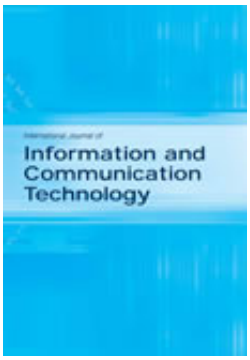
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Bidirectional attention network for real-time segmentation of forest fires based on UAV images

Zhuangwei Ji, Xincheng Zhong

Article History:

Received:	02 July 2024
Last revised:	23 July 2024
Accepted:	23 July 2024
Published online:	12 September 2024

Bidirectional attention network for real-time segmentation of forest fires based on UAV images

Zhuangwei Ji* and Xincheng Zhong

Computer Science Department,
Changzhi College,
Changzhi, 046000, Shanxi, China
Email: 17835852216@163.com
Email: cy12016596@czc.edu.cn
*Corresponding author

Abstract: For the purpose of monitoring hill fires, a bidirectional attention module is designed, which not only allocates the weights between different features reasonably, but also pays attention to the relationship between neighbouring pixels of the same feature to accurately segment the boundaries of flames and forest background. A feature fusion module is designed to effectively fuse deep and shallow feature information, and automatically adjust the input feature weights to improve the model's segmentation ability for small fire targets. Finally, we conducted experiments on the public dataset Flame, and the results show that the designed model outperforms other state-of-the-art methods in segmentation accuracy and computational efficiency.

Keywords: UAV images; semantic segmentation; hill fire detection.

Reference to this paper should be made as follows: Ji, Z. and Zhong, X. (2024) 'Bidirectional attention network for real-time segmentation of forest fires based on UAV images', *Int. J. Information and Communication Technology*, Vol. 25, No. 6, pp.38–51.

Biographical notes: Zhuangwei Ji graduated from Harbin Institute of Technology in 2018 with a Master's degree in Computer Technology. He is working as a Lecturer in Changzhi College. He mainly teaching data structures, databases and other specialised core courses, and the direction of scientific research is semantic segmentation.

Xincheng Zhong graduated from University of Chinese Academy of Sciences in 2016 with a Master's degree in Computer Science and Technology. He is working as a Lecturer in Changzhi College. He mainly teaching machine learning, discrete mathematics and other courses, the scientific research direction is semantic and instance segmentation.

1 Introduction

In recent years, forest fires have been occurring frequently, not only destroying the ecological environment, but also wasting a large amount of human and material resources and requiring firefighters to risk their lives to combat the disasters. If fires can be monitored in a timely manner, losses can be effectively reduced. The monitoring of forest

fires can be basically categorised into two types, one is ground monitoring and the other is aerial monitoring. Ground monitoring technology (Aliser and Duranay, 2024; Wang et al., 2024) is through the deployment of temperature, smoke sensor network in the forest, can be real-time collection of environmental data, once the parameter exceeds the preset threshold that is to issue an alarm, this method can be a wide range of fire monitoring, to achieve early warning, but the sensor's maintenance costs are high, difficult, and the transmission of data by the complex terrain obstacles to the real-time and integrity of the data. Airborne monitoring technology (Gupta and Nihalani, 2024; Akyol, 2024), on the other hand, is real-time monitoring of fires through aerial equipment, which can be categorised into remote sensing satellite monitoring and drone monitoring. Remote sensing satellite monitoring is to draw thermal radiation remote sensing images of forests in real-time from space through satellites, and then neural networks detect and segment fires on the images, but the operating cost of satellites is high, the problems involved are complex, and the remote sensing images are obstructed by cloud cover, which affects the accuracy of monitoring. Drone monitoring is achieved by deploying drones to capture forest images in real-time for fire monitoring, which has low operating costs and high flexibility compared to large aircraft and satellites. Moreover, the images taken are not affected by cloud cover and are not limited by terrain, so monitoring forest fires by taking images from drones has been widely utilised.

In this paper, we propose a semantic segmentation network for processing images captured by UAVs to help the monitoring department to monitor forest fires in a timely manner. The network chooses STDCNet as the benchmark model (Fan et al., 2021), which is an optimisation of the classical real-time semantic segmentation network BiseNet (Xu et al., 2021), where the STDC module is a feature extraction module specially designed for the semantic segmentation task of the image, which can then perform multi-scale feature extraction on the image without increasing the amount of computation. Through comparative experiments, compared with VGG16 (Simonyan and Zisserman, 2014), DeepLabV3 (Chen et al., 2017), which are common benchmark models, STDCNet better meets the requirements of the semantic segmentation task of the model in this paper. Other attention mechanism networks only focus on the connection between different features (Agarwal et al., 2024; Wang et al., 2024) by assigning different weights to different features in order to enhance the influence of important features on the result, which improves the segmentation accuracy of the target but locates the boundary pixels ambiguously. In this paper, a two-way attention module is proposed to optimise the extracted features, which is divided into two paths, one path assigns appropriate weights to different channel features to improve the accuracy of judging individual pixel categories, which is called the attention channel module (Karandikar et al., 2024; Zedda et al., 2024), and the other path pays attention to the strength of the connection between adjacent pixels in the same channel to improve the model's accurate judgments of boundary pixels and optimise the segmentation boundaries between different features, thus improving the segmentation accuracy of small targets, and is referred to as the SAM. Since shallow spatial information and deep semantic information cannot be directly fused, this paper proposes a specialised feature fusion module (FFM), which makes the shallow spatial information can get the guidance of the deep semantic information, so as to improve the segmentation accuracy of the model, due to the fact that the small segmentation target fire may be treated as a noise, in order to improve the segmentation ability of the small target, we expand the fusion feature output stage by adding a fully connected conditional random field , which transforms the segmentation

problem of the image into an optimisation problem and effectively prevents the problem of losing small targets (Xue et al., 2023; Liang et al., 2021; Ou et al., 2024).

This is how the rest of the paper is structured. The study pertaining to forest fire segmentation is briefly described in the second section. The third station section describes the design of each module proposed in this paper. Section 4 conducts experiments on Flame dataset to evaluate the enhancement effect of the proposed modules on fire segmentation and gives some results of fire segmentation. Section 5 summarises the work and gives the possible future work plan.

2 Related work

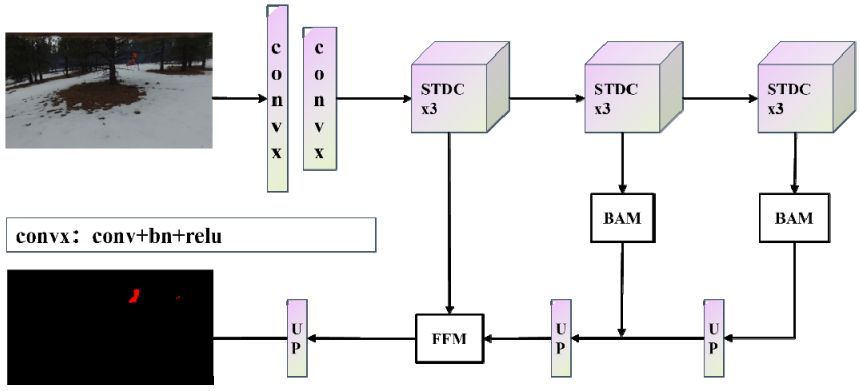
Forest fire segmentation is used to extract the fire region from the fire image by training the network so as to find a suitable segmentation threshold at pixel level. With the development of semantic segmentation technology, more and more segmentation networks have been applied to forest fire detection with good results, but fire segmentation of UAV images still faces many challenges.

In real scenarios, the flame's in the image may be mixed with a variety of complex backgrounds, such as lighting, smoke, vegetation, etc., which can easily lead to false or missed detection of fires, Chen et al. (2024) designed a multimodal fusion network, which combines the features of visible and thermal infrared images, and utilises the dual-modal information for the refined segmentation of the fire region, which effectively enhances the network's background occlusion in the presence of the detection ability. The features of the flame in the video, such as morphology, brightness, colour, etc., change rapidly with time and environment, which also puts higher requirements on the model's generalisation ability, so Arlovi et al. (2024) use a time series model to capture the temporal information of the flame (Goay et al., 2023; Arlovi et al., 2024), combined with the relationship between the video frames to stabilise the segmentation effect; Turker and Eksioglu (2022) dynamically adjust the network's weights, for different scenarios and flame states to fine-tune the parameters of the model to address the effect of flame changes on the segmentation effect of the model. Fire detection requires fast response, so it puts real-time requirements on the network, and more and more lightweight real-time segmentation networks (Qu et al., 2024; Hu et al., 2024; Zhu et al., 2023) are applied in fire segmentation. Li et al. (2021) optimise the network structure, and improve the network inference speed by reducing the network level and complexity. Commonly used lightweight networks MobileNetV3 and ShuffleNetV2 also appeared in fire segmentation, designed for real-time semantic segmentation network ENet (Paszke et al., 2016), also often used as a benchmark model in fire segmentation network. Often the percentage of pixels of flame in an image will be very low, which causes the class imbalance problem, thus affecting the accuracy of segmentation. Zhao et al (2022) explored how to combine the conditional random fields (CRFs) and data enhancement strategies to solve the class imbalance problem in semantic segmentation, which is an inspiration for the segmentation of fires. Zhou and Feng (2017) used the idea of integrated learning to combat the class imbalance problem.

It is difficult to achieve the balance between speed and accuracy in forest fire segmentation network nowadays, and the use of deep convolutional networks such as U-Net (Weng and Zhu, 2021) and DeepLabV3 to segment the fire region can ensure the segmentation accuracy, but its detection speed is slow and it is difficult to ensure

real-time. And for real-time, the method of pruning the model more or less affects the segmentation accuracy. Therefore, in this paper, a specialised semantic segmentation network STDCNet is chosen as the basic model, which can fuse multi-scale feature information without increasing the quantity of parameters, and guarantee the precision of the segmentation while improving the segmentation speed. In fire images, the correct classification of the boundary pixels between the flame and the background can effectively improve the segmentation accuracy of the model for the flame, so this paper proposes a two-way attention module, which not only focuses on the weight distribution between different channels, but also takes into account the connection between neighbouring pixels in the same channel, so as to improve the segmentation accuracy from two directions. A fully connected conditional random field is incorporated into this paper's specialised FFM, which not only effectively fuses deep and shallow feature information but also automatically adjusts the input feature weights to prevent small flames from being lost as noise. Small flames are frequently seen in fire images, and this paper aims to prevent the model from treating them as noise.

Figure 1 Overall architecture diagram of the proposed network basic module (see online version for colours)



3 Methods

In this chapter, a deep learning network with a bidirectional attention mechanism is proposed to realise real-time segmentation of forest fires. The general structure of the network model is shown in Figure 1, the input image is first convolved with two layers, and then the obtained feature image is fed into STDC blocks, 3 blocks as a group, we choose STDC as the encoder because each STDC block can extract features at multiple scales and splice them together, and at the same time has a high computational efficiency. The feature maps obtained by the encoder are optimised by the bidirectional attention module (BAM), which assigns the weights between different features and the strength judgment of the links between neighbouring pixels. The deep semantic feature information and shallow spatial feature information are fused by the FFM, which modifies the input multi-level feature weights automatically, effectively fuses multi-level feature information, and enhances the network's segmentation capability for small targets.

3.1 Basic module

The backbone network for semantic segmentation is usually based on existing image classification backbone networks, such as vggNET, DEEPLAB, etc. These networks are designed for image classification tasks, and it is difficult to meet the segmentation network needs to have scalable receptive field and multi-scale information at the same time, and the classification usually does not have real-time requirements, and the network has a rather large number of parameters, and we need lightweight backbone networks to improve the segmentation speed. STDCNet is specifically designed for semantic segmentation tasks, using short-time dense concatenated modules (STDC) to simultaneously acquire sensory field and multi-scale feature information, the specific structure is shown in Figure 2, each Bblock block in the STDC module includes a conv-bn-relu operator, and given the output channel N of the module, we can calculate the number of convolution kernels of each convolutional layer in the middle, except that the number of convolution kernels of the last block is the same as that of the previous layer, the number of convolution kernels in the i^{th} block is n/2i, and the output of the network is the spliced fusion of the output channels of each convolutional layer in the middle, so that the module obtains the multiscale feature fusion information and the scalable sensory field, and improves the segmentation accuracy. The number of convolution kernels in the middle convolutional layer of the STDC module is geometrically reduced, reducing the computational accuracy. geometrically reduced, which reduces the computational complexity, and the total number of parameters of the STDC module is mainly determined by the sizes of the input and output channels, and the output channel n has little effect on the parameter sizes, and the number of parameters remains almost unchanged when n reaches the maximum limit. The relationship between the number of module parameters and channel n is shown in equation (1). The reduction of computational complexity and the stabilisation of the number of parameters guarantee the speed of segmentation.

$$S_{param} = \frac{N * M}{2} + \frac{3N^2}{2} * \left(1 + \frac{1}{2^{2n-3}}\right) \tag{1}$$

3.2 Bidirectional attention module

The bi-directional attention module is divided into channel attention module (CAM) and spatial attention module (SAM) as shown in Figure 3. The same input feature map is divided into two paths for feature optimisation. CAM measures the degree of importance between each channel and assigns different weights to different features. SAM performs horizontal pooling and vertical pooling for the same feature, and fuses to obtain a feature representation optimised by the correlation of adjacent pixels of the same feature. The optimised feature maps output from the two paths are spliced and output after 1*1 convolutional fusion in order to enable the network to take into account information from both channel features and pixel space.

Figure 2 Diagram illustrating the structure of the STDC module bidirectional attention module (see online version for colours)

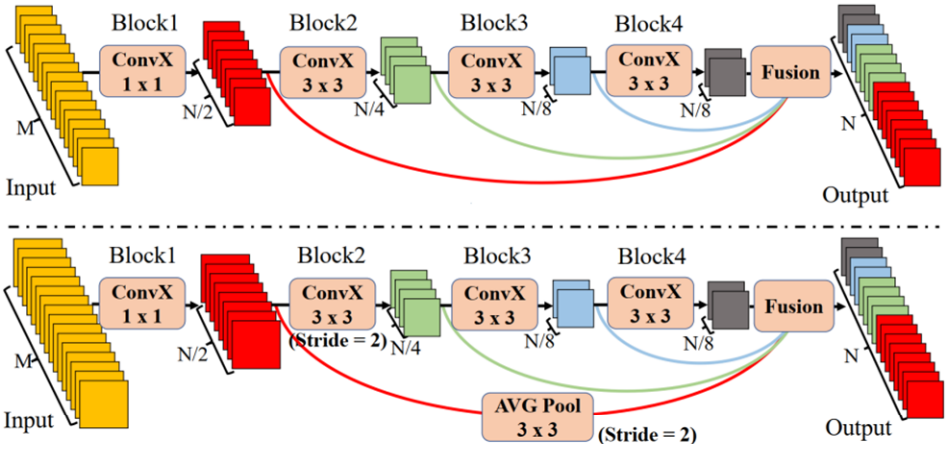
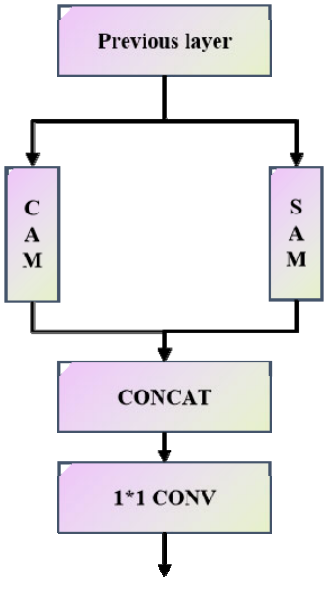


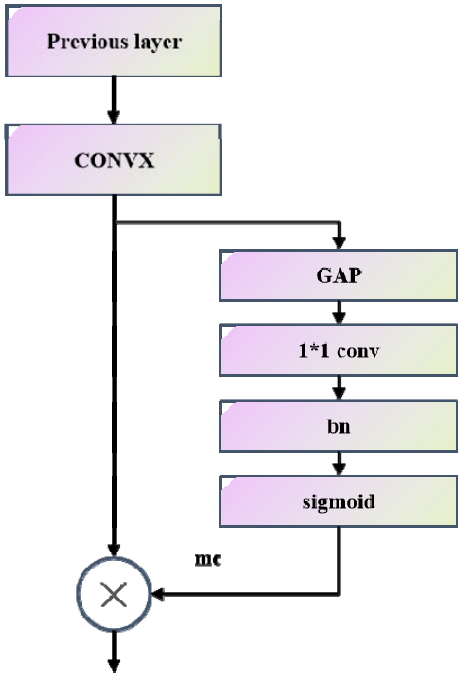
Figure 3 Diagram of the bidirectional attention module (see online version for colours)



Due to the small percentage of fire source pixels that need to be segmented in the forest fire image, which belongs to the small-scale target segmentation, as well as the backbone network is a lightweight and shallow network, which makes the network limited to capture semantic information due to the small target of the fire source. Therefore, before fusing the high and low dimensional information, it is necessary to use the focus of the channel module CAM to maximise each dimensional graph's attributes, according to the importance of each feature channel, re-assigning the weight value to it, and improving the feature graph's capacity to capture semantic information, as shown in Figure 4. Firstly, the feature map is input into the 3*3 convolution operator of bn and relu to equilibrium the quantity of channels in various dimensional feature maps, then the resolution of the

feature map is reduced by using global average pooling, and the 1*1 resolution feature map is input into the 1*1 convolution to adjust the channel weights, and finally the channel attention mask MC is generated via bn and sigmod, and the output of the channel attention branch is produced by multiplying the MC by the input feature map, which creates the channel attention mask MC.

Figure 4 Schematic of the CAM (see online version for colours)

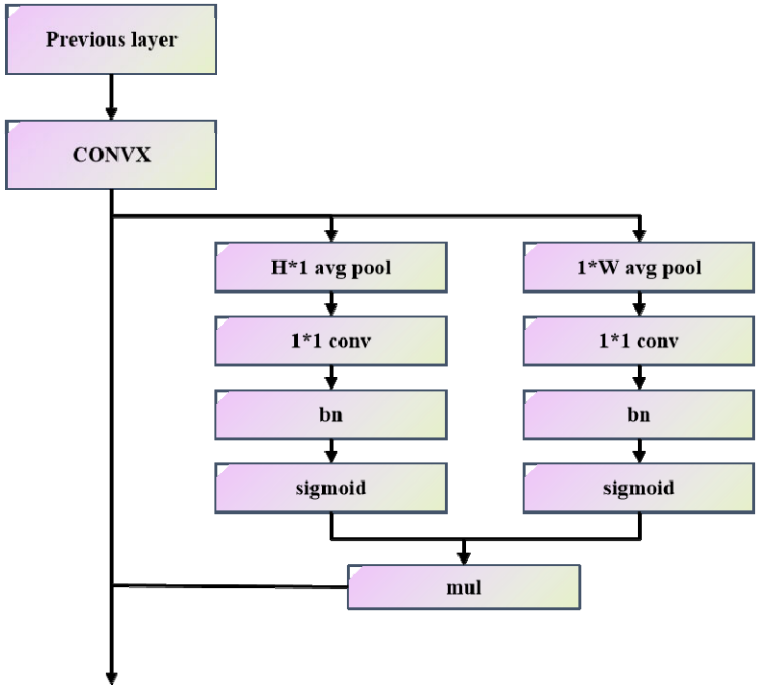


Although the CAM improves the information capturing ability of the feature map, the boundary blurring problem of segmenting the target and the forest background still cannot be solved, and the network is still unable to accurately segment small-scale targets. Therefore, we continue to optimise the feature map using the SAM, which performs average pooling on the same feature in the horizontal and vertical axes, respectively, so that the model focuses on the spatial relationship between neighbouring pixels of the same feature. The overall structure of SAM is shown in Figure 5, where the feature map of H*W resolution is pooled by the average pooling of H*1 and 1*W after one conv-bn-relu operator partitioned into two paths, one focusing on the correlation between horizontal pixels and the other on the correlation between vertical pixels. Following 1*1 convolution and sigmod calculation, the two routes provide weight vectors in their respective orientations. These vectors are then multiplied by the original feature map to produce the spatial attention branch's output.

The output splicing of the BAMintegrates the outputs of the two branches, so that the model can pay attention to the small target of the fire while also paying enough attention to the segmentation boundaries of the fire target and the background to improve the model's segmentation ability and segmentation accuracy for the small target. And the two-way attention module has no up-sampling operation from top to bottom, which

shows that the computational cost is quite low and does not affect the overall segmentation speed of the network.

Figure 5 Schematic of the SAM (see online version for colours)

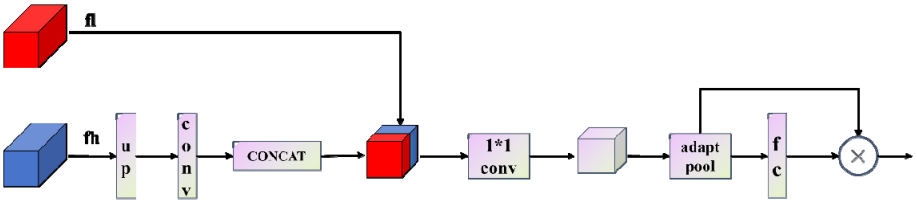


3.3 Feature fusion module

The shallow spatial feature information and the deep semantic feature information are at different feature representation levels, and we can't splice the feature maps directly, so We suggest the FFM, which improves the shallow feature map's capacity to extract abstract information by incorporating deep feature map data into the shallow feature map. This allows the shallow feature map to extract small-scale semantic information with greater accuracy. FFM The general arrangement is displayed in Figure 6. The fusion module has two inputs, shallow feature input fl and deep feature input fh, where the resolution of the deep feature map FH is 1/2 of the resolution of the shallow feature map FL, which can limit the information gap between the two feature maps and promote the information fusion. The deep feature map FH is first up-sampled and adjusted to the same resolution as the shallow feature map, and then a splice between the shallow feature map and the optimised deep feature map and aggregated using the up-sampled feature map generated by conv-bn-relu optimisation. The aggregated feature maps are first optimised by a 1*1 convolution operation, and then pass through an adaptive average pooling layer to produce channel metrics, which enhances the adaptive information features and suppresses the useless features, so that the model has the ability to automatically adjust the weights of the multilevel features. Finally, the C*1*1 statistics are fed into a fully connected conditional random field to further optimise the network's localisation of

individual pixels and improve the network's capacity to divide up targets into smaller segments.

Figure 6 Schematic of the FFM (see online version for colours)



Through the FFM, not only the network's ability to capture the deep semantic information is improved, and the shallow spatial information can be guided by the deep semantic information, the network obtains rich contextual and spatial information at the same time, so that some of the small fire targets in the image will not be discarded as noise, and the segmentation of the network on small targets is improved. The picture segmentation issue is changed into an optimisation problem by the completely linked conditional random field, solving the problem that the local minimum of the network is transmitted to the output layer to affect the segmentation effect.

4 Experiments

In order to validate the effectiveness of the module proposed in this paper, we do experiments on the Flame dataset. This section, describes the specific details of the experiments and compares the performance with today's state-of-the-art fire segmentation networks.

4.1 Datasets

Flame (Shamsoshoara et al., 2020) is a dataset containing video and images publicly released by Northern Arizona University and others to facilitate advances in hill fire detection technology. The dataset is video recordings and heat maps captured by infrared cameras, and the captured videos and images are annotated and labelled by frame. The semantic segmentation dataset contains 2003 images with 3,840 × 2,160 pixels, we divided the dataset into training and test sets in the ratio of 0.85:0.15, the training set contains 1,702 images and the test set contains 301 images. The specific case of fire images in this dataset is shown in Figure 7. The experiment was built using pytorch and run on a computer with win10 system and NVIDIA RTX 2080 TI.

During the training of the network, the original and labelled images are first resized to 512*512 before they are fed into the network to start training. The training process is optimised using the Adam optimiser with a learning rate of 0.001, which demonstrates good overall performance compared to other optimisers. Since the edge pixels of the fire and the background account for very few pixels relative to the non-edge pixels, which belongs to the class imbalance problem, we use the binary cross entropy as the loss function, which is insensitive to the number of background pixels, and can optimise the

edge learning of the model as shown in equation (2). The batch size parameter during training is set to 32, and a total of 30 epochs are trained.

Figure 7 Case of flame dataset (see online version for colours)

$$L = -\left[y \log\left(\hat{y}\right) + (1-y) \log\left(1-\hat{y}\right) \right], \tag{2}$$

4.1.1 Evaluation indicators

For an efficient assessment of the model's performance, we use the mean pixel accuracy (MPA), mean intersection ratio (MIOU) and frames per second (FPS) for performance evaluation, and the detailed definitions of these evaluation metrics are as follows.

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$MPA = \frac{1}{n} \tag{4}$$

$$Miou = \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right) * \frac{1}{2} \tag{5}$$

In the above formula, TP denotes the number of pixels predicting positive examples as positive examples, TN denotes the number of pixels predicting counterexamples as counterexamples, FP denotes the number of pixels predicting counterexamples as positive examples, FN denotes the number of pixels predicting positive examples as counterexamples, PAi denotes the pixel accuracy of the i^{th} category.

Among these 3 performance metrics, the network model's segmentation accuracy is measured using MPA and MIOU, and its processing speed is measured using FPS to see whether it is operating at a real-time level.

4.2 Performance evaluation

In order to prove the effectiveness of the network model proposed in this paper, we compare it with the existing state-of-the-art hill fire segmentation methods, and Figure 8 displays the results of the segmentation. It can be seen that The paper's suggested network model can better segment the boundary between the hill fire and the background,

that is because of the bidirectional attention module, which not only assigns the weights of the different segmentation categories, but also pays attention to the degree of correlation between the neighbouring pixels, and the FFM also takes into account the localisation accuracy of the individual pixels, which is why it achieves a better effect of boundary segmentation.

Figure 8 Example segmentation effect diagram (see online version for colours)

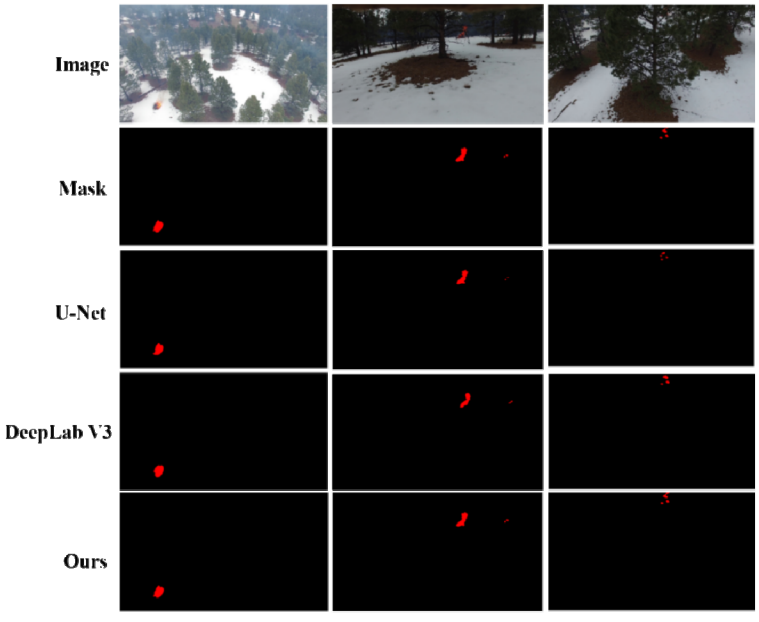


Table 1 shows the performance comparison in terms of MPA, MIOU and FPS metrics using the methods in this paper and other state-of-the-art methods. As we can see from the table, MPA and MIOU can reach the segmentation accuracy of traditional semantic segmentation networks while the segmentation speed is improved by a factor of 1.7, and compared with Xception, a lightweight network, MPA and MIOU are improved by 3 percentage points and 1 percentage point, respectively, while the segmentation speed is also slightly improved. Thus the method proposed in this paper achieves higher segmentation accuracy with the fastest segmentation speed. In other words, our network model can not only segment the fire and forest background effectively, but also ensure the segmentation speed of the image. This is because the STDC block does not increase a large number of parameters while acquiring multi-scale feature information, while the design of the BAMand the FFM does not take up too much computational resources.

Table 1 Segmentation performance metrics for different networks

Models	*MPA (%)*	*MIOU (%)*	*FPS*
U-Net	90.02	86.79	26.3
DeeplabV3	92.09	87.75	24.4
Xception + deeplabV3	91.40	86.49	62.2
Ours	94.4	87.40	65.4

In order to verify the performance enhancement effect of each module, we did ablation experiments. We use three different backbone networks, VGG16, RESNet and STDCNet, to perform the segmentation test, respectively, and the segmentation results are shown in Table 2, which shows that STDCNet performs slightly better than the other methods in MPA and MIOU, but the segmentation speed is 10 frames higher than that of the lightweight network, ResNet, and is therefore more suitable for the real-time application environment of hill fire monitoring in this paper.

Table 2 Ablation experiments for different backbone networks

Basic module	*MPA (%)*	*MIOU (%)*	*FPS*
VGG16	93.15	87.03	28.6
ResNet	94.20	86.75	55.2
STDCNet	94.40	87.40	65.4

Table 3 Ablation experiments with two-way attention modules

Methods	*MPA (%)*	*MIOU (%)*	*FPS*
STDCNet + FFM	90.31	85.60	65.1
STDCNet + BAM + FFM	94.40	87.40	65.4

Table 4 Ablation experiments with feature fusion module

Methods	*MPA (%)*	*MIOU (%)*	*FPS*
STDCNet + BAM	91.26	86.36	65.6
STDCNet + BAM + FFM	94.40	87.40	65.4

Tables 3 and 4 show the results of the ablation experiments for the BAM and the FFM, respectively. from Table 3, it can be seen that the BAM module did not reduce the FPS of the model and that the MIOU was improved by 2.1% and the MPA by 4.1%. From Table 4, it can be seen that the FFM module did not decrease the FPS of the model, and the MIOU was improved by 1.1% and the MPA by 3.2%. Therefore, it shows that the BAM module and the FFM module can improve the accuracy of fire segmentation while keeping the segmentation speed.

4 Conclusions

The network proposed in this paper can realise real-time segmentation of forest fire images with high accuracy in order to effectively reduce the losses caused by forest fires. First, the network uses a specialised real-time semantic segmentation model STDCNet as a benchmark model. Second, a BAMis proposed, which can simultaneously take into account information from both channel features and pixel space, enhancing the network's boundary segmentation of flame and forest background, and improving the segmentation ability for small-scale targets. Finally, the FFM is utilised to efficiently fuse the deep semantic information and shallow spatial information, and at the same time, fully connected CRFs are added to the output of the fused features to further improve the segmentation capability of the network for small targets. In this paper, from the selection of the benchmark model, as well as the design of each module, we avoid increasing the

number of parameters of the model too much, so as to ensure the segmentation speed of the model and meet the real-time requirements of fire segmentation. Comparing the model suggested in this research to the current state-of-the-art techniques, the experimental findings demonstrate that the model achieves higher segmentation accuracy with reduced time consumption. In fire segmentation, it is difficult to correctly segment the fire seedlings obscured by vegetation, and in future work, we can consider fusing time series in the model to further enhance the segmentation performance of the model.

References

Agarwal, P.S., Ghadge, P.M., Malapure, R.P. and Hedau, S.J. (2024) 'Elevating large-scale forest surveillance: a deep learning analysis of inception V3 and Efficientnet for IoT-driven fire detection', *Pattern Recognition and Image Analysis*, Vol. 56, No. 8, pp.653–688.

Akyol, K. (2024) 'A comprehensive comparison study of traditional classifiers and deep neural networks for forest fire detection', *Pattern Recognition and Image Analysis*, Vol. 27, No. 2, pp.1201–1215.

Aliser, A. and Duranay, Z.B. (2024) 'Fire/flame detection with attention-based deep semantic segmentation', *Pattern Recognition and Image Analysis*, Vol. 48, No. 3, pp.705–717.

Arlovi, M., Patel, M., Balen, J. et al. (2024) 'F2M: ensemble-based uncertainty estimation model for fire detection in indoor environments', *Engineering Applications of Artificial Intelligence*, Vol. 133, No. 5, p.108428.

Chen, H., Wang, Z., Qin, H. and Mu, X. (2024) 'DHFNet: decoupled hierarchical fusion network for RGB-T dense prediction tasks', *Neurocomputing*, Vol. 583, No. 10, p.127594.

Chen, L.C., Papandreou, G., Schroff, F. et al. (2017) *Rethinking Atrous Convolution for Semantic Image Segmentation*, arxiv preprint arxiv: 1706.05587.

Fan, M., Lai, S., Huang, J. et al. (2021) 'Rethinking BiSeNet, for real-time semantic segmentation', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vol. 576, No. 12, pp.9716–9725.

Goay, C.H., Ahmad, N.S. and Goh, P. (2023) 'Temporal convolutional networks for transient simulation of high-speed channels', *Alexandria Engineering Journal*, Vol. 74, No. 3, pp.643–663.

Gupta, H. and Nihalani, N. (2024) 'An efficient fire detection system based on deep neural network for real-time applications', *Pattern Recognition and Image Analysis*, Vol. 1, No. 2, pp.1–14.

Hu, X. J. Feng, J. Gong, LFFNet: lightweight feature-enhanced fusion network for real-time semantic segmentation of road scenes. Pattern Recognition and Image Analysis, 27, (2024).

Karandikar, A.M., Agrawal, A.J. and Welekar, R.R. (2024) 'Decadal forest cover change analysis of the tropical forest of Tadoba-Andhari India', *Pattern Recognition and Image Analysis*, Vol. 18, No. 2, pp.1705–1714.

Li, Y., Zhang, W., Liu, Y. et al. (2022) 'A visualized fire detection method based on convolutional neural network beyond anchor', *Applied Intelligence*, Vol. 52, No. 13, pp.13280–13295.

Liang, Y., Hang, T., Chen, J. et al. (2021) 'MSPPNet: a lightweight network for real-time semantic image segmentation', *Journal of Physics: Conference Series.*

Ou, Z., Bai, J., Chen, Z. et al. (2024) 'RTSeg-net: a lightweight network for real-time segmentation of fetal head and pubic symphysis from intrapartum ultrasound images', *Pattern Recognition and Image Analysis*, Vol. 175, No. 1, p.108501.

Paszke, A., Chaurasia, A., Kim, S. and Culurciello, E. (2016) 'ENet: a deep neural network architecture for real-time semantic segmentation', arxiv preprint arxiv: 1606.02147.

Qu, S., Wang, Z., Wu, J. et al. (2024) 'FBRNet: a feature fusion and border refinement network for real-time semantic segmentation', *Pattern Recognition and Image Analysis*, Vol. 27, No. 1, p.2.

Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P.Z. and Blasch, E. (2020) 'Aerial imagery pile burn detection using deep learning: the FLAME dataset', *Computer Networks*, Vol. 193, No. 5, p.108001.

Simonyan, K. and Zisserman, A. (2014) 'Very deep convolutional networks for large-scale image recognition', *Computer Science*, arxiv preprint arxiv: 1409.1556.

Turker, A. and Eksioglu, E.M. (2022) 'A fully convolutional encoder-decoder network for moving object segmentation', *International Conference on INnovations in Intelligent Systems and Applications* (INISTA), pp.1–6.

Wang, Y., Wang, Y., Xu, C. et al. (2024) 'Computer vision-driven forest wildfire and smoke recognition via IoT drone cameras', *Wireless Networks*, Vol. 18, No. 10, pp.1–14.

Wang, Y., Wang, Y., Xu, C. et al. (2024) 'Computer vision-driven forest wildfire and smoke recognition via IoT drone cameras', *Pattern Recognition and Image Analysis*, Vol. 18, No. 10, pp.1–14.

Weng, W. and Zhu, X. (2021) 'INet: convolutional networks for biomedical image segmentation', *IEEE Access*, Vol. 9, No. 2, pp.16591–16603.

Xu, Q., Ma, Y., Wu, J. et al. (2021) 'Faster BiSeNet: a faster bilateral segmentation network for real-time semantic segmentation', in *2021 International Joint Conference on Neural Networks* (IJCNN), Vol. 11, No. 3, pp.1–8.

Xue, J., Dai, Y., Wang, Y. and Aili, Q. (2023) 'Multiscale feature extraction network for real-time semantic segmentation of road scenes on the autonomous robot', *International Journal of Control, Automation and Systems*, Vol. 21, No. 6, pp.1993–2003.

Zedda, L., Loddo, A. and Di Ruberto, C. (2024) 'FIRESTART: fire ignition recognition with enhanced smoothing techniques and real-time tracking', *Pattern Recognition and Image Analysis*, Vol. 16, No. 1, pp.282–293.

Zhao, H., Jin, J. and Wang, L. (2022) 'A pulsar search method combining a new feature representation and convolutional neural network', *The Astrophysical Journal*, Vol. 929, No. 1, pp.18–25.

Zhou, Z.H. and Feng, J. (2017) 'Deep forest: towards an alternative to deep neural networks', *Twenty-Sixth International Joint Conference on Artificial Intelligence*.

Zhu, Y., Zhu, B., Chen, Y. and Wang, J. (2023) 'Uncertainty-aware boundary attention network for real-time semantic segmentation', *Pattern Recognition and Image Analysis*, Vol. 825, No. 7, pp.388–400.