Adaptive Multi-Scale Fusion Blind Deblurred Generative Adversarial Network Method for Sharpening Image Data

Baoyu Zhu [1,2,3,†], **Qunbo Lv** [1,2,3,†] **and Zheng Tan** [1,3,*]

1   Aerospace Information Research Institute, Chinese Academy of Sciences, No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China
2   School of Optoelectronics, University of Chinese Academy of Sciences, No.19(A) Yuquan Road, Shijingshan District, Beijing 100049, China
3   Department of Key Laboratory of Computational Optical Imagine Technology, CAS, No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China
*   Correspondence: tanzheng@aircas.ac.cn; Tel.: +86-1591-063-9123
†   These authors contributed equally to this work.

Abstract: Drone and aerial remote sensing images are widely used, but their imaging environment is complex and prone to image blurring. Existing CNN deblurring algorithms usually use multi-scale fusion to extract features in order to make full use of aerial remote sensing blurred image information, but images with different degrees of blurring use the same weights, leading to increasing errors in the feature fusion process layer by layer. Based on the physical properties of image blurring, this paper proposes an adaptive multi-scale fusion blind deblurred generative adversarial network (AMD-GAN), which innovatively applies the degree of image blurring to guide the adjustment of the weights of multi-scale fusion, effectively suppressing the errors in the multi-scale fusion process and enhancing the interpretability of the feature layer. The research work in this paper reveals the necessity and effectiveness of a priori information on image blurring levels in image deblurring tasks. By studying and exploring the image blurring levels, the network model focuses more on the basic physical features of image blurring. Meanwhile, this paper proposes an image blurring degree description model, which can effectively represent the blurring degree of aerial remote sensing images. The comparison experiments show that the algorithm in this paper can effectively recover images with different degrees of blur, obtain high-quality images with clear texture details, outperform the comparison algorithm in both qualitative and quantitative evaluation, and can effectively improve the object detection performance of blurred aerial remote sensing images. Moreover, the average PSNR of this paper's algorithm tested on the publicly available dataset RealBlur-R reached 41.02 dB, surpassing the latest SOTA algorithm.

Keywords: drone and aerial remote sensing; image deblurring; generative adversarial networks; multi-scale; image blur level; object detection; deep learning

1. Introduction

Drone and aerial remote sensing images have been widely used in many fields such as land and mineral resource management and monitoring, traffic and road network safety monitoring, and geological disaster early warning system and national defense system construction [1–5]. However, as the geometric resolution of aerospace optical cameras continues to increase, high-speed spatial motion and random vibration of the camera platform can cause image shift blurring [6,7], and rapid motion of the target can create additional blurring [8]. In addition, factors such as the spatially varying characteristics of the blur kernel, the imaging depth of field, and detector noise can also increase the complexity of image blurring to varying degrees [9], resulting in degraded image quality. This not only severely affects the visualization of the images and reduces the perceptual

quality, but also has a significant impact on all levels of visual tasks and increases the difficulty of applying analysis tasks.

Since image motion blur is difficult to avoid and drone aerial images are highly motion blurred [10], the study of deblurring is very important, especially for tasks such as object detection and object recognition, where image blurring reduces metrics such as precision and recall [11].

Aerial remote sensing cameras usually produce image shift blur because the CMOS or CCD detector is displaced by more than one image element with respect to the ground target during the exposure time. The blur kernel is an unknown quantity and requires blind deblurring. Image non-uniform blur models [12] are usually modelled as:

$$B = S * k(M) + N, \tag{1}$$

where B is the blurred image, S is the corresponding clear image, $k(M)$ denotes the unknown blur kernel, M denotes a sparse matrix in which each row contains a local blur kernel, * is the discrete convolution, and N is the noise.

Blind image deblurring algorithms can be divided into two categories: those based on deconvolution and those based on neural network operations. The deconvolution-based image deblurring algorithm uses methods such as the natural image prior [13–16] to estimate the blur kernel and perform a deconvolution operation on the blurred image to restore it to a clear image. However, accurate estimation of the blur kernel is a difficult task, and the algorithm modelling relies on a priori constraints. The algorithm is computationally intensive and vulnerable to noise and computational errors, which severely the overall performance of the algorithm. Consequently, deconvolution algorithms are difficult to apply to complex deblurring tasks [17]. Deep learning algorithms based on convolutional neural networks (CNNs) have now achieved remarkable results in image deblurring tasks [18,19]. The algorithms use pairs of blurred and clear images to train a deblurring model that implicitly learns knowledge of the mapping between blurred and clear images through large-scale natural images to achieve recovery from blurred to clear images. Deep learning algorithms widely use coarse-to-fine network architectures to extract information from blurred images and improve the deblurring performance of models through multi-scale fusion [20]. In particular, compared to other types of neural network algorithms such as DeepDeblur [21] and STFAN [22], generative adversarial networks (GAN) [23] are able to more reliably preserve texture details in images after multi-scale fusion, obtaining results closer to the real image and recovering images more clearly [9,24,25]. Multiscale methods can simultaneously use multiple layers of features to deal with different degrees of blurring in images [20,21,26]. However, feature misalignment inevitably occurs in the multi-scale fusion process, leading to fusion errors and reducing the deblurring effect.

In a multi-scale deblurring network model, as the network model continues to deepen, the feature map size decreases, the pixel offset corresponding to the feature map becomes smaller and smaller, reflecting less and less detailed information [27], and the distribution of information in the feature layer becomes more and more concentrated. The process of multi-scale fusion can be described as a reorganization of the information in the shallow high-resolution feature layers with the deeper low-resolution feature layers. There are two main types of recombination: 1. low-resolution feature maps are up-sampled and fused with high-resolution feature maps, and 2. high-resolution feature maps are down-sampled and fused with low-resolution feature maps [28,29]. The existing CNN algorithms do not consider the possible inconsistency of information distribution between the low-resolution feature maps and the high-resolution feature maps. If the high-resolution feature layers with large pixel offsets are fused with the low-resolution feature layers with small pixel offsets, the problem of feature misalignment will inevitably arise, resulting in large fusion errors, and such errors will be superimposed layer by layer in the network structure from coarse to fine. Currently, network models are trained in a data-driven manner [30], and the final multiscale fusion scheme is a balanced weighting strategy that fits the training

dataset [31,32]. Using the same multiscale weights for all images with different degrees of blurring also causes the fusion error to increase layer by layer.

To solve the above problems, this paper combines the physical characteristics of image blurring with network models to propose an improved multi-scale model scheme more suitable for image deblurring tasks. The physical properties of image blurring are that the energy from the point source is distributed diffusely in the image plane according to the trajectory of the motion blur kernel; the greater the energy diffusion, the greater the pixel shift, and the more blurred the image. Therefore, for multi-scale feature fusion, images with a low level of blurring can be recovered by using shallow high-resolution feature layers due to their small energy spread, while if too many deep low-resolution feature layers are fused, the recovered image will become blurred due to fusion errors, and it is more appropriate to use shallow high-resolution feature layers for fusion to reduce fusion errors. Similarly, images with a high degree of blurring have a large energy spread. The higher resolution feature layers with shallower layers do not reflect a sufficiently concentrated energy distribution with large pixel shifts due to image blur diffusion, while the lower resolution feature layers with deeper layers reflect a more concentrated energy distribution with small pixel shifts. In this case, it is more appropriate to use the individual low-resolution feature layers with deeper layers for fusion to reduce the pixel offset and reduce the fusion error.

Based on the above analysis of the physical characteristics of image blurring and the adaptability of the network model, this paper proposes an adaptive multi-scale fusion blind deblurred generative adversarial network (AMD-GAN) that can perform high-quality restoration of remote sensing blurred images. The model combines the physical characteristics of image blurring, makes full use of the characteristics of different blurred images with different energy diffusion and different information distribution, and uses the image blurring degree to guide the network model for multi-scale fusion of feature maps with adaptive weights at different scales, which can effectively suppress the alignment errors in the process of multi-scale fusion. This is equivalent to enhancing the interpretability of the multi-scale network structure for blurry image feature extraction and fusion at the physical level. Unlike the estimation of blur kernels, when evaluating the degree of image blur, this paper treats the estimation of the degree of blur as the estimation of the pixel offset in the image, and the remote sensing image with different blur levels drives the image blur degree description model for training and completes the task of image blur estimation.

The innovations in this paper are summarized as follows:

1. This paper combines the physical property of the image blurring degree with multi-scale feature fusion for image deblurring processing and proposes an adaptive multi-scale feature fusion method for high-quality restoration of drone and aerial remote sensing blurred images, which effectively suppresses errors in the multi-scale fusion process, enhances the interpretability of the deblurring model for blurred image feature extraction and fusion, makes the function model represented by the network model more physically meaningful for image deblurring, and enhances the reliability of the model. To the best of our knowledge, this is the first application of a multi-scale fusion strategy to adjust the degree of image blur.
2. This paper proposes a model for describing the degree of image blur. This paper provides some research and exploration on the degree of image blur, using classification algorithms in deep learning for image blur estimation, revealing the necessity and effectiveness of image blur information in deblurring tasks.
3. This paper explores the impact of drone and aerial remote sensing image blurring and image deblurring on object detection tasks, where image deblurring models that are not adapted to the image being detected can lead to a reduction in object detection accuracy.
4. In this paper, we construct our own aerial remote sensing image dataset with different levels of blur, including both camera multi-exposure time shots and multi-level random motion blur kernel simulations, at both 1 M and 2 M resolutions.

2. Related Work

2.1. Image-Deblurring-Based Deep Learning

Existing deep learning approaches focus on training deblurred models using pairs of blurred and clear images, learning knowledge implicitly through large-scale data. For example, DeepDeblur [21] pioneered techniques to recover blurred images in a trainable end-to-end manner, proposing a coarse-to-fine processing pipeline that achieves better performance by stacking multiple sub-networks. DeblurGAN-v2 [9] constructs a new cGAN framework that introduces a feature pyramid network (FPN) with multi-scale feature fusion. DMPHN [33] proposed a novel stacking paradigm for deblurring, where increasing the depth in the horizontal direction (stacking multiple network models) achieves better deblurring results. MPRNet [34] proposed a multi-stage architecture network model for collaborative design, decomposing the entire recovery process to progressively learn the degraded features of the inputs. MAXIM [35] proposes a multi-axis MLP-based architecture where each module has a global/local perceptual field, improving the learning capability of the model. Ref-MFFDN [36] proposed a new reference-based multi-level feature fusion deblurring network for remote sensing image deblurring, which extracts textures from clear reference images at different moments of the same location to help recover blurred images. NSRN [37] proposed a noise-suppression-based restoration network for turbulent degraded images. SR-DeblurUGAN [38] considers the differences in the levels of features extracted from different perceptual layers and uses a generative adversarial network based on weighted perceptual loss to deblur the drone images. Inspired by the success of the vision transformer (ViT) [39] on several computer vision tasks, several transformer-based models have been developed [40–42] with significant performance. However, transformer-based models for image deblurring networks are computationally complex.

Currently, the network model uses the same multi-scale fusion weights for all input images.

2.2. Multi-Scale Fusion

The idea of multi-scale fusion is widely used in the overall construction of various neural network model architectures [34,36,43], and in particular, multi-scale modules have been used as a plug-and-play base module for various computer vision tasks.

FPN [44] can exploit both the high resolution of shallow features and the high semantic information of deep features, including bottom-up, top-down and lateral connection structures, to achieve the desired effect by fusing these different layers of features. PANet [45] is an extension of FPN, additionally augmented by adding a bottom-up route after FPN. Bi-FPN [46] proposed a weighted bi-directional feature pyramid network in EfficientDet that allows simple and fast multi-scale feature fusion. NAS-FPN [47] takes advantage of neural architecture search, using augmented learning to select the best cross-connections and learn a better pyramidal network architecture for object detection features with high accuracy and latency trade-offs.

Coarse-to-fine multi-scale fusion methods are widely used in deep-learning-based image deblurring network model architectures, where multi-scale input images and feature extraction sub-networks are usually stacked and the resolution of the images and sub-networks is gradually increased from the bottom to the top [20]. DeblurGAN-v2 [9] constructs a feature pyramid network (FPN)-based deblurring framework that introduces a multi-scale feature fusion FPN in the generator, containing five scales of feature outputs that are upsampled to the original input quarter size and stitched together as a whole (containing multi-scale information), followed by two upsampling modules to recover to the original image size. DBCPeNet [48] proposes a multi-scale network architecture for full-scale utilization of images, which can make full use of images of different scales and maximize the information flow between different scales by training from coarse to fine and fine to coarse simultaneously, resulting in better recovery performance. MIMO-UNet [20] revisits the coarse-to-fine scheme, using a single encoder with multiple inputs and a single decoder with multiple outputs, to propose a new deblurring network, called multi-input multi-output UNet, which can handle multi-scale blurring with low computational complexity.

These model architectures all use a similar coarse-to-fine approach in order to improve image deblurring performance. The coarse-to-fine network design principle combined with multi-scale networks has been shown to be an effective image deblurring method. However, these network models do not consider the effect of image blurring on the multi-scale fusion approach, and the same multi-scale fusion weights are used for input images with different blurring levels. The blur kernel diffusion is smaller for images with lower blur degrees and larger for images with higher blur degrees. If the input images are not differentiated and features from shallow and deeper layers of images with different blur degrees are fused according to uniform weights, fusion errors will inevitably occur due to unaligned features.

Therefore, the multi-scale network structure using uniform weights makes it difficult to accurately describe image features with different degrees of blurring, leading to poor interpretability of existing multi-scale models for blurred features, which is the main source of error in the models.

2.3. Description of the Degree of Image Blur

There are three main types of image blur description: fully referenced, semi-referenced and unreferenced, with the unreferenced image description method being the most applicable.

There are many methods of evaluating the degree of image blur in traditional image processing, using information such as the image gradient and entropy to measure the degree of image blur, and the following algorithms are commonly used to describe the degree of blurring of images:

Brenner gradient function, Laplacian gradient function, SMD (grey scale difference) function, SMD2 (product of grey scale differences) function, variance function, energy function (energy gradient), Vollath function, and entropy function.

The Brenner gradient function is the simplest gradient evaluation function, which simply calculates the square of the difference between the grey levels of two adjacent pixels, and is defined as follows:

$$D(f) = \sum_y \sum_x |f(x+2,y) - f(x,y)|^2, \tag{2}$$

The definition of image blur based on the Laplacian gradient function is as follows:

$$D(f) = \sum_y \sum_x G(x,y) \qquad (G(x,y) > T), \tag{3}$$

$$L = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}, \tag{4}$$

where $G(x,y)$ is the convolution of the Laplacian operator at pixel point (x,y).

The SMD2 function is to multiply the difference between two grey levels in each pixel field of the image and then accumulate them pixel by pixel. In a blurred image, the difference between the adjacent values of the pixel points will be small, so the larger the accumulation value, the clearer the image, which is calculated as shown below:

$$D(f) = \sum_y \sum_x |f(x,y) - f(x+1,y)| * |f(x,y) - f(x,y+1)|, \tag{5}$$

Information such as the kernel size of the blur kernel can also be used to determine how blurred an image is, but any method that can accurately estimate the blur kernel is complex.

The degree of image blur is also an important indicator of image quality [49], and Wang, et al [50] use image roughness as a measure of image blur when assessing image quality. The smaller the image roughness, the better the image quality, defined as:

$$R = \frac{\|h * e\|_1 + \|h^T * e\|_1}{\|e\|_1}, \tag{6}$$

where e is the evaluation image and h is the mask.

3. Methods

As shown in Figure 1, the adaptive multi-scale single image blind deblurring network AMD-GAN designed in this paper consists of two main parts: 1. a generator (AMD GAN_G) and 2. A discriminator (AMD GAN_D). The generator generates a clear image by processing the blurred image and the discriminator discriminates the clear image generated by the generator. MAP0–MAP3 correspond to four different resolutions of blurred images in turn, BMD is the model proposed in this paper to describe the degree of image blur, and AMS Fusion represents the adaptive multi-scale fusion strategy proposed in this paper.
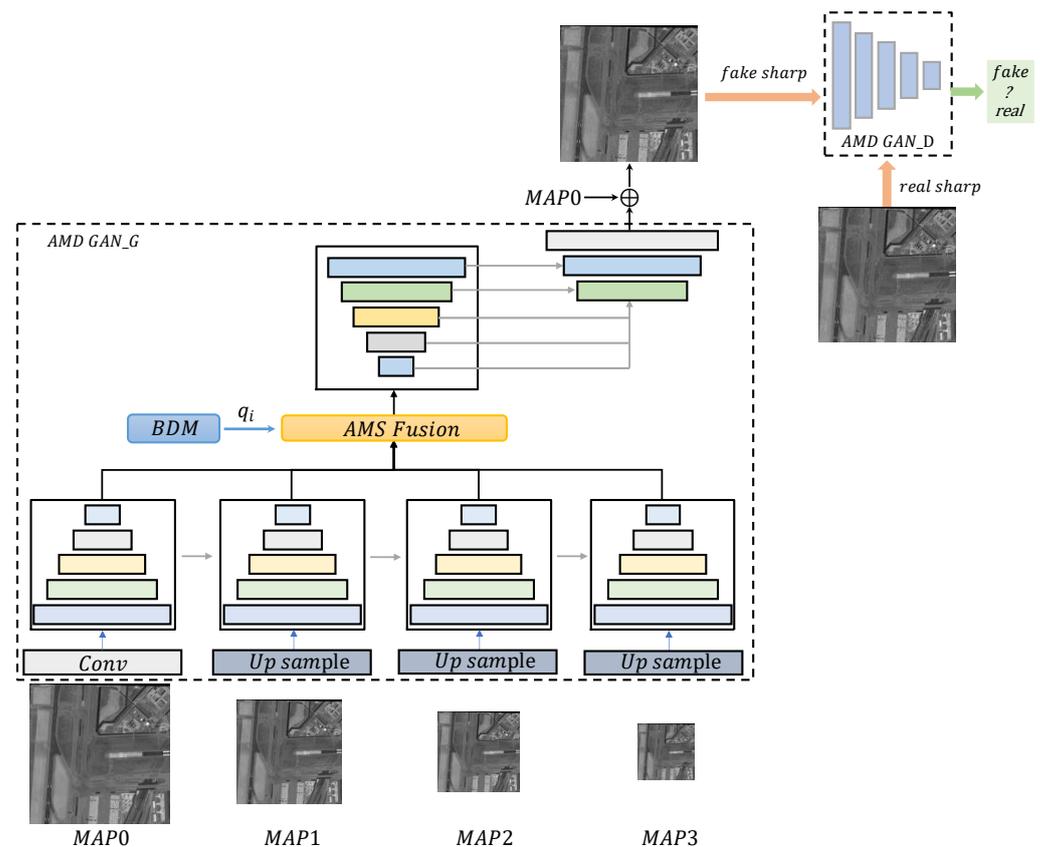
Figure 1. The AMD-GAN framework.

The image deblurring process in this paper is shown in Figure 2, where the blurred image is represented by the blurred degree description model (BDM). If the degree of blurring is less than the threshold, it is not processed and is output directly; if the degree of blurring is greater than the threshold, it is deblurred. In the process of deblurring, AMD-GAN uses the characteristics of different blurred images with different information distribution and different energy diffusion to adaptively adjust the weights of the model's multi-scale fusion according to the degree of image blurring. Blurs with small image shifts use more shallow high-resolution layer features to retain more detailed texture; blurs with large image shifts use more deeper low-resolution layers to reduce the spread of energy distribution. The adaptive fusion weighting strategy can effectively suppress the alignment error of layer-by-layer superposition in the multi-scale fusion process and improve the deblurring effect. Moreover, the processing flow of this paper is more suitable for the application scenario where multiple blurred images coexist with clear images in the real environment of drone and aerial remote sensing.
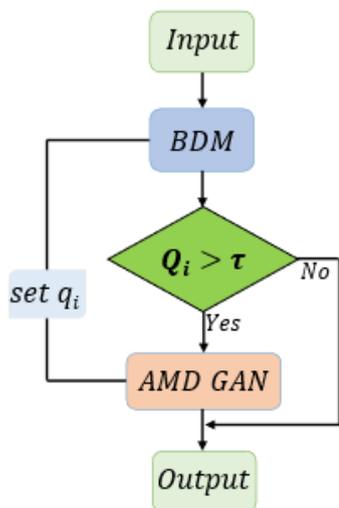
Figure 2. The AMD-GAN framework.

3.1. Blurred Degree Description Model (BDM)

Based on the idea of adaptation to image deblurring tasks, this paper investigates models that can describe the degree of image blurring and combines them with existing multi-scale network structures to enable deblurring models to account for the physical characteristics of blurred images, a technical way of solving existing problems.

In this paper, the deblurring model requires only the blurring offset degree of an image. Inspired by the image classification task, the image blur degree description task is considered as a classification task for the image blur degree. Drawing on the main idea of EfficientNet [51], an image classification network, the network is suitably extended by combining the width, depth, and resolution. Considering the speed and model size, MobileNet [52] was chosen as the baseline model to design the image blur description model BDM (blurred degree description model). Using multi-resolution hybrid training, the data-driven model learns the offset of the blurred image, reducing the influence of other factors on the model's determination of the blur level. To our knowledge, this is the first attempt to use a classification algorithm from deep learning for image blur estimation and to apply it to a deblurring task.

A schematic diagram of the structure of the image blur description model (BDM) is shown in Figure 3, where MBConv is an alternating hybrid module of MBConv1 and MBConv2.
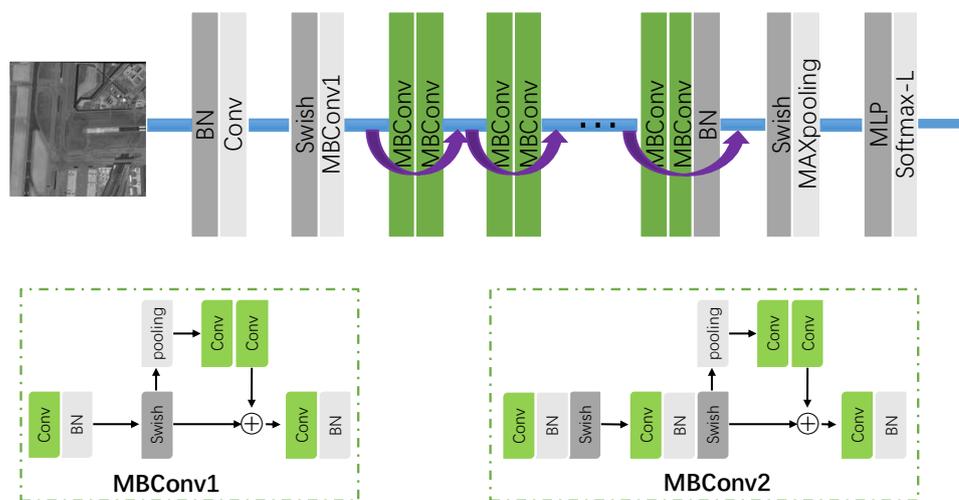
Figure 3. BDM (blurred degree description model) structure diagram.

The model uses the Swish activation function, which helps to prevent the gradient from approaching zero and leading to saturation during slow training, and provides superior performance compared to other activation functions. Currently, Softmax activation functions are commonly used for multi-category classification models. In order to achieve multi-blur category training, the output results in continuous blur degree values. Based on Softmax, we propose that the activation function Softmax-L for the image blurring level can be defined as follows:

$$Q = \sum_i w_i * \frac{e^{x_i}}{\sum x}, \tag{7}$$

where w is the weighting factor for each blur level, x is the degree of blurring of the image, and the interval of the output blurring degree is [0, 1].

3.2. Adaptive Multi-Scale Model

DeblurGAN-v2 [9] introduces feature pyramid networks (FPNs) [44] into the field of image deblurring, where FPNs use top-down and lateral connectivity structures to combine lower resolution features with higher resolution features.

However, blurred images in real scenes are complex and variable, and the degree of blurring varies dramatically. Although FPN is simple and effective, it may not be the best architectural design. For images with different levels of blurring, bottom-up and top-down convolution by FPN alone is not sufficient to extract image features. Moreover, the sufficiently coarse image obtained after multiple downsampling of the blurred image is approximately the low-resolution image corresponding to the clear image [32]. Therefore, this paper designs an image multi-scale feature pyramid network (MFPN) based on FPN, and the sketch of the structure is shown in (a) in Figure 4. MAP0-MAP3 corresponds to four different resolutions of blurred images in order to improve the feature extraction and fusion of images with different blurring levels in order to achieve better deblurring results. First, the blurred images are gradually downsampled using bi-triple interpolation at a ratio of 1/2, and four scaled pairs of images are generated with resolutions of (H × W, H/2 × W/2, H/4 × W/4, and H/8 × W/8). Each scaled blurred image is then used as four inputs to the multi-scale model, with the subsequent network structure shown in Figure 1, and a clear image at the original resolution is considered the final output.
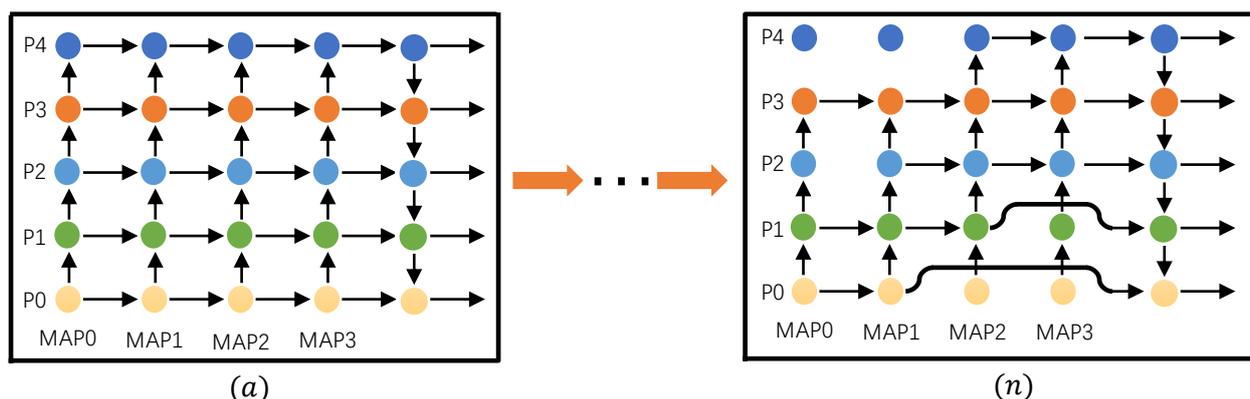
Figure 4. Multi-scale feature pyramid model structure sketch.

This paper uses neural architecture search (NAS) for the optimization of the multi-scale model of images. The search space used in this paper includes all connections between scales directly, indirectly, and across scales. The controller uses a recurrent neural network to select sub-models in the search space, using the deblurring metric PSNR as a feedback signal for updating the parameters. After selecting some structures in a search space covering all scale connections, 200 epochs were trained, and the results were tested for

comparison. By cyclic attempts, the search revealed a new multi-scale feature pyramid structure (NAS MFPN). As shown in Figure 4, from (*a*) to (*n*).

The physical mechanism of image blurring is to distribute the point source energy on the image plane according to the diffusion characteristics of the blur kernel, and the shape and size of the blur kernel are not the same for images with different degrees of blurring. However, the current network model has fixed multi-scale weights for images with different blurring degrees in the multi-scale fusion process. This uniform weight fusion method is not adaptable to the deblurring task, which limits the performance and adaptation range of the model. This paper proposes an adaptive multi-scale fusion strategy, which combines a blur description module with an image multi-scale fusion module to build an adaptive multi-scale fusion model. A sketch of the structure is shown in Figure 5.
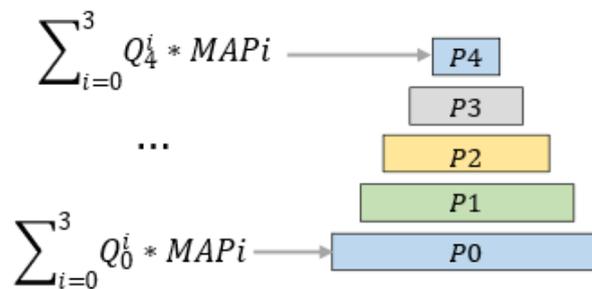
Figure 5. Adaptive multi-scale fusion strategy.

The model combines the physical characteristics of image blurring, makes full use of the characteristics of different blurred images with different energy diffusion and different information distribution, adaptively adjusts the multi-scale fusion weights according to the results of blur description, and fuses the four scales of MAP0-MAP3 feature layers according to the adjusted weights. It can effectively suppress the alignment errors in the process of multi-scale fusion and optimize the information reorganization strategy in the process of multi-scale fusion. It is possible to make less blurred images use more layers of shallow, high-resolution features, reducing the likelihood that too many deep, low-resolution features will affect their recovery, and similarly to make more blurred images use more deep, low-resolution features, thereby reducing the pixel offset of the recovered image.

3.3. Loss Function

Using cross-entropy loss in the training process of the image blur description model:

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic}\log(p_{ic}), \tag{8}$$

where M denotes the number of classes; y_{ic} denotes the discriminant function (0 or 1), which takes 1 when the true label class of sample i is c and 0 otherwise; and p_{ic} denotes the probability that sample *i* is predicted to be in class c.

During the training of the deblurring model, in order to compare the image under training, the reconstructed completed image and the original paired clear image are compared under a number of metrics. In this paper, as in SDD-GAN [53], a loss function consisting of four components, including adversarial loss, perceptual loss, reconstruction loss and colour feature loss, is used and minimized during the training of the deblurred model to achieve optimal results for the model. The total loss is defined as:

$$L = w_g\mathcal{L}_g + w_p\mathcal{L}_p + w_r\mathcal{L}_r + w_c\mathcal{L}_c, \tag{9}$$

where w_g, w_p, w_r, w_c are used as trade-off coefficients to adjust the importance of the different component losses to the total loss function.

The use of RaGAN-LS loss as an adversarial loss function $\mathcal{L}_g$ for global and local discriminators, helps the training process to be smooth and efficient [9], prompting the deblurring network model to produce clear, visually appealing images.

$$
\begin{aligned}
\mathcal{L}_g = E_{x \sim I_X\,(x)} & \left[\left(D(x) - E_{y \sim I_Y(y)} D(G(y)) - 1 \right)^2 \right] \\
+ E_{y \sim I_Y(y)} & \left[\left(D(G(y)) - E_{x \sim I_X\,(x)} D(x) + 1 \right)^2 \right],
\end{aligned}
\tag{10}
$$

where x is the sample image for the blurred image domain I_X, y is the sample image for the clear image domain I_Y, D is the discriminator, and G is the generator.

Perceptual loss [54] $\mathcal{L}_P$ improves the visual quality of the generated image by computing the L2 loss of the CNN feature map between the model-generated image and the target image, so that the generated image approximates perceptually to the real image. The perceptual loss used in this paper is the difference between the clear image and the recovered blurred image in the conv3.3 feature map of VGG-19:

$$
\mathcal{L}_P = E \left[\| VGG(Y) - VGG(G(X)) \|_2^2 \right],
\tag{11}
$$

In this paper, we choose the mean square error (MSE) loss as $\mathcal{L}_r$ and minimized $\mathcal{L}_r$ so that the model generates images with less texture distortion, defined as:

$$
\mathcal{L}_r = \frac{1}{n} \sum_{i=1}^{n} (y_i - G(x_i))^2,
\tag{12}
$$

The error between the target image and the generated image is minimized using the color feature loss function $\mathcal{L}_c$, which drives the model to produce an image with the same color distribution as the target image, and the color feature loss function is defined as ollows:

$$
\mathcal{L}_c = E_{x \sim I_X\,(x), y \sim I_Y\,(y),} \left[\sum_p \angle((G(x))_p, (y)_p) \right] + E_{y \sim I_Y\,(y)} [\| G(y) - y \|_1],
\tag{13}
$$

where (,) is the operator that calculates the angle between two images in the vector space and (.) p denotes the pixel point of the image.

3.4. Multi-Blur Level Aerial Remote Sensing Image Dataset

The existing public dataset lacks image data with multiple levels of blur, so a self-built multi-blur level remote sensing image dataset was constructed, one from data captured by the camera and one from the blur kernel simulation.

Aerial remote sensing dataset with multi-blur levels taken by the camera(RSML_C): The initial data are sampled from the aerial remote sensing open dataset DOTA [55], selecting images with aircraft and downsampling the data to 1 M and 2 M based on their actual resolution. As shown in Figure 6, in order to make the blurred image as close as possible to the real scene, we dithered the picture on a 144HZ high frame rate screen, using a DMK 33GX290-GigE black and white industrial camera; after aligning to take pictures, the output is a three-channel black and white image, and we choose four exposure times from short to long as the blur level standard for the generated picture. We chose 50 scenes, generating an average of 36 blurred images under one blurring level for each scene. The four blurring levels plus their corresponding clear images generated a total of 18,000 sets of data at two resolutions, with an image size of 640×640, and the training and test sets were randomly assigned in the ratio of 8:2.
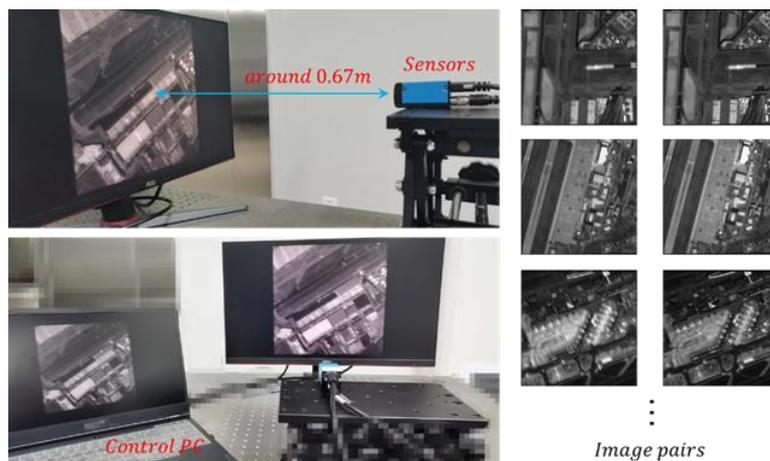
Figure 6. Diagram of the camera's shooting environment.

According to Section 4.2, SMD2 can sort and classify multi-blur levels under the same clear image within a certain range. In order to eliminate possible errors in the camera shooting process, the images under each of the four blur levels corresponding to each scene are sorted according to the SMD2 values, and the middle 2/5 under each scene are taken as the standard dataset for training the image blur description model, with an average of 72 images for each scene. All data are used for training the deblurring model.

Aerial remote sensing multi-blur level image dataset generated by motion blur kernel (RSML_S): Based on the construction of a real remote sensing motion blur image dataset by SDD-GAN, a multi-level remote sensing motion blur dataset was constructed in this paper, with the initial data sampled from the publicly available dataset DOTA [55]. The random trajectory generation method proposed in DeblurGAN [18] was used to generate motion blur kernels with randomly varying intervals of blur kernel size as [(1,1)-(5,5)), [(5,5)-(9,9)), [(9,9)-(17,17)), [(17,17)-(31,31)], and some of the blur kernels are shown in Figure 7, which are combined with SDD-GAN's motion blur image construction method to generate multi-blur level images, and the blur images are shown in Figure 8. As with RSML_C, 50 scenes were selected to generate a total of 18,000 sets of data at two resolutions, with an image size of 640*640, and the training set and test set were randomly assigned in the ratio of 8:2. The blurred images generated by the motion blur kernel are exactly as expected and free from error, so the full data can be used to train the image blur description model on multi-level blur levels.
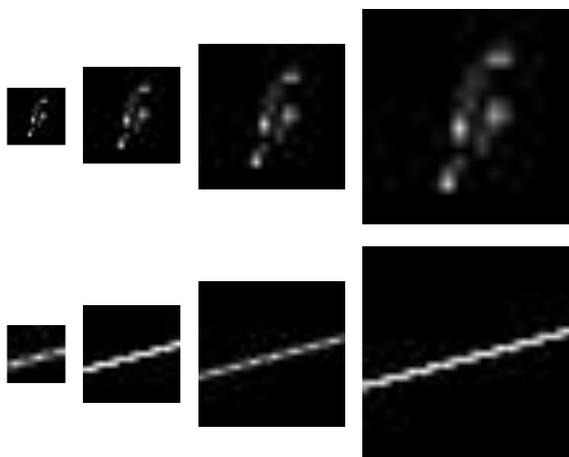
Figure 7. Schematic diagram of a random multi-level image motion blur kernel.
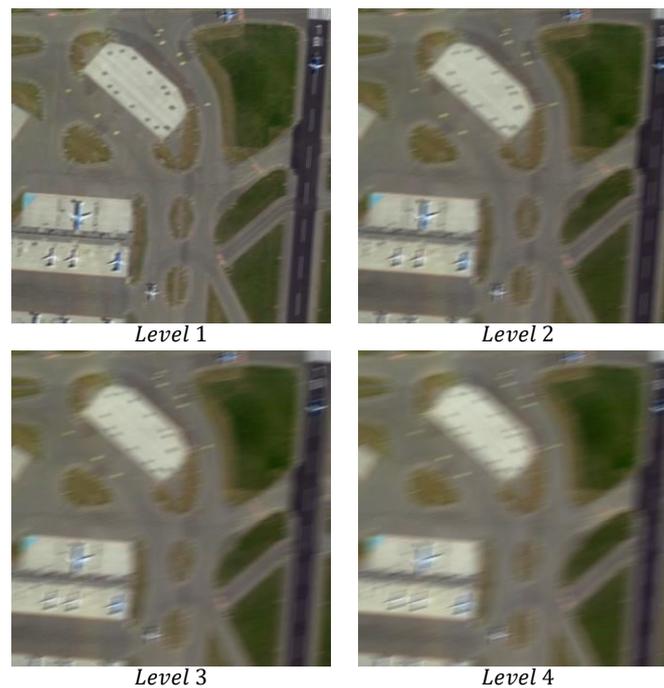
Figure 8. Example diagram of RSML_S image dataset.

4. Results

4.1. Datasets

- RSML_C dataset: For a self-built dataset of drone and aerial remote sensing images at multi-blur levels taken by the cameras in this paper, please read Section 3.4 for details.
- RSML_S dataset: For a self-built multi-blur level drone and aerial remote sensing image dataset using motion blur kernel simulation, please read Section 3.4 for details.
- GoPro dataset [21]: Two hundred and forty frames per second (fps) video sequences were captured using a GoPro Hero 4 camera, and then blurred images were generated by averaging consecutive short exposure frames. The GoPro dataset is common benchmark data in the field of image motion blur and contains 3214 blurred/clear image pairs. This paper follows the same division, using 2103 pairs for training and the remaining 1111 pairs for test evaluation.
- RealBlur dataset [56]: RealBlur-R is generated from raw camera images. Images are captured that are blurred by camera shake and captured in dimly lit environments (e.g., streets and indoor rooms at night) to cover the most common cases of motion blur and they are blurred images that are more in line with the real world. RealBlur-R has a total of 4,556 blurred and sharp image pairs, with 3,758 image pairs randomly selected as the training set and 980 image pairs as the test set.
- Visdrone dataset [57]: The dataset is collected using different drone platforms (different models of drones) in different scenarios, under different weather and lighting conditions, covering a wide range of locations (from 14 different cities in China separated by thousands of kilometers), objects (pedestrians, vehicles, bicycles, etc.), and densities (sparse and crowded scenes). This paper uses 548 validation data from Visdrone for test experiments and the motion blur kernel in Section 3.4 to construct the drone motion blur dataset.

4.2. Image Blur Description Test

For tests carried out using a number of traditional algorithms on a self-built multi-blur level dataset alone, none of the results could be accurately graded and the test results were strongly correlated with the clear image of each blurred image. The Brenner gradient function, the Laplacian gradient function, and the SMD2 (product of grey scale differences)

function can roughly classify the dataset with no change in the underlying image; SMD2 is relatively good; the remaining functions have confusing test values and cannot be clearly graded on the same clear image. The test result graph is shown in Figure A1.

As shown in Figure 9 for some of the images (14 scenes) on the SMD2 function, calculated as shown in equation (5), the right panel shows the test result curves for all images of the four blur levels at the same scale, and the left panel shows the results of comparing a clear image and its corresponding four blur levels in the same scene, the larger the value of the vertical coordinate, the clearer the image. It can be intuitively seen that the SMD2 function can distinguish the images roughly correctly between the four levels in the same scene (the same clear image), but it is obviously not possible to adapt all images using the same threshold and can only be applied in a single scene.
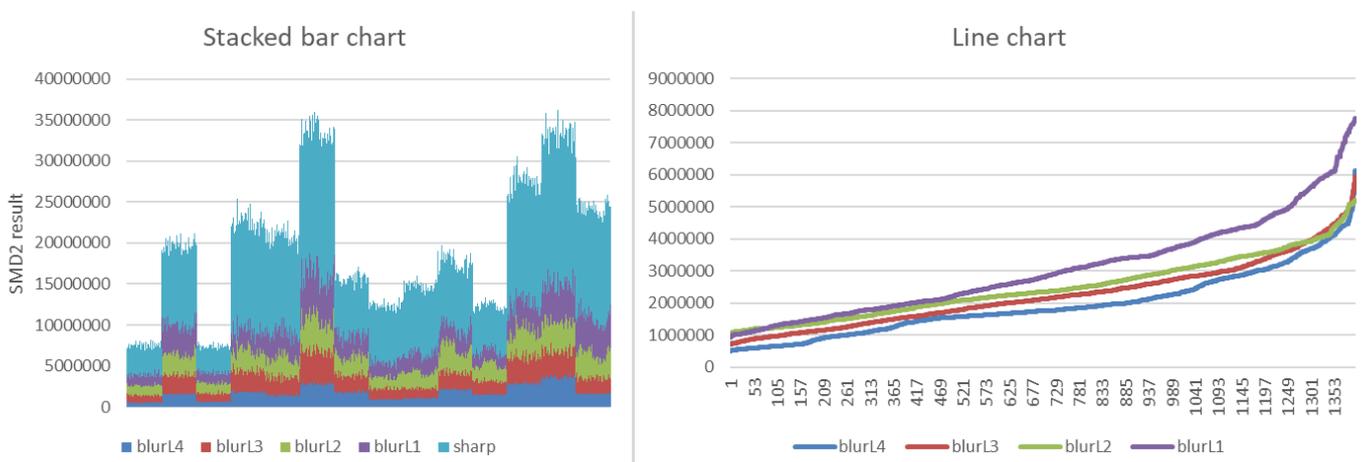
Figure 9. SMD2 blur description results for some images. The larger the SMD2 result, the clearer the image.

The blur metric of a blurred image is closely related to the clear image itself. Traditional algorithms can broadly determine the degree of blurring on the same clear image, but they cannot be adapted to a wide range of clear image situations, but in practice, the vast majority are complex environments, making them difficult to apply directly.

Because SMD2 is relatively effective in grading under the same clear picture, the four blur levels in each scene of the RSML_C dataset can be sorted according to SMD2 values, and the middle 2/5 of the image data for each blur level in each scene are used as the standard dataset for training the blur level description model for classification.

Trained on a standard dataset for blur classification, the test results are shown in Table 1 with excellent results: 99.5% accuracy on the RSML_S dataset test set and 98.9% accuracy on the RSML_C dataset test set, with some of the test results shown in Figure 10.

Table 1. The results of the blur level description model in this paper on a test set of different datasets.

Train	Test	top1 Accuracy
RSML_S	RSML_S	0.995
RSML_C(2/5)	RSML_C(2/5)	0.989

The larger the value of q, the more blurred the image is, $q \in [0, 1]$.

The trained blur level description model on the RSML_C dataset is directly migrated to the GoPro dataset and RealBlur dataset, and images with different degrees of blurring are selected for detection. Some of the detection results are shown in Figures A2 and A3. The model performs well in testing and can detect the degree of blurring of images, and the larger the pixel offset, the greater the image detection result q.
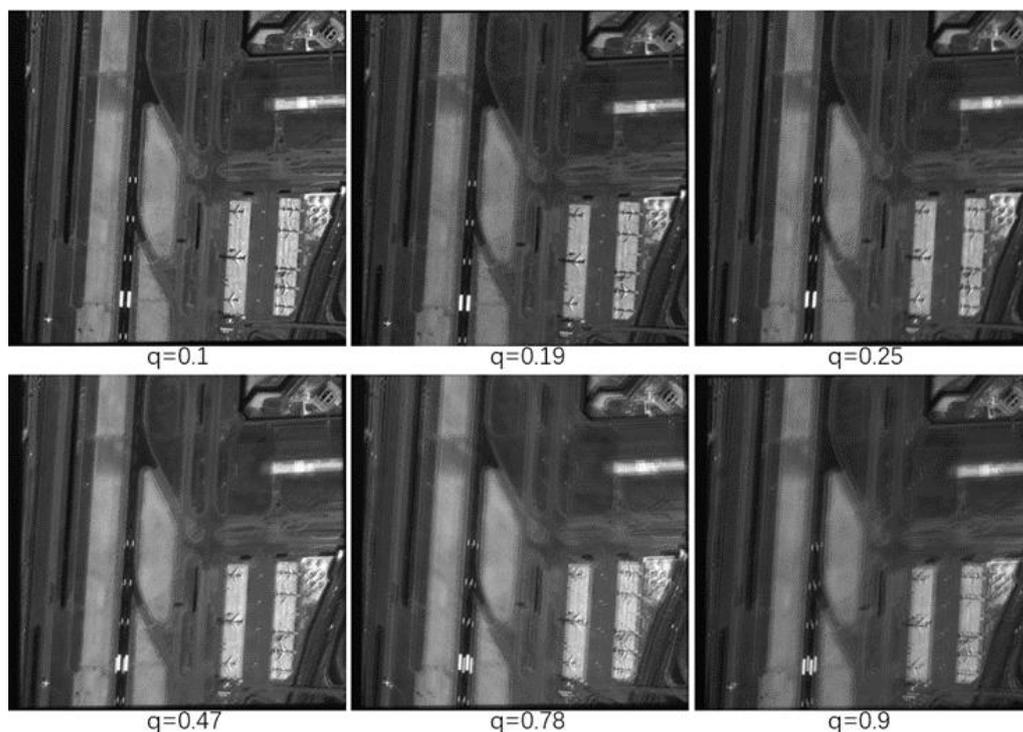
Figure 10. Results of blur level description for a scene in the RSML_C test set.

4.3. Image Deblurring Comparison Results

In order to evaluate the performance of the deblurring method proposed in this paper, the algorithm of this paper is compared with other deblurring methods on four datasets for analysis. The codes of the deblurring algorithms of the other methods and their test results on public datasets are obtained from the authors' official websites.

The RSML_S dataset and the RSML_C dataset:The results of DeblurGAN-v2 [9], MAXIM [35], MPRNet [34], DeepRFT [58], Restormer [42], and this paper were analyzed quantitatively on the self-built dataset. As the official websites of the compared algorithms all provide the training parameters of the algorithms on the GoPro dataset, the method of training on the GoPro dataset and testing on the self-built dataset was chosen for the comparison in this paper.

The average PSNR and average SSIM of different deblurring algorithms on the RSML_S dataset are shown in Table 2. The average PSNR and average SSIM of the algorithms in this paper are better than the comparison algorithms.

Table 2. The mean PSNR and SSIM for different deblurring methods on the RSML_S dataset (Trained on the GoPro dataset). The highest scores are highlighted in bold.

Method	PSNR(dB)	SSIM
DeblurGAN-v2 [9]	23.83	0.722
MAXIM [35]	24.58	0.786
MPRNet [34]	24.74	0.785
DeepRFT [58]	25.39	0.826
Restormer [42]	25.11	0.809
AMD-GAN	**27.89**	**0.869**

The test results of different defuzzification algorithms on the RSML_C dataset are shown in Table 3. The algorithm in this paper achieves a PSNR of 24.37 dB, which is an improvement of 4.67 dB compared to DeblurGAN-v2.

Table 3. The mean PSNR and SSIM for different deblurring methods on the RSML_C dataset (Trained on the GoPro dataset). The highest scores are highlighted in bold.

Method	PSNR(dB)	SSIM
DeblurGAN-v2 [9]	19.70	0.670
MAXIM [35]	21.78	0.685
MPRNet [34]	22.99	0.705
DeepRFT [58]	23.06	0.706
Restormer [42]	23.22	0.712
AMD-GAN	**24.37**	**0.786**

The results of the RSML_C dataset recovery are shown in Figure 11. As can be seen from the comparative results in the figure, the algorithm in this paper performs well in these samples, basically recovering the original details of the images and reducing a large number of artefacts compared to other algorithms.
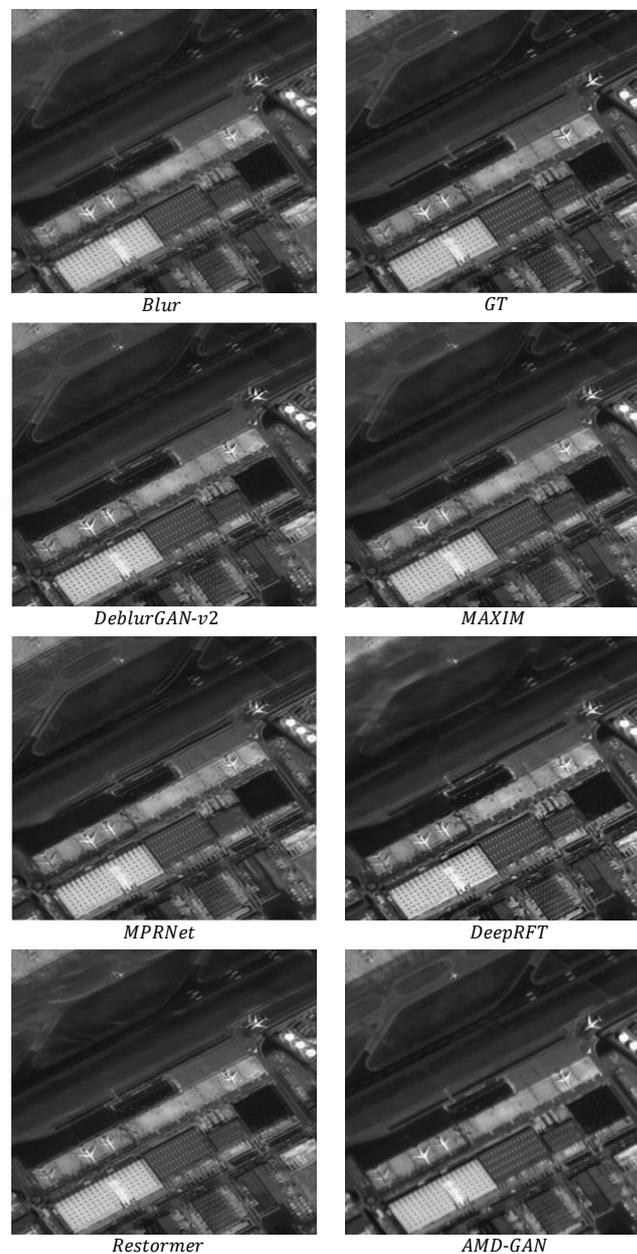
Figure 11. Recovery comparison results for the RSML_C dataset.

The method proposed in this paper is trained and tested on a self-built dataset, and the test results are shown in Table 4. Compared with those trained on the GoPro dataset, the test results are substantially improved, indicating that there is a large difference between the non-remote sensing image dataset and the remote sensing blur dataset; therefore, a new aerial remote sensing blur dataset is a necessary work.

Table 4. Training test results of this paper's method AMD-GAN on a self-built dataset.

Train	Test	PSNR(dB)	SSIM
RSML_C	RSML_C	27.35	0.876
RSML_S	RSML_S	31.32	0.917

The GoPro dataset: The results of this paper were quantitatively analyzed against other methods on the GoPro dataset.

The test results of different deblurring algorithms are shown in Table 5. The algorithm in this paper outperforms other deblurring methods in terms of average PSNR and average SSIM. The test results of this paper's algorithm on the GoPro dataset achieved a PSNR of 32.32 and an improvement of 2.77 dB compared to DeblurGAN-v2 [9], ranking first among deblurring models using the GAN method and 1.22 dB higher than the DBGAN model.

Table 5. Average PSNR and SSIM for different deblurring methods on the GoPro dataset. The highest scores are highlighted in red.

Method	PSNR(dB)	SSIM
DeepDeblur [21]	29.08	0.913
SRN [59]	30.26	0.934
DeblurGAN-v2 [9]	29.55	0.934
STFAN [22]	28.59	0.861
BlurredItV [60]	30.58	0.941
ESTRNN [61]	31.07	0.902
DBGAN [62]	31.10	0.942
DBCPeNet [48]	31.10	0.945
SDWNet [63]	31.36	-
LEBMD [64]	31.79	0.949
AMD-GAN	32.32	0.952

RealBlur-R dataset: The results of this paper were quantitatively analyzed against other methods on the RealBlur-R dataset.

Training was performed on each of the two datasets. RealBlur-R (GoPro) and RealBlur-R denote training on the GoPro training set and on the RealBlur-R training set, respectively, and both methods were tested on the test set of RealBlur-R.

The test results of different deblurring algorithms in the RealBlur-R dataset (GoPro) are shown in Table 6. The average PSNR and average SSIM of the algorithms in this paper outperformed the other algorithms. The PSNR of the algorithm in this paper reached 36.74 dB, which is 1.5 dB better compared to DeblurGAN-v2.

Table 6. Average PSNR and SSIM for different deblurring methods on the RealBlur-R dataset (GoPro). The he highest scores are highlighted in red.

Method	PSNR(dB)	SSIM
DeepDeblur [21]	32.51	0.841
DeblurGAN-v2 [9]	35.26	0.944
SRN [59]	35.66	0.947
DMPHN [33]	35.70	0.948
MAXIM [35]	35.78	-
MPRNet [34]	35.99	0.952
MSSNet [65]	35.93	0.953
DeepRFT [58]	36.06	0.954
Uformer-B [41]	36.19	0.957
Restormer [42]	36.22	0.957
AMD-GAN	36.76	0.959

The average PSNR and average SSIM of different deblurring algorithms on the RealBlur-R dataset are shown in Table 7. The algorithm in this paper outperforms the other defuzzification methods in terms of average PSNR and average SSIM. The PSNR of the algorithm in this paper reaches 40.58 dB, which is 4.14 dB higher compared to DeblurGAN-v2.

Table 7. Mean PSNR and SSIM for different deblurring methods on the RealBlur-R dataset. The highest scores are highlighted in red.

Method	PSNR(dB)	SSIM
SRN [59]	38.65	0.96
DeblurGAN-v2 [9]	36.44	0.93
MPRNet [34]	39.31	0.972
MAXIM [35]	39.45	0.961
BANet [66]	39.55	0.971
MSSNet [65]	39.76	0.972
DeepRFT [58]	39.84	0.972
Stripformer [67]	39.84	0.974
AMD-GAN	40.58	0.975

The results of the RealBlur-R dataset recovery are shown in Figure 12. Three typical scenes are chosen to compare the recoveries. The target image scenes are rich in content, and the images contain distinct edges and complex textures. Therefore, such images of complex scenes can test the recovery ability of the method in this paper very well. From the comparative results in the figure, we can see that the algorithm in this paper performs well in these samples and basically recovers the original details and colors of the image.
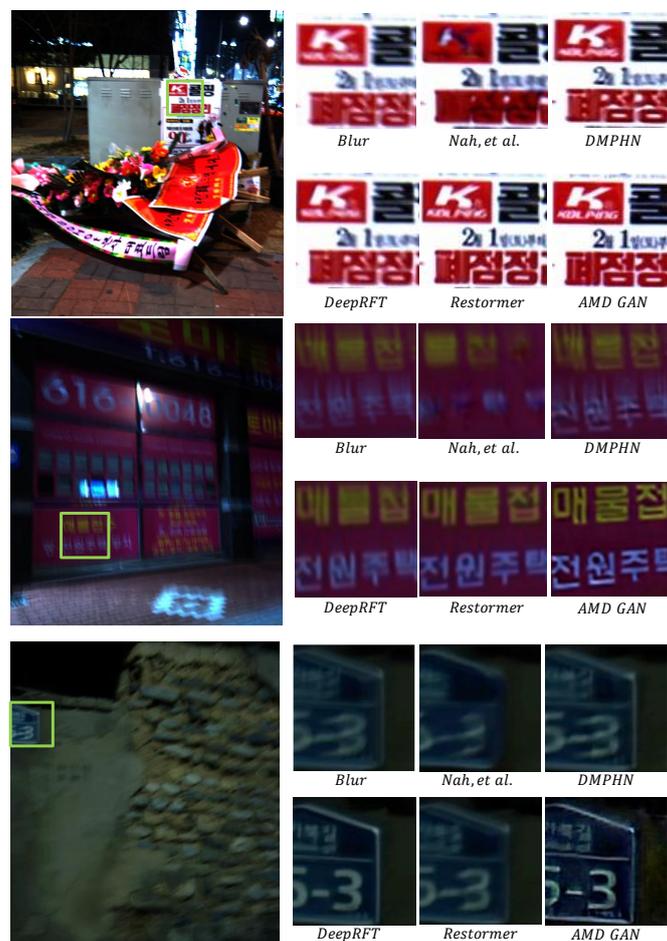
Figure 12. Comparison of recovery results for the RealBlur-R dataset.

Among the existing image deblurring algorithms, the algorithm in this paper achieves state-of-the-art (SOTA) results on the RealBlur-R dataset.

4.4. Object Detection Experiments

Image blurring not only reduces the quality of human perception, but also increases the difficulty of subsequent computer vision analysis tasks, especially for tasks that widely employ real-time image processing. To test the impact of image deblurring on other computer vision tasks, this paper uses object detection as an example for comparative testing, including the self-built RSML_C remote sensing dataset and the Visdrone drone dataset.

RSML_C dataset: The YOLOV5 [68] model is used to detect objects on blurred aerial remote sensing images, images after deblurring by other methods, and images after deblurring by this method. Objects with a confidence level of 0.5 or higher are calculated, and since it is difficult to mark the anchor box accurately for blurred images, as long as the detection box and the object area IOU are greater than 0.1, they are considered correct. The experimental results are shown in Table 8, and the comparison algorithm models are taken from their official websites. None of the compared algorithms effectively improved the object detection performance of the blurred images, while the method in this paper improved the object detection index of aerial remote sensing blurred images by 26%, which proved the effectiveness and practicality of the method in this paper.

Table 8. MAP of different methods on the RSML_C dataset.

Method	MAP(%)	P(%)	R(%)
YOLOV5	46.1	56.3	55.1
DeepRFT+YOLOV5	45.4(-1.5%)	55.0	53.1
Restormer+YOLOV5	46.6(+1%)	55.2	54.2
MPRNet+YOLOV5	46.9(+1.7%)	55.7	54.2
AMD-GAN+YOLOV5	58.1(+26%)	61.2	59.8

The test result graph is shown in Figure 13, where GT indicates the real object anchor frame and Blur indicates detection directly on the blurred image. It can be clearly seen that the image object detection accuracy using this paper's method of deblurring is high.

Visdrone dataset: The same as RSML_C dataset, YOLOV5 is used to detect the object of the drone blurred image and the image after deblurring by this method. The Visdrone blurred images were deblurred using the method in this paper. The average PSNR was 31.176 dB and the SSIM was 0.918. The test comparison results for object detection are shown in Table 9, and the effect graph is shown in Figure 14. GT indicates a sharp image, and Blur indicates detection directly on a blurred image. Using the image deblurring method in this paper can significantly improve the object detection index, and the MAP improves by 0.108 after image deblurring.

Table 9. Comparison results of object detection on the Visdrone dataset.

Method	MAP(%)	P(%)	R(%)
GT + YOLOV5	37.5	61.8	36.5
Blur + YOLOV5	19.7	44.9	20
AMD-GAN + YOLOV5	30.5	50.3	31.3
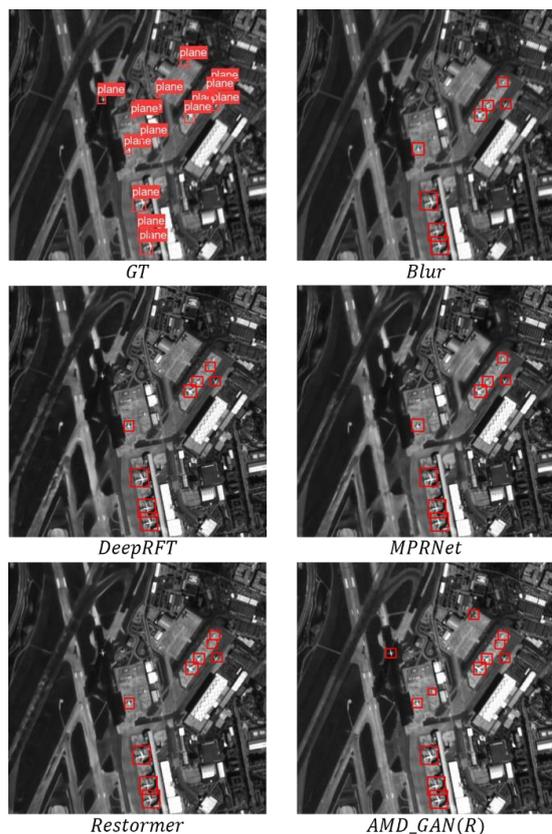
Figure 13. Comparison of object detection results on the RSML_C dataset.
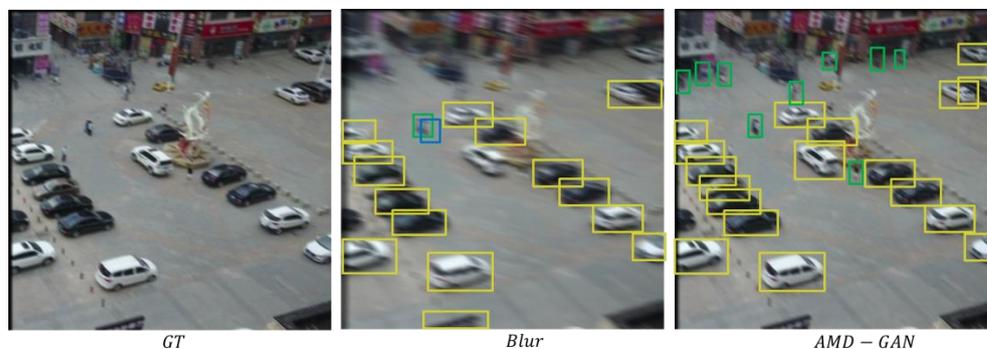
Figure 14. Comparison of object detection results on the Visdrone dataset.

4.5. Ablation Study

4.5.1. The Effectiveness of Multi-Scale Feature Pyramid Network Structures

The multi-scale fusion structure using neural network search (NAS MFPN) is removed from the generator and tested on the self-built dataset RSML_C using the structure shown in (a) in Figure 4 to verify the effectiveness of NAS MFPN. The results are shown in Table 10, with an average PSNR improvement of 0.74 dB.

Table 10. Results of ablation experiments on the RSML_C dataset for PSNR and SSIM (NAS MFPN).

Method	PSNR(dB)	SSIM
AMD-GAN without NAS MFPN	24.68	0.794
AMD-GAN with NAS MFPN	**25.24**	**0.835**

4.5.2. The Effectiveness of Adaptive Multi-Scale Fusion

The adaptive multi-scale fusion structure (AMS Fusion) in the generator structure is removed, and the multi-scale fusion structure (NAS MFPN) in (n) in Figure 4 is used to test on the self-built dataset RSML_C to verify the effectiveness of AMS Fusion. The results are shown in Table 11, with an average PSNR improvement of 2.11 dB.

Table 11. Results of ablation experiments on the RSML_C dataset for PSNR and SSIM (AMS Fusion).

Method	PSNR(dB)	SSIM
AMD-GAN without AMS Fusion	25.24	0.835
AMD-GAN with AMS Fusion	**27.35**	**0.876**

5. Conclusions

This paper proposes a single-image blind deblurring generative adversarial network (AMD-GAN) using adaptive multi-scale fusion of physical properties of blurred images, which takes full advantage of the different energy diffusion and information distribution of images with different degrees of blurring to enhance the deblurring effect. Experiments show that the method in this paper has good recovery effects on self-built remote sensing datasets, drone datasets, and publicly deblurred datasets, which is conducive to mitigating the impact of image blurring on other computer vision analysis tasks and can significantly improve the accuracy of object detection on blurred images.

In future work, we will investigate methods to fuse features more finely for multi-scale feature fusion in order to improve the performance of image deblurring.

Author Contributions: Conceptualization, B.Z. and Q.L.; methodology, B.Z.; software, B.Z.; validation, B.Z.; writing—original draft preparation, B.Z.; writing—review and editing, B.Z. and Z.T.; project administration, Q.L.; funding acquisition, Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Program Project of Science and Technology Innovation of Chinese Academy of Sciences (no. KGFZD-135-20-03-02).

Data Availability Statement: Not applicable.

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

Appendix A
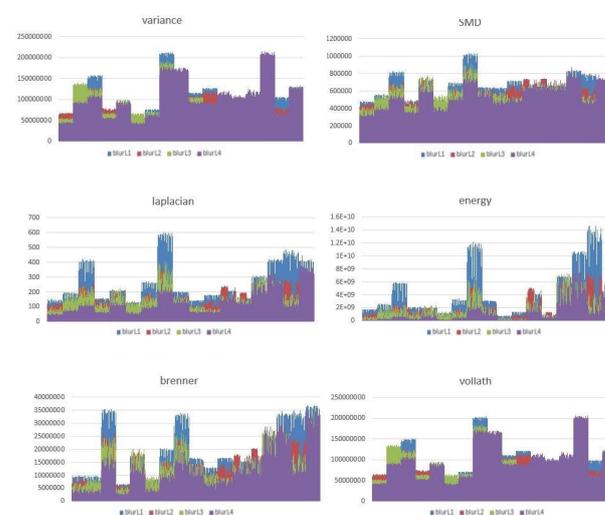
Figure A1. Results of blur degree description in different algorithms for some images of the self-built multi-blur level dataset.

Figure A2. GoPro dataset blur degree description results. The larger the value of q, the more blurred the image is, $q \in [0, 1]$.

Figure A3. RealBlur dataset blur degree description results. The larger the value of q, the more blurred the image is, $q \in [0, 1]$.

References

1. Himeur, Y.; Rimal, B.; Tiwary, A.; Amira, A. Using Artificial Intelligence and Data Fusion for Environmental Monitoring: A Review and Future Perspectives. *Inf. Fusion* **2022**, *86–87*, 44–75. [CrossRef]
2. Zhang, J.; Xu, S.; Zhao, Y.; Sun, J.; Xu, S.; Zhang, X. Aerial Orthoimage Generation for UAV Remote Sensing: Review. *Inf. Fusion* **2023**, *89*, 91–120. [CrossRef]
3. Miller, Z.; Hupy, J.; Hubbard, S.; Shao, G. Precise Quantification of Land Cover before and after Planned Disturbance Events with UAS-Derived Imagery. *Drones* **2022**, *6*, 52. [CrossRef]
4. Shelekhov, A.; Afanasiev, A.; Shelekhova, E.; Kobzev, A.; Tel'minov, A.; Molchunov, A.; Poplevina, O. Low-Altitude Sensing of Urban Atmospheric Turbulence with UAV. *Drones* **2022**, *6*, 61. [CrossRef]
5. Park, G.; Park, K.; Song, B.; Lee, H. Analyzing Impact of Types of UAV-Derived Images on the Object-Based Classification of Land Cover in an Urban Area. *Drones* **2022**, *6*, 71. [CrossRef]
6. Liu, S.; Tong, X.; Li, L.; Ye, Z.; Lin, F.; Zhang, H.; Jin, Y.; Xie, H. Geometric Modeling of Attitude Jitter for Three-Line-Array Imaging Satellites. *Opt. Express* **2021**, *29*, 20952–20969. [CrossRef]
7. Simsek, B.; Bilge, H.S. A Novel Motion Blur Resistant VSLAM Framework for Micro/Nano-UAVs. *Drones* **2021**, *5*, 121. [CrossRef]
8. Zhang, K.; Ren, W.; Luo, W.; Lai, W.-S.; Stenger, B.; Yang, M.-H.; Li, H. Deep Image Deblurring: A Survey. *Int. J. Comput. Vis.* **2022**, *130*, 2103–2130. [CrossRef]
9. Kupyn, 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.

10. Chen, S.; Lan, J.; Liu, H.; Chen, C.; Wang, X. Helmet Wearing Detection of Motorcycle Drivers Using Deep Learning Network with Residual Transformer-Spatial Attention. *Drones* **2022**, *6*, 415. [CrossRef]

11. Bertasius, G.; Torresani, L.; Shi, J. Object Detection in Video with Spatiotemporal Sampling Networks. In *Lecture Notes in Computer Science*; Springer; Cham, Switzerland, 2018; pp. 331–346.

12. Bai, Y.; Cheung, G.; Liu, X.; Gao, W. Graph-Based Blind Image Deblurring from a Single Photograph. *IEEE Trans. Image Process.* **2019**, *28*, 1404–1418. [CrossRef]

13. Ren, D.; Zhang, K.; Wang, Q.; Hu, Q.; Zuo, W. Neural Blind Deconvolution Using Deep Priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3341–3350.

14. Dong, J.; Roth, S.; Schiele, B. DWDN: Deep Wiener Deconvolution Network for Non-Blind Image Deblurring. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9960–9976. [CrossRef] [PubMed]

15. Zhang, Z.; Zheng, L.; Piao, Y.; Tao, S.; Xu, W.; Gao, T.; Wu, X. Blind Remote Sensing Image Deblurring Using Local Binary Pattern Prior. *Remote Sens.* **2022**, *14*, 1276. [CrossRef]

16. Sun, S.; Duan, L.; Xu, Z.; Zhang, J. Blind Deblurring Based on Sigmoid Function. *Sensors* **2021**, *21*, 3484. [CrossRef] [PubMed]

17. Zhang, Z.; Zheng, L.; Xu, W.; Gao, T.; Wu, X.; Yang, B. Blind Remote Sensing Image Deblurring Based on Overlapped Patches' Non-Linear Prior. *Sensors* **2022**, *22*, 7858. [CrossRef] [PubMed]

18. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. Deblurgan: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8183–8192.

19. Feng, H.; Guo, J.; Xu, H.; Ge, S.S. SharpGAN: Dynamic Scene Deblurring Method for Smart Ship Based on Receptive Field Block and Generative Adversarial Networks. *Sensors* **2021**, *21*, 3641. [CrossRef]

20. Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; Ko, S.-J. Rethinking Coarse-to-Fine Approach in Single Image Deblurring. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, 10–17 October 2021; IEEE: New York, NY, USA, 2021; pp. 4621–4630.

21. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA; 2017; pp. 3883–3891.

22. Zhou, S.; Zhang, J.; Pan, J.; Xie, H.; Ren, J. Spatio-Temporal Filter Adaptive Network for Video Deblurring. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.

23. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems 27 (nips 2014); Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Neural Information Processing Systems (nips): La Jolla, CA, USA, 2014; Volume 27, pp. 2672–2680.

24. Hong, Y.; Hwang, U.; Yoo, J.; Yoon, S. How Generative Adversarial Networks and Their Variants Work: An Overview. *ACM Comput. Surv.* **2019**, *52*, 10. [CrossRef]

25. Liu, P.; Li, J.; Wang, L.; He, G. Remote Sensing Data Fusion With Generative Adversarial Networks: State-of-the-Art Methods and Future Research Directions. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 295–328. [CrossRef]

26. Liu, S.; Wang, H.; Wang, J.; Pan, C. Blur-Kernel Bound Estimation From Pyramid Statistics. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 1012–1016. [CrossRef]

27. Zhang, Q.; Zeng, Z.; Lie, Y.; Tane, K.; Wang, J. Dynamic Scene Deblurring Using Enhanced Feature Fusion and Multi-Distillation Mechanism. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021.

28. Suin, M.; Purohit, K.; Rajagopalan, A. Spatially-Attentive Patch-Hierarchical Network for Adaptive Motion Deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3606–3615.

29. Gao, H.; Tao, X.; Shen, X.; Jia, J. Dynamic Scene Deblurring with Parameter Selective Sharing and Nested Skip Connections. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 15–20 June 2019; pp. 3843–3851.

30. Liu, P.; Wang, L.; Ranjan, R.; He, G.; Zhao, L. A Survey on Active Deep Learning: From Model Driven to Data Driven. *ACM Comput. Surv.* **2022**, *54*, 221. [CrossRef]

31. Tadros, T.; Cullen, N.C.; Greene, M.R.; Cooper, E.A. Assessing Neural Network Scene Classification from Degraded Images. *ACM Trans. Appl. Percept.* **2019**, *16*, 21. [CrossRef]

32. Bai, Y.; Jia, H.; Jiang, M.; Liu, X.; Xie, X.; Gao, W. Single-Image Blind Deblurring Using Multi-Scale Latent Structure Prior. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2033–2045. [CrossRef]

33. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.

34. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Multi-Stage Progressive Image Restoration. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.

35. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxim: Multi-Axis Mlp for Image Processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5769–5780.

36. Li, Z.; Guo, J.; Zhang, Y.; Li, J.; Wu, Y. Reference-Based Multi-Level Features Fusion Deblurring Network for Optical Remote Sensing Images. *Remote Sens.* **2022**, *14*, 2520. [CrossRef]

37. Shu, J.; Xie, C.; Gao, Z. Blind Restoration of Atmospheric Turbulence-Degraded Images Based on Curriculum Learning. *Remote Sens.* **2022**, *14*, 4797. [CrossRef]

38. Xiao, Y.; Zhang, J.; Chen, W.; Wang, Y.; You, J.; Wang, Q. SR-DeblurUGAN: An End-to-End Super-Resolution and Deblurring Model with High Performance. *Drones* **2022**, *6*, 162. [CrossRef]

39. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.

40. Liang, J.; Cao, J.; Fan, Y.; Zhang, K.; Ranjan, R.; Li, Y.; Timofte, R.; Van Gool, L. Vrt: A Video Restoration Transformer. *arXiv* **2021**, arXiv:2201.12288.

41. 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.

42. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H. Restormer: Efficient Transformer for High-Resolution Image Restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.

43. Zhou, Q.; Ding, M.; Zhang, X. Image Deblurring Using Multi-Stream Bottom-Top-Bottom Attention Network and Global Information-Based Fusion and Reconstruction Network. *Sensors* **2020**, *20*, 3724. [CrossRef] [PubMed]

44. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Honolulu, HI, USA, 21–26 June 2017, pp. 2117–2125.

45. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; IEEE: New York, NY, USA, 2018; pp. 8759–8768.

46. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.

47. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. In Proceedings of the 2019 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (cvpr 2019), Long Beach, CA, USA, 16–17 June 2019; IEEE: New York, NY, USA, 2019; pp. 7029–7038.

48. Cai, J.; Zuo, W.; Zhang, L. Dark and Bright Channel Prior Embedded Network for Dynamic Scene Deblurring. *IEEE Trans. Image Process.* **2020**, *29*, 6885–6897. [CrossRef]

49. Gu, K.; Zhou, J.; Qiao, J.-F.; Zhai, G.; Lin, W.; Bovik, A.C. No-Reference Quality Assessment of Screen Content Pictures. *IEEE Trans. Image Process.* **2017**, *26*, 4005–4018. [CrossRef]

50. Wang, T.; Yin, Q.; Cao, F.; Li, M.; Lin, Z.; An, W. Noise Parameter Estimation Two-Stage Network for Single Infrared Dim Small Target Image Destriping. *Remote Sens.* **2022**, *14*, 5056. [CrossRef]

51. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the International Conference on Machine Learning*; Chaudhuri, K., Salakhutdinov, R., Eds.; Jmlr-Journal Machine Learning Research: San Diego, CA, USA, 2019; Volume 97.

52. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

53. Zhu, B.; Lv, Q.; Yang, Y.; Sui, X.; Zhang, Y.; Tang, Y.; Tan, Z. Blind Deblurring of Remote-Sensing Single Images Based on Feature Alignment. *Sensors* **2022**, *22*, 7894. [CrossRef]

54. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.

55. Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.

56. Rim, J.; Lee, H.; Won, J.; Cho, S. Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 184–201.

57. Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and Tracking Meet Drones Challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7380–7399. [CrossRef] [PubMed]

58. Mao, X.; Liu, Y.; Shen, W.; Li, Q.; Wang, Y. Deep Residual Fourier Transformation for Single Image Deblurring. *arXiv* **2021**, arXiv:2111.11745.

59. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-Recurrent Network for Deep Image Deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182.

60. Purohit, K.; Shah, A.; Rajagopalan, A. Bringing Alive Blurred Moments. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6830–6839.
61. Zhong, Z.; Gao, Y.; Zheng, Y.; Zheng, B. Efficient Spatio-Temporal Recurrent Neural Network for Video Deblurring. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 191–207.
62. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Stenger, B.; Liu, W.; Li, H. Deblurring by Realistic Blurring. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 2734–2743.
63. Zou, W.; Jiang, M.; Zhang, Y.; Chen, L.; Lu, Z.; Wu, Y. SDWNet: A Straight Dilated Network with Wavelet Transformation for Image Deblurring. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW 2021), Montreal, BC, Canada, 11–17 October 2021; IEEE Computer Soc: Los Alamitos, CA, USA, 2021; pp. 1895–1904.
64. Jiang, Z.; Zhang, Y.; Zou, D.; Ren, J.; Lv, J.; Liu, Y. Learning Event-Based Motion Deblurring. In Proceedings of the 2020 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: New York, NY, USA, 2020; pp. 3317–3326.
65. Kim, K.; Lee, S.; Cho, S. MSSNet: Multi-Scale-Stage Network for Single Image Deblurring. In Proceedings of the ECCVW (AIM), Tel Aviv, Israel, 23–23 October 2022.
66. Tsai, F.-J.; Peng, Y.-T.; Tsai, C.-C.; Lin, Y.-Y.; Lin, C.-W. BANet: A Blur-Aware Attention Network for Dynamic Scene Deblurring. *IEEE Trans. Image Process.* **2022**, *31*, 6789–6799. [CrossRef] [PubMed]
67. Tsai, F.-J.; Peng, Y.-T.; Lin, Y.-Y.; Tsai, C.-C.; Lin, C.-W. Stripformer: Strip Transformer for Fast Image Deblurring. In Proceedings of the ECCV, Tel Aviv, Israel, 23–27 October 2022; pp. 146–162.
68. Ultralytics. Yolov5. Available online: Https://Github.Com/Ultralytics/Yolov5 (accessed on 28 November 2022).