Context-Aware Residual Network with Promotion Gates for Single Image Super-Resolution

Xiaozhong Ji[1], Yirui Wu[2], and Tong Lu[1](✉)

[1] National Key Lab for Novel Software Technology, Nanjing University,
Nanjing, China
shawn_ji@163.com, lutong@nju.edu.cn

[2] College of Computer and Information, Hohai University, Nanjing, China
wuyirui@hhu.edu.cn

Abstract. Deep learning models have achieved significant success in quantities of vision-based applications. However, directly applying deep structures to perform single image super-resolution (SISR) results in poor visual effects such as blurry patches and loss in details, which are caused by the fact that low-frequency information is treated equally and ambiguously across different patches and channels. To ease this problem, we propose a novel context-aware deep residual network with promotion gates, named as G-CASR network, for SISR. In the proposed G-CASR network, a sequence of G-CASR modules is cascaded to transform low-resolution features to high informative features. In each G-CASR module, we also design a dual-attention residual block (DRB) to capture abundant and variant context information by dually connecting spatial and channel attention scheme. To improve the informative ability of extracted context information, a promotion gate (PG) is further applied to analyze inherent characteristics of input data at each module, thus offering insight for how to enhance contributive information and suppress useless information. Experiments on five public datasets consisting of Set5, Set14, B100, Urban100 and Manga109 show that the proposed G-CASR has achieved averagely 1.112/0.0255 improvement for PSNR/SSIM measurements comparing with the recent methods including SRCNN, VDSR, lapSRN and EDSR. Simultaneously, the proposed G-CASR requires only about 25% memory cost comparing with EDSR.

Keywords: Context-aware residual network · Channel and spatial attention scheme · Promotion gate · Single image super-resolution

1 Introduction

Recently, numerous deep learning methods have been proposed to reconstruct high-resolution images based on single low-resolution images in multimedia. However, these methods still suffer from drawbacks in visual effects. We show

Y. M. Ro et al. (Eds.): MMM 2020, LNCS 11962, pp. 136–147, 2020.
https://doi.org/10.1007/978-3-030-37734-2_12

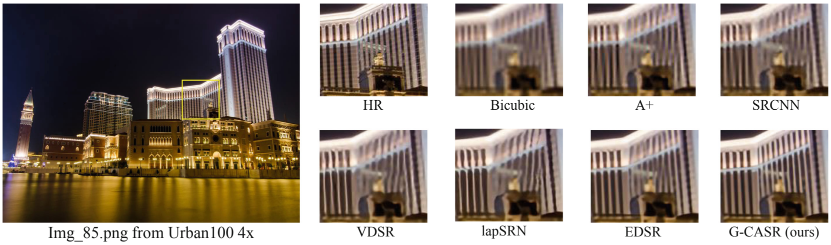

Fig. 1. Comparisons on SISR results achieved by G-CASR and comparative methods, where HR refers to the high-resolution image of the yellow rectangle region. (Color figure online)

examples of reconstructing a high-resolution image as in Fig. 1 (Left) to show these unpleasant effects. Based on the comparisons between the ground truth (HR in Fig. 1) and the generated high-resolution images of different methods in Fig. 1, we can observe blurry patches, failures in reconstructing high-frequency image details, and loss of low-frequency features like straight lines for the existing methods consisting of Bicubic, A+, SRCNN, VDSR, lapSRN and EDSR. The reason for unpleasant visual effects lies in the fact that the existing methods lack context information to capture the unique characteristics of low-resolution images. Essentially, the lack of context descriptor is one of the main drawbacks in most deep residual networks.

Based on these limitations of existing methods for single image super-resolution (SISR), we propose a novel Gated Context-Aware Super-Resolution network, which is named as G-CASR. The proposed G-CASR network consists of two main parts, namely, a dual-attention residual block (DRB) and a promotion gate (PG). By modeling channel-wise and spatial attention information to describe the inherent property of context information, the proposed DRB can restore high-frequency features and maintain low-frequency features simultaneously. On the other side, the proposed PG is used to enhance informativeness of context information with an adaptive gating signal.

By involving these two parts, we conduct a light-scale deep residual network to capture the unique and informative context characteristics. The proposed G-CASR network learns an end-to-end mapping between low-resolution image and reconstructed high-resolution image. As shown in G-CASR in Fig. 1, we can see the result of the same input is greatly improved by G-CASR.

The contributions of this paper are three-fold:

- We propose a novel and context-aware residual network G-CASR for SISR, in which dual-attention residual structure and promotion gate mechanism are proposed to enhance feature representative ability based on multi-level features and context information.
- We design a new dual-attention residual block (DRB) by involving channel and spatial attention scheme to modeling context information. Furthermore, we are the first to propose promotion gate (PG) for attention-based residual

networks, which effectively enhance high contribution features and meanwhile suppress redundant ones.
- Experiments on five benchmark datasets show that the proposed G-CASR achieves averagely 1.112/0.0255 improvement for PSNR/SSIM measurements compared with recent methods. Additionally, our model requires only about 25% memory cost.

2 Related Work

We category the existing deep learning methods for SISR into two types, i.e., convolutional neural networks (CNN) and generative adversarial networks. CNN-based SISR methods are quite larger in the amount due to more years of development and their impressive high-resolution reconstruction results. The first work to solve SISR problem, i.e., SRCNN, is introduced by Dong et al. [3]. Their proposed three-layer CNN network directly learns an end-to-end mapping between interpolated low-resolution image and the corresponding high-resolution output image. Inspired by the success of very deep networks like Res-Net, Kim et al. [8] propose very deep convolutional networks (VDSR), in which global residual learning is utilized to recover high-frequency details. Moreover, VDSR stacks 20 convolutional layers to construct a very deep network for accurate SISR and thus has an impressive property of fast convergence.

To pursue a deeper network for SISR task, Tong et al. [15] present a novel SISR method by introducing dense skip connections in a very deep network. By propagating feature maps of each layer into all the subsequent layers and allowing dense skip connection, their model combines low-level and high-level features in a reasonable way to boost reconstruction performance. Lim et al. [11] develop an enhanced deep super-resolution network (EDSR) with its performance exceeding the current state-of-the-art SISR methods. Their method performs optimization by removing unnecessary modules in convolutional residual networks and expanding model depth with a stable training procedure.

Most recently, Kim et al. [9] propose a novel channel-wise and spatial attention mechanism specially optimized for super-resolution, which prefers to fuse spatial and channel attention for a unity representation before assigning weights, rather than two separate weight schemes. However, their work is only tested on two simplified attention schemes. Woo et al. [16] construct convolutional block attention module (CBAM) as a lightweight and general attention module, which sequentially infers attention maps along spatial and channel dimensions at first and then multiply attention maps to the input feature map for adaptive feature refinement. Their proposed light-scale attention module has achieved excellent performance in lots of recognition and classification tasks.

3 The Proposed Method

In this section, we describe the network architecture of G-CASR, and the structures of the proposed DRB and PG.

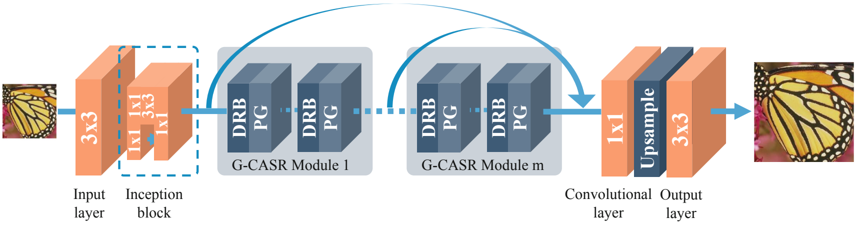

Fig. 2. The framework of proposed G-CASR method, which consists of an input layer, an inception block, G-CASR modules, a convolutional layer, an upsampling layer, and an output layer.

3.1 Network Architecture Design

As shown in Fig. 2, the proposed G-CASR network mainly consists of six parts, that is, an input layer, an inception block, G-CASR modules, a convolutional layer, an upsampling layer, and an output layer. The input low-resolution image I_L is firstly processed by a convolutional kernel of the input layer to generate shallow feature and then enhanced by an inception block, which can be formulated as

$$F_I = H_I(H_S(I_L)) \tag{1}$$

where function $H_S()$ denotes convolutional operation in the input layer, and $H_I()$ refers to multi-branch operations of inception block.

After that, the first G-CASR module is adopted to generate deep feature F_{G_1} based on enhanced feature F_I:

$$F_{G_1} = H_{G_1}(F_I) \tag{2}$$

where function $H_{G_1}()$ denotes the operation of the first G-CASR module. Take G-CASR as the basic module of the whole network, we thus construct deeper network by cascading a quantity of DRBs and PGs. The generated feature after processing of the mth G-CASR can be represented as

$$F_{G_m} = H_{G_m}(F_{G_{m-1}}) \tag{3}$$

To increase the width of the network and generate global features, the convolutional layer accepts input from different modules. G-CASR network obtains a high-resolution image I_H by firstly performing the upsampling layer and then generating the image after operation of the output layer:

$$I_H = H_O(H_U(H_C([F_I, H_{G_1}(F_I), \cdots, H_{G_m}(H_{G_{m-1}}(\cdots H_{G_1}(F_I)\cdots))]))) \tag{4}$$

where function $H_C()$, $H_U()$ and $H_O()$ represent the convolutional layer, upsampling, and the output layer, respectively, while [,] denotes concatenation along channel dimension.

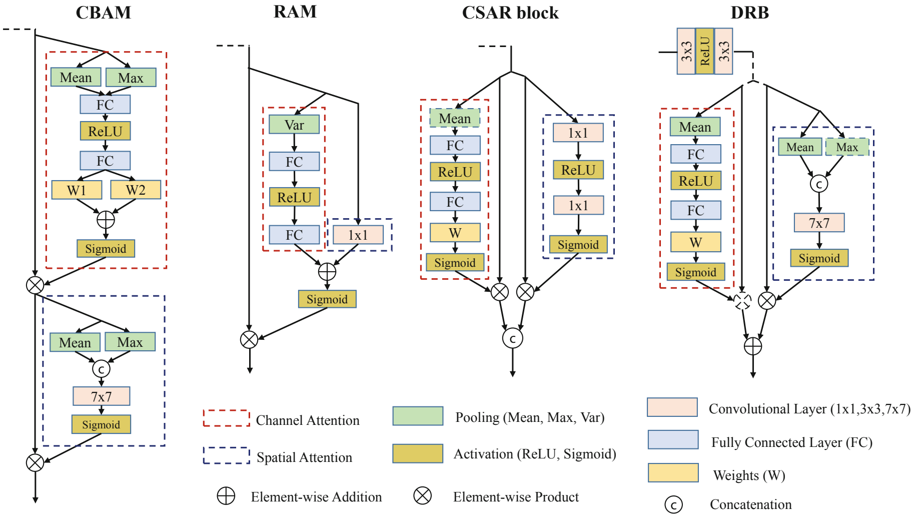

Fig. 3. Structure of the proposed DRB and three other different attention schemes, where the rightmost structure is the proposed DRB.

3.2 Structure of Dual-Attention Residual Block

Inspired by attention schemes applied in other domains and related SISR work based on attention scheme, we propose a lightweight DRB structure, which combines channel-wise and spatial attention in a dual form to adaptively modulate feature representations with context information among feature channels and different regions.

We show structures of various attention schemes including CBAM [16], RAM [9], CSAR block [6] and the proposed DRB in Fig. 3. We can observe DRB is different from other schemes in structure design and combination form. Applying an additional max-pooling to construct spatial attention scheme is adopted by CBAM and DRB, which exploits maximal context characteristics of the feature map to enhance its representative ability. Essentially, regions with the maximal values can be edges, corners or places with high gradient values, which are more salient than other regions and require more attention to their high-frequency details reconstruction. Meanwhile, we use mean-pooling operation in the construction of channel-wise attention scheme, due to the fact that maximal information is weak to exploit the inter-channel relationship.

We prefer dual form to combine the channel and spatial attention scheme rather than a cascade form. This is because usually for recognition and classification tasks, feature information needs to be highly compressed to resolve high-level and semantic information; however, a SISR task requires to restore high-frequency details based on generated feature maps. Cascade combination form often leads passed-by information to be compressed, while dual form can increase bandwidth for information transmission to obtain abundant information for high-frequency detail reconstruction.

Take the first G-CASR module H_{G_1} as an example and assume only one DRB and PG inside, we firstly adopt two convolutional layers and an activation layer to extract feature $\tilde{F_I} = H_E(F_I)$. Then, we construct the dual form of attention scheme to extract informative part of the generated feature, which can be represented as

$$F_D = C(\tilde{F_I}) \odot \tilde{F_I} + S(\tilde{F_I}) \odot \tilde{F_I} \tag{5}$$

where $\odot$ denotes element-wise multiplication, $C()$ and $S()$ represent channel and spatial attention scheme, respectively.

Channel Attention Scheme. Considering that a convolutional layer consists of different channel filters, each 2D slice of the output 3D feature map essentially encodes spatial-visual responses raised by a channel filter. By stacking different layers, CNN extracts image features through a hierarchical representation of visual abstractions [17]. Therefore, features extracted from CNN structure are essentially channel-wise and multi-layer. However, not all the channel-wise features are equally important and informative for recovering high-frequency details. We thus utilize channel attention scheme to compute task-specified feature map for SISR by exploiting the cross-channel relationship.

As shown in Fig. 3, a global mean-pooling is firstly performed on input feature map $\tilde{F_I}$ to output global mean-pooled feature map F_c with size $C \times 1 \times 1$. Then, F_c will be fed into a multi-layer perception with two hidden layers. It is noted that the first hidden layer is used to perform dimension reduction for compact feature representation. Finally, a sigmoid activation function is applied to squeeze the output, thus generating channel attention weight as follows:

$$C(\tilde{F_I}) = sig(W_1 * (relu(W_0 * P_a(\tilde{F_I})))) \tag{6}$$

where function $P_a()$, $sig()$ and $relu()$ refer to the global mean-pooling operation, Sigmoid and ReLU activation function respectively, W_0 and W_1 are learnable parameter matrices and defined with size $\frac{C}{r} \times C$ and $C \times \frac{C}{r}$ respectively, and r is a pre-defined dimension reduction parameter and we set it as 16 by experiments.

Spatial Attention Scheme. We observe the information contained in feature maps and low-resolution images is diverse over spatial positions. For example, edge or texture regions usually contain high-frequency information, while smooth areas have low-frequency information. To better recover high-frequency details and maintain low-frequency parts for a SISR task, we thus propose a spatial attention scheme to adaptively optimize feature map in different regions with suitable operations. Spatial attention scheme is constructed based on the difference of feature map of different positions, which essentially explores the spatial relationship to construct context descriptor.

As shown in Fig. 3, a global max-pooling operation is first performed on input feature map $\tilde{F_I}$ to output max-pooled feature map F_m with size $1 \times m \times n$. Then, we perform mean-pooling operation along channel dimension to generate

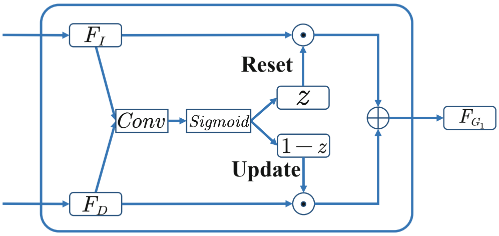

Fig. 4. Architecture of the proposed PG for residual network.

mean-pooled feature map F_a. Finally, a convolutional layer and a sigmoid activation function are performed on the concatenated feature map of F_a and F_m to generate spatial attention weight:

$$S(\tilde{F_I}) = sig(Conv([P_m(\tilde{F_I}), P_{ac}(\tilde{F_I})])) \tag{7}$$

where function $P_m()$, $P_{ac}()$ and $Conv()$ refer to the global max-pooling operation, mean-pooling operation along channel dimension and convolutional layer with 7×7 kernel, respectively.

3.3 Promotion Gate for Residual Network

During modeling a SISR task, missing pixels of high-resolution images can be generated from clues by analyzing low-frequency information from the input low-resolution images. Such highly non-linear processing can be properly achieved by constructing deep neural network to learn from a large training set. However, gradient disappearance from layer to layer leads to shallow structure, thus preventing to obtain deep CNN-based structure.

Inspired by GRU [2] and LSTM [5] for time-varying signal processing, we design PG to work on residual network for deeper layers, where we show its architecture in Fig. 4. Essentially, a GRU or LSTM-based network can build long-term dependencies based on complicated and time-varying information due to their unique gate design, which allows to efficiently update memory with the useful part of a signal. This inspires us to borrow the most important concept of GRU, i.e., gating mechanism, to help build a deeper residual network by enhancing informative part of features, thus relieving the burden of gradient disappearance.

As shown in Fig. 4, the feature F_D is computed by the proposed DRB block for reconstruction, and also the input of the proposed PG. By comparing between the original signal F_I and F_D, the proposed PG decides the proportion of enhancing and forgetting information with a simple but effective gating signal z, which is constructed as a lightweight structure of a convolutional layer with 1×1 kernel and a sigmoid activation function:

$$z = sig(Conv([F_I, F_D])) \tag{8}$$

Table 1. Comparisons on PSNR/SSIM measurement with or without DRB and PG. It is noted that the scale factor is 2×, $\surd$ and × represent network design with or without structure, respectively.

DRB	PG	Set5	Set14	B100	Urban100	Manga109
×	×	37.97/0.9604	33.54/0.9169	32.17/0.8996	31.99/0.9270	38.40/0.9767
√	×	37.98/0.9605	33.49/0.9163	32.15/0.8994	32.06/0.9275	38.61/0.9769
×	√	38.03/0.9606	33.62/0.9179	32.20/0.8999	32.10/0.9283	38.52/0.9769
√	√	38.01/0.9606	33.68/0.9186	32.19/0.9000	32.19/0.9288	38.70/0.9772

Since z is of the same size as F_I and F_D, it can be directly applied as weight to process both features:

$$F_{G_1} = z \odot F_I + (1 - z) \odot F_D \tag{9}$$

Essentially, z resets F_I by assisting to forget useless information, meanwhile $1 - z$ updates F_D by selectively enhancing valuable information. Since z and $1-z$ change synchronously, the PG allows the residual network to keep balance in remembering and forgetting, thus relieving the burden of gradient disappearance.

4 Experimental Results

In this section, we firstly introduce datasets. Then, we conduct four groups of ablation studies to demonstrate the proposed DRB and PG are effective for SISR task. After that, we show performance of our final model on five benchmark datasets. Finally, we describe implementation details for readers' convenience.

4.1 Datasets and Metrics

We conduct experiments on five datasets, i.e., Set5 [1], Set14 [18], B100 [12], Urban100 [7] and Manga109 [4]. Note that Set5, Set14 and B100 consist of natural scenes, Urban100 contains challenging urban scenes images with details, and Manga109 is a dataset of Japanese cartoon drawing. Besides these benchmark datasets, DIV2K [13], which served as the benchmark for NTIRE 2017 challenge, is adopted as a part of the training set. We achieve pairs of low-resolution and high-resolution images by a bicubic operator on high-resolution images. Above all, we obtain 800 images for training and 100 images to perform cross-validation for evaluating SISR methods. Peak signal to noise ratio (PSNR) and structural similarity index (SSIM) are used to measure reconstruction performances for SISR.

4.2 Ablation Study

To verify the effect of the proposed DRB and PG, we conduct four ablation experiments with different network design. To be clear, we define the number of

Table 2. Quantitative evaluation of state-of-the-art SISR algorithms, where average PSNR/SSIM for scale factors 2×, 3×, 4× are listed. Best results are **highlighted**.

Methods	Scale	Set5	Set14	B100	Urban100	Manga109
Bicubic	2×	33.66/0.9299	30.24/0.87688	29.56/0.8431	26.88/0.8403	30.80/0.9339
A+ [14]	2×	36.54/0.9544	32.28/0.9056	31.21/0.8863	29.20/0.8938	35.57/0.9663
SRCNN [3]	2×	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
VDSR [8]	2×	37.53/0.9590	33.05/0.9130	31.90/0.8960	30.77/0.9140	37.22/0.9750
LapSRN [10]	2×	37.52/0.9591	33.08/0.9130	31.80/0.8950	30.41/0.9101	37.27/0.9740
EDSR [11]	2×	38.11/0.9601	33.92/0.9195	32.32/0.9013	**32.93/0.9351**	39.10/0.9773
G-CASR	2×	**38.22/0.9614**	**33.94/0.9214**	**32.33/0.9015**	32.79/0.9344	**39.24/0.9782**
Bicubic	3×	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
A+ [14]	3×	32.58/0.9088	29.13/0.8188	28.29/0.7835	26.03/0.7973	29.93/0.9089
SRCNN [3]	3×	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
VDSR [8]	3×	33.67/0.9210	29.78/0.8320	28.83/0.7990	27.14/0.8290	32.01/0.9340
LapSRN [10]	3×	33.82/0.9227	29.87/0.8320	28.82/0.7980	27.07/0.8280	32.21/0.9350
EDSR [11]	3×	34.65/0.9282	30.52/0.8462	29.25/0.8093	**28.80/0.8653**	34.17/0.9476
G-CASR	3×	**34.66/0.9294**	**30.55/0.8464**	**29.26/0.8094**	28.76/0.8637	**34.18/0.9480**
Bicubic	4×	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
A+ [14]	4×	30.28/0.8603	27.32/0.7491	26.82/0.7087	24.32/0.7183	27.03/0.8439
SRCNN [3]	4×	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
VDSR [8]	4×	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
LapSRN [10]	4×	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7560	29.09/0.8900
EDSR [11]	4×	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	31.02/0.9148
G-CASR	4×	**32.54/0.8996**	**28.88/0.7882**	**27.72/0.7424**	**26.69/0.8038**	**31.14/0.9163**

G-CASR modules as m, the number of DRB with PG in each G-CASR module as n and filter number of each convolutional layer as k. The setting for ablation experiments is $m = 4$, $n = 4$ and $k = 64$. For parameter balancing, two convolutional layers with an activation function replace the proposed structure to construct original network.

By comparing the first and second rows of Table 1, we can observe the effectiveness of DRB structure since PSNR/SSIM values achieved by G-CASR with DRB structure are higher than those of the original network on most datasets. It is noted that G-CASR with DRB fails to improve reconstruction effect on Set14 and B100 datasets. This is caused by unsuccessful modeling of complicated and multi-type context information embedded in natural scene scenario. This conclusion can be further proved by tests on Manga109 dataset, where G-CASR with and without DRB achieve results of 38.61/0.9769 and 38.40/0.9767, respectively. Manga109 dataset contains only cartoon drawings, which makes it easy to model context information. By comparing the first and third rows of Table 1, we can observe the effectiveness of PG since PSNR/SSIM values achieved by G-CASR are higher than those of the original network on all the datasets including B100 and Urban100.

Between the results of the first and last rows, we can notice the network with DRB and PG achieves improvements on all listed datasets, which proves the effectiveness of the proposed DRB and PG for feature map enhancement. Moreover, PG enhances reconstruction performance based on the network only with DRB, which can be proved by the fact that DRB with PG achieves 0.13/0.0013 and 0.09/0.0003 improvement on Urban100 and Manga109.

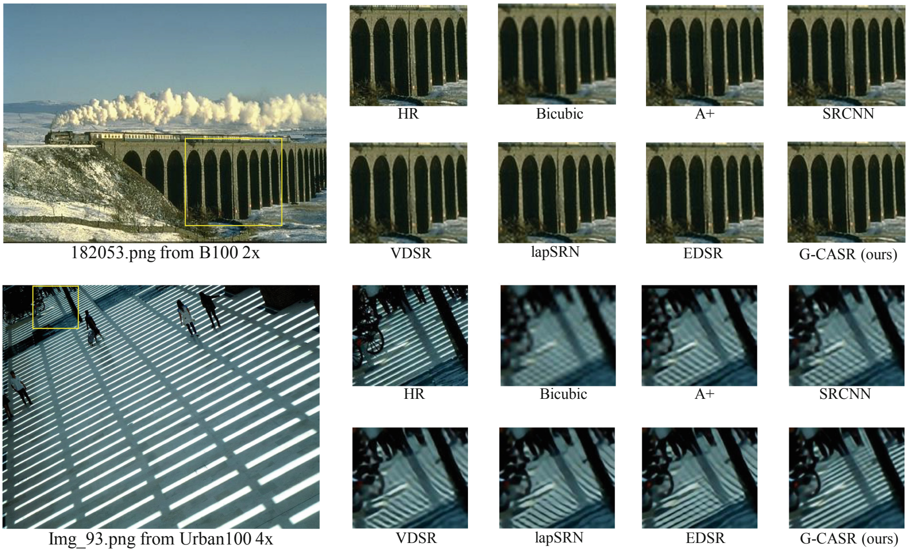

Fig. 5. Visual comparisons for SISR on B100 and Urban100 dataset, where the yellow rectangle represents enlarged regions for comparisons. (Color figure online)

4.3 SISR Performance and Analysis

Table 2 shows quantitative comparative results with 6 SISR algorithms for 2×, 3× and 4× SISR, respectively. It is noted that we obtain results of all comparative methods on five public datasets directly from their published papers. Among these methods, we pay special attention to EDSR since it is the current state-of-the-art algorithm for SISR. We test G-CASR by setting $m = 4$, $n = 8$ and $k = 128$.

From Table 2, we can notice G-CASR achieves better performance on Set5, Set14, B100 and Manga109 datasets compared with EDSR at 2×, 3× and 4×. This is due to their context information is easy to be captured and described by the proposed structure, thus enhancing feature map by context information. In fact, less performance improvement on Urban100 can be viewed by comparing G-CASR with EDSR because the former suffers from fewer model parameters. With only 25% model size of EDSR, G-CASR still produces superior SISR results. For example, the PSNR/SSIM value of G-CASR and EDSR on Manga109 are 31.14/0.9163 and 31.02/0.9148, respectively. The proposed structure acts well in most cases to partly describe context information embedded in complex urban and natural scenes, which are difficult to completely modeling. This can be proved by the fact that G-CASR obtains all the best performance during testing on B100 and Urban100 at 4×.

Figure 5 shows comparisons of visual effects achieved by G-CASR and comparative methods on B100 and Urban100. We can notice that G-CASR accurately reconstructs straight lines and parallel grid patterns on building surface and ground texture. This is because the proposed G-CASR network well preserves low-frequency features. We notice blurry effects and loss of image details achieved by other comparative methods for testing on image containing hair and beards since they fail to achieve clear focus and restore high-frequency details through learning on abundant low-frequency features. In contrast, our approach effectively suppresses these methods by removing redundant information but remember high contributive information.

4.4 Implementation Details

The upsampling layer contains a convolutional layer with 3 × 3 kernel and a pixel-shuffle operation afterward. The number of feature channels after the convolutional operation is s times the input so that the pixel-shuffle operation can generate an enlarged feature map, where s refers to scale factor. To make full usage of training data, we used a data augmentation method, in which each training picture is rotated 90°, 180°, 270° with a probability of 0.5, or flipped along a horizontal position. The input patch size is set as $48 \times 48 \times 3$. We adopt the Adam optimizer by setting its hyperparameters with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. We adopt L_1 loss function and set the initial learning rate as 0.0001. It is noted the learning rate decays by 0.5 for every 100 epochs and the total number of training epoch is 300. All of these experiments are performed on a single GTX 1080Ti GPU with 12 GB memory.

5 Conclusion

In this work, we propose a deep and lightweight context-aware residual network named as G-CASR, which appropriately encodes channel and spatial attention information to construct a context-aware feature map for SISR. Comparative results show that G-CASR not only achieves superior SISR performances than the current state-of-the-art method, i.e., EDSR, but also has the advantages of fewer parameters and less memory requirement. Our future work includes explorations to achieve real-time performance and better visual effects with extreme imaging situations.

Acknowledgment. This work is supported by the Natural Science Foundation of China under Grant 61672273, Grant 61832008, and Grant 61702160, Scientific Foundation of State Grid Corporation of China (Research on Ice-wind Disaster Feature Recognition and Prediction by Few-shot Machine Learning in Transmission Lines), National Key R&D Program of China under Grant 2018YFC0407901, and the Science Foundation of Jiangsu under Grant BK20170892.

References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of BMVC (2012)

2. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
3. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
4. Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., Aizawa, K.: Manga109 dataset and creation of metadata. In: Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding, p. 2 (2016)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
6. Hu, Y., Li, J., Huang, Y., Gao, X.: Channel-wise and spatial feature modulation network for single image super-resolution. arXiv preprint arXiv:1809.11130 (2018)
7. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of CVPR, pp. 5197–5206 (2015)
8. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of CVPR, pp. 1646–1654 (2016)
9. Kim, J.H., Choi, J.H., Cheon, M., Lee, J.S.: Ram: residual attention module for single image super-resolution. arXiv preprint arXiv:1811.12043 (2018)
10. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of CVPR (2017)
11. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of CVPR, vol. 1, p. 4 (2017)
12. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of ICCV, vol. 2, pp. 416–423 (2001)
13. 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 Computer Vision and Pattern Recognition Workshops, pp. 114–125 (2017)
14. Timofte, R., De Smet, V., Van Gool, L.: A+: adjusted anchored neighborhood regression for fast super-resolution. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 111–126. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16817-3_8
15. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: Proceedings of ICCV, pp. 4809–4817 (2017)
16. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
17. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
18. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., et al. (eds.) Curves and Surfaces 2010. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47