IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Dual-streams Edge Driven Encoder-decoder Network for Image Super-Resolution

FENG LI[1, 2], HUIHUI BAI[1, 2], LIJUN ZHAO[1, 2] , and YAO ZHAO[1, 2], (Senior Member, IEEE)
[1]Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[2]Beijing Key Laboratory of Advanced Information Science and Network Technoloy, Beijing 100044, China

Corresponding author: Huihui Bai (e-mail: hhbai@bjtu.edu.cn).

This work was supported in part by National Natural Science Foundation of China (No. 61672087), Key Innovation Team of Shanxi 1331 Project (KITSX1331), and the Fundamental Research Funds for the Central Universities (2017YJS050).

ABSTRACT Single image super-resolution (SR) aims at reconstructing high-resolution (HR) images from low-resolution (LR) ones. One of the most key issues is to recover finer image details of LR images. In this paper, considering the importance of edge prior for image SR, we propose a dual-streams edge driven encoder-decoder network (Dual-EEDN), which combines edge stream based encoder-decoder network (edge-EDN) and color stream based encoder-decoder network (color-EDN) to reconstruct HR images with more image details. Instead of utilizing two sub-networks to learn edge information and color image contents respectively, a multitask learning framework is developed to jointly train edge-EDN and color-EDN. Therefore, as the structure prior, the reconstructed HR edge maps are fused with learned features of color stream to refine the HR color images. To reconstruct HR images with better visual quality, a total loss function combining edge loss and color loss is designed to make an optimal trade-off between the image fidelity and texture details. Our extensive benchmark evaluations demonstrate that our method for SR performs better both on high objective quality and human visual perception compared with several state-of-the-art SR methods.

INDEX TERMS Dual-streams, edge-driven, image super-resolution (SR), edge stream, color stream.

I. INTRODUCTION

SINGLE image super-resolution (SR) is a classic problem in computer vision, which aims at reconstructing the corresponding high resolution (HR) images from observed low resolution (LR) images. Since image SR can overcome the limitations of image resolution in a small scale, it has been widely used in many applications, such as medical image analysis [1], video surveillance [2], and face hallucination [3]– [4].

Since a multiplicity of methods could generate same LR images from HR ones, most SR methods resolve the typically ill-posed problem with various strong priors. Considering that image patches are a set of data with multiview characteristics and spatial organization, Yang *et al.* [5] propose a dual-geometric neighbor embedding (DGNE) approach for single image SR. In [6], a general framework is proposed for "blind" super-resolution, which exploits the inherent recurrence property of small patches across scales of LR image. Wang *et al.* [8] introduce an additional edge constraint to reduce the undesired artifacts brought by the traditional interpolation algorithm. Zhang *et al.* [9] propose an image SR method by learning both non-local and local regularization priors from a given LR image. In order to achieve simple and efficient performance for real-time SR applications, Li *et al.* [10] propose an edge-directed interpolation algorithm for natural images, which estimates the local covariance coefficients from an LR image and then uses these coefficients to adapt the interpolation at an HR image. To deal with the image blurring caused by fixed weights for interpolation, two adaptive interpolation methods are proposed in [11]– [12].

In view of the well performance by using internal or external data to guide image restoration, most state-of-the-art methods adopt the example-based strategy, which learns the correspondence between LR and HR image patches from a huge database. In [13], based on the observation that patches in a natural image redundantly recur many times inside the image, Glasner *et al.* introduce a unified framework for single image SR. In [14], Freedman *et al.* propose a high-

quality and efficient single image upscaling technique which follows a local self-similarity assumption on natural images and extracts patches from extremely localized regions in input images. By investigating the application of the clustered sparse coding scheme into the SR problem, Yang *et al.* [15] propose a multiple-geometric-dictionaries-based clustered sparse coding scheme for image SR. Huang *et al.* [16] extend the self-similarity based SR, which expands the internal patch searching space by allowing geometric variations and incorporating additional affine transformations to accommodate local shape variations. Timofte *et al.* [17] propose an adjusted anchored neighborhood regression method, which solves the problem of image upscaling in the form of single image SR based on a dictionary of LR and HR exemplars. Different from the existing methods that figure out the whole gradient profile structure and locate the edge points, Song *et al.* [18] propose a new approach which sharpens the gradient field adaptively only based on the pixels in a small neighborhood. Besides, there are still other external example-based methods that focus on learning the dictionaries or mapping functions [19]– [21].

Recently, deep convolutional neural networks (CNNs) have shown great performance in computer vision field. Due to the powerful learning ability, CNNs are widely used to tackle the image restoration tasks, such as image denoising [22], JPEG deblocking [23], and image inpainting [24]. By directly learning an end-to-end mapping between LR and HR images, Dong *et al*. [25] propose a deep CNN for single image super-resolution (SRCNN), which achieves a well trade-off between performance and speed. The authors further redesign the SRCNN and introduce a compact hourglass-shape structure (FSRCNN) for faster and better SR [26]. However, the two models both fail to obtain superior performance in very deep networks.

Inspired by VGG-net [27], Kim *et al.* [28] propose a 20-layers CNN (VDSR) for multi-scale factor image SR, which requires plenty of parameters. In [29], a deeply-recursive convolutional network (DRCN) is presented to improve the performance without introducing new parameters for additional convolutions. Zeng *et al.* [30] develop a data driven model coupled deep autoencoder (SRCDA) for single image SR, which learns the intrinsic representations of LR and HR image patches and corresponds the LR representations to their HR representations through a mapping function. Motivated by the promising performance of autoencoder based image denoising, Mao *et al.* [31] propose a very deep full convolutional encoder-decoder framework for image restoration, which symmetrically links the convolutional and deconvolutional layers with skip connections. Although the aforementioned CNN based models have achieved great performance in accuracy and speed of single image SR, they do not take the importance of natural image priors into consideration in neural networks, which is important to recover finer image details.

In addition, multitask learning is an approach to learn tasks in parallel while using a shared representation, what is learned for each task can help other tasks to be learned better. In image SR, Liang *et al.* [32] present a multitask learning framework based on the deep neural network for image SR, which jointly considers the image SR process and the image degeneration process. By sharing parameters between the two highly relevant tasks, the framework improves the obtained neural network-based mapping model between HR and LR patches. In [33], Zhao *et al.* propose a color-depth conditional generative adversarial network (CDcGAN) to concurrently resolve the problem of depth SR and color SR in 3D videos, which adopts the mutual information of color image and depth image to enhance each other in consideration of the geometry structural dependency of color-depth image in the same scene.

Considering that the edge knowledge can contribute to producing sharp edges and compensate the high-frequency details of reconstructed HR images, we embed the edge prior into a deep network for image SR. Besides, in the light of the performance of multitask learning framework in image SR, we propose a dual-streams edge driven encoder-decoder network (Dual-EEDN), which utilizes the edge prior and color image contents simultaneously to reconstruct an HR image with better human visual perception. In this work, we fully utilize the HR edge information for single image SR, where two types of image components are jointly learned, *i.e.*, the basic color image contents, and the extracted edge contents. Instead of utilizing two sub-networks to learn edge contents and color image contents respectively, the edge contents are firstly pre-trained by edge streams based encoder-decoder networks (edge-EDN), then color stream based encoder-decoder network (color-EDN) is imposed to learn color image contents. Furthermore, the reconstructed HR edge maps are fused with color images which are predicted from color-EDN to recover HR images with well image details. Due to the performance degradation problem caused by very deep networks, the skip connections in edge-EDN and color-EDN are used for passing information from previous layers to bottom layers, which also make train the deep network easier. In order to reconstruct HR images with better visual quality, we investigate a novel total loss function combining the edge loss and color loss for the best trade-off to balance the color image contents and edge contents. Extensive experiments on four benchmark datasets demonstrate that Dual-EEDN outperforms several state-of-the-art methods on both quantitative metrics and image details.

The remainder of this paper is organized as follows. Dual-EEDN is presented in detail in Section II. The experimental results and comparisons with other state-of-the-art methods are demonstrated in Section III. The conclusion of this paper and future research work are presented in Section IV.

II. PROPOSED METHOD

In this section, we present each component of our framework in detail. As shown in Figure 1, our proposed Dual-EEDN consists of five parts: a feature extraction network (FENet), edge stream based encoder-decoder network (edge-EDN),

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access

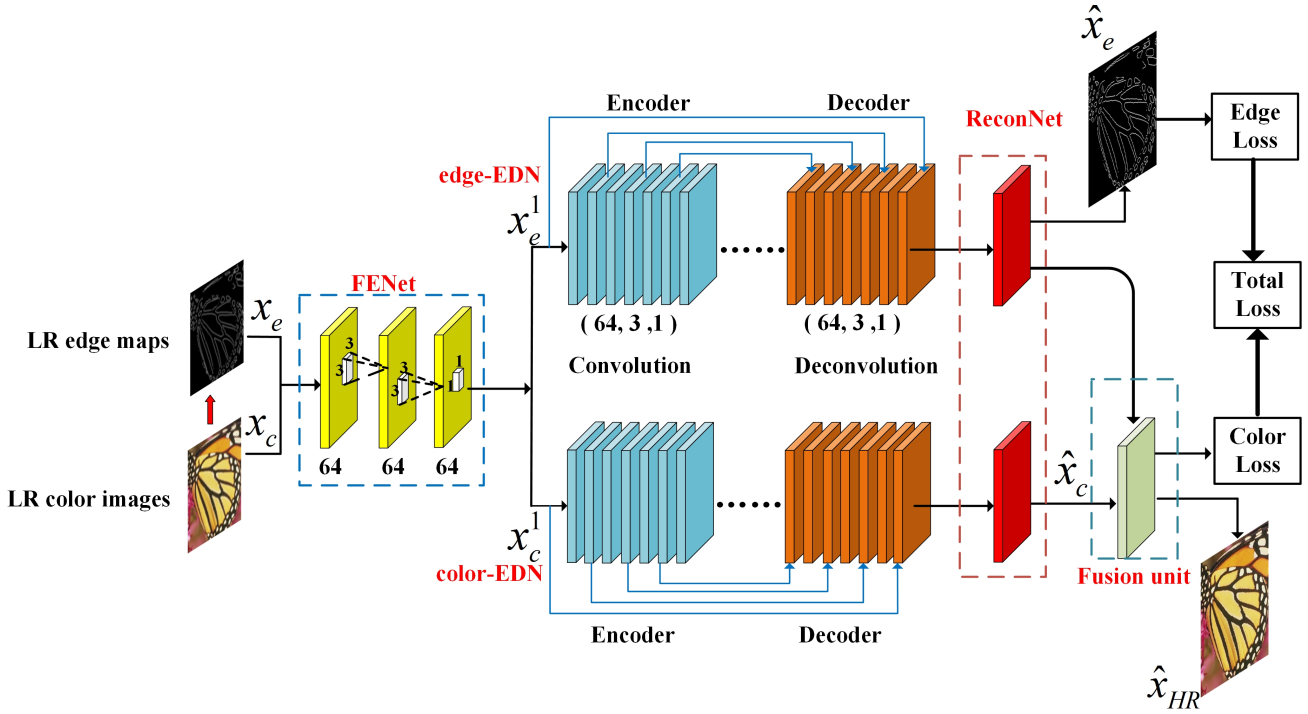


FIGURE 1: The architecture of our dual-streams edge driven encoder-decoder network for single image super resolution. It consists of five parts: feature extraction network (FENet) which includes three convolutional layers for feature extraction, edge-EDN and color-EDN are used for predicting the edge details and color contents respectively, a reconstruction network (ReconNet) in each stream, and a fusion unit. (64, 3, 1) represent the 64 chanels and 3x3 kernel size with stride 1.

color stream based encoder-decoder network (color-EDN), two reconstruction networks (ReconNet) in edge-EDN and color-EDN, and a fusion unit for combining the edge contents and color contents to jointly train our Dual-EEDN.

A. FEATURE EXTRACTION

In image SR, the degradation process of a LR image x from the HR image $\tilde{x}$ can be formulated as

$$x = D(\tilde{x}) + n. \tag{1}$$

where $D(\cdot)$ denotes the degradation function and n is an additive noise. To reconstruct HR images while preserving photo-realistic image details, we combine the benefits of edge-directed SR method and CNN based SR method to super-resolve the ill-posed problem. Different from most CNN based methods directly predicting HR images from LR images, the goal of our Dual-EEDN is to reconstruct HR images based on two inputs, *i.e.*, the LR color image x_c and the LR edge map x_e. Given an input LR image x_c, we firstly employ an edge extractor such as Sobel operator, Canny operator *et al.*, to obtain the edge map as another input x_e. Then the predicted LR edge maps and color images are delivered to the FENet for feature extraction,

$$\begin{aligned} f_{ext}(f_0) &= [x_e^1, x_c^1] \\ f_0 &= [x_e, x_c] \end{aligned} \tag{2}$$

where the input f_0 to Dual-EEDN is a concatenation of the LR image x_e and x_c. $[x_e^1, x_c^1]$ are the extracted features from the LR input $[x_e, x_c]$ by the FENet. And $f_{ext}(\cdot)$ denotes the feature extraction function. Specifically, in FENet, the first two layers with kernel size 3×3 are designed to capture image information and achieve better efficiency. The third layer with 1×1 kernel size is introduced as a bottleneck layer, which is responsible to reduce the feature dimensions, and thus to improve the computational efficiency. In our proposed Dual-EEDN, as shown in Figure 1, to formulate the SR problem by considering both the color image contents and edge contents, the the extracted features $[x_e^1, x_c^1]$ are sent to edge-EDN and color-EDN respectively.

B. EDGE STREAM BASED ENCODER-DECODER NETWORK

We now present our edge-EDN, which is used to learn a direct mapping from the LR patches of edge maps to the desired HR patches. And then the reconstructed HR edge contents are fused with the HR color contents predicted by color-EDN to jointly recover HR images with well texture details. In the encoder of edge-EDN, supposing m convolutional layers are stacked to constantly learn local features and preserve primary components of edge maps,

$$\begin{aligned} y_e^1 &= f_e^m(f_e^{m-1}(...f_e^2(f_e^1(x_e^1))...)) \\ &= f_e^m(f_e^{m-1}(...f_e^2(f_e^1(f_{ext}(x_e)))...)) \end{aligned} \tag{3}$$

where $f_e^m(\cdot)$ denotes the mapping function of the m^{th} convolutional layer, and y_e^1 denotes the learnt edge features through the m convolutional layers. As well known, deconvolution

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access

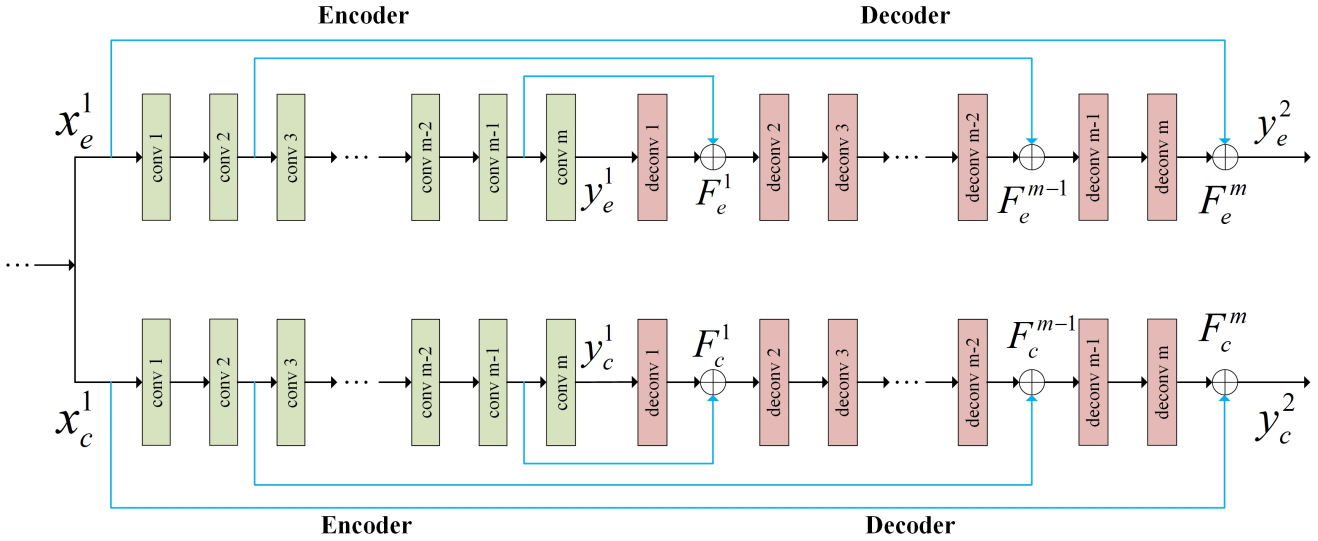


FIGURE 2: **The details of our two streams encoder-decoder networks: edge-EDN (top) and color-EDN (bottom). Each stream consists of m convolutional layers (green) and m deconvolutional layers (light red). Skip connections are used for passing the previous states to current state and back-propagating the gradients to bottom layers.**

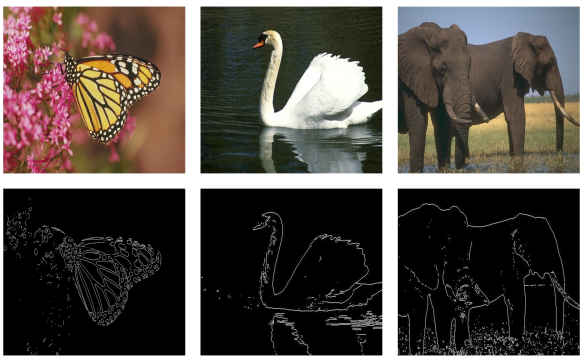


FIGURE 3: **Example images with edge information. Top: original images; bottom: extracted edge maps.**

operation is usually used as upsampling, which can be regarded as an inverse process of convolution. In a sense, supposing that the input stride is s, it can be found that upsampling with factor s is the convolution operation with a fractional input stride of $1/s$. Therefore, the deconvolutional layer is usually adopted to learn the upsampling kernels with an output stride of s. Nevertheless, in edge-EDN, we utilize the same number of deconvolutional layers as decoder to compensate the information of edge maps and make the feature maps reach to a finer level,

$$y_e^2 = d_e^m(d_e^{m-1}(...d_e^2(d_e^1(y_e^1))...)) \tag{4}$$

where d_e^m denotes the deconvolution operation of the m^{th} deconvolutional layer, and y_e^2 is the output of edge-EDN. To keep the output size of encoder same with decoder, we use the same kernel size 3×3 and 64 channels in both encoder and decoder. At the end of edge-EDN, a convolutional layer with kernel size 3×3 is used as ReconNet to reconstruct HR edge maps. The basic reconstruction of HR edge maps can be represented as

$$\begin{aligned} \hat{x}_e &= f_{recon}(y_e^2) + x_e \\ &= f_{recon}(d_e^m(d_e^{m-1}(...d_e^2(d_e^1(y_e^1))...))) + x_e \end{aligned} \tag{5}$$

where $f_{recon}(\cdot)$ denotes the reconstruction function of ReconNet, and $\hat{x}_e$ is the recovered HR edge map. However, with the network depth increasing, adding more layers can cause the input information and gradient information weaken, which will make the performance of the model degrade rapidly. In our framework, to solve this problem, as shown in Figure 2, we not only adopt deconvolutional layers to compensate the details but also employ skip connections to correspond the convolutional features to deconvolutional features to pass previous states to current state, which can ensure the sufficient information flow between the layers in the network. Furthermore, the skip connections also contribute to back-propagating the gradient to bottom layers that make train very deep networks easier. For m convolutional layers and deconvolutional layers in edge-EDN, the output after m^{th} skip connection F_e^m is

$$\begin{aligned} y_e^2 =& F_e^m(y_e^1) \\ =& d_e^m(d_e^{m-1}(F_e^{m-1}(y_e^1))) + x_e^1 \\ =& d_e^m(d_e^{m-1}(...(d_e^2(F_e^1(x_e^1)))...)) + x_e^1 \end{aligned} \tag{6}$$

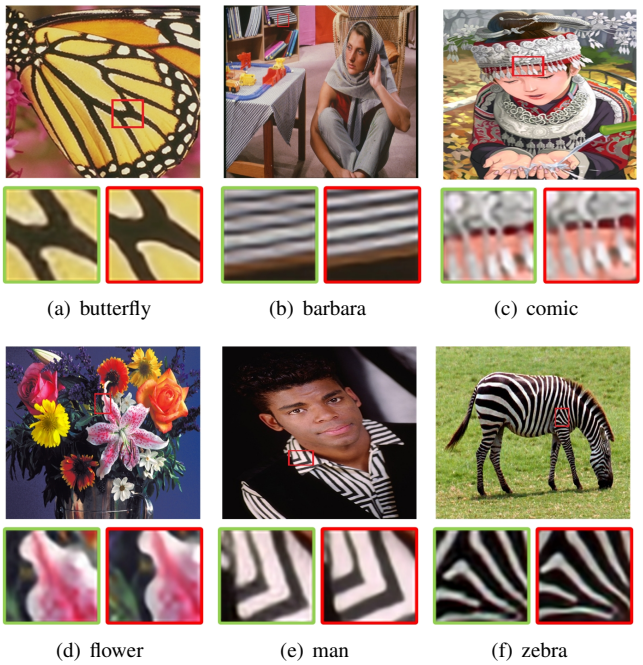
where $F_e^1(x_e^1) = d_e^1(x_e^1) + f_e^{m-1}(x_e^1)$. And the reconstructed HR edge maps $\hat{x}_e$ can be reformulated as

$$\begin{aligned} \hat{x}_e =& f_{recon}(y_e^2) + x_e \\ =& f_{recon}(F_e^m(y_e^1)) + x_e \\ =& f_{recon}(F_e^m(f_{ext}(x_e))) + x_e \end{aligned} \tag{7}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access

TABLE 1: Parameters Setting of Each Component in Our Framework

Componet	FENet		edge-EDN		color-EDN		ReconNet	Fusion unit
layer-name	2-convs	1-conv	5-convs	5-deconvs	5-convs	5-deconvs	conv	conv
filter	3×3	1×1	3×3	3×3	3×3	3×3	3×3	1×1
stride	1	1	1	1	1	1	1	1
channel	64	64	64	64	64	64	1	1



(a) butterfly (b) barbara (c) comic

(d) flower (e) man (f) zebra

FIGURE 4: Visual comparisons on several images from the benchmark datasets with the scaling factor 3. The green rectangles denote the results produced by color-EDN, and the red rectangles are the corresponding results produced by Dual-EEDN. The contents and boundary from Dual-EEDN are much clearer and sharper, whereas the color-EDN gives blurry boundary.

C. COLOR STREAM BASED ENCODER-DECODER NETWORK

In our Dual-EDN, the networks of edge-EDN and color-EDN share the same structure and parameter size. After extracting features by FENet, the output x_c^1 is sent to color-EDN. Like edge-EDN, m convolutional layers are stacked to encode the extracted color image features while preserving clearer contents. We use y_c^1 represents the feature maps learned from the convolutional layers,

$$\begin{aligned} y_c^1 &= f_c^m(f_c^{m-1}(...f_c^1(x_c^1)...)) \\ &= f_c^m(f_c^{m-1}(...f_c^1(f_{ext}(x_c))...)) \end{aligned} \tag{8}$$

where $f_c^m(\cdot)$ represents the mapping function of the m^{th} convolutional layer. The m^{th} deconvolutional layer then decodes the color image primary components to recover HR color image contents with better performance. The mapping functions of the stacked deconvolutional layers can be represented as $[d_c^1, d_c^2, ..., d_c^{m-1}, d_c^m]$. Skip connections are adopted to pass previous information to current layer for addressing the degradation problem caused by the deep network, which can compensate the lost high-frequency information at latter layers. Similar to edge-EDN, in color-EDN, the corresponding relationship between the encoder and decoder can be formulated as

$$\begin{aligned} y_c^2 =& F_c^m(y_c^1) \\ =& d_c^m(d_c^{m-1}(...(d_c^2(F_c^1(x_c^1)))...)) + x_c^1 \end{aligned} \tag{9}$$

where $F_c^1(x_c^1) = d_c^1(x_c^1) + f_c^{m-1}(x_c^1)$. Finally, our color-EDN uses a convolutional layer in ReconNet to reconstruct HR color image content,

$$\begin{aligned} \hat{x}_c =& f_{recon}(y_c^2) + x_c \\ =& f_{recon}(F_c^m(x_c^1)) + x_c \\ =& f_{recon}(F_c^m(f_{ext}(x_c))) + x_c \end{aligned} \tag{10}$$

where $\hat{x}_c$ is the reconstructed HR color image contents.

D. FUSION UNIT

To combine the benefits of multitask learning with edge prior for jointly recovering HR images with more texture details, a fusion unit is utilized to fuse the reconstructed HR edge contents with the HR color image contents predicted by color-EDN. It is worth emphasizing that we do not calculate the predicting loss between the reconstructed HR color image representations and the ground truth as edge-EDN. Specifically, a convolutional layer with kernel size 1×1 is adopted to learn the joint representations of the HR edge contents and color contents to guide our proposed Dual-EEDN recovering HR images with better structural details and sharper edges. The fusion function can be represented as

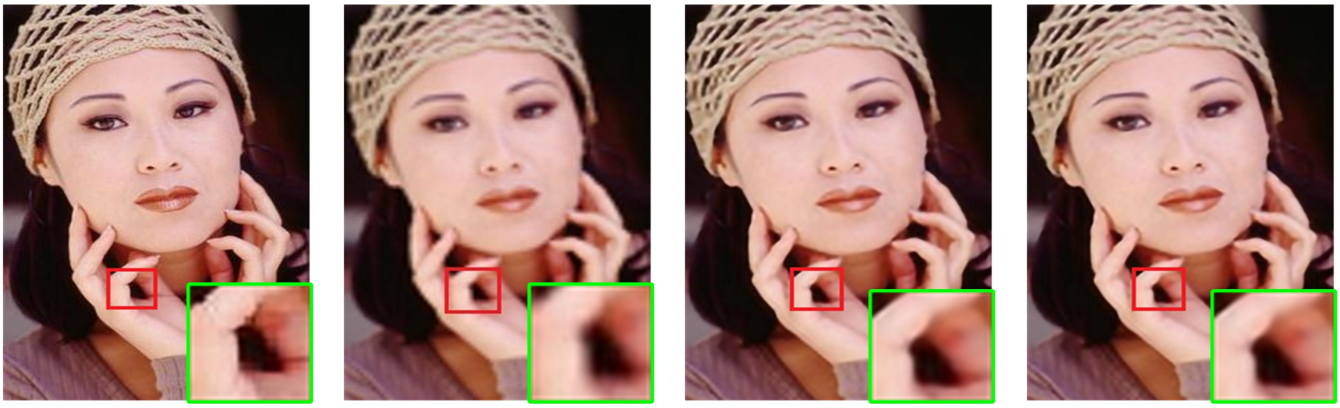
$$\hat{x}_{HR} = f_u(\hat{x}_e + \hat{x}_c) \tag{11}$$

where $\hat{x}_{HR}$ denotes the final HR images with better performance by jointly learning of Dual-EEDN.

E. FRAMEWORK TRAINING

As mentioned above, our proposed Dual-EEDN is jointly optimized from " LR edge maps, LR images" to "HR edge maps, HR images". For convenience, we use $[\tilde{x}_e, \tilde{x}_c]$ to represent the ground truth of LR edge maps and LR images respectively. Given a training set $[x_e^i, x_c^i, \tilde{x}_e^i, \tilde{x}_c^i]$, the object of this work is to recover a HR image while preserving high-frequency details. For weights initialization, we use the method described in [34]. To make Dual-EEDN learn more complicated features and ensure the nonlinear mapping, the rectified linear unit (ReLU) [35] is used as the activation function. For each convolutional and deconvolutional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access



(a) Ground Truth (PSNR/SSIM) (b) Bicubic (28.56/0.889) (c) A+ [17] (31.20/0.929) (d) SRF [20] (30.58/0.920)

(e) SelfEx [16] (31.45/0.932) (f) SRCNN [25] (30.89/0.923) (g) FSRCNN [26] (31.80/0.934) (h) Our (31.82/0.936)

FIGURE 5: Visual comparisons on the "woman" image from *Set5* [38] for $\times 3$ scale. The boundary of finger is sharper in our results, whereas other methods produce blurry boundary.

TABLE 2: Quantitative Comparisons of PSNR (dB) and SSIM on *Set5* and *BSD100* with the Scaling Factor of 2, 3, 4. Text Indicates the Best Performance.

Dataset	Scale	color-EDN	Dual-EDN
Set5	2	36.93/0.955	**37.13/0.958**
	3	33.04/0.914	**33.25/0.917**
	4	30.57/0.865	**30.92/0.876**
BSD100	2	31.53/0.890	**31.90/0.893**
	3	28.65/0.791	**28.69/0.794**
	4	26.98/0.715	**27.14/0.720**

layer, we define a composite function of two consecutive operations: convolution/deconvolution, followed by a ReLU activation function. In our proposed model, the loss function $L_1(x_e, \Theta_1)$ of edge-EDN can be formulated as

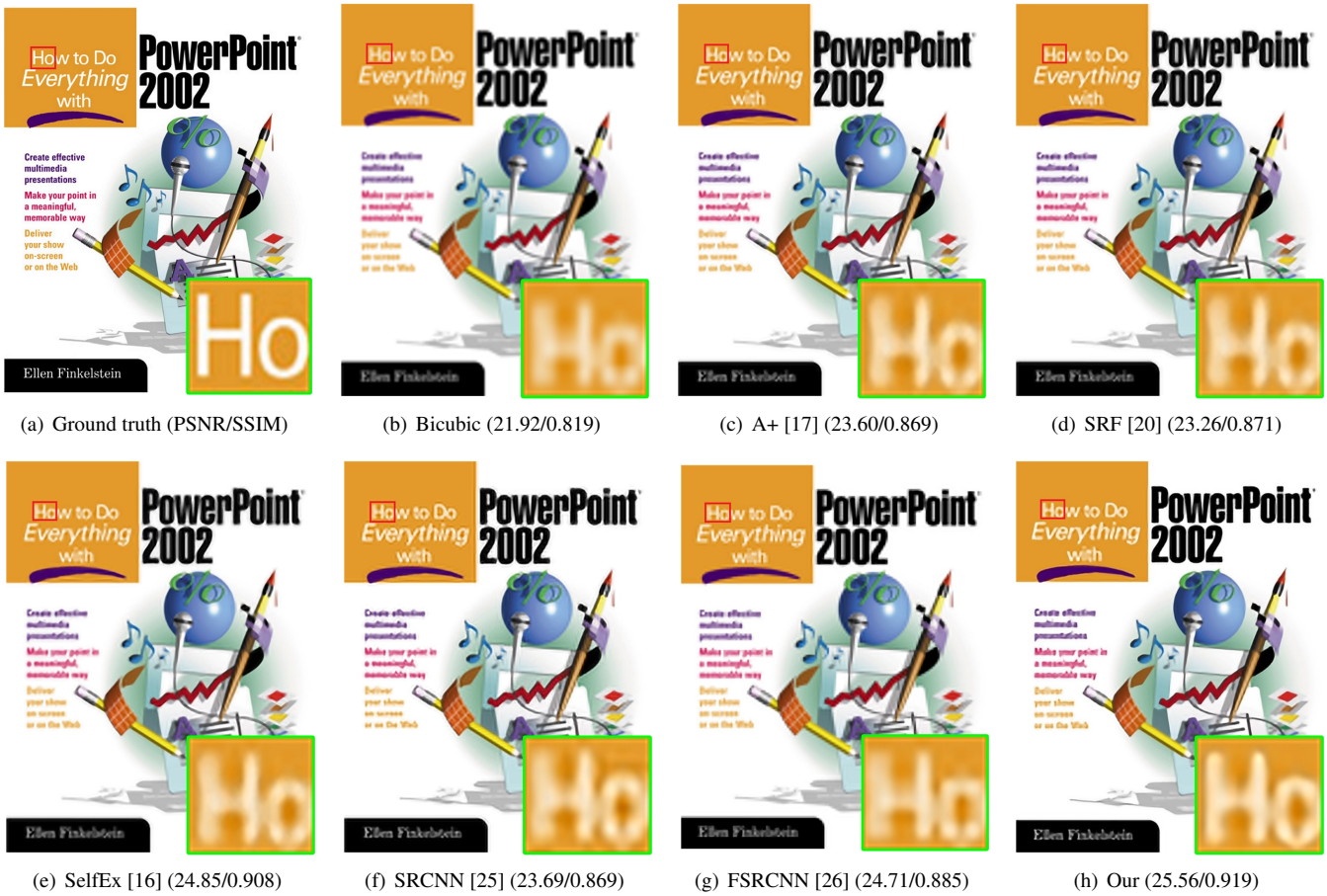
$$\begin{aligned} L_1(x_e, \Theta_1) &= \frac{1}{N}\sum_{i=1}^{N} \left\| (f(x_e^i, \Theta_1) - \tilde{x}_e^i) \right\|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \left\| (\hat{x}_e^i - \tilde{x}_e^i) \right\|^2 \end{aligned} \tag{12}$$

where Θ_1 represents the learning parameters of edge-EDN, and $f(x_e^i, \Theta_1)$ represents the learning mapping function of edge-EDN. Here, $\hat{x}_e^i$ is the reconstructed HR edge map. Furthermore, since we utilize a fusion unit to learn the joint representations of HR edge contents and color contents, the loss function between HR color images predicted by Dual-EEDN and original HR images can be formulated as

$$\begin{aligned} L_2(x_c, \Theta_2) &= \frac{1}{N}\sum_{i=1}^{N} \left\| (f(x_e^i, x_c^i, \Theta_2) - \tilde{x}_c^i) \right\|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \left\| f_u(\hat{x}_c^i, \hat{x}_e^i) - \tilde{x}_c^i \right\|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \left\| \hat{x}_{HR}^i - \tilde{x}_c^i \right\|^2 \end{aligned} \tag{13}$$

where Θ_2 denotes the learning parameters of our model, and $f(x_e^i, x_c^i, \Theta_2)$ is the joint learning function of our Dual-EEDN. To achieve an optimal trade-off between the image fidelity and image details, we develop a trade-off parameter λ to balance the importance of color image contents and edge details,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access



(a) Ground truth (PSNR/SSIM) (b) Bicubic (21.92/0.819) (c) A+ [17] (23.60/0.869) (d) SRF [20] (23.26/0.871)

(e) SelfEx [16] (24.85/0.908) (f) SRCNN [25] (23.69/0.869) (g) FSRCNN [26] (24.71/0.885) (h) Our (25.56/0.919)

FIGURE 6: Visual comparisons on the "ppt3" image from *Set14* [39] for $\times 4$ scale. The text in Dual-EEDN is clearer and sharper, whereas in other methods, the edges of character are blurry.

$$
\begin{aligned}
&L_{total}(x_e, x_c, \Theta_1, \Theta_2) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| (f(x_c^i, x_e^i, \Theta_2) - \tilde{x}_c^i \right\|^2 \\
&\quad + \lambda \cdot \frac{1}{N}\sum_{i=1}^{N} \left\| (f(x_e^i, \Theta_1) - \tilde{x}_e^i \right\|^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| (\hat{x}_{HR}^i - \tilde{x}_c^i) \right\|^2 + \lambda \cdot \frac{1}{N}\sum_{i=1}^{N} \left\| (\hat{x}_e^i - \tilde{x}_e^i) \right\|^2
\end{aligned} \tag{14}
$$

where $L_{total}(x_e, x_c, \Theta_1, \Theta_2)$ represents the total loss of our Dual-EEDN. In this paper, we set λ as 1 to get well performance according to a number of experimental results with different values of λ, which is demonstrated in Section III. The weights are shared in the edge-EDN and color-EDN.

III. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our model on several datasets. Here, we first describe datasets used for training and testing our model. Then the implementation details of this work are given. To demonstrate the importance of edge prior for image SR, we compare the performance of Dual-EEDN with the color-EDN which can be seen as the network without embedding edge information. Based on the basic network settings, we investigate the different value of λ to research an optimal trade-off between image fidelity and texture details. Finally, the proposed method is compared with several state-of-the-art.

TABLE 3: Quantitative Comparisons of PSNR (dB) on *BSD100* with Different Values of λ. The Text Indicates the Best Performance.

λ	0.5	0.7	1.0	1.5	2	4
$\times 2$	31.68	31.69	**31.70**	31.68	31.66	31.62
$\times 3$	28.68	**28.69**	**28.69**	**28.69**	28.67	28.65
$\times 4$	27.12	27.12	**27.14**	**27.14**	27. 12	27.08

A. DATASETS

Following the experimental setting in [28] , we use a training dataset of 291 images, which includes 200 training images from Berkeley Segmentation Dataset (*BSD200*) [36] and 91 images from Yang *et al.* [37]. For comparing our Dual-EEDN with recent SR methods, we evaluate our model on four widely used benchmarks: *Set5* [38], *Set14* [39], *BSD100* [36] and *Urban100* [16] with the scaling factors of 2, 3, 4.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access

TABLE 4: Quantitative Comparisons of State-of-the-art Methods in Term of PSNR (dB) with Three Scales ($\times 2$, $\times 3$, $\times 4$). Red Text Indicates The Best Performance and Text Indicates the Second Best Performance.

Dataset	Scale	Bicubic	A+ [17]	SRF [20]	SelfEx [16]	SRCNN [25]	FSRCNN [26]	CSCN [44]	AGST [18]	Dual-EEDN
Set5	2	33.65	36.55	36.89	36.34	36.49	**36.94**	36.93	36.69	**37.13**
	3	30.39	32.59	32.72	32.58	32.59	33.06	**33.10**	32.77	**33.25**
	4	28.42	30.28	30.35	30.31	30.48	30.55	**30.86**	30.45	**30.92**
Set14	2	30.23	32.28	32.52	32.22	32.45	**32.54**	32.42	32.45	**32.73**
	3	27.54	29.07	29.23	29.16	29.30	29.37	**29.41**	29.37	**29.60**
	4	26.00	27.32	27.41	27.40	27.20	27.50	**27.64**	27.55	**27.76**
BSD100	2	29.56	31.22	31.16	31.18	31.36	**31.51**	31.24	31.38	**31.70**
	3	27.21	28.29	28.22	28.29	28.41	**28.65**	28.54	28.45	**28.69**
	4	25.96	26.82	26.75	26.84	26.91	**26.97**	26.87	26.88	**27.14**
Urban100	2	26.88	29.23	29.13	29.54	29.52	**29.87**	29.50	-/-	**30.33**
	3	24.46	25.58	25.86	26.44	26.24	26.55	**26.57**	-/-	**26.79**
	4	23.15	24.34	24.20	**24.79**	24.53	24.61	24.52	-/-	**24.88**

TABLE 5: Quantitative Comparisons of State-of-the-art Methods in Term of SSIM with Three Scales ($\times 2$, $\times 3$, $\times 4$). Red Text Indicates The Best Performance and Text Indicates the Second Best Performance.

Dataset	Scale	Bicubic	A+ [17]	SRF [20]	SelfEx [16]	SRCNN [25]	FSRCNN [26]	CSCN [44]	AGST [18]	Dual-EEDN
Set5	2	0.930	0.954	0.954	0.954	0.952	0.955	**0.957**	0.955	**0.958**
	3	0.868	0.909	0.906	0.909	0.909	**0.913**	0.911	0.910	**0.917**
	4	0.810	0.860	0.853	0.862	0.863	0.865	**0.875**	0.861	**0.876**
Set14	2	0.869	0.906	0.904	0.903	0.904	0.908	**0.909**	0.905	**0.911**
	3	0.774	0.819	0.817	0.820	0.815	0.823	**0.825**	0.822	**0.828**
	4	0.702	0.749	0.745	0.752	0.750	0.753	**0.754**	0.752	**0.758**
BSD100	2	0.843	0.886	0.884	0.886	0.888	**0.891**	0.885	0. 888	**0.893**
	3	0.739	0.784	0.781	0.784	0.786	**0.789**	0.789	0.783	**0.794**
	4	0.668	0.709	0.705	0.711	0.710	**0.714**	0.710	0.710	**0.720**
Urban100	2	0.840	0.894	0.890	0.897	0.895	**0.901**	0.896	-/-	**0.908**
	3	0.735	0.797	0.781	0.790	0.809	**0.817**	0.817	-/-	**0.818**
	4	0.658	0.718	0.710	**0.737**	0.725	0.727	0.726	-/-	**0.740**

We evaluate the SR images with two popular metrics: PSNR [40], SSIM [41]. Since the image SR reconstruction is performed on the luminance component only in YCbCr color space as similarly practice in previous methods, the PSNR and SSIM are calculated on the Y-channel of images.

B. IMPLEMENTATION DETAILS

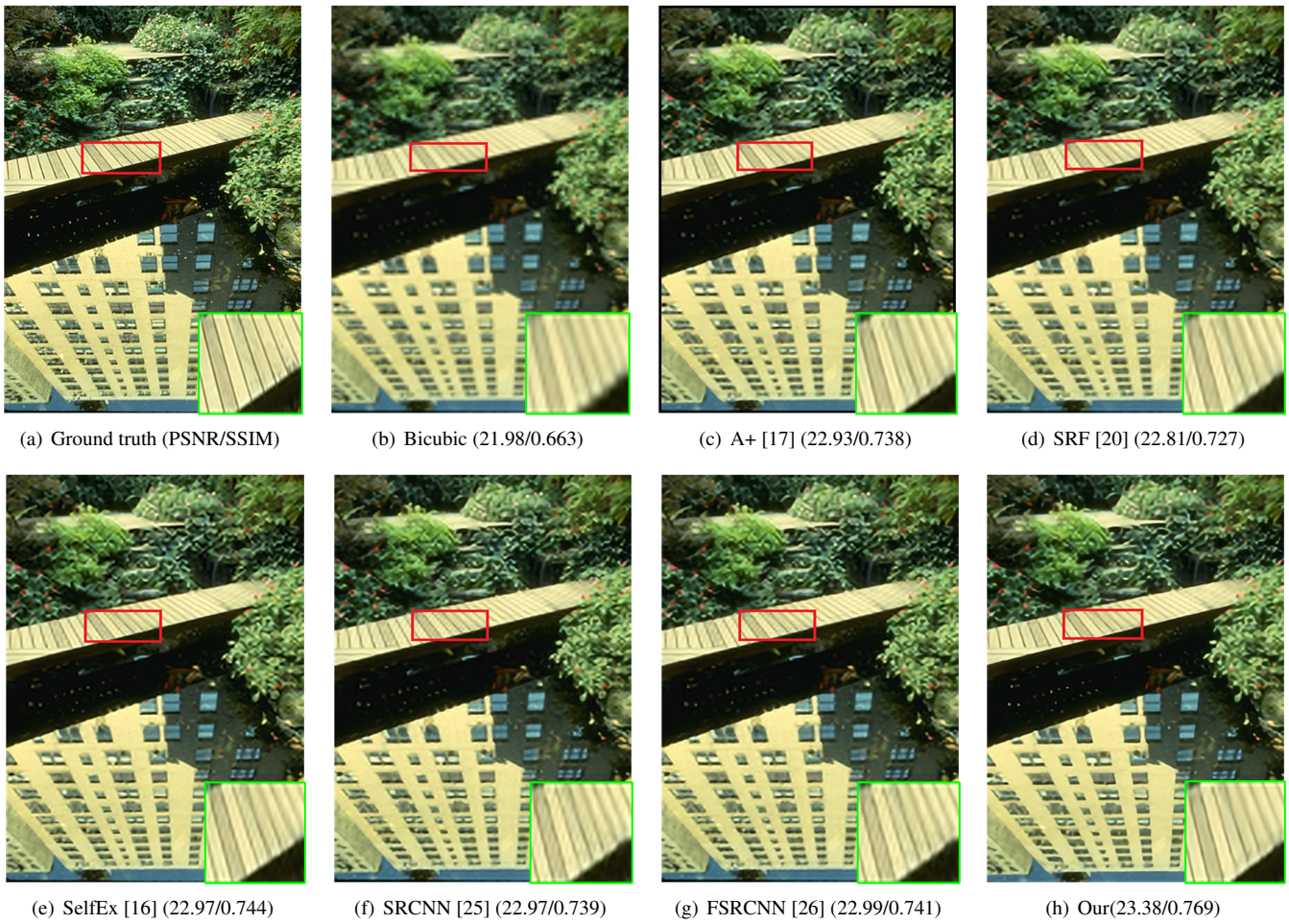
For data augmentation, the training images are rotated by $90°$, $180°$, $270°$ and flipped horizontally. Meanwhile, we augment the training data with different scales, so that the training images with different scales ($\times 2$, $\times 3$, and $\times 4$) are also included in the training set. In the training phase, we first detect the edge maps from LR images by Sobel extractor, and the examples are shown in Fig. 3. The input LR images are generated from original images and downsampled with bicubic interpolation.

In this work, our proposed Dual-EEDN uses a framework with 26 convolutional layers. As illustrated in Table 1, in the FENet, 3 convolutional layers are adopted to extract the hybrid features of LR color images and edge maps. Besides, in edge-EDN and color-EDN, our two streams encoder-decoder networks both consist of 5 convolutional layers and 5 deconvolutional layers. Then a convolutional layer is adopted as ReconNet in each stream to reconstruct the HR edge contents and color image contents. Finally, 1 convolutional layer is utilized in the fusion unit to combine the edge contents and color image contents for higher accuracy reconstruction. In our training process, the training images are split into 31×31 patches with the stride of 21. We use Adam solver [42] with a mini-batch size of 64. The weights are initialized as described in [34]. Learning rate is initially set to 0.001 and then decreased by a factor of 10 every 30 epochs. Momentum and weight decay parameters are set to 0.9 and 0.0001, respectively. We implement and train our Dual-EEDN using Caffe platform [43] with 1 Titan X Pascal GPU.

C. IMPORTANCE OF EDGE PRIOR FOR IMAGE SR

In this work, considering that edge knowledge can contribute to producing sharp edges and compensating the high-frequency details of reconstructed HR images, the edge prior is integrated into the Dual-EEDN and performs the image SR problem. To validate the advance of edge prior for image SR, we compare the performance of Dual-EEDN with color-EDN, which can be seen as the network without embedding edge information. Table 2 presents the objective quality of the two methods on *Set5* and *BSD100*. Besides, in Figure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access



(a) Ground truth (PSNR/SSIM) (b) Bicubic (21.98/0.663) (c) A+ [17] (22.93/0.738) (d) SRF [20] (22.81/0.727)

(e) SelfEx [16] (22.97/0.744) (f) SRCNN [25] (22.97/0.739) (g) FSRCNN [26] (22.99/0.741) (h) Our(23.38/0.769)

FIGURE 7: Visual comparisons on the "148026" image from *BSD100* [36] for $\times 3$ scale. Line is straightened and sharpened in our result, whereas other methods give blurry lines. Our result seems more visually pleasing.

4, we show visual comparisons on several images from the benchmark datasets to demonstrate the importance of edge prior for high-frequency details preserving. As shown in Figure 4, by utilizing the edge prior for jointly reconstruction, the Dual-EEDN can recover more high-frequency details compared to the color-EDN, which validates the importance of edge prior for image SR.

D. IMAGE FIDELITY AND TEXTURE DETAILS TRADE-OFF

For reconstructing HR images with better visual quality, we investigate a novel total loss function combining the edge loss and color loss for the best trade-off of image fidelity and edge information. According to (14), the parameter λ is developed to balance the importance of color image contents and edge details. The larger value of λ indicates that more consideration of image edge feature and less consideration of the of the main contents of HR images. We evaluate the performance of our proposed Dual-EEDN trained with different values of λ on *BSD100*. As shown in Table 3, the $\lambda = 0.7, 1.0, 1.5$ give more superior performance but the λ with larger value produce poorer results.

TABLE 6: Summary of PSNR (dB) and SSIM Results of Several Test Images for $\times 3$ Magnification. The Text Indicates the Best Performance.

Images	DGNE	Ours
Butterfly	27.06/0.890	**29.48/0.936**
Parrot	29.98/0.909	**31.35/0.930**
Girl	33.55/0.827	**33.80/0.829**
Bike	24.43/0.793	**26.75/0.826**
Hat	30.82/0.861	**32.36/0.900**

E. COMPARISONS WITH STATE-OF-THE-ART METHODS

We provide quantitative and qualitative comparisons with several state-of-the-arts methods, which include, A+ [17], SRF [19], SelfEx [16], SRCNN [25], FSRCNN [26], CSCN [44] and AGST [18]. We provide a summary of quantitative evaluation in Table 4 and Table 5, in which our method achieve the best performance both on PSNR and SSIM. Besides, in DGNE [5], multiview features and local spatial neighbors of patches are explored to find a feature-spatial manifold embedding for images. We compare our model to DGNE on several images from DGNE, and the results are illustrated in Table 6. To fully investigate how our proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2846363, IEEE Access

model perform in terms of visual quality, some promising results from several state-of-the-art methods with larger scales on *Set5* [38], *Set14* [39], and *Urban100* [16] are visualized in Figure 5, 6, and 7.

F. RUNNING TIME

We evaluate the running time of our proposed model and compare its efficiency with several state-of-the-art methods. All running time of the methods are evaluated by their original codes on the same machine: 3.4 GHz Intel i7 CPU (128G RAM) and NVIDIA Titan X Pascal GPU. Figure 8 shows the trade-off between the running time and performance on *Set14* for $\times 4$ SR. It is demonstrated that the speed of the proposed Dual-EEDN is faster than all the existing methods except FSRCNN but achieve PSNR gain of 0.26 dB compared to FSRCNN.

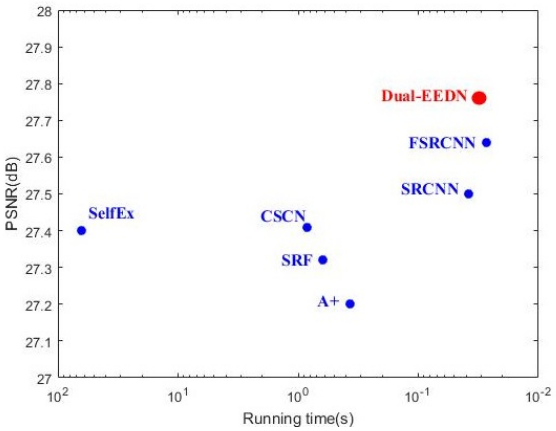


FIGURE 8: The efficiency comparison between our proposed Dual-EEDN and several state-of-the-art models. The results are evaluated on *Set14* [39] with the scale factor $\times 4$.

IV. CONCLUSION

In this paper, we proposed a novel dual-streams edge driven encoder-decoder network for single image SR (Dual-EEDN). Our framework utilizes edge contents and color contents to reconstruct HR images with well image details. The edge information is extracted from the color images to drive the reconstruction of HR images. In Dual-EEDN, instead of utilizing two sub-networks that learn edge information and color image contents respectively, an optimized HR edge maps are recovered by edge-EDN, then we impose color stream based encoder-decoder network (color-EDN) to learn color image contents. The reconstructed HR edge contents are fused with color contents predicted from color-EDN to recover HR images with much clearer texture details. Extensive benchmark experiments and analysis have shown that Dual-EEDN is a superior framework for single image SR. Although our model has achieved very promising results for image SR, we will extend our framework to other image restoration tasks, such as image denoising and JPEG artifacts reduction and demonstrate the effectiveness of our framework on real-world image restoration.

REFERENCES

[1] W. Shi, J. Caballero, C. Ledig, X. Zhang, W. Bai, K. Bhatia, A. Marvao, T. Dawes, D. ORegan, and D. Ruechkert, "Cardiac image super-resolution with global correspondence using multi-atlas patch match," in *MICCAI*, 2013, pp. 9–16.

[2] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Process.*, vol. 90, no. 3, pp. 848–859, 2010.

[3] W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, 2012.

[4] T. Lu, Z. X. Xiong, Y. D. Zhang, B. Wang, and T. Lu, "Robust Face Super-Resolution via Locality-Constrained Low-Rank Representation," *IEEE Access*, vol. 5, pp. 13103–13117, 2017.

[5] S. Yang, Z. Wang, L. Zhang, and M. Wang, "Dual-Geometric Neighbor Embedding for Image Super Resolution With Sparse Tensor," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2793–2803, 2014.

[6] T. Micheal and M. Irani, "Nonparametric blind super-resolution," in *Proc. IEEE Int'l Conf. Comput. Vis.*, 2013, pp. 945–952.

[7] Y. -W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2400–2407.

[8] L. Wang, S. Xiang, G. Meng, H.-Y. Wu, and C. Pan, "Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation." *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1289–1299, 2013.

[9] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4544–4556, 2012.

[10] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, 2001.

[11] J. Chu, J. Liu, J. Qiao, X. Wang, and Y. Li, "Gradient-based adaptive interpolation in super-resolution image restoration," in *Proc. Int'l Conf. Signal Process.*, 2008, pp. 1027–1030.

[12] S. J. van der Walt and B. M. Herbst, "A polygon-based interpolation operator for super-resolution imaging," *arXive: 1210.3404*, 2012.

[13] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int'l Conf. Comput. Vis.*, 2009, pp. 349–356.

[14] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graphics*, vol. 30, no. 2, 2011, pp. 1–11.

[15] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-Image Super-Resolution Reconstruction via Learned Geometric Dictionaries and Clustered Sparse Coding," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4016–4028, 2012.

[16] J. B. Huang, A. Singh, and N. Ahuja, "Single image superresolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.

[17] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. IEEE Asia. Conf. Comput. Vis.*, 2014, pp. 111–126.

[18] Q. Song, R. Xiong, D. Liu, Z. Xiong, F. Wu, and W. Gao, "Fast Image Super-Resolution via Local Adaptive Gradient Field Sharpening Transform," *IEEE Trans. Image. Process.*, vol. 27, no. 4, pp. 1966–1980, 2018.

[19] K. Jia, X. Wang, and X. Tang, "Image transformation based on learning dictionaries across image spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 367–380, 2013.

[20] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3791–3799.

[21] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput Vis.*, 2013, pp. 1920–1927.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3124-3155, 2017.

[23] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int'l Conf. Comput. Vis.*, 2015, pp. 576–584.

[24] C. Yang, X, Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4076–4084.

[25] C. Dong, 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, 2016.

[26] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," in *Proc. Euro. Conf. Comput. Vis.*, 2016, pp. 391–407.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXive: 1409.1556*, 2014.
[28] 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.*, 2016, pp. 1646–1654.
[29] 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.*, 2016, pp. 1637–1645.
[30] K. Zeng, J. Yu, R. Wang, C. Li, and D. Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 27–37, 2017.
[31] X. J. Mao, C. Shen, and Y. B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016.
[32] Y. Liang, J. Wang, S. Zhang, and Y. Gong, "Incorporating image degeneration modeling with multitask learning for image super-resolution," in *Proc. IEEE Int'l Conf. Image Process.*, 2015, pp. 2110âĂŞ2114.
[33] L. Zhao, H. Bai, J. Liang, B. Zeng, A. Wang, and Y. Zhao, "Simultaneously Color-Depth Super-Resolution with Conditional Generative Adversarial Network," *arXiv: 1708.09105*, 2017.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[35] V. Nair, and G. E. Hinton, âĂIJRectified linear units improve restricted boltzmann machines,âĂİ in *Proc. Int'l Conf. Mach. Learn.*, 2010, pp. 807-814.
[36] 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. IEEE Int'l Conf. Comput. Vis.*, 2001, pp. 416–423.
[37] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image superresolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.
[38] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. British Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.
[39] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int'l Conf. Curves Surfaces*, 2012, pp. 711–730.
[40] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.
[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, âĂIJImage quality assessment: From error visibility to structural similarity,âĂİ *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[42] D. P. Kingma and J. Ba. Adam, "A method for stochastic optimization," in *arXiv:1412.6980*, 2014.
[43] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv:1408.5093*, 2014.
[44] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int'l Conf. Comput. Vis.*, 2015, pp. 370–378.

FENG LI received his B.S. degree in Anhui Normal University, China, in 2012. Now, he is pursuing his Ph. D degree in Institute of Information Science, Beijing Jiaotong University, Beijing, China. His research interests are image and video compression, video super resolution, and image restoration, such as image super-resolution, and compression reduction.

HUIHUI BAI received her B.S. degree from Beijing Jiaotong University, China, in 2001, and her Ph.D. degree from Beijing Jiaotong University, China, in 2008. She is currently a professor in Beijing Jiaotong University. She has been engaged in R&D work in video coding technologies and standards, such as HEVC, 3D video compression, multiple description video coding (MDC), and distributed video coding (DVC).

LIJUN ZHAO received his M.E. degrees in Taiyuan University of Science and Technology (TYUST) 2015. Now he is pursuing his Ph.D degree in Institute of Information Science, Beijing Jiaotong University. His research interests include image coding, 3D image processing, pattern recognition, and computer vision.

YAO ZHAO received the BS degree from Fuzhou University, China, in 1989, and the ME degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996. He became an associate professor at BJTU in 1998 and became a professor in 2001. From 2001 to 2002, he was a senior research fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He serves on the editorial boards of several international journals, including as associate editors of IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, and an area editor of Signal Processing: Image Communication (Elsevier), etc. He was named a distinguished young scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of Ministry of Education of China in 2013. He is a senior member of the IEEE.