IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 18, 2021, accepted March 27, 2021, date of publication March 30, 2021, date of current version April 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069777

Mixed High-Order Non-Local Attention Network for Single Image Super-Resolution

XIAOBIAO DU[1], SAIBIAO JIANG[1,2], YUJUAN SI[1], LINA XU[1,3], AND CHONGJIN LIU[1]

[1]Zhuhai College of Jilin University, Zhuhai 519041, China
[2]Department of Electromechanical Engineering, University of Macau, Zhuhai 999078, China
[3]College of Instrument Science and Electrical Engineering, Jilin University, Changchun 130061, China

Corresponding author: Saibiao Jiang (jiangsb827@163.com)

This work was supported by the Special Project in Key Areas of Universities in Guangdong Province under Grant 2020ZDZX3016.

ABSTRACT Attention has been diffusely used in many tasks since it can guide network concentrating on the most important regions of an input pattern. Nevertheless, many advanced works focus on first-order attention design, e.g. channels and spatial attention, but ignore higher-order attention mechanisms. In this work, we propose the Mixed High-Order Attention (MHA) module to model the complex and high-order information in the attention mechanism, which captures the subtle texture and outputs the discriminative attention map. Besides, the region of the convolution is local, which can't capture global context and long-range dependencies. Therefore, we propose a non-local block to obtain global attention features. We also propose the Mixed High-Order Non-local Attention Network (MHNAN) to improve the richness of attention. Extensive experiments are conducted to demonstrate the superiority of our MHNAN for super-resolution over several state-of-the-art models.

INDEX TERMS Super resolution, deep neural network, deep learning.

I. INTRODUCTION

Single image super-resolution (SISR) has attracted a lot of attention. The task of SISR is to recovery clear high-resolution (HR) images given its low-resolution (LR) images. Nevertheless, many high-resolution solutions can map to any LR input, thus this is an ill-posed problem. Therefore, researchers proposed a lot of SR models, ranging from model-based and interpolation-based [1]–[3], to current learning-based methods [4]. These conventional methods are efficient, but they still exist some drawbacks: the accuracy may quickly decrease; these methods are time-consuming.

Deep learning has achieved unexpected results in many different fields [5], [6]. Recently, a lot of CNN-based SR models concentrate on learning to map HR images given LR images. CNNs have obtained satisfactory results in SISR [7]–[16] through exploiting the image patterns. Dong *et al.* [7] first try to employ CNN to tackle SISR. They proposed a CNN-based model named SRCNN to learn a non-linear transformation from LR to HR images. [10], [17] proposed much deeper networks through integrating recursive learning and residual learning, obtaining significant results over the SRCNN. Nevertheless, the above methods use the interpolation method to firstly upscale an low-resolution image to the desired output size and then fed it into the network, thus suffering from reconstruction artifacts and extra computational cost. In order to deal with this problem, some researchers proposed new methods, which upscale the final low-resolution feature in the tail of the network. Dong *et al.* [8] proposed a deconvolution module at the tail of the architecture to upsample the final LR information. Shi *et al.* [18] proposed a novel upscale module named sub-pixel convolution layer to upsample the LR feature to the HR feature at the tail of the network. Therefore, more and more researchers employ this efficient post-processing method, which not only reduces computational load but also deepens the network. Recently, Lim *et al.* [9] proposed a network by integrating residual learning, named EDSR, which won the NTIRE2017 [19] SR Challenge championship [19].

Nevertheless, a lot of CNN-based super-resolution models still have some drawbacks: (1) most of the CNN-based models ignore the high-order information, leading to not get the utmost out of the information from the blurry LR pattern, thus resulting in unsatisfying performance; (2) the region of the

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

convolution is local, which can't capture long-range dependencies and rarely exploiting the global feature correlations, thus limiting the learning ability of CNNs.

In order to tackle the above issues, we introduce a deep Mixed High-Order Non-local Attention Network (MHNAN) for mixed high-order feature extraction. Especially, a mixed high-order attention (MHA) module is introduced to extract and mix high-order information. Through exploiting mixed high-order feature information, our MHA adaptively learns feature inter-dependencies. Such MHA guides our network to focus on the significant pattern and improve performance. Besides, a non-local enhanced group (NEG) is proposed to integrate non-local operations to model long-range dependencies. We extract the pattern from the LR images by stacking the non-local residual channel attention groups (NRCAG) structure. To sum up, our contributions are as follow:

- We introduce a Mixed High-Order Non-local Attention Network (MHNAN) to tackle the image super-resolution task. Extensive experiments on multiple datasets show that our MHNAN is superior to many advanced models.
- We propose mixed high-order attention (MHA) module, which is deep plug-and-play and can adaptively capture features through considering feature information higher than first-order. Thus, this MHA mechanism makes the network enhance discriminative learning ability and concentrate on the more important features.
- We propose a non-local enhanced group (NEG), which further integrates non-local operations to extract long-range pattern information.

II. MIXED HIGH-ORDER NON-LOCAL ATTENTION NETWORK

A. NETWORK FRAMEWORK

As can be seen from Figure 2, our MHNAN primarily comprises the following parts: shallow feature extractor, non-local enhanced group (NEG), up-scale layer, and reconstruction layer. Give I_{LR} and I_{SR} as the input and output of our MHNAN. Following the [9], [20], we use a convolution layer to extract the shallow feature F_0 from the low-resolution input

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

where H_{SF} stands for the convolution operation. Then we feed the shallow feature F_0 in NEG, which obtains the deep feature as

$$F_{DF} = H_{NEG}(F_0) \tag{2}$$

where H_{NEG} denotes the NEG based non-local enhanced group, which comprises multiple non-local residual channel attention groups to extract the long-range information and channel information. Then the deep feature F_{DF} is upsampled by the upscale layer through

$$F_{\uparrow} = H_{\uparrow}(F_{DF}) \tag{3}$$

where $F_{\uparrow}$ and $H_{\uparrow}$ are upsampled feature and upsample layer respectively. We have many methods to apply as an upscale module, like deconvolution [8], ESPCN [18]. We embed the upscaled module in the tail of the network to achieve high performance. This method is preferable in current super-resolution models [8], [9], [20]. The upscaled feature fed in one convolution layer to reconstruct high-resolution information

$$I_{SR} = H_R(F_{\uparrow}) = H_{MHNAN}(I_{LR}) \tag{4}$$

where H_R, $H_{\uparrow}$, and H_{MHNAN} are the reconstruction layer, upsample layer, and the function of MHNAN, respectively.

Then we use a loss function to optimize MHNAN. There are some widely used loss functions, such as L2, L1, perceptual loss. To demonstrate the effectiveness of our MHNAN, we use the L1 loss. Given a training set with N low-resolution images and high-resolution images denoted by $\{I_{HR}, I_{HR}\}^N$, the purpose of the MHNAN is to optimize the loss function:

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} ||I_{HR} - I_{SR}||_1 \tag{5}$$

where θ denotes the parameter of MHNAN. We apply Adam optimizer to minimize the loss.

B. NON-LOCAL ENHANCED GROUP (NEG)

We here describe the proposed non-local enhanced group (NEG) (see Figure. 1). The NEG is composed of several non-local residual channel attention groups (NRCAG) and an MHA module. The non-local block can extract the long-range feature. Each NRCAG further contains M simplified residual channel attention blocks [4].

We here detailedly describe the non-local block. We give an input image with shape (H, W, C_{in}). For simplicity, we have omitted the batch dimension. It can be expressed as:

$$O_h = \text{Soft max}(\frac{(XW_q)(XW_k)^T}{\sqrt{d_k^h}})(XW_v) \tag{6}$$

where W_q, $W_k \in \mathbb{R}^{C_{in}\times d_k^h}$ and $W_v \in \mathbb{R}^{C_{in}\times d_v^h}$ transform the X to values $V = XW_v$, keys $K = XW_k$, and queries $Q = XW_q$ through learned linear transformations.

Stacking several blocks may be a helpful and simple way to improve network [9], [11]. However, the network constructs in this method would result in training difficulty and performance bottleneck due to the problem of gradient exploding and vanishing. Simply stacking repeated blocks may not achieve better results. To handle this problem, we propose the NEG to bypass abundant low-frequency features and facilitate the training of the network. Then NRCAG in the g-th group can be denoted as:

$$F_g = H_g(F_{g-1}) \tag{7}$$

where H_g is the operation of the g-th NRCAG. F_g, F_{g-1} represent the output and input of the g-th NRCAG. Then deep feature can be extracted as:

$$F_{DF} = F_0 + F_g \tag{8}$$

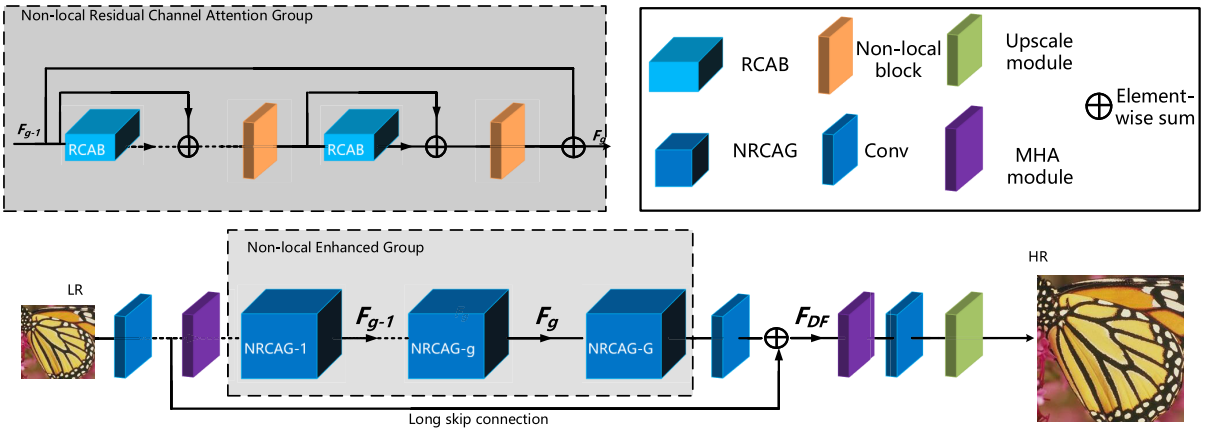

FIGURE 1. Framework of the proposed Mixed High-Order Non-local Attention Network (MHNAN).

C. MIXED HIGH-ORDER ATTENTION

Attention is a tool that biases the allocation of available resources towards the most useful part of the pattern. In CNN, it is widely used to weight convolution response graphs to highlight significant parts and suppress irrelevant information, like channel [21]and spatial attention [22]–[28]. Given input image $\mathcal{X} \in \mathbb{R}^{C\times H\times W}$,We briefly formulate spatial and channel attention to a general case:

$$F = A(I) \odot I \tag{9}$$

where $\odot$ represents the Hadamard Product, $A(I) \in \mathbb{R}^{C\times H\times W}$. As $A(I)$ serves as a attention map, the $A(I)$ is in the range $[0, 1]$. Considering the representation of attention, we describe $A(I)$ with many ways. For example, if $A(I) = rep[V]|^{H,W}$ where $rep[V]|^{H,W}$ means copy this scale vector along width and height dimensions by W and H times respectively and $V \in \mathbb{R}^C$ is a scale vector. Therefore, Eq. 9 is the implementation of *channel attention*. And if $A(I) = rep[M]|^C$ where $rep[M]|^C$ means copy this spatial mask M along channel dimension by C times and $M \in \mathbb{R}^{H\times W}$ is a spatial mask. Therefore, Eq. 9 represents the implementation of *spatial attention*.

Nevertheless, *channel attention* or *spatial attention* cannot learn the high-order information, leading to failure in capturing the subtle texture in the image. Therefore, we concentrate on modeling $A(I)$ with high-order feature representation.

$$A(i) = sigmoid(\sum_{r=1}^{R} \hat{\alpha}^{r^T} \sigma(\mathrm{z}^r)) \tag{10}$$

where α^r is the weight vector and σ represents an non-linear activation function, like ReLU function. We denote R as the number of order. $A(i)$ in Eq.10 is used as the required mixed high-order attention feature for the i.

III. EXPERIMENTS

A. SETUP

Following [9], [15], we use 800 training images from DIV2K dataset [19] as training data. In order to demonstrate the performance of our MHNAN, we use 5 public datasets: Set5, Set14 BSD100, Urban100, and Manga109. The Matlab resize function with the bicubic operation was adopted by us as a degradation model. The PSNR and SSIM were evaluated as SR results.

In the training stage, the LR images are augmented by horizontally flipping and randomly rotating 90°, 180°, 270°. We set R as 4, which means our MHA with order $\{1, 2, 3, 4\}$. We set 16 low-resolution image patches with the size of 48×48 as mini-batch. The ADAM algorithm was applied to optimize MHNAN with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. The learning rate was initialized as 10^{-4} and then reduced to half every 200 epochs.

B. IMPLEMENTATIONS

We fix the NRCAG number as $G = 18$ in the NEG and embed two MHA modules at the head and tail of NEG. In each NRCAG, we set $M = 10$ residual channel attention blocks. In addition to the shallow extract layer and upscale layer, we set the number of the filter as $C = 64$. For the upscale layer, we follow the works in [4], [15] and use sub-pixel convolution [18] to upscale and reconstruct the deep feature, followed by a 1 x 1 convolution with three filters to output RGB images.

C. ABLATION STUDY

As can be seen from Figure 1, our MHNAN contains two primary components, including non-local enhanced group (NEG) and mixed high-order attention (MHA) modules. To demonstrate the effectiveness of the various modules, we train and test MHNAN with its variants for comparison. Specific results are shown in Table 1.

TABLE 1. Effects of different modules. We report the best PSNR (DB) values on BSD100 (4X) in 5.6 X 10^5 iterations.

	R_{base}	R_a	R_b	R_c	R_e	R_d
MHA(before NEG)		✓			✓	✓
MHA(after NEG)			✓		✓	✓
Non-local block				✓		✓
PSNR	27.43	27.57	27.59	27.61	27.69	27.75

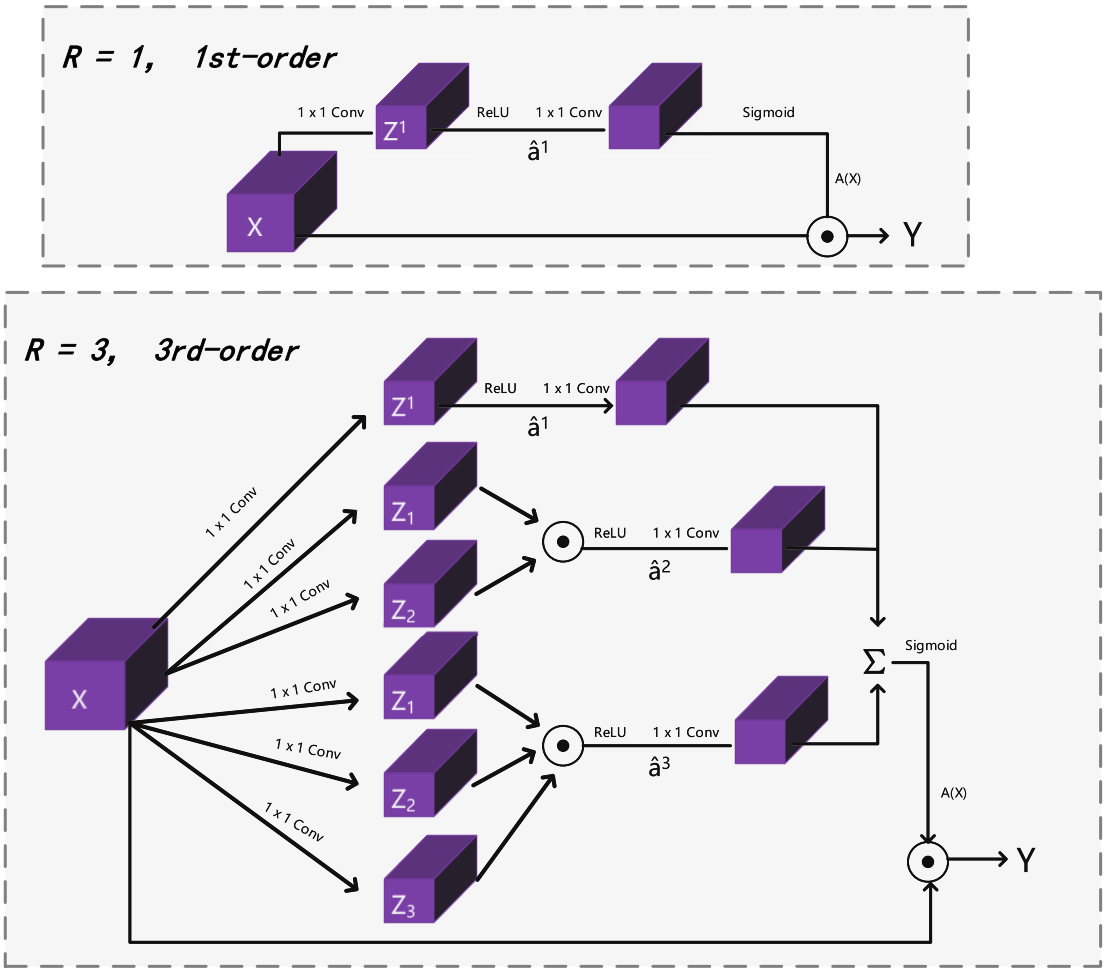

FIGURE 2. Illustration of Mixed High-Order Attention (MHA) modules.

We set R_{base} as a basic baseline, which only contains the convolutional layer containing 18 residual channel attention groups RCAGs and 10 residual channel attention blocks [4] in each NRCAG. It can be found that the performance of the R_{base} only reaches 27.43 PSNR. When we embed the MHA module before NEG, the performance of R_a reaches 27.57 PSNR. If we only embed the MHA module after NEG, the performance of R_b reaches 27.59 PSNR. This phenomenon demonstrates the effect of our MHA from the results of R_a and R_b. Specifically, R_c means that non-local blocks are employed to capture long-range dependencies by a self-attention mechanism. It can be found that the performance of the R_c can reach 27.61 PSNR when we employ the non-local block to our model. R_e means that the result of both applying MHA before and after NEG. It can be found that both of R_e achieve better results than methods of R_a to R_c, which reach 27.69 PSNR. When we combine MHA and non-local block, R_d can reach 27.75 PSNR.

D. RESULTS WITH BICUBIC DEGRADATION (BI)

We conduct a comparison test with model 14 state-of-the-art CNN-based SR methods: SRCNN [7], FSRCNN [8], VDSR [10], LapSRN [14], UDM-F3 [29], MemNet [13], HDRN [30], EDSR [9], SRMD [31], DBPN [32], RDN [15], RCAN [4], SAN [20] and ESRN [33] to verify the effectiveness of MHNAN. Table 2 demonstrates the quantitative results for each scaling factor. Compared to other models, our MHNAN achieves the best performance. Compared to RCAN, our MHNAN has satisfactory performance for datasets with rich texture information, It should be noted that textures are higher-order information with more complex statistical properties. However, edges are first-order information. Therefore, our MHA based on mixed high-order feature information and non-local operators works better on images with higher-order information (such as textures).

As can be seen from Figure 3, we show the visualization of different methods. Many super-resolution models fail to reconstruct the lattice and have severe blurring artifacts, but our MHNAN achieves sharper results and reconstructs more texture details. In the case of "img005", most models output severe blurring artifacts. Compared to ground-truth, MHNAN can restore more image details and get more reliable results. Although recovering high-frequency texture on the

TABLE 2. Quantitative results with bi degradation model.

Method	Scale	Set5	Set14	BSD100	Urban100	Manga109
Bicubic	2	33.66/.9299	30.24/.8688	SSIM	26.88/.8403	30.80/.9339
SRCNN	2	36.66/.9542	32.45/.9067	31.36/.8879	29.50/.8946	35.60/.9663
FSRCNN	2	37.05/.9560	32.66/.9090	31.53/.8920	29.88/.9020	36.67/.9710
VDSR	2	37.53/.9590	33.05/.9130	31.90/.8960	30.77/.9140	37.22/.9750
LapSRN	2	37.52/.9591	33.08/.9130	31.08/.8950	30.41/.9101	37.27/.9740
MemNet	2	37.78/.9597	33.28/.9142	32.08/.8978	31.31/.9195	37.72/.9740
UDM-F3	2	37.53/.9580	33.16/.9132	31.71/.8978	31.80/.9155	37.74/.9740
HDRN	2	37.75/.9550	33.49/.915	32.03/.8978	31.87/.9155	38.07/.9770
EDSR	2	38.11/.9602	33.92/.9195	32.32/.9013	32.93/.9351	39.10/.9773
SRMD	2	37.79/.9601	33.32/.9159	32.05/.8985	31.33/.9204	38.07/.9761
DBPN	2	38.09/.9600	33.85/.9190	32.27/.9000	32.55/.9324	38.89/.9775
RDN	2	38.24/.9614	34.01/.9212	32.34/.9017	32.89/.9353	39.18/.9780
RCAN	2	38.27/.9614	34.11/.9216	32.41/.9026	33.34/.9384	39.43/.9786
SAN	2	38.31/.9620	34.07/.9213	32.42/.9028	33.10/.9370	39.32/.9792
ESRN	2	38.04/.9607	33.71/.9185	32.23/.9005	32.37/.9310	-
MHNAN	2	**38.39/.9652**	**34.15/.9231**	**32.47/.9121**	**33.14/.9481**	**39.43/.9891**
Bicubic	3	39.32/.9792	27.55/.7742	27.21/.7385	24.46/.7349	26.95/.8556
SRCNN	3	32.75/.9090	29.30/.8215	28.41/.7863	26.24/.7989	30.48/.9117
FSRCNN	3	33.18/.9140	29.37/.8240	28.53/.7910	26.43/.8080	31.10/.9210
VDSR	3	33.67/.9210	29.78/.8320	28.83/.7990	27.14/.8290	32.01/.9340
LapSRN	3	33.82/.9227	29.87/.8320	28.82/.7980	27.07/.8280	32.21/.9350
MemNet	3	34.09/.9248	30.01/.8350	28.96/.8001	27.56/.8376	32.51/.9369
UDM-F3	3	33.49/.9228	29.97/.8340	28.73/.8001	27.13/.8336	32.30/.9359
HDRN	3	34.24/.9240	30.23/.8440	28.96/.8041	27.93/.8496	33.17/.9459
EDSR	3	34.65/.9280	3.52/ .8462	29.25/.8093	28.80/.8653	34.17/.9476
SRMD	3	34.12/.9254	30.04/.8382	28.97/.8025	27.57/.8398	33.00/.9403
RDN	3	34.71/.9296	30.57/.8468	29.26/.8093	28.80/.8653	34.13/.9484
RCAN	3	34.74/.9299	30.64/.8481	29.32/.8111	29.08/.8702	34.43/.9498
SAN	3	34.75/.9300	30.59/.8476	29.33/.8112	28.93/.8671	34.30/.9494
ESRN	3	34.46/.9281	30.43/.8439	29.15/.8072	28.42/.8579	-
MHNAN	3	**34.77/.9321**	**30.63/.8521**	**29.44/.8221**	**28.98/.8771**	**34.41/.9477**
Bicubic	4	28.42/.8104	26.00/.7027	25.96/.6675	23.14/.6577	24.89/.7866
SRCNN	4	30.48/.8628	27.50/.7513	26.90/.7101	24.52/.7221	27.58/.8555
FSRCNN	4	30.72/.8660	27.61/.7550	26.98/.7150	24.62/.7280	27.90/.8610
VDSR	4	31.35/.8830	28.02/.7680	27.29/.0726	25.18/.7540	28.83/.8870
LapSRN	4	31.54/.8850	28.19/.7720	27.32/.7270	25.21/.7560	29.09/.8900
MemNet	4	31.74/.8893	28.26/.7723	27.40/.7281	25.50/.7630	29.42/.8942
UDM-F3	4	31.67/.8860	28.22/.7723	27.15/.7271	25.31/.7610	29.10/.8912
HDRN	4	32.23/.8860	28.58/.7823	27.53/.7271	26.09/.7870	30.43/.9012
EDSR	4	32.46/.8968	28.80/.7876	27.71/.7420	26.64/.8033	31.02/.9148
SRMD	4	31.96/.8925	28.35/.7787	27.49/.7337	25.68/.7731	30.09/.9024
DBPN	4	32.47/.8980	28.82/.7860	27.72/.7400	26.38/.7946	30.91/.9137
RDN	4	32.47/.8990	28.81/.7871	27.72/.7419	26.61/.8028	31.00/.9151
RCAN	4	32.62/.9001	28.86/.7888	27.76/.7435	26.82/.8087	31.21/.9172
SAN	4	32.64/.9003	28.92/.7888	27.78/.7436	26.79/.8068	31.18/.9169
ESRN	4	32.26/.8957	28.63/.7818	27.62/.7378	26.24/.7912	-
MHNAN	4	**32.64/.9014**	**28.93/.7894**	**27.81/.7438**	**26.81/.8093**	**31.23/.9188**

input information of the limited LR is hard, our MHNAN can make the best use of the limited low-resolution pattern to transfer mixed high-order non-local attention and has a more powerful pattern representation, which results in more accuracy.

E. COMPONENT ANALYSIS

To explore the effect of the different number of orders in the MHA module, we perform quantitative comparisons on MHNAN. As shown in Table 3, we can observe that MHNAN enhances the performance of the model for single image super-resolution tasks over both RCAN and SAN. Especially, comparing MHNAN-2 with RCAN and SAN, it can be found that using higher-order attention patterns indeed enhances the learning ability of the network. Moreover, the performance will further improve with the number of order. When increasing the number of order 2 to 4, the performance of MHNAN-4 is higher than MHNAN-2. This phenomenon demonstrates that employing a more large number of order to MHNAN is

TABLE 3. Effect of attention modules. MHNAN-R denote the number of order for the MHA module.

Method	Scale	BSD100	Urban100	Manga109
RCAN	4	27.76/.7435	26.82/.8087	31.21/.9172
SAN	4	27.78/.7436	26.79/.8068	31.18/.9169
MHNAN-2	4	27.79/.7453	26.84/.8127	31.12/.9028
MHNAN-4	4	**27.81/.7538**	**26.97/.8193**	**31.23/.9188**
MHNAN-6	4	27.80/.7489	26.91/.8176	31.17/.9125

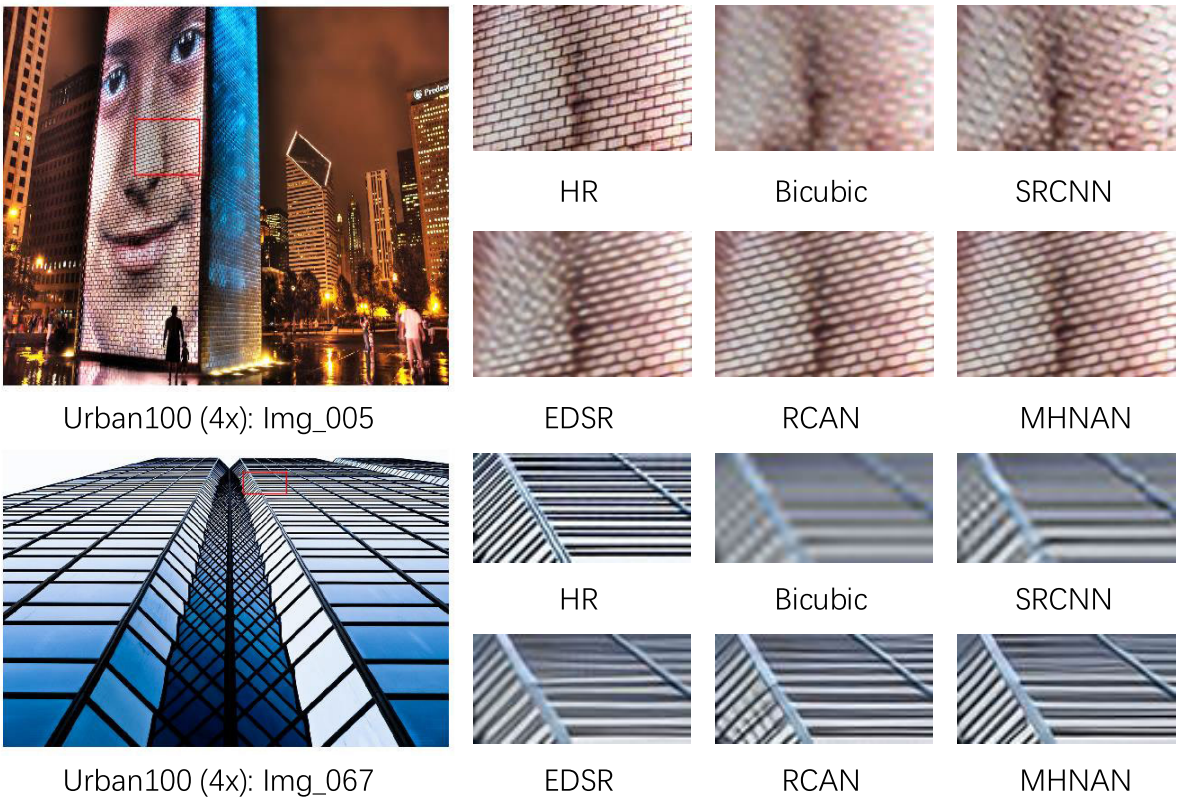

FIGURE 3. Visual comparison for 4× SR with BI model on Urban100 dataset.

beneficial to capture high-order information, and MHNAN-4 outperforms all the baseline models, showing the effectiveness of our network. Nevertheless, when increasing the number of order to 6, there was little performance improvement. Therefore, we're not reporting here.

F. MODEL SIZE ANALYSES

Table 4 shows the model size and performance of the current CNN SR model. In these methods, MemNet and NLRG contain far fewer parameters, which reduces performance. Not only does MHNAN have fewer parameters than RDN and RCAN, but it also gets the better performance, which means that MHNAN can achieve a good performance compromise between model complexity and performance.

TABLE 4. Computational and parameter comparison (2X) set5).

	EDSR	MemNet	NLRG	DBPN	RDN	RCAN	MHNAN
Para.	43M	677k	330k	10M	22.3M	16M	15.2M
PSNR	38.11	37.78	38.00	38.09	38.24	38.27	38.39

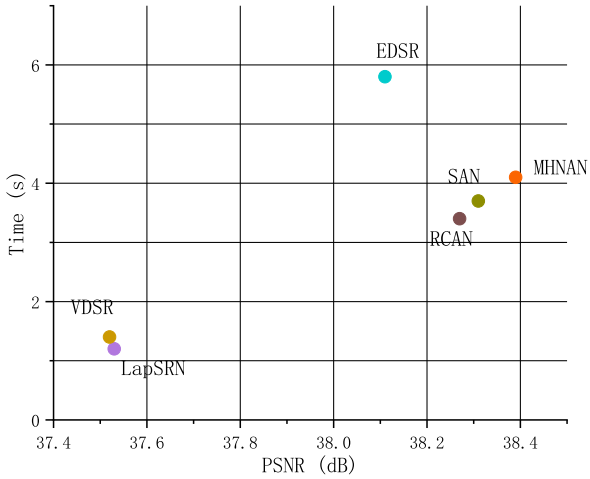

FIGURE 4. Speed comparison for 2× SR with BI model on Set5 dataset.

G. INFERENCE ANALYSES

Figure 4 shows the speed comparison for 2× scale factor with several state-of-the-art models on the Set5 dataset. It can be found that our MHNAN achieves best results compared with other models. Moreover, the inference time of MHAN is largely reduced compared with EDSR. Although the inference time of MHAN is a little slower than RCAN [15] and SAN [20], our MHNAN achieves better results, which demonstrates our MHNAN obtain a trade-off between inference time and performance.

H. VISUAL ANALYSES OF MHA

In order to demonstrate the effectiveness of our proposed MHA, we visualize the average feature maps for the input

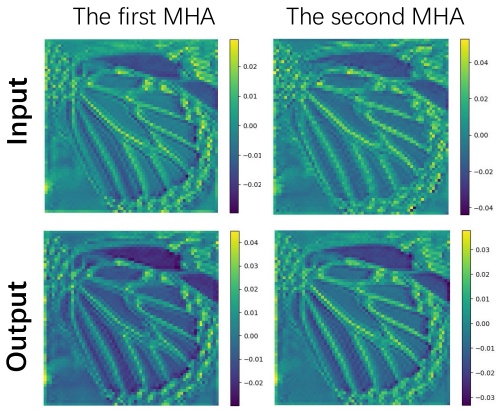

FIGURE 5. The visual of average feature maps on MHA.

and output of MHA. As can be seen from Figure 5, we can find the bottom line, the visual of the output of the MHA has more rich texture than the input of MHA, which indicates our proposed MHA can extract high-order feature, like texture. Besides, we can find the visual of the output of the MHA have more negative values than the input, which indicates our proposed MHA can suppress smoothness, such that extract more texture.

IV. CONCLUSION

In this work, we propose the mixed high-order attention (MHA) module to further enhance the discrimination of attention proposals through capturing the complex and high-order information. We also propose the Mixed High-Order Non-local Attention Network (MHNAN) to tackle the SISR task. Importantly, MHNAN achieved promising results through applying the MHA module and NEG to single image super-resolution task. Extensive experiments demonstrate our superiority of our MHNAN.

REFERENCES

[1] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 20–23, Jan. 2010.

[2] X. Chen and C. Qi, "Low-rank neighbor embedding for single image super-resolution," *Signal Process.*, vol. 94, no. 1, pp. 6–22, 2014.

[3] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[4] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[5] S. B. Jiang, P. K. Wong, and Y. C. Liang, "A fault diagnostic method for induction motors based on feature incremental broad learning and singular value decomposition," *IEEE Access*, vol. 7, pp. 157796–157806, 2019.

[6] S. B. Jiang, P. K. Wong, R. Guan, Y. Liang, and J. Li, "An efficient fault diagnostic method for three-phase induction motors based on incremental broad learning and non-negative matrix factorization," *IEEE Access*, vol. 7, pp. 17780–17790, 2019.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[8] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 391–407.

[9] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[10] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[11] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 517–532.

[12] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[13] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[14] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[15] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[16] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," 2019, *arXiv:1903.10082*. [Online]. Available: http://arxiv.org/abs/1903.10082

[17] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.

[18] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[19] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 114–125.

[20] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11065–11074.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[22] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[23] D. M. Vo and S.-W. Lee, "Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 18689–18707, Jul. 2018.

[24] J. Li, Z. L. Yu, Z. Gu, H. Liu, and Y. Li, "Dilated-inception net: Multi-scale feature aggregation for cardiac right ventricle segmentation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3499–3508, Dec. 2019.

[25] N. Wang, Y. Zhang, and L. Zhang, "Dynamic selection network for image inpainting," *IEEE Trans. Image Process.*, vol. 30, pp. 1784–1798, 2021.

[26] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.

[27] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y.-K. Wong, "Learning spatial attention for face super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021.

[28] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1423–1438, 2021.

[29] P. Yi, Z. Wang, K. Jiang, Z. Shao, and J. Ma, "Multi-temporal ultra dense memory network for video super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2503–2516, Aug. 2020.

[30] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320320302788

[31] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.
[32] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.
[33] D. Song, C. Xu, X. Jia, Y. Chen, C. Xu, and Y. Wang, "Efficient residual dense block search for image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 12007–12014.

XIAOBIAO DU is currently pursuing the bachelor's degree with the Zhuhai College of Jilin University. His research interests include super-resolution, image segmentation, and machine learning.

SAIBIAO JIANG received the M.Sc. degree from the Guangdong University of Technology, in 2011. He is currently pursuing the Ph.D. degree in electromechanical engineering with the University of Macau. He is also a Lecturer with the Zhuhai College of Jilin University. His research interests include data-driven fault diagnosis, feature extraction, and artificial intelligence.

YUJUAN SI received the master's and Ph.D. degrees in engineering from the Jilin University of Technology, in 1988 and 1996, respectively. She is currently a Professor with the Institute of Communication Engineering, Jilin University. She is also a Lecturer with the Zhuhai College of Jilin University. Her research interests include embedded systems and biomedical signal processing and recognition.

LINA XU is currently pursuing the Ph.D. degree with the College of Instrument Science and Electrical Engineering, Jilin University. Her research interests include small signal processing and biomedical image signal processing.

CHONGJIN LIU is currently pursuing the bachelor's degree with the Zhuhai College of Jilin University. His research interests include signal processing, image recognition, and machine learning.