forests

MDPI

Article

Fire in Focus: Advancing Wildfire Image Segmentation by Focusing on Fire Edges

Guodong Wang [1], Fang Wang [2,*], Hongping Zhou [1] and Haifeng Lin [1,*]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; guodong@njfu.edu.cn (G.W.); hpzhou@njfu.edu.cn (H.Z.)
[2] College of Electronic Engineering, Nanjing XiaoZhuang University, Nanjing 211171, China
* Correspondence: wangfang0182217@njxzu.edu.cn (F.W.); haifeng.lin@njfu.edu.cn (H.L.); Tel.: +86-25-8542-7827 (H.L.)

Abstract: With the intensification of global climate change and the frequent occurrence of forest fires, the development of efficient and precise forest fire monitoring and image segmentation technologies has become increasingly important. In dealing with challenges such as the irregular shapes, sizes, and blurred boundaries of flames and smoke, traditional convolutional neural networks (CNNs) face limitations in forest fire image segmentation, including flame edge recognition, class imbalance issues, and adapting to complex scenarios. This study aims to enhance the accuracy and efficiency of flame recognition in forest fire images by introducing a backbone network based on the Swin Transformer and combined with an adaptive multi-scale attention mechanism and focal loss function. By utilizing a rich and diverse pre-training dataset, our model can more effectively capture and understand key features of forest fire images. Through experimentation, our model achieved an intersection over union (IoU) of 86.73% and a precision of 91.23%. This indicates that the performance of our proposed wildfire segmentation model has been effectively enhanced. A series of ablation experiments validate the importance of these technological improvements in enhancing model performance. The results show that our approach achieves significant performance improvements in forest fire image segmentation tasks compared to traditional models. The Swin Transformer provides more refined feature extraction capabilities, the adaptive multi-scale attention mechanism helps the model focus better on key areas, and the focal loss function effectively addresses the issue of class imbalance. These innovations make the model more precise and robust in handling forest fire image segmentation tasks, providing strong technical support for future forest fire monitoring and prevention.

Keywords: forest fire; Swin Transformer; adaptive multi-scale attention mechanism (ASA); semantic segmentation; wildfire monitoring

Citation: Wang, G.; Wang, F.; Zhou, H.; Lin, H. Fire in Focus: Advancing Wildfire Image Segmentation by Focusing on Fire Edges. *Forests* **2024**, *15*, 217. https://doi.org/10.3390/f15010217

Academic Editor: Palaiologos Palaiologou

Received: 11 December 2023
Revised: 5 January 2024
Accepted: 20 January 2024
Published: 22 January 2024

1. Introduction

With the intensification of global climate change, the frequent occurrence of wildfires has become a major challenge worldwide [1]. These fires not only cause destruction to forest ecosystems [2] but also pose a severe threat to human living environments. In the process of addressing this challenge, traditional forest fire prevention and control methods, such as fire monitoring, source localization, and manual and physical fire extinguishing, are effective to some extent [1,3]. However, they still face many challenges in large-scale or hard-to-reach areas. One of the main challenges is the accurate identification and analysis of the fire front of surface fires, i.e., the boundary of the fire spreading on the ground [4]. This boundary is key to understanding fire behavior and developing effective response measures. The challenges in identifying the fire front include the irregularity of the flame shape, changes in size, and interference from complex backgrounds [5]. These factors make effective fire detection and management crucial to mitigate the impacts of these disasters [6,7].

In recent years, deep learning technology has shown immense potential in forest fire detection. Its efficient image processing and pattern recognition capabilities are particularly suited for analyzing large volumes of remote sensing data [8], such as satellite imagery and photos taken by drones [9], which are crucial for monitoring and responding to forest fires. Satellite images offer a wide-range monitoring perspective and are an important tool for quickly assessing the extent and impact of fires. Utilizing deep learning models, especially convolutional neural networks (CNNs), can effectively detect fire hotspots in satellite imagery [10]. On the other hand, drones provide a flexible observation angle and high-resolution images which are helpful in identifying early signs of fires and monitoring flame spread. Deep learning models, particularly those based on CNN architectures, can perform real-time analysis of drone images and rapidly identify fire-affected areas [11]. These models excel in fire detection because they can learn complex features from vast amounts of training data and handle uncertainties and variations in images, such as different lighting conditions and complex terrain backgrounds. Furthermore, applying deep learning technology not only improves the early detection rate of fires but also helps in devising effective firefighting strategies. By analyzing fire spread patterns and potential risk areas, deep learning models provide valuable information for fire departments to optimize resource allocation and decision making [12].

Despite significant achievements in fire detection using deep learning technology [1], there remains a substantial research gap in the field of wildfire image segmentation, particularly in the critical aspect of accurately segmenting the fire front. Compared to traditional fire detection methods, wildfire image segmentation technology can provide more detailed information about the fire, such as precisely locating the flames, assessing the extent of the fire, and predicting its spread rate. This information is crucial for devising effective prevention and control strategies and for the rational allocation of fire prevention resources [13].

However, the detailed processing involved in wildfire image segmentation faces numerous challenges, with the goal being to accurately distinguish between flame and non-flame areas at the pixel level [14]. Current research in this field is still in its early stages, partly due to the relatively limited application of advanced technologies like semantic segmentation and instance segmentation in forest fire detection [15]. The high precision requirements of these technologies demand higher standards of data quality and also pose challenges to the performance and generalization capabilities of models [1–16]. Given that the identification of the fire front is crucial for early warning systems and the formulation of firefighting strategies, developing technologies that can more precisely identify and analyze the fire front has become an urgent need in this field [17].

To address these challenges, this study proposes several technological innovations: First, we adopt a transfer learning approach. In the field of wildfire image segmentation, the applicability of transfer learning is evident in its ability to utilize the learning outcomes of existing models on large datasets, thereby accelerating and enhancing the model's performance on specific wildfire images. Second, we employ the Swin Transformer as the model's backbone to initially extract features. This novel network structure is more suited for processing complex patterns in images, particularly in identifying irregular and fine flame edges. Finally, we incorporated an adaptive multi-scale attention mechanism module and fine-tuned the model using focal loss. This helps the model focus more on hard-to-identify areas during flame segmentation, improving overall segmentation precision and robustness.

This study aims to enhance the identification accuracy of the fire front in forest fires by improving existing semantic segmentation models through deep learning technology. Our model, with its refined flame edge segmentation technique, accurately identifies the fire line in complex wildfire images. This improvement not only increases the detection rate of the fire front but also enhances the model's ability to estimate the size of the fire and its spreading tendency. This provides a powerful tool for early detection, behavior analysis, and emergency response in forest fire situations.

2. Materials and Methods

2.1. Dataset Preparation

In the development of the wildfire image segmentation model, we first recognized that the quality of the model's training and prediction is closely related to the quality of the dataset. To ensure high-accuracy results, our research team manually annotated about 600 high-quality labeled images. Regarding the ratio of fire pixels to the background, we have maintained a balance where fire pixels constitute approximately 20%–30% of the image. This ratio was chosen to provide a realistic representation of wildfire scenarios, capturing both the fire and its surrounding environment effectively for segmentation analysis. Through this detailed annotation process, we were able to mark the edges of the flames more meticulously, as shown in Figure 1. For the training, testing, and validation of our model, we divided the dataset as follows: 70% of the images were used for training, 20% for validation, and the remaining 10% for testing. The details of the forest fire classification dataset are shown in Table 1. Such annotation work is crucial for subsequent model training and performance enhancement.

Figure 1. Detailed annotation process of flame edges.

Table 1. Details of our dataset.

Dataset	Resolution	Train	Val	Test
forest fire dataset	512 × 512	370	106	52
not containing wildfire forest data	512 × 512	50	14	8

However, training a new network structure from scratch often leads to suboptimal results. To address this issue, we adopted a transfer learning strategy. Specifically, we used the Common Objects in Context (COCO) public dataset as the pre-training dataset. The COCO Stuff Segmentation dataset has already demonstrated excellent performance in semantic segmentation, so a model pre-trained on the COCO dataset can learn to extract key visual features from various natural landscapes. By retaining these general visual features learned during the pre-training phase, we were able to accelerate model training and enhance its generalization capability on a new task—wildfire image segmentation [18].

2.2. Transfer Learning Strategy

In our research, transfer learning is one of the core strategies employed to enhance the performance of wildfire image segmentation tasks. We have chosen to start with a model pre-trained on the COCO dataset, leveraging its excellent performance in semantic segmentation. Pre-training on the COCO dataset enables the model to learn the ability to extract key visual features from various natural landscapes. This means that during the transfer learning process, the primary layers of the model (usually those closer to the input) are "frozen"; i.e., the weights of these layers remain unchanged during further training. These primary layers are typically responsible for extracting more general features, such as edges, textures, and colors. By freezing these layers, the model can retain the general visual features learned during the pre-training phase, thereby speeding up training and enhancing the model's generalization ability on new tasks. Subsequently, we transitioned the model to the wildfire image segmentation task.

2.3. FFDeepLab Model with Swin Transformer Backbone

In this study, we propose a deep learning model, named Fire-in-Focus DeepLab (FFDeepLab), specifically designed for wildfire image segmentation tasks. The FFDeepLab model is based on an encoder–decoder architecture [1]. In this architecture, the encoder part is responsible for extracting features from the input image, while the decoder part reconstructs the spatial details of the image, based on these features, for precise segmentation. Following the feature extraction stage of the model is the atrous spatial pyramid pooling (ASPP) module [19]. ASPP, by applying a series of convolutions with different dilation rates, captures image information at various scales, effectively increasing the model's receptive field. Additionally, FFDeepLab incorporates short connections in its architecture, which combine low-level image details and high-level semantic information extracted from the encoder, to generate the final image segmentation result [20]. This design aids the model in performing effective semantic segmentation while maintaining image details. Figure 2 illustrates the structure of the FFDeepLab model, with a detailed introduction to its key components provided in subsequent subsections.

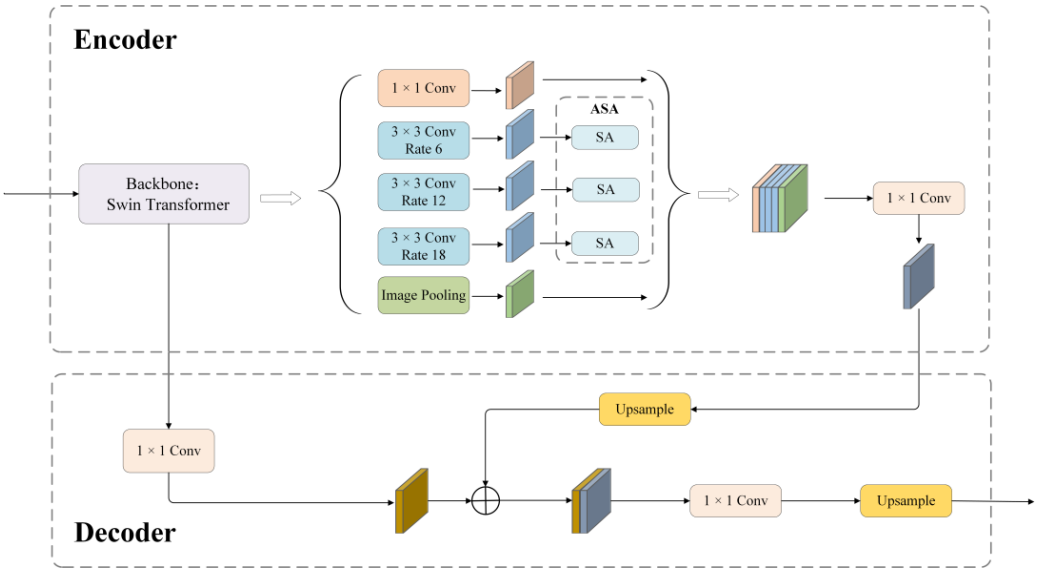

Figure 2. Structure of the FFDeepLab model.

2.3.1. Backbone

In our research, we developed the FFDeepLab model specifically for wildfire image segmentation tasks, utilizing the Swin Transformer as the backbone network for feature extraction [21]. The Swin Transformer, an advanced visual transformer, breaks through the limitations of traditional convolutional neural networks and has proven effective in various visual tasks such as image classification, object detection, and semantic segmentation. One of its core innovations is the shifted window multi-head self-attention (SW-MSA) mechanism, which implements self-attention calculation across consecutive layers through shifted windows. This not only calculates self-attention within local windows but also allows information transfer between windows, significantly reducing computational complexity and enhancing the model's capability to capture image features.

In the FFDeepLab model, the Swin Transformer divides the input RGB images into non-overlapping patches, each serving as a "token", with its features composed of the original pixel RGB values. This patch-based approach allows for more efficient processing of image data, as each patch is treated independently, enabling parallel processing and reducing the overall computational load. These patch tokens are then processed through shifted window-based multi-head self-attention (SW-MSA), enabling the model to capture rich contextual information. The SW-MSA is pivotal in enhancing the model's ability to discern finer details within each patch, thus improving its accuracy in identifying and

segmenting fire in images. Each Swin Transformer block primarily consists of SW-MSA and processes the self-attention mechanism through the shifted window strategy, reducing computation while maintaining performance. This unique combination of SW-MSA and shifted window strategy is key to balancing computational efficiency with high segmentation performance. With this approach, the Swin Transformer in FFDeepLab generates hierarchical feature representations at different resolutions, meeting the requirements for feature extraction at various scales [22]. For instance, in wildfire image segmentation tasks, the model can capture detailed features at lower levels, such as the edges and shapes of flames, while understanding the distribution and contextual relationships of the flames in the overall scene at higher levels. This dual-level analysis is particularly effective in differentiating between fire and non-fire elements in varied environmental backgrounds. Figure 3a presents a detailed schematic of the backbone structure, while Figure 3b focuses on a simplified depiction of the Swin Transformer Block, which is a crucial element of the backbone. Additionally, Figure 3c,d specifically illustrate the configurations of the patch partition module and the patch merging module, respectively, showcasing their integral roles in the model's architecture. The patch partition module is responsible for the initial division of the image into patches, whereas the patch merging module plays a key role in combining features from these patches to form a comprehensive image representation.

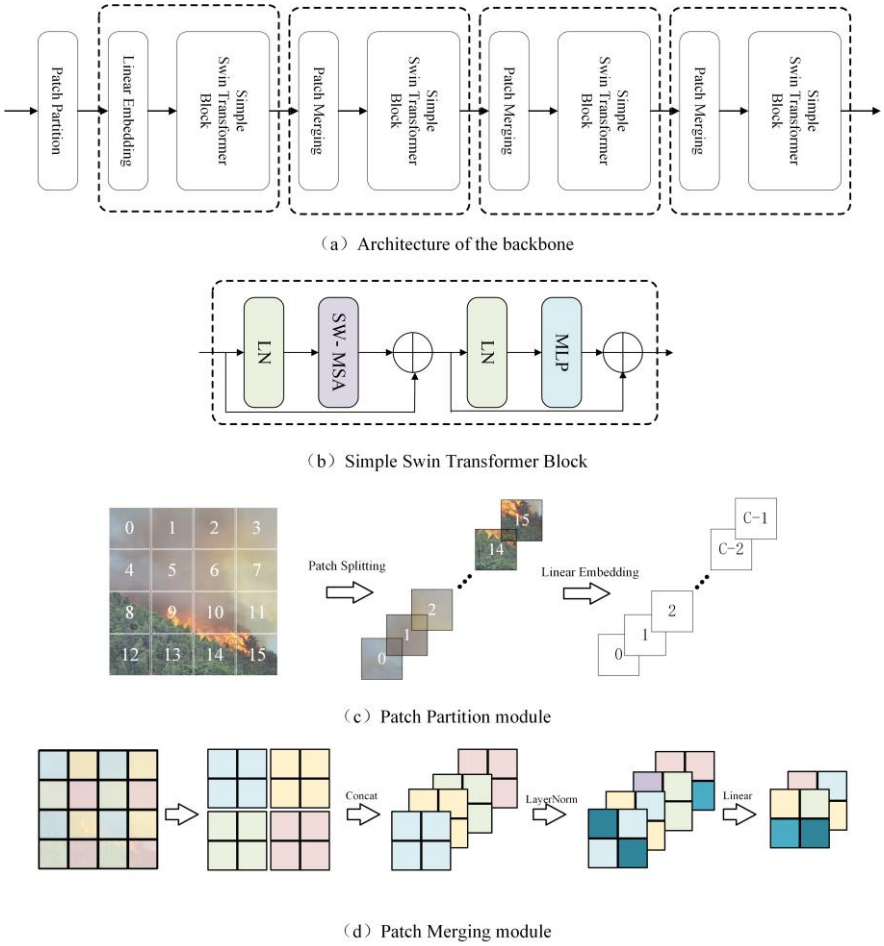

Figure 3. Structure of the backbone.

2.3.2. Adaptive Multi-Scale Attention Mechanism (ASA)

Further, to strengthen the model's capability in handling multi-scale information, we integrated an adaptive multi-scale attention module at the end of each branch of the atrous spatial pyramid pooling (ASPP). ASPP, a core component of the DeepLab architecture series, effectively captures multi-scale information of images using atrous convolutions

with different dilation rates while maintaining the original resolution. After each feature map produced by a specific dilation rate, we introduce an adaptive attention module. The adaptive attention module tailors the model's focus to the most relevant features at each scale, significantly improving segmentation precision. This module utilizes a small neural network to calculate attention weights for the corresponding scale feature map. For each branch of the ASPP module with a different dilation rate, a feature map, F_i, is produced, where i denotes different dilation rates. For each generated F_i, separate convolutional layers are used to generate the Q_i, K_i, V_i values. These values are instrumental in the adaptive attention process, enabling the model to dynamically adjust its focus based on the unique characteristics of each feature map. Attention scores are calculated using Q_i and K_i, and the Softmax function is applied to normalize the scores. The attention-weighted values are then used. To preserve the original feature information, a residual connection is added to combine the original feature map, retaining the input feature map's information and incorporating the features weighted by self-attention. This residual connection is crucial for maintaining the integrity of the original data while enhancing them with the model's learned attention-focused features. The feature maps processed through self-attention are fused to form the final output of the ASPP module. The formula represents the calculation process for the attention scores [23]. Figure 4 illustrates the specific workflow of the ASPP process after integrating the ASA module. Equations (1) and (2) detail the calculation methodology for the ASA module:

$$Q_i = \mathrm{Conv}_{Q_i}(F_i),\ K_i = \mathrm{Conv}_{K_i}(F_i),\ V_i = \mathrm{Conv}_{V_i}(F_i) \tag{1}$$

$$\mathrm{Attention}(Q_i, K_i, V_i) = \mathrm{softmax}\left(\frac{Q_i {K_i}^T}{\sqrt{d_k}}\right) V \tag{2}$$

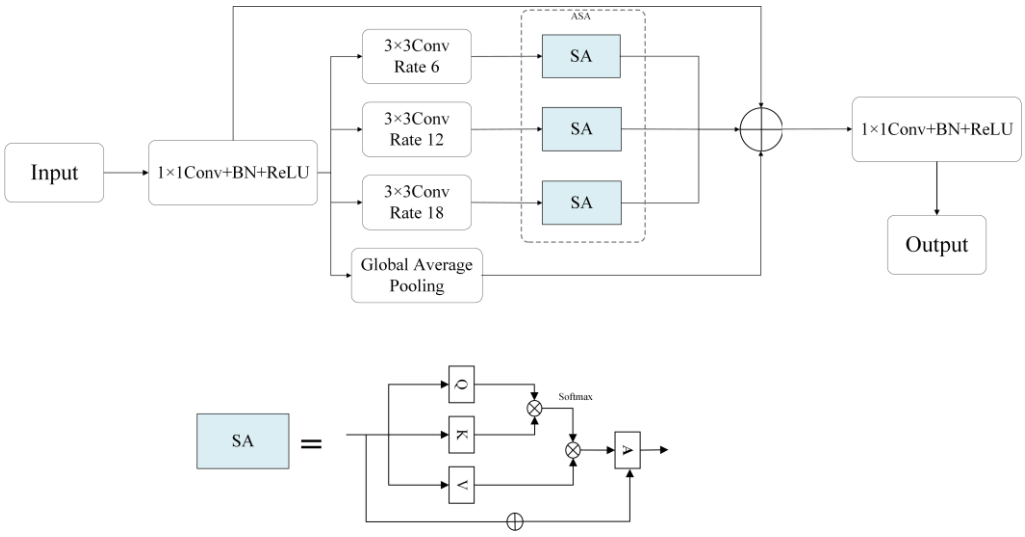

Figure 4. ASPP integrating the ASA module.

2.3.3. Focal Loss

Traditional image segmentation models like DeepLabV3 typically use cross-entropy loss to optimize the model [24]. This loss function measures the difference between the probability distribution predicted by the model and the actual labels at each pixel level. However, when dealing with imbalanced data like wildfire images, cross-entropy loss might not be sufficient to make the model focus on less frequent but more important classes (such as the flame areas). The focal loss is designed to address a specific but crucial problem [25]: how to make the model pay more attention to less numerous classes in the presence of extremely imbalanced class distributions. Its core idea is to reduce focus on easily classified samples, thereby directing more attention to those that are difficult to

classify. Specifically, it modifies the cross-entropy loss function by adding a modulating factor that reduces the weight of easily classified samples. In the context of wildfire image segmentation, the flame areas might occupy only a small part of the entire image, with the rest being mostly non-flame areas. In such cases, traditional cross-entropy loss might lead to the model overfitting the large non-flame areas while neglecting the relatively smaller flame areas. Equation (3) is provided for the improved loss function, where p_t is the model's probability of correctly classifying each pixel, a_t is a balancing parameter used to balance the importance between different classes (flame and non-flame), and γ is the modulating factor used to reduce the weight of easily classified samples. Equation (3) outlines the calculation method for focal loss:

$$FL(p_t) = -\alpha_t(1 - p_t)^{\gamma}\log(p_t) \quad (3)$$

3. Results

3.1. Pre-Training Results

In this section, we present the loss curves of the model with and without the pre-trained weights from the COCO dataset. The loss curves demonstrate the model's learning progress, indicating its gradual adaptation to the general features of semantic segmentation tasks in natural landscapes. This is a crucial step before applying the pre-trained weights to wildfire image segmentation, providing insights into the challenges and adjustments that might arise in the subsequent transfer learning phase. By thoroughly examining these results, we lay a solid foundation for the model's precise transition to wildfire segmentation tasks [1,26].

Based on Figure 5, we observe significant improvements in the model during the pre-training phase, especially in terms of reduction in training loss. This indicates that the model has made significant progress in recognizing and processing image features in complex datasets. From these analyses, we can reasonably infer that the model will exhibit higher efficiency and accuracy in future transfer learning processes, particularly in semantic segmentation specifically targeted at forest fire images.

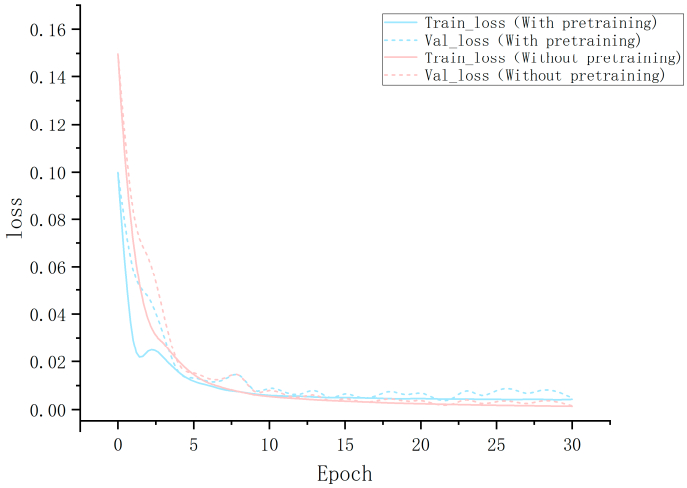

Figure 5. Comparison of loss curves during training.

3.2. Wildfire Image Segmentation Performance

In this section, we focus on showcasing the performance of the model on the specially prepared wildfire image dataset, including the results of key indicators and a comparison with baseline performance.

3.2.1. Evaluation Metrics

In this part of our study, we focus on evaluating the performance of the improved model on a dataset specifically prepared for wildfire image segmentation tasks. To com-

prehensively understand the model's performance, we have chosen a range of key performance indicators: intersection over union (IoU), precision, and recall. These metrics collectively provide detailed insights into the model's ability to accurately identify and segment wildfire areas.

Intersection over union (IoU): IoU is one of the most important metrics for evaluating image segmentation tasks. It measures accuracy by calculating the overlap ratio between the predicted segmentation area and the actual segmentation area. In the context of wildfire image segmentation, a high IoU value indicates that the model can accurately identify flame areas.

Precision: precision refers to the proportion of pixels correctly identified as flame out of all pixels identified as flame. High precision means the model has a lower false positive rate when marking flame areas.

Recall: recall measures the proportion of pixels correctly identified as flame out of all actual flame pixels. A high recall indicates that the model can capture most of the actual flame areas, reducing missed detections.

The calculation equations are as follows:

$$IoU = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{4}$$

$$\text{Precision} = \frac{TP_i}{TP_i + FP_i} \tag{5}$$

$$\text{Recall} = \frac{TP_i}{TP_i + FN_i} \tag{6}$$

3.2.2. Performance Analysis and Comparison

In this section, we delve into a detailed analysis and comparison of the model's performance on the wildfire image dataset. By thoroughly evaluating the model's performance on key metrics, we aim to reveal the specific effects of model improvements and compare these results with baseline performances. This comparative analysis not only showcases the overall efficacy of our model in wildfire detection and segmentation tasks but also provides insights into how various technical enhancements collectively contribute to improving the model's performance.

In the comparative experiment section, we will present a performance comparison of FFDeepLab with other advanced models in wildfire image segmentation tasks. This will help in assessing our model's relative performance in the current technological landscape.

Original DeepLabV3: a benchmark for semantic image segmentation that utilizes atrous convolution and atrous spatial pyramid pooling (ASPP) to capture multi-scale context [26].

FCN (fully convolutional network): A foundational model for semantic segmentation that replaces fully connected layers with convolutional layers to enable pixel-level predictions [27].

SegNet: a network primarily designed for tasks like road scene understanding and indoor object segmentation, known for its efficient use of memory during training [28].

PSPNet (pyramid scene parsing network): distinguished by its pyramid pooling module, PSPNet is effective in aggregating global context information for scene parsing tasks [29].

By comparing these models, we aim to highlight the strengths and weaknesses of our model in the specific context of wildfire image segmentation. The comparison will focus on how each model performs in terms of IoU, precision, and recall, and will provide a comprehensive understanding of where our model stands in relation to these well-established segmentation architectures. This comparative study is crucial for validating the improvements made in our model and for identifying potential areas for further advancements. To provide a clear visual comparison of the performance disparities among

various models, we compared the prediction results of several models, as shown in Figure 6. Table 2 details the evaluation metrics for different models.

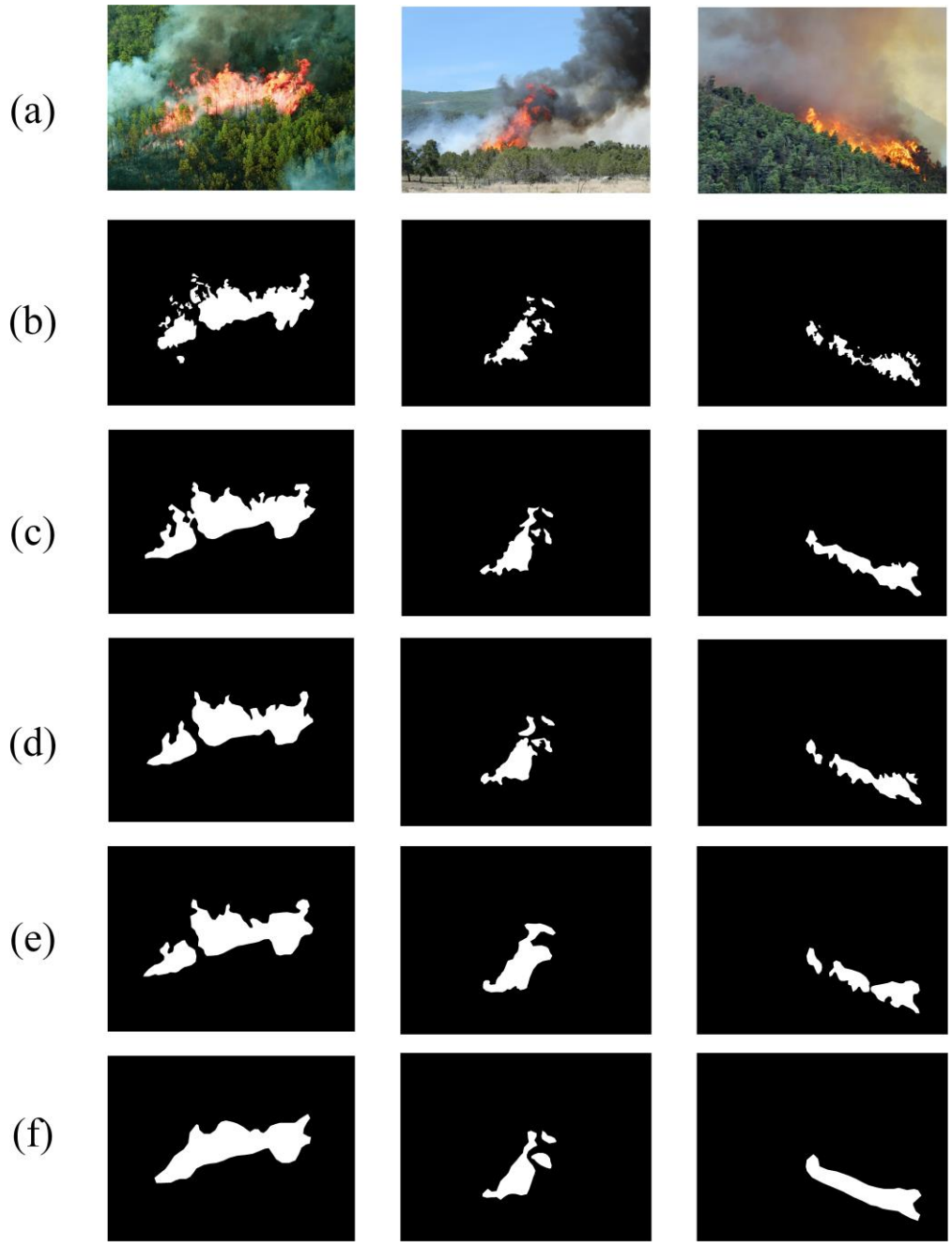

Figure 6. Visual comparison of prediction results from various models: (**a**) Image; (**b**) FFDeepLab (ours); (**c**) PSPNet; (**d**) DeepLabV3; (**e**) SegNet; (**f**) FCN.

Table 2. Performance comparison of different models on fire segmentation task.

Model	Metrics		
	IoU	Precision	Recall
DeepLabV3	84.50%	89.00%	85.50%
FCN	73.20%	78.73%	80.50%
SegNet	77.80%	82.30%	78.90%
PSPNet	85.00%	88.50%	86.00%
FFDeepLab (Ours)	86.73%	91.23%	87.94%

3.3. Ablation Experiments

To comprehensively evaluate the contributions and effectiveness of each component in our proposed wildfire image segmentation model, this study has designed a series of exhaustive ablation experiments. These experiments aim to individually explore the impact of the Swin Transformer backbone network, the introduction of the adaptive multi-scale attention mechanism, and the replacement of the traditional cross-entropy loss function with the focal loss function on the model's performance.

Specifically, we first compare the use of the Swin Transformer with traditional convolutional neural networks as the backbone network. This comparison is meant to demonstrate the utility of the Swin Transformer in wildfire image segmentation tasks. The effectiveness of the Swin Transformer is gauged by its ability to handle the complexities of wildfire images more efficiently than standard CNN architectures. Next, we investigate the impact of incorporating an adaptive multi-scale attention mechanism into the ASPP module on the model's segmentation performance. This step aims to assess how this attention mechanism improves the model's ability to focus on relevant features at different scales, thus enhancing its segmentation accuracy. Finally, to validate the advantages of the newly adopted focal loss function over standard cross-entropy loss in addressing the issue of class imbalance, we conduct comparative experiments on the loss functions. This comparison is critical to demonstrate how the focal loss function better equips the model to focus on less frequent but crucial classes, such as flame areas in wildfire images. Through these ablation experiments, we can meticulously showcase how each improvement enhances the overall performance of the model, thereby providing solid empirical evidence for our design choices. In the following sections, we will present detailed results of these experiments, along with corresponding analyses and discussions, to underscore the efficacy of each component in our model. As shown in Table 3, the performance of our model further confirms the effectiveness and superior performance of our approach.

Table 3. Performance of different model configurations on fire segmentation.

Backbone	ASA	Focus Loss	Metrics		
			IoU	Precision	Recall
MobileNetV2			78.43%	82.16%	79.58%
Xception			80.67%	84.22%	81.39%
Swin Transformer			82.35%	85.78%	83.21%
Swin Transformer	√		84.56%	87.03%	85.47%
Swin Transformer	√	√	86.73%	91.23%	87.94%

4. Discussion

This study introduces an innovative approach to wildfire image segmentation, leveraging the Swin Transformer and an adaptive multi-scale attention mechanism. These technological advancements mark a significant step forward in accurately identifying

and analyzing fire fronts in complex wildfire scenarios. The Swin Transformer, with its novel network structure, excels in processing intricate patterns in images, especially in identifying irregular and fine flame edges. The adaptive multi-scale attention mechanism further enhances the model's precision by focusing on challenging areas during flame segmentation.

The performance of the proposed model, FFDeepLab, shows a notable improvement over traditional semantic segmentation models. In comparison with established models such as Original DeepLabV3, FCN, SegNet, and PSPNet, FFDeepLab demonstrates superior capabilities in key metrics like IoU, precision, and recall. This superiority is particularly evident in the context of flame area identification and segmentation, highlighting the effectiveness of the integrated Swin Transformer and the adaptive attention mechanism.

The introduction of focal loss function in the FFDeepLab model addresses a critical challenge in wildfire image segmentation—the class imbalance between flame and non-flame areas. By focusing on less frequent but crucial classes, the model avoids overfitting to dominant non-flame areas, thereby enhancing its effectiveness in detecting and segmenting flame areas.

While this study marks a significant advancement, it acknowledges certain limitations. The dataset used, although meticulously annotated, is limited in size, potentially impacting the model's generalizability to diverse wildfire scenarios. Expanding this dataset to cover a broader spectrum of fire incidents, including those in different vegetation types and weather conditions, would enhance the robustness of our model. Future research could expand the dataset, incorporating a wider range of fire types and environmental conditions. Moreover, the computational demands of the Swin Transformer and adaptive multi-scale attention highlight the need for optimization to facilitate real-time application in wildfire monitoring and response systems. Efforts to optimize these components could lead to more efficient models suitable for deployment in real-time systems, where rapid processing is essential.

Lastly, integrating this model with other data sources, such as thermal imaging and real-time environmental data, could further improve accuracy and utility in practical applications.

5. Conclusions

This study aims to significantly improve wildfire image segmentation through a series of innovative technological advancements. Our approach focuses on three main aspects: introducing the Swin Transformer as the backbone network of the model, integrating an adaptive multi-scale attention mechanism, and employing the focal loss function to address the issue of class imbalance:

(1) Swin Transformer as the backbone network: by incorporating the Swin Transformer as the backbone network, our model can more effectively process and recognize complex features in wildfire images. This transformer-based approach demonstrates significant advantages in capturing image details, particularly in identifying small flames within images.

(2) Adaptive multi-scale attention mechanism: the addition of an adaptive multi-scale attention mechanism further enhances the model's ability to focus on important areas. This mechanism enables the model to more accurately differentiate between flames and background during segmentation, especially in cases with blurred edges and varying sizes.

(3) Focal loss function for class imbalance: we have adopted the focal loss function to optimize the training process of the model. This strategy effectively addresses the common issue of class imbalance in wildfire image segmentation tasks, improving the accuracy of the model in segmenting small flame areas.

Overall, our research not only provides an efficient method for wildfire image segmentation but also demonstrates how technological innovations can tackle challenges in complex visual tasks. Moreover, this model holds significant practical importance for forest fire management. More accurate identification of the fire front can help in early warning and prevention of forest fires, minimizing the damage to ecosystems and human settle-

ments. Additionally, the advancements in this technology offer new possibilities for future simulation and prediction of fire behavior, providing valuable guidance for improving disaster response strategies and resource allocation.

Author Contributions: G.W. was responsible for program design and drafting the initial manuscript. F.W. assisted with data collection and analysis. H.L. and H.Z. designed the project and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Research and Development Plan of Jiangsu Province (Grant No. BE2021716).

Data Availability Statement: Available on request.

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

References

1. Diffenbaugh, N.S.; Konings, A.G.; Field, C.B. Atmospheric Variability Contributes to Increasing Wildfire Weather but Not as Much as Global Warming. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2117876118. [CrossRef] [PubMed]
2. Overpeck, J.T.; Breshears, D.D. The Growing Challenge of Vegetation Change. *Science* **2021**, *372*, 786–787. [CrossRef] [PubMed]
3. Lin, H.; Qian, J.; Di, B. Learning for Adaptive Multi-Copy Relaying in Vehicular Delay Tolerant Network. *IEEE Trans. Intell. Transp. Syst.* **2023**. [CrossRef]
4. Ferreira, L.M.; Coimbra, A.P.; De Almeida, A.T. Autonomous System for Wildfire and Forest Fire Early Detection and Control. *Inventions* **2020**, *5*, 41. [CrossRef]
5. Resco De Dios, V.; Nolan, R.H. Some Challenges for Forest Fire Risk Predictions in the 21st Century. *Forests* **2021**, *12*, 469. [CrossRef]
6. Lin, H.; Han, Y.; Cai, W.; Jin, B. Traffic Signal Optimization Based on Fuzzy Control and Differential Evolution Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8555–8566. [CrossRef]
7. Lin, H.; Lin, J.; Wang, F. An innovative machine learning model for supply chain management. *J. Innov. Knowl.* **2022**, *7*, 100276. [CrossRef]
8. Priya, R.S.; Vani, K. Deep Learning Based Forest Fire Classification and Detection in Satellite Images. In Proceedings of the 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 18–20 December 2019; pp. 61–65.
9. Perez-Saura, D.; Fernandez-Cortizas, M.; Perez-Segui, R.; Arias-Perez, P.; Campoy, P. Urban Firefighting Drones: Precise Throwing from UAV. *J. Intell. Robot. Syst.* **2023**, *108*, 66. [CrossRef]
10. Kang, Y.; Jang, E.; Im, J.; Kwon, C. A Deep Learning Model Using Geostationary Satellite Data for Forest Fire Detection with Reduced Detection Latency. *GISci. Remote Sens.* **2022**, *59*, 2019–2035. [CrossRef]
11. Filkov, A.; Cirulis, B.; Penman, T. Quantifying Merging Fire Behaviour Phenomena Using Unmanned Aerial Vehicle Technology. *Int. J. Wildland Fire* **2021**, *30*, 197. [CrossRef]
12. Kim, B.; Lee, J. A Video-Based Fire Detection Using Deep Learning Models. *Appl. Sci.* **2019**, *9*, 2862. [CrossRef]
13. Ghali, R.; Akhloufi, M.A.; Souidene Mseddi, W.; Jmal, M. Wildfire Segmentation Using Deep-RegSeg Semantic Segmentation Architecture. In Proceedings of the International Conference on Content-Based Multimedia Indexing, Graz, Austria, 14 September 2022; pp. 149–154.
14. Frizzi, S.; Bouchouicha, M.; Moreau, E. Comparison of Two Semantic Segmentation Databases for Smoke Detection. In Proceedings of the 2021 22nd IEEE International Conference on Industrial Technology (ICIT), Valencia, Spain, 10 March 2021; pp. 856–863.
15. Guan, Z.; Miao, X.; Mu, Y.; Sun, Q.; Ye, Q.; Gao, D. Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model. *Remote Sens.* **2022**, *14*, 3159. [CrossRef]
16. Hu, X.; Ban, Y.; Nascetti, A. Uni-Temporal Multispectral Imagery for Burned Area Mapping with Deep Learning. *Remote Sens.* **2021**, *13*, 1509. [CrossRef]
17. De Andrade, R.B.; Mota, G.L.A.; Da Costa, G.A.O.P. Deforestation Detection in the Amazon Using DeepLabv3+ Semantic Segmentation Model Variants. *Remote Sens.* **2022**, *14*, 4694. [CrossRef]
18. Caesar, H.; Uijlings, J.; Ferrari, V. COCO-Stuff: Thing and Stuff Classes in Context. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1209–1218.
19. Qayyum, A.; Ahmad, I.; Mumtaz, W.; Alassafi, M.O.; Alghamdi, R.; Mazher, M. Automatic Segmentation Using a Hybrid Dense Network Integrated With an 3D-Atrous Spatial Pyramid Pooling Module for Computed Tomography (CT) Imaging. *IEEE Access* **2020**, *8*, 169794–169803. [CrossRef]
20. Yurtkulu, S.C.; Şahin, Y.H.; Unal, G. Semantic Segmentation with Extended DeepLabv3 Architecture. In Proceedings of the 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 24–26 April 2019; pp. 1–4. [CrossRef]
21. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 9992–10002.

22. Luo, H.; Zhang, S.; Lei, M.; Xie, L. Simplified Self-Attention for Transformer-Based end-to-end Speech Recognition. In Proceedings of the 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 19–22 January 2021; pp. 75–81. [CrossRef]
23. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. *arXiv* **2018**, arXiv:1803.02155.
24. Ho, Y.; Wookey, S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access* **2019**, *8*, 4806–4813. [CrossRef]
25. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]
26. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
27. Ataş, İ. Performance Evaluation of Jaccard-Dice Coefficient on Building Segmentation from High Resolution Satellite Images. *Balk. J. Electr. Comput. Eng.* **2023**, *11*, 100–106. [CrossRef]
28. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
29. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.