{t+1} = z_t + v_{t+1}, \tag{2}$$



Clear image E =0.1 E =0.2 E =0.5

Figure 1: Examples of different levels of motion blur. The bottom-left corner of each image shows the visualization of the corresponding blur kernel.

where v_t is the velocity, Δv_t is a perturbation incorporating Gaussian noise, inertial pull, and occasional abrupt changes, L denotes the total desired trajectory length, which governs the spatial extent of motion blur, T represents the number of discrete time steps (i.e., total trajectory iterations). The quotient $\frac{L}{T}$ ensures that the trajectory has a consistent and controlled step size.

The perturbation term Δv_t encapsulates random variations in velocity and is computed as:

$$\Delta v_t = P \cdot (\gamma_t - \alpha z_t), \tag{3}$$

where $\gamma_t \sim \mathcal{N}(0, I)$ is a Gaussian noise vector representing random camera shake, and αz_t is a centripetal force term that pulls the trajectory toward the origin to avoid unbounded drift. The scalar parameter P (set to $P = 0.00005$ in our implementation) controls the magnitude of motion randomness. A small value of P produces near-linear motion trajectories that closely match the smooth motion patterns typically observed in anti-drone infrared imaging scenarios.

The PSF $K(x, y)$ is computed by aggregating weighted contributions from the trajectory positions:

$$K(x, y) = \sum_{t=1}^{\lfloor E \cdot T \rfloor} w_t \cdot \delta(x - \mathrm{Re}(z_t), y - \mathrm{Im}(z_t)), \tag{4}$$

Here, $\mathrm{Re}(z_t)$ and $\mathrm{Im}(z_t)$ denote the real and imaginary parts of the complex-valued position z_t, corresponding to the x- and y-coordinates of the trajectory, respectively. This formulation allows the trajectory to be represented in a compact and continuous 2D form. The Dirac delta function $\delta(\cdot)$ is approximated via bilinear interpolation, and w_t represents the temporal contribution of step t to the kernel energy. And $E \in \left\{\frac{1}{10}, \frac{1}{5}, \frac{1}{2}\right\}$ is the exposure factor that controls the fraction of the trajectory used (thus simulating different blur levels), and w_t is a bilinear interpolation weight. The kernel K is normalized to satisfy $\sum_{x,y} K(x, y) = 1$, and applied to the image using FFT-based convolution for computational efficiency.

2.2 Selection of Real Blurred Infrared UAV Images.

Similarly, we manually select real blurred infrared UAV images from the Anti-UAV410 (Huang et al. 2024) dataset by identifying scenes with rapid camera motion. In total, 4,118 real blurred infrared UAV images were collected through this process.

2.3 Statistics and Analysis of the IRBlurUAV Dataset

Figure 2 illustrates several representative scenarios from the IRBlurUAV dataset. For clarity of comparison, the first row presents clean images from the synthetic subset, while the second row shows corresponding real-world blurred images. These scenes encompass two different lighting conditions (day and night), two seasons (autumn and winter), and a variety of backgrounds, including buildings (30%), mountains (20%), forests (5%), urban areas (30%), clouds (10%), and water surfaces (3%) (Huang et al. 2024).

Scale variation of infrared UAV targets. As shown in Figure 3, in the IRBlurUAV-syn dataset, tiny-scale UAV targets (Tiny, (0, 100)) account for approximately 26.4% (9,852 targets), mini-scale UAV targets (Mini, [100, 400)) account for approximately 44.0% (16,406 targets), small-scale UAV targets (Small, [400, 900)) account for approximately 8.3% (3,110 targets), and medium & normal-scale UAV targets (Medium & Normal, [900, ∞)) account for 1.7% (632 targets).

In contrast, in the IRBlurUAV-real dataset, tiny-scale UAV targets account for approximately 28.4% (1,427 targets), mini-scale UAV targets account for approximately 41.5% (2,088 targets), small-scale UAV targets account for approximately 9.2% (462 targets), and medium & normal-scale UAV targets account for 2.8% (141 targets).

The IRBlurUAV-real dataset effectively captures the complexities of real-world anti-UAV conditions. The high prevalence of tiny and mini-scale UAV targets underscores the inherent difficulty in detecting small objects, making the dataset highly representative for real-world detection and tracking tasks.

Comprehensiveness of Target Position Distribution. Figure 4 illustrates the spatial distribution of UAV targets in the IRBlurUAV dataset, including both the real-world (IRBlurUAV-real) and synthetic (IRBlurUAV-syn) subsets. As seen in the heatmaps, the real-world subset exhibits continuous motion trajectories with high-density regions along

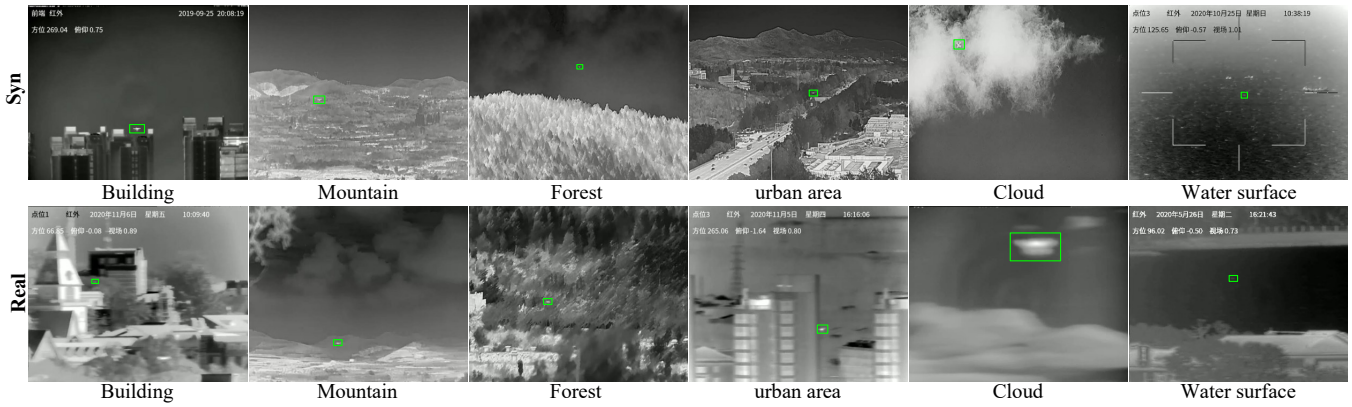


Figure 2: Six representative background types in the synthetic and real datasets. The first row shows examples from IRBlur-UAV-syn, while the second row presents corresponding examples from IRBlur-UAV-real.

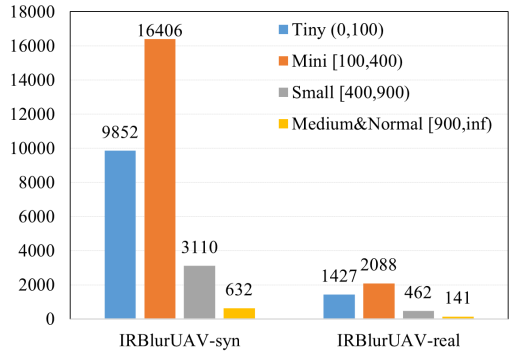
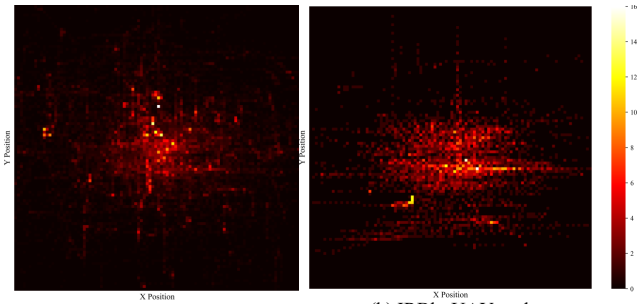


Figure 3: Targets area distribution in IRBlurUAV.



(a) IRBlurUAV-syn (b) IRBlurUAV-real

Figure 4: Position distribution of targets in IRBlurUAV.

specific paths, capturing realistic UAV motion patterns.

In contrast, the synthetic subset shows a more evenly distributed pattern across the image space, offering extensive spatial coverage. Together, these two subsets provide a complementary dataset for evaluating UAV detection under both real-world and controlled synthetic conditions.

Strategy	Module			Metrics			
	Deblurring	Detection	Pub'Year	AP_{50}	AR_{50}	AP	AR
Direct	/	YOLO11-N	2024	0.510	0.530	0.213	0.258
		YOLO11-L	2024	0.551	0.565	0.232	0.282
		RT-DETR	CVPR'24	0.716	<u>0.811</u>	<u>0.369</u>	<u>0.400</u>
		D-FINE	ICLR'25	<u>0.722</u>	0.795	0.347	0.382
		DEIM	CVPR'25	0.654	0.734	0.290	0.330
		MSHNet	CVPR'24	0.358	0.428	0.099	0.148
		PConv	AAAI'25	0.581	0.592	0.227	0.278
Separate	DeepRFT	RT-DETR	AAAI'23	0.673	0.749	0.284	0.329
		D-FINE		0.660	0.743	0.255	0.303
	MDT	RT-DETR	CVPR'25	0.195	0.297	0.062	0.082
		D-FINE		0.181	0.268	0.052	0.069
	EVSSM	RT-DETR	CVPR'25	0.636	0.713	0.244	0.285
		D-FINE		0.589	0.662	0.194	0.234
	MaIR	RT-DETR	CVPR'25	0.342	0.471	0.118	0.150
		D-FINE		0.292	0.403	0.084	0.114 1
	AdaRevD	RT-DETR	CVPR'24	0.064	0.111	0.019	0.024
		D-FINE		0.103	0.254	0.025	0.039
	DiffPIR	RT-DETR	CVPRW'25	0.092	0.184	0.029	0.041
		D-FINE		0.097	0.189	0.029	0.039
Joint	DREB-Net		TGRS'25	0.710	0.754	0.300	0.357
	Our JFD^3		-	**0.767**	**0.850**	**0.428**	**0.458**

Table 1: Performance comparison of more methods on IRBlurUAV-syn dataset. **Bold** and <u>underline</u> indicate the best and the second best results, respectively.

3 Further Quantitative Experiment Results

3.1 More Quantitative Comparison Results

As shown in Table 1, we compare our proposed method against a variety of baselines under three different strategies. In the direct strategy, detection models such as YOLO11 (N and L vesion)(Jocher and Qiu 2024), RT-DETR (ResNet18 version) (Zhao et al. 2024), D-FINE (N version) (Peng et al. 2025), and DEIM (D-FINE-N version)(Huang et al. 2025), MSHNet(Liu et al. 2024), and PConv (YOLOv8-N version)(Yang et al. 2025), are directly applied to blurred images. In the separate strategy, we combine six deblur-

Method	PSNR	SSIM
IRBlurUAV-syn	21.531	0.767
DeepRFT	26.795	0.851
MDT	22.523	0.851
EVSSM	24.170	0.820
MaIR	22.187	0.786
AdaRewD	21.940	0.591
DiffPIR	21.323	0.704

Table 2: Deblurring quality (PSNR / SSIM) of different deblurring methods on IRBlurUAV-syn.

ring methods: DeepRFT (Mao et al. 2023), MDT (Chen et al. 2025), EVSSM citekong2025EVSSM, MaIR (Li et al. 2025a), and two newly added methods, AdaRevD(Mao, Li, and Wang 2024) and DiffPIR(Zhu et al. 2023) (diffusion-based). Lastly, in the joint strategy, we include DREB-Net(Li et al. 2025b) and our proposed JFD³, which integrates deblurring and detection into a unified end-to-end framework. Furthermore, we report their deblurring performance on the IRBlurUAV-syn dataset using PSNR and SSIM metrics, as summarized in Table 2.

Interestingly, we observe a clear trend: methods achieving better deblurring performance in Table 2 tend to yield better detection performance after deblurring, as seen in Table 1. This experiment demonstrates that the Separate strategy heavily relies on the quality of the deblurring module. In some cases, suboptimal deblurring even leads to performance degradation, indicating that deblurring may introduce adverse artifacts. In contrast, our proposed Joint strategy achieves more stable and robust detection performance.

3.2 Further Experiments on Public Infrared Target Dataset

To evaluate the generalization ability of our method, we retrain and test several top-performing baselines from Table 1 on a new public infrared dataset, HIT-UAV (Suo et al. 2023), which contains five categories of ground targets: *Person*, *Car*, *Bicycle*, *OtherVehicle*, and *DontCare*. We apply the same blur generation strategy described in Section 1 to simulate motion blur for this dataset. The evaluation metrics are the average over all five categories: mAP_{50}, mAR_{50}, mAP, and mAR. The results are summarized in Table 3.

As shown in Table 3, our proposed method JFD^3 achieves the highest scores across all metrics, with mAP_{50} = 0.573, mAR_{50} = 0.833, mAP = 0.286, and mAR = 0.140. In comparison, other strong methods such as RT-DETR and D-FINE achieve lower performance, while combinations like DeepRFT + D-FINE and DREB-Net also fall short. These results clearly demonstrate that our joint framework not only excels in the original dataset but also generalizes effectively to new infrared UAV scenarios involving different object categories.

4 Further Results of Qualitative Comparison

As illustrated in Figure 5, we present additional experimental results for subjective visual comparison. The first to third rows display detection results on the IRBlurUAV-syn, while the fourth and fifth rows correspond to the IRBlurUAV-real. In each case, the third and fourth columns show images deblurred using DeepRFT, whereas the remaining columns use the original blurry inputs. A magnified view of the region near the ground truth (GT) target is shown in the bottom-right corner of each image, with green boxes denoting GT and red boxes denoting the detection results.

It can be observed that most methods exhibit reduced discriminative ability between target and background under motion blur, leading to false alarms, missed detections, or low confidence scores. In contrast, our method consistently avoids false or missed detections and maintains the highest confidence scores across both synthetic and real-world scenarios.

Method	$mAP_{50}\uparrow$	$mAR_{50}\uparrow$	$mAP\uparrow$	$mAR\uparrow$
RT-DETR	0.433	0.808	0.226	<u>0.129</u>
D-FINE	<u>0.512</u>	<u>0.830</u>	<u>0.242</u>	0.126
YOLO11-N	0.157	0.317	0.077	0.056
DeepRFT + RT-DETR	0.454	0.810	0.228	0.112
DeepRFT + D-FINE	0.466	0.823	0.215	0.112
DREB-Net	0.490	0.707	0.206	0.120
Our JFD^3	**0.573**	**0.833**	**0.286**	**0.140**

Table 3: Generalization results on the public HIT-UAV (Suo et al. 2023) dataset with simulated motion blur.

5 Further Ablation Studies

5.1 Impact of Different Structural Guidance Modules

Our proposed FSGM can be interpreted as a cross-attention mechanism guided by structural information. To systematically evaluate the impact of different structural guidance strategies and parameter settings, we present a comparative study in Table 4.

We begin by employing ContMix (Lou and Yu 2025)as our baseline, as its architecture closely resembles our own framework. The first row reports results where no structural prior is involved; only the Stem component is used to generate the Query, Key, and Value. This setup lacks structure-aware modeling, resulting in slightly inferior detection performance.

In the second row, we replace our FSGM with the original ContMix. The comparison clearly reveals that incorporating structural prior knowledge significantly enhances detection performance, particularly in the presence of image blur.

In the third row, we remove the frequency feature refine block (FFRB) from FSGM to assess its contribution. While this version achieves lower parameter count and computational cost, it comes at the expense of a noticeable drop in detection accuracy.

The final row presents our full FSGM model. It achieves the best overall performance across all evaluation metrics, while maintaining a competitive computational footprint. Notably, FSGM demonstrates superior robustness and generalization capability for detecting blurred targets, validating

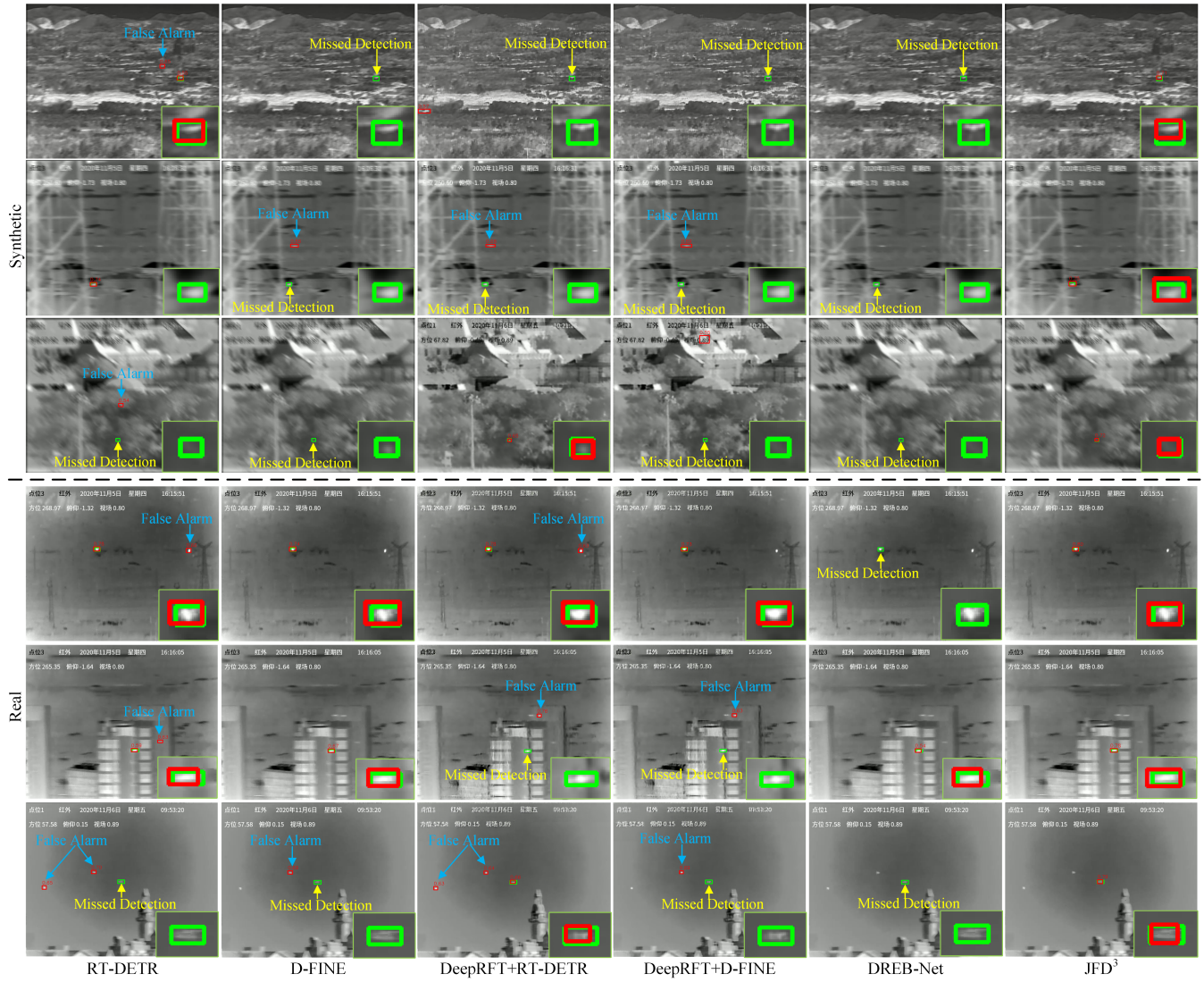


Figure 5: Qualitative detection results on IRBlurUAV-syn (top three rows) and IRBlurUAV-real (bottom two rows). Columns 3–4 show DeepRFT deblurred results; others use original blurry images. Red boxes denote predictions, green boxes indicate ground truth. Insets show zoomed regions around targets. Our method achieves accurate detection without false alarms or missed targets, outperforming others under motion blur.

the effectiveness of both its structure-guided mechanism and the frequency-enhanced refinement module.

5.2 Impact of FDD on Different Layers of the Backbone

As shown in Table 5, we analyze the effect of incorporating feature-domain deblurrin(FDD) at different stages of the backbone by evaluating PSNR and SSIM metrics on intermediate feature maps, both with and without FDD. The results clearly show that applying FDD at the shallow layers (Stem and Stage1) leads to a substantial improvement in both PSNR and SSIM. For instance, in the Stem layer, PSNR increases from 14.66 to 15.64 and SSIM rises from 0.5977 to 0.6201. Similarly, Stage1 sees PSNR improve from 15.29 to 18.18, and SSIM from 0.5189 to 0.6464. These improvements demonstrate that FDD enhances the representational

Method	$AP_{50}\uparrow$	$AR_{50}\uparrow$	$AP\uparrow$	$AR\uparrow$	Param (K)	FLOPs (M)
ContMix-w/o Strucute Prior	0.759	0.835	0.416	0.449	14.16	621.78
ContMix	<u>0.763</u>	<u>0.846</u>	<u>0.419</u>	<u>0.451</u>	14.16	621.78
FSGM w/o FFRB	0.737	0.823	0.349	0.384	**2.40**	**150.13**
FSGM	**0.767**	**0.850**	**0.428**	**0.458**	<u>10.63</u>	<u>440.66</u>

Table 4: Comparison of structure-guided attention methods. The proposed frequency-enhanced prior leads in all metrics.

Module	w/o FDD		w/ FDD	
	PSNR	SSIM	PSNR	SSIM
Stem	14.66	0.5977	**15.64** ↑	**0.6201** ↑
Stage1	15.29	0.5189	**18.18** ↑	**0.6464** ↑
Stage2	**18.05**	**0.6470**	16.42 ↓	0.6233 ↓

Table 5: Effect of feature restoration at different backbone positions.

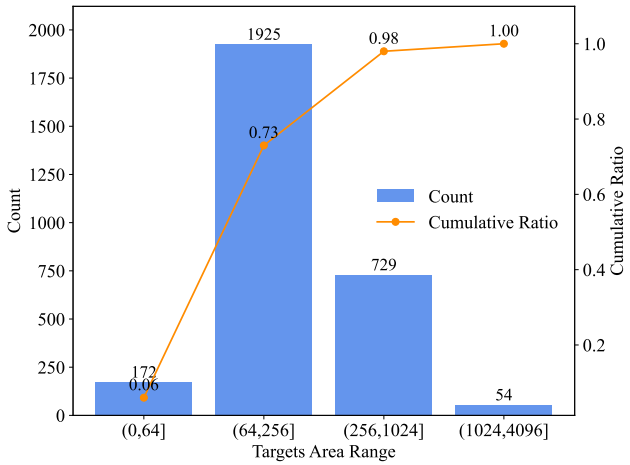


Figure 6: Targets area distributsaion in IRBlurUAV-syn test set.

quality of early-stage features, enabling more discriminative and robust learning in the shallow backbone.

In contrast, performance at Stage2 slightly decreases when FDD is applied. This can be attributed to the nature of deeper-stage feature maps. At Stage2, with a downsampling factor of 8, the spatial resolution is significantly reduced. Given that 73%of targets in the IRBlurUAV-syn test set have a spatial size smaller than 16×16, and 98% are below 32×32, the corresponding target regions in Stage2 feature maps are often less than 2×2 or 4×4 pixels, as shown in Figure 6. As a result, the target occupies only a minimal portion of the global feature map, and the majority of pixels correspond to background. Consequently, global metrics like PSNR and SSIM—computed over all pixels—are dominated by background content and fail to accurately reflect improvements in target-related feature quality. Therefore, the slight drop in PSNR and SSIM at Stage2 does not imply degraded performance in target representation, but rather highlights the limitations of these metrics at deeper, lower-resolution stages.

5.3 Visualization of Feature Maps with FDD and FSGM

To investigate the impact of the proposed components on intermediate feature representations, we visualize the feature maps from Stage1 to Stage4 of the backbone network, as shown in Figure 7. The input is a blurred infrared UAV image, and the green bounding box indicates the ground-truth

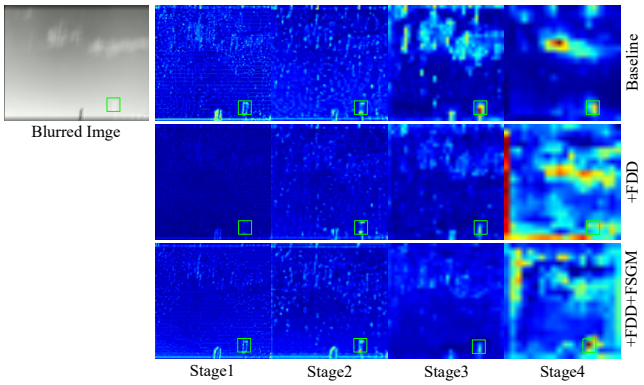


Figure 7: Targets area distributsaion in IRBlurUAV-syn test set.

target region.

The first row corresponds to the baseline model without any enhancement modules. It can be observed that the feature maps, especially in the deeper stages (Stage3 and Stage4), contain significant background clutter. The network's attention is dispersed, often focusing on irrelevant background regions instead of the target.

The second row illustrates the feature maps after integrating the feature-domain deblurring (FDD) module. With reduced background interference, the feature responses appear cleaner. However, due to the lack of structural guidance, certain target details are suppressed during deblurring, leading to diminished target focus in deeper layers such as Stage4.

The third row shows the results of our final proposed method, which incorporates both FDD and the frequency structure guided module (FSGM). This configuration significantly enhances the target structure and boundary clarity, enabling the network to maintain accurate focus on the ground-truth target across all stages. The background clutter is effectively suppressed, while the target signal becomes more distinguishable, especially in deeper layers.

These visualizations confirm that our proposed modules work synergistically: FDD reduces noise and blur, while FSGM provides explicit structural cues to strengthen target awareness under challenging blurred conditions.

6 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:

$$\mathrm{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b}, \tag{5}$$

where μ_t and μ_b represent the average pixel values of the target region and the surrounding neighboring region, respectively. σ_b is the standard deviation of the pixel values in the surrounding neighboring region of the target. As shown in Figure 8, we assume the size of the small UAV target is $a \times b$, and then the size of its background region is

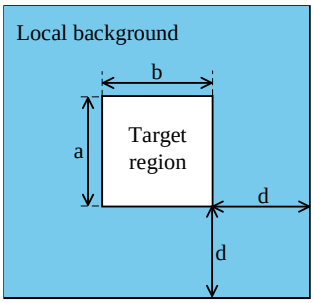


Figure 8: The bounding box of a small target and the adjacent background box.

$(a+2d)\times(b+2d)$, where d is the pixel width of neighboring area. We set $d = 40$ pixels in our experiment.

References

Chen, D.; Zhou, S.; Pan, J.; Shi, J.; Qu, L.; and Yang, J. 2025. A Polarization-Aided Transformer for Image Deblurring via Motion Vector Decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 28061–28070.

Huang, B.; Li, J.; Chen, J.; Wang, G.; Zhao, J.; and Xu, T. 2024. 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.

Huang, S.; Lu, Z.; Cun, X.; Yu, Y.; Zhou, X.; and Shen, X. 2025. DEIM: DETR with Improved Matching for Fast Convergence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15162–15171.

Jocher, G.; and Qiu, J. 2024. Ultralytics YOLO11.

Li, B.; Zhao, H.; Wang, W.; Hu, P.; Gou, Y.; and Peng, X. 2025a. MaIR: A Locality- and Continuity-Preserving Mamba for Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7491–7501.

Li, Q.; Zhang, Y.; Fang, L.; Kang, Y.; Li, S.; and Xiang Zhu, X. 2025b. DREB-Net: Dual-Stream Restoration Embedding Blur-Feature Fusion Network for High-Mobility UAV Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–18.

Liu, Q.; Liu, R.; Zheng, B.; Wang, H.; and Fu, Y. 2024. Infrared Small Target Detection with Scale and Location Sensitivity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17490–17499.

Lou, M.; and Yu, Y. 2025. OverLoCK: An Overview-first-Look-Closely-next ConvNet with Context-Mixing Dynamic Kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 128–138.

Mao, X.; Li, Q.; and Wang, Y. 2024. AdaRevD: Adaptive Patch Exiting Reversible Decoder Pushes the Limit of Image Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25681–25690.

Mao, X.; Liu, Y.; Liu, F.; Li, Q.; Shen, W.; and Wang, Y. 2023. Intriguing Findings of Frequency Selection for Image Deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 1905–1913.

Peng, Y.; Li, H.; Wu, P.; Zhang, Y.; Sun, X.; and Wu, F. 2025. D-FINE: Redefine Regression Task of DETRs as Fine-grained Distribution Refinement. In *Proceedings of the International Conference on Representation Learning (ICLR)*, volume 2025, 44015–44031.

Sayed, M.; and Brostow, G. 2021. Improved Handling of Motion Blur in Online Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1706–1716.

Suo, J.; Wang, T.; Zhang, X.; Chen, H.; Zhou, W.; and Shi, W. 2023. HIT-UAV: A high-altitude infrared thermal dataset for Unmanned Aerial Vehicle-based object detection. *Scientific Data*, 10(1): 227.

Yang, J.; Liu, S.; Wu, J.; Su, X.; Hai, N.; and Huang, X. 2025. Pinwheel-shaped Convolution and Scale-based Dynamic Loss for Infrared Small Target Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9202–9210.

Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; and Chen, J. 2024. DETRs Beat YOLOs on Real-time Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16965–16974.

Zhu, Y.; Zhang, K.; Liang, J.; Cao, J.; Wen, B.; Timofte, R.; and Van Gool, L. 2023. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1219–1229.