Soft-Edge Assisted Network for Single Image Super-Resolution

Faming Fang, Juncheng Li, and Tieyong Zeng

***Abstract*— The task of single image super-resolution (SISR) is a highly ill-posed inverse problem since reconstructing the high-frequency details from a low-resolution image is challenging. Most previous CNN-based super-resolution (SR) methods tend to directly learn the mapping from the low-resolution image to the high-resolution image through some complex convolutional neural networks. However, the method of blindly increasing the depth of the network is not the best choice because the performance improvement of such methods is marginal but the computational cost is huge. A more efficient method is to integrate the image prior knowledge into the model to assist the image reconstruction. Indeed, the soft-edge has been widely applied in many computer vision tasks as the role of an important image feature. In this paper, we propose a Soft-edge assisted Network (SeaNet) to reconstruct the high-quality SR image with the help of image soft-edge. The proposed SeaNet consists of three sub-nets: a rough image reconstruction network (RIRN), a soft-edge reconstruction network (Edge-Net), and an image refinement network (IRN). The complete reconstruction process consists of two stages. In Stage-I, the rough SR feature maps and the SR soft-edge are reconstructed by the RIRN and Edge-Net, respectively. In Stage-II, the outputs of the previous stages are fused and then fed to the IRN for high-quality SR image reconstruction. Extensive experiments show that our SeaNet converges rapidly and achieves excellent performance under the assistance of image soft-edge. The code is available at https://gitlab.com/junchenglee/seanet-pytorch.**

***Index Terms*— Edge assistance, soft-edge, convolutional neural network, single image super-resolution, image restoration.**

I. Introduction

SINGLE image super-resolution (SISR) is an extremely hot topic in the field of computer vision, which aims to reconstruct a super-resolution (SR) image from a single low-resolution (LR) one (Fig. 1). It has been widely used in computer vision tasks such as medical image enhancement [1], [2], video super-resolution [3], [4], and facial illusion [5], [6]. Meanwhile, the quality of reconstructed images significantly affects the accuracy of high-level tasks, such as image classification, objective detection, and image segmentation. Although SISR has a wide range of applications, it is still considered as a highly ill-posed problem due to information loss.

Recently, convolutional neural networks (CNNs) have achieved remarkable success in many computer vision tasks and greatly promoted the development of SISR. To handle the SISR problem, plenty of CNN-based models have been investigated, including [7]–[17]. Among them, Dong *et al.* [7] proposed the Super-Resolution Convolutional Neural Network (SRCNN), which was the first successful model adopting CNN to the SISR problem and achieved significant improvement. Later, Kim et al. [10] extended the depth of the network and addressed the VDSR with residual learning. The aforementioned models both used the preprocessed LR image as input and amplify it to the HR dimension by a Bicubic interpolation. However, it was argued that using the preprocessed LR image as input will increase computational complexity and produce visible artifacts [9]. Therefore, ESPCN [9] introduced an efficient sub-pixel convolutional layer that can learn an array of upscaling filters to directly upscale the final LR feature maps into the SR image. FSRCNN [8] adopted a deconvolutional layer at the end of the model to directly learn the mapping between the original LR and SR images.

After that, CNN-based SR models have been blooming and constantly refreshing the best results. For example, Kim *et al.* [11] investigated a deeply-recursive convolutional network for SISR, which introduced recursive-supervision learning to reduce model parameters; Lai *et al.* [13] proposed a Laplacian pyramid network to progressively reconstruct the sub-band residuals of HR images; Li *et al.* [16] considered a multi-scale residual network to extract rich image features for high-quality SR images reconstruction; He *et al.* [18] used the residual blocks to build an extremely wide and deep EDSR network, which achieved state-of-the-art results. However, there are still two potential problems in the aforementioned SR models: (1) Since all these methods tend to use deeper networks, the training of these networks becomes more difficult and requires more training data, time, and memory. (2) Most CNN-based SR models directly learn the mapping between LR and HR images by minimizing some loss function. However, it is rather difficult for them to reconstruct realistic high-frequency details due to the lack of prior knowledge of natural images.

Manuscript received August 6, 2018; revised December 19, 2018 and November 11, 2019; accepted February 9, 2020. Date of publication February 24, 2020; date of current version February 27, 2020. The work of Faming Fang was supported by the Key Project of the National Natural Science Foundation of China under Grant 61731009, in part by the National Natural Science Foundation of China under Grant 61871185, and in part by the "Chenguang Program" supported by the Shanghai Education Development Foundation and Shanghai Municipal Education Commission under Grant 17CG25. The work of Tieyong Zeng was supported by the NSFC under Grant 11671002, in part by the CUHK Start-Up, and in part by the CUHK DAG under Grant 4053296, Grant 4053342, Grant RGC 14300219, and Grant NSFC/RGC N_CUHK 415/19. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaolin Wu.

Faming Fang and Juncheng Li are with the Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200062, China, and also with the School of Computer Science and Technology, East China Normal University, Shanghai 200062, China (e-mail: fmfang@cs.ecnu.edu.cn; cvjunchengli@gmail.com).

Tieyong Zeng is with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong (e-mail: zeng@math.cuhk.edu.hk).

Digital Object Identifier 10.1109/TIP.2020.2973769

1057-7149 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

Fig. 1. An example SR result of our SeaNet (×4). (Lower left corner: LR image; Left: SR image (SeaNet); Right: HR image (ground-truth).

Various previous works have pointed out that the application of prior knowledge of images can effectively assist image reconstruction. Accordingly, many image priors have been proposed for image reconstruction, among which the total variation (TV) prior [19], [20], sparse prior [21]–[23], and edge prior [24], [25] are extremely popular. Inspired by these studies, recent works [26]–[28] have attempted to introduce TV or sparse priors into deep neural networks for SR image reconstruction. However, the TV prior will smooth texture details in the restored images and the sparse priors are usually difficult to model because it requires other domain knowledge. After that, Yang *et al.* [29] integrated the edge prior with recursive networks and proposed the Edge Guided Recurrent Residual Network for SISR. Although those methods have introduced images priors into the deep networks, how to accurately extract image priors and how to effectively use the image priors to assist image reconstruction are still challenging.

It is universally acknowledged that an image can be divided into two components: low-frequency and high-frequency ones. The low-frequency component refers to the region where the intensity of the image is varying gently, that is, where the large color patches are located. The high-frequency component refers to the region where the intensity of the image changes drastically, that is, the edge of the image. Driven by these facts, we aim to explore a unified framework that can automatically extract and integrate the soft-edge prior for image super-resolution. To achieve this, we build a Soft-edge assist Network (SeaNet) for image super-resolution, which consists of a rough image reconstruction network (RIRN), a soft-edge reconstruction network (Edge-Net), and an image refinement network (IRN). Specifically, the reconstruction process has two stages. In Stage-I, we reconstruct the rough image feature maps and the image soft-edge by the RIRN and Edge-Net, respectively. Both sub-nets take the LR image as input and obtain the corresponding maps in high-dimensional space. In Stage-II, we first fuse the reconstructed image feature maps and the image soft-edge by a bottleneck layer. Then we send the fused feature maps to the IRN for the final SR image reconstruction. All of these subnets constitute the complete Soft-edge assisted Network (SeaNet). Extensive experiments and ablation analysis demonstrate that under the assistance of soft-edge prior, our SeaNet can converge quickly and reconstruct realistic SR images with high-frequency details. The main contributions are as follows:

(i). We propose a soft-edge reconstruction network (Edge-Net), which is the first CNN model used to reconstruct the image soft-edge directly from the LR image. The Edge-Net can work independently for the image soft-edge reconstruction, or be embedded as a subnet into any SR model to provide image soft-edge prior for high-quality SR image reconstruction.

(ii). We propose an efficient and accurate Soft-edge assisted Network (SeaNet), which is a well-designed network that introduces the Edge-Net to provide image soft-edge prior.

The rest of this paper is organized as follows. Related works are reviewed in Section II. A detailed explanation of the proposed SeaNet is given in Section III. The experimental results and ablation analysis are presented in Section IV and V, respectively. Finally, we draw a conclusion in Section VI.

II. Related Works

A. Single Image Super-Resolution

Image super-resolution, especially single image super-resolution (SISR) has been extremely popular in the past several decades. Thorough reports can be found in [30] and [31]. In this paper, we focus on the SISR task and all comparison methods are based on SISR.

The development of SISR can be simply divided into two-stages: (1) Earlier methods used the interpolation techniques based on the sampling theory, such as linear or Bicubic methods, which were fast and flexible. However, these methods failed to reconstruct high-frequency texture details, thus can not reconstruct realistic SR images. (2) Later, more powerful methods based on the learning techniques have been developed to establish a complex mapping between LR and HR images. The sparsity-based SR methods [32], [33] assume that any natural image can be sparsely represented in a dictionary learned from a database. Neighborhood embedding methods [34], [35] upsample a LR image patch by finding similar patches in a low dimensional manifold and combining their corresponding HR patches for the SR image reconstruction. In addition to the methods mentioned above, other learning-based methods have been proposed to learn the mapping between LR and HR images, including the convolutional neural network (CNN) [7] and random forest [36]. Among them, CNN-based methods have shown outstanding performance and become the current mainstream means. Most CNN-based methods tend to use bigger and deeper network to reconstruct SR images. However, these models often gain marginal improvement while they fail to reconstruct high-frequency details and result in significant resource overhead. To solve this problem, we aim to explore an efficient and accurate SR model that can introduce image priors for high-frequency feature learning.

B. Edge-Assisted Image Reconstruction

In the past few decades, image reconstruction based on image priors has grown tremendously. Among all image priors,

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

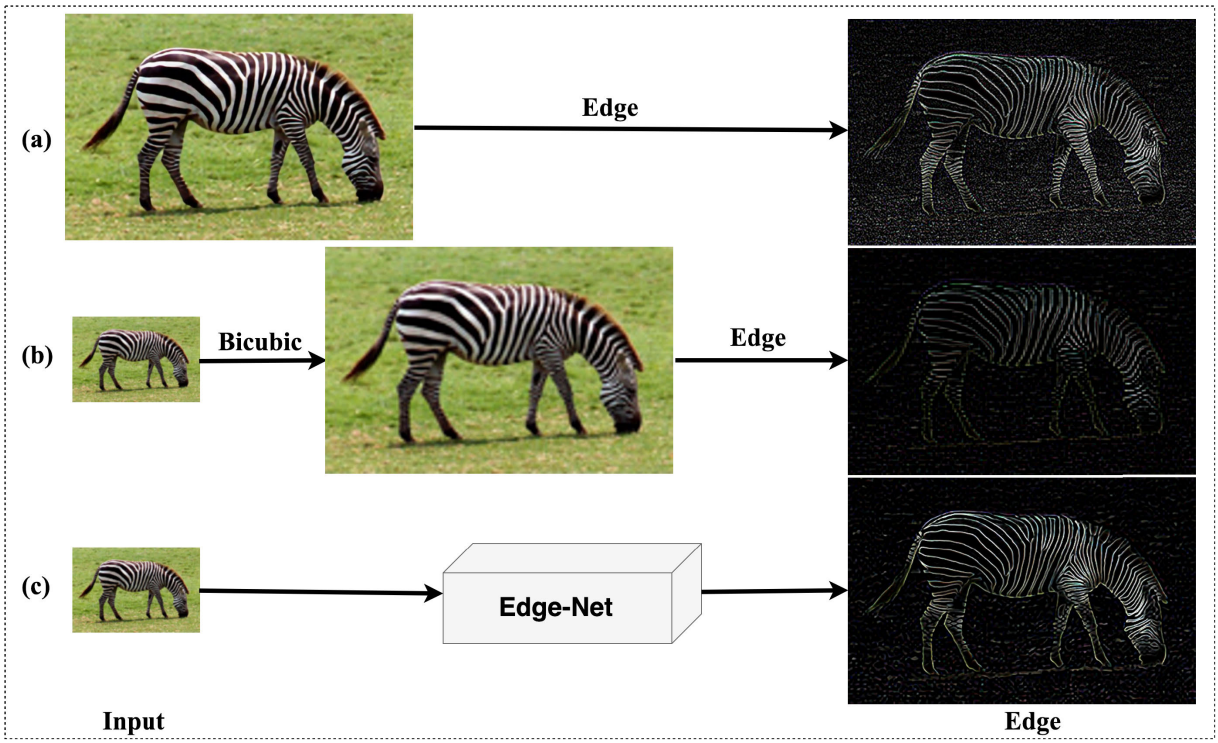

Fig. 2. (a) The image soft-edge directly extracted from the HR image. (b) The image soft-edge extracted from the upsampled LR image. (c) The image soft-edge reconstructed by the Edge-Net (ours). The SR soft-edge extracted by method (b) is blurred and carries lots of noise, while the soft-edge extracted by method (c) is clear and sharp.

image edge prior is one of the most effective priors since image edges are important image features. Extensive edge-assisted or edge-guided image processing methods [24], [25] have verified the feasibility and necessity of the image edge prior. However, the implementation of these methods is complex and exist certain limitations. For instance, Yang *et al.* [29] introduced the image edges into the CNN model and proposed the Edge Guided Recurrent Residual (DEGREE) Network. As a ground breaking job, this method still has some shortcomings resulting in sub-optimal performance, e.g. (a) DEGREE uses Bicubic preprocessed LR images as input. This will bring extra noise and produce visible artifacts; (b) DEGREE applies an off-the-shelf edge detector (e.g. Sobel detector) on the preprocessed LR image to get the image edges. This may introduce additional noise and cause blurred image edges (Fig. 2 (b)); (c) DEGREE directly adds the learned image edge feature to the LR image to obtain the final SR image; This is essentially a residual learning and the addition method can not maximize the usage of image edge prior; (d) DEGREE is a recurrent network. The recurrent mechanism can reduce model parameters but can not reduce the execution time. In addition, training a recurrent network needs more training tricks. In order to solve the aforementioned issues and make full use of image edge prior, we aim to explore an efficient edge reconstruction network (Fig. 2 (c)) to directly reconstruct clear image edges from the LR image and build a two-stage framework to fully use image edge prior for SR image reconstruction.

C. Image Soft-Edge

The points where the brightness of an image changes drastically are usually organized into a set of curve segments called image edges. A variety of methods have been raised to extract image edges, including **Sobel**, **Prewitt**, **Roberts**, and **Canny**. However, we find that these methods have some limitations and are difficult to apply to arbitrary images. Meanwhile, these off-the-shelf edge detectors use the binarization measurement to convert all the values of the edges to 0 and 1, which results in the loss of a great number of image features and the appearance of false edges. To avoid these problems, we suggest using image soft-edge instead of image edge. The soft-edge is acquired by eliminating the binarization strategy in order to retain accurate image edge information. Detailed acquisition methods will be introduced in Section III-B.

III. Soft-Edge Assisted Network

As shown in Fig. 3, our Soft-edge assisted Network (SeaNet) includes a rough image reconstruction network (RIRN), a soft-edge reconstruction network (Edge-Net), and an image refinement network (IRN). Specifically, the SeaNet can be divided into two stages. In Stage-I, we use RIRN to extract low-frequency features from the LR image and reconstruct rough SR feature maps. In addition, we build an efficient Edge-Net to directly reconstruct clear and sharp super-resolution soft-edge from the LR image. In Stage-II, the outputs of Stage-I are concatenated and fused by a bottleneck layer. Then the fused image features are sent to the IRN for the final SR image reconstruction. We define I_{LR} and I_{SR} as the input and output of SeaNet, respectively. Besides, f_{rough} and f_{edge} represent the output of the RIRN and the Edge-Net, respectively. Therefore, the rough SR image feature maps extraction and the image soft-edge reconstruction in Stage-I

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

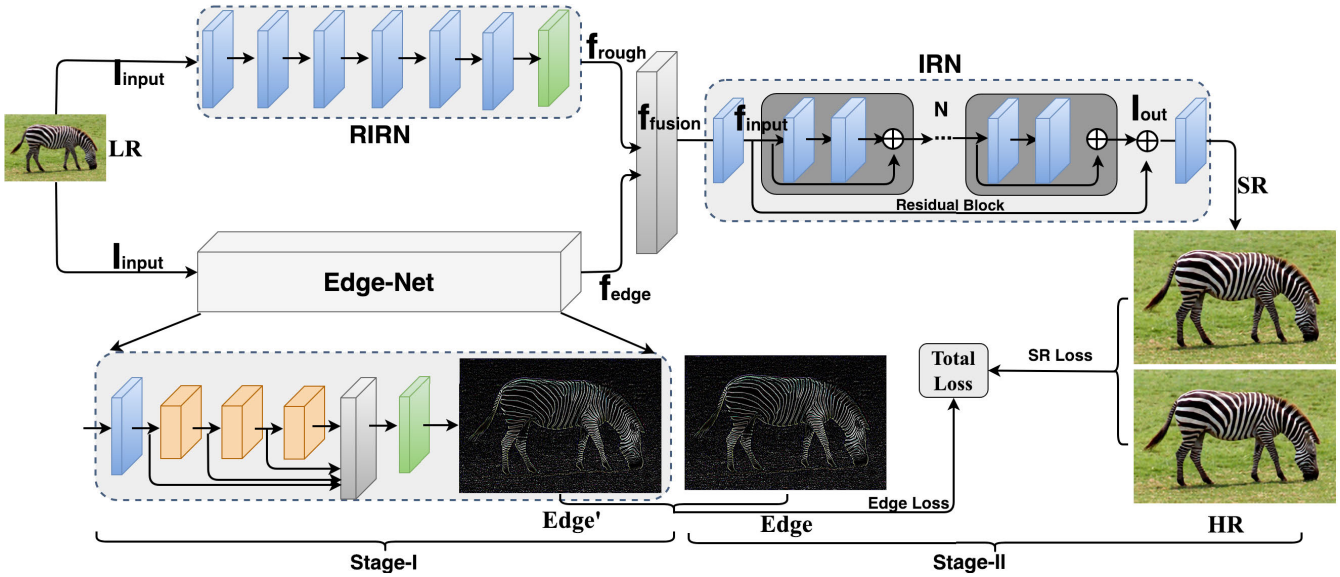

Fig. 3. The complete architecture of SeaNet. The model can be divided into three parts: rough image reconstruction network (RIRN), soft-edge reconstruction network (edge-net), and image refinement network (IRN). The blue block, green box, and gray box denote the convolutional layer, upsample module, and bottleneck layer, respectively.

can be expressed as:

$$\mathrm{f_{rough}} = F_{RIRN}(\mathrm{I_{LR}}), \tag{1}$$

$$\mathrm{f_{edge}} = F_{Edge-Net}(\mathrm{I_{LR}}), \tag{2}$$

where $F_{RIRN}(\cdot)$ and $F_{Edge-Net}(\cdot)$ denote the RIRN and Edge-Net, respectively. The outputs $\mathrm{f_{rough}}$ and $\mathrm{f_{edge}}$ represent the low-frequency images features (the rough SR feature maps) and the high-frequency image features (the image soft-edge), respectively. They are merged through a fusion layer

$$\mathrm{f_{fusion}} = F_{fusion}([\mathrm{f_{rough}}, \mathrm{f_{edge}}]), \tag{3}$$

where [ ] is the concatenate operation and $F_{fusion}(\cdot)$ denotes the fusion layer. The fusion layer is essentially a bottleneck layer, which can realize feature fusion and increase the non-linear relationship between features. After that, the merged image features are sent to the IRN for high-quality SR image reconstruction in Stage-II

$$I_{SR} = F_{IRN}(\mathrm{f_{fusion}}), \tag{4}$$

where $F_{IRN}(\cdot)$ denotes the IRN and I_{SR} is the final SR image.

Different from previous works which learn the mapping between LR and HR images directly, we propose an edge-assisted loss function. The edge-assisted loss function consists of a content loss and an edge loss. Therefore, given a training dataset $\left\{I^i_{LR}, I^i_{HR}, I^i_{Edge}\right\}_{i=1}^{N}$, we need to solve

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N}\sum_{i=1}^{N} \left\| F(I^i_{LR}) - I^i_{HR} \right\|_1 + \lambda \left\| E(I^i_{LR}) - I^i_{Edge} \right\|_1, \tag{5}$$

where λ is a hyperparameter, θ denotes the parameter set of our SeaNet, $F(\cdot)$ and $E(\cdot)$ denote the SeaNet and Edge-Net, respectively. It is worth noting that the Edge-Net is served as a subnet of the SeaNet and trained jointly with the whole model in an end-to-end manner.

A. Rough Image Reconstruction Network (RIRN)

In Stage-I, we use RIRN for rough SR image feature maps reconstruction. As we know, with the powerful representation capabilities of CNN, the image low-frequency features can be easily detected. Therefore, a shallow CNN model is sufficient for this task. Similar to previous works, we first use a 3×3 convolutional layer to transform the image from the RGB channel to a higher dimension. Then, five convolutional layers are applied to extract low-frequency image features from the LR image, each layer can be expressed as

$$\mathrm{f_m} = \max(0, W_m * \mathrm{f_{m-1}} + b_m), \tag{6}$$

where $\mathrm{f_{m-1}}$ is the output of the previous layer and is also considered as the input of the current layer. $\mathrm{f_m}$, W_m, and b_m are the output, weight, and bias of the current layer, respectively. Meanwhile, we set layer's ID $m = 1, 2, \cdots, 5$. Finally, the extracted low-frequency features are upscaled into HR dimension via an upsample module

$$\mathrm{f_{rough}} = F_{UP}(\mathrm{f_5}), \tag{7}$$

where $F_{UP}(\cdot)$ denotes the upsample module which consists of two convolutional layers and a sub-pixel layer [9]. The output $\mathrm{f_{rough}}$ represents the reconstructed rough SR image features. It contains three feature maps, essentially a representation of low-frequency features in the SR space.

B. Soft-Edge Reconstruction Network (Edge-Net)

Image edge prior is one of the most widely used and easily collected prior knowledge. As shown in Fig.5, directly upsampling the image soft-edge extracted from the LR image

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

into HR space or extracting the image soft-edge from the upsampled LR image will cause blurred edges and introduce additional noise. To settle this issue, we designed a soft-edge reconstruction network (Edge-Net) to reconstruct super-resolution image soft-edge from the LR image directly.

In Section II-C, we introduce the difference between the image soft-edge and image edge. The image soft-edge can retain more accurate image edge information, thus we use the following curvature formula to obtain the corresponding soft-edge I_{Edge} from I_{HR} since it can accurately describe the change in the gradient domain

$$I_{Edge} = \mathrm{div}(u_x, u_y), \tag{8}$$

where $u_i = \frac{\nabla_i I_{HR}}{\sqrt{1+|\nabla I_{HR}|^2}}, i \in \{x, y\}$, x and y represent horizontal and vertical directions, respectively. Meanwhile, ∇ and $\mathrm{div}(\cdot)$ denote the gradient and divergence operations, respectively. Although there are other soft-edge detection methods that can be used to construct the label of the training dataset, extensive experiments show that the image soft-edge detected by our proposed curvature formula is good enough in this work.

For Edge-Net, we use a modified version of multi-scale residual network (MSRN) [16] as its structure. MSRN is an efficient network that can detect the image features at different scales adaptively. This characteristic is beneficial for image soft-edge extraction. However, we find that the original MSRN is too complicated. In order to adapt MSRN to the soft-edge reconstruction task, we make the following modifications: (1). reducing M, the number of multi-scale residual blocks. (2). learning the mapping between I_{LR} and I_{Edge} instead of the mapping between I_{LR} and I_{HR}.

The Edge-Net can be used as part of any SR model to provide image soft-edge, or works independently to reconstruct super-resolution image soft-edge from the LR image directly. Its ultimate goal is to learn a reconstruction function that can reconstruct a SR soft-edge from the corresponding LR input. Thus, we define the edge loss as

$$\mathcal{L}_{edge} = \left\| E(I_{LR}) - I_{Edge} \right\|_1, \tag{9}$$

where $E(\cdot)$ denotes the Edge-Net, $E(I_{LR})$ represents the reconstructed soft-edge, and I_{Edge} is the soft-edge detected from the corresponding HR image. It is worth noting that, the reconstructed image soft-edge has the same dimension as the HR image since we also introduce an upsampling module in the Edge-Net.

C. Image Refinement Network (IRN)

In Stage-I, we use the RIRN and Edge-Net to reconstruct the rough SR image features and image soft-edge, respectively. The reconstructed rough SR image features contain plenty of low-frequency features and the reconstructed soft-edge contains rich high-frequency details. However, these two subnets are independent and the outputs of these modules cannot interact with each other. A common and simple method is to add the outputs of these two parts directly to get the final SR image. Although this method is feasible, we find that its result is sub-optimal. In order to take full advantage of image features and achieve soft-edge guidance, an image refinement network (IRN) is built for the final SR image reconstruction. In detail, we apply a fusion layer to fuse the low and high-frequency features, and then send the fused feature maps to the IRN to reconstruct SR images.

The IRN is a well designed image refinement network which contains two convolutional layers and N residual blocks. The residual blocks [18] are widely used in a variety of computer vision tasks, and the principle of the residual learning can be explained as follows:

Remark 1: Consider $H(x)$ as an underlying mapping to be fitted by some stacked convolutional layers, x denotes the input of this block. If multiple nonlinear layers can asymptotically approximate complicated functions, we can also hypothesize that they can asymptotically approximate the residual functions, i.e., $H(x)-x$. Thus, rather than expecting stacked layers to approximate $H(x)$, we let these layers to approximate a residual function $F(x) = H(x)-x$. Then the original function can be defined as $H(x) = F(x) + x$.

Our residual block contains two convolutional layers and one ReLU function, the ReLU function is only applied in the first layer. Different from the original residual block, we remove all batch normalization layers to reduce memory usage. Therefore, the output $\mathrm{f}_{\mathrm{rb}}^{\mathrm{n}}$ of each residual block can be defined as

$$\mathrm{f}_{\mathrm{rb}}^{\mathrm{n}} = (W_2 * \mathcal{R}(W_1 * \mathrm{f}_{\mathrm{rb}}^{\mathrm{n}-1} + b_1) + b_2) + \mathrm{f}_{\mathrm{rb}}^{\mathrm{n}-1}, \tag{10}$$

where W_1 and W_2 denote the weights of the first and the second convolutional layer, respectively. b_1 and b_2 are the corresponding biases, $\mathcal{R}$ denotes the ReLU function, and $\mathrm{f}_{\mathrm{rb}}^{\mathrm{n}-1}$ is the output of the previous blocks which is also considered as the input of current block.

In addition, except for using the residual block for local residual learning, we also apply a long skip connection for global residual learning. The long skip connection can solve the problem of gradient disappearance while increasing the flow of information and improving the model performance. The global residual learning can be defined as

$$\mathrm{f}_{\mathrm{out}} = \mathrm{f}_{\mathrm{rb}}^{\mathrm{N}} + \mathrm{f}_{\mathrm{input}}, \tag{11}$$

where $\mathrm{f}_{\mathrm{input}}$ and $\mathrm{f}_{\mathrm{rb}}^{\mathrm{N}}$ represent the input of the first residual block and the output of the last residual block, respectively.

During training, we use the L_1 loss as the content loss to minimize the difference between the SR and HR images

$$\mathcal{L}_{content} = \left\| I_{SR} - I_{HR} \right\|_1. \tag{12}$$

In summary, under the assistance of the image soft-edge, SeaNet can efficiently reconstruct high-quality SR images with sharp edges and rich texture details. Then RIRN, Edge-Net, and IRN form the complete Soft-edge assisted Network (SeaNet). The content loss $\mathcal{L}_{SR}$ and edge loss $\mathcal{L}_{Edge}$ form the complete edge-assisted loss $\mathcal{L}_{total}$,

$$\mathcal{L}_{total} = \mathcal{L}_{content} + \lambda \mathcal{L}_{edge}, \tag{13}$$

where λ is a hyper-parameter used to control the composition of the edge loss. The special emphasis here is that although SeaNet consists of three subnets, it can also achieve end-to-end training.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

TABLE I

QUANTITATIVE COMPARISONS OF THE STATE-OF-THE-ART SR METHODS. ALL OF THESE METHODS ARE TRADITIONAL MATHEMATICAL MODELS OR MODELS THAT INTRODUCE IMAGE PRIORS INTO CNN FOR SR IMAGE RECONSTRUCTION. NOTICE THAT, DEGREE-MV USES THE MULTI-VIEW TESTING STRATEGY TO IMPROVE PERFORMANCE. BEST RESULTS ARE **HIGHLIGHTED**

Algorithm	Scale	Set5 [37] PSNR / SSIM	Set14 [38] PSNR / SSIM	BSDS100 [39] PSNR / SSIM	Image Priors
Bicubic	×2	33.69 / 0.9284	30.34 / 0.8675	29.57 / 0.8434	-
A+ [32]	×2	36.60 / 0.9542	32.42 / 0.9059	31.24 / 0.8870	Neighbor embedding
SelfExSR [40]	×2	36.60 / 0.9537	32.46 / 0.9051	31.20 / 0.8863	Transformed self-exemplars
CSCN-MV [27]	×2	37.14 / 0.9567	32.56 / 0.9074	31.40 / 0.8840	Sparse prior
DEGREE-MV [29]	×2	37.61 / 0.9589	33.11 / 0.9129	31.84 / 0.8951	Edge prior
SeaNet (Ours)	×2	**38.08 / 0.9609**	**33.75 / 0.9190**	**32.27 / 0.9008**	Edge prior
Bicubic	×3	30.41 / 0.8655	27.64 / 0.7722	27.21 / 0.7344	-
A+ [32]	×3	32.63 / 0.9085	29.25 / 0.8194	28.31 / 0.7828	Neighbor embedding
SelfExSR [40]	×3	32.66 / 0.9089	29.34 / 0.8222	28.30 / 0.7839	Transformed self-exemplars
CSCN-MV [27]	×3	33.26 / 0.9167	29.55 / 0.8271	28.50 / 0.7885	Sparse prior
DEGREE-MV [29]	×3	33.70 / 0.9212	29.77 / 0.8309	28.76 / 0.7956	Edge prior
SeaNet (Ours)	×3	**34.55 / 0.9282**	**30.42 / 0.8445**	**29.17 / 0.8071**	Edge prior
Bicubic	×4	28.43 / 0.8022	26.10 / 0.6936	25.97 / 0.6517	-
A+ [32]	×4	30.33 / 0.8565	27.44 / 0.7450	26.83 / 0.6999	Neighbor embedding
SelfExSR [40]	×4	30.34 / 0.8593	27.55 / 0.7511	26.84 / 0.7032	Transformed self-exemplars
CSCN-MV [27]	×4	31.04 / 0.8775	27.76 / 0.7620	27.11 / 0.7191	Sparse prior
DEGREE-MV [29]	×4	31.30 / 0.8968	27.92 / 0.7637	27.18 / 0.7207	Edge prior
SeaNet (Ours)	×4	**32.33 / 0.8970**	**28.72 / 0.7855**	**27.65 / 0.7388**	Edge prior

Overall speaking, we have designed a model called SeaNet according to the fact that an image consists of low and high-frequency features. Specifically, SeaNet is a well-designed model that contains three subnets on the basis of specific learning purposes. In addition, these subnets are well-defined and work synergistically, which greatly improves the interpretability of the proposed network.

IV. EXPERIMENTS

A. Datasets

The DIV2K [42] is a new high-quality dataset for image restoration, which contains 800 training images, 100 validation images, and 100 test images. Following previous works [16], [17], we train all of our models and compared models on DIV2K (1-800) dataset. For testing, we choose Set5 [37], Set14 [38], BSDS100 [39], Urban100 [40], and Manga109 [41] as our test datasets. All of them are the most widely used test benchmark datasets, which contain a variety of scenarios that can fully validate the model performance.

B. Implementation Details

1) Model Setting: In this work, we design three different versions of SeaNet. The baseline SeaNet contains 3 multi-scale residual blocks (MSRB) in the Edge-Net and 20 residual blocks (RB) in the IRN, denoted as **SeaNet (Baseline)**. The final SeaNet contains 5 MSRBs in the Edge-Net and 40 RBs in the IRN, denoted as **SeaNet**. We also introduce the self-ensemble strategy to further improve our **SeaNet** and denote the self-ensembled version as **SeaNet+**. The kernel size of all convolutional layers is set to 3×3 except for the feature fusion layer, whose kernel size is 1×1. Meanwhile, the input and output channel of each MSRB and RB are set to 64.

2) Training Setting: In this work, we first apply the Equation (8) on 800 training images to obtain their corresponding image soft-edge. Then, we generate LR images by applying Bicubic interpolation to HR images. Finally, we use the RGB image as input and augment the training data with random horizontal flips and vertical. Following previous works [16], [17], we randomly extract 16 LR patches with the size of 48×48 as input for **SeaNet** (32×32 for **SeaNet (Baseline)**) and 1,000 iterations of back-propagation constitute an epoch. We implement our SeaNet with the PyTorch framework and update it with Adam optimizer. The learning rate is initialized as 10^{-4} and halved every 200 epochs. Meanwhile, we set $\lambda = 0.1$ based on a lot of experience. Besides, to further verify the performance of the Edge-Net, we also train the independent Edge-Net for SR soft-edge reconstruction and all of them use 3 MSRBs in this work. All the models are implemented with the PyTorch framework and trained on NVIDIA Titan Xp GPU.

C. Comparisons With the State-of-the-Art Methods

As shown in TABLEs I-III, we compare our SeaNet with more than 14 SR methods to fully verify the model effectiveness, including Bicubic, A+ [32], SelfExSR [40], CSCN-MV [27], DEGREE-MV [29], SRCNN [7], ESPCN [9], FSRCNN [8], VDSR [10], DRCN [11], LapSRN [13], DRRN [12], MSRN [16], and EDSR [17]. These methods include traditional mathematical models and CNN-based models. The CNN-based methods are further divided into methods with and without image priors. Among them, EDSR [17] wins the first place in NTIRE 2017 Super-Resolution Challenge [43] and achieves the state-of-the-art results. All of the SR results are evaluated with PSNR and SSIM on the Y channel of the transformed YCbCr space.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

TABLE II

QUANTITATIVE COMPARISONS OF THE STATE-OF-THE-ART SR METHODS. ALL OF THESE METHODS ARE BASED ON CNN WITHOUT IMAGE PRIORS. BEST RESULTS ARE **HIGHLIGHTED** AND SECOND BEST RESULTS ARE UNDERLINED

Algorithm	Scale	Set5 [37] PSNR / SSIM	Set14 [38] PSNR / SSIM	BSDS100 [39] PSNR / SSIM	Urban100 [40] PSNR / SSIM	Manga109 [41] PSNR / SSIM
SRCNN [7]	×2	36.71 / 0.9536	32.32 / 0.9052	31.36 / 0.8880	29.54 / 0.8962	35.74 / 0.9661
ESPCN [9]	×2	37.00 / 0.9559	32.75 / 0.9098	31.51 / 0.8939	29.87 / 0.9065	36.21 / 0.9694
FSRCNN [8]	×2	37.06 / 0.9554	32.76 / 0.9078	31.53 / 0.8912	29.88 / 0.9024	36.67 / 0.9694
VDSR [10]	×2	37.53 / 0.9583	33.05 / 0.9107	31.92 / 0.8965	30.79 / 0.9157	37.22 / 0.9729
DRCN [11]	×2	37.63 / 0.9584	33.06 / 0.9108	31.85 / 0.8947	30.76 / 0.9147	37.63 / 0.9723
LapSRN [13]	×2	37.52 / 0.9581	33.08 / 0.9109	31.80 / 0.8949	30.41 / 0.9112	37.27 / 0.9855
DRRN [12]	×2	37.74 / 0.9590	33.23 / 0.9140	32.05 / 0.8970	31.23 / 0.9190	37.92 / 0.9760
SeaNet (Ours)	×2	38.08 / 0.9609	33.75 / 0.9190	32.27 / 0.9008	32.50 / 0.9318	38.76 / 0.9774
SeaNet+ (Ours)	×2	**38.15 / 0.9611**	**33.86 / 0.9198**	**32.31 / 0.9013**	**32.68 / 0.9332**	**38.97 / 0.9779**
SRCNN [7]	×3	32.47 / 0.9067	29.23 / 0.8201	28.31 / 0.7832	26.25 / 0.8028	30.59 / 0.9107
ESPCN [9]	×3	33.02 / 0.9135	29.49 / 0.8271	28.50 / 0.7937	26.41 / 0.8161	30.79 / 0.9181
FSRCNN [8]	×3	33.20 / 0.9149	29.54 / 0.8277	28.55 / 0.7945	26.48 / 0.8175	30.98 / 0.9212
VDSR [10]	×3	33.68 / 0.9201	29.86 / 0.8312	28.83 / 0.7966	27.15 / 0.8315	32.01 / 0.9310
DRCN [11]	×3	33.85 / 0.9215	29.89 / 0.8317	28.81 / 0.7954	27.16 / 0.8311	32.31 / 0.9328
LapSRN [13]	×3	33.82 / 0.9207	29.89 / 0.8304	28.82 / 0.7950	27.07 / 0.8298	32.21 / 0.9318
DRRN [12]	×3	34.03 / 0.9240	29.96 / 0.8350	28.95 / 0.8000	27.53 / 0.7640	32.74 / 0.9390
SeaNet (Ours)	×3	34.55 / 0.9282	30.42 / 0.8444	29.17 / 0.8071	28.50 / 0.8594	33.73 / 0.9463
SeaNet+ (Ours)	×3	**34.65 / 0.9290**	**30.53 / 0.8461**	**29.23 / 0.8081**	**28.68 / 0.8620**	**34.02 / 0.9478**
SRCNN [7]	×4	30.50 / 0.8573	27.62 / 0.7453	26.91 / 0.6994	24.53 / 0.7236	27.66 / 0.8505
ESPCN [9]	×4	30.66 / 0.8646	27.71 / 0.7562	26.98 / 0.7124	24.60 / 0.7360	27.70 / 0.8560
FSRCNN [8]	×4	30.73 / 0.8601	27.71 / 0.7488	26.98 / 0.7029	24.62 / 0.7272	27.90 / 0.8517
VDSR [10]	×4	31.36 / 0.8796	28.11 / 0.7624	27.29 / 0.7167	25.18 / 0.7543	28.83 / 0.8809
DRCN [11]	×4	31.56 / 0.8810	28.15 / 0.7627	27.24 / 0.7150	25.15 / 0.7530	28.98 / 0.8816
LapSRN [13]	×4	31.54 / 0.8811	28.19 / 0.7635	27.32 / 0.7162	25.21 / 0.7564	29.09 / 0.8845
DRRN [12]	×4	31.68 / 0.8888	28.21 / 0.7722	27.38 / 0.7240	25.44 / 0.7640	29.46 / 0.8960
SeaNet (Ours)	×4	32.33 / 0.8970	28.72 / 0.7855	27.65 / 0.7388	26.32 / 0.7942	30.74 / 0.9129
SeaNet+ (Ours)	×4	**32.44 / 0.8981**	**28.81 / 0.7872**	**27.70 / 0.7399**	**26.50 / 0.7976**	**31.05 / 0.9154**

1) Objective Evaluation: TABLE I shows the quantitative comparison between the SeaNet and other SR methods. All the report methods are traditional mathematical-based models or the models introducing image priors into CNN. Obviously, our SeaNet achieves the best results among these methods. It is worth noting that, the DEGREE-MV [29] also introduces image edge prior into CNN model and uses the multi-view testing strategy to improve the model performance while the results of DEGREE are still unsatisfactory. The main reasons may lie in the following: (i) DEGREE uses Bicubic interpolated LR image as input; (ii) DEGREE applies an off-the-shelf edge detector for edge detection; (iii) the network structure of DEGREE is sub-optimal. In order to solve these shortcomings, we propose the SeaNet. Extensive experiments demonstrate that the performance of our SeaNet is better than the DEGREE.

In TABLE II, we show the performance comparison between the SeaNet and some classical CNN-based SR models. All of these CNN-based models use a well-designed network to learn the mapping function between LR and HR images directly and all of them achieve the best results at the time. However, they ignore the importance of image priors for image reconstruction. Therefore, the reconstructed SR images often lack texture details. Different from these models, we design an Edge-Net for image soft-edge reconstruction and embed the Edge-Net as part of the SeaNet to provide image soft-edge prior. With the assistance of the image soft-edge prior, the SeaNet achieves the best results in all test datasets.

Recently, a series of large SR models have been proposed, including but not limited to MSRN [16] and EDSR [17]. Among them, the EDSR is one of the most famous large-scale SR model and MSRN is the prototype network of our Edge-Net. In TABLE III, we show the comparison of MSRN, EDSR, EDSR+, SeaNet (baseline), SeaNet and SeaNet+. Obviously, the performance of SeaNet(+) and EDSR(+) are close. It is worth noting that the parameter quantity of EDSR are 40121k, 42481k and 45430k in x2, x3, and x4, respectively. While the parameter quantities of SeaNet are only 7102k, 7471k and 7397k in x2, x3, and x4 respectively, which is about one-sixth (**1/6**) of that of EDSR. Besides, compared with MSRN, SeaNet has achieved better results on different upsampling factors. Considering the parameter quantities of our SeaNet are more than that of MSRN, we show the comparison between SeaNet (Baseline) and MSRN to further verify the performance. SeaNet (baseline) only has 3 MSRBs and 20 RBs, its parameter quantity is about two-thirds (2/3)

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

TABLE III
QUANTITATIVE COMPARISONS OF MSRN, EDSR, EDSR+, SEANET (BASELINE), SEANET (FINAL), AND SEANET+ (FINAL)

Algorithm	Scale	Parameters	Set5 [37] PSNR/SSIM	Set14 [38] PSNR/SSIM	BSDS100 [39] PSNR/SSIM	Urban100 [40] PSNR/SSIM	Manga109 [41] PSNR/SSIM
MSRN [16]	×2	5926k	38.08/0.9605	33.74/0.9170	32.23/0.9013	32.22/0.9326	38.82/0.9868
SeaNet (Baseline, Ours)	×2	4194k	37.99/0.9607	33.60/0.9174	32.18/0.8995	32.08/0.9276	38.48/0.9768
EDSR [17]	×2	**40121k**	38.11/0.9602	33.92/0.9195	32.32/0.9013	32.93/0.9351	39.10/0.9773
EDSR+ [17]	×2	**40121k**	38.20/0.9606	34.02/0.9204	32.37/0.9018	33.10/0.9363	– / –
SeaNet (Final, Ours)	×2	7102k	38.08/0.9609	33.75/0.9190	32.27/0.9008	32.50/0.9318	38.76/0.9774
SeaNet+ (Final, Ours)	×2	7102k	38.15/0.9611	33.86/0.9198	32.31/0.9013	32.68/0.9332	38.97/0.9779
MSRN [16]	×3	6110k	34.38/0.9262	30.34/0.8395	29.08/0.8041	28.08/0.8554	33.44/0.9427
SeaNet (Baseline, Ours)	×3	4563k	34.36/0.9280	30.34/0.8428	29.09/0.8053	28.17/0.8527	33.40/0.9444
EDSR [17]	×3	**42481k**	34.65/0.9280	30.52/0.8462	29.25/0.8093	28.80/0.8653	34.17/0.9476
EDSR+ [17]	×3	**42481k**	34.76/0.9290	30.66/0.8481	29.32/0.8104	29.02/0.8685	– / –
SeaNet (Final, Ours)	×3	7471k	34.55/0.9282	30.42/0.8444	29.17/0.8071	28.50/0.8594	33.73/0.9463
SeaNet+ (Final, Ours)	×3	7471k	34.65/0.9290	30.53/0.8461	29.23/0.8081	28.68/0.8620	34.02/0.9478
MSRN [16]	×4	6073k	32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896	30.17/0.9034
SeaNet (Baseline, Ours)	×4	4224k	32.18/0.8948	28.61/0.7822	27.57/0.7359	26.05/0.7896	30.44/0.9088
EDSR [17]	×4	**45430k**	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	31.02/0.9148
EDSR+ [17]	×4	**45430k**	32.62/0.8984	28.94/0.7901	27.79/0.7437	26.67/0.8041	– / –
SeaNet (Final, Ours)	×4	7397k	32.33/0.8970	28.72/0.7855	27.65/0.7388	26.32/0.7942	30.74/0.9129
SeaNet+ (Final, Ours)	×4	7397k	32.44/0.8981	28.81/0.7872	27.70/0.7399	26.50/0.7976	31.05/0.9154

of MSRN. Obviously, SeaNet (Baseline) achieves competitive results with fewer parameters. In addition, compared with MSRN, SeaNet achieves better results on large scale (x3, x4). This further demonstrates the importance of image edge prior on large scale SR problem.

2) Subjective Evaluation: In Fig. 4, we show the visual comparisons on ×2, ×3, and ×4, respectively. We can clearly see that most SR methods cannot recover clean and right image edges. However, with the image soft-edge assistance, our SeaNet can reconstruct high-quality SR images with sharper and more accurate image edges. All the reconstructed SR images can be downloaded from http://t.cn/EUnPgu6.

V. ANALYSIS AND DISCUSSION

A. Effectiveness of the Edge-Net

According to our knowledge, the Edge-Net is the first CNN model used to directly reconstruct the image soft-edge from the LR image. The Edge-Net can be used independently or embedded as a part of SR models to provide image soft-edge prior. In this part, we provide a series of ablation analysis to illustrate the effectiveness of the Edge-Net.

(1). We train an Edge-Net alone to observe its ability to reconstruct image soft-edge from the LR image. In Fig. 5, we show the visual comparisons of three different soft-edge reconstruction methods, including 'Bicubic+Eq.(8)', 'Eq.(8)+Bicubic', and our Edge-Net. The 'Bicubic+Eq.(8)' represents that we adopt Eq.(8) to extract soft-edge from the preprocessed LR which is upsampled to HR space using Bicubic. 'Eq. (8)+Bicubic' means that we adopt Bicubic to upsample the soft-edge which is extracted from the original LR using Eq.(8). The 'Edge-Net' represents using our proposed Edge-Net to directly reconstruct the SR soft-edge from LR image and the 'Ground Truth' denotes using Eq. (8) to directly detect the soft-edge from the HR image. One can see that the edge extracted by 'Bicubic+Eq.(8)' and 'Eq.(8)+Bicubic' have been severely damaged and contain a lot of noise while our Edge-Net can reconstruct clean and accurate soft-edge with more texture details.

(2). In this work, we embed the Edge-Net as part of the SeaNet to provide image soft-edge for SR image reconstruction. To further verify the effectiveness of Edge-Net, we build a new model named SRN, which removes the Edge-Net from the SeaNet. Considering the difference in the amount of parameters between these two models, we use the SeaNet (Baseline) as the benchmark model because it contains only a few parameters in Edge-Net. TABLE IV and Fig.6 present the performance comparison of SRN and SeaNet (Baseline). According to the experimental results, we can find that SRN is inferior to the SeaNet (Baseline) both qualitatively and quantitatively. Especially, as shown in Fig.6, with the help of the Edge-Net, our SeaNet (Baseline) can build high-quality SR images with sharp image edges.

(3). We proposed an Edge-Net to learn the mapping between LR images and the soft-edge of HR images, thus our Edge-Net can directly reconstruct the SR soft-edge from the LR image. In TABLE V, we show the PSNR comparison of model effect under the guidance of different soft-edge provide by 'Bicubic+Eq.(8)', 'Eq.(8)+Bicubic', and the Edge-Net. Obviously, when the Edge-Net is embedded as a part of the SeaNet to provide the image soft-edge prior, the model achieves the best results.

Therefore, the aforementioned experiments fully demonstrate the effectiveness of the Edge-Net. Meanwhile, experiments show that high-quality image soft-edge is useful for image reconstruction.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

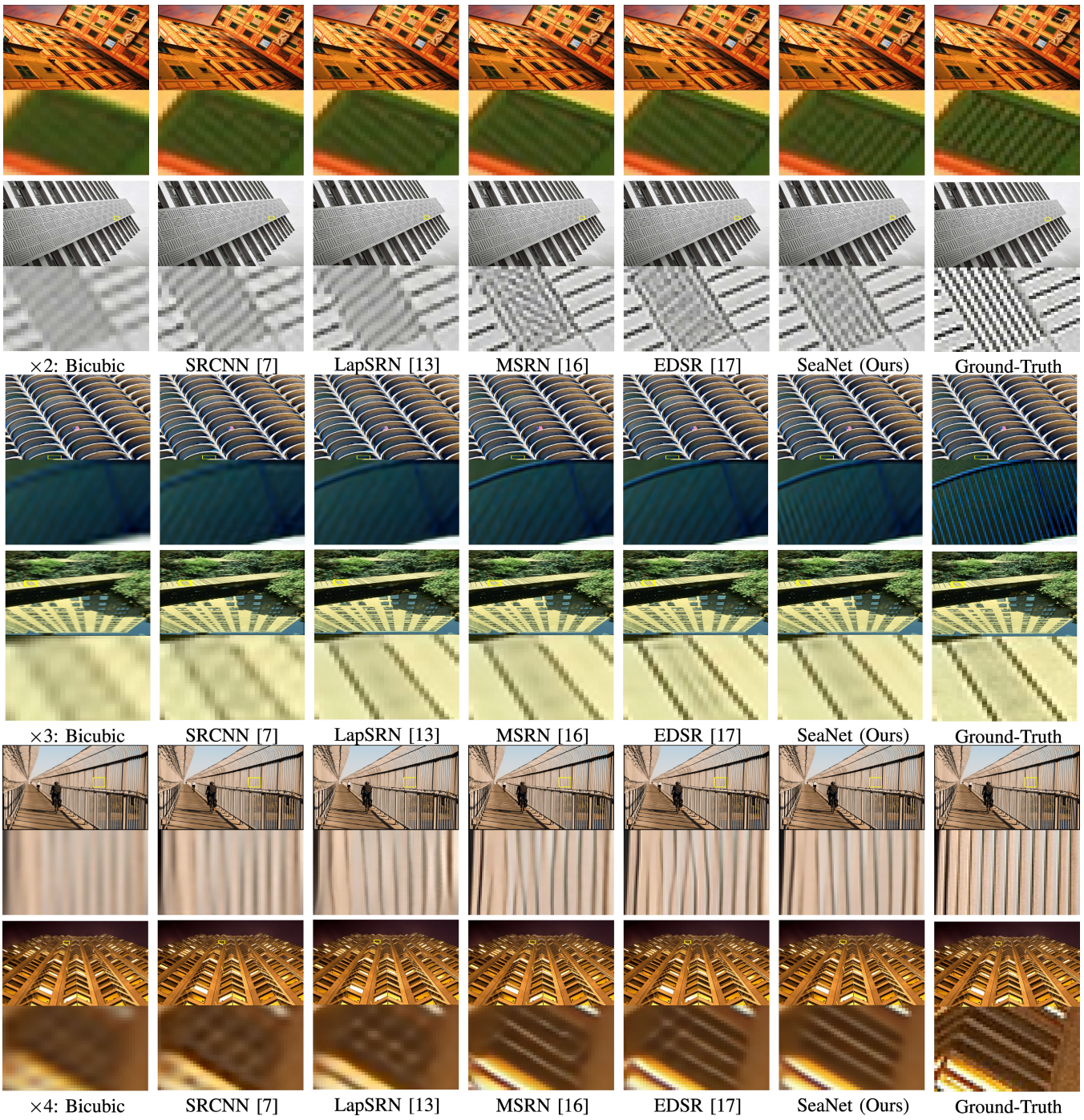

Fig. 4. Visual comparison for x2, x3, and x4 SR images. Our SeaNet can reconstruct more realistic SR images with sharper edges.

B. Effectiveness of the Image Soft-Edge Prior

Most CNN-based SR models tend to use deep and wide models to directly learn the mapping between LR and HR images. However, blindly increasing the depth of the network is not the best choice because each model has its performance bottleneck. Furthermore, increasing the depth of the network will increase model parameters, execution time, and memory. Therefore, we aim to explore a SR framework guided by image priors. To achieve this, we suggest using the image soft-edge to assist SR images reconstruction.

To evaluate the effectiveness of the proposed image soft-edge assistance, a classical SR model, ESPCN [9], is adopted as our benchmark model. The ESPCN is a shallow model that contains only 5 convolutional layers. We here build a new model termed 'ESPCN+ISE' that using ESPCN as the backbone and introducing the image soft-edge as additional prior knowledge. For a fair comparison, we retrain the ESPCN model because we need to make slight adjustments to the network structure when adding the image soft-edge. The detailed experimental design is as follows: (a) We pre-trained a simplified version of Edge-Net, which has 3 MSRBs and

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

Fig. 5. Visual comparison of different soft-edge extraction methods. The 'Bicubic+Eq. (8)' represents that we adopt Eq. (8) to extract soft-edge from the preprocessed LR which is upsampled to HR space using Bicubic. 'Eq. (8)+Bicubic' means that we adopt Bicubic to upsample the soft-edge which is extracted from the original LR using Eq. (8). The 'Edge-Net' represents using our proposed Edge-Net to directly reconstruct the SR soft-edge from LR image and the 'Ground Truth' denotes using Eq. (8) to directly detect the soft-edge from the HR image. Obviously, our SeaNet can reconstruct clear and accurate soft-edge.

the input and output channels of each MSRB are set to 16. (b) In 'ESPCN+ISE' we introduce a fusion layer at the end of the model to fuse the feature maps reconstructed by the ESPCN and the image soft-edge provided by the Edge-Net. (c) During training, the parameters of the Edge-Net are fixed. This means that the Edge-Net is only used to provide the soft-edge without participating in model training. Therefore, compared to ESPCN (20k parameters), the 'ESPCN+ISE' only increase a fusion layer (18 parameters). The PSNR results of ESPCN and 'ESPCN+ISE' are presented in the TABLE VI. Obviously, with the image soft-edge assistance, the performance of 'ESPCN+ISE' has been greatly improved compared to ESPCN.

Therefore, we can draw a conclusion that the soft-edge prior provided by the Edge-Net is effective and the introduction of additional information can greatly improve the performance of the existing SR models. The soft-edge prior can be flexibly introduced into any SR model to assist high-quality SR image reconstruction. All these experiments fully demonstrate the validity of the soft-edge prior.

C. Study of λ

The λ is a hyper-parameter used to control the composition of the edge loss. According to our experience, SeaNet is robust to the choice of λ and the best performance is provided by setting $0 < \lambda < 1$. In Fig. 7, we show the impact of different

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

QUANTITATIVE COMPARISONS OF SRN AND SEANET (BASELINE)

Dataset	Scale	SRN	SeaNet (Baseline)
	×2	37.78/0.9597	**37.99/0.9607**
Set5 [37]	×3	34.11/0.9249	**34.36/0.9280**
PSNR/SSIM	×4	32.01/0.8919	**32.18/0.8948**
	×2	33.42/0.9158	**33.60/0.9174**
Set14 [38]	×3	30.12/0.8378	**30.34/0.8428**
PSNR/SSIM	×4	28.42/0.7771	**28.61/0.7822**
	×2	32.04/0.8974	**32.18/0.8995**
BSDS100 [39]	×3	28.95/0.8006	**29.09/0.8053**
PSNR/SSIM	×4	27.43/0.7304	**27.57/0.7359**
	×2	31.56/0.9223	**32.08/0.9276**
Urban100 [40]	×3	27.74/0.8415	**28.17/0.8527**
PSNR/SSIM	×4	25.74/0.7718	**26.05/0.7896**
	×2	37.98/0.9756	**38.48/0.9768**
Manga109 [41]	×3	32.98/0.9405	**33.40/0.9444**
PSNR/SSIM	×4	30.00/0.9022	**30.44/0.9088**

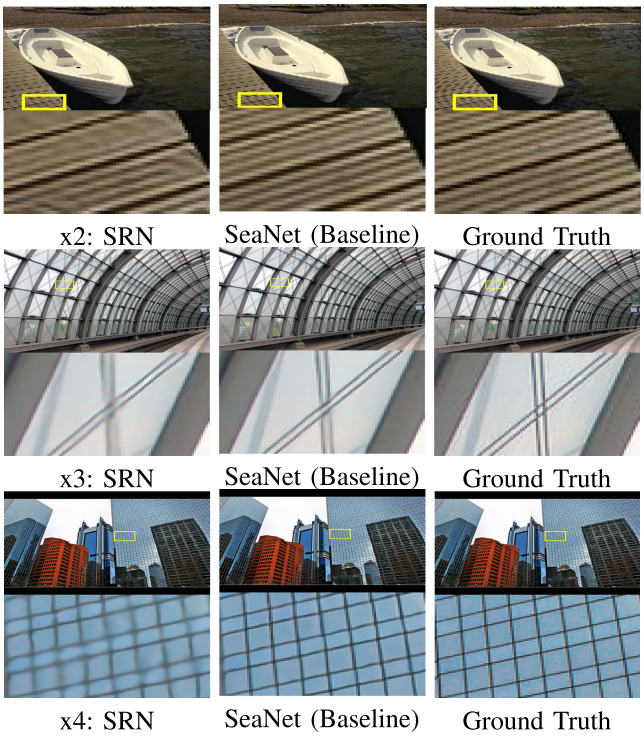

Fig. 6. Visual comparison of SRN and SeaNet (baseline) for x2, x3, and x4 SR images. SRN is a simplified model obtained by removing the Edge-Net from the SeaNet (baseline).

λ on the model performance. The result shows that the model achieves best results when $\lambda = 0.1$. Therefore, we set $\lambda = 0.1$ in our final model.

D. Study of Model Size

In Fig. 8, we show the comparison of model performance and parameters between SeaNet and other SR models. The blue dots represent classical SR models with fewer parameters, the green dot represents EDSR [17], and the red dots represent our 3 different versions of SeaNet: SeaNet (Baseline) (M=3, N=20), SeaNet (M=5, N=40), and SeaNet+ (M=5, N=40). It is clear seen that the parameter quantity of SeaNet is only 1/7 of EDSR, but their performances are rather close. Meanwhile, SeaNet (Baseline) achieves the same performance as MSRN with fewer parameters. All of these studies show that SeaNet is an efficient and accurate SR model that strikes a good balance between model complexity and performance.

TABLE V

PSNR COMPARISON OF MODEL EFFECT UNDER THE GUIDANCE OF DIFFERENT IMAGE SOFT-EDGE PROVIDED BY 'BICUBIC+EQ.(8)', 'EQ.(8)+BICUBIC', AND THE EDGE-NET, RESPECTIVELY

Method	Set14			Urban100		
	× 2	× 3	× 4	× 2	× 3	× 4
'Bicubic + Eq.(8)'	33.33	30.15	28.45	31.60	27.77	25.76
'Eq.(8) + Bicubic'	33.35	30.18	28.49	31.70	27.84	25.82
Edge-Net	**33.60**	**30.34**	**28.61**	**32.08**	**28.17**	**26.05**

TABLE VI

QUANTITATIVE COMPARISONS OF ESPCN [9] AND 'ESPCN+ISE'

Dataset	Scale	ESPCN	ESPCN+ISE
	×2	37.00/0.9559	**37.50/0.9580**
Set5 [37]	×3	33.02/0.9135	**33.63/0.9190**
PSNR/SSIM	×4	30.66/0.8646	**31.32/0.8780**
	×2	32.75/0.9098	**33.01/0.9100**
Set14 [38]	×3	29.49/0.8271	**29.78/0.8291**
PSNR/SSIM	×4	27.71/0.7562	**28.07/0.7600**
	×2	31.51/0.8939	**31.88/0.8956**
BSDS100 [39]	×3	28.50/0.7937	**28.77/0.7960**
PSNR/SSIM	×4	26.98/0.7124	**27.20/0.7156**
	×2	29.87/0.9065	**30.66/0.9145**
Urban100 [40]	×3	26.41/0.8161	**27.09/0.8299**
PSNR/SSIM	×4	24.60/0.7360	**25.08/0.7430**
	×2	36.21/0.9694	**37.18/0.9711**
Manga109 [41]	×3	30.79/0.9181	**31.89/0.9288**
PSNR/SSIM	×4	27.70/0.8560	**28.78/0.8756**

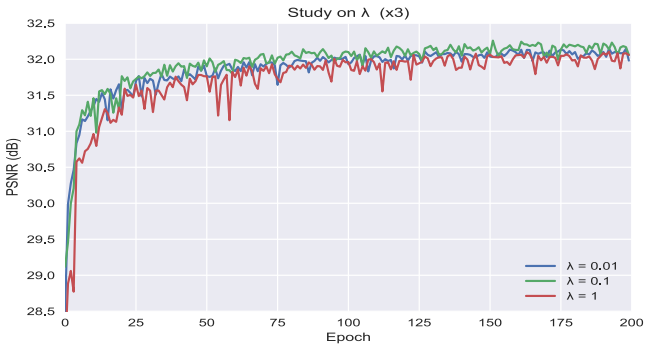

Fig. 7. Study on λ. When λ = 0.1, the model achieves best results.

E. Limitations and Future Works

The proposed SeaNet performs well in SISR. However, it also has some limitations: (1) During training, we use Equation (8) to extract soft-edge from HR images as labels.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

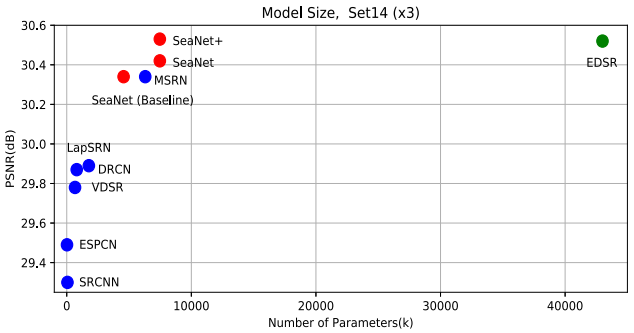

Fig. 8. Study of model size on the test dataset Set14 (x3). SeaNet strikes a good balance between model size and performance.

Equation (8) is a curvature formula, which is an efficient method but maybe not the best. (2) At the beginning of State-II, we use a fusion layer to fuse the reconstructed image features and soft-edge provided by Stage-I. Extensive experiments have shown that it is an effective method. However, we believe that there exists some better way to integrate image features which can make the soft-edge play a bigger effect. We will discuss those in future works. Moreover, our Edge-Net can be used in many applications, such as image denoising and deblurring. We also leave it as our future work.

VI. Conclusion

In this paper, we proposed a new edge-guided image restoration framework and developed an efficient and accurate Soft-edge assisted Network (SeaNet) for image super-resolution. The SeaNet consists of three sub-nets: a rough image reconstruction network (RIRN), a soft-edge reconstruction network (Edge-Net), and an image refinement network (IRN). In State-I, we use the RIRN and Edge-Net for the rough SR image features and image soft-edge reconstruction, respectively. In State-II, we first fuse the reconstructed image feature maps and soft-edge by a bottleneck layer and then send the fused feature maps to the IRN for the final SR image reconstruction. Among them, the Edge-Net is the first CNN model that can directly reconstruct super-resolution soft-edge from the LR image. Furthermore, the Edge-Net is a flexible model that can work independently for image soft-edge reconstruction or embedded as a part of any SR model to provide soft-edge prior. Extensive benchmark evaluations well demonstrated that with the assistance of image soft-edge prior, our SeaNet achieves competitive results with fewer parameters.

References

[1] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Phys. Med. Biol.*, vol. 63, no. 13, Jun. 2018, Art. no. 135007.

[2] D. Lee, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 15–18.

[3] M. Ben-Ezra, A. Zomet, and S. K. Nayar, "Video super-resolution using controlled subpixel detector shifts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 977–987, Jun. 2005.

[4] 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, Feb. 2010.

[5] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. 4th IEEE Int. Conf. Automat. Face Gesture Recognit.*, Mar. 2000, pp. 83–88.

[6] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

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

[8] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 391–407.

[9] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

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

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

[12] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.

[13] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[14] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[15] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[16] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 517–532.

[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] H. Shen, "Zoom-based super-resolution reconstruction approach using prior total variation," *Opt. Eng.*, vol. 46, no. 12, Dec. 2007, Art. no. 127003.

[20] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 1–16, Jun. 2007.

[21] 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.

[22] 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.

[23] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[24] Q. Zhou, S. Chen, J. Liu, and X. Tang, "Edge-preserving single image super-resolution," in *Proc. 19th ACM Int. Conf. Multimedia (MM)*, 2011, pp. 1037–1040.

[25] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 428–438, Jan. 2016.

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

[27] 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. (ICCV)*, Dec. 2015, pp. 370–378.

[28] N. Basty, D. McClymont, I. Teh, J. E. Schneider, and V. Grau, "Reconstruction of 3D cardiac MR images from 2D slices using directional total variation," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. Cham, Switzerland: Springer, 2017, pp. 127–135.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.

[29] W. Yang *et al.*, "Deep edge guided recurrent residual learning for image super-resolution," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5895–5907, Sep. 2017.
[30] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, Jun. 2014.
[31] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 372–386.
[32] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 111–126.
[33] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[34] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, pp. 275–282.
[35] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.
[36] J. Salvador and E. Perez-Pellitero, "Naive Bayes super-resolution forest," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 325–333.
[37] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.
[38] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, 2010, pp. 711–730.
[39] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[40] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.
[41] Y. Matsui *et al.*, "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, Nov. 2016.
[42] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jul. 2017, pp. 1110–1121.
[43] R. Timofte *et al.*, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1110–1121.

Faming Fang received the Ph.D. degree in computer science from East China Normal University, Shanghai, China, in 2013. He is currently a Professor with the School of Computer Science and Technology, East China Normal University. His main research area is image processing using variational methods, PDEs, and deep learning.

Juncheng Li received the B.Sc. degree in computer science from Jiangxi Normal University, Nanchang, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, East China Normal University, Shanghai, China.

His main research interests are image processing, machine learning, and deep learning.

Tieyong Zeng received the B.S. degree from Peking University, Beijing, China, in 2000, the M.S. degree from École Polytechnique, Palaiseau, France, in 2004, and the Ph.D. degree from the Université of Paris XIII, Paris, France, in 2007. He is currently an Associate Professor with the Department of Mathematics, The Chinese University of Hong Kong. He worked as a Postdoctoral Researcher with ENS de Cachan from 2007 to 2008 and an Assistant/Associate Professor with Hong Kong Baptist University from 2008 to 2018. His research interests are image processing, machine learning, and scientific computing.

Authorized licensed use limited to: East China Normal University. Downloaded on February 28,2020 at 03:22:09 UTC from IEEE Xplore. Restrictions apply.