sensors

MDPI

Article

Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation

Rafik Ghali [1], Moulay A. Akhloufi [1,*] and Wided Souidene Mseddi [2]

[1] Perception, Robotics and Intelligent Machines Research Group (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada; rafik.ghali@ept.rnu.tn
[2] SERCOM Laboratory, Ecole Polytechnique de Tunisie, Université de Carthage, BP 743, La Marsa 2078, Tunisia; wided.souidene@ept.rnu.tn
* Correspondence: moulay.akhloufi@umoncton.ca

Abstract: Wildfires are a worldwide natural disaster causing important economic damages and loss of lives. Experts predict that wildfires will increase in the coming years mainly due to climate change. Early detection and prediction of fire spread can help reduce affected areas and improve firefighting. Numerous systems were developed to detect fire. Recently, Unmanned Aerial Vehicles were employed to tackle this problem due to their high flexibility, their low-cost, and their ability to cover wide areas during the day or night. However, they are still limited by challenging problems such as small fire size, background complexity, and image degradation. To deal with the aforementioned limitations, we adapted and optimized Deep Learning methods to detect wildfire at an early stage. A novel deep ensemble learning method, which combines EfficientNet-B5 and DenseNet-201 models, is proposed to identify and classify wildfire using aerial images. In addition, two vision transformers (TransUNet and TransFire) and a deep convolutional model (EfficientSeg) were employed to segment wildfire regions and determine the precise fire regions. The obtained results are promising and show the efficiency of using Deep Learning and vision transformers for wildfire classification and segmentation. The proposed model for wildfire classification obtained an accuracy of 85.12% and outperformed many state-of-the-art works. It proved its ability in classifying wildfire even small fire areas. The best semantic segmentation models achieved an F1-score of 99.9% for TransUNet architecture and 99.82% for TransFire architecture superior to recent published models. More specifically, we demonstrated the ability of these models to extract the finer details of wildfire using aerial images. They can further overcome current model limitations, such as background complexity and small wildfire areas.

Keywords: wildfire detection; fire classification; fire segmentation; vision transformers; UAV; aerial images

Citation: Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep Learning and Transformers Approaches for UAV Based Wildfire Detection and Segmentation. *Sensors* **2022**, *22*, 1977. https://doi.org/10.3390/s22051977

Academic Editor: Chiman Kwan

Received: 13 February 2022
Accepted: 1 March 2022
Published: 3 March 2022

1. Introduction

Forest fire accidents are one of the most dangerous risks due to their frightening loss statistics. The fires cause human, financial, and environmental losses, including the death of animals and the destruction of wood, houses, and million acres of land worldwide. In 2021, forest fires have occurred in several countries such as the European Union countries, the US (United States), central and southern Africa, the Arabian Gulf, and South and North America [1]. They affect 350 million to 450 million hectares every year [2]. In the western US alone, the frequency of wildfires and the total area burned increased by 400% and 600%, respectively, in the last decade [3]. In addition, approximately 8000 wildfires affected 2.5 million hectares each year in Canada [4].

Generally, wildfires are detected using various sensors such as gas, smoke, temperature, and flame detectors. Nevertheless, these detectors have a variety of limitations such as delayed response and small coverage areas [5]. Fortunately, the advancement of computer vision techniques has made it possible to detect fire using visual features

collected with cameras. However, traditional fire detection tools are being replaced by vision-based models that have many advantages such as accuracy, large coverage areas, small probability of errors, and most importantly the ability to work with existing camera surveillance systems. Through the years, researchers have proposed many innovative techniques based on computer vision in order to build accurate fire detection systems [6–9].

In recent years, Unmanned Aerial Vehicles (UAV) or drone systems were deployed in various tasks such as traffic monitoring [10], precision agriculture [11], disaster monitoring [12], smart cities [13], cover mapping [14], and object detection [15]. They are also very practical and well developed for wildfire fighting and detection. UAV-based systems help with precise fire management and provide real-time information to limit damage from fires thanks to their low cost and ability to cover large areas whether during the day or night for a long duration [16,17]. The integration of UAVs with visual and/or infrared sensors help in finding potential fires at daytime and nighttime [18]. Furthermore, fire detection and segmentation showed impressive progress thanks to the use of deep learning (DL) techniques. DL-based fire detection methods are used to detect the color of wildfire and its geometrical features such as angle, shape, height, and width. Their results are used as inputs to the fire propagation models. Thanks to the promising performances of DL approaches in wildfire classification and segmentation [19], researchers are increasingly investigating this family of methods. The existing methods use input images captured by traditional visual sensors to localize wildfire and to detect the precise shape of fire; they achieved high results [20–22]. However, it is not yet clear that these methods will perform well in detecting and segmenting forest fire using UAV images, especially in the presence of various challenges such as small object size, background complexity, and image degradation.

To address these problems, we present in this paper a novel deep ensemble learning method to detect and classify wildfire using aerial images. This method employs EfficientNet-B5 [23] and DenseNet-201 [24] models as a backbone for extracting forest fire features. In addition, we employed a deep model (EfficientSeg [25]) and two vision transformers (TransUNet [26] and TransFire) in segmenting wildfire pixels and detecting the precise shape of fire on aerial images. Then, the proposed wildfire classification method was compared to deep convolutional models (MobileNetV3-Large -Small [27], DenseNet-169 [24], EfficientNet-B1-5 [23], Xception [28,29], and InceptionV3 [29]), which showed excellent results for object classification. TransUNet, TransFire, and EfficientSeg are also evaluated with U-Net [28].

More specifically, three main contributions were reported in this paper. First, a novel DL method was proposed to detect and classify wildfire using aerial images in order to improve detection and segmentation of wildland fires. Second, vision transformers were adopted for UAV wildfire segmentation to segment fire pixels and identify the precise shape of the fire. Third, the efficiencies of CNN methods and vision transformers are demonstrated in extracting the finer details of fire using aerial images and overcoming the problems of background complexity and small fire areas.

The remainder of the paper is organized as follows: Section 2 presents recent DL approaches for UAV-based fire detection and segmentation. Section 3 describes the methods and materials used in this paper. In Section 4, experimental results and discussion are presented. Finally, Section 5 concludes the paper.

2. Related Works

DL approaches are employed for fire detection and segmentation using aerial images. They proved their ability to detect and segment wildfires [6,20]. They can be grouped into three categories: DL approaches for UAV-based fire classification, DL approaches for UAV-based fire detection, and DL approaches for UAV-based fire segmentation.

2.1. Fire Classification Using Deep Learning Approaches for UAV Images

Convolutional Neural Networks (CNNs) are the most popular AI models for images classification tasks. They extract feature maps from input images and then predict their

correct classes (two classes in our case: Fire and Non-Fire). Three main types of layers, which are convolutional layers, pooling layers, and fully connected layers, are employed to build a classical CNN architecture:

- Convolution layers are a set of filters designed to extract basic and complex features such as edges, corners, texture, colors, shapes, and objects from the input images. Then, activation functions are used to add the non-linearity transformation. It helps CNN to learn complex features in the input data. Various activation functions were employed, such as Rectified Linear Unit (ReLU) function [30], Leaky ReLU (LReLU) function [31], parametric ReLU (PReLU) function [32], etc.
- Pooling layers reduce the size of each feature map resulting from the convolutional layers. The most used pooling methods are average pooling and max pooling.
- The fully connected layer is fed by the final flattened pooling or convolutional layers' output, and the class scores for the objects present in the input image are computed.

CNNs showed good results for object classification and recognition [33]. Motivated by their great success, researchers presented numerous CNN-based contributions for fire detection and classification using aerial images in the literature, and these are summarized in Table 1.

Table 1. Deep learning methods for UAV-based fire classification.

Ref.	Methodology	Smoke/Flame	Dataset	Accuracy (%)
[34]	CNN-17	Flame/Smoke	Private dataset: 2100 images	86.00
[35]	AlexNet	Flame	Private dataset: 23,053 images	94.80
	GoogLeNet			99.00
	Modified GoogLeNet			96.90
	VGG13			86.20
	Modified VGG13			96.20
[28]	Xception	Flame	FLAME dataset: 48,010 images	76.23
[36]	Fire_Net	Flame	UAV_Fire dataset: 1540 images	98.00
	AlexNet			97.10
[29]	VGG16	Flame	FLAME dataset: 8617 images	80.76
	VGG19			83.43
	ResNet50			88.01
	InceptionV3			87.21
	Xception			81.30
[37]	Fog computing and simple CNN	Flame	Private dataset: 2964 images	95.07
[38]	Fire_Net	Flame/Smoke	Private dataset: 2096 images	97.50
	AlexNet			95.00
	MobileNetv2			99.30

Chen et al. [34,39] proposed two CNNs to detect wildfire in aerial images. The first CNN contains nine layers [39]. It consists of a convolutional layer with Sigmoid function, max-pooling layer, ReLU activations, Fully connected layer, and Softmax classifier. Using 950 images collected with a six-rotor drone (DJI S900) equipped with a SONYA7 camera, the experimental results showed improvements in accuracy compared to other detection methods [39]. The second includes two CNNs for detecting fire and smoke in aerial images [34]. Each CNN contains 17 layers. The first CNN classifies Fire and Non-Fire, and the second detects the presence of smoke in the input images. Using 2100 aerial images, great performance (accuracy of 86%) was achieved, outperforming the first method and the classical method, which combines LBP (Local Binary Patterns) and SVM [34]. Lee et al. [35] employed five deep CNNs, which included AlexNet [40], GoogLeNet [41], VGG13 [42], a modified GoogLeNet, and a modified VGG13 to detect forest fires in aerial images:

- AlexNet includes eleven layers: five convolutional layers with ReLU activation function, three max-pooling layers, and three fully connected layers;
- VGG13 is a CNN with 13 convolutional layers;

- GoogLeNet contains 22 inception layers, which employ, simultaneously and in parallel, multiple convolutions with various filters and pooling layers;
- Modified VGG13 is a VGG13 model with a number of channels of each convolutional layer and fully connected layers equal to half of that of the original VGG13;
- Modified GoogLeNet is a GoogLeNet model with a number of channels of each convolutional layer and fully connected layer equal to half of that of the original GoogLeNet.

GoogLeNet and the modified GoogLeNet achieved high accuracies thanks to data augmentation techniques (cropping, vertical, and horizontal flip). They showed their ability in detecting wildfires in aerial images [35]. Shamsoshoara et al. [28] proposed a novel method based on the Xception model [43] for wildfire classification. Xception architecture is an extension of the Inceptionv3 model [44] with the modified depth-wise separable convolution, which contains 1×1 convolution followed by a $n \times n$ convolution and no intermediate ReLU activations. Using 48,010 images of the FLAME dataset [45] and data augmentation techniques (horizontal flip and rotation), this method achieved an accuracy of 76.23%. Treneska et al. [29] also adopted four deep CNNs, namely InceptionV3, VGG16, VGG19, and ResNet50 [46], to detect wildfire in aerial images. ResNet50 achieved the best accuracy with 88.01%. It outperformed InceptionV3, VGG16, and VGG19 and the recent state-of-the-art model, Xception, using transfer learning techniques and the FLAME dataset as learning data. Srinivas et al. [37] also proposed a novel method, which integrates CNN and Fog computing to detect forest fire using aerial images at an early stage. The proposed CNN consists of six convolutional layers followed by the ReLU activation function and max-pooling layers, three fully connected layers, and a sigmoid classifier that determines the output as Fire or Non-Fire. This method showed a great performance (accuracy of 95.07% and faster response time) and proved its efficiency to detect forest fires. Zhao et al. [36] presented a novel model called "Fire_Net" to extract fire features and classified them as Fire and Non-Fire. Fire_Net is a deep CNN with 15 layers. It consists of eight convolutional layers with ReLU activation functions, four max-pooling layers, three fully connected layers, and a softmax classifier. Using the UAV_Fire dataset [36], Fire_Net achieved an accuracy of 98% and outperformed previous methods. Wu et al. [38] used a pretrained MobileNetv2 [47] model to detect both smoke and fire. MobileNetv2 is an extended version of MobileNetv1 [48], which is a lightweight CNN with depth-wise separable convolutions. It requires small data and reduces the number of parameters of the model and its computational complexity. It employs inverted residuals and linear bottlenecks to improve the performance of MobileNetv1. Using transfer learning and data augmentation strategies, this method achieved an accuracy of 99.3%. It outperformed published state-of-the-art methods such as Fire_Net and AlexNet and proved its suitability in detecting forest fire on aerial monitoring systems [38].

2.2. Fire Detection Using Deep Learning Approaches for UAV

Region-based CNNs are used to detect, identify, and localize objects in an image. They determine the detected objects' locations in the input image using bounding boxes. These techniques are divided into two categories: one-stage detectors and two-stage detectors [49]. One-stage detectors detect and localize objects as a simple regression task in an input image. Two-stage detectors generate the ROI (Region of Interest) in the first step using the region proposal network. Then, the generated region is classified and its bounding box is determined. Region-based CNNs showed excellent accuracy for object detection problems. They are also employed to reveal the best performance in detecting fires on aerial images.

Table 2 presents deep learning methods for UAV-based fire detection. Jiao et al. [50] exploited the one-stage detector, YOLOv3 [51], to detect forest fires. YOLOv3 is the third version of YOLO deep object detectors. It was proposed to improve the detection performance of older versions by making detections at three different scales and using the Darknet-53 model, which contains 53 convolutional layers as a backbone [51]. Testing results revealed great performances and high speed [50]. Jiao et al. [52] also proposed the UAV-FFD (UAV forest fire detection) platform, which employs YOLOv3 to detect smoke

and fire by using UAV-acquired images. YOLOv3 showed high performance with reduced computational time (F1-score of 81% at a frame rate of 30 frames per second). It proved its potential in detecting smoke/fire with high precision in real-time UAV applications [52]. Alexandrov et al. [53] adopted two one-stage detectors (SSD [54] and YOLOv2 [55]) and a two-stage detector (Faster R-CNN [56]) to detect wildfires. Using large data of real and simulated images, YOLOv2 showed the best performance compared to Faster R-CNN, SSD, and hand-crafted classical methods. It proved its reliability in detecting smoke at an early stage [53]. Tang et al. [57] also presented a novel application to detect wildfire using 4K images, which have a high resolution of 3840 × 2160 pixels collected by UAS (Unmanned Aerial Systems). A coarse-to-fine strategy was proposed to detect fires that are sparse, small, and irregularly shaped. At first, an ARSB (Adaptive sub-Region Select Block) was employed to select subregions, which contain the objects of interest in 4K input images. Next, these subregions were zoomed to maintain the area bounding box's size. Then, YOLOv3 was used to detect the objects. Finally, the bounding boxes in the subregions were combined. Using 1400 4K aerial images, this method obtained a mean average precision (mAP) of 67% at an average speed of 7.44 frames per second.

Table 2. Fire detection using Deep learning methods for UAVs.

Ref.	Methodology	Smoke/Flame	Dataset	Results (%)
[50]	YOLOv3	Flame	Private dataset: 3,840,000 images	F1-score = 81.0
[53]	YOLOv2	Smoke	Private dataset: 12,000 images	Accuracy = 98.3
	Faster R-CNN			Accuracy = 95.9
	SSD			Accuracy = 81.1
[52]	YOLOv3	Flame/Smoke	Private dataset: 3,684,000 images	F1-score = 81.0
[57]	YOLOv3 and ARSB method	Flame	Private dataset: 1400 K images	mAP = 67.0

2.3. Fire Segmentation Using Deep Learning Approaches for UAV

Image segmentation is very important in computer vision. It determines the exact shape of the objects in the images. With the progress of deep learning models, numerous problems were tackled and a variety of solutions was proposed with good results.

Deep learning models are also used to segment fire pixels and detect the precise shape of smoke and/or flame using aerial images. Table 3 shows deep learning methods for UAV-based fire segmentation. For example, Barmpoutis et al. [58] proposed a 360-degree remote sensing system to segment both fire and smoke using RGB 360-degree images, which were collected from UAV. Two DeepLab V3+ [59] models that are encoder–decoder detectors with ASPP (Atrous Spatial Pyramid Pooling) were applied to identify smoke and flame regions. Then, an adaptive post-validation scheme was employed to reject smoke/flame false-positive regions, especially regions with similar characteristics with smoke and flame. Using 150 360-degree images of urban and forest areas, experiments achieved an F1-score of 94.6% and outperformed recent state-of-the-art methods such as DeepLabV3+. These results showed the robustness of the proposed method in segmenting smoke/fire and reducing the false-positive rate [58]. Similarly to wildfire classification, Shamsoshoara et al. [28] proposed a method based on the encoder–decoder U-Net [60] for wildfire segmentation. Using a dropout strategy and the FLAME dataset, U-Net obtained an F1-score of 87.75% and proved its ability to segment wildfire and identify the precise shapes of flames [28]. Frizzi et al. [61] also proposed a method based on VGG16 to segment both smoke and fire. This method showed good results (accuracy of 93.4% and segmentation time per image of 21.1 s) using data augmentation techniques such as rotation, flip, changing brightness/contrast, crop, and adding noises. It outperformed previous published models and proved its efficiency in detecting and classifying fire/smoke pixels [61].

Table 3. Fire segmentation using deep learning methods for UAVs.

Ref.	Methodology	Smoke/Flame	Dataset	Results (%)
[58]	DeepLabV3+	Flame/Smoke	Fire detection 360-degree dataset:	F1-score = 81.4
	DeepLabV3+		150 360-degree images	F1-score = 94.6
	+ validation approach			
[60]	U-Net	Flame	FLAME dataset: 5137 images	F1-score = 87.7
[61]	U-Net	Flame/Smoke	Private dataset: 366 images	Accuracy = 90.2
	CNN based on VGG16			Accuracy = 93.4

3. Materials and Methods

In this section, we first introduced our proposed methods for wildfire classification and segmentation. Then, we describe the dataset used in training and testing. Finally, we present the evaluation metrics employed in this work.

3.1. Proposed Method for Wildfire Classification

To detect and classify fire, we propose a novel method based on deep ensemble learning using EfficientNet-B5 [23] and DenseNet-201 [24] models. EfficientNet models proved their efficiency to reduce the parameters and Floating-Point Operations Per Second using an effective scaling method that employs a compound coefficient to uniformly scale model depth, resolution, and width. EfficientNet-B5 showed excellent accuracy and outperformed state-of-the-art models such as Xception [43], AmoebaNet-A [62], PNASNet [63], ResNeXt-101 [64], InceptionV3 [44], and InceptionV4 [65]. DenseNet (Dense Convolutional Network) connects each layer to all preceding layers to create very diversified feature maps. It has several advantages, including feature reuse, elimination of the vanishing-gradient problem, improved feature propagation, and a reduction in the number of parameters. Using extracted features of all complexity levels, DenseNet shows interesting results in various competitive object recognition benchmark tasks such as ImageNet, SVHN (Street View House Numbers), CIFAR-10, and CIFAR-100 [24].

Figure 1 presents the architecture of the proposed method. First, this method is fed with RGB aerial images. EfficientNet-B5 and DenseNet-201 models were employed as a backbone to extract two feature maps. Next, the feature maps of the two models are concatenated. The concatenated map was then fed an average pooling layer. Then, a dropout of 0.2 was employed to avoid overfitting. Finally, a Sigmoid function was applied to classify the input image into Fire or Non-Fire classes.

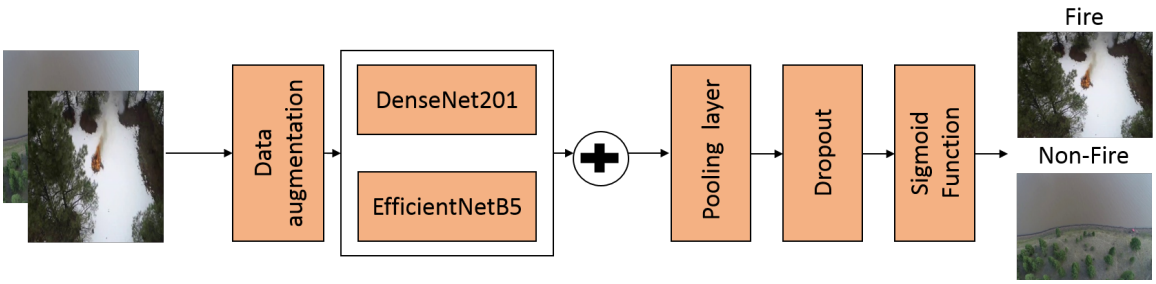

Figure 1. The proposed architecture for wildfire classification.

3.2. Proposed Methods for Wildfire Segmentation

To segment wildfires, we used a CNN model, EfficientSeg [25], and two vision transformers, which are TransUNet [26] and TransFire.

3.2.1. TransUNet

TransUNet [26] is a vision transformer based on U-Net architecture. It employs global dependencies between inputs and outputs using self-attention methods. It is an encoder–decoder. The encoder uses a hybrid CNN-transformer architecture consisting of ResNet-50

and pretrained ViT (Vision Transformer) to extract feature maps. It contains MLP (Multi-Layer Perceptron) and MSA (Multihead Self-Attention) blocks. The decoder employs CUP (cascaded up-sampler) blocks to decode the extracted features and outputs the binary segmentation mask. Each CUP includes a 3 × 3 convolutional layer, ReLU activation function, and two upsampling operators. Figure 2 depicts the architecture of TransUNet.

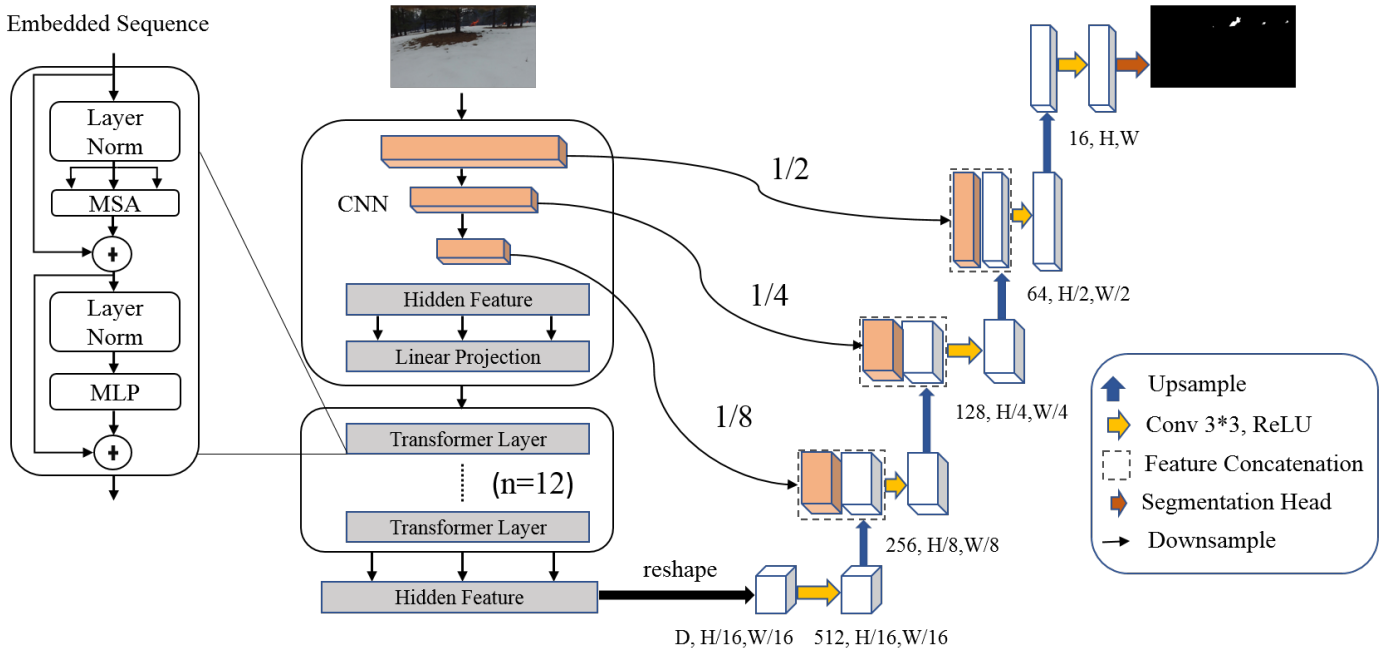

Figure 2. The proposed TransUNet architecture.

3.2.2. TransFire

TransFire is based on MedT (Medical Transformer) architecture. MedT [66] was proposed in order to segment medical images with no requirement of a large dataset for training. Two concepts, gated position-sensitive axial attention and LoGo (Local-Global) training methodology, were employed to improve segmentation performance. Gated position-sensitive axial attention was used to determine long-range interactions between the input features with high computational efficiency. LoGo training methodology used two branches, which are global branch and local branch, to extract feature maps. The first branch works on the image's original resolution. It consists of 2 encoders and 2 decoders. The second operates on image patches. It contains 5 encoders and 5 decoders. The input to both of these branches is the feature extracted using a convolutional block, which includes 3 convolutional layers with ReLU activation function and batch normalization.

TransFire is a modified MedT architecture. It includes one encoder and one decoder in the global branch. It also employs a dropout strategy in the local branch (after the fourth first encoders and the last decoder), in the global branch (after the decoder), and in each input of both of these branches. TransFire was developed to overcome the memory problem of MedT and to prevent overfitting. Figure 3 illustrates the architecture of TransFire.

3.2.3. EfficientSeg

EfficientSeg [25] is a semantic segmentation method, which is based on a U-Net structure and uses MobileNetV3 [27] blocks. It showed impressive results and outperformed U-Net in some medical image segmentation tasks [25].

Figure 4 depicts the architecture of EfficientSeg. It is an encoder–decoder with 4 concatenation shortcuts. It includes five types of blocks, which are MobileNetV3 blocks (Inverted Residual blocks), Downsampling operator, Upsampling operator, and 1 × 1 and 3 × 3 convolutional blocks with ReLU activation function and batch normalization layer.

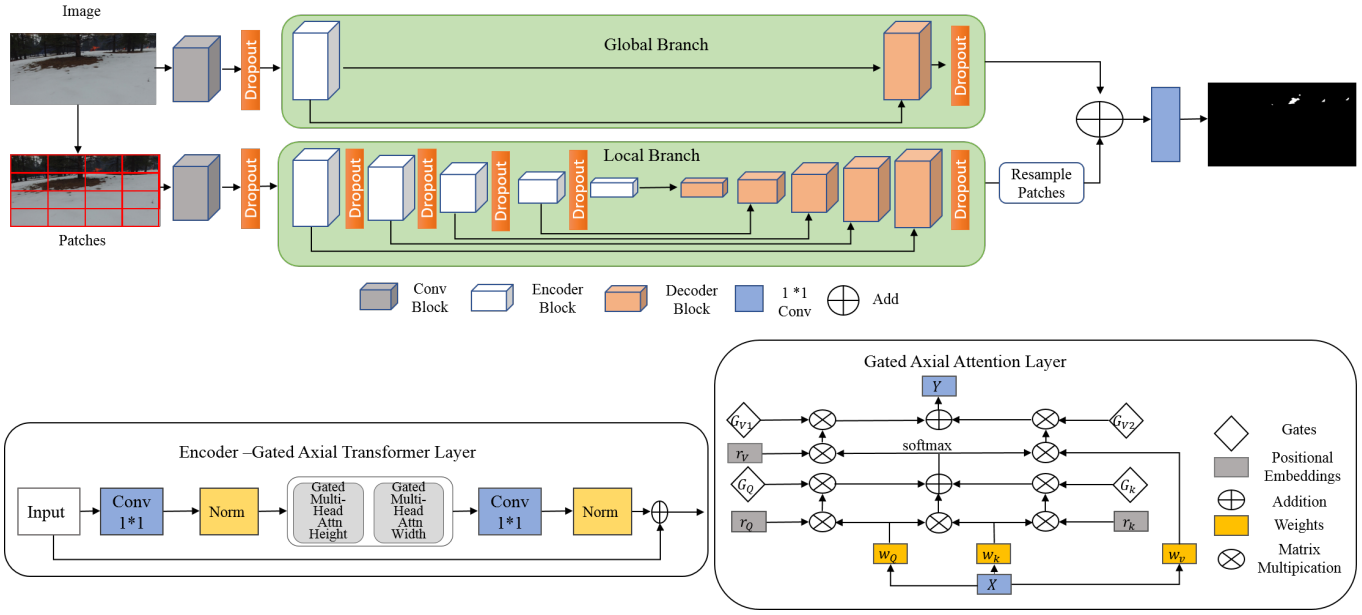

Figure 3. The proposed TransFire architecture.

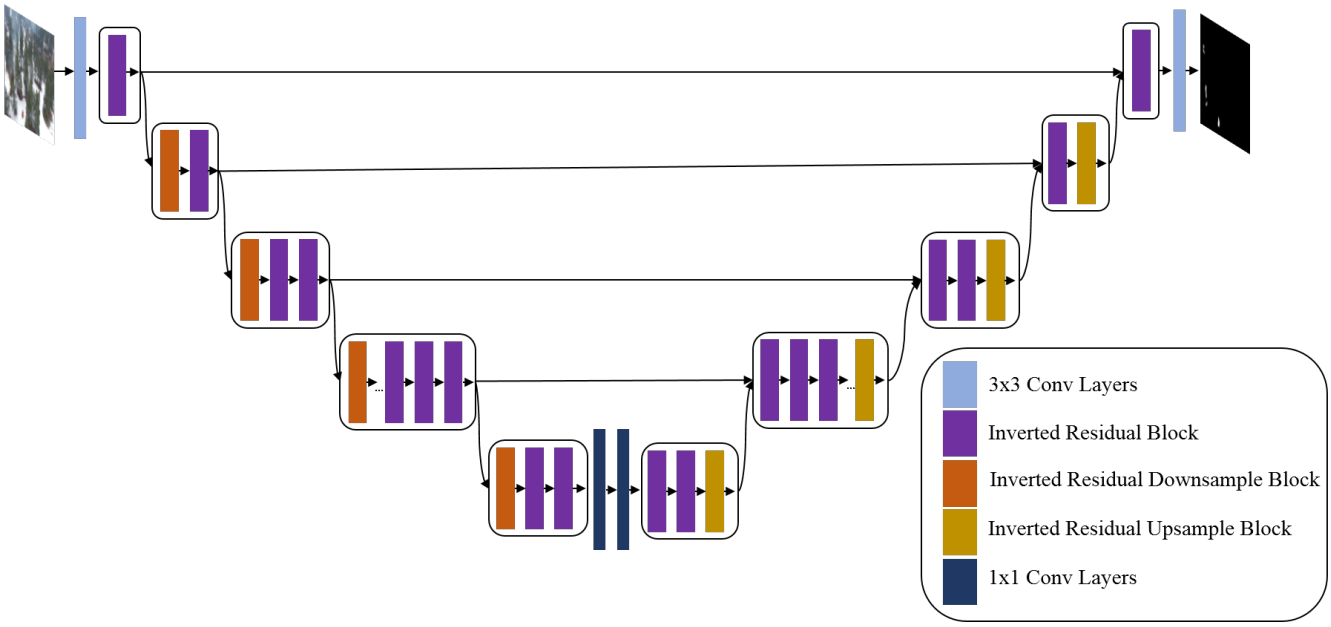

Figure 4. The proposed EfficientSeg architecture.

3.3. Dataset

In the area of deep learning, many large datasets are available for researchers to train their models and perform benchmarking by making comparisons with other methods. However, until recently, there was a lack of a UAV dataset for fire detection and segmentation. In this work, we use a public database called FLAME dataset (Fire Luminosity Airborne-based Machine learning Evaluation) [45] to train and evaluate our proposed methods. The FLAME dataset contains aerial images and raw heat-map footage captured by visible spectrum and thermal cameras onboard a drone. It consists of four types of videos, which are a normal spectrum, white-hot, fusion, and green-hot palettes.

In this paper, we focus on RGB aerial images. We used 48,010 RGB images, which are split into 30,155 Fire images and 17,855 Non-Fire images for wildfire classification task. Figure 5 presents some samples of the FLAME dataset for fire classification. On the other hand, we used 2003 RGB images and their corresponding masks for fire segmentation task. Figure 6 illustrates some examples of RGB aerial images and their corresponding binary masks.

Figure 5. Examples from the FLAME dataset. Top line: Fire images and bottom line: Non-Fire images.

Figure 6. Examples from the FLAME dataset. Top line: RGB images; bottom line: their corresponding binary masks.

3.4. Evaluation Metrics

We used F1-score, accuracy, and inference time to evaluate our proposed approaches for fire classification and segmentation:

- F1-score combines precision and recall metrics to determine the ability of the model in detecting wildfire pixels (as shown by Equation (1)):

$$\text{F1-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

where TP is the true positive rate, FP is the false positive rate, and FN is the false negative rate.

- Accuracy is the proportion of correct predictions over the number of total ones, achieved per the proposed model (as given by Equation (4)):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

where TN is the true negative rate, FN is the false negative rate, TP is the true positive rate, and FP is the false positive rate.

- Inference time is the average time of segmentation or classification using our testing images.

4. Results and Discussion

For wildfire classification, we used TensorFlow [67] and trained the proposed models on a machine with NVIDIA Geforce RTX 2080Ti GPU. The learning data were split as follows: 31,515 images for training, 7878 images for validation, and 8617 images for testing as presented in Table 4.

Table 4. Dataset subsets for classification.

Dataset	Fire Images	Non-Fire Images
Training set	20,015	11,500
Validation set	5003	2875
Testing set	5137	3480

We employed categorical cross-entropy loss (CE) [68], which measures the probability of the presence of a wildfire in the input image (as shown in Equation (5)):

$$CE = -\sum_{c=1}^{M} z_{b,c} \log\left(p_{b,c}\right) \tag{5}$$

where M is the number of classes (in our case two classes (Fire and Non-Fire)), p is the predicted probability, and z is the binary indicator.

For our experiments, we used input RGB images with 254 × 254 resolution, a batch size of 16, and Adam as an optimizer. We also employed the following data augmentation techniques: rotation, shear, zoom, and shift with random values.

For wildfire segmentation, we developed the proposed methods using Pytorch [69] on an Nvidia V100l GPU. Learning data were divided into three sets: 1401 images for training, 201 images for validation, and 401 images for testing. We employed dice loss [70] to measure the difference between the predicted binary mask and the corresponding input mask (as given by Equation (6)). We also used two data augmentation methods, which are a horizontal flip and a rotation of 15 degrees:

$$DC = 1 - \frac{2|Z \bigcap W|}{|Z| + |W|} \tag{6}$$

where Z is the input aerial image, W is the predicted image, and $\bigcap$ is the intersection of the input and the predicted images.

The input data are RGB aerial images with a 512 × 512 resolution and their corresponding binary mask. The TransFire Transformer was trained from scratch (no pretraining) using a hybrid CNN-Transformer as a backbone, patch sizes of 16, and a learning rate of 10^{-3}. TransUNet is evaluated using a learning rate of 10^{-3}, patch size of 16, and two backbones that include a pretrained ViT and a hybrid backbone, which includes ResNet50

(R-50) and pretrained ViT. EfficientSeg also was tested from scratch using a learning rate of 10^{-1}.

We analyzed the proposed methods' performance (accuracy and F1-score) as well as their speed (inference time). In addition, we compared our novel wildfire classification method to state-of-the-art models (Xception [28,29] and InceptionV3 [29]) and deep CNNs (MobileNetV3-Large [27], MobileNetV3-Small [27], DenseNet-169 [24], and EfficientNet-B1-5 [23]), which already showed excellent results for object classification. We also compared the proposed wildfire segmentation methods, including TransUNet, TransFire, and EfficientSeg, to U-Net [28].

4.1. Wildfire Classification Results

We trained wildfire classification methods on aerial images collected using the Matrice 200 drone with a Zenmuse X4S camera. Testing data are collected using the Phantom drone with a Phantom camera.

Table 5 reports a comparative analysis of our proposed method and deep CNN methods using the test data. We can observe that our proposed method achieved the best performance (accuracy of 85.12% and F1-score of 84.77%) thanks to scaled and diversified feature maps extracted by EfficientNet-B5 and DenseNet-201 models. It outperformed recent models for object classification (MobileNetV3-Large, MobileNetV3-Small, DensNet-169, and EfficientNet models (EfficientNet-B2, -B3, -B4, and -B5)) and inception models (Xception and InceptionV3). It proved its good ability to detect and classify forest fires on aerial images. However, it needed a high inference time with 0.018 s.

Figure 7 presents the confusion matrix on test data. We can see that the rate of true positives (classifying Fire as Fire) and the rate of true negatives (classifying No-Fire as No-Fire) are higher than the rate of the false positives (classifying Fire as No-Fire) and the rate of false negatives (classifying No-Fire as Fire), respectively. Our proposed method showed interesting results in detecting and classifying fires, even for very small fire areas. It proved its efficiency to overcome challenging problems such as uneven object intensity and background complexity.

To conclude, our proposed method revealed the best result based on the trade-off between performance and inference time. It showed an excellent capacity to classify forest fires in aerial images and managed to overcome the problems of small fire areas and background complexity.

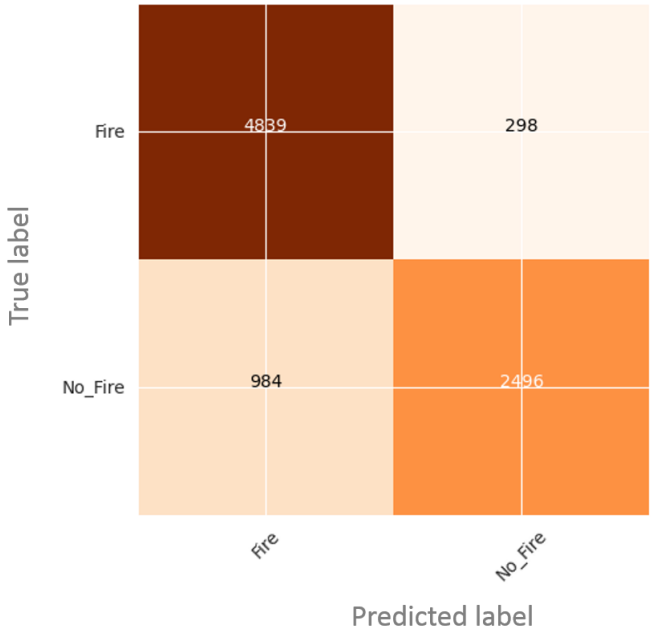

Figure 7. Confusion matrix for fire classification.

Table 5. Performance evaluation of wildfire classification models.

Models	Accuracy (%)	F1-Score (%)	Inference Time (s)
Xception	78.41	78.12	0.002
Xception [28]	76.23	—	—
EfficientNet-B5	75.82	73.90	0.010
EfficientNet-B4	69.93	65.51	0.008
EfficientNet-B3	65.81	64.02	0.004
EfficientNet-B2	66.04	60.71	0.002
InceptionV3	80.88	79.53	0.002
DenseNet169	80.62	79.40	0.003
MobileNetV3-Small	51.64	44.97	**0.001**
MobileNetV3-Large	65.10	60.91	**0.001**
Proposed ensemble model	**85.12**	**84.77**	0.018

4.2. Wildfire Segmentation Results

Table 6 illustrates the quantitative results of fire segmentation using the FLAME dataset. We can see that TransUNet, TransFire, and EfficientSeg obtained excellent results and outperformed U-Net used as a baseline model.

Vision Transformers (TransUNet and TransFire) obtained higher performances compared to deep CNN models (EfficientSeg and U-Net) due to their ability to determine long-range interactions within input features and extract the finer details of the input images. TransUNet-R50-ViT achieved the best performance with an accuracy of 99.9% and an F1-score of 99.9% thanks to local and global features extracted using a hybrid backbone, which includes a CNN, R-50, and pretrained ViT Transformer.

Figure 8 depicts examples of the segmentation of TransUNet-R50-ViT. We can see that this model accurately detected the finer details of fire and distinguished between wildfire and background. In addition, TransUNet-R50-ViT showed its efficiency in localizing and detecting the precise shape of wildfire, especially with respect to small fire areas on aerial images.

TransUNet-ViT also showed excellent performances (accuracy of 99.86% and F1-score of 99.86%) and high speeds (inference time of 0.4 s) compared to TransFire and EfficientSeg. We can see in Figure 8 that TransUNet with ViT transformer accurately segmented wildfire pixels and detected wildfire regions even for small fire areas.

TransUNet models proved their ability in segmenting wildfire, in detecting the exact shape of fire areas, and in overcoming challenging problems such as small fire areas and background complexity. However, they still depend on a pretrained vision transformer (ViT) on a large dataset.

TransFire also showed a higher accuracy with 99.83% and an F1-score of 99.82% due to high-level information and finer features extracted in the global branch and local branch, respectively. It outperformed EfficientSeg and U-Net. It proved its excellent capacity in segmenting wildfire pixels and detecting the exact fire areas, especially small fire areas as shown in Figure 8. It also segmented forest fire pixels under the presence of smoke.

EfficientSeg also obtained a high accuracy with 99.63% and an F1-score of 99.66% thanks to its extracted finer details. It outperformed U-Net. It showed its efficiency in segmenting fire pixels and in detecting the precise shape of fire areas as depicted in Figure 8. However, It had a higher inference time with 1.38 s compared to vision transformers.

To conclude, TransUNet, TransFire, and EfficientSeg showed excellent performances. They proved an impressive potential in segmenting wildfire pixels and determining the precise shape of fire. Based on the F1-score, TransFire showed great performance and outperformed deep convolutional models (EfficientSeg and U-Net) and was very close to the performance of vision transformer (TransUNet). In addition, it demonstrated its reliability in detecting and segmenting wildland fires; in particular, it was the best performing in detecting small fire areas under the presence of smoke, as observed in Figure 9.

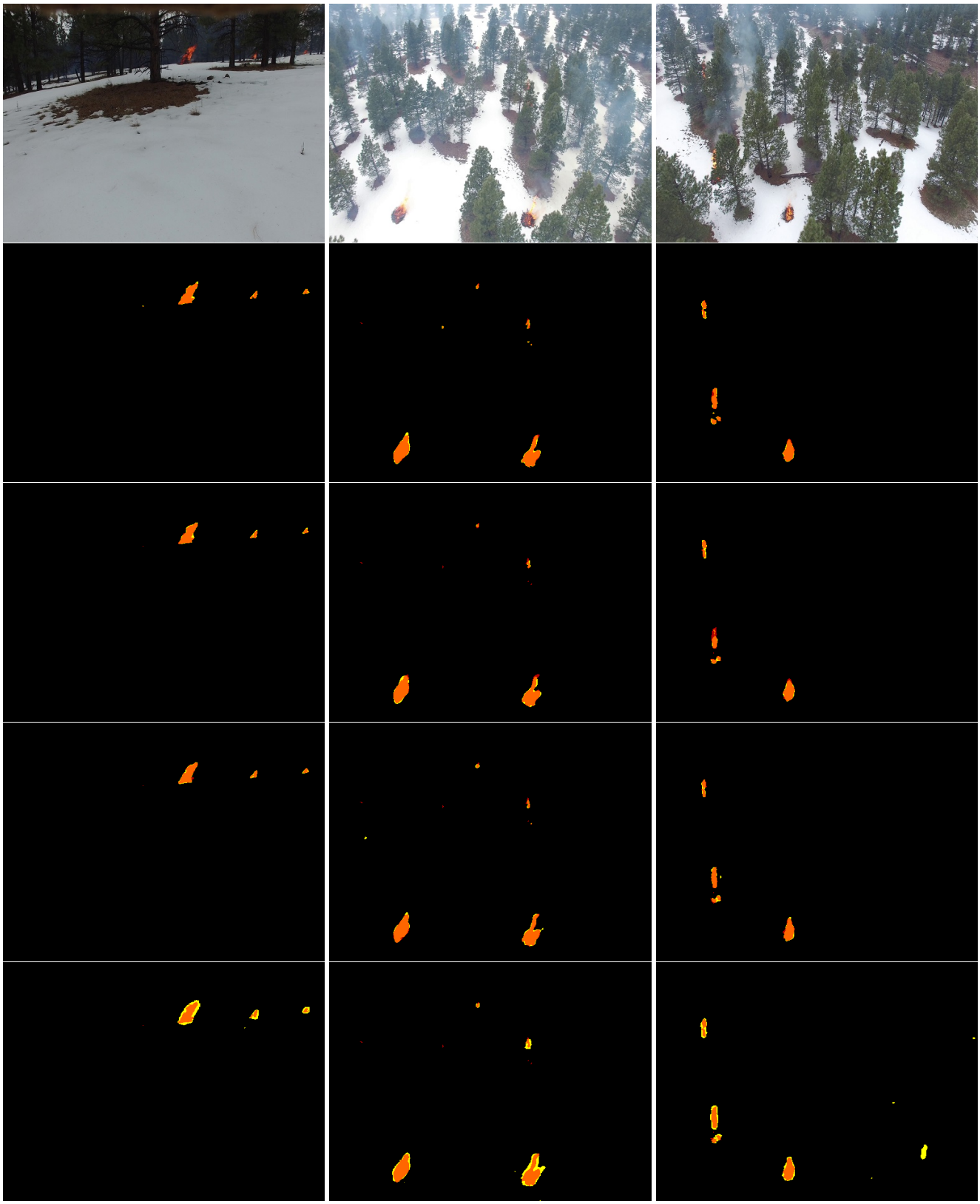

Figure 8. Segmentation results of the proposed models. From top to bottom: RGB aerial images and the results of TransUNet-R50-ViT, TransUNet-ViT, TransFire, and EfficientSeg. Orange represents *TP* (true positives), yellow depicts *FP* (false positives), and red shows *FN* (false negatives).

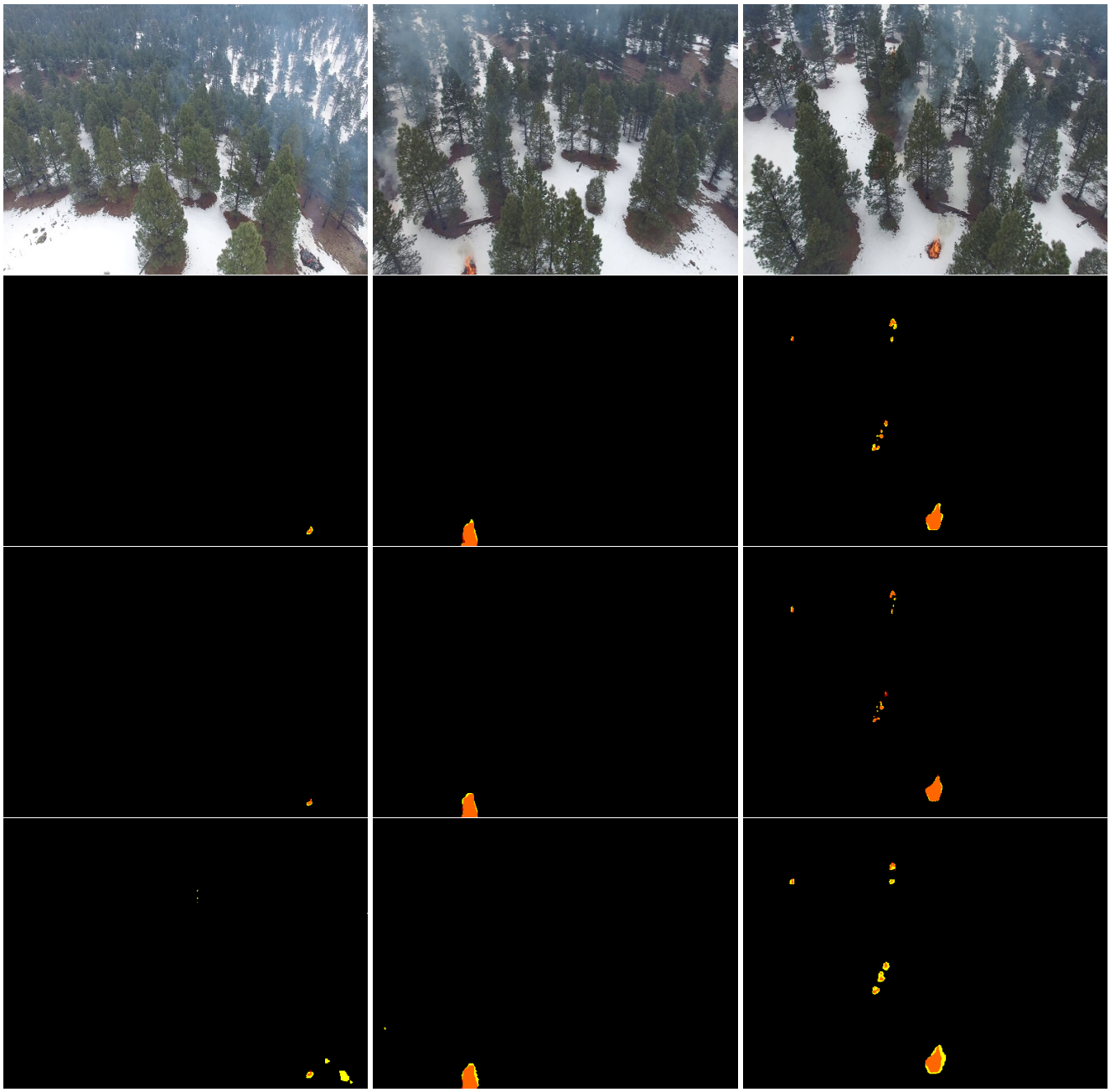

Figure 9. Results of TransFire, TransUNet-R50-ViT, and EfficientSeg. From top to bottom: RGB aerial images and the results of TransFire, TransUNet-R50-ViT, and EfficientSeg. Orange represents *TP*, yellow depicts *FP*, and red shows *FN*. We can see the interesting results of TransFire in determining the precise size of small wildfire areas under the presence of smoke compared to TransUNet and EfficientSeg models.

Table 6. Performance evaluation of wildfire segmentation models.

Models	Accuracy (%)	F1-Score (%)	Inference Time (s)
TransUNet-R50-ViT	**99.90**	**99.90**	0.51
TransUNet-ViT	99.86	99.86	0.40
TransFire	99.83	99.82	1.00
EfficientSeg	99.63	99.66	1.38
U-Net	99.00	99.00	**0.29**

5. Conclusions

In this paper, we address the problem of wildfire classification and segmentation on aerial images using deep learning models. A novel ensemble learning method, which combines EfficientNet-B5 and DenseNet-201 models, was developed to detect and classify wildfires. Using the FLAME dataset, experimental results showed that our proposed method was the most reliable in wildfire classification tasks, presenting a higher performance than recent state-of-the-art models. Furthermore, two vision transformers (TransUNet and TransFire) and a deep CNN (EfficientSeg) are developed to segment wildfires and detect the region of fire areas on aerial images. This is the first proposed approach (in our knowledge) using Transformers for UAV wildfire image segmentation. These models showed impressive results and outperformed recent published methods. They proved their ability in segmenting wildfire pixels, detecting the precise shape of fire. Based on the F1-score, TransFire obtained great performance, outperforming deep models such as EfficientSeg and U-Net. It also showed its excellent potential in detecting and segmenting forest fires and in overcoming challenging problems such as small fire areas and background complexity.

Author Contributions: Conceptualization, M.A.A. and R.G.; methodology, R.G. and M.A.A.; software, R.G.; validation, R.G. and M.A.A.; formal analysis, R.G., M.A.A. and W.S.M.; writing—original draft preparation, R.G.; writing—review and editing, M.A.A. and W.S.M.; funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was enabled in part by support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2018-06233.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This work uses a publicly FLAME dataset, which is available on IEEE-Dataport [45]. More details about the data are available under Section 3.3.

Acknowledgments: The authors would like to thank the support of WestGrid (www.westgrid.ca/) and Compute Canada (www.computecanada.ca).

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

References

1. Aytekin, E. Wildfires Ravaging Forestlands in Many Parts of Globe. 2021. Available online: https://www.aa.com.tr/en/world/wildfires-ravaging-forestlands-in-many-parts-of-globe/2322512 (accessed on 20 November 2021).
2. Dimitropoulos, S. Fighting fire with science. *Nature* **2019**, *576*, 328–329. [CrossRef]
3. Westerling, A.L.; Hidalgo, H.G.; Cayan, D.R.; Swetnam, T.W. Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. *Science* **2006**, *313*, 940–943. [CrossRef] [PubMed]
4. Canadian Wildland Fire Information System. Canada Wildfire Facts. 2021. Available online: https://www.getprepared.gc.ca/cnt/hzd/wldfrs-en.aspx (accessed on 20 November 2021).
5. Gaur, A.; Singh, A.; Kumar, A.; Kulkarni, K.S.; Lala, S.; Kapoor, K.; Srivastava, V.; Kumar, A.; Mukhopadhyay, S.C. Fire Sensing Technologies: A Review. *IEEE Sens. J.* **2019**, *19*, 3191–3202. [CrossRef]
6. Ghali, R.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Recent Advances in Fire Detection and Monitoring Systems: A Review. In Proceedings of the 18th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Genoa, Italy, 18–20 December 2018; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; Volume 1, pp. 332–340.
7. 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]
8. Dao, M.; Kwan, C.; Ayhan, B.; Tran, T.D. Burn scar detection using cloudy MODIS images via low-rank and sparsity-based models. In Proceedings of the IEEE Global Conference on Signal and Information Processing GlobalSIP), Washington, DC, USA, 7–9 December 2016; pp. 177–181.
9. Töreyin, B.U.; Dedeoğlu, Y.; Güdükbay, U.; Çetin, A.E. Computer vision based method for real-time fire and flame detection. *Pattern Recognit. Lett.* **2006**, *27*, 49–58. [CrossRef]
10. Zhang, J.S.; Cao, J.; Mao, B. Application of deep learning and unmanned aerial vehicle technology in traffic flow monitoring. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 9–12 July 2017; Volume 1, pp. 189–194.

11. Chen, C.J.; Huang, Y.Y.; Li, Y.S.; Chang, C.Y.; Huang, Y.M. An AIoT Based Smart Agricultural System for Pests Detection. *IEEE Access* **2020**, *8*, 180750–180761. [CrossRef]
12. Geraldes, R.; Gonçalves, A.; Lai, T.; Villerabel, M.; Deng, W.; Salta, A.; Nakayama, K.; Matsuo, Y.; Prendinger, H. UAV-Based Situational Awareness System Using Deep Learning. *IEEE Access* **2019**, *7*, 122583–122594. [CrossRef]
13. Lee, H.; Jung, S.; Kim, J. Distributed and Autonomous Aerial Data Collection in Smart City Surveillance Applications. In Proceedings of the IEEE VTS 17th Asia Pacific Wireless Communications Symposium (APWCS), Osaka, Japan, 30–31 August 2021; pp. 1–3.
14. Giang, T.L.; Dang, K.B.; Toan Le, Q.; Nguyen, V.G.; Tong, S.S.; Pham, V.M. U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery. *IEEE Access* **2020**, *8*, 186257–186273. [CrossRef]
15. Aposporis, P. Object Detection Methods for Improving UAV Autonomy and Remote Sensing Applications. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), The Hague, The Netherlands, 7–10 December 2020; pp. 845–853.
16. Akhloufi, M.A.; Castro, N.A.; Couturier, A. UAVs for wildland fires. In *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*; International Society for Optics and Photonics: Orlando, FL, USA, 3 May 2018; pp. 134–147.
17. Khennou, F.; Ghaoui, J.; Akhloufi, M.A. Forest fire spread prediction using deep learning. In *Geospatial Informatics XI*; Palaniappan, K., Seetharaman, G., Harguess, J.D., Eds.; International Society for Optics and Photonics: Bellingham, WA, USA, 2021; pp. 106–117.
18. Akhloufi, M.A.; Couturier, A.; Castro, N.A. Unmanned Aerial Vehicles for Wildland Fires: Sensing, Perception, Cooperation and Assistance. *Drones* **2021**, *5*, 15. [CrossRef]
19. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Wildfire Segmentation Using Deep Vision Transformers. *Remote Sens.* **2021**, *13*, 3527. [CrossRef]
20. Yuan, C.; Zhang, Y.; Liu, Z. A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques. *Can. J. For. Res.* **2015**, *45*, 783–792. [CrossRef]
21. Mseddi, W.S.; Ghali, R.; Jmal, M.; Attia, R. Fire Detection and Segmentation using YOLOv5 and U-NET. In Proceedings of the 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 741–745.
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 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 2109–2114.
23. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
24. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
25. Yesilkaynak, V.B.; Sahin, Y.H.; Unal, G.B. EfficientSeg: An Efficient Semantic Segmentation Network. *arXiv* **2020**, arXiv:2009.06469.
26. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
27. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.
28. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Comput. Netw.* **2021**, *193*, 108001. [CrossRef]
29. Treneska, S.; Stojkoska, B.R. Wildfire detection from UAV collected images using transfer learning. In Proceedings of the 18th International Conference on Informatics and Information Technologies, Skopje, North Macedonia, 6–7 May 2021.
30. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; Volume 15, pp. 315–323.
31. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the ICML, Atlanta, GA, USA, 16–21 June 2013; p. 3.
32. Jin, X.; Xu, C.; Feng, J.; Wei, Y.; Xiong, J.; Yan, S. Deep Learning with S-shaped Rectified Linear Activation Units. *arXiv* **2015**, arXiv:1512.07030.
33. Zhao, B.; Feng, J.; Wu, X.; Yan, S. A survey on deep learning-based fine-grained object classification and semantic segmentation. *Int. J. Autom. Comput.* **2017**, *14*, 119–135. [CrossRef]
34. Chen, Y.; Zhang, Y.; Xin, J.; Wang, G.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. UAV Image-based Forest Fire Detection Approach Using Convolutional Neural Network. In Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 2118–2123.
35. Lee, W.; Kim, S.; Lee, Y.T.; Lee, H.W.; Choi, M. Deep neural networks for wild fire detection with unmanned aerial vehicle. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Taipei, Taiwan, 12–14 June 2017; pp. 252–253.
36. Zhao, Y.; Ma, J.; Li, X.; Zhang, J. Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery. *Sensors* **2018**, *18*, 712. [CrossRef]

37. Srinivas, K.; Dua, M. Fog Computing and Deep CNN Based Efficient Approach to Early Forest Fire Detection with Unmanned Aerial Vehicles. In Proceedings of the International Conference on Inventive Computation Technologies, Coimbatore, India, 26–28 February 2020; pp. 646–652.
38. Wu, H.; Li, H.; Shamsoshoara, A.; Razi, A.; Afghah, F. Transfer Learning for Wildfire Identification in UAV Imagery. In Proceedings of the 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 18–20 March 2020; pp. 1–6.
39. Chen, Y.; Zhang, Y.; Xin, J.; Yi, Y.; Liu, D.; Liu, H. A UAV-based Forest Fire Detection Algorithm Using Convolutional Neural Network. In Proceedings of the 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 10305–10310.
40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst. (NIPS)* **2012**, *25*, 1097–1105. [CrossRef]
41. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
43. Chollet, F. Xception: Deep Learning With Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
44. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
45. 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]
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
47. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
48. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
49. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
50. 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 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 23–27 July 2019; pp. 1–5.
51. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
52. Jiao, Z.; Zhang, Y.; Mu, L.; Xin, J.; Jiao, S.; Liu, H.; Liu, D. A YOLOv3-based Learning Strategy for Real-time UAV-based Forest Fire Detection. In Proceedings of the Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 4963–4967.
53. Alexandrov, D.; Pertseva, E.; Berman, I.; Pantiukhin, I.; Kapitonov, A. Analysis of Machine Learning Methods for Wildfire Security Monitoring with an Unmanned Aerial Vehicles. In Proceedings of the 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 8–12 April 2019; pp. 3–9.
54. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October2016; pp. 21–37.
55. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
56. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.
57. Tang, Z.; Liu, X.; Chen, H.; Hupy, J.; Yang, B. Deep Learning Based Wildfire Event Object Detection from 4K Aerial Images Acquired by UAS. *AI* **2020**, *1*, 10. [CrossRef]
58. Barmpoutis, P.; Stathaki, T.; Dimitropoulos, K.; Grammalidis, N. Early Fire Detection Based on Aerial 360-Degree Sensors, Deep Convolution Neural Networks and Exploitation of Fire Dynamic Textures. *Remote Sens.* **2020**, *12*, 3177. [CrossRef]
59. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
60. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
61. Frizzi, S.; Bouchouicha, M.; Ginoux, J.M.; Moreau, E.; Sayadi, M. Convolutional neural network for smoke and fire semantic segmentation. *IET Image Process.* **2021**, *15*, 634–647. [CrossRef]

62. Real, E.; Aggarwal, A.; Huang, Y.; Le, Q.V. Regularized Evolution for Image Classifier Architecture Search. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 4780–4789. [CrossRef]
63. Liu, C.; Zoph, B.; Neumann, M.; Shlens, J.; Hua, W.; Li, L.J.; Fei-Fei, L.; Yuille, A.; Huang, J.; Murphy, K. Progressive Neural Architecture Search. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 19–34.
64. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
65. Szegedy, C.; Ioffe, S.; Vanhoucke, V. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
66. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.10662.
67. Dillon, J.V.; Langmore, I.; Tran, D.; Brevdo, E.; Vasudevan, S.; Moore, D.; Patton, B.; Alemi, A.; Hoffman, M.D.; Saurous, R.A. TensorFlow Distributions. *arXiv* **2017**, arXiv:1711.10604.
68. Ma, Y.; Liu, Q.; Qian, Z. Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; pp. 743–746.
69. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
70. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2017; pp. 240–248.