remote sensing

MDPI

Article

FireMambaNet: A Multi-Scale Mamba Network for Tiny Fire Segmentation in Satellite Imagery

Bo Song [1,2], Bo Li [2], Hong Huang [2], Zhiyong Zhang [2], Zhili Chen [1,3,4,*], Tao Yue [2] and Yun Chen [2]

[1] College of Environmental Science and Engineering, Guilin University of Technology, Guilin 541006, China; bosong@glut.edu.cn
[2] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China; 2120242157@glut.edu.cn (B.L.); hh@glut.edu.cn (H.H.); 2120242221@glut.edu.cn (Z.Z.); yuetao@glut.edu.cn (T.Y.); chenyun@glut.edu.cn (Y.C.)
[3] Guangxi Key Laboratory of Environmental Pollution Control Theory and Technology, Guilin University of Technology, Guilin 541006, China
[4] Key Laboratory of Carbon Emission and Pollutant Collaborative Control of Education, Department of Guangxi Zhuang Autonomous Region, Guilin University of Technology, Guilin 541006, China
* Correspondence: 2017075@glut.edu.cn

Highlights

What are the main findings?

- A multi-scale Mamba-based segmentation network, termed FireMambaNet, is proposed to address the challenges of extremely small-scale, sparsely distributed fire points in complex satellite remote sensing imagery.
- Three key modules (CG-RSU, M2AM, and FDCM) significantly enhance the representation of tiny fire targets (less than 5 pixels), achieving state-of-the-art performance on the Oceania and Asia4 datasets.

What are the implications of the main findings?

- The proposed lightweight framework shows strong robustness under extreme pixel sparsity and cross-region distribution differences.
- The integration of multi-scale Mamba-based long-range dependency modeling with CG-RSU provides a new perspective for tiny-target segmentation in remote sensing.

Abstract

Satellite remote sensing plays an essential role in wildfire monitoring due to its large-scale observation capability. However, fire targets in satellite imagery are typically extremely small, sparsely distributed, and embedded in complex backgrounds, making accurate segmentation highly challenging for existing methods. To address these challenges, this paper proposes a multi-scale Mamba-based network for tiny fire segmentation, named FireMambaNet. The network adopts a nested U-shaped encoder-decoder architecture, primarily consisting of three modules: the Cross-layer Gated Residual U-shaped module (CG-RSU), the Fire-aware Directional Context Modulation module (FDCM), and the Multi-scale Mamba Attention Module (M2AM). The CG-RSU, as the core building block, adaptively suppresses background redundancy and enhances weak fire responses by extracting multi-scale features through cross-layer gating. The FDCM explicitly enhances the network's ability to perceive anisotropic expansion features of fire points, such as those along the wind direction and terrain orientation, by modeling multi-directional context. The M2AM model employs a Mamba state-space model to suppress background interference through global context modeling during cross-scale feature fusion, while enhancing

Academic Editor: Xiaoyang Zhang

Received: 11 February 2026
Revised: 9 March 2026
Accepted: 27 March 2026
Published: 29 March 2026

consistency among sparsely distributed tiny fire targets. In addition, experimental validation is conducted using two subsets from the Active Fire dataset, which have significant pixel-level sparse features: Oceania and Asia4. The results show that the proposed method significantly outperforms various mainstream CNN, Transformer, and Mamba baseline models on both datasets. It achieves an IoU of 88.51% and F1 score of 93.76% on the Oceania dataset, and an IoU of 85.65% and F1 score of 92.26% on the Asia4 dataset. Compared to the best-performing CNN baseline model, the IoU is improved by 1.81% and 2.07%, respectively. Overall, the FireMambaNet demonstrates significant advantages in detecting tiny fire points in complex backgrounds.

Keywords: satellite imagery; fire point segmentation; tiny objects; deep learning; Mamba

1. Introduction

Wildfires are among the most severe and frequent natural disasters, posing significant threats to climate change, agricultural production, and human life and property safety [1–3]. Therefore, proactive and accurate wildfire monitoring is of critical importance, as it can effectively reduce the losses caused by fire events.

Satellite remote sensing has attracted extensive attention due to its capability for long-term and large-scale observation, providing crucial information for wildfire monitoring, including fire detection, burn scar mapping, and fire impact assessment [4,5]. At present, remote sensing imagery acquired from satellites such as Landsat, Sentinel, Gaofen, and Himawari has been widely used for fire detection tasks [6–8]. As a key component of satellite-based wildfire monitoring, fire point segmentation enables timely and accurate identification of active fire regions. However, fire segmentation in satellite imagery remains highly challenging, mainly due to the extremely small target scale, sparse spatial distribution, and complex and dynamically changing backgrounds.

Existing fire point segmentation methods can be broadly categorized into traditional methods and deep learning-based methods. Traditional fire detection approaches primarily rely on spectral, spatial, and temporal characteristics of remote sensing imagery, identifying anomalous fire signals through threshold-based decision rules [9,10]. Early studies focused on the distinctive radiative responses of fires in the mid- and far-infrared bands and developed fire detection algorithms based on fixed spectral thresholds [11,12], which were widely applied to sensors such as AVHRR and MODIS [13–16]. Subsequently, considering the spatial heterogeneity and temporal variability of fire backgrounds, researchers proposed multi-temporal thresholding and spatial context-based methods. The former detects fire anomalies by analyzing brightness temperature or radiance differences across time [17–21], while the latter identifies fire pixels by comparing statistical differences in thermal properties, reflectance, and other features between candidate pixels and their surrounding background regions [22–27]. These methods typically rely on local statistical measures (e.g., mean, standard deviation, and variance) to determine adaptive thresholds and improve robustness under varying surface conditions [28–30]. Nevertheless, traditional approaches remain sensitive to threshold selection and background interference, leading to limited generalization performance in complex scenes and weak fire scenarios [31].

In recent years, the rapid advancement of deep learning has significantly promoted the application of artificial intelligence in remote sensing and has gradually become a dominant research direction for active fire detection [32]. Compared with traditional methods, deep learning models can automatically learn hierarchical discriminative features through end-to-end training, alleviating the limitations of handcrafted feature design and

fixed threshold strategies. For instance, Pereira et al. [33] constructed a large-scale remote sensing dataset for active fire detection based on Landsat-8 imagery and validated the effectiveness of convolutional neural networks (CNNs) using the U-Net architecture [34]. Teymoor et al. [35] proposed the Fire-Net framework for automatic identification of active fire regions. Kang et al. [31] combined random forests (RF) and CNNs for early-stage active fire detection. Fang et al. [36] further proposed the FPS-U^{2}Net model, which enhances the original U^{2}Net [37] by introducing multi-feature aggregation strategies, achieving improved performance in detecting single tiny fire targets. However, this approach mainly incorporates multi-layer aggregation modules in the encoding and decoding stages, resulting in limited global contextual modeling capability and a tendency to miss detections in complex scenes with multiple tiny fire targets.

To overcome the limitations of CNNs in modeling long-range dependencies, several studies have introduced Transformer architectures, which treat two-dimensional images as one-dimensional sequences to enhance global feature representation [38–40]. For example, Zhang et al. [41] proposed a hybrid CNN-Transformer framework for fire detection, significantly improving performance in complex scenarios. However, Transformers generally require substantial computational resources and incur high inference latency, making them less suitable for time-sensitive active fire monitoring applications. Recently, the Mamba state-space model has emerged as an efficient alternative, offering strong global modeling capability comparable to Transformers while significantly reducing computational overhead. Mamba-based models have demonstrated promising performance across various segmentation tasks [42–44]. Motivated by these advantages, this paper integrates U^{2}Net with the Mamba architecture and proposes a multi-scale Mamba network for tiny fire segmentation, named FireMambaNet.

Within the FireMambaNet framework, the backbone RSU module of U^{2}Net is redesigned to form a Cross-layer Gated Residual U-shaped module (CG-RSU), which enhances the discriminative feature extraction capability for tiny fire targets with weak responses. To further address the weak local responses and spatial sparsity of multiple tiny fire pixels, a Multi-scale Mamba Attention Module (M2AM) is embedded into both the encoding and decoding stages, strengthening the network's ability to model global consistency across scales. Moreover, considering that wildfire propagation is influenced by factors such as wind direction and terrain, resulting in pronounced directional patterns in remote sensing imagery, a fire-oriented directional context modulation module is introduced to explicitly model multi-directional contextual information, thereby further improving fire pixel segmentation accuracy.

In summary, the main contributions of this paper are as follows:

(1) An active fire detection network, named FireMambaNet, is proposed for monitoring extremely small fire targets in satellite imagery under complex background conditions.
(2) To address the insufficient capability of existing networks in extracting features of tiny fire targets, a Cross-layer Gated Residual U-block (CG-RSU) is designed, which enhances the representation of fire target features through cross-layer feature fusion and adaptive gating mechanisms.
(3) To alleviate the issue of inconsistent local responses in scenarios where multiple tiny fire targets coexist, a Multi-scale Mamba Attention Module (M2AM) is introduced, enabling the network to establish global contextual consistency among multiple sparsely distributed fire targets while suppressing interference from background areas.
(4) A Fire-aware Directional Context Modulation (FDCM) is proposed to explicitly model multi-directional contextual information of fire targets, thereby further improving

the recognition accuracy of fire targets exhibiting pronounced directional diffusion characteristics.

2. Materials and Methods

2.1. Overall Architecture

The overall architecture of the FireMambaNet network proposed in this paper is shown in Figure 1. The FireMambaNet adopts a nested U-shaped encoder-decoder framework and consists of three core components: the Cross-layer Gated Residual U-block (CG-RSU), the Fire-aware Directional Context Modulation (FDCM), and the Multi-scale Mamba Attention Module (M2AM). The network takes raw satellite imagery as input, and its encoder is composed of six cascaded CG-RSU modules to progressively extract multi-scale contextual features via downsampling operations. Specifically, CG-RSU7, CG-RSU6, CG-RSU5, and CG-RSU4 are employed in Stage 1 to Stage 4 to model low- and mid-level contextual features, while lightweight CG-RSU4F structures are adopted in Stage 5 and Stage 6 for high-level semantic feature extraction. To enhance the network's capability to perceive directional expansion characteristics of the fire point, FDCM is inserted after Stage4 of the encoder to perform multi-directional context modeling on mid-to-high-level features, thereby strengthening directional structural responses. During the decoding, the decoder adopts a structure symmetrical to the encoder and fuses contextual features through skip connections. Notably, the proposed M2AM is embedded into the skip-connection pathway to enhance the global consistency of extremely small and spatially sparse fire points across large-scale scenes. Furthermore, a deep supervision strategy with six auxiliary output branches (s1–s6) is employed to guide multi-scale feature learning during training. Ultimately, the network generates fire point segmentation results, enabling robust detection of extremely small satellite fire targets under complex backgrounds.

2.2. Cross-Layer Gated Residual U-Block (CG-RSU)

To address the challenge of fire points being extremely small and difficult to effectively characterize using traditional convolutional neural networks in satellite remote sensing imagery, this paper proposes the Cross-Layer Gated Residual U-Block (CG-RSU). Which can adaptively select feature responses at different scales, effectively suppressing redundant noise background information while significantly enhancing the representation capability of faint fire points. As shown in Figure 2, CG-RSU adopts a nested encoder-decoder architecture, serving as the core building block of FireMambaNet.

In the encoding stage, the CG-RSU progressively extracts contextual information from local to global scales through a multi-layer downsampling structure, effectively capturing the spatial and semantic difference features between the fire points and their surrounding background. This multi-feature extraction process can be defined as:

$$X_i = f_i(\mathcal{P}(X_{i-1})), i = 1, \ldots, L \tag{1}$$

where X_i is the feature extracted from the $i - th$ layer, $\mathcal{P}(\cdot)$ represents the max pooling operation, $f_i(\cdot)$ is the convolution mapping, and L is the number of layers dynamically adjusted by the network.

In the decoding stage, CG-RSU performs progressive upsampling and concatenates with the corresponding scale encoding features, achieving a full fusion of high-level semantic information and low-level fine spatial information. This process can be defined as:

$$X_i' = Concat(\mathcal{U}(F_{i+1}), X_i) \tag{2}$$

where X_i' represents the fused feature of the $i-th$ layer, and $\mathcal{U}(\cdot)$ represents the upsampling operation.

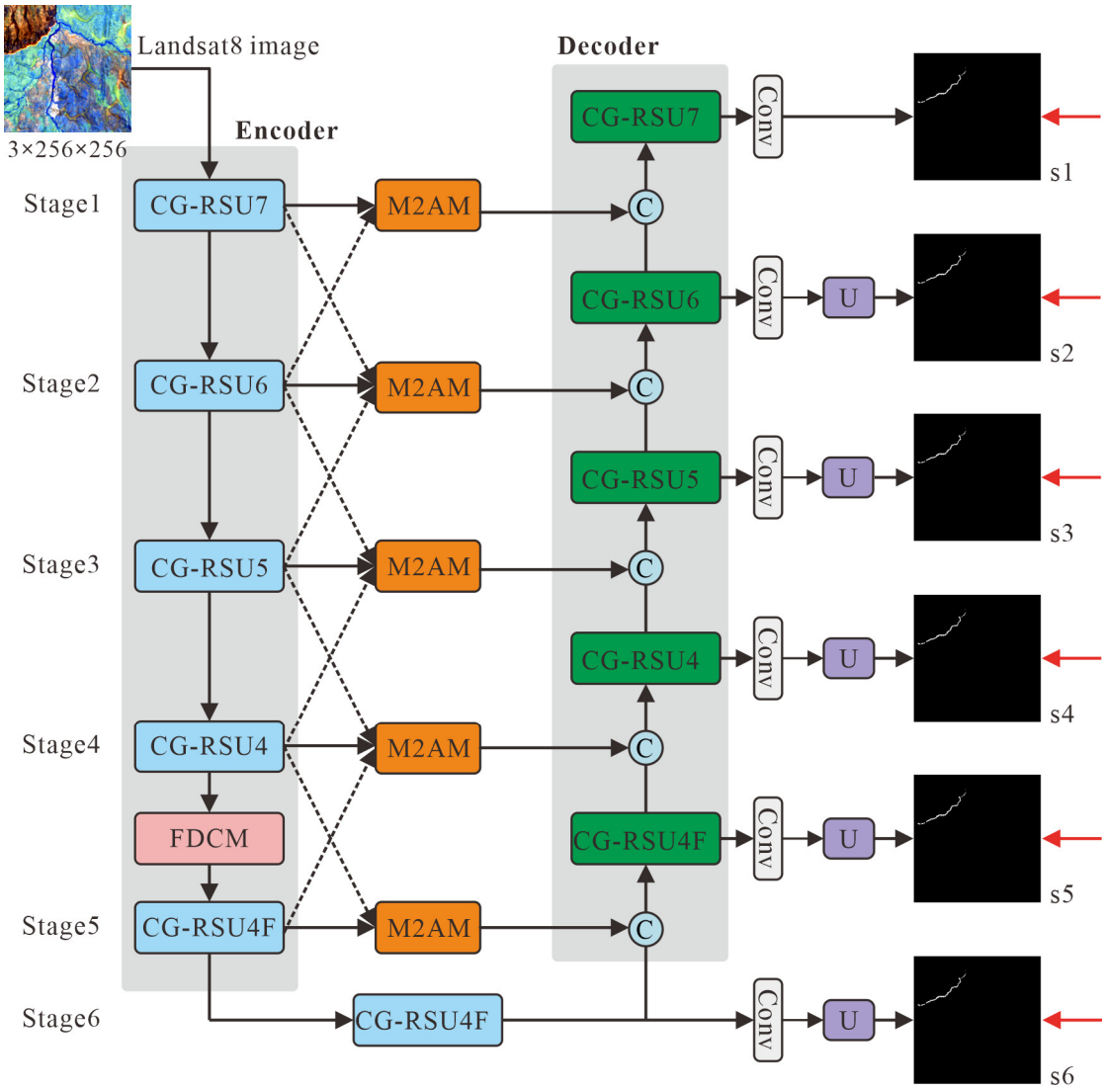

Figure 1. Overall structure of the network.

Unlike traditional RSU or single-layer gating structures, this paper innovatively designs a cross-layer gating mechanism. Specifically, CG-RSU establishes a hierarchical information passing path for gating between adjacent decoding layers, allowing high-level semantic features to explicitly guide the feature selection process at lower levels. This continuously strengthens the discriminative information related to fire points and suppresses background redundant responses during the multi-scale fusion process. The cross-layer gating process can be defined as:

$$X_i^G = X_i' \otimes (G_i + G_{i+1}) \tag{3}$$

where X_i^G represents the gated-modulated features, $\otimes$ denotes channel-wise multiplication, and G_i represents the gating weight. The computation process can be defined as:

$$G_i = \sigma(W_i GAP(X_i')) \tag{4}$$

where $GAP(\cdot)$ represents global average pooling, W_i denotes the 1×1 convolution weights, and $\sigma(\cdot)$ represents the Sigmoid function.

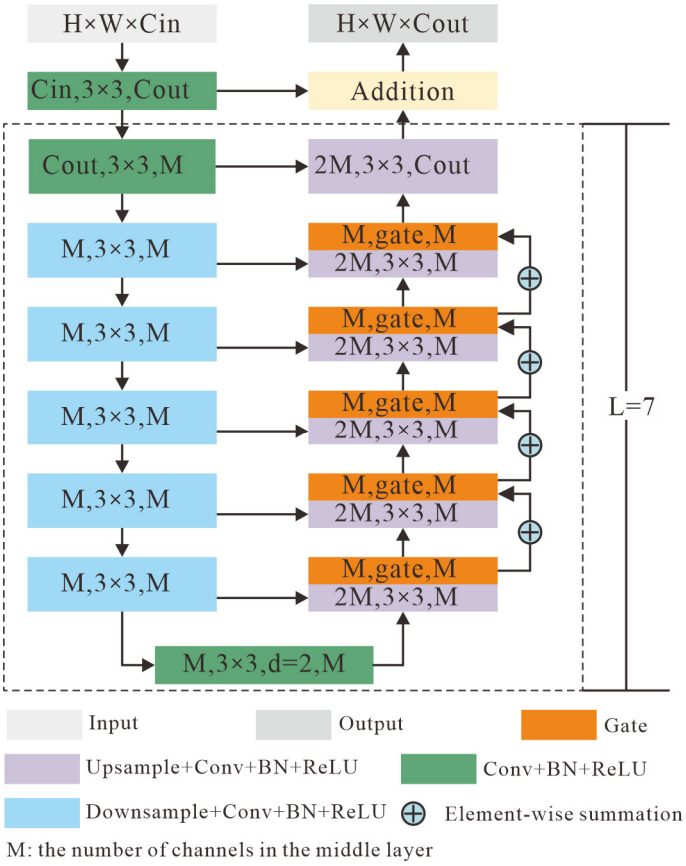

Figure 2. Structure of the CG-RSU. Cin and Cout denote the input and output channel numbers, respectively. M represents the intermediate feature channel dimension within the CG-RSU, and 2M denotes channel expansion to twice this dimension. Gate indicates the cross-layer gating mechanism. The symbol $\oplus$ indicates element-wise summation. L denotes the depth of the nested U-shaped structure. Down-sampling is implemented using a stride-2 convolution layer, while up-sampling is performed via bilinear interpolation followed by a 3×3 convolution.

2.3. Fire-Aware Directional Context Modulation (FDCM)

To enhance the network's ability to recognize the distinct directional diffusion characteristics of fire points in spatial configurations, this paper specifically designs a Fire-aware Directional Context Modulation (FDCM). Through direction-sensitive context modeling, this module significantly improves the network's perception of weak fire points and their surrounding spatial contextual information. As shown in Figure 3, the FDCM consists of three complementary context modeling branches. The first branch uses global average pooling combined with a 1×1 convolution to perform global context compression and channel recalibration on the feature map, capturing the global response strength and semantic prior constraints of the fire point across the entire image. This process can be defined as:

$$X_{s4}^{g} = \mathcal{U}(f_{1\times1}(GAP(X_{s4}))) \tag{5}$$

where X_{s4}^{g} represents the features after modeling by the first branch, X_{s4} represents the features output by the CG-RSU4 module, $GAP(\cdot)$ represents global average pooling, $f_{1\times1}(\cdot)$ represents the 1×1 convolution mapping, and $\mathcal{U}(\cdot)$ represents bilinear interpolation to the original spatial dimensions.

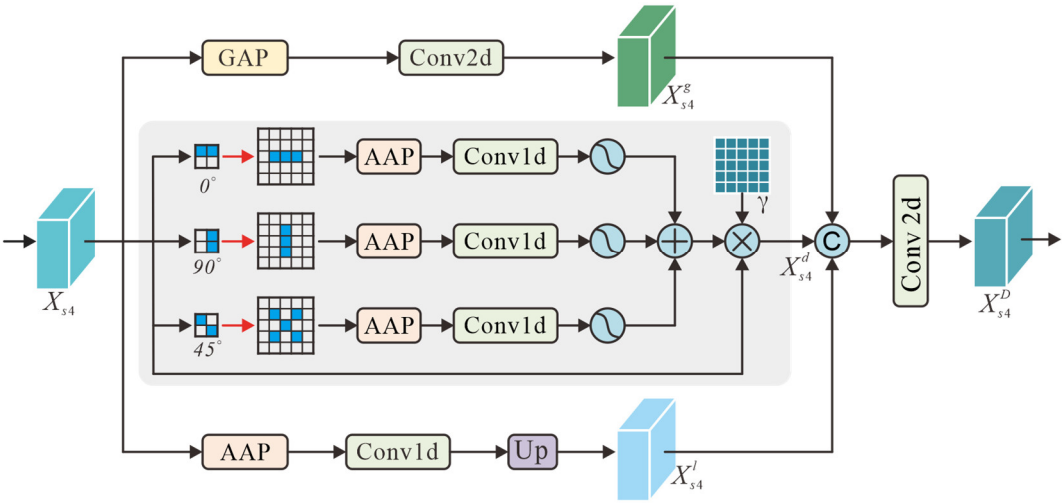

Figure 3. Structure of the FDCM.

The second branch is a direction-aware branch based on Strip Pooling (Strip Pooling Branch), which performs directional aggregation of features along the horizontal (0°), vertical (90°), and diagonal (45°) directions, respectively. It models the strip-shaped spatial dependencies via depth-wise separable 1D convolutions, effectively enhancing the network's capability to perceive the directionally extended features of fire points constrained by wind directions and surface structures, while suppressing the interference of unstructured noise in background regions during this process. Moreover, within the direction-aware branch, strip responses from different directions are fused after normalization via the Sigmoid function. A learnable gating parameter γ is also introduced to achieve adaptive modulation of directional attention. This design helps maintain network stability during the early stages of training and gradually strengthens the guidance of directional context on fire point features during the convergence process. The perception process of this branch can be defined as:

$$X_{s4}^{d} = \gamma\left(\frac{1}{|\Theta|}\sum_{\theta\in\Theta}\sigma_{\theta}(f_{\theta}(AvgPool(R_{\theta}(X_{s4})))) + X_{s4}\right) \tag{6}$$

where X_{s4}^{d} represents the features after modeling by the second branch, γ denotes the learnable gating parameters, Θ represents the set of modeled directions, with a value of $\Theta = \{0°, 45°, 90°\}$; $\sigma_{\theta}(\cdot)$ represents the Sigmoid activation function; $f_{\theta}(\cdot)$ denotes the per-channel 1D convolution along direction θ; $AvgPool(\cdot)$ represents the adaptive average pooling operation along the strip direction θ; and $R_{\theta}(\cdot)$ is the rotation transformation matrix.

The third branch uses local average pooling combined with convolutional operations to complement the modeling of local contextual information within the fire point's neighborhood, in order to compensate for the lack of local detail representation in global and directional modeling. This process can be defined as:

$$X_{s4}^{l} = f_{1\times1}(AvgPool_{3\times3}(X_{s4})) \tag{7}$$

where X_{s4}^{l} represents the features after modeling by the third branch, and $AvgPool_{3\times3}(\cdot)$ represents the local average pooling operation.

Further, the output features X_{s4}^{g}, X_{s4}^{d}, X_{s4}^{l} from the three branches are concatenated along the channel dimension. A 1×1 convolution is then applied for feature compression and fusion, forming a comprehensive representation that complements directional

perception and multi-scale context. Finally, a residual connection is used to add the output features to the original features, producing the output feature X_{s4}^{D}, which effectively enhances the feature representation ability without disrupting the original discriminative information. This process can be defined as:

$$X_{s4}^{D} = X_{s4} + f_{1\times1}\left(Concat\left(X_{s4}^{g}, X_{s4}^{d}, X_{s4}^{l}\right)\right) \tag{8}$$

2.4. Multi-Scale Mamba Attention Module (M2AM)

In complex remote sensing scenes, fire points are typically sparsely distributed and occupy only a very small proportion of pixels, while background regions dominate the spatial structure of the image. To address the challenges caused by extreme pixel-level imbalance and background interference, this paper proposes a Multi-scale Mamba Attention Module (M2AM). As illustrated in Figure 1, M2AM is embedded between the encoder and decoder to facilitate cross-scale feature interaction. By fusing adjacent high-level and low-level features, a Mamba module based on state-space modeling is introduced to establish global contextual dependencies across the entire scene. This design enables the network to enhance consistency among sparsely distributed fire targets while suppressing false responses from background regions with similar spectral characteristics, thereby improving segmentation stability in large-scale complex environments.

As shown in Figure 4a, M2AM mainly consists of three parts: the spatial attention weighting branch, the Mamba global modeling branch, and the feature fusion layer. The spatial attention weighting branch centers around the current scale feature $X_{s(i)}$, and through the introduction of spatial attention guidance from adjacent high-level feature $X_{s(i-1)}$ and low-level feature $X_{s(i+1)}$, it enables effective interaction of cross-scale contextual information. Specifically, spatial alignment is achieved through upsampling or downsampling operations, and the spatial attention mechanism is used to adaptively emphasize potential fire point areas while suppressing complex background interference. This process can be defined as:

$$\begin{cases} X_{s(i)}^{h} = SA\left(\mathcal{D}\left(X_{s(i-1)}\right)\right) \otimes X_{s(i)} \\ X_{s(i)}^{l} = SA\left(\mathcal{U}\left(X_{s(i+1)}\right)\right) \otimes X_{s(i)} \end{cases} \tag{9}$$

where $\mathcal{U}(\cdot)$ and $\mathcal{D}(\cdot)$ represent the upsampling and downsampling operations, $SA(\cdot)$ represents the spatial attention mechanism, and $\otimes$ denotes element-wise multiplication.

The Mamba global modeling branch performs global dependency modeling on the spatial correlation of extremely tiny fire points in large-scale scenes via the state space model. First, to model the spatial correlations of tiny fire points in large-scale scenes, the two-dimensional feature map $X_{s(i)} \in \mathbb{R}^{C\times H\times W}$ is restructured into a sequential form:

$$X_{s(i)} = Reshape\left(X_{s(i)}\right) \in \mathbb{R}^{HW\times C} \tag{10}$$

Then, the Mamba module based on state space modeling is introduced to capture long-range spatial dependencies, defined as:

$$X'_{s(i)} = L\left(SSM\left(\sigma_s\left(f_{1\times1}\left(L\left(X_{s(i)}\right)\right)\right)\right) \circ \sigma_s\left(L\left(X_{s(i)}\right)\right)\right) \tag{11}$$

where $L(\cdot)$ represents the linear mapping layer, $SSM(\cdot)$ represents the state space model operator, $\sigma_s(\cdot)$ represents the SiLU activation function, and $\circ$ represents the Hadamard product.

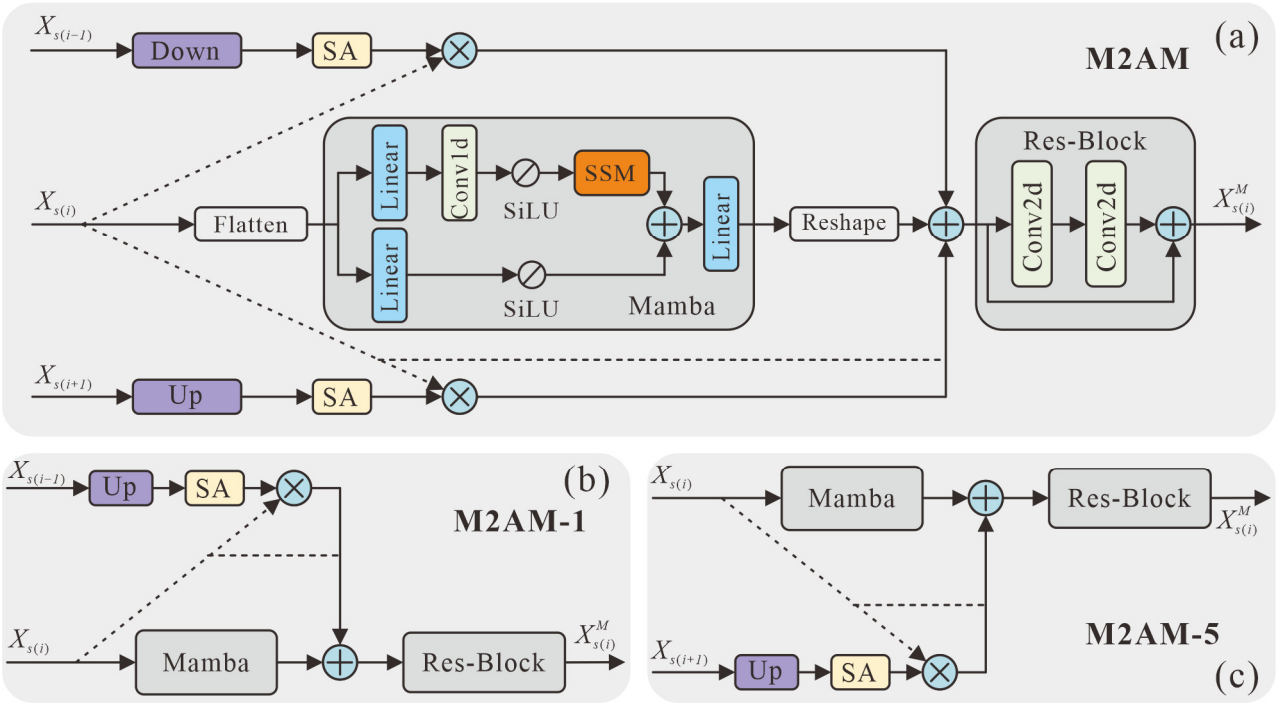

Figure 4. Structure of the M2AM. (**a**) General structure of M2AM, (**b**) M2AM variant at Stage 1 (M2AM-1), (**c**) M2AM variant at Stage 5 (M2AM-5).

Finally, the output features are reconstructed back into the two-dimensional feature space. This process can be defined as:

$$X^{m}_{s(i)} = Reshape\left(X'_{s(i)}\right)\mathbb{R}^{C\times H\times W} \tag{12}$$

In the feature fusion layer, M2AM stacks the current feature $X_{s(i)}$, cross-scale attention feature $X^{h}_{s(i)}$, $X^{l}_{s(i)}$, and global dependency modeling result $X^{m}_{s(i)}$. These are then fused and channel-adjusted through a lightweight residual bottleneck structure (Res-Bottleneck), outputting more discriminative multi-scale feature $X^{M}_{s(i)}$. This process can be defined as:

$$X^{M}_{s(i)} = \mathcal{F}\left(X_{s(i)} + X^{h}_{s(i)} + X^{l}_{s(i)} + X^{m}_{s(i)}\right) \tag{13}$$

where $\mathcal{F}(\cdot)$ represents the residual bottleneck fusion module.

Additionally, considering the special structure of the first and fifth layers, as shown in Figure 4b,c, this paper further designs simplified versions of M2AM: M2AM-1 and M2AM-5. These are used for feature fusion scenarios that only include inputs from the previous or subsequent stage, ensuring model performance while balancing structural flexibility and computational efficiency.

2.5. Loss Function

To mitigate the severe foreground-background imbalance caused by extremely sparse fire-point pixels, a multi-scale weighted hybrid loss is adopted as the network optimization objective. This loss function is composed of Binary Cross-Entropy (BCE) loss and IoU loss, which respectively constrain the network training at the pixel-level discrimination and region-level consistency. The BCE loss provides pixel-level supervision to distinguish fire pixels from the background. However, due to the dominance of background pixels in satellite images, BCE alone tends to bias the model toward background regions, reducing its sensitivity to tiny fire points. To address this limitation, the IoU loss is introduced to

optimize the overlap between predictions and ground truth, which emphasizes region-level structure and is more suitable for extremely small and sparse fire point targets. Therefore, the single-scale supervision loss is defined as:

$$L_i = \mathcal{L}_{BCE}(S_i, G) + \mathcal{L}_{IoU}(\sigma(S_i), G) \tag{14}$$

where $\mathcal{L}_{BCE}$ represents the BCE loss, $\mathcal{L}_{IoU}$ represents the IoU loss, S_i denotes the network output at the $i-th$ layer, G represents the corresponding ground truth fire point mask, and $\sigma(\cdot)$ is the Sigmoid activation function.

Furthermore, to improve the localization of tiny fire points, a deep supervision strategy is employed to jointly optimize multi-scale outputs with level-dependent weights. Higher weights are assigned to shallow outputs with finer spatial resolution to enhance precise localization, while deeper outputs receive gradually reduced weights to ensure training stability. Accordingly, the overall loss function $\mathcal{L}_{total}$ is defined as:

$$\mathcal{L}_{total} = L_1 + \frac{1}{2}L_2 + \frac{1}{4}L_3 + \frac{1}{8}L_4 + \frac{1}{16}L_5 + \frac{1}{32}L_6 \tag{15}$$

3. Results

3.1. Experimental Environment

3.1.1. Dataset

To evaluate the effectiveness of the proposed method, experiments are conducted on two subsets of the Active Fire dataset [33] with pronounced pixel-level sparsity characteristics, namely Oceania and Asia4. The dataset is constructed based on officially released Landsat-8 OLI scenes, and the images consist of 10-band multispectral patches cropped from preprocessed Landsat-8 data. The original Landsat-8 images were acquired with standard processing, including radiometric calibration and geometric correction prior to dataset construction. Cloud and cloud-shadow handling were addressed in the dataset preparation, and invalid/no-data pixels were masked accordingly. The Oceania and Asia4 subsets contain 2200 and 4900 images, respectively. To better illustrate the distribution of fire points, fire pixels are grouped into five intervals (0–5, 6–10, 11–20, 21–50, >50), as shown in Figure 5.

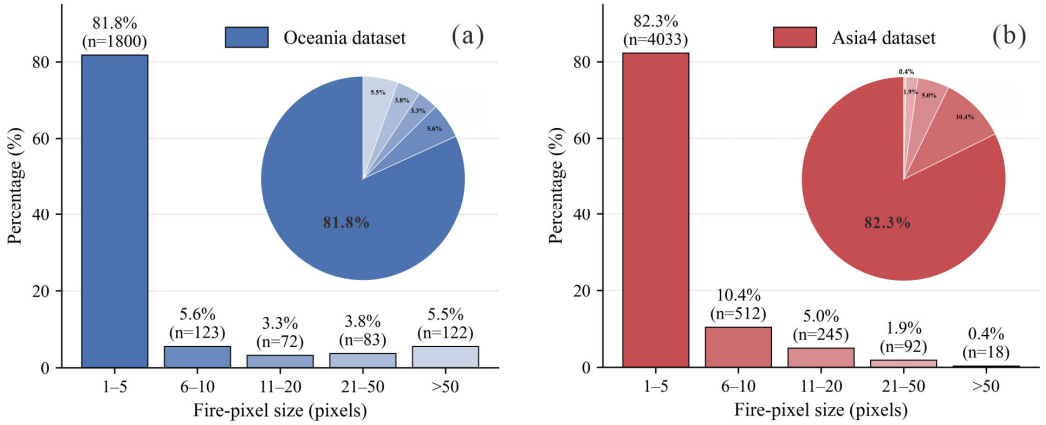

Figure 5. Fire pixel distribution statistics of the Oceania (**a**) and Asia4 (**b**) datasets.

As shown in Figure 5a, in the Oceania dataset, 1800 images (81.8%) contain fewer than five fire pixels, while only 205 images (approximately 9.3%) contain more than 20 fire pixels, indicating a highly imbalanced fire pixel distribution. Similarly, as shown in Figure 5b, in the Asia4 dataset, about 82.3% of the images (4033 images) contain fewer than five fire pixels, whereas only 18 images (0.4%) contain more than 50 fire pixels, exhibit-

ing an even more severe imbalance. Although both datasets are dominated by small fire targets, differences in fire pixel distributions still exist. The Asia4 dataset contains more medium-sized fire targets (6–20 pixels), while the Oceania dataset includes relatively more large fire targets (>20 pixels). Such cross-region differences further increase the difficulty of model generalization. Overall, the fire distributions in both subsets exhibit the typical characteristics of small-scale and sparsely distributed fire targets in satellite imagery, which can effectively evaluate the performance of the proposed model in tiny fire point segmentation tasks.

3.1.2. Experimental Details

Experiments were conducted on the Oceania and Asia4 datasets. Each dataset was divided into training, validation, and test sets based on the fire pixel values in each image, following a 7:1:2 ratio. Specifically, the Oceania dataset contains 1540, 220, and 440 images in the training, validation, and test sets, respectively, while the Asia4 dataset contains 3430, 490, and 980 images in the corresponding subsets.

All network models were implemented using the PyTorch 2.1.0+cu121 framework on a single NVIDIA GeForce RTX 4070 Super GPU (NVIDIA Corporation, Santa Clara, CA, USA). To ensure fair comparison, all baseline models and the proposed FireMambaNet were trained using the exact same multispectral input bands (Landsat-8 B7, B6, and B5), which were determined to be the optimal configuration in Section 3.3.2. No external pre-trained weights were utilized for any model. Regarding training strategies, all models employed identical settings: Adam optimizer, initial learning rate of 0.005, weight decay of 1×10^{-4}, batch size of 4, and a polynomial learning rate decay scheme over 100 epochs. No data augmentation techniques were applied during training in order to maintain a consistent experimental setup across all models. For the proposed method, no additional training epochs or special hyperparameter tuning effort was applied. For the baseline models, hyperparameters were initialized according to their original publications and were only minimally adjusted within reasonable ranges to ensure stable convergence.

3.1.3. Evaluation Metrics

In our work, IoU and the F1 score are adopted to evaluate model performance. IoU is used to measure the spatial consistency between the predicted fire regions and the ground-truth masks, while the F1 score comprehensively reflects the accuracy and detection capability of the model in fire point segmentation. The corresponding formulations are defined as follows:

$$\mathrm{IoU} = \frac{TP}{TP + FP + FN} \tag{16}$$

$$\mathrm{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

where TP denotes the number of true positive pixels correctly predicted as fire pixels, FP denotes the number of false positive pixels incorrectly predicted as fire pixels, and FN denotes the number of fire pixels that are incorrectly classified as background.

3.2. Comparison with State-of-the-Art Segmentation Methods

To further validate the effectiveness of the proposed method, we compare it with a range of state-of-the-art segmentation approaches, including: (1) CNN-based methods, such as UNet [34], U^{2}Netp [37], PSPNet [45], CorrNet [46], SeaNet [47], and FPSU^{2}Net [36]; (2) Transformer-based methods, including SegFormer [38] and PGNet [48]; and (3) Mamba-based methods, such as AfaMamba [49] and P-Mamba [50].

3.2.1. Experimental Results on the Oceania Dataset

As shown in Table 1, the quantitative comparison results of the proposed method and other networks on the Oceania dataset are presented. Compared to the best-performing CNN-based segmentation network (FPSU^{2}Net), the proposed method improves IoU by 1.81% and F1 score by 0.80%, demonstrating that the constructed network has stronger advantages in both the overall consistency of fire point regions and pixel-level classification accuracy. Compared to Transformer-based segmentation networks (such as SegFormer and PGNet), both the IoU and F1 scores are significantly lower than those of the proposed method. The Transformer structure focuses on global modeling capabilities, which have advantages in large-scale target and contextual relationship modeling. However, when facing the numerous tiny fire point targets with extremely small scales and low pixel proportions in the Oceania dataset, global features tend to overwhelm local salient information, leading to limited recognition ability for tiny fire points. Furthermore, compared to Mamba-based segmentation networks (such as AfaMamba and P-Mamba), while Mamba performs better than Transformer in long sequence modeling and feature propagation efficiency, its overall performance still lags behind the proposed method but is noticeably superior to pure Transformer architectures. This indicates that the Mamba structure alleviates, to some extent, the issue of global modeling being unfriendly to small targets. However, it still struggles with capturing fire point boundaries and fine-grained structures in complex backgrounds.

Table 1. Quantitative comparison results for the Oceania dataset.

Method	Type	Backbone	Oceania	
			IoU	F1
UNet	C	-	81.67	89.98
U^{2}Netp	C	RSU	83.91	91.27
PSPNet	C	Resnet50	36.88	53.92
CorrNet	C	VGG16	82.34	90.38
SeaNet	C	MobileNet	60.88	75.67
FPSU^{2}Net	C	RSU	86.91	93.01
SegFormer	C-T	MiT-b3	59.91	74.91
PGNet	C-T	Res18+Swin-B	52.74	68.95
AfaMamba	C-M	Res18	74.68	85.5
P-Mamba	C-T-M	Res18+Swin-B	58.42	73.75
Ours	C-M	CG-RSU	88.51	93.76

Notes: C, C-T, C-M, and C-T-M denote convolution-based, convolution–Transformer hybrid, convolution–Mamba hybrid, and convolution–Transformer–Mamba hybrid models, respectively; RSU denotes Residual U-block; MiT-b3 denotes Mix Transformer (b3); Swin-B denotes Swin Transformer Base; Res18 denotes ResNet-18; and VGG16 denotes the 16-layer VGG network.

From the perspective of the network backbone structure, the CG-RSU backbone used in this paper achieves more thorough multi-scale feature fusion while maintaining fewer network layers and a smaller parameter scale. This effectively enhances the network's ability to perceive local fire point shapes and weak response regions. As a result, the network not only ensures accuracy but also demonstrates better stability and generalization performance.

To better compare the network's segmentation performance on fire points of different scales, as shown in Figure 6, we selected scenes from the Oceania dataset featuring fire points of varying scales and shapes. The fire point pixel sizes range from pixels = 120 to pixels = 4. As shown in Figure 6a,b, for fire point samples with larger pixel sizes (pixels = 120 and 93), most CNN-based methods are able to detect the main body of the fire points. However, there are deficiencies in maintaining the continuity of elongated structures, with the predicted results commonly showing breaks or local omissions. In

contrast, the prediction results of Transformer- and Mamba-based segmentation networks (SegFormer, AfaMamba, and P-Mamba) contain more errors in prediction. In comparison, the proposed method performs better in maintaining the overall connectivity of the fire points. The predicted results are highly consistent with the ground truth in terms of shape and spatial distribution, and the number of false positive pixels is significantly reduced.

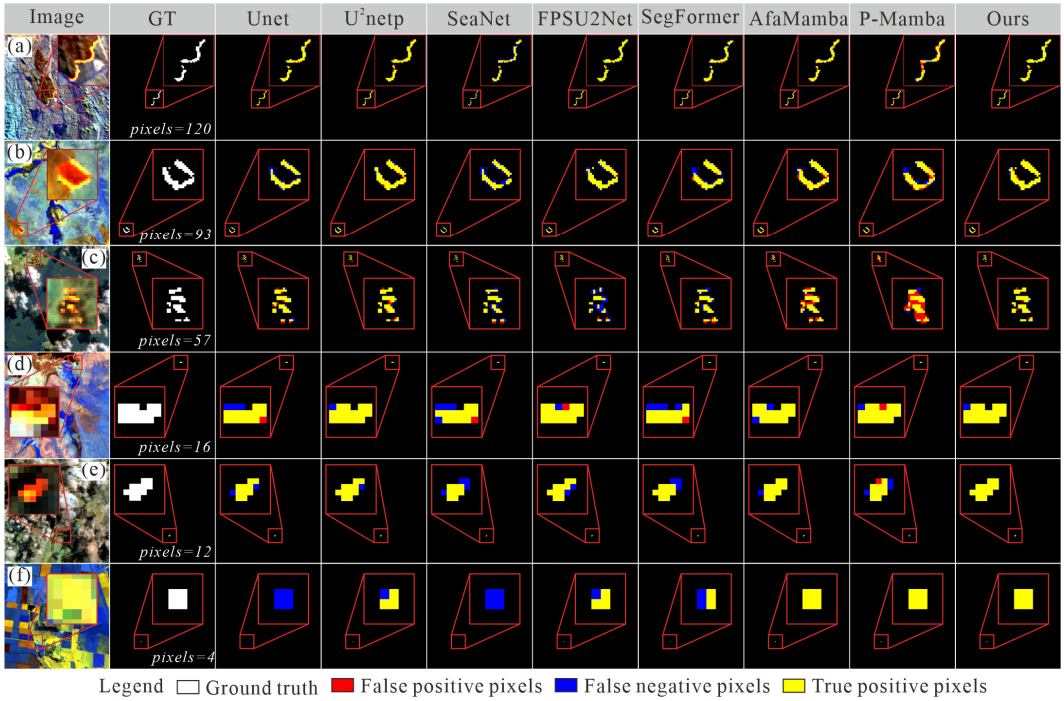

Figure 6. Visualization of segmentation results for fire points at different scales on the Oceania dataset. (**a**–**f**) represent six fire point samples with increasing scales.

As shown in Figure 6c,d, when the fire point scale is reduced to medium size (pixels = 57 and 16), the differences between methods become more apparent. Some methods (such as SeaNet and FPSU^{2}Net) generate more false negatives under complex background interference, leading to incomplete fire point structures. Mamba-based methods (AfaMamba and P-Mamba) exhibit significant false positives in local regions, severely affecting detection accuracy. In contrast, the proposed method is still able to accurately delineate fire point contours at this scale, maintaining a good balance between false positives and false negatives.

As shown in Figure 6e,f, in the most challenging tiny fire point scenarios (pixels = 12 and 4), most of the comparison methods suffer from varying degrees of false negatives, with some methods (such as UNet and SeaNet) completely failing to detect the fire points. Mamba-based methods and the proposed method, however, are able to consistently identify the fire point locations, with the proposed method achieving nearly perfect predictions.

Overall, the visualization results on the Oceania dataset show that the proposed method has significant advantages in maintaining the continuity of elongated fire point structures, detecting small-scale targets, and suppressing false positives in complex backgrounds. This further validates its robustness and effectiveness under different fire point shapes and scale conditions.

In summary, the experimental results fully validate the effectiveness and superiority of the proposed method in fire point segmentation tasks at different scales, particularly in application scenarios where complex backgrounds and tiny fire points coexist, such as in the Oceania dataset.

3.2.2. Experimental Results on the Asia4 Dataset

To better assess the model's performance, as shown in Table 2, we also quantitatively compared the proposed method with several representative segmentation networks on the Asia4 dataset. As shown in Table 2, the proposed method achieved optimal performance in both IoU and F1 score metrics, reaching 85.65% and 92.26%, respectively. This further validates the effectiveness of the method under different regional and data distribution conditions. Compared to the best-performing CNN-based segmentation network, FPSU2Net, the proposed method improves IoU by 2.07% and F1 score by 1.21%. This performance improvement is even more significant compared to the Oceania dataset, indicating that the proposed method has stronger robustness in maintaining the integrity of fire point regions and suppressing false positives. Compared to Transformer-based segmentation networks (SegFormer, PGNet), their performance further declines on the Asia4 dataset, with both IoU and F1 scores significantly lower than those of the proposed method. This indicates that when the Asia4 dataset contains a large amount of high background interference, low contrast, and fire points with significant scale variations, Transformer architectures that rely on global modeling struggle to effectively highlight the local salient features of fire points, thereby limiting their segmentation performance. For Mamba-based segmentation networks, such as AfaMamba and P-Mamba, their overall performance lies between that of CNN and Transformer methods, showing better stability. However, they still fall significantly behind the proposed method in terms of both IoU and F1 scores.

Table 2. Quantitative comparison results for the Asia4 dataset.

Method	Type	Backbone	Asia4	
			IoU	F1
UNet	C	-	69.53	82.03
U^2Netp	C	RSU	81.16	89.6
PSPNet	C	Resnet50	28.43	44.28
CorrNet	C	VGG16	72.06	83.76
SeaNet	C	MobileNet	64.39	78.33
FPSU^2Net	C	RSU	83.58	91.05
SegFormer	C-T	MiT-b3	50.84	67.41
PGNet	C-T	Res18+Swin-B	49.24	65.99
AfaMamba	C-M	Res18	71.95	83.69
P-Mamba	C-T-M	Res18+Swin-B	53.27	69.51
Ours	C-M	CG-RSU	85.65	92.26

To better assess the model's performance, we selected different scenes with multiple fire points at various scales for display, as shown in Figure 7. As shown in Figure 7a–c, in scenes with multiple fire points at different scales, CNN-based methods (such as UNet and SeaNet) exhibit a significant number of blue pixels. This indicates that these methods have a serious issue with false negatives, and as the number of fire pixels decreases, the false negative rate for these methods increases. The Transformer-based methods also exhibit a certain degree of false negatives, though their false negative rate is lower compared to CNN-based methods. Mamba-based methods, on the other hand, have fewer blue pixels but more red pixels. This indicates that although these methods have a lower false negative rate, they suffer from a higher false positive rate, making it difficult to identify tiny fire point targets. In contrast, our method predominantly shows yellow pixels, demonstrating stable performance across different scales and scenes with multiple fire points. For single fire points, as shown in Figure 7d–f, CNN-based methods and Transformer-based methods exhibit obvious false negatives, and they even fail to effectively distinguish fire points from complex backgrounds. Although U^2Net and FPSU^2Net have a lower false negative rate,

they have a higher false positive rate compared to our method. Mamba-based methods, in this scenario, suffer from both false positives and false negatives, making it difficult to consistently identify fire point targets. In contrast, our method demonstrates more stable performance. As shown in Figure 7d, the proposed method is able to fully segment the fire points.

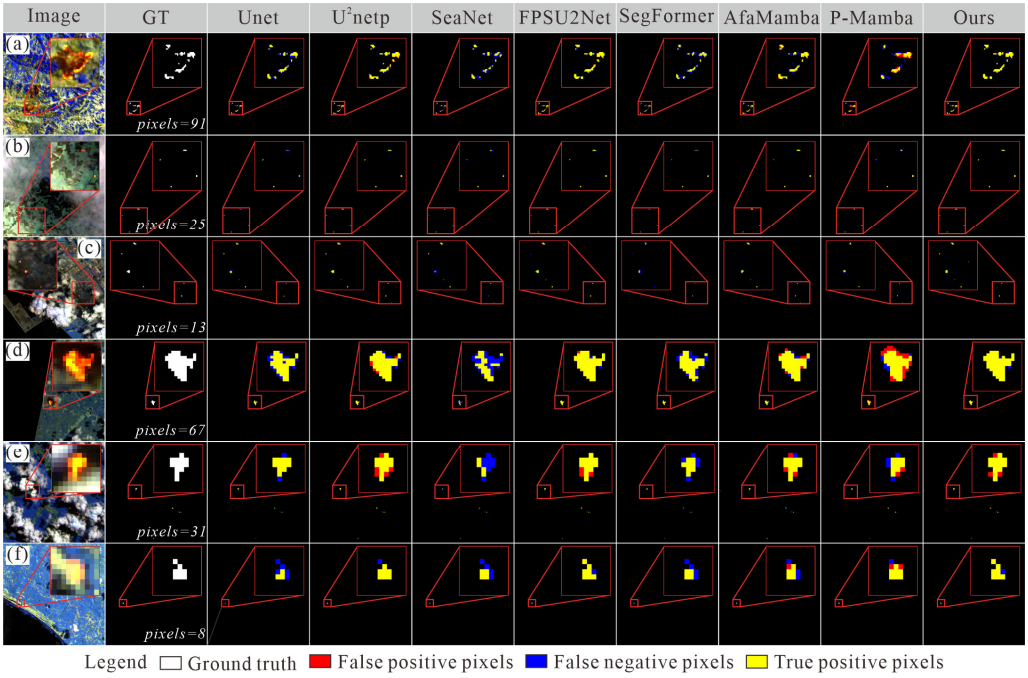

Figure 7. Visualization of segmentation results for fire points at different scales on the Asia4 dataset. (**a**–**f**) represent six fire point samples with increasing scales.

In summary, under complex backgrounds and in scenes with multiple fire points at different scales, the proposed method demonstrates stronger robustness and fine segmentation capabilities. Moreover, the proposed method not only excels in overall statistical metrics but also has a clear advantage in pixel-level details and the recognition of multiple tiny targets.

3.2.3. Experimental Results on Cross-Region Generalization

To further evaluate the cross-region generalization capability of the proposed method, two additional experiments were conducted. Specifically, the models were trained on the Asia4 dataset and tested on Oceania, and the reverse setting was also considered. The results are summarized in Table 3. As shown in Table 3, under cross-region evaluation, most methods experience a noticeable performance degradation due to differences in background characteristics and fire distribution patterns between the two datasets. Traditional convolutional neural network (CNN)-based models, such as UNet and PSPNet, exhibit relatively limited generalization capability, with their IoU and F1 scores decreasing by an average of 35.15% and 22.42%, respectively. Although several more advanced architectures, such as CorrNet and FPSU^{2}Net, achieve improved performance, their cross-region accuracy remains lower than that obtained in the within-dataset experiments. In contrast, the proposed method demonstrates the best overall cross-region performance under both evaluation settings. When trained on Asia4 and evaluated on Oceania, the proposed model achieves an IoU of 87.94%, showing only a slight decrease compared with the original experiment while still outperforming the competing methods. Similarly, when trained on Oceania and evaluated on Asia4, our model still maintains strong performance, achieving an IoU of 81.59% and an F1 score of 89.86%, corresponding to performance decreases of

4.74% and 2.60%, respectively, compared with the original results. Moreover, the proposed method still surpasses the second-best method (FPSU^{2}Net) by 3.34% IoU.

Table 3. Cross-Region Generalization Results Between Asia4 and Oceania.

Train		**Asia4**		**Oceania**	
Test		**Oceania**		**Asia4**	
Method	**Type**	**IOU**	**F1**	**IOU**	**F1**
UNet	C	56.01	71.79	56.88	72.51
U^{2}Netp	C	83.44	90.97	66.67	80.00
PSPNet	C	30.95	47.27	30.43	46.66
CorrNet	C	77.41	87.27	72.95	84.36
SeaNet	C	47.16	64.10	48.59	65.40
FPSU^{2}Net	C	86.26	92.63	78.25	87.80
SegFormer	C-T	45.56	62.60	44.73	61.81
PGNet	C-T	36.02	52.96	58.27	73.63
AfaMamba	C-M	59.89	74.68	70.25	82.53
P-Mamba	C-T-M	49.46	66.19	47.05	63.99
Ours	C-M	87.94	93.58	81.59	89.86

Overall, the results in Table 3 demonstrate that the proposed network achieves superior robustness and cross-region generalization ability compared with existing approaches. These results further indicate that the proposed modules exhibit effective synergy, enabling the network to better capture fire targets under complex background conditions.

3.3. Ablation Experiments

To validate the contribution of each key module to satellite fire point detection performance, this paper conducted ablation experiments on the CG-RSU, FDCM, and M2AM modules, as well as ablation experiments on different spectral bands of the Landsat 8 satellite. The details of these ablation experiments are described below.

3.3.1. Ablation Experiment of the Module

To verify the effectiveness of each proposed module, a total of eight module ablation configurations were evaluated on the Oceania and Asia4 datasets, with the quantitative results summarized in Table 4. In addition, to provide a more intuitive understanding of how different modules influence feature representation, we selected representative scenes containing a single fire point and multiple coexisting fire points from both datasets and visualized the feature maps at different network stages. The corresponding qualitative results are presented in Figure 8.

Table 4. Ablation Experiment Results for the Module.

No.	**Baseline**	**CG-RSU**	**M2AM**	**FDCM**	**Oceania**		**Asia4**	
					IoU	**F1**	**IoU**	**F1**
1	✔				83.91	91.27	81.16	89.60
2	✔	✔			85.52	92.19	83.34	90.91
3	✔		✔		84.84	91.80	83.24	90.85
4	✔			✔	85.83	92.37	83.63	91.09
5	✔	✔	✔		86.61	92.83	83.87	91.23
6	✔	✔		✔	87.32	93.23	83.48	91.00
7	✔		✔	✔	86.13	92.57	84.25	91.45
8	✔	✔	✔	✔	88.51	93.76	85.65	92.26

Notes: ✔ indicates that the corresponding module is included.

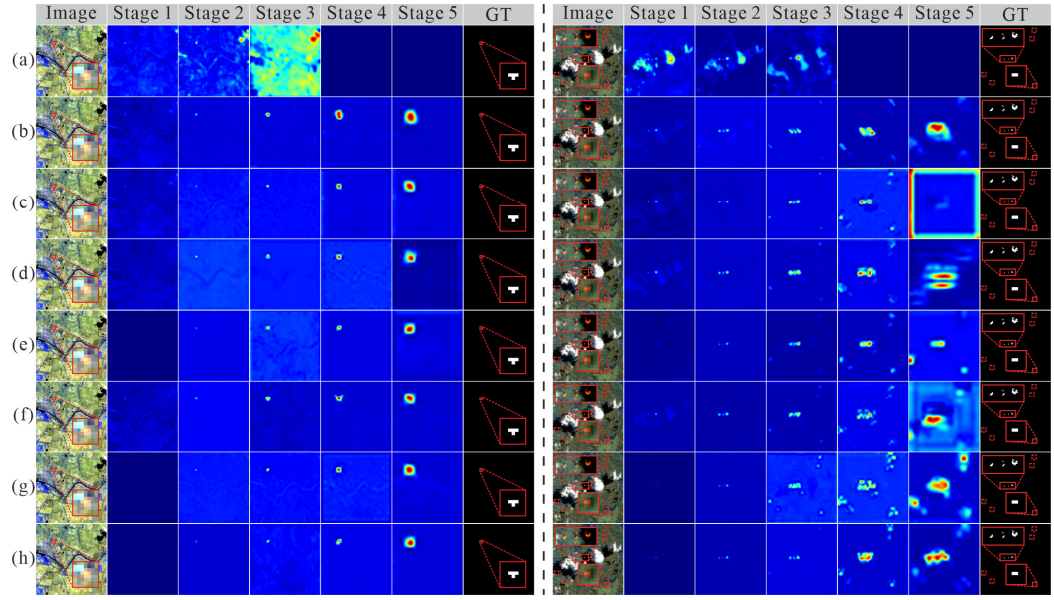

Figure 8. Visualization results of multi-stage feature representations under different module ablation configurations, where (**a**–**h**) correspond to the feature heatmaps of the model under different module combinations No.1–No.8 in Table 3. The color gradient from blue to red indicates increasing feature response intensity.

As shown in Table 4 (No.1), the baseline model achieves IoU scores of 83.91% and 81.16% on the Oceania and Asia4 datasets, respectively. Based on this baseline, the individual introduction of CG-RSU, M2AM, and FDCM modules (No.2–No.4) consistently improves detection performance on both datasets, with IoU gains ranging from 1.9% to 2.6%. As illustrated in Figure 8a–d, compared with the baseline model, the single-module configurations produce feature maps that can more clearly activate fire regions, demonstrating the effectiveness of each module in multi-scale stable modeling, long-range dependency perception, and direction-aware contextual compensation.

Furthermore, Table 4 (No.5–No.7) reports the results of different dual-module combinations. As observed in Figure 8e–g, in scenarios with multiple coexisting fire points, any combination of two modules outperforms the corresponding single-module configuration, enabling the detection of more fire points with improved completeness. This observation indicates strong complementarity among multi-scale stable modeling, long-range dependency perception, and structure-aware modulation. Specifically, the combination of CG-RSU and FDCM (No.6) improves the IoU by 3.94% on the Oceania dataset, while the combination of M2AM and FDCM (No.7) achieves superior performance on the Asia4 dataset, with an IoU gain of 3.70%. These results suggest that direction-aware contextual information exhibits stronger generalization capability in complex regions and heterogeneous background scenarios.

When all three modules are jointly integrated (No.8), the proposed method achieves the best overall performance, obtaining an IoU of 88.51% and an F1 score of 93.76% on the Oceania dataset, and 85.65% IoU and 92.26% F1 score on the Asia4 dataset. Compared with the baseline model, the proposed framework yields IoU improvements of 4.60% and 4.49% on the two datasets, respectively. Overall, these results demonstrate that the proposed modules exhibit significant synergistic gains for satellite fire detection, effectively enhancing the robustness and generalization capability of the model under complex backgrounds and cross-regional scenarios.

3.3.2. Effect Analysis of the FDCM Module

Since FDCM serves as a critical feature modulation module bridging different networSince the proposed FDCM serves as a critical feature modulation component that bridges different network stages, its placement within the network may significantly influence the overall detection performance. To investigate this effect, ablation experiments were conducted by inserting FDCM at different stages of the network (Stage 1–Stage 5), as summarized in Table 5. As shown in Table 5, placing FDCM in the shallow layers (Stage 1) leads to a significant performance degradation, with the IoU on the Oceania dataset dropping to only 42.95%. This indicates that directional context modeling fails to effectively capture high-level semantic information when applied to low-level feature representations. As FDCM is progressively moved toward the intermediate stages, the model performance improves substantially. Among all configurations, the Stage 4 placement achieves the best results on both datasets, yielding IoU scores of 88.51% and 85.65% on the Oceania and Asia4 datasets, respectively. In contrast, when FDCM is further shifted to the deepest layer (Stage 5), a noticeable performance decline is observed. This suggests that features at very deep layers are already highly semantically aggregated, thereby limiting the benefits of directional context modulation. Consequently, FDCM is deployed at Stage 4 in the proposed framework to maximally exploit its role in feature modulation between low-level and high-level representations.

Table 5. Ablation Experiment Results for FDCM Location.

No.	Location	Oceania		Asia4	
		IoU	F1	IoU	F1
1	Stage 1	42.95	59.96	54.54	70.59
2	Stage 2	87.82	93.36	83.84	91.21
3	Stage 3	88.05	93.52	81.67	89.91
4	Stage 4	88.51	93.76	85.65	92.26
5	Stage 5	87.86	93.42	83.51	91.02

In addition, to further evaluate the effectiveness of FDCM for detecting tiny fire targets, we conducted comparative experiments across five fire-size groups based on the number of fire pixels. The experimental results are summarized in Table 6. As shown in Table 6, on the Oceania subset, for the 0–5 pixel group, the IoU increases from 69.60% to 73.12%, corresponding to a relative improvement of 5.06%, while the F1 score improves by 2.92%. Similarly, for the 11–20 pixel group, the IoU increases by 6.07% and the F1 score improves by 3.33%. These results indicate that directional feature modeling helps capture the structural characteristics of small fire targets. For medium and large targets (21–50 pixels and >50 pixels), the performance changes are relatively small, suggesting that FDCM mainly benefits scenarios involving small fire targets. In addition, as shown in Table 6, FDCM also demonstrates consistent improvements on the Asia4 subset. For the 0–5 pixel group, the IoU improves by 2.61%, and the F1 score increases by 1.44%. For the 6–10 pixel group, the improvements reach 1.79% and 0.95%, respectively. These results suggest that the directional modeling capability introduced by FDCM enhances the network's ability to capture anisotropic fire patterns and improves the segmentation performance for sparse tiny fire pixels. For larger fire regions (>50 pixels), the effect of FDCM becomes limited, as large fire areas are already sufficiently prominent and easier for the network to distinguish.

Table 6. Performance comparison across fire-size groups (w/ and w/o FDCM).

Dataset	Metric	0–5 Pixels		6–10 Pixels		11–20 Pixels		21–50 Pixels		>50 Pixels	
		IOU	F1	IOU	F1	IOU	F1	IOU	F1	IOU	F1
Oceania	w/o FDCM	69.60	82.07	86.79	92.93	77.88	87.56	88.68	94.00	91.77	95.71
	w/ FDCM	73.12	84.47	87.20	93.16	82.61	90.48	90.19	94.84	91.76	95.70
	Δ (%)	5.06	2.92	0.47	0.25	6.07	3.33	1.70	0.89	−0.01	−0.01
Asia4	w/o FDCM	80.02	88.90	85.96	92.45	87.90	93.56	85.75	92.33	87.95	93.59
	w/ FDCM	82.11	90.18	87.50	93.33	87.46	93.31	86.58	92.80	89.87	94.67
	Δ (%)	2.61	1.44	1.79	0.95	−0.50	−0.27	0.97	0.51	2.18	1.15

3.3.3. Ablation Study on Different Spectral Band Combinations

To analyze the impact of different spectral bands on fire point detection performance, ablation experiments with multiple spectral band combinations were conducted on the Oceania and Asia4 datasets. The experimental results are presented in Table 7. It can be observed that different spectral bands contribute unequally to fire point detection performance. Specifically, when only visible bands (432) are used, the model exhibits extremely poor performance, achieving IoU values of only 15.79% and 11.25% on the Oceania and Asia4 datasets, respectively. This result indicates that relying solely on visible-spectrum information makes it difficult to effectively distinguish fire pixels from complex backgrounds. In contrast, as shown in Table 7 (No. 3–7), the introduction of near-infrared (B5) and shortwave infrared (B6 and B7) bands leads to a substantial improvement in detection performance. Among these configurations, the combination of bands 7, 6, and 5 consistently demonstrates stable and superior detection capability across both datasets. In particular, the 765 bands combination achieves the best overall performance, with IoU and F1 scores of 88.51% and 93.76% on the Oceania dataset, and 85.65% and 92.26% on the Asia4 dataset, respectively. Furthermore, when all available bands (ALL) are jointly modeled, the performance does not improve significantly. This phenomenon suggests that the inclusion of redundant or weakly discriminative spectral information may limit the effectiveness of the network. Therefore, the 765 bands combination is selected as the optimal spectral configuration for fire point detection in this study.

Table 7. Ablation Experiment Results of Different Band Combinations.

No.	Band	Oceania		Asia4	
		IoU	F1	IoU	F1
1	432	15.79	30.15	11.25	20.28
2	654	66.83	80.12	55.48	71.37
3	762	87.92	93.48	79.13	88.35
4	763	87.75	93.33	79.95	88.86
5	764	86.58	92.72	78.64	88.05
6	765	88.51	93.76	85.65	92.26
7	766	84.52	91.60	54.52	70.57
8	ALL	85.26	92.02	84.66	91.69

3.3.4. Ablation Study on Different Random Seeds

To further evaluate the training stability of the proposed method, additional experiments were conducted using five different random seeds. The results are reported as mean ± standard deviation, as summarized in Table 8. As shown in Table 8, on the Oceania dataset, FireMambaNet achieves an Intersection over Union (IoU) of 88.82 ± 0.59 and an F1 score of 94.04 ± 0.35. On the Asia4 dataset, the model obtains an IoU of 85.48 ± 0.90 and an F1 score of 92.17 ± 0.52. The relatively small standard deviations indicate that the proposed

method maintains stable performance under different random seeds, demonstrating strong training robustness and reproducibility. In addition, 95% confidence intervals (CI) were computed to further quantify the statistical reliability of the results. As shown in Table 8, the relatively narrow confidence intervals suggest that the proposed model exhibits stable performance across different random initializations.

Table 8. Ablation Results under Different Random Seeds.

Metric	Oceania		Asia4	
	IoU	**F1**	**IoU**	**F1**
Mean ± Std	88.82 ± 0.59	94.04 ± 0.35	85.48 ± 0.90	92.17 ± 0.52
95% CI	[88.09, 89.55]	[93.61, 94.48]	[84.36, 86.59]	[91.52, 92.81]

3.3.5. Ablation Study on Different Dataset Splitting Strategies

To address the potential risk of spatial or temporal data leakage caused by random patch-level splitting, we further conducted additional experiments using a leakage-safe data partition protocol. Specifically, the dataset was split at the scene level, ensuring that all patches originating from the same scene appear exclusively in either the training, validation, or testing set. The datasets were still divided according to the 7:1:2 ratio. After the scene-level split, the Oceania dataset contains 1549 training images, 331 validation samples, and 320 testing images, while the Asia4 dataset contains 3214, 567, and 1119 images, respectively.

We then re-conducted the experiments on both datasets under the same experimental settings and environment. The results are summarized in Table 9. As shown in the table, on the Oceania dataset, the proposed network shows only slight performance decreases under the scene-level split compared with the random patch split, with the IoU and F1 scores decreasing by 0.69% and 0.30%, respectively. On the Asia4 dataset, the IoU decreases by 0.14%, while the F1 score decreases by 0.09%. These results indicate that the performance of the proposed model is not significantly affected by potential image leakage caused by random splitting, and the model still maintains stable performance under the more rigorous scene-level evaluation protocol. This further demonstrates the robustness and reliability of the proposed method.

Table 9. Performance Comparison Under Different Data Splitting Strategies.

Split Strategy	Oceania		Asia4	
	IOU	**F1**	**IOU**	**F1**
Patch-random	88.51	93.76	85.65	92.26
Scene-level	87.82	93.46	85.51	92.17
Δ (%)	−0.78	−0.32	−0.16	−0.10

3.4. Computational Complexity and Inference Efficiency

To further evaluate the proposed model in terms of computational complexity and inference efficiency, comparison experiments were conducted on a range of representative semantic segmentation models under a unified experimental environment. All models were tested on NVIDIA GeForce RTX 4070 Super GPUs with input image dimensions of 256 × 256 and a batch size of 4. The input data consisted of preprocessed TIFF images, and the reported FPS was calculated based solely on the forward inference time, excluding data loading, preprocessing, and postprocessing operations. Additionally, all models were evaluated in inference mode with gradient computation disabled. The results are summarized in Table 10.

Table 10. Comparison of Model Complexity and Size.

Methods	Type	Backbone	GFlOPs	Param	FPS
UNet	C	-	81.70	39.42	108.07
U^2Netp	C	RSU	37.65	1.13	70.80
PSPNet	C	Resnet50	44.44	52.49	128.41
CorrNet	C	VGG16	21.31	4.08	66.65
SeaNet	C	MobileNet	1.43	2.76	113.99
FPSU^2Net	C	RSU	16.09	1.33	61.49
SegFormer	C-T	MiT-b3	1.69	3.71	136.94
PGNet	C-T	Res18+Swin-B	28.91	72.71	76.82
AfaMamba	C-M	Res18	7.01	13.48	14.92
P-Mamba	C-T-M	Res18+Swin-B	4.54	28.76	23.08
Ours	C-M	CG-RSU	16.10	1.56	42.80

As shown in Table 10, traditional CNN methods such as UNet and PSPNet have relatively large parameter counts, reaching 39.42 M and 52.49 M, respectively. Although these models can achieve high inference speeds on the GPU, they suffer from considerable model redundancy. By comparison, lightweight networks such as SeaNet, U^2Netp, and FPSU^2Net substantially reduce parameter size, with U^2Netp and FPSU^2Net containing only approximately 1.5 M parameters. However, these lightweight models are still limited in their capacity to model complex features effectively. Transformer-based or hybrid architectures (e.g., SegFormer, PGNet, and P-Mamba) demonstrate advantages in capturing global context, but their parameter sizes and inference speeds vary widely. For instance, PGNet has a parameter count of 72.71 M, while P-Mamba, despite a lower computational load, still suffers from reduced inference speed. In contrast, the proposed FireMambaNet achieves a more balanced performance across model size, computational complexity, and inference efficiency. Built upon the CG-RSU architecture, it contains only 1.56 M parameters with a computational cost of 16.10 GFLOPs, and achieves an inference speed of 42.80 FPS on the NVIDIA GeForce RTX 4070 Super. While maintaining low model complexity, it exhibits stable inference efficiency, demonstrating that the proposed approach strikes a favorable balance between lightweight design and feature representation capability, indicating high potential for practical deployment.

In practical applications, Landsat-8 satellite images are typically much larger than the input patches used for model evaluation. Therefore, large-area inference is commonly performed using a sliding-window strategy, in which the model processes image patches sequentially and then stitches the prediction results to obtain the final segmentation map. To estimate the processing time of the proposed model, a standard Landsat scene with an image size of 10,000 $\times$ 10,000 pixels is considered as an example. Based on the measured inference speed of 42.80 FPS, the estimated processing time for a full Landsat scene is approximately 40–45 s under the same experimental settings. This corresponds to roughly 80–85 scenes per hour, indicating that the proposed FireMambaNet maintains practical efficiency for large-scale wildfire monitoring tasks. It should be noted that the actual runtime may vary depending on the sliding-window stride, image input/output operations, and post-processing procedures.

4. Discussion

As described in Section 3.1.1 for the Oceania and Asia4 datasets, satellite fire point segmentation tasks generally face challenges such as extremely small target scales, imbalanced pixel distribution, and significant regional differences in practical applications. Under these challenging conditions, as shown in Tables 1 and 2, the proposed method achieves stable and leading performance on both datasets, indicating that the proposed network

demonstrates strong robustness and generalization ability in complex backgrounds and cross-regional scenarios. Compared to traditional CNN methods, the CG-RSU backbone introduced in this paper effectively alleviates the issue of detail information loss caused by multiple downsampling steps through multi-scale stable feature fusion, allowing tiny fire points and elongated structures to be more fully expressed in high-level semantic features. In contrast, Transformer-based methods, due to their reliance on global modeling mechanisms, tend to be dominated by large-scale background areas in pixel-level highly imbalanced scenes, limiting their ability to recognize local weak-response fire points. Mamba-based methods, on the other hand, demonstrate certain advantages in long-range dependency modeling but still face challenges in capturing fire point boundaries and fine-grained structures under complex background conditions. This paper introduces the M2AM and FDCM modules, achieving effective synergy between long-range dependency perception and directional context modeling, significantly improving the model's stability and fine-grained expression ability in fire point segmentation tasks at different scales. Additionally, the band ablation experiment results shown in Table 7 further validate the crucial role of near-infrared and shortwave infrared bands in enhancing the spectral separability between fire points and the background. A reasonable band combination (e.g 765 band) is more beneficial for improving the overall segmentation performance of the model compared to simply increasing the input dimensions.

Although the proposed method demonstrates good overall performance, it still has certain limitations. As shown in Table 11, this paper presents the main limitations of the proposed method and corresponding improvement strategies.

Table 11. Summary of limitations and planned mitigation strategies.

Limitation	Planned Mitigation
Single-temporal imagery only	Incorporate multi-temporal or spatiotemporal modeling
False positives under complex backgrounds	Integrate terrain, meteorological, or multi-source remote sensing data
Fixed input resolution (256 × 256)	Explore adaptive multi-scale inference strategies

5. Conclusions

This paper addresses key challenges, such as extremely small fire point target scales, sparse spatial distribution, and complex backgrounds in satellite remote sensing imagery, by proposing a multi-scale Mamba network, FireMambaNet, for tiny fire point segmentation. The method effectively enhances segmentation accuracy for tiny fire points in complex remote sensing scenarios through the collaborative design of multi-scale feature encoding, directional context modeling, and long-range dependency modeling.

In terms of network architecture design, this paper constructs a nested encoder-decoder backbone network composed of 6 Cross-layer Gated Residual U-blocks (CG-RSU). The encoder extracts multi-scale contextual features from local to global scales through progressive downsampling operations. Meanwhile, the cross-layer gating modulation mechanism designed in the CG-RSU adaptively suppresses redundant responses in complex backgrounds and significantly enhances the features of tiny and weakly responding fire points, providing a cleaner and more discriminative feature foundation for subsequent directional modeling and global dependency learning. Building on this, the paper designs the Fire-aware Directional Context Modulation (FDCM) module, which explicitly models the anisotropic spatial expansion features of fire points under wind direction and terrain constraints by aggregating structured features along the horizontal, vertical, and diagonal directions. This effectively enhances the network's ability to perceive the directional propa-

gation and continuity of fires. Furthermore, the proposed Multi-scale Mamba Attention Module (M2AM) leverages the advantages of state space models in long sequence modeling, achieving cross-scale long-range dependency modeling while maintaining computational efficiency. This significantly improves the global consistency representation ability in sparse fire point regions.

Extensive experimental results on the Oceania and Asia4 subsets of the Active Fire dataset show that the proposed FireMambaNet outperforms various mainstream CNN, Transformer, and Mamba-based methods in both IoU and F1 score evaluation metrics. The module ablation experiments further validate the effectiveness and complementarity of the CG-RSU, FDCM, and M2AM modules in enhancing tiny fire point features, directional context modeling, and global dependency learning. The position sensitivity analysis of FDCM shows that deploying it in the mid-to-high layers of the encoder (Stage 4) allows it to fully exert its role in modulating features between lower and higher layers, suppressing background interference while strengthening the response to directional expansion structures of fire points along wind direction, terrain orientation, and other factors.

Additionally, the spectral band ablation experiments further reveal the critical role of the near-infrared (NIR) and shortwave infrared (SWIR) bands in fire point discrimination. Specifically, the B7-B6-B5 (765) band combination achieved optimal and stable detection performance on both datasets. In terms of model complexity and inference efficiency, FireMambaNet, with only 1.56 M parameters and a computational complexity of 16.10 GFLOPs, still maintains stable inference speed, demonstrating a good balance between lightweight design and feature representation capability, with high potential for practical deployment.

Author Contributions: Conceptualization, B.S. and Z.C.; methodology, B.S. and B.L.; validation, B.S., H.H. and B.L.; formal analysis, B.S.; investigation, B.S. and T.Y.; resources, Z.Z. and Y.C.; data curation, B.S. and H.H.; writing—original draft preparation, B.S.; writing—review and editing, Z.C., B.L. and Z.Z.; funding acquisition, Z.C. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Science and Technology Project (grant No. GuikeAB25069093) and partly by the National Natural Science Foundation of China (General Program, Grant No. 21976043).

Data Availability Statement: The data used in this study are from the publicly available Active Fire dataset published by de Almeida Pereira G. H. et al. (2021) [33] and are available at: https://github.com/pereira-gha/activefire (accessed on 1 March 2026).

Acknowledgments: The experiments in this study are conducted using the Active Fire dataset released by de Almeida Pereira G. H. et al. (2021) [33], with the Oceania and Asia4 subsets selected for evaluation. The authors would like to thank the dataset creators for providing high-quality data for satellite-based fire detection research. This work also benefits from previous studies and open-source resources in the fields of remote sensing and deep learning.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhao, J.; Yue, C.; Wang, J.; Hantson, S.; Wang, X.; He, B.; Li, G.; Wang, L.; Zhao, H.; Luyssaert, S. Forest fire size amplifies postfire land surface warming. *Nature* **2024**, *633*, 828–834. [CrossRef]
2. Jiao, Q.; Fan, M.; Tao, J.; Wang, W.; Liu, D.; Wang, P. Forest fire patterns and lightning-caused forest fire detection in Heilongjiang Province of China using satellite data. *Fire* **2023**, *6*, 166. [CrossRef]
3. Ali, A.W.; Kurnaz, S. Optimizing Deep Learning Models for Fire Detection, Classification, and Segmentation Using Satellite Images. *Fire* **2025**, *8*, 36. [CrossRef]

4. Wang, H.; Zhang, G.; Yang, Z.; Liu, F.; Xie, S. Satellite Remote Sensing False Forest Fire Hotspot Excavating Based on Time-Series Features. *Remote Sens.* **2024**, *16*, 2488. [CrossRef]
5. Zhang, X.; Zhao, J.; Chen, G.; Wang, T.; Wang, Q.; Wang, K.; Miao, T. Spatio-temporal fusion of landsat and modis data for monitoring of high-intensity fire traces in karst landscapes: A case study in china. *Remote Sens.* **2025**, *17*, 1852. [CrossRef]
6. Yang, S.; Huang, Q.; Yu, M. Advancements in remote sensing for active fire detection: A review of datasets and methods. *Sci. Total Environ.* **2024**, *943*, 173273. [CrossRef]
7. Ramsey, S.; Jones, S.; Reinke, K. Review of approaches and challenges for the validation of satellite-based active fire products in savannah ecosystems. *Int. J. Wildland Fire* **2024**, *33*, WF23202. [CrossRef]
8. Chen, J.; Wu, Y.; Wu, S.; Xie, L.; Tang, J.; Xu, Z.; Han, X.; Ma, X.; Zheng, W.; Sun, T.; et al. Application of FY-4B geostationary meteorological satellite in grassland fire dynamic monitoring. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1000409. [CrossRef]
9. Yin, J.; He, R.; Zhao, F.; Ye, J. Research on forest fire monitoring based on multi-source satellite remote sensing images. *Spectrosc. Spectr. Anal.* **2023**, *43*, 917–926.
10. Giglio, L.; Justice, C.O. Early polar-orbiting satellite-based fire remote sensing during the 1960s. *Int. J. Remote Sens.* **2024**, *45*, 5605–5615. [CrossRef]
11. Kondratyev, K.Y.; Dyachenko, L.N.; Binenko, V.I.; Chernenko, A.P. Detection of small fires and mapping of large forest fires by infrared imagery. *Remote Sens. Environ.* **1972**, *VIII*, 1297.
12. Matson, M.; Holben, B. Satellite detection of tropical burning in Brazil. *Int. J. Remote Sens.* **1987**, *8*, 509–516. [CrossRef]
13. Morisette, J.T.; Giglio, L.; Csiszar, I.; Justice, C.O. Validation of the MODIS active fire product over Southern Africa with ASTER data. *Int. J. Remote Sens.* **2005**, *26*, 4239–4264. [CrossRef]
14. Maier, S.W.; Russell-Smith, J.; Edwards, A.C.; Yates, C. Sensitivity of the MODIS fire detection algorithm (MOD14) in the savanna region of the Northern Territory, Australia. *ISPRS J. Photogramm. Remote Sens.* **2013**, *76*, 11–16. [CrossRef]
15. Schroeder, W.; Oliva, P.; Giglio, L.; Csiszar, I.A. The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sens. Environ.* **2014**, *143*, 85–96. [CrossRef]
16. Giglio, L.; Schroeder, W.; Justice, C.O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **2016**, *178*, 31–41. [CrossRef]
17. Filizzola, C.; Corrado, R.; Marchese, F.; Mazzeo, G.; Paciello, R.; Pergola, N.; Tramutoli, V. RST-FIRES, an exportable algorithm for early-fire detection and monitoring: Description, implementation, and field validation in the case of the MSG-SEVIRI sensor. *Remote Sens. Environ.* **2016**, *186*, 196–216. [CrossRef]
18. Laneve, G.; Castronuovo, M.M.; Cadau, E.G. Continuous monitoring of forest fires in the Mediterranean area using MSG. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2761–2768. [CrossRef]
19. Xie, Z.; Song, W.; Ba, R.; Li, X.; Xia, L. A spatiotemporal contextual model for forest fire detection using Himawari-8 satellite data. *Remote Sens.* **2018**, *10*, 1992. [CrossRef]
20. Hally, B.; Wallace, L.; Reinke, K.; Jones, S.; Skidmore, A. Advances in active fire detection using a multi-temporal method for next-generation geostationary satellite data. *Int. J. Digit. Earth* **2019**, *12*, 1030–1045. [CrossRef]
21. Mazzeo, G.; Falconieri, A.; Filizzola, C.; Genzano, N.; Pergola, N.; Marchese, F. Wildfire detection and mapping by satellite with an enhanced configuration of the Normalized Hotspot Indices: Results from Sentinel-2 and Landsat 8/9 data integration. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5606121. [CrossRef]
22. Giglio, L.; Descloitres, J.; Justice, C.O.; Kaufman, Y.J. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.* **2003**, *87*, 273–282. [CrossRef]
23. Liew, S.C. Detecting active fires with Himawari-8 geostationary satellite data. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019*; IEEE: New York, NY, USA, 2019; pp. 9984–9987.
24. Parto, F.; Saradjian, M.; Homayouni, S. MODIS brightness temperature change-based forest fire monitoring. *J. Indian Soc. Remote Sens.* **2020**, *48*, 163–169. [CrossRef]
25. Wooster, M.J.; Xu, W.; Nightingale, T. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. *Remote Sens. Environ.* **2012**, *120*, 236–254. [CrossRef]
26. Hally, B.; Wallace, L.; Reinke, K.; Jones, S.; Engel, C.; Skidmore, A. Estimating fire background temperature at a geostationary scale—An evaluation of contextual methods for AHI-8. *Remote Sens.* **2018**, *10*, 1368. [CrossRef]
27. Zhang, H.; Sun, L.; Zheng, C.; Ge, S.; Chen, J.; Li, J. A weighted contextual active fire detection algorithm based on Himawari-8 data. *Int. J. Remote Sens.* **2023**, *44*, 2400–2427. [CrossRef]
28. Wickramasinghe, C.H.; Jones, S.; Reinke, K.; Wallace, L. Development of a multi-spatial resolution approach to the surveillance of active fire lines using Himawari-8. *Remote Sens.* **2016**, *8*, 932. [CrossRef]
29. Yan, J.; Qu, J.; Ran, M.; Zhang, F. Himawari-8 AHI fire detection in clear sky based on time-phase change. *J. Remote Sens.* **2020**, *24*, 571–577.
30. Chen, J.; Zheng, W.; Wu, S.; Liu, C.; Yan, H. Fire monitoring algorithm and its application on the geo-kompsat-2A geostationary meteorological satellite. *Remote Sens.* **2022**, *14*, 2655. [CrossRef]

31. Kang, Y.; Jang, E.; Im, J.; Kwon, C. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *GISci. Remote Sens.* **2022**, *59*, 2019–2035. [CrossRef]
32. Muhammad, K.; Ahmad, J.; Lv, Z.; Bellavista, P.; Yang, P.; Baik, S.W. Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *49*, 1419–1434. [CrossRef]
33. de Almeida Pereira, G.H.; Fusioka, A.M.; Nassu, B.T.; Minetto, R. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 171–186. [CrossRef]
34. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
35. Seydi, S.T.; Saeidi, V.; Kalantar, B.; Ueda, N.; Halin, A.A. Fire-Net: A Deep Learning Framework for Active Forest Fire Detection. *J. Sens.* **2022**, *2022*, 8044390. [CrossRef]
36. Fang, W.; Fu, Y.; Sheng, V.S. FPS-U^{2}Net: Combining U^{2}Net and multi-level aggregation architecture for fire point segmentation in remote sensing images. *Comput. Geosci.* **2024**, *189*, 105628. [CrossRef]
37. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U^2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]
38. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
39. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, 11–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 7262–7272.
40. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
41. Zhang, L.; Zhang, Q.; Yang, Q.; Yue, L.; He, J.; Jin, X.; Yuan, Q. Near-real-time wildfire detection approach with Himawari-8/9 geostationary satellite data integrating multi-scale spatial-temporal feature. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *137*, 104416. [CrossRef]
42. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In *Proceedings of the First Conference on Language Modeling (COLM 2024), Philadelphia, PA, USA, June 2024*; OpenReview: Philadelphia, PA, USA, 2024.
43. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. Vmamba: Visual state space model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063.
44. Zhu, E.; Chen, Z.; Wang, D.; Shi, H.; Liu, X.; Wang, L. Unetmamba: An efficient unet-like mamba for semantic segmentation of high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 6001205. [CrossRef]
45. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017*; IEEE: New York, NY, USA, 2017; pp. 2881–2890.
46. GongyangLi, Z.; Bai, Z.; Lin, W.; Ling, H. Lightweight salient object detection in optical remote sensing images via feature correlation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5617712.
47. Li, G.; Liu, Z.; Zhang, X.; Lin, W. Lightweight salient object detection in optical remote-sensing images via semantic matching and edge alignment. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5601111. [CrossRef]
48. Xie, C.; Xia, C.; Ma, M.; Zhao, Z.; Chen, X.; Li, J. Pyramid grafting network for one-stage high resolution saliency detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022), New Orleans, LA, USA, 18–24 June 2022*; IEEE: New York, NY, USA, 2022; pp. 11717–11726.
49. Chen, H.; Luo, H.; Wang, C. AfaMamba: Adaptive Feature Aggregation With Visual State Space Model for Remote Sensing Images Semantic Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 8965–8983. [CrossRef]
50. Wang, L.; Li, D.; Dong, S.; Meng, X.; Zhang, X.; Hong, D. PyramidMamba: Rethinking pyramid feature fusion with selective space state model for semantic segmentation of remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *144*, 104884. [CrossRef]