







Structure-Texture Dual Preserving for Remote Sensing Image Super Resolution

Kanghui Zhao , Tao Lu , *Member, IEEE*, Yanduo Zhang , Junjun Jiang , *Senior Member, IEEE*, Zhongyuan Wang , *Member, IEEE*, and Zixiang Xiong , *Fellow, IEEE*

Abstract—Most of the existing remote sensing image super-resolution (SR) methods based on deep learning tend to learn the mapping from low-resolution (LR) images to high-resolution (HR) images directly. But they ignore the potential structure and texture consistency of LR and HR spaces, which cause the loss of high-frequency information and produce artifacts. A structure-texture dual preserving method is proposed to solve this problem and generate pleasing details. Specifically, we propose a novel edge prior enhancement strategy that uses the edges of LR images and the proposed interactive supervised attention module (ISAM) to guide SR reconstruction. First, we introduce the LR edge map as a prior structural expression for SR reconstruction, which further enhances the SR process with edge preservation capability. In addition, to obtain finer texture edge information, we propose a novel ISAM in order to correct the initial LR edge map with high-frequency information. By introducing LR edges and ISAM-corrected HR edges, we build LR–HR edge mapping to preserve the consistency of LR and HR edge structure and texture, which provides supervised information for SR reconstruction. Finally, we explore the salient features of the image and its edges in the ascending space, and restored the difference between LR and HR images by residual and dense learning. A large number of experimental results on Draper and NWPU-RESISC45 datasets show that our model is superior to several advanced SR algorithms in both objective and subjective image quality.

Index Terms—Edge enhanced, interactive supervised attention module (ISAM), remote sensing image, super-resolution (SR) reconstruction.

Manuscript received 12 October 2023; revised 14 January 2024 and 1 February 2024; accepted 2 February 2024. Date of publication 6 February 2024; date of current version 6 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62072350, Grant 62171328, Grant 62171327, Grant 62071339, and Grant 61971165, in part by Central Government Guides Local Science and Technology Development Special Projects under Grant ZYYD2022000021, and in part by the National Natural Science Foundation of Hubei under Grant 2023AFB158. (*Corresponding author: Tao Lu.*)

Kanghui Zhao and Tao Lu are with the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China (e-mail: zhaokanghui1024@gmail.com; lutxyl@gmail.com).

Yanduo Zhang is with the Computer School, Hubei University of Arts and Science, Xiangyang 441021, China, and also with the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430079, China (e-mail: zhangyanduo@hotmail.com).

Junjun Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: jiangjunjun@hit.edu.cn).

Zhongyuan Wang is with the National Engineering Research Center for Multimedia Software (NERCMS), School of Computer Science, Wuhan University, Wuhan 430079, China (e-mail: wzy_hope@163.com).

Zixiang Xiong is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zx@ece.tamu.edu).

Digital Object Identifier 10.1109/JSTARS.2024.3362880

I. INTRODUCTION

REMOTE sensing high-resolution (HR) images can observe clear ground objects, so as to serve the subsequent image classification [1], [2], detection [3], [4], and recognition [5], [6] tasks. Limited by remote sensing equipment and internet bandwidth, satellite observation images usually have low resolution (LR). In addition, the resolution of satellite images captured by optical sensors cannot be directly applied to the analysis of subsequent high-level visual tasks of satellite images due to the influence of undersampling by optical sensors.

Super resolution (SR) [7], [8], [9], [10], [11], [12], [13], [14] can reconstruct the HR image that is most similar to the original LR image by using the LR image captured and the prior knowledge learned from the sample library, which can effectively enhance the resolution of low-quality image and recover the image feature details. The use of the SR method to improve the resolution of observation targets has always been the focus of remote sensing research. Subsequently, convolutional neural networks (CNNs) have achieved success in SR task. A large number of CNNs-based models have been studied, including the models in [15], [16], [17], and [18]. As we all know, an image includes low-frequency information component and high-frequency information component. The high-frequency information component refers to the region of the images with great changes, which contains rich structural information and texture details of images, such as image edges and gradients. Inspired by these facts, SR methods based on image edge priors have been developed greatly [19], [20], [21], [22].

However, these methods are complex and limited. For example, Yang et al. [23] first introduced the edge of images into CNNs as an auxiliary for SR reconstruction. As a pioneering work, this method still has some drawbacks that lead to poor performance. First, DEGREE preprocessed the image by bicubic interpolation before input, which introduces noise and artifacts. Second, DEGREE directly used the existing edge detector to obtain the edge of the images, which obtained edge images are rough. Finally, the edge features learned by DEGREE are essentially a direct residual learning process. Similar to [23], Fang et al. [22] introduced the same LR edge map as a complement to texture information, and then, directly calculated the L_1 loss of SR_{edge} and HR_{edge} as a constraint on the total loss function to generate SR. However, the SR images generated by this approach are often blurred or too smooth, resulting in missing details and obvious blurring.

EEGAN [20] constructed the edge enhancement subnet (EESN) to extract and enhance image contours by using mask processing to purify noise pollution components, and combine the restored intermediate image with the enhanced edge. SPSR [21] uses an adversarial learning mechanism to first introduce LR gradient branches in the generative network, synthesize the gradients of HR images by the gradients of LR images, and fuse them to improve the boundary quality of the generated images. However, in SPSR [21] and EEGAN [20], the authors used perceptually driven methods to design the gradient loss, but the results of such methods often suffer from geometric distortion and the generated textures are unnatural or even distorted. In addition, GAN-based SR could generate some high-fidelity results, but it also introduces geometric distortions, especially at edges and fine textures and the stable training of GAN is still a problem.

GEDRN [19] uses shared source residual structure and non-local operations to learn rich low-frequency information and long-distance spatial correlations, and uses gradient loss and perceptual loss to further improve the perceptual quality. However, it performs multiscale upsampling of images before feature extraction and extracts gradient information for the intermediate step of fusion. However, this preprocessing approach increases the model computation on the one hand, and introduces erroneous edge and texture information on the other hand.

Instead, we know that deep learning has an excellent ability to handle probabilistic transformations of pixel distributions, so we build up a mapping of LR_{edge} to HR_{edge} by CNNs. This edge mapping can be seen as another image, and thus, the image to image conversion technique can be used to learn the mapping between the two modalities. The conversion process is equivalent to transforming the spatial distribution from LR_{edge} sharpness to HR_{edge} sharpness. Since most of the region of the edge map is close to zero, CNNs can focus more on the spatial relationship of contours. As a result, the network may more easily capture the structural dependencies, and thus, generate approximate edge maps for SR images. Therefore, based on the aforementioned considerations, we propose a novel edge enhancement strategy that differs from previous work in that we establish a mapping relationship between LR edges and HR edges. Specifically, LR edges are first introduced as a complement to texture details, and subsequently, an intermediate process edge map is obtained after edge enhancement branching pairs. This intermediate process edge map is edge corrected by our proposed interactive supervised attention module (ISAM) and summed with the results of LR edge upsampling to obtain the final reconstructed edge. This reconstructed edge is finally used to assist in generating SR results with high confidence and clear content.

In this article, we propose a novel structure texture dual preservation (STP) method to solve this problem, which provides supervised information for SR reconstruction through edges, thereby reconstructing high-quality remote sensing images. Specifically, our model consists of structural branches and texture branches. The structure branch is the SR branch, which is directly mapped from LR to SR through CNNs to obtain preliminary reconstruction results. Subsequently, we proposed a novel edge prior enhancement strategy that does not require the

introduction of additional prior loss. By directly establishing a prior mapping between LR_{edge} and HR_{edge} using CNNs, precise edge images can be obtained for SR reconstruction assistance. Due to the fact that edge images represent the high-frequency components of the image and are rich in texture information, they can serve as an important supplement to SR reconstruction. We introduced LR edge image and designed a texture branch for it to convert the edge images of LR images into HR images as auxiliary SR problems. In addition, we have designed an ISAM to monitor LR_{edge} and supplement high-frequency information to obtain more accurate HR_{edge} . Through texture branching, we established edge mapping for LR_{edge} and HR_{edge} , and the final reconstructed edge was fused with the SR intermediate result to obtain the final SR reconstructed image. For structural and texture branches, we use the same structure to map the low-dimensional space of the image to the multiscale high-dimensional space, and extract multiscale global structural prior information from the multiscale high-dimensional space of the image as the structural prior representation of the SR task. Through this guidance, we not only established a mapping between LR and HR, but also established a mapping between LR_{edge} and HR_{edge} , maintaining consistency between LR and HR in spatial and texture structure.

The contributions and innovations of this article can be summarized as follows.

- 1) We propose a novel edge enhancement strategy, which differs from previous work in establishing a mapping relationship between LR_{edge} and HR_{edge} . Specifically, LR_{edge} are first introduced as a supplement to texture details, and then an intermediate process edge image is obtained after edge enhancement branch pairing. The edge image of this intermediate process is edge corrected by our proposed ISAM and added to the sampled results on the LR_{edge} to obtain the final reconstructed edge. It can effectively restore high-frequency information, remove artifacts, and maintain the sharpness and details of edges.
- 2) We design two new feature extraction modules called multiscale global prior extraction block (MGPB) and residual dense-in-dense block (RDDDB) as the basic architecture of the proposed model. Among them, the MGPB maps LR images from low-dimensional space to high-dimensional space for structure and texture information extraction and multiscale feature fusion, while the RDDDB makes full use of interblock local and global features to assist SR reconstruction.
- 3) We construct a model for SR reconstruction of remote sensing images, called STP. The proposed SR model consists of two branches, one branch is a structure branch for generating intermediate SR reconstruction results, and the other branch, called the texture branch, is used to reconstruct fine HR edge images and provide texture supervision for SR reconstruction.

II. RELATED WORK

A. Structure Preserving Remote Sensing Image SR

Structure preserving SR methods express and synthesize the input whole image as a variable, and utilize the overall

structural nature of the image to enhance the SR reconstruction performance, which is essentially a global structure-based approach, such as attention mechanism and multiscale model. Zhang et al. [24] designed the first SR method based on attention mechanism that biases the allocation of processing resources toward the most informative part of the input, and introduced a channel attention mechanism to propose a residual channel attention network. Afterwards, second-order attention network [25], holistic attention network [26], non-local sparse attention [27], efficient long-range attention [28], and hybrid attention-based U-shaped network (HAUNet) [29] have been developed for remote sensing image SR. Zhang et al. [30] proposed a mixed high-order attention network (MHAN) for solving the problem of under-representation of high-frequency details and high computational memory cost of first-order attention mechanisms. Wang et al. [29] proposed a new HAUNet that efficiently explores multiscale features and enhances global feature representation through hybrid convolution-based attention. While all of these methods have achieved some results, the reconstruction results are often too smooth and lack textural detail.

Multiscale modeling is another strategy. Although multiscale feature strategies are widely used in the field of remote sensing SR, the specific technical details are different. MSRN [31], MRNN [32], and MEN [33] use a combination of convolutional layers with multiple convolutional kernel sizes to refine the extraction of multiscale features and learn image multiscale features adaptively. However, this does not take full advantage of the multiscale features of the image, instead the larger the convolution kernel size increases the computational effort of the model. MSDNN [34] processes input images mainly in two different downscaling spaces, thus greatly reducing GPU memory usage. However, for the SR task, the use of downsampling operations in the feature extraction module results in a severe loss of structural and texture information of the image, and thus, this greatly limits the utilization of multiscale information. LapSRN [35] uses progressive upsampling to directly output the reconstruction results at multiple scales. However, this is still essentially a reuse of single-scale features, and does not achieve multiscale feature fusion. Unlike the aforementioned approaches, our multiscale strategy specifically maps LR images from low-dimensional space to high-dimensional space, performs multiscale feature extraction in the upscaling spaces of the image, and then, performs multiscale feature fusion. This not only avoids the loss of structure and texture information in the downscaling spaces, but also preserves the original multiscale structure and texture features, and at the same time fuses the multiscale features of the image, which is robust to SR reconstruction of images.

B. Texture Preserving Remote Sensing Image SR

In order to better mine the texture information of the image itself, some scholars have begun to assist in SR reconstruction by introducing edge or gradient priors of the image. Jiang et al. [20] proposed a new generative adversarial networks (GANs) framework based on HR edge vivid enhancement,

called EEGAN. EEGAN uses the edge image reconstructed from LR to SR as a prior extraction, and guides the final SR reconstruction through its designed edge enhancer subnet and edge enhancement mask. SPSR [21] uses an adversarial learning mechanism to first introduce LR gradient branch in the generator, synthesize the gradients of HR images by the gradients of LR images, and fuse them to improve the boundary quality of the generated images. However, GANs-based GANs would generate some high-fidelity results, but it also introduces geometric distortions, especially at edges and fine textures. In addition, the stable training of GANs is still a problem. Yang et al. [23] and Fang et al. [22] all used the same idea to design the loss function. They directly calculated the losses of SR_{edge} and HR_{edge} as constraints on the total loss function to assist SR. In addition, for the design of the loss function, the L_1 loss is also directly used as a constraint. However, this loss function is still essentially a peak signal-to-noise ratio (PSNR)-oriented method, and the SR images generated by such methods tend to be blurred or too smooth, resulting in missing details and obvious blurring. GEDRN [19] used shared source residual structure and nonlocal operations to learn rich low-frequency information and long-distance spatial correlations, and used gradient loss and perceptual loss to further improved the perceptual quality. However, it performed multiscale upsampling of images before feature extraction and extracts gradient information for the intermediate step of fusion. However, this preprocessing approach increases the model computation on the one hand, and introduces erroneous edge and texture information on the other hand.

III. STP REMOTE SENSING IMAGE SR

In this section, we will describe STP in detail. We first introduce the overall structure of STP and its discarding function. Then, the detailed structure design of the ISAM and structure/texture branch is introduced in detail. It should be noted that in this article, the network structure of the structure branch and the texture branch is the same.

A. Whole Network Architecture of STP

We show the overall framework diagram of STP in detail in Fig. 1. Let us denote the input LR image as I_{LR} (the size of I_{LR} is $h \times w \times c$), and its corresponding original HR size is $sh \times sw \times c$, where sh and sw , respectively, represent the height and width of HR, s is the scale factor, and c is the number of image bands. The purpose of our model is to reconstruct image I_{SR} from the input degraded image I_{LR} by the end-to-end manner.

The edge extraction [36] obtains a preliminary edge image $I_{LR_{edge}}$ through the Laplacian sharpening filter, and the input image and the edge image enter the network in parallel. We can define the Laplace operator of LR image $I_{LR}(x, y)$ as the second derivative as follows:

$$I_{LR_{edge}}(x, y) = \frac{\partial^2 I_{LR}(x, y)}{\partial x^2} + \frac{\partial^2 I_{LR}(x, y)}{\partial y^2} \quad (1)$$

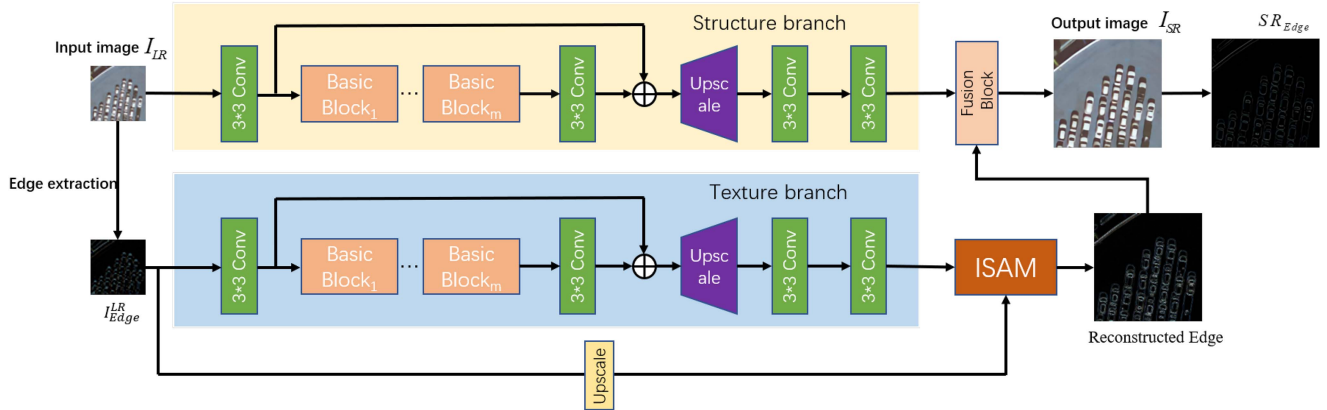


Fig. 1. Proposed STP framework. I_{LR} , I_{Edge}^{LR} , I_{SR} , and SR_{Edge} , respectively, represent input LR image, LR edge image, final SR output, and the SR edge image. Our architecture consists of two branches, the structure branch and the texture branch. The structure of the structure branch and the texture branch are the same. The edge image is extracted by the Laplacian operator. The texture branch aims to super resolve LR edge maps to the HR counterparts.

where $I_{LR_{edge}}(x, y)$ represents the LR edge image obtained from the aforementioned formula. The image edge obtained by modifying the second derivative can produce a steep zero crossing because of the isotropy and rotation invariance of the Laplace operator.

The obtained edge image passes through the texture branch to obtain the reconstructed edge image Re_{Edge} . The reconstructed edge image Re_{Edge} and the output of the structure branch are fused in the fusion block to obtain the final reconstructed SR image. The fusion block splices the feature maps of the two branches in spatial dimensions according to the channels, and uses the 1×1 convolutional layer to perform spatial dimensionality reduction to obtain the image I_{SR} reconstructed by the network SR. This process can be expressed by the formula

$$I_{SR} = \text{Conv}_{1 \times 1}(\text{Concat}(\text{Structure}_{\text{branch}}, \text{Texture}_{\text{branch}})) \quad (2)$$

where $\text{Conv}_{1 \times 1}$ denotes 1×1 convolutional layer, $\text{Structure}_{\text{branch}}$ and $\text{Texture}_{\text{branch}}$ represent the output of the two branches, and Concat represents the fusion operation. We choose the classic $L1$ loss as our loss function to train and optimize our model, which is as follows:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|F_{STP}^i(I_{LR}^i) - I_{HR}^i\|_1 \quad (3)$$

where $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$ is training set containing N pairs of LR and HR, and θ represents the parameter set.

B. Structure and Texture Branch

Structure and texture branch are the main parts of STP, and their design determines the performance of STP. Each branch consists of three parts. First, a 3×3 convolutional layer, which is used to extract shallow and rough features of the image. Then, there is the main part of the model composed of several basic blocks, which is used to extract the deep and fine features of the image. Finally, there is the upsampling and image reconstruction module, which is used to reconstruct the final HR image. The upsampling reconstruction module is composed of an upsampling layer and two 3×3 convolutional layers.

After this part, a preliminary SR image can be obtained. Overall structure diagram of the model is shown in Fig. 1. Next, we will focus on the backbone of our model. The trunk of our model is stacked by six basic blocks. Each basic block includes an MGPB and RDDB.

1) *Multiscale Global Prior Extraction Block (MGPB)*: For the image SR task, it is important to recover as much structural and texture information as possible, i.e., the high-frequency information of the image. The LR image has more low-frequency information, while the HR image has more high-frequency information. Based on this reality, we propose a new multiscale feature extraction strategy that is completely different from the previous multiscale strategies. We consider upsampling the image into HR space for feature extraction in the feature extraction stage, and mapping the feature vector from the low-dimensional space to the high-dimensional space. The feature vectors are first multiscale upsampled and mapped to multiscale high-dimensional space, followed by feature extraction through convolutional layers, and finally, the dimensionality is restored to be consistent for fusion to obtain multiscale features that are rich in structural texture features. In addition, remote sensing imaging distances are different, the conditions are complex, and the scales are different, resulting in local texture detail information often being lost, and the picture is too smooth. To solve this problem, we designed an MGPB, as shown in Fig. 2.

The MGPB is mainly used to map LR image from low-dimensional space to multiscale high-dimensional space of image, and finally, fuse multiscale features obtained. The MGPB consists of three parallel upsampling and downsampling units (UDN) of different scales, they can simultaneously obtain satellite LR image features at the scales of $2\times$, $4\times$, and $8\times$, respectively. For different scale UDNs, we use different sizes of convolution kernels to obtain different sizes of receptive fields, which can adapt to scale changes. Multiscale features can be fully extracted by adjusting the size of the convolution kernel of different scales.

Given an input feature f_{input} , it first passes through the 3×3 convolutional layer, and then, it is feature upsampled. Then,

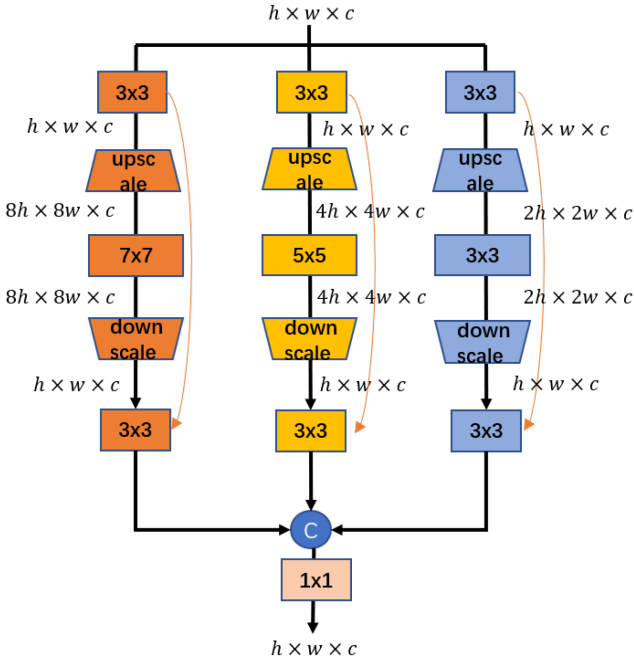


Fig. 2. Proposed MGPD structure. Multiscale features are extracted in the high-dimensional space so that the structural consistency between the original LR image and the edge image can be maintained to the greatest extent.

we use a convolutional layer to extract upscaled features (i.e., feature extraction in the high-dimensional space of the image), and conduct downsampling at the same scale. Finally, the output and input obtained by sampling features under the convolutional layer are added at the element level to obtain the final output result. It is worth noting that the upsampling and downsampling scales can be set by themselves to extract the features of images at different scales. This process can be expressed by publicity as follows:

$$f_{UDN} = (f_{input}) \oplus \text{Conv}(\text{Down}(\text{Up}(f_{input}))) \quad (4)$$

where f_{input} denotes the input feature, Down denotes the downsampling operation, Up denotes upsampling, Conv denotes convolutional layer, and \oplus denotes element addition.

Finally, multiscale feature fusion is carried out, and the dimension is reduced through the 1×1 convolutional layer after fusion to obtain the output result. The proposed MGPD uses UDNs to extract image features of different scales through residual learning, which can be expressed as

$$f_{MGPD} = \text{Conv}_{1 \times 1}(\text{Concat}(UDN_2, UDN_4, UDN_8)) \quad (5)$$

where Concat represents feature fusion, and UDN_2 , UDN_4 , and UDN_8 represents upsampling and downsampling of 2, 4, and 8, respectively.

2) *Residual Dense-in-Dense Block (RDDDB)*: For SISR, the use of the hierarchical features of LR images can effectively improve the performance of SR. The hierarchical features of the CNN will provide more clues for reconstruction. Most of the existing SISR methods (for example, EDSR [37], SRResNet [38], and SRDenseNet [39]) ignore the use of hierarchical features for reconstruction. RDN [40] and RRDBNet [41] used residual

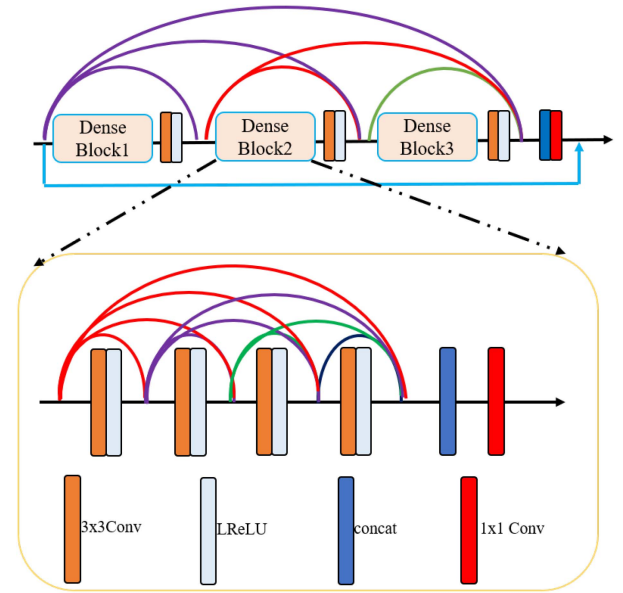


Fig. 3. Proposed RDDDB structure. An inner dense block consists of four convolutional layers and the LReLU function, and feature fusion is used between all layers instead of feature addition. The external RDDDB is composed of three dense blocks.

dense blocks (RDB) to extract rich local features. However, for the previously obtained multiscale fusion features, it can only extract the local features within the blocks, and the local and global features between the blocks have not been fully utilized.

In order to make full use of the obtained multiscale global fusion features, we optimize the network structure by integrating dense block and RDB, and designed a new RDDDB. As can be seen from Fig. 3, an internal dense block consists of four 3×3 convolutional layers and four LReLU functions. Here, all jump connections are not simple feature addition, but feature fusion so that the original multiscale fusion features obtained before can be retained to the greatest extent, and finally, a 1×1 convolution is used for dimensionality reduction. The external RDDDB is composed of three dense blocks, all of which are feature fusion. In order to stabilize the training, we add a convolutional layer and LReLU function after each dense block to calculate the fused features. It can be represented as

$$f_{Concat} = \text{Concat}(\text{Dense}(\text{dense}_1, \text{dense}_2, \text{dense}_3)) \quad (6)$$

where dense_1 , dense_2 , and dense_3 represent three dense blocks, Dense represents a dense connection of three dense blocks, and f_{Concat} represents an output feature after three dense blocks.

Finally, the dimensionality of the fusion features is reduced by a 1×1 convolution, and the input features and output features are added element level. We use the following formula as the output feature of RDDDB:

$$f_{RDDDB} = f_{MGPD} \oplus \text{Conv}_{1 \times 1}(f_{Concat}) \quad (7)$$

where f_{RDDDB} represents output characteristics of the RDDDB, f_{MGPD} represents the output characteristics, and \oplus represents element-level addition.

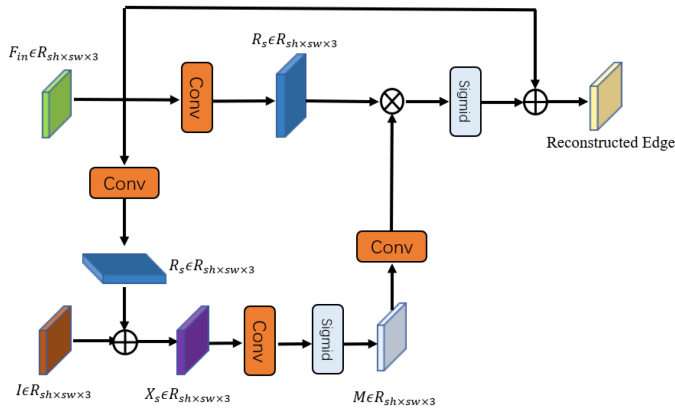


Fig. 4. Proposed ISAM structure. The high-frequency information is corrected through the interaction between the reconstructed SR edge and the LR edge, so as to supervise the final SR reconstruction.

C. Interactive Supervision Attention Module (ISAM)

The SR edge map obtained by directly extracting the edge map from the LR image after the texture branch may still be inaccurate. Therefore, we hope to obtain a fine edge map to provide supervision information, so as to obtain the final high-quality SR image. Inspired by [42], we designed an ISAM. The schematic diagram of the ISAM is shown in Fig. 4. Its contribution is mainly to provide high-frequency information correction for the edge provided by the texture branch.

As illustrated in Fig. 4, the ISAM takes the reconstructed SR edge map $F_{in} \in R_{h \times w \times 3}$ in the previous stage, and first uses a simple 3×3 convolution to generate the residual image $R_s \in R_{h \times w \times 3}$. The residual image is added to the SR edge image I obtained by directly upsampling the LR edge to obtain the restored image $X_s \in R_{h \times w \times 3}$. For this predicted image X_s , we provide clear supervision of the real image. Next, use 3×3 convolution and sigmoid activation to generate a per-pixel attention mask $M \in R_{h \times w \times 3}$ from image X_s . After talking about the attention mask feature through a 3×3 convolutional layer, add it to the previous residual image $R_s \in R_{h \times w \times 3}$, and then, recalibrate the converted local feature to guide the attention. The feature of is added to the original path. Finally, the attention enhancement reconstructed edge generated by the ISAM indicates that four is passed to the next stage for further processing.

For [42], the main application of SAM is in d multistage image restoration tasks, such as image defogging, deblurring, and denoising. It is inserted between each two stages to compute attention maps using the predictions of the previous stage, and these attention maps are used to refine the features of the previous stage before passing to the next stage for progressive learning. In other words, SAM is to provide supervised information for each stage of progressive image restoration while generating attention maps to suppress the less informative features of the current stage and allow only the useful features to propagate to the next stage.

Unlike [42], the main role of our proposed ISAM is to correct the LR edge maps in the edge branches to obtain finer

texture information to guide SR reconstruction. We analyze the shortcomings of existing edge enhancement and propose a novel edge enhancement strategy, which is to establish a mapping between LR edges and HR edges by CNNs, which is equivalent to converting the spatial distribution from LR edge sharpness to HR edge sharpness. Since most regions of the edge map are close to zero, CNNs can focus more on the spatial relationship of contours. As a result, the network can more easily capture the structural correlations, and thus, generate an approximate edge map of the SR image. Specifically, we edge correct the edge maps generated by the proposed ISAM for the intermediate process and add them to the results of the LR edge upsampling to obtain the final reconstructed edges. This reconstructed edge is ultimately used to help generate SR results with high confidence and clear content.

D. Discussion

In deep-learning-based remote sensing image SR, the introduction of the edge prior of the image by a fixed Laplace edge detection operator is a common measure, which has been applied in several articles such as [20], [22], and [23]. But in fact, different edges introduce their strategies differently. For example, in [22] and [23], the authors first introduced the same LR edge map as a complement to texture information, and then, directly calculated the L1 loss of SR_{edge} and HR_{edge} as a constraint on the total loss function to generate SR. However, the SR images generated by this approach are often blurred or too smooth, resulting in missing details and obvious blurring. In [20] and [21], the authors used an EESN approach for edge enhancement. Specifically, the authors perform edge extraction of the intermediate acquired HR images by the EESN, and then, extract and enhance the image contours by mask processing. Finally, the recovered intermediate images are combined with the enhanced edges to obtain the final SR results. However, due to its use of adversarial learning approach of the GAN architecture for training, on the one hand, the training is unstable, and on the other hand, the results of this approach often suffer from geometric distortion, and the generated textures are unnatural or even distorted.

Instead, we know that deep learning has an excellent ability to handle probabilistic transformations of pixel distributions, so we build up a mapping of LR_{edge} to HR_{edge} by CNNs. This edge mapping can be seen as another image, and thus, the image-to-image conversion technique can be used to learn the mapping between the two modalities. The conversion process is equivalent to transforming the spatial distribution from LR_{edge} sharpness to HR_{edge} sharpness. Since most of the region of the edge image is close to zero, CNNs can focus more on the spatial relationship of contours. As a result, the network may more easily capture the structure dependencies, and thus, generate approximate edge images for SR images. Therefore, based on the aforementioned considerations, we propose a novel edge enhancement strategy that differs from previous work in that we establish a mapping relationship between LR_{edge} and HR_{edge} . Specifically, LR edges are first introduced as a complement to texture details, and subsequently, an intermediate process edge image is obtained

after edge enhancement branch pairs. This intermediate process edge image is edge corrected by our proposed ISAM and summed with the results of LR_{edge} upsampling to obtain the final reconstructed edge. This reconstructed edge is finally used to assist in generating SR results with high confidence and clear content. We can extract the edge information of an image by introducing a learning-based approach, and we can achieve good performance. But introducing a learnable implicit edge detection module will add extra computation, and the model improvement performance is not significant compared to using the laplace operator directly. Therefore, in order to achieve a better tradeoff between computational effort and model, we directly use the laplace operator to extract the edges of images.

IV. EXPERIMENT RESULTS AND DISCUSSION

In this section, we first introduce the dataset, evaluation indicators, and implementation details. Then, the effectiveness of this method is verified by ablation experiments. Finally, the method is compared with the existing method, and experimental analysis is carried out.

A. Datasets

We conduct experiments on two common satellite image datasets Draper and NWPU-RESISC45 [43], and ensure that all algorithms use the same amount of training data. The Draper dataset contains more than 1000 HR aerial photos taken in Southern California. The original image size and resolution is 3099×2329 , we crop it to 192×192 pixels as HR, of which 1000 are used for training and verification, and 200 are used for testing. We conduct $\times 4$ down sampling of HR through bicubic down sampling, so the corresponding LR resolution is 48×48 .

The NWPU-RESISC45 dataset is a publicly available benchmark for remote sensing image scene classification created by Northwestern Polytechnic University. The dataset covers 45 categories, each with 700 images. We randomly selected 52 pictures from each category, a total of 2340. The size of the HR image is 256×256 pixels, of which 2250 is for training and 90 is for testing. We conduct $\times 4$ down sampling of HR through bicubic down sampling, so the corresponding LR resolution is 64×64 .

Real scenes have different sensors, degradation environments, resolutions, and times of image capture, so it may lead to a model trained on synthetic data that can generate artifacts on real data. However, we take this problem into account when considering the data for the model, and both of the data we use in this article take into account the aforementioned problem. As a result, our model has the ability to generalize in real application scenarios where HR images are not available.

B. Parameter Settings and Implementation Details of the Model

For the base line of our proposed model, the number of the proposed basic block is set to 6. We use the Adam [44] algorithm to train our model. The initial learning rate is 2×10^{-4} and decays continuously. Finally, we choose structural similarity (SSIM) [45], visual information fidelity (VIF) [46], feature

TABLE I
VERIFY THE EFFECT OF EDGE ENHANCEMENT

Methods	PSNR/dB	SSIM [45]	FSIM [47]	VIF [46]
STP without ISAM	34.12	0.8984	0.9154	0.5383
STP without edge	34.05	0.8955	0.9139	0.5386
STP	34.21	0.9014	0.9177	0.5390

STP without ISAM" indicates our model removes ISAM. "STP without edge" indicates that our model removes texture branch.

The bold value denote the best result.

similarity (FSIM) [47], and PSNR as our objective evaluation indexes. We compare our model with popular deep-learning-based methods such as MSRN [31], RCAN [24], EEGAN [20], SeaNet [22], MHAN [30], CTNet [48], and HSENet [49]. We adjust the hyperparameters of these methods and use the same data distribution to maximize their good performance.

C. Ablation Study

In this section, we perform ablation experiments to verify the effectiveness of the proposed method on the Draper dataset. It is mainly divided into two parts. The first is to verify the influence of edge enhancement on the reconstruction result, and last is to verify the influence of the number of the basic block on the reconstruction result. We will analyze this in detail as follows.

1) *Effect of Edge Enhancement and ISAM*: In order to verify the effectiveness of introducing LR edge enhancement and ISAM, we sequentially removed ISAM and texture branch, then retrained and tested the model. The results are shown in Table I. It can be seen that when the ISAM is removed alone, the PSNR is reduced from 34.21 to 34.12 dB, and when the entire texture branch is removed, the PSNR is reduced to 34.05 dB, which is a decrease of 0.16 dB.

We visualized the results of SR with the edges removed as well as the edge branches, as shown in Fig. 5. Fig. 5(a) shows the reconstructed results with ISAM removed, Fig. 5(b) shows the reconstructed results with edge branches removed, Fig. 5(c) shows the reconstructed results of our final model, and Fig. 5(d) shows the HR. As can be seen from the figure, when the ISAM and edge branching are removed, the visual effect of the reconstruction appears to degrade, losing texture and detail information in the sector, this confirms the effectiveness of our proposed ISAM and edge branching. This is because the edge of the image is the most essential feature of the image. Edges widely exist in the image between objects and background, between objects and objects. In many applications, the main focus of imaging is the feature details of the object field, such as the edge contour, etc. This requires edge enhancement processing on the object field to highlight the required detail information such as the edge contour. In the spatial domain, the edge part of the object field generally refers to the boundary of the object field or the place where the complex amplitude changes drastically. Looking at the edge part of the object field in the spectral domain, it refers to the high-frequency information of the spatial spectrum of the object field. Satellite images have a high degree of structure, so their SR reconstruction tasks must make good use of the existing structural prior information and pay attention to the reconstruction of their detailed parts. Therefore, the introduction of edge prior information can bring more accurate detail information

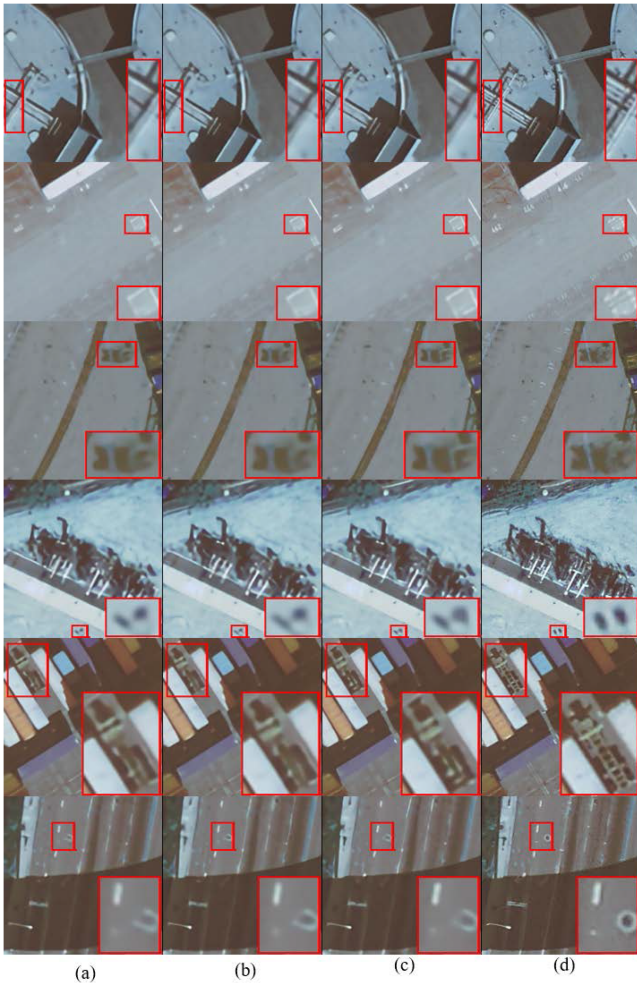


Fig. 5. Visualized the SR results with or without the texture branch. From left to right: (a) reconstructed results with ISAM removed, (b) reconstructed results with edge branches removed, and (c) reconstructed results of our final model. (d) HR.

TABLE II
VERIFY THE EFFECT OF MGPB AND RDDB

Methods	PSNR/dB	SSIM [45]	FSIM [47]	VIF [46]
STP without MGPB	33.92	0.8964	0.9130	0.5368
STP without RDDB	33.75	0.8925	0.9095	0.5330
STP	34.21	0.9014	0.9177	0.5390

“STP without MGPB” indicates our model removes MGPB. “STP without RDDB” indicates that our model removes RDDB.
The bold value denote the best result.

to the reconstructed image, which has more natural and true external characteristics.

2) *Effect of MGPB and RDDB*: In order to verify the effect of MGPB and RDDB, we executed a set of ablation experiments by removing the MGPB and RDDB modules and testing the experimental results on the Draper test set, respectively, the results of which are shown in the Table II. As seen in the table, when we remove the MGPB and RDDB modules, respectively, the performance of the model shows a significant decline in PSNR values by 0.29 and 0.46 dB, which demonstrates the role of our proposed MGPB and RDDB in our model.

TABLE III
EXPLORE THE IMPACT OF THE NUMBER OF BASIC BLOCKS ON THE RECONSTRUCTION RESULTS ON THE DRAPER TEST DATASET

Number of m	Scales	PSNR	SSIM [45]	FSIM [47]	VIF [46]
$m = 3$	×4	34.08	0.8989	0.9155	0.5364
$m = 4$		34.12	0.8981	0.9152	0.5367
$m = 5$		34.01	0.8945	0.9116	0.5349
$m = 6$		34.21	0.9014	0.9177	0.5390
$m = 7$		34.12	0.8989	0.9152	0.5389
$m = 8$		34.05	0.8952	0.9126	0.5362
$m = 9$		34.05	0.8951	0.9123	0.5370

The bold value denote the best result.

TABLE IV
VERIFY THE EFFECT OF THE KERNEL SIZE

kernel size	PSNR/dB	SSIM [45]	FSIM [47]	VIF [46]
5×5	34.09	0.8984	0.9153	0.5369
7×7	34.15	0.9000	0.9171	0.5387
16×16	34.15	0.8996	0.9165	0.5382
3×3 (final)	34.21	0.9014	0.9177	0.5390

The bold values denote the best results.

3) *Effect of m* : In this part, we will explore the impact of the network depth on the final reconstructed image effect by controlling the number of basic blocks. We use M to represent the number of basic blocks and test the performance of image reconstruction when the number of basic blocks increases from 3 to 9 with stride 1. The experimental results are shown in Table III. From the data in the table, the model achieves the best performance when m is set to 6.

4) *Effect of Kernels Size in MGPB*: We conducted ablation experiments on the effect of different convolution kernel sizes on the reconstruction effect in MGPB, and the experimental results are shown in the following table. From the Table IV, we can see that the reconstruction performance of the network is the best when the convolutional kernel size is 3×3 . And when the convolutional kernel size is too large, not only the performance decreases, but also leads to an increase in the computational cost of the model.

In general, we believe that the performance of the model improves as the width of the depth increases. However, as the width of the depth increases, the information flow is transmitted more, the feature loss increases, and the computational effort of the model increases. Therefore, we explore the actual impact of network depth and width on network performance through ablation experiments, while reaching a compromise with the amount of computation.

D. Compared on Draper Dataset and NWPU-RESISC45 Dataset

We compare our methods with popular deep learning based methods (MSRN [30], RCAN [24], EEGAN [20], SeaNet [22], MHAN [30], CTNet [48], and HSENet [49]). We adjust the superparameters of these methods and use the same data distribution to optimize their performance.

Tables V and VI show the average performance of the proposed method and other competitive algorithms based on deep learning on the two open datasets, where red font represents the optimal result and blue font represents the suboptimal result. It

TABLE V
COMPARISON RESULTS OF AVERAGE OBJECTIVE EVALUATION INDEX RESULTS FOR $\times 4$ SATELLITE IMAGE SR ON DRAPER DATASET AND PERFORM EXPERIMENTS ON SIMILAR SIZE

Methods	PSNR	SSIM [45]	FSIM [47]	VIF [46]
Bicubic	30.84	0.8217	0.8481	0.4223
MSRN [31]	33.57	0.8877	0.9078	0.5221
RCAN [24]	33.62	0.8919	0.9126	0.5195
EEGAN [20]	33.17	0.8781	0.9021	0.4985
SeaNet [22]	33.29	0.8907	0.9089	0.5164
MHAN [30]	33.14	0.8819	0.9035	0.5002
CTNet [48]	33.03	0.8810	0.8997	0.5128
HSENet [49]	33.49	0.8901	0.9094	0.5240
Ours	34.21	0.9014	0.9177	0.5390

The red values denote the best results, and blue values denote the suboptimal results.

TABLE VI
COMPARISON RESULTS OF AVERAGE OBJECTIVE EVALUATION INDEX RESULTS FOR $\times 4$ SATELLITE IMAGE SR ON NWPU-RESISC45 DATASET AND PERFORM EXPERIMENTS ON SIMILAR SIZE

Methods	PSNR	SSIM [45]	FSIM [47]	VIF [46]
Bicubic	27.24	0.6785	0.7807	0.3673
MSRN [31]	28.61	0.7537	0.8431	0.4119
RCAN [24]	28.65	0.7556	0.8451	0.4126
EEGAN [20]	28.09	0.7358	0.8352	0.3909
SeaNet [22]	28.59	0.7483	0.8371	0.4107
MHAN [30]	28.41	0.7395	0.8311	0.4126
CTNet [48]	28.57	0.7499	0.8373	0.4139
HSENet [49]	28.75	0.7577	0.8430	0.4208
Ours	28.93	0.7618	0.8464	0.4250

The red values denote the best results, and blue values denote the suboptimal results.

can be seen from the table that the proposed method is obviously superior to other algorithms in all objective evaluation indicators. Specifically, on the Draper dataset, the PSNR value of the proposed method is 0.59 dB higher than that of the suboptimal method; on the NWPU-RESISC45 dataset, the PSNR value of this method is 0.18 dB higher than that of the suboptimal method.

In order to show the performance of our methods more intuitively, we show the subjective visual effects of different methods in Figs. 6 and 7. As can be seen from the figure, the bicubic interpolation method cannot produce additional details. For CNN-based technologies, such as CTNet, HSENet, EEGAN, and MSRN, they can infer some texture details, but their global optimization schemes and low feature utilization lead to blurred image contours. Some methods based on attention mechanism, such as RCAN and MHAN, produce too smooth results, artifacts will be generated at the edges of the generated image, and texture details are also very fuzzy. In the figure, we use red boxes to mark the reconstruction results of some details. It can be seen that our method can reconstruct images with more realistic texture details, and generate few artifacts.

In addition, we also visualized the number of parameters and PSNR of several algorithmic models against our own algorithmic model, as shown in Fig. 8. As can be seen from the figure, the number of parameters of our model is larger than CTNet, HSENet, SeaNet, EEGAN, and MSRN, and smaller than RCAN and MHAN. However, compared to them, our performance is the best. In summary, our model achieves a tradeoff between the number of parameters and model performance.

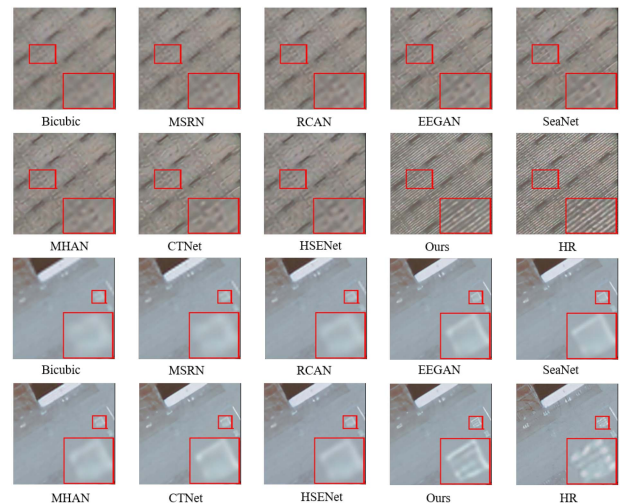


Fig. 6. Result of the subjective visual effect comparison between our method and other methods on the Draper remote sensing dataset, we use the red wire frame to mark out where the details of our reconstruction are better than other methods.

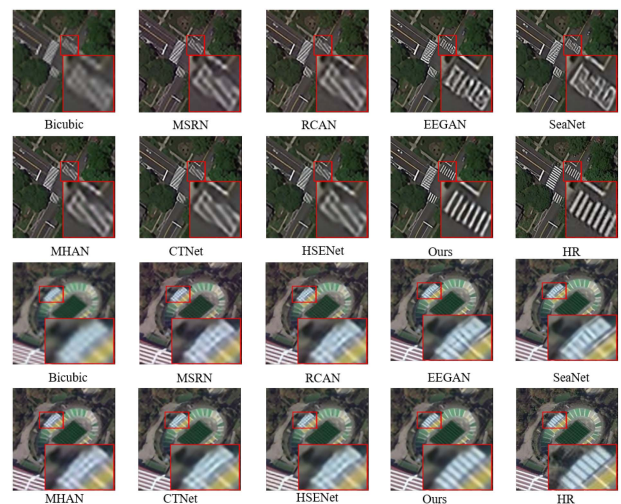


Fig. 7. Result of the subjective visual effect comparison between our method and other methods on the NWPU-RESISC45 remote sensing dataset, we use the red wire frame to mark out where the details of our reconstruction are better than other methods.

We also show the MSE between SR and HR. we observe that the proposed approach outperforms other competing algorithms, as demonstrated by its ability to better recover textures and structures, e.g., the contours of curb lines. Specifically, satellite images with complex textures and dense objects complex textures and dense objects are easily contaminated by artifacts. The proposed model leads to accurate contours, as shown in Fig. 9.

E. Validation Multiscale Robustness

In order to verify the effectiveness and robustness of our method to different input scales reconstruction results, we perform some multiscale experiments in this part. We choose the OpenBayes dataset for this set of experiments. The OpenBayes

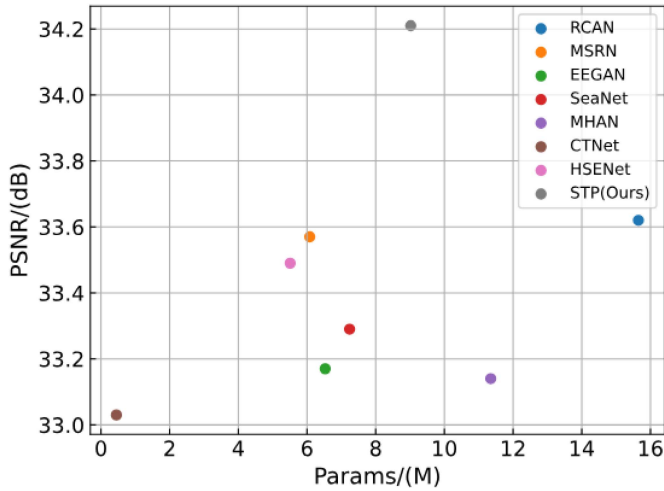


Fig. 8. Comparison with other models in terms of parameters and performance.

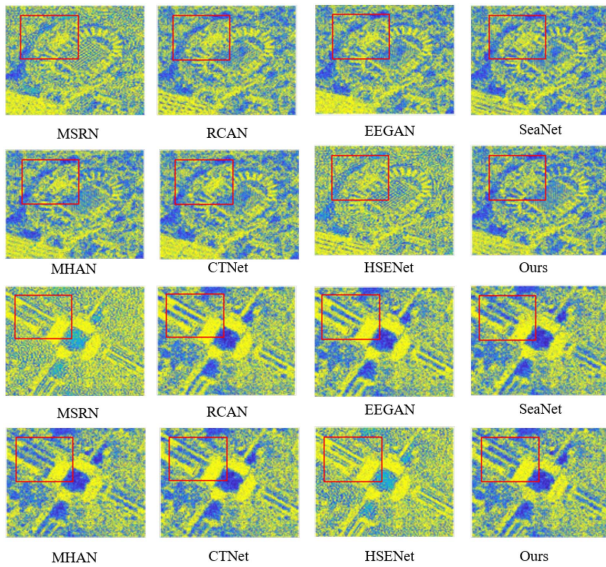


Fig. 9. Result of MSE visual effect comparison between our method and other methods, we use the red wire frame to mark out where the details of our reconstruction are better than other methods.

dataset is a small-scale land classification dataset that contains seven common categories captured from Google Earth. The seven categories include buildings, roads, bare soil, water, grasslands, playgrounds, and cultivated land. The OpenBayes database contains 303 images with a spatial resolution of 1.2 m (250 images for training and verification, and 53 images for testing). We cropped the HR image into a patch of $560 \times 560 \times 3$ pixels. We downsample the HR by $2\times$, $4\times$, and $8\times$ to get LR images with resolutions of $280 \times 280 \times 3$, $140 \times 140 \times 3$, $70 \times 70 \times 3$, and $35 \times 35 \times 3$. We choose HSENet [49], CTNet [48], and MHAN [30] as the comparison objects. The objective evaluation results are shown in Table VII.

As we can see that our method has achieved very good reconstruction results on different scales, and achieved the best results in objective evaluation indicators. At the same time, we show the competent visual effects of some reconstruction results

TABLE VII
COMPARISON RESULTS OF AVERAGE OBJECTIVE EVALUATION INDEX RESULTS FOR MULTISCALE SATELLITE IMAGE SR WITH OTHER APPROACHES ON OPENBAYES DATASET

Method	Scale	PSNR/dB	SSIM [45]	FSIM [47]	VIF [46]
bicubic	x2	31.27	0.8868	0.9932	0.5792
HSENet [49]		34.54	0.9391	0.9984	0.6590
CTNet [48]		33.21	0.9224	0.9971	0.6148
MHAN [30]		33.51	0.9264	0.9979	0.6287
Ours		34.94	0.9444	0.9986	0.6726
bicubic	x4	26.43	0.6888	0.9081	0.3758
HSENet [49]		28.01	0.7677	0.9546	0.4340
CTNet [48]		27.25	0.7330	0.9404	0.3939
MHAN [30]		27.63	0.7509	0.9483	0.4158
Ours		28.23	0.7758	0.9576	0.4429
bicubic	x8	23.51	0.5127	0.7582	0.2100
HSENet [49]		24.45	0.5703	0.8323	0.2408
CTNet [48]		24.01	0.5387	0.8123	0.2075
MHAN [30]		24.34	0.5648	0.8237	0.2343
Ours		24.67	0.5865	0.8398	0.2548

The red values denote the best results, and blue values denote the suboptimal results.

at different scales, as shown in Fig. 10. In Fig. 10, we show three different scale reconstruction subjective visual effects. From top to bottom, there are $2\times$, $4\times$, and $8\times$ SR reconstructions results. In the figure, we can see that our method has shown advantages on three different scales, and it is better than other methods in the restoration of texture details. But when the reconstruction scale is too large ($8\times$), we can see that almost all methods have failed, and the recovery results on texture details are far from HR, far lower than the desired results. Therefore, designing a large-scale satellite image SR method is still a challenging problem.

F. Effectiveness of Postprocessing

We use iterative self-organizing data analysis technology algorithm (ISODATA), which is a classic unsupervised semantic segmentation of satellite images, to evaluate the results from different SR methods. We set the number of classification categories to 5 and the maximum number of iterations to 5, as in [50]. Fig. 11 shows the SR and classification results. The white box highlight areas where the proposed method is superior to other methods.

Specifically, bicubic's results contain significant blur patterns and spectral distortions; therefore, it is difficult to obtain fine classification results through ISODATA. In particular, we noticed that in the classification results in CTNet [48] and HSENet [49], we find that the car classification in the upper left corner has obvious errors. We also observed the lack of texture information in the results of MSRN [31], RCAN [24], and SeaNet [22], while EEGAN [20] had a classification error in the rightmost truck in the white box, and MHAN [30] failed to recover the trucks and cars that were close to the left in

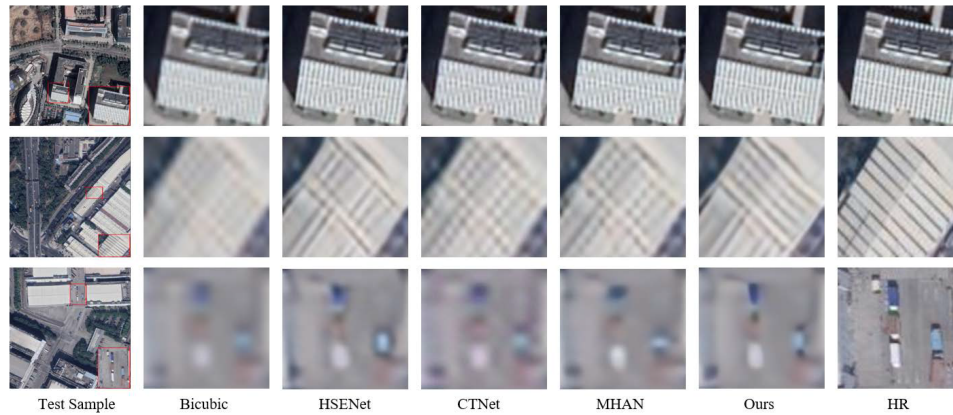


Fig. 10. Qualitative comparison of the proposed method with four counterparts on a typical satellite image from the OpenBayne dataset. From top to bottom, the reconstruction results are $2\times$, $4\times$, and $8\times$, respectively.

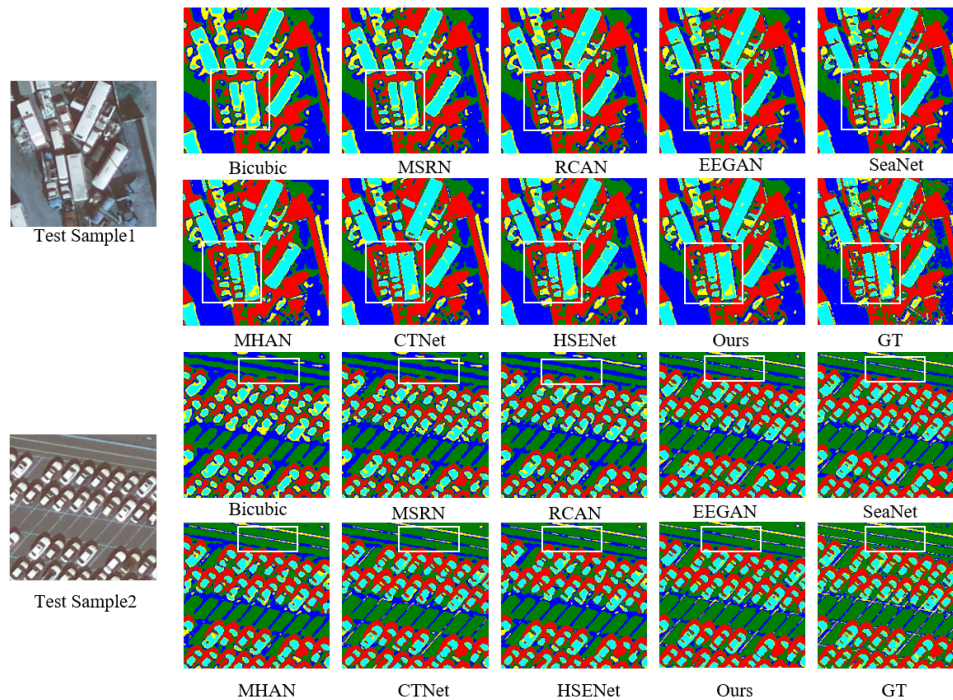


Fig. 11. Classification result through the ISODATA classification method. The white boxes highlight areas in the proposed model that are superior to other methods. Zoom in to see more details.

the white box. On the contrary, our method achieves a finer texture, which is demonstrated at the junction of car and truck in the white box, and the ISODATA classification result of the proposed method includes more accurate and finer details. But at the same time we should also note that due to the limitations of unsupervised semantic segmentation methods, all methods, including HR, have some misclassifications, but in general our method is the closest to ground truth visually. This shows that the proposed method can achieve high fidelity.

G. Compared on SuperView-1 Satellite Imagery

In order to further verify the efficiency of the proposed method in SuperView-1 satellite images, we compared the proposed

method with EEGAN [20], MHAN [30], CTNet [48], and HSENet [49] (four methods dedicated to satellite image SR) on SuperView-1 satellite images (1080×1080 pixels) collected from the SuperView-1 satellite. The spatial resolution of images captured by the SuperView-1 satellite is 0.5 m, which is lower than that of many HR remote sensing images. We use the center area ($256 \times 256 \times 3$ pixels) of the cropped image of typical urban land cover categories (including roads, buildings, vegetation, and bare soil) as the test image. We use the classic nonreference image evaluation metric, NIQE [51], where a smaller value indicates a better model performance. The assessment and the results are shown in Table VIII. It can be seen that our model achieves the best performance. Fig. 12 shows the SR image reconstructed from the proposed method and other

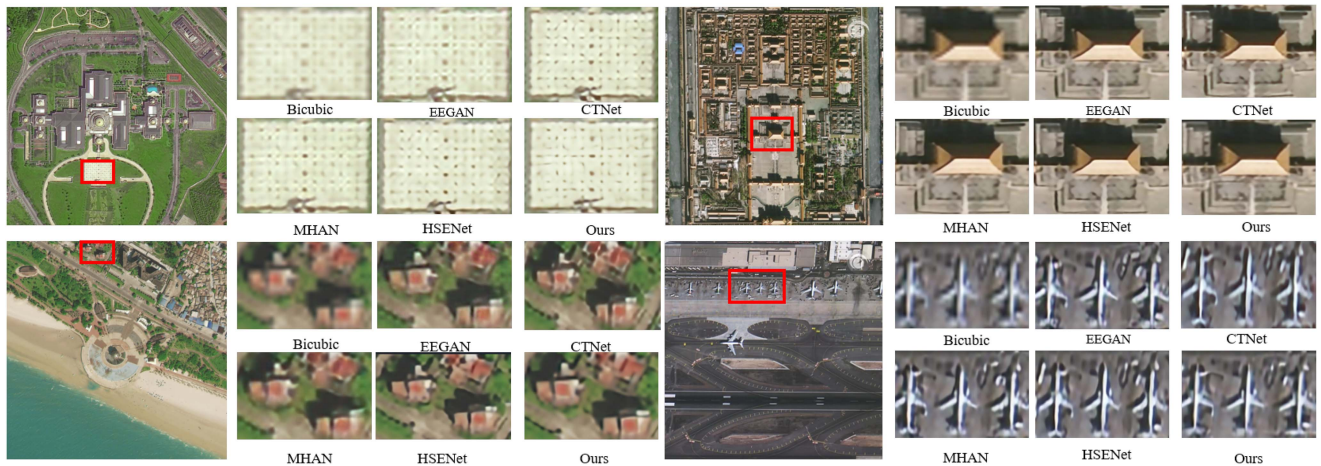


Fig. 12. Visual effects of SuperView-1 satellite images are compared.

TABLE VIII
COMPARISON RESULTS OF NIQE OF $\times 4$ SR RECONSTRUCTED IMAGES ON SUPERVIEW-1

Method	NIQE
Bicubic	6.815951
MHAN [30]	5.117883
CTNet [48]	5.425047
HSENet [49]	5.285778
EEGAN [33]	5.322917
Ours	5.098734

The bold value denote the best result.

competing methods on the SuperView-1 satellite image, with an upsampling factor of 4. There are significant artifacts in the reconstruction results of EEGAN, MHAN, CTNet, and HSENet, even in low-frequency regions (see the lawn and bare soil in the image). These artifacts pose a huge challenge to the subsequent remote sensing image processing steps. On the contrary, the proposed method achieves fine-grained texture details at the edges with clearer image content. The aforementioned results from video satellite images further reveal the effectiveness of this method and highlight its powerful artifacts removal ability.

H. Limitations and Future Work

Our algorithm also has limitations, that is, our method can only perform high-quality reconstruction of single-frame images, but cannot reconstruct LR satellite videos well. This is because, compared with static satellite images, imaging jitter and inconsistency in irradiation intensity caused by high-speed satellite motion reduce the availability of imaging. Second, compared with still satellite images, video satellites have the characteristics of continuous observation in time. Finally, the compression ratio of video satellite data is large, which further reduces the quality of video images. Therefore, in future research, we will focus on SR reconstruction algorithms for video satellite application scenarios.

V. CONCLUSION

In this article, we proposed an STP method for remote sensing image SR. Specifically, we propose a novel edge prior enhancement strategy that uses the edges of LR images and the proposed ISAM to guide SR reconstruction. First, we introduce the LR edge map as a priori structural expression for SR reconstruction, which further enhances the SR process with edge preservation capability. In addition, to obtain finer texture edge information, we propose a novel ISAM in order to correct the initial LR edge map with high-frequency information. By introducing and LR edges and ISAM-corrected HR edges, we build LR–HR edge mapping to preserve the consistency of LR and HR edge structure and texture, which provides supervised information for SR reconstruction. We designed two novel feature extraction modules called MGPB and RDDB. With these two modules, we explore the salient features of the image and its edges in the ascending space, and restored the difference between LR and HR images by residual and dense learning. We conducted a large number of experiments including SR and classification on three RGB remote sensing image datasets. Comprehensive experimental results show that, compared with the most advanced methods, our method can reconstruct sharp edges and clean image content, which is more realistic and faithful to the ground truth. It is expected that the method we proposed can be effectively transferred to other remote sensing image restoration fields, such as remote sensing image denoising and deblurring.

REFERENCES

- [1] Z. Zhang, H. Guo, J. Yang, X. Wang, and Y. Du, "Adversarial network with higher order potential conditional random field for PolSAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1795–1812, Oct. 2024.
- [2] W. He, Y. Yang, S. Mei, J. Hu, W. Xu, and S. Hao, "Configurable 2D-3D CNNs accelerator for FPGA-based hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9406–9421, Oct. 2023.
- [3] Y. Choi, J.-H. Choi, and Y. Choi, "Fit-for-purpose approach for the detection and analysis of earthquake surface ruptures using satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9574–9589, Oct. 2023.

- [4] X. Guo, M. Anisetti, M. Gao, and G. Jeon, "Object counting in remote sensing via triple attention and scale-aware network," *Remote Sens.*, vol. 14, no. 24, 2022, Art. no. 6363.
- [5] X. Geng, G. Dong, Z. Xia, and H. Liu, "SAR target recognition via random sampling combination in open-world environments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 331–343, Dec. 2023.
- [6] J. Qin, Z. Liu, L. Ran, R. Xie, J. Tang, and Z. Guo, "A target SAR image expansion method based on conditional Wasserstein deep convolutional GAN for automatic target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7153–7170, Aug. 2022.
- [7] M. Shi, Y. Gao, L. Chen, and X. Liu, "Double prior network for multi-degradation remote sensing image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3131–3147, Feb. 2023.
- [8] M. Shi, Y. Gao, L. Chen, and X. Liu, "Multisource information fusion network for optical remote sensing image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3805–3818, Feb. 2023.
- [9] J. Shang, M. Gao, Q. Li, J. Pan, G. Zou, and G. Jeon, "Hybrid-scale hierarchical transformer for remote sensing image super-resolution," *Remote Sens.*, vol. 15, no. 13, 2023, Art. no. 3442.
- [10] J. Wang, T. Lu, X. Huang, R. Zhang, and X. Feng, "Pan-sharpening via conditional invertible neural network," *Inf. Fusion*, vol. 101, 2024, Art. no. 101980.
- [11] J. Wang, Q. Zhou, X. Huang, R. Zhang, X. Chen, and T. Lu, "Pan-sharpening via intrinsic decomposition knowledge distillation," *Pattern Recognit.*, vol. 149, 2024, Art. no. 110247.
- [12] T. Lu, K. Zhao, Y. Wu, Z. Wang, and Y. Zhang, "Structure-texture parallel embedding for remote sensing image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2022.
- [13] Y. Wang, T. Lu, Y. Yao, Y. Zhang, and Z. Xiong, "Learning to hallucinate face in the dark," *IEEE Trans. Multimedia*, vol. 26, pp. 2314–2326, Jul. 2024.
- [14] Y. Wang, T. Lu, Y. Zhang, Z. Wang, J. Jiang, and Z. Xiong, "Faceformer: Aggregating global and local representation for face hallucination," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2533–2545, Jun. 2023.
- [15] T. Lu, Y. Wang, Y. Zhang, J. Jiang, Z. Wang, and Z. Xiong, "Rethinking prior-guided face super-resolution: A new paradigm with facial component prior," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 2, 2022, doi: [10.1109/TNNLS.2022.3201448](https://doi.org/10.1109/TNNLS.2022.3201448).
- [16] J. Wang, Z. Shao, X. Huang, T. Lu, R. Zhang, and J. Ma, "Enhanced image prior for unsupervised remote sensing super-resolution," *Neural Netw.*, vol. 143, pp. 400–412, 2021.
- [17] K. Zeng, Z. Wang, T. Lu, J. Chen, J. Wang, and Z. Xiong, "Self-attention learning network for face super-resolution," *Neural Netw.*, vol. 160, pp. 164–174, 2023.
- [18] D. Mishra and O. Hadar, "Accelerating neural style-transfer using contrastive learning for unsupervised satellite image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, Sep. 2023.
- [19] Z. Zhang et al., "Gradient enhanced dual regression network: Perception-preserving super-resolution for multi-sensor remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Dec. 2022.
- [20] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.
- [21] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7766–7775.
- [22] F. Fang, J. Li, and T. Zeng, "Soft-edge assisted network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 4656–4668, Feb. 2020.
- [23] W. Yang et al., "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
- [24] 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.*, 2018, pp. 286–301.
- [25] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11065–11074.
- [26] B. Niu et al., "Single image super-resolution via a holistic attention network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 191–207.
- [27] Y. Mei, Y. Fan, and Y. Zhou, "Image super-resolution with non-local sparse attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3517–3526.
- [28] X. Zhang, H. Zeng, S. Guo, and L. Zhang, "Efficient long-range attention network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 649–667.
- [29] 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, pp. 1–15, Jun. 2023.
- [30] D. Zhang, J. Shao, X. Li, and H. T. Shen, "Remote sensing image super-resolution via mixed high-order attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5183–5196, Jun. 2021.
- [31] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 517–532.
- [32] T. Lu, J. Wang, Y. Zhang, Z. Wang, and J. Jiang, "Satellite image super-resolution via multi-scale residual deep neural network," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1588.
- [33] Y. Wang, Z. Shao, T. Lu, C. Wu, and J. Wang, "Remote sensing image super-resolution via multiscale enhancement network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, Feb. 2023.
- [34] S. Gao and X. Zhuang, "Multi-scale deep neural networks for real image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2006–2013.
- [35] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5835–5843.
- [36] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, May 2007.
- [37] 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*, 2017, pp. 136–144.
- [38] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.
- [39] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4799–4807.
- [40] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.
- [41] X. Wang et al., "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.
- [42] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14816–14826.
- [43] 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.
- [44] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [45] 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.
- [46] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [47] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [48] S. Wang, T. Zhou, Y. Lu, and H. Di, "Contextual transformation network for lightweight remote-sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Dec. 2022.
- [49] S. Lei and Z. Shi, "Hybrid-scale self-similarity exploitation for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, Apr. 2022.
- [50] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2018.
- [51] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.



Kanghui Zhao received the B.S. degree in computer science and technology from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, in 2019, and the master's degree in computer technology from the Wuhan Institute of Technology, Wuhan, in 2022. He is currently working toward the Ph.D. degree in industrial engineering under the supervision of Prof. Tao Lu in the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology.

His research interests include image/video processing and computer vision.



Tao Lu (Member, IEEE) received the B.S. and M.S. degrees in computer science and technology from the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China, in 2003 and 2008, respectively, and the Ph.D. degree in communication and information systems from National Engineering Research Center For Multimedia Software, Wuhan University, in 2013.

He is currently a Professor with the School of Computer Science and Engineering, Wuhan Institute of Technology, and a Research Member with Hubei Provincial Key Laboratory of Intelligent Robot. He was a Postdoc with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, from 2015 to 2017. His research interests include image/video processing, computer vision, and artificial intelligence.

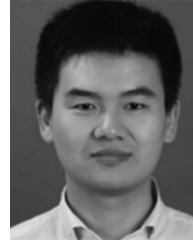


Yanduo Zhang received the Ph.D. degree in aircraft design from the Harbin Institute of Technology, Harbin, China, in 1999.

He was a Visiting Professor with the Center for Vision, Cognition, Learning, and Art, University of California at Los Angeles, Los Angeles, from 2012 to 2013. He is currently a Professor with the Wuhan Institute of Technology, Wuhan, China. He is also the President with the Hubei University of Arts and Science and the Director of the Hubei Province Key Laboratory of Intelligent Robot. He has coauthored

two books and more than 100 research articles. His research interests include computer vision, robot control, and intelligent computing.

Dr. Zhang is a Member of the ACM.



Junjun Jiang (Senior Member, IEEE) received the B.S. degree in computer science and technology from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree in communication and information systems from the School of Computer, Wuhan University, Wuhan, China, in 2014.

From 2015 to 2018, he was an Associate Professor with the China University of Geosciences, Wuhan. He was a Project Researcher with the National Institute of Informatics, Tokyo, Japan, from 2016 to 2018. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision.

Dr. Jiang was the recipient of the Finalist of the World's FIRST 10 K Best Paper Award at IEEE International Conference on MultiMedia and expo (ICME) 2017, Best Student Paper Runner-Up Award at International Conference on MultiMedia Modeling (MMM) 2015, Best Paper Award at International Forum of Digital Multimedia Communication (IFTC) 2018, 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award, and 2015 Association for Computing Machinery (ACM) Wuhan Doctoral Dissertation Award.



Zhongyuan Wang (Member, IEEE) received the Ph.D. degree in communication and information systems from Wuhan University, Wuhan, China, in 2008.

He is currently a Professor with the School of Computer Science, Wuhan University. He is also directing three projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications.



Zixiang Xiong (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana—Champaign, Champaign, IL, USA, in 1996.

Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, where he is currently a Professor.

Dr. Xiong has served as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1999 to 2005, IEEE TRANSACTIONS ON IMAGE PROCESSING from 2002 to 2005, IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2006, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B: CYBERNETICS from 2005 to 2009, and the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2013. He is also an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA.