ITC 2/55
Information Technology and Control
Vol. 55 / No. 2/ 2026
pp. 605-633
DOI 10.5755/j01.itc.55.2.42555

Received 2025/08/14 | Accepted after revision 2026/01/31

HOW TO CITE: Feng, X., Yang, X. (2026) RI-Mamba: A Dual-branch Residual Interaction Network for Lightweight Insulator Segmentation. *Information Technology and Control*, 55(2), 605-633. https://doi.org/10.5755/j01.itc.55.2.42555

RI-Mamba: A Dual-branch Residual Interaction Network for Lightweight Insulator Segmentation

Xiaowei Feng
Gansu Shining Science & Technology Co., Ltd; Lanzhou, China; 730070; e-mail: fxw417221015@sina.com

Xijun Yang*
Gansu Keyuan Electric Power Group Co., Ltd; Lanzhou, China; 730070; phone: 13919194577;
e-mail: 3600697705@qq.com

Corresponding author: 3600697705@qq.com

In aerial images, slender and multi-scale insulator targets are often set against complex backgrounds, which imposes higher demands on the model's ability to retain details and capture long-range dependencies. To address this challenge, we propose a lightweight Dual-branch Residual Interaction Network (RI-Mamba) for insulator segmentation. Unlike typical single-branch architectures, RI-Mamba adopts a parallel dual-branch to prioritize the extraction of detail-semantic features. Specifically, the detail branch enhances the representation of local fine structures through feature expansion, while the semantic branch strengthens the correlation between positional and contextual semantics from a coordinate perspective. During the feature fusion stage, the proposed Residual Mamba Block (R-Mamba) discards traditional convolutional methods for processing local features and instead directly reuses detail-semantic information from the original features via a residual structure, achieving efficient feature reuse while avoiding parameter growth. Furthermore, a Gated Global-local Interaction Module (G2LIM) is introduced, which uses a learnable gated self-attention mechanism to dynamically adjust the fusion weights of global context and local details, thereby reducing redundant computation and enhancing feature adaptability. Experimental results on two public datasets demonstrate that RI-Mamba, with only 4.15M Params and 5.31G FLops, achieves mIoU of 86.28% and 91.59%, respectively, outperforming 17 lightweight insulator segmentation models. These results highlight that RI-Mamba achieves a favorable balance between computational efficiency and segmentation accuracy, making it well-suited for deployment in real-world power inspection scenarios. The code is available at: https://github.com/yaoshuang-yaobo/RI-Mamba.

KEYWORDS: Insulator Segmentation, Parallel Dual-branch, Detail-semantic Features; R-Mamba, Gated Self-attention Mechanism

1. Introduction

Insulators are devices installed between different electrical conductors and grounded structural components, capable of withstanding voltage and mechanical stress. During operation, they often susceptible to defects such as zero-value, broken strings, and corrosion, all of which can seriously compromise the safety and reliability of overhead transmission lines [12, 18, 34]. With the rapid development of power inspection technology based on unmanned aerial vehicle (UAV), insulator detection based on aerial images has emerged as a crucial method for ensuring the secure operation of power grids [36, 42]. However, insulators often exhibit elongated and slender from aerial perspectives, making them vulnerable to interference from complex backgrounds, such as transmission wires and towers. This makes it challenging for traditional detection methods to simultaneously balance the consistency between detailed and semantic information. Deep learning technology, which can automatically identify insulator regions in images, offers advantages such as low cost, high efficiency, and high accuracy [32, 41]. Current insulator segmentation methods base on deep learning can be broadly categorized into three main types: CNNs-based methods, Transformer-based methods, and Mamba-based methods.

CNN-based methods utilize convolutional sliding windows to process the slender, grid-arranged insulators and employ end-to-end convolutional layers to automatically learn multi-level feature representations. Benefiting from this self-learning mechanism, such methods demonstrate strong segmentation for large-sized insulator [1, 41, 49]. However, the local receptive field inherent in standard convolutions makes it difficult for these models to fully capture the overall structural information of insulators, which restricts the modeling of global context and long-range spatial dependencies, thereby affecting the accurate understanding of the insulator's complete morphology. Transformer-based models leverage self-attention mechanisms to model global dependencies between pixels, effectively capturing the long-range contextual information of insulators across different spatial scales. Consequently, numerous studies have introduced Transformer and its variant architectures into insulator segmentation tasks [3, 27, 40]. Nevertheless, the computational complexity of the self-attention mechanism grows quadratically with the input image size, leading to substantial computational and memory resource consumption in insulator segmentation tasks, making them difficult to deploy in resource-constrained environments such as UAV-based power inspection tasks. In recent years, the Mamba based on State Space Models (SSM), has shown significant potential in long-sequence modeling tasks. Compared to Transformer, Mamba achieves linear computational complexity through a selective state space mechanism while maintaining excellent global modeling capabilities [7, 30]. Leveraging this characteristic, Mamba and its variants are gradually being applied to visual tasks and have demonstrated promising performance in the field of insulator segmentation [20, 57].

Although the three aforementioned methods have achieved notable performance in insulator segmentation tasks to varying degrees, the accurate segmentation of insulators in aerial images still faces a series of significant technical challenges. These challenges limit both the performance ceiling and the deployment feasibility of existing models in real-world power inspection scenarios. These challenges primarily stem from the coupling between target objects and complex environments:

1 **Morphological Specificity:** Insulators typically exhibit extreme slenderness ratios and variable directional angles, which demands the segmentation models possess strong detail preservation capabilities and perceptual performance across

all directions. Traditional convolutions, due to its inherent locality and isotropy, struggle to capture the global structure of such slender and mutil-directional targets effectively, often leading to issues such as breakage or blurred edges in the segmentation results.

2 **Complex Background Interference:** Aerial images are characterized by complex backgrounds containing a wide variety of object types, often including distractors similar in color and texture to insulators, such as transmission lines, pylons, and buildings. This requires the model to possess not only sharp local discriminative ability but also an accurate understanding of global semantics.

3 **Balance Between Performance and Efficiency:** Power inspection imposes stringent requirements on real-time performance, necessitating the processing of aerial images in real-time on embedded or mobile devices. However, models capable of effectively modeling long-range dependencies to address the above issues often have high computational complexity, resulting in prohibitively high computational costs.

In this challenging task, previous studies have applied Mamba-based variant models in this domain, their approaches mainly modified scanning strategies to capture semantic features more comprehensively, without effectively balancing local details and long-range dependencies [5, 13, 15, 24]. Moreover, existing global-local fusion mechanisms struggle to distinguish shallow appearance information from deep semantic features, because of the slender morphology of insulators and their high similarity to other interfering objects [22, 25, 26]. To address these issues, this paper proposes a lightweight Dual-branch Residual Interaction Network (RI-Mamba) for insulator segmentation. Unlike commonly adopted single-branch architectures in image segmentation, the proposed model employs a parallel dual-branch to separately extract detail-semantic features. Specifically, the detail branch uses a Detail Collection Expansion Unit (DCEU) to enhance edge representation and fine-grained local structures, while the semantic branch employs a Coordinate Focused Module (CFM) to explicitly encode spatial coordinates, thereby improving the localization and discriminative capacity of semantic features. During the feature fusion stage, the Residual-Mamba Block (R-Mamba) integrates residual structures with a 4D selective scanning strategy (SS4D), enabling bidirectional feature interaction and effectively reducing the reliance on convolutional operations. Furthermore, the Gated Global-local Interaction Module (G^2LIM) employs a gated self-attention mechanism to dynamically integrate multi-scale features, significantly enhancing the model's ability to adaptively select and fuse global contextual information with local details.

RI-Mamba is designed by integrating several key ideas from lightweight models to achieve high-performance segmentation with minimal computational and parameter overhead: (1) a parallel dual-branch lightweight architecture is adopted to avoid the computational redundancy commonly introduced by traditional serial or heavy multi-scale fusion designs; (2) the residual interaction mechanism in R-Mamba substantially reduces the parameter overhead introduced by additional convolutional layers; and (3) the gated self-attention mechanism enables dynamic feature selection, preventing fixed or redundant multi-scale fusion operations. Collectively, compared with the 17 existing insulator segmentation models, RI-Mamba achieves a favorable balance between efficiency and performance with only 4.15M Params and 5.13G FLops.

The main contributions of this work are summarized as follows:

1 We propose a lightweight insulator segmentation model, RI-Mamba, which achieves slightly higher mean performance with only 5.31GFLops and 4.15M Params.

2 To effectively capture the slender morphology and multi-scale characteristics of insulators, RI-Mamba adopts a parallel dual-branch architecture. Specifically, the detail branch integrates the DCEU unit to strengthen the representation of edges and local structures, while the semantic branch employs the CFM unit to explicitly encode spatial coordinate information, thereby improving the localization accuracy of semantic features.

3 To efficiently fuse detail-semantic features and capture long-range dependencies, the R-Mamba block combines residual connections with the SS4D. This design generates interactive weights between the two feature streams, improving gradient flow and feature reuse while maintaining linear computational complexity.

4 To further improve feature discrimination in complex backgrounds, the G^2LIM module introduces a learnable gated self-attention mechanism that dynamically adjusts fusion weights between global context and local details during upsampling. This adaptive feature enhancement highlights the slender morphology of insulators more effectively in the decoding stage.

The remainder of this paper is organized as follows. Section 2 reviews related work on insulator segmentation, the Mamba architecture, and global–local feature interaction. he overall architecture and working principles of RI-Mamba are detailed in Section 3. Experimental analyses and visualizations are presented in Section 4. The limitations of this work and potential directions for future research are discussed in Section 5, and Section 6 concludes the paper.

2. Related Work

2.1. Insulator Segmentation Model

As a key component in the stable operation of power systems, insulators have experienced a paradigm shift in segmentation techniques, evolving from traditional methods to deep learning approaches. Early studies primarily relied on potential texture, color, and other visual cues in aerial images to achieve insulator segmentation [43, 48]. However, such shallow information is highly sensitive to unstable factors such as illumination and background variations, often leading to frequent false positives and missed detections. With advances computer vision, CNN-based and Transformer-based methods have achieved remarkable progress in insulator segmentation tasks [8, 35]. Leveraging the end-to-end autonomous learning capability of these models, such models can effectively explore low-level semantic representations, reduce reliance on handcrafted features, and significantly improve adaptability to complex scenarios.

Despite the evident strengths of existing methods in local feature extraction and long-range dependency modeling, neither can effectively balance the two. CNN-based methods are inherently constrained by their local receptive fields, whereas Transformer-based methods suffer from quadratic computational complexity. In contrast, RI-Mamba is built on the Mamba architecture, which maintains linear computational complexity while employing an innovative dual-branch design. This architecture enables the adaptive integration of local details and global semantics, and establishes a continuous representational space for dual-stream features through feature reuse. As a result, the proposed model achieves a favorable balance between lightweight efficiency and segmentation accuracy.

2.2. Mamba

Compared with the self-attention mechanism of Transformer, Mamba preserves strong global context modeling capabilities while achieving linear computational complexity via the SSM, offering a significant advantage for vision tasks with limited computational resources and leading to its gradual adoption in image segmentation [21, 31, 37]. However, existing Mamba architectures typically rely on single-stream sequence modeling, limiting spatial awareness, and directly introducing bidirectional state propagation may compromise linear computational efficiency. Furthermore, the slender morphology of insulators requires the model to dynamically adjust its receptive field, whereas Mamba variants mainly rely on fixed-scale state transitions, reducing adaptability in the complex scenarios encountered in power inspection. To address these challenges, the R-Mamba block is introduced, which incorporates residual connections into the Mamba architecture to enhance gradient flow and feature reuse, thereby improving modeling flexibility while preserving linear computational complexity.

2.3. Global-local Feature Interaction

In the field of computer vision, global-local feature interaction mechanisms have long been recognized as crucial for enhancing model performance. Conventional approaches typically employ a sequential architecture to extract local details and global semantics, but these strategies often struggle to dynamically balance the contributions of features at different levels [25, 56]. To overcome this limitation, researchers have proposed various innovative interaction strategies that establish dynamic connections between local features and global context, enabling models to adaptively focus on key information across multiple scales [22, 29, 50]. Notably, in broader computer vision applications, global-local interaction mechanisms have been shown to significantly enhance model robustness and representational capacity. For instance, in salient object detection, the Co-Compensation Transformer network [2] alternates computations within and across visual tokens to explicitly capture the relationship between multi-level global context and fine-grained structural representations. In multi-modal image fusion, the Equivariant Multi-modal Image Fusion method (EMMA) [51] enhances perceptual realism and information integrity by maintaining geometric consistency across modalities. These studies further confirm the universality and effectiveness of global-local feature interaction in visual tasks. Accordingly, for segmenting slender insulators, an effective global-local interaction mechanism can substantially improve edge detail preservation while maintaining overall structural coherence. Motivated by this, the proposed G^2LIM module employs learnable gated weights to automatically adjust the fusion ratio between global and local features, further facilitating interaction between the dual-stream features in the R-Mamba block and effectively addressing varying levels of background complexity and morphological variation.

3. RI-Mamba

The proposed lightweight RI-Mamba is an end-to-end network that adopts an ''Encoding-Interaction-Decoding'' paradigm, as illustrated in Figure 1. The encoding stage of RI-Mamba consists of four hierarchical layers, each employing a parallel dual-branch to extract both fine-grained details and semantic features of insulators. Specifically, the detail branch employs the DCEU module, constructed with depth-wise separable convolution (D-DWConv), to capture subtle porcelain gaps and edge characteristics within insulator strings. Meanwhile, the semantic branch, designed from a coordinate-aware perspective, utilizes the CFM module to significantly enhance the localization accuracy in the connection regions between insulators and transmission lines. During the feature interaction stage, the proposed R-Mamba block is introduced to further refine cross-branch feature fusion. By integrating the unique mechanisms of Mamba and residual structures, R-Mamba effectively enhances feature correlation and representation capability. In the decoding stage, a G^2LIM module is incorporated to adaptively balance global contextual information and local detail representation through a learnable gating mechanism. This dynamic adjustment highlights the elongated morphology of insulators against complex backgrounds, thereby enhancing the segmentation accuracy of small-scale objects.

As discussed in the Introduction, the design of RI-Mamba aims to address two key challenges in insulator segmentation: preserving fine-grained details and balancing long-range dependency modeling with computational efficiency. The remainder of this section elaborates on the core components developed to achieve these objectives. First, the parallel dual-branch for extracting detail and semantic features is introduced (Contribution 2). Then, the R-Mamba block, which enables bidirectional feature interaction and fusion, is described in detail (Contribution 3). Finally, the G2LIM module, employed in the decoding stage for adaptive feature selection, is presented (Contribution 4). The overall lightweight design and comparable performance of the RI-Mamba network (Contribution 1) are achieved through the synergistic integration of these components.

3.1. Dual-branch Encoding

Detail-branch: The objective of the detail branch is to flexibly adjust the receptive field to capture fine-grained features at multiple scales. To this end, the DCEU module is introduced in the detail branch, as illustrated in Figure 2(a). The module first splits the feature map equally along the channel dimension into two parts, which are subsequently processed by D-DWConv layers with different configurations. Specifically, the left branch employs a 3×1 and 1×3 D-DWConv, while the right branch uses a 1×3 and 3×1 D-DWConv. This design reduces the number of parameters while maintaining an equivalent receptive field. The processed feature maps from both branches are then refined via a spatial attention mechanism (SA) to accurately identify the insulator regions. Subsequently, the outputs are concatenated to restore the original channel dimension, and a channel attention mechanism (CA) is applied to further emphasize insulator-related channels. To balance the contributions of the original and processed features, a residual connection followed by element-wise addition is employed, achieving holistic feature reassembly. Finally, a channel shuffle operation is applied to promote cross-channel interaction, yielding multi-stage, fine-grained features D_α, α 1, 2, 3, 4 of the insulator regions.

Semantic-branch: The precise location of insulators in an image can be effectively represented using coordinate information. Encoding two-dimensional spatial coordinates explicitly enhances the geometric awareness of the extracted features. To this end, the CFM module is constructed by combining Coord-Conv [19] with a parallel coordinate attention mechanism (PCAT) [9]. This module first generates position-sensitive feature embeddings based on the coordinate information, which are subsequently used by the PCAT component to compute location-specific attention weights. In this way, the module simultaneously emphasizes the semantic content and accurately localizes insulator regions. The underlying mechanism of the CFM module is illustrated in Figure 2(b).

Specifically, the CFM module first splits the feature map into two branches via Coord-Conv, with one branch augmented by coordinate i and the other by coordinate j. Both branches are then passed through a 1×1 Conv block to refine the features. Next, the features in both branches undergo *Avg Pool* and *Max Pool* along their respective directions to extract direction-sensitive spatial characteristics. These feature maps are concatenated to obtain two intermediate feature maps: SF_i with embedded coordinate information i, and SF_j with embedded coordinate information j. Subsequently, a 1×1 Conv block followed by channel splitting is applied to enable deep interaction between the two coordinate-enhanced features: the vertical feature i_{mean} from the SF_i is multiplied with the horizontal feature j_{max} from the SF_j, and the horizontal feature i_{max} from the SF_i is multiplied with the vertical feature j_{mean} from the SF_j. This operation yields the recalibrated horizontal-vertical coordinate features SF_i' and SF_j'. Meanwhile, a

Figure 1

The architecture of the proposed RI-Mamba.

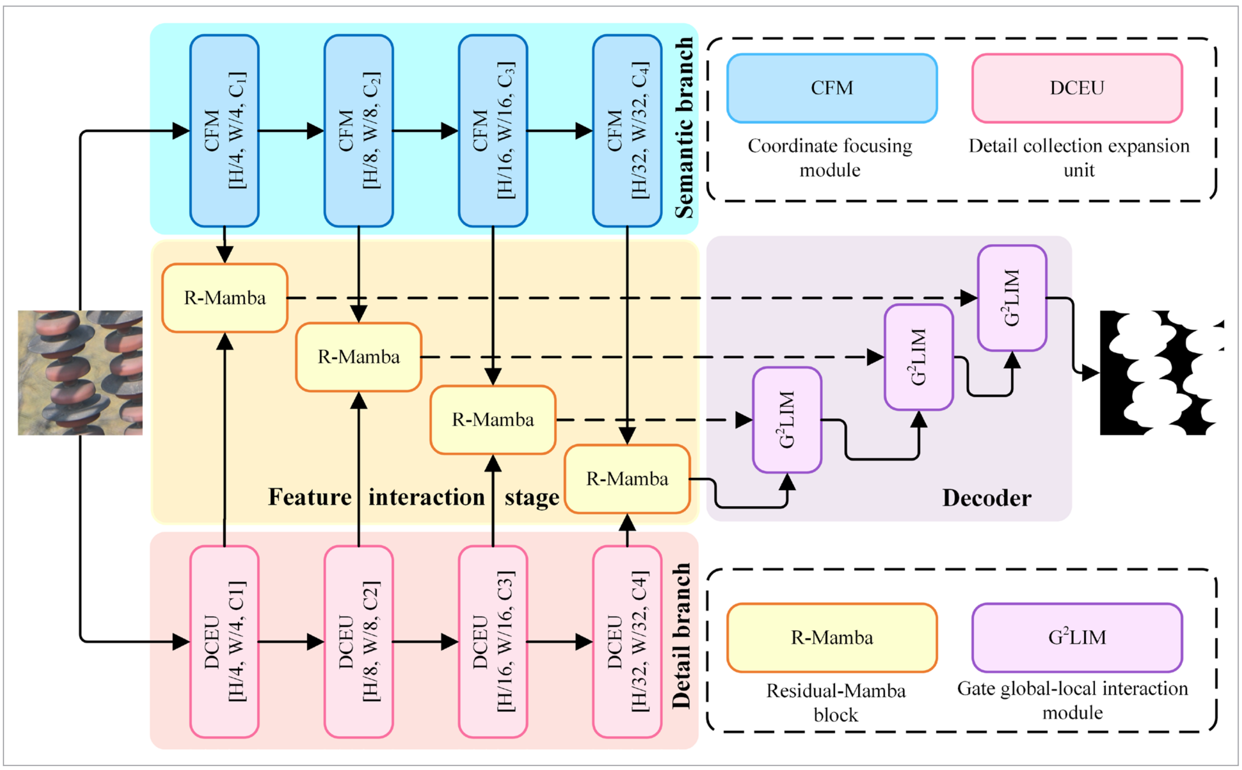
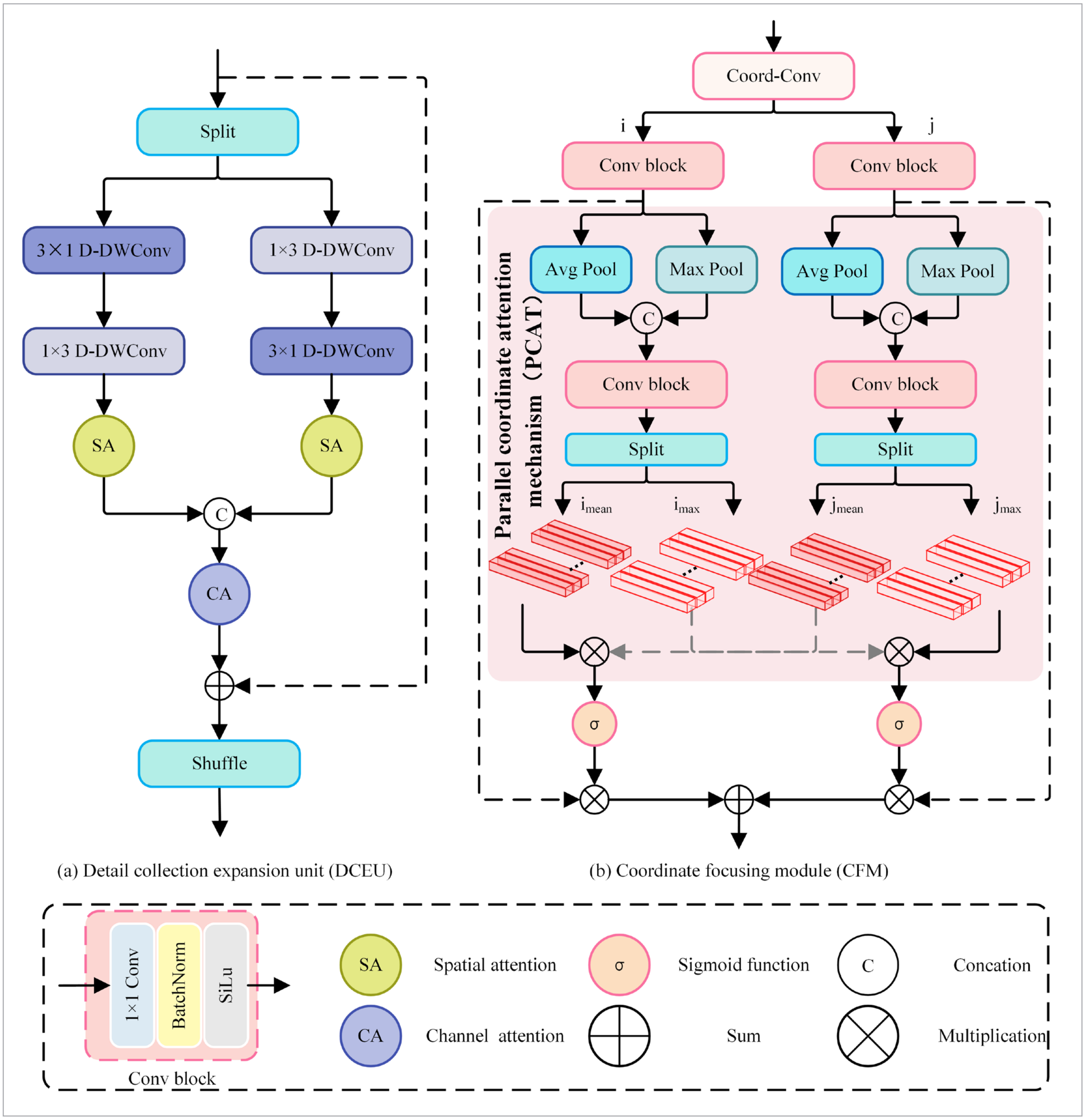


Figure 2
Schematic of the DCEU unit and the CFM unit.



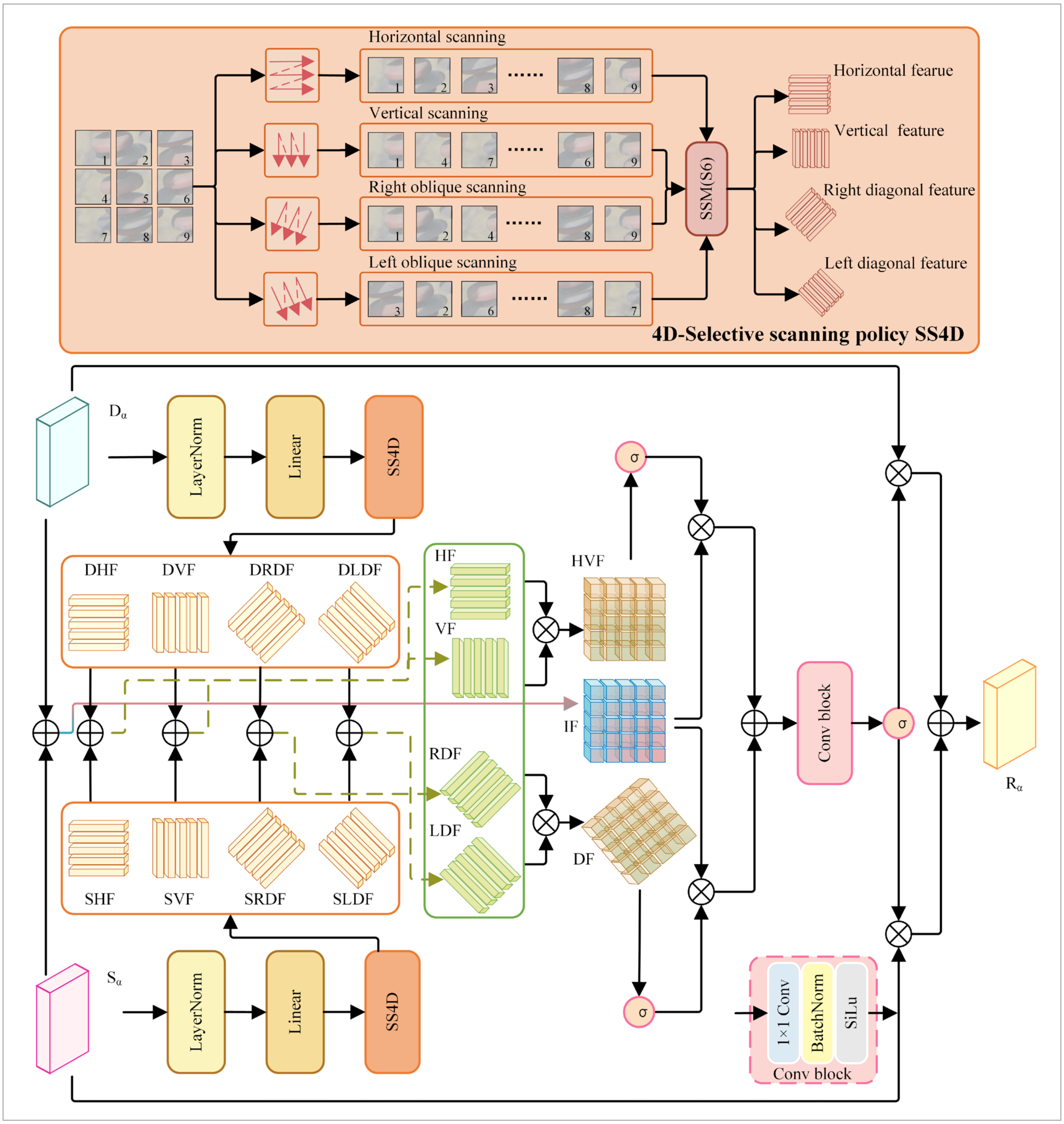
Sigmoid gating mechanism adaptively adjusts the relative importance between SF_i' and SF_j', effectively mitigating bias caused by fixed fusion strategies. Finally, the two types of coordinate features are combined via element-wise addition to produce the output S_α, $\alpha \in 1, 2, 3, 4$ of the semantic branch.

3.2. R-Mamba Block

The Mamba architecture, based on the SSM, has shown great potential in computer vision owing to its linear computational complexity and efficient long-sequence modeling capability, leading to several representative variants such as DC-Mamba, SD-Mamba, and RS³Mamba [21, 24, 46]. However, a key challenge emerges when Mamba processes long sequences: its scanning mechanism tends to overemphasize distant dependencies, thereby neglecting fine-grained details near the central region. Existing studies address this issue by incorporating convolutions to extract local features, which are subsequently fused with the long-range dependencies captured by Mamba [17, 45]. While effective, the introduction of convolution substantially increases the model size and computational burden, especially when processing deep-layer feature maps. In contrast, the proposed R-Mamba em-

Figure 3

Schematic diagram of the R-Mamba block.



ploys only the well-established SS4D [24] for long-sequence modeling. Rather than using excessive convolutions, the local component is directly derived from the original feature map through a residual structure, which injects the original detail-semantic features into the global representations obtained by SS4D. This design uses a limited number of convolutions while preserving intrinsic local information from the original features, effectively compensating for the detail loss inherent in standard Mamba.

As shown in Figure 3, he four pairs of detail features D_α, $\alpha \in 1, 2, 3, 4$ and semantic features S_α, $\alpha \in 1, 2, 3,$ 4extracted from the dual-branch encoder are first processed through the Layer Normalization (*LN*) and Linear layer (*Linear*) built into Mamba. Each pair is then projected into the dimensional space required by the SS4D, yielding the transformed feature representations D'_α and S'_α.

$$\begin{cases} D'_\alpha = Linear(LN(D_\alpha)) \\ S'_\alpha = Linear(LN(S_\alpha)) \end{cases} \quad \alpha \in 1, 2, 3, 4 \tag{1}$$

The SS4D scanning mechanism is adopted to comprehensively traverse the entire feature map through four scanning directions: horizontal, vertical, left-diagonal, and right-diagonal. Each direction is independently processed through a dedicated SSM block that captures direction-specific characteristics, including horizontal features (*DHF*, *SHF*), vertical features (*DVF*, *SVF*), left-diagonal features (*DLDF*, *SLDF*), and right-diagonal features (*DRDF*, *SRDF*). For each directional pair, the SS4D-generated features are fused through element-wise addition to form the complete directional feature flows: horizontal flow *HF*, vertical flow *VF*, left-diagonal flow *LDF*, and right-diagonal flow *RDF*.

Subsequently, the horizontal flow *HF* is multiplied by the vertical flow *VF*, and the left-diagonal flow *LDF* is multiplied by the right-diagonal flow *RDF*, generating two comprehensive directionally fused gradient flows: *HVF* and *DF*. These gradient flows are then normalized via the Sigmoid function, constraining their values within the range [0, 1] to ensure stable convergence. At this stage, the detail feature D_α, $\alpha \in 1, 2, 3, 4$ and the semantic feature S_α, $\alpha \in 1, 2, 3,$ 4 are combined via element-wise addition to obtain the interaction feature *IF*, which simultaneously preserves both fine-grained and semantic information. The *IF* is then modulated by the normalized directional values through element-wise multiplication, yielding IF_{HV} and IF_D. These two feature flows are aggregated to form the refined interaction feature IF', allowing the central regions to focus on critical information from all directional perspectives.

$$\begin{cases} IF_{HV} = IF \times \sigma(HVF) \\ IF_D = IF \times \sigma(DF) \end{cases}, \tag{2}$$

where, σ denotes Sigmoid function.

Next, the refined interaction feature IF' is processed through a 1×1 Conv followed by a Sigmoid function, decomposing the key information along each directional. Through a residual connection, the original features D_α, $\alpha \in 1, 2, 3, 4$ and S_α, $\alpha \in 1, 2, 3, 4$ are element-wise multiplied with the decomposed directional key information. This process leverages the intrinsic detail-semantic information from the original feature maps to compensate for local characteristics overlooked by SS4D, effectively restoring fine-grained details within the global feature representation.

$$\begin{cases} F_D = D_\alpha \times \sigma(Conv_{1\times1}(IF')) \\ F_S = S_\alpha \times \sigma(Conv_{1\times1}(IF')) \end{cases}, \tag{3}$$

where, $Conv_{1\times1}$ denotes 1×1 Conv block.

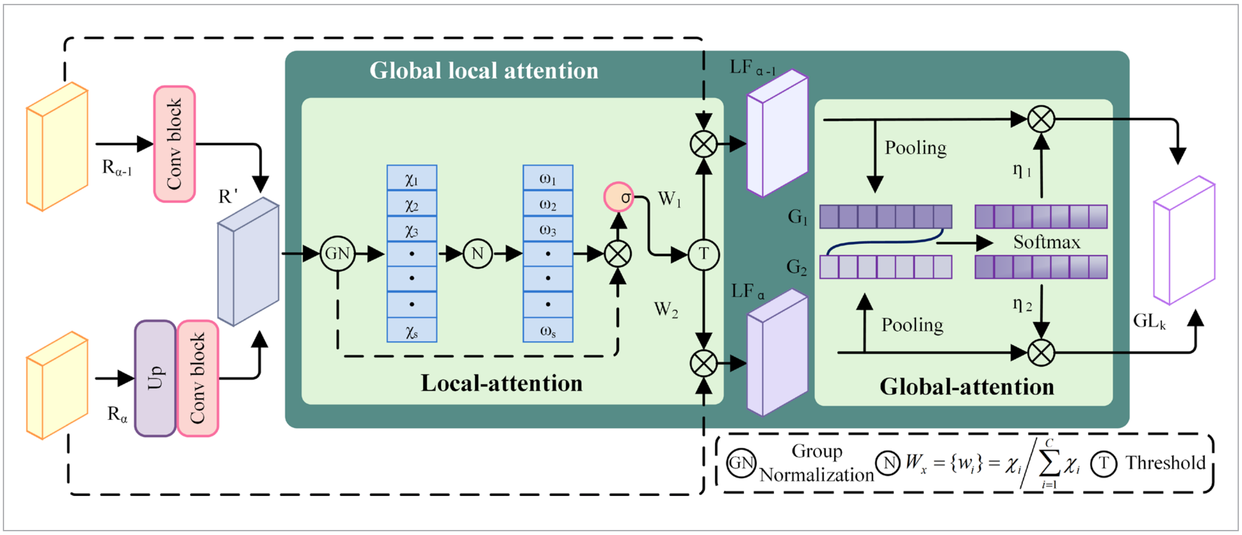
Finally, the two feature maps, which encode local characteristics and long-sequence modeling information, denoted as F_D and F_s, are combined via element-wise addition to produce the dynamically fused deep feature R_α, $\alpha \in 1, 2, 3, 4$.

3.3. G^{2}LIM Module

To further realize global-local fusion of R_α, $\alpha \in 1, 2, 3,$ 4, the G^{2}LIM module is introduced in the decoding stage. It establishes a dynamic connection between local features and the global context, allowing the model to adaptively attend to critical information across multiple scales. As shown in Figure 4, the shallow interactive feature $R_{\alpha-1}$ and the deep interactive feature R_α are first processed through a 1×1 Conv block, an upsampling operation and another 1×1 Conv block, respectively. Both $R_{\alpha-1}$ and R_α are then resized to $H_{\alpha-1} \times W_{\alpha-1} \times 64$. The resized features are subsequently combined via element-wise addition to generate the fused feature representation R'.

Figure 4

Diagram of the G²LIM module.



$$R'= Conv_{1\times 1}(R_{\alpha\text{-}1})+Conv_{1\times 1}(Up(R_{\alpha})), \tag{4}$$

where, Up denotes upsampling operation.

Subsequently, R' is fed into the local attention mechanism, where the scaling factors in Group Normalization (GN) are used to assess the information content of spatial pixels. By subtracting the mean μ from R' and dividing by the standard deviation ρ, the normalized representation R is obtained.

$$R=GN(R')=\chi\frac{R'\text{-}\,\mu}{\sqrt{\rho^2+v}}+\xi\,, \tag{5}$$

where, v is a small constant (typically 10^{-5}) introduced to ensure numerical stability during division; χ and ξ are learnable affine transformation parameters initialized to 1 and 0, respectively. In this context, the learnable parameter χ in the GN layer is used to quantify the variance of each batch and spatial pixel. Larger values of χ correspond to richer spatial cues, indicating greater variation in the associated spatial pixels. The normalized correlation weights ω_s represent the relative importance of different spatial regions.

$$\omega_s=\{\omega_i\}=\frac{\chi_i}{\sum_{j=1}^{C}\chi_j}\quad i,j=1,2,\ldots,C, \tag{6}$$

where, C denotes the number of channels. The correlation weights ω_s are then used to reweight the values of R within the range (0, 1) via a Sigmoid function followed by a threshold-based dynamic gating mechanism. Specifically, weights ewwxceeding the threshold T are set to 1, corresponding to the spatial weights W_1, whereas those below T are set to 0, corresponding to the non-spatial weights W_2 (the T set to 0.5 in the experiments). The thresholded weights W_1 and W_2 are subsequently multiplied with $R_{\alpha-1}$ and R_α, respectively, to obtain the local features $LR_{\alpha-1}$ and LR_α.

$$W=T(\sigma(\omega_s(GN(R'))\times R')) \tag{7}$$

$$\begin{cases} LR_{\alpha\text{-}1}=W_1\times R_{\alpha\text{-}1} \\ LR_{\alpha}=W_2\times R_{\alpha} \end{cases}. \tag{8}$$

Finally, $LR_{\alpha-1}$ and LR_α are fed into the global attention mechanism, where global average pooling ($Pooling$) is applied to extract the global descriptive factors G_1 and G_2 from the two feature maps. These factors are then concatenated and passed through a global soft-attention operation using a Softmax function to generate the global vectors η_1 and η_2. The global vectors η_1 and η_2 are learnable parameters produced by the Softmax function, with values constrained to the range (0, 1).

$$\begin{cases}\eta_1 = \dfrac{e^{c1}}{e^{c1}+e^{c2}} \\ \eta_2 = \dfrac{e^{c2}}{e^{c1}+e^{c2}}\end{cases} \quad \eta_1+\eta_2=1 \tag{9}$$

Guided by η_1 and η_2, $LR_{\alpha-1}$ and LR_α are merged along the spatial dimension to produce the global-local refined features GL_k, $k \in 1, 2, 3$.

$$GL_k=\eta_1 LR_{\alpha\text{-}1}+\eta_2 LR_\alpha \quad k \in 1,\ 2,\ 3 \tag{10}$$

3.4. Efficiency of RI-Mamba

To rigorously illustrate the lightweight nature of RI-Mamba from a theoretical perspective, we conduct a per-component computational complexity analysis and compare it with Transformer to highlight the overall efficiency advantage of RI-Mamba.

Computational Complexity of the dual-branch encoding: Let the input feature map have dimensions $H\times W\times C$. In the detail branch, the D-DWConv layers use kernels of size 3×1 and 1×3. The computational cost for a single forward pass can be expressed as:

$$O_{D\text{-}DWConv} = H \times W \times C \times (K_h+K_w), \tag{11}$$

where, K_h and K_w denote the height and width of the convolutional kernels, respectively. As the original feature map is split into two halves for parallel processing, the total computational cost of the detail branch remains linearly proportional to H, W, and C. In the semantic branch, Coord-Conv and PCAT are utilized. Owing to the lightweight nature of the 1×1Conv and simple coordinate transformations, the additional computational overhead is negligible. Therefore, the overall computational complexity of the dual-branch encoder can be expressed as $O(HWC)$.

Computational Complexity of the R-Mamba block: For an input feature map of size $H\times W\times C$, the SS4D scanning mechanism unfolds the feature map into sequences along the horizontal, vertical, and two diagonal directions. Each directional sequence has a length $L=Max(H, W)$. The computational complexity of the SSM for processing a sequence of length L is $O(LCN)$, where N denotes the fixed state dimension. Considering all four scanning directions, the total computational complexity of SS4D becomes:

$$O_{SS4D} = 4 \times L \times C \times N = O(LD) \tag{12}$$

Computational Complexity of the G²LIM module: The G²LIM module is based on a gated self-attention mechanism. First, the local gating mechanism computes trainable weights based on the variance in GN, with a computational complexity of $O(HWC)$. Second, the global attention mechanism generates fusion weights through global average pooling, which also has a complexity of $O(HWC)$. Therefore, the overall computational complexity of the G²LIM module remains linear with respect to the feature map dimensions, $O(HWC)$.

Based on the above analyses, all core components of RI-Mamba maintain linear computational complexity, in stark contrast to Transformer-based methods. Consequently, the total computational cost (FLops) of RI-Mamba can be expressed as the sum of the costs of all stages and components:

$$\begin{aligned}FLops_{total} = {} & \textstyle\sum_i^s O(H_iW_iC_i)+O(H_iW_iC_i) \\ & +O(L_iD_i)+O(H_iW_iC_i)=O(HWC)\end{aligned}, \tag{13}$$

where, i denotes the individual layers of the model. This linear growth in computational complexity provides the theoretical foundation for RI-Mamba's capability to achieve high-performance segmentation with minimal Params and FLops.

4. Experiments

4.1. Datasets

IS2020 dataset [10]: The dataset originates from the 8th "Teddy Cup" Data Mining Challenge (2020) and was provided by Guangzhou Intelligent Equipment Research Institute Co., Ltd. The original dataset consists of 2,859 pairs of aerial images and their corresponding insulator annotations. After data cleaning, 127 samples with inconsistent annotations or blurred images were removed, and the remaining 2,732 images were uniformly resized to 512×512. To enhance data diversity and improve model generalization, several data augmentation strategies such as translation, mirror flipping, and brightness adjustment were applied, expanding the dataset to 5319. Finally, the dataset was randomly divided into training, testing, and validation sets in a ratio of 6:3:1.

SISD dataset [53]: The super-large-scale insulator segmentation dataset (SISD) contains of 44,000 original images. A total of 15,175 invalid samples with poor annotation quality, incomplete targets, or severe occlusion were first excluded. The remaining 28,825 images were standardized and uniformly resized to 512×512. Given the large scale of the dataset, no additional data augmentation operations were applied. Finally, the dataset was randomly divided into training (17,307 images), testing (8,633 images), and validation (2,885 images) sets in a ratio of 6:3:1. This dataset encompasses diverse weather conditions, shooting angles, and complex backgrounds, offering strong scene diversity and serving as a comprehensive benchmark for evaluating the generalization capability of segmentation models.

4.2. Evaluation Metrics

For the quantitative evaluation of RI-Mamba's segmentation, we adopt commonly used metrics from recent studies [10, 24, 45, 46], namely Precision (P), Recall (R), $F1$ and mIoU. These metrics provide an objective assessment of the model's ability to accurately identify insulator regions. Furthermore, to emphasize the lightweight advantage of RI-Mamba, we evaluate model complexity and inference efficiency using Floating-point operations (*FLops*), Frames per second (*FPS*), and Parameters (*Params*). *FLops* are measured in *G*, representing one billion floating-point operations; *FPS* are measured in *FPS*, denotes the number of frames processed per second; and *Params* are measured in *M*, represents the model's memory footprint. Lower *FLops* and *Params* indicate a reduced computational burden and a more compact architecture, while higher *FPS* values correspond to greater real-time inference capability.

$$OA=\frac{TP+TN}{TP+TN+FP+TN} \tag{14}$$

$$P=\frac{TP}{TP+FP} \tag{15}$$

$$R=\frac{TP}{TP+TN} \tag{16}$$

$$F1=\frac{2\times Pre\times Rec}{Pre+Rec} \tag{17}$$

$$mIoU=\frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i+FP_i+FN_i}, \tag{18}$$

where, C denotes the number of classes; TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

4.3. Experimental details

All models were trained on an NVIDIA RTX 4060Ti GPU and implemented using the PyTorch framework with Python 3.8. AdamW was selected as the optimizer, with β_1 and β_2 set to 0.9 and 0.999, respectively. Following commonly used hyperparameter settings in insulator segmentation models[24, 26, 31, 44], the initial learning rate was set to 0.0018, with a batch size of 4, weight decay of 10^{-4}, and a total of 400 training epochs.

To ensure a fair evaluation of all models, a unified training strategy was adopted: no pretrained weights were used, and all models were trained from scratch on the same two insulator segmentation datasets. An early stopping strategy was employed during training: if the validation loss did not decrease for 50 consecutive epochs, training was terminated to prevent overfitting and improve generalization. In addition, to mitigate the impact of randomness on the experimental results, all models in the primary performance analyses were trained and evaluated using five different random seeds (42, 1234, 3047, 4096, and 2024). The performance metrics reported in this paper are presented as the meanstandard deviation across the five runs. The standard deviation reflects the sensitivity of each model to different random initializations.

4.4. Comparison Models

To comprehensively evaluate the segmentation of RI-Mamba, we compared it with 17 state-of-the-art segmentation models, including CNN-based methods, Transformer-based methods, and Mamba-based methods.

CNN-based Methods:

PSPNet [52]: This method introduces a pyramid pooling module to incorporate global context, producing high-quality results in scene parsing tasks. CALNet [6]: This method employs a conditional attention mechanism to efficiently fuse multi-dimensional features and enhance the model's representational capacity. UCM-Net[44]: By collaboratively leveraging multi-layer perceptrons (MLP) and CNN, this method significantly reduces the number of parameters and improves learning efficiency. PMFSNet [54]: This model simplifies the hierarchical structure of UNet and uses an attention-based multi-scale feature enhancement module to capture long-range dependencies. nnU-Net [11]: A stable and adaptive 2D/3D U-Net-based framework, designed to simplify network design while emphasizing key factors affecting performance and generalization. RTCNet [16]: This method utilizes a boundary refinement module to effectively aggregate intermediate features from different branches.

Table 1

Performance comparison of RI-Mamba and other methods on the IS2020 dataset, the best results are highlighted in bold.

Methods	IS2020 dataset				
	mIoU/%	OA/%	F1/%	P/%	R/%
CNN-based methods					
PSPNet	80.94±0.68	93.93±0.52	88.97±0.73	89.83±0.81	88.18±0.76
nnU-Net	81.25±1.83	93.68±1.42	88.45±1.65	89.72±1.88	87.35±1.72
CALNet	81.53±0.72	94.30±0.48	89.35±0.69	91.54±0.75	87.49±0.78
UCM-Net	82.02±0.35	94.45±0.22	89.68±0.30	91.65±0.33	87.98±0.37
PMFSNet	83.40±0.32	94.87±0.20	90.58±0.28	91.98±0.31	89.32±0.34
RTCNet	83.87±0.59	94.95±0.41	90.83±0.61	92.12±0.66	89.65±0.69
Transformer-based methods					
SegFormer	81.71±0.74	94.43±0.47	89.46±0.70	92.35±0.76	87.13±0.79
nnFormer	82.18±2.15	94.12±1.68	89.23±1.92	90.85±2.18	87.85±1.98
UNETR++	82.67±2.38	94.35±1.85	89.72±2.12	91.24±2.42	88.45±2.25
RepViT	82.97±0.67	94.58±0.44	90.32±0.65	90.47±0.71	90.17±0.73
CasaFormer	83.54±0.31	94.88±0.19	90.67±0.27	91.70±0.30	89.73±0.33
AFFormer	84.14±0.29	95.12±0.18	91.05±0.26	92.68±0.29	89.85±0.32
Mamba-based methods					
SMM-UNet	83.99±0.66	95.11±0.40	90.95±0.64	92.74±0.69	89.39±0.72
LightM-UNet	85.76±0.58	95.71±0.36	92.06±0.59	**95.16±0.64**	89.55±0.67
MF-Mamba	85.92±0.56	95.68±0.37	92.18±0.58	94.87±0.63	89.72±0.66
ECMNet	86.05±0.74	95.75±0.21	92.25±0.16	93.78±0.81	90.88±0.24
DBCNet	86.14±0.25	95.79±0.14	92.30±0.22	93.60±0.24	91.13±0.27
RI-Mamba*	**86.28±0.32**	**95.81±0.16**	**92.40±0.26**	93.29±0.30	**91.56±0.28**

Table 2

Performance comparison of RI-Mamba and other methods on the SISD dataset, the best results are highlighted in bold.

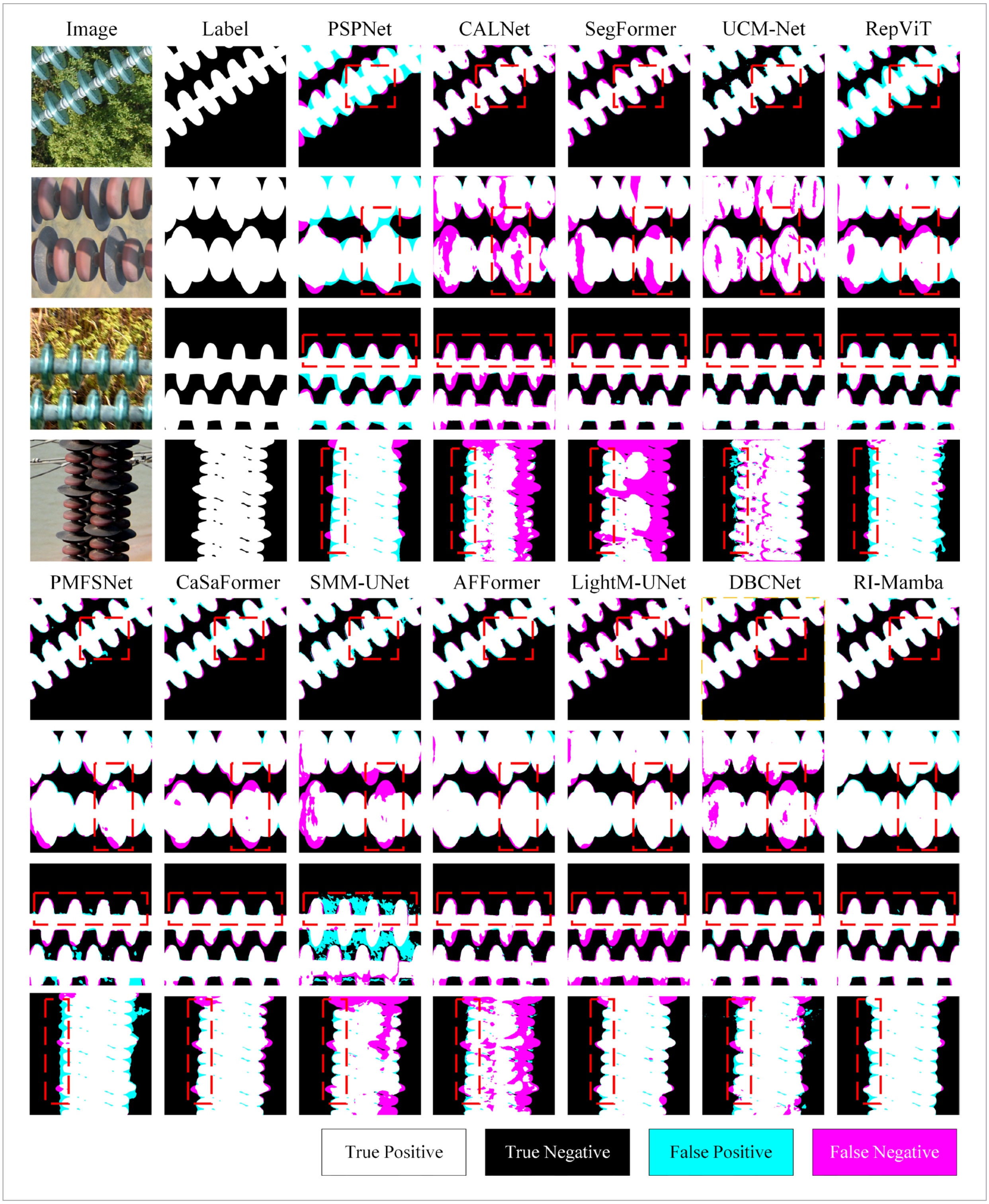
Methods	SISD dataset				
	mIoU/%	OA/%	F1/%	P/%	R/%
CNN-based methods					
PSPNet	84.99±0.75	97.30±0.45	91.42±0.72	92.39±0.78	90.51±0.81
nnU-Net	85.34±2.25	97.12±1.78	91.78±2.08	92.67±2.35	91.05±2.18
CALNet	86.03±0.72	97.63±0.42	92.09±0.69	95.52±0.75	89.23±0.79
UCM-Net	87.55±0.32	97.78±0.16	93.06±0.26	93.41±0.28	92.71±0.31
PMFSNet	88.27±0.30	97.95±0.15	93.50±0.24	94.38±0.68	92.66±0.29
RTCNet	88.76±0.63	98.03±0.37	93.82±0.61	94.25±0.66	93.45±0.69
Transformer-based methods					
SegFormer	87.02±0.73	97.85±0.43	92.71±0.70	96.99±0.76	89.28±0.80
nnFormer	87.89±2.48	97.65±1.92	92.95±2.28	93.82±2.58	92.25±2.42
UNETR++	88.02±2.75	97.78±2.18	93.28±2.52	94.15±2.85	92.65±2.68
RepViT	88.33±0.29	97.86±0.14	93.54±0.23	92.24±0.25	94.94±0.28
CasaFormer	88.98±0.27	98.17±0.13	93.92±0.21	97.23±0.23	91.14±0.26
AFFormer	90.53±0.60	98.34±0.36	94.86±0.59	94.77±0.64	94.96±0.68
Mamba-based methods					
SMM-UNet	88.50±0.69	98.11±0.42	93.63±0.66	**97.74±0.72**	90.30±0.75
LightM-UNet	89.67±0.26	98.27±0.13	94.34±0.20	96.91±0.56	92.10±0.55
MF-Mamba	90.11±0.61	98.18±0.37	94.62±0.53	92.56±0.66	**96.94±0.69**
ECMNet	90.35±0.59	98.32±0.39	94.78±0.61	96.45±0.66	93.25±0.70
DBCNet	91.02±0.18	98.39±0.67	94.95±0.59	93.82±0.84	92.18±0.47
RI-Mamba*	**91.59±0.30**	**98.56±0.11**	**95.48±0.18**	96.04±0.20	94.93±0.22

Transformer-based Methods:

SegFormer [39]: This method employs a lightweight MLP-based decoder to fuse multi-level features, combining local and global attention mechanisms to achieve powerful feature representations. RepViT [33]: This method integrates an efficient architecture based on lightweight Vision Transformers (ViT), progressively enhancing the computational efficiency and mobile-friendliness of MobileNetV3. CaSaFormer [14]: The proposed cross-attention gate fusion module (CAGF) effectively integrates complementary information from global semantic and local spatial features. AFFormer [4]: This method replaces the traditional decoder with prototype-based representations serving as learnable local descriptors, thereby preserving rich semantic information within high-resolution feature maps. nnFormer [55]: This method incorporates skip attention within a U-Net framework, replacing conventional concatenation or summation operations in the skip connections. UNETR++ [23]: This method shares the weights of the query and key-value mapping

Figure 5

Visualization results of RI-Mamba with other models on the IS2020 datasets, the red boxes are drawn to highlight the advantages of RI-Mamba.



functions to achieve complementary benefits while reducing the overall number of parameters.

Mamba-based Methods:

SMM-Net [13]: This method combines selective fusion Mamba (SF-Mamba) and multi-scale fusion Mamba (MF-Mamba) to achieve dynamic and adaptive integration of local and global features. LightM-UNet [15]: The pure Mamba-based architecture extracts deep semantic features while modeling long-range spatial dependencies with linear computational complexity. DBCNet [47]: This method incorporates an enhanced VMamba module as a contextual extraction branch to achieve more accurate pixel-level segmentation. MF-Mamba [38]: This method utilizes an eight-directional selective scanning mechanism combined with multi-kernel parallel convolution to capture rich global-local contextual information. ECMNet [5]: This method integrates a Mamba-enhanced feature fusion module in capsule framework, significantly improving segmentation accuracy.

4.5. Comparison Study

Quantitative Analysis: Tables 1 and 2 summarize the experimental results of RI-Mamba on the IS2020 and SISD datasets, respectively. Compared with the three types of baseline methods, RI-Mamba achieves comparable performance in mIoU, OA, and F1, demonstrating slightly higher segmentation accuracy and robustness. Specifically, on the relatively small but complex-background IS2020 dataset, RI-Mamba attains an mIoU of 86.28%, while on the larger and more diverse SISD dataset, its mIoU increases to 91.59%, indicating strong generalization capability. Notably, RI-Mamba exhibits high R and F1 on both datasets, indicating that it significantly enhances R while maintaining a balanced relationship between P and R, with discrepancies remaining within a reasonable range. Compared with PMFS-Net, RI-Mamba shows performance improvements ranging from 0.94% to 3.63% across multiple metrics on both datasets, with particularly pronounced gains in overall segmentation quality indicators such as mIoU and R. This indicates that RI-Mamba's parallel dual-branch effectively extracts detail-semantic features of the targets, providing a higher-quality basis for subsequent feature fusion. In addition, the standard deviation of RI-Mamba's performance across both datasets is below 0.32%, indicating low sensitivity to parameter initialization and robust training behavior. This ensures stable and consistent performance across various hardware environments and training cycles.

To further evaluate the effectiveness of RI-Mamba, we analyzed the training curves of all models on both datasets. As shown in Figure A1(a) and Figure A2(a), the mIoU curves of RI-Mamba display a rapid and stable upward trend, eventually converging to high values, indicating its strong representation learning capability. Although the P-R curve areas of RI-Mamba are relatively smaller on both datasets (Figure A1(c) and Figure A2(c)), the corresponding F1 curves (Figure A1(b) and Figure A2(b)) show that the model maintains a well-balanced trade-off between P and R. In addition, the loss curves (Figure A1(d) and Figure A2(d)) show that RI-Mamba undergoes a rapid and stable decrease in loss, indicating an efficient optimization process that is less prone to local minima. This superior convergence behavior can be attributed to the residual structure and gating mechanism, which jointly enhance gradient flow and improve model stability during training.

Qualitative Analysis. Figures 5-6 illustrate the qualitative comparison results of RI-Mamba and other models on the two datasets. On the IS2020 dataset, for regions such as insulator edges, porcelain gaps, and adhesive areas between the insulator and background, CNN-based methods tend to produce blurred or fragmented boundaries due to their limited receptive fields. In contrast, Transformer-based methods are prone to introduce background noise due to similar texture patterns. Benefiting from the detail branch, RI-Mamba explicitly enhances edge and local structural representations, effectively distinguishing insulators from complex backgrounds (the last three rows in Figure 5). Moreover, variations in camera viewpoints often cause insulator strings to appear tilted rather than strictly horizontal or vertical. Such geometric variations demand a model with strong spatial perception (the first row in Figure 5, the third row in Figure 6). Conventional convolutional networks and unidirectional sequence modeling struggle to capture the global structural continuity, resulting in fragmented or partially segmented outputs. RI-Mamba addresses this issue by employing

Figure 6

Visualization results of RI-Mamba with other models on the SISD datasets, the red boxes are drawn to highlight the advantages of RI-Mamba.

Table 3

Ablation study of RI-Mamba on the two datasets, the best results are highlighted in bold.

RI-Mamba				IS2020/%		SISD/%		FLops/G	Params/M
Detail branch	Semantic branch	R-Mamba	G²LIM	mIoU	F1	mIoU	F1		
√	--	--	--	80.70±0.45	88.80±0.52	83.97±0.38	90.75±0.41	2.18	2.08
--	√	--	--	77.89±0.61	86.87±0.58	81.89±0.55	89.35±0.49	2.29	2.05
√	√	--	--	81.73±0.32	89.47±0.29	86.22±0.36	92.21±0.33	4.04	4.07
√	√	--	√	83.14±0.28	90.41±031	88.17±0.27	93.43±0.25	4.23	4.08
√	√	√	--	84.09±0.26	91.02±0.24	89.06±0.22	93.97±0.21	5.11	4.14
√	√	√	√	**86.28±0.32**	**92.40±0.26**	**91.59±0.30**	**95.48±0.18**	5.31	4.15

an SS4D scanning strategy that assigns independent SSM to each scanning direction for feature extraction. This design enables unbiased and coherent modeling of long-range insulator structures with arbitrary orientations, yielding complete and geometrically consistent segmentation outputs. Compared with the IS2020 dataset, the SISD dataset presents denser insulator distributions, causing most models to misidentify multiple adjacent insulators as a single connected object, thereby blurring inter-insulator boundaries. In contrast, RI-Mamba leverages global-local feature interactions with adaptive gating weights to balance semantic and structural contributions under varying environmental conditions (the second and fourth rows in Figure 6), enabling more precise boundary delineation and improved instance separability.

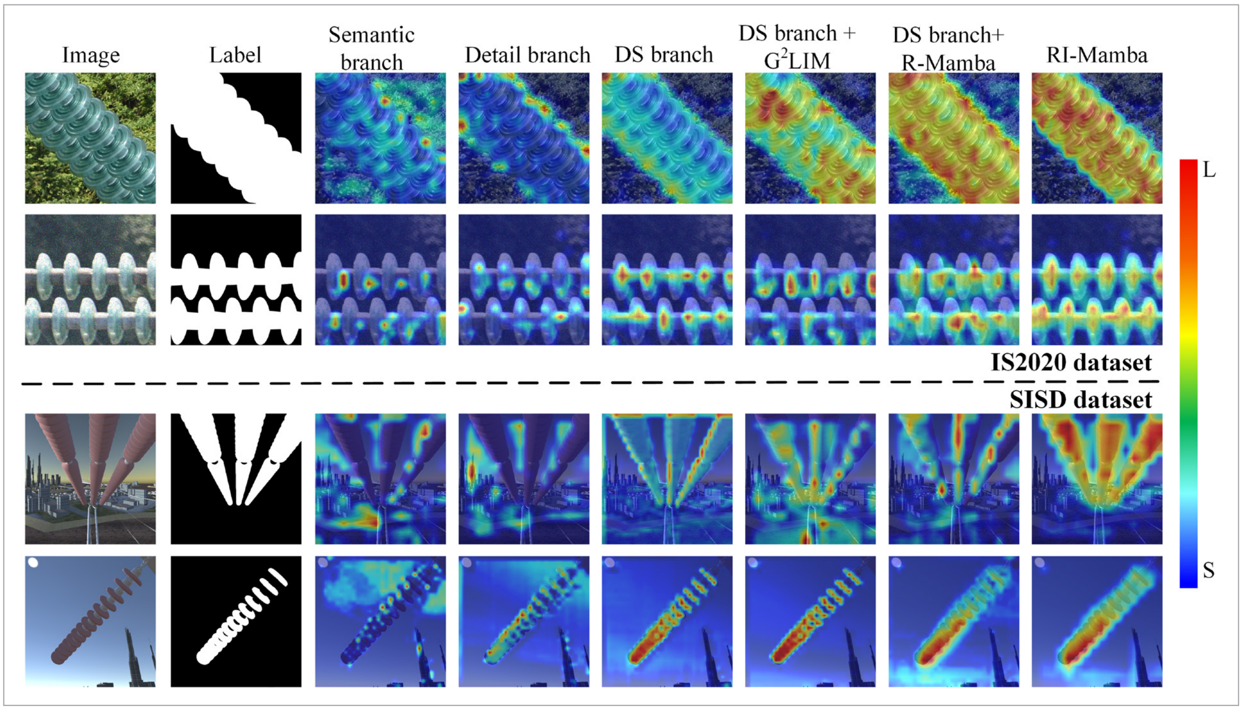
4.6. Ablation Study

We conducted systematic ablation experiments to evaluate the effectiveness and efficiency of each component in RI-Mamba, with the results summarized in Table 3. When only the semantic branch is used, the model attains the lowest number of Params and FLops, but also exhibits the weakest performance, achieving mIoU of 77.89% and 81.89% on the two datasets. After introducing the detail branch, the Params increases by only 2.02M, while the F1 improves by 2.60% and 2.86%, respectively. This demonstrates that both semantic and detail information are equally critical for insulator segmentation tasks. In particular, for slender-shaped insulators, the detail branch effectively captures edge characteristics and fine-grained local structures that are challenging for the semantic branch to extract. Furthermore, integrating the R-Mamba block and G^{2}LIM module into the dual-branch architecture yields distinct performance gains. The R-Mamba block delivers slightly higher the mean improvement, increasing mIoU by 2.36% and 2.84%, while adding only 0.82M Params and 1.3 GFLops. This confirms that incorporating the residual interaction mechanism into Mamba variants to reduce the frequency of convolution usage not only compensates for the limitations of the Mamba in feature modeling but also enables effective dual-stream feature interaction with minimal computational overhead. In comparison, the G^{2}LIM module contributes a smaller yet consistent improvement over the baseline dual-branch backbone, adding only 0.31M Params and 0.7GFLops. This indicates that its gated dynamic fusion mechanism enhances global-local feature interactions efficiently, achieving improved contextual consistency with negligible computational cost. Ultimately, the complete RI-Mamba, integrating all modules, achieves comparable segmentation performance with only 4.15M Params and 5.31 GFLops. These findings confirm that RI-Mamba effectively integrates semantic and detail features through multi-level interaction mechanisms, achieving a favorable balance between accuracy and efficiency within constrained model complexity and parameter budgets.

To further investigate the contribution of each component in RI-Mamba from the perspective of the feature space, we visualized the feature representations of different configurations using heatmaps, as shown

Figure 7

Visualization of the contribution of each component to RI-Mamba on two datasets.



in Figure 7. The final-layer activation maps were superimposed on the original images to highlight the extracted insulator features. When only the semantic branch is used, RI-Mamba can roughly localize the main body of insulators; however, its attention remains scattered and is easily affected by background. After incorporating the detail branch, the activation responses around insulator edges and porcelain gaps become significantly stronger, indicating that the detail branch effectively captures fine-grained local features that complement semantic cues. Moreover, after embedding the R-Mamba block into the dual-branch architecture, the heatmaps exhibit more continuous and uniform activation distributions, particularly along elongated insulator strings, where response intensity increases notably. This improvement arises from the residual interaction mechanism, which efficiently integrates semantic and detail features, thereby enhancing long-range structural modeling without introducing additional parameter overhead. Finally, with the integration of the G^2LIM module, RI-Mamba assigns higher adaptive weights to the previously captured insulator regions through its gated self-attention mechanism. This further strengthens the model's focus on both global structures and local details, resulting in more discriminative and coherent feature representations.

4.7. Influence of Different Dilation Rates

To further examine how different dilation rate in the D-DWConv of the DCEU unit influence the detail branch's ability to segment insulator regions, we conducted comparative experiments using six sets of dilation rate, as summarized in Table 4. The segmentation performance declines when identical dilation rates are applied across all layers. This degradation occurs because shallow and deep layers require receptive fields of different sizes; a uniform dilation rate cannot effectively capture the diverse detail distributions across stages. Moreover, different dilation yield noticeably varied evaluation outcomes. Specifically, when the dilation rates are set to (2, 4, 6, 8), the model exhibits the lowest performance. The smaller dilation span restricts receptive field expansion, resulting in overly localized detail features that difficult to effectively integrated with semantic representations. As the dilation span increases, the mIoU on the IS2020 dataset improves from 80.64% to 86.28%. However,

Table 4

Performance of RI-Mamba with different dilation rates in the DCEU unit on the two datasets, the best results are highlighted in bold.

Dilation rates	IS2020/%			SISD/%		
	mIoU	F1	R	mIoU	F1	R
(1, 3, 5, 7)	85.59±0.85	91.97±0.62	91.40±0.74	90.42±0.63	94.80±0.78	96.37±0.35
(2, 2, 2, 2)	84.64±0.69	91.36±0.57	89.41±0.44	87.99±0.42	93.32±0.90	91.17±0.71
(4, 4, 4, 4)	85.43±0.45	91.88±0.34	92.16±0.66	88.78±0.78	93.81±0.56	92.19±0.43
(2, 4, 6, 8)	80.64±0.77	88.78±0.42	88.18±0.58	87.95±0.36	93.29±0.61	90.21±0.72
(3, 7, 9, 13)	86.28±0.32	92.40±0.26	91.56±0.28	91.59±0.30	95.48±0.18	94.93±0.22
(4, 8, 12, 16)	84.71±0.81	91.42±0.64	90.17±0.81	88.25±0.42	93.48±0.39	91.17±0.51

when the dilation gap becomes excessively large, feature extraction in the shallow layers is affected by the ″gridding effect″ causing a decline in segmentation accuracy. Based on these observations, we ultimately set the D-DWConv dilation rates to (3, 7, 9, 13), achieving the favorable balance between receptive field coverage and fine-detail preservation.

4.8. Effectiveness of R-Mamba

To further evaluate the performance differences among advanced Mamba variants on the insulator segmentation task, we replaced the R-Mamba block with several representative variants, including CroMB[28], ConMB[28], SRCM[45], and T-Mamba[24]. The corresponding results are summarized in Table 5. We observed that although CroMB and ConMB have Params similar to R-Mamba, their FLops increase significantly, accompanied by a notable decline in segmentation accuracy. Specifically, the mIoU of ConMB drops to 81.91% and 87.42%, while its FLops increase to 1.25 times those of R-Mamba. This indicates that simple structural modifications of existing Mamba variants are insufficient to achieve an optimal balance between performance and efficiency. In contrast, SRCM and T-Mamba, which incorporate convolutional operations, demonstrate enhanced segmentation sensitivity, validating the benefits of convolution for local feature extraction. However, this improvement comes at the cost of increased model complexity, as their Params increase by approximately 3.5%-9.8%. By comparison, the proposed R-Mamba employs a residual interaction strategy that mitigates the limited local feature modeling ability of the standard Mamba architecture without introducing additional convolutional operations. Consequently, R-Mamba maintains lower Params and FLops while achieving comparable segmentation results. These findings confirm the effectiveness of integrating re-

Table 5

Comparison of different mamba variants integrated into RI-Mamba on the two datasets, the best results are highlighted in bold.

	FLops/G	Params/M	IS2020/%			SISD/%		
			mIoU	F1	OA	mIoU	F1	OA
ConMB	6.67	4.25	81.91±0.31	89.24±0.35	93.91±0.28	87.42±0.33	92.59±0.38	96.91±0.42
CroMB	5.95	4.25	82.93±0.28	90.30±0.32	94.43±0.25	88.36±0.29	93.54±0.34	98.08±0.27
SRCM	5.58	4.40	83.57±0.22	90.71±0.26	94.73±0.41	88.55±0.54	93.68±0.38	97.89±0.33
T-Mamba	5.56	4.67	84.96±0.21	91.58±0.45	95.33±0.19	89.81±0.44	94.43±0.30	98.25±0.58
R-Mamba	**5.31**	**4.15**	**86.28±0.32**	**92.40±0.26**	**95.81±0.16**	**91.59±0.30**	**95.48±0.18**	**98.56±0.11**

Table 6

Comparison of different global-local interaction modules integrated into RI-Mamba on the two datasets, the best results are highlighted in bold.

	FLops/G	Params/M	IS2020/%			SISD/%		
			mIoU	F1	OA	mIoU	F1	OA
GLFIM	5.89	4.19	82.66±0.26	90.15±0.33	94.15±0.47	87.42±0.46	92.59±0.52	96.91±0.83
GLFCFN	6.76	4.20	83.55±0.65	91.33±0.58	95.05±0.53	88.36±0.47	93.54±0.68	98.08±0.55
GLFB	9.44	4.30	84.71±0.37	91.41±0.49	95.35±0.34	88.55±0.27	93.68±0.31	97.89±0.25
GLSTM	8.45	4.24	85.66±0.22	92.01±0.26	95.57±0.21	89.81±0.24	94.43±0.28	98.25±0.22
G²LIM	5.31	4.15	**86.28±0.32**	**92.40±0.26**	**95.81±0.16**	**91.59±0.30**	**95.48±0.18**	**98.56±0.11**

sidual structures into Mamba architectures and offer valuable insights for the developing future vision-oriented Mamba models.

4.9. Evaluation of G²LIM

To comprehensively assess the feature fusion capability of the G²LIM module, we replaced it within the RI-Mamba with four representative global-local interaction modules, including GLFIM[56], GLFCFN[25], GLFB[26], and GLSTM[22]. The corresponding results are presented in Table 6. In terms of Params, all modules remained below 4.5M, satisfying the design requirements of a lightweight segmentation model. Among them, GLFIM exhibited FLops most comparable to G²LIM; however, due to its multimodal-oriented architecture, it was unable to fully exploit its potential in a single-modality scenario, leading to inferior segmentation accuracy. In contrast, GLFB, which employs a self-attention mechanism, achieved relatively high segmentation performance but at the expense of computational efficiency, its FLops increased to 1.78 times of G²LIM. By comparison, the proposed G²LIM dynamically adjusts the fusion weights between global contextual and local detailed features through a learnable gating mechanism, achieving favorable segmentation performance while maintaining high computational efficiency. Specifically, G²LIM achieved OA of 95.81% and 98.56% on the two datasets, respectively. These results clearly demonstrate that integrating a dynamic gating strategy into the global-local attention mechanism effectively balances accuracy and efficiency, making G²LIM a more adaptable and robust design for lightweight segmentation tasks.

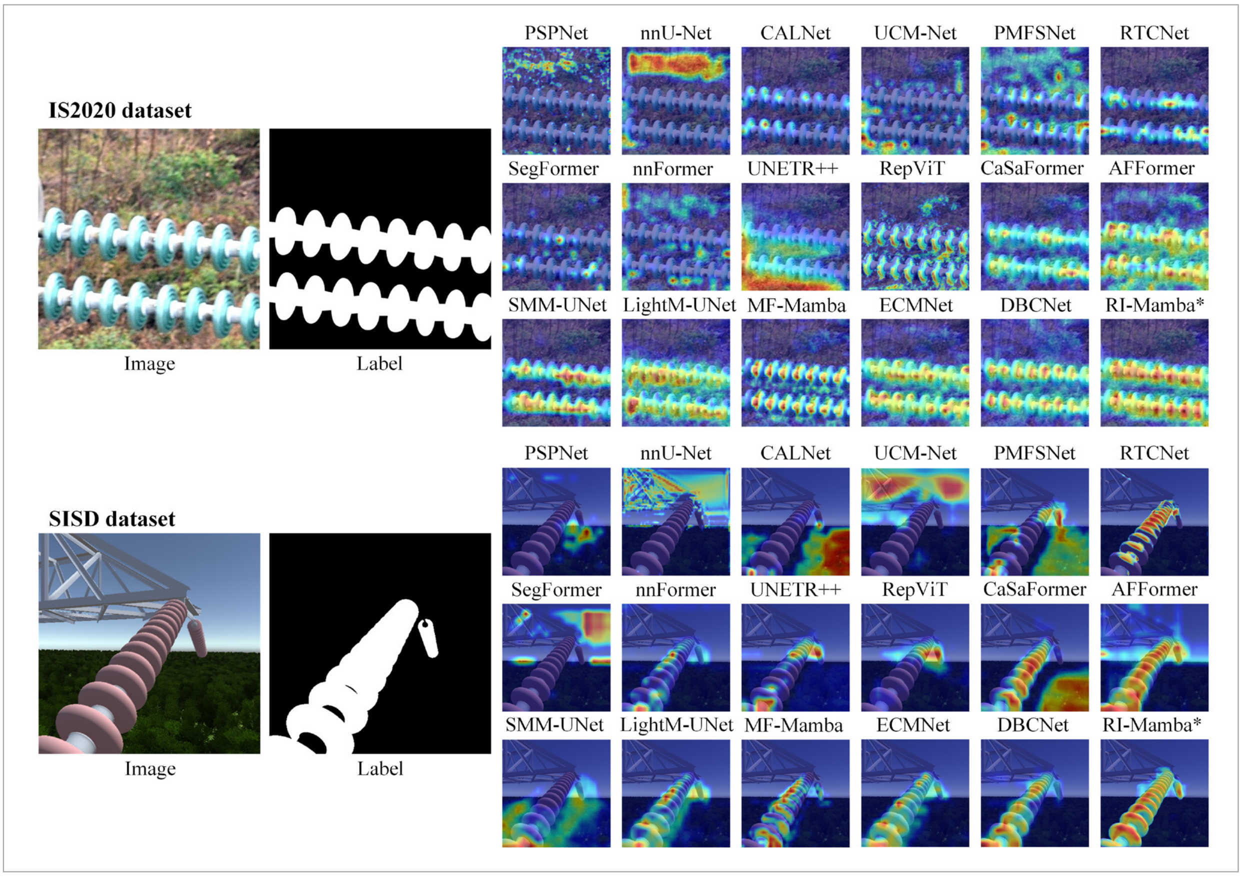
5. Discussion

5.1. Insulator Class Analysis of Each Model

To further elucidate the decision-making mechanisms of RI-Mamba compared with other lightweight models in the insulator segmentation task, we systematically analyzed and compared their attention distributions, as illustrated in Figure 8. For CNN-based methods, the attention maps primarily concentrated on high-frequency regions of the insulators such as porcelain edges, due to their inherently limited receptive fields. However, these methods exhibit insufficient continuity in capturing the overall structure, leading to response discontinuities when insulators appear in non-horizontal or non-vertical orientations. Transformer-based methods, while capable of capturing global context via self-attention, tend to produce diffuse activations in complex backgrounds, often introducing interference patterns resembling ″salt-and-pepper noise″. In contrast, Mamba-based methods demonstrate superior long-range modeling capabilities, generating more continuous responses along the insulator backbone. Nevertheless, lacking effective synergy between detail and semantic information, they still suffer from blurred boundaries and partial omission of fine structural. Benefiting from the residual interaction mechanism embedded in the R-Mamba block, RI-Mamba achieves highly confident and coherent activations across the entire spatial structure of insulators, while maintaining fine-grained responses in detailed regions such as porcelain gaps and connector areas. Moreover, the G²LIM module, through its gated self-attention

Figure 8

Visualized heatmaps of RI-Mamba and other lightweight segmentation models.



mechanism, further strengthens the weighting of key regions, enabling the model to retain strong feature discriminability even under complex background conditions. This demonstrates RI-Mamba's superior robustness and balanced capability in both global perception and local detail representation.

5.2. Lightweight Analysis

To investigate the lightweight characteristics of RI-Mamba, we conducted a comprehensive comparison with several state-of-the-art lightweight segmentation models, considering both theoretical computational complexity and practical inference speed, as summarized in Table 7. In terms of theoretical computational burden, RI-Mamba achieves significantly lower Params and FLops than most CNN-based and Transformer-based methods, and even surpasses several Mamba-based methods. Regarding practical inference efficiency, RI-Mamba attains 82.11 FPS, second only to DBCNet (93.76 FPS), which has a substantially larger Params, demonstrating excellent real-time processing capability. Notably, RI-Mamba achieves the "low Params, low FLops, high performance" characteristics through several lightweight design strategies, including: a parallel dual-branch lightweight architecture; linear-complexity modeling based on Mamba; a residual interaction mechanism in R-Mamba that replaces the need for numerous convolutional operations, and the gated dynamic feature selection. These design choices effectively reduce redundant computations while maintaining efficient feature extraction, thereby fully validating the effectiveness of RI-Mamba's lightweight design.

5.3. Analysis of Radom Seeds

To further quantify the impact of random seeds on the performance of RI-Mamba and verify its stability, we

Table 7

Performance comparison of RI-Mamba with lightweight segmentation models, the best results are highlighted in bold.

	FLops/G	Params/M	FPS/s
CNN-based methods			
PSPNet	20.13	25.35	68.12
nnU-Net	29.11	45.20	55.10
CALNet	15.96	7.26	47.46
UCM-Net	6.27	4.55	38.22
PMFSNet	6.44	4.33	78.55
RTCNet	7.16	12.74	71.38
Transformer-based methods			
SegFormer	13.24	13.68	53.23
nnFormer	28.95	52.16	32.50
UNETR++	25.60	48.88	38.51
RepViT	18.75	4.82	57.82
CasaFormer	5.99	7.33	26.20
AFFormer	11.60	5.02	38.72
Mamba-based methods			
SMM-UNet	30.35	18.88	42.71
LightM-UNet	6.01	8.20	64.62
MF-Mamba	8.17	9.22	58.96
ECMNet	10.79	14.62	57.27
DBCNet	29.18	19.73	93.76
RI-Mamba*	5.31	4.15	82.11

conducted an additional analysis using the same five random seeds (42, 1234, 3047, 4096, and 2024) employed in the main experimental evaluations. The results are presented in Table 8. The model exhibits remarkable stability across all seeds, with mIoU values ranging from 85.25% to 86.28% on the IS2020 dataset and 90.58% to 91.59% on the SISD dataset. Moreover, the standard deviation of all evaluation metrics remains below 0.5%, indicating that RI-Mamba is largely insensitive to parameter initialization and data randomness, thus demonstrating strong robustness and reproducibility. Notably, when the random seed is set to 2024, the model consistently achieves comparable performance on both datasets.

5.4. Limitation and Future Work

Although RI-Mamba has demonstrated outstanding segmentation performance across various complex scenarios, it still exhibits certain limitations under extreme conditions, as illustrated in Figure 9. In the IS2020 dataset, RI-Mamba tends to misclassify overexposed regions as insulators, suggesting that its robustness to illumination variations remains limited. In particular, under strong reflections or severe overexposure, maintaining consistency between local details and semantic context becomes challenging. In addition, a common challenge observed in both datasets is the mis-segmentation of insulators that are severely occluded or exhibit high texture similarity with the background. For instance, in the

Figure 9

Visualization limitations of RI-Mamba on the two datasets.

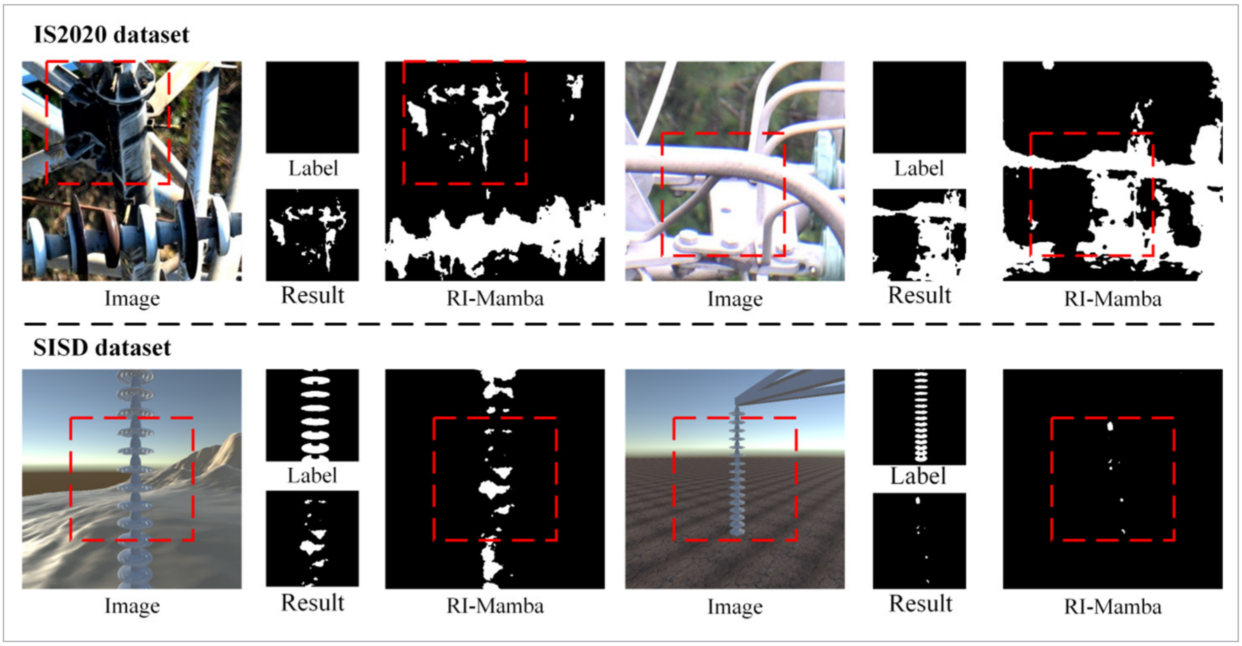


Table 8

Performance stability analysis of RI-Mamba under different random seeds on two datasets, the best results are highlighted in bold.

Radom seeds	IS2020/%			SISD/%		
	mIoU	F1	OA	mIoU	F1	OA
42	86.15±0.29	92.31±0.24	95.77±0.45	91.47±0.22	95.39±0.26	98.53±0.21
1234	85.33±0.35	91.89±0.30	94.75±0.18	90.65±0.38	94.55±0.22	97.59±0.13
3047	86.20±0.31	92.36±0.27	95.79±0.31	91.52±0.29	95.43±0.18	98.48±0.10
4096	85.25±0.33	91.42±0.47	94.82±0.19	90.58±0.44	94.50±0.21	97.55±0.12
2024	**86.28±0.32**	**92.40±0.26**	**95.81±0.16**	**91.59±0.30**	**95.48±0.18**	**98.56±0.11**

SISD dataset, insulators partially covered by sand are frequently misclassified due to class confusion, leading to incomplete segmentation.

To overcome these limitations, future research could incorporate illumination-invariant feature learning mechanisms to enhance the model's generalization capability under extreme lighting conditions. Furthermore, for scenarios involving severe occlusion or structural overlap, developing more discriminative and adaptive dynamic feature selection mechanisms could further strengthen the model's ability to distinguish between inter-class similarity and improve segmentation completeness in highly complex environments.

6. Conclusion

In this paper, we propose a lightweight dual-branch residual interaction network (RI-Mamba) based on the Mamba architecture, for insulator segmentation. The model employs a parallel dual-branch structure to extract complementary detail-semantic representations of insulator regions. Specifically, the detail branch aggregates and expands feature categories to enhance the perception of fine-grained local structures, whereas the semantic branch, from a coordinate-aware perspective, captures the positional relationships and semantic correlations among insulators. To enable long-range modeling between the dual feature streams, the R-Mamba block facilitates effective feature reuse via its residual-enhanced gradient flow. Furthermore, the G^{2}LIM module introduces a gated self-attention mechanism to dynamically adjust the fusion weights between global and local representations during upsampling, thereby enhancing feature discriminability in complex visual environments. Compared with existing insulator segmentation models, RI-Mamba achieves a well-balance between detection accuracy and computational efficiency. It precisely delineates the overall structure of insulators while meeting the requirements for lightweight deployment, making it both accurate and practical for real-world applications.

Acknowledgement

This work was supported by the Youth Science Foundation of Gansu Province (Grant No. 23JRRA16 95) and the Youth Science Foundation of Lanzhou Jiaotong University (Grant No. 2023009).

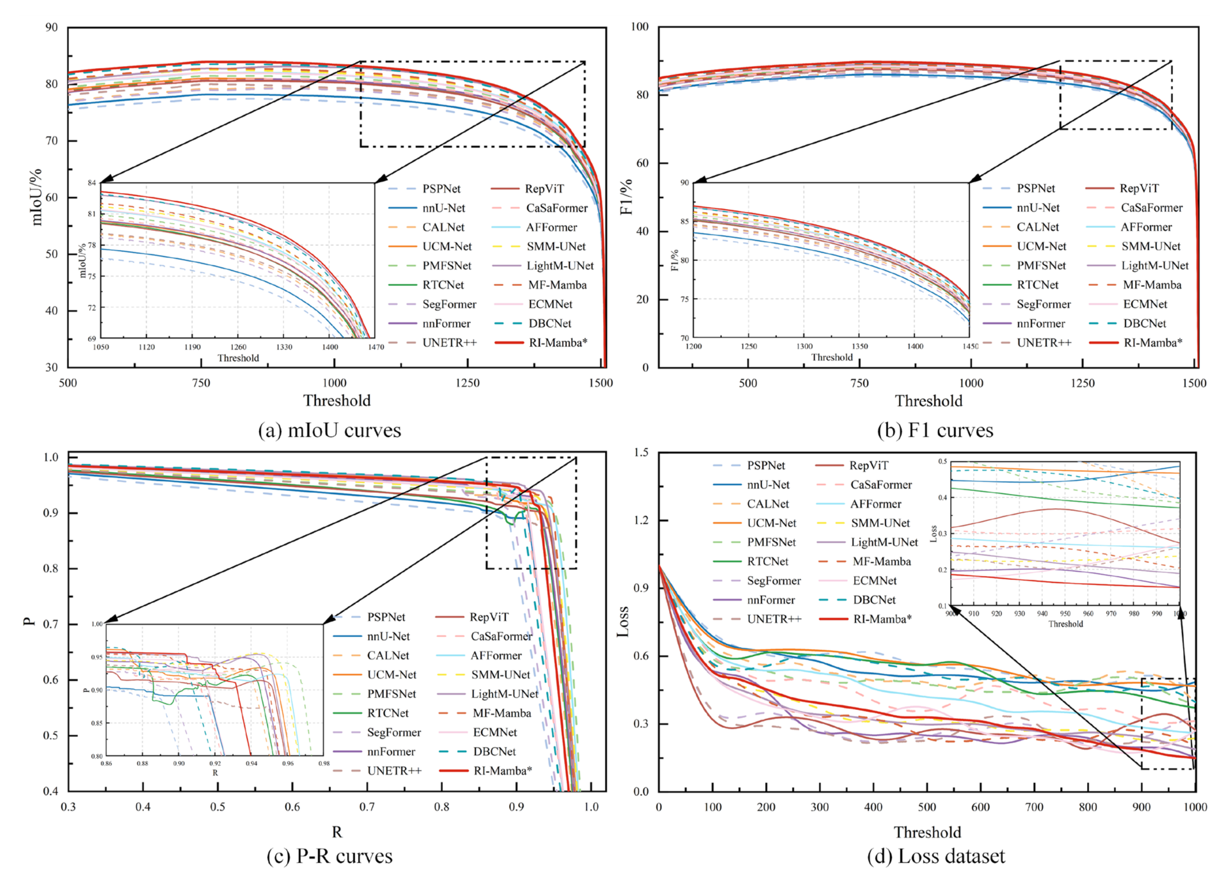
Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this paper.

Appendix A.
Learning curves of various indicators on the IS2020 and SISD datasets

Figure A1

mIoU curve, F1 curve, Pre-Rec curve and Loss curve of RI-Mamba with other lightweight methods on the IS2020 dataset.

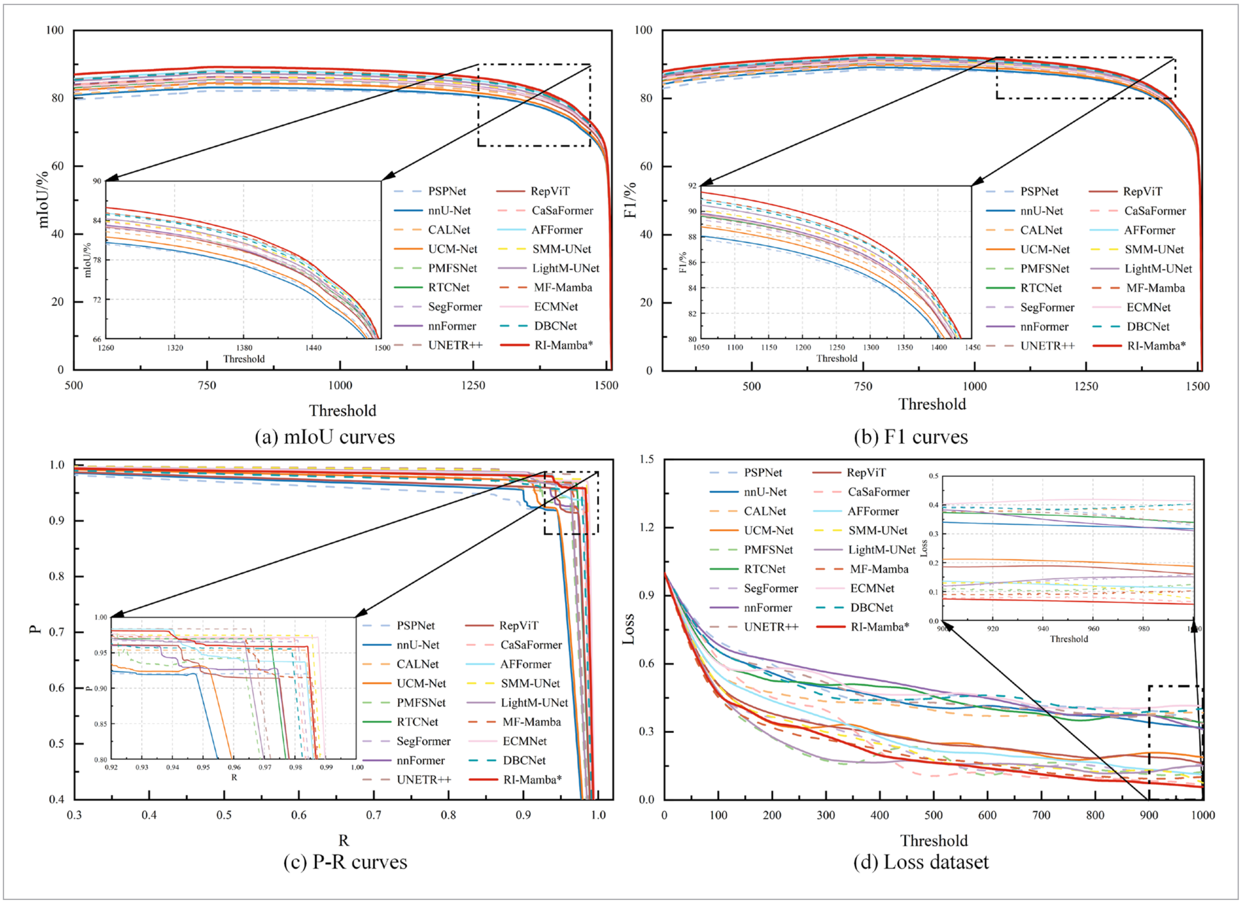


(a) mIoU curves (b) F1 curves

(c) P-R curves (d) Loss dataset

Figure A2

mIoU curve, F1 curve, Pre-Rec curve and Loss curve of RI-Mamba with other lightweight methods on the SISD dataset.



References

1. Chen, J., Fu, Z., Cheng, X. A Method for Power Lines Insulator Defect Detection with Attention Feedback and Double Spatial Pyramid. Electric Power Systems Research, 2023, 218, 109175. https://doi.org/10.1016/j.epsr.2023.109175
2. Chen, J., Zhang, H., Gong, M. Collaborative Compensative Transformer Network for Salient Object Detection. Pattern Recognition, 2024, 154, 110600. https://doi.org/10.1016/j.patcog.2024.110600
3. Cheng, Y., Liu, D. AdIn-DETR: Adapting Detection Transformer for End-to-End Real-Time Power Line Insulator Defect Detection. IEEE Transactions on Instrumentation and Measurement, 2024. https://doi.org/10.1109/TIM.2024.3420265
4. Dong, B., Wang, P., Wang, F. Head-Free Lightweight Semantic Segmentation with Linear Transformer. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(1), 516-524. https://doi.org/10.1609/aaai.v37i1.25126
5. Du, F., Wu, S. ECMNet: Lightweight Semantic Segmentation with Efficient CNN-Mamba Network. arXiv Preprint arXiv:2506.08629, 2025. https://doi.org/10.1177/00368504261419245
6. Fu, Y., Li, W., Fan, S. CAL-Net: Conditional Attention Lightweight Network for In-Orbit Landslide Detection. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61, 1-15. https://doi.org/10.1109/TGRS.2023.3321716

7. Gu, A., Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv Preprint arXiv:2312.00752, 2023. https://doi.org/10.48550/arXiv.2312.00752

8. Guo, J., Li, T., Du, B. Segmentation Head Networks with Harnessing Self-Attention and Transformer for Insulator Surface Defect Detection. Applied Sciences, 2023, 13(16), 9109. https://doi.org/10.3390/app13169109

9. Gupta, S., Tripathi, A. K. Flora-NET: Integrating Dual Coordinate Attention with Adaptive Kernel Based Convolution Network for Medicinal Flower Identification. Computers and Electronics in Agriculture, 2025, 230, 109834. https://doi.org/10.1016/j.compag.2024.109834

10. Han, G., Zhang, M., Wu, W. Improved U-Net Based Insulator Image Segmentation Method Based on Attention Mechanism. Energy Reports, 2021, 7, 210-217. https://doi.org/10.1016/j.egyr.2021.10.037

11. Isensee, F., Petersen, J., Klein, A. nnU-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation. arXiv Preprint arXiv:1809.10486, 2018. https://doi.org/10.1007/978-3-658-25326-4_7

12. Lee, Y. J., Kim, Y., Gim, H. Nanoelectronics Using Metal-Insulator Transition. Advanced Materials, 2024, 36(5), 2305353. https://doi.org/10.1002/adma.202305353

13. Li, G., Huang, Q., Wang, W. Selective and Multi-Scale Fusion Mamba for Medical Image Segmentation. Expert Systems with Applications, 2025, 261, 125518. https://doi.org/10.1016/j.eswa.2024.125518

14. Li, J., Hu, Y., Huang, X. CaSaFormer: A Cross-and Self-Attention Based Lightweight Network for Large-Scale Building Semantic Segmentation. International Journal of Applied Earth Observation and Geoinformation, 2024, 130, 103942. https://doi.org/10.1016/j.jag.2024.103942

15. Liao, W., Zhu, Y., Wang, X. LightM-UNet: Mamba Assists in Lightweight U-Net for Medical Image Segmentation. arXiv Preprint arXiv:2403.05246, 2024. https://doi.org/10.48550/arXiv.2403.05246

16. Liu, B., Kang, J., Guan, H. RTCNet: A Novel Real-Time Triple Branch Network for Pavement Crack Semantic Segmentation. International Journal of Applied Earth Observation and Geoinformation, 2025, 136, 104347. https://doi.org/10.1016/j.jag.2024.104347

17. Liu, D., Zhang, T., Wei, Y. Speech Conv-Mamba: Selective Structured State Space Model with Temporal Dilated Convolution for Efficient Speech Separation. IEEE Signal Processing Letters, 2025. https://doi.org/10.1109/LSP.2025.3560237

18. Liu, K., Zheng, J., Sha, Y. Spontaneous Broken-Symmetry Insulator and Metals in Tetralayer Rhombohedral Graphene. Nature Nanotechnology, 2024, 19(2), 188-195. https://doi.org/10.1038/s41565-023-01558-1

19. Liu, R., Lehman, J., Molino, P. An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution. Advances in Neural Information Processing Systems, 2018, 31. https://doi.org/10.48550/arXiv.1807.03247

20. Liu, X. ISMU-Net: Insulator Image Segmentation with Mamba-Based U-Net. International Conference on Optics, Electronics, and Communication Engineering (OECE 2024). SPIE, 2024, 13395, 641-648. https://doi.org/10.1117/12.3049373

21. Ma, X., Zhang, X., Pun, M. O. RS3Mamba: Visual State Space Model for Remote Sensing Image Semantic Segmentation. IEEE Geoscience and Remote Sensing Letters, 2024, 21, 1-5. https://doi.org/10.1109/LGRS.2024.3414293

22. Ni, Y., Liu, J., Chi, W. CGGLNet: Semantic Segmentation Network for Remote Sensing Images Based on Category-Guided Global-Local Feature Interaction. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62, 1-17. https://doi.org/10.1109/TGRS.2024.3379398

23. Shaker, A., Maaz, M., Rasheed, H. UNETR++: Delving Into Efficient and Accurate 3D Medical Image Segmentation. IEEE Transactions on Medical Imaging, 2024, 43(9), 3377-3390. https://doi.org/10.1109/TMI.2024.3398728

24. Shen, Y., Yao, S., Qiang, Z. SD-Mamba: A Lightweight Synthetic-Decompression Network for Cross-Modal Flood Change Detection. International Journal of Applied Earth Observation and Geoinformation, 2025, 136, 104409. https://doi.org/10.1016/j.jag.2025.104409

25. Song, R., Sun, L., Gao, Y. Global-Local Feature Cross-Fusion Network for Ultrasonic Guided Wave-Based Damage Localization in Composite Structures. Sensors and Actuators A: Physical, 2023, 362, 114659. https://doi.org/10.1016/j.sna.2023.114659

26. Sun, S., Fu, C., Xu, S. GLFNet: Global-Local Fusion Network for the Segmentation in Ultrasound Images. Computers in Biology and Medicine, 2024, 171, 108103. https://doi.org/10.1016/j.compbiomed.2024.108103

27. Tian, Y., Ahmad, R. B., Abdullah, N. A. B. Accurate and Efficient Insulator Maintenance: A DETR Algorithm for Drone Imagery. PLOS ONE, 2025, 20(2), e0318225. https://doi.org/10.1371/journal.pone.0318225

28. Wan, Z., Zhang, P., Wang, Y. Sigma: Siamese Mamba Network for Multi-Modal Semantic Segmentation. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025, 1734-1744. https://doi.org/10.1109/WACV61041.2025.00176

29. Wang, C., Nie, J., Yin, P. Unknown Fault Detection of Rolling Bearings Guided by Global-Local Feature Coupling. Mechanical Systems and Signal Processing, 2024, 213, 111331. https://doi.org/10.1016/j.ymssp.2024.111331

30. Wang, F., Wang, J., Ren, S. Mamba-Reg: Vision Mamba Also Needs Registers. Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, 14944-14953. https://doi.org/10.1109/CVPR52734.2025.01392

31. Wang, L., Zhao, R., Feng, W. DCAMamba: Mamba-Based Rapid Response DC Arc Fault Detection. arXiv Preprint arXiv:2503.01264, 2025. https://doi.org/10.48550/arXiv.2503.01264

32. Wang, Q., Li, Y., Cui, S. Enhanced Recognition of Insulator Defects on Power Transmission Lines via Proposal-Based Detection Model with Integrated Improvement Methods. Engineering Applications of Artificial Intelligence, 2024, 136, 109078. https://doi.org/10.1016/j.engappai.2024.109078

33. Wang, A., Chen, H., Lin, Z. RepViT: Revisiting Mobile CNN From ViT Perspective. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, 15909-15920. https://doi.org/10.1109/CVPR52733.2024.01506

34. Wei, S., Xia, X., Bi, S. Metal-Insulator-Semiconductor Photoelectrodes for Enhanced Photoelectrochemical Water Splitting. Chemical Society Reviews, 2024, 53(13), 6860-6916. https://doi.org/10.1039/D3CS00820G

35. Wu, J., Deng, Q., Xian, R. An Instance Segmentation Method for Insulator Defects Based on an Attention Mechanism and Feature Fusion Network. Applied Sciences, 2024, 14(9), 3623. https://doi.org/10.3390/app14093623

36. Xia, H., Yang, B., Li, Y. An Improved CenterNet Model for Insulator Defect Detection Using Aerial Imagery. Sensors, 2022, 22(8), 2850. https://doi.org/10.3390/s22082850

37. Xia, P., Peng, L., Di, X. S^{3}Mamba: Arbitrary-Scale Super-Resolution via Scaleable State Space Model. arXiv Preprint arXiv:2411.11906, 2024. https://doi.org/10.48550/arXiv.2411.11906

38. Xiao, P., Dong, Y., Zhao, J. MF-Mamba: Multi-Scale Convolution and Mamba Fusion Model for Semantic Segmentation of Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing, 2025. https://doi.org/10.1109/TGRS.2025.3593410

39. Xie, E., Wang, W., Yu, Z. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Advances in Neural Information Processing Systems, 2021, 34, 12077-12090.

40. Xu, J., Liao, H., Li, K. Multi-Scale Feature Fusion Transformer with Hybrid Attention for Insulator Defect Detection. IEEE Transactions on Instrumentation and Measurement, 2025. https://doi.org/10.1109/TIM.2025.3568984

41. Yang, Y., Yang, S., Li, C. Insulator Defect Detection Under Extreme Weather Based on Synthetic Weather Algorithm and Improved YOLOv7. High Voltage, 2025, 10(1), 69-77. https://doi.org/10.1049/hve2.12513

42. You, X., Zhao, X. An Insulator Defect Detection Network Based on Improved YOLOv7 for UAV Aerial Images. Measurement, 2025, 253, 117410. https://doi.org/10.1016/j.measurement.2025.117410

43. Yu, Y., Cao, H., Wang, Z. Texture-and-Shape Based Active Contour Model for Insulator Segmentation. IEEE Access, 2019, 7, 78706-78714. https://doi.org/10.1109/ACCESS.2019.2922257

44. Yuan, C., Zhao, D., Agaian, S. S. UCM-Net: A Lightweight and Efficient Solution for Skin Lesion Segmentation Using MLP and CNN. Biomedical Signal Processing and Control, 2024, 96, 106573. https://doi.org/10.1016/j.bspc.2024.106573

45. Zhang, H., Chen, K., Liu, C. CDMamba: Incorporating Local Clues Into Mamba for Remote Sensing Image Binary Change Detection. IEEE Transactions on Geoscience and Remote Sensing, 2025. https://doi.org/10.1109/TGRS.2025.3545012

46. Zhang, J., Chen, R., Liu, F. DC-Mamba: A Novel Network for Enhanced Remote Sensing Change Detection in Difficult Cases. Remote Sensing, 2024, 16(22), 4186. https://doi.org/10.3390/rs16224186
47. Zhang, J., Li, D., Zeng, Z. Dual-Branch Crack Segmentation Network with Multi-Shape Kernel Based on Convolutional Neural Network and Mamba. Engineering Applications of Artificial Intelligence, 2025, 150, 110536. https://doi.org/10.1016/j.engappai.2025.110536
48. Zhang, K., Yang, L. Insulator Segmentation Algorithm Based on K-Means. 2019 Chinese Automation Congress (CAC). IEEE, 2019, 4747-4751. https://doi.org/10.1109/CAC48633.2019.8996273
49. Zhang, Q., Zhang, J., Li, Y. IL-YOLO: An Efficient Detection Algorithm for Insulator Defects in Complex Backgrounds of Transmission Lines. IEEE Access, 2024, 12, 14532-14546. https://doi.org/10.1109/ACCESS.2024.3358205
50. Zhang, Y., Xie, M., Kong, Z. An Adaptive Global-Local Interactive Non-Local Boosting Network for Mixed Noise Removal. Expert Systems with Applications, 2025, 268, 126300. https://doi.org/10.1016/j.eswa.2024.126300
51. Zhao, Z., Bai, H., Zhang, J. Equivariant Multi-Modality Image Fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, 25912-25921. https://doi.org/10.1109/CVPR52733.2024.02448
52. Zhao, H., Shi, J., Qi, X. Pyramid Scene Parsing Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 2881-2890. https://doi.org/10.1109/CVPR.2017.660
53. Zhao, Y., Wu, J., Chen, W. A Small Object Real-Time Detection Method for Power Line Inspection in Low-Illuminance Environments. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024. https://doi.org/10.1109/TETCI.2024.3378651
54. Zhong, J., Tian, W., Xie, Y. PMFSNet: Polarized Multi-Scale Feature Self-Attention Network for Lightweight Medical Image Segmentation. Computer Methods and Programs in Biomedicine, 2025, 261, 108611. https://doi.org/10.1016/j.cmpb.2025.108611
55. Zhou, H. Y., Guo, J., Zhang, Y. nnFormer: Interleaved Transformer for Volumetric Segmentation. arXiv Preprint arXiv:2109.03201, 2021. https://doi.org/10.48550/arXiv.2109.03201
56. Zhou, T., Liu, F., Ye, X. CCGL-YOLOV5: A Cross-Modal Cross-Scale Global-Local Attention YOLOV5 Lung Tumor Detection Model. Computers in Biology and Medicine, 2023, 165, 107387. https://doi.org/10.1016/j.compbiomed.2023.107387
57. Zhu, J., Ren, Y., Zhou, W. LLM-MambaFormer: Integrating Mamba and Transformer for Crystalline Solids Properties Prediction. Materials Today Communications, 2025, 44, 112029. https://doi.org/10.1016/j.mtcomm.2025.112029