Contrastive Degradation-Aware Mamba Network for Blind Image Super-Resolution*

Guangzi Ye[1], Gang Ke[1*]

[1]School of Electronic Information, Dongguan Polytechnic, Dongguan 523808, China
[*]Corresponding author: kegang95@126.com

Abstract: Blind image super-resolution (BISR) aims to recover high-resolution images from low-resolution inputs with unknown degradations. Existing degradation representation learning methods usually construct positive pairs from local patches of the same degraded image, which inevitably introduces strong content correlations and limits the modeling capability for complex degradation patterns. To address this issue, we propose a Contrastive Degradation-Aware Mamba Network (CDAM-Net) for blind image super-resolution. Specifically, we design a contrastive degradation representation learning strategy that generates positive pairs from different images with identical degradation parameters, enabling the network to learn content-independent degradation representations. Furthermore, a degradation-aware Mamba restoration network is proposed to adaptively incorporate degradation information into the image reconstruction process. In particular, we introduce a degradation-guided feature modulation mechanism within the Spatial Mamba Module to dynamically guide feature restoration under different degradation conditions. In addition, a gated convolution feed-forward block is utilized to enhance local nonlinear feature representation and texture recovery capability. Extensive experiments on multiple benchmark datasets demonstrate that the proposed method achieves competitive performance against state-of-the-art blind image super-resolution methods in both quantitative and visual comparisons.

Keywords: Blind image super-resolution;degradation representation learning; contrastive learning;Mamba, image restoration.

I Introduction

Single image super-resolution (SISR) aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) observation. With the development of deep learning, convolutional neural network (CNN)-based methods have achieved remarkable progress in image super-resolution. SRCNN first introduced an end-to-end CNN framework for SISR [1]. Later, deeper and more powerful networks, such as VDSR [2], EDSR [3], and RCAN [4], further improved reconstruction performance through residual learning and attention mechanisms.

However, most traditional SISR methods are trained under predefined degradation assumptions, such as bicubic downsampling. In real-world scenarios, LR images usually suffer from unknown and complex degradations, including blur, noise, compression artifacts, and sensor distortions. The mismatch between synthetic degradation and real degradation often leads to severe performance degradation. Therefore, blind image super-resolution (Blind SR), which aims to restore HR images from LR inputs with unknown degradation, has attracted increasing attention.

Existing Blind SR methods usually focus on degradation estimation or degradation representation learning. Some methods explicitly estimate degradation kernels to guide the SR process. For example, IKC iteratively corrects the estimated blur kernel for blind SR [5], while

DAN formulates degradation estimation and image restoration in an end-to-end alternating optimization framework [6]. Although these methods provide interpretable degradation priors, their performance may be limited when real-world degradations cannot be accurately represented by simple blur kernels.

Recently, implicit degradation representation learning has become an effective solution for Blind SR. DASR introduces contrastive learning to learn degradation representations from LR images without explicit kernel supervision [7]. AirNet further demonstrates the effectiveness of contrastive degradation representation learning for unknown image corruption restoration [8]. Although these methods achieve promising performance, most existing methods construct positive pairs from local patches of the same degraded image. Such strategies inevitably introduce strong content correlations into degradation representation learning, which limits the discrimination capability for complex degradation patterns.

To address this issue, we propose a content-decoupled contrastive degradation representation learning strategy for Blind SR. Specifically, positive pairs are constructed from different HR images degraded with identical degradation parameters, enabling the network to focus on degradation characteristics rather than image content. In this way, the proposed method can effectively suppress content interference and learn more discriminative degradation-aware representations.

In addition, existing Mamba-based image restoration methods mainly focus on general low-level vision tasks and pay limited attention to degradation-guided feature restoration for Blind SR. Therefore, we further introduce a degradation-aware Mamba restoration network, where degradation representations are adaptively injected into the Spatial Mamba Module through degradation-guided feature modulation. By combining degradation-aware modulation with efficient state-space modeling, the proposed method can better handle complex unknown degradations while maintaining computational efficiency.

The main contributions of this paper are summarized as follows:

• We propose a content-decoupled contrastive degradation representation learning strategy for blind image super-resolution. Different from previous methods that construct positive pairs from local patches of the same image, the proposed method generates positive pairs from different images with identical degradation parameters, which effectively reduces content interference during degradation representation learning.

• We design a degradation-aware Mamba restoration network, where degradation representations are adaptively injected into the Spatial Mamba Module through a degradation-guided feature modulation mechanism for adaptive image reconstruction.

• Extensive experiments demonstrate that the proposed method achieves competitive performance against state-of-the-art blind image super-resolution methods on multiple benchmark datasets.

II Related Work

A. Single Image Super-Resolution

Single image super-resolution has been widely studied in low-level vision. Early deep-learning-based methods mainly used CNNs to learn the mapping from LR images to HR images. SRCNN is one of the first works to apply deep CNNs to SISR [1]. VDSR further increases network depth and adopts residual learning to improve reconstruction accuracy [2]. EDSR removes unnecessary modules from residual networks and enlarges model capacity for better SR performance [3]. RCAN introduces channel attention and residual-in-residual structures to

enhance feature representation [4].

Although these methods achieve impressive results, they usually assume a fixed degradation model. When applied to real-world images with unknown degradations, their performance often drops significantly. This limitation motivates the development of blind image super-resolution methods.

B. Blind Image Super-Resolution

Blind image super-resolution aims to recover HR images from LR inputs with unknown degradation. Existing methods can be roughly divided into explicit degradation estimation methods and implicit degradation representation learning methods.

Explicit degradation estimation methods attempt to estimate degradation kernels and use them as priors for SR reconstruction. IKC proposes an iterative kernel correction strategy to refine the estimated blur kernel [5]. DAN unfolds the alternating optimization process of kernel estimation and image restoration into an end-to-end trainable network [6]. Real-ESRGAN and BSRGAN design more practical degradation models to synthesize realistic LR-HR training pairs for real-world blind SR [9, 10].

Different from explicit kernel estimation, implicit degradation representation learning methods aim to learn degradation-aware features directly from LR images. DASR introduces contrastive learning to distinguish different degradations in the latent representation space [7]. AirNet uses a contrastive degradation encoder and a degradation-guided restoration network to handle unknown image corruptions [8]. These methods show that degradation representations can effectively guide image restoration. However, effective degradation-aware fusion in modern restoration backbones remains underexplored.

C. State Space Models for Image Restoration

Transformer-based models have shown strong capability in image restoration due to their ability to model long-range dependencies. SwinIR adopts the Swin Transformer as a strong image restoration backbone and achieves promising performance [11]. However, self-attention-based models usually suffer from high computational complexity, especially for high-resolution image restoration.

State space models provide an efficient alternative for long-range dependency modeling. Mamba introduces selective state space modeling and achieves linear-time sequence modeling [12]. Inspired by its efficiency and global modeling ability, MambaIR applies Mamba to image restoration and improves local feature enhancement for low-level vision tasks [13]. However, most existing Mamba-based restoration methods are not specifically designed for blind SR. In this work, we introduce degradation-aware feature modulation into a Mamba-based restoration network to improve reconstruction under unknown degradations.

III Proposed Method

In this section, we present the proposed Contrastive Degradation-Aware Mamba Network (CDAM-Net) for blind image super-resolution. As illustrated in Fig. 1, the proposed framework mainly consists of two components: 1) a contrastive degradation representation learning module for extracting degradation-aware representations, and 2) a degradation-aware Mamba restoration network for adaptive image reconstruction.

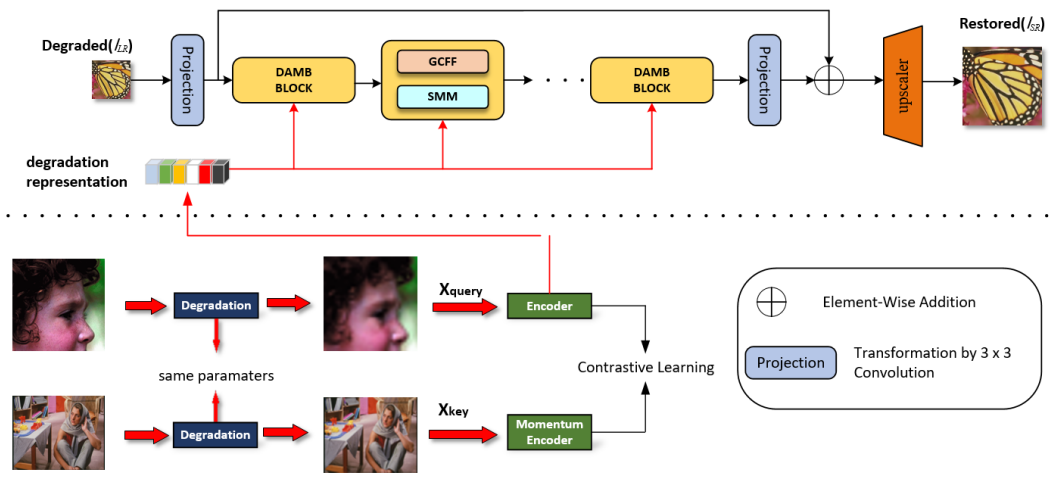

Figure 1. Overall architecture of the proposed CDAM-Net. The framework mainly consists of a contrastive degradation representation learning module and a degradation-aware Mamba restoration network.

A. Overall Architecture

Blind image super-resolution aims to recover a high-resolution (HR) image from a degraded low-resolution (LR) observation with unknown degradation. Existing methods usually estimate degradation information from local image patches, which often suffer from content interference and limited degradation modeling capability. To address this issue, we propose a contrastive degradation representation learning strategy to obtain content-independent degradation representations.

Given a degraded image ILR, the proposed network first extracts shallow features through a projection layer. The extracted features are then fed into multiple degradation-aware Mamba blocks (DAMBs) for progressive restoration. Meanwhile, a degradation representation encoder is introduced to learn degradation-aware embeddings through contrastive learning. The learned degradation representation is further injected into the restoration process through a degradation-guided feature modulation mechanism. Finally, the reconstructed HR image ISR is generated by the upsampling reconstruction module.

B. Contrastive Degradation Representation Learning

Existing degradation representation learning methods commonly construct positive pairs from cropped patches of the same degraded image. Although such strategies reduce degradation inconsistency, the learned representations may still contain strong content correlations, which limits the modeling capability for complex degradation patterns.

To alleviate this issue, we propose a contrastive degradation representation learning strategy based on degradation-consistent image pairs. Specifically, two different HR images are degraded using identical degradation parameters to generate positive pairs with different content but consistent degradation characteristics. In this way, the encoder is encouraged to focus on degradation information instead of image content.

Given two HR images $I_{HR}^{a}$and I_{HR}^{b}, we apply the same degradation operator $\mathcal{D}(\cdot)$ to obtain:

$$I_{LR}^{a} = \mathcal{D}(I_{HR}^{a}) \quad (1)$$

$$I_{LR}^{b} = \mathcal{D}(I_{HR}^{b}) \quad (2)$$

The degraded images are then fed into the degradation encoder to extract degradation representations:

$$z_a = E(I_{LR}^{a}), z_b = E(I_{LR}^{b}) \quad (3)$$

where E(·) denotes the degradation encoder.

To improve degradation discrimination ca-

pability, we employ the InfoNCE loss for contrastive optimization:

$$\mathcal{L}_{cl} = -log\frac{exp(sim(z_a,z_b)/\tau)}{\sum_{i=1}^{N}exp(sim(z_a,z_i)/\tau)} \quad (4)$$

where sim(·) denotes cosine similarity and τ is the temperature parameter.

Through the proposed contrastive learning strategy, the network can effectively learn degradation-aware representations while suppressing content interference.

C. Degradation-Aware Mamba Block

To effectively utilize degradation representations for blind image restoration, we propose a degradation-aware Mamba block (DAMB), which mainly consists of a Spatial Mamba Module (SMM) and a Gated Convolution Feed-Forward block (GCFF). In addition, a degradation-guided feature modulation (DGFM) mechanism is embedded into the SMM to adaptively inject degradation representations into the restoration process.

1) Spatial Mamba Module

The Spatial Mamba Module is designed to capture long-range spatial dependencies for image restoration. Given an input feature $F_x \in \mathbb{R}^{H\times W\times C}$,we first apply layer normalization and point-wise convolution to project the feature into a high-dimensional representation space. Subsequently, depth-wise convolution is employed to enhance local spatial interaction.

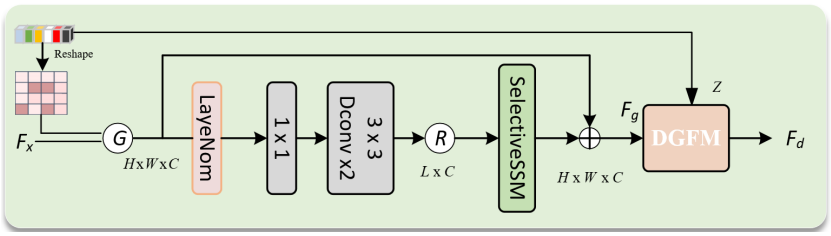

Figure 2. Structure of the proposed Spatial Mamba Module (SMM).

Following previous Mamba-based vision architectures, the spatial feature is reshaped into a sequence representation and fed into the selective state-space model (SelectiveSSM) for global dependency aggregation:

$$F_s = SSM(F_x) \quad (5)$$

where SSM(·) denotes the selective state-space modeling operation.

Compared with conventional convolution operations, the proposed SMM can effectively model long-range contextual information with linear computational complexity, which is beneficial for recovering degraded structures and textures in blind super-resolution tasks.

2) Degradation-Guided Feature Modulation

To adaptively incorporate degradation information into the restoration process, we introduce a degradation-guided feature modulation (DGFM) mechanism within the Spatial Mamba Module.

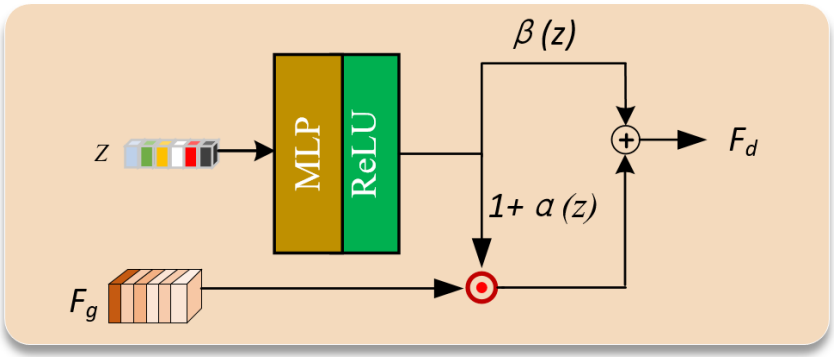

Figure 3. Structure of the proposed Degradation-Guided Feature Modulation (DGFM) module.

Specifically, the degradation representation z is first projected through a multilayer perceptron (MLP) to generate channel-wise modulation parameters α(z) and β(z):

$$[\alpha(z), \beta(z)] = MLP(z) \quad (6)$$

The restoration feature is then adaptively modulated as:

$$F_d = (1 + \alpha(z)) \odot F_g + \beta(z) \quad (7)$$

where ⊙ denotes element-wise multiplication, Fg represents the input restoration feature, and Fd denotes the modulated feature.

Through the proposed DGFM mechanism, the learned degradation representation can dynamically guide the restoration process under different degradation conditions, thereby improving degradation adaptability and reconstruction quality.

3) Gated Convolution Feed-Forward Block

Although the Spatial Mamba Module can effectively capture global contextual dependencies, local nonlinear feature refinement is still essential for recovering fine textures and structural details. Therefore, we further introduce a Gated Convolution Feed-Forward block (GCFF) for local feature enhancement.

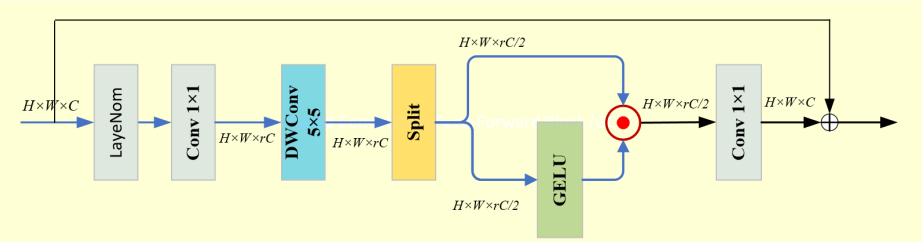

Figure 4. Structure of the proposed Gated Convolution Feed-Forward block (GCFF).

Specifically, the input feature is first expanded through point-wise convolution and depth-wise convolution. The expanded feature is then split into two branches, where one branch is activated by the GELU function and subsequently multiplied with the other branch through element-wise gating operation:

$$F_{gated} = F_1 \odot GELU(F_2) \quad (8)$$

where F1 and F2 denote the split feature branches.

Finally, a point-wise convolution is utilized to project the refined feature back to the original feature dimension. The proposed GCFF can effectively enhance local feature interaction and nonlinear representation capability, thereby improving texture restoration performance in blind image super-resolution.

D. Loss Function

The overall loss function consists of the reconstruction loss and the contrastive degradation learning loss:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{cl} \quad (9)$$

where λ denotes the balancing coefficient.

For image reconstruction, we employ the ℓ_1 loss between the reconstructed image ISR and the ground-truth HR image IHR:

$$\mathcal{L}_{rec} = \|I_{SR} - I_{HR}\|_1 \quad (10)$$

IV Experiments

A. Experimental Settings

1) Datasets

Following previous blind image super-resolution methods, we conduct experiments on both synthetic degradation datasets and real-world degradation datasets. For training, DIV2K and Flickr2K are adopted as high-resolution image sources. The low-resolution images are generated by applying random blur kernels, bicubic downsampling, Gaussian noise, and JPEG compression. For evaluation, we use five widely used benchmark datasets, including Set5, Set14, BSD100, Urban100, and Manga109.

2) Implementation Details

The proposed method is implemented using PyTorch. During training, the input low-resolution patches are randomly cropped with a size of 64 × 64. The batch size is set to 16. Adam optimizer is adopted with $\beta1=0.9$ and $\beta2=0.999$. The initial learning rate is initialized as 2×10^{-4} and gradually decayed using cosine annealing. The number of degradation-aware Mamba blocks (DAMBs) is set to 6, and the feature channel dimension is set to 64. The balancing coefficient of the contrastive learning loss is empirically set to $\lambda=0.1$.

3) Evaluation Metrics

We evaluate the reconstruction quality using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). Following previous works, PSNR and SSIM are calculated on the Y channel of the YCbCr color space.

B. Comparison with State-of-the-Art Methods

We compare the proposed method with several representative blind image super-resolution methods, including Bicubic, IKC [5], DANv2 [6], CDCN [15], and other representative degradation-aware restoration methods. The quantitative comparison results are reported in Table 1.

Table 1. Quantitative comparison of different blind super-resolution methods on synthetic degradation datasets. The best results are highlighted in bold.

Method	Scale	PSNR(dB)/SSIM on Synthetic Degradation Datasets				
		Set5	Set14	BSD100	Urban100	Manga109
Bicubic	×3	26.96 / 0.7848	24.54 / 0.6800	24.91 / 0.6584	21.97 / 0.6510	23.78 / 0.7783
IKC [5]	×3	33.35 / 0.9177	29.66 / 0.8281	28.69 / 0.7929	27.55 / 0.8342	32.47 / 0.9333
DANv2 [6]	×3	34.31 / 0.9238	30.29 / 0.8356	29.09 / 0.7992	28.04 / 0.8459	33.48 / 0.9426
CDCN [15]	×3	34.29 / 0.9238	30.28 / 0.8356	29.10 / 0.7996	28.11 / 0.8471	33.47 / 0.9424
Ours	×3	34.36 / 0.9241	30.33 / 0.8360	29.12 / 0.7998	28.15 / 0.8480	33.54 / 0.9430
Bicubic	×4	24.39 / 0.6869	24.30 / 0.6325	23.38 / 0.5671	20.34 / 0.5478	21.30 / 0.6834
IKC [5]	×4	31.67 / 0.8829	28.31 / 0.7643	27.34 / 0.7192	25.33 / 0.7504	28.91 / 0.8782
DANv2 [6]	×4	32.00 / 0.8885	28.50 / 0.7715	27.56 / 0.7277	25.94 / 0.7748	30.45 / 0.9037
CDCN [15]	×4	32.04 / 0.8883	28.47 / 0.7695	27.52 / 0.7257	25.92 / 0.7731	30.55 / 0.9036
Ours	×4	32.08 / 0.8890	28.54 / 0.7720	27.59 / 0.7283	26.01 / 0.7762	30.61 / 0.9050

As shown in Table 1, the proposed method achieves superior reconstruction performance on most benchmark datasets. Benefiting from the proposed content-decoupled degradation representation learning and degradation-guided Mamba restoration, our method can more effectively recover structural details and suppress degradation artifacts under complex degradation conditions.

In particular, the proposed method achieves better performance on Urban100 and Manga109, which contain abundant repetitive structures and high-frequency textures. This demonstrates the effectiveness of degradation-guided state-space modeling for blind image super-resolution.

C. Ablation Study

To verify the effectiveness of different components in the proposed framework, we conduct ablation studies on the contrastive degradation representation learning strategy and the degradation-guided feature modulation mechanism. The results are shown in Table 2.

Table 2. Ablation study of different proposed components.

Baseline	Same-image CL	Proposed CDRL	DGFM	PSNR
√				**31.86**
√	√			32.01
√		√		32.12
√		√	√	32.21

As shown in Table 2, introducing conventional same-image contrastive learning improves the reconstruction performance compared with the baseline model. However, the proposed content-decoupled contrastive degradation representation learning strategy achieves further improvement, demonstrating that degradation-consistent positive pairs can effectively reduce content interference during degradation representation learning.

Moreover, the proposed degradation-guided feature modulation (DGFM) further enhances restoration performance by adaptively incorporating degradation representations into the Mamba restoration process.

D. Visual Comparison

Fig. 5 presents visual comparison results on synthetic degradation datasets.

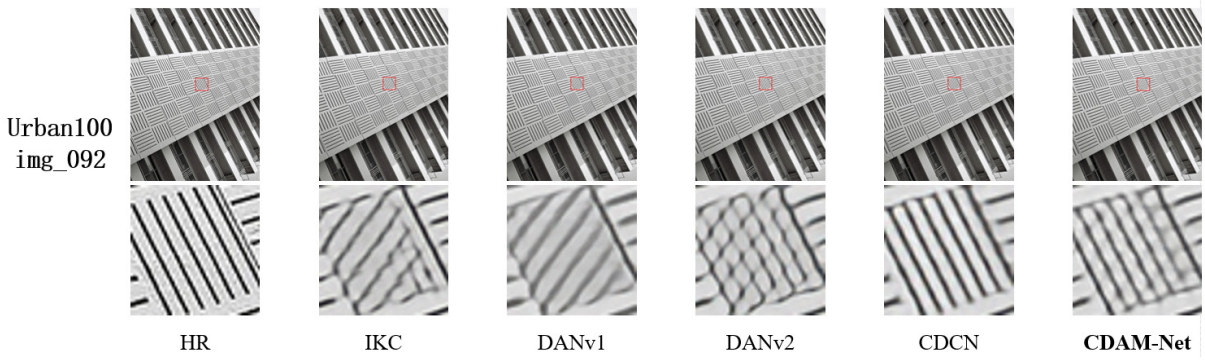

Figure 5. Visual comparison with state-of-the-art blind image super-resolution methods.

Compared with existing methods, the proposed method recovers sharper edges and more faithful texture details. In particular, for images with repetitive structures and complex textures, such as buildings and manga images, our method can better preserve structural consistency and suppress restoration artifacts. This further demonstrates the effectiveness of the proposed degradation-aware state-space restoration mechanism.

V Conclusion

In this paper, we propose a Contrastive Degradation-Aware Mamba Network (CDAM-Net) for blind image super-resolution. Different from existing degradation representation learning methods that construct positive pairs from the same degraded image, the proposed method introduces a content-decoupled contrastive degradation learning strategy by generating positive pairs from different images with identical degradation parameters.

Furthermore, a degradation-guided Mamba restoration network is designed to adaptively incorporate degradation representations into the image restoration process through degradation-guided feature modulation. Extensive experiments demonstrate that the proposed method achieves competitive reconstruction performance and effectively handles complex unknown degradations.

VI Acknowledgment

This research was funded by the Scientific Research Fund of Dongguan Polytechnic (2025a23), and the 2026 Annual Research Platform Project of Institutions of Higher Education, Department of Education of Guangdong Province.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2014, pp. 184-199.

[2] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 1646-1654.

[3] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017, pp. 136-144.

[4] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 286-301.

[5]J. Gu, H. Lu, W. Zuo, and C. Dong, “Blind super-resolution with iterative kernel correction,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 1604-1613.

[6] Z. Luo, Y. Huang, S. Li, L. Wang, and T. Tan, “End-to-end alternating optimization for blind super resolution,” in Proc. Int. Joint Conf. Artif. Intell. (IJCAI), 2021, pp. 876-883.

[7] L. Wang, Y. Wang, X. Dong, Q. Xu, J. Yang, W. An, and Y. Guo, “Unsupervised degradation representation learning for blind super-resolution,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 10581-10590.

[8] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, “All-in-one image restoration for unknown corruption,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 17452-17462.

[9] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), 2021, pp. 1905-1914.

[10] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, “Designing a practical degradation model for deep blind image super-resolution,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 4791-4800.

[11] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “SwinIR: Image restoration using Swin Transformer,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), 2021, pp. 1833-1844.

[12] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” arXiv preprint arXiv:2312.00752, 2023.

[13] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S. T. Xia, “MambaIR: A simple baseline for image restoration with state-space model,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2024.

[14] Z. Luo, H. Huang, L. Yu, Y. Li, H. Fan, and S. Liu, “Deep constrained least squares for blind image super-resolution,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 17642-17652.

[15]Y. Wu, Z. Li, Y. Wang, H. Liu, and X. Zhang, “Bridging component learning with degradation modelling for blind image super-resolution,” IEEE Trans. Multimedia, 2022.