

## Article

# Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks

Ziqi Wang <sup>1</sup>, Tao Peng <sup>1,\*</sup> and Zhaoyou Lu <sup>2</sup>

<sup>1</sup> College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China; zqwang@nefu.edu.cn

<sup>2</sup> School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan 430205, China; 18346741020@163.com

\* Correspondence: pt@nefu.edu.cn; Tel.: +86-13945108517

**Abstract:** In recent years, frequent forest fires have plagued countries all over the world, causing serious economic damage and human casualties. Faster and more accurate detection of forest fires and timely interventions have become a research priority. With the advancement in deep learning, fully convolutional network architectures have achieved excellent results in the field of image segmentation. More researchers adopt these models to segment flames for fire monitoring, but most of the works are aimed at fires in buildings and industrial scenarios. However, there are few studies on the application of various fully convolutional models to forest fire scenarios, and comparative experiments are inadequate. In view of the above problems, on the basis of constructing the dataset with remote-sensing images of forest fires captured by unmanned aerial vehicles (UAVs) and the targeted optimization of the data enhancement process, four classical semantic segmentation models and two backbone networks are selected for modeling and testing analysis. By comparing inference results and the evaluation indicators of models such as mPA and mIoU, we can find out the models that are more suitable for forest fire segmentation scenarios. The results show that the U-Net model with Resnet50 as a backbone network has the highest segmentation accuracy of forest fires with the best comprehensive performance, and is more suitable for scenarios with high-accuracy requirements; the DeepLabV3+ model with Resnet50 is slightly less accurate than U-Net, but it can still ensure a satisfying segmentation performance with a faster running speed, which is suitable for scenarios with high real-time requirements. In contrast, FCN and PSPNet have poorer segmentation performance and, hence, are not suitable for forest fire detection scenarios.

**Keywords:** forest fires; deep learning; image processing; neural network; flame segmentation



**Citation:** Wang, Z.; Peng, T.; Lu, Z. Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks. *Forests* **2022**, *13*, 1133. <https://doi.org/10.3390/f13071133>

Academic Editor: Costantino Sirca

Received: 6 June 2022

Accepted: 15 July 2022

Published: 19 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The major causes of socio-economic losses and human casualties include traffic accidents [1,2], forest fires [3,4], and natural and geological disasters [5,6]. Among them, the ecological damage caused by forest fires is huge, irreversible and long-lasting compared to the economic damage caused to human society, even endangering the fragile global ecosystem. Forest fires occur suddenly and are wide-ranging; however, due to the lack of observation methods, forest fires are always detected after they have spread over a large scale, making them arduous to control and extinguish. Therefore, since the end of the last century, many scholars have carried out research on forest fire image recognition and monitoring technology.

Traditional forest fire detection methods include man-powered watchtowers, patrol aircraft, and remote-sensing satellites for woodland monitoring [7,8]. Among them, although the satellites have a wide range of applications, their detection periods are long and lack flexibility. In general, these methods not only require a large amount of resource input, but also are severely limited by the weather and environmental conditions, making it difficult

to ensure the quality of forest fire monitoring. With the advancement in image-processing technology, in recent years, authorities have increased the use of unmanned aerial vehicles (UAVs) and surveillance for aerial monitoring of forest areas [9,10], collecting and real-time-processing fire images for earlier warning and intervention. In order to identify forest fire areas in the images, flame detection techniques are introduced in some studies. In the early stage of the research, Habiboglu et al. [11] proposed a forest fire flame detection algorithm based on a covariance matrix by using color, space and event information. Wang et al. [12] proposed an adaptive fire-monitoring algorithm based on random testing and robust features to construct a forest fire classification model for fire identification. Although the above methods have achieved certain results in the field of forest fire identification, the overall process is complex, the accuracy of flame area detection is relatively low, and the scale and change of forest fire is difficult to quantify. In view of this, more researchers adopt flame segmentation algorithms for forest fire detection. Such algorithms can identify all the pixels of forest fire areas in an image and better detect and quantify a fire situation, and can be applied to fire trend prediction and intervention analysis to provide timely data support for forestry departments and authorities. Flame segmentation techniques mainly contain two categories: classical methods and deep-learning-based methods. Classical segmentation methods mainly rely on features such as color, spatial structure and texture to label the pixels of the flame area. TLIG et al. [13] proposed a corresponding image segmentation method based on principal component analysis and Gabor filter, which proposed a new superpixel extraction strategy, but it was insensitive to noise and needed to rely on manual feature design. The applicability in different scenarios was poor. T Toulouse et al. [14] extracted 200 million fire pixels and 700 million non-fire pixels in 500 images, and segmented the flame area based on two new rules and detection methods. Meanwhile, this work shows that a machine-learning-based segmentation technique outperforms the state-of-the-art techniques at the time, thus opening new ideas for fire detection research.

Semantic segmentation techniques based on deep learning can train network models based on annotation data and predict the category of each pixel to achieve end-to-end image segmentation. Among them, fully convolutional network models are widely used in the field. By testing the performances of different semantic segmentation models on public datasets, it was found that the detection accuracy of the fully convolutional network models is much better than that of previous methods [15–17]. Hence, researchers have introduced FCN and FCN-based models into the field of flame segmentation. Choi H. S. et al. [18] applied the FCN model to fire detection based on a FiSmo dataset. Cai C et al. [19] used the U-Net model to segment flames of a chain furnace. Although the above work successfully applied fully convolutional models to fire segmentation tasks, since the performance of the same model on different targets will be quite different, these models that perform well in respective scenarios are not necessarily suitable for forest fire segmentation. Hossain F. M. A. and Ghali R. [20,21] combined the U-Net model with UAV aerial photography technology, and proposed two efficient up-sampling methods to monitor and segment forest fires. Harkat H et al. [22] applied the DeepLabV3+ model to the forest fire detection field. However, most of the data used above are collected by traditional means based on optical sensors and surveillance cameras, and the data quality is lower than that captured by UAVs. Zhang et al. [23] developed an effective SqueezeNet-based asymmetric encoder–decoder U-shape architecture, mainly functioning as an extractor and a discriminator of a forest fire, which is relatively reliable with good accuracy and prediction time. M Shahid et al. [10] introduced the spatio-temporal network and proposed a two-stage cascaded architecture to improve accuracy, achieving better performance than the existing state-of-the-art methods. In general, these works only analyze from the perspective of a single model and backbone network, lacking comparisons between models, and therefore cannot compare the performance of different models in forest fire segmentation tasks. In addition, due to the relative complexity of some of the methods, it is difficult to promote them in practical forest fire detection applications due to the limited computational resources of the relevant departments in various regions. Großkopf J et al. [24] used DeepLabV3+ to conduct a

comparative analysis in industrial fire environments with different backbone networks, experimenting in specific scenarios. Although this study carried out a comparative study on different backbones, it still lacks a comparison with other models, and is only applicable to industrial scenarios. At present, due to the shortage of remote-sensing images for forest fire segmentation, there is little related research on forest fire scenarios, and the comparative experiments of different semantic models applied in forest fire scenarios are inadequate; hence, the comprehensive performances of these models applied to forest fire scenarios cannot be evaluated from multiple perspectives.

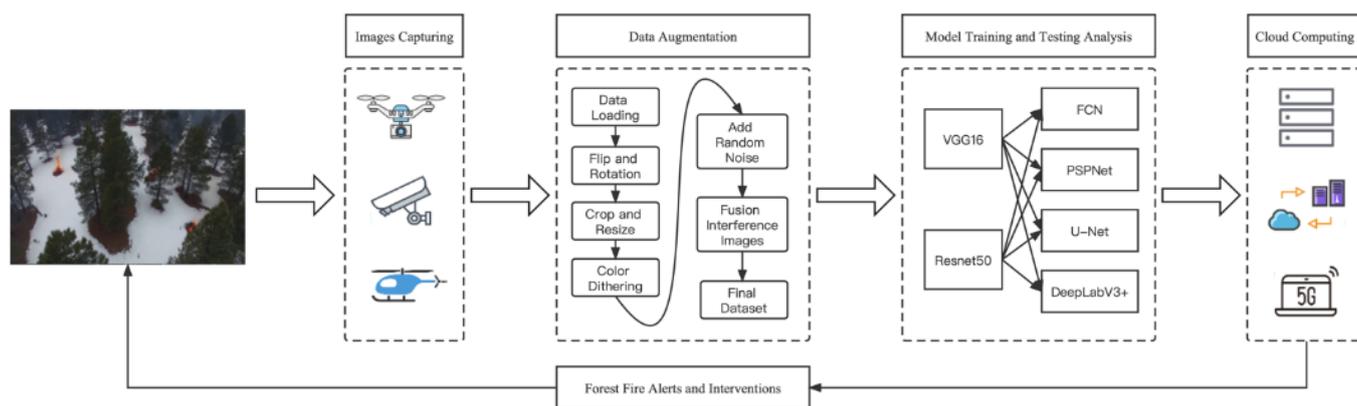
In view of the above problems, in this study, we collected a large number of remote-sensing images of forest fires taken by UAVs, manually annotated some of the data, constructed a flame segmentation dataset, and established a data enhancement process applicable to forest fire images to improve the data quality. Four classical semantic segmentation models and two backbone networks were selected for theoretical research, experimental design and comparative analysis. From different perspectives, we tested the performance of each model for forest fire segmentation tasks, explored the most suitable semantic segmentation model for forest fire scenarios, and provided forestry departments and authorities with more effective application means, data and decision support.

## 2. Materials and Methods

At present, the systems adopted by most countries for forest fire identification and segmentation include five stages, which are image capturing, data processing, model training and testing analysis, cloud computing, and alerts and intervention. This work focuses on the optimization and comparative research of data processing and model analysis stage. The overall process of this experiment is shown in Figure 1, and the specific contents of each stage are as follows:

- Stage 1: Image capturing. Usually, when a forest fire occurs, UAVs, forest surveillance and cruise helicopters capture and collect images of the fire and upload them after collation, thus obtaining a real-time forest fire scene.
- Stage 2: Data augmentation. In this paper, the traditional data-processing method is optimized, and a set of data enhancement processes for forest fire images is proposed to improve the data quality and the accuracy of the model. In this stage, the images collected by UAVs are firstly imported. During the augmentation process, the images are first flipped and rotated, and then cropped and resized; subsequently, the images are subjected to color dithering in batches. In order to improve the generalization capacity of the model, random noise is added to the image, and then the interference images are fused with the dataset and divided.
- Stage 3: Model training and testing analysis. In this part, four widely used fully convolutional network models are selected and trained based on two backbone networks, respectively. Meanwhile, we adopt four different indicators to evaluate the models' segmentation performances, record segmentation results of the models on forest fire and interference images, and compare and analyze from various perspectives.
- Stage 4: Cloud computing. In the actual forest-fire-monitoring application, the main computing tasks such as model training and real-time segmentation of forest fire images can be performed in the cloud computing center after 5 G high-speed data transmission, which greatly reduces the computing tasks of single-point servers to adapt to the problem of insufficient server resources in some regions.
- Stage 5: Alerts and intervention. Once the cloud server completes the segmentation task, it can detect whether a forest fire has occurred in the currently captured area. If a fire area is segmented, an alarm is immediately raised to notify the relevant forestry and fire departments to intervene in a timely manner and take measures on the fire site.

On the basis of improving the data augmentation module to obtain high-quality data, this work focuses on the comparative analysis of common semantic segmentation algorithms, and explores models that are more suitable for forest fire scenarios.



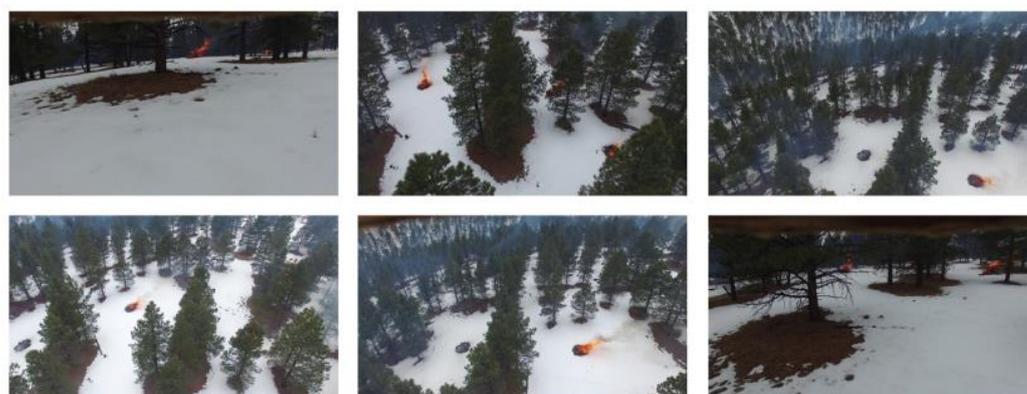
**Figure 1.** Overall process of the experiment.

### 2.1. Construction of Forest Fire Dataset

In this study, 4200 remote-sensing images of a forest fire were collected as the basis of the flame segmentation dataset. Among them, a large proportion of the images were from Northern Arizona University’s forest fire dataset FLAME [25], captured by UAVs. However, most of these forest fire images were taken in sunny or clear weather, and the environmental conditions were relatively simple. In order to improve the robustness of the system under various conditions, images taken under four common environmental conditions such as foggy, stormy, blizzard, and smoky were selected and added to the dataset. These data came from the Internet and the forest fire videos were intercepted frame-by-frame, which were captured by forest surveillance and firefighting helicopters. In addition, in order to improve the anti-interference ability of the models, 200 interference images similar to forest fires and background colors were also added to the dataset. As shown in Table 1, 95.23% of the 4200-image dataset contains forest fires, and 4.76% of the data contain similar interferences; the respective proportions of these images taken in various conditions are also shown. Figure 2 shows part of the forest fire images with a resolution of  $3840 \times 2160$ ; Figure 3 shows part of the interference images with a resolution of  $800 \times 450$ .

**Table 1.** Distribution and proportion of dataset.

Proportion	Clear	Foggy	Stormy	Blizzard	Smoky	Total
Forest fire images	63.49%	8.48%	4.22%	3.62%	15.42%	95.23%
Interferences	3.42%	0.56%	—	—	0.78%	4.76%



**Figure 2.** Part of the forest fire images.

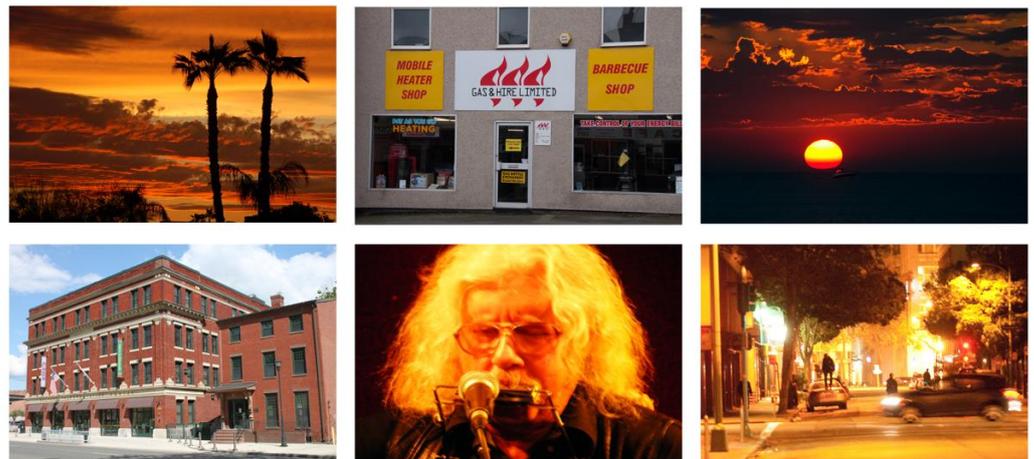


Figure 3. Part of the interference images.

## 2.2. Forest Fire Segmentation Models

In this study, four fully convolutional network models, FCN, U-Net, PSPNet and DeepLabV3+, which are widely used in image semantic segmentation tasks, were selected for comparative analysis. Among them, FCN (fully convolutional networks) [26] applied deep-learning technology to the field of image segmentation for the first time, achieved end-to-end pixel-level processing, and opened a new era of semantic segmentation. U-Net (unity-networking) [27] solved the problem of partial pixel spatial information loss through the encoder–decoder structure. PSPNet (pyramid scene parsing network) [28], a semantic segmentation model based on feature fusion, improved segmentation accuracy through feature fusion of global and local information. DeepLabV3+ [29] fused spatial pyramid pooling and a decoder to further refine the segmentation results. In summary, the four models adopted four different core architectures which are representative.

### 2.2.1. FCN

In 2014, Long proposed fully convolutional networks, which was the first time deep-learning technology was used to fundamentally solve the image semantic segmentation tasks. The FCN structure is shown in Figure 4.

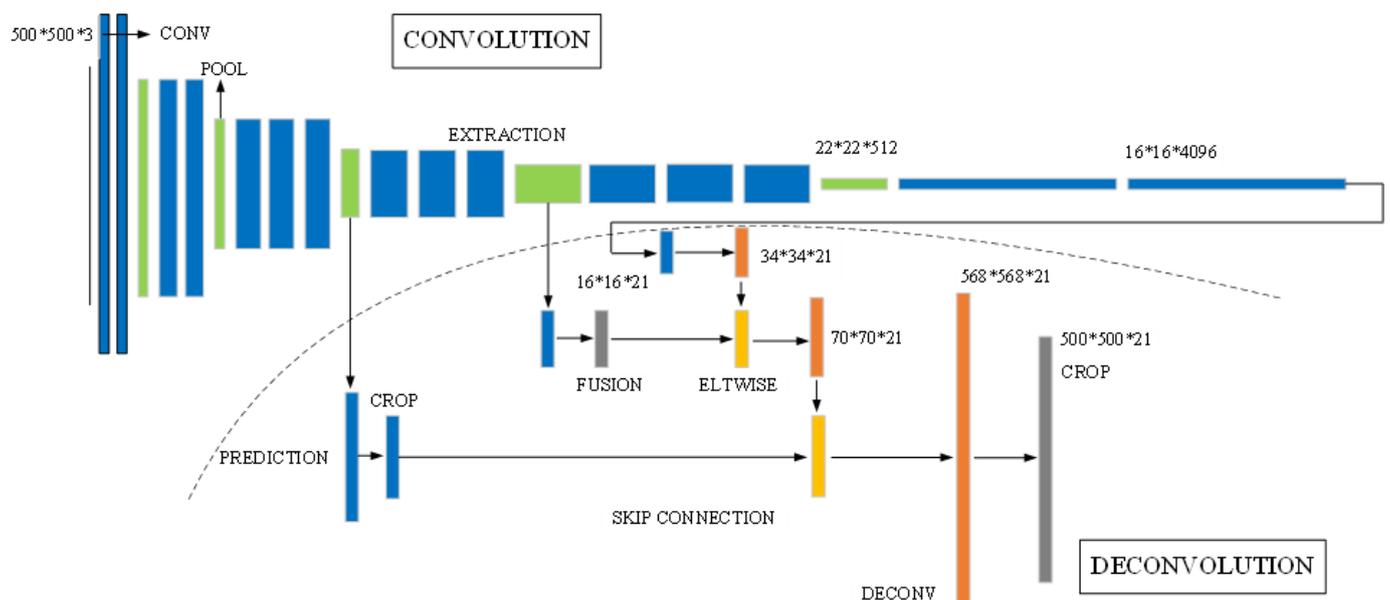


Figure 4. FCN structure.

The convolutional neural network (CNN) structure [30,31] is more used to solve target classification, which uses fully connected layers after convolution layers to obtain a fixed-length feature vector. However, the FCN structure uses convolutional layers to replace the fully connected layer for the first time, uses the skip-layer connection to fuse the features extracted by different layers, then uses deconvolution, bilinear interpolation and other means to up-sample and restore the image size, thereby achieving the end-to-end processing of images.

Compared with CNN, the FCN model can accept image inputs of any size without pre-sizing. Secondly, the FCN model is much more efficient, which can effectively solve the problems related to repeated storage and convolution calculations caused by the use of pixel blocks.

The FCN structure transforms the classification network into a structure for semantic segmentation for the first time. The FCN model's segmentation accuracy is greatly improved compared with previous methods, which drives the image semantic segmentation technology into a new era and provides inspirations for subsequent research on semantic segmentation algorithms.

### 2.2.2. U-Net

In the field of image segmentation, the encoder–decoder structure [27,32] is often used to solve the problem of spatial pixel loss. The structure mainly contains two parts: the encoder and decoder. The encoder mainly includes convolutional layers and down-sampling layers, which gradually reduce the size of feature maps and capture higher-level semantic information through convolution. The decoder includes an up-sampling layer, a convolution layer and a fusion layer, and gradually recovers the target detail information and spatial dimension through deconvolution. The whole structure exploits the multi-scale features of the autoencoder and recovers the spatial resolution from the decoder.

In 2015, the U-Net (unity-networking) structure proposed by Ronneberger O became one of the classic representatives of the encoder–decoder structure.

The U-Net structure is shown in Figure 5, which mainly includes a contraction path to capture contextual information and a symmetric expansion path for precise localization. The encoder part on the left side of the model consists of four sub-modules; each sub-module contains two convolutional layers, and is followed by a max-pooling layer to achieve down-sampling. The decoder part on the right side of the model also consists of four sub-modules, which recover the target details and spatial dimensions through up-sampling. The skip connection structure in U-Net connects the up-sampling results with the features of the same resolution in the encoder, and performs cross-layer feature fusion as the input to the next sub-module of the decoder. U-Net achieves multi-scale feature fusion and can complete end-to-end processing with a small amount of data.

### 2.2.3. PSPNet

In 2017, Zhao H proposed the PSPNet structure, which is shown in Figure 6. Based on FCN, it makes full use of global semantic connection to improve the reliability of prediction by aggregating contextual information in different regions. It is characterized by the use of a pyramid pooling module, which fuses four feature maps with different scales to extract local and global information of the image, cascading and integrating the features of the same level.

In the pyramid pooling module, the outputs of different scale levels contain feature maps of different sizes, and the contextual dimension is reduced to  $1/N$  by using a  $1 \times 1$  convolutional layer to ensure the global region weight. Then, the low-dimensional feature maps are up-sampled by bilinear interpolation to obtain features of equal size, these features of different levels are spliced into bit pyramid pooling global features, and, finally, pixel-by-pixel prediction results are outputted through a convolutional layer.

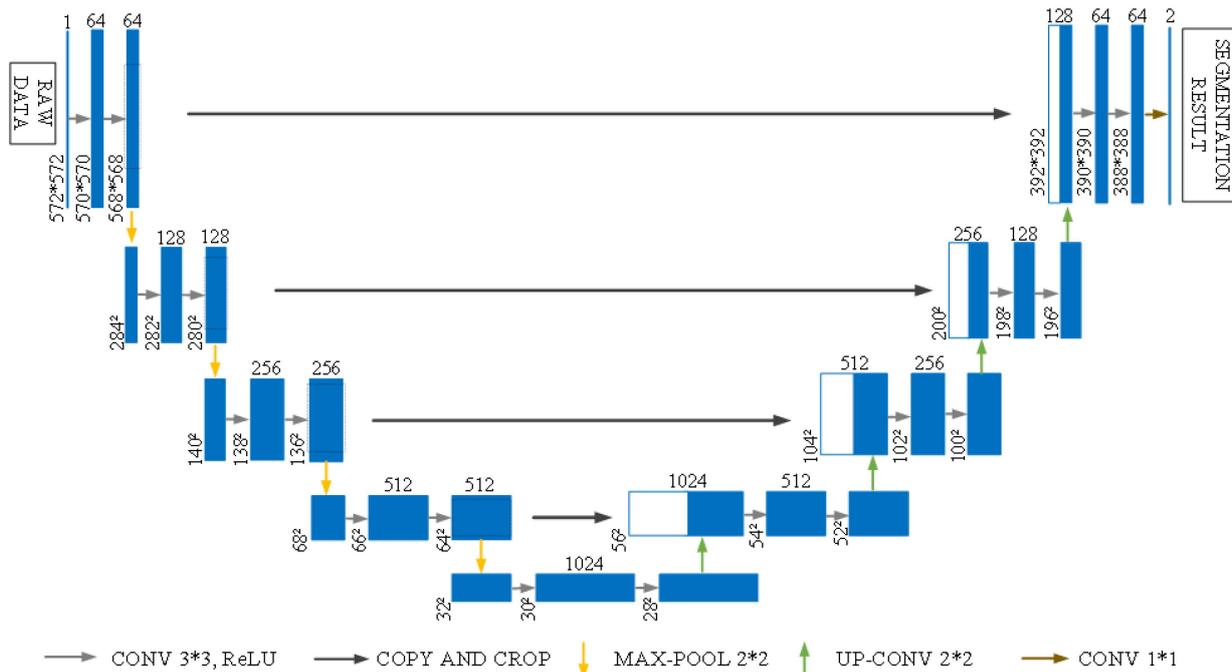


Figure 5. U-Net structure.

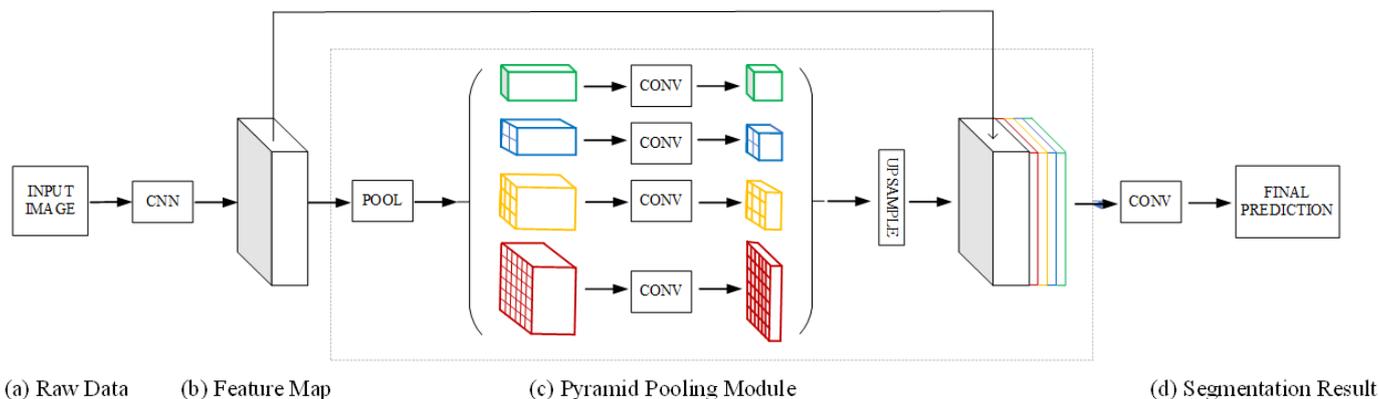


Figure 6. PSPNet structure.

To a certain extent, PSPNet overcomes the main disadvantage of the FCN structure for image semantic segmentation without considering contextual information, thereby extracting a complete feature representation and further improving the segmentation accuracy.

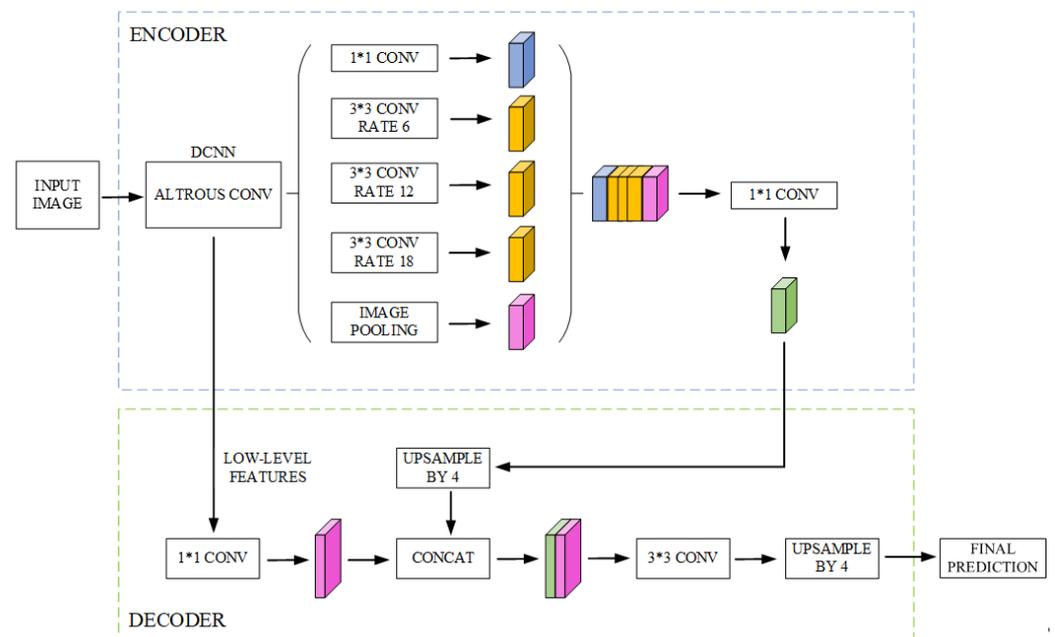
### 2.2.4. DeepLabV3+

In 2018, on the basis of deeplabV3, Chen proposed DeepLabV3+, which added a decoder structure. The structure is shown in Figure 7. DeepLabV3 is used as the encoder structure to extract rich contextual information, and the added new decoder optimizes the segmentation results and restores semantic information of the object.

The core aim of the encoder is to build an atrous spatial pyramid pooling module (ASPP) which can process the features of DCNN in 5 different ways, including 3 kinds of convolution with different hole rates, pooling and dimensionality reduction, and then fuse the features. The decoder connects the shallow-layer features of the DCNN through the features extracted by the ASPP module, and performs deconvolution to obtain the final result.

The DeeplabV3+ structure adds a separable depthwise separable convolution to the ASPP module and decoder, which improves the network running speed and robustness,

and greatly improves the segmentation accuracy, so as to achieve a relative balance between model accuracy and algorithm time complexity.



**Figure 7.** DeepLabV3+ structure.

### 2.3. Methods

This study used the forest fire dataset constructed above, and randomly selected 80% of the data for training, 10% for validation, and 10% for testing.

FCN, U-Net, PSPNet, and DeeplabV3+ were used to conduct comparative segmentation research on forest fires, and the backbones were VGG16 and Resnet50, respectively. The structure of VGG16 is relatively simple, and the stacked small convolution kernels were used to replace the larger ones in the classical network, with fewer parameters and a faster running speed; Resnet50 has a deeper structure, and adopts a residual module to ameliorate the degradation of the model, which can build more convolutional layers to extract deeper features.

In the experiments, the binary cross-entropy loss function was used as Loss. The experimental environment and configuration are shown in Table 2, and the network parameter settings are shown in Table 3.

**Table 2.** Experimental environment and configuration.

Content	Configuration
CPU	Intel(R) Core(TM) i7-7700
GPU	NVIDIA GeForce 2080Ti
Operating System	Win 10
Pytorch	1.11.0

**Table 3.** Network parameter settings.

Parameter	Value
Learning rate	0.001
Batchsize	8
Epoch	600

#### 2.4. Evaluation Indicators

The evaluation indicators recorded during model testing include PA, mPA, mIoU and FWIoU. In the following formulas, TP (true positive) refers to the number of flame pixels correctly predicted by the model; FP (false positive) refers to the number of pixels predicted to be flames, but the true value is the woodland background; FN (false negative) indicates that the woodland background is predicted, but the true value is the number of pixels of the flame; TN (true negative) represents the number of correctly predicted woodland background pixels.

##### 1. Pixel Accuracy (PA):

PA refers to the proportion of correctly classified flame and woodland background pixels in the total pixels in the image, which represents the accuracy of the model's classification of pixels. The greater the PA, the more accurate the model will be to segment the flame and woodland background.

$$PA = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{\text{Total}} \quad (1)$$

##### 2. Mean Pixel Accuracy (mPA)

mPA is to calculate the prediction accuracy of the flame and woodland background pixels separately, and take the average value of the two, which can better reflect the segmentation accuracy of the model for the overall semantics. The higher the mPA, the higher the segmentation accuracy of the flame and the woodland background, and the more precise the model.

$$mPA = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (2)$$

##### 3. Mean Intersection over Union (mIoU)

The IoU refers to the ratio of the intersection and union of the target ground-truth pixel set and the prediction set. The mIoU is to calculate the IoU of the flame and woodland background, respectively, and take the average value. The larger the mIoU, the higher the predicted coincidence between the flame and the woodland background and the real area.

$$mIoU = \frac{1}{2} \left( \frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right) \quad (3)$$

##### 4. Frequency-Weighted Intersection over Union (FWIoU):

The frequency-weighted IoU is a weighted summation of the various intersection ratios according to the frequency of the flame and forest background pixels. Compared with mIoU, FWIoU sets weights for each category according to the frequency of occurrence, which better improves the overall semantic evaluation. The larger the FWIoU, the stronger the overall segmentation performance.

$$FWIoU = \frac{TP + FN}{\text{Total}} \times \frac{TP}{TP + FP + FN} + \frac{TN + FP}{\text{Total}} \times \frac{TN}{TN + FN + FP} \quad (4)$$

### 3. Results

The test set contains a total of 420 pictures, including various environmental conditions such as clear, foggy, stormy, blizzard and smoky, and also includes interference images similar to the background color of flames and woodland. Based on the trained models above, we aimed to test and record the segmentation results and evaluation indicators from multiple perspectives.

### 3.1. Forest Fire Segmentation

Shown in Figure 8 are the segmentation results of each model with VGG16 as the backbone network; shown in Figure 9 are the segmentation results of each model with Resnet50 as the backbone network.

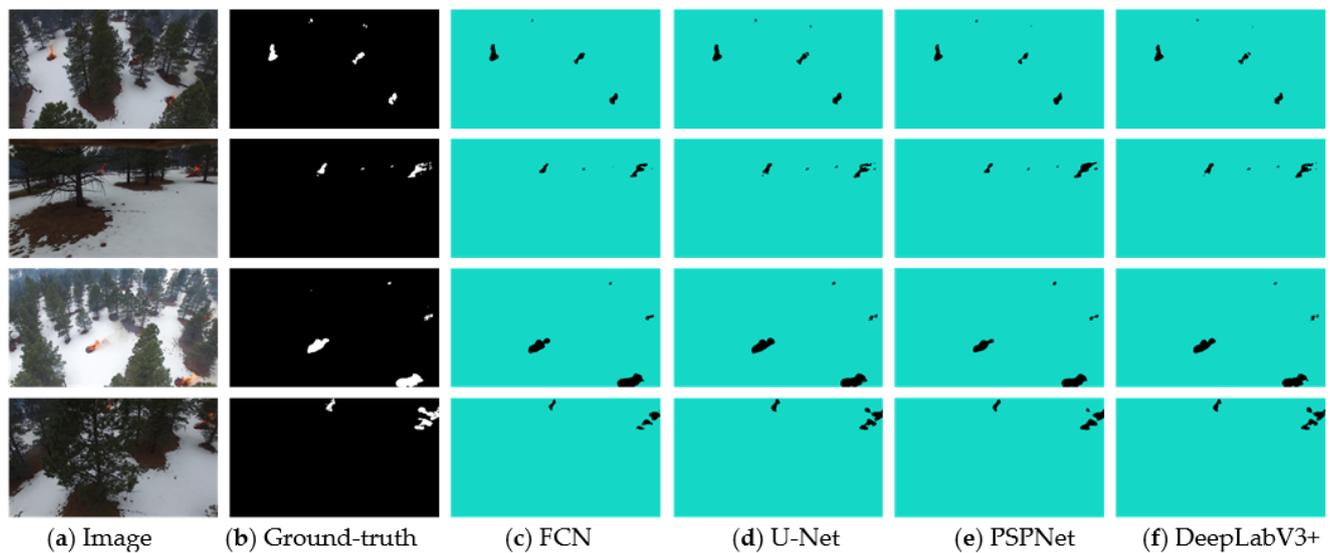


Figure 8. Results of forest fire image segmentation (VGG16).

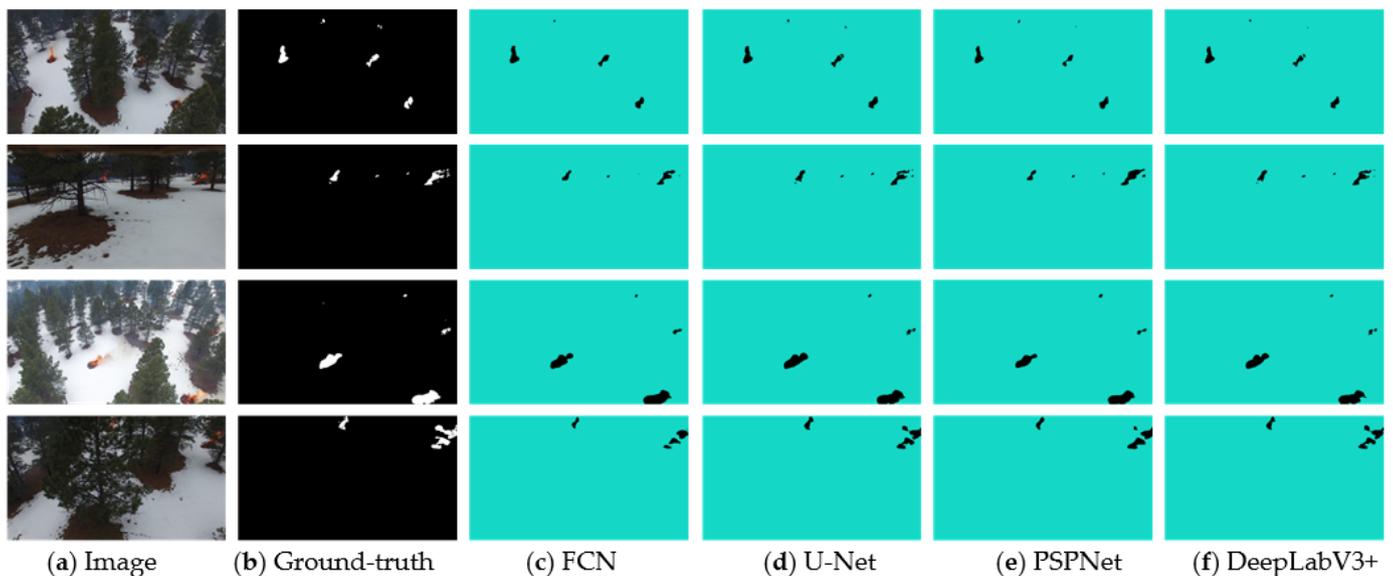


Figure 9. Results of forest fire image segmentation (Resnet50).

In the figure, column (a) is the original image of the forest fire; column (b) is the ground-truth label of the forest fire image; column (c) is the segmentation result of the FCN structure; column (d) is the segmentation result of the U-Net structure; column (e) is the segmentation result of the PSPNet structure; and column (f) is the segmentation result of the DeepLabV3+ structure.

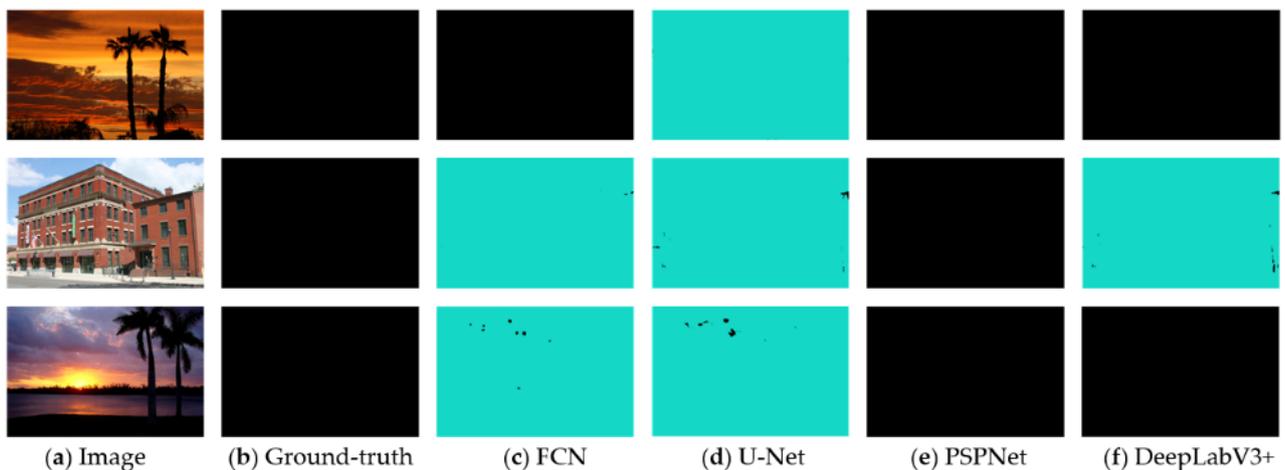
From the perspective of different fully convolutional network models, for the forest fire area, it can be seen from column (d) that the U-Net model has the most precise segmentation results of the flame edge, with the highest degree of overlap with the flame area of ground-truth. U-Net also has a satisfying segmentation result on the outer flame, and can well segment the flame area occluded by the trees. Column (c) shows that the FCN model has

the worst flame segmentation results among all models, whose segmentation results of the flame contour and outer flame are poor, and only the inner flame can be identified. However, the FCN model cannot segment the flame occluded by trees well. Columns (e) and (f) show that the PSPNet and DeepLabV3+ structure can better identify the main body and part of the outer flame. However, for the irregular outer flame shape, the forest background will be incorrectly divided into flames, and the segmentation of flame details is not as good as the U-Net structure, causing errors.

### 3.2. Anti-Interference Segmentation

In the test with interference images similar to the background colors of flames and forest background, the models with VGG16 as the backbone predict the interference images precisely, the background pixels are not incorrectly predicted as flames, and the segmentation results are completely coincident with the label map which is in black.

However, models with Resnet50 as the backbone had slightly different segmentation of interference images. Figure 10 shows that the PSPNet model with Resnet50 was not affected by the interference images, the segmentation result is no flame in the figure which is a full black image, the same as ground-truth. U-Net misidentified flames in all three interfering images, causing errors. The FCN and DeepLabV3+ models performed similarly, they also had the correct segmentation performance for different interference images, and their false-pixel-prediction frequency was lower than U-Net.



**Figure 10.** Results of Interference image segmentation.

### 3.3. Comparison of Models

From the data in Table 4, it can be seen that the evaluation indicators of the four models are generally high, and the above four models have good performance in the forest fire segmentation. U-Net with Resnet50 has higher indicators than the other models in different forest fire scenarios, and the segmentation result is the best. The FCN model with VGG16 is inferior to other models in all indicators, and the segmentation result is the poorest. The overall performance of PSPNet and DeepLabV3+ with Resnet50 is moderate and has a relatively good segmentation result.

Table 5 shows the size of the weight files of each model. Combined with the experimental results in Table 4, it can be seen that the weight files of FCN and PSPNet are small, but the segmentation effect is relatively poor. U-Net's weight file is moderate and the segmentation is the best. DeepLabV3+ has the largest weight file, but its segmentation results are slightly lower than U-Net. Table 6 shows the time consumption of respective trained models when inferring a single image when running on CPU.

**Table 4.** Evaluation indicators of models.

Model	Backbone	PA	mPA	mIoU	FWIoU
FCN	VGG16	0.9967	0.8456	0.8162	0.9952
	Resnet50	0.9972	0.8474	0.8279	0.9963
U-Net	VGG16	0.9986	0.9226	0.8875	0.9983
	Resnet50	0.9991	0.9281	0.8924	0.9989
PSPNet	VGG16	0.9978	0.8986	0.8438	0.9965
	Resnet50	0.9981	0.8998	0.8571	0.9969
DeepLabV3+	VGG16	0.9986	0.9112	0.8728	0.9973
	Resnet50	0.9989	0.9147	0.8752	0.9979

**Table 5.** Weight file of models.

Backbone	FCN	U-Net	PSPNet	DeepLabV3+
VGG16	168.2 MB	384.1 MB	231.3 MB	371.2 MB
Resnet50	276.5 MB	462.3 MB	356.6 MB	423.7 MB

**Table 6.** Time consumption of a single image inference.

Backbone	FCN	U-Net	PSPNet	DeepLabV3+
VGG16	0.41 s	0.73 s	0.46 s	0.56 s
Resnet50	0.43 s	0.82 s	0.49 s	0.68 s

#### 4. Discussion

In forest fire detection, more and more systems adopt flame segmentation algorithms to replace classification algorithms. Table 7 shows some representative algorithms for fire classification and segmentation and their results. In comparison, the fire segmentation algorithms have higher accuracy and stronger detection performance. Secondly, for the forest background, the segmentation algorithm can better locate the fire site, then analyze and predict the fire situation in combination with the surrounding conditions. In addition, classification algorithms mainly focus on detecting flames, while most segmentation algorithms can combine flames together with smoke to further improve the detection performance.

**Table 7.** Comparison between fire classification and segmentation method.

Mode	Ref	Methodology	Smoke/Flame	Dataset	Accuracy (%)
Classification	Treneska S and Stojkoska B.R. [33]	VGG16	Flame	FLAME: 8617 images	80.76
		VGG19			83.43
		Resnet50			88.01
	Chen Y. et al. [34]	CNN-17	Flame/Smoke	Private: 2100 images	86.00
	Shamsoshoara A. et al. [35]	Xception	Flame	FLAME: 48,010 images	76.23
Segmentation	Bochkov V.S. and Kataeva L.Y. [36]	wUUNet	Flame	Private: 6250 images	94.09
	Harkat H. et al. [37]	DeepLabV3+	Flame/Smoke	Corsian: 1775 images	97.53
	Frizzi S. et al. [38]	U-Net	Flame/Smoke	Private: 366 images	90.20
CNN based on VGG16		93.40			

Comparing the segmentation results of each model with different backbone networks in Figures 8 and 9, on the whole, all of the above models could well segment the flame from

the forest background, but there are still some differences in the accuracy of the outline. It can be seen that the U-Net with Resnet50 as the backbone network was the most accurate for forest fire segmentation and described the details the best; the FCN model with VGG16 as the backbone network was the least accurate for flame segmentation, but it could still identify the flame's main body.

From the perspective of different backbones, compared with VGG16, each model with Resnet50 as the backbone had more refined segmentation results, which could better segment the flame from the forest background, and the segmentation results were more complete and closer to ground-truth.

In terms of the segmentation of interference images, U-Net led to the misidentification of flames for interference images in the test due to its excessive sensitivity to fire features. In contrast, DeepLabV3+ outperformed the U-Net model in terms of segmentation performance against interference images, and its flame recognition misidentification rate was lower than that of the U-Net model. Although PSPNet had the lowest false-identification rate for interfering images, it is not suitable to be used because of its relatively poor segmentation accuracy.

From the analysis of model evaluation index level, PA and FWIoU can better reflect the model's overall segmentation performance, that is, the prediction of all pixels including the flames and forest background. However, because the forest fire pictures used in the experiment had high resolution and many pixels, most of the pixels were forest background, and just about 5%–10% of the pixels in the image were forest fire. Therefore, in the test, these two indicators tended to account for the larger portion of the forest background pixels, resulting in higher values.

In contrast, mPA and mIoU can better reflect the model's segmentation performance of the flame area. As can be seen from the above, each model had a good segmentation of the forest background whose pixels accounted for a large proportion. The above two indicators were calculated after averaging the forest fire area segmentation results. Therefore, mPA and mIoU had a strong positive correlation with the flame area segmentation accuracy.

The main reason for the differences in the segmentation results of each model is that different semantic segmentation network models have different fusion methods and utilization levels of deep and shallow features.

The FCN structure only performs two cross-layer feature fusions for the extracted features, and lacks multi-scale operations on features, so the segmentation results were poor. The core of PSPNet is to pool the extracted features at four different scales, and then merge them with the previous features. Compared with other models, PSPNet lacks the fusion of shallow features and has a low feature utilization rate. As an encoder–decoder structure, U-Net performs multiple down-sampling in the encoder, and the collected deep features can obtain the overall contextual semantic information. It performs up-sampling operations multiple times in the decoder, and connects the extracted features with the down-sampling results of the same stage, instead of directly back-propagating on the high-level semantic features, which ensures that more shallow features can be fused into the feature map. Therefore, U-Net retains and combines both deep and shallow features, so it achieves better results in forest fire segmentation tasks. DeepLabV3+ performs multi-dimensional operations on deep features in the encoder, and combines them with shallow features. It also takes into account both deep and shallow features, so its indicators are better than FCN and PSPNet, second only to U-Net.

Based on the above data, U-Net with Resnet50 performs the best with the highest indicators, but its weight file is the largest with a relatively slow running speed. It can be used in forest fire segmentation scenarios that do require high-precision flame details, such as fire attribute analysis, trend prediction for forestry departments, and related scientific research. The FCN and PSPNet models have performed relatively poorer with lower indicators, so they are not suitable for forest fire segmentation scenarios. The indicators of DeepLabV3+ with Resnet50 are relatively satisfying, only slightly lower than the U-Net, but its weight file is smaller with a faster running speed. It is suitable for scenarios with

high real-time requirements, such as the real-time detection of forest fires and monitoring, fire real-time warning systems, etc.

## 5. Conclusions

In this paper, we compile remote-sensing forest fire images captured by UAVs in multiple conditions to construct a dataset, and optimize the data enhancement process for forest fire image features. Based on the enhancement of data quality, four full convolutional networks and two backbone network models are adopted for combined training, analysis and testing, respectively. The focus is on the comparative research of the semantic segmentation methods applied to the forest fire segmentation scenarios, analyzing and discussing their application value and functional focus.

The results show that, among the four widely used deep semantic segmentation models, the U-Net model has the best performance in forest fire segmentation, with a better description of semantic details such as outflames and contours, but the inference speed for a single instance is relatively slow, which is more suitable for application scenarios requiring high accuracy, such as forest fire attribute analysis and trend prediction; the DeepLabV3+ model has relatively faster inference speed while ensuring high segmentation accuracy. Despite the description of flame details being slightly lower than that of U-Net model, it is still satisfactory, which is suitable for scenarios requiring high real-time performance, such as forest fire real-time detection and alerts. The FCN and PSPNet models are relatively unsuitable for forest fire segmentation scenarios because of their low segmentation test and evaluation index. In addition, the comparative research concluded that the detection accuracy of forest fire segmentation methods is generally higher than that of the classification methods; therefore, the U-Net and DeepLabV3+ are more applicable in forest fire detection systems.

Image segmentation is the key area for forest fire detection. Forest fire segmentation and recognition in remote sensing woodland images with faster, more accurate and lower resource cost is the focus of current research in this field. Although the existing techniques have achieved certain results, improvements are still needed in the following aspects in future research: combining smoke and forest fires for identification and analysis, the optimization of identification for forest fires under complex backgrounds and occlusion, and the adoption of edge computing technology to improve inference speed.

**Author Contributions:** Conceptualization, Z.W.; Methodology, Z.W.; Software, Z.W.; Validation, Z.W.; Formal Analysis, Z.W.; Investigation, Z.W.; Resources, T.P.; Data Curation, Z.W. and Z.L.; Writing—Original Draft Preparation, Z.W.; Writing—Review and Editing, T.P. and Z.L.; Visualization, Z.W.; Supervision, T.P. and Z.L.; Project Administration, T.P. and Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the China National College Student Innovation Training Program under Grant 202110225117.

**Data Availability Statement:** Data are available from the author upon reasonable request.

**Acknowledgments:** Dedicated to the 70th anniversary of Northeast Forestry University.

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

## References

1. Wang, L.; Li, H.; Li, S.; Bie, Y. Gradient illumination scheme design at the highway intersection entrance considering driver's light adaption. *Traffic Inj. Prev.* **2022**, *23*, 266–270. [[CrossRef](#)] [[PubMed](#)]
2. Lu, Z.; Ding, N.; Lu, L.; Tian, Z. Optimizing signal timing of the arterial-branch intersection: A fuzzy control and nonlinear programming approach. *Asian J. Control* **2022**, *Early View*. [[CrossRef](#)]
3. Cisneros, R.; Schweizer, D.; Navarro, K.; Veloz, D.; Procter, C.T. Climate change, forest fires, and health in California. In *Climate Change and Air Pollution*; Springer: Cham, Switzerland, 2018; pp. 99–130.
4. Boer, M.M.; Resco de Dios, V.; Bradstock, R.A. Unprecedented burn area of Australian mega forest fires. *Nat. Clim. Chang.* **2020**, *10*, 171–172. [[CrossRef](#)]

5. Cavallo, E.A.; Noy, I. The Economics of Natural Disasters: A Survey. 2009. Available online: <https://ssrn.com/abstract=1817217> (accessed on 21 April 2011).
6. Toya, H.; Skidmore, M. Economic development and the impacts of natural disasters. *Econ. Lett.* **2007**, *94*, 20–25. [[CrossRef](#)]
7. Alkhatib, A.A.A. A review on forest fire detection techniques. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 597368. [[CrossRef](#)]
8. Szpakowski, D.M.; Jensen, J.L.R. A review of the applications of remote sensing in fire ecology. *Remote Sens.* **2019**, *11*, 2638. [[CrossRef](#)]
9. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors* **2022**, *22*, 1977. [[CrossRef](#)]
10. Shahid, M.; Virtusio, J.J.; Wu, Y.-H.; Chen, Y.-Y.; Tanveer, M.; Muhammad, K.; Hua, K.-L. Spatio-temporal self-attention network for fire detection and segmentation in video surveillance. *IEEE Access* **2021**, *10*, 1259–1275. [[CrossRef](#)]
11. Habiboğlu, Y.H.; Günay, O.; Çetin, A.E. Covariance matrix-based fire and flame detection method in video. *Mach. Vis. Appl.* **2012**, *23*, 1103–1113. [[CrossRef](#)]
12. Wang, D.-C.; Cui, X.; Park, E.; Jin, C.; Kim, H. Adaptive flame detection using randomness testing and robust features. *Fire Saf. J.* **2013**, *55*, 116–125. [[CrossRef](#)]
13. Tlig, L.; Bouchouicha, M.; Tlig, M.; Sayadi, M.; Moreau, E. A fast segmentation method for fire forest images based on multiscale transform and PCA. *Sensors* **2020**, *20*, 6429. [[CrossRef](#)]
14. Toulouse, T.; Rossi, L.; Celik, T.; Akhloufi, M. Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods. *Signal Image Video Process.* **2016**, *10*, 647–654. [[CrossRef](#)]
15. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
16. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. [[CrossRef](#)]
17. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens.* **2021**, *13*, 4712. [[CrossRef](#)]
18. Choi, H.-S.; Jeon, M.; Song, K.; Kang, M. Semantic fire segmentation model based on convolutional neural network for outdoor image. *Fire Technol.* **2021**, *57*, 3005–3019. [[CrossRef](#)]
19. Cai, C.; Zhang, P.; Tan, J.; Liu, H. An automatic information extraction method for the combustion flame of chain furnace based on U-net. In Proceedings of the 2021 IEEE Conference on Telecommunications, Optics and Computer Science, Shenyang, China, 10–11 December 2021.
20. Hossain, F.M.A.; Zhang, Y. Development of new efficient transposed convolution techniques for flame segmentation from UAV-captured images. In Proceedings of the 2021 International Conference on Industrial Artificial Intelligence, Shenyang, China, 18–21 August 2021.
21. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Mseddi, W.S.; Attia, R. Forest fires segmentation using deep convolutional neural networks. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics, Melbourne, Australia, 17–20 October 2021.
22. Harkat, H.; Nascimento, J.M.P.; Bernardino, A. Fire detection using Deeplabv3+ with mobilenetv2. Proceedings of 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021.
23. Zhang, J.; Zhu, H.; Wang, P.; Ling, X. ATT squeeze U-Net: A lightweight network for forest fire detection and recognition. *IEEE Access* **2021**, *9*, 10858–10870. [[CrossRef](#)]
24. Großkopf, J.; Matthes, J.; Vogelbacher, M.; Waibel, P. Evaluation of deep learning-based segmentation methods for industrial burner flames. *Energies* **2021**, *14*, 1716. [[CrossRef](#)]
25. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.; Blasch, E. The FLAME dataset: Aerial Imagery Pile burn detection using drones (UAVs). *IEEE Dataport* **2020**. [[CrossRef](#)]
26. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7 June 2015.
27. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
28. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017.
29. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
30. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [[CrossRef](#)]
32. 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](#)] [[PubMed](#)]

33. Treneska, S.; Stojkoska, B.R. Wildfire detection from UAV collected images using transfer learning. In Proceedings of the 18th International Conference on Informatics and Information Technologies, Skopje, North Macedonia, 6–7 May 2021.
34. Chen, Y.; Zhang, Y.; Xin, J.; Wang, G.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. UAV image-based forest fire detection approach using convolutional neural network. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications, Xi'an, China, 19–21 June 2019.
35. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Comput. Netw.* **2021**, *193*, 108001. [[CrossRef](#)]
36. Bochkov, V.S.; Kataeva, L.Y. WUUNet: Advanced fully convolutional neural network for multiclass fire segmentation. *Symmetry* **2021**, *13*, 98. [[CrossRef](#)]
37. Harkat, H.; Nascimento, J.; Bernardino, A. Fire segmentation using a DeepLabv3+ architecture. *Image Signal Process. Remote Sens. XXVI SPIE* **2020**, *11533*, 134–145.
38. Frizzi, S.; Bouchouicha, M.; Ginoux, J.; Moreau, E.; Sayadi, M. Convolutional neural network for smoke and fire semantic segmentation. *IET Image Process.* **2021**, *15*, 634–647. [[CrossRef](#)]