IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 August 2023, accepted 3 September 2023, date of publication 11 September 2023,
date of current version 14 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3314196

RESEARCH ARTICLE

Single Image Super Resolution via Multi-Attention Fusion Recurrent Network

QIQI KOU[1], DEQIANG CHENG[2], (Member, IEEE), HAOXIANG ZHANG[2], (Graduate Student Member, IEEE), JINGJING LIU[2], XIN GUO[3], AND HE JIANG[2]

[1]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[2]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[3]Huawei Hangzhou Research Institute, Hangzhou 310007, China

Corresponding author: He Jiang (jianghe@cumt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52204177 and Grant 52304182, and in part by the Fundamental Research Funds for the Central Universities under Grant 2020QN49.

ABSTRACT Deep convolutional neural networks have significantly enhanced the performance of single image super-resolution in recent years. However, the majority of the proposed networks are single-channel, making it challenging to fully exploit the advantages of neural networks in feature extraction. This paper proposes a Multi-attention Fusion Recurrent Network (MFRN), which is a multiplexing architecture-based network. Firstly, the algorithm reuses the feature extraction part to construct the recurrent network. This technology reduces the number of network parameters, accelerates training, and captures rich features simultaneously. Secondly, a multiplexing-based structure is employed to obtain deep information features, which alleviates the issue of feature loss during transmission. Thirdly, an attention fusion mechanism is incorporated into the neural network to fuse channel attention and pixel attention information. This fusion mechanism effectively enhances the expressive power of each layer of the neural network. Compared with other algorithms, our MFRN not only exhibits superior visual performance but also achieves favorable results in objective evaluations. It generates images with sharper structure and texture details and achieves higher scores in quantitative tests such as image quality assessment.

INDEX TERMS Super resolution, multiplexing-based, attention fusion mechanism, recurrent network.

I. INTRODUCTION

Single Image Super Resolution (SISR) has found widespread applications, including object tracking and detection [1], scene classification and reconstruction [2], medical image analysis [3], [43], remote sensing [39], [40], [41], [46], [47], [48], [49] and face hallucination [42], [44]. Notably, the quality of the reconstructed images has a direct bearing on the accuracy of the aforementioned applications. By employing sophisticated algorithms, SISR can upsample Low Resolution (LR) images to produce High Resolution (HR) images, thereby enhancing image quality in a hardware-independent manner. Given its significant practical utility, SISR has emerged as an active area of research.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

With the rapid advancement of convolutional neural networks and deep learning technology, traditional models such as [7] and [13] have been shown to be inadequate in feature extraction. The introduction of SRCNN [4] represents the first attempt to incorporate a convolutional neural network in SISR. Subsequently, FSRCNN [5] is proposed to expedite the training process of SRCNN. ESPCN [6] is explored to upsample LR images using subpixel-based techniques. Despite improving reconstruction performance, these algorithms suffer from limited network depth and significant information loss. Deep networks are capable of extracting deep features, but training them is challenging due to the problem of gradient information vanishing during transmission. The residual network [22] effectively addresses this issue. VDSR [8] utilizes the residual network to generate SISR results that produce clear images under different

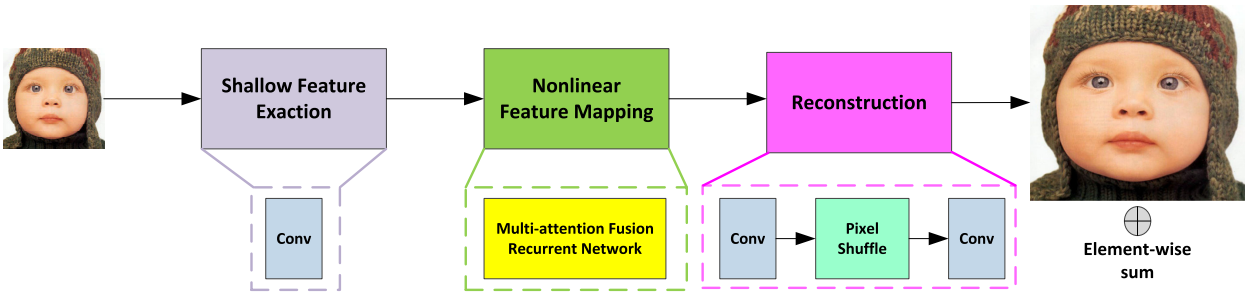


FIGURE 1. The simplified network architecture of the SISR system based on multi-attention fusion recurrent network (MFRN).

magnification factors. The Deep Laplacian network, namely LapSRN [14], gradually zooms images to avoid feature loss caused by direct scaling. EDSR [15] is a refinement of the residual network that eliminates the unnecessary batch normalization component to enhance the model's feature representation capabilities. Owing to its simplicity and superior expressiveness, EDSR has become the benchmark model for the SISR task.

Despite the impressive performance of deep convolutional neural networks in SISR, existing algorithms face several unresolved issues. Firstly, to achieve superior performance, many algorithms add more convolutional layers to their network structure, which increases the training time significantly. Secondly, single-channel network architecture may result in the unreasonable utilization of information resources. Thirdly, most deep models process features indiscriminately, and this equal processing results in the waste of valuable feature information extracted from previous layers, leading to high computational overhead and low efficiency. To address these limitations, researchers have developed various methods, such as Recurrent Neural Network (RNN) [16], [17], [19], [20], [25], [31], [32], [35], [45], feature attention mechanisms [9], [10], [34], [37], sparse representation [33], and knowledge distillation [23], [36] to improve the information utilization and feature representation.

RNN offers the advantage of parameter sharing across different time steps, resulting in a more lightweight model architecture. Furthermore, RNNs have the ability to capture complex feature dependencies, allowing for spatial interactions between different layers. These characteristics make RNNs well-suited for tackling SISR problems, and it is adopted in DRCN [20], DRRN [17], RDRN [25], DRUDN [19], MCSR [16], CARN [31], IMDN [32] LBNet [35] and HDRN [45]. These models are distinguished by their use of different recurrent structures. For instance, DRUDN [19] and LBNet [35] employ up-down sampling blocks and transformer blocks as the recurrent structure, respectively. Thus, the design of the recurrent structure not only serves as an effective means to distinguish different models but also represents a critical innovation point for diverse models.

The attention mechanism [9] has become a widely utilized technique in the field of SISR, as evidenced by its extensive applications in various studies [10], [34], [37]. This mechanism is introduced by RCAN [10] with the aim of enhancing the channel representation characteristics of SISR models. However, RCAN [10] tends to overlook the interplay between local and global features, which can potentially lead to a reduction in the brightness and fidelity of the reconstructed images. To address this issue, pixel attention [37] has emerged as a viable solution. In addition to mitigating the aforementioned challenges, pixel attention can also enhance the interpretability and robustness of the model, thereby enabling it to effectively handle diverse scenes. The organic combination of these two attention mechanisms can enable the neural network to selectively allocate computational resources to modules that have a greater impact on reconstruction performance, resulting in a significant enhancement of the model's overall reconstruction capability.

In this study, a novel Multi-attention Fusion Recurrent Network, namely MFRN, is proposed. Our contributions can be summarized as follows. Firstly, the issue of improper utilization of information in single-channel network structures is addressed by introducing a multiplexing-based network structure. This approach can acquire more comprehensive feature information and enhance the reconstruction capability. Secondly, the concept of recurrent networks is leveraged to optimize the training time of neural networks by reusing feature extraction modules. Thirdly, the attention fusion mechanism is integrated into the multiplexing-based network. Specifically, channel and pixel attention mechanisms are incorporated synergistically to facilitate feature information transfer in deep networks.

II. METHODS

The network's structure is depicted in Figure 1 and can be broadly categorized into three parts, a shallow feature extraction component comprising the first convolutional layer, a non-linear feature mapping element realized by the Multi-attention Fusion Recurrent Network (MFRN), and the image reconstruction component that employs the pixel shuffle technique.

A. SHALLOW FEATURE EXTRACTION

The shallow feature extraction part is the first step in the neural network. For instance, when dealing with an input image I, a convolutional layer is used to extract its shallow

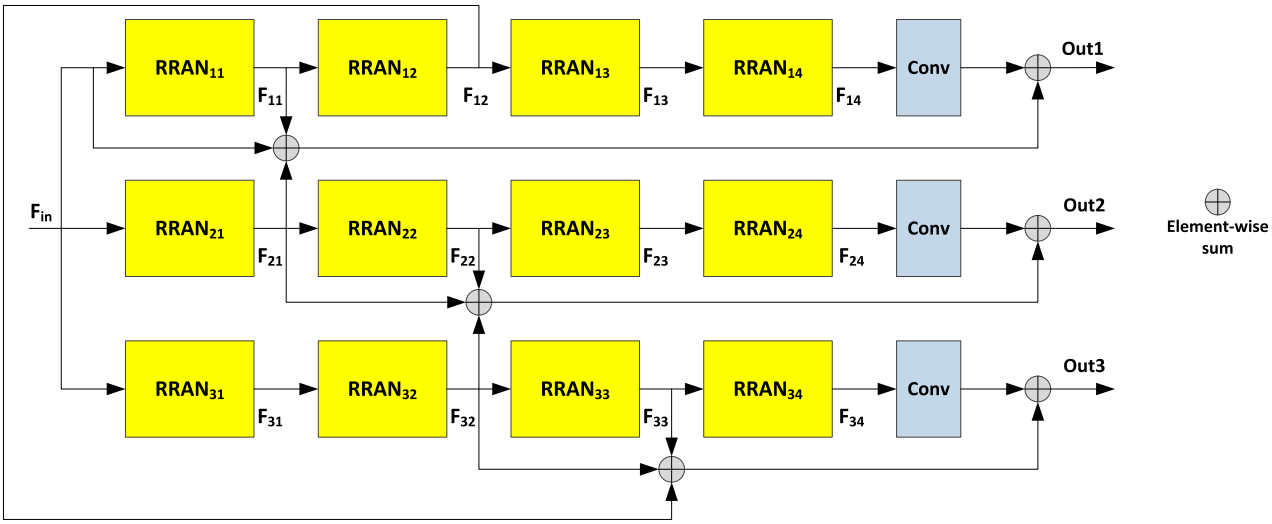


FIGURE 2. The net architecture of multi-attention fusion recurrent network (MFRN).

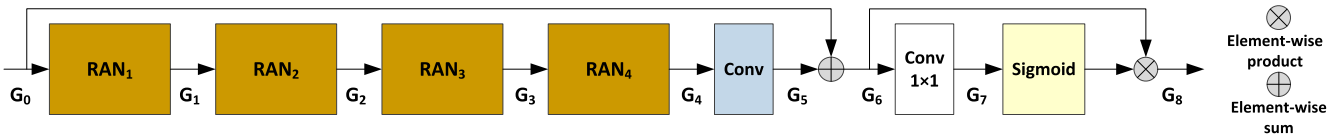


FIGURE 3. The net architecture of recurrent residual attention network (RRAN).

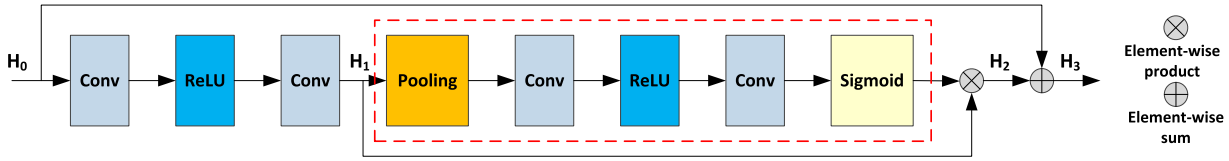


FIGURE 4. The net architecture of residual attention network (RAN).

features, which can be represented by Equation 1. In this part, $Conv(\cdot)$ stands for the convolution operation, whose kernel size is 3×3 , and F_{in} is the output feature map.

$$F_{in} = Conv(I) \tag{1}$$

B. MULTI-ATTENTION FUSION RECURRENT NETWORK

Feature extraction primarily captures low-frequency features, while more complex high-frequency features require a nonlinear feature mapping approach. In this study, MFRN is utilized to fit the nonlinear feature mapping part.

Based on Figure 2, MFRN can be characterized as a deep learning network comprising three channels, each of which encompasses four Recursive Residual Attention Network (RRAN) layers along with one 3×3 convolutional layer. In Figure 3, the RRAN layer is constructed from four Residual Attention Networks (RANs), a 3×3 convolutional layer, a 1×1 convolutional layer, and a feature activation unit $Sigmoid(\cdot)$. The 1×1 convolutional layer and feature activation unit $Sigmoid(\cdot)$ are particularly noteworthy, as they function as pixel attention mechanism that facilitates the system's ability to learn the supervised mask of features, thereby reducing the loss function value and reconstructing superior high-frequency information. The internal structure of the RAN layer is depicted in Figure 4, where the red dashed box denotes the Channel Attention Block (CAB) [10] structure. This structural element enables the system to process channel features differently, thus augmenting the ability of channel information representation.

To provide a clear depiction of each network's structure, signal transmission expressions are provided for each network module. In Equations 2 to 4, the following notations are used: $GPl(\cdot)$ represents global pooling, $Conv_{1\times1}(\cdot)$ and $Conv(\cdot)$ denote the convolution operations using 1×1 and 3×3 kernel size, $\delta(\cdot)$ means the nonlinear activation unit $ReLU(\cdot)$ [21] and $Sm(\cdot)$ is the activation function $Sigmoid(\cdot)$, which outputs values between 0 and 1. $\oplus$ and $\otimes$ are adopted as element-wise sum and product symbols. Additionally, $H_0 \sim H_3$ represent high-dimensional feature maps and $CAB(\cdot)$ is short for Channel Attention Block. Specifically, $H_2 = H_1 \otimes CAB(H_1)$ and $H_3 = RAN(H_0)$.

$$H_1 = Conv(\delta(Conv(H_0))) \tag{2}$$

$$CAB(H_1) = Sm(Conv(\delta(Conv(GPl(H_1))))) \tag{3}$$

$$H_3 = RAN(H_0) = H_0 \oplus (H_1 \otimes CAB(H_1)) \tag{4}$$

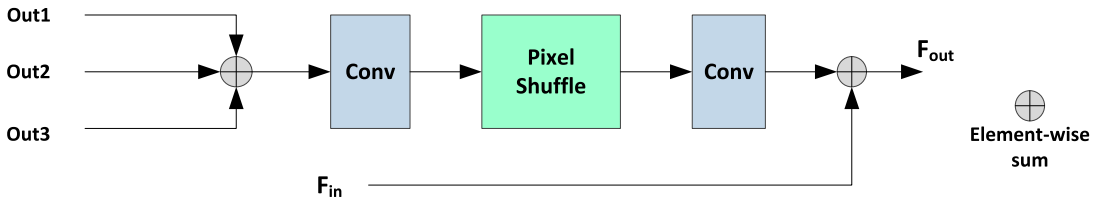


FIGURE 5. The net architecture of image reconstruction.

Figure 3 shows the transmission details in a RRAN, and $G_{k+1} = RAN_{k+1}(G_k)$ when k ranges from 0 to 3. Specially, $G_5 = Conv(G_4)$, $G_6 = G_0 \oplus G_5$, $G_7 = Conv_{1\times1}(G_6)$ and $G_8 = G_6 \otimes Sm(G_7)$. In particular, when the RRAN is the first network module of each channel, $G_0 = F_{in}$, and when the RRAN network locates in row i and column j in MFRN in Figure 2, $G_0 = F_{i,j-1}$, $G_8 = F_{i,j}$, and $F_{i,j} = RRAN_{ij}(F_{i,j-1})$. In this subsection, the value of i and j ranges from $1 \sim 3$ and $1 \sim 4$, respectively.

In convolutional neural networks, using too many layers can lead to loss of feature information and degraded reconstruction quality. Additionally, the convolutional layer's computationally-intensive nature is a primary reason for slow program execution. This paper presents a novel multiplexing-based information reuse method called MFRN, which retains a more comprehensive set of feature information and enables parameter sharing between different channels to avoid redundant computations. Equations $5 \sim 7$ illustrate the signal multiplexing technique, where F_{in} is the input signal of the MFRN, and F_{11}, F_{12}, F_{14}, F_{21}, F_{22}, F_{24}, F_{31}, F_{32}, and F_{34} are node signals in the MFRN network used to fuse feature information across different channels, thereby enabling the reuse of information from other channels. $Out1$, $Out2$, and $Out3$ are the MFRN's three-way signal outputs.

$$Out1 = Conv(F_{14}) \oplus F_{in} \oplus F_{11} \oplus F_{21} \tag{5}$$
$$Out2 = Conv(F_{24}) \oplus F_{21} \oplus F_{22} \oplus F_{32} \tag{6}$$
$$Out3 = Conv(F_{34}) \oplus F_{32} \oplus F_{33} \oplus F_{12} \tag{7}$$

The RRAN not only preserves more feature information but also enhances network training efficiency. The experiments indicate that the recurrent approach is about 30% faster than that of non-recurrent networks in training time.

C. RECONSTRUCTION MODULE

Figure 5 showcases the process of image reconstruction, which is described mathematically in Equation 8. In this equation, $Out1$, $Out2$, and $Out3$ denote the three outputs of the MFRN network, while F_{in} refers to the shallow feature map of the input image I. The upsampling operation is carried out using pixel shuffle $PS(\cdot)$ and F_{out} represents the super-resolution result of I.

$$F_{out} = F_{in} \oplus Conv(PS(Conv(Out1 \oplus Out2 \oplus Out3))) \tag{8}$$

Shi et al. mentioned in their paper [6] that using deconvolution layers for image upsampling can generate a significant amount of redundant information between pixels. This can result in a suboptimal upsampling effect or even negatively impact the gradient descent algorithm in severe cases. Therefore, pixel shuffle is adopted for image reconstruction.

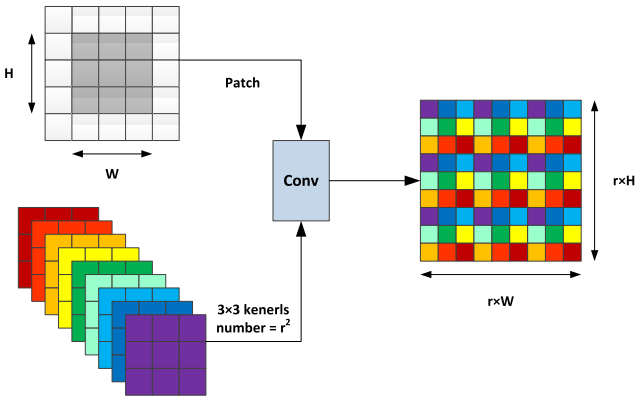


FIGURE 6. The pixel shuffle mechanism used for image reconstruction.

Figure 6 depicts the usage of the pixel shuffle for upsampling. For an input feature map with resolution $H \times W$, the number of convolution kernels is r^2, and the output image resolution is $rH \times rW$. For SISR tasks with magnification factors of 3 or 4, the method proposed in paper [15] is utilized. By leveraging the pre-trained network, the training efficiency of the model can be improved for larger factors, and the overall model's operation time can be reduced.

D. LOSS FUNCTION

In the training process, the L_1 loss function is used to constrain the learning of the MFRN network, as it is sensitive to data fluctuations and can effectively guide the update of model parameters. The L_1 loss function with parameter set Θ is represented by Equation 9.

$$Loss(\Theta) = \frac{1}{N}\sum_{i=1}^{N} ||F_{SR}(x_i) - X_i||_1 \tag{9}$$

In Equation 9, $[x_i, X_i]_{i=1}^{N}$ is the training set, where N is the number of training image patches. X_i represents one high-resolution image patch, which is regarded as the ground truth of the low-resolution patch x_i. $F_{SR}(x_i)$ represents the reconstructed super-resolution image patch of x_i.

TABLE 1. Quantitative tests: The average PSNR/SSIM values for ×2, ×3, ×4 SISR results are reported on Set5 [26], Set14 [27], BSD100 [28], Urban100 [29] and Manga109 [38], comparing the performance of MFRN with other methods. The best and second-best results are shown in black bold and blue bold, respectively.

Method	Scale	Set5 [26]	Set14 [27]	BSD100 [28]	Urban100 [29]	Manga109 [38]
SRCNN (TPAMI 2014) [4]	2	36.66/0.9542	32.43/0.9073	31.34/0.8879	29.50/0.8953	35.74/0.9661
DRCN (CVPR 2016) [20]	2	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723
LapSRN (TPAMI 2017) [14]	2	37.53/0.9590	32.99/0.9124	31.80/0.8953	30.45/0.9130	37.27/0.9740
EDSR (CVPR 2017) [15]	2	37.68/0.9582	33.28/0.9143	31.29/0.8974	31.29/0.9192	38.54/0.9769
DRRN (CVPR 2017) [17]	2	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.92/0.9758
CARN (ECCV 2018) [31]	2	37.76/0.9591	33.52/0.9164	32.09/0.8974	31.92/0.9254	-/-
RDRN (NC 2019) [25]	2	37.73/**0.9610**	33.25/0.9154	32.08/0.8983	31.25/0.9202	38.43/0.9761
IMDN (MM 2019) [32]	2	38.00/0.9604	33.63/0.9173	32.19/0.8994	32.17/0.9283	**38.88**/0.9774
DRUDN (NC 2020) [19]	2	37.68/0.9591	33.31/0.9150	32.02/0.8968	31.53/0.9224	38.08/0.9707
RFDN (ECCV 2020) [36]	2	**38.05**/0.9606	**33.68/0.9184**	32.16/0.8994	32.12/0.9274	**38.88**/0.9773
HDRN (PR 2020) [45]	2	37.75/0.9590	33.49/0.9150	32.03/0.8980	31.87/0.9250	38.07/0.9770
SMSR (CVPR 2021) [33]	2	38.00/0.9601	33.64/0.9173	32.17/0.8990	32.19/0.9284	38.76/0.9771
MCSR (IG 2021) [16]	2	38.03/0.9603	33.58/0.9172	32.18/0.8992	31.94/0.9262	38.45/0.9747
ARRFN (NC 2022) [34]	2	38.01/0.9603	33.66/0.9174	**32.20/0.8999**	32.37/**0.9295**	38.86/0.9768
LBNet (IJCAI 2022) [35]	2	**38.05**/0.9604	33.65/**0.9176**	32.15/0.8993	**32.30**/0.9283	**38.88/0.9775**
Our MFRN	2	**38.04/0.9614**	**33.67/0.9184**	**32.21/0.8997**	**32.28/0.9300**	**38.89/0.9779**
SRCNN (TPAMI 2014) [4]	3	32.75/0.9090	29.30/0.8219	28.41/0.7863	26.25/0.7989	30.59/0.9017
DRCN (CVPR 2016) [20]	3	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328
LapSRN (TPAMI 2017) [14]	3	33.82/0.9232	29.87/0.8324	28.83/0.7980	27.08/0.8281	32.21/0.9351
EDSR (CVPR 2017) [15]	3	34.13/0.9240	30.14/0.8377	28.98/0.8012	27.69/0.8407	33.45/0.9419
DRRN (CVPR 2017) [17]	3	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.42/0.9352
CARN (ECCV 2018) [31]	3	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	-/-
RDRN (NC 2019) [25]	3	34.10/0.9251	29.99/0.8362	28.96/0.8010	27.53/0.8380	33.15/0.9379
IMDN (MM 2019) [32]	3	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
DRUDN (NC 2020) [19]	3	34.25/0.9252	30.20/0.8383	29.01/0.8021	27.89/0.8463	33.25/0.9380
RFDN (ECCV 2020) [36]	3	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8524	33.67/0.9449
HDRN (PR 2020) [45]	3	34.24/0.9240	30.23/0.8400	28.96/0.8040	27.93/0.8490	33.17/0.9420
SMSR (CVPR 2021) [33]	3	34.40/0.9270	30.33/0.8412	29.10/0.8050	28.25/0.8536	33.68/0.9445
MCSR (IG 2021) [16]	3	34.44/0.9262	30.37/**0.8422**	29.11/0.8051	28.10/0.8511	33.19/0.9416
ARRFN (NC 2022) [34]	3	34.38/0.9272	30.36/**0.8422**	29.09/0.8050	28.22/0.8533	33.32/0.9439
LBNet (IJCAI 2022) [35]	3	**34.47/0.9277**	**30.38**/0.8417	**29.13/0.8061**	**28.42/0.8559**	**33.82/0.9460**
Our MFRN	3	**34.49/0.9278**	**30.39/0.8423**	**29.14/0.8071**	**28.45/0.8562**	**33.84/0.9463**
SRCNN (TPAMI 2014) [4]	4	30.48/0.8628	27.50/0.7503	26.90/0.7110	24.53/0.7212	27.66/0.7866
DRCN (CVPR 2016) [20]	4	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
LapSRN (TPAMI 2017) [14]	4	31.53/0.8850	28.19/0.7720	27.33/0.7270	25.20/0.7560	29.29/0.8900
EDSR (CVPR 2017) [15]	4	31.91/0.8893	28.44/0.7773	27.45/0.7307	25.65/0.7714	30.35/0.9067
DRRN (CVPR 2017) [17]	4	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8966
CARN (ECCV 2018) [31]	4	32.13/0.8937	28.60/0.7806	**27.58**/0.7349	26.07/0.7837	-/-
RDRN (NC 2019) [25]	4	31.77/0.8902	28.26/0.7743	27.43/0.7312	25.43/0.7633	28.65/0.8965
IMDN (MM 2019) [32]	4	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
DRUDN (NC 2020) [19]	4	32.17/0.8944	28.56/0.7801	27.54/0.7344	25.99/0.7832	30.37/0.9048
RFDN (ECCV 2020) [36]	4	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
HDRN (PR 2020) [45]	4	32.23/0.8960	28.58/0.7810	27.53/0.7370	26.09/0.7870	30.43/0.9080
SMSR (CVPR 2021) [33]	4	32.12/0.8932	28.55/0.7808	27.55/0.7351	26.11/0.7868	30.54/0.9085
MCSR (IG 2021) [16]	4	32.19/0.8941	28.63/0.7823	**27.58**/0.7362	26.04/0.7830	30.45/0.9039
ARRFN (NC 2022) [34]	4	32.22/0.8952	28.60/0.7817	27.57/0.7355	26.09/0.7858	30.57/0.9092
LBNet (IJCAI 2022) [35]	4	**32.29/0.8960**	**28.68/0.7832**	**27.62/0.7376**	**26.27/0.7904**	**30.76/0.9111**
Our MFRN	4	**32.31/0.8961**	**28.69/0.7841**	**27.62/0.7392**	**26.29/0.7911**	**30.78/0.9119**

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENT SETTINGS

The programming framework is Pytorch1.0 in Ubuntu18.04. The processor is Intel(R) Core (TM) i7-7800 with 32GB memory. The graphics card is GTX1080Ti and the CUDA version is 8.0. The optimizer is Adam, with its parameters set to empirical values, specially $\epsilon = 10^{-8}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Last but not least, the initial learning rate is set to 0.0001, with a reduction to half of the previous value every 200 epochs, and the number of the training epoch is 1000.

Our model is compared with sixteen state-of-the-art methods, and they are SRCNN [4], DRCN [20], LapSRN [14], EDSR [15], DRRN [17], CARN [31], RDRN [25], IMDN [32], DRUDN [19], RFDN [36], HDRN [43], SMSR [33], MCSR [16], ARRFN [34], LBNet [35], respectively. Their codes are available for free download from Github, and their default parameter settings are followed.

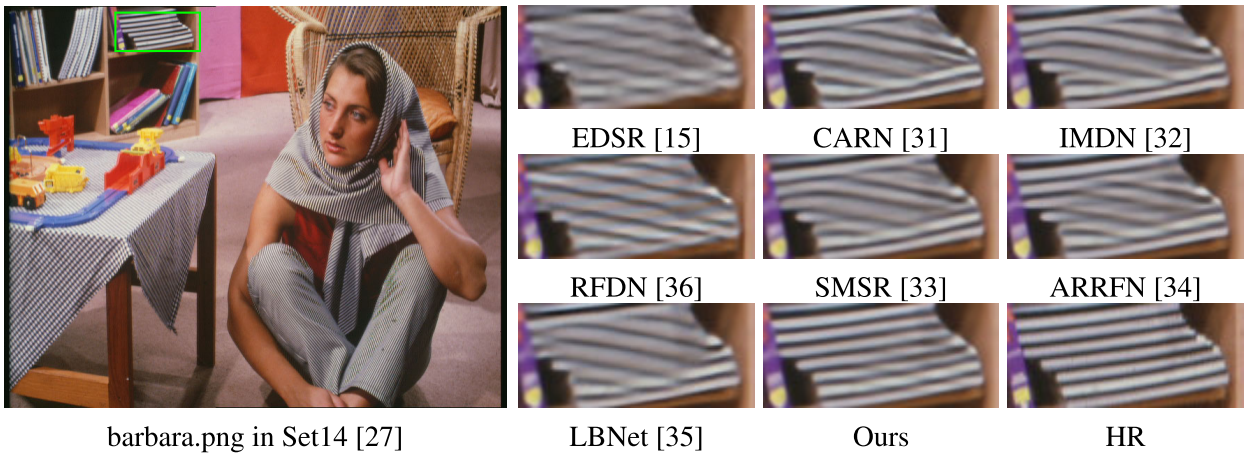


FIGURE 7. The green rectangle marks a 118 × 52 × 3 patch from barbara.png in the dataset Set14 [27]. On the right side are its ×4 super-resolution results, and the names of the corresponding super-resolution algorithms are marked below the results. HR is the high resolution patch, or known as the ground truth.

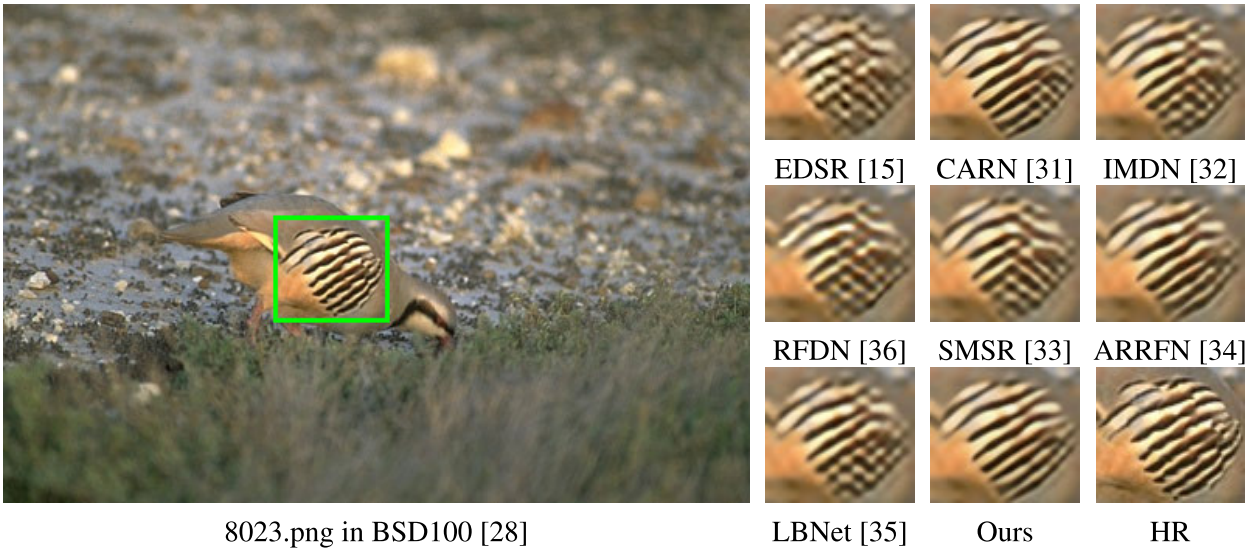


FIGURE 8. The green rectangle marks a 69 × 63 × 3 patch from 8023.png in the dataset BSD100 [28]. On the right side are its ×4 super-resolution results, and the names of the corresponding super-resolution algorithms are marked below the results. HR is the high resolution patch, or known as the ground truth.

B. DATASETS

The training dataset is DIV2K with data augmentations. The DIV2K is a high-quality 2K dataset that contains 800 training images, 100 validation images, and 100 testing images. Four public benchmark datasets, namely Set5 [26], Set14 [27], BSD100 [28], Urban100 [29] and Manga109 [38] are used to verify the training model. The use of the datasets is in accordance with internationally accepted guidelines, and the images in these datasets contain complex textures, challenging structures, and rich information in the frequency domain.

C. COMPARISONS WITH OTHER ALGORITHMS

1) PSNR AND SSIM COMPARISONS

In this study, Peak Signal-to-Noise Ratio (PSNR [11]) and Structural SIMilarity (SSIM [12]) are primarily employed to evaluate image reconstruction quality. PSNR is a widely used metric for image reconstruction evaluation, which measures pixel loss between two images. SSIM is also used to assess image quality and mainly measures the structural similarity between two images. In our objective quality assessment tests, only the Y-channel component of a single image is used, and the results are presented in Table 1. The PSNR and SSIM are calculated as shown in Eq. 10 and Eq. 11, where $m \times n$ represents the size of the images, specifically the ground truth image I_{gt} and the result image I'. μ_x, μ_y and δ_x^2, δ_y^2 are the mean and variance of I_{gt} and I' respectively, δ_{xy} is the covariance of I_{gt} and I', and c_1 and c_2 are very small constants and their values are set to 0.001 in the paper.

$$PSNR = 20log_{10}\frac{255 \times mn}{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{gt}(i,j) - I'(i,j))^2} \tag{10}$$

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\delta_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\delta_x^2 + \delta_y^2 + c_2)} \tag{11}$$

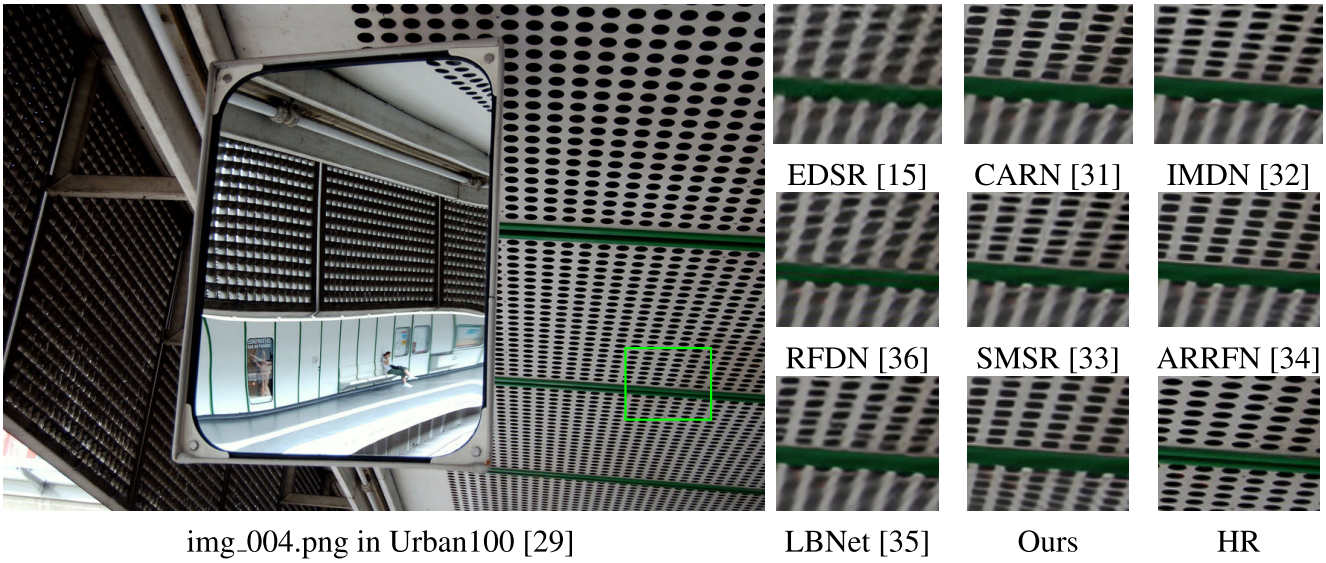


FIGURE 9. The green rectangle marks a 112 × 93 × 3 patch from img_004.png in the Urban100 [29]. On the right side are its ×4 super-resolution results, and the names of the corresponding super-resolution algorithms are marked below the results. HR is the High Resolution patch, or known as the ground truth.

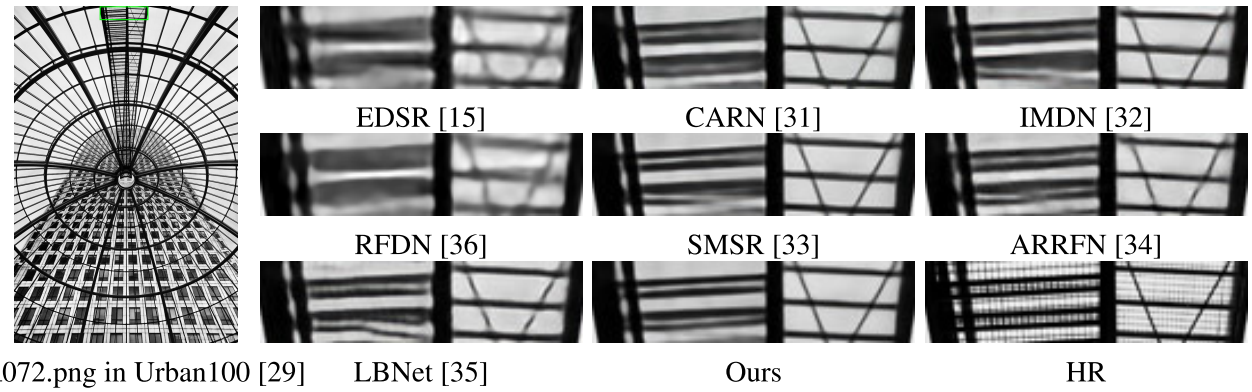


FIGURE 10. The green rectangle marks a 142 × 36 × 3 patch from the img_072. png in the Urban [29]. On the right side are its ×4 super-resolution results, and the names of the corresponding super-resolution algorithms are marked below the results. HR is the High Resolution patch, or known as the ground truth.

As shown in Table 1, our algorithm outperforms most of the other methods in terms of both PSNR and SSIM on the four commonly-used datasets. This indicates that our algorithm exhibits superior pixel recovery and structure preservation capabilities for low-resolution input images. Furthermore, as the texture complexity increases, from × 2 to × 4 and from the Set5 [26] to Urban100 [29], the advantages of our algorithm over other methods are further enhanced.

It is worth noting that when the magnification is 2, both local and global information play equally important roles in the image reconstruction process. The LBNet model, based on transformers, and the RFDN model with enhanced spatial attention demonstrate a stronger ability to describe global features. However, due to the limited receptive field, MFRN struggles to capture effective global features, leading to slightly weaker reconstruction performance. Conversely, as the magnification increases to 3 or 4, the importance of local information in the image reconstruction process becomes more prominent. The utilization of a multiplexed structure, along with channel attention and pixel attention, further enhances the representation of features. In such case, MFRN exhibits a greater capability to characterize spatial features, resulting in superior performance.

2) VISUAL PERFORMANCE COMPARISONS

The human eye is the primary means of obtaining information, and thus visual effect serves as a crucial measure of quality. Test results demonstrate that our proposed algorithmic model is more effective in preserving image texture without distortion. For instance, in Figure 7, while the orientation of the book edge is horizontal, only our results correctly maintain this orientation, while other algorithms erroneously alter it, which is clearly untenable. The same holds in Figure 8, where the feather texture of the chickadee is obliquely upward, and the reconstructed texture in other algorithms exhibits crossed patterns. It is worth emphasizing that our algorithmic model is robust even when dealing with

TABLE 2. MOS comparisons. Top 5 algorithms for scale factor × 2, × 3, ×4 on datasets Set5 [26], Set14 [27], BSD100 [28], Urban100 [29] and Manga109 [38]. Our model MFRN are shown in black bold.

Dataset	Scale	Top 5 algorithms
Set5 [26]	2	ARRFN >**MFRN** >LBNet >SMSR >MCSR
Set5 [26]	3	**MFRN** >ARRFN >LBNet >SMSR >IMDN
Set5 [26]	4	**MFRN** >LBNet >ARRFN >RFDN >MCSR
Set14 [27]	2	RFDN >**MFRN** >LBNet >ARRFN >SMSR
Set14 [27]	3	**MFRN** >LBNet >ARRFN >MCSR >RFDN
Set14 [27]	4	**MFRN** >LBNet >MCSR >ARRFN >RFDN
BSD100 [28]	2	**MFRN** >ARRFN >LBNet >MCSR >SMSR
BSD100 [28]	3	**MFRN** >LBNet >MCSR >SMSR >ARRFN
BSD100 [28]	4	**MFRN** >LBNet >MCSR >CARN >ARRFN
Urban100 [29]	2	**MFRN** >LBNet >ARRFN >MCSR >SMSR
Urban100 [29]	3	**MFRN** >MCSR >DRUDN >RDRN >EDSR
Urban100 [29]	4	**MFRN** >LBNet >SMSR >ARRFN >MCSR
Manga109 [38]	2	LBNet >**MFRN** >IMDN >RFDN >ARRFN
Manga109 [38]	3	**MFRN** >LBNet >SMSR >RFDN >ARRFN
Manga109 [38]	4	**MFRN** >LBNet >MCSR >RFDN >ARRFN

repetitive and dense textures. As illustrated in Figure 9, only our algorithm can accurately recover the texture below the green line, whereas in other models, this texture appears blurred. Likewise, the same holds in the building support in Figure 10, further substantiating the superiority and robustness of our algorithm.

3) MOS COMPARISONS

The Mean score Of the System (MOS) is a subjective score evaluation metric widely used in visual tasks on an international scale. This metric involves selecting individuals with and without professional backgrounds in proportion and requesting them to evaluate images provided to them for rating. The evaluation criterion relies solely on the comfort level of human eyes during the observation of these images. After the removal of extreme scores, the remaining scores are averaged in descending order to obtain the final result.

As depicted in Table 2, our algorithm demonstrates a top-ranking performance in most cases, and a second-place performance in some cases in the MOS test. This not only serves as evidence of our algorithm's ability to generate natural images with excellent subjective performance but also illustrates its robustness and proficiency in recovering most textures present in the datasets.

4) COMPARISONS WITH THE BENCHMARK MODEL IMDN [32]

IMDN [32] is regarded as the benchmark algorithm for SISR models, and its effectiveness is evaluated based on the quantified metrics of PSNR and SSIM. The histogram in Figure 11 illustrates the ΔPSNR or ΔSSIM values between our MFRN and IMDN. The results indicate that MFRN outperforms IMDN across all five commonly-used international texture datasets and exhibits superior performance on more complex texture datasets, such as BSD100 [28], Urban100 [29] and Manga109 [38].

5) COMPARISONS WITH THE RECURRENT-BASED METHODS

As our algorithm is a recurrent-based method, it is necessary to provide a brief overview of recurrent algorithms. Recurrent

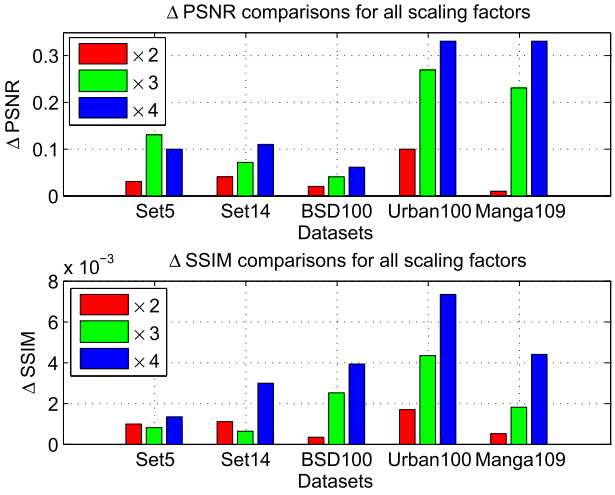


FIGURE 11. Comparisons with the model IMDN [32]. × 2, × 3 and × 4 tests in Set5 [26], Set14 [27], BSD100 [28], Urban100 [29] and Manga109 [38]. The objective evaluation metrics are ΔPSNR and ΔSSIM. The test results are presented as histograms with the names of datasets directly below them.

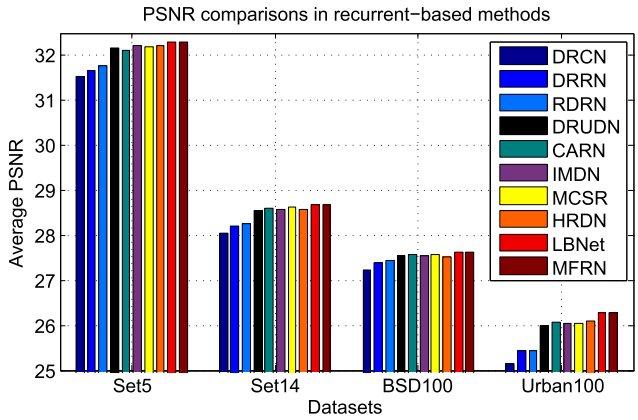


FIGURE 12. × 4 SISR performance comparisons of recurrent algorithms. The objective evaluation metric is PSNR, and Set5 [26], Set14 [27], BSD100 [28] and Urban100 [29] are test datasets.

algorithms are designed to reduce training time through module reuse and effectively alleviate the loss of features in the transmission process, ensuring the optimal utilization of information. This subsection presents comparisons of recurrent networks from two perspectives: network performance and structures. To make comparisons, our MFRN is evaluated against eight SISR recurrent networks, namely DRCN [20], DRRN [17], RDRN [25], DRUDN [19], CARN [31], IMDN [32], MCSR [16], HDRN [45] and LBNet [35]. Given the unavailability of data from the Manga109 [38] dataset for the CARN model, the comparisons are limited to utilizing results solely from the first four commonly-used datasets, as shown in Figure 12, with MFRN being the most stable and superior among them.

Figure 13 depicts a simplified version of eight recurrent networks that reuse the depth feature extraction module. The shallow feature extraction and reconstruction modules are represented in gray and pink, while the remaining modules form the recurrent network. Notably, Figure 13

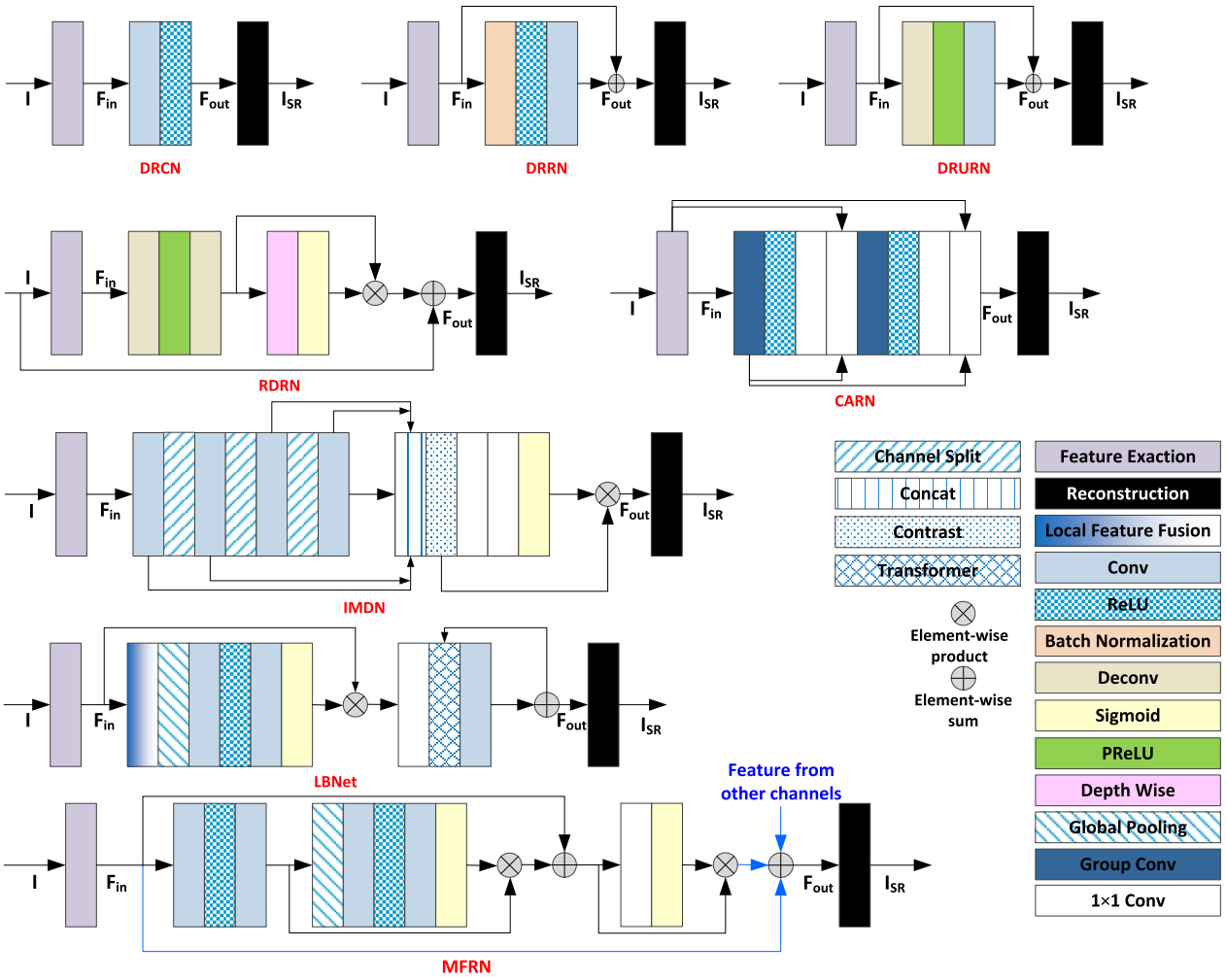


FIGURE 13. Simplified recurrent network structures of DRCN [20], DRRN [17], DRUDN [19],RDRN [25], CARN [31], IMDN [32], LBNet [35] and our model MFRN. All functional modules are color-coded and annotated. In particular, I is the input image, F_{in} is the shallow feature, F_{out} is the output of the recurrent network, and I_{SR} is the reconstruction result of I.

shows significant differences between the MFRN and other networks in three aspects. Firstly, MFRN stands out due to its adoption of a multiplexing-based structure, which is distinct from the single-channel structure used by other networks. Secondly, MFRN utilizes a different type and order of functional modules in each channel. Lastly, unlike other recurrent networks, MFRN employs an attention fusion mechanism to process information from each part differently. This mechanism enhances the feature expression of the neural network, making it the most significant point that sets MFRN apart from MCSR [16]. In conclusion, MFRN stands out from other recurrent networks due to its unique performance and network structure.

TABLE 3. Model parameters comparisons.

	SRCNN [4]	LapSRN [14]	EDSR [15]
PSNR	24.53dB	25.20dB	25.65dB
FLOP	52.7G	29.9G	316.2G
Running time	297ms	189ms	315ms
Parameter size	57K	815K	40.73M
	CARN [31]	IMDN [32]	SMSR [33]
PSNR	26.07dB	26.04dB	26.11dB
FLOP	222.8G	158.8G	351.5G
Running time	335ms	38ms	309ms
Parameter size	1.59M	0.7M	0.98M
	ARRFN [34]	LBNet [35]	Our MFRN
PSNR	26.09dB	26.27dB	26.28dB
FLOP	-	-	171.9G
Running time	234ms	679ms	361ms
Parameter size	0.98M	0.73M	0.86M

6) COMPARISONS OF MODEL PARAMETERS

Table 3 illustrates recent SISR models renowned for their exceptional performance, along with corresponding metrics such as FLOating-Point operations (FLOP, also known as Multi-adds), Running time, and Parameter size. The test images are × 4 SISR results in dataset Urban100 [29], and the statistics are the mean values of the test images. Among the different models used, SRCNN [4], LapSRN [14], CARN [31], IMDN [32], SMSR [33], ARRFN [34], LBNet [35] and our MFRN are lightweight models since they contain less than 1M parameters. However, the other models contain more than 1M parameters, especially EDSR [15], which

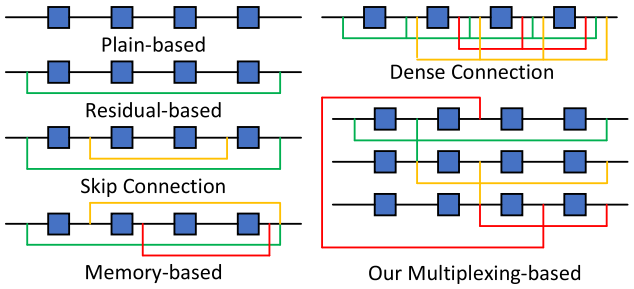


FIGURE 14. Topologies of information flow, and they are plain-based, residual -based [22], skip connection [24], memory-based [18], dense connection [30], and our multiplexing-based method.

has 47 times more parameters than MFRN. Furthermore, our MFRN not only ensures top-notch performance but also resides within a reasonable range in terms of speed and FLOP metrics. This substantiates the fact that the multiplexed structure successfully strikes a commendable balance between model performance and efficiency.

7) COMPARISONS OF TOPOLOGIES OF INFORMATION FLOW

Various connection methods exist between network modules. Considering each network module as a transmission node of information flow, the connection methods between nodes possess different topologies in space, leading to different forms of information computation and propagation, thereby forming distinct features.

In Figure 14, five common information flow topologies are illustrated. The blue square nodes represent information flow nodes, and the black lines depict the information transmission routes. Each topology employs unique forms of information reuse to generate diverse features. While Densenet [30] is a common method for information reuse in single-channel network topologies, it overlooks the potential of multiplexing-based feature fusion. Our approach uses information from multiple channels recursively, reducing training time while enhancing feature information capacity. Furthermore, the attention fusion mechanism effectively coordinates the contributions of each module in the network, leading to more realistic reconstructed images.

8) REAL WORLD COMPARISONS AND DISCUSSIONS

The intricate textures and fine details found in real-world images present a substantial challenge for SISR tasks. Real-world images are frequently affected by various forms of noise and artifacts, such as signal noise, compression artifacts, and motion blur, all of which have a detrimental impact on the quality of SISR results. Moreover, the acquisition of high-quality training data has emerged as a significant hurdle in real-world SISR tasks, primarily due to the associated costs and difficulties involved in obtaining HR images. These inherent challenges have served as strong motivation for researchers to strive towards the development of more effective and accurate SISR techniques.

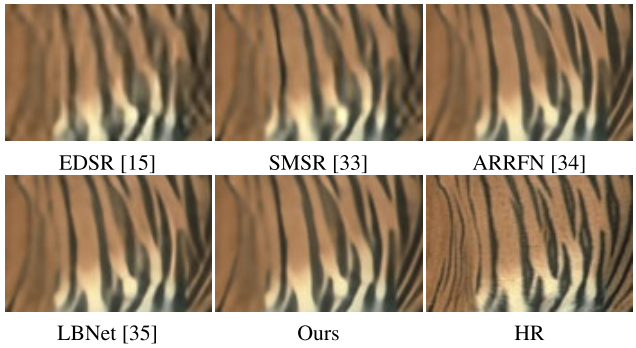


FIGURE 15. 1st visual performance comparisons. HR is the High Resolution patch from the real-world image, or known as the ground truth. Other images are ×4 SISR results, with the name of the algorithm marked at the bottom of each image.

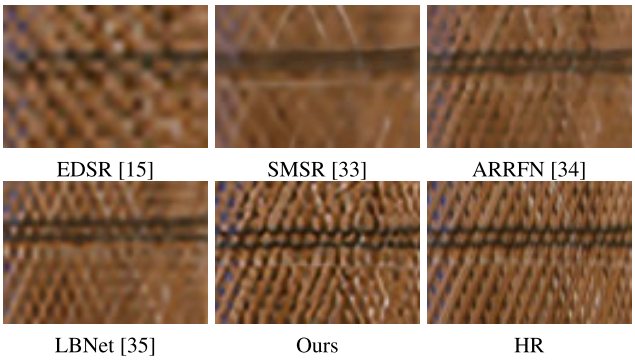


FIGURE 16. 2nd visual performance comparisons. HR is the high resolution patch from the real-world image, or known as the ground truth. Other images are ×4 SISR results, with the name of the algorithm marked at the bottom of each image.

By analyzing Figures 15 and 16, it can be inferred that our proposed SISR algorithm, namely MFRN, demonstrates remarkable performance in real-world image processing. Firstly, MFRN exhibits proficiency in capturing and restoring intricate textures and details, thereby preserving the authenticity of the images. This proficiency is clearly evidenced by the precise rendering of the tiger's fur in Figure 15 and the intricate texture of the cooler in Figure 16. Secondly, MFRN upholds exceptional efficiency and accuracy while processing large-scale images, successfully executing complex computational operations within a reasonable timeframe. As a result, MFRN proves to be well-suited for effectively handling real-world images.

D. ABLATION STUDY

Our model MFRN consists of three key components: the shallow feature extraction, nonlinear feature mapping and reconstruction parts. Omitting any of these components would render the network unable to complete the SISR task. Consequently, ablation studies are conducted to evaluate the impact of the Recurrent Architechture (RA), MultiPlexing-based structure (MP) and the Attention Fusion mechanism (AF) on the overall network performance.

Table 4 illustrates three cases that we evaluated by modifying the structure of channels and incorporating an

TABLE 4. Ablation studies of the network parts of MFRN.

Case	RA	MP	AF	ΔPSNR	ΔSSIM	ΔFLOP	Δ Running time
1	✓	×	×	−0.96dB	−0.0400	−71.40G	+79ms
2	×	✓	×	−0.12dB	−0.0050	−15.11G	+20ms
3	✓	×	✓	−0.33dB	−0.0140	−44.99G	+52ms

TABLE 5. Ablation studies of structurally-similar methods with MFRN, and the test images are × 4 SISR results in dataset Urban100 [29]. MC, MP, CA, PA, AF are short for multiple channels, multiplexing-based structure, channel attention, pixel-attention, and attention fusion.

Method	MC	MP	CA	PA	AF	PSNR	SSIM
CARN [31]	✓	×	×	×	×	26.07dB	0.7837
RDRN [25]	✓	×	×	×	×	25.43dB	0.7633
ARRFN [34]	✓	×	×	×	×	26.09dB	0.7858
RFDN [36]	✓	×	×	×	×	26.11dB	0.7858
DRUDN [19]	✓	×	×	×	×	25.99dB	0.7832
IMDN [32]	✓	×	✓	×	×	26.04dB	0.7838
MCSR [16]	✓	×	✓	×	×	26.04dB	0.7830
SMSR [33]	✓	×	×	✓	×	26.11dB	0.7868
LBNet [35]	✓	×	✓	×	✓	26.27dB	0.7904
Our MFRN	✓	✓	✓	✓	✓	26.29dB	0.7911

attention fusion mechanism. The test results of these cases demonstrate variations in PSNR, SSIM, FLOP (also known as Multi-adds), and running time, highlighting the essentiality of both a multiplexing-based structure and an attention fusion mechanism for accurate feature extraction. The test images are × 4 SISR results in dataset Urban100 [29], and the statistics are the mean values of the test images.

Table 5 effectively illustrates several models that have a similar structure to our MFRN. Notably, CARN [31], RDRN [25], AFFRN [34], RFDN [36], DRUDN [19] and other models leverage multiple channel structures for optimizing information transmission, but they all ignore to fuse information between multiplexes. In comparison, our MFRN employs a multiplexed-based information flow transmission structure and finally generates superior features. In addition, these models are not capable enough of distinguishing features due to the lack of an effective attention mechanism. For example, IMDN [32], MCSR [16], LBNet [35], and other models utilize channel attention mechanisms to learn differences in information across channels. However, they do not pay adequate attention to pixel-level features within each channel. Furthermore, SMSR [33] prioritizes the sparsity of each feature but indiscriminately treats features of different channels, which is unreasonable. Our MFRN, however, uses channel and pixel attention mechanisms to learn better features globally and locally. Besides, multiplexed-based architectures and feature fusion mechanisms are explored to process these features efficiently, and global arithmetic power is allocated reasonably to achieve better output results.

Classical attention mechanisms encompass channel attention, pixel attention, and spatial attention. These mechanisms can be utilized individually or in combination. For instance, in image classification, the CBAM model [50] integrates channel attention and spatial attention. However, many attention studies predominantly focus on high-level visual problems, while disparities may arise in low-level problems such as super-resolution. In essence, the performance of the same attention module can vary between low-level and high-level problems. This discrepancy arises because high-level problems emphasize image semantics, whereas low-level problems focus on individual pixel values.

TABLE 6. Ablation studies of attention mechanisms. The test images are × 4 SISR results in dataset Urban100 [29], and the statistics are the mean values of the test images. CA, PA, SA are short for Channel Attention, Pixel Attention, and Spatial Attention.

Case	CA	PA	SA	ΔPSNR	ΔSSIM	ΔFLOP	Δ Running time
1	✓	✓	×	-	-	-	-
2	✓	×	✓	−0.02dB	−0.0007	+23.62G	+31ms
3	✓	✓	✓	+0.01dB	+0.0002	+36.75G	+57ms

Table 6 illustrates three cases: case 1 represents the model proposed in this study, case 2 replaces pixel attention with spatial attention, i.e., the CBAM model, and case 3 incorporates all three types of attention simultaneously. Statistical data demonstrates that in case 2 compared to case 1, image reconstruction performance declines, computational complexity increases, and inference time rises. Case 3 only exhibits a marginal performance improvement of 0.038% over case 1, which is quite limited, while simultaneously increasing running time by 15.79% and FLOP by 21.27%. Consequently, the overall system performance decreases. Considering the trade-off between performance and efficiency, this study adopts the fusion of channel attention and pixel attention.

The underlying reason for this phenomenon is that the proposed multiplexed structure in this study already offers spatial features of ample richness. Consequently, the simultaneous utilization of the multiplexed structure and spatial attention would lead to redundant spatial information, which hampers the improvement of system performance. Furthermore, the coarse nature of spatial features poses challenges in recovering fine-grained textures. However, with the aid of pixel attention, the system can effectively prioritize pixel-level features, resulting in exceptional image reconstruction performance.

IV. CONCLUSION AND FUTURE WORK

In this study, a Multi-attention Fusion Recurrent Network, namely MFRN, is proposed. Compared with most SISR algorithms, this paper adopts a multiplexing-based architecture to improve the information representation ability and reduce the loss rate of features. At the same time, the recurrent reuse of neural networks greatly reduces the training time. Finally, it is worth mentioning that the attention fusion mechanisms, including the channel and pixel attention mechanisms, help to process the feature information of each layer differently. Numbers of experiments show that the proposed SISR algorithm can not only output visually nature-looking reconstructed images but also achieve remarkable results in the test of objective metrics.

We acknowledge that the model proposed in this study primarily focuses on local features while overlooking the crucial role of global features. Additionally, it remains uncertain whether there exists an optimal geometric topology in space for multiplexed architectures. Regrettably, the current study lacks theoretical investigations in this specific domain. Consequently, both of these aspects represent essential areas for future research, with the goal of expanding and refining the model.

V. DECLARATION OF COMPETING INTEREST

The authors declare that they are not aware of the possibility of competing for financial interests or personal relationships affecting the work reported in this study.

REFERENCES

[1] Q. Li, L. Qiu, B. Qi, and G. Liang, "Adaptive weighting for estimation of the mean of the merged measurement for multi-target bearing tracking," *Electron. Lett.*, vol. 57, no. 10, pp. 412–414, May 2021.

[2] C. Zhou, Y. Zhou, Z. Suo, and Z. Li, "Voxel area sculpturing-based 3D scene reconstruction from single-pass CSAR data," *Electron. Lett.*, vol. 56, no. 11, pp. 566–567, May 2020.

[3] S. Peled and Y. Yeshurun, "Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging," *Magn. Reson. Med.*, vol. 45, no. 1, pp. 29–35, Jan. 2001.

[4] 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.

[5] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 391–407.

[6] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.

[7] H. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 6, pp. 508–517, Dec. 1978.

[8] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.

[9] K. Xu, J. L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 2048–2057.

[10] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 294–310.

[11] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

[12] 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.

[13] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. 12th Asian Conf. Comput. Vis. (ACCV)*, Singapore, 2015, pp. 111–126.

[14] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.

[15] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[16] D. Cheng, X. Guo, L. Chen, Q. Kou, K. Zhao, and R. Gao, "Image super-resolution reconstruction from multi-channel recursive residual network," *J. Image Graph.*, vol. 26, no. 3, pp. 605–618, 2021.

[17] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.

[18] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4549–4557.

[19] Z. Li, Q. Li, W. Wu, J. Yang, Z. Li, and X. Yang, "Deep recursive up-down sampling networks for single image super-resolution," *Neurocomputing*, vol. 398, pp. 377–388, Jul. 2020.

[20] 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. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1637–1645.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Mach. Learn. (ICML)*. Madison, WI, USA: Omnipress, 2010, pp. 807–814.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[23] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 723–731.

[24] X. Mao, C. Shen, and Y. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, Jan. 2016, pp. 1–9.

[25] F. Li, H. Bai, and Y. Zhao, "Detail-preserving image super-resolution via recursively dilated residual network," *Neurocomputing*, vol. 358, pp. 285–293, Sep. 2019.

[26] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, Surrey, U.K., Sep. 2012, pp. 135.1–135.10.

[27] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Cof. Curves Surf.*, Avignon, France, Jun. 2010, pp. 711–730.

[28] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[29] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5197–5206.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[31] N. Ahn, B. Kang, and K. A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 256–272.

[32] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2024–2032.

[33] L. Wang, X. Dong, Y. Wang, X. Ying, Z. Lin, W. An, and Y. Guo, "Exploring sparsity in image super-resolution for efficient inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 4915–4924.

[34] J. Qin and R. Zhang, "Lightweight single image super-resolution with attentive residual refinement network," *Neurocomputing*, vol. 500, pp. 846–855, Aug. 2022.

[35] G. Gao, Z. Wang, J. Li, W. Li, Y. Yu, and T. Zeng, "Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive transformer," in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 1–8.

[36] L. Jie, J. Tang, and G. Wu, "Residual feature distillation network for lightweight image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 41–55.

[37] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 56–72.

[38] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using Manga109 dataset," *Multimedia Tools Appl.*, vol. 76, pp. 21811–21838, Nov. 2017.

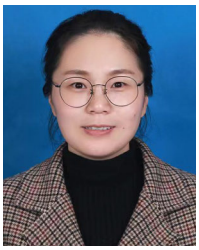
[39] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.

[40] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5610819.
[41] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, and L. Zhang, "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, pp. 297–311, Aug. 2023.
[42] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, "Dual-path deep fusion network for face image hallucination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 378–391, Jan. 2022.
[43] M.-I. Georgescu, R. T. Ionescu, A.-I. Miron, O. Savencu, N.-C. Ristea, N. Verga, and F. S. Khan, "Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2023, pp. 2194–2204.
[44] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2734–2747, Oct. 2020.
[45] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475.
[46] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509417.
[47] K. Zheng, L. Gao, D. Hong, B. Zhang, and J. Chanussot, "NonRegSRNet: A nonrigid registration hyperspectral super-resolution network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520216.
[48] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.
[49] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.
[50] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.

QIQI KOU received the B.S. and M.S. degrees from the Anhui University of Science and Technology, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Information and Control Engineering, China University of Mining and Technology, in 2019. He is currently a Lecturer with the School of Computer Science and Technology, China University of Mining and Technology. His research interests include image processing, computer vision, and pattern recognition.

DEQIANG CHENG (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and information engineering from the China University of Mining and Technology. He is currently a Professor with the China University of Mining and Technology. His research interests include machine learning, video coding, image processing, and pattern recognition.

HAOXIANG ZHANG (Graduate Student Member, IEEE) received the B.S. degree from the University of Electronic Science and Technology, in 2018, and the M.S. degree from the China University of Mining and Technology, in 2021, where he is currently pursuing the Ph.D. degree. His research interests include image retrieval and super resolution algorithms.



JINGJING LIU received the master's degree in electrical engineering from the China University of Mining and Technology, in 2013, where she is currently pursuing the Ph.D. degree with the School of Information and Control Engineering. Her research interests include image processing and pattern recognition.

XIN GUO received the B.S. degree from Anhui University, in 2018, and the M.S. degree from the China University of Mining and Technology, in 2022. He is currently with the Huawei Hangzhou Research Institute, Hangzhou. His research interest includes single image super-resolution reconstruction.

HE JIANG received the B.S. degree in telecommunication engineering from the Nanjing University of Posts and Telecommunications and the M.S. degree in telecommunication engineering and the Ph.D. degree in control science and engineering from Shanghai Jiao Tong University, in 2021. He is currently a Lecturer with the School of Information and Control Engineering, China University of Mining and Technology. His main research interests include machine learning and deep learning-based low-level vision tasks, such as super resolution, detail enhancement, video frame interpolation, and de-noising.