

A Review of Image Restoration Based on Deep Learning

Hongze Zuo *

College of Informatics, Huazhong Agricultural University, Wuhan, China

* Corresponding Author Email: 2022317220611@webmail.hzau.edu.cn

Abstract. Image restoration is a key task in the field of computer vision, which aims to recover high-quality clear images from degraded images. Degraded images may be produced for various reasons, such as noise, blurring, compression, etc. In recent years, the rapid development of deep learning technology has brought new breakthroughs in the field of image restoration, and various image restoration methods based on deep learning continue to emerge, and their performance is better than the previous GAN-based methods. Nevertheless, comprehensive and enlightening reviews of image restoration based on diffusion models are still scarce. This paper summarizes the latest research progress in image restoration in recent years, including cue learning, model scaling, generative prior, state space model, basic model and so on, and compares and analyzes their experimental results. Although the image restoration methods based on deep learning have achieved remarkable progress, there are still some challenges. Therefore, this paper analyzes the challenges in the field of image restoration and looks forward to the future research direction.

Keywords: Image Restoration; Deep Learning; Cue Learning; State Space Model.

1. Introduction

As one of the core tasks in the field of computer vision, image inpainting is committed to reconstructing high-quality visual content from degraded observations. Its technical value has deeply penetrated into key fields such as remote sensing monitoring, medical diagnosis, and cultural heritage protection. With the increasing complexity of imaging environment, the degradation form gradually evolves from single noise and blur to multi-factor coupling (such as coexistence of rain and fog, motion blur and low light interleaved), which poses a severe challenge to traditional inpainting methods. Although early methods based on filters and sparse coding are effective in specific scenarios, they are limited by the ability of hand-crafted feature representation and are difficult to deal with complex degradation patterns in real scenes. In recent years, deep learning technology has promoted the field of image inpainting into a new stage of development with its powerful data-driven characteristics, forming a diversified technical system with convolutional neural Network (CNN), Generative adversarial network (GAN), Transformer and diffusion model (DM) as the core. Significant breakthroughs have been made in model architecture innovation, prior knowledge fusion, and computational efficiency optimization.

Early studies mainly focus on CNN architecture design, which builds end-to-end mapping relationships by stacking convolutional layers. For example, SRCNN first introduces the three-convolutional layer structure into super-resolution tasks, but its limited receptive field leads to insufficient detail recovery ability [1]. Subsequently, mechanisms such as Residual Learning (EDSR) and Dense connection (DenseNet) were introduced [2][3], which significantly improved the feature reuse efficiency of deep networks. However, the local inductive bias property of CNNs makes them a bottleneck in modeling long-range dependencies, which drives researchers to turn to Transformer architectures with stronger global modeling capabilities. The IPT model realizes cross-region feature association in denoising tasks by introducing a multi-head self-attention mechanism, but its quadratic computational complexity restricts the efficiency of high-resolution image processing [4]. To balance performance and computational cost, hybrid architectures (such as SwinIR) have been proposed to reduce the computational load through the local window attention mechanism [5].

The rise of generative models has further expanded the technical boundaries of image inpainting. GAN generates high perceptual quality images through adversarial training. For example, ESRGAN uses relative discriminator to enhance texture authenticity, but its mode collapse problem easily leads to structural distortion [6]. With the progressive denoising process, the diffusion model achieves a better balance between fidelity and generation quality. For example, DiffBIR significantly improves the robustness of blind inpainting tasks through two-stage degradation peeling and detail reconstruction [7]. It is worth noting that the integration of multi-modal technologies is becoming a new trend: PromptIR realizes dynamic restoration of degradation perception through prompt learning [8], and SUPIR realizes user-controllable image enhancement by combining text prompts and generation prior [9], which indicates the evolution of restoration technology to the direction of intelligence and interaction.

In recent years, the field of image inpainting has made substantial progress in several key technical dimensions. In terms of model architecture, the introduction of state Space Model (SSM) brings innovation in computational efficiency. VmambaIR breaks through the unidirectional modeling limitation of traditional SSM by Omni Selective Scan mechanism, and achieves 0.36dB PSNR improvement compared with SwinIR in super-resolution tasks while maintaining linear computational complexity [10]. Large kernel convolutional networks (such as OKNet) enhance the multi-scale representation ability through 63×63 convolution kernels, and achieve 3.47dB PSNR gain in deblurring tasks [11]. The development of the base model paradigm is equally compelling. UniFMIR improves the FRC index by 15% in fluorescence microscopy image reconstruction through cross-modal pre-training [12], which verifies the potential of general representation learning.

Fusion strategies that generate priors continuously push performance boundaries. DiffBIR innovates to decoupling the diffusion process into the degradation removal and detail regeneration stages, and achieves 28.74dB PSNR on the BSR50 dataset, which is 4.2dB higher than the traditional method [7]. By integrating Stable Diffusion XL priors, SUPIR outperforms existing methods by 23% on perceptual quality metrics such as NIQE and supports semantically guided repair [9]. These breakthroughs show that hybrid paradigms combining physical degradation models with data-driven priors are becoming a mainstream direction.

This review systematically combs the development context of image inpainting technology driven by deep learning, focuses on the innovation and expansion of several representative methods, and discusses the core technology system based on CNN, Transformer and generative model. The latest methods were compared and evaluated from the dimensions of model architecture and efficiency optimization. The common challenges such as data dependence, computational efficiency, and cross-domain generalization were analyzed. The frontier directions such as lightweight design, multimodal learning, and 3D reconstruction are prospected. This paper aims to provide researchers with a panoramic view of the technology evolution and promote the application of image inpainting technology in real complex scenes.

2. Review of Main Models

2.1. PromptIR

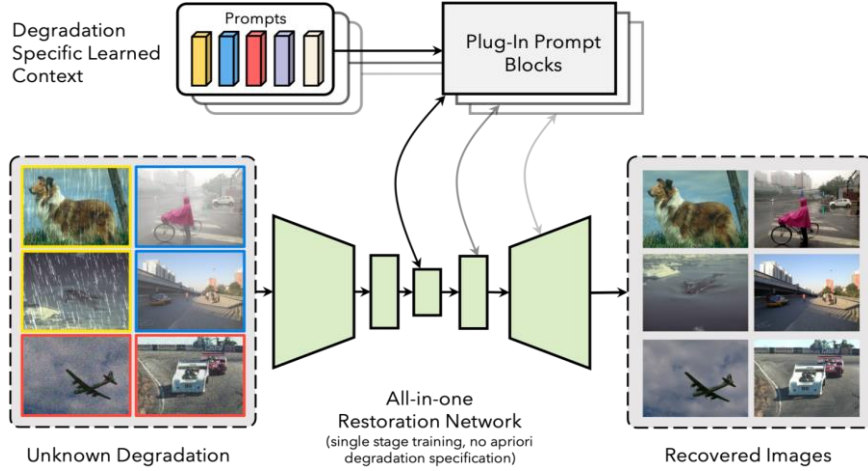


Figure 1. The network of PromptIR model [8]

In Figure 1, PromptIR is a global image restoration method based on cue learning. By introducing a hint block, it encodes the image degradation information into the hint parameters, and dynamically guides the inpainting network to repair, so as to achieve one-stop image inpainting. This allows PromptIR to generalize to different degradation types and degrees, while still achieving state-of-the-art results on tasks such as image denoising, rain removal, and haze removal. In the experiments, PromptIR achieves better performance than the existing methods in the All-in-One setting, and achieves significant performance improvement on the tasks of image haze removal, image rain removal, and image denoising. Meanwhile, PromptIR still outperforms the other methods in the single-task setting, demonstrating the effectiveness of the hint block [8].

2.2. SUPIR

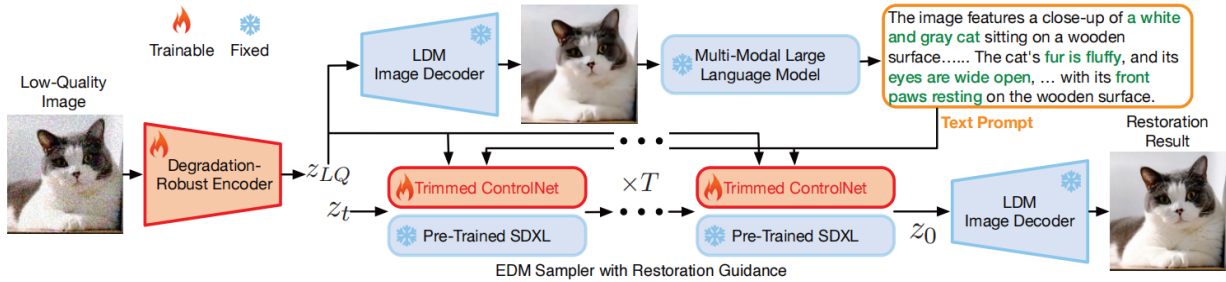


Figure 2. This figure briefly shows the workflow of the proposed SUPIR model [9]

As shown in Figure 2, SUPIR is an image restoration method based on model scaling and generative priors aiming to explore the visual effect and intelligence of image restoration. We use Stable Diffusion-XL (SDXL) as a powerful generative prior with 2.6 billion parameters and are able to generate high-quality images. By fine-tuning the image encoder to make it robust to image degradation, we can more accurately understand the content of low-quality images. We introduce a new adapter, including pruned ControlNet and ZeroSFT connectors, to efficiently apply SDXL to image inpainting tasks and reduce computational cost. It uses LLaVA multi-modal language model to understand the content of low-quality images and uses them as text prompts to guide the inpainting process. By using negative quality hints, the model is guided to avoid generating low-quality images, thereby improving image quality. In addition, this paper proposes a new sampling method to ensure that the inpainting result is faithful to the input low-quality image by controlling the generation process. In our experiments, SUPIR performs well in terms of perceptual quality and is able to control the inpainting process using text cues, such as controlling object inpainting, material textures, and semantic properties [9].

2.3. DiffBIR

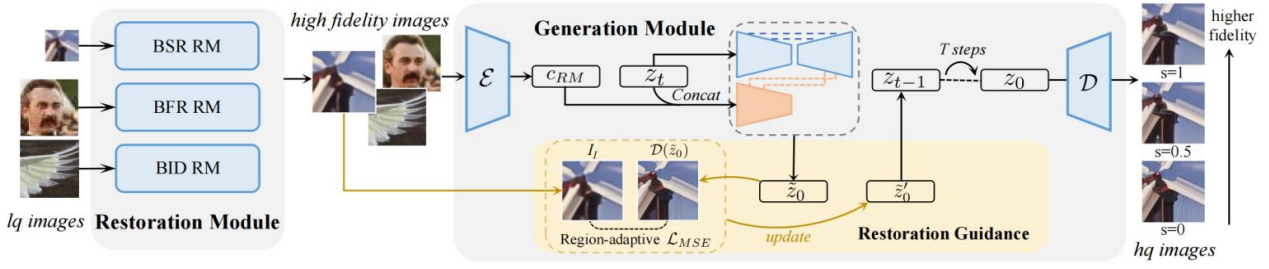


Figure 3. The two-stage pipeline of DiffBIR. 1) Restoration Module (RM) for degradation removal; 2) Generation Module (GM) for realistic image reconstruction with optional region-adaptive restoration guidance for a trade-off between quality and fidelity [7]

In Figure 3, DiffBIR is a blind image restoration method based on generative diffusion priors, such as blind image super-resolution (BSR), blind image denoising (BID), and blind face restoration (BFR). It decomposes the BIR problem into two stages, denoising and information regeneration, which are processed using independent modules. In the degraded denoising part, the model removes content-independent degradations in the image such as noise, blur, etc., using a specific restoration module (RM). The RM can be selected according to different BIR tasks, such as BSRGAN for BSR, SwinIR for BFR, and SCUNet for BID. The purpose of RM is to generate more reliable prior images to provide input for the generation module in the next stage. In the information regeneration part, IRControlNet is used to generate the missing image content, such as details, textures, etc. IRControlNet is trained on pre-trained text-to-image latent Diffusion models (e.g., Stable Diffusion) and leverages techniques such as conditional coding, conditional networks, and feature modulation to generate realistic and visually pleasing image content.

In addition, a region-adaptive recovery guidance is introduced to the model, which allows users to control the authenticity and clarity of the generated results through an adjustable guidance scale to achieve a trade-off between fidelity and quality. The guidance mechanism guides the generation result towards the prior image by calculating the region-adaptive MSE loss between the generated result and the prior image, which is optimized by the gradient descent algorithm. In the experiments, DiffBIR outperforms the existing methods on BSR, BFR and BID tasks and achieves a balance between fidelity and quality [7].

2.4. Omni-Kernel Network

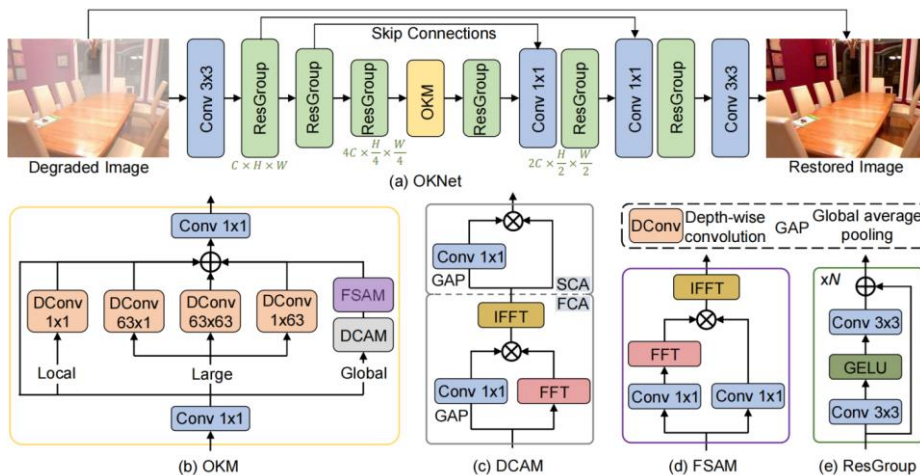


Figure 4. The architecture of OKNet. FFT and IFFT denote fast Fourier transform and its inverse operation, respectively [11]

As shown in Figure 4. OKNet is an image restoration network based on large kernel convolution. It improves image restoration performance by enhancing multi-scale representation learning. The core

component of OKNet, Omni-Kernel Module (OKM), consists of three branches, which are used to learn local, large-scale and global feature representations respectively. Large-scale branch: This branch uses large-size deep convolutions of different shapes (63x63, 1x63, 63x1) to obtain receptive fields of different shapes to better capture large-scale features. Global branch: In this branch, the dual-domain Channel Attention Module (DCAM) and Frequency-based Spatial attention module (FSAM) are used to realize the global receptive field, and the global features are effectively extracted by dual-domain processing. Small-scale branch: This branch uses 1x1 depth convolutions to supplement local information to better handle small-scale features. In the experiments, OKNet achieves excellent performance on tasks such as image dehazing, image deblurring, and image snow removal, which outperforms other state-of-the-art methods [11].

2.5. VmambaIR

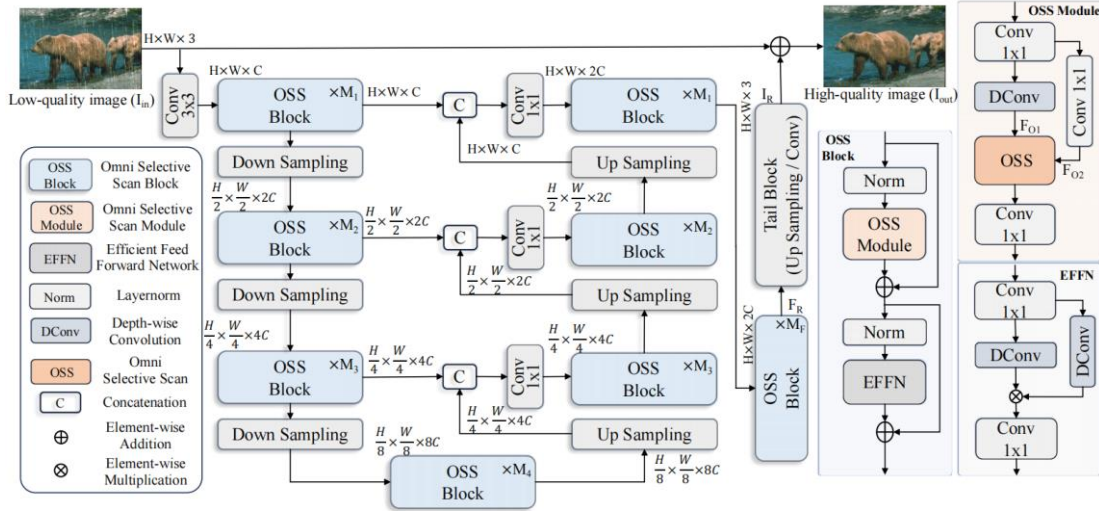


Figure 5. Overview of our VmambaIR [10]

In Figure 5, VmambaIR is an image restoration network based on state Space Model (SSM). It utilizes the multi-scale feature extraction ability of Unet, and combines the Omni Selective Scan (OSS) mechanism and EFFN technology to achieve efficient and accurate image restoration. In the experiments, VmambaIR achieves better performance than the existing SOTA methods on the tasks of single image super-resolution, actual image super-resolution and image rain removal, showing finer image details. At the same time, OSS mechanism has linear computational complexity, which is more suitable for processing large images [10].

2.6. UniFMIR

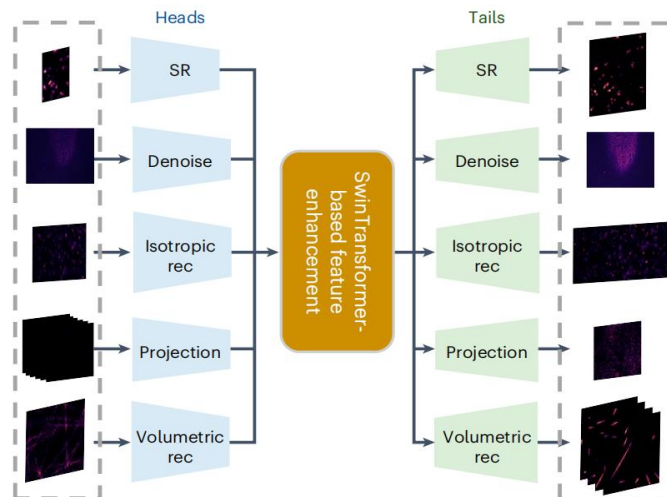


Figure 6. The network of UniFMIR model [12]

In Figure 6, UniFMIR is a basic model-based fluorescence microscopy image restoration (FMIR) method. It adopts the architecture of multi-head, multi-tail and Swin Transformer feature augmentation modules, and uses techniques such as pruning and quantization to reduce model size and computational cost. In experiments, UniFMIR evaluates the performance of the model on five different FMIR tasks, including super-resolution, isotropic reconstruction, 3D images denoising, surface projection, and volume reconstruction, and achieves excellent performance. The advantage of this model is that UniFMIR learns general image restoration knowledge through pre-training, which reduces the dependence on task-specific training data. It can quickly adapt to new tasks through fine-tuning, which improves the training efficiency. It can be used to solve a variety of image restoration problems and simplify the model development process [12].

2.7. MambaIR

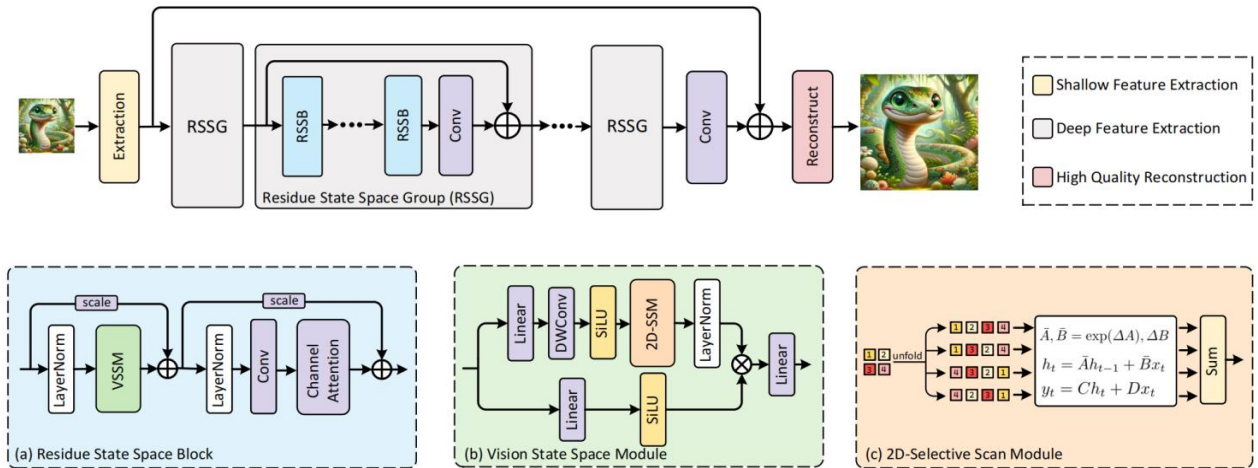


Figure 7. The overall network architecture of our MambaIR, as well as the (a) Residual State-Space Block (RSSB), the (b) Vision State-Space Module (VSSM), and the (c) 2D Selective Scan Module (2D-SSM) [13]

MambaIR is a comprehensive image inpainting network based on SSM. It introduces the Omni Selective Scan (OSS) mechanism to overcome the limitations of one-way modeling of SSM and can effectively model the image information flow. As shown in Figure 7, in terms of model features, MambaIR adopts SSM model to model global dependencies, while using convolutional layers to enhance local features, so as to obtain a larger receptive field while maintaining efficient computation. Residual state Space Block (RSSB) is the core component of MambaIR, which improves the standard Mamba model by introducing local convolution and channel attention mechanism to alleviate the problems of local pixel forgetting and channel redundancy. The Visual state Space Module (VSSM) uses SSM equations to capture long-distance spatial dependencies, and adopts a multi-direction scanning strategy to better utilize 2D spatial information. The Two-dimensional Selective Scanning Module (2D-SSM) unrolls the image feature map into a 1D sequence and scans along four directions to capture long-range dependencies and finally recover the 2D structure.

In the experiments, MambaIR achieves excellent performance on various image inpainting tasks, outperforming existing methods such as SwinIR with less computational cost [13].

2.8. AutoDIR

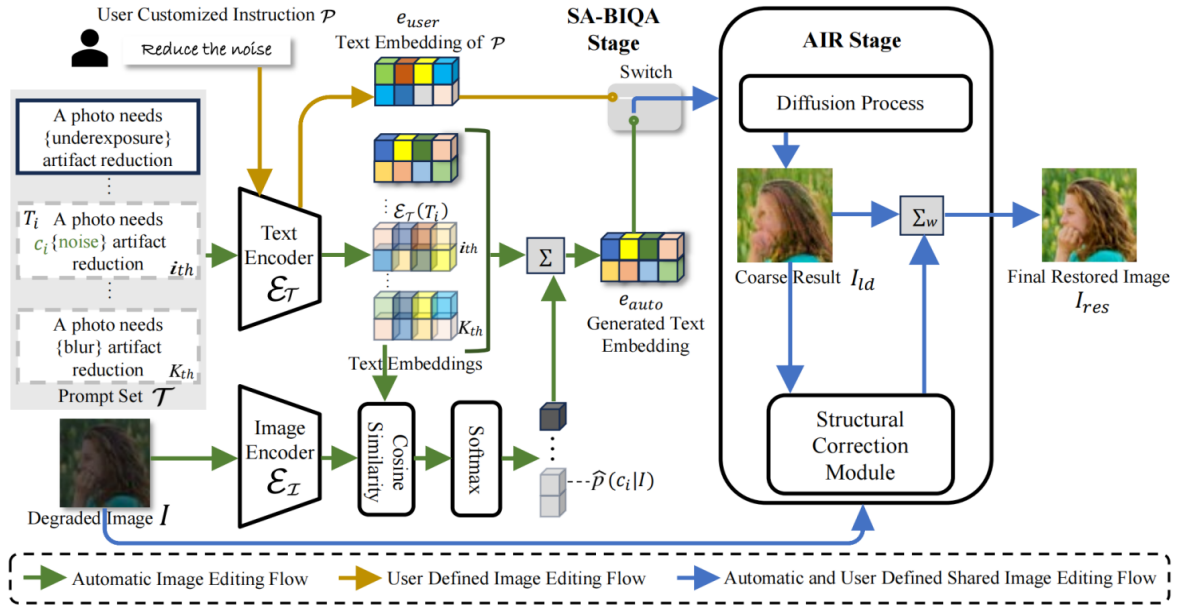


Figure 8. Diagram of the proposed All-in-One Image Restoration with Latent Diffusion (AutoDIR) [14]

In Figure 8, AutoDIR is an innovative image inpainting system that combines latent diffusion models and is able to automatically identify and inpaint various unknown image degradations. It consists of two key stages: a semantically oblivious Blind Image Quality Assessment (SA-BIQA) stage: This stage automatically detects unknown image degradations present in the input image using the SA-CLIP model based on a semantically oblivious visual language model. The A-BIQA stage generates the corresponding text prompt e_{auto} , which is used to guide the image inpainting process. Omnipotent image inpainting (AIR) stage: This stage utilizes the structure corrected Latent diffusion model (SC-LDM) to handle multiple types of image degradation. The SC-LDM model performs image inpainting based on text embeddings e_{auto} or user-defined open vocabulary instructions e_{user} generated in the SA-BIQA stage. In our experiments, AutoDIR achieves comparable or even better performance than state-of-the-art methods on multiple image inpainting tasks and supports inpainting of unseen types of image degradation [14]. Table 1 shows the comparison of different models.

Table 1. Models' comparison

Model Name	Key innovation point	Advantage	Disadvantage	Applicable Scene
PromptIR	The Prompt mechanism is introduced to guide the repair process	Flexible to adapt to different degradation types and has strong user interaction	Relies on a large number of data annotations and has low real-time performance	Custom repair
SUPIR	Well adapted to StableDiffusion-XL (SDXL), and integrates multi-modal language model LLaVA	High recovery effect Support for text control and high fidelity results	High computational resource demand Requires the accuracy of text prompts	Image restoration, image enhancement and image editing
DiffBIR	Generative Inpainting with Diffusion	The results are natural and support high quality texture synthesis	High computational resource consumption and slow inference speed	Artistic Image Inpainting and Super-Resolution
OKNet	Combining multi-scale attention mechanism with residual learning	It has strong detail ability and supports multi-level degradation repair	The number of parameters is large, and the training complexity is high	High resolution image inpainting
VmambaIR	Design of visual Architecture based on State Space Model (SSM)	It has high computational efficiency and strong fusion ability of global and local features	Adaptability to complex degradation types may be limited	Lightweight real-time repair tasks
UniFMIR	The unified base framework fuses multiple repair tasks	Multi-task compatibility reduces the cost of repeated training	Performance on specific tasks may be weaker than dedicated models	Biological application scenarios such as imaging repair of multiple microscopes
MambaIR	Improved Mamba architecture	Global effective receptive field, linear computational complexity	The number of parameters is large and the training time is long	Image super-resolution, image denoising and JPEG compression artifact removal
AutoDIR	Automated degradation modeling and reverse recovery	Without manually designing degenerate priors, it has strong generalization	It takes a long time to train and relies on large-scale datasets	Blind restoration of unknown degradation types

3. Core Challenges of current models

3.1. VmambaIR and MambaIR: Tradeoff between Efficiency and Local Details

Based on the state space model (SSM), VmambaIR and MambaIR have achieved efficient global modeling with linear computational complexity (such as the Omni Selective Scan mechanism of VmambaIR). However, their capacity for local detail restoration is still constrained by the unidirectional modeling characteristic of SSM. For example, although MambaIR outperforms SwinIR by 0.36 dB in PSNR in the image super-resolution task [13], there are still problems of edge blurring when reconstructing complex textures. Moreover, the hardware adaptability of SSM has not been fully verified, which might restrict its deployment on mobile terminals.

3.2. PromptIR and AutoDIR: The Cost of Dynamic Adaptation

PromptIR achieves unified processing of multiple degradations through the prompt mechanism and outperforms AirNet by 8.13 dB PSNR in the defogging task [8]. However, it relies on a large amount of degradation-labeled data to generate prompt parameters and is susceptible to unknown degradations in real scenarios. Similarly, although the blind image quality assessment (BIQA) module of AutoDIR can automatically detect degradation types, it has a high misjudgment rate for low-light or mixed degradations, leading to unstable restoration results [14].

3.3. DiffBIR and SUPIR: The Contradiction between Generation and Fidelity

DiffBIR achieves high-fidelity restoration by leveraging diffusion priors, but its two-stage process (initialization + refinement) leads to a significant increase in computational cost, with the processing time for a single image reaching several minutes. Although SUPIR enhances generalization through large-scale pre-training, it is prone to structural distortion in specific domains (such as medical imaging) due to differences in data distribution.

3.4. OKNet and UniFMIR: The Battle of Complexity and Generalization in Architectural Design

OKNet achieves multi-scale modeling through multi-branch large kernel convolution (63×63), and outperforms the Fourmer model by 3.47 dB PSNR in the deblurring task. However, it has a staggering 21.5M parameters and only a 12 FPS inference speed [11]. UniFMIR exhibits excellent performance in multimodal data fusion (such as an 15% improvement in the FRC metric for fluorescence microscopy images), but its training requires paired data with cross-modal alignment, which limits its application in unsupervised scenarios [12].

4. Common Technical Bottlenecks

Regarding the data dependency dilemma, current models generally rely on high-quality paired datasets (such as the millions of data in Real-ESRGAN), but the degradation types in real scenarios are complex and the annotation cost is high. For instance, DiffBIR shows a 4.2dB drop in PSNR when not trained at the noise level ($\sigma=100$), exposing the sensitivity issue of the model to data distribution. Regarding real-time performance and computational resource limitations, it is not difficult to see that VmambaIR achieves linear complexity but still requires 2.1GB of VRAM to process 4K images; the LAControlNet sub-network of DiffBIR accounts for 67% of the overall computation. On the mobile side, MambaIR-light is lighter than SwinIR-light (with 18% fewer parameters), but its PSNR still drops by 0.3dB. Moreover, insufficient cross-task and cross-modal generalization is also a significant problem: a single model (such as PromptIR) shows large performance differences among tasks when simultaneously handling denoising, rain removal, and super-resolution. For multi-modal models (such as UniFMIR), the SSIM fluctuation in cross-device data (such as different imaging parameters of different microscopes) is also quite obvious.

5. Future Research Directions

5.1. Lightweight Design and Hardware Co-design

Model Compression: Borrowing the channel attention mechanism from MambaIR, develop a dynamic weight pruning strategy to reduce redundant computations while maintaining the global receptive field. For instance, the dynamic pruning method based on the SENet module assesses the importance of channels through the attention mechanism and combines sparse regularization to achieve dynamic pruning, thereby significantly reducing the model's parameters and computational load [15]. **FPGA/ASIC Adaptation:** Targeting the sequential modeling characteristics of SSM [16], design parallelized hardware architectures (such as the pulse array of Huawei's Da Vinci NPU), enabling real-time processing (e.g., 30 FPS @ 1080p).

5.2. Multimodal Fusion

Cross-modal Alignment: Utilize contrastive learning (such as CLIP) to construct a joint embedding space for degradation-repair, reducing reliance on paired data [17]. For instance, associate text descriptions (such as "motion blur") with image features to enhance the zero-shot generalization ability of PromptIR. **Uncertainty-aware Fusion:** Draw inspiration from the QMF framework and dynamically weight multimodal inputs through the DS evidence theory to enhance robustness under low-quality data (such as joint restoration of noisy images and voice descriptions) [18][19]. This method can effectively combine information from different modalities and explicitly model uncertainty, thereby improving the accuracy and reliability of restoration. In scenarios of joint restoration of noisy images and voice descriptions, this method can fully utilize the additional information provided by the voice description, helping the model better understand the content of the noisy image and perform more precise restoration.

5.3. Three-dimensional Repair Driven by Neural Radiance Field (NeRF)

NeRF Prior Enhancement: By integrating the diffusion prior of DiffBIR with the geometric constraints of NeRF [20], consistent 3D scene restoration is achieved (e.g., the generalization error of HyP-NeRF is reduced by 22%). **Dynamic Object Removal:** Based on multi-view segmentation of Spinc-NeRF, seamless restoration of occluded objects in complex scenes (such as street views) is realized, with a significant improvement in PSNR [21].

6. Conclusion

In recent years, the rapid development of deep learning has significantly advanced the progress of image restoration technology. Methods based on convolutional neural networks (CNN), generative adversarial networks (GAN), Transformer, and diffusion models (DM) have achieved breakthroughs in tasks such as image denoising, super-resolution, and deblurring. Representative works like PromptIR achieve dynamic degradation perception through prompt learning, VmambaIR balances efficiency and global modeling ability using state space models (SSM), DiffBIR enhances fidelity by combining diffusion priors, and UniFMIR enhances cross-modal generalization through base models. All of these have broken through the limitations of traditional methods in different dimensions. However, current technologies still face multiple challenges: contradictions between generation and fidelity (such as the high computational cost of diffusion models), data dependency dilemmas (high annotation costs, complex degradation in real-world scenarios), insufficient cross-task generalization (significant performance differences among multi-task models), and real-time limitations (difficulties in deploying large models to mobile devices). The core contradiction of current technical bottlenecks lies in the data-computation-quality triangle dilemma: annotation costs limit model generalization, computing resources constrain real-time deployment, and generation and fidelity are difficult to achieve simultaneously. Future research needs to integrate lightweight design, multimodal learning, and 3D reconstruction and other emerging technologies to build a more universal and efficient

restoration framework, promoting its deep application in fields such as healthcare, autonomous driving, and cultural heritage protection.

References

- [1] M. Chris, et al. Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN). *Applications of Digital Image Processing XL*. Vol. 10396. SPIE, 2017.
- [2] M. Kuriakose, EDSR: Empowering super-resolution algorithms with high-quality DIV2K images. *Intelligent Decision Technologies* 17.4 (2023): 1249 - 1263.
- [3] Y. T. Hou, et al. The application of improved densenet algorithm in accurate image recognition. *Scientific Reports* 14.1 (2024): 8645.
- [4] H. T. Chen, et al. pre-trained image processing transformer. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.
- [5] J. Y. Liang, et al. Swinir: Image restoration using swin transformer. *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
- [6] V. H. G. Correa. Application of Real-ESRGAN in improving IR sensor Images for use in SAR operations. 2024.
- [7] X. Q. Lin, et al. Diffbir: Toward blind image restoration with generative diffusion prior. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024.
- [8] P. Vaishnav, et al. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv: 2306.13090* 2023.
- [9] F. H. Yu, et al. "Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
- [10] Y. Shi, et al. Vmambair: Visual state space model for image restoration. *IEEE Transactions on Circuits and Systems for Video Technology* (2025).
- [11] Y. N. Cui, et al. Omni-kernel network for image restoration. *Proceedings of the AAAI conference on artificial intelligence*. Vol. 38. No. 2. 2024.
- [12] C. X. Ma, et al. Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration. *Nature Methods* 21.8 (2024): 1558 - 1567.
- [13] H. Guo, et al. Mambair: A simple baseline for image restoration with state-space model. *European conference on computer vision*. Cham: Springer Nature Switzerland, 2024.
- [14] Y. T. Jiang, et al. Autodir: Automatic all-in-one image restoration with latent diffusion. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024.
- [15] J. Kang, et al. Research on weed detection algorithm based on dynamic pruning neural network. *Transactions of the Chinese Society for Agricultural Machinery*. 54.04 (2023): 268 - 275.
- [16] E. Alcaín, et al. Hardware architectures for real-time medical imaging. *Electronics* 10.24 (2021): 3118.
- [17] S. Shen, et al. How much can clip benefit vision-and-language tasks? *arXiv preprint arXiv: 2107.06383* (2021).
- [18] Q. Y. Zhang, et al. Provable dynamic fusion for low-quality multimodal data. *International conference on machine learning*. PMLR, 2023.
- [19] K. Y. Zhao, et al. A survey: Optimization and applications of evidence fusion algorithm based on Dempster-Shafer theory. *Applied Soft Computing* 124 (2022): 109075.
- [20] B. Mildenhall, et al. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65.1 (2021): 99 - 106.
- [21] H. H. Chen, et al. Mvip-nerf: Multi-view 3d inpainting on nerf scenes via diffusion prior. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.