Learning Correction Filter via Degradation-Adaptive Regression for Blind Single Image Super-Resolution

Hongyang Zhou[1], Xiaobin Zhu[1]*, Jianqing Zhu[2], Zheng Han[1], Shi-Xue Zhang[1]
Jingyan Qin[1], Xu-Cheng Yin[1]
[1]University Of Science and Technology Beijing, [2]College of Engineering, Huaqiao University
ilaopi@foxmail.com, zhuxiaobin@ustb.edu.cn, jqzhu@hqu.edu.cn, han970421@foxmail.com
zhangshixue111@163.com, qinjingyanking@foxmail.com, xuchengyin@ustb.edu.cn

Abstract

Although existing image deep learning super-resolution (SR) methods achieve promising performance on benchmark datasets, they still suffer from severe performance drops when the degradation of the low-resolution (LR) input is not covered in training. To address the problem, we propose an innovative unsupervised method of Learning Correction Filter via Degradation-Adaptive Regression for Blind Single Image Super-Resolution. Highly inspired by the generalized sampling theory, our method aims to enhance the strength of off-the-shelf SR methods trained on known degradations and adapt to unknown complex degradations to generate improved results. Specifically, we first conduct degradation estimation for each local image region by learning the internal distribution in an unsupervised manner via GAN. Instead of assuming degradation are spatially invariant across the whole image, we learn correction filters to adjust degradations to known degradations in a spatially variant way by a novel linearly-assembled pixel degradation-adaptive regression module (DARM). DARM is lightweight and easy to optimize on a dictionary of multiple pre-defined filter bases. Extensive experiments on synthetic and real-world datasets verify the effectiveness of our method both qualitatively and quantitatively. Code can be available at: https://github.com/edbca/DARSR.

1. Introduction

Single image super-resolution (SISR) aims to recover a photo-realistic HR image from its LR counterpart. As a fundamental computer vision task, SISR has been widely used in various applications, such as enhancing the images' visual perception on high-resolution devices and improving the image quality for other vision tasks [7]. In recent years,

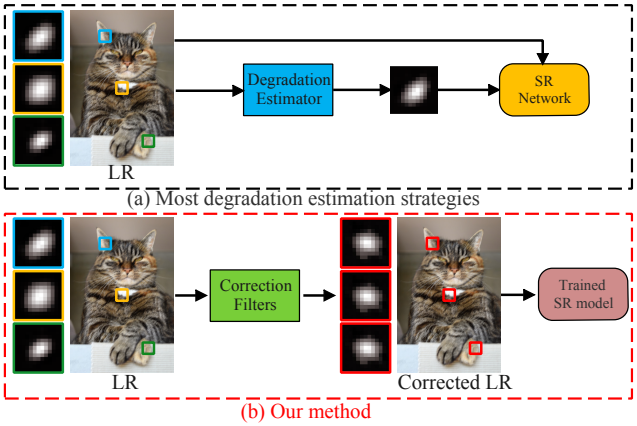


Figure 1. Illustration of most degradation estimation strategies for blindSR and our method. (a) Most methods assume degradations are spatially invariant and estimate a single degradation for the whole image. (b) Our method can learns correction filters to adjust the degradations to a known degradation in a spatially variant way.

a multiplicity of deep learning-based SISR methods have been proposed [33, 18, 25, 52, 51, 8, 29, 54] and achieved substantial performance improvement.

Typically, training and testing LR images in existing SISR approaches are simulated from a large collection of HR images. As a pioneer work, Dong *et al.* [9] first introduced a three-layer convolutional neural network in SISR to learn LR-HR mapping. Zhang *et al.* [51] proposed a channel attention mechanism to build a deep model for SISR. Liang *et al.* [22] proposed a baseline model SwinIR for SISR based on the transformer mechanism. In these methods, the LR images are often generated by limited numbers of fixed degradations (e.g., Bicubic or blur-downscale). However, the degradation process in real-world scenarios is extremely complicated and unpredictable. Although existing methods [25, 51, 29, 21, 18, 8] have achieved promising performance, they often encounter severe performance drops when the degradation is not well covered in training.

*Corresponding author (zhuxiaobin@ustb.edu.cn).

Recently, many works have attempted to address the challenging degradation adaptation problem by explicit or implicit degradation modeling in a blind or non-blind manner. SRMD [48] learns an SR model with training data covering a large set of degradation maps, which are modeled with blur kernel and noise. Following SRMD, UDVD [43] proposes a single unified dynamic network trained for various degradations to improve performance. However, in most practical applications, blur kernels are not available. They can hardly deal with the LR inputs with degradations out of their modeling scope. To tackle this problem, some methods [34, 35, 2, 11] estimate a blur kernel from each input LR image for adapting to complex degradations. IKC [12] proposes to correct kernel estimation in an iterative way to approach a satisfactory result gradually. ZSSR [34] learns degradation kernels of LR images via internal statistics for every image in a self-supervised manner. Pseudo-supervision [53] combines the forward SR reconstruction path in CinCGAN [44] architecture with degradation learning to deal with the domain shift problem. In addition, Hussein *et al.* [14] proposed a correction filter strategy on the signal that is sampled by a certain basis can be reconstructed by a different basis, for modifying an LR input to match the SR model with a pre-defined degradation. These methods' generalization ability still greatly relies on the accuracy of degradation estimation or training data distribution. In addition, as shown in Fig. 1 (a), they assume degradations are spatially invariant and only estimate a single degradation for the whole image. Such an assumption is rarely applicable for real images whose regions (e.g., edges, corners) are discriminative for degradation estimation.

To fully take advantage of the off-the-shelf super-resolvers, we try to robustify the degradation of LR input without hurting useful information for efficiently adapting to practical applications. A brief illustration and comparison to the existing degradation estimation strategy for blindSR are presented in Fig. 1. Specifically, we propose an innovative unsupervised method of learning correction filter via a linearly-assembled pixel degradation-adaptive regression module (DARM) for correcting the LR image's degradation to the known one in a spatially variant way. Our DARM is highly lightweight and can be easily optimized over a dictionary of multiple pre-defined filter bases. In this way, our method can conduct correction filtering and help off-the-shelf SISR models to adapt to complex degradations on every testing image efficiently. In addition, we propose a degradation metric (DM) loss to learning to map the unquantifiable degradation in a metric space to guarantee the accuracy of degradation correction.

In summary, our main contributions are three-fold:

- We propose an innovative unsupervised strategy by learning correction filters for Blind Single Image Super-Resolution in a spatially variant way, greatly enhancing the adapting ability of trained super-resolvers to unknown complex degradations.
- We propose a lightweight degradation-adaptive regression module (DARM) that could accurately yet efficiently learn correction filters over a dictionary of multiple pre-defined filter bases.
- We propose a novel degradation metric (DM) loss via metric learning to map the unquantifiable degradation in a metric space to effectively improve the accuracy of degradation correction.
- Extensive experiments on synthetic datasets and real images verify that our method can consistently improve the performance of the off-the-shelf super-resolvers both qualitatively and quantitatively.

2. Related Work

2.1. Predefined-Degradation based SR

Dong *et al.* [9] firstly proposed a three consecutive convolutional neural network in SISR to learn a complex LR-HR mapping. Kim *et al.* [15] increased the depth of the network via a skip connection for stable training. To achieve promising performance, most methods adopt residual architecture [38, 52, 26]. EDSR [25] removes the batch normalization layer in a residual block and increasing the network depth. RCAN [51], SAN [8] and HAN [29] introduce the attention mechanism to draw more high-frequency information to further improve the performance. Furthermore, SwinIR [22], IPT [4], GRL [20] employ the transformer mechanism for constantly refreshing the state-of-the-art of SISR. However, these methods perform promisingly under a single degradation model (e.g., bicubic and blur-downscale degradation models); they tend to produce over-sharped or blurry results if the degradations present in the testing deviate from training.

2.2. Degradation-Learning based SR

Multiple degradation modeling based SR. By modeling multiple degradations, this type of method can produce good results for LR inputs covered by the training dataset. SRMD [48] proposes to directly concatenate an LR input image with its degradation map as a unified input to the SR model, thus allowing feature adaptation according to the specific degradation and covering multiple degradation types in a single model. DPSR [49] incorporates a super-resolution network into a MAP-based iterative optimization scheme which solves the HR image by minimizing objective functions. Similar to DPSR, USRNet [46] proposes an end-to-end trainable unfolding network that leverages both learning-based methods and model-based methods by unfolding the MAP inference via a half-quadratic splitting algorithm. Luo *et al.* [13] adopted an alternating optimization

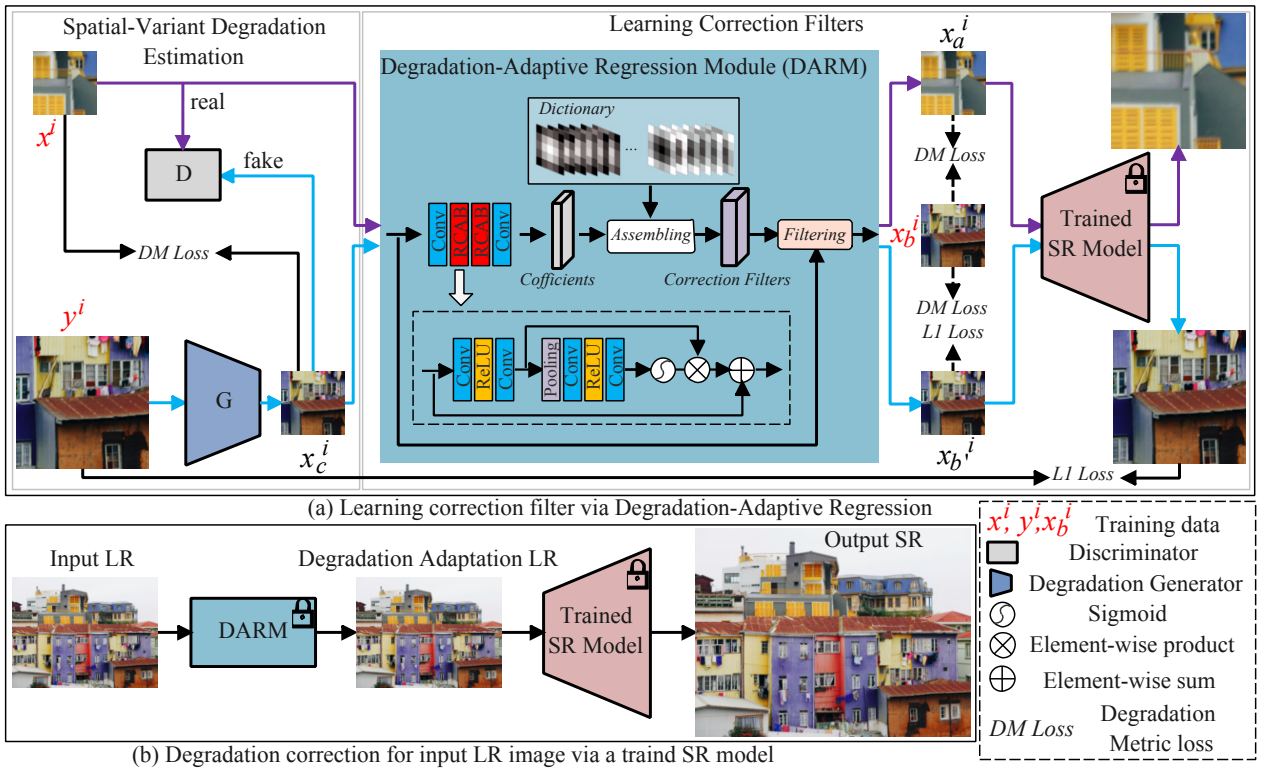


Figure 2. Architecture of our method. (a) Illustration of the learning of correction filters. (b) Illustration of the degradation correction.

algorithm to estimate blur kernel and restore SR image in a single model. However, the performance of these methods still relies on accurate estimation of input kernel, which is challenging on arbitrary LR images.

Degradation estimation based SR. Gu *et al.* [12] proposed a general framework for blind SR by using self-similarity to recover the blur kernel and the HR image jointly. KernelGAN [2] interprets the maximization of patch recurrence within a single image as a data distribution learning problem and generates downscaled LR images by learning the internal patch distribution of the testing LR image. Shocher *et al.* [34] proposed a zero-shot super-resolution (ZSSR) method which trains image-specific on the testing LR image to take the degradation parameters and estimate blur kernel for improving the performance. Afterward, Soh *et al.* [35] proposed to learn an image-specific degradation simulation network together with a zero-shot image-specific super-solution network which exploits the depth information of the image to extract an unpaired high/low resolution patch collection to train the network. Liang *et al.* [24] proposed a novel method to incorporate a flow-based kernel prior into the framework by learning an invertible mapping between the anisotropic Gaussian kernel distribution and a tractable latent distribution. While these methods consider degradation are spatially invariant and only estimate a single degradation for input image. Cornillere *et al.* [5] proposed a degradation-aware SR network to adapt spatially variant degradations. MANet [23] uses mutual affine convolution in their network for spatially variant kernel estimation. However, these methods greatly rely on the coverage of degradation modeling or training data distribution.

3. Our Method

3.1. Preliminaries

The SISR degradation model can be mathematically formulated as:

$$x = (y \otimes k) \downarrow_s, \tag{1}$$

where y denotes the HR image; x denotes the LR image; k is the blur kernel; $\otimes$ denotes convolution operator; $\downarrow_s$ denotes sub-sampling operator with stride of s. Note that Eq. 1 can be written as:

$$x = Dy, \tag{2}$$

where D represents degradation operator. As assumed by Hussein *et al* [14], the signal y can be perfectly recovered from its samples x by the operator $U(DU)^{-1}$ as:

$$y = U(DU)^{-1}x = U(DU)^{-1}Dy, \tag{3}$$

where U is an upsampling operator which is the adjoint operator of D. Eq. 3 indicates that the image y resides in the linear subspace can be spanned by the known degradation.

Therefore, it can be perfectly recovered from the observations x_k by applying the pseudo inverse of U_k as:

$$y' = U_k(D_kU_k)^{-1}D_ky = U_k(D_kU_k)^{-1}x_k, \quad (4)$$

where D_k denotes known degradation operation; x_k denotes known degraded image; U_k denotes a DNN super-resolver trained under the assumption of known degradation and it can handle the x_k quite well with a learned prior.

According to the generalization of classical Whittaker-Shannon sampling theorem [17, 30, 32, 42], some works [3, 10, 39] consider that a signal sampled by a certain basis can be reconstructed by different basis. By combining a correction filter with a reconstruction kernel, a sampled signal can be reconstructed. The correction filter can transform the sampling coefficients associated with the sampling kernel to coefficients that fit the reconstruction kernel. Given a degraded LR image x_u obtained by a unknown degradation D_u, $x_u = D_uy$. Then, the y' can be reconstructed as:

$$y' = U_kCx_u = U_kCD_uy, \quad (5)$$

where C is a correction operator for x_u. Combining Equation 4 and Eq. 5, we can find the expression of C as follows:

$$C = (D_uU_k)^{-1} \Rightarrow CD_u = U_k^{-1}. \quad (6)$$

The operator C can be applied simply as a convolution with a correction filter. Following Eq. 6, we first conduct spatial-variant degradation estimation to model D_u, which learns the spatially variant degradations in x_u. Then, we use our DARM to learn the correction filters C for correcting the degradations of x_u to pre-defined degradation. Finally, we employ a trained U_k to super-resolve the corrected image.

Our method mainly distinguishes from the work [14] in two respects: (1) The performance of the work [14] greatly relies on the accuracy of degradation estimation. In our method, we coherently combine degradation estimation with correction filter learning into a unified network, greatly improving the accuracy and adaptiveness; (2) The work [14] only considers the spatially invariant degradation situation, failing to adopt a correction filter to solve a natural image with complex degradations. In our method, we estimate degradations in a spatially variant way.

3.2. Overview

As shown in Fig. 2 (a), we learn correction filters via DARM in an unsupervised manner. Then, as shown in Fig. 2 (b), we use learned correction filters to correct the degradation of the input LR image to the pre-defined degradation. Finally, an off-the-shelf super-resolver trained by known degradation is used for upsampling the degradation adapted LR image for SR. The learning of correction filters mainly consists of three components: generation of training data from input LR image, spatial-variant degradation estimation, and learning correction filters through proposed degradation adaptation regression module.

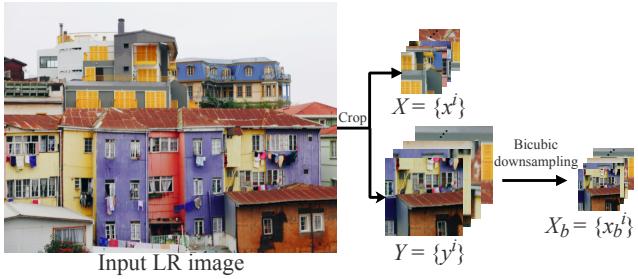


Figure 3. Illustration of the generation of training data from the input LR image. $X = \{x^i\}$, $Y = \{y^i\}$, and $X_b = \{x_b^i\}$ are used to train our proposed degradation-adaptive regression module.

3.3. Generation of Training Data

Given the input LR image, our goal is to correct complex degradations of all regions to the degradation (e.g. bicubic) in the off-the-shelf super-resolver. As shown in Fig. 3, we first crop the input LR image to generate the patches $X = \{x^i\}$ where each patch x^i has the degradation of the corresponding area and X can cover the degradation of each region in the input LR image. Then, we crop the input LR image to generate the patches $Y = \{y^i\}$, which aims to learn a downscaled image (e.g., by a factor of 2 or 4) such that its degradation is approximate to that of the patches X (see Sec. 3.4). Finally, we use known downsampling operation to degrade patches Y for generating patches $X_b = \{x_b^i\}$ of the same size with x^i for degradation-adaptive regression learning. After obtaining training data X, X_b, and Y, we will train our network to learn correction filters as shown in Fig. 2 (a).

3.4. Spatial-Variant Degradation Estimation.

X and X_b are not paired, which leads to the correction filters are unable to be learned directly. Thus, we adopt the GAN-based method to learn $X_c = \{x_c^i\}$ with the same degradation as X through a GAN. Specially, we employ a degradation generator G and a discriminator D, as shown in Fig. 2 (a). Different from KernelGAN [2] which only learns one global invariant degenerate kernel, we will learn spatially variant degradation for each local patch x_c^i. In this way, we can learn X_c to cover the degradation of each region for improving performance. In addition, we propose a novel degradation metric (DM) loss (see Sec. 3.6) to learn more accurate degradation in x_i.

3.5. Learning Correction Filters via DARM

As shown in Fig. 2 (a), we adopt paired training data X_c and X_b to learn correction filters by training our DARM through L1 loss and DM loss. We also train our network on unpaired training data X and X_b by DM loss. The output of X_c and X through DARM can be expressed as $X_{b'} = \{x_{b'}^i\}$ and $X_a = \{x_a^i\}$, respectively. Notably, our DARM learns a

correction filter for each input pixel.

Our key idea lies on the total probability theorem and the additivity of normal distribution. We can obtain the correction filter by combining multiple Gaussian filters. Specifically, a correction filter F^i can be represented as a linear combination of a set of underlying base Gaussian filters:

$$F^i = \Phi^i \mathbb{D} = \sum_{j=1}^{N} \theta_j^i f_j, \tag{7}$$

where $F = \mathbb{R}^{HW \times (k \times k)}$ (k is the kernel size which is emprically set to 5 in our experiments) denotes the learned correction filters for each pixel in the input image of size $H \times W$;$\Phi = \mathbb{R}^{HW \times N}$ denotes the learned linear combination coefficients for each pixel; The dictionary $\mathbb{D} = \mathbb{R}^{N \times (k \times k)}$ consists of 72 Gaussian and difference of Gaussians (DoG) filters [19], which can help to increase the visibility of details accurately learn correction filters.

For the purpose of efficiency, we adopt a lightweight residual network for DARM via learning linear combination coefficients Φ. It mainly consists of two residual channel attention blocks (RCAB) [51] to learn deep feature, the convolutional layers in the head to extract the shallow feature of the input image, and the final convolutional layers to generate combination coefficients.

3.6. Loss Functions

Adversarial loss. We adopt GAN-based method to learn spatially variant degradations of LR image. The adversarial loss for the degradation generator G can be formulated as:

$$\mathcal{L}_{gan} = \mathbb{E}[D(G(y^i)) - 1]^2 + \mathcal{R}, \tag{8}$$

where $\mathcal{R}$ denotes the regularization term which applies realistic explicit priors [2]. The discriminator D tries to distinguish degradations generated by G from x^i of input LR image and the constrain of D can be formulated as:

$$\mathcal{L}_D = \mathbb{E}[D(x^i) - 1)^2] + \mathbb{E}[D(G(y^i))^2]. \tag{9}$$

Pixelwise loss. It is used to accelerate the convergence of DRAM. Besides, it can effectively compensate for the structural distortion caused by GAN. We use L1 loss to constrain the results generated by DASR and SR model. The mathematic formulation is:

$$\mathcal{L}_{pix} = ||x_b^i - x_{b'}^i||_1 + ||y^i - SR(x_{b'}^i)||_1, \tag{10}$$

where $SR(*)$ denotes an off-the-shelf super-resolver.

Degradation metric loss. It is difficult to estimate complex degradations accurately via single adversarial learning in real-world scenarios. To overcome this problem, we propose a novel degradation metric (DM) loss which maps the complex degradation into a metric space for effective

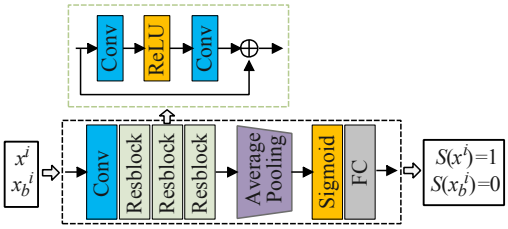


Figure 4. Architecture of degradation score estimation module (DSEM), which mainly consist of three Resblocks [25], an average pooling layer, a sigmoid layer, and a full connection layer.

constraint. Specially, we use a degradation score estimation module (DSEM) to model the distance between target degradation and estimated degradation. We use x^i and ${x_b}^i$ as anchor samples to train the metric space as shown in Fig. 4. The mathematical formulation is elaborated as:

$$\mathcal{L}_{ac} = \frac{1}{N} \sum_{i}^{N} (||S(x_b^i)||_2^2 + ||S(x^i) - 1||_2^2), \tag{11}$$

where $S(*)$ denotes the degradation score estimation operation, which is implemented by DSEM. Then, we use DM loss from two parts, one is the result generated by downscale generator G, and the other one is the result generated by DARM. The mathematical formulation is elaborated as:

$$\begin{aligned} \mathcal{L}_{dml} = ||S(x_b^i) - S(x_{b'}^i)||_2^2 + ||S(x^i) - S(G(y^i))||_2^2 \\ + ||S(x^i) - S({x_a}^i))||_2^2. \end{aligned} \tag{12}$$

Full objective. Mathematically, the total loss function for our method can be formulated as:

$$\mathcal{L}_{All} = \mathcal{L}_{gan} + \lambda_1 \mathcal{L}_{pix} + \lambda_2 \mathcal{L}_{dml}, \tag{13}$$

where λ_1 and λ_2 are two empirical hyper-parameters that are set to 2.5 and 0.5, respectively.

4. Experiments

4.1. Implementation Details

We conduct experiments on real-world datasets (RealSRSet [47]) and synthetic datasets (NTIRE2017 [36] and NTIRE2018 [37]). The results are evaluated by the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [41], and learned perceptual image patch similarity (LPIPS) [50]. PSNR and SSIM are evaluated on the Y channel of transformed YCbCr, while LPIPS is on RGB. We train our network by each iteration with 16 batches which consist of paired patches from the LR input. The patches 32×32 and 128×128 (x^i and y^i in Fig. 3) are for scaling ×4. The patches 64×64 and 128×128 are for scaling ×2. The iteration number is set to $3,000$, and the optimizer is the Adam [16] with $\beta 1 = 0.9$, $\beta 2 = 0.999$. The initial

Table 1. Quantitative results on NTIRE2017 [36] and NTIRE2018 [37] testing datasets for scaling factor $\times 4$. $\uparrow$ denotes the larger the better. $\downarrow$ denotes the smaller the better. Red color indicates the best performance, and blue color indicates the second best performance.

Methods	2017Track2			2018Track2			2018Track3			2018Track4		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SRGAN [18]	22.99	0.5889	0.4669	18.99	0.4582	0.6127	19.75	0.4706	0.6241	19.50	0.4687	0.6150
SRFBN-s [21]	23.07	0.6043	0.4845	18.87	0.4886	0.6085	19.78	0.5043	0.6251	19.51	0.5020	0.6112
KernelGAN [2] + ZSSR [34]	22.49	0.5800	0.4951	18.43	0.5450	0.5739	19.75	0.5750	0.6000	19.36	0.5747	0.5774
BlindSR [5]	21.61	0.5483	0.4820	18.35	0.4096	0.6088	19.31	0.4243	0.6192	18.90	0.4171	0.6102
IKC [12]	22.28	0.5738	0.4683	18.72	0.5488	0.5686	19.81	0.5716	0.5950	19.42	0.5713	0.5723
SRGAN [18] + Correction Filter [14]	20.98	0.5776	0.5821	18.73	0.5555	0.6141	19.07	0.5571	0.6288	18.95	0.5661	0.6147
SRFBN-s [21] + Correction Filter [14]	20.98	0.5851	0.6023	18.74	0.5631	0.6255	19.10	0.5659	0.6499	18.98	0.5746	0.6313
USRNet [46]	22.85	0.6029	0.5110	19.00	0.5641	0.5700	19.97	0.5825	0.6016	19.67	0.5853	0.5795
MANet [23]	21.41	0.5579	0.4911	18.66	0.4744	0.6090	19.31	0.4833	0.6288	19.01	0.4762	0.6154
DIP-FKP [24]	20.49	0.5336	0.5050	18.11	0.5528	0.5781	18.72	0.5609	0.6086	18.45	0.5605	0.5869
DASR [40]	22.92	0.5968	0.4780	18.92	0.5579	0.5690	19.83	0.5748	0.5929	19.53	0.5782	0.5712
BSRDM [45]	21.63	0.5797	0.5581	18.72	0.5483	0.5946	19.57	0.5657	0.6124	19.18	0.5596	0.6001
Ours + SRGAN [18]	23.09	0.6017	0.4637	20.24	0.5741	0.6088	21.04	0.5845	0.6173	21.18	0.5887	0.6021
Ours + SRFBN-s [21]	23.35	0.6126	0.5063	20.32	0.5787	0.6225	21.11	0.5859	0.6291	21.34	0.5948	0.6021

Table 2. Quantitative results on NTIRE2017 Track2 [36] testing dataset for scaling factor $\times 2$.

Methods	PSNR↑	SSIM↑	LPIPS↓
SRFBN-s [21]	26.43	0.7246	0.3437
KernelGAN [2] + ZSSR [34]	26.34	0.7288	0.3295
BlindSR [5]	27.22	0.7539	0.3054
SRFBN-s [21] + Correction Filter [14]	23.44	0.6125	0.3482
USRNet [46]	26.52	0.7368	0.3710
MANet [23]	27.92	0.7818	0.3086
DIP-FKP [24]	26.29	0.7492	0.3188
DASR [40]	26.38	0.7222	0.3335
BSRDM [45]	26.65	0.7303	0.4432
Ours + SRFBN-s [21]	27.99	0.7822	0.3513

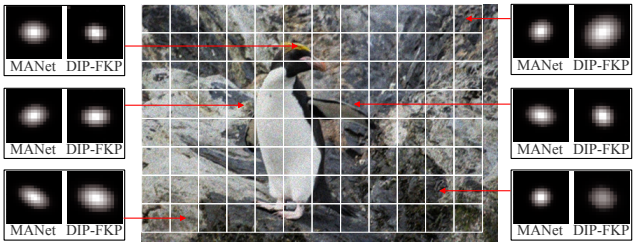


Figure 5. Illustration of degradation estimation for MANet [23] and DIP-FKP [24] on "*0801×4m.png*" in NTIRE2018 track2.

learning rate is 10^{-3} which will be decayed by 0.5 after each 750 iterations. Data augmentation includes randomly rotating 90, 180, 270 degrees, and horizontally flipping. We implement experiments with PyTorch1.12.0 on an NVIDIA Geforce RTX 3090 GPU.

4.2. Experiments on Synthetic Datasets.

We conduct experiments on four synthetic benchmarks: track2 of NTIRE2017 [36] and track2, track3, and track4 of NTIRE2018 [37]. The degradations in these datasets are complex and unknown. Notably, the official uses same degradation operators within an image. But, as shown in Fig. 5, we use MANet [23] and DIP-FKP [24] to estimate the degradation of different regions in LR image, we find that the degradation of degraded image is spatially variant. We think that the given HR images are spatially variant degraded from the real HR ones due to camera acquisition in complex scenes, which make the LR images are spatially variant. We compare with unsupervised and supervised methods, including SRGAN [18], SRFBN-s [21], KernelGAN [2], ZSSR [34], BlindSR [5], IKC [12], Correction_Filter [14], USRNet [46], DIP-FKP [24], MANet [23] and BSRDM [45].

4.2.1 Quantitative Evaluation.

The detailed experimental results are listed in Tab. 1 and Tab. 2. Note that our method does not have a design module for denoising, so we employ CBM3D [6] to conduct on track2, track3, and track4 of NTIRE2018 datasets firstly. We also use the same operation for other methods without the ability of denoising for a fair comparison.

Comparing with trained SR methods. As listed in Tab. 1 and Tab. 2, our method can improve the performance of SRGAN in all testing datasets on PSNR/SSIM/LPIPS. Our method also improve the performance of SRFBN-s in all testing datasets on PSNR/SSIM for scaling factor $\times 2$ and $\times 4$. We noticed that our method did not improve the LPIPS values of SRFBN-s. This phenomenon often occurs in PSNR-Oriented methods (e.g., EDSR, RCAN). Our method can improve the performance of perceptual-driven methods (e.g., SRGAN, ESRGAN) in all indicators, including LPIPS. PSNR-Oriented methods focus on achiev-

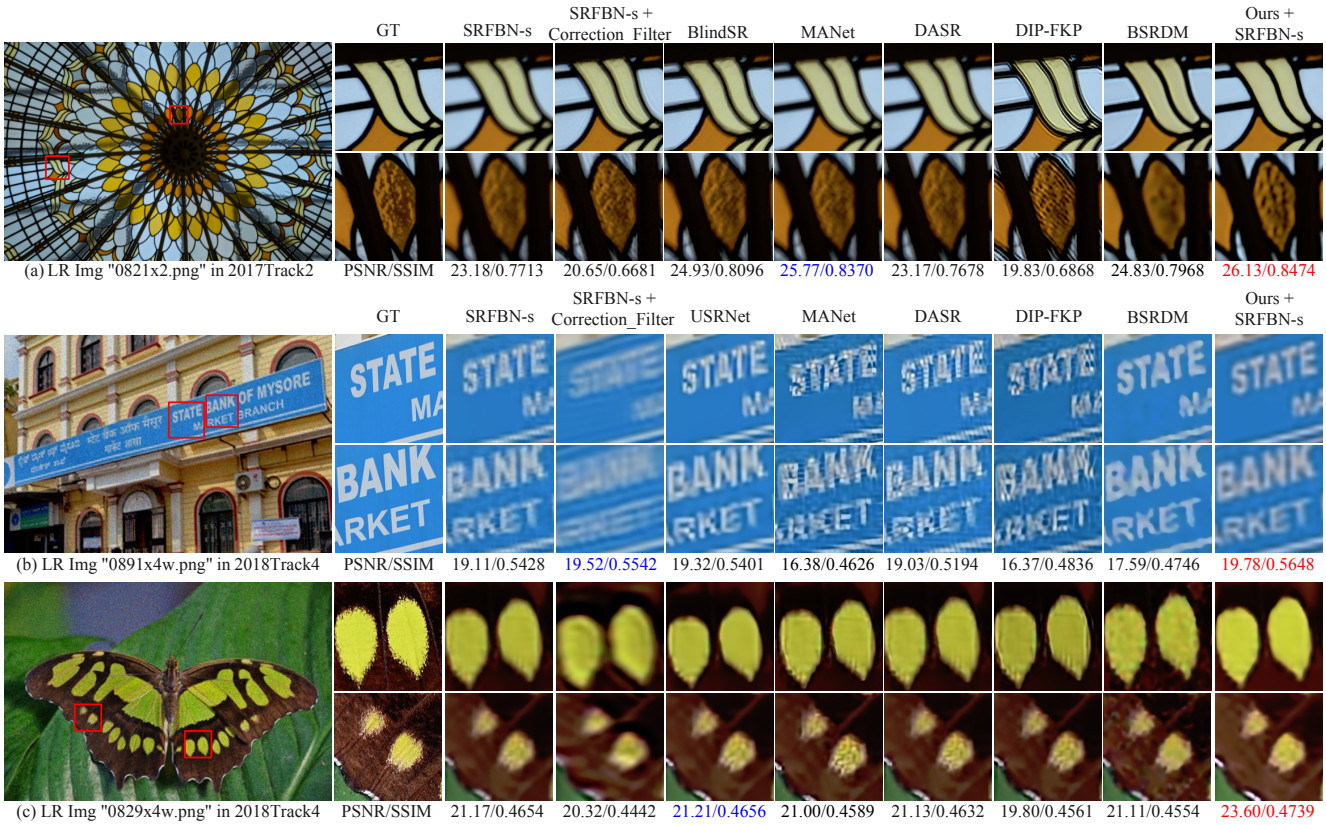


Figure 6. Visual comparisons of different methods on 2017Track2 [36], 2018Track2 [37], and 2018Track4 [37]. Zoom in for best view.

Figure 7. Visual comparisons of different method for $\times$ 4 on RealSRSet [47] dataset. Zoom in for best view.

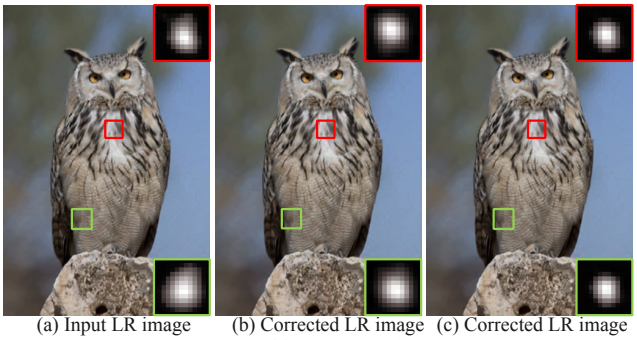
ing high PSNR and tend to generate images with too smooth and lost details (More analysis in **Supp. Materials**). Although our method can effectively recover LR images, the characteristics of these methods still lead to LPIPS degradation. But in general, our method can consistently improve the performance of off-the-shelf super-resolvers.

Comparing with Correction_Filter. As shown in Tab. 1 and Tab. 2, Correction_Filter failed to improve the performance of SRFBN-s and SRGAN. The performance of Correction_Filter greatly relies on the accuracy of degradation estimation. However, it is difficult to solve complex degradation by their iterative approximation strategies. Besides, they only considers the spatially invariant degradation, failing to tackle spatially ones (More comparison are in **Supp. Materials**). The experimental results show that our method is more effective than Correction_Filter.

Comparing with spatially-variant SR methods. From Tab. 2, we can see that the performance of SRFBN-s is far lower than BlindSR and MANet in all indicators. While SRFBN-s equipped our method outperforms BlindSR and MANet by 0.77 dB/0.0283 and 0.07 dB/0.0004 in terms of PSNR/SSIM, respectively. In Tab. 1, our method can help SRGAN and SRFBN-s outperforms Blind and MANet on all testing datasets in terms of PSRN/SSIM. The experi-

mental results verify that our method is superior in solving spatial-variant degradation.

Comparing with other blind methods. As listed in Tab. 1, Ours+SRFBN-s achieves the best SSIM and PSNR on all datasets. Ours+SRFBN-s outperforms BSRDM by 1.72 dB/0.0329 and 2.16 dB/0.0352 on 2017Track2 and 2018Track4 on PSNR/SSIM, respectively. Ours+SRFBN-s outperforms DASR by 1.4 dB/0.0208 and 1.81 dB/0.0166 on 2018Track2 and 2018Track4 on PSNR/SSIM, respectively. Tab. 2 also verify that our method outperforms other blind methods. We conclude that our method is effective and has strong degradation-adaptive ability.



(a) Input LR image (b) Corrected LR image (without DM Loss) (c) Corrected LR image

Figure 8. Visual ablation study about DM loss.

4.2.2 Qualitative Evaluation.

Fig. 6 are representative visual results. As shown in Fig. 6 (a), our method can generate sharper and cleaner SR images, but most other methods generate details with blurry or artifacts. Although the SR images generated by BSRDM seem to be clear, they lose some important details. In Fig. 6 (b), Ours+SRFBN-s can correctly restore font textures, while most other methods will produce incorrect textures. Although BSRDM generates font texture with good visual effects, many distortions exist. In Fig. 6 (c), Ours+SRFBN-s recovers clear and natural details, while other methods are prone to generate detailed textures with visual aliasing and artifacts. Notably, our method can greatly improves the visual quality of SRFBN-s, demonstrating our method's effectiveness.

4.3. Experiments on Real-World Dataset.

We also conduct experiments on RealSRSet [47] dataset. (Due to most images in RealSRSet are small size, we reduce the cropped image size in Sec 4.1.) As mentioned in [47, 45], current non-reference metrics (e.g., NIQE [28], NRQM [27], and PI [1]) are not consistent with the perceptual visual system. Therefore, we only compare visual results. Fig. 7 shows the visual results on two representative images. Apparently, Ours+SRFBN-s generates improved visual results than other methods. In the first examples in Fig. 7, most other methods tend to generate noises or artifacts. USRNet generated observable visual results but with too many white lines that should not exist. BSRDM generates unrealistic and discontinuous lines with some distortions. In the second example in Fig. 7, most other methods cannot finely deal with the hair and tend to produce unrealistic detailed textures with visual aliasing. In contrast, our Ours+SRFBN-s reconstructs clear and natural details.

Table 3. Ablation study of DM Loss on NTIRE2017 Track2 testing dataset for scaling factor $\times 4$.

Method	$\mathcal{L}_{gan}$	$\mathcal{L}_{pix}$	$\mathcal{L}_{dml}$	PSNR↑	SSIM↑	LPIPS↓
Ours+SRFBN-s	✓			16.24	0.3757	0.7231
Ours+SRFBN-s		✓		21.89	0.5624	0.6532
Ours+SRFBN-s	✓	✓		22.06	0.5704	0.6430
Ours+SRFBN-s	✓	✓	✓	**23.35**	**0.6126**	**0.5063**

4.4. Ablation Study

About DM Loss. To verify the effectiveness of DM loss, we conduct experiments on the 2017Track2 dataset. As listed in Tab. 3, Ours+SRFBN-s outperforms Ours +SRFBN-s (without $\mathcal{L}_{dml}$) on scale $\times 4$ by 1.29dB on PSNR, 0.0423 on SISM, and 0.1367 on LPIPS, respectively. Besides, our DM loss can clearly improve the accuracy of degradation correction of corrected LR image (Fig. 8 (c)), compared with our model without DM loss (Fig. 8 (b)).

Table 4. Ablation study of filter dictionary on NTIRE2017 Track2 testing dataset for scaling factor $\times 4$.

Filter Type	Number	PSNR↑	SSIM↑	LPIPS↓
G + DoG [19]	72	**23.35**	**0.6126**	**0.5063**
RAISR [31]	72	22.12	0.5334	0.6430
Learned	72	22.44	0.5343	0.6266
G + DoG [19]	24	23.17	0.6114	0.5145

About the base Filter Dictionary. We conducted experiments on the 2017Track2 dataset to verify the chosen of the base filter dictionary. We compare with RAISR filters [31] and filters directly learned by networks. As listed in Tab. 4, it is clear that 72 Gaussian and DoG combination achieves the best result. Tab. 4 also shows that the result with 72 basic filters is better than the dictionary with 24 basic filters. Besides, the filters size with 5×5 for each pixel is sufficient to restore complex real-world images. Various kernel sizes definitely can enhance performance but introduce great computational costs.

Table 5. Ablation study on trained SR with different degradations.

Method	2018Track2	2018Track3
Ours + SRFBN-s(Bicubic)	20.32/0.5787	21.11/05859
Ours + SRFBN-s(Gaussian)	20.31/0.5787	21.11/0.5858
Ours + SRFBN-s(Lanczos)	20.32/0.5787	21.13/0.5859

About trained SR with different degradation. We choose two different degradations (Gaussian blur and Lanczos) to train SRFBN-s on DIV2K dataset for scaling factor $\times 4$, respectively. Tab. 5 lists the detailed results on NTIRE 2018 track2 and NTIRE 2018 track3. The results on the two degradation models are nearly consistent with bicubic degradation, which demonstrates that our method can adapt to any known degeneration in a trained model.

5. Conclusion

In this paper, we propose a novel blind image super-resolution method that corrects the degradations of an LR input to that of the training data in an off-the-shelf method. Specifically, we propose an innovative unsupervised method of learning correction filter via a linearly-assembled pixel DARM for correcting the LR image's degradation to the known one in a spatially variant way for each testing image. Besides, we propose a degradation metric loss to map the unquantifiable degradation to guarantee the accuracy of degradation correction. Extensive experiments verify the superior performance of our method both quantitatively and qualitatively.

6. Acknowledge

The research is supported by National Key Research and Development Program of China (2020AAA0109700), National Science Fund for Distinguished Young Scholars(62125601), National Natural Science Foundation of China (62172035, 62076024, 62006018).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 8

[2] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019. 2, 3, 4, 5, 6

[3] Paul Leo Butzer. A survey of the whittaker-shannon sampling theorem and some of its extensions. In *Journal of Mathematical Research and Exposition*, volume 3, 1983. 4

[4] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 2

[5] Victor Cornillere, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. 3, 6

[6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *IEEE International Conference on Image Processing*, volume 1, pages I–313. IEEE, 2007. 6

[7] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. Is image super-resolution helpful for other vision tasks? In *IEEE Winter Conference on Applications of Computer Vision*, pages 1–9, 2016. 1

[8] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019. 1, 2

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of the European conference on computer vision*, pages 184–199. Springer, 2014. 1, 2

[10] Yonina C Eldar. *Sampling theory: Beyond bandlimited systems*. Cambridge University Press, 2015. 4

[11] Mohammad Emad, Maurice Peemen, and Henk Corporaal. Dualsr: Zero-shot dual learning for real-world super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1630–1639, 2021. 2

[12] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1604–1613, 2019. 2, 3, 6

[13] Yan Huang, Shang Li, Liang Wang, Tieniu Tan, et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, 33:5632–5643, 2020. 2

[14] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1428–1437, 2020. 2, 3, 4, 6

[15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2

[16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Represent*, 2015. 5

[17] Vladimir Aleksandrovich Kotelnikov. On the transmission capacity of the 'ether'and of cables in electrical communications. In *Proceedings of the first All-Union Conference on the technological reconstruction of the communications sector and the development of low-current engineering. Moscow*, 1933. 4

[18] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017. 1, 6

[19] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Advances in Neural Information Processing Systems*, 33:20343–20355, 2020. 5, 8

[20] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 2

[21] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019. 1, 6

[22] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 1, 2

[23] Jingyun Liang, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4096–4105, 2021. 3, 6

[24] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10601–10610, 2021. 3, 6

[25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017. 1, 2, 5

[26] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2359–2368, 2020. 2

[27] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 8

[28] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012. 8

[29] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *Proceedings of the European Conference on Computer Vision*, pages 191–207. Springer, 2020. 1, 2

[30] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928. 4

[31] Yaniv Romano, John Isidoro, and Peyman Milanfar. Raisr: rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging*, 3(1):110–125, 2016. 8

[32] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. 4

[33] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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*, pages 1874–1883, 2016. 1

[34] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3118–3126, 2018. 2, 3, 6

[35] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3516–3525, 2020. 2, 3

[36] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125, 2017. 5, 6, 7

[37] Radu Timofte, Shuhang Gu, Jiqing Wu, and Luc Van Gool. Ntire 2018 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 852–863, 2018. 5, 6, 7

[38] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4799–4807, 2017. 2

[39] Michael Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000. 4

[40] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, 2021. 6

[41] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[42] ET Whitaker. On the functions which are represented by the expansion of interpolating theory. In *Proc. Roy. Soc. Edinburgh*, volume 35, pages 181–194, 1915. 4

[43] Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12496–12505, 2020. 2

[44] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition Workshops, pages 701–710, 2018. 2

[45] Zongsheng Yue, Qian Zhao, Jianwen Xie, Lei Zhang, Deyu Meng, and Kwan-Yee K Wong. Blind image super-resolution with elaborate degradation modeling on noise and kernel. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2128–2138, 2022. 6, 8

[46] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3217–3226, 2020. 2, 6

[47] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 5, 7, 8

[48] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018. 2

[49] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019. 2

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 5

[51] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision*, pages 286–301, 2018. 1, 2, 5

[52] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 1, 2

[53] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017. 2

[54] Xiaobin Zhu, Zhuangzi Li, Jungang Lou, and Qing Shen. Video super-resolution based on a spatio-temporal matching network. *Pattern Recognition*, 110:107619, 2021. 1