

Article

Infrared Image Super Resolution Method Based on Stochastic Degradation Modeling

Lihong Yang , Kai Hu *, Hang Ge, Zhi Zeng and Shurui Ge

School of Opto-Electrical Engineering, Xi'an Technological University, Xi'an 710021, China

* Correspondence: hukai@st.xatu.edu.cn

Abstract

Infrared images hold significant application value in fields such as military reconnaissance, security surveillance, and medical diagnosis. However, issues like low resolution, high noise, and complex degradation characteristics severely hinder their practical application effectiveness. This paper introduces an infrared super-resolution reconstruction algorithm based on a random degradation model and Generative Adversarial Networks (GANs), addressing the diversity of infrared image degradation processes. The primary contribution lies in explicitly modeling key degradation parameters of infrared images (such as blur kernel and noise distribution) using a random degradation model, generating diverse low-resolution to high-resolution image pairs, and significantly enhancing the model's generalization ability for complex degradations. Experiments on the Airo infrared dataset and a self-built infrared dataset demonstrate that when the resolution is increased by a factor of 2, 4, and 8, the $4\times$ resolution reconstructed images exhibit notable advantages in terms of texture detail and noise suppression. Especially in $4\times$ super-resolution reconstruction, compared to three typical deep learning algorithms, our algorithm improves the Peak Signal-to-Noise Ratio (PSNR) by 3.046 dB, 1.8489 dB, and 0.2108 dB, respectively, and the Structural Similarity Index (SSIM) by 0.0387 (4.76%), 0.0287 (3.48%), and 0.0131 (1.56%), respectively, with perceptual similarity decreasing by 0.2465, 0.13344, and 0.0514 (lower values indicate better perceptual quality). Subjective visual assessments further validate the algorithm's significant advantages in noise reduction and weak texture restoration. This study proposes an infrared image super-resolution reconstruction method based on random degradation modeling, which holds significant theoretical and practical value in complex infrared degradation scenarios.

Keywords: infrared image; super-resolution; random degradation; super-resolution reconstruction



Received: 3 January 2026

Revised: 27 January 2026

Accepted: 3 February 2026

Published: 5 February 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Infrared imaging technology, with its unique advantages such as all-weather capability, interference resistance, and smoke penetration [1], plays an irreplaceable role in fields such as military reconnaissance, security surveillance, industrial inspection, medical diagnosis, and autonomous driving. However, due to limitations in infrared detector manufacturing processes, imaging physical mechanisms (such as thermal diffusion effects), and complex environmental factors (such as atmospheric attenuation and lens aberrations), the infrared images obtained generally suffer from inherent defects such as low spatial resolution, blurred details, lack of texture, and poor contrast [2], severely limiting the performance of advanced visual tasks such as precise target identification, feature extraction,

and scene understanding. In recent years, with the rapid advancement of deep learning technologies—particularly the successful application of convolutional neural networks and Transformer architectures in computer vision—research on infrared image super-resolution reconstruction has undergone a paradigm shift from traditional interpolation methods to learning-based end-to-end mapping [3]. Early research drew heavily from approaches used in natural image Super-Resolution. However, it gradually became apparent that infrared images possess unique characteristics such as low signal-to-noise ratio, weak contrast, and sparse texture information. Directly applying models designed for visible light often resulted in overly smoothed reconstructions or unnatural, physically implausible textures. Consequently, the current research frontier and hotspots have shifted toward exploring solutions that better align with the intrinsic properties of infrared images. Traditional super-resolution [4] (SR) methods based on interpolation or shallow models often perform poorly when processing such complex degraded infrared images, with reconstruction results prone to issues such as edge jaggedness, artifact amplification, and detail loss. While deep learning-driven SR methods have made significant progress in recent years, their performance highly depends on the precise modeling of image degradation processes [5]. Most existing methods adopt simplified deterministic degradation models (e.g., bicubic downsampling plus fixed noise), which fail to adequately reflect the complex and significantly random degradation factors in infrared imaging chains (e.g., non-uniform noise, random blurring kernels, nonlinear responses, etc.), leading to limited generalization capabilities in real complex degradation scenarios [6] and reconstructed images that may introduce physically unrealistic pseudo-details or overly smooth real structures.

Despite significant progress, this field faces a series of profound challenges and unresolved issues [7]. First, the degradation process of infrared images in practical applications is complex and poorly understood, making it extremely difficult to construct a universal degradation model and limiting the performance ceiling of blind super-resolution algorithms. Second, when utilizing visible light images for cross-modal guidance, significant modal differences and stringent alignment requirements pose risks of introducing semantic errors or artifacts during texture transfer. Furthermore, most existing advanced models rely on substantial computational resources, making real-time processing difficult on resource-constrained embedded or mobile platforms and limiting their engineering applications. Finally, the inconsistency between objective evaluation metrics and subjective visual perception creates a gap between algorithm optimization directions and the ultimate application outcomes.

The core challenge in infrared super-resolution lies in the scarcity of true paired data and the complexity of the degradation process, which has spurred the emergence of unsupervised and semi-supervised methods. Unsupervised approaches entirely eliminate reliance on paired data through degradation modeling and cyclic consistency, intra-image learning, or true degradation estimation. However, they suffer from poor stability and potential artifact introduction risks. Semi-supervised methods combine sparse paired data with vast unlabeled data, guided by embedding physical priors or leveraging visible light cross-modal knowledge distillation. This achieves a better balance between performance and data cost, making it the most promising direction for practical applications. Future trends will focus on integrating more refined physical models, leveraging transfer learning from visual foundation models, and co-optimizing with downstream perception tasks. This will drive the technology toward high-fidelity, strongly generalized practical applications.

This study focuses on the core challenge of insufficient modeling of the randomness of the degradation process in infrared image super-resolution reconstruction and proposes a comprehensive solution based on probabilistic degradation modeling. Through theoretical analysis, model construction, network design, and experimental validation,

it systematically demonstrates that modeling the degradation process as a probabilistic distribution and utilizing it to drive data generation and network learning can effectively enhance the reconstruction performance, generalization capability, and robustness of infrared SR methods in complex real-world scenarios [8]. The study systematically analyzes the random characteristics of key degradation stages in the infrared imaging chain. It proposes a probabilistic degradation generation model for infrared images, capable of simulating the process of generating low-resolution (LR) images with complex, random degradation characteristics from any high-resolution (HR) infrared image. This model not only includes random spatial blurring kernels but also complex signal-related noise and non-signal-related noise, as well as potential non-uniformity effects and compression artifacts. This model lays the theoretical and data foundation for subsequent development of SR methods with strong generalization capabilities [9]. This study provides new insights into infrared image super-resolution reconstruction in complex degradation scenarios, offering significant theoretical and practical value. Super-resolution reconstruction technology aims to recover high-resolution information from low-resolution images [10,11]. Traditional methods (such as bicubic interpolation and sparse representation) rely on fixed degradation assumptions and struggle to adapt to the complex degradation processes in infrared images, which involve multiple coupled factors (such as thermal noise nonlinearity and time-varying blurring kernels).

In recent years, deep learning-based super-resolution algorithms (such as SRCNN and ARCNN) have achieved breakthroughs in visible light images, but their direct application to infrared images faces two major challenges:

- (1) Infrared degradation processes exhibit high randomness, and a single degradation model struggles to capture the diversity of real-world scenarios.
- (2) Infrared images have low signal-to-noise ratios and sparse textures, making traditional convolutional networks susceptible to noise interference, resulting in artifacts and detail distortion in reconstructed images. Additionally, while generative adversarial networks (GANs) can enhance visual quality (e.g., SRGAN), they lack explicit modeling of degradation processes.

To address the above issues, this paper proposes an infrared image super-resolution generative adversarial network based on a random degradation model. The core innovation lies in introducing a probabilistic degradation model, which infers the normal distribution of degradation parameters through learning, generates diverse training samples, and enhances the model's adaptability to complex degradation. Experiments show that compared to existing methods, the improved algorithm model has relative advantages in noise suppression and detail fidelity.

The main contributions of this paper are as follows:

- (1) It proposes a random degradation model that applies degradation variable modeling to the infrared image degradation process, addressing the issue of insufficient degradation diversity in training data;
- (2) It provides new solutions for infrared image enhancement under complex degradation conditions.
- (3) It presents experiments on the Airui infrared dataset and a self-built infrared dataset, which demonstrate that when the resolution is increased by factors of 2, 4, and 8, the $4\times$ resolution reconstructed images exhibit notable advantages in terms of texture detail and noise suppression.

2. Principles of Super-Resolution Reconstruction

This paper builds upon generative adversarial networks (GANs). Reference [12] proposed SRGAN, which employs GANs for image super-resolution reconstruction, intro-

ducing adversarial loss functions and perceptual loss functions into deep learning-based image super-resolution techniques [13]. GANs primarily consist of a generator and a discriminator [14]. The generator synthesizes high-resolution (HR) images, while the discriminator determines whether a given image originates from the generator or actual samples through adversarial competition [15]. This process enables the generator to ultimately reconstruct low-resolution (LR) images into HR images. Figure 1 illustrates the principle of GAN-based image super-resolution reconstruction. Initially, the generator obtains LR images from the validation set and generates corresponding HR images through over-reconstruction. LR images are obtained by downsampling the corresponding HR images in the training set [16–18]. Subsequently, the discriminator evaluates the integrity of the HR images generated by the generator by comparing them with the actual HR images in the training set. The generator is optimized based on the loss value calculated by the discriminator for these two images, and this process is iterated in an adversarial manner. Ultimately, the generator learns to transform a given LR image into an HR image while simultaneously deceiving the discriminator [19,20].

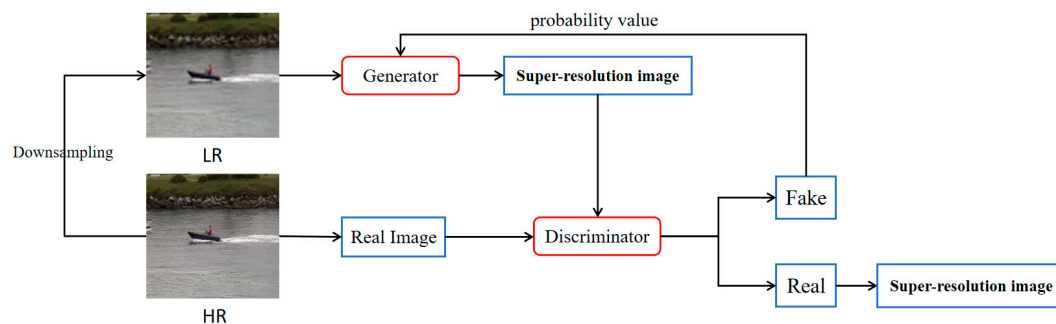


Figure 1. Principles of Image Super-Resolution Reconstruction Using SRGAN.

3. Improved Super-Resolution Reconstruction Network

3.1. Degradation Model

3.1.1. Principles of Degraded Models

The random degradation model is a model used in image super-resolution tasks, aiming to simulate the degradation process of an image from high resolution (HR) to low resolution (LR), as shown in the green processing block in Figure 2. The core idea of this model is to model the degradation process as a distribution of random variables, which can include blur kernels, downsampling operations, additive noise, and so on. Synthetic high-resolution (HR) and low-resolution (LR) images are widely used in existing resolution methods. To avoid domain gaps between synthetic images and test images [21], most previous methods have attempted to adaptively learn the synthetic (degradation) process using deterministic models. However, some degradation in real-world scenarios is random and cannot be determined by the image content. These deterministic models may fail to model the random factors and content-independent components of degradation, thereby limiting the performance of subsequent SR models. In the degradation model, degradation D is treated as a random variable, and its distribution is learned by modeling the mapping from the prior random variable z to D . Compared to previous deterministic degradation models, this model can randomly model more diverse degradations, and the generated HR-LR pairs can better cover various degradations in test images, thereby preventing SR models from overfitting to specific images [22].

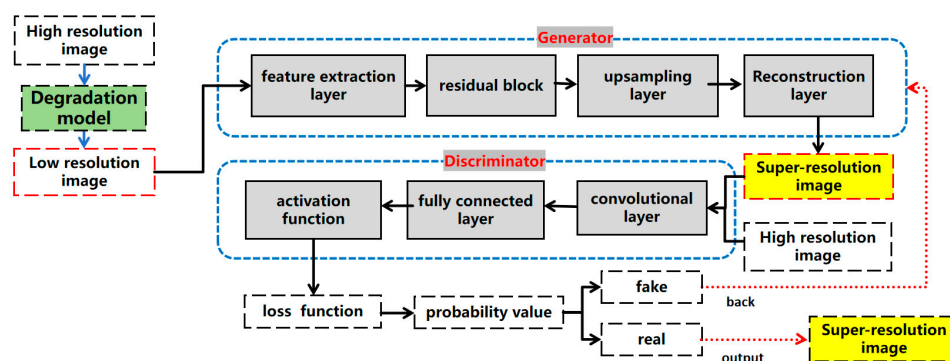


Figure 2. Network model framework diagram.

Super-resolution aims to reconstruct high-resolution images (HR) from low-resolution images (LR). In recent years, learning-based SR methods have achieved significant results, often requiring HR-LR image pairs as training data. However, obtaining paired images in real-world scenarios is extremely challenging. As an alternative, synthetic HR-LR image pairs are widely used in existing super-resolution methods. These methods degrade HR images using predefined degradation designs, such as bilinear down sampling, blurring kernel convolution, or adding noise within a certain range [23]. However, in blind super-resolution, the degradation of test images is unknown, and predefined degradation settings may differ from those of test images. This discrepancy significantly degrades the performance of these methods in real-world scenarios. To avoid domain mismatches, some studies have attempted to adaptively learn degradation settings, with degradation models supervised through adversarial learning [24] and encouraged to generate LR images in the same domain as the test images. In Kernel GAN, the degradation model is designed as a linear convolution layer, while in CycleSR, a nonlinear degradation model is learned within the Cycle GAN framework. These learned degradation models can generate HR-LR pairs with a smaller gap from the test cases and can thus be used to test a better SR model, as shown in Figure 3.

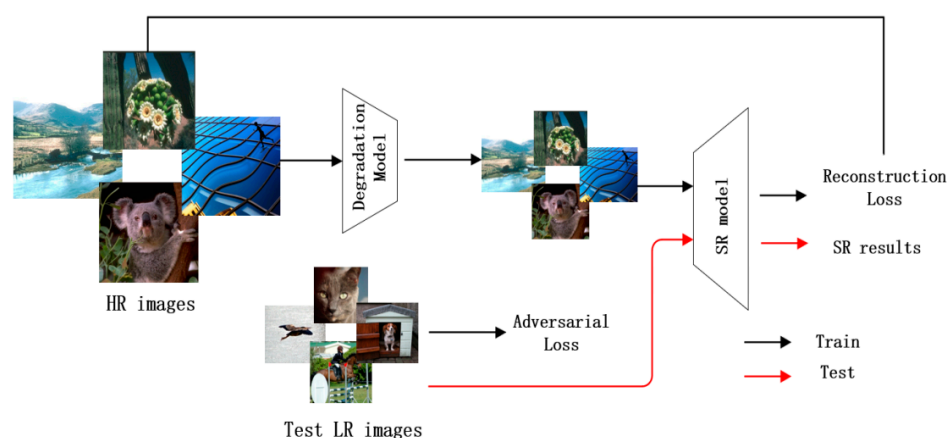


Figure 3. Schematic diagram of the random degradation model.

However, most previous image degradation learning methods share a common drawback: their degradation models are deterministic, meaning that each high-resolution (HR) image can only be degraded into a specific low-resolution (LR) image. This implies the assumption that degradation is entirely dependent on the content of the image; however, this may not hold true in most cases. Some degradations are content-independent and random, such as random noise or blurring caused by random jitter. These random factors and the content-independent nature of degradation cannot be simulated by these deterministic models. A better assumption is that degradation is influenced by a distribution, which can

be better modeled using a probabilistic model [25]. This model can learn the degradation distribution of blind image super-resolution. Specifically, we parameterize degradation using two random variables: the blur kernel k and random noise n . The degradation process is expressed as a linear function, as shown below:

$$D(x) = (x \otimes k) \downarrow_s + n \quad (1)$$

D represents the degradation equation. x denotes the HR image, $\otimes$ denotes the convolution operation, s denotes downsampling and the scale factor, k denotes the blur kernel, and n denotes the added noise. Therefore, the distribution of D can be represented as the joint distribution of k and n . This can be modeled by learning the mapping from the prior random variable z to k and n , and then training the model in an adversarial framework. During training, the distribution of D is automatically learned, enabling the model the random factors in the degradation process.

3.1.2. Degradation Process

The degradation process in Equation (2) actually consists of two linear steps:

$$\begin{cases} y^{clean} = (x \otimes k) \downarrow_s \\ y = y^{clean} + n \end{cases} \quad (2)$$

y^{clean} represents the image after convolution with the blur kernel and subsampling. y represents the image after undergoing a degradation process.

Intuitively, these two steps are independent of each other, because the blur kernel mainly depends on the characteristics of the camera lens, while noise is mainly related to the characteristics of the sensor. Therefore, the degradation distribution can be modeled as

$$P_D(D) = P_{k,n}(k, n) = P_k(k)P_n(n) \quad (3)$$

In this way, the distributions of k (nuclear blur) and n (noise blur) can be modeled independently to represent the distribution of D .

3.1.3. Nuclear Fuzziness

In order to model the distribution of fuzzy kernels k , we define a prior random variable z_k , which is a multidimensional normal distribution. Then, we use a generative module to learn the mapping from z to k , as shown in the equation:

$$k = netK(Z_k), Z_k \sim N(0, 1) \quad (4)$$

Among them, $netK$ is a generative module represented by a convolutional neural network without loss of generality. We first consider spatially varying blur kernels, i.e., the blur kernel for each pixel of x is different. In this case, we have the following equation:

$$Z_k \in R^{f_k * h * w}, k \in R^{(k * k) * h * w} \quad (5)$$

f_k is the dimension of the normal distribution z_k to k ; k is the scale of the blur kernel; and h and w are the height and width of x , respectively. We added a softmax layer at the end to ensure that the sum of all elements of k is 1. The size of the convolution weights in $netK$ is generally set to 3×3 . This indicates that the learned fuzzy kernels are spatially related. Otherwise, if the spatial size of the convolution weights is set to 1×1 , each pixel's fuzzy kernel will be learned independently.

3.1.4. Noise Fuzziness

The second step of degradation is to add noise to the blurred and downsampled image y_{clean} . Previous degradation models only considered additive Gaussian white noise, which is unrelated to the content of clean. In this case, the distribution of n can also be represented by a normal generative model:

$$n = netN(Z_n) \sim N(0, 1) \quad (6)$$

where $netN$ is a convolutional neural network, and we denote the height, width, and channel number of y as h , w , and c , respectively, in the noise domain:

$$Z_n \in R^{f_n * h * w}, n \in R^{h * w * c} \quad (7)$$

f_n is the dimension of the normal distribution z_n . In other methods, noise in the original space n_{raw} is modeled as a combination of shot noise and readout noise, and n_{raw} can be approximated by an anisotropic Gaussian distribution:

$$n_{raw} \sim N(0, \sigma_{read} + \sigma_{shot} y_{clean}), Z_n \in R^{f_n * h * w}, n \in R^{h * w * c} \quad (8)$$

Among them, σ_{read} and σ_{shot} are determined by the analog and digital gain of the camera sensor. Since n comes from n_{raw} , this indicates that n is related to the image content, and the distribution of n should be represented by a conditional generation module as follows:

$$n = netN(Z_n), Z_n \sim N(0, 1) \quad (9)$$

Therefore, we can also adjust the size of the convolution weights in $netN$ to determine whether the noise has spatial correlation.

$$\begin{cases} y_{ref}^{clean} = (x_{ref} \otimes netK(z_k)) \downarrow_s, z_k \sim N(0, 1) \\ y_{ref} = y_{ref}^{clean} + netN(z_n, y_{ref}^{clean}), z_n \sim N(0, 1) \end{cases} \quad (10)$$

x_{ref} is the reference high-resolution image, and $\{x_{ref}, y_{ref}\}$ form a paired training sample for the SR model. Our degradation model is optimized through adversarial training, which encourages y_{ref} to be similar to all low-resolution test images. Additionally, we assume that the mean of noise n is 0, so in addition to the adversarial loss l_{adv} , we add an extra regularization term for noise n :

$$l_{reg} = \|n\|_2^2 \quad (11)$$

The total loss of the degradation model is

$$l_{reg} = l_{adv} + \lambda l_{reg} \quad (12)$$

Together, the kernel module and noise module discussed above form our random degradation model, which is used to synthesize HR-LR image pairs, as shown in Figure 4.

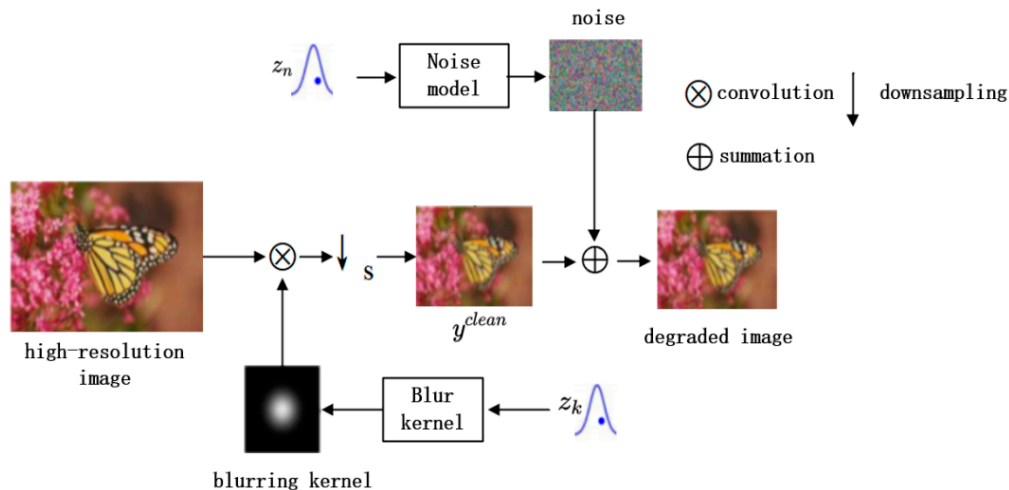


Figure 4. Schematic diagram of super-resolution reconstruction based on probabilistic degradation.

This method does not directly sample degradation parameters from a fixed-parameter normal distribution. Instead, it adopts a “prior variable + generative network” paradigm: prior variables z_k and z_n are set as multidimensional normal distributions to facilitate sampling and ensure training stability. Generative networks net_k and net_N map these prior variables to complex degradation parameter spaces. net_k learns mappings from z_k to blur kernels k , enabling generation of spatially variable, content-dependent blur kernels. net_N learns mappings from z_n (and potentially clean image content y_{clean}) to noise fields n , simulating signal-dependent and independent mixed noise. The final k and n are generated from complex distributions learned from data. The size and intensity range of the fuzzy kernels are set based on analysis of the point spread function (PSF) of real infrared optical systems. For noise parameters, we reference infrared sensor noise models (e.g., shot noise and readout noise models), implicitly introducing a variance-signal strength dependency in net_N 's design, as guided:

$$\sigma^2 = \sigma_{read}^2 + \sigma_{shot}^2 \cdot y_{clean} \quad (13)$$

By employing an adversarial training framework, our degradation model is compelled to generate samples indistinguishable from genuine low-resolution infrared images. This ensures the learned degradation distribution approximates real-world complex scenarios rather than overfitting to a limited training set.

The network architecture of net_K takes the prior variable z_k (64 dimensions) as input. After reshaping into an $8 \times 8 \times 64$ feature map, it sequentially undergoes five convolutional layers, including two $2 \times$ upsampling operations. The channel count evolves as follows: $64 \rightarrow 128 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 1$. ReLU activation functions are applied to all layers except the final one. The final output undergoes Softmax normalization to yield the spatial blur kernel.

The net_N architecture takes the conditional image y_{clean} and prior variable z_n (replicated and concatenated to 33 channels) as inputs. It sequentially processes these through 6 layers of 3×3 convolutions with channel changes: $33 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 1$. The first five layers employ the LeakyReLU activation function (negative slope 0.2). Dropout connections are introduced between the third and fifth layers to enhance feature fusion. The final layer outputs the noise field n directly without an activation function. Both modules adopt lightweight designs and utilize Instance Normalization and Group Normalization respectively to improve training stability. The training flowchart for the regression model is shown in Figure 5.

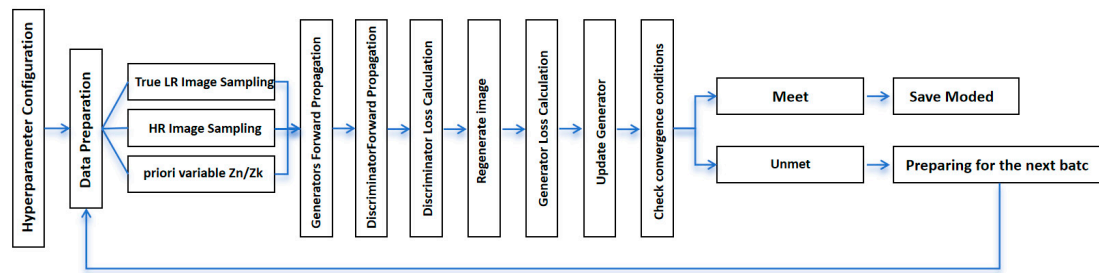


Figure 5. Training flowchart for degenerative models.

In super-resolution tasks, the generator network [26] captures the features of low-resolution images through a feature extraction module and uses these features to reconstruct high-resolution images. This process involves multi-scale feature mapping [27], i.e., learning the correspondence between low-resolution features and high-resolution features at different scales. Generator Network (Generator): This is the core component of SRGAN, responsible for receiving low-resolution images as input and generating high-resolution images. Low-resolution images are input into the feature extraction layer, where 9×9 convolutional kernels are used to extract features from the low-resolution images. These features are then input into the residual block, which contains multiple residual units. Each unit consists of two 3×3 convolutional layers and a PReLU activation function, which helps train deeper networks and mitigate the vanishing gradient problem [28]. The features are then input into the upsampling layer, where sub-pixel convolutional layers are used for upsampling to enlarge the feature map size to the target size. Finally, the features enter the reconstruction layer, where a 9×9 convolutional kernel is used to reconstruct the final high-resolution image. The resulting high-resolution image is input into the discriminator along with the original image. The discriminator network's task is to distinguish whether the input image is a genuine high-resolution image or a synthetic high-resolution image generated by the generator. The discriminator network typically includes the following: multiple convolutional layers, which progressively extract image features; fully connected layers, which map features to binary classification outputs; Sigmoid activation function, which outputs the probability that the image is a real high-resolution image. This probability is used to determine whether the high-resolution image generated by the generator is similar to the original image features, and the probability value is fed back to the generator. As the number of iterations increases, the high-resolution image output by the discriminator becomes the super-resolution image when the discriminator can no longer distinguish between real and fake images, and the image reconstruction quality is ultimately evaluated.

3.2. Image Resolution Analysis

The essence of super-resolution reconstruction algorithms lies in overcoming the resolution limitations imposed by physical sensor sampling, achieving an expansion of the image's spatial dimensions through computational methods [29]. The reconstruction process inevitably results in a significant increase in pixel count (the total number of pixels in the output image is r^2 times that of the input), but a mere increase in pixel count does not guarantee an improvement in image quality. Its core value is that the algorithm models and learns from prior knowledge of natural images (such as structural continuity, texture statistical patterns, and semantic context) to generate high-fidelity information at the newly added pixel locations that aligns with visual perception. This process maps the limited information from low-resolution input to the rich details of high-resolution output, essentially achieving the intelligent reconstruction of high-frequency components lost due to undersampling. However, reconstruction quality is constrained by the upscaling factor, training data distribution, and model architecture—excessively high upscaling rates may

induce artifacts or over-smoothing, while perception-driven methods (such as adversarial loss), though enhancing visual realism, may inherently conflict with pixel-level fidelity objectives. Therefore, the effectiveness of super-resolution technology should be measured by the “information entropy of newly added pixels,” i.e., its contribution to edge sharpness, texture details and structural consistency, rather than relying solely on the numerical increase in resolution. Ultimately, the quality of an algorithm is reflected in its ability to balance pixel expansion with information quality optimization, enabling the reconstructed image to approach the characteristics of a real high-resolution scene in both quantitative metrics (PSNR/SSIM) and subjective perception (MOS/LPIPS) [30].

4. Experimental Process and Analysis of Results

4.1. Experimental Data

The infrared dataset used in this paper is sourced from the open-source infrared dataset within the AiRui database, as shown as Table 1. Data Set Classification and Data Tensors. This dataset comprises 7000 high-resolution infrared images spanning diverse scenarios. Given its moderately small scale, the dataset was partitioned into training, validation, and test sets at a ratio of 7:1.5:1.5. The dataset used for training is a hybrid dataset. One portion consists of a self-built dataset derived from raw videos captured using infrared cameras across various scenarios (including outdoor, daytime, and nighttime conditions). These videos underwent keyframe extraction, denoising, and non-uniformity correction before being filtered to obtain infrared images. The other portion utilizes urban scene images provided by the Airo Red open-source platform (primarily at a resolution of 640×512). This balance ensures sufficient data for reliable evaluation in the validation and test sets while maximizing the training set size to enable the model to learn robust features. Since the training model incorporates random degradation processing, no additional image preprocessing is performed outside the training process.

Table 1. Data Set Classification and Data Tensors.

Airui Dataset	Tensor
Training set	4900
Val set	1050
Test set	1050

4.2. Experimental Equipment Parameters and Evaluation Criteria

The experimental development language used in this paper is composed of Python 3.11.9 and Matlab 2019b, with the development platform being PyCharm 2024.3. The hardware specifications of the desktop computer used are as follows: CPU is i5-14600KF from China, GPU is NVIDIA RTX 4060 Ti, and memory is 32 GB; the imaging device used is the DJI Mini 2 drone from China.

4.2.1. Peak Signal-to-Noise Ratio

The Peak Signal-to-Noise Ratio (PSNR) is a metric used to assess the quality of images or videos, representing the peak signal-to-noise ratio between the original signal and the distorted signal. MSE represents the average of the squares of the differences between the two signals. PSNR is expressed by mapping MSE to a logarithmic scale to better align with the human perceptual system’s sensitivity to signal differences. MAX represents the maximum possible pixel value in the image.

The formula for calculating PSNR is as follows:

$$PSNR = 10 \cdot \log_{10}((MAX^2)/MSE) \quad (14)$$

4.2.2. Structural Similarity Index

SSIM (Structural Similarity Index) is a metric used to measure the similarity of image quality. The calculation of SSIM is based on three components: brightness similarity, contrast similarity, and structural similarity. The results of these three components are combined into a comprehensive SSIM index, which ranges from 0 to 1, where 1 indicates that the two images are identical and 0 indicates that the two images are completely dissimilar. The formula for calculating SSIM is as follows:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (15)$$

In this context, x and y represent two images, while $l(x, y)$, $c(x, y)$, and $s(x, y)$ denote brightness similarity, contrast similarity, and structural similarity, respectively. α , β , and γ are weighting parameters used to adjust the relative importance of different components.

4.2.3. Perceptual Similarity

Perceptual similarity LPIPS, short for Learned Perceptual Image Patch Similarity, is a metric used to measure the perceptual similarity between two images. Unlike traditional PSNR and SSIM, LPIPS is learned through deep learning methods, enabling it to better simulate human visual perception. The calculation process involves first dividing the two images into multiple local patches, using a pre-trained deep neural network model to extract features from each patch, calculating the distance between corresponding patches in a specific space, aggregating the distances of all patches, and then calculating the score. The formula is as follows:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h, w} w_l \odot \|\hat{y}_{lhw}^x\|^2 \quad (16)$$

$d(x, x_0)$ represents the perceived distance or perceived difference value. x represents the image to be compared (e.g., the generated image or the distorted image). x_0 represents the reference image (original image). l represents the l th layer of the pre-trained convolutional neural network. H_l represents the height of the l th layer feature map. W_l represents the width of the l th layer feature map. y_{lhw}^x represents the feature vector at the spatial coordinates (h, w) in the l th layer feature map of image x . $y_{lhw}^{x_0}$ represents the feature vector at the same position in the l th layer feature map for image x_0 . w_l represents the learnable weight vector for layer l . $\odot$ denotes element-wise multiplication. The weight vector w_l is multiplied in an element-wise manner with the normalized feature differences, thereby weighting the differences for each channel. $\|\cdot\|^2$ The square of the L2 norm of the weighted feature difference vector is calculated. The average differences across multiple layers is determined to obtain the final LPIPS distance.

4.3. Experimental Results

4.3.1. Subjective Evaluation

In terms of subjective evaluation, we compared the reconstructed images and their residual local images obtained using the bicubic interpolation method, the SRCNN algorithm, SRGAN and the improved algorithm. From the example shown in Figure 6, their reconstructed images, and the corresponding residual local images and original local images, it can be observed that the bicubic interpolation method yields unsatisfactory reconstruction results, with severe loss of high-frequency textures and information in its local images; SRCNN, as the pioneering algorithm for super-resolution image reconstruction, demonstrates significantly improved reconstruction performance compared to the bicubic interpolation method, with enhanced capability in capturing high-frequency textures and

information. The reconstructed images and residual local images of the proposed algorithm are more closely aligned with the original images and their residual local images. Compared to the reconstructed images of the previous two algorithms, the proposed algorithm exhibits clearer edges and richer texture details. Figures 6–8 show the original image, interpolation method, SRCNN, SRGAN and improved algorithm reconstruction images and their residual local images at $\times 4$ magnification, along with their reconstruction quality curves, as shown in Figure 9.

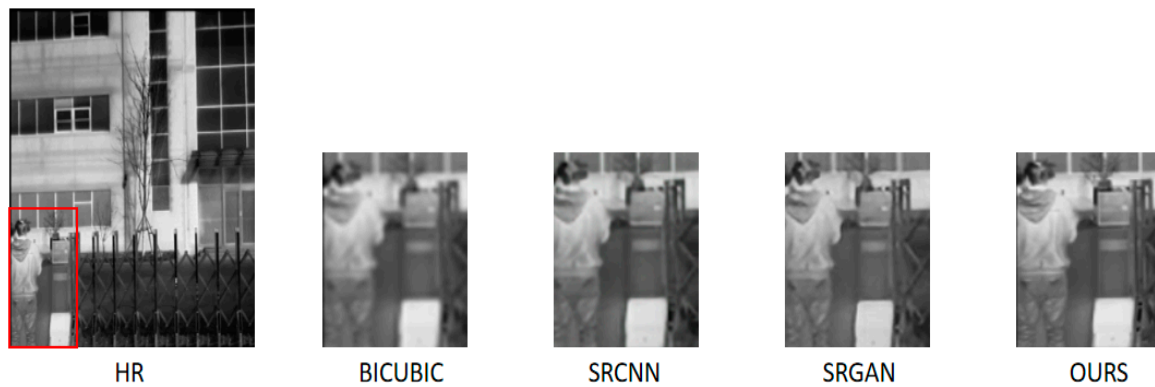


Figure 6. Infrared image test example 1 (partial human figure reconstruction).

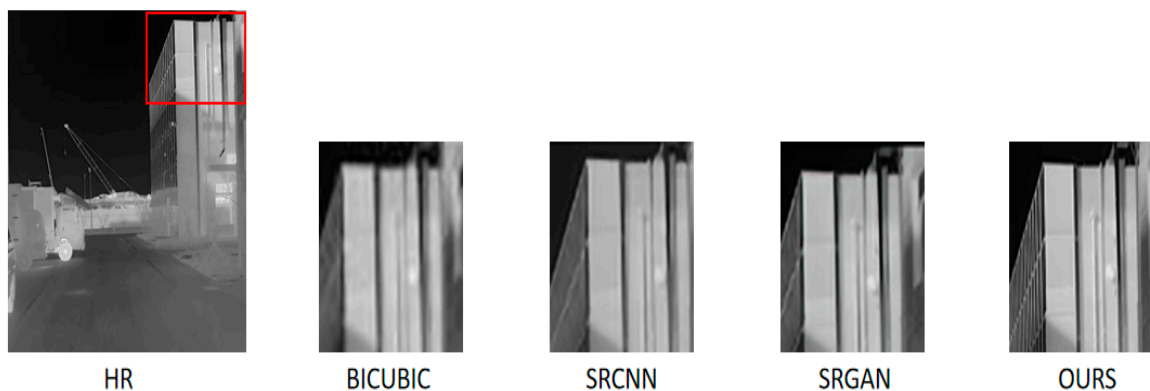


Figure 7. Infrared image test example 2 (partial building image reconstruction).

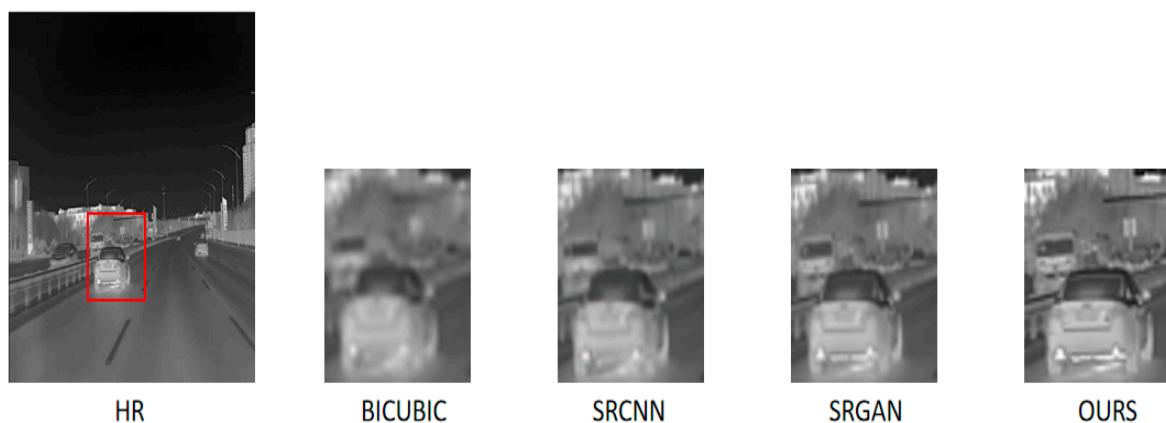


Figure 8. Infrared image test example 3 (vehicle partial image reconstruction).

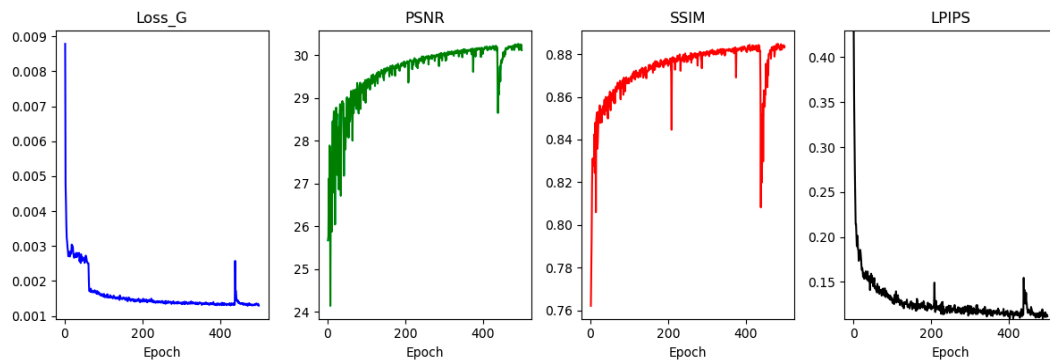


Figure 9. Fourfold reconstruction index curve diagram.

4.3.2. Objective Evaluation

Comparing $2\times$ Table 2, $4\times$ Table 3, and $8\times$ Table 4 super-resolution reconstructed images with different algorithm models (Bicubic, SRCNN, SRGAN), Table 1 clearly shows that under $2\times$ super-resolution, the improved network model achieves certain advantages in terms of PSNR, SSIM values, and Lpips. Compared to $2\times$ super-resolution, $4\times$ super-resolution still exhibits a significant numerical advantage, while the advantage further narrows at $8\times$ super-resolution reconstruction.

Table 2. Comparison of $\times 2$ reconstruction model.

Scale	Methods	Enter Dimensions	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
$\times 2$	Bicubic	320×256	28.2404	0.8455	0.3264
	SRCNN		29.7601	0.7954	0.2188
	SRGAN		30.7503	0.8826	0.1233
	Ours		31.4581	0.8981	0.1174

Note: Bold indicates optimal data. The table below follows the same convention.

Table 3. Comparison of $\times 4$ reconstruction model.

Scale	Methods	Enter Dimensions	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
$\times 4$	Bicubic	160×128	26.7863	0.8136	0.3773
	SRCNN		27.9834	0.8236	0.2642
	SRGAN		28.6215	0.8392	0.1822
	Ours		29.8323	0.8523	0.1308

Table 4. Comparison of $\times 8$ reconstruction model.

Scale	Methods	Enter Dimensions	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
$\times 8$	Bicubic	80×64	25.3425	0.7936	0.4151
	SRCNN		26.6434	0.7951	0.2894
	SRGAN		26.6221	0.8035	0.2739
	Ours		26.7531	0.8111	0.2673

5. Conclusions

This paper addresses the challenges of low resolution, high noise, and complex degradation processes in infrared images by proposing a super-resolution reconstruction method based on stochastic degradation modeling and Generative Adversarial Networks (GANs). The core innovation lies in modeling key degradation parameters in infrared

imaging—such as blur kernels and noise distributions—as random variables. By simulating diverse degradation processes closer to real-world scenarios through probabilistic distributions, the method significantly enhances the model’s generalization capability under complex degradation conditions. Experiments on the Airo infrared dataset and custom datasets demonstrate that the proposed method achieves outstanding performance in $2\times$, $4\times$, and $8\times$ super-resolution reconstruction tasks. Specifically, in $4\times$ super-resolution reconstruction, compared to three typical deep learning algorithms, our algorithm improves the Peak Signal-to-Noise Ratio (PSNR) by 3.046 dB, 1.8489 dB, and 0.2108 dB, respectively, and the Structural Similarity Index (SSIM) by 0.0387 (4.76%), 0.0287 (3.48%), and 0.0131 (1.56%), respectively, while perceptual similarity decreased by 0.2465, 0.13344, and 0.0514 (lower values indicate better perceptual quality). Under $8\times$ super-resolution reconstruction, information bottlenecks, limitations of the generator architecture, and the extreme nature of degenerate coupling may all become factors restricting further improvements in model performance. The primary contribution of this study is that it proposes a random degradation model that effectively enhances degradation diversity, offering a novel solution for super-resolution reconstruction of infrared images in complex scenes. Although the algorithm achieves favorable results, its generalization capability under extreme degradation scenarios remains subject to improvement. Future work will focus on model optimization, reducing reliance on large-scale training data, and exploring the method’s potential applications in broader tasks such as image denoising and enhancement.

Author Contributions: Conceptualization, L.Y. and K.H.; methodology, K.H. and H.G.; software, K.H.; validation, Z.Z. and S.G.; formal analysis, H.G.; investigation, S.G.; data curation, K.H.; writing—original draft preparation, K.H.; writing—review and editing, L.Y.; visualization, K.H.; supervision, L.Y.; project administration, L.Y.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported in part by the Key Research and Development Program of Shaanxi [2025SF-YBXM-287], the “Scientist + Engineer” Team Construction Project of Qinchuangyuan in Shanxi Province [2025QCY-KXJ-049], the Shanxi Provincial Science and Technology Plan Project [2025JC-YBMS-082], and the Shanxi Provincial Science and Technology Plan Project [2025JC-YBMS-744].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

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

References

1. Ren, Z.; Zhao, J.; Wang, C.; Ma, X.; Lou, Y.; Wang, P. Research on Key Technologies of Super-Resolution Reconstruction of Medium and Long Wave Maritime Infrared Image. *Appl. Sci.* **2022**, *12*, 10871. [\[CrossRef\]](#)
2. Chen, H.; He, X.; Qing, L.; Wu, Y.; Ren, C.; Sheriff, R.E.; Zhu, C. Real-world single image super-resolution: A brief review. *Inf. Fusion* **2022**, *79*, 124–145. [\[CrossRef\]](#)
3. Li, X.; Liu, R.; Yang, Y. Infrared image super-resolution reconstruction based on residual fast fourier transform. *Multimed. Tools Appl.* **2024**, *84*, 6805–6823. [\[CrossRef\]](#)
4. Su, H.; Li, Y.; Xu, Y.; Fu, X.; Liu, S. A review of deep-learning-based super-resolution: From methods to applications. *Pattern Recognit.* **2025**, *157*, 110935. [\[CrossRef\]](#)
5. Izumi, T.; Amagasaki, M.; Ishida, K.; Kiyama, M. Super-resolution of sea surface temperature with convolutional neural network-and generative adversarial network-based methods. *J. Water Clim. Change* **2022**, *13*, 1673–1683. [\[CrossRef\]](#)
6. Dong, Y.; Li, Y.; Li, Z. Research on Detection and Recognition Technology of a Visible and Infrared Dim and Small Target Based on Deep Learning. *Electronics* **2023**, *12*, 1732. [\[CrossRef\]](#)

7. Fan, K.; Hong, K.; Li, F. Infrared Image Super-Resolution via Progressive Compact Distillation Network. *Electronics* **2021**, *10*, 3107. [CrossRef]
8. Du, Y.-B.; Sun, H.-M.; Zhang, B.; Cui, Z.; Jia, R.-S. A multi-scale mixed convolutional network for infrared image super-resolution reconstruction. *Multimed. Tools Appl.* **2023**, *82*, 41895–41911. [CrossRef]
9. Yang, C.; Lu, G. Unsupervised image blind super resolution via real degradation feature learning. *IET Comput. Vis.* **2024**, *18*, 485–498. [CrossRef]
10. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516117. [CrossRef]
11. Li, J.; Zheng, K.; Li, Z.; Gao, L.; Jia, X. X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5518317. [CrossRef]
12. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 105–114.
13. Wu, C.; Jing, Y. Unsupervised super resolution using dual contrastive learning. *Neurocomputing* **2025**, *630*, 129649. [CrossRef]
14. Dempsey, G.T.; Vaughan, J.C.; Chen, K.H.; Bates, M.; Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **2011**, *8*, 1027–1036. [CrossRef]
15. Wang, Y.; Xing, K.; Wang, B.; Hai, S.; Li, J.; Deng, M. Image super-resolution reconstruction method based on residual mechanism. *J. Electron. Imaging* **2022**, *31*, 033010. [CrossRef]
16. Yang, Y.; Ma, Z.; Zhou, X.; Zheng, F.; Huang, H. Spatial Adaptive and Frequency Fusion Network for Super-Resolution of Single Remote Sensing Images. *Opt. Precis. Eng.* **2025**, *33*, 1238–1258. [CrossRef]
17. Liu, Y.; Zhu, J.; Zheng, D.; Zhang, D. A lightweight low-light image enhancement method based on multi-attention feature fusion. *J. Data Acquis. Process.* **2025**, 1–18. Available online: <https://link.cnki.net/urlid/32.1367.TN.20250711.1339.037> (accessed on 2 February 2026).
18. Li, M.; Hao, Y.; Xu, S. Research on infrared and visible light image fusion algorithms. *Firepower Command Control* **2025**, *50*, 165–177. [CrossRef]
19. Zhu, X.; Jiang, N. Multi-dimensional Attention Feature Fusion Image Dehazing Algorithm. *Adv. Lasers Optoelectron.* **2025**, *62*, 385–395. [CrossRef]
20. Chauhan, K.; Patel, S.N.; Kumhar, M.; Bhatia, J.; Tanwar, S.; Davidson, I.E.; Mazibuko, T.F.; Sharma, R. Deep learning-based single-image super-resolution: A comprehensive review. *IEEE Access* **2023**, *11*, 21811–21830. [CrossRef]
21. Zhou, Y.; Liu, Y.; Yuan, L.; Chen, Q.; Gu, G.; Sui, X. Real-infraredSR: Real-world infrared image super-resolution via thermal imager. *Opt. Express* **2023**, *31*, 36171–36187. [CrossRef]
22. Zhang, H.; Hu, Y. Lightweight dynamic attention network for single thermal image super-resolution. *Signal Image Video Process.* **2024**, *18*, 2195–2206. [CrossRef]
23. Hao, Y.; Yu, F. Super-resolution degradation model: Converting high-resolution datasets to optical zoom datasets. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 6374–6389. [CrossRef]
24. Jiang, N.; Zhao, W.; Wang, H.; Luo, H.; Chen, Z.; Zhu, J. Lightweight super-resolution generative adversarial network for SAR images. *Remote Sens.* **2024**, *16*, 1788. [CrossRef]
25. Xu, X.; Ma, Y.; Sun, W.; Yang, M.-H. Exploiting raw images for real-scene super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1905–1921. [CrossRef]
26. Xu, Z.; Gao, J.; Wang, X.; Kang, C. Inf-OSRGAN: Optimized Blind Super-Resolution GAN for Infrared Images. *Appl. Sci.* **2024**, *14*, 7620. [CrossRef]
27. Daihong, J.; Sai, Z.; Lei, D.; Yueming, D. Multi-scale generative adversarial network for image super-resolution. *Soft Comput.* **2022**, *26*, 3631–3641. [CrossRef]
28. Chen, W.; Qu, S.; Luo, L.; Lu, Y. Dual Attention Fusion Enhancement Network for Lightweight Remote-Sensing Image Super-Resolution. *Remote Sens.* **2025**, *17*, 1078. [CrossRef]
29. Han, X.; Wang, L.; Wang, X.; Zhang, P.; Xu, H. A multi-scale recursive attention feature fusion network for image super-resolution reconstruction algorithm. *Sensors* **2023**, *23*, 9458. [CrossRef]
30. Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Image quality assessment: Unifying structure and texture similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2567–2581. [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.