

Learning a Single Convolutional Super-Resolution Network for Multiple Degradations

Kai Zhang^{1,2,3}, Wangmeng Zuo^{1,*}, Lei Zhang²

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

²Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China

³DAMO Academy, Alibaba Group

caskaizhang@gmail.com, wzmzuo@hit.edu.cn, cslzhang@comp.polyu.edu.hk

Abstract

Recent years have witnessed the unprecedented success of deep convolutional neural networks (CNNs) in single image super-resolution (SISR). However, existing CNN-based SISR methods mostly assume that a low-resolution (LR) image is bicubically downsampled from a high-resolution (HR) image, thus inevitably giving rise to poor performance when the true degradation does not follow this assumption. Moreover, they lack scalability in learning a single model to non-blindly deal with multiple degradations. To address these issues, we propose a general framework with dimensionality stretching strategy that enables a single convolutional super-resolution network to take two key factors of the SISR degradation process, i.e., blur kernel and noise level, as input. Consequently, the super-resolver can handle multiple and even spatially variant degradations, which significantly improves the practicability. Extensive experimental results on synthetic and real LR images show that the proposed convolutional super-resolution network not only can produce favorable results on multiple degradations but also is computationally efficient, providing a highly effective and scalable solution to practical SISR applications.

1. Introduction

Single image super-resolution (SISR) aims to recover a high-resolution (HR) version of a low-resolution (LR) input. As a classical problem, SISR is still an active yet challenging research topic in the field of computer vision due to its ill-posedness nature and high practical values [2]. In the typical SISR framework, an LR image y is modeled as the output of the following degradation process:

$$y = (x \otimes k) \downarrow_s + n, \quad (1)$$

where $x \otimes k$ represents the convolution between a blur kernel k and a latent HR image x , \downarrow_s is a subsequent downsampling operation with scale factor s , and n usually is additive white Gaussian noise (AWGN) with standard deviation (noise level) σ .

SISR methods can be broadly classified into three categories, i.e., interpolation-based methods, model-based optimization methods and discriminative learning methods. Interpolation-based methods such as nearest-neighbor, bilinear and bicubic interpolators are simple and efficient but have very limited performance. By exploiting powerful image priors (e.g., the non-local self-similarity prior [11, 32], sparsity prior [52] and denoiser prior [4, 13, 57]), model-based optimization methods are flexible to reconstruct relative high-quality HR images, but they usually involve a time-consuming optimization procedure. Although the integration of convolutional neural network (CNN) denoiser prior and model-based optimization can improve the efficiency to some extent, it still suffers from the typical drawbacks of model-based optimization methods, e.g., it is not in an end-to-end learning manner and involves hand-designed parameters [57]. As an alternative, discriminative learning methods have attracted considerable attentions due to their favorable SISR performance in terms of effectiveness and efficiency. Notably, recent years have witnessed a dramatic upsurge of using CNN for SISR.

In this paper, we focus on discriminative CNN methods for SISR so as to exploit the merits of CNN, such as the fast speed by parallel computing, high accuracy by end-to-end training, and tremendous advances in training and designing networks [16, 18, 21, 28]. While several SISR models based on discriminative CNN have reported impressive results, they suffer from a common drawback: their models are specialized for a single simplified degradation (e.g., bicubic degradation) and lack scalability to handle multiple degradations by using a single model. Because the practical degradation of SISR is much more complex [40, 51], the performance of learned CNN models may deteriorate seriously when the assumed degradation deviates from the true one, making them less effective in practical scenarios. It has been pointed out that the blur kernel plays a vital role for the success of SISR methods and the mismatch of blur kernels will largely deteriorate the final SISR results [12]. However, little work has been done on how to design a CNN to address this crucial issue.

*Corresponding author.

Given the facts above, it is natural to raise the following questions, which are the focus of our paper: (i) Can we learn a single model to effectively handle multiple and even spatially variant degradations? (ii) Is it possible to use synthetic data to train a model with high practicability? This work aims to make one of the first attempts towards answering these two questions.

To answer the first question, we revisit and analyze the general model-based SISR methods under the maximum a posteriori (MAP) framework. Then we argue that one may tackle this issue by taking LR input, blur kernel and noise level as input to CNN but their dimensionality mismatch makes it difficult to design a single convolutional super-resolution network. In view of this, we introduce a dimensionality stretching strategy which facilitates the network to handle multiple and even spatially variant degradations with respect to blur kernel and noise. To the best of our knowledge, there is no attempt to consider both the blur kernel and noise for SISR via training a single CNN model.

For the second question, we will show that it is possible to learn a practical super-resolver using synthetic data. To this end, a large variety of degradations with different combinations of blur kernels and noise levels are sampled to cover the degradation space. In a practical scenario, even the degradation is more complex (*e.g.*, the noise is non-AWGN), we can select the best fitted degradation model rather than the bicubic degradation to produce a better result. It turns out that, by choosing a proper degradation, the learned SISR model can yield perceptually convincing results on real LR images. It should be noted that we make no effort to use specialized network architectures but use the plain CNN as in [9, 41].

The main contributions of this paper are summarized in the following:

- We propose a simple yet effective and scalable deep CNN framework for SISR. The proposed model goes beyond the widely-used bicubic degradation assumption and works for multiple and even spatially variant degradations, thus making a substantial step towards developing a practical CNN-based super-resolver for real applications.
- We propose a novel dimensionality stretching strategy to address the dimensionality mismatch between LR input image, blur kernel and noise level. Although this strategy is proposed for SISR, it is general and can be extended to other tasks such as deblurring.
- We show that the proposed convolutional super-resolution network learned from synthetic training data can not only produce competitive results against state-of-the-art SISR methods on synthetic LR images but also give rise to visually plausible results on real LR images.

2. Related Work

The first work of using CNN to solve SISR can be traced back to [8] where a three-layer super-resolution network (SRCNN) was proposed. In the extended work [9], the authors investigated the impact of depth on super-resolution and empirically showed that the difficulty of training deeper model hinders the performance improvement of CNN super-resolvers. To overcome the training difficulty, Kim *et al.* [24] proposed a very deep super-resolution (VDSR) method with residual learning strategy. Interestingly, they showed that VDSR can handle multiple scales super-resolution. By analyzing the relation between CNN and MAP inference, Zhang *et al.* [56] pointed out that CNN mainly model the prior information and they empirically demonstrated that a single model can handle multiple scales super-resolution, image deblocking and image denoising. While achieving good performance, the above methods take the bicubically interpolated LR image as input, which not only suffers from high computational cost but also hinders the effective expansion of receptive field.

To improve the efficiency, some researchers resort to directly manipulating the LR input and adopting an upscaling operation at the end of the network. Shi *et al.* [41] introduced an efficient sub-pixel convolution layer to up-scale the LR feature maps into HR images. Dong *et al.* [10] adopted a deconvolution layer at the end of the network to perform upsampling. Lai *et al.* [27] proposed a Laplacian pyramid super-resolution network (LapSRN) that takes an LR image as input and progressively predicts the sub-band residuals with transposed convolutions in a coarse-to-fine manner. To improve the perceptual quality at a large scale factor, Ledig *et al.* [29] proposed a generative adversarial network [16] based super-resolution (SRGAN) method. In the generator network of SRGAN, two sub-pixel convolution layers are used to efficiently upscale the LR input by a factor of 4.

Although various techniques have been proposed for SISR, the above CNN-based methods are tailored to the widely-used settings of bicubic degradation, neglecting their limited applicability for practical scenarios. An interesting line of CNN-based methods which can go beyond bicubic degradation adopt a CNN denoiser to solve SISR via a model-based optimization framework [4, 34, 57]. For example, the method proposed in [57] can handle the widely-used Gaussian degradation as in [11]. However, manually selecting the hyper-parameters for different degradations is not a trivial task [39]. As a result, it is desirable to learn a single SISR model which can handle multiple degradations with high practicability. This paper attempts to give a positive answer.

Due to the limited space, we can only discuss some of the related works here. Other CNN-based SISR methods can be found in [6, 22, 23, 30, 37, 42, 44, 45, 46, 53, 58].

3. Method

3.1. Degradation Model

Before solving the problem of SISR, it is important to have a clear understanding of the degradation model which is not limited to Eqn. (1). Another practical degradation model can be given by

$$\mathbf{y} = (\mathbf{x} \downarrow_s) \otimes \mathbf{k} + \mathbf{n}. \quad (2)$$

When \downarrow is the bicubic downsampler, Eqn. (2) corresponds to a deblurring problem followed by a SISR problem with bicubic degradation. Thus, it can benefit from existing deblurring methods and bicubic degradation based SISR methods. Due to limited space, we only consider the more widely assumed degradation model given in Eqn. (1). Nevertheless, our method is general and can be easily extended to handle Eqn. (2). In the following, we make a short discussion on blur kernel \mathbf{k} , noise \mathbf{n} and downsampler \downarrow .

Blur kernel. Different from image deblurring, the blur kernel setting of SISR is usually simple. The most popular choice is isotropic Gaussian blur kernel parameterized by standard deviation or kernel width [11, 51]. In [38], anisotropic Gaussian blur kernels are also used. In practice, more complex blur kernel models used in deblurring task, such as motion blur [5], can be further considered. Empirical and theoretical analyses have revealed that the influence of an accurate blur kernel is much larger than that of sophisticated image priors [12]. Specifically, when the assumed kernel is smoother than the true kernel, the recovered image is over-smoothed. Most of SISR methods actually favor for such case. On the other hand, when the assumed kernel is sharper than the true kernel, high frequency ringing artifacts will appear.

Noise. While being of low-resolution, the LR images are usually also noisy. Directly super-resolving the noisy input without noise removal would amplify the unwanted noise, resulting in visually unpleasant results. To address this problem, the straightforward way is to perform denoising first and then enhance the resolution. However, the denoising pre-processing step tends to lose detail information and would deteriorate the subsequent super-resolution performance [43]. Thus, it would be highly desirable to jointly perform denoising and super-resolution.

Downsampler. Existing literatures have considered two types of downsamplers, including direct downsampler [11, 17, 36, 51, 55] and bicubic downsampler [7, 12, 14, 15, 47, 52]. In this paper, we consider the bicubic downsampler since when \mathbf{k} is delta kernel and the noise level is zero, Eqn. (1) turns into the widely-used bicubic degradation model. It should be pointed out that, different from blur kernel and noise which vary in a general degradation model, downsampler is assumed to be fixed.

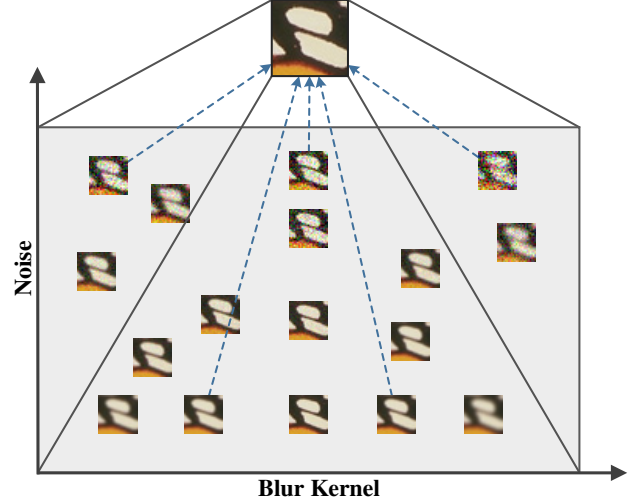


Figure 1. An illustration of different degradations for SISR. The scale factor is 2. The general degradation models of Eqns. (1) and (2) assume an HR image actually can degrade into many LR images, whereas bicubic degradation model assumes an HR image corresponds to a single LR image.

Though blur kernel and noise have been recognized as key factors for the success of SISR and several methods have been proposed to consider those two factors, there has been little effort towards simultaneously considering blur kernel and noise in a single CNN framework. It is a challenging task since the degradation space with respect to blur kernel and noise is rather large (see Figure 1 as an example). One relevant work is done by Zhang *et al.* [57]; nonetheless, their method is essentially a model-based optimization method and thus suffers from several drawbacks as mentioned previously. In another related work, Riegler *et al.* [38] exploited the blur kernel information into the SISR model. Our method differs from [38] on two major aspects. First, our method considers a more general degradation model. Second, our method exploits a more effective way to parameterize the degradation model.

3.2. A Perspective from MAP Framework

Though existing CNN-based SISR methods are not necessarily derived under the traditional MAP framework, they have the same goal. We revisit and analyze the general MAP framework of SISR, aiming to find the intrinsic connections between the MAP principle and the working mechanism of CNN. Consequently, more insights on CNN architecture design can be obtained.

Due to the ill-posed nature of SISR, regularization needs to be imposed to constrain the solution. Mathematically, the HR counterpart of an LR image \mathbf{y} can be estimated by solving the following MAP problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2\sigma^2} \|(\mathbf{x} \otimes \mathbf{k}) \downarrow_s - \mathbf{y}\|^2 + \lambda \Phi(\mathbf{x}) \quad (3)$$

where $\frac{1}{2\sigma^2} \|(\mathbf{x} \otimes \mathbf{k}) \downarrow_s - \mathbf{y}\|^2$ is the data fidelity term, $\Phi(\mathbf{x})$ is the regularization term (or prior term) and λ is the trade-off parameter. Simply speaking, Eqn. (3) conveys two points: (i) the estimated solution should not only accord with the degradation process but also have the desired property of clean HR images; (ii) $\hat{\mathbf{x}}$ is a function of LR image \mathbf{y} , blur kernel \mathbf{k} , noise level σ , and trade-off parameter λ . Therefore, the MAP solution of (non-blind) SISR can be formulated as

$$\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y}, \mathbf{k}, \sigma, \lambda; \Theta) \quad (4)$$

where Θ denotes the parameters of the MAP inference.

By treating CNN as a discriminative learning solution to Eqn. (4), we can have the following insights.

- Because the data fidelity term corresponds to the degradation process, accurate modeling of the degradation plays a key role for the success of SISR. However, existing CNN-based SISR methods with bicubic degradation actually aim to solve the following problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x} \downarrow_s - \mathbf{y}\|^2 + \Phi(\mathbf{x}). \quad (5)$$

Inevitably, their practicability is very limited.

- To design a more practical SISR model, it is preferable to learn a mapping function like Eqn. (4), which covers more extensive degradations. It should be stressed that, since λ can be absorbed into σ , Eqn. (4) can be reformulated as

$$\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y}, \mathbf{k}, \sigma; \Theta). \quad (6)$$

- Considering that the MAP framework (Eqn. (3)) can perform generic image super-resolution with the same image prior, it is intuitive to jointly perform denoising and SISR in a unified CNN framework. Moreover, the work [56] indicates that the parameters of the MAP inference mainly model the prior; therefore, CNN has the capacity to deal with multiple degradations via a single model.

From the viewpoint of MAP framework, one can see that the goal of SISR is to learn a mapping function $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y}, \mathbf{k}, \sigma; \Theta)$ rather than $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y}; \Theta)$. However, it is not an easy task to directly model $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{y}, \mathbf{k}, \sigma; \Theta)$ via CNN. The reason lies in the fact that the three inputs \mathbf{y} , \mathbf{k} and σ have different dimensions. In the next subsection, we will propose a simple dimensionality stretching strategy to resolve this problem.

3.3. Dimensionality Stretching

The proposed dimensionality stretching strategy is schematically illustrated in Figure 2. Suppose the inputs consist of a blur kernel of size $p \times p$, a noise level σ and an LR image of size $W \times H \times C$, where C denotes the number of channels. The blur kernel is first vectorized into a vector

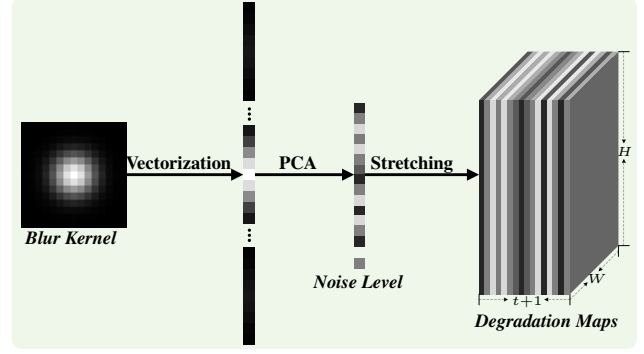


Figure 2. Schematic illustration of the dimensionality stretching strategy. For an LR image of size $W \times H$, the vectorized blur kernel is first projected onto a space of dimension t and then stretched into a tensor \mathcal{M} of size $W \times H \times (t + 1)$ with the noise level.

of size $p^2 \times 1$ and then projected onto t -dimensional linear space by the PCA (Principal Component Analysis) technique. After that, the concatenated low dimensional vector and the noise level, denoted by \mathbf{v} , are stretched into degradation maps \mathcal{M} of size $W \times H \times (t + 1)$, where all the elements of i -th map are \mathbf{v}_i . By doing so, the degradation maps then can be concatenated with the LR image, making CNN possible to handle the three inputs. Such a simple strategy can be easily exploited to deal with spatially variant degradations by considering the fact that the degradation maps can be non-uniform.

3.4. Proposed Network

The proposed super-resolution network for multiple degradations, denoted by SRMD, is illustrated in Figure 3. As one can see, the distinctive feature of SRMD is that it takes the concatenated LR image and degradation maps as input. To show the effectiveness of the dimensionality stretching strategy, we resort to plain CNN without complex architectural engineering. Typically, to super-resolve an LR image with a scale factor of s , SRMD first takes the concatenated LR image and degradation maps of size $W \times H \times (C + t + 1)$ as input. Then, similar to [24], a cascade of 3×3 convolutional layers are applied to perform the non-linear mapping. Each layer is composed of three types of operations, including Convolution (Conv), Rectified Linear Units (ReLU) [26], and Batch Normalization (BN) [20]. Specifically, “Conv + BN + ReLU” is adopted for each convolutional layer except the last convolutional layer which consists of a single “Conv” operation. Finally, a sub-pixel convolution layer [41] is followed by the last convolutional layer to convert multiple HR subimages of size $W \times H \times s^2 C$ to a single HR image of size $sW \times sH \times C$.

For all scale factors 2, 3 and 4, the number of convolutional layers is set to 12, and the number of feature maps in each layer is set to 128. We separately learn models for each scale factor. In particular, we also learn the models for

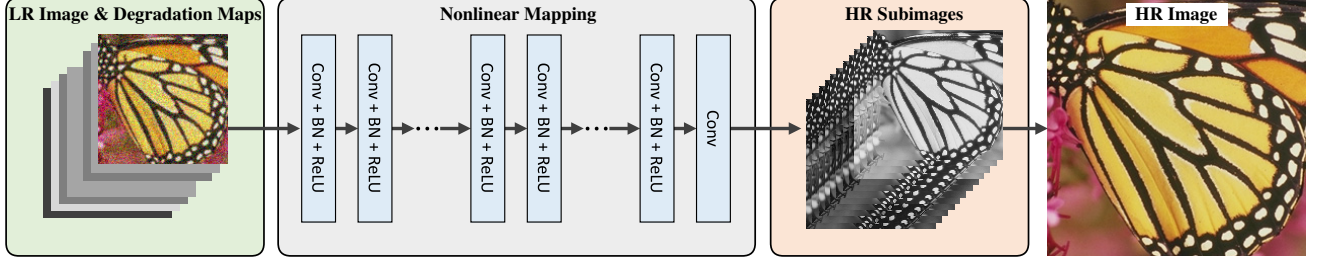


Figure 3. The architecture of the proposed convolutional super-resolution network. In contrast to other CNN-based SISR methods which only take the LR image as input and lack scalability to handle other degradations, the proposed network takes the concatenated LR image and degradation maps as input, thus allowing a single model to manipulate multiple and even spatially variant degradations.

noise-free degradation, namely SRMDNF, by removing the connection of the noise level map in the first convolutional filter and fine-tuning with new training data.

It is worth pointing out that neither residual learning nor bicubically interpolated LR image is used for the network design due to the following reasons. First, with a moderate network depth and advanced CNN training and design such as ReLU [26], BN [20] and Adam [25], it is easy to train the network without the residual learning strategy. Second, since the degradation involves noise, bicubically interpolated LR image would aggravate the complexity of noise which in turn will increase the difficulty of training.

3.5. Why not Learn a Blind Model?

To enhance the practicability of CNN for SISR, it seems the most straightforward way is to learn a blind model with synthesized training data by different degradations. However, such blind model does not perform as well as expected. First, the performance deteriorates seriously when the blur kernel model is complex, *e.g.*, motion blur. This phenomenon can be explained by the following example. Given an HR image, a blur kernel and corresponding LR image, shifting the HR image to left by one pixel and shifting the blur kernel to right by one pixel would result in the same LR image. Thus, an LR image may correspond to different HR images with pixel shift. This in turn would aggravate the pixel-wise average problem [29], typically leading to over-smoothed results. Second, the blind model without specially designed architecture design has inferior generalization ability and performs poorly in real applications.

In contrast, non-blind model for multiple degradations suffers little from the pixel-wise average problem and has better generalization ability. First, the degradation maps contain the warping information and thus can enable the network to have spatial transformation capability. For clarity, one can treat the degradation maps induced by blur kernel and noise level as the output of a spatial transformer as in [21]. Second, by anchoring the model with degradation maps, the non-blind model generalizes easily to unseen degradations and has the ability to control the tradeoff between data fidelity term and regularization term.

4. Experiments

4.1. Training Data Synthesis and Network Training

Before synthesizing LR images according to Eqn. (1), it is necessary to define the blur kernels and noise level range, as well as providing a large-scale clean HR image set.

For the blur kernels, we follow the kernel model of isotropic Gaussian with a fixed kernel width which has been proved practically feasible in SISR applications. Specifically, the kernel width ranges are set to $[0.2, 2]$, $[0.2, 3]$ and $[0.2, 4]$ for scale factors 2, 3 and 4, respectively. We sample the kernel width by a stride of 0.1. The kernel size is fixed to 15×15 . To further expand the degradation space, we also consider a more general kernel assumption, *i.e.*, anisotropic Gaussian, which is characterized by a Gaussian probability density function $\mathcal{N}(\mathbf{0}, \Sigma)$ with zero mean and varying covariance matrix Σ [38]. The space of such Gaussian kernel is determined by rotation angle of the eigenvectors of Σ and scaling of corresponding eigenvalues. We set the rotation angle range to $[0, \pi]$. For the scaling of eigenvalues, it is set from 0.5 to 6, 8 and 10 for scale factors 2, 3 and 4, respectively.

Although we adopt the bicubic downsampler throughout the paper, it is straightforward to train a model with direct downsampler. Alternatively, we can also include the degradations with direct downsampler by approximating it. Specifically, given a blur kernel \mathbf{k}_d under direct downsampler \downarrow^d , we can find the corresponding blur kernel \mathbf{k}_b under bicubic downsampler \downarrow^b by solving the following problem with a data-driven method

$$\mathbf{k}_b = \arg \min_{\mathbf{k}_b} \|(\mathbf{x} \otimes \mathbf{k}_b) \downarrow_s^b - (\mathbf{x} \otimes \mathbf{k}_d) \downarrow_s^d\|^2, \quad \forall \mathbf{x}. \quad (7)$$

In this paper, we also include such degradations for scale factor 3.

Once the blur kernels are well-defined or learned, we then uniformly sample substantial kernels and aggregate them to learn the PCA projection matrix. By preserving about 99.8% of the energy, the kernels are projected onto a space of dimension 15 (*i.e.*, $t = 15$). The visualization of some typical blur kernels for scale factor 3 and some PCA eigenvectors is shown in Figure 4.

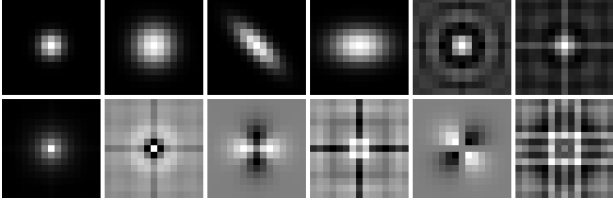


Figure 4. Visualization of six typical blur kernels (**first row**) of isotropic Gaussian (first two), anisotropic Gaussian (middle two) and estimated ones for direct downsampler (last two) for scale factor 3 and PCA eigenvectors (**second row**) for the first six largest eigenvalues.

For the noise level range, we set it as $[0, 75]$. Because the proposed method operates on RGB channels rather than Y channel in YCbCr color space, we collect a large-scale color images for training, including 400 BSD [33] images, 800 training images from DIV2K dataset [1] and 4,744 images from WED dataset [31].

Then, given an HR image, we synthesize LR image by blurring it with a blur kernel k and bicubically downsampling it with a scale factor s , followed by an addition of AWGN with noise level σ . The LR patch size is set to 40×40 which means the corresponding HR patch sizes for scale factors 2, 3, and 4 are 80×80 , 120×120 and 160×160 , respectively.

In the training phase, we randomly select a blur kernel and a noise level to synthesize an LR image and crop $N = 128 \times 3,000$ LR/HR patch pairs (along with the degradation maps) for each epoch. We optimize the following loss function using Adam [25]

$$\mathcal{L}(\Theta) = \frac{1}{2N} \sum_{i=1}^N \|\mathcal{F}(\mathbf{y}_i, \mathcal{M}_i; \Theta) - \mathbf{x}_i\|^2. \quad (8)$$

The mini-batch size is set to 128. The learning rate starts from 10^{-3} and reduces to 10^{-4} when the training error stops decreasing. When the training error keeps unchanged in five sequential epochs, we merge the parameters of each batch normalization into the adjacent convolution filters. Then, a small learning rate of 10^{-5} is used for additional 100 epochs to fine-tune the model. Since SRMDNF is obtained by fine-tuning SRMD, its learning rate is fixed to 10^{-5} for 200 epochs.

We train the models in Matlab (R2015b) environment with MatConvNet package [48] and an Nvidia Titan X Pascal GPU. The training of a single SRMD model can be done in about two days. The source code can be downloaded at <https://github.com/cszn/SRMD>.

4.2. Experiments on Bicubic Degradation

As mentioned above, instead of handling the bicubic degradation only, our aim is to learn a single network to handle multiple degradations. However, in order to show the advantage of the dimensionality stretching strategy, the

proposed method is also compared with other CNN-based methods specifically designed for bicubic degradation.

Table 1 shows the PSNR and SSIM [50] results of state-of-the-art CNN-based SISR methods on four widely-used datasets. As one can see, SRMD achieves comparable results with VDSR at small scale factor and outperforms VDSR at large scale factor. In particular, SRMDNF achieves the best overall quantitative results. Using ImageNet dataset [26] to train the specific model with bicubic degradation, SRResNet performs slightly better than SRMDNF on scale factor 4. To further compare with other methods such as VDSR, we also have trained a SRMDNF model (for scale factor 3) which operates on Y channel with 291 training images. The learned model achieves 33.97dB, 29.96dB, 28.95dB and 27.42dB on Set5, Set14, BSD100 and Urban100, respectively. As a result, it can still outperform other competing methods. The possible reason lies in that the SRMDNF with multiple degradations shares the same prior in the MAP framework which facilitates the implicit prior learning and thus benefits to PSNR improvement. This also can explain why VDSR with multiple scales improves the performance.

For the GPU run time, SRMD spends 0.084, 0.042 and 0.027 seconds to reconstruct an HR image of size $1,024 \times 1,024$ for scale factors 2, 3 and 4, respectively. As a comparison, the run time of VDSR is 0.174 second for all scale factors. Figure 5 shows the visual results of different methods. One can see that our proposed method yields very competitive performance against other methods.

4.3. Experiments on General Degradations

In this subsection, we evaluate the performance of the proposed method on general degradations. The degradation settings are given in Table 2. We only consider the isotropic Gaussian blur kernel for an easy comparison. To further show the scalability of the proposed method, another widely-used degradation [11] which involves 7×7 Gaussian kernel with width 1.6 and direct downsampler with scale factor 3 is also included. We compare the proposed method with VDSR, two model-based methods (*i.e.*, NC-SR [11] and IRCNN [57]), and a cascaded denoising-SISR method (*i.e.*, DnCNN [56]+SRMDNF).

The quantitative results of different methods with different degradations on Set5 are provided in Table 2, from which we have observations and analyses as follows. First, the performance of VDSR deteriorates seriously when the assumed bicubic degradation deviates from the true one. Second, SRMD produces much better results than NCSR and IRCNN, and outperforms DnCNN+SRMDNF. In particular, the PSNR gain of SRMD over DnCNN+SRMDNF increases with the kernel width which verifies the advantage of joint denoising and super-resolution. Third, by setting proper blur kernel, the proposed method delivers good per-

Table 1. Average PSNR and SSIM results for bicubic degradation on datasets Set5 [3], Set14 [54], BSD100 [33] and Urban100 [19]. The best two results are highlighted in red and blue colors, respectively.

Dataset	Scale Factor	Bicubic	SRCNN [9]	VDSR [24]	SRResNet [29]	DRRN [44]	LapSRN [27]	SRMD	SRMDNF
		PSNR / SSIM							
Set5	×2	33.64 / 0.929	36.62 / 0.953	37.56 / 0.959	–	37.66 / 0.959	37.52 / 0.959	37.53 / 0.959	37.79 / 0.960
	×3	30.39 / 0.868	32.74 / 0.908	33.67 / 0.922	–	33.93 / 0.923	33.82 / 0.922	33.86 / 0.923	34.12 / 0.925
	×4	28.42 / 0.810	30.48 / 0.863	31.35 / 0.885	32.05 / 0.891	31.58 / 0.886	31.54 / 0.885	31.59 / 0.887	31.96 / 0.893
Set14	×2	30.22 / 0.868	32.42 / 0.906	33.02 / 0.913	–	33.19 / 0.913	33.08 / 0.913	33.12 / 0.914	33.32 / 0.915
	×3	27.53 / 0.774	29.27 / 0.821	29.77 / 0.832	–	29.94 / 0.834	29.89 / 0.834	29.84 / 0.833	30.04 / 0.837
	×4	25.99 / 0.702	27.48 / 0.751	27.99 / 0.766	28.49 / 0.780	28.18 / 0.770	28.19 / 0.772	28.15 / 0.772	28.35 / 0.777
BSD100	×2	29.55 / 0.843	31.34 / 0.887	31.89 / 0.896	–	32.01 / 0.897	31.80 / 0.895	31.90 / 0.896	32.05 / 0.898
	×3	27.20 / 0.738	28.40 / 0.786	28.82 / 0.798	–	28.91 / 0.799	28.82 / 0.798	28.87 / 0.799	28.97 / 0.803
	×4	25.96 / 0.667	26.90 / 0.710	27.28 / 0.726	27.58 / 0.735	27.35 / 0.726	27.32 / 0.727	27.34 / 0.728	27.49 / 0.734
Urban100	×2	26.66 / 0.841	29.53 / 0.897	30.76 / 0.914	–	31.02 / 0.916	30.82 / 0.915	30.89 / 0.916	31.33 / 0.920
	×3	24.46 / 0.737	26.25 / 0.801	27.13 / 0.828	–	27.38 / 0.833	27.07 / 0.828	27.27 / 0.833	27.57 / 0.840
	×4	23.14 / 0.657	24.52 / 0.722	25.17 / 0.753	–	25.35 / 0.758	25.21 / 0.756	25.34 / 0.761	25.68 / 0.773

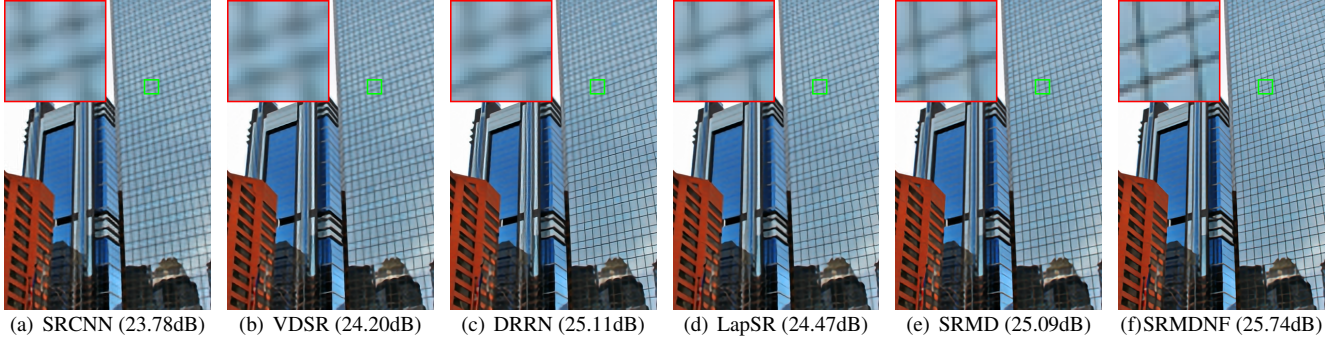


Figure 5. SISR performance comparison of different methods with scale factor 4 on image “*Img_099*” from Urban100.

Table 2. Average PSNR and SSIM results of different methods with different degradations on Set5. The best results are highlighted in red color. The results highlighted in gray color indicate unfair comparison due to mismatched degradation assumption.

Degradation Settings			VDSR [24]	NCSR [11]	IRCNN [57]	DnCNN [56]+SRMDNF	SRMD	SRMDNF
Kernel Width	Down-sampler	Noise Level	PSNR (×2/×3/×4)					
0.2	Bicubic	0	37.56/33.67/31.35	– /23.82/–	37.43/33.39/31.02	–	37.53/33.86/31.59	37.79/34.12/31.96
0.2	Bicubic	15	26.02/25.40/24.70	–	32.60/30.08/28.35	32.47/30.07/28.31	32.76/30.43/28.79	–
0.2	Bicubic	50	16.02/15.72/15.46	–	28.20/26.25/24.95	28.20/26.27/24.93	28.51/26.48/25.18	–
1.3	Bicubic	0	30.57/30.24/29.72	– /21.81/–	36.01/33.33/31.01	–	37.04/33.77/31.56	37.45/34.16/31.99
1.3	Bicubic	15	24.82/24.70/24.30	–	29.96/28.68/27.71	27.68/28.78/27.71	30.98/29.43/28.21	–
1.3	Bicubic	50	15.89/15.68/15.43	–	26.69/25.20/24.42	24.35/25.19/24.39	27.43/25.82/24.77	–
2.6	Bicubic	0	26.37/26.31/26.28	– /21.46/–	32.07/31.09/30.06	–	33.24/32.59/31.20	34.12/33.02/31.77
2.6	Bicubic	15	23.09/23.07/22.98	–	26.44/25.67/24.36	– /21.33/23.85	28.48/27.55/26.82	–
2.6	Bicubic	50	15.58/15.43/15.23	–	22.98/22.16/21.43	– /19.03/21.15	25.85/24.75/23.98	–
1.6	Direct	0	– /30.54/ –	– /33.02/ –	– /33.38/ –	–	– /33.74/ –	– / 34.01 / –

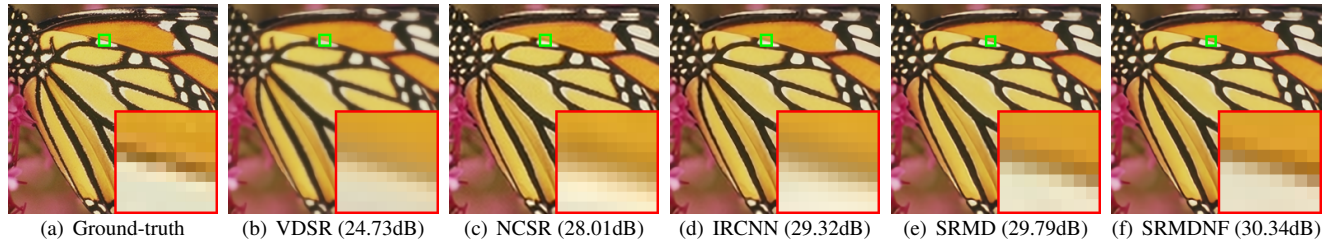


Figure 6. SISR performance comparison on image “*Butterfly*” from Set5. The degradation involves 7×7 Gaussian kernel with width 1.6 and direct downsampler with scale factor 3. Note that the comparison with VDSR is unfair because of degradation mismatch.

formance in handling the degradation with direct downsampler. The visual comparison is given in Figure 6. One can see that NCSR and IRCNN produce more visually pleasant results than VDSR since their assumed degradation matches the true one. However, they cannot recover edges as sharper as SRMD and SRMDNF.

4.4. Experiments on Spatially Variant Degradation

To demonstrate the effectiveness of SRMD for spatially variant degradation, we synthesize an LR images with spatially variant blur kernels and noise levels. Figure 7 shows the visual result of the proposed SRMD for the spatially

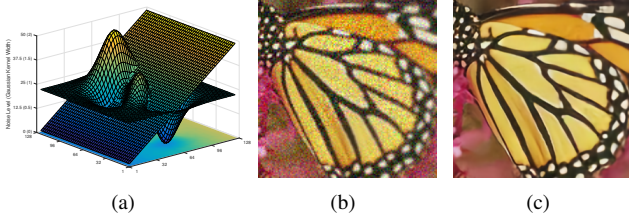


Figure 7. An example of SRMD on dealing with spatially variant degradation. (a) Noise level and Gaussian blur kernel width maps. (b) Zoomed LR image. (c) Results of SRMD with scale factor 2.

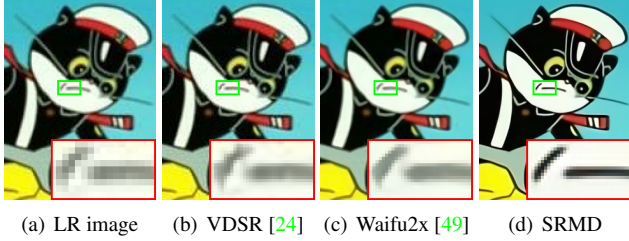


Figure 8. SISR results on image “Cat” with scale factor 2.

variant degradations. One can see that the proposed SRMD is effective in recovering the latent HR image. Note that the blur kernel is assumed to be isotropic Gaussian.

4.5. Experiments on Real Images

Besides the above experiments on LR images synthetically downsampled from HR images with known blur kernels and corrupted by AWGN with known noise levels, we also do experiments on real LR images to demonstrate the effectiveness of the proposed SRMD. Since there are no ground-truth HR images, we only provide the visual comparison.

As aforementioned, while we also use anisotropic Gaussian kernels in training, it is generally feasible to use isotropic Gaussian for most of the real LR images in testing. To find the degradation parameters with good visual quality, we use a grid search strategy rather than adopting any blur kernel or noise level estimation methods. Specifically, the kernel width is uniformly sampled from 0.1 to 2.4 with a stride of 0.1, and the noise level is from 0 to 75 with stride 5.

Figures 8 and 9 illustrate the SISR results on two real LR images “Cat” and “Chip”, respectively. The VDSR [24] is used as one of the representative CNN-based methods for comparison. For image “Cat” which is corrupted by compression artifacts, Waifu2x [49] is also used for comparison. For image “Chip” which contains repetitive structures, a self-similarity based method SelfEx [19] is also included for comparison.

It can be observed from the visual results that SRMD can produce much more visually plausible HR images than the competing methods. Specifically, one can see from Fig-

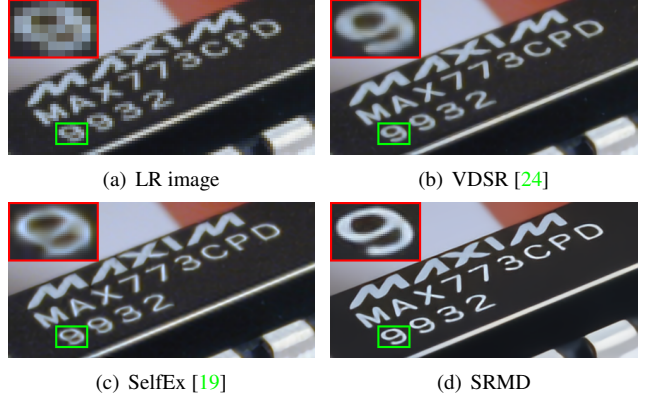


Figure 9. SISR results on real image “Chip” with scale factor 4.

ure 8 that the performance of VDSR is severely affected by the compression artifacts. While Waifu2x can successfully remove the compression artifacts, it fails to recover sharp edges. In comparison, SRMD can not only remove the unsatisfying artifacts but also produce sharp edges. From Figure 9, we can see that VDSR and SelfEx both tend to produce over-smoothed results, whereas SRMD can recover sharp image with better intensity and gradient statistics of clean images [35].

5. Conclusion

In this paper, we proposed an effective super-resolution network with high scalability of handling multiple degradations via a single model. Different from existing CNN-based SISR methods, the proposed super-resolver takes both LR image and its degradation maps as input. Specifically, degradation maps are obtained by a simple dimensionality stretching of the degradation parameters (*i.e.*, blur kernel and noise level). The results on synthetic LR images demonstrated that the proposed super-resolver can not only produce state-of-the-art results on bicubic degradation but also perform favorably on other degradations and even spatially variant degradations. Moreover, the results on real LR images showed that the proposed method can reconstruct visually plausible HR images. In summary, the proposed super-resolver offers a feasible solution toward practical CNN-based SISR applications.

6. Acknowledgements

This work is supported by National Natural Science Foundation of China (grant no. 61671182, 61471146), HK RGC General Research Fund (PolyU 152240/15E) and PolyU-Alibaba Collaborative Research Project “Quality Enhancement of Surveillance Images and Videos”. We gratefully acknowledge the support from NVIDIA Corporation for providing us the Titan Xp GPU used in this research.

References

- [1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, volume 3, pages 126–135, July 2017. 6
- [2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002. 1
- [3] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *British Machine Vision Conference*, 2012. 7
- [4] S. A. Bigdeli, M. Jin, P. Favaro, and M. Zwicker. Deep mean-shift priors for image restoration. In *Advances in Neural Information Processing Systems*, 2017. 1, 2
- [5] G. Boracchi and A. Foi. Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, 21(8):3502–3517, Aug 2012. 3
- [6] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5261–5269, 2015. 2
- [7] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep network cascade for image super-resolution. In *European Conference on Computer Vision*, pages 49–64, 2014. 3
- [8] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199, 2014. 2
- [9] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016. 2, 7
- [10] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*, pages 391–407, 2016. 2
- [11] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013. 1, 2, 3, 6, 7
- [12] N. Efzrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin. Accurate blur models vs. image priors in single image super-resolution. In *IEEE International Conference on Computer Vision*, pages 2832–2839, 2013. 1, 3
- [13] K. Egiazarian and V. Katkovnik. Single image super-resolution via BM3D sparse coding. In *European Signal Processing Conference*, pages 2849–2853, 2015. 1
- [14] W. Freeman and C. Liu. Markov random fields for super-resolution and texture synthesis. *Advances in Markov Random Fields for Vision and Image Processing*, 1:155–165, 2011. 3
- [15] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE International Conference on Computer Vision*, pages 349–356, 2009. 3
- [16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 1, 2
- [17] H. He and W.-C. Siu. Single image super-resolution using Gaussian process regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 449–456, 2011. 3
- [18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1
- [19] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 7, 8
- [20] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015. 4, 5
- [21] M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015. 1, 5
- [22] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711, 2016. 2
- [23] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016. 2
- [24] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2, 4, 7, 8
- [25] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference for Learning Representations*, 2015. 5, 6
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012. 4, 5, 6
- [27] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 624–632, July 2017. 2, 7
- [28] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 1
- [29] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, July 2017. 2, 5, 7
- [30] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, July 2017. 2
- [31] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2017. 6
- [32] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2272–2279, 2009. 1

- [33] 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 *IEEE International Conference on Computer Vision*, volume 2, pages 416–423, July 2001. 6, 7
- [34] T. Meinhardt, M. Möller, C. Hazirbas, and D. Cremers. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *IEEE International Conference on Computer Vision*, pages 1781–1790, 2017. 2
- [35] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring text images via L0-regularized intensity and gradient prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2908, 2014. 8
- [36] T. Peleg and M. Elad. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE transactions on Image Processing*, 23(6):2569–2582, 2014. 3
- [37] J. S. Ren, L. Xu, Q. Yan, and W. Sun. Shepard convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2015. 2
- [38] G. Riegler, S. Schuler, M. Ruther, and H. Bischof. Conditioned regression models for non-blind single image super-resolution. In *IEEE International Conference on Computer Vision*, pages 522–530, 2015. 3, 5
- [39] Y. Romano, M. Elad, and P. Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. 2
- [40] Y. Romano, J. Isidoro, and P. Milanfar. RAISR: rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging*, 3(1):110–125, 2017. 1
- [41] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016. 2, 4
- [42] Y. Shi, K. Wang, C. Chen, L. Xu, and L. Lin. Structure-preserving image super-resolution via contextualized multi-task learning. *IEEE Transactions on Multimedia*, 2017. 2
- [43] A. Singh, F. Porikli, and N. Ahuja. Super-resolving noisy images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2846–2853, 2014. 3
- [44] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3147–3155, 2017. 2, 7
- [45] Y. Tai, J. Yang, X. Liu, and C. Xu. Memnet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision*, pages 4539–4547, 2017. 2
- [46] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125, July 2017. 2
- [47] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126, 2014. 3
- [48] A. Vedaldi and K. Lenc. MatConvNet: Convolutional neural networks for matlab. In *ACM Conference on Multimedia Conference*, pages 689–692, 2015. 6
- [49] Waifu2x. Image super-resolution for anime-style art using deep convolutional neural networks. <http://waifu2x.udp.jp/>. 8
- [50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6
- [51] C.-Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In *European Conference on Computer Vision*, pages 372–386, 2014. 1, 3
- [52] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. 1, 3
- [53] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, and S. Yan. Deep edge guided recurrent residual learning for image super-resolution. *IEEE Transactions on Image Processing*, 26(12):5895–5907, 2017. 2
- [54] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730, 2010. 7
- [55] K. Zhang, X. Zhou, H. Zhang, and W. Zuo. Revisiting single image super-resolution under internet environment: blur kernels and reconstruction algorithms. In *Pacific Rim Conference on Multimedia*, pages 677–687, 2015. 3
- [56] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, pages 3142–3155, 2017. 2, 4, 6, 7
- [57] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep CNN denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, July 2017. 1, 2, 3, 6, 7
- [58] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2