Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution

Wenhan Yang, Jiashi Feng, Jianchao Yang, Fang Zhao, Jiaying Liu, *Senior Member, IEEE*,
Zongming Guo, *Member, IEEE*, and Shuicheng Yan, *Fellow, IEEE*

***Abstract*— In this paper, we consider the image super-resolution (SR) problem. The main challenge of image SR is to recover high-frequency details of a low-resolution (LR) image that are important for human perception. To address this essentially ill-posed problem, we introduce a Deep Edge Guided REcurrent rEsidual (DEGREE) network to progressively recover the high-frequency details. Different from most of the existing methods that aim at predicting high-resolution (HR) images directly, the DEGREE investigates an alternative route to recover the difference between a pair of LR and HR images by recurrent residual learning. DEGREE further augments the SR process with edge-preserving capability, namely the LR image and its edge map can jointly infer the sharp edge details of the HR image during the recurrent recovery process. To speed up its training convergence rate, by-pass connections across the multiple layers of DEGREE are constructed. In addition, we offer an understanding on DEGREE from the view-point of sub-band frequency decomposition on image signal and experimentally demonstrate how the DEGREE can recover different frequency bands separately. Extensive experiments on three benchmark data sets clearly demonstrate the superiority of DEGREE over the well-established baselines and DEGREE also provides new state-of-the-arts on these data sets. We also present addition experiments for JPEG artifacts reduction to demonstrate the good generality and flexibility of our proposed DEGREE network to handle other image processing tasks.**

***Index Terms*— Super-resolution, edge guidance, recurrent residual network, sub-band recovery.**

I. Introduction

IMAGE super-resolution (SR) aims at recovering a high resolution (HR) image from low resolution (LR) observations. Although it has seen wide applications, such as surveillance video recovery [40], face hallucination [20], medical image enhancement [12], the SR problem, or more concretely the involved inverse signal estimation problem therein, is essentially ill-posed and still rather difficult to solve. In order to relieve ill-posedness of the problem, most of recent SR methods propose to incorporate various prior knowledge about natural images to regularize the signal recovery process. This strategy establishes a standard maximum *a posteriori* (MAP) image SR framework [2], [25], where an HR image is estimated by maximizing its fidelity to the target with kinds of *a priors*.

Most of existing MAP based image SR methods [29], [37] associate the data fidelity term with the mean squared error (MSE), in order to ensure consistency between the estimated HR image and the ground truth when learning model parameters. However, solely considering minimizing MSE usually fails to recover the sharp or high-frequency details such as textures and edges. This phenomenon is also observed in much previous literature [4], [26], [27]. To address this problem, bandpass filters – that are commonly used to extract texture features – were employed to preserve sharp details in the image SR process [4], [11], [26], [27]. The bandpass filters decompose an LR image into several sub-band images and build hierarchical fidelity terms to steer recovery of those sub-band images. The hierarchical fidelity consideration is shown to be able to help preserve moderate-frequency details and thus improve quality of the produced HR images.

Besides data fidelity, another important aspect for MAP based image SR methods is priors on HR images, which are effective in relieving ill-posedness of the problem. Commonly used priors describing natural image properties include sparseness [6], [32], spatial smoothness [1], [21] and nonlocal similarity [24], which help produce more visually pleasant HR images. Among those priors, the edge prior [5], [30], [41] is a very important one. In contrast to textures that are usually difficult to recover after image degradation, edges are much easier to detect in LR images and thus more informative for recovering details of HR images. Thus, separating edges from the image signal and modeling them separately would benefit image SR substantially.

Recently, several deep learning based SR methods have been developed, in order to utilize the strong capacity of deep neural networks in modeling complex image contents and details. The image super-resolution CNN (SRCNN) [8] is the seminal work that has introduced a deep convolutional network model to image SR. The proposed SRCNN consists of three convolutional layers and is equivalent to performing a sparse

Manuscript received August 9, 2016; revised February 16, 2017; accepted August 18, 2017. Date of publication September 8, 2017; date of current version September 21, 2017. This work was supported by the National Natural Science Foundation of China under Contract 61772043 and CCF-Tencent Open Research Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Feng Wu. *(Corresponding author: Jiaying Liu.)*

W. Yang, J. Liu, and Z. Guo are with the Institute of Computer Science and Technology, Peking University, Beijing 100080, China (e-mail: yangwenhan@pku.edu.cn; liujiaying@pku.edu.cn; guozongming@pku.edu.cn).

J. Feng and F. Zhao are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: elefjia@nus.edu.sg; elezhf@nus.edu.sg).

J. Yang is with Snapchat Inc., Los Angeles, CA USA (e-mail: jyang29@ifp.uiuc.edu).

S. Yan is with the Artificial Intelligence Institute, Qihoo 360 Technology Company, Ltd., Beijing 100015, China (e-mail: eleyans@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2750403

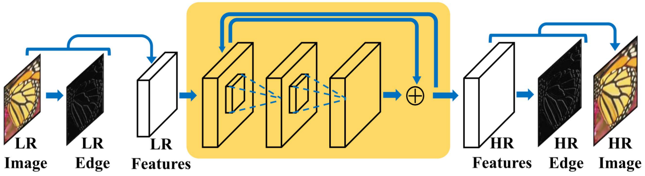


Fig. 1. The framework of the proposed DEGREE network for image SR. The DEGREE network takes the raw LR image as well as prior map (LR edge here) as its inputs and outputs the predicted HR feature maps and HR edge maps – which are integrated to produce the HR image. The recurrent residual network (highlighted in orange color) recovers sub-bands of the HR image features from the LR input iteratively and actively utilizes edge feature guidance in image SR for preserving sharp details.

reconstruction to generate HR images. Benefiting from being end-to-end trainable, SRCNN improves the quality of image SR significantly. However, SRCNN only aims at minimizing the MSE loss without exploiting natural image priors and suffers from losing sharp details. Following SRCNN, several recent works [23], [34] propose to embed sparsity priors into the deep networks for image SR, offering more visually pleasant results. However, much domain knowledge and extra effort are needed for designing a suitable architecture to model the sparsity priors. A simple and adaptive method to embed various priors into standard CNN networks for image SR is still absent.

Motivated by the fact that edge features can provide valuable guidance for image SR and the success of deep neutral network models, we propose a Deep Edge Guided REcurrent rEsidual (DEGREE) network to progressively perform image SR with properly modeled edge priors. Instead of trying to predict HR images from LR ones directly, the DEGREE model takes an alternative route and focuses on predicting the *residue* between a pair of LR and HR images, as well as the edges in HR images. Combining these predictions together give a recovered HR image with high quality and sharp high-frequency details. An overview on the architecture of the DEGREE model is provided in Fig. 1. Given an LR image, DEGREE extracts its edge features and takes the features to predict edges of the HR image via a deep recurrent network. To recover details of an HR image progressively, DEGREE adopts a recurrent residual learning architecture that recovers details of different frequency sub-bands at multiple recurrence stages. Bypass connections are introduced to fuse recovered results from previous stages and propagate the fusion results to later stages. In addition, adding bypass connections enables a deeper network trained with faster convergence rate.

In summary, our contributions to image SR can be summarized as:

1) We introduce a novel DEGREE network model to solve image SR problems. The DEGREE network integrates edge priors and performs image SR recurrently, and improves the quality of produced HR images in a progressive manner. Moreover, DEGREE is end-to-end trainable and thus effective in exploiting edge priors for both LR and HR images. With the recurrent residual learning and edge guidance, DEGREE outperforms well-established baselines significantly on three benchmark datasets and provides new state-of-the-arts.
2) The proposed DEGREE also introduces a new framework that is able to seamlessly integrate useful prior knowledge into a deep network to facilitate solving various image processing problems in a principled way. By letting certain middle layers in the DEGREE alike framework learn features reflecting the priors, our framework gets rid of hand-crafting new types of neurons for different image processing tasks based on domain knowledge, and thus is highly flexible to integrate useful priors into deep SR and other tasks.
3) To the best of our knowledge, we are the first to apply recurrent deep residual learning for SR, and we establish the relation between it and the classic sub-band recovery. Our extensive experimental results demonstrate that the recurrent residual structure is more effective in image SR than the standard feed forward architecture used in the modern CNN models. This is promising for providing new ideas for the community on how to design an effective network for SR or other tasks based on well-built traditional methods.

The rest of this paper is organized as follows. Related works are briefly reviewed in Section II. Section III illustrates how to build a recurrent residual network to perform sub-band image SR progressively. Section IV constructs DEGREE network via embedding the edge prior into the recurrent residual network. Experimental results are presented in Section V. Concluding remarks and future directions for research are presented in Section VI.

II. Related Work

A. Single Image Super-Resolution

The key issue of image SR is to recover the missing high-frequency details. However, the mean square error (MSE), adopted by most SR methods as the constraint or loss function, naturally leads to losing very high frequency details due to its oversmoothing property. To address the high-frequency information loss issue in purely minimizing the MSE, sub-band decomposition based methods propose to recover information at different frequency bands of the image signal separately [4], [11], [26], [27]. In [11], interpolation to high-frequency sub-band images by discrete wavelet transform (DWT) was performed for image SR. Song *et al.* [27] proposed a joint sub-band-based neighbor-embedding SR with a constraint on each sub-band, achieving more promising SR results.

Some works also explore how to preserve edges in application of image SR, denoising and deblurring. Total variation (TV) [1], [21], focusing on modeling the intensity change of image signals, was proposed to guide the SR recovery by suppressing the excessive and possibly spurious details in the HR estimation. Bilateral TV (BTV) [10], [38] was then developed to preserve sharp edges. Sparsity prior [9], [37] constraining the transformation coefficients was introduced to enhance salient features. As a kind of sparsity prior, the gradient prior [19], [28], [42] was proposed to enforce the gradient distribution of the denoised image to fit distribution estimated from the original image. By embedding these regularizations, sharper and finer edges of HR images are restored.

B. Deep Image Super-Resolution

Many recent works have exploited deep learning for solving low level image processing problems including image denoising [33], image completion [36] and image super-resolution [7]. Particularly, Dong *et al.* [8] proposed a three layer CNN model for image SR through equally performing sparse coding. Instead of using a generic CNN model, Wang *et al.* [34] incorporated the sparse prior into CNN by exploiting a learned iterative shrinkage and thresholding algorithm (LISTA), which provided better reconstruction performance.

Compared with those deep CNN based image SR models, DEGREE is more general in the sense of being able to embed various priors systematically. It has quite unique architectures for achieving this: the bypass connections and priori (or more specifically edge in this context) embedding components. The bypass connections and induced residual learning help progressively recover the high-frequency details at a fast convergence rate. The edge embedding shows to be beneficial for SR, which however is not easy for deep CNN models to simulate.

Recently, two deep-learning based image SR approaches [17], [18] achieve very promising results. In [17], a very deep network is cascaded and trained with multi-scale augmentation. In [18], the difficulty of training a recursive convolutional network is addressed by the recursive-supervision and skip-connection. They propose to handle image SR problems purely from the perspective of machine learning without employing image processing domain knowledge. Different from them, we propose to construct a very deep architecture – the DEGREE network – by considering the frequency sub-band combination and exploiting low-level edge information.

Specifically, we are the first to explore a principled way to embed priors into a recurrent convolutional network for image SR. The network is designed based on the intuition of sub-band recovery. The embedded edge guidance and sub-band recovery jointly lead to sharper edges and better detail recovery, which is the important supplement to the existing very deep network approaches. Specifically, the network design follows sub-band recovery, where high-frequency information is inferred and recovered progressively. Thus, the details are reconstructed more accurately. The LR edge is extracted as the network input, making the network converge in less training epochs. The HR edges are embedded in the penultimate layer as guidance, which boosts objective results and helps provide better subjective results in edge and texture regions. These designs effectively help construct a very deep network even not using advanced training skills, lead to faster convergence in the training and provide sharper edge details.

III. Deep Recurrent Residual Learning for Image SR

In this section we first review the sub-band reconstruction methods [11], [26] for image SR. Then we illustrate how to build a recurrent residual network that can learn to perform sub-band reconstruction and recover HR images progressively.

A. Sub-Band Reconstruction for Image SR

In most cases, quality degradation of an HR image $\mathbf{x}$ to an LR image $\mathbf{y}$ is caused by blurring and down sampling, and the degradation process can be modeled as

$$\mathbf{y} = DH\mathbf{x} + \upsilon, \tag{1}$$

where H and D depict the blurring and down-sampling effects respectively. The additive noise in the imaging process is denoted as υ. Given an observed LR image $\mathbf{y}$, image SR aims at estimating the original HR $\mathbf{x}$. Most of image SR methods obtain an estimation of HR by solving the following MAP problem:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|DH\mathbf{x} - \mathbf{y}\|_2^2 + p(\mathbf{x}), \tag{2}$$

where $p(\cdot)$ is a regularization term induced by priors on $\mathbf{x}$. However, directly learning a one-step mapping function from $\mathbf{y}$ to $\mathbf{x}$ usually ignores some intrinsic properties hidden in different frequency bands of $\mathbf{x}$, such as the high-frequency edge and textural details. This is because the recovery function needs to fit the inverse mapping from the low-frequency component of the LR image to that of the HR one. It by nature neglects some high-frequency details with small energy.

To address this problem, a sub-band based image reconstruction method is proposed to recover images at different frequency bands separately. It separates the image signal into multiple components of different intrinsic frequencies, which are called sub-bands, and models them individually. In this way, the sub-bands with small energy can still gain sufficient "attention" and sharper image details can be preserved during image SR. Formally, let $\mathbf{y}_i$ be the i-th sub-band of the LR image $\mathbf{y}$ out of in total n sub-bands, *i.e.*, $\mathbf{y} = \sum_{i=1}^{n} \mathbf{y}_i$. $\mathbf{y}_i$ is used for estimating the i-th corresponding sub-band $\mathbf{x}_i$ of the HR image $\mathbf{x}$. The sub-band based method recovers different sub-bands individually and outputs the recovered HR image as follows,

$$\begin{aligned} \widehat{\mathbf{x}}_i &= \arg\min_{\mathbf{x}_i} \|DH\mathbf{x}_i - \mathbf{y}_i\|_2^2 + p(\mathbf{x}_i), \quad i = 1, 2, \ldots, n; \\ \widehat{\mathbf{x}} &= \sum_{i=1}^{n} \widehat{\mathbf{x}}_i. \end{aligned} \tag{3}$$

However, recovering each sub-band separately in (3) neglects the dependency across sub-bands. To fully model the dependencies both in the corresponding sub-bands and across sub-bands, we relax (3) into a progressive recovery process. It performs an iterative sub-band recovery implicitly and utilizes the useful information from lower-frequency sub-bands to recover higher-frequency ones.

For ease of explanation, we introduce an auxiliary signal $\mathbf{s}_i$ that approximates the signal $\mathbf{x}$ up to the i-th sub-band, *i.e.*, $\mathbf{s}_i = \sum_{j=1}^{i} \widehat{\mathbf{x}}_j$. Then, the sub-band image $\mathbf{x}_i$ and HR image $\mathbf{x}$ can be estimated through recovering $\widehat{\mathbf{x}}_i$ and $\mathbf{s}_i$ progressively. We here use $F_i(\cdot)$ and $G_i(\cdot)$ to denote the generating functions of $\mathbf{s}_i$ and $\widehat{\mathbf{x}}_i$ respectively, *i.e.*,

$$\widehat{\mathbf{x}}_i = F_i(\mathbf{s}_{i-1}), \quad \mathbf{s}_i = G_i(\widehat{\mathbf{x}}_i, \mathbf{s}_{i-1}), \tag{4}$$

where $\mathbf{s}_0 = \mathbf{y}$ is the input LR image and $\mathbf{s}_n$ eventually re-produces the HR image $\mathbf{x}$. Fig. 2(a) gives an overall

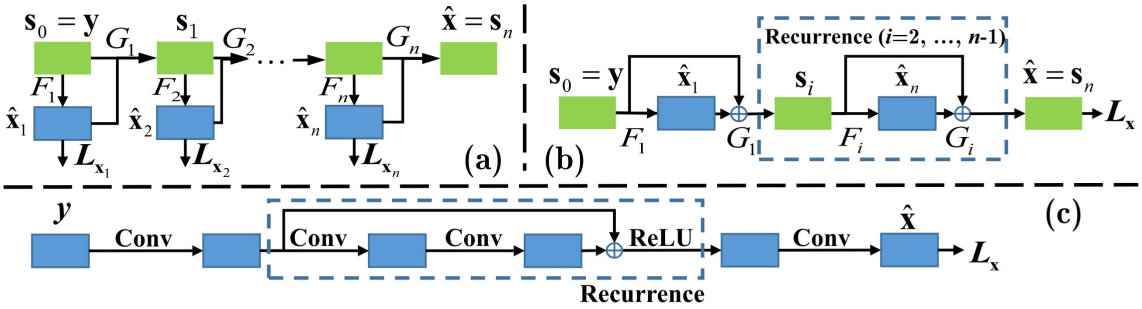


Fig. 2. (a) The flowchart of the sub-band reconstruction for image super-resolution. (b) A relaxed version of (a). G_i is set as the element-wise summation function. In this framework, only the MSE loss is used to constrain the recovery. (c) The deep network designed with the intuition of (b). G_i is the element-wise summation function and F_i is modeled by two layer convolutions.

illustration on this process. The functions F_i and G_i usually take linear transformations as advocated in [26] and [27]. F_i learns to recover high frequency detail, estimating the i-th sub-band component based on the accumulated recovered results from previous $(i-1)$ sub-bands. G_i fuses $\widehat{\mathbf{x}}_i$ and $\mathbf{s}_{i-1}$ in order to balance different sub-bands. In the figure, $\boldsymbol{L}_{\mathbf{x}_i}$ is the loss term corresponding to the data fidelity in (3). The progressive sub-band recovery can be learned in a supervised way [4], [26], where the ground truth sub-band signal $\mathbf{x}_i$ is generated by applying band filters on $\mathbf{x}$. In our proposed method, we choose the element-wise summation function to model G_i in the proposed network, following the additive assumption for the sub-bands of the image signal that is generally implied in previous methods [11], [26].

B. Learning Sub-Band Decomposition by Recurrent Residual Net

The sub-band paradigm mentioned above learns to recover HR images through minimizing a hierarchical loss generated by applying hand-crafted frequency domain filters, as shown in Fig. 2(a). However, this paradigm suffers from following two limitations. First, it does not provide an end-to-end trainable framework. Second, it suffers from the heavy dependence on the choice of the frequency filters. A bad choice of the filters would severely limit its capacity of modeling the correlation between different sub-bands, and recovering the HR $\mathbf{x}$.

To handle these two problems, by employing a summation function as G_i, we reformulate the recover process in (4) into:

$$\mathbf{s}_i = \mathbf{s}_{i-1} + F_i(\mathbf{s}_{i-1}). \tag{5}$$

In this way, the intermediate estimation $\widehat{\mathbf{x}}_i$ is not necessary to estimate explicitly. An end-to-end training paradigm can then be constructed as shown in Fig. 2(b). The MSE loss $\boldsymbol{L}_\mathbf{x}$ imposed at the top layer is the only constraint on $\widehat{\mathbf{x}}$ for the HR prediction. Motivated by (5) and Fig. 2(b), we further propose a recurrent residual learning network whose architecture is shown in Fig. 2(c). To increase the modeling ability, F_i is parameterized by two layers of convolutions. To introduce nonlinearities into the network, G_i is modeled by an element-wise summation connected with a non-linear rectification. Training the network to minimize the MSE loss gives the functions F_i and G_i adaptive to the training data. Then, we stack n recurrent units into a deep network to perform a progressive sub-band recovery. Our proposed recurrent residual network follows the intuition of gradual sub-band recovery process. The proposed model is equivalent to balancing the contributions of each sub-band recovery. Benefiting from the end-to-end training, such deep sub-band learning is more effective than the traditional supervised sub-band recovery. Furthermore, the proposed network indeed has the ability to recover the sub-bands of the image signal recurrently, as validated in Section V-F(3).

IV. DEGREE Network for Edge Preserving SR

We have presented how to construct a recurrent residual network to perform deep sub-band learning. In this section, we proceed to explain how to embed the edge guidance into the recurrent residual network, in order to predict high-frequency details better for image SR.

A. Edge Extraction

An HR image $\mathbf{x}$ can be separated into low-frequency and high-frequency components, as $\mathbf{x} = \mathbf{x}_L + \mathbf{x}_H$, where the high-frequency component $\mathbf{x}_H$ contains subtle details of the image, such as edges and textures. Patterns contained in $\mathbf{x}_H$ are usually irregular and have smaller magnitude compared with $\mathbf{x}_L$. Thus, in image degradation, the component of $\mathbf{x}_H$ is more fragile and easier to be corrupted, which is also difficult to recover. To better recover $\mathbf{x}_H$, we propose to extract extra prior knowledge about $\mathbf{x}_H$ from the LR image $\mathbf{y}$ as a build-in component in the deep recurrent residual network to regularize the recovery process. Among all the statistical priors about natural images, edge is one of the most informative priors. Therefore, we propose to model edge guidances and develop a deep edge guided recurrent residual network, which is introduced in the following section. However, our proposed network architecture can also embed other statistical priors extractable from LR inputs for image SR. To extract edges, we first apply an off-the-shelf edge detector (such as the Sobel one) on $\mathbf{y}$ and $\mathbf{x}$ to get its high-frequency component $\mathbf{y}_H$ and $\mathbf{x}_H$. Then we train the model to predict $\mathbf{x}_H$ based on both $\mathbf{y}$ and $\mathbf{y}_H$. Please note that $\mathbf{x}_H$ is the high-frequency residual of $\mathbf{x}$.

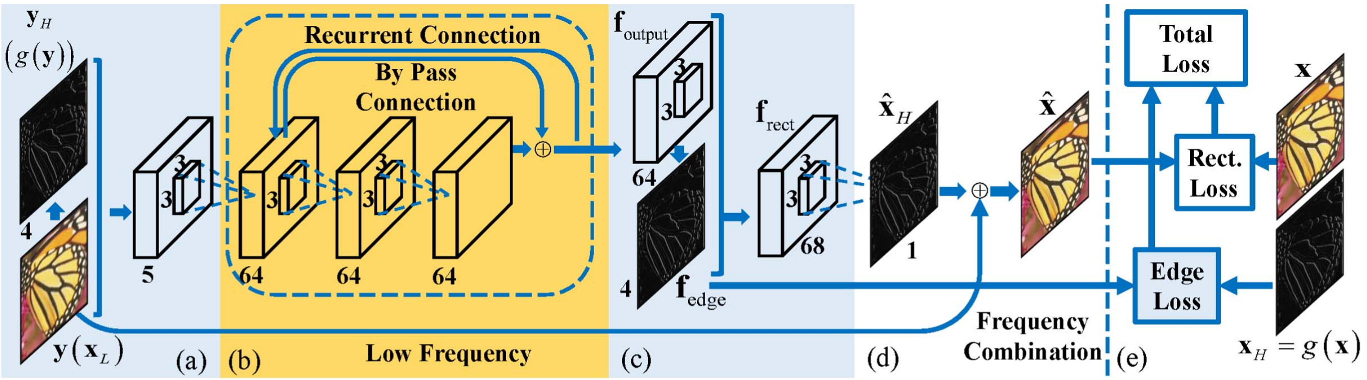


Fig. 3. The architecture of the DEGREE network for image SR. (a) The LR edge maps $\mathbf{y}_H(g(\mathbf{y}))$ of the LR image are part of the input features. (b) Recurrent residual learning network for sub-band recovery. (c) Part of the feature maps $\mathbf{f}_{\text{edge}}$ in the penultimate layer aim at generating HR edges. (d) Combining the high-frequency estimation and the LR image by $\widehat{\mathbf{x}} = \mathbf{x}_L + \widehat{\mathbf{x}}_H$. (e) The total loss is the combination of the edge loss and reconstruction loss, which constrain the recovery of HR edges and HR images respectively. Our main contributions, the edge guidance and recurrent residual learning, are highlighted with blue and orange colors.

B. DEGREE Network

We propose an end-to-end trainable deep edge guided recurrent residual network (DEGREE) for image SR. The network is constructed based on the following two intuitions. First, as we have demonstrated, a recurrent residual network is capable of learning sub-band decomposition and reconstruction for image SR. Second, modeling edges extracted from the LR image would benefit recovery of details in the HR image. An overview on the architecture of the proposed DEGREE network is given in Fig. 3. As shown in the figure, DEGREE contains following components. **a) LR Edge Extraction**. An edge map of the input LR image is extracted by applying a hand-crafted edge detector and is fed into the network together with the raw LR image, as shown in Fig. 3(a). **b) Recurrent Residual Network.** The mapping function from LR images to HR images is modeled by the recurrent residual network as introduced in Section III-B, Instead of predicting the HR image directly, DEGREE recovers the residual image at different frequency sub-bands progressively and combine them into the HR image, as shown in Fig. 3(b). **c) HR Edge Prediction**. DEGREE produces convolutional feature maps in the penultimate layer, part of which ($\mathbf{f}_{\text{edge}}$) are used to reconstruct the edge maps of the HR images and provide extra knowledge for reconstructing the HR images, as shown in Fig. 3(c). **d) Frequency Combination For Residue.** Since the LR image contains necessary low-frequency details, DEGREE only focuses on recovering the high-frequency component, especially several high-frequency sub-bands of the HR image, which are the differences or *residue* between the HR image and the input LR image. Combining the estimated residue with sub-band signals and the LR image gives an HR image, as shown in Fig. 3(d). **e) Training Loss.** We consider the reconstruction loss of both the HR image and HR edges simultaneously for training DEGREE as shown in Fig. 3(e). We now explain each individual part of the proposed network in details.

1) Recurrent Residual Network: The recurrent residual network aims to refine SR images progressively through producing the residue image at different frequency. We follow the notations in Section III. To provide a formal description, let $\mathbf{f}^k_{\text{in}}$ denote the input feature map for the recurrent sub-network at the k-th time step. The output feature map $\mathbf{f}^k_{\text{out}}$ of the recurrent sub-network is progressively updated as follows,

$$\begin{aligned}\mathbf{f}^k_{\text{out}} &= \max\left(0, \mathbf{W}^k_{\text{mid}} * \mathbf{f}^k_{\text{mid}} + \mathbf{b}^k_{\text{mid}}\right) + \mathbf{f}^k_{\text{in}},\\ \mathbf{f}^k_{\text{mid}} &= \max\left(0, \mathbf{W}^k_{\text{in}} * \mathbf{f}^k_{\text{in}} + \mathbf{b}^k_{in}\right),\end{aligned} \tag{6}$$

where $\mathbf{f}^k_{\text{in}} = \mathbf{f}^{k-1}_{\text{out}}$ is the output features by the recurrent sub-network at $(k-1)$-th time step. Please note the by-pass connection here between $\mathbf{f}^k_{\text{in}}$ and $\mathbf{f}^k_{\text{out}}$. In the context of sub-band reconstruction, the feature map $\mathbf{f}^k_{\text{out}}$ can be viewed as the recovered k-th sub-band of the image signal. Let K be the total recurrence number of the sub-networks. Then, the relation between $\mathbf{f}^1_{\text{in}}$, $\mathbf{f}^K_{\text{out}}$ and the overall network is

$$\begin{aligned}\mathbf{f}^1_{\text{in}} &= \max(0, \mathbf{W}_{\text{input}} * \mathbf{f}_{\text{input}} + \mathbf{b}_{\text{input}}),\\ \mathbf{f}_{\text{output}} &= \mathbf{f}^K_{\text{out}},\end{aligned} \tag{7}$$

where $\mathbf{W}_{\text{input}}$ and $\mathbf{b}_{\text{input}}$ denote the filter parameter and basis of the convolution layer before the recurrent sub-network. Thus, $\mathbf{f}_{\text{output}}$ is the output features of the recurrent residual network, which are used to reconstruct both the HR features and images.

2) Edge Modeling: We here illustrate how to embed the edge information into the proposed deep network. This can also generalize to modeling other natural image priors. In particular, the proposed network takes edge features extracted from the LR image as another input, and aims to predict edge maps of the HR image as a part of its output features which are then used for recovering the HR image.

The input feature $\mathbf{f}_{\text{input}}$ to the network is a concatenation of the raw LR image $\mathbf{y}$ and its edge map $g(\mathbf{y})$,

$$\mathbf{f}_{\text{input}} = \left[\mathbf{y}, g(\mathbf{y})\right]. \tag{8}$$

To recover the HR image, DEGREE outputs two types of features at its penultimate layer. One is for HR image recovery and the other one is for edge prediction in the HR image. More specifically, let $\mathbf{f}_{\text{output}}$ denote the features used to reconstruct HR images and let $\mathbf{f}_{\text{edge}}$ denote the edge feature computed by

$$\mathbf{f}_{\text{edge}} = \max\left(0, \mathbf{W}_{\text{edge}} * \mathbf{f}_{\text{output}} + \mathbf{b}_{\text{edge}}\right), \tag{9}$$

where $\mathbf{W}_{\text{edge}}$ and $\mathbf{b}_{\text{edge}}$ are the filter and the bias of the convolution layer to predict the HR edge map. Thus, the features $\mathbf{f}_{\text{recon}}$ ("recon" signifies reconstruction) in the penultimate layer for reconstructing the HR image with the edge guidance are given as follows,

$$\mathbf{f}_{\text{recon}} = \left[\mathbf{f}_{\text{output}}, \mathbf{f}_{\text{edge}}\right]. \tag{10}$$

3) Frequency Combination: In sub-band based image SR methods, the low-frequency and high-frequency components of an image signal are usually extracted at different parts in a hierarchical decomposition of the signal. DEGREE network also models the low-frequency and high-frequency components of an image jointly. Denote the high-frequency and low-frequency components of an HR image $\mathbf{x}$ as $\mathbf{x}_H$ and $\mathbf{x}_L$ respectively. We have $\mathbf{x} = \mathbf{x}_H + \mathbf{x}_L$. Here, we use the notation $\mathbf{y}$ to denote both the original LR image and its up-scaled version of the same size as $\mathbf{x}$, if it causes no confusion. Obviously, $\mathbf{y}$ is a good estimation of the low frequency component $\mathbf{x}_L$ of the HR image $\mathbf{x}$. The retained high-frequency component $\mathbf{y}_H$ of $\mathbf{y}$, *i.e.*, the edge map of $\mathbf{y}$, is estimated by applying an edge extractor (we use Sobel) onto $\mathbf{y}$. In our proposed DEGREE network, as shown in Fig. 3, the low-frequency component $\mathbf{x}_L \approx \mathbf{y}$ is directly passed to the last layer and combined with the predicted high-frequency image $\widehat{\mathbf{x}}_H$ to produce an estimation $\widehat{\mathbf{x}}$ of the HR image $\mathbf{x}$: $\widehat{\mathbf{x}} = \mathbf{x}_L + \widehat{\mathbf{x}}_H$. Here, $\widehat{\mathbf{x}}_H$, an estimation of the high-frequency component $\mathbf{x}_H$, is generated by

$$\widehat{\mathbf{x}}_H = \max\left(0, \mathbf{W}_{\text{recon}} * \mathbf{f}_{\text{recon}} + \mathbf{b}_{\text{recon}}\right), \tag{11}$$

where $\mathbf{f}_{\text{recon}}$ is the features learned in the penultimate layer to reconstruct $\mathbf{x}_H$. The filters and biases involved in the layer are denoted as $\mathbf{W}_{\text{recon}}$ and $\mathbf{b}_{\text{recon}}$.

4) Training: Let $\mathbf{F}(\cdot)$ represent the learned network for recovering the HR image $\mathbf{x}$ based on the input LR image $\mathbf{y}$ and the LR edge map $\mathbf{y}_H$. Let $\mathbf{F}_{\text{edge}}(\cdot)$ denote the learned HR edge predictor which outputs $\mathbf{f}_{\text{edge}}$. We use $\boldsymbol{\Theta}$ to collectively denote all the parameters of the network,

$$\begin{aligned}\boldsymbol{\Theta} = \big\{&\mathbf{W}_{\text{input}}, \mathbf{b}_{\text{input}}, \mathbf{W}_{\text{in}}, \mathbf{b}_{\text{in}}, \mathbf{W}_{\text{mid}}, \mathbf{b}_{\text{mid}}, \\ &\mathbf{W}_{\text{edge}}, \mathbf{b}_{\text{edge}}, \mathbf{W}_{\text{recon}}, \mathbf{b}_{\text{recon}}\big\}.\end{aligned} \tag{12}$$

Given n pairs of HR and LR images $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n}$ for training, we first extract the high-frequency components of LR and HR images, $\{\mathbf{y}_{i,H}\}$ and $\{\mathbf{x}_{i,H}\}$, by applying Sobel operator on the image $\mathbf{x}_i$ and $\mathbf{y}_i$ respectively. We adopt the following joint mean squared error (MSE) to train the network parameterized by $\boldsymbol{\Theta}$ such that it can jointly estimate the HR images and HR edge maps:

$$\begin{aligned}L(\boldsymbol{\Theta}) = \frac{1}{n}\sum_{i=1}^{n}(&||\mathbf{F}\left(\mathbf{y}_i, \mathbf{y}_{i,H}, \mathbf{x}_i, \mathbf{x}_{i,H}; \boldsymbol{\Theta}\right) - \mathbf{x}_i||^2 \\ &+ \lambda||\mathbf{F}_{\text{edge}}\left(\mathbf{y}_i, \mathbf{y}_{i,H}, \mathbf{x}_i, \mathbf{x}_{i,H}; \boldsymbol{\Theta}\right) - \mathbf{x}_{i,H}||^2).\end{aligned} \tag{13}$$

Here λ is a trade-off parameter that balances importance of the data fidelity term and the edge guidance term. We empirically set λ as 1 throughout the paper because we observe that our method performs similarly for different values of λ in a large range, as mentioned in Section V and validated in supplementary material.

V. Experiments

Datasets: Following the experimental setting in [8] and [34], we compare the proposed method with recent SR methods on three popular benchmark datasets: Set5 [3], Set14 [39] and BSD100 [22] with scaling factors of 2, 3 and 4. The three datasets contain 5, 14 and 100 images respectively. Among them, the Set5 and Set14 datasets are commonly used for evaluating traditional image processing methods, and the BSD100 dataset contains 100 images with diverse natural scenes. We train our model using a training set created in [37], which contains 91 images. For fair comparison with other methods [34], we do not train our models with a larger dataset.

Baseline Methods: We compare our DEGREE SR network (DEGREE) with Bicubic interpolation and the following six state-of-the-art SR methods: ScSR (Sparse coding) [37], A+ (Adjusted Anchored Neighborhood Regression) [31], SRCNN [8], TSE-SR (Transformed Self-Exemplars) [14], CSCN (Deep Sparse Coding) [34] and JSB-NE (Joint Sub-Band Based Neighbor Embedding) [27]. It is worth noting that CSCN and JSB-NE are the most recent deep learning and sub-band recovery based image SR methods respectively.

Implementation Details: We evaluate our proposed model with 10 and 20 layers respectively. The bypass connections are set with an interval of 2 convolution layers, as illustrated in Fig. 3. The number of channels in each convolution layer is fixed as 64 and the filter size is set as 3 × 3 with a padding size of 1. All these settings are consistent with the one used in [13]. The edge extractor is applied along four directions (up-down, down-up, left-right and right-left) for extracting edge maps. Following the experimental setting in [8], we generate LR images by applying Bicubic interpolation on the HR images. The training and validation images are cropped into small sub-images with a size of 33 × 33 pixels. We use flipping (up-down and left-right) and clockwise rotations ($0°, 90°, 180°$ and $270°$) for data augmentation. For each training image, 16 augmented images are generated. The final training set contains around 240,000 sub-images. The weighting parameter λ for balancing the losses is empirically set as 1. We empirically show that the DEGREE network is robust to the choice of λ in the supplementary material and the best performance is provided by setting $\lambda \leq 1$. Following the common practice in many previous methods, we only perform super-resolution in the luminance channel (in YCrCb color space). The other two chrominance channels are bicubically upsampled for displaying the results. We train our model on the Caffe platform [16]. Stochastic gradient descent (SGD) with standard back-propagation is used for training the model. In particular, in the optimization we set momentum as 0.9, the initial learning rate as 0.0001 and change it to 0.00001 after 76 epochs. We only allow at most 270 epochs. All the

TABLE I

COMPARISON AMONG DIFFERENT IMAGE SR METHODS ON THREE TEST DATASETS WITH THREE SCALE FACTORS (×2, ×3 AND ×4). THE BOLD NUMBERS DENOTE THE BEST PERFORMANCE AND THE UNDERLINED NUMBERS DENOTE THE SECOND BEST PERFORMANCE

Dataset		Set5			Set14			BSD100		
Method	Metric	×2	×3	×4	×2	×3	×4	×2	×3	×4
Bicubic	PSNR	33.66	30.39	28.42	30.13	27.47	25.95	29.55	27.20	25.96
	SSIM	0.9096	0.8682	0.8105	0.8665	0.7722	0.7011	0.8425	0.7382	0.6672
ScSR	PSNR	35.78	31.34	29.07	31.64	28.19	26.40	30.77	27.72	26.61
	SSIM	0.9485	0.8869	0.8263	0.8990	0.7977	0.7218	0.8744	0.7647	0.6983
A+	PSNR	36.56	32.60	30.30	32.14	29.07	27.28	30.78	28.18	26.77
	SSIM	0.9544	0.9088	0.8604	0.9025	0.8171	0.7484	0.8773	0.7808	0.7085
TSE-SR	PSNR	36.47	32.62	30.24	32.21	29.14	27.38	31.18	28.30	26.85
	SSIM	0.9535	0.9092	0.8609	0.9033	0.8194	0.7514	0.8855	0.7843	0.7108
JSB-NE	PSNR	36.59	32.32	30.08	32.34	28.98	27.22	31.22	28.14	26.71
	SSIM	0.9538	0.9042	0.8508	0.9058	0.8105	0.7393	0.8869	0.7742	0.6978
CNN	PSNR	36.34	32.39	30.09	32.18	29.00	27.20	31.11	28.20	26.70
	SSIM	0.9521	0.9033	0.8530	0.9039	0.8145	0.7413	0.8835	0.7794	0.7018
CNN-L	PSNR	36.66	32.75	30.49	32.45	29.30	27.50	31.36	28.41	26.90
	SSIM	0.9542	0.9090	0.8628	0.9067	0.8215	0.7513	0.8879	0.7863	0.7103
CSCN	PSNR	36.88	33.10	30.86	32.50	29.42	27.64	31.40	28.50	27.03
	SSIM	0.9547	0.9144	0.8732	0.9069	0.8238	0.7573	0.8884	0.7885	0.7161
CSCN-MV	PSNR	37.14	33.26	31.04	32.71	29.55	27.76	31.54	28.58	27.11
	SSIM	0.9567	0.9167	0.8775	0.9095	0.8271	0.762	0.8908	0.791	0.7191
DRCN	PSNR	**37.63**	**33.82**	**31.53**	33.04	29.76	28.02	31.85	28.80	27.23
	SSIM	**0.9588**	**0.9226**	**0.8854**	0.9118	0.8311	0.7670	0.8942	0.7963	0.7233
VDSR	PSNR	37.53	33.66	31.35	33.03	29.77	28.01	**31.90**	**28.82**	**27.29**
	SSIM	0.9587	0.9213	0.8838	0.9124	0.8314	**0.7674**	0.8960	**0.7976**	**0.7251**
DEGREE-1	PSNR	37.29	33.29	30.88	32.87	29.53	27.69	31.66	28.59	27.06
	SSIM	0.9574	0.9164	0.8726	0.9103	0.8265	0.7574	**0.8962**	0.7916	0.7177
DEGREE-2	PSNR	37.40	33.39	31.03	32.96	29.61	27.73	31.73	28.63	27.07
	SSIM	0.9580	0.9182	0.8761	0.9115	0.8275	0.7597	0.8937	0.7921	0.7177
DEGREE-3	PSNR	37.54	33.72	31.43	33.01	29.78	28.02	31.76	28.69	27.14
	SSIM	0.9584	0.9204	0.8818	0.9118	0.8317	0.7646	0.8939	0.7937	0.7200
DEGREE-MV	PSNR	37.61	33.70	31.30	**33.11**	29.77	27.92	31.84	28.76	27.18
	SSIM	0.9589	0.9212	0.8807	**0.9129**	0.8309	0.7637	0.8951	0.7956	0.7207
DEGREE-MV-2	PSNR	37.58	33.76	31.47	33.06	**29.82**	**28.10**	31.80	28.74	27.20
	SSIM	0.9587	0.9211	0.8837	0.9123	**0.8326**	0.7669	0.8947	0.7950	0.7216

experimental results related to GPU are obtained based on TITAN X.

A. Objective Evaluation

We use DEGREE-1, DEGREE-2, DEGREE-3, DEGREE-MV and DEGREE-MV-2 to denote three versions of the proposed model when we report the results. DEGREE-1 has 10 layers and 64 channels. DEGREE-2 and DEGREE-3 have 20 layers and 64 channels. DEGREE-1 and DEGREE-2 are trained by 91 images [8], and DEGREE-3 is trained by 291 images with scale augmentation [17]. The results of DEGREE-MV and DEGREE-MV-2 are generated by the multi-view testing strategy via fusing and boosting the results generated by DEGREE-2 and DEGREE-3, similar to CSCN-MV [34] to further investigate the effect of improving the quality of prior edge maps adopted in DEGREE on the final performance (see following texts for more details). In the multi-view testing strategy, we first generate different versions of LR images via geometric transformations, such as flipping (left to right, top to bottom), transpose and rotation $(90°, 180°, 270°$ in the clockwise direction). Then, their HR results are obtained by the enlargement by DEGREE-2. Finally, these HR results are transformed inversely and are fused together to give the final HR result. The quality of the HR images produced by different SR methods is measured by the Peak Signal-to-Noise Ratio (PSNR) [15] and the perceptual quality metric Structural SIMilarity (SSIM) [35], which are two widely used metrics in image processing. The results of our proposed DEGREE-1, DEGREE-2 and DEGREE-MV as well as the baselines are given in Table I.

From the table, it can be seen that the our proposed DEGREE models consistently outperform those well-established baselines with significant performance gains. DEGREE-MV performs the best for all the three scaling factors on the three datasets. Comparing the performance of DEGREE-1 and DEGREE-2 clearly demonstrates that increasing the depth of the network indeed improves the performance, but we observe that further increasing the depth leads to no performance gain. We also list the concrete performance gain brought by the proposed DEGREE-MV model over the state-of-the-art (CSCN-MV). One can observe that, our DEGREE-MV improves significantly over CSCN-MV. When enlarging the image by a factor of 2, our proposed method outperforms CSCN-MV with a rather large margin of 0.47dB (PSNR) and 0.0022 (SSIM) on Set5. For other scaling factors, our method still consistently performs better. For example, on the Set5 dataset, DEGREE-MV improves the performance by 0.45dB and 0.26dB for ×3 and ×4 settings respectively. Our models are more competitive for a small scale factor. This might be because edge features are more salient and are easier to be predicted in small scaling enlargements. This is also consistent with the observation

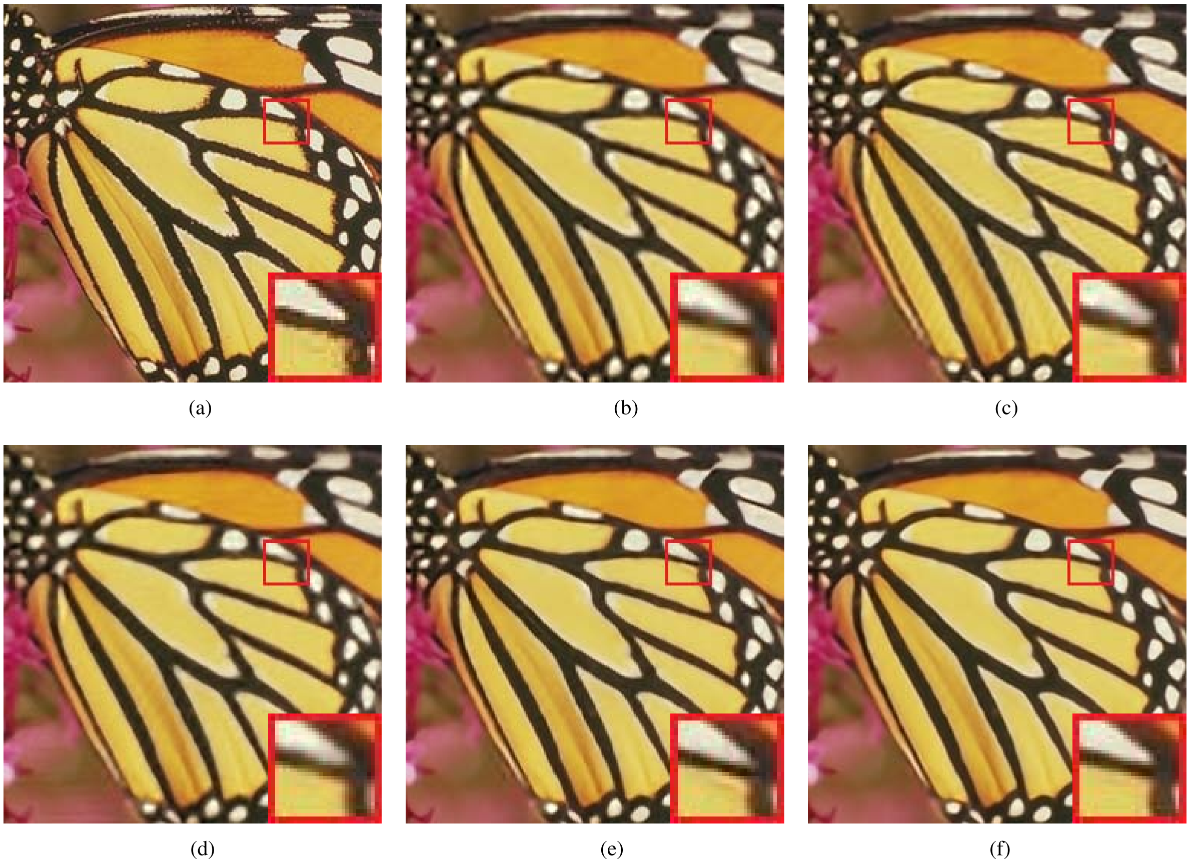


Fig. 4. Visual comparisons between different algorithms for the image *butterfly* (4×). The DEGREE avoids the artifacts near the corners of the white and yellow plaques, which are present at the results produced by other state-of-the-art methods.

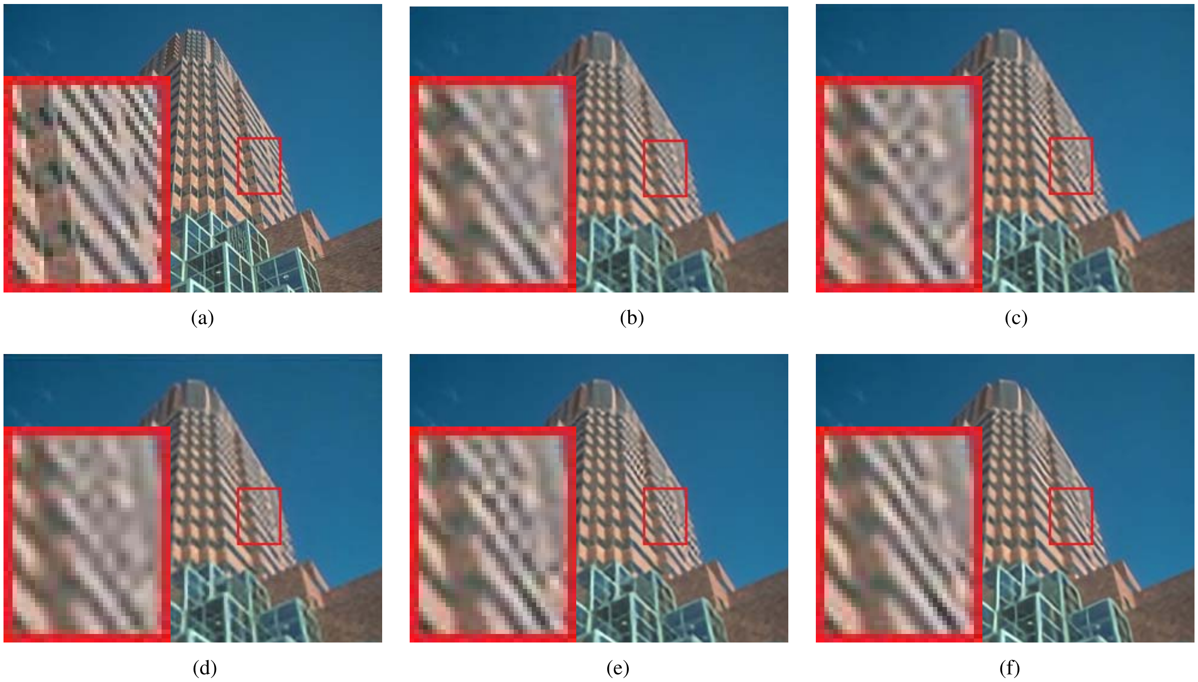
made for the gradient statistics in the previous edge-guided SR method [30]. It is worth noting that, although multi-view testing is a general post-processing strategy, it brings more significant performance gain for our DEGREE network than CSCN. And we also try to utilize it to boost the performance of VDSR and DRCN but obtain no gains. It is because that those two models are not designed for well utilizing the edge information. It also implies that it is not trivial to embedding the edge information into deep networks design is not trivial.

B. Subjective Evaluation

We also present some visual results in Figs. 4, 5 and 6 to investigate how the methods perform in terms of visual quality. These results are generated by our proposed network with 20 layers, *i.e.* DEGREE-2. Since our method is significantly better than baselines for the scaling factor of 2, here we in particular focus on comparing the visual quality of produced images with larger scaling factors. Fig. 4 displays the SR results on the image of *Butterfly* from Set5 for ×4 enlargement. From the figure, one can observe that the results generated by A+, SRCNN and JSB-NE contain artifacts or blurred details. CSCN provides fewer artifacts. But there are still a few remained, such as the corners of yellow and white plaques as shown in the enlarged local result in Fig. 4. Our method generates a more visually pleasant image with clean details and sharp edges. For the image *86000* from BSD100, as shown in Fig. 5, our method produces an image with the cleanest window boundary. For the image *223061* from BSD100 in Fig. 6 that contains a lot of edges and texture, most of methods generate the results with severe artifacts. Benefiting from explicitly exploiting the edge guidance, our method produces complete and sharp edges as desired. Note that more visual results are presented in the supplementary material due to space limitation.

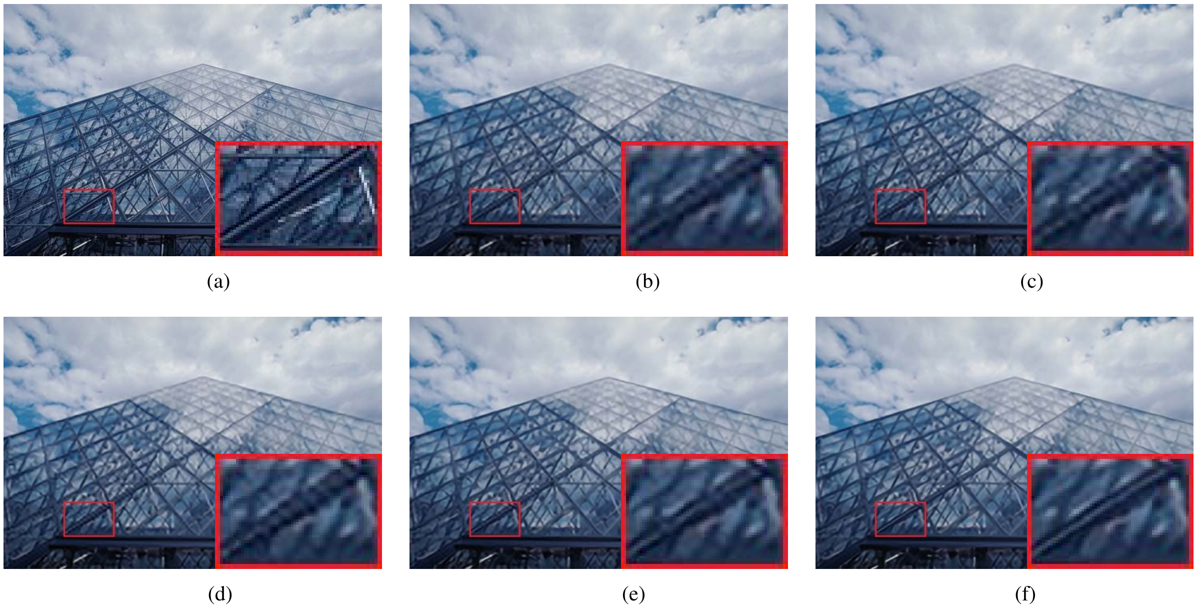
C. Effect of Loss Weighting Parameter

For the training loss (13), the loss weighting parameter λ balances the effect of the data fidelity and the edge enhancement functionality of the feature maps in the penultimate layer. A larger λ makes the feature maps in the penultimate layer better at reconstructing edges and high-frequency details, and meanwhile makes the network neglect details about main structures of HR images. We evaluate the performance of our proposed DEGREE network trained with different weighting parameter λ on the dataset Set5 with 91 images. The results are given in Table II. It shows that the performance of DEGREE-2 is robust to the choice of λ in a wide range. Setting $0.1 \leq \lambda \leq 1$ gives slightly better performance – the data fidelity term deserves a relatively larger weight than the edge prior modeling term for getting better HR images. In particular, totally ignoring the edge prior ($\lambda = 0$) leads to a degraded performance the performance drops from 37.39, 33.39 and 31.03dB to 37.25, 33.22 and 30.91 dB.


(a) (b) (c)

(d) (e) (f)

Fig. 5. Visual comparisons between different algorithms for the image *86000* (3×). The denotation 'D' signifies the results of our DEGREE network. The DEGREE presents less artifacts around the window boundaries.


(a) (b) (c)

(d) (e) (f)

Fig. 6. Visual comparisons between different algorithms for the image *223061* (3×). The DEGREE produces more complete and sharper edges. (a) High-res. (b) A+. (c) SRCNN. (d) TSE-SR. (e) CSCN. (f) DEGREE.

TABLE II

PERFORMANCE OF DEGREE (IN PSNR) ON SET5 WITH DIFFERENT VALUES OF THE WEIGHTING PARAMETER λ, WHERE S DENOTES THE SCALING FACTOR

λ	0	0.01	0.1	0.5	1	3	5	7
2	37.25	37.35	37.39	37.39	37.39	37.37	37.36	37.35
3	33.22	33.36	33.39	33.39	33.39	33.38	33.35	33.32
4	30.91	30.98	31.03	31.03	31.03	31.03	31	30.98

D. Running Time

We report time cost of our proposed model and compare its efficiency with other methods. Fig. 7 plots their running time (in secs.) against performance (in PSNR). All the compared methods are implemented using the public available codes from the authors. We implement our method using Caffe with its Matlab wrapper. We evaluate the running time of all the algorithms with following machine configuration: Intel X5675 3.07GHz, 24 GB memory and TITAN X GPU. The GPU version of our method (DEGREE-2) costs 1.81 seconds for performing SR on all the images of Set5, while other methods are significantly slower than ours in orders. The CPU version of our method (DEGREE-2) is comparable to other deep learning-based SR methods, including CSCN and CNN-L.

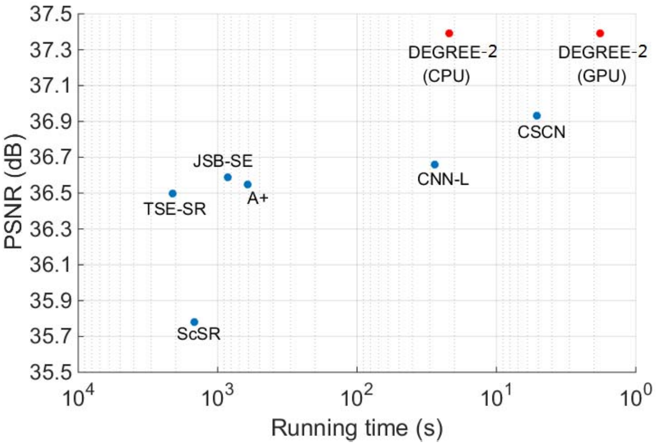


Fig. 7. The performance comparison between our proposed DEGREE model with state-of-the-art methods, including the final performance (y-axis) and time complexity (x-axis), in 2× enlargement on dataset Set5.

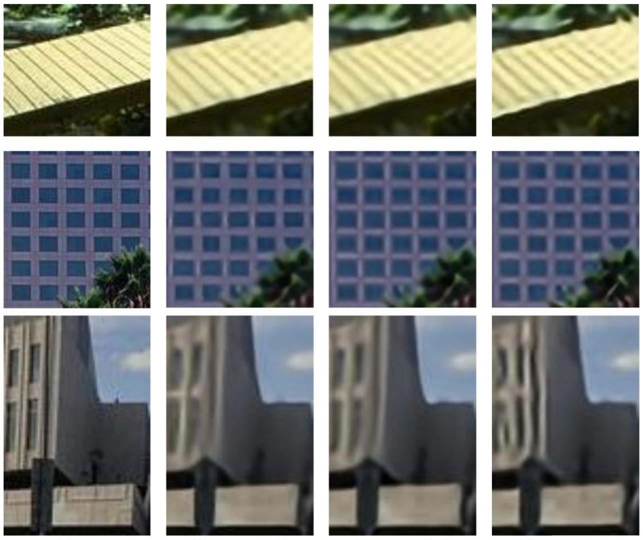

(a) HR (b) VDSR [17] (c) DRCN [18] (d) DEGREE-3

Fig. 8. Visual comparisons between VDSR, DRCN and DEGREE (4×). From top to bottom: No. *148026*, *119082* and *78004* in BSD100. DEGREE produces more complete and sharper edges.

E. Comparing With Very Deep Image SR

To illustrate the effectiveness of our DEGREE, we further compare with recent proposed very deep image SR approaches, VDSR [17] and DRCN [18], in both objective and subjective evaluations. Our previous versions of DEGREE are trained based on 91 images, which is different from VDSR [17]. For a fair comparison, we implement two new versions, denoted as DEGREE-3 and DEGREE-MV-2. Besides the training data, other configurations are exactly the same as DEGREE-2 and DEGREE-MV, respectively.

The performance comparison is reported in Table I and Fig. 8. In general, our approach is comparable to VDSR and DRCN in objective quality. Moreover, our method achieves better subjective quality with more accurate detail recovery and sharper edges as shown in Fig. 8. For objective evaluation, DRCN performs slightly better on Set5. However, it uses more parameters than our method. The channel numbers of DEGREE and DRCN are 64 and 256, respectively. But it just shows similar performance to our model and VDSR on Set5 and BSD100.

TABLE III

PSNR Results of Vanilla VDSR and That With Edge Guidance

Method		VDSR	VDSR+Edge	VDSR	VDSR+Edge
Layer		10L		20L	
2×	Set5	37.06	37.30	37.24	37.38
3×		33.22	33.28	33.34	33.35
4×		30.89	30.88	31.11	31.09
2×	Set14	32.71	32.8	32.79	32.9
3×		29.54	29.53	29.67	29.69
4×		27.7	27.69	27.87	27.88
2×	BSD100	31.62	31.7	31.66	31.74
3×		28.59	28.61	28.6	28.63
4×		27.06	27.07	27.11	27.09

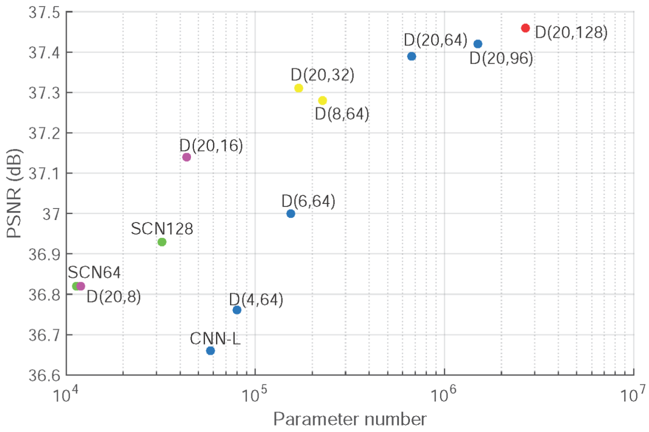


Fig. 9. PSNR for 2× SR on Set5 with various parameter numbers, compared with CSCN and CNN. The denotation 'D' signifies the results of our DEGREE network. The plot clearly demonstrates the high parameter-efficiency of our proposed DEGREE model.

F. Evaluation on VDSR With Edge Guidance

To evaluate the effectiveness of the proposed edge guidance, we also compare VDSR with VDSR embedded with edge guidance. For a fair comparison, the results presented are based on the model we trained because of the necessary retraining when adding edge guidance. Our training totally follows the details in [17]. The PSNR results (dB) are presented in Table III.

It is observed that, the edge guidance benefits the performance more in small scale enlargement (2×) and in the shallow case (10L). In small scaling enlargements, edge features are more salient and are easier to be predicted [30]. The performance gap between the shallower and deeper cases is caused by that edge guidance and deeper layer have the overlapped functionality for the detail recovery. Thus, the larger gain in the shallower case (10L) also proves the effectiveness of the proposed edge guidance.

G. Discussions

We further provide additional investigations on our model in depth, aiming to give more transparent understandings on its effectiveness.

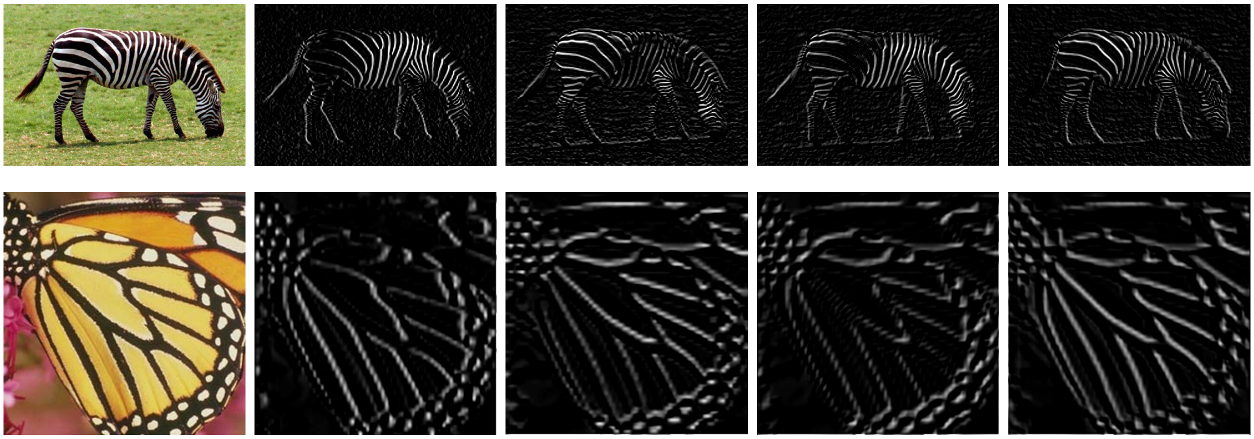


Fig. 10. The visualization of the learned edges. From top to bottom: *Zebra* (3× enlargement) in Set14 and *Butterfly* (4× enlargement) in Set5. From left to right: the HR image and four recovered edge maps.

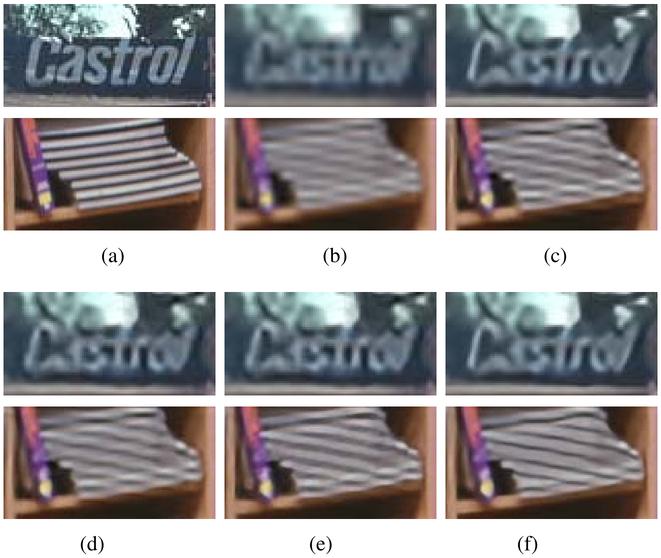


Fig. 11. Failure case examples in 4× enlargement. Top panel in each subfigure: fail to recover the correct letters. Bottom panel in each subfigure: reconstructed edges are in the wrong direction.

1) Model Size: We investigate how the size of the model, including number of layers and size of channels within each layer, influences the final performance. We compare performance of our model with different pairs of (# layers, # channels) in Fig. 9. It can be seen that a large model with more than $(20, 32) \times 10^5$ and $(8, 64) \times 10^5$ parameters (shown as yellow points) is necessary for achieving reasonably good performance. The combination of $(20, 8) \times 10^4$ (the purple point) results in a model with the same size of SCN64 (the green point where its dictionary size is equal to 64) and achieves almost the same performance. Further increasing the model size to $(20, 16) \times 10^4$ (the higher purple point) gives a better result than SCN128 (with a dictionary size of 64), whose model size is slightly smaller.

2) Visualization of Recovered Edge Through Learning: We visualize the learned edge in the penultimate layer in Fig. 10. DEGREE successfully reconstructs the edge maps in four directions.

3) Failure Case Analysis: Our proposed DEGREE as well as previous methods may fail in the cases when performing SR with a large scaling factor (4×) on the LR images conveying limited edge guidance, as illustrated in the two examples of Fig. 11. For observing, the information loss and changes of edge characteristics (*e.g.* edge direction changes) make almost all SR methods fail to recover the accurate HR details. Compared with previous methods, our DEGREE still generates clearer and sharper details, which provide better visual quality. To handle this problem in the future, the edge statistic across the scale needs to be modeled and some content semantic information should be utilized to construct more effective and targeted priors.

VI. Conclusions and Future Work

In this paper, we proposed a deep edge guided recurrent residual network for image SR. The edge information is separated out from the image signal to guide the recovery of the HR image. The extracted LR edge maps are used as parts of the input features and the HR edge maps are utilized to constrain the learning of parts of feature maps for image reconstruction. The recurrent residual learning structure with by-pass connections enables the training of deeper networks. Extensive experiments have validated the effectiveness of our method for producing HR images with richer details. Furthermore, this paper presented a general framework for embedding various natural image priors into image processing tasks. The experimental results for JPEG artifacts reduction demonstrate the good generality and flexibility of our proposed DEGREE network. Although achieving very promising results for both image SR and JPEG artifacts reduction, our DEGREE model still has some limitations that motivate our future work. First, the model has a very deep network structure, which makes the testing phrase rather time-consuming. It is our next aim to compress the model, decreasing the network layer and parameter number, and further speedup our algorithm to facilitate real applications. Second, our performance gain will decrease in the large scale enlargement. To provide the obvious edge guidance in the case of a large scaling factor is also our future research goal. Third, the performance gain from the multi-view testing shows the potential of multi-orientation information for edge guidance-directed methods. It is an interesting question to create an end-to-end trainable network that embeds both edge guidances and multi-orientation predictions automatically instead of an offline boosting.

References

[1] H. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1647–1659, Oct. 2005.

[2] S. P. Belekos, N. P. Galatsanos, and A. K. Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1451–1464, Jun. 2010.

[3] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 135.

[4] P. Chatterjee, V. P. Namboodiri, and S. Chaudhuri, "Super-resolution using sub-band constrained total variation," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.*, 2007, pp. 616–627.

[5] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[7] 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. 2016.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. IEEE Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 184–199.

[9] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[10] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Proc. SPIE's 48th Annu. Meet. Int. Symp. Opt. Sci. Technol.*, vol. 5203. San Diego, CA, USA, Aug. 2003, pp. 121–130.

[11] A. Gholamreza and D. Hasan, "Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image," *ETRI J.*, vol. 32, no. 3, pp. 390–394, Jun. 2010.

[12] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, Jan. 2009.

[13] K. He, X. Zhang, S. Ren, and J. Sun. (2015). "Deep residual learning for image recognition." [Online]. Available: https://arxiv.org/abs/1512.03385

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

[15] 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, Jun. 2008.

[16] Y. Jia *et al.* (2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/1408.5093

[17] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654 .

[18] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1637–1645.

[19] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE CVPR*, Jun. 2011, pp. 2657–2664.

[20] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 115–134, 2007.

[21] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, Dec. 2008.

[22] 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. (ICCV)*, vol. 2. Jul. 2001, pp. 416–423.

[23] C. Osendorfer, H. Soyer, and P. Smagt, "Image super-resolution with fast approximate convolutional sparse coding," in *Proc. Int. Conf. Neural Inf. Process.*, 2014, pp. 250–257.

[24] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Proc. IEEE Eur. Conf. Comput. Vis.*, Aug. 2008, pp. 57–68.

[25] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2006, pp. 1089–1096.

[26] A. Singh and N. Ahuja, "Sub-band energy constraints for self-similarity based super-resolution," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 4447–4452.

[27] S. Song, Y. Li, J. Liu, and Z. Quo, "Joint sub-band based neighbor embedding for image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 1661–1665.

[28] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[29] 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.

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

[31] 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.*, Nov. 2015, pp. 111–126.

[32] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[33] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[34] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 370–378.

[35] 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.

[36] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2012, pp. 341–349.

[37] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[38] G. Ye, Y. Wang, J. Xu, G. Herman, and B. Zhang, "A practical approach to multiple super-resolution sprite generation," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct. 2008, pp. 70–75.

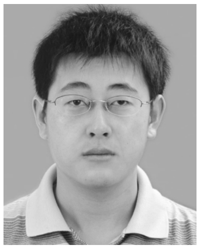
[39] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Conf. Curves Surfaces*, 2012, pp. 711–730.

[40] 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.

[41] Q. Zhou, S. Chen, J. Liu, and X. Tang, "Edge-preserving single image super-resolution," in *Proc. ACM Trans. Multimedia*, 2011, pp. 1037–1040.

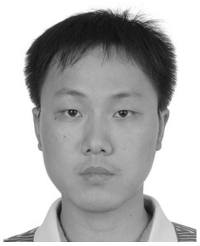
[42] W. Zuo, L. Zhang, C. Song, D. Zhang, and H. Gao, "Gradient Histogram Estimation and Preservation for Texture Enhanced Image Denoising," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2459–2472, Jun. 2014.

Wenhan Yang received the B.S degree in computer science from Peking University, Beijing, China, in 2012, where he is currently pursuing the Ph.D. degree with the Institute of Computer Science and Technology. He was a Visiting Scholar with the National University of Singapore, from 2015 to 2016. His current research interests include image processing, sparse representation, image restoration, and deep learning-based image processing.



Jiashi Feng received the Ph.D. degree from the National University of Singapore (NUS) in 2014. He was a Post-Doctoral Research Follow with UC Berkeley. He joined NUS as a faculty member, where he is currently an Assistant Professor with the Department of Electrical and Computer Engineering. His research areas include computer vision, machine learning, object recognition, detection, segmentation, robust learning, and deep learning.

Jianchao Yang (M'12) received the M.S. and Ph.D. degrees from the ECE Department, University of Illinois at Urbana–Champaign, under the supervision of Prof. T. Huang. He was a Research Scientist with Adobe Research. He is currently a Lead Research Scientist with Snapchat Inc. He has authored over 80 technical papers over a wide variety of topics on top tier conferences and journals, with Google scholar citation over 12 000 times. His research focuses on computer vision, deep learning, and image and video processing. He received the Best Student Paper award from ICCV 2010, the Classification Task Prize in PASCAL VOC 2009, first position for object localization using external data for ILSVRC ImageNet 2014, and third place in the WebVision Challenge 2017. He serves as the Workshop Chair of the ACM MM 2017.



Fang Zhao received the B.Sc. degree in telecommunication engineering from the Xidian University, Xi'an, China, in 2009, the M.Sc. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2012, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2015.

He has been a Research Fellow with the Learning and Vision Laboratory, National University of Singapore, Singapore, since 2015. His current research interests include deep learning, recurrent neural network, face recognition, and super-resolution.

Jiaying Liu (S'08–M'10–SM'17) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, in 2005, and the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2010, respectively.

She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored over 90 technical articles in refereed journals and proceedings, and holds 19 granted patents. Her current research interests include image/video processing, compression, and computer vision.

She was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She was a Visiting Researcher with Microsoft Research Asia in 2015, supported by the Star Track for Young Faculties. She is a CCF Senior Member. She has also served as a TC member in the IEEE CAS MSA and APSIPA IVM, and APSIPA Distinguished Lecture from 2016 to 2017.

Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication.

Dr. Guo is currently the Executive Member of the China-Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award in 2008, the Chia Tai Teaching Award in 2008, the Government Allowance granted by the State Council in 2009, and the Distinguished Doctoral Dissertation Advisor Award of Peking University in 2012 and 2013.

Shuicheng Yan (F'16) is currently the Vice President and the Chief Scientist of Qihoo 360 Technology Company Ltd., and the Head of the 360 Artificial Intelligence Institute. He is also a tenured Associate Professor with the National University of Singapore. His research areas include computer vision, machine learning, and multimedia analysis. He has authored/co-authored about 500 high quality technical papers, with Google Scholar citation over 25 000 times and an h-index 70. He is an IAPR Fellow and the ACM Distinguished Scientist. His team received seven times winner or honorable-mention prizes in five years over PASCAL, VOC, and ILSVRC competitions, which are core competitions in the field of computer vision, along with over ten times best (student) paper awards and especially a Grand Slam at the ACM MM, the top conference in the field of multimedia, including the Best Paper Award, the Best Student Paper Award, and the Best Demo Award. He is a TR Highly Cited Researcher of 2014, 2015, and 2016.