Progressive low/high-resolution Space Attention Fusion Network for Single Image Super-Resolution

Chengzu Zhong and Yue Zhou*
Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China
*Email: zhouyue@sjtu.edu.cn

Abstract. The general single image super-resolution methods mainly extract features from the high-resolution (HR) space by the pre-upscaling step at the beginning of the network or from the low-resolution (LR) space before the post-upscaling step at the end of the network. The former way requires high computation as well as misleading the network by wrong artificial priors. The latter way cannot learn mapping well by only conducting simple operations in HR space. In this paper, we aim to utilize the features from LR and HR space more efficiently and propose the novel network, which applies a frequency-slicing mechanism to divide features into LR and HR space, a direction-aware fusion residual group to extract distinctive features in LR space and an attention fusion module to recalibrate features in HR space. The experimental results demonstrate that our model is superior to the state-of-the-art methods upon quantitative metrics and visual quality.

1. Introduction
The purpose of the Single Image Super-Resolution (SISR) is to restore the high-resolution (HR) image with abundant textures and structures from its low-resolution (LR) counterpart. Based on the setting of the up-scaling factor $\alpha$, the model needs to learn accurately $\alpha \times \alpha$ times pixels in HR space. It can be observed that SISR is an ill-posed inverse problem, many HR images correspond to the same LR image.

In recent years, the development of deep learning has made great progress in computer vision, especially since the introduction of convolution neural networks (CNN) has a powerful learning ability to promote much better performance than conventional methods in the SISR study. The first representative method with CNN architecture, called SRCNN proposed by Dong et al. [1], employs bicubic interpolation to upscale the LR input image to the desired size and adopt three convolutional layers to learn the non-linear mapping between LR and HR image pairs.

The shallow layers of SRCNN are difficult to learn the finer information in HR space leading to slow convergence. To improve this defect, VDSR [2] expands the network depth from 3 to 20 layers to learn more informative feature maps by larger receptive fields. LR and HR image pairs retaining the relevant low-frequency information look similar to each other, however, unclear LR images still lose details from HR images. In other words, recovering high-frequency information is the key in the most SISR literature. VDSR exploits residual learning to directly reconstruct the residual between pre-upscaled LR image and its HR image. The subsequent SISR methods start to utilize this advantage of sparse residual restoration so as to the training efficiency raises significantly. DRCN [3] introduces global residual learning strategy and recursive layers to iteratively share the same parameters. Tai et al.
The overall architecture of the proposed Progressive low/high-resolution Space Attention Fusion Network (PAFN) divides into two regions depending on the size of feature maps, which are the LR input image size in low-resolution feature space and the HR output image size in high-resolution feature space (Mark as light yellow and light pink background respectively).

[4,5] propose two modified network architectures, where DRRN [4] applies deeper recursive layers with local-and-global residual learning and MemNet [5] employs memory blocks consisting of a recursive unit and a gate unit to preserve all iterative outputs. The above-mentioned methods all use the bicubic-upscaled image as input, which inevitably requires high computation and high memory requirement as well as misleading the network by wrong artificial priors. Besides networks do not utilize the features in the LR space, which highly correlated to the HR image.

To solve the problem of the pre-upscaled input image, many methods inspired by FSRCNN [6] and ESPCN [7] use a deconvolution layer or a sub-pixel convolution layer respectively at the end of the network to upscale the images. Benefiting from this improvement, the network can use smaller feature maps extracted from LR space to accelerate training process and get a better performance by deeper layers. EDSR [8] designs a very deep and wide network with multiple simplified residual blocks. SRDesNet [9] introduces dense connections to efficiently utilize hierarchical features. Combining the advantages of EDSR and SRDesNet, RDN [10] employs residual dense blocks comprised of dense feature fusion and local residual learning to enhance the transfer of low-level features and gradient flows. However, those methods that extract features from LR space achieve excellent performance largely by the increase of network scale. Conducting only simple operations after the up-scaling step cannot learn mapping well in HR space.

In summary, it is necessary for the trade-off between LR and HR space so that we propose the novel network to make full use of different features from these two spaces by frequency-slicing mechanisms.

2. Related Work
The deep learning networks of single image super-resolution can roughly classify into two types. Both devote to explore more distinguishing features, but the difference is whether to extract features in LR space or HR space mainly. According to our perspective, these two kinds of methods have their advantages and weaknesses:

Firstly, we discuss the methods of extracting features in HR space. DCSR [11] mixes standard convolutions and dilated convolutions in each layer to capture larger scale contextual information. BySEDenseSR [12] adopts bypass convolutions to inhibit noise, which improves the stability of middle-layer features. CMSC [13] utilizes stacked multi-scale cross (MSC) modules to fuse complementary information under different receptive fields. Due to the input image is in the same size as the output image, those methods can apply pixel-to-pixel mapping between the corresponding location accurately. However, the wrong artificial priors by the bicubic upsampling operation will limit the performance.

Secondly, the methods of extracting features in LR space, MSDN [14] employs different dilated convolutions to gather multi-scale information. MSRN [15] uses different convolutional kernels to exploit local features. MRFN [16] is mainly comprised of multi-receptive-field modules, which extract features from different receptive fields. Thanks to the small size of feature maps, those studies can
make network structure deeper by stacking more basic building blocks. However, conducting only a convolutional layer or weighted averaging predictions in HR space cannot learn accurate mapping.

Finally, based on the above discussion, we propose a novel network to balance these two spaces. The frequency-slicing mechanisms that outputs the recovered features to HR space, and unrecovered features still retain in LR space. In so doing, we can operate both features step by step. For LR space, introducing the direction-aware fusion residual groups extract distinctive features. For HR space, we mainly use the attention fusion modules to recalibrate the importance of recovered features without complex computation.

3. Proposed Method
In this paper, degradation factors are not our research emphasis. To compare performance with previous methods, we only apply bicubic degradation to the ground-truth images \((I_{HR})\) to generate the LR-HR image pairs. Our view for the SISR task is to utilize the features from LR and HR space more efficiently. The reason why these two spaces are so important: For LR space, its feature maps highly correlated to HR feature maps. For HR space, its feature maps accurately match pixels to the HR image. Based on these insights, we design the novel architecture called Progressive low/high-resolution space Attention Fusion Network (PAFN), which mainly consists of five parts: feature extraction layer, frequency-slicing mechanisms (FSM), direction-aware fusion residual groups (DFRG), attention fusion modules (AFM) and reconstruction stage. As shown in Figure 1, we define the bicubic-downscaled input image as \(I_{LR}\) and the super-resolved output image as \(I_{SR}\). In our PAFN, we first apply the feature extraction layer to extract shallow feature maps \(F_0\) from \(I_{LR}\), denoted as:

\[
F_0 = H_{FE}(I_{LR})
\]

where \(H_{FE}(\cdot)\) is the \(3 \times 3\) convolutional layer with the filter size of \(3 \times 3\). Unless explicitly stated in this paper, all convolutional layers have 64 filters. \(F_0\) then fed into the first FSM (i.e., \(X^{FSM}_1\)). The details of the other components are described in the following subsections.

3.1. Frequency-Slicing Mechanism (FSM)
The frequency-slicing mechanism (FSM) indicated by the red box in Figure 1 is comprised of the up-scaling unit and the down-scaling unit. The \(i\)-th up-scaling unit \(F^{up}_i\) aims at retaining recovered information in the HR space, expressed as:

\[
F^{up}_i = H_{c}(H_{\alpha}(X^{FSM}_i)), i = 1, 2, \cdots, N
\]

where \(H_{\alpha}(\cdot)\) denotes the nearest-neighbor upsampling layer with the scale factor \(\alpha\), and \(H_c(\cdot)\) is the \(3 \times 3\) convolutional layer followed by Relu. \(X^{FSM}_i\) means the input of \(i\)-th FSM. The two reasons why we use the nearest-neighbor upsampling layer followed by two convolutions are relatively fewer parameters against a sub-pixel convolution layer [7] and avoid the problem of artifacts caused by a deconvolution layer [17]. In our experiment, this way still obtains similar performance to a sub-pixel convolution layer or a deconvolution layer.

The \(i\)-th down-scaling unit \(F^{down}_i\) reduces the size of feature maps from the \(i\)-th up-scaling unit to LR space and then minuses the input of \(i\)-th FSM. It aims at conveying unrecovered residuals into the \(i\)-th DFRG for extracting more details from feature maps in LR space, expressed as:

\[
F^{down}_i = H_{\alpha}(F^{up}_i) - X^{FSM}_i, i = 1, 2, \cdots, N - 1
\]

where \(H_{\alpha}(\cdot)\) denotes the convolutional layer downsampling the feature maps to LR space.

Inspired by EBRN [18], the frequency-slicing mechanism can gradually split different frequencies of information. In LR space, \(F^{down}_i\) delivers higher-frequency information to \(i\)-th DFRG.
In HR space, $F_{i+1}^{up}$ restores lower-frequency information to combine with more details $F_i^{up}$ by AFM. Notably, the last FSM only contains the up-scaling unit.

3.2. Direction-aware Fusion Residual Group (DFRG)

To extract better deep features in LR space, we propose the direction-aware fusion residual group (DFRG) that is constructed with three hierarchical feature fusion modules (HFFM) and each HFFM contains four direction-aware blocks (DAB), as shown in Figure 2. Many improved SISR methods utilize larger kernel convolutions to exploit local features resulting in high computation [15,13,19] or employ dilated convolution to enlarge receptive fields by inserting zeros in the convolution kernels, and this way causes the loss of local features [11,14,20]. The main component DAB is two combinations of $1 \times k$ and $k \times 1$ convolution by a series of connections, which can extract distinctive features (i.e., horizontal, vertical, and diagonal directions) with fewer parameters compared to $k \times k$ convolution but preserve more local information. In this manner, the convolutions are not overlapped with each other that ignore relations between adjacent pixels. Thus, we further use two $3 \times 3$ convolutions with Relu to solve this problem. For fusing the direction-aware features efficiently, all the outputs of DAB with different $k$ are summed hierarchically before being concatenated with them.

3.3. Attention Fusion Module (AFM)

As illustrated in Figure 3, we design the attention fusion module (AFM) to fuse different up-scaling units by channel attention and spatial attention. Our two attention structures are the same as CBAM [21] but adopt different connection way to achieve the fusion purpose.

Channel Attention: The first and the rest $F_i^{low}$ of $i$-th AFM are the shallow up-scaling unit $F_i^{up}$ and $F_{i-1}^{AFM}$ (i.e., the output of $i$-1 AFM) restoring low-frequency feature maps, which need to properly discriminate their importance for reconstruction by channel attention (marked as the orange dashed box in Figure 3) The inter-channel feature maps are not equally important through convolution operations. Channel attention recalibrates the weights among channels to heighten the significant
Figure 3. Attention Fusion Module (AFM).

feature maps and depress redundant ones. In this way, the network can efficiently learn more principal features so as to enhance the representational ability and improve the performance. We use global max pooling and global mean pooling operations to aggregate information. The channel attention can be formulated as:

$$CA(F) = \sigma(H_U \delta(H_D(\text{MaxPool}(F)))) + H_U \delta(H_D(\text{MeanPool}(F))))$$

(4)

where $CA(\cdot)$ is channel attention operation identical to CBAM [21]. $\sigma(\cdot)$ and $\delta(\cdot)$ denote the sigmoid function and Relu, respectively. $H_U$ and $H_D$ are the $1 \times 1$ convolutional layer to form a bottleneck structure with a reduction ratio of $r$.

Then, we conduct channel attention on the concatenated features to adjust the channel weight of $F_{i,\text{low}}$, represented as:

$$F_{i,\text{CA}} = CA([F_{i,\text{low}}, F_{i,\text{high}}]) \otimes F_{i,\text{low}}, i = 1,2, \cdots, N$$

(5)

where $F_{i,\text{low}}$ is channel-wise refined features, and $[\cdot, \cdot]$ means concatenation operation along channel axis. $\otimes$ is element-wise multiplication.

**Spatial Attention:** The $F_{i,\text{high}}$ of $i$-th AFM is the deep up-scaling unit $F_{i,\text{up}}$ containing more high-frequency information, where using spatial attention (marked as the purple dashed box in Figure 3) focuses on informative regions, such as edges, corners, etc. Considering that channel attention employs global pooling operations to obtain global descriptor leading to the loss of spatial information, spatial attention complements this weakness by applying channel-axis pooling operations to get channel information that can learn how to highlight or reduce the spatial weights, represented as:

$$SA(F) = \sigma(H_S ([\text{MaxPool}(F), \text{MeanPool}(F)]))$$

(6)

$$F_{i,\text{SA}} = SA([F_{i,\text{low}}, F_{i,\text{high}}]) \otimes F_{i,\text{high}}, i = 1,2, \cdots, N$$

(7)

where $SA(\cdot)$ is spatial attention operation identical to CBAM [21], $H_S(\cdot)$ means the $7 \times 7$ convolutional layer, and $F_{i,\text{high}}$ denotes pixel-wise modified features.

The output of $i$-th AFM (i.e., $F_{i,\text{AFM}}$) is the element-wise summation of improved features ($F_{i,\text{CA}}$ and $F_{i,\text{SA}}$) that captures major information in the channel and spatial dimensions at the same time. Finally, all the outputs of AFM are concatenated to preserve long-term information and then operated by the $1 \times 1$ convolutional layer $H_f$ for adaptive fusion and dimensionality reduction. We also apply
global residual learning strategy by summation of the first up-sampling unit that can promote a fast convergence. The reconstruction of SR image process is formulated as:

\[
I_{SR} = H_{PAFN}(I_{LR}) = H_f(H_f([F_{AFM}, F_{AFM}^2, \ldots, F_{AFM}^N]) + F_f)
\]  

(8)

where \(H_f(\cdot)\) denotes the convolutional layer with three filters to produce super-resolved output image (RGB channels), and \(H_{PAFN}(\cdot)\) is the function of PAFN.

For a fair comparison, our network is optimized with the same loss function as previous works (e.g., L1 loss function). Given a training set with N LR-HR image pairs denoted by \{\(I_{LR}^i, I_{HR}^i\}\}_{i=1}^N. Hence, the objective function of our network can be written as:

\[
L(\theta) = \frac{1}{N} \|H_{PAFN}(I_{LR}^i) - I_{HR}^i\|_1
\]  

(9)

where \(\theta\) denotes the parameter set of the PAFN network.

4. Experimental Results

In this section, we systematically evaluate the performance of the PAFN network and clearly explain the details of the training process. The experiments demonstrate the effectiveness of network structure and are superior to the state-of-the-art methods upon quantitative metrics and visual quality.

4.1. Training Settings

Following recent CNN-based methods, our training set is 800 training images from DIV2K dataset and five benchmark datasets as our testing sets, including Set5, Set14, BSD100, Urban100 and Manga109. We expand the training set by random horizontal and vertical flips as well as 90° rotations. In each training batch, 16 LR patches randomly cropped from LR image as input with the size of 48×48. Our model is trained by ADAM optimizer [22] with the parameters of \(\beta_1 = 0.9\), \(\beta_2 = 0.999\), and \(\varepsilon = 10^{-8}\). The learning rate is initialized as \(10^{-4}\) and halved at every 2×10^5 iterations. We use Pytorch framework to implement the proposed PAFN network on an NVIDIA TITAN Xp GPU and N is set to 5. To evaluate SISR performance, we calculate peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [23] metrics on the Y channel of transformed YCbCr space.

4.2. Ablation Study

We also employ the ablation study to verify the effectiveness of three proposed components. Table 1 is divided into two parts: (left) the network structure without FSM, which is transformed into a single post-upscaling step at the end of the network; (right) Our proposed network structure with FSM, which can more fully exploits the features from LR and HR space. The model with FSM gets better performance (+0.25 dB) than the model without FSM. This observation can prove that our main idea in this paper is effective. Moreover, adopting DAB can greatly improve the performance from 32.03 dB to 32.40 dB by the extraction of features in different directions. Notably, we replace DAB by dilated convolutions with different dilation rates so as to prove the effectiveness of DAB and the weakness of dilated convolution.

Table 1. Ablation study on the performance of our proposed components in terms of average PSNR on Urban100 (×2).

| Components | General post-upsampling network | Our proposed network structure |
|------------|---------------------------------|-------------------------------|
| FSM        | ✗ ✗ ✗ ✗                          | ✓ ✓ ✓ ✓                        |
| AFM        | ✗ ✓ ✗ ✓                          | X ✓ X ✓                        |
| DAB        | ✗ ✓ ✓ ✓                          | X X ✓ ✓                        |
| PSNR       | 32.03 32.07 32.40 32.43           | 32.28 32.36 32.71 32.78        |
4.3. Comparisons with State-of-the-art Methods
To test the effectiveness of our model, we compare the proposed PAFN with 6 state-of-the-art SISR methods: SRCNN [1], VDSR [2], DRCN [3], MemNet [5], MRFN [16] and MSRN [15]. As illustrated in Table 2, our proposed model outperforms other methods in all scales. The PAFN+ model achieves the best performance due to the self-ensemble strategy [8]. As shown in Figure 4, our PAFN model can restore sharper and clearer edges, unlike the other methods that generate blurred-curve lines with burrs. Moreover, our model maintains similar textures and structures to the HR image without any blurring artifacts. According to this visual contrast, PAFN obtains more faithful results and can restore sharper and clearer features in LR space and reweighting the importance of features in HR space all enhance the representation ability of our model.

Table 2. Quantitative results on five benchmark datasets based on average PSNR/SSIM for scale factor ×2, ×3 and ×4. The red and blue indicate the best and second-best performance respectively. (+ denotes self-ensemble strategy)

| Method      | α | Set5   | Set14    | B100     | Urban100 | Manga109 |
|-------------|---|--------|----------|----------|----------|----------|
| Bicubic     |   | 33.66/0.929 | 30.24/0.868 | 29.56/0.843 | 26.88/0.840 | 30.80/0.933 |
| SRCNN [1]   |   | 36.66/0.954 | 32.45/0.906 | 31.36/0.887 | 29.50/0.894 | 35.60/0.966 |
| VDSR [2]    |   | 37.53/0.958 | 33.03/0.912 | 31.90/0.896 | 30.76/0.914 | 37.22/0.975 |
| DRCN [3]    |   | 37.63/0.958 | 33.04/0.911 | 31.85/0.894 | 30.75/0.913 | 37.63/0.972 |
| MemNet [5]  |   | 37.78/0.959 | 33.28/0.914 | 32.08/0.897 | 31.31/0.919 | 37.72/0.974 |
| MRFN [16]   |   | 37.98/0.961 | 33.41/0.916 | 32.14/0.900 | 31.45/0.922 | 38.29/0.976 |
| MSRN [15]   |   | 38.08/0.960 | 33.74/0.917 | 32.23/0.901 | 32.22/0.932 | 38.82/0.986 |
| PAFN (Ours) |   | 38.17/0.961 | 33.93/0.920 | 32.31/0.901 | 32.78/0.934 | 39.10/0.978 |
| PAFN+ (Ours)|   | 38.24/0.962 | 34.03/0.921 | 32.37/0.902 | 32.97/0.936 | 39.30/0.979 |

| Method      | α | Set5   | Set14    | B100     | Urban100 | Manga109 |
|-------------|---|--------|----------|----------|----------|----------|
| Bicubic     | 2 | 30.39/0.868 | 27.55/0.774 | 27.21/0.738 | 24.46/0.734 | 26.95/0.855 |
| SRCNN [1]   | 2 | 32.75/0.909 | 29.30/0.821 | 28.41/0.786 | 26.24/0.798 | 30.48/0.911 |
| VDSR [2]    | 2 | 33.66/0.921 | 29.77/0.831 | 28.82/0.797 | 27.14/0.827 | 32.01/0.934 |
| DRCN [3]    | 2 | 33.82/0.921 | 29.76/0.831 | 28.80/0.796 | 27.15/0.827 | 32.31/0.932 |
| MemNet [5]  | 2 | 34.09/0.924 | 30.01/0.835 | 28.96/0.800 | 27.56/0.837 | 32.51/0.936 |
| MRFN [16]   | 2 | 34.21/0.927 | 30.03/0.836 | 28.99/0.803 | 27.53/0.839 | 32.82/0.940 |
| MSRN [15]   | 2 | 34.38/0.926 | 30.34/0.839 | 29.08/0.804 | 28.08/0.855 | 33.44/0.942 |
| PAFN (Ours) | 2 | 34.67/0.929 | 30.58/0.847 | 29.24/0.809 | 28.77/0.865 | 34.13/0.948 |
| PAFN+ (Ours)| 2 | 34.79/0.930 | 30.67/0.848 | 29.30/0.810 | 28.95/0.868 | 34.42/0.950 |

| Method      | J | Set5   | Set14    | B100     | Urban100 | Manga109 |
|-------------|---|--------|----------|----------|----------|----------|
| Bicubic     | 4 | 28.42/0.810 | 26.00/0.702 | 25.96/0.667 | 23.14/0.657 | 24.89/0.786 |
| SRCNN [1]   | 4 | 30.48/0.862 | 27.50/0.751 | 26.90/0.710 | 24.52/0.722 | 25.78/0.855 |
| VDSR [2]    | 4 | 31.35/0.883 | 28.01/0.767 | 27.29/0.725 | 25.18/0.752 | 28.83/0.887 |
| DRCN [3]    | 4 | 31.53/0.885 | 28.02/0.767 | 27.23/0.723 | 25.14/0.751 | 28.98/0.881 |
| MemNet [5]  | 4 | 31.74/0.889 | 28.26/0.772 | 27.40/0.728 | 25.50/0.763 | 29.42/0.894 |
| MRFN [16]   | 4 | 31.90/0.892 | 28.31/0.775 | 27.43/0.731 | 25.46/0.765 | 29.57/0.896 |
| MSRN [15]   | 4 | 32.07/0.890 | 28.60/0.775 | 27.52/0.727 | 26.04/0.789 | 30.17/0.903 |
| PAFN (Ours) | 4 | 32.45/0.898 | 28.81/0.787 | 27.71/0.741 | 26.60/0.802 | 31.01/0.915 |
| PAFN+ (Ours)| 4 | 32.59/0.900 | 28.90/0.789 | 27.77/0.743 | 26.79/0.806 | 31.32/0.918 |

5. Conclusion
In this paper, we propose a novel CNN architecture network for utilizing the features from LR and HR space more efficiently. Different from the previous SISR methods considering only one space, our
model employs the frequency-slicing mechanisms to progressively output features in LR and HR space. In this way, we utilize the convolutions with horizontal and vertical kernels to extract directional features in LR space as well as apply spatial and channel attention to recalibrate the importance of different-level features and fuse them step by step in HR space. The experimental results demonstrate that our model can restore accurate structures and clear edges. Based on the ablation study, it can verify the essentiality of balancing these two spaces and the effectiveness of extracting valuable features in LR space. Therefore, our future work is to explore the features from LR and HR space more comprehensively to achieve better performance.

Figure 4. The visual contrast experimental results of SISR methods on Urban100 (×4).

References
[1] Dong C, et al 2014 Learning a Deep Convolutional Network for Image Super-Resolution (ECCV Springer International Publishing) pp. 184–199
[2] Kim J, Lee J K and Lee K M 2016 Accurate Image Super-Resolution Using Very Deep Convolutional Networks 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV) pp. 1646-1654
[3] Kim J., Lee J. K. and Lee K. M. 2016 Deeply-Recursive Convolutional Network for Image Super-Resolution,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV) pp. 1637-1645

[4] Tai Y, Yang J, Liu X 2017 Image Super-Resolution via Deep Recursive Residual Network IEEE Conference on Computer Vision & Pattern Recognition pp 3147–3155

[5] Tai Y, Yang J, Liu X. and Xu C 2017 MemNet: A Persistent Memory Network for Image Restoration 2017 IEEE International Conference on Image Vision (ICCV) (Venice, 2017) pp. 4549-4557

[6] Dong C, Loy C C, Tang X 2016 Accelerating the Super-Resolution Convolutional Neural Network European Conference on Computer Vision (ECCV) pp. 391-407.

[7] Shi W. et al. 2016 Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV) pp. 1874-1883

[8] Lim B., Son S., Kim H., Nah S. and Lee K. M. 2017 Enhanced Deep Residual Networks for Single Image Super-Resolution 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Honolulu, HI) pp. 1132-1140

[9] Tong T, Li G, Liu X, et al. 2017 Image Super-Resolution Using Dense Skip Connections,” IEEE International Conference on Computer Vision. IEEE Computer Society pp. 4799–4807

[10] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y 2018 Residual dense network for image super-resolution IEEE Conference on Computer Vision and Pattern Recognition pp. 2472–2481

[11] Zhang Z., Wang X. and Jung C. 2019 DCSR: Dilated Convolutions for Single Image Super-Resolution IEEE Transactions on Image Processing, 28(4) 1625-35

[12] Pan Z, Zhang Y, Guo X. 2019 Image super-resolution model based on bypass convolution and feature weighting IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) pp. 1866–1871

[13] Hu Y, Gao X, Li J, et al. 2018 Single image super-resolution via cascaded multi-scale cross network arXiv preprint arXiv:1802.08808

[14] Chang C. and Chien S. 2019 Multi-scale Dense Network for Single-image Super-resolution IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Brighton, United Kingdom) pp. 1742-1746

[15] Li J, Fang F, Mei K, et al. 2018 Multi-scale Residual Network for Image Super-Resolution European Conference on Computer Vision (ECCV) pp. 517–532

[16] He Z. et al. 2020 MRFN: Multi-Receptive-Field Network for Fast and Accurate Single Image Super-Resolution IEEE Transactions on Multimedia 22(4) 1042-54

[17] Odena, et al. 2016 Deconvolution and Checkerboard Artifacts Distill, 2016. http://doi.org/10.23915/distill.00003

[18] Qiu Y, Wang R, Tao D and Cheng J 2019 Embedded Block Residual Network: A Recursive Restoration Model for Single-Image Super-Resolution IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, South Korea) pp. 4179-4188

[19] Wu H. et al. 2019 Multi-grained Attention Networks for Single Image Super-Resolution IEEE Transactions on Circuits and Systems for Video Technology arXiv:1909.11937

[20] Muqet A, Iqbal M T B and Bae S 2019 HRAN: Hybrid Residual Attention Network for Single Image Super-Resolution IEEE Access 7 pp. 137020-9

[21] Woo S, Park J, Lee J Y, et al 2018 CBAM: Convolutional Block Attention Module European Conference on Computer Vision (Springer, Cham) pp. 3–19

[22] Kingma D P, Ba J 2014 Adam: A Method for Stochastic Optimization Computer Science, 2014

[23] Zhou W, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity IEEE Transactions on Image Processing 13(4) pp. 600-612