IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 22, 2020, accepted April 5, 2020, date of publication April 14, 2020, date of current version May 1, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2987911*

# Robust Prior-Based Single Image Super Resolution Under Multiple Gaussian Degradations

**WENYI WANG, (Member, IEEE), GUANGYANG WU, (Graduate Student Member, IEEE), WEITONG CAI, (Graduate Student Member, IEEE), LIAOYUAN ZENG, (Member, IEEE), AND JIANWEN CHEN, (Senior Member, IEEE)**

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Guangyang Wu (wuguangyang@std.uestc.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61701094.

**ABSTRACT** Although SISR (Single Image Super Resolution) problem can be effectively solved by deep learning based methods, the training phase often considers single degradation type such as bicubic interpolation or Gaussian blur with fixed variance. These priori hypotheses often fail and lead to reconstruction error in real scenario. In this paper, we propose an end-to-end CNN model RPSRMD to handle SR problem in multiple Gaussian degradations by extracting and using as side information a shared image prior that is consistent in different Gaussian degradations. The shared image prior is generated by an AED network RPGen with a rationally designed loss function that contains two parts: consistency loss and validity loss. These losses supervise the training of AED to guarantee that the image priors of one image with different Gaussian blurs to be very similar. Afterwards we carefully designed a SR network, which is termed as PResNet (Prior based Residual Network) in this paper, to efficiently use the image priors and generate high quality and robust SR images when unknown Gaussian blur is presented. When we applied variant Gaussian blurs to the low resolution images, the experiments prove that our proposed RPSRMD, which includes RPGen and PResNet as two core components, is superior to many state-of-the-art SR methods that were designed and trained to handle multi-degradation.

**INDEX TERMS** Single image super-resolution, convolution neural network, Gaussian blur, multiple degradations.

## I. INTRODUCTION

As one of the most active research topics, single image super-resolution (SISR) has attracted increasingly attention. SISR aims at estimating a high-resolution (HR) image from its degraded low-resolution (LR) observation. Currently, various deep learning based methods with different network architectures and training strategies have been proposed to improve the SR performance [1]–[7]. Since most of the existing SR methods assumed that the LR image is down-sampled by some pre-defined downsampler (e.g. bicubic interpolation with/without a known Gaussian blur), these priori hypotheses make the SR methods suffer from a common defect: their models were specialized for a single degradation and lack scalability to handle multiple degradations by a single model.

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

As was revealed in [8]–[10], learning-based SR methods easily suffer performance drop when the assumed degradation deviates from the real one (especially the mismatch of blur kernels). As shown in Fig. 1, the quality of SR images deteriorates due to the mismatch between the actual and the assumed degradations in LR images during the training phase. In this figure, we present the 4 times scale SR results by using EDSR [3]. Although EDSR assumed that the LR image was generated from HR image by bicubic interpolation with no other degradation, we used the EDSR to restore the LR images that were generated from the HR images by bicubic down-sampling along with different types of degradations. In the left column, the LR-HR relation obeys the EDSR assumption, therefore the SR result is visually plausible. In the middle and the right columns, the LR-HR relations do not obey the EDSR assumption, therefore the blurry results and over-sharped artifacts are observed in the middle and right columns

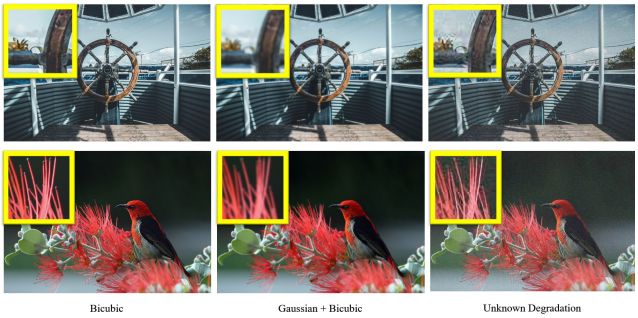



**FIGURE 1. SR results (X4) by using EDSR method [3]. We present the SR results of LR images that are generated from DIV2K [21] HR images by bicubic down-sampling along with different degradation methods. From the left to the right, the applied degradations are: none, Gaussian blur with kernel width 2.6, and unknown degradation.**

respectively. Given the aforementioned facts, the SR methods considered single degradation type are usually less effective in practical scenarios that the degradation types are hardly predictable. The SR problem with unknown degradations, also known as *blind SR*, presents great challenge to most deep learning based SR methods since multiple-degradation makes the LR-HR mapping relation becomes complicate [11], [12].

Efficient and powerful image priors are widely exploited in model-based optimization methods (e.g., the non-local self-similarity prior [13], [14], sparsity prior [15], denoiser prior [16], [17], etc.). However, these methods were usually not end-to-end and involved sophisticated optimization procedures. Even though the CNN denoiser prior was integrated with model-based optimization in [17] and it improved the efficiency to some extent, the hand-designed parameters made it less practical in real-world applications.

Given the correct degradation information as additional input, some blur-aware methods (e.g. CAB [18], SRMD [19], SFTMD [20]) could handle multiple degradations and achieved satisfactory performance. Therefore, the hybrid method that combined blur estimation and blur-aware SR could be a solution of blind SR problem. Due to the ill-posed property of blur estimation, the blur kernel estimated from one degraded LR image is hardly identical to the real one. Therefore, the SR methods that rely on accurate blur estimation are very likely to be unstable in real cases [20].

Given the facts above, this paper focuses on using one end-to-end CNN to robustly solve SR problem with unknown Gaussian degradations. Examining the existing SR methods that offered optimal working blur kernel, a sharper kernel often leads to the ringing effects and a smoother kernel often leads to blurry output images. These phenomena empirically inspires us that the SR model needs robust image prior as an anchor to stabilize the SR results in different Gaussian blurs. As shown in Fig. 2, the second row presents our generated image priors which are highly consistent to different Gaussian blur kernels without losing the primary contents in original LR image. It's noted that the robustness of prior here means that different degradations applied on one image only introduce subtle changes on output prior. In practical, we propose an innovative training strategy to train an auto-encoder-decoder, named RPGen (Robust Prior Generator), for generating robust priors.

Instead of estimating the kernel maps as SR prior and then super-resolve the LR image like SRMD did [19], we integrate the image content prior into a CNN based SR model due to the following considerations. First, the kernel maps do not actually contain the image information. Manipulating the blur kernel and the LR image at the same time with convolutional operation introduces interference that is not related to the image [20]. Oppositely, image prior can act as feature maps and naturally integrate with CNNs. Secondly, the degraded texture in LR image depends on blur kernel and original HR texture at the same time. Therefore, region-aware processing, which can be offered by an image-content prior, is preferable. For example, Gaussian blur kernel is a low-pass filter, the width of Gaussian blur kernel does not significantly affect the output of degraded image in flat region. In conclusion, a robust image prior can be more efficient and robust compared with an estimated blur kernel, since it carries both degradation and image content information in together. In order to make effective use of the robust prior, we carefully designed a residual-structure based SR network PResNet. The complete framework named RPSRMD (robust prior based super resolution for multiple degradation) combines the RPGen and PResNet, and achieves the state-of-the-art (SOTA) performance on SR problem under multiple Gaussian degradations.

In the view of above, the main contributions of this paper are summarized as follows:

1: We propose an effective end-to-end deep learning framework, referred to as RPSRMD, to robustly deal with SR problem in multiple Gaussian blur cases. The proposed method extends the widely-used bicubic degradation assumption to bicubic + unknown Gaussian blur. This improvement makes the SR solution more practical in real applications. The proposed framework contains a prior generator (i.e., RPGen) that generates the robust image priors and a SR network (i.e., PResNet) that integrates the prior with LR image to reconstruct visually plausible HR images.
2: In order to generate shared primary image content in LR images with different Gaussian blur strength, we designed a new loss function to control the output consistency of the proposed RPGen. The trained network is able to produce stable image priors as illustrated in Fig. 2. Even though the training strategy is proposed for SR problem, it can be a general approach and can be extended to other tasks such as deblurring and denoising in multiple degradations.

## II. RELATED WORKS

Recently, CNN based models applied on SR problem made efficient use of large dataset to learn the mapping from the synthetic LR image to the ground truth HR images in end-to-end manner. Plenty of CNN models have been proposed

**FIGURE 2.** Illustration of robust priors. The first row presents LR images degraded from one HR image by different blur kernels, the second row presents the corresponding priors. The SSIM value between the three priors are at least 0.99 which indicates the expected robustness to different blur kernels.

to enhance the performance and efficiency. SRCNN [22] first trained a shallow CNN to solve SR problem and it significantly outperformed the traditional methods at that time. In order to go deeper in network structure and achieve better performance, VDSR [23] was proposed using skip connection learning strategy. Moreover, they also showed that VDSR could handle multiple SR scales with single CNN model. Although it achieved good performance, VDSR suffered from high computational budgets because the input images were interpolated to the same spatial size as the output images via bicubic method. To improve the computational efficiency, deconvolution layer [24] and pixel-shuffling layer [25] were introduced in SR network to upscale the image at the end of the network. LapSRN [26] further took the LR image as input and progressively predicted the sub-band residuals with a coarse-to-fine strategy. SRResNet [27] first introduced residual blocks in SR networks. EDSR [3] then improved it by removing the BN (batch normalization) layers and applied residual scaling layer in residual blocks to accelerate the training procedure and expand the model size. SRGAN [27] and ESRGAN [28] applied GAN methods to improve the perceptual quality of the reconstructed images. Some other works focused on special scenarios, e.g., SFT layer was introduced in [29] to introduce the semantic prior as additional input of SR network. SRMD [19] proposed a stretching strategy to integrate the degradation information in the SR network and achieved visually plausible results in reconstructing real LR image.

## III. PROPOSED METHOD

### A. PROBLEM FORMULATION

In SISR problem, the LR image $I^{LR}$ generated from the HR image $I^{HR}$ can be formulated by a degradation model as follows:

$$I^{LR} = (k \otimes I^{HR}) \downarrow_s + n, \tag{1}$$

where $\otimes$ denotes the convolution operation, $k$ is the blur kernel, $\downarrow_s$ represents the downsampler and $n$ is the additive noise. The main problem we focus on in this paper is to restore the down-sampled blurry image without additive noise. Similar as many SR methods considered LR degradations [8], [19], [20], we adopt the combination of Gaussian blur and bicubic downsampling as our LR generation model:

$$I^{LR} = (g \otimes I^{HR}) \downarrow_{bs}, \tag{2}$$

where $\downarrow_{bs}$ is the bicubic downsampler and $g$ is the Gaussian blur kernel:

$$g = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \tag{3}$$

and $\sigma$ is also referred to as Gaussian blur kernel width or blur strength in this paper.

### B. OVERALL FRAMEWORK

As shown in Fig. 3, the proposed RPSRMD framework consists of two CNN (convolutional neural network) models: a prior estimator $P$ (RPGen in Fig. 3) and a SR model $F$ (PResNet in Fig. 3). We suppose the LR image $I^{LR}$ has the shape of $H \times W \times C$, where $C$ denotes the number of color channels, $H$ and $W$ denote the height and weight of the image.

The prior estimator $P$ aims at providing robust image prior with the given degraded LR image by the function:

$$prior = P_{\theta_p}(I^{LR}), \tag{4}$$

where $prior$ is a feature map with the shape of $H \times W \times 1$, and $\theta_p$ represents the trained CNN parameters of estimator $P$.

The SR model $F$ integrates the LR image and its corresponding prior together as the input to reconstruct the high quality HR image. It can be formulated as:

$$I^{SR} = F_{\theta_f}(I^{LR}, prior), \tag{5}$$

where $\theta_f$ represents the trained CNN parameters of $F$. Model $F$ upscales the input $I^{LR}$ at the output instead of applying the upscale at the input. Therefore, the SR model $F$ can apply convolution on feature maps in small spatial size.

The entire framework is a FCN (full convolutional neural network) so as to exploit the advanced merits of CNN, such as the fast speed by parallel computing and high accuracy through recursive learning. It's worth noting that even though we train two parts of the framework (i.e. $P$ and $F$) individually, the whole structure is still an end-to-end CNN model in the test phase.

### C. RPGEN: HOW TO GENERATE IMAGE PRIOR

#### 1) RPGen STRUCTURE

As shown in Fig.4, the RPGen network is a shallow CNN model to preserve the primary contents in the image and eliminate the randomly generated corruptions. It consists of several symmetric downsample and upsample modules with detailed information in the right bottom part of Fig. 3. The downsample modules act as the feature extractor and encode the primary contents in the degraded LR images. The upsample modules decode the image abstraction to restore image details. Moreover, skip connections from a downsample module to its corresponding upsample module pass the feature

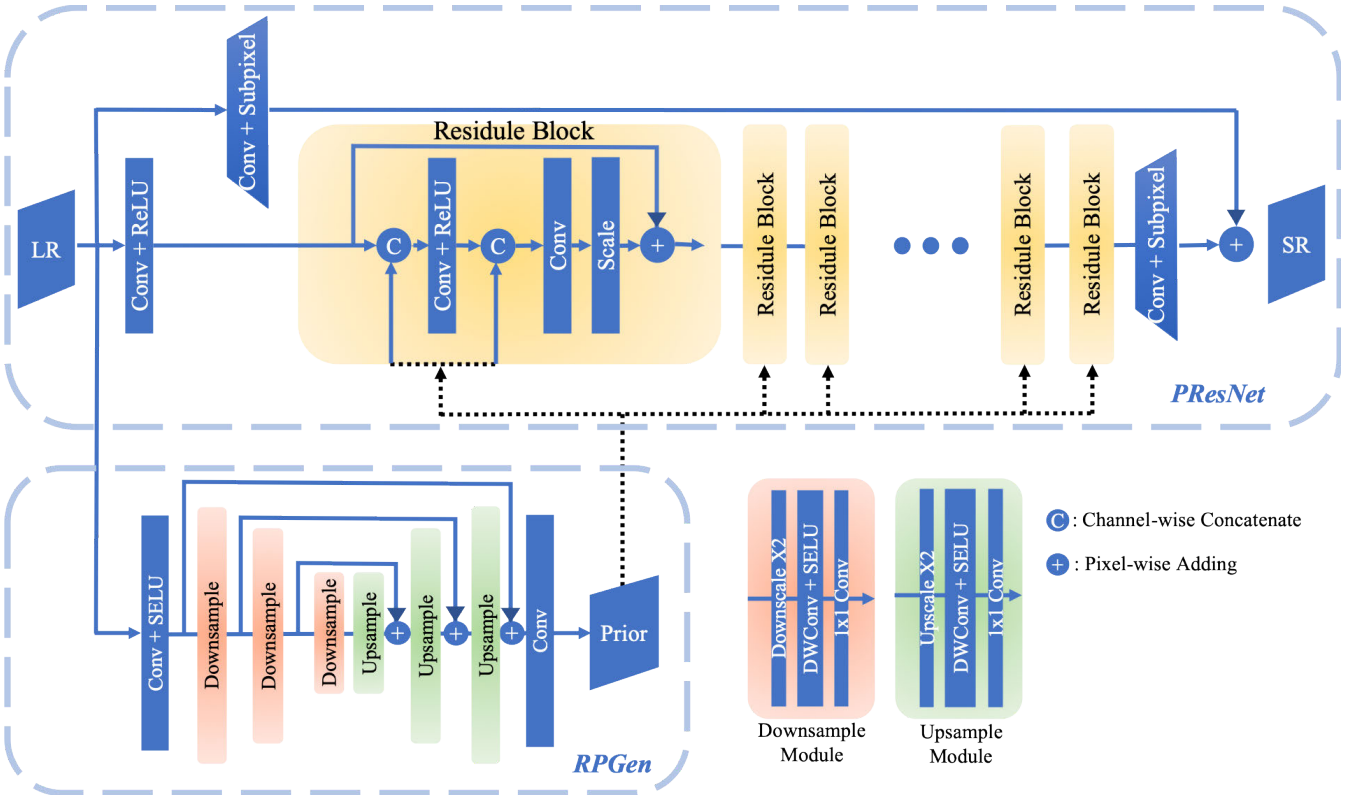



**FIGURE 3.** The structure of the RPSRMD framework. The whole framework consists of two CNN networks, the PResNet and the RPGen model. The feature maps of the priors are plugged into residual blocks, concatenated in channel-wise with output features of each layer in the residual block. The skip connection upscale the inputs and add with the outputs of the last residual block in pixel-wise.

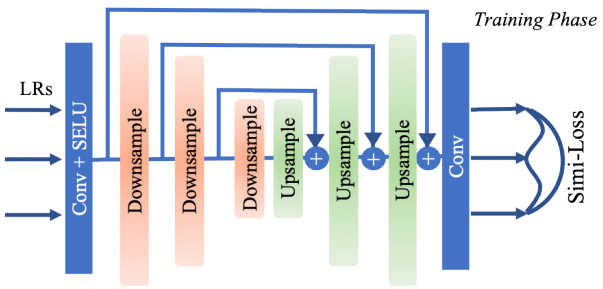



**FIGURE 4.** Illustration of the simi-loss of the RPGen. We minimize the consistency loss of the priors extracted from different LR images which are generated by different degradations from the single HR image.

maps through and add them element-wise. The output has same spatial size (i.e., width and height) as the input image.

In SR processing, we prefer using neither pooling nor unpooling in the network since pooling usually discards useful texture information which could be essential for restoring details [30]. Thus we use pixelshuffle operation to down/upscale the spatial size of the feature maps without information loss. Given the scale factor $r$, the upsample layer $L^u$ reshapes a $H \times W \times C$ feature matrix to a $rH \times rW \times C/r^2$ feature matrix and the downsample layer $L^d$ reshapes a $H \times W \times C$ feature matrix to a $H/r \times W/r \times r^2C$ feature matrix as introduced in [25]

Each down/upsample module consists of a down/upscale layer with scale factor $r = 2$, a convolution layer with the activation function RELU. As shown in Fig.4, the symmetric skip connections pass the feature maps of downscale module to the upscale module thus the network provides low-level features for better restoration and help the back propagation of gradient [30].

### 2) LOSS FUNCTION

In order to enhance the robustness of image priors generated by RPGen, we introduce an innovative loss function to guarantee that the image priors of one image with different Gaussian blurs to be very similar. In practical, we generate $n$ LR images ($I_k^{LR}, k = 0, 1, 2, \ldots, n$) by down sampling one HR image with $n$ different Gaussian blur kernels. We further assume that the $n$ LR images share similar image prior estimations from our proposed RPGen without losing primary image contents. Accordingly, we introduce a novel loss function to exploit the required properties of the network, which consists of two parts:
consistency loss:

$$L_{simi} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} mse(prior^i, prior^j), \tag{6}$$

and validity loss:

$$L_{mmse} = \frac{1}{n} \sum_{i=1}^{n} mse(prior^i, I^{anchor}), \tag{7}$$

where $prior^k, k \in [1, 2, \ldots, n]$ is the $k$th output of RPGen $P_{\theta_p}(I_k^{LR})$, and $mse$ denotes the mean-squared-error. $I^{anchor}$ is the LR image that is bicubic down-sampled from the HR image. The final cost function is a linear combination of this two parts by hyper-parameter $\alpha$:

$$L_{rp} = L_{simi} + \alpha L_{mmse}. \tag{8}$$

In a network with multiple inputs, $L_{simi}$ enforces the outputs to be pixel-wise similar. However, the network often shrinks the value range of the output to reduce the loss energy if only $L_{simi}$ is presented in the loss function. We therefore add a regularization term $L_{mmse}$ to address the above issue. The anchor $I^{anchor}$ plays an important role here, it guides the texture and structure restoration and it avoids the value shrinkage problem. We empirically set $I^{anchor}$ as the Y channel (in YCbCr color space) of LR image bicubic down-sampled from the HR image.

It is worth noticing that there is a trade-off between $L_{simi}$ and $L_{mmse}$. Specifically, the consistency of output gets worse with large $\alpha$ while the RPGen output tends to be blurry with very small $\alpha$. In order to keep the balance between consistency and clarity, we set $n = 2$ and $\alpha = 2$ in the training phase. In specific, we optimize:

$$\begin{aligned}\min_{\theta_p} \mathbb{E}\{&[P_{\theta_p}(I_0^{LR}) - P_{\theta_p}(I_1^{LR})]^2 \\ &+ [P_{\theta_p}(I_0^{LR}) - I^{anchor}]^2 \\ &+ [P_{\theta_p}(I_1^{LR}) - I^{anchor}]^2\},\end{aligned} \tag{9}$$

where $I_0^{LR}$ and $I_1^{LR}$ are two LR images blurred by two different random Gaussian kernels.

### D. PResNet: HOW TO USE IMAGE PRIOR

PResNet is essentially a SR network to reconstruct high quality images with the degraded LR image and its corresponding content prior. We follow the trends of using residual structure based network for its capability of generalization. There are alternative ways to introduce priors to a SR network. For instance, one can concatenate priors with the input LR images as a joint input to the network [19]. However, directly concatenating them together only applies the prior information in the first layer, and the deep layers are difficult to be affected by prior information [20]. Hence, we feed the priors into each residual blocks to employ the prior information efficiently and exploit the deep residual structure to improve performance. In practice, we concatenate the prior and feature maps in each residual block before convolution, as shown in Fig.5. In the red boxes of Fig.5, the dimension of the feature maps are denoted as $W \times H \times C$, where $W$ is the feature width, $H$ is the feature height and $C$ is the number of feature channels.

Specifically, we pass the input LR image (i.e. the RGB channels) into two branches in parallel, one is fed to a pre-trained RPGen to obtain the robust image prior, and the other is fed to the PResNet. At first, the original LR image with dimension $W \times H \times 3$ is expanded to feature maps with

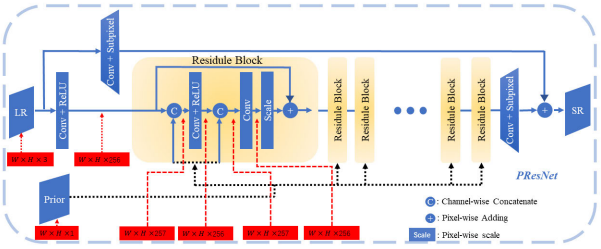



**FIGURE 5.** The proposed PResNet. The numbers in the red boxes indicate the feature dimension at each step.

dimension $W \times H \times 256$, which is the size of feature maps fed into every residual block. In each residual block, the input feature maps are first concatenated with the image content prior channel-wise (i.e. in $C$ dimension) and then are convolved and rectified. The corresponding feature maps sizes are presented in the red boxes of Fig.5. In each residual block, we apply twice concatenating+convolution operations to add into PResNet the information of image prior. Moreover, to avoid the numerical unstable in training procedure, we adopt the residual scaling with factor 0.1 [3] in each residual block. Overall, we stack 32 residual blocks, each of which takes the output of the previous module and the robust prior as input. In the up-sampling and the skip connection part, we use pixel-shuffling layer [25] to upsample the output of the last residual block and the LR image respectively and add them pixel-wise to obtain the final HR image. With the pre-trained RPGen network, the loss function of PResNet can be formulated as:

$$\min_{\Theta_f} \mathbb{E}\{I^{HR} - F_{\theta_f}[I^{LR}, P_{\theta_p}(I^{LR})]\}^2, \tag{10}$$

where $P_{\theta_p}$ is the pre-trained RPGen estimating image content prior, $F_{\theta_f}$ is the PResNet trained to generate SR image with the LR image $I^{LR}$ and its corresponding prior $P_{\theta_p}(I^{LR})$.

### E. WHY NOT LEARN NETWORK DIRECTLY

Compared with our proposed method, one similar but straightforward method is to train a neural network with the dataset which is augmented to cover multiple degradations. However, direct training often leads to the following defects: 1) It tends to over-fitting or under-fitting some degradation types so that the network lacks robustness; 2) In order to generalize the ability to reconstruct all degraded images, the network has pool average reconstruction quality. The above problems lead to the lack of practicality in real applications. Moreover, we will show in Sec. IV that our proposed method can achieve better performance both on robustness and reconstruction quality than direct training.

## IV. EXPERIMENTS

### A. DATA PREPARATION

We combined DIV2K training set [21] and Flicker2K dataset [3] to obtain 3450 high quality HR training images and trained the RPGen and the PResNet in our proposed RPSRMD framework. We randomly cropped each training

image into 100 patches with the size of $48S \times 48S$ where $S$ was the scale factor. The pixel values were first normalized within [0,1] and the mean value of each channel was subtracted. In order to present multiple Gaussian degradations in our training dataset, we defined the degradation domain $\mathbb{D} = \{g_{\sigma_i}(\cdot)\downarrow_{bs} | \sigma_i \in [0.2, 0.3, \ldots, S]\}$, where $g_{\sigma_i}(\cdot)$ was the Gaussian blur in Eq. 3 with variance $\sigma_i$ and $\downarrow_{bs}$ represented the bicubic downsampling. In RPGen's training phase, according to Eq. (9), we generated two LR images as joint inputs from one HR image by randomly applying two degradation types in $\mathbb{D}$, and $I^{anchor}$ was the Y channel (in YCbCr color space) of LR image bicubic down-sampled from the HR image. In PResNet's training phase, according to Eq. (10), $I_{LR}$ was degraded by randomly choosing the degradation in set $\mathbb{D}$. All models were trained with Adam [31] as optimizer, the initial learning rate was set to be $lr^0 = 0.0001$, and gradually decayed by:

$$lr^k = max(\frac{lr^0}{2}(cos(\frac{k\pi}{N}) + 1), 10^{-5}) \tag{11}$$

where $lr^k$ is the learning rate in the $kth$ epoch and $N$ is the max epoch number. We set $N = 300$ in RPGen's training and $N = 600$ in PResNet's training respectively.

### B. ROBUSTNESS OF PRIORS

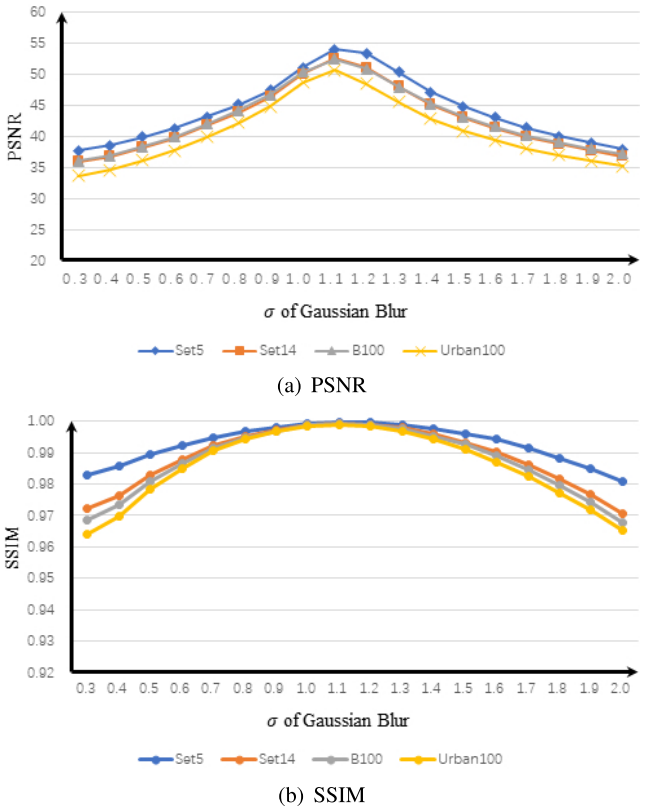
In Sec. III-C, we introduced the underlying assumptions of shared content in SR and the corresponding training strategies of the proposed prior generator $P_{\theta_p}$. In this part, we will show by simulation experiments that our trained prior generator $P_{\theta_p}$ can generate robust image priors from LR images derived from different Gaussian blur kernels.

Given each HR image in the widely used datasets (i.e. Set5, Set14, BSD100, and Urban100), we generated multiple LR images $I^{LR}_{\sigma_i}$ by using Eq. (2) and Eq. (3) with $\sigma_i \in \{0.3, 0.4, \ldots, 2.0\}$. Afterwards, the image priors $I^{prior}_{\sigma_i}$ of each LR image $I^{LR}_{\sigma_i}$ was generated from the trained $P_{\theta_p}$ and the prior similarity in one HR image is evaluated by calculating PSNR and SSIM between $I^{prior}_{\sigma_i}$ and

$$I^{prior}_{mean} = \frac{1}{M}\sum_{i=1}^{M} I^{prior}_{\sigma_i}, \tag{12}$$

where $M = 18$ is the number of different Gaussian $\sigma$ we applied here.

In Fig. 6, we present the PSNR and SSIM values of the image priors derived from dataset Set5, Set14, BSD100, and Urban100. The x axis represents the Gaussian $\sigma$ of the blur kernel applied in LR image generation, the y axis represents the PSNR/SSIM value of the image prior corresponding to a specified Gaussian kernel versus the average image prior. It is clear that the image priors shared very similar contents and textures even if their corresponding LR images were blurred very differently (i.e $\sigma$ varies from 0.3 to 2.0). Therefore, this prior can be a very useful and robust side information in SR restoration, especially when the LR images are blurred by variant or unknown Gaussian kernels.




(a) PSNR

(b) SSIM

**FIGURE 6. Prior robustness. The PSNR and SSIM values of the image priors in different Gaussian blurs versus the average image prior.**

### C. SR QUALITY COMPARISON

We evaluate and compare the performance of our proposed method based on the widely used datasets: Set5, Set14, BSD100, and Urban100. First we compare the SR results between EDSR [3], SRMD [19] and the proposed RPSRMD when the LR images are blurred by varying Gaussian $\sigma$.

In Fig. 7, the x axis represents the Gaussian $\sigma$ of the blur kernel applied in LR image generation, the y axis represents the PSNR value of the reconstructed SR image versus original HR image in the dataset. Specifically, Fig. 7 (a) compares the PSNR performance in scale 2 on Set5, and Fig. 7 (b) compares the PSNR performance in scale 4 on Set14. It is noted that the performance of EDSR drops fast as the Gaussian $\sigma$ increases because the network is trained with single type of degradation and it is fragile when kernel mismatch happens. The SRMD takes the complete kernel information and LR image as joint input, therefore it should be more robust when the strength of Gaussian blur changes. According to Fig. 7, SRMD is indeed more robust than EDSR as we expected, its performance does not significantly drop when Gaussian $\sigma$ increases. Different from SRMD that needs accurate information of the blur kernel, our proposed RPSRMD only takes the LR images as inputs and its robustness to Gaussian $\sigma$ changing outperforms SRMD by means of PSNR. Therefore, this experiment shows that our proposed method can stably deal with the SR problem when the LR image suffers varying Gaussian blur and the blur strength is unknown.

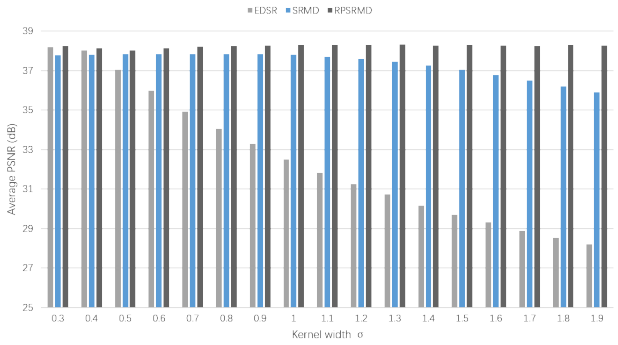
**TABLE 1.** The PSNR (dB) and SSIM of multi-degradation SR methods (x2) with different kernel width ($\sigma$).

| | sigma | Bicubic | EDSR [3] | SRMD [19] | EDSR-MD | RPSRMD |
|---|---|---|---|---|---|---|
| | | x2(PSNR/SSIM) | | | | |
| Set5 | 0.3 | 33.69/0.9305 | 38.19/0.9611 | 37.79/0.9601 | 38.09/0.9620 | 38.25/0.9630 |
| Set14 | | 30.37/0.8700 | 33.95/0.9200 | 33.33/0.9154 | 34.31/0.9180 | 34.18/0.9230 |
| BSD100 | | 29.53/0.8436 | 32.35/0.9018 | 32.05/0.8985 | 32.17/0.9000 | 32.40/0.9030 |
| Urban100 | | 26.85/0.8407 | 32.98/0.9359 | 31.33/0.9204 | 31.92/0.9290 | 33.07/0.9370 |
| Set5 | 1.2 | 30.37/0.8739 | 31.25/0.8941 | 37.60/0.9574 | 37.97/0.9580 | 38.31/0.9620 |
| Set14 | | 27.73/0.7842 | 28.30/0.8141 | 33.29/0.9124 | 34.08/0.9110 | 34.33/0.9230 |
| BSD100 | | 27.31/0.7488 | 26.92/0.7810 | 32.06/0.8962 | 32.30/0.9000 | 32.56/0.9030 |
| Urban100 | | 24.58/0.7466 | 25.08/0.7845 | 31.01/0.9146 | 31.37/0.9190 | 32.84/0.9350 |
| Set5 | 1.8 | 28.28/0.8187 | 28.53/0.8271 | 36.21/0.9436 | 38.05/0.9580 | 38.30/0.9620 |
| Set14 | | 26.13/0.7157 | 26.27/0.7263 | 32.27/0.8877 | 34.19/0.9110 | 34.24/0.9220 |
| BSD100 | | 26.01/0.6799 | 25.82/0.6905 | 31.07/0.8659 | 32.36/0.9020 | 32.58/0.9040 |
| Urban100 | | 23.21/0.6728 | 23.34/0.6859 | 29.79/0.8898 | 31.28/0.9180 | 32.54/0.9320 |

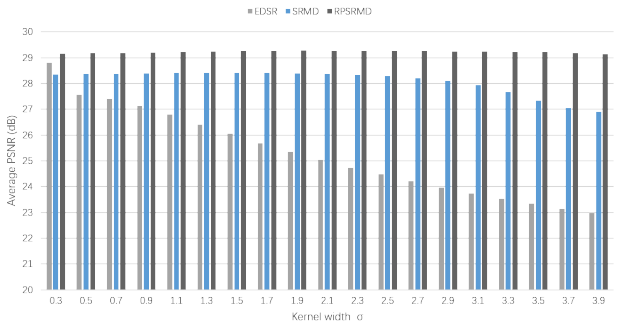
**TABLE 2.** The PSNR (dB) and SSIM of multi-degradation SR methods (x4) with different kernel width ($\sigma$).

| | sigma | Bicubic | EDSR [3] | SRMD [19] | EDSR-MD | RPSRMD |
|---|---|---|---|---|---|---|
| | | x4(PSNR/SSIM) | | | | |
| Set5 | 0.3 | 28.45/0.8112 | 32.48/0.8986 | 31.96/0.8924 | 32.24/0.9000 | 32.94/0.9080 |
| Set14 | | 26.09/0.7050 | 28.82/0.7874 | 28.35/0.7771 | 28.65/0.7830 | 29.17/0.7950 |
| BSD100 | | 25.87/0.6682 | 27.72/0.7417 | 27.49/0.7337 | 27.57/0.7380 | 28.02/0.7520 |
| Urban100 | | 23.11/0.6582 | 26.65/0.8035 | 25.68/0.7731 | 25.73/0.7900 | 27.61/0.8260 |
| Set5 | 1.2 | 27.73/0.7907 | 29.96/0.8633 | 31.99/0.8923 | 32.24/0.8990 | 33.05/0.9090 |
| Set14 | | 25.60/0.6824 | 26.63/0.7505 | 28.41/0.7772 | 28.64/0.7790 | 29.23/0.7950 |
| BSD100 | | 25.50/0.6465 | 26.11/0.7057 | 27.53/0.7336 | 27.61/0.7360 | 28.09/0.7520 |
| Urban100 | | 22.70/0.6345 | 23.55/0.7394 | 25.70/0.7720 | 25.62/0.7800 | 27.66/0.8260 |
| Set5 | 1.8 | 26.91/0.7657 | 28.01/0.8153 | 32.01/0.8915 | 32.25/0.9000 | 33.12/0.9090 |
| Set14 | | 25.04/0.6572 | 25.53/0.7020 | 28.41/0.7761 | 28.68/0.7800 | 29.27/0.7960 |
| BSD100 | | 25.07/0.6233 | 25.44/0.6618 | 27.54/0.7330 | 27.65/0.7370 | 28.13/0.7530 |
| Urban100 | | 22.24/0.6082 | 22.59/0.6676 | 25.66/0.7695 | 25.60/0.7790 | 27.63/0.8250 |
| Set5 | 2.6 | 25.74/0.7256 | 26.11/0.7514 | 31.77/0.8849 | 32.22/0.8990 | 33.11/0.9100 |
| Set14 | | 24.20/0.6209 | 24.33/0.6427 | 28.25/0.7677 | 28.70/0.7800 | 29.26/0.7960 |
| BSD100 | | 24.44/0.5908 | 24.59/0.6095 | 27.43/0.7245 | 27.67/0.7380 | 28.14/0.7540 |
| Urban100 | | 21.57/0.5704 | 21.63/0.5965 | 25.46/0.7596 | 25.53/0.7770 | 27.53/0.8230 |
| Set5 | 3.7 | 24.30/0.6728 | 24.40/0.6822 | 29.95/0.8474 | 32.14/0.8980 | 32.94/0.9080 |
| Set14 | | 23.15/0.5774 | 23.15/0.5845 | 27.04/0.7179 | 28.63/0.7780 | 29.17/0.7940 |
| BSD100 | | 23.64/0.5540 | 23.70/0.5597 | 26.51/0.6751 | 27.65/0.7370 | 28.08/0.7530 |
| Urban100 | | 20.73/0.5269 | 20.71/0.5350 | 24.18/0.6994 | 25.38/0.7710 | 27.23/0.8170 |

We then conduct thorough comparisons between our proposed RPSRMD and other SOTA methods in different Gaussian blur kernels. Table 1 and Table 2 compare the PSNR of SR images on four widely-used datasets (i.e., Set5, Set14, BSD100 and Urban100) with scale factor 2 and 4. The original EDSR (denoted as EDSR) suffers severe performance drop when the downsampling kernel is different to the predefined one. In order to make fair comparison, we also trained the EDSR model (denoted as EDSR-MD) with our PResNet training set which contains multiple degradation types. The SR results from EDSR-MD is still not as good as our proposed RPSRMD. It proves that the robust prior can enhance the SR results when multiple Gaussian blur types are presented in the training set. Although SRMD takes the complete kernel information as input, our proposed RPSRMD requires less blur information and generates better quality SR images than SRMD. It should be noted that this does not mean comprehensive information leads to negative effect. Actually, acquiring complete blur information would be always preferable but difficult. Therefore a SR method that




(a) set5 x2




(b) set14 x4

**FIGURE 7.** The PSNR values of the reconstructed SR images from LR images that suffer different Gaussian blur. We test the scale factor 2 on Set5 and scale factor 4 on Set14 using EDSR, SRMD and the proposed RPSRMD.

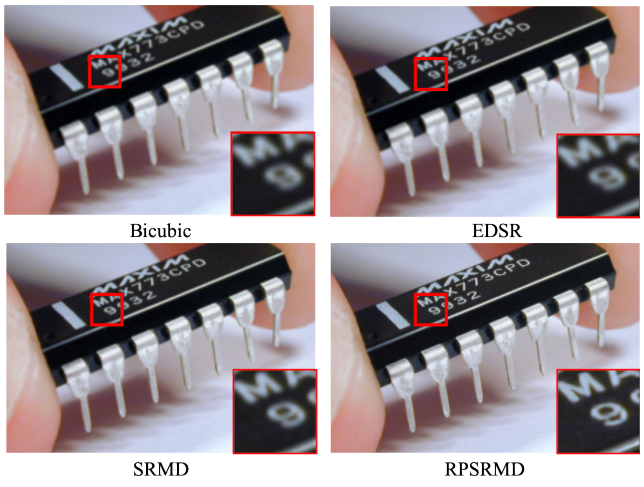



**FIGURE 8.** Visual comparison between SR methods on real LR image 'chips'.

requires less blur information would be useful and practical. In our method, the side information "image prior" is trained by our designed network and the cascade SR network is jointly designed. Therefore, our proposed SR method can make good use of the "image prior" and generated high quality SR images.

Furthermore, we compare the visual quality of different methods on real and synthetic images in Fig.8 - 10.

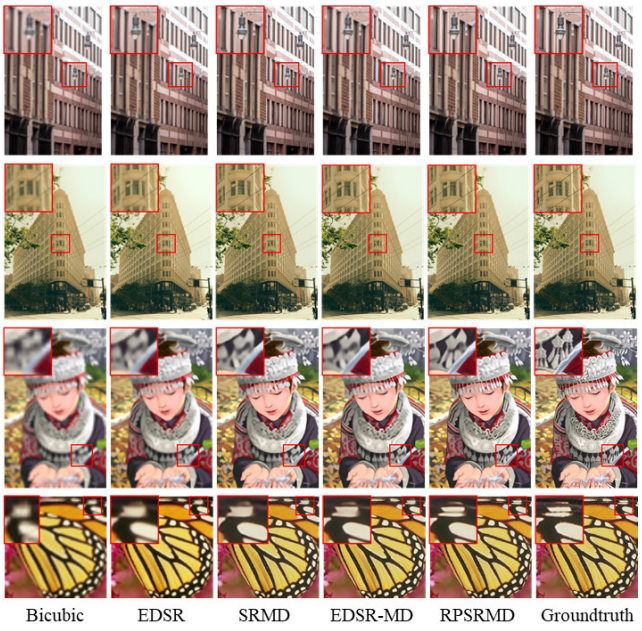



**FIGURE 9.** Visual comparison between SR methods on synthetic LR images with Gaussian blur strength $\sigma = 1.2$.

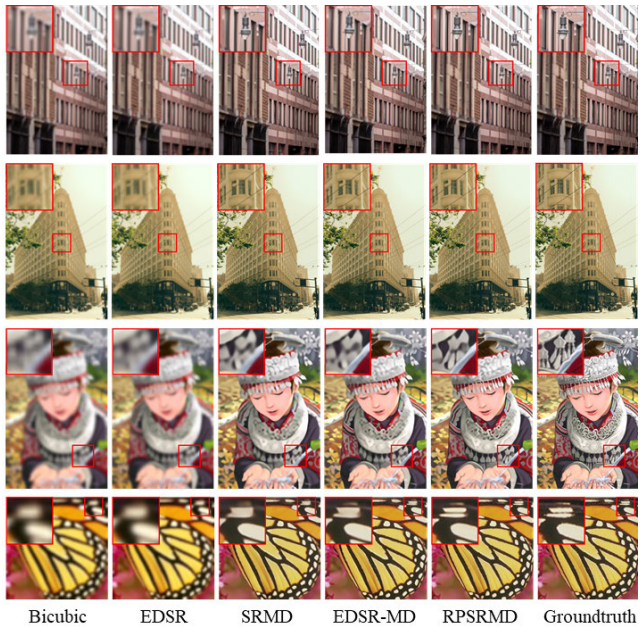



**FIGURE 10.** Visual comparison between SR methods on synthetic LR images with Gaussian blur strength $\sigma = 2.6$.

In Fig. 8, we compare our methods to Bicubic interpolation, EDSR and SRMD on the real LR image 'chips'. The performance of EDSR is severely affected by the unknown degradation. Because SRMD takes known degradation information as input, it reconstructs better SR image compared with EDSR but the reconstruction result still looks blurry. In comparison, our method can not only produces sharper edges but also provides good recognizability of the characters printed on the chip.

Afterwards, we generate LR images from high quality HR images by applying different Gaussian blurs ($\sigma = 1.2, \sigma = 2.6$) and bicubic downsampler, and restore them by Bicubic

interpolation, EDSR, SRMD, EDSR-MD and our proposed RPSRMD, as shown in the enlarged parts in Fig. 9 and 10. The EDSR reconstructs blurry SR images under unexpected degradation type in Fig. 9, and the results become worse when kernel width rises in Fig. 10. The SRMD which takes known degradation type as part of the input can produce better results than the EDSR. Meanwhile EDSR-MD produces similar results as SRMD does though without known degradation type as input because of the great representational capacity of the deep network. Our proposed RPSRMD outperforms the aforementioned methods when different Gaussian kernel widths are applied in Fig. 9 and Fig. 10 respectively. In Fig. 9 and Fig. 10, we can find in the first two rows that RPSRMD produces less blurry results and sharper textures, and we can find in the last two rows that RPSRMD is able to recover plausible image details which is similar with them in the groundtruth image. Both the qualitative and quantitative results prove the RPSRMD achieves the SOTA in unknown degradations which is highly applicable in real scenario.

## V. CONCLUSIONS

In order to address the common SR problem that the degradation priori hypotheses often fail and lead to reconstruction error in real scenario, we propose the RPSRMD framework to handle SISR problem in multiple Gaussian degradations scenarios. By designing an AED network called as PRGen, we can robustly extract from the degraded LR images a shared image content prior that is highly consistent when the LR images are blurred by different Gaussian kernels. Afterwards, we propose a deep ResNet called as PResNet to take the image content prior as side information to recover the SR image from its LR observations with different Gaussian blur strength. The experiments prove that our method is superior to many state-of-the-art SR methods that were designed and trained to handle multi-degradation. In this paper, we use the Gaussian blur as the blur model since it is one of the most common blur types in many photography cases. In the future work, the underlying principles used in this paper are promising to extend in more complicated and even mixture blur types.

## REFERENCES


[1] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.

[2] 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.

[3] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[4] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang, "Wide activation for efficient and accurate image super-resolution," 2018, *arXiv:1808.08718*. [Online]. Available: http://arxiv.org/abs/1808.08718

[5] Y. Zhao, G. Li, W. Xie, W. Jia, H. Min, and X. Liu, "GUN: Gradual upsampling network for single image super-resolution," *IEEE Access*, vol. 6, pp. 39363–39374, 2018.

[6] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *IEEE Access*, vol. 7, pp. 31959–31970, 2019.

[7] Z. Lu, Z. Yu, P. Yali, L. Shigang, W. Xiaojun, L. Gang, and R. Yuan, "Fast single image super-resolution via dilated residual networks," *IEEE Access*, vol. 7, pp. 109729–109738, 2019.

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

[9] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2832–2839.

[10] G. Wu, L. Zhao, W. Wang, L. Zeng, and J. Chen, "PRED: A parallel network for handling multiple degradations via single model in single image super-resolution," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2881–2885.

[11] I. Begin and F. R. Ferrie, "Blind super-resolution using a learning-based approach," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, 2004, pp. 85–89.

[12] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. CVPR*, Jun. 2011, pp. 449–456.

[13] 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.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2272–2279.

[15] 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.

[16] S. A. Bigdeli, M. Zwicker, P. Favaro, and M. Jin, "Deep mean-shift priors for image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 763–772.

[17] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

[18] G. Riegler, S. Schulter, M. Ruther, and H. Bischof, "Conditioned regression models for non-blind single image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 522–530.

[19] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.

[20] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1604–1613.

[21] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 114–125.

[22] 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.* Cham, Switzerland: Springer, 2014, pp. 184–199.

[23] 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.

[24] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 391–407.

[25] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "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.

[26] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.

[27] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[28] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 1–16.

[29] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.
[30] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*. [Online]. Available: http://arxiv.org/abs/1606.08921
[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

**WENYI WANG** (Member, IEEE) received the B.S. degree from Wuhan University, China, in 2009, and the M.S. and Ph.D. degrees from the University of Ottawa, Ottawa, ON, Canada, in 2011 and 2016, respectively. Since February 2017, he has been a Lecturer with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), where he joined the MediaLab. His current research interests include computer vision, pattern recognition, and video processing.

**GUANGYANG WU** (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018, where he is currently pursuing the M.S. degree in electronic and communication engineering. His research interests include single image super-resolution and deep learning.

**WEITONG CAI** (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2017, where he is currently pursuing the M.S. degree in information and communication engineering.

From February 2015 to June 2015, he was an Exchange Student with National Chiao Tung University (NCTU), Hsinchu, Taiwan. His research interests include super-resolution and machine learning.

**LIAOYUAN ZENG** (Member, IEEE) received the B.Eng. degree in telecommunication engineering from Southwest Jiaotong University, Chengdu, China, in 2005, the M.Eng. degree in computer and communication engineering from the University of Limerick, Limerick, Ireland, in 2006, and the Ph.D. degree in electrical engineering from the University of Limerick, in 2011.

He is currently an Associate Professor with the MediaLab, University of Electronic Science and Technology of China, Chengdu. His current research interests include radio resource assignment in cognitive radio systems, especially multicarrier, and low-power-based systems. He is an active member of the research community, participating in activities that help facilitate the exchange of ideas between members within the community. He is a member of the IEEE Communications Society and the IEEE Signal Processing Society. He received the Best Student Paper Award in ICETE 2008 and the Best Paper Award in WebTel 2009. He serves as the Track Co-Chair of BWCCA 2012, and he was a member of the Technical Program Committee of VTC2012-Fall, AICT 2012/2011/2010, and COCORA 2012/2011.

**JIANWEN CHEN** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China. From 2007 to 2010, he was a Staff Researcher in IBM Research. Since September 2010, he has been with the Image Communications Lab, University of California, Los Angeles (UCLA), where he focused on the research of high-efficiency video coding (HEVC) techniques, high-performance computing architecture and video analysis applications. Since October 2012, he has been with the Eye Research Institute, Harvard University, where he focused on the video quality assessment and visual application research. Since December 2015, he has also been a Full Professor with the University of Electronic Science and Technology of China (UESTC), where he is currently acting as the Director of the MediaLab. His research interests are in general area of signal processing algorithms and systems, and in particular video signal algorithms, video content analysis, and video coding. He has more than 100 publications in these areas. He has acted as TPC Member of many academic conferences. He has been serving as the Chairman of the MPEG Internet Video Codec Adhoc Group, since February 2012. He was nominated as the Chancellor's Postdoctoral Researcher of UCLA in 2012.