Crafting Training Degradation Distribution for the Accuracy-Generalization Trade-off in Real-World Super-Resolution

Ruofan Zhang [1] **Jinjin Gu** [2 3] **Haoyu Chen** [4] **Chao Dong** [2 5] **Yulun Zhang** [6] **Wenming Yang** [1]

Abstract

Super-resolution (SR) techniques designed for real-world applications commonly encounter two primary challenges: generalization performance and restoration accuracy. We demonstrate that when methods are trained using complex, large-range degradations to enhance generalization, a decline in accuracy is inevitable. However, since the degradation in a certain real-world applications typically exhibits a limited variation range, it becomes feasible to strike a trade-off between generalization performance and testing accuracy within this scope. In this work, we introduce a novel approach to craft training degradation distributions using a small set of reference images. Our strategy is founded upon the binned representation of the degradation space and the Fréchet distance between degradation distributions. Our results indicate that the proposed technique significantly improves the performance of test images while preserving generalization capabilities in real-world applications.

1. Introduction

Image Super-Resolution (SR) is focused on reconstructing high-resolution (HR) images from their corresponding low-resolution (LR) or degraded observations. SR has a rich history of utilizing deep learning techniques, dating back to the groundbreaking work of SRCNN (Dong et al., 2015). With the advanced modeling capacity of deep networks, the performance of SR networks has experienced rapid progress. Nevertheless, it is widely recognized that the efficacy of SR models in practical applications is heavily influenced by their generalization performance and the training degradations (Liu et al., 2022a). The complex and diverse degradation scenarios encountered in real-world applications present considerable challenges to the successful implementation of SR techniques.

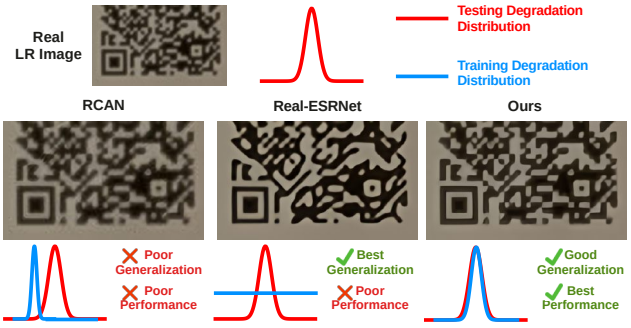

Figure 1. The figure shows the different effects of different training degradation distributions (shown in blue lines) on the target test distribution (shown in red lines). (a) The generalization performance of SR models limits their application when the training distributions are insufficient or mismatched. (b) When the training distribution is too large, the generalization of the SR model is better, but the overall accuracy will drop dramatically. (c) The proposed method improves the accuracy of test images as much as possible while ensuring the model's generalization performance.

A potential solution to tackle real-world SR challenges is the adoption of blind SR methods. These approaches generally rely on a predefined degradation model, while assuming that certain parameters remain unknown, such as the extent of downsampling blur or noise level. Consequently, blind SR methods are capable of addressing SR problems within a specific degradation range. However, the utilization of predefined degradation models limits their applicability to a narrow range of degradations, rendering them incapable of generalizing to mixed, intricate degradation cases encountered in real-world applications. As a result, these methods continue to exhibit sub-optimal generalization performance.

Recently, a new SR paradigm that employs a vast array of complex or mixed degradation data for training has increasingly gained interest within the research community. Notable methods include BSRGAN (Zhang et al., 2021) and Real-ESRGAN (Wang et al., 2021c). By leveraging

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University [2]Shanghai AI Laboratory [3]The University of Sydney [4]The Hong Kong University of Science and Technology (Guangzhou) [5]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences [6]ETH Zürich. Correspondence to: Jinjin Gu <jinjin.gu@sydney.edu.au>, Wenming Yang <yang.wenming@sz.tsinghua.edu.cn>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

large and diverse synthetic training datasets in an end-to-end manner, these techniques achieve substantial generalization performance. In practice, the degradation distribution used for training is determined by a manually configured stochastic process. Generally, the distribution range is set to be extensive in order to enhance generalization performance across degradations. However, there are inevitable trade-offs. In exchange for improved generalization performance, the accuracy of these models experiences a significant decline. Common issues include the removal of texture details, the generation of blurred or inaccurate edges, and the unwarranted sharpening of blurred backgrounds.

In this paper, we investigate the trade-off between accuracy and generalization from a more pragmatic standpoint. We contend that in the majority of application scenarios, degradation range tends to be limited, for instance, images captured by a specific type of image sensor, different frames from the same old movie, or old photos originating from the same era. Although these degradations are complex, which renders blind methods based on predefined degradation models ineffective, the range of degradation variation is relatively small compared to existing data synthesis strategies. The excessive portion of training degradation in conventional practices adversely impacts SR performance within the target range. Given the availability of images requiring super-resolution in practical applications (despite the unknown degradation process responsible for these images), we can customize the degradation distribution used for training to better suit the target test images. We illustrate this process in Figure 1. In summary, we explore a novel SR problem wherein, *given access to some test images, the training degradation distribution is modified to enhance performance within the target test degradation range while preserving generalization performance*.

To address this novel problem, we introduce two primary technical designs. Firstly, to determine a distribution within the space of potential degradations and sample training degradations from it, we require a suitable representation of the degradation distribution. We employ the binning method to partition the feasible degradation space into multiple distinct intervals. Uniform sampling is conducted within an interval, while weighted sampling is performed between different intervals (bins). This approach allows us to obtain a simple, parameterized sampling method for the degradation space. Secondly, we propose measuring the distance between two degradation distributions by calculating the Fréchet distance of deep features, even when the content of the two image sets differs. Based on the obtained degradation distribution distance, we calculate the weight for the degradation distribution bins, thereby parameterizing the training degradation distribution. We conduct extensive experiments for the proposed SR problem and method using both synthetic data and real-world low-resolution images. Our method demonstrates robust quantitative performance while maintaining strong generalization capabilities.

2. Related Work

Super-Resolution. Single image super-resolution (SR), which aims at reconstructing a high-resolution (HR) image from its low-resolution (LR) observations, is a long-standing problem in the low-level vision field. Since SRCNN (Dong et al., 2015), which is the pioneering work of employing deep learning in SR, deep SR networks have brought prosperous development in this field. Plenty of deep learning based SR methods have been proposed, including deeper networks (Kim et al., 2016b; Shi et al., 2016), light-weight networks (Dong et al., 2016; Gu et al., 2022; Zhou et al., 2023), recurrent architectures (Kim et al., 2016a; Tai et al., 2017), residual architectures (Ledig et al., 2017; Wang et al., 2018; Li et al., 2022), attention networks (Zhang et al., 2018; Dai et al., 2019), and Transformer networks (Chen et al., 2021; 2022; Liang et al., 2021; Chen et al., 2023b). However, most of these methods are aimed at the laboratory environment with pre-defined degradations, and the effect is limited in real applications.

Blind SR methods are proposed to solve the problem of SR model failure in real-world applications. The community has already reached a relatively clear conclusion for the reasons of the failure, that is, the mismatch between training and testing degradations (Liu et al., 2022a). Early blind SR methods usually assume a pre-defined degradation model with some unknown parameters, *e.g.*, the parameters of the blur kernel and the noise level (Gu et al., 2019; Huang et al., 2020; Cornillere et al., 2019). These methods still fail in a wide range of complex situations because real-world degradations are very complex. Simple degradation models cannot represent these situations. And the generalization ability of these methods is also not enough to make them applicable in the wild. Then, methods with implicit modeling are proposed and do not depend on any explicit parameterization. Utilizing data distribution within the external datasets, they often learn the underlying SR model implicitly, *e.g.*, CinCGAN (Yuan et al., 2018), DASR (Wang et al., 2021a), FS-SRGAN (Zhou et al., 2020), and FSSR (Fritsche et al., 2019). More recently, these methods have evolved further, with successful training on complex, large-range degradation data. Wang et al. (2021c) propose a novel data synthesizing method called high-order degradation model and train Real-ESRGAN. Zhang et al. (2021) propose BSRGAN that randomly applies different degradation operations during data synthesizing. Despite progress in visual effects, these methods rely on handset training degradation distributions. Both Real-ESRGAN and BSRGAN suffer from significant accuracy degradation when the training distribution is too wide.

More related to this work, there are also methods that con-

sider the case where a part of the test input image can be obtained as a reference. Wang et al. (2021b) propose an unpaired SR training framework based on feature domain adaptation. Luo et al. (2022) use adversarial training methods to generate specific training degradations.

Generalization Performance of SR models. The generalization performance of an SR network is crucial for its effectiveness on unseen data (Chen et al., 2023a; Gu et al., 2023). To date, limited research has focused on explaining, evaluating, and improving the generalization performance of SR networks. One study Liu et al. (2021) discovered that SR networks tend to overfit to degradations and exhibit characteristic degradation "semantics" (DDR) within the network, which often leads to a decrease in generalization ability. Building on this finding, another study developed a generalization assessment index for SR networks called SRGA Liu et al. (2022b). This non-parametric, non-learning metric measures generalization ability by examining the statistical characteristics of deep features within SR networks, rather than output images. As the generalization performance of SR gains increasing attention, specialized methods for enhancing SR generalization performance have emerged. For instance, one study demonstrated that the appropriate use of dropout benefits SR networks and improves generalization ability Kong et al. (2022). The goal of this paper is to explore ways to achieve higher accuracy while ensuring robust generalization performance.

3. Methodology

3.1. Problem Formulation

The proposed SR problem can be formulated as follows. Suppose we are designing an SR model for a new application, and we have obtained a set of reference degraded images $X_{ref} = \{x_i^r\}_{i=1}^n$ relevant to the application, and a set of test degraded images $X_{test} = \{x_i^t\}_{i=1}^N$ for evaluation. We can assume that these images are generated from the corresponding high-quality images Y_{ref} and Y_{test} with different degradations sampled from the same degradation distribution $P_r(d)$, where d denotes a random degradation and P_r represents the testing degradation distribution. This process can be formulated as $X = \mathcal{D}(Y, P_r)$, indicating that $x_i = d(y_i)$ for $x_i \in X$, $y_i \in Y$, and $d \sim P_r(d)$.

We now introduce a new degradation distribution P_t for training. Given a set of high-quality training images Y_{trn}, we synthesize the training degraded images $X_{trn} = \mathcal{D}(Y_{trn}, P_t)$ and then obtained the SR model F_θ. Here, θ is determined by P_t via a conditional distribution $\theta \sim P(\theta|P_t)$, as different training data will produce different SR models. Our goal is to maximize the performance of the obtained SR model F_θ on the target test degradation P_r by changing P_t. This can be formulated as:

$$\max_{P_t} \mathbb{E}_{\theta \sim P(\theta|P_t)} \mathcal{S}(F_\theta(\mathcal{D}(Y, P_r)), Y), \tag{1}$$

where $\mathcal{S}$ represents an image similarity metric used to evaluate the image reconstruction, *e.g.*, PSNR, SSIM.

3.2. Motivation

We then review the importance of the training with appropriate degradation distribution P_t using experiments on synthetic data. We assume a simple degradation model $x = d(y) = (y \otimes k) \downarrow$, where k is the blur kernel and $\downarrow$ denotes downsampling. We set P_r as the degradation distribution formed by sampling blur kernel widths σ_r from 1.5 to 2.5 uniformly, denoted as $\sigma_r \sim \mathcal{U}_{[1.5,2.5]}$. We train three SR models with different training distributions P_1, P_2 and P_3 and observe their behavior, where $\sigma_1 = 2$, $\sigma_2 \sim \mathcal{U}_{[0,4]}$ and $\sigma_3 \sim \mathcal{U}_{[1.5,2.5]}$. The testing average PSNR is shown in Figure 2 (a). We also test the performance of these models under different blur conditions separately, shown in Figure 2 (b). It can be seen that the model trained with a single degradation cannot handle other degradations within P_r except the training degradation. The SR model trained with P_2 degradation distribution can handle a large range of blurs, and is similar to the existing practices such as Real-ESRGAN and BSRGAN. This approach does bring about excellent generalization ability even beyond the target blur range, but in order to take into account a larger range of inputs, the restoration accuracy of all degradations in the range is reduced when the network capacity remains unchanged. The final model uses a matched training distribution and achieves the best PSNR performance while maintaining its generalization performance in the target range.

This experiment justifies our problem from three aspects: First, training within a certain degradation *range* is necessary because it can provide generalization performance to avoid performance drop when the degradation model is slightly mismatched. Second, this range is not the bigger, the better. An SR model training with a larger range may generalize to more images but suffer a performance drop on all images. And third, the closer the training degradation distribution P_t is to the test degradation distribution P_r, the better the final result. We next describe the technical designs to achieve our goal.

3.3. Degradation Distribution Binning and Sampling

The representation and quantification of the degradation distribution are essential to determine the training distribution P_t. Existing methods represent the degradation process as a pipeline, *e.g.*, the high-order degradation in Real-ESRGAN. In the pipeline, operations such as random blurring, noise, and compression are sequentially performed on the image. The parameter for each operation is given by a pre-set distribution. This practice makes it extremely difficult to quantify and control its distribution. First, its different operations are performed in sequence, and the previous operations will affect the subsequent operations and get very different results. If we want to update the parameter distribution, it is

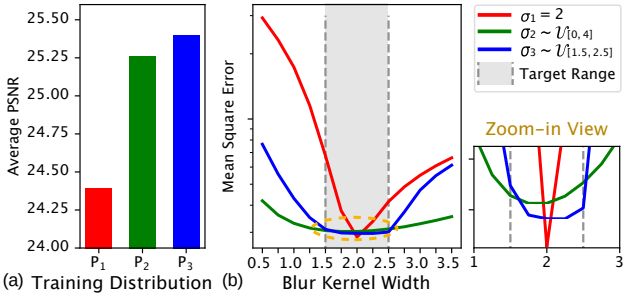

Figure 2. (a) shows the average PSNR performance. (b) shows the performance under different blur conditions. These figures show the importance of the training with appropriate degradation distribution P_t.

difficult for us to attribute the contribution of each step in the degradation. Second, its degradation parameters follow the continuous distributions, which poses challenges for us in conducting statistical estimation or inference.

We propose a binning methodology for discretizing the joint distribution of multiple degradation parameters. This simplifies the representation and sampling process of the degradation distribution. We describe our binning method based on a widely used image degradation model:

$$x = \mathcal{D}(y, d) = \mathcal{C}_q \circ \mathcal{E}_l \circ \mathcal{B}_\sigma(y) = [(y \otimes k) \downarrow + \epsilon]_{\texttt{JPEG}}, \quad (2)$$

where σ, l, and q are parameters, *i.e.*, kernel width for blurring $\mathcal{B}$, noise level for noising $\mathcal{E}$ and quality level for compression $\mathcal{C}$. The distribution of degradation is described by a joint distribution of these parameters $P(d) = p(q, l, \sigma)$. Typically, these parameters are sampled from uniform distributions. In our work, $\sigma \sim \mathcal{U}_{[0,5]}$, $l \sim \mathcal{U}_{[0,50]}$ and $q \sim \mathcal{U}_{[30,90]}$. We divide these continuous distributions evenly into discrete intervals. For example, the compression parameter q is partitioned into three bins as $q_1 \sim \mathcal{U}_{[30,50]}$, $q_2 \sim \mathcal{U}_{[50,70]}$ and $q_3 \sim \mathcal{U}_{[70,90]}$. We use the same binning method for noise and blur and divide it into five equal bins, respectively. Binning these three components of degradation yields $3 \times 5 \times 5 = 75$ possible degradation bins. Then sampling from the entire degradation space can be viewed as first sampling a bin from the set of bins, and then sampling a degradation from this bin. It can be formulated as $P(d) = P(q, l, \sigma) = \sum_b p(q, l, \sigma|b)p(b)$, where b is the random variable of the bin and $p(b)$ is its distribution.

We next change the distribution of the degradations $P(d)$ by changing $p(b)$. Mathematically, we assign a sampling weight (importance) to each bin and formulate $p(b)$ as $p(b_i) = w_i$ for $i \in \{1, 2, \ldots, N_{bin}\}$, where N_{bin} is the number of bins. In the initial stage, we give each bin the same uniform sampling weight, and the result of sampling at this time is equivalent to uniform sampling over the entire interval. By updating the weight vector $\mathbf{w} \in \mathbb{R}^{N_{bin}}$, we can shape the degradation distribution $P(d)$. Figure 3 (a) and (b) show a schematic illustration of this process. It is worth noting that when the number of steps in the degradation process increases, such as using the high-order degradation model, the number of binning increases exponentially. However, we found that lower-order degradation models can already provide good generalization performance and cover most situations, as long as their distributions are well-matched.

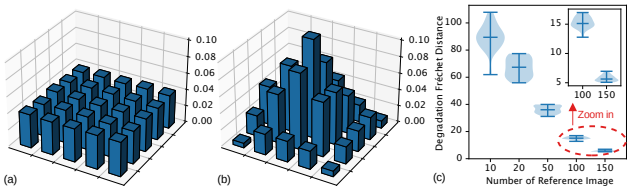

Figure 3. (a) Perform binning operations on multiple degradation parameters and assign uniform sampling weights at the beginning. (b) Updating the sampling weights of the bins in order to sample in the joint distribution of multiple degradation parameters. (c) The estimated degradation distance using different amounts of images.

3.4. Obtaining Weight Using Feature Fréchet Distance

We next describe the method to obtain this binning sampling weight given a set of real degraded images X_{ref}. Recall that in the studied problem, we have access to a small number of real degraded images as references to help estimate the test degradation distribution P_r. Since we only need to adjust $\mathbf{w}$, the probability of sampling from each bin, we need to calculate a kind of distance between the degradation $p(d|b)$ in each bin b and the degradation of P_r. The main difficulty is that we don't have high-quality images with the same content as the X_{ref}. Thus, we introduce another set $Y_w = \{y_i^w\}_{i=1}^n$ of high-quality images to synthesize degraded images $X_w^b = \mathcal{D}(Y_w, p(d|b))$ according to the degradation in each bin b. Then the distance of $p(d|b)$ to P_r is given by the Fréchet distance (Fréchet, 1957) between the deep features $\phi(X_w^b)$ and $\phi(X_{ref})$ of X_w^b and X_{ref}, where $\phi(\cdot)$ is a deep feature extractor, and the features are of size $\mathbb{R}^{n \times c}$ with feature dimension c. Following the common practice of using the Fréchet distance to measure the distance between two deep features (Heusel et al., 2017), we assume that these c dimensional deep features follow their respective c dimensional Gaussian distributions. We give the solution of the Fréchet distance between them as

$$D_F(\phi(X_w^b), \phi(X_{ref}))^2 = \quad (3)$$
$$\|\mu_b - \mu_{ref}\|_2^2 + \mathbf{tr}\big(\Sigma_b + \Sigma_{ref} - 2(\Sigma_b^{\frac{1}{2}} \Sigma_{ref} \Sigma_b^{\frac{1}{2}})^{\frac{1}{2}}\big),$$

where $\mathcal{N}(\mu_b, \Sigma_b)$ is the fitted Gaussian distribution using $\phi(X_w^b)$ and $\mathcal{N}(\mu_{ref}, \Sigma_{ref})$ is fitted using $\phi(X_{ref})$.

We calculate this distance for all the bins and obtain $\mathbf{D} \in \mathbb{R}^{N_{bin}}$, where $\mathbf{D}_i = D_F(\phi(X_w^{bi}), \phi(X_{ref}))^2$. We first normalize the vector $\mathbf{D}_{norm}$ linearly into interval $[0, 1]$. We assign sampling weights to these bins based on $\mathbf{D}_{norm}$ us-

Table 1. The settings of the testing degradation distribution involved in our study.

Degradation	Setting ①	Setting ②	Setting ③	Setting ④
Blur σ	$\mathcal{U}_{[0,1]}$	$\mathcal{U}_{[0.5,1.5]}$	$\mathcal{U}_{[1.5,2.5]}$	$\mathcal{U}_{[2.5,3.5]}$
Noise level l	$\mathcal{U}_{[0,10]}$	$\mathcal{U}_{[15,25]}$	$\mathcal{U}_{[5,15]}$	$\mathcal{U}_{[25,35]}$
JPEG quality q	$\mathcal{U}_{[80,90]}$	$\mathcal{U}_{[75,85]}$	$\mathcal{U}_{[75,85]}$	$\mathcal{U}_{[65,75]}$

ing the following function:

$$w_i = \frac{\exp((1-\mathbf{D}_{norm}[i])^{\alpha}) - 1}{\sum_{j=1}^{N_{bin}}[\exp((1-\mathbf{D}_{norm}[j])^{\alpha}) - 1]}, \tag{4}$$

where α is a hyper-parameter that controls the kurtosis of the distribution. When α is set to be a large value, the resulting distribution will be concentrated in a small range. On the contrary, the formed distribution will be wider, and it degenerates to a uniform distribution when $\alpha = 0$. Equation (4) also ensures that $\sum_{i}^{N_{bin}} p(b_i) = 1$.

3.5. Discussion

We have gotten the full picture of our solution, but some issues are still worth discussing.

The choice of the deep feature extractor ϕ. We use a deep network to extract features to assist in computing the Fréchet distance between degradation distributions. But is this distance computation robust to the choice of feature extractor? Do we need to train a deep feature extractor specifically for this distance? In this subsection, we verify the effectiveness of the proposed degradation Fréchet distance and the difference between feature extractors. We selected the following representative deep feature extractors: $\text{VGG}_{2,2}$[1] is a commonly used extractor for extracting low-level features; $\text{VGG}_{5,4}$, which is often used for extracting deeper features; the DASR degradation representation (Wang et al., 2021a), which is specially used to learn the latent representation for degradation; AlexNet trained using ImageNet dataset (Deng et al., 2009), MINC texture classification dataset (Bell et al., 2015); and a randomly initialized AlexNet. The results are visualized in Figure 4. We arrange these 75 bins regularly and visualize their calculated sampling weights. Weights closer to ground truth are better. It can be seen that these methods can identify the four bins that appear in the ground truth and give them higher weights. However, these feature extractors perform differently on the weight calculation of the remaining bins. Both AlexNet and VGG trained on ImageNet can predict weights with better results. AlexNet trained with MINC can also achieve good results. These methods only assign minor weight to non-target degradation bins. Although DASR is a method specially designed for degradation, its prediction assigns too much weight to irrelevant bins, which may affect its performance. Randomly initialized AlexNet can only show limited effectiveness in matching degradation distributions. In the following research, we mainly use AlexNet because of its good prediction effect and easy availability.

The Use of the Fréchet distance. In this work, the Fréchet distance is used to overcome the influence of different image content on degenerate distance estimation. In the case of image content changes, the traditional element-wise comparison is not applicable anymore. However, the Fréchet distance can produce reasonable results. The Fréchet distance is based on statistics on features, so the number of samples used to estimate this distance is important. We made estimations using different numbers of images and studied the variance of different measurements. We tested each image quantity 25 times with different images and recorded their values. The size of the image is 72×72. The results are shown in Figure 3 (c). It can be seen that when 100 images are used, the randomness of the results obtained by the algorithm is greatly reduced. Although more images can bring better stability, 100 – 150 images can already give good estimation results. In contrast, the previous works that allow reference testing images are based on adversarial learning (Luo et al., 2022), or domain adaptation matching (Wang et al., 2021b). They usually require a large number of reference images and produce unsatisfactory results when the number of reference images is insufficient.

The choice of the distribution range α. Another important parameter in our method is α, which scales the Fréchet distance by a power function to adjust the range of the final degenerate distribution. α is the only parameter in our method that needs to be adjusted manually. A smaller α means a larger degradation range and better generalization performance. But smaller α also leads to lower SR performance. A larger α means a narrower degradation range, which will improve the resulting final performance when the target degradation range is also small. However, the performance degradation faced when exceeding this range is also more severe. α controls the accuracy-generalization trade-off. We show the visualized weights using different αs in Figure 4. In order to verify the impact of different αs on performance, we set a synthetic test degradation interval. And use different αs to obtain the training degradation weights. The baseline is the result of uniform sampling among the bins. The upper bound is the result of direct training on target degradation. As can be seen, as α increases, the range of training degradation becomes smaller, and the test accuracy within the target range is improved. We also include Real-ESRNet and BSRNet for comparison. Their training degradation range is larger than the baseline, so even though they use a better network, their only achieve lower performance than the proposed method.

[1] $\text{VGG}_{i,j}$ is defined as the feature map obtained by the jth convolution (after activation) and before the ith max-pooling layer within the VGG19 (Simonyan & Zisserman, 2014) network

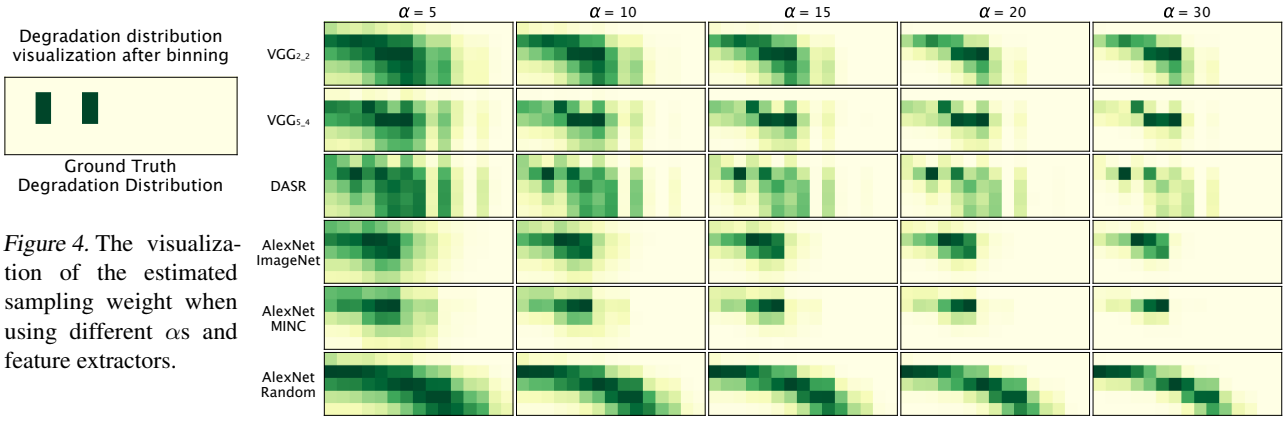

Figure 4. The visualization of the estimated sampling weight when using different αs and feature extractors.

Table 2. The quantitative comparison of different methods with respect to different settings. The upper bound results are marked as grey color to show that a direct comparison of it is unfair to other methods.

Methods	Type	Setting ①			Setting ②			Setting ③			Setting ④		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
RCAN	non-blind	23.65	0.5612	0.4674	18.86	0.2700	0.6509	24.22	0.5186	0.5812	16.54	0.1450	0.7152
KernelGAN	blind	16.04	0.3303	0.7420	12.56	0.1268	0.7675	17.79	0.3048	0.8097	11.34	0.0670	0.7828
IKC	blind	24.76	0.6174	0.4554	20.75	0.3606	0.6235	24.89	0.5607	0.5867	18.87	0.2323	0.6807
DASR	blind	24.81	0.6195	0.4505	20.11	0.3312	0.6432	24.75	0.5557	0.5938	17.64	0.1855	0.7103
FSSR	blind	20.83	0.4166	0.4194	16.14	0.1521	0.6989	20.75	0.3338	0.5189	14.75	0.0872	0.7686
USR-DA	reference	24.98	0.6712	0.3563	22.41	0.4816	0.5154	23.50	0.6076	0.4922	20.73	0.3428	0.6120
PDM-SR	reference	26.96	0.7071	0.3249	24.93	0.6208	0.4251	25.78	0.6433	0.4033	23.47	0.5576	0.4808
BSRNet	blind	27.88	0.7252	0.3431	26.80	0.6908	0.3964	27.48	0.6968	0.3987	25.01	0.6269	0.4916
Real-ESRNet	blind	27.75	0.7316	0.3410	26.75	0.6955	0.3896	27.44	0.6986	0.3994	25.36	0.6309	0.4916
Real-ESRNet-Dropout	blind	26.94	0.7123	0.3783	26.40	0.6777	0.4125	26.88	0.6879	0.4207	24.84	0.6109	0.5040
SRResNet (baseline)	blind	28.14	0.7368	0.3370	27.01	0.6950	0.3873	27.53	0.6980	0.4067	25.34	0.6276	0.4992
Ours (ϕ =VGG$_{2,2}$)	reference	28.85	0.7508	0.3173	27.32	0.7021	0.3861	27.77	0.7039	0.4011	25.50	0.6315	0.4952
Ours (ϕ =VGG$_{5,4}$)	reference	28.82	0.7492	0.3192	27.36	0.7034	0.3824	27.73	0.7037	0.4005	25.43	0.6310	0.4950
Ours (ϕ =DASR)	reference	28.88	0.7518	0.3147	27.27	0.7007	0.3866	27.79	0.7041	0.4005	25.48	0.6307	0.4962
Ours (ϕ =AlexNet-ImageNet)	reference	28.83	0.7511	0.3166	27.42	0.7040	0.3836	27.89	0.7054	0.4002	25.21	0.6158	0.4871
Ours (ϕ =AlexNet-MINC)	reference	28.85	0.7502	0.3193	27.44	0.7042	0.3815	27.99	0.7051	0.4048	25.53	0.6314	0.4959
Ours (ϕ =AlexNet-random)	reference	28.77	0.7467	0.3256	27.44	0.7042	0.3815	27.85	0.7044	0.4020	25.54	0.6316	0.4957
Upper bound	supervised	28.86	0.7529	0.3099	27.44	0.7054	0.3792	27.97	0.7067	0.4002	25.59	0.6321	0.4975
RRDB (baseline)	blind	28.70	0.7532	0.3035	27.45	0.7090	0.3648	27.91	0.7096	0.3875	25.63	0.6358	0.4881
Ours (ϕ =VGG$_{2,2}$)	reference	29.18	0.7615	0.2970	27.66	0.7128	0.3658	28.07	0.7123	0.3863	25.74	0.6381	0.4826
Ours (ϕ =VGG$_{5,4}$)	reference	29.16	0.7610	0.2973	27.73	0.7141	0.3640	28.07	0.7126	0.3856	25.68	0.6375	0.4824
Ours (ϕ =DASR)	reference	29.18	0.7621	0.2957	27.65	0.7125	0.3663	28.08	0.7123	0.3861	25.73	0.6377	0.4832
Ours (ϕ =AlexNet-ImageNet)	reference	29.16	0.7614	0.2963	27.77	0.7142	0.3649	28.17	0.7132	0.3862	25.76	0.6378	0.4834
Ours (ϕ =AlexNet-MINC)	reference	29.18	0.7613	0.2978	27.77	0.7142	0.3645	28.24	0.7133	0.3891	25.79	0.6381	0.4838
Upper bound	supervised	29.15	0.7607	0.2979	27.75	0.7135	0.3680	28.23	0.7135	0.3888	25.84	0.6384	0.4892

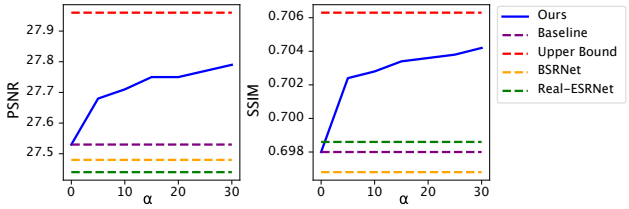

Figure 5. Average PSNR and SSIM performance at different αs. This experiment uses Setting ③ of Table 1 as the target degradation distribution. The network structure is SRResNet. Its feature extractor is DASR.

4. Experiments

4.1. Experiments on the Synthesized Images

Set up. We first evaluate the performance of the proposed method on the synthetic test images. We set four different cases with four different degradation distributions. The detail of these settings is shown in Table 1. We carefully designed these settings to include relatively clean test images (setting ① contains small blur and noise); two cases of moderate degeneration (setting ② contains smaller blur and larger noise and ③ contains larger blur with smaller noise; and severely degeneration (setting ④). We use the PIPAL dataset (Gu et al., 2020) as the reference set Y_{ref} and Y_w. Following the previous study on the number of images used to calculate the Fréchet distance, we set the number of reference images as $n = 100$, and the size of the degraded images is 72×72. Although the degradation is randomly sampled from a distribution, we fix the degraded images during the experiment to eliminate the effect of randomness. In our experiments, we included a total of 75 degradation

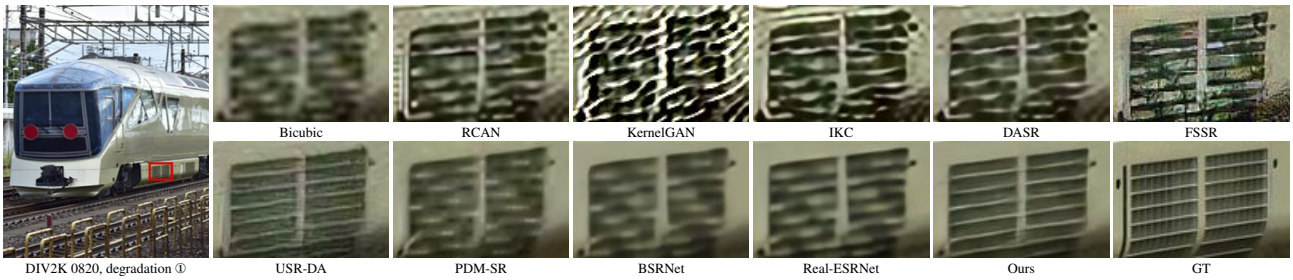

Figure 6. SR results of images from the DIV2K dataset with scale factor $\times 4$.

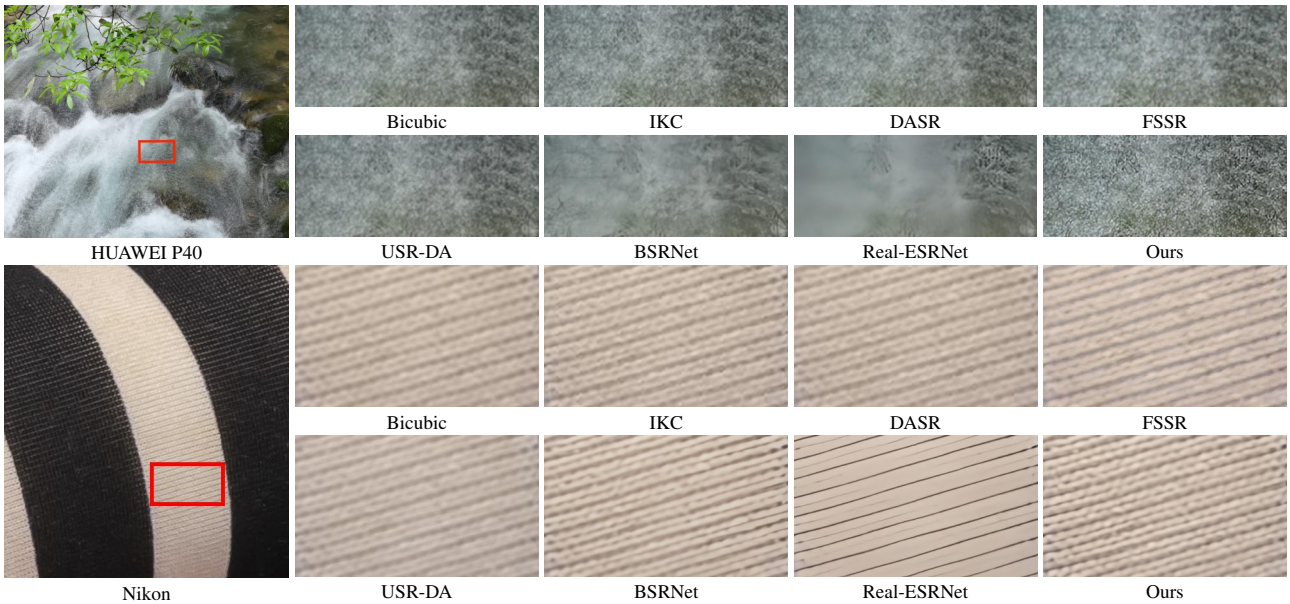

Figure 7. SR results of real cases with scale factor $\times 4$. The first case is from a smartphone camera. The rest two cases are from old films.

distribution bins. For the image blurring, we divide the kernel width into five equal parts between 0 and 5. For the noise, we divide the noise level into five equal parts between 0 and 50. For image compression, divide the quality level into three equal parts between 30 and 90. We set $\alpha = 25$.

Results. We compare the proposed method with several existing methods, including a non-blind method RCAN (Zhang et al., 2018), two blind SR method IKC (Gu et al., 2019) and KernelGAN (Bell-Kligler et al., 2019), DASR with a pre-defined degradation model, the FSSR model trained to maximize the performance on the blurry and noisy dataset, the BSRNet and Real-ESRNet trained with a large range of complex degradations and the Dropout method that proposed to improve generalization performance. We also include two methods with reference images as input, USR-DA (Wang et al., 2021b), and PDM-SR (Luo et al., 2022). Additionally, we also compared the following methods: (1) the baseline model with a uniform sampling from all bins, (2) models with different feature extractors, and (3) the model trained with corresponding test degradation for each setting, which is used to show the upper bound performance in each setting. We test two different backbone architectures: SRResNet and RRDB (Wang et al., 2018). The results are shown in Table 2. It can be seen that, despite the careful design, existing blind problems do not perform well under these stochastic degradations. The reasons include the two aspects mentioned earlier: insufficient generalization for cases beyond the pre-defined degradation models degradation, and the use of a too-large training degradation distribution which leads to a drop in overall accuracy. As for our method, we first provide a comparison of the performance of the baseline model with that of our method. We can see that our method has a substantial performance improvement over the baseline without the degradation range crafting. This is more evident for the clean image test dataset (the setting ①), as training with large degradation is unnecessary for these situations. Methods using different feature extraction all provide good performance improvement compared to the baseline model and other competitive methods, which shows that our method is easy to use. AlexNet trained on ImageNet and MINC outperform others, which is in line with the conclusion in Figure 4.

The qualitative results agree with the conclusions of the numerical results. We show a set of comparisons in the Figure 6. It can be seen that some blind methods are almost

Figure 8. Due to the large range of training blur, Real-ESRNet and BSRNet may unreasonably sharpen the background blur. However, our method does not have this problem.

completely ineffective in this case, such as IKC, KernelGAN and FSSR. Because the degradation models assumed by these methods' design do not match the test degradation. BSRNet and Real-ESRNet have good generalization ability due to training on large-scale and complex degraded datasets, thus obtaining smooth and clear results. However, the training degradation range used by these methods is too large. Degraded training data outside the target range will affect the training, making them unable to generate accurate texture details. Based on the proposed method, our method correctly recovers the pattern and produces a pleasing result. Please refer to the appendix for more results.

4.2. Experiments for Real-World Applications

Our method is designed for real-world scenarios. We demonstrate the value of our research using two valuable scenarios: SR processing of images taken by DSLR (digital single-lens reflex) cameras or mobile phones and old films. The processing of these images is an admittedly difficult task due to the complexities of noise and blur. But we argue that, even though complex, these degradations are not as wide-ranging as we thought. Especially when we limit our processing to only a certain class of sensors or lenses. The range of this test degradation will be further narrowed. This is very suitable for the method proposed in this paper. We test our method on the RealSR dataset (Cai et al., 2019) and our collected mobile phone camera dataset. RealSR contains images taken by Canon 5D3 DSLR cameras. The mobile data is captured by Huawei P40 mobile phone. Furthermore, we also extracted the image frames of two old films as the testing sets. These two films are "Groundhog Day" and "My Fair Lady". We show some visual results in Figure 7.

As one can see, our results show excellent sharpness and detail restoration. Some methods, such as IKC and DASR, lead to ambiguous results due to the mismatch of their degradation models. An example is the image from "Groundhog Day" in Figure 7. The results obtained by these two methods

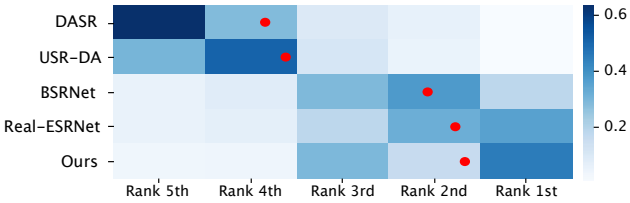

Figure 9. This figure shows the normalized histogram of votes in the user study. The average score is also shown in red dots.

are still blurry, while our method is able to recover the sharp edges. BSRNet and Real-ESRNet can handle a wide range of degradations at the cost of reduced accuracy. However, due to the larger range of noise and blur used for training, the network tends to reduce all noise-like textures. This makes it unable to recover some subtle textures and generate over-smooth results. This drawback is evident in the case of the image captured by Huawei P40 in Figure 7. A large area of the dense texture is removed by BSRNet and Real-ESRNet, resulting in over-smooth results, while our method can restore sufficient texture. For the testing image from "My Fair Lady", since with closer training degradation distribution, only our method recovers the correct pattern of the image, other methods all result in incorrect patterns, especially USR-DA and BSRNet, which directly recover the circular pattern into lines. Another problem with the existing methods is the processing of background blur. As an artistic technique, background blur often appears on film screens. Due to the large range of training blur, Real-ESRNet and BSRNet may unreasonably sharpen the background blur, as shown in Figure 8. However, our method does not have this problem. These results demonstrate that our method achieves pleasing edges and effects while preserving detail, and also matches the correct pattern of the image.

Due to the lack of reasonable quantitative measures for comparing real images. We conducted a user study for some representative methods. We show the results of the five methods and ask the user to rank them. In total, our research contains more than 20 images from different scenes and sources. More than 30 users participated in our user study shown in Figure 9. Our method was ranked first most times, and its average score also outperformed other methods.

5. Conclusion and Limitation

This paper describes a method to craft the training degradation distribution for real-world SR applications. Our method can improve its performance while maintaining the generalization ability of SR. One of the limitations of our work is that the number of bins increases exponentially as the degradation model becomes more complex.

Acknowledgements This work was partly supported by the National Natural Science Foundation of China (Nos. 62276251, 62171251 and 62311530100), the Special Foundations for the Development of Strategic Emerging In-

dustries of Shenzhen (Nos.JCYJ20200109143010272 and CJGJZD20210408092804011), Oversea Cooperation Foundation of Tsinghua, the Joint Lab of CAS-HK, and in part by the Youth Innovation Promotion Association of Chinese Academy of Sciences (No. 2020356).

References

Bell, S., Upchurch, P., Snavely, N., and Bala, K. Material recognition in the wild with the materials in context database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3479–3487, 2015.

Bell-Kligler, S., Shocher, A., and Irani, M. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019.

Cai, J., Zeng, H., Yong, H., Cao, Z., and Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3086–3095, 2019.

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., and Gao, W. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12299–12310, 2021.

Chen, H., Gu, J., Liu, Y., Magid, S. A., Dong, C., Wang, Q., Pfister, H., and Zhu, L. Masked image training for generalizable deep image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023a.

Chen, X., Wang, X., Zhou, J., Qiao, Y., and Dong, C. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22367–22377, 2023b.

Chen, Z., Zhang, Y., Gu, J., Kong, L., Yuan, X., et al. Cross aggregation transformer for image restoration. *Advances in Neural Information Processing Systems*, 35:25478–25490, 2022.

Cornillere, V., Djelouah, A., Yifan, W., Sorkine-Hornung, O., and Schroers, C. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics*, 38(6):1–13, 2019.

Dai, T., Cai, J., Zhang, Y., Xia, S.-T., and Zhang, L. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11065–11074, 2019.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

Dong, C., Loy, C. C., He, K., and Tang, X. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015.

Dong, C., Loy, C. C., and Tang, X. Accelerating the super-resolution convolutional neural network. In *Proceedings of the European Conference on Computer Vision*, pp. 391–407. Springer, 2016.

Fréchet, M. Sur la distance de deux lois de probabilité. *Comptes Rendus Hebdomadaires des Seances de L Academie des Sciences*, 244(6):689–692, 1957.

Fritsche, M., Gu, S., and Timofte, R. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop*, pp. 3599–3608. IEEE, 2019.

Gu, J., Lu, H., Zuo, W., and Dong, C. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019.

Gu, J., Cai, H., Chen, H., Ye, X., Ren, J., and Dong, C. Pipal: a large-scale image quality assessment dataset for perceptual image restoration. In *Proceedings of the European Conference on Computer Vision*, pp. 633–651. Springer, 2020.

Gu, J., Cai, H., Dong, C., Zhang, R., Zhang, Y., Yang, W., and Yuan, C. Super-resolution by predicting offsets: An ultra-efficient super-resolution network for rasterized images. In *Proceedings of the European Conference on Computer Vision*, pp. 583–598. Springer, 2022.

Gu, J., Ma, X., Kong, X., Qiao, Y., and Dong, C. Networks are slacking off: Understanding generalization problem in image deraining. *arXiv preprint arXiv:2305.15134*, 2023.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30, 2017.

Huang, Y., Li, S., Wang, L., Tan, T., et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, 33: 5632–5643, 2020.

Kim, J., Kwon Lee, J., and Mu Lee, K. Deeply-recursive convolutional network for image super-resolution. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1637–1645, 2016a.

Kim, J., Kwon Lee, J., and Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646–1654, 2016b.

Kong, X., Liu, X., Gu, J., Qiao, Y., and Dong, C. Reflash dropout in image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6002–6012, 2022.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 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*, pp. 4681–4690, 2017.

Li, Z., Liu, Y., Chen, X., Cai, H., Gu, J., Qiao, Y., and Dong, C. Blueprint separable residual network for efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 833–843, 2022.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844, 2021.

Liu, A., Liu, Y., Gu, J., Qiao, Y., and Dong, C. Blind image super-resolution: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022a.

Liu, Y., Liu, A., Gu, J., Zhang, Z., Wu, W., Qiao, Y., and Dong, C. Discovering" semantics" in super-resolution networks. *arXiv preprint arXiv:2108.00406*, 2021.

Liu, Y., Zhao, H., Gu, J., Qiao, Y., and Dong, C. Evaluating the generalization ability of super-resolution networks. *arXiv preprint arXiv:2205.07019*, 2022b.

Luo, Z., Huang, Y., Li, S., Wang, L., and 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.

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

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Tai, Y., Yang, J., and Liu, X. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3147–3155, 2017.

Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., and Guo, Y. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10581–10590, 2021a.

Wang, W., Zhang, H., Yuan, Z., and Wang, C. Unsupervised real-world super-resolution: A domain adaptation perspective. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4318–4327, 2021b.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision workshops*, 2018.

Wang, X., Xie, L., Dong, C., and Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1905–1914, 2021c.

Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., and Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 701–710, 2018.

Zhang, K., Liang, J., Van Gool, L., and Timofte, R. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4791–4800, 2021.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision*, pp. 286–301, 2018.

Zhou, L., Cai, H., Gu, J., Li, Z., Liu, Y., Chen, X., Qiao, Y., and Dong, C. Efficient image super-resolution using vast-receptive-field attention. In *Proceedings of the European Conference on Computer Vision Workshops*, pp. 256–272. Springer, 2023.

Zhou, Y., Deng, W., Tong, T., and Gao, Q. Guided frequency separation network for real-world super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 428–429, 2020.

A. More Results

We provide more qualitative results to show the effectiveness of our method.

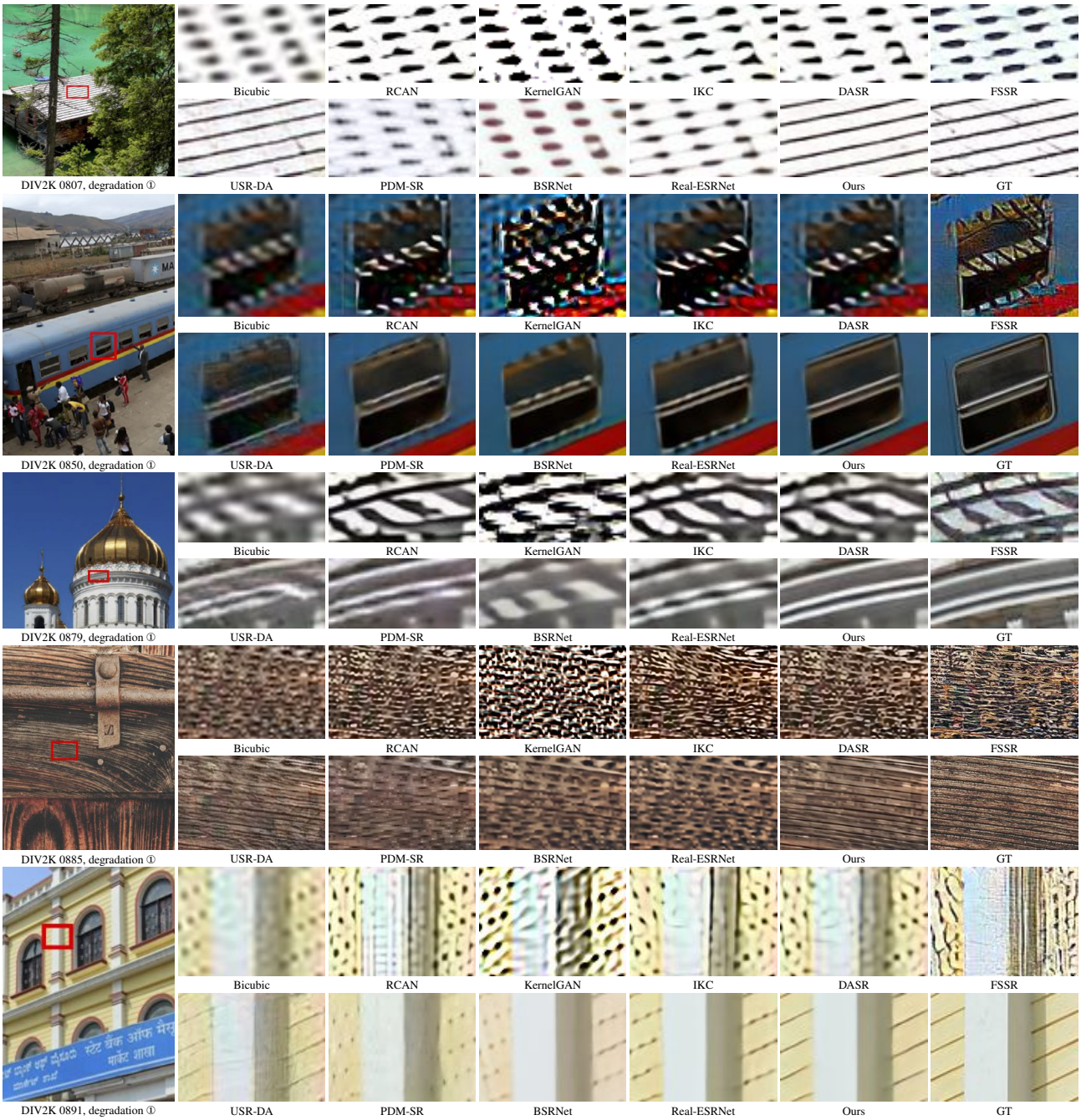

Figure 10. SR Results of synthesized testing images with scale factor $\times 4$.

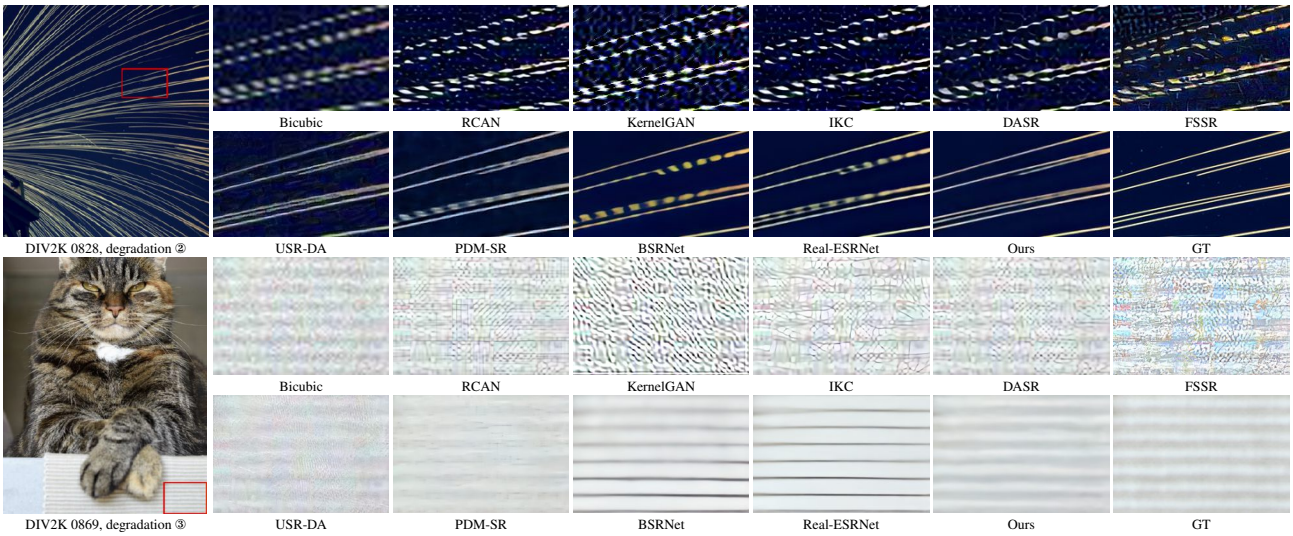

Figure 11. SR Results of synthesized testing images with scale factor $\times 4$.

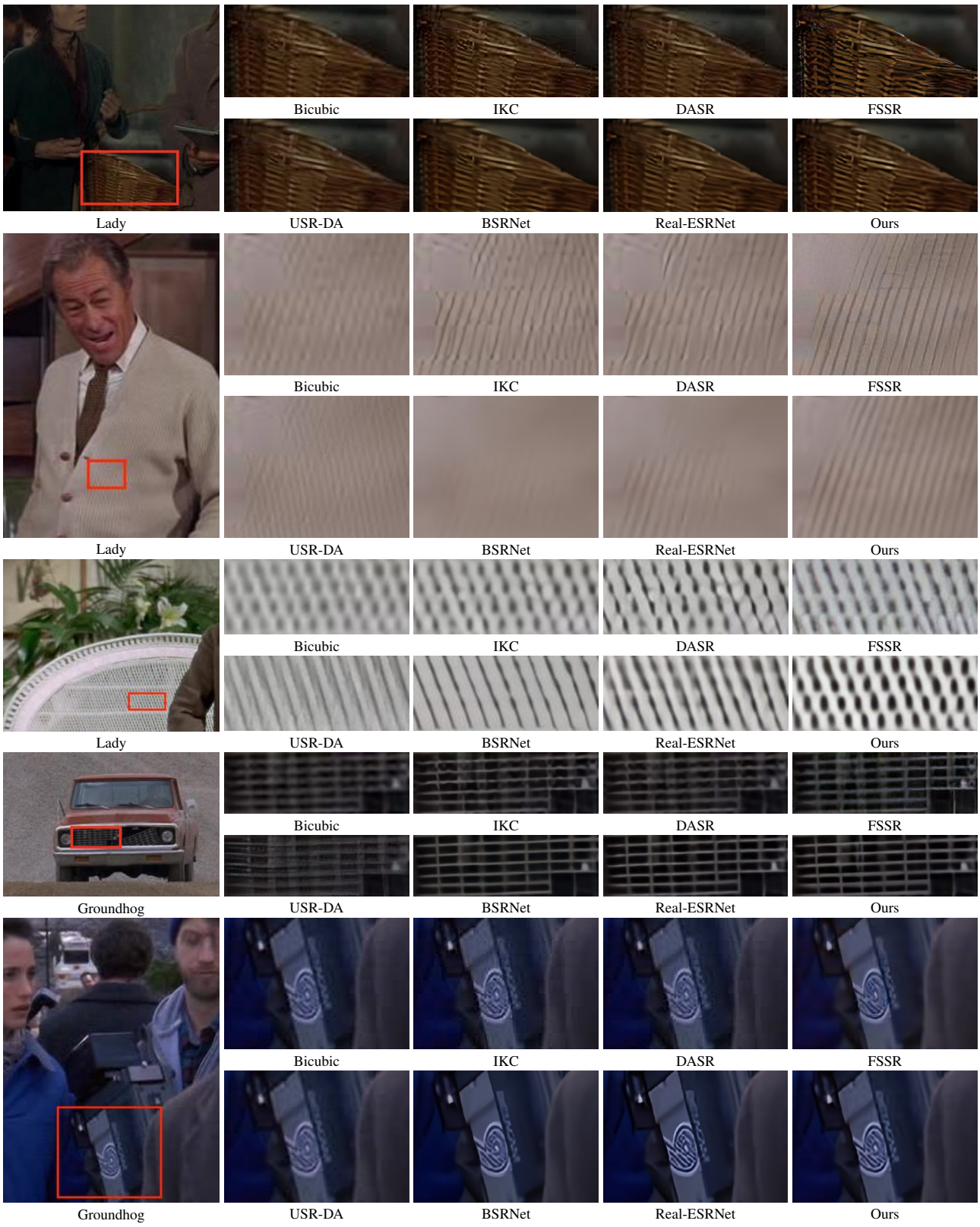

Figure 12. SR results of real-world images with scale factor $\times 4$.

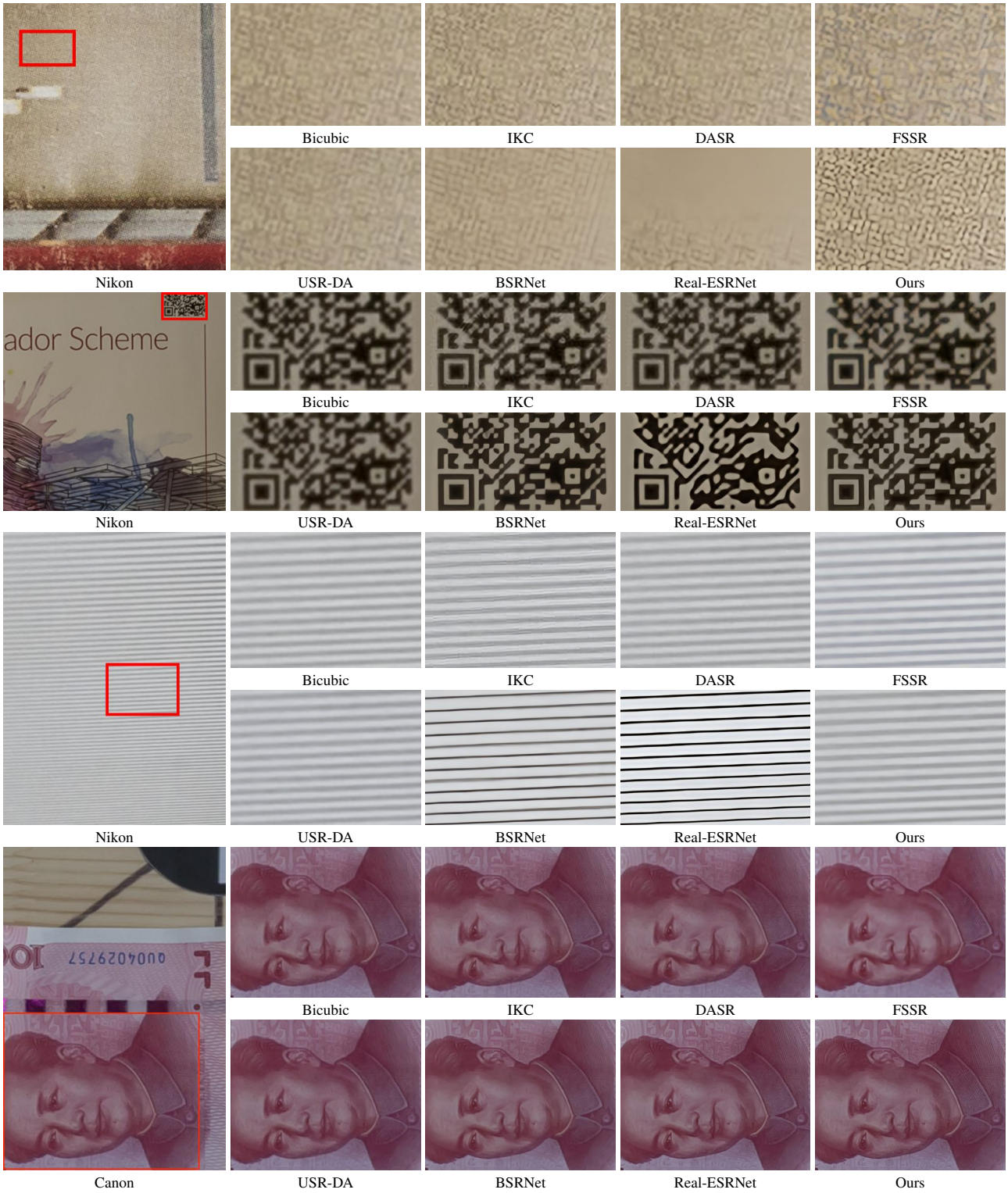

Figure 13. SR results of real-world images with scale factor $\times 4$.