Towards Information Diversity Through Separable Cascade Modules for Image Super Resolution

Wei Chen†
Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, China
e-mail: chw@hust.edu.cn
† These authors contributed equally.

Zhenxing Huang†
Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, China
e-mail: huangzx@hust.edu.cn
† These authors contributed equally.

Dongdong Liu
Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, China
e-mail: ddliu@hust.edu.cn

Ping Lu
School of Computer Science & Technology
Huazhong University of Science and Technology
Wuhan, China
e-mail: luping06@hust.edu.cn

Jincai Chen*
Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, China
* Corresponding author. e-mail: jcchen@hust.edu.cn

***Abstract*—Convolution neural networks (CNNs) have been widely applied to learn the mapping function between low-resolution (LR) images and high-resolution (HR) images. With the aid of deep networks operating through cascading collaborative modules, several works have effectively improved image quality. In these works, the local cascading module is usually treated as an atomic term to extend, which is a convenient way to deepen network structures for feature extraction and information fusion. However, this limits the representation of later reconstruction layers because of its lack of abundant features if local atomic modules can output only one type of information. For instance, residual blocks only contain residual information. In this paper, we propose a separable mechanism in module design to increase the representation capability of deep neural networks. For the sake of diverse features, we obtain two-stream features with four separable modules based on residual learning and attention mechanisms. Through a contiguous memory (CM) mechanism, it ultimately combines low-level features with high-level features. Similar to the residual-in-residual (RIR) structure, we propose an attention-in-attention (AIA) framework to deepen our networks. Experimental results demonstrate the effectiveness of our method when applied to several image datasets.**

Keywords—Image super-resolution, separable cascade modules, attention-in-attention, residual network, attention mechanism

I. INTRODUCTION

As one of the most challenging problems, single-image super-resolution (SISR), is to estimate a high-resolution (HR) image from its low-resolution (LR) counterpart Since image details are in high demand in some applications, such as facial images [1], medical images [2], and video restoration [3], image super-resolution (SR) has attracted substantial attention in the computer vision research community. Convolution neural networks (CNNs) have been the basis for the recent rapid development in computer vision tasks. In particular, SR methods[4] based on CNNs have achieved notable performance improvements over several previous traditional methods, including interpolation [5], neighbor embedding methods [6], and sparse coding methods [7].

As proven by the fact that deeper networks have better performance [4], designing scalable modules to extract more useful features and better information fusion methods are two key insights to consider when attempting to build deeper networks. However, the vanishing gradient problem hinders reconstructed image quality as the convolution layers become increasingly deep. In addition, deepening neural networks with large network parameters entails the problem of increased computing costs. Since He et al. proposed ResNet [8], the residual module has been an excellent choice to cascade local residual information through the element-wise add operation, which has effectively alleviated the vanishing gradient problem. On the basis of [8], several works [4, 9] improve performance by increasing skip-connection paths among residual modules. In these works, the local modules can only transmit one kind of feature consisting of residual features and the input feature, which results in a lack of information diversity because of this homogeneous approach. Another research topic for SR methods based on deep learning is information fusion for reconstruction functions. To fuse more features from low-level layers, the complex Densenet[10] was introduced into the SR domain, and it has fewer parameters and obtains helpful contextual information in large regions. SRDenseNet [11] fully explores the advantages of skip connections by linking all layers in the networks. If we substantially deepen the convolution layers, it becomes difficult to train a wider network with dense blocks because of their abundant skip connections and huge squeeze convolution layers[12].

To solve these problems, we propose our separable cascade modules (SCMs) to seek both a diversity of information and better feature fusion to build a deeper network. In the separable module, the output streams can be divided into two parts: one main stream that contains residual information for the next cascading module and the another

2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS) | 978-1-6654-1538-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACCTCS52002.2021.00034

978-1-6654-1538-5/21/$31.00 ©2021 IEEE
DOI 10.1109/ACCTCS52002.2021.00034

extra stream that is utilized to convey shallow-level features to high-level layers. Considering the information fusion for the separable part, we adopt the contiguous memory (CM) cascade structure to combine each local module and the ultimate reconstruction layer, where the CM cascading structure enhances the computational efficiency and has fewer parameters than the dense structure [10]. Recently, the embedded residual structure, the so-called residual-in-residual (RIR), has become popular for deepening networks with global and local skip connections [12]. Inspired by RIR, we propose a attention-in-attention (AIA) framework to combine global and local parts. Furthermore, we propose our separable residual attention network (SRAN) on the basis of separable cascade modules and the AIA framework.

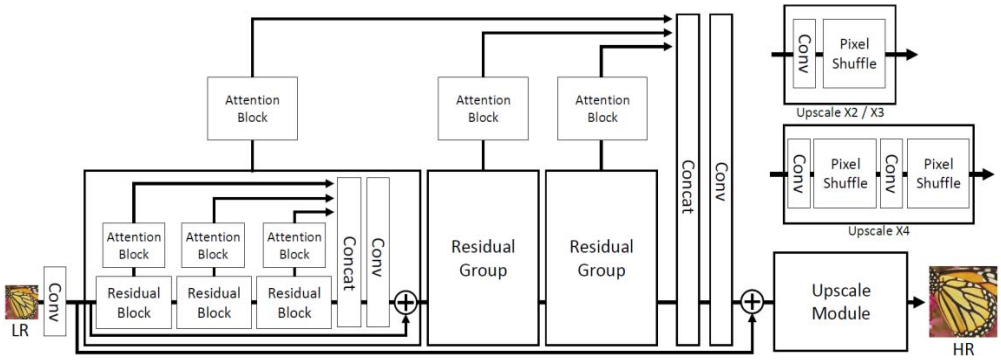


Fig. 1. Main network framework of our separable residual attention network (SRAN).

II. METHODS

A. Network Architecture

We demonstrate our proposed network framework in Fig.1. We divide our model into two parts: one that handles the feature extraction and one that reconstructs HR images. In the feature extraction part, the primary coarse feature extraction block consists of one convolution layer with a kernel size of 3×3 to attain coarse features for input images. To obtain more context information, SRAMs are cascaded with the CM cascade connection. Moreover, two-stream outputs convey residual and attention information. For outputs, the residual stream is utilized as the input data for the next cascade module, while the attention stream can transmit the local low-level features to the fusion layer. Similarly, we designed SRAGs on the basis of SRAMs. Because most streams for the fusion layer from each attention stream in SRAM and SRAG, we name this network framework the AIA framework. Through the AIA framework, useful information is directly conducted into the reconstruction layers. For the reconstruction part, the upscale module adopts a pixel shuffle to amplify the image size. Our model can complete image SR with scaling factors ×2, ×3, and ×4.

For LR input image ILR , the ultimate HR output image IH R can be formulated as follows:

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

$$I_{HR} = H_{UP}\,(H_{AI\,A}(F_0) + F_0) \tag{2}$$

Where H_0 denotes the coarse feature extracting block with one convolution layer and F_0 denotes coarse features. H_{UP} denotes the upscale operation, and H_{AIA} denotes the proposed deep attention in the attention framework.

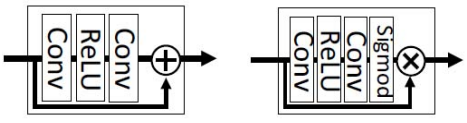
Several previous methods [13, 14] for mapping LR images and HR images adopt mean squared error (MSE) to minimize loss. MSE usually avails the peak signal-to-noise ratio (PSNR) while it would result in over-smoothed constructed image [15]. To trade off the image construction quality and PSNR, we utilize mean absolute error (MAE) to define our loss function. With network parameter Θ preparing for the end-to-end mapping function F, our loss function can be represented as:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \| F(Y_i;\Theta) - X_i \|_1 \tag{3}$$

Where $L(\cdot)$ represents the mean loss between the estimated HR result and the ground truth. $F(Y_i; \Theta)$ denotes the reconstructed HR image through our network model. X_i stands for the ground truth, and Y_i denotes the degenerated counterpart.

B. Separable Cascade Modules (SCMs)

We propose a separable module that can separate multi-type information from one module such that local information can be propagated to a high-level layer to improve image reconstruction quality. By doing so, it will augment helpful information for a local module.



(a) Basic residual block (b) Basic attention block

Fig 2. Basic residual block and attention block. We use two convolution layers and the element-wise add "⊕" operation in the residual block. For the attention block, the sigmod function is utilized to produce the attention mask, and the element-wise product "⊗"helps to produce attention output.

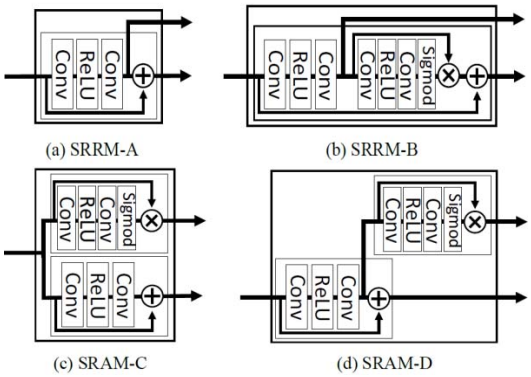


Fig 3. Four separable residual attention modules. (a) SRRM-A module: Single separable residual stream from the residual module. (b)SRRM-B module: Separable residual stream from the residual attention module. (c) SRAM-C module: Separable attention stream and the residual stream from the input, where the residual stream will be transferred to the next module. (d) SRAM-D module: Separable attention stream from residual output

Due to the successful application of the attention mechanism in certain domains, we attempt to adopt an attention mechanism to produce a separable stream for image SR. Similar to the human visual system's attention to local interesting areas, we hope to obtain attention region context information to recover HR details. In addition, skip-connection networks manifest superb performance in the SR problem, which is another stream for which we want to produce residual stream outputs. Recently, modular structures for neural networks have become popular due to their scalability and collaborative learning for feature extraction. Considering the benefits of modular coordination, we design our local SRAMs based on the basic residual block and attention block shown in Fig.2. As shown in Fig.3, four separable modules are demonstrated to gain local residual information and attention information as follows:

SRRM-A. we directly output the residual path for the basic residual block.

SRRM-B. the residual path is directly produced after letting the attention block follow the residual part.

SRAM-C. For the input features, we use two streams to extract residual and attention information through the residual and attention block.

SRAM-D. For the residual path, we employ the attention block to gain attention output. As for SRRM-A and SRRM-B, the main stream actually contains separable residual information. Instead, the SRAM-C and SRAM-D output two types of information. After ablation studies on the four separable modules, we find that SRAM-C and SRAM-D perform better in the diversity of features than redundant separable modules SRRM-A and SRRM-B.

We obtain a soft attention feature with a sigmoid activation function and element-wise product operations. Our attention block can be represented as follows:

$$F^a = M(F^i) * F^i \tag{4}$$

Where $M(\cdot)$ represents attention masks ranging from [0,1] and F^i denotes the input features.

C. Attention-in-Attention (AIA)

We now describe our proposed AIA structure, which contains G SRAGs and the long skip connection (LSC). Furthermore, each SRAG contains M SRAMs with a short connection (SSC). With this framework, we train a deep convolution network (over 200 layers) to gain better performance.

It has been demonstrated that residual blocks can be cascaded to achieve more than 1000-layer networks [8]. However, such a system will suffer from training difficulty and be unlikely to gain further performance improvement by cascading a very deep network [4]. To solve this problem, Zhang et al. [4] propose the RIR structure to achieve much deeper networks of over 400 convolution layers. Inspired by RIR, we build our deep network through the CM cascade connection to capture sufficient knowledge for recovering high-frequency details.

Through the AIA framework, the residual stream $F^r_{g,m}$ and attention stream $F^a_{g,m}$ in the m-th SRAM of the g−th SRAG can be formulated as:

$$F^r_{g,m} = K_{g,m}(F^r_{g,m-1}) = K_{g,m}(K_{g,m-1}(\ldots, K_{g,1}(F_{g-1}))) \tag{5}$$

$$F^a_{g,m} = M_{g,m}(F_{g,m-1}) * F_{g,m-1} \tag{6}$$

Where $K_{g,m}$ denotes the operation to extract a residual stream output by two convolution layers and $M_{g,m}$ denotes the attention mask produced by the attention path.

For the g-th SRAG, M SRAMs are cascaded through the CM cascade connection. The final output is combined with each module in this group. Thus, the g-th SRAG could be formulated as:

$$F^r_g = Conv(Cat(F^r_{g,M}, F^a_{g,M}, F^a_{g,M} - 1, \ldots, F_{g,1})) + F^r_{g-1} \tag{7}$$

$$F^a_g = M_g(F_{g-1}) * F_{g-1} \tag{8}$$

Where "Conv" denotes the fusion convolution layer and "Cat" is the concatenation function to combine high-level and low-level features from each SRAM. To preserve the majority of information in this group, the short path is used. For each SRAG, we also apply the attention block to manage separable group attention information.

By cascading each SRAG, we ultimately build the global AIA framework. On the basis of the constructed SRAG, the output of the AIA structure is formulated as:

$$F_{AIA} = Conv(Cat(F^r_G, F^a_G, F^a_{G-1}, \ldots, F_1)) = H_{AIA}(F_0) \tag{9}$$

Where "Conv" denotes the fusion convolution layer and "Cat" is the concatenation function to combine high-level and low-level features for each SRAG.

In the total framework, the attention information contains the majority of paths for the reconstruction part. In addition, the residual stream, which is frequently used to cascade through the residual connection to a local short path, could preserve more context texture information for rehabilitating image details. We find that attention maps clearly depict that the principal part of the input image and the residual stream output gains as much texture as possible. Regarding the pixel value, the attention map usually presents a wider distribution than the residual output.

D. Reconstruction Module

The most popular ways to enlarge feature maps for image SR are the transposed convolution and sub-pixel magnification methods. We choose the sub-pixel magnification because its computational cost is lower than the former method. By enlarging, we can employ LR images as inputs instead of magnified images. This magnification process can be formulated as follows:

$$\begin{aligned} \mathrm{Dim}(I) &= I_H \times I_W \times I_{C0} \\ &= I_H \times I_W \times (s \times s \times I_{C1}) \\ &= (s \times I_H) \times (s \times I_W) \times I_{C1} \end{aligned} \quad (10)$$

Where Dim(·) denotes the dimension of a tensor and s denotes the scaling factor. For the input tensor I with shape $I_H \times I_W \times I_{C0}$, the channel I_{C0} can be represented as $s \times s \times I_{C1}$. Then, we reshape the input tensor as $(s \times I_H) \times (s \times I_W) \times I_{C1}$. When enlarging the input tensor by the scaling factors ×2 and ×3, we directly set s at 2 and 3, respectively. Specifically, we require the ×4 tensor to cascade twice as much as the ×2 upscale module and the latter thrice for magnifying ×8.

III. Experiments

In this section, we evaluate the performance of our model with several benchmark test results. First, we describe public training datasets and several test datasets. Then, we show network parameters for the training model. Third, we provide an ablation analysis on different SCMs and validate the effect of this method. Next, we compare our SRAN models with several state-of-the-art methods. Finally, we present local texture details of the example output results.

A. Experimental Datasets

The new public DIV2K dataset [16] is adopted as our training dataset, and it contains has 800 HR images for training and 100 validation images as well as 100 test images. In terms of testing, we test our model in standard benchmark datasets, including Set5 [6], Set14 [17], BSD100 [18] and Urban100 [19]. Before training, we use the bicubic interpolation method to separately downsample training images as input images with scaling factors ×2, ×3, and ×4.

B. Training Implementation

We adopt data augmentation metrics including randomly rotating 90◦, 180◦, 270◦ and flipping both in the horizontal and vertical directions. For the input data for our model, we randomly cut LR color images into several patches with image shapes 48 × 48, and the batch size is set at 16. To hold size fixed, we use a 3 × 3 filter with a zero-padding strategy to exchange contextual information. In the AIA structure, we set our AG number at G = 12 and the SRAM number at 12 so that our model exceeds 200 convolution layers. In addition, we utilize channel-downsampling to reduce parameter convolution layers in attention parts where the reduction ratio is set at 16, similar to RCAN. The channel of the first layer and the last layer is set at 3, and the other channels are 64. The details of the network parameters are shown in TABLE I. We train our model with the ADAM optimizer where $\beta1 = 0.9$, $\beta2 = 0.999$. We finally actualize our model in the PyTorch framework on Ubuntu 16.04 with a Titan XP GPU.

TABLE I. DETAILS OF NETWORK PARAMETERS

Componets	Layer	Filter Size	Input channel	Output channel
Feature Coarse	1	3 × 3	64	64
Residual Block	1	3 × 3	64	64
ReLu	-	-	64	64
Residual Block	2	3 × 3	64	64
Attention Block	1	3 × 3	64	64
ReLu	-	-	64	64
Attention Block	2	3 × 3	64	64
Sigmoid	-	-	64	64
Concatenation	-	-	64 × (M+1)	64 × (M+1)
Fusion block in SRAG	1	3 × 3	64 × (M+1)	64
Concatenation	-	-	64 × (G+1)	64 × (G+1)
Fusion block in SRAG	1	3 × 3	64 × (G+1)	64

IV. Results and Disscussions

We analyze the effect of separable residual attention and the AIA framework in this subsection. We cascade residual blocks as a baseline model without separable streams and an SSC. To control the parameters in the ablation studies, we set G = 12 and M = 12, and we test the results of the models in the 200-th epoch. In terms of SRRM-A, we study the effect by employing an SRAG and a separable residual group (SRRG). The ablation results on the Set5 and DIV2K validation datasets are shown in TABLE.II for the scaling factor ×2.

To validate the effect of SRAG, we compare results on SRRM-A+SRRG and SRRM-A+SRAG. As shown in TABLE. II, both PSNR and the structural similarity index (SSIM) perform better when SRAGs are engaged. Because of the attention mechanism, the attention output can filter the results of residual features, which extracts more helpful information for the global reconstruction layer. We find that the other models, except SRAN-C and SRAN-D, underperform the baseline model. In SRRM and SRRG, the mainstream can be divided into the summation of the extra residual stream and the input signal. Thus, separable residual information is redundant for origin outputs, which is deleterious to image quality in these approaches.

Our SRAM-C and SRAM-D both have a residual stream and an attention stream to extract dual-stream information in a local module. Compared with the SRRM-A+SRAG and SRRm-B+SRAG, SRAN-C and SRAN-D have better performance. For SRAN-C and SRAN-D, the CM cascading structure is adopted for separable attention streams in local and global networks. The comparative results from the SRAN-C, SRAN-D, and baseline prove the benefit of this design.

TABLE II. ABLATION RESULTS (PSNR(DB)/SSIM) WITH SCALE FACTORS ×2. WE HIGHLIGHT THE BEST PERFORMING RESULT IN BOLD.

Models	Baseline	SRRM-A	SRRM-A+SRRG	SRRN-A+SRAG	SRRN-B+SRAG	SRAN-C	SRAN-D
SRRM-A	×	√	√	√	×	×	×
SRRM-B	×	×	×	×	√	×	×
SRRM-C	×	×	×	×	×	√	×
SRRM-D	×	×	×	×	×	×	√
SRRG	×	×	√	×	×	×	×
SRAG	×	×	×	√	√	√	√
PSNR on DIV2K	36.26	36.22	36.22	36.24	26.21	36.29	36.28
PSNR on DIV2K	0.9465	0.9464	0.9463	0.9464	0.9463	0.9468	0.9468
PSNR on Set5	38.04	38.04	38.02	38.05	38.03	38.08	38.08
SSIM on Set5	0.9606	0.9607	0.9605	0.9606	0.9606	0.9608	0.9608

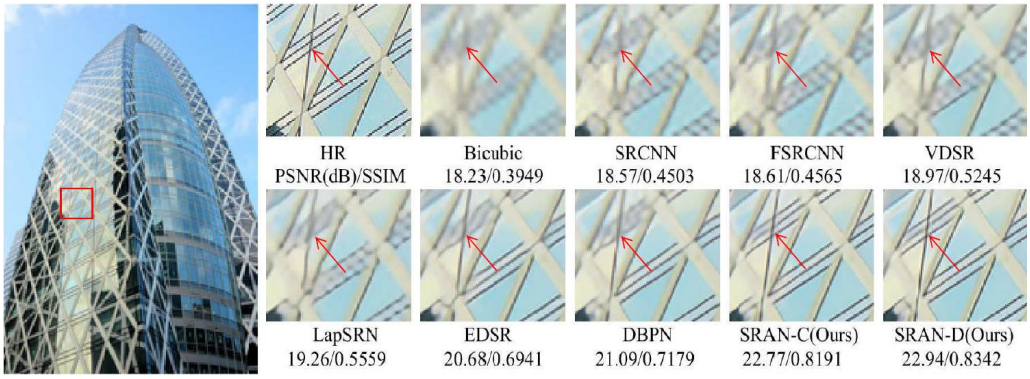


Fig. 4. Super-resolution results of "img039" from the Urban100 dataset with scale factor ×4.

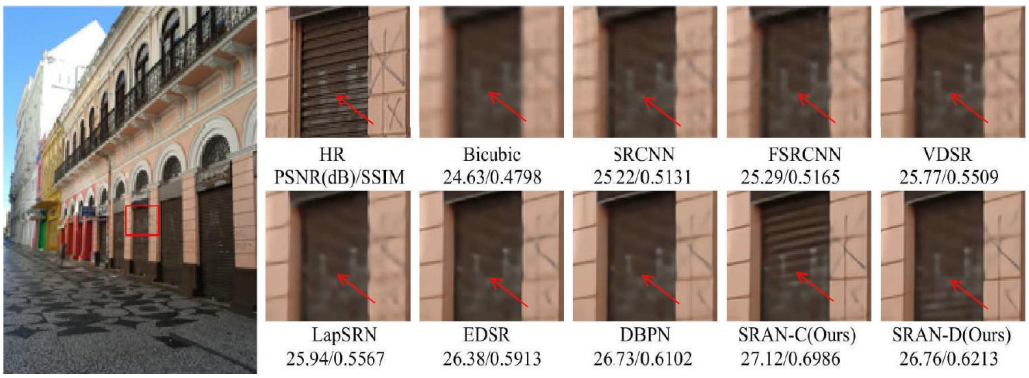


Fig. 5. Super-resolution results of "img089" from the Urban100 dataset with scale factor ×4.

As common image quality metrics, PSNR and SSIM are adopted to evaluate our estimated HR images with homologous images in the Y channel of the YCbCr color space. For a fair comparison, we apply our method and some state-of-the-art methods, including Bicubic, SRCNN [20], FSRCNN [21], VDSR [13], LapSRN [22], EDSR [15], and D-BPN [23] in the same test datasets mentioned in the previous paragraph. TABLE. III demonstrates SR image quality in Set5, Set14, BSD100, and Urban100 with scaling factors ×2, ×III, and ×4. We find that our models perform better than the other methods in most test datasets. In particular, we gain 0.51 dB in PSNR improvement over the second-best alternative with a scale ×2 in the Urban100 dataset. Because the input features of the attention block come from the preview cascade module in SRAN-C, the original input stream usually contains more information than the residual stream in SRAN-D. Thus, the results from SRAN-C are better than those from SRAN-D in most cases.

Several SR results are shown in Fig.4 and Fig.5. Compared with other methods, our output image can restore more image details. For example, textures of shutter doors reconstructed by SRAN-C and SRAN-D in image "img089" (in the Urban100 dataset with scale factor ×4) are clearer than others. Another example shows that local textures constructed by other methods are blurry, while the results of our method are clear in image "img039".

V. CONCLUSION

In this paper, we propose a separable mechanism and design our separable residual attention network (SRAN) for image super-resolution. Through the separable cascade modules (SCMs), more information, not tedious information, can be extracted. Based on the residual learning and attention mechanism, we designed our separable residual attention modules to separate residual information and attention regions from a single cascade module. In this way, it considerably enriches the output features for reconstruction

layers. Similar to a residual-in-residual structure, we propose the attention-in-attention (AIA) framework to deepen the networks to improve the representation of our model. In the AIA structure, we utilize a contiguous memory cascade structure to combine more latent context information from low-level layers with high-level layers. Compared with several state-of-the-art algorithms, our network shows superior performance in both PSNR and SSIM. In addition, our super-resolution results appear to be clearer. In the future, we will seek more useful separable information and integrate it into a unitary framework.

TABLE III. PUBLIC DATASETS (SET5, SET14, BSD100, URBAN100) ON TEST RESULTS (PSNR(DB)/SSIM) WITH SCALE FACTORS ×2, ×3, AND ×4. WE HIGHLIGHT THE BEST PERFORMING RESULT IN BOLD.

Algorithm	Scale	Set5 PSNR(dB)/SSIM	Set14 PSNR(dB)/SSIM	BSD100 PSNR(dB)/SSIM	Urban100 PSNR(dB)/SSIM
Bicubic	×2	33.66 / 0.9299	30.24 / 0.8688	29.56 / 0.8431	26.88 / 0.8403
SRCNN [20]	×2	36.66 / 0.9542	32.45 / 0.9067	31.36 / 0.8879	29.50 / 0.8946
FSRCNN [21]	×2	37.05 / 0.9560	32.66 / 0.9090	31.53 / 0.8920	29.88 / 0.9020
VDSR [13]	×2	37.53 / 0.9590	33.05 / 0.9130	31.90 / 0.8960	30.77 / 0.9140
LapSRN [22]	×2	37.52 / 0.9591	33.08 /0.9130	31.80 / 0.8950	30.41 / 0.9101
EDSR [15]	×2	38.11 / 0.9602	33.92 / 0.9195	31.32 / 0.9013	32.93 / 0.9351
DBPN [23]	×2	38.09 / 0.9600	33.85 / 0.9190	32.27 / 0.9000	32.55 / 0.9324
SRAN-C(Ours)	×2	**38.22 / 0.9614**	**34.03 / 0.9210**	**32.37 / 0.9021**	**33.06 / 0.9369**
SRAN-D(Ours)	×2	**38.21 / 0.9612**	**33.97 / 0.9208**	**32.36 / 0.9020**	**32.92 / 0.9357**
Bicubic	×3	30.39 / 0.88682	27.55 / 0.7742	27.21 / 0.7385	24.46 / 0.7349
SRCNN [20]	×3	32.75 / 0.9090	29.30 / 0.8215	28.41 / 0.7863	26.24 / 0.7989
FSRCNN[21]	×3	33.18 / 0.9140	29.37 / 0.8240	28.53 / 0.7910	26.43 / 0.8080
VDSR [13]	×3	33.67 / 0.9210	29.78 / 0.8320	28.83 / 0.7990	27.14 / 0.8290
LapSRN [22]	×3	33.82 / 0.9227	29.87 / 0.8320	28.82 / 0.7980	27.07 / 0.8280
EDSR [15]	×3	34.65 / 0.9280	30.52 / 0.8462	29.25 / 0.8093	28.80 / 0.8653
DBPN [23]	×3	- / -	- / -	- / -	- / -
SRAN-C(Ours)	×3	**34.72 / 0.9297**	**30.61 / 0.8478**	**29.29 / 0.8102**	**28.91 / 0.8677**
SRAN-D(Ours)	×3	**34.67 / 0.9293**	**30.51 / 0.8462**	**29.25 / 0.8095**	**28.77 / 0.8654**
Bicubic	×4	28.42 / 0.8404	26.00 / 0.7027	25.96 / 0.6675	23.14 / 0.6577
SRCNN [20]	×4	30.48 / 0.8628	27.50 / 0.7513	26.90 / 0.7101	24.52 / 0.7221
FSRCNN[21]	×4	30.72 / 0.8660	27.61 / 0.7550	26.98 / 0.7150	24.62 / 0.7280
VDSR [13]	×4	31.35 / 0.8830	28.02 / 0.7680	27.29 / 0.7250	25.18 / 0.7540
LapSRN [22]	×4	31.54 / 0.8850	28.19 / 0.7720	27.32 / 0.7270	25.21 / 0.7560
EDSR [15]	×4	31.46 / 0.8968	28.80 / 0.7876	27.71 / 0.7420	26.64 / 0.8033
DBPN [23]	×4	32.47 / 0.8980	28.82 / 0.7860	27.72 / 0.7400	26.38 / 0.7946
SRAN-C(Ours)	×4	**32.49 / 0.8985**	**28.78 / 0.7871**	**27.73 / 0.7418**	**26.69 / 0.8048**
SRAN-D(Ours)	×4	**32.50 / 0.8984**	**28.78 / 0.7869**	**27.71 / 0.7413**	**26.69 / 0.8019**

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (61672246, 61272068). We gratefully acknowledge the support from NVIDIA Corporation for providing us the Titan XP GPU used in this research.

REFERENCES

[1] A. Agrawal and N. Mittal, "Using cnn for facial expression recognition:a study of the effects of kernel size and number of filters on accuracy," The Visual Computer, vol. 36, pp. 405–412, 2020.

[2] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis," Computerized Medical Imaging and Graphics, vol. 71, pp. 30 – 39, 2019.

[3] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang, "Convolutional neural network-based block up-sampling for intra frame coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 9, pp. 2316–2330, 2018.

[4] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image superresolution using very deep residual channel attention networks," in ECCV, 2018, pp. 294–310.

[5] E. Meijering, "A chronology of interpolation: from ancient astronomy to modern signal and image processing," Proceedings of the IEEE, vol. 90, no. 3, pp. 319–342, 2002.

[6] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Lowcomplexity single-image super-resolution based on nonnegative neighbor embedding," BMVC, 2012.

[7] W. Yang, Y. Tian, F. Zhou, Q. Liao, H. Chen, and C. Zheng, "Consistent coding scheme for single-image super-resolution via independent dictionaries," IEEE Transactions on Multimedia, vol. 18, no. 3, pp. 313–325, 2016.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016, pp. 770–778.

[9] Y. Wang, L. Wang, H. Wang, and P. Li, "Resolution-aware network for image super-resolution," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 5, pp. 1259–1269, 2018.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." in CVPR, vol. 1, no. 2, 2017, p. 3.

[11] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in ICCV, 2017, pp. 4809–4817.

[12] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in CVPR, 2018.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in CVPR, 2016, pp. 1646–1654.

[14] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, ´ D. Rueckert, and Z. Wang, "Real-time single image and video superresolution using an efficient sub-pixel convolutional neural network," in CVPR, 2016, pp. 1874–1883.

[15] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in The IEEE conference on computer vision and pattern recognition (CVPR) workshops, vol. 1, no. 2, 2017, p. 4.

[16] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee et al., "Ntire 2017 challenge on single image super-resolution: Methods and results," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1110–1121.

[17] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in International conference on curves and surfaces. Springer, 2010, pp. 711–730.

[18] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 5, pp. 898–916, 2011.

[19] J. B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in CVPR, 2015, pp. 5197–5206.

[20] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in ECCV, 2014, pp. 184–199.

[21] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in ECCV, 2016, pp. 391–407.

[22] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate superresolution,” in CVPR , vol. 2, no. 3, 2017, p. 5.

[23] M. Haris, G. Shakhnarovich, and N. Ukita, “Deep back-projection networks for super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1664–1673.