

Improving Optical Remote Sensing Image Quality Through Random Degradation and Adaptive Fusion Super-Resolution Networks

Lijun Sun^{ID}, Jiping Bi^{ID}, Yongchao Song^{ID}, Zhaowei Liu^{ID}, Member, IEEE,
and Xuan Wang^{ID}, Senior Member, IEEE

Abstract— High-resolution (HR) optical remote sensing images are the guarantee for remote sensing image analysis and application. However, many images suffer from blurring, distortion, and low resolution (LR) due to camera hardware limitations and unstable image transmission. To address these challenges, we propose a random degradation and adaptive fusion-based super-resolution (SR) network [RDAF-generative adversarial network (GAN)] for improving the clarity and detail of images. Specifically, unlike the traditional single degradation method, we design a comprehensive simulation model for remote sensing image degradation. It aims to generate LR remote sensing images that are closer to the real scene. Subsequently, these generated LR images are fed into RDAF-GAN for reconstruction to recover finer and more accurate image details. In addition, we propose an image fusion method based on local contrast. By adaptively adjusting the fusion weights, the perceived clarity and visual quality of the images are further enhanced. The experimental results validate that RDAF-GAN outperforms other state-of-the-art (SOTA) methods and consistently produces excellent results in a variety of situations.

Index Terms— Adaptive fusion, generative adversarial network, optical remote sensing image, random degradation, super-resolution (SR).

I. INTRODUCTION

OPTICAL remote sensing images are critical in various applications such as environmental monitoring, urban planning, agricultural management, and other fields. Nevertheless, they are frequently characterized by lower spatial resolution due to challenges such as sensor limitations or bandwidth restrictions. It limits its application to detailed recognition and quantitative analysis, such as target detection [1]. Super-resolution (SR) reconstruction techniques are critical in the realm of image processing [2]. It improves the spatial resolution of an image by recovering high-resolution (HR) detailed information from low resolution (LR). By doing so, it enhances the detailed visibility and information richness

Received 21 May 2025; revised 7 July 2025; accepted 8 July 2025. Date of publication 11 July 2025; date of current version 21 July 2025. This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR2022QF037, in part by the Graduate Innovation Foundation of Yantai University under Grant GGIFYTU2542 and Grant KGIFYTU2514, and in part by the School and Locality Integration Development Project of Yantai City under Grant 2022. (Corresponding author: Yongchao Song.)

The authors are with the School of Computer and Control Engineering, Yantai University, Yantai 264005, China (e-mail: sunlijun@s.ytu.edu.cn; bijiping@s.ytu.edu.cn; yesong@ytu.edu.cn; lzw@ytu.edu.cn; xuanwang91@ytu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2025.3588168

of the image [3], [4]. While SR techniques have demonstrated remarkable success in medical imaging [5], [6] and surveillance systems [7], remote sensing SR faces unique challenges that demand specialized solutions [8]. Many approaches primarily focus on maximizing the peak signal-to-noise ratio (PSNR) [9], often overlooking visual quality and the recovery of high-frequency textures. Therefore, the main motivation of this work is to strike a balance between achieving a high PSNR and enhancing the perceptual quality of images. Improving SR techniques for remote sensing images not only increases spatial resolution, but also ensures that the reconstructed images more accurately represent real-world textures and details.

Different strategies have been used to enhance image quality [10], [11], [12], [13], [14]. SRCNN [11] uses neural networks to learn the relationship between bicubic interpolated LR images and HR images, improving reconstruction but with high computational overhead. Other models like VDSR [12] and EDSR [13] also rely on large datasets to learn the LR–HR mapping. RCAN [14] improves performance by using residual in residual structures, but it negatively affects image texture details.

The quality of reconstructed images is, to some degree, impacted by image degradation [17]. Despite existing image degradation methods, they cannot accurately simulate the degradation of remote sensing images in the real world. Bicubic downsampling is an established pattern that generates synthetic detail, and its degradation is unrealistic. Liu et al. [18] proposed a method of self-supervised degradation through comparative learning to enhance the robustness and discriminatory nature of the embedding space. Wang et al. [19] pioneered a model that employs a higher order modeling process and integrates the U-Net discriminator to enhance the discriminator's capacity. An unsupervised adaptive degradation-aware network was introduced by the seminal work of degradation-aware super-resolution [20]. It draws on representation learning techniques and is adept at managing various degradation scenarios. The respected article BSRDM [21] suggested an innovative patch-based noise model, which amplifies the flexibility of noise representation. And it employs an effective kernel generator for novel creation intending to enhance the proficiency of the model.

Even though all the above methods achieved some reconstruction results, the details recovered from Fig. 1(b)–(d) are somewhat distorted. Substantial opportunities for advancement

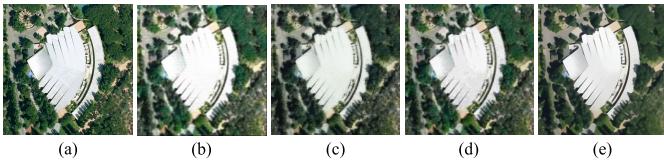


Fig. 1. Comparison of the effects of different reconstruction methods. (a) HR. (b)-(d) Reconstruction results of Bicubic [12], SRGAN [15], and EDiffSR [16], respectively. (e) Our approach.

still exist in remote sensing image reconstruction. First, the degradation of remote sensing images in the real world is affected by many factors: atmospheric conditions, sensor vibrations, and seasonal variations [22]. The various factors that contribute to the blurring of images and the limitations on resolution can lead to reductions in the amount of information present and the accuracy of remote sensing images. Second, remote sensing images contain a wide range of complex and informative backgrounds. In addition to providing a priori information for the SR task, it can also potentially render the task more difficult. In essence, significant strides in the research and development of degradation simulation and reconstruction algorithms are of paramount importance for enhancing the overall quality and restoring the authentic details of remote sensing images.

To tackle these challenges, we draw inspiration from BSRGAN [23] and propose a novel stochastic degradation strategy designed for optical remote sensing imagery. To simulate degradation and generate LR, various techniques such as blurring, downsampling, and noise are employed randomly. The stochastic degradation strategy is not merely a combination of images. These degradation operations aim to mimic the changes in image quality, which can be attributed to various environmental factors, such as varying light intensities, noise levels, and image resolutions, which can be found in remote sensing images. By leveraging this methodology, we can generate the necessary training data to subsequently train the neural network.

Regarding architectural innovation, RDAF-generative adversarial network (GAN) differs significantly from BSRGAN [23] in two key areas. First, in the degradation modeling stage, we design a content-aware stochastic degradation strategy based on the physical characteristics of remote sensing. Combining the atmospheric turbulence model, sensor noise characteristics, and platform motion, a degradation space that is more in line with the real remote sensing scene is constructed. Second, in the reconstruction stage, random degradation and adaptive fusion-based SR network (RDAF-GAN) introduces an adaptive fusion mechanism. This mechanism utilizes dynamic weight allocation to achieve synergistic optimization of low-frequency structures and high-frequency textures. As a result, significant progress has been achieved in image reconstruction quality and detail recovery. These innovations effectively enhance the processing capability of remote sensing image SR tasks.

For the reconstruction, we trained a novel model constructed with GAN. Given a pretrained HR image, it undergoes a stochastic degradation process that generates a corresponding

simulated LR image. The data pairs are, then, utilized for network training. The reconstructed image is adaptively fused to improve the perceptual quality of the image. Moreover, we expand the number of network layers within the generator network to improve its ability to learn the mapping relationship between HR images and LR images. Through this joint training strategy, the network can build a more profound and intricate understanding of the interplay between the two, subsequently boosting the precision of image reconstruction. Conventional methods [18], [24] typically utilize only LRs as inputs, which impedes the learning capabilities and reconstruction quality of the network. The present method, on the other hand, makes full use of the original HR image as a reference, providing richer information to guide the reconstruction process. Through continuous iteration and optimization, the network gradually learns how to recover the HR image details and features from the LR image. Utilizing image fusion technology enables more accurate and detailed remote sensing images, satisfying the requirements of advanced-level applications.

The main contributions of this article are given as follows.

- 1) A more suitable simulation model is proposed for efficiently simulating the degradation of remote sensing images. Three major factors, namely, ambiguity, down-sampling, and noise, are considered, and a stochastic degradation strategy is applied to improve the robustness of the model.
- 2) We propose the RDAF-GAN for the SR of optical remote sensing images. Adaptive fusion is applied to the initial output to achieve high-quality and visually appealing results.

II. RELATED WORK

This work investigates the application of GAN-based SR processing in remote sensing and the usage of models to simulate their degradation. This section will introduce the main concepts and approaches used to set the scene for this article.

A. Conventional Image Degradation Methods

As covered in Section I, bicubic downsampling and a few other standard approaches are the mainstays of current image degradation techniques. An innovative dyadic regression approach [25] is proposed to limit the search space for possible functions by imposing additional constraints on LR data. Differently, KernelGAN [26] utilizes a deep linear network and certain regularization loss to limit the dimensionality of the kernel space. Flow-based kernel prior (FKP) [27] approaches flow models in deep generative models by developing an invertible transformation to model the kernel.

A batch normalization (BN) layer is usually included in traditional residual networks (RNs) to accelerate convergence and prevent the problem of vanishing or exploding gradients. It was discovered in EDSR [13], though, as the BN layer may occasionally restrict the model's capacity to learn. Consequently, EDSR improves the SR performance by increasing the network depth and removing the BN layer from the residual block.

Liang et al. [28] employed a transformer machine mechanism to advance single-image SR technology. However, when the image degradation patterns in the test data do not match those in the training data, they may produce problems such as oversharpened or blurred results. The phenomenon is because the model is overly dependent on specific degradation patterns during training and does not adapt well to other types of degradation. For example, models trained under common degradation models such as bicubic interpolation and fuzzy reduction may perform well under these models but may produce undesirable results under other types of degradation models.

B. Unknown Methods of Degradation

Although these methods have yielded some success in addressing the degradation problem in image processing, they also have limitations, such as inadequate model robustness and generalization performance, the requirement for additional computational resources and training time, and the complexity involved in learning the parameters.

In recent years, scholars have been continuously striving to develop effective methods to solve various degradation problems. Degradation-adaptive regression SR (DARSR) [29], an innovative unsupervised approach, has been devised to learn correction filters. It adjusts the unknown degradation to a known degradation through filter correction. Degradation-adaptive SR (DASR) [30] utilizes a very compact regression network to predict the degradation parameters of the input image. Real-ESRGAN [19], a high-order degeneracy model based on the classic first-order degeneracy process, also makes significant progress. This model generates images that simulate real degradation and utilizes them as training data for the network. A framework was proposed by Ji et al. [31] to estimate kernel and noise distributions by learning the degradation factors of real images. In deep constrained least squares for blind image SR (DCLS-SR) [32], the authors introduce dynamic deep linear filters and redefine the degeneracy model. The reconstructed kernel weights are adaptively generated based on the filter and LR. A color-guided domain mapping network was introduced by Wei et al. [33]. It incorporates spatial attention and adaptive instance normalization, intending to alleviate the color bias problem caused by GAN.

Despite the significant advances made by these methods, some complex degradation models, such as [34] and [35], still struggle to accurately represent the noise distribution in remote sensing images. The actual remote sensing images, acquired via sensors, often contain noise. This noise can originate from a variety of sources, including the sensor itself and atmospheric conditions. Typically, these noises exhibit high correlation and nonuniformity. A mismatch in noise representation can result in performance degradation, irrespective of whether the fuzziness can be precisely represented or not.

C. SR Reconstruction Methods

Recent advancements in deep learning have led to the development of various SR models. Nonblind SR models [11], [12] achieve better reconstruction by utilizing HR images,

but may suffer from overfitting due to HR and reference data constraints. The blind SR model is more suitable for situations where no HR images or reference images are available, as opposed to the nonblind SR model. An earlier study [36] utilized an end-to-end learning method, employing a backpropagation algorithm, to automatically discover and identify image features. Deep learning advances have spurred numerous blind SR models that employ CNNs to extract hierarchical image features, enabling high-quality reconstruction with preserved structures and details.

The MPRNet model [37] uses progressive regression to learn image details and structure layer by layer. TTST [38] improves multiscale feature aggregation and global contextual attention. Additionally, Xiao et al. [16] developed an SR algorithm for remote sensing images based on the diffusion probabilistic model (EDiffSR). Regarding architectural innovations, HAUNet [39] hybridizes attention with U-Net to efficiently capture multiscale global features. SPIFFNet [40] introduces transformer-based cross-space pixel integration and cross-stage feature fusion for enhanced contextual modeling. However, these methods may still fail to adequately recover details when dealing with complex or blurred images. However, these methods still struggle to recover sufficient detail when dealing with complex or blurred images. In hyperspectral imaging, M2U-Net [41] combines deep image prior and degradation modeling, and EDIP-Net [42] introduces a two-stage enhanced prior network for better learning. Despite these advancements, challenges remain in recovering details from complex or blurred images.

To mitigate the risk that misleading details generated by GANs may undermine the overall realism of the generated image, Li et al. [43] developed a novel strategy called region-aware adversarial learning. It allows the model to fine-tune detail generation, specifically within textured regions, to deliver a more comprehensive and detailed output. SAGAN [44] employs a novel second-order channel attention mechanism and regional nonlocal modules to fully utilize the a priori information in LR and suppress artifacts. DRGAN [45] proposes a new dense RN to recover image details.

Unlike these methods, we introduce a novel approach of stochastic degradation and adaptive fusion to guide the network to recover the realistic details of the image. The random degradation increases the diversity of degradation compared to fixed degradation kernels. A wider range of data variations can be simulated, improving the robustness and generalization of the algorithm. Additionally, adaptive fusion improves visualization by increasing image quality and detail retention.

III. METHOD

In this section, we will focus on the proposed RDAF-GAN network, and the overall framework of the network model is shown in Fig. 2. The network is primarily composed of three key elements: random degradation, reconstruction network, and adaptive fusion. First, an LR image is generated from an HR image using image preprocessing techniques. The resulting LR images are, then, input into the generator section for reconstruction. Next, a discriminator is introduced

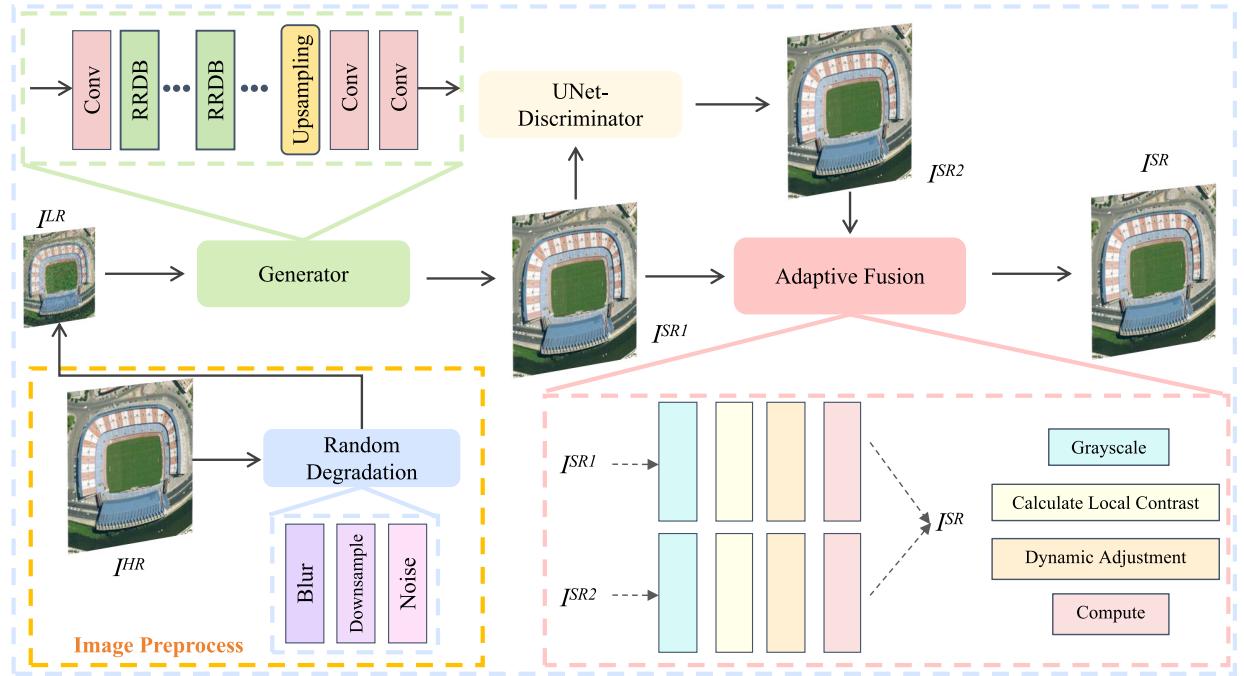


Fig. 2. Overall framework of the proposed RDAF-GAN.

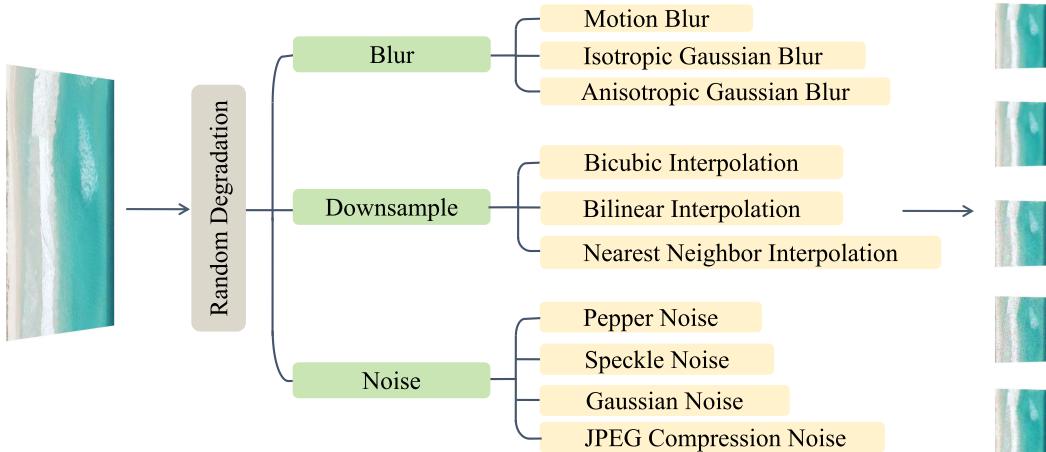


Fig. 3. Several simulation processes are used to recreate the degradation of remote sensing images. These processes are applied to the image at random, creating a wide range of LR that may accurately represent the natural deterioration of actual remote sensing images.

to evaluate the reconstruction results, ensuring the quality of the generated images. Finally, the image produced by the generator is combined with the image assessed by the discriminator through an adaptive fusion mechanism to optimize the quality of the final output. With this structure, the RDAF-GAN network ensures high-quality generated images while effectively utilizing stochastic degradation and adaptive fusion techniques to enhance image reconstruction.

A. Random Degradation

A random strategy is applied to mimic the image degradation process, providing the training data needed to create LR. The operations that are included in the framework are

illustrated in Fig. 3. The specific method includes a series of degradation operations on the input image such as blurring, downsampling, and adding random noise to generate an LR that matches the actual scene. The stochastic degradation strategy is not merely the combination of different images. These degradation operations reflect the effects of diverse factors on the quality changes of images under varied environments. The degradation process is described as follows:

$$D(I_{HR}, \theta) = \mathcal{R}(B, N, S)(I_{HR}, \theta) \quad (1)$$

$$\mathcal{R}(B, N, S) = \prod_{i=1}^M O_i \quad (2)$$

where $D(\cdot, \theta)$ is the degeneracy function and θ is the degeneracy parameter. \mathcal{R} denotes the function of random selection and superimposed degradation operations. B , N , and S denote the blur, noise, and downsampling addition operations, respectively. M is the number of degenerate operations, which can be randomly generated. \prod denotes the combination of operations. O_i is one or more degenerate operations chosen at random from B , N , and S .

The fuzzy operation of this method employs several methods including motion blur, isotropic Gaussian blur, and anisotropic Gaussian blur. These methods enable the simulation of various blurring effects that occur during the capture, transmission, or display of real-world images. For the motion blur effect, an image is simulated as it moves by utilizing a blurring kernel to create a blurring effect. This is achieved by first establishing the blur direction and length, after which a blur kernel is calculated and applied to the input image. When using Gaussian blur, the image is blurred using a smooth Gaussian kernel, and the amount of blurring is adjustable by varying the kernel's size and standard deviation. The process is shown as follows:

$$B(I_{\text{HR}}, \theta_B) = \sum_{k \in K} \omega_k \cdot (I_{\text{HR}} * k) \quad (3)$$

where K is the set of fuzzy kernels. K consists of three parts: the motion fuzzy kernel, the isotropic Gaussian kernel, and the anisotropic Gaussian kernel. The motion fuzzy kernel is designed based on the satellite's motion characteristics, with adjustable fuzzy direction angles and lengths for physical accuracy. The isotropic Gaussian kernel assumes uniform blurring in all directions, ideal for atmospheric turbulence effects. In contrast, the anisotropic Gaussian kernel accounts for directional blurring differences, providing a more accurate representation of blurring in imaging systems. ω_k is the weight of each fuzzy kernel, which can be generated randomly. $*$ denotes the convolution operation.

Noise is a typical component of picture deterioration, as it not only degrades the clarity of images but also has the potential to cause fine detail loss. Pretzel noise, speckle noise, Gaussian noise, JPEG compression noise, and other forms of noise are included to mimic the deterioration of photographs. By adding these sounds, simulations that mimic different disruptions and deteriorations that the image is normally exposed to will become far more realistic. In particular, it can be defined by

$$N(I, \theta_N) = I + \sum_{n \in N} \omega_n \cdot n \quad (4)$$

where I can be an image processed by a degradation operation or an original HR image (I_{HR}). N is the set of noise types. n is some type of noise. ω_n is the weight of each noise type. The weights ω_n for each noise type are first sampled from a uniform distribution $U(0, 1)$ and then normalized by softmax to ensure that $\sum \omega_n = 1$.

An interpolation technique is executed throughout the down-sampling procedure to create a new, reduced image from the original image. Using the original pixel that is closest to each target pixel and using that value as the target pixel

itself, nearest neighbor interpolation is a quick and easy technique. Two popular techniques for approximating data are bilinear and bicubic interpolation. They use a thorough computation method, taking into account the data from both their closer-by and farther-flung neighbors. The goal is to guarantee the preservation of the image's general composition and texture details. The image's resolution can be changed by up sampling and downsampling methods, which can also increase or decrease it

$$S(I, \theta_S) = \text{Interp}(I, f) \quad (5)$$

where $\text{Interp}(\cdot)$ denotes the interpolation function. f is the randomly selected interpolation method. We randomly select an interpolation method from nearest neighbor, bilinear, and bicubic interpolation with probabilities of 0.2, 0.3, and 0.5, respectively. Bicubic interpolation is preferred to reduce the occurrence of aliasing artifacts. For downsampling, we use uniform sampling from a range of 2–4 to address common resolution gaps in remote sensing scenes.

Our stochastic degradation strategy simulates the degradation of remote sensing images by combining various operations, resulting in diverse LR images that enhance model generalization. Unlike single bicubic downsampling and BSRGAN [23], our method accurately models the degradation process by incorporating blurring, downsampling, and noise addition, making it more flexible and comprehensive.

Regarding the physical modeling, we have designed a fuzzy model that meets the actual situation for the degradation characteristics of remote sensing images. Compared with the generalized fuzzy kernel used in BSRGAN, it is more effective in modeling platform motion and atmospheric effects. For the noise model, we innovatively integrate pretzel noise, speckle noise, and sensor characteristic noise. Compared with the single Gaussian noise used by BSRGAN, it is closer to the actual scene of remote sensing images. Meanwhile, for adaptive adjustment, the dynamic adjustment of degradation intensity is realized by random weights in (3) and (4), which break through the limitation of traditional fixed parameters.

B. Image Reconstruction Network

The ideal reconstruction method should be efficient and effective. However, the degradation of images can be complex, particularly when considering the limitations of computational resources. The majority of current SR methods are not versatile enough to handle different types and degrees of image degradation. Our reconstruction module is built on residual in residual dense block (RRDB) [46], which effectively boosts the model's performance and convergence time by combining the concepts of dense join and residual join.

In the process of creating effective and efficient reconstruction models, we have developed a novel RDAF-GAN network. It blends stochastic degeneracy with adaptive fusion, incorporating advanced technologies such as GAN to generate improved SR results. Specifically, our approach entails the generation of highly relevant LR remote sensing images through a process of stochastic degradation. This degradation process creates a more realistic input for our generator, thus

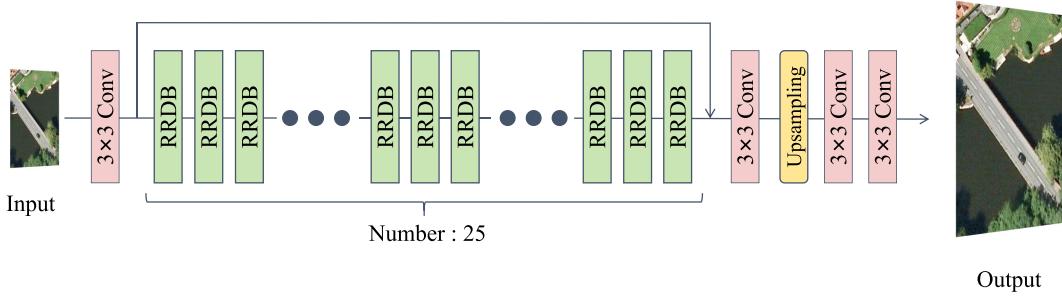


Fig. 4. Network framework of the generator framework for RDAF-GAN.

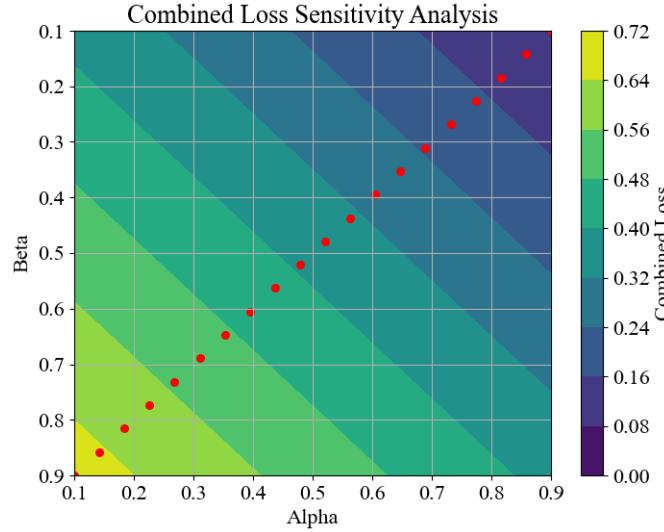


Fig. 5. Sensitivity analysis of two weighting parameters in the combined loss function.

leading to better training results. In parallel, we utilize the LR images to train our generator and GAN model, ensuring that the network can learn and generalize effectively. Once the training is completed, we use fusion techniques to improve the overall image quality of the reconstructed images, resulting in a more realistic and visually pleasing output.

We construct the generator using RRDB as a benchmark. Fig. 4 presents the structure of this approach. Shallow feature extraction is performed by a 3×3 convolution. Then, stacking the RRDB is performed for deep feature extraction. By doing so, we can provide more in-depth learning of the image features and detailed information. Each subblock can be viewed as a feature extractor, capable of generating higher level features as we stack more subblocks. These features are, then, combined to produce SR images with improved quality.

For the discriminator, a UNet [47] with a unique encoding-decoding structure is used. It can extract and fuse features at different scales, preserving rich spatial information. Its skip connections enhance the ability to recognize subtle differences between the generated image and the real image. This not only improves the discriminator's accuracy, but also enhances the overall performance and stability of the generated model.

C. Combined Loss Function

The training process of the generator employs a unique blend of the L_1 loss function and Charbonnier loss function, which combines the detail fidelity and smoothness of the image. This method can enhance the quality of the resulting HR image. The L_1 loss function encourages the generator to generate the highest level of detail and texture information possible, aiming to match the real image. On the other hand, the Charbonnier loss function aids in reducing noise and unreality and boosting the smoothness of the image.

Integrating these two loss functions can enable better control over the generator's feature learning and subtle alterations in remote sensing images, thereby enhancing the efficiency of SR reconstruction. It also boosts the diversity and robustness of the generator training process, thereby boosting the adaptability of the generator to various types of remote sensing images. The application of this combined loss function is a significant improvement to SR algorithms, offering a more advanced solution for the fields of remote sensing image processing and computer vision.

The Charbonnier loss function formula is

$$L_{\text{Char}} = \sqrt{\epsilon + (I_{\text{HR}} - I_{\text{SR}})^2} \quad (6)$$

where I_{HR} represents the real HR image, I_{SR} represents the SR result of the LR output by the generator, and ϵ is a very small constant that is used to avoid the case where the denominator of the square root is calculated to be zero.

The L_1 loss function equation is

$$L_{L1} = \|I_{\text{HR}} - I_{\text{SR}}\| \quad (7)$$

where $\|\cdot\|$ denotes the sum of absolute values of each element in the vector. It measures the difference between the generated image and the real HR image at pixel level.

The loss function of the combination is

$$L_{\text{total}} = \alpha L_{L1} + \beta L_{\text{Char}} \quad (8)$$

where α and β represent the proportion of weights of the two loss functions, respectively. L_{L1} and L_{Char} denote the L_1 and Charbonnier loss values, respectively.

According to Fig. 5, the variation of the combined loss function can be observed for different combinations of alpha and beta values. The red dots on the graph indicate the evaluation of the effect under various combinations of parameters. Various hue blocks represent correspondingly varying

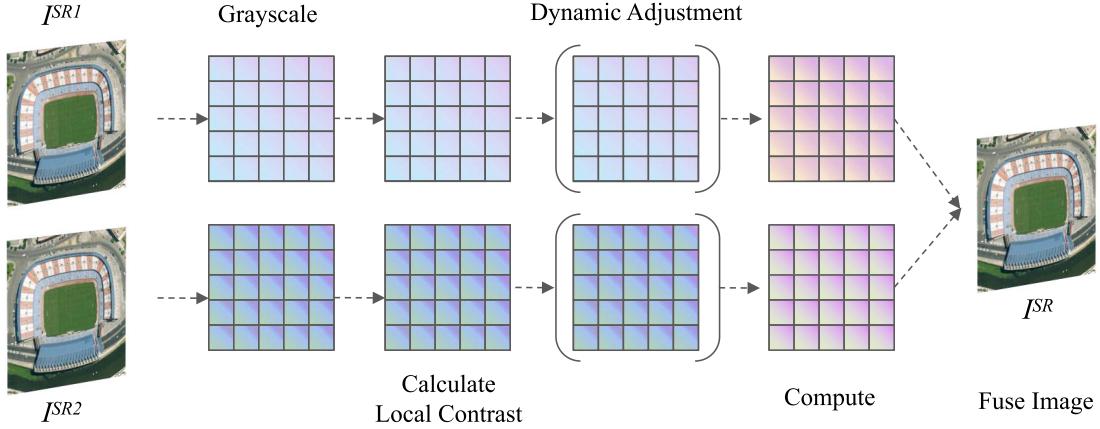


Fig. 6. Adaptive fusion framework for RDAF-GAN.

loss values. Other combinations of α and β may lead to better or worse training results. For outliers, the Charbonnier loss is more robust. Hence, the Charbonnier loss is more resilient to the majority of noise. $L1$ loss, however, can offer more precise information regarding the gradient. Thus, it is possible to balance the requirements for gradient guidance and robustness while optimizing the training impact by the appropriate adjustment of α and β value combinations.

D. Adaptive Fusion

To create more diverse and visually appealing images, we innovatively fused the outputs of the generator and the GAN model. As shown in Fig. 6, I^{SR1} and I^{SR2} represent the images generated by the generator and RDAF-GAN, respectively. To do this, we performed the following steps for adaptive fusion: first, we convert the input color image to grayscale. Next, Gaussian blurring is applied to the grayscale image to reduce the effect of noise. Then, the local contrast is calculated by weighted summation. The fusion weights are dynamically adjusted by calculating the mean value of the local contrast. In particular, the fusion weights of the images I^{SR1} and I^{SR2} can be defined as

$$\text{Adaptive}_{\alpha} = \frac{\text{mean}(lc_{I^{SR1}})}{\text{mean}(lc_{I^{SR1}}) + \text{mean}(lc_{I^{SR2}})} \quad (9)$$

$$\text{Adaptive}_{\beta} = 1 - \text{Adaptive}_{\alpha} \quad (10)$$

where Adaptive_{α} and Adaptive_{β} denote the adaptive fusion weights of images I^{SR1} and I^{SR2} , respectively. The two averages, $\text{mean}(lc_{I^{SR1}})$ and $\text{mean}(lc_{I^{SR2}})$, denote the mean contrast values of images I^{SR1} and I^{SR2} , respectively.

Ultimately, the computed localized contrast image is the outcome. To create a more flexible and manageable picture fusion effect, we can use this approach to alter the brightness and contrast based on the local elements of the image.

The advantage of this approach is that it can fully utilize the advantages of the generator and the GAN model to improve the quality and realism of the image. By fusing different reconstruction results, we can obtain more characteristic and diverse images, further expanding the scope of applications in image processing and computer vision. By dynamically varying the fusion weights, local contrast-weighted fusion

maintains information based on the image's level of local contrast. In addition to avoiding over-averaging, it enhances image quality and visual impact. The design offers a broad range of applications and great adaptivity.

Within the framework of the GAN model, the generator and the discriminator are two collaborative entities responsible for the reconstruction of the input image. Therefore, the adaptive fusion of the reconstruction results from the generator and the GAN model can synthesize the characteristics of the two images to obtain a richer and more diverse image.

IV. EXPERIMENTS

To comprehensively evaluate the SR performance of our RDAF-GAN with the state-of-the-art (SOTA) remote sensing image methods, we select several models for comparison, including Bicubic [12], SRGAN [15], ESRGAN [46], BebyGAN [43], BSRGAN [23], EDiffSR [16], and TTST [38].

A. Datasets and Metrics

1) *Datasets:* The ITCVD dataset [48] and the DLR Munich vehicle dataset [49] were used for training. It contains 135 images of 5616×3744 . The test set comprises of the RSC11 [50], AID [51], WHU-RS19 [52], and NWPU-RESISC45 [53]. The RSC11 dataset provides remote sensing scene images of 11 diverse categories, while the AID dataset consists of more than 10 000 images from 30 different scene categories. The WHU-RS19 dataset contains 1005 images from 19 scene classes. The datasets were subjected to a uniform treatment using a stochastic degradation strategy. The processed datasets were subsequently compared using different reconstruction methods. NWPU-RESISC45 was used to validate the performance of the model on real-world remote sensing images. The specific experiments are described in detail in Sections IV-B and IV-C.

2) *Metrics:* Four evaluation metrics were used to assess the quality of the reconstructed images, which are: PSNR, structural similarity (SSIM) index [9], Color Fidelity [54], and learned perceptual image patch similarity (LPIPS) [55]. PSNR and SSIM are used to evaluate the fidelity of the image, and the higher the value indicates the better image

TABLE I
PSNR AND SSIM RESULTS OF THE RECONSTRUCTION OF DIFFERENT MODELS USING A UNIFIED CONSEQUENT DEGRADATION STRATEGY ON EACH CATEGORY OF THE RSC11 DATASET

Metrics \ Methods	Bicubic	SRGAN	ESRGAN	Beby-GAN	BSRGAN	TTST	EDiffSR	Ours
Categories	PSNR↑ SSIM↑							
Denseforest	23.05 0.4337	24.25 0.4475	23.20 0.4069	22.82 0.3322	25.13 0.4901	23.11 0.4084	23.41 0.4165	26.14 0.5222
Grassland	22.36 0.3811	24.27 0.4494	23.83 0.4168	23.97 0.3551	26.21 0.5168	24.31 0.4321	24.42 0.4237	26.95 0.5475
Harbor	17.68 0.3659	18.01 0.3748	17.53 0.3411	15.53 0.1884	17.58 0.3892	17.78 0.3636	17.74 0.3525	18.03 0.3986
Highbuildings	18.94 0.4092	20.73 0.4918	20.19 0.4459	18.41 0.2762	20.10 0.5534	20.42 0.4697	20.37 0.4534	20.79 0.5887
Lowbuildings	17.81 0.3030	19.55 0.3934	18.91 0.3467	18.18 0.2444	19.73 0.4649	19.01 0.3710	19.00 0.3502	20.23 0.4965
Overpass	18.73 0.3361	20.40 0.3994	19.89 0.3677	18.84 0.2426	20.56 0.4562	20.03 0.3865	20.02 0.3689	20.99 0.4792
Railway	19.16 0.3317	21.43 0.4300	21.03 0.3973	20.10 0.2815	21.73 0.5036	21.14 0.4177	21.05 0.3952	22.20 0.5213
Residentialarea	18.82 0.3622	19.72 0.3896	19.23 0.3635	17.93 0.2220	19.52 0.4188	19.41 0.3811	19.43 0.3660	19.92 0.4455
Roads	19.13 0.3674	21.02 0.4398	20.53 0.4073	19.44 0.2687	21.11 0.4947	20.71 0.4277	20.64 0.4050	21.63 0.5222
Sparseforest	19.98 0.3573	20.94 0.3231	21.00 0.3095	21.64 0.2757	21.22 0.3638	20.89 0.3164	20.97 0.3092	21.94 0.4036
Stroagetanks	18.20 0.3311	19.73 0.4017	19.24 0.3647	17.86 0.2298	19.54 0.4403	19.40 0.3870	19.32 0.3645	20.01 0.4711
Average	19.44 0.3573	20.91 0.4127	20.41 0.3788	19.52 0.2651	21.13 0.4628	20.56 0.3964	20.57 0.3822	21.71 0.4905

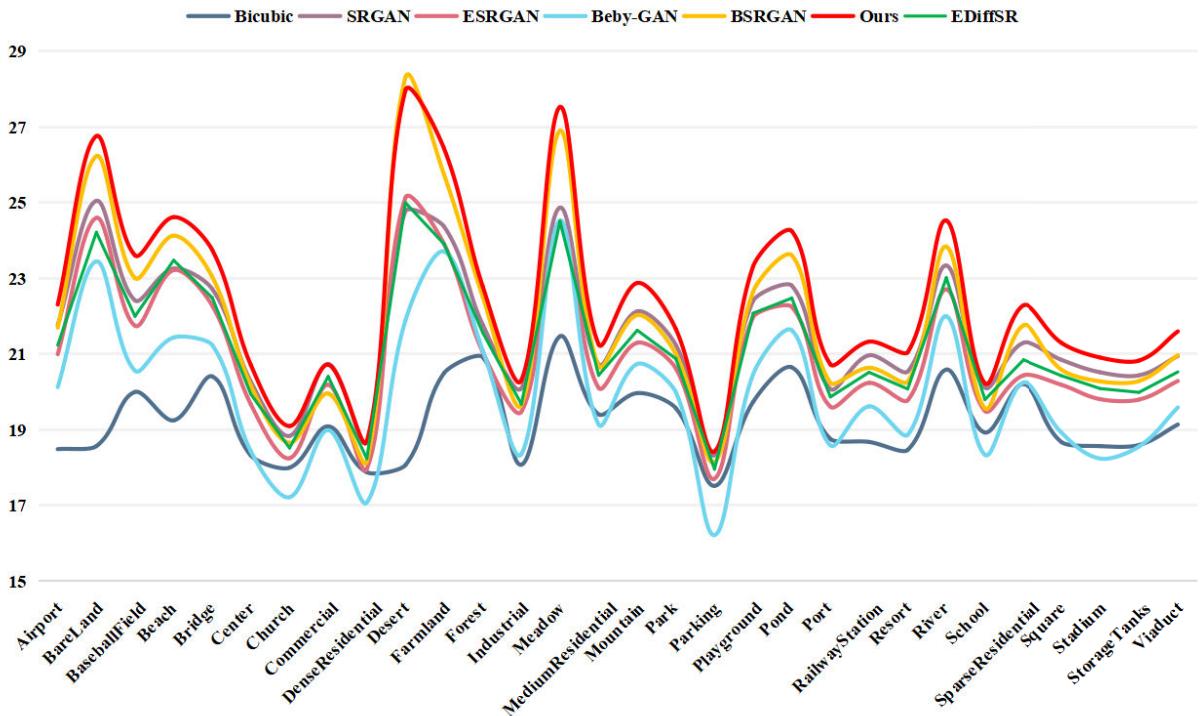


Fig. 7. Line plots of PSNR values for each model for each category in the AID dataset.

quality. Color fidelity is a measure of the degree of color similarity between the original image and the reconstructed or processed image. Higher values represent better reconstruction. On the other hand, LPIPS calculates the difference between the original and reconstructed images in the feature domain. The lower the value of this metric, the more similar the images are and thus the better the image reconstruction.

In addition to this, we used natural image quality evaluator (NIQE) and average gradient (AG) to test the quality of reconstructed images from real-world remote sensing images.

PSNR is a common measure of image quality, which is calculated from the mean square error (mse) to calculate the

degree of image distortion. The formula is shown as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{mse}} \right) \quad (11)$$

where MAX is the maximum possible value of the image pixel value, usually 255 (for 8-bit images). And mse is the mean square error. Larger values of PSNR indicate less distortion in the image.

The SSIM metric provides a comprehensive assessment of image quality by considering both distortion and similarity. It evaluates the image based on factors like brightness, color variation, and structure, aligning with the way humans



Fig. 8. Images reconstructed using advanced perception-oriented methods were qualitatively compared. The first column on the left shows a localized enlargement of the original image, while the rest zooms in on the same locations of the reconstructed image for a clearer view and comparison.

TABLE II
SSIM AND LPIPS RESULTS FOR DIFFERENT MODELS RECONSTRUCTED ON THE RSC11 AND AID DATASETS. THE RESULTS ARE OBTAINED BY AVERAGING THE SUM OF THE INDIVIDUAL CATEGORIES

Metrics \ Datasets	RSC11		AID	
	SSIM ↑	LPIPS ↓	SSIM ↑	LPIPS ↓
Methods				
Bicubic	0.3573	0.7530	0.3529	0.8131
SRGAN	0.4127	0.6022	0.4472	0.5684
ESRGAN	0.3788	0.5704	0.4026	0.5786
Beby-GAN	0.2651	0.8407	0.2848	0.8903
BSRGAN	0.4728	0.4512	0.4983	0.4532
EDiffSR	0.3822	0.5741	0.4091	0.5469
Ours	0.4992	0.5202	0.5305	0.5190

perceive visual information, which is defined as follows:

$$\text{SSIM} = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (12)$$

where α , β , and γ are the weighting parameters that represent the share of three different features in the SSIM measure: brightness, contrast, and structure, respectively. $l(x, y)$ represents the brightness comparison, $c(x, y)$ is the difference comparison, and $s(x, y)$ is the texture comparison. Larger SSIM values indicate less distortion and greater similarity to the original image.

LPIPS focuses on capturing the perceptual structure of an image, making it stable under different transformations while

disregarding pixel-level similarities. The formula is shown as follows:

$$L(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)\|_2^2 \quad (13)$$

where $L(x, x_0)$ denotes the distance between the image x and the reference image x_0 . The features extracted from the image x at layer l are denoted by \hat{y}_{hw}^l , while the features of the reference image x_0 extracted at layer (l) are denoted by \hat{y}_{0hw}^l . A weight w_l is applied to weigh the contribution of different layer features. The symbol \odot denotes the element-by-element multiplication. The height and width of the layer (l) feature map are represented by H_l and W_l , respectively. A lower LPIPS value indicates greater image similarity.

Color fidelity is measured by comparing the color difference between the original image and the reconstructed image. The formula for its calculation is as follows:

$$\text{Color Fidelity} = \frac{1}{1 + \text{Mean Delta E}} \quad (14)$$

where Mean Delta E is the mean color variance, which is the average of all pixel color variances. A larger color fidelity value usually means better color quality in the reconstructed image. It means that the color of the reconstructed image is closer to the original image.

3) *Implementation Details*: During model training, we focus on the $\times 4$ SR task. This study divides the training process into two key phases. First, we focus on the stage of training the generator, which is a key step in achieving

TABLE III

QUANTITATIVE PSNR AND SSIM COMPARISONS WITH THE SOTA SR MODELS IN 30 SCENE CATEGORIES OF THE AID TEST SET. THE RED COLOR INDICATES THE HIGHEST LEVEL OF PERFORMANCE, WHEREAS BLUE INDICATES THE SECOND-HIGHEST LEVEL OF PERFORMANCE

Metrics \ Methods	Bicubic		SRGAN		ESRGAN		Beby-GAN		BSRGAN		EDiffSR		Ours	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Airport	18.49	0.3217	21.77	0.4702	21.00	0.4156	20.13	0.2936	21.70	0.5222	21.24	0.4312	22.31	0.5576
BareLand	18.57	0.2765	25.05	0.5466	24.60	0.5051	23.45	0.3591	26.23	0.6431	24.22	0.4763	26.76	0.6804
BaseballField	20.00	0.3809	22.42	0.4981	21.75	0.4548	20.56	0.3212	23.01	0.5838	22.00	0.4566	23.61	0.6313
Beach	19.25	0.3488	23.26	0.5630	23.22	0.5084	21.44	0.3693	24.13	0.6390	23.48	0.5038	24.62	0.6802
Bridge	20.41	0.4189	22.72	0.5308	22.28	0.4978	21.24	0.3667	23.04	0.6272	22.49	0.4951	23.75	0.6572
Center	18.32	0.3426	20.27	0.4437	19.67	0.3873	18.43	0.2585	20.17	0.4891	19.94	0.4029	20.71	0.5183
Church	18.00	0.3276	18.84	0.3600	18.25	0.3192	17.22	0.2089	18.66	0.4012	18.52	0.3372	19.10	0.4150
Commercial	19.09	0.3899	20.74	0.4265	20.19	0.3831	18.99	0.2472	19.96	0.4407	20.41	0.3960	20.72	0.4761
DenseResidential	17.88	0.3364	18.68	0.3394	17.97	0.2976	17.08	0.1857	18.17	0.3393	18.24	0.3119	18.70	0.3626
Desert	18.07	0.2567	24.78	0.6116	25.13	0.5494	21.89	0.3585	28.31	0.7707	25.00	0.5127	27.97	0.8006
Farmland	20.50	0.3621	24.37	0.5207	23.91	0.4841	23.71	0.3785	25.74	0.6312	23.91	0.4752	26.43	0.6748
Forest	20.92	0.3519	21.78	0.3383	21.06	0.2916	21.10	0.2406	22.52	0.3291	21.57	0.3171	22.79	0.3365
Industrial	18.08	0.3320	20.10	0.4041	19.48	0.3541	18.36	0.2368	19.66	0.4257	19.68	0.3656	20.31	0.4564
Meadow	21.47	0.3592	24.87	0.4731	24.48	0.4396	24.53	0.3742	26.90	0.5480	24.51	0.4338	27.53	0.5927
MediumResidential	19.41	0.3448	20.73	0.3939	20.11	0.3570	19.14	0.2469	20.68	0.4211	20.44	0.3699	21.26	0.4501
Mountain	19.97	0.3742	22.13	0.4204	21.30	0.3738	20.74	0.2744	22.03	0.4125	21.63	0.3858	22.88	0.4587
Park	19.57	0.3773	21.24	0.4119	20.62	0.3686	19.99	0.2662	21.00	0.4365	20.88	0.3815	21.66	0.4700
Parking	17.52	0.3366	18.33	0.4147	17.71	0.3654	16.22	0.2338	18.15	0.4762	17.96	0.3823	18.43	0.4838
Playground	19.74	0.3787	22.41	0.5055	21.98	0.4702	20.47	0.3318	22.65	0.5897	22.08	0.4708	23.32	0.6276
Pond	20.65	0.4052	22.81	0.4965	22.25	0.4589	21.63	0.3514	23.61	0.5911	22.48	0.4610	24.25	0.6226
Port	18.76	0.4047	20.08	0.4940	19.61	0.4535	18.59	0.3211	20.24	0.5958	19.87	0.4613	20.73	0.6206
RailwayStation	18.68	0.3332	20.97	0.3991	20.24	0.3486	19.62	0.2526	20.64	0.4197	20.52	0.3592	21.33	0.4496
Resort	18.47	0.3487	20.54	0.4210	19.78	0.3725	18.87	0.2594	20.27	0.4540	20.08	0.3852	21.05	0.4979
River	20.59	0.3731	23.34	0.4695	22.71	0.4286	22.00	0.3176	23.84	0.5120	23.01	0.4369	24.53	0.5504
School	18.93	0.3823	20.12	0.4131	19.50	0.3727	18.34	0.2376	19.57	0.4267	19.80	0.3866	20.23	0.4626
SparseResidential	20.21	0.3120	21.30	0.3264	20.44	0.2877	20.25	0.2249	21.76	0.3456	20.85	0.3029	22.29	0.3630
Square	18.68	0.3562	20.85	0.4446	20.19	0.3962	18.93	0.2626	20.59	0.4834	20.43	0.4071	21.30	0.5244
Stadium	18.57	0.3663	20.52	0.4628	19.81	0.4015	18.24	0.2636	20.28	0.5076	20.09	0.4193	20.91	0.5416
StorageTanks	18.60	0.3396	20.44	0.4203	19.80	0.3782	18.56	0.2530	20.29	0.4627	19.99	0.3830	20.83	0.4951
Viaduct	19.14	0.3488	20.95	0.3968	20.29	0.3569	19.59	0.2486	20.97	0.4267	20.53	0.3669	21.60	0.4578
Average value	19.21	0.3530	21.54	0.4472	20.97	0.4026	19.97	0.2848	21.82	0.4983	21.19	0.4091	22.39	0.5305

TABLE IV

COLOR FIDELITY OF THE IMAGES IN THE RSC11 DATASET. LARGER VALUES INDICATE BETTER COLOR FIDELITY

Metrics \ Methods	Color Fidelity Metrics							
	Bicubic	SRGAN	ESRGAN	Beby-GAN	BSRGAN	TTST	EDiffSR	Ours
Categories								
Denseforest	0.0930	0.0968	0.0936	0.0919	0.0952	0.0921	0.0920	0.0984
Grassland	0.0924	0.0979	0.0976	0.0963	0.1057	0.1012	0.0999	0.1069
Harbor	0.0920	0.0929	0.0938	0.0892	0.0962	0.0957	0.0950	0.0955
Highbuildings	0.0928	0.0985	0.0978	0.0907	0.0871	0.1006	0.1001	0.1076
Lowbuildings	0.0912	0.0991	0.0963	0.0920	0.0869	0.0987	0.0981	0.1064
Overpass	0.0922	0.0987	0.0961	0.0926	0.0871	0.0988	0.0982	0.1042
Railway	0.0933	0.1009	0.1001	0.0952	0.0876	0.1028	0.1017	0.1112
Residentialarea	0.0915	0.0952	0.0934	0.0901	0.0870	0.0959	0.0955	0.0972
Roads	0.0921	0.0988	0.0966	0.0925	0.0871	0.1001	0.0991	0.1060
sparseforest	0.0872	0.0871	0.0870	0.0855	0.0868	0.0875	0.0874	0.0891
Stroagetanks	0.0911	0.0973	0.0953	0.0962	0.0867	0.0975	0.0970	0.1024

high-quality image SR. Then, the perceptual quality-oriented GAN model is trained. Optimization was performed using

the Adam algorithm. The initial learning rate is set to 5×10^{-5} . We train by minimizing a weighted combination of

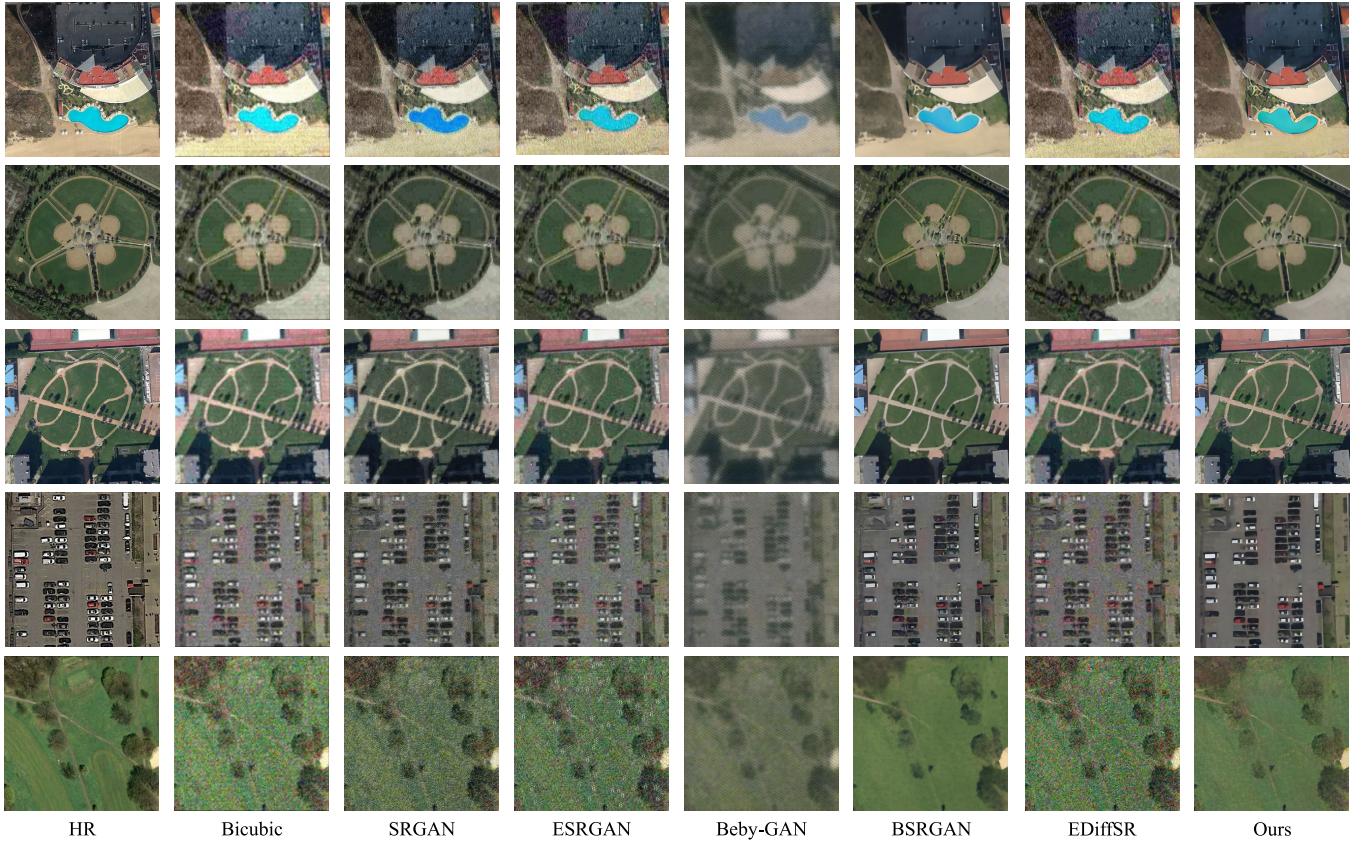


Fig. 9. Visual comparison of the visualization of different methods on the AID dataset.

L_1 loss, VGG perceptual loss, and PatchGAN loss. To train our RDAF-GAN, we performed 1 600 000 iterations with a small batch size of 8. We implement our models with the PyTorch framework and Python programming language and train them using an NVIDIA GTX 3090 GPU.

B. Experimental Results and Analysis

1) *Experiments on Simulated Datasets*: To ensure a fair comparison, all SR methods considered in this study were retrained using the same training set without any pretraining or fine-tuning process. The same level of degradation was applied uniformly across each category of the RSC11, AID, and WHU-RS19 datasets to ensure a fair comparison, thereby demonstrating the models' stability and generalization capacity.

2) *Quantitative Comparison*: The reconstructed metrics data for this article and the comparison model on the various categories of the RSC11 dataset are presented in Table I. As can be seen from the table, the model proposed in this article shows significant advantages on the RSC11 dataset. In comparison to the other comparative models, there was an average improvement of 6.7% in PSNR. This demonstrates its effectiveness and superiority in handling image reconstruction tasks. The results of EDiffSR and TTST are slightly lower. The main reason for this could be the diverse degradation in remote sensing images, making it difficult to generalize across various degradations.

Table II tests the performance of each model on both SSIM and LPIPS metrics. The model in this article performs well under SSIM metrics. Compared to the suboptimal values, it improves by 6.46% and 5.58% on RSC11 and AID, respectively. It shows that there is a high SSIM between its generated images and the original images. Under the LPIPS metric, the model in this article performs next best, outperforming most models. Although our RDAF-GAN shows strong performance in SSIM and LPIPS evaluations, there is still room for further optimization, especially in perceptual detail processing.

Table III presents the experimental findings of other approaches for the two indicators. We computed the metrics for every scenario on the AID dataset and presented the average outcomes for each method. Our proposed RDAF-GAN method yielded the highest average values for both PSNR and SSIM metrics, surpassing the second-best value by 0.57 and 0.032, respectively. These results demonstrate that RDAF-GAN can produce robust data distributions across various remote sensing scenarios, highlighting its strong generative capabilities.

The color fidelity metrics on the RSC11 dataset are displayed in Table IV, which assesses the models' capacity to preserve the original image's color. Based on the evaluation results, the TTST color retention ability is suboptimal compared to other models. The RDAF-GAN model proposed has demonstrated a remarkable ability to maintain the original

TABLE V

THREE METRICS PSNR, SSIM, AND LPIPS ON THE WHU-RS19 DATASET FOR OUR METHODOLOGY AND COMPARATIVE APPROACH (PSNR ↑, SSIM ↑, AND LPIPS ↓)

Categories	Metrics	Bicubic	SRGAN	ESRGAN	Beby-GAN	BSRGAN	EDiffSR	Ours
Airport	PSNR	18.00	20.48	19.84	17.41	20.26	20.00	20.84
	SSIM	0.3380	0.4492	0.3960	0.2786	0.4772	0.4013	0.5166
	LPIPS	0.7927	0.5199	0.5167	0.7344	0.3837	0.5008	0.4429
Beach	PSNR	20.27	24.34	24.02	17.90	25.95	25.43	26.10
	SSIM	0.5130	0.7491	0.5819	0.5077	0.7917	0.5728	0.8570
	LPIPS	0.6875	0.4108	0.5466	0.6913	0.4858	0.5313	0.4184
Bridge	PSNR	20.88	23.68	23.13	20.99	24.40	23.68	24.90
	SSIM	0.4228	0.5999	0.5344	0.4639	0.7288	0.5541	0.7523
	LPIPS	0.7731	0.5039	0.5283	0.6950	0.3917	0.5224	0.3633
Commercial	PSNR	17.68	18.21	17.71	14.81	17.38	17.80	18.15
	SSIM	0.3771	0.3655	0.3340	0.1906	0.3638	0.3345	0.3907
	LPIPS	0.7049	0.5294	0.5122	0.7914	0.4023	0.5045	0.4781
Desert	PSNR	17.90	24.48	24.13	22.40	28.43	24.62	28.46
	SSIM	0.2179	0.6014	0.4677	0.4475	0.7991	0.4805	0.8397
	LPIPS	0.8996	0.5784	0.6038	0.6660	0.4017	0.5936	0.4300
Farmland	PSNR	20.83	24.58	23.97	24.25	26.32	24.02	27.41
	SSIM	0.3677	0.5003	0.4472	0.4493	0.6482	0.4437	0.6895
	LPIPS	0.7831	0.5658	0.6015	0.7384	0.4653	0.6061	0.5327
footballField	PSNR	18.57	20.78	20.23	16.92	20.22	20.51	20.79
	SSIM	0.3847	0.4980	0.4373	0.2895	0.5255	0.4514	0.5534
	LPIPS	0.6720	0.4664	0.4477	0.6952	0.3605	0.4477	0.4304
Forest	PSNR	20.94	21.99	21.32	19.84	21.99	21.93	22.20
	SSIM	0.3358	0.3320	0.2975	0.2286	0.3050	0.3225	0.3340
	LPIPS	0.7868	0.5786	0.5775	0.7338	0.5742	0.5607	0.7073
Industrial	PSNR	17.87	19.14	18.60	15.80	18.28	18.67	19.03
	SSIM	0.3830	0.4081	0.3622	0.2290	0.4069	0.3627	0.4390
	LPIPS	0.6842	0.4838	0.4560	0.7602	0.3645	0.4638	0.4353
Meadow	PSNR	21.07	25.21	24.88	24.02	27.56	25.23	27.89
	SSIM	0.3479	0.5352	0.4676	0.4630	0.6374	0.4911	0.6954
	LPIPS	0.7442	0.5170	0.5072	0.5920	0.3913	0.5082	0.4677
Mountain	PSNR	18.93	19.55	19.11	17.75	19.28	19.41	19.65
	SSIM	0.2970	0.2836	0.2597	0.1745	0.2641	0.2680	0.2643
	LPIPS	0.8534	0.6470	0.6157	0.7325	0.5513	0.5904	0.7049
Park	PSNR	19.66	20.63	20.25	18.45	20.42	20.35	21.03
	SSIM	0.3873	0.3937	0.3607	0.2831	0.4234	0.3612	0.4535
	LPIPS	0.7366	0.5500	0.5254	0.7595	0.4830	0.5278	0.5602
Parking	PSNR	17.48	19.03	18.39	16.14	18.54	18.51	19.02
	SSIM	0.3518	0.4516	0.3955	0.2656	0.5114	0.3990	0.5294
	LPIPS	0.6880	0.4491	0.4717	0.8030	0.3522	0.4752	0.4243
Pond	PSNR	20.73	23.45	22.68	21.03	23.93	23.05	24.70
	SSIM	0.4174	0.5364	0.4815	0.4123	0.6338	0.4984	0.6694
	LPIPS	0.7985	0.5503	0.5583	0.6715	0.3724	0.5391	0.4114
Port	PSNR	18.33	19.23	18.73	16.29	19.43	18.91	19.92
	SSIM	0.3912	0.4518	0.3961	0.3353	0.5684	0.4156	0.5929
	LPIPS	0.7111	0.5383	0.5231	0.7664	0.3675	0.5030	0.4185
railwayStation	PSNR	17.75	18.89	18.43	17.51	18.70	18.44	19.49
	SSIM	0.3091	0.3107	0.2800	0.2193	0.3327	0.2746	0.3563
	LPIPS	0.8303	0.6182	0.5780	0.7694	0.4470	0.5338	0.5360
Residential	PSNR	17.10	17.79	17.27	14.30	16.99	17.40	17.59
	SSIM	0.3605	0.376	0.3425	0.1843	0.3755	0.3463	0.4044
	LPIPS	0.7036	0.5064	0.5080	0.7803	0.3660	0.4987	0.4489
River	PSNR	19.63	21.37	20.49	18.83	21.34	21.03	21.12
	SSIM	0.3484	0.4023	0.3471	0.2689	0.4165	0.3724	0.4578
	LPIPS	0.7817	0.5430	0.5251	0.7075	0.4558	0.5182	0.5488
Viaduct	PSNR	17.71	18.90	18.35	15.89	18.58	18.42	19.08
	SSIM	0.3461	0.3638	0.3334	0.1976	0.3769	0.3302	0.3961
	LPIPS	0.7430	0.5274	0.5097	0.7563	0.3758	0.4832	0.4599
Average value	PSNR	19.01	21.14	20.60	18.44	21.47	20.91	21.96
	SSIM	0.3629	0.4530	0.3959	0.3099	0.5045	0.4042	0.5349
	LPIPS	0.7565	0.5307	0.5322	0.7286	0.4206	0.5215	0.4852

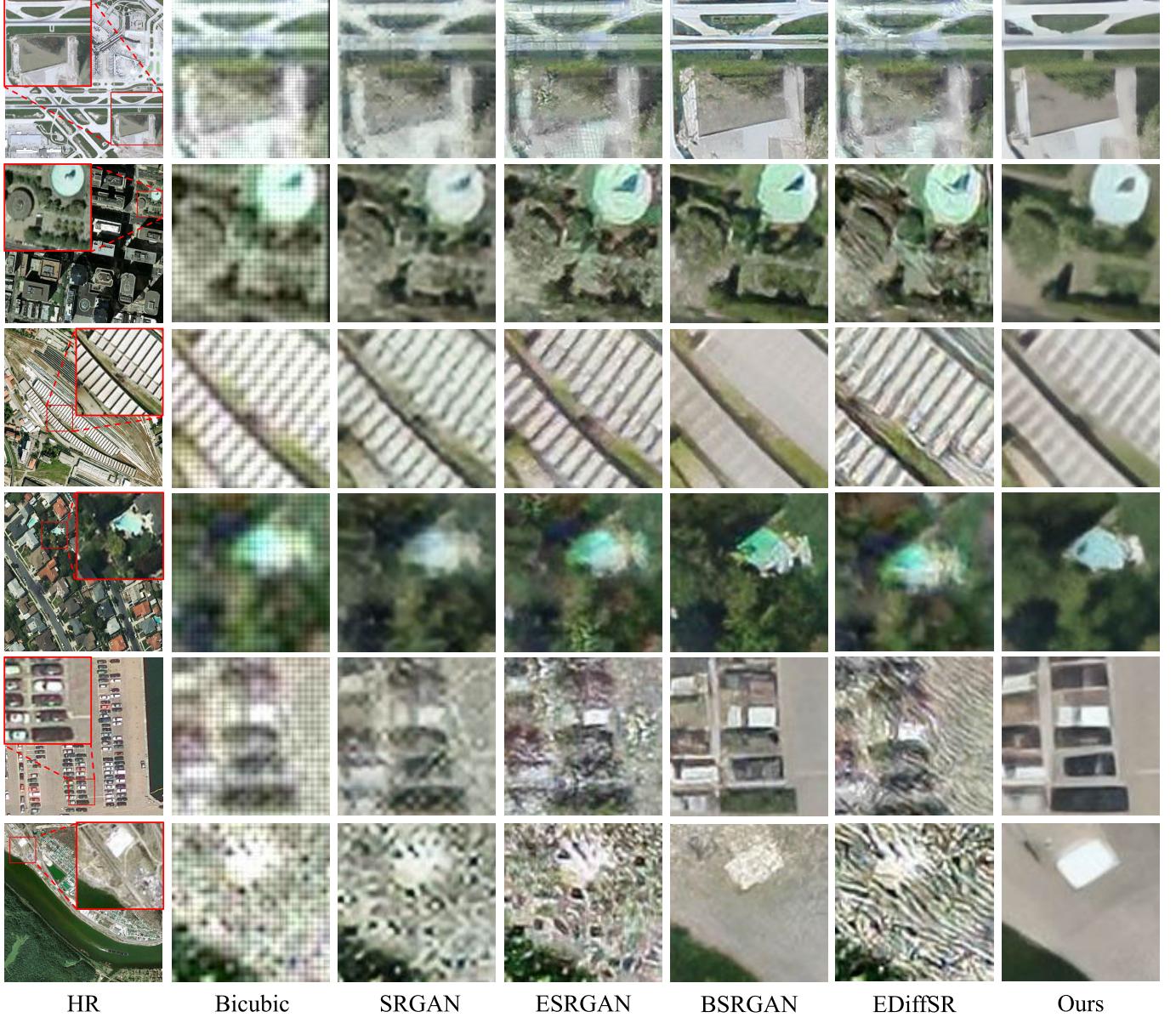


Fig. 10. Visualization of six representative scenes from the WHU-RS19 dataset by various methods.

image's color with superior color fidelity compared to other models.

Additionally, Table V presents the experimental results of six methods concerning three indicators. Among them, our RDAF-GAN performs the best on the SSIM metric, showcasing remarkable SSIM. Overall, our proposed RDAF-GAN outperforms the other methods in terms of PSNR and SSIM, with scores of 0.49 and 0.03 higher than the second-best method, respectively.

3) Qualitative Comparison: Fig. 7 displays each model's PSNR for each class in the AID dataset. The model presented in this study performs well across a wide range of categories, particularly Meadow and River, where the model's PSNR values are noticeably higher than those of the other models. This suggests that the model presented in this article is more capable of reconstructing images in these categories and can

better preserve the structural information and details of the images.

The reconstruction outcomes of different techniques applied to the RSC11 dataset under identical degradation circumstances are displayed in Fig. 8. Of particular note is the first row of the highbuilding category, where our approach outperforms other methods by processing the image with greater clarity and smoothness, resulting in less noise and more recognizable details. Other techniques produce images with noise and blurring that can compromise visual quality.

Our method stands out in the second and fifth rows of Fig. 8, as it successfully enhances the image's clarity and authentic details, unlike other methods that produce intricate patterns with noticeable flaws and distortions. Although TTST and EDiffSR recover some detail, they introduce severe artifacts that do not correspond to the ground truth. We must highlight

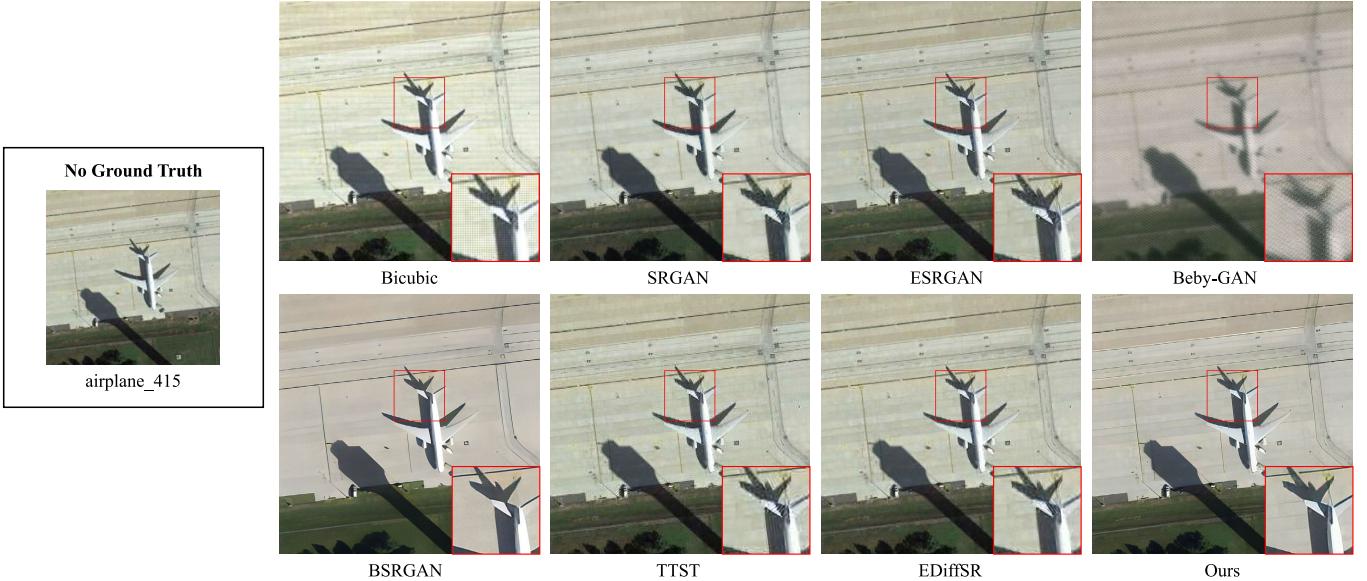


Fig. 11. $4\times$ visual comparison with SOTA SR model local zoomed-in view on NWPU-RESISC45.

TABLE VI

RESULTS OF THE $\times 4$ TRUE DEGRADATION COMPARISON ON THE NWPU-RESISC45. THE BEST PERFORMANCE VALUE IS HIGHLIGHTED IN RED, AND THE SECOND-BEST PERFORMANCE VALUE IS HIGHLIGHTED IN BLUE

Methods	Bicubic	SRGAN	ESRGAN	Beby-GAN	BSRGAN	TTST	EDiffSR	Ours
NIQE \downarrow	20.806	15.098	16.887	23.336	14.497	14.590	17.932	11.931
AG \uparrow	6.924	2.770	3.594	6.219	3.229	3.100	3.323	3.978

that our method significantly improves the image's visual appeal, which is a testament to its efficacy.

Our RDAF-GAN demonstrates exceptional accuracy in recovering the textures of the vegetation and the boat, as evidenced by the third and fourth rows of Fig. 8. Other methods, however, often fail to produce accurate textures. Despite the visually impressive textures generated by BSRGAN [23], it still suffers from significant distortion issues.

In Fig. 9, we also visualize some SR results on AID. In the first row of examples, the majority of the methods produce noise and artifacts. BSRGAN [23] shows rougher recovery results in the detail edge part while Beby-GAN [43] loses more color information, resulting in color distortion in the SR outcomes. Conversely, our method stands out with better color recovery results and more natural SR results compared to other methods.

The task of image reconstruction involves various models that exhibit unique characteristics. The BaseballField and Square scenes are depicted in the second and third rows of Fig. 9, respectively, showcasing the reconstruction outcomes. It is evident from the illustration that models with greater depth can effectively capture intricate image features and structures, leading to precise and sophisticated reconstruction outputs.

The last two rows of the example shown in Fig. 9 demonstrate that most other methods are unable to process details such as cars and grass accurately. They usually generate unrealistic textures with jagged edges and artifacts. Beby-GAN exhibits severe distortion, which deviates from the ground-truth distribution. In contrast, our RDAF-GAN suc-

cessfully reconstructs clear and natural details and presents a more natural perception.

The detailed analysis of six different scene types from WHU-RS19 is shown in Fig. 10. For each scene, we selected a random image as a demonstration. Each column of images is labeled at the end, e.g., the first column is an HR image. Each scene has unique structural details that can significantly contribute to more precise assessments of the model's performance. The result of the image recovery by our RDAF-GAN is more realistic in the first row. As can be seen from the fourth line, the color of the image recovered by RDAF-GAN is closer to HR. On the last row, the RDAF-GAN eliminates noise and artifacts, providing a superior-quality image.

4) *Experiments on Real-World Data:* To assess the performance of our RDAF-GAN on real-world remote sensing images, we conducted experimental analysis on the NWPU-RESISC45 dataset without any simulated processing. For efficiency, we randomly selected 135 images from this dataset. In Table VI, we compared RDAF-GAN with SOTA methods using NIQE and AG metrics. A lower NIQE score indicates higher image quality, representing images closer to their natural appearance. While higher AG values denote richer texture detail in complex, realistic scenes. The results demonstrated that our proposed method achieved the best results in NIQE, indicating alignment with human perception in real-world scenarios. RDAG-GAN is competitive in solving the single image super-resolution (SISR) challenge of remote sensing images with real degradation.

TABLE VII

MODEL EFFICIENCY ANALYSIS WITH THE SOTA SR MODELS. THE BEST PERFORMANCE VALUE IS HIGHLIGHTED IN RED

Methods	BSRGAN	ESRGAN	EDiffSR	TTST	Ours
Param. (M)	20.09	29.78	26.79	18.94	21.47
FLOPs (G)	76.51	75.23	2.89	76.84	79.40
Inference time (s)	25.79	96.66	68.79	40.21	22.60
PSNR \uparrow	18.48	18.44	18.47	18.64	19.35
SSIM \uparrow	0.4868	0.5498	0.4316	0.4356	0.5321
LPIPS \downarrow	0.5116	0.6261	0.5989	0.6204	0.5251

TABLE VIII

ABLATION ANALYSIS OF DIFFERENT NUMBERS OF RRDBS, WITH THE BEST PERFORMANCE VALUE HIGHLIGHTED IN RED

Num.	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
22	19.25	0.5415	0.6324
23	19.34	0.5529	0.6268
24	19.46	0.5530	0.6265
25	19.69	0.5535	0.6172
26	19.63	0.5531	0.6189

The results of visualizing each model on the NWPU-RESISC45 dataset are presented in Fig. 11. The diffusion-based approach has limitations in recovering edge details. GAN-based methods exhibit varying degrees of artifact problems. In contrast, our method excels in recovering sharp texture details and produces minimal artifacts.

5) *Model Efficiency*: To reduce costs, we randomly selected 90 images from the AID dataset to create the AID part and test the speed of the model’s inferences. The chosen images included all the categories. Table VII shows that our RDAF-GAN, with a moderate number of parameters, has the fastest inference speed. This indicates that it performs well with large-scale data while maintaining efficient computational performance. Compared to other models, RDAF-GAN can produce high-quality results while ensuring lower computational overhead.

Our model demonstrates exceptional ability in balancing efficiency and performance, achieving superior results in both PSNR (19.35) and SSIM (0.5321). It boasts the shortest inference time of 22.60 s, striking a commendable equilibrium between efficiency and image quality. Although the floating point operations per second (FLOP) is slightly higher, RDAF-GAN produces the highest PSNR gain per giga floating-point operations per second (GFLOP) in most cases, indicating higher computational resource utilization. In contrast, EDiffSR, while registering the lowest FLOPs (2.89G), exhibits inferior image quality and a longer inference time of 68.79 s. This suggests that EDiffSR compromises overall performance by opting for a simplified network structure. Overall, our model outperforms by effectively harmonizing efficiency, inference speed, and high-quality output.

C. Ablation Studies

In the ablation section, we extensively discuss the model design and key components of RDAF-GAN. It is worth noting that we trained these models on the same training set and tested them on the AID part. PSNR, SSIM, and LPIPS are all averaged over the AID part.

TABLE IX

ABLATION ANALYSIS OF ADAPTIVE FUSION. THE BEST PERFORMANCE VALUES ARE MARKED IN RED. NUM. INDICATES A DIFFERENT NUMBER OF RRDBS

Type	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
w/o adaptive fusion	19.69	0.5535	0.6172
w/ adaptive fusion	20.23	0.5978	0.6019

TABLE X

ABLATION ANALYSIS OF DIFFERENT COMPONENTS OF RDAF-GAN

Type	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
random degradation	18.56	0.4653	0.5394
optimal stacking	18.64	0.4878	0.5336
adaptive fusion	19.35	0.5321	0.5251

1) *Validity of RRDB*: We investigated the effect of different numbers of RRDBs on the reconstruction effect at a single degradation. As shown in Table VIII, we can see that when the number of RRDBs increases, the reconstruction metrics improve. However, it is not true that more modules are better. When the number of RRDBs is 26, the performance decreases slightly. To strike a balance between model size and performance, we set the number of RRDBs at 25. Compared to a number of 23, PSNR improves by 1.8%, SSIM slightly by 0.1%, and LPIPS by 1.6%.

2) *Effectiveness of Adaptive Fusion*: To demonstrate how adaptive fusion improves SR reconstruction, we compared the reconstruction results with and without adaptive fusion in Table IX. The findings indicate that the performance with adaptive fusion is significantly superior to that without it. Specifically, PSNR, SSIM, and LPIPS showed improvements of 2.7%, 8.0%, and 2.4%, respectively. These results emphasize the crucial role of adaptive fusion strategies in enhancing the performance of RDAF-GAN.

3) *Different Components of RDAF-GAN*: We conducted ablation experiments on various components of the RDAF-GAN, as detailed in Table X. Initially, we introduced random degradation to mimic a real remote sensing scenario. As we incorporated additional components, the quality of the images improved further. Specifically, the metrics PSNR, SSIM, and LPIPS increased by 0.43%, 0.10%, and 1.07%, respectively. Moreover, the use of adaptive fusion significantly enhanced the reconstruction outcomes, with PSNR, SSIM, and LPIPS improving by 3.8%, 9.08%, and 1.59%, respectively.

V. CONCLUSION

This article aims to improve the resolution of remote sensing images and proposes an advanced model called RDAF-GAN. It incorporates stochastic degradation and adaptive fusion to correctly learn the mapping from LR to HR to produce clear images. Specifically, we develop a stochastic degradation method based on physical constraints to simulate multiple degradation factors in real remote sensing scenarios through controlled image deformation operations. Meanwhile, the adaptive fusion technique dynamically adjusts the fusion of image features to adapt to different image contents. Furthermore, for the loss function, a comprehensive loss function

is designed to guide generator learning. Finally, the effectiveness of our model in recovering remote sensing images was confirmed through rich comparison experiments, which included quantitative evaluation metrics such as PSNR, SSIM, and qualitative visual effects. The results demonstrate that our model excels in image quality and detail recovery.

Nevertheless, the RDAF-GAN model has limitations. The degradation models may not fully capture the complexity of the real world, resulting in poor performance when dealing with extreme types of degradation. Additionally, recovering fine textures and high-frequency details remains challenging, especially in complex scenes. These limitations could impact critical applications like environmental monitoring and disaster management. Future work will focus on enhancing the degradation model with more realistic patterns and improving texture recovery through advanced techniques like attention mechanisms. Addressing these issues will further strengthen the model's applicability in remote sensing image analysis.

REFERENCES

- [1] B. Wang, Y. Zhao, and X. Li, "Multiple instance graph learning for weakly supervised remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5613112.
- [2] Y. Song, J. Bi, L. Sun, Z. Liu, Y. Jiang, and X. Wang, "Freq-3DLane: 3D lane detection from monocular images via frequency-aware feature fusion," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–13, 2025, doi: 10.1109/TITS.2025.3565272.
- [3] J. Zhou et al., "UGIF-Net: An efficient fully guided information flow network for underwater image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4206117.
- [4] J. Bi et al., "Lane detection for autonomous driving: Comprehensive reviews, current challenges, and future predictions," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 5, pp. 5710–5746, May 2025.
- [5] H. Lu et al., "Information sparsity guided transformer for multi-modal medical image super-resolution," *Expert Syst. Appl.*, vol. 261, Feb. 2025, Art. no. 125428.
- [6] H. Li, N. M. Lokman Hossain, M. Trocan, D. Galayko, and M. Sawan, "CMISR: Circular medical image super-resolution," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108222.
- [7] D. Berardini, L. Migliorelli, A. Galdelli, and M. J. Marín-Jiménez, "Edge artificial intelligence and super-resolution for enhanced weapon detection in video surveillance," *Eng. Appl. Artif. Intell.*, vol. 140, Jan. 2025, Art. no. 109684.
- [8] X. Wang, L. Sun, J. Yi, Y. Song, Q. Zheng, and A. Chehri, "Efficient degradation representation learning network for remote sensing image super-resolution," *Comput. Vis. Image Understand.*, vol. 249, Dec. 2024, Art. no. 104182.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [10] A. M. Jawad, L. Alamro, L. F. Abdulrazak, A. A. AbdelRahman, and I. Bezklubenko, "Study of camera efficiency and image resolution," in *Proc. 35th Conf. Open Innov. Assoc. (FRUCT)*, Apr. 2024, pp. 294–306.
- [11] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [13] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
- [14] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
- [15] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [16] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, "EDiffSR: An efficient diffusion probabilistic model for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5601514.
- [17] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.
- [18] Z. Liu, Z. Li, X. Wu, Z. Liu, and W. Chen, "DSRGAN: Detail prior-assisted perceptual single image super-resolution via generative adversarial networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7418–7431, Nov. 2022.
- [19] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.
- [20] L. Wang et al., "Unsupervised degradation representation learning for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10576–10585.
- [21] Z. Yue, Q. Zhao, J. Xie, L. Zhang, D. Meng, and K. K. Wong, "Blind image super-resolution with elaborate degradation modeling on noise and kernel," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2118–2128.
- [22] X. Wang, L. Sun, A. Chehri, and Y. Song, "A review of GAN-based super-resolution reconstruction for optical remote sensing images," *Remote Sens.*, vol. 15, no. 20, p. 5062, Oct. 2023.
- [23] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4791–4800.
- [24] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [25] Y. Guo et al., "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5407–5416.
- [26] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-GAN," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2019, pp. 284–293.
- [27] J. Liang, K. Zhang, S. Gu, L. V. Gool, and R. Timofte, "Flow-based kernel prior with application to blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10596–10605.
- [28] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops*, Oct. 2021, pp. 1833–1844.
- [29] H. Zhou et al., "Learning correction filter via degradation-adaptive regression for blind single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12331–12341.
- [30] J. Liang, H. Zeng, and L. Zhang, "Efficient and degradation-adaptive network for real-world image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 574–591.
- [31] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, "Real-world super-resolution via kernel estimation and noise injection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2020, pp. 466–467.
- [32] Z. Luo, H. Huang, L. Yu, Y. Li, H. Fan, and S. Liu, "Deep constrained least squares for blind image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17621–17631.
- [33] Y. Wei, S. Gu, Y. Li, R. Timofte, L. Jin, and H. Song, "Unsupervised real-world image super-resolution via domain-distance aware training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13380–13389.
- [34] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, and L. Zhang, "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, pp. 297–311, Aug. 2023.
- [35] J. Zhang, T. Xu, J. Li, S. Jiang, and Y. Zhang, "Single-image super-resolution of remote sensing images with real-world degradation modeling," *Remote Sens.*, vol. 14, no. 12, p. 2895, Jun. 2022.
- [36] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.
- [37] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14821–14831.

- [38] Y. Xiao, Q. Yuan, K. Jiang, J. He, C.-W. Lin, and L. Zhang, "TTST: A top-k token selective transformer for remote sensing image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 738–752, 2024.
- [39] J. Wang, B. Wang, X. Wang, Y. Zhao, and T. Long, "Hybrid attention-based U-shaped network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5612515.
- [40] Y. Lu et al., "Cross-spatial pixel integration and cross-stage feature fusion-based transformer network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5625616.
- [41] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot, "Model-informed multistage unsupervised network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5516117.
- [42] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, and J. Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5504218.
- [43] W. Li, K. Zhou, L. Qi, L. Lu, and J. Lu, "Best-buddy GANs for highly detailed image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 2, pp. 1412–1420.
- [44] J. Zhao et al., "SA-GAN: A second order attention generator adversarial network with region aware strategy for real satellite images super resolution reconstruction," *Remote Sens.*, vol. 15, no. 5, p. 1391, Mar. 2023.
- [45] Y. Song, L. Sun, J. Bi, S. Quan, and X. Wang, "DRGAN: A detail recovery-based model for optical remote sensing images super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5602113.
- [46] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Jan. 2019, pp. 63–79.
- [47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [48] M. Y. Yang, W. Liao, X. Li, Y. Cao, and B. Rosenhahn, "Vehicle detection in aerial images," *Photogramm. Eng. Remote Sens.*, vol. 85, no. 4, pp. 297–304, 2019.
- [49] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.
- [50] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *J. Appl. Remote Sens.*, vol. 10, no. 3, Jul. 2016, Art. no. 035004.
- [51] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
- [52] W. Huang, Q. Wang, and X. Li, "Feature sparsity in convolutional neural networks for scene classification of remote sensing image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 3017–3020.
- [53] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [54] C.-H. Chou and K.-C. Liu, "A fidelity metric for assessing visual quality of color images," in *Proc. 16th Int. Conf. Comput. Commun. Netw.*, Aug. 2007, pp. 1154–1159.
- [55] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.



Lijun Sun was born in Yantai, Shandong, China, in 2001. She received the B.S. degree from the School of Computer and Control Engineering, Yantai University, Yantai, in 2023, where she is currently pursuing the master's degree.

Her research interests include deep learning, remote sensing image super-resolution, and artificial intelligence.



Jiping Bi was born in Weifang, Shandong, China, in 2001. He received the B.S. degree from the School of Computer and Control Engineering, Yantai University, Yantai, China, in 2023, where he is currently pursuing the master's degree.

His research interests include lane detection, deep learning, and automatic control.



Yongchao Song was born in Weihai, Shandong, China, in 1990. He received the B.S. and Ph.D. degrees from the School of Electronic and Control Engineering, Chang'an University, Xi'an, China, in 2015 and 2020, respectively.

He is currently an Associate Professor with the School of Computer and Control Engineering, Yantai University, Yantai, China. His research interests include remote sensing information processing and deep learning.



Zhaowei Liu (Member, IEEE) received the Ph.D. degree from Shandong University, Jinan, China, in 2018.

He is currently a Professor with Yantai University, Yantai, China. His main research interests include distributed artificial intelligence, machine learning, remote sensing information processing, virtual reality, and collaboration with proposed software companies in directions, including digital twin and industrial meta-universe-related research.

Dr. Liu is a member of China Computer Federation (CCF).



Xuan Wang (Senior Member, IEEE) was born in Weihai, Shandong, China, in 1991. She received the B.S. and Ph.D. degrees in traffic information engineering and control from Chang'an University, Xi'an, China, in 2013 and 2018, respectively.

She is currently an Associate Professor with the School of Computer Science and Control Engineering, Yantai University, Yantai, China. Her research interests include intelligent traffic control, artificial intelligence, and computer vision.