International Journal of Machine Learning and Cybernetics (2024) 15:2309–2321
https://doi.org/10.1007/s13042-023-02030-1

ORIGINAL ARTICLE

# Single image super-resolution via global aware external attention and multi-scale residual channel attention network

**Mingming Liu[1,2] · Sui Li[2,3] · Bing Liu[2,3] · Yuxin Yang[2,3] · Peng Liu[4] · Chen Zhang[2,3]**

Received: 24 October 2022 / Accepted: 25 October 2023 / Published online: 29 November 2023


## Abstract

Recently, deep convolutional neural networks (CNNs) have shown significant advantages in improving the performance of single image super-resolution (SISR). To build an efficient network, multi-scale convolution is commonly incorporated into CNN-based SISR methods via scale features with different perceptive fields. However, the feature correlations of the same sample are not fully utilized by the existing multi-scale SISR approaches, impeding the further improvement of reconstruction performance. In addition, the correlations between different samples are still left unexplored. To address these problems, this paper proposes a deep-connected multi-scale residual attention network (DMRAN) by virtue of the feature correlations of the same sample and the correlations between different samples. Specifically, we propose a deep-connected multi-scale residual attention block (DMRAB) to take fully advantage of the multi-scale and hierarchical features, which can effectively learn the local interdependencies between channels by adjusting the channel features adaptively. Meanwhile, a global aware external attention (GAEA) is introduced to boost the performance of SISR by learning the correlations between all the samples. Furthermore, we develop a deep feature extraction structure (DFES), which seamlessly combines the stacked deep-connected multi-scale residual attention groups (DMRAG) with GAEA to learn deep feature representations incrementally. Extensive experimental results on the public benchmark datasets show the superiority of our DMRAN to the state-of-the-art SISR methods.

**Keywords** Single image super-resolution · Deep feature extraction structure · Deep-connected multi-scale residual attention block · Local aware channel attention · Global aware external attention

## 1 Introduction

Single Image Super-Resolution (SISR) has attracted considerable attention in the field of computer vision, which aims to accurately construct a high-resolution (HR) image from the given low-resolution (LR) image. It has been widely used in various computer vision tasks, such as public safety monitoring [1], medical imaging [2], and satellite imaging [3]. However, due to the lack of high-frequency information, multiple different versions of HR images could be generated for the same LR input, so SISR is a highly ill-posed problem. A large amount of SISR methods have been proposed to overcome this problem, which can generally be divided into three categories: interpolation-based [4, 5], reconstruction-based [6, 7] and learning-based methods [8–11].

Recently, benefiting from the excellent feature representation ability and end-to-end training paradigm, deep convolutional neural network (CNN) based SR methods [8–10, 12–18] have become the dominant approaches to

Mingming Liu and Sui Li are co-first authors.

✉ Bing Liu
liubing@cumt.edu.cn

✉ Peng Liu
liupeng@cumt.edu.cn

[1] School of Intelligent Manufacturing, Jiangsu Vocational Institute of Architectural Technology, Xuzhou 221000, Jiangsu Province, China

[2] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, Jiangsu Province, China

[3] Mine Digitization Engineering Research Center of Ministry of Education of the People's Republic of China, Xuzhou, China

[4] National Joint Engineering Laboratory of Internet Applied Technology of Mines, Xuzhou 221008, Jiangsu Province, China

SISR. The CNN-based SR methods try to capture a non-linear mapping relationship between LR image and its HR counterpart, which have shown remarkable superiority compared with the conventional SR methods. SRCNN [10] was a shallow three-layer model for SISR, which is the first successful scheme utilizing CNN for the SISR reconstruction. Later, FSRCNN [18] improved the recovery speed with higher restoration quality by introducing a deconvolution layer in the network. Following these pioneer works, most of the SISR methods based on CNN framework tend to design deeper or wider networks to reconstruct the HR images more accurately. Inspired by ResNet [19], Kim et al. introduced the residual learning in VDSR [9] and DRCN [17] to further increase the depth of the network. DRRN [16] later extended the network depth to more layers via a recursive structure. By stacking multiple simplified residual blocks, Lim et al. devised a deeper and wider model EDSR [20]. Motivated by DenseNet [21], some researchers built more effective SISR models, such as MemNet [22], SRDenseNet [23] and RDN [15].

Although considerable progress has been made, the existing CNN-based SISR methods still have some limitations. Firstly, the scale and hierarchical features obtained from intermediate layers of networks are not fully utilized by most CNN-based SISR approaches, resulting in relatively low performance. Secondly, to learn more discriminative high-frequency information, the CNN-based SISR methods mostly pay more attention to the deeper or wider networks, but neglect the inherent feature correlations between channels for the same sample. Finally, the correlations between different samples, which have been shown to be beneficial for the deep feature representations in various visual tasks [24], are still left unexplored in the existing SISR methods.

To solve these problems, we propose a deep-connected multi-scale residual attention network (DMRAN) to learn more powerful feature representations. Specifically, we present a deep-connected multi-scale residual attention block (DMRAB) consisting of the multi-scale and channel attention mechanisms. Our DMRAB can leverage both the multi-scale and hierarchical features through the multi-scale convolution operations. At the same time, our DMRAB can rescale the channel features adaptively to capture the local interdependencies between channels of the same sample. In addition, we devise a deep feature extraction structure (DFES) to learn the deep feature representations incrementally. Among them, the deep-connected multi-scale residual attention group (DMRAG) is used as the basic component, and the long range skip connection (LRSC) is used to perform the residual learning. Meanwhile, a global aware external attention (GAEA) is introduced into DFES to learn the correlations between different samples, which can boost the performance of image reconstruction further. Extensive experimental evaluations on five public benchmark datasets and ablation analysis demonstrate the effectiveness of our proposal.

Our contributions can be summarized as follows:

- We build a deep-connected multi-scale residual attention network (DMRAN) to improve the performance of SISR. To our knowledge, this is the first SISR model that introduces the correlations learning between different samples to facilitate the image reconstruction.
- We propose a deep-connected multi-scale residual attention block (DMRAB), which exploits the mulit-scale and hierarchical features simultaneously, and rescales the channel features adaptively to capture the local interdependencies between channels for the same sample. In addition, the short range skip connection in DMRAB helps to bypass abundant low-frequency information.
- We propose a deep feature extraction structure (DFES) to construct a deep trainable network, which further incorporates a global aware external attention (GAEA) module to learn the correlations between different samples. The feature representation ability of our network is further enhanced by this GAEA module.

## 2 Related works

### 2.1 CNN-based SISR methods

SRCNN [10] first introduced the convolutional neural network (CNN) into single image SR. It up-scales LR image to the HR image of the desired size through the bicubic interpolation method, then fits the non-linear mapping by utilizing only three convolutional layers to output the HR image. SRCNN was built on a lightweight but efficient network and achieved superior performance compared to traditional methods. Later, FSRCNN [18] accelerated SRCNN by embedding a deconvolution layer in the network, which does not need any pre-processing operations and even achieves better restoration quality. ESPCN [25] incorporated the sub-pixel convolutional layer to improve the performance of SISR. However, the networks of all the models mentioned above are relatively shallow, with a depth of less than 5 layers, and their learning ability is limited.

After that, most CNN-based SISR approaches began to build deeper or wider networks to reconstruct more accurate HR images. For example, inspired by ResNet [19], Kim et al. introduced residual learning in VDSR [9] and DRCN [17], which extends the depth of network to 20 layers and enhances the accuracy of SISR. To further improve the performance, DRRN [16] built a much deeper network with a larger receptive field by adopting a recursive structure, extending its depth to 52 layers. Combining traditional image algorithm Laplacian pyramid with deep learning,

LapSRN [26] was proposed to gradually predict the residuals from coarse to fine based on a cascade structure of CNNs and reconstruct the HR images by progressive up-sampling. By stacking simplified residual blocks, Lim et al. then constructed a very deep and wide model EDSR [20] to further boost the performance of SISR. Some recent methods, such as MemNet [22] and RDN [15], paid more attention to taking full advantage of all hierarchical features based on dense blocks [21]. RCAN [14] introduced the channel attention into the residual blocks and further increased the network depth to more than 400 layers. The significant improvement of performance shows that network depth is crucial for SISR. Although increasing network depth can greatly improve performance of SISR, it could be extremely difficult to train due to the large number of parameters and may suffer from the problem of gradient vanishing or gradient exploding during the training process. In addition, when the network depth increases to a certain extent, it is invalid to increase the network depth further to improve performance.

### 2.2 Multi-scale SR methods

In the feature extraction modules, one of the key problems is the size of the receptive fields of convolution kernels. Small kernels are beneficial for the extraction of low-frequency components, while large kernels are helpful to extract high-frequency components. Multi-scale features are usually defined as the features acquired by multi-scale convolution, which is consistent with the concept of capturing more features of different receptive fields simultaneously to extract more abundant information. Multi-scale feature fusion networks can be roughly classified as the serial skip connection structure network and the parallel multi-branch network.

On the one hand, hierarchical features [15], i.e., the features of different network depths in CNN networks, can be considered as the scale features since they have different receptive fields. Therefore, the CNN-based networks with skip connections, such as DRCN [17], RED [27], DRRN [16], SRDenseNet [23], RDN [15], etc., can be regarded as the serial skip connection structure networks to some extent. On the other hand, multi-scale feature extraction mostly chooses the structure of parallel multi-streams, such as Inception [28]. A multi-scale residual network named MSRN [13] was proposed to fully extract multi-scale spatial features. For the sake of extracting features of different scales, MSRN incorporates two convolution kernels of different sizes ($3\times3$, $5\times5$) into each block corresponding to two branches. Moreover, MSRN is a simple and effective SISR model, which makes good use of local multi-scale features as well as hierarchical features. Later, plenty of multi-scale SISR networks have been proposed, such as MSFFRN [12], PMRN [29], AAMN [30], etc. However, the methods mentioned above treat the scale features from parallel multi-branches equally, while ignoring their dissimilarities and redundancy.

### 2.3 Attention mechanism

High-frequency information and details play a critical role in image super-resolution. Due to the limited information processing resources, attention mechanism is introduced into computer vision for the purpose of achieving optimal performance by allocating available processing resources more reasonably. The attention mechanism essentially imitates the activities of the human brain in a simplified way by selectively focusing on some of the most important features.

In recent years, plenty of CNN based models have achieved satisfactory results by means of attention mechanism. The squeeze-and-excitation network (SENet) designed by Hu et al. [31] enhanced the discriminability of CNN by explicitly modeling the interdependence between feature channels. SENet automatically learns the importance of each channel feature, then emphasizes useful features and suppresses less useful features according to the importance of channels. Zhang et al. [14] combined SENet with simplified residual blocks (RB) to construct a very deep residual channel attention network (RCAN), and achieved the state-of-the-art performance at the time. Since SENet only exploited first-order statistical information (global average pooling), Dai et al. [32] proposed a second-order attention network (SAN), which introduces higher-order feature statistics for better feature selection. The non-local neural network built by Wang et al. [33] was originally used to explore semantic relationships in high-level tasks, such as object detection and image classification. Many other methods, such as RNAN [34], SAN [32], CSNLN [35], etc., incorporated the non-local attention mechanism into their proposal to fully leverage clues in image structures by introducing long-range feature correlations. Although non-local attention boosts the performance of SISR, this mechanism requires matrix multiplication to calculate the affinity between features to gain long-term dependencies, which usually leads to large memory overhead and high computational complexity.

## 3 Proposed method

### 3.1 Network architecture

As shown in Fig. 1, our DMRAN is mainly composed of four components: the shallow features extraction part (SFEP), the deep feature extraction part (DFEP), the upscale module part (UP), and the reconstruction part (REC). Given an LR image $I_{LR}$ as input, we denote the output of DMRAN as $I_{SR}$. First, $I_{LR}$ is input into SFEP to extract the shallow feature $F_0$,

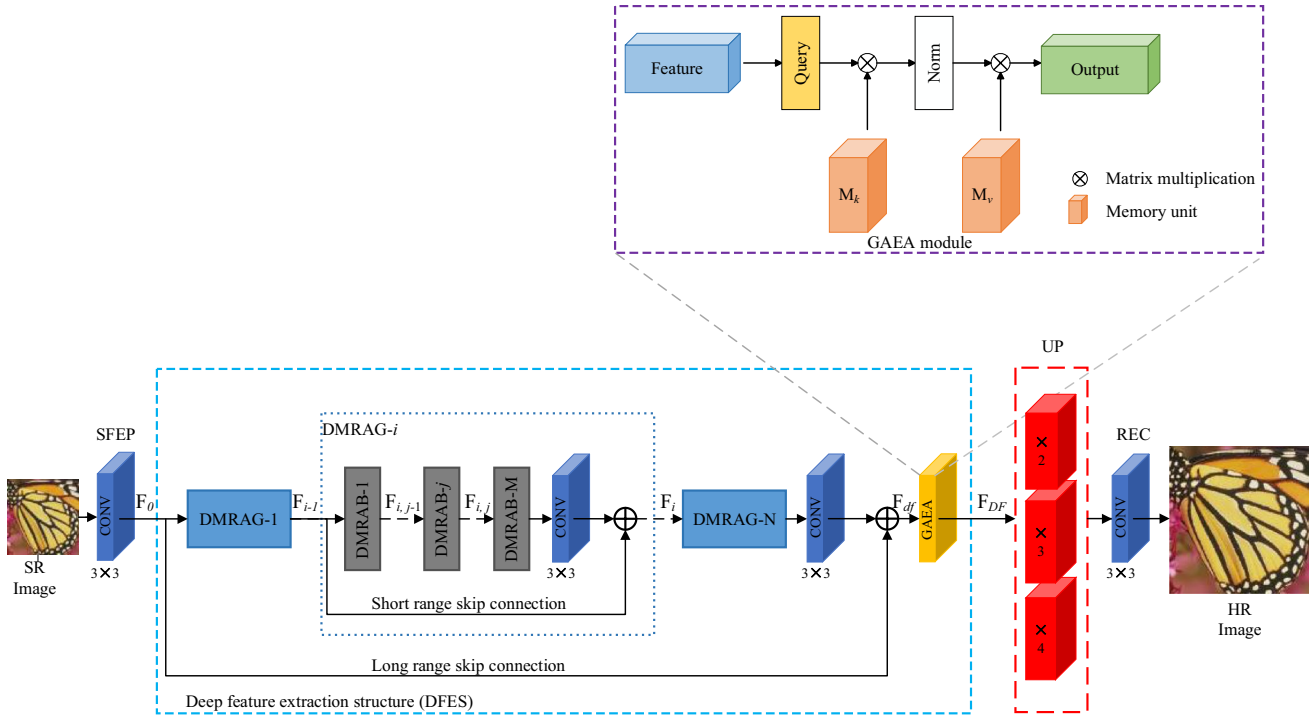


**Fig. 1** Architecture of the proposed deep-connected multi-scale residual attention network (DMRAN)

where SFEP is actually a convolutional layer here. Formally, we have

$$F_0 = f_{SFEP}(I_{LR}), \tag{1}$$

where $f_{SFEP}(.)$ represents the convolution operation. The obtained shallow feature $F_0$ is then fed into the next deep feature extraction part (DFEP). We can further have

$$F_{DF} = f_{DFES}(F_0), \tag{2}$$

where $f_{DFES}(.)$ denotes our proposed deep feature extraction structure, which consists of $N$ deep-connected multi-scale residual attention groups (DMRAGs), a long range skip connection (LRSC) and a global aware external attention (GAEA). Therefore, our proposed DFES contributes to constructing a very deep and trainable network architecture. Next, the obtained deep feature $F_{DF}$ is upscaled by virtue of an upscale module

$$F_{UP} = f_{UP}(F_{DF}), \tag{3}$$

where $f_{UP}(.)$ and $F_{UP}$ denote the upscale module and the upscaled feature, respectively. The upscale modules can be constructed by using the deconvolution layer [18] or ESPCN [25]. Compared with the pre-upsampling SR method, such post-upsampling SR method with the upscale modules can achieve a better balance between the model complexity and performance. Therefore, it is widely adopted in the recent CNN-based SR models [15, 32, 36]. Then one convolutional layer is applied to convert the upscaled feature $F_{UP}$ into the SR image

$$I_{SR} = f_{REC}(F_{UP}) = f_{DMRAN}(I_{LR}), \tag{4}$$

where $f_{REC}(.)$ and $f_{DMRAN}(.)$ represent the reconstruction layer and the function of DMRAN respectively.

Some loss functions, such as $L_1$ [15, 16, 20, 26, 32, 37, 38] $L_2$, perceptual losses [39] and adversarial losses [40], have been widely used to train the SISR networks. For fair comparison, the commonly used $L_1$ loss function is utilized to optimize our DMRAN. Given a training dataset $\{I^i_{LR}, I^i_{HR}\}_{i=1}^N$, which contains $N$ LR- HR image pairs, the $L_1$ loss function is formulated as:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \| f_{DMRAN}(I^i_{LR}) - I^i_{HR} \|_1, \tag{5}$$

where $\theta$ denotes the learnable parameter set of DMRAN.

### 3.2 Local aware channel attention (LACA)

LR images contain abundant low-frequency information, but limited high-frequency information, such as sharp contrast edges, textures, and other details. Therefore, it is critical to extract more limited but valuable high-frequency information from LR images. However, previous CNN-based SISR methods neglect the interdependencies of feature channels for

the same sample, which is not conducive to the extraction of high-frequency information. To this end, we exploit a local aware channel attention (LACA) mechanism to capture the local cross-channel interaction for the same sample and rescale the previously acquired features.

As shown in Fig. 2, suppose that $X = [x_1, \ldots, x_C]$ is the set of $C$ feature maps, whose spatial dimension is $h \times w$. We first apply the global average pooling along the spatial dimension to obtain the global statistics, which can be expressed as $Z = [z_1, \ldots, z_C]$. The $c$-th element of $Z$ is obtained by

$$z_c = f_{GAP}(x_c) = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} x_c(i,j) \tag{6}$$

where $x_c(i,j)$ is the element at the position $(i,j)$ of $c$-th feature map $x_c$ and $f_{GAP}(.)$ represents the global average pooling (GAP) function. Of note, the global average pooling can also be replaced by other more sophisticated aggregation techniques. Then a fast $1D$ convolution is implemented

$$S = \sigma\left(C1D_s(Z)\right), \tag{7}$$

where $C1D_s(.)$ indicates $1D$ convolution operation, whose kernel size is $s$, and $\sigma(.)$ is the sigmoid function responsible for generating the final channel-wise weights $S$, which plays a key role in the LACA mechanism. As analyzed in [41], the LACA model with only $s$ parameters has much lower model complexity compare with SENet [31]. Specifically, the kernel size $s$ can be adaptively obtained by

$$s = \varphi(C) = \left| \frac{\log_2(C) + b}{\delta} \right|_{odd}, \tag{8}$$

where $|e|_{odd}$ represents the nearest odd number of $e$. Here, we set $b$ and $\delta$ to 1 and 2 respectively. It is obvious from the mapping function $\varphi$ that the larger the value of C, the wider the range of local interdependencies between channels captured by the LACA. Finally, the learned weight set $S$ is used to reweight the input $X$

$$\hat{X} = S \otimes X, \tag{9}$$

where $\otimes$ represents the element-wise product operation, and $\hat{X} = [\hat{x}_1, \cdots, \hat{x}_c, \cdots, \hat{x}_C]$ denotes the output of the LACA mechanism. With the LACA module, the multi-scale residual features in the DMRAB can be adaptively rescaled.

### 3.3 Global aware external attention (GAEA)

In contrast to the LACA mechanism that reweights each feature map by appropriately capturing local cross-channel interaction within a single sample, the global aware external attention (GAEA) pays more attention to the potential correlations between different samples, which contributes to the better representation of features. Previous SR algorithms [14, 32, 35] based on attention mechanism mainly investigate channel attention, spatial attention and their combination or deformation. Although the non-local attention can capture long-range interactions, it is difficult to apply the non-local attention directly to the original image due to the quadratic complexity of the number of input pixels. Therefore, the previous SR method [32] applies the non-local attention to patches instead of pixels to reduce model parameters. Moreover, all of the above attention mechanisms concentrate on learning the attention within a single image and ignore the correlations learning between different samples. Therefore, it's necessary to capture the correlations between different samples to further improve the performance of SISR.

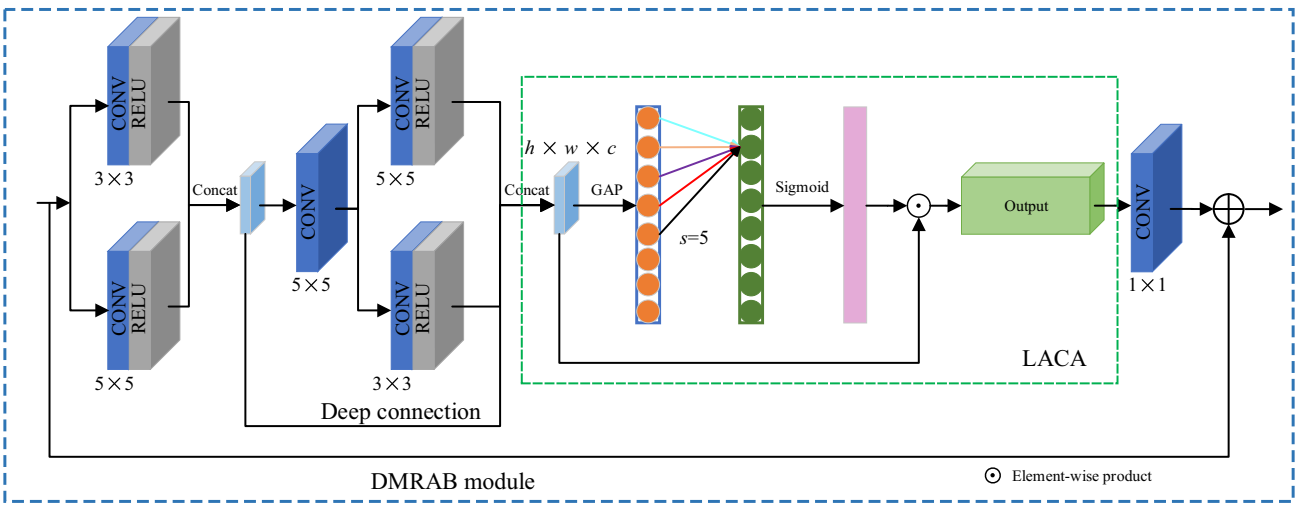


**Fig. 2** Local Aware Channel Attention (LACA) and DMRAB module

Based on the discussion mentioned above, we propose to learn the potential correlations between different samples by introducing the lightweight global aware external attention (GAEA) module, as shown in Fig. 1. Suppose $F \in R^{N\times d}$ is the input feature map, where $N = h \times w$ denotes the number of pixels and $d$ is the feature dimension. Inspired by self-attention, we use two external and learnable memory units $M_k$ and $M_v$ as the key and value. The purpose of the two external memories is to capture the most discriminative features of the entire dataset and exclude interference information from other samples. The GAEA can be formulated as

$$A = (\alpha)_{i,j} = Norm\left(FM_k^T\right), \tag{10}$$

$$F_{out} = AM_v, \tag{11}$$

Here, $(\alpha)_{i,j}$ in Eq. (10) is the similarity between the $i$-th pixel and the $j$-th rows of the external memory unit $M_k$. $M_k$ ,$M_v \in R^{S\times d}$ are two different and learnable parameters independent of individual samples, which serve as two shared memories for the entire training dataset. $A$ is an attention map learned from the prior knowledge, which is normalized in a manner similar to self-attention. Finally, we utilize the similarities in $A$ to update the features from $M_v$. As discussed in [24], the two memories are easy to implement in practice with linear layers, and the computational complexity of GAEA is linear with the number of pixels. Combined with the lightweight GAEA mechanism, our DMRAN can adaptively learn the attention between different samples across the whole training dataset.

### 3.4 Deep feature extraction structure (DFES)

We now elaborate our proposed deep feature extraction structure (DFES) (see Fig. 1), which consists of $N$ deep-connected multi-scale residual attention groups (DMRAG), one long range skip connection (LRSC) and one global aware external attention (GAEA) module. Each DMRAG further contains $M$ deep-connected multi-scale residual attention blocks (DMRAB) with short range skip connection (SRSC). This DMRAG structure is conducive to training a very deep high-performance network for SISR. In addition, the lightweight GAEA module is embedded in the end of our DFES to capture the correlations between different samples.

Previous studies [15, 20] have proved the superiority of stacked residual blocks in constructing deeper CNNs. However, very deep SR networks simply constructed in this way are prone to cause the problems of training difficulty and performance bottleneck. Inspired by the method in [20], we utilize the deep-connected multi-scale residual attention group (DMRAG) as the basic unit. A DMRAG in the $i$-th group can be denoted as

$$F_i = g_i\left(F_{i-1}\right) = g_i\left(g_{i-1}\left(\cdots g_1\left(F_0\right)\cdots\right)\right), \tag{12}$$

where $g_i(.)$ represents the function of the $i$-th DMRAG. $F_{i-1}$ and $F_i$ are the input and output for the $i$-th DMRAG respectively. Considering the simple stacking of DMRAGs is difficult to improve the performance, we apply LRSC in DFES to facilitate the stable training of deep network and bypass abundant low-frequency information. Therefore, the deep feature is obtained as

$$F_{df} = F_0 + W_{LRSC}F_N = F_0 + W_{LRSC}g_i\left(g_{i-1}\left(\cdots g_1\left(F_0\right)\cdots\right)\right), \tag{13}$$

where $W_{LRSC}$ denotes the weights of the convolutional layer after DMRAG-$N$. For simplicity, the bias is omitted here. As mentioned in Sect. 3.3, we embed the GAEA module in our DFES to adaptively learn the attention between different samples across the whole training dataset, which further improves the discriminability of our network combined with the LACA mechanism. Then, the deep feature obtained by DFES can be denoted as

$$F_{DF} = f_{GAEA}\left(F_{df}\right), \tag{14}$$

where $f_{GAEA}(.)$ denotes the operation of GAEA mechanism, and $F_{DF}$ is the output of GAEA.

### 3.5 Deep-connected multi-scale residual attention group (DMRAG)

We stack $M$ deep-connected multi-scale residual attention blocks (DMRAB) in each DMRAG to go a further step towards residual learning. The $j$-th DMRAB in the $i$-th DMRAG can be denoted as

$$F_{i,j} = g_{i,j}\left(F_{i,j-1}\right) = g_{i,j}\left(g_{i,j-1}\left(\cdots g_{i,1}\left(F_{i-1}\right)\cdots\right)\right), \tag{15}$$

where $F_{i,j-1}$ and $F_{i,j}$ are the input and output of the $j$-th DMRAB in $i$-th DMRAG respectively, and the $g_{i,j}(.)$ denotes the corresponding function. To enable the network to focus more on informative features, SRSC is added to get the block output

$$F_i = F_{i-1} + W_{SRSC}F_{i,M} = F_{i-1} + W_{SRSC}g_{i,M}\left(g_{i,M-1}\left(\cdots g_{i,1}\left(F_{i-1}\right)\cdots\right)\right), \tag{16}$$

where $W_{SRSC}$ represents the weight of the convolutional layer at the end of the $i$-th DMRAG. With LRSC and SRSC, it is easier to bypass abundant information in the process of training. In order to obtain a more discriminative representation, we rescale the channel features from the local cross-channels as mentioned in Sect. 3.2.

### 3.6 Deep-connected multi-scale residual attention block (DMRAB)

To fully utilize the image features at different scales and the feature correlations for the same sample, we propose deep-connected multi-scale residual attention block (DMRAB). Our DMRAB contains three parts (as shown in Fig. 1): deep-connected multi-scale features fusion, local aware channel attention (LACA) mechanism and local residual connection.

#### 3.6.1 Deep-connected multi-scale features fusion

To obtain the rich features of different scales, we devise a two-branch network composed of different convolution kernels. Therefore, the information between two branches can be shared with each other and the operation can be described as

$$B_{11} = \gamma\left(\omega^1_{3\times3} * H_{s-1} + b^1\right), \tag{17}$$

$$B_{12} = \gamma\left(\omega^1_{5\times5} * H_{s-1} + b^1\right), \tag{18}$$

$$B_2 = \omega^2_{1\times1} * \left[B_{11}, B_{12}\right] + b^2, \tag{19}$$

$$B_{31} = \gamma\left(\omega^3_{3\times3} * B_2 + b^3\right), \tag{20}$$

$$B_{32} = \gamma\left(\omega^3_{5\times5} * B_2 + b^3\right), \tag{21}$$

$$B_4 = \left[B_{11}, B_{12}, B_{31}, B_{32}\right], \tag{22}$$

where $\omega$ and $b$ denote the weights and bias respectively. The subscripts denote the size of convolution kernels, while the superscripts denote the number of layers in which they are reside. $\gamma(x) = max(0, x)$ represents the ReLU function and $[\bullet]$ represents the concatenation operation. The features $B_{11}, B_{12}, B_{31}$ and $B_{32}$ are concatenated through a deep connection (DC) to obtain the feature $B_4$, which aims to provide more sufficient information to the LACA mechanism and further improve the representation capability of our network.

#### 3.6.2 Local aware channel attention (LACA) mechanism

As discussed in Sect. 3.2, we assign different weights to each channel feature on the basis of the learned statistics of local cross-channel interdependencies, so as to utilize the informative features as efficiently as possible.

$$R = W_R\left(LACA\left(B_4\right)\right), \tag{23}$$

where $LACA(.)$ represents the operation of LACA mechanism, and $W_R$ is the weight of the $1 \times 1$ convolution layer, which serves as the function of reducing the number of the feature maps to be the same as the number of input features of DMRAB. $R$ denotes the deep-connected multi-scale residual.

#### 3.6.3 Local residual learning

The efficiency of the network is improved by performing residual learning for each DMRAB. Finally, a deep-connected multi-scale residual attention block (DMRAB) can be denoted as

$$H_s = R \oplus H_{s-1}, \tag{24}$$

where $H_{s-1}$ and $H_s$ represent the input and output of the DMRAB respectively. The operation $\oplus$ denotes the addition of elements.

### 3.7 Implementation details

Now we introduce the implementation details of DMRAN. In the DFES network, we use $N = 5$ DMRAGs with single GAEA module. In each DMRAG, the number of DMRAB is set as$M = 10$. The output feature maps of each convolutional layer are set as 64 expect for the LACA mechanism and the last reconstruction part. We set the kernel size of all convolutional layers except DMRAB and GAEA to $3 \times 3$, where zero-padding strategy is used to fix the feature size. We use ESPCN [25] as our upscale module, followed by the final convolutional layer with three filters to output color images.

## 4 Experiments

### 4.1 Setup

Following [14, 15, 20, 38], we choose 800 HR images from the DIV2K dataset [42] as our training set. Bi-cubic downsample is performed on the training set to obtain the LR images, and we carry out data augmentation on all training images by horizontally flipping and random rotation of $90^\circ$, $180^\circ$ and $270^\circ$. For each training mini-batch, 16 randomly cropped LR color patches with the size of $48 \times 48$ are used as inputs. Our model is optimized by ADAM [43] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$.

For testing, five standard benchmark datasets are used: Set5 [44], Set14 [45], BSD100 [46], Urban100 [47] and Manga109 [48], which differ in contents and styles. Table 1 shows detailed information about 6 benchmark datasets for training and test. All the SR outputs of our proposed DMRAN are evaluated by the PSNR and SSIM metrics on the luminance channel (also known as $Y$ channel) of

**Table 1** Specification of 6 benchmark datasets

| Dataset | Number of samples | Number of classes | Samples per class | Number of test samples |
|---|---|---|---|---|
| DIV2K | 1000 | 8 | ≈ 120 | 100 |
| Set5 | 5 | 5 | 1 | 5 |
| Set14 | 14 | 14 | 1 | 14 |
| BSD100 | 100 | 1 | 100 | 100 |
| Urban100 | 100 | 1 | 100 | 100 |
| Manga109 | 21,142 | 109 | 194 | 109 |

the transformed YCbCr space. The proposed DMRAN are conducted on the PyTorch framework with a NVIDIA RTX 2080Ti GPU.

We implement the proposed DMRAN with Python 3.8.5 and Pytorch1.7.0 and update it with Adam optimizer [12]. 1,000 iterations of back-propagation constitute an epoch. The learning rate is initialized to $10^{-4}$ for all layers and decreases half for every 200 epochs. The training setup consists of an Intel i7-10700 K central processing unit (CPU) and an NVIDIA GeForce RTX 2080Ti graphics processing unit (GPU).

### 4.2 Ablation experiment

In this section, we disassemble our DMRAN to validate the effectiveness of each component, namely multi-scale features fusion (Multi-scale), local aware channel attention (LACA), the deep connection (DC), short range skip connection (SRSC), long range skip connection (LRSC) and global aware external attention (GAEA). Our base model (represented as BASE) is obtained by removing these six parts and replacing the two-branch structure with a single path. Thus, in the BASE scheme, each DMRAB will become a residual block with a kernel size of $3 \times 3$. The numbers of DMRAG and DMRAB are set as 5 and 10 respectively. For fair comparison, the scaling factor is set to 2 and the models are trained with 200 epochs.

**Table 3** Effects of GAEA on Set14, BSD100, Urban100 and Manga109

| GAEA | Set14 | BSD100 | Urban100 | Manga109 |
|---|---|---|---|---|
| × | 33.75/0.9193 | **32.25**/0.9005 | 32.54/0.9323 | 38.88/**0.9776** |
| √ | **33.80/0.9200** | **32.25/0.9006** | **32.59/0.9324** | **38.92**/0.9775 |

The best PSNR (dB) and SSIM values (2 ×) are marked in bold

Table 2 lists the best PSNR (dB) results on Set5. We can see that the performance of BASE network is poor and its PSNR is only 37.84 dB. This shows that simple stacking of residual blocks is not reasonable to construct an effective deep SR network. Then we gradually add each component to the BASE to verify the effectiveness of six components corresponding to the A, B, C, D, E, F, G and H schemes in Table 2. Compare to the BASE, A and B increase PSNR to 37.92 dB and 37.94 dB respectively. The network shows better performance when both the SRSC and LRSC are exploited simultaneously, whose PSNR = 37.98 dB (Table 2, column 5). This suggests that the SRSC and LRSC play a vital role in our DMRAN because they are able to bypass the abundant information during training and test. This also proves the positive effect of our DFES on very deep network.

As shown in Table 2, the effectiveness of our DMRAB module is also demonstrated according to the results of D, E, F and G. Specifically, D improves the performance of C from 37.98 dB to 38.01 dB, this is due to the fact that the rich scale features can be captured by the multi-scale structure, which provide more sufficient clues for the recovery of information. Furthermore, the model E with LACA acheves about 0.03 dB improvement on PSNR compared with the C model, indicating that adaptive attention to channel-wise features is very important to improve the performance of deep SR network. Performance can be further boosted when using both the multi-scale and LACA mechanisms (corresponding to the F scheme). The G scheme achieves obvious quantitative performance gains (38.11 dB PSNR) compare to C. The results of D, E, F and G show that our proposed DMRAB has a significant effect on boosting the

**Table 2** Effects of different components

| Schemes | BASE | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|---|
| Multi-scale | × | × | × | × | √ | × | √ | √ | √ |
| LACA | × | × | × | × | × | √ | √ | √ | √ |
| DC | × | × | × | × | × | × | × | √ | √ |
| SRSC | × | × | √ | √ | √ | √ | √ | √ | √ |
| LRSC | × | √ | × | √ | √ | √ | √ | √ | √ |
| GAEA | × | × | × | × | × | × | × | × | √ |
| PSNR | 37.84 | 37.92 | 37.94 | 37.98 | 38.01 | 38.01 | 38.05 | 38.11 | 38.11 |

The best PSNR (dB) values on Set5 (2 ×) are marked in bold

**Table 4** Quantitative results with BI degradation model for scaling factors × 2, × 3 and × 4

| Methods | Scale | Set5 PSNR/SSIM | Set14 PSNR/SSIM | BSD100 PSNR/SSIM | Urban100 PSNR/SSIM | Manga109 PSNR/SSIM |
|---|---|---|---|---|---|---|
| Bicubic | × 2 | 33.66/0.9299 | 30.24/0.8688 | 29.56/0.8431 | 26.88/0.8403 | 30.80/0.9339 |
| SRCNN [10] | × 2 | 36.66/0.9542 | 32.45/0.9067 | 31.36/0.8879 | 29.50/0.8946 | 35.60/0.9663 |
| FSRCNN [18] | × 2 | 37.05/0.9560 | 32.66/0.9090 | 31.53/0.8920 | 29.88/0.9020 | 36.67/0.9710 |
| VDSR [9] | × 2 | 37.53/0.9590 | 33.05/0.9130 | 31.90/0.8960 | 30.77/0.9140 | 37.22/0.9750 |
| LapSRN [26] | × 2 | 37.52/0.9591 | 33.08/0.9130 | 31.08/0.8950 | 30.41/0.9101 | 37.27/0.9740 |
| MemNet [22] | × 2 | 37.78/0.9597 | 33.28/0.9142 | 32.08/0.8978 | 31.31/0.9195 | 37.72/0.9740 |
| DRRN [16] | × 2 | 37.74/0.9591 | 33.23/0.9136 | 32.05/0.8973 | 31.23/0.9188 | 37.60/0.9736 |
| EDSR [20] | × 2 | 38.11/0.9602 | 33.92/0.9195 | 32.32/0.9013 | 32.93/0.9351 | 39.10/0.9773 |
| DBPN [38] | × 2 | 38.09/0.9600 | 33.85/0.9190 | 32.27/0.9000 | 32.55/0.9324 | 38.89/0.9775 |
| RDN [15] | × 2 | **38.24/0.9614** | 34.01/0.9212 | 32.34/0.9017 | 32.89/0.9353 | 39.18/0.9780 |
| SeaNet [49] | × 2 | 38.08/0.9609 | 33.75/0.9190 | 32.27/0.9008 | 32.50/0.9318 | 38.76/0.9774 |
| DMRAN | × 2 | 38.20/0.9612 | 33.94/0.9213 | 32.33/0.9015 | 32.93/0.9350 | 39.20/0.9780 |
| DMRAN + | × 2 | **38.24/0.9614** | **34.08/0.9222** | **32.38/0.9020** | **33.13/0.9368** | **39.39/0.9785** |
| Bicubic | × 3 | 30.39/0.8682 | 27.55/0.7742 | 27.21/0.7385 | 24.46/0.7349 | 26.95/0.8556 |
| SRCNN [10] | × 3 | 32.75/0.9090 | 29.30/0.8215 | 28.41/0.7863 | 26.24/0.7989 | 30.48/0.9117 |
| FSRCNN [18] | × 3 | 33.18/0.9140 | 29.37/0.8240 | 28.53/0.7910 | 26.43/0.8080 | 31.10/0.9210 |
| VDSR [9] | × 3 | 33.67/0.9210 | 29.78/0.8320 | 28.83/0.7990 | 27.14/0.8290 | 32.01/0.9340 |
| LapSRN [26] | × 3 | 33.82/0.9227 | 29.87/0.8320 | 28.82/0.7980 | 27.07/0.8280 | 32.21/0.9350 |
| MemNet [22] | × 3 | 34.09/0.9248 | 30.00/0.8350 | 28.96/0.8001 | 27.56/0.8376 | 32.51/0.9369 |
| DRRN [16] | × 3 | 34.03/0.9244 | 29.96/0.8349 | 28.95/0.8004 | 27.53/0.8378 | 32.42/0.9359 |
| EDSR [20] | × 3 | 34.65/0.9280 | 30.52/0.8462 | 29.25/0.8093 | 28.80/0.8653 | 34.17/0.9476 |
| DBPN [38] | × 3 | -/- | -/- | -/- | -/- | -/- |
| RDN [15] | × 3 | 34.71/0.9296 | 30.57/0.8468 | 29.26/0.8093 | 28.80/0.8653 | 34.13/0.9484 |
| SeaNet [49] | × 3 | 34.55/0.9282 | 30.42/0.8444 | 29.17/0.8071 | 28.50/0.8594 | 33.73/0.9463 |
| DMRAN | × 3 | 34.67/0.9293 | 30.56/0.8468 | 29.26/0.8090 | 28.83/0.8656 | 34.23/0.9484 |
| DMRAN + | × 3 | **34.73/0.9297** | **30.66/0.8480** | **29.31/0.8100** | **29.02/0.8684** | **34.50/0.9498** |
| Bicubic | × 4 | 28.42/0.8104 | 26.00/0.7027 | 25.96/0.6675 | 23.14/0.6577 | 24.89/0.7866 |
| SRCNN [10] | × 4 | 30.48/0.8628 | 27.50/0.7513 | 26.90/0.7101 | 24.52/0.7221 | 27.58/0.8555 |
| FSRCNN [18] | × 4 | 30.72/0.8660 | 27.61/0.7550 | 26.98/0.7150 | 24.62/0.7280 | 27.90/0.8610 |
| VDSR [9] | × 4 | 31.35/0.8830 | 28.02/0.7680 | 27.29/0.0726 | 25.18/0.7540 | 28.83/0.8870 |
| LapSRN [26] | × 4 | 31.54/0.8850 | 28.19/0.7720 | 27.32/0.7270 | 25.21/0.7560 | 29.09/0.8900 |
| MemNet [22] | × 4 | 31.74/0.8893 | 28.26/0.7723 | 27.40/0.7281 | 25.50/0.7630 | 29.42/0.8942 |
| DRRN [16] | × 4 | 31.68/0.8888 | 28.21/0.7721 | 27.38/0.7284 | 25.44/0.7638 | 29.18/0.8914 |
| EDSR [20] | × 4 | 32.46/0.8968 | 28.80/0.7876 | 27.71/0.7420 | 26.64/0.8033 | 31.02/0.9148 |
| DBPN [38] | × 4 | 32.47/0.8980 | 28.82/0.7860 | 27.72/0.7400 | 26.38/0.7946 | 30.91/0.9137 |
| RDN [15] | × 4 | 32.47/0.8990 | 28.81/0.7871 | 27.72/0.7419 | 26.61/0.8028 | 31.00/0.9151 |
| SeaNet [49] | × 4 | 32.33/0.8970 | 28.72/0.7855 | 27.65/0.7388 | 26.32/0.7942 | 30.74/0.9129 |
| DMRAN | × 4 | 32.42/0.8977 | 28.76/0.7865 | 27.72/0.7416 | 26.58/0.8021 | 30.94/0.9147 |
| DMRAN + | × 4 | **32.58/0.8995** | **28.87/0.7883** | **27.77/0.7428** | **26.81/0.8067** | **31.31/0.9179** |

Best and second best results are bold and underlined, respectively

**Table 5** Comparison of model size and quantitative results (× 2, Set5)

| Methods | DRRN | MemNet | EDSR | DBPN | RDN | SeaNet | DMRAN |
|---|---|---|---|---|---|---|---|
| Parameters | 0.3 M | 0.7 M | 43 M | 10 M | 22.3 M | 7 M | 8.6 M |
| PSNR/dB | 37.74 | 37.78 | 38.11 | 38.09 | 38.24 | 38.08 | 38.20 |

**Table 6** Comparison of running time and quantitative results (× 2, Urban100)

| Methods | FSRCNN [18] | EDSR [20] | MSRN [13] | RCAN [14] | DMRAN |
|---|---|---|---|---|---|
| Time(s) | 0.69 | 0.72 | 1.65 | 8.69 | 6.1 |
| PSNR/dB | 29.88 | 32.93 | 32.22 | 33.13 | 32.93 |

performance of deep SR network, which demonstrate the effectiveness of our DMRAB module.

From Table 2, we can observe that the PSNR of H and G are both 38.11 dB, which seems to indicate that GAEA does not boost the performance of our network. However, when we increase the number of epochs from 200 to 600, H improves the performance of G from 38.18 dB to 38.20 dB, which shows that GAEA may lead to slower network convergence speed, but proves the effectiveness of GAEA. In addition, we further evaluate the models G and H on Set14, BSD100, Urban100 and Manga109 to validate the effectiveness of GAEA, whose best PSNR (dB) and SSIM values are listed in Table 3 ("×" denotes the model G without GAEA, and "$\surd$" denotes the model H with GAEA). As can be seen from columns 2, 4, and 5 of Table 3, GAEA significantly improves the performance of the network. But, as shown in column 3 of Table 3, there is little improvement in network performance. The reason is that the correlations between different samples in different datasets are variable. This proves that GAEA does learn the potential correlations between different samples. Consequently, all of the above comparisons firmly validate the effectiveness of six components of the DMRAN.

### 4.3 Comparisons with state-of-the-art methods

We quantitatively compare our DMRAN with 11 state-of-the-art CNN-based SR approaches on five standard test datasets. These approaches include Bicubic, SRCNN [10], FSRCNN [18], VDSR [9], LapSRN [26], MemNet [22], DRRN [16], EDSR [20], DBPN [38], RDN [15] and SeaNet [49]. Following [15, 20], the self-ensemble strategy is applied to further improve the proposed DMRAN, denoted as DMRAN+. The results of other models are obtained from the published models or papers. Quantitative comparisons for three scaling factors (× 2, × 3, × 4) are reported in Table 4. Our DMRAN+ obtains the best results on all datasets for three scaling factors compared with other methods. Even without using the self-ensemble strategy, our proposed DMRAN can still be comparable to EDSR [20], DBPN [38] and RDN [15] with much lower model parameters (as shown in Table 5), indicating that DMRAN can achieve a good balance between the model complexity and the quantitative performance. In addition, our model outperforms SeaNet [49] in terms of PSNR and SSIM on all datasets. The average PSNR of the five test datasets is improved by 0.25 dB, 0.78 dB and 0.13 dB for three upscaling factors, respectively. Of note, compared with other methods, the DMRAN can achieve significantly enhanced results on five datasets for three scaling factors. The reasons are as follows. First, the deep-connected multi-scale feature fusion structure enables DMRAB to explore more texture dependencies and diverse structure within a single sample, which can provide more sufficient scale features and hierarchical features to guide information recovery. Second, the local aware channel attention (LACA) structure can adaptively rescale the channel features, so that the DMRAB network can focus on more informative features. Third, the global aware external attention (GAEA) structure is able to capture the potential correlations between different samples, which is functionally complementary to DMRAB. Finally, the LRSC and SRSC structures enable our DMRAN to bypass the abundant information, leading to the feature extraction of more effective information.

### 4.4 Model complexity and running time

Table 5 reports the model size and quantitative performance of some mainstream SR methods with the scaling factor × 2 on Set5. Among these methods, DRRN and MemNet have much less parameters at the expense of performance degradation. However, our DMRAN with much fewer parameters can obtain comparable performance to EDSR [20], DBPN [38] and RDN [15]. Compared with SeaNet, our DMRAN achieves much better performance with a little more parameters. As a result, the proposed DMRAN can obtain a good balance between the model complexity and SR performance.

In Table 6, we compare the running time and PSNR scores of DMRAN with those of some strong baselines for the scaling factor × 2 on Urban100. We can see that our model runs slower than FSRCNN and MSRN, but DMRAN can generate the images of higher quality than FSRCNN and MSRN. DMRAN also runs slower than EDSR due to the long range skip connection and the global aware external attention applied in DFES. However, DMRAN has less parameters than EDSR. The PSNR score of DMRAN is slightly lower than that of RCAN, while DMRAN runs faster than it.

### 4.5 Qualitative analysis

To further verify the effectiveness of DMRAN, we conduct a qualitative analysis of the results. In Fig. 3, we visualize the SISR results on three test datasets for the scaling factor × 4. For the image "monarch", it can be seen that the earlier Bicubic method suffers from the serious blurring artifacts and most other methods cannot recover the texture details

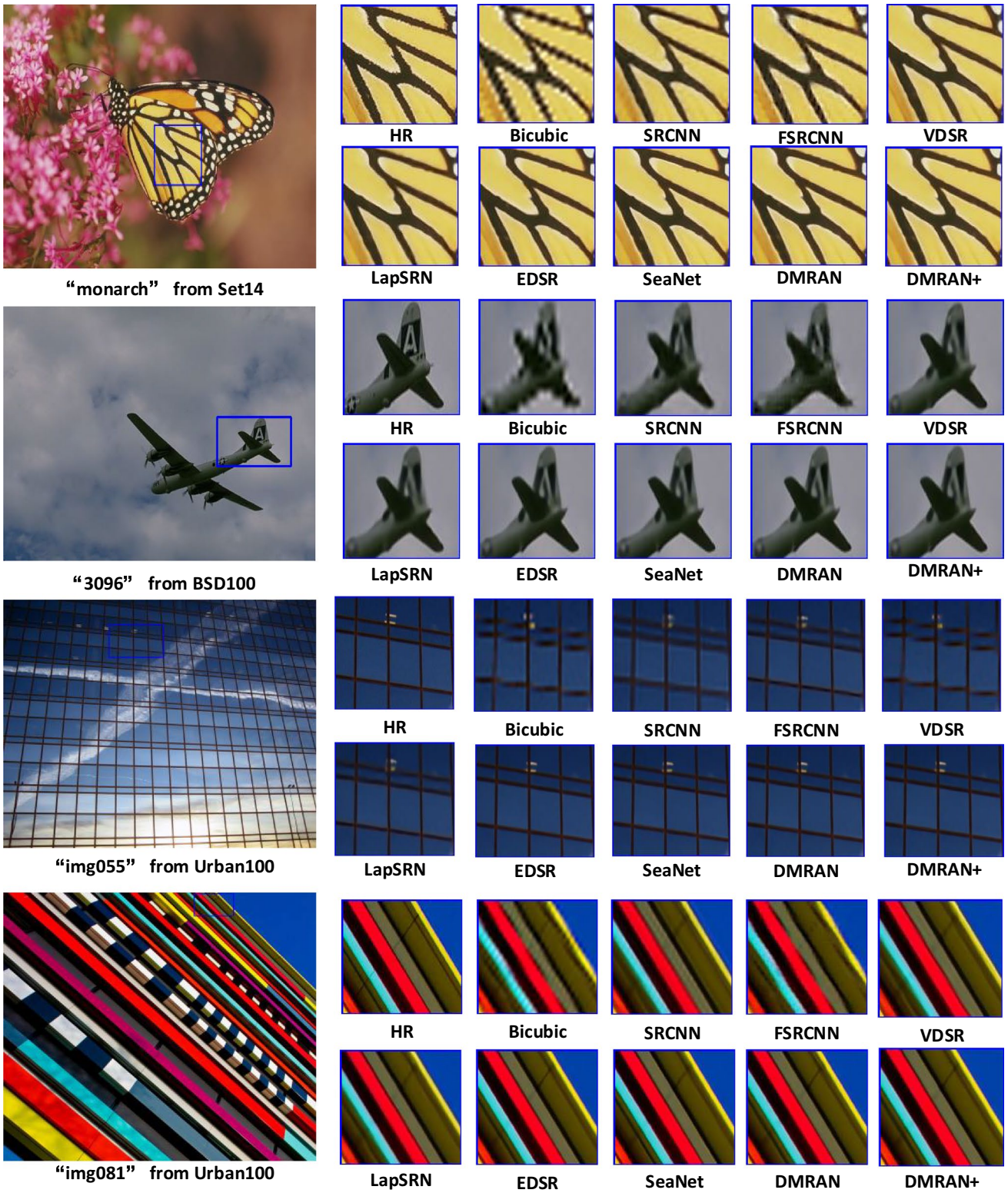


**Fig. 3** Visual comparisons on the test datasets for the scaling factor × 4

well. In contrast, our DMRAN and DMRAN + can generate sharper images with more fine details. For the image "3096" from the BSD100 dataset, the competing methods mostly produce the images with blurring artifacts. Particularly, Bicubic, SRCNN and FSRCNN even fail to recover the "A" and "star" markings on the plane. The "img055" and "img081" are from Urban100, a challenging dataset that contains the abundant contexts of the urban environment. Obviously, compared with other SISR methods, the HR images produced by our methods are more visually comfortable with much clearer grids and lines. Take the "img055" as an example, SRCNN, FSRCNN and LapSRN can recover blurring lines at least, while Bicubic and VDSR even cannot reconstruct lines. However, our DMRAN and DMRAN + not only recover the clear lines, but also clearly reconstruct the lights inside the building, and thus output more faithful results. Therefore, these qualitative comparisons further verify the effectiveness of our proposed DMRAN on SISR.

## 5 Conclusion

In this paper, a deep-connected multi-scale residual attention network (DMRAN) is proposed for accurate SISR. Specifically, the proposed deep-connected multi-scale residual attention block (DMRAB) encourages DMRAN to fully utilize the multi-scale and hierarchical features. Meanwhile, DMRAB enables DMRAN to rescale the channel features adaptively to learn the inherent local interdependencies between channels. In addition to mining the inherent feature correlations of the same sample through local aware channel attention (LACA), we devise a deep feature extraction structure (DFES) to capture the correlations between different samples by incorporating the global aware external attention (GAEA) in the network. The quantitative and qualitative experiments on the benchmark datasets demonstrate that our model has superior performance of SISR to the state-of-the-art approaches.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grants 61801198 and 62276266.

**Data availability** The authors confirm that the data supporting the findings of this study are available within the article.

## References

1. Rasti P, Uiboupin T, Escalera S et al (2016) Convolutional neural network super resolution for face recognition in surveillance monitoring. In: Perales FJ, Kittler J (eds) International conference on articulated motion and deformable objects. Springer International Publishing, Cham, pp 175–184
2. Oktay O, Bai W, Lee M et al (2016) Multi-input cardiac image super-resolution using convolutional neural networks. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W (eds) Medical image computing and computer-assisted intervention. Springer International Publishing, Cham, pp 246–254
3. Luo Y, Zhou L, Shu W et al (2017) Video satellite imagery super resolution via convolutional neural networks. IEEE Geosci Remote Sens Lett 14:2398–2402
4. Keys RG (2003) Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Proces 29:1153–1160
5. Romano Y, Protter, et al (2014) Single image interpolation via adaptive nonlocal sparsity-based modeling. IEEE Trans Image Process 23:3085–3098
6. Zhang M, Desrosiers C (2018) High-quality image restoration using low-rank patch regularization and global structure sparsity. IEEE Trans Image Process 28:868–879
7. Ren C, He X, Pu Y et al (2019) Enhanced non-local total variation model and multi-directional feature prediction prior for single image super resolution. IEEE Trans Image Process 28:3778–3793
8. Kim JH, Lee JS (2018) Deep residual network with enhanced upscaling module for super-resolution. IEEE/CVF Conf Comput Vis Patt Recogn Workshops. https://doi.org/10.1049/ell2.12689
9. Kim J, Lee JK, Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. IEEE Conf Comput Vis Patt Recogn. https://doi.org/10.1109/CVPR.2016.182
10. Dong C, Loy CC, He K et al (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38:295–307
11. Chang H, Yeung DY, Xiong Y (2004) Super-resolution through neighbor embedding. IEEE Comput Soc Conf Comput Vis Patt Recogn 34:275–282
12. Qin J, Huang Y, Wen W (2020) Multi-scale feature fusion residual network for single image super-resolution. Neurocomputing 379:334–342
13. Li J, Fang F, Mei K et al (2018) Multi-scale residual network for image super-resolution. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) European conference on computer vision. Springer International Publishing, Cham, pp 527–542
14. Zhang Y, Li K, Li K et al (2018) Image super-resolution using very deep residual channel attention networks. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) European conference on computer vision. Springer International Publishing, Cham, pp 294–310
15. Zhang Y, Tian Y, Kong Y et al (2018) Residual dense network for image super-resolution. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) IEEE/CVF Conference on Computer Vision and Pattern Recognition. Springer International Publishing, Cham, pp 2472–2481
16. Ying T, Jian Y, Liu X (2017) "Image Super-Resolution via Deep Recursive Residual Network," in IEEE Conference on Computer Vision & Pattern Recognition., , pp. 2790–2798
17. Kim J, Lee J K, Lee K M (2016) "Deeply-Recursive Convolutional Network for Image Super-Resolution," in IEEE Conference on Computer Vision and Pattern Recognition., , pp.1637–1645
18. Chao D, Chen CL, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: Leibe B, Matas J, Sebe N, Welling M (eds) European conference on computer vision. Springer International Publishing, Cham, pp 391–407
19. He K, Zhang X, Ren S, et al (2016) "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition, pp.770–778
20. Lim B, Son S, Kim H, et al (2017) "Enhanced Deep Residual Networks for Single Image Super-Resolution," in IEEE Conference

on Computer Vision and Pattern Recognition Workshops. IEEE, pp.1132–1140
21. Huang G, Liu Z, Laurens V, et al (2016) "Densely Connected Convolutional Networks," in IEEE Conference on Computer Vision and Pattern Recognition., , pp. 2261–2269
22. Tai Y, Yang J, Liu X, et al (2017) "MemNet: A Persistent Memory Network for Image Restoration," in IEEE International Conference on Computer Vision., , pp. 4549–4557
23. Tong T, Li G, Liu X, et al (2017) "Image Super-Resolution Using Dense Skip Connections," in IEEE International Conference on Computer Vision., , pp. 4809–4817
24. Guo, M.H., et al (2021) "Beyond Self-attention: External Attention using Two Linear Layers for Visual Tasks," CoRR, vol. abs/2105.02358
25. Shi W, Caballero J, F Huszár, et al (2016)"Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883
26. Lai W S, Huang J B, Ahuja N, et al (2017) "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," in IEEE Conference on Computer Vision & Pattern Recognition., , pp.5835–5843
27. Xiao M, Chuhua S, Yubin Y (2016) "Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections," CoRR, vol.abs/1606.08921
28. Szegedy C, Liu W, Jia Y, et al (2014) "Going Deeper with Convolutions," in IEEE Computer Society., 2014, pp. 1–9
29. Liu Y, Zhang X, Wang S, et al (2020) "Progressive Multi-Scale Residual Network for Single Image Super-Resolution," CoRR, vol.abs/2007.09552
30. Xiong C, Shi X, Gao Z et al (2020) Attention augmented multi-scale network for single image super-resolution. Appl Intell 51:935–951
31. Jie H, Li S, Gang S, et al (2017) "Squeeze-and-Excitation Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 7132–7141
32. Dai T, Cai J, Zhang Y, et al (2019) "Second-order Attention Network for Single Image Super-Resolution," in IEEE Conference on Computer Vision and Pattern Recognition., pp. 11065–11074
33. Wang X, Girshick R, Gupta A, et al (2017) "Non-local Neural Networks," CoRR, vol.abs/1711.07971
34. Zhang Y, K Li, K Li, et al (2019) "Residual non-local attention networks for image restoration," CoRR, vol.abs/1903.10082
35. Mei Y, Fan Y, Zhou Y, et al (2020) "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," In IEEE/CVF conference on computer vision and pattern recognition., , pp. 5689–5698
36. Liu Z, Huang J, Zhu C et al (2021) Residual attention network using multi-channel dense connections for image super-resolution. Appl Intell 51:85–99
37. Hu X, Mu H, Zhang X, et al (2020) "Meta-SR: A magnification-arbitrary network for super-resolution," In: IEEE conference on computer vision and pattern recognition., 2020, pp. 1575–1584
38. Haris M, Shakhnarovich G, Ukita N (2018) "Deep back-projection networks for super-resolution," *arXiv. arXiv*, , pp. 1664–1673
39. Sajjadi M, Scholkopf B, Hirsch M (2017) "EnhanceNet: single image super-resolution through automated texture synthesis," In: IEEE International Conference on Computer Vision., , pp. 4501–4510
40. Ledig C, Theis L, F Huszar, et al (2016) "Photo-realistic single image super-resolution using a generative adversarial network," In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 105–114
41. Wang Q, et al (2020) "ECA-Net: efficient channel attention for deep convolutional neural networks," In: IEEE conference on computer vision and pattern recognition., 2020, pp. 11531–11539
42. Agustsson E, Timofte R (2017) "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.1122–1131
43. Kingma D, Ba J (2015) "Adam: a method for stochastic optimization," in international conference on learning representations
44. Bevilacqua M, Roumy A, Guillemot C, et al (2012) "Neighbor embedding based single-image super-resolution using Semi-Nonnegative Matrix Factorization," in IEEE International Conference on Acoustics, pp.1289–1292
45. Zeyde R, Elad M, Protter M (2010) On single image scale-up using sparse-representations. In: Boissonnat J-D, Chenin P, Cohen A, Gout C, Lyche T, Mazure M-L, Schumaker L (eds) International conference on curves and surfaces. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 711–730
46. Martin D, Fowlkes C, Tal D, et al (2002) "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," In: IEEE International Conference on Computer Vision, pp. 416–425
47. Huang J B, Singh A, Ahuja N (2015) "Single image super-resolution from transformed self-exemplars," In: IEEE Conference on Computer Vision and Pattern Recognition., pp. 5197–5206
48. Matsui Y, Ito K, Aramaki Y et al (2017) Sketch-based manga retrieval using manga109 dataset. Multimed Tools Appl 76:21811–21838
49. Fang F, Li J, Zeng T (2020) Soft-edge assisted network for single image super-resolution. IEEE Trans Image Process 29:4656–4668