Results in Physics 56 (2024) 107262

Contents lists available at ScienceDirect

Results in Physics

journal homepage: www.elsevier.com/locate/rinp

A method for remote sensing image restoration based on the system degradation model

Pengfei Zhang, Jinnan Gong, Shikai Jiang *, Tianjun Shi, Jiawei Yang, Guangzhen Bao, Xiyang Zhi

Research Center for Space Optical Engineering, Harbin Institute of Technology, Harbin 150001, China

A R T I C L E I N F O

Keywords:
Optical remote sensing
Image restoration
Deep learning
Multi-stage network
Loss function design

A B S T R A C T

High-definition remote sensing images have been widely used in many fields such as urban planning and resource exploration. Due to the impact of imaging links, the quality of images produced by remote sensing platforms often deteriorates. Therefore, high-performance remote sensing image restoration processing methods are of great significance for improving their application efficiency. However, existing deep learning methods have not been matched and adjusted based on the imaging characteristics and degradation mechanism of remote sensing systems, and lack representation and constraints on various prior information of remote sensing platforms, which leads to false information easily and is not conducive to interpretation application. To overcome these problems, in this work, we conduct research on a multi-stage remote sensing image restoration network. First, we propose a multi-stage network framework of "denoising deblurring detail enhancement", in which structures at different stages are designed to address the multi-scale characteristics of remote sensing images. Secondly, in order to avoid the degradation of the multi-stage network, we design a differentiated intermediate supervision module and an adaptive structural adjustment module. Finally, based on the degradation characteristics of the remote sensing platform imaging system, we propose a loss function from the prior term of the modulation transfer function. We verify our method by conducting comprehensive comparisons on a new dataset by the full-link imaging simulation. Experiments show that our method is superior to other comparison methods in remote sensing image restoration, and shows competitive performance in terms of restoration performance and speed. Compared with the basic network, the proposed modules improve PSNR and SSIM by 7.18 % and 4.74 % respectively.

Introduction

Remote sensing technology is a comprehensive technology that utilizes sensors to capture electromagnetic wave information emitted by remote targets and perform image processing to achieve remote observation, imaging, and analysis of targets. According to the detection band of the sensors used, remote sensing technology can be divided into various types, such as visible light remote sensing, infrared remote sensing, ultraviolet remote sensing, and multispectral remote sensing. With the help of remote sensing technology, observation equipment in outer space can achieve high-precision detection and recognition of the ground. High-resolution remote sensing images have been widely applied in multiple fields, such as urban planning, resource exploration, and military observation.

The basic principle of traditional methods is to deconvolute the imaging results and search for the inverse transformation operator H^{-1} of the linear convolution degradation operator H. Due to the mathematical irreversibility of convolution operations, image restoration problems are essentially pathological. The focus and difficulty of traditional image restoration algorithms lie in overcoming the ill-posed problem of image deconvolution and suppressing noise while restoring images. These methods (e.g., [1–6]) usually formulate the task as a Bayesian problem, solving under a unified MAP (maximizing a posterior) framework. Regularization methods are used to alleviate the ill-posed problem by enforcing desired property, which involves sophisticated priors, e.g., total variation [7], sparse representation [8–10], low-rank [11], and self-similarity [1,2]. However, the representation ability of handcrafted design is limited, leading to unstable results, and they are usually time-

* Corresponding author.
E-mail addresses: 22B921003@stu.hit.edu.cn (P. Zhang), gongjinnan@hit.edu.cn (J. Gong), jiangshikai@stu.hit.edu.cn (S. Jiang), 22B921004@stu.hit.edu.cn (T. Shi), 23S021001@hit.edu.cn (J. Yang), 22B921005@stu.hit.edu.cn (G. Bao), zhixiyang@hit.edu.cn (X. Zhi).

https://doi.org/10.1016/j.rinp.2023.107262
Received 13 October 2023; Received in revised form 23 November 2023; Accepted 8 December 2023
Available online 9 December 2023

consuming in inference compared to deep learning methods.

With the development of deep learning, the performance of image restoration methods improved significantly. Deep learning methods based on Deep Convolutional Neural Networks (CNNs) [12,13] learn the regression between a blurry input image and the corresponding sharp image in an end-to-end manner [14,15]. Up to now, numerous function units have been proposed. [16] proposed a memory strategy to broadcast useful information in different layers. [17–21] utilized hourglass-shaped architectures to explore multiscale features. Some non-local methods [22–24] were proposed to enlarge the receptive field. However, deep learning-based image restoration methods often do not design network structures specifically based on remote sensing images' characteristics and degradation process, resulting in limited restoration effects.

In this paper, we propose a multi-stage adaptive remote sensing image restoration network, focusing on solving the problems in deep learning-based restoration algorithms. We adjust the network structure according to the characteristics of remote sensing images and apply the system degradation prior of remote sensing images. Firstly, we design a combined network structure based on the image degradation process, which includes three stages of "denoising, deblurring, and detail enhancement". At each stage, we design a matching network structure according to the current stage's task requirements to improve the network's overall restoration performance. Secondly, in order to avoid multi-stage network degradation, we design an inter-stage information transmission method based on the differentiated intermediate supervision and the adaptive structure adjustment module. Finally, in response to the existing deep learning-based restoration methods being unable to utilize the system degradation prior to remote sensing platforms, we propose a modulation transfer function (MTF) loss function to improve the overall restoration performance of the network.

In the experiment, we construct a remote sensing image restoration dataset to evaluate the restoration performance of the network. We proved the effectiveness of each improvement module based on objective evaluation indicators such as peak signal-to-noise ratio and structural similarity. The experimental results show that the proposed method has good robustness and effectiveness in remote sensing image restoration. The main contributions of this work are summarized as follows:

- A combined network structure based on the image degradation process that includes three stages of "denoising, deblurring, and detail enhancement". In each stage, a matching network structure is designed according to the task requirements of the current stage to improve the overall restoration performance of the network.
- An interstage information transfer method based on the idea of differentiated intermediate supervision. According to the specific tasks to be solved in each stage, the attention supervision module is used to monitor and modulate the results of each stage to ensure the transmission of effective feature information between stages. In order to adapt to the tasks of each stage of the network, we also propose an adaptive network structure adjustment method to further improve the resilience and generalization ability of the network.
- A loss function group based on remote sensing imaging modulation transfer function. It introduces system degradation information from remote sensing platforms in the training process of the network, effectively utilizing prior information of the imaging system and improving the authenticity of network restoration results.

The rest of this article is organized as follows. Section 2 introduces the related work and existing problems of remote sensing image restoration. Section 3 illustrates the rationale and details of the proposed method. In Section 4, we compare the performance of the proposed method with typical methods. Finally, we summarize and conclude in Section 5.

Related works

In the process of remote sensing imaging links, image quality is affected by various degradation factors, leading to a decrease in image quality in various forms, such as image blur, detail loss, and increased noise. At present, the existing remote sensing image restoration algorithms can be divided into two types: traditional restoration methods and deep learning methods.

Traditional restoration methods

According to whether the fuzzy kernel is known or not, the algorithm can be divided into two types: nonblind restoration and blind restoration. For nonblind restoration algorithms, Richardson et al., [25] proposed the Richardson-Lucy algorithm, which is based on Bayesian statistics and maximum likelihood estimation principles, using Point Spread Function (PSF) to iteratively estimate the original image. However, the Richardson-Lucy algorithm is sensitive to noise, so it may encounter artifacts or excessive noise amplification when processing noisy images. Dong et al., [26] introduced the wavelet frame contraction in the field of mathematics into the calculation process of the image restoration algorithm and used the wavelet frame to explain the partial differential equation, which improved the convergence speed and restoration performance of the algorithm. For blind restoration algorithms, due to the lack of system degradation information, it is necessary to combine the prior knowledge of the physical model of the degradation process to constrain the deconvolution process. The core problem lies in the construction of regularization terms. You et al., [27] proposed a widely used regularization method in 1996, and then Osher et al., [28] introduced a new iterative regularization term using Bregman distance as the inverse problem, improving the algorithm's performance for image denoising tasks. Lefkimmiatis et al., [29] used a norm regularization method based on the Hessian matrix, using the spectral radius of the Hessian matrix as the regularization term and adding it to the objective function of the restoration problem. However, the traditional method often has a large amount of calculation, and in complex and changeable scenes, it will introduce false details and amplify noise.

Deep learning methods

Deep learning has exploded in CV fields over the past decade. With strong abilities in feature extraction and characterization, deep learning is in wide usage across remote sensing scenarios. The image restoration method based on deep learning adaptively learns the deep semantic features of the image, which improves the restoration effect while suppressing noise and artifacts. In 2015, Sun et al., [30] first introduced the convolutional neural network into the field of image restoration. This method decomposes degraded images into gradient and residual images and uses CNN to learn the mapping relationship between gradient and residual images to achieve restoration of non-uniform motion blurred images. Kupyn et al., [31,32] proposed the DeblurGAN network in 2018, introducing the generative adversarial network into the field of image restoration, and introducing a step-by-step training method to improve the stability of the model. Hui et al., [35,36] utilized variational physics-informed and dual-target mechanisms, combined with deep learning methods, to effectively improve the restoration effect of synthetic aperture optical systems. Wang et al., proposed the methods [37–39] based on smart protocol and reinforcement learning, which effectively improves the performance of image enhancement. Zamir et al., [21] adopted a multi-stage progressive approach to design a restoration network MPRNet with a multi-stage network structure. This network decomposes tasks into multiple subtasks to gradually remove noise and improve the network's restoration performance. However, current deep learning-based image restoration methods often do not utilize prior information from remote sensing imaging degradation models, resulting in performance limitations. The lack of relatively complete image

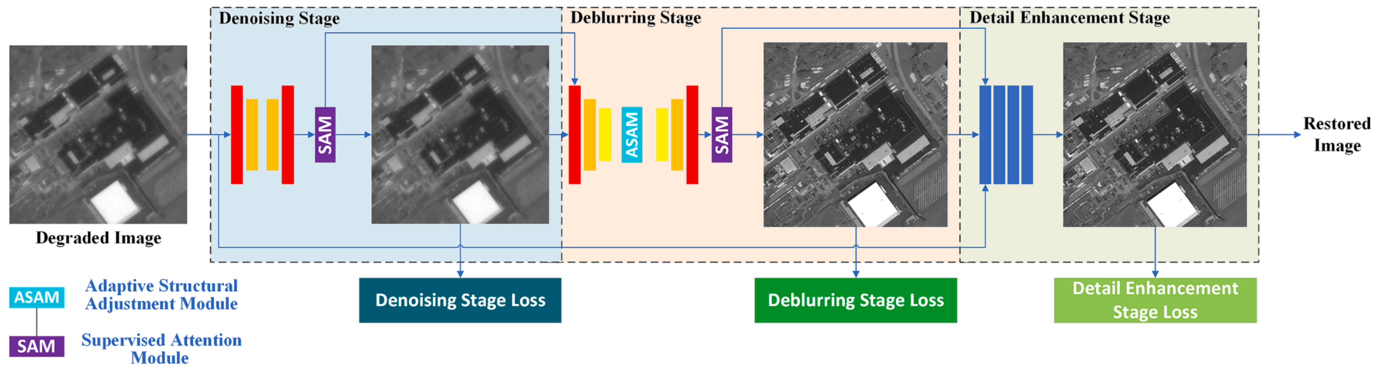

Fig. 1. The structure of MARRNet.

restoration data sets based on real degradation models based on multi-remote sensing platforms has also caused difficulties in the training of deep learning methods.

Proposed method

In view of the above problems, we innovatively propose a multi-stage adaptive remote-sensing restoration network (MARRNet) based on the combination of deep learning and prior information of remote sensing optical systems. Based on the degradation model of remote sensing imaging and imaging priors, we adjust the network structure and loss function to improve the algorithm's restoration performance for complex remote sensing images. At the same time, the network can adaptively adjust its structure based on remote sensing image information to improve the restoration effect of remote sensing platform imaging under different conditions and enhance the network's generalization ability.

Method overview

In traditional image restoration tasks, due to the relatively single factor of image degradation, in the same scene, it is generally only necessary to select one type of denoising, deblurring, or rain removal based on the corresponding degradation factors (motion blur, rain, etc.) to restore the degraded image. During the imaging process of remote sensing images, there are many degradation factors that can affect them, such as atmospheric disturbances, optical system aberrations, platform high and low-frequency vibrations, and detector noise. As a result, remote sensing image restoration tasks require simultaneous denoising and deblurring of input degraded images. Therefore, we introduce the influence of image degradation models and propose a multi-stage image restoration network model based on the three-stage idea of "denoising, deblurring, and detail enhancement" to achieve end-to-end restoration tasks for remote sensing images. At the same time, we introduce a differentiated intermediate supervision mechanism, an adaptive network structure adjustment, and a novel loss function group based on MTF to further improve the network's restoration performance and generalization ability. See (Fig. 1).

Three-stages network structure based on image degradation process

The basic idea of the network is to construct a network structure consisting of multiple stages, where each stage focuses on solving degradation problems at different levels. This phased and step-by-step restoration strategy allows the network to handle relatively simple degradation problems in the initial stage, and then gradually solve more complex degradation phenomena in subsequent stages. This network can solve specific degradation problems in various stages and has good restoration performance and robustness. In addition, this multi-stage restoration strategy also helps to enhance the network's generalization ability.

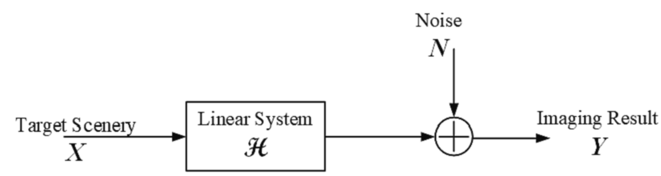

Fig. 2. Typical remote sensing image degradation model.

And typical remote sensing image degradation model suggests that the process of imaging the target scenery through a remote sensing system can be seen as a convolutional process, as shown in Fig. 2. During the imaging process, the scene is first convolved with the fuzzy kernel of the imaging system, and then linearly superimposed with random noise. It is equivalent to sequentially blurring and denoising the input image. Therefore, in order to restore the original image from a degraded image, it is necessary to perform inverse transformation on the above operations, that is, perform denoising and deblurring operations on the input image in sequence.

Therefore, we propose a multi-stage image restoration network model based on the three-stage concept of "denoising, deblurring, and detail enhancement", combining the above multi-stage network design concept with the image degradation model, to achieve end-to-end restoration tasks for remote sensing images.

Firstly, for the network structure in the denoising stage (as shown in Fig. 3(a)), we adopted a basic structure composed of encoder-decoder and skip connection, and introduced a channel attention module (CAB) to enhance the network's feature learning ability, in order to better fit the complex physical model of remote sensing image degradation process. The basic structure of the encoder and decoder consists of convolution and CAB. Adopting bilinear upsampling instead of the original transposed convolution method to eliminate the checkerboard effect caused by transposed convolution.

Secondly, in order to achieve information transmission between stages, we improve the structure of the encoder and decoder in the deblurring stage of the network (as shown in Fig. 3(b)). Due to the complexity of the remote sensing image degradation process that needs to be modeled in the deblurring stage of the network, we introduce a multi-scale encoder module and a nested skip connection module on the basis of the above network to enhance the network's feature learning and representation capabilities and enhance the network's restoration effect on remote sensing images. We also introduce an adaptive structural adjustment module in the network, which will be described in detail later.

The structure of the multi-scale feature encoder module is shown in Fig. 3(c). The basic principle is to scale the input image by using step-by-step downsampling, generate a series of images with different resolutions and scales, and then input these images together with the multi-channel feature maps of the upper level into the convolution layer of

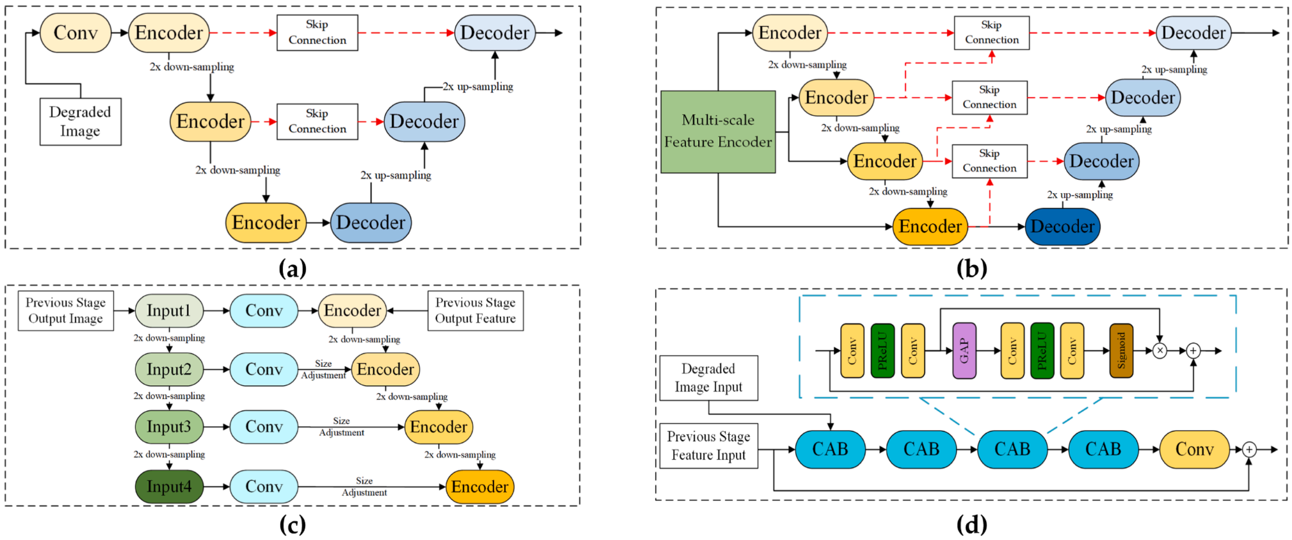

Fig. 3. Design of the three-stage restoration module. (a) Denoising stage structure. (b) Deblurring stage structure. (c) Design of multi-scale feature encoder. (d) Detail enhancement stage structure.

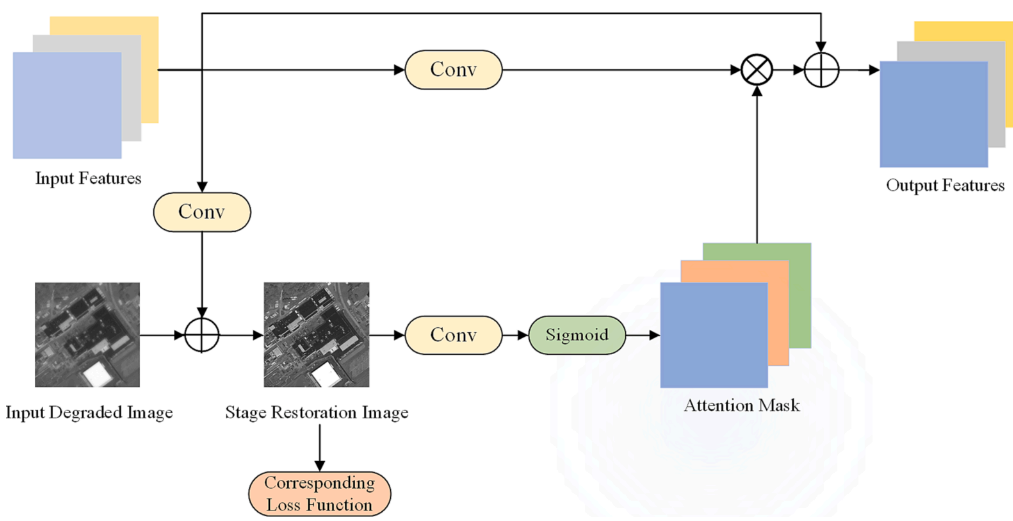

Fig. 4. Design of the attention supervision module.

the encoder network of the next level through edge cutting, so as to integrate the image information of different sensory fields into the encoder. This multi-scale information fusion strategy enables the network to capture and extract image features of different scales at each level, so as to adapt to different image degradation degrees and scale characteristics and improve the generalization and robustness of the network when processing various complex scenes and degradation phenomena in remote sensing images. The introduction of nested jump connections enables feature fusion at different depth levels, enhancing the network's ability to learn multi-scale information, and enabling subsequent decoders to better extract and learn the spatial feature information of input multi-scale remote sensing images.

Finally, the up-and-down sampling operation in the previous stage will inevitably lead to the loss of image details. At the same time, in the process of processing the output of the network in the deblurring stage, the false details generated by the network may be amplified, thus affecting the reliability of the whole multi-stage network for remote sensing image restoration. Therefore, in the network design of the detail enhancement stage (as shown in Fig. 3(d)), we adopt the residual network structure as the basic model and abandon the use of up-and-down sampling operation. We use the original degraded image and the output of the previous stage as the network input in this stage at the same time, increasing the real image spatial information that the network can learn, so as to improve the reliability of the final multi-stage network for remote sensing image detail restoration.

Differentiated intermediate supervision and adaptive structural adjustment module

Intermediate Supervision is a training strategy for convolutional neural networks, which can improve the training efficiency and accuracy of network models. The basic idea of intermediate supervision is to add some additional supervisory nodes in the middle layer of convolutional neural networks to introduce additional objective functions and loss functions in the training process, so as to help the network better learn the features and structure of input data, reduce the network depth, and accelerate the convergence speed in the training process of network models.

Therefore, we introduce the attention supervision module (as shown in Fig. 4), which improves the overall accuracy of the network by using intermediate supervision for each stage of the sub-network in a multi-stage progressive network. Secondly, the attention grid generated by

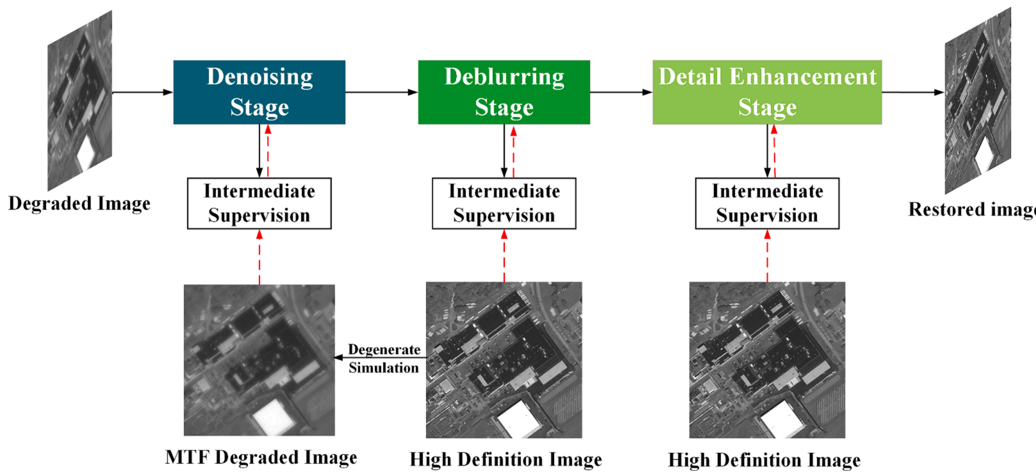

Fig. 5. Design of the differentiated intermediate supervision.

this module can suppress redundant features in the feature images generated in the previous step, improving the convergence speed of the network during the training process.

In order to enable each stage of the network to independently learn and train remote sensing image restoration tasks, we adopted a differentiated intermediate supervision method in the above-supervised attention module to adopt different supervision and optimization objectives for different stages of the network, so as to improve the overall restoration performance of the network. The principle of the differentiated intermediate supervision method is shown in Fig. 5. This method uses different images obtained from the degradation simulation method to calculate the loss function and backpropagation of parameters for different restoration stages.

Firstly, an imaging simulation method based on real remote sensing image degradation links is used to obtain final degraded images and MTF degraded images from the original high-definition images. Among them, the MTF degraded image refers to the degraded image before adding random noise during the simulation process. Afterward, through the intermediate supervision mechanism in the supervised attention module, an intermediate supervision based on MTF degraded images is applied to the denoising stage, and the original high-resolution images are used for intermediate supervision in both the deblurring and detail enhancement stages. The intermediate supervision process includes obtaining the restored image of the network in this stage based on the residual feature map output by the network, calculating the loss function, and backpropagation of weights.

The role of differentiated intermediate supervision is to ensure the completion of tasks in each stage of the image and to enable each stage to achieve different restoration tasks according to the design. Therefore, based on the prior information on the degradation process of remote sensing images, targeted adjustments are made to the network structure and network model parameters at different restoration stages in the network to improve the overall network's optimizability.

In addition, in actual remote sensing image restoration tasks, the remote sensing images to be restored may come from a large number of different remote sensing platforms, and their imaging process is also affected by different degradation factors. Therefore, it is difficult to determine the network structure of each image through manual judgment or feedback from the next level task. Meanwhile, the existing image restoration network parameters are determined during the loading process of the network model, so they cannot change with the changes in the characteristics of the input degraded remote sensing image during the training process. In order to solve the above problems, we propose an adaptive structure adjustment module based on dynamic network design ideas and design using the gating mechanism to achieve adaptive structure adjustment for each input remote sensing image

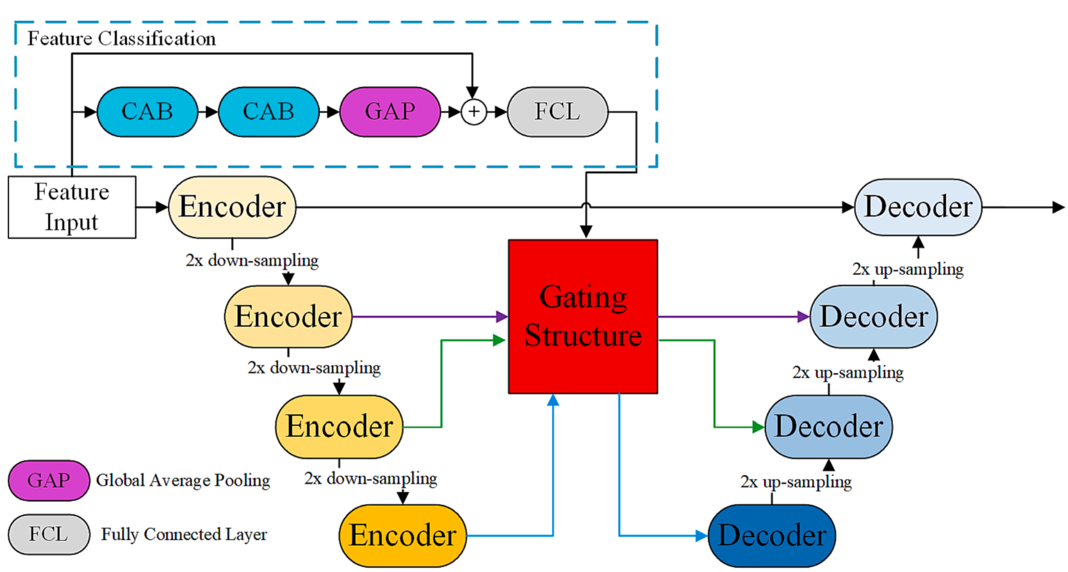

Fig. 6. Design of the adaptive structural adjustment module.

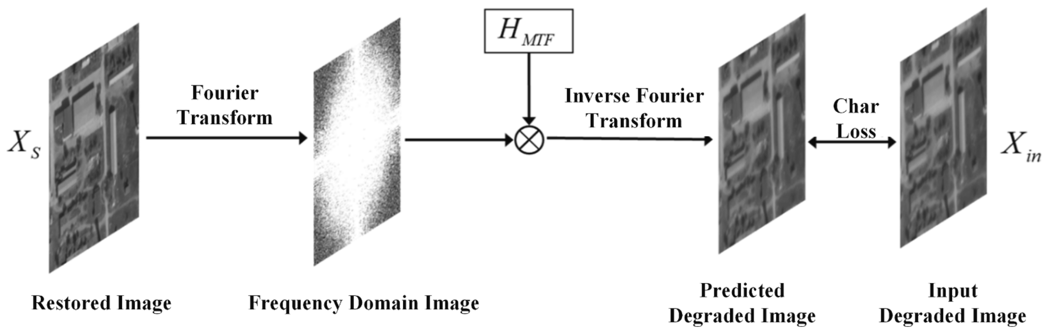

Fig. 7. Calculation principle of MTF loss function.

during training and testing.

We apply the module to the denoising stage and deblurring stage. This module adopts a multi-channel convolutional layer based on the channel attention module to learn advanced semantic information of remote sensing images from input feature maps (as shown in Fig. 6). Then, the image information extracted from the network is subjected to feature dimensionality reduction through global pooling, and the final classification result is generated by inputting residual connections into the fully connected layer to achieve classification of the input image based on the degradation characteristics. The feature classification results include three categories: priority denoising, priority deblurring, and equalization. Based on the classification results, we adopt a gating mechanism to adaptively adjust the network structure.

The gating structure consists of the fully connected layer and the softmax activation function, which generates corresponding weights to modulate the output of encoders at different levels and determines whether to transmit features to the corresponding decoders. In this process, the module adaptively determines the processing flow of feature information based on the characteristics of input features, achieving adaptive adjustment of the network structure and improving the generalization performance of the multi-stage restoration network.

Design of loss function based on modulation transfer function

Due to the inability of existing methods to utilize the system degradation prior information of remote sensing platforms, such as point spread function and modulation transfer function information, the restoration performance of the network significantly decreases when the degradation process of the input image is more complex and the degree of blur is high. Therefore, we design a loss function group based on optical priors to introduce prior information of the transfer function into the training process of the network through the MTF loss function, in order to improve the network's restoration performance.

Remote sensing platforms need to conduct regular onboard calibration to calibrate the sensors of remote sensing satellite payload in orbit, ensuring the accuracy and consistency of remote sensing data. The commonly used onboard calibration method currently uses a ground calibration pattern. During the calibration process, the fuzzy kernel of the remote sensing platform can be extracted using methods such as the knife-edge method, which makes the remote sensing image restoration task a type of non-blind restoration task.

In order to enable the above multi-stage restoration network to utilize the system degradation information, accelerate the convergence speed of network model training, and improve the image restoration effect of the network, we design an MTF loss function based on the system degradation prior, which is defined as follows:

$$L_{\mathrm{MTF}} = \sqrt{\left\| ifft(fft(X_{\mathrm{S}})\cdot H_{\mathrm{MTF}}) - X_{\mathrm{in}} \right\|^{2} + \varepsilon^{2}} \tag{1}$$

where *fft* and *ifft* represent the Fourier transform and inverse Fourier transform of the image, respectively, H_{MTF} represents the transfer function of the optical system. X_{S} is the image after the previous stage of network restoration, X_{in} is the low definition image output in the previous stage, and ε in the equation is a constant.

The principle of this loss function is to assume that the input degraded image is derived from the original image without degradation factors, which has undergone imaging degradation by the remote sensing optical system. The restored image output by the network is an estimate of the original image. According to the definition of fuzzy kernel function in remote sensing optical systems, there is a relationship between the two as follows:

$$X_{\mathrm{in}} = X_{\mathrm{S}} * h_{\mathrm{PSF}} \tag{2}$$

where them, h_{PSF} is the fuzzy kernel function of the remote sensing system. In practical algorithms, the transfer function of remote sensing systems is usually obtained. Therefore, by using the Fast Fourier Transform to map the process of image convolution to the frequency domain, the formula is transformed into:

$$X_{\mathrm{in}} = ifft(fft(X_{\mathrm{S}})\cdot H_{\mathrm{MTF}}) \tag{3}$$

Therefore, when evaluating the restoration effect of network restored image X_{S} in the restoration task, one can choose to calculate its Charbonnier loss with the input image X_{in}, and add a constant term ε to prevent potential gradient vanishing problems during the loss function calculation process, thus obtaining the definition formula of the MTF loss function mentioned above. The calculation principle is shown in Fig. 7:

After completing the design of the MTF loss function, it is necessary to combine other loss functions to constrain the network training to improve the performance of image restoration. We design a loss function group based on Charbonnier loss function, edge loss function, and MTF loss function, and optimize the loss function group according to the characteristics of different restoration tasks at different stages to further optimize the training process of the network model.

The Charbonnier loss function is defined as follows:

$$L_{\mathrm{char}} = \sqrt{\left\| X_{\mathrm{S}} - Y \right\|^{2} + \varepsilon^{2}} \tag{4}$$

where ε is a constant which is generally taken as 1×10^{-3}based on experience, X_{S} and Y are the restored image and real image respectively. This loss function is similar to traditional L1 and L2 loss functions, focusing on the pixel-by-pixel differences between the restored image and the real image. Its advantage is that it does not excessively smooth the image and to some extent avoids gradient vanishing and gradient explosion.

The Charbonnier loss function is defined as follows:

$$L_{\mathrm{edge}} = \sqrt{\left\| \Delta(X_{\mathrm{S}}) - \Delta(Y) \right\|^{2} + \varepsilon^{2}} \tag{5}$$

where Δ is the Laplacian operator, ε is a constant which is generally taken as 1×10^{-3}based on experience. This loss function is to protect the high-frequency texture structure information in the image to ensure the

Table 1
Parameter μ value of each stage loss function group.

Network Stage	μ Value
Denoising Stage	0.4
Deblurring Stage	0.8
Detail Enhancement Stage	0

restoration effect of image details.

Finally, we combine our proposed MTF loss function with the above loss functions to obtain the final network loss function group, which is defined as follows:

$$L = L_{\text{char}}(X_S, Y) + \lambda L_{\text{edge}}(X_S, Y) + \mu L_{MTF}(X_S, X_{\text{in}}) \tag{6}$$

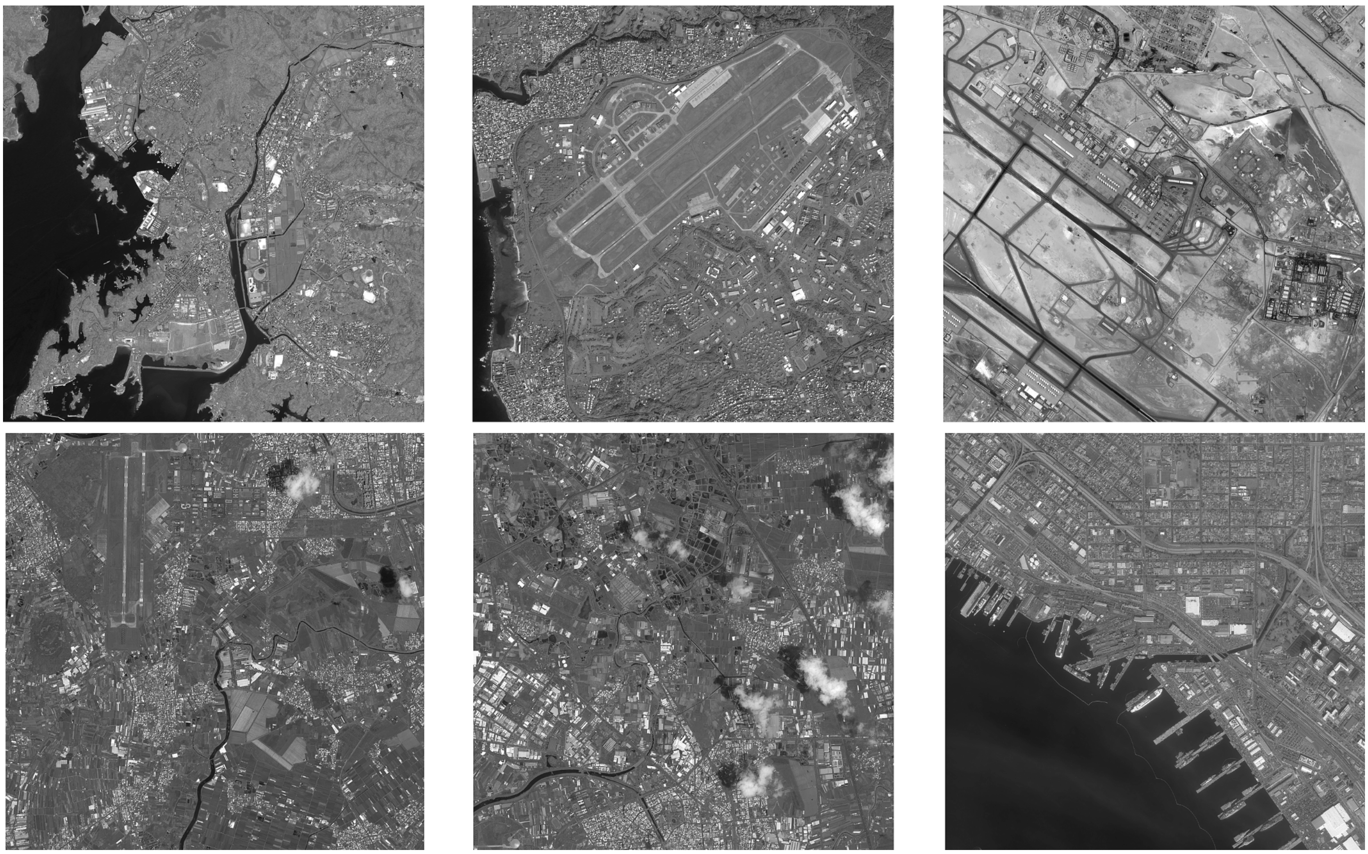

Fig. 8. Typical multi-resolution remote sensing images from different data sources.

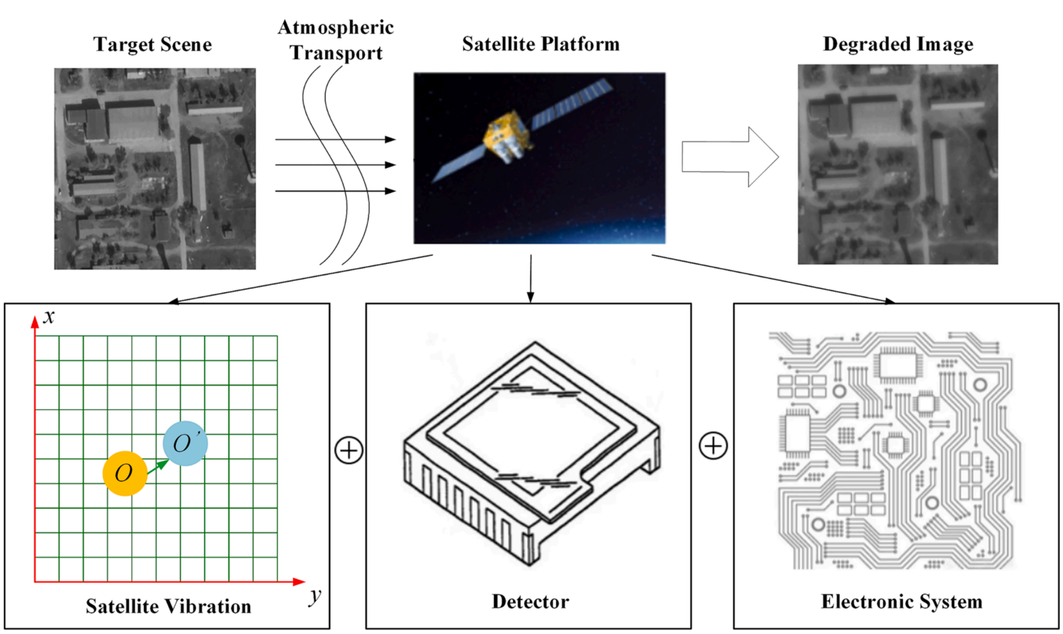

Fig. 9. Remote sensing platform imaging degradation process.

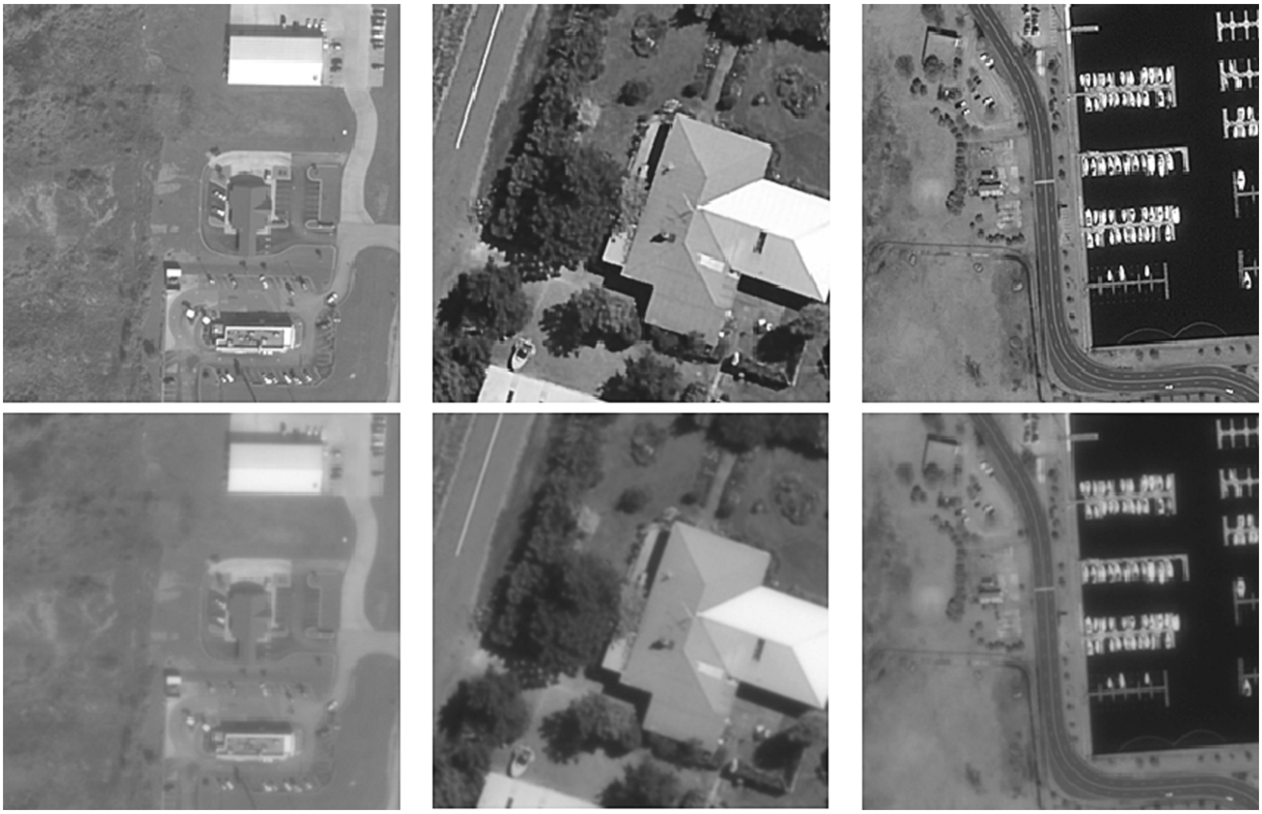

Fig. 10. Example of degradation simulation algorithm results. The upper images show the original high-definition images, while the lower images show the simulated degraded images.

where λ, μ are hyperparameters, determined by the needs of the restoration task in each stage and feedback during the actual training process. The parameter λ in Eq. (6) controls the relative importance of the Charbonnier loss, which is set to 0.05 as in [34]. In the detail enhancement stage, in order to improve the sharpening effect of image restoration, we set λ to 0.1. For parameter μ, it controls the proportion of the MTF loss, which value is shown in Table 1. In subsequent ablation experiments, we will explain how the value of μ is determined.

Experimental results

In this section, we introduce the experimental settings, including the dataset, evaluation metrics, and some implementation specifics. Then, based on the simulated remote sensing image dataset of the remote sensing platform, we conduct comparison experiments and ablation experiments to verify the effectiveness and reliability of the proposed method.

Datasets and evaluation metrics

In order to solve the problem of lacking a multi-platform comprehensive image restoration dataset based on real degradation model simulation in the field of remote sensing images, we construct a simulation model that includes the atmospheric transmission process, optical imaging process, and platform vibration process, and use image sources from various high-resolution remote sensing satellites to construct a remote sensing image restoration dataset that includes multiple scene types.

We use high-resolution visible light remote sensing satellites such as WorldView-2, WorldView-3, PL, Jilin 1A, Jilin 1B, and IKONOS as data sources for remote sensing image restoration datasets. The typical imaging results are shown in Fig. 8.

In order to improve the simulation performance of remote sensing image restoration datasets, we adopt a degraded image simulation method based on the real physical degradation model of the remote sensing platform. The degradation factors in the imaging process of remote sensing platforms include atmospheric transmission, optical system aberrations, detector and imaging electronics, and platform vibrations, as shown in Fig. 9.

And in order to increase the authenticity of the simulation process, a random MTF perturbation is added to each image during each MTF degradation process to simulate the complexity of the real degradation scene, thereby increasing the generalization and robustness of the network for image restoration tasks in different complex situations.

Using the above process to perform degradation simulation on high-resolution remote sensing images, the image size is set to 256 × 256 pixels, with a total of 5123 pairs of images. The training and testing sets are divided in a 9:1 ratio. An example of simulation results is shown in Fig. 10:

In order to evaluate the effectiveness of restoration algorithms on degraded remote sensing images, we use peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as evaluation metrics for algorithm performance. Among them, PSNR is defined as:

$$PSNR = 10\lg\left(L^2/MSE\right) \tag{7}$$

where L is the dynamic range of image pixel brightness, and MSE is the mean square error between two images.

SSIM is defined as:

$$SSIM(f,g) = \frac{(2\mu_f\mu_g + c_1)(2\sigma_{fg} + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)} \tag{8}$$

where $\mu_f\mu_f$ and σ_f^2 σ_g^2 respectively represent the mean and variance of the image f and g, represent the covariance of the two images; c_1 and c_2 are adjustable parameters.

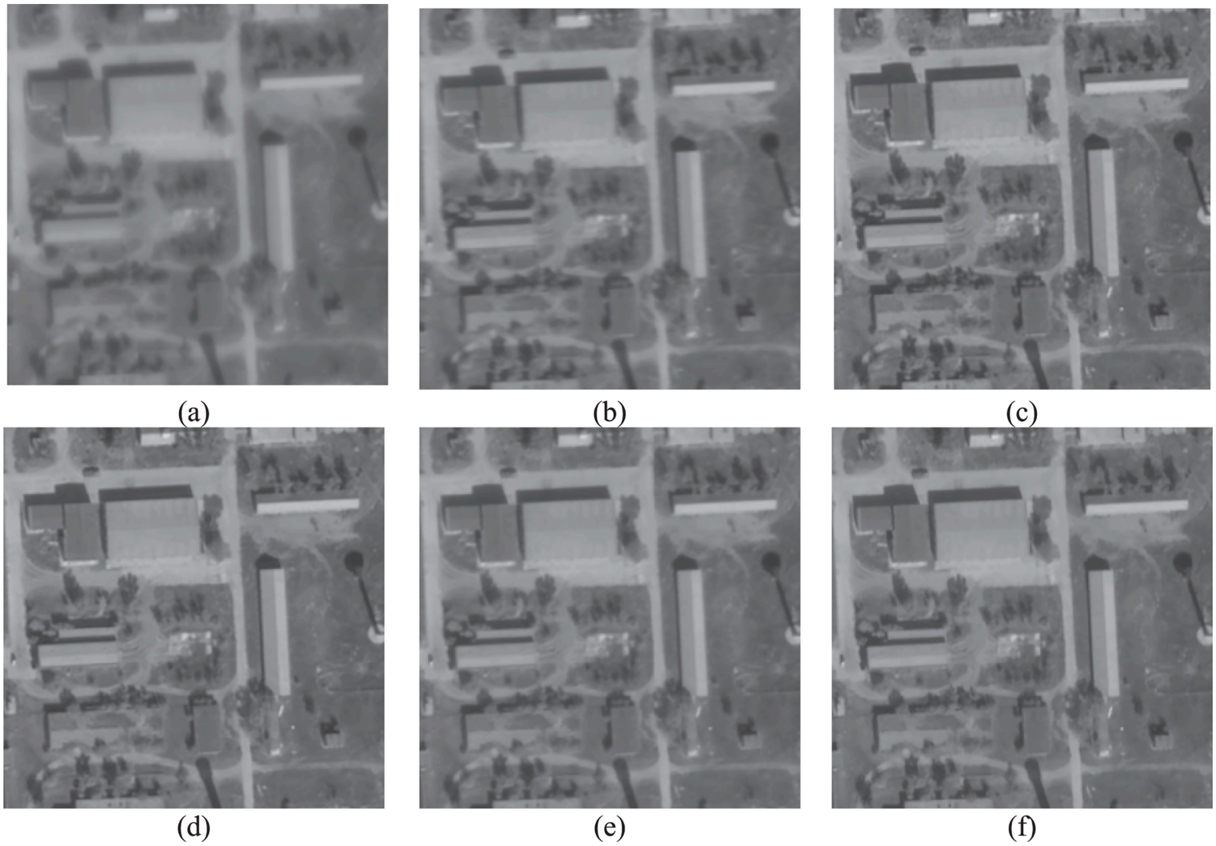

Fig. 11. Remote sensing image restoration results (background scene, 1.0 m image resolution). (a) Degraded Image. (b)Result of Wavelet Regularization. (c) Result of MIMO-Unet. (d) Result of DeblurGANv2. (e) Result of MPRNet. (f) Result of MARRNet(Ours).

Implementation details

Our MARRNet is implemented using the PyTorch framework and trained end-to-end on an Nvidia RTX 3080 GPU. In the comparison experiments, we set the batch size to 8 and trained the model with 500 epochs. We use the Adam optimizer with the initial learning rate of 2×10^{-4}, which is steadily decreased to 1×10^{-6} using the cosine annealing strategy. The random number seed of the model is fixed in the ablation experiment.

Comparison experiments

To evaluate the restoration performance of the proposed network under different resolutions and image source conditions, we test the restoration performance of our method under input image resolutions of 0.2 m, 0.7 m, and 1.0 m based on the aforementioned remote sensing image restoration dataset. The comparison experiments use the wavelet regularization method [26] as a traditional method and MIMO-Unet [33], DeblurGANv2 [32], and MPRNet [21] as deep learning methods to compare with our method. The experimental results at certain resolutions are shown in Fig. 11. and Fig. 12.

The evaluation metrics of each restoration algorithm are shown in Table 2. The processing ability of different image resolutions characterizes the generalization performance of the algorithm at different orbital heights.The experimental results show that our proposed MARRNet can effectively restore remote sensing images affected by complex degradation factors, and the restoration effect is subjectively superior to methods such as hybrid filtering algorithm and MIMO-UNet, and slightly superior to other deep learning methods. For typical targets at different resolutions, such as airplanes, roads, ships, etc. (as shown in Fig. 13.), our method can effectively restore image features, enhance the texture details of the targets, and thus effectively improve the performance of subsequent work such as target detection and recognition. Our proposed method achieves the optimal restoration effect under different resolutions, verifying the generalization ability of our method at different orbital heights and the enhancement ability for targets at different resolutions.

Table 3 shows the model parameter quantity and inference time of different comparison methods. The wavelet regularization method has no parameter result because it is a traditional method. From the table, our proposed method effectively improves restoration accuracy while maintaining good lightweight, achieving a good balance between performance and speed.

According to the above objective evaluation metrics, the restoration performance of the proposed algorithm under different resolutions is significantly better than the wavelet regularization method, MIMO-Unet, and DeblurGANv2, and slightly better than the MPRNet method based on the same multi-stage progressive network. The restoration effect of this network improves significantly with the decrease of image resolution and the gradual improvement of multi-scale characteristics, proving the effectiveness of the proposed restoration network and the multi-scale module in the network.

Ablation experiments

In order to evaluate the effectiveness of the algorithm modules proposed in this paper, the proposed modules are configured on the basic architecture. Their contribution to the final results is analyzed by comparing the image restoration performance of these modules before and after use. The experimental results are shown in Table 4:

According to the calculation results of the objective evaluation metrics obtained in the above experiments, the performance of the basic network has been improved to some extent after loading the above-improved modules. Among them, the differential intermediate

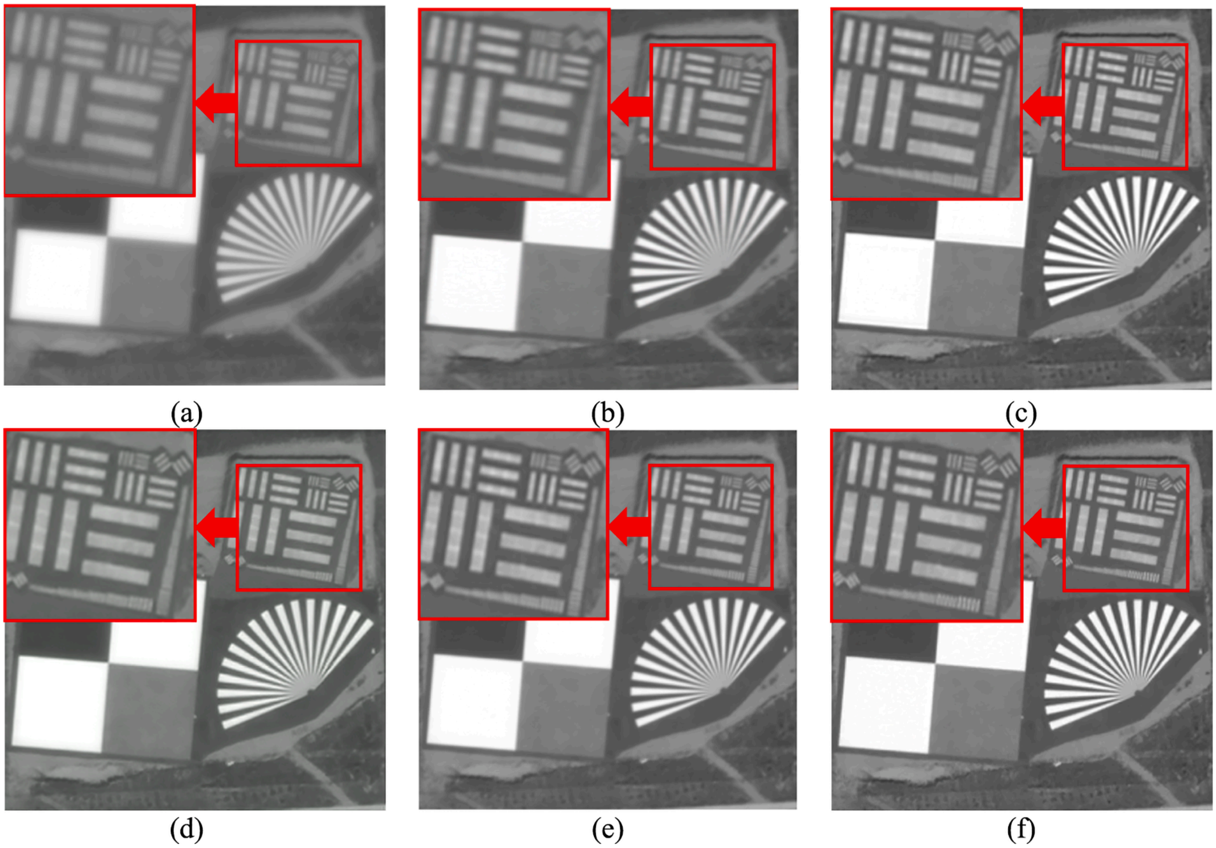

Fig. 12. Remote sensing image restoration results (ground calibration pattern, 0.7 m image resolution). (a) Degraded Image. (b)Result of Wavelet Regularization. (c) Result of MIMO-Unet. (d) Result of DeblurGANv2. (e) Result of MPRNet. (f) Result of MARRNet(Ours).

Table 2
Statistical results of evaluation metrics at different resolutions.

Method	0.2 *m*		0.7 *m*		1.0 *m*	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Degraded Image	26.23	0.8190	26.11	0.8193	26.03	0.8166
Wavelet Regularization	32.84	0.9117	32.79	0.9132	32.97	0.9126
MIMO-Unet	35.42	0.9419	35.62	0.9432	36.28	0.9475
DeblurGANv2	36.56	0.9468	36.79	0.9437	37.12	0.9487
MPRNet	**38.94**	0.9511	38.81	0.9553	39.03	0.9636
MARRNet(Ours)	38.71	**0.9629**	**39.27**	**0.9652**	**39.39**	**0.9713**

supervision (DIS) has the greatest improvement, fully proving the effectiveness of the multi-stage restoration structure and differential supervision strategy. The introduction of MTF Loss has also brought a certain improvement in restoration performance, achieving 1.72 % and 2.36 % improvements respectively in PSNR and SSIM, and proving the effectiveness of introducing system degradation prior. The adaptive network structure adjustment module (ASAM) has a relatively small improvement on the multi-stage progressive network when loaded separately. The reason is that due to the lack of DIS constraints, the network only focuses on the denoising task in the first stage, which reduces the whole restoration performance of the network. When the two modules are used together, the performance is improved significantly.

Finally, through the combination of three modules, a significant performance improvement is achieved compared to the baseline, with PSNR and SSIM improving by 7.18 % and 4.74 %, respectively. The experiments show that the algorithm after adding the modules can better restore image information and effectively improve image clarity, which proves the effectiveness of each algorithm module.

For the selection of the optimal value of μ in the loss function in each stage, we compare the generated results obtained by using different parameter μ in the current stage with the corresponding ground truth. The experimental results are shown in Table 5:

As shown in Table 5, since the deblurring stage is mainly responsible for restoring the impact of system degradation, the proportion of MTF loss in this stage is the highest, and the optimal parameter μ is 0.8. The denoising stage, as the previous stage of the deblurring stage, also has the impact of system degradation. Experiments show that using parameter μ of 0.4 in this stage can achieve the best restoration effect. In the detail enhancement stage, the main purpose is to restore the texture details of the image. At this point, adding MTF loss will actually dilute the attention to the details. Experiments show that using 0 as the parameter μ in this stage has the best restoration effect.

Conclusions

In this work, we propose a multi-stage adaptive image restoration method based on the degradation mechanism of remote-sensing images. Firstly, based on the degradation process of remote sensing images, a multi-stage network framework of “denoising deblurring detail enhancement” is constructed, which enhances the network’s ability to gradually restore spatial details of remote sensing images. Secondly, based on the remote sensing image degradation model, we introduce differential intermediate supervision and adaptive structural adjustment module to achieve targeted adjustment of the network based on image features, further improving the generalization performance for complex images. Finally, based on the system degradation prior to remote sensing images, a loss function based on MTF is proposed, which improves the network’s performance in restoring remote sensing images. Experiments show that our method has a significant performance improvement in remote sensing image restoration compared to other comparison

(a) 0.2m image resolution, airplanes (b) 0.7m image resolution, roads (c) 1.0m image resolution, ships

Fig. 13. Typical target restoration results at different resolutions. The upper images show the degraded images, while the lower images show the restoration results of our method.

Table 3
Model parameter quantity and inference time for different comparison methods.

Methods	Wavelet Regularization	MIMO-Unet	DeblurGANv2	MPRNet	MARRNet (Ours)
PSNR	32.97	36.28	37.12	39.03	39.39
#Params (M)	–	6.8	60.9	20.1	24.1
Time (s)	7.52	0.02	0.23	0.18	0.20

Table 4
Ablation experimental results on individual components.

Module combination	PSNR	Relative Change	SSIM	Relative Change
baseline	36.62	0 %	0.9214	0 %
+DIS	37.41	2.16 %	0.9440	2.45 %
+ASAM	37.18	1.53 %	0.9428	2.32 %
+MTF Loss	37.25	1.72 %	0.9432	2.36 %
+DIS & ASMA	38.42	4.92 %	0.9595	4.14 %
+ DIS & ASAM & MTF Loss	**39.25**	**7.18 %**	**0.9651**	**4.74 %**

Table 5
Ablation experimental results on the parameter μ value of each stage.

μ Value	Denoising Stage		Deblurring Stage		Detail Enhancement Stage	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
0	31.82	0.8842	32.94	0.9002	**38.03**	**0.9536**
0.2	32.55	0.8976	33.62	0.9053	37.88	0.9502
0.4	**33.17**	**0.9132**	34.47	0.9112	35.65	0.9325
0.6	32.94	0.9013	35.29	0.9263	34.84	0.9273
0.8	30.44	0.8839	**36.79**	**0.9437**	34.13	0.9205
1.0	29.32	0.8697	36.42	0.9395	32.72	0.9063

methods, and can better adapt to the problem of image degradation caused by remote sensing imaging processes.

However, our method currently relies more on system degradation prior information in the remote sensing imaging process and has poor adaptability to image restoration tasks in other fields or blind restoration tasks of remote sensing images. In future research, we will further optimize information input and network structure to enhance our processing capabilities for various restoration tasks.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) (62305086 62101160).

CRediT authorship contribution statement

Pengfei Zhang: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Jinnan Gong:** Writing – review & editing, Supervision, Investigation. **Shikai Jiang:** . **Tianjun Shi:** Validation. **Jiawei Yang:** Validation, Supervision. **Guangzhen Bao:** Validation. **Xiyang Zhi:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shikai Jiang reports financial support was provided by National Natural Science Foundation of China.

Data availability

The dataset established is available upon requests from the corresponding author.

References

[1] Antoni B, Bartomeu C, Morel J-M. A nonlocal algorithm for image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2005**, pp. 60–65.
[2] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 2007;16(8):2080–95.
[3] Dong W, Zhang L, Shi G, Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans Image Process 2011;20(7):1838–57.
[4] He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 2010;33(12):2341–53.
[5] Kim KI, Kwon Y. Single-image superresolution using sparse regression and natural image prior. IEEE Trans Pattern Anal Mach Intell 2010;32(6):1127–33.
[6] Rudin LI, Osher S, Fatemi E. Nonlinear Total Variation Based Noise Removal Algorithms 1992;60:259–68.
[7] Osher S, Burger M, Goldfarb D, Jinjun Xu, Yin W. An iterative regularization method for total variation-based image restoration. Multiscale Model Simul 2005;4(2):460–89.
[8] Weisheng D, Xin L, Lei Z, Guangming S. Sparsity-based image denoising via dictionary learning and structural clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2011**, pp. 457–464.
[9] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 2006;15(12):3736–45.
[10] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration. IEEE Trans Image Process 2007;17(1):53–69.
[11] Shuhang G, Lei Z, Wangmeng Z, Xiangchu F. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2014**, pp. 2862–2869.
[12] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. Adv. Neural Inf. Process. Syst; 2012. p. 1097–105.
[13] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, **2014**.
[14] Nah S, Kim TH, Lee KM. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proc. IEEE Conf. Comp. Vis. Patt. Recogn; 2017. p. 257–65.
[15] Tao X, Gao H, Shen X, Wang J, Jia J. Scale-recurrent network for deep image deblurring. In: Proc. IEEE Conf. Comp. Vis. Patt. Recogn; 2018. p. 8174–82.
[16] Ying T, Jian Y, Xiaoming L, Chunyan X. MemNet: A persistent memory network for image restoration. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), **2017**, pp. 4539–4547.
[17] Tim B, Ben M, Tianfan X, Jiawen C, Dillon S, Jonathan TB. Unprocessing images for learned raw denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2019**, pp. 11036–11045.
[18] Shi G, Zifei Y, Kai Z, Wangmeng Z, Lei Z. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2019**, pp. 1712–1722.
[19] Yoonsik K, Jae WS, Gu YP, Nam IC. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2020**, pp. 3482–3492.
[20] Zongsheng Y, Hongwei Y, Qian Z, Deyu M, Lei Z. Variational denoising network: Toward blind noise modeling and removal. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), **2019**, pp. 1688–1699.
[21] Syed WZ, Aditya A, Salman K, Munawar H, Fahad SK, Ming-Hsuan Y, Ling S. Multi-stage progressive image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2021**, pp. 14821–14831.
[22] Ding L, Bihan W, Yuchen F, Chen CL, Thomas SH. Non-local recurrent network for image restoration. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), **2018**, pp. 1673–1682.
[23] Mou C, Zhang J, Fan X, Liu H. and Ronggang Wang. Cola-net: Collaborative attention network for image restoration. IEEE Transactions on Multimedia; 2021.
[24] Chong M, Jian Z, Zhuoyuan W. Dynamic attentive graph learning for image restoration. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), **2021**, pp. 4328–4337.
[25] Richardson WH. Bayesian-based iterative method of image restoration. JoSA 1972;62(1):55–9.
[26] Dong B, Jiang Q, Shen Z. Image restoration: Wavelet frame shrinkage, nonlinear evolution pdes, and beyond. Multiscale Model Simul 2017;15(1):606–60.
[27] You YL, Kaveh M. A regularization approach to joint blur identification and image restoration. IEEE Trans Image Process 1996;5(3):416–28.
[28] Osher S, Burger M, Goldfarb D, et al. An iterative regularization method for total variation-based image restoration. Multiscale Model Simul 2005:460–89.
[29] Lefkimmiatis S, Bourquard A, Unser M. Hessian-based norm regularization for image restoration with biomedical applications. IEEE Trans Image Process 2011;21(3):983–95.
[30] Sun J, Cao W, Xu Z, Ponce J. Learning a convolutional neural network for non-uniform motion blur removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2015**, pp.769–777.
[31] 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 (CVPR), **2018**, pp. 8183-8192.
[32] Kupyn O, Martyniuk T, Wu J, Wang Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2019**, pp. 8878-8887.
[33] Cho SJ, Ji, SW, Hong JP, Jung SW, Ko, SJ. Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), **2021**, pp. 4641-4650.
[34] Kui J, Zhongyuan W, Peng Y, Baojin H, Yimin L, Jiayi M, Junjun J. Multi-scale progressive fusion network for single image deraining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2020**. pp. 2-6.
[35] Ning Bu, Hui M, Liu M, Dong L, Kong L, Zhao Y. Image restoration for optical synthetic aperture system via variational physics-informed network. Results Phys 2023;52:106878.
[36] Hui M, Li X, Zhang H, Liu M, Dong L, Kong L, et al. Image restoration of optical sparse aperture systems based on a dual target network. Results Phys 2020;19:103429.
[37] Wang H, Sun S, Ren P. Meta underwater camera: A smart protocol for underwater image enhancement. ISPRS J Photogramm Remote Sens 2023;195:462–81.
[38] Wang H, Sun S, Bai X, Wang J, Ren P. A reinforcement learning paradigm of configuring visual enhancement for object detection in underwater scenes. IEEE J Ocean Eng 2023;48(2):443–61.
[39] Wang H, Sun S, Ren P. Underwater color disparities: Cues for enhancing underwater images toward natural color consistencies. IEEE Trans Circuits Syst Video Technol 2023.