Non-local Second-Order Attention Network for Single Image Super Resolution

Jiawen Lyn, Sen Yan

Non-Local Second-Order Attention Network For Single Image Super Resolution

Jiawen Lyn[1][0000−0003−2119−4468] and Sen Yan[0000−0002−6860−3962]

Trinity College Dublin, Dublin 2, D02PN40 Ireland
linj1@tcd.ie

Abstract. Single image super-resolution is a ill-posed problem which aims to characterize the texture pattern given a blurry and low-resolution image sample. Convolution neural network recently are introduced into super resolution to tackle this problem and further bringing forward progress in this field. Although state-of-the-art studies have obtain excellent performance by designing the structure and the way of connection in the convolution neural network, they ignore the use of high-order data to train more power model. In this paper, we propose a non-local second-order attention network for single image super resolution, which make the full use of the training data and further improve performance by non-local second-order attention. This attention scheme does not only provide a guideline to design the network, but also interpretable for super-resolution task. Extensive experiments and analyses have demonstrated our model exceed the state-of-the-arts models with similar parameters.

Keywords: Super resolution · Deep neural network · Deep learning.

1 Introduction

Because the rapid growth of the video and image data, super-resolution (SR) has enjoyed the current advances in deep learning and has attracted more attention in recent years. In real life, super-resolution techniques can be applied to many applications such as satellite and medical image processing [16], facial image improvement [2] and aerial imaging [33], etc. Obtaining high-resolution images given low-resolution images can be an ill-posed problem, but convolution neural network had a huge impact in this field, making the result images detailed and natural. In this work, we mainly tackle single image super-resolution task.

Considering the successful experience of convolutional neural network (CNN) in high-level vision tasks such as images segmentation, Dong et al. [7] proposed CNN based SR algorithm namely SRCNN. So from then on CNN have attract more attention for researchers to tackle super-resolution tasks [8, 18, 19, 22, 25, 30]. Although the performance is largely improved, exist some problems. Firstly, and most importantly, previous researches focus on introducing deeper convolutional neural network to improve performance and ignore the computation overhead. A large number of calculations make the algorithm difficult to be

applied in practice. Secondly, with the depth of network increasing, the training procedure will become more unstable [22, 24], which means that need more training tricks to train network for improving performance. Thirdly, most of the previous methods did not make full use of the training data to reconstruct the super-resolution image. Practically, we have some train data did not use in training.

To tackle the above problems, we propose applying non-local second-order attention mechanism to super-resolution network designs. First, we focus on the non-local block to make the network to learn self-attention by capturing long-range dependency. Second, we develop the network to make full use of the training data for reconstructing super-resolution image. To the best of our knowledge, this is the first proposed non-local second-order training strategy into single image super-resolution network design, providing a special viewpoint of the data enhancement for deep learning and a extraordinary guidance on network design. In this work, we propose a both lightweight and efficient networks using the proposed schemes to design. Experimental results on benchmark datasets demonstrate our methods is superior most of state of the art models.

2 Related work

2.1 Single image super-resolution

Single image super resolution is a low-level computer vision task. The popular method in our literatures is learning the mapping function from low-resolution images to high-resolution images to reconstruct. Traditional machine learning techniques are widely applied in super-resolution, including kernel method[4], PCA [3], sparse-coding [29], learning embedding [5], etc. There are a powerful method take full sue of the image self-similarity without extra data. In order to obtain a super-resolution image, [11] use the patch redundancy to produce. Freedam et al. [9] further develop a localized searching method. [13] extend this algorithm to guide the patch search through using detected perspective geometry.

Current advances in SISR make full use of the powerful representation capability of convolution neural network. Dong et al. [7] first proposed SRCNN to recovery high-resolution image. They interpret the architecture in CNN as extraction layer, non-linear mapping layer, and reconstruction layer, corresponding these steps in sparse coding [29]. DRCN [18] further these steps through firstly interpolating the low-resolution image to the desired size so that suffers from the huge computational complexity and some detail lost. Kim el al. [17, 18] adopt the deep residual convolutional neural network to achieve better performance, which use the bicubic interpolation to upsample the low-resolution image to the desired size and then fed in network to output super-resolution image. Since then, the deeper CNN-based super resolution models is a trend to obtain superior performance, such as LapSRN [19], DRRN [25], SRResNet [20], EDSR [22] and RCAN [30].

Nevertheless, the depth of the network bring a huge amount of the computation and increase the process time. In order to solve this problem, Dong et

al. adopt smaller filter sizes and a deeper network namely FSRCNN [8], which remove the bicubic interpolation layer in SRCNN and embedding the deconvolution layer at the tail of the FSRCNN. To reduce parameters, DRRN [25] proposed the combination of the residual skip connection and the recursive so that compromise the runtime speed. Currently, CARN [1] exploit the multiple shortcut connections and multiple -level representation to obtain a cascading mechanism upon a residual network. In order to utilize the multi-scale feature, [21] proposed MSRN model to capture the multi-scale feature at different scale size. In order to improve the performance, Dai et al. proposed SAN [6] and He et al proposed ODE-inspired network [**?**], which both introduced high-order feature extractor to capture high-order statistic, but they ignored the operate of the convolution layer is local so we combined both the non-local operate and high-order statistic extractor to improve our network.

Although most of the CNN-based super-resolution methods strongly promote progress in this field, most of the advanced model blindly increase the depth and parameters of the network and they ignore the operate of the convolution is local. It is clear that these method increase the running time and not necessarily improve accuracy .

2.2 Attention model

To human perception, attention generally means human visual systems focus on salient areas [15] and adaptively process visual information. Currently, several study have proposed embedding attention mechanism processing to improve the performance of CNNs for different tasks, such as image segmentation, image and video classification [12, 28]. Wang et al. proposed non-local neural network [28] for video classification, which incorporate non-local process to spatially attention long range feature. Hu et al. proposed SENet [12] to capture channel-wise feature relationships to obtain better performance for image classification. Li et al. proposed expectation-maximization attention network for semantic segmentation, which borrowed EM algorithm to iteratively optimize parameters and decrease the complex of the operation in non-local block. Huang et al. proposed criss-cross attention [14] for semantic segmentation, which can efficiently capture contextual pattern from long-range dependencies. Fu et al. proposed a dual attention network (DANet) [10], which mainly consists of the position attention module and the channel attention module. They use position attention module to learn the spatial interdependencies. The channel attention module is designed to model channel interdependencies. It largely improves the segmentation results through capturing rich contextual dependencies. Zhang et al. proposed residual channel attention network (RCAN) [30] for single image super resolution, which adopt channel attention (CA) mechanism to adaptively capture channel-wise pattern information through considering interdependencies among channels. Zhang et al. is first introduce non-local block in single image super resolution. They proposed residual non-local attention learning [31] to capture more detailed information through preserving more low-level features, being more suitable for super resolution image reconstruction. The network pursue better network representational

ability and achieve high-quality image reconstruction results. Dai et al. proposed non-locally enhanced residual group (NLRG) [6] to capture spatial contextual information so that hugely improve the performance of the model.

3 Non-local Second-order Attention Network

3.1 Network framework

As show in Figure 1, our NSAN mainly consists of four parts: shallow feature extractor, high-order enhanced group (HEG) based deep feature extraction, upscale layer and reconstruction layer. Give I_{LR} and I_{SR} as the input and output of our NSAN. Following the [22, 6], we apply one convolution layer to capture the shallow feature F_0 from the LR input

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

where H_{SF} represents the convolution operation. Then the shallow feature F_0 fed in HEG based deep feature extraction, which thus obtains the deep feature as

$$F_{DF} = H_{HEG}(I_0) \tag{2}$$

where H_{HEG} stands for the HEG based non-local enhanced feature extraction module, which consists of two RL-NL modules to capture the long-range information and G residual channel attention groups. So our proposed HEG achieve very deep depth and can capture more information. Then the extracted deep feature F_{DF} is upsample through the upsacale module via

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

where $H_{\uparrow}$ and $F_{\uparrow}$ are a upsample layer and upsampled feature respectively. In the previous works, there are several choices to perform as upscale part, such as transposed convolution [8], ESPCN [23]. Embedding upscaling feature in the last few layers achieve a good trade off between performance and computational burden, thus is preferable in recent SR models [8, 6, 22]. Then upscaled feature is through one convolution layer

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

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

Then NSAN will be optimized with a loss function. Some loss functions have been widely used, such as L2, L1, perceptual losses. In order to verify the effectiveness of our NSAN, we adopt the L1 loss functions followed previous works. Given a training set with N low-resolution images and high-resolution images denoted by $\{I_{HR}, I_{HR}\}^N$, the purpose of the NSAN is to optimize the loss function:

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

where θ represents the parameter set of NSAN. We choose Adam algorithm to optimize the loss function.

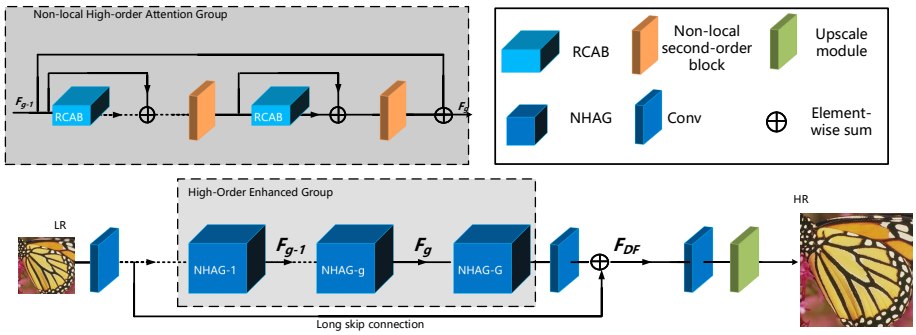

Fig. 1. Framework of the proposed non-local second-order attention network (NSAN).

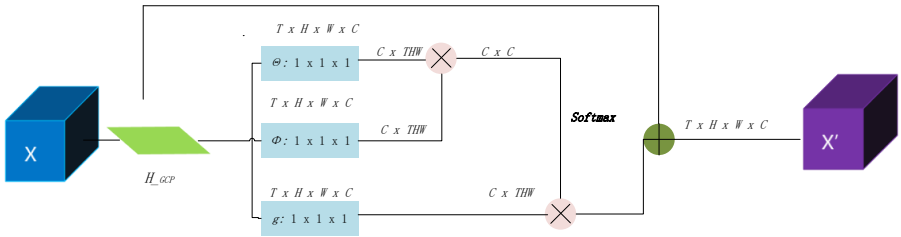

Fig. 2. Framework of the proposed non-local second-order attention module (NSA).

3.2 High-order Enhanced Group (HEG)

We now describe our edge enhanced (HEG) (see Figure. 1), which can be divided into the main branch and the edge enhanced branch. The main branch consists of two region-level non-local (RL-NL) modules [6] and G non-local residual channel attention groups (NRCAG) structure. The RL-NL can capture the long-range information. Each NRCAG further contains M simplified residual channel blocks with local skip connection, followed by a non-local channel attention (NCA) module to exploit feature interdependencies. The edge enhanced branch consists of the padding module and V NRCAG, which can make full use of the edge information and use edge information to enhance channel feature attention.

Stacking residual blocks has been verified that is helpful way to form a deep network in [22, 21, 6]. Nevertheless, deeper network built in such way would lead in performance bottleneck and training difficulty during the problem of gradient vanishing and exploding in deep network. It is known simply stacking repeated block may not to obtain better performance. In order to address this issue, we introduce the NHAG to not only to bypass abundant low-frequency information from LR images, but also facilitate the training of our deep network. Then a HEG in the g-th group is represented as:

$$F_g = H_g(F_{g-1}) \tag{6}$$

where F_g, F_{g-1} denote the output and input of the g-th HEG. The bias term is omitted for simplicity. H_g is the function of the g-th HEG.

Then deep feature is obtained as:

$$F_{DF} = F_0 + F_G \tag{7}$$

3.3 Non-local Second-order Attention

Most previous CNN-based SR models ignore the feature interdependencies. In order to take full use of these information, SENet [12] introduced CNNs to rescale the channel-wise features for image SR. Nevertheless, SENet only exploits first-order statistics features through global average pooling, while ignoring non-local statistics more rich than local, thus hindering the discriminative ability of the network.

Inspired by the above works, we propose a non-local second-order attention (NSA) module (see Figure 2) to capture high-order feature interdependencies through considiering non-local features. Now we will describe how to exploit such non-local information. We reshape the feature map $\mathrm{F} = [\mathrm{f}_1, \cdots, \mathrm{f}_C]$ with C feature maps with size of $H \times W$ to a feature matrix X with $s = WH$ features of C-dimension. Then compute the sample covariance matrix as

$$\Sigma = \mathrm{X}\overline{I}\mathrm{X}^T$$

where $\overline{I} = \frac{1}{s}(\mathrm{I} - \frac{1}{s}1)$, I and 1 are the $s \times s$ identity matrix and manix of all ones, respectively.

It is shown that covariance normalization plays a critical role for more discriminative representations. For this reason, we first perform covariance normalization for the obtained covariance matrix Σ, which is symmenic positive semi-definite and thus has eigenvalue decomposition (EIG) as follows

$$\Sigma = \mathrm{U}\Lambda\mathrm{U}^T$$

where U is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_C)$ is diagonal matrix with eigenvalues in non-increasing order. Then convariance normalization can be converted to the power of eigenvalues:

$$\hat{\mathrm{Y}} = \Sigma^{\alpha} = \mathrm{U}\Lambda^{\alpha}\mathrm{U}^T$$

where α is a positive real number, and $\Lambda^{\alpha} = \mathrm{diag}(\lambda_1^{\alpha}, \cdots, \lambda_C^{\alpha})$. When $\alpha = 1$, there is no normalization; when $\alpha < 1$, it nonlinearly shrinks the eigenvalues larger than 1.0 and streches those less than 1.0. As explored in [**?**], $\alpha = 1/2$ works well for more discriminative representations. Thus, we set $\alpha = 1/2$ in the following.

The normalized covariance manix characterizes the correlations of channel-wise features. We then take such normalized covariance matrix as a channel descriptor by global covariance pooling. As illustrated in Fig. 2, let $\hat{\mathrm{Y}} = [\mathrm{y}_1, \cdots, \mathrm{y}_C]$, the channel-wise statistics $\mathrm{z} \in R^{C\times 1}$ can be obtained by shrinking $\hat{\mathrm{Y}}$. Then the c-th dimension of z is computed as

$$z_c = H_{GCP}(\mathrm{y}_c) = \frac{1}{C}\sum_{i}^{C}\mathrm{y}_c(i)$$

where H_{GCP} denotes the global covariance pooling function. Compared with the commonly used first-order pooling (*e.g.*, global average pooling), our global covariance pooling explores the feature distribution and captures the feature statistics higher than first-order for more discriminative representations.

To fully exploit feature interdependencies from the aggregated information by global covariance pooling, we also introduce non-local block to capture long range pattern. Inspire [10], which proposed non-local channel attention block to extract more useful feature, but it do not use second-order feature. In our network, we combine the second order extractor and non-local attention, which can extract second-order feature and then obtain non-local attention map

$$\mathrm{X}^{'} = H_{NSB}(z)$$

where H_{NSB} denote the non-local second-order block. The detail of the second-order block can see figure 2.

4 Experiments

4.1 Setup

Following [22, 31], we train on 800 training images in DIV2K dataset [27]. In order to verify the effectiveness of our network, we choose 5 benchmark datasets:

Set5, Set14, BSD100, Urban100 and Manga109. For the degradation model, we adopt Matlab resize function with bicubic operation. For the metrics, we use PSNR and SSIM to evaluate SR result.

For the training, the low-resolution images are augmented through horizontally flipping and randomly rotating $90°, 180°, 270°$. For the each min-batch, we set 16 low resolution image patches with size 48×48 as inputs. We use ADAM algorithm to optimize our model with $\beta_1 = 0.9, \beta_2 = 0.99$, and $\epsilon = 10^{-8}$ and initialize learning rate as 10^{-4} and then reduced to half every 200 epochs. We use Pytorch framework to train our proposed NSAN on an Nvidia 1080Ti GPU.

4.2 Ablation Study

As show in Figure 1, our NSAN contains two main components, including High-order Enhanced Group (HEG) and non-local second-order attention module (NSA). In order to test the effectiveness of the various modules, we train and test NSAN with its variants on the Set5 dataset for comparison. Specific performance is shown in Table 1.

We set R_{BASE} as a basic baseline, which only contains the convolutional layer containing 20 NHAG and 10 remaining blocks in each NHAG. Following the [32], we also added long skip and short skip connections to the base model. R_a and R_b mean embedding second-order feature extractor and non-local block in base structure, respectively. R_c means that the result of combining second-order feature extractor and non-local block. It can be found that both of R_c obtain better performance than methods of R_a to R_b.

Table 1. Effect of different module. We report the result in Set5 dataset on 200 epoch

	Base	Ra	Rb	Rc
Second-order Feature		True		
Non-local Block			True	
Non-local Block with Non-local Block				True
PSNR	31.97	32.04	32.08	32.23

4.3 Results with Bicubic Degradation (BI)

We set up a comparison test with model 12 state-of-the-art CNN-based SR methods: SRCNN [7], FSRCNN [8], VDSR [17], LapSRN [19], MemNet [26], EDSR [22], RDN [32] and RCAN [30] for verifying the effectiveness of the NSAN. The quantitative results of each scale factor are shown in Figure 3. Compared to other methods, our NSAN performed best on all datasets, with different scaling factors. Without self-integration, NSAN and SAN can achieve very similar results and are superior to other approaches. This is mainly because both of them use high-order feature to learn the interdependence between features, which makes the network pay more attention to information features.

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
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
NSAN	2	38.43/.9634	34.17/.9233	32.47/.9038	33.12/.9350	39.42/.9893
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
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
NSAN	3	34.85/.9321	30.63/.8576	29.54/.8121	28.99/.8771	34.41/.9497
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
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
NSAN	4	32.67/.90021	28.95/.7894	27.81/.7456	26.87/.8087	31.23/.9188

Fig. 3. Quantitative results with BI degradation model.

	EDSR	MemNet	NLRG	DBPN	RDN	RCAN	NSAN
Para.	43M	677k	330k	10M	22.3M	16M	15.5M
PSNR	38.11	37.78	38.00	38.09	38.24	38.27	38.43

Fig. 4. Computational and parameter comparison (2X) Set5).

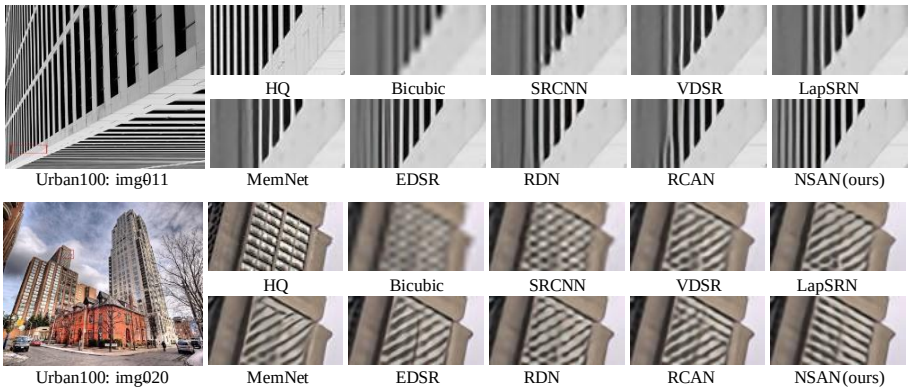

Fig. 5. Visual comparison for 4x SR with BI model on Urban100 dataset.

Compared to RCAN, our NSAN got satisfactory performance for data sets with rich texture information, such as Set5, Set14, and BSD100, and slightly worse results for data sets, such as Manga109 and BSD100 with rich reprocessing edge information. As we all know, texture is a higher-order pattern with more complex statistical properties, while edge is a first-order pattern that can be extracted by a first-order operator. Therefore, our NSA based on second-order feature statistics and non-local operator works better on images with more higher-order information like texture.

We also show the visual results of different methods as shown in Figure 2. We find that most SR models cannot accurately reconstruct the lattices and have severe fuzzy artifacts. On the contrary, our NSAN achieve clearer results and reconstruct more high-frequency details such as high contrast and sharp edges. In the case of "img011", most of the comparison methods output heavily fuzzy artifacts. The early developments of bicubic, SRCNN, FSRCNN and LapSRN even lost their main structures.

Compared with the ground-truth, NSAN gets more reliable results and restores more image detail. Although reconstructing high frequency information is hard during the limited input information of LR, our NSAN can still take full advantage of the limited LR information through second-order non-local attention, while taking advantage of the spatial feature of both high-order characteristics associated with more powerful pattern representation, resulting in more refined results.

4.4 Model Size Analyses

The Figure 4 shows the model size and performance of current CNN SR models. In these methods, MemNet and NLRG contain far fewer parameters for the cost of performance degradation. Not only NSAN had fewer parameters than RDN, RCAN and SAN, but also achieved better performance, which means

NSAN could have a great performance trade-off between model complexity and performance.

5 Conclusions

We propose a deep non-local second-order attention network (NSAN) for SISR. Specifically, the high-order enhanced group allows NSAN to capture the structural information and long-range dependencies through embedding non-local operations. Meanwhile, NHAG allows abundant low-frequency information from the LR images to be bypassed through local skip connections. Not only NSAN exploiting the spatial feature correlations, but also learn high-order feature interdependencies by global covariance pooling for more discriminative representations through second-order non-local attention (NSA) module. Extensive experiment on SR with BI demonstrate the effectiveness of our NSAN in terms of quantitative and visual results.

References

1. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 252–268 (2018)
2. Bo, L., Hong, C., Shan, S., Chen, X.: Low-resolution face recognition via coupled locality preserving mappings. IEEE Signal Processing Letters **17**(1), 20–23 (2009)
3. Capel, D., Zisserman, A.: Super-resolution from multiple views using learnt image models. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 2, pp. II–II. IEEE (2001)
4. Chakrabarti, A., Rajagopalan, A., Chellappa, R.: Super-resolution of face images using kernel pca-based prior. IEEE Transactions on Multimedia **9**(4), 888–892 (2007)
5. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 1, pp. I–I. IEEE (2004)
6. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11065–11074 (2019)
7. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2), 295–307 (2015)
8. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. pp. 391–407. Springer (2016)
9. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. ACM Transactions on Graphics (TOG) **30**(2), 12 (2011)
10. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3146–3154 (2019)

11. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: 2009 IEEE 12th international conference on computer vision. pp. 349–356. IEEE (2009)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
13. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5197–5206 (2015)
14. Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 603–612 (2019)
15. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis & Machine Intelligence (11), 1254–1259 (1998)
16. Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R., Azhari, H.: Super-resolution in pet imaging. IEEE transactions on medical imaging **25**(2), 137–147 (2006)
17. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1646–1654 (2016)
18. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1637–1645 (2016)
19. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 624–632 (2017)
20. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)
21. Li, J., Fang, F., Mei, K., Zhang, G.: Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 517–532 (2018)
22. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
23. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016)
24. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
25. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3147–3155 (2017)
26. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: Proceedings of the IEEE international conference on computer vision. pp. 4539–4547 (2017)
27. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 114–125 (2017)

28. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7794–7803 (2018)
29. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE transactions on image processing **19**(11), 2861–2873 (2010)
30. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301 (2018)
31. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. arXiv preprint arXiv:1903.10082 (2019)
32. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472–2481 (2018)
33. Zhang, Y.: Problems in the fusion of commercial high-resolution satelitte as well as landsat 7 images and initial solutions. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences **34**(4), 587–592 (2002)