Efficient and Degradation-Adaptive Network for Real-World Image Super-Resolution

Jie Liang[1,2], **Hui Zeng**[2] and **Lei Zhang**[1,2,⋆]

[1]The HongKong Polytechnic University, [2]OPPO Research
{*liang27jie, cshzeng*}*@gmail.com*; *cslzhang@comp.polyu.edu.hk*

Abstract. Efficient and effective real-world image super-resolution (Real-ISR) is a challenging task due to the unknown complex degradation of real-world images and the limited computation resources in practical applications. Recent research on Real-ISR has achieved significant progress by modeling the image degradation space; however, these methods largely rely on heavy backbone networks and they are inflexible to handle images of different degradation levels. In this paper, we propose an efficient and effective degradation-adaptive super-resolution (DASR) network, whose parameters are adaptively specified by estimating the degradation of each input image. Specifically, a tiny regression network is employed to predict the degradation parameters of the input image, while several convolutional experts with the same topology are jointly optimized to specify the network parameters via a non-linear mixture of experts. The joint optimization of multiple experts and the degradation-adaptive pipeline significantly extend the model capacity to handle degradations of various levels, while the inference remains efficient since only one adaptively specified network is used for super-resolving the input image. Our extensive experiments demonstrate that DASR is not only much more effective than existing methods on handling real-world images with different degradation levels but also efficient for easy deployment. Codes, models and datasets are available at https://github.com/csjliang/DASR.

Keywords: Real-world image super-resolution, degradation-adaptive, efficient super-resolution

1 Introduction

Single image super-resolution (SISR) [1–5] is an active research topic in low-level vision, aiming at reconstructing a high-resolution (HR) version of a degraded low-resolution (LR) image. Since the seminal work of SRCNN [6], many convolutional neural network (CNN) based SISR methods [7–11] have been proposed, most of which assume a pre-defined degradation process (*e.g.*, bicubic downsampling) from HR to LR images. Despite the great success, the performance

⋆ Corresponding author.

⋆⋆ This work is supported by the Hong Kong RGC RIF grant (R5001-18) and the PolyU-OPPO Joint Innovation Lab.

of these non-blind SISR methods will deteriorate a lot when facing real-world images [12] because of the mismatch of degradation models between the training data and the real-world test data [13].

The blind image super-resolution (BISR) methods [12,14–17] have been proposed to address the problems of non-blind SISR methods by considering more complex degradation kernels extracted from real-world images. However, the degradation space of these methods is actually restricted to a set of pre-collected kernels, such as the DPED kernel pool [17, 18]. For real-world images, their degradation space can be much larger, including more types and more complex kernels than the DPED kernel pool, more complex and stronger noise, and other degradation operations such as compression. Therefore, many recent researches have been focused on the real-world image super-resolution (Real-ISR) tasks [19–26] by modeling and synthesizing the complex degradation process of real-world images [27, 28]. The representative works include BSRGAN [13] and Real-ESRGAN [29], which introduce comprehensive degradation operations such as blur, noise, down-sampling, and JPEG compression, and control the severity of each operation by randomly sampling the respective hyper-parameters. Random shuffle of degradation orders [13] and second-order degradation [29] are also employed to better simulate the real-world complex degradations, respectively.

Despite the remarkable progress of BSRGAN [13] and Real-ESRGAN [29] on improving the image perceptual quality, they have several limitations for practical usage. On one hand, they are basically designed to work on severely degraded LR images. While BSRGAN and Real-ESRGAN can generate a certain amount of details on some tough LR images, they are difficult to generate fine details on mildly degraded LR inputs. It is highly anticipated to develop Real-ISR models which can handle images with different degradation levels. On the other hand, the BSRGAN and Real-ESRGAN methods rely on heavy backbone networks (*e.g.*, RRDB [2]), which make them difficult to be deployed on devices with limited computational resources [30–34]. It is also anticipated to develop efficient Real-ISR models to meet the requirement of high efficiency.

To tackle the above problems, in this paper, we propose a degradation-adaptive super-resolution (DASR) network whose parameters are adaptively specified to the given image according to its degradation. Our DASR consists of a tiny regression network to estimate the degradation parameters of the input image and multiple light-weight super-resolution experts, which are jointly optimized on a balanced degradation space. For each input image, an adaptive network is constructed via a non-linear mixture of experts, whose adaptive weighting factors are specified by the estimated degradation parameters. The multiple super-resolution experts and the degradation-aware mixture significantly improve the model capacity for handling images of different degradations. Meanwhile, the whole pipeline of DASR is highly efficient against existing methods to meet the requirement of Real-ISR tasks, as only one adaptive network is employed to super-resolve the image during inference and the cost of mixing experts is negligible.

The contributions of this paper are two-fold. First, we propose a degradation-adaptive super-resolution network, which significantly improves the model capacity to super-resolve images of various degradation levels. Second, the pipeline of our DASR network is highly efficient against existing methods, providing a good solution to perform Real-ISR in practical applications. Extensive experiments verified the effectiveness and efficiency of the proposed method.

2 Related Work

2.1 Real-World Image Super-Resolution

How to reproduce effectively and efficiently HR images from low-quality real-world LR images is a challenging issue in SISR research. The distribution of real-world images can differ dramatically due to the varying image degradation process, different imaging devices, and image signal processing methods [12, 28]. Researchers [19, 35] have tried to capture real-world HR-LR image pairs by adapting the focal length of the camera, yet the collection of data is tedious and this can only describe a limited subspace of image degradation. Some unsupervised methods [23, 28] have also been proposed to explore the domain adaptation between the synthesized LR image and the real one, yet the domain gap is still big, deteriorating the SR performance [21, 22].

Recently, several Real-ISR methods such as BSRGAN [13], Real-ESRGAN [29] and SwinIR [36] have achieved remarkable progress by introducing comprehensive degradation models to effectively synthesize real-world images. However, they rely on a heavy and computationally intensive backbone network, *e.g.*, RRDB [2] and Swin transformer [37], and are not flexible to process images of different degradation levels. In this paper, we propose a degradation-adaptive framework to address this issue, targeting at an effective and efficient network for the challenging Real-ISR task.

2.2 Image Degradation Modeling

In many non-blind SISR methods [1–4, 38–40], the degradation model is simply assumed as bicubic down-sampling or blurred down-sampling with a Gaussian kernel. The performance of these non-blind methods can be dramatically undermined when applied to images with different degradations [12]. As a remedy, SRMD [14], UDVD [41] and some other methods [42, 43] extend the degradation space to cover more blur kernels and noise levels, and use the degradation map as additional input to perform conditional SISR. While these methods can handle multiple degradations with a single model, they rely on accurate degradation estimation, which itself is also a challenging task.

A few blind SISR methods have been proposed for handling unknown degradations [27, 28, 44–48]. In KMSR [17], a kernel pool is constructed from real photographs using generative adversarial network [49], and training pairs are synthesized in a more realistic way. Some methods like IKC [16] and VBSR [50]

incorporate a blur kernel estimator into the SISR framework to be adaptive to images degraded with different blur kernels [15, 51]. However, most of the blind SISR methods are trained with a pre-collected kernel pool [17, 18], and hence they are not really blind and can hardly be generalized to real-world images.

Recent Real-ISR methods such as BSRGAN [13] and Real-ESRGAN [29] further extend the degradation modeling space by incorporating comprehensive degradation types with randomly sampled degradation parameters to enhance the variation. The larger degradation space helps the trained Real-ISR model to improve the perceptual quality of some tough LR inputs. However, the degradation parameter sampling in BSRGAN and Real-ESRGAN is unbalanced to train a flexible network, limiting the trained model in generating fine details, especially for inputs with mild degradations. In this work, we propose to balance the degradation space by partitioning it into three levels with balanced samplings. Such a balanced degradation space facilitates the optimization of our degradation-adaptive model on different degradation levels and brings a better approximation to the real-world LR images.

2.3 Mixture of Experts and Dynamic Convolution

The mixture of experts (MoE, [52–55]) is a long-standing method that calculates the weighted sum of multiple expert networks to improve the performance. A trainable gating network is employed to compute the weight for activating each expert [56], usually based on an explicit (*e.g.*, labeled classes) or implicit (content clustering) partition of the data. In this paper, we calculate the adaptive weight of experts according to the degradation of the image for the Real-ISR tasks. Besides, instead of activating all experts and calculating the weighted sum of outputs as in previous MoE methods [57], we adaptively mix the network parameters, resulting in only one adapted network for inference. Such a pipeline is effective and efficient due to the increased non-linearity and the fast inference.

Dynamic convolution [58, 59] or conditional convolution [60, 61] aims to enhance the feature representation capacity by making the convolutional parameters sample-adaptive. Most of the existing methods optimize multiple sets of convolutional parameters and learn feature self-attention to linearly combine the parameters. However, this pipeline introduces many computations to obtain self-attention, causing a trade-off between efficiency and effectiveness. In this paper, we achieve the non-linear mixture of experts via an adapted conditional convolution, where the conditions are the degradation parameters and the weighting factors are calculated once for all layers to keep efficiency.

3 Methodology

This section presents our degradation-adaptive network for real-world image super-resolution, *i.e.*, DASR. As shown in Figure 1, DASR mainly consists of a degradation prediction network and a CNN-based SR network with multiple experts. In the following sections, we first provide the details of the proposed

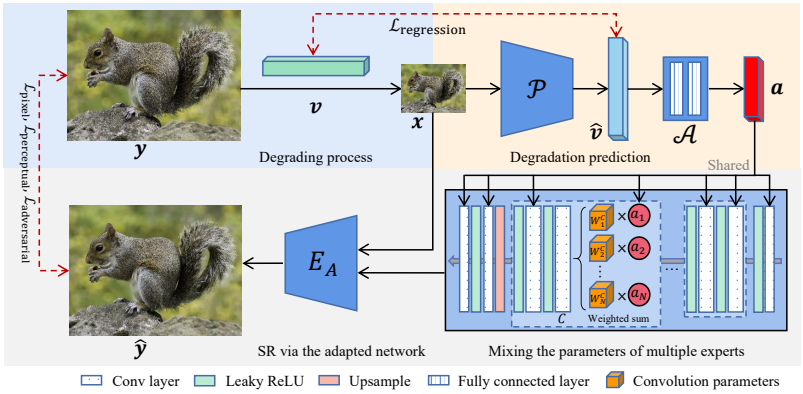

Fig. 1. Overall pipeline of the proposed DASR. Here, $\boldsymbol{x}, \boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ denote the LR image, the ground truth HR image and the super-resolved result, respectively. For each convolution layer C, the parameters W_i^C of N experts are mixed according to the weighting factors in $\boldsymbol{a}$. The input $\boldsymbol{x}$ is super-resolved to $\hat{\boldsymbol{y}}$ by the adapted network E_A.

DASR framework and then introduce our degradation modeling to set degradation parameters and generate training pairs.

3.1 Degradation-Adaptive Super-Resolution

Degradation prediction network. To allow efficient and degradation-adaptive super-resolution, we propose to estimate the degradation parameters $\boldsymbol{v} \in \mathbb{R}^{1\times n}$ of each input $\boldsymbol{x}$ via a regression network $\mathcal{P}$, *i.e.*, $\hat{\boldsymbol{v}} = \mathcal{P}(\boldsymbol{x})$, where $\hat{\boldsymbol{v}}$ denotes the estimation of $\boldsymbol{v}$. We employ a set of parameters $\boldsymbol{v}$ to elaborately describe the degradation space. The details of degradation space modeling will be discussed in Section 3.2. To make the estimation process efficient, we design a light-weighted network $\mathcal{P}$ to predict $\boldsymbol{v}$. Specifically, $\mathcal{P}$ consists of 6 convolution layers with Leaky ReLU activation, followed by a global average pooling layer. We first use convolution layers to extract image spatial degradation features and then use the global average pooling layer to estimate the degradation parameters.

To optimize the network $\mathcal{P}$, we introduce a regression loss between the estimated degradation parameters $\hat{\boldsymbol{v}}$ and the ground-truth $\boldsymbol{v}$ using the ℓ_1-norm distance as follows:

$$\mathcal{L}_{\text{regression}} = \|\hat{\boldsymbol{v}} - \boldsymbol{v}\|_1. \tag{1}$$

According to the degradation model, each parameter in $\boldsymbol{v}$ is randomly sampled to specify the degradation process to generate the LR-HR image pairs.

Image super-resolution network. An ideal Real-ISR method is expected to be both effective and efficient. On one hand, in real-world SR tasks, the computation resources are usually limited, especially for edge devices. On the other hand, the model should be able to effectively handle images with various kinds of degradations. Nevertheless, most of the current SR methods [13, 29, 36, 38, 62] can only trade-off between efficiency and effectiveness, and they are inflexible to handle images with different degradation types and levels.

To develop an effective and efficient Real-ISR model, we propose a degradation-adaptive SR network to boost the model capacity via non-linear mixture of experts (MoE), whose additional cost is negligible during inference. In specific, we employ N convolutional experts, denoted by $\boldsymbol{E} = [E_1, E_2, \cdots, E_N]$, where each expert E_i is a light-weighted SR network, *e.g.*, SRResNet [38] or EDSR-M [62], with independent parameters $\boldsymbol{\Phi}_{E_i}$. All the E_i share the same network topology, and they are optimized jointly with the supervision of the same loss. Our idea is to implicitly train each expert to handle images falling into a sub-space of the degradation space so that they can work together to process images with various kinds of degradations in the whole space.

A vector of weighting factors $\boldsymbol{a} \in \mathbb{R}^{1\times N}$, which is adaptive to the degradation of the input $\boldsymbol{x}$, is then calculated to adaptively mix the N experts. We calculate $\boldsymbol{a}$ conditioned on the estimated $\hat{\boldsymbol{v}}$ via a tiny network $\mathcal{A}$ with two fully-connected layers, *i.e.*, $\boldsymbol{a} = \mathcal{A}(\hat{\boldsymbol{v}})$. As both $\hat{\boldsymbol{v}}$ and $\boldsymbol{a}$ are of low dimension ($n = 33$ and $N = 5$ in our experiments), the network $\mathcal{A}$ is highly efficient. Note that if $\boldsymbol{a}$ is constrained to be a one-hot vector, only one expert will be activated for super-resolving the input $\boldsymbol{x}$, and this will degrade our framework to a competitive MoE [56], which may perform well on tasks whose sample distribution space can be partitioned with clear boundaries, yet it can hardly work well for the Real-ISR task with a large and continuous degradation space.

With the multiple experts $\boldsymbol{E}$ and their adaptive weighting factors $\boldsymbol{a}$, we mix the experts adaptively in a non-linear manner. For each convolution layer C of the desired network, we employ the dynamic convolution technique [58, 61] to parameterize the convolutional kernels as follows:

$$\boldsymbol{f}_{\text{output}} = \sigma((a_1 \cdot W_1^C + a_2 \cdot W_2^C + \cdots + a_N \cdot W_N^C) * \boldsymbol{f}_{\text{input}}). \tag{2}$$

where $\boldsymbol{f}_{\text{input}}$ and $\boldsymbol{f}_{\text{output}}$ denote the input and the output features, a_i indicates the i^{th} value of $\boldsymbol{a}$, W_i^C denotes the layer C parameters for expert E_i and σ is the activation function. That is, we adaptively fuse the parameters of each layer among all experts, resulting in an adaptive network, denoted as E_A.

Note that in classic dynamic convolution, the weighting factor of each layer is calculated by an independent network conditioned on the feature map of the last layer, thus introducing non-negligible computational costs. In contrast, we learn a single set of degradation-adaptive weighting factors $\boldsymbol{a}$ for all convolution layers, which is very efficient. Our framework follows the spirit of MoE but in a non-linear manner due to the activation operation in intermediate layers. The non-linearity and the degradation-adaptive mixture of multiple experts significantly extend the model capacity to handle degradations of various levels.

Our DASR is very efficient. For each convolutional layer, the model only deploys one adapted network E_A in the inference stage, rather than deploying N models as done in the classic MoE methods [52, 53]. The degradation prediction network $\mathcal{P}$ and the weighting module $\mathcal{A}$ are also very light-weighted. Therefore, the cost of inference is of the same order as one single expert network. The computational overhead caused by the mixture operation is negligible. Specifically, the mixture process consists of multiplications and additions operations on the

parameters of N experts. For a light-weighted backbone network like SRResNet or EDSR-M, the number of parameters of each expert is only $1.52M$, and they are independent of the size of input images. Therefore, compared with the calculation of multiple feature maps, the complexity of the mixture of parameters is several orders of magnitude lower and thus can be neglected.

3.2 Degradation Modeling

Since high-quality real-world LR-HR pairs are hard to be collected due to the misalignment issue [19, 35], the degradation modeling is very important to synthesize real-world LR inputs $\boldsymbol{x}$ from a given HR image $\boldsymbol{y}$ for Real-ISR model training. A degradation space, denoted by S, should be pre-defined to synthesize training pairs and perform degradation-adaptive optimization. The quality of an LR sample $\boldsymbol{x}$ in S is controlled by a degradation parameter vector $\boldsymbol{v} = [v_1, v_2, \cdots, v_n]$, where v_i specifies the type or severity of a degrading operation and n denotes the number of degradation parameters. In our DASR, $\boldsymbol{v}$ also serves as the ground-truth for training the degradation prediction network.

The image degradation model has been recently improved significantly from the simple bicubic down-sampling [2, 6] to shuffling [13] and second-order [29] pipelines. We adopt the degradation operations of blurring (both isotropic and anisotropic Gaussian blur), resizing (both down-sampling and up-sampling with area, and bilinear and bicubic operations), noise corruption (both additive Gaussian and Poisson noise), and JPEG compression in our modeling. In $\boldsymbol{v}$, we use a one-hot code to quantify the degradation operation type and use a single value to record the degradation level normalized by its respective dynamic range.

It is worth mentioning that different from the methods [14, 16] which quantify a blur kernel by its kernel coefficients, we quantify a blurring degradation by its kernel size s, the standard deviation σ_1, σ_2 along the two principal axes, and the rotation degree θ. In this way, the degradation parameters are more interpretable to specify the degradation types and levels, and can better support the degradation-aware mixture of experts. Meanwhile, the parameter vector $[s, \sigma_1, \sigma_2, \theta]$ has only 4 dimensions, while the kernel vector $\boldsymbol{k}$ will have much higher dimensions to estimate. Benefiting from the interpretability and compactness of the degradation space, our DASR allows explicit user control towards degradation parameters during inference. This can facilitate many user-interactive applications to customize the desired super-resolving effect.

Though the shuffling degradation method in BSRGAN [13] and the second-order degradation pipeline in Real-ESRGAN [29] can generate a sufficiently large degradation space, it is hard for them to train a model which can adaptively handle images with different levels of degradations. Our DASR is designed to be adaptive to a wide range of real-world inputs with multiple light-weight expert networks, each of which is expected to handle a subspace of images of different degradation levels. Therefore, we partition the whole degradation space S into 3 levels $[S_1, S_2, S_3]$ by specifying the parameters $\boldsymbol{v}$ accordingly. Among them, S_1 and S_2 are generated with first-order degradation with small and large parameter ranges, respectively, while S_3 is generated by the second-order degradation.

Due to space limitation, more details of the degradation operations and the specification of $[S_1, S_2, S_3]$ are provided in the supplimentary material.

3.3 Training Losses

The learnable modules of our DASR network include $[\boldsymbol{E}, \mathcal{P}, \mathcal{A}]$. As mentioned in Section 3.1, the $\mathcal{L}_{\text{regression}}$ loss is used to optimize $\mathcal{P}$ to predict the degradation parameters. To optimize the overall framework, following the many works in literature [2, 13, 29], we adopt the L_1-norm pixel-wise loss $\mathcal{L}_{\text{pixel}}$, the perceptual loss $\mathcal{L}_{\text{perceptual}}$ and the adversarial loss $\mathcal{L}_{\text{adversarial}}$. The total loss is defined as follows (more details are provided in the supplimentary material):

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pixel}} + \lambda_1 \mathcal{L}_{\text{regression}} + \lambda_2 \mathcal{L}_{\text{perceptual}} + \lambda_3 \mathcal{L}_{\text{adversarial}}, \tag{3}$$

where λ_1, λ_2 and λ_3 denote the balancing parameters.

4 Experiments

4.1 Training Details

Following previous works [2,29], we employ DIV2K, Flickr2K, and OutdoorScene-Training datasets for training our DASR model. For efficiency, we employ the SRResNet [38] as our backbone. The weights of the N experts are initialized by the model pre-trained with pixel-wise loss. The Adam [63] optimizer is employed to train the network. The learning rate is set to $1 \times e^{-4}$, the total batch size is 24 and the training iteration is set to $500K$. We balance the training loss with $\lambda_1 : \lambda_2 : \lambda_3 = 1 : 1 : 0.1$. Without loss of generality and for a fair comparison, we conduct Real-ISR experiments with the scale factor of 4 by following the setting in BSRGAN [13] and Real-ESRGAN [29]. In our experiment, the dimension of degradation parameters is $n = 33$ and the number of experts is $N = 5$. The LR patch size is set to 64×64.

4.2 Evaluation and Compared Methods

We evaluate our DASR method both quantitatively and qualitatively. For quantitative evaluation, as in BSRGAN [13] we synthesize 300 LR-HR pairs by applying the 3 levels of degradations to the 100 validation images in the DIV2K dataset, *i.e.*, 100 LR-HR pairs for each level. We also make the comparison on the original DIV2K dataset with bicubic downsampling. An illustration of images with different degradations is shown in Fig. 2, where more samples are shown in the supplementary material. For qualitative evaluation, we also employ the images in the RealSRSet [13, 29], where the input images are corrupted by various blur, noise, or other real degradation operations.

We compare the proposed DASR with representative and state-of-the-art SR methods, including RRDB [2], ESRGAN [2], IKC [16], BSRGAN [13], Real-ESRGAN [29] and Real-SwinIR (-M and -L) [36]. Among them, RRDB is trained

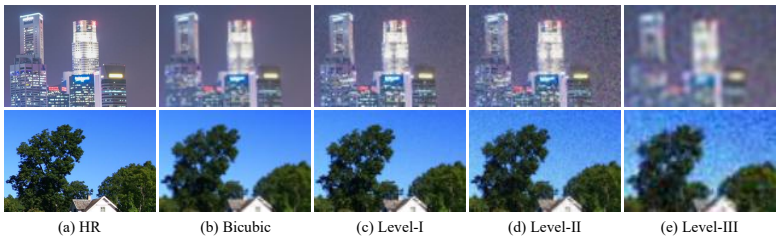

Fig. 2. Sample images with different levels of degradations in our datasets.

on bicubic degradation with pixel-wise loss; ESRGAN is trained on bicubic degradation with pixel-wise, perceptual and adversarial losses; IKC is a representative BISR method trained on various isotropic Gaussian blur kernels; BSRGAN and Real-ESRGAN are state-of-the-art Real-ISR methods with a heavy RRDB backbone; Real-SwinIR is trained on the degradation space of BSRGAN with the computationally expensive SwinIR backbone.

For a more comprehensive and fair comparison, we also re-train those commonly used backbone networks, including SRResNet, EDSR, RRDB, and SwinIR, with our constructed training dataset. Following the common practice [13, 29], we employ PSNR (the larger the better) and LPIPS (learned perceptual image patch similarity, the smaller the better) to quantitatively compare the performance of different methods on synthetic datasets, and make visual comparisons on real-world images since there are no reference images.

4.3 Quantitative Comparison

Effectiveness. In Table 1 and Table 2, we quantitatively compare the performance of competing methods in terms of PSNR and LPIPS on datasets with different levels of degradations. Specifically, Table 1 compares the methods trained with their own degradation models, while Table 2 compares the methods retrained on our proposed degradation space.

As shown in Table 1, existing methods can only achieve satisfactory performance on datasets with a specific type of degradation, yet show weakness in other cases. For example, RRDB and ESRGAN can respectively achieve good fidelity and perceptual quality on the bicubic-downsampled dataset, yet their performance drops dramatically when handling images with other degradations, even for the 'Level-I' degradation with mild noise and blurs. Real-ESRGAN, BSRGAN, and Real-SwinIR perform well on the most severely degraded dataset. However, their performance deteriorates much on the other three datasets.

In contrast, our DASR achieves stable and significant improvement against other methods under the first three types of degradations, which cover the majority of real-world images, while achieving highly competitive (among the best two) results for the last type of degradation. For example, DASR outperforms Real-ESRGAN by about 1.7dB in PSNR and 26% in LPIPS on the 'Level-I' dataset. On the 'Level-III' dataset with severely degraded images (as shown in Fig. 2 (d)), DASR achieves almost the same PSNR and LPIPS indices as BSRGAN. These observations clearly demonstrate that our DASR can generalize well to images with a wide range of degradations.

Table 1. Quantitative comparison of different methods on datasets with different degradations (D-Level). 'Bicubic' denotes the DIV2K validation set with bicubic degradation, while 'Level I', 'II', and 'III' denote the datasets with mild, medium, and severe degradations, respectively. For the compared methods, we employ their officially released pre-trained models. The PSNR is calculated on the Y channel of YCbCr space.

D-Level	Metric	RRDB	ESRGAN	IKC	BSRGAN	Real-ESRGAN	Real-SwinIR-M	Real-SwinIR-L	DASR
Bicubic	PSNR	30.92	28.17	28.01	27.32	26.65	26.83	27.21	28.55
	LPIPS	0.2537	0.1154	0.2695	0.2364	0.2284	0.2221	0.2135	0.1696
Level-I	PSNR	26.27	21.16	24.09	26.78	26.17	26.21	26.45	27.84
	LPIPS	0.3419	0.4727	0.3805	0.2412	0.2312	0.2247	0.2161	0.1707
Level-II	PSNR	26.46	22.77	25.39	26.75	26.16	26.12	26.39	27.58
	LPIPS	0.4441	0.4900	0.4531	0.2462	0.2391	0.2313	0.2213	0.2126
Level-III	PSNR	23.91	23.63	22.91	24.05	23.81	23.34	23.46	23.93
	LPIPS	0.7631	0.7314	0.7583	0.3995	0.3901	0.3844	0.3765	0.4144

To further validate the effectiveness of our degradation-adaptive strategy, in Table 2 we re-train the backbones of popular SR models on our proposed degradation space. Note that the heavy RRDB backbone is adopted in both BSRGAN and RealESRGAN, and the lightweight SRResNet is adopted in our DASR as the backbone. As can be seen from this table, with the same network topology and similar computational overhead, our DASR outperforms the baseline SRResNet on all datasets by a large margin, *e.g.*, improving 0.5db of PSNR on the bicubic-downsampled dataset and about 5% of LPIPS on the Level-II dataset. This demonstrates that the degradation-adaptive mixture of multiple experts can significantly extend the model capacity while keeping the efficiency.

Compared to RRDB and SwinIR backbones that are adopted in recent state-of-the-art methods [13,29,36], our DASR consumes much less computational resources, *e.g.*, about 1/3 and 1/12 latency of RRDB and SwinIR, respectively. At the same time, DASR outperforms these heavy models in terms of reconstruction fidelity on all datasets, demonstrating its effectiveness of degradation-adaptive super-resolution and high efficiency to deploy in practice.

Efficiency. The inference efficiency is a crucial factor in Real-ISR tasks due to the limited computational resources in practical applications. We compare different backbone networks in terms of multiple efficiency-related metrics and depict the results in the bottom rows of Table 2.

As shown in the table, the computational overhead of different backbone networks differs dramatically. For example, RRDB [2], which is employed in recent Real-ISR methods [13, 29], consumes about 7 times the FLOPs and more than 4 times the inference time than SRResNet [38]. In other words, the RRDB based Real-ISR methods achieve superior performance at the price of applicability. The recent transformer-based method SwinIR has an acceptable number of FLOPs, however, it actually consumes much more inference time due to the heavy computation of attentions and frequent IO consumption.

Benefiting from the light SRResNet-based backbone and the efficient degradation prediction and parameter fusion, our DASR is very efficient against recent

Table 2. Quantitative comparison of different backbone networks re-trained on our proposed degradation space and the efficiency comparison (the bottom rows). The evaluation datasets are the same as in Table 1. For efficiency evaluation, the input-dependent metric FLOPs is calculated on images with 256×256 pixels; the Latency and Memory are the average inference time and the maximum GPU memory allocation on the DIV2K validation dataset (most LR inputs are with 510×339 pixels). Statistics are collected following the implementation of [64,65] by using an NVIDIA V100 GPU.

Data & Metrics		SRResNet	EDSR	SwinIR	RRDB	DASR
Bicubic	PSNR	28.05	28.26	28.28	27.92	28.55
	LPIPS	0.1747	0.1807	0.1488	0.1473	0.1696
Level-I	PSNR	27.60	27.79	27.78	27.84	27.84
	LPIPS	0.1772	0.1834	0.1531	0.1569	0.1707
Level-II	PSNR	27.34	27.53	27.45	27.29	27.58
	LPIPS	0.2228	0.2284	0.1854	0.1886	0.2126
Level-III	PSNR	23.71	23.87	23.60	23.54	23.93
	LPIPS	0.4419	0.4351	0.3869	0.3847	0.4144
Latency (ms)		113	105	1719	460	142
#FLOPs (G)		166	130	539	1176	184
#Params (M)		1.52	1.52	11.72	16.70	8.07
#Memory (M)		2359	2169	2699	2417	2452

methods. In specific, the degradation prediction network $\mathcal{P}$ and the weighting module $\mathcal{A}$ consume 18G FLOPs, 18ms latency, 0.47M parameters and 111M GPU memory in total for $N = 5$. Besides, the consumption on parameter fusion operation is negligible, as there are only $N \times 1.52$M multiplications and additions respectively and they can be calculated in parallel. Compared with the classical MoE methods that mix the feature maps of all experts [52,53,57,66], our DASR only conducts one forward pass. As a result, the computational cost increases slightly with a larger N, which supports a flexible extension of model capacity.

It is worth mentioning that although our model has more parameters, the maximum GPU memory consumption does not increase much as shown in the row of #Memory in Table 2, since the deployment of model parameters costs much less space than storing input-dependent feature maps. On the other hand, the increased model parameters do not demand much storage space, which is much easier to afford than the computing power.

4.4 Qualitative Comparison

Fig. 3 shows the visual comparisons between different methods on images with different degradations. One can see that DASR can stably restore sharp and realistic details and remove artifacts for a wide range of degradations. In specific, the first sample image is degraded with bicubic downsampling and suffers from the aliasing issue. Both BSRGAN and Real-ESRGAN cannot generate satisfactory texture details even with the heavy RRDB backbone. This is because these two methods are trained on pairs with relatively severe degradations so that their denoising capacity is strengthened yet the detail-generation capacity is limited. Similar observations can be made on all the four samples in Fig. 3.

The RRDB backbone trained with pixel-wise loss performs well on the first two samples in generating textures details, yet it cannot be generalized to the

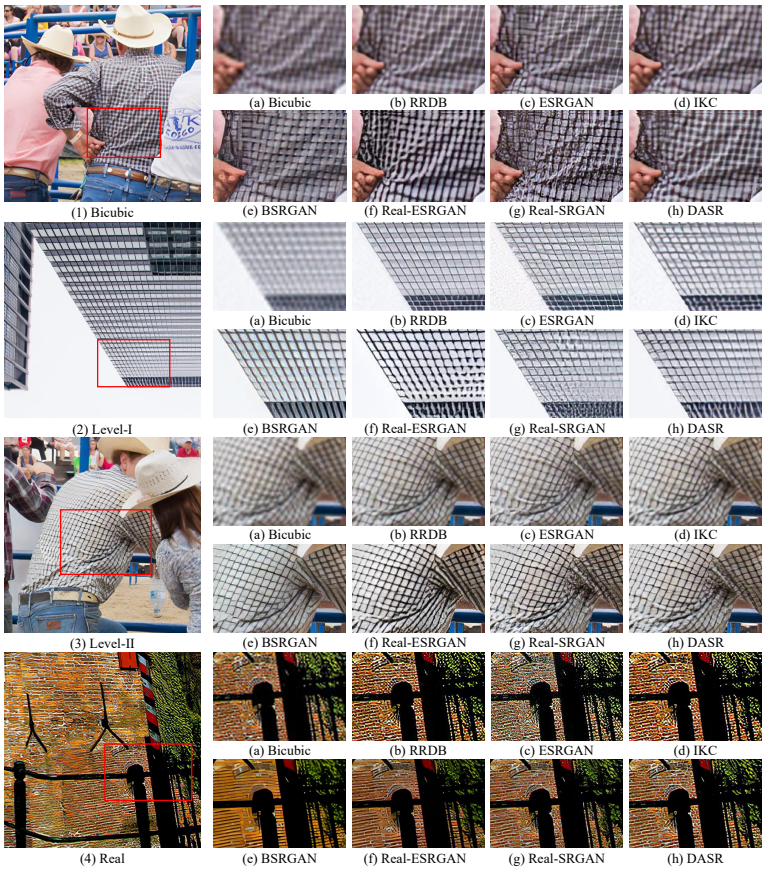

Fig. 3. Qualitative comparison of competing methods on images with different degradations. The results of (b-f) are generated by using the officially released models, while the output of (g) is obtained by re-training the SRResNet backbone with our proposed degradation model. Better zoom in for details.

last two samples whose degradations are severe. This is reasonable since all its training pairs are generated by bicubic downsampling. In addition, the results of RRDB in the first and third samples are blurry, which is a well-known side-effect of pixel-wise loss. By applying perceptual and adversarial losses, ESRGAN achieves sharper results yet introduces many visual artifacts due to the instability of training generative adversarial networks. The ESRGAN also amplifies the noise as shown in the second sample. By considering different blur kernels, IKC can restore rich textures on most images, yet bring overshoot artifacts when facing unseen kernels in real-world images (the fourth sample). It also lacks the capacity to remove noise as shown in the second sample.

The results of Real-SRGAN are obtained by re-training the SRResNet on our proposed degradation space with the same loss as in Real-ESRGAN [29]. It can be observed that due to the insufficient feature representation capacity, Real-SRGAN cannot perform well on all the four samples compared to our DASR.

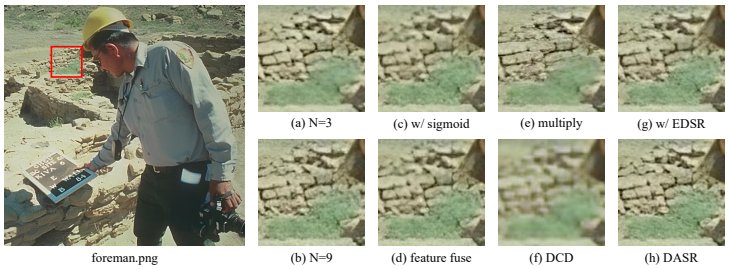

Fig. 4. Ablation study. (a) and (b) validate the models with different N; (c) appends a sigmoid layer to the weighting module $\mathcal{A}$; (d) conducts classical MoE [52, 53, 57, 66] where the output of multiple experts are fused; (e) performs dynamic convolution with a single expert by learning a mapping matrix and multiplying it to the parameters; (f) conducts dynamic convolution following the work [59]; (g) applies EDSR-M backbone to DASR; (h) denotes our default DASR model.

In the first three samples, the Real-SRGAN generates messy details or artifacts, as the light-weighted model limits its capacity to achieve degradation-adaptive super-resolution. On the last sample, which is a real-world image, Real-SRGAN fails to reconstruct rich details. In contrast, DASR reproduces realistic details and inhibits artifacts. For hard cases in the real world that our degradation space may not cover, DASR would give a reasonable 'guess' on these images or regions (*e.g.*, synthesize less details to transit smoothly) rather than outputting fatal artifacts (*e.g.*, generate false structures or details), which may happen in other methods. The stability of DASR comes from its comprehensive degradation space and the fusing strategy.

4.5 Ablation Study

We conduct comprehensive ablation studies on our proposed DASR model by using real-world images and depict the visual results in Fig. 4. More results can be found in the supplementary material.

Effectiveness of N. Models in Figs. 4(a) and (b) evaluate the selection of N. It can be seen that using 3 experts leads to relatively smooth results, while models of $N = 5$ in (h) and $N = 9$ in (b) enhance the generation of details. As $N = 9$ shows similar visual quality to $N = 5$, we consider that $N = 5$ is sufficient to model the proposed degradation space.

Effectiveness of model design. Figs. 4(c) and (d) validate the effectiveness of our model design. The result in (c) demonstrates that adding a sigmoid layer to the weighting module $\mathcal{A}$ cannot improve the performance. As we mix different experts in terms of model parameters, there is no need to ensure positive weights by a sigmoid layer. The experts in Fig. 4(d) are fused by following the strategy of classical MoE [52, 53, 57, 66], where the outputs of all experts are fused. We can see that the result of classical MoE in (d) lacks fine details compared to (h), yet its computational cost is N times heavier than our DASR.

Effectiveness of different dynamic convolutions. Figs. 4(e) and (f) compare different dynamic convolutions [41, 59] without introducing many ad-

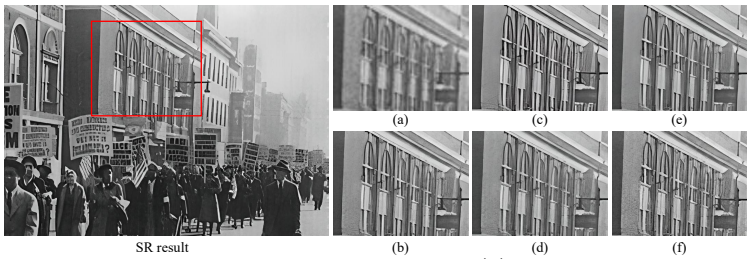

Fig. 5. Example of user-interactive super-resolution. (a) is the input image with bicubic upsampling; (b) is the result of DASR where the degradation parameters are estimated automatically by model $\mathcal{P}$; (c) and (d) are generated by manually increasing and decreasing the scale of blur kernel, respectively; (e) and (f) are the super-resolution results by manually increasing and decreasing the level of noise, respectively.

ditional parameters. While the inference latency and FLOPs are increased, the performance of those methods drops, *e.g.*, the artifacts generated in (e). We believe it is the joint optimization of multiple experts and the degradation-adaptive mixture that make our DASR more effective than other methods.

Generalization to different backbone. Fig. 4(g) applies the EDSR-M backbone to DASR. The satisfactory perceptual quality of (g) demonstrates the generalization capacity of our proposed DASR to different backbone networks.

4.6 User-Interactive Super-resolution

One interesting advantage of our DASR over other Real-ISR methods is that it supports easy user-interactive super-resolution during inference, owing to its interpretable and compact degradation representation.

We depict an example of user-interactive super-resolution in Fig. 5. As can be seen, the proposed DASR allows explicit user control to customize the super-resolution effects. Manually setting larger values to the blur-related parameters (*e.g.*, kernel scale) leads to sharper super-resolution results, as shown in Fig. 5(c), while adjusting the level of noise can flexibly balance between image details and noise, as shown in Figs. 5(e) and (f). Such an advantage of flexible user control makes our DASR very attractive in practical Real-ISR tasks.

5 Conclusion

In this paper, we proposed an efficient degradation-adaptive network, namely DASR, for the real-world image super-resolution (Real-ISR) task. In order to improve the modeling capacity and flexibility of various degradation levels, we jointly learned multiple super-resolution experts and adaptively mixed them into one expert in a degradation-aware manner. The proposed DASR was not only degradation adaptive but also efficient during inference. Extensive quantitative and qualitative experiments were conducted. The results demonstrated that DASR not only achieved superior performance on images with a wide range of degradation levels but also kept good efficiency for easy deployment. In addition, DASR allowed easy user control for customized super-resolution results.

References

1. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1, 3
2. 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 *ECCVW*, 2018. 1, 2, 3, 7, 8, 10
3. Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 1, 3
4. Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *CVPR*, 2020. 1, 3
5. Jian Sun, Zongben Xu, and Heung-Yeung Shum. Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Transactions on Image Processing*, 20(6):1529–1542, 2010. 1
6. Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 1, 7
7. Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *ICCV*, 2017. 1
8. Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *CVPR*, 2016. 1
9. Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In *CVPR*, 2019. 1
10. Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, 2018. 1
11. Younghyun Jo, Seoung Wug Oh, Peter Vajda, and Seon Joo Kim. Tackling the ill-posedness of super-resolution through adaptive target generation. In *CVPR*, 2021. 1
12. Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond. *arXiv preprint arXiv:2107.03055*, 2021. 2, 3
13. Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, 2021. 2, 3, 4, 5, 7, 8, 9, 10
14. Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2018. 2, 3, 7
15. Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020. 2, 4
16. Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 2, 3, 7, 8
17. Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *ICCV*, 2019. 2, 3, 4
18. Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In *ICCV*, 2017. 2, 4
19. 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 *ICCV*, 2019. 2, 3, 7

20. Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *ECCV*, 2020. 2
21. Andreas Lugmayr, Martin Danelljan, and Radu Timofte. NTIRE 2020 challenge on real-world image super-resolution: Methods and results. In *CVPRW*, 2020. 2, 3
22. Andreas Lugmayr, Martin Danelljan, Radu Timofte, Manuel Fritsche, Shuhang Gu, Kuldeep Purohit, Praveen Kandula, Maitreya Suin, AN Rajagoapalan, Nam Hyung Joon, et al. AIM 2019 challenge on real-world image super-resolution: Methods and results. In *ICCVW*, 2019. 2, 3
23. Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *ICCVW*, 2019. 2, 3
24. Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *ICCVW*, 2019. 2
25. Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *CVPRW*, 2020. 2
26. Haoyu Ren, Amin Kheradmand, Mostafa El-Khamy, Shuangquan Wang, Dongwoon Bai, and Jungwon Lee. Real-world super-resolution using generative adversarial networks. In *CVPRW*, 2020. 2
27. Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. To learn image super-resolution, use a GAN to learn how to do image degradation first. In *ECCV*, 2018. 2, 3
28. Yunxuan Wei, Shuhang Gu, Yawei Li, Radu Timofte, Longcun Jin, and Hengjie Song. Unsupervised real-world image super resolution via domain-distance aware training. In *CVPR*, 2021. 2, 3
29. Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *ICCVW*, 2021. 2, 3, 4, 5, 7, 8, 9, 10, 12
30. Chao Dong, Change Loy Chen, and Tang Xiaoou. Accelerating the super-resolution convolutional neural network. In *ECCV*, 2016. 2
31. Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *ACM Multimedia*, 2021. 2
32. Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, 2018. 2
33. Wenming Yang, Wei Wang, Xuechen Zhang, Shuifa Sun, and Qingmin Liao. Lightweight feature fusion network for single image super-resolution. *IEEE Signal Processing Letters*, 26(4):538–542, 2019. 2
34. Dehua Song, Yunhe Wang, Hanting Chen, Chang Xu, Chunjing Xu, and DaCheng Tao. Addersr: Towards energy efficient image super-resolution. In *CVPR*, 2021. 2
35. Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *CVPR*, 2019. 3, 7
36. Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *ICCVW*, 2021. 3, 5, 8, 10
37. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 2021. 3
38. 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 *CVPR*, 2017. 3, 5, 6, 8, 10
39. Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 2018. 3
40. Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution. In *ICCV*, 2021. 3
41. Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *CVPR*, 2020. 3, 13
42. Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR*, 2020. 3
43. Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *CVPR*, 2019. 3
44. Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *CVPR*, 2021. 3
45. Zheng Hui, Jie Li, Xiumei Wang, and Xinbo Gao. Learning the non-differentiable optimization for blind super-resolution. In *CVPR*, 2021. 3
46. Pengju Liu, Hongzhi Zhang, Yue Cao, Shigang Liu, Dongwei Ren, and Wangmeng Zuo. Learning cascaded convolutional networks for blind single image super-resolution. *Neurocomputing*, 417:371–383, 2020. 3
47. Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. In *CVPR*, 2020. 3
48. 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 *CVPRW*, 2018. 3
49. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014. 3
50. Victor Cornillere, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. 3
51. Soo Ye Kim, Hyeonjun Sim, and Munchurl Kim. KOALAnet: Blind super-resolution using kernel-oriented adaptive local adjustment. In *CVPR*, 2021. 4
52. Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991. 4, 6, 11, 13
53. Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214, 1994. 4, 6, 11, 13
54. Sam Gross, Marc'Aurelio Ranzato, and Arthur Szlam. Hard mixtures of experts for large scale weakly supervised vision. In *CVPR*, 2017. 4
55. Rahaf Aljundi, Punarjay Chakravarty, and Tinne Tuytelaars. Expert gate: Lifelong learning with a network of experts. In *CVPR*, 2017. 4
56. Shunta Maeda. Fast and flexible image blind denoising via competition of experts. In *CVPRW*, 2020. 4, 6
57. Yifan Wang, Lijun Wang, Hongyu Wang, Peihua Li, and Huchuan Lu. Blind single image super-resolution with a mixture of deep networks. *Pattern Recognition*, 102:107169, 2020. 4, 11, 13

58. Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu. Dynamic convolution: Attention over convolution kernels. In *CVPR*, 2020. 4, 6
59. Yunsheng Li, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Ye Yu, Lu Yuan, Zicheng Liu, Mei Chen, and Nuno Vasconcelos. Revisiting dynamic convolution via matrix decomposition. In *ICLR*, 2021. 4, 13
60. Chao Li, Aojun Zhou, and Anbang Yao. Omni-dimensional dynamic convolution. In *ICLR*, 2021. 4
61. Brandon Yang, Gabriel Bender, Quoc V Le, and Jiquan Ngiam. Condconv: Conditionally parameterized convolutions for efficient inference. *NeurIPS*, 2019. 4, 6
62. Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017. 5, 6
63. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 8
64. Kai Zhang, Martin Danelljan, Yawei Li, Radu Timofte, et al. Aim 2020 challenge on efficient super-resolution: Methods and results. In *ECCVW*, 2020. 11
65. Kai Zhang, Shuhang Gu, Radu Timofte, et al. Aim 2019 challenge on constrained super-resolution: Methods and results. In *ICCVW*, 2019. 11
66. Mohammad Emad, Maurice Peemen, and Henk Corporaal. MoESR: Blind super-resolution using kernel-aware mixture of experts. In *WACV*, 2022. 11, 13