

Deep Degradation Prior for Real-World Super-Resolution

Kyungdeuk Ko¹
kdko@korea.ac.kr

Bokyeung Lee¹
bksain@korea.ac.kr

Jonghwan Hong¹
jhong2661@korea.ac.kr

David Han²
dkh42@drexel.edu

Hanseok Ko¹
hsko@korea.ac.kr

¹ Korea University
Seoul, Republic of Korea

² Drexel University
Philadelphia, USA

Abstract

Real-world Super-Resolution (SR) is a very challenging task to reconstruct a higher resolution image from a real-world image which generally has unexpected artifacts and distortions. Most of the high performance SR methods rely on availability of LR-HR paired datasets, target domain images, or degradation priors. These information, however, are usually not available in real world, thus these methods are not often practical for real world obtained images. Recent studies related to real-world SR mainly focus on constructing the LR-HR paired dataset. The methods estimate the noises and the blur kernels from real-world images to generate a new training set. However, these methods use the degradation only for dataset construction. In this paper, we propose a novel real-world SR method called Deep Degradation Prior-based SR (DDP-SR). Upon completion of training, denoising network and kernel estimation network within DDP-SR becomes capable of extracting degradation representation of any given input image. Thus, the model works regardless of whether the input image from the same or the different domain of the source images. As such, DDP-SR achieves generalization performance on images from different domain while it also outperforms the state-of-the-art methods qualitatively and quantitatively.

1 Introduction

Super-Resolution (SR) is a low-level image processing task to increase resolution and to enhance the quality of images. In general, it is assumed that Low-Resolution (LR) images are obtained from High-Resolution (HR) images through the following degradation process:

$$I_L = (I_H \otimes k) \downarrow_s + n, \quad (1)$$

where I_H is an HR image and I_L is an LR image. k and n represent a blur kernel and additive noise, respectively. s denotes a scale factor for downsampling. The goal of SR is to find an inverse operator to convert I_L to I_H . With the development of deep learning, many studies using LR-HR image pairs made of Eq.(1) have emerged [6, 6, 20, 24, 25, 26, 41, 45, 51, 52]. However, HR images are generally downsampled by a bicubic kernel in these methods, and the SR models are trained by minimizing the difference between the reconstructed images and the original images as:

$$\arg \min_{SR} \|SR(I_L) - I_H\|. \quad (2)$$

Consequently, since these SR models are optimized for the bicubic kernel, they generalize poorly on images degraded by other causes.

Several learning-based methods have attempted to solve this problem. One of these methods applies the idea of the domain transfer. In these efforts, LR images reside in the source domain with unknown degradation while the target domain contains clean HR images. Cycle-in-cycle Generative Adversarial Network (CinCGAN), applied by Yuan *et al.* [47], Lugmayr *et al.* [27], and Kim *et al.* [49], sequentially applied two CycleGANs [54] with the first CycleGAN transferring the noisy LR domain to a clean LR domain, and the second CycleGAN transferring the clean LR domain to a clean HR domain. These SR models need target domain images for training, but the target domain images generally could not be acquired in real world.

Another method is to train by self-supervised learning using only test images without any training images. Zero-Shot Super-Resolution (ZSSR) [53] is trained by the internal recurrence of information within a single image. Therefore, ZSSR can be applied to images degraded from a variety of kernels, including bicubic kernels. Since the ground truth kernel of the test image is not known, ZSSR would exhibit poor performance if the predicted kernel and the actual kernel turned out to be different. To compensate for these shortcomings, a blur kernel corresponding to the test image is obtained by using kernel estimation methods such as KernelGAN [2]. With the difficulty of estimating the kernel, ZSSR implemented with KernelGAN often shows poor performance, as we have observed in our experiments.

SR methods relying on LR-HR paired dataset or prior knowledge on degradation have progressed and now are able to deliver some impressive results. However, when the process associated with transforming HR to LR, or the source of noise added in LR images are not known, SR becomes much more challenging. A typical SR method based on supervised learning by paired images is not possible in such cases. In New Trends in Image Restoration and Enhancement (NTIRE) 2020 challenge [50], track 2 of real-world SR is one of these problems without the target domain and without any prior knowledge on noise and blur kernel in LR images. The task provides real-world images taken by smartphones, not by high-performance cameras such as DSLRs. As such, the images are more noisy and blurry due to the phone’s limited hardware and simpler circuitry. RealSR [46], which is trained by creating a degradation pool composed of the noises and kernels extracted from variance based noise cropping method and KernelGAN, won the NTIRE 2020 real-world SR competition. Since the training dataset is created by reflecting the noise and blur kernel of the source domain, the SR model trained with the generated dataset is optimized for the source domain. However, the degradation pool is only used when constructing the training set, so it is difficult for RealSR to reflect the prior for actual testing. If an image with domains different from the trained source domain is tested, poor performance is achieved.

To solve the aforementioned problem in the SR task, we propose a novel method of deep degradation prior to represent the degradation. We developed a novel structure of a

deep prior generator containing a denoising network and a kernel estimation network. The output of the deep prior generator, termed deep degradation prior, is an informative feature representation that strongly corresponds dominant components of the degradation. The representation composed of noise and kernel estimations allow the SR model to generate realistic and clean SR images by an improved understanding of the degradation in the observed images. The novelty we stress is that once the training is done, the deep prior generator is capable of composing representation of degradation involved in any given image. Thus, the trained deep prior generator is capable of dealing with input image domains different from that of the training images. Our proposed deep prior generator, therefore, would generalize well to a variety of input domains. The deep prior generator is trained with the degradation pool to extract the deep degradation prior from the observation image. Our proposed network is implemented for the SR task through an adaptive method based on Adaptive Instance Normalization (AdaIN) [12, 45, 44] and channel attention [10]. We term the overall SR architecture proposed here "Deep Degradation Prior-based SR (DDP-SR)."

Our overall contributions can be summarized as follows:

- We propose a novel real-world SR method called DDP-SR to improve the performance by using the deep degradation prior corresponding to noise and kernel of observation image.
- We propose a novel deep prior generator capable of extracting degradation components of any image on the fly once its training is completed. Our design is rooted on enabling the model to generalize to any given input domain.
- We developed a novel learning architecture for the deep prior generator composed of the denoising network and the kernel estimation network.
- DDP-SR achieves higher performance both quantitatively and qualitatively than the state-of-the-art real-world SR methods in experiments using public datasets.

2 Related Work

2.1 Noise Estimation

Recently, many studies on Convolutional Neural Network (CNN)-based denoising has been conducted. Convolutional Blind Denoising Network (CBDNet) [8] is made up of a noise estimation network and a non-blind denoising network. The noise estimation network is simply composed of five layer convolutional layers without pooling layer and normalization operation and extracts noise level map from the input image. The non-blind denoising network has residual learning [10] and U-Net [57] architecture which takes both the original noisy image and the noise level map extracted by the noise estimation network. Consequently, the non-blind denoising network outputs the clean image. Adaptively Tuned Denoising Network (ATDNet) [21] has an overall similar architecture to the CBDNet, but the noise level map is applied to the denoising subnetwork differently from the CBDNet. Since the noise level map is used as input in the middle of the denoising network by gate-residual block, the gate residual block makes the denoising network adaptive to the change of the noise level. Adaptive Instance Normalization Denoising Network (AINDNet) [22] also has an architecture similar to the the ATDNet. Instead of the gate residual block to apply the noise level map, the AINDNet uses adaptive instance normalization residual block based on AdaIN. The

denoising networks mentioned above have in common that they have residual learning [48]. However, these denoising models are not suitable for real-world image noise because they learn Gaussian noise-based noise levels. Based on these previous studies, we design the deep prior generator that learns the noise information of the real-world image and apply the deep degradation prior to the SR model.

2.2 Blind Super-Resolution

Blind SR estimates the blur kernels from the dataset, and enhance the SR performance by using the blur kernels. Since the blur kernel is an important factor in SR performance, there have been some efforts focused on finding an accurate kernel. Michaeli *et al.* [50] first introduced the concept of the blind SR. They estimate the blur kernels based on Maximum a Posteriori (MAP). Xu *et al.* [46] first proposed the blind SR used deep learning. They use GAN for the blind SR of face image and text image without the kernel estimation. Kernel Modeling Super-Resolution (KMSR) [53] uses blur kernel estimation algorithm based on dark channel prior [9, 54]. But this algorithm is very computationally heavy, and target domain images are required to construct the LR-HR pairs. Iterative Kernel Correction (IKC) [7] estimates the blur kernel using predictor network, and updates the estimated kernel using corrector network. However, IKC needs a prior to predict the accurate blur kernel. KernelGAN based on Internal-GAN (InGAN) [39] estimates the blur kernel. The generator simply consists of 6 convolutional layers without a non-linear activation function. The discriminator compares the patch of the input image with the patch of the image downsampled by the generator.

2.3 Real-World Super-Resolution

Real-world SR is a task that makes a blurry and noisy image taken with a real camera or smartphone into a clean HR image. Therefore, there is no LR-HR paired dataset. Furthermore, super-resolution without target domain images is a very challenging task. Many studies have been conducted to solve this problem. Super-Resolution network for Multiple Degradations (SRMD) [49] reduces the dimension of the blur kernel and then concatenates it with the noise level. The degradation representation is stretched and concatenated with the input image. SRMD uses a grid search strategy to estimate the blur kernel and noise level of the real image. However, the strategy is inefficient because it requires a lot of computation to calculate, and it is also inaccurate because it is estimating among the pre-set blur kernel and noise. Ren *et al.* [56] uses various downsampling, blur kernel, and noise to make synthetic LR images similar to real-world LR images. However, in order to construct these LR images, the target domain dataset is required. On the other hand, RealSR composes the LR-HR paired dataset using only the source domain dataset. However, the degradation information is used only to compose the data set, and the information is not used as an important prior. We proposed DDP-SR that reflects the deep degradation prior to the input real-world image to the SR model. Section 3 describes DDP-SR in detail.

3 Proposed Method

Figure 1 displays our proposed method called DDP-SR. The method is divided into two main parts in dealing with a situation where target domain images don't exist. The first is

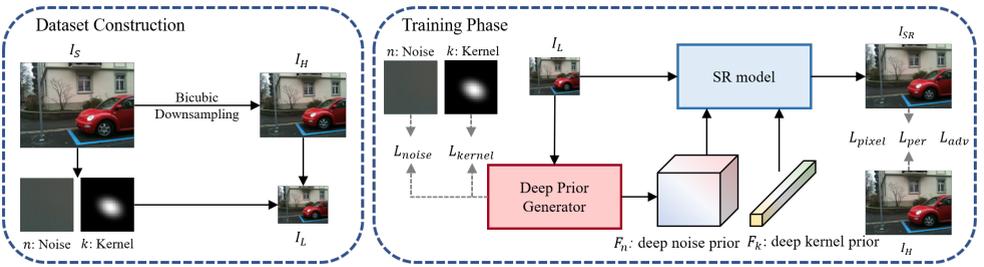


Figure 1: Illustration of our proposed method called DDP-SR. The figure on the left shows the process of constructing a dataset, and the figure on the right shows the process of training DDP-SR based on the constructed data.

to construct the training dataset using only source images. The other is to train the network based on the constructed data.

3.1 Dataset Construction

In the dataset construction, the LR-HR pairs are obtained by using only source images I_S . First, the HR image I_H is produced by using bicubic downsampling that removes noise and makes the source image sharper. The noise n and the kernel k of the source image can be estimated by variance based noise cropping [17] and KernelGAN. In order to generate the LR image I_L with a similar domain to the source image, the noises and the blur kernels collected from the source image are randomly selected, and I_L is obtained by Eq.(1) The constructed LR-HR pairs and the degradation information estimated from source images are used to train the SR model and the deep degradation prior generator.

3.2 Deep Degradation Prior Generator

The deep degradation prior generator consists of two modules and delivers prior information into Residual Dense Block (RDB) [52] as shown in Figure 2. One is a deep noise prior generator and the other is a deep kernel prior generator.

3.2.1 Deep Noise Prior

The deep noise prior generator is designed as the CNN based on residual learning. The CNN is composed of five convolutional layers, Rectified Linear Unit (ReLU) activation function [33], and the skip connection for the residual learning. The input of the deep noise prior generator is the LR image, and the label $I_{clean} = (I_H \otimes k) \downarrow_s$ is the corresponding image before applying the noise prior to the dataset construction stage. For training the deep noise prior generator, we define the noise prior loss, $L_{noise} = \|NG(I_L) - I_{clean}\|_2$, where NG denotes the deep noise prior generator. By applying residual structure for denoising, CNN layers are trained to estimate residual noise. The estimated raw noise map is not adequate as prior knowledge because it is too simple information. So we use encoded feature map which contains noise information.

The deep noise prior $F_n \in \mathbb{R}^{W_I \times H_I \times C}$, where W_I and H_I are the width and height of the input image, and C is the number of channels, is applied to the SR model through AdaIN

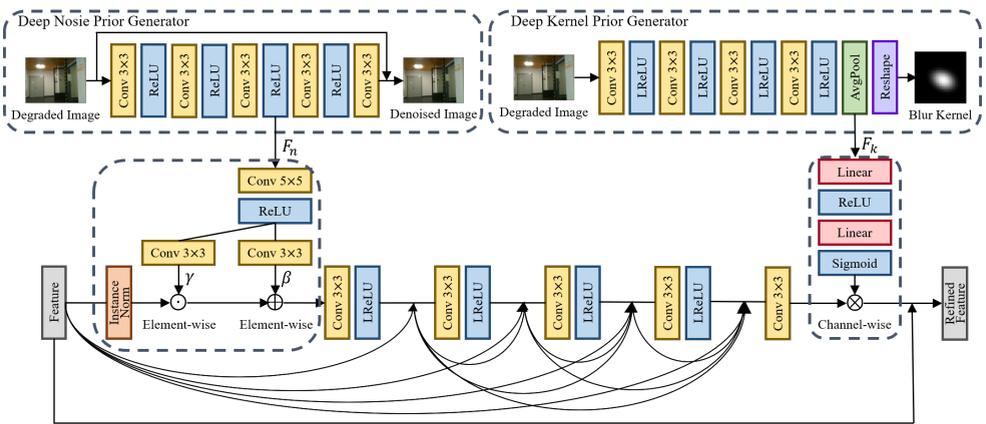


Figure 2: Illustration of the deep prior generator and the example where the deep degradation prior extracted from the deep prior generator is applied to the RDB.

[17, 65, 42, 44]. injecting the deep noise prior into the SR model via AdaIN can prevent overfitting to the noise on the training set and regularize the network. For the deep noise prior, AdaIN is composed of convolutional layers and instance normalization (IN) with $h'_j = \gamma \odot IN(h_j) + \beta$ where h_j denotes j -th-mid level features, IN means IN operation, and \odot represents element-wise multiplication. Also, γ and β are obtained through the deep noise prior and the convolutional layers as in Figure 2.

3.2.2 Deep Kernel Prior

The deep kernel prior generator is composed of four convolutional layers, Leaky ReLU (LReLU) activation function [29], and a global average pooling. The size of the pooling result is the same as $W_k \times H_k$ of the blur kernel used in the dataset construction stage. After the pooling result is reshaped to the shape of the blur kernel, we use the kernel prior loss, $L_{kernel} = \|KG(I_L) - k\|_2$, where KG represents the deep kernel prior generator and k means the blur kernel that is used when generating the degraded image I_L . For deep kernel prior $F_k \in \mathbb{R}^{1 \times 1 \times (W_k \times H_k)}$, channel attention is similar to AdaIN in the deep noise prior. Channel attention is used to apply the deep kernel prior to global areas of features. In our case, the kernel size is set to 33×33 .

3.3 Deep Degradation Prior for Super-Resolution

The deep degradation prior extracted from the deep prior generator can be applied to various SR models. In this case, Enhanced Super-Resolution GAN (ESRGAN) [45] is used as the SR model. The performance of ESRGAN has not only been proven in several studies [17, 66], but also empirically showed the best in our case. The generator of ESRGAN is composed of Residual-in-Residual Dense Block (RRDB) which combines multi-level residual network and dense network. The deep degradation previously mentioned in Section 3.2 is applied to each RDB in each RRDB. Based on the inverse operation of Eq. 1, the deep noise prior is applied to the beginning of the RDB and the deep kernel prior is applied to the end of RDB. For efficient learning, the pre-trained SR generator and discriminator are used. Pixel loss,



Figure 3: Qualitative results on the DPED dataset

perceptual loss [47], and adversarial loss are used to train the generator. The pixel loss is the distance between I_H and the output of the SR model I_{SR} , and the perceptual loss is the distance in the feature space of CNN. The pixel loss preserves the overall HR image content, and the perceptual loss improves high-frequency components. In this case, both the pixel loss and the perceptual loss use L1 distance, and VGG-19 [48] is used for perceptual loss. The adversarial loss makes the generated image more realistic by enhancing texture details. The final loss is defined as the weighted sum of the aforementioned losses as follows:

$$L_{total} = \lambda_{per}L_{per} + \lambda_{pixel}L_{pixel} + \lambda_{noise}L_{noise} + \lambda_{kernel}L_{kernel} + \lambda_{adv}L_{adv}, \quad (3)$$

where λ_{per} and λ_{adv} are set as 1 and 0.005, and λ_{pixel} , λ_{noise} , and λ_{kernel} are all set as 0.01.

Patch discriminator [45, 54] is used to determine whether each patch is real or fake, instead of the original discriminator of ESRGAN. The patch discriminator makes a final decision through average operation after determining whether it is real or fake for each patch of the image. Consequently, it can work regardless of the size of the image. The SR generator and the discriminator are fine-tuned for 60000 iterations by using Adam optimizer [23] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ without weight decay. The learning rate is initially set to 0.0001 and halved at {5000, 10000, 20000, 30000} iterations.

4 Experiment

4.1 Datasets

The DSLR Photo Enhancement Dataset (DPED) [43] is used for real-world SR experiments. The DPED is the dataset taken with four different cameras, but only pictures taken with iPhone 3GS are used this time. The training set contains only 5614 source domain images. The test set includes 100 images cropped from the original images. Since there is no Ground Truth (GT), the performance needs to be evaluated based on the no-reference method. Therefore, Natural Image Quality Evaluator (NIQE) [52], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [53], Perception-based Image Quality Evaluator

Dataset	Method	PNSR \uparrow	SSIM \uparrow	NIQE \downarrow	BRISQUE \downarrow	PIQE \downarrow	NRQM \uparrow	PI \downarrow	DISTS \downarrow
DPED	Bicubic	-	-	5.5426	53.2307	84.9406	3.1220	6.2103	-
	SRMD	-	-	5.4809	50.2295	75.7359	4.0941	5.6934	-
	IKC	-	-	5.6268	51.2808	76.4175	4.2140	5.7064	-
	ZSSR	-	-	6.7644	42.1918	51.3160	5.2201	5.7722	-
	K-ZSSR	-	-	10.2671	43.8575	44.2234	4.7748	7.7462	-
	RealSR	-	-	4.1223	23.3344	14.0526	6.5217	3.8003	-
	Ours	-	-	3.1137	21.4566	13.6320	6.6546	3.5296	-
	RealSR-V3	Bicubic	27.2291	0.8206	6.0751	56.2801	91.5781	2.7174	6.6789
SRMD		27.5849	0.8303	6.1688	51.5007	90.2800	3.0248	6.5720	0.1287
IKC		27.1788	0.8331	5.0329	48.7129	83.7245	3.8288	5.6021	0.1117
ZSSR		25.5300	0.6603	7.2244	43.4438	60.3120	4.7165	6.2540	0.1285
K-ZSSR		23.1509	0.6262	7.2699	43.4680	56.2521	4.8461	6.2119	0.1385
RealSR		24.0146	0.7618	3.3585	30.4957	23.5409	6.4980	3.4303	0.1218
Ours		24.2664	0.7595	3.0353	28.2168	21.9210	6.6330	3.2012	0.1192
DIV2K		Bicubic	25.7313	0.7932	6.0341	58.5037	93.0483	2.6416	6.6963
	SRMD	26.0145	0.8014	6.0426	56.9209	89.9418	2.8700	6.5863	0.0825
	IKC	27.6948	0.8535	5.2750	46.9631	81.5228	4.2163	5.5294	0.0286
	ZSSR	24.1092	0.6718	6.5096	42.0896	60.5725	4.9242	5.7927	0.0822
	K-ZSSR	24.8064	0.7090	5.9446	43.2806	55.6396	5.1476	5.3989	0.0345
	RealSR	26.4570	0.8181	3.3536	24.5535	21.0390	6.4448	3.4544	0.0293
	Ours	26.2967	0.8112	3.1636	22.9385	19.9166	6.5047	3.3295	0.0289
	CelebA-HQ	Bicubic	34.1167	0.9550	5.8163	56.2974	97.9132	2.2373	6.7895
SRMD		34.4920	0.9571	5.7378	57.3826	95.6500	2.4405	6.6487	0.0777
IKC		36.7090	0.9691	5.0668	49.2875	80.0878	3.5369	5.7650	0.0305
ZSSR		27.6223	0.7142	7.3876	42.8626	66.6122	5.2340	6.0768	0.0756
K-ZSSR		27.8847	0.7239	6.1327	44.1258	61.1679	5.3011	5.4158	0.0678
RealSR		33.2764	0.9413	3.3758	13.1447	12.1853	6.8005	3.2877	0.0402
Ours		33.8797	0.9450	3.1111	10.4859	12.1332	6.7738	3.1687	0.0346

Table 1: Quantitative results on the DPED dataset, RealSR-V3 dataset, DIV2K dataset, and CelebA-HQ dataset.

(PIQE) [13], No-Reference Quality Metric (NRQM) [18], and Perceptual Index (PI) [14] are used in our experiments on the DPED.

Additionally, in order to verify the generalization performance of DDP-SR, the DPED-trained models are evaluated through RealSR-V3 [3], DIV2K [11], and CelebA-HQ [18]. The RealSR-V3 provides a 100 LR-HR paired test set taken with cameras different from the one used in the DPED. The images collected on different scales by zooming the lens of the camera are made of the LR-HR paired dataset by distortion correction, central region crop, and iterative pixel-wise registration algorithm considering luminance adjustment. DIV2K and CelebA-HQ are used to evaluate generalization performance on datasets with different domain and high quality images. In DIV2K, 100 images from the validation set are used, and in CelebA-HQ, 100 images are randomly selected and tested. Both datasets are tested by applying blur kernel and noise. Consequently, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Map (SSIM), and Deep Image Structure and Texture Similarity (DISTS) [11], which is a recently proposed perceptual similarity evaluation metric, can be used for the performance evaluation while comparing the reconstructed image and GT.

4.2 Experimental Results

To compare DDP-SR performance, we perform baseline experiments such as bicubic, SRMD, IKC, ZSSR, K-ZSSR which means a combination of KernelGAN and ZSSR, and RealSR that is the state-of-the-art method for real-world SR. As shown in Table 1, DDP-SR achieves the best performance in all evaluation metrics. Except for RealSR, the baseline performance is quantitatively very poor. Especially, NIQE values of ZSSR, K-ZSSR, and IKC are worse than the bicubic method. Figure 3 displays the outputs of DPP-SR and the baseline methods.

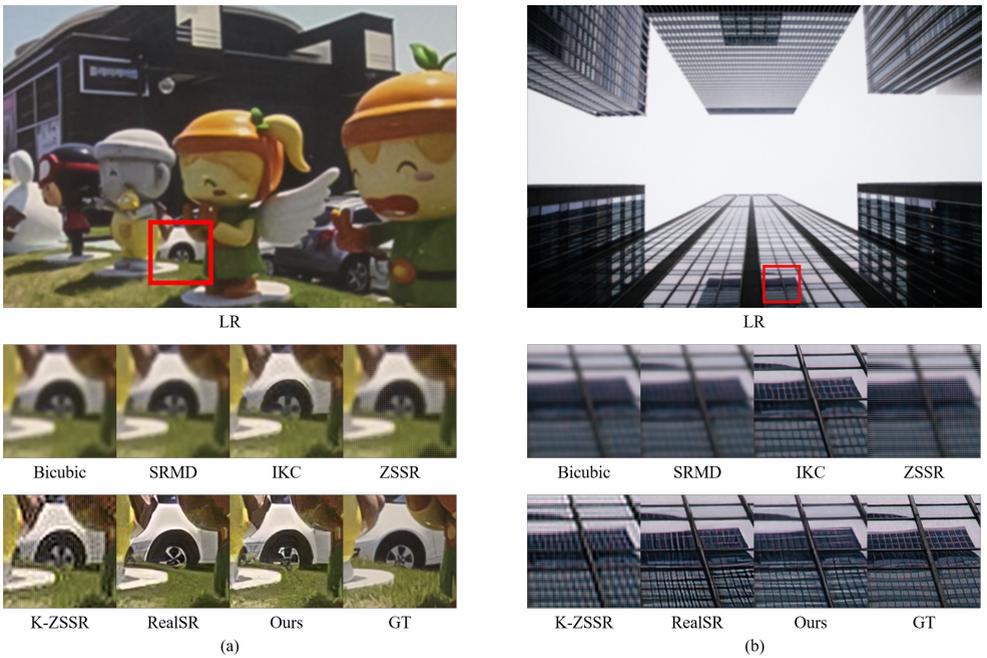


Figure 4: Qualitative results on other dataset different from the training dataset domain: (a) RealSR-V3 dataset, (b) DIV2K dataset

ZSSR produces the LR-HR pairs without considering the source domain, so the blurred outputs are generated. Also, K-ZSSR makes unwanted textures that look like artifacts. Real-SR shows good performance compared to other baselines, but over-sharpening occurs at edges such as letters or digits. On the other hand, DPP-SR produces relatively clean and sharp-edged results. In particular, DPP-SR creates brighter results that are similar to the original brightness and expresses a pattern such as a tree branch well. Consequently, it can be seen that DPP-SR shows higher performance than the state-of-the-art real-world SR methods.

To demonstrate generalization performance of our proposed method, we use the RealSR-V3 dataset, DIV2K dataset, and CelebA-HQ dataset with the networks trained with the DPED dataset. In Table 1, the quantitative results are summarized. DPP-SR shows higher performance than other methods in the no-reference evaluation metrics. On the other hand, since the degraded images were generated synthetically, IKC which is supervised learning based method shows higher performance than DPP-SR in all reference evaluation metrics. However, it can be seen that the results of IKC are more blurry than DPP-SR in Figure 4. It can be seen that distortion has occurred in the results of RealSR. In contrast, DPP-SR produces the high-quality images without distortion. Figure 5 shows the experimental results for CelebA-HQ. DPP-SR converts degraded LR face images into clear high-resolution face images. As a result, the experimental results demonstrate that our proposed method shows high generalization performance qualitatively and quantitatively.

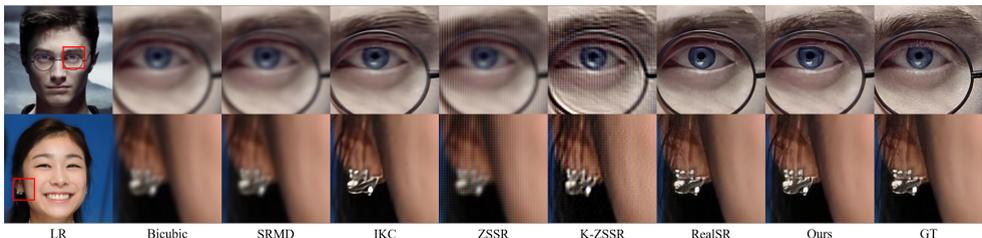


Figure 5: Qualitative results on the CelebA-HQ dataset

Type	Noise	Kernel	Front	Back	Reverse	Conv1	Conv5	Ours
NIQE↓	3.9824	3.9854	3.8445	4.1088	3.8427	3.8302	3.8436	3.7137
BRISQUE↓	22.1859	22.8842	21.9890	22.4550	22.2226	21.3545	20.6951	21.4566
PIQE↓	12.7949	14.0710	13.7051	12.6076	13.7598	13.1032	12.9753	13.6320
NRQM↑	6.5898	6.5737	6.5434	6.5698	6.6268	6.6155	6.6954	6.6546
PI↓	3.6963	3.7059	3.6506	3.7695	3.6080	3.6074	3.5741	3.5296

Table 2: Quantitative results of the ablation study.

4.3 Ablation Study

Several experiments are additionally conducted to analyze how deep degradation prior affects SR performance. In Table 2, the results of the ablation study are summarized. First, we compare when only the deep noise prior is used, when only the deep kernel prior is used, and when both are used. Table 2 shows better NIQE, BRISQUE, NRQM, and PI when using both than when using only one, but PIQE is highest when using only deep noise prior. Next, there is the experiment according to the position where the deep noise prior and the deep kernel prior are applied. In Table 2, *front* is when both priors are in front, *back* is when they are in the back, and *reverse* is when the two positions are swapped. Our proposed position achieves the best NIQE, BRISQUE, NRQM and PI, but PIQE is best at *back*. The last experiment is the performance evaluation according to the position where the deep noise prior is extracted from the deep noise prior generator. BRISQUE, PIQE, and NRQM are the best when extracted from *Conv5*. However, the result of *Conv3* shows the best NIQE, PI, and visually the best results. As a result, the deep noise prior extracted from *Conv3* has the highest performance both in terms of NIQE and PI, and qualitatively. The qualitative results about the ablation study are displayed in the supplementary material.

5 Conclusion

In this paper, a novel real-world SR method called DDP-SR was proposed. In DDP-SR, the deep degradation prior extracted from the deep prior generators is applied to the SR model. The SR model delivered high performance by reflecting the degradation prior. To evaluate the performance of DDP-SR, we compared with state-of-the-art methods. As a result, DDP-SR outperformed the state-of-the-art methods. Additionally, when testing with the datasets different from the domain of DPED used for the training, the DDP-SR outperformed the baseline qualitatively and quantitatively.

Acknowledgements

This research was supported by Deep Machine Lab (Q2109331).

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.
- [2] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *arXiv preprint arXiv:1909.06581*, 2019.
- [3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. 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*, pages 3086–3095, 2019.
- [4] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *arXiv preprint arXiv:2004.07728*, 2020.
- [5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.
- [6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- [7] 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.
- [8] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1712–1722, 2019.
- [9] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353, 2010.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [12] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

- [13] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3277–3285, 2017.
- [14] Andrey Ignatov, Radu Timofte, Thang Van Vu, Tung Minh Luu, Trung X Pham, Cao Van Nguyen, Yongwoo Kim, Jae-Seok Choi, Munchurl Kim, Jie Huang, et al. Pirm challenge on perceptual image enhancement on smartphones: Report. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.
- [15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [16] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 466–467, 2020.
- [17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [18] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [19] Gwantae Kim, Jaihyun Park, Kanghyu Lee, Junyeop Lee, Jeongki Min, Bokyeung Lee, David K Han, and Hanseok Ko. Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 456–457, 2020.
- [20] 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.
- [21] Yoonsik Kim, Jae Woong Soh, and Nam Ik Cho. Adaptively tuning a convolutional neural network by gate process for image denoising. *IEEE Access*, 7:63447–63456, 2019.
- [22] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3482–3492, 2020.
- [23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [24] 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.
- [25] Junyeop Lee, Jaihyun Park, Kanghyu Lee, Jeongki Min, Gwantaek Kim, Bokyeung Lee, Bonhwa Ku, David K Han, and Hanseok Ko. Fbrnn: feedback recurrent neural network for extreme image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 488–489, 2020.
- [26] 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.
- [27] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3408–3416. IEEE, 2019.
- [28] 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.
- [29] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013.
- [30] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 945–952, 2013.
- [31] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.
- [32] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.
- [33] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.
- [34] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.
- [35] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019.
- [36] Haoyu Ren, Amin Kheradmand, Mostafa El-Khamy, Shuangquan Wang, Dongwoon Bai, and Jungwon Lee. Real-world super-resolution using generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 436–437, 2020.
- [37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

- [38] 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.
- [39] Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. Ingan: Capturing and retargeting the “dna” of a natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4492–4501, 2019.
- [40] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [41] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017.
- [42] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6924–6932, 2017.
- [43] N Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 Twenty First National Conference on Communications (NCC)*, pages 1–6. IEEE, 2015.
- [44] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018.
- [45] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.
- [46] Xiangyu Xu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming-Hsuan Yang. Learning to super-resolve blurry face and text images. In *Proceedings of the IEEE international conference on computer vision*, pages 251–260, 2017.
- [47] 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.
- [48] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.
- [49] 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.

- [50] Kai Zhang, Shuhang Gu, and Radu Timofte. Ntire 2020 challenge on perceptual extreme super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 492–493, 2020.
- [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 (ECCV)*, pages 286–301, 2018.
- [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.
- [53] Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2433–2443, 2019.
- [54] 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.