Deep Learning Approach for UAV-Based Wildfire Detection and Segmentation

K. Sreenath[1], G. Hemalatha Reddy[2], K. Venkata Ranga Reddy[3], G. SnehaSandhya[4],

MD. Mahaboob Basha[5]

[1]Assistant Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India
[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Corresponding author.
Correspondence: K. Sreenath
E-mail: sreenathd@gmail.com

Abstract

Wildfires are a worldwide natural disaster causing important economic damages and loss of lives . Early detection and prediction of fire spread can help reduce affected areas and improve firefighting. 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 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.

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 . They affect 350 million to 450 million hectares every year. In the western US alone, the frequency of wildfires and the total area burned increased by 400% and 600%, respectively, in the last decade . In addition, approximately 8000 wildfires affected 2.5 million hectares each year in Canada . 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 . 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 . In recent years, Unmanned Aerial Vehicles (UAV) or drone systems were deployed in various tasks such as traffic monitoring , precision agriculture , disaster monitoring , smart cities , cover mapping , and object detection . 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 . The integration of UAVs with visual and/or infrared sensors help in finding potential fires at daytime and nighttime . 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 , 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 . 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 and DenseNet-201 models as a backbone for extracting forest fire features. In addition, we employed a deep model (EfficientSeg) and two vision transformers (TransUNet 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 , DenseNet169 , EfficientNet-B1-5 , Xception [28,29], and InceptionV3), which showed excellent results for object classification. TransUNet, TransFire, and EfficientSeg are also evaluated with U-Net . 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.

1.2 Related Works

DL approaches are employed for fire detection and segmentation using aerial images. They proved their ability to detect and segment wildfires . 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.

1. 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 , Leaky ReLU (LReLU)

function [31], parametric ReLU (PReLU) function , 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 . 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.

Table 3. Deep learning methods for UAV-based fire classification.

Ref.	Methodology	Smoke/Flame	Dataset	Accuracy (%)
[14]	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
[19]	Xception	Flame	FLAME dataset 48,010 images	76.23
[34]	Fire_Net	Flame	UAV Fire dataset 1540 images	98.00
	AlexNet			97.50
[13]	VGG16	Flame	FLAME dataset 3647 images	80.76
	VGG19			80.48
	ResNet50			88.01
	InceptionV3			87.21
	Xception			81.30
[37]	Fog computing and simple CNN	Flame	Private dataset 2961 images	95.07
[39]	Fire_Net	Flame/Smoke	Private dataset 7896 images	97.50
	AlexNet			65.60
	MobileNet2			90.30

Chen et al. proposed two CNNs to detect wildfire in aerial images. The first CNN contains nine layers. 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 . 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 . Lee et al. employed five deep CNNs, which included AlexNet , GoogLeNet , VGG13 , 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 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 Shamsoshoara et al. proposed a novel method based on the Xception model for wildfire classification. Xception architecture is an extension of the Inceptionv3 model with the modified depth-wise separable convolution, which contains 1 × 1 convolution followed by a n × n convolution and no intermediate ReLU activations. Using 48,010 images of the FLAME dataset and data augmentation techniques (horizontal flip and rotation), this method achieved an accuracy of 76.23%. Treneska et al. also adopted four deep CNNs, namely InceptionV3, VGG16, VGG19, and ResNet50 , 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. 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. 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 , Fire_Net achieved an accuracy of 98% and outperformed previous methods. Wu et al. used a pretrained MobileNetv2 model to detect both smoke and fire. MobileNetv2 is an extended version of MobileNetv1 , 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 .

1.2.2 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 . 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. exploited the one-stage detector, YOLOv3 , 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 . Testing results revealed great performances and high speed Jiao et al. 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 . Alexandrov et al.adopted two one-stage detectors (SSD and YOLOv2) 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 . Tang et al 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 (%)
[56]	YOLOv3	Flame	Private dataset: 3,868,000 images	F1-score = 81.1
[51]	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 = 81.1
[57]	YOLOv3 and AHSB method	Flame	Private dataset: 1400 K images	mAP = 67.0

1.2.3 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. 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+ 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. proposed a method based on the encoder–decoder U-Net 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 . Frizzi et al. 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 . Sensors 2022, 22, 1977 6 of 1

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 = 91.4
	DeepLabV3+ + validation approach		150 360-degree images	F1-score = 94.6
[61]	U-Net	Flame	FLAME dataset: 5137 images	F1-score = 87.7
[11]	U-Net	Flame/Smoke	Private dataset: 366 images	Accuracy = 98.2
	CNN based on VGG16			Accuracy = 93.4

2.PROPOSED SYSTEM

To deal with the limitations, we adapted and optimized Deep Learning methods to detect wildfire at an early stage. A Deep Learning ensemble method, which combines EfficientNeB5 and DenseNet-201 models, is proposed to identify and classify wildfire using aerial imageThe obtained results are promising and show the efficiency of using Deep Learning and vision transformers for wildfire classification and segmentation. It proved its ability in classifying wildfire even small fire areas.

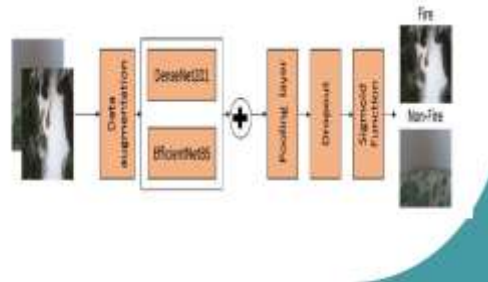**Data Collection** - FLAME dataset (Fire Luminosity Airborne-based Machine learning Evaluation)

Data Processing (Handling missing data, filtering, Train-Test Split)

Model Selection (Deep Learning - EfficientNet-B5 and DenseNet-201)

Train and Evaluate the Model.

Prediction and Classification (Wild fire)

SYSTEM ARCHITECTURE :

 DenseNet 201:DenseNet-201 is a convolutional neural network that is 201 layers deep. we can load a pretrained version of the network trained on more than a million images from the   ImageNet database.

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient.

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

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

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

4.Conclusions

In this paper, 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.

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.
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.
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.
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.; Dedeo ˇglu, Y.; Güdükbay, U.; Çetin, A.E. Computer vision based method for real-time fire and flame detection. Pattern Recognit. Lett. 2006, 27, 49–58.
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
11. S. Jafar Ali Ibrahim et al, "An Overview on Network Representation", Learning Journal of University of Shanghai for Science and Technology ISSN: 1007-6735 Vol.23, Issue 01 Page 60-69, January 2021, https://jusst.org/wp-content/uploads/2021/01/Network-Representation-Learning.pdf