Kernel Attention Network for Single Image Super-Resolution

DONGYANG ZHANG, JIE SHAO, and HENG TAO SHEN, Center for Future Media, School of Computer Science and Engineering, University of Electronic Science and Technology of China, China and Sichuan Artificial Intelligence Research Institute, China

Recently, attention mechanisms have shown a developing tendency toward convolutional neural network (CNN), and some representative attention mechanisms, i.e., channel attention (CA) and spatial attention (SA) have been fully applied to single image super-resolution (SISR) tasks. However, the existing architectures directly apply these attention mechanisms to SISR without much consideration of the nature characteristic, resulting in less strong representational power. In this article, we propose a novel kernel attention module (KAM) for SISR, which enables the network to adjust its receptive field size corresponding to various scales of input by dynamically selecting the appropriate kernel. Based on this, we stack multiple kernel attention modules with group and residual connection to constitute a novel architecture for SISR, which enables our network to learn more distinguishing representations through filtering the information under different receptive fields. Thus, our network is more sensitive to multi-scale features, which enables our single network to deal with multi-scale SR task by predefining the upscaling modules. Besides, other attention mechanisms in super-resolution are also investigated and illustrated in detail in this article. Thanks to the kernel attention mechanism, the extensive benchmark evaluation shows that our method outperforms the other state-of-the-art methods.

CCS Concepts: • **Computing methodologies → Reconstruction**;

Additional Key Words and Phrases: Image super-resolution, kernel attention, receptive field, multi-scale features

ACM Reference format:
Dongyang Zhang, Jie Shao, and Heng Tao Shen. 2020. Kernel Attention Network for Single Image Super-Resolution. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 3, Article 90 (July 2020), 15 pages.
https://doi.org/10.1145/3398685

1 INTRODUCTION

Single image super-resolution (SISR), which aims to reconstruct high-resolution (HR) images from its low-resolution (LR) counterparts, is a fundamental low-level computer vision task. Since the

This work was supported by the National Natural Science Foundation of China (No. 61832001, No. 61672133, and No. 61632007), and Sichuan Science and Technology Program (No. 2019YFG0535).
Authors' addresses: D. Zhang, J. Shao (corresponding author), and H. T. Shen, Center for Future Media, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, 611731, Sichuan Artificial Intelligence Research Institute, Yibin, China, 644000; emails: dyzhang@std.uestc.edu.cn, {shaojie, shenhengtao}@uestc.edu.cn.

1551-6857/2020/07-ART90 $15.00
https://doi.org/10.1145/3398685

nature of SISR is to improve image perceptual quality by restoring the missing high-frequency part in LR images, it has a great demand in the applications where the acquisition of high-quality images is arduous and expensive. SR also does a favor for other real-world applications, such as video surveillance [12].

In literature, a few years ago, SISR was dominated by a diversity of classical methods based on feature engineering, such as sparse representation methods [54], prediction-based methods [24], and edge-based methods [10]. Recently, with the flourish development of deep learning, some advanced techniques in convolutional neural networks (CNN) have been successfully extended to tackle SR tasks. For example, ResNets [16] and DenseNets [20] architectures have shown outstanding performance in image classification, and the residual learning strategy is also applied in the SR network [2, 8, 30, 32, 36], which shows a remarkable effect on accuracy improvement and model training. Besides, inspired by recent discoveries that deep feed-forward structures can be represented as recurrent neural networks (RNNs) with finite unfoldings [35], naturally, recursive learning strategy is introduced into SR where the same convolutional module is repeatedly applied for feature extraction; thus, the number of parameters do not increase along with the recursion depth, such as deeply-recursive convolutional network (DRCN) [26], deep recursive residual network (DRRN) [46], and dual-state recurrent network (DSRN) [13]. Moreover, other techniques and their combinations are also fully explored in SR, such as dilated convolution [58], channel attention [60], back-projection [37], and group convolution [2].

Although fascinating results have been obtained by the above methods, some drawbacks also should be noted. As to the existing CNN-based methods that treat all types of information equally, accordingly, it is barely for these models to effectively distinguish the informative content, such as low and high frequency information. Recently, the attention mechanism has become increasingly prevalent among the design of deep neural networks, which allows the model to recalibrate the feature maps based on their importance. As shown in Figure 1, Zhang et al. [60] proposed residual channel attention block (RACB) to form residual channel attention network (RCAN) by simply introducing the channel attention (CA) mechanism for SR. In addition, CA and spatial attention (SA) are jointly integrated in multi-path adaptive modulation network (MAMNet) [27] and channel-wise and spatial attention residual CSAR [19] to exploit the relationship of both inter- and intra-channels. However, these applied mechanisms to SR are directly borrowed from the high-level computer vision tasks without much modification. The substantive characteristics between SR and other tasks need further consideration to selectively extract the representative feature, such as multi-scale information. Moreover, it is universally acknowledged that most SR models are customized where a trained model is only for a specific scale factor. In other words, the existing SR models with different scale factors are regarded as independent tasks; thus, we have to train and store multiple models for different scale factors (e.g., 2×, 3×, and 4×), resulting in inefficient computing. Therefore, it is imperative to develop a novel architecture for multiple scale factors.

To overcome these drawbacks, we first introduce the kernel attention module (KAM) for SISR. The existing attention modules applied in SR all focus on CA and SA. Different from them, by observing that image features on different scales and receptive fields are critical for high-quality image restoration [31], we develop a dynamic selection mechanism, which allows the multiple kernels in KAM to automatically adapt to input information with different scales. The structure of KAM is shown in Figure 1, and we can find that KAM contains multiple branches with different kernel sizes to aggregate multi-scale information. Inspired by Ref. [60], we arrange a number of KAMs with residual groups (RG), and a series of RG further constitute the residual architecture with long skip connection, rendering our model a residual kernel attention network (RKAN). Besides, it is a common practice for most SR models [9, 30, 60, 62] to adopt the post-upsampling strategy, which first extracts features from the original LR inputs, and then upscales spatial resolu-

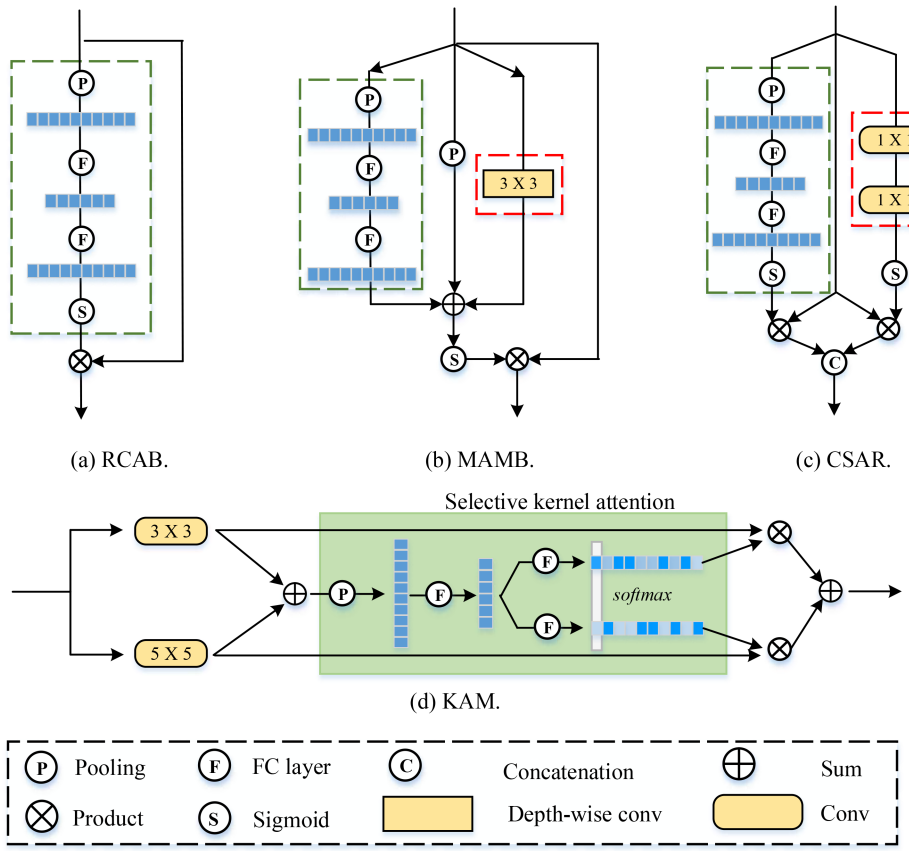


Fig. 1. Illustration of our proposed kernel attention module (KAM) and other attention mechanisms. The green-dotted box denotes the channel attention (CA) and the red-dotted box denotes spatial attention (SA).

tion at the tail of network with sub-pixel convolution [43] or transposed convolution [9]. Actually, different scale factors depend on the particular configuration of the upsampling layers applied at the tail of the network. However, regarding to the feature extraction process in our model, due to the dynamic kernel selection mechanism, there is no substantial difference among SR models with specific scale factor. Thus, we design a novel model that treats the feature extraction process as the communal function, and an upsampling layer corresponding to specific scale factor is plugged with extracted feature to obtain the HR images. From the above point of view, just one model is capable of dealing with multi-scale SR as long as the specific scale factor is given during testing, leading to efficient calculation.

In general, the contributions of the article can be summarized in the following points:

—We propose the kernel attention mechanism for SR in consideration of the nature of the task. Based on the KAM, a novel architecture named RKAN is also proposed.
—The feature extraction process for different scale factors is integrated together as one function module. By specifying the various scale factors, a single model can deal with the multi-scale SR task.

—Extensive experiments have been conducted to verify the proposed method, and the results demonstrate that our method is more effective and efficient than existing state-of-the-art SR methods.

2 RELATED WORK

Lots of literatures have contributed to address SISR, in which Wang et al. [51] publish a very comprehensive survey on deep learning based SR, including benchmark datasets, network design, and assessment methods. In this section, we give a brief description on attention mechanism, as well as supervised and unsupervised methods for SR.

2.1 Attention Mechanism

Recently, the attention mechanism, reallocating the available resources based on the importance to suppress the less informative feature expressions [41] has been successfully applied in CNN and facilitated the development across a range of tasks, such as visual captioning and visual question answering [3, 5, 11, 15, 45, 48], as well as cross-modal retrieval [49, 53]. Hu et al. [18] proposed squeeze-and-excitation (SE) block to exploit the relationship between channel-wise, which benefits the image classification a lot. Apart from channel attention, BAM [42] and CBAM [52] sequentially applied channel and spatial attentions in the same manner. Similar to our work, SKNets, proposed by Li et al. [33] first introduced the attention mechanism into neurons to make the receptive fields of neurons more adaptive to input information. However, the above studies focus on high-level vision task (image classification), and only a few works [19, 27, 62] pay attention to applying the attention mechanism for SR task compared with the conventional CNN-based methods. For better understanding, the existing attention mechanisms for SR are shown in Figure 1, and we can find that there exists an obvious difference between the proposed kernel attention and other attentions. Specifically, the exiting attentions mainly focus on the channel attention or spatial attention, while our method concentrates on selective kernel attention.

2.2 Supervised Super-resolution

Until now, most existing methods address the SR issue in a supervised manner, i.e., training with both LR images and ground-truth HR counterparts. super-resolution convolutional neural network (SRCNN) [7] is a seminal work in the history of SR, which proposed a simple but efficient CNN-based architecture with three layers for the first time. Later works [14, 38] gradually increase the network depth and width to improve the capacity of the model, which is the key factor for the progress in SR.

Recently, with residual and dense connection, Zhang et al. [62] proposed residual dense network (RDN) by making full use of the hierarchical features from all the convolutional layers. The generative adversarial networks (GANs) are also introduced in SR, such as Refs [30] and [57], which not only show remarkable ability in restoring photo-realistic natural images, but also bring significant gains in perceptual quality than any other state-of-the-art methods. Inspired by the error feedback mechanism, dense deep back-projection network (D-DBPN) [14] is proposed to address the mutual dependencies of HR and LR images. However, due to the dense connection and HR feature maps during running, it suffers from vast computational overhead. Super-resolution feedback network (SRFBN) [34] introduced a novel feedback block to generate powerful high-level representations. Besides, recursive learning is also widely studied in SR. Kim et al. [26] proposed the 16-recursions DRCN employing a single convolutional layer as the recursive unit. Yang et al. [55] proposed a lightweight network named deep recurrent fusion network (DRFN), which adopts the deep recurrence learning strategy to enlarge the receptive field. Some other variants performing recursion on residual block are also proposed, such as DRRN [46], MemNet [47], and CARN [2].

Moreover, tentative works based on attention mechanism have also been investigated recently. Borrowing the idea from the squeeze-and-excitation (SE) block [18], Zhang et al. [60] proposed the RCAN for SR, which adaptively recalibrates each channel-wise feature by exploiting inter-channel relations. Beyond the CA, spatial attention (SA) mechanism is also studied by CSAR [19] and MAMNet [27], both of which achieve improved SR performance as more informative features can be learned. Instead of conventional CA, information multi-distillation network (IMDN) [22] employs the contrast-aware channel attention layer, which is good at capturing the structure information. In addition, Zhang et al. [61] proposed a residual non-local attention network (RNAN), which shows superior results over leading methods for not only super-resolution, but also other classical image restoration tasks such as denoising and demosaicing. Recently, Dai et al. [6] developed a second-order attention network (SAN) by exploring second-order statistics of features for SR.

2.3 Unsupervised Super-Resolution

Generally speaking, the LR images for supervised methods are often obtained by performing predefined degradation on existing HR images, which shows poor generalization performance in real-world scenarios. Since it is difficult to collect paired images in the real world, the SR methods in an unsupervised manner have attracted more attention recently.

Shocher et al. [44] proposed a CNN-based method named zero-shot super-resolution (ZSSR) in an unsupervised manner, which exploits the internal recurrence of information inside a single image by extracting image pairs from the input image itself. However, ZSSR is an image-specific method where each image corresponds to a single network during test time. Bulat et al. [4] proposed a two-stage process to bypass the paired data requirement, in which an HR-to-LR GAN is trained to learn degradation first using unpaired LR-HR images; then, an LR-to-HR GAN is trained for image SR using paired LR-HR images conducted based on the first GAN. The structure of cycle-in-cycle from CycleGAN [64] shows strong power on mapping two different domains in an unsupervised manner. Motivated by CycleGAN, Yuan et al. [56] proposed CinCGAN with two CycleGANs, consisting of four generators and two discriminators to build two mapping functions (noisy LR $\Longleftrightarrow$ clean LR and clean LR $\Longleftrightarrow$ clean HR). CinCGAN achieves comparable performance to other supervised methods, even under very harsh conditions.

3 RESIDUAL KERNEL ATTENTION NETWORK

3.1 Kernel Attention Module

Figure 1 shows the existing attention mechanisms applied in SR, including RCAB [60], CSAR [19], and multi-path adaptive modulation block (MAMB) [27]. Although these methods are different in architecture, they explore the channel attention (CA) in the same way, which can be summarized as three steps: *squeeze-excitation-scaling*. The *squeeze* step mainly generates a brief descriptor for a feature by using pooling methods, which can be formulated as:

$$S_{avg} = pooling_{avg}(X), \tag{1}$$

where $X \in \mathbb{R}^{C\times H\times W}$, $S_{avg} \in \mathbb{R}^{C\times 1\times 1}$, C, H, and W denote the channel number, height, and width of X. The *excitation* step mainly learns the nonlinear interactions among the descriptors extracted from the *squeeze* step to capture the importance between channels, which can be formulated as:

$$M = \sigma(FC_2(FC_1(S_{avg}))), \tag{2}$$

where FC and σ denote the fully connected layer and sigmoid activation, respectively. The two fully connected layers play the role as a bottleneck structure with a reduction ratio of r. As the

attention map M is generated, finally, the *scaling* step is to recalibrate the input feature maps by multiplying M and X through channel-wise, which can be formulated as:

$$\tilde{X} = M \otimes X. \tag{3}$$

As to the SA, CSAR [19] adopts two convolutional layers with a 1×1 kernel. However, MAMB [27] uses depth-wise convolution [17] with a 3×3 kernel.

The *squeeze-excitation-scaling* manner is also incorporated in our KAM, but behaves in a totally different way. Inspired by Ref. [31], showing that multi-scale information is crucial to reconstruct high-quality images, we not only design multiple branches with different convolution kernels to extract features under different receptive fields, but also provide a soft attention across channels to adaptively adjust their receptive field sizes. Looking carefully at Figure 1(d), a two-branch case is shown. Given a feature map X, we first conduct two convolution operations with 3×3 and 5×5 kernels:

$$\widetilde{M} = Conv_{3\times3}(X), \quad \widehat{M} = Conv_{5\times5}(X), \tag{4}$$

where zero padding is used to ensure feature map size of each layer is unchanged. To integrate the information from different branches, a summation operation is used:

$$M = \widetilde{M} + \widehat{M}. \tag{5}$$

Next is the usual *squeeze-excitation-scaling* operation. Note that after the first FC layer performs channel-downscaling with reduction ratio r, the low-dimension feature is then restored to the original scale by multiple FC layers. Here, two compact feature descriptors, $A \in \mathbb{R}^{C\times1}$ and $B \in \mathbb{R}^{C\times1}$, are obtained, which correspond to different branches. To provide the selection mechanism, a softmax operator is applied where each channel-wise signal is normalized to a range of 0 and 1, which can be formulated as:

$$a_c = \frac{e^{A_c}}{e^{A_c} + e^{B_c}}, \quad b_c = \frac{e^{B_c}}{e^{A_c} + e^{B_c}}, \tag{6}$$

where A_c is the c-th element of A, likewise B_c. a_c and b_c are the attention weights on various kernels used to recalibrate the feature from the two branches:

$$O_c = a_c \otimes \widetilde{M}_c + b_c \otimes \widehat{M}_c, \quad a_c + b_c = 1, \tag{7}$$

where O_c is the c-th element of O, likewise $\widetilde{M}_c$ and $\widehat{M}_c$, $O_c \in \mathbb{R}^{H\times W}$ and $O \in \mathbb{R}^{C\times H\times W}$.

3.2 Network Architecture

An overall architecture of our residual kernel attention network (RKAN) is shown in Figure 2. It clearly shows that the architecture is composed of two parts, feature extraction part and upscaling part. Firstly, a convolution layer is applied to the input image I^{LR}. Note that, we employ big kernel size 7×7 to extract initial feature maps. Inspired by Ref. [60], residual in residual (RIR) structure shows fascinating performance on high-frequency information learning and suppresses the redundant low-frequency information through multiple skip connections. Therefore, we arrange our network with a residual group (RG), which also is illustrated in Figure 2. Two convolution layers are before and after a KAM, and these three parts form the basic attention unit (BAU). A series of BAUs with short skip connection constitute an RG. Together with long skip connection, the feature extraction part can be formulated as:

$$F_n = F_0 + G_n(G_{n-1}(\cdots G_1(F_0)\cdots)), \tag{8}$$

where F_0 and F_n are the initial and final feature maps, respectively, G denotes the RG, and the subscript n denotes the number of RG. In this paper, three BAUs are employed in RG and three RGs make up the feature extraction part.

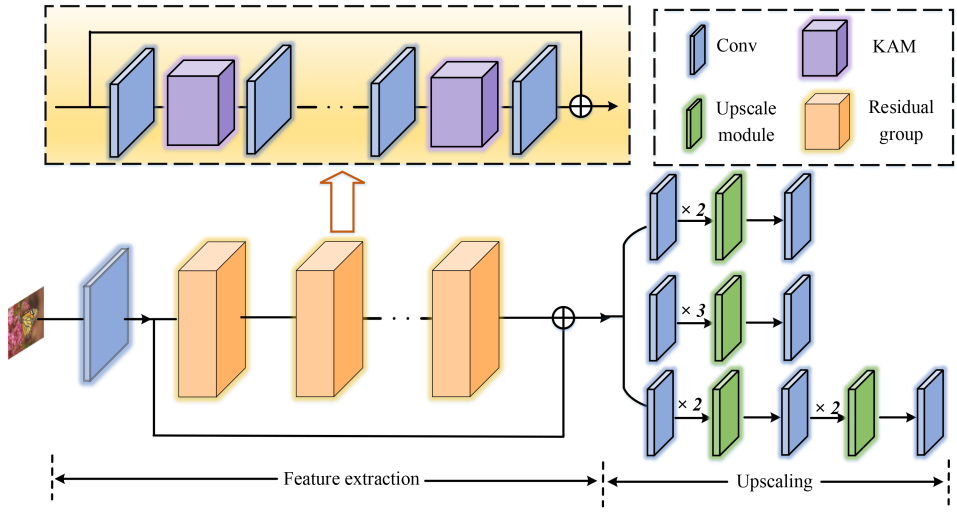


Fig. 2. Illustration of the proposed RKAN. The network is composed of feature extraction module and upscaling module. The upscaling module for different scale factors is predefined, so our single network can deal with multi-scale SR task by specifying the upscaling factor.

As to the upscaling part, in contrast to most existing methods, three modules with different scale factors (×2, ×3, and ×4) are predefined. In this sense, we have to give the specific scale factor one forward pass. Therefore, the process of SR can be formulated as:

$$I^{SR} = F_{SR}(I^{LR}, r), \tag{9}$$

where r denotes the scale factor. By doing so, we can train one model for different scale factors, leading to high efficiency. In contrast, most existing methods are only capable of one specific scale factor.

3.3 Loss Functions

In this work, our goal is to learn the mapping function F_{SR} via neural network to generate an HR image I^{SR} that is close to the ground-truth image I^{SR}. Since mean squared error (MSE), also called $L2$ loss, is related to PSNR, which is an important evaluation criterion for SR, it is wildly used to supervise the training phase for general image restoration. However, through comprehensive experiments on an image restoration task [63] shows that mean absolute error (MAE), which is also called $L1$ loss, is full of more superiority than $L2$ loss for super-resolution. First, compared with $L1$ loss, $L2$ loss has a strong penalty for large errors and a low penalty for small errors, ignoring the impact of image content itself and leading to inferior visual quality. Second, the convergence performance of $L2$ loss is worse than $L1$ loss. So far, almost all existing SR and other image restoration methods [2, 36, 60, 62] employ $L1$ loss as the loss function. MAE is formulated as follows:

$$\mathcal{L}_{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left\|F_{SR}\left(I_i^{LR}, r\right) - I_i^{HR}\right\|_1. \tag{10}$$

Consequently, we employ $L1$ loss to supervise our RKAN during training.

4 EXPERIMENTS

In this section, we demonstrate the details about experiments. Following the common practice, we mainly perform SR to restore high-quality images on 2×, 3×, and 4×. Five commonly used

Table 1. Performance Comparison of Different Settings in Kernel Size and SKA

Model	K3	K5	K7	SKA	#Params (K)	PSNR
1	✓				1,340	32.05
2		✓			1,930	32.09
3			✓		2,815	32.13
4	✓	✓			2,267	32.17
5	✓	✓		✓	2,277	32.22
6	✓		✓		3,152	32.19
7	✓		✓	✓	3,162	32.25
8		✓	✓		3,752	32.20
9		✓	✓	✓	3,742	32.27
10	✓	✓	✓		4,074	32.21
11	✓	✓	✓	✓	4,089	32.32

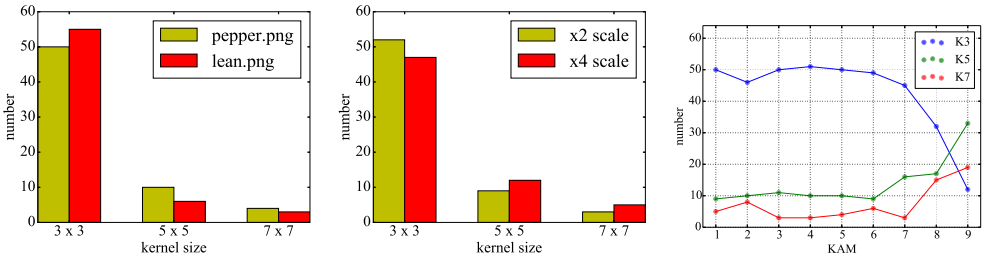
benchmarks, including Set5, Set14, BSDS100 [39], URBAN100 [21], and MANGA109 [40], which have 5, 14, 100, 100, and 109 images, respectively, are used for comparison evaluation with other state-of-the-art methods. In addition, we employ two common image quality metrics, higher PSNR, and structural similarity (SSIM) index [50] for numerical evaluation.

4.1 Implementation and Training Details

As more and more recent SR methods train their networks on DIV2K [1], which is a big dataset containing 800 high-quality images with 2K resolution, following the practice, our network is also trained on DIV2K. During training, the Adam optimization method [28] is used to update the model parameters where β_1 and β_2 are set to 0.9 and 0.999, respectively. Note that, RGB color channels are used for input and output images during training; however, for numerical evaluation, the PSNR and SSIM evaluations are calculated on the Y channel of the YCbCr color space. We implement our RKAN with the PyTorch framework and a NVIDIA 1080Ti GPU is used during training. In KAM, we set three branches containing 3×3, 5×5, and 7×7 kernels to fully exploit the selective kernel attention. To render a lightweight model named RKAN-L, dilated convolution is employed where a 3×3 convolution with Dilation-2 can approximate a 5×5 filter, likewise, a 3×3 convolution with Dilation-4 and 7×7 filter. With the help of dilated convolution, the parameters and computational costs are reduced substantially at the expense of little performance loss.

4.2 Analysis and Discussion

To validate the effect of the kernel attention mechanism in SR, we design several models with different branches. Referring to Table 1, these models are the combination of different kernels and SKA (selective kernel attention). By removing the SKA (the green box in Figure 1(d)), we simply sum up the results from different kernels. By checking Table 1, we have the following three observations: (1) Comparing the first three models, all of which only have one kernel, the model with the larger kernel size obtains better results. This is because a large kernel size leads to a wide receptive field, which can extract rich contextual features for SR reconstruction. (2) Comparing models 4–9, which contain two branches, SKA is also another control variable. The results indicate that the model with SKA achieves higher PSNR than the simple aggregation from each branch. (3) Comparing a 3-branch model 10 in absence of SKA with 2-branch models 5, 7, and 9, we can



(a) Different images on the same scale. (b) Same image for different scales. (c) Kernel number for each KAM.

Fig. 3. Analysis of kernel attention on Set14 dataset. (a) The attention result for different images with 2×. (b) The attention result for the same image with 2× and 4×. The results in (a) and (b) are based on the 5th KAM in the network. (c) The attention results are counted on each of the 9 KAMs. K3 denotes the 3 × 3 kernel, K5 denotes the 5 × 5 kernel, and K7 denotes the 7 × 7 kernel.

find that SKA with 2-branch models 5, 7, and 9 indeed improves the SR performance, in spite of the smaller number of parameters in 2-branch models.

To further investigate the internal mechanism in SKA, we counted the number of kernels with the maximum attention weight in the 5th KAM (in total, there are nine KAMs in our final 3-branch model), the total number of which is 64 as the channel number is 64. Because the 5th KAM is the center of the network, the statistical results in this block are most intuitive. First, we take two different images with the same size as input, and Figure 3(a) shows that for different textures, SKA assigns different numbers on different size kernels to focus on the informative feature. We then feed the same input for 2× and 4× scale factors. The statistical result in the 5th KAM shown in Figure 3(b) demonstrates that even for the same picture, the choice of kernel also makes a difference. Figure 3(c) shows the average choice of kernel in each KAM on Set14, and we can find that the 3 × 3 kernel is dominant in the initial stage; subsequently, there is a significant drop as the network becomes deeper. Through the analysis on Figure 3(c), we can find that the network prefers the small kernel (3 × 3) at the beginning stage, but the large kernel (5 × 5 and 7 × 7) gains increasing popularity in the later stage, which can be regarded as a guideline for deep network design in SR. Generally speaking, the results in Table 1 and Figure 3 indicate that the combination of multiple kernels and selection mechanism is indeed beneficial to SR reconstruction.

4.3 Ablation Study

We also conduct a comprehensive ablation study to demonstrate the effect of model design, and the results are shown in Figure 4. From Figure 4(a), it can be observed that increasing the number of KAMs and channel number is an effective way to improve the accuracy. Nevertheless, the aim of this work is to investigate the kernel selection mechanism in SR; thus, we only construct a lightweight network for fast training and testing. Figure 4(b) further demonstrates that the combination of multiple kernels is of great help for high quality SR reconstruction. In addition, Figure 4(c) shows the convergence curve of models with different pooling operations, and we can find that pooling operation is not a key issue in our method and has little impact on the final results during training. As average pooling still yields slightly higher PSNR in the final results, we adopt the average pooling in our KAM.

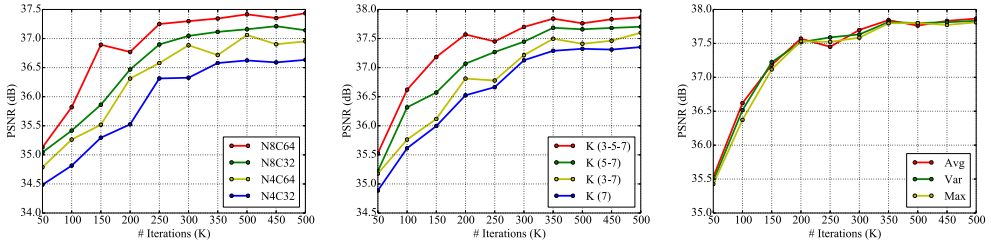
4.4 Comparisons with State-of-the-Arts

Finally, we compare the performance of our RKAN with the other state-of-the-art SR methods from two aspects: visual quality and numerical index. For numerical comparisons, Table 2 gives

Table 2. Quantitative Evaluation on Benchmark

Method	Scale	#Params (K)	Set5	Set14	BSDS100	Urban100	MANGA109
Bicubic	2×	—	33.69 / 0.931	30.25 / 0.870	29.57 / 0.844	26.89 / 0.841	30.86 / 0.936
SRCNN [8]	2×	57	36.72 / 0.955	32.51 / 0.908	31.38 / 0.889	29.53 / 0.896	35.76 / 0.968
VDSR [25]	2×	666	37.53 / 0.959	33.05 / 0.913	31.90 / 0.896	30.77 / 0.914	37.22 / 0.975
LapSRN [29]	2×	251	37.52 / 0.959	33.08 / 0.913	31.80 / 0.895	30.41 / 0.910	37.27 / 0.974
DRRN [46]	2×	297	37.74 / 0.959	33.23 / 0.914	32.05 / 0.897	31.23 / 0.919	37.92 / 0.976
SRMDNF [59]	2×	1,511	37.79 / 0.960	33.32 / 0.916	32.05 / 0.898	31.33 / 0.920	38.07 / 0.976
IDN [23]	2×	553	37.83 / 0.960	33.30 / 0.915	32.08 / 0.898	31.27 / 0.919	38.01 / 0.974
EDSR (base) [36]	2×	1,370	37.99 / 0.960	33.57 / 0.917	32.16 / 0.899	31.98 / 0.927	38.54 / 0.976
MSRN [31]	2×	5,930	38.08 / 0.960	33.74 / 0.917	32.23 / 0.901	32.22 / 0.932	38.82 / 0.986
DRFN [55]	2×	385	37.71 / 0.959	33.29 / 0.914	32.02 / 0.897	31.08 / 0.917	—/—
CARN [2]	2×	1,592	37.76 / 0.959	33.52 / 0.916	32.09 / 0.897	31.92 / 0.925	—/—
RKAN-L	2×	1,564	37.78 / 0.960	33.49 / 0.917	32.08 / 0.899	31.81 / 0.926	38.33 / 0.977
RKAN (ours)	2×	4,089	**37.98 / 0.961**	**33.70 / 0.919**	**32.12 / 0.899**	**32.08 / 0.928**	**38.45 / 0.976**
Bicubic	3×	—	30.39 / 0.868	27.55 / 0.774	27.21 / 0.738	24.46 / 0.734	26.95 / 0.855
SRCNN [8]	3×	57	32.75 / 0.909	29.30 / 0.821	28.41 / 0.786	26.24 / 0.798	30.48 / 0.911
VDSR [25]	3×	666	33.67 / 0.921	29.78 / 0.832	28.83 / 0.799	27.14 / 0.829	32.01 / 0.934
LapSRN [29]	3×	502	33.82 / 0.922	29.87 / 0.832	28.82 / 0.798	27.07 / 0.828	32.21 / 0.935
DRRN [46]	3×	297	34.03 / 0.924	29.96 / 0.835	28.95 / 0.800	27.53 / 0.764	32.74 / 0.939
SRMDNF [59]	3×	1,528	34.12 / 0.925	30.04 / 0.838	28.97 / 0.802	27.57 / 0.839	33.00 / 0.940
IDN [23]	3×	553	34.11 / 0.925	29.99 / 0.835	28.95 / 0.801	27.42 / 0.836	32.71 / 0.938
EDSR (base) [36]	3×	1,555	34.37 / 0.927	30.28 / 0.841	29.09 / 0.805	28.15 / 0.852	33.45 / 0.944
MSRN [31]	3×	6,114	34.38 / 0.926	30.34 / 0.839	29.08 / 0.804	28.08 / 0.855	33.44 / 0.942
DRFN [55]	3×	385	34.01 / 0.923	30.06 / 0.836	28.93 / 0.801	27.43 / 0.835	—/—
CARN [2]	3×	1,592	34.29 / 0.925	30.29 / 0.840	29.06 / 0.803	28.06 / 0.849	—/—
RKAN-L	3×	1,564	34.27 / 0.925	33.31 / 0.841	29.04 / 0.803	27.98 / 0.848	33.43 / 0.943
RKAN (ours)	3×	4,089	**34.45 / 0.927**	**30.43 / 0.844**	**29.07 / 0.804**	**28.18 / 0.852**	**33.54 / 0.944**
Bicubic	4×	—	28.43 / 0.811	26.01 / 0.704	25.97 / 0.670	23.15 / 0.660	24.93 / 0.790
SRCNN [8]	4×	57	30.50 / 0.863	27.52 / 0.753	26.91 / 0.712	24.53 / 0.725	27.66 / 0.858
VDSR [25]	4×	666	31.35 / 0.883	28.02 / 0.768	27.29 / 0.726	25.18 / 0.754	28.83 / 0.887
LapSRN [29]	4×	502	31.54 / 0.885	28.19 / 0.772	27.32 / 0.727	25.21 / 0.756	29.09 / 0.890
DRRN [46]	4×	297	31.68 / 0.888	28.21 / 0.772	27.38 / 0.728	25.44 / 0.764	29.18 / 0.891
SRMDNF [59]	4×	1,552	31.96 / 0.892	28.35 / 0.778	27.49 / 0.733	25.68 / 0.773	30.09 / 0.902
IDN [23]	4×	553	31.82 / 0.890	28.25 / 0.773	27.41 / 0.729	25.41 / 0.763	29.41 / 0.894
EDSR (base) [36]	4×	1,555	32.09 / 0.893	28.58 / 0.781	27.57 / 0.735	26.04 / 0.784	30.35 / 0.906
IMDN [22]	4×	715	32.21 / 0.894	28.58 / 0.781	27.56 / 0.735	26.04 / 0.783	30.45 / 0.907
MSRN [31]	4×	6,114	32.07 / 0.890	28.60 / 0.775	27.52 / 0.727	26.04 / 0.789	30.17 / 0.903
DRFN [55]	4×	385	31.55 / 0.886	28.30 / 0.773	27.39 / 0.729	25.45 / 0.762	—/—
CARN [2]	4×	1,592	32.13 / 0.893	28.60 / 0.780	27.58 / 0.734	26.07 / 0.783	—/—
RKAN-L	4×	1,564	32.14 / 0.894	28.57 / 0.780	27.55 / 0.734	26.00 / 0.781	30.42 / 0.907
RKAN (ours)	4×	4,089	**32.32 / 0.896**	**28.72 / 0.783**	**27.58 / 0.735**	**26.18 / 0.787**	**30.50 / 0.907**

Average PSNR and SSIM values for scale factors 4× and 8× on datasets Set5, Set14, BSDS100, Urban100 and Manga109. Bold text indicates the best performance.



(a) Training processes on KAMs. (b) Training processes on kernel. (c) Training processes on pooling.

Fig. 4. Convergence analysis of our model with various configurations. (a) Training processes for models with different numbers of KAMs (N) and channel numbers (C). (b) Training processes for models with different combinations of kernels. (c) Training processes for models with different pooling operations in KAM. Avg, Var, and Max denote the different global pooling methods, which are the average, variance, and maximum.

Table 3. Comparison of Computation Cost with Recent Methods

Method	EDSR [36]	RDN [62]	RCAN [60]	SRFBN [34]	SAN [6]	RKAN-L	RKAN
#Para. (K)	21,257	16,372	15,444	2,140	11,833	1,564	4,089
#FLOPs (M)	1,114,379	429,423	402,203	57,370	309,179	24,141	102,901
Time (ms)	155.36	105.12	315.76	166.71	405.56	29.23	57.34

The number of parameters, average running time and FLOPs of each method are evaluated on Set5 dataset with 2× upscaling factor.

the results in terms of PSNR and SSIM on five common-used benchmarks, where our method achieves satisfactory results on all enlargements. In addition, the number of parameters for each model is also given in Table 2. Although our RKAN performs the best on most datasets with all scaling factors, the parameters are relatively large due to the use of large kernels. To make the tradeoff between performance and network parameters, we also propose a lightweight model by employing the dilated convolution, where a 3×3 kernel with Dilation-2 approximates a 5×5 kernel, and a 3×3 kernel with Dilation-4 approximates 7×7 kernel. By doing so, the model benefits from significant parameter reduction at the expense the little PSNR loss. In a qualitative manner, Figure 5 gives some visual examples of reconstructed images on 4× factor. We can find that our RKAN can better alleviate blurring artifacts and the results from our model are as faithful as the ground-truth. Some visual details are missed with other methods, but our method is capable of constructing sharp and detailed structures. Especially the second row in Figure 5, except for our method, other methods hardly recover the accurate pattern. Through the comparison, the results suggest that the kernel attention mechanism is effective in estimating better visual details.

Besides, regarding the running efficiency, the comparison with recent leading SR methods is shown in Table 3. Although the compared methods in Table 3 achieve higher PSNR than our RKAN, all of them have a large number of parameters, resulting in very slow inference speed and taking up vast GPU memory. However, our method embraces the lightweight model size and fast running time. Especially, RKAN nearly achieves the real time. On the other hand, almost all the SR methods need to train multiple independent models to deal with the multi-scale SR task. As to our model, just by specifying the different scale factors during the test phase, a single model can deal with the multi-scale SR task, which greatly facilitates applications in real-world scenarios.

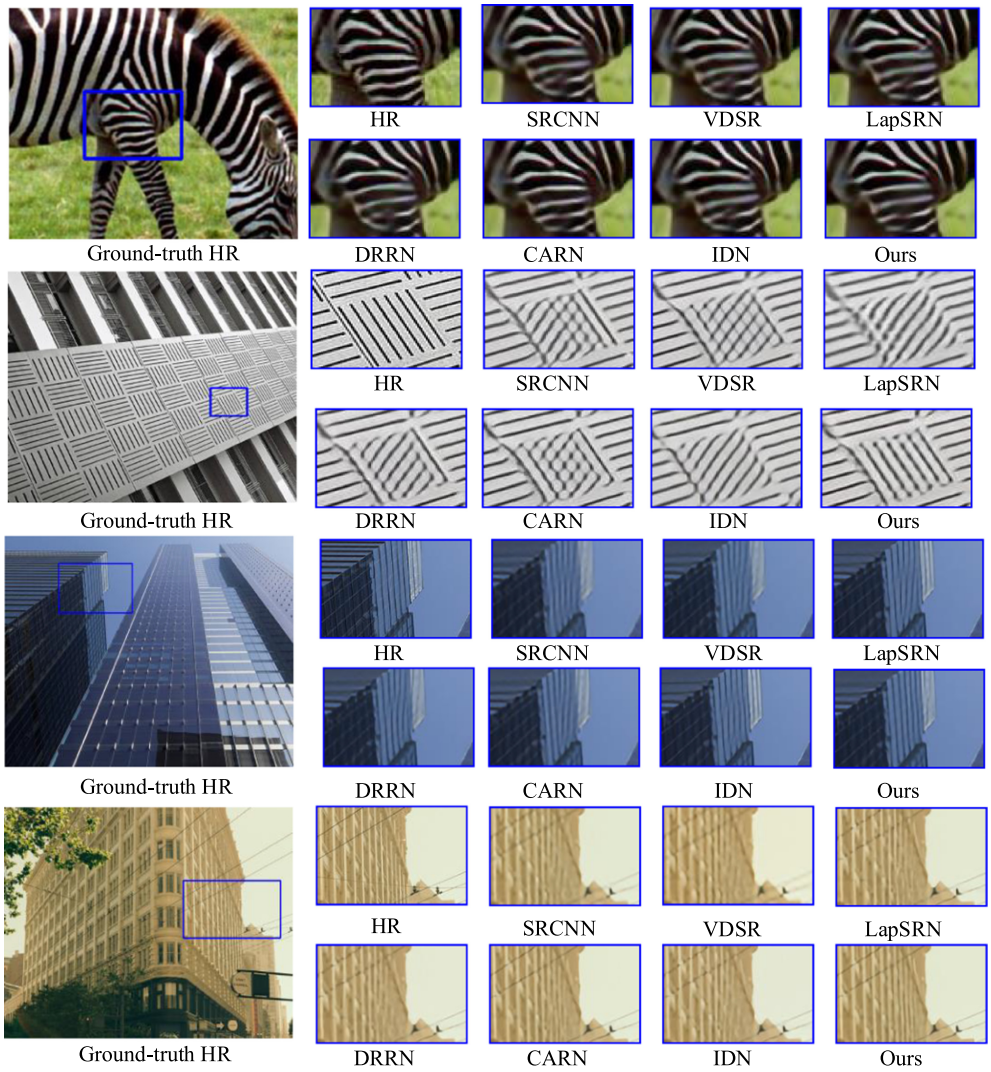


Fig. 5. Visual comparison of each method on 4× super-resolution.

5 CONCLUSION

In this article, we investigated the kernel attention mechanism in SR. By combining multiple branches containing different kernel sizes and selective kernel attention, the novel KAM is proposed to constitute the RKAN. To fully exploit the ability of KAM, which aggregates multi-scale information, we predefine three upscaling modules with different scale factors for joint training. As a result, in contrast to most existing models, our RKAN can deal with the multi-scale SR task by specifying the various scale factors in a forward pass. For the better tradeoff between model size and performance, RKAN-L is also proposed by adopting dilated convolution. In addition, we also give a comprehensive analysis on kernel selection, leading to a new understanding of architectural design for the SR task.

REFERENCES

[1] Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, July 21–26, 2017*. 1122–1131.

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. 2018. Fast, accurate, and lightweight super-resolution with cascading residual network. In *15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, September 8–14, 2018, Part X*. 256–272.

[3] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2018. Bottom-up and top-down attention for image captioning and visual question answering. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018*. 6077–6086.

[4] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. 2018. To learn image super-resolution, use a GAN to learn how to do image degradation first. In *15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, September 8–14, 2018, Part VI*. 187–202.

[5] Jie Chen, Jie Shao, and Chengkun He. 2020. Movie fill in the blank by joint learning from video and text with adaptive temporal attention. *Pattern Recognit. Lett.* 132 (2020), 62–68.

[6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. 2019. Second-order attention network for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, June 16–20, 2019*. 11065–11074.

[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In *13th European Conference on Computer Vision, ECCV 2014, Zurich, Switzerland, September 6–12, 2014, Part IV*. 184–199.

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 2 (2016), 295–307.

[9] Chao Dong, Chen Change Loy, and Xiaoou Tang. 2016. Accelerating the super-resolution convolutional neural network. In *14th European Conference on Computer Vision, ECCV 2016, Amsterdam, The Netherlands, October 11–14, 2016, Part II*. 391–407.

[10] Gilad Freedman and Raanan Fattal. 2011. Image and video upscaling from local self-examples. *ACM Trans. Graph.* 30, 2 (2011), 12:1–12:11.

[11] Lianli Gao, Xiangpeng Li, Jingkuan Song, and Heng Tao Shen. 2020. Hierarchical LSTMs with adaptive attention for visual captioning. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 5 (2020), 1112–1131.

[12] Jianting Guo, Peijia Zheng, and Jiwu Huang. 2017. An efficient motion detection and tracking scheme for encrypted surveillance videos. *TOMCCAP* 13, 4 (2017), 61:1–61:23.

[13] Wei Han, Shiyu Chang, Ding Liu, Mo Yu, Michael Witbrock, and Thomas S. Huang. 2018. Image super-resolution via dual-state recurrent networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018*. 1654–1663.

[14] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. 2018. Deep back-projection networks for super-resolution. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018*. 1664–1673.

[15] Chen He and Haifeng Hu. 2019. Image captioning with visual-semantic double attention. *TOMCCAP* 15, 1 (2019), 26:1–26:16.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, June 27–30, 2016*. 770–778.

[17] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. *CoRR* abs/1704.04861. arxiv:1704.04861

[18] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018*. 7132–7141.

[19] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao. 2019. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Trans. Circuits Syst. Video Techn.* DOI: https://doi.org/10.1109/TCSVT.2019.2915238

[20] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, July 21–26, 2017*. 2261–2269.

[21] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. 2015. Single image super-resolution from transformed self-exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, June 7–12, 2015*. 5197–5206.

[22] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. 2019. Lightweight image super-resolution with information multi-distillation network. In *27th ACM International Conference on Multimedia, MM 2019, Nice, France, October 21–25, 2019*. 2024–2032.

[23] Zheng Hui, Xiumei Wang, and Xinbo Gao. 2018. Fast and accurate single image super-resolution via information distillation network. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018*. 723–731.

[24] Michal Irani and Shmuel Peleg. 1991. Improving resolution by image registration. *CVGIP: Graphical Model and Image Processing* 53, 3 (1991), 231–239.

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, June 27–30, 2016*. 1646–1654.

[26] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Deeply-recursive convolutional network for image super-resolution. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, June 27–30, 2016*. 1637–1645.

[27] Jun-Hyuk Kim, Jun-Ho Choi, Manri Cheon, and Jong-Seok Lee. 2020. MAMNet: Multi-path adaptive modulation network for image super-resolution. *Neurocomput* 402 (2020), 38–49. DOI: https://doi.org/10.1016/j.neucom.2020.03.069

[28] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, May 7–9, 2015*.

[29] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2019. Fast and accurate image super-resolution with deep Laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 11 (2019), 2599–2613.

[30] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, July 21–26, 2017*. 105–114.

[31] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. 2018. Multi-scale residual network for image super-resolution. In *15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, September 8–14, 2018, Part VIII*. 527–542.

[32] Xianguo Li, Yemei Sun, Yanli Yang, and Changyun Miao. 2019. Symmetrical residual connections for single image super-resolution. *TOMCCAP* 15, 1 (2019), 19:1–19:10.

[33] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. 2019. Selective kernel networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, June 16–20, 2019*. 510–519.

[34] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. 2019. Feedback network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, June 16–20, 2019*. 3867–3876.

[35] Qianli Liao and Tomaso A. Poggio. 2016. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. *CoRR* abs/1604.03640 (2016).

[36] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, July 21–26, 2017*. 1132–1140.

[37] Heng Liu, Jungong Han, Shudong Hou, Ling Shao, and Ruan Yue. 2018. Single image super-resolution using a deep encoder-decoder symmetrical network with iterative back projection. *Neurocomput.* 282 (2018), 52–59.

[38] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. 2016. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain*. 2802–2810.

[39] David R. Martin, Charless C. Fowlkes, Doron Tal, and Jitendra Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *8th International Conference On Computer Vision (ICCV-01), Vancouver, British Columbia, Canada, July 7–14, 2001, Volume 2*. 416–425.

[40] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2017. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools Appl.* 76, 20 (2017), 21811–21838.

[41] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. 2014. Recurrent models of visual attention. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada*. 2204–2212.

[42] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon. 2018. BAM: Bottleneck attention module. In *British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3–6, 2018*. 147.

[43] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, June 27–30, 2016*. 1874–1883.

[44] Assaf Shocher, Nadav Cohen, and Michal Irani. 2018. "Zero-shot" Super-resolution using deep internal learning. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018.* 3118–3126.

[45] Jingkuan Song, Yuyu Guo, Lianli Gao, Xuelong Li, Alan Hanjalic, and Heng Tao Shen. 2019. From deterministic to generative: Multimodal stochastic RNNs for video captioning. *IEEE Trans. Neural Networks Learn. Syst.* 30, 10 (2019), 3047–3058.

[46] Ying Tai, Jian Yang, and Xiaoming Liu. 2017. Image super-resolution via deep recursive residual network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, July 21–26, 2017.* 2790–2798.

[47] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. 2017. MemNet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017.* 4549–4557.

[48] Anqi Wang, Haifeng Hu, and Liang Yang. 2018. Image captioning with affective guiding and selective attention. *TOMCCAP* 14, 3 (2018), 73:1–73:15.

[49] Bokun Wang, Yang Yang, Xing Xu, Alan Hanjalic, and Heng Tao Shen. 2017. Adversarial cross-modal retrieval. In *2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, October 23–27, 2017.* 154–162.

[50] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 4 (2004), 600–612.

[51] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. 2019. Deep learning for image super-resolution: A survey. *CoRR* abs/1902.06068.

[52] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional block attention module. In *15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, September 8–14, 2018, Part VII.* 3–19.

[53] Xing Xu, Fumin Shen, Yang Yang, Heng Tao Shen, and Xuelong Li. 2017. Learning discriminative binary codes for large-scale cross-modal retrieval. *IEEE Trans. Image Process.* 26, 5 (2017), 2494–2507.

[54] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. 2008. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2008, June 24–26, 2008, Anchorage, AK.*

[55] Xin Yang, Haiyang Mei, Jiqing Zhang, Ke Xu, Baocai Yin, Qiang Zhang, and Xiaopeng Wei. 2019. DRFN: Deep recurrent fusion network for single-image super-resolution with large factors. *IEEE Trans. Multimedia* 21, 2 (2019), 328–337.

[56] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. 2018. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, June 18–22, 2018.* 701–710.

[57] Dongyang Zhang, Jie Shao, Gang Hu, and Lianli Gao. 2017. Sharp and real image super-resolution using generative adversarial network. In *24th International Conference on Neural Information Processing, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Part III.* 217–226.

[58] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. 2017. Learning deep CNN denoiser prior for image restoration. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, July 21–26, 2017.* 2808–2817.

[59] Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2018. Learning a single convolutional super-resolution network for multiple degradations. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018.* 3262–3271.

[60] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. 2018. Image super-resolution using very deep residual channel attention networks. In *15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, September 8–14, 2018, Part VII.* 294–310.

[61] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. 2019. Residual non-local attention networks for image restoration. In *International Conference on Learning Representations, ICLR 2019, New Orleans, LA.*

[62] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. 2018. Residual dense network for image super-resolution. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, June 18–22, 2018.* 2472–2481.

[63] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. 2017. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* 3, 1 (2017), 47–57.

[64] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017.* 2242–2251.

Received May 2019; revised April 2020; accepted May 2020