SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería RIMNI

In cooperation with CIMNE

The Segmentation of Debris Flow Fans Based on Spatial Coordinate Attention Mechanism

Xin Song and Baoyun Wang*

School of Mathematics, Yunnan Normal University, Kunming, 650500, China

*Corresponding Author: Baoyun Wang. Email: bywang@ynnu.edu.cn

Received: 24 July 2024 Accepted: 27 September 2024

ABSTRACT

In response to the low accuracy and poor performance of traditional machine learning methods in identifying debris flow fans. This paper proposes an optimized Simple, Parameter-Free Attention Module (SimAM) attention mechanism named Spatial Coordinate Attention Module. It combines with convolutional neural networks to achieve precise segmentation of debris flow fans. Firstly, the energy function of the SimAm is improved to retain the spatial coordinate information of features. Secondly, the closed-form solution of the module is obtained through optimization theory to ensure lightweightness, resulting in the Spatial Coordinate Attention Module. Finally, the Spatial Coordinate Attention Module is embedded into classic segmentation network models to compare with mainstream attention mechanisms. Experimental results demonstrate that the proposed method outperforms mainstream attention mechanisms in various classic models, yielding more complete segmentation results. This approach effectively enhances the segmentation performance of the network models in the task of debris flow fans segmentation.

OPEN ACCESS

Accepted: 27/09/2024

Submited: 24/07/2024

DOI
110.23967/
j.rimni.2024.10.56508

Keywords: Semantic segmentation; convolutional neural network; debris flow fans; SimAM

1 Introduction

Debris flow fans are not only direct products of debris flow activities but also key indicators for identifying debris flow gullies, studying the development history of debris flows, and understanding geomorphological evolution. Additionally, they are crucial for assessing the potential hazard range of debris flows. Furthermore, debris flow fans are closely related to national economic development. With the growth of population and expansion of spatial needs, the relationship between humans and land has become increasingly strained, making debris flow fans important areas for local development and utilization. Surrounding debris flow fans are often residential areas, factories, major transportation routes, or fertile farmland. In land planning and construction in mountainous regions, accurately identifying the fan-shaped alluvial landforms formed at the mouths of gullies is essential.

In the semantic segmentation task of debris flow fans, the challenge lies in the close connection of the fan's boundary with the surrounding land, leading to high similarity in appearance and environment, resulting in weak boundary distinctiveness. Traditional methods for debris flow fans

identification heavily rely on visual recognition, which is inefficient and time-consuming. With the rapid development of digital image processing and machine learning technologies, some researchers have employed Support Vector Machine (SVM) classification [1] and pixel-based classification methods [2] to identify debris flow fans in remote sensing images. However, these methods still face issues such as incomplete segmentation, excessive breakpoints, and noise points. In comparison to the mentioned methods, the rapid progress of deep learning technology provides a new solution to the segmentation problem of debris flow fans. The application of Convolutional Neural Networks (CNNs) [3–7] has greatly addressed challenges in segmentation, such as difficulty in distinguishing boundaries and excessive noise.

In the research on the application of convolutional neural networks for image segmentation, various network structures have emerged, such as Fully Convolutional Networks (FCN) [8], U-net architecture (UNet) [9], and Pyramid Scene Parsing Network (PSPNet) [10]. FCN, proposed by Long and Darrell in 2015, achieved pixel-level image classification, thereby solving semantic-level image segmentation tasks, and can be considered a foundational work in the field of deep learning for semantic segmentation. In 2015, Ronneberger et al. introduced UNet to address medical image segmentation problems, and due to UNet's outstanding performance, it later found widespread applications in various scenarios. In 2017, Zhao et al. proposed PSPNet, which used a pyramid pooling module to build a scene parsing network. By fusing features at multiple scales and aggregating context information from different regions, the model gained the ability to understand global contextual information. These studies indicate that excellent convolutional neural network structures can significantly improve the performance of image segmentation tasks. Therefore, designing well-crafted convolutional neural network structures is crucial. However, when the above-mentioned network architectures are used for debris flow fans segmentation tasks, issues still persist, such as inaccurate localization and incomplete segmentation. Constructing a more suitable network architecture is highly complex, and therefore, it is necessary to introduce attention mechanism modules [11–15] to help the network focus on key information.

Human visual attention, as a crucial selection mechanism, contributes to allocating limited information processing capacity by focusing on information relevant to the target task while attenuating the interference of irrelevant information. Inspired by visual attention mechanisms, researchers have designed similar attention modules in convolutional neural networks, including the Squeeze-and-Excitation Networks (SE) [16], Convolutional Block Attention Module (CBAM) [17], Effcient Channel Attention (ECA) [18], Coordinate Attention (CA) [19], and Simple, Parameter-Free Attention Module (SimAM) [20]. The SE module enhances the model's sensitivity to channel features by compressing and activating them. The CBAM module strengthens the model's sensitivity to both channel and spatial features. The ECA module, proposed on the foundation of the SE module, acts as a local cross-channel interaction strategy, enhancing the model's information extraction capability on channel features. The CA module enables the network to accurately capture spatial position information, increasing the model's focus on spatially important features. However, the aforementioned attention mechanism modules have limitations as they can only refine features in the channel or spatial dimension, thereby restricting the flexibility to learn attention weights for cross-channel and spatial transformations. To address this issue, Yang et al. proposed the SimAM attention mechanism module based on mature neural science theories. It defines an energy function that simultaneously considers overall spatial and channel features, maintaining a lightweight design while improving model performance. These attention mechanism modules have been widely applied to enhance the performance of convolutional neural networks. Debris flow fans segmentation is an intensive prediction task that requires predicting the category for each pixel while simultaneously

learning high-level semantic information, overall spatial information, and spatial coordinate information for each feature. The SimAM module determines the importance of the current neuron by linearly separating it from other neurons, emphasizing overall spatial information, but it does not retain the spatial coordinate information of neurons. Therefore, we propose an improved strategy for the SimAM attention mechanism, preserving feature spatial coordinate information to achieve precise segmentation of debris flow fans.

In order to achieve precise segmentation of debris flow fans, this paper combines convolutional neural networks and proposes a Spatial Coordinate Attention Mechanism (SCAM). The main contributions of this paper are as follows: (1) Based on neuroscience, a new energy function is proposed to enhance the deficiencies of the SimAM module in extracting spatial coordinate information; (2) The closed-form solution of the energy function is derived through mathematical theory, ensuring the lightweight nature of the module; (3) The model is embedded into different semantic segmentation network architectures and evaluated on the task of debris flow fans segmentation.

2 Spatial Coordinate Attention Mechanism

2.1 SimAM

The SimAM attention mechanism is inspired by the spatial inhibition mechanism in neuroscience theory. By defining an energy function and utilizing the closed-form solution of the energy function as weights, the module is lightweight and enhances interpretability.

The energy function provided by SimAM is expressed as Eq. (1).

$$e_t\left(w_t, b_t, y, x_i\right)=\frac{1}{M-1}\sum_{i=1}^{M-1}\left(y_0-\left(w_t x_i+b_t\right)\right)^2+\left(y_t-\left(w_t t+b_t\right)\right)^2+\lambda w_t^2 \tag{1}$$

In the formula, $y_t=1$ and $y_0=-1$ represent two different label values, t and x_i are the target neuron and other neurons in a single channel of input features, i is the index of spatial coordinates, M is the number of neurons on that channel, w_t and b_t are the weights and biases of the transformation.

Solve the partial derivatives of Eq. (1) with respect to w_t and b_t and set them equal to 0. Assuming that all neurons in a single channel follow the same distribution, substituting w_t and b_t back into Eq. (1) yields the minimum energy as:

$$e_t^*=\frac{4\left(\hat{\sigma}^2+\lambda\right)}{\left(t-\hat{u}\right)^2+2\hat{\sigma}+2\lambda}. \tag{2}$$

In the formula, $\hat{u}=\frac{1}{M}\sum_i^M x_i$ and $\hat{\sigma}^2=\frac{1}{M}\sum_i^M\left(x_i-\hat{u}\right)^2$. The lower energy e_t^* is, the greater the difference between the target neuron and other neurons, indicating increased importance for visual processing. Therefore, the importance of each neuron can be represented by $1/e_t^*$. Additionally, the attention regulation mechanism in the mammalian brain often manifests as a scaling effect on neurons. Hence, a scaling operator is used to weight and refine features, as illustrated in Eq. (3).

$$\tilde{t}=sigmoid\left(\frac{1}{e_t^*}\right)\otimes t \tag{3}$$

The addition of the sigmoid function in the formula is to prevent e_t^* from becoming too large. Since sigmoid is a monotonic function, it doesn't affect the relative importance of each neuron.

2.2 SCAM

The SimAM algorithm evaluates the current neuron and other neurons based on linear separability to determine the importance of neurons, highlighting overall spatial information. However, in the process of handling neuron information, the algorithm does not preserve the spatial coordinate information of neurons, which is crucial for segmentation tasks. Therefore, this paper proposes the following improvement strategy.

Observing all neurons on a single channel, from Eq. (1), we can derive the following expression:

$$\frac{1}{M}\sum_{i=1}^{M}\left(y_0-\hat{x}_i\right)^2=\frac{1}{W\times H}\sum_{j=1}^{H}\sum_{i=1}^{W}\left(y_0-\hat{x}_{j,i}\right)^2. \tag{4}$$

In the formula, $\hat{x}_{j,i}=w_t x_{j,i}+b_t$. $x_{j,i}$ represents a neuron with height j and width i in a single channel, where W is the maximum width and H is the maximum height of the spatial dimension in that channel. To obtain accurate spatial position information and capture long-range information, we decompose it. Specifically, given the input feature X, each neuron in the feature undergoes a linear transformation $\hat{X}=w_t X+b_t$, followed by taking the squared difference with the label y_0, resulting in $y_0-\hat{X}^2$. Finally, using pooling kernels (H,1) and (1,W) along the horizontal and vertical coordinates, respectively, each channel is encoded. Therefore, the output at height h in the channel can be represented as Eq. (5).

$$z_h(h)=\frac{1}{W}\sum_{i=1}^{W}\left(y_0-\hat{x}_{h,i}\right)^2 \tag{5}$$

Similarly, the output at width w can be represented as Eq. (6).

$$z_w(w)=\frac{1}{H}\sum_{i=1}^{H}\left(y_0-\hat{x}_{j,w}\right)^2 \tag{6}$$

The two mentioned transformations target feature aggregation and implementation of corresponding transformations for the two spatial directions (height and width), generating a pair of direction-aware feature maps. These transformations allow our attention to retain accurate position information while capturing long-range feature relationships, aiding the network in more accurately locating the target of interest.

Therefore, this paper presents a new energy formula, as shown in Eq. (7).

$$e_t'\left(w_t,b_t,y,x_{h,i},x_{j,w}\right)=\left(y_t-\hat{t}\right)^2+\frac{1}{W}\sum_{i=1}^{W}\left(y_0-\hat{x}_{h,i}\right)^2+\frac{1}{H}\sum_{j=1}^{H}\left(y_0-\hat{x}_{j,w}\right)^2 \tag{7}$$

In the formula, $\hat{t}=w_t t+b_t$. $x_{h,i}$ and $x_{j,w}$ are other neurons located at the same height and width as the target neuron t, and all values in Eq. (7) are scalars.

Adding a regularization term in Eq.(7), the final energy function is:

$$e_t'\left(w_t, b_t, y, x_{h,i}, x_{j,w}\right) = \left(1-\left(w_t t + b_t\right)\right)^2 + \frac{1}{W}\sum_{i=1}^{W}\left(-1-\left(w_t x_{h,i} + b_t\right)\right)^2 + \frac{1}{H}\sum_{j=1}^{H}\left(-1-\left(w_t x_{j,w} + b_t\right)\right)^2 + \lambda w_t^2. \tag{8}$$

In the formula, λ is the regularization coefficient.

Eq. (8) assesses the importance of the current neuron based on the value of the neuron after a linear transformation. Specifically, if the value of the current neuron t is close to 1, and the values of neurons at the same height and width as the current neuron are close to -1, it indicates that the current neuron is more important. Therefore, the lower the value of Eq. (8), the more significant neuron t is in the current task.

In theory, each channel has M energy functions, and solving all equations computationally using iterative methods like Stochastic Gradient Descent (SGD), Newton's method, etc., can be computationally burdensome. Fortunately, Eq. (8) can be solved for the minimum value through optimization theory, leading to the following unconstrained optimization problem $\min_{w_t,b_t} e_t'\left(w_t, b_t, y, x_{h,i}, x_{j,w}\right)$.

The function $e_t'\left(w_t, b_t, y, x_{h,i}, x_{j,w}\right)$ is twice differentiable, and the Hessian matrix is positive semi-definite. Therefore, $e_t'\left(w_t, b_t, y, x_{h,i}, x_{j,w}\right)$ is a convex function, making this optimization problem a convex optimization problem with a global minimum.

Setting $\frac{\partial e_t'}{\partial b_t} = 0$ and $\frac{\partial e_t'}{\partial w_t} = 0$, the solution is:

$$b_t = -\frac{1}{3} - \frac{1}{3}w_t\left(u+v+t\right), \tag{9}$$

$$w_t = -\frac{2\left(u+v-2t\right)}{\left(u-v\right)^2 + \left(u-t\right)^2 + \left(v-t\right)^2 + 3\left(\alpha^2+\beta^2+\lambda\right)}. \tag{10}$$

In the formula, $u = \frac{1}{W}\sum_{i=1}^{W} x_{h,i}$, $v = \frac{1}{H}\sum_{i=1}^{H} x_{j,w}$, $\alpha^2 = \frac{1}{W}\sum_{i=1}^{W}\left(x_{h,i}-u\right)^2$ and $\beta^2 = \frac{1}{H}\sum_{j=1}^{H}\left(x_{j,w}-v\right)^2$.

Substituting Eqs. (9) and (10) into Eq. (8), we obtain the minimum energy formula:

$$e_t'^{*} = \frac{4\left(2F - \left(u+v-2t\right)^2\right)}{3F}. \tag{11}$$

In the formula, $F = \left(u-v\right)^2 + \left(u-t\right)^2 + \left(v-t\right)^2 + 3\left(\alpha^2+\beta^2+\lambda\right)$.

After obtaining the minimum energy, $1/e_t'^{*}$ is used to represent the importance of each neuron's position. Additionally, Eq. (11) is for extracting spatial coordinate information of the target neuron, so it needs to be combined with SimAM's description of overall features. The final result is shown in Eq. (12).

$$\gamma_t = \frac{1}{e_t'^{*}} + \frac{1}{e_t^{*}} \tag{12}$$

Similar to the SimAM algorithm, the method proposed in this paper also utilizes a scaling operator to weight and refine features.

3 Experimental Results and Analysis

3.1 Data Introduction

In this study, the debris flow fans data was obtained through the following steps: first, provinces with frequent debris flows were selected, and relevant local disaster reduction yearbooks such as "Sichuan Disaster Reduction Yearbook" and "Yunnan Disaster Reduction Yearbook" were consulted. Secondly, relevant papers and news reports on debris flows from 2000 to the present were collected. After sorting, a total of 749 gullies with debris flows were identified, of which the fans of 31 gullies remained undamaged. Using the coordinates of the debris flow fans, 31 high-resolution remote sensing images with a size of 8192 $\times$ 4585 were extracted and saved.

We determined the specific extent of debris flow fans in each image through manual identification and field surveys, and processed them using the Labelme tool to obtain accurate labels for the debris flow fan images. Subsequently, data augmentation methods were applied to both the images and the labels, resulting in 90 usable images and their corresponding label data. Fig. 1 shows a data example.

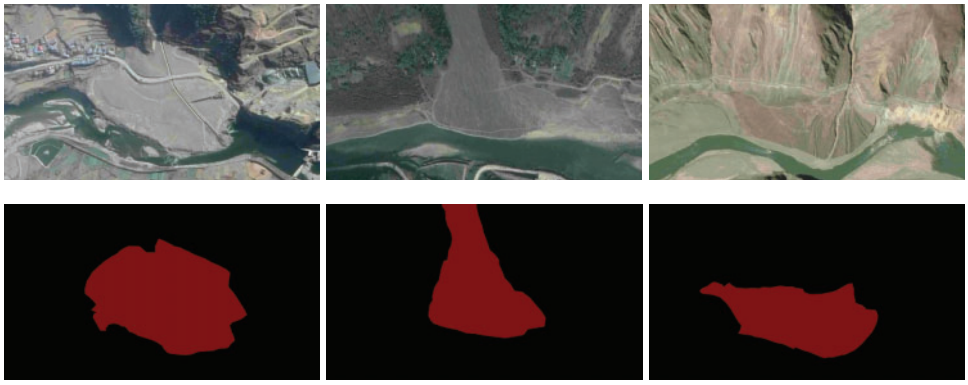

Figure 1: Example data of debris flow fans

3.2 Parameter Design and Evaluation

Metrics for ease of comparison, all experiments in this study were conducted using the PyTorch framework. Hardware environment: CPU-Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50 GHz, GPU-NVIDIA GeForce RTX 2080 Ti; Software environment: Ubuntu 18.04, Python 3.8, CUDA 11.3, cuDNN 8, NVCC, PyTorch 1.11.0, torchvision 0.12.0, torchaudio 0.11.0. All experiments in this paper use the cross-entropy loss function.

The Mean Intersection over Union (mIoU), Pixel Accuracy (PA), Mean Pixel Accuracy (mPA) and F1 score (F1) are four metrics used to evaluate the classification performance of different classification methods.

The mIoU is calculated by taking the ratio of the intersection to the union for each predicted category and true label. The values are summed, averaged, and then translated into a confusion matrix, as shown in Eq. (13).

$$mIoU = \frac{\sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i + FN_i}}{n} \tag{13}$$

In the formula, TP_i represents the number of true positive predictions for the i-th class, FP_i represents the number of false positive predictions for the i-th class, FN_i represents the number of false negative predictions for the i-th class, and n represents the number of classes.

The specific meaning of PA is the percentage of correctly predicted pixel values out of the total pixel values, as shown in Eq. (14).

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

In the formula, TP represents the quantity of true positive predictions, FP represents the quantity of false positive predictions, FN represents the quantity of false negative predictions, and TN represents the quantity of true negative predictions.

The specific meaning of mPA is the average of the pixel accuracy for all classes. Firstly, the pixel accuracy for each class is determined, and then the sum of these accuracies is averaged, as shown in Eq. (15).

$$mPA = \frac{\sum_{i=1}^{n} \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}}{n} \tag{15}$$

In the formula, TN_i represents the quantity of true negative predictions for the i-th class.

The F1 is a metric that measures the similarity between two sets, with values ranging from 0 to 1. For semantic segmentation tasks, it is used to evaluate the similarity between the network's predicted segmentation results and manually annotated results. The F1 score is equal to twice the product of precision and recall divided by the sum of precision and recall. It can be converted to the confusion matrix, as shown in Eq. (16).

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{16}$$

The mIoU is the average Intersection over IoU of all classes, providing an overall measure of the quality of the segmentation results. It comprehensively assesses the model's performance on different classes. PA and mPA gauge the overall accuracy of predictions, while the F1 score measures the overlap between predicted results and true labels. The importance of these four metrics in semantic segmentation varies, with mIoU being the most crucial, followed by mPA and PA, and F1 score considered least important.

3.3 Experimental Results and Analysis

To validate the effectiveness of SCAM, this study employed three typical deep learning semantic segmentation models—FCN , UNet, and PSPNet—as backbone networks. The effects of applying common attention mechanisms such as SE, CBAM, CA, and the proposed SCAM in the backbone networks were explored. Through a comparative analysis of four evaluation metrics and the segmentation results on test images, the study revealed differences between the introduction of different attention mechanisms in different backbone networks.

SCAM, along with other attention mechanism modules, was embedded in the skip-connection part of UNet and before each downsampling in FCN and PSPNet. The segmentation results of the test images are illustrated in Fig. 2.

Comparing the segmentation results between images with and without attention mechanisms, it is evident that incorporating attention mechanisms is more beneficial for the network to identify crucial regions. Observing images in the 3rd, 4th, and 5th groups, there are still issues with breakpoints and noise points. In contrast, the SCAM method exhibits the optimal performance, producing more precise, complete segmentation results with reduced noise points. Looking at the 2nd group of images, when using FCN, the CA and CBAM modules almost fail to recognize the presence of the accumulated fan. Embedding SimAM and SCAM achieves better segmentation results. In the 6th group of images, regardless of the module used, the segmentation of the debris flow fans is relatively good, but there are still issues with inaccurate localization and incomplete segmentation. The attention mechanism proposed in this study effectively addresses these problems. Examining the 1st group of images, the segmentation results of all methods are suboptimal, but the approach proposed in this study still demonstrates relatively better performance.

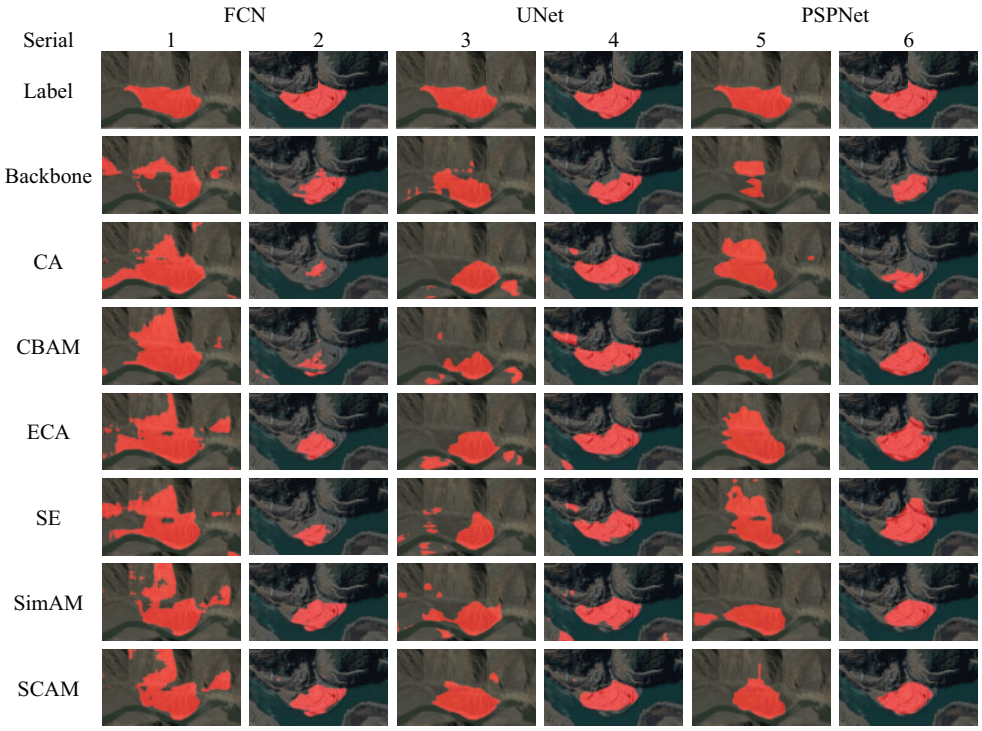

Figure 2: Comparison diagram of segmentation results

The mIoU, mPA, PA, and F1 scores were calculated for the predicted images generated using the mentioned methods on the test set, and the summarized results are presented in Table 1. While embedding SCAM in various backbone networks, although the F1 score in UNet and mPA and F1 in PSPNet are not the highest, they are close to the highest values, and all other metrics reach their peak levels. Specifically, compared to the worst results after adding attention mechanism modules in each backbone network, in FCN, mIoU increased by 3.7%, mPA increased by 5.3%, PA increased by 11.1%, and F1 score increased by 6.4%. In UNet, mIoU increased by 4.1%, mPA increased by 4.3% and PA

increased by 1.2%. In PSPNet, mIoU increased by 9.0% and PA increased by 3.7%. When comparing the three backbone networks without attention mechanisms, adding different attention mechanisms may cause fluctuations in network performance. For example, in FCN, adding CA, CBAM, ECA, SE, and SimAM attention mechanism modules all led to a decrease in the mIoU metric. However, regardless of the backbone network, embedding the SCAM module proposed in this study improves the network model's performance to a certain extent.

Table 1: Evaluation results using different backbone networks and attention mechanisms

Backbone	Attention mechanism	mIoU (%)	mPA (%)	PA (%)	F1 (%)
FCN	/	79.1	86.5	93.3	87.7
	CA	77.4	86.5	93.0	86.5
	CBAM	76.1	87.1	82.3	85.6
	ECA	76.6	88.7	92.2	86.0
	SE	71.0	85.8	89.5	81.8
	SimAM	77.3	89.0	92.3	86.5
	SCAM	**79.8**	**91.1**	**93.4**	**88.2**
UNet	/	80.5	87.7	94.2	88.8
	CA	81.7	87.8	94.7	87.7
	CBAM	82.7	89.1	95.0	89.0
	ECA	82.9	88.4	95.2	86.1
	SE	81.7	88.7	94.8	**89.1**
	SimAM	81.2	90.5	94.2	87.5
	SCAM	**84.6**	**92.0**	**95.4**	89.0
PSPNet	/	76.5	88.9	92.0	87.0
	CA	73.5	84.9	91.2	84.8
	CBAM	79.4	87.7	93.7	89.0
	ECA	81.8	**90.6**	94.4	**90.3**
	SE	73.7	84.1	91.5	85.0
	SimAM	80.3	87.1	94.2	89.2
	SCAM	**82.5**	89.3	**94.9**	90.1

The above results indicate that in the task of debris flow fans segmentation, SCAM exhibits significant advantages compared to similar attention mechanism algorithms. It can effectively enhance segmentation accuracy and overall performance.

3.4 Further Discussion

In order to systematically investigate the superiority of the SCAM attention mechanism, we comprehensively evaluated its performance through the loss change curve during training, heat maps generated by Grad-CAM, the specific values of parameter A in the energy function and its impact on network performance, as well as the computational and parameter size of SCAM. This thorough analysis provides insights into the practical performance of SCAM.

According to Figs. 3 and 4, we observe a high consistency in the training set loss curve of the network embedded with SCAM during the training process. However, on the validation set, the network embedded with SCAM exhibits lower loss values and greater stability.

Grad-CAM highlights the image areas the network focuses on by computing gradient weights. In the heatmap, the redder the color, the more attention it receives. According to Fig. 5, we observe that after integrating SCAM, the network focuses on significantly larger key regions. It pays more attention to the area where the debris flow fans is located and presents it in a fan-shaped pattern. This alignment is consistent with the shape of the debris flow fans. This result indicates that the introduction of SCAM effectively improves the network's ability to recognize key areas.

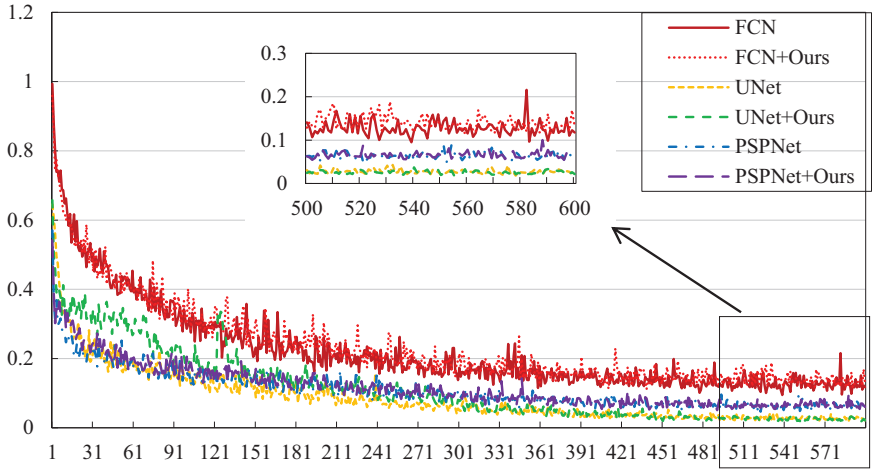

Figure 3: Comparison diagram of training set loss

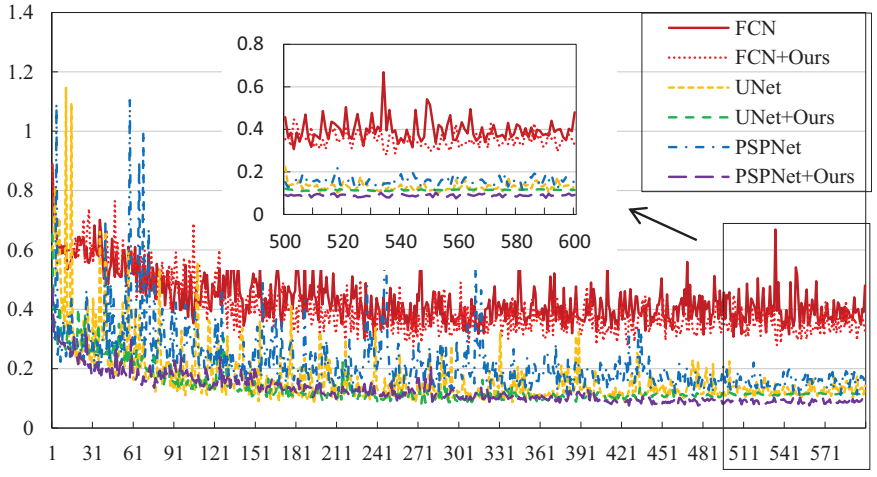

Figure 4: Comparison diagram of validation set loss

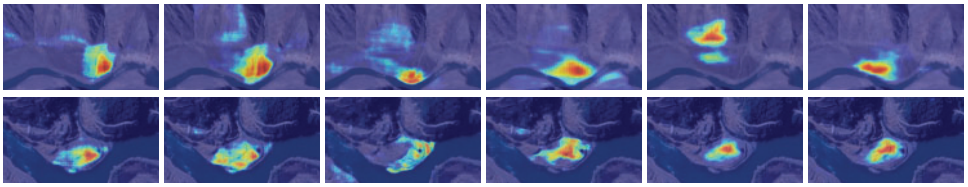

Figure 5: Comparison diagram of Grad-CAM

According to the data in Table 2, we can clearly see that embedding SCAM does not lead to an increase in the number of network parameters. However, it does result in a slight increase in computational complexity. Considering the significant improvement in network performance with SCAM, this minor increase in computational complexity can be considered negligible.

Table 2: FLOPs and Params of SCAM and typical attention mechanisms

Attention mechanisms	FLOPs (G)	Params (M)
CA	0.142	1.917
CBAM	0.041	2.516
ECA	0.029	0
SE	0.034	5.030
SimAM	0	0
SCAM	0.120	0

Fig. 6 presents the results of four evaluation metrics on the training set using different values for λ. Due to the difficulty in validating all possible values of λ, we set the range of λ values from 10^{-1} to 10^{-5}. Based on Fig. 6, it can be clearly concluded that utilizing this module significantly enhances network performance.

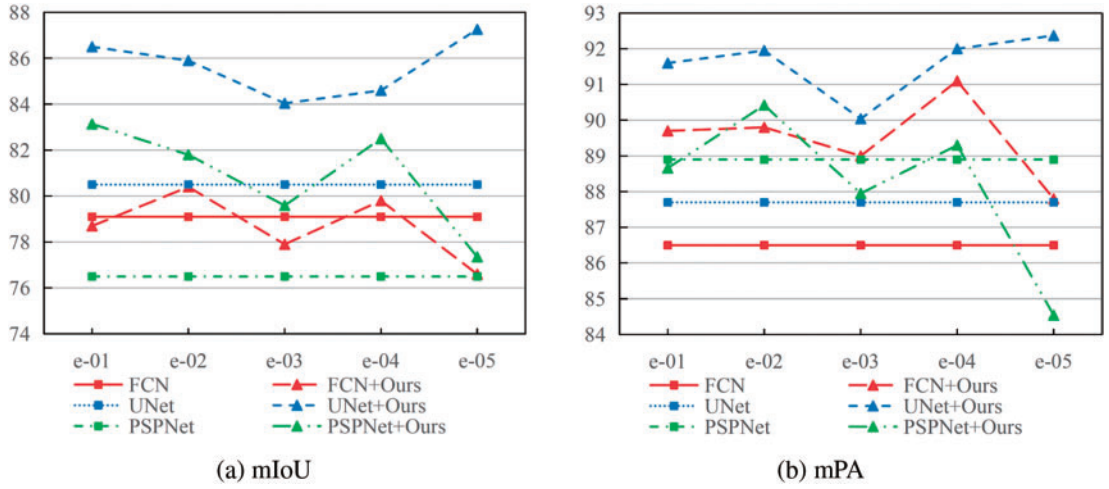

Figure 6: (Continued)

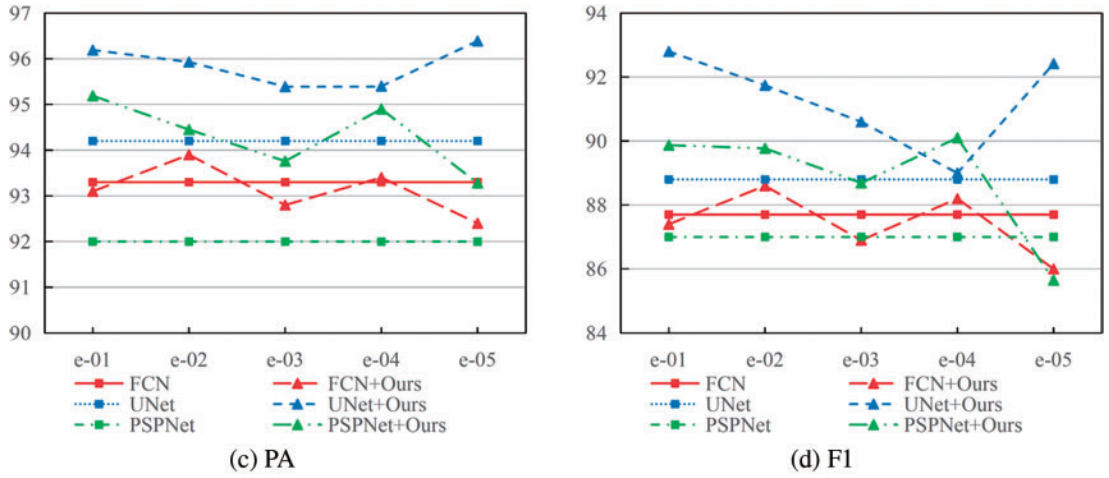

Figure 6: Comparison diagram of evaluation indicators with different values of λ

4 Conclusion

This paper addresses the limitations of the SimAM attention mechanism in extracting spatial information from feature spaces. It proposes an optimization strategy that incorporates spatial coordinate information, making it easily implementable through optimization theory while ensuring lightweight module design and high efficiency.

In the task of debris flow fan segmentation, six attention mechanisms (CA, CBAM, ECA, SE, SimAM, and SCAM) are embedded into different backbone networks. Through experimental observations, SCAM demonstrates superior performance. SCAM effectively addresses noise and discontinuity issues, achieving precise localization of debris flow fans and generating more complete and accurate segmentation results. Detailed discussions and studies on SCAM are also conducted in this paper. SCAM still has certain limitations, as its performance heavily relies on the effectiveness of the appropriate selection of parameter λ. Training with inappropriate parameters may significantly deviate from expected results. Addressing these issues, we have identified them as key directions for future improvements.

The proposed method extends the segmentation and recognition techniques for debris flow fans, contributing to the development of theories in this field. It has profound implications for the study of debris flow geomorphic processes, as well as for the assessment of debris flow hazards and the selection of monitoring and warning areas.

Acknowledgement: We thank all reviewers for their kindly and constructive suggestions.

Funding Statement: This work was supported by National Natural Science Foundation of China, Grant Number 61966040.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Xin Song, Baoyun Wang; data collection: Xin Song; analysis and interpretation of results: Xin Song; essay writing: Xin Song; dissertation review: Baoyun Wang, Xin Song. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The raw data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006. doi: 10.1038/nbt1206-1565 .
2. G. Li and Y. Wan, "A new combination classification of pixel-and object-based methods," *Int. J. Remote Sens.*, vol. 36, no. 23, pp. 5842–5868, 2015. doi: 10.1080/01431161.2015.1109728 .
3. P. Wang *et al.*, "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2018, pp. 1451–1460. doi: 10.1109/WACV.2018.00163 .
4. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90 .
6. Z. Zhou, and M. Yuan, "Steel surface defect detection network based on improved YOLOv5s," *Métodos numéricos para cálculo y diseño en ingeniería: Revista internacional*, vol. 40, no. 2, pp. 1–11, 2024. doi: 10.23967/j.rimni.2024.05.011 .
7. F. Xu and B. Wang, "Debris flow susceptibility mapping in mountainous area based on multi-source data fusion and CNN model-taking Nujiang Prefecture, China as an example," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 1966–1988, 2022. doi: 10.1080/17538947.2022.2142304 .
8. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440. doi: 10.48550/arXiv v.1411.4038 .
9. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Intervent.-MICCAI 2015: 18th Int. Conf.*, Munich, Germany, Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28 .
10. H. Zhao, J. Shi, X. Qi, and X. Wang, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890. doi: 10.48550/arXiv v.1612.01105 .
11. Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, "Global context networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 6881–6895, 2020. doi: 10.1109/TPAMI.2020.3047209 .
12. Z. Yang, L. Zhu, Y. Wu, and Y. Yang, "Gated channel transformation for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11794–11803. doi: 10.48550/arXiv v.1909.11519 .
13. M. Nawaz, R. Qureshi, M. A. Teevno, and A. R. Shahid, "Object detection and segmentation by composition of fast fuzzy C-mean clustering based maps," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 6, pp. 7173–7188, 2023. doi: 10.1007/s12652-021-03570-6 .
14. Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021. doi: 10.1016/j.neucom.2021.03.091 .
15. H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10705–10714. doi: 10.1109/cvpr.2019.01096 .
16. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141. doi: 10.48550/arXiv v.1709.01507 .

17. S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CB AM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19. doi: 10.48550/arXi v.1807.06521 .
18. Q. Wang, B. Wu, P. Zhu, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542. doi: 10.48550/arXi v.1910.03151 .
19. Q. Hou, D. Zhou, and J. Feng, "Coor dinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722. doi: 10.48550/arXi v.2103.02907 .
20. L. Yang, R. Y. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.