


## Article

# Forest Fire Smoke Detection Based on Multiple Color Spaces Deep Feature Fusion

Ziqi Han <sup>1,†</sup>, Ye Tian <sup>1,†</sup>, Change Zheng <sup>1,\*</sup>  and Fengjun Zhao <sup>2,\*</sup>

<sup>1</sup> School of Technology, Beijing Forestry University, Beijing 100083, China; ziqi\_han@bjfu.edu.cn (Z.H.); tytoemail@bjfu.edu.cn (Y.T.)

<sup>2</sup> Key Laboratory of Forest Protection of National Forestry and Grassland Administration, Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing 100091, China

\* Correspondence: zhengchange@bjfu.edu.cn (C.Z.); zhaofj@caf.ac.cn (F.Z.)

† These authors contributed equally to this work.

**Abstract:** The drastic increase of forest fire occurrence, which in recent years has posed severe threat and damage worldwide to the natural environment and human society, necessitates smoke detection of the early forest fire. First, a semantic segmentation method based on multiple color spaces feature fusion is put forward for forest fire smoke detection. Considering that smoke images in different color spaces may contain varied and distinctive smoke features which are beneficial for improving the detection ability of a model, the proposed model integrates the function of multi-scale and multi-type self-adaptive weighted feature fusion with attention augmentation to extract the enriched and complementary fused features of smoke, utilizing smoke images from multi-color spaces as inputs. Second, the model is trained and evaluated on part of the FlgLib dataset containing high-quality smoke images from watchtowers in the forests, incorporating various smoke types and complex background conditions, with a satisfactory smoke segmentation result for forest fire detection. Finally, the optimal color space combination and the fusion strategy for the model is determined through elaborate and extensive experiments with a superior segmentation result of 86.14 IoU of smoke obtained.

**Keywords:** forest fire smoke segmentation; color spaces; features fusion; self-adaptive weights



**Citation:** Han, Z.; Tian, Y.; Zheng, C.; Zhao, F. Forest Fire Smoke Detection Based on Multiple Color Spaces Deep Feature Fusion. *Forests* **2024**, *15*, 689. <https://doi.org/10.3390/f15040689>

Academic Editor: Manuel Marey-Pérez

Received: 5 March 2024

Revised: 31 March 2024

Accepted: 8 April 2024

Published: 11 April 2024



**Copyright:** © 2024 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 global increase in extreme weather events has led to a significant rise in the frequency and size of wildfires in recent years. Wildfires pose a critical threat to natural resources, economic interests, and the security of human societies [1–6]. Early intervention in the growth of fires through vision-based early forest fire automated monitoring and detection technologies is essential to minimize losses.

Forest fire monitoring and detection involve the detection of flames and smoke. While there has been considerable research on flame detection [7–13], smoke detection is more important in the early stages of a forest fire because smoke tends to appear earlier than flames [14]. There have been numerous research works published on smoke detection [15–17].

The methodology of smoke detection consists of traditional methods and deep learning methods [18]. Traditional methods require manually designing and selecting appropriate features based on the characteristics of forest fire smoke, which can significantly affect the behavior of the classifier. However, this manual process is time-consuming and relies heavily on the expertise and understanding of the designer. Additionally, the appearance of smoke can vary depending on factors such as fuel type, burning conditions, and airflow [19], making it challenging to design generalized and robust features. As a result, traditional methods may have limited generalization capability and can lead to false positive and false negative detections in new scenarios.

On the other hand, deep learning methods offer automatic extraction of richer features [20], making it easier to leverage a wider variety of data types and fuse their corresponding features for better performance. Deep learning-based smoke detection (including forest fire smoke) has been widely studied and applied in the field. Compared to smoke classification [21–26] and smoke detection tasks [27–33], smoke segmentation not only detects the presence of smoke and its location but also provides information about the approximate smoke area and boundary contour [2,3,34–39]. This additional information can be valuable in assessing the scale of the fire and predicting its potential spread. Researchers have proposed various methods for smoke segmentation, including a concentration weighting-based approach to address the challenges posed by the transparency, fuzzy contour, and concentration diversity of smoke. Wang et al. [40] proposed a forest fire smoke semantic segmentation method based on concentration weighting to address the problem of smoke label uncertainty leading to the degradation of smoke segmentation capability caused by the transparency, fuzzy contour, and concentration diversity possessed by smoke in the supervised smoke segmentation task. The method built the mathematical relationship between smoke concentrations and its pixel values utilizing the clue that different concentrations of smoke reflect different pixel values in an image. Then, the authors jointly trained the model with concentration weighting labels and basic labels, making the network realize the distinct role of forest fire smoke pixels, which reduced the influence on the smoke detection model brought by the uncertainty introduced from smoke data annotation.

Feature fusion [41] has been widely applied in image fusion [42] and visual recognition. It combines the extracted features of images with similar content and objects but different attributes, such as RGB and IR image pairs, near-focus and far-focus image pairs, and low-light and over-exposed image pairs. By recovering the fused features, deep learning-based image fusion produces a fused image that simultaneously exhibits the corresponding attributes of the original two images. This fused image enhances and complements the characteristics of images with different attributes, enabling downstream visual recognition tasks. For example, in the case of forest fire detection, the presence of smoke in the early stages can obstruct the detection of the fire source and hinder firefighting operations. To address this issue, Liu et al. [43] fused RGB and IR image pairs captured by a UAV drone to obtain fused images that contain both environmental information and a clear view of the burning fire source. Then, the fused image was input into an object detection network to detect smoke and obscured fire. Similarly, deep learning-based feature fusion can be directly applied to object recognition tasks. Compared to the complex two-stage recognition process, direct fusion of features extracted from multi-modal data simplifies the task in a visual recognition model. This approach allows the feature extractor to capture more diverse multi-modal features and retain more information from the original image data, resulting in improved final predictions.

In remote sensing imagery semantic segmentation, Zhao et al. [44] achieved superior segmentation results by directly fusing features from RGB images and depth images into their semantic segmentation model. Additionally, Chen et al. [45] conducted unified forest fire and smoke detection using RGB and IR image pairs captured by a UAV. They utilized the feature fusion method to combine the characteristics of both RGB and IR images. The fusion methods were divided into early fusion, which involved concatenating images as input into the model, and late fusion, which involved directly fusing the extracted features for the recognition task. The resulting features were then fed into a classifier to predict the presence of fire and smoke. The late fusion approach yielded significantly better classification results compared to single-channel input.

In addition to the feature fusion, fire detection has been studied in different color spaces. Color space refers to a system used to describe and represent colors in computer vision. Common color spaces, such as RGB, HSV, and YCbCr, offer different ways of representing colors and can capture specific color details of target objects, such as fire. Fire detection based on color spaces has been extensively studied in previous research [16]. Flame detection, in particular, has been widely explored using traditional and deep learning-

based methods due to the distinct color of flames [7]. For example, Zeineb D et al. [46] utilized the PJF color space to better highlight fire regions in flame images as a preprocessing step before feeding them into a lightweight network they had designed. Haridasan et al. [47] applied multiple color space transforms to RGB fire images with flame areas and fused the resulting features using concatenation to improve classification performance. Xing D et al. [48] achieved refined smoke segmentation results by merging smoke segmentation regions using HSV and LAB color spaces. Similarly, Prema E et al. [49] and Pundir, A.S. et al. [50] used color spaces (YUV and YCbCr, respectively) as color criteria and combined them with other feature extraction analyses to detect smoke presence in video frames.

However, using strong neural networks for efficient feature extraction from multiple color spaces is still lacking in these studies. Moreover, while the research on flame detection based on color spaces is well-established, the research on smoke detection using deep feature fusion from multiple color spaces requires further development. With multi-color spaces and feature fusion, it is possible to highlight distinct smoke features and enhance the final detection results. This research direction represents an advancement in the field.

This paper introduces SUFN (Smoke U-Shape Fusion Network), a semantic segmentation network for forest fire smoke. The proposed method fuses features from multiple color spaces to enhance the visual deep learning-based approach into the forest fire smoke segmentation direction in the hope of supplementing and completing the research field of forest fire smoke detection. The method leveraged the idea of feature fusion based on deep learning semantic segmentation in order to obtain a superior smoke detection performance under various and complex situations in real application by more conveniently and efficiently satisfying the enrichment and complementarity of distinctive features of smoke images in multiple color spaces.

The main contributions of this paper are as follows:

- (1) Compared with the smoke data selected or captured by humans in previous studies, this research utilizes a dataset of forest fire smoke derived from real-time forest fire monitoring cameras. This dataset encompasses a variety of real-world scenarios, providing a more impartial assessment of forest fire smoke detection models.
- (2) Our study introduces SUFN, a semantic segmentation model explicitly designed for forest fire smoke detection. SUFN builds upon the U-Net foundation and innovatively integrates features from multiple color spaces using three specialized fusion modules: multi-scale feature encoding, deep feature fusion, and multi-scale shallow feature fusion.
- (3) We present a novel local fusion strategy employing element-wise self-adaptive weighted addition, with an adaptive fusion weight policy devised for different local contexts and dependencies within the model. It enables the extraction of comprehensive, complementary smoke features from the diverse color spaces, significantly enhancing the model's detection performance.

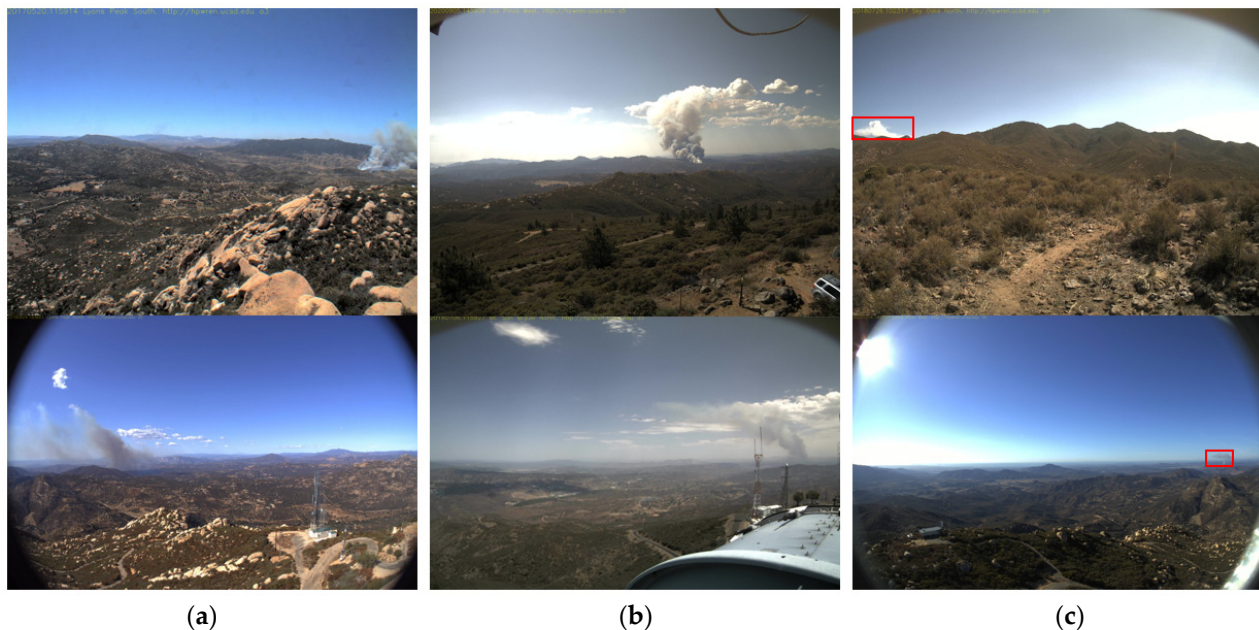
## 2. Materials and Methods

### 2.1. Dataset and Preprocessing

The forest fire smoke image dataset used in the study was from part of the Fire Ignition Library—FIgLib [4]. FIgLib is a large, publicly available wildfire smoke detection dataset containing wildfire image sequences selected from images taken by the fixed-view cameras of the High Performance Wireless Research and Education Network—HPWREN on remote mountain tops in Southern California, USA. This research was one of the few to have used such a high-quality and authentic forest fire smoke dataset for detection by semantic segmentation.

In this study, there were 14 fire videos selected from the FIgLib database from June 2016 through to December 2021. These videos have varying degrees of smoke scale and scene complexity. We extracted sequences of smoke images from the start of ignition to the smoke generation process to construct the dataset. To evaluate better the performance of the model, in this paper, we defined three types of smoke in the FIgLib database after thorough

observation and statistical analysis: Normal clear smoke (NCS), normal interference smoke (NSI), and inconspicuous smoke (IS). The NCS represented easily visible smoke, occupying a relatively large proportion of the image without any surrounding interference. The NSI represented similar conspicuous smoke but difficult to recognize the smoke contours with some interferences. Interferences usually include clouds, glare, smoke, or a background mixed with smoke, etc. The IS represented smoke occupied a very small proportion in an image due to the incipient stage of fire ignition or a very distant fire shot from the camera, or thin faint smoke with low concentration. Figure 1 shows three examples of images of smoke.



**Figure 1.** Examples of images of smoke. (a) Normal clear smoke as NCS. (b) Normal Interference interference smoke as NSI. (c) Inconspicuous smoke as IS. The red boxes in sub-figure (c) indicate the smoke occurrence area.

Hence, the dataset comprised 3 NCS, 9 NSI, and 2 IS fire video clips, representing the complexity of the data. We randomly selected 3 of the 14 forest fire videos and sampled 71 images based on smoke type to form the test set consisting of 29 NCS, 22 NSI, and 20 IS smoke images. The remaining 11 videos sampled 281 smoke images (including all three smoke types) as the training set. The ratio of the training set to the test set is approximately 8:2, while ensuring that the training and test sets do not overlap. The dataset consisted of 352 high-resolution forest fire smoke RGB images, with resolutions of  $2048 \times 3072$ ,  $1536 \times 2048$ , or  $1200 \times 1600$  pixels. The details of the dataset are shown in Table 1.

It is important to note that the ratio of NCS, NSI, and IS types in our dataset was 3:8:3. This ratio closely aligned with the statistical ratio of the original FlgLib dataset. During the statistical process, we observed that the NSI type of fire situations occurred more frequently than the other two types, especially compared to the IS type. The IS type of fire situations occurred the least, resulting in a relatively small amount of fire samples of this type in our dataset. Considering that the NCS type, as an easy sample, would not significantly contribute to the model's performance or accurately represent real-world scenarios, we intentionally reduced the number of NCS samples in the dataset.

Unfortunately, we could not obtain the corresponding ground truth labels for the images in the original dataset. We annotated the images in the dataset using the online annotation site CVAT [51] to obtain smoke segmentation masks. In training, we performed geometric transformations such as random cropping, random scaling, and random flipping on the original training images. The image size remained at  $512 \times 512$  pixels.



Table 1. Fires data details in the dataset.

Data	Datetime of the Fire	Location of the Fire	Smoke Type	#Images	Total
Training	11:19 on 20 May 2017	Lyons Peak South	NCS	38	281
	14:19 on 13 August 2019	L. A. Co. F. D Helibase 69 Bravo East	NCS	34	
	11:30 on 6 July 2018	Mt. San Miguel North	NSI	24	
	13:31 on 27 July 2018	Mt. Woodson North	NSI	29	
	15:03 on 29 May 2019	Otay Mountain North	NSI	23	
	13:03 on 16 July 2019	Mesa Grande North	NSI	21	
	18:34 on 6 August 2020	Otay Mountain North	NSI	24	
	10:56 on 29 August 2020	Cuyamaca Peak South	NSI	24	
	14:25 on 5 September 2020	Los Pinos West	NSI	23	
	09:43 on 26 July 2018	Sky Oaks North	IS	24	
	13:34 on 16 December 2020	Lyons Peak West	IS	17	
Testing	14:56 on 24 September 2019	Lyons Peak North	NCS	29	71
	14:29 on 5 September 2020	Mt. San Miguel East	NSI	22	
	11:55 on 25 July 2018	High Point North	IS	20	

## 2.2. Model Architecture

Inspired by the idea of feature fusion of multi color spaces of images, we chose the semantic segmentation network U-Net [52] as baseline model and modified it to achieve multi-space feature fusion. The overview of the proposed forest fire smoke segmentation model SUFN (Smoke U-Shape Fusion Network) is in Figure 2.

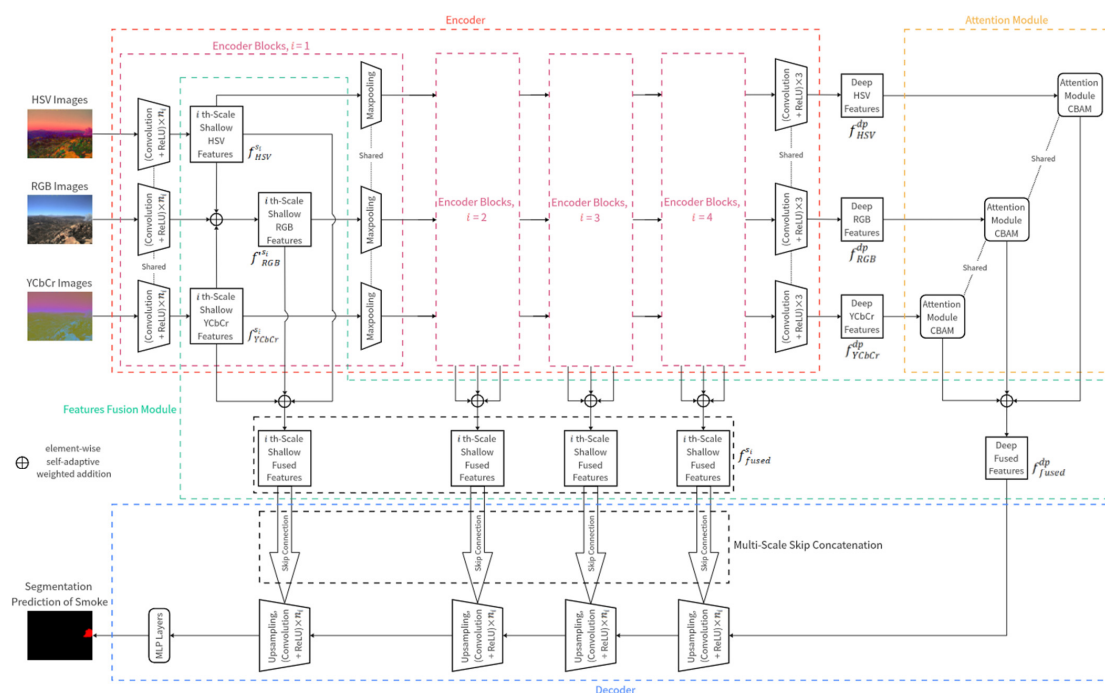


Figure 2. Overview of the proposed SUFN architecture.

The U-Net semantic segmentation model can segment the small objects in images as mentioned in previous research, which was consistent with our observation of segmenting smoke at the early stage in our case. Therefore, we designed a U-shape encoder–decoder segmentation model for smoke segmentation based on the paradigm of U-Net architecture, combined with the feature fusion method of multi color spaces of images, namely the smoke U-shape fusion network (SUFN). The proposed model contained four basic modules, named as encoder module, attention module, feature fusion module, and decoder module.

### 2.2.1. Encoder Module

In the inputs and encoder module, the input images of HSV and YCbCr color spaces  $x_{HSV}$ ,  $x_{YCbCr}$  with shape  $512 \times 512 \times 3$  were fed into the encoder part of the model to extract the corresponding deep features  $f_{HSV}^{dp}$ ,  $f_{YCbCr}^{dp}$  with equal shape  $32 \times 32 \times 512$ .

The backbone used in the study as model encoder was based on VGG-16 [53] with fully-connected layers removed for future operations. Additionally, to enhance the features extracted from the RGB space which was easy for collecting, observing, and utilizing in reality, we fused the RGB features at multiple middle scales with the corresponding middle scale features from HSV and YCbCr spaces and made the RGB space dominant with larger fusion weights when fusing its features. Specifically, we first input the images of RGB space with the equivalent size to the encoder and obtained the non-fused shallow features  $f_{RGB}^{si}$  at the first scale as Equation (1):

$$f_{RGB}^{si} = \theta_{ec}^i(x_{RGB}), \quad i = 1 \quad (1)$$

Then we fused the multi-scale shallow features, the backbone output from HSV, and the YCbCr space  $f_{HSV}^{si}$ ,  $f_{YCbCr}^{si}$  with those from the RGB space in the corresponding feature scales by the element-wise self-adaptive weighted addition method as in Equation (2) to enrich and complement the RGB space shallow features with other two spaces as the fused RGB features  $f_{RGB}^{si}$ :

$$f_{RGB}^{si} = W_{RGB}^{ec} \times f_{RGB}^{si} + W_{HSV}^{ec} \times f_{HSV}^{si} + W_{YCbCr}^{ec} \times f_{YCbCr}^{si}, \quad i = 1, 2, 3, 4, \quad (2)$$

where  $W_{RGB}^{ec}$ ,  $W_{HSV}^{ec}$ , and  $W_{YCbCr}^{ec}$  represented the self-adaptive encoding fusion weights for the RGB, HSV, and YCbCr spaces, respectively.

The fused RGB features  $f_{RGB}^{si}$  were input to the subsequent part of the encoder to obtain the non-fused shallow RGB features at the next scale followed by the next similar encoding fusion.

The deep RGB features  $f_{RGB}^{dp}$  were non-fused at the last scale  $i_{max} = 5$  while the shallow features from the HSV and YCbCr space remained non-fused during the entire encoding process. As a result, the model encoder output two types of non-fused deep features from the HSV and YCbCr space and a type of deep feature RGB space containing information of three spaces, shown in Figure 3.

### 2.2.2. Attention Module

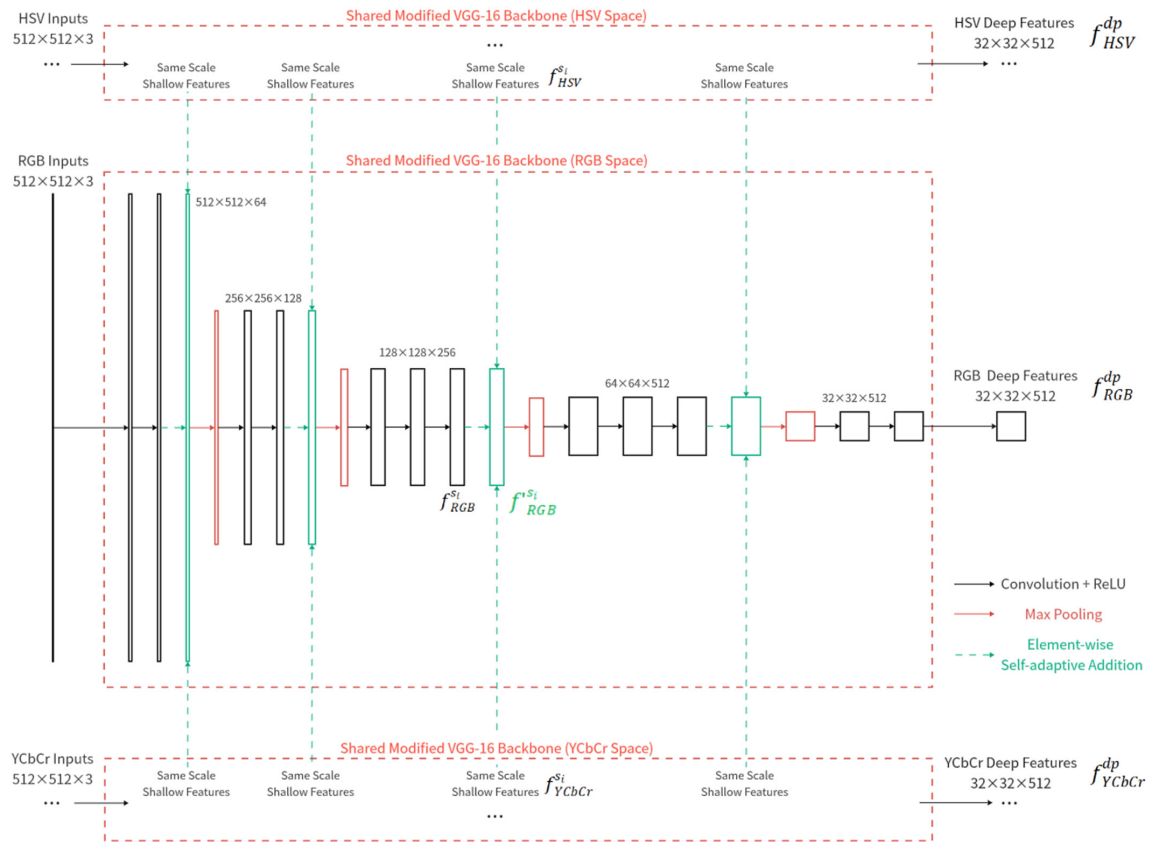
Considering both the background in forest fire images as being generally complicated with early smoke mostly occupying a small fraction in an image in the smoke dataset, and the inputs of the proposed model containing various multi-channel representations, we separately fed deep features from three spaces extracted by the encoder into the attention module CBAM [54] (Figure 4) to augment the spatial and channel information of the smoke object in images.

### 2.2.3. Feature Fusion Module

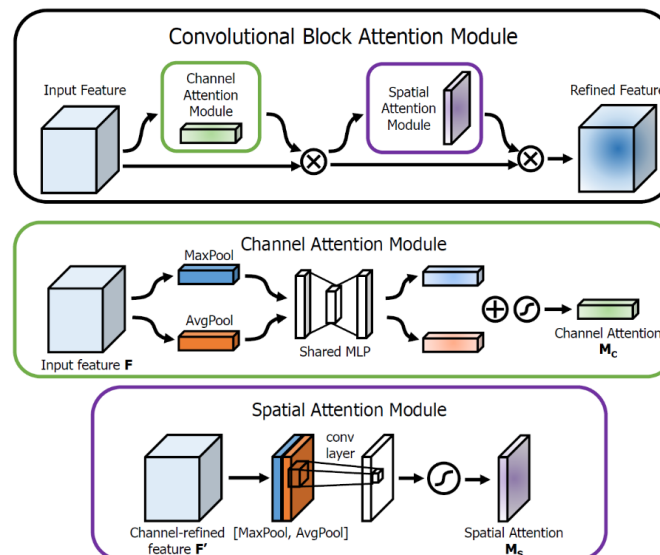
The feature fusion module in this paper contained three parts of fusion operation, namely features encoding fusion, deep feature fusion and shallow feature fusion. Features encoding the fusion part overlapped with the corresponding part in the encoder module, which was discussed in Section 2.2.1. For the deep feature fusion, the features of multi spaces with the attention augmented were fused by the element-wise self-adaptive weighted addition method similar to the encoder fusion part to obtain the deep fused features  $f_{fused}^{dp}$  as shown in Equation (3):

$$f_{fused}^{dp} = W_{RGB}^{dp} \times f_{RGB}^{dp} + W_{HSV}^{dp} \times f_{HSV}^{dp} + W_{YCbCr}^{dp} \times f_{YCbCr}^{dp} \quad (3)$$

where  $W_{RGB}^{dp}$ ,  $W_{HSV}^{dp}$ ,  $W_{YCbCr}^{dp}$  are the self-adaptive fusion weights for fusing the deep features from the corresponding space.



**Figure 3.** The SUFN encoder.



**Figure 4.** The CBAM attention mechanism as the attention module in the model [54].

Similarly, for the shallow feature fusion, the multi scale shallow features of multi spaces  $f_{RGB}^{s_i}$ ,  $f_{HSV}^{s_i}$ ,  $f_{YCbCr}^{s_i}$  from the multiple color space inputs were fused in the corresponding

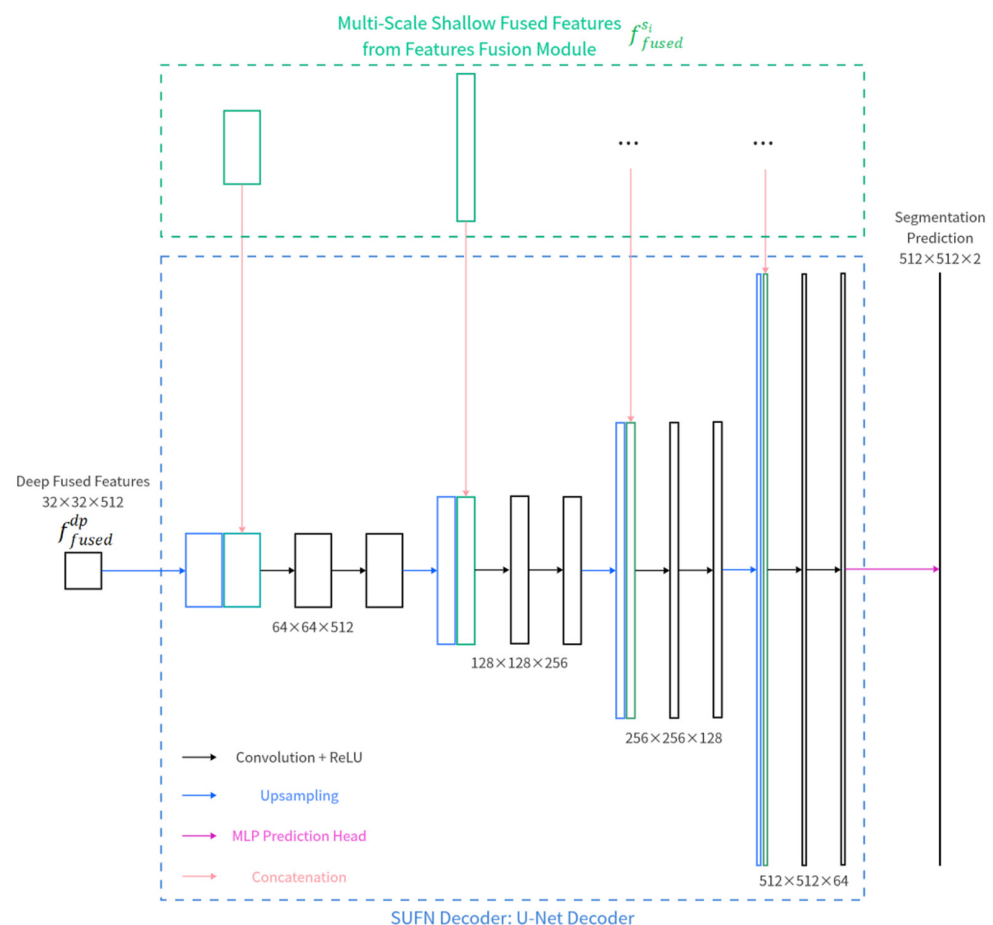
feature scale using the similar fusion strategy to obtain the multi-scale shallow fused features  $f_{fused}^{s_i}$  for later fusions as shown in Equation (4):

$$f_{fused}^{s_i} = W_{RGB}^{dc} \times f_{RGB}^{s_i} + W_{HSV}^{dc} \times f_{HSV}^{s_i} + W_{YCbCr}^{dc} \times f_{YCbCr}^{s_i}, i = 1, 2, 3, 4, \quad (4)$$

where  $W_{RGB}^{dc}$ ,  $W_{HSV}^{dc}$ ,  $W_{YCbCr}^{dc}$  are the self-adaptive fusion weights of shallow feature fusion for concatenation to the decoder.

#### 2.2.4. Decoder Module

In the decoder module, according to the characteristics of the original U-Net architecture, the extracted deep fused features  $f_{fused}^{dp}$  by the encoder, attention, and feature fusion module were up-sampled and restored by the decoder. And the shallow fused features in multi scales obtained by the same encoder  $f_{fused}^{s_i}$  were fused with the shallow features decoded in the corresponding scales, respectively, by concatenation to retain more spatial and positional information, named as multi-scale skip concatenation. This was similar to the skip connection known in the original U-Net architecture, that completed the entire “U” shape encoding and decoding. The decoder of the proposed model was the original U-Net decoder, shown in Figure 5. Finally, the fused map from the decoder was fed into the pixel classifier to classify the smoke and non-smoke pixel to achieve the smoke segmentation prediction.



**Figure 5.** The SUFN decoder.

The pseudocode of the proposed method is shown in the below Algorithm 1.



**Algorithm 1** The Pseudo-Code of the Proposed SUFN Method

---

**Input:** original RGB images  $x_{RGB}$ , smoke GT images  $\hat{y}$   
**Output:** The smoke segmentation prediction  $y$   
**Initialized:** the pretrained VGG16 backbone on VOC-2007 dataset  $\theta_{ec}$ , and the pretrained U-Net decoder  $\theta_{dc}$

**for**  $i$  in  $1, 2, \dots, N_{epoch}$  **do**  
    **for**  $j$  in  $1, 2, \dots, N_{batch}$  **do**  
         $x_{HSV}, x_{YCbCr} \leftarrow \text{ColorTransform}_{RGB \rightarrow HSV}(x_{RGB}), \text{ColorTransform}_{RGB \rightarrow YCbCr}(x_{RGB})$   
        **for**  $k$  in  $1, 2, 3, 4$   $Block_{ec}$  **do**  
            **if**  $k = 1$  **do**  
                 $f_{RGB}^k, f_{HSV}^k, f_{YCbCr}^k \leftarrow \theta_{ec}^k(x_{RGB}), \theta_{ec}^k(x_{HSV}), \theta_{ec}^k(x_{YCbCr})$   
            **else do**  
                 $f_{RGB}^k, f_{HSV}^k, f_{YCbCr}^k \leftarrow \theta_{ec}^k(\text{Maxpool}(f_{RGB}^{k-1})), \theta_{ec}^k(\text{Maxpool}(f_{HSV}^{k-1})),$   
                 $\theta_{ec}^k(\text{Maxpool}(f_{YCbCr}^{k-1}))$   
            **end if**  
             $f_{RGB}^k \leftarrow \text{Norm}(W_{RGB}^{ec} f_{RGB}^k + W_{HSV}^{ec} f_{HSV}^k + (1 - W_{RGB}^{ec} - W_{HSV}^{ec}) f_{YCbCr}^k)$   
             $f_{fused}^k \leftarrow \text{Norm}(W_{RGB}^{dc} f_{RGB}^k + W_{HSV}^{dc} f_{HSV}^k + (1 - W_{RGB}^{dc} - W_{HSV}^{dc}) f_{YCbCr}^k)$   
            **end for**  
             $f_{RGB}^{dp_5}, f_{HSV}^{dp_5}, f_{YCbCr}^{dp_5} \leftarrow \theta_{ec}^5(\text{Maxpool}(f_{RGB}^{s_4})), \theta_{ec}^5(\text{Maxpool}(f_{HSV}^{s_4})), \theta_{ec}^5(\text{Maxpool}(f_{YCbCr}^{s_4}))$   
             $f_{RGB}^{dp_5}, f_{HSV}^{dp_5}, f_{YCbCr}^{dp_5} \leftarrow \text{CBAM}(f_{RGB}^{dp_5}), \text{CBAM}(f_{HSV}^{dp_5}), \text{CBAM}(f_{YCbCr}^{dp_5})$   
             $f_{fused}^{dp_5} \leftarrow \text{Norm}(W_{RGB}^{dp} f_{RGB}^{dp_5} + W_{HSV}^{dp} f_{HSV}^{dp_5} + (1 - W_{RGB}^{dp} - W_{HSV}^{dp}) f_{YCbCr}^{dp_5})$   
             $f_{dc}^4 \leftarrow \theta_{dc}^4(\text{Cat}(\text{Upsample}(f_{fused}^{dp_5}), f_{fused}^{s_4}))$   
            **for**  $m$  in  $3, 2, 1$  **do**  
                 $f_{dc}^m \leftarrow \theta_{dc}^m(\text{Cat}(\text{Upsample}(f_{dc}^{m+1}), f_{fused}^m))$   
            **end for**  
             $y \leftarrow \text{FC}(f_{dc}^1)$   
             $L \leftarrow \text{CE}(y, \hat{y})$   
            Using gradient descent with Adam for loss backward  
            Update  $\theta_{ec}, \theta_{dc}, W_{RGB}^{ec}, W_{HSV}^{ec}, W_{RGB}^{dp}, W_{HSV}^{dp}, W_{RGB}^{dc}$  and  $W_{HSV}^{dc}$   
        **end for**  
    **end for**

---

**3. Results and Analyses****3.1. Evaluation Protocol**

The mean intersection-over-union (mIoU) metric, as a commonly used evaluation protocol in semantic segmentation tasks, was computed to evaluate the forest fire smoke segmentation performance by the models experimented in this study. The mIoU is defined as follows:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^N \text{IoU}_i, \quad (5)$$

where  $N$  is the number of categories and  $\text{IoU}_i$  score of categories  $i$ . And  $N = 1$  in our study since only the IoU score of smoke was focused, which is defined as follows:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (6)$$

where TP, FP, FN are, respectively, the number of true positives, false positives, false negatives of smoke in the segmentation results.

The false positive rate FPR and false negative rate FNR were also applied for measuring the false segmentation of and miss segmentation of smoke, respectively, which were defined as follows:

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}, \quad (7)$$

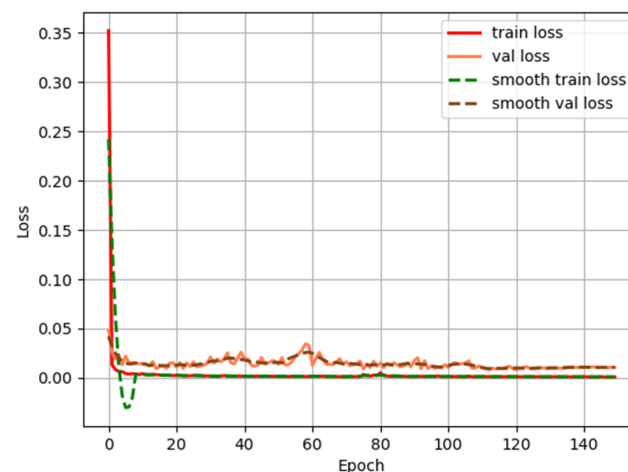
$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}. \quad (8)$$

### 3.2. Experimental and Training Settings

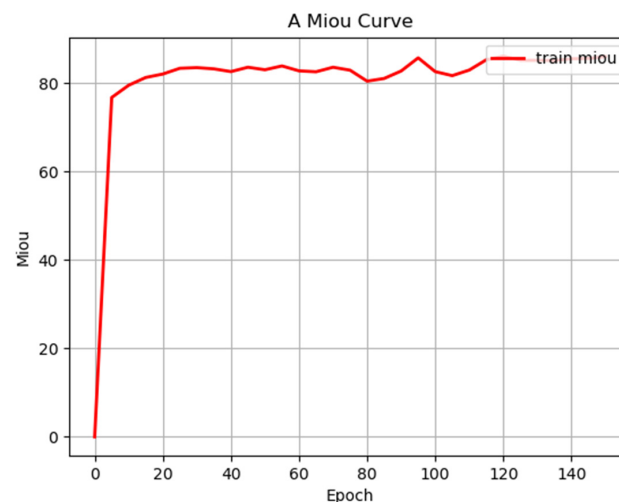
The hardware configuration in the study included a GPU of GeForce GTX 1080ti and a CPU of E5-2620. The experimental environment was Ubuntu 16.04 with PyTorch deep learning framework based on Python 3.6.

The input size of images to the model was  $512 \times 512$ . The network was trained end to end with the cross-entropy loss function. The backbone network of the modified VGG-16 was pretrained on the VOC dataset. The initial learning rate and momentum for training were, respectively, 0.0001 and 0.9 with Adam optimization. The batch size was 2 for 150 training epochs. In the feature fusion module, the richer and more prominent image information of the smoke object was in the RGB color space based on observation and our primary experiment on the smoke images. We set empirically larger initial trainable fusing weights of RGB space to 0.6 during the element-wise self-adaptive weighted addition fusion process, and two series of fusing weights of the other two spaces equal to 0.2. If only two spaces were input with the RGB space involved, the fusion weights were set to 0.7 and 0.3 for the other. Otherwise, the two series of weights were equally set to 0.5. These three series of fusion weights were automatically optimized during the training process so that it could obtain the most superior fused features and final segmentation prediction by obtaining the most appropriate self-adaptive weights.

The progress of loss decline and IoU score boosting in training are shown in Figures 6 and 7. The training ends up with a good trend and number as well as acceptable stability.



**Figure 6.** The loss curve of the proposed model during training process.



**Figure 7.** The IoU score curve of the proposed model during training process.

### 3.3. Results

In this study, we aimed to evaluate the effectiveness of fusing features from different color spaces for smoke segmentation. We conducted experiments to compare the performance of our proposed model with other segmentation models when incorporating features from various color spaces. The results are shown in Table 2.

**Table 2.** Testing results (in IoU) obtained from different spaces. The values in bold represent the optimal numerical result for the proposed SUFN in every multiple spaces combination.

Models	Spaces						
	RGB	HSV	YCbCr	RGB HSV	RGB YCbCr	HSV YCbCr	RGB HSV YCbCr
DeepLabv3+ Based Basic Fusion Model	68.03	66.89	65.14	70.86	67.99	68.41	73.44
PSPNet Based Basic Fusion Model	74.60	71.91	70.32	74.36	72.52	73.78	77.53
U-Net Based Basic Fusion Model	83.12	79.33	78.46	82.61	78.08	82.19	82.14
U-Net Based Fusion Model w. CBAM	83.54	79.32	79.47	82.29	79.86	83.05	84.05
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights	—	—	—	84.44	81.44	85.88	85.04
SUFN w/o CBAM nor Self-Adaptive Weights	83.12	79.33	78.46	84.89	82.83	—	84.57
SUFN w/o Self-Adaptive Weights	83.54	79.32	79.47	85.21	83.73	—	84.63
SUFN	—	—	—	<b>85.44</b>	<b>84.86</b>	—	<b>86.14</b>

In Table 2, the DeepLabv3+/PSPNet Based Basic Fusion Model is based on the original semantic segmentation models DeepLabv3+ [55]/PSPNet [56] with the least modifications according to the method that we reconstructed for the original U-Net as two segmentation models with the function of multi-color spaces feature extraction and fusion. It is worth noting that SUFN is unavailable for the HSV and YCbCr space inputs due to the absence of principal RGB space, and the proposed self-adaptive fusion weights for the U-Net Based Basic Fusion Model and SUFN are disabled for the single color space input since the self-adaptive weights can be defined only when the feature fusion of multiple spaces is implemented.

The results indicate that using the RGB color space as the principal space yields superior segmentation performance compared to other single-color spaces. The reason for this is the use of RGB images in the adopted VGG pre-training model, which makes the model more adaptable to RGB images. For the fusion of features from multiple color spaces, we observed that the fusion of RGB and HSV spaces resulted in higher IoU scores than any single space model. This improvement was further enhanced when incorporating a self-adaptive weighted fusion strategy, indicating the efficacy of multi-color space feature fusion. Similar improvements were observed in the proposed SUFN model, where the fusion of RGB and YCbCr spaces outperformed any single space model. It supports the conclusion that multi-color space feature fusion is efficient for smoke segmentation. Moreover, when fusing features from RGB, HSV, and YCbCr spaces together, the results were overall enhanced compared to any single space or double space fusion combination, except for the triple space fusion result of the U-Net Based Basic Fusion Model. Specifically, the RGB and HSV and YCbCr fusion by the SUFN model achieved an IoU score of 86.14, outperforming other fusion combinations. It demonstrates the superiority of the proposed multi-color space feature fusion and segmentation approach.

In terms of network architecture, applying the attention mechanism CBAM before deep feature fusion generally improved the segmentation performance. In addition, employing a self-adaptive weighted fusion strategy instead of constant empirical fusion weights further harmonized the features from different spaces and led to superior segmentation results. When comparing our SUFN model with the DeepLabv3+ Basic Fusion Model and the PSPNet Based Basic Fusion Model, the SUFN model consistently exhibited superior

segmentation performance, regardless of the color space used for feature extraction, fusion, and segmentation.

In addition to the displayed results in the IoU score of smoke, which indicated the superiority and the upper limit the model could achieve, the result to measure the degree of detection failures, including smoke false segmentation and miss segmentation of the obtained optimal models by giving the FPR and FNR percentage at the pixel level, is shown in Table 3 below.

**Table 3.** Testing results for measuring detection failures by the obtained optimal models with the corresponding color space combination shown.

Models	FPR (%)	FNR (%)
DeepLabv3+ Based Basic Fusion Model (RGB & HSV & YCbCr)	14.85	15.33
PSPNet Based Basic Fusion Model (RGB & HSV & YCbCr)	12.27	11.12
U-Net Baseline Model	8.10	10.06
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights (HSV & YCbCr)	7.28	7.65
SUFN (RGB & HSV & YCbCr)	4.93	6.13

The proposed SUFN significantly reduced the false segmentation and miss segmentation rate by 3.17 and 3.93, respectively, compared to the original U-Net baseline in Table 3. And the optimal modified DeepLabv3+ Based Basic Fusion Model tended to have a relatively high miss detection rate of smoke while the optimal modified PSPNet Based Basic Fusion Model tended to have a relatively high false detection rate of smoke on the contrary. The FPR and FNR of two optimal comparative models were, respectively, 9.92 and 9.20 weaker for the former, and 7.34 and 4.99 weaker for the latter, both compared to the proposed SUFN. The results demonstrated leveraging the diverse and distinct smoke information from multiple color spaces could enhance the smoke characteristics and suppress the interferences of the smoke-like object or the complex background, which is advantageous for the robust detection in the challenging real-world forest fire monitoring scenarios.

Overall, the results validate the effectiveness of our proposed multi-color space feature fusion.

### 3.4. Self-Adaptive Weighted Fusion Strategies and Coefficients

Feature fusion in our model is a critical component for accurate segmentation results. We considered different fusion strategies based on the location where the fusion occurred. In the U-Net Based Fusion Model, feature fusion was only performed in the deep feature fusion and the multi-scale shallow feature fusion. We designed three fusion schemes: scheme A, which used a single set of self-adaptive weights for all the fusion operations; scheme B, which used a double set of self-adaptive weights for the deep feature fusion and the entire multi-scale shallow feature fusion separately; and scheme C, which used a quintuple set of self-adaptive weights for the deep feature fusion and the quadruple shallow feature fusion. Similar fusion schemes were also applied to the proposed SUFN model, denoted as schemes D, E, and F. The results on the testing set are shown in Table 4.

**Table 4.** Testing results (in IoU) with various fusion strategies of self-adaptive weights. The values in bold represent the optimal numerical result for the designed models and fusion strategies.

Models & Fusion Hyper-Parameters Type	Spaces						
	RGB	HSV	YCbCr	RGB HSV	RGB YCbCr	HSV YCbCr	RGB HSV YCbCr
U-Net Based Basic Fusion Model	83.12	79.33	78.46	82.61	78.08	82.19	82.14
U-Net Based Fusion Model w. CBAM	83.54	79.32	79.47	82.29	79.86	83.05	84.05
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights (A)	—	—	—	83.47	80.93	84.16	82.74
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights (B)	—	—	—	<b>84.44</b>	<b>81.44</b>	<b>85.88</b>	<b>85.04</b>
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights (C)	—	—	—	81.79	78.90	84.56	83.33
SUFN w/o CBAM nor Self-Adaptive Weights	83.12	79.33	78.46	84.89	82.83	—	84.57
SUFN w/o Self-Adaptive Weights	83.54	79.32	79.47	85.21	83.73	—	84.63
SUFN (D)	—	—	—	84.81	84.00	—	85.22
SUFN (E)	—	—	—	<b>85.44</b>	<b>84.86</b>	—	<b>86.14</b>
SUFN (F)	—	—	—	85.30	83.29	—	83.32

To determine the best fusion strategy, we compared the results of these fusion schemes on the testing set. As presented in Table 4, scheme B in the U-Net Based Basic Fusion Model and scheme E in the SUFN model achieved optimal segmentation results, outperforming the models with constant empirical fusion weights.

The sub-optimal strategies were scheme A for the U-Net Based Basic Fusion Model and scheme D for the SUFN model. Schemes A and D, while outperforming the constant empirical weights in most cases, do not perform as well as Schemes B and E. It may be because they control only one set of fusion weights throughout the fusion operation, which restricts the optimization of fusion weights for each fusion stage.

On the other hand, schemes C and F, which used multiple sets of fusion weights, yielded the worst segmentation results. This could be due to the excessive independence between the sets of fusion weights, leading to increased model complexity and making it challenging to optimize the fusion weights for each phase and position during training.

Based on these findings, we selected scheme B for the U-Net Based Basic Fusion Model and Scheme E for the SUFN model as the ideal self-adaptive weighted fusion scheme. These schemes consistently produced superior results, as shown in Table 5, when the self-adaptive weights were activated.

**Table 5.** The self-adaptive weighted fusion parameters derived from the optimal fusion strategies.

Models	Features Fusion Phases								
	Features Encoding Fusion			Deep Features Fusion			Shallow Features Fusion		
	RGB	HSV	YCbCr	RGB	HSV	YCbCr	RGB	HSV	YCbCr
U-Net Based Fusion Model w. CBAM & Self-Adaptive Weights (B)	—	—	—	—	0.491	0.509	—	0.437	0.563
SUFN (E)	0.592	0.170	0.238	0.584	0.197	0.219	0.543	0.142	0.315

### 3.5. Results for Different Smoke Type and the Visualization Performance

To verify the generalization of the model to smoke segmentation, the segmentation performance of the derived optimal model for the various types of smoke image in the NCS, NSI, and IS was also tested, shown in Table 6.



**Table 6.** Results (in IoU) for testing different smoke types.

Model	Testing Smoke Type		
	NCS	NSI	IS
SUFN(E) in RGB & HSV & YCbCr Spaces	90.31	85.79	80.47

In Table 6, the NCS (normal-scale and clearly seen smoke without obvious interferences) type has the highest IoU score of 90.31, which is reasonable because this type of smoke has well-defined contour and relatively few surrounding disturbances, which makes it easier to recognize and accurately classify. The NSI (normal-scale and interference smoke) type achieves an IoU score of 85.79, representing the average performance of the model. This type is more common in real smoke detection scenarios and is often accompanied by various external interferences such as clouds, haze, fog, and glare, making the smoke contour less distinct. However, the model still provides accurate smoke segmentation masks to indicate the presence of smoke in the image, albeit with a slightly lower accuracy. The IS (inconspicuous smoke) type achieves an IoU score of 80.47, which is comparatively weaker than the other two types. This is due to the difficulty in identifying and delineating the small and inconspicuous smoke regions in the image, which poses a challenge for the model and human observers. In addition, the limited number and proportion of IS images in the dataset may contribute to the lower performance of this type. It may be necessary to increase the number and proportion of IS-type training samples in the dataset to improve the model's performance on the IS type. Despite these challenges, the model still demonstrated the ability to segment IS-type smoke. Furthermore, visual detections of challenging forest fire smoke images from the complementary set of FlgLib, which were not included in our dataset, are presented in Figure 8 to showcase the effectiveness of our model in smoke detection and segmentation.

As seen in Figure 8, the proposed SUFN could segment the smoke in challenging background scenarios in most cases, such as strong background glare interference in the NSI case (a), the IS cases (c) and (d), cloud disturbance in the IS case (c), background blurriness caused by saturation near the horizon in the IS cases (b) and (d). Notably, the three IS cases (b–d) are, respectively, the very early stage of fire with little smoke spreading outwards, the small visible smoke with surrounding obstruction, and the relatively faint smoke due to the long distance to the camera, which leads to the difficulty of segmentation. However, the smoke detection visualizations from our model are satisfactory when dealing with the above challenges. It could be seen that the IS case (c), that is partly blocked by the hill with a small proportion uncovered in the image, is nearly perfectly segmented and the main body of smoke in the other three cases in the figure is almost detected accurately with little edge of smoke acceptably under-segmented for (a) and (b) or over-segmented for (d). In comparison to our method, the optimal comparative DeepLabv3+ Based Basic Fusion Model tends to have higher false negative due to the missing detection of smoke contour under-segmentation for IS cases (b) and (d), and the failing detection for NSI case (a) and IS case (c). The optimal comparative PSPNet based basic fusion model tends to be the opposite of the higher false positive except for the same failing detection for the IS case (c), which is over-segmented for the NSI case (a), the IS case (b) and (d), with low accuracy especially for the case (d). In all, the detection visualization demonstrates that the proposed SUFN have strong detection ability to face the challenging conditions brought about by ambiguous smoke itself or its complex surroundings in completely new scenarios.



**Figure 8.** The visualization of the forest fire smoke detection. (a) NSI type. (b–d) IS type.

#### 4. Discussion

We propose the SUFN model, optimized for forest fire smoke detection. We introduced feature fusion to the original segmentation baseline model to incorporate multiple inputs from different color spaces. We added features encoding fusion blocks for the principal RGB space and employed the CBAM attention mechanism for feature augmentation. Moreover, we improved the fusion method by applying self-adaptive weighted addition at local fusion positions. Experimental results demonstrate the effectiveness of our approach, with an IoU of 86.14 achieved for smoke detection.

Regarding the single dataset used for testing and training in this paper, a brilliant smoke detection model should have good detection results when testing on diverse data in different datasets to prove the generalizability. This still needs our further verification on additional types of data to be collected. However, the quality of training data is critical for the model to acquire favorable performance. Compared to the dataset used in many previous studies which was not from actual fire monitoring scenarios or even not from actual forest fire occurrence with low image resolution (e.g., experimental simulations of fire [32], synthetic forest fire smoke images [14,32,33]), the dataset we used for model training was from an extremely rare publicly available database so far containing authentic forest fire smoke from the actual fire ignition with high data quality of high image resolution, high image content clarity, rich and multifarious smoke pattern and background. This made the model achieve superior detection results for testing and visualization on various types of authentic smoke images with the help of our fine smoke annotations.

Nevertheless, some objective limitations certainly exist in this research:

- (1) The volume of data samples used for training and testing are relatively small due to the lack of smoke annotation, especially for the IS smoke samples, which would lead to a decrease in the model accuracy, robustness, and reliability in real application.

- (2) The CBAM attention mechanism and the way it was introduced into the proposed model may not be the optimal one. More attention mechanisms are needed to be taken into consideration, and how to incorporate them more properly into the baseline model requires further attempts.
- (3) Only three color spaces are involved for multiple color spaces feature fusion, which limits the research refinement. More distinct color spaces and combinations could be considered and the appropriate number of multiple inputs for multiple color spaces needs further study.

Subsequent studies will focus on the above problems by adding more smoke samples of various categories defined in the paper with fine annotations to further stabilize the detection results. The methodology of the improvement to the model structure and the refinement to the multiple color space study will also be developed in the future. In addition, the idea of leveraging multiple color space information combined with the extracted feature fusion could obtain richer and more complementary smoke features. However, how do the features from multiple color spaces emphasize and distinguish smoke features from other smoke-like phenomena being suppressed (e.g., fog, clouds, haze) share similar visual characteristics and how to understand, control, and magnify this advantage further to reduce the false positives for better forest fire smoke detection is worth future exploration. In addition to multiple color space information of the image, more static information such as texture and shape of smoke, or dynamic information such as the direction of smoke motion, contour change in temporal consecutive frames, or videos could also be included in the future for richer feature fusion study.

## 5. Conclusions

Our proposed method introduces a semantic segmentation approach for fire smoke detection using self-adaptive weighted feature fusion from multiple color spaces. The model (i.e., SUFN) is a U-Net extension and integrates self-adaptive weighted fusion strategies for features extracted from RGB, HSV, and YCbCr color spaces. In addition, an attention mechanism is integrated to enhance deep features from these spaces before fusion. Through extensive experimentation, we demonstrated the effectiveness of our method in smoke segmentation, achieving an optimal IoU score of 86.14. The performance of our proposed model surpasses that of other models modified for weighted feature fusion. Fusing multiple color spaces improves segmentation results, regardless of the specific model architecture. Notably, the optimal color space combinations for fusion are RGB, HSV, and YCbCr. Our experiments were conducted with a high-quality and diverse forest fire smoke dataset called FFigLib, which contains challenging smoke forms in various backgrounds. The test results and visualization performance in unknown complex scenes further demonstrated the superiority of our proposed method.

**Author Contributions:** Conceptualization, Z.H., Y.T. and C.Z.; methodology, Z.H. and C.Z.; software, Z.H.; validation, Z.H.; formal analysis, Z.H.; investigation, Z.H.; resources, C.Z.; data curation, Z.H.; writing—original draft preparation, Z.H.; writing—review and editing, Y.T. and C.Z.; visualization, Z.H.; supervision, C.Z. and Y.T.; project administration, C.Z. and Z.H.; funding acquisition, C.Z. and F.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Key R&D Program of China under Grant 2023YFC3006805, and in part by the National Natural Science Foundation of China under Grant 31971668.

**Data Availability Statement:** Data available on request due to restrictions of privacy.

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

## References

1. Cao, Y.; Tang, Q.; Wu, X.; Lu, X. EFFNet: Enhanced Feature Foreground Network for Video Smoke Source Prediction and Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 1820–1833. [\[CrossRef\]](#)
2. Zhu, G.; Chen, Z.; Liu, C.; Rong, X.; He, W. 3D video semantic segmentation for wildfire smoke. *Mach. Vis. Appl.* **2020**, *31*, 50. [\[CrossRef\]](#)
3. Lu, N. Dark convolutional neural network for forest smoke detection and localization based on single image. *Soft Comput.* **2022**, *26*, 8647–8659. [\[CrossRef\]](#)
4. Dewangan, A.; Pande, Y.; Braun, H.-W.; Vernon, F.; Perez, I.; Altintas, I.; Cottrell, G.W.; Nguyen, M.H. FlgLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection. *Remote Sens.* **2022**, *14*, 1007. [\[CrossRef\]](#)
5. Muhammad, K.; Ahmad, J.; Lv, Z.; Bellavista, P.; Yang, P.; Baik, S.W. Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications. *IEEE Trans. Syst. Man, Cybern. Syst.* **2019**, *49*, 1419–1434. [\[CrossRef\]](#)
6. Singh, P.K.; Sharma, A. An insight to forest fire detection techniques using wireless sensor networks. In Proceedings of the 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 21–23 September 2017; pp. 647–653. [\[CrossRef\]](#)
7. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [\[CrossRef\]](#)
8. Jin, C.; Wang, T.; Alhusaini, N.; Zhao, S.; Liu, H.; Xu, K.; Zhang, J. Video Fire Detection Methods Based on Deep Learning: Datasets, Methods, and Future Directions. *Fire* **2023**, *6*, 315. [\[CrossRef\]](#)
9. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.W.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [\[CrossRef\]](#)
10. Abid, F. A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. *Fire Technol.* **2021**, *57*, 559–590. [\[CrossRef\]](#)
11. Szpakowski, D.M.; Jensen, J.L.R. A Review of the Applications of Remote Sensing in Fire Ecology. *Remote Sens.* **2019**, *11*, 2638. [\[CrossRef\]](#)
12. Khan, F.; Xu, Z.; Sun, J.; Khan, F.M.; Ahmed, A.; Zhao, Y. Recent Advances in Sensors for Fire Detection. *Sensors* **2022**, *22*, 3310. [\[CrossRef\]](#)
13. Bouguettaya, A.; Zarzour, H.; Taberkit, A.M.; Kechida, A. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Process.* **2022**, *190*, 108309. [\[CrossRef\]](#)
14. Mao, J.; Zheng, C.; Yin, J.; Tian, Y.; Cui, W. Wildfire Smoke Classification Based on Synthetic Images and Pixel- and Feature-Level Domain Adaptation. *Sensors* **2021**, *21*, 7785. [\[CrossRef\]](#)
15. Memane, S.E.; Kulkarni, V.S. A Review on Flame and Smoke Detection Techniques in Videos. *Int. J. Adv. Res. Electr. Electron. Instrum. Energy* **2015**, *4*, 855–859.
16. Gaur, A.; Singh, A.; Kumar, A.; Kumar, A.; Kapoor, K. Video Flame and Smoke Based Fire Detection Algorithms: A Literature Review. *Fire Technol.* **2020**, *56*, 1943–1980. [\[CrossRef\]](#)
17. Garg, S.; Verma, A.A. Review Survey on Smoke Detection. *Imp. J. Interdiscip. Res.* **2016**, *2*, 935–939.
18. Chaturvedi, S.; Khanna, P.; Ojha, A. A survey on vision-based outdoor smoke detection techniques for environmental safety. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 158–187. [\[CrossRef\]](#)
19. Altun, M.; Celenk, M. Smoke Detection in Video Surveillance Using Optical Flow and Green's Theorem. In Proceedings of the IPCV 2013: International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, NV, USA, 22–25 July 2013.
20. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [\[CrossRef\]](#)
21. Govil, K.; Welch, M.L.; Ball, J.T.; Pennypacker, C.R. Preliminary Results from a Wildfire Detection System Using Deep Learning on Remote Camera Images. *Remote Sens.* **2020**, *12*, 166. [\[CrossRef\]](#)
22. Zhang, Z.; Jin, Q.; Wang, L.; Liu, Z. Video-based Fire Smoke Detection Using Temporal-spatial Saliency Features. *Procedia Comput. Sci.* **2022**, *198*, 493–498. [\[CrossRef\]](#)
23. Zhang, H.; Yang, S.; Wang, H.; Li, J.; Liu, H. Unified Smoke and Fire Detection in An Evolutionary Framework with Self-Supervised Progressive Data Augment. In Proceedings of the 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Chengdu, China, 19–21 August 2022; pp. 574–581.
24. Fernandes, A.M.; Utkin, A.B.; Chaves, P. Automatic Early Detection of Wildfire Smoke with Visible Light Cameras Using Deep Learning and Visual Explanation. *IEEE Access* **2022**, *10*, 12814–12828. [\[CrossRef\]](#)
25. Bhamra, J.K.; Ramaprasad, S.A.; Baldota, S.; Luna, S.; Zen, E.; Ramachandra, R.; Kim, H.; Schmidt, C.; Arends, C.; Block, J.; et al. Multimodal Wildland Fire Smoke Detection. *Remote Sens.* **2023**, *15*, 2790. [\[CrossRef\]](#)
26. Cheng, G.; Chen, X.; Gong, J. Deep Convolutional Network with Pixel-Aware Attention for Smoke Recognition. *Fire Technol.* **2022**, *58*, 1839–1862. [\[CrossRef\]](#)
27. Hasan, S.B.; Rahman, S.; Khaliluzzaman, M.; Ahmed, S. Smoke Detection from Different Environmental Conditions Using Faster R-CNN Approach Based on Deep Neural Network. In *Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, 15–16 February 2020, Proceedings 2*; Springer International Publishing: Cham, Switzerland, 2020; pp. 705–717.



28. Guede-Fernández, F.; Martins, L.; de Almeida, R.V.; Gamboa, H.; Vieira, P. A Deep Learning Based Object Identification System for Forest Fire Detection. *Fire* **2021**, *4*, 75. [\[CrossRef\]](#)
29. Hu, Y.; Zhan, J.; Zhou, G.; Chen, A.; Cai, W.; Guo, K.; Hu, Y.; Li, L. Fast forest fire smoke detection using MVMNet. *Knowl. Based Syst.* **2022**, *241*, 108219. [\[CrossRef\]](#)
30. Li, J.; Zhou, G.; Chen, A.; Wang, Y.; Jiang, J.; Hu, Y.; Lu, C. Adaptive linear feature-reuse network for rapid forest fire smoke detection model. *Ecol. Inform.* **2022**, *68*, 101584. [\[CrossRef\]](#)
31. Choi, M.; Kim, C.; Oh, H. A video-based SlowFastMTB model for detection of small amounts of smoke from incipient forest fires. *J. Comput. Des. Eng.* **2022**, *9*, 793–804. [\[CrossRef\]](#)
32. Zheng, X.; Chen, F.; Lou, L.; Cheng, P.; Huang, Y. Real-Time Detection of Full-Scale Forest Fire Smoke Based on Deep Convolution Neural Network. *Remote Sens.* **2022**, *14*, 536. [\[CrossRef\]](#)
33. Chen, G.; Cheng, R.; Lin, X.; Jiao, W.; Bai, D.; Lin, H. LMDFS: A Lightweight Model for Detecting Forest Fire Smoke in UAV Images Based on YOLOv7. *Remote Sens.* **2023**, *15*, 3790. [\[CrossRef\]](#)
34. Yuan, F.; Zhang, L.; Xia, X.; Huang, Q.; Li, X. A Gated Recurrent Network with Dual Classification Assistance for Smoke Semantic Segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 4409–4422. [\[CrossRef\]](#)
35. Perrolas, G.; Niknejad, M.; Ribeiro, R.; Bernardino, A. Scalable Fire and Smoke Segmentation from Aerial Images Using Convolutional Neural Networks and Quad-Tree Search. *Sensors* **2022**, *22*, 1701. [\[CrossRef\]](#)
36. Khan, S.; Muhammad, K.; Hussain, T.; Del Ser, J.; Cuzzolin, F.; Bhattacharyya, S.; Akhtar, Z.; de Albuquerque, V.H.C. DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments. *Expert Syst. Appl.* **2021**, *182*, 115125. [\[CrossRef\]](#)
37. Ding, Z.; Zhao, Y.; Li, A.; Zheng, Z. Spatial–Temporal Attention Two-Stream Convolution Neural Network for Smoke Region Detection. *Fire* **2021**, *4*, 66. [\[CrossRef\]](#)
38. Muksimova, S.; Mardieva, S.; Cho, Y.-I. Deep Encoder–Decoder Network-Based Wildfire Segmentation Using Drone Images in Real-Time. *Remote Sens.* **2022**, *14*, 6302. [\[CrossRef\]](#)
39. Martins, L.; Guede-Fernández, F.; Valente de Almeida, R.; Gamboa, H.; Vieira, P. Real-Time Integration of Segmentation Techniques for Reduction of False Positive Rates in Fire Plume Detection Systems during Forest Fires. *Remote Sens.* **2022**, *14*, 2701. [\[CrossRef\]](#)
40. Wang, Z.; Zheng, C.; Yin, J.; Tian, Y.; Cui, W. A Semantic Segmentation Method for Early Forest Fire Smoke Based on Concentration Weighting. *Electronics* **2021**, *10*, 2675. [\[CrossRef\]](#)
41. Deng, J.; Bei, S.; Shaojing, S.; Zhen, Z. Feature Fusion Methods in Deep-Learning Generic Object Detection: A Survey. In Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 11 December 2020; pp. 431–437.
42. Tang, L.; Zhang, H.; Xu, H.; Ma, J. Deep Learning-Based Image Fusion: A Survey. *J. Image Graph.* **2023**, *28*, 3–36.
43. Liu, Y.; Zheng, C.; Liu, X.; Tian, Y.; Zhang, J.; Cui, W. Forest Fire Monitoring Method Based on UAV Visual and Infrared Image Fusion. *Remote Sens.* **2023**, *15*, 3173. [\[CrossRef\]](#)
44. Zhao, J.; Zhou, Y.; Shi, B.; Yang, J.; Zhang, D.; Yao, R. Multi-Stage Fusion and Multi-Source Attention Network for Multi-Modal Remote Sensing Image Segmentation. *ACM Trans. Intell. Syst. Technol.* **2021**, *12*, 1–20. [\[CrossRef\]](#)
45. Chen, X.; Hopkins, B.; Wang, H.; O’neill, L.; Afghah, F.; Razi, A.; Fulé, P.; Coen, J.; Rowell, E.; Watts, A. Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset. *IEEE Access* **2022**, *10*, 121301–121317. [\[CrossRef\]](#)
46. Daoud, Z.; Ben Hamida, A.; Ben Amar, C. FireClassNet: A deep convolutional neural network approach for PJF fire images classification. *Neural Comput. Appl.* **2023**, *35*, 19069–19085. [\[CrossRef\]](#)
47. Haridasan, S.; Rattani, A.; Demissie, Z.; Dutta, A. Multispectral Deep Learning Models for Wildfire Detection. In Proceedings of the International Workshop on Data-driven Resilience Research, Leipzig, Germany, 6 July 2022.
48. Xing, D.; Zhongming, Y.; Lin, W.; Jinlan, L. Smoke Image Segmentation Based on Color Model. *J. Innov. Sustain.* **2015**, *6*, 130–138. [\[CrossRef\]](#)
49. Prema, C.E.; Vinsley, S.S.; Suresh, S. Multi Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection. *Fire Technol.* **2016**, *52*, 1319–1342. [\[CrossRef\]](#)
50. Pundir, A.S.; Raman, B. Deep Belief Network for Smoke Detection. *Fire Technol.* **2017**, *53*, 1943–1960. [\[CrossRef\]](#)
51. CVAT. Available online: <http://www.cvat.ai> (accessed on 4 March 2024).
52. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351. [\[CrossRef\]](#)
53. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
54. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11211. [\[CrossRef\]](#)



55. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11211. [[CrossRef](#)]
56. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.

**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.