IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 31, 2021, accepted August 26, 2021, date of publication August 30, 2021, date of current version September 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3108998

Edge Profile Super Resolution

JIUN LEE[1], (Graduate Student Member, IEEE), INYONG YUN[2], AND JAEKWANG KIM[3], (Member, IEEE)

[1]Department of Electronic, Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea
[2]Big Data & AI Lab, Hana Institute of Technology, Hana TI, Seoul 06133, Republic of Korea
[3]School of Convergence, Sungkyunkwan University, Seoul 03063, Republic of Korea

Corresponding author: Jaekwang Kim (linux@skku.edu)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant through the Ministry of Science and ICT (MSIT), Government of Korea, under Grant NRF·2019R1C1C1008174, and in part by the Ministry of Science and ICT (MSIT), South Korea, through the ICT Creative Consilience Program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2020-0-01821.

ABSTRACT The purpose of single-image super resolution (SISR) is to reconstruct an accurate high-resolution image from a degraded low-resolution image. Owing to the lack of information in low-resolution images, SISR is a challenging problem. In particular, it is difficult to represent details, including high-frequency components, such as texture and structural information. We propose the edge profile super-resolution (EPSR) method to preserve structural information and restore texture. EPSR is achieved by stacking modified fractal residual network (mFRN) structures hierarchically and repeatedly. Each mFRN is composed of many residual edge profile blocks (REPBs) that extract features and preserve the edge, structure, and texture information of the image. For implementing REPB, we design three main modules: Residual Efficient Channel Attention Block(RECAB) module, Edge Profile(EP) module, and Context Network(CN) module. By repeating the procedure in the mFRN structure, the EPSR method could be used to extract high-fidelity features, thus preventing texture loss and preserving the structure with appropriate sharpness. Experimental results show that EPSR achieves competitive performance against state-of-the-art methods in terms of the peak signal-to-noise ratio(PSNR) and structural similarity index measure(SSIM) evaluation metrics, as well as visual results.

INDEX TERMS Image super-resolution, contextual information, edge, image structure, texture.

I. INTRODUCTION

Single Image Super-Resolution(SISR) [1] has been the focus of recent research. Generally, SISR targets the reconstruction of an accurate high-resolution (HR) image from a degraded low-resolution (LR) image. Image super-resolution(SR) is usually applied to diverse computer vision tasks (e.g. security and surveillance imaging [2], object recognition [3], image generation [4], and medical imaging [5]). Although there are many methods for the reconstruction of any LR inputs, their solutions to the ill-posed problem [6] still demonstrate limitation with regard to details such as structure and texture. For high-fidelity image, it is necessary to represent such details as high frequency components such as texture and structural information. To address this issue, numerous SR methods have been proposed, such as conventional methods [7]–[14], deep learning methods [15]–[23] and the perceptual-driven method [3], [24]–[28].

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

A. RELATED WORK

In conventional methods, edge-based models [7]–[14] enhance the sharpness of super-resolved image by utilizing edge statistics. They model edge statistical dependencies by estimating the structural connectivity between HR and LR. Fattal [7] proposed a method for learning the prior dependencies among the edge statistics of image gradients. Sun *et al.* [8] proposed a gradient field transformation to control the HR gradient fields and enhance the sharpness. Yan *et al.* [9] proposed a method based on the gradient profile sharpness extracted from gradient description models. Tai *et al.* [10] proposed an approach to combine edge-directed SR with detail from an image and texture examples. Zhu *et al.* [11] proposed an SISR method based on gradient reconstruction by collecting a dictionary of gradient patterns. The edge distribution tends to depend heavily on the similarities between training and test datasets. Therefore, the results are determined by the similarities, and the performance lacks consistency. In addition, the processes are modeled point by point and are therefore complicated and inflexible.

Furthermore, because they focus on the sharpness of SR image, they face difficulties with improving texture restoration.

On the other hand, general deep learning methods are more flexible and effective in handling probability transformations, including pixel distribution. They produce outstanding results compared with previous methods [29], [30]. Deep learning methods generally approach the SISR problem by utilizing influential feature representation and a deep end-to-end structure. SRCNN proposed by Dong *et al.* [15] achieved noteworthy performance using a three-layer convolutional network. Later, VDSR [16] and DRCN [17] improved the accuracy by stacking convolutional networks deeply through residual learning. Tai *et al.* [18] introduced DRRN, which is a recursive learning model based on parameter sharing, and proposed MemNet [19], which consists of memory blocks for a deep network. EDSR and MDSR by Lim *et al.* [20] significantly improved the performance by stacking residual blocks deeply and widely. As a result, the depth of the network becomes a key point in the image SR. After the development of deep networks, RDN was designed by Zhang *et al.* [21] as a deep network based on the dense block to utilize all the hierarchical features from all convolutional layers. Zhang *et al.* [22] and Dai *et al.* [23] considered not only increasing the depth of the network, but also applying feature correlations in the spatial and channel dimensions.

Most of these were optimized to measure the pixel distance between the SR and its corresponding HR by mean square error(MSE) or L_1. These optimizing methods tend to cause the networks to generate an image based on the statistical information of possible HR solutions. Even though they reach high numerical value evaluation in terms of peak signal-to-noise ratio(PSNR), general deep learning models show blurry with texture loss and structural trouble results. To represent texture and preserve the image structure, some methods [6] simply apply edge information in a deep learning model by utilizing it as an assistant device. However, they designed their model to achieve a higher PSNR evaluation metric — that is, the structural information is utilized inadequately.

For perceptual improvement, some methods [3], [24]–[28] utilize the generative adversarial network(GAN) with perceptual loss to generate a photo realistic image. As mentioned previously, all general deep-learning methods concentrate on achieving a high PSNR. However, their results are blurry and unstable structural SR images. To recover SR images more toward realistic direction, Johnson *et al.* [24] proposed perceptual loss to enhance the visual quality of SR images. Ledig *et al.* [25] designed SRGAN based on adversarial loss, and it is the first model that can generate photo-realistic HR images. EnhanceNet by Sajjadi *et al.* [3] produced high-fidelity textures SR images by applying texture loss. Wang *et al.* [26] proposed ESRGAN, which enhances the previous frameworks by constructing Residual-in-Residual Dense Block. However, Wang *et al.* [27] generated more natural textures for specific categories by exploiting semantic segmentation maps as priors. In addition, SROBB by Rad *et al.* [28] involves objective perceptual loss based on the labels of the object, background and boundary. These perceptual-driven methods bring perceptual enhancement by restoring texture information related to blurriness. However, they leave structural distortion problems and fail to recover details such as texture. To overcome the structural limitations of images, some models [13], [14] utilize structural information by designing additional modules for preserving the structure. Ma *et al.* [13] utilized edge information in perceptual-driven methods by feeding explicit guidance to the established model, and Nazari *et al.* [14] proposed an edge-informed SR method based on an image inpainting task.

Although these methods compensate for the structural defects of the GAN-based model, they do not achieve the visual quality of HR images. Furthermore, because the discriminators may cause unstable factors during the optimization procedure, GAN-based models face difficulties with the stability of the learning process and maintaining structural consistency.

B. CONTRIBUTIONS

The EPSR method is designed to alleviate the issues mentioned above. In the SISR problem, to generate high-quality SR images, it is important to represent high-frequency details, such as structure and texture information. Since these components have frequent pixel variations, they have contextual properties, and thus displaying them is crucial for high- quality results. To achieve this goal, we modify Fractal Residual Network(FRN) as a network structure to utilize various information in the learning process. we call it modified Fractal Residual Network(mFRN) structure. To draw high-frequency components from diverse information, we construct Residual Edge Profile Blocks(REPBs) as basic blocks. An REPB consists of an Residual Efficient Channel Attention Block(RECAB) module, Edge Profile(EP) module and Context Network(CN) module. To extract high-fidelity features, it is necessary to utilize informative features that contain detail information. Hence, by referring to previous methods [20], [22] and recent research [31], we apply efficient channel attention(ECA) to feature extraction. This systemically organized feature provides abundant information to the EP module. The EP module feeds structural information on the features by generating the edge profile itself from the informative features. This module is based on the principle of conventional edge extraction. The EP module contributes to preserving the image structure. However, exploiting high frequency components, such as sharpness and textures, should be considered for high-fidelity results. These contextual details contain complex variations in specific regions(i.e. high frequency regions such as edge and texture). Thus, it is difficult to maintain the detail information in the process. To exploit high-frequency components, we construct a Context Network(CN) module. By exposing contextual information, this module captures pixel variation, and thus, the sharpness of the results can be enhanced

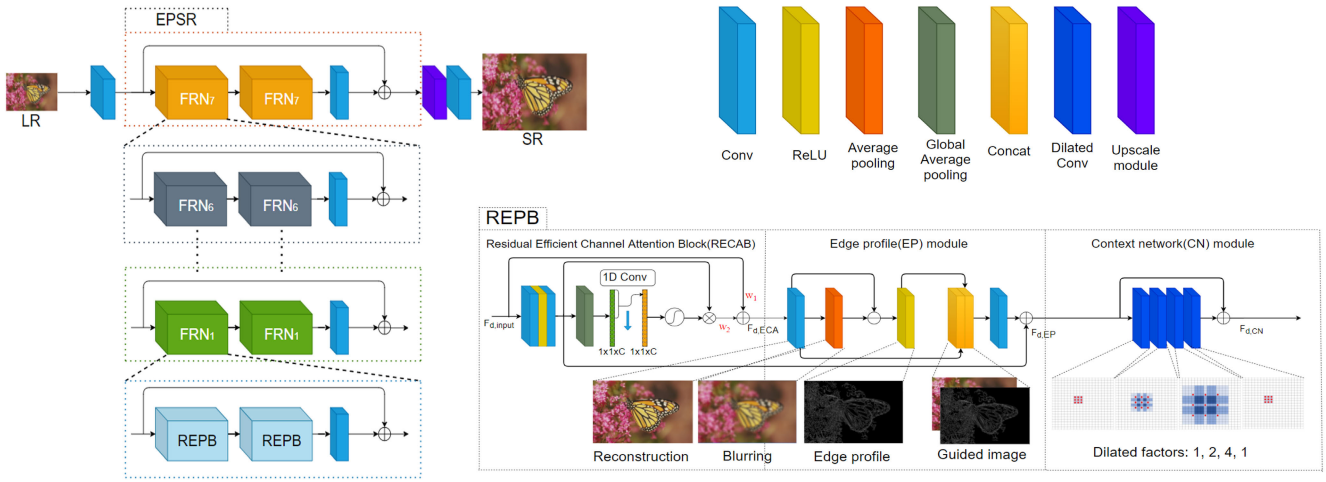


FIGURE 1. The architecture of the proposed edge profile super-resolution (EPSR) method follows a modified fractal residual network (mFRN) structure consisting of residual edge profile blocks (REPBs) as basic blocks. An REPB consists of the residual efficient channel attention block (RECAB), edge profile (EP), and context network (CN) modules. we also include the explanation of components of all blocks.

properly and texture loss can be restored. By proceeding repeatedly with this process in the network, the SR results show structural stability and represent details with reduced texture loss and structural distortions. Experimental results on benchmark datasets demonstrate that our EPSR improves the SR quality.

The contribution of this work can be summarized as follows:

- We propose our edge profile super resolution (EPSR) for high-quality image SR using image properties. The network is designed to learn diverse features from the original LR image.
- Our EPSR can preserve structural information stably by extracting the edge components itself.
- By exploiting contextual information, our EPSR can recover such details as structure and texture.

The remainder of this paper is organized as follows. In Section II,, the proposed method is explained. In Section III, the proposed method is compared experimentally with previous methods. The conclusions of the study are presented in Section IV.

II. METHODOLOGY

In this section, an overview of EPSR is presented. Then, the details of REPBs are introduced. Informative features are formed by utilizing structural information and exploiting high-frequency components. Finally, the objective functions are described.

A. OVERVIEW

The overall structure is described in Fig.1. As researched in [21], [32], we apply one convolution layer to extract the shallow feature from the LR input. To utilize diverse information in the process, we modify the FRN skip connection structure proposed by Kwak and Son [33] as shown in Fig.1. we call it mFRN. mFRN consists of REPBs. Since the self-similarity property of the mFRN structure gains deep depth and provides a large receptive field, REPBs can obtain diverse information and generate informative features effectively, which include high-frequency components containing such details as structural information and texture. Then, the deep features from the mFRN structure are upscaled by the upscale module. We apply this upscaling module as in previous work [21], [34]. According to the process, the upscaled feature is converted into an SR image using one convolution layer.

B. RESIDUAL EDGE PROFILE BLOCK (REPB)

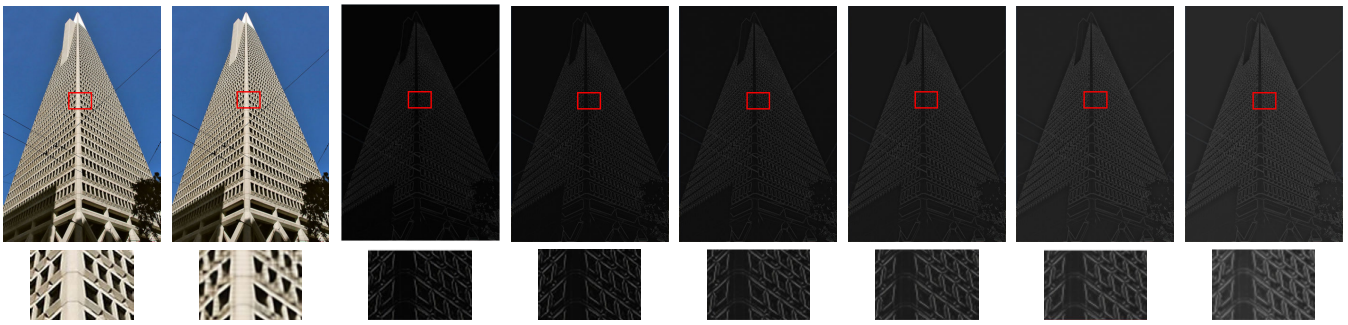
Due to self-similarity of the mFRN structure, abundant diverse frequency information can be bypassed. Based on this information, our proposed REPBs can focus on exploiting high-frequency components by utilizing influential features with structural information and exposing contextual information. The REPB consists of three parts: the RECAB, EP, and CN module.

1) RESIDUAL EFFICIENT CHANNEL ATTENTION BLOCK (RECAB)

As proposed for EDSR and MDSR [20], the feature is extracted by removing batch normalization layers. Thus, the range flexibility of the EPSR can be maintained and the feature extraction can be formulated as

$$F_{FE} = H_{FE}(F_{input}), \tag{1}$$

where the output F_{FE} and $H_{FE}(\cdot)$ stand for the feature and function from the feature extraction of the REPB block, respectively. Here, F_{input} is the input feature of the REPB block. In the SISR problem, RCAN proposed by Zhang *et al.* [22] considers feature interdependencies and utilizes mutual independence by applying the channel attention process from SENet [35]. However, this process has shown that dimensionality reduction results in side effects in channel prediction. By messing up the direct correspondence between its channel and weight, it captures



(a) HR (b) SR (c) 24th Edge (d) 48th Edge (e) 72th Edge (f) 96th Edge (g) 112th Edge (h) Final edge

FIGURE 2. Extracted edges of "img_048" by edge profile modules.

unnecessary dependencies across all channels empirically. To avoid this problem, we use efficient channel attention(ECA) by [31]. ECA captures local cross-channel interaction by using 1D convolution of size k, where kernel size k implies the coverage of local cross-channel interaction and the number of neighbors involved in the attention prediction of one channel. This process can be expressed as

$$w_{eca} = \sigma(C1D_k(g(F_{FE}))), \tag{2}$$

where $C1D_k$ denotes 1D convolution, w_{eca} is the scale statistics of the channel, and $g(\cdot)$ stands for global average pooling. Then, F_{FE} is rescaled as

$$\hat{F}_{FE} = w_{eca} \cdot F_{FE}, \tag{3}$$

where $\hat{F}_{FE}$ stands for the rescaled feature.

To utilize the informative features from ECA, we apply a residual block to the network. We transform residual block by applying a weighted summation.

$$F_{RECAB} = \frac{w_1}{\epsilon + \sum_{i=1}^{2} w_i} F_{input} + \frac{w_2}{\epsilon + \sum_{i=1}^{2} w_i} \hat{F}_{FE}, \tag{4}$$

where F_{RECAB} is the final extracted feature and w_i is a learnable weight, which is a scalar per feature. By applying ReLU to each w_i, it is ensured that $w_i \geq 0$, and the ϵ value is fixed as 0.00001 to avoid numerical instability. Similar to the interpolation, the values of each weight range from 0 to 1. Since these two weight values are learnable parameters, they find more proper values for producing well-balanced features in every training process. From the process, the informative feature is generated by considering the interdependencies among feature channels, thereby providing connectivity among channels and discriminative ability in network.

2) EDGE PROFILE (EP) MODULE

In SISR, it is important point to maintain structure for high-quality SR images. To consider structural information, we construct an EP module based on the conventional image processing principle. This module extracts the edge profile itself from the systemically organized feature using RECAB. Intuitively, the edge area has a rapid variance in pixels as shown in Fig.3. This means that there are large pixel gradient

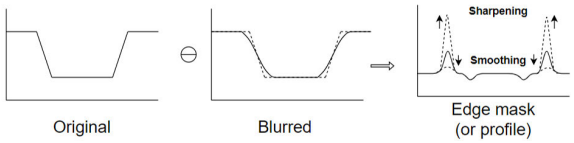


FIGURE 3. Extracting edge profile process.

values in the edge area. Next, the onset and end of discontinuities (e.g. step and ramp discontinuities) in the image are described as edge areas. To extract the edge profile of an image, we consider utilizing the discontinuous property of the edge. As shown in Fig.3, the blurred image was subtracted from the original image to obtain the edge mask (or profile). Therefore, this process can be formulated as:

$$g(x, y) = f(x, y) - \bar{f}(x, y), \tag{5}$$

where $g(x, y)$, $f(x, y)$ and $\bar{f}(x, y)$ are the edge mask, original image and blurred image, respectively. We convert this process to a deep learning method. First, an image was generated from the feature F_{RECAB}, which comes from feature extraction, using one convolution layer.

$$I_{blockSR} = H_{blockSR}(F_{RECAB}), \tag{6}$$

where $I_{blockSR}$ is an image produced from feature F_{RECAB}, and $H_{blockSR}(\cdot)$ can be denoted as image reconstruction in the EP module, which generates an RGB-channel image from the 64-channel feature. To form a blurred image, we transfer the arithmetic mean filter concept using average pooling. Let's denote S_{xy} as the set of coordinates in a rectangular subimage window of size $m \times n$ where the center point is (x, y). Then, this filter computes the average values of the original image $i(x, y)$ in the area defined by S_{xy}. In other words,

$$\hat{i}(x, y) = \frac{1}{mn} \sum_{(s,t) \in S_{xy}} i(s, t), \tag{7}$$

where $\hat{i}(x, y)$ is a blurred image of $i(x, y)$. Based on this operation, if one defines the window size as 3×3, it can also be an average pooling operation. Thus, a blurred image is formed by using it.

$$I_{blockblur} = H_{blur}(I_{blockSR}), \tag{8}$$

where $I_{blockblur}$ is the blurred image from $I_{blockSR}$, and $H_{blur}(\cdot)$ denotes average pooling with a kernel size of 3×3 and padding margin of 1. From the operation in Eq.5, to obtain an edge profile(or mask), $I_{blockSR}$ is subtracted by $I_{blockblur}$. Then, we apply the ReLU operation on the edge profile(or mask) to obtain the outer line.

$$M = ReLU(I_{blockSR} \ominus I_{blockblur}), \tag{9}$$

where M denotes edge profile(or mask) in the REPB, and $\ominus$ is element-wise subtraction. To guide the edge in the training process, $I_{blockSR}$ is concatenated with M

$$I_{guided} = Concat(I_{blockSR}, M), \tag{10}$$

where I_{guided} and $Concat(\cdot)$ denote a guided image and concatenation operation, respectively. Finally, to generate the feature of the EP module, one convolution layer is used, and then the feature F_{FE} information is provided by using the residual structure.

$$F_{EP} = F_{FE} + H_{EP}(I_{guided}), \tag{11}$$

where F_{EP} stands for the feature from the EP module with a channel size is 64, and $H_{EP}(\cdot)$ denotes the edge profile module of the REPB. By extracting the structural information, we can obtain structure-preserving effects.

3) CONTEXT NETWORK (CN) MODULE

From the EP module, we can obtain informative feature with structural information. These features could be beneficial for preserving the structure. However, this module has limitations in handling high-frequency components, such as the texture and sharpness of the structure. Since the details have frequent pixel variations, they can be difficult to capture. To reveal these contextual components, we construct a CN module.

Inspired by a previous study [36], we design a CN module that is based on dilated convolutions. We apply the CN module following the EP module. As described in the CN part of Fig.1, the CN module consists of four 3×3 dilated convolution networks, whose dilated factors are 1, 2, 4, and 1 in order. To prevent loss of resolution or coverage, we consider the expansion of the receptive field to set up dilated factors exponentially. Intuitively, the CN module can improve the learning of the feature maps by passing them through multiple layers that expose contextual information. Subsequently, the output feature is added to the input feature as a residual block.

$$F_{CN} = F_{EP} + H_{f=1} \circ H_{f=4} \circ H_{f=2} \circ H_{f=1}(F_{EP}), \tag{12}$$

where F_{CN} is the output feature of the CN module, and $H_{f=n}(\cdot)$ denotes a dilated convolution whose dilated factor f is n. Because this operation captures contextual information from the feature of the EP module F_{EP}, EPSR can minimize the loss of texture and recover sharpness. In other words, recovering high-frequency components can be ensured by minimizing the side effects and damages.

C. OBJECTIVE FUNCTIONS

Our EPSR is optimized with the set-up loss functions. Generally, L_1 [20], [21], [32], [37], L_2 [15], [16], [18], [19],and adversarial and perceptual losses [3], [24] have been used in the SR methods. To establish the effect of EPSR, we choose two loss functions L_1 and $L_{gradient}$. As proposed in previous works, we select L_1 to guarantee stable convergence. A given training set with N LR images and their HR counterparts is denoted as $\{I^i_{LR}, I^i_{HR}\}^N_{i=1}$. Then, we can formulate L_1 loss as:

$$L_1 = \frac{1}{N} \sum_{n=1}^{N} ||H_{EPSR}(I^i_{LR}) - (I^i_{HR})||_1 \tag{13}$$

Since our EPSR utilizes diverse features, REPBs generate edge profiles depending on feature information from their feature input. To provide a consistent standard for EP modules in the learning process, we consider a loss function to guide them. Using Sobel filter [38], we can extract the gradient maps of HR and SR and formulate gradient loss function as

$$L_{gradient} = \frac{1}{N} \sum_{n=1}^{N} ||S(H_{EPSR}(I^i_{LR})) - S((I^i_{HR}))||_1, \tag{14}$$

where $S(\cdot)$ is the gradient function based on the Sobel filter [38]. By adding $L_{gradient}$ to L_1, we can achieve an end-to-end network without additional module training. Therefore, the goal of training the EPSR is to optimize the total loss function:

$$L_{Total}(\theta) = L_1 + 10^{-1} L_{gradient}, \tag{15}$$

where θ is the parameter set of EPSR. We set the coefficient as 10^{-1} empirically. The loss function is optimized by ADAM gradient descent algorithm.

III. EXPERIMENT RESULTS

A. SETTINGS

We state the settings of experiment about datasets, degradation models, evaluation, and training settings.

1) DATASETS

Following [20]–[22], we set up 800 high resolution images from DIV2K dataset [39] as a training set. For testing, we use 5 standard benchmark datasets: Set5 [40], Set14 [41], B100 [42], Urban100 [43], and Manga109 [44].

2) DEGRADATION MODELS

In order to prove the effectiveness of our EPSR, we use 3 degradation models to generate LR images. First, we generate LR images with scaling factor $\times 2$, $\times 3$, $\times 4$ by using Bicubic Interpolation(BI) operation. BI degradation model are generally used for checking the effect of super resolution as standard. Second, by using Gaussian kernel of size 7×7 with standard deviation 1.6, we blur HR image and downsample it with scaling factor $\times 3$. We denote this process as BD [45]. BD degradation model is utilized to verify the resolution ability on blurry problem. At last, we downsample

(a) RECAB (b) RECAB + EP (c) RECAB + CN (d) RCAB+EP+CN (e) RECAB+EP+CN

FIGURE 4. Ablation study with Bicubic(BI) degradation(×4) on "img_033" from Urban100.

TABLE 1. Comparisons of models with different components. The best results are highlighted.

	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
RECAB	32.26/0.8937	28.46/0.7802	27.26/0.7327	26.29/0.7934	30.55/0.9017
RECAB+EP	32.20/0.8932	28.51/0.7823	27.34/0.7356	26.31/0.7943	30.66/0.9062
RECAB+CN	31.44/0.8786	22.75/0.5959	21.33/0.5206	19.40/0.5848	27.05/0.8274
RECAB+EP+CN	**32.28/0.8945**	**28.55**/0.7828	**27.34/0.7362**	**26.43/0.7983**	**30.82/0.9084**
RCAB+EP+CN	32.25/0.8939	28.53/**0.7831**	27.32/0.7349	26.39/0.7975	30.73/0.9076

HR image with scaling factor ×3 using bicubic interpolation and then add Gaussian noise with level 30. This process is denoted as DN for short. DN degradation model is utilized to prove the denoising ability.

3) EVALUATION METRICS

The SR results are evaluated with PSNR and SSIM [46] on Y channel(i.e. luminance) of YCbCr space.

4) TRAINING SETTINGS

In training process, the training images are augmented by randomly rotating 90°,180°,270°, and horizontally flipping. In each training batch, 8 LR color patches with size 48×48 are extracted as input. Our model is trained by ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 1e-8$. We set learning rate as 10^{-4} initially and then it is reduced to half every 200 epochs. We implement our proposed EPSR using Pytorch [47] on a Tesla V100 GPU.

B. ABLATION STUDY

As discussed above, our EPSR concentrates on the structure preservation and representation details. To demonstrate the effectiveness of EPSR, we focus on showing the influence of the EP and CN modules, which could affect the quality of SR results. Therefore, we set three comparisons by decomposing REPB, and two comparisons are performed by feature extractions based on RECAB or RCAB according to previous research [22].

First, to establish a criterion, we construct a basic block without an EP module and a CN module. In other words, by using only RECAB, SR images are generated directly. As shown in Fig.4, only using RECAB is effective in representing texture information. However, it is difficult to recover the image details and edge components. We proceed with an experiment by connecting the EP module to feature extraction to examine the effect of the edge profile. Even though the edge profile is only provided the on network, we can check the enhancement of image reconstruction in aspect of structure preservation. Subsequently, by adding a CN, we build a full REPB. As explained in section II, the CN helps to capture hidden information that includes image details. In Fig.4 (e), which is generated by EPSR, edge and texture information are reconstructed more stable than two images.

In terms of PSNR and SSIM evaluations(See Table.1) on all datasets, utilizing edge properties brings significant overall benefits in each evaluations. This implies that the EP module is helpful in preserving the image structure in the reconstruction process as shown in the SSIM evaluations. Furthermore, by exploiting contextual information as image details such as texture and edge, the CN module shows a synergistic effect with the EP module. As a result, the efficacy of the EP and CN modules is verified in image results and numerical value evaluations. Furthermore, when we remove the EP module in our EPSR, some problems occur in recovering texture and edge information like as shown in Fig.4 (c). This shows that, even if the CN module provides benefits to capture contextual information, it could experience difficulty in exploiting the overall features, which can be checked in the numerical value evaluation. This indicates the rationality of plugging the CN module into the combination of RECAB and EP modules because it concentrates on capturing contextual information that contains image details and not tendency of features.

Additional experiments are conducted to investigate the relationship between the EP module and feature extraction. In our EPSR, we choose to use ECA for extracting features in RECAB. To verify the effect of this, we conduct an experiment by substituting ECA for feature extraction with Channel Attention(CA) taken from a previous research [22]. This substitution is RCAB. As shown in Fig.4 (d), the EPSR based on RCAB generates a good SR image. However, the direction of the edge lines is incorrect. As mentioned in section II, since CA has problems with channel predictions, it generates unclear features, and it seems that the EP module has

TABLE 2. Quantitative results with BI degradation model. Highlight stands for the best performance, red indicates the second, and blue is the third.

Method	scale	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×2	33.66/0.9229	30.24/0.8688	29.560.8431	26.88/0.8403	30.80/0.9339
SRCNN [15]	×2	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
DEGREE [12]	×2	37.40/0.9580	32.96/0.9115	31.73/0.8937	- / -	- / -
VDSR [16]	×2	37.53/0.9587	33.05/0.9127	31.90/0.8960	30.77/0.9141	37.16/0.9740
LapSRN [37]	×2	37.52/0.9591	32.99/0.9124	31.80/0.8949	30.41/0.9101	37.53/0.9740
EDSR [20]	×2	37.99/0.9587	33.57/0.9175	32.16/0.8994	31.98/0.9272	39.10/0.9773
MemNet [19]	×2	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740
IDN [48]	×2	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.02/0.9749
SRMDNF [45]	×2	37.79/0.9601	33.32/0.9159	32.05/0.8985	31.33/0.9204	38.07/0.9761
CARN [49]	×2	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9764
RDN [21]	×2	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353	39.18/0.9780
RCAN [22]	×2	38.27/0.9614	34.12/0.9216	**32.41**/0.9027	33.34/0.9384	39.44/0.9786
EPSR(Ours)	×2	**38.29/0.9618**	**34.13/0.9227**	32.38/**0.9046**	**33.36/0.9401**	**39.57/0.9788**
Bicubic	×3	30.40/0.8686	27.54/0.7741	27.21/0.7389	24.46/0.7349	26.95/0.8556
SRCNN [15]	×3	32.75/0.9090	29.29/0.8215	28.41/0.7863	26.24/0.7991	30.48/0.9117
DEGREE [12]	×3	33.39/0.9182	29.61/0.8275	28.63/0.7921	- / -	- / -
VDSR [16]	×3	33.66/0.9213	29.78/0.8318	28.83/0.7976	27.14/0.8279	32.01/0.9340
LapSRN [37]	×3	33.82/0.9227	29.79/0.8320	28.82/0.7973	27.07/0.8271	32.21/0.9350
EDSR [20]	×3	34.37/0.9270	30.28/0.8418	29.09/0.8052	28.15/0.8527	34.17/0.9476
MemNet [19]	×3	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369
IDN [48]	×3	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.69/0.9378
SRMDNF [45]	×3	34.12/0.9254	30.04/0.8382	28.97/0.8025	27.57/0.8398	33.00/0.9403
CARN [49]	×3	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.49/0.9440
RDN [21]	×3	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653	34.13/0.9484
RCAN [22]	×3	**34.74/0.9299**	**30.65**/0.8482	**29.32**/0.8111	**29.09/0.8702**	34.44/**0.9499**
EPSR(Ours)	×3	34.73/0.9297	30.52/**0.8491**	29.15/**0.8139**	28.96/**0.8702**	**34.46**/0.9486
Bicubic	×4	28.43/0.8109	26.00/0.7023	25.96/0.6678	23.14/0.6574	25.15/0.7890
SRCNN [15]	×4	30.48/0.8628	27.50/0.7513	26.90/0.7103	24.52/0.7226	27.66/0.8580
DEGREE [12]	×4	31.03/0.8761	27.73/0.7597	27.07/0.7177	- / -	- / -
VDSR [16]	×4	31.35/0.8838	28.02/0.7678	27.29/0.7252	25.18/0.7525	28.82/0.8860
LapSRN [37]	×4	31.54/0.8866	28.09/0.7694	27.32/0.7264	25.21/0.7553	29.09/0.8900
EDSR [20]	×4	32.09/0.8938	28.58/0.7813	27.57/0.7357	26.04/0.7849	31.02/0.9148
MemNet [19]	×4	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	29.42/0.8942
IDN [48]	×4	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.40/0.8936
SRMDNF [45]	×4	31.96/0.8925	28.35/0.7787	27.49/0.7337	25.68/0.7731	30.09/0.9024
SRGAN [25]	×4	32.05/0.8910	28.53/0.7804	27.57/0.7354	26.07/0.7839	- / -
NatSR [50]	×4	32.20/0.8939	28.54/0.7808	27.60/0.7366	26.21/0.7904	- / -
SPSR [13]	×4	31.52/0.8827	27.74/0.7828	27.21/0.7276	24.80/0.8021	30.12/0.9037
CARN [49]	×4	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.40/0.9082
RDN [21]	×4	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
RCAN [22]	×4	**32.63/0.9002**	**28.87/0.7889**	**27.77/0.7436**	**26.82/0.8087**	**31.22/0.9173**
EPSR(Ours)	×4	32.42/0.8969	28.65/0.7867	27.45/0.7403	26.64/0.8038	31.16/0.9127

difficulty finding the right edge lines. In contrast, EPSR based on ECA feature extraction reconstructs the edge and texture successfully. It is revealed visually in the results generated by EPSR based on ECA feature extraction and in numerical value evaluations on PSNR and SSIM. This indicates that the formation of proper features is important key for extracting the correct edge profile to preserve the structure in SR.

C. RESULT WITH BI DEGRADATION

1) QUANTITATIVE COMPARISON

To compare the effectiveness of our network with other methods, we investigate 14 state-of-the-art SR methods including general deep learning models, perceptual-driven models and edge-related models: SRCNN [15], DEGREE [12], VDSR [16], LapSRN [37], EDSR [20], MemNet [19], IDN [48], SRMDNF [45], CARN [49], RDN [21], RCAN [22], SRGAN [25], NatSR [50], SPSR [13]. All the quantitative comparisons for ×2, ×3 and ×4 SR are shown in Table.2. With rich texture information datasets, such as Set5, Set14, and BSD100, our EPSR obtains better SSIM results compared with other networks. NatSR obtains high results, and shows limitations on the BSD100 dataset specifically. However, our EPSR shows very well-balanced results compared with NatSR and achieves high performance on all datasets. Furthermore, in PSNR, it obtains comparable results with those of RCAN and RDN, whose main

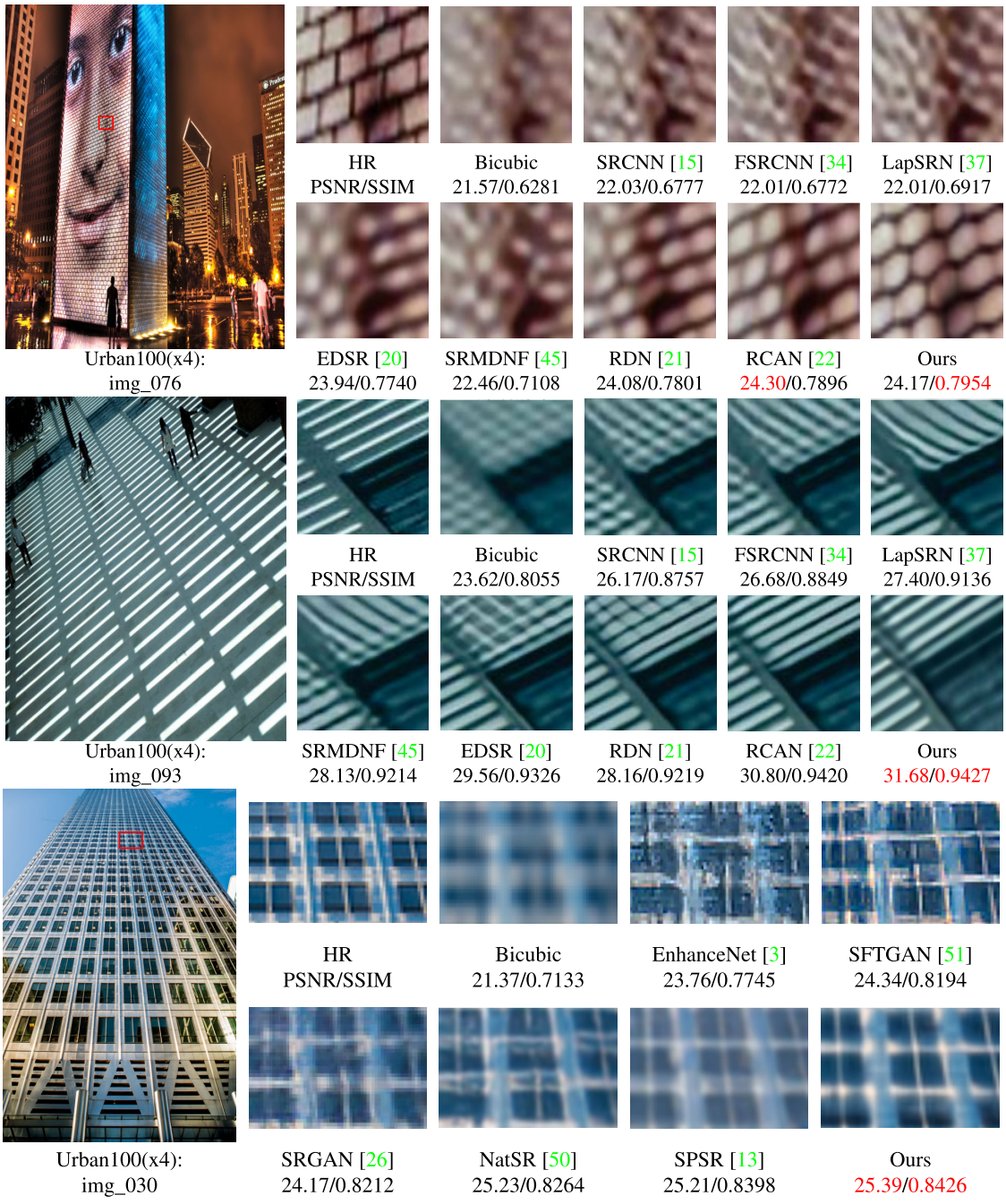


FIGURE 5. Visual Comparison for SR(×4) with BI model on Urban100. The best results are highlighted.

target is the PSNR evaluation metric. In the Urban100 and Manga109 datasets that contain rich repeated edge information, our EPSR achieves competitive results in both PSNR and SSIM. Subsequently, we compare our EPSR with SPSR and DEGREE, which utilize structure information in the SR method. They depend on artificial edge extraction and exhibit good improvement in structure preservation. However, the results are inferior to those of our EPSR. Overall, our EPSR shows high or competitive performance in terms of the PSNR and SSIM evaluation metrics.

2) QUALITATIVE COMPARISON

We present visual comparison on scale ×4. Fig.5 shows that the EPSR results are stronger in preserving the structure and recovering texture than those of other methods. For "img_076" and "img_093", we can observe that most of the compared models cannot reconstruct the lattices and have trouble with blurring effects. Other methods generate twisted lines and squashed lattices. However, EPSR shows strength in recovering structural properties. We can see the capabilities of capturing the structural characteristics of objects in an

TABLE 3. Quantitative results with BD degradation model.

Method	Scale	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×3	28.78/0.8308	26.38/0.7271	26.33/0.6918	26.88/0.8403	25.46/0.8149
SRMSR [52]	×3	32.21/0.9001	28.89/0.8105	28.13/0.7740	25.84/0.7856	29.64/0.9003
SRCNN [15]	×3	32.05/0.8944	28.80/0.8074	28.13/0.7736	25.70/0.7770	29.47/0.8924
FSRCNN [34]	×3	26.23/0.8124	24.44/0.7106	24.86/0.6832	22.04/0.6745	23.04/0.7927
VDSR [16]	×3	33.25/0.9150	29.46/0.8244	28.57/0.7893	26.61/0.8136	31.06/0.9234
IRCNN_G [53]	×3	33.38/0.9182	29.63/0.8281	28.65/0.7922	26.77/0.8154	31.15/0.9245
IRCNN_C [53]	×3	33.17/0.9157	29.55/0.8271	28.49/0.7886	26.47/0.8081	31.13/0.9236
SRMDNF [45]	×3	34.01/0.9242	30.11/0.8364	28.98/0.8009	27.50/0.8370	32.97/0.9391
RDN [21]	×3	34.58/0.9280	30.53/0.8447	29.23/0.8079	28.46/0.8582	33.97/0.9465
RCAN [22]	×3	**34.70/0.9288**	**30.63**/0.8462	**29.32**/0.8093	28.81/0.8645	34.38/**0.9483**
EPSR(Ours)	×3	34.68/**0.9288**	30.56/**0.8484**	29.14/**0.8130**	**28.83/0.8667**	**34.51**/0.9476

TABLE 4. Quantitative results with DN degradation model.

Method	Scale	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×3	24.01/0.5369	22.87/0.4724	22.92/0.4449	21.63/0.4687	23.01/0.5381
SRCNN [15]	×3	25.01/0.6950	23.78/0.5898	23.76/0.5538	21.90/0.5737	23.75/0.7148
FSRCNN [34]	×3	24.18/0.6932	23.02/0.5856	23.41/0.5556	21.15/0.5682	22.39/0.7111
VDSR [16]	×3	25.20/0.7183	24.00/0.6112	24.00/0.5749	22.22/0.6096	24.20/0.7525
IRCNN_G [53]	×3	25.70/0.7379	24.45/0.6305	24.28/0.5900	22.90/0.6429	24.88/0.7765
IRCNN_C [53]	×3	27.48/0.7925	25.92/0.6932	25.55/0.6481	23.93/0.6950	26.07/0.8253
RDN [21]	×3	28.47/0.8151	**26.60**/0.7101	**25.93**/0.6573	24.92/0.7364	28.00/0.8591
EPSR(Ours)	×3	**28.53/0.8142**	26.57/**0.7105**	25.86/**0.6588**	**25.16/0.7477**	**28.20/0.8634**

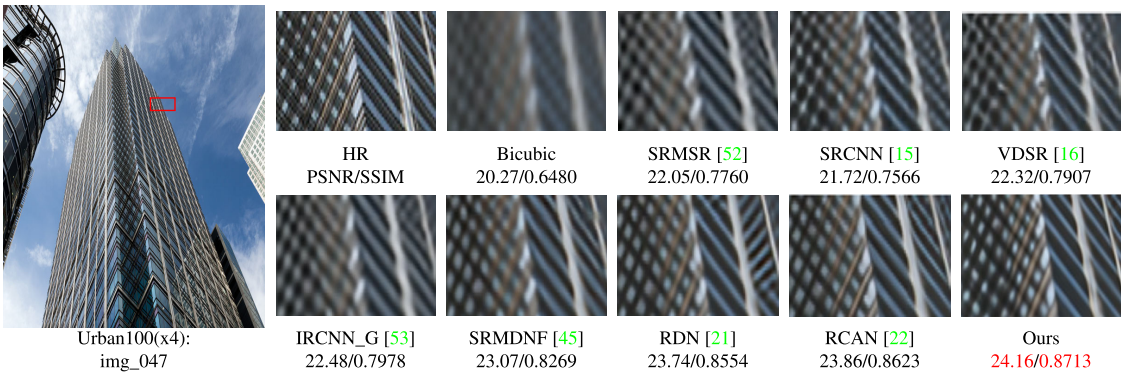


FIGURE 6. Visual Comparison for SR(×3) with BD model on Urban100. The best results are highlighted.

image, which contributes to preserving the structural information in the image. Our EPSR captures image details well, including high-frequency components. In "img_030" our EPSR shows a clear structure in images without damage and distortion, while most of the other methods fail to reconstruct the fine appearance of the objects. The qualitative comparison verifies that our EPSR generates a geometrically more stable image for perceptions by utilizing structural information extracted autonomously and exploiting contextual components.

D. RESULT WITH BD AND DN DEGRADATION

1) QUANTITATIVE COMPARISON

We apply our EPSR with the BD degradation model, which was used in a recent [22], Following other research [21], we further compare various SR methods on images with the DN degradation model. We compare our EPSR with 8-state-of-the-art SR methods with ×3 scaling factors: SRMSR [52], SRCNN [15], FSRCNN [34], VDSR [16], IRCNN [53], SRMDNF [45], RDN [21], and RCAN [22]. In Table.3 and Table.4, all the results are presented explicitly. We can observe that our EPSR exhibits a higher performance than the other methods, indicating that our EPSR is an effective method for various types of degradation models.

2) QUALITATIVE COMPARISON

We also show visual comparisons for the challenging problem of blurring(BD) and noising(DN) degradation. First, in BD, there are difficulties in restoring the definite texture and structural information. Fig.6 shows that the results of EPSR are clearer and more natural than those of the other methods. Most methods suffer from heavy blurring problem, but

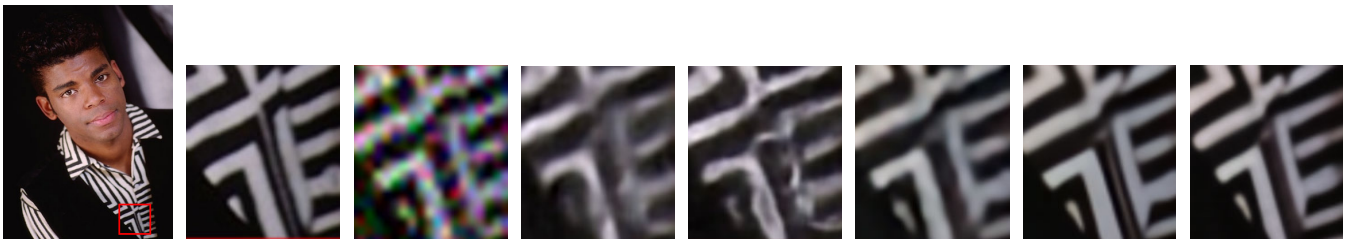


FIGURE 7. Visual Comparison for SR(×3) with DN model on BSD100. The best results are highlighted.

our EPSR recovers texture more clearly than other methods. In particular, the structure in our results is well-preserved without serious distortions. In DN, since there is a heavy loss of information in the LR, it is difficult to reconstruct an ordinary image. In Fig.7, because of heavy damages of the input, other methods have difficulties in overcoming a lack of information and restoring distortions. However, EPSR can restore edge information while preventing texture loss.

The results indicates that our EPSR can cope with damage and distortion of texture and structure by utilizing diverse information effectively. Our EPSR alleviates these problems significantly and can reconstruct more details compared with those reconstructed by other methods.

IV. CONCLUSION

In this paper, we propose Edge Profile Super Resolution (EPSR) method to preserve the structural information and to restore texture in SISR. We construct EPSR by building modified-Fractal Residual Network (mFRN) structures hierarchically and repeatedly. mFRN is composed of residual Edge Profile Blocks (REPBs) consisting of Residual Efficient Channel Attention Block (RECAB), Edge Profile (EP), and Context Network (CN) modules. RECAB generates more informative features with high-frequency components. From this feature, EP module produces structure-informed features by generating the edge profile itself. Finally, the CN module captures details by exploiting high-frequency information, such as texture and structure with proper sharpness. By repeating the procedure in mFRN structure, our EPSR can extract high-fidelity features, thus preventing texture loss and preserving the structure with appropriate sharpness. Because our EPSR considers texture loss and structural information by applying conventional principle to the deep learning method, high-quality results are obtained. Extension experiments on SR with BI, BD, and DN degradation models demonstrate the effectiveness of our EPSR.

REFERENCES

[1] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.
[2] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.
[3] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4491–4500.
[4] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: http://arxiv.org/abs/1710.10196
[5] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, A. M. S. M. de Marvao, T. Dawes, D. O'Regan, and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2013, pp. 9–16.
[6] M. Ebrahimi and E. R. Vrscay, "Solving the inverse problem of image zooming using 'self-examples,'" in *Proc. Int. Conf. Image Anal. Recognit.* Springer, 2007, pp. 117–130.
[7] R. Fattal, "Image upsampling via imposed edge statistics," in *Proc. ACM SIGGRAPH Papers (SIGGRAPH)*, 2007, p. 95.
[8] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1529–1542, Jun. 2011.
[9] Q. Yan, Y. Xu, X. Yang, and T. Q. Nguyen, "Single image superresolution based on gradient profile sharpness," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3187–3202, Oct. 2015.
[10] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2400–2407.
[11] Y. Zhu, Y. Zhang, B. Bonev, and A. L. Yuille, "Modeling deformable gradient compositions for single-image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5417–5425.
[12] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, and S. Yan, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Dec. 2017.
[13] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7769–7778.
[14] K. Nazeri, H. Thasarathan, and M. Ebrahimi, "Edge-informed single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3275–3284.
[15] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
[16] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[17] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.
[18] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.
[19] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.
[20] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
[21] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[22] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[23] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11065–11074.
[24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 694–711.
[25] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
[26] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 63–79.
[27] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.
[28] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and J.-P. Thiran, "SROBB: Targeted perceptual loss for single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2710–2719.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[30] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun./Jul. 2004, pp. I–I.
[31] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020.
[32] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.
[33] J. Kwak and D. Son, "Fractal residual network and solutions for real super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2114–2121.
[34] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 391–407.
[35] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[36] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[37] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.
[38] J. R. Parker, *Algorithms for Image Processing and Computer Vision*. Hoboken, NJ, USA: Wiley, 2010.
[39] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2017, pp. 114–125.
[40] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
[41] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.* Springer, 2010, pp. 711–730.
[42] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, vol. 2, Jul. 2001, pp. 416–423.
[43] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.
[44] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based Manga retrieval using Manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, 2017.
[45] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.
[46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[47] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, A. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," 2017.
[48] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.
[49] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 252–268.
[50] J. W. Soh, G. Y. Park, J. Jo, and N. I. Cho, "Natural and realistic single image super-resolution with explicit natural manifold discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8122–8131.
[51] Y. Zhang, X. Li, and J. Zhou, "SFTGAN: A generative adversarial network for pan-sharpening equipped with spatial feature transform layers," *J. Appl. Remote Sens.*, vol. 13, no. 2, 2019, Art. no. 026507.
[52] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.
[53] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

JIUN LEE (Graduate Student Member, IEEE) was born in Seoul, South Korea, in 1994. He received the B.S. degree from Sungkyunkwan University (SKKU), Suwon, South Korea, in 2019, where he is currently a graduate student from 2019 with the College of Electrical and Computer Engineering. His research interests include computer vision, image processing which are related to super-resolution, optical flow, and 3-D cloud point.

INYONG YUN received the B.S. degree in ocean engineering from Jeju National University, South Korea, in 2010, and the Ph.D. degree in electronic engineering from Sungkyunkwan University, South Korea, in 2020. He is currently a Senior Researcher with Hana Institute of Technology, South Korea. His main research interests include image and video processing, computer vision, and machine learning.

JAEKWANG KIM (Member, IEEE) was born in Seoul, South Korea, in 1980. He received the B.S., M.S., and Ph.D. degrees from Sungkyunkwan University (SKKU), Suwon, South Korea, in 2004, 2006, and 2014, respectively. He is currently an Assistant Professor with the College of Computing and Informatics, SKKU. His research interests include recommender systems, machine learning algorithm, and intelligent systems. He was elected as a Bronze Medal Superior Student from the Brain Korea 21 Enterprise Department, Sungkyunkwan University, and received the Best Presentation Paper Award at ICUIMC, Suwon, in January 2009.