

Article

MCFI-Net: Multi-Scale Cross-Layer Feature Interaction Network for Landslide Segmentation in Remote Sensing Imagery

Jianping Liao ^{1,*} and Lihua Ye ² ¹ Department of Information Engineering, Quzhou College of Technology, Quzhou 324000, China² School of Computer Science, Jiaxing University, Jiaxing 314001, China; ylh9604@zjxu.edu.cn

* Correspondence: liaojianping@qzct.edu.cn

Abstract

Accurate and reliable detection of landslides plays a crucial role in disaster prevention and mitigation efforts. However, due to unfavorable environmental conditions, uneven surface structures, and other disturbances similar to those of landslides, traditional methods often fail to achieve the desired results. To address these challenges, this study introduces a novel multi-scale cross-layer feature interaction network, specifically designed for landslide segmentation in remote sensing images. In the MCFI-Net framework, we adopt the encoder–decoder as the foundational architecture, and integrate cross-layer feature information to capture fine-grained local textures and broader contextual patterns. Then, we introduce the receptive field block (RFB) into the skip connections to effectively aggregate multi-scale contextual information. Additionally, we design the multi-branch dynamic convolution block (MDCB), which possesses both dynamic perception ability and multi-scale feature representation capability. The comprehensive evaluation conducted on both the Landslide4Sense and Bijie datasets demonstrates the superior performance of MCFI-Net in landslide segmentation tasks. Specifically, on the Landslide4Sense dataset, MCFI-Net achieved a Dice score of 0.7254, a Matthews correlation coefficient (Mcc) of 0.7138, and a Jaccard score of 0.5699. Similarly, on the Bijie dataset, MCFI-Net maintained high accuracy with a Dice score of 0.8201, an Mcc of 0.8004, and a Jaccard score of 0.6951. Furthermore, when evaluated on the optical remote sensing dataset EORSSD, MCFI-Net obtained a Dice score of 0.7770, an Mcc of 0.7732, and a Jaccard score of 0.6571. Finally, ablation experiments carried out on the Landslide4Sense dataset further validated the effectiveness of each proposed module. These results affirm MCFI-Net’s capability in accurately identifying landslide regions from complex remote sensing imagery, and it provides great potential for the analysis of geological disasters in the real world.

Keywords: remote sensing; multi-scale cross-layer; receptive field block; multi-branch dynamic convolution block; landslide segmentation



Academic Editor: Massimiliano Pieraccini

Received: 9 September 2025

Revised: 20 October 2025

Accepted: 21 October 2025

Published: 27 October 2025

Citation: Liao, J.; Ye, L. MCFI-Net: Multi-Scale Cross-Layer Feature Interaction Network for Landslide Segmentation in Remote Sensing Imagery. *Electronics* **2025**, *14*, 4190. <https://doi.org/10.3390/electronics14214190>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides represent a significant category of natural disasters, characterized by the downward and outward movement of slope-forming materials such as rock, soil, and debris. These events typically occur when the structural integrity of a slope is compromised due to natural or human-caused factors. In recent decades, due to the acceleration of climate change and the increase in extreme weather phenomena, the frequency and severity of landslides have significantly increased. This trend has led to a wider geographical distribution and larger scale of landslides, which can cause a large number of casualties,

economic disruptions and infrastructure damage. Consequently, rapid and accurate identification of areas prone to landslides is of vital importance for emergency response and post-disaster recovery.

In the early stage of landslide research, the exploration and mapping work mainly relied on traditional field investigations. These methods required disaster assessment personnel to conduct on-site investigations and manually record the key features of landslides, such as their location, extent, form and related topographic features. However, it has been demonstrated to be labor-intensive, time-consuming, and usually entails logistical challenges, especially in remote areas with rugged terrain, dense vegetation, or limited accessibility. With the advancement of remote sensing technology, significant progress has been made in the ability to monitor and analyze geohazards over broad geographic areas. Due to its characteristics such as strong macroscopic observation ability, short observation cycle and low sensitivity to ground factors, remote sensing technology can well replace manual observation. Meanwhile, high-resolution remote sensing images have richer geometric details, texture features and complex contextual relationships, which are very suitable for the interpretation of newly formed landslides. However, for large-scale applications, relying solely on manual visual interpretation of these images remains inefficient and impractical. For the past few years, the advent of convolutional neural networks (CNNs) has significantly advanced the field of remote sensing by enabling the extraction of intricate spatial patterns and contextual information from high-resolution imagery. By leveraging these capabilities, CNNs not only enhance the efficiency and accuracy of disaster management work, but also provide valuable insights for early warning, urban planning and post-disaster recovery.

Nowadays, a growing number of researchers have devoted considerable attention to the task of landslide segmentation [1–5] in remote sensing imagery. Among them, Liu et al. [6] proposed to integrate a spatial detail enhancement module into the skip connections of their network architecture to address the challenge of recovering the fine-grained spatial information that was lost during the encoding process. Yu et al. [7] presented introducing a multi-scale convolutional attention mechanism in the encoder stage, which can capture global context cues and fine-grained local details. Şener et al. [8] integrated an efficient hybrid attention convolution mechanism into the architecture, which can adaptively prioritize significant regions and key features while suppressing redundant or irrelevant information in the input data. Jin et al. [9] introduced the dilated and efficient multi-scale attention block, which was strategically embedded into the network to enhance its perceptual capabilities. Zhou et al. [10] developed an encoder–decoder architecture to capture hierarchical feature representations and enhance the discriminative ability of the network. Wang et al. [11] employed a multi-scale feature fusion strategy to strengthen the semantic representation of low-level features within the network. Ling et al. [12] introduced a selective kernel attention module aimed at refining the integration of multi-resolution features within the network.

Recently, an increasing number of innovative technologies have been integrated into deep learning frameworks, including attention mechanisms [13,14], transformer-based modules [15,16], multi-scale feature extraction methods [17,18] and dense connection networks [19,20]. Among them, Chen et al. [21] proposed a strategy to effectively capture multi-scale contextual information by employing dilated convolutions to flexibly adjust the receptive field of encoder features. In addition, they incorporated depthwise separable convolutions into both the atrous spatial pyramid pooling and decoder modules, which maintained the segmentation accuracy and improved the boundary refinement effect. Xie et al. [22] proposed a novel hierarchical Transformer encoder architecture that progressively aggregates multi-level features through a structured representation learning process.

Zhang et al. [23] introduced the cross-perception mamba module aimed at enhancing the model's capacity to capture spatial dependencies from multiple orientations. Yang et al. [24] combined attention with a pyramid structure, which enhanced the model's ability to detect features at various scales. Zheng et al. [25] developed the directional convolutional edge branches to enhance the model's sensitivity to boundary structures. Zhou et al. [26] integrated the boundary extraction module, the boundary guidance module and the boundary supervision module to form a comprehensive boundary-aware segmentation strategy, which enhances the network's ability to capture fine boundary details. Li et al. [27] proposed a synergistic attention module that effectively balances the learning of global and local feature dependencies. Li et al. [28] developed a dual-branch encoder that integrates a local–global interaction module to effectively capture both detailed spatial information and high-level semantic context.

In this study, we propose a novel multi-scale cross-layer feature interaction network, specifically designed for landslide segmentation in high-resolution remote sensing images. Different from the traditional segmentation methods, the proposed MCFI-Net is built upon an encoder–decoder backbone, enriched with advanced multi-scale and cross-layer feature extraction strategies to improve both segmentation precision and robustness. The primary contributions of this paper are outlined as follows:

- (1) By constructing a structure with tightly connected layers through cross-layer feature interaction, we can more effectively integrate hierarchical information. This design promotes the dissemination of low-level spatial details and high-level semantic representations, enhancing the model's ability to understand the complex structure of scenes.
- (2) The receptive field block is embedded into the skip connection path to enhance the preservation and transfer of spatially informative features from the encoder to the decoder. This module enriches the context flow in different network stages and makes the depiction of the target area clearer.
- (3) The multi-branch dynamic convolution block is introduced, equipped with dynamic perception capabilities and powerful multi-scale feature extraction. This adaptive module is designed to selectively respond to varying spatial patterns and object scales by dynamically adjusting convolutional kernels.

2. Materials and Methods

2.1. Overall Architecture of MCFI-Net

As illustrated in Figure 1, MCFI-Net is strategically built upon an enhanced encoder–decoder framework. The encoder component employs a series of specialized Encoder blocks while the decoder incorporates Decoder blocks that are symmetrically aligned with the encoder. Specifically, during the encoding phase, each Encoder block incorporates a multi-branch dynamic convolutional module, which can enhance the network's ability to perceive and adapt to different spatial patterns. At each hierarchical stage, the receptive field block receives the feature maps generated by the corresponding Encoder block and enhances the network's ability to perceive spatial structures across varying scales. During the decoding stage, feature reconstruction is achieved through a combination of transposed convolution layers and double convolution blocks. To further optimize the integration of semantic and spatial information across different layers, the feature maps produced by the receptive field block and the corresponding Decoder block at each layer are weighted and combined with the decoded features from the next layer. In the final stage of the decoding process, the output feature map generated by the highest-layer Decoder block passes through a Sigmoid function, which obtains the possibility that the pixels belong to the target category. Compared with the conventional UNet++ architecture [29], MCFI-Net

adopts a more explicit and efficient multi-level cross-feature interaction strategy rather than relying on densely nested skip connections. In UNet++, feature propagation mainly depends on cascaded convolutional nodes, which may lead to redundant information flow and increased computational burden. In contrast, MCFI-Net embeds receptive field blocks into the encoder–decoder pathways, allowing the network to capture richer contextual dependencies and enlarge the effective receptive field. Moreover, the cross-layer feature interaction in MCFI-Net facilitates direct semantic exchange between deep and shallow layers, which significantly strengthens boundary localization and structural consistency. By integrating multi-scale contextual cues with spatially refined decoding, MCFI-Net achieves more discriminative feature representation and exhibits superior segmentation performance, particularly in complex background scenarios where traditional UNet++ structures tend to suffer from feature attenuation and blurred edges.

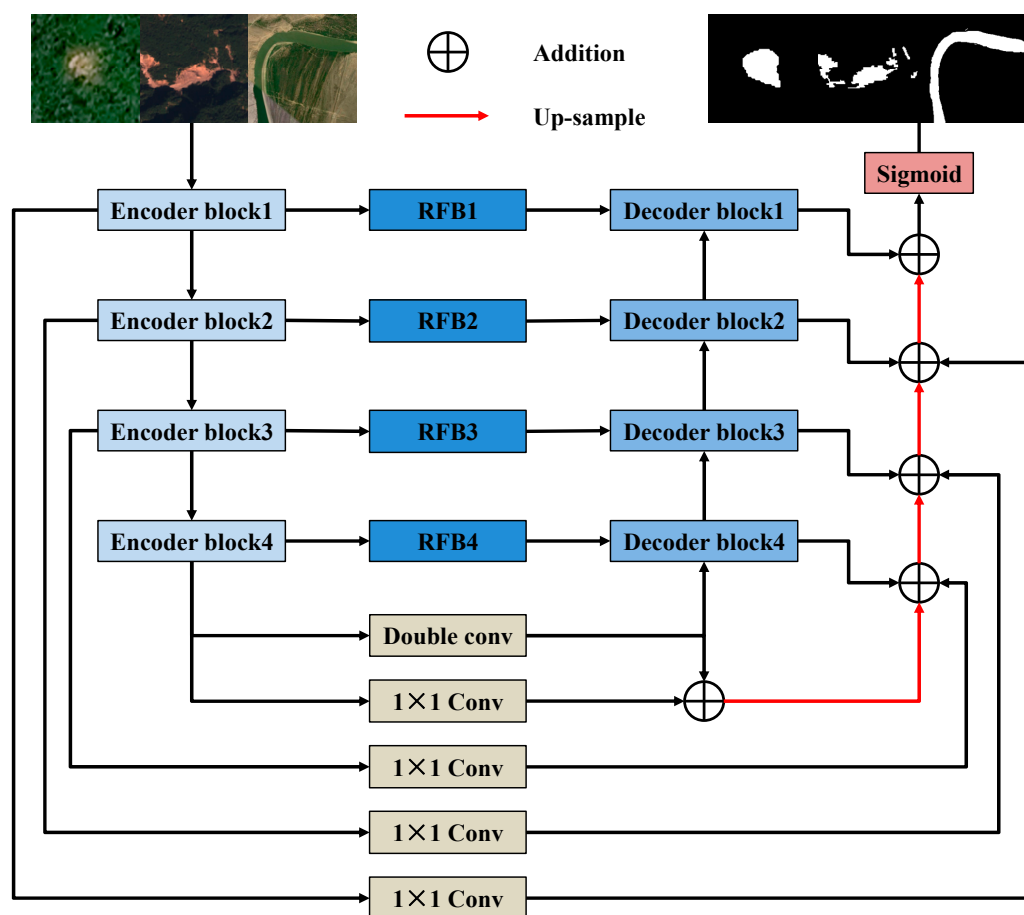


Figure 1. Overall architecture of MCFI-Net.

2.2. Encoder Block

Traditional encoders usually rely on two sequential 3×3 convolutional layers to extract image features. Although this approach is effective for capturing fine-grained local patterns, it has obvious limitations when applied to complex remote sensing images. To address these shortcomings, the Encoder block in MCFI-Net introduces a more advanced design, as illustrated in Figure 2. Specifically, instead of using fixed convolutional filters, the architecture incorporates a multi-branch dynamic convolutional block. This block dynamically adjusts its receptive field and convolutional kernels based on input content, enabling it to adaptively perceive spatial features at multiple scales. Following the dynamic convolutional operation, a 1×1 convolution is applied to compress channel dimensions and facilitate more efficient feature integration. Finally, the MaxPooling operation is

performed to further down-sample the feature map, which allows deeper layers to capture more abstract semantic information.

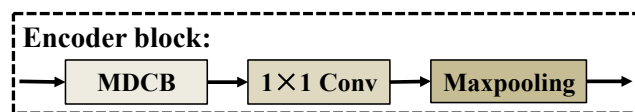


Figure 2. The structure of the Encoder block.

As illustrated in Figure 3, the multi-branch dynamic convolution block is specifically designed to improve the model's feature representation capabilities by capturing spatial information at multiple scales and dynamically adapting to the input features. Specifically, the MDCB comprises two parallel processing branches, each tailored to extract different scales of spatial information. In the first branch, the input undergoes two consecutive 3×3 dynamic convolutional layers, which adaptively adjust their kernels based on the content of the input feature map. The second branch applies two 5×5 standard convolutional layers, which are capable of capturing a broader spatial context and are particularly effective for modeling large-scale structures or more dispersed patterns. After the independent feature transformations in each branch, their outputs are fused via element-wise addition, allowing the block to merge the dynamically modulated local features with the global context information. This combination enhances the model's representational capacity and improves its robustness in handling complex and diverse visual scenes. Compared with the convolutional blocks used in U-Net [30], the proposed multi-branch dynamic convolution block introduces two major advantages. Firstly, the dynamic convolution mechanism enables the kernels to be adaptively generated according to the input features, allowing the network to flexibly adjust its receptive patterns and better respond to local variations and heterogeneous textures. Secondly, the multi-branch design integrates both fine-grained dynamic feature extraction and large-scale contextual perception. Consequently, MDCB provides a richer and more discriminative feature representation, especially in regions with irregular boundaries or complex background interference.

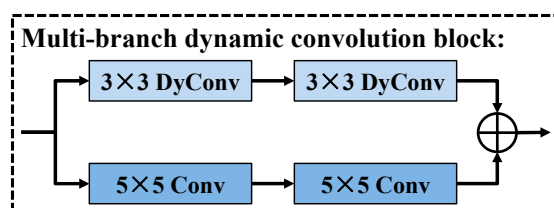


Figure 3. The structure of a multi-branch dynamic convolution block.

As illustrated in Figure 4, the process of the dynamic convolution block [31] begins with a global average pooling operation, which compresses the spatial dimensions of the input feature map to capture holistic, image-wide contextual information. The resulting global descriptor is then passed through two fully connected (FC) layers with an intervening ReLU activation, effectively projecting the compressed feature into a latent space that corresponds to the number of candidate convolution kernels K . Subsequently, the output of the second FC layer undergoes a Softmax normalization, producing a set of attention weights $\pi_1, \pi_2, \dots, \pi_k$, which are non-negative and sum to one. Followed this, these weights serve as dynamic coefficients for a weighted summation of the K pre-defined convolution kernels $\text{Conv}_1, \text{Conv}_2, \dots, \text{Conv}_k$. Furthermore, the adaptively generated kernel is applied to the input features through a standard convolution operation, followed by batch normalization (BN) and a non-linear activation function. This dynamic adjustment ensures that the network is capable of selecting and combining convolutional patterns most

relevant to the input, thereby enhancing its flexibility and feature representation capacity across diverse image contexts.

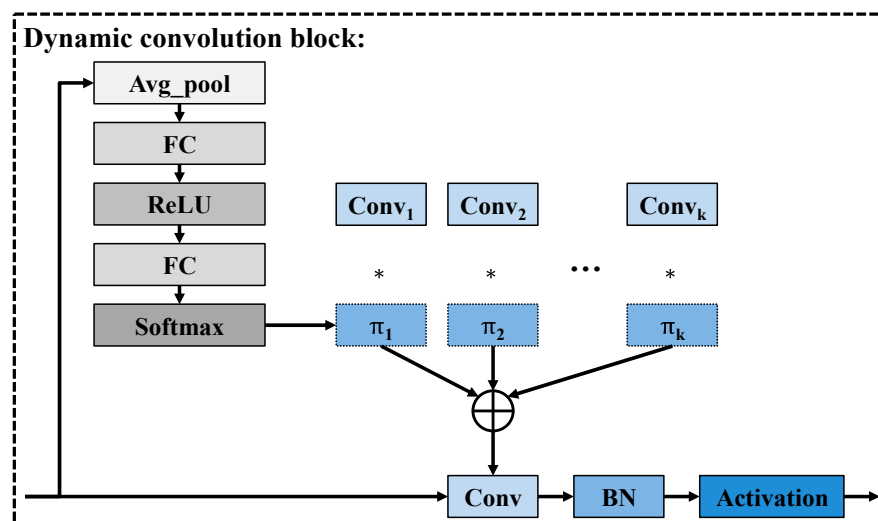


Figure 4. The structure of the dynamic convolution block.

2.3. Decoder Block

As illustrated in Figure 5, the input to each Decoder block is composed of two critical components. The first input originates from the output of the previous Decoder block (i.e., Decoder block-($n - 1$)), which has been processed through a transposed convolution (ConvTranspose) layer to up-sample the feature maps. The second input comes from the corresponding RFB- n located in the encoder pathway, which carries rich, high-level semantic information. These two feature maps are combined through element-wise addition. This fusion strategy enables the network to effectively integrate multi-scale information, enhancing its ability to accurately reconstruct fine-grained structural details. After fusion, the resulting feature map is passed through a Double Convolutional block. This block consists of two sequential 3×3 convolutional layers, where each convolution operation is immediately followed by Batch Normalization (BN) to stabilize the learning process and a ReLU activation function to introduce non-linearity. The design of this Decoder block not only enhances the representational capacity of the network but also accelerates the convergence speed of training and significantly improves the overall stability of training.

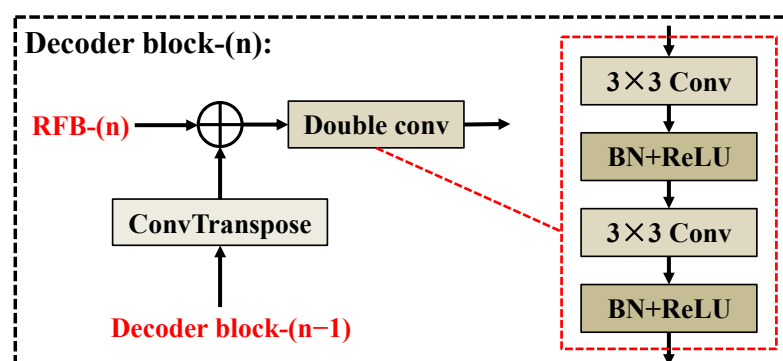


Figure 5. The structure of the Decoder block.

2.4. Receptive Field Block

As illustrated in Figure 6, the receptive field block is architecturally composed of four parallel branches, each designed to extract and integrate features at multiple scales. The

first branch follows a straightforward pathway, applying a 1×1 convolution directly to the input feature map. In the second branch, the input first undergoes a 1×1 convolution to reduce its dimensionality. It then passes through sequential 1×3 and 3×1 convolutions, which independently capture spatial dependencies along the horizontal and vertical directions. Following these asymmetric convolutions, a 3×3 dilated convolution with a dilation rate of 3 is applied, significantly expanding the receptive field and allowing the branch to integrate broader contextual information without a proportional increase in parameters. The third and fourth branches mirror the structure of the second, but with larger convolution kernels and increased dilation rates. Specifically, the third branch utilizes 1×5 and 5×1 convolutions, followed by a 3×3 dilated convolution with a dilation rate of 5, while the fourth branch employs 1×7 and 7×1 convolutions with a subsequent dilation rate of 7. This progressive expansion across branches enables the RFB to capture context from various spatial scales, ranging from local fine details to more global structures. After independent processing, the outputs from all four branches are concatenated along the channel dimension. Subsequently, the concatenated feature map passes through a standard 3×3 convolution layer, which serves both to integrate the multi-scale features and to refine the feature dimensions. Finally, the refined output is element-wise added to the original input feature map via a residual connection. Unlike ASPP in DeepLab [32] and the Inception network [33], the RFB enhances spatial adaptability by introducing dilated convolutions with increasing rates (3, 5, 7) and directionally asymmetric convolutions ($1 \times n$ and $n \times 1$) before the expansion convolutions. In essence, the proposed RFB achieves a more flexible and denser receptive field than ASPP, and also has a more adaptable and more efficient multi-scale fusion mechanism than Inception.

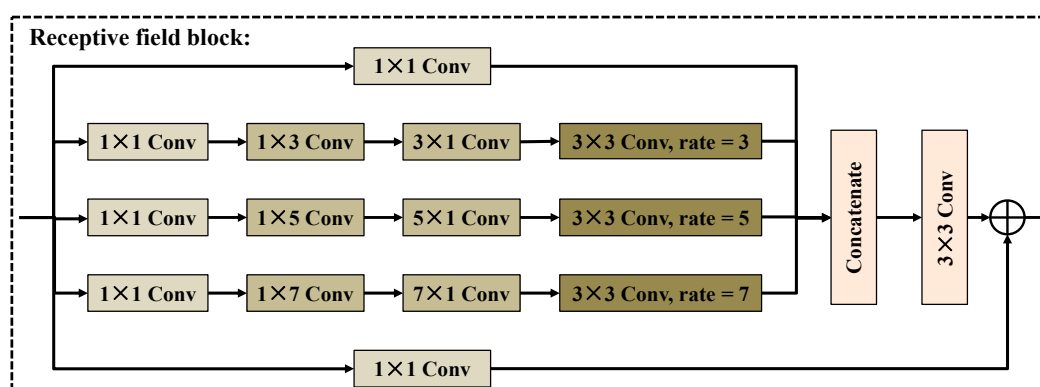


Figure 6. The structure of the receptive field block.

3. Results

In this section, we begin by detailing the experimental setup, including the configuration of the dataset preparation, training environment, implementation parameters and evaluation metrics. Subsequently, the MCFI-Net was tested on the three benchmark datasets: Landslide4Sense, Bijie and EORSSD. In addition, a series of ablation studies was conducted to systematically verify the contribution of each key module in the MCFI-Net.

3.1. Dataset

Landslide4Sense dataset [34]: The Landslide4Sense dataset is a dedicated resource curated by the International Institute of Artificial Intelligence, specifically designed to support research in landslide detection. This dataset encompasses a wide variety of landslide events, spanning diverse terrain types and geographic regions. Each image within the collection is meticulously annotated to ensure high precision in ground truth labeling. In total, the dataset includes 1980 high-quality images, which are systematically

divided into three subsets: 1385 images for training, 396 for validation, and 199 for testing. Representative sample images from the dataset are illustrated in Figure 7.

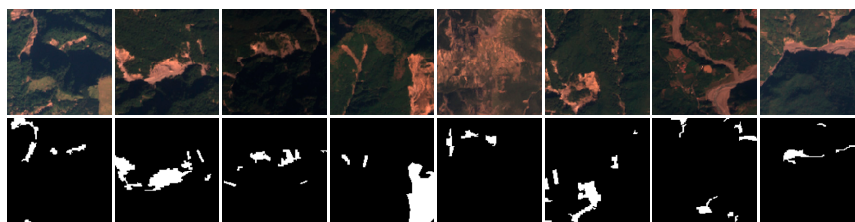


Figure 7. Landslide samples on the Landslide4Sense dataset.

Bijie dataset [35]: The Bijie dataset is composed of 770 carefully selected samples originating from Bijie County in Guizhou Province, China. The samples encompass a variety of landslide-related phenomena, including rock falls, rock slides, and debris flows, providing a comprehensive representation of the region's diverse hazard types. The dataset is partitioned into three subsets: 464 images for training, 153 for validation, and 153 for testing. Representative sample images from the dataset are illustrated in Figure 8.

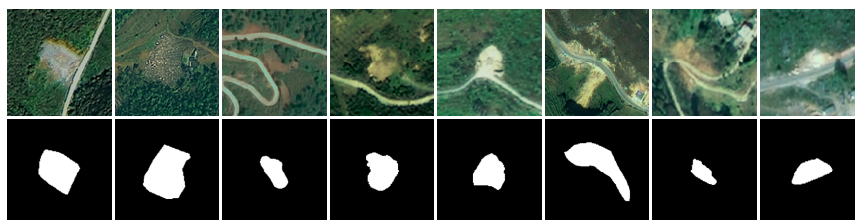


Figure 8. Landslide samples on the Bijie dataset.

EORSSD dataset [36]: The EORSSD dataset proposed by Zhang et al. is an enhanced and more comprehensive extension of the ORSSD dataset. It was meticulously compiled to address the increasingly complex and diverse requirements of remote sensing scenarios in the real world. Compared with the previous version, EORSSD integrates a broader geographical context, including densely populated urban areas, vast agricultural zones, complex coastal regions and rugged mountainous areas. The dataset contains 2000 high-resolution optical remote sensing images, divided into 1200 for training, 400 for validation, and 400 for testing. Representative samples of EORSSD are shown in Figure 9.

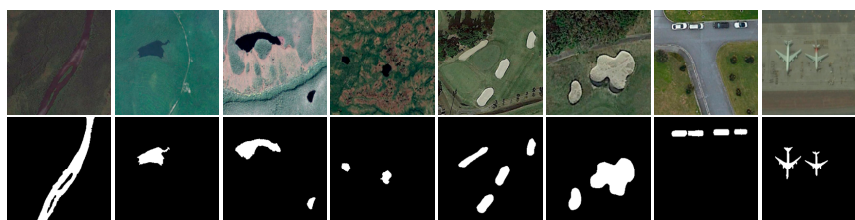


Figure 9. Representative samples on the EORSSD dataset.

3.2. Model Configuration and Evaluation Metrics

The experiment was implemented using the PyTorch 2.0 framework to ensure flexibility and efficiency during the training phases. All experiments were executed on a workstation equipped with an NVIDIA RTX A6000 GPU, which provided ample computational power for handling complex model architectures. During the training process, the Dice loss function [37] was employed to effectively mitigate class imbalance and enhance segmentation accuracy along object boundaries. In addition, our network was trained for

200 epochs with a batch size of 32 using the Adam optimizer, where the learning rate was initialized to 0.001, and the weight decay was set to 1×10^{-4} . The optimizer hyperparameters were set as $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-8}$. It is worth noting that no data augmentation techniques were applied during training to ensure that the evaluation purely reflected the intrinsic generalization ability of all models. Considering the computer configuration, all input images will be uniformly adjusted to 256×256 pixels.

To comprehensively evaluate the segmentation performance of different models, three common evaluation metrics were used: Dice [38], Mcc [39], and Jaccard [40]. The formal definitions for these metrics are provided below:

$$Dice = \frac{2TP}{2TP + FN + FP} \quad (1)$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (2)$$

$$Jaccard = \frac{TP}{TP + FN + FP} \quad (3)$$

3.3. Ablation Experiment

To thoroughly evaluate the contribution of each individual component within the MCFI-Net, Table 1 presents the quantitative results of a series of ablation experiments conducted on the Landslide4Sense dataset. As shown in the table, the U-Net is first considered, achieving a Dice score of 67.31%, an Mcc of 65.92%, and a Jaccard score of 50.96%. Building upon this, the introduction of our refined Baseline leads to a notable improvement, with Dice increasing to 70.91%, Mcc rising to 69.69%, and Jaccard improving to 55.03%. These results suggest that the modifications in the Baseline provide a more effective feature extraction and representation ability. When the RFB is integrated into the Baseline, further gains are observed: Dice reaches 71.94%, Mcc advances to 70.80%, and Jaccard increases to 56.26%. These enhancements demonstrate that RFB effectively enlarges the receptive field, allowing the model to capture richer contextual information at multiple scales. Similarly, when adding the MDCB to the Baseline, comparable improvements are achieved, with a Dice score of 71.80%, an Mcc of 70.81%, and a Jaccard score of 56.16%. This result highlights the contribution of MDCB in adaptively adjusting the convolutional kernels based on input features. Finally, the combination of both RFB and MDCB modules results in the best performance across all metrics: Dice improves to 72.54%, Mcc rises to 71.38%, and Jaccard reaches 56.99%. This performance boost verifies that the complementary strengths of the RFB and MDCB synergistically enhance the network's feature modeling capabilities. Overall, the ablation experiments clearly demonstrate that each component of MCFI-Net plays a critical role in progressively boosting the segmentation performance on the Landslide4Sense dataset.

Table 1. Ablation experiment on the Landslide4Sense dataset.

Method	Dice	Mcc	Jaccard
U-Net [30]	0.6731	0.6592	0.5096
Baseline	0.7091	0.6969	0.5503
Baseline + RFB	0.7194	0.7080	0.5626
Baseline + MDCB	0.7180	0.7081	0.5616
Baseline + RFB + MDCB (MCFI-Net)	0.7254	0.7138	0.5699

To further validate the effectiveness of the receptive field block, a series of ablation experiments was conducted on the Landslide4Sense dataset, as shown in Table 2. When the

RFB module was removed from the MCFI-Net, the overall segmentation accuracy noticeably declined, with the Dice decreasing from 72.54% to 69.80%, the Mcc dropping from 71.38% to 68.81%, and the Jaccard falling from 56.99% to 54.16%. This demonstrates that the RFB plays a critical role in enhancing multi-scale contextual perception and improving feature representation. Furthermore, when the RFB was replaced by the ASPP module from DeepLab or the Inception structure, the model exhibited only marginal improvements compared to the version without RFB. Specifically, the Dice scores of 70.58% and 70.61% indicate that while both alternatives enhance multi-scale feature extraction to some extent, they fail to achieve the same level of adaptive receptive field adjustment and feature fusion as the RFB. In contrast, the proposed RFB demonstrates a more flexible and fine-grained capability for integrating hierarchical contextual information.

Table 2. Ablation experiment of RFB on the Landslide4Sense dataset.

Method	Dice	Mcc	Jaccard
MCFI-Net with ASPP	0.7058	0.6946	0.5458
MCFI-Net with Inception	0.7061	0.6948	0.5475
MCFI-Net without RFB	0.6980	0.6881	0.5416
MCFI-Net with RFB	0.7254	0.7138	0.5699

To investigate the contribution of each branch and the impact of different convolutional in the multi-branch dynamic convolution block, a series of ablation experiments were conducted on the Landslide4Sense dataset, as presented in Table 3. When the first branch adopted dynamic convolution and the second branch used standard convolution, the MDCB achieved the best overall performance. This result indicates that combining dynamic and standard convolutions allows the MDCB to effectively balance adaptive receptive field learning and stable local feature extraction. In contrast, when both branches were replaced by dynamic convolutions, the performance declined notably, suggesting that excessive dynamic filtering might introduce redundant parameter updates and hinder feature stability. Similarly, using standard convolutions in both branches led to suboptimal results, as the lack of adaptive capacity limited the model's ability to capture multi-scale contextual information. Furthermore, when only one branch (whether it is the dynamic convolution branch or the standard convolution branch) is retained, the segmentation performance further deteriorates. Overall, these results demonstrate that the hybrid configuration of dynamic convolution in the first branch and standard convolution in the second branch yields the most effective balance between adaptability and stability.

Table 3. Ablation experiment of MDCB on the Landslide4Sense dataset.

The First Branch	The Second Branch	Dice	Mcc	Jaccard
Dynamic convolutional	Standard convolutional	0.7254	0.7138	0.5699
Dynamic convolutional	Dynamic convolutional	0.7098	0.6988	0.5518
Standard convolutional	Standard convolutional	0.7177	0.7053	0.5609
/	Standard convolutional	0.7126	0.6997	0.5545
Dynamic convolutional	/	0.7143	0.7021	0.5566

3.4. Results for Landslide4Sense Dataset

To assess the performance and capability of the proposed MCFI-Net for landslide segmentation, we conducted a series of comparative experiments on the Landslide4Sense dataset against several state-of-the-art segmentation models, and the quantitative results are summarized in Table 4. These competing methods include U-Net [30], SeaNet [41], HST-UNet [42], MsanlfNet [43], IMFF-Net [44], SwinT [45], DECENet [46], BRRNet [47],

EGENet [48], UNetFormer [49], and DCSwin [50]. Among the compared models, SeaNet and DCSwin exhibit the weakest performance, with Dice scores of 63.10% and 62.52%, Mcc values of 61.72% and 61.18%, and Jaccard scores of 46.48% and 46.13%. These results indicate that SeaNet and DCSwin have significant limitations in terms of segmentation accuracy and boundary positioning. Similarly, DECENet, EGENet and UNetFormer also show limited segmentation accuracy, with Dice values of 66.16%, 66.94%, and 66.46%, respectively, suggesting that their feature representation capacity is constrained when dealing with heterogeneous landslide regions. U-Net is a widely used benchmark architecture, with Dice, Mcc, and Jaccard scores of 67.31%, 65.92%, and 50.96%. HST-UNet and MsanlfNet show slightly better performance, with Dice scores of 68.84% and 68.64%, respectively. SwinT is a method based on transformers, which achieves a Dice score of 70.16% and a Jaccard score of 54.22%. More competitive results are observed with IMFF-Net and BRRNet, with Dice scores of 71.26% and 71.34%, and Jaccard scores both exceeding 55%. Notably, MCFI-Net achieves the best performance across all evaluated metrics, with a Dice score of 72.54%, an Mcc of 71.38%, and a Jaccard score of 56.99%. Based on the comparative analysis in Table 4, MCFI-Net consistently outperforms other state-of-the-art approaches, demonstrating its strong potential for reliable and precise landslide segmentation in remote sensing imagery.

Table 4. The quantitative results of MCFI-Net and other models on the Landslide4Sense dataset.

Method	Dice	Mcc	Jaccard
U-Net [30]	0.6731	0.6592	0.5096
SeaNet [41]	0.6310	0.6172	0.4648
HST-UNet [42]	0.6884	0.6757	0.5257
MsanlfNet [43]	0.6864	0.6735	0.5244
IMFF-Net [44]	0.7126	0.7012	0.5548
SwinT [45]	0.7016	0.6909	0.5422
DECENet [46]	0.6616	0.6484	0.4971
BRRNet [47]	0.7134	0.7010	0.5554
EGENet [48]	0.6694	0.6610	0.5094
UNetFormer [49]	0.6646	0.6527	0.5044
DCSwin [50]	0.6252	0.6118	0.4613
MCFI-Net	0.7254	0.7138	0.5699

Figure 10 presents the visualization results of various deep learning models on the Landslide4Sense dataset, showcasing their ability to segment landslide regions in complex remote sensing imagery. Each row corresponds to a different landslide sample, while the columns represent the input image, ground truth mask, and predictions from competing models. Among these approaches, U-Net demonstrates moderate performance in landslide segmentation. While it can detect the general location and shape of some landslide regions, the boundaries are often coarse and incomplete. SeaNet performs less favorably than U-Net. The segmented results appear sparse and less aligned with the ground truth, missing many fine-grained details and often failing to capture the full extent of landslide areas. HST-UNet shows a modest improvement over SeaNet and U-Net. It captures more landslide areas and preserves general shapes better, though it still suffers from irregular boundary delineation and occasional false positives in non-landslide regions. MsanlfNet shows better continuity along landslide edges and improved segmentation in more textured areas, though it still lacks precision in finely fragmented landslide zones. IMFF-Net exhibits a noticeable enhancement in segmentation accuracy. The predicted regions more closely align with the ground truth masks, especially in terms of shape fidelity and boundary sharpness. The results of SwinT sometimes appear to be slightly overly smooth, resulting

in the loss of details in complex edges or scattered areas. BRRNet yields competitive results, comparable to IMFF-Net. It performs robustly in diverse terrain settings and preserves landslide shape and size well. DECENet underperforms compared to most other models. It misses several landslide regions, especially smaller or more irregular ones, and its output appears fragmented. EGENet achieves moderate segmentation quality, correctly identifying large landslide regions but often missing subtle boundaries and smaller patches. Due to its hybrid convolution-transformer design, UNetFormer produces smoother and more coherent prediction results compared to EGENet, but it still fails to capture some complex boundary transitions. Although DCSwin is based on the converter architecture, it often results in overly smooth output and the loss of fine structural details. However, MCFI-Net delivers the most accurate and visually consistent results among all compared methods. It shows strong agreement with the ground truth masks, effectively capturing the full extent of landslide areas with high boundary precision and minimal noise.

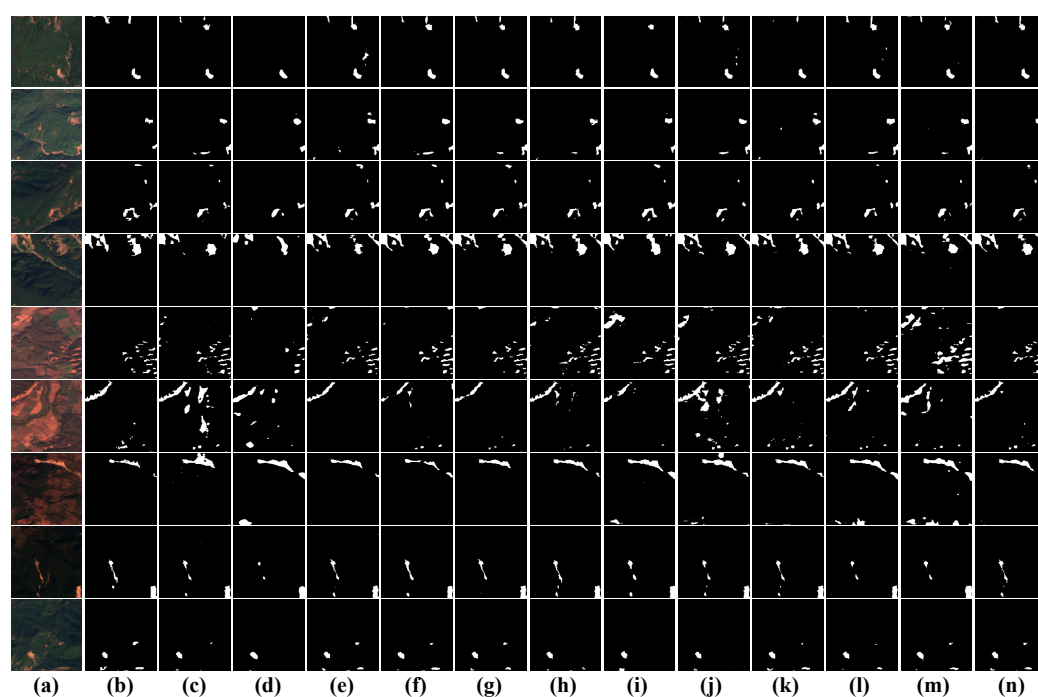


Figure 10. Visualization results of different models on the Landslide4Sense dataset. (a) Original images; (b) Mask images; (c) U-Net; (d) SeaNet; (e) HST-UNet; (f) MsanlfNet; (g) IMFF-Net; (h) SwinT; (i) DECENet; (j) BRRNet; (k) EGENet; (l) UNetFormer; (m) DCSwin, (n) MCFI-Net.

3.5. Results for Bijie Dataset

Table 5 presents the quantitative results of MCFI-Net and several state-of-the-art models on the Bijie dataset, and visual examples of the segmentation outcomes are illustrated in Figure 11. As shown in Table 5, MCFI-Net achieves the highest performance across all three evaluation metrics, with a Dice score of 82.01%, an Mcc of 80.04%, and a Jaccard score of 69.51%. These results underscore MCFI-Net's strong ability to accurately delineate landslide boundaries, maintain high spatial consistency with ground truth, and generalize well to complex geospatial features in the Bijie region. In comparison, DECENet also performs competitively, achieving a Dice score of 81.32%, an Mcc of 79.20%, and a Jaccard score of 68.53%. Although close in performance, it slightly lags behind MCFI-Net in each metric, suggesting its contextual encoding approach may not capture as comprehensive multi-scale spatial features as MCFI-Net. As illustrated in Figure 11, the segmentation results produced by MCFI-Net exhibit a high degree of consistency with the corresponding ground truth annotations. The predicted masks closely match the actual landslide regions in both shape

and extent, effectively capturing fine-grained boundaries and complex terrain features. These visual outcomes not only validate the model's robust segmentation capabilities but also reinforce the quantitative gains highlighted in Table 5.

Table 5. The quantitative results of MCFI-Net and other models on the Bijie dataset.

Method	Dice	Mcc	Jaccard
U-Net [30]	0.7599	0.7336	0.6130
SeaNet [41]	0.7962	0.7741	0.6616
HST-UNet [42]	0.7855	0.7620	0.6476
MsanlfNet [43]	0.8019	0.7796	0.6693
IMFF-Net [44]	0.8073	0.7858	0.6770
SwinT [45]	0.7896	0.7666	0.6525
DECENet [46]	0.8132	0.7920	0.6853
BRRNet [47]	0.7922	0.7691	0.6561
EGENet [48]	0.7755	0.7520	0.6348
UNetFormer [49]	0.7996	0.7789	0.6672
DCSwin [50]	0.7577	0.7330	0.6114
MCFI-Net	0.8201	0.8004	0.6951

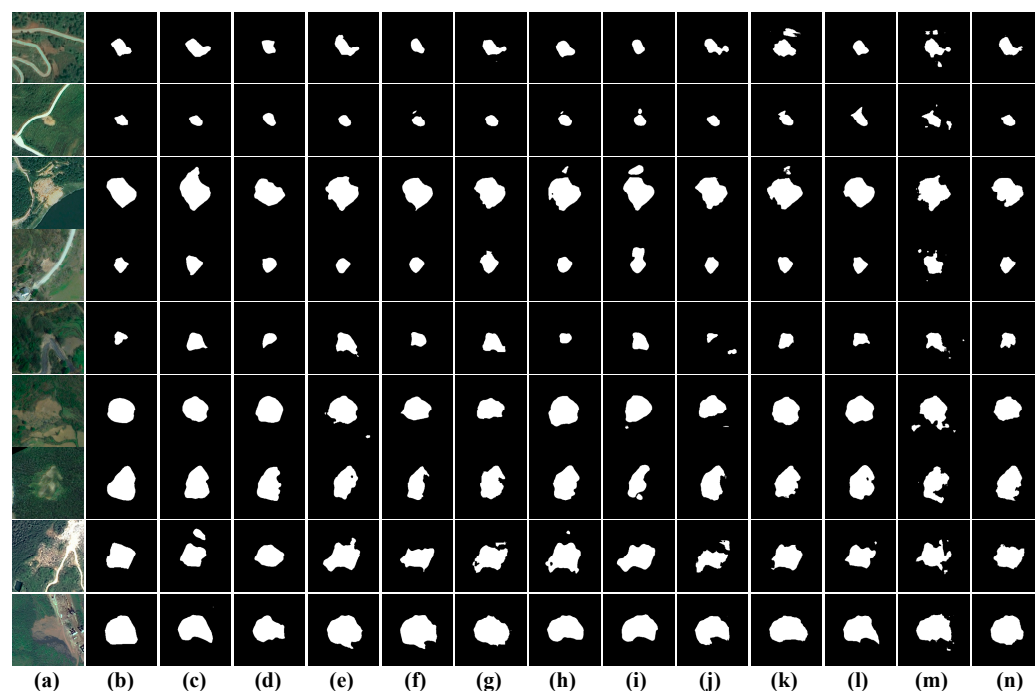


Figure 11. Visualization results of different models on the Bijie dataset. (a) Original images; (b) Mask images; (c) U-Net; (d) SeaNet; (e) HST-UNet; (f) MsanlfNet; (g) IMFF-Net; (h) SwinT; (i) DECENet; (j) BRRNet; (k) EGENet; (l) UNetFormer; (m) DCSwin, (n) MCFI-Net.

3.6. Results for EORSSD Dataset

Table 6 presents the quantitative comparison results of the aforementioned models on the EORSSD dataset, while Figure 12 provides the corresponding visual segmentation results. As shown in the table, MCFI-Net achieves the best performance among all methods, obtaining a Dice score of 77.70%, an Mcc of 77.32%, and a Jaccard score of 65.71%. These superior results demonstrate the model's strong capability in accurately identifying ship regions and preserving boundary integrity, even in scenarios with varying lighting, backgrounds, and scales. As illustrated in Figure 12, the segmentation masks generated by MCFI-Net exhibit excellent consistency with the ground truth. In contrast,

conventional convolution-based models such as U-Net and SeaNet tend to produce fragmented or incomplete predictions, while transformer-based models like SwinT and DCSwin suffer from missing details due to inadequate local spatial encoding. These qualitative results further validate the quantitative improvements reported in Table 6, confirming that MCFI-Net achieves superior accuracy, robustness, and generalization capability on the EORSSD dataset.

Table 6. The quantitative results of MCFI-Net and other models on the EORSSD dataset.

Method	Dice	Mcc	Jaccard
U-Net [30]	0.6433	0.6536	0.4813
SeaNet [41]	0.7067	0.7036	0.5588
HST-UNet [42]	0.7166	0.7109	0.5766
MsanlfNet [43]	0.7622	0.7572	0.6293
IMFF-Net [44]	0.7106	0.7026	0.5712
SwinT [45]	0.5773	0.5617	0.4155
DECENet [46]	0.7630	0.7589	0.6326
BRRNet [47]	0.7083	0.6979	0.5637
EGENet [48]	0.5896	0.5846	0.4375
UNetFormer [49]	0.7549	0.7525	0.6290
DCSwin [50]	0.6775	0.6703	0.5278
MCFI-Net	0.7770	0.7732	0.6571

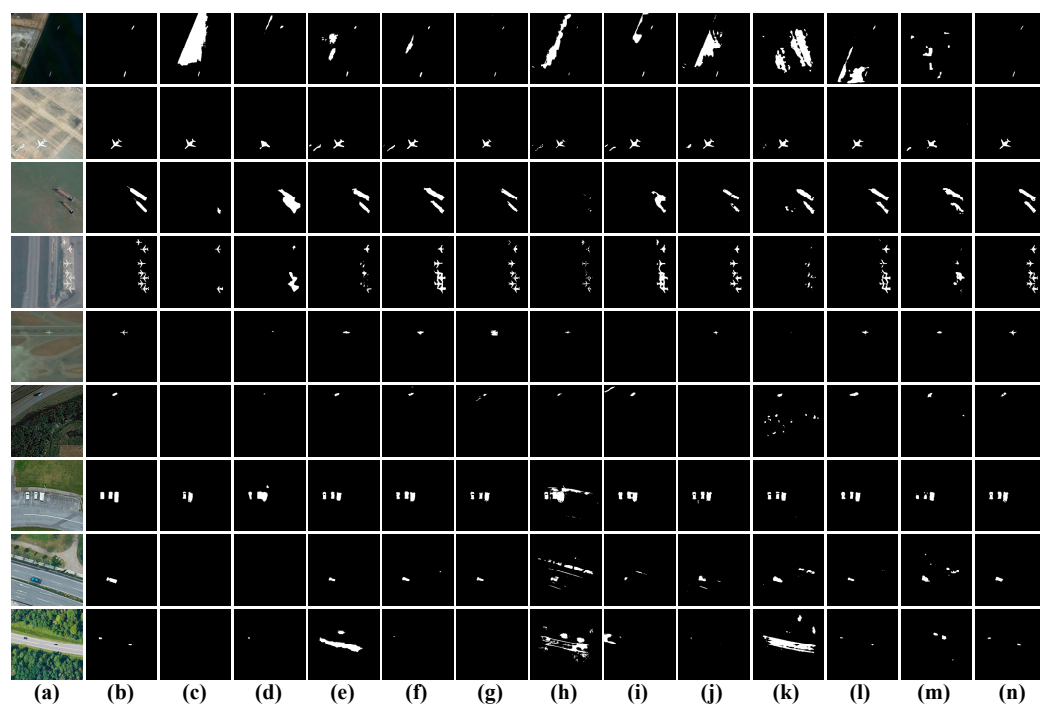


Figure 12. Visualization results of different models on the EORSSD dataset. (a) Original images; (b) Mask images; (c) U-Net; (d) SeaNet; (e) HST-UNet; (f) MsanlfNet; (g) IMFF-Net; (h) SwinT; (i) DECENet; (j) BRRNet; (k) EGENet; (l) UNetFormer; (m) DCSwin, (n) MCFI-Net.

3.7. Complexity Analysis

To comprehensively evaluate the computational efficiency of the proposed MCFI-Net, a complexity analysis was conducted in terms of GFLOPs, parameter count, and frames per second (FPS), as summarized in Table 7. Among all the compared models, SeaNet exhibits the lowest computational burden, with only 1.43 GFLOPs, 2.75 M parameters, and 73.96 FPS, indicating limited efficiency despite its lightweight design. U-Net and

DECENet maintain moderate complexity, with 3.49 GFLOPs and 3.17 GFLOPs. However, their representational capability remains constrained, as reflected in their lower segmentation accuracy. In contrast, models such as HST-UNet and MsanlfNet significantly increase computational costs, reaching 10.96 GFLOPs/28.61 M and 6.95 GFLOPs/32.24 M, while achieving relatively lower FPS values (22.02 and 28.09). This suggests that their complex hierarchical feature extraction structures impose heavier computational demands without a proportional performance gain. Similarly, due to its transformer-heavy architecture, the computational requirements of DCSwin are 34.46 GFLOPs and 118.87 M parameters, which result in its lowest FPS (27.20), making it less suitable for real-time or resource-constrained applications. Conversely, BRRNet and IMFF-Net exhibit high GFLOPs (58.79 and 71.96) but deliver notably faster inference speeds (150.62 FPS and 133.44 FPS), benefiting from efficient network parallelization and streamlined feature flow design. Conversely, BRRNet and IMFF-Net exhibit high GFLOPs (58.79 and 71.96) but deliver notably faster inference speeds (150.62 FPS and 133.44 FPS), benefiting from efficient network parallelization and streamlined feature flow design. The proposed MCFI-Net achieves 48.36 GFLOPs, 19.37 M parameters, and an inference speed of 60.42 FPS, striking an effective balance between computational cost and segmentation accuracy. Although not the lightest model, MCFI-Net's design ensures superior representational power without excessive resource consumption.

Table 7. The complexity analysis of all models on the Landslide4Sense dataset.

Method	GFLOPs	Params(M)	FPS
U-Net [30]	3.49	1.95	227.32
SeaNet [41]	1.43	2.75	73.96
HST-UNet [42]	10.96	28.61	22.02
MsanlfNet [43]	6.95	32.24	28.09
IMFF-Net [44]	71.96	34.68	133.44
SwinT [45]	7.38	2.78	39.86
DECENet [46]	3.17	1.16	44.58
BRRNet [47]	58.79	14.93	150.62
EGENet [48]	15.53	2.98	80.42
UNetFormer [49]	2.93	11.68	104.20
DCSwin [50]	34.46	118.87	27.20
MCFI-Net	48.36	19.37	60.42

3.8. Selection of the Optimizer

To investigate the impact of different optimization algorithms on the segmentation performance of MCFI-Net, we conducted a series of comparative experiments using five widely adopted optimizers: Adam, Nadam, RMSprop, Rprop, and SGD. The quantitative results on the Landslide4Sense dataset are summarized in Table 8. Among all the tested optimizers, Adam achieves the best overall performance, with a Dice score of 72.54%, an Mcc of 71.38%, and a Jaccard index of 56.99%. This demonstrates that Adam provides more stable and effective convergence during training, leading to more accurate and robust segmentation outcomes. In contrast, SGD yields the lowest performance across all three metrics, with Dice, Mcc, and Jaccard values of only 67.96%, 66.70%, and 51.65%, respectively. This suggests that the basic stochastic gradient descent algorithm may struggle with the complex optimization landscape required for landslide segmentation tasks. The results of Nadam, RMSprop, and Rprop fall in between, with Nadam slightly outperforming the others, achieving a Dice score of 69.89%, while RMSprop and Rprop yield Dice scores of 69.47% and 68.85%. Although these optimizers improve performance compared to SGD, they still fall short of Adam in all metrics. Overall, Adam emerges as the most effective optimization strategy for training MCFI-Net on the Landslide4Sense dataset.

Table 8. The selection experiment of the optimizer on the Landslide4Sense dataset.

Method	Dice	Mcc	Jaccard
Adam	0.7254	0.7138	0.5699
Nadam	0.6989	0.6859	0.5391
RMSprop	0.6947	0.6824	0.5339
Rprop	0.6885	0.6764	0.5265
SGD	0.6796	0.6670	0.5165

3.9. Selection of the Loss Function

To determine the most suitable loss function for training MCFI-Net on the Landslide4Sense dataset, we conducted a series of experiments using four widely utilized loss functions. The quantitative performance for each setting is reported in Table 9. Among the tested loss functions, Dice Loss achieves the highest segmentation performance across all three metrics, with a Dice score of 72.54%, an Mcc of 71.38%, and a Jaccard score of 56.99%. This indicates that Dice loss provides an effective optimization objective for landslide segmentation tasks by directly maximizing the overlap between predicted and ground truth regions. In contrast, Tversky loss performs the weakest, yielding a Dice score of only 53.2, an Mcc of 56.84%, and a Jaccard score of 36.44%. The underperformance may be attributed to its sensitivity to imbalanced classes or difficulty in optimizing sharp boundaries within landslide regions. FocalTversky and IoU loss provide similar performance, with Dice of 54.44% and 54.15%, and Jaccard of 37.57% and 37.26%, respectively. In summary, the comparative results clearly demonstrate that Dice loss is the most appropriate choice for the MCFI-Net architecture when applied to landslide segmentation tasks, and it offers the most balanced and robust performance in terms of both overlap accuracy and overall prediction reliability.

Table 9. The selection experiment of the loss function on the Landslide4Sense dataset.

Method	Dice	Mcc	Jaccard
Dice	0.7254	0.7138	0.5699
FocalTversky	0.5444	0.5786	0.3757
IoU	0.5415	0.5753	0.3726
Tversky	0.5327	0.5684	0.3644

4. Conclusions

In this study, we present MCFI-Net, a multi-scale cross-layer feature interaction network specifically developed for precise landslide segmentation in complex remote sensing imagery. Built upon an encoder–decoder architecture, MCFI-Net integrates multi-level contextual and spatial information by incorporating innovative components such as the receptive field block and the multi-branch dynamic convolution block. The RFB enhances the network’s ability to capture multi-scale features through skip connections, while the MDCB simultaneously enables dynamic receptive perception and robust multi-scale representation. Through extensive comparative and ablation experiments on the Landslide4Sense, Bijie and EORSSD datasets, MCFI-Net demonstrates strong segmentation accuracy and generalization capability. These results validate the robustness and practical applicability of MCFI-Net in real-world landslide detection scenarios. Overall, MCFI-Net offers a highly accurate, adaptable, and efficient solution for landslide segmentation, showing significant promise in the domain of geohazard monitoring and disaster mitigation.

Author Contributions: Conceptualization, J.L.; methodology, L.Y.; validation, J.L. and L.Y.; writing—original draft preparation, J.L.; writing—review and editing, L.Y.; visualization, L.Y.; supervision, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Landslide4Sense, Bijie and EORSSD are public datasets, and their links are mentioned in the paper.

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

References

1. Yang, Y.; Miao, Z.; Li, X.; Zhang, H.; Chen, S. LSL-SS-Net: Level set loss-guided semantic segmentation networks for landslide extraction. *GISci. Remote Sens.* **2024**, *61*, 2416727. [\[CrossRef\]](#)
2. Han, Z.; Fu, B.; Fang, Z.; Li, Y.; Li, J.; Jiang, N.; Chen, G. Dynahead-YOLO-Otsu: An efficient DCNN-based landslide semantic segmentation method using remote sensing images. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2398103. [\[CrossRef\]](#)
3. Lv, J.; Zhang, R.; Wu, R.; Bao, X.; Liu, G. Landslide detection based on pixel-level contrastive learning for semi-supervised semantic segmentation in wide areas. *Landslides* **2025**, *22*, 1087–1105. [\[CrossRef\]](#)
4. Kaushal, A.; Gupta, A.K.; Sehgal, V.K. A semantic segmentation framework with UNet-pyramid for landslide prediction using remote sensing data. *Sci. Rep.* **2024**, *14*, 30071. [\[CrossRef\]](#)
5. Zhu, X.; Zhang, Z.; He, Y.; Wang, W.; Yang, S.; Hou, Y. LandslideNet: A landslide semantic segmentation network based on single-temporal optical remote sensing images. *Adv. Space Res.* **2024**, *74*, 4616–4638. [\[CrossRef\]](#)
6. Liu, B.; Wang, W.; Wu, Y.; Gao, X. Attention Swin Transformer UNet for Landslide Segmentation in Remotely Sensed Images. *Remote Sens.* **2024**, *16*, 4464. [\[CrossRef\]](#)
7. Yu, C.C.; Chen, Y.D.; Cheng, H.Y.; Jiang, C.L. Semantic segmentation of satellite images for landslide detection using foreground-aware and multi-scale convolutional attention mechanism. *Sensors* **2024**, *24*, 6539. [\[CrossRef\]](#)
8. Şener, A.; Ergen, B. LandslideSegNet: An effective deep learning network for landslide segmentation using remote sensing imagery. *Earth Sci. Inform.* **2024**, *17*, 3963–3977. [\[CrossRef\]](#)
9. Jin, Y.; Liu, X.; Huang, X. EMR-HRNet: A multi-scale feature fusion network for landslide segmentation from remote sensing images. *Sensors* **2024**, *24*, 3677. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Zhou, N.; Hong, J.; Cui, W.; Wu, S.; Zhang, Z. A multiscale attention segment network-based semantic segmentation model for landslide remote sensing images. *Remote Sens.* **2024**, *16*, 1712. [\[CrossRef\]](#)
11. Wang, K.; He, D.; Sun, Q.; Yi, L.; Yuan, X.; Wang, Y. A novel network for semantic segmentation of landslide areas in remote sensing images with multi-branch and multi-scale fusion. *Appl. Soft Comput.* **2024**, *158*, 111542. [\[CrossRef\]](#)
12. Ling, Y.; Wang, Y.; Lin, Y.; Chan, T.O.; Awange, J. Optimized landslide segmentation from satellite imagery based on multi-resolution fusion and attention mechanism. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6015505. [\[CrossRef\]](#)
13. Wang, Z.; Wen, Y.; Ma, Y.; Peng, W.; Lu, Y. Optimizing under-sampling in Fourier single-pixel imaging using GANs and attention mechanisms. *Opt. Laser Technol.* **2025**, *187*, 112752. [\[CrossRef\]](#)
14. Xu, W.; Wan, Y.; Zhao, D. SFA: Efficient attention mechanism for superior CNN performance. *Neural Process. Lett.* **2025**, *57*, 38. [\[CrossRef\]](#)
15. Cheng, D.; Zhou, Z.; Li, H.; Zhang, J.; Yang, Y. A morphological difference and statistically sparse Transformer-based deep neural network for medical image segmentation. *Appl. Soft Comput.* **2025**, *174*, 113052. [\[CrossRef\]](#)
16. Hasannezhad, A.; Sharifian, S. Explainable AI enhanced transformer based UNet for medical images segmentation using gradient weighted class activation map. *Signal Image Video Process.* **2025**, *19*, 321. [\[CrossRef\]](#)
17. Wang, L.; Wu, G.; Tossou, A.I.H.C.; Liang, Z.; Xu, J. Segmentation of crack disaster images based on feature extraction enhancement and multi-scale fusion. *Earth Sci. Inform.* **2025**, *18*, 55. [\[CrossRef\]](#)
18. Guo, H.; Shi, L.; Liu, J. An improved multi-scale feature extraction network for medical image segmentation. *Quant. Imaging Med. Surg.* **2024**, *14*, 8331–8346. [\[CrossRef\]](#)
19. Hu, Y.; Guo, Y.; Xu, X.; Xie, S. Medical image segmentation of gastric adenocarcinoma based on dense connection of residuals. *J. Appl. Clin. Med. Phys.* **2024**, *25*, e14233. [\[CrossRef\]](#)
20. Zhang, J.; Zhang, Y.; Qiu, H.; Wang, T.; Li, X.; Zhu, S.; Huang, M.; Zhuang, J.; Shi, Y.; Xu, X. Constrained multi-scale dense connections for biomedical image segmentation. *Pattern Recognit.* **2025**, *158*, 111031. [\[CrossRef\]](#)
21. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8 September 2018; pp. 801–818.

22. 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.
23. Zhang, J.; Li, D.; Zeng, Z.; Zhang, R.; Wang, J. Dual-branch crack segmentation network with multi-shape kernel based on convolutional neural network and Mamba. *Eng. Appl. Artif. Intell.* **2025**, *150*, 110536. [\[CrossRef\]](#)
24. Yang, G.; Yang, J.; Fan, W.; Yang, D. Neural network for underwater fish image segmentation using an enhanced feature pyramid convolutional architecture. *J. Mar. Sci. Eng.* **2025**, *13*, 238. [\[CrossRef\]](#)
25. Zheng, J.; Fu, Y.; Chen, X.; Zhao, R.; Lu, J.; Zhao, H.; Chen, Q. EGCM-UNet: Edge guided hybrid CNN-Mamba UNet for farmland remote sensing image semantic segmentation. *Geocarto Int.* **2025**, *40*, 2440407. [\[CrossRef\]](#)
26. Zhou, T. Boundary-aware and cross-modal fusion network for enhanced multi-modal brain tumor segmentation. *Pattern Recognit.* **2025**, *165*, 111637. [\[CrossRef\]](#)
27. Li, X.; Xu, F.; Liu, F.; Lyu, X.; Tong, Y.; Xu, Z.; Zhou, J. A synergistical attention model for semantic segmentation of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5400916. [\[CrossRef\]](#)
28. Li, X.; Xu, F.; Liu, F.; Tong, Y.; Lyu, X.; Zhou, J. Semantic segmentation of remote sensing images by interactive representation refinement and geometric prior-guided inference. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 5400318. [\[CrossRef\]](#)
29. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. In *International Workshop on Deep Learning in Medical Image Analysis*; Springer: Cham, Switzerland, 2018; pp. 3–11.
30. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
31. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020; pp. 11030–11039.
32. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. [\[CrossRef\]](#)
33. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
34. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5633017. [\[CrossRef\]](#)
35. Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [\[CrossRef\]](#)
36. Zhang, Q.; Cong, R.; Li, C.; Cheng, M.M.; Fang, Y.; Cao, X.; Zhao, Y.; Kwong, S. Dense attention fluid network for salient object detection in optical remote sensing images. *IEEE Trans. Image Process.* **2021**, *30*, 1305–1317. [\[CrossRef\]](#)
37. Yuan, H.J.; Chen, L.N.; He, X.F. MMUNet: Morphological feature enhancement network for colon cancer segmentation in pathological images. *Biomed. Signal Process. Control* **2024**, *91*, 105927. [\[CrossRef\]](#)
38. Selvaraj, A.; Nithiyaraj, E. CEDRNN: A convolutional encoder-decoder residual neural network for liver tumour segmentation. *Neural Process. Lett.* **2023**, *55*, 1605–1624. [\[CrossRef\]](#)
39. Zhu, Q. On the performance of Matthews correlation coefficient (Mcc) for imbalanced dataset. *Pattern Recognit. Lett.* **2020**, *136*, 71–80. [\[CrossRef\]](#)
40. Yeung, M.; Rundo, L.; Nan, Y.; Sala, E.; Schönlieb, C.B.; Yang, G. Calibrating the dice loss to handle neural network overconfidence for biomedical image segmentation. *J. Digit. Imaging* **2023**, *36*, 739–752. [\[CrossRef\]](#)
41. 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. [\[CrossRef\]](#)
42. Zhou, H.; Xiao, X.; Li, H.; Liu, X.; Liang, P. Hybrid Shunted Transformer embedding UNet for remote sensing image semantic segmentation. *Neural Comput. Appl.* **2024**, *36*, 15705–15720. [\[CrossRef\]](#)
43. Bai, L.; Lin, X.; Ye, Z.; Xue, D.; Yao, C.; Hui, M. MsanlfNet: Semantic segmentation network with multiscale attention and nonlocal filters for high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6512405. [\[CrossRef\]](#)
44. Liu, M.; Wang, Y.; Wang, L.; Hu, S.; Wang, X.; Ge, Q. IMFF-Net: An integrated multi-scale feature fusion network for accurate retinal vessel segmentation from fundus images. *Biomed. Signal Process. Control* **2024**, *91*, 105980. [\[CrossRef\]](#)
45. Xu, Y.; Yu, K.; Qi, G.; Gong, Y.; Qu, X.; Yin, L.; Yang, P. Brain tumour segmentation framework with deep nuanced reasoning and Swin-T. *IET Image Process.* **2024**, *18*, 1550–1564. [\[CrossRef\]](#)
46. Tang, Q.; Min, S.; Shi, X.; Zhang, Q.; Liu, Y. DESENet: A bilateral network with detail-enhanced semantic encoder for real-time semantic segmentation. *Meas. Sci. Technol.* **2024**, *36*, 015425. [\[CrossRef\]](#)
47. Shao, Z.; Tang, P.; Wang, Z.; Saleem, N.; Yam, S.; Sommai, C. BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images. *Remote Sens.* **2020**, *12*, 1050. [\[CrossRef\]](#)
48. Chen, W.; Wang, Q.; Wang, D.; Xu, Y.; He, Y.; Yang, L.; Tang, H. A lightweight and scalable greenhouse mapping method based on remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103553. [\[CrossRef\]](#)

49. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 196–214. [[CrossRef](#)]
50. Wang, L.; Li, R.; Duan, C.; Zhang, C.; Meng, X.; Fang, S. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6506105. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.