Deep Objective Quality Assessment Driven Single Image Super-Resolution

Bo Yan, *Senior Member, IEEE*, Bahetiyaer Bare, Chenxi Ma, Ke Li, and Weimin Tan

***Abstract*—Single-image super-resolution (SISR) is a classic problem in the image processing community, which aims at generating a high-resolution image from a low-resolution one. In recent years, deep learning based SISR methods emerged and achieved a performance leap than previous methods. However, because the evaluation metrics of SISR methods is peak signal-to-noise ratio (PSNR), previous methods usually choose L2-norm as the loss function. This leads to a significant improvement in the final PSNR value but little improvement in perceptual quality. In this paper, in order to achieve better results in both perceptual quality and PSNR values, we propose an objective quality assessment driven SISR method. First, we propose a novel full-reference image quality assessment approach for SISR and employ it as a loss function, namely super-resolution image quality assessment (SR-IQA) loss. Then, we combine SR-IQA loss with L2-norm to guide our proposed SISR method to achieve better results. Besides that, our proposed SISR method consists of several proposed highway units. Furthermore, in order to verify the generalization ability of our new kind of loss function, we integrate SR-IQA loss to generative adversarial networks based SR method and achieve better perceptual quality. Experimental results prove that our proposed SISR method achieves better performance than other methods both qualitatively and quantitatively in most of the cases.**

***Index Terms*—Single image super-resolution, full-reference quality assessment, generative adversarial networks, image enhancement.**

I. Introduction

In order to upsample a low resolution (LR) image to become a high-resolution (HR) one, we need to use single image super-resolution (SISR) [1] technique, which is a hot research topic in image processing community. Compared to LR images, HR images have clear detail and high image quality. Using HR camera is the most direct way to obtain HR images. However, considering the high production cost and manpower, we cannot use an HR camera in many situations. Therefore, research on SISR is very important and practical, which can be used to replace HR cameras to obtain HR images from LR images.

Manuscript received November 2, 2018; revised February 25, 2019 and April 14, 2019; accepted April 15, 2019. Date of publication May 3, 2019; date of current version October 24, 2019. This work was supported by the National Natural Science Foundation of China under Grant 61772137. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abdulmotaleb El Saddik. *(Corresponding author: Bo Yan.)*

The authors are with the School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, China (e-mail: byan@fudan.edu.cn; 16110240015@fudan.edu.cn; 17210240039@fudan.edu.cn; 15110240012@fudan.edu.cn; 14110240025@fudan.edu.cn).

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/TMM.2019.2914883

For addressing SISR, there are many solutions have been developed. Bicubic interpolation and Lanczos resampling [2] are the representative methods of early solutions, which aim to only use the information of an LR image to get an HR image. Since SISR is an ill-posed problem, that is, one pixel from a LR image corresponds to multiple pixels in the HR version. Therefore, early methods cannot address the SISR problem very well. In order to handle this problem well, we need to employ strong prior knowledge to aid the algorithm to find the right mapping. Thus, in order to get better HR results, recent methods prefer to adopt example-based [3] strategy. These methods can be classified into two categories based on whether it depends on external examples or not. Inner-example based SR methods [4]–[6] exploit internal similarities of the same image. External-example based methods [7]–[12] learn a mapping function from external low- and high-resolution example pairs. It is worth noting that image and video vectorization [13], [14] can also be used to upsample image and videos.

Recently, with the success of deep learning in image recognition tasks, there are many deep learning based SISR methods emerged, especially based on convolutional neural networks (CNNs) [16]. These methods [17]–[20] have better performance than previous methods. Among them, SRCNN [17] is a pioneer work, which proves the superiority of CNNs and a large amount of LR-HR training samples in solving the SISR problem. In [24], Kim *et al.* [18] developed a 20 layer CNNs model for SISR problem. Through their experimental results, we found that performance of their very deep SR (VDSR) model can be about 1.0 dB higher than SRCNN. In our previous work [21], we developed a better CNNs-model than VDSR with same parameter capacity.

However, due to the evaluation metrics of SISR methods are peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [22], most of the CNNs-based SISR methods employ L2-norm as the loss function. As revealed in [23], this will lead to the significant increase in PSNR or SSIM value but little improvement in the perceptual quality. Thus, in order to improve the perceptual quality, some methods [24], [25] employed perceptual loss (difference between the feature maps of VGG-Net [26]) as loss function. Besides that, Ledig *et al.* [25] employed generative adversarial networks (GAN) to further improve the perceptual quality of the generated image. As proved by the authors of [23], GAN-based SISR methods have the best perceptual quality accompanied by poor PSNR and SSIM value.

In SISR benchmark study [3], such as PSNR and SSIM are proved to have lower consistency with human visual system (HVS). So, it is very important to develop a specific quality

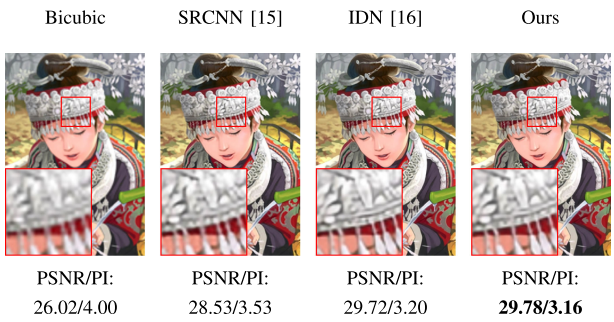


Fig. 1. Illustration of our proposed method. We propose a balanced SISR method, which can achieve higher PSNR value and perceptual quality than compared methods. Our proposed SISR method is guided by the proposed CNNs-based FR-IQA method, which is designed for the quality assessment of SISR methods. We test the perceptual quality of the generated results with perceptual index (PI) value (lower is better) that used in PIRM2018-SR Challenge [15].

assessment method for SISR. In [3], information fidelity criterion (IFC) [27] is proved to be more consistent with HVS than PSNR and SSIM. No-reference image quality assessment (NR-IQA) for SISR methods is very challenging and more practical than full-reference image quality assessment (FR-IQA) methods. To develop an NR-IQA metric for SISR methods, Ma *et al.* [28] built a database at first and then employed a two-stage regression model to learn a mapping between 138 hand-crafted features and the quality score of each training image. They achieved better performance than previous methods including IFC. In [29], Fang *et al.* proposed a CNNs-based NR-IQA method for SISR and achieved better performance than commonly used methods. However, as revealed in the experimental results in [28], the performance of shallow CNNs-based methods may lower than Ma *et al.*'s method. In order to address mentioned problems, authors of [30] developed a deeper CNNs-model and achieved state-of-the-art performance. Inspired by this method, we improve it to have better performance when it used as a part of loss function for SISR methods.

In this paper, in order to address the mentioned problems of the existing methods, we propose a balanced SISR method, which has better PSNR value and perceptual quality (Fig. 1). Inspired by previous objective quality assessment for SISR methods, we develop a deep FR-IQA method for SISR methods and employ this method as a part of the loss function of our proposed SISR method. Our proposed method is benefited from the proposed highway unit and the novel loss function. Specifically, the main contributions of our work lie in three aspects:

- We propose a deep FR-IQA method for SISR methods, which can handle images with arbitrary sizes. Then, we employ this IQA method to be a part of loss function of SISR methods to improve the perceptual quality of the generated HR results while keeping the PSNR values do not drop too much.
- We propose a CNNs-based SISR method, which is guided by our proposed novel loss function. Besides that, our proposed SISR method consists of several highway units, which are designed for combining the input signal and output signal with trainable weights.
- As proved by experimental results, by the guidance of the proposed FR-IQA method, our proposed SISR method outperforms most of the methods in both PSNR value and perceptual quality. Furthermore, we show that GAN-based SISR method can be guided by the designed novel loss function to achieve better performance.

The remainder of the paper is organized as follows. In Section II, we introduce related works. We describe our proposed SISR method in Section III. In Section IV, we present and discuss experimental results. Finally, we draw conclusions in Section V.

II. Related Works

A. CNNs-Based SISR Methods

In recent years, deep learning based SISR methods emerged and achieved better performance than previous methods. SRCNN [17] is the first CNNs-based SISR method. Then, Kim *et al.* [18] proposed a deeper VDSR model and introduced the concept of residual learning for CNNs-based SISR methods. In [19], authors developed an SISR method based on deep recursive convolutional network (DRCN). Inspired by previous works, Tai *et al.* [31] proposed a deep recursive residual networks (DRRN) for SISR. However, SRCNN, VDSR, DRCN, and DRRN have high computation cost due to bicubic interpolation process of the first step. So, in order to decrease the processing time of CNN-based SR methods, some methods tried to process images at low scale and upsample images at the end with sub-pixel convolution [20] or transpose convolution [32]. Lai *et al.* [33] proposed a deep laplacian pyramid networks for SISR and achieved pleasing results with fast speed. In order to achieve the same goal, Hui *et al.* [34] proposed a fast and accurate SISR by utilizing information distillation network (IDN). NTIRE challenge [35] boosts the community to develop deeper models with better performance when neglecting the running time. EDSR [36] utilized optimized residual block to train a very deep model and won the first place in NTIRE 2017 challenge [35]. Authors of [37] developed a residual dense networks and achieved better performance than EDSR. In [38], Zhang *et al.* proposed a deep SISR method based on channel attention residual block and achieved better performance than previous deep methods.

In order to generate visually pleasing HR results, authors in [24] proposed a perceptual loss for SISR. In [25], authors tried to solve the SR problem with the help of GAN (SRGAN) and achieved the best-looking SR results ever. Inspired by SRGAN, Wang *et al.* [39] proposed an enhanced SRGAN (ESRGAN) and won the first place in PIRM2018-SR Challenge [15]. Although SISR methods, which are based on perceptual loss and GAN, achieve better perceptual quality than previous methods, they always generate lower PSNR values than previous methods. This is because these methods generate much high frequency information, which may not necessarily correspond to the original image.

B. Objective Quality Assessment Metrics

Objective image quality assessment methods can be classified into three categories: NR, FR, and reduced reference (RR) IQA methods. FR-IQA methods need the full information of reference image, while NR-IQA methods predict the quality of

distorted images with no access to the reference image. RR methods [40], [41] provide a trade-off between FR and NR methods, which only need the feature or partial information of the reference image.

PSNR and mean squared error (MSE) are the simplest and widely used FR-IQA methods. However, the performances of these two methods are not consistent with the HVS because of only taking pixel level difference into account. In [22], Wang *et al.* proposed SSIM algorithm, which integrates structural similarity with the pixel-level difference and achieved better consistency with HVS. In [42], Zhang *et al.* proposed a feature-based similarity (FSIM) index based on the fact that humans usually understand images based on low-level features. Besides FSIM, many other FR-IQA methods [43]–[45] have been developed based on the ideas of SSIM.

For deep learning based FR-IQA methods, Gao *et al.* utilized the different level feature maps of a pre-trained deep CNNs model for similarity calculations. Liang *et al.* [46] trained a dual-path Siamese network for predicting the quality score for a distorted image. In [47], Kim *et al.* developed a deep image quality assessment (DeepQA) model, which seeks the optimal visual weight based on the understanding of database information itself without any prior knowledge of the HVS. Bosse *et al.* [48] proposed two deep FR-IQA models. The main difference between these two models is the calculation of the final quality score from small patch scores.

Different from FR and RR IQA methods, NR-IQA is very challenging because it does not take any information about reference image into account when predicting quality for a distorted image. A common solution for NR-IQA problem is to extract features and train a regression model between extracted features and the subjective score from the quality assessment databases. Therefore, the main difference between different NR-IQA methods is how to extract quality-related features. Among various NR-IQA methods, a popular approach is to extract features using natural scene statistics (NSS). Typical NSS-based NR-IQA methods extract features from discrete wavelet transform (DWT) [49], discrete cosine transform (DCT) [50], or spatial domain [51]. In [52], Mittal *et al.* proposed a completely blind image quality analyzer using NSS features.

The other trend of NR-IQA is pure data-driven methods. Ye *et al.* [53] proposed code book representation for no-reference image assessment (CORNIA), which is the first pure data-driven method. Kang *et al.* [54] employed a CNNs-model to learn the mapping between the input image and the subjective score. Bosse *et al.* [48] proposed a deeper CNNs model for solving the NR-IQA problem. Talebi *et al.* [55] proposed a novel method, which predicts the distribution of human opinion scores using a CNNs model. Also, some other kind of NR-IQA methods are proposed in the literature. For example, these methods try to solve the NR-IQA problem with a k-nearest-neighbor (KNN)-based quality prediction model [56], a novel rank-order regularized regressor [57], a novel local learning method [58], an improved multi-scale local binary pattern [59], or from perspective of color information processing in the brain [60], etc. In [61], Wu *et al.* proposed a perceptually weighted rank correlation indicator for evaluation of IQA methods.

Above mentioned IQA methods are general-purpose methods, which are designed based on image signal and noise, while quality assessment of SISR methods should be designed based on visual perception. There are many specific SR-IQA methods proposed in the literature. Previous methods try to address SR-IQA using subjective evaluation [62]–[64], NSS [65]–[67], or energy change and texture variation [68]. In [28], Ma *et al.* proposed an NR-IQA method for SISR with a two-stage regression model. More importantly, for training the regression model, they built a SISR image quality assessment database. With the help of this database, Fang *et al.* [29] and Bare *et al.* [30] proposed different CNNs models for NR-IQA problem of SISR. Fang *et al.* extract low-level features using shallower network, whereas Bare *et al.* [30] extract high-level features using deeper network. Zhang *et al.* [69] built a huge patch similarity dataset and proved that deep features can be used as a perceptual metric. Among various specific IQA methods for SISR, Bare *et al.* achieve state-of-the-art performance. In this paper, we improve the method of Bare *et al.* [30] to adapt for the training of CNNs-based SISR methods.

C. Applications of Objective Quality Assessment Metrics

Objective quality assessment is a very active sub-discipline of image processing community, and many of the resulting methods have begun to benefit image processing algorithms. A direct application of IQA methods is to optimize and monitor the image processing methods. Many image processing algorithms such as denoising [70], [71], image and video coding [72], digital watermarking [73], [74], and various algorithms of other areas (e.g., fingerprint verification [75] and fake biometric detection [76]) have proved the usefulness of IQA methods. Besides, IQA methods play an important role in visual communication applications for the monitoring of the quality-of-service [77]–[79].

Same to other applications of IQA methods, we develop a CNNs-based FR-IQA method for SISR methods at the beginning of this paper. Then, we utilize this FR-IQA method to guide our proposed CNNs-based SISR method. The reason for using the proposed FR-IQA method to guide the proposed SISR method is to improve the perceptual quality of the SISR results while maintaining the PSNR and SSIM values of the result do not drop too much. The loss function used in previous methods does not improve the subjective quality when the PSNR value is increased, or significantly decreases the PSNR value when the subjective quality is improved. The common part of our proposed CNNs-based SISR method to previous CNNs-based methods is that our method also utilizes fully convolutional network architecture, residual learning, and transpose convolution to get the HR result. The different part between our proposed method and previous CNNs-based methods is that our proposed method is guided by the designed novel loss function and our CNNs architecture consists of several proposed highway units.

III. Our Proposed Method

We demonstrate the framework of our proposed method in Fig. 2. As demonstrated in this figure, our proposed SISR method is guided by the proposed FR-IQA method. Thus, we introduce

Fig. 2. Framework of our proposed method. Our proposed SISR method takes an LR image as input and outputs the HR result by the guidance of our proposed deep FR-IQA method for SR images and L2-norm.

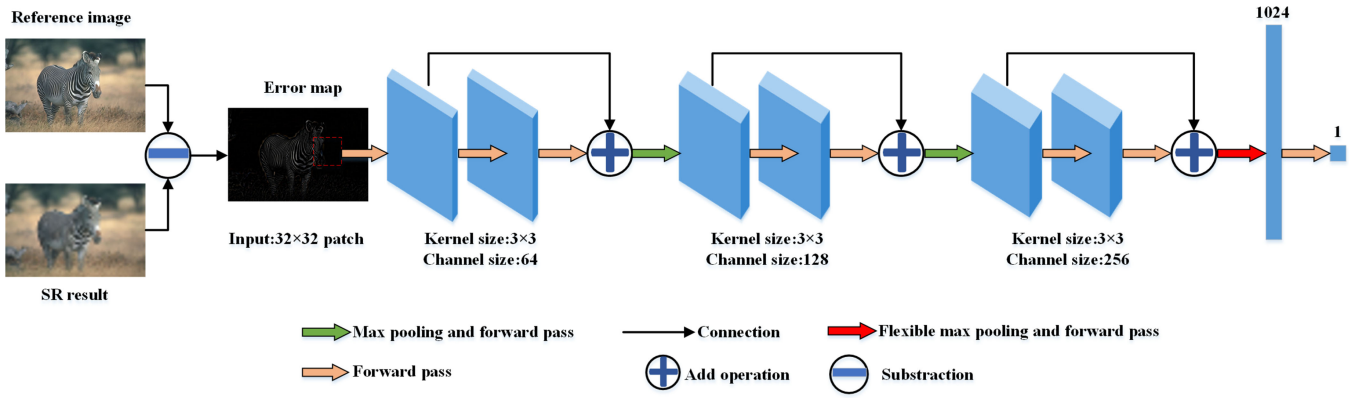


Fig. 3. Network structure of our proposed FR-IQA method. The input of this method is a 32×32 patch from the error map between an SR result and its corresponding ground truth reference image. This method contains three residual blocks. The channel number of each residual block is gradually doubled. The output of the third residual block is flattened and is sent to a fully connected layer to get a 1024-dimension feature vector. The final quality score is obtained based on the feature vector.

the proposed FR-IQA method at first. Then, we introduce our proposed SISR method and the loss function of our proposed SISR method, respectively.

A. Our Proposed FR-IQA Method

The reason of developing a deep learning based FR-IQA method for SISR is that reference images are always available in the training of a CNNs-based SISR method. To the best of our knowledge, there is no deep learning based specific FR-IQA model for SISR methods is proposed in the literature. Therefore, we develop a deep learning based FR-IQA method for SISR methods. For developing this FR-IQA method, we choose to improve the state-of-the-art deep learning based SR-IQA method [30] to adapt for the training of different SISR methods. We improve the method of Bare *et al.* [30] from three aspects:

- Firstly, we change the input of [30] with the error map between reference image and the SR result.
- Secondly, we remove the normalization process that used in [30], namely the input of our proposed FR-IQA method is the difference between the luminance map of reference image and SR result.
- Thirdly, in order to adapt the proposed FR-IQA method to guide the different SISR methods without resizing process, we adopt flexible max pooling layer between the first fully connected layer and the third sum layer. The pooling window size of the flexible max pooling is equal to the output feature map size of the previous layer. Therefore, flexible max pooling layer changes the input to the fixed size and send it to the first fully connected layer.

Network structure of the proposed FR-IQA method is shown in Fig. 3. As demonstrated in this figure, the input of our proposed model is a 32×32 small patch. This small patch is from the error map, which is calculated by subtracting the SR result from the reference image. Specifically, our proposed method extracts 32×32 small patches without overlaps from the error map firstly. Then, our proposed FR-IQA method predicts quality score for each small patch. Finally, our method calculates the average score of each small patch to be the final quality score of the tested image.

Except for the mentioned difference between Bare *et al.* [30], other part of the network architecture is same to Bare *et al.* [30]. Therefore, we briefly describe the network architecture in this paper. As demonstrated in Fig. 3, our proposed FR-IQA method employs three residual blocks to extract features. Each residual block contains two convolutional layers. The first convolutional layer in each residual block takes rectified linear unit (ReLU) as activation function. After each residual block, we use max pooling to enlarge the receptive fields, thus the feature map size of convolutional layers in residual block becomes half of the

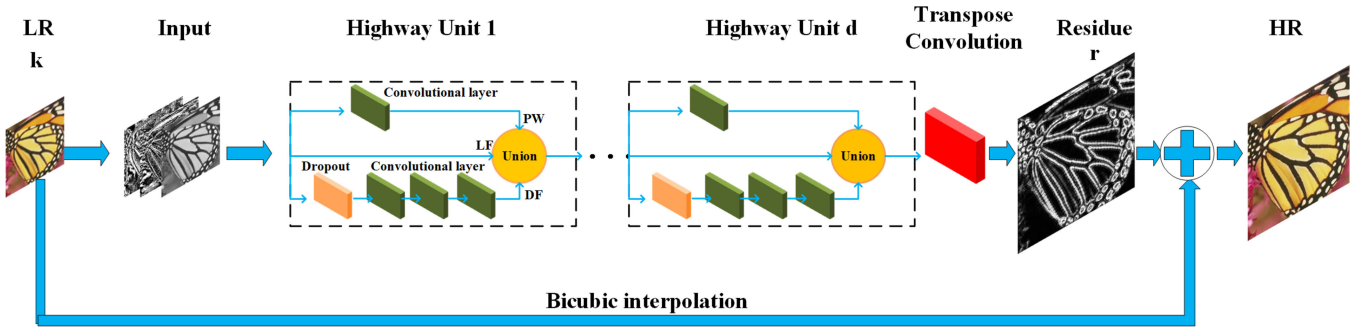


Fig. 4. Network architecture of our proposed SISR method. At first, we preprocess an LR image k to obtain input of our network. Then, new input go through an input convolutional layer, cascaded highway units, and a transpose convolutional layer to get the residual map r. Finally, the HR version of LR image is obtained by adding r to the bicubic interpolated version of k. It should be noted that the first two convolutional layers of the lower branch of the highway unit take ReLU as the activation function and the third convolutional layer has no non-linear activation because of residual learning. Besides, the convolutional layer of upper branch of highway unit takes sigmoid as the activation function.

convolutional layer of previous residual block. So, in order to extract features better, we enlarge the channel size to become the double of the previous one. After the first and the second residual blocks, we add max pooling layer to enlarge the receptive fields. Because we set the pooling window size to 2×2 with stride 2, the feature map size of the convolutional layer becomes half of the previous one after max pooling. The flexible max pooling layer is able to change the input to the fixed size so as to adapt for the training of different SISR methods without resizing process.

B. Our Proposed SISR Method

Based on our previous work [21], in order to fairly compare our method with state-of-the-arts, we expand the channel size of each convolutional layers from 32 to 64. Then, significantly different from previous SISR methods, we employ our proposed novel loss function to guide our proposed SISR network to generate results with better perceptual quality and competitive PSNR and SSIM values. In order to fasten the running speed, we change the input of our previous work to LR image and employ transpose convolution to upsample it at the end, which is another difference from [21].

1) Network Architecture: The framework of our proposed network is demonstrated in Fig. 4. For faster convergence, our network preprocesses an LR image k and obtains three images: horizontal gradient map, vertical gradient map, and luminance map. Then, our proposed network combines these three maps to create a new input signal. The combined new signal passes through an input convolutional layer, several highway units, and a transpose convolution to obtain a residual image r. Finally, the output HR image is obtained by combining the bicubic upsampled version of k and predicted residual image r. This process can be formulated as:

$$r = C_{out}(HU_d(HU_{d-1}...(HU_1(C_{in}(Input)))...)) \quad (1)$$

where HU denotes the highway unit, d denotes the total number of highway unit, C_{in} denotes the input convolutional layer, C_{out} denotes the output transpose convolutional layer, $Input$ denotes the combination of horizontal gradient map, vertical gradient map, and luminance map. Final HR image can be obtained by:

$$HR = Bicubic(LR) + r \quad (2)$$

where r denotes the learned residual image, $Bicubic(LR)$ denotes the upsampled LR image with bicubic interpolation.

The main novelty of this SISR network is the highway unit. Unlike image recognition task, the purpose of SISR is to generate visually pleasing and clear HR result. Because each pixel in an image is not perceptually equal, we design a highway unit to learn individual weight value for each pixel, then utilize this weight value to weightily add the input and output signal of highway unit. By sending the processed LR image to cascaded highway units, we can get accurate residual image r. Thus, benefit from cascaded highway unit architecture, the final HR result is also visually pleasing. We demonstrate the specific architecture of highway unit in Fig. 4. As shown in this figure, each highway unit contains upper and lower two branches. Lower branch contains three convolutional layers to handle the input lower feature LF to become deeper feature DF. Each convolutional layer has 64 kernels size of 5×5. It should be noted that the first two convolutional layers of the lower branch take ReLU as an activation function and the third convolutional layer has no non-linear activation because of residual learning. In order to prevent overfitting problem, we add dropout [80] to the beginning of lower branch. Upper branch aims to generate weight value PW for each pixel. This branch contains a convolutional layer, which also has 64 kernels size of 5×5, and a sigmoid layer. The reason of adopting sigmoid as the non-linear activation in the convolutional layer of upper branch is that the upper branch is responsible for generating weight values range in [0,1]. LF and DF are combined by weight value through union layer, this process can be formulated as:

$$CO = PW \times DF + (1 - PW) \times LF \quad (3)$$

where CO denotes the combined output from the union layer.

2) Novel Loss Function: We design a new loss function for CNNs-based SISR. Previous CNNs-based methods [17]–[19] employ L2-norm as loss function. As revealed in [23], as the L2-norm decreases, the perceptual quality will get worse. Therefore, in order to improve the PSNR value and perceptual quality of the

generated images simultaneously, we design a new loss function. Our designed new loss function is a weighted sum of the L2-norm and SR-IQA loss. Because the value of our proposed SR-IQA method is between 0 and 1, 1 represents the best, and the 0 represents the worst, we adopt $|1 - SRIQA|$ as SR-IQA loss. Thus the L2-norm loss L_2, SR-IQA loss L_{SRIQA}, and the total loss function $Loss$ can be defined as:

$$L_2 = ||G(LR;\theta) - GT||_2 \tag{4}$$

$$L_{SRIQA} = ||1 - SRIQA(GT - G(LR;\theta))||_1 \tag{5}$$

$$Loss = L_{SRIQA} + \alpha * L_2 \tag{6}$$

$$\theta = \arg\min Loss \tag{7}$$

where $G(LR;\theta)$ denotes the generated HR patch from the LR by our proposed SISR network G with weights θ, GT denotes the ground truth HR patch , $SRIQA()$ denotes our proposed FR-IQA method for SISR, and α denotes the weight of L2-norm. In order to make the two loss at the same order of magnitude, we take $\alpha = 0.1$ in our experiments.

IV. Experiments

In this section, we introduce conducted experiments. First, we introduce the training details of our proposed FR-IQA method and SISR method. Then, we present ablation study and the parameter selection experiments of our proposed SISR method. Moreover, we present qualitative and quantitative experiment results. Finally, we show that our novel loss function can be used in GAN-based SISR method to further improve the perceptual quality.

A. Training of Our Proposed FR-IQA Method

Dataset: We train and test our proposed FR-IQA method by using Ma *et al.*'s [28] dataset. This dataset is built by processing 30 images from BSD [81] with 9 different SISR methods: bicubic interpolation (Bicubic), back projection (BP) [1], Shan08 [82], Glasner09 [6], Yang10 [9], Dong11 [83], Yang13 [11], Timofte13 [8], and SRCNN [17]. Results of these 9 different SISR methods is obtained by setting the downsample factor and the corresponding kernel width to 6 different values. Specifically, downsample factors $s \in \{2, 3, 4, 5, 6, 8\}$ and corresponding kernel width factors $\sigma \in \{0.8, 1.0, 1.2, 1.6, 1.8, 2.0\}$. Overall, this dataset has 1620 results from 9 different SISR methods and each result has subjective perceptual score, which is obtained by averaging the subjective scores from trained volunteers.

Evaluation metric: In order to fairly compare our method with previous methods, following Ma *et al.* [28] and Bare *et al.* [30], we utilize spearman rank correlation coefficients (SROCC) value to assess the correlation degree between predicted objective scores rank and the subjective perceptual scores rank. This value indicates how much the two ranks can be described by a monotonic function. This value is between -1 and 1. 1 denotes two ranks are totally same, and -1 denotes that two ranks are totally different.

Training details: At the beginning of our model training, we extract non-overlapped 32×32 patches from the difference map of each image and the corresponding reference image in the training set. Then, we begin the training of our model. Learning rate begins from 1e-2 and ends at 1e-5. It becomes the one-tenth of the previous one every 10 epochs. In order to better train the proposed model, we adopt gradient clip to control the gradient in range $[-\beta/\gamma, \beta/\gamma]$, where γ denotes current learning rate and β denotes the clip control value, which is set to 0.1. We stop the training of our model after 40 epochs. We train our model by applying deep learning toolbox MatConvNet [84] and it takes 30 minutes to train a model with GTX1070 GPU.

B. Training of Our Proposed SISR Method

Training dataset: Following previous works [18], [19], [31], [33], [34], we train our proposed SISR model on 291 images, which are collected from 91 images by Yang *et al.* [9] and additional 200 images from Berkeley segmentation dataset (BSD) [81]. We extract 36×36 patches with stride 24 from each training image. In order to obtain more training data, we use rotation, horizontal flip, and downsampling to augment the training data. After data augmentation, our training set has about 1,400,000 image patches.

Testing dataset: We use four benchmark datasets to evaluate our proposed SISR method. These datasets are: "Set5" [7], "Set14" [10], "B100" [81], and "Urban100" [85]. Among them, "Set5" [7] and "Set14" [10] are generic datasets used in previous works [17], [18], [34]. "Urban100" [85] contains many challenging images failed by many of the existing methods. "B100" is the testing set of BSD [81].

Training details: In order to fairly compare our proposed model with state-of-the-art methods, we adopt 7 highway units in our experiments. This assures our proposed SISR method has similar or lower parameter size with state-of-the-arts. We train two networks for each scale. The first one is our proposed SISR method, which is trained only using L2-norm. The second one is our proposed method, which is trained on L2-norm and our proposed SR-IQA loss. In the training of our proposed SISR method, we fixed the parameters of our proposed FR-IQA model. In our experiments, the batch size is set to 64, weight decay is set to 0.0001, momentum is set to 0.9, and dropout ratio is set to 0.1. The kernel size and stride of the output transpose convolutional layer for $\times 2$, $\times 3$, and $\times 4$ models are set to $(4 \times 4, 2)$, $(6 \times 6, 3)$, and $(8 \times 8, 4)$, respectively. We train all the networks for 30 epochs. The learning rate begins with 1e-2 and divided by 10 after every 10 epochs. We also use gradient clip to limit the parameters' gradient in $[-\beta/\gamma, \beta/\gamma]$, where γ denotes current learning rate and β is set to 0.1. It takes about 16 hours to train a model with GTX1070 GPU by applying deep learning tool box MatConvNet [84].

C. Validation of Our Proposed FR-IQA Method

Following previous methods [28], [30] , we employ 5-fold validation to verify the performance of our model. Specifically, we randomly divide the training dataset to 5 folds. Then, we utilize 4 folds for training and the remaining one is for testing. Repeat this process 5 times to ensure that each fold is selected as a testing set once. After 5 iterations, we can obtain the predicted

TABLE I
SROCC VALUE COMPARISON OF OUR PROPOSED FR-IQA METHOD WITH STATE-OF-THE-ARTS. WE DEMONSTRATE THE OVERALL SROCC VALUE ON THE WHOLE SISR IQA DATASET AND THE SROCC VALUES OF IQA METHODS ON EACH INDEPENDENT SISR METHOD

	Bicubic	Bp	Shan08	Glasner09	Yang10	Dong11	Yang13	Timofte13	SRCNN	Overall
PSNR	0.572	0.620	0.564	0.605	0.625	0.634	0.631	0.620	0.645	0.604
FSIM [42]	0.706	0.770	0.648	0.778	0.757	0.765	0.768	0.756	0.780	0.747
SSIM [23]	0.588	0.657	0.560	0.648	0.649	0.649	0.652	0.656	0.660	0.635
IFC [28]	0.884	0.880	0.934	0.890	0.866	0.865	0.870	0.881	0.885	0.810
DIVINE [50]	0.784	0.842	0.653	0.426	0.525	0.763	0.537	0.122	0.625	0.589
CNNIQA [55]	0.926	0.956	0.832	0.914	0.943	0.921	0.927	0.924	0.908	0.904
CORNIA [54]	0.889	0.932	0.907	0.918	0.908	0.912	0.923	0.911	0.898	0.919
BLIINDS [51]	0.886	0.931	0.664	0.862	0.901	0.811	0.864	0.903	0.843	0.853
BRISQUE [52]	0.850	0.917	0.667	0.738	0.886	0.783	0.784	0.843	0.812	0.802
Ma *et al.* [29]	0.933	0.966	0.891	0.931	0.968	0.954	0.958	0.930	0.949	0.931
Bare *et al.* [31]	0.973	0.977	0.926	0.950	0.971	0.955	0.971	0.934	0.953	0.958
Ours	**0.980**	**0.982**	**0.937**	**0.975**	**0.976**	**0.970**	**0.982**	**0.947**	**0.977**	**0.967**

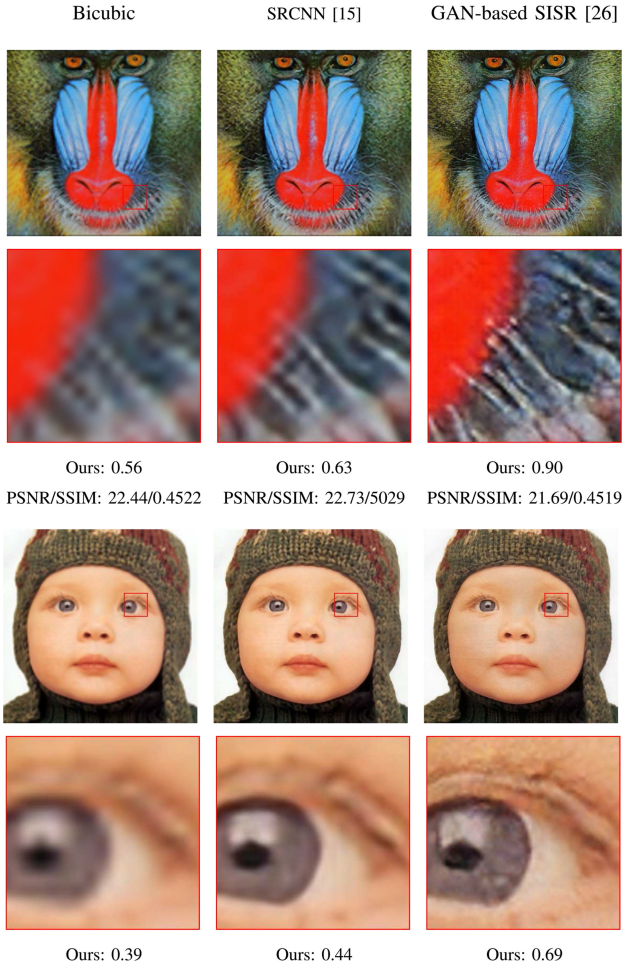


Fig. 5. An example of the quality prediction ability of the proposed FR-IQA method. From left to right: Bicubic, SRCNN, and GAN-based SISR. It is obvious that the perceptual quality is improved from left to right. Our proposed FR-IQA method (higher is better) is able to accurately predict the perceptual quality of the images than PSNR and SSIM (both higher is better).

score by our method for each image in dataset. We run 5-fold validation for 50 times and demonstrate the average value of 50 test results.

We compare our proposed method with PSNR, SSIM [22], IFC [27], FSIM [42], DIVINE [49], CNNIQA [54], CORNIA [53], BLIINDS [50], BRISQUE [51], Ma *et al.* [28], and Bare *et al.* [30]. Among them, the first four methods are the FR-IQA methods and the other methods are the NR-IQA methods. All data-driven methods are retrained on the Ma *et al.*'s database.

TABLE II
AVERAGE RUNNING TIME COMPARISON BETWEEN OUR PROPOSED FR-IQA METHOD AND OTHER TWO SR-IQA METHODS ON "SET5" [7] DATASET

	Running Time (s)
Ma *et al.* [29]	13.80
Bare *et al.* (CPU) [31]	0.95
Bare *et al.* (GPU) [31]	0.29
Our proposed method (CPU)	0.64
Our proposed method (GPU)	0.17

We compare our results with other methods on overall SROCC value and the separate SROCC value on each different SISR methods in the dataset. The SROCC values are computed on whole dataset and the whole set of each specific SISR method. Experimental results are displayed in Table I. As we can observe from this table, our proposed FR-IQA method outperforms other compared methods on each SROCC value comparison. It is worth noting that our proposed FR-IQA method achieves 0.967 in terms of overall SROCC value, which indicates a high correlation between our proposed FR-IQA method and the HVS.

Besides quantitative evaluation, we provide an example of the quality prediction ability of our proposed FR-IQA method. This example is demonstrated in Fig. 5. From this figure, we can easily find that the perceptual quality of the given images are improved from left to right. Our model can provide accurate perceptual quality prediction. We also provide the results of PSNR and SSIM for better demonstrating the quality prediction ability of our method. As we can observe from Fig. 5, PSNR and SSIM make a wrong decision when encounter with the results of GAN-based SISR, which is not consistent with HVS.

In Table II, we compare the running time of our proposed FR-IQA model with Ma *et al.*'s [28] method and the Bare *et al.*'s method [30]. This test is implemented on GTX1070 GPU and Intel i5-8400 CPU 2.80 GHZ. As shown in Table II, speed of our proposed model is 21 times faster on CPU and 81 times faster on GPU. Because Ma *et al.*'s method is not a deep learning

TABLE III
PERFORMANCE COMPARISON OF OUR SISR METHOD WHEN THE DROPOUT RATIO IS SET TO DIFFERENT VALUES

	Set5		Set14		B100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ours w/ dropout ratio=0.5	31.58	0.8855	28.17	0.7702	27.34	0.7265	25.25	0.7555
ours w/ dropout ratio=0.4	31.67	0.8867	28.20	0.7705	27.36	0.7267	25.31	0.7569
ours w/ dropout ratio=0.3	31.64	0.8869	28.15	0.7697	27.35	0.7270	25.27	0.7567
ours w/ dropout ratio=0.2	31.66	0.8873	28.20	0.7711	27.37	0.7274	25.33	0.7585
ours w/ dropout ratio=0.1	**31.70**	**0.8875**	**28.21**	**0.7714**	**27.38**	**0.7278**	**25.36**	**0.7596**
ours w/o dropout	31.70	0.8873	28.18	0.7696	27.36	0.7267	25.33	0.7575

TABLE IV
PERFORMANCE COMPARISON OF OUR SISR METHOD WHEN THE NETWORK ARCHITECTURE CHANGES TO DIFFERENT. WE REMOVE RELU AND RESIDUAL LEARNING FROM THE NETWORK STRUCTURE, IN TURN, TO VERIFY WHETHER THESE COMPONENTS ARE USEFUL. WE ALSO REPLACE THE PROPOSED HIGHWAY UNIT WITH THE RESBLOCK AND INCEPTION LAYER TO VERIFY WHETHER THE UNIT CAN BRING MORE PERFORMANCE IMPROVEMENTS. FINALLY, WE ADJUST THE SIZE OF THE KERNEL OF EACH CONVOLUTIONAL LAYER FROM 5×5 TO 3×3 TO COMPARE THE PERFORMANCE

	Set5		Set14		B100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Ours w/o ReLU	31.46	0.8816	27.97	0.7651	27.27	0.7226	25.06	0.7456
Ours w/o residual learning	31.66	0.8863	28.17	0.7701	27.35	0.7265	25.34	0.7562
Ours w/ inception	30.56	0.8582	27.36	0.7496	26.96	0.7122	24.46	0.7166
Ours w/ resblock	31.68	0.8873	28.22	0.7715	27.35	0.7268	25.28	0.7568
Ours w/ convolution kernel=3×3	**31.73**	**0.8882**	**28.23**	**0.7717**	27.36	0.7274	25.31	0.7587
Ours	31.70	0.8875	28.21	0.7714	**27.38**	**0.7278**	**25.36**	**0.7596**

based method, we are not able to test it on the GPU. It is worth noting that, our proposed FR-IQA method is faster than [30], this is because we adopt flexible max pooling layer in our method, whereas Bare *et al.* [30] do not adopt it in their method.

In order to apply our proposed FR-IQA method to aid the SISR methods to achieve better perceptual quality, we retrain this FR-IQA model on the whole SISR quality assessment dataset with the original images from Ma *et al.* [28]. In the experiment section of the paper, we train our proposed SISR method and the GAN-based SISR method with the aid of our proposed FR-IQA method.

D. Network Parameter Study of Our Proposed SISR Method

1) Parameter Selection For Dropout Layer: Our designed highway unit contains a dropout layer at the beginning. This design is different from the dropout layer that used in famous image recognition CNNs models. In image recognition models, dropout layers are usually placed after fully-connected layers. This is because fully connected layers have a lot of parameter than convolutional layers. However, dropout ratio of the dropout layers that used in image recognition models usually are set to larger values such as 0.5. Since convolutional layers have fewer parameters than fully connected layers, we set the dropout ratio from 0.5 to 0.1 for finding an optimal value for SISR task. Also, we remove the dropout layers from our SISR method for comparison. Experimental results are demonstrated in Table III. As can be seen from this table, the performance of our SISR method is the best when the dropout ratio is set to 0.1. Thus, we set the dropout ratio to 0.1 in our other experiments.

2) Ablation Study for Network Architecture: In order to verify whether the final network architecture is the best, we conduct an ablation study about network architecture. The experimental results are listed in Table IV. As can be seen from this table, we firstly remove the ReLU from our SISR method to check whether activation functions are necessary for SISR methods. Secondly, we remove the residual learning from our network architecture, namely the output of our network architecture is the output of the transpose convolutional layer. Then, we replace the proposed highway unit with resblock and inception layer to check whether the proposed highway unit is the optimal for SISR task. Finally, we change the kernel size of convolutional layers from 5×5 to 3×3 to check whether the 5×5 is the optimal parameter for convolutional layers. Our final model has ReLU activation function, residual learning, highway unit, and convolutional layers with 5×5 size. As demonstrated in Table IV, ReLU, residual learning, and highway unit increase the performance of our proposed model with different degree. As for convolution kernel size, although convolution with 3×3 size achieves better performance on smaller datasets, we choose 5×5 size because of the performance is better than 3×3 on larger datasets.

3) Loss Function Selection: In order to verify the proposed novel loss function whether can achieve better perceptual quality while keeping the PSNR and SSIM values, we compare our novel

TABLE V
PERFORMANCE COMPARISON WHEN THE LOSS FUNCTION OF OUR SISR METHOD IS SET TO DIFFERENT. WE DEMONSTRATE THE PERFORMANCE OF OUR METHOD WHEN ADOPTING L2-NORM, PERCEPTUAL LOSS, AND OUR PROPOSED LOSS FUNCTION AS THE LOSS FUNCTION, RESPECTIVELY

	Set5			Set14			B100			Urban100		
	PSNR	SSIM	PI	PSNR	SSIM	PI	PSNR	SSIM	PI	PSNR	SSIM	PI
Our model with L2-norm+perceptual loss	24.36	0.8371	**4.78**	21.44	0.7028	**4.15**	22.30	0.6587	**4.05**	19.88	0.6822	**4.22**
Our model with L2-norm	**31.70**	**0.8875**	6.42	**28.21**	**0.7714**	5.90	**27.38**	**0.7278**	5.72	**25.36**	**0.7596**	5.59
Our model with L2-norm+SR-IQA loss (proposed)	31.61	0.8860	6.05	28.14	0.7699	5.84	27.32	0.7268	5.60	25.27	0.7559	5.52

TABLE VI
MEAN PSNR, SSIM, AND PERCEPTUAL INDEX (PI) VALUES COMPARISON OF OUR PROPOSED SISR METHOD WITH STATE-OF-THE-ARTS ON "SET5" [7] AND "SET14" [10] DATASETS. WE ALSO CALCULATE THE P-VALUE BETWEEN OUR PROPOSED METHOD AND EACH COMPARED METHOD ON EACH QUANTITATIVE VALUES. FOR THE CONVENIENCE OF OBSERVATION, WE HIGHLIGHT THE TWO BEST RESULTS

		Set5						Set14					
		PSNR		SSIM		PI		PSNR		SSIM		PI	
method	scale	mean	p-value	mean	p-value	mean	p-value	mean	p-value	mean	p-value	mean	p-value
Bicubic	×2	33.66	5.18E-02	0.9299	1.41E-01	5.54	1.19E-03	30.24	2.29E-02	0.8690	3.84E-02	4.78	5.63E-04
SRCNN	×2	36.66	2.83E-01	0.9542	4.10E-01	4.05	1.88E-01	32.38	2.72E-01	0.9027	3.00E-01	3.87	1.95E-01
VDSR	×2	37.53	4.36E-01	0.9587	4.80E-01	3.74	4.30E-01	33.05	4.31E-01	0.9127	4.60E-01	3.67	4.36E-01
DRCN	×2	37.63	4.55E-01	0.9588	4.82E-01	3.88	3.00E-01	33.06	4.32E-01	0.9121	4.49E-01	3.71	3.80E-01
LapSRN	×2	37.52	4.20E-01	0.9591	4.71E-01	3.83	3.34E-01	32.99	4.08E-01	0.9124	4.41E-01	3.76	3.20E-01
DRRN	×2	37.74	4.76E-01	0.9591	4.86E-01	3.89	3.19E-01	33.25	4.81E-01	0.9137	4.77E-01	3.74	3.56E-01
IDN	×2	37.83	4.93E-01	0.9600	5.00E-01	3.85	3.61E-01	33.30	4.92E-01	0.9148	4.98E-01	3.68	4.27E-01
HNSR	×2	37.81	4.89E-01	0.9597	4.95E-01	3.71	4.53E-01	**33.40**	5.18E-01	**0.9153**	5.07E-01	3.65	4.65E-01
Ours (LR_input, w/o SR_IQA_Loss)	×2	37.83	4.93E-01	0.9590	4.84E-01	3.78	3.97E-01	33.33	4.99E-01	0.9143	4.88E-01	3.64	4.80E-01
Ours (LR_input)	×2	37.67	4.62E-01	0.9578	4.66E-01	3.74	4.27E-01	33.26	4.83E-01	0.9136	4.75E-01	**3.63**	4.47E-01
Ours (Bicubic_input, w/o SR_IQA_Loss)	×2	**37.90**	5.06E-01	**0.9601**	5.02E-01	**3.69**	4.79E-01	**33.37**	5.11E-01	**0.9152**	5.04E-01	3.65	4.63E-01
Ours (Bicubic_input)	×2	**37.86**	5.00E-01	**0.9600**	5.00E-01	**3.67**	5.00E-01	33.33	5.00E-01	0.9149	5.00E-01	**3.62**	5.00E-01
Bicubic	×3	30.39	6.34E-02	0.8682	6.37E-02	7.10	8.28E-04	27.55	4.91E-02	0.7743	7.45E-02	6.66	3.03E-11
SRCNN	×3	32.66	2.19E-01	0.9088	3.09E-01	5.47	9.30E-02	29.19	2.95E-01	0.8155	3.10E-01	5.06	9.52E-03
VDSR	×3	33.66	4.00E-01	0.9213	4.54E-01	5.07	2.94E-01	29.78	4.40E-01	0.8318	4.60E-01	4.48	4.15E-01
DRCN	×3	33.82	4.36E-01	0.9226	4.69E-01	5.19	2.16E-01	29.77	4.38E-01	0.8314	4.56E-01	4.64	2.08E-01
LapSRN	×3	33.81	4.28E-01	0.9220	4.58E-01	5.42	9.11E-02	29.79	4.39E-01	0.8325	4.54E-01	4.68	1.71E-01
DRRN	×3	34.03	4.81E-01	0.9244	4.90E-01	4.93	4.63E-01	29.96	4.91E-01	0.8349	4.91E-01	**4.45**	4.64E-01
IDN	×3	34.11	5.04E-01	**0.9253**	5.01E-01	4.96	4.34E-01	29.99	4.94E-01	0.8356	4.97E-01	4.52	3.55E-01
HNSR	×3	34.03	4.84E-01	0.9248	4.95E-01	4.98	4.01E-01	30.00	4.95E-01	0.8358	5.00E-01	4.46	4.45E-01
Ours (LR_input, w/o SR_IQA_Loss)	×3	**34.15**	5.11E-01	0.9244	4.91E-01	5.18	2.34E-01	29.94	4.82E-01	0.8345	4.86E-01	4.67	1.99E-01
Ours (LR_input)	×3	33.91	4.57E-01	0.9221	4.63E-01	**4.91**	4.78E-01	29.91	4.74E-01	0.8332	4.74E-01	4.49	4.05E-01
Ours (Bicubic_input, w/o SR_IQA_Loss)	×3	**34.16**	5.13E-01	**0.9258**	5.07E-01	4.97	4.08E-01	**30.04**	5.05E-01	**0.8363**	5.04E-01	4.49	4.12E-01
Ours (Bicubic_input)	×3	34.10	5.00E-01	0.9252	5.00E-01	**4.90**	5.00E-01	**30.02**	5.00E-01	**0.8358**	5.00E-01	**4.43**	5.00E-01
Bicubic	×4	28.42	7.98E-02	0.8105	3.17E-02	7.18	2.67E-02	26.01	5.45E-02	0.7029	8.94E-02	6.90	3.40E-08
SRCNN	×4	30.49	2.52E-01	0.8628	2.19E-01	6.71	2.28E-01	27.46	2.88E-01	0.7476	3.09E-01	6.06	6.20E-03
VDSR	×4	31.35	4.00E-01	0.8838	4.31E-01	6.25	4.25E-01	28.02	4.29E-01	0.7678	4.54E-01	5.75	1.92E-01
DRCN	×4	31.53	4.41E-01	0.8854	4.50E-01	6.32	3.68E-01	28.03	4.33E-01	0.7673	4.50E-01	5.99	2.01E-02
LapSRN	×4	31.54	4.39E-01	0.8852	4.49E-01	6.43	2.91E-01	28.09	4.48E-01	0.7700	4.61E-01	6.00	2.15E-02
DRRN	×4	31.68	4.73E-01	0.8888	4.87E-01	5.98	4.88E-01	28.21	4.83E-01	0.7720	4.88E-01	5.51	4.82E-01
IDN	×4	**31.82**	5.06E-01	**0.8903**	5.03E-01	6.12	4.48E-01	28.25	4.91E-01	0.7731	4.95E-01	5.67	2.94E-01
SRGAN	×4	27.36	9.98E-03	0.8337	8.74E-02	**3.51**	9.99E-01	24.45	4.75E-03	0.6780	5.63E-02	**2.99**	1.00E+00
HNSR	×4	31.65	4.69E-01	0.8887	4.86E-01	6.21	4.42E-01	28.22	4.84E-01	0.7730	4.94E-01	5.80	1.30E-01
Ours (LR_input, w/o SR_IQA_Loss)	×4	31.70	4.79E-01	0.8875	4.72E-01	6.42	2.89E-01	28.21	4.81E-01	0.7714	4.82E-01	5.90	7.28E-02
Ours (LR_input)	×4	31.61	4.59E-01	0.8860	4.56E-01	6.05	4.62E-01	28.14	4.62E-01	0.7699	4.70E-01	5.84	1.96E-01
Ours (Bicubic_input, w/o SR_IQA_Loss)	×4	**31.79**	5.01E-01	**0.8903**	5.03E-01	6.23	4.29E-01	**28.29**	5.04E-01	**0.7740**	5.02E-01	5.74	1.92E-01
Ours (Bicubic_input)	×4	31.79	5.00E-01	0.8900	5.00E-01	**5.95**	5.00E-01	**28.28**	5.00E-01	**0.7737**	5.00E-01	**5.45**	5.00E-01

loss function with perceptual loss [24]. Same as [24], we utilize feature maps of conv2,2 layer of VGG19 network as perceptual loss. Other training parameters are same with our proposed SISR method. In order to better comparison, we also demonstrate the experimental results when only using L2-norm as loss function. Experimental results are listed in Table V. As can be seen from this table, although perceptual loss can achieve very promising perceptual quality, PSNR and SSIM values are dropped severely. This result does not meet our expectations. The performance of the proposed novel loss function is between the L2-norm and the perceptual loss. Our novel loss function can improve the perceptual quality when keeping the PSNR/SSIM values do not drop much. Therefore, we employ our proposed novel loss function instead of perceptual loss.

E. Validation of Our Proposed SISR Method

For fair comparison with the previous works, we convert RGB images to YCbCr and apply our method to luminance component. The color components for HR images are acquired from applying bicubic interpolation on LR images. We compute PSNR and SSIM values on luminance component. Since most of the compared methods take the bicubic interpolated version of an LR image as input, we also demonstrate the results of our method with bicubic input. We replace the transpose convolution with a convolutional layer with 7×7 kernel size in the bicubic input version of our proposed SISR method.

We compare the performance of our method with SRCNN [17], VDSR [18], DRCN [19], LapSRN [33], DRRN [31], and IDN [34]. We also compare our proposed SISR method with our

TABLE VII
MEAN PSNR, SSIM, AND PERCEPTUAL INDEX (PI) VALUES COMPARISON OF OUR PROPOSED SISR METHOD WITH STATE-OF-THE-ARTS ON "B100" [81] AND "URBAN100" [85] DATASETS. WE ALSO CALCULATE THE P-VALUE BETWEEN OUR PROPOSED METHOD AND EACH COMPARED METHOD ON EACH QUANTITATIVE VALUES. FOR THE CONVENIENCE OF OBSERVATION, WE HIGHLIGHT THE TWO BEST RESULTS

		B100						Urban100					
		PSNR		SSIM		PI		PSNR		SSIM		PI	
method	scale	mean	p-value	mean	p-value	mean	p-value	mean	p-value	mean	p-value	mean	p-value
Bicubic	×2	29.56	3.98E-06	0.8432	6.13E-08	4.05	3.07E-17	26.88	4.51E-12	0.8403	1.36E-14	4.96	3.94E-17
SRCNN	×2	30.82	1.24E-02	0.8719	1.55E-03	**3.02**	4.91E-01	29.00	1.69E-04	0.8796	2.19E-06	4.07	2.90E-01
VDSR	×2	31.90	3.73E-01	0.8960	3.93E-01	3.21	2.99E-01	30.77	2.37E-01	0.9141	2.67E-01	4.08	2.51E-01
DRCN	×2	31.85	3.43E-01	0.8942	3.19E-01	3.22	2.45E-01	30.76	2.34E-01	0.9133	2.38E-01	4.09	2.23E-01
LapSRN	×2	31.80	2.96E-01	0.8952	3.13E-01	3.29	6.50E-02	30.41	9.91E-02	0.9103	1.17E-01	4.18	5.09E-02
DRRN	×2	32.05	4.70E-01	0.8973	4.50E-01	3.26	1.55E-01	31.23	4.94E-01	0.9188	4.82E-01	4.00	4.70E-01
IDN	×2	32.08	4.91E-01	0.8985	5.08E-01	3.24	2.13E-01	31.27	5.18E-01	0.9196	5.24E-01	**3.96**	4.93E-01
HNSR	×2	**32.11**	5.12E-01	**0.8987**	5.17E-01	3.17	4.71E-01	**31.43**	6.10E-01	**0.9211**	5.98E-01	4.02	4.84E-01
Ours (LR_input, w/o SR_IQA_Loss)	×2	32.04	4.60E-01	0.8970	4.39E-01	3.20	3.58E-01	31.21	4.80E-01	0.9181	4.51E-01	4.05	3.77E-01
Ours (LR_input)	×2	32.00	4.39E-01	0.8965	4.14E-01	3.07	4.97E-01	31.15	4.43E-01	0.9177	4.28E-01	4.03	4.13E-01
Ours (Bicubic_input, w/o SR_IQA_Loss)	×2	**32.12**	5.19E-01	**0.8986**	5.13E-01	3.17	4.84E-01	**31.34**	5.54E-01	**0.9200**	5.44E-01	4.01	4.63E-01
Ours (Bicubic_input)	×2	32.09	5.00E-01	0.8984	5.00E-01	**3.01**	5.00E-01	31.24	5.00E-01	0.9191	5.00E-01	**3.94**	5.00E-01
Bicubic	×3	27.21	4.50E-04	0.7386	3.75E-05	6.47	1.88E-71	24.46	5.10E-08	0.7350	6.01E-12	6.63	4.20E-72
SRCNN	×3	28.24	8.96E-02	0.7776	5.40E-02	4.40	9.11E-03	26.08	6.74E-03	0.7910	3.30E-04	4.74	5.78E-02
VDSR	×3	28.83	4.01E-01	0.7976	4.08E-01	4.28	1.61E-01	27.14	2.76E-01	0.8279	2.58E-01	4.67	2.39E-01
DRCN	×3	28.80	3.85E-01	0.7963	3.74E-01	4.29	1.31E-01	27.15	2.81E-01	0.8277	2.52E-01	4.74	7.49E-02
LapSRN	×3	28.82	3.88E-01	0.7980	3.77E-01	4.30	1.09E-01	27.07	2.33E-01	0.8275	2.21E-01	4.79	2.55E-02
DRRN	×3	28.95	4.87E-01	0.8004	4.83E-01	4.34	6.76E-02	**27.53**	5.22E-01	**0.8378**	5.42E-01	**4.54**	4.70E-01
IDN	×3	28.95	4.86E-01	0.8013	5.08E-01	4.26	2.38E-01	27.42	4.49E-01	0.8360	4.87E-01	4.57	4.64E-01
HNSR	×3	28.95	4.89E-01	**0.8014**	5.11E-01	4.24	2.86E-01	27.49	4.91E-01	0.8362	4.95E-01	4.63	3.98E-01
Ours (LR_input, w/o SR_IQA_Loss)	×3	28.91	4.60E-01	0.7991	4.48E-01	4.45	5.19E-03	27.32	3.84E-01	0.8321	3.68E-01	4.79	2.90E-02
Ours (LR_input)	×3	28.89	4.43E-01	0.7989	4.43E-01	**4.14**	4.51E-01	27.34	3.93E-01	0.8316	3.54E-01	4.66	2.89E-01
Ours (Bicubic_input, w/o SR_IQA_Loss)	×3	**28.98**	5.13E-01	**0.8018**	5.20E-01	4.20	4.20E-01	**27.54**	5.23E-01	**0.8377**	5.41E-01	4.63	3.95E-01
Ours (Bicubic_input)	×3	**28.97**	5.00E-01	0.8010	5.00E-01	**4.08**	5.00E-01	27.50	5.00E-01	0.8364	5.00E-01	**4.51**	5.00E-01
Bicubic	×4	25.96	1.84E-03	0.6676	4.62E-04	6.77	3.35E-42	23.15	3.74E-06	0.6579	9.17E-10	6.83	1.21E-60
SRCNN	×4	26.84	1.29E-01	0.7058	9.26E-02	5.69	1.65E-03	24.47	2.60E-02	0.7181	2.63E-03	5.58	2.53E-04
VDSR	×4	27.29	3.98E-01	0.7252	4.07E-01	5.57	8.27E-02	25.18	2.78E-01	0.7525	2.40E-01	5.50	1.42E-02
DRCN	×4	27.24	3.62E-01	0.7233	3.67E-01	5.79	2.08E-05	25.14	2.51E-01	0.7511	2.14E-01	5.77	2.68E-09
LapSRN	×4	27.32	4.16E-01	0.7275	4.15E-01	5.73	1.10E-03	25.21	2.93E-01	0.7562	2.80E-01	5.58	6.02E-04
DRRN	×4	27.38	4.70E-01	0.7284	4.78E-01	5.49	3.19E-01	25.45	4.54E-01	0.7639	4.96E-01	5.20	4.83E-01
IDN	×4	27.41	4.89E-01	**0.7295**	5.03E-01	5.52	2.17E-01	25.41	4.27E-01	0.7630	4.75E-01	5.32	4.63E-01
HNSR	×4	27.38	4.67E-01	0.7291	4.94E-01	5.67	6.32E-03	25.41	4.28E-01	0.7609	4.24E-01	5.57	8.18E-04
SRGAN	×4	24.30	1.17E-10	0.6260	8.58E-08	**2.51**	1.00E+00	22.56	1.31E-08	0.6656	1.59E-08	**3.52**	1.00E+00
Ours (LR_input, w/o SR_IQA_Loss)	×4	27.38	4.66E-01	0.7278	4.64E-01	5.72	1.08E-03	25.36	3.92E-01	0.7596	3.91E-01	5.59	6.24E-04
Ours (LR_input)	×4	27.32	4.25E-01	0.7268	4.41E-01	5.60	3.89E-01	25.27	3.30E-01	0.7559	3.07E-01	5.52	1.24E-02
Ours (Bicubic_input, w/o SR_IQA_Loss)	×4	**27.43**	5.05E-01	**0.7297**	5.07E-01	5.58	7.34E-02	**25.51**	4.98E-01	**0.7641**	5.01E-01	5.51	1.27E-02
Ours (Bicubic_input)	×4	**27.42**	5.00E-01	0.7294	5.00E-01	**5.44**	5.00E-01	**25.51**	5.00E-01	**0.7640**	5.00E-01	**5.13**	5.00E-01

previous work HNSR [21] and retrained SRGAN [25] (training details see Section IV-F). Table VI and Table VII demonstrate the average PSNR, SSIM, and perceptual index (PI) performance on Set5, Set14, B100 and Urban100 datasets for magnification factors: ×2, ×3, and ×4. PI value is the combination of Ma *et al.* [28] and NIQE [52], namely $PI = ((10 - Ma) + NIQE)/2$ (lower is better). As can be seen from these two tables, our proposed SISR method with LR input can achieve competitive performance among various methods in terms of PI value and PSNR/SSIM values. Except for LapSRN, other compared SISR methods take bicubic interpolated version of LR image as input. Therefore, we also train our SISR method with bicubic input, which has an output convolutional layer with 7×7 kernel size instead of the transpose convolution. As can be seen from Table VI and Table VII, our proposed SISR method with bicubic input, which is trained without SR-IQA loss, achieves the best PSNR and SSIM values among compared methods most of the time, but the PI values are worse than IDN and DRRN in most cases. Although the PSNR and SSIM values are lower than the version which is only trained on L2-norm, our proposed method achieves better PI values when compared with state-of-the-arts in most of the cases. Although SRGAN [25] has very promising PI values, but PSNR/SSIM values are lower than every compared methods, even the bicubic interpolation.

In order to verify the significant improvement with respect to state-of-the-art methods, we conduct t-test between our proposed SISR method with bicubic input and each compared method to calculate the p-value for each PSNR, SSIM, and PI values. From p-values reported in Table VI and Table VII, we can find that the perceptual performance of our proposed SISR method has significant difference with compared methods in larger datasets most of the time.

In Fig. 6, we show the comparison of different methods' results on Urban100 for magnification ×4. As shown in this figure, our proposed method can generate more real results with more details. At this time, we have achieved our goal, which is the proposal of a method that has competitive PSNR/SSIM values and has the best perceptual quality than existing methods that only employs pixel-wise loss function as loss function. This is because our proposed method is obtained by optimizing the combined loss including the proposed novel SR-IQA loss. This is very different from the previous methods. We believe

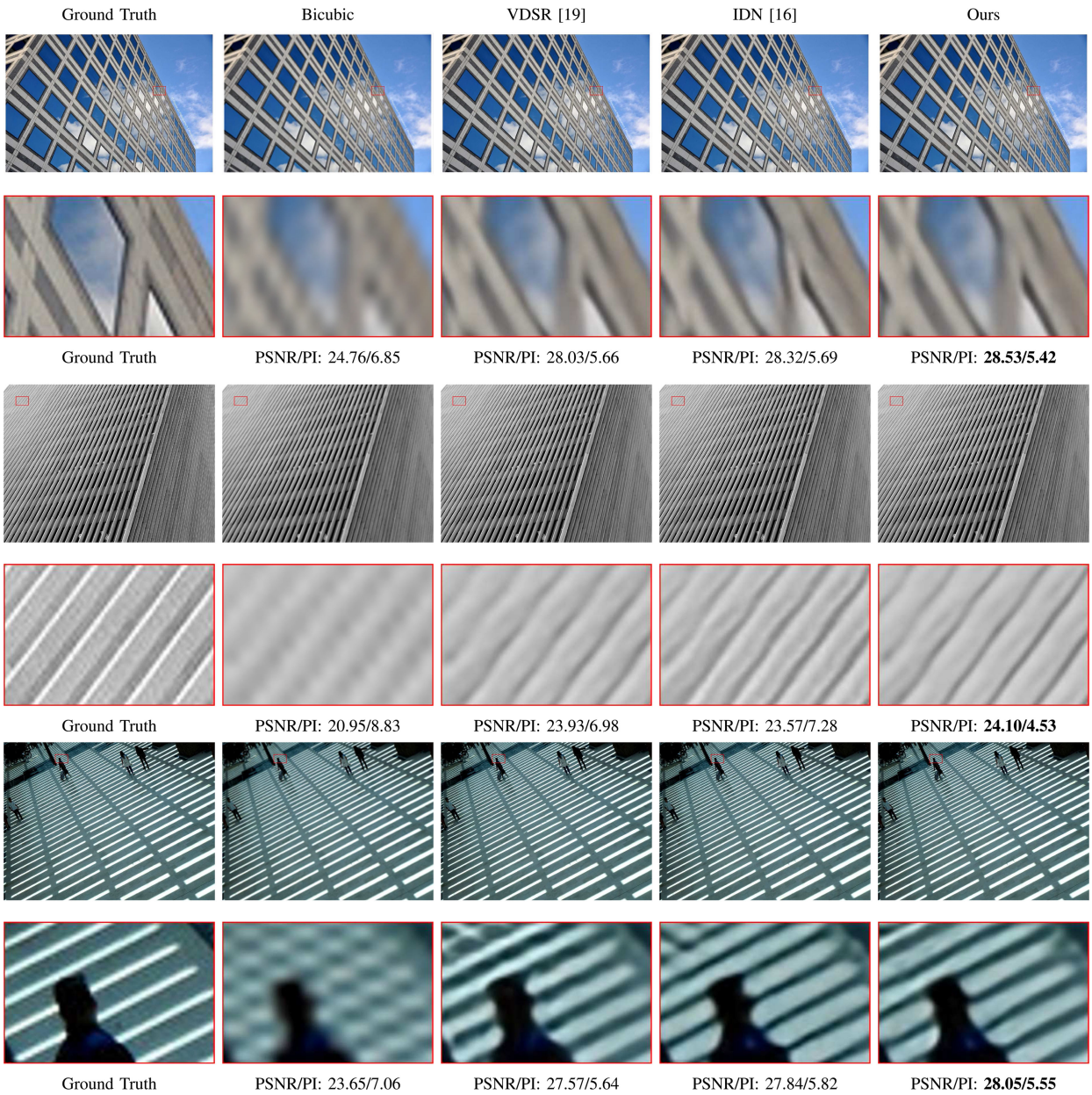


Fig. 6. Visual results for $\times 4$ on dataset "Urban100" [85]. We compare the visual results of our method with Bicubic, VDSR [18], and IDN [34]. In order to better compare the performance, we also put the enlarged patch to the bottom of each image, and put the PSNR and perceptual index (PI) values to the bottom of each enlarged patch.

that other SISR methods can also find a balance between PSNR value and perceptual quality by using our designed novel loss function. In next subsection, we explore whether our designed SR-IQA loss can help the GAN-based SISR method to further enhance the perceptual quality.

F. Application of Our Novel Loss Function to GAN-Based SISR Method

Following [25], we use a popular implementation from GitHub[1] to retrain GAN-based SISR on DIV2K Dataset [86]. For each training step, we randomly crop the 224×224 size patch from the images from the DIV2K dataset and down-sample cropped patch to 56×56 size to form LR-HR pair. At training, we firstly train the generative network with only L2-norm for 100 epochs, and then jointly train the generative network and discriminator network for 2000 epochs. In the jointly training step, generative networks optimized by L2-norm, perceptual loss (difference between the outputs of the conv5,4 layer of VGG19 networks), and adversarial loss. In order to verify the impact of our novel loss function to the GAN-based SISR method, we use the same training strategy to train the GAN-based SISR method with SR-IQA loss. Because our proposed FR-IQA model is trained on MatConvNet [84], we copy the

[1]https://github.com/tensorlayer/srgan

TABLE VIII
MEAN PI VALUE AND MA *et al.*'S SCORE COMPARISON BETWEEN GAN-BASED SISR WITH AND WITHOUT SR-IQA LOSS. BESIDES MEAN VALUES, WE ALSO DEMONSTRATE THE P-VALUE BETWEEN EACH PI AND MA *et al.*'S SCORE OF TWO METHODS

	GAN-based SISR				GAN-based SISR+SR-IQA loss			
	PI		Ma *et al.* [29]		PI		Ma *et al.* [29]	
	mean	p-value	mean	p-value	mean	p-value	mean	p-value
Set5	3.51	2.02E-01	7.78	1.66E-01	**2.95**	5.00E-01	**8.39**	5.00E-01
Set14	2.99	9.72E-02	7.88	2.50E-01	**2.58**	5.00E-01	**8.14**	5.00E-01
B100	2.51	1.05E-04	8.60	2.31E-02	**2.21**	5.00E-01	**8.75**	5.00E-01
Urban100	3.52	6.94E-03	6.77	4.14E-02	**3.26**	5.00E-01	**6.91**	5.00E-01

parameters of FR-IQA model to the TensorFlow module with same architecture. In the training, we add our designed SR-IQA loss as a part of loss function and set the weight value of SR-IQA loss to 1. The weight values of other losses are: 1e-3 for adversarial loss, 2e-6 for perceptual loss and 1 for pixel-wise L2-norm.

We compare the performance of the GAN-based SISR with and without SR-IQA loss on benchmark datasets and evaluate the perceptual quality with PI value and Ma *et al.*'s score [28]. Experimental results are listed in Table VIII. The values listed in this table are the mean value over each dataset. As we can observe from this table, our designed SR-IQA loss can help the GAN-based SISR method to improve the perceptual quality. We also conduct t-test between these two methods to verify the significant improvement of the GAN-based SISR with SR-IQA loss to the original one. P-values that are listed in Table VIII are lower than 0.05 in larger datasets, which indicate significant difference between compared two methods. Therefore, it can be concluded from the experimental results from Table VIII that our proposed SR-IQA loss can help the GAN-based SISR method to significantly improve the perceptual quality.

We also demonstrate the visual comparison results in Fig. 7. As demonstrated in this figure, with the help of our proposed SR-IQA loss, GAN-based SISR method can achieve better perceptual quality than the original one.

G. Discussion

Experimental results from previous subsections demonstrate that the combination of the proposed SR-IQA loss with L2-norm is able to significantly improve the perceptual quality while keeping the PSNR and SSIM values. This is benefited from our proposed deep FR-IQA method, which is a specific IQA method for accurately predicting the perceptual quality of SISR results. We believe that other CNNs-based image and video enhancement methods, such as denoising [87], inpainting [88], color dequantization [89], and quality enhancement of compressed video [90] can also employ our proposed SRIQA loss to further improve the perceptual quality while keeping the quantitative results do not drop too much. It is obvious that the development of a specific image quality assessment database for other image enhancement tasks can boost better IQA method than our proposed IQA method. Therefore, our proposed method can provide a solution for other image enhancement tasks to accurately improve the perceptual quality.

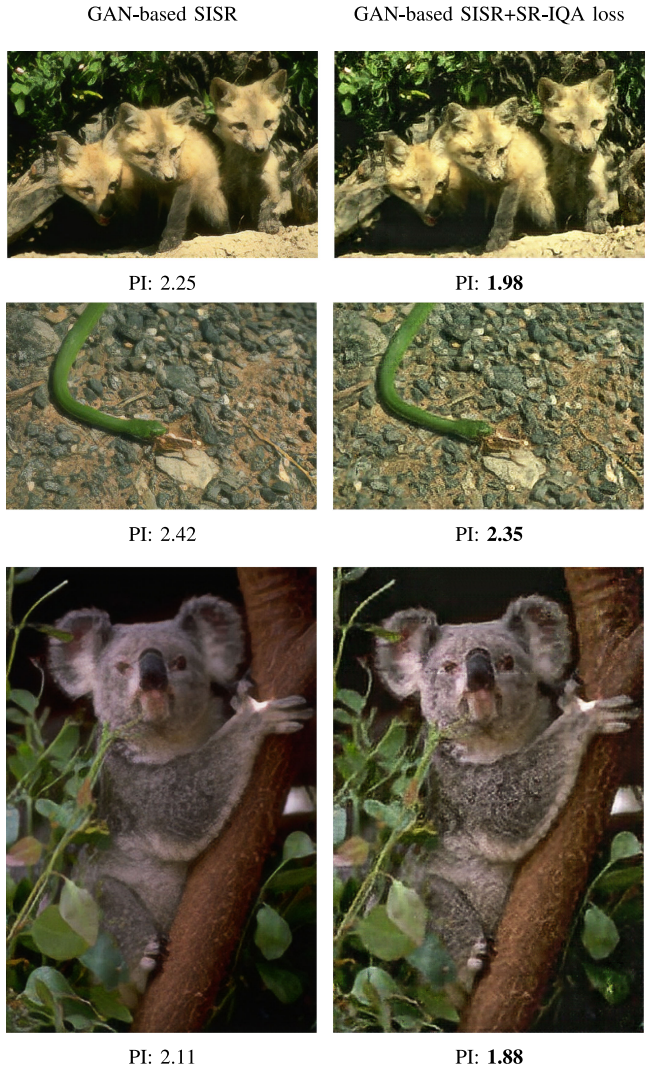


Fig. 7. Visual results comparison of GAN-based SISR with and without SR-IQA loss. Our proposed FR-IQA method can be a part of the loss function of GAN-based SISR method to further improve the perceptual quality.

V. CONCLUSION

In this paper, we propose an objective image quality assessment method guided single image super-resolution method. First, we train a full-reference image quality assessment method

for single image super-resolution methods. Our quality assessment method achieves state-of-the-art performance among various quality assessment methods. Second, we employ the proposed novel loss function to train our proposed SISR method. With the help of our proposed highway unit and the proposed novel loss function, our proposed SISR method surpasses previous methods both on PSNR value and perceptual quality in most of the cases. Finally, we apply our designed novel loss function to generative adversarial networks based super-resolution method and achieve better perceptual quality than the original one. This proves the generalization ability of our designed novel loss function and this loss function can be used in other super-resolution methods to improve the perceptual quality.

REFERENCES

[1] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[2] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, 1979.

[3] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 372–386.

[4] M.-C. Yang and Y.-C. F. Wang, "A self-learning approach to single image super-resolution," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 498–508, Apr. 2013.

[5] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.

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

[7] 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. Vision Conf.*, 2012, pp. 135.1–135.10.

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

[9] 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.

[10] 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.

[11] J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 1059–1066.

[12] Z. Xiong, D. Xu, X. Sun, and F. Wu, "Example-based super-resolution with soft information and decision," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1458–1465, Oct. 2013.

[13] J. Sun, L. Liang, F. Wen, and H.-Y. Shum, "Image vectorization using optimized gradient meshes," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 11.

[14] C. Wang, J. Zhu, Y. Guo, and W. Wang, "Video vectorization via tetrahedral remeshing," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1833–1844, Apr. 2017.

[15] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 334–355.

[16] Y. LeCun, B. Boser, J. Denker, and D. Henderson, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[17] 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.

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

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

[20] 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. Vision Pattern Recognit.*, 2016, pp. 1874–1883.

[21] K. Li, B. Bare, B. Yan, B. Feng, and C. Yao, "HNSR: Highway networks based deep convolutional neural networks model for single image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 1478–1482.

[22] 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.

[23] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 6228–6237.

[24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 694–711.

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

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015.

[27] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[28] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vision Image Understanding*, vol. 158, pp. 1–16, 2017.

[29] Y. Fang, C. Zhang, W. Yang, J. Liu, and Z. Guo, "Blind visual quality assessment for image super-resolution by convolutional neural network," *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29829–29846, Nov. 2018.

[30] B. Bare, K. Li, B. Yan, B. Feng, and C. Yao, "A deep learning based no-reference image quality assessment model for single-image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 1223–1227.

[31] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, vol. 1, no. 2, pp. 2790–2798.

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

[33] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate superresolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, vol. 2, no. 3, pp. 5835–5843.

[34] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 723–731.

[35] R. Timofte *et al.*, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, Jul. 2017, pp. 1110–1121.

[36] 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. Vision Pattern Recognit. Workshops*, 2017, vol. 1, no. 2, pp. 1132–1140.

[37] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 2472–2481.

[38] Y. Zhang *et al.*, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 286–301.

[39] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vision Workshops*, 2018, pp. 63–79.

[40] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[41] Y. Liu *et al.*, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.

[42] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[43] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Thirty-Seventh Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1398–1402.

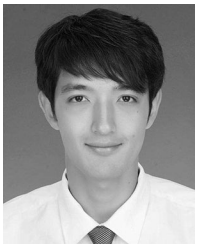
[44] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[45] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[46] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, "Image quality assessment using similar scene as reference," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 3–18.
[47] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1969–1977.
[48] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[49] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[50] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[51] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[52] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[53] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 1098–1105.
[54] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2014, pp. 1733–1740.
[55] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[56] Q. Wu *et al.*, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.
[57] Q. Wu *et al.*, "Blind image quality assessment based on rank-order regularized regression," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2490–2504, Nov. 2017.
[58] Q. Wu, H. Li, K. N. Ngan, and K. Ma, "Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2078–2089, Sep. 2018.
[59] G. Yue, C. Hou, K. Gu, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2722–2732, Oct. 2018.
[60] G. Yue, C. Hou, K. Gu, S. Mao, and W. Zhang, "Biologically inspired blind quality assessment of tone-mapped images," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2525–2536, Mar. 2018.
[61] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2499–2513, May 2018.
[62] A. R. Reibman, R. M. Bell, and S. Gray, "Quality assessment for super-resolution image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 2017–2020.
[63] T. Lukeš, K. Fliegel, and M. Klíma, "Performance evaluation of image quality metrics with respect to their use for super-resolution enhancement," in *Proc. Int. Workshop Qual. Multimedia Experience*, 2013, pp. 42–43.
[64] G. Wang *et al.*, "Perceptual evaluation of single-image super-resolution reconstruction," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3145–3149.
[65] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment for image super-resolution: A natural scene statistics approach," in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 1481–1484.
[66] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment of interpolated natural images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4651–4663, Nov. 2015.
[67] T. R. Goodall, I. Katsavounidis, Z. Li, A. Aaron, and A. C. Bovik, "Blind picture upscaling ratio prediction," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1801–1805, Dec. 2016.
[68] Y. Fang, J. Liu, Y. Zhang, W. Lin, and Z. Guo, "Quality assessment for image super-resolution based on energy change and texture variation," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2057–2061.
[69] R. Zhang, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 586–595.
[70] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.
[71] A. Rehman, M. Rostami, Z. Wang, D. Brunet, and E. R. Vrscay, "SSIM-inspired image restoration using sparse representation," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, 2012, Art. no. 16.
[72] Z. Wang, A. C. Bovik, and L. Lu, "Wavelet-based foveated image quality measurement for region of interest image coding," in *Proc. IEEE Int. Conf. Image Process.*, 2001, vol. 2, pp. 89–92.
[73] A. Koz and A. A. Alatan, "Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 326–337, Mar. 2008.
[74] I. G. Karybali and K. Berberidis, "Efficient spatial image watermarking via new perceptual masking and blind detection schemes," *IEEE Trans. Inf. Forensics Secur.*, vol. 1, no. 2, pp. 256–274, Jun. 2006.
[75] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain, "Incorporating image quality in multi-algorithm fingerprint verification," in *Proc. IEEE Int. Conf. Biometrics*, 2006, pp. 213–220.
[76] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.
[77] Z. Wang *et al.*, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.
[78] M. C. Farias, S. Mitra, M. Carli, and A. Neri, "A comparison between an objective quality measure and the mean annoyance values of watermarked videos," in *Proc. IEEE Int. Conf. Image Process.*, 2002, vol. 3, pp. III–III.
[79] A. Rehman and Z. Wang, "Reduced-reference SSIM estimation," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 289–292.
[80] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, arXiv:1207.0580.
[81] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vision*, Jul. 2001, vol. 2, pp. 416–423.
[82] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, "Fast image/video upsampling," *ACM Trans. Graph.*, vol. 27, no. 5, 2008, Art. no. 153.
[83] 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.
[84] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 689–692.
[85] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2015, pp. 5197–5206.
[86] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, Jul. 2017, pp. 1122–1131.
[87] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[88] C. Wang, H. Huang, X. Han, and J. Wang, "Video inpainting by jointly learning temporal structure and spatial details," in *Proc. 33th AAAI Conf. Artif. Intell.*, 2019.
[89] Y. Wang *et al.*, "Gif2video: Color dequantization and temporal interpolation of GIF images," 2019, arXiv:1901.02840.
[90] X. Meng *et al.*, "MGANET: A robust model for quality enhancement of compressed video," 2018, arXiv:1811.09150.

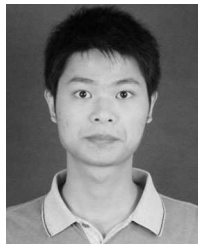
Bo Yan (SM'10) received the B.E. and M.E. degrees in communication engineering from Xi'an Jiaotong University, Xi'an, China, in 1998 and 2001 respectively, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2004. From 2004 to 2006, he was a Postdoctoral Guest Researcher with the National Institute of Standards and Technology, Gaithersburg, MD, USA. He is currently a Professor with the School of Computer Science, Fudan University, Shanghai, China. His research interests include video processing, computer vision, and multimedia communications.

Dr. Yan was the Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the Guest Editor of Special Issue on "Content-Aware Visual Systems: Analysis, Streaming and Retargeting" for the IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS.



Bahetiyaer Bare received the master's degree from the School of Computer Science, Fudan University, Shanghai, China, where he is currently working toward the Ph.D. degree. His research interests include digital image and video processing.

Chenxi Ma received the B.E. degree from the School of Computer Science, Shandong University, Shandong, China. She is currently working toward the Ph.D. degree with the School of Computer Science, Fudan University, Shanghai, China. Her research interests include digital image and video processing.



Ke Li received the B.E. and Ph.D. degrees from the School of Computer Science, Fudan University, Shanghai, China. His research interests include digital image and video processing.

Weimin Tan received the master's degree from the College of Communication Engineering, Chongqing University, Chongqing, China, and the Ph.D. degree from the School of Computer Science, Fudan University, Shanghai, China. He is currently a Postdoctoral Researcher with Fudan University. His research interests include digital image and video processing.