

Edge Maps Assisted Image Super-Resolution using Spatial Feature Transform

Ahmad Waseem, Ye Liu and Liang Wan*

College of Intelligence and Computing, Tianjin University, Tianjin, China

ABSTRACT

Deep learning has revolutionized every field of computer vision, including single image super-resolution because of its remarkable performance pertaining to effectiveness and efficiency. Various recent methods try to predict the SR image by incorporating different prior knowledge of images. In this paper, we propose a new method that utilizes not only internal image features but also multi-level edge prior knowledge with richer information. Holding the intuition that edge information helps to deal with blurry edges and try to generate sharper results, we present a residual edge and channel attention super-resolution network to handle LR images, named *RECAN*. Our architecture consists of two basic modules: the first module is *EdgeNet*, which generates multi-level edge maps from the input image; and the second module takes advantage of significant information in input image along with edge maps, called *SRNet*. Specifically, the *SRNet* uses channel attention technique and spatial feature transform (*SFT*) layers to super-resolve an image. Qualitative and quantitative comparisons are presented with state-of-the-art methods, which show promising results of our method together with improved image quality.

Keywords: Single Image Super-Resolution, Edge Information, Channel Attention, Spatial Feature Transform

1. INTRODUCTION

Single image super-resolution (SISR) is an active research topic and has attracted a lot of researchers to work on it. It aims to enhance an image from low resolution to high resolution, which has been studied widely in computer vision. This SISR has many important applications in different fields such as remote imaging,¹ medical imaging,² etc. It is a vital yet challenging research topic in digital image processing due to its ill-posed nature³ since multiple solutions can map to any LR image. As a classical problem, there are many solutions based on different techniques, and learning-based methods have been widely used in recent years. As with the rapid development of deep learning methods recently, deep learning models have shown impressive improvements in super-resolution over all other techniques because of its astonishing feature representation and training exemplar. A variety of deep learning methods, such as SRCNN,⁴ LapSRN⁵ and EDSR,⁶ have been proposed to solve super-resolution problems. Although these methods have achieved promising results, SISR still has some limitations, like, new deep learning-based methods just try to focus on designing a deeper architecture by stacking more residual blocks to learn more high-level features while not utilizing and ignoring the correlated features in intermediate or deeper layers, which hinders the real power of CNNs.

Various recent methods try to resolve super-resolution by incorporating different prior knowledge of natural images, which establishes maximum a posteriori (MAP) framework. In this framework, a high resolution (HR) image is predicted by maximizing its fidelity to the target with various kinds of priors. There are different aspects regarding MAP based SR methods,⁷ such as data fidelity or priors knowledge. Multiple priors have been used to generate more plausible output images. These priors describe the properties of natural images, such as smoothness, non-local similarity or sparseness. Textures are challenging to recover, especially when an image is degraded, but edges are easier. Intuitively, these edges can help to produce a better HR image, which uplifts the importance of edge prior. Thus, separating edges from an input image and then modeling them with input image in deep learning network would benefit SR image performance.

To deal with SR problems using edge prior, we propose a deep residual edge and channel attention network, named *RECAN*. This model is divided into two parts, *EdgeNet* and *SRNet* as shown in Fig. 1. *EdgeNet* is an edge detection part of *RECAN* that focuses on taking an input image and producing multi-level edge maps,

*corresponding author

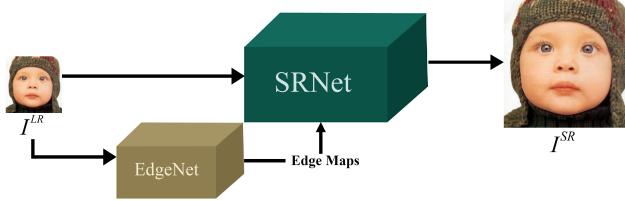


Figure 1. Illustration of proposed RECAN which includes two modules EdgeNet and SRNet

while SRNet module is the main part that takes LR image along with edge maps, and generates SR image. This SRNet module includes channel attention (CA) block and spatial feature transform (SFT) block. RECAN uses residual learning architecture similar to the residual channel attention network.⁸

The main contributions of this paper are summarized as follow:

- We propose a novel *RECAN* model to solve single image super-resolution problem. Extensive experiments on benchmark datasets images show promising results both quantitatively and qualitatively.
- We develop a method for embedding natural image prior into LR image, which uses edge information as a prior to overcome the edge artifacts and generate a better SR image.
- We employ the spatial feature transform layers to merge edge maps with a channel attention mechanism. These two techniques are used in one model to utilize prior knowledge with image features maps. Other priors, such as depth maps or segmentation maps, can also be applied.

2. RELATED WORK

Researchers have proposed a lot of image super-resolution work since this field had initiated a few decades ago including example-based methods, interpolation-based methods and learning-based methods. During this decade, deep learning-based methods emerged from the learning-based method and revolutionized the SISR field.

Deep Learning-based SISR. In the last few years, deep learning-based methods have been focused in SISR because of its non-linear representational ability. Dong et al.⁴ started working on CNNs and chose a three-layer CNN to learn an end-to-end mapping to generate SR image from an LR image. Xuhui et al.¹² tried to use SRCNN and successfully have created JLH-SRCNN model for better performance. Later, VDSR¹³ and DRCN¹⁴ are proposed based on residual learning with 16 layers that try to handle multiple scales. In the same vein, Lim et al.⁶ attempted to improve SISR by stacking residual blocks in a very deep network. Some methods choose to improve LR image first, then do the upscaling like Shi et al.¹⁵ adopted a sub-pixel upscaling to improve network receptive field. Laplacian pyramid super-resolution network (LapSRN)⁵ takes an input image and uses transposed convolution to gradually generate sub-band residuals. Generative adversarial network (GAN) based method SRGAN¹⁶ focuses on the perceptual quality of an image and uses sub-pixel convolution layers to upscale the input.

Attention is a kind of direction where we can try to utilize the informative resource of an image. Multiple methods have been proposed using the attention mechanism in different computer vision fields. Residual channel attention is proposed by Wang et al.¹⁷ for image classification with the attention mechanism. Zhang et al.⁸ took inspiration from¹⁷ and applied it to SISR. Our work is inspired by these two attention mechanism methods. We also use spatial feature transform (SFT) that is influenced by Xintao et al. work.¹⁸

Edge Detection. Edge detection is also a significant research field in computer vision. This field is started from Sobel operator¹⁹ that produces an edge map. Later, Gaussian smoothing²⁰ is used to generate more robust edge maps. It is one of the most efficient methods of edge detection that is being used till now. After the arrival of deep learning, CNN achieves significant gains in edge detection performance. Such deep learning-based methods include CSCNN,²¹ HED²² and CaseNet.²³

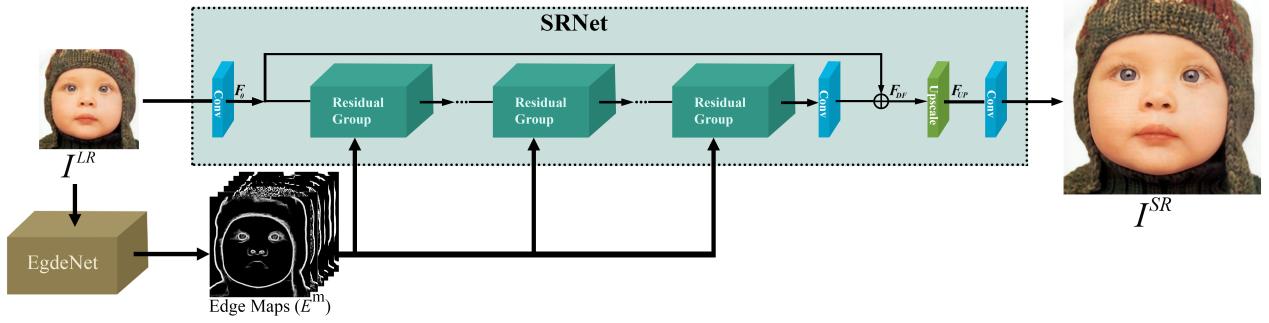


Figure 2. The proposed network architecture RECAN which includes two modules EdgeNet and SRNet while SRNet is composed of multiple residual groups, convolutional layers and upsampling layer.

Joint estimation methods. There are three types of tasks in computer vision: low-level, mid-level, and high-level vision tasks. Joint estimation methods try to solve the image processing problems by joining these tasks. Liu et al.²⁴ joined two different methods, image denoising that is considered as low-level vision task and image segmentation that is a high-level vision task relatively, using deep learning to improve the performance. Similarly, Makwana et al.²⁵ proposed a high-level vision task in image super-resolution where they use Canny edge detector²⁰ for edge detection and use iterative feedback technique. Later, Yang et al.²⁶ tried to improve the SR results using Sobel operator¹⁹ to extract an edge map and combine it with interpolated input. Kamyar et al.²⁷ proposed an SR method that is inspired by their inpainting method, where they consider missing information in an LR image as missing pixels just like the inpainting technique. To overcome the issues in previous methods and improve SISR performance, we propose this RECAN method.

3. METHODOLOGY

This section introduces our residual network based model that uses edge maps with channel attention and spatial transform layer modules. As shown in Fig. 2, the proposed method consists of two major modules, EdgeNet, and SRNet. We first present the EdgeNet, and then discuss the proposed SISR network architecture.

3.1 EdgeNet

Edge detection is one of the significant fields in image processing. There are different types of edge detectors consisting of classical and CNNs-based methods. Edges can be used as a prior knowledge of natural images to help other computer vision tasks. Among other priors, edge information is one of the most informative priors. In contrast to the texture prior, which is challenging to recover in LR image, edge information is uncomplicated

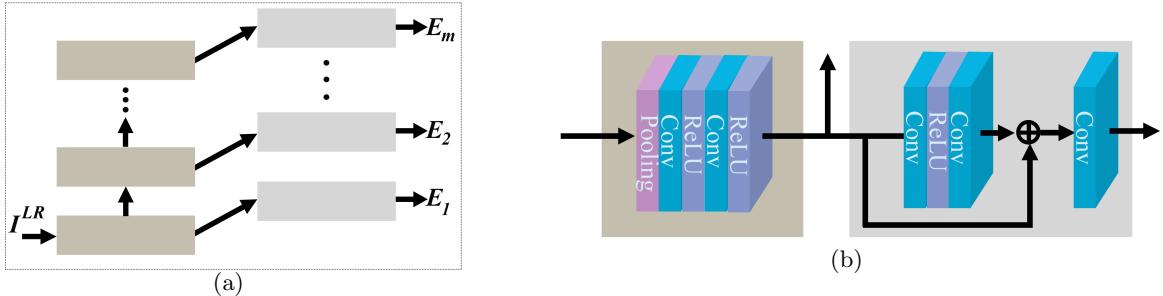


Figure 3. Illustration of proposed EdgeNet. (a) shows the whole EdgeNet architecture that has multiple blocks generating multiple outputs. (b) represents the interior of one block in (a).

to detect after image degradation. These edges are more informational for recovering the details of HR images. Therefore, we utilize this edge prior to reconstruct SR image in our method. Unlike other methods, our method

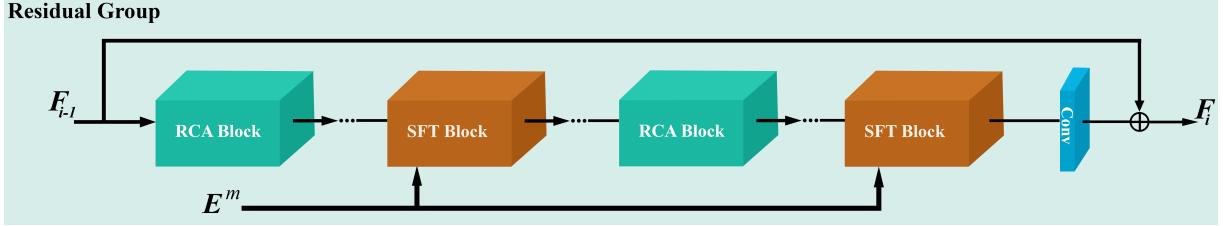


Figure 4. Illustration of Residual Group structure that includes Channel attention block and SFT block.

uses multiple edge maps instead of only one edge image. To get these multiple edge maps, EdgeNet is proposed. The edge prior can be represented by these edge maps E^m as

$$E^m = (E_1, E_2, \dots, E_m) = E^{net}(I^{LR}). \quad (1)$$

where I^{LR} is an LR input image. E^{net} takes I^{LR} as input and generates multiple edge maps. m in E^m is the total number of maps. These generated maps are sent to SRNet module of the proposed architecture. Most of the methods have used one edge map as prior information, while our method tries to use multi-level edge maps. These multi-level maps are resourceful maps with different scales that consist of rich information, which can be utilized in SRNet to achieve better performance.

EdgeNet architecture is shown in Fig. 3. This architecture is inspired by HED²² because HED tries to generate multi-level features perceptually. Fig. 3 consists of two parts, where Fig. 3(a) represents the complete EdgeNet architecture having multiple blocks and producing multiple edge maps. Fig. 3(b) is the illustration of a block from Fig. 3(a) interiorly that includes pooling layer block, residual block and a convolutional layer. The Left side of Fig. 3(b) represents a block from the left side of Fig. 3(a), while the right side of Fig. 3(b) is the interior of a block of Fig. 3(a). We adopt the HED architecture with the following modifications: (1) In EdgeNet, we modify the last part by adding residual blocks between the first part and the last convolutional layer. We try to go more deeper to take the benefit of the residual block in detecting the edges and using the receptive field to produce an enhanced feature map. As shown in Fig. 3(b), there is an additional residual block that is taking feature maps and generating output using skip connection. (2) Instead of using the only final output of HED, we change the architecture to use the side outputs, which can be seen in Fig. 3(a) generating n maps. By changing this, EdgeNet can produce multiple edge maps that can be used in SRNet. EdgeNet is being trained separately using I^{HR} instead of I^{LR} , because I^{LR} image contains less information than I^{HR} . After training EdgeNet, pre-trained model is used to generate edge maps E^m using I^{LR} . These E^m are used in SRNet for better performance.

3.2 Proposed Method

In this section, we describe our proposed method, including SRNet module. Given as, I^{LR} for LR image for our framework, the relation can be modeled as

$$I^{SR} = RECAN(I^{LR}), \quad (2)$$

where $RECAN$ is our proposed method that takes I^{LR} input and generates the output I^{SR} image. For SRNet, the equation can be

$$I^{SR} = SRNet(I^{LR}, E^m), \quad (3)$$

where E^m are the edge maps from equation. 1. $SRNet$ is the second module that generates the output using I^{LR} and E^m . This SRNet consists of four parts: feature extraction, residual group, upscaling module and reconstruction module. For the first part, we can say

$$F_0 = \Phi_{FE}(I^{LR}), \quad (4)$$

where Φ_{FE} is the convolutional operation and F_0 represents the feature extracted. This first part of SRNet takes the input image and generate the feature maps and send it next. For the next part, we can have

$$F_{DF} = \Phi_{RG}(F_0, E^m), \quad (5)$$

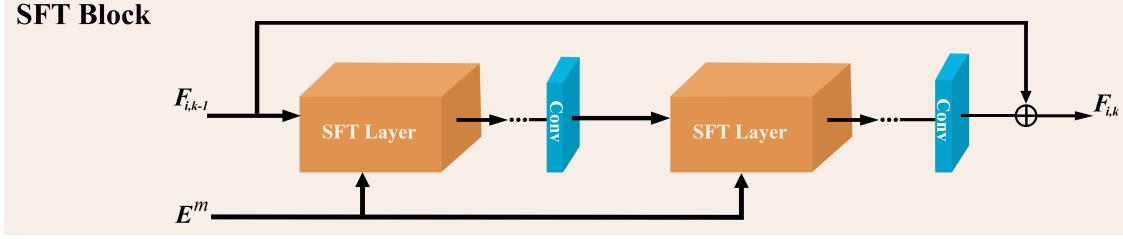


Figure 5. Representation SFT Block that includes Spatial Feature Transform (SFT) layers with convolutional layers and a skip connection.

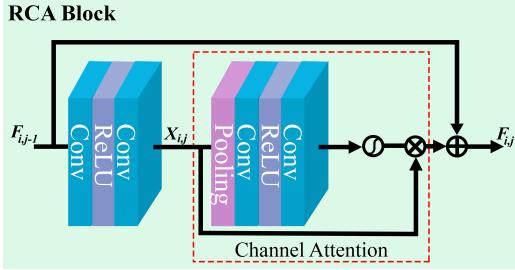


Figure 6. RCA Block with Channel Attention Structure.

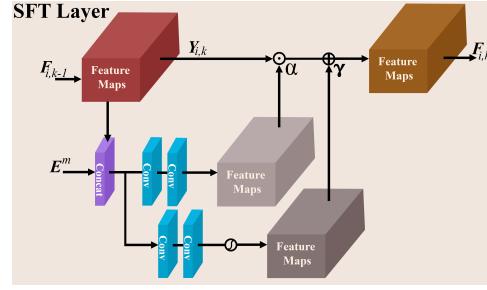


Figure 7. Spatial Feature Transform (SFT) layer Structure

where Φ_{RG} is the very deep residual group, which consists of multiple residual blocks. Later, these deep features are sent to next part as

$$F_{UP} = \Phi_{UP}(F_{DF}), \quad (6)$$

where Φ_{UP} is the upscaling module that upscale the deep features and sends it further for the reconstruction.

Residual Group. The major part in the SRNet is a residual group that has a long skip connection and contains channel attention blocks with short skip connection. This is different from residual in residual in RCAN because in this residual group, we add another spatial feature transform (SFT) block next to channel attention block. This SFT block can transform the features inside this deep network that can improve the performance. This residual group architecture can be seen in Fig. 4. Such structures help the model to train very deep to get better performance. For the residual group, i -th group can be written as

$$F_i = \Phi_i(F_{i-1}, E^m) = \Phi_i(\Phi_{i-1}(\dots\Phi_1(F_0, E^m)\dots)), \quad (7)$$

where Φ_i is the function for the i -th residual group, and F_{i-1} and F_i are the input and output for that group, respectively. This residual group takes the convolutional features (F_0) from initial part along with edge maps (E^m) from EdgeNet and pass them to internal residual channel attention (RCA) blocks and SFT blocks for further operations.

Channel Attention Some recent deep learning-based methods consider to treat LR image channels differently. Such methods^{8,28} try to produce different attention for each channel-wise feature. This technique rescales these features by modeling the feature interdependencies. This mechanism is depicted in Fig. 6, where the feature maps are extracted from the channel network using a global average pooling layer to obtain a global context embedding vector. Then two layers are applied to learn nonlinear interactions with the help of a gating mechanism that is actually a sigmoid function. Finally, the last channel scaling statistics is obtained to rescale the feature maps.

This channel attention is merged into a residual block. For the j -th RCA block in the i -th residual group, it can be

$$F_{i,j} = F_{i,j-1} + CA_{i,j}(X_{i,j}), \quad (8)$$

where $CA_{i,j}$ is the function for the channel attention. $F_{i,j-1}$ and $F_{i,j}$ are the input and the output of this block, respectively. While $X_{i,j}$ is obtained after two convolutional layers operations on the input $F_{i,j-1}$. Inside a residual

group, the input is passed to the RCA block where channel attention mechanism is implemented for channel-wise operation and the obtained feature maps are shifted towards the SFT block.

Spatial Feature Transform. Spatial Feature Transform (SFT) block consists of SFT layers as shown in Fig. 5. These SFT layers produce the output parameters based on some prior information by learning a mapping function. These parameters affect the output feature maps when the affine transformation is applied to the input feature maps. Based on the prior information, it can be modeled as

$$(\gamma^E, \beta^E) = f(E^m + F_{i,k-1}), \quad (9)$$

where $f(\cdot)$ denotes the mapping function that generates two parameters γ and β and takes concatenated edge maps with feature maps as input; alternatively, it can take any other prior. After getting the parameters from mapping function, this transformation can be

$$F_{i,k} = SFT(\alpha \odot Y_{i,k} + \gamma), \quad (10)$$

where $SFT(\cdot)$ is the function that transforms the feature maps. $Y_{i,k}$ and $F_{i,k}$ denotes the input and output of SFT whose dimensions are same as α and γ while \odot is Hadamard product which is used for element-wise multiplication. To take the benefit of a prior knowledge in the deeper network, these SFT layers are an advantageous technique used in our proposed method. We apply the affine transformation to the edge maps.

As shown in Fig. 7, where Spatial Transform Layer (SFT) structure is illustrated, SFT can take edge maps as input along with feature maps and concatenate them. The input feature maps of these layers are the output of the previous RCA block. After concatenation, convolution operations have been performed that generate two parameters, as mentioned above. Later, these parameters are merged with feature maps and produce the output feature maps which are passed on to the next RCA block.

After all the operations from RCA block and SFT block, the output is convoluted again. Element-wise sum with a skip connection from the initial feature maps have been performed here, and the output is passed to the upscaling module. After upscaling, the image is reconstructed by the last part of our method and generates the final SR image output.

Table 1. Results Comparisons with SR models that use edge prior in their method. Best results are **highlighted**

Method	PSNR								
	Set5			Set14			B100		
	$\times 2$	$\times 4$	$\times 8$	$\times 2$	$\times 4$	$\times 8$	$\times 2$	$\times 4$	$\times 8$
DEGREE	37.40	31.03	-	32.96	27.73	-	31.73	27.07	-
Edge-SISR	37.70	31.15	-	33.28	28.15	-	31.98	27.35	-
EdgeInfo	33.60	28.59	23.73	29.24	25.19	23.73	28.12	24.25	21.63
SREdge	-	31.02	27.12	-	27.24	27.12	-	27.06	24.84
SeaNet	38.08	32.33	-	33.75	28.72	-	32.27	27.65	-
RECAN(ours)	38.28	32.66	27.27	34.19	28.89	25.21	32.41	27.76	24.92

4. EXPERIMENTS

4.1 Datasets and Network Training

We have used DIV2K dataset²⁹ for the training of our models of SRNet. This dataset is provided by NTIRE Challenge 2017 that contains high-quality images. It consists of 800 training images along with 100 test images. While EdgeNet is being trained using NYUD³⁰ dataset with ground truth by Canny edge detector. For testing, we use standard benchmark datasets which includes Set5,³¹ Set14,³² BSD100,³³ Urban100,³⁴ and Manga109.³⁵ We conduct the experiments using bicubic degradation.

The training dataset is augmented with random rotations of 90° , 180° , 270° , and random horizontal flips. We set the patch size to 48×48 during each training batch. This model is trained with ADAM optimizer with internal parameters as $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is set to 10^{-4} at the start, which decreases

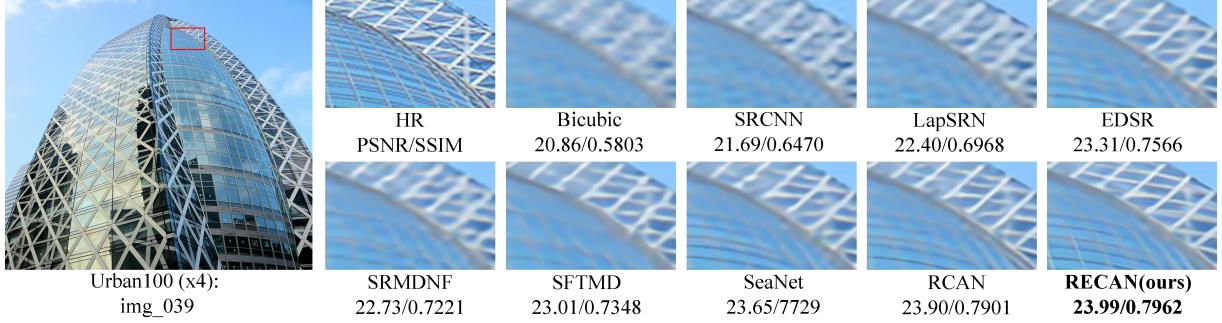


Figure 8. Visual comparison for $\times 4$ scaling factor on Urban100 dataset.

every 2×10^5 iterations to half. We set the residual group to 10 where each residual group contains 20 residual blocks that comprise 10 RCA blocks and 10 SFT blocks. The results are evaluated with PSNR and SSIM.³⁶ We implement our models using PyTorch³⁷ with NVIDIA RTX 2080Ti GPU.

4.2 Results and Comparisons

To test the effectiveness of our method, we present RECAN model (residual edge and channel attention network) that uses edge prior with channel attention mechanism using residual network architecture. The comparison is listed with 12 different state-of-the-art methods: SRCNN,⁴ LapSRN,⁵ DEGREE,²⁶ EDSR,⁶ SRMD,³⁸ SREdge,³⁹ SFTMD,⁴⁰ RCAN,⁸ Edge-SISR,⁴¹ EdgeInfo,²⁷ SeaNet,⁴² and SAN.²⁸ The comparison with these methods is divided into two tables. Our comparison is directly related to the methods of Table. 1 because these methods use edge information as a prior in their methods. Nevertheless, we also compare our method with general state-of-the-art frameworks in Table. 2. All the quantitative results for different scaling factors are presented in Table. 1 and 2. As shown in Table. 1, we compare the results using three benchmark datasets (Set5, Set14 and BSD100) and presented by PSNR. It can be seen clearly that our RECAN method has surpassed all the methods with a significant margin. In Table. 2, we compare the results using five benchmark datasets (Set5, Set14, BSD100, Urban100, and Manga109) and shown by PSNR/SSIM. In this table, the comparisons with other methods show a clear difference. Our method has superior results among all these methods. There are two methods, RCAN and SAN, which are closer to our method mainly because both of them also use the channel attention mechanism to learn feature interdependencies that make the method to focus on distinctive features. Thus, RECAN proves its effectiveness on images while using more information over other methods.

This RECAN uses only edge information alongside the channel attention mechanism and produces promising results. We have used multiple edge maps as prior information instead of one map because multiple maps provide more edge information comparative to one edge map, and these multiple maps better improve the SISR result than a single map. As we can see the difference between one map and multiple maps from Table. 2 where RECAN^e presents results using one edge map while RECAN^m shows the results for multiple maps.

We have also shown the zoomed visual results of different methods with scaling factor $\times 4$ and $\times 8$. From these visual results, it can be seen that most of the methods cannot recover HR image efficiently. In contrast, our proposed method produces more high frequency details in the final output. For example, in Fig. 8, most of the compared models generate blurring artifacts and cannot recover sharp edges. Earlier methods like bicubic, SRCNN and LapSRN, even fail to recover the main structure while the recent methods (e.g., SRMDNF, SFTMD and EDSR) recover main outlines. SeaNet and RCAN successfully generate a more detailed structure. Compared with ground truth, our method obtains more closer results and recovers a better image with sharper detail among all other methods. Similar results can also be seen in Fig. 9 with scaling factor $\times 8$. These results demonstrate the authenticity and superiority of our proposed method. Recovering the high frequency detail from an LR image is difficult due to the availability of less information, but our method can utilize the limited LR information through edge prior and channel feature correlations that help to produce finer results.

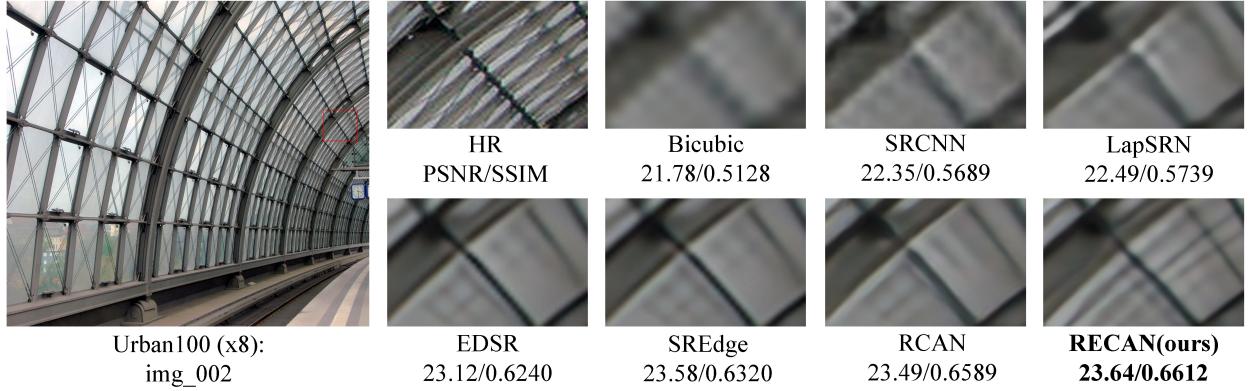


Figure 9. Visual comparison for $\times 8$ scaling factor on Urban100 dataset.

Table 2. Results Comparisons with general SR models. Best results are **highlighted** and second best are underlined

Method	Scale	Set5		Set14		B100		Urban100		Manga100	
		PSNR	SSIM								
Bicubic	$\times 2$	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
SRCCNN	$\times 2$	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
LapSRN	$\times 2$	37.52	0.9591	33.08	0.9130	31.80	0.8950	30.82	0.9101	37.27	0.9740
EDSR	$\times 2$	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.92	0.9351	39.10	0.9773
SRMDNF	$\times 2$	37.79	0.9601	33.32	0.9159	32.05	0.8985	31.33	0.9204	38.07	0.9761
SFTMD	$\times 2$	35.44	0.9617	31.27	0.8676	30.54	0.8946	27.80	0.8464	08.75	0.9074
RCAN	$\times 2$	38.25	0.9614	34.05	0.9213	32.38	0.9022	33.19	0.9374	39.45	0.9783
SAN	$\times 2$	38.30	0.9620	<u>34.07</u>	<u>0.9213</u>	<u>32.39</u>	<u>0.9025</u>	33.10	0.9370	39.32	0.9790
RECAN ^e (ours)	$\times 2$	38.24	0.9612	34.01	0.9211	32.35	0.9023	<u>33.21</u>	0.9373	39.40	0.9784
RECAN ^m (ours)	$\times 2$	<u>38.28</u>	<u>0.9618</u>	34.19	0.9224	32.41	0.9026	33.37	0.9389	39.49	<u>0.9786</u>
Bicubic	$\times 4$	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCCNN	$\times 4$	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
LapSRN	$\times 4$	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560	29.09	0.8900
EDSR	$\times 4$	32.46	0.8968	28.80	0.7876	27.70	0.7420	26.64	.8033	31.02	0.9148
SRMDNF	$\times 4$	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731	30.09	0.9024
SFTMD	$\times 4$	29.29	0.9014	26.40	0.7137	26.16	0.7648	22.97	0.6722	24.24	0.7950
RCAN	$\times 4$	32.60	0.9000	28.83	0.7886	27.72	0.7434	26.78	0.8085	31.21	0.9170
SAN	$\times 4$	32.64	0.9003	28.92	0.7888	<u>27.76</u>	<u>0.7436</u>	26.79	0.8086	31.18	<u>0.9169</u>
RECAN(ours)	$\times 4$	32.66	0.9005	<u>28.89</u>	0.7890	27.76	0.7437	26.83	0.8089	31.23	0.9172
Bicubic	$\times 8$	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160	21.47	0.6500
SRCCNN	$\times 8$	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440	22.46	0.6950
LapSRN	$\times 8$	26.15	0.7380	24.35	0.6200	34.54	0.5860	21.81	0.5810	23.39	0.7349
EDSR	$\times 8$	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221	24.69	0.7841
SAN	$\times 8$	27.22	0.7840	25.13	0.6476	24.88	0.6011	22.70	0.6314	24.85	0.7906
RCAN	$\times 8$	27.29	0.7876	<u>25.20</u>	<u>0.6510</u>	<u>24.90</u>	0.6049	<u>22.97</u>	<u>0.6448</u>	<u>25.24</u>	0.8029
RECAN(ours)	$\times 8$	<u>27.27</u>	<u>0.7876</u>	25.21	0.6510	24.92	<u>0.6040</u>	22.98	0.6451	25.26	<u>0.8027</u>

5. CONCLUSION

In this paper, we propose a residual edge and channel attention network, RECAN, to produce better SISR results using prior knowledge. This method is divided into two modules, EdgeNet and SRNet. EdgeNet provides edge maps from the input image by detecting edges from the images. These maps are used in SRNet by spatial feature transform (SFT) layers. SRNet uses residual groups that consist of channel attention blocks with SFT blocks. Such groups allow to pass low-frequency information through multiple blocks with skip connections that help to recover high frequency information. Using edge maps by SFT help to recover a finer SISR output with sharper edges. SFT layers help to incorporate these maps with feature maps, which leads to a superior SISR method. Extensive experiments on SISR using RECAN demonstrate the effectiveness of our proposed method for producing high-resolution images with richer details.

6. ACKNOWLEDGMENT

The work is supported by the research fund for the Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China (Grant No. 2019ASP-TJ01).

REFERENCES

- [1] Belov, A. M. and Denisova, A. Y., “Algorithm for spectral-spatial remote sensing image super-resolution: multi-sensor case,” in [*Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*], Li, C., Yu, H., Pan, Z., and Pu, Y., eds., **11069**, 1030 – 1040, International Society for Optics and Photonics, SPIE (2019).
- [2] Xu, Z., Chen, Z., and Hou, W., “A self-similarity based method for medical image resolution enhancement,” in [*Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*], Li, C., Yu, H., Pan, Z., and Pu, Y., eds., **11069**, 804 – 808, International Society for Optics and Photonics, SPIE (2019).
- [3] Baker, S. and Kanade, T., “Limits on super-resolution and how to break them,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(9), 1167–1183 (2002).
- [4] Dong, C., Loy, C. C., He, K., and Tang, X., “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence* **38**(2), 295–307 (2015).
- [5] Lai, W.-S., Huang, J.-B., Ahuja, N., and 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*], 624–632 (2017).
- [6] Lim, B., Son, S., Kim, H., Nah, S., and 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*], 136–144 (2017).
- [7] Yang, J., Wright, J., Huang, T. S., and Ma, Y., “Image super-resolution via sparse representation,” *IEEE transactions on image processing* **19**(11), 2861–2873 (2010).
- [8] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y., “Image super-resolution using very deep residual channel attention networks,” in [*Proceedings of the European Conference on Computer Vision (ECCV)*], 286–301 (2018).
- [9] Song, X., Liu, W., Liu, J., Liu, C., Lu, C., and Gao, H., “Deep cnn jointing low-high level feature for image super-resolution,” in [*Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*], **11069**, 110693O, International Society for Optics and Photonics (2019).
- [10] Kim, J., Kwon Lee, J., and Mu Lee, K., “Accurate image super-resolution using very deep convolutional networks,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1646–1654 (2016).
- [11] Kim, J., Kwon Lee, J., and Mu Lee, K., “Deeply-recursive convolutional network for image super-resolution,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1637–1645 (2016).
- [12] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and 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*], 1874–1883 (2016).
- [13] 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*], 4681–4690 (2017).
- [14] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., and Tang, X., “Residual attention network for image classification,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3156–3164 (2017).
- [15] Wang, X., Yu, K., Dong, C., and Change Loy, C., “Recovering realistic texture in image super-resolution by deep spatial feature transform,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 606–615 (2018).
- [16] Feldman, J. A., Feldman, G., Falk, G., Grape, G., Pearlman, J., Sobel, I., and Tenenbaum, J. M., “The stanford hand-eye project.,” in [*IJCAI*], 521–526 (1969).

- [17] Canny, J., “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence* (6), 679–698 (1986).
- [18] Hwang, J.-J. and Liu, T.-L., “Pixel-wise deep learning for contour detection,” *arXiv preprint arXiv:1504.01989* (2015).
- [19] Xie, S. and Tu, Z., “Holistically-nested edge detection,” in [*Proceedings of the IEEE international conference on computer vision*], 1395–1403 (2015).
- [20] Yu, Z., Feng, C., Liu, M.-Y., and Ramalingam, S., “Casenet: Deep category-aware semantic edge detection,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 5964–5973 (2017).
- [21] Liu, D., Wen, B., Jiao, J., Liu, X., Wang, Z., and Huang, T. S., “Connecting image denoising and high-level vision tasks via deep learning,” *IEEE Transactions on Image Processing* **29**, 3695–3706 (2020).
- [22] Makwana, R. R. and Mehta, N. D., “Single image super-resolution via iterative back projection based canny edge detection and a gabor filter prior,” *International Journal of Soft Computing & Engineering* **3**(1), 379–384 (2013).
- [23] Yang, W., Feng, J., Yang, J., Zhao, F., Liu, J., Guo, Z., and Yan, S., “Deep edge guided recurrent residual learning for image super-resolution,” *IEEE Transactions on Image Processing* **26**(12), 5895–5907 (2017).
- [24] Nazeri, K., Thasarathan, H., and Ebrahimi, M., “Edge-informed single image super-resolution,” in [*Proceedings of the IEEE International Conference on Computer Vision Workshops*], 0–0 (2019).
- [25] Dai, T., Cai, J., Zhang, Y., Xia, S.-T., and Zhang, L., “Second-order attention network for single image super-resolution,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 11065–11074 (2019).
- [26] Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and 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*], 114–125 (2017).
- [27] Nathan Silberman, Derek Hoiem, P. K. and Fergus, R., “Indoor segmentation and support inference from rgbd images,” in [*ECCV*], (2012).
- [28] Bevilacqua, M., Roumy, A., Guillemot, C., and Alberi-Morel, M. L., “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” (2012).
- [29] Zeyde, R., Elad, M., and Protter, M., “On single image scale-up using sparse-representations,” in [*International conference on curves and surfaces*], 711–730, Springer (2010).
- [30] Martin, D., Fowlkes, C., Tal, D., and Malik, J., “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in [*Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*], **2**, 416–423, IEEE (2001).
- [31] Huang, J.-B., Singh, A., and Ahuja, N., “Single image super-resolution from transformed self-exemplars,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 5197–5206 (2015).
- [32] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K., “Sketch-based manga retrieval using manga109 dataset,” *Multimedia Tools and Applications* **76**(20), 21811–21838 (2017).
- [33] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing* **13**(4), 600–612 (2004).
- [34] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., “Automatic differentiation in pytorch,” (2017).
- [35] Zhang, K., Zuo, W., and Zhang, L., “Learning a single convolutional super-resolution network for multiple degradations,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3262–3271 (2018).
- [36] Kim, K. and Chun, S. Y., “Sredgenet: Edge enhanced single image super resolution using dense edge detection network and feature merge network,” *arXiv preprint arXiv:1812.07174* (2018).
- [37] Gu, J., Lu, H., Zuo, W., and Dong, C., “Blind super-resolution with iterative kernel correction,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1604–1613 (2019).
- [38] Seif, G. and Androutsos, D., “Edge-based loss function for single image super-resolution,” in [*2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 1468–1472, IEEE (2018).
- [39] Fang, F., Li, J., and Zeng, T., “Soft-edge assisted network for single image super-resolution,” *IEEE Transactions on Image Processing* **29**, 4656–4668 (2020).