Unsupervised Degradation Representation Aware Transform for Real-World Blind Image Super-Resolution

Sen Chen[1*], Hongying Liu[2,3*], Chaowei Fang[1], Fanhua Shang[4†], Yuanyuan Liu[1†], Liang Wan[2,4], Dongmei Jiang[3], Yaowei Wang[3]

[1]School of Artificial Intelligence, Xidian University, China
[2]Medical College, Tianjin University, Tianjin, China
[3]Peng Cheng Lab, Shenzhen, China
[4]College of Intelligence and Computing, Tianjin University, Tianjin, China
chensen@stu.xidian.edu.cn; hyliu2009@tju.edu.cn; fhshang@tju.edu.cn; yyliu@xidian.edu.cn

Abstract

Blind image super-resolution (blind SR) aims to restore a high-resolution (HR) image from a low-resolution (LR) image with unknown degradation. Many existing methods explicitly estimate degradation information from various LR images. However, in most cases, image degradations are independent of image content. Their estimations may be influenced by the image content resulting in inaccuracy. Unlike existing works, we design a dual-encoder for degradation representation (DEDR) to preclude the influence of image content from LR images. This benefits in extracting the intrinsic degradation representation more accurately. To the best of our knowledge, this paper is the first work that estimates degradation representation through filtering out image content. Based on the degradation representation extracted by DEDR, we present a novel framework, named **degradation representation aware transform network (DRAT)** for blind SR. We propose global degradation aware (GDA) blocks to propagate degradation information across spatial and channel dimensions, in which a degradation representation transform module (DRT) is introduced to render features degradation-aware, thereby enhancing the restoration of LR images. Extensive experiments are conducted on three benchmark datasets (including Gaussian 8, DIV2KRK, and real-world datasets) under large scaling factors with complex degradations. The experimental results demonstrate that DRAT surpasses state-of-the-art supervised kernel estimation and unsupervised degradation representation methods.

Code — https://github.com/KKKc3231/DRAT

1 Introduction

In single image super-resolution (SISR), the aim is to restore a given low-resolution (LR) image to a high-resolution (HR) image. Traditional SISR approaches (Lim et al. 2017; Zhang et al. 2018a,b; Wang et al. 2023; Xia et al. 2022) often hinge on predetermined downsampling operations. However, these methods show limitations when facing the dynamic and varied nature of degradations. Hence, our study concentrates on blind Super-Resolution (blind SR), specifically targeting multiple degradations. The degradation model for blind SR task assumes that the LR image y is obtained as follows:

$$y = (x * k_b)\downarrow_s + n, \tag{1}$$

where x is the HR image, k_b denotes the blur kernel, $*$ denotes convolution operation, $\downarrow_s$ represents the downsampling with scale factor s, and n denotes the random noise.

The study of blind SR is particularly challenging due to complex degradation and degradation estimation from LR image (Zhang et al. 2021; Luo et al. 2022b; Wang et al. 2021b). Supervised Kernel Estimation (SKE) methods (Liang et al. 2021b; Tao et al. 2021; Huang et al. 2020) employ a two-step process where an estimated blur kernel is first derived and then applied within the SR network to facilitate image restoration. The main issue is the inaccuracy of the blur kernel estimation, which may lead to significant performance decline and the artifacts in the restored images. To overcome these limitations, recent trends, such as the methods (Wang et al. 2021a; Xia et al. 2023;

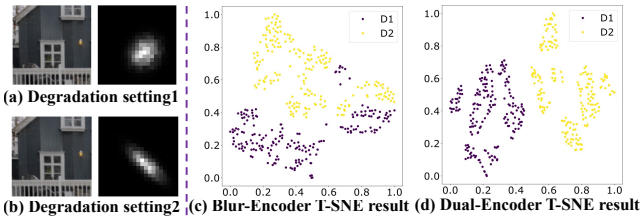

Figure 1: T-SNE visualization for the features extracted by conventional blur-encoder and our dual-encoder. Given an image, two different blur kernels are used to degrade it as shown in (a) and (b). The features extracted by conventional blur-encoder (Wang et al. 2021a; Xia et al. 2023) and our proposed dual-encoder on the degraded images are visualized in (c) and (d), respectively. D1 and D2 indicate features extracted from images degraded with kernels in (a) and (b), respectively. T-SNE is used to reduce the feature dimension. We can see that the features extracted by our dual-encoder are more discriminative to the degradation type than those features extracted by the conventional blur encoder.

*These authors contributed equally.
†Corresponding author.

Zhou et al. 2022) have leaned towards unsupervised learning paradigms, aiming to internalize degradation representation without direct supervision. However, existing unsupervised degradation learning methods (Wang et al. 2021a; Xia et al. 2023) usually adopt a blur-encoder for feature extraction. They can hardly discriminate different degradations as shown in **Figure** 1 (c). Inspired by the study of (Luo et al. 2022b; Liu et al. 2021), we consider that image degradation is usually relevant to factors like sensor limitations, transmission errors, noise interference, and processing methods, and it is independent of image content. Existing methods inadvertently incorporate image content in the learning, which undermines the quality and effect of the blind SR process.

Addressing the gaps identified in existing methodologies, we present a novel method named degradation representation aware transform network (DRAT) for blind SR. This approach enables learning more discriminative degradation representation. The cornerstone of our approach is the development of a dual-encoder for degradation representation (DEDR) that significantly reduces the incorporation of image content into degradation representation. This innovation allows for the extraction of accurate degradation by subtracting image content captured by an auxiliary encoder from LR embeddings. This is verified in **Figure** 1 (d).

Moreover, we propose a global degradation aware (GDA) block. This component enhances the propagation of degradation information throughout the network. The GDA block incorporates a degradation representation transform (DRT) module, which ingeniously integrates degradation aware information into the feature enhancement process. This not only ensures effective handling of degradation but also improves the quality of the blind SR output. To the best of our knowledge, this is the first work that filtering out image content to estimate degradation representation and using it to transform features for blind SR. The main contributions are summarized as follows:

- We propose a new framework DRAT for blind SR, in which a dual-encoder is designed to reduce the influence from image content in degradation representation, enabling more discriminative degradation representation.
- We present a global degradation aware block that propagates degradation information in spatial and channel-wise, enhancing the network's ability to handle various complex degradations.
- We also propose a degradation representation transform module which integrates degradation aware information for superior feature enhancement.
- Through extensive experiments and comparisons, we demonstrate that our proposed degradation representation aware transform network surpasses existing methods, achieving state-of-the-art performance in blind SR even more complex real-world degradation.

2 Related Work

2.1 Supervised Kernel Estimation for Blind SR

Since SRCNN (Dong et al. 2014) was first proposed for image SR, more works have focused on deep learning techniques (Dai et al. 2019; Haris, Shakhnarovich, and Ukita 2018; Kim, Lee, and Lee 2016a,b; Lai et al. 2017; Liang et al. 2021a). However, most SR methods are based on a predefined single degradation setting (e.g., bicubic downsampling). When confronting more complex degradations, they are likely ineffective. Recently, an increasing number of works employ SKE methods to enable the model for a wider range of degradations. SRMD (Zhang, Zuo, and Zhang 2018) is the first method using the concatenated input of LR images with a stretched blur kernel map to target multiple degradations. Gu et al. (2019) proposed to apply spatial feature transform (SFT) and iterative kernel correction (IKC) to estimate the blur kernel from LR and iteratively correct the estimated kernel using intermediate SR results. Luo et al. (2021, 2023) introduced deep alternating network (DAN) for the first time to use blur kernel estimation and image restoration as an end-to-end training process. Luo et al. (2022a) estimated the degradation kernel and involved it in the deep constrained least squares (DCLS) convolution in the feature domain to fully utilize the blur kernel to restore clean SR. However, all of the above methods for blur kernel estimation require the supervision of the blur kernel. Once the blur kernel is estimated inaccurately, it may cause distortion and artifacts.

2.2 Unsupervised Degradation Prediction for Blind SR

The unsupervised degradation prediction methods do not require supervision with blur kernels and learn the degradation representation from LR images using an unsupervised learning strategy. Compared to SKE methods, UDP methods are able to solve SR tasks with more complex degradations. Wang et al. (2021a) proposed DASR, which used contrastive learning to learn potential degradation representation in LR and reshaped the degradation representation into degradation kernel to guide blind SR. Li et al. (2022) proposed to optimize the HR image detail and latent structural context in an iterative manner without the incorporation of the blur kernel prior. CDCN (Wu et al. 2022) decoupled the image into structure and detail components and supervised the image detail and structure restoration in a degradation-driven manner. It attained the SR result without blur kernel information. MRDA (Xia et al. 2023) introduced meta-learning into SR network to implicitly extract degradation information for restoration. ResShift (Yue, Wang, and Loy 2024) shifted the residual between HR and LR, improved the transition efficiency for diffusion-based SR methods. Unlike above methods, by filtering out image content, we focus on enhancing the accuracy of degradation representation and employ it to improve the performance of blind SR.

3 Proposed Method

Figure 2 shows the framework of our proposed method for blind SR. It mainly consists of three parts: a dual-encoder for degradation representation, a degradation representation transform and global degradation aware blocks. The detailed explanation is in the following subsections.

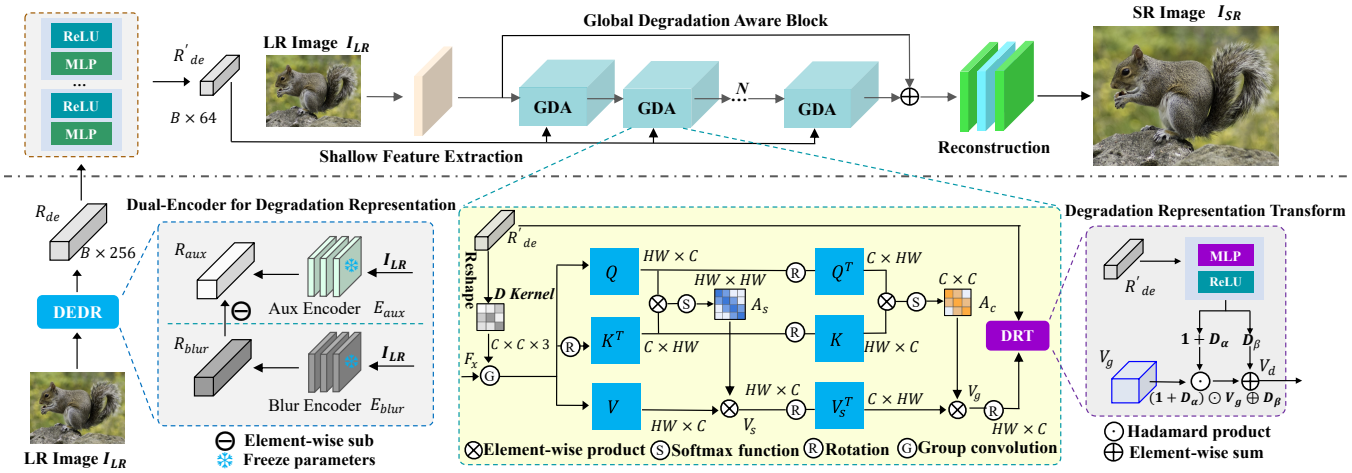

Figure 2: The overview of our proposed DRAT network, which mainly consists of a dual-encoder for degradation representation, global degradation aware blocks and the degradation representation transform. The auxiliary encoder E_{aux} is trained on HR patches. Once trained, E_{aux} is frozen and it is capable of extracting the content representation from the LR image.

3.1 Degradation Representation Aware Transform Network

In our architecture, an LR image I_{LR} is input into the DEDR module, and the DEDR filters out the influence of image content on degradation information in LR image. Then we compress degradation representation in order to efficiently utilize it and reduce the parameters of model. By inputting the degradation representation into the DRT module to attain a set of degradation transform parameters, we transform the features to make them degradation-aware. Furthermore, through the degradation transformation of DRT and the efficient propagation of GDA, global features with degradation awareness are obtained. Finally, through reconstruction and upsample, the super-resolved image I_{SR} is yielded by our network. We optimize the blind SR network as follows:

$$\mathcal{L} = \|I_{SR} - I_{HR}\|_1 + \alpha \|FFT(I_{SR}) - FFT(I_{HR})\|_1, \quad (2)$$

where I_{SR} and I_{HR} are SR and HR images, respectively. The $\|\|_1$ represents L_1 loss, FFT denotes Fast Fourier Transform, and α is a weight coefficient.

3.2 Dual-Encoder for Degradation Representation

Generally, the random noise and blur are generated by camera shake and sensor limitation. Luo et al. (2022b) also indicate that in some cases, the degradation has no relation to the image content. In conventional works, supervised and unsupervised methods directly estimate or extract degradation information from LR images. The extracted degradation representation contains image content which may influence the degradation, and makes it inaccurate.

Our DEDR is designed for extracting more accurate degradation representation, then we can restore SR images with higher quality. Since SKE methods (Gu et al. 2019; Luo et al. 2021, 2022a) may suffer from inaccuracy of degradation estimation, we consider using the unsupervised strategy to learn the degradation distribution. To make degradation representation more discriminative, we train auxiliary encoder E_{aux} on HR images to extract the image content representation and blur encoder E_{blur} on LR image to extract mixed representation (content and degradation representation). By filtering out the content representation from mixed representation, we can obtain more accurate degradation representation. We adopt contrastive learning to train the dual-encoder, and the loss function $\mathcal{L}_c$ is:

$$\mathcal{L}_c = -log\Big[\frac{\exp(I_q \cdot I_{k^+}/\tau)}{\sum_{i=0}^{K} \exp(I_q \cdot I_{k_i}/\tau)}\Big], \quad (3)$$

where I_{k+} is a positive sample of the I_q, all samples except I_{k+} are negative samples. K is the number of samples and τ is the temperature coefficient. The E_{aux} and E_{blur} are trained respectively. For training E_{aux}, the positive sample is a transform of the HR image, and the negative samples are other HR images with different content. While for E_{blur}, the positive sample is one LR image with same degradation, and negative samples are the LR images with different degradations. Once E_{aux} and E_{blur} are trained, we freeze their parameters and use them to extract the content representation and mixed representation from LR input.

Taking a blind SR task for example, the degradation representation R_{de} is given below:

$$R_{aux} = E_{aux}(I_{LR}), \quad (4)$$

$$R_{blur} = E_{blur}(I_{LR}), \quad (5)$$

$$R_{de} = R_{blur} - R_{aux}, \quad (6)$$

where I_{LR} is the LR input, R_{aux} and R_{blur} are the content and mixed representation from E_{blur} and E_{aux}, respectively. For two LR images with the same degradation but different content, DEDR can extract the same degradation representation R_{de} according Eq.(6) while R_{aux} differs. Compared with other methods, such as (Wang et al. 2021a; Zhou et al. 2022; Xia et al. 2023), the dual-encoder extracts degradation representation guided by auxiliary encoder, exhibiting more discriminative degradation representation across complex degradation LR images.

3.3 Degradation Representation Transform

The proposed degradation representation transform blocks can transform features and make them degradation-aware. We devise DRT to transform features into corresponding degradation space, and use residuals and skip connections to better utilize the degradation representation to the blind SR restoration process. We employ Multi-Layer Perceptron (MLP) and ReLU operation to compress the degradation representation for more effectively utilization. We convolve the degradation representation extracted by DEDR module to generate a pair of degradation transformation parameters D_α and D_β, and use them to transform degradatation information on intermediate feature maps. This procedure can be described as:

$$D_\alpha, D_\beta = C(MLP(R_{de})), \tag{7}$$

where R_{de} denotes the degradation representation, MLP is the fully connected layers, and C represents the chunk operation. The degradation transformation parameters D_α and D_β are obtained by MLP and C operation, then we adopt the learning style of SFT to transform the input feature F_s.

$$F_{de} = (1 + D_\alpha) \cdot F_s + D_\beta, \tag{8}$$

where F_{de} is the feature obtained by DRT module. The DRT can perform degradation transform for features. The initial dimension of the degradation representation is $R^{B\times 256}$, in order to reduce the computational effort of the model, we compress it to $R^{B\times 64}$ to transform features.

Unlike our DRT, previous methods do not fully utilize the degradation information to restore LR, even destroying the internal structure of degradation presentation. In (Gu et al. 2019), a set of affine variation parameters is predicted by concatenating the stretched kernel estimation vectors with the feature maps to transform the features, which leads to the estimation of affine variation parameters have a relation with the image content. The problem also exists in (Wang et al. 2021a), which brings inaccurate degradation representation in the subsequent reconstruction of the blind SR process.

3.4 Global Degradation Aware Block

Unlike previous methods in which the degradation information is only used for transmission in channel domain, we consider that features should be globally transmitted either in spatial domain. Therefore, we present a global degradation aware block to transmit degradation representation in both spatial and channel-wise domains. Because the information on the spatial features and the channel features are related, It is efficient to transform degradation information both spatial and channel-wise.

Specifically, suppose $F_x \in R^{H\times W\times C}$ is the output from shallow feature extraction module, then through a series of GDA blocks, we can get the refined degradation-aware feature to reconstruct blind SR. First, we embed the F_x to query, key and value matrices Q, K and $V \in R^{HW\times C}$ through the linear projection, given by:

$$Q = F_x \cdot \mathcal{W}_q,\ K = F_x \cdot \mathcal{W}_k,\ V = F_x \cdot \mathcal{W}_v, \tag{9}$$

where $\mathcal{W}_q$, $\mathcal{W}_k$, and $\mathcal{W}_v$ denote the linear projection matrices for the query, key and value, respectively. Then we obtain the spatial attention map $A_s \in R^{HW\times HW}$ by Q and K^T, and multiply it with V to obtain the feature map V_s adjusted in spatial attention defined by:

$$V_s = A_s \cdot V,\ A_s = Softmax(Q \cdot K^T), \tag{10}$$

Moreover, we rotate Q and V_s to attain Q^T and $V_s^T \in R^{C\times HW}$ given by:

$$Q^T = R(Q),\ V_s^T = R(V_s), \tag{11}$$

where $R(.)$ denotes the rotation operation. We obtain the channel attention map $A_c \in R^{C\times C}$ through Q^T and K. Note that, V_s^T contains the information from spatial domain, then we multiply the channel attention map A_c with V_s^T to yield the global feature V_g. The process is described as:

$$V_g = A_c \cdot V_s^T,\ A_c = Softmax(Q^T \cdot K), \tag{12}$$

Then we use DRT to transform the feature V_g, so that the feature has the global degradation awareness. The process for the final global feature is obtained by Eq.(8) as follows:

$$V_d = (1 + D_\alpha) \cdot V_g + D_\beta, \tag{13}$$

where V_d is the output of GDA blocks. Compared to the conventional SKE and UDP methods, the features obtained from our DRT module are more sensitive to the degradation information and can be better optimized. This is also verified by experimental results in Section 4.

4 Experimental Results

4.1 Datasets and Experimental Setting

In the experiments, we use 3450 HR images from DIV2K (Agustsson and Timofte 2017) and Flickr2K (Timofte et al. 2017) as training datasets. For testing, we synthesize degradation data using isotropic and anisotropic Gaussian kernels with five datasets: Set5 (Bevilacqua et al. 2012), Set14 (Zeyde, Elad, and Protter 2012), BSD100 (Martin et al. 2001), Urban100 (Huang, Singh, and Ahuja 2015), and Manga109 (Matsui et al. 2017). We also adopt three realwolrd datasets: RealSR (Cai et al. 2019), DRealSR (Wei et al. 2020) and ImageNet-Test (Yue, Wang, and Loy 2024). All the methods are evaluated in terms of PSNR and SSIM (Wang et al. 2004) exclusively in YCbCr color space.

Implementation details. For our dual-encoder training, we utilize MoCo (Chen, Xie, and He 2021) with the parameter settings as: $\tau = 0.07$, queue number $N = 12228$ and the degradation representation channel is set to 256. For blind SR training, we employ 10 GDA blocks to construct DRAT network, and each GDA block with input channel number 64. The LR patch size is uniformly set to 64 $\times$ 64 for each scale factor and batchsize is 16. We adopt the Adam optimizer (Kingma and Ba 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ to optimize our model. The initial learning rate is set to 2×10^{-4}, and it decays continuously during training, being reset to 2×10^{-4} every 2.5×10^5 iterations. The weight coefficient α of FFT loss is set to 0.05.

Method	Scale	PSNR(dB)/SSIM on Synthetic Degradation Datasets				
		Set5	Set14	BSD100	Urban100	M109
Bicubic	×3	26.96 / 0.7848	24.54 / 0.6800	24.91 / 0.6584	21.97 / 0.6510	23.78 / 0.7783
IKC (Gu et al. 2019)		33.35 / 0.9177	29.66 / 0.8281	28.69 / 0.7929	27.55 / 0.8342	32.47 / 0.9333
DANv1 (Huang et al. 2020)		34.19 / 0.9225	30.16 / 0.8331	29.00 / 0.7963	27.81 / 0.8405	33.30 / 0.9404
DANv2 (Luo et al. 2021)		34.31 / 0.9238	30.29 / 0.8356	29.09 / 0.7992	28.04 / 0.8459	33.48 / 0.9426
DCLS (Luo et al. 2022a)		34.38 / 0.9243	30.34 / 0.8368	29.13 / 0.8003	28.22 / 0.8503	33.66 / 0.9435
DSSR (Li et al. 2022)		34.22 / 0.9206	30.17 / 0.8318	29.03 / 0.7966	27.81 / 0.8406	33.28 / 0.9407
CDCN (Wu et al. 2022)		34.29 / 0.9238	30.28 / 0.8356	29.10 / 0.7996	28.11 / 0.8471	33.47 / 0.9424
MRDA (Xia et al. 2023)		33.99 / 0.9229	30.23 / 0.8348	29.02 / 0.7991	27.92 / 0.8434	33.37 / 0.9412
DRAT (Ours)		**34.38 / 0.9245**	**30.39 / 0.8370**	**29.16 / 0.8009**	**28.33 / 0.8506**	**33.82 / 0.9447**
Bicubic	×4	24.39 / 0.6869	24.30 / 0.6325	23.38 / 0.5671	20.34 / 0.5478	21.30 / 0.6834
IKC (Gu et al. 2019)		31.67 / 0.8829	28.31 / 0.7643	27.34 / 0.7192	25.33 / 0.7504	28.91 / 0.8782
DANv1 (Huang et al. 2020)		31.89 / 0.8864	28.42 / 0.7687	27.51 / 0.7248	25.86 / 0.7721	30.50 / 0.9037
DANv2 (Luo et al. 2021)		32.00 / 0.8885	28.50 / 0.7715	27.56 / 0.7277	25.94 / 0.7748	30.45 / 0.9037
DCLS (Luo et al. 2022a)		32.12 / 0.8890	28.54 / 0.7728	27.60 / 0.7285	26.15 / 0.7809	30.86 / 0.9086
DSSR (Li et al. 2022)		31.79 / 0.8870	28.43 / 0.7679	27.49 / 0.7229	25.73 / 0.7668	30.44 / 0.9023
CDCN (Wu et al. 2022)		32.04 / 0.8883	28.47 / 0.7695	27.52 / 0.7257	25.92 / 0.7731	30.55 / 0.9036
MRDA (Xia et al. 2023)		31.98 / 0.8872	28.42 / 0.7671	27.55 / 0.7254	25.90 / 0.7734	30.51 / 0.9088
DRAT (Ours)		**32.15 / 0.8916**	**28.72 / 0.7777**	**27.66 / 0.7326**	**26.36 / 0.7880**	**31.03 / 0.9123**
Bicubic	×6	21.87 / 0.5760	20.83 / 0.4950	21.92 / 0.4946	18.81 / 0.4603	19.17 / 0.5990
IKC (Gu et al. 2019)		27.88 / 0.7961	25.74 / 0.6666	25.40 / 0.6259	22.82 / 0.6310	24.76 / 0.7757
DANv1 (Huang et al. 2020)		28.00 / 0.8011	25.43 / 0.6656	25.45 / 0.6241	22.95 / 0.6362	24.90 / 0.7777
DANv2 (Luo et al. 2021)		27.96 / 0.7998	25.68 / 0.6644	25.43 / 0.6369	22.98 / 0.6399	24.93 / 0.7786
DCLS (Luo et al. 2022a)		28.03 / 0.7979	25.74 / 0.6682	25.46 / 0.6297	23.17 / 0.6476	25.54 / 0.7941
DSSR (Li et al. 2022)		27.99 / 0.8010	25.72 / 0.6664	25.44 / 0.6273	22.97 / 0.6482	25.32 / 0.7920
CDCN (Wu et al. 2022)		28.02 / 0.7993	25.78 / 0.6688	25.46 / 0.6290	23.05 / 0.6417	25.42 / 0.7620
MRDA (Xia et al. 2023)		28.01 / 0.7989	25.74 / 0.6691	25.44 / 0.6296	23.07 / 0.6453	25.53 / 0.7983
DRAT (Ours)		**28.09 / 0.8012**	**25.85 / 0.6731**	**25.52 / 0.6343**	**23.29 / 0.6543**	**25.87 / 0.8044**

Table 1: Quantitative comparison on datasets with isotropic Gaussian 8 kernels on scale factors 3, 4 and 6. All metrics are computed on Y channels. Bold indicates the best.

Method	Param	DIV2KRK	
		PSNR	SSIM
Scale ×2			
Bicubic	-	28.73	0.8040
DANv2 (Luo et al. 2021)	4.71	32.58	0.9048
DCLS (Luo et al. 2022a)	13.66	32.75	0.9094
DASR (Wang et al. 2021a)	5.80	32.24	0.8960
DSSR (Li et al. 2022)	8.90	32.46	0.9018
CDCN (Wu et al. 2022)	11.70	32.73	0.9051
DAN-Real (Luo et al. 2023)	1.95	32.96	0.9114
DRAT (Ours)	9.59	**33.31**	**0.9170**
Scale ×4			
Bicubic	-	25.33	0.6795
DANv2 (Luo et al. 2021)	4.71	28.74	0.7893
DCLS (Luo et al. 2022a)	13.66	28.99	0.7946
DASR (Wang et al. 2021a)	5.80	28.41	0.7813
DSSR (Li et al. 2022)	8.90	28.78	0.7905
CDCN (Wu et al. 2022)	11.70	28.92	0.7929
DAN-Real (Luo et al. 2023)	1.95	28.90	0.7964
DRAT (Ours)	9.59	**29.21**	**0.8017**

Table 2: Quantitative comparisons with anisotropic Gaussian kernels on DIV2KRK for scale factors 2 and 4. Bold indicates the best.

4.2 Comparison with State-of-the-art Methods

Evaluation with isotropic Gaussian kernels. We evaluate our method on the synthetic datasets by Gaussian 8 kernels following IKC (Gu et al. 2019). We compare our method with state-of-the-art (SOTA) SKE methods: IKC (Gu et al. 2019), DANv1 (Huang et al. 2020), DANv2 (Luo et al. 2021), and DCLS (Luo et al. 2022a). Additionally, we compare DRAT with UDP methods, including DSSR (Li et al. 2022), CDCN (Wu et al. 2022) and MRDA (Xia et al. 2023). For fair comparison, all experiments use official implementations and weights for testing. The quantitative results are shown in **Table** 1. It can be observed that our method outperforms DCLS, and surpasses the latest SOTA UDP method MRDA. In **Figure** 3, our DRAT achieves the best visual results and highest numerical values in terms of PSNR/SSIM metrics. SKE methods restore LR images by embedding the estimated blur kernels into blind SR network, but inaccurate degradation estimation introduces artifacts in the super-resolved images in **Figure** 3 (a). In addition, the UDP methods, such as DSSR, CDCN, and MRDA, fail to consider the influence of the image content in LR image on degradation representation, which results in the extraction of inaccurate degradation representations for reconstruction and insufficient utilization of degradation representations and thus per-

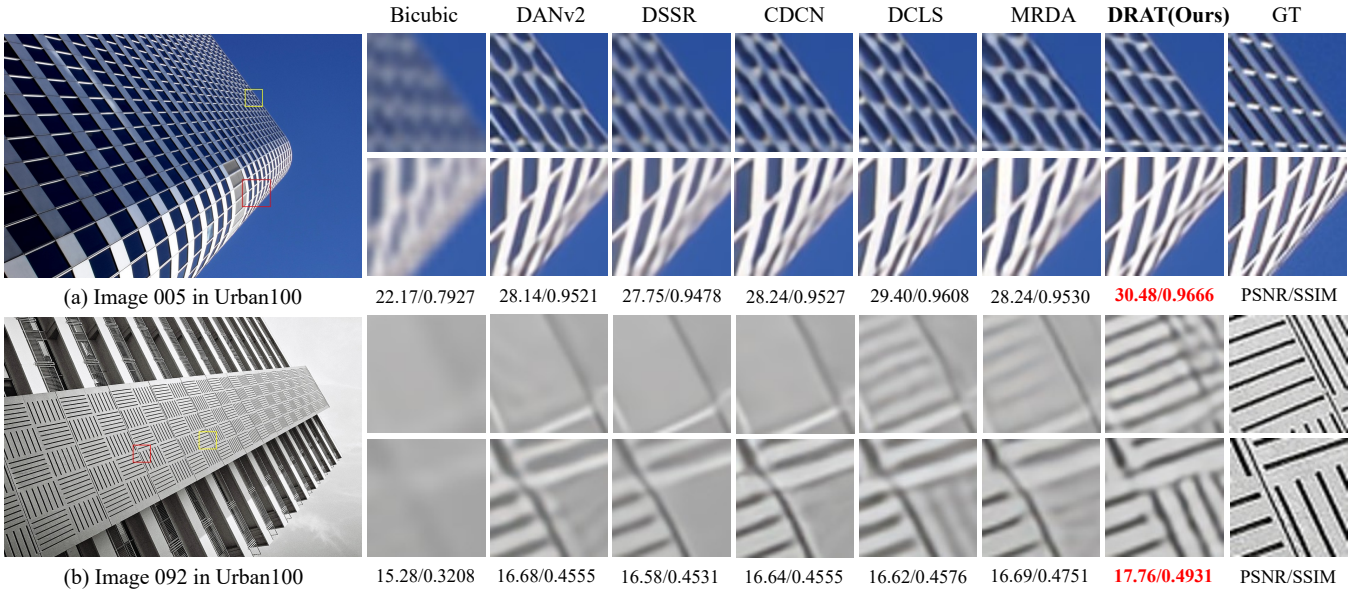

Figure 3: Visual results of Img 005 and Img 092 from Urban100 dataset with isotropic kernels. (a) Img 005 $\times$ 4 with 1.8 kernel width; (b) Img 092 $\times$ 6 with 4.2 kernel width.

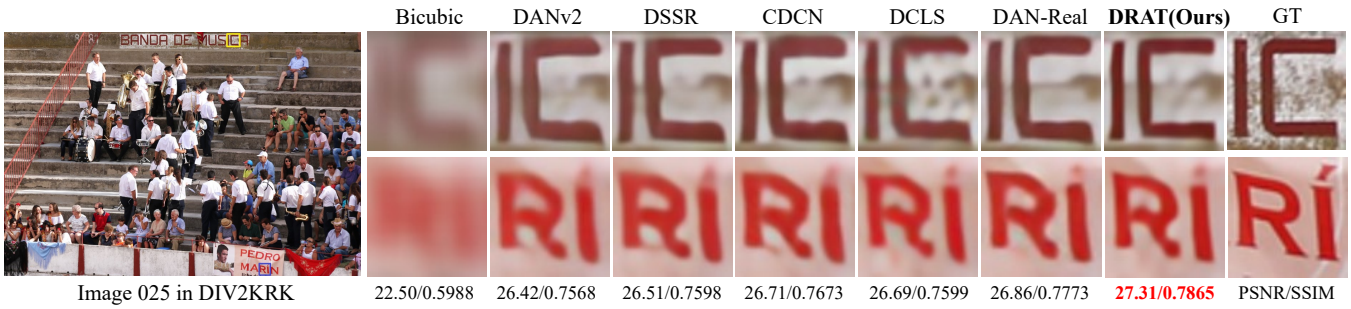

Figure 4: Visual results of Img 025 from DIV2KRK dataset for $\times$ 4 random anisotropic kernel super-resolution.

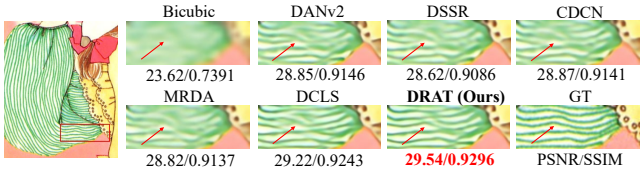

Figure 5: Visual results of Img YumeirooCooking with scale factor 4 and kernel width 2.0 on Manga109 dataset.

form worse than our method. **Figure** 5 is the visual results on Manga109 dataset. DRAT restores the sharpest line texture and it increases the PSNR by almost 0.9 dB compared with DSSR, which has similar number of parameters.

Evaluation with anisotropic Gaussian kernels. In **Table 2**, we evaluate our method with anisotropic Gaussian kernels on DIV2KRK (Bell-Kligler, Shocher, and Irani 2019) dataset, which is more complex and challenging. The visual results are demonstrated in **Figure 4**. We first compare the SKE methods, Methods such as DANv2, DCLS and DAN-Real often struggle to recover sharp lines and may exhibit artifacts. This issue arises from the inaccurate estimation of complex blur kernels, which affects the methods' ability to restore fine details in the resulting images.

In addition, some SOTA UDP methods, such as DASR, DSSR, and CDCN, restore poorly results. It is probably because they extract inaccurate degradation representations and do not make good use of these representations to adjust features, which is influenced by the image content in the LR image. As shown in **Figure** 4, our DRAT yields clearer textures. DRAT learns more detailed degradation representations by DEDR and propagates degradation globally into features by DRT, achieving the best PSNR/SSIM. In the anisotropic Gaussian kernel setting for scale factor 2, DRAT outperforms SOTA SKE method DAN-Real by 0.35 dB and UDP method CDCN by 0.58 dB.

Evaluation on large scale factor and unseen degradations. As can be seen in **Figure** 3, we compare the performance of the models on $\times$6, where the kernel width ranges from 3.8 to 5.2. It is challenging for the model to handle large scale factor and unseen complex degradations. In **Figure** 3 (b), only DRAT recovers texture details similar to GT,

Method	Param	PSNR(dB)/SSIM on Real-World Degradation Datasets			
		RealSR-Nikon	RealSR-Canon	DRealSR	ImageNet-Test
HAT (Chen et al. 2023)	20.80	27.47 / 0.7676	27.80 / 0.7902	30.61 / 0.8610	23.86 / 0.6447
CDCN (Wu et al. 2022)	11.70	27.63 / 0.7804	28.03 / 0.7981	30.58 / 0.8606	24.18 / 0.6558
DCLS (Luo et al. 2022a)	13.66	27.63 / 0.7803	28.03 / 0.7983	30.61 / 0.8612	24.62 / 0.6675
DAN-Real (Luo et al. 2023)	1.95	26.75 / 0.7879	27.62 / 0.8198	29.06 / 0.8550	25.62 / **0.7062**
ResShift (Yue, Wang, and Loy 2024)	118.59	25.45 / 0.7213	26.13 / 0.7665	26.91 / 0.7426	25.01 / 0.6806
DRAT(Ours)	9.59	**27.79 / 0.7960**	**28.34 / 0.8206**	**31.00 / 0.8737**	**25.82** / 0.7035

Table 3: Quantitative comparisons on RealSR, DRealSR, and ImageNet-Test datasets for scale factor 4. Bold indicates the best. Param (M) indicates the parametric quantity in Millions.

Model	GDA	DEDR	DRT	DIV2KRK ×2	
				PSNR	SSIM
Model1	✗	✔	✔	32.69	0.9047
Model2	✔	✗	✔	32.82	0.9081
Model3	✔	✔	✗	32.88	0.9085
DRAT	✔	✔	✔	**33.31**	**0.9170**

Table 4: Ablation studies of the proposed DRAT on DIV2KRK dataset for ×2 blind super-resolution.

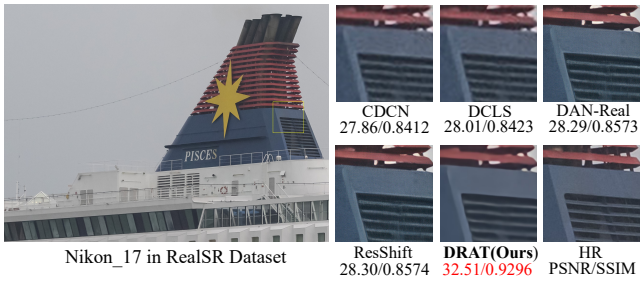

Figure 6: Comparison on a real-world image from RealSR dataset for ×4 blind super-resolution.

while the other methods suffer from severe artifacts or even do not recover the correct texture.

Evaluation on the real-world degradations. For real-world test datasets, the quantitative results are shown in **Table** 3. Our DRAT consistently achieves the highest metric results across those three datasets. Notably, on the DRealSR dataset, even when encountering more complex real-world degradations, it continues to deliver the best PSNR/SSIM scores of 31.00/0.8737. The numbers of parameters of DCLS and ResShift are 13.66M and 118.59M, while our DRAT model reduces the number of parameters by 42% and 92%, respectively. But we outperform the PSNR metric on the ImageNet-Test dataset by 1.20dB and 0.81dB. It indicates that our DRAT gains the highest PSNR while consuming relative small number of model parameters, demonstrating the performance of DRAT on complex real-world degradation. We test DRAT on the real-world LR images from RealSR dataset (Cai et al. 2019) for ×4 super-resolution in **Figure** 6. Compared with other SOTA methods, such as DAN-Real and ResShift, our DRAT demonstrates clear texture details. This also indicates the effectiveness of DRAT on the real-world degradations.

4.3 Ablation Study and Discussion

We conduct ablation studies to validate the effectiveness of vital components in our method: DEDR, DRT, and GDA modules. The quantitative results on DIV2KRK dataset for ×2 super-resolution are shown in **Table** 4. It can be seen Model1 without GDA module, the PSNR is 32.69, and it falls 0.62 dB compared with our final model DRAT. As the main component of the blind SR, GDA plays an important role in generating robust features to restore LR image, indicating the efficacy of GDA. The GDA blocks integrate information globally, by introducing GDA, features are enhanced through the modulation of degradation representation, propagating degradation globally. Moreover, in Model2 it can be observed that when the DEDR module is removed from DRAT, namely replaced with a conventional blur encoder. The PSNR/SSIM drops from 33.31/0.9170 to 32.82/0.9081, which indicating the importance of DEDR in accurately extracting degradation representations. While Model3 does not incorporate the DRT module, the PSNR/SSIM is 32.88/0.9085. It validates the effectiveness of the DRT module, which enables the features to be aware of complex degradation and enhances their ability to handle various degradation types more effectively.

5 Conclusion

In this study, we explore methods for extracting more accurate degradation representations for blind SR tasks. To tackle these challenges, we propose a DRAT network. Initially, we present a DEDR module to learn the degradation representations of LR images unsupervised, effectively filtering out the influence of image content on degradation representation. Subsequently, we propose DRT that converts degradation information into features, making the features with degradation awareness capabilities. We further develop GDA blocks to propagate degradation information in both spatial and channel domains and efficiently leverage the extracted degradation representation to guide blind SR. Extensive experiments on various blur kernels and noise levels demonstrate that the proposed method DRAT outperforms both state-of-the-art SKE and UDP methods for blind SR. In the future, we will focus on more complex scenarios and extend DRAT to real-world video super-resolution.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62276182, 62376206, and 62476196), and Peng Cheng Lab Program (No. PCL2023A08).

References

Agustsson, E.; and Timofte, R. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 126–135.

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

Bevilacqua, M.; Roumy, A.; Guillemot, C.; and Alberi-Morel, M. L. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; and Zhang, L. 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.

Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22367–22377.

Chen, X.; Xie, S.; and He, K. 2021. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9640–9649.

Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; and Zhang, L. 2019. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11065–11074.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, 184–199. Springer.

Gu, J.; Lu, H.; Zuo, W.; and Dong, C. 2019. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1604–1613.

Haris, M.; Shakhnarovich, G.; and Ukita, N. 2018. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1664–1673.

Huang, J.-B.; Singh, A.; and Ahuja, N. 2015. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5197–5206.

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

Kim, J.; Lee, J. K.; and Lee, K. M. 2016a. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1646–1654.

Kim, J.; Lee, J. K.; and Lee, K. M. 2016b. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1637–1645.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lai, W.-S.; Huang, J.-B.; Ahuja, N.; and Yang, M.-H. 2017. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 624–632.

Li, F.; Wu, Y.; Bai, H.; Lin, W.; Cong, R.; Zhang, C.; and Zhao, Y. 2022. Learning Detail-Structure Alternative Optimization for Blind Super-Resolution. *IEEE Transactions on Multimedia*, 1–1.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021a. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1833–1844.

Liang, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021b. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4096–4105.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 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.

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

Luo, Z.; Huang, H.; Yu, L.; Li, Y.; Fan, H.; and Liu, S. 2022a. Deep constrained least squares for blind image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17642–17652.

Luo, Z.; Huang, Y.; ; Li, S.; Wang, L.; and Tan, T. 2022b. Learning the Degradation Distribution for Blind Image Super-Resolution. In *CVPR*.

Luo, Z.; Huang, Y.; Li, S.; Wang, L.; and Tan, T. 2023. End-to-end alternating optimization for real-world blind super resolution. *International Journal of Computer Vision*, 131(12): 3152–3169.

Luo, Z.; Huang, Y.; Shang, L.; Wang, L.; and Tan, T. 2021. End-to-end Alternating Optimization for Blind Super Resolution. *Cornell University - arXiv,Cornell University - arXiv*.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, 416–423. IEEE.

Matsui, Y.; Ito, K.; Aramaki, Y.; Fujimoto, A.; Ogawa, T.; Yamasaki, T.; and Aizawa, K. 2017. Sketch-based manga retrieval using manga109 dataset. *Multimedia tools and applications*, 76: 21811–21838.

Tao, G.; Ji, X.; Wang, W.; Chen, S.; Lin, C.; Cao, Y.; Lu, T.; Luo, D.; and Tai, Y. 2021. Spectrum-to-kernel translation for accurate blind image super-resolution. *Advances in neural information processing systems*, 34: 22643–22654.

Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; and Zhang, L. 2017. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 114–125.

Wang, H.; Chen, X.; Ni, B.; Liu, Y.; and Liu, J. 2023. Omni aggregation networks for lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22378–22387.

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

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021b. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1905–1914.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wei, P.; Xie, Z.; Lu, H.; Zhan, Z.; Ye, Q.; Zuo, W.; and Lin, L. 2020. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, 101–117. Springer.

Wu, Y.; Li, F.; Bai, H.; Lin, W.; Cong, R.; and Zhao, Y. 2022. Bridging Component Learning with Degradation Modelling for Blind Image Super-Resolution. *IEEE Transactions on Multimedia*, 1–16.

Xia, B.; Hang, Y.; Tian, Y.; Yang, W.; Liao, Q.; and Zhou, J. 2022. Efficient non-local contrastive attention for image super-resolution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 2759–2767.

Xia, B.; Tian, Y.; Zhang, Y.; Hang, Y.; Yang, W.; and Liao, Q. 2023. Meta-learning based degradation representation for blind super-resolution. *IEEE Transactions on Image Processing*.

Yue, Z.; Wang, J.; and Loy, C. C. 2024. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36.

Zeyde, R.; Elad, M.; and Protter, M. 2012. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, 711–730. Springer.

Zhang, K.; Liang, J.; Van Gool, L.; and Timofte, R. 2021. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. In *IEEE International Conference on Computer Vision*, 4791–4800.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3262–3271.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018a. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, 286–301.

Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. 2018b. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2472–2481.

Zhou, Y.; Lin, C.; Luo, D.; Liu, Y.; Tai, Y.; Wang, C.; and Chen, M. 2022. Joint learning content and degradation aware feature for blind super-resolution. In *Proceedings of the 30th ACM international conference on multimedia*, 2606–2616.