IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 October 2023, accepted 19 November 2023, date of publication 21 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335382

RESEARCH ARTICLE

Feature Preserving and Enhancing Network for Image Super-Resolution

MINGLAN SU[1], XINCHI LI[1], JIAOYANG XU[2], MINGCHUAN YANG[1], AND CHAOYING ZHANG[1]
[1]Institute of Big Data and Artificial Intelligence, China Telecom Corporation Ltd. Beijing Research Institute, Beijing 102209, China
[2]Beijing Normal University, Beijing 100091, China

Corresponding author: Minglan Su (suml@chinatelecom.cn)

ABSTRACT Single image super-resolution (SISR) with deep convolutional neural networks has recently attracted increasing attention due to its potentials to generate rich details. To obtain better fidelity and visual quality, most of existing methods are of heavy design with the depth of network. However, the lack of high-frequency information in the deep network and the limit of non-local context relations of pixels remain the core issue. In addition, there are many arbitrary-scale image super-resolution tasks, so it is a focus that provides an effective and efficient model for super resolution with arbitrary-scale. To this end, we propose a Feature Preserving and Enhancing Network (FPEN) based on implicit representation, which is aims to preserving high-frequency information in deep network and enhancing non-local contextual features of pixels to output more realistic images at arbitrary scales. In particular, our proposed High Frequency Preserving Block (HFPB) can divide the features into high-frequency components and low-frequency components, and allocate more operations to high-frequency components to ensure that high-frequency components can be preserved in the network. Since high-frequency information contains the details and texture of the image, it can restore the finer details of the image. Moreover, Pixel Continuity Attention (PiCA) module our proposed utilizes visual cues observed from pixels to adaptively recalibrate the pixel range that needs attention in the image to better achieve feature enhancement and generate smoother images. Extensive experiments conducted on benchmark SISR models and datasets show that Feature Preserving and Enhancing Network can be employed for various SISR tasks with arbitrary-scale to obtain the better visual quality than other state-of-the-art SR algorithms.

INDEX TERMS Image super resolution with arbitrary-scale, high frequency preserving single methods, non-local attention methods, implicit neural representation.

I. INTRODUCTION

Single image super resolution reconstruction (SISR) is an important research topic [1], whose goal is to reconstruct high-resolution images from low-resolution images as shown in Figure 1. SISR has been widely used in many fields, including medical imaging [2], image compression [3], small object detection [4], etc. However, SISR is an ill-posed problem. Due to the information loss in low resolution images, there may be multiple high resolution images corresponding to the same low resolution image. This ambiguity makes the SISR problem uncertain. Currently, many SR methods have been proposed to solve this problem, and these methods are mainly of three types: interpolation-based methods, reconstruction-based methods, and learning-based methods [5], [6]. Among them, methods based on deep learning have greatly improved the performance of SISR [7], [8], [9], [10].

Super resolution convolutional neural wetwork(SRCNN) is one of the earliest methods to introduce CNN for super resolution reconstruction [11]. It learns the mapping from low resolution images to high resolution images through a deep convolutional neural network, and the reconstruction effect far exceeds traditional algorithms. The powerful feature representation and end-to-end training mode of CNN make it an effective method for SISR. Subsequently, in order to further explore the impact of deep features on super

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

FIGURE 1. The visual comparison of SR results by our menthods. The top is the HR, the middle is the LR(x30), and the bottom is the SR(our method).

resolution reconstruction, researchers introduced residual learning [12], [13] to solve the gradient disappearance and degradation problems in deep network training and utilized dense connections [7], [14] increase the depth and width of the network. Although existing CNN-based methods have achieved good reconstruction results, high-resolution images still suffer from varying degrees of detail blurring and super-resolution reconstruction with arbitrary-scale cannot be achieved.

According to Fourier Transform, natural images are composed of different frequency signals [15]. The low frequency signal consists of global structure and can be forwarded directly to the final high resolution image without extensive computation. High frequency information consists of fine details. As the network deepens, high frequency information is gradually lost during network transmission. Therefore, how to ensure the transmission of high frequency information in the network has become one of the key factors in restoring and reconstructing details. Previous work introduced attention mechanisms [16], [17] into super resolution reconstruction to simulate spatial positions, channels, or the interdependence between the two to solve this problem. Recent frequency-based methods [18] use more complex convolution operations to preserve high frequency information in the network. Although these methods achieve good reconstruction results to a certain extent, the attention-based method does not distinguish between low frequency and high frequency features, while the frequency-based method simply uses the residual branch as high frequency information. Therefore, how to accurately estimate high frequency information and sustainably transmit high frequency information is still a challenge today.

In addition, there are certain correlations and similarities between pixels in the image, such as similar textures, structures or edges. Therefore, some scholars introduce non-local features [19]. Non-local features refer to the correlation between pixels at different locations in the image, which can provide more contextual information, allowing the algorithm to better understand the structure and texture in the image, thereby better restoring details and edges. Although non-local block reconstruction works well, the attention map computed by non-local blocks is often unstable, which means that it is easy to focus on less relevant pixel locations (such as locations with very large pixel color changes) and ignore pixel continuity during reconstruction.

To solve these problems, we propose a novel Feature Preserving and Enhancing Network (FPEN). It is worth noting FPEN is a hybrid architecture that uses a pattern of feature extraction and implicit reconstruction to perform super resolution reconstruction at arbitrary scale. In order to better convey high frequency features and achieve accurate texture detail reconstruction, we introduced the High Frequency Preserving Block (HFPB) in the feature extraction module to estimate the high frequency information in the feature map. Secondly, we propose a pixel continuous attention mechanism (PiCA) in the implicit reconstruction module, which extends non-local operations in two ways. First, a normalized self-similarity matrix (NSSM) is used to capture the similarity and dissimilarity structures in surrounding pixels to reveal the continuity of pixels. Second, we treat NSSM as prior knowledge and combine it with the attention map generated by non-local operations to adaptively recalibrate the range of attention needed in surrounding pixels.

Based on implicit reconstruction, we propose a Feature Preserving and Enhancing Network (FPEN). Our network achieves significant improvements over single image super resolution, and produces more competitive SR results. In summary, these are the main contributions of this paper:

- We propose a High Frequency Preserving Block (HFPB), which can estimate high frequency information more accurately. While retaining low frequency information, we perform more complex operations on high-frequency information so that high-frequency information can be propagated in the deep network to generate images with richer details and improve reconstruction quality.
- We propose a Pixel Continuous Attention mechanism (PiCA), which extends non-local operations and adaptively recalibrates the range that needs attention in pixels to captures pixel continuity dependencies. It delivers richer elements for the reconstruction of HR images, and allows the network to focus for more information features and outputing smoother images.
- Combined with implicit reconstruction, we propose the Feature Preserving and Enhancing Network (FPEN),

which can achieve super-resolution reconstruction at arbitrary-scale. Extensive experiments on a variety of public datasets demonstrate that the proposed architecture outperforms state-of-the-art models in terms of quantitative and visual quality.

II. RELATED WORKS

This section provides a brief overview of recent work on various super-resolution methods, non-local patterns, and implicit neural representations relevant to our work.

A. CNN-BASED SUPER RESOLUTION RECONSTRUCTION

Super resolution reconstruction aims to recover high resolution details and sharpness from low resolution input images.Traditional super-resolution reconstruction techniques include interpolation-based methods, edge-based methods, and statistical modeling-based methods. These methods can improve the resolution of images to a certain extent, but are generally less effective at reconstructing details and textures than CNN-based methods.

The first super-resolution reconstruction model based on CNN is superResolution convolutional neural network (SRCNN) proposed by Chao Dong et al., which learns the mapping function relationship from low resolution images to high resolution images by training a deep convolutional neural network [11]. Subsequently, the FSRCNN model has significantly improved performance and speed by fine-tuning the SRCNN structure [20]. The CPDS model proposed by Sedighe adopts a deep and shallow convolutional neural network framework for single image super resolution, achieving state-of-the-art performance with a small sample set and low computational cost [21]. Yu et al. proposed a multi-stage lightweight network enhancement method by using the enhanced high-resolution output as additional supervision to improve the performance of image super-resolution while achieving faster inference time [22]. The SRNO model proposed by Min et al. in 2023 achieved the super-resolution reconstruction task in a semi-supervised manner through a small number of LR-HR samples, which takes into account both generalization and efficiency [23].

Inspired by the Residual Network [24], Kim et al. proposed the Very Deep Super-Resolution (VDSR) model based on the deep residual network, which achieved better results in retaining detailed information and enhancing texture, and also has a significant improvement in calculation speed [12]. Subsequently, Enhanced Deep Super-Resolution (EDSR) proposed by Lim et al. removed the redundant batch normalization layer based on the VDSR model, and maked the model more compact and achieving better results [7]. The Multi-scale Feature Fusion Residual (MSFFR) model constructs residual blocks of multiple intertwined paths to adaptively detect and fuse image features of different scales, and achieves remarkable results [25].

In addition, the CNN-based super-resolution reconstruction models still have good results in three-dimensional super-resolution reconstruction tasks. Jinglong et al. proposed the mDCSRN model using a 3D convolutional neural network, which achieved outstanding results in medical MRI imaging data [26]. The NVSR model performs super-resolution reconstruction of multi-view consistent views of invisible three-dimensional scenes [27].

Overall, with the continuous development of deep learning technology, super-resolution reconstruction methods will achieve better results in retaining image details and textures, and will play a more important role in high-quality image generation, video processing, medical images and other fields.

B. ATTENTION MECHANISM

The core goal of the attention mechanism is to select information that is more critical to the current task from numerous information. As for deep learning, attention mechanisms are usually applied to the processing of sequence data. Attention mechanisms are roughly divided into self-attention [28], spatial attention [29] and temporal attention [30] mechanisms according to different weighted targets.

Self-attention is the most widely used attention mechanism, which allows each element to adjust its representation according to the importance of other elements, thereby capturing long-range dependencies between elements. The Transformer model based on multi-head self-attention has attracted the attention of a large number of scholars due to its extremely high versatility and ability to capture long-range correlations [31], [32] and has also achieved good results in super resolution reconstruction. Subsequently, Wang et al. [33]showed that self-attention is an instantiation of non-local mean, and proposed a non-local block for the CNN to capture long-range dependencies.

The non-local attention mechanism can establish global non-local correlations in feature maps, and captures the dependencies between distant pixels. Its idea is to obtain global contextual information by calculating the similarity between each position in feature map with all other positions. The application fields of non-local attention mechanism in image tasks include text recognition [34], image segmentation [35], image super-resolution reconstruction [36], etc. In super-resolution tasks, non-local attention mechanisms can be used to capture long-range dependencies within images. When processing images, there are often regions that span large spatial distances but have obvious correlations, such as similar textures in the image. Non-local attention can help the model search for these regions with similar features in the entire image, and better recover texture and structural details. Because it is important to recover as much high-frequency detail as possible during the conversion of low-resolution images to high-resolution images. Mei et al. combine non-local operations and sparse representation to ensure the super-resolution reconstruction effect while reducing the amount of calculation [36]. Reference [37] based on non-local operations, directly extracted long-range

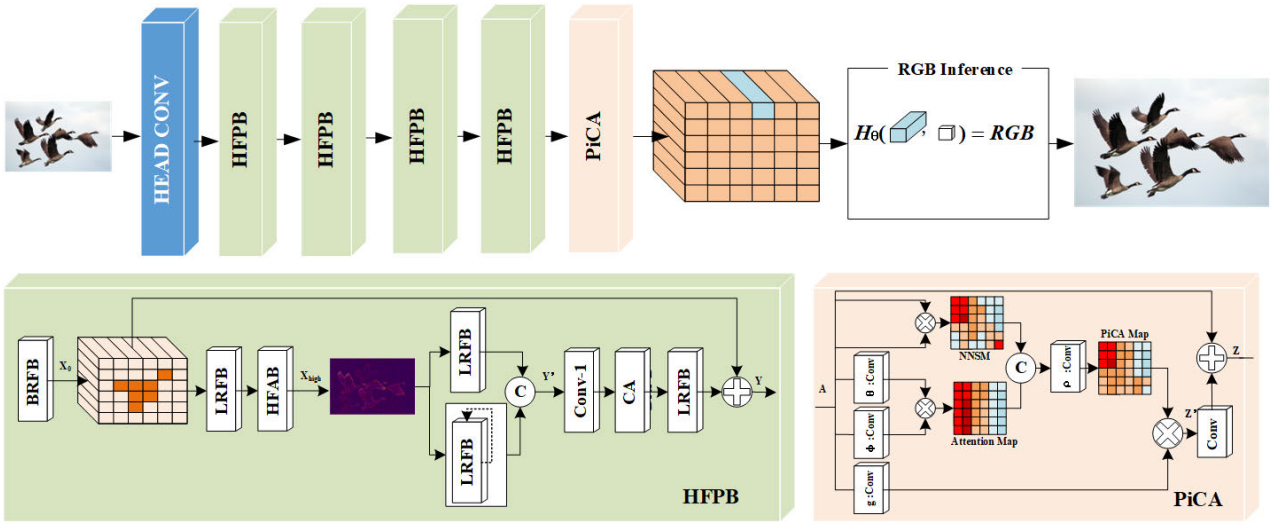

FIGURE 2. Proposed Feature Preserving and Enhancing Network (FPEN) for SISR, which consists of feature extraction module and implicit reconstruction modules. H_θ is an implicit function to predict the set of RGB colors of pixels.

spatio-temporal correlation, replacing traditional motion estimation and motion compensation to improve video super-resolution reconstruction.

C. IMPLICIT NEURAL REPRESENTATION

Implicit representation is a method of representing data by learning implicit functions, which map high-dimensional inputs to low-dimensional space. By learning implicit functions, it can achieve efficient encoding and generation of data. Implicit representations are often modeled through neural networks. These models can be trained to map input data to a low-dimensional representation in a latent space and then reconstruct the data from this low-dimensional representation. Implicit representation has been widely used in many fields.

Implicit representations can play an important role in three-dimensional shape construction during the 3D modeling process. By learning to map point cloud or voxel data to implicit functions, 3D shape reconstruction and generation can be achieved. For example, Mescheder et al. proposed a method based on implicit representation in 2019 for reconstructing the geometry of three-dimensional objects from point cloud data [38]. In recent years, some methods such as union implicit function (UNIF) [39] and deep implicit function (DIF) [40] have improved the accuracy of implicit functions to a certain extent.

In addition, implicit representation can also play an excellent role in super resolution reconstruction. Arbitrary super-resolution reconstruction of images can be achieved by learning implicit functions that map low resolution images to high resolution images. Local implicit image function (LIIF) explores the image super-resolution reconstruction from another angle and opens a new era of image super-resolution [41]. Li et al. proposed an adaptive local image function to improve structural distortion and ringing artifacts around edges during image reconstruction [42]. McGinnis et al. proposed a method using implicit neural representation (INR) to achieve high-fidelity anatomical reconstruction [43]. MoTIF proposed by Chen et al. uses local implicit neural functions to achieve continuous spatio-temporal video super-resolution reconstruction [44].

Taken together, the super-resolution algorithms have achieved significant performance improvements by using deep learning models to learn the mapping relationship of images, and combining attention mechanisms and other technical means. These algorithms also provide important theoretical and practical foundations for image enhancement, visual recognition and other fields.

III. PROPOSED METHODS

In this section, we first describe the overall network architecture. Next, we introduce the proposed High Frequency Preserving Block (HFPB) in detail. Finally, we discuss the proposed Pixel Continuous Attention module (PiCA).

A. NETWORK OVERVIEW

Taking a low-resolution image as an input, our goal is to reconstruct a high-resolution image with arbitrary scale. To achieve this goal, we propose Feature Preserving and Enhancing Network (FPEN) in Figure 2. Let us define I_{LR} and I_{SR} as the input and output of the network respectively.

First, we extract the input feature F_0 through one 3×3 convolutional layer(H_0) :

$$F_0 = H_0(I_{LR}). \tag{1}$$

Then the obtained image features F_0 is used as the input of several stacked HFPBs to estimate and preserve high frequency features. It can be formulated as

$$F_k = H_k(F_{k-1}), k = 1, \ldots, m. \tag{2}$$

where H_k represents mapping function of the k-th HFPB. F_k denotes the feature from the previous HFPB, and m is the total number of HFPBs.

The feature block F_m can be regarded as a matrix composed of several pixel feature vectors, so it can be reshaped into a sequence of pixel features, which is defined as $A = \{\alpha_i\}_{i=1}^{P}$. α_i represents the feature vector of the pixel and P is the total numble of pixels. Then the extracted feature A is as an input of the proposed PiCA module to calculate non-local feature representation $Z = \{z_i\}_{i=1}^{P}$. The operation is formulated as

$$Z = H_{PiCA}(A). \tag{3}$$

In the implicit reconstruction module, the RGB value of every pixel $q(i,j)$ in the I_{SR} can be obtained by fusing pixel values of its closest key point $k(i',j')$. Therefore, in order to achieve super resolution reconstruction with arbitrary scale, we define H_θ (θ is a parameter) as an implicit function to predict the set of RGB colors of pixels, which can be shown as:

$$I_{SR}(x_q) = H_\theta(z^*, (x_q - v^*, c)). \tag{4}$$

Among them, H_θ is a mapping function, which can be composed of MLP. z^* is the nearest latent code from x_q which is obtained by projecting the high-resolution pixel point into the low-resolution space. v^* is the coordinate of latent code z^* in the I_{LR}. c is the conditions of the query pixel q in I_{SR}.

Here we set v_t^* ($t \in \{00, 01, 10, 11\}$) to be the four nearest points in top-left, top-right, bottom-left, bottom-right of x_q, and z_t^* is the nearest latent code of v_t^*. We extend Equation (4) to:

$$I_{SR}(x_q) = \sum_{t\in\{00,01,10,11\}} \frac{S_t}{S} \cdot H_\theta(z_t^*, (x_q - v_t^*, c)). \tag{5}$$

where S_t is the area of the rectangle between x_q and v_t^*, and $S = \sum_t S_t$ is the total area of the four rectangles.

In order to achieve more accurate super resolution reconstruction, we use L_1 loss for training. The training set contains N pairs of I_{LR} and corresponding I_{HR}. The network is optimized to minimize the L_1 loss funxtion:

$$L_1 = \frac{1}{N}\sum_{i=1}^{N} \|I_{SR} - I_{HR}\|. \tag{6}$$

B. HIGH FREQUENCY PRESERVING BLOCK

According to Fourier Transform, natural images contain low-frequency signals that describe the image structure and high-frequency signals that describe detailed textures. We assume that the output feature map from the previous convolutional layer also contains low frequency component and high frequency component. High frequency information is seriously lost during the down-sampling process, and recovering high frequency information requires more complex feature processing. In order to preserve as much high-frequency detail information as possible in the network, it is very important to assign the feature to the appropriate branch according to the frequency strength, so that the high frequency signal can be further enhanced through more complex operations, while the low frequency part is passed through the lower complexity operation to compensate for the increased calculations. To this end, we proposed High Frequency Preserving Block(HFPB) to achieve more accurate reconstruction by learning discriminants to separate high frequency and low frequency information, as shown in Figure 2.

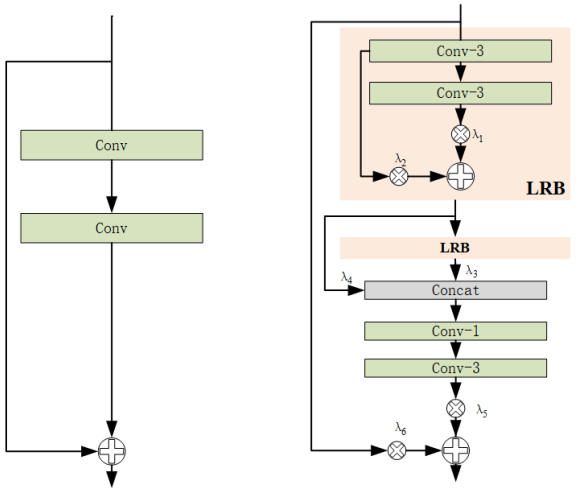

FIGURE 3. The right is architecture of the Basic Residual Feature Block (BRFB), and the left is architecture of the Learnable Residual Feature Block (LRFB).

Given the input $X \in \mathbb{R}^{C\times H\times W}$, where C is the number of channels, and $H \times W$ is denoted spatial demensions, HFPB can be modeled as

$$Y = X_0 + \mathcal{F}_{res}(X_0). \tag{7}$$

where X_0=$\mathcal{F}_B(X)$, $\mathcal{F}_B$ represents the Basic Residual Feature Block(BRFB).$\mathcal{F}_{res}$ uses residual learning to matigate the gradient vanishing problem and enhance the representation ability of the model.

In the residual branch, we first extract the features of X_0 based on the Learnable Residual Feature Block(LRFB). High Frequency Adaptive Block(HFAB) is applied to estimate the high frequency information X_{high} from the low frequency space. Then We enhance high frequency information through two branches, where both branches contain a LRFB. For the purpose of exploring feature representations of different sizes, we reduce the feature size in the lower branch and deepen the network by sharing weights. After the feature extraction, the features are upsampled, and the upper and lower branch features are mixed through the concate operation to obtain the feature Y'. Its operation can be expressed as

$$Y' = [\mathcal{F}_L(X_{high}), \uparrow \mathcal{F}_L^{\circlearrowleft n}(\downarrow X_{high})]. \tag{8}$$

where $\uparrow$ and $\downarrow$ denote the upsampling and downsampling operations. $\mathcal{F}_L$ means the LRFB. And we use n LRFBs to deepen the network.

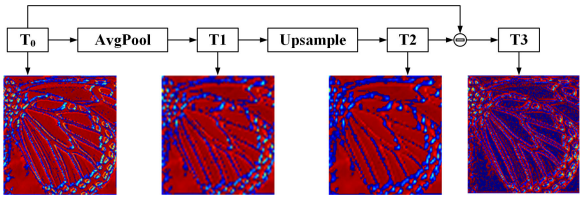

FIGURE 4. The visual activation feature maps of HFAB.

As for Y', we adopt an 1×1 convolution to reduce the number of channels and Channel Attention(CA) to emphasize important channels. Finally, the feature from the last LRFB is added to X_0 to get Y.

The residual block is shown in the left of Figure 3, which consists of a residual branch and an identity branch, and the weights of the two branches are equal. In order to adjust the importance of branches more flexibly, we propose Learnable Residual Feature Block(LRFB), as shown in the right of Figure 3. LHRB is composed of two Learnable Residual Blocks(LRBs) and two convolutional layers. Each LRB has learnable parameters λ to control the weights of the residual branch and the identity branch. After extracting local dense features through LRBs, in order to explore global hierarchical features, we concatenate the outputs of two LRBs through learnable parameters λ. Then, a 1×1 convolution layer fuses hierarchical features, and a 3×3 convolution is used to further extract features and reduce the number of channels. Next, we implement global residual learning by controlling the identity branch and residual branch through learnable parameters λ. The learnable parameters λ of each residual block are different.Through the learnable parameter λ, we have implemented adaptive control of different branch weights, so that the network can dynamically allocate attention according to the characteristics of the input data.

The target of High Frequency Adaptive Block(HFAB) is to estimate the high frequency information of the feature space. As shown in Figure 4, for the input $T_0 \in \mathbb{R}^{C\times H\times W}$, we first uses an average pooling layer to obtain the intermediate feature map T_1. Each value in T_1 can be viewed as the average intensity of each specified small area of T_0. Then, T_1 is upsampled through bicubic interpolation to obtain a new feature map T_2. T_2 can represent the average smoothness of input T_0. Therefore, the high frequency information T_3 can be expressed as:

$$T_3 = T_0 - T_2. \tag{9}$$

As shown in Figure 4, it can be observed that compared to T_0, T_2 is smoother. At the same time, T_3 retains more details and edges.

C. PIXEL CONTINUOUS ATTENTION

To capture pixel continuity dependence, we propose a Pixel Continuous Attention module(PiCA) that leverages visual cues observed from surrounding pixels to adaptively recalibrate the range that require attention. Our proposed PiCA module is shown in Figure 2. Given the sequence of pixel features $A \in \mathbb{R}^{P\times 64}$, the goal of PiCA is to obtain a non-local content representation Z', which aims to capture pixel continuity dependencies throughout the representation sequence by weighted sum of features at all pixel positions,

$$Z' = \rho([f(A, A), f(\theta(A), \phi(A))])g(A). \tag{10}$$

$\phi(\cdot)$,$\theta(\cdot)$,$g(\cdot)$ are implemented by using 1×1 convolutional layer.

$f(\cdot,\cdot)$ is a function of pairwise that computes the affinity for all positions, as shown in Equation (11).

$$f(\theta(A), \phi(A)) = \theta(A) \cdot \phi(A)^T. \tag{11}$$

And the size of the resulting pairwise function $f(\theta(A),\phi(A))$ is denoted as $\mathbb{R}^{P\times P}$, which encodes the mutual similarity between pixel positions under transformed feature representation sequence. Then we use the softmax function to normalize it into an attention map.

We find that attention map provides opportunities to understand the dependencies between surrounding pixels, but sometimes also lead to attention on pixel with less correlation but large feature differences, such as locations with large differences in color. Therefore, we introduce the Normalized Self-Similarity Matrix (NSSM) into the PiCA to learn to focus attention on a more appropriate pixel sequence range.

For NSSM, we take the pixel feature sequence as input and construct $f(A,A) = AA^T \in \mathbb{R}^{P\times P}$ through the self-similarity matrix. It can more directly reveal the continuity relationship between pixels. Finally, the self-similarity matrix is normalized through the softmax operation to obtain NSSM.

We combine NNSM as prior knowledge with attention map, and then use the $\rho(\cdot)$ which is a 1×1 convolution to recalibrated the attention map. In order to obtain the PiCA Map, the reault of $\rho(\cdot)$ is normalized by the softmax peration. Finally, the non-local content relationship Z' can be can be cal culated from the linear combination between the matrices resulted from $\rho(\cdot)$ and $g(\cdot)$.

At last, we use residual connections to generate the pixel feature representation sequence $Z \in \mathbb{R}^{P\times 64}$.$W_z$ is a 1×1 convolution. And Z can be can be viewed as an enhanced A.

$$Z = Z'W_z + A. \tag{12}$$

IV. EXPERIMENTAL RESULTS

In this section, we will use 6 benchmark datasets to evaluate the performance of our algorithm. First, we will introduce the experimental settings, evaluation metrics, and implementation details, and then systematically compare our method with state-of-the-art SISR algorithms on benchmark datasets.

A. IMPLEMENTATION AND TRAINING DETAILS

1) DATASETS AND METRICS

we used the 800 images of DIV2K dataset for training[51]. The DIV2K dataset provide a robust basis for learning

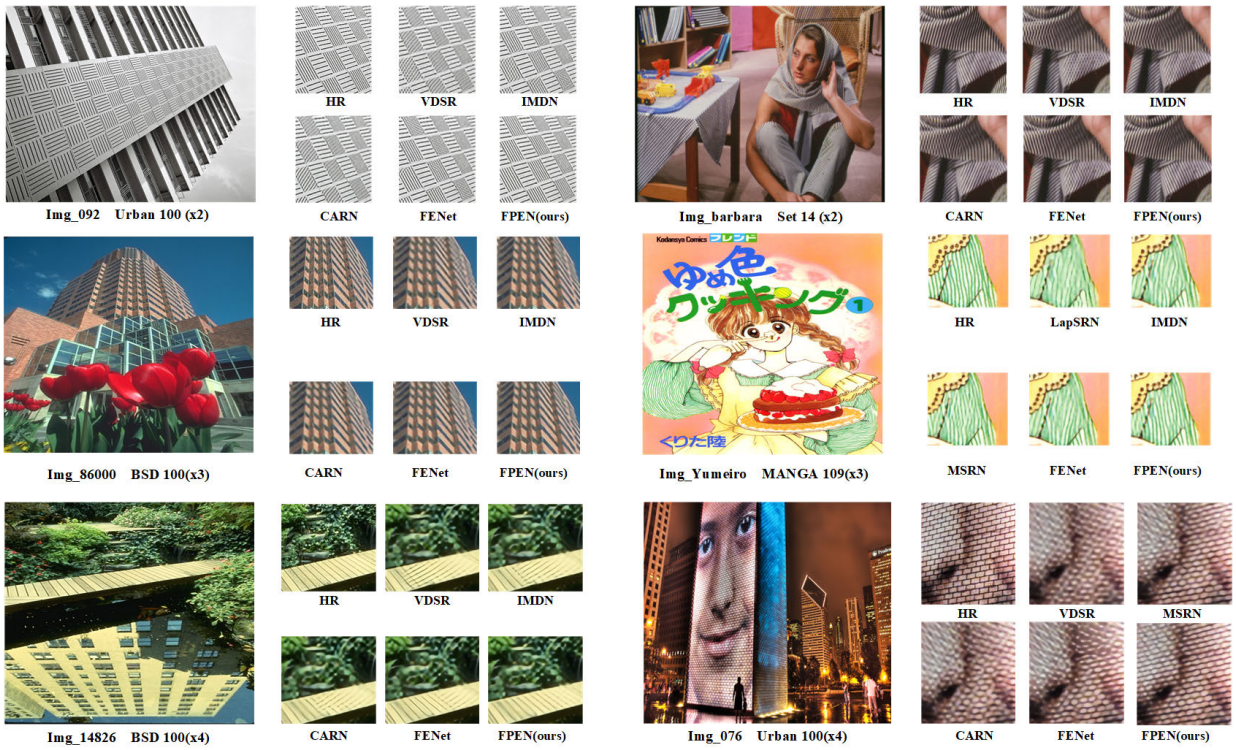

FIGURE 5. The visual results for ×2,×3,and ×4 super resolution. Our method can reconstruct images with more correct texture.

fine-grained details and textures, which is crucial for an effective super resolution model. For testing the performance of our model, we used the validation dataset from DIV2K and the well-known Set5 [6], Set14 [62], B100 [4], Urban100 [23], and Manga109 dataset which contain natural images, urban scenes and Japanese manga.

The evaluation of the super resolution performance was conducted using the widely accepted metrics: Peak Signal to Noise Ratio (PSNR).PSNR compares the maximum possible power of a signal against the power of corrupting noise. A higher PSNR indicates better reconstruction quality.

By calculating PSNR as evaluation indicators, we comprehensively evaluate the performance and effect of our proposed super-resolution algorithm.

2) IMPLEMENTATION DETAILS

During training, we set the batchsize to B, and first sample B random scales $r_{1\sim B}$ in uniform distribution $\mathcal{U}(1,4)$. Then we crop B pathches with size $\{48r_i \times 48r_i\}_{i=1}^{B}$ from training images. The ADAM optimizer is used to optimize the parameters, and the learning rate was initialized to 0.0003 and decayed by half per 200 epochs. We adopted L1 loss to train the network.

The operation system is Centos, and the GPU is NVIDIA V100. All experiments were completed using the deep learning framework Pytoroch 1.13.0 and the accelerator library CUDA Tookit 11.7.

B. COMPARISON WITH STATE-OF-THE-ART METHODS

1) RESULTS OF IN-DISTRIBUTION

For in-distribution scales(×2, ×3, ×4), we compare our method with previous work on Set5, Set14, B100, Urban100, and Manga109 dataset. From Table 1, it can be seen that our method has higher evaluation indicators in PSNR for scales 2, 3, and 4. For instancce, the PSNR of our algorithm on all datasets ×2 is improved by 0.14dB in avarage, and especially yields a performance increase of about 0.29dB on Urban100 dataset. Experimental results show that our proposed method has obvious superiority in super-resolution reconstruction effect. Besides, it can also prove that our method can adapt to the reconstruction of various scenarios. In particular, our training set does not contain any manga images, but MANGA109 shows an excellent reconstruction effect beyond other algorithms.

In addition to the quantitative results, we visualize the super-resolution images in Figure 5. It can be seen that the details reconstructed by our method are more refined, especially in edges and textures. For example, our method achieves more correct reconstruction of scarves in images of Set14 with ×2. This further verifies the remarkable effectiveness of our method in retaining high-frequency information and reconstructing more correctly detailed textures.

2) RESULTS OF OUT-OF-DISTRIBUTION

Although the training data we use only contains in-distribution scales(×1-×4), our algorithm can still achieve

TABLE 1. Comparisons with the State-of-the-arts. Performance is shown for scale factors 2, 3 and 4. The best and second best results are highlighted in red and blue respectively.

Algorithm	Scale	Set5	Set14	BSD100	Urban100	MANGA109
Bicubic	2	33.65	30.34	29.56	26.88	30.84
VDSR [9]	2	37.53	33.03	31.90	30.76	37.16
DRCN [45]	2	37.63	33.06	31.85	30.76	37.57
SEINet [46]	2	37.89	33.61	32.08	-	-
CARN [47]	2	37.76	33.52	32.09	31.92	38.36
SRFBN-S [48]	2	37.78	33.35	32.00	31.41	38.06
A2F-S [49]	2	37.79	33.32	31.99	31.44	38.11
CBPN [50]	2	37.90	33.60	32.17	32.14	-
MADNet [51]	2	37.94	33.46	32.10	31.74	-
FALSR-A [52]	2	37.82	33.55	32.12	31.93	-
HDRN [14]	2	37.75	33.49	32.03	31.87	38.07
DPN [53]	2	37.52	33.08	31.89	30.82	-
LAPAR-A [54]	2	38.01	33.62	32.19	32.10	38.67
IMDN [55]	2	38.00	33.63	32.19	32.17	38.88
OISR-RK2 [56]	2	38.02	33.62	32.20	32.21	-
FENet [18]	2	38.08	33.70	32.20	32.18	38.89
FPEN(ours)	2	38.10	33.88	32.26	32.50	39.07
Bicubic	3	30.39	27.55	27.21	24.46	26.95
VDSR [9]	3	33.66	29.77	28.82	27.14	32.01
DRCN [45]	3	33.82	29.76	28.80	27.15	32.24
CARN [47]	3	34.29	30.29	29.06	28.06	33.50
SRFBN-S [48]	3	34.20	30.10	28.96	27.66	33.02
A2F-S [49]	3	34.06	30.08	28.92	27.57	32.86
MADNet [51]	3	34.26	30.29	29.04	27.91	-
HDRN [14]	3	34.24	30.23	28.96	27.93	33.17
MSRN [57]	3	34.38	30.34	29.08	28.08	33.44
DPN [53]	3	33.71	29.80	28.84	27.17	-
LAPAR-A [54]	3	34.36	30.34	29.11	28.15	33.51
IMDN [55]	3	34.36	30.32	29.09	28.17	33.61
OISR-RK2 [56]	3	34.39	30.35	29.11	28.24	-
FENet [18]	3	34.40	30.36	29.12	28.17	33.52
FPEN(ours)	3	34.56	30.43	29.18	28.49	33.96
Bicubic	4	28.43	26.00	25.96	23.14	25.15
VDSR [9]	4	31.35	28.01	27.29	25.18	28.83
DRCN [45]	4	31.54	29.19	27.32	25.12	29.09
SEINet [46]	4	32.05	28.49	27.44	-	-
SRResNet [58]	4	32.05	28.53	27.57	26.07	-
CARN [47]	4	32.13	28.60	27.58	26.07	30.47
SRFBN-S [48]	4	31.98	28.45	27.44	25.71	29.91
A2F-S [49]	4	31.87	28.36	27.41	25.58	29.77
CBPN [50]	4	32.21	28.63	27.58	26.14	-
MADNet [51]	4	32.11	28.52	27.52	25.89	-
HDRN [14]	4	32.23	28.58	27.53	26.09	30.43
DPN [53]	4	31.42	28.07	27.30	25.52	-
LAPAR-A [54]	4	32.15	28.61	27.61	26.14	30.42
IMDN [55]	4	32.21	28.58	27.56	26.04	30.45
OISR-RK2 [56]	4	32.14	28.63	27.60	26.04	30.45
FENet [18]	4	32.24	28.61	27.63	26.20	30.46
FPEN(ours)	4	32.41	28.73	27.66	26.42	30.89

TABLE 2. Performance is shown for DIV2K dataset out of distribution with PSNR. Our method achieves the best reconstruction results.

Algorithm	×2	×3	×4	×6	×12	×18	×24	×30
Bicubic	31.01	28.22	26.66	24.82	22.27	21.00	20.19	19.59
EDSR-baseline-MetaSR [59]	34.64	30.93	28.92	26.61	23.55	22.03	21.06	20.37
EDSR-baseline-LIIF [41]	34.67	30.96	29.00	26.75	23.71	22.17	21.18	20.48
A-LIIF [42]	34.65	30.95	28.99	26.75	23.71	22.18	21.18	20.48
FPEN(ours)	34.77	31.05	29.08	26.83	23.76	22.22	21.21	20.51

reconstruction of out-of-distribution dimensions (×6, ×12,×18, ×24, ×30). Table 2 shows our reconstruction results of out-of-distribution scales. It can be seen that compared with other methods, our method achieves the highest PSNR. This shows that our method can not only achieve super-resolution reconstruction at in-distribution scales, but also have a good reconstruction effect at arbitrary scales. In order to achieve super-resolution reconstruction at arbitrary scales, our method focuses on exploring the feature representation of any pixel point on the high resolution image,

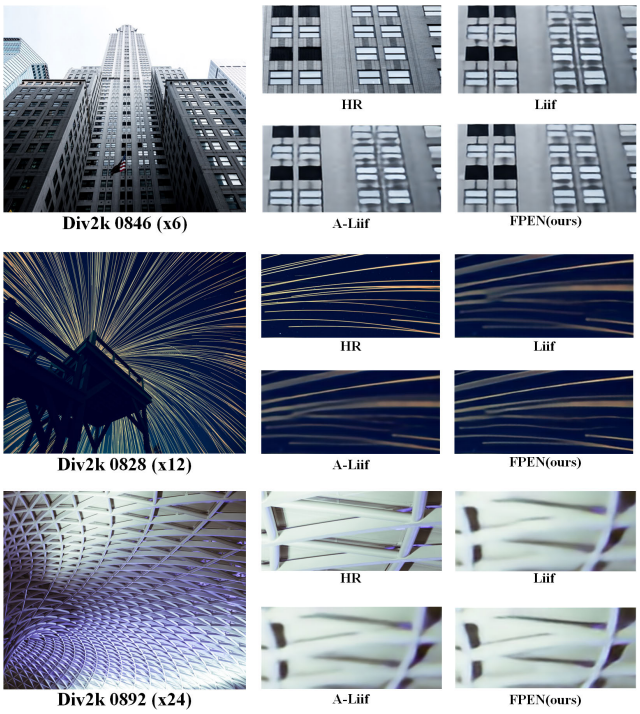

FIGURE 6. The visual results for×6, ×12, and ×24 super resolution. Our method can reconstruct images with of out-of-distribution scales.

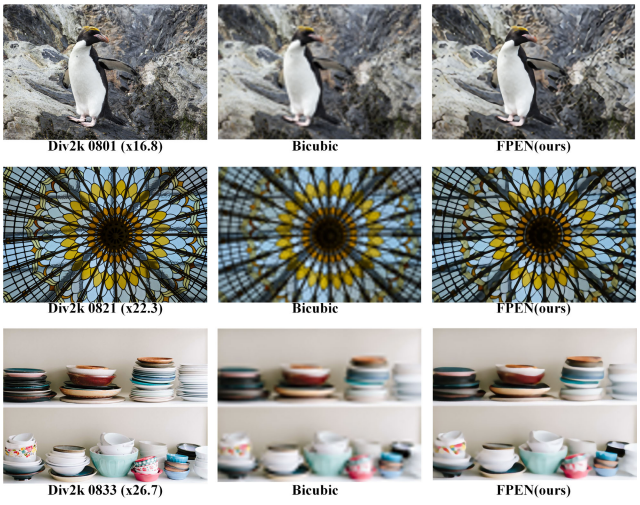

FIGURE 7. The visual results for ×16.8, ×22.3 and×26.7 super resolution. Our method can reconstruct images with non-integer scales.

TABLE 3. Average PSNR obtained with ablation study on DIV2K dataset.

Algorithm	×3	×6	×18	×30
Bicubic	28.22	24.82	21.00	19.59
EDSR-baseline-LIIF [41]	30.96	26.75	22.17	20.48
EDSR-baseline-LIIF+PICA	30.98	26.76	22.18	20.49
FPEN(ours)	31.05	26.83	22.22	20.51

and then takes the image coordinates and surrounding 2D features as input, and uses implicit prediction to obtain RGB values of the given coordinates.

For qualitative analysis, it can be seen in Figure 6 that our method can still retain more details and textures in large-scale reconstruction (×12,×24), and make the image visually clearer and sharper. This shows that our method can not only preserve more high-frequency information which consists of detailed textures, but also adaptively integrate non-local pixel information to achieve more accurate reconstruction.

3) RESULTS OF NON-INTEGER RECONSTRUCTION

Our super-resolution reconstruction algorithm can not only achieve reconstruction at integer scale, but also successfully achieve reconstruction at non-integer scale. This means that our algorithm has higher flexibility and adaptability, and can produce reconstruction results with various scaling factors in different application scenarios. In Figure 7, we show the reconstruction results under different non-integer scales respectively. We observe that reconstructions of non-integer scales present visually satisfactory results. Compared with traditional interpolation methods, our algorithm is better able to preserve image details and structure, generating more realistic and clearer reconstructed images.

C. ABLATION STUDY

In the ablation study, we explore the influence of the HFPB and PICA module. First, we use EDSR-baseline-LIIF [41] as our baseline, and then add the PICA module and HFPB module in sequence. All models are trained for 1000 iterations.

The experimental results are shown in Table 3, and we observed a significant improvement in performance when the PICA module is added. The EDSR-baseline-LIIF algorithm artificially defines the range of pixels to focus on and gives equal weight to each pixel. To further explore the impact of non-local pixels in super-resolution reconstruction, we use the PICA module to adaptively correct pixel ranges and weights. As can be seen from the Table 3, our PICA module can indeed further improve the reconstruction effect. The EDSR module in EDSR-baseline-LIIF is used for feature extraction, but it does not distinguish between low-frequency and high-frequency features. We replaced the EDSR module with the HFPB module as FPEN(ours). As can be seen from Table 3, our method achieves the best reconstruction. This shows that high-frequency information is very important in super-resolution reconstruction. Our method distinguishes high-frequency and low-frequency components from the feature dimension, assigns more complex operations to high-frequency components to preserve high-frequency components, which can better restore the texture of the image.

V. CONCLUSION

We propose a Feature Preserving and Enhancing Network (FPEN) based on implicit representation to achieve super-resolution reconstruction at arbitrary-scale. We utilize High Frequency Preserving Block (HFPB) to distinguish high-frequency components and low-frequency components from the feature dimension, and assign complex operations to

high-frequency regions to preserve high-frequency information, so that the details and texture of the image can be better reconstructed. Furthermore, our proposed Pixel Continuity Attention (PiCA) module adaptively recalibrates the range of attention required to better capture the continuity dependence of pixels. In addition, we combine implicit representation to learn implicit functions that map low resolution images to high-resolution images, in order to achieve super-resolution reconstruction with arbitrary scales. Through extensive experiments, our method achieves remarkable image reconstruction results at arbitrary scales. Compared with other state-of-the-art super-resolution algorithms, our method can provide richer details and texture information by preserving and enhancing image features. In the future, we will further explore decomposing low-frequency and high-frequency information at the pixel level to achieve more accurate image reconstruction.

REFERENCES

[1] H. Lin and J. Yang, "Light weight IBP deep residual network for image super resolution," *IEEE Access*, vol. 9, pp. 93399–93408, 2021.
[2] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, and G. Wang, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020.
[3] Y. Tan, J. Cai, S. Zhang, W. Zhong, and L. Ye, "Image compression algorithms based on super-resolution reconstruction technology," in *Proc. IEEE 4th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2019, pp. 162–166.
[4] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "SOD-MTGAN: Small object detection via multi-task generative adversarial network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 206–221.
[5] H. Yang and Y. Wang, "An effective and comprehensive image super resolution algorithm combined with a novel convolutional neural network and wavelet transform," *IEEE Access*, vol. 9, pp. 98790–98799, 2021.
[6] Z. Yang, D. Pan, and P. Shi, "Joint image dehazing and super-resolution: Closed shared source residual attention fusion network," *IEEE Access*, vol. 9, pp. 105477–105492, 2021.
[7] 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.
[8] 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)*, Oct. 2017, pp. 4539–4547.
[9] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[10] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
[11] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Zurich, Switzerland, Dec. 2015, pp. 57–65.
[12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[13] P. Behjati, P. Rodríguez, A. Mehri, I. Hupont, C. F. Tena, and J. Gonzalez, "OverNet: Lightweight multi-scale super-resolution with overscaling network," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2693–2702.
[14] 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.
[15] W. Xie, D. Song, C. Xu, C. Xu, H. Zhang, and Y. Wang, "Learning frequency-aware dynamic network for efficient super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4288–4297.
[16] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1273–1284, Mar. 2022.
[17] A. Muqeet, J. Hwang, S. Yang, J. Kang, Y. Kim, and S.-H. Bae, "Multi-attention based ultra lightweight image super-resolution," in *Computer Vision—ECCV 2020*. Cham, Switzerland: Springer, 2020, pp. 103–118.
[18] P. Behjati, P. Rodriguez, C. F. Tena, A. Mehri, F. X. Roca, S. Ozawa, and J. Gonzàlez, "Frequency-based enhancement network for efficient super-resolution," *IEEE Access*, vol. 10, pp. 57383–57397, 2022.
[19] F. Zhu, C. Fang, and K.-K. Ma, "PNEN: Pyramid non-local enhanced networks," *IEEE Trans. Image Process.*, vol. 29, pp. 8831–8841, 2020.
[20] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision—ECCV 2016*. Cham, Switzerland: Springer, 2016, pp. 391–407.
[21] S. Dargahi, A. Aghagolzadeh, and M. Ezoji, "Single image super-resolution by cascading parallel-structure units through a deep-shallow CNN," *Optik*, vol. 286, Sep. 2023, Art. no. 171001.
[22] L. Yu, X. Li, Y. Li, T. Jiang, Q. Wu, H. Fan, and S. Liu, "DIPNet: Efficiency distillation and iterative pruning for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1692–1701.
[23] M. Wei and X. Zhang, "Super-resolution neural operator," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18247–18256.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[25] J. Qin, Y. Huang, and W. Wen, "Multi-scale feature fusion residual network for single image super-resolution," *Neurocomputing*, vol. 379, pp. 334–342, Feb. 2020.
[26] J. Wang, Y. Chen, Y. Wu, J. Shi, and J. Gee, "Enhanced generative adversarial network for 3D brain MRI super-resolution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3616–3625.
[27] Y. Bahat, Y. Zhang, H. Sommerhoff, A. Kolb, and F. Heide, "Neural volume super-resolution," 2022, *arXiv:2212.04666*.
[28] X. Liu, S. Chen, L. Song, M. Woźniak, and S. Liu, "Self-attention negative feedback network for real-time image super-resolution," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6179–6186, Sep. 2022.
[29] W. Chen, S. Ouyang, W. Tong, X. Li, X. Zheng, and L. Wang, "GCSANet: A global context spatial attention deep learning network for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1150–1162, 2022.
[30] C. Tan, Z. Gao, L. Wu, Y. Xu, J. Xia, S. Li, and S. Z. Li, "Temporal attention unit: Towards efficient spatiotemporal predictive learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18770–18782.
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–13.
[32] K. Wang, X. Liao, J. Li, D. Meng, and Y. Wang, "Hyperspectral image super-resolution via knowledge-driven deep unrolling and transformer embedded convolutional recurrent neural network," *IEEE Trans. Image Process.*, vol. 32, pp. 4581–4594, 2023.
[33] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.
[34] N. Lu, W. Yu, X. Qi, Y. Chen, P. Gong, R. Xiao, and X. Bai, "MASTER: Multi-aspect non-local network for scene text recognition," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107980.
[35] G. Zhang, J.-H. Xue, P. Xie, S. Yang, and G. Wang, "Non-local aggregation for RGB-D semantic segmentation," *IEEE Signal Process. Lett.*, vol. 28, pp. 658–662, 2021.
[36] Y. Mei, Y. Fan, and Y. Zhou, "Image super-resolution with non-local sparse attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3516–3525.
[37] P. Yi, Z. Wang, K. Jiang, J. Jiang, T. Lu, and J. Ma, "A progressive fusion generative adversarial network for realistic and consistent video super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2264–2280, May 2022.

[38] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3D reconstruction in function space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4455–4465.
[39] S. Qian, J. Xu, Z. Liu, L. Ma, and S. Gao, "UNIF: United neural implicit functions for clothed human reconstruction and animation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 121–137.
[40] T. Li, X. Wen, Y.-S. Liu, H. Su, and Z. Han, "Learning deep implicit functions for 3D shapes with dynamic code clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12830–12840.
[41] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8624–8634.
[42] H. Li, T. Dai, Y. Li, X. Zou, and S.-T. Xia, "Adaptive local implicit image function for arbitrary-scale super-resolution," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 4033–4037.
[43] J. McGinnis, S. Shit, H. Bran Li, V. Sideri-Lampretsa, R. Graf, M. Dannecker, J. Pan, N. Stolt Ansó, M. Mühlau, J. S. Kirschke, D. Rueckert, and B. Wiestler, "Single-subject multi-contrast MRI super-resolution via implicit neural representations," 2023, *arXiv:2303.15065*.
[44] Y.-H. Chen, S.-C. Chen, Y.-H. Chen, Y.-Y. Lin, and W.-H. Peng, "MoTIF: Learning motion trajectories with local implicit neural functions for continuous space-time video super-resolution," 2023, *arXiv:2307.07988*.
[45] 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)*, Jun. 2016, pp. 1637–1645.
[46] J.-S. Choi and M. Kim, "A deep convolutional neural network with selection units for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1150–1156.
[47] 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. 252–268.
[48] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3862–3871.
[49] X. Wang, Q. Wang, Y. Zhao, J. Yan, L. Fan, and L. Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–20.
[50] F. Zhu and Q. Zhao, "Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2453–2460.
[51] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021.
[52] X. Chu, B. Zhang, H. Ma, R. Xu, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 59–64.
[53] Y. Liang, R. Timofte, J. Wang, S. Zhou, Y. Gong, and N. Zheng, "Single-image super-resolution—When model adaptation matters," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107931.
[54] W. Li, K. Zhou, L. Qi, N. Jiang, J. Lu, and J. Jia, "LAPAR: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 20343–20355.
[55] 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.
[56] X. He, Z. Mo, P. Wang, Y. Liu, M. Yang, and J. Cheng, "ODE-inspired network design for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1732–1741.
[57] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 517–532.
[58] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[59] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "Meta-SR: A magnification-arbitrary network for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1575–1584.

MINGLAN SU received the B.S. degree from Beijing Jiaotong University, in 2017, and the M.S. degree from the Beijing University of Posts and Telecommunications, in 2020. She is currently an Artificial Intelligence Engineer with the Institute of Big Data and Artificial Intelligence and the China Telecom Research Institute. Her research interests include super-resolution reconstruction, 3D human body reconstruction, and intelligent recommendation.

XINCHI LI received the degree in computer application from the Computer Network Information Center, Chinese Academy of Sciences. She is currently the Director of the Big Data Center of China, Telecom Research Institute. She is also a Senior Engineer. She has published more than ten papers and written two monographs. She is mainly engaged in technical research and product development in the fields of big data and blockchain.

JIAOYANG XU received the B.S. degree from the School of Mathematics and Statistics, Central China Normal University, Hubei, in 2021. He is currently pursuing the M.S. degree with the School of Systems Science, Beijing Normal University, Beijing. His research interests include deep learning, brain-like intelligence, and intelligent robotics.

MINGCHUAN YANG received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, in 2002. He is currently the Director of the Institute of Big Data and Artificial Intelligence and the China Telecom Research Institute. His research interests include big data and artificial intelligence.

CHAOYING ZHANG received the B.S. degree from the Taiyuan University of Technology, in 2016, and the M.S. degree from the Beijing Institute of Technology, in 2019. She is currently an Artificial Intelligence Engineer with the Institute of Big Data and Artificial Intelligence and the China Telecom Research Institute. Her research interests include 3D human body reconstruction and image processing.