

Article

FPGA-Based Real-Time Deblurring and Enhancement for UAV-Captured Infrared Imagery

Jianghua Cheng, Lehao Pan, Tong Liu *, Bang Cheng and Yahui Cai

College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China; jianghua_cheng@nudt.edu.cn (J.C.); lehaopan22@nudt.edu.cn (L.P.); chengbang@nudt.edu.cn (B.C.); caiyh95@nudt.edu.cn (Y.C.)

* Correspondence: liutong1129@nudt.edu.cn; Tel.: +86-0731-87003158

Highlights

Main Findings:

- A novel deep learning network is proposed for simultaneous blind deblurring and enhancement of UAV infrared images, integrating feature extraction, fusion, and simulated diffusion modules, along with a region-specific pixel loss and progressive training strategy.
- The method achieves significant performance improvements, including a 10.7% increase in PSNR, 25.6% reduction in edge inference time, and 18.4% decrease in parameter count.

Implications:

- Offers an efficient and lightweight solution for real-time infrared image restoration on mobile platforms such as UAVs.
- Enhances image quality and inference speed, providing a reliable foundation for downstream high-level vision tasks.

Abstract

In response to the inherent limitations of uncooled infrared imaging devices and the image degradation caused by UAV (Unmanned Aerial Vehicle) platform motion, resulting in low contrast and blurred details, a novel single-image blind deblurring and enhancement network is proposed for UAV infrared imagery. This network achieves global blind deblurring and local feature enhancement, laying a foundation for subsequent high-level vision tasks. The proposed architecture comprises three key modules: feature extraction, feature fusion, and simulated diffusion. Furthermore, a region-specific pixel loss is introduced to strengthen local feature perception, while a progressive training strategy is adopted to optimize model performance. Experimental results on public infrared datasets demonstrate that the presented method outperforms state-of-the-art methods HCTIRdeblur, reducing parameter count by 18.4%, improving PSNR by 10.7%, and decreasing edge inference time by 25.6%. This work addresses critical challenges in UAV infrared image processing and offers a promising solution for real-world applications.

Keywords: UAV; infrared image blind deblurring; image enhancement; FPGA



Academic Editor: Li Li

Received: 26 August 2025

Revised: 4 October 2025

Accepted: 13 October 2025

Published: 15 October 2025

Citation: Cheng, J.; Pan, L.; Liu, T.; Cheng, B.; Cai, Y. FPGA-Based Real-Time Deblurring and Enhancement for UAV-Captured Infrared Imagery. *Remote Sens.* **2025**, *17*, 3446. <https://doi.org/10.3390/rs17203446>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Infrared (IR) imaging has become an indispensable technology in various applications, including surveillance, remote sensing, and military operations. However, the quality of IR images acquired from Unmanned Aerial Vehicle (UAV) platforms is often compromised due to inherent limitations of uncooled infrared sensors and platform motion [1]. These factors contribute to low contrast, blurred details, and overall degradation of image quality, which significantly impedes subsequent high-level vision tasks, like object detection, as illustrated in Figure 1. This figure showcases representative examples of low-quality infrared images, highlighting the challenges faced in UAV-Captured infrared imaging. Recent advancements [2] in deep learning have shown promising results in image restoration and enhancement. However, the unique characteristics of UAV IR imagery present several challenges that are not adequately addressed by existing methods [3]. First, the complex motion patterns of UAV platforms induce spatially varying blur, which is difficult to model and remove using traditional deblurring techniques. Second, the low signal-to-noise ratio (SNR) of uncooled IR sensors exacerbates the problem of feature preservation during the enhancement process. Third, the lack of large-scale, high-quality IR image datasets hinders the development of robust learning-based solutions.



Figure 1. Poor visual quality and inadequate detection results of low-quality infrared images acquired from UAV platforms. (a) Acquired low-quality images; (b) Object detection results, where red arrows indicate false detections and blue arrows indicate missed detections.

To address these challenges, we propose a novel single-image blind deblurring and enhancement network specifically designed for UAV IR imagery. Our approach integrates global blind deblurring with local feature enhancement in a unified framework, leveraging the strengths of deep learning to overcome the limitations of traditional methods. The proposed network architecture incorporates three key modules: feature extraction, feature fusion, and simulated diffusion, each tailored to address specific aspects of the IR image restoration problem.

Furthermore, we introduce a region-specific pixel loss to enhance local feature perception, addressing the critical issue of preserving fine details in IR images. To optimize the learning process, we employ a progressive training strategy that gradually increases the complexity of the tasks, allowing the model to build a robust representation of IR image characteristics.

The main contributions of this work can be summarized as follows:

1. A novel network architecture is proposed for single-image blind deblurring and enhancement of UAV IR imagery, addressing the unique challenges posed by uncooled sensors and platform motion.

2. A region-specific pixel loss is proposed together with a progressive training strategy, aiming to improve local feature preservation and enhance the model's overall performance.
3. Extensive evaluations on public IR datasets confirm that the method not only achieves superior performance to existing approaches but also does so with fewer parameters and faster inference speeds.

The remainder of this paper is organized as follows: Section 2 reviews related work in IR image enhancement and deep learning-based image restoration. Section 3 describes the proposed network architecture and training methodology in detail. Section 4 presents experimental results and comparisons with existing methods. Finally, Section 5 concludes the paper and discusses future research directions.

2. Related Work

The enhancement and restoration of IR images, particularly those captured from UAV or airborne platforms, have been extensively studied in recent years. This section reviews the most relevant works in the fields of IR image deblurring, enhancement, and deep learning-based image restoration, highlighting the development trends and the innovative contributions of our work. Additionally, to highlight the significance of our work, a systematic review of UAV Image Rapid Processing Technologies and Applications was also included.

2.1. Infrared Image Enhancement

IR image enhancement has been a subject of intense research due to its crucial role in various applications, including surveillance, medical imaging, and autonomous systems. Traditional methods, such as histogram equalization and its variants [4], have been widely used for contrast enhancement. However, these methods often struggle with preserving fine details and handling complex scenes.

Recent years have witnessed a shift towards more sophisticated enhancement techniques. Ma et al. [5] proposed a novel method that synergizes the advantages of Multi-Scale Retinex (MSR) and Adaptive Gamma Correction with Weighting Distribution (AGCWD). Their approach first preserves the merits of both techniques in a detailed image using illumination-based weighting. The final image, constructed by combining the detailed and original images, effectively maintains contrast in high-luminance regions while enhancing details in low-luminance areas, resulting in a more balanced enhancement across varying illumination conditions.

Ren et al. [6] proposed a novel Region Super-Resolution Generative Adversarial Network (RSRGAN) to address the challenges in infrared small target detection. Their approach incorporates a Region Context Network (RCN) as the backbone for efficient region proposal, and a Generative Adversarial Network (GAN) structure for distribution transformation and super-resolution enhancement. This method effectively converts blurry infrared small targets to clearer ones with similar distributions to the training set, followed by resolution enhancement. The discriminator aids in generating higher-quality super-resolution images. Their approach demonstrated superior performance in detecting small infrared targets compared to state-of-the-art methods, particularly when tested on a challenging custom dataset of small infrared drone targets.

More recently, attention mechanisms have been incorporated into enhancement networks to focus on salient regions. Fan et al. [7] introduced a Lightweight Attention-guided ConvNeXt Network (LACN) for low-light image enhancement. Their method incorporates a novel Attention ConvNeXt Module (ACM) and a Selective Kernel Attention Module (SKAM) into a lightweight network structure. This approach effectively improves im-

age brightness and contrast while suppressing noise amplification. The network's multi-attention mechanism and feature fusion strategy enable adaptive adjustment to various lighting conditions, demonstrating superior performance in enhancing low-light images compared to state-of-the-art methods.

2.2. Deep Learning-Based Image Restoration

Deep learning has revolutionized image restoration tasks, including denoising, deblurring, and super-resolution. Convolutional Neural Networks (CNNs) have been at the forefront of this revolution, with architectures like U-Net [8] setting benchmarks in various restoration tasks.

In the context of image deblurring, Kupyn et al. [9] introduced DeblurGAN-v2, a conditional GAN that achieves state-of-the-art performance in motion deblurring. Their approach incorporates a feature pyramid network and a relativistic discriminator, enabling efficient and high-quality blur removal.

For image super-resolution, the field has seen significant advancements. Liu et al. [10] proposed two enhancements to the ESRGAN algorithm for super-resolution image synthesis. They incorporated channel attention mechanisms (SENet and ECA-Net) into the generator to model cross-channel feature dependencies, and integrated LPIPS as an image quality assessment metric in the discriminator. Their approach demonstrated improved image naturalness and perceptual quality across multiple benchmark datasets, as evaluated by NIQE and LPIPS metrics.

Transformer-based architectures have recently gained traction in image restoration tasks. Liang et al. [11] proposed SwinIR, which effectively incorporates Swin Transformer blocks into a three-part architecture for tasks like super-resolution and denoising, demonstrating the potential of self-attention for global dependency modeling. Building upon this, recent advancements have explored more specialized and efficient designs. The LucidFlux framework adapts a large-scale diffusion Transformer (DiT) for universal image restoration without relying on text captions. It introduces a dual-branch conditioner to anchor geometry from the degraded input and suppress artifacts, achieving robust performance across various degradation types through a timestep-aware modulation schedule [12]. Another significant trend is the hybridization of CNNs and Transformers to leverage their complementary strengths. For instance, the MDDA-former architecture strategically embeds multi-dimensional dynamic attention blocks (inspired by CNNs) for local feature extraction and efficient Transformer blocks for global context modeling in a U-Net framework. This design has proven effective for complex degradations like rain and haze, achieving a balance between performance and computational efficiency [13]. Furthermore, the MSCSCformer introduces a multi-scale convolutional sparse coding-based Transformer for pansharpening, which employs both spatial and spectral self-attention mechanisms to capture long-range dependencies and inter-band correlations simultaneously [14].

2.3. Infrared Image Deblurring

Deblurring IR images presents unique challenges due to the low signal-to-noise ratio and the complex blur patterns often encountered in practical scenarios. Traditional deblurring methods, such as Wiener filtering and Richardson-Lucy deconvolution [15], often struggle with the spatially varying blur common in IR imagery.

Deep learning approaches have shown promising results in addressing these challenges. Zhao et al. [16] proposed a GAN-based method incorporating channel prior discrimination for infrared image deblurring. Their approach uniquely combines traditional and learning-based blind deblurring techniques, addressing both uniform and non-uniform blur. The method considers blur caused by object motion, camera movement, and scene

depth variations, demonstrating effectiveness on various datasets for single infrared image deblurring. Wu et al. [17] introduced a multi-scale network for single image deblurring, addressing the issues of blurring scale and context information utilization. Their approach incorporates an Ensemble Learning Model with weak learners and a Spatial Feature Enhancement Module to preserve spatial details. The proposed method demonstrated superior performance on benchmark datasets compared to state-of-the-art deblurring techniques, including NAFNet and Restormer.

2.4. UAV Image Rapid Processing Technologies and Applications

The rapid advancement of UAV in remote sensing has driven increasing demand for real-time, high-efficiency image processing pipelines, especially in mission-critical applications such as precision agriculture, urban monitoring, and disaster response. Recent efforts have focused on optimizing the entire photogrammetric workflow—from feature matching to pose reconstruction—under challenging conditions including low image overlap, repetitive textures, and large-format sensor outputs.

Xiao et al. [18] proposed a novel real-time matching and pose reconstruction framework specifically tailored for low-overlap agricultural UAV imagery. By introducing an adaptive map initialization strategy, optimized feature extraction with global texture awareness, and a multi-model adaptive tracking mechanism, their method achieves robust visual localization at approximately 3.4 frames per second, significantly outperforming traditional SfM and SLAM systems in both speed and reliability. This work underscores the importance of algorithmic resilience in real-time UAV photogrammetry, particularly when dealing with weakly textured or repetitive scenes—a common challenge also encountered in infrared imaging.

Further expanding on geometric and semantic integration, Li et al. [19] introduced a spatio-temporal data fusion framework for digital twin cities, leveraging cross-domain data integration through a unified geographic entity model. Their system enables real-time association of heterogeneous urban data via a knowledge graph powered by deep learning-based address matching, achieving over 99% F1-score in entity alignment. This highlights a growing trend toward intelligent, context-aware processing architectures that go beyond pixel-level enhancement to incorporate semantic understanding, a direction highly relevant for advanced UAV-based surveillance and decision support systems.

In another line of work, Xiao et al. [20] addressed the challenge of affine-invariant feature matching in oblique urban UAV imagery by proposing the Nicer Affine Invariant Feature (NAIF) algorithm. By exploiting rough exterior orientation (EO) data to guide affine rectification before SIFT matching, NAIF drastically reduces computational overhead while maintaining high correspondence accuracy. This sensor-guided fusion of geometric priors with traditional computer vision techniques exemplifies an efficient paradigm for accelerating image processing pipelines—an approach highly compatible with FPGA-based hardware acceleration, where predictable dataflow and early-stage regularization can maximize throughput and minimize latency.

Collectively, these studies reflect a shift from purely software-centric optimization toward tightly coupled algorithm-hardware co-design strategies for UAV image processing. While much of the existing literature emphasizes visible-spectrum imagery and geometric reconstruction, there remains a critical gap in real-time enhancement of infrared data—particularly under motion blur and low signal-to-noise conditions. To this end, FPGA-based implementations offer a promising avenue by enabling low-latency, power-efficient execution of deblurring and enhancement kernels, thereby complementing the algorithmic advances in real-time UAV vision systems.

2.5. Summary

The deployment of image restoration algorithms on UAV platforms imposes a unique set of constraints that are often at odds with the high computational demands of modern deep learning models. As reviewed, existing methods have made significant strides in enhancing IR image quality [5–7], deblurring [16,17], and even leveraging transformative architectures like Transformers [11–14]. Furthermore, the push for real-time processing in UAV photogrammetry is evident [18,20]. However, a critical gap remains in providing a computationally efficient, end-to-end solution that delivers simultaneous blind deblurring and feature enhancement specifically for power- and latency-sensitive UAV infrared vision systems.

The prevailing trend towards larger models and more complex architectures [12,21–23] often leads to high parameter counts and inference latency, rendering them unsuitable for real-time processing on UAV-borne edge computing devices. While some lightweight networks exist [24], they frequently lack the specialized design to handle the severe, spatially-varying blur and low contrast inherent in UAV-captured IR imagery.

Therefore, our work directly addresses this gap by introducing a novel network that is co-designed with the UAV's operational constraints from the ground up.

In summary, the principal contribution of this work is not merely another incremental improvement in image quality, but the presentation of a holistically designed, application-aware solution that effectively balances restoration performance with the stringent real-world requirements of UAV platforms, as validated by our state-of-the-art results in accuracy, speed, and power efficiency on edge hardware.

3. Method

This section is divided into three parts, which introduce the overall structure and principles of the model, the construction method of the required infrared dataset, and the design of loss functions and training strategies, respectively.

3.1. Network Architecture

UAV infrared imaging faces unique challenges, including image blur and degradation caused by platform vibration, atmospheric turbulence, and rapid target movement. Additionally, the inherent low contrast and lack of detail in infrared images further exacerbate image quality deterioration [25]. To address these challenges, this paper proposes a single-image blind deblurring and enhancement network for UAV platforms, designed to simultaneously achieve blind deblurring and target feature enhancement of infrared images. The network architecture is illustrated in Figure 2.

Considering the limited computational resources and real-time processing requirements in UAV environments, the network adopts an efficient, lightweight design with an overall encoder–decoder structure. This architecture effectively captures multi-scale image features while maintaining computational efficiency. The network comprises three key components: a feature extraction module, a feature fusion module, and a diffusion module.

The feature extraction module consists of three cascaded residual blocks, effectively mitigating the vanishing gradient problem in deep networks. Unlike mainstream approaches that reduce image spatial dimensions, this network avoids using max pooling to better preserve image details and edge information, facilitating the enhancement of infrared target features. This design choice is based on the critical role of target edge information in infrared images for object detection and recognition.

The feature fusion module is composed of CUR modules, employing nearest neighbor interpolation for upsampling. Compared to other upsampling methods such as bilinear interpolation or transposed convolution, nearest neighbor interpolation is computation-

ally less expensive and better preserves edge information in lower resolution infrared images [26]. Concurrently, feature outputs from the extraction module are concatenated, fusing multi-level image features without significantly increasing computational complexity. This multi-scale feature fusion strategy has proven effective in various computer vision tasks [27].

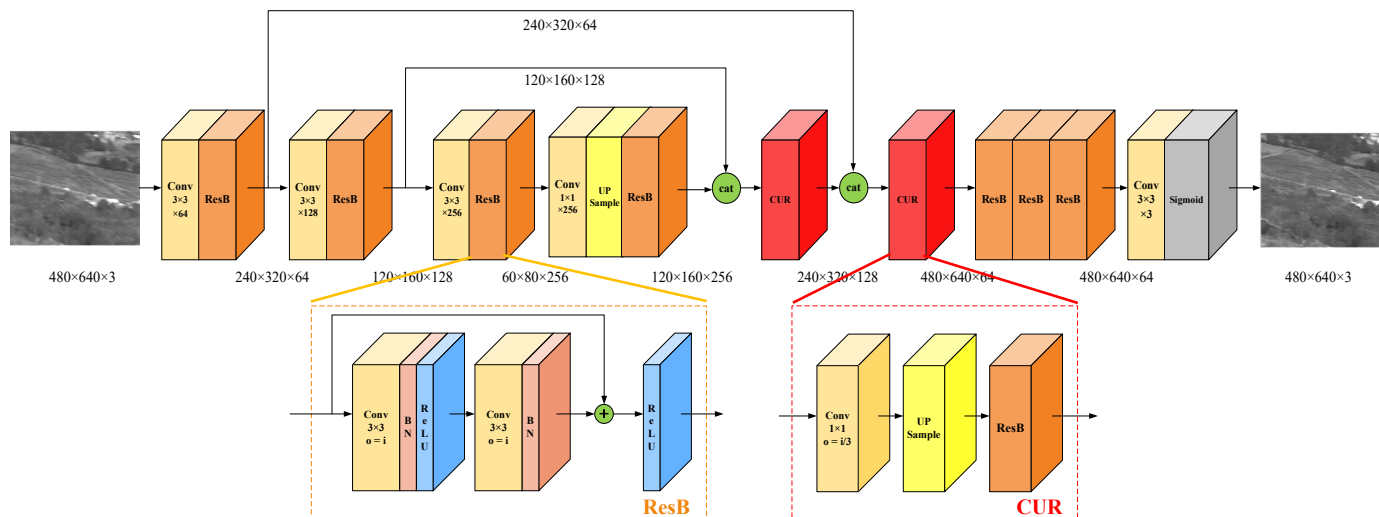


Figure 2. The architecture of the Single-Image Infrared Blind Deblurring and Enhancement Network for UAV Platforms is presented.

Notably, the diffusion module employs three cascaded ResB blocks to simulate the progressive refinement characteristic of diffusion models, rather than replicating their full multi-step noising-denoising mechanism. The choice of three ResB blocks represents a design trade-off, offering sufficient capacity for progressive refinement while maintaining low computational overhead suitable for UAV deployment, as validated by our edge inference results in Section 4.5. Inspired by the iterative enhancement process in diffusion-based methods [28,29], each ResB stage progressively refines feature representations—effectively simulating a coarse-to-fine recovery trajectory analogous to multi-stage denoising. While classical diffusion models explicitly model noise transitions across hundreds of steps, our design captures a condensed, three-stage refinement process: each ResB incrementally restores structural details and suppresses artifacts. Additionally, the skip connections in ResB help maintain information flow throughout the generation process, preserving gradient flow and information content to prevent loss—a mechanism widely proven effective in deep learning. This progressive learning strategy enables the model to handle complex degradation patterns efficiently, aligning with the conceptual framework of diffusion without incurring excessive computational overhead.

3.2. Dataset Construction

Deep learning models driven by data rely on high-quality image deblurring datasets to learn the correspondence between blurred and sharp images for image restoration and enhancement. While there are various datasets available for visible light image blind deblurring, there is a lack of publicly available datasets specifically designed for infrared image blind deblurring and feature enhancement in air-to-ground scenarios.

To address this issue, we utilized the infrared track dataset from the 2021 Aerospace Cup competition organized by the National University of Defense Technology as the foundation for our custom dataset construction. This dataset comprises 21,750 infrared images captured under diverse meteorological conditions, featuring weak and small targets.

The images have a resolution of 640×480 pixels, with each image containing up to six small vehicles as targets. Figure 3 presents representative examples from this dataset.



Figure 3. Infrared dataset from the 2021 Aerospace Cup competition organized by the National University of Defense Technology.

We introduced global mixed blur noise to the original images and intensified the inherent noise in target regions of infrared images to generate degraded images. These degraded images were then used as input, with the original images serving as ground truth (GT) for network training. By comparing the differences between the degraded input and the original image, the network learned to perform image restoration and target feature enhancement. In practical applications, inputting blurred or original images into the trained network yields clear images with enhanced features, laying the foundation for subsequent advanced vision tasks.

During the flight of UAVs platforms, infrared image acquisition is affected by high-speed flight, attitude changes, and rotor vibrations, resulting in defocus blur, motion blur, and mixture blur [30]. The blur noise modeling is as follows:

Motion blur, caused by the rapid movement of the UAV platform, can be simulated by convolving the original infrared image with an affine transformation matrix, as shown in Equation (1).

$$\mathbf{I}_m = \mathbf{I}_o * \mathbf{W}_A \quad (1)$$

where $\mathbf{I}_m$ represents the motion-blurred infrared image, $\mathbf{I}_o$ denotes the original infrared image, $*$ signifies the convolution operation, and $\mathbf{W}_A$ represents the affine transformation matrix.

Defocus blur in infrared images can result from improper focusing of the infrared imaging equipment. This phenomenon can be simulated by convolving the original infrared image with a Gaussian filter, as expressed in Equation (2).

$$\mathbf{I}_d = \mathbf{I}_o * \mathbf{G}_F \quad (2)$$

where $\mathbf{I}_d$ represents the infrared image with defocus blur, and $\mathbf{G}_F$ signifies the Gaussian filter. Mixed blur can be simulated by sequentially convolving the original infrared image with a Gaussian filter and an affine transformation matrix [31], as shown in Equation (3).

$$\mathbf{I}_{\text{mix}} = \mathbf{I}_o * \mathbf{G}_F * \mathbf{W}_A \quad (3)$$

Furthermore, due to inherent limitations, infrared imaging devices generate both fixed pattern noise and random noise during image acquisition [32]. These noise types can be modeled as follows:

Fixed pattern noise, resulting from non-uniformity in the infrared detector's response, imaging defects, and clutter interference, is modeled as multiplicative noise, as expressed in Equation (4).

$$\mathbf{x}_i = \mathbf{a}_i \mathbf{I} + \mathbf{b}_i \quad (i = 1, 2, \dots, N) \quad (4)$$

where $\mathbf{I}$ represents the incident infrared radiation, $\mathbf{x}_i$ denotes the response of the detector element, and N is the number of detector elements in the array. It is evident that the gain ($\mathbf{a}_i$) and offset ($\mathbf{b}_i$) of each unit reflect the non-uniformity of individual detector elements.

Random noise, caused by photon fluctuations in infrared background radiation, photo-electric conversion noise in the infrared detector, and additional noise from signal readout and processing circuits, is modeled as Poisson noise. Its probability density function (PDF) is shown in Equation (5).

$$P(x = m) = \lambda^m e^{-\lambda} / m! \quad (m = 0, 1, 2, \dots; \lambda > 0) \quad (5)$$

To ensure the reproducibility of our blur simulation, we specify the parameter ranges employed for each blur type. For motion blur, the affine transformation matrix $\mathbf{W}_A$ incorporated rotation angles uniformly sampled from $[-15^\circ, +15^\circ]$ and translation displacements along both axes within $[-20, +20]$ pixels. Defocus blur was simulated using a Gaussian filter $\mathbf{G}_F$ with kernel sizes varying between 3×3 and 5×5 , and standard deviations σ ranging from 0.5 to 4.0. For mixed blur, we sequentially applied the Gaussian filter and affine transformation matrix, with parameters independently drawn from the aforementioned ranges. Additionally, to emulate real-world infrared sensor imperfections, fixed pattern noise was introduced with gain $\mathbf{a}_i$ and offset $\mathbf{b}_i$ varying by $\pm 10\%$ around the nominal response, while Poisson noise was added to simulate photon fluctuations with λ scaled proportionally to the local image intensity.

3.3. Loss Function and Training Strategy

To comprehensively evaluate the local and global feature differences between the degraded and original images, we designed a multi-scale loss function. This function incorporates regional pixel loss, deep feature loss and gradient mixed loss.

3.3.1. Regional Pixel Loss

In infrared image enhancement tasks, methods relying on L2 loss to calculate image differences primarily focus on disparities between individual pixel values, neglecting the relationships among pixels. This limitation leads to restricted perception of local features. To address this challenge, as shown in Equation (6), we introduce a novel loss function that incorporates pixel relationships into the calculation of image differences, thereby enabling a more accurate assessment of image disparities. This approach effectively integrates the connectivity of regional pixels into the loss calculation by adjusting weights $\omega(i, j)$ based on local pixel similarities. This loss not only reflects differences between individual pixels but also contributes to a more comprehensive evaluation of overall image differences, particularly emphasizing visually continuous or highly similar regions.

$$L_{\text{pd}} = \frac{1}{N} \sum_{i,j} \omega(i, j) |\mathbf{I}_o(i, j) - \mathbf{I}_d(i, j)|^2 \quad (6)$$

The calculation of $\omega(i, j)$ is shown in Equation (7), where α is a regulatory parameter controlling the degree of influence that similarity has on the weight. The regulatory parameter α was empirically determined through subsequent ablation studies. We evaluated a range of values and found that $\alpha = 0.5$ effectively balances the contribution of local structural similarity against the global pixel-wise fidelity. A lower α undervalues the local pixel relationships, reverting the loss towards a standard L2 norm, while a higher α over-penalizes minor intensity variations, potentially introducing artifacts. The chosen value of 0.5 consistently yielded optimal performance across our validation set, demonstrating a robust compromise that enhances local feature perception without compromising overall image integrity. $D(i, j)$ represents the average difference of the 8 surrounding pixels at corresponding positions in the two images, as specifically illustrated in Equation (8).

$$\omega(i, j) = e^{-\alpha D(i, j)} \quad (7)$$

$$D(i, j) = \frac{1}{8} \sum_{u=i-1}^{i+1} \sum_{v=j-1}^{j+1} |\mathbf{I}_1(u, v) - \mathbf{I}_2(u, v)|, (u \neq i, v \neq j) \quad (8)$$

3.3.2. Deep Feature Loss

The deep feature loss is defined as shown in Equation (9). The degraded enhanced image output by the network and the original image (GT) are separately input into a pre-trained VGG-16 model to extract their feature representations. Mean Squared Error (MSE) is used as the loss metric to calculate the difference between these two feature representations. This ensures that the degraded enhanced image and the original image maintain consistency at the perceptual level, which is crucial for the visual quality of the image.

$$L_{df} = \frac{1}{N} \sum_{i=1}^N |VGG(\mathbf{I}_{oi}) - VGG(\mathbf{I}_{ei})|^2 \quad (9)$$

where $\mathbf{I}_{oi}$ represents the original image serving as the GT, $\mathbf{I}_{ei}$ represents the degraded enhanced image output by the network, $VGG(\cdot)$ represents the convolution operation of the VGG-16 network, and N represents the training batch size.

3.3.3. Gradient Mixed Loss

The gradient mixed loss is shown in Equation (10), which comprehensively considers both the gradient information and pixel-level differences of the image. Gradient information represents the rate of change in pixel values of the infrared image, manifesting as edge information and detailed textures in the image. By minimizing the gradient difference between the degraded enhanced image and the original image, the network can learn to enhance image edges and textures. To avoid introducing excessive noise and causing distortion while enhancing image details, we incorporate the L1 loss function on this basis. The L1 loss helps the network learn pixel-level consistency of the image, thereby maintaining consistency between the enhanced image and the original image in terms of overall brightness and visual effect.

$$L_{gm} = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{\partial \mathbf{I}_{oi}}{\partial x} - \frac{\partial \mathbf{I}_{ei}}{\partial x} \right| + \left| \frac{\partial \mathbf{I}_{oi}}{\partial y} - \frac{\partial \mathbf{I}_{ei}}{\partial y} \right| + \beta |\mathbf{I}_{oi} - \mathbf{I}_{ei}| \right) \quad (10)$$

where $\frac{\partial \mathbf{I}}{\partial x}$ and $\frac{\partial \mathbf{I}}{\partial y}$ represent the gradients of image $\mathbf{I}$ in the horizontal and vertical directions, respectively. β is a weight parameter, adjusted based on experimental results, with a default value of 1. The weight parameter β is set to 1 to establish an equal contribution between the gradient loss and the L1 pixel loss. This balanced weighting was concluded from systematic observation during our experimental phase. We found that assigning equal importance to both terms ensures that the network simultaneously prioritizes edge

sharpness (through gradient alignment) and global intensity consistency (through L1 minimization). This 1:1 ratio provided the most stable convergence behavior and superior quantitative results compared to other tentative values (e.g., 0.5, 2), avoiding the need for an additional hyperparameter that could lead to over-fitting on a specific dataset.

3.3.4. Progressive Training Strategy

In the realm of infrared image enhancement, the design of the training strategy is crucial, as it directly impacts the model's learning effectiveness and ultimate enhancement performance. Inspired by diffusion models, which simulate the gradual evolution of data akin to the progressive enhancement of images from low to high quality, we implement a stage-wise training strategy as a practical realization of diffusion model principles in our infrared image enhancement task.

Our approach involves adjusting the loss function across different stages to gradually guide the model in learning deblurring and feature enhancement capabilities. This progressive training strategy allows the model to focus on different learning objectives at various stages, incrementally improving its overall performance. Such a gradual learning mechanism also helps prevent overfitting or local optima traps in the early training phases.

The specific training strategy is as follows. In the initial phase (0–0.2 epochs), we employ deep feature loss as the total loss function. This enables the model to learn high-level visual features, which is crucial for preliminary deblurring and enhancing image structure. In the early training phase (0.2–0.4 epochs), we introduce regional pixel loss as the total loss function. This significantly enhances pixel-level consistency, ensuring the enhanced image maintains higher visual similarity to the original. In the mid-training phase (0.4–0.6 epochs), gradient mixed loss is used for optimization, allowing the network to learn the ability to enhance image edges and textures. In the late training phase (0.6–1 epoch), we utilize a weighted combination of the aforementioned loss functions as the total loss. This comprehensively improves the image's pixel-level accuracy, perceptual quality, and structural consistency.

Through this progressive training strategy, the model gradually acquires deblurring and feature enhancement capabilities, achieving high-quality infrared image enhancement. The total loss function incorporates regional pixel loss, deep feature loss, and gradient mixed loss, each playing a crucial role in different stages of the training process.

4. Experiments

4.1. Experiment Environment

To evaluate the effectiveness of the proposed method, ablation studies and comparative experiments were designed. The hardware platform settings applied in these experiments are shown in Table 1. Our experiments were conducted under the Python 3.10.12 and PyTorch 2.1.1+cu121 framework. For both training and testing phases, we utilized a computing platform equipped with an NVIDIA GeForce RTX 3090 GPU.

Table 1. Applications in each class.

Names	Related Configurations
GPU	NVIDIA GeForce RTX 3090
CPU	Inter(R) Core™ i9-13900K
Operating system	Ubuntu 22.04
Deep learning framework	Pytorch 2.1.1 + CUDA 12.1

4.2. Assessment Indicators

In this study, we selected three commonly used metrics to evaluate the performance of infrared image blind deblurring. Specifically, these include Root Mean Square Error (*RMSE*), Peak Signal-to-Noise Ratio (*PSNR*), and Structural Similarity Index (*SSIM*). Additionally, two no-reference metrics were chosen to assess feature enhancement performance in practical applications: Metric of Enhancement (*EME*) and Signal-to-Noise Ratio based Measure of Enhancement (*SNRME*). Furthermore, inference time on edge devices (Time) was used to evaluate the practical application inference speed of the algorithm.

RMSE represents the average pixel difference between the degraded enhanced image and the corresponding original image. A lower *RMSE* indicates a smaller discrepancy between the degraded enhanced image and the corresponding original image, suggesting stronger image stabilization and feature enhancement performance of the network. It is defined as shown in Equation (11).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_{oi} - I_{ei})^2} \quad (11)$$

PSNR (Peak Signal-to-Noise Ratio) is used to evaluate the ratio between the maximum pixel value and the mean squared error of deblurring in an image. A higher *PSNR* value indicates better image deblurring performance. Specifically, *PSNR* objectively reflects the degree of image quality improvement, especially in cases where blur is introduced by motion and noise. The calculation method is shown in Equation (12), where *P*MAX represents the maximum pixel value in the image, which is 255 for infrared images.

$$PSNR = 20 \log_{10} \left(\frac{P_{MAX}}{RMSE} \right) \quad (12)$$

SSIM (Structural Similarity Index Measure) comprehensively assesses the similarity between a clear image and its corresponding deblurred image from three aspects: luminance, contrast, and structure. A higher *SSIM* value indicates better image deblurring performance. The *SSIM* ranges from 0 to 1, where 1 represents two identical images. This metric provides a more perceptually relevant evaluation of image quality compared to pixel-based measures. The calculation method is shown in Equation (13).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (13)$$

where μ_x and μ_y represent the mean pixel values of the degraded-enhanced image and the original image, respectively. σ_x , σ_y , and σ_{xy} denote the variance of the degraded-enhanced image, the variance of the original image, and the covariance between the degraded-enhanced image and the original image, respectively. C_1 and C_2 are two parameters used to maintain stability, which are set to 2.55 and 7.65 in this study.

In this study, we employ two no-reference image quality metrics to evaluate the enhancement performance: the Metric of Enhancement (*EME*) and a Signal-to-Noise Ratio based Measure of Enhancement (*SNRME*). The *EME* is calculated as Equation (14).

$$EME = \frac{1}{N_b} \sum_{k=1}^{N_b} 20 \log_{10} \left(\frac{I_{\max,k}}{I_{\min,k}} \right) \quad (14)$$

where N_b is the total number of 8×8 pixel blocks in the image, and $I_{\max,k}$ and $I_{\min,k}$ are the maximum and minimum pixel values in the k -th block, respectively. A higher *EME* value indicates better local contrast.

The *SNRME*, based on the image's signal-to-noise ratio, is computed as Equation (15).

$$SNRME = \frac{\mu}{\sigma} \quad (15)$$

where μ is the mean pixel value of the image, and σ is the standard deviation of pixel values. A higher *SNRME* value suggests a better balance between signal strength and noise level in the image.

4.3. Ablation Experiments

To validate the effectiveness of the proposed modules, ablation experiments were conducted, and comparative studies with various state-of-the-art algorithms were performed under identical conditions. All experimental models were trained from scratch without pre-trained weights. The input image size was adjusted to 480×640 , with a batch size of 8 and training duration of 100 epochs.

The results of the ablation experiments are presented in Table 2, with visual examples shown in Figure 4. The analysis reveals several key insights regarding the contribution of each component.

Table 2. Comparison of metrics for ablation experiments.

Diffusion Module	Regional Pixel Loss	Progressive Training	PSNR/dB	SSIM	RMSE	EME	SNRME
			31.35	0.8678	7.16	1.32	3.38
✓			34.38	0.8721	5.87	1.45	3.39
✓	✓		35.56	0.8715	4.33	1.53	3.42
✓		✓	36.44	0.8919	3.89	1.59	3.44
	✓	✓	36.93	0.9006	3.68	1.64	3.47
✓	✓	✓	36.95	0.9012	3.67	1.66	3.49

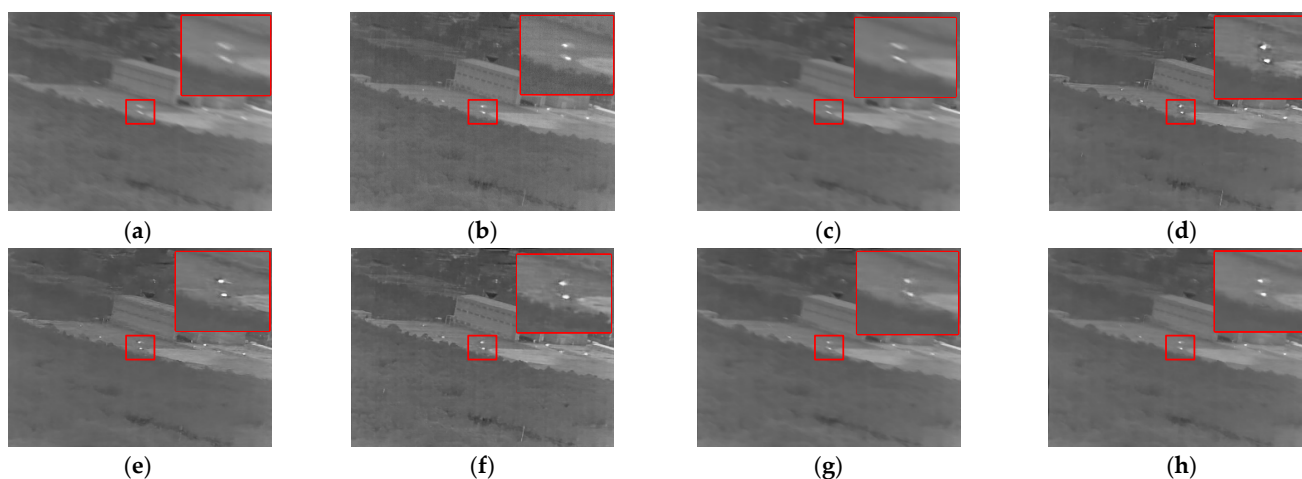


Figure 4. Comparison of ablation experiments results. (a) Blurred image, (b) Original image, (c) No configuration, (d) Diffusion Module only, (e) Diffusion Module + Regional Pixel Loss, (f) Diffusion Module + Progressive Training, (g) Regional Pixel Loss + Progressive Training, and (h) Full configuration. The red box is an enlarged detail illustration.

First, the significant performance leap from the baseline (Row 1) to the model with only the Diffusion Module (Row 2) across all metrics—especially PSNR (31.35 to 34.38 dB) and EME (1.32 to 1.45)—underscores its role as the foundational component for global deblurring and contrast improvement. It effectively initiates the restoration process by suppressing high-frequency noise and recovering low-frequency structural information.

Second, the efficacy of the Regional Pixel Loss and Progressive Training Strategy as powerful independent enhancers is evident. Rows 3 and 4 demonstrate that introducing

either one alongside the Diffusion Module yields substantial further gains. Notably, the configuration employing only these two modules, without the Diffusion Module (Row 5), achieves performance (PSNR 36.93 dB, EME 1.64) that is remarkably close to the full model (Row 6). This indicates that the Regional Pixel Loss and Progressive Training Strategy together possess a strong capability for local detail recovery and overall optimization, to the extent that they can, to a large degree, compensate for the absence of the dedicated Diffusion Module. The synergy between them drives high pixel-level accuracy and superior perceptual quality.

Ultimately, the full model (Row 6), which integrates all three components, achieves the most balanced and robust performance. While its quantitative metrics are comparable to Row 5, its incorporation of the Diffusion Module ensures a more stable and generalized restoration process, as observed qualitatively in Figure 4h, which exhibits superior clarity and artifact suppression compared to Figure 4g. This validates that the Diffusion Module provides a crucial deblurring prior, while the other two components focus on enhancement and optimization, resulting in a synergistic effect that delivers both high-quality and reliable outputs.

4.4. Comparison with State-of-the-Art Methods

To evaluate the applicability and effectiveness of the proposed method for infrared image deblurring and feature enhancement in UAV scenarios, comparative studies were conducted with various state-of-the-art algorithms under identical conditions. The experimental results are presented in Table 3, where the bolded data represent the optimal values.

Table 3. Performance comparison with other algorithms.

Model	Parameters	PSNR/dB	SSIM	RMSE
IREGAN [33]	1.05 M	31.87	0.8619	7.16
PWStableNet [21]	7.91 M	30.89	0.8542	7.50
Improved U-Net [22]	7.29 M	31.85	0.9065	6.89
Uformer-T [23]	5.23 M	34.33	0.9144	6.78
HCTIRdeblur [30]	4.77 M	33.38	0.9367	5.43
Ours	3.89 M	36.95	0.9012	3.67

The results demonstrate that our method achieves significant advantages across multiple evaluation metrics. Specifically, compared to the latest advancement in [30], our method reduces the parameter count by 18.4% while improving PSNR by 10.7%. This can be attributed to several factors. While PWStableNet and improved U-Net show good performance, they have a large number of parameters and high computational complexity. IREGAN adopts a lightweight design with fewer parameters but struggles to handle complex noise, resulting in lower PSNR and SSIM values.

In contrast, our approach employs a lightweight feature extraction module that fuses multi-level features at low computational cost. The diffusion module, utilizing cascaded ResB models, progressively processes noise, enhancing image restoration capabilities. Furthermore, the introduction of a regional pixel loss function and a staged training strategy further improves the model's ability to deblur images and enhance features.

It is noted that our method's SSIM is lower than some competitors, despite superior PSNR/RMSE. This stems from a fundamental design choice: our network performs joint deblurring and active enhancement. The regional pixel loss deliberately enhances local features and contrasts beyond the original image to improve target saliency for detection. This enhancement alters local image statistics, slightly reducing structural similarity with the GT, but yields significantly better pixel-level accuracy and, most importantly, provides

a more effective input for downstream vision tasks, as confirmed by our object detection results presented in Section 4.5.

The experimental results of different algorithms are illustrated in Figure 5. It can be observed that the method in [33], with its shallow convolutional layers, exhibits limited feature extraction capability and fails to effectively restore image information. The approach in [21] successfully recovers the contour information of targets but lacks overall contrast. Methods in [22,23] restore target contour information with good contrast, but suffer from significant loss of detail information.

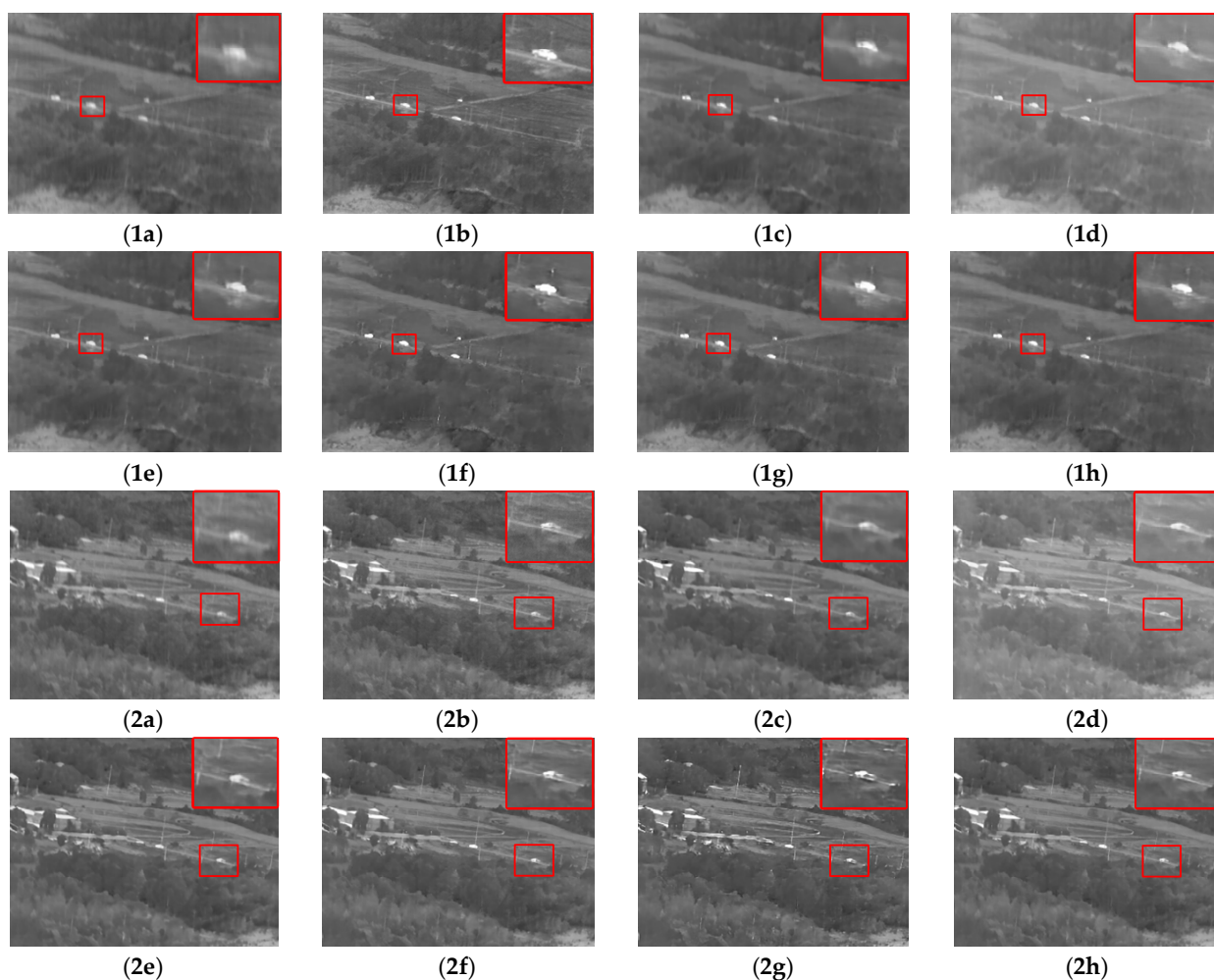


Figure 5. Two sets of comparative experimental results with state-of-the-art methods. From (1a–1h): blurred image, original image, method of [33], method of [21], method of [22], method of [23], method of [30], and ours. From (2a–2h) are the same.

In comparison, our proposed method achieves effective global blind deblurring of infrared images and local enhancement of target features. The restored images demonstrate superior clarity, improved contrast, and preservation of fine details. This comprehensive improvement in image quality is particularly evident in the enhanced visibility of target features and the overall sharpness of the image.

4.5. Real-World Scenario Evaluation

To validate the practical application value of our method, we implemented algorithm compilation and hardware deployment using the Vitis-AI framework, and conducted real-world scenario testing for infrared image deblurring and feature enhancement on an UAV

platform DJI MATRICE 300 RTK, equipped with Master 600 uncooled Vanadium Oxide (VOx) IR camera as its payload (Figure 6).

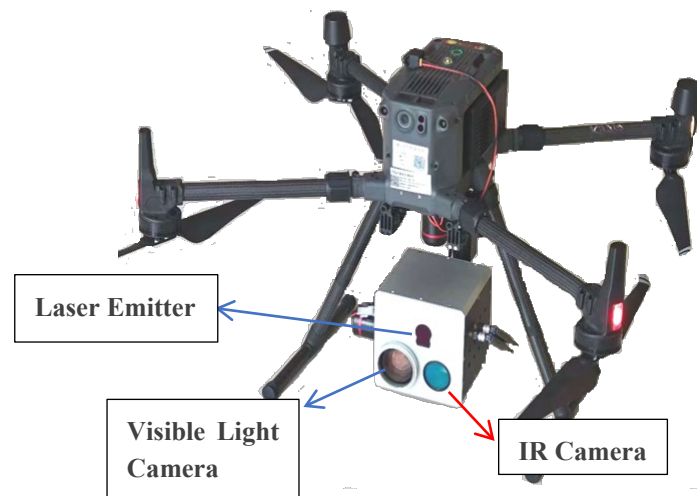


Figure 6. DJI Drone Payload.

In the AI full-stack deployment framework for edge computing devices, Vitis-AI serves as the compiler and backend, receiving network parameters trained by frontend Deep Neural Network (DNN) frameworks, optimizing and compiling them, and then transferring the results to the backend for edge device utilization. The algorithm deployment process is illustrated in Figure 7.

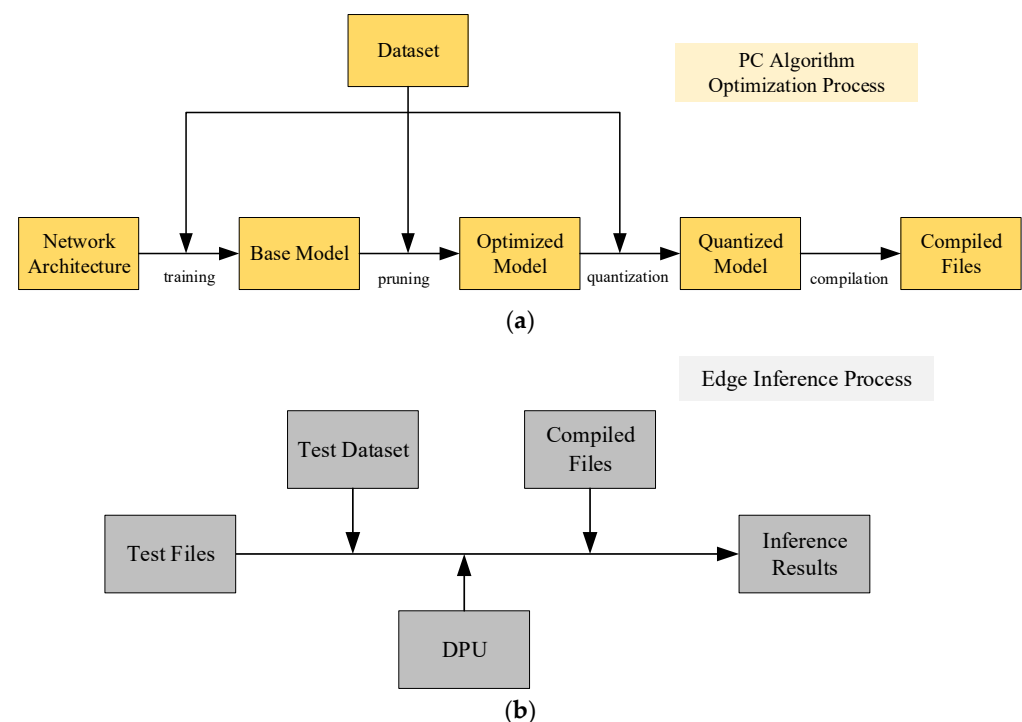


Figure 7. Optimization and deployment process for deep neural network models using the Vitis-AI, with (a) representing the host-side optimization process and (b) denoting the edge-side inference process.

On the host side, the process begins with constructing and training the base model using the PyTorch deep learning framework. Subsequently, the Vitis-AI tool is employed for model pruning to obtain an optimized model. Further quantization is performed on the

model, converting floating-point weights to fixed-point representations to align with FPGA computational characteristics. Finally, the Vitis-AI compiler compiles the model to generate compilation files. Concurrently, the compiler optimizes the network model for specific Xilinx hardware architectures, such as the ZU5EV chip, to achieve optimal performance and resource utilization. On the edge side, the optimized model can be loaded onto the Zynq UltraScale+ MPSoC ZU5EV chip for execution.

The quantitative performance metrics of various enhancement methods are presented in Table 4, with the corresponding visual results illustrated in Figure 8. The bolded data in Table 4 represent the optimal values. Our proposed method demonstrates notable improvements in image information content, exhibiting increases of 1.8% and 0.6% in EME and SNRME, respectively, compared to state-of-the-art approaches. Visually, our method achieves superior clarity and contrast enhancement, significantly improving target edge definition and detail preservation. In contrast, other algorithms struggle with detail blurring or insufficient contrast, particularly in complex background scenarios.

Table 4. Performance metrics comparison of various algorithms in real-world scenarios.

Model	EME	SNRME	Time/Ms	Power/W	mAP/%
IREGAN [33]	1.25	3.45	5.6	1.36	83.3
PWStableNet [21]	1.40	3.52	17.3	4.50	85.0
Improved U-Net [22]	1.51	3.50	16.8	4.24	85.9
Uformer-T [23]	1.54	3.52	9.2	2.58	86.8
HCTIRdeblur [30]	1.63	3.47	8.2	2.16	86.5
Ours	1.66	3.49	6.1	1.95	87.2

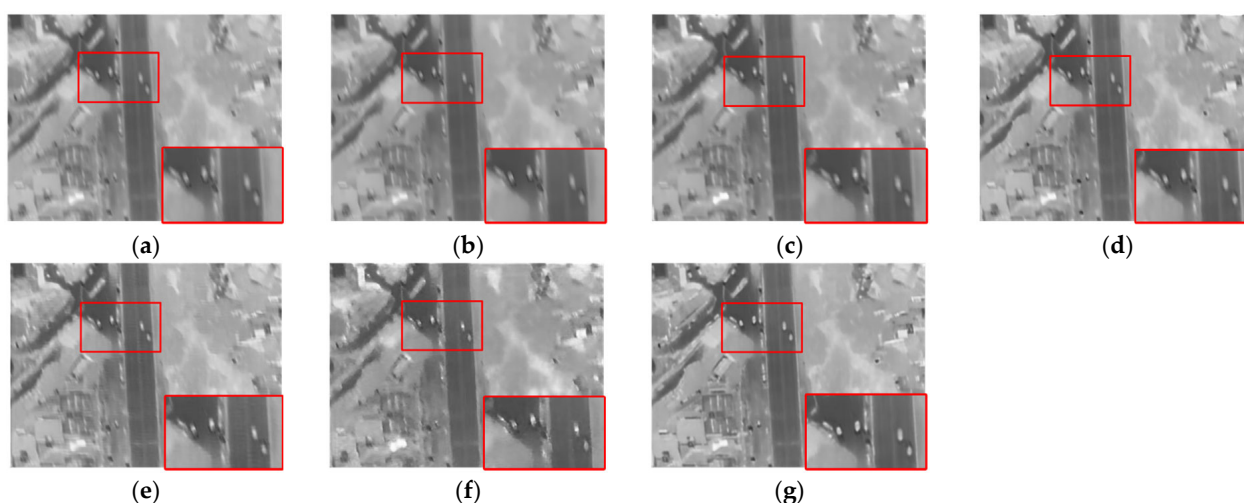


Figure 8. Comparison of experimental results from various algorithms in real-world scenarios. From (a–g): original image, method of [33], method of [21], method of [22], method of [23], method of [30], and ours.

Although the methods proposed in [21,22] achieve relatively high SNRME values, their complex model structures pose challenges for edge deployment and inference. Our proposed method, however, strikes a balance between accuracy and computational efficiency, offering faster inference speeds. This makes it more suitable for practical applications, especially in resource-constrained environments. To quantitatively validate the claim that our enhancement method facilitates subsequent high-level vision tasks, we evaluated object detection performance on the enhanced imagery using a pre-trained YOLOv11 network [34]. As presented in Table 4, our method achieves the highest mAP of 87.2%, outperforming all competing approaches. This result provides concrete evidence that the superior deblurring

and contrast enhancement of our model directly translates into more reliable and accurate object detection, as demonstrated by its high EME and the visual results in Figure 8. The corresponding comparisons in Figure 9 further show that our enhanced images enable the detector to identify targets with greater confidence and reduce false negatives in cluttered backgrounds. Thus, this quantitative analysis confirms that our real-time deblurring network not only improves perceptual image quality but also establishes a more robust foundation for critical downstream applications in UAV infrared vision.

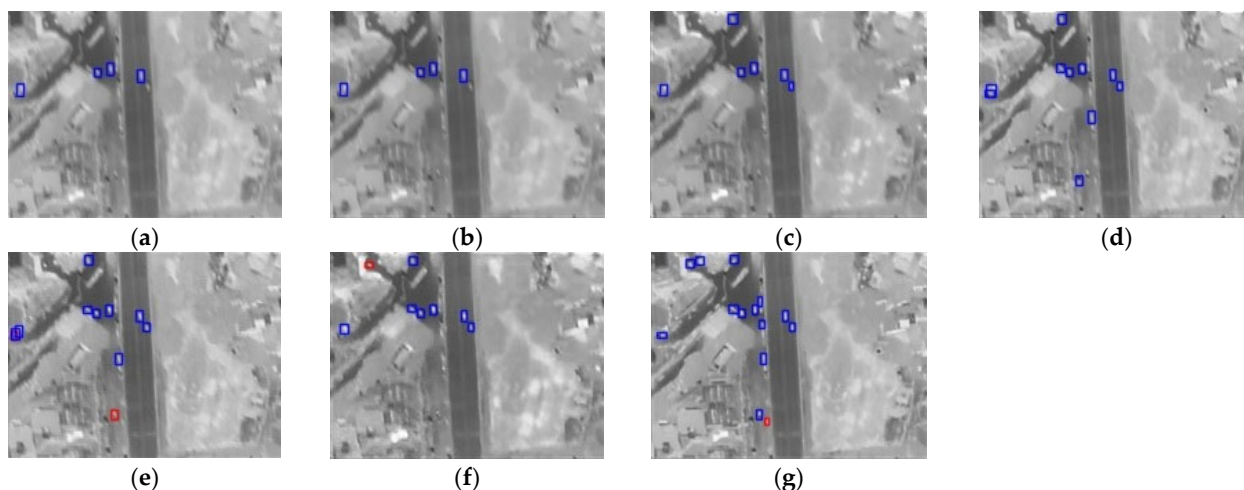


Figure 9. Visual comparison of object detection performance on images deblurred by different algorithms, demonstrating the practical benefit of our method for downstream tasks. Detections are generated by a pre-trained YOLOv11 model. From (a–g): original image, method of [33], method of [21], method of [22], method of [23], method of [30], and ours. In the figure, the blue bounding boxes represent car targets, while the red bounding boxes denote person targets.

Furthermore, the power consumption of the FPGA implementation is a critical consideration for UAV-based edge platforms, where energy efficiency directly impacts mission endurance. As summarized in Table 4, our method achieves a competitive power consumption of only 1.95 W during real-time inference, significantly lower than several recent approaches such as PWStableNet (4.50 W) and Improved U-Net (4.24 W). This efficiency stems from our lightweight architecture design—with only 3.89 M parameters—effectively balancing computational complexity and inference quality. Such low-power operation, combined with faster processing speed and superior enhancement performance, underscores the practical viability of our system in resource-constrained aerial imaging scenarios.

The experimental results validate the efficacy of our proposed approach in enhancing infrared images, demonstrating its superior capability in preserving fine details while significantly improving overall image quality. Our method achieves this through efficient feature extraction and fusion strategies, coupled with the synergistic effects of the diffusion module and region-specific pixel loss. This combination enables more effective deblurring and feature enhancement of infrared images in UAV scenarios. Furthermore, the computational efficiency of our method addresses the growing demand for real-time processing in various infrared imaging applications. By providing higher quality, enhanced infrared images, our approach lays a more reliable foundation for subsequent advanced vision tasks, such as object detection and tracking. This improvement in image quality and processing speed represents a significant step forward in infrared image processing, particularly for applications requiring rapid, on-board analysis in dynamic environments.

To further analyze the system latency, we deconstructed the total inference time into three distinct stages: (1) the pre-processing stage, executed by the ARM processing unit,

which handles image decoding and resizing; (2) the network inference stage, accelerated by the dedicated DPU kernel, which performs convolutional operations; and (3) the post-processing stage, again handled by the ARM unit, responsible for image display. The profiling results, detailed in Table 5, reveal a significant disparity in the time distribution: the pre-processing stage dominates the latency, accounting for 64% of the total time, while the core network inference and post-processing stages constitute 15% and 21%, respectively. This imbalance can be primarily attributed to the architectural mismatch between the tasks and the processing units. The pre- and post-processing stages, involving memory-intensive and sequential operations, are executed on the general-purpose ARM processor, which becomes a bottleneck. In contrast, the network inference stage is highly optimized and off-loaded to the specialized DPU, a domain-specific architecture designed for efficient parallel computation of convolutional layers, resulting in its markedly lower latency. Consequently, these findings indicate that the primary target for subsequent optimization should be the acceleration of the pre- and post-processing stages, potentially through algorithmic refinements, instruction-level optimizations on the ARM cores, or the integration of lightweight hardware accelerators for specific data handling tasks.

Table 5. Breakdown of the end-to-end inference time.

Stage	Processing Unit	Key Operations	Time Consumption (%)
Pre-processing	ARM	Image Decoding, Resizing	64
Network Inference	DPU Kernel	Convolutional Operations	15
Post-processing	ARM	Image Display	21

Additionally, to evaluate the hardware efficiency and feasibility for deployment on resource-constrained UAV platforms, the post-implementation resource utilization of our design on the Zynq UltraScale+ MPSoC is summarized in Table 6. The results demonstrate a balanced resource consumption profile. The LUT utilization of 55% is expected, as it primarily drives the compute-intensive operations of the Deep Processing Unit (DPU). A moderate Flip-Flop usage of 33% reflects an efficient pipeline design, while the 22% consumption of Block RAM is adequate for storing the parameters of our model after quantization and the intermediate feature maps. This resource footprint, combined with a total power consumption of 1.95 W, confirms the practical viability of our system for embedded aerial applications.

Table 6. Hardware Resource Utilization Analysis.

	LUTs	FFs	BRAM
Total Available	117,120	234,240	144
Amount Utilized	64,416	77,299	32
Utilization Ratio	55%	32.9%	22.2%
Primary Function	Compute, Control Logic	Pipeline Registers	Feature Map, Weight Storage

5. Conclusions

In this article, we proposed a novel single-image blind deblurring and enhancement network for UAV infrared imaging systems. This approach effectively addresses the inherent limitations of uncooled infrared sensors and the image degradation caused by platform vibrations, which typically result in low contrast and detail loss. Our network architecture, comprising feature extraction, fusion, and simulated diffusion modules, achieves global blind deblurring and local feature enhancement of infrared images.

The introduction of a region-specific pixel loss function significantly improved the model's ability to capture local features, while the progressive training strategy further enhanced overall performance. Experimental results demonstrate that our method reduces parameter count by 18.4%, improves PSNR by 10.7%, and decreases FPGA-based edge-device inference time by 25.6% compared to state-of-the-art approaches such as HCTIRdeblur [30]. These improvements not only enhance the quality of UAV infrared images but also provide a more reliable foundation for subsequent advanced vision tasks.

Despite the promising results, certain limitations of the proposed methodology warrant discussion. Firstly, the model's performance is inherently tied to the diversity and quality of the training dataset. While our custom dataset covers various blur types, it may not encompass all possible real-world degradation scenarios encountered by UAVs in extreme conditions. Secondly, as a supervised learning approach, our method relies on the availability of paired sharp-degraded data, which can be challenging to acquire at scale for specific operational environments. Future work will focus on exploring self-supervised or test-time adaptation techniques to mitigate this dependency and enhance the model's generalization capability across a broader spectrum of unseen degradation patterns.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, J.C.; data curation, software, validation, formal analysis, L.P.; investigation, resources, writing—review and editing, T.L.; visualization, writing—review & editing, B.C.; project administration, supervision, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data related to this study are available from the corresponding author upon reasonable request. The codes used during this study are available from the corresponding author upon request. The data are not publicly available due to the specific policies of the institution.

Acknowledgments: We would like to thank the National University of Defense Technology for providing the infrared small target detection and tracking image dataset through the “Aerospace-Cup” competition.

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

References

1. Shu, S.; Fu, Y.; Liu, S.; Zhang, Y.; Zhang, T.; Wu, T.; Gao, X. A correction method for radial distortion and nonlinear response of infrared cameras. *Rev. Sci. Instrum.* **2024**, *95*, 034901. [\[CrossRef\]](#)
2. Tsagkatakis, G.; Aidini, A.; Fotiadou, K.; Giannopoulos, M.; Pentari, A.; Tsakalides, P. Survey of deep-learning approaches for remote sensing observation enhancement. *Sensors* **2019**, *19*, 3929. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Zhang, Z.; Zhu, L. A review on unmanned aerial vehicle remote sensing: Platforms, sensors, data processing methods, and applications. *Drones* **2023**, *7*, 398. [\[CrossRef\]](#)
4. Dhal, K.G.; Das, A.; Ray, S.; Gálvez, J.; Das, S. Histogram equalization variants as optimization problems: A review. *Arch. Comput. Methods Eng.* **2021**, *28*, 1471–1496. [\[CrossRef\]](#)
5. Ma, S.; Yang, C.; Bao, S. Contrast enhancement method based on multi-scale retinex and adaptive gamma correction. *J. Adv. Comput. Intell. Intell. Inform.* **2022**, *26*, 875–883. [\[CrossRef\]](#)
6. Ren, K.; Gao, Y.; Wan, M.; Gu, G.; Chen, Q. Infrared small target detection via region super resolution generative adversarial network. *Appl. Intell.* **2022**, *52*, 11725–11737. [\[CrossRef\]](#)
7. Fan, S.; Liang, W.; Ding, D.; Yu, H. LACN: A lightweight attention-guided convnext network for low-light image enhancement. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105632. [\[CrossRef\]](#)
8. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; MICCAI 2015. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [\[CrossRef\]](#)

9. Kupyń, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8878–8887. [\[CrossRef\]](#)
10. Liu, J.; Chandrasiri, N.P. CA-ESRGAN: Super-resolution image synthesis using channel attention-based ESRGAN. *IEEE Access* **2024**, *12*, 25740–25748. [\[CrossRef\]](#)
11. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 1833–1844. [\[CrossRef\]](#)
12. Fei, S.; Ye, T.; Wang, L.; Zhu, L. LucidFlux: Caption-free universal image restoration via a large-scale diffusion transformer. *arXiv* **2025**, arXiv:2509.22414.
13. Zhang, H.; Zhang, X.; Cai, N.; Di, J.; Zhang, Y. Joint multi-dimensional dynamic attention and transformer for general image restoration. *Comput. Struct. Biotechnol. J.* **2025**, *25*, 102162. [\[CrossRef\]](#)
14. Ye, Y.; Wang, T.; Fang, F.; Zhang, G. MSCSCformer: Multi-scale convolutional sparse coding-based transformer for pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5405112. [\[CrossRef\]](#)
15. Chobola, T.; Müller, G.; Dausmann, V.; Theileis, A.; Taucher, J.; Huisken, J.; Peng, T. LucyD: A feature-driven richardson-lucy deconvolution network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; Springer Nature: Cham, Switzerland, 2023; pp. 656–665. [\[CrossRef\]](#)
16. Zhao, Y.; Fu, G.; Wang, H.; Zhang, S.; Yue, M. Infrared image deblurring based on generative adversarial networks. *Int. J. Opt.* **2021**, *2021*, 9946809. [\[CrossRef\]](#)
17. Wu, W.B.; Pan, Y.; Su, N.; Wang, J.; Wu, S.; Xu, Z.; Yu, Y.; Liu, Y. Multi-scale network for single image deblurring based on ensemble learning module. *Multimed. Tools Appl.* **2025**, *84*, 9045–9064. [\[CrossRef\]](#)
18. Xiao, X.; Qu, W.; Xia, G.S.; Xu, M.; Shao, Z.; Gong, J.; Li, D. A novel real-time matching and pose reconstruction method for low-overlap agricultural UAV images with repetitive textures. *ISPRS J. Photogramm. Remote Sens.* **2025**, *226*, 54–75. [\[CrossRef\]](#)
19. Li, Y.; Chen, S.; Hwang, K.; Ji, X.; Lei, Z.; Zhu, Y.; Ye, F.; Liu, M. Spatio-temporal data fusion techniques for modeling digital twin City. *Geo-Spat. Inf. Sci.* **2024**, *28*, 541–564. [\[CrossRef\]](#)
20. Xiao, X.; Guo, B.; Shi, Y.; Gong, W.; Li, J.; Zhang, C. Robust and rapid matching of oblique UAV images of urban area. In *MIPPR 2013: Pattern Recognition and Computer Vision*; SPIE: Bellingham, WA, USA, 2013; Volume 8919, pp. 223–230. [\[CrossRef\]](#)
21. Zhao, M.; Ling, Q. Pwstabilenet: Learning pixel-wise warping maps for video stabilization. *IEEE Trans. Image Process.* **2020**, *29*, 3582–3595. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Ahmed, Z.; Tanim, S.A.; Prity, F.S.; Rahman, H.; Maisha, T.B.M. Improving biomedical image segmentation: An extensive analysis of U-Net for enhanced performance. In Proceedings of the International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Amaravati, India, 22–23 February 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
23. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693. [\[CrossRef\]](#)
24. Luo, X.; Qu, Y.; Xie, Y.; Zhang, Y.; Li, C.; Fu, Y. Lattice network for lightweight image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4826–4842. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Qi, J.; Abera, D.E.; Fanose, M.N.; Wang, L.; Cheng, J. A deep learning and image enhancement based pipeline for infrared and visible image fusion. *Neurocomputing* **2024**, *578*, 127353. [\[CrossRef\]](#)
26. Huang, W.; Xue, Y.; Hu, L.; Liuli, H. S-EEGNet: Electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation. *IEEE Access* **2020**, *8*, 131636–131646. [\[CrossRef\]](#)
27. Pan, L.; Mo, C.; Li, J.; Wu, Z.; Liu, T.; Cheng, J. Design of lightweight infrared image enhancement network based on adversarial generation. In Proceedings of the International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 9–11 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 995–998. [\[CrossRef\]](#)
28. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.-H. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.* **2023**, *56*, 1–39. [\[CrossRef\]](#)
29. Wu, S.; Gao, X.; Wang, F.; Hu, X. Variation-guided condition generation for diffusion inversion in few-shot image classification. In Proceedings of the International Conference on New Trends in Computational Intelligence (NTCI), Qingdao, China, 3–5 November 2023; IEEE: Piscataway, NJ, USA, 2023; Volume 1, pp. 318–323. [\[CrossRef\]](#)
30. Yi, S.; Li, L.; Liu, X.; Li, J.; Chen, L. HCTIRdeblur: A hybrid convolution-transformer network for single infrared image deblurring. *Infrared Phys. Technol.* **2023**, *131*, 104640. [\[CrossRef\]](#)
31. Cao, S.; He, N.; Zhao, S.; Lu, K.; Zhou, X. Single image motion deblurring with reduced ringing effects using variational Bayesian estimation. *Signal Process.* **2018**, *148*, 260–271. [\[CrossRef\]](#)

32. Li, M.; Nong, S.; Nie, T.; Han, C.; Huang, L.; Qu, L. A novel stripe noise removal model for infrared images. *Sensors* **2022**, *22*, 2971. [[CrossRef](#)]
33. Cheng, J.H.; Pan, L.h.; Liu, T.; Cheng, J. Lightweight infrared image enhancement network based on adversarial generation. *J. Signal Process.* **2024**, *40*, 484–491. [[CrossRef](#)]
34. Khanam, R.; Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [[CrossRef](#)]

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