Trifusion-RWKV for Complex Degradation Restoration in Library and Archive Environments

Yong Cui
Library of Central South University
ChangSha, China
cuiyong@csu.edu.cn

Ruilin Deng*
Library of Central South University
ChangSha, China
224139@csu.edu.cn

Liyang Yu*
Library of Central South University
ChangSha, China
222161@csu.edu.cn

***Abstract*—Historical library images suffer from low illumination, blur, and color fading. We propose TriFusion-RWKV, which includes three parallel branches: the Frequency-domain Processing Branch corrects illumination via learnable filtering and amplitude adjustment; the Spatial Processing Branch suppresses blur using multi-scale dilated convolution and channel attention; the Color Restoration Branch restores colors with dynamic LUTs and pixel-wise adaptive mapping. Experiments on four datasets show it achieves 23.99dB PSNR and 0.89 SSIM on LOL-v2-real with only 9.17M parameters, outperforming state-of-the-art methods in multi-degradation restoration while maintaining efficiency.**

***Index Terms*—Image Restoration, library, archive, RWKV**

I. Introduction

In historical research on libraries and archives, in addition to the collections themselves, a vast number of historical photographs documenting their interior scenes—such as shelf arrays, reading desk arrangements, and overall spatial layouts—constitute critical visual documents for understanding the material evolution and cultural functions of these institutions. However, due to the limitations of early photographic technology and preservation conditions, these invaluable images commonly suffer from three major types of degradation: low-light conditions, ghosting and blur, and color fading. The superimposition of multiple degradations severely constrains their academic utility. Consequently, developing efficient image restoration methods capable of simultaneously addressing multiple types of degradation holds significant research importance and practical application value.

Early image restoration techniques relied heavily on manually designed features, suffering from problems such as time and labor consumption, weak generalization ability, and limited applicable scenarios [1]. To address these issues, researchers have introduced CNNs into the field of image restoration and continuously optimized their performance. For instance, Cui, Yuning et al. developed the lightweight image restoration model IRNeXt using low-cost convolution operators [2]. Yulu Bai et al. utilized regularization terms to counteract the equivariance problem of convolutional layers while meeting the requirements of high-precision representation and flexible symmetry adaptation for image restoration [3]. However, the limited receptive field of CNNs imposes constraints on model accuracy. Consequently, Transformers have begun to attract the attention of researchers. For example, Jingyun Liang et al. built the high-quality image reconstruction model SwinIR based on Swin Transformer blocks [4]; Wang, Qiong et al. integrated the complementary advantages of CNNs and Transformers to achieve efficient image rain streak removal [5]. Nevertheless, the high computational complexity of Transformers makes them unsuitable for resource-constrained environments. To balance performance and efficiency, a series of lightweight methods based on state-space models (e.g., Mamba, RWKV) have emerged. For instance, Boyun Li et al. proposed a Mamba structure that combines local continuity preservation and sequence aggregation mechanisms [6]; Baocai Chang et al. constructed a Mamba-based attention wavelet network, integrating multi-scale features through 2D selective scanning [7]; Zhiwen Yang et al. applied RWKV to medical image restoration for the first time and designed a WKV attention mechanism with linear complexity to capture global dependencies [8].

Despite significant advancements in image restoration technology, research targeting the specific scenarios of libraries and archives remains insufficient. Images in such contexts often suffer from multiple concurrent degradations—low illumination, ghosting and blur, and color fading—and frequently need to be processed on ordinary computing devices, imposing higher demands on both performance and efficiency. To address these challenges, we propose TriFusion-RWKV, an image restoration model tailored for library and archive scenes. The core of our model is the Fusion-RWKV module, which comprises three parallel branches—a frequency-adaptive branch for illumination correction via learnable filtering and amplitude adjustment, a multi-scale dilated attention branch for ghosting and blur suppression through expanded receptive fields, and a dynamic LUT branch for color restoration via pixel-wise adaptive color mapping. These branches are integrated by an adaptive fusion module, followed by RWKV-style gating and feed-forward networks to refine features. Built upon this core module, the overall architecture effectively handles multiple degradations while maintaining low computational complexity, making it well-suited for practical library and archive applications. Experimental results on public benchmarks demonstrate its superiority in restoring images with mixed degradation types.

* Corresponding author

979-8-3315-8470-2/26/$31.00 ©2026 IEEE

II. Related Work

Image restoration refers to the reconstruction of clear images from degraded images. Existing research can be categorized into three types based on applicable scenarios: models dedicated to low-light image enhancement (LLIE models), models that jointly handle low-light image enhancement and deblurring (LLIE-deblur models), and unified restoration models.

Low-light image enhancement technology aims to address issues such as blurriness, noise, and detail loss in images captured under low-light conditions. Early deep learning methods primarily relied on Convolutional Neural Networks (CNNs), but the locality of convolution operations limited their ability to model long-range dependencies. To tackle this, Xu R et al. proposed a convolution-Transformer collaborative perception block, which significantly improved the effect of image degradation restoration by fusing local and global information [9]. Yuanhao Cai et al. further introduced intermediate features from illumination prediction as guiding variables for the Transformer, enhancing the interpretability and performance of the method [10]. Despite the excellent enhancement capability demonstrated by Transformers, their quadratic computational complexity still poses an efficiency bottleneck in high-resolution image processing. To resolve this issue, Tao Wang et al. proposed an axial-based multi-head self-attention (A-MSA) mechanism; by calculating self-attention along the height and width directions of the image separately, the complexity was reduced to a linear level [11]. Jiesong Bai et al. designed an illumination feature embedding module based on Retinex theory and, for the first time, introduced a state-space model (Mamba) into low-light enhancement tasks, effectively balancing global modeling and computational efficiency [12].

In real-world scenarios, degradation of low-light images is usually complex. Apart from insufficient illumination and noise, such images are often accompanied by motion blur caused by camera shake or object movement. If existing Low-Light Image Enhancement (LLIE) models are applied directly, they tend to amplify blur artifacts, resulting in poor visual effects. To address this, researchers have focused on developing end-to-end models that jointly perform enhancement and deblurring. Yuezhou Li et al. proposed decomposing degraded images into two perceptual dimensions—"brightness" and "details"—and fusing features from auxiliary tasks through a parallel architecture to improve the restoration performance of the main task [13]. Daniel Feijoo et al. constructed an asymmetric encoder-decoder structure based on Metaformer, assigning low-light enhancement and deblurring tasks to different modules, and leveraging cross-domain information interaction to enhance reconstruction quality [14]. Shangchen Zhou et al. designed a data synthesis process that better aligns with real-world degradation scenarios, and trained a dedicated model to effectively learn the degradation patterns within [15].

Unlike specialized models designed for specific types of degradation, unified image restoration models aim to address multiple image degradation issues through a single model. Syed Waqas Zamir et al. proposed an encoder-decoder structure based on Transformer modules, which restores images with different degradation types by modeling long-range dependencies [16]. Zhendong Wang et al. introduced an improved local enhancement window mechanism to optimize the Transformer structure, reducing computational complexity while maintaining performance [17]. Hang Guo et al. were the first to introduce the state-space model (Mamba) into the field of image restoration; they designed a residual state-space block (RSSB) and combined it with local convolution and channel attention mechanisms, significantly enhancing the model's expressive capability [18]. Rui Xu et al. adapted the RWKV architecture to the image restoration task, optimized its information interaction method, and designed an SSF component that supports cross-module information fusion, further improving the restoration effect by enhancing feature transmission [19].

III. Methodology

A. Overall Architecture

As shown in Figure 1, the TriFusion-RWKV model adopts an encoder-decoder architecture. The encoder extracts multi-scale features and compresses resolution through input projection followed by multiple rounds of "Fusion-RWKV block and downsampling". the decoder gradually restores resolution via "upsampling and Fusion-RWKV block" and finally obtains the result through output projection. The Fusion-RWKV block is the core of the model, consisting of three parallel branches (Frequency-domain Processing Branch, Spatial Processing Branch, and Color Restoration Branch) and an adaptive fusion module. It achieves in-depth feature optimization through gating mechanisms and residual connections. Input projection and output projection consist of simple convolutional layers.

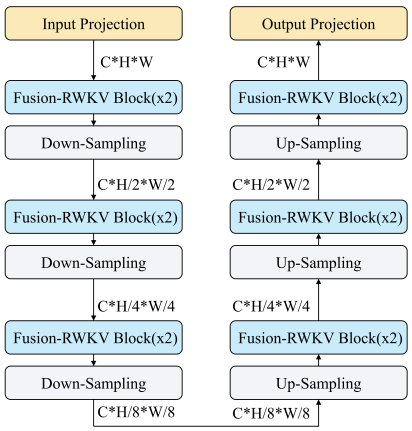

Fig. 1. The architectural details of TriFusion-RWKV

B. Fusion-RWKV

Frequency-domain Processing Branch. The essence of low-light enhancement lies in the adjustment of amplitude in the frequency domain [20]. This branch converts features to

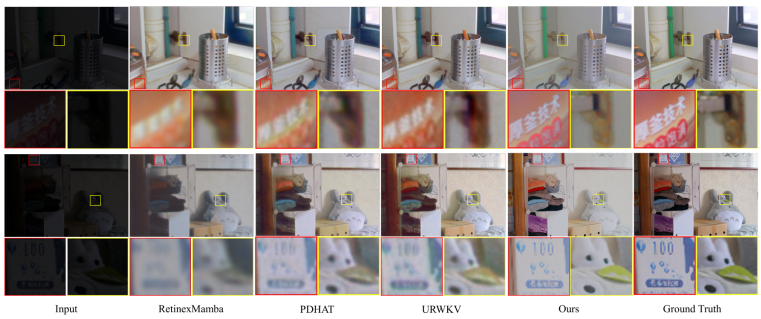

Fig. 2. Visual comparison of state-of-the-art models on the LOL-v2-real

TABLE I
QUANTITATIVE COMPARISON RESULTS ON THE LOL-V2-REAL, LOL-V2-SYN, SID, SMID DATASETS.IN THE "METHODS" COLUMN, THE BLUE-MARKED ONES ARE LLIE MODELS, THE RED-MARKED ONES ARE LLIE-DEBLUR MODELS, AND THE BLACK ONES ARE UNIFIED MODELS.

Methods	LOL-v2-real		LOL-v2-syn		SID		SMID		Complexity	
	PSNR↑	*SSIM*↑	*PSNR*↑	*SSIM*↑	*PSNR*↑	*SSIM*↑	*PSNR*↑	*SSIM*↑	*FLOPs (G)*	*Params (M)*
BiFormer [9]	22.67	0.86	24.87	0.92	22.49	0.63	28.24	0.81	5.93	0.83
Retinexformer [10]	22.79	0.83	25.67	0.92	24.44	0.68	29.15	0.81	15.57	1.61
LLFormer [11]	21.63	0.80	24.13	0.90	22.83	0.65	28.42	0.79	22.52	24.52
RetinexMamba [12]	22.45	0.84	25.88	0.93	22.45	0.65	28.62	0.80	34.75	24.1
PDHAT [13]	20.16	0.84	24.94	0.93	21.93	0.65	29.19	0.81	208.19	7.83
LEDNet [15]	19.02	0.83	24.79	0.93	21.47	0.63	28.42	0.80	38.57	7.41
Restormer [16]	18.60	0.78	21.41	0.83	22.01	0.64	28.58	0.80	140.99	26.11
MambaIR [17]	20.45	0.84	25.65	0.93	22.02	0.65	28.41	0.80	60.66	4.30
Uformer [18]	18.82	0.77	19.66	0.87	18.54	0.57	27.20	0.79	12.00	5.29
URWKV [19]	23.11	0.87	26.36	0.94	23.11	0.67	29.44	0.826	18.34	2.25
Ours	23.99	0.89	27.13	0.95	24.18	0.69	29.89	0.83	22.19	9.17

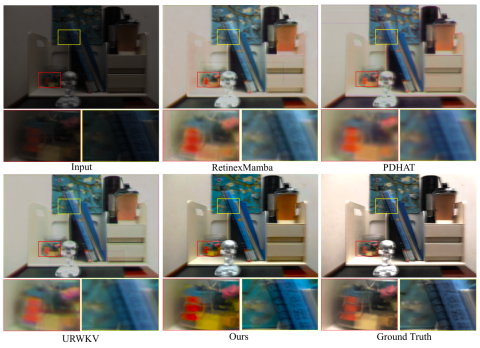

Fig. 3. Visual comparison of state-of-the-art models on an image with complex degradations

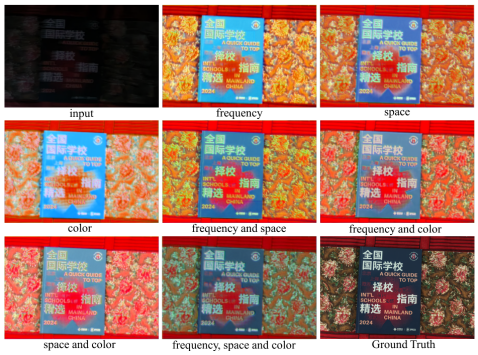

Fig. 4. Visual comparison of Image Restoration Results in Ablation Experiments. "frequency" represents the Frequency-domain Processing Branch, "spatial" represents the Spatial Processing Branch, and "color" represents the Color Restoration Branch.

the frequency domain via Fourier transform and optimizes frequency-domain features using a learnable filter and an amplitude adjustment mechanism. The specific steps are as follows:

(a) Frequency-domain transformation: First, reshape the spatial features $x \in \mathrm{R}^{B\times H\times W\times d}$ into $x \in \mathrm{R}^{B\times d\times H\times W}$, then convert them to the frequency domain using 2D Fourier transform:

$$X_{\text{freq}} = \text{FFT2}(x_{\text{norm}}) \tag{1}$$

where x_{norm} denotes the feature normalized by LayerNorm, and FFT2$(\cdot)$ represents the orthonormally sampled 2D Fast Fourier Transform (FFT).

(b) Learnable frequency-domain filtering: The output from the previous step is filtered using a learnable frequency-domain filter. This filter consists of learnable real-part parameters $F_{\text{r}} \in \mathrm{R}^{1\times d\times k\times k}$ and imaginary-part parameters $F_{\text{i}} \in \mathrm{R}^{1\times d\times k\times k}$ (where k is the filter size), After adjusting the filter to a size matching the input features via bilinear interpolation, complex multiplication is performed on the frequency-domain features:

$$X^{\text{real}}_{\text{filtered}} = X^{\text{real}}_{\text{freq}} \cdot F_r - X^{\text{imag}}_{\text{freq}}(F_{i,r}) \tag{2}$$

$$X^{\text{imag}}_{\text{filtered}} = X^{\text{real}}_{\text{freq}} \cdot F_i - X^{\text{imag}}_{\text{freq}}(F_{i,r}) \tag{3}$$

TABLE II
RESULTS OF ABLATION EXPERIMENTS."FRE" REPRESENTS THE FREQUENCY-DOMAIN PROCESSING BRANCH, "SPA" REPRESENTS THE SPATIAL PROCESSING BRANCH, AND "COL" REPRESENTS THE COLOR RESTORATION BRANCH.

Methods			LOL-v2-real		Complexity	
fre	*spa*	*col*	*PSNR* ↑	*SSIM* ↑	*FLOPs (G)*	*Params (M)*
✓			22.94	0.770	5.85	4.17
	✓		22.82	0.761	3.70	4.12
		✓	23.41	0.822	18.98	8.36
✓	✓		23.17	0.800	6.38	4.55
✓		✓	23.76	0.861	21.66	8.79
	✓	✓	23.64	0.852	21.76	6.49
✓	✓	✓	23.99	0.891	22.19	9.17

where $X_{\text{filtered}}^{\text{real}}$ and $X_{\text{filtered}}^{\text{imag}}$ are the real part and imaginary part of the frequency-domain features, respectively, and the filtered feature X_{filtered} can be calculated as:

$$X_{\text{filtered}} = X_{\text{filtered}}^{\text{real}} + j \cdot X_{\text{filtered}}^{\text{imag}} \tag{4}$$

(c) Amplitude adjustment: Extract the amplitude $A = |X_{\text{filtered}}|$ and phase $\phi = \angle X_{\text{filtered}}$ of the filtered frequency-domain features, and adjust the amplitude using learned parameters:

$$\hat{A} = A \cdot (1 + \tanh(\alpha)) + \beta \tag{5}$$

where $\alpha, \beta \in \mathrm{R}^{B\times d\times 1\times 1}$ are obtained by linear projection of the global mean of the features.

(d) Frequency domain to spatial domain conversion: Reconstruct frequency-domain features based on the adjusted amplitude and original phase, convert them back to the spatial domain via inverse Fourier transform: $x_{\text{freq}} = \text{Real}(\text{IFFT2}(\hat{A} \cdot e^{j\phi}))$, and finally control the information flow of frequency-domain features through a gating mechanism:

$$F_{\text{freq}} = x_{\text{freq}} \cdot \sigma(\text{Linear}(x_{\text{glob}})) \tag{6}$$

where σ denotes the Sigmoid function, and x_{glob} represents the mean value of spatial dimensions.

Spatial Processing Branch. Image deblurring relies on a large receptive field to capture global information [21]. This branch expands the receptive field through multi-scale dilated convolution and enhances key features by integrating channel attention.

(a) Multiple dilated convolutions with different dilation rates $d \in \{1, 2, 4\}$ are employed to expand the receptive field without increasing parameters:

$$\mathbf{F}_{\text{spatial}} = \sum_{d\in D} \text{Conv2D}_{\text{dilated}=d}(\mathbf{F}_{\text{in}}) \tag{7}$$

(b) Channel attention mechanism: Channel weights are learned through adaptive average pooling and convolutional layers to enhance important feature channels:

$$F_{\text{spatialout}} = F_{\text{spatial}} \otimes \text{CA}(F_{\text{spatial}}) \tag{8}$$

where $\otimes$ denotes element-wise multiplication, and the attention map is generated by the squeeze-and-excitation (SE) structure.

Color Restoration Branch. The Lookup Table (LUT), as a classic color mapping tool, efficiently maps an input color space to a target color space and is widely used in color correction and style transfer [22]. However, traditional static LUTs lack adaptability to image content, making them unable to handle images with varying degrees of degradation. Moreover, a globally uniform mapping struggles to correct local color anomalies. To overcome these limitations, we design a learnable dynamic LUT mechanism that generates multiple base LUTs using global context and combines them with local attention maps for pixel-wise adaptive fusion, thereby improving the flexibility and accuracy of color restoration while maintaining efficiency. The specific steps are as follows:

(a) Color space projection. First, the input feature $F_{\text{in}} \in \mathbb{R}^{B\times C\times H\times W}$ is projected into the RGB color space via a 1×1 convolution, yielding $I_{\text{rgb}} \in \mathbb{R}^{B\times 3\times H\times W}$, which facilitates subsequent color mapping.

(b) Dynamic LUT generation. To accommodate different color degradation patterns, we leverage global features to generate multiple LUTs. Global average pooling is applied to I_{rgb} to obtain a global feature $g \in \mathbb{R}^{B\times C}$. Then, two fully connected layers predict the parameters of K LUTs, each represented as a trilinear interpolation grid $\text{LUT}_k \in \mathbb{R}^{N\times N\times N\times 3}$, where N is the grid resolution (set to $N = 33$ in experiments):

$$\{\text{LUT}_k\}_{k=1}^{K} = \text{MLP}(g), \quad \text{LUT}_k = \mathcal{T}(\theta_k) \tag{9}$$

where $\mathcal{T}$ denotes the reshaping of flattened parameters into the grid structure, and θ_k are learnable parameters.

(c) Local adaptive weight prediction. To achieve spatially varying color enhancement, a convolutional network predicts pixel-wise weights for the different LUTs from I_{rgb}. Two convolutional layers (with ReLU activation) generate an initial weight map $W_0 \in \mathbb{R}^{B\times K\times H\times W}$, which is then normalized by Softmax to obtain the spatial weight map W_1:

$$W_1 = \frac{\exp(\hat{W}_1)}{\sum_{j=1}^{K} \exp(\hat{W}_1)} \tag{10}$$

(d) Adaptive LUT transformation. For each pixel position (h, w), the RGB value $I_{\text{rgb}}(h, w)$ is trilinearly interpolated in each of the K LUTs, and the resulting colors are weighted by the corresponding weights to produce the output color:

$$I_{\text{out}}(h, w) = \sum_{k=1}^{K} W_k(h, w) \cdot \text{LUT}_k\big(I_{\text{rgb}}(h, w)\big) \tag{11}$$

where $\text{LUT}_k(\cdot)$ denotes the trilinear interpolation operation on the 3D grid, returning the corresponding RGB vector.

(e) Gated fusion and residual connection. To prevent structural distortion caused by over-enhancement, a gating mechanism adaptively fuses the original color and the enhanced result. First, I_{rgb} and I_{out} are concatenated along the channel dimension, and a 3×3 convolution followed by a Sigmoid function generates the gating map G:

$$G = \sigma\big(\text{Conv}_{3\times 3}([I_{\text{rgb}}, I_{\text{out}}])\big) \tag{12}$$

Then the branch output is computed as:

$$F_{\text{color}} = G \odot I_{\text{out}} + (1 - G) \odot I_{\text{rgb}} \tag{13}$$

Finally, a 1×1 convolution expands F_{color} back to the original channel number C, yielding the final feature representation of this branch.

Adaptive Feature Fusion. To synergistically leverage the advantages of the three branches, we propose an adaptive fusion module instead of a simple summation.

(a) Adaptive Weight Learning

This module first learns the adaptive weights $W \in \mathrm{R}^{B\times 3\times 1\times 1}$ for the features ($F_{\text{freq}}, F_{\text{spatialout}}, F_{\text{color}}$) of the three branches:

$$F_{\text{cat}} = [F_{\text{freq}}, F_{\text{spatialout}}, F_{\text{color}}] \tag{14}$$

$$W = \text{Softmax}(\text{Conv2D}(\text{ReLU}(\text{Conv2D}(\text{GAP}(F_{\text{cat}}))))) \tag{15}$$

then, W is split into three weight maps w_1, w_2, w_3 along the channel dimension.

(b) Weighted Fusion and Cross-Branch Interaction

First, a weighted summation is performed, and then a gating mechanism is used to facilitate information interaction between branches.

$$F_{\text{weighted}} = w_1 \cdot F_{\text{freq}} + w_2 \cdot F_{\text{spatialout}} + w_3 \cdot F_{\text{color}} \tag{16}$$

$$G_{\text{cross}} = \sigma(\text{Conv2D}(F_{\text{cat}})) \tag{17}$$

$$F_{\text{fused}} = F_{\text{weighted}} \otimes G_{\text{cross}} \tag{18}$$

The finally obtained fused feature F_{fused} will be fed into the subsequent RWKV-style gating mechanism and feed-forward network (FFN) to complete the remaining intra-block computations.

IV. Experiment

A. Experimental Setup

The experimental setup was implemented with Python 3.10.13 and Torch 1.13.1. The execution environment featured a 13th Gen Intel(R) CPU running at 2.50 GHz, an Nvidia GeForce RTX 3090 GPU equipped with 24 GB of memory, and the Windows 11 operating system. For hyperparameters, the epoch count was set to 100, the batch size to 16, and the learning rate to 0.0001. In this study, the widely - used PSNR and SSIM metrics were employed to assess the experimental outcomes.

B. Comparative Experiments

We conducted a comprehensive comparison between TriFusion-RWKV and current state-of-the-art methods on four public datasets (LOL-v2-real [23], LOL-v2-syn [23], SID [24], and SMID [24]). As shown in Table I, TriFusion-RWKV achieves optimal or near-optimal performance across most evaluation metrics. On the LOL-v2-real dataset, our method reaches a PSNR of 23.99 dB and an SSIM of 0.89, outperforming all specialized low-light image enhancement models, while utilizing only 9.17M parameters and 22.19 GFLOPs.

Figure 2 compares the restoration results of different methods on a typical low-light image from the LOL-v2-real dataset. RetinexMamba produces a relatively dark output with blurred text. Although PDHAT and URWKV improve brightness and sharpness to some extent, blurring remains in the corner of the windowsill. In contrast, TriFusion-RWKV delivers more natural and clearer visual results, effectively mitigating the issues mentioned above. Figure 3 further presents the restoration outcome of a complex degraded image suffering from low light, motion blur, ghosting, and color fading. RetinexLambda and PDHAT yield dark outputs with blurred text on the clips and book titles inside the box, resulting in an overall unnatural appearance. While URWKV enhances the sharpness in the clip area, the overall image brightness remains insufficient. TriFusion-RWKV excels in brightness, sharpness, and overall color harmony, achieving the clearest restoration of details such as book titles and clips.

This advantage benefits from the collaborative mechanism of the model's three branches: the Frequency-domain Processing Branch effectively restores global illumination and frequency-domain structures through learnable filters and adaptive amplitude modulation; the Spatial ProcessingBranch significantly expands the receptive field and enhances the suppression of ghosting, blur, and noise by employing multi-scale dilated convolutions combined with channel attention; the Color Restoration Branch accurately reconstructs color-missing regions and restores faded information through dynamic lookup tables and pixel-wise adaptive color mapping.

C. Ablation Experiments

To verify the necessity and effectiveness of the three core branches, we conducted systematic ablation experiments on the LOL-v2-real dataset, with the results shown in Table II. Figure 4 presents the restoration effects of each single branch on an image with artifacts, blur, extremely low illumination, and color fading. It can be observed that the Frequency-domain Processing Branch, while effectively enhancing overall image brightness, struggles to restore color contrast; the Spatial ProcessingBranch significantly reduces blur and slightly improves the color contrast of the flower region, yet its enhancement remains limited; the Color Restoration Branch substantially improves overall color contrast but fails to effectively eliminate blur and artifacts. These results indicate that a single branch alone cannot simultaneously address multiple degradation types, underscoring the necessity of multi-branch collaborative restoration.The performance of any dual-branch combination outperforms that of single branches, demonstrating significant complementarity. The frequency-domain processing branch combined with the spatial processing branch (PSNR 23.17, SSIM 0.800) achieves synergistic enhancement in noise suppression and illumination adjustment. As shown in Figure 4, the restored image is brighter overall, and the flower colors are effectively recovered. The frequency-domain processing branch combined with the Color Restoration Branch (PSNR 23.76, SSIM 0.861) delivers the best performance among dual-branch combinations: the frequency-domain processing branch optimizes illumination, while the pixel completion branch restores structures, leading to a significant improvement in

overall quality with bright images and clear details. The spatial processing branch combined with the Color Restoration Branch (PSNR 23.64, SSIM 0.852) realizes deblurring and detail restoration based on spatial information; however, due to the lack of frequency-domain priors, its brightness restoration is slightly insufficient. Although all dual-branch combinations show visually superior results to single branches, the restored images still lack sufficient distinction from the red background in the results, revealing a common limitation. The full model achieves the optimal performance (PSNR 23.99, SSIM 0.891), indicating that the adaptive fusion module effectively integrates the advantages of the three branches. As shown in Figure 4, the final output image exhibits significant improvements in overall brightness, detail clarity, and color distinguishability—it has a clear contrast with the red background, and the details of text and flowers are also clearly distinguishable.

V. Conclusion

This study addresses the complex degradation issues of low illumination, blur, and color fading commonly found in library and archive images by proposing the TriFusion-RWKV model. Through the collaborative mechanism of three parallel branches—frequency-domain processing, spatial processing, and color restoration—the model achieves joint restoration of multiple degradations. Experimental results demonstrate that TriFusion-RWKV achieves optimal or near-optimal performance on four public datasets, reaching 23.99dB PSNR and 0.89 SSIM on the LOL-v2-real dataset, outperforming state-of-the-art methods. Ablation studies validate the necessity and complementarity of the three branches: the frequency-domain branch effectively restores global illumination, the spatial branch significantly suppresses blur and ghosting, and the color branch accurately reconstructs faded information, with their synergy enabling high-quality restoration. Meanwhile, the model maintains excellent lightweight characteristics with only 9.17M parameters and 22.19 GFLOPs, making it suitable for deployment on ordinary computing devices. This research provides an efficient and practical solution for restoring complex degraded images in library and archive scenarios.

References

[1] Yubin Gu, Yuan Meng, Jiayi Ji, and Xiaoshuai Sun, "Acl: Activating capability of linear attention for image restoration," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 17913–17923.

[2] Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll, "Irnext: Rethinking convolutional network design for image restoration," 2023.

[3] Yulu Bai, Jiahong Fu, Qi Xie, and Deyu Meng, "A regularization-guided equivariant approach for image restoration," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 2300–2310.

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844.

[5] Qiong Wang, Kui Jiang, Jinyi Lai, Zheng Wang, and Jianhui Zhang, "Hpcnet: A hybrid progressive coupled network for image deraining," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023, pp. 2747–2752.

[6] Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng, "Mair: A locality-and continuity-preserving mamba for image restoration," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 7491–7501.

[7] Baocai Chang, Genji Yuan, and Jinjiang Li, "Mamba-enhanced spectral-attentive wavelet network for underwater image restoration," *Engineering Applications of Artificial Intelligence*, vol. 143, pp. 109999, 2025.

[8] Zhiwen Yang, Jiayin Li, Hui Zhang, Dan Zhao, Bingzheng Wei, and Yan Xu, "Restore-rwkv: Efficient and effective medical image restoration with rwkv," *IEEE Journal of Biomedical and Health Informatics*, 2025.

[9] Rui Xu, Yuezhou Li, Yuzhen Niu, Huangbiao Xu, Yuzhong Chen, and Tiesong Zhao, "Bilateral interaction for local-global collaborative perception in low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 26, pp. 10792–10804, 2024.

[10] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12504–12513.

[11] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proceedings of the AAAI conference on artificial intelligence*, 2023, vol. 37, pp. 2654–2662.

[12] Jiesong Bai, Yuhao Yin, Qiyuan He, Yuanxian Li, and Xiaofeng Zhang, "Retinexmamba: Retinex-based mamba for low-light image enhancement," in *International Conference on Neural Information Processing*. Springer, 2024, pp. 427–442.

[13] Yuezhou Li, Rui Xu, Yuzhen Niu, Wenzhong Guo, and Tiesong Zhao, "Perceptual decoupling with heterogeneous auxiliary tasks for joint low-light image enhancement and deblurring," *IEEE Transactions on Multimedia*, vol. 26, pp. 6663–6675, 2024.

[14] Daniel Feijoo, Juan C Benito, Alvaro Garcia, and Marcos V Conde, "Darkir: Robust low-light image restoration," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 10879–10889.

[15] Shangchen Zhou, Chongyi Li, and Chen Change Loy, "Lednet: Joint low-light enhancement and deblurring in the dark," in *European conference on computer vision*. Springer, 2022, pp. 573–589.

[16] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[17] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia, "Mambair: A simple baseline for image restoration with state-space model," in *European conference on computer vision*. Springer, 2024, pp. 222–241.

[18] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17683–17693.

[19] Rui Xu, Yuzhen Niu, Yuezhou Li, Huangbiao Xu, Wenxi Liu, and Yuzhong Chen, "Urwkv: Unified rwkv model with multi-state perspective for low-light image restoration," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 21267–21276.

[20] Guoqing Wang, Bo Qiu, Ali Luo, Xiao Kong, Zhiren Pan, Qi Li, Fuji Ren, and Guanlong Cao, "Detection and restoration of abnormal band data in photometric images," *Computers and Electrical Engineering*, vol. 121, pp. 109871, 2025.

[21] Jin Cao, Xiangyu Rui, Li Pang, Deyu Meng, and Xiangyong Cao, "Latenthsi: Restore hyperspectral images in a latent space," *Information Fusion*, vol. 117, pp. 102848, 2025.

[22] Canlin Li, Haowen Su, Xin Tan, Xiangfei Zhang, and Lizhuang Ma, "Wv-lut: Wide vision lookup tables for real-time low-light image enhancement," *IEEE Transactions on Multimedia*, pp. 1–14, 2025.

[23] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 30, pp. 2072–2086, 2021.

[24] Hai Jiang, Binhao Guan, Zhen Liu, Xiaohong Liu, Jian Yu, Zheng Liu, Songchen Han, and Shuaicheng Liu, "Learning to see in the extremely dark," *arXiv preprint arXiv:2506.21132*, 2025.