MULTI-CONTEXT AND ENHANCED RECONSTRUCTION NETWORK FOR SINGLE IMAGE SUPER RESOLUTION

Jiqing Zhang[1], *Chengjiang Long*[2*], *Yuxin Wang*[1], *Xin Yang*[1,4*], *Haiyang Mei*[1], *Baocai Yin*[1,3]

[1] Dalian University of Technology, [2] Kitware Inc,
[3] Peng Cheng Laboratory, [4] Beijing Technology and Business University

{zhangjiqing, mhy666}@mail.dlut.edu.cn, chengjiang.long@kitware.com, {wyx,xinyang,ybc}@dlut.edu.cn

ABSTRACT

Most existing single image super-resolution (SISR) methods continually increase the depth or width of networks, without adequately exploring contextual features which are essential for reconstruction. Moreover, such existing methods pay little attention to the final high-resolution(HR) image reconstruction step and therefore hinder the desired SR performance. In this paper, we propose a multi-context and enhanced reconstruction network (MCERN) for SISR. Specifically, a novel model named Multi-Context Block (MCB) which extracts more image contextual features with multi-branch dilated convolution. Applying multiple MCBs with residual and dense connections, we can effectively extract contextual and hierarchical features for obtaining the coarse super-resolution result. Then an enhanced reconstruction block (ERB) is followed to extract essential spatial features on the high-resolution image to refine the coarse result to a better result. Extensive benchmark evaluations demonstrate the efficacy of our proposed MCERN in terms of metric accuracy and visual effects.

Index Terms— single image super-resolution, deep learning, multi-context block, enhanced reconstruction block.

1. INTRODUCTION

Single-image super-resolution (SISR) is a computer vision task that reconstructs a high-resolution (HR) image from a low-resolution (LR) image. It could be used in a variety of applications such as medical imaging, security, and surveillance imaging. The quality of the reconstructed HR image depends on how to extract and use the information from LR image. Since there are multiple HR images that can be downsampled to the same LR image and this is a one-to-many mapping relation to recover HR images from a LR image, SISR is an ill-posed and still challenging problem in the community.

Recently, convolutional neural networks (CNNs) [1, 2, 3, 4, 5, 6, 7] have been widely used to handle SISR owing to the powerful learning ability. In spite of remarkable progress achieved in SISR, we observe that existing methods are still faced with two main limitations, first, for feature extraction, most methods blindly increase the depth or width of the network in order to enhance the ability of feature extraction of the network but ignore taking full use of the contextual features from LR image. The contextual feature information gradually disappears with the increase of the network depth; Second, for the HR image reconstruction process, most models only use LR spatial information and directly use an upsampling layer at the end of the network to reconstruct the HR image. These methods only focus on LR spatial information and make the network failed to explore useful cues in HR space for reconstructing visually pleasant SR image.

Such observations inspire us to make full use of contextual features and simultaneously distill features in LR and HR space for SISR. We propose a multi-context and enhanced reconstruction network (MCERN) to fully explore the contextual features in LR space for upsampling and further explore a reconstruction process after upsampling in a coarse-to-fine fashion with two stages for SISR. As illustrated in Figure 1, we design a multi-context block (MCB) and multiple multi-context integration blocks (MCIBs) with a skip connection for upsampling to obtain a coarse HR result at stage 1, and then an enhanced reconstruction block (ERB) is proposed to effectively extract essential features in HR space for reconstruction more details to refine the coarse HR result at stage 2.

Different from the existing methods, our MCB employs multi-branch dilated convolution to increase the receptive field for extracting more abundant contextual features, without introducing additional parameters. Combining multiple MCBs with a residual connection and dense connections, each MCIB is good at integrating rich contextual features from MCBs, and multiple MCIBs are stacked with residual connection to combine hierarchical fusion features for the recovery of missing local details with a convolution layer. Therefore, the combination of an MCB and multiple MCIBs can guarantee rich contextual features extracted in LR space for upsampling to achieve a good initial coarse HR result. Unlike MCB and MCIBs extracting LR features, we propose a simple but effective ERB which can explore local details in

[*] This work was co-supervised by Chengjiang Long and Xin Yang. The corresponding author is Xin Yang.

978-1-7281-1331-9/20/$31.00 ©2020 IEEE

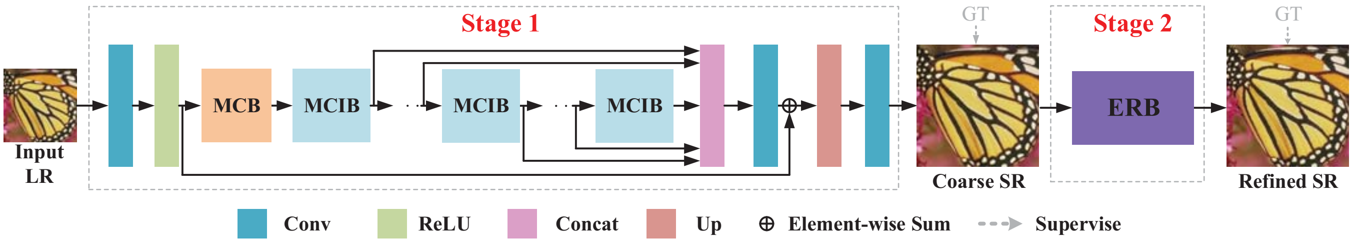

Fig. 1. The overview of our proposed MCERN network. First, a convolutional layer followed by the ReLU activation function is used to extract shallow features from the input RGB image. Second, the shallow features pass through MCB and multiple MCIBs and the output of each MCIB are concatenated to generate multi-level context-aware residual features. Third, the sub-pixel convolutional layer [8] followed by a convolutional layer is used to upsample the LR features to HR space and reconstruct a coarse SR image. Finally, the coarse SR image is fed to the ERB to obtain the final refined SR image.

HR space for reconstructing fine-detailed HR image. The intuition behind the ERB is that the information in LR space is limited and we believe features extracted in HR space could benefit a better recovery of local details. Such a coarse-to-fine network can robustly harvest abundant hierarchical and contextual features in both LR and HR spaces. Also, the well-designed structure and the use of dilation convolution enable our proposed MCERN to be a light-weight network.

To sum up, the main contributions of this paper are: (1) our proposed two-stage MCERN network is able to robustly extract rich hierarchy and contextual features in both LR and HR spaces for recovering a visually pleasing HR image in a coarse-to-fine fashion; (2) an ERB with multiple convolution layers and ReLU activations is proposed to produce a residual map to further refine the initial SR result in HR space, which has not been sufficiently explored before; and (3) the well-designed MCB with multi-branch dilation convolutions and the simple yet effective ERB can guarantee our MCERN as a light-weighted network. We evaluate the proposed MCERN network on four benchmark datasets and the experiments demonstrate the effectiveness of our proposed MCERN in terms of metric accuracy and visual effects.

2. RELATED WORK

In this section, we briefly review recent deep learning methods based on pre-upsampling, post-upsampling, and sampling, which are developed to solve the SR problems.

Pre-upsampling based methods use bicubic interpolation to upsample LR image before the network extracts features, including SRCNN [1], VDSR [2], DRCN [3], DRFN [9] and MemNet [10]. As these methods learn the mapping in HR space, the raw features cannot be extracted from original LR images and the computation complexity of the network grows dramatically with the increase of the specific size of HR images. Moreover, these methods often produce visible reconstruction artifacts due to a lack of information from the LR space.

Post-upsampling based methods like ESPCN [8], EDSR [4], MSRN [11], RCAN [6], RDN [5], and CARN [12], directly extract features from input LR images and then use the features extracted in LR space to obtain HR images by a transposed/sub-pixel convolution layer. Compared with pre-upsampling based methods, the computational complexity of these methods is insensitive to the SR magnification scale. However, most of these methods, such as EDSR [4] and RDN [5], blindly increase the depth or width of the network to enhance the ability of feature extraction of the network and ignore taking full use of the contextual features from LR image. Moreover, these methods only focus on extracting information from LR space and pay no attention to utilizing HR space information for the reconstruction process.

Regarding sampling based methods [13, 14, 15], they adopt sampling methods with some strategies. For example, DBPN [14] exploits iterative up- and down- sampling layers, and provides an error feedback mechanism for projection errors at each stage.

It worths mentioning that our proposed two-stage MCERN network belongs to a post-upsampling method because we use MCB and multiple MCIBs to extract abundant contextual features on the input LR image at the first stage. Note that our ERB is also very similar to a pre-upsampling method because we further feed the coarse HR image to the ERB to recover more local details in the second stage.

3. PROPOSED METHOD

3.1. Network Architecture

As shown in Figure 1, our proposed MCERN consists of two stages to solve the SISR problem in a coarse-to-fine fashion. We design a **multi-context block** (MCB) and **multiple multi-context integration blocks** (MCIBs) to reconstruct a coarse SR result at stage 1. And at stage 2, we propose an **enhance reconstruction block** (ERB) in order to extract essential feature in HR space.

At stage 1, given an input low-resolution image I^{LR}, we first extract shallow features H_0 by

$$H_0 = f_{MCB}(H_{raw}), \quad H_{raw} = \delta(C^{64}_{1\times 1}(I^{LR})), \qquad (1)$$

where $C^{c}_{k\times k}$ represents convolution operation where kernel size is $k \times k$ and the number of output channels is c; δ denotes the rectified linear unit (ReLU) activation function; H_{raw} is the raw features directly extracted from input LR image; and f_{MCB} denotes the further feature extraction function by a MCB. Then, H_0 is fed to the contextual features extraction and fusion component for further feature extraction,

$$H_d = f_d(H_{d-1}) = f_d(f_{d-1}(\cdots f_1(H_0)\cdots)), \qquad (2)$$

where f_d denotes the d-th multi-context integration function and H_d and H_{d-1} are the output and input of the d-th MCIB, respectively. We set $d = 3$, that is, only three MCIBs are used in our actual network. Later, all $H_i, i \in [1, d]$, will be fused by applying convolution layers upon the concatenation of all the previous outputs of each MCIB, *i.e.*,

$$H_{fusion} = C^{64}_{1\times1}([H_1, H_2, \cdots, H_d]), \qquad (3)$$

where $[\cdot]$ denotes the concatenation operation. With the fused feature H_{fusion} which has incorporated hierarchical information, we can get the coarse super-resolution result by applying a convolution on contacting the upsampling upon H_{fusion} and the skip connection from the raw features H_{raw}, *i.e.*,

$$I^{SR}_{coarse} = C^{3}_{1\times1}(f_{up}(H_{fusion} + H_{raw})), \qquad (4)$$

where f_{up} denotes the transposed convolution operation.

At stage 2, we design an ERB (denoted as f_{refine}) to model image details (residuals) to get a better super-resolution result:

$$I^{SR}_{refined} = f_{refine}(I^{SR}_{coarse}), \qquad (5)$$

Finally, the overall loss function $L_{overall}$ is defined as:

$$L_{overall} = w_c L(I^{SR}_{coarse}, I^{HR}) + w_r L(I^{SR}_{refined}, I^{HR}), \quad (6)$$

where I^{HR} is the ground truth image, L is the mean absolute error (MAE) loss, w_c and w_r are the balancing parameters.

3.2. Multi-Context Block

In order to make sure our method can leverage more context to predict image details and simultaneously extract features with different contextual characteristics for reconstructing visually pleasing HR image, we propose a novel structure named Multi-Context Block (MCB).

A common operation to increase the receptive field is to cascade several convolution layers, as shown in Figure 2(a). In the cascading structure, as the depth of the network increases, the receptive field gradually increases. The output of every layer is concatenated to utilize multiple scales of receptive fields. As shown in Figure 2(b), a variety of contextual features can be obtained with parallel structure. In this parallel structure, to sample the input with different contextual information, multiple layers accept the same input and their outputs are fused.

As shown in Figure 2(c), we push the boundaries of cascading and parallel strategies to a novel compact structure to simultaneously distill features with different receptive fields and contextual characteristics. Our proposed MCB contains two branches, which are used to extract different contextual features. Each branch consists of two cascaded convolutions, which are used to extract features with different receptive fields. We adopt the dilated convolution for widening the receptive field without additive parameters, which maintains the lightweight structure. Finally, we concatenate all features of different branches and depths and fuse them via a 1×1 convolution layer.

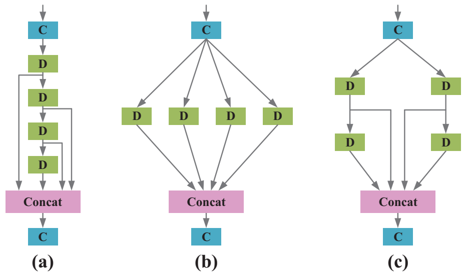

Fig. 2. Comparison of feature extraction structures for SR: (a) cascading structure, (b) parallel structure, and (c) our proposed compact structure. "C" and "D" denote traditional and dilated convolution (both with a ReLU activation function).

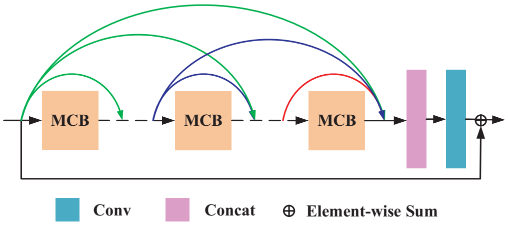

Fig. 3. The structure of MCIB.

3.3. Multi-Context Integration Block

As MCB can effectively extract multiple contextual information from input features, inspired by [5], we further use a multi-context integration block (MCIB) based on MCB to harvest rich contextual features of different levels. Specifically, we stack multiple MCBs in a dense connection manner as shown in Figure 3. By doing so, each MCB in MCIB has access to all the previous MCB's output and thus could fully utilize them to further distill higher level contextual features. We then concatenate the outputs of each MCB and feed them into a 1×1 convolution to distill information that needs to be preserved, from contextual features of different levels. Here, we adopt the residual learning strategy to ease the difficulty of training. MCIB can be expressed as follows,

$$H_d = C^{64}_{1\times1}([H_{d-1}, H_{d,1}, \cdots, H_{d,e}]) + H_{d-1}, \qquad (7)$$

where $H_{d,e}$ denotes the output of *e-th* MCB in *d*-th MCIB. Each MCIB contains six MCBs, *i.e.*, $e = 6$, and the dilation rates of these MCBs are set to 1, 2, 3, 3, 2, and 1, respectively.

As shown in Figure 1, we used three MCIBs and combined the outputs of each MCIB to obtain hierarchical contextual features used for reconstructing the initial HR image.

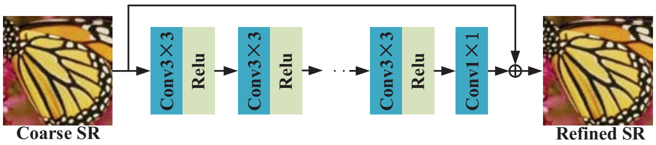

Fig. 4. The structure of our proposed ERB.

3.4. Enhance Reconstruction Block

We propose an enhance reconstruction block (ERB) to reconstruct a refined SR result from a coarse SR result. As shown in Figure 4, our proposed ERB contains six convolution layers with ReLU activation functions and then applies a 1×1 convolution layer to reduce the number of dimensions to three. We use the residual connection to merge the recovered details with the coarse SR result to recover a better SR result. Our network can fully extract image features information in both the LR space and the HR space with this coarse-to-fine fashion.

3.5. Implementation Details

We implement our approach in Pytorch and run experiments with a NVIDIA Titan V GPU. For training, we use 48×48 patches cropped from LR image as input and its corresponding HR patches as ground truth. Following [11, 12, 16, 17], we pre-process all the images by subtracting the mean RGB value of the DIV2K dataset [18] and augment the training data with random horizontal flips and 90° rotations. We train our model with ADAM optimizer [19] by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$. The mini-batch size is set to 16. The learning rate is initialized as 0.0001 and decreases to half every 200 epoch. The number of total epochs is 1000. The balancing parameters w_c and w_r in Equation 6 are set to 1.

4. EXPERIMENT

4.1. Experimental Settings

For a fair comparison, we evaluate all methods with two commonly used metrics, *i.e.*, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [23] on Y channel of transformed YCbCr space. Following [11, 12, 16, 17], we use the DIV2K [18] dataset for training and four datasets - Set5 [24], Set14 [20], BSDS100 [21], and Urban100 [22] for evaluation.

Table 1. MCB architecture analysis. "Cascading" and "Parallel" denote the cascading and parallel architecture shown in Figure 2(a) and Figure 2(b), respectively; And MCB is our proposed multi-branch architecture shown in Figure 2(c).

Archi–tecture	Set14 [20]		BSDS100 [21]		Urban100 [22]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Cascading	33.67	0.9184	32.18	0.8987	32.37	0.9305
Parallel	33.68	0.9184	32.20	0.8993	32.32	0.9304
MCB(s=1)	**33.77**	**0.9194**	**32.25**	**0.9006**	**32.47**	**0.9312**
MCB(s=x)	**33.83**	**0.9196**	**32.27**	**0.9014**	**32.67**	**0.9336**

Table 2. The effectiveness of ERB.For a fair comparison, we move the ERB to the front of the upsampling to ensure that the depth of MCERN w/o and w/ ERB are the same.

MCERN	Set5 [24]		Set14 [20]		Urban100 [22]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
w/o ERB	38.10	0.9608	33.74	0.9191	32.33	0.9304
w/ ERB	**38.20**	**0.9612**	**33.83**	**0.9196**	**32.67**	**0.9336**

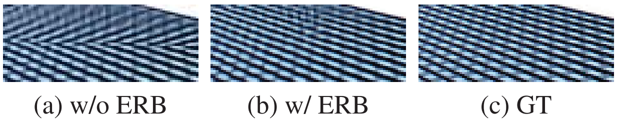
(a) w/o ERB (b) w/ ERB (c) GT

Fig. 5. Qualitative comparisons of the effectiveness of ERB.

4.2. Effectiveness of MCB

In order to verify the effectiveness of our proposed MCB structure, first, we replace our MCB structure with the cascading structure (see Figure 2 (a)) or the parallel structure (see Figure 2 (b)). As shown in Table 1, we set $d = 3$, $e = 6$, and $s = 1$, the results suggest the multiple branches structure in the MCB is more efficient than cascading the structure and parallel structure structures under the same parameters. Then, to demonstrate the effectiveness of the dilated convolution with the dilation rate described in Section 3.3 (*i.e.*, the dilation rates of these six MCBs are set to 1, 2, 3, 3, 2, and 1, respectively), we compare dilation rate s are various ($s = x$) and all are 1 ($s = 1$). Similarly, we set $d = 3$ and $e = 6$. The results are summarized in Table 1.

The above two experiments show that MCB is an effective structure, which can extract different contextual features through two branches, and extract features in different receptive fields through two cascaded convolutional layers in each branch. And combining multiple MCBs into one MCIB, the MCIB can adaptively integrate different contextual features from the MCB.

4.3. Effectiveness of ERB

To prove that the HR space information can improve the reconstruction results and the effectiveness of ERB, we move the ERB to the front of the upsampling to ensure that the depth of MCERN w/o and w/ ERB are the same. The results are summarized in Table 2, from which we can clearly observe that the performance with EBR works better than that without ERB. This observation suggests that our ERB is able to refine a coarse SR result to a more detailed one since it can continue to extract useful features from HR space. We also show the visual comparison in Figure 5. As we can, our ERB is able to correct the direction for black lines.

Table 3. Performance comparison to seven current state-of-the-art proposed methods with light-weighted model ($<$ 6M parameters) in terms of PSNR and SSIM on four benchmarks with scale factors of ×2,×3,×4. We denote the best performance and the second-best performance in red and blue, respectively.

Datasets	Scale	bicubic interpolation	VDSR [2] (CVPR'16)	LapSRN [13] (CVPR'17)	CARN [12] (CVPR'18)	SRMDNF [16] (CVPR'18)	NLRN [25] (NIPS'18)	MSRN [11] (ECCV'18)	FRSR [17] (CVPR'19)	Ours MCERN
Set5 [24]	×2	36.66/0.9542	37.53/0.9583	37.52/0.9591	37.76/0.9590	37.79/0.9601	38.00/0.9603	38.08/0.9605	37.95/0.9594	38.20/0.9612
	×3	30.39/0.8682	33.68/0.9201	33.82/0.9227	34.29/0.9255	34.12/0.9254	34.27/0.9266	34.38/0.9262	34.38/0.9262	34.52/0.9282
	×4	28.42/0.8104	31.36/0.8796	31.54/0.8850	31.92/0.8903	31.96/0.8925	31.92/0.8916	32.07/0.8903	32.22/0.8950	32.24/0.8965
Set14 [20]	×2	30.24/0.8688	33.05/0.9107	33.08/0.9130	33.52/0.9166	33.32/0.9159	33.46/0.9195	33.74/0.9170	33.45/0.9195	33.83/0.9196
	×3	27.55/0.7742	29.86/0.8312	29.87/0.8320	30.29/0.8407	30.04/0.8382	30.16/0.8374	30.34/0.8395	30.27/0.8411	30.42/0.8441
	×4	26.00/0.7027	28.11/0.7624	28.19/0.7720	28.42/0.7762	28.35/0.7787	28.36/0.7745	28.60/0.7751	28.64/0.7830	28.68/0.7844
BSDS100 [21]	×2	29.56/0.8431	31.92/0.8965	31.80/0.8950	32.09/0.8978	32.05/0.8985	32.19/0.8992	32.23/0.9013	32.17/0.8991	32.27/0.9014
	×3	27.21/0.7385	28.83/0.7966	28.82/0.7980	29.06/0.8034	28.97/0.8025	29.06/0.8026	29.08/0.8041	29.11/0.8050	29.17/0.8071
	×4	25.96/0.6675	27.29/0.7167	27.32/0.7270	27.44/0.7304	27.49/0.7337	27.48/0.7306	27.52/0.7273	27.60/0.7370	27.64/0.7382
Urban100 [22]	×2	26.88/0.8403	30.79/0.9157	30.41/0.9101	31.51/0.9312	31.33/0.9204	31.82/0.9249	32.22/0.9326	32.23/0.9290	32.67/0.9336
	×3	24.46/0.7349	27.15/0.8315	27.07/0.8280	27.38/0.8404	27.57/0.8398	27.93/0.8453	28.08/0.8554	28.33/0.8556	28.46/0.8589
	×4	23.14/0.6577	25.18/0.7543	25.21/0.7560	25.63/0.7688	25.68/0.7731	25.79/0.7729	26.04/0.7896	26.21/0.7910	26.32/0.7934

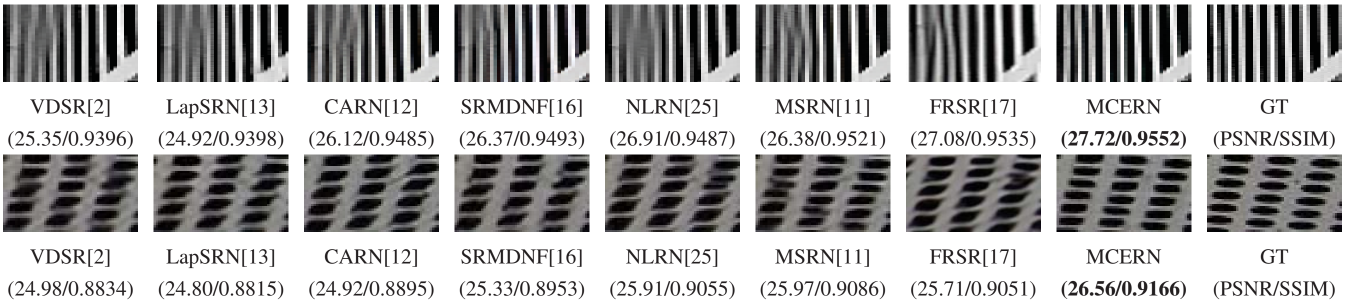

Fig. 6. Visual comparison between different algorithms on different datasets with different scale factors ×2 and ×3.

Table 4. Comparison with deep CNN-based state-of-the-arts.$1M = 10^6$, and $1G = 10^9$.

Methods	Param M(ratio)	Multi-Adds G(ratio)	Set5 [24] PSNR/SSIM
EDSR [4](CVPRW'17)	40.7(10.7)	93.8(9.3)	38.11/0.9602
RDN [5](CVPR'18)	22.1(2.8)	51.0(5.1)	38.24/0.9614
RCAN [6](ECCV'18)	15.4(4.1)	35.3(3.5)	38.27/0.9614
SAN [26](CVPR'19)	15.7(4.1)	36.0(3.6)	38.31/0.9620
DBPN [14](CVPR'18)	10.0(2.6)	34.7(3.4)	38.09/0.9600
RNAN [27](ICLR'19)	8.3(2.2)	16.6(1.6)	38.17/0.9611
MCERN	3.8(1.0)	10.1(1.0)	38.20/0.9612

4.4. Comparisons with State-of-the-art Methods

To confirm the ability of the proposed network, we compare our proposed MCERN model with 7 current state-of-the-art light-weighted methods (with $<$ 6M parameters): VDSR [2], LapSRN [13], CARN [12], SRMDNF [16], NLRN [25], MSRN [11], and FRSR [17].

We show the quantitative results in Table 3. Our proposed MCERN model outperforms the existing methods by a large margin on different datasets and upsampling scales. We also visualize two examples with different scales in Figure 6. Qualitatively, MCERN is able to generate a more visually pleasant image with clean details and sharp edges, while the SR images generated by other methods exhibit visible artifacts. It shows that our coarse-to-fine framework can fully extract rich contextual features in both LR and HR space,

We further compare our method with six state-of-the-art methods with large parameters or heavy complicated calculations in Table 4. It can be seen that our proposed MCERN achieves comparable performance more efficiently (*e.g.* $\geq$ 3 times faster than EDSR [4] and RDN [5]) with a much lighter network. For example, the number of parameters and Multi-Adds of EDSR [4] are 10.7 and 9.3 times than ours, but MCERN obtains 38.20 dB which is 0.09 dB better than EDSR [4]. The results show that our network can effectively extract contextual information, even though it is lightweight.

5. CONCLUSION

In this paper, we propose a novel and light-weighted MCERN network for SISR in a coarse-to-fine fashion to utilize the contextual information and focus on the reconstruction process after upsampling. Our well-designed MCB is good at increasing the receptive field and extracting rich contextual features. We combine multiple MCBs with residual connections and dense connections to form MCIBs for further extracting hierarchical and contextual features before upsampling to obtain the coarse result. Also, an ERB is proposed to focus on extracting essential HR space features after upsampling to refine the coarse result. Extensive evaluations on the benchmark datasets have demonstrated the efficacy of our proposed MCERN in terms of metric accuracy and visual effects. Our future work includes extending it for video SR, and applying it to solve multiple vision applications [28, 29].

ACKNOWLEDGEMENTS. This work was supported in part by the National Natural Science Foundation of China under Grant 91748104, Grant 61972067, Grant 61632006, Grant U1811463, Grant U1908214, Grant 61751203, Grant 61902053, in part by the National Key Research and Development Program of China under Grant 2018AAA0102003, Grant 2018YFC0910506, in part by the Open Research Fund of Beijing Key Laboratory of Big Data Technology for Food Safety (Project No. BTBD-2018KF).

6. REFERENCES

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Image super-resolution using deep convolutional networks,” *TPAMI*, 2016.

[2] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in *CVPR*, 2016.

[3] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in *CVPR*, 2016.

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *CVPRW*, 2017.

[5] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, “Residual dense network for image super-resolution,” in *CVPR*, 2018.

[6] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, “Image super-resolution using very deep residual channel attention networks,” in *ECCV*, 2018.

[7] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang, “Toward real-world single image super-resolution: A new benchmark and a new model,” in *ICCV*, 2019.

[8] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *CVPR*, 2016.

[9] Xin Yang, Haiyang Mei, Jiqing Zhang, Ke Xu, Baocai Yin, Qiang Zhang, and Xiaopeng Wei, “Drfn: Deep recurrent fusion network for single-image super-resolution with large factors,” *IEEE TMM*, 2018.

[10] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu, “Memnet: A persistent memory network for image restoration,” in *ICCV*, 2017.

[11] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang, “Multi-scale residual network for image super-resolution,” in *ECCV*, 2018.

[12] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in *ECCV*, 2018.

[13] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *CVPR*, 2017.

[14] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita, “Deep back-projection networks for super-resolution,” in *CVPR*, 2018.

[15] Yajun Qiu, Ruxin Wang, Dapeng Tao, and Jun Cheng, “Embedded block residual network: A recursive restoration model for single-image super-resolution,” in *ICCV*, 2019.

[16] Kai Zhang, Wangmeng Zuo, and Lei Zhang, “Learning a single convolutional super-resolution network for multiple degradations,” in *CVPR*, 2018.

[17] Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho, “Natural and realistic single image super-resolution with explicit natural manifold discrimination,” in *CVPR*, 2019.

[18] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang, “Ntire 2017 challenge on single image super-resolution: Methods and results,” in *CVPRW*, 2017.

[19] Diederik P. Kingma and Jimmy Lei Ba, “Adam: A method for stochastic optimization,” in *ICLR*, 2015.

[20] Roman Zeyde, Michael Elad, and Matan Protter, “On single image scale-up using sparse-representations,” in *ICCS*, 2010.

[21] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik, “Contour detection and hierarchical image segmentation,” *TPAMI*, 2011.

[22] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja, “Single image super-resolution from transformed self-exemplars,” in *CVPR*, 2015.

[23] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al., “Image quality assessment: from error visibility to structural similarity,” *TIP*, 2004.

[24] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” in *BMVC*, 2012.

[25] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang, “Non-local recurrent network for image restoration,” in *NIPS*, 2018.

[26] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang, “Second-order attention network for single image super-resolution,” in *CVPR*, 2019.

[27] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu, “Residual non-local attention networks for image restoration,” in *ICLR*, 2019.

[28] Chengjiang Long and Gang Hua, “Correlational gaussian processes for cross-domain visual recognition,” in *CVPR*, 2017.

[29] Gang Hua, Chengjiang Long, Ming Yang, and Yan Gao, “Collaborative active visual recognition from crowds: A distributed ensemble approach,” *TPAMI*, 2018.