fire

MDPI

Article

Attention-Enhanced Segmentation for Vegetation and Snow Cover Extraction Supporting Grassland Fire Danger Factor Monitoring

Weiping Liu [1,2], Shuye Chen [1,2], Yun Yang [1,2] and Yili Zheng [1,2,3,*]

[1] School of Technology, Beijing Forestry University, Beijing 100083, China; lwp123@bjfu.edu.cn (W.L.); chenshuye@bjfu.edu.cn (S.C.); yyfelmina@bjfu.edu.cn (Y.Y.)
[2] Key Laboratory of National Forestry and Grassland Administration on Forestry Equipment and Automation, Beijing 100083, China
[3] State Key Laboratory of Efficient Production of Forest, Beijing 100083, China
* Correspondence: zhengyili@bjfu.edu.cn

Abstract

Grassland fire is one of the major disasters threatening regional ecological security. Its occurrence, development, and spread are closely related to the spatial distribution and coverage of surface vegetation and snow cover across grassland areas. As the primary combustible fuel source, higher vegetation coverage increases fuel load and continuity, thereby directly determining grassland fire danger levels and accelerating fire spread velocity. In contrast, snow cover imposes an indirect regulatory effect on the spatiotemporal pattern of fire danger factors: it lowers surface temperature, raises near-surface humidity, and restricts the germination and growth of herbaceous vegetation in cold seasons, which effectively reduces available combustible materials and weakens regional fire hazard conditions. Therefore, accurately obtaining the coverage status of vegetation (direct combustible fuel factor) and snow cover (indirect fire-regulating factor) in complex grassland scenarios is the essential premise for reliable grassland fire danger monitoring, early warning, disaster prevention and control, and regional ecological management. Aiming at the practical problems in complex grassland scenarios (such as undulating terrain, uneven vegetation growth, large differences in snow depth, and complex lighting conditions), including difficulty in extracting vegetation and snow-covered areas, blurred and confusing boundaries, and low accuracy in coverage calculation, which seriously restrict the technical bottleneck of precise monitoring of grassland fire danger factors, this study takes near-ground images collected by grassland fire danger factor monitoring stations as the core data source, and proposes an improved UNet image segmentation model combined with image segmentation technology and deep learning methods to realize precise extraction of vegetation and snow-covered areas and efficient calculation of coverage in complex scenarios. To improve the model's feature extraction ability, boundary localization accuracy, and reduce model parameters and computational overhead, the CBAM-ASPP (Convolutional Block Attention Module—Atrous Spatial Pyramid Pooling) module is integrated at the end of the encoding path. The attention mechanism is used to enhance the weight of key features, and the multi-scale receptive field of atrous spatial pyramid pooling is utilized to strengthen the model's ability to fuse features of vegetation and snow areas of different scales. The residual attention mechanism is introduced in the upsampling stage to effectively alleviate the gradient disappearance problem, improve the model's ability to accurately locate the boundaries of vegetation and snow areas, and reduce segmentation errors. In the training process, a dynamically weighted hybrid loss function is adopted to dynamically adjust the weights according to the segmentation difficulty of different types of samples during

Check for updates

Academic Editors: Mingyu Wang, Jili Zhang, Xiaodong Liu and Feng Chen

Received: 4 April 2026
Revised: 11 May 2026
Accepted: 18 May 2026
Published: 20 May 2026

Copyright: © 2026 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.

training, optimize the model training effect, and improve the segmentation accuracy and generalization ability. Experiments were conducted using near-ground images of typical complex grassland scenarios as the dataset, and the performance of the proposed model was verified through comparative experiments. The results show that in the vegetation segmentation task, the mean Intersection over Union (mIoU) of the model reaches 84.70%, and the accuracy rate is 91.28%, which are 1.48 and 1.58 percentage points higher than those of the standard UNet model, respectively. In the snow segmentation task, the mIoU of the model reaches 92.74%, and the accuracy rate is 94.19%, which are 2.39 and 2.36 percentage points higher than those of the standard UNet model, respectively. At the same time, the number of parameters of the model is reduced by 12.85% compared with the standard UNet. Also, its comprehensive performance is significantly better than that of mainstream image segmentation models such as FCN, SegNet, and DeepLabv3+. Based on the standardized time-series data retrieved by the optimized segmentation model, this study further constructs a Grassland Fire Risk Index (GFRI) using the Analytic Hierarchy Process (AHP). Pearson correlation verification confirms that the GFRI has an extremely significant positive correlation with historical fire frequency, accurately capturing the seasonal dynamic rhythm of regional grassland fire occurrence. This integrated framework of intelligent segmentation and fire risk quantification provides a reliable technical solution for grassland fire factor monitoring, dynamic fire risk assessment, early warning systems, and refined regional ecological management.

Keywords: image segmentation; multi-scale feature fusion; residual attention mechanism; fire risk monitoring station; fire risk index

1. Introduction

Grasslands are vital natural resources, providing climate regulation and carbon-oxygen balance maintenance, serving as irreplaceable terrestrial ecosystems. China boasts the world's largest grassland area, nearly 4 billion mu (approximately 266.7 million hectares), yet its ecosystems are inherently fragile. Grassland fires [1] are widespread, recurring natural disasters with rising frequency, extent, and intensity globally, severely threatening ecological balance, socioeconomic development, and human safety. Annually, large-scale fires burn vast amounts of vegetation, releasing 300–400 million tons of carbon worldwide, disrupting atmospheric circulation and climate stability. During fires, massive pollutants are emitted, destroying habitats and ecological balance. Post-fire, grasslands degrade into bare land within months to years, accelerating desertification and triggering sandstorms and erosion [2]. In China's key grassland regions like Inner Mongolia and Xinjiang, a single medium-to-large fire can degrade tens of thousands of hectares of pasture, causing tens of millions of yuan in direct losses and endangering local livelihoods. Beyond ecological degradation, fires reduce grassland productivity, damage livestock bases and infrastructure, and hinder carbon peaking and neutrality goals. Thus, accurate grassland fire risk identification is critical for disaster prevention and mitigation [3]. The occurrence, development, and spread of grassland fires are closely related to the distribution characteristics and coverage status of surface vegetation and snow in grasslands. The occurrence and spread of grassland fires are closely related to surface vegetation and snow cover conditions. Vegetation serves as the fundamental combustible source for grassland fires, and the coverage of withered and dry grass largely determines surface fuel abundance and continuity. Sufficient dry vegetation coverage creates favorable combustion conditions, elevates regional fire danger levels, and accelerates fire propagation. In comparison,

snow cover indirectly modulates grassland fire hazards. It lowers surface temperature and increases atmospheric humidity to constrain fire ignition, while limiting cold-season vegetation growth and reducing flammable fuel accumulation [4]. These dual effects effectively suppress fire risks. Therefore, accurate identification of vegetation and snow cover is critical for evaluating fire risk levels in grassland regions and provides an essential basis for disaster prevention and mitigation in agricultural and pastoral production, as well as for the scientific management of grassland resources.

Currently, vegetation and snow cover are primarily monitored through satellite remote sensing and near-surface observation, forming a spatially complementary data acquisition system. Near-surface monitoring, as an important supplement to satellite remote sensing, offers advantages such as high spatial resolution, high temporal frequency, and immunity to atmospheric cloud interference [5]. By capturing micro-level details at centimeter-scale resolution [6], it provides a benchmark for validating the accuracy of satellite remote sensing data and calibrating models [7]. Donager et al. [8] employed LiDAR systems to obtain high-temporal-frequency snow cover datasets and, after extensive data preprocessing, successfully achieved rapid assessment of dynamically changing snow cover; however, this approach is limited by high equipment costs and constrained operational range. Compared with LiDAR, visible-light cameras offer lower cost and reduced data preprocessing requirements [9]. Furukawa et al. [10] collected vegetation data using high-resolution cameras and, through comparison with multispectral cameras, demonstrated that visible-light cameras provide higher spatial resolution and can capture finer variations when monitoring vegetation cover.

China has established an integrated "space–air–ground" observation network for grassland ecosystems, supporting dynamic monitoring of grassland resources and providing critical data for assessing grassland fire risk levels [11]. The monitoring stations within this network have collected a substantial volume of vegetation and snow imagery; however, methods that rely on manual visual estimation or interactive software interpretation of cover have limitations, including low efficiency and significant subjective bias [12], making them insufficient for the demand for efficient and accurate quantification of cover [13]. Traditional automated methods for identifying vegetation and snow cover mainly rely on features such as color, shape, and texture [14], which are often inadequate for capturing the complex information within images and generally lack robust generalization capabilities. Moreover, conventional image recognition algorithms are highly sensitive to image scale and deformation, limiting their adaptability to the complex and variable conditions of field imagery [15]. Consequently, due to the diversity of vegetation structures and unstable lighting conditions, traditional algorithms struggle to capture fine details accurately, resulting in low precision in cover calculation [16].

Compared with traditional machine learning methods, deep learning-based image recognition does not require manual feature description or selection for target images. Using frameworks based on convolutional neural networks (CNNs) or visual Transformers (ViTs), deep learning integrates feature extraction and recognition tasks [17], enabling end-to-end model training for automatic learning and adaptation to diverse cover identification scenarios [18]. Guo [19] proposed a transfer learning-based approach for automatic extraction of snow cover, achieving promising segmentation results on small-sample datasets through an improved Xception model and DeepLabv3+ architecture. Kalinaki [20] employed the DeepLabv3+ model to identify vegetation and estimate cover along highways, achieving a high mean Intersection over Union (mIoU) of 0.92. Wu [21] trained a multi-task CNN model to perform urban vegetation cover recognition, effectively mitigating shadow interference caused by solar angle and building occlusion. These studies demonstrate the significant potential and advantages of deep learning in cover identification research.

However, deep learning models require substantial computational resources, and their generalization capability in complex environments remains limited, constraining their effectiveness in field monitoring [22]. Therefore, enhancing model robustness and accuracy across diverse scenarios remains one of the key challenges in deep learning-based research for vegetation and snow cover identification.

This study focuses on a near-surface image dataset collected from grassland fire risk monitoring stations. Building upon a classical image segmentation framework, it incorporates a multi-scale attention mechanism and structural optimization strategy, and optimizes the loss function to achieve accurate segmentation of grassland surface vegetation and snow cover. Existing grassland fire risk evaluation is largely constrained by inaccurate acquisition of underlying surface environmental factors, which reduces the reliability of risk assessment results. To solve this problem, this study quantitatively extracts vegetation and snow cover information based on the improved segmentation model, and further constructs a targeted Grassland Fire Risk Index (GFRI) on the basis of conventional assessment frameworks to realize refined grassland fire risk evaluation. The quantitative monitoring of vegetation and snow cover conditions and the establishment of the fire risk indicator system can effectively support dynamic grassland resource supervision and sustainable management, thereby facilitating grassland ecological security protection and promoting regional sustainable development.

2. Materials and Methods

2.1. Dataset Construction

2.1.1. Data Collection

The vegetation and snow data were collected at grassland fire risk monitoring stations in the Xinjiang Uygur Autonomous Region. As illustrated in Figure 1, a total of 332 monitoring sites are spatially distributed across the study area, spanning a longitude range of 75.59° E to 95.54° E and a latitude range of 36.07° N to 48.36° N. These stations cover 14 prefecture-level administrative regions, including 5 autonomous prefectures and 9 districts. The spatial distribution spans from semi-humid zones to extremely arid zones, ensuring systematic representation of regional ecological diversity.

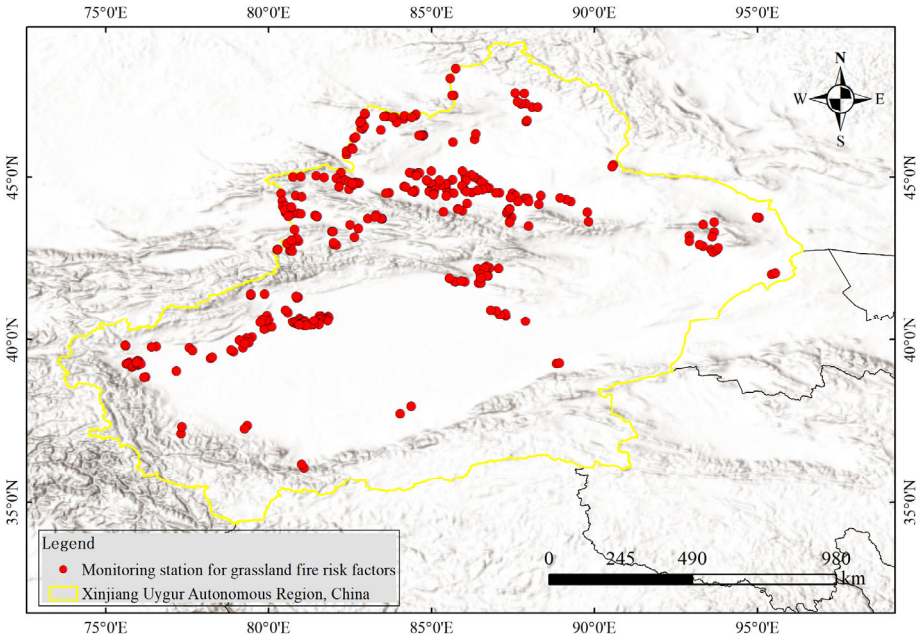


Figure 1. Layout of monitoring stations for grassland fire risk factors.

The monitoring stations are installed in field environments and powered continuously by 100 W solar panels. As shown in Figure 2, the stations are equipped with high-definition phenological cameras, which, through edge computing controllers, enable fixed-point, time-controlled image acquisition with automatic focusing, supporting multi-angle and multi-mode shooting. The cameras are fixed in position with a vertical nadir (top-down) view, installed at a height of 3 m above the ground.



Figure 2. Monitoring station for grassland fire risk factors.

A total of 1500 vegetation images were collected to ensure sufficient sample representativeness for model training and validation. All samples belong entirely to the grass family, which is the dominant vegetation type in the study area. Images were captured under diverse weather conditions, including sunny, cloudy, rainy, snowy and foggy days, with variable lighting conditions and vegetation growth status. The sampling period covered August to December 2024 (vegetation senescence period) and March to May 2025 (vegetation growth period). Although the March–May vegetation growth season generally witnesses a relatively low overall probability of large-scale grassland fires, this period is selected for targeted data acquisition due to the vast geographic coverage of Xinjiang. Some regions show lush grassland vegetation, while others still retain extensive bare land. During this period, there is a considerable amount of bare soil, which can enrich the background data and assist in distinguishing between vegetation and snow-covered areas.

Based on the predefined 1 m × 1 m plot area and imaging scale, all the original images were cropped to retain only the valid coverage range within each actual plot boundary. Examples of images from areas with complex interference are shown in Figure 3.



a. Object interference area image

b. Snow homogeneous color interference area image



c. Shadow interference area image



d. Haze interference area image

Figure 3. Diverse scenarios.

2.1.2. Data Preprocessing

Vegetation and snow in the images were annotated separately using the open-source labeling software Labelme (v5.10.1), and the resulting JSON files were converted into label images (BMP format) via OpenCV (v4.8.1).

This study used two datasets, each containing 1500 images, which were randomly divided into 667 images for an expandable sample set and 833 images for the test set. After data augmentation, the ratio of training to test images was maintained at 8:2. This partitioning strategy aims to train the model with a relatively small amount of original data while testing it under more diverse and complex scenarios, thereby evaluating the model's generalization ability and robustness.

To increase data diversity and further enhance model generalization and robustness, the training images were augmented using the OpenCV library through operations such as horizontal flipping, brightness adjustment, noise addition, and rotation. These augmentations simulated varying lighting conditions and shadow regions. After augmentation (Figure 4), each of the vegetation and snow datasets was expanded to 3335 images and subsequently split into training and validation sets at a 9:1 ratio. Only the augmented datasets were used to train the vegetation and snow segmentation models.

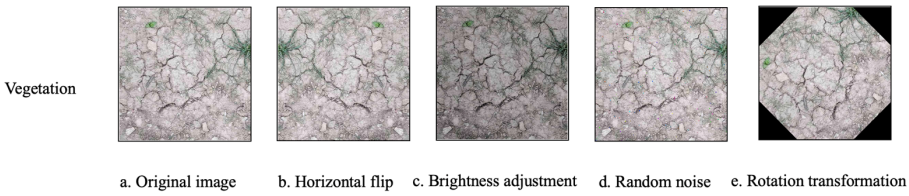


Figure 4. Sample images with data augmentation.

Additionally, statistical analysis of the original image data revealed the following pixel proportion distributions: for vegetation, 0–30% of pixels accounted for 23% of images, 30–60% for 37%, and 60–100% for 40%; for snow, 0–30% of pixels accounted for 22% of images, 30–60% for 35%, and 60–100% for 43%.

2.2. Model Construction

2.2.1. Overall Framework

The improved model, optimized and extended based on the UNet architecture, is shown in Figure 5. In this design, a structurally optimized CBAM-ASPP module is introduced at the end of the UNet encoding path to enhance multi-scale feature extraction. A residual attention mechanism is embedded in the decoding stage to improve feature reconstruction and localization accuracy. Additionally, a dynamically weighted hybrid loss function is incorporated in the final decoding layer to further optimize the model's segmentation performance and robustness.

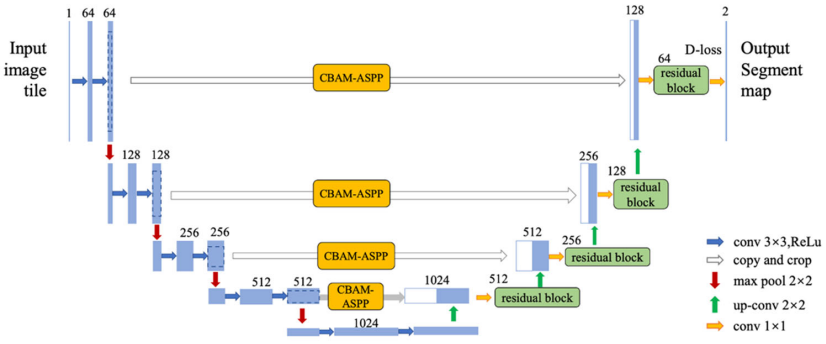


Figure 5. Improved model structure diagram.

2.2.2. Feature Extraction Network

To improve the efficiency of multi-scale feature fusion, this study reconstructs the standard UNet skip connection mechanism by embedding a structurally optimized CBAM-ASPP module within the feature interaction paths between the encoder and decoder. This module reorganizes feature-rich information from shallower layers into deeper layers. Atrous Spatial Pyramid Pooling (ASPP) is a classical module that applies multi-scale convolutions and pooling to feature maps. Its atrous convolution layers can analyze objects of different scales and sizes in the original image from multiple complementary receptive field dimensions [23], thereby enhancing the model's ability to recognize objects across varying scales.

In this study, depthwise separable convolution (DSC) is used to replace the standard convolution layers in the ASPP module, achieving an optimized design for the ASPP architecture. This modification effectively reduces the number of parameters while significantly improving model training efficiency. The ASPP structure comprises a 1×1 convolution layer, three 3×3 atrous convolution branches with dilation rates of 3, 6, and 12, and a global average pooling layer. The global average pooling layer integrates features from multiple branches, enabling efficient fusion and concatenation of image information and thereby enhancing segmentation accuracy.

However, during multi-scale feature extraction, traditional ASPP can lose critical information due to excessive receptive field expansion [24], which may degrade segmentation precision. To address this issue, a CBAM attention mechanism is further incorporated into the ASPP module, forming the CBAM-ASPP module. Different from simple parallel superposition, this study adopts a sequential embedding integration strategy: after completing multi-scale feature extraction via the optimized ASPP branches, the fused multi-scale feature map is input into the CBAM module for secondary feature recalibration. This hierarchical integration mode enables the attention mechanism to dynamically suppress noisy background features and amplify effective vegetation and snow feature responses specifically for multi-scale fused outputs, rather than performing blind feature weighting. This module enhances the saliency of target features—specifically snow and vegetation—while suppressing background interference, by leveraging spatial, channel, and multi-scale interactions, thereby improving the model's ability to perceive complex targets. The structure is illustrated in Figure 6.

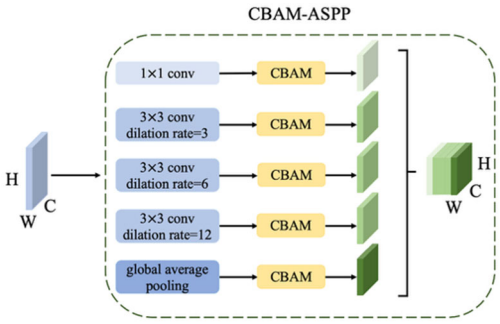


Figure 6. CBAM-ASPP module structure.

The CBAM module consists of two submodules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). Following a sequential attention computation process—channel attention first, then spatial attention—the module weights features along both the channel and spatial dimensions of the feature maps. The weighted features are then fused, enhancing the representational capability of the convolutional neural network [25]. This dual-dimensional attention mechanism enables the model to adaptively

learn and focus on informative features across both channel and spatial dimensions, thereby improving network accuracy.

The core function of the Channel Attention Module (CAM) is to capture dependencies among different feature channels. The input feature map is first processed through parallel max pooling and average pooling branches to extract spatial information. The outputs of these two branches are then passed through a multilayer perceptron (MLP), compressing the channel dimension to 1/6 of the original value (r = 6 in this study) and subsequently restoring it to the original channel number. The compressed-and-restored features are fused via element-wise addition, and a Sigmoid activation function generates the channel attention weights. Finally, these weights are multiplied element-wise with the original input feature map to dynamically optimize the channel features.

The computation is shown in Equation (1):

$$M_c(F) = \sigma\left[P\left(F^c_{avg}\right) + P(F^c_{max})\right] = \sigma\left\{W_1\left[W_0\left(F^c_{avg}\right) + W_1[W_0(F^c_{max})]\right]\right\} \quad (1)$$

In the equation, F denotes the input feature map; P represents the multilayer perceptron (MLP), with W_0 and W_1 corresponding to individual fully connected layers within P; and σ is the Sigmoid activation function.

The Spatial Attention Module (SAM) focuses on capturing the relationships among different spatial positions within a feature map. Unlike the channel attention mechanism, this module first applies max pooling and average pooling along the channel dimension of the input feature map to generate two spatial feature maps. These maps are then concatenated along the channel dimension and passed through a 7×7 convolution layer to compress the fused features into a single-channel feature map. A Sigmoid activation function is subsequently applied to generate the spatial attention map. Finally, this attention map is multiplied element-wise with the original input feature map, producing a feature map with spatially weighted features.

The computation is expressed in Equation (2):

$$M_s(F) = \sigma\left[f^{7\times7}\left(F^s_{avg}; F^s_{max}\right)\right] \quad (2)$$

In the equation, F denotes the input feature map; $f^{7\times7}$ represents the convolution operation with a 7×7 kernel; and σ is the Sigmoid activation function.

2.2.3. Residual Attention Mechanism

To address the gradient vanishing problem caused by increasing network depth [26], a residual attention module is introduced in the upsampling stage to integrate high-level semantic features with low-level details.

To enhance the network's ability to capture critical foreground information while avoiding parameter redundancy, the CBAM attention mechanism is embedded within the residual unit to improve feature representation. The proposed residual attention block adopts a standard residual learning framework, consisting of a main residual learning branch and a shortcut connection branch. In the main branch, the input feature map sequentially passes through one 3×3 convolutional layer, batch normalization (BN), ReLU activation, a second 3×3 convolutional layer, and another batch normalization layer; subsequently, the refined feature map is fed into the CBAM attention module for adaptive channel and spatial feature recalibration. For the shortcut branch, a 1×1 convolutional layer with batch normalization is utilized to match the channel dimension and spatial resolution of the main branch output, eliminating feature dimension mismatch during skip connection fusion. Finally, the attention-weighted main branch features and shortcut

residual features are fused via element-wise summation, followed by a ReLU activation to obtain the final output of the residual attention block.

The improved residual attention module, shown in Figure 7, consists of two 3 × 3 convolution layers and a CBAM module, with each convolution layer followed by batch normalization (BN) and ReLU activation. Additionally, a 1 × 1 convolution layer is added at the input of the residual module to unify the channel dimensions of the feature maps, effectively preventing dimension mismatches during skip connections.

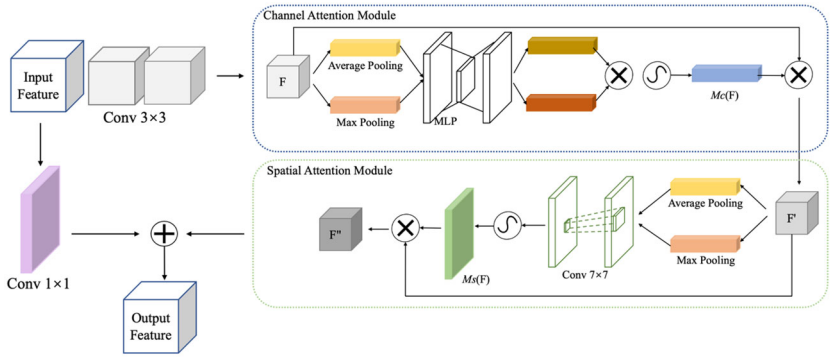


Figure 7. Improved residual channel attention block.

Leveraging the residual connection design, this module reduces the overall model complexity while preserving the original input features. As a result, the CBAM module does not need to learn the complete feature mapping; it only needs to fit the residual adjustment of the features. This design not only minimizes reliance on redundant parameters but also significantly improves the model's parameter efficiency.

2.2.4. Hybrid Loss Function

To collaboratively improve the regional integrity and boundary accuracy of segmentation results, this study experimentally adopts a hybrid loss function composed of the Dice loss and boundary loss, whose relative weights are dynamically adjusted during training to adapt to different learning stages. The Dice loss focuses on target pixels [27] and addresses the class imbalance problem by optimizing the intersection over union (IoU), thereby ensuring the accuracy of the overall target contour. Its calculation formula is shown in Equation (3):

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \times \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon} \tag{3}$$

In the equation, y_i denotes the ground-truth label of the i-th pixel (0 or 1); $\hat{y}_i$ denotes the predicted probability of the i-th pixel; N is the total number of pixels; and ϵ is a smoothing term to avoid division by zero.

To address the loss saturation issue inherent in the Dice loss, a boundary loss is incorporated to enhance boundary discrimination. Weighted by distance transform, this loss penalizes deviations between the predicted edges and the ground-truth boundaries, improving the transitional distinction between target pixels and the background. Its calculation formula is shown in Equation (4):

$$L_{Boundary} = \sum_{\rho \in \Omega} \Phi_G(\rho) \times s_\theta(\rho) \tag{4}$$

In the equation, ρ denotes an arbitrary pixel point in the image spatial domain Ω; Φ is the distance weighting function based on the ground-truth boundary G, derived from the signed distance function (SDF) of G (negative inside the boundary and positive

outside); and S denotes the probability that pixel ρ belongs to the background as output by the network.

Meanwhile, considering fluctuations in the background ratio across different batches, fixed weights may lead to dominance by one loss term and impair training performance. To further improve the adaptability of the loss function to varying class ratios and learning stages, we introduce a lightweight gating network to dynamically assign weights to L_{dice} and $L_{Boundary}$. This gating network is a single-layer fully connected network with a sigmoid activation, which takes as input a concatenated feature vector h_t, and outputs the weights $\alpha(t)$ for L_{dice} and $\beta(t)$ for $L_{Boundary}$, as shown in Equations (5) and (6).

$$\alpha(t) = \sigma(W \times h_t + b), \beta(t) = 1 - \alpha(t) \tag{5}$$

$$h_t = Concat\left(\frac{t}{T}, L_{Dice}, L_{Boundary}, M_c, R_b\right) \tag{6}$$

In Equations (5) and (6), t is the current training epoch; T is the total number of training epochs; R_b is the pixel ratio of foreground to background in the current batch; M_c is the channel mean of a high-level feature map from the decoder, computed as:

$$M_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F(i,j) \tag{7}$$

In the equation, $F(i,j)$ denotes the feature map at the spatial position (i,j); W and H are its height and width.

During training, the gating network is updated jointly with the segmentation network via backpropagation, using the same optimizer. The dynamic weights are computed at each iteration and applied to the two loss terms in the final hybrid loss:

$$Loss = \alpha(t) \times L_{dice} + \beta(t) \times L_{Boundary} \tag{8}$$

2.3. Model Training and Network Parameter Settings

This study implemented model development and training using Anaconda Python 3.9 (Anaconda Inc., Austin, Texas, USA) under the Windows 10 operating system. The experimental high-performance server was manufactured by Lenovo Group Limited, Beijing, China, and it was equipped with an NVIDIA® GeForce RTX™ 3090 Ti graphics card (23.68 GiB of video memory). The network training environment was constructed based on the PyTorch deep learning framework.

The training hyperparameters were configured as follows: the input images were resized to 640 × 640 pixels, the batch size was set to 16, the number of training epochs was 300, and the learning rate was 0.0001. No pre-trained weights were loaded, and the Adam optimizer was employed.

In this experiment, Mean Intersection over Union (mIoU), F1-score, and Precision were adopted as the performance evaluation metrics for the segmentation algorithm. The definitions of the metrics are given as follows:

1. Let the number of true positives, false positives, and false negatives be denoted as TP, FP, and FN, respectively, with the number of categories K = 2·mIoU reflecting the average segmentation accuracy of the model across all categories; a higher value indicates better model performance. Its calculation formula is shown in Equation (9):

$$mIou = \frac{1}{K} \sum_{k=1}^{K} \frac{TP_k}{TP_k + FP_k + FN_k} \tag{9}$$

2. The F1-score (F1-measure) is the harmonic mean of Precision and Recall, which comprehensively reflects the model's performance on both positive and negative samples. Its calculation formula is shown in Equation (10):

$$F_1\text{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

3. Precision refers to the proportion of actual positive samples among those predicted as positive. A higher value indicates a higher precision rate of the model. Its calculation formula is shown in Equation (11):

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

3. Results

3.1. Comparative Experiments of Different Segmentation Models

This section presents comparative experiments of the improved model on the vegetation and snow datasets. Several mainstream and high-performance classical image segmentation networks—FCN, SegNet, DeepLabv3+, and UNet [28–30]—were selected as baseline models for comparison.

The four aforementioned models were compared with the improved model proposed in this study. The segmentation results for the vegetation dataset are presented in Table 1, and the results for the snow dataset are shown in Table 2.

Table 1. Comparison of recognition performance among different models on the grassland dataset.

Models	mIoU/%	F1 Score/%	Precision/%	Parameters/M
FCN	76.90	82.57	80.13	71.10
SegNet	81.24	82.91	85.74	48.50
DeepLabv3+	82.34	83.93	89.10	58.00
UNet	83.22	84.84	89.70	51.11
Our model	84.70	86.12	91.28	44.54

Table 2. Comparison of recognition performance among different models on the snow-covered dataset.

Models	mIoU/%	F1 Score/%	Precision/%	Parameters/M
FCN	81.18	86.81	84.84	71.10
SegNet	84.97	89.68	90.50	48.50
DeepLabv3+	89.83	92.63	91.86	58.00
UNet	90.35	92.96	91.83	51.11
Our model	92.74	94.79	94.19	44.54

As shown in Table 1, the proposed method demonstrates significant advantages across all evaluation metrics, including mIoU, F1 score, and accuracy. For the vegetation dataset, the mIoU scores of FCN, SegNet, DeepLabv3+, and UNet were 76.90%, 81.24%, 82.34%, and 83.22%, respectively. The proposed model achieved an mIoU of 84.70%, representing improvements of 7.80, 3.46, 2.36, and 1.48 percentage points over the respective baseline models, demonstrating superior segmentation performance.

For the F1 score, FCN, SegNet, DeepLabv3+, and UNet scored 82.57%, 82.91%, 83.93%, and 84.84%, respectively. The proposed model achieved an F1 score of 86.12%, correspond-

ing to improvements of 3.55, 3.21, 2.19, and 1.28 percentage points over the respective baseline models.

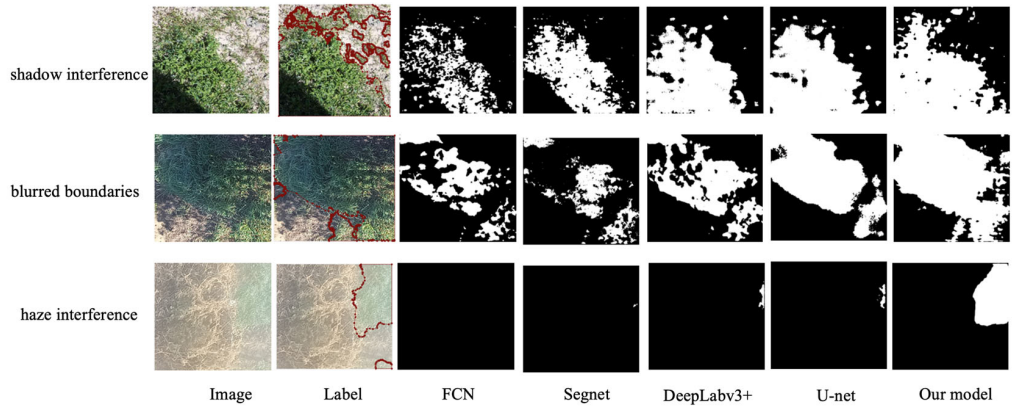
In terms of accuracy, FCN, SegNet, DeepLabv3+, and UNet achieved 80.13%, 85.74%, 89.10%, and 89.70%, respectively. The proposed model attained an accuracy of 91.28%, significantly outperforming the comparison models, with increases of 11.15, 5.54, 2.18, and 1.58 percentage points, respectively.

The evaluation results for the snow dataset (Table 2) indicate that the proposed method achieves significant advantages across mIoU, F1 score, and accuracy metrics. For the mIoU metric, FCN, SegNet, DeepLabv3+, and UNet scored 81.18%, 84.97%, 89.83%, and 90.35%, respectively. The proposed model achieved an mIoU of 92.74%, representing improvements of 11.56, 7.77, 2.91, and 2.39 percentage points over the respective baseline models.

For the F1 score, FCN, SegNet, DeepLabv3+, and UNet achieved 86.81%, 89.68%, 92.63%, and 92.96%, respectively. The proposed model attained an F1 score of 94.79%, improving by 7.98, 5.11, 2.16, and 1.83 percentage points compared with the baseline models.

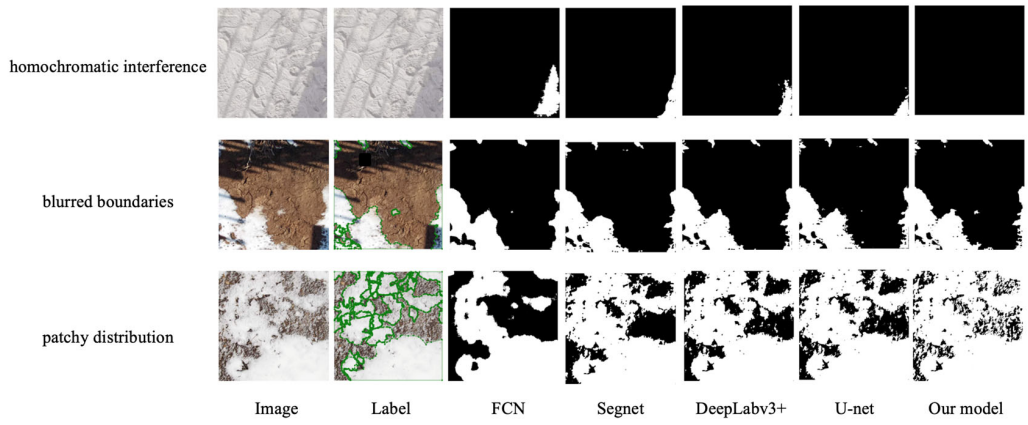
Regarding accuracy, FCN, SegNet, DeepLabv3+, and UNet achieved 84.84%, 90.50%, 91.86%, and 91.83%, respectively. The proposed model reached 94.19% accuracy, significantly outperforming the comparison models, with increases of 9.35, 3.69, 2.33, and 2.36 percentage points, respectively.

To visually demonstrate the advantages of the improved model, the segmentation results of all models on the test datasets were visualized. The segmentation results for the vegetation dataset are shown in Figure 8, and the results for the snow dataset are presented in Figure 9.



Note: The area enclosed by red lines indicates vegetation.

Figure 8. Recognition performance results for different models on grassland dataset.



Note: The area enclosed by green lines indicates snow.

Figure 9. Recognition performance results for different models on the snow-covered dataset.

Based on the segmentation performance results and the visualizations in Figures 8 and 9, the Fully Convolutional Network (FCN) performs poorly on the test datasets. Although FCN uses transposed convolution for upsampling, which can restore the spatial dimensions of feature maps, it cannot fully recover the original feature values. This limitation reduces the utilization of higher-level features, thereby constraining segmentation accuracy.

Comparative analysis shows that SegNet and DeepLabv3+ exhibit superior segmentation performance. SegNet preserves the locations of maximum values through unpooling operations, effectively enhancing feature utilization, while DeepLabv3+ captures contextual information at multiple scales using atrous convolution modules, strengthening edge detail extraction [30,31].

However, all the baseline models share a common limitation: their segmentation performance is suboptimal in shadowed regions and complex backgrounds, especially when the background color is similar to the target objects.

3.2. Ablation Study

To validate the effectiveness of each component in the proposed model—including the optimized CBAM-ASPP module, the residual attention mechanism (residual block), and the hybrid loss function—an ablation study was designed. The performance of each configuration was evaluated using mIoU, F1 score, and accuracy. The ablation results for the vegetation dataset, which features more complex backgrounds, are presented in Table 3.

Table 3. Comparison of ablation study results.

CBAM-ASPP	Residual Block	Hybrid Loss	Fixed-Ratio Hybrid Loss	mIoU/%	F1 Score/%	Precision/%	Parameters/M
×	×	×	×	83.22	84.84	89.70	51.11
√	×	×	×	84.01	85.16	89.91	56.79
×	√	×	×	83.05	84.67	91.11	43.65
×	×	√	×	84.05	85.68	90.59	51.11
×	×	×	√	80.21	85.38	90.81	51.11
√	√	×	×	83.68	85.07	90.88	44.54
×	√	√	×	83.87	85.10	91.07	43.65
√	√	√	×	84.70	86.12	91.28	44.54

Note: "×" indicates that the module is not used, and "√" indicates that the module is used.

As shown by the ablation results in Table 3, individually introducing the CBAM-ASPP module, the residual attention mechanism, and the dynamically weighted hybrid loss function led to varying degrees of improvement in the model's mIoU, F1 score, and accuracy. Each module demonstrated a significant optimization effect for addressing specific challenges.

When only the CBAM-ASPP module was introduced, the model exhibited the most significant improvement in mIoU, reaching 84.01%, while the F1 score increased to 85.16%. At the same time, the parameter count was reduced compared with using the standard ASPP module alone. This indicates that the CBAM-ASPP module effectively addresses the trade-off between accuracy improvement and computational complexity in traditional multi-scale feature extraction methods, balancing feature extraction efficiency and effectiveness.

When only the residual attention mechanism was introduced, the model exhibited the most significant improvement in accuracy, reaching 91.11%, second only to the final integrated model. Although the mIoU (83.05%) and F1 score (84.67%) were slightly lower than those of the baseline UNet model, this result still demonstrates that the mechanism

can accurately focus on key target region features. It effectively addresses the traditional decoder's limitations in recognizing critical regions, making it particularly suitable for segmentation tasks that require high pixel-level classification accuracy.

When only the dynamically weighted hybrid loss function was applied, the model achieved simultaneous improvements in the three core metrics: mIoU, F1 score, and accuracy, reaching 84.05%, 85.68%, and 90.59%, respectively. In comparison, the fixed-proportion hybrid loss strategy, which assigns equal weights to different loss terms, only improved the F1 score (85.38%) and accuracy (90.81%), while mIoU dropped to 80.21%. This indicates that the fixed-weight scheme may induce false detections or omission errors when merely optimizing overlapping regions. By contrast, the dynamically weighted hybrid loss outperformed the baseline model across all three evaluation metrics. It effectively addresses the segmentation bias caused by single-objective optimization in conventional loss functions. Furthermore, the proposed strategy dynamically balances multiple loss objectives throughout the training process, enabling precise optimization of boundary segmentation performance without increasing the model's parameter count.

When any two of the three modules were combined, the model's segmentation metrics exceeded those achieved by using a single module alone, further demonstrating the complementary optimization effects between different modules [32]. When the CBAM-ASPP module, residual attention mechanism, and dynamically weighted hybrid loss function were all integrated, the model achieved optimal performance with mIoU of 84.70%, F1 score of 86.12%, and accuracy of 91.28%, while maintaining a relatively low parameter count of 44.54 M. Compared with using any single module, the three core metrics improved on average by more than 1.3 percentage points; compared with the baseline UNet model, the improvement was even more pronounced; and compared with any two-module combination, the average improvement exceeded 0.8 percentage points.

These results indicate that the three modules can generate a synergistic effect, not only addressing specific bottlenecks in the segmentation task individually, but also collectively enhancing the model's feature extraction capability, key-region recognition ability, and loss optimization accuracy. This synergy ultimately achieves an effective balance between segmentation precision and recognition efficiency, significantly improving the model's overall segmentation performance.

The proposed model addresses this issue effectively due to several factors:

1. The model's skip connection structure ensures that the final reconstructed feature maps retain lower-level features [33].

2. The CBAM-ASPP module captures multi-scale feature information.

3. The residual attention mechanism improves parameter efficiency and reduces model complexity.

The hybrid loss function enhances the model's ability to capture boundary information.

Experimental results demonstrate that the proposed method achieves superior segmentation performance and offers greater effectiveness in vegetation and snow cover identification.

4. Construction of Grassland Fire Risk Index (GFRI)

4.1. Fire Risk Factors

To validate the practical application value of the vegetation and snow cover segmentation results in this study, we adopt the multi-factor coupling framework of existing fire dynamic risk indices (FDDI [34] and FFDI [35]). Combined with the formation mechanism of grassland fire hazards and the underlying surface characteristics of the study area, four core influencing factors are selected: vegetation coverage, green chromatic coordinate (GCC), snow coverage, and elevation. Using observational data from a large number of near-surface fixed grassland fire risk monitoring stations across the region [36], this study

develops a Grassland Fire Risk Index (GFRI) tailored for fine-scale field grassland fire risk assessment. All fire risk factors are derived entirely from the image segmentation results and basic field station information obtained in this study.

Among them, vegetation coverage (VC) is defined as the ratio of the number of pixels in segmented vegetation areas to the total number of pixels in the entire image. It characterizes the coverage degree and load level of regional combustible materials. Higher vegetation coverage corresponds to better continuity of combustibles and greater fire hazard potential, making it a positive fire risk factor.

The green chromatic coordinate (GCC) [15] is calculated solely from vegetation regions extracted by the improved algorithm. Invalid pixels are excluded to accurately reflect vegetation growth status and indirectly characterize vegetation moisture content to a certain extent. The GCC is defined as:

$$GCC = \frac{G}{R + G + B} \tag{12}$$

In Equation (12), R, G, and B represent the digital number values of red, green, and blue bands in visible light imagery, respectively. Lower GCC values indicate higher vegetation withering degree and greater flammability, rendering GCC a positive fire risk factor.

Snow coverage (SC) is defined as the ratio of snow pixels obtained from segmentation to the total pixels of the entire image. By effectively covering combustible materials, reducing surface temperature, and alleviating vegetation drying, snow coverage markedly restrains the ignition and spread of grassland fires, thus acting as a negative fire risk factor.

Elevation (E) refers to the vertical distance from a given point to the absolute datum along the plumb line. As one of the key external environmental factors affecting plant growth [37], variations in elevation influence vegetation growth and development, functional structure, as well as leaf area and stomatal density. Therefore, elevation exerts a substantial impact on vegetation growth. In addition, meteorological and climatic characteristics change with increasing altitude. Generally, higher elevations are associated with greater relative humidity and higher vegetation moisture content, which reduces flammability. For this reason, elevation serves as a negative fire risk factor.

To eliminate dimensional differences among fire risk factors and unify their contribution directions to grassland fire hazard, all indicators are processed using range standardization. All values are normalized uniformly to the interval of [0, 1]. After standardization, a higher value indicates a stronger promoting effect of the corresponding factor on grassland fire risk [38]. However, different fire risk factors adopt distinct normalization methods. Positive fire risk factors are normalized using Equation (13), while negative fire risk factors are processed via Equation (14). Here, x_{min} denotes the minimum value of the factor, and x_{max} represents the maximum value.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{13}$$

$$x' = \frac{x_{min} - x}{x_{max} - x_{min}} \tag{14}$$

4.2. Dynamic Grassland Fire Risk Index Based on AHP-Improved Weighting

To reasonably determine the weight of each fire risk factor, avoid the limitation of equal weighting that ignores the differentiated contributions of individual factors to fire hazard, and clarify the intrinsic relationship between fire risk factors and grassland fire occurrence, this study adopts the Analytic Hierarchy Process (AHP) for weight calculation, so as to improve the accuracy of conventional fire early warning models. The Analytic Hierarchy Process (AHP) is a systematic mathematical decision-making method that addresses

complex multi-factor decision problems. It decomposes multiple influencing factors into hierarchical levels, including target, criterion, and alternative layers. By combining qualitative indicators with quantitative calculation, AHP derives the weight sequence of factors at each level, thereby providing a rational basis for scheme optimization and scientific decision-making [39]. The weight calculation process of AHP consists of the following three steps [40].

First, combined with existing research on forest and grassland fire hazards and local fire occurrence patterns in the study area, vegetation coverage and snow cover are regarded as the dominant surface conditions determining grassland fire risk, followed by vegetation greenness. Elevation is treated as a relatively stable background factor with the weakest influence. On this basis, the judgment matrix $A_{4\times4}$ is established.

Second, the maximum eigenvalue is obtained using the sum-product method, with the specific steps as follows [41]:

The judgment matrix $A_{4\times4}$ is normalized according to the following formula to obtain the matrix $B_{4\times4}$:

$$b_{ij} = \frac{a_{ij}}{\sum_{i=1}^{4} a_{ij}} \ (i, j = 1, 2, 3, 4) \tag{15}$$

The vector $C = (c_1, c_2, c_3, c_4)^T$ is derived from the matrix $B_{4\times4}$.

$$c_i = \sum_{j=1}^{4} b_{ij} \ (i, j = 1, 2, 3, 4) \tag{16}$$

The vector C is then normalized to obtain the eigenvector $W = (w_1, w_2, w_3, w_4)^T$:

$$w_i = \frac{c_i}{\sum_{i=1}^{4} c_i} \ (i = 1, 2, 3, 4) \tag{17}$$

Finally, the maximum eigenvalue λ_{max} is calculated as:

$$\lambda_{max} = \frac{1}{4} \times \sum_{i=1}^{4} \frac{(Aw)_i}{w_i} \ (i = 1, 2, 3, 4) \tag{18}$$

Third, a consistency test is performed on the judgment matrix to determine whether the calculated weights are reasonable as a decisive indicator. The criterion is: the weights are considered acceptable when the consistency ratio $CR < 0.1$. The definition of CR is:

$$CR = \frac{CI}{RI} \tag{19}$$

In Equation (19), RI represents the random consistency index, which varies with the order of the matrix. In a 4th-order square matrix, its value is 0.89.

Consistency index (CI) is given by:

$$CI = \frac{\lambda_{max} - 4}{4 - 1} = \frac{\lambda_{max} - 4}{3} \tag{20}$$

Through matrix normalization, weight calculation, and consistency testing, the following results are obtained: the maximum eigenvalue $\lambda_{max} \approx 4.01$, the consistency index $CI \approx 0.003$, and the consistency ratio $CR \approx 0.003 < 0.1$. The matrix passes the consistency test, indicating that the weight distribution is reasonable and valid. The final weights for each factor are: $w_{gcc} = 0.205$, $w_{VC} = 0.351$, $w_{SC} = 0.351$, and $w_E = 0.093$.

Drawing on the multiplicative coupling form of the classic FDDI index, and incorporating the mechanism-based weights derived from AHP, the Grassland Fire Risk Index (GFRI) is constructed as follows:

$$GFRI = \left(GCC'\right)^{0.205} \times \left(VC'\right)^{0.351} \times \left(SC'\right)^{0.351} \times \left(E'\right)^{0.093} \quad (21)$$

4.3. Correlation Analysis

To verify the effectiveness of vegetation coverage, snow coverage, and green chromatic coordinate (GCC) derived from near-surface image segmentation in grassland fire risk assessment, and further examine the capability of the constructed GFRI in characterizing the seasonal dynamics of regional fire risk, Pearson correlation analysis is conducted using historical fire records from 2002 to 2018 (Figure 10) [42] and monthly fire risk index values. Forest and grassland fires in the Altay region are jointly controlled by seasonal snowmelt, vegetation phenological succession, and the periodic strong winds in spring and autumn. The intra-annual fire distribution presents a stable seasonal pattern: fires occur most frequently in spring, followed by autumn, with low incidence in summer and rare occurrences in winter. Interannual minor climatic fluctuations only affect the total number of fire incidents without altering the inherent monthly fire distribution pattern of the region.

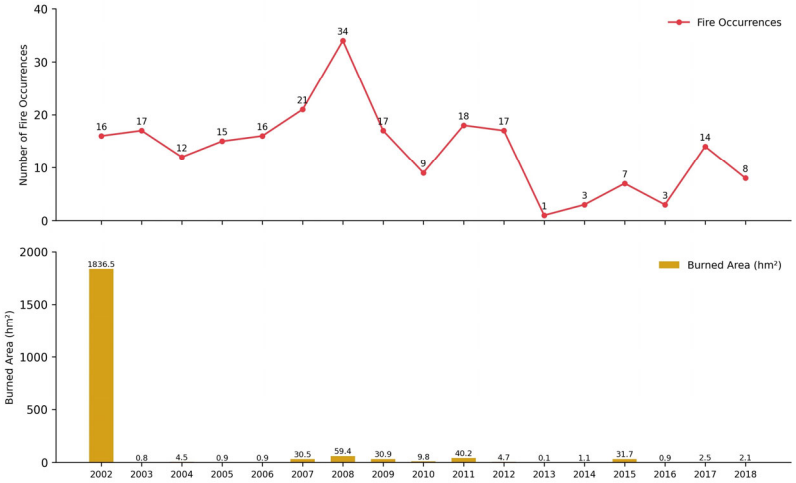


Figure 10. Historical fire incidents in Altay Prefecture, Xinjiang, from 2002 to 2018.

Based on valid monthly image data from all near-surface monitoring stations in the Altay region for 2025, this study calculated the vegetation coverage, snow coverage, and green chromatic coordinate (GCC) for each image. Representative monthly values of these vegetation cover characteristics were obtained by arithmetic averaging. Combined with the fixed elevation data of each station, all indicators were standardized using range normalization and then substituted into the GFRI model to derive the monthly dynamic grassland fire risk index for the Altay region from January to December 2025.

Combined with the historical statistical data of the Altay region, the monthly historical fire occurrence counts and corresponding GFRI index values within the year in the study area are summarized in Table 4.

Table 4. Historical statistical data of the Altay region.

Month	Cumulative Number of Historical Fire Incidents (2002–2018)	Monthly GFRI Index (2025)
January	0	0.13
February	0	0.16
March	1	0.29

Table 4. *Cont.*

Month	Cumulative Number of Historical Fire Incidents (2002–2018)	Monthly GFRI Index (2025)
April	37	0.70
May	22	0.65
June	29	0.74
July	58	0.82
August	47	0.79
September	25	0.67
October	6	0.38
November	1	0.20
December	0	0.11

The Pearson correlation coefficient is used to quantify the linear correlation between the monthly GFRI index and historical fire occurrence frequency, with the formula given as follows:

$$r = \frac{\sum_i^n (x_i - \overline{x}) \times (y_i - \overline{y})}{\sqrt{\sum_i^n (x_i - \overline{x})^2} \times \sqrt{\sum_i^n (y_i - \overline{y})^2}} \quad (22)$$

In Equation (22), r is the Pearson correlation coefficient, with a value range of $[-1, 1]$; n is the sample size, which is taken as 12 months within the year in this study ($n = 12$); x_i is the GFRI fire risk index in the i-th month; $\overline{x}$ is the average value of the GFRI index for 12 months of the whole year; y_i is the cumulative number of historical fires in the i-th month from 2002 to 2018; $\overline{y}$ is the monthly average number of historical fires for the whole year.

Pearson correlation coefficient $r > 0$ indicates a positive correlation between the two variables; the higher the GFRI index, the more the monthly fires, which is consistent with the fire mechanism. The closer r is to 1, the better the representation effect of the index on the seasonal rhythm of regional fire risk. The significance test adopted a confidence level of 95% ($p < 0.05$) as the criterion for significance judgment.

Pearson correlation analysis of the 12 monthly samples shows that the GFRI index is extremely significantly positively correlated with the historical monthly fire occurrence counts ($r = 0.924$, $p < 0.001$). This indicates that the fire risk index constructed based on vegetation and snow cover derived from near-surface image segmentation can effectively capture the intra-annual seasonal dynamics of grassland fire risk in the Altay region. It exhibits a high degree of coupling with the regional fire occurrence rhythm, and can well characterize the seasonal variation characteristics of grassland fire risk in the study area.

Combined with the historical fire records of the Altay region, it can be found that the GFRI index can effectively capture the seasonal driving mechanism of grassland fire risk in this area, and its overall variation trend is highly consistent with the rhythm of historical fire occurrence.

In winter (December–February), the entire study area is covered by snow, and combustible materials are completely suppressed with no fire occurrences; accordingly, the fire risk index remains at its lowest level throughout the year. In March, snow melts gradually, a small amount of withered grass is exposed on the surface, sporadic fires emerge, and fire risk rises slightly. In April, extensive snow retreat exposes abundant surface combustibles, causing fire risk to increase rapidly and fire occurrences to grow remarkably. During summer (June–August), the region features high temperatures and dry air. Although vegetation remains green, prolonged high temperatures and drought reduce the surface moisture content of vegetation and enhance its flammability [43]. Coupled with persistent

dry conditions, this period forms the annual peak of fire activity, corresponding to the highest GFRI index of the whole year. In September, temperature drops and humidity rises, vegetation gradually withers and turns yellow, and both fire occurrences and fire risk decline synchronously [44]. From October to December, temperatures continue to decrease and snow gradually covers the land surface, leading fire activities to fade rapidly until they disappear completely. Meanwhile, slight deviations between the fire risk index and fire frequency exist in certain months. This is because grassland fire risk is governed not only by surface cover factors but also interfered with by stochastic factors such as instantaneous meteorological conditions and human activities, which conforms to the actual occurrence pattern of regional fires.

This result fully verifies that the vegetation coverage, snow coverage, and GCC retrieved by the improved image segmentation algorithm in this paper can accurately respond to the seasonal variations in regional surface cover and climate. The indicators can effectively characterize the background state of regional grassland fire risk and possess favorable practical application potential in field scenarios.

5. Discussion

The proposed multi-scale residual attention network demonstrates outstanding performance in image segmentation tasks, achieving an effective balance between accuracy and computational efficiency. Its advantages primarily stem from three core modules and the design of a hybrid loss function, which work synergistically to accurately capture critical target information in complex backgrounds while maintaining low computational overhead, which provides high-precision, reliable basic data support for the subsequent construction of the Grassland Fire Risk Index (GFRI).

Second, the residual attention module optimizes the fusion strategy of high-level semantic features and low-level detailed features during the upsampling stage. By embedding the CBAM attention mechanism into residual units, the proposed network strengthens the extraction of effective foreground information, such as vegetation and snow cover, while effectively suppressing irrelevant background interference from terrain and bare land. This design not only improves the model's multi-level feature perception ability but also reduces redundant parameters and computational cost. It notably optimizes the edge clarity and structural integrity of segmentation results, ensuring the high-precision acquisition of vegetation coverage (VC) and snow coverage (SC) datasets required for GFRI modeling.

Additionally, the dynamically weighted hybrid loss function further improves the model's comprehensive segmentation performance. Different from conventional fixed-proportion hybrid loss functions that suffer from unbalanced optimization objectives, the proposed strategy adaptively adjusts the weight of each loss term at different training stages. The Dice loss constrains the overall structural integrity of segmentation results and alleviates the class imbalance problem between vegetation, snow and background pixels; the boundary loss adopts distance transform weighting to penalize edge deviation errors, which effectively improves the identification accuracy of fuzzy boundary regions between vegetation, snow cover and bare land. Comparative experiments verify that the dynamic weighted loss scheme outperforms the fixed-proportion loss strategy across all core metrics. It avoids the mIoU decline and omission/false detection risks caused by fixed-weight optimization, and achieves synchronous optimization of overall segmentation accuracy and boundary precision without increasing model parameters.

More importantly, the high-quality vegetation and snow cover segmentation results obtained by the proposed network lay a solid foundation for the construction and application of the GFRI model. The accurate time-series VC, SC, and green chromatic coordinate

(GCC) data retrieved from segmented images, combined with stable topographic elevation factors, enable reliable monthly grassland fire risk quantification. The Pearson correlation verification further proves that the segmentation products generated by the proposed model can accurately respond to seasonal variations in surface cover and climate, and the derived GFRI index is highly coupled with the actual seasonal fire occurrence rhythm in the study area.

In summary, the proposed segmentation method effectively balances segmentation accuracy, boundary refinement capability and computational efficiency. The synergistic effect of multi-scale feature fusion, residual attention optimization and dynamic hybrid loss function ensures the reliability of ground cover parameter inversion. This not only validates the comprehensive effectiveness of the designed network modules in complex grassland image segmentation, but also provides a feasible, efficient and refined technical support for regional grassland fire risk dynamic assessment, realizing the organic combination of intelligent image segmentation and ecological fire risk evaluation.

6. Conclusions

Based on the UNet architecture, this study proposes an improved image segmentation model. The model replaces the original multi-scale feature fusion module with a CBAM-ASPP module, introduces a residual attention mechanism in the upsampling stage, and adopts a dynamically weighted hybrid loss function as a collaborative optimization strategy. These improvements significantly enhance the model's segmentation accuracy and generalization capability. The main conclusions are as follows:

1. The improved UNet model demonstrates excellent segmentation performance on both vegetation and snow datasets. Specifically, for the vegetation segmentation task, the model achieved an accuracy of 91.28%, representing an improvement of 1.58 percentage points over the standard UNet. For the snow segmentation task, the accuracy reached 94.19%, an increase of 2.36 percentage points compared with UNet. These results fully demonstrate the feasibility of reducing the parameters and improving the accuracy of the UNet backbone network, and provide a reliable model foundation for efficiently obtaining the environmental background data required for grassland fire risk assessment.

2. Under the same experimental conditions, the improved UNet model outperforms mainstream segmentation models such as FCN, SegNet, and DeepLabv3+ in both vegetation and snow segmentation tasks. For vegetation segmentation, the accuracy improved by 11.15, 5.54, and 2.18 percentage points, respectively, compared with the three baseline models. For snow segmentation, the accuracy increased by 9.35, 3.69, and 2.33 percentage points, respectively. Leveraging the advantages of multi-module collaborative optimization, the improved model effectively addresses the limitations of traditional models and enhances segmentation accuracy in complex field scenarios.

3. The ablation study further validates the effectiveness and synergistic benefits of each improved module. The final model, integrating all proposed enhancements, outperforms the standard UNet across the three core metrics: mIoU, F1 score, and accuracy, with improvements of 1.48, 1.28, and 1.58 percentage points, respectively. At the same time, the model's parameter count is reduced by 12.85%, achieving a balance between segmentation accuracy and computational efficiency, and laying the foundation for deployment on resource-constrained edge devices.

4. The improved model enables consistent and stable extraction of time-series vegetation coverage, snow coverage, and GCC from near-surface monitoring images, providing unified and reliable basic parameters for regional fire risk evaluation. On this basis, the constructed GFRI index integrates mature fire risk assessment frameworks and actual surface environmental characteristics, forming a mechanism-driven fire risk quantification model

with robust result reliability and scientific rationality. Correlation verification confirms that the GFRI index can effectively characterize the seasonal dynamic rhythm of grassland fire risks and match the actual regional fire occurrence rules.

Although the proposed improved model achieves notable progress in vegetation and snow cover identification, further research is needed in the following areas. First, the current segmentation models are designed for single tasks. In the future, by constructing a unified multi-task segmentation model, we can improve efficiency and performance. Meanwhile, future optimization can further improve the GFRI's universality and accuracy by expanding multi-year, multi-vegetation-type monitoring datasets and combining field-measured combustible moisture data. Second, the model relies heavily on high-quality annotated data, highlighting the need to explore semi-supervised learning approaches to reduce annotation costs. Additionally, combining techniques such as model quantization and pruning could enhance inference speed, enabling real-time deployment and providing more precise technical support for sustainable grassland management and fire risk early warning.

Author Contributions: Methodology, S.C. and Y.Y.; formal analysis, W.L., S.C. and Y.Y.; investigation S.C. and Y.Y.; writing—original draft preparation, W.L., S.C. and Y.Y.; writing—review and editing Y.Z. and S.C.; supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No: 32471858) and the Beijing Forestry University "5.5 Project" Scientific Research and Innovation Team Project (No: BLRC2023C05).

Data Availability Statement: All data and materials will be available on request. The experiment conducted at this location was approved by the relevant government authorities. If anyone wishes to request the data from this study, they should contact the first author.

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

References

1. Li, L. Research on Forest and Grassland Fire Warning. *China For. By-Prod.* **2026**, 88–90. [CrossRef]
2. Wu, R.; Hong, Z.; Du, W.; Shan, Y.; Ying, H.; Wu, R.; Gantumur, B. A generalized spatiotemporally weighted boosted regression to predict the occurrence of grassland fires in the Mongolian Plateau. *Remote Sens.* **2025**, *17*, 1485. [CrossRef]
3. Guo, G.Z.; Liao, H.Q.; Sun, N. Multi-time scale features of fire weather in two major forests in China during 1961–2020. *Disaster Reduct. China* **2022**, *3*, 36–39.
4. Wang, X. Study on the Effects of Snow Characteristics on Wildfire Occurrence in the Mid- and High-Latitudes of the Northern Hemisphere. Ph.D. Thesis, Northeast Normal University, Changchun, China, 2024.
5. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
6. Khanal, S.; Kc, K.; Fulton, J.P.; Shearer, S.; Ozkan, E. Remote sensing in agriculture—Accomplishments, limitations, and opportunities. *Remote Sens.* **2020**, *12*, 3783. [CrossRef]
7. Hu, T.; Zhao, D.; Zeng, Y.; Guo, Q.; He, H. Advances in multi-source data fusion for ecosystem assessment. *Acta Ecol. Sin.* **2023**, *43*, 542–553. [CrossRef]
8. Donager, J.; Sankey, T.T.; Meador, A.J.S.; Sankey, J.B.; Springer, A. Integrating airborne and mobile lidar data with UAV photogrammetry for rapid assessment of changing forest snow depth and cover. *Sci. Remote Sens.* **2021**, *4*, 100029. [CrossRef]
9. Ersoz, B.A.; Pekcan, O. UAV-based automated earthwork progress monitoring using deep learning with image inpainting. *Autom. Constr.* **2025**, *175*, 106211. [CrossRef]
10. Furukawa, F.; Laneng, L.A.; Ando, H.; Yoshimura, N.; Kaneko, M.; Morimoto, J. Comparison of RGB and multispectral unmanned aerial vehicle for monitoring vegetation coverage changes on a landslide area. *Drones* **2021**, *5*, 97. [CrossRef]
11. Zhu, Z.; Zhao, F.; Wang, Q.H.; Gao, Z.L.; Deng, X.F.; Huang, P.G. Driving factors of forest fire occurrence and fire risk zoning in Gansu Province. *Sci. Silvae Sin.* **2025**, *61*, 63–71.

12. Zhang, H.; Tang, Z.; Wang, B.; Meng, B.; Qin, Y.; Sun, Y.; Lv, Y.; Zhang, J.; Yi, S. A non-destructive method for rapid acquisition of grassland aboveground biomass for satellite ground verification using UAV RGB images. *Glob. Ecol. Conserv.* **2022**, *33*, E01999. [CrossRef]
13. Ide, R.; Oguma, H. A cost-effective monitoring method using digital time-lapse cameras for detecting temporal and spatial variations of snowmelt and vegetation phenology in alpine ecosystems. *Ecol. Inform.* **2013**, *16*, 25–34. [CrossRef]
14. Fu, S.; Feng, Q.S.; Dang, J.Y.; Lei, K.X.; Qiao, W.X.; Liang, T.G.; Pan, D.R.; Sun, B.; Jiang, J.C. Assessing alpine grassland vegetation coverage based on UAV and Sentinel-2 data. *Pratacultural Sci.* **2022**, *39*, 455–464.
15. Zheng, Y.L.; Tian, Z.; Guan, P.; Shi, H. Development of an intelligent monitoring system for vegetation coverage and phenology in grassland. *Trans. Chin. Soc. Agric. Eng. (Trans. CASE)* **2023**, *39*, 162–171.
16. Wang, Y.; Wan, H.; Gao, J.; Sun, H.; Hu, Z.; Zhang, Z. Comparative study of grassland vegetation coverage estimation based on deep learning semantic segmentation models. *Res. Environ. Sci.* **2024**, *37*, 2299–2309.
17. Mahendra, H.N.; Mallikarjunaswamy, S.; Subramoniam, S.R. An assessment of vegetation cover of Mysuru City, Karnataka State, India, using deep convolutional neural networks. *Environ. Monit. Assess.* **2023**, *195*, 526. [CrossRef]
18. Wang, S.; Cui, D.; Wang, L.; Peng, J. Applying deep-learning enhanced fusion methods for improved NDVI reconstruction and long-term vegetation cover study: A case of the Danjiang River Basin. *Ecol. Indic.* **2023**, *155*, 111088. [CrossRef]
19. Luo, J.; Dong, C.; Lin, K.; Chen, X.; Zhao, L.; Menzel, L. Mapping snow cover in forests using optical remote sensing, machine learning and time-lapse photography. *Remote Sens. Environ.* **2022**, *275*, 113017. [CrossRef]
20. Kalinaki, K.; Malik, O.A.; Lai, D.T.C.; Sukri, R.S.; Wahab, R.B.H.A. Spatial-temporal mapping of forest vegetation cover changes along highways in Brunei using deep learning techniques and Sentinel-2 images. *Ecol. Inform.* **2023**, *77*, 102193. [CrossRef]
21. Wu, J.; Meng, Q.; Gao, L.; Zhang, L.; Zhao, M.; Su, C. A deep learning framework for 3D vegetation extraction in complex urban environments. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103798. [CrossRef]
22. Zhang, L.; Wang, W.; Dong, Y.; Wu, L. Vegetation coverage monitoring model design based on deep learning. *Sci. Program.* **2022**, *2022*, 4818985. [CrossRef]
23. Qi, X.; Du, H.B.; Zhu, L.J. Image semantic segmentation algorithm based on ASPP module and weighted fusion of feature maps. *Comput. Appl. Softw.* **2025**, *42*, 227–232+267.
24. Cao, X.Y.; Zhang, T.H.; Zhao, Y.J. Improved ASPP and semantic segmentation algorithm with fused multi-scale features. *Comput. Simul.* **2025**, *42*, 208–215, 476.
25. Yiyang, L.; Jiayi, Z.H.A.O.; Ruoyi, Y.U.; Huixiang, L.I.U.; Yu, G.U. Colon polyp detection based on multi-scale and multi-level feature fusion and lightweight convolutional neural network. *J. Biomed. Eng.* **2024**, *41*, 911–918.
26. Wang, W.W.; Luo, Y.; Li, X. Semantic segmentation method integrating attention mechanism and multi-level features. *Ind. Control. Comput.* **2025**, *38*, 93–95.
27. Rezvani, S.; Fateh, M.; Jalali, Y.; Fateh, A. FusionLungNet: Multi-scale fusion convolution with refinement network for lung CT image segmentation. *Biomed. Signal Process. Control.* **2025**, *107*, 107858. [CrossRef]
28. Qi, H.Z.; Wu, J.B.; Guo, R.Q. Semantic Segmentation method of lawn area based on improved DeepLabv3+. *Autom. Instrum.* **2025**, *40*, 132–136, 141.
29. Chang, M.H.; Li, S.H.; Peng, S.F.; Liu, Z.; Zhao, T.; Zhang, Y.; Lin, J. Multi-feature progressive fusion for cropland extraction in complex scenes. *Trans. Chin. Soc. Agric. Eng.* **2025**, *41*, 294–303.
30. Zhang, Y.; Wang, T.; You, Y.; Wang, D.; Lu, M.; Wang, H. A novel estimation method of grassland fractional vegetation cover based on multi-sensor data fusion. *Comput. Electron. Agric.* **2024**, *225*, 109310. [CrossRef]
31. Araji, T.M.; Waqas, A.; Ali, R. Utilizing deep learning towards real-time snow cover detection and energy loss estimation for solar modules. *Appl. Energy* **2024**, *375*, 124201. [CrossRef]
32. Liu, X.; Tian, M.; Liang, J.Y. Prediction of cotton yield densely planted in Xinjiang of China using RCH-UNet model. *Trans. Chin. Soc. Agric. Eng. (Trans. CASE)* **2024**, *40*, 230–239.
33. Wang, Z.; Ma, W.; Xiang, B.; Yang, L.; Shen, X.; Liu, J. Locating branch bifurcation points to select the pruning position of walnut trees using improved UNet model. *Trans. Chin. Soc. Agric. Eng. (Trans. CASE)* **2025**, *41*, 165–172.
34. Li, X.L. Research on Forest Fire Prediction and Monitoring Based on Multi-factor Synergistic Effects Using MODIS Data. Ph.D. Thesis, University of Science and Technology of China, Heifei, China, 2016.
35. *GB/T 36743-2018*; National Technical Committee for Meteorological Disaster Prevention and Mitigation Standards. Forest Fire Danger Meteorological Level. China Standards Press: Beijing, China, 2018.
36. Ren, S.J.; Deng, J.; Wang, J.R.; Wang, C.P.; Xiao, Y.; Kang, F.R.; Li, H.J.; Dou, L.H. Porosity inversion and determination in gob unconsolidated materials based on acoustic signal response. *Fuel* **2026**, *426*, 139614. [CrossRef]
37. Liu, W.; Zheng, L.; Qi, D. Variation in leaf traits at different altitudes reflects the adaptive strategy of plants to environmental changes. *Ecol. Evol.* **2020**, *10*, 8166–8175. [CrossRef]
38. Dang, L.B.; Jia, X.Y. Design of Coalfield Fire Monitoring and Control System in Xinjiang. *West. Resour.* **2017**, *5*, 188–190.

39. Lamat, R.; Kumar, M.; Kundu, A.; Lal, D. Forest fire risk map using analytical hierarchy process (AHP) and earth observation datasets: A case study in the mountainous terrain of Northeast India. *SN Appl. Sci.* **2021**, *3*, 425. [CrossRef]
40. Tolche, A.D.; Gurara, M.A.; Pham, Q.B.; Anh, D.T. Modelling and accessing land degradation vulnerability using remote sensing techniques and the analytical hierarchy process approach. *Geocarto Int.* **2022**, *37*, 7122–7142. [CrossRef]
41. Yang, J.H. Research on Forest Fire Risk Warning in Liaoning Province Based on MODIS. Master's Thesis, Liaoning University of Technology, Jinzhou, China, 2024.
42. Wang, J.; Wang, Q.H.; Liang, Y.; Zhang, W.W.; Long, T.T.; Gao, Z.L.; Wang, R.C. Study on Spatial and Temporal Distribution of Forest and Grassland Fires in Altay Region of Xinjiang from 2002 to 2018. *Shandong For. Sci. Technol.* **2021**, *51*, 47–52.
43. Zhou, N.Y.; Li, X.X.; Zhao, F.J.; Si, L.Q.; Wang, M.Y.; Li, W.K.; Yan, K.D.; Li, W. Study on Fire Behavior Simulation of the Extra-large Grassland Fire in Chabaertu Banner Based on BehavePlus. *Grassl. Sci.* **2025**, *42*, 2392–2402.
44. Xu, W.; Zhou, Q.; Ma, W.; Liu, F.; Li, L. Research on Grassland Fire Prevention Capabilities and Influencing Factors in Qinghai Province, China. *Earth* **2025**, *6*, 101. [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.