Kernel Agnostic Real-world Image Super-resolution

Hu Wang  Congbo Ma  Chunhua Shen
The University of Adelaide

ABSTRACT
Recently, deep neural network models have achieved impressive results in various research fields. Come with it, an increasing number of attentions have been attracted by deep super-resolution (SR) approaches. Many existing methods attempt to restore high-resolution images from directly down sampled low-resolution images or with the assumption of Gaussian degradation kernels with additive noises for their simplicities. However, in real-world scenarios, highly complex kernels and non-additive noises may be involved, even though the distorted images are visually similar to the clear ones. Existing SR models are facing difficulties to deal with real-world images under such circumstances. In this paper, we introduce a new kernel agnostic SR framework to deal with real-world image SR problem. The framework can be hanged seamlessly to multiple mainstream models. In the proposed framework, the degradation kernels and noises are adaptively modeled rather than explicitly specified. Moreover, we also propose an iterative super-resolution process and frequency-attended objective from orthogonal perspectives to further boost the performance. The experiments validate the effectiveness of the proposed framework on multiple real-world datasets.

CCS CONCEPTS
• Computing methodologies → Computer Vision; • Image processing; • Reconstruction;

KEYWORDS
Image Super-resolution, Computer Vision, Kernel Agnostic

ACM Reference Format:
Hu Wang Congbo Ma Chunhua Shen. 2018. Kernel Agnostic Real-world Image Super-resolution. In Preprint ’2021: Kernel Agnostic Real-world Image Super-resolution. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3122445.3122456

1 INTRODUCTION
In the image super-resolution task, low-resolution images will be inputted into a SR model and the corresponding high-resolution images are expected to be restored. SR techniques can be applied to a wide range of applications, such as high-definition (HD) television display, zooming processes on phones, security and surveillance of cameras, etc. With the emerging of applying deep neural networks in the computer vision field, super-resolution research has experienced a big surge. In 2014, Dong et al. [6] applied a three-layer convolutional neural network to the SR task for the first time. Since then, super-resolution research has entered a new era. Many powerful models have been proposed [11, 13, 14, 23, 29].

However, how to perform SR models on a real-world image dataset remains a big challenge to be solved. Bicubic interpolation or Gaussian blurry kernel combined with additive noises are still mostly adopted to acquire low-resolution images for SR dataset building because of their simplicities. Nevertheless, in real-world image super-resolution scenarios, highly non-additive noises are generated by different CMOS image sensors in cameras. The distribution of low-level pixels within testing images are significantly different from its training samples. Thus, it inevitably causes the poor generalization of the SR model trained on Bicubic downsampled or Gaussian degenerated LR images.

In 2000, Hui et al.[20] analysed the temporal noises caused by CMOS image sensors. Similarly, in 2006, Liu et al.[15, 16] conducted research attempting to figure out the relations between the changes of noise level with respect to brightness. They also committed to infer the noise level function from one image with Bayesian MAP inference. However, the noises tried to model turns out to be extremely non-additive, as well as the noises vary across different CCD digital cameras. Some existing super-resolution models [2, 19] attempt to alleviate this problem to estimate the kernel from a few-shot learning perspective. Another typical solution [10] is to maintain a finite set of estimated degradation kernels and additive noises to accommodate the SR process. However, they neither can be applied directly in the inference phase, nor the estimated kernels still retain linearity that cannot represent a complex degradation process in the real world.

In this paper, we introduce a generic framework for real-world image super-resolution without prior knowledge or explicitly specify the degradation kernels, named Kernel Agnostic Super-resolution (KASR). KASR is able to be hanged on multiple mainstream seamlessly and it is end-to-end trainable. For intuitive purpose, the images generated by KASR and the differences with the original ones are shown in Fig. 1 and 3. Besides, we introduce a complement optimization process termed Iterative Supervision to refine the SR images gradually. What is more, a Frequency-attended Loss function is adopted to further excavate the high-frequency areas within SR images. The experimental results demonstrate that, when compared to multiple strong baselines on prevalent real-world datasets, the models equipped with KASR is able to generate high-quality SR images in real-world images SR scenarios.

Our contributions can be summarized as follow.

- We propose a Kernel Agnostic Super-resolution (KASR) framework from an adversarial training perspective to deal with real-world image super-resolution task. KASR is able to model the degradation process implicitly in an end-to-end training manner.
Figure 1: The figure shows (a) the original low-resolution image, (b) the LR image down-sampled by the KASR model and (c) the corresponding differences of these two images. The differences are highlighted in pink. We can perceive that the differences are mainly colour changes combined with some texture distortions.

• To facilitate the KASR training process, we introduce a complementary Iterative Supervision optimization process to refine the SR image gradually. Moreover, towards the nature of high-frequency preference of SR task, a Frequency-attended Loss function is adopted to assign more attention on high-frequency regions of the images.

• The extensive experiments demonstrate that the models equipped with KASR is able to receive superior performances over competitive methods across multiple real-world datasets.

2 RELATED WORK

2.1 Neural Network based Image Super-resolution

With the spring up of deep neural networks, super-resolution research has gain momentum in recent years. After the first deep super-resolution model proposed by Dong et al.[5, 6], a variety of convolutional neural models have been studied. After the proposal of SRCNN, a revised version called FSRCNN introduced by Dong et al.[7]. In this model, a low-resolution image can be inputted directly. Kim et al.[11, 12] brought residual learning into super-resolution task for deeper network training. An encoder-decoder framework is introduced by the RED model [17]. Later on, ESPCN [18] is proposed to optimize the SR model by learning sub-pixel convolutional filters. Ledig et al.[13] introduced a Generative Adversarial Networks (GANs) SR model named SRGAN, along with which the generator is named SRResNet. Tong et al.[22] presented SRDenseNet model by first adopting dense-blocks into the SR task. The model has the advantages of diminishing the problem of gradient disappearance, strengthening the feature propagation, and reducing the number of parameters at the same time. Lim et al.[14] proposed a model named EDSR. Along with the EDSR model, the MDSR model is proposed as well, who is the winner of the NTIRE2017 challenge [1]. EDSR trades cancelling the batch-normalization layer with stacking more convolutional layers and broader residual blocks to obtain more powerful representational ability. Residual channel attention is introduced by Zhang et al.[29] to overcome the gradient vanishing problem in very deep SR networks.

2.2 Real-world Image Super-resolution

Low-resolution images can be viewed as degraded high-resolution images after blurring, down-sampling and noise interference. An important premise of deep learning is that the distribution of the test image should have consistent distribution with the training data. When the degradation process in the training phase is inconsistent with the inference phase, the result will often drop significantly. Therefore, how to simulate extremely complex degradation process as real world is challenging.

Thinking from a few-shot learning perspective, [19] proposed a kernel estimation approach by further down-sampling the given low-resolution image and learn a super-resolution function from these low-high resolution image pairs. However, it cannot be applied directly in the inference phase since training is required on each test image, which is time-consuming. A similar idea has been adopted by KernelGAN [2]. Nevertheless, the assumption of Gaussian blur kernels still exists to train the model.

Gaussian kernel is the most widely adopted blur kernel [8, 25]. But it differs from the real-world images significantly. Therefore, some research works attempted to simulate the degradation process in a more complicated manner by considering a variety of degradation kernels and noise levels. Zhang et al.[26] examine the performance of SR models under a set of Gaussian blur kernels. Another typical idea is to estimate a finite set of down-sample kernels/noises and inject them into training images as [10]. However, a finite set of kernels and noises is still a very strong assumption.

2.3 Adversarial Training

As proven by the existing work [9], deep neural models are very sensitive towards small perturbations of the inputs. Tiny pixel-wise distortions will lead the deep neural networks to entirely different results. Adversarial training is an effective method for the model against attacks by dynamically and adaptively adding perturbations into the training samples.

In real-world SR scenario, models usually suffer from poor generalization performance due to unknown degradation kernels and noises in the inference phase. Highly non-additive noise would
Figure 2: The overall framework of the proposed KASR framework. The framework consists of three parts: (a) Kernel Agnostic Noise Simulation to adaptively simulate the image degradation process. More specifically, a high-resolution image $\mathbf{I}_{HR}$ is fed into the kernel agnostic noise simulation network to generate a low-resolution image $\mathbf{I}'_{LR}$ containing injected non-additive noises. We constrain the $\mathbf{I}'_{LR}$ to be visually similar with the original low-resolution image $\mathbf{I}_{LR}$, but the injected perturbations will interfere with the downstream super-resolution image restoration process. (b) Frequency-attended Objective forces the model to pay more attention to high-frequency regions within images due to the high-frequency restoration nature of the super-resolution task. Specifically speaking, we first extract frequency maps from the high-resolution and super-resolution images respectively and then normalized them into 0 to 1 as the frequency attention masks $\mathbf{F}_{SR}$ and $\mathbf{F}_{HR}$. Later on, Hadamard products between frequency attention masks and high-resolution/super-resolution images are performed to get the attended images $\mathbf{I}'_{SR}$ and $\mathbf{I}'_{HR}$ for loss computation. (c) By stacking multiple KANS blocks, Iterative Supervision can leverage the supervision signals in a better manner for the refinement of SR images.

be injected by different CMOS image sensors. Adversarial training can be adopted here against unknown kernels and extremely non-additive noises. As far as we know, we are the first to introduce adversarial training into SR task for implicit simulation of the degradation process.

3 THE PROPOSED METHOD

In this section, we introduce the proposed Kernel Agnostic Real-world Image Super-resolution framework as shown in Fig. 2.

3.1 Preliminary and Motivation

Put the super-resolution process into a simple mathematical model, the obtaining of a low-resolution image can be viewed as a degradation process by combining a degradation kernel with noise. The degradation process of the super-resolution problem can be formulated as follow:

$$\mathbf{I}_{LR} = (\mathbf{I}_{HR} \ast k) \downarrow_s + \delta,$$

where LR image $\mathbf{I}_{LR}$ are formed from HR image $\mathbf{I}_{HR}$ by down-sampling and adding perturbations. The $\downarrow_s$ denotes down-sampling operation with scale $s$ and $\ast$ denotes the convolutional operation. $k$ and $\delta$ are the blur kernel and the noise, respectively.

However, in real-world scenarios, super-resolution models are usually deployed on phones or HD TVs. Arbitrary non-additive distortions could exist, may include motion blur, defocused, sensor noises or deformations from image compression. Poisson noises or Gaussian noises can be simply added on RAW images to model the process. Nevertheless, when it comes to modelling sRGB noises, it is much more difficult, since the real noises in sRGB images are processed by an image signal processing (ISP) pipeline, where a series of non-linear processing operations are performed. The noises in the original sensor space are signal-dependent, and they are spatially and colour-related. Therefore, additional Poisson noises or Gaussian noises are able to model the process well. But after ISP, the noises cannot simply be treated as additive noises.

Many existing SR works [6, 12–14, 18, 29] assume the low-resolution images are clean images fetched from naive bicubic down-sampling.
Following this strong assumption, the problem of inconsistent blur kernels and noises between training and application exists according to Eqn. 1. In order to conquer this issue, more generally, some works [8, 10, 25, 26] adopt Gaussian kernel, which is the most widely used blur kernel, combined with down-sampling to assist the super-resolution training process. But in real-world scenarios, degradation kernels and noises are more complicated then this assumption. The aforementioned degradation thus problem persists.

Motivated by the idea of adversarial learning, we propose to generate the degraded images dynamically with the power of deep neural networks for the sake of super-resolution models.

### 3.2 Overall Framework

The proposed KASR framework can be mainly divided into three parts: Kernel Agnostic Noise Simulation to dynamically simulate the image degradation process for robust SR model training. The Frequency-attended Objective constrains the model to focus on high-frequency regions within images. Iterative Supervision leverages the provided supervision signals in a better manner by repeatedly stacking the KANS block for SR results refinement. These parts work complementarily and are able to hang seamlessly with multiple mainstream SR models.

As shown in Fig. 2, a high-resolution image is first inputted into the kernel agnostic noise simulation network to generate a low-resolution image with non-additive simulated noises. The generated LR image is forced to be visually similar with the original low-resolution image, but the distortion process will interfere super-resolution image restoration process. After the super-resolution image is generated, the Frequency-attended Objective constrains the model to focus on high-frequency regions within images due to the high-frequency preference of the SR task. We first extract frequency maps from the HR and SR images and then normalized them into 0 to 1 as the frequency attention masks. Later on, Hadamard products between frequency attention masks and HR/SR images are performed to get the converted images for loss computing. Moreover, an iterative aforementioned process is adopted by Iterative Supervision for further results refinement as shown in Fig. 2.

### 3.3 Kernel Agnostic Noise Simulation

How to adaptively optimize the SR model to gain promising performance by giving arbitrary unknown degradation kernels and noises is a challenging problem. Inspired by the adversarial training process, we propose a Kernel Agnostic Noise Simulation for real-world SR problem.

Adversarial training techniques were proposed to defend against adversarial attacks initially. While in our context, adversarial training is adopted for adaptively and actively generate non-additive simulated noises and injected them into the low-resolution image by maximizing the super-resolution image restoration loss. But the generated low-resolution image should be visually similar to the original one. Therefore, the problem can be naturally transferred into a minimization and maximization problem to search the suitable subtle noise $\delta$ in the $\epsilon$-ball space ($\|\delta\|_p \leq \epsilon$):

$$
\begin{align*}
\min_{k,\theta} \max_{\|\delta\|_p \leq \epsilon} & \|\eta((I_{HR} \otimes k) \downarrow_s + \delta; \theta) - I_{HR}\|_p \\
\min_{k,\theta} & \|((I_{HR} \otimes k) \downarrow_s + \delta) - I_{LR}\|_p
\end{align*}
$$

where $k$ and $s$ are the degradation kernel and scale respectively, $(I_{HR} \otimes k) \downarrow_s$ is the degradation process and $\eta(\cdot)$ is the super-resolution network parametrized by $\theta$, $\| \cdot \|_p$ denotes the p-norm operation.

In super-resolution, the optimization of Eqn. 2 is not trivial and the restriction can be relaxed in our task. Thus, the aforementioned problem can be resolved by simply generating subtle noises and down-sampling images with a neural network, such that the noise can be generated dynamically with little computation increment. The training framework is thus successfully transferred into an adversarial training structure. More formally, the equation after relaxation is:

$$
\begin{align*}
\min_{\theta} & \max_{\phi} \|\eta(\phi(I_{HR}; \theta_{\phi}); \theta) - I_{HR}\|_p \\
\min_{\theta} & \|\phi(I_{HR}; \theta) - I_{LR}\|_p
\end{align*}
$$

where the $\phi(\cdot)$ is the degradation neural function with parameter $\theta_{\phi}$.

By doing so, tremendously non-additive noise can be injected into the generated low-resolution image. More specifically to achieve this, as mentioned above, a data-driven approach is adopted to solve the unknown kernel and highly Non-Additive noise (NA-$\delta$) injection problem by minimizing the reconstruction error over the generated and real low-resolution image pairs and maximizing the super-resolution image restoration error.

$$
\begin{align*}
k_{NA-\delta}^{xs} = \arg \min_k (\alpha \|\phi(I_{HR}; \theta) - I_{LR}\|_p - \gamma \|I_{SR} - I_{HR}\|_p).
\end{align*}
$$

where $\phi(\cdot)$ is to conduct the degradation process $(I_{HR} \otimes k) \downarrow_s$ with a neural network $\phi$ for degradation purposes parametrized by $\theta$. $\gamma$ is the trade-off factor to balance between the objectives.

### 3.4 Training Process and Objectives

**Iterative Supervision.** In standard SR models, high-resolution images are directly generated from low-resolution images. Different from that, we propose to perform iterative optimization by repeating the degradation and the super-resolution processes. The gradients are able to flow through the chain by providing extra supervisions to the iteratively generated images. Therefore, supervision signals can be excavated in a better manner to constrain the blurry images generated by the latter blocks.

**Frequency-attended Objective.** In the nature of super-resolution tasks, high-frequency areas within an image ought to be retained for better visualization quality. Naive pixel-wise losses (e.g. MSE loss) assign the same weights to each pixel and average the errors that come from these pixels. It will result in blurry SR images. Thus, pixels should be assigned with different weights according to their importance in the objectives. Due to the high-frequency preference of super-resolution task, pixels located in high-frequency regions require more attention; vice versa. Based on this theory, we adopt
the Frequency-attended Objective to attend to the “important” regions within an image. Formally, the Frequency-attended Objective can be represented as:

\[ L_F = \| \nabla \psi(f(I_{HR})) \odot I_{HR} - \psi(f(I_{SR})) \odot I_{SR} \|_p. \]  

where \( f(\cdot) \) is the high-frequency filter and \( \psi(\cdot) \) denotes the normalization operation. \( \odot \) represents the element-wise hadamard product. \( p \) norm is performed in the equation as well.

**Objective functions.** Besides the aforementioned objectives of Kernel Agnostic Noise Simulation and Frequency-attended Objective, the pixel-wise loss function has been adopted as well. The total objective function can be formally presented as below. SR image loss:

\[ L_I = \| I_{SR} - I_{HR} \|_p. \]  

Therefore, the total objective function for super-resolution network is:

\[ L_{SR} = L_I + wL_F, \]  

where \( w \) is the trade-off factor between this two objectives.

In the adversarial training loss perspective, a generative adversarial objective is adopted to judge whether the generated LR images are real images by a discriminator to further ensure the quality of generated images. So the objective for KANS can be presented:

\[ L_{KANS} = \alpha \| \psi(I_{HR} ; 0) - I_{LR} \|_p + \beta L_G - \gamma \| I_{SR} - I_{HR} \|_p. \]  

where \( L_G \) is a generative adversarial objective. Trade-off factors \( \alpha, \beta \) and \( \gamma \) balance different types of objectives functions. In general, we empirically discover that finding suitable trade-off factors is crucial for the model training.

4 EXPERIMENTS

4.1 Experimental Setup

In the experiments, we examine the effectiveness of the proposed Kernel Agnostic Real-world Image Super-resolution framework combined with multiple mainstream super-resolution models across three prevalent real-world datasets. Later on, comparisons between the proposed model with existing SR models are performed. In the ablation study, the effectiveness of each component is validated.

**Evaluation Metrics.** For the evaluation of the proposed method and other methods, we adopt Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [24] as our evaluation metrics. Besides that, since the PSNR and SSIM metrics only take pixel-wise distance into consideration, we further include widely adopted LPIPS [27] as another evaluation metric for perceptual evaluation. LPIPS values particularly resonate with the human observations by depicting the perceptual similarities between SR images and real HR ones.

**Datasets.** We conduct the model training on multiple real-world image super-resolution datasets and testing on the low-resolution images of the corresponding dataset. The datasets are RealSR [3] (including both \( \times 2 \) and \( \times 4 \) up-scaling), City100 [4] and SR-RAW [28]. There are 506 and 500 low-resolution/high-resolution image pairs taken by Canon and Nikon cameras for \( \times 2 \) and \( \times 4 \) up-scaling of the RealSR dataset. Within the dataset, we adopt 406 images for \( \times 2 \) model training and 400 images for \( \times 4 \) model training, respectively. City100 dataset contains 100 real-world low-resolution/high-resolution image pairs taken by a Nikon camera with \( \times 3 \) up-scaling. We adopt the first 95 images for model training and the rest 5 images for model evaluation. W.r.t. the SR-RAW dataset, we adopt 449 sRGB images for model training. In the dataset, 400 images are adopted for model training and 49 images for model evaluation.

**Implementation details.** During the training of models, 32 batch-size is adopted and Adam optimizer is chosen for model optimization. Following the original training scheme of the EDSR model, we randomly flip images vertically/horizontally and rotate them by 90 degrees for data augmentation. The initial learning rate is set to \( 1 \times 10^{-4} \). The models are trained for 300 epochs with a multi-step learning rate reduction strategy. The trade-off hyper-parameter \( w \) is set to 0.5; \( \alpha \) is set to 1.0; \( \beta \) is set to 0.5; \( \gamma \) is set to 1.0, respectively.

For the implementation of the Kernel Agnostic Noise Simulation Network, three convolutional layers combined with the LeakyReLU activation function and Max-pooling operation are stacked. \( 2 \times 10^{-1} \) is selected as the negative slope of LeakyReLU activation. 32 neural units are set for each hidden layers of the Kernel Agnostic Noise Simulation Network. In our experiments, \( L_1 \) norm is adopted for norm-required operations and two KASR blocks are placed for the iterative supervision training. We consider Sobel filter as the high-frequency filter \( f(\cdot) \) and Max-min normalization as \( \psi(\cdot) \) due to the simplicity. The models are implemented with PyTorch on one NVIDIA GTX 1080 Ti graphic card. The experimental settings are kept the same across all models for fair comparisons.

4.2 Overall Performance

The overall performance of the proposed KASR model is generally shows in Tab. 1, Tab. 2 and Tab. 3.

**Experiments on RealSR [3].** We equip the proposed KASR framework on three prevalent super-resolution models, including EDSR [14], SRResNet [13] and SRGAN [13]. From the Tab. 1, we can discover the models equipped with KASR are able to gain better results steadily when compared with the original models. Under \( \times 2 \) up-scaling, SRResNet-KASR increase 0.322 for PSNR compared with the original SRResNet model. SRGAN-KASR gain 0.301 more performance for PSNR. For \( \times 4 \) up-scaling, EDSR-KASR increase the PSNR results of the EDSR model from 27.494 to 27.850 and SSIM from 0.813 to 0.824. Similarly, 0.294 increments for the SRResNet-KASR model and 0.279 for SRGAN backboned model on PSNR, respectively. The performance improvements are not only on pixel-level evaluation metrics (PSNR and SSIM), the models equipped with KASR are able to receive better results on perceptual metric LPIPS as well.

**Experiments on City100 [4].** Similar with the experiments on RealSR, we conduct experiments to equip the KASR framework on EDSR [14], SRResNet [13] and SRGAN [13] for City100 dataset. The quantitative comparisons are shown in Tab. 2. On the City100 dataset, 0.743 increment gained by the EDSR-KASR model. We find from the table that the perceptual performance improvements of the model equipped with the KASR framework are more obvious compared with that on RealSR. EDSR-KASR improves the LPIPS index from 0.138 to 0.112; both SRResNet-KASR and SRGAN-KASR improves the LPIPS from 0.117 to 0.109. This phenomenon may be caused by the different distribution of data in different datasets.
Figure 3: The figure illustrates the differences between original images $I_{LR}$ and KASR downgraded images $I'_{LR}$ through (a)-(e). $I_{LR}$ and $I'_{LR}$ are visually similar, but small perturbations are contained in $I'_{LR}$ to misled the deep super-resolution models. (a) Compared with $I_{LR}$, $I'_{LR}$ is modified in both colour and texture. (b) Texture modifications are included in $I'_{LR}$. (c) Colour changes are performed on $I'_{LR}$. (d) Blurs are added to $I'_{LR}$ compared with $I_{LR}$. (e) Some fine details are added to $I'_{LR}$.

**Table 1:** The comparison between original models and the ones equipped with the proposed Kernel Agnostic Real-world Super-resolution framework on RealSR dataset [3] with $\times2$ and $\times4$ up-scaling. The $\uparrow$ indicates the higher value the better; while the $\downarrow$ represents the lower value the better.

| Methods               | Scale | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |
|-----------------------|-------|-----------------|-----------------|---------------------|
| EDSR [14]             | $\times2$ | 32.378          | 0.922           | 0.081               |
| EDSR-KASR (Ours)      | $\times2$ | 32.454          | 0.922           | 0.080               |
| SRResNet [13]         | $\times2$ | 31.630          | 0.915           | 0.086               |
| SRResNet-KASR (Ours)  | $\times2$ | 31.952          | 0.915           | 0.083               |
| SRGAN [13]            | $\times2$ | 31.611          | 0.915           | 0.086               |
| SRGAN-KASR (Ours)     | $\times2$ | 31.912          | 0.915           | 0.081               |
| EDSR [14]             | $\times4$ | 27.494          | 0.813           | 0.157               |
| EDSR-KASR (Ours)      | $\times4$ | 27.850          | 0.824           | 0.148               |
| SRResNet [13]         | $\times4$ | 27.366          | 0.819           | 0.153               |
| SRResNet-KASR (Ours)  | $\times4$ | 27.660          | 0.820           | 0.153               |
| SRGAN [13]            | $\times4$ | 27.355          | 0.819           | 0.156               |
| SRGAN-KASR (Ours)     | $\times4$ | 27.634          | 0.818           | 0.151               |

**Experiments on SR-RAW [28]** For the SR-RAW dataset, we equip KASR to EDSR and RCAN and examine the model performance. KASR improves the EDSR model by increasing 0.63 of the PSNR and boosts the RCAN model performance by gaining 0.376 of the PSNR result.

| Methods               | Scale | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |
|-----------------------|-------|-----------------|-----------------|---------------------|
| EDSR [14]             | $\times3$ | 29.971          | 0.765           | 0.138               |
| EDSR-KASR (Ours)      | $\times3$ | 30.714          | 0.824           | 0.112               |
| SRResNet [13]         | $\times3$ | 30.305          | 0.817           | 0.117               |
| SRResNet-KASR (Ours)  | $\times3$ | 30.309          | 0.816           | 0.109               |
| SRGAN [13]            | $\times3$ | 30.181          | 0.815           | 0.117               |
| SRGAN-KASR (Ours)     | $\times3$ | 30.253          | 0.815           | 0.109               |

**Table 2:** The comparison between original models and the ones equipped with the proposed Kernel Agnostic Real-world Super-resolution framework on City100 dataset [4] with $\times3$ up-scaling.

| Methods               | Scale | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |
|-----------------------|-------|-----------------|-----------------|---------------------|
| EDSR [14]             | $\times2$ | 20.482          | 0.793           | 0.179               |
| EDSR-KASR             | $\times2$ | 21.112          | 0.795           | 0.172               |
| RCAN [29]             | $\times2$ | 22.683          | 0.804           | 0.184               |
| RCAN-KASR             | $\times2$ | 23.059          | 0.806           | 0.178               |

SR models [14], we also present our model performance with self-ensemble [21] as Ours+ in the table. As shown in Tab. 4 and Tab. 5, the experiments are conducted on RealSR dataset and City100 dataset.

**4.3 Comparison with Existing SR Models**

We compare the proposed KASR model with existing super-resolution models. In this section, the EDSR model equipped with KASR is selected as our model. Following the mainstream evaluation of
the most popular baselines. The Noise-injection model is a state-of-the-art real-world super-resolution model. It estimates a finite set of degradation kernels and noises from pre-collected real images by adopting KernelGAN [2] and ZSSR [19]. In order to deal with Real-world SR noises, the model performs degradation operation on the degradation pool during training. However, extra training time for kernel collection is required.

On the RealSR dataset, our model can receive better results under multiple evaluations. Compared with ESRGAN and RCAN, our models outperform the baseline models by a large margin across three evaluations, especially on LPIPS evaluations. The Noise-injection model is able to receive much better perceptual results than ESRGAN and RCAN models, but still not as good as the proposed KASR model. Compared with the Noise-injection model, our model improves the LPIPS metric from 0.215 to 0.150.

Interestingly, from the table, we find out that the models with self-ensemble can be boosted the performance under both PSNR and SSIM, but not on LPIPS. This may be the self-ensemble operation performs pixel-level ensemble that can not effectively increase the perceptual quality of the images.

| Methods          | Scale | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
|------------------|-------|--------|--------|---------|
| ESRGAN [23]      | ×4    | 27.569 | 0.774  | 0.415   |
| RCAN [29]        | ×4    | 27.647 | 0.780  | 0.442   |
| Noise-injection [10] | ×4    | 25.768 | 0.772  | 0.215   |
| Ours             | ×4    | 27.850 | 0.824  | 0.148   |
| Ours+            | ×4    | 27.991 | 0.827  | 0.150   |

Table 5: The comparison of our proposed Kernel Agnostic Real-world Image Super-resolution model and existing super-resolution models on City100 dataset [4] with ×4 up-scaling. The best results of a column are bolded.

4.4 Ablation Study

Effect of different components. This section examines the contribution of different components of the proposed KASR framework, including the Kernel Agnostic Noise Simulation network (KANS), Frequency-attended Objective (FAO) and Iterative Supervision (IS). The quantitative results are shown in the Tab. 6.

| Models | KANS | FAO | IS | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
|--------|------|-----|----|--------|--------|---------|
| 1      |      |     |    | 27.494 | 0.813  | 0.157   |
| 2      | ✓    |     |    | 27.718 | 0.823  | 0.151   |
| 3      | ✓    | ✓   |    | 27.762 | 0.823  | 0.151   |
| 4      | ✓    | ✓   | ✓  | 27.850 | 0.824  | 0.148   |

From the table, we empirically find out the effectiveness of proposed each component. With KANS, model #2 raises the PSNR from 27.494 to 27.718; SSIM from 0.813 to 0.823; LPIPS from 0.157 to 0.151, respectively. By equipping with FAO and IS, the performance of the model is able to be further boosted to our best #4 model with 27.850 on PSNR, 0.824 on SSIM and 0.148 on LPIPS.

Sensitivity test of γ. The trade-off factor γ controls the intensity of adversarial training and the choice of γ play an important role in the model performance. Tab. 7 generally shows the sensitivity of gradually adding the trade-off factor γ. Interesting to note that the results show the KASR model performs relatively stably w.r.t. the choice of γ. This may be caused by adopting the KASR framework, the SR models are able to learn distinctive features for super-resolution even with small γ values. However, there is an increasing trend with the increment of γ. In general, 1.0 are recommended for the γ setting of KASR to achieve effective super-resolution performance.

Similar results are shown in Tab. 5 on City100 dataset. The comparing methods include RCAN [29], CamSR-SRGAN [4] and CamSR-VDSR [4]. Our KASR model improves the PSNR value from 30.260 to 30.714 compared to the second place model (CamSR-VDSR). When compared with CamSR-based models, the KASR model increases the LPIPS evaluation by a large margin.

4.5 Visualization

Visualization of generated LR images with KANS. As shown through sub-figures (a) to (e) in Fig. 3, more examples of LR images generated by KANS network. The distortions including texture changes, colour modification and blurriness.

Visualization of SR images. From the visualization presented in Fig. 4, we identify that the KASR model (RCAN-KASR) can generate
Figure 4: Visual comparison between HR image, Bicubic interpolation, RCAN and the proposed KASR model.

Figure 5: Visual comparison between HR image, Bicubic interpolation, Noise-injection and the proposed KASR model.

much clearer images than the Bicubic method and RCAN model on real-world images.

Similar results are shown in Fig. 5, our KASR model (EDSR-KASR) can generate clear SR images than other competitors. Worth noting that the Noise-injection [10] model can generate clear SR images in some cases, but it will create small artifacts as the character and the man’s beard shown in figures. This may be because its estimated kernels and noises cannot cover all the real scenes.

5 CONCLUSION

In this paper, we propose a Kernel Agnostic Real-world Image Super-resolution to assist the SR model to simulate the complex degradation process implicitly. The framework is able to be combined with multiple mainstream SR models seamlessly. The proposed KASR model receives promising results on real-world image SR scenarios across multiple evaluation metrics. Furthermore, an extensive ablation study is performed to examine the effectiveness of each component. From the visualization of SR images, our model can generate clear SR images compared with its opponents.
REFERENCES

[1] Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

[2] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. 2019. Blind super-resolution kernel estimation using an internal-gan. arXiv preprint arXiv:1909.08581 (2019).

[3] Jiajun Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. 2019. 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. 3086–3095.

[4] Chang Chen, Zhiwei Xiong, Ximmei Tian, Zheng-Jun Zha, and Feng Wu. 2019. Camera lens super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1652–1660.

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaou Tang. 2014. Learning a deep convolutional network for image super-resolution. In European conference on computer vision. Springer, 184–199.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaou Tang. 2015. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38, 2 (2015), 295–307.

[7] Chao Dong, Chen Change Loy, and Xiaou Tang. 2016. Accelerating the super-resolution convolutional neural network. In European conference on computer vision. Springer, 391–407.

[8] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. 2012. Nonlocally centralized sparse representation for image restoration. IEEE transactions on Image Processing 22, 4 (2012), 1620–1630.

[9] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014).

[10] Xiaohong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Fuyue Huang. 2020. Real-world super-resolution via kernel estimation and noise injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 466–467.

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1644–1654.

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1637–1645.

[13] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition. 4681–4690.

[14] BeeLim, SangHyunSon, HeewonKim, SeungjunNah, and KyoungMuLee. 2017. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 136–144.

[15] Ce Liu, William T Freeman, Richard Szeliski, and Sing Bing Kang. 2006. Noise estimation from a single image. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), Vol. 1. IEEE, 901–908.

[16] Ce Liu, Richard Szeliski, Sing Bing Kang, C Lawrence Zitnick, and William T Freeman. 2007. Automatic estimation and removal of noise from a single image. IEEE transactions on pattern analysis and machine intelligence 30, 2 (2007), 299–314.

[17] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. 2016. Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint arXiv:1606.08921 (2016).

[18] Wenhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. 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. 1874–1883.

[19] Assaf Shocher, Nadav Cohen, and Michal Irani. 2018. “zero-shot” super-resolution using deep internal learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 3118–3126.

[20] Hui Tian, Boyd Fowler, and Abbas E Gamal. 2001. Analysis of temporal noise in CMOS photodiode active pixel sensor. IEEE Journal of Solid-State Circuits 36, 1 (2001), 92–101.

[21] Radu Timofte, Rasmus Rothe, and Luc Van Gool. 2016. Seven ways to improve example-based single image super resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1865–1873.

[22] Tong Tong, Gen Li, Xiejie Liu, and Qiongquan Gao. 2017. Image super-resolution using dense skip connections. In Proceedings of the IEEE international conference on computer vision. 4799–4807.

[23] Xintao Wang, Ke Yu, Shixiang Wu, Jinjia Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Ergan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) 0–0.

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

[25] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. 2014. Single-image super-resolution: A benchmark. In European conference on computer vision. Springer, 372–386.

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

[27] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition. 586–595.

[28] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. 2019. Zoom to learn, learn to zoom. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 3762–3770.

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

Preprint '2021, April 18, 2021,