



Deep Fourier Kernel Exploitation in Blind Image Super-Resolution

Yu Fu, Xiaoyun Zhang^(✉), Yixuan Huang, Ya Zhang, and Yanfeng Wang

Cooperative Medianet Innovation Center, Shanghai Jiao Tong University,
Shanghai, China

{fyuu11,xiaoyun.zhang,huangyixuan,ya_zhang,wangyanfeng}@sjtu.edu.cn

Abstract. Blind image super-resolution (SR) has achieved great progress through estimating and utilizing blur kernels. However, current predefined dimension-stretching strategy based methods trivially concatenate or modulate the vectorized blur kernel with the low-resolution image, resulting in raw blur kernels under-utilized and also limiting generalization. This paper proposes a deep Fourier kernel exploitation framework to model the explicit correlation between raw blur kernels and images without dimensionality reduction. Specifically, based on the acknowledged degradation model, we decouple the effects of downsampling and the blur kernel, and reverse them by the upsampling and deconvolution modules accordingly, via introducing a transitional SR image. Then we design a novel Kernel Fast Fourier Convolution (KFFC) to filter the image feature of the transitional image with the raw blur kernel in the frequency domain. Extensive experiments show that our methods achieve favorable and robust results.

Keywords: Blind image super-resolution · Blur kernel · Fourier convolution network

1 Introduction

Single Image Super-Resolution (SISR) aims to reconstruct the High-Resolution (HR) image from the given Low-Resolution (LR) counterpart. In past decades, it has been widely used in surveillance imaging, astronomy imaging, and medical imaging. As SISR is an extremely ill-posed task, early learning-based methods [8, 15, 37] make plain exploration, *i.e.* assume that the blur kernel in the degradation is predefined/known (e.g. Bicubic kernel). For more powerful generalization [3, 27], the blind SR task [5, 10, 21, 25, 30, 36], which handles LR images underlying unknown blur kernels, draws increasing attention. Most of these methods focus on accurate blur kernel estimation [3, 18], fully kernel exploitation [25, 33, 36], or combining both [10, 12]. This work focuses on fully exploiting the blur kernel in the blind SR task.

Given explicit raw blur kernels, the core of blind SR lies in how to properly introduce and fully utilize the information in them. The existing methods adopt

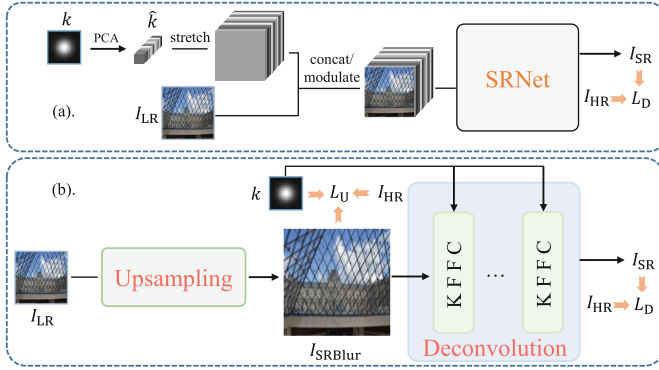


Fig. 1. Framework comparison on kernel exploitation super-resolution task, assuming the blur kernel is estimated or handcrafted. (a) Existing dimensionality reduction based framework; (b) Our deep Fourier kernel exploitation framework.

dimensionality reduction manners, ignoring the physical meaning of the blur kernel. As shown in Fig. 1(a), they use a predefined principal component analysis (PCA) to project the raw blur kernel into a discriminative domain vector \hat{k} , then \hat{k} is stretched and trivially concatenated [25, 36] or modulated [10, 12, 30, 33] with the LR image. Such dimensionality reduction based methods suffer from two major flaws. (1) The PCA matrix inevitably discards part of the blur kernel information, and the predefined strategy severely limits generalization, since it is helpless when dealing with unseen kernels [7, 19]. (2) Simply relying on the SR network to implicitly model the correlation between the blur kernel and the image makes the blur kernel under-utilized, and also lacks interpretability.

To tackle these problems and fully exploit the blur kernel information, without predefined dimensionality reduction, we revisit the degradation model, to decouple the effects of the raw blur kernel and the downsampling. Specifically, in the degradation model, we observe that the blur kernel and downsampling are independent. Hence, we propose to decouple the blind SR task into **upsampling** and **deconvolution** modules, via introducing a transitional SR image. As shown in Fig. 1(b), the upsampling module is only fed with the LR image, to produce the transitional SR image I_{SRBlur} , which has the same size as the HR image, but with coarse details. I_{SRBlur} is the approximation of convolving the HR image with the blur kernel k , thus it eliminates the effects of downsampling in the LR image, and naturally avoids poor generalization caused by introducing the predefined PCA. Given the transitional SR image, the deconvolution module aims to reverse the effects of k . To fully explore the kernel information, we here model the explicit correlation between the raw blur kernel and the image, instead of concatenating or modulation. We design a novel Kernel Fast Fourier Convolution (KFFC) to achieve efficient filtering via Fourier transform, considering that the convolving can be well reversed in the frequency domain with good interpretability [9]. Specifically, the proposed KFFC explicitly filters

the image features of I_{SRBlur} through k in the frequency domain. We train the whole framework end-to-end and conduct extensive experiments and comprehensive ablation studies in both non-blind and blind settings to evaluate the state-of-the-art performance.

2 Related Work

2.1 Kernel Exploitation

Kernel exploitation in blind super-resolution (SR) aims to reconstruct super-resolved (SR) images from given low-resolution(LR) images with known degradation, which is also called the non-blind SR setting. Generally speaking, there are two definitions of the non-blind SR setting in the literature.

The one is training an SR model for one single degradation blur kernel and evaluating the model with the synthetic dataset based on the predefined kernel [2, 8, 15, 31, 37]. For instance, Dong et al. [8] proposed the first learning-based CNN model for image super-resolution. Kim et al. [15] and Zhang et al. [37] employed a very deep CNN and attention mechanism, which improved the SISR performance by a large margin. Although they show superiority compared with traditional interpolation-based and model-based methods, all of them assume the LR images are degraded from the HR images through bicubic downsampling, therefore, they construct paired LR-HR training dataset through bicubic downsampling and conduct experiments with the predefined bicubic kernel for evaluation. One can see that degradation in a real-world scenario is complicated and unknown which can not be described by a simple bicubic kernel, and the models trained with a bicubic dataset will suffer from severe performance drop when the degradation is different from the bicubic kernel.

The other is that reconstruct the SR image from both the LR image and the corresponding blur kernel [10, 25, 33, 36]. Following the acknowledged degradation model 1, these types of kernel exploitation methods focus on fully utilizing blur kernel for reconstructing SR images. Compared with former methods which are trained with a paired bicubic dataset, these types of methods take blur kernels into account and obtain better performance than the former methods when given an estimated or handcrafted blur kernel. Nevertheless, these kernel exploitation methods introduce the blur kernel through a dimension-stretching strategy, which underutilizes the blur kernel and limits the generalization. Specifically, they use a predefined principal component analysis (PCA) matrix to project the raw blur kernel into a discriminative domain vector and model the correlation between the blur kernel and the image through direct concatenation or modulation. In this paper, we focus on modeling the explicit correlation of the raw blur kernel and the image, which can remedy the kernel underutilization and limited generalization problems.

2.2 Blind SR

To tackle the blind SR task, which reconstructs SR image from only LR image with unknown degradation, a lot of works are proposed and achieve encouraging

performance [3, 7, 10, 12, 14, 18, 28]. Most blind SR methods follow a two-step manner. Firstly, estimating the blur kernel from the given LR image. Bell-Kligler et al. [3] estimates the degradation kernel by utilizing an internal generative adversarial network (GAN) on a single LR image. Tao et al. [28] estimate the blur kernel in the frequency domain. Liang et al. [18] and Xiao et al. [32] try to estimate a spatial variant blur kernel for real-world images.

Given the estimated blur kernel and the LR image, the core of blind SR lies in utilizing blur kernels. As mentioned before, the kernel exploitation approaches can solve the blind SR task by introducing the estimated blur kernel. There also exist two types of methods that can combine the kernel estimating network. The one is to synthesis paired training datasets based on the estimated blur kernels, and then train a new non-blind SR model to enhance the performance on real datasets [3, 28]. However, one can see that these methods are time-consuming for training a new model and indirectly use the estimated blur kernel.

The other is that introduce the estimated blur kernel to the kernel exploiting SR network directly [10, 25, 33, 36]. Furthermore, some approaches tried to solve the two-step blind SR task in an end-to-end manner [7, 10, 12, 30] and proposed some new training strategies to enhance the performance. For example, Gu et al. [10] proposed an iterative mechanism to refine the estimated blur kernel based on the super-resolved image. Luo et al. [12] alternatively optimized the blur kernel estimation and kernel exploitation. It is worth pointing out that the proposed method focuses on fully exploiting the blur kernel for blind SR and could be combined with most kernel estimation and training strategies.

2.3 Other SR Methods

Beyond these methods aforementioned, there are other related methods that do not explicitly estimate blur kernels, such as degradation-modeling SR [20]. They try to introduce more types of degradation, like jpeg compression, and design a sophisticated degradation model. Then they synthesize paired datasets for training, the whole process is time-consuming and still tends to generate over-smoothed results when meeting slight degradation [20]. Unpaired SR [13, 21] learns the degradation in real-world images based on the cycleGAN [38], which usually suffer from unstable training and pixel misalignment problems. Zero-shot SR [27] utilizes the similarity of patches across scales to reconstruct SR image from single LR image input, which has in-born problems that it can not learn a good prior of HR datasets.

3 Methodology

3.1 Blur Kernels in Blind Super-Resolution Task

Given the low-resolution image I_{LR} and the corresponding blur kernel k , the core of the blind SR task lies in fully exploiting the blur kernel information

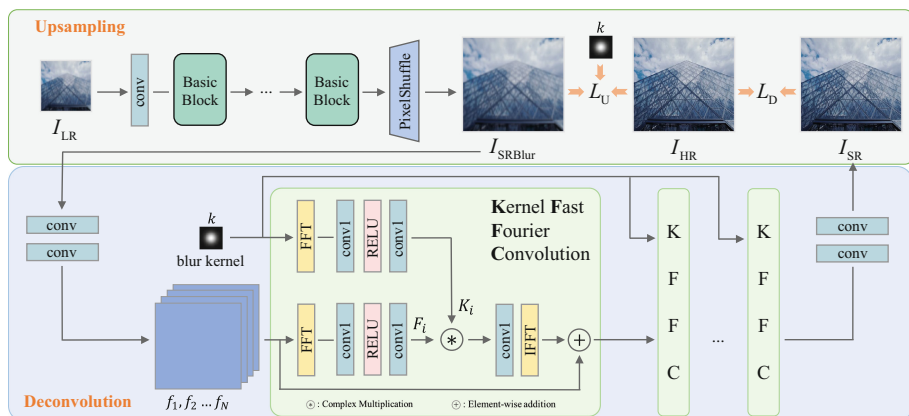


Fig. 2. Framework Overview. Given the LR image and the corresponding blur kernel, we decouple the kernel exploitation task into upsampling and deconvolution modules based on the degradation model. The **Upsampling** module reconstructs the SRBlur image under the constraints of the HR image and the blur kernel. The **Deconvolution** module is fed with the resultant image SRBlur and the blur kernel, and recovers the super-resolved image through our Kernel Fast Fourier Convolution which can model the explicit correlation between the blur kernel and the SRBlur image in the frequency domain. The whole framework is trained end-to-end.

to reconstruct the super-resolved image I_{SR} , compared with general SR tasks. Firstly, the acknowledged degradation model can be formulated as:

$$I_{LR} = (I_{HR} * k) \downarrow_{\alpha} + n, \quad (1)$$

where $*$ is a convolution operation, \downarrow_{α} is a downsampling operation with the scale factor α , and n is the additive Gaussian noise. Following [10, 12, 25], we assume noise can be reduced by [34]. And the degradation model can be divided into: convolving ($I_{HR} * k$) and independent downsampling \downarrow_{α} . The former means information blending of I_{HR} via k , and the latter denotes irreversible information lost, which is irrelevant with k . Thus, k only has a direct effect on the HR scale.

3.2 Deep Fourier Kernel Exploitation Framework

Motivation: Although existing dimension-stretching strategy based kernel exploitation methods in blind SR show superiority, they still pose two main flaws.

Firstly, the used PCA matrix inevitably discards part of the blur kernel information, and the predefined strategy severely limits generalization. To solve this, without using predefined PCA, we introduce the raw blur kernel to the SR network. The degradation model in Sect. 3.1 shows the independent downsampling impedes applying the raw blur kernel to the LR image directly. We hence decouple the effects of downsampling and the blur kernel, and reverse them by the upsampling and deconvolution modules accordingly. The two modules are bridged through a transitional SR image.

Secondly, simply relying on the SR network to implicitly interact the blur kernel and the image, under-utilizes the blur kernel, and also lacks interpretability. To address these problems, in the deconvolution module, we model the explicit correlation between the transitional SR images and raw blur kernels through the proposed module Kernel Fast Fourier Convolution (KFFC), to filter the image feature of the transitional image with the blur kernel in the frequency domain.

Therefore, as shown in Fig. 2, the raw blur kernel is explicitly introduced to the SR network through the decoupling framework and the KFFC module. Firstly, the LR image I_{LR} is fed into the upsampling module and generates transitional SR image I_{SRBlur} . Then, the deconvolution module reconstructs SR image I_{SR} from I_{SRBlur} and the raw blur kernel k through the proposed KFFC.

Upsampling Module is designed to generate the transitional HR-size image I_{SRBlur} , which is an approximation of convolving the HR image and the blur kernel k . Compared to directly recovering the original SR image I_{SR} from I_{LR} , the objective of the upsampling module is easier. Hence, for simplicity, we adopt the trivial SISR network as the module architecture. Specifically, we use several convolution layers to extract features from I_{LR} ; then forward features to several cascaded basic blocks, which follows Residual in Residual Dense Block (RRDB) [31], to enhance the non-linear ability; and finally, upsample features via the PixelShuffle layer [26]. Note that, the architecture of upsampling module is flexible to the existing SISR network.

To supervise the upsampling module \mathcal{U} , we first calculate $(I_{HR} * k)$ as the ground-truth of I_{SRBlur} ; then use the L_1 constraint to achieve reverse downsampling:

$$L_U = L_1(I_{HR} * k, I_{SRBlur}) = L_1(I_{HR} * k, \mathcal{U}(I_{LR})) \quad (2)$$

Deconvolution Module aims to reverse $(I_{HR} * k)$, *i.e.* generate the SR image I_{SR} , given the transitional image I_{SRBlur} and blur kernel k . A trivial way is to concatenate or modulate k into the SR network like existing methods, we experimentally find that it only sharpens images without reconstructing high-frequency details, thus under-utilizes blur kernel information (see Sect. 4.4). To solve this problem, we propose to model an explicit correlation between k and I_{SRBlur} . Since I_{SRBlur} is the approximation of convolving I_{HR} and k , we can regard deconvolution as deblurring I_{SRBlur} given k . Considering that, there exists a huge gap between I_{SRBlur} and ground-truth $(I_{HR} * k)$, directly restoring SR images by the inverse filter is impracticable. We hence regard the gap as independent noise, which is termed as $I_{SRBlur} = (I_{HR} * k) + n$; then, treat the task of the deconvolution module as deblurring images with noise.

Like inverse filters and Wiener filters, considering that convolution can be well reversed in the frequency domain, which has good interpretability [9], we here handle the deconvolution task via Fourier Transform. Inspired by Fast Fourier Convolution (FFC) [6] which is designed to capture a large reception field in shallow layers of the network for high-level tasks, we design a module called Kernel Fast Fourier Convolution (KFFC), to filter the transitional image with corresponding blur kernel k in the frequency domain.

Concretely, as shown in Fig. 2, we first use some CNN layers to extract the image feature of I_{SRBlur} , denoted as $\{f_i\}_{i=1}^N$; and then transform both feature maps $\{f_i\}_{i=1}^N$ and the blur kernel k into frequency domain through fast Fourier transform \mathcal{F} ; next adopt two individual compilations of 1×1 convolution and Leaky ReLU $\mathcal{T}_F, \mathcal{T}_K$ to achieve non-linearly transformation. Based on the property of Fourier Transform, we can conduct filtering through complex multiplication in the frequency domain. Sequentially, we empirically use an extra 1×1 convolution to refine the filtered feature. Eventually, the filtered feature maps are converted back to the spatial domain through Inverse Fourier Transform \mathcal{F}^{-1} . The whole procedure of KFFC can be formulated as follows:

$$\begin{aligned} \text{KFFC}(f_i, k) = & \mathcal{F}^{-1}[(\Re(\mathbf{F}_i) \odot \Re(\mathbf{K}_i) - \Im(\mathbf{F}_i) \odot \Im(\mathbf{K}_i)) \\ & + i(\Im(\mathbf{F}_i) \odot \Re(\mathbf{K}_i) + \Re(\mathbf{F}_i) \odot \Im(\mathbf{K}_i))] \end{aligned} \quad (3)$$

where $\mathbf{F}_i = \mathcal{T}_F(\mathcal{F}(f_i))$, $\mathbf{K}_i = \mathcal{T}_K(\mathcal{F}(k))$, \odot denotes the Hadamard product. \Re and \Im denotes the real and the imaginary part respectively. Note that we adopt real Fourier Transform to simplify computation because applying Fourier Transform on real signal is perfectly conjugate symmetric. To enhance performance, we also cascade multiple KFFC structures, then use two convolution layers to output the SR image I_{SR} .

To supervise the deconvolution module \mathcal{D} , we leverage the L_1 constraint, which can be formulated as:

$$L_D = L_1(I_{\text{HR}}, I_{\text{SR}}) = L_1(I_{\text{HR}}, \mathcal{D}(I_{\text{SRBlur}}, k)) \quad (4)$$

The whole framework is end-to-end trained with a total loss of $L_D + \lambda L_U$, and λ is a weight term which is set as 1 in our experiments.

4 Experiments

4.1 Datasets and Implementations

The training data is synthesized based on Eq. 1. 3450 2K HR images are collected from DIV2K [1] and Flickr2K [29] for training. We use isotropic blur kernels following [10, 12, 33, 36]. The kernel width is uniformly sampled in $[0.2, 2.0]$, $[0.2, 3.0]$ and $[0.2, 4.0]$ for scale factors 2, 3, 4, respectively. The kernel size is fixed to 21×21 . We also augment data by randomly horizontal flipping, vertical flipping, and 90° rotating.

Implementation Details: For all experiments, we use 6 RRDB blocks [31] in the upsampling module and 7 KFFCs in the deconvolution module with 64 channels. During training, we crop and degrade HR images to 64×64 LR patches for all scale factors. For optimization, we use Adam [16] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a mini-batch size of 32. The learning rate is initialized to 4×10^{-4} and decayed by half at 1×10^5 iterations. We evaluate on five benchmarks: Set5 [4], Set14 [35], BSDS100 [22], Urban100 [11] and Manga109 [23]. All models are trained on RGB space, PSNR and SSIM are evaluated on Y channel of transformed YCrCb space.

Table 1. Quantitative performance comparison of the proposed method with other SOTAs on non-blind SR setting. The comparison is conducted using three different isotropic Gaussian blur kernels on Set5, Set14, and BSD100 datasets. We provide the ground-truth kernel(GTker) for all the listed methods in this table, as the original setting in SRCNN-CAB [25], SRMDNF [36] and UDVD [33]. As for SFTMD [10] and ZSSR [27], we also compare them by giving the GTker although they can solve the blind SR task through a kernel estimation network or bicubic downsampling. The best two results are highlighted in red and blue colors.

Method	Kernel Width	Set5 [4]			Set14 [35]			BSD100 [22]		
		x2	x3	x4	x2	x3	x4	x2	x3	x4
GTker+ZSSR [27]	0.2	34.94	29.29	28.87	31.04	28.05	27.15	31.42	28.24	26.68
SRCNN-CAB [25]		33.27	31.03	29.31	30.29	28.29	26.91	28.98	27.65	25.51
SRMDNF [36]		37.79	34.13	31.96	33.33	30.04	28.35	32.05	28.97	27.49
SFTMD [10]		38.00	34.57	32.39	33.68	30.47	28.77	32.09	29.09	27.58
UDVD [33]		38.01	34.49	32.31	33.64	30.44	28.78	32.19	29.18	27.70
Ours		38.17	34.63	32.45	33.82	30.52	28.85	32.33	29.23	27.74
GTker+ZSSR [27]	1.3	33.37	28.67	27.44	31.31	27.34	26.15	30.31	27.30	25.95
SRCNN-CAB [25]		33.42	31.14	29.50	30.51	28.34	27.02	29.02	27.91	25.66
SRMDNF [36]		37.44	34.17	32.00	33.20	30.08	28.42	31.98	29.03	27.53
SFTMD [10]		37.46	34.53	32.41	33.39	30.55	28.82	32.06	29.15	27.64
UDVD [33]		37.36	34.52	32.37	33.39	30.50	28.85	32.00	29.23	27.75
Ours		37.77	34.71	32.56	33.70	30.58	28.93	33.20	29.32	27.81
GTker+ZSSR [27]	2.6	29.89	27.80	27.69	27.72	26.42	26.06	27.32	26.47	25.92
SRCNN-CAB [25]		32.21	30.82	28.81	29.74	27.83	26.15	28.35	26.63	25.13
SRMDNF [36]		34.12	33.02	31.77	30.25	29.33	28.26	29.23	28.35	27.43
SFTMD [10]		34.27	33.22	32.05	30.38	29.63	28.55	29.35	28.41	27.47
UDVD [33]		33.74	33.15	31.99	30.08	29.58	28.55	28.93	28.49	27.55
Ours		33.67	33.31	32.18	29.92	29.64	28.68	28.81	28.48	27.64

4.2 Experiments on Non-Blind Setting

We first evaluate the performance of the proposed method on the non-blind SR setting, which reconstructs the super-resolved image under the given low-resolution image and a corresponding known blur kernel. Following [10, 33, 36], we only consider isotropic Gaussian blur for simplicity, the kernel widths are set to 0.2, 1.3, 2.6, and the kernel sizes are 21×21 for all scale factors. Table 1 makes a comparison with SOTA methods under different kernel settings. We provide ground-truth kernels for all the listed methods in Table 1, as the original setting in SRCNN-CAB [8], SRMDNF [36], and UDVD [33]. As for SFTMD [10] and ZSSR [27], we also compare them by giving the ground-truth kernel, although they solve the blind SR task through a kernel estimation network [10] or bicubic downsampling [27].

The results in Table 1 show that our method performs best on most benchmarks, regardless of the degradation degree and scale factors. We surpass all competitors in most experimental settings and have a large gain of 0.31 dB for Set14 with kernel width 1.3. Interestingly, as the scale factor increases, the gain

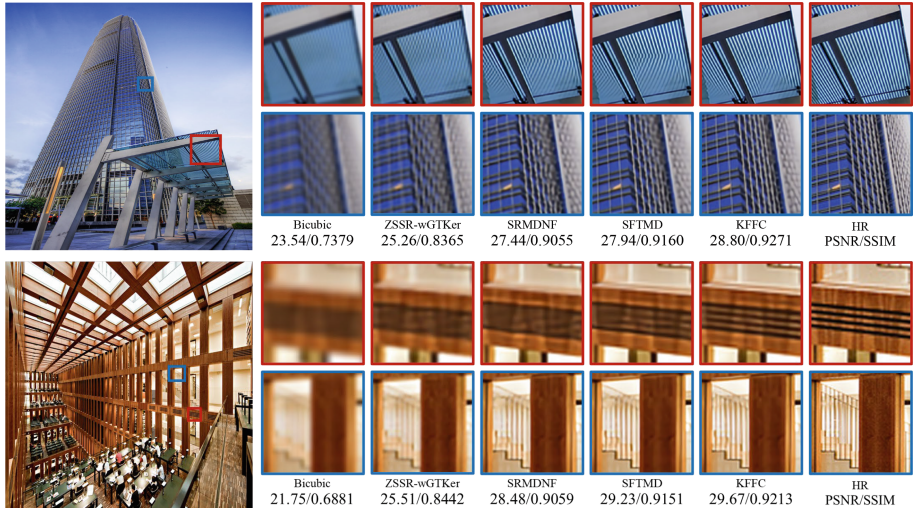


Fig. 3. Visual results of *img046*, *img049* from Urban100, for different methods in non-blind setting. The SR factor is 2 and kernel width is 1.3.

of the proposed method gradually declines. We conjecture this is because the increasing scale factor reduces the effectiveness of blur kernels. In other words, the information loss in the downsampling process impedes the reconstruction. Figure 3 also shows visual comparisons. Our method yields better results on the recurring and regular textures such as the fine stripes, and reconstructs high quality fine details with fewer artifacts. All the above results reveal the superiority of our method quantitatively and qualitatively.

4.3 Experiments on Blind Setting

We also conduct blind SR settings to evaluate the effectiveness of the proposed method, which reconstruct super-resolved image under unknown degradation. Since determined blur kernels are needed for reasonable comparison, following in [10, 12, 14], we uniformly sample 8 kernels from range [1.8, 3.2], [1.35, 2.40] and [0.8, 1.6] for scale factor $\times 4$, $\times 3$, and $\times 2$ respectively, which is also referred to *Gaussian8*. We compare our method with other SOTAs blind SR approaches including ZSSR [27] (with bicubic kernel), IKC [10], DAN [12], AdaTarget [14]. Following [10, 12], we also conduct comparisons with CARN [2] and its variants of performing the deblurring method before and after CARN. For all compared methods, we use their official implementation and pre-trained model except special remarks. It is worth pointing out that owing to our proposed method assuming the blur kernel is known, we conduct a comparison in a blind SR setting through integrating our method into IKC [10] and DAN [12]. Specifically, we replace the kernel exploitation part in the original methods with our proposed method and preserve the original training strategy such as iterative refining.

Table 2. Quantitative comparison of the proposed method with other SOTAs on various datasets in blind SR setting. The comparison is conducted using *Gaussian8* kernels on five benchmark datasets. We integrate our proposed method into IKC [10] and DAN [12], which means replacing the kernel exploitation part of the original network and preserving their training strategy. The best two results are highlighted in red and blue colors respectively.

Method	Scale	Set5 [4]		Set14 [35]		BSD100 [22]		Urban100 [11]		Manga109 [23]	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic		28.82	0.8577	26.02	0.7634	25.92	0.7310	23.14	0.7258	25.60	0.8498
CARN [2]		30.99	0.8779	28.10	0.7879	26.78	0.7286	25.27	0.7630	26.86	0.8606
Bicubic+ZSSR [27]		31.08	0.8786	28.35	0.7933	27.92	0.7632	25.25	0.7618	28.05	0.8769
[24]+CARN [2]		24.20	0.7496	21.12	0.6170	22.69	0.6471	18.89	0.5895	21.54	0.7946
CARN [2]+ [24]	x2	31.27	0.8974	29.03	0.8267	28.72	0.8033	25.62	0.7981	29.58	0.9134
IKC [10]		37.19	0.9526	32.94	0.9024	31.51	0.8790	29.85	0.8928	36.93	0.9667
DAN [12]		37.34	0.9526	33.08	0.9041	31.76	0.8858	30.60	0.9060	37.23	0.9710
DASR [30]		37.00	0.9508	32.61	0.8958	31.59	0.8813	30.26	0.9015	36.20	0.9686
IKC [10] + Ours		37.60	0.9562	33.31	0.9105	31.84	0.8880	30.66	0.9090	38.07	0.9718
DAN [12] + Ours	37.55	0.9547	33.26	0.9074	31.96	0.8897	31.20	0.9144	37.83	0.9726	
Bicubic		26.21	0.7766	24.01	0.6662	24.25	0.6356	21.39	0.6203	22.98	0.7576
CARN [2]		27.26	0.7855	25.06	0.6676	25.85	0.6566	22.67	0.6323	23.85	0.7620
Bicubic+ZSSR [27]		28.25	0.7989	26.15	0.6942	26.06	0.6633	23.26	0.6534	25.19	0.7914
[24]+CARN [2]		19.05	0.5226	17.61	0.4558	20.51	0.5331	16.72	0.5895	18.38	0.6118
CARN [2]+ [24]	x3	30.31	0.8562	27.57	0.7531	27.14	0.7152	24.45	0.7241	27.67	0.8592
IKC [10]		33.06	0.9146	29.38	0.8233	28.53	0.7899	24.43	0.8302	32.43	0.9316
DAN [12]		34.04	0.9199	30.09	0.8287	28.94	0.7919	27.65	0.8352	33.16	0.9382
DASR [30]		33.53	0.9150	29.64	0.8143	28.64	0.7825	27.26	0.8269	32.05	0.9290
IKC [10] + Ours		33.51	0.9170	30.11	0.8313	28.84	0.7931	27.53	0.8344	32.90	0.9369
DAN [12] + Ours	34.11	0.9205	30.15	0.8303	28.98	0.7925	27.78	0.8380	33.20	0.9389	
Bicubic		24.57	0.7108	22.79	0.6032	23.29	0.5786	20.35	0.5532	21.50	0.6933
CARN [2]		26.57	0.7420	24.62	0.6226	24.79	0.5963	22.17	0.5865	21.85	0.6834
Bicubic+ZSSR [27]		26.45	0.7279	24.78	0.6268	24.97	0.5989	22.11	0.5805	23.53	0.7240
[24]+CARN [2]		18.10	0.4843	16.59	0.3994	18.46	0.4481	15.47	0.3872	16.78	0.5371
CARN [2]+ [24]	x4	28.69	0.8092	26.40	0.6926	26.10	0.6528	23.46	0.6597	25.84	0.8035
IKC [10]		31.67	0.8829	28.31	0.7643	27.37	0.7192	25.33	0.7504	28.91	0.8782
DAN [12]		31.89	0.8864	28.42	0.7687	27.51	0.7248	25.86	0.7721	30.50	0.9037
AdaTarget [14]		31.58	0.8814	28.14	0.7626	27.43	0.7216	25.72	0.7683	29.97	0.8955
DASR [30]		31.55	0.8822	28.13	0.7597	27.36	0.7186	25.35	0.7538	29.80	0.8929
IKC [10] + Ours	31.96	0.8858	28.35	0.7661	27.39	0.7210	25.68	0.7634	30.29	0.8994	
DAN [12] + Ours	31.98	0.8877	28.42	0.7676	27.54	0.7249	25.90	0.7729	30.41	0.9016	

Table 2 shows the PSNR and SSIM results on five widely-used benchmark datasets. As one can see, the traditional interpolation-based method and non-blind SR methods suffer severe performance drop when the degradation is unknown or different from the predefined bicubic kernel. Although ZSSR also assumes the bicubic kernel, it achieves better performance than CARN because ZSSR trains a specific network for each single test image by utilizing the internal patch recurrence. However, ZSSR has in-born flaws: it can not learn a good prior of HR datasets because the train samplings are sampled from one single LR

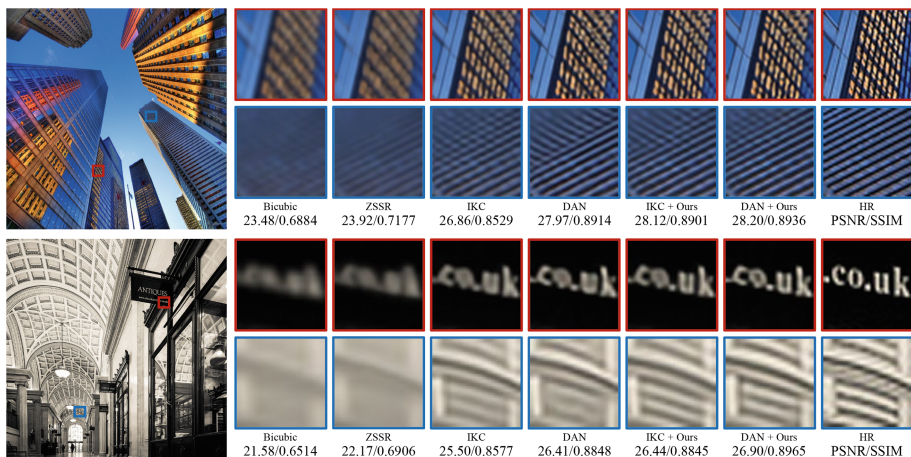


Fig. 4. Visual results of *img012*, *img083* from Urban100, for different methods in blind setting. The σ of blur kernel is 1.25.

image. By taking the degradation kernel into account, the performance of CARN is largely improved when combined with a deblurring method. IKC and DAN are both two-step blind SR methods and can largely improve the results. IKC refined the estimated kernel through an iterative manner based on the last estimated kernel and super-resolved image. Furthermore, DAN optimizes the blur kernel estimation and super-resolution alternatively, which obtains better performance than IKC. Nevertheless, both of them adopt a kernel-stretching strategy for introducing the estimated blur kernel to the reconstruction network, which under-utilizes the blur kernel and limits generalization. Therefore, we integrate our proposed method into IKC and DAN to remedy the kernel underutilizing problem. As one can see, when combined with our proposed model, both the performance of IKC and DAN improve by a large margin. Our method leads to the best performance over most datasets, which proves that our methods achieve better kernel utilization. The qualitative results are shown in Fig. 4, which illustrates that our method can reconstruct sharp and clear repetitive texture or edges with fewer artifacts.

4.4 Ablation Study

We conduct ablation studies on vital components of the proposed method: the decoupling framework and the proposed KFFC module. The quantitative results are listed in Table 3. Following [10, 36], we first project the raw blur kernel k of size $p \times p$ to a t -dimensional kernel vector \hat{k} through a predefined PCA matrix. Note that p is 21 and t is set to 15 by default. And then we introduce \hat{k} to the reconstruction network and interact the blur kernel and the LR image by directly concatenating or modulating, which is denoted as Base_Cat and Base_Mod respectively. We adopt our proposed decoupling architecture but still concatenate or modulate corresponding \hat{k} in the deconvolution module for

Table 3. Ablation study results of network architectures. Results are reported in non-blind setting and the test dataset is Urban100. With both the decouple architecture and KFFC module, our method performs best.

Method	Decouple	Cat	Mod	KFFC	PSNR \uparrow	SSIM \uparrow
Base_Cat		✓			30.84	0.9111
Base_Mod			✓		31.61	0.9189
Dec_Cat	✓	✓			31.64	0.9193
Dec_Mod	✓		✓		31.68	0.9196
Ours _{Sw/o}	✓			✓	29.98	0.9136
Ours	✓			✓	31.96	0.9233

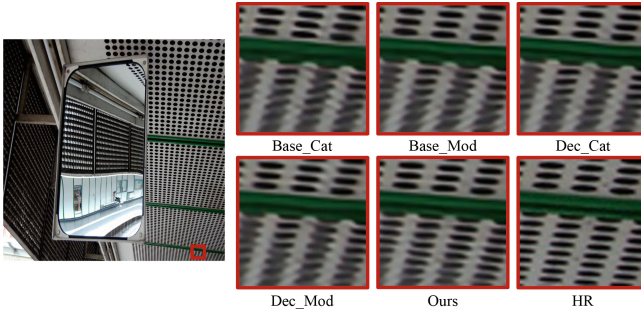


Fig. 5. Ablation visualization on Urban100. Our complete method reconstructs more clear and correct texture.

introducing the blur kernel, and term these methods as Dec_Cat and Dec_Mod respectively in Table 3. As one can see, compared to introducing the blur kernel in the LR scale, the decoupling architecture improves the performance quantitatively, which is consistent with our analysis that the blur kernel should only have an impact on the high-resolution scale based on the degradation model Sect. 3.1. Furthermore, one can see modeling the explicit correlation between the raw blur kernel and images in the frequency domain based on the proposed KFFC module gains significant improvements by about 0.35db by a large margin quantitatively, which demonstrates that the proposed KFFC utilizes the blur kernel more effectively. Moreover, we also evaluate our method without the blur kernel (denoted as Ours w/o). Its inferior is strong evidence that gains come from both the proposed method and the utilization of blur kernels. Qualitative results are shown in Fig. 5, our proposed method reconstructs more clear and correct texture when enabling all the vital components.

4.5 Generalization Evaluation

As mentioned in Sect. 3.2, the predefined PCA matrix will limit the generalization. Specifically, the previous dimension-stretching-strategy based kernel

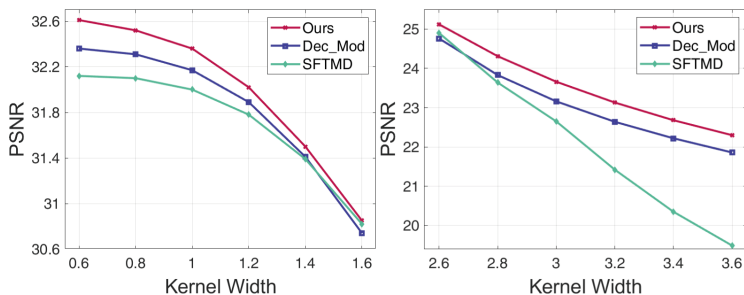


Fig. 6. In-distribution (left) and out-of-distribution (right) degradation results. Our method performs robustly under both settings.

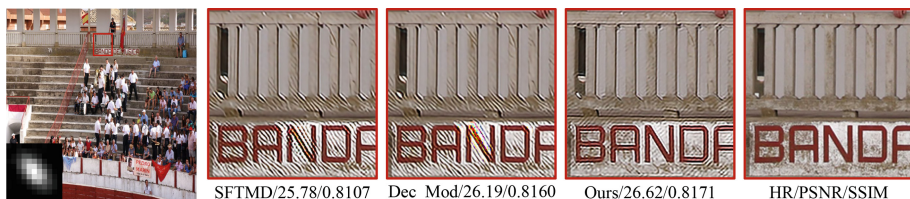


Fig. 7. Visualization on unseen blur kernel. The quantitative results of the whole benchmark DIV2KRRK [3] are listed below directly. Our method generates visual pleasant result with less artifacts.

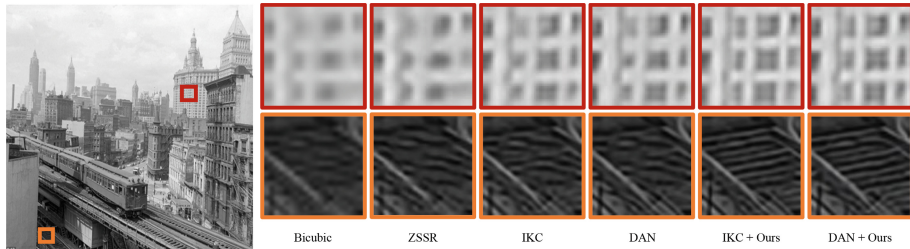


Fig. 8. SISR performance comparison of different methods with SR factor 2 on a real historical image [17].

exploitation methods [10, 12, 25, 33, 36] project the raw blur kernel into a kernel vector through a predefined PCA matrix. When the predefined PCA matrix meets unknown raw blur kernels, it will generate a wrong kernel vector, which means the whole SR network needs to be retrained again. Therefore, we conduct two experiments to evaluate the generalization of the proposed method: *i.e.*, out-of-distribution degradation, and unseen blur kernels. At last, we conduct an experiment on a real degradation dataset where the ground-truth HR image and the blur kernel are not available.

Out-of-Distribution Degradation: During training, we shrink the kernel widths of seen blur kernels from range $[0, 2, 2]$, and then evaluate the model on two parameter ranges of blur kernel widths. The in-distribution kernel width ranges from 0.6 to 1.6, while the out-of-distribution kernel width ranges from 2.6 to 3.6. Figure 6 shows quantitative results. The decoupling architecture limits the effect of the blur kernel on the HR scale, it alleviates the influence of the wrong kernel vector. As one can see, for both in and out distribution settings, our method performs better than other methods because we utilize the raw blur kernel and explicitly interact with raw blur kernels and images in the frequency domain.

Unseen Blur Kernels: We investigate the proposed method on DIV2KRRK [3]. It sets the blur kernel to an anisotropic Gaussian kernel, randomly samples the kernel width from $[0.6, 5]$, and adopts uniform multiplicative noise. As shown in Fig. 7, all the performance suffers from severe performance drop because blur kernels in DIV2KRRK are different from the isotropic Gaussian kernel in training settings. Nonetheless, our method still achieves better performance than existing methods both quantitatively and qualitatively. Here, quantitative results of the total dataset are recorded under the image directly.

Performance on Real Degradation: To further demonstrate the effectiveness of the proposed method. We conduct an experiment on a real degradation dataset where the ground-truth image and the blur kernel are not available. The qualitative comparison is shown in Fig. 8. Compared with other blind SR methods, our proposed method enhance the performance of IKC and DAN, and recover clear and sharp texture.

5 Conclusion

In this paper, we demonstrate the flaws in existing blind SR methods. Current blind SR methods adopt a two-step framework including kernel estimation and kernel exploitation. However, they adopt a dimension-stretching strategy for kernel exploitation when the blur kernel is estimated. Specifically, the estimated raw blur kernel will be projected into a kernel vector through a predefined PCA matrix, and they interact the blur kernel with LR images by direct concatenation or modulation, which underutilizes the raw blur kernel and limit the generalization. To tackle these problems, we decouple the kernel exploitation task into the upsampling and deconvolution modules. To fully explore the blur kernel information, we also design a feature-based kernel fast Fourier convolution, for filtering image features in the frequency domain. Comprehensive experiments and ablation studies in both non-blind and blind settings are conducted to show our gratifying performance.

Acknowledgement. This work is supported by National Natural Science Foundation of China (62271308), STCSM (No. 22511105700, No. 18DZ2270700), 111 plan (No. BP0719010), and State Key Laboratory of UHD Video and Audio Production and Presentation.

References

1. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1122–1131 (2017)
2. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision, pp. 252–268 (2018)
3. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-GAN. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
4. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the British Machine Vision Conference, pp. 135.1–135.10 (2012)
5. Chen, H., et al.: Real-world single image super-resolution: a brief review. *Inf. Fusion* **79**, 124–145 (2021)
6. Chi, L., Jiang, B., Mu, Y.: Fast Fourier convolution. In: Advances in Neural Information Processing Systems, pp. 4479–4488 (2020)
7. Cornillere, V., Djelouah, A., Yifan, W., Sorkine-Hornung, O., Schroers, C.: Blind image super-resolution with spatially variant degradations. *ACM Trans. Graph.* **38**(6), 1–13 (2019)
8. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2015)
9. Dong, J., Roth, S., Schiele, B.: Deep wiener deconvolution: wiener meets deep learning for image deblurring. In: Advances in Neural Information Processing Systems, vol. 33, pp. 1048–1059 (2020)
10. Gu, J., Lu, H., Zuo, W., Dong, C.: Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1604–1613 (2019)
11. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206 (2015)
12. Huang, Y., Li, S., Wang, L., Tan, T., et al.: Unfolding the alternating optimization for blind super resolution. In: Advances in Neural Information Processing Systems, vol. 33, pp. 5632–5643 (2020)
13. Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: Real-world super-resolution via kernel estimation and noise injection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 466–467 (2020)
14. Jo, Y., Oh, S.W., Vajda, P., Kim, S.J.: Tackling the ill-posedness of super-resolution through adaptive target generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16236–16245 (2021)
15. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
16. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations (2015)
17. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 624–632 (2017)

18. Liang, J., Sun, G., Zhang, K., et al.: Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4096–4105 (2021)
19. López-Tapia, S., de la Blanca, N.P.: Fast and robust cascade model for multiple degradation single image super-resolution. *IEEE Trans. Image Process.* **30**, 4747–4759 (2021)
20. Luo, Z., Huang, Y., Li, S., Wang, L., Tan, T.: Learning the degradation distribution for blind image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6063–6072 (2022)
21. Maeda, S.: Unpaired image super-resolution using pseudo-supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 291–300 (2020)
22. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision, vol. 2, pp. 416–423 (2001)
23. Matsui, Y., et al.: Sketch-based manga retrieval using manga109 dataset. *Multimed. Tools Appl.* **76**(20), 21811–21838 (2017)
24. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Deblurring images via dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 2315–2328 (2017)
25. Riegler, G., Schuler, S., Ruther, M., Bischof, H.: Conditioned regression models for non-blind single image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 522–530 (2015)
26. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
27. Shocher, A., Cohen, N., Irani, M.: “Zero-shot” super-resolution using deep internal learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3118–3126 (2018)
28. Tao, G., et al.: Spectrum-to-kernel translation for accurate blind image super-resolution. In: Advances in Neural Information Processing Systems, vol. 34, pp. 22643–22654 (2021)
29. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: NTIRE 2017 challenge on single image super-resolution: methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 114–125 (2017)
30. Wang, L., et al.: Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10581–10590 (2021)
31. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision Workshops (2018)
32. Xiao, J., Yong, H., Zhang, L.: Degradation model learning for real-world single image super-resolution. In: ACCV (2020)
33. Xu, Y.S., Tseng, S.Y.R., Tseng, Y., Kuo, H.K., Tsai, Y.M.: Unified dynamic convolutional network for super-resolution with variational degradations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12496–12505 (2020)
34. Zamir, S.W., et al.: CycleISP: real image restoration via improved data synthesis. In: CVPR (2020)

35. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., et al. (eds.) *Curves and Surfaces 2010*. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47
36. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3262–3271 (2018)
37. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 286–301 (2018)
38. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232 (2017)