MFmamba: A Multi-function Network for Panchromatic Image Resolution Restoration Based on State-Space Model

Qian Jiang[1], Qianqian Wang[1], Xin Jin[1*]
Michał Woźniak[2], Shaowen Yao[1], Wei Zhou[3*]
[1]School of software, Yunnan University, Kunming, Yunnan, China
[2]Faculty of Information and Communication Technology, Wroclaw University of Science and Technology, Wroclaw, Poland
[3]School of Engineering, Yunnan University, State Key Laboratory of Vegetation Structure, Function and Construction (VegLab), Kunming, Yunnan, China
jiangqian_1221@163.com, qianqianwang1325@163.com, xinxin_jin@163.com, michal.wozniak@pwr.edu.pl, yaosw@ynu.edu.cn, zwei@ynu.edu.cn

Abstract

Remote sensing images are becoming increasingly widespread in military, earth resource exploration. Because of the limitation of a single sensor, we can obtain high spatial resolution grayscale panchromatic (PAN) images and low spatial resolution color multispectral (MS) images. Therefore, an important issue is to obtain a color image with high spatial resolution when there is only a PAN image at the input. The existing methods improve spatial resolution using super-resolution (SR) technology and spectral recovery using colorization technology. However, the SR technique cannot improve the spectral resolution, and the colorization technique cannot improve the spatial resolution. Moreover, the pansharpening method needs two registered inputs and can not achieve SR. As a result, an integrated approach is expected. We designed a novel multi-function model (MFmamba) to realize the tasks of SR, spectral recovery, joint SR and spectral recovery through three different inputs. Firstly, MFmamba utilizes UNet++ as the backbone, and a Mamba Upsample Block (MUB) is combined with UNet++. Secondly, a Dual Pool Attention (DPA) is designed to replace the skip connection in UNet++. Finally, a Multi-scale Hybrid Cross Block (MHCB) is proposed for initial feature extraction. Many experiments show that MFmamba is competitive in evaluation metrics and visual results and performs well in the three tasks when only the input PAN image is used.

Code —
https://github.com/QianqianWang1325/MFmamba.git

Introduction

Remote sensing images are widely used in many fields (Ma et al. 2024; Park, Soh, and Cho 2025; Zhang et al. 2024). PAN and MS images are different types of remote sensing images with information from different wavelengths of the same area (Feng et al. 2022). PAN image has a high spatial resolution. Still, due to them represented as grayscale, the spectral resolution is low (Jin et al. 2024). The spectral resolution of MS image is higher, but due to physical limitations, the spatial resolution is lower. A single sensor cannot simultaneously capture remote sensing images with high spatial and spectral resolution. However, high-resolution (HR) color remote sensing images are more beneficial for interpreting remote sensing scenes, such as image fusion (Li et al. 2022b), semantic segmentation (An et al. 2024). Therefore, improving the spatial and spectral resolution of remote sensing images becomes a worthwhile research topic in the case of input-only PAN image.

The existing research uses SR technology to improve the spatial resolution of low-resolution (LR) images and colorization technology to recover the spectral resolution of grayscale images. Since the potential of CNN in image processing has been discovered, subsequent studies have introduced more CNN-based SR methods, such as (Li et al. 2023a). There are also many other methods to optimize image SR performance by different model structure, such as LGC-GDAN (Li et al. 2024) was based on a generative adversarial network (GAN), MAT (Xie et al. 2025) was based on Transformer, StableSR (Wang et al. 2024) was based on diffusion model. The automatic image colorization models based on deep learning (DL) have been gradually designed. For instance, CycleGAN-Color (Huang et al. 2021) was based on GAN, (Kumar, Weissenborn et al. 2021) was based on Transformer.

The purpose of PAN image joint SR and colorization is to promote the image's spatial resolution and spectral resolution synchronously. We designed a multi-function network (MFmamba) to solve these problems. Using different inputs, it can achieve three tasks, including PAN image SR, spectral recovery, joint SR and spectral recovery. The main contributions of our work are as follows:

- We design an efficient PAN image resolution restoration network (MFmamba) to produce colorized HR images. The proposed Mamba Upsample Block (MUB) adopts a state space model for resolution restoration.
- We design a Multi-scale Hybrid Cross Block (MHCB) for shallow feature extraction, which can detect local and multi-scale features effectively and improve the ability of detail feature extraction.
- We introduce a novel Dual Pool Attention (DPA) to im-

*Corresponding authors.

prove feature representation by dynamically adjusting channel weights so that the model may focus on more important feature channels.

Related Work

Image Super-Resolution (SR)

In recent years, image SR has made a breakthrough due to the strong learning capabilities of DL. Li *et al*. (Li et al. 2020) designed an SR model, MDCN, which utilized dense connection and residual strategy to improve performance and used fewer parameters. Li *et al*. (Li et al. 2023b) proposed a MulSR, which used a symmetric guidance encoder and direction-sensing spatial context aggregation module to realize multiscale information interaction. Li *et al*.(Li et al. 2025) released the power of deep image prior to the hyperspectral image SR and achieved high-quality reconstruction results. Zhu *et al*. (Zhu et al. 2023) enhanced GAN using residual blocks in the generator and attention mechanisms for feature fusion. It introduces RaGAN in the discriminator to improve edge and texture learning. Chen *et al*. (Chen et al. 2023) reported a novel Hybrid Attention Transformer (HAT). They were inspired by Swin Transformer and proposed a window-based self-attention, an overlapping cross-attention module for better cross-window interaction. Xie *et al*. (Xie et al. 2025) proposed a Multi-Range Attention Transformer for SR tasks, which harnesses the computational efficiency of dilation operations and integrates sparse multi-range attention (SMA) through self-attention mechanisms to capture features.

Li *et al*. (Li et al. 2022a) proposed the first diffusion-based model (SRDiff) for single-image SR, which transforms Gaussian noise into SR images through a Markov chain and utilizes residual prediction to accelerate convergence. Wang *et al*. (Wang et al. 2024) introduced a controllable feature wrapping module (StableSR) for balancing quality and fidelity via a scalar value and a progressive aggregation sampling strategy to adapt pre-trained diffusion models to any resolution. However, the above methods still have some problems. For example, extracting and retaining all textures is difficult, so error texture information will be generated in some places with complex details, and the detailed features extracted are insufficient.

Image colorization

In recent years, automatic colorization methods are gradually designed. Feng *et al*. (Feng et al. 2021b) proposed an end-to-end CNN network for remote sensing image colorization, using multi-scale residual receptive fields for feature extraction and adopting residual, attention, and pixel-shuffle blocks. Saeed *et al*. (Anwar et al. 2025) reviewed the image colorization techniques based on DL, pointed out the limitations of existing datasets, and introduced a new dataset for image colorization. Shafiq *et al*. (Shafiq and Lee 2024) designed an image colorization network using a hybrid architecture based on Transformer and GANs. They improved the encoder to generate color features and then used self-attention to capture long-range dependencies. Wu *et al*. (Wu et al. 2021) proposed a new model for PAN image colorization by utilizing DCGAN with an auto-encoder-based generator and a residual network discriminator. Zhao *et al*. (Zhao et al. 2021) proposed SCGAN. This fully automatic framework uses saliency maps and two hierarchical discriminators for perceptually accurate image colorization with minimal data, leveraging global features from a pre-trained VGG-16-Gray network.

Jiang *et al*. (Jiang et al. 2025) designed a multiscale colorization network using a hybrid architecture based on Transformer and CNN. They used the attention and mask mechanism to enhance image features. Liang *et al*. (Liang et al. 2024) developed a framework called CtrlColor, which leverages the pre-trained Stable Diffusion model to handle both unconditional and conditional colorization. The CtrlColor approach includes methods for encoding user-provided strokes, controlling color distribution and so on. However, these methods have some shortcomings. For example, it cannot address color distortion problems and depends on the color distribution, making color results unreasonable(Zhao et al. 2021). Therefore, grayscale image colorization methods still need further exploration.

Proposed Method

To achieve SR and spectral recovery tasks of PAN images, we propose a multi-function model (MFmamba) based on UNet++ (Zhou et al. 2018) and mamba, which can realize the tasks of SR, spectral recovery, and joint SR and spectral recovery of PAN image. Fig. 1 is the overall structure of our proposition. We combine the Mamba Upsample Module (MUB) constructed utilizing the state space model with UNet++ (Zhou et al. 2018) and design a novel Dual Pool Attention (DPA) to replace the original skip connection in UNet++ for information transfer between the same-level feature maps. Multi-scale Hybrid Cross Block (MHCB) is designed to extract multi-scale features through convolution operations with different convolution kernel sizes.

Multi-scale Hybrid Cross Block (MHCB)

We design a Multi-scale Hybrid Cross Block (MHCB) to extract multi-scale features information through parallel convolutional pathways with different kernel sizes, then skip connections with dense residual grouping to maintain gradient flow and reinforce critical feature persistence throughout the network hierarchy. Fig. 2 shows the MHCB concept. We use 5×5 focuses more on global feature extraction than a small convolution kernel. We use 3×3 kernel with fewer parameters to obtain local features and fully utilize the detailed information in the PAN image. We also employ 1×1 kernel at the module's end to fuse features and combine the features of each layer. MHCB can be represented as follows:

$$\begin{cases} X_1 = ReLU(3 \times 3Conv(X)) \oplus X, \\ X_2 = ReLU(5 \times 5Conv(X)) \oplus X, \end{cases} \tag{1}$$

$$X_3 = 1 \times 1Conv(Concat(X_1, X_2, X)), \tag{2}$$

$$\begin{cases} X_4 = ReLU(3 \times 3Conv(X_3)), \\ X_5 = ReLU(5 \times 5Conv(X_3)), \end{cases} \tag{3}$$

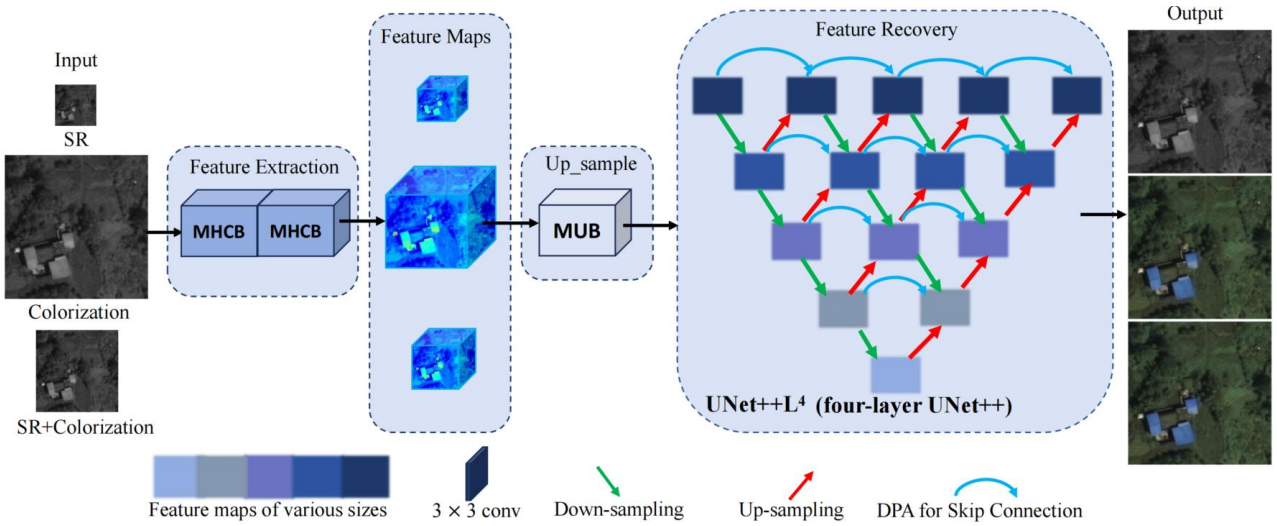

Figure 1: General structure of the proposed solution (MFmamba).

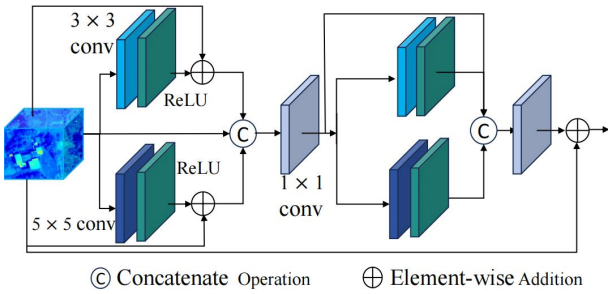

Figure 2: The structure of Multi-scale Hybrid Cross Block (MHCB).

$$MHCB_{out} = 1 \times 1Conv((Concat(X_3, X_4, X_5)) \oplus X, \tag{4}$$

where X represents the input feature of MHCB. X_1, X_2, X_3, X_4, and X_5 denote the intermediate output features. And $Concat$ represents the concatenate operation.

Dual Pool Attention (DPA)

In SR and the spectral recovery of PAN images, the attention mechanism is crucial in helping the model capture correlation and semantic information effectively. Therefore, we propose a novel Dual Pool Attention (DPA) as Fig. 3. DPA employs a dual-stream architecture for comprehensive channel-wise feature calibration. DPA utilizes both adaptive global average pooling (Eq. 5) and maximum pooling to compress global spatial information into channel descriptors. Then, the sigmoid function is used to fully obtain the channel-wise dependencies and generate weight feature maps of size $1 \times 1 \times C$, see Eq. 6. These maps are applied to the input to evaluate channel importance and then summed (Eq. 7). DPA introduces maximum pooling to capture significant features, enhancing model performance by focusing

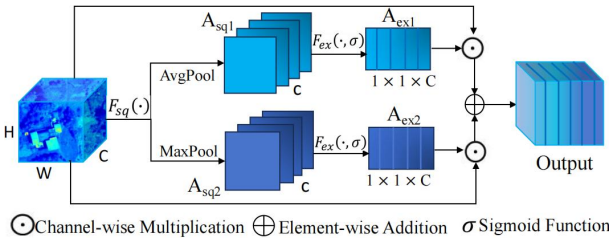

Figure 3: The structure of Dual Pool Attention (DPA).

on fine details and important information for more comprehensive feature extraction. Combining the two pooling operations helps the model focus on key details and relevant information, helping the model better retain the features of the input image.

$$\begin{cases} A_{sq1} = AP_c(i,j) = \dfrac{1}{H \times W} \displaystyle\sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,c}, \\ A_{sq2} = MP_c(i,j) = \displaystyle\max_{p,q \in \{1,\ldots,k\}} X_{p,q,c}, \end{cases} \tag{5}$$

$$\begin{cases} A_{ex1} = sigmoid(A_{sq1}), \\ A_{ex2} = sigmoid(A_{sq2}), \end{cases} \tag{6}$$

$$DPA_{out} = (X \odot A_{ex1}) \oplus (X \odot A_{ex1}), \tag{7}$$

where $X_{i,j,c}$ is the value at the feature map position (i,j) on the channel c. $MP_c(i,j)$ indicates the value of position (i,j) on channel c of the pooled output feature map. $X_{p,q,c}$ is the pixel value in the pooled window $k \times k$.

Mamba Upsampling Block (MUB)

The structured state-space sequence model has made a breakthrough in recently, which maps a 1D sequence $x(t) \in \mathbb{R} \rightarrow y(t) \in \mathbb{R}$ through an latent state $h(t) \in \mathbb{R}^M$. Based on this, mamba model was proposed to solve the

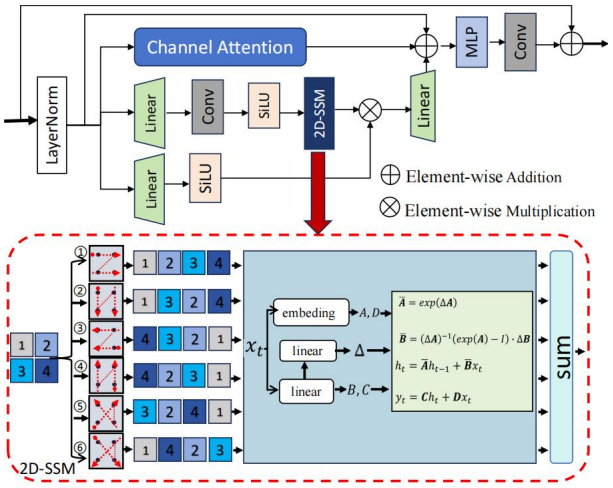

Figure 4: The structure of Mamba Upsample Module (MUB).

problem of low computational efficiency of Transformers on long sequences. Specifically, the state space sequence model is defined with five parameters $(\Delta, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, which sequence-to-sequence transformation is defined as a two-stage process. Formally, this system is represented as a linear ordinary differential equation (ODE) (Guo et al. 2024):

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \quad y(t) = \mathbf{C}h(t) + \mathbf{D}x(t), \tag{8}$$

where M represents the state size, $\mathbf{A} \in \mathbb{R}^{M\times M}$, $\mathbf{B} \in \mathbb{R}^{M\times 1}$, $\mathbf{C} \in \mathbb{R}^{1\times M}$, and $\mathbf{D} \in \mathbb{R}$.

The discretization process is then used to integrate Eq. 8 into an actual DL model. The parameter Δ controls the timescale, converting the continuous parameters $\mathbf{A}$ and $\mathbf{B}$ into their discrete counterparts, $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$. The widely used discretization approach is the zero-order hold (ZOH) defined as follows:

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}. \tag{9}$$

After discretized with a time step size of Δ, Eq. 8 can be reformulated as follows:

$$h_\tau = \bar{\mathbf{A}}h_{\tau-1} + \bar{\mathbf{B}}x_\tau, \quad y_\tau = \mathbf{C}h_\tau + \mathbf{D}x_\tau, \tag{10}$$

where τ represents a discrete time step, the original continuous time step t is converted to the discrete time step τ.

Eq. 10 describes a linear time-invariant (LTI) system that remains static with different input parameters, and the system's equation of state never changes with the input. However, this static system has certain limitations. To overcome this problem, (Yue Liu 2024) proposed a selective scan mechanism that dynamically adjusts matrices $\mathbf{B}$, $\mathbf{C}$, and Δ based on input data, improving performance. Inspired by the good performance of this selective scan mechanism, we employ a 2D selective scan mechanism (2D-SSM) to promote the awareness of contextual information embedded in the input. We designed the Mamba Upsample Module (MUB). The MUB structure diagram is shown in Fig. 4. We increase

	MHCB	DPA	MUB	PSNR↑	SSIM↑	MSE↓	MAE↓
w/o DPA	✓	✗	✓	39.980	0.965	7.457	76.791
w/o MUB	✓	✓	✗	39.794	0.965	7.679	**68.398**
w/o MHCB	✗	✓	✓	40.061	0.966	7.202	71.327
MHCB-1	✓	✓	✓	40.072	0.966	7.210	75.845
MHCB-2 (ours)	✓	✓	✓	**40.148**	**0.967**	**7.096**	73.499
MHCB-3	✓	✓	✓	40.080	0.966	7.204	72.915

Table 1: Ablation experiments on Potsdam. Black bold font marks the best performance.

Depth	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓
2	39.548	0.965	7.926	77.036	0.039	0.038
3	39.986	0.965	7.382	75.545	0.030	0.039
4 (ours)	**40.148**	**0.967**	**7.096**	**73.499**	**0.029**	**0.036**

Table 2: Different UNet++ depth on Potsdam.

the origina 2D image feature scanning from four different horizontal directions to six scanning directions, adding two diagonal scanning directions to capture data information more comprehensively. Then, we integrated and flattened the 2D feature maps into 1D sequences, and the long-range dependency of each sequence is computed by Eq. 10. Finally, the results are combined into a 2D feature map as the final result of MUB moudle (Guo et al. 2024). As shown the red box in Fig. 4. By utilizing complementary 1D traversal paths, the 2D-SSM framework allows each pixel to efficiently collect information from other pixels in multiple directions. Thus, it is convenient to establish the global receptive fields in 2D space (Yue Liu 2024).

Loss Function

We optimize our network using L_1 loss, which calculates the sum of absolute differences between the predicted and ground truth images as follows:

$$L_1(y, \hat{y}) = \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{11}$$

where n denotes the number of batch training samples, y_i represents the corresponding label image, and $\hat{y}_i$ stands for the output image generated by MFmamba.

Configuration	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓
3 × 3 patch	39.266	0.966	7.828	75.391	0.039
w/o MUB	39.794	0.965	7.679	**68.398**	0.034
MUB (⑤ scan)	40.071	0.966	7.185	73.530	0.031
MUB (⑥ scan)	40.007	0.966	7.340	74.477	0.032
2 × 2 MUB (ours)	**40.148**	**0.967**	**7.096**	73.499	**0.029**

Table 3: Different scanning methods and patch settings on Potsdam.

Experiments and Analysis

Dataset and Setting

We evaluated MFmamba on five remote sensing datasets: NWPU, Potsdam, QuickBird, GF2, and RSSCN7. The depth of d UNet++ (Zhou et al. 2018) for MFmamba was set at 4, and the growth rate of g was 32. In all experiments, the batch size was 1, and the number of training epochs was 32. The random seed was set to 10. We optimized the proposed MFmamba using the Adam optimizer with a learning rate of 1×10^{-4}, and adjusted the learning rate every 10 iterations using StepLR. β_1 = 0.9 and β_2 = 0.99 were chosen. All experiments were based on PyTorch, and all models were trained by a 80 GB memory GPU. For the SR x2 task, we downsampled the original PAN image to 128×128 as the input. For the SR x4, we downsampled the original PAN image to 64×64 as the input. In joint SR and colorization tasks, we downsampled the original PAN image to 128×128 as the input.

Ablation Study

To verify the performance of the proposed modules (MHCB, DPA, and MUB), we designed an ablation experiment on the Potsdam dataset for the joint SR and spectral recovery task of remote sensing images. The SR scale factor was set to x2 and all experiments used the same configuration. MHCB, DPA, and MUB are independent of each other. First, we test the performance of the DPA module and MUB module. As seen in Table 1, without the DPA module and MUB module led to degraded performance. This demonstrates that the DPA and MUB module helps the model focus on critical information, preserves key image features, improves the upsampling process, and enhances overall performance. Second, we estimated the impact of the MHCB module. Our model employed two MHCBs for initial feature extraction. To determine the optimal number, we conducted an ablation study using 1, 2, and 3 MHCBs, finally finding that using two MHCBs brings the best performance, as shown in the third to sixth rows of Table 1.

We also evaluated the performance of different UNet++ (Zhou et al. 2018) depths in this study. We test the commonly used two-layer, three-layer, and four-layer UNet++ architectures. Table 2 shows that the four-layer network we adopted demonstrated superior performance, excelling in feature extraction and recovery. In addition. We used inputs with 4 patches and 9 patches in MUB to verify the impact of different numbers of patches on the model's performance. We also conducted ablation experiments on the two scanning directions (referred to as ⑤ and ⑥) designed by us in the 2D-SSM module of MUB, in addition to the traditional four horizontal scanning directions. All results are shown in the Table 3. It can be observed that the 4-patch configuration used in our model achieves better performance than the 9-patch configuration. Moreover, the MUB model cannot achieve optimal performance when only the ⑤ or ⑥ scanning direction is used. The model's performance is effectively improved when both scanning directions are incorporated, demonstrating the effectiveness of our proposed two scanning directions.

Experimental Comparison and Discussions

Joint SR and Spectral Recovery Our primary goal is to achieve joint PAN image SR and spectral recovery for remote sensing images. We compare our model with some existing models for joint SR and colorization, including RSI (Feng et al. 2022), CSRDNN (Feng et al. 2021a), CASR (Liu et al. 2022), and MBPRR (Jin et al. 2024), using a x2 scale factor for the SR task. HAT (Chen et al. 2023) + CIR (Feng et al. 2021b) means that HAT is used first for SR operation and then the CIR is used for colorization. CIR (Feng et al. 2021b) + HAT (Chen et al. 2023) means using the CIR colorization model first and then HAT SR model. SwinIR (Liang et al. 2021) + CIR (Feng et al. 2021b) and CIR (Feng et al. 2021b) + SwinIR (Liang et al. 2021) have the same meaning as above. Comparative experiments and analyses were conducted on two datasets to thoroughly assess the generalization capabilities and performance of MFmamba across different datasets. The experimental results on Potsdam and NWPU are presented in Tables 4, experimental results on QuickBird, GF2 and RSSCN dataset in the Appendix for details. As shown in these tables, while MFmamba may be slightly inferior in a few metrics, it achieves the best overall performance. This demonstrates MFmamba's superiority in jointly accomplishing SR and spectral recovery tasks. It also indicates that MFmamba excels at extracting details and critical information in remote sensing images, enhancing feature recovery performance.

The convergence curves of the loss function values of the proposed MFmamba with the number of epochs on five datasets as shown in Fig. 6. To better illustrate our experimental results, we selected one image from each dataset to compare the performance of several existing methods with our proposed MFmamba, as shown in Fig. 5. The 'Error Intensity' in the last column represents the error intensity between the generated results of our model and the label. The color changes gradually from blue to red. Among them, blue represents the area with relatively small errors, while yellow and red represent the areas with relatively large errors.

Super-Resolution For the image SR task, we compared MFmamba with bicubic interpolation as the baseline and comparison methods including RSI (Feng et al. 2022), MDCN (Li et al. 2020), HAT (Chen et al. 2023), SwinIR (Liang et al. 2021), and MulSR (Li et al. 2023b). We conducted comparative experiments and analyses exclusively on the Potsdam dataset. We performed experiments using x2 and x4 upscale factors for the SR task. The results are shown in Table 5. To illustrate the result further, we select two images from the Potsdam dataset to present the results of the comparison methods and MFmamba on x2 and x4 upscale factors, as shown in Fig. 7. These results demonstrate that MFmamba excels at jointly performing SR and spectral recovery and achieves outstanding performance when focused solely on SR tasks. MFmamba performs better on images SR of trees, roofs, buildings, etc. (see the red box in Fig. 7). In Fig. 7, the last row is the error intensity heatmap of the SR results generated by various comparative methods for the image of SRx4 and the Label image, which can intuitively display the error distribution and detail differences of

Method	Potsdam Dataset						NWPU Dataset					
	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓
MBPRR(Jin et al. 2024)	34.953	0.943	28.653	103.165	**0.023**	0.064	28.504	0.913	101.887	128.501	0.077	0.096
CASR(Liu et al. 2022)	32.685	0.946	51.054	105.339	0.029	0.079	28.200	0.906	109.594	120.439	**0.049**	0.105
RSI(Feng et al. 2022)	34.857	0.910	25.760	99.911	0.054	0.052	32.354	0.913	59.534	**104.111**	0.074	**0.068**
CSRDNN(Feng et al. 2021a)	31.461	0.954	56.084	96.923	0.081	0.049	31.478	0.905	62.645	127.096	0.087	0.070
HAT(Chen et al. 2023)+CIR(Feng et al. 2021b)	32.853	0.958	34.774	103.314	0.062	0.039	22.760	0.533	393.884	125.429	0.299	0.279
CIR(Feng et al. 2021b)+HAT(Chen et al. 2023)	27.777	0.720	375.405	113.637	0.147	0.082	24.519	0.672	257.197	126.742	0.163	0.467
SwinIR(Liang et al. 2021)+CIR(Feng et al. 2021b)	32.838	0.957	34.907	103.308	0.062	0.038	22.789	0.536	390.958	125.484	0.198	0.270
CIR(Feng et al. 2021b)+SwinIR(Liang et al. 2021)	27.777	0.720	375.272	113.547	0.147	0.082	24.518	0.672	257.232	126.674	0.163	0.469
Ours	**40.148**	**0.966**	**7.096**	**73.499**	0.029	**0.036**	**33.183**	**0.927**	**46.606**	106.451	0.066	0.074

Table 4: Contrast experiments results on Potsdam and NWPU, black bold font marks the best performance.

Method	x2						x4					
	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓
Bicubic	35.591	0.908	21.307	112.893	0.049	0.130	31.145	0.788	59.929	117.396	0.082	0.410
RSI(Feng et al. 2022)	37.525	0.936	13.626	**66.786**	0.038	0.055	31.886	**0.810**	49.420	**68.907**	0.072	0.189
MDCN(Li et al. 2020)	34.718	0.859	313.392	99.817	0.096	0.0706	26.187	0.668	314.211	101.733	0.091	0.071
HAT(Chen et al. 2023)	34.963	0.861	311.239	106.593	0.095	0.0705	30.209	0.752	328.867	113.026	0.121	0.073
SwinIR(Liang et al. 2021)	34.921	0.861	311.239	106.423	0.095	0.0704	30.006	0.746	331.536	110.928	0.122	0.073
MulSR(Li et al. 2023b)	25.593	0.903	343.201	144.138	0.089	0.075	23.975	0.885	393.033	102.323	0.130	0.116
Ours	**38.458**	**0.954**	**10.129**	77.288	**0.035**	**0.033**	**32.257**	0.809	**46.230**	95.163	**0.072**	**0.019**

Table 5: Experimental results of SR at different up_scale factors on Potsdam, black bold font marks the best performance.

Figure 5: Visual results of MFmamba and benchmark methods on five different datasets.

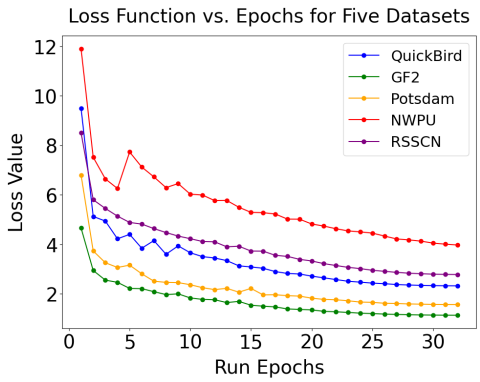

Figure 6: The curves of the loss function with the number of epochs on the five datasets.

Method	PSNR↑	SSIM↑	MSE↓	MAE↓	SAM↓	LPIPS↓
RSI(Feng et al. 2022)	33.135	0.983	47.805	103.962	0.068	0.052
CIR(Feng et al. 2021b)	31.990	0.957	42.428	95.134	0.069	0.077
SEGAN(Wu et al. 2019)	32.513	0.632	396.840	122.054	0.083	0.640
Wu(Wu et al. 2019)	31.558	0.537	458.715	127.489	0.085	0.634
Iizukas(Iizuka et al. 2016)	11.072	0.190	507.459	125.625	0.465	0.110
Huang(Huang et al. 2021)	32.025	0.968	72.866	105.335	0.066	0.175
Ours	**35.569**	**0.984**	**28.477**	**88.036**	**0.052**	**0.048**

Table 6: Colorization experimental results on Potsdam.

different methods.

Spectral Recovery The propose image colorization method was compared with the most advanceding algorithms include RSI (Feng et al. 2022), CIR (Feng et al. 2021b), SEGAN (Wu et al. 2019), Wu (Wu et al. 2019), Iizukas (Iizuka et al. 2016) and Huang (Huang et al. 2021) on Potsdam. The results are presented in Table 6. Although image colorization is an additional function of MFmamba model, it still performs better than other methods, and the generated results have more visual advantages. We selected five images from the Potsdam dataset to show the results of several comparison methods adopted and the results of MFmamba, as shown in Fig. 8. MFmamba performed the best in image colorization tasks, generating images with more realistic and natural colors, rich details, and very close to real labeled images. Both color transitions and edge details have been effectively restored, demonstrating excellent spectral recovery ability and outstanding visual effects.

Comparison Between Mamba and Transformer We conducted experiments by replacing the mamba module in the proposed MUB module with the Transform module to compare and analyse the differences between mamba and Transform in terms of running time, number of parameters and memory usage on the Potsdam dataset. The experimental results are shown in Fig. 9, from which we can see that the mamba-based model we designed runs faster and has fewer parameters then Transformer, demonstrating the powerful advantages of Mamba.

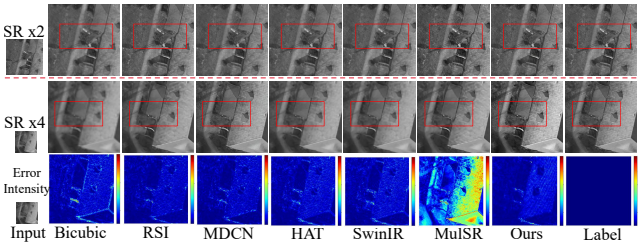

Figure 7: Visual SR results of MFmamba and comparison methods on Potsdam dataset.

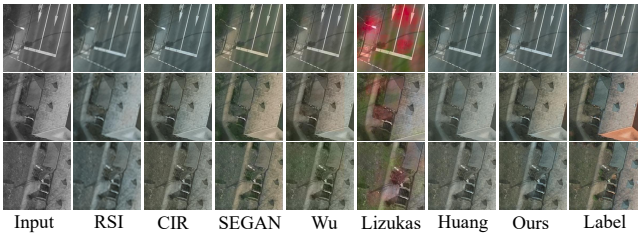

Figure 8: Visual colorization results of MFmamba and benchmark methods on Potsdam dataset.

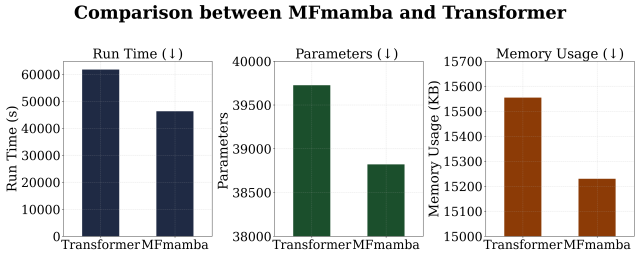

Figure 9: The bar chart for comparison between Mamba and Transformer on Potsdam.

Conclusion

We proposed a multi-function resolution recovery model called MFmamba to enhance the spatial and spectral resolution of remote sensing images. MFmamba employs a state space model to develop the MUB module, which was used to enhance the model's contextual modeling capabilities. The proposed MHCB extracts multiscale features, effectively conveying detailed image information. We introduced a novel attention mechanism, DPA, to enable the model to focus on the most important feature channels. Extensive experiments and visualization results demonstrate that MFmamba, combined with these modules, achieves impressive results in spatial resolution recovery, spectral resolution recovery, and joint recovery in PAN images. Furthermore, our method outperforms most existing super-resolution and colorization models. Extensive experiments also validate the effectiveness of integrating multiple tasks within a single framework.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (Nos. 62261060), Yunnan Fundamental Research Projects (Nos. 202503AG380006, 202301AW070007, 202301AU070210, 202401AT070470), and Yunnan Province Expert Workstations (202305AF150078), Yunnan Province Special Project (Grant No.202403AP140021), and Xingdian Talent Project in Yunnan Province of China.

References

An, S.; Liao, Q.; Lu, Z.; et al. 2024. Dual Correlation Network for Efficient Video Semantic Segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(3): 1572–1585.

Anwar, S.; Tahir, M.; Li, C.; et al. 2025. Image colorization: A survey and dataset. *Information Fusion*, 114: 102720.

Chen, X.; Wang, X.; Zhou, J.; et al. 2023. Activating more pixels in image super-resolution transformer. In *ICCV*, 22367–22377.

Feng, J.; Jiang, Q.; Jin, X.; et al. 2021a. CSRDNN: An Integrated Scheme for Single Satellite Image Colorization and Super-Resolution Using Deep Neural Networks. In *IJCNN*, 1–9.

Feng, J.; Jiang, Q.; Jin, X.; et al. 2021b. Remote sensing image colorization based on deep neural networks with multiscale residual receptive filed. *Journal of Computer-Aided Design & Computer Graphics*, 33(11): 1658–1667.

Feng, J.; Jiang, Q.; Tseng, C.-H.; et al. 2022. A deep multitask convolutional neural network for remote sensing image super-resolution and colorization. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15.

Guo, H.; Li, J.; Dai, T.; et al. 2024. Mambair: A simple baseline for image restoration with state-space model. *arXiv preprint arXiv:2402.15648*.

Huang, S.; Jin, X.; Jiang, Q.; et al. 2021. A fully-automatic image colorization scheme using improved CycleGAN with skip connections. *Multimedia Tools and Applications*, 80(17): 26465–26492.

Iizuka, S.; et al. 2016. Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics*, 35(4): 1–11.

Jiang, Q.; Wang, Q.; Miao, S.; et al. 2025. SRColorNet: Multipath attention aggregated and mask enhanced network for the super resolution and colorization of panchromatic image. *Expert Systems with Applications*, 276: 127091.

Jin, X.; Liu, L.; Ren, X.; et al. 2024. A Restoration Scheme for Spatial and Spectral Resolution of Panchromatic Image Using Convolutional Neural Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 3379–3393.

Kumar, M.; Weissenborn, D.; et al. 2021. Colorization transformer. *arXiv preprint arXiv:2102.04432*.

Li, H.; Deng, W.; Zhu, Q.; et al. 2024. Local-Global Context-Aware Generative Dual-Region Adversarial Networks for Remote Sensing Scene Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Li, H.; Yang, Y.; Chang, M.; et al. 2022a. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479: 47–59.

Li, J.; Fang, F.; Li, J.; et al. 2020. MDCN: Multi-scale dense cross network for image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(7): 2547–2561.

Li, J.; Zheng, K.; Gao, L.; et al. 2025. Enhanced Deep Image Prior for Unsupervised Hyperspectral Image Super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*.

Li, J.; Zheng, K.; Li, Z.; et al. 2023a. X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*.

Li, J.; Zheng, K.; Yao, J.; et al. 2022b. Deep unsupervised blind hyperspectral and multispectral data fusion. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Li, Q.; et al. 2023b. Multi-scale factor joint learning for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*.

Liang, J.; Cao, J.; Sun, G.; et al. 2021. Swinir: Image restoration using swin transformer. In *ICCV*, 1833–1844.

Liang, Z.; Li, Z.; Zhou, S.; et al. 2024. Control Color: Multimodal Diffusion-based Interactive Image Colorization. *arXiv preprint arXiv:2402.10855*.

Liu, L.; Jiang, Q.; Jin, X.; et al. 2022. CASR-Net: A color-aware super-resolution network for panchromatic image. *Engineering Applications of Artificial Intelligence*, 114: 105084.

Ma, C.; Du, L.; Zhuo, L.; et al. 2024. MPLA-Net: Multiple Pseudo Label Aggregation Network for Weakly Supervised Video Salient Object Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5): 3905–3918.

Park, K.; Soh, J. W.; and Cho, N. I. 2025. Efficient Attention-Sharing Information Distillation Transformer for Lightweight Single Image Super-Resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39.

Shafiq, H.; and Lee, B. 2024. Transforming color: A novel image colorization method. *Electronics*, 13(13): 2511.

Wang, J.; Yue, Z.; Zhou, S.; et al. 2024. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 1–21.

Wu, M.; Jiang, Q.; Jin, X.; et al. 2021. Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space. *The Visual Computer*, 37: 1707–1729.

Wu, M.; Jin, X.; Jiang, Q.; et al. 2019. Remote sensing image colorization based on multiscale SEnet GAN. In *CISP-BMEI*, 1–6.

Xie, C.; Zhang, X.; Li, L.; et al. 2025. MAT: Multi-Range Attention Transformer for Efficient Image Super-Resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–13.

Yue Liu, Y. Z., Yunjie Tian. 2024. VMamba: Visual State Space Model. *arXiv preprint arXiv:2401.10166*.

Zhang, X.; Yu, H.; Qin, Y.; et al. 2024. Video-Based Multi-Camera Vehicle Tracking via Appearance-Parsing Spatio-Temporal Trajectory Matching Network. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10): 10077–10091.

Zhao, Y.; Po, L.-M.; Cheung, K.-W.; Yu, W.-Y.; and Rehman, Y. A. U. 2021. SCGAN: Saliency Map-Guided Colorization With Generative Adversarial Network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(8): 3062–3077.

Zhou, Z.; Rahman Siddiquee, M. M.; Tajbakhsh, N.; et al. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *DLMIA*, 3–11.

Zhu, F.; Wang, C.; Zhu, B.; et al. 2023. An improved generative adversarial networks for remote sensing image super-resolution reconstruction via multi-scale residual block. *The Egyptian Journal of Remote Sensing and Space Science*, 26(1): 151–160.