# A Recurrently Complicated Lightweight Network for Superresolution of Remote Sensing Images

Duwei Hua, Kunping Yang, *Member, IEEE*, Jianchong Wei, Liang Chen, *Member, IEEE*, Dingli Xue, and Yi Wu, *Member, IEEE*

***Abstract*—Convolutional neural network (CNN) has made significant progress in image superresolution (SR), which could thrash the limits of image spatial resolution. Recently, abundant CNN-based methods have been proposed for the remote sensing image SR; however, the usages of complex structures and coarse manners could introduce excessive learnable parameters and ignorance of heterogeneous image details, respectively. In this article, we propose a recurrently complicated lightweight network (RCL-Net) for SR image recovery, through procedures of recurrent fluctuated complexity. We design a serial of the progressive complicated block (PC-Block) in the RCL-Net, and each PC-Block is composed of three complicated lightweight branches (CL-Branches) with increasing complexities in order, for recovering heterogeneous image details. Meanwhile, the CL-Branch is integrated with a multireceptive field module (MRF-Module) to more efficiently recover intact images through forward propagation paths of heterogeneous routes and lengths, where the excessive interactive calculations between feature subparts are constrained to reduce learnable parameters. In this manner, the proposed RCL-Net achieves a tradeoff between model complexities of traditional powerful structures, such as coarse-to-fine manners, and SR performances. Plentiful experiments with excellent results grounded on popular datasets exactly demonstrate the superiority of our proposed network, which even surpasses the advanced large SR model with less than 3% learnable parameters, compared to the state-of-the-art lightweight methods.**

***Index Terms*—Lightweight network, recurrent complicated structures, remote sensing, superresolution (SR).**

Manuscript received 13 January 2024; revised 9 April 2024 and 3 May 2024; accepted 29 May 2024. Date of publication 13 June 2024; date of current version 1 July 2024. This work was supported by the National Natural Science Foundation of China under Grant 62301161, Grant U1805262, Grant 61901117, and Grant 62201152 in part by the Special Funds of the Central Government Guiding Local Science and Technology Development under Grant 2021L3010, in part by the Key Provincial Scientific and Technological Innovation Projects under Grant 2021G02006, and in part by the Natural Science Foundation of Fujian Province, China, under Grant 2022J01169 and Grant 2020J01157. *(Corresponding author: Kunping Yang.)*

Duwei Hua, Kunping Yang, Liang Chen, and Yi Wu are with the Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, Fujian Provincial Key Laboratory of Photonics Technology, Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350117, China (e-mail: qbx20200096@yjs.fjnu.edu.cn; kunpingyang@fjnu.edu.cn; cl_0827@126.com; wuyi@fjnu.edu.cn).

Jianchong Wei is with the College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350117, China, and with the College of Electronics and Information Science, Fujian Jiangxia University, Fuzhou 350108, China (e-mail: jjasonway@fjjxu.edu.cn).

Dingli Xue is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100864, China (e-mail: xuedl@aircas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3413838

## I. Introduction

As a long-standing low-level computer vision problem, single-image superresolution (SISR) [1], [2], [3], [4] refers to reconstructing the high-resolution (HR) image from its low-resolution (LR) counterpart, which has been widely utilized to assist advanced computer vision tasks, such as person re-identification [5], [6], medical imaging [7], remote sensing investigations [8], [9], etc. In the research on remote sensing images, the acquisition of high-quality HR images is essential for subsequent studies, such as object detection [10], urban planning [11], semantic labeling [12], etc. However, the methods with complicated structures would be dramatically constrained by high model complexities in real-world applications. Therefore, studying how to effectively improve the remote sensing image spatial resolution through lightweight models is of great significance.

In the early stage, SISR mainly concentrated on natural images, among which the interpolation-based SR methods [13], [14] and reconstruction-based SR methods were successively proposed [15], [16]. However, the oversmoothing results and poor reconstruction accuracy began to disturb the researchers. Recently, ResDiff [17] combined convolutional neural network (CNN) and diffusion probabilistic models (DPM), where the CNN is utilized to restore primary low-frequency components and the DPM is utilized to predict the residual information. In addition, ESRGAN [18] and Real ESRGAN [19] have been proposed to utilize synthetic data for SR. Meanwhile, deep-learning-based SR methods [1], [2], [3], [4], [20], [21], [22] have gradually become popular due to the remarkable performances. For instance, VDSR [23] deepened the architecture to obtain promising results while RCAN [24] introduced attention mechanisms [25] for better performances. Afterward, SwinIR [26] introduced transformer structures [27] to fully excavate the feature information, while HAT [28] combined channel attention and window-based self-attention to introduce more global context into the transformer. Moreover, to overcome the shortcomings of transformer-based SISR methods, a novel transformer architecture HIPA [29] that progressively recovers the HR image using a hierarchical patch partition has been designed. Nevertheless, the aforementioned methods mainly employed complicated structures to restore SR images, which would greatly increase the model parameters. To solve the excessive parameter issue, several lightweight networks [30], [31] introduced recursive learning to accomplish reconstructions while still accompanying

high computations. Meanwhile, CARN [32] further introduced the cascading mechanism to exploit hierarchical features. Moreover, IDN [33] and IMDN [34] employed distillation strategies to decrease model complexities while ECDP [35] designed a continuous time conditional diffusion model for image super-resolution (SR) to enable the use of probability flow sampling for reducing the time consumption of DPM.

Besides the natural image SR research, remote sensing image SR has drawn increasing attractions among the community attributing to the significantly wide applications. In the beginning, sparse-representation-based methods with few parameters were the leading techniques, which could recover remote sensing SR images with seriously built overcomplete dictionaries. For example, Song et al. [36] combined the global dictionary and sparse coding to restore LR Landsat images. Meanwhile, for the complicated imaging conditions of remote sensing images, Shao et al. [37] designed a novel coupled sparse autoencoder to effectively learn the mapping functions between LR and HR images. Considering the information scarcity of raw image data, Li et al. [38] designed two multitemporal dictionary learnings to restore the image details. However, even though the sparse representation is naturally lightweight, the aforementioned methods could encounter difficulties when recovering image details due to limited feature representation abilities.

Recently, deep-learning-based SR methods have become the primary architectures for remote sensing image SR and achieved remarkable performances. On one hand, CNN-based networks mainly designed various flexible structures in an end-to-end manner. For instance, LGCNet [39] first designed the CNN-based network to reconstruct remote sensing SR images while HSENet [40] exploited the internal recurrences within images to obtain powerful features. On the other hand, GAN-based networks mainly employed adversarial relationships between the generator and discriminator to improve the subjective effects of SR images. Specifically, EEGAN [41] and CDGANs [42] employed the adversarial learning strategy to reconstruct realistic SR images. Moreover, considering the scale diversity and redundant token representation in challenging remote sensing scenarios, TTST [43] has been designed to select the most critical tokens to make the long-range modeling more effective and proposed the MFL to aggregate more multiscale cues. However, the aforementioned methods mainly exploit complicated structures, where dramatic model complexities could limit the applications in real-world environments. Therefore, the community has drawn attention on lightweight structures to reduce the model complexity. Specifically, CTN [44] designed the contextual transformation layer to decrease the network complexity. Similarly, ReFDN [45] employed the reparameterization and distillation strategies to compose a lightweight network. Meanwhile, addressing the poor visual quality, EDiffSR [46] introduced the DPM for efficient remote sensing image SR, which designed ESAnet to achieve favorable noise prediction performance with low computation costs. However, although existing lightweight structures could reduce the model parameters and accomplish image SR with low computation costs, the employed coarse-grained manners could ignore diverse image details, when the optimal tradeoff between powerful performances of complicated structures, such as coarse-to-fine manners, and low model complexities is hard to be achieved.

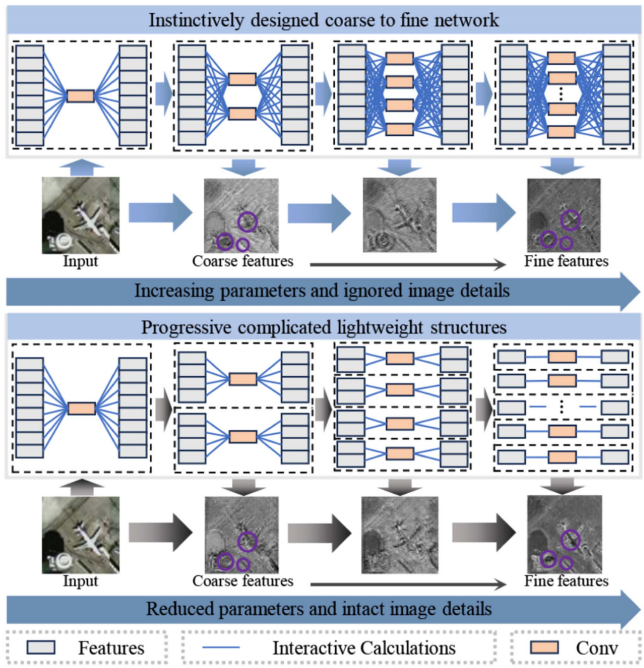


Fig. 1. Brief displaying features from the PC-Block in RCL-Net, which recover diverse image details ignored by the intuitively designed dense coarse-to-fine structures. Meanwhile, RCL-Net retains powerful performances of complicated structures and greatly reduces learnable parameters simultaneously by separating image features and constraining excessive calculations between each subpart.

*To solve the aforementioned problem, we propose a recurrently complicated lightweight network (RCL-Net) to efficiently implement remote sensing image reconstructions through procedures of recurrent fluctuated complexity*. Specifically, we design a serial of progressive complicated blocks (PC-Blocks) to recover image details little by little, where the recurrent usages of structure from simple to complicated help to recover image details in deeper layers that may be ignored by shallow layers. In each PC-Block, three complicated lightweight branches (CL-Branches) presenting increasing complexities in order are successively exploited, *where designed multireceptive field modules (MRF-Modules) are integrated for providing diverse forward paths to be adapted to heterogeneous image details during the training procedure*. Importantly, in each CL-Branch, we separate features into several subparts and *constrain excessive interactive calculations between each subpart to greatly reduce learnable parameters, which reaches the tradeoff between model complexities of traditional powerful structures, such as coarse-to-fine manners, and SR performances.* As illustrated in Fig. 1, compared to intuitively designed coarse-to-fine structures, the special progressive complicated structures, i.e., PC-Block, in RCL-Net help to better recover image details with more simple designs. The major contributions of this article are as follows.

1) To efficiently reconstruct pleasuring remote sensing SR images with diverse details, we propose a novel image SR lightweight network RCL-Net with recurrent usages

of designed progressive complicated structures to recover image details ignored by shallow layers.
2) To obtain powerful features, we delicately design PC-Blocks with three CL-Branch of increasing complexities in order, where the MRF-Module is incorporated to obtain heterogeneous forward propagation paths for recovering diverse details with heterogeneous contents.
3) To reduce learnable parameters, we separate features in the CL-Branch along the channel dimension and constrain excessive interactive calculations between distinct subparts, by which we greatly reduce the learnable parameters and retain powerful performances simultaneously.

## II. Related Work

### A. CNN-Based SR Methods

Due to the excellent performances, deep-learning-based SR methods have gradually become the prevalent approaches to address SISR problems. Specifically, SRCNN [21] first designed the three-layer CNNs to reconstruct SR images, while ESPCN [22] designed subpixel convolution to reduce computation costs with postupscaling. Afterward, RDN [47] and RCAN [24] introduced more complex structures to further improve the model performance. However, although the aforementioned methods have achieved excellent performances, the model complexities caused by complicated structures still limit the practical applications. To solve the problem, DRCN [30] employed recursive strategies to greatly reduce the model parameters. Afterward, to better balance the model performances and complexities, DRRN [31] further introduced recursive-supervision and skip-connection to alleviate the training difficulties of DRCN while still accompanying high computation costs. Although the recursive manner did reduce the model parameters, it also increased the model depth to trade off the model performance. Inspired by the distillation methods, IDN [33] and IMDN [34] leveraged distillation approaches to implement reconstructions with low computation costs, which processed deep features by information distillation to reduce the model computation costs simultaneously. To further improve the efficiency of IMDN, RFDN designed the FDC operation [48] to obtain distilled features and achieved promising results. Further to simplify the network structure of RFDN, BSRN [49] introduced a BSConv to process features with fewer parameters.

### B. SR for Remote Sensing Images

Compared to natural images, remote sensing images usually accompany larger scenes with diverse details [50]. In the early stage, sparse-representation-based methods were the predominant remote sensing image SR solutions with few parameters. For instance, Pan et al. [51] first introduced the sparse representation to reconstruct HR remote sensing images while Hou et al. [52] further designed a global joint dictionary model to acquire the internal relationships among given image patches. Recently, owing to excellent performances, deep learning methods have become mainstream approaches for remote sensing image SR tasks. As the pioneering method, LGCNet [39] first designed the CNN-based SR network to learn multilevel representations while DCM [53] further combined residual units and skip connections to enhance the deep features. Afterward, to improve SR image qualities, TransENet [54] designed the multistage enhancement structures. TTST [43] has been designed to select the most critical tokens to model long-range information more effectively. Besides, a self-supervised degradation-guided adaptive network [55] has been proposed to mitigate the domain gap between simulation and reality while a DDRN [56] has been designed to compensate for the high-frequency details lost during information propagation. However, excessive parameters and computation costs of the aforementioned methods still limit the applications in real-world environments. To reduce model parameters, CTN designed the contextual transformation layer [44] to yield efficient contextual features with few parameters. Correspondingly, ReFDN [45] designed a ReFDB module to extract stronger features, which could greatly decrease the model complexities and enhance the SR image qualities. Nevertheless, although the CTN and ReFDN designed lightweight networks with distillation and reparameterization approaches to reduce the model complexities, the heterogeneity of diverse details in remote sensing images could still be ignored, which might limit the model performances.

### C. Coarse-to-Fine Manners

In order to obtain powerful features for recovering image details of distinct characteristics, several researchers have designed coarse-to-fine structures, where the coarse-level information can be gradually refined into a fine level for enhancing model results. In particular, as for semantic segmentation tasks during the cross-domain segmentation stage, coarse-level categories could be refined into finer categories in a coarse-to-fine manner [57]. For example, CCDA [58] introduced a novel coarse-to-fine knowledge distillation constraint to solve the semantic segmentation tasks. Ma et al. [59] proposed a novel coarse-to-fine domain adaptive semantic segmentation pipeline that seamlessly combines coarse image-level alignment with finer category-level feature distribution regularization. Furthermore, as for the image SR, CFSRCNN [60] designed a feature refinement block to refine the obtained coarse features for reconstructing the HR image in the final stage, which might ignore the intermediate rich features and degrade the model performance. Although coarse-to-fine manners have achieved promising results, the high complexities caused by fine model parts can limit the practical applications.

Grounded on the thorough analyses of image SR models and coarse-to-fine structures, we design a lightweight RCL-Net to recover diverse image details of heterogeneous contents by recurrently exploiting progressively complicated structures. Meanwhile, we separate features and constrain the excessive interactive calculations between distinct subparts to reduce learnable parameters and retain powerful performances simultaneously. As we shall see, plentiful experiments with excellent results would display the superiority of RCL-Net, which even dramatically surpasses powerful dense SR models.

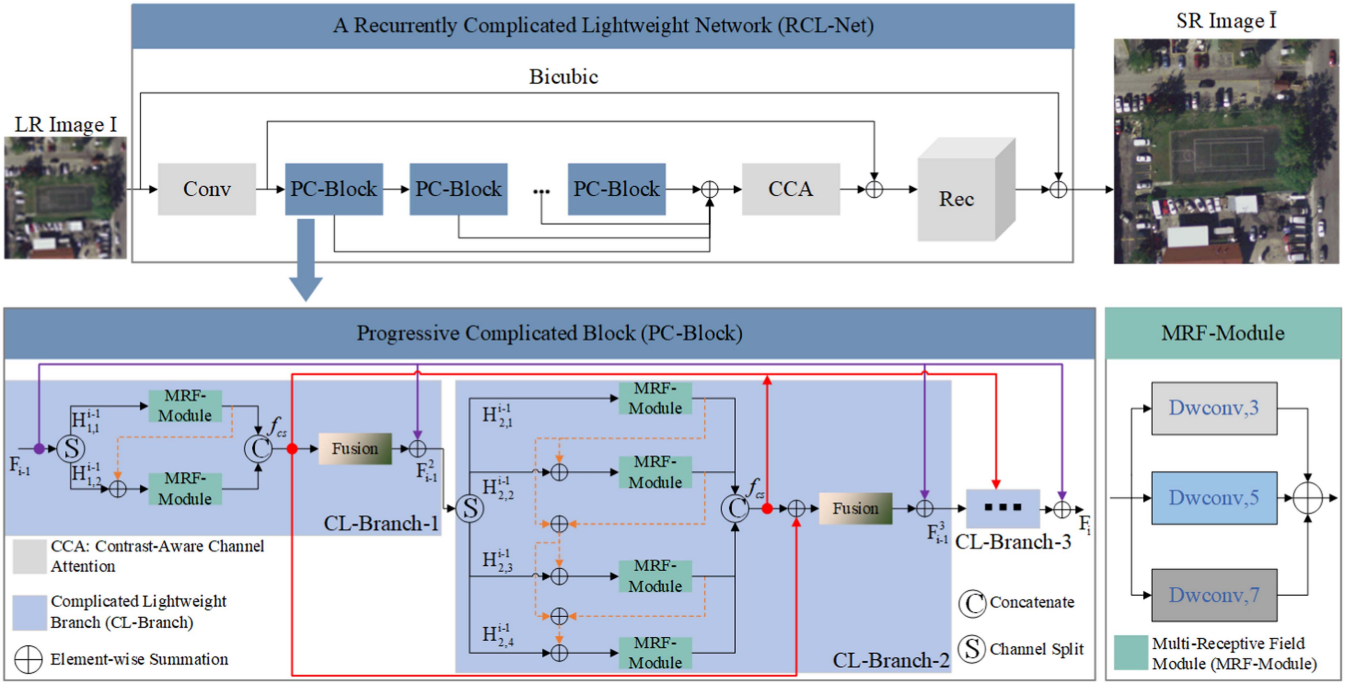


Fig. 2. Overview architecture of RCL-Net. CL-Branch-3 denotes the CL-Branch with eight branches, which is represented in ellipsis. Each designed PC-Block in RCL-Net consists of three CL-Branches of increasing complexities in order. We recurrently exploit PC-Blocks with structures from simple to complicated for recovering diverse image details. Lines of different colors denote skip connections in cascade structures to provide abundant propagation paths. Importantly, we separate features in the CL-Branch to greatly reduce parameters while powerful performances can be retained simultaneously. Particulars of CL-Branch will be detailed in the following sections.

## III. Proposed Method

Although intuitively designed coarse-to-fine structures can capture heterogeneous image details, the excessive calculations can be raised. Motivated by such observations, we build an RCL-Net for the remote sensing image SR to recover image details little by little in recurrent coarse-to-fine manners, where we reform the multibranch module by constraining excessive interaction calculations between distinctly separated feature subparts.

### A. Overview of RCL-Net

To restore remote sensing SR images with diverse image details, we build an RCL-Net with designed serial PC-Blocks, where *the usages of structures recurrently from simple to complicated can recover diverse image details*. During the model training, structures with increasing complexities in each PC-Block of RCL-Net can be automatically adapted to image details of diverse characteristics. The overview structure of RCL-Net is shown in Fig. 2. For simplicity, we denote the input and restored SR image as $I$ and $\bar{I}$ respectively. Given the LR image $I$, we have

$$F_0 = h_{\text{conv}}(I) \tag{1}$$

where $h_{\text{conv}}$ indicates the function induced by a concise $3 \times 3$ convolution, and $F_0$ denotes the obtained low-level features.

Then, to enhance the feature representation abilities, $F_0$ would be further processed through successive CL-Branch modules of distinct complexities, which compose recurrently exploited PC-Blocks with coarse-to-fine designs. Specifically in each PC-Block, the intermediate features among CL-Branch-1 would be separated into two subparts to obtain the coarse features, which would then be integrated and again separated into four subparts to be processed among CL-Branch-2. Finally, the output features of CL-Branch-2 would be separated into eight subparts to be processed among CL-Branch-3. Consequently, each PC-Block is of coarse-to-fine structures, where rich information could be obtained little by little. By recurrently exploiting PC-Blocks, we obtain powerful deep features to recover heterogeneous image details. The backbone could be formulated as

$$\begin{aligned} F_i &= h^i_{\text{PC-Block}}(F_{i-1}) \\ &= h^{i,3}_{\text{CL}}(h^{i,2}_{\text{CL}}(h^{i,1}_{\text{CL}}(F_{i-1}))), i = 1, \ldots, N \\ F_{\text{Rec}} &= h_{\text{CCA}}(F_1 + F_2 + \cdots + F_N) + F_0 \end{aligned} \tag{2}$$

where the $i$th PC-Block is denoted as $h^i_{\text{PC-Block}}$. Meanwhile, $\{h^{i,j}_{\text{CL}}\}^{i=1,2,\ldots,N}_{j=1,2,3}$ denotes successive CL-Branch-$j$ modules in the $i$th PC-Block. Besides $F_i$ denotes the obtained intermediate features translated from the PC-Block, $h_{\text{CCA}}$ denotes the CCA module [34], and $N$ represents the number of PC-Block. All the intermediate features of PC-Block output will be integrated with elementwise addition to obtain powerful features $F_{\text{Rec}}$ through $h_{\text{CCA}}$. Further particulars of CL-Branch will be detailed in the next section.

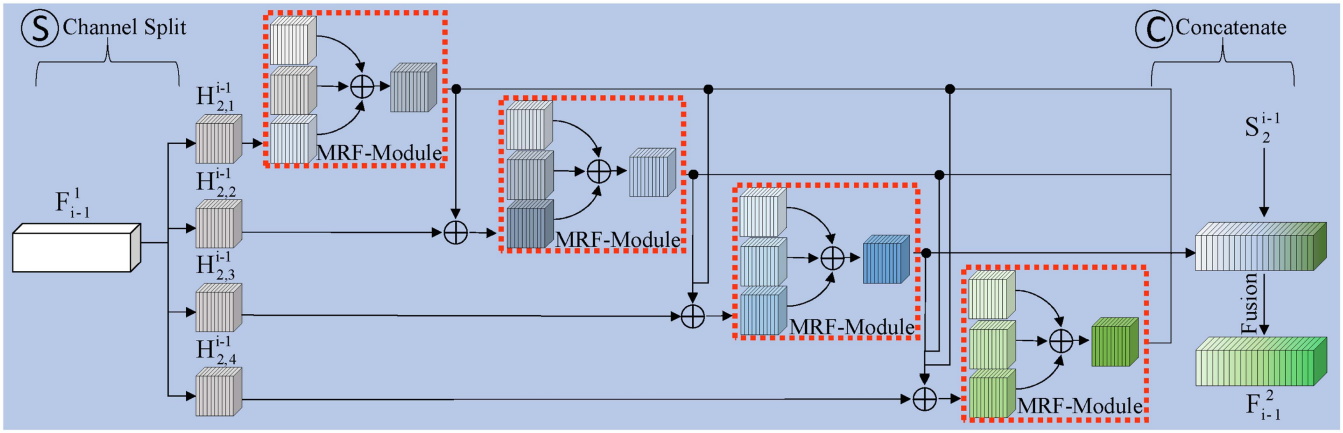


Fig. 3. Overview structures of designed CL-Branch-2 while CL-Branch-1 and CL-Branch-3 are designed in similar frameworks with two and eight branches, respectively. Designed cascade structures in CL-Branch incorporated with simple MRF-Module can provide diverse propagation paths, which can be automatically adapted to heterogeneous image details during the model training. We only allow summations with no extra parameters and eliminate all excessive interactive calculations between distinct subparts to reduce learnable parameters.

Finally, to efficiently reconstruct the realistic SR image $\bar{I}$, $F_{\text{Rec}}$ would be upscaled by the nonparametric subpixel convolution followed by several simple residual blocks (RBs) to further improve the SR image quality in HR space with few computation costs. We have

$$\begin{aligned}\bar{I} &= h_{\text{Rec}}(F_{\text{Rec}}, I) \\ &= h_{\text{RB}}(\cdots(h_{\text{RB}}(h_{u1}(F_{\text{Rec}})))) + h_{u2}(I)\end{aligned} \tag{3}$$

where $h_{u1}$ and $h_{u2}$ denote the subpixel convolution and bicubic interpolation, respectively. Besides, $h_{\text{RB}}$ denotes the residual block. We employ the simple interpolation operation to directly recover the coarse SR image, which makes our model focus on learning the residual features.

For a fair comparison, we also employ $L_1$ loss to optimize our model by minimizing it following previous works [40], [54]. Specifically, for the given paired HR and LR images with $M$ pairs, which are referred by $\{I^{(q)}, I^{(q)}_{\text{HR}}\}_{q=1}^{M}$, the loss function can be obtained as

$$L_1(\delta) = \frac{1}{M}\sum_{q=1}^{M}\left\| I^{(q)}_{\text{HR}} - \text{RCL-Net}(I^{(q)})\right\|_1 \tag{4}$$

where $\delta$ denotes the corresponding learnable parameter sets of our proposed network RCL-Net. Meanwhile, RCL-Net$(I^{(q)})$ is the $q$th SR image $\bar{I}^{(q)}$ and $M$ is the number of training images. And then, $I$ and $I_{\text{HR}}$ refer to the given LR image and corresponding HR reference, respectively.

### B. Complicated Lightweight Branch

To obtain powerful features, we carefully design the CL-Branch, which incorporates cascade designs with simple MRF-Module to provide various propagation paths for recovering heterogeneous image contents. *Suitable calculation processes provided by distinct propagation paths could be automatically adapted to diverse image details during the model training.*

As illustrated in Fig. 3, deep features in each CL-Branch would be separated into several subparts and compose cascade structures, which would then be processed with the MRF-Module. Specifically, outputs of all previous branches are regarded as the inputs of current branch followed by MRF-Module structures, *which could generate heterogeneous forward propagation paths for diverse image details.* Formally, given the input features $F_{i-1}$ of CL-Branch-1 in $i$th PC-Block for instance, we first averagely separate $F_{i-1}$ into two subparts, which would be processed with the counterpart MRF-Module. Considering in various sizes of objects among large coverage areas in remote sensing images, we employ three depthwise separable convolutions (DWconv) with distinct convolutional kernel sizes to capture diverse image details in MRF-Module. For excavating relationships between each subpart, we introduce branch-by-branch addition operations with no parameters. Afterward, separated deep features derived from the MRF-Module in each branch would be fused with concatenation and channel shuffle operations, where a general fusion module is introduced for the feature integration. Meanwhile, we introduce local residual learning to stabilize the network training. Afterward, similar frameworks are employed to process the separated four and eight subparts among CL-Branch-2 and CL-Branch-3, respectively. In summary, the CL-Branch-$j$ in $i$th PC-Block $h^{i,j}_{\text{CL}}$ could be formulated as

$$\begin{aligned}&H^{i-1}_{j,1}, H^{i-1}_{j,2}, \ldots, H^{i-1}_{j,2^j} = h_{\text{split}}(F^j_{i-1}) \\ &D^{i-1}_j = f_{\text{cat}}(\{m^{i-1}_{j,k}(H^{i-1}_{j,k} + \cdots + m^{i-1}_{j,1}(H^{i-1}_{j,1}))\}_{1\le k\le 2^j}) \\ &S^{i-1}_j = f_{\text{cs}}(D^{i-1}_j) \\ &F^{j+1}_{i-1} = f^{i-1}_{\text{Fus}}(S^{i-1}_1 + S^{i-1}_2 + \cdots + S^{i-1}_j) + F^j_{i-1}\end{aligned} \tag{5}$$

where index $\{j = 1, 2, 3\}$ indicate the $j$th CL-Branch-$j$ and index $i$ indicates the $i$th PC-Block. Meanwhile, $F^j_{i-1}$ denotes the input of CL-Branch-$j$ (especially $F^1_{i-1} = F_{i-1}$) as illustrated in Figs. 2 and 3. Besides, index $\{k = 2, \ldots, 2^j\}$ indicate distinct

branches in CL-Branch-$j$ corresponding to each feature subpart while $h_{\text{split}}$ denotes the separating operation to generate several subparts $H_{j,1}^{i-1}, H_{j,2}^{i-1}, \ldots, H_{j,2^j}^{i-1}$. Then, $m_{j,k}^{i-1}$ represents MRF-Module, $f_{cs}$ *denotes the channel shuffle operation to enhance subpart relationships without extra parameters* and $f_{\text{cat}}$ denotes the concatenation operation along channel dimensions. Meanwhile, $f_{\text{Fus}}^{i-1}$ refers to a general fusion module and finally $F_{i-1}^{j+1}$ denotes powerful fused features generated by CL-Branch-$j$, which would be regarded as the input of following CL-Branch.

Specifically, we employ the efficient PATM [61] as the general fusion module $f_{\text{Fus}}^{i-1}$ to assist feature fusions. Moreover, we set the largest convolution kernel size in the MRF-Module as $7 \times 7$ to make the half-side length of receptive fields boosted by 42 in each PC-Block for covering almost the whole spatial dimension of features (default size of $48 \times 48$). Furthermore, as shown in Fig. 2, we design three DWconv operations with $3 \times 3$, $5 \times 5$, and $7 \times 7$ kernel sizes in the MRF-Module to obtain various receptive fields and heterogeneous forward propagation paths, where the smallest kernel size is set as $3 \times 3$ to avoid the constraint within local information. With the help of multiscale settings in MRF-Module, the designed procedures of recurrent fluctuated complexity consisting of CL-Branch can assign more various receptive fields to different feature subparts and thus improve the model performance.

### C. Reconstruction Processes

Direct upscale of LR features to obtain SR images would ignore image details and degrade model performances [60]. Inspired by CFSRCNN [60], we propose the feature refinement module to further improve remote sensing image qualities with simpler structures compared to CFSRCNN. Specifically, we first employ a subpixel convolution to upscale LR features. Then, we adopt four simpler residual blocks (only a $3 \times 3$ convolution and a GELU [62] activation function) and a plain $3 \times 3$ convolution to refine HR features, which could improve the SR image quality with few computation costs. Given arbitrary deep feature $F$, the residual block $h_{\text{RB}}$ could be formulated as

$$h_{\text{RB}}(F) = h_{\text{gelu}}(h_{\text{conv}}(F)) + F \tag{6}$$

where $h_{\text{RB}}$ denotes the residual block in (3). Besides, $h_{\text{gelu}}$ and $h_{\text{conv}}$ denote the GELU function and $3 \times 3$ convolution.

### D. Analyses of Parameter Quantities

To illustrate the superiority of separation operations in CL-Branch, we provide detailed analyses of learnable parameters taking CL-Branch-2 as the instance, whose input and output contain the same channel number $C$. *With the feature separating in CL-Branch, we constrain the excessive interactive calculations between distinct subparts* $H_{j,k}^{i-1}$ *with distinct index* $k$ *except for elementwise summation, which greatly reduces learnable parameters compared to instinctively designed coarse-to-fine structures.* Specifically, we list the following four types of parallel convolution structures with MRF-Module:

I) The structure with standard convolutions in MRF-Module to generate deep features with $C$ channels in all branches, where output features of each branch would be directly added to obtain the module output.
II) The structure with standard convolutions in MRF-Module to generate deep features with $0.25C$ channels in each branch, where output features of each branch would be concatenated to maintain the equivalent representation abilities.
III) The standard convolutions in type-(II) structure are replaced with counterpart DWConv to generate deep features, while other settings are the same as type-(II) structure.
IV) The input of type-(III) structure is directly separated into four subparts to generate deep features for four branches, respectively, while other settings are the same as type-(III) structure.

TABLE I
ANALYSIS OF NETWORK PARAMETERS

| Type | Parameters | $C$=32 | $C$=64 |
|---|---|---|---|
| I | $332C^2 + 12C$ | 340.4K | 1.4M |
| II | $83C^2 + 3C$ | 85.1K | 340.2K |
| III | $3C^2 + 335C + 12$ | 13.8K | 33.7K |
| IV(ours) | $0.75C^2 + 86C + 12$ | 3.5K | 8.6K |

The channel dimensions of input and output are C. Type-(IV) represents the reformed multi-branch structures with separating operations proposed in this article.

As shown in Table I, our method greatly reduces the learnable parameters. For example, when $C = 32$, our reformed multi-branch structure with MRF-Module only has 1%, 4%, and 26% parameters compared to the type-(I), type-(II), and type-(III) structures; similarly, when $C = 64$, our reformed multibranch structure with MRF-Module only has 0.6%, 2.5%, and 25.4% parameters compared to the type-(I), type-(II), and type-(III) structures.

## IV. EXPERIMENTS

In this section, we will provide abundant experiments to make in-depth studies of our model and illustrate the superiority of our conscientiously designed module. Specifically, we first describe the relevant experimental datasets and then exhibit the effectiveness of the proposed RCL-Net. Afterward, we compare our results with several classical representative state-of-the-art (SOTA) methods together with advanced powerful large SR models to demonstrate the superiority of RCL-Net. Finally, we present several ablation studies to give throughout analyses.

### A. Experimental Datasets and Settings

For fair comparisons, following previous works [39], [53], [54] for remote sensing image SR, we employ two popular benchmark datasets: 1) UCMerced [63] and 2) AID [64], to evaluate our model performance. Specifically, the UCMerced dataset totally contains 21 classes of remote sensing scenes, such as river, runway, etc. On the other hand, the AID dataset with 10 000 images totally contains 30 classes of remote sensing scenes, such as school, square, stadium, viaduct, etc. Specifically, following the work in [54], we employ the same strategy to construct and split the train, test, and validation sets from UCMerced and AID

datasets. Meanwhile, we employ the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [65] as metrics to evaluate the model performances.

### B. Implementation Details

As for the remote sensing image SR, in this article, we draw attentions to improving SR image qualities with $\times 2$, $\times 3$, and $\times 4$ upscaling factors. In the training phase, following the work in [54], the input LR image of our network will be randomly cropped into $48 \times 48$ image patches, and corresponding reference patches from counterpart HR images will be obtained. Besides, to further improve the model robustness, we introduce data augmentations for training samples, such as random rotation with $90°, 180°, 270°$, and horizontal flipping. Additionally, the number of PC-Block is set to 10 and the global residual learning is introduced with bicubic interpolation operation to stabilize our network training.

In order to efficiently optimize our network, we adopt Adam optimizer [66] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$ to optimize our proposed RCL-Net. Specifically, we train the RCL-Net with 2000 epochs and the initial learning rate $1 \times 10^{-3}$ will be multiplied by 0.5 at the 1500th epoch. Besides, the minibatch size is set to 16 and our earnestly designed network is implemented by PyTorch without any postprocessing. In addition, all the experiments are accomplished on an NVIDIA GeForce RTX 3090.

### C. Effectiveness Analyses

In this section, we directly provide visualizations of deep features to analyze the effectiveness of our proposed RCL-Net. Specifically, grounded on the UCMerced test dataset with $\times 4$ upscaling factor, we visualize features to study the recurrent usages of coarse-to-fine structures. Specifically, the first two rows of Fig. 4 display features of a given image from the first and third PC-Blocks, respectively. As we can see, output features from the third PC-Block recover more image details, such as shadows, compared to the ones from first PC-Block, which illustrates the merits of the recurrent usages of coarse-to-fine structures. Moreover, the last two rows show the features of a given image from the first PC-Block, CCA module, and reconstruction procedures. As we can see, with all features derived from recurrent coarse-to-fine structures integrated in the CCA module and reconstruction procedures, image details can be distilled. In summary, visualized features exactly illustrate the effectiveness of our designed RCL-Net with structures recurrently from simple to complicated.

### D. Comparison With SOTA Methods

In order to demonstrate the effectiveness of RCL-Net, we compare our method with several representative lightweight models, such as VDSR [23], LGCNet [39] together with CTN [44], and SOTA lightweight networks, such as DCM [53] and ReFDN [45]. Further to illustrate the superiority of our method, we also compare our results with the advanced powerful large models, such as TransENet [54], HSENet [40], SwinIR [26], and TTST [43]. All the algorithms are trained on the same dataset settings.

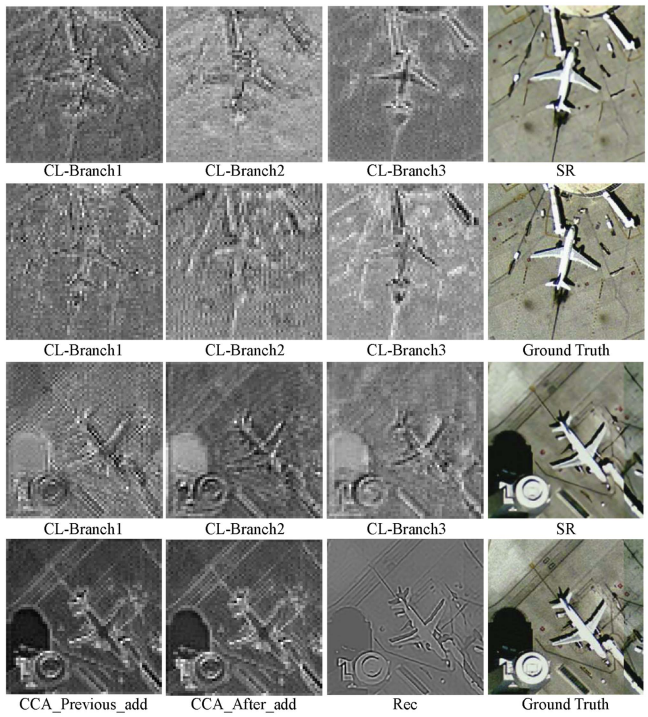


Fig. 4. Visualization of intermediate features in the PC-Block. Among the first three columns, the first two rows display features in the CL-Branch from first and third PC-Blocks, respectively. Meanwhile, the last two rows display features in the first PC-Block, CCA module, and reconstruction procedures. The last column displays recovered SR images and ground-truth images. We can conclude that RCL-Net is able to recover image details through serial PC-Block structures, which are recurrently from simple to complicated.

*1) Results on UCMerced Test Dataset:* Table II displays the average PSNR and SSIM values of recovered SR images on $\times 2$, $\times 3$, and $\times 4$ upscale factors of the aforementioned methods. As we can see from Table II, our method performs better than not only the classical methods but also SOTA lightweight networks. Specifically, when the upscale factor is set to 2, designed RCL-Net outperforms SOTA lightweight networks DCM and ReFDN by 0.76 and 0.35 in PSNR, respectively, while our results are also superior to DCM and ReFDN by 0.0066 and 0.0036 in SSIM. Meanwhile, RCL-Net obviously outperforms classical lightweight networks, such as LGCNet and CTN, with at least 0.82 in PSNR and 0.0085 in SSIM, which displays the superiority of the designed RCL-Net. Especially compared to DCM, our proposed RCL-Net only consumes 57% parameters and 34% FLOPs, attributing to feature separations in the CL-Branch. Furthermore, to illustrate the effectiveness of RCL-Net, we compare the results with advanced powerful large models. It could be found that RCL-Net outperforms TransENet and HSENet by 0.38 and 0.19 in PSNR while the improvements are 0.0039 and 0.0013 in SSIM. It is worth mentioning that RCL-Net only consumes 3% and 23% parameters compared to TransENet and HSENet while the FLOPs ratios are 57% and 23%, respectively. Meanwhile, our proposed RCL-Net achieves comparative results compared to powerful large models, such

TABLE II
AVERAGE PSNR (DB) AND SSIM OVER THE UCMERCED TEST DATASET

| Scale | Bicubic PSNR/SSIM | DCM [53] PSNR/SSIM | LGCNet [39] PSNR/SSIM | CTN [44] PSNR/SSIM | ReFDN [45] PSNR/SSIM | HSENet [40] PSNR/SSIM | TransENet [54] PSNR/SSIM | SwinIR [26] PSNR/SSIM | TTST [43] PSNR/SSIM | Ours PSNR/SSIM |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 30.76/0.8789 | 33.65/0.9274 | 33.48/0.9235 | 33.59/0.9255 | 34.06/0.9304 | 34.22/0.9327 | 34.03/0.9301 | 34.24/0.9314 | 34.50/0.9343 | 34.41/0.9340 |
| 3 | 27.46/0.7631 | 29.52/0.8394 | 29.28/0.8238 | 29.44/0.8319 | 29.85/0.8405 | 30.00/0.8420 | 29.92/0.8408 | 29.80/0.8382 | 30.21/0.8465 | 30.15/0.8470 |
| 4 | 25.65/0.6725 | 27.22/0.7528 | 27.02/0.7333 | 27.41/0.7512 | 27.68/0.7596 | 27.73/0.7623 | 27.40/0.7559 | 27.50/0.7602 | 27.77/0.7630 | 27.79/0.7656 |
| Params | - | 2177K | 767K | 413K | 918K | 5433K | 37459K | 11900K | 18367K | 1243K |
| FLOPs | - | 7.3G | 1.7G | 1.0G | 3.5G | 10.8G | 4.4G | 7.1G | 18.4G | 2.5G |

The FLOPs is calculated with input 48 × 48 under ×4 factor.

TABLE III
AVERAGE PSNR (DB) OF EACH CLASS FOR ×3 UPSCALING FACTOR ON THE UCMERCED TEST DATASET, WHICH TOTALLY CONTAINS 21 CLASSES[1]

| Class | Bicubic | HSENet [40] | CTN [44] | LGCNet [39] | DCM [53] | TransENet [54] | SwinIR [26] | TTST [43] | Ours |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 26.86 | 27.64 | 31.79 | 27.66 | 29.06 | 28.02 | 27.33 | 27.55 | 28.10 |
| 2 | 26.71 | 30.09 | 28.22 | 29.12 | 30.77 | 29.94 | 30.04 | 30.29 | 30.22 |
| 3 | 33.33 | 35.05 | 29.37 | 34.72 | 33.76 | 35.04 | 34.96 | 35.43 | 35.26 |
| 4 | 36.14 | 37.69 | 34.73 | 37.37 | 36.38 | 37.53 | 37.22 | 37.61 | 37.69 |
| 5 | 25.09 | 28.95 | 37.39 | 27.81 | 28.51 | 28.81 | 28.70 | 29.32 | 29.21 |
| 6 | 25.21 | 26.70 | 28.01 | 26.39 | 26.81 | 26.69 | 26.47 | 26.78 | 26.78 |
| 7 | 25.76 | 29.24 | 26.42 | 28.25 | 28.79 | 29.11 | 28.92 | 29.47 | 29.42 |
| 8 | 27.53 | 28.59 | 28.41 | 28.44 | 28.16 | 28.59 | 28.23 | 28.55 | 28.62 |
| 9 | 27.36 | 30.63 | 28.43 | 29.52 | 30.45 | 30.38 | 30.48 | 31.01 | 30.80 |
| 10 | 35.21 | 36.62 | 29.67 | 36.51 | 34.43 | 36.68 | 36.63 | 36.93 | 36.77 |
| 11 | 21.25 | 24.88 | 36.24 | 23.63 | 26.55 | 24.72 | 24.79 | 25.12 | 25.07 |
| 12 | 26.48 | 29.21 | 23.99 | 28.29 | 29.28 | 29.03 | 29.09 | 29.49 | 29.39 |
| 13 | 25.68 | 28.55 | 28.42 | 27.76 | 27.21 | 28.47 | 28.33 | 28.73 | 28.74 |
| 14 | 22.25 | 25.70 | 27.86 | 24.59 | 26.05 | 25.64 | 25.56 | 26.05 | 25.94 |
| 15 | 24.59 | 28.22 | 24.99 | 26.58 | 27.77 | 27.83 | 28.03 | 28.30 | 28.34 |
| 16 | 21.75 | 24.66 | 27.48 | 23.69 | 24.95 | 24.45 | 24.81 | 25.23 | 24.84 |
| 17 | 28.12 | 29.22 | 23.63 | 29.12 | 28.89 | 29.25 | 28.90 | 29.42 | 29.35 |
| 18 | 29.30 | 31.15 | 29.03 | 31.15 | 32.53 | 31.25 | 31.25 | 31.59 | 31.24 |
| 19 | 28.34 | 31.64 | 30.68 | 30.53 | 29.81 | 31.57 | 31.33 | 31.72 | 31.73 |
| 20 | 29.97 | 32.95 | 31.18 | 32.17 | 29.02 | 32.71 | 32.55 | 32.86 | 32.98 |
| 21 | 29.75 | 32.71 | 32.43 | 31.58 | 30.76 | 32.51 | 32.19 | 32.97 | 32.74 |
| Avg | 27.46 | 30.00 | 29.44 | 29.28 | 29.52 | 29.92 | 29.80 | 30.21 | 30.15 |

All these 21 classes of UCMerced dataset: 1-Agricultural, 2-Airplane, 3-Baseballdiamond, 4-Beach, 5-Buildings, 6-Chaparral, 7-Denseresidential, 8-Forest, 9-Freeway, 10-Golfcourse, 11-Harbor, 12-Intersection, 13-Mediumresidential, 14-Mobilehomepark, 15-Overpass, 16-Parkinglot, 17-River, 18-Runway, 19-Sparseresidential, 20-Storagetanks 21-Tenniscourt.

as SwinIR and TTST, with less than 12% parameters. Similar improvements raised by our designed RCL-Net could also be seen when the upscale factor is set to 3 and 4, where we can conclude that RCL-Net outperforms both classical and SOTA lightweight models. *More importantly, the RCL-Net is the only lightweight network that surpasses advanced powerful large models TransENet and HSENet with all upscale settings, which demonstrates that RCL-Net can retain powerful performances of complicated structures and reduce learnable parameters simultaneously*.

Following the work in [54], we further provide detailed results about each class of the UCMerecd dataset with ×3 upsampling factor, where 21 classes are totally involved. As shown in Table III, the PSNR value of our network greatly exceeds other methods except for SwinIR and TTST among 14 scenes, such as beach, forest, overpass, river, etc. Meanwhile, our average PSNR value for all classes is beyond advanced large model TransENet while RCL-Net still obtains comparative results compared to SwinIR and TTST with less than 12% parameters. The abundant quantitative results exactly demonstrate the superiority of our model. Moreover, for the visual comparison, we list SR images obtained by the current best methods in Fig. 5, where the superiority of RCL-Net could be found. Specifically, the first row in Fig. 5 demonstrates that the results of RCL-Net are with less blur and clear textures compared to large models TransENet and HSENet. Additionally, the last row in Fig. 5 also illustrates that the proposed RCL-Net could obtain recovered results with fewer jaggies and accurately reconstruct the object edges. Meanwhile, RCL-Net achieves comparative results with less than 7% parameters compared to TTST.

*2) Results on AID Test Dataset:* To demonstrate the robustness of proposed RCL-Net, we perform abundant extra experiments on the AID dataset, where we provide quantitative and qualitative comparisons with other methods. As illustrated in Table IV, when the upscale factor is set to 4, designed RCL-Net outperforms SOTA lightweight networks DCM and ReFDN by 0.29 and 0.26 in PSNR, respectively, while our results are also superior to DCM and ReFDN by 0.0096 and 0.0085 in SSIM. Meanwhile, RCL-Net obviously outperforms classical lightweight networks, such as VDSR and CTN, at least 0.28 in PSNR and 0.0092 in SSIM. For the validation, we also compare the results with the advanced powerful large model TransENet on the AID dataset. It could be found that RCL-Net outperforms TransENet by 0.08 in PSNR and 0.0011 in SSIM with only 3% parameters and 57% FLOPs. Meanwhile, our proposed RCL-Net achieves comparative results compared to powerful large models, such as SwinIR and TTST, with less than 12% parameters. Similar improvements raised by our designed RCL-Net could

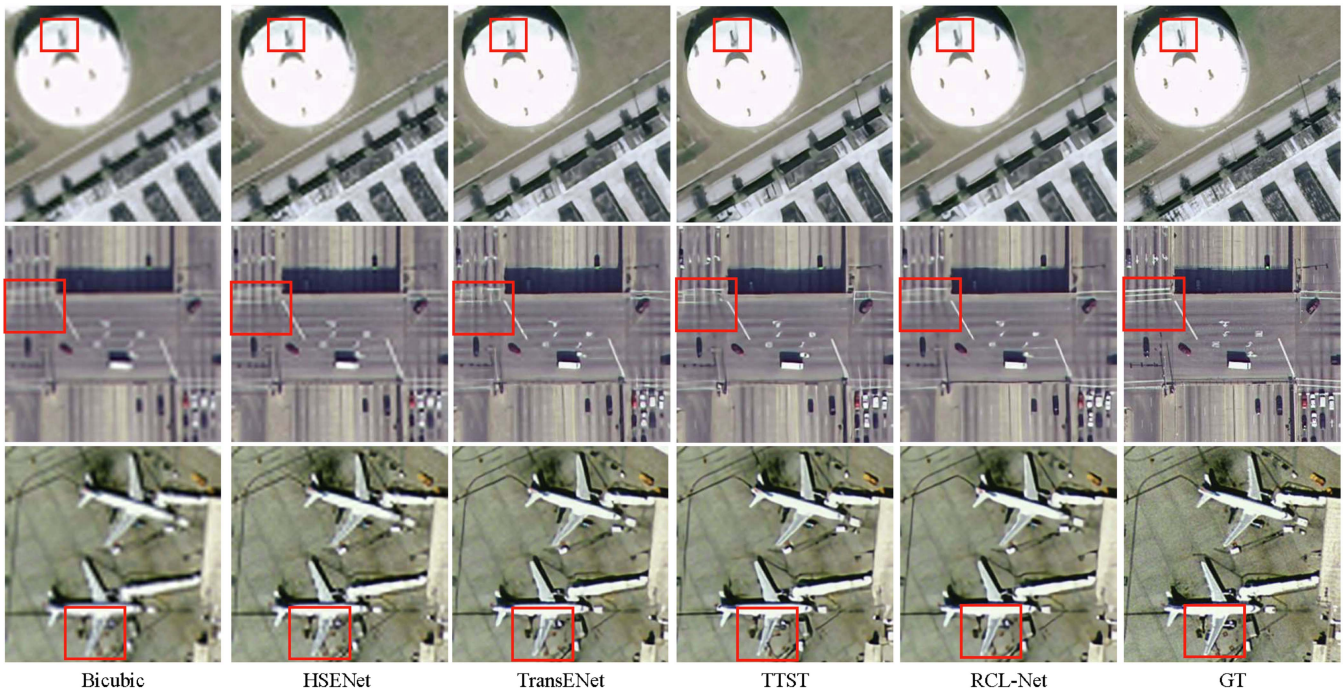


Fig. 5. Visual comparisons of RCL-Net with some representative methods on the UCMerced test dataset for ×4 upscaling factor.

TABLE IV
AVERAGE PSNR (DB) AND SSIM OVER THE AID TEST DATASET

| Scale | Bicubic PSNR/SSIM | CTN [44] PSNR/SSIM | ReFDN [45] PSNR/SSIM | LGCNet [39] PSNR/SSIM | DCM [53] PSNR/SSIM | VDSR [23] PSNR/SSIM | TransENet [54] PSNR/SSIM | SwinIR [26] PSNR/SSIM | TTST [43] PSNR/SSIM | Ours PSNR/SSIM |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 32.39/0.8906 | 35.22/0.9369 | 35.24/0.9368 | 34.80/0.9320 | 35.21/0.9366 | 35.05/0.9346 | 35.28/0.9374 | 35.68/0.9402 | 35.82/0.9414 | 35.60/0.9405 |
| 3 | 29.08/0.7863 | 31.31/0.8561 | 31.35/0.8566 | 30.73/0.8417 | 31.31/0.8561 | 31.15/0.8522 | 31.45/0.8595 | 35.68/0.8637 | 31.71/0.8650 | 31.77/0.8624 |
| 4 | 27.30/0.7036 | 29.18/0.7828 | 29.20/0.7835 | 28.61/0.7626 | 29.17/0.7824 | 28.99/0.7753 | 29.38/0.7909 | 29.53/0.7940 | 29.57/0.7964 | 29.46/0.7920 |
| Params | - | 413K | 918K | 767K | 2177K | 666K | 37459K | 11900K | 18367K | 1243K |
| FLOPs | - | 1.0G | 3.5G | 1.7G | 7.3G | 24.6G | 4.4G | 7.1G | 18.4G | 2.5G |

The FLOPs is calculated with input 48 × 48 under ×4 factor.

be seen when the upscale factor is set to 2 and 3, where we can conclude that RCL-Net outperforms both classical and SOTA lightweight models. *More importantly, the proposed RCL-Net is the only lightweight network that surpasses advanced powerful large model TransENet with all upscale settings*. Following the work in [54], to further demonstrate the effectiveness of our proposed method, we also provide detailed results about each class of the AID dataset with ×4 upscaling, where 30 classes are totally involved. As shown in Table V, the PSNR value of our network obviously exceeds other methods except for SwinIR and TTST among all scenes while RCL-Net still achieves comparative results compared to SwinIR and TTST with less than 12% parameters. *The abundant quantitative results exactly demonstrate the superiority of our model, which reaches the tradeoff between powerful performances and low model complexities.*

In addition to quantitative comparisons on the AID dataset, the qualitative visual results are also provided for subjective comparisons. As shown in Fig. 6, RCL-Net could better recover the remote sensing image textures. For instance, as for the second row in Fig. 6, the bicubic interpolation operation, VDSR, and TransENet networks hardly recover the precise roof textures, which accompany terrible distortions and blurs. However, our method could exactly restore the roof textures with few distortions and insignificant blurs, which clearly implies the superiority of RCL-Net. Similarly, as for the last row in Fig. 6, the proposed RCL-Net could accurately reconstruct the boarding bridge while TransENet, VDSR, and the bicubic interpolation operation fail to generate accurate structures. Meanwhile, RCL-Net achieves comparative results with less than 7% parameters compared to TTST. In sum, the excellent quantitative and qualitative results on the AID dataset properly demonstrate the effectiveness and robustness of our method.

### E. Ablation Study

In this section, to illustrate the effectiveness of our proposed method, we would make plentiful ablation experiments grounded on the UCMerced dataset with ×4 upscaling factor. Specifically, we first analyze the influence of PC-Block numbers and provide the corresponding qualitative and quantitative comparisons. Besides, we seriously study the designs of the proposed PC-Block and demonstrate the superiority with subjective and objective results.

TABLE V
AVERAGE PSNR (DB) OF EACH CLASS FOR ×4 UPSCALING FACTOR ON THE AID TEST DATASET, WHICH TOTALLY CONTAINS 30 CLASSES[2]

| Class | Bicubic | SRCNN [21] | LGCNet [39] | VDSR [23] | DCM [53] | TransENet [54] | SwinIR [26] | TTST [43] | Ours |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 27.03 | 28.17 | 28.39 | 28.82 | 28.99 | 29.23 | 29.37 | 29.38 | 29.29 |
| 2 | 34.88 | 35.63 | 35.78 | 35.98 | 36.17 | 36.20 | 36.41 | 36.48 | 36.27 |
| 3 | 29.06 | 30.51 | 30.75 | 31.18 | 31.36 | 31.59 | 31.71 | 31.93 | 31.62 |
| 4 | 31.07 | 31.92 | 32.08 | 32.29 | 32.45 | 32.55 | 32.71 | 32.80 | 32.57 |
| 5 | 28.98 | 30.41 | 30.67 | 31.19 | 31.39 | 31.63 | 31.80 | 31.86 | 31.69 |
| 6 | 25.26 | 26.59 | 26.92 | 27.48 | 27.72 | 28.03 | 28.23 | 28.29 | 28.12 |
| 7 | 22.15 | 23.41 | 23.68 | 24.12 | 24.29 | 24.51 | 24.63 | 24.67 | 24.59 |
| 8 | 25.83 | 27.05 | 27.24 | 27.62 | 27.78 | 27.97 | 28.11 | 28.12 | 28.07 |
| 9 | 23.05 | 24.13 | 24.33 | 24.70 | 24.87 | 25.13 | 25.25 | 25.26 | 25.22 |
| 10 | 38.49 | 38.84 | 39.06 | 39.13 | 39.27 | 39.31 | 39.50 | 39.58 | 39.34 |
| 11 | 32.30 | 33.48 | 33.77 | 34.20 | 34.42 | 34.58 | 34.74 | 34.79 | 34.66 |
| 12 | 27.39 | 28.15 | 28.20 | 28.36 | 28.47 | 28.56 | 28.62 | 28.62 | 28.59 |
| 13 | 24.75 | 26.00 | 26.24 | 26.72 | 26.92 | 27.21 | 27.38 | 27.38 | 27.30 |
| 14 | 32.06 | 32.57 | 32.65 | 32.77 | 32.88 | 32.94 | 33.02 | 33.15 | 32.97 |
| 15 | 26.09 | 27.37 | 27.63 | 28.06 | 28.25 | 28.45 | 28.63 | 28.65 | 28.57 |
| 16 | 28.04 | 28.90 | 28.97 | 29.11 | 29.18 | 29.28 | 29.34 | 29.35 | 29.27 |
| 17 | 26.23 | 27.25 | 27.37 | 27.69 | 27.82 | 28.01 | 28.11 | 28.17 | 28.07 |
| 18 | 22.33 | 24.01 | 24.40 | 25.21 | 25.74 | 26.40 | 26.71 | 26.75 | 26.62 |
| 19 | 27.27 | 28.72 | 29.04 | 29.62 | 29.92 | 30.30 | 30.53 | 30.67 | 30.40 |
| 20 | 28.94 | 29.85 | 30.00 | 30.26 | 30.39 | 30.53 | 30.61 | 30.81 | 30.56 |
| 21 | 24.69 | 25.82 | 26.02 | 26.43 | 26.62 | 26.91 | 27.07 | 27.00 | 26.98 |
| 22 | 26.31 | 27.55 | 27.76 | 28.19 | 28.38 | 28.61 | 28.75 | 28.76 | 28.67 |
| 23 | 25.98 | 27.12 | 27.32 | 27.71 | 27.88 | 28.08 | 28.24 | 28.18 | 28.17 |
| 24 | 29.61 | 30.48 | 30.60 | 30.82 | 30.91 | 31.00 | 31.08 | 31.12 | 31.02 |
| 25 | 24.91 | 26.13 | 26.34 | 26.78 | 26.94 | 27.22 | 27.38 | 27.38 | 27.31 |
| 26 | 25.41 | 26.16 | 26.27 | 26.46 | 26.53 | 26.63 | 26.70 | 26.70 | 26.68 |
| 27 | 26.75 | 28.13 | 28.39 | 28.91 | 29.13 | 29.39 | 29.56 | 29.60 | 29.47 |
| 28 | 24.81 | 26.10 | 26.37 | 26.88 | 27.10 | 27.41 | 27.60 | 27.58 | 27.50 |
| 29 | 24.18 | 25.27 | 25.48 | 25.86 | 26.00 | 26.20 | 26.31 | 26.26 | 26.27 |
| 30 | 25.86 | 27.03 | 27.26 | 27.74 | 27.93 | 28.21 | 28.37 | 28.42 | 28.28 |
| Avg | 27.30 | 28.40 | 28.61 | 28.99 | 29.17 | 29.38 | 29.53 | 29.57 | 29.46 |

All these 30 classes of AID dataset: 1-Airport, 2-Bareland, 3-Baseballdiamond, 4-Beach, 5-Bridge, 6-Center, 7-Church, 8-Commercial, 9-Denseresidential, 10-Desert, 11-Farmland, 12-Forest, 13-Industrial, 14-Meadow, 15-Mediumresidential, 16-Mountain, 17-Park, 18-Parking, 19-Playground, 20-Pond, 21-Port, 22-Railwaystation, 23-Resort, 24-River, 25-School, 26-Sparseresidential, 27-Square, 28-Stadium, 29-Storagetanks, 30-Viaduct.

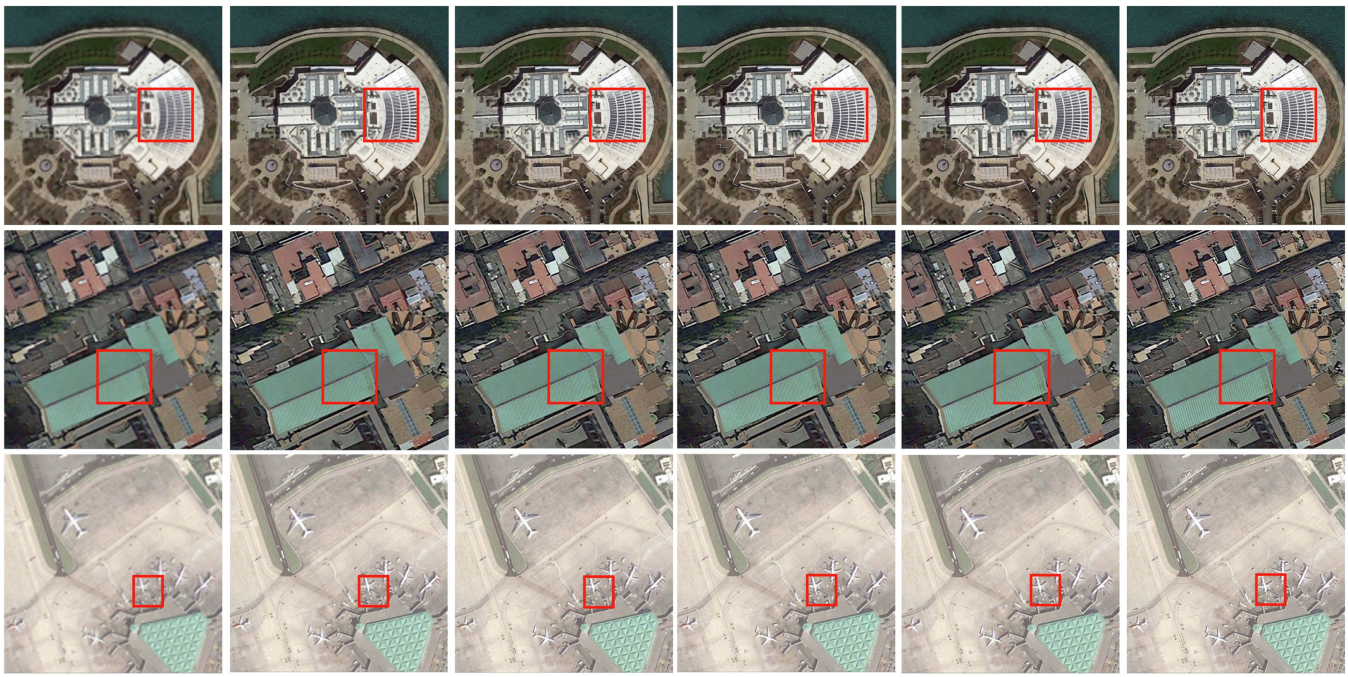


Fig. 6. Visual comparisons of RCL-Net with some representative methods on the AID dataset for ×4 upscaling factor.

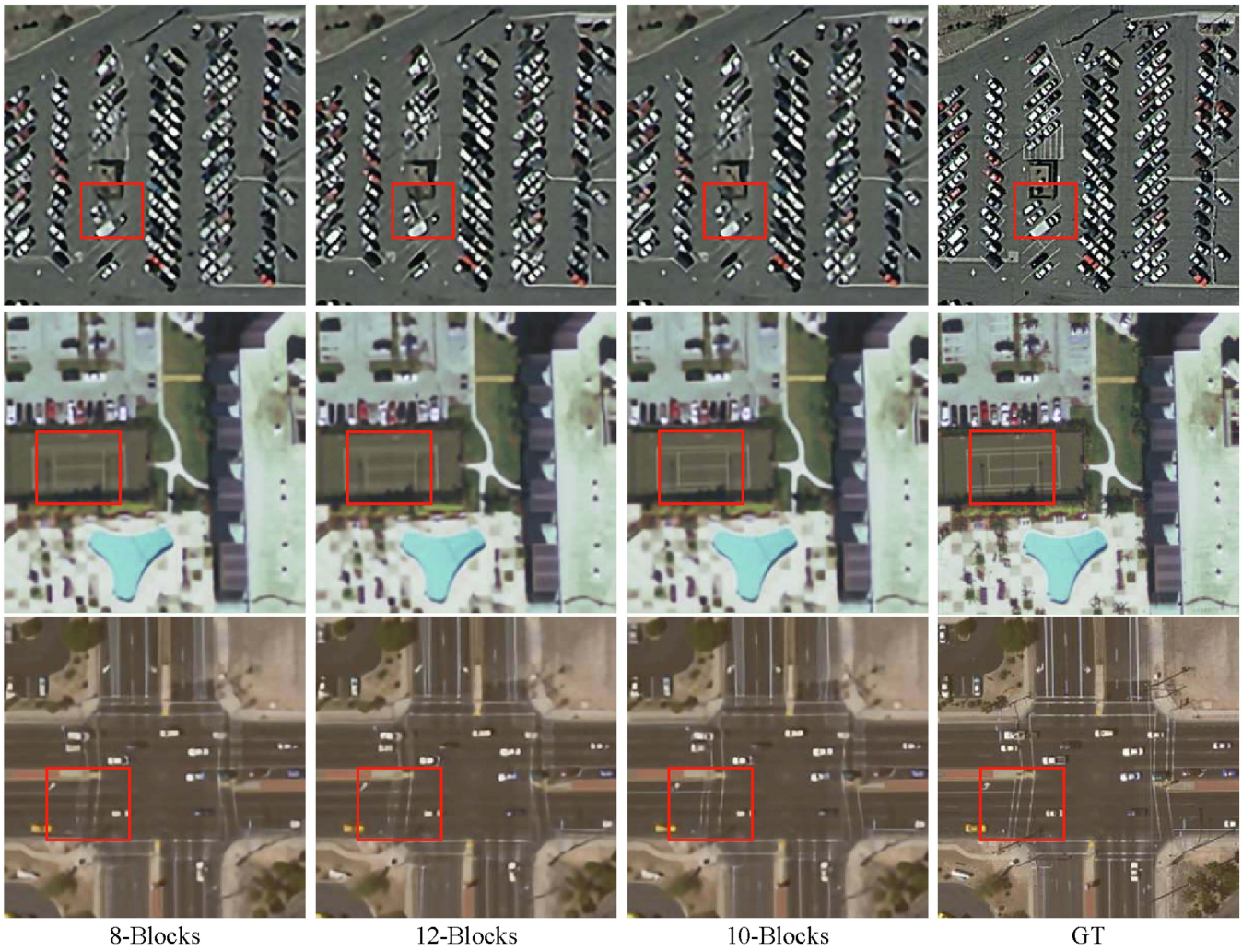


Fig. 7. Visual comparisons of RCL-Net with different numbers of PC-Block on the UCMerced dataset for $\times 4$ upscaling factor. Each column represents the network with the counterpart number PC-Block, where the last column denotes the ground-truth images.

TABLE VI
PSNR (dB) AND SSIM VALUES OF RCL-NET WITH DIFFERENT NUMBERS OF PC-BLOCK

| Number | PSNR | SSIM |
|---|---|---|
| 8 | 27.78 | 0.7642 |
| 10 | 27.79 | 0.7656 |
| 12 | 27.74 | 0.7643 |

*1) Number of PC-Block:* In order to balance the model complexity and reconstruction accuracy, we provide a deep analysis of our network with different numbers of PC-Block. As shown in Table VI, when we increase the number of PC-Block from 8 to 10 with enlarged model complexity, the reconstruction accuracy will be further promoted. However, when the number of PC-Block is increased from 10 to 12, the reconstruction accuracy is slipped. Through the exact exploitation, we conclude that RCL-Net would be somehow overfitting with twelve PC-Blocks. Specifically, the PSNR/SSIM values of 10 PC-Blocks are higher than 8 and 12 blocks by 0.01/0.0014 and 0.05/0.0013, respectively.

Additionally, the qualitative visual comparisons are provided in Fig. 7, which apparently illustrates that the network with 10 PC-Blocks could exactly recover remote sensing SR image details. Specifically, the third row in Fig. 7 shows that our network with 10 PC-Blocks can clearly restore the zebra crossing at the intersection while the images restored by the networks with 8 and 12 PC-Blocks contain blurry artifacts. Similarly, for the tennis court in the second-row images in Fig. 7, the recovered results of our network are almost without any distortions while the networks with 8 and 12 PC-Blocks recover images with blurry artifacts on the edge. Therefore, considering the quantitative and qualitative results, we can conclude that 10 times the usages of PC-Block structures from simple to complicated can better recover image details. Consequently, we employ 10 PC-Blocks as the default setting in our experiments.

*2) Structure of PC-Block:* To obtain realistic SR remote sensing images, the structures of PC-Block are progressively complicated as the network deepens, where deep features are gradually separated into two, four, and eight subparts, to recover images in structures from simple to complicated. In order to

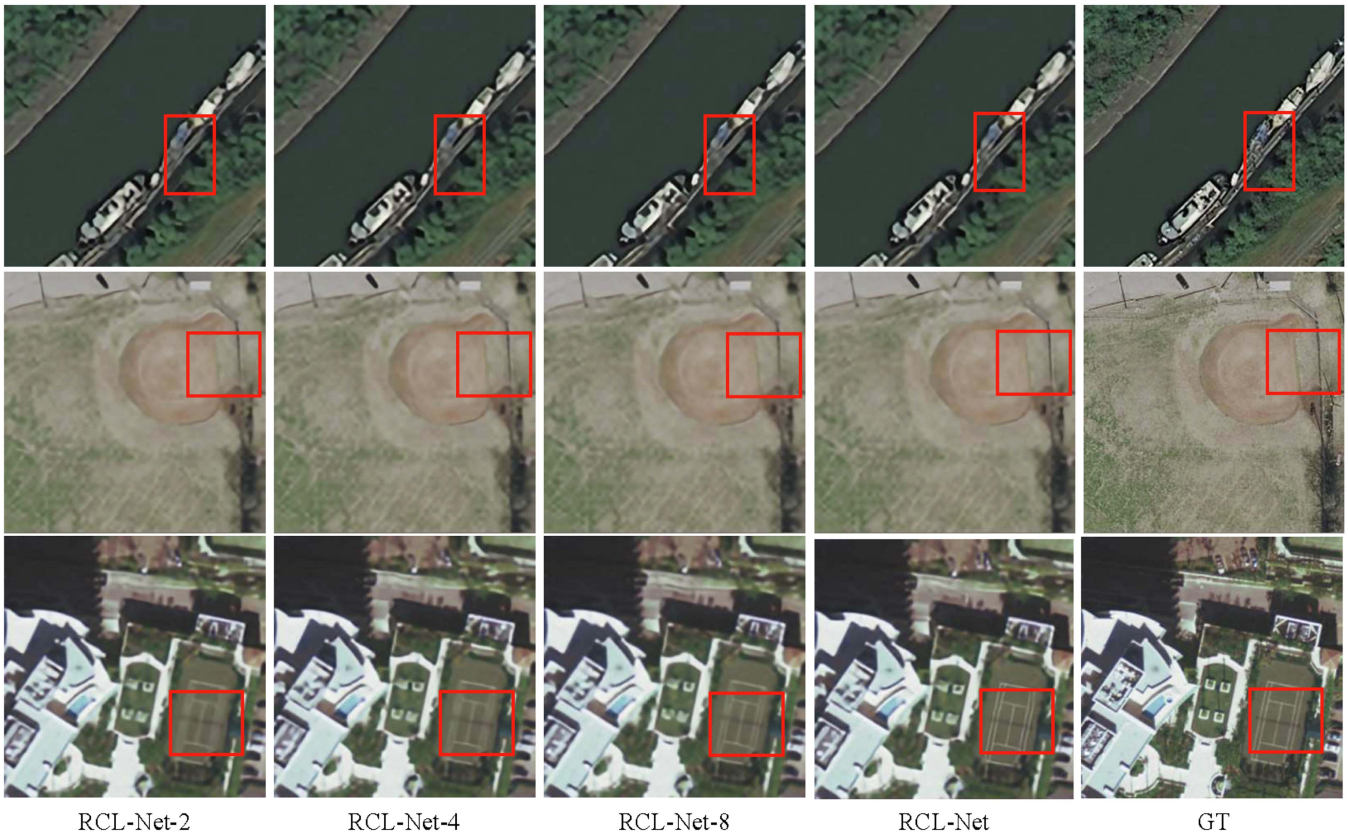


Fig. 8. Visual comparisons of RCL-Net with different CL-Branch structures on the UCMerced dataset for $\times 4$ upscaling factor. Each column represents the network with corresponding CL-Branch structures, where the last column refers to ground-truth images.

TABLE VII
PSNR (dB) AND SSIM VALUES OF RCL-NET WITH DIFFERENT CL-BRANCH STRUCTURES

| Structure | PSNR | SSIM |
|---|---|---|
| RCL-Net-2 | 27.74 | 0.7631 |
| RCL-Net-4 | 27.78 | 0.7650 |
| RCL-Net-8 | 27.79 | 0.7650 |
| RCL-Net | 27.79 | 0.7656 |

analyze the effectiveness of PC-Block, we first replace the four and eight branches with the same two branch structures in PC-Block, respectively, which is denoted as RCL-Net-2. Meanwhile, we build the network RCL-Net-4, where the two and eight branches of PC-Block are replaced with the same four branch structures. Similarly, RCL-Net-8 is also built only with eight branch structures to replace the two and four branches in PC-Block. Afterward, we provide quantitative and qualitative results for the comparisons. As we can conclude from Table VII, the increasingly complicated designs of PC-Block to recover image details through structures recurrently from simple to complicated would promote the model results, where RCL-Net obtains the best PSNR value and the SSIM result is beyond other designs simultaneously.

Several qualitative results are also given in Fig. 8, which illustrates that RCL-Net could accurately recover image details. For example, RCL-Net could more clearly restore the footpath in the field in the second row of Fig. 8. Meanwhile, as for the tennis court contour in the third row of Fig. 8, RCL-Net reconstructs more accurate textures compared to RCL-Net-2, RCL-Net-4, and RCL-Net-8. Although RCL-Net-8 with eight branches in all CL-Branch contains more complicated structures, our designed RCL-Net with the simpler framework still obtains better reconstruction performances. Therefore, the quantitative and qualitative results clearly demonstrate the superiority of designed RCL-Net with structures recurrently from simple to complicated.

*3) MultiScale Settings of MRF-Module:* In order to further explore the designed procedures of recurrent fluctuated complexity in RCL-Net, we replace the multiscale designs in MRF-Module with different single-scale designs to analyze model results. As we can conclude from Table VIII, RCL-Net with 5 $\times$ 5 and 7 $\times$ 7 scale settings perform slightly worse than the one with 3 $\times$ 3 scale setting, where the ignorance of small receptive fields could cause model performance losses. Furthermore, model results with 3 $\times$ 3 scale setting achieve the PSNR as 27.75 and the SSIM as 0.7639, which are better than HSENet and comparative to TransENet with less than 3% parameters. Such phenomenon indicates the effectiveness of designed procedures of recurrent fluctuated complexity even without multiscale designs. To further explore the effectiveness of multiscale designs in MRF-Module, we fix the MRF-Module with 3 $\times$ 3 scale setting in different CL-Branch, where the receptive fields are

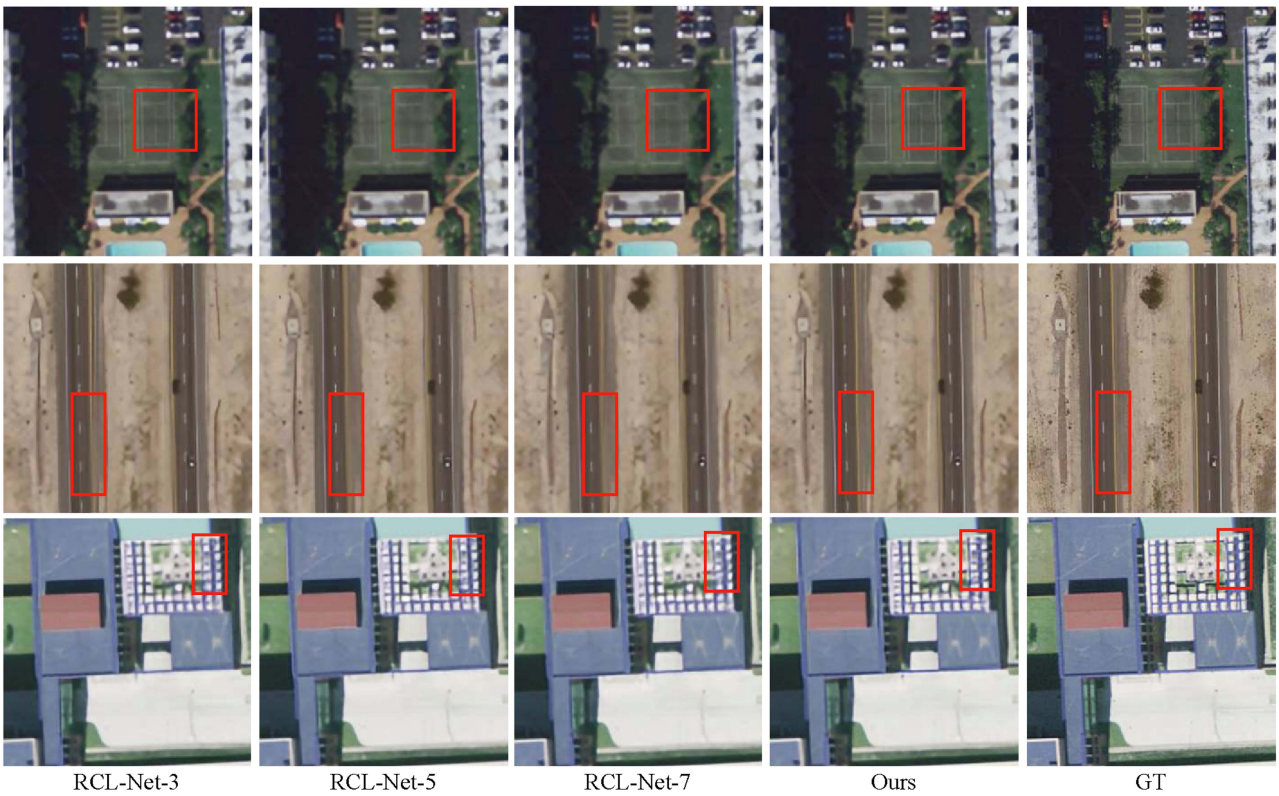


Fig. 9. Comparison of visual results of the recovered SR images corresponding to RCL-Net with single-scale settings in MRF-Module. RCL-Net followed by different numbers indicate the visual results corresponding to different single-scale settings.

TABLE VIII
PSNR (dB) AND SSIM VALUES OF RCL-NET WITH DIFFERENT SCALE SETTINGS IN MRF-MODULE

| Structure | PSNR | SSIM |
|---|---|---|
| RCL-Net-3 | 27.75 | 0.7639 |
| RCL-Net-5 | 27.71 | 0.7624 |
| RCL-Net-7 | 27.68 | 0.7606 |
| RCL-Net-1-fix | 27.78 | 0.7643 |
| RCL-Net-2-fix | 27.72 | 0.7652 |
| RCL-Net-3-fix | 27.75 | 0.7630 |
| RCL-Net | 27.79 | 0.7656 |

The numbers 3, 5, and 7 in the structure name from the second to the fourth rows denote different single-scale settings in MRF-Module. The numbers 1, 2, and 3 in the structure name from the fifth to seventh rows indicate which CL-Branch to be with a fixed-scale setting.

constrained to different degrees in the designed procedures of recurrent fluctuated complexity. As it is shown in Table VIII, with the fixed scale setting in CL-Branch-1, RCL-Net suffers slight performance losses of 0.01 in PSNR and 0.0013 in SSIM while 0.04 drops in PSNR and 0.0026 declines in SSIM can be caused by the absence of multiscale setting in CL-Branch-3. Consequently, we can conclude that the multiscale setting in MRF-Module can help to improve the effectiveness of designed procedures of recurrent fluctuated complexity in RCL-Net.

Several qualitative results can be seen in Fig. 9. We can find out in the first two rows that RCL-Net with single-scale settings can still recover image details such as straight lines while the intact one with multiscale settings in MRF-Module obtains the best results. Moreover, in the third row of Fig. 9, RCL-Net with $3 \times 3$ scale setting achieves comparative results in recovering square objects while the intact one still performs better in recovering object edges. Furthermore, in Fig. 10, the effectiveness of multiscale setting to improve the performance of designed procedures of recurrent fluctuated complexity can be observed. Therefore, the quantitative and qualitative results clearly demonstrate the effectiveness of both multiscale designs in MRF-Module and designed procedures of recurrent fluctuated complexity in RCL-Net.

### F. Robustness Experiment

Furthermore, to demonstrate the superiority and robustness of our network, we perform robustness experiments on the UCMerced dataset with $\times 4$ upscaling factor compared to advanced powerful large models. Specifically, we inject Gaussian noises with different variances and default mean values 0 on the UCMerced test dataset, which would then serve for the evaluations over the trained models. As shown in Table IX, where the bold represents the best results, our designed lightweight RCL-Net still evidently exceeds the advanced large models HSENet and TransENet when the noise increases. For instance, with variances of 0.0002, the PSNR/SSIM values of RCL-Net surpass TransENet and HSENet by 0.39/0.0279 and 0.61/0.0308, respectively. Furthermore, with variances increasing to 0.001, the designed RCL-Net also significantly outperforms TransENet and HSENet by 0.97/0.22 in PSNR and 0.0704/0.0172 in SSIM, respectively.

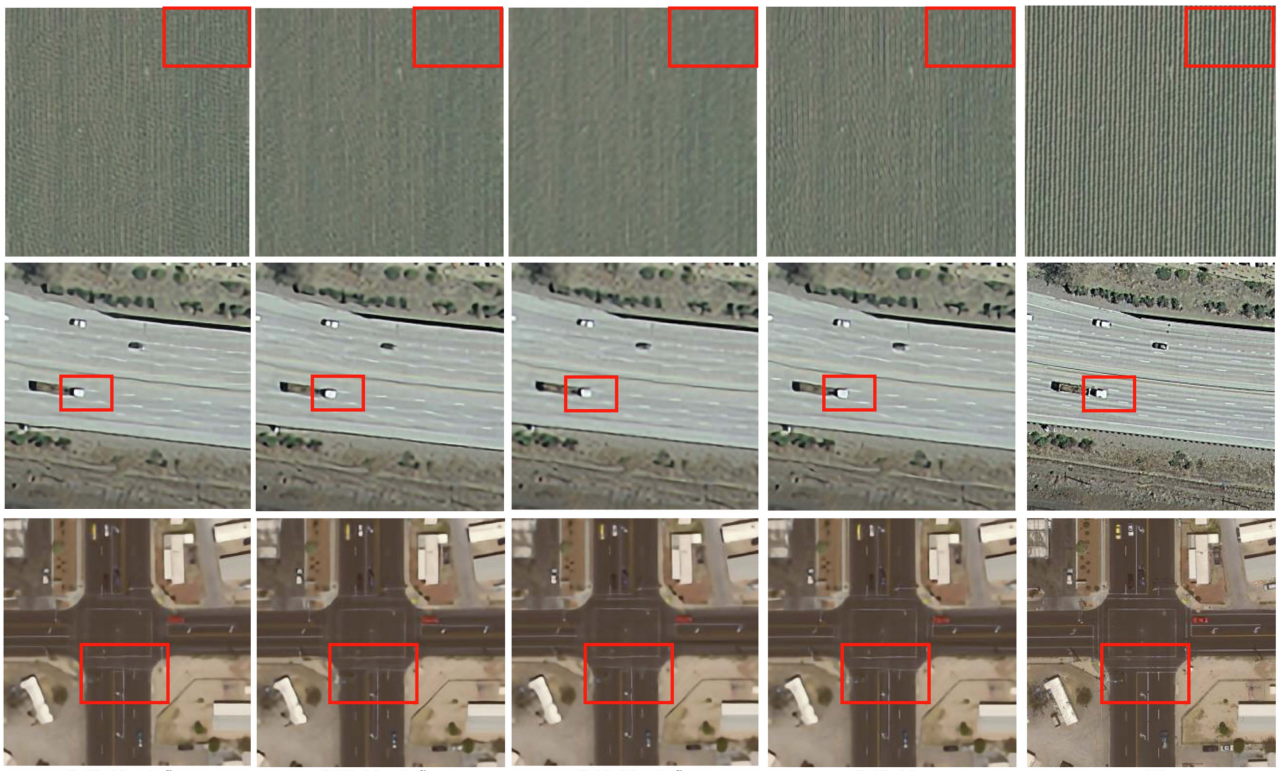


Fig. 10. Comparison of visual results of the recovered SR images corresponding to RCL-Net with fixed 3 × 3 scale setting of MRF-Module in particular CL-Branch. The numbers following RCL-Net indicate which CL-Branch to be with a fixed-scale setting.

TABLE IX
AVERAGE PSNR (DB) AND SSIM OVER THE ROBUSTNESS EXPERIMENTS ON THE UCMERCED TEST DATASET, WHICH IS INJECTED DIVERSE GAUSSIAN NOISES WITH DIFFERENT VARIANCES

| Method | Var=0.0002 PSNR/SSIM | Var=0.0003 PSNR/SSIM | Var=0.0005 PSNR/SSIM | Var=0.0007 PSNR/SSIM | Var=0.0009 PSNR/SSIM | Var=0.001 PSNR/SSIM |
|---|---|---|---|---|---|---|
| HSENet [40] | 25.83/0.6771 | 25.52/0.6588 | 24.95/0.6262 | 24.50/0.5998 | 24.11/0.5765 | 23.92/0.5654 |
| TransENet [54] | 26.05/0.6800 | 25.50/0.6491 | 24.63/0.5983 | 23.96/0.5583 | 23.41/0.5262 | 23.17/0.5122 |
| Ours | **26.44/0.7079** | **26.03/0.6866** | **25.35/0.6503** | **24.81/0.6204** | **24.35/0.5945** | **24.14/0.5826** |

Besides, the qualitative comparisons are also provided in Fig. 11, where each row represents the recovered images derived from counterpart methods with different noise settings denoted at the bottom. Specifically, with variances as 0.0002, RCL-Net could clearly recover the guidepost with sharper edges compared to results restored by large models HSENet and TransENet. Even with the injected noises increasing to 0.001, RCL-Net could still reconstruct the guidepost. In summary, the superiority of our designed lightweight RCL-Net compared to advanced large models HSENet and TransENet could be steadily observed with excellent quantitative and qualitative robust results, which demonstrate the tradeoff between powerful performances and low model complexities achieved by RCL-Net.

### G. Visual Quality on Real-World Scenes

In addition to PSNR and SSIM, we further apply TransENet, TTST, and RCL-Net on the real-world scene, where natural image quality evaluator (NIQE) [67] is exploited as the evaluation metric to assess the visual quality of recovered images derived from each method. As a reference-free image quality evaluation method, NIQE aims to measure whether the model result looks like natural scenes by measuring the difference between built multivariate normal distributions. Quantitative and visual comparisons are listed in Fig. 12, where we can conclude the effectiveness of the proposed RCL-Net. Specifically, RCL-Net achieves at least 1.42 performance gain compared to TransENet and TTST on the first row, where RCL-Net recovers a more straight harbor line on the top of the image. As for the second row, at least 0.04 performance gain can be raised by RCL-Net, where the border of recovered traffic markings is more straight in the middle of the image. Finally, as for the third row, RCL-Net performs better by at least 1.16 in terms of NIQE, where RCL-Net obtains more natural results on the top of buildings in the marked red rectangle. In summary, RCL-Net can also obtain better results in terms of visual comparison on real-world scenes.

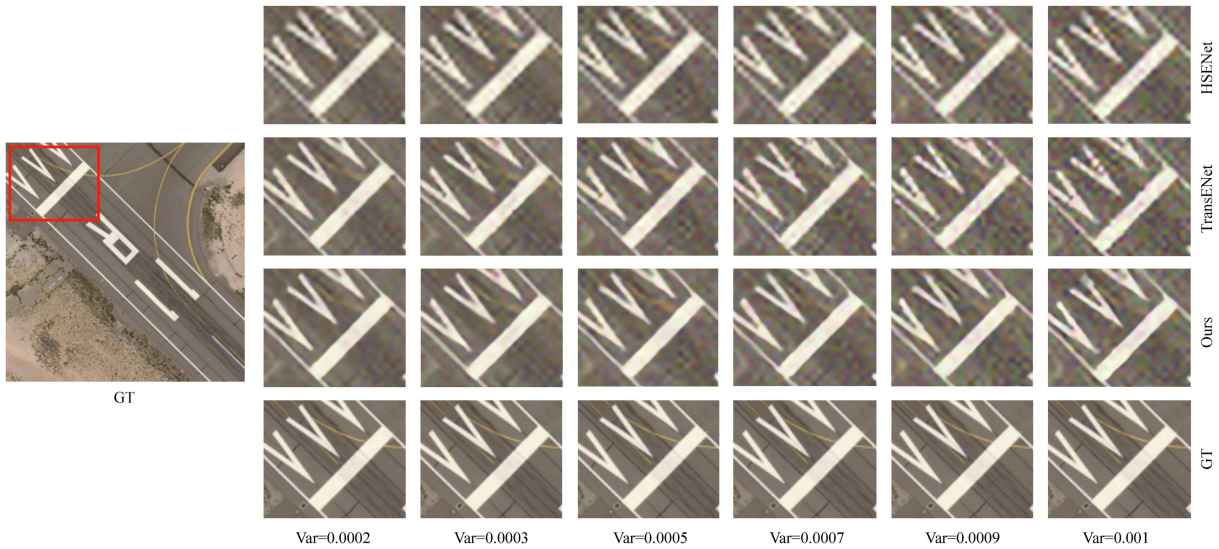


Fig. 11. Visualization of the recovered SR images, which are initially inserted Gaussian noises with different variances represented in the bottom row. Besides, each column image represents the reconstructed results of corresponding methods with different noises.


Fig. 12. Quantitative and visual comparison on the real-world scenes, where NIQE is applied as the evaluation metric.

## V. Conclusion

In this article, we design an RCL-Net for remote sensing image SR in structures recurrently from simple to complicated. Specifically, we design a serial of PC-Blocks in RCL-Net, where each PC-Block is composed of three CL-Branches with progressive complexities in order to recover image details little by little. Meanwhile, we design a cascade structure in each CL-Branch integrated with MRF-Module to obtain distinct forward paths for recovering heterogeneous image details, where constrained calculations between feature subparts help to reform the multibranch structures for reducing learnable parameters. We implement abundant experiments on the popular UCMerced dataset and AID dataset. The experimental results demonstrate that RCL-Net is superior to both classical and SOTA lightweight networks while RCL-Net even surpasses the advanced powerful large model with less than 3% parameters. Meanwhile, the ablation studies and feature visualizations validate the effectiveness of designed structures in our proposed RCL-Net. However, the employed multiscale modeling and coarse-to-fine strategy can still increase the parameter quantities and model complexities when aiming to improve the model performance. More efficient procedures, such as dictionary-based methods with less employed parameters, can be explored more deeply in the future.

## References


[1] B. Ping, F. Su, X. Han, and Y. Meng, "Applications of deep learning-based super-resolution for sea surface temperature reconstruction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 887–896, 2021.

[2] Z. Ren, L. He, and J. Lu, "Context aware edge-enhanced GAN for remote sensing image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1363–1376, 2024.

[3] X. Ma, H. Li, and Z. Chen, "Feature-enhanced deep learning network for digital elevation model super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5670–5685, Jun. 2023.

[4] A. Ulrichsen et al., "Operational neural networks for parameter-efficient hyperspectral single-image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1470–1484, 2024.

[5] X. Jing et al., "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1363–1378, Mar. 2017.

[6] G. Zhang, Y. Ge, Z. Dong, H. Wang, Y. Zheng, and S. Chen, "Deep high-resolution representation learning for cross-resolution person re-identification," *IEEE Trans. Image Process.*, vol. 30, pp. 8913–8925, Oct. 2021.

[7] J. Hatvani, A. Basarab, J. Tourneret, M. Gyöngy, and D. Kouamé, "A tensor factorization method for 3-D super resolution with application to dental CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 6, pp. 1524–1531, Jun. 2019.

[8] J. Shermeyer and A. V. Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1432–1441.

[9] X. Xu, B. Pan, Z. Chen, Z. Shi, and T. Li, "Simultaneously multiobjective sparse unmixing and library pruning for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3383–3395, Apr. 2021.

[10] Z. Dong, M. Wang, Y. Wang, Y. Zhu, and Z. Zhang, "Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2104–2114, Mar. 2020.

[11] R. Mathieu, C. Freeman, and J. Aryal, "Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery," *Landscape Urban Plan.*, vol. 81, no. 3, pp. 179–192, 2007.

[12] S. Lei, Z. Shi, X. Wu, B. Pan, X. Xu, and H. Hao, "Simultaneous super-resolution and segmentation for remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3121–3124.

[13] 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.

[14] D. Rajan and S. Chaudhuri, "Generalized interpolation and its application in super-resolution imaging," *Image Vis. Comput.*, vol. 19, no. 13, pp. 957–969, 2001.

[15] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. Katsaggelos, "SoftCuts: A soft edge smoothness prior for color image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 969–981, May 2009.

[16] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[17] S. Shang et al., "ResDiff: Combining CNN and diffusion model for image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, pp. 8975–8983.

[18] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshop*, 2018, vol. 11133, pp. 63–79.

[19] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 1905–1914.

[20] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

[21] 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.

[22] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.

[23] 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.*, 2016, pp. 1646–1654.

[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] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[26] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 1833–1844.

[27] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[28] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern. Recognit.*, 2023, pp. 22367–22377.

[29] Q. Cai et al., "HIPA: Hierarchical patch transformer for single image super resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 3226–3237, May 2023.

[30] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[31] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.

[32] N. Ahn, B. Kang, and K. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 252–268.

[33] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 723–731.

[34] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. ACM Multimedia Conf.*, 2019, pp. 2024–2032.

[35] Y. Yuan and C. Yuan, "Efficient conditional diffusion model with probability flow sampling for image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, pp. 6862–6870.

[36] H. Song, B. Huang, Q. Liu, and K. Zhang, "Improving the spatial resolution of landsat TM/ETM+ through fusion with SPOT5 images via learning-based super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1195–1204, Mar. 2015.

[37] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019.

[38] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7086–7098, Nov. 2014.
[39] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.
[40] S. Lei and Z. Shi, "Hybrid-scale self-similarity exploitation for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5401410.
[41] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.
[42] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3633–3643, May 2020.
[43] Y. Xiao, Q. Yuan, K. Jiang, J. He, C. Lin, and L. P. Zhang, "TTST: A top-k token selective transformer for remote sensing image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 738–752, Jan. 2024.
[44] 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, 2022, Art. no. 5615313.
[45] T. Zhang, C. Bian, X. Zhang, H. Chen, and S. Chen, "Lightweight remote-sensing image super-resolution via re-parameterized feature distillation network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Mar. 2023, Art. no. 6003905.
[46] 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.
[47] 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.
[48] J. Liu, J. Tang, and G. Wu, "Residual feature distillation network for lightweight image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 41–55.
[49] Z. Li et al., "Blueprint separable residual network for efficient image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 833–843.
[50] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.
[51] Z. Pan et al., "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4864–4876, Sep. 2013.
[52] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2312–2327, Apr. 2018.
[53] J. Haut, M. Paoletti, R. Fernández-Beltran, J. Plaza, A. Plaza, and J. Li, "Remote sensing single-image superresolution based on a deep compendium model," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1432–1436, Sep. 2019.
[54] S. Lei, Z. Shi, and W. Mo, "Transformer-based multistage enhancement for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5615611.
[55] 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, 2023.
[56] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, and Y. Yao, "Deep distillation recursive network for remote sensing imagery super-resolution," *Remote. Sens.*, vol. 10, no. 11, 2018, Art. no. 1700.
[57] D. Zihao, N. Sijie, G. Xizhan, and S. Xiuli, "Coarse-to-fine online latent representations matching for one-stage domain adaptive semantic segmentation," *Pattern Recognit.*, vol. 146, 2024, Art. no. 110019.
[58] D. Shenaj, F. Barbato, U. Michieli, and P. Zanuttigh, "Continual coarse-to-fine domain adaptation in semantic segmentation," *Image Vis. Comput.*, vol. 121, 2022, Art. no. 104426.
[59] H. Ma, X. Lin, Z. Wu, and Y. Yu, "Coarse-to-fine domain adaptive semantic segmentation with photometric alignment and category-center regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4050–4059.
[60] C. Tian, Y. Xu, W. Zuo, B. Zhang, L. Fei, and C. Lin, "Coarse-to-fine CNN for image super-resolution," *IEEE Trans. Multimedia.*, vol. 23, pp. 1489–1502, 2021.
[61] Y. Tang et al., "An image patch is a wave: Phase-aware vision MLP," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10925–10934.
[62] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUS)," 2016, *arXiv:1606.08415*.
[63] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
[64] 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.
[65] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
[67] 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.

**Duwei Hua** received the B.S. degree in electronic information engineering from Xinyang Normal University, Xinyang, China, in 2017. He is currently working toward the Ph.D. degree in optical engineering with the College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou, China.

His research interests include image superresolution and related vision tasks.

**Kunping Yang** (Member, IEEE) received the B.S. and M.S. degrees in mathematics from Wuhan University, Wuhan, China, in 2014 and 2016, respectively, and the Ph.D. degree in communication and information system from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, in 2022.

His research interests include deep learning techniques, semantic segmentation, and change detection.

**Jianchong Wei** received the Ph.D. degree in optical engineering from the College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou, China, in 2023.

He is currently a Lecturer with the College of Electronics and Information Science, Fujian Jiangxia University, Fuzhou. His research interests include image dehazing, image restoration, and related vision problems.

**Liang Chen** (Member, IEEE) received the Ph.D. degree in communication and information systems from NERCMS, Wuhan University, Wuhan, China, in 2017, and a joint Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, China, in 2018.

She is currently an Associate Professor with Fujian Normal University, Fuzhou, China. Her research interests include image superresolution and related vision problems.

**Dingli Xue** received the B.S. degree in information engineering from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2014, and the Ph.D. degree in electrical engineering from Florida State University, Tallahassee, FL, USA, in 2020.

From 2021 to 2023, he was with Tsinghua University, where he left as a Postdoctoral Fellow of Electrical Engineering. He has been an Assistant Researcher (Professor) with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, since September 2023. His research interests include Hirschman theory, entropic uncertainty, time–frequency analysis, target detection, and efficient FPGA filter.

**Yi Wu** (Member, IEEE) received the B.S. degree in radio technology from Southeast University, Nanjing, China, in 1991, the M.S. degree in communications and information systems from Fuzhou University, Fuzhou, China, in 2004, and the Ph.D. degree in information and communication engineering from Southeast University in 2013.

She is currently a Professor with the College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou. Her research interests include 5G networks, vehicular ad hoc networks, and communication protocols.