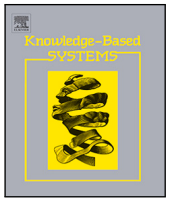
Knowledge-Based Systems 297 (2024) 111973

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Efficient blind super-resolution imaging via adaptive degradation-aware estimation

Haoran Yang [a], Qilei Li [b], Bin Meng [c], Gwanggil Jeon [d], Kai Liu [e], Xiaomin Yang [a,*]

[a] College of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, 610065, China
[b] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, United Kingdom
[c] School of Cyber Science and Engineering, Sichuan University, Chengdu, Sichuan, 610207, China
[d] Department of Embedded Systems Engineering, Incheon, 22012, Republic of Korea
[e] College of Electrical Engineering, Sichuan University, Chengdu, Sichuan, 610065, China

ARTICLE INFO

Keywords:
Blind super-resolution imaging
Degradation estimation
Efficient super-resolution imaging
Non-local feature similarity learning
Feature representation

ABSTRACT

Single image super-resolution (SISR) has achieved prominent success based on deep learning. However, most SISR methods based on the specific degradation pattern, *e.g.,* bicubic interpolation with a bulky network structure, are unsuitable for real-world images reconstructed on edge devices. To mitigate the above limitations, unlike the majority of blind super-resolution methods, which consist of a separate complex degradation kernel projection network and a SR reconstruction network, we propose an end-to-end lightweight multi-degradation oriented network that not only can estimate degradation kernel but also reconstructs HR image fast and accurately. The core components of the designed network adopt a lightweight SR model with region non-local feature similarity learning, *i.e.,* Region Non-Local Feature Block (RNLFB), to establish region-wise global feature correlation. Concretely, more and fewer RNLFBs are used to learn more sophisticated feature representation and relatively simple degradation representation. Then, a degradation kernel projector is equipped to fulfill adaptive degradation-aware estimation from the degradation representation. With the help of the encoded degradation kernel, the reconstruction model can learn to restore a high-resolution image from the feature representation. By doing so, the degradation kernel estimation task and the HR image reconstruction task can be accomplished without designing two complex networks separately, which guides more stable and effective network training under low computational resource costs. Extensive experiments on synthetic and real-world datasets demonstrate that our method can fulfill precise degradation kernel prediction and HR image reconstruction compared with other state-of-the-art SR methods with lower model complexity.

1. Introduction

The target of Single image super-resolution (SISR) technology is reconstructing a high-resolution (HR) image from its low-resolution (LR) observation, which has been broadly applied in multifarious fields, such as medical image processing [1,2], remote sensing [3–5], and as the pre-tasks of object detection or recognition [6] for security monitoring. In the past decade, due to the promising success of deep learning, various SISR methods have achieved a favorable SR performance based on convolutional neural networks (CNNs) or transformers. In general, the degradation process can be expressed by the following formula

$$y = (x \otimes k)\downarrow, \tag{1}$$

where the LR image y is degraded by the following steps: (1). Using convocational operation between degradation kernel k and HR image x to obtain the blur version of x, *i.e.,* $x \otimes k$. (2). Using a down-sampled operation on the blur version of x to obtain the LR image y. To simplify the degradation process, many researchers choose to ignore k in Eq. (1) and only use bicubic interpretation to simulate the degradation process, *i.e.,* $y = x\downarrow$. Based on this assumption, Dong et al. [7], as a pioneer, propose a CNN-based SISR method, whose datasets are synthesized by bicubic operation degraded on HR images. VDSR [8] and EDSR [9] are introduced in succession, constantly refreshing SR performance. Yang et al. [10] explore the contextual and semantic information by building a hierarchical accumulation network with grid attention for more accurate SR images. Xia et al. [11] introduce a novel non-local attention mechanism with contrastive learning to distinguish feature relationships. Chen et al. [12] propose a hybrid attention model to solve the issue that Transformer-based methods cannot fully exploit the spatial range of input information. The aforementioned SR methods successfully explored the possibilities of deep learning deployment in

* Corresponding author.
E-mail address: arielyang@scu.edu.cn (X. Yang).

https://doi.org/10.1016/j.knosys.2024.111973
Received 30 January 2024; Received in revised form 6 May 2024; Accepted 17 May 2024
Available online 20 May 2024

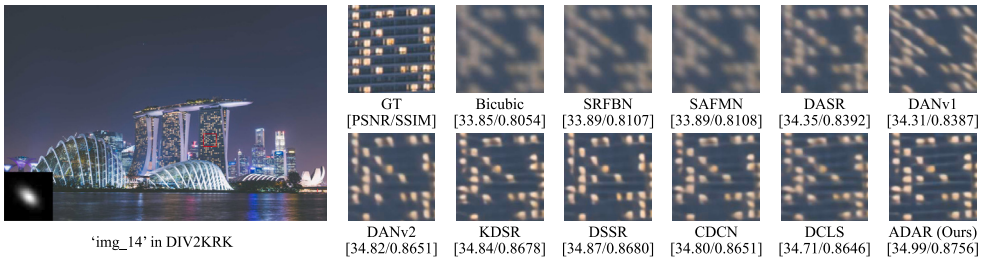


Fig. 1. Subjective visual comparison on ×4 image SR in 'img_14' of DIV2KRK. Since the degradation mode of the LR image, as shown in the bottom left corner of the Figure, deviates from the bicubic degradation mode assume degradation by the Non-Blind SR methods *i.e.,* Bicubic, SRFBN, and SAFMN, causing the SR performance collapse, while Blind SR methods *i.e.,* the remaining SR methods, achieve a favorable SR results.

the field of SR. However, these methods suffer from limitations such as difficulty in deployment on edge devices and single degradation pattern of the input LR image. To alleviate the first limitation and reduce the computational costs and inference time of the SR model, Liu et al. [13] utilize group convolutional kernel with a global-skip cascading mechanism for saving network parameters. Kong et al. [14] propose a novel residual local feature block to reduce the inference time while maintaining the model capacity of extracting features. Wang et al. [15] design a lightweight Transformer structure where the self-attention mechanism capture feature relationship at the local, meso, and global levels, respectively. The above lightweight SR methods typically are tailored on LR images with the single degradation pattern, as shown in Fig. 1, once the degradation pattern in real-world LR image is inconsistent with the hypothesis, *e.g.,* bicubic-interpretation degradation, the SR performance of the above models [16,17] decreases disastrously. Therefore, to alleviate the second limitation, blind Super-Resolution Imaging aims to reconstruct a high-resolution image from its low-resolution observation that degraded from multiple degradations according to Eq. (1). How to correctly estimate the degradation kernel k is of paramount importance in blind SR Imaging. IKC [18] uses a predictor network to predict the degradation kernel and a corrector network to calibrate the predicted kernel. DASR [19] and IDMBSR [20] learn a degradation representation space by degradation estimator to distinguish various degradation kernels with the help of contrastive learning. Zhou et al. [21] design a patch-based encoder and codebook-based space compression module to complete degradation kernel estimation. Although these blind SR methods accurately estimate the degradation kernel that assists the SR network to fulfill better SR reconstruction under multiple degradations, it is necessary to design two or three networks, *e.g.,* degradation kernel predictor network, SR reconstruction network, separately, which significantly increase the consumption of computing recourse, as well as the collaborative training of these networks is sophisticated. In a nutshell, the above lightweight SR methods and blind SR methods have partially alleviated the adverse effects of bulky networks and the single degradation pattern of the input LR image on SR model performance. However, simultaneously achieving both lightweight SR and blind SR increases the difficulty of training the SR model since it requires substantial computational resources to estimate and leverage the degradation information. Hence, researching an efficient blind super-resolution algorithm is warranted.

In this work, to realize an economically blind super-resolution imaging, we introduce the Region Non-Local Feature Block (RNLFB) for non-local feature similarity learning with the traits of lightweight and practical based on the lightweight SR model, Residual Local Feature Block (RLFB). Different from RLFB, to cope with more sophisticated feature extraction in blind SR imaging, we replaced the Enhanced Spatial Attention in RLFB with the introduced Region-Kernel Attention (RKA) Model for elevating the model representation ability with low computational consumption. In the RKA model, the input feature maps are firstly fed into the proposed Region Non-Local Attention (RNLA) with two loops to indirectly establish the region-wise feature correlation in loop 2 based on the generated pixel-to-region feature correlation in loop 1, in which RNLA can achieve a comparable SR performance compared with traditional Non-Local Attention along with much lower computational complexity than Non-Local Attention. Then, the Large Kernel Attention Block [22] is complemented after two RNLAs in the RKA model to further improve the feature extraction ability in both spatial and channel dimensions. The RKA model can provide the facility basis for simultaneously accomplishing kernel estimation and image reconstruction tasks. Concretely, to predict the degradation kernel, an adaptive degradation-aware estimation model is equipped by a few RNLFBs to learn degradation representation, from which a degradation kernel projector precisely predicts the degradation kernel. Subsequently, to reconstruct the HR image, more RNLFBs are required to extract and learn feature representation with the help of the predicted degradation kernel since feature learning is more sophisticated than degradation kernel learning. In this way, the kernel prediction task and the HR image reconstruction task can be accomplished without designing two complex networks separately, which guides more stable and effective network training under low computational resource costs. Extensive experiments on synthetic and real-world datasets demonstrate that our method can fulfill precise degradation kernel prediction and HR image reconstruction compared with state-of-the-art methods with lower model complexity. The primary contributions of this work are summarized as follows:

- We introduce an end-to-end efficient multi-degradation oriented network, which not only can project degradation kernel but also reconstructs SR image efficiently. Extensive experiments demonstrate the favorable SR performance and efficiency of our method on both synthetic and real-world datasets.
- We introduce a region non-local similarity learning that is applied in the feature extractor to economically accumulate region-wise global feature correlations based on the pixel-to-region feature correlations, which can help the feature extractor synchronously learn the feature and degradation representation instead of designing a specific kernel prediction network and a reconstruction network.
- We introduce an adaptive degradation-aware estimator, in which the proposed degradation kernel projector achieves fast and accurate prediction of degradation kernel from the degradation representation, providing definite kernel prior knowledge for SR reconstruction.

The following sections are arranged as follows. Section 2 reviews two hot topics, *i.e.,* efficient SR imaging and blind SR imaging in SISR. Section 3 demonstrates the proposed algorithm process. Section 4.1 verifies the performance of the proposed method. Finally, Section 5 concludes our work.

2. Related work

2.1. Efficient super-resolution imaging

The flourishment of deep learning technology has become an effective tool for super-resolution imaging. As a pioneer, Dong et al. [7]

firstly use a convolutional neural network to fulfill the SISR task with favorable SR performance. Since Kim et al. [8] propose a perception that deep and/or wide network structure can improve SR performance, several methods [9,10,23–25] constantly promote SR performance via deep CNNs with cascading mechanisms and/or attention mechanisms. However, deep CNNs are hard to embed on edge devices, *e.g.,* mobile phone or digital camera, so efficient super-resolution imaging is introduced to alleviate this limitation. For example, IDN [26] explores an information distillation mechanism that can distill the redundancy feature channels to save network parameters and the amount of calculation. Information distillation appears in some research [27–29] as well, whose effectiveness can be demonstrated in Efficient SR Imaging. CARN [30] utilizes a cascading mechanism to repeatedly connect previous features to reduce the number of convolutional kernels. LESR-CNN [31] design a lightweight network structure based on local and global cascading mechanisms as well. RLFN [14] presents a residual local feature network with contrastive loss, which can cut down network fragments while maintaining model capacity. MREN [32] refines features using a multi-level spatial-channel attention mechanism, which guides the model to focus on the high-frequency feature details. Since the limited utilization of global feature information and the substantial increase in model parameters and computational complexity caused by simply stacking CNNs, the algorithms based on the non-local feature similarities exploration, have been widely applied in SR tasks. For instance, Zhang et al. [33] integrate an efficient regularization term using a non-local steering kernel regression model into the standard back-projection framework. This optimization effectively leverages the non-local redundancy to assist the SR model in preserving the sharp edges of SR images. Mei et al. [34] introduce a Cross-Scale Non-Local (CS-NL) attention to capture similar patches in non-local regions. Wu et al. [35] achieve more long-range feature dependencies by the proposed Multi-Scale Non-Local Attention Block, in which multi-scale patches within the same features can establish relevant non-local correlations to lessen the redundant computations. Recently, Transformer architecture has made a breakthrough in SR performance in the field of SR due to its capacity to efficiently exploit non-local feature similarities. Liang et al. [36] introduce SwinIR that obtains better SR performance with fewer model parameters compared with previous efficient SR models because Swin Transformer takes advantage of both CNN and transformers, which can capture long-range feature dependency and weight content-based features. More recently, SAFMN [17] optimizes the network parameters by adding the proposed Faster Fourier transform-based frequency loss function on the mean absolute error loss. Furthermore, in [17], a compact convolutional channel mixer is designed to perform local contextual feature mixing. Using the N-Gram context on SwinIR, NGswin [37] expands the receptive field during HR image reconstruction to improve SR performance in terms of Transformer-based efficient SR imaging. MSID [29] exploits multi-scale convolutions with multiple receptive fields for accumulating diverse level features, which is suitable in resource-constrained scenarios. However, the above SR methods are based on the specific degradation assumption, *i.e.,* LR images are degraded by using bicubic interpretation downsampled operation on HR images, which is inconsistent with LR images in real-world scenarios. Hence, it is necessary to consider more degradation patterns in SR modeling, avoiding the SR performance drop due to the input LR image of the SR model under multiple degradations.

2.2. Blind super-resolution imaging

Towards exploring super-resolution imaging technology under multiple degradations, a variety of Blind SR methods have emerged. Zhang et al. [38] point out that CNN-based SR algorithms are on the basis of maximum a posterior (MAP) framework. By analyzing CNN architecture under MAP inference, Zhang et al. model an intrinsic relationship between blind SR algorithms and the MAP principle. In other words, the parameters of model parameters can be estimated in the case where the LR image and the degradation kernel are known. With the known degradation kernel and noise level, SRMD [38] and VBSR [39] stretch them onto the same dimension as the LR image so that LR input with degradation prior can be fed into the SR model in the meantime. Regarding the LR image as the reference image, ZSSR [40] fits the non-linear mapping between the LR image and its corresponding multi-degraded down-sampled image to exploit interior information of the LR image. Gu et al. [18] propose a predictor network that estimates kernel degradation with principal component analysis for improving network generalization capacity. Some blind SR methods [41,42] concentrate on building a large pool that consists of various degradation kernels and noises as the degradation condition to synthesize HR-LR training datasets. Therefore, in the phase of training the SR model, it just requires a sample non-blind reconstruction network under supervised learning with the synthetic datasets. Luo et al. [43] observe that simultaneously optimizing both the predictor and reconstruction networks is difficult to converge. Motivated by this observation, they introduce an end-to-end network, DAN, that takes Estimator and Restorer to fulfill degradation kernel prediction and HR image reconstruction, respectively. UDVD [44] incorporate dynamic convolution to handle numerous cross-image and/or spatial variational degradations flexibly. Instead of estimating an encoded degradation kernel, DASR [19] and IDMBSR [20] utilize contrastive learning that projects LR patches into implicit degradation representations, in which the encoded patches with the same degradation are pulled together while the different ones are pushed away. Then, DASR and IDMBSR reshape the encoded degradation representation as a convolutional kernel to complete depthwise convolution operation on the LR features, which accurately restores the HR image with degradation kernel information. DSSR [45] alternatively exploits detail features and structure features with recursive strategy, allowing multi-level feature components for accurate blind-SR reconstruction. Since the above blind SR methods rely on two or three elaborate network structures to achieve degradation kernel prediction, degradation kernel correction, and HR image reconstruction, or a time-consuming recursive strategy, their network structure invariably suffers from parameters and the amount of calculation redundancy, which are difficult to embed on edge devices. This observation provides us with a direction to research efficient blind SR algorithms for fast and accurate Blind SR Imaging.

3. Adaptive degradation-aware estimation for efficient blind super-resolution imaging

In this Section, we first introduce the feature extractor, which plays a role in learning degradation representation and feature representation at the phases of degradation estimation and SR image reconstruction, respectively. We then introduce the framework of the adaptive degradation-aware estimator in Section 3.2, whose function is to predict degradation kernel from degradation representation. Finally, we introduce the framework of the adaptive degradation-aware restorer, aiming at achieving blind SR reconstruction with the help of the encoded estimated degradation kernel information in Section 3.3.

3.1. Feature extractor for degradation and feature representation learning

Since previous blind methods [18–20,46,47] tend to design two or more networks to complete the tasks of degradation kernel prediction, correction, and SR image reconstruction, which is difficult to cooperative training between networks. Besides, It is hard to embed on edge devices because the intricate kernel prediction network and the SR reconstruction network require to consume a huge of computing resources. To realize efficient blind super-resolution imaging, we introduce a lightweight Region Non-Local Feature Block as the core component in the feature extractor which is simultaneously equipped for degradation and feature representation learning. The former one

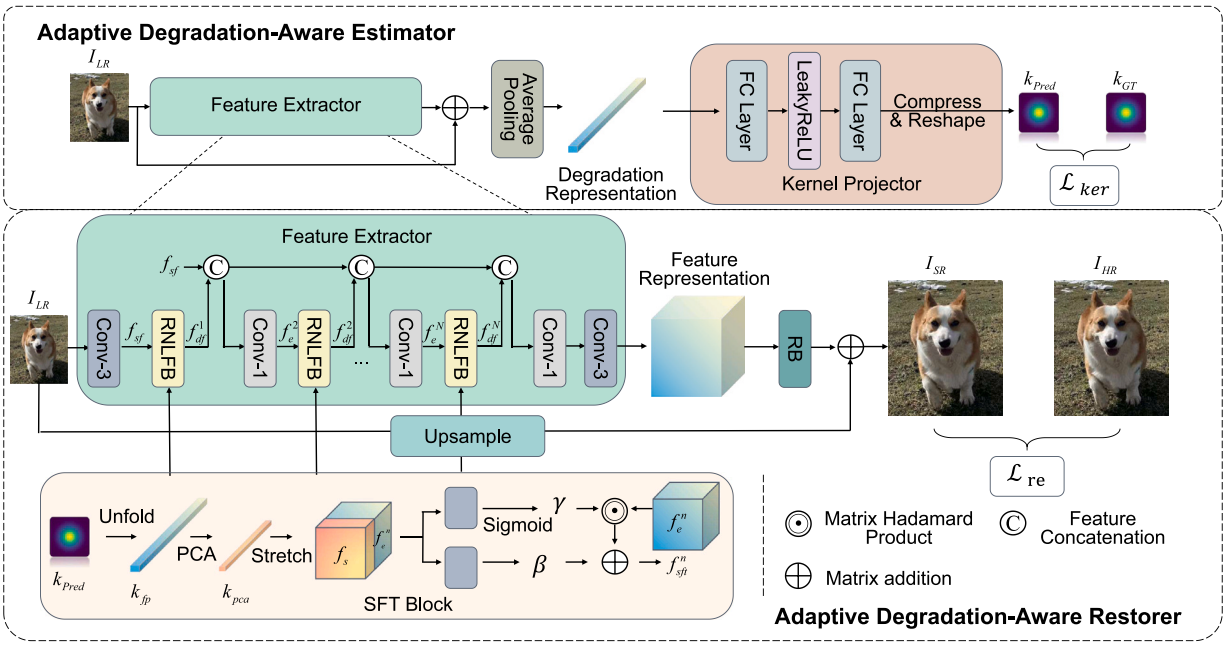


Fig. 2. The framework of Adaptive Degradation-Aware Estimator and Adaptive Degradation-Aware Restorer.

is used for degradation kernel estimation, and the function of feature representation is to achieve SR image reconstruction. Specifically, as demonstrated in Fig. 2, given a LR image I_{LR}, the feature extractor first utilizes a shallow feature extractor to map the I_{LR} with 3 channels onto the feature space, the output regards as f_{sf} with C channels, which can be expressed as follows:

$$f_{sf} = \mathcal{H}_{SFE}(I_{LR}), \tag{2}$$

where $\mathcal{H}_{SFE}(\cdot)$ represents the function of shallow feature extractor that consists of a 3×3 convolutional kernel. Then, we introduce Region Non-Local Feature Block (RNLFB) for deep feature similarity learning

$$f_{df}^{n} = \begin{cases} \mathcal{H}_R(f_{sf}), & n = 1 \\ \mathcal{H}_R(\mathcal{H}_{comp}(\mathcal{H}_{fuse}([f_{sf}, f_{df}^{1}, \ldots, f_{df}^{n-1}]))), & n \in \{2, 3, \ldots, N\} \end{cases}, \tag{3}$$

where $\mathcal{H}_R$, $\mathcal{H}_{comp}$ and $\mathcal{H}_{fuse}$ stands for the function of RNLFB, the operation of feature channel compression and the operation of feature fusion by 1×1 convolutional kernel, respectively. $[f_{sf}, f_{df}^{1}, \ldots, f_{df}^{n-1}]$ represents the feature concatenation of shallow and deep features, f_{sf} *and* $\{f_{df}^{1}, \ldots, f_{df}^{n-1}\}$. f_{df}^{n} denotes the output of the nth RNLFB and N indicates the number of RNLFBs. By using RNLFB, the extracted deep features exhibit the long-range feature interdependences with flying colors, and the operation of feature fusion helps the feature extractor preserve the previous feature information without causing additional network parameters redundancy. Different from the baseline structure residual local feature block (RLFB), as demonstrated in Fig. 3, the RNLFB exploits a cascading mechanism to aggregate current and previous features, which increases the optional features for depth feature extraction. To capture global feature correlation, the lightweight Region Non-Local Attention (RNLA) with layer normalization is introduced in RNLFB. The inference of RNLA is shown as Fig. 4. The dimension of input feature maps is $\mathbb{R}^{H,W,C}$. First, RNLA exploits 1×1 convolutional kernel on the input feature maps to reduce the feature channel to C/r, yielding $Q_c \in \mathbb{R}^{H,W,C/r}$ and $K_c \in \mathbb{R}^{H,W,C/r}$. Then, $K_p \in \mathbb{R}^{H/s,W/s,C/r}$ is obtained by the average pooling operation with region scale factor s on K_c. Next, the reshape operation is applied on Q_c and K_p to generate $Q_r \in \mathbb{R}^{H\times W,C/r}$ and $K_r \in \mathbb{R}^{C/r,H\times W/s^2}$. Through multiplying Q_r by K_r, we can obtain the Attention Map $A \in \mathbb{R}^{H\times W\times H\times W/s^2}$, in which each position indicates the similarity between pixels and regions. Finally, the output feature maps of RNLA with the same dimensions as the input feature maps can be obtained by weighting Attention Map A after the operation of softmax on the reshaped input feature maps $V_r \in \mathbb{R}^{H\times W/s^2,C}$ with residual learning. How RNLA leverages long-range feature interdependences by non-local feature similarity learning in detail is demonstrated in Fig. 5. Note that the residual connections are ignored for clear illustration. Compared with traditional Non-Local feature similarity learning that focus on capturing the pixel-wise feature correlation, our RNLA in Fig. 5 indirectly prefers to capture the region-wise feature correlation through establishing pixel-to-region feature correlation, which can significantly economize computational resources consumption in time and space form $\mathcal{O}((H \times W) \times (H \times W))$ to $\mathcal{O}((H \times W) \times (H/s \times W/s))$. Assuming the region scale factor is equal to 2, We provide an instance to explain pixel-to-region and region-wise feature correlation as demonstrated in Fig. 6. Specifically, Position P_i can establish the feature correlation with region R_a in the input feature maps by using RNLA in Loop 1:

$$p_i^1 = f(A_1, p_i, R_a) \cdot p_i + p_i, \tag{4}$$

where $f(A_1, p_i, R_a)$ denotes the correlation coefficient between p_i and R_a according to the attention map A_1 generated by RNLA in Loop 1, yielding the output position p_i^1. Similarly, we can obtain the position p_c^1 that is weighted by the feature correlation between position p_c and region R_m. By executing another RNLA in Loop 2, the output position p_i^1 in Loop 1 can be weighted according to the feature correlation between the position p_i^1 and the region R_a, in which the positions p_a^1, p_b^1, p_c^1, and p_d^1 contains the feature correlation with the region R_m:

$$\begin{aligned} p_i^2 &= f(A_2, p_i^1, R_a) \cdot p_i^1 + p_i^1 \\ &= f(A_2, p_i^1, P(p_a^1, p_b^1, p_c^1, p_d^1)) \cdot p_i^1 + p_i^1 \\ &= f(A_2, p_i^1, P(p_a^1, p_b^1, f(A_1, p_c, R_m) \cdot p_c + p_c, p_d^1)) \cdot p_i^1 + p_i^1, \end{aligned} \tag{5}$$

where A_2 represents the attention map generated by RNLA in the Loop 2, $P(\cdot)$ denotes the operation of pooling between positions in the region. Therefore, by feeding the input feature maps to the RNLA in the Loop 1, the model can achieve the pixel-to-region feature correlation. After proceeding with another RNLA in the Loop 2, the model can indirectly establish the region-wise feature correlation. Furthermore, to enhance the model representation ability in both channel and spatial dimensions, we adopt the large kernel attention block [22] behind the region non-local similarity learning, integrating as Region-Kernel Attention

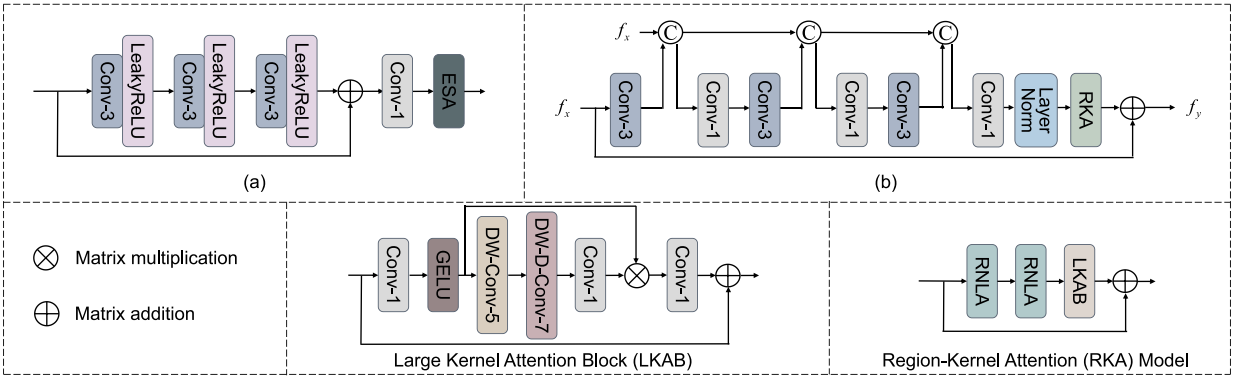


Fig. 3. The frameworks of (a). Residual Local Feature Block and (b). Region Non-Local Feature Block. DW-Conv-5 represents a Depth-Wise Convolution with the kernel size of 5 and DW-D-Conv-7 stands for a Depth-Wise Dilation Convolution with the kernel size of 7.

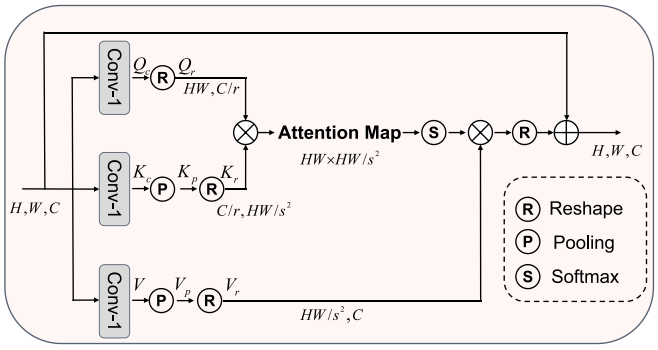


Fig. 4. The inference of Region Non-Local Attention.

Model to replace the ESA in the RLFB. By feeding the LR image to the feature extractor introduced above, we can leverage the extracted features to achieve the following degradation kernel estimation and the SR image reconstruction, respectively.

3.2. Kernel estimation from degradation representation

Previous SOTA blind SR methods [18–20,46,47] routinely introduce a sophisticated degradation kernel estimation network to predict the degradation kernel of LR image and a reconstruction network to utilize the estimated degradation kernel information. The limitation of separately designing two networks in different vision tasks is that it significantly increases the computational costs of the entire SR framework. Besides, Simultaneously training two networks demands meticulous attention to select the hyper-parameters to control the convergence of the loss function, which further complicates the training of the entire SR framework. Therefore, in our work, we leverage the feature extractor introduced in Section 3.1 to separately learn the degradation representation and the feature representation in both trainings of the Adaptive Degradation-Aware Estimator (ADAE) and the Adaptive Degradation-Aware Estimator Restorer (ADAR). Since the powerful ability of feature extraction by the core component, RNLFB, in feature extractor, both ADAE and ADAR can economize a huge of network parameters and calculations. Furthermore, a unified feature extractor in the ADAE and ADAR increases the training stability which can achieve higher performance compared with designing two separate feature extractors and it is unnecessary to validate the effectiveness of different feature extractors from two networks. In this section, how to utilize the ADAE achieving kernel estimation from degradation representation is as follows and the next section will describe how to utilize the ADAR achieving network reconstruction from feature representation with the help of estimated degradation kernel information. As illustrated in Fig. 2, the Adaptive Degradation-Aware Estimator utilizes the average pooling to gain the Degradation Representation from the extracted features with residual learning. The mathematical formulation of this process can be described as

$$\mathcal{D}_{\mathcal{R}} = \mathcal{H}_{pool}(\mathcal{F}_{\mathcal{R}}) + I_{LR}, \tag{6}$$

where $\mathcal{H}_{pool}(\cdot)$ represents the function of average pooling. $\mathcal{D}_{\mathcal{R}}$ denotes the degradation representation, and the $\mathcal{F}_{\mathcal{R}}$ stands for the feature representation encoded by the fusion of previously extracted shallow and depth features which are proceeded by a channel compression unit and a fusion feature extraction unit. The process can be formulated as follows.

$$\mathcal{F}_R = \mathcal{H}_{DFE}(\mathcal{H}_{comp}(\mathcal{H}_{fuse}([f_{sf}, f_{df}^1, \ldots, f_{df}^N]))), \tag{7}$$

where $\mathcal{H}_{DFE}(\cdot)$ represents the depth feature extractor using a 3×3 convolutional kernel. Then, the degradation kernel projector is equipped for degradation kernel estimation under the supervision of the corresponding ground-truth kernel

$$\mathcal{K}_{pred} = \mathcal{H}_{pred}(\mathcal{D}_{\mathcal{R}}) = \mathcal{H}_{trans}(\mathcal{H}_{fc}(\mathcal{H}_{act}(\mathcal{H}_{fc}(\mathcal{D}_{\mathcal{R}})))), \tag{8}$$

where $\mathcal{H}_{fc}(\cdot)$ and $\mathcal{H}_{act}(\cdot)$ denote the function of the fully connected layer and the non-linear activation layer, LeakyReLU, respectively. $\mathcal{H}_{trans}(\cdot)$ represents the function of dimension transformation. In the feature dimension analysis, $\mathcal{F}_{\mathcal{R}} \in \mathbb{R}^{H,W,C}$ is first projected onto a C-dimensional linear space by the operation of average pooling to obtain $\mathcal{D}_{\mathcal{R}} \in \mathbb{R}^{C}$. After obtaining the logits from $\mathcal{D}_{\mathcal{R}}$ through two fully connected layers and a non-linear activation layer, we utilize linear expansion operation to project the logits onto a l^2-dimensional linear space and utilize reshape operation to gain the output predicted kernel $\mathcal{K}_{pred} \in \mathbb{R}^{l,l}$. The degradation kernel estimator can be optimized by minimizing the $l1$ distance as

$$\begin{aligned} \omega_p &= \underset{\omega_p}{argmin} \|\mathcal{L}_{ker}\|_1 = \underset{\omega_p}{argmin} \|\mathcal{K}_{GT} - \mathcal{K}_{pred}\|_1 \\ &= \underset{\omega_p}{argmin} \|\mathcal{K}_{GT} - \mathcal{P}_{ke}(I_{LR}; \omega_p)\|_1, \end{aligned} \tag{9}$$

where $\mathcal{P}_{ke}(I_{LR}; \omega_p)$ denotes the function of the proposed adaptive degradation-aware estimator with the parameter ω_p and the input I_{LR}. $\mathcal{L}_{ker}$ represents the Loss of degradation kernel estimator.

3.3. Network reconstruction from feature representation

After degradation kernel estimation, the proceed mission is to incorporate the predicted kernel information in the Adaptive Degradation-Aware Restorer (ADAR) to accomplish the SR image reconstruction, which is illustrated in Fig. 2. For each RNLFB, we utilize a spatial feature transform (SFT) block that provides affine transformation [18] for applying the degradation kernel information in the input feature. Concretely, as demonstrated in Fig. 2, given a predicted kernel $\mathcal{K}_{pred} \in$

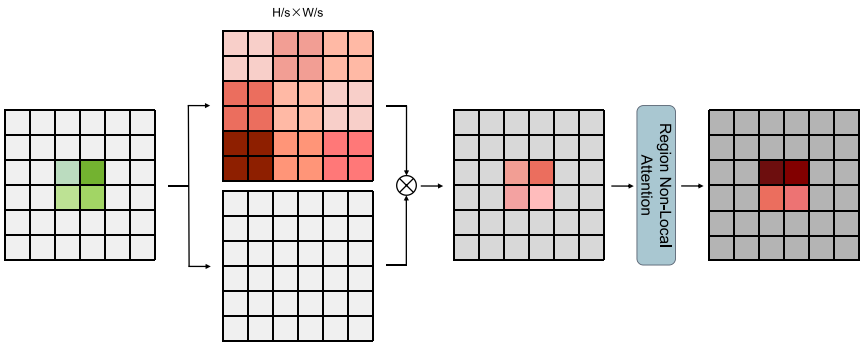


Fig. 5. The illustration of region non-local feature similarity learning. Every position (*e.g.,* the green-marked positions at different levels) is weighted according to the attention map $A \in \mathbb{R}^{H/s \times W/s}$ (*e.g.,* the red-marked regions at different levels indicate the degree of similarity to the green-marked position) that consists of pixel-to-region feature correlation. The attention map is generated by Region Non-Local Attention (RNLA). After another RNLA, every position in the output can indirectly accumulate the region-wise correlation since the position in the middle map generated by the first RNLA collects the pixel-to-region feature correlation.

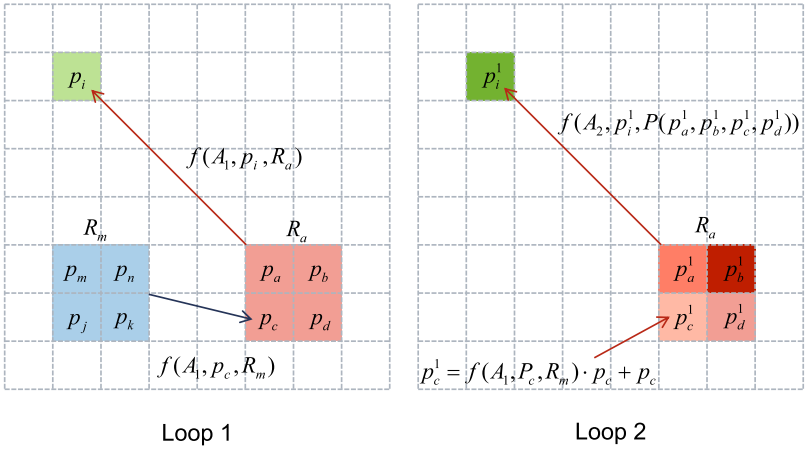


Fig. 6. An example of establishing region-wise feature correlation by 2 RNLAs.

$\mathbb{R}^{l,l}$, SFT block first unfold it onto l^2-dimensional linear space to gain the unfolded predicted kernel $\mathcal{K}_{fp} \in \mathbb{R}^{l^2}$. Then, we utilize principal component analysis (PCA) to compress the kernel onto b-dimensional linear space via multiplying PCA matrix $P \in \mathbb{R}^{l^2,b}$ by $\mathcal{K}_{fp} \in \mathbb{R}^{l^2}$. The compressed kernel information is expressed as $\mathcal{K}_{pca} \in \mathbb{R}^b$, which effectively reduces the computational costs. To combine the kernel information and the input feature of the nth RNLFB $f_e^n \in \mathbb{R}^{H,W,C}$ into the same dimension, we utilize the dimensionality stretching strategy [38] to project the $\mathcal{K}_{pca} \in \mathbb{R}^b$ onto the same dimension as f_e^n to gain $f_s \in \mathbb{R}^{H,W,b}$. The concatenation of f_e^n and f_s can be formulated as $[f_e^n, f_s] \in \mathbb{R}^{H,W,(C+b)}$. $[f_e^n, f_s]$ is subsequently processed by two convolutional layers and a sigmoid layer for learning the scaling parameter γ and the shifting parameter β in the affine transformation. Therefore, we can obtain the output feature f_{sft}^n of nth SFT block that contains the degradation kernel information by the incorporation between RNLFB and SFT block

$$f_{sft}^n = \gamma \odot (f_e^n) + \beta = \begin{cases} \gamma \odot (f_{sf}) + \beta, & n = 1 \\ \gamma \odot \mathcal{H}_{comp}(\mathcal{H}_{fuse}([f_{df}^1, \ldots, f_{df}^{n-1}])) + \beta, & n \in \{2, 3, \ldots, N\} \end{cases}, \quad (10)$$

where f_{sft}^n is fed into nth RNLFB for feature extraction with the help of degradation kernel information. Different from the phase of degradation kernel estimation, in the reconstruction step, we utilize the reconstruction block (RB) that consists of a pixel-shuffle upsampled layer to obtain the SR image I_{SR} from the feature representation with the residual learning. The reconstruction network can be optimized by minimizing the $l1$ distance as

$$\omega_r = \underset{\omega_r}{argmin} \|\mathcal{L}_{re}\|_1 = \underset{\omega_r}{argmin} \|I_{HR} - I_{SR}\|_1 = \underset{\omega_r}{argmin} \|I_{HR} - \mathcal{P}_{re}(I_{LR}; \omega_r)\|_1, \quad (11)$$

where $\mathcal{P}_{re}(I_{LR}; \omega_r)$ denotes the function of the Restorer with the parameter ω_r and the input I_{LR}. $\mathcal{L}_{re}$ represents the Loss of the Restorer.

4. Experiments

4.1. Implementation details

Datasets and evaluation index. We combine 3450 High-Resolution (HR) images from DIV2K [48] and Flickr2K [49] as the training data of our model, of which data augmentation as [9] do is adopted to make the most. As the testing data, we adopt five standard benchmark datasets, Set5 [50], Set14 [51], B100 [52], Urban100 [53], and Manga109 [54] to evaluate the SR performance of our model. We utilize two evaluation indexes, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [55] on the luminance (Y) channel as previous SR methods did.

Degradation models. Following [43,56], the synthetic LR images for the training model are generated according to Eq. (1) by two degradation kernels, *i.e.,* isotropic and anisotropic Gaussian degradation kernels with a kernel size of 21. In **Setting 1**, *i.e.,* the degradation model by isotropic degradation kernel, the kernel width is uniformly sampled from the range $[0.2, 2.0]$, $[0.2, 3.0]$, and $[0.2, 4.0]$ for training $\times 2$, $\times 3$, and $\times 4$ SR model, respectively. For testing the SR performance of the SR model, as [18] does, the kernel sets named *Gaussian8*, consisting of 8 isotropic Gaussian degradation kernels from kernel width range $[0.8, 1.6]$, $[1.35, 2.4]$, and $[1.8, 3.2]$ with uniform stride, are used for generating the HR-LR testing data pairs. In **Setting 2**, *i.e.,* the degradation model by anisotropic degradation kernel, for training the SR model, the length of degradation kernel at each axis is determined by widths λ_1

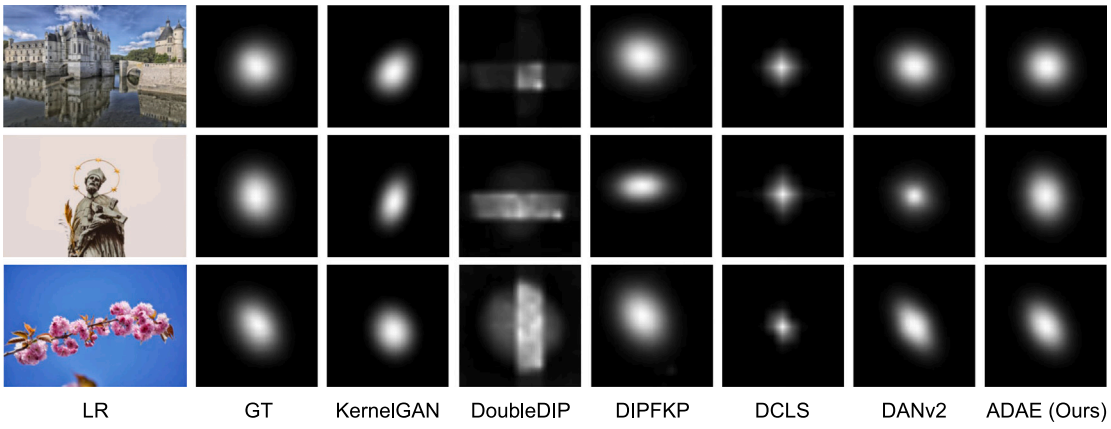


Fig. 7. Visual comparison results generated by different degradation kernel estimators on Img_65, Img_68, and Img_98 from DIV2KRK [56].

Table 1
Average LR-PSNR/LR-SSIM values on DIV2KRK in the task of Kernel Estimation of different methods in **Setting 2**. The best LR-PSNR/LR-SSIM values are **bold**.

Options	Zero-Shot Learning with Internal Statistics			Degradation kernel Learning with External Datasets		
	KernelGAN [56]	DoubleDIP [59]	DIPFKP [60]	DCLS [47]	DANv2 [61]	ADAE
Performance	34.38/0.9615	37.64/0.9679	29.70/0.9366	32.10/0.9230	48.80/0.9973	**50.78/0.9987**
#Params (M)	0.54	0.56	0.50	6.70	4.74	0.64
#Flops (G)	544.90	30 279.19	30 279.12	17.43	1088.15	38.27
#Times (s)	34.17	156.34	156.17	0.44	0.63	0.48

and λ_2 in the random range $[0.6, 5.0]$ with a random angle $\theta \sim U[-\pi, \pi]$. To get closer to the real degradation scenario, the multiplicative noise is added to the degradation kernel (up to 25% of each pixel value of the kernel) and normalized to sum to one. For testing the SR model, we also choose the dataset DIV2KRK [56] which consists of 100 LR images degraded by a surge of anisotropic degradation kernels as most SOTA blind SR methods [43,45,47,57,58] do. To further validate the model generalization capacity, we choose the eight representative anisotropic kernels to make up the kernel sets, named *AnisoGaussian8*, in which we randomly and uniformly sample the degradation angle in range $[0, \pi/4, \pi/2, 3\pi/4]$ with the kernel width set of $\lambda_1 = 0.8$, $\lambda_2 = 1.6$ and $\lambda_1 = 2.0$, $\lambda_2 = 4.0$. We use AnisoGaussian8 to generate LR images on datasets Set5, Set14, B100, Urban100, Manga109 for the **Setting 2**. Since **Setting 2** is closer to the degradation mode of LR images in real-world scenarios, the testing datasets for all ablation studies is determined by **Setting 2**.

Training Settings. The number of RNLFBs is set as 3 and 6 to learn the degradation representation and feature representation, respectively for the best trade-off between model computational complexity and SR performance. During training the Adaptive Degradation-Aware Estimator (ADAE), the patch size is set as 64×64, the ADAE is trained for 200 epochs with the learning rate initialized to 1×10^{-4}. During the training of the Adaptive Degradation-Aware Estimator (ADAR), the parameters of the ADAE are frozen, extracting degradation kernel information as prior for blind SR reconstruction. The HR patch size is set as 256×256 for all up-sampled scale factors, the SR model is trained for 600 epochs with the learning rate initialized to 4×10^{-4}. For training all models, the batch size is 64 and the learning rate is first set to one-tenth of the initial learning rate, rising uniformly to the value of the initial learning rate in the first 10 epochs, and then decreasing to 1.25×10^{-5} with a cosine annealing in the remaining epochs [62]. We exploit Adam [63] as the optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$. All models are trained on the Pytorch framework with an RTX3090 GPU.

4.2. Study of kernel estimation

To demonstrate the effectiveness of the introduced Adaptive Degradation-Aware Estimator (ADAE), we conduct both subjective and objective experiments to appraise the kernel estimation performance compared to other SOTA kernel estimators, including KernelGAN [56], DoubleDIP [59], DIP-FKP [60], DANv2 [61], and DCLS [47]. The subjective visual comparison results generated by different degradation kernel estimators on some test images from DIV2KRK [56] are demonstrated in Fig. 7. It can be evidently perceived that the degradation kernels predicted by the proposed ADAE have distinct contours, and whose intensity distribution is closest to the Ground-Truth. For the objective evaluation criteria, we adopt the LR-PSNR and LR-SSIM [64] to validate the accuracy of the estimated kernel. LR-PSNR and LR-SSIM are obtained by calculating the PSNR and SSIM value between the ground-truth LR image and the fake LR image that is degraded by the estimated kernel. The Flops and Running times of the model are tested on an LR size of 320 × 180 with an RTX3090 GPU. As demonstrated in Table 1, since the network training is done during test time of the zero-shot learning with internal statistics, *e.g.,* KernelGAN, DoubleDIP and DIP-FKP, there are no actual #Flops for generating degradation kernels. The #Flops of the methods with zero-shot learning in Table 1 is calculated by multiplying the #Flops generated in a single iteration with the total number of iterations. Compared with the degradation kernel estimators trained with the external datasets, estimators using zero-shot learning cannot obtain favorable degradation kernel estimated results due to the limited feature information of the input test image itself. In addition, estimators using zero-shot learning require a significant amount of time to complete degradation kernel estimation, which is not suitable for resource-constrained platforms. Therefore, we are inclined to compare estimators trained with the external datasets, *e.g.,* DANv2 [61], DCLS [47], and the proposed ADAE. Compared with the SOTA estimators by degradation kernel learning with external datasets, our ADAE gains the best LR-PSNR/LR-SSIM performance, which means that the degradation kernel estimated by ADAE is closet to the Ground-Truth degradation kernel. Specifically, compared with DANv2 [61] whose kernel performance is sub-optimal, the ADAE can save an amount of computational resources costs, *e.g.,* ADAE has only 35% of the Flops of DANv2, while still outperforming DANv2 in kernel estimation. In comparison to the fastest method for kernel estimation, DCLS [47],

Table 2
Average LR-PSNR/LR-SSIM values on DIV2KRK for investigations of the components in Feature Extractor on the task of kernel estimation in **Setting 2**. The best LR-PSNR/LR-SSIM values are **bold**.

Options	Baseline	Baseline+Projector (Model A)	A+CA	A+ESA	A+LKA	A+RNLA	ADAE
Kernel Projector		✓	✓	✓	✓	✓	✓
CA			✓				
ESA				✓			
LKA					✓		✓
RNLA						✓	✓
#Params (K)	547.96	556.28	558.02	577.40	608.31	587.48	639.51
#Flops (G)	29.83	29.83	29.84	30.34	32.77	35.33	38.27
Performance	43.80/0.9980	46.97/0.9981	47.79/0.9981	46.85/0.9978	47.09/0.9983	48.45/0.9985	**50.78/0.9987**

Table 3
Average PSNR/SSIM values on six datasets for investigations of the components in Feature Extractor on the task of SR reconstruction in **Setting 2**. The best PSNR/SSIM values are **bold**.

Options	Set5	Set14	B100	Urban100	Manga109	DIV2KRK	#Params (M)	#Flops (G)
Baseline	31.23/0.8803	28.10/0.7638	27.23/0.7210	25.19/0.7557	29.35/0.8902	28.73/0.7904	2.11	119.44
Baseline+Projector (Model A)	31.16/0.8806	28.12/0.7640	27.24/0.7214	25.24/0.7572	29.44/0.8919	28.77/0.7911	2.12	119.44
A+CA	31.42/0.8809	28.14/0.7621	27.30/0.7202	25.41/0.7601	29.67/0.8936	28.48/0.7805	2.12	119.48
A+ESA	31.42/0.8814	28.16/0.7635	27.27/0.7201	25.25/0.7574	29.56/0.8931	28.73/0.7917	2.18	120.96
A+LKA	31.51/0.8827	28.23/0.7650	27.34/0.7223	25.47/0.7626	29.64/0.8938	28.54/0.7848	2.27	128.27
A+RNLA	31.61/0.8847	28.28/0.7672	27.38/0.7240	25.60/0.7680	29.96/0.8993	28.96/0.7943	2.21	135.95
A+RKA (ADAR)	**31.77/0.8864**	**28.36/0.7687**	**27.47/0.7265**	**25.87/0.7754**	**30.29/0.9039**	**29.05/0.7982**	2.37	144.77

Table 4
Average LR-PSNR/LR-SSIM values on DIV2KRK in the task of Kernel Estimation of different Feature Extractors in ADAE in **Setting 2**. The best LR-PSNR/LR-SSIM values are **bold**.

Options	Residual Block (RB) in ADAE	Residual Local Feature Block (RLFB) in ADAE	Region Non-Local Feature Block (RNLFB) in ADAE
Performance	44.71/0.9977	46.74/0.9984	**50.78/0.9987**
#Params (K)	703.48	648.70	639.51
#Flops (G)	38.32	33.98	38.27
#Times (ms)	452.06	460.74	481.15

ADAE achieved a breakthrough in kernel estimation performance with only a slight increase in model running time.

4.3. Study of the components in feature extractor

Since all the Feature Extractors in different experimental settings are designed to both learn the degradation representation and feature representation for the task of degradation kernel estimation and SR reconstruction, respectively, we have conducted several ablation experiments to investigate the effectiveness of the proposed kernel projector and region non-local attention mechanism in Feature Extractor in the following tasks.

Degradation Kernel Estimation. The experimental setting 'Baseline' in Table 2, represents the proposed Feature Extractor as shown in Fig. 2 does not consist of kernel projector and any attention mechanisms. When we applied the proposed kernel projector in the 'Baseline', *i.e.,* 'Baseline+Projector (model A)', it can be observed that a 1.32 dB increase in the PSNR value, accompanied by a negligible increase in the number of model parameters. This demonstrates that the kernel projector can efficiently predict the degradation kernel from the degradation representation. When we applied the proposed Region Non-Local Attention (RNLA) mechanism in the model A, *i.e.,* A+RNLA, compared with several traditional attention mechanisms that are widely employed in the field of super-resolution [14,65,66], *e.g.,* 'A+CA', 'A+ESA', 'A+LKA', our RNLA achieves a favorable PSNR value improvement in the task of degradation kernel estimation. With the help of LKA, the performance of the experimental setting 'A+RNLA' can be further improved, which verifies the ADAE can capture the feature correlation in both channel and spatial dimensions to help the model improve the capacity of feature extraction.

SR Reconstruction. The aforementioned task of kernel estimation is relatively simple, while the subsequent task of SR Reconstruction, due to the sophisticated textures reconstruction and the interference from degradation, further validates the effectiveness of the proposed Feature Extractor and its components. The ablation experiments result in the task of SR Reconstruction are demonstrated in Table 3. Comparing the SR performance of Model A and the Baseline in SR reconstruction, we can draw a conclusion that more accurate degradation kernel information projected by the kernel projector can help the baseline model better reconstruct SR images. To investigate the impact of different attention mechanisms on the performance of Feature Extractor, we incorporate several attention mechanisms, *e.g.,* Channel Attention (CA) mechanism, Enhanced Spatial Attention (ESA) mechanism, Large Kernel Attention (LKA) mechanism, the proposed Region Non-Local Attention (RNLA) mechanism, and Region Kernel Attention (RKA) mechanism on Model A. By comparing the SR reconstruction performance of the model 'A+RNLA' with other attention models, *e.g.,* 'A+CA', 'A+ESA', it demonstrated the powerful feature extraction performance of RNLA. Especially on Urban100 and DIV2KRK datasets whose image resolutions are larger than other datasets, due to the ability to capture global feature correlation, the 'A+RNLA' model outperformed the best-performing model 'A+LKA' among traditional attention models with PSNR improvements of 0.13 dB, 0.25 dB, and 0.42 dB on Urban100, Manga109, and DIV2KRK, respectively. Based on RNLA, we introduce the RKA mechanism, which can further help the Feature Extractor to improve the SR performance since we supply the feature correlation in channel dimension into consideration with little computational resources costs.

4.3.1. Study of different feature extractors

We combine the proposed RNLFB with a cascading mechanism to make up the Feature Extractor for both Degradation Kernel Estimation and SR Reconstruction. To validate the powerful feature extraction capacity of the proposed Feature Extractor, we choose two SOTA Feature Extractors as the compared block, *i.e.,* Residual Block (RB) [9] and Residual Local Feature Block (RLFB) [14]. RB is one of the most successful Feature Extractors in both Blind [18,43,45,64] and Non-Blind [9,11,24,65] SR fields and RLFB is the winning Feature Extractor of the runtime track and the second place in the Overall Performance Track in the NTIRE challenge on Efficient SR [67]. The experimental

Table 5
Average PSNR/SSIM values on six datasets for investigations of different Feature Extractors on the task of SR reconstruction in **Setting 2**. The best PSNR/SSIM values are **bold**.

Options	Set5	Set14	B100	Urban100	Manga109	DIV2KRK	#Params (M)	#Flops (G)	#Times (ms)
RB in SR model	31.37/0.8816	28.17/0.7645	27.30/0.7219	25.34/0.7595	29.66/0.8938	28.72/0.7871	2.61	147.91	469.83
RLFB in SR model	31.45/0.8830	28.21/0.7648	27.31/0.7232	25.34/0.7595	29.60/0.8947	28.76/0.7926	2.76	152.98	490.91
RNLFB in SR model	**31.77/0.8864**	**28.36/0.7687**	**27.47/0.7265**	**25.87/0.7754**	**30.29/0.9039**	**29.05/0.7982**	2.37	144.77	565.18

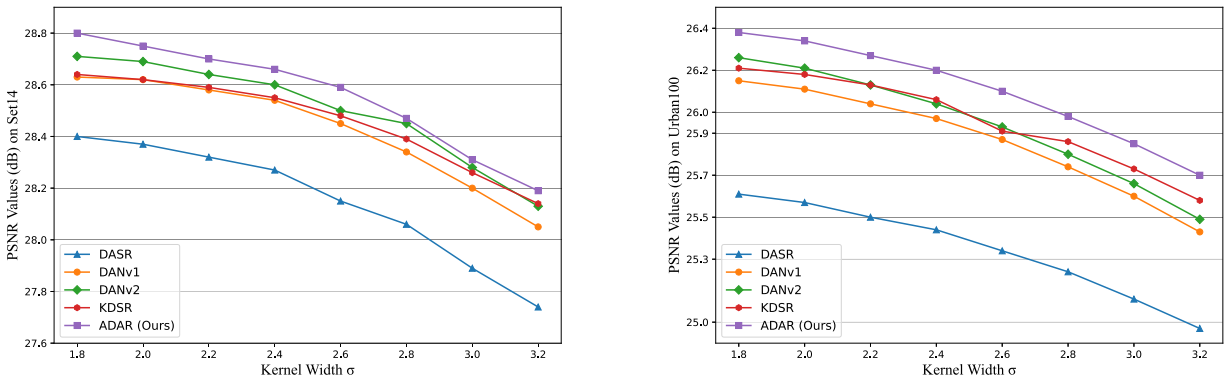


Fig. 8. The PSNR performance comparison on the setting of kernel width σ from 1.8 to 3.2. (a) ×4 SR performance validation on Set14. (b) ×4 SR performance validation on Urban100.

results were categorized into Degradation Kernel Estimation and SR Reconstruction.

Degradation Kernel Estimation. For a reasonable comparison, we substitute the Feature Extractor in ADAE with multiple RB and RLFB, while maintaining all other aspects of ADAE unchanged. Additionally, we standardized all experimental models to have the same parameter and computational complexity for fair performance evaluation. It can be demonstrated in Table 4 that our RNLFB outperforms RB and RLFB by a larger margin in kernel estimation performance while only sacrificing a small amount of model execution speed. The reason for this phenomenon is that both RLFB and RNLFB have the ability to capture long-range feature correlations, and RNLFB has a stronger feature representative ability than RLFB according to the comparison between the model 'A+RKA' in RNLFB and the model 'A+ESA' in RLFB from Table 3.

SR Reconstruction. For all feature extractors in the following experiments, we utilize a spatial feature transform (SFT) block that provides affine transformation to make full use of degradation kernel prior information generated by themselves. During training the experiment under the setting 'RLFB in SR model', we observe that simply stacking RLFB will cause a vanishing gradient. Therefore, we performed residual learning every two RLFBs to ensure the correct back-propagation of gradients. Compared with the RLFB with the sub-optimal SR performance in Table 5, our RNLFB exhibits the powerful feature representative ability and achieves a breakthrough in SR performance, *e.g.,* 0.53 dB PSNR improvement on Urban100 and 0.69 dB PSNR improvement on Manga109.

The above ablation investigation further validates the effectiveness of RNLFB which realizes more accurate kernel degradation estimation and utilizes the kernel prior information more sufficiently to help the SR model complete SR reconstruction with better performance, compared with other SOTA feature extractors.

4.4. Comparison with state-of-the-arts

Setting 1 : Isotropic Gaussian Kernels. For **setting 1**, following the SOTA blind-SR methods [18,43,45,58,68], our method is validated on the five datasets, whose HR observations are degraded by *Gaussian8* kernels to obtain the corresponding LR counterparts. We choose twelve SOTA SR methods, including non-blind SR methods, *i.e.,* bicubic method and RCAN [65], Blind SR methods, *i.e,* ZSSR [40], IKC [18], DANv1 [43], DANv2 [61], DASR [19], DSSR [45], CDCN [58], SRDRL [46], AdaTarget [69], and KDSR-M [68], to make comparisons on model SR performance, model parameters, and model calculation for ×2, ×3, and ×4 SR. The comparison results are demonstrated in Table 6. The SR performance of the non-blind SR method RCAN is only higher than that of the Bicubic method due to the degradation mode of tested LR images deviating from the assumed mode of the model. The blind SR method, ZSSR, can only utilize the limited internal statistics of the input LR image with Zero-Shot learning, leading to a lower SR performance, even inferior to that of the non-blind SR method RCAN trained on external datasets. Furthermore, since ZSSR undergoes iterative training while validating images, the runtime of the model far exceeds that of methods using external datasets. By combining the detail and structure feature information into SR reconstruction, DSSR [45] and CDCN [58] improve blind SR performance by a larger margin against many blind SR methods [18,19,43,46,69]. However, CDCN and DSSR suffer from a large amount of computational resource costs, which are not suitable for real-world application scenarios. Benefiting from the powerful yet efficient feature extractor, the proposed method, ADAR, exhibits comparable blind SR performance overall up-sampled scales with fewer model parameters and calculations. Especially on ×2 SR, the blind SR performance of our ADAR outperforms other SOTA methods. In the most common ×4 blind SR performance comparison, our ADAR has model parameters about 59% less than the KDSR [68], while ADAR gains +0.14 dB PSNR value improvement on Urban100. Although SRDRL [46] has fewer model parameters than other blind SR methods, the blind SR performance of the model is low owing to the influx of a large amount of redundant feature information. Besides, our ADAR only consumes 18% of the Flops required by SRDRL, yet it obtains a PSNR value improvement of 2.51 dB on Urban100. To further verify the effectiveness and generalization ability of the SR model, we also conducted blind SR performance comparison experiments under different kernel width settings. As illustrated in Fig. 8, our ADAR achieves the best PSNR value on Set14 and Urban100 under the setting of eight different kernel widths from the range of 1.8 to 3.2.

Setting 2 : Anisotropic Gaussian Kernels. For **setting 2**, following the SOTA blind-SR methods [43,45,47,58], our method is validated on the six datasets. Among them, DIV2KRK [56] is a widely used

Table 6
Quantitative evaluation for ×2, ×3 and ×4 SR in **Setting 1**. The best PSNR/SSIM values is shown in **bold** and the second-best PSNR/SSIM values is underline.

Method	Scale	Set5	Set14	B100	Urban100	Manga109	Params (M)	Flops (G)
Bicubic	×2	28.82/0.8577	26.02/0.7634	25.92/0.7310	23.14/0.7528	25.60/0.8498	–	–
RCAN	×2	31.37/0.8918	28.63/0.8144	28.16/0.7822	25.58/0.7863	28.48/0.8933	15.44	3529.73
ZSSR+Bicubic	×2	31.08/0.8786	28.35/0.7933	27.92/0.7632	25.25/0.7618	28.05/0.8769	0.23	–
IKC	×2	36.82/0.9534	32.81/0.9074	31.62/0.8884	30.21/0.9019	36.15/0.9660	5.32	6,847.51
DANv1	×2	37.34/0.9526	33.08/0.9041	31.76/0.8858	30.60/0.9060	37.23/0.9710	4.18	3833.16
DANv2	×2	37.60/0.9544	33.44/0.9094	32.00/0.8904	31.43/0.9174	38.07/0.9734	4.57	3771.51
DASR	×2	37.01/0.9504	32.61/0.8959	31.59/0.8815	30.25/0.9016	36.20/0.9687	5.84	593.37
DSSR	×2	37.46/0.9529	33.17/0.9044	31.95/0.8901	31.03/0.9118	37.40/0.9720	8.31	10 125.48
CDCN	×2	37.62/0.9550	33.45/0.9084	32.02/0.8905	31.44/0.9164	37.57/0.9729	11.26	2559
ADAR (Ours)	×2	**37.66/0.9552**	**33.58/0.9106**	**32.06/0.8914**	**31.69/0.9199**	**38.08/0.9738**	2.34	619.10
Bicubic	×3	26.21/0.7766	24.01/0.6662	24.25/0.6356	21.39/0.6203	22.98/0.7576	–	–
RCAN	×3	28.28/0.8158	26.13/0.7133	26.06/0.6772	23.31/0.6750	25.16/0.8057	15.63	1563.71
ZSSR+Bicubic	×3	28.25/0.7989	26.11/0.6942	26.06/0.6633	23.26/0.6534	25.19/0.7914	0.23	–
IKC	×3	33.06/0.9146	29.38/0.8233	28.53/0.7899	27.43/0.8302	32.43/0.9316	5.32	3180.68
DANv1	×3	34.04/0.9199	30.09/0.8287	28.94/0.7919	27.65/0.8352	33.16/0.9382	4.37	1754.81
DANv2	×3	**34.19**/0.9209	30.20/0.8309	29.03/0.7948	27.83/0.8395	33.28/0.9400	4.74	1904.27
DASR	×3	33.48/0.9125	29.64/0.8154	28.63/0.7829	26.88/0.8149	32.03/0.9286	5.84	279.05
DSSR	×3	34.05/0.9197	30.09/0.8270	28.98/0.7923	27.64/0.8349	33.07/0.9384	9.34	7126.24
CDCN	×3	34.12/0.9209	30.19/0.8313	29.04/0.7952	27.91/0.8410	**33.30**/0.9400	11.81	1177.91
ADAR (Ours)	×3	34.06/**0.9212**	**30.28/0.8323**	**29.07/0.7966**	**28.00/0.8442**	33.11/**0.9409**	2.35	266.83
Bicubic	×4	24.57/0.7108	22.79/0.6032	23.29/0.5786	20.35/0.5532	21.50/0.6933	–	–
RCAN	×4	26.60/0.7598	24.85/0.6513	25.01/0.6170	22.19/0.6078	23.52/0.7428	15.59	917.60
ZSSR+Bicubic	×4	26.45/0.7279	24.78/0.6268	24.97/0.5989	21.11/0.5805	23.53/0.7240	0.23	–
IKC	×4	31.67/0.8829	28.31/0.7643	27.37/0.7192	25.33/0.7504	28.91/0.8782	5.32	2528.03
DANv1	×4	31.89/0.8864	28.42/0.7687	27.51/0.7248	25.86/0.7721	30.50/0.9037	4.33	1098.33
DANv2	×4	32.00/0.8885	28.50/0.7715	27.56/0.7277	25.94/0.7748	30.45/0.9037	4.71	1088.14
DASR	×4	31.46/0.8789	28.11/0.7603	27.44/0.7214	25.36/0.7506	29.39/0.8861	5.84	185.66
DSSR	×4	31.97/0.8870	28.43/0.7679	27.49/0.7229	25.73/0.7668	30.44/0.9023	8.90	5549.74
CDCN	×4	**32.04**/0.8883	28.47/0.7695	27.52/0.7257	25.92/0.7731	30.55/0.9036	11.70	745.25
SRDRL	×4	29.06/0.8312	26.55/0.7118	26.28/0.6742	23.59/0.6769	25.86/0.8126	1.34	790.43
AdaTarget	×4	31.58/0.8814	28.14/0.7626	27.43/0.7216	25.72/0.7683	29.97/0.8955	16.70	1032.59
KDSR-M	×4	32.02/0.8892	28.46/**0.7761**	27.52/0.7281	25.96/0.7760	**30.58**/0.9026	5.80	191.42
ADAR (Ours)	×4	32.01/**0.8893**	**28.56**/0.7731	**27.58/0.7295**	**26.10/0.7814**	30.53/**0.9073**	2.37	144.77

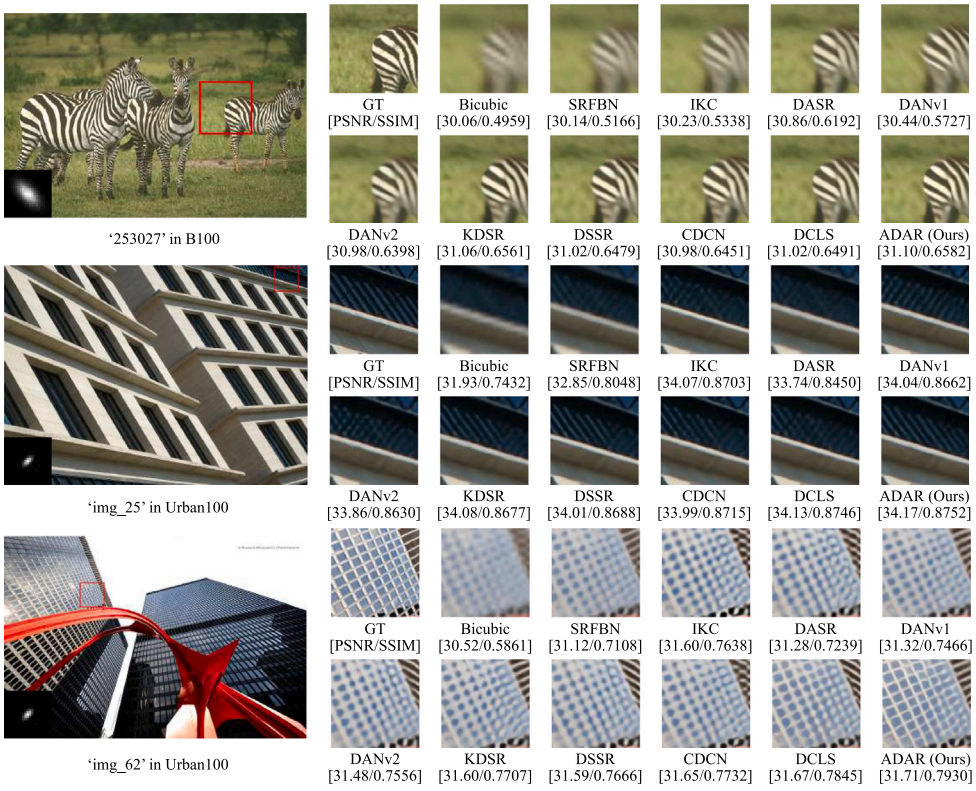


Fig. 9. Qualitative comparison of ADAR with other SOTA methods on ×4 image SR of B100 and Urban100 in **Setting 2**.

Fig. 10. Qualitative comparison of ADAR with other SOTA methods on ×4 image SR of Set14 and Manga109 in **Setting 2**.

validation dataset, whose HR observations are degraded by 100 different anisotropic Gaussian Kernels. We also provided blind SR performance comparison on the other five datasets, consisting of Set5, Set14, B100, Urban100, and Manga109, whose HR observations are degraded by *AnisoGaussian8* degradation kernels to validate the generalization ability of SR models. Similar to **setting 1**, we choose eleven SOTA SR methods, including non-blind SR methods, *i.e.,* bicubic method, SRFBN [16], and SAFMN-L [17], Blind SR methods, *i.e,* IKC [18], DANv1 [43], DANv2 [61], DASR [19], DSSR [45], CDCN [58], DCLS [47], and KDSR-M [68], to make comparisons on model SR performance, model parameters, and model calculation. For a fair comparison, we retrained IKC, DASR, and KDSR by their official code and options in **Setting 2**. To save the training time, like KDSR [68], we only make a blind SR performance comparison in ×4 SR. The comparison results are reported in Table 7. From the Table, we can see that our ADAR achieves the best PSNR/SSIM values with much lower computational resources costs in the common dataset, *i.e.,* DIV2KRK compared with other SR methods. Especially, when comparing our ADAR with the sub-optimal method DCLS, ADAR has model parameters about 88% and Flops about 67% less than the DCLS [47], while our ADAR achieves a favorable blind SR performance comparable to DCLS. We also provided the runtime comparison of different ×4 Blind SR models on reconstructing a 1280 × 720 image. As demonstrated in Table 8, our ADAR gains the third fastest model runtime. With a small increase in model runtime, the blind SR performance of ADAR is much higher than the first two fastest methods *i.e.,* DASR and KDSR-M (+1.27 dB and +1.56 dB PSNR value improvement, respectively). We further conduct the experiments to contrast our algorithm with other SOTA Blind SR algorithms in terms of Memory Cost. Besides, we combined this analysis with Flops (Floating Point Operations) to validate the feasibility of deploying our algorithm and other SOTA Blind SR algorithms on AI edge devices. For instance, NVIDIA Jetson Nano stands out as one of the most common AI edge devices, featuring an NVIDIA Maxwell architecture with 128 NVIDIA CUDA cores, 4 GB 64-bit LPDDR4 memory running at 1600MHz with a bandwidth of 25.6 GB/s, and a computational power of 472 GFLOPS (Floating Point Operations per Second). As shown in Table 9, our algorithm achieves high performance with relatively low memory cost (3.17 GB) in the task of ×4 SR on Set5 datasets, meeting the memory requirements of NVIDIA Jetson Nano. In contrast, algorithms such as DANv2, CDCN, and DSSR have memory costs exceeding 4 GB, making them unsuitable for deployment on NVIDIA Jetson Nano. Furthermore, our algorithm exhibits lower computational complexity compared to other SOTA Blind SR algorithms, with only 82.65G Flops. This value is below the computational power supported by NVIDIA Jetson Nano (472 GFLOPS). Therefore, under ideal conditions without considering data communication, data loading, and GPU utilization, our algorithm embedded in NVIDIA Jetson Nano can achieve ×4 super-resolution of Set5 datasets within 1 s. Based on the above analysis, compared to other SOTA Blind SR algorithms, ADAR achieves the highest reconstruction performance with nearly lower Flops and Memory Costs, demonstrating its efficiency and potential for deployment on AI edge devices. To further validate the generalization ability of the proposed ADAR, following [70], the SOTA SR algorithm in the field of remote sensing (RS), we adopt the three common RS datasets, i.e., NWPU45 [71], RSC11 [72], and UC-Merced [73] for validation of the SR performance of our ADAR in the field of RS. The experimental results are demonstrated in the Table 10. It can be illuminated that our ADAR achieves the best and the second-best SR performance on DIV2K+Flickr2K and DIV2K datasets, respectively, compared with the SOTA RS-tailored algorithm and other common SR methods. Concretely, in terms of the RSC11 test dataset, which contains more sophisticated scene categories and higher resolution compared with the other two datasets, with the complex degradation setting, i.e., the degradation kernel width of 3.6, our ADAR gains 1.18 dB PSNR value improvement on the same training datasets against of the suboptimal RS algorithm, RAN. Since there are plenty of similar regions in RS images, the introduced RNLA can play a significant role due to its ability to capture long-range region similarities. Besides, the introduced ADAR can efficiently leverage prior information on the degradation kernel accurately generated by the ADAE, hence assisting the model in reconstructing favorable RS images under complex degradation scenarios. Figs. 9 and 10 demonstrate the subjective visual comparisons on B100, Urban100, Set14, and Manga109 for ×4 SR. Our ADAR generates texture, parallel straight lines, letters, and grids more clearly and accurately than other SOTA SR methods. The subjective visual comparison on DIV2KRK [56] is illustrated in Fig. 11, our ADAR also achieves the clearest and sharpest SR reconstruction results among the listed SR methods. To further validate the model performance of our ADAR, we conducted a subjective visual comparison experiment between our ADAR and other SOTA blind-SR methods on the reconstruction of the real-world images, which are without any degradation assumption and paired Ground-Truth HR images. As demonstrated in

Table 7
Quantitative evaluation for ×4 SR in **Setting2**. The best PSNR/SSIM values is shown in **bold** and the second-best PSNR/SSIM values is underline.

Method	Set5	Set14	B100	Urban100	Manga109	DIV2KRK	#Params (M)	#Flpos (G)
Bicubic	26.33/0.7454	24.62/0.6411	24.85/0.6099	21.95/0.5920	23.17/0.7257	25.33/0.6795	–	–
SRFBN	27.72/0.7907	25.67/0.6834	25.59/0.6465	23.04/0.6530	24.85/0.7841	25.66/0.6945	3.63	7466.13
SAFMN-L	27.73/0.7911	25.66/0.6836	25.59/0.6469	23.02/0.6520	24.83/0.7834	25.66/0.6946	5.60	320.83
IKC	30.12/0.8494	27.29/0.7379	26.67/0.6945	24.57/0.7215	27.71/0.8492	27.26/0.7420	5.32	2528.03
DANv1	29.22/0.8338	26.60/0.7206	25.01/0.6841	24.10/0.7054	26.35/0.8272	27.55/0.7582	4.33	1098.33
DASR	31.04/0.8735	27.65/0.7480	27.03/0.7070	24.61/0.7258	28.43/0.8699	27.49/0.7478	5.84	185.66
DSSR	31.58/0.8836	28.24/0.7657	27.34/0.7220	25.41/0.7603	29.86/0.8957	28.78/0.7905	8.90	5549.74
CDCN	31.65/0.8840	28.28/0.7667	27.39/0.7231	25.63/0.7664	30.03/0.8967	28.92/0.7929	11.70	745.25
DANv2	31.50/0.8817	28.14/0.7620	27.28/0.7190	25.30/0.7570	29.64/0.8909	28.74/0.7893	4.71	1088.14
KDSR-M	31.58/0.8826	28.22/0.7648	27.38/0.7236	25.59/0.7646	29.95/0.8963	27.78/0.7900	5.80	191.42
DCLS	31.68/0.8844	28.34/**0.7691**	27.46/**0.7271**	**25.90/0.7763**	**30.29**/0.9031	28.99/0.7947	19.05	436.57
ADAR (Ours)	**31.77/0.8864**	**28.36**/0.7687	**27.47**/0.7265	25.87/0.7754	**30.29/0.9039**	**29.05/0.7982**	2.37	144.77

Table 8
Trade-off comparison between SR performance and runtime of different SR models in **Setting 2**. SR performance is validated on ×4 DIV2KRK and the best PSNR/SSIM values are **bold**.

Options	IKC	DASR	KDSR-M	DANv1	DSSR	CDCN	DANv2	DCLS	ADAR (Ours)
Performance	27.26/0.7420	27.49/0.7478	27.78/0.7900	27.55/0.7582	28.78/0.7905	28.92/0.7929	28.74/0.7893	28.99/0.7947	**29.05/0.7982**
Runtime	6842.92	410.75	473.07	8108.61	924.88	596.18	1106.35	635.08	561.42

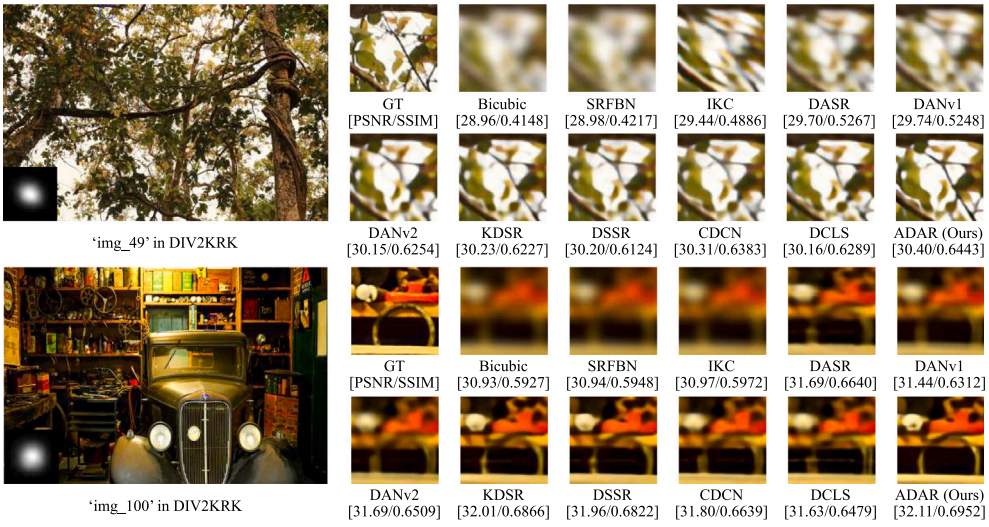


Fig. 11. Qualitative comparison of ADAR with other SOTA methods on ×4 image SR of DIV2KRK in **Setting 2**.

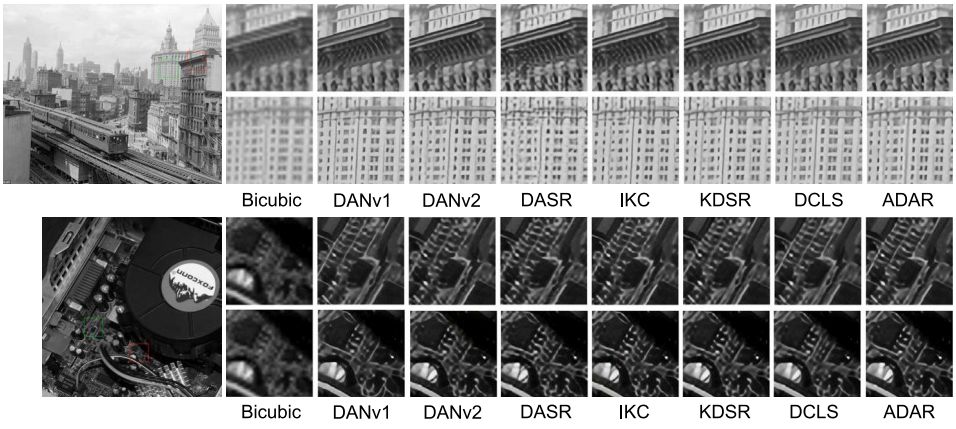


Fig. 12. Qualitative comparison of ADAR with other SOTA methods on real-world images without any degradation assumptions.

Fig. 12, in the upper image, our ADAR accurately generated the correct fence orientation and the number of windows, while other methods exhibited artifacts. In the lower image, compared to other methods, our ADAR produced more obvious chip pins.

Table 9
Trade-off comparison between SR performance and Memory Costs of different SR models. SR performance is validated on ×4 Set5 and the best PSNR/SSIM values are **bold**.

Options	IKC	DASR	KDSR-M	DANv1	DSSR	CDCN	DANv2	DCLS	ADAR (Ours)
Memory Cost (G)	2.63	1.98	1.58	1.60	4.60	4.08	5.58	3.11	3.17
Flops (G)	263.60	113.52	140.08	169.45	706.62	459.38	561.19	269.35	82.65
Performance	30.12/0.8494	31.04/0.8735	31.58/0.8826	29.22/0.8338	31.58/0.8836	31.65/0.8840	31.50/0.8817	31.68/0.8844	**31.77/0.8864**

Table 10
Quantitative evaluation for ×4 SR on three remote sensing datasets. The LR images are degraded by isotropic Gaussian degradation kernels with different kernel width followed by [70]. The best PSNR value is shown in **bold** and the second-best PSNR value is underline.

options	Training datasets	RSC11				UC-Merced				NWPU45			
		0	1.2	2.4	3.6	0	1.2	2.4	3.6	0	1.2	2.4	3.6
Bicubic	DIV2K	26.48	26.45	25.41	24.45	26.69	26.68	24.81	23.48	26.90	26.86	25.52	24.35
RCAN	DIV2K	27.98	26.76	26.11	24.64	**29.14**	27.15	25.81	23.76	28.68	27.33	26.27	24.52
IKC	DIV2K	26.56	26.63	26.54	25.87	27.56	27.60	27.42	26.41	27.25	27.29	27.15	26.42
DASR	DIV2K	27.16	26.87	26.06	25.88	27.73	27.71	27.43	26.43	27.55	27.27	27.16	26.51
RAN	DIV2K	27.96	27.71	27.59	26.95	28.43	28.38	28.27	27.48	28.29	28.33	28.23	27.48
ADAR (Ours)	DIV2K	28.79	28.86	28.79	28.13	28.79	28.81	28.67	27.80	29.02	29.06	28.95	28.19
ADAR (Ours)	DIV2K+Flickr2K	**29.08**	**29.14**	**29.00**	**28.23**	**29.14**	**29.13**	**28.92**	**27.98**	**29.21**	**29.23**	**29.08**	**28.27**

5. Conclusion

In this paper, an adaptive degradation-aware estimation algorithm is introduced for efficient blind super-resolution. Concretely, we propose a lightweight Region Non-Local Attention (RNLA) to successively build up pixel-to-region and region-wise feature correlation in two loops. Compared with traditional Non-Local attention mechanism for global feature accumulation, our RNLA tremendously economizes computational resource consumption. Based on RNLA, we introduce the Region Non-Local Feature Block (RNLFB) that play the role of feature extractor for both degradation and feature representation learning. The number of RNLFBs is determined by the difficulty of the feature extraction task to save the computational resources. Based on the degeneration representation, the introduced kernel projector can accurately estimate degeneration kernel as the prior information for blind SR reconstruction. Extensive experiments on both synthetic and real-world datasets demonstrate that our method achieves faster kernel estimation and blind SR reconstruction with negligible SR performance loss compared with the SOTA blind SR methods.

CRediT authorship contribution statement

Haoran Yang: Writing – review & editing, Writing – original draft, Validation, Methodology. **Qilei Li:** Visualization. **Bin Meng:** Validation. **Gwanggil Jeon:** Formal analysis. **Kai Liu:** Investigation. **Xiaomin Yang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The research in our paper is sponsored by National Press and Publication Administration of China (No. UHD-ZD-202306) and Science Foundation of Sichuan Science and Technology Department (No. 24NSFSC2159).

References

[1] Z. Chen, L. Yang, J.-H. Lai, X. Xie, CuNeRF: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 21185–21195.
[2] M.-I. Georgescu, R.T. Ionescu, A.-I. Miron, O. Savencu, N.-C. Ristea, N. Verga, F.S. Khan, Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 2195–2205.
[3] Y. Wang, W. Liu, W. Sun, X. Meng, G. Yang, K. Ren, A progressive feature enhancement deep network for large-scale remote sensing image super-resolution, IEEE Trans. Geosci. Remote Sens. 61 (2023).
[4] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, L. Zhang, EDiffSR: An efficient diffusion probabilistic model for remote sensing image super-resolution, IEEE Trans. Geosci. Remote Sens. 62 (2023).
[5] P. Wang, E. Sertel, Multi-frame super-resolution of remote sensing images using attention-based GAN models, Knowl.-Based Syst. 266 (2023) 110387.
[6] A. Kabiraj, D. Pal, D. Ganguly, K. Chatterjee, S. Roy, Number plate recognition from enhanced super-resolution using generative adversarial network, Multimedia Tools Appl. 82 (9) (2023) 13837–13853.
[7] C. Dong, C.C. Loy, K. He, X. Tang, 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, Springer, 2014, pp. 184–199.
[8] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.
[9] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.
[10] Y. Yang, Y. Qi, Hierarchical accumulation network with grid attention for image super-resolution, Knowl.-Based Syst. 233 (2021) 107520.
[11] B. Xia, Y. Hang, Y. Tian, W. Yang, Q. Liao, J. Zhou, Efficient non-local contrastive attention for image super-resolution, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 3, 2022, pp. 2759–2767.
[12] X. Chen, X. Wang, J. Zhou, Y. Qiao, C. Dong, Activating more pixels in image super-resolution transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22367–22377.
[13] C. Liu, P. Lei, An efficient group skip-connecting network for image super-resolution, Knowl.-Based Syst. 222 (2021) 107017.
[14] F. Kong, M. Li, S. Liu, D. Liu, J. He, Y. Bai, F. Chen, L. Fu, Residual local feature network for efficient super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 766–776.
[15] H. Wang, X. Chen, B. Ni, Y. Liu, J. Liu, Omni aggregation networks for lightweight image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22378–22387.
[16] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, W. Wu, Feedback network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3867–3876.
[17] L. Sun, J. Dong, J. Tang, J. Pan, Spatially-adaptive feature modulation for efficient image super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.
[18] J. Gu, H. Lu, W. Zuo, C. Dong, Blind super-resolution with iterative kernel correction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1604–1613.

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

[20] Y. Zhang, L. Dong, H. Yang, L. Qing, X. He, H. Chen, Weakly-supervised contrastive learning-based implicit degradation modeling for blind image super-resolution, Knowl.-Based Syst. 249 (2022) 108984.

[21] Y. Zhou, C. Lin, D. Luo, Y. Liu, Y. Tai, C. Wang, M. Chen, Joint learning content and degradation aware feature for blind super-resolution, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 2606–2616.

[22] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, S.-M. Hu, Visual attention network, Comput. Vis. Media 9 (4) (2023) 733–752.

[23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481.

[24] J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual feature aggregation network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2359–2368.

[25] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, H. Shen, Single image super-resolution via a holistic attention network, in: European Conference on Computer Vision, Springer, 2020, pp. 191–207.

[26] Z. Hui, X. Wang, X. Gao, Fast and accurate single image super-resolution via information distillation network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 723–731.

[27] J. Liu, J. Tang, G. Wu, Residual feature distillation network for lightweight image super-resolution, in: European Conference on Computer Vision, Springer, 2020, pp. 41–55.

[28] Z. Hui, X. Gao, Y. Yang, X. Wang, Lightweight image super-resolution with information multi-distillation network, in: Proceedings of the 27th Acm International Conference on Multimedia, 2019, pp. 2024–2032.

[29] Y. Hu, Y. Huang, K. Zhang, Multi-scale information distillation network for efficient image super-resolution, Knowl.-Based Syst. (2023) 110718.

[30] N. Ahn, B. Kang, K.-A. Sohn, Fast, accurate, and lightweight super-resolution with cascading residual network, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 252–268.

[31] C. Tian, R. Zhuge, Z. Wu, Y. Xu, W. Zuo, C. Chen, C.-W. Lin, Lightweight image super-resolution with enhanced CNN, Knowl.-Based Syst. 205 (2020) 106235.

[32] Y. Cai, H. Lai, Z. Jia, L. Wang, Lightweight spatial-channel adaptive coordination of multilevel refinement enhancement network for image reconstruction, Knowl.-Based Syst. 256 (2022) 109824.

[33] K. Zhang, X. Gao, J. Li, H. Xia, Single image super-resolution using regularization of non-local steering kernel regression, Signal Process. 123 (2016) 53–63.

[34] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T.S. Huang, H. Shi, Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5690–5699.

[35] X. Wu, K. Zhang, Y. Hu, X. He, X. Gao, Multi-scale non-local attention network for image super-resolution, Signal Process. 218 (2024) 109362.

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

[37] H. Choi, J. Lee, J. Yang, N-gram in swin transformers for efficient lightweight image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 2071–2081.

[38] K. Zhang, W. Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3262–3271.

[39] V. Cornillere, A. Djelouah, W. Yifan, O. Sorkine-Hornung, C. Schroers, Blind image super-resolution with spatially variant degradations, ACM Trans. Graph. 38 (6) (2019) 1–13.

[40] A. Shocher, N. Cohen, M. Irani, "zero-shot" super-resolution using deep internal learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3118–3126.

[41] R. Zhou, S. Susstrunk, Kernel modeling super-resolution on real low-resolution images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2433–2443.

[42] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, F. Huang, Real-world super-resolution via kernel estimation and noise injection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 466–467.

[43] Y. Huang, S. Li, L. Wang, T. Tan, et al., Unfolding the alternating optimization for blind super resolution, Adv. Neural Inf. Process. Syst. 33 (2020) 5632–5643.

[44] Y.-S. Xu, S.-Y.R. Tseng, Y. Tseng, H.-K. Kuo, Y.-M. Tsai, Unified dynamic convolutional network for super-resolution with variational degradations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12496–12505.

[45] F. Li, Y. Wu, H. Bai, W. Lin, R. Cong, C. Zhang, Y. Zhao, Learning detail-structure alternative optimization for blind super-resolution, IEEE Trans. Multimed. (2022).

[46] Z. He, Z. Jin, Y. Zhao, Srdrl: A blind super-resolution framework with degradation reconstruction loss, IEEE Trans. Multimed. 24 (2021) 2877–2889.

[47] Z. Luo, H. Huang, L. Yu, Y. Li, H. Fan, S. Liu, Deep constrained least squares for blind image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17642–17652.

[48] E. Agustsson, R. Timofte, Ntire 2017 challenge on single image super-resolution: Dataset and study, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 126–135.

[49] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, Ntire 2017 challenge on single image super-resolution: Methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 114–125.

[50] M. Bevilacqua, A. Roumy, C. Guillemot, M.L. Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: British Machine Vision Conference, 2012.

[51] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: International Conference on Curves and Surfaces, 2010.

[52] D. Martin, C. Fowlkes, D. Tal, J. Malik, 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, Vol. 2, ICCV 2001, IEEE, 2001, pp. 416–423.

[53] J.-B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5197–5206.

[54] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, K. Aizawa, Sketch-based manga retrieval using manga109 dataset, MTAP 76 (20) (2017) 21811–21838.

[55] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[56] S. Bell-Kligler, A. Shocher, M. Irani, Blind super-resolution kernel estimation using an internal-gan, Adv. Neural Inf. Process. Syst. 32 (2019).

[57] M. Emad, M. Peemen, H. Corporaal, Moesr: blind super-resolution using kernel-aware mixture of experts, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3408–3417.

[58] Y. Wu, F. Li, H. Bai, W. Lin, R. Cong, Y. Zhao, Bridging component learning with degradation modelling for blind image super-resolution, IEEE Trans. Multimed. (2022).

[59] Y. Gandelsman, A. Shocher, M. Irani, "Double-DIP": unsupervised image decomposition via coupled deep-image-priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11026–11035.

[60] J. Liang, K. Zhang, S. Gu, L. Van Gool, R. Timofte, Flow-based kernel prior with application to blind super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10601–10610.

[61] Z. Luo, Y. Huang, S. Li, L. Wang, T. Tan, End-to-end alternating optimization for blind super resolution, 2021, arXiv preprint arXiv:2105.06878.

[62] I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts, 2016, arXiv preprint arXiv:1608.03983.

[63] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

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

[65] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 286–301.

[66] H. Feng, L. Wang, Y. Li, A. Du, LKASR: Large kernel attention for lightweight image super-resolution, Knowl.-Based Syst. 252 (2022) 109376.

[67] Y. Li, K. Zhang, R. Timofte, L. Van Gool, et al., NTIRE 2022 challenge on efficient super-resolution: Methods and results, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022.

[68] B. Xia, Y. Zhang, Y. Wang, Y. Tian, W. Yang, R. Timofte, L. Van Gool, Knowledge distillation based degradation estimation for blind super-resolution, in: ICLR, 2023.

[69] Y. Jo, S.W. Oh, P. Vajda, S.J. Kim, Tackling the ill-posedness of super-resolution through adaptive target generation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16236–16245.

[70] B. Liu, L. Zhao, S. Shao, W. Liu, D. Tao, W. Cao, Y. Zhou, RAN: Region-aware network for remote sensing image super-resolution, IEEE Trans. Geosci. Remote Sens. (2023).

[71] G. Sheng, W. Yang, T. Xu, H. Sun, High-resolution satellite scene classification using a sparse coding based multiple feature combination, Int. J. Remote Sens. 33 (8) (2012) 2395–2412.

[72] L. Zhao, P. Tang, L. Huo, Feature significance-based multibag-of-visual-words model for remote sensing image scene classification, J. Appl. Remote Sens. 10 (3) (2016) 035004.

[73] Y. Yang, S. Newsam, Bag-of-visual-words and spatial extensions for land-use classification, in: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010, pp. 270–279.