forests

Article

FlameTransNet: Advancing Forest Flame Segmentation with Fusion and Augmentation Techniques

Beiqi Chen [1], Di Bai [2,*], Haifeng Lin [1,*] and Wanguo Jiao [1]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; momura@njfu.edu.cn (B.C.); wgjiao@njfu.edu.cn (W.J.)

[2] College of Information Management, Nanjing Agricultural University, Nanjing 210095, China

* Correspondence: baidi000@njau.edu.cn (D.B.); haifeng.lin@njfu.edu.cn (H.L.); Tel.: +86-25-8542-7827 (H.L)

Abstract: Forest fires pose severe risks, including habitat loss and air pollution. Accurate forest flame segmentation is vital for effective fire management and protection of ecosystems. It improves detection, response, and understanding of fire behavior. Due to the easy accessibility and rich information content of forest remote sensing images, remote sensing techniques are frequently applied in forest flame segmentation. With the advancement of deep learning, convolutional neural network (CNN) techniques have been widely adopted for forest flame segmentation and have achieved remarkable results. However, forest remote sensing images often have high resolutions, and relative to the entire image, forest flame regions are relatively small, resulting in class imbalance issues. Additionally, mainstream semantic segmentation methods are limited by the receptive field of CNNs, making it challenging to effectively extract global features from the images and leading to poor segmentation performance when relying solely on labeled datasets. To address these issues, we propose a method based on the deeplabV3+ model, incorporating the following design strategies: (1) an adaptive Copy-Paste data augmentation method is introduced to learn from challenging samples (Images that cannot be adequately learned due to class imbalance and other factors) effectively, (2) transformer modules are concatenated and parallelly integrated into the encoder, while a CBAM attention mechanism is added to the decoder to fully extract image features, and (3) a dice loss is introduced to mitigate the class imbalance problem. By conducting validation on our self-constructed dataset, our approach has demonstrated superior performance across multiple metrics compared to current state-of-the-art semantic segmentation methods. Specifically, in terms of IoU (Intersection over Union), Precision, and Recall metrics for the flame category, our method has exhibited notable enhancements of 4.09%, 3.48%, and 1.49%, respectively, when compared to the best-performing UNet model. Moreover, our approach has achieved advancements of 11.03%, 9.10%, and 4.77% in the same aforementioned metrics as compared to the baseline model.

Keywords: forest flame; CBAM; semantic segmentation; transformer; adaptive Copy-Paste

check for updates

Citation: Chen, B.; Bai, D.; Lin, H.; Jiao, W. FlameTransNet: Advancing Forest Flame Segmentation with Fusion and Augmentation Techniques. *Forests* **2023**, *14*, 1887. https://doi.org/10.3390/f14091887

Academic Editors: Luis A. Ruiz, Nopparat Kaakkurivaara and Stelian Alexandru Borz

Received: 4 August 2023
Revised: 4 September 2023
Accepted: 15 September 2023
Published: 17 September 2023

1. Introduction

The forest ecosystem plays a significant role in the global ecosystem and human society. Forests provide habitat for numerous species and serve as the foundation of food chains [1,2]. They also regulate climate, maintain water sources, and prevent soil erosion. However, forest fires, as a severe ecological disturbance, have profound impacts on forest ecosystems and human society [3]. Forest fires give rise to a range of issues. Firstly, they disrupt the structure and functioning of forest ecosystems, leading to loss of biodiversity, habitat degradation, and disruption of ecological processes [4]. Secondly, forest fires release a substantial amount of carbon into the atmosphere, exacerbating global climate change [5]. Thirdly, fires trigger soil erosion and water source contamination, negatively affecting the sustainable utilization of water resources and ecosystem health. Moreover, forest fires lead to significant economic losses, safety risks, and health issues in human society. For instance, the recent Lahina Fire in the United States resulted in the highest death toll since 1900, causing extensive casualties and property damage [6]. Therefore, accurate identification and effective monitoring of forest fires are crucial. In this regard, the importance of forest flame segmentation and recognition becomes evident. Accurate segmentation and recognition of forest flames contribute to real-time monitoring of forest fires, providing crucial information for emergency response and forest management decisions [7]. By applying advanced computer vision and deep learning techniques such as convolutional neural networks (CNNs) [8], the accuracy of forest flame recognition and segmentation can be enhanced, offering robust support for fire management. This helps in early detection and control of forest fires, minimizing damage to ecosystems and human society, and ensuring the sustainability of forest resources [7].

With the rapid advancement of remote sensing technology [9], it plays a crucial role in the segmentation of forest fires, offering significant advantages and significance [10,11]. Remote sensing technology provides high-resolution remote sensing images that capture detailed information about the shape and boundaries of flames, enabling accurate segmentation of fire regions [12,13]. This is of paramount importance for assessing the scale, intensity, and impact of wildfires on forest ecosystems [13]. Furthermore, remote sensing technology allows for the acquisition of multi-temporal image data, facilitating the observation and monitoring of fire dynamics. By analyzing the temporal changes in flames, researchers can investigate fire propagation patterns, predict potential fire spread paths, and provide more accurate guidance for wildfire suppression operations [14]. In addition, remote sensing technology possesses extensive spatial coverage capabilities, enabling the coverage of large forested areas [15]. It acquires images from different angles and heights, providing comprehensive information for flame segmentation. Moreover, remote sensing technology offers real-time capabilities, enabling the timely acquisition and rapid analysis of fire images, thereby facilitating prompt response to fire incidents and the implementation of effective firefighting and rescue measures. Given the various advantages offered by remote sensing technology, researchers have endeavored to utilize forest remote sensing images for forest flame segmentation [16,17].

Building upon the advantages of remote sensing technology in forest fire segmentation, the application of Convolutional Neural Networks (CNNs) has exhibited remarkable potential in enhancing the accuracy and efficiency of fire detection and segmentation [18]. CNNs, as a class of deep learning algorithms, have revolutionized computer vision tasks, including object recognition, image classification, and semantic segmentation. Leveraging the power of CNNs, researchers have made significant strides in effectively analyzing and extracting features from remote sensing images, enabling more precise and automated fire segmentation [17]. In their work, Eleni Tsalera et al. [19] propose a method that utilizes lightweight CNNs, such as SqueezeNet, ShuffleNet, MobileNetv2, and ResNet50, for wildfire identification. Performance evaluation is conducted on multiple datasets with cross-dataset analysis, comparing computational resources and costs to ResNet-50. For contextualization purposes, ResNet-18 is employed for image semantic segmentation. The experimental results demonstrate a high accuracy of 96% and satisfactory performance across datasets. Furthermore, five classes from the CamVid dataset are identified for contextualizing wildfires. Zhihao Guan et al. [20] propose a novel approach for forest fire detection and segmentation. They introduce a channel domain attention mechanism for image classification, achieving an impressive classification accuracy of 93.65%. Additionally, they develop MaskSU R-CNN, a novel instance segmentation method, which exhibits a precision of 91.85%, recall of 88.81%, F1-score of 90.30%, and mean intersection over union (mIoU) of 82.31%.

However, despite the good performance of convolutional neural network (CNN)-based semantic segmentation techniques on forest remote sensing datasets for forest flame applications, they have not considered some challenges inherent in remote sensing datasets and limitations of CNNs themselves.

Challenge 1: The limited receptive field of CNNs prevents the comprehensive extraction and utilization of information from the entire image, further exacerbating the neglect of flame features [21].

Challenge 2: As shown in Figure 1, due to the high resolution of remote sensing datasets, the flame region usually occupies a small proportion, resulting in insufficient attention from the model towards the flame region and incomplete learning of flame features.

Challenge 3: The scarcity of flame instances and the extremely imbalanced class distribution lead to long training time and elevated dataset requisites (encompassing a larger number of training images or images with more pronounced flame characteristics).

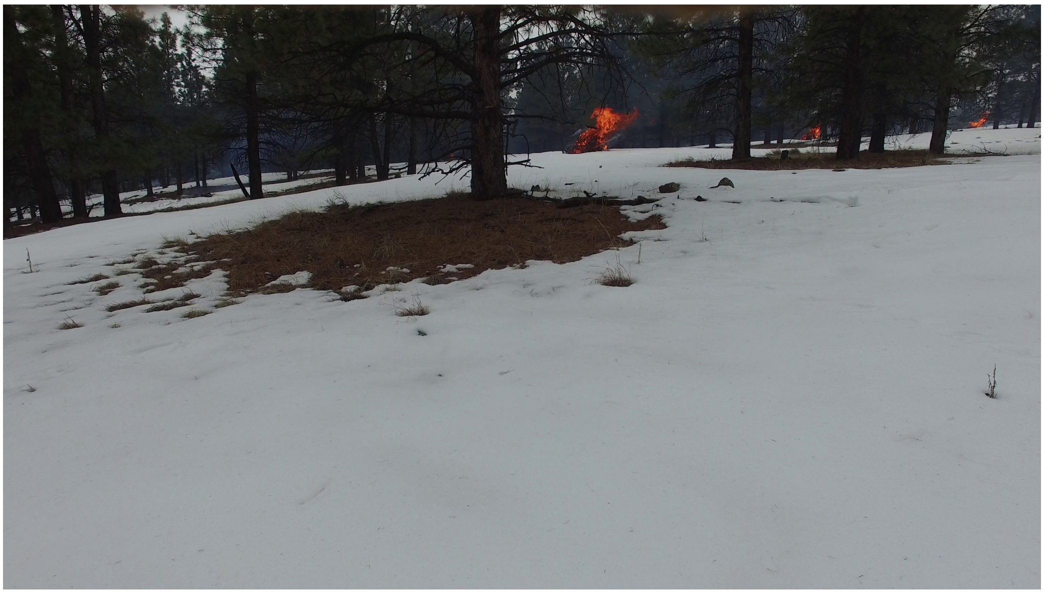

Figure 1. Visualization of Forest Remote Sensing Dataset Images.

In response to these challenges, we propose corresponding designs to enhance the performance of the model and fully leverage the training data. To address Challenge 1, we incorporate a simple transformer architecture in the encoder part of the network to capture global features in a parallel and serial manner, and introduce the CBAM(Convolutional Block Attention Module) attention mechanism in the decoder part to enable comprehensive learning of the image and improve segmentation accuracy and detail preservation. For Challenge 2, we introduce an adaptive Copy-Paste data augmentation method to increase the presence of poorly learned classes, allowing for sufficient learning of these classes. For Challenge 3, we introduce the dice loss, which emphasizes the flame region rather than the non-flame region, thereby improving model training speed.

As shown in Figure 2, our model is based on an encoder-decoder architecture, where the encoder part considers both speed and performance, and we choose MobileNetV2, while the decoder part adopts DeepLabV3+. Specifically, the approach involves initially selecting the image with the minimum confidence score from the current batch. Subsequently, based on the confidence scores transformed into probabilities for the images within the current batch, another image is randomly chosen. All pixels belonging to the flame category in the second image are then copied and pasted onto the first image. Subsequently, the batch images are passed through the transformer and further feature extraction by the encoder, where in the transformed features are concatenated with the features post transformer. This step serves to enhance the feature richness and accuracy. The final stage encompasses decoder operations for label prediction and incorporates the dice loss for facilitating the backpropagation process.

In this research paper, we have developed a novel network model based on unmanned aerial vehicle (UAV) remote sensing imagery, aimed at enhancing forest fire management and assessment. This can be categorized into two specific aspects:

- Accurate Flame Detection and Localization: Our approach enables the direct segmentation of UAV-acquired remote sensing images, accurately identifying the presence of flames within the images. Simultaneously, it provides information regarding the shape and size of the flames. Even relatively small flames can be accurately recognized using our method, facilitating early flame detection and timely firefighting measures.
- Fire Monitoring and Management: Managers can assess the fire situation and make informed decisions by analyzing the images segmented by our model. This facilitates the timely and accurate development of firefighting plans.

FlameTransNet, by integrating our approach with UAV technology, it provides managers with a convenient and efficient means of obtaining insights into forest conditions, reducing the labor costs associated with manual on-site inspections. Taking into account the forest environment and the dataset we have utilized, we believe that our technology holds significant potential for effective application in extensive forest regions such as Northern Arizona.

In the following section, we will introduce the work of previous researchers in Section 2 and compare our work with theirs. Then we provide a detailed description of our model design and the methods employed in Section 3. We will specifically describe our self-built dataset, the selected evaluation metrics, performance comparison with mainstream semantic segmentation methods, and ablation experiments of each module in Section 4. Finally, we will summarize our work and provide future research directions in Section 5.

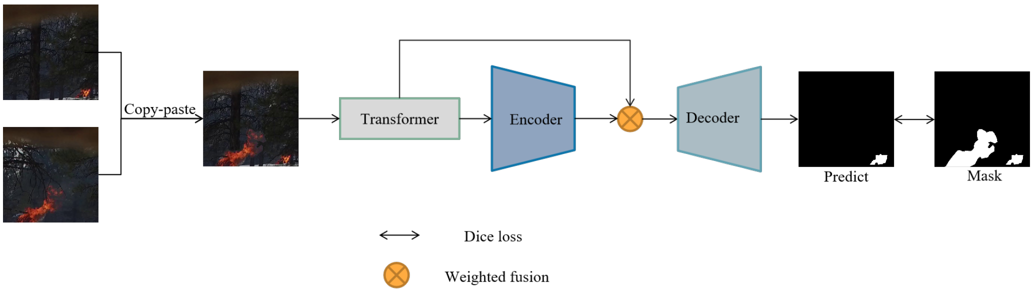

Figure 2. Model Pipeline for Semantic Segmentation: MobileNetV2 Encoder and DeepLabV3+ Decoder.

2. Related Work

2.1. Forest Fire Segmentation

Currently, numerous researchers have conducted explorations in the field of forest fire segmentation. For example, Rafik Ghali et al. [22] present a novel approach using deep convolutional networks and customized loss functions to accurately segment forest fire pixels and detect fire areas. This method holds promise for enhancing forest fire monitoring and response efforts. Lin Zhang et al. [23] proposed the FBC-ANet network, combining boundary enhancement and context-aware modules in a lightweight structure. This innovative approach achieved impressive results, with a segmentation accuracy of 92.19%, F1 score of 90.76%, and IoU of 83.08% on UAV images from the FLAME dataset. The FBC-ANet effectively extracts fire-related features, enhancing forest fire area segmentation accuracy. Rafik Ghali et al. [24] introduces a novel forest fire monitoring framework based on convolutional neural networks (CNNs). Their innovative approach effectively detects early forest fires, as demonstrated through experiments on a self-generated dataset and real monitoring videos. Although some studies have explored forest fire segmentation, most researchers have focused on evaluating different semantic segmentation methods or introducing new modules to enhance specific datasets or scenarios. However, these efforts often fail to address the limitations of CNNs comprehensively. Moreover, their studies often require a substantial amount of data. In contrast, our research delves into the limitations of CNNs and better utilization of datasets, seeking to overcome these challenges.

2.2. Fire Detection Systems

Fire Detection Systems utilize advanced technologies, including AI algorithms, to identify and promptly alert about fires, enhancing early detection and response capabilities.

Hamood Alqourabah et al. [25] proposed a smart fire detection system that integrates heat, smoke, and flame sensors to detect fires and alert property owners, emergency services, and police stations. The system minimizes false alarms, enhancing reliability, with positive affordability and effectiveness results demonstrated in experiments using the Ubidots platform. Giacomo Peruzzi et al. [26] developed a low-power Video Surveillance Unit (VSU) with embedded Machine Learning (ML) algorithms to detect forest fires using audio and image inputs. The combined ML approach achieved higher accuracy, precision, recall, and F1 score (96.15%, 92.30%, 100.00%, and 96.00%). Remote signaling through LoRaWAN protocol enables swift response to detected events. Diyana Kinaneva et al. [27] introduced a platform using UAVs equipped with AI and onboard image processing for forest fire detection. The approach involves continuous monitoring of fire-prone areas using drones and computer vision techniques to detect smoke or fire from images or video captured by the drones' cameras. Zhentian Jiao et al. [7] proposed a forest fire detection algorithm using UAV-based aerial images by utilizing YOLOv3. Their method achieved a recognition rate of about 83% and a frame rate of over 3.2 fps, showcasing its effectiveness for real-time forest fire detection using UAVs. In our research, we comprehensively considered both performance and model size, and selected MobileNetV2 as the feature extraction network. To enhance dataset utilization, we devised an adaptive Copy-Paste method, resulting in significant performance improvements for our model trained on relatively limited datasets. Our model demonstrates accurate segmentation even for smaller fire instances, providing robust support for early fire detection. In the future, we aim to explore its deployment on edge devices to realize the design of a fire monitoring system.

3. Method

3.1. Proposed Framework

As shown in Figure 3, our proposed network is built upon an encoder-decoder architecture. In the encoder part, we employ the MobileNetV2 network for feature extraction. Prior to the MobileNetV2 network, we integrate a Transformer module to capture deep image information, while simultaneously incorporating a parallel Transformer module to preserve the spatial context of the image, thereby alleviating the limited receptive field issue of CNNs. This approach maximizes the extraction of image features. Moreover, during the fusion stage of low-level features in both the encoder and decoder parts, we utilize the CBAM attention mechanism to further extract informative details from the lower-level features, enabling the model to pay more attention to the flame region and further enhance the model's performance.

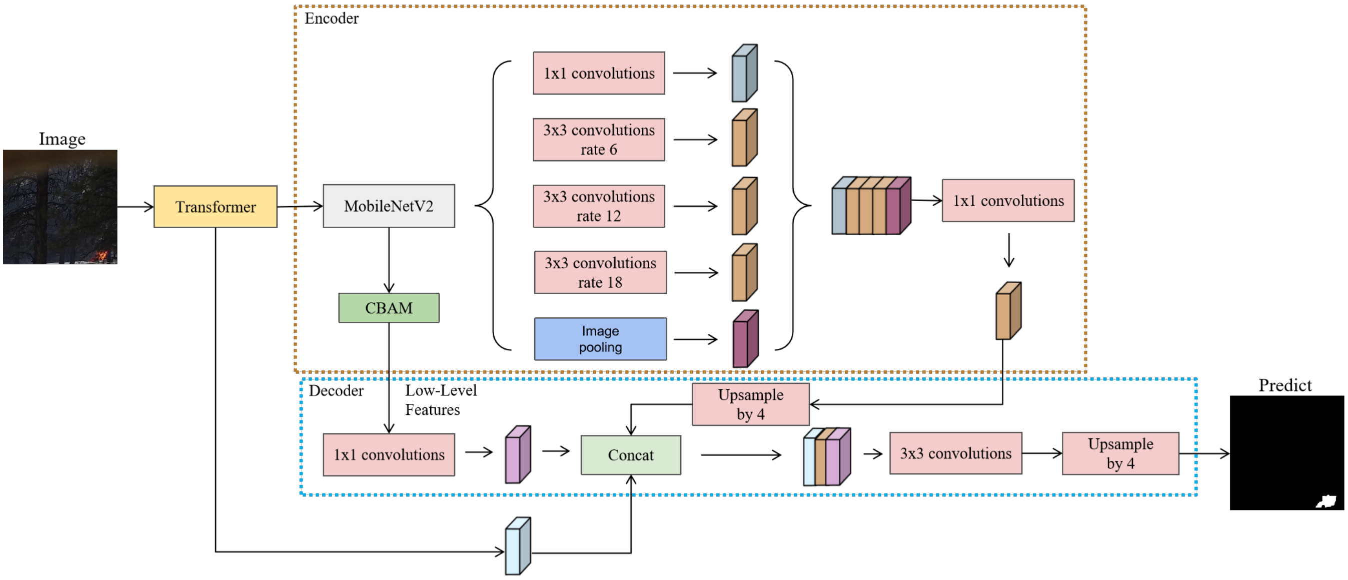

Figure 3. Architecture Design Diagram of the Proposed Model.

3.1.1. Encoder (MobileNetV2 Based)

MobileNet, a pioneering lightweight deep neural network devised by Google, was crafted to meet the demands of mobile and embedded devices. As illustrated in Figure 4, MobileNetV2 [28] represents a refined iteration of MobileNet, introducing a pivotal enhancement known as the Inverted Residual Block. This distinctive feature anchors the entirety of the MobileNetV2 architecture, facilitating its efficiency and effectiveness.

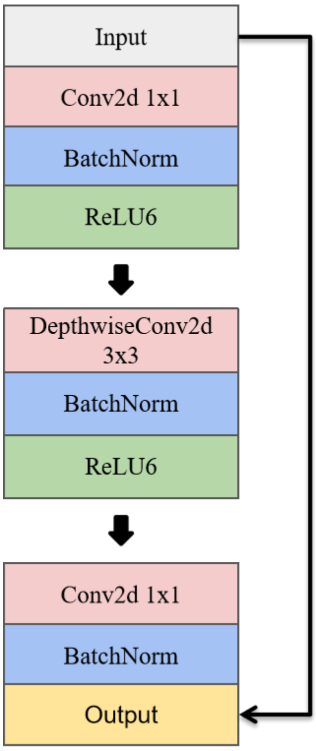

Figure 4. Architecture Diagram of MobileNetV2.

The Inverted Residual Block, a cornerstone of MobileNetV2, is carefully engineered to strike a balance between robust feature extraction and model lightweightness. Comprising two interconnected components, this innovative design leverages the strengths of MobileNetV2:

Main Branch (Left Side): This segment initiates with a 1 × 1 convolution, strategically employed to expand dimensionality without a significant surge in computational complexity. Following this, a 3 × 3 depthwise separable convolution is deployed for capturing intricate features, enhancing the network's capacity to discern fine-grained patterns. The sequence culminates with another 1 × 1 convolution, skillfully tailored to compress dimensionality while retaining crucial information.

Residual Connection (Right Side): A defining aspect of the Inverted Residual Block, this pathway establishes a direct connection between input and output, thereby fostering information flow and facilitating gradient propagation. This architectural innovation significantly contributes to both model performance and training efficiency.

Given our commitment to maintaining robust feature extraction capabilities while minimizing model overhead, our selection of MobileNetV2 as the encoder aligns seamlessly with our objectives. By leveraging the strengths of the Inverted Residual Block, we can harness the advantages of MobileNetV2's efficient and lightweight design, ensuring that our model strikes an optimal balance between computational efficiency and representation power.

3.1.2. Enhancing Feature Extraction and Expanding Receptive Field Using Transformer

In the context of flame semantic segmentation, where the flame regions are typically small and require accurate feature extraction, we propose a method that leverages the Transformer architecture [29,30] to capture a broader range of contextual information and enhance the representation of flame semantics.

The Transformer module(as shown in Figure 5), integrated into our flame semantic segmentation framework, consists of multiple TransformerEncoderLayer modules. These modules enable the network to effectively process the input data and extract discriminative features relevant to flame semantics. During the feature extraction process, the TransformerEncoderLayer module utilizes a self-attention mechanism to capture long-range dependencies between different regions in the input image. By attending to the entire image simultaneously, the Transformer can effectively capture the spatial context of the flame region and its surroundings, even when the flame region is small. During the feature extraction process, the TransformerEncoderLayer module utilizes a self-attention mechanism to capture long-range dependencies between different regions in the input image. By attending to the entire image simultaneously, the Transformer can effectively capture the spatial context of the flame region and its surroundings, even when the flame region is small. By incorporating the Transformer architecture into our flame semantic segmentation framework, our method can effectively extract flame-specific features by capturing extensive contextual information. This enables the model to better understand the spatial relationship between the flame region and its surroundings, leading to improved segmentation accuracy and performance.

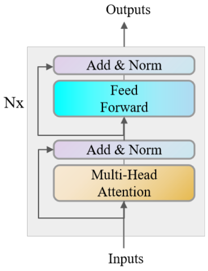

Figure 5. Structure Diagram of Transformer Encoder.

In practical usage, we employ a combined approach by both concatenating and parallelizing Transformer modules in the encoder phase, aiming to effectively extract flame semantic features and address the limited receptive field issue commonly encountered in traditional convolutional neural networks.

In the encoder phase, we first concatenate a Transformer module to extract deep-level information from the images. By utilizing the self-attention mechanism, this Transformer module captures global contextual relationships, aiding in the understanding of spatial characteristics within the flame regions. However, considering that flame regions are typically small, relying solely on a single Transformer module may struggle to accurately capture subtle features.

To overcome this limitation, we further introduce a parallel Transformer module in the encoder phase. The parallel Transformer module aims to preserve the extensive spatial information of the images and provide a broader receptive field. By incorporating both concatenated and parallel Transformer modules, we can leverage the complementary aspects of different layers in feature representation, enabling a more comprehensive capture of the flame region's semantic information.

By simultaneously concatenating and parallelizing Transformer modules, our proposed method harnesses the benefits of deep-level information and an expanded receptive field, thereby enhancing the capability to extract flame semantic features. This architectural design proves valuable in practical applications, augmenting the model's understanding and accuracy in flame region analysis.

The self-attention mechanism serves as the pivotal component within the Transformer encoder, playing a vital role in directing the model's focus towards salient image regions based on their respective significance. This enables the network to emphasize critical information and tailor the extracted features to align with identified targets. In this mechanism, embedded patch vectors are transformed into three distinct vectors: query (Q), key (K), and value (V), which are computed through dot product operations. The correlation between K and Q is assessed via dot product calculation. Subsequent to normalization through scaling and softmax functions, the computed similarity values are utilized to weight the value vector, thereby obtaining semantic importance. Aggregation of all semantic weights facilitates the generation of the self-attention feature. Ultimately, a feature map enriched with substantial information is derived through subsequent processing via a Multi-Layer Perceptron (MLP). This self-attention computation process can be represented as follows:

$$Z = Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_K}})V, \tag{1}$$

where Z is the self-attention feature; d_K is the scaling factor; Q is the query vector; K is the key vector; V is a value vector.

3.1.3. CBAM (Convolutional Block Attention Module) Attention Mechanism

The Convolutional Block Attention Module (CBAM) is an attention mechanism module that combines spatial and channel attention [31] in the convolutional blocks [32]. By integrating both spatial and channel attention mechanisms, CBAM offers improved performance compared to attention mechanisms that focus solely on channel attention, such as SENet [33]. Figure 6 illustrates the overall structure after incorporating the CBAM module. It can be observed that the output of the convolutional layers undergoes a channel attention module, which generates weighted results. Subsequently, the output passes through a spatial attention module before obtaining the final weighted results. The introduction of CBAM aims to enhance the features specifically related to the flame region.

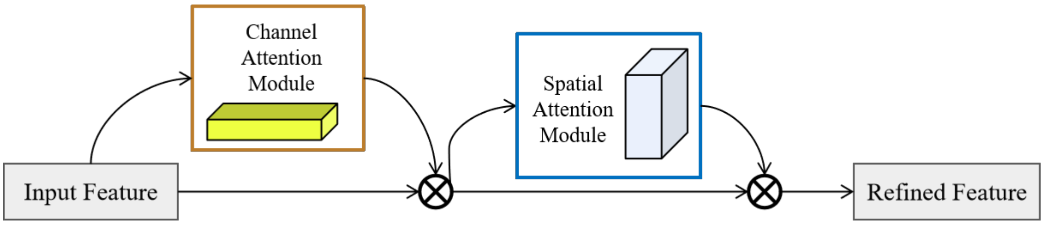

Figure 6. Overview of CBAM Module: Spatial and Channel Attention Mechanisms.

The channel attention module processes the input feature map by applying global max pooling and global average pooling operations based on width and height. Subsequently, each pooled feature is fed through a Multi-Layer Perceptron (MLP). The output features from the MLPs are element-wise summed and passed through a sigmoid activation function to generate the final channel attention feature map. This channel attention feature map is then multiplied element-wise with the input feature map to produce the input features required for the spatial attention module.

The spatial attention module takes the output feature map from the channel attention module as its input. Firstly, a global max pooling and global average pooling operation are performed based on the channels. The results of these operations are then concatenated along the channel dimension. Subsequently, a convolutional operation is applied to reduce the dimensionality to a single channel. The resulting feature map is passed through a sigmoid function to generate the spatial attention feature. Finally, this feature map is multiplied element-wise with the input feature map of this module, yielding the final generated feature.

As illustrated in Figure 7, the integration of the CBAM attention mechanism results in a model that focuses more on the flame's characteristic regions. The fine details of the features become more pronounced, while attention to the background is reduced.

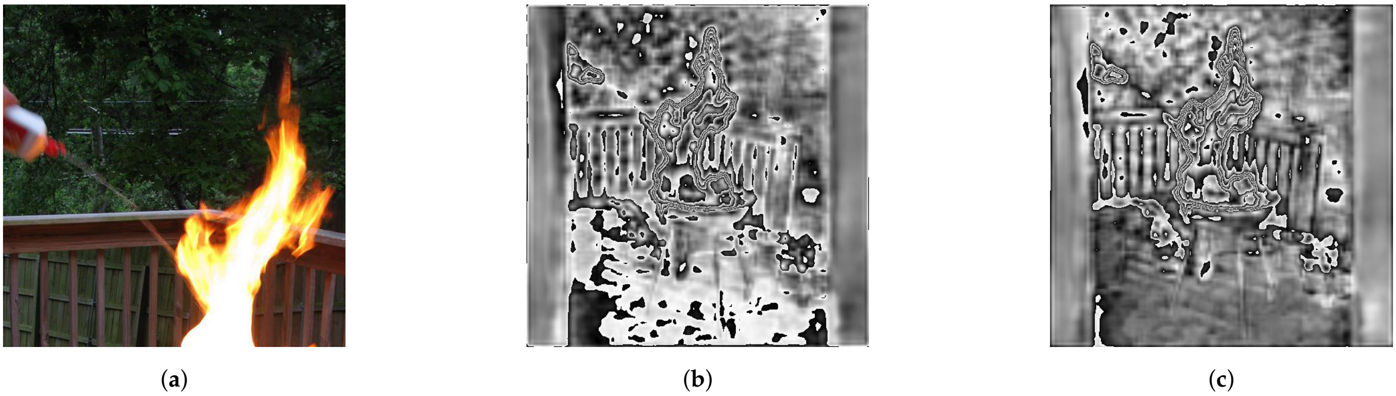

Figure 7. Comparison of Feature Maps before and after Integration of CBAM Attention Mechanism: (**a**) Original Image (**b**) Feature Maps before CBAM Integration (**c**) Feature Maps after CBAM Integration where in (**c**), greater emphasis is placed on capturing flame-specific features, while in (**b**), background features are partly misclassified as flame characteristics.

Discussion: In order to better extract fire-related features and improve model performance, we compared the effectiveness of various attention mechanisms including SE (Squeeze-and-Excitation attention), CAM (Channel Attention Module), SAM (Spatial Attention Module), and CBAM (Convolutional Block Attention Module), as shown in Table 1 and Figure 8. Introducing different attention mechanisms proved beneficial for enhancing the model's learning and segmentation of fire features, with CBAM exhibiting a better focus on fire-related characteristics. Considering factors such as model parameters and overfitting, we opted to solely employ the CBAM attention mechanism in this study.

Table 1. Comparison of Various Attention Mechanism Modules (✓denotes the incorporation of this module and the bolded sections represent the optimal values).

Method				Metric		
SE	CAM	SAM	CBAM	IoU	Recall	Precision
✓				77.06%	85.68%	**88.45%**
	✓			76.74%	87.80%	85.90%
		✓		76.80%	85.93%	85.85%
			✓	**78.12%**	**87.89%**	86.34%

3.1.4. Decoder (DeepLabV3+ Based)

In DeeplabV3+, the enhanced feature extraction network can be divided into two parts:

In the Encoder, the preliminary effective feature maps that have been compressed by a factor of four are processed using parallel Atrous Convolutions. These Atrous Convolutions are performed with different rates to extract features at multiple scales. The resulting feature maps are then merged and further compressed using 1 × 1 convolutions.

In the Decoder, the preliminary effective feature maps that have been compressed by a factor of two are adjusted in terms of channel dimensions using 1 × 1 convolutions. These adjusted feature maps are then stacked with the upsampled feature maps from the output of the Atrous Convolutions. Once the stacking is complete, two rounds of depth-wise separable convolution blocks are applied.

Additionally, DeeplabV3+ incorporates other important components such as the use of dilated convolutions (Atrous Convolutions) to capture multi-scale context information, the application of skip connections to combine features at different levels, and the utilization of depth-wise separable convolutions for efficient computation. These elements collectively contribute to the overall performance improvement and semantic segmentation accuracy achieved by DeeplabV3+.

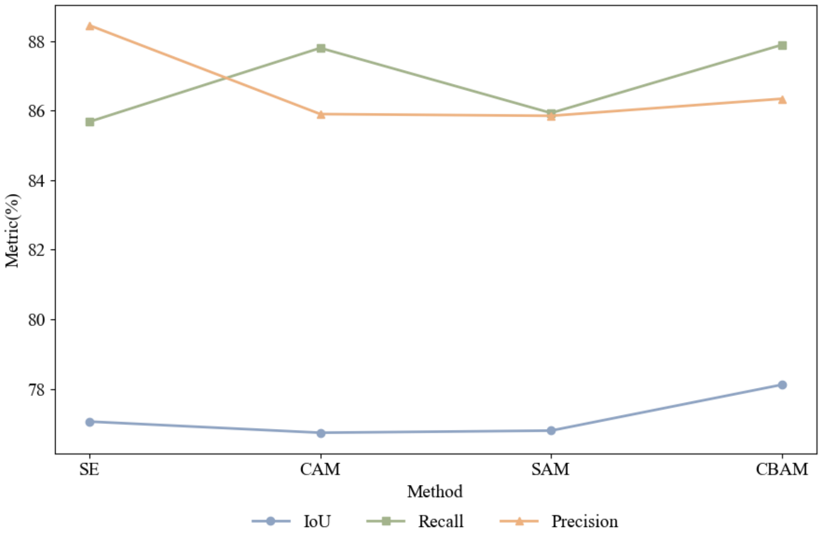

Figure 8. Performance Comparison of Various Attention Mechanisms.

3.2. Adaptive Copy-Paste

The Copy-Paste augmentation method [34] involves the process of pasting objects from one image onto another image, resulting in a diverse set of training data with various choices of source images, object instances, and paste locations. This simple strategy of randomly selecting and pasting objects at random locations has shown significant improvements in model performance across multiple settings.

However, high-resolution and large-scale characteristics of remote sensing images result in a limited proportion of flame regions in general images. This leads to insufficient learning of flame-specific features and the inability of random Copy-Paste methods to augment flame-related features, thereby hindering the improvement of model performance. To address this issue, we propose an adaptive Copy-Paste data augmentation method, which further trains the underrepresented flame regions and enhances the model performance. Compared to traditional techniques such as resampling and undersampling, our method eliminates the need for manual hyperparameter tuning, reduces training time, and does not affect the size of the dataset.

As shown in Figure 9, we introduce a global confidence bank to store the confidence values for each image. Considering that our task aims to segment flame and non-flame regions, we utilize the Intersection over Union(IoU) metric for the flame category as the confidence measure for each image. Specifically, for each image, we initialize the confidence value to 0 and update it during each training iteration using an exponential moving average (EMA) parameter transfer, as described in Equation (2).

$$Con_i = \theta \times Con_i + (1 - \theta) \times iou \tag{2}$$

where Con_i represents the confidence value for the i-th image, while iou represents the IoU metric specific to the flame category. The parameter θ is set to a value of 0.98.

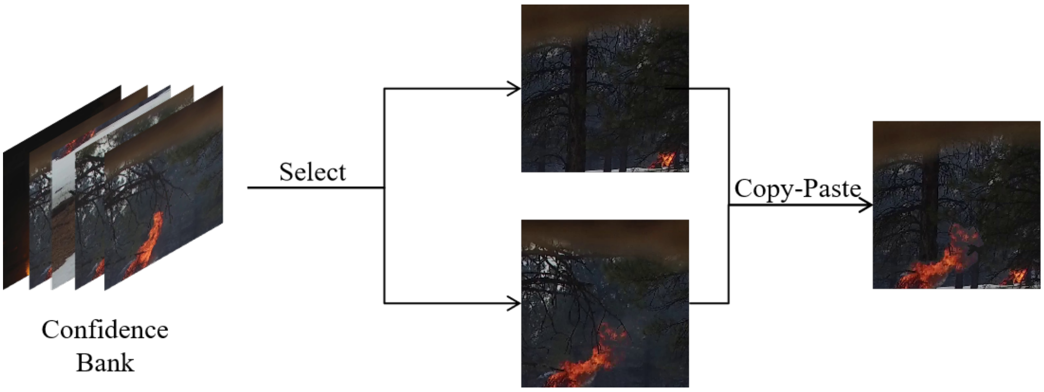

Figure 9. Overview Diagram of the Adaptive Copy-Paste Method.

Upon obtaining the confidence bank, during each training iteration, the confidences within the current confidence bank are initially normalized to probabilities. Subsequently, a random image is selected based on the probabilistic transformation of confidences for each image in the current batch. In this selected image, all pixels belonging to the flame category are then superimposed onto the image with the lowest confidence within the batch. Finally, Gaussian filtering [35] is applied to achieve edge-smoothing effects.

Simultaneously, following the validation of the batch, the confidence values corresponding to the images within the batch are updated using the post-validation IoU metric. This comprehensive approach ensures that the training process incorporates probabilistic image selection, targeted flame category augmentation, and refinement through confidence-based IoU updates.

The pseudo-code for the adaptive Copy-Paste method is shown in Algorithm 1.

Algorithm 1 Adaptive Copy-Paste Augmentation

```
Require: Batch of images
Ensure: Updated batch with pasted flame pixels
Initialize/Update confidence bank $Con$
Normalize all confidences in $Con$ to probabilities
Select image $I_{\text{min}}$ with the lowest confidence in the batch
for each image $I$ in the batch do
  if $I = I_{\text{min}}$ then
    Continue to the next image
  end if
  Randomly select a flame image $I_{\text{flame}}$
  Paste all flame pixels from $I_{\text{flame}}$ onto $I_{\text{min}}$
  Apply Gaussian filter to smooth the edges of $I_{\text{min}}$
end for
```

3.3. Dice Loss

In the context of fire segmentation, where the fire regions constitute a small proportion of the overall image, we introduce the Dice loss as a means to focus on and learn the fire-specific features.

The Dice loss is a widely used loss function in segmentation tasks, aiming to optimize the similarity between the predicted fire segmentation and the ground truth fire mask. It is derived from the Dice coefficient, which measures the overlap or similarity between two binary masks.

The Dice loss [36] is defined as 1 minus the Dice coefficient, and it serves as an objective function to guide the model towards producing more accurate fire segmentations. The Dice coefficient is computed as twice the intersection of the predicted fire mask and the ground truth fire mask, divided by the sum of their areas.

By incorporating the Dice loss during the training process, we encourage the model to focus on and accurately capture the fire regions. The Dice loss penalizes the discrepancies between the predicted and ground truth fire masks, guiding the model to better learn the fire-specific features and improve the segmentation performance.

Since the fire regions are sparse in each image, the Dice loss is particularly beneficial as it can effectively handle class imbalance. It emphasizes the intersection between the predicted and ground truth fire masks, enabling the model to learn the subtle details and boundaries of the fire regions, even in the presence of significant background regions.

The introduction of the Dice loss in our fire segmentation framework addresses the challenge of imbalanced class distribution and enables the model to effectively learn and focus on the fire regions. By optimizing the Dice loss, our model can achieve more accurate and precise fire segmentations, contributing to improved fire detection and analysis tasks.

The Dice loss can be described as follows:

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

where y_i and $\hat{y}_i$ represent the label value and predicted value, respectively, for pixel i in an image. The parameter N represents the total number of pixels, which is equal to the number of pixels in a single image multiplied by the batch size.

4. Data and Experiments

4.1. Data Description

The quality of the dataset and labels significantly influence the training results in the context of fire semantic segmentation tasks. Therefore, we extracted a portion of the publicly available flame dataset and preprocessed it. Additionally, we collected forest fire images from remote sensing sources to create a custom dataset. This dataset not only preserves the flame features but also includes diverse background scenarios, enabling effective segmentation of complex forest conditions.

Specifically, we randomly selected 500 images from the flame dataset [37] and resized them to 512 × 512 dimensions. Images without flame features were removed, and an additional 500 forest fire images of size 512 × 512 were collected from various regions using online sources. In total, the dataset comprises 1000 images. To better validate the effectiveness of our approach, we partitioned the dataset into training, validation, and testing sets in an 8:1:1 ratio. The specific distribution quantities are presented in Table 2.

Table 2. Partitioning of the Dataset.

Dataset	Train	Validation	Test	Summary
Number	800	100	100	1000

Based on the visualization shown in Figure 10, our dataset exhibits severe class imbalance, which provides an opportunity to validate the effectiveness of our proposed method. Furthermore, we present several visualized images from our dataset in Figure 11.

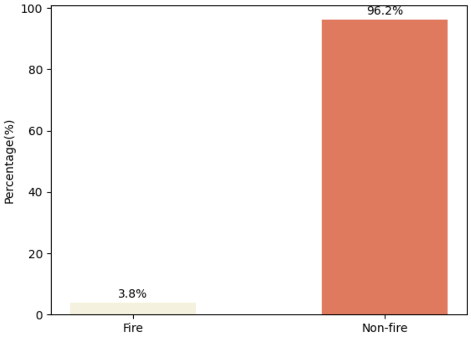

Figure 10. Visualization of Fire and Non-Fire Pixel Proportions in the Dataset.

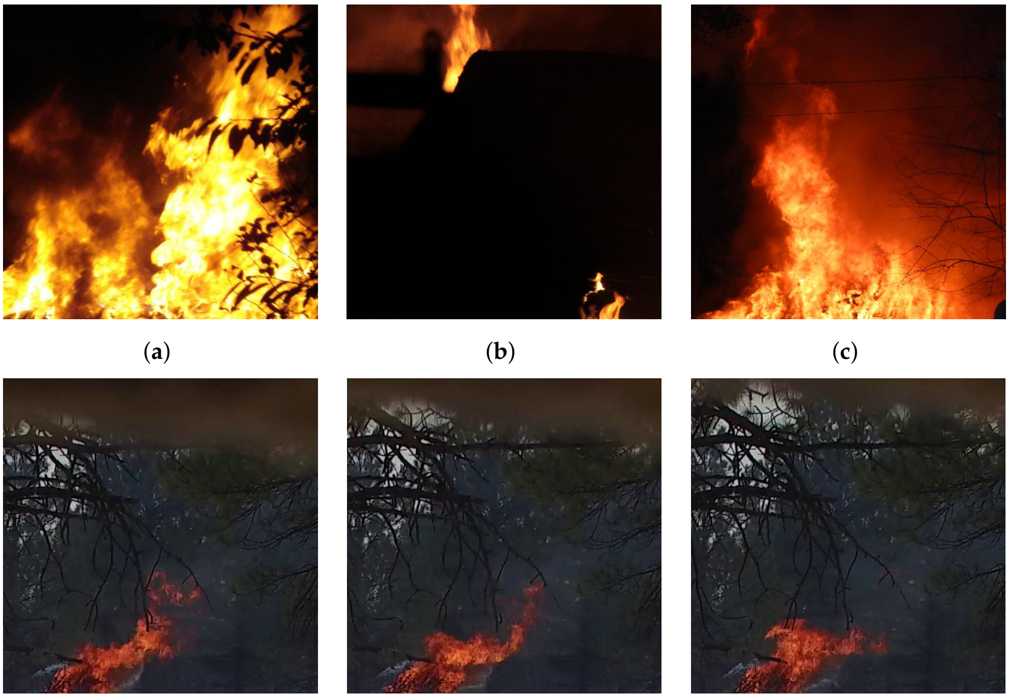
(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 11. Visualization of Several Typical Images in the Dataset ((**a**–**c**) Obtained through network acquisition with complex terrains as the background, (**d**–**f**) Obtained through processing the flame dataset).

Flame: The FLAME dataset is a vital resource for wildfire research, offering aerial imagery captured by UAVs and drones during controlled burns in Northern Arizona. It includes raw drone videos and thermal heatmaps. Aimed at fire classification and segmentation tasks, it provides 39,375 labeled frames for training, 8617 for testing, and 2003 pixel-annotated frames for segmentation. This dataset empowers advanced image analysis, aiding in understanding wildfire behavior for improved management, risk reduction, and ecological preservation.

4.2. Experimental Settings

The experimental settings were carefully configured as follows:

The input images were resized to a shape of [512, 512]. A batch size of 4 was used during training. The initial learning rate was set to 5×10^{-4}, and a minimum learning rate of 0.01 times the initial learning rate was defined for learning rate decay. The optimization algorithm employed was Adam [38] (Adam is an optimization algorithm that combines momentum and RMSProp techniques to dynamically adjust learning rates for individual model parameters, making it effective for a variety of optimization tasks) with a momentum value of 0.9. No weight decay was applied in the training process. The learning rate decay strategy used was cosine annealing, where the learning rate decreases gradually over the course of training. These settings were chosen to ensure a balanced trade-off between model performance and computational efficiency. It is worth noting that we used the same experimental settings when comparing our approach with other state-of-the-art semantic segmentation models.

4.3. Evaluation Metrics

To assess the effectiveness of our proposed method in forest fire segmentation, we employed various evaluation metrics, such as Intersection over Union (IoU) [39], *precision*, and *recall* specifically for the fire class.

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

We computed evaluation metrics using the confusion matrix generated by our improved model, which includes the pixel counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Specifically, IoU measures the similarity between the predicted forest/non-forest areas and the ground truth. Precision and recall evaluate the completeness and accuracy of our method. Our results demonstrate the superior performance of our proposed method in accurately segmenting forest flames, as evidenced by the higher values of these evaluation metrics.

4.4. Results and Analysis

In this section, we compare the performance of our proposed method with several state-of-the-art semantic segmentation approaches, specifically focusing on the IoU, Precision, and Recall metrics for the fire class. For our comparative experiments, we selected a set of representative semantic segmentation networks, which are described in detail below:

FCN (Fully Convolutional Network): FCN [40] is a semantic segmentation network that replaces fully connected layers with convolutional layers, enabling end-to-end pixel-level prediction. It utilizes upsampling and skip connections to capture both local and global context information, resulting in accurate and detailed segmentation maps.

PSPNet (Pyramid Scene Parsing Network): PSPNet [41] is a semantic segmentation model that incorporates a pyramid pooling module to capture multi-scale contextual information. By aggregating features from different pyramid levels, PSPNet effectively captures context at various scales, allowing for robust and precise segmentation of objects in complex scenes.

U-Net: U-Net [42] is a popular network architecture for biomedical image segmentation. It consists of an encoder-decoder structure with skip connections. The encoder captures contextual information, while the decoder recovers spatial details using skip connections. U-Net is known for its ability to handle limited training data and produce accurate segmentation results.

DeepLabV3+: DeepLabV3+ [43] is an advanced semantic segmentation model that combines the strengths of DeepLabV3 and a modified encoder-decoder architecture. It utilizes atrous convolution and a multi-scale feature fusion module to capture fine-grained details and context information. DeepLabV3+ also incorporates a spatial pyramid pooling module to handle objects at different scales. This network achieves state-of-the-art performance in semantic segmentation tasks.

The validation results are shown in Table 3 and Figure 12. Despite achieving state-of-the-art performance in current mainstream semantic segmentation tasks, networks such as deeplabV3+ struggle in the specific context of forest fire segmentation due to the extreme class imbalance of the fire class. This hinders the effective learning of fire-specific features by these advanced networks. On the other hand, Unet, with its unique architecture, is capable of handling limited training data and producing accurate segmentation results. Consequently, in the comparison of base models, Unet outperforms other base models in all metrics. Our proposed model, built upon the deeplabV3+ framework, addresses these limitations through various design improvements. The validation results are shown in Table 3. Despite achieving state-of-the-art performance in current mainstream semantic segmentation tasks, networks such as deeplabV3+ struggle in the specific context of forest fire segmentation due to the extreme class imbalance of the fire class. This hinders the effective learning of fire-specific features by these advanced networks. On the other hand, Unet, with its unique architecture, is capable of handling limited training data and producing accurate segmentation results. Consequently, in the comparison of base models, Unet outperforms other base models in all metrics. Our proposed model, built upon the deeplabV3+ framework, addresses these limitations through various design improvements. As a result, our model achieves a 6.67% improvement in IoU, a 5.23% improvement in Precision, and a 3.27% improvement in Recall compared to the base model. Furthermore, our model also surpasses Unet in all metrics.

Table 3. Comparison of Semantic Segmentation Methods' Performance (The bolded sections represent the optimal values).

Model	IoU	Precision	Recall
FCN	62.90%	82.91%	63.00%
Unet	79.63%	88.40%	88.92%
PSPNet	71.16%	71.73%	82.46%
DeeplabV3+	72.69%	82.78%	85.64%
FlameTransNet (Ours)	**83.72%**	**91.88%**	**90.41%**

In Figure 13, we provide visual representations of the prediction results obtained from different models. To ensure the inclusion of diverse scenarios, we carefully selected four typical situations to evaluate the models' performance. In the first column, which corresponds to images with a higher proportion of fire, our model demonstrates superior accuracy compared to the other models. Even the Unet model shows noticeable false detections, whereas our model consistently produces relatively accurate predictions. Moving to the second column, where the fire class is less prominent, our model showcases remarkable completeness in comparison to the competing models. In the third column, depicting fire-absent scenarios, our model effectively avoids false detections altogether. Furthermore, in the fourth column, which presents fire images with complex backgrounds involving objects such as humans, trees, and smoke, our model accurately delineates the fire regions. Taking all these distinct scenarios into consideration, our model consistently outperforms mainstream semantic segmentation networks both in terms of quantitative analysis and qualitative evaluation, thereby establishing its superior performance and reliability.

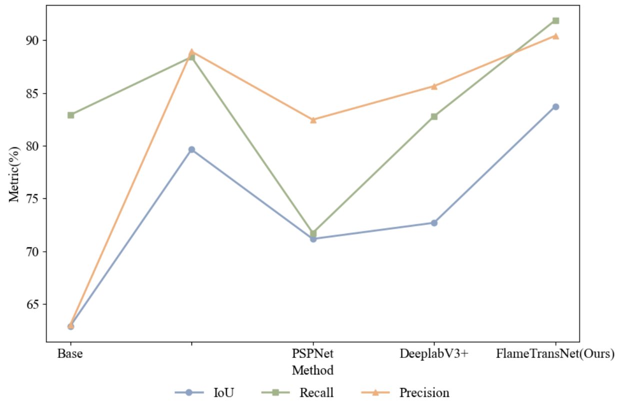

Figure 12. Visualization of Performance Comparison Among Various Semantic Segmentation Methods.

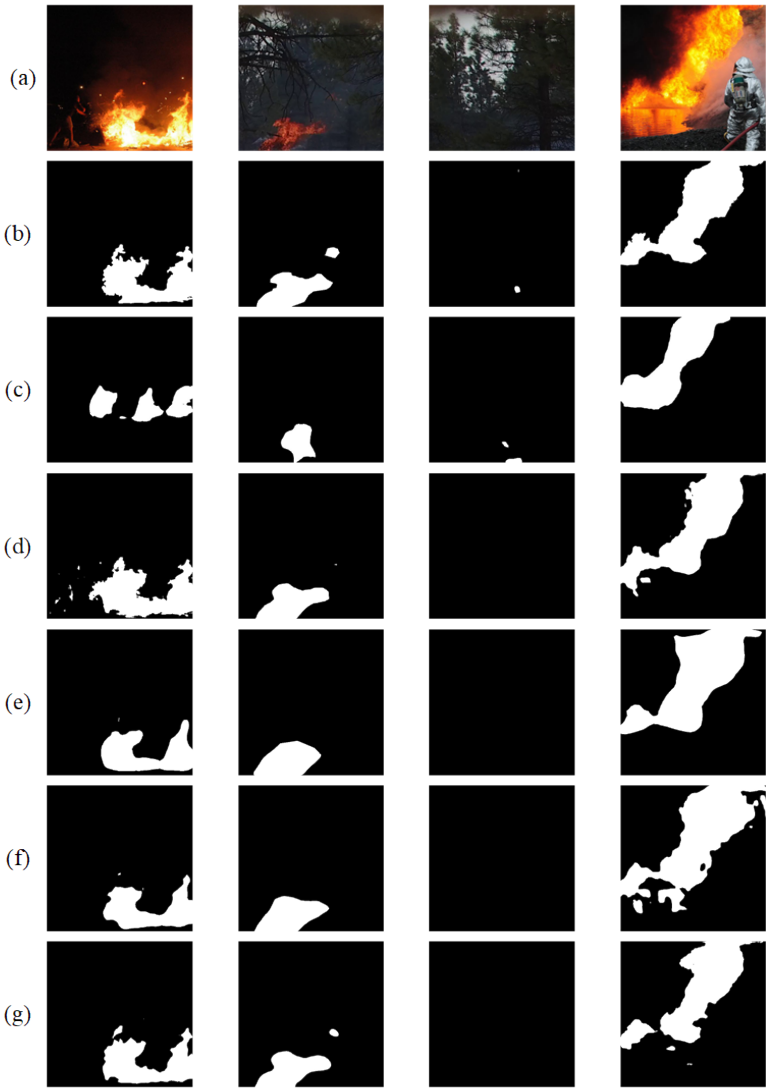

Figure 13. Visual Comparison of Prediction Results from Various Models: (**a**) Image, (**b**) Ground Truth, (**c**) FCN, (**d**) Unet, (**e**) PSPNet, (**f**) DeeplabV3+, (**g**) FlameTransNet (Ours).

4.5. Ablation Experiments

To further validate the effectiveness of our proposed method, we conducted ablation experiments. Specifically, our method consists of the fusion of Transformer, the incorporation of CBAM, the adoption of the adaptive Copy-Paste data augmentation method (ACP), and the integration of Dice loss (DL).

In Table 4, we demonstrate the impact of incorporating each method on various performance metrics of the model. It can be observed that our proposed methods contribute to the improvement of model performance. Specifically, the inclusion of the CBAM results in a 5.43% increase in IoU. Furthermore, the introduction of Transformer Module leads to an additional 2.42% improvement in IoU. Subsequently, with the incorporation of the adaptive Copy-Paste method and Dice loss, the performance of the model is further enhanced.

Table 4. Ablation Experiments of the Proposed Methods on the Dataset (✓denotes the incorporation of this module and the bolded sections represent the optimal values).

Method					**Metric**		
Base	**CBAM**	**Transformer**	**ACP**	**DL**	**IoU**	**Precision**	**Recall**
✓					72.69%	82.78%	85.64%
✓	✓				78.12%	86.34%	87.89%
✓	✓	✓			80.54%	87.64%	88.38%
✓	✓	✓	✓		82.93%	88.71%	89.98%
✓	✓	✓	✓	✓	**83.72%**	**91.88%**	**90.41%**

5. Conclusions

In conclusion, our study focused on semantic segmentation of forest flames using advanced techniques such as the fusion of Transformer, the incorporation of CBAM, the adoption of the adaptive Copy-Paste data augmentation method, and the integration of Dice loss. We demonstrated the effectiveness of our proposed method through comprehensive evaluations and comparisons with state-of-the-art semantic segmentation models. The accurate segmentation and recognition of forest flames are of great importance for real-time monitoring, emergency response, and forest management decisions. Our method offers robust support for early detection and control of forest fires, minimizing ecological damage and ensuring the sustainable utilization of forest resources. By leveraging computer vision and deep learning techniques, we contribute to the development of effective solutions for forest fire management, thus mitigating the adverse impacts of forest fires on ecosystems and human society. Future research directions can include exploring the application of our method in real-time fire monitoring systems and extending its capabilities to handle different types of fire scenarios with improved accuracy and efficiency.

Although our proposed method has achieved promising results, we believe there are several directions for further research in the future:

- When considering the dynamic update of the confidence bank using the Exponential Moving Average (EMA) method, we used a standard configuration with the parameter value θ set to 0.98. However, we did not extensively explore this parameter further. In the future, there is potential to investigate more refined updating methods and optimal parameter values.
- During both training and testing, we employed an approach of resizing images to a specific size to reduce training duration and enhance training effectiveness. This resized configuration was used to validate the effectiveness of our method. However, for real-world applications, post-processing is required to restore the predicted results to the original image size. In future work, there is potential to explore techniques for effectively handling high-resolution image training and addressing the challenges associated with such data.

- The detection of smoke holds significant importance in fire detection as well. Efforts can be made to transfer the methodologies to smoke detection or to integrate smoke detection techniques, thereby enabling further explorations in this domain.

Author Contributions: B.C. devised the programs, drafted the initial manuscript, and contributed to writing enhancements. D.B. and W.J. helped with data collection and data analysis. H.L. and D.B. designed the project and revised the manuscript. All authors have read and agree 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: All data generated or presented in this study are available upon request from the corresponding author. Furthermore, the models and code used during the study cannot be shared at this time as the data also form part of an ongoing study.

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

References

1. Ryu, J.H.; Han, K.S.; Hong, S.; Park, N.W.; Lee, Y.W.; Cho, J. Satellite-Based Evaluation of the Post-Fire Recovery Process from the Worst Forest Fire Case in South Korea. *Remote Sens.* **2018**, *10*, 918. [CrossRef]
2. Houle, G.P.; Kane, E.S.; Kasischke, E.S.; Gibson, C.M.; Turetsky, M.R. Recovery of carbon pools a decade after wildfire in black spruce forests of interior Alaska: Effects of soil texture and landscape position. *Can. J. For. Res.* **2018**, *48*, 1–10.
3. White, J.C.; Wulder, M.A.; Hermosilla, T.; Coops, N.C.; Hobart, G.W. A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat time series. *Remote Sens. Environ.* **2017**, *194*, 303–321. [CrossRef]
4. Attri, V.; Dhiman, R.; Sarvade, S. A review on status, implications and recent trends of forest fire management. *Arch. Agric. Environ. Sci.* **2020**, *5*, 592–602. [CrossRef]
5. Yun, T.; Jiang, K.; Li, G.; Eichhorn, M.P.; Fan, J.; Liu, F.; Chen, B.; An, F.; Cao, L. Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach. *Remote Sens. Environ.* **2021**, *256*, 112307. [CrossRef]
6. Li, X.Y.; Jin, H.J.; Wang, H.W.; Marchenko, S.S.; Shan, W.; Luo, D.L.; He, R.X.; Spektor, V.; Huang, Y.D.; Li, X.Y.; et al. Influences of forest fires on the permafrost environment: A review. *Adv. Clim. Chang. Res.* **2021**, *12*, 48–65.
7. Jiao, Z.; Zhang, Y.; Xin, J.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. A Deep Learning Based Forest Fire Detection Approach Using UAV and YOLOv3. In Proceedings of the 2019 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 23–27 July 2019; pp. 1–5. [CrossRef]
8. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [CrossRef]
9. Khanal, S.; Kc, K.; Fulton, J.P.; Shearer, S.; Ozkan, E. Remote sensing in agriculture—Accomplishments, limitations, and opportunities. *Remote Sens.* **2020**, *12*, 3783. [CrossRef]
10. Winberg, O.; Pyörälä, J.; Yu, X.; Kaartinen, H.; Kukko, A.; Holopainen, M.; Holmgren, J.; Lehtomäki, M.; Hyyppä, J. Branch information extraction from Norway spruce using handheld laser scanning point clouds in Nordic forests. *ISPRS Open J. Photogramm. Remote Sens.* **2023**, *9* , 100040. [CrossRef]
11. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Wildfire Segmentation Using Deep Vision Transformers. *Remote Sens.* **2021**, *13*, 3527. [CrossRef]
12. Zheng, W.; Chen, J.; Fan, J.; Li, Y.; Liu, C. Wildfire Monitoring Using Infrared Bands and Spatial Resolution Effects. In *Vegetation Fires and Pollution in Asia*; Vadrevu, K.P., Ohara, T., Justice, C., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 21–33. [CrossRef]
13. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [CrossRef]
14. Zheng, S.; Gao, P.; Zhou, Y.; Wu, Z.; Wan, L.; Hu, F.; Wang, W.; Zou, X.; Chen, S. An Accurate Forest Fire Recognition Method Based on Improved BPNN and IoT. *Remote Sens.* **2023**, *15*, 2365. [CrossRef]
15. Zheng, H.; Dembélé, S.; Wu, Y.; Liu, Y.; Chen, H.; Zhang, Q. A lightweight algorithm capable of accurately identifying forest fires from UAV remote sensing imagery. *Front. For. Glob. Chang.* **2023**, *6*, 1134942. [CrossRef]
16. Shiklomanov, A.N.; Bradley, B.A.; Dahlin, K.M.; Fox, A.M.; Gough, C.M.; Hoffman, F.M.; Middleton, E.M.; Serbin, S.P.; Smallman, L.; Smith, W.K. Enhancing global change experiments through integration of remote-sensing techniques. *Front. Ecol. Environ.* **2019**, *17*, 215–224.
17. Wang, Z.; Peng, T.; Lu, Z. Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks. *Forests* **2022**, *13*, 1133. [CrossRef]

18. Avula, S.B.; Badri, S.J.; Reddy, P, G. A Novel Forest Fire Detection System Using Fuzzy Entropy Optimized Thresholding and STN-based CNN. In Proceedings of the 2020 International Conference on COMmunication Systems and NETworkS (COMSNETS), Bengaluru, India, 7–11 January 2020; pp. 750–755. [CrossRef]
19. Tsalera, E.; Papadakis, A.; Voyiatzis, I.; Samarakou, M. CNN-based, contextualized, real-time fire detection in computational resource-constrained environments. *Energy Rep.* **2023**, *9*, 247–257.
20. 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]
21. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
22. 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 (SMC), Melbourne, Australia, 17–20 October 2021; pp. 2109–2114. [CrossRef]
23. Zhang, L.; Wang, M.; Ding, Y.; Wan, T.; Qi, B.; Pang, Y. FBC-ANet: A Semantic Segmentation Model for UAV Forest Fire Images Combining Boundary Enhancement and Context Awareness. *Drones* **2023**, *7*, 456. [CrossRef]
24. Wang, G.; Zhang, Y.; Qu, Y.; Chen, Y.; Maqsood, H. Early Forest Fire Region Segmentation Based on Deep Learning. In Proceedings of the 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 6237–6241. [CrossRef]
25. Alqourabah, H.; Muneer, A.; Fati, S.M. A Smart Fire Detection System using IoT Technology with Automatic Water Sprinkler. *Int. J. Electr. Comput. Eng.* **2021** , *11*, 2994–3002. [CrossRef]
26. Peruzzi, G.; Pozzebon, A.; Van Der Meer, M. Fight Fire with Fire: Detecting Forest Fires with Embedded Machine Learning Models Dealing with Audio and Images on Low Power IoT Devices. *Sensors* **2023**, *23*, 783. [CrossRef] [PubMed]
27. Kinaneva, D.; Hristov, G.; Raychev, J.; Zahariev, P. Early Forest Fire Detection Using Drones and Artificial Intelligence. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 20–24 May 2019; pp. 1060–1065. [CrossRef]
28. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
29. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in Transformer. In *Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Montreal, QC, Canada 2021; Volume 34, pp. 15908–15919.
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Montreal, QC, Canada, 2017; Volume 30.
31. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
34. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
35. Afshari, H.H.; Gadsden, S.A.; Habibi, S. Gaussian filters for parameter and state estimation: A general review of theory and recent trends. *Signal Process.* **2017**, *135*, 218–238. [CrossRef]
36. Zhao, R.; Qian, B.; Zhang, X.; Li, Y.; Wei, R.; Liu, Y.; Pan, Y. Rethinking Dice Loss for Medical Image Segmentation. In Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2020; pp. 851–860. [CrossRef]
37. 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: New York, NY, USA, 2020. [CrossRef]
38. Jais, I.K.M.; Ismail, A.R.; Nisa, S.Q. Adam Optimization Algorithm for Wide and Deep Neural Network. *Knowl. Eng. Data Sci.* **2019**, *2*, 41–46. [CrossRef]
39. van Beers, F.; Lindström, A.; Okafor, E.; Wiering, M.A. Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation. In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods—ICPRAM, Prague, Czech Republic, 19–21 February 2019; SciTePress: Setubal, Portugal, 2019; pp. 438–445. [CrossRef]
40. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA , 7–12 June 2015.
41. 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 (CVPR), Honolulu, HI, USA, 21–26 July 2017.

42. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
43. 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.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.