Review

Deep Learning Approaches for Wildland Fires Remote Sensing: Classification, Detection, and Segmentation

Rafik Ghali and **Moulay A. Akhloufi** *

Perception, Robotics and Intelligent Machines Research Group (PRIME), Department of Computer Science, Université de Moncton, Moncton, NB E1A 3E9, Canada
* Correspondence: moulay.akhloufi@umoncton.ca

Abstract: The world has seen an increase in the number of wildland fires in recent years due to various factors. Experts warn that the number of wildland fires will continue to increase in the coming years, mainly because of climate change. Numerous safety mechanisms such as remote fire detection systems based on deep learning models and vision transformers have been developed recently, showing promising solutions for these tasks. To the best of our knowledge, there are a limited number of published studies in the literature, which address the implementation of deep learning models for wildland fire classification, detection, and segmentation tasks. As such, in this paper, we present an up-to-date and comprehensive review and analysis of these vision methods and their performances. First, previous works related to wildland fire classification, detection, and segmentation based on deep learning including vision transformers are reviewed. Then, the most popular and public datasets used for these tasks are presented. Finally, this review discusses the challenges present in existing works. Our analysis shows how deep learning approaches outperform traditional machine learning methods and can significantly improve the performance in detecting, segmenting, and classifying wildfires. In addition, we present the main research gaps and future directions for researchers to develop more accurate models in these fields.

Keywords: wildland fire detection; wildland fire segmentation; wildland fire classification; forest fire; wildfire; drone

1. Introduction

Fires are one of the most dangerous natural risks that present a big threat to the safety of humans, properties, and the environment. They can occur in various environments such as forests, grasslands, bushes, and deserts. They can rapidly spread in the presence of strong wind. If they are not detected early, these fires may cause devastation to forests and other areas of vegetation. For that, permanent monitoring of the situation is required to avoid such disasters. Fires should be detected in the quick and accurate manner possible.

Generally, fires are detected using sensors such as temperature detectors, gas detectors, smoke detectors, and flame detectors. Nevertheless, these detectors have a number of limitations like small coverage areas, delayed response, and availability issues for the general public [1]. Fortunately, the advancement in image processing and computer vision techniques has made it possible to detect fire with no equipment required other than cameras. Traditional ground fire detection tools, such as flame and smoke sensors, are being replaced by vision-based models that have many advantages like accuracy, less prone to errors, robustness towards the environment, lower cost, and large coverage areas compared to these sensors [1].

Through the years, researchers attempted to propose many innovative techniques based on image processing and computer vision in order to set up a fire detection system as accurately as possible. As fire is distinguished by its color, the color analysis technique has been widely used. This technique transforms the image into another color space such as

YCbCr [2,3], Y component is the luminance or luma (brightness) of the color, Cb and Cr are the blue component and red component, respectively, related to the chroma (the color itself), YUV [3–5], which determines the luma of the color using the Y components and the chroma using U and V components, and then classifies its pixels into fire or non-fire based on comparing pixel values to some thresholds. Efficient Forest Fire Detection Index (EFFDI) [6] was also used to detect the tonalities of fires based on the color index. It distinguishes fire areas from the forest background more quickly and efficiently than color space methods using the thresholding process [6]. To model fire, many techniques extracted high-level features like optical flow [7] or dynamic texture [8] and used them either independently or combined [9]. Feature-based techniques presented decent performances in the task of fire detection but they are outperformed by machine learning (ML) techniques [10,11]. Popular methods of fire classification, which predict the presence of a wildfire in an input image include SVM (Support Vector Machine) [12], Markov models [13], Instance-Based Learning classifiers [14], and Bayesian classifiers [15].

The major difficulty of the above-mentioned techniques is the extraction of relevant features that best represent the problem we are dealing with. Instead, it is possible to use a network that learns relevant features on its own. Deep learning (DL) networks have been successfully used in countless areas across all industries such as automatic machine translation [16], image/video classification [17,18], speech recognition [19], video captioning [20], face recognition [21], self-driving cars [22], health diseases detection [23], medical image segmentation [24], and drug discovery [25]. For all of the aforementioned applications, deep learning proved its efficiency in object detection, segmentation, and classification thanks to both automatic extraction and classification of features from data within the same network. In fact, with the availability of huge amounts of data and higher computational power, deep learning has been employed for wildfire classification, detection, and segmentation tasks. It showed its potential to tackle these problems faced by classical ML methods for both ground and aerial images [10]. In addition, automated detection systems using deep learning can be extremely beneficial for the development of fire detection AI agents. Utilizing these models can help detect and track fires quickly and accurately within the view of the camera. If the DL system detects a fire in one view of a camera, it can alert other cameras in the network to adapt their angle and zoom to get a second or third perspective on the fire, which can help more accurately geolocate the fire and its propagation, expediting the dispatch of resources to begin suppression operations. As far as we know, only a few reviews have been introduced in the literature presenting deep learning models used for forest fire classification and detection tasks. Among them, Bouguettaya et al. [26] reviewed DL techniques used for early forest fire detection from UAVs (Unmanned Aerial Vehicles). Akhloufi et al. [27] provided a review of the development of UAV systems in the forest fire context, highlighting existing fire perception models, sensing instruments, cooperative autonomous systems for forest fires, and different coordination strategies. Barmpoutis et al. [28] reviewed fire detection systems (terrestrial, spaceborne, and airborne systems) as well as the smoke and flame detection models used on each system. Geetha et al. [29] presented a survey of smoke and fire detection methodologies, including classical ML and DL models. In contrast, recognizing, detecting, and segmenting wildfires in aerial and terrestrial images are less explored research areas. Bot et al. [30] reviewed recent machine learning methods (published between 2019 and 2022) for forest fire management decision support, presenting the used datasets and evaluation metrics, as well as the main application of these methods, such as pre-fire prevention, active wildfires, and post-fire. In addition, many studies in the literature addressed the detection of wildland fire smoke [31,32]. For such, we provide, in this paper, a comprehensive and up-to-date review only of wildland fire (not wildland smoke) classification, detection, and segmentation tasks based on deep learning techniques as well as the popularly used datasets of each task.

In the process of reviewing the literature, we searched on Google Scholar databases using the keywords: "forest fire detection", "forest fire classification", "forest fire segmentation",

"wildfire", "deep learning" and "vision Transformer". Then, we selected the most relevant articles for the topics of wildland fire classification, detection, and segmentation from 2017 to 2022, where we obtained a total number of 48 articles (23 wildland fire classification articles, 13 wildland fire detection articles, and 22 wildland fire segmentation articles).

The main contributions of this paper are:

- We explore and analyze recent advanced methods (between 2017 and 2022) for wildfire recognition, detection, and segmentation based on deep learning including vision transformers using aerial and ground images.
- We present the most widely used public datasets for forest fire classification, detection, and segmentation tasks.
- We discuss various challenges related to these tasks, highlighting the interpretability of deep learning models, data labeling, and preprocessing.

The rest of the paper is organized as follows: Sections 2–4 review, respectively, the previous research based on deep learning including vision transformers for wildland fire classification, detection, and segmentation. Section 5 presents the most used and public datasets for these tasks. In Section 6, their main challenges are discussed. Finally, Section 7 summarizes the paper.

2. Deep Learning Approaches for Wildland Fire Classification

Convolutional neural networks (CNN) are the most employed and successful image classification models [25]. In general, a CNN takes an input image and predicts the presence of objects (wildland fire in our case) as output as depicted in Figure 1. CNN is composed of two stages: feature extraction and feature classification which are achieved using three types of layers [33]:

- Convolutional layers extract the features from the input data. Activation functions are then applied in order to add the nonlinear transformation to the network and increase its complexity. Numerous activation functions are used in the literature such as ReLU (Rectified Linear Unit) [34], PReLU (parametric ReLU) [35], LReLU (Leaky ReLU) [36], Sigmoid, etc. The resulting output of this layer is called a feature map or activation map.
- The feature maps feed the pooling layer to reduce its size. Among them, max-pooling and average pooling are the most used pooling methods [37].
- Fully connected layers convert the results of the feature extraction stage to 1-D vector and predict the suitable labels for objects in the input image by computing a confidence score.
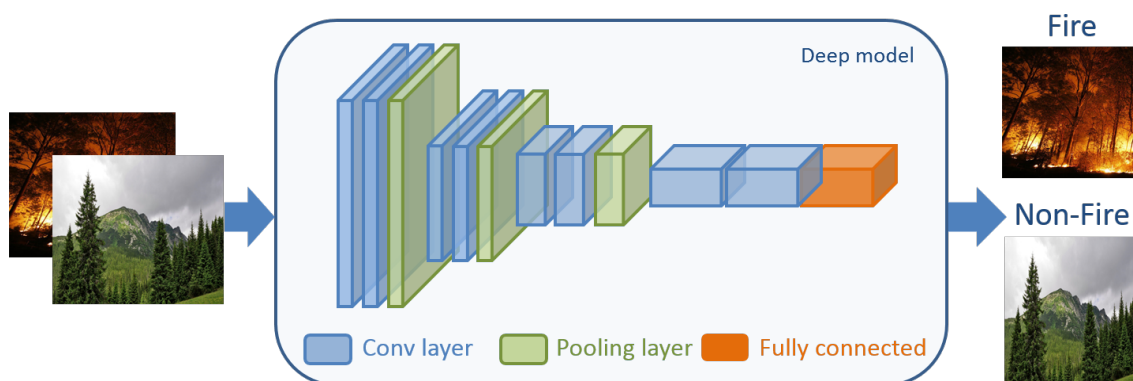
Figure 1. Wildland fire classification based on CNN.

Motivated by the great success of deep learning, the number of contributions dealing with CNN for wildland fire classification in the literature is growing exponentially. Several contributions have been proposed to explore these methods and compare them (see Table 1). Lee et al. [38] reviewed deep neural networks for wildfire detection with UAVs. Their comparative study showed that GoogleNet [39] and the modified GoogleNet [38] present

better performances than AlexNet [40], modified VGG13 [38], and VGG13 [41]. In the same direction, Zhao et al. [42] proposed a 15-layered self-learning DCNN (Deep Convolutional Neural Network) architecture named 'Fire_Net'. It is working as a fire feature extractor and classifier. It has the advantage of using data augmentation techniques that guarantee feature loss avoidance caused by direct resizing. Furthermore, their proposed framework shows a good trade-off between performance and computational cost on the standard aerial fire image dataset 'UAV_Fire'. Srinivas et al. [43] also proposed a forest fire recognition method to identify fires with a drone. This method employed fog computing and a deep CNN, which is based on AlexNet. It was tested using 2964 wildfire images and deployed on a fog device. Results showed a higher accuracy of 95.7% and an accurate response time. Wang et al. [44] used the AlexNet model with an adaptive pooling method, which is integrated with color features for the wildfire recognition task. This method achieved a high recognition rate with an accuracy of 93.75% using the CorsicanFire dataset [45]. Chen et al. [46] proposed a hybrid method to detect both smoke and fire at an early stage. For such, they first combined Local Binary Patterns (LBP) extraction and SVM classifier for smoke detection. Then, two CNN models were designed to detect fire. Shamsoshoara et al. [47] also adopted Xception model [48] to identify flame on aerial images. Using a large dataset (FLAME [47]: 47,992 aerial images) collected by drones in an Arizona pine forest, this model achieved an accuracy of 76.23%. ResNet and VGG models were also evaluated in Arteaga et al. [49] to identify fires. The used dataset includes 1800 images downloaded from the web. The obtained results demonstrated the ability of ResNet models (ResNet18-34-50-101-152) in classifying wildfires using a Raspberry Pi nano-computer. Rahul et al. [50] utilized a modified ResNet50 by adding a convolutional layer, batch normalization, ReLU activation, and a softmax function to classify fires. Dropout and transfer learning techniques were adopted to evaluate this model. An accuracy of 92.27% was achieved, surpassing DenseNet121 and VGG16 by 3.23% and 7.06%, respectively. Sousa et al. [51] used Inception v3 [52] to solve the fire misclassification problem. Data augmentation techniques (cropping and resizing) were employed to remove the logo in the wildfire images and balance the learning dataset. Inception V3 achieved a promising accuracy of 98.6% using the CorsicanFire dataset. Park et al. [53] proposed a deep learning-based forest fire detection method, which adopted the DenseNet model [54] to detect wildfires from surveillance cameras. The CycleGAN (Cycle-consistent Generative Adversarial Network) data augmentation technique [55] was utilized to solve the imbalanced data problem, giving a total of 1395 fire images and 4959 non-fire images. DenseNet showed the best accuracy with 98.27% in comparison to VGG16 and ResNet50. To overcome the problem of lack of data and high computation time, Wu et al. [56] used a pretrained MobilNet v2 [57] in detecting forest fires on aerial images. MobileNet v2 is a lightweight CNN framework designed to operate on embedded or mobile devices. It uses depth-separable convolution, which reduces the computational time for training. It also requires a small dataset for training. The proposed model showed superior performance, achieving the best accuracy with 99.7% compared to Fire_Net and AlexNet. ResNet50 was also used by Tang et al. [58] to develop a deep learning method, called ForestResNet, for forest fire image classification. ForestResNet was trained and evaluated with learning data (175 images) collected from the web and two data augmentation techniques (crop and horizontal flip), obtaining an accuracy of 92%. Dutta et al. [59] combined a simple CNN with an image processing technique to identify wildfires on aerial images. The simple CNN contains six convolutional layers, ReLU activations, max-pooling layers, and three regularization techniques (L2 regularization, batch normalization, and dropout). The image processing technique uses the multi-channel binary thresholding method and the HSV color space thresholding method. Using the FLAME dataset [47], a high sensitivity of 98.10% was achieved, even in the presence of fog and smoke. Ghali et al. [60] developed an ensemble learning using DenseNet201 [54] and EfficientNet-B5 [61] models to identify wildfires on aerial images. Interesting results (accuracy of 85.12%, F1-score of 84.77%, and test time of 0.018 s) were reached using the FLAME dataset and data augmentation techniques (shift with random values, rotation, zoom, and shear). Treneska et al. [62] ex-

plored a deep learning method for the wildfire classification task. Five Finetuned CNNs (VGG16, VGG19, Inception, ResNet50 and Xception) were trained and evaluated on the FLAME dataset. ResNet50 showed the best accuracy of 88.01% and a faster test time of 0.12 s. Zhang et al. [63] proposed a deep learning method, FT-ResNet50, for recognizing wildfires in drone images. FT-ResNet50 is a modified ResNet50 by adding a Mish activation function instead of the ReLU function and a focal loss in place of a binary cross entropy loss. Using the mix-up data augmentation technique, this model revealed an accuracy of 79.48%, surpassing the accuracy of VGG16 and ResNet50 by 6.22% and 3.87%, respectively. Khan et al. [64] studied the VGG19 model in recognizing wildfires on aerial images. VGG19 was trained based on transfer learning using the DeepFire dataset [64], which comprises 950 fire images and 950 non-fire images. It achieved an accuracy of 95%, outperforming existing models, Inception v3 [51] and ForestResNet [58] by 1.40% and 3.0%, respectively, as well as traditional machine learning models, naive Bayes, KNN (k-Nearest Neighbors), SVM, and random forest by 15.53%, 8.95%, 3.69%, and 6.58%, respectively. Dogan et al. [65] developed an ensemble learning method based on pretrained residual networks to identify wildfires. First, four pretrained networks (ResNet19, ResNet50, ResNet101, and InceptionResNet v2) extracted the wildfire features using their last pooling and fully connected layers. Next, the NCA (Neighborhood Component Analysis) was used as a feature selector to choose the best 256 features generated by these models. Finally, the SVM model was adopted to produce the output of the wildfire classification. The proposed method achieved an accuracy of 99.15% using the 10-fold cross-validation strategy and 1650 images. Yandouzi et al. [66] studied eleven deep CNNs (VGG16, ResNet50, MobileNet, VGG19, NASNetMobile, InceptionResNet v2, Xception, Inception v3, ResNet50 v2, DenseNet, and MobileNet v2) in detecting forest fires. ResNet50 showed the best accuracy of 99.94% using numerous data augmentation techniques that are random rotation, gaussian blur, pixel level augmentation, and horizontal and vertical mirroring. Ghosh and Kumar [67] proposed a hybrid deep learning method, which combines a simple CNN and an RNN (Recurrent Neural Network) to identify forest fires. First, the proposed CNN, which consists of six convolutional layers and three maxpooling layers, extracts the low- and high-level features. Second, the proposed RNN generates sequential and continuous features. Finally, two fully connected layers detect wildfires as output. Various data augmentation techniques, such as random zooming, and horizontal and vertical flipping, were applied to augment the training data. An accuracy of 99.62% and 99.10% was achieved using the Mivia [68] and FIRE [69] datasets, respectively. Zheng et al. [70] proposed a dynamic CNN, DCN_Fire, to identify wildfire risk. DCN_Fire consists of 15 layers, including eight convolution layers, four max-pooling layers, and two fully connected layers. The analysis of the performance demonstrated that DCN_Fire had an excellent accuracy of 98.3%, thereby confirming its potential in the recognition of wildfire risks. Mohammed [71] used a pretrained Inception-ResNet v2 for wildfire recognition on the images collected from the Raspberry Pi camera. The result showed that the InceptionResNet v2 obtained a high accuracy of 99.9% utilizing data augmentation techniques (scaling, horizontal and vertical flip). Chen et al. [72] evaluated five deep learning models (Xception, LeNet5, VGG16, MobileNet v2, and ResNet18) in identifying wildfires in aerial images. These models were tested on a large learning dataset (53,451 RGB images: 25,434 Fire/Smoke images, 14,317 Fire/non-smoke images, and 13,700 non-fire/non-smoke images) [73], obtaining F1-score values of 99.92%, 99.36%, 98.46%, 94.53%, and 95.39% for VGG16, MobileNet v2, ResNet18, Xception, and LeNet5, respectively. Guan et al. [74] proposed DSA-ResNet (Dual Semantic attention ResNet) method for wildfire image classification tasks on aerial images. An attention module was added in the ResNet method to dynamically select and merge wildfire characteristics from different scales of convolutional layers. DSA-ResNet obtained an accuracy of 93.65%, thus better than the results of VGGNet, GoogLeNet, ResNet, SE-ResNet [75] models by 8.79%, 5.42%, 2.37%, and 1.19%, respectively.

Table 1. CNN models for wildland fire classification.

Ref.	Methodology	Object Detected	Dataset	Results (%)
[38]	AlexNet, GoogleNet, VGG13	Flame	Private: 23,053 images	Accuracy = 99.00
[42]	Fire_Net	Flame/Smoke	UAV_Fire: 3561 images	Accuracy = 98.00
[43]	Deep CNN	Flame	Private: 2964 images	Accuracy = 95.70
[44]	AlexNet with an adaptive pooling method	Flame	CorsicanFire: 500 images	Accuracy = 93.75
[47]	XCeption	Flame	FLAME: 47,992 images	Accuracy = 76.23
[49]	ResNet152	Flame	Private: 1800 images	Accuracy = 99.56
[50]	Modified ResNet50	Flame	Private: numerous images	Accuracy = 92.27
[51]	Inception v3	Flame	CorsicanFire: 500 images	Accuracy = 98.60
[53]	DenseNet	Flame	Private: 6345 images	Accuracy = 98.27
[56]	MobileNet v2	Flame	Private: 2096 images	Accuracy = 99.70
[58]	ForestResNet	Flame	Private: 175 images	Accuracy = 92.00
[59]	Simple CNN and image processing technique	Flame	FLAME: 8481 images	Sensitivity = 98.10
[60]	EfficientNet-B5, DenseNet-201	Flame	FLAME: 48,010 images	Accuracy = 85.12
[62]	ResNet50	Flame	FLAME: 47,992 images	Accuracy = 88.01
[63]	FT-ResNet50	Flame	FLAME: 31,501 images	Accuracy = 79.48
[64]	VGG19	Flame	DeepFire: 1900 images	Accuracy = 95.00
[65]	ResNet19, ResNet50, ResNet101, InceptionResNet v2, NCA, SVM	Flame	DeepFire & Fire:1650 images	Accuracy = 99.15
[66]	VGG16, ResNet50, MobileNet ,VGG19, NASNet-Mobile , InceptionResNet v2, Xception, Inception v3, ResNet50 v2, DenseNet, MobileNet v2	Flame	Private: 4661 images	Accuracy = 99.94
[67]	CNN,RNN	Flame	FIRE: 1000 images Mivia: 15,750 images	Accuracy = 99.10 Accuracy = 99.62
[70]	DCN_Fire	Flame/Smoke	Private: 1860 images	Accuracy = 98.30
[71]	InceptionResNet v2	Flame/Smoke	Private: 1102 images	Accuracy = 99.90
[72]	Xception, LeNet5, VGG16, MobileNet v2, ResNet18	Flame	FLAME2: 53,451 images	F1-score = 99.92
[74]	DSA-ResNet	Flame	FLAME: 8000 images	Accuracy = 93.65

3. Deep Learning Approaches for Wildland Fire Detection

While an object classification model can predict if an input image contains an object (wildland fires in our case) or not, it cannot determine where the fire is located. Object detection-based DL models determine both the class of objects detected in the input image and its bounding box that localizes this detected object, as illustrated in Figure 2.
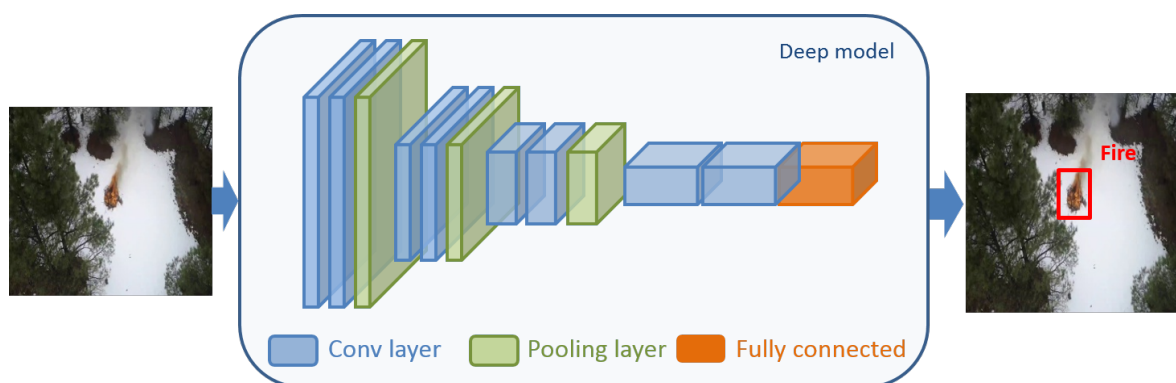
Figure 2. Wildland fire detection based on DL models.

There exist two categories of region-based CNNs: single-stage (also called one-stage) and two-stage object detectors [76]. One-stage detectors treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates. Two-stage detectors generate, first, region proposals using fully convolution architectures, which are then used for object classification and bounding boxes regression. As a result of their tremendous success, researchers introduced various

region-based CNN contributions to wildfire detection in the literature, as summarized in Table 2.

Table 2. Deep learning based wildland fire detection methods.

Ref.	Methodology	Object Detected	Dataset	Results (%)
[77]	Modified Yolo v3	Flame/Smoke	Private: various images & videos	Accuracy = 83.00
[78]	Faster R-CNN, Yolo v1,2,3, SSD	Flame/Smoke	Private: 1000 images	Accuracy = 99.88
[79]	Yolo v3	Flame	Private UAV data	Precision = 84.00
[80]	ARSB, zoom, Yolo v3	Flame	Private: 1400 4k images	mAP = 67.00
[81]	Yolo v5, EfficientDet, EfficientNet	Flame	BowFire, FD-dataset, ForestryImages, VisiFire	AP = 79.00
[82]	Yolo v4 with MobileNet v3	Flame/Smoke	Private: 1844 images	Accuracy = 99.35
[83]	Yolo v4 tiny	Flame	Private: more than 100 images	Accuracy = 91.00
[84]	Yolo v5, U-Net	Flame	CorsicanFire and fire-like objects images: 1300 images	Accuracy = 99.60
[85]	Fire-YOLO	Flame/Smoke	Private: 19,819 images	F1-score = 91.50
[86]	Yolo v5, CBAM, BiFPN, SPPFP	Flame	Private: 3320 images	mAP = 70.30
[87]	FCDM	Flame	Private: 544 images	mAP = 86.90
[88]	Faster R-CNN with multidimensional texture analysis method	Flame	CorsicanFire, Pascal VOC: 1050 images	F1-score = 99.70
[89]	STPM_SAHI	Flame	Private: 3167 images	AP = 89.40

3.1. One Stage Detectors

Various one-stage detectors such as Yolo (You only look once) v1, 2, 3,4, & 5 [90–94], EfficientDet [95], and Single Shot MultiBox Detector (SSD) [96] have been proposed. Yolo models were proposed for the object detection task as a regression method in place of a classification method. They used advanced CNN frameworks such as DarkNet and CSPDarkNet to focus on the most interesting regions of the images and improved the classification performance as well as the accuracy of two-stage detectors. They also showed fast processing time, which is suitable for real-time applications [97]. The first version of Yolo used a pretrained backbone on ImageNet-1000 database [98] and was able to achieve a reliable result suitable for real-time detection. However, this model had multiple errors when localizing neighboring objects. It also misclassified small objects and ones with various sizes in input. To solve this problem, Yolo v2 applied the following process to detect objects:

- Batch normalization.
- Image size change: 448 × 448 instead of 224 × 224 used by Yolo v1.
- Use of anchor boxes to visualize numerous predicted objects.
- Use of multi-scale training image ranging from 320 × 320 to 608 × 608.

This model used DarkNet19 as the backbone which contains five max-pooling layers and nineteen convolutional layers. Yolo v2 proved to be better, faster and stronger than its previous version as well as several state-of-the-art approaches. Yolo v3 came later to improve the performance of Yolo v2 by adding three prediction scales for each input image and bounding box predictions to determine the score for each of the detected bounding boxes. This model employed DarkNet53 containing 53 convolutional layers, as a feature extractor. Thanks to its performances, Yolo v3 was employed in [77] to detect forest fires in real-time using UAV. Jiao et al. [79] also adopted Yolo v3 as a forest fire detector in the UAV-FFD (UAV-based forest fire detection) platform. This model showed a high precision of 84% and a low time cost on the data collected by the UAV. Tang et al. [80] proposed a method for wildfire detection using 4K aerial images. First, an ARSB (adaptive sub-region select block) was applied in extracting a rough area, which includes the objects. Next, a zoom technique was used to find the small objects. Then, Yolo v3 was adopted to detect forest fires. Finally, the result of Yolo v3 was merged with the input 4k images. The proposed method improved the mAP (mean Average Precision) of Yolo v3 from 29% to 67% using 1400 4k aerial images. Yolo v4 [93] is a modified version of Yolo v3 by adding

CSPDarkNet53 as a backbone and PANet (Path Aggregation Network) instead of FPNs (Feature Pyramid Network). It improved the Yolo v3 performance by 10% and its speed by 12%. In [82], a modified version of Yolo v4 was used to identify and locate wildfires. MobileNet v3 [99] was adopted as a backbone to extract forest fire features and decrease the number of parameters. A fast inference speed, low memory requirement, and accuracy of 99.35% were obtained compared to the original version of Yolo v4. Kasyap et al. [83] introduced a mixed learning method based on Yolo v4 tiny and LiDAR techniques to recognize and locate wildfires. Yolo v4 tiny was trained on a desktop computer utilizing more than 100 images and two data augmentation techniques (mix-up and mosaic). It was also tested on a UAV-CPU, achieving an accuracy of 91%. Yolo v5 [94] achieved higher performance using Microsoft COCO [100] and Pascal VOC [101] datasets outperforming all previous versions of Yolo. It improved its performance by using CSPDarkNet as a backbone, PANet as its neck, and Yolo layer as its head. CSPDarkNet integrates CSPNet (Cross Stage Partial Network) [102] into DarkNet. CSPNet minimizes model parameters and size by incorporating gradient information into the feature map and solving the problems of repeating gradient information in large-scale backbones. PANet [103] adopts a new FPN to improve the low-level features' propagation. Xu et al. [81] proposed an ensemble learning approach that combines two detectors (Yolo v5 and EfficientDet [95]) and the classifier EfficientNet [61]. EfficientDet adopts Bi-FPN (Bi-directional Feature Pyramid Network) to extract multi-scale features. EfficientNet employs a novel compound scaling method based on the balance between model width, model depth, and input image resolution. Experimental results showed the potential of this ensemble learning to detect wildfires in different scenarios. Zhao et al. [85] developed a one-stage detection method, Fire-YOLO, for wildfire detection. Fire-YOLO employs EfficientNet as a feature extractor. A large dataset consisting of 19,819 images collected from the web was used to train Fire-YOLO, achieving an F1-score of 91.50% better than Yolo v3 and Faster R-CNN [104]. Mseddi et al. [84] proposed a novel ensemble learning to sequentially comb U-Net [105] and Yolo v5 to detect forest fire and localize the pixels in the image that represent the fire. A reliable detection (accuracy of 99.6%) for wildfire and small fire areas without false alarms of fire-like objects was shown using learning data combined CorsicanFire [45] and fire-like object images collected from the web. Xue et al. [86] proposed an improved wildfire detection method based on Yolo v5. The modified Yolo v5 used SPPFP (Spatial Pyramid Pooling-Fast-Plus) instead of SPPF (Spatial Pyramid Pooling-Fast), CBAM (Convolutional Block Attention Module) to improve the identification of small wildfire targets and BiFPN (Bi-directional Feature Pyramid Network) in place of the PANet (Path Aggregation Network). Based on a transfer learning strategy, this model showed high performance (mAP of 70.3%) in detecting small fires, outperforming the original Yolo v5 model by 10.1%. Xue et al. [87] introduced a wildfire detection method (named FCDM) based on improved Yolo v5. The PANet layer and bounding box loss function were replaced by BiFPN and SIoU loss, respectively. The CBAM attention module was also applied. The experimental results showed that the FCDM method achieved an mAP of 86.9% higher than Yolo v5. The SSD detector [96] applied a multi-resolution feature map to detect and recognize objects at various scales. In [78], SSD was compared to many region-based CNN architectures (Faster R-CNN, Yolo v2, Tiny Yolo v3, and Yolo v3), when applied to solve the problem of forest fire detection at an early stage. Based on the trade-off between accuracy and speed, SSD showed that it is the most suitable for this task.

3.2. Two Stage Detectors

In recent research studies [25,76] object detectors having the highest accuracies are based on a two-stage approach popularized by R-CNN, where a classifier was applied to a sparse set of candidate object locations. A plethora of region-based CNNs models with competitive performances has been proposed in recent years. R-CNN framework [106] was first applied to solve the problem of extracting a large number of regions and selecting 2000 regions using selective search methods. However, R-CNN is slow since it performs a

ConvNet forward pass for each object proposal, without sharing computation. To solve this limitation, Girshick et al. [107] proposed Fast R-CNN, a network that produces a convolutional feature map by processing the whole image with several convolutional and max-pooling layers. Afterward, the Faster R-CNN model [104] was proposed to apply the region proposal network (RPN) in order to generate the regions on the feature map. This last network was employed in [88] with a multidimensional texture analysis method to detect and localize wildfires. An efficient F1-score of 99.7% was obtained using ResNet101 as a backbone and wildfire and non-fire images. It outperformed existing works such as Yolo v3 and SSD. It showed the reliability of Faster R-CNN in detecting forest fires with negligible classification errors. Lin et al. [89] proposed an improved small target wildfire detection method, STPM_SAHI, integrating attention mechanism to model global features of wildfires. STPM_SAHI is composed of a modified Mask R-CNN by adding Swin Transformer [108] as its backbone and PAFPN [109] in place of its feature-fusion network as well as SAHI (Slicing Aided Hyper Inference) technology to overcome the difficulty in detecting small wildfire areas. STPM_SAHI reached an average precision (AP) of 89.4%, outperforming Yolo v5 and EfficientDet.

4. Deep Learning Approaches for Forest Fire Segmentation

Image segmentation or pixel-level classification is one of the main operations in computer vision tasks. It aims to group or cluster similar parts of images or videos together under their respective object class [110]. Wildland fire segmentation is the process of grouping the parts of an input image that belong to the fire class to generate a binary mask as output highlighting the location and shape of the fire (the visible surface of fire), as shown in Figure 3.
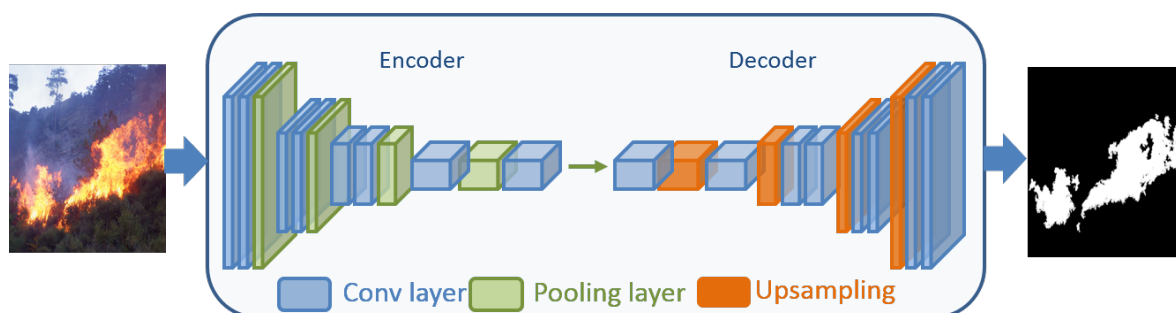
Figure 3. Wildland fire segmentation.

Recently, deep learning-based forest fire segmentation methods showed interesting results in detecting the boundaries of fire as well as its geometrical features such as height, angle, width, and shape. These important results are employed as inputs to develop metrology tools and wildfire propagation models [3,111]. Many studies tackled the wildland fire segmentation problem using deep learning techniques including vision transformers, as shown in Table 3.

For example, Gonzalez et al. [112] proposed an FCN architecture, SFEwAN-SD (Simple Feature Extraction with FCN AlexNet, Single Deconvolution), to detect forest fires. SFEwAN-SD includes two CNNs that are AlexNet and a simple CNN. AlexNet detects the texture and the shape to determine the fire regions. The simple CNN, which consists of numerous 3×3 convolutional layers followed by ReLU activation extract fire features (color and texture). Based on the F1-score and processing time, good results are obtained outperforming the state-of-the-art method proposed by Frizzi et al. [113]. Wang et al. [114] developed a deep learning model, which consists of SqueezeNet [115] as a backbone and a multi-scale context aggregation module to detect and segment wildfires. This model was tested with numerous videos of real forest fires. It showed great potential with a false positive rate (FP) of 5% to identify the region of forest fires at an early stage. Choi et al. [116] presented a novel semantic fire image segmentation method based on FusionNet [117].

An accuracy of 97.46% was attained on CorsicanFire and FiSmo [118] datasets surpassing existing fire and object segmentation deep learning models. Akhloufi et al. [119] employed the encoder-decoder, U-Net, in segmenting wildfires and detecting their regions. The encoder extracts the wildfire features from fire input images. It contains convolutional layers of size 3×3, a max-pooling layer, and a ReLU activation. The decoder decodes the produced feature maps to generate the fire mask. It includes convolutional layers of size 3×3, Upsampling layer, up-convolution layer, and ReLU activation. Finally, a convolutional layer of size 1×1 generates the binary mask of input images. U-Net obtained an F1-score value of 97.09% and showed its ability in segmenting forest fire using the CorsicanFire dataset (419 images) and Dice loss [120]. Shamsoshoara et al. [47] also employed U-Net to detect and segment wildfires on aerial images. An F1-score of 87.75% was reached using the FLAME dataset and showed the U-Net efficiency in determining the fire regions on UAV images. Bochkov et al. [121] proposed a novel encoder-decoder, wide-UUNet concatenative (wUUNet), to determine the region of the fire. wUUNet includes the modernization of U-Net, UUNet, which comprises two U-Net architectures with a skip connection between the decoder of the first U-Net and the encoder of the second U-Net. The first determines the binary segmentation of fire areas. The second detects the specific fire colors (yellow, red, and orange) as a multi-class segmentation. Excellent performance was achieved outperforming the U-Net model in both binary and multi-class segmentation. Ghali et al. [122] explored three deep models that are U^2-Net [123], U-Net, and EfficientSeg [124] in segmenting fire pixels and detecting fire regions thanks to their excellent results in medical image task. EfficientSeg with Dice loss showed the best F1-score of 95% surpassing deep learning models (U-Net and U^2-Net) and classical method (color space fusion method [125]). It confirmed its reliability to overcome the false detection of wildfire pixels. Song et al. [126] developed a deep learning-based binary segmentation method (SFBSNet) to segment fire areas. SFBSN adopted a modified version of FusionNet [117] by replacing the residual block with the confusion block and the conventional convolution with a depth separable convolution. The obtained results showed that SFBSNet performed better on the CorsicanFire dataset, reaching an IoU of 90.76%. Ghali et al. [127] introduced a forest fire segmentation method based on deep learning (Deep-RegSeg). First, they used the RegNet model as a backbone to encode the wildfire characteristics, followed by three Deconv blocks, Conv block, and 1×1 convolutional to determine the output fire mask. Each Deconv block includes a transposed convolution layer, a batch normalization layer, and a ReLU activation. The Conv block comprises a 3×3 convolutional layer, a batch normalization, and a ReLU function. Deep-RegSeg achieved an F1-score of 94.46%, allowing accurate identification of wildfire zones very close to the input mask.

Table 3. Deep learning based wildland fire segmentation methods.

Ref.	Methodology	Object Segmented	Dataset	Results (%)
[112]	SFEwAN-SD	Flame	Private: 560 images	F1-score = 90.31
[116]	Encoder-decoder based on FusionNet	Flame	CorsicanFire, FiSmo: 212 images	Accuracy = 97.46
[114]	CNN based on SqueezeNet	Flame	Private: various videos	FP = 5.00
[119]	U-Net	Flame	CorsicanFire: 419 images	Accuracy = 97.09
[47]	U-Net	Flame	FLAME: 5137 images	F1-score = 87.70
[121]	wUUNet	Flame	Private: 6250 images	Accuracy = 95.34
[122]	U-Net, U^2-Net, EfficientSeg	Flame	CorsicanFire: 1135 images	F1-score = 95.00

Table 3. *Cont.*

Ref.	Methodology	Object Segmented	Dataset	Results (%)
[126]	SFBSNet	Flame	CorsicanFire: 1135 images	IoU = 90.76
[127]	Deep-RegSeg	Flame	CorsicanFire: 1135 images	F1-score = 94.46
[128]	DeepLab v3+	Flame	CorsicanFire: 1775 images	Accuracy = 97.67
[129]	DeepLab v3+ + validation approach	Flame/Smoke	Fire detection 360-degree dataset: 150 360-degree images	F1-score = 94.60
[130]	DeepLab v3+ with Xception	Flame	CorsicanFire: 1775 images	Accuracy = 98.48
[131]	DeepLab v3+	Flame	CorsicanFire, FLAME, private: 4241 images	Accuracy = 98.70
[132]	SqueeZeNet, U-Net, Quad-Tree search	Flame	CorsicanFire, private: 2470 images	Accuracy = 95.80
[133]	FireDGWF	Flame/Smoke	Private: 4856 images	Accuracy = 99.60
[134]	U-Net, DeepLab v3+, FCN, PSPNet	Flame	FLAME: 4200 images	Accuracy = 99.91
[135]	ATT Squeeze U-Net	Flame	CorsicanFire& Private: 6135 images	Accuracy = 90.67
[136]	Encoder-decoder with attention mechanism	Flame	CorsicanFire: 1135 images+ various non-fire images	Accuracy = 98.02
[137]	TransUNet, MedT	Flame	CorsicanFire: 1135 images	F1-score = 97.70
[60]	TransUNet, TransFire	Flame	FLAME: 2003 images	F1-score = 99.90
[74]	MaskSU R-CNN	Flame	FLAME: 8000 images	F1-score = 90.30
[138]	Improved DeepLab v3+ with MobileNet v3	Flame	FLAME: 2003 images	Accuracy = 92.46

DeepLab models (DeepLab v3 [139] and DeepLab v3+[140]) are proposed to address the problem of reducing the spatial resolution of the generated feature maps. The two detectors showed important performance in the object segmentation task owing to the rich and multi-scale extracted features. DeepLab v3 [139] encodes multi-scale information using atrous convolution as pooling operations and ASPP (atrous spatial pyramid pooling) as multiple effective fields-of-view. DeepLab v3+ [139] adds a simple and efficient parameter, atrous rate, into DeepLab v3 decoder to improve segmentation results and enlarge the field of views without increasing the parameters number and computation time. Harkat et al. [128] employed DeepLab v3+ in segmenting forest fire and detecting fire regions. The capability of DeepLab v3+ was evaluated with various loss functions (Tversky and Dice loss) and two backbones (ResNet50 and ResNet18) using infrared and RGB images. The obtained results (accuracy of 97.67%) are very promising and demonstrated the efficiency of DeepLab v3+ in forest fire segmenting tasks. Harkat et al. [130] also applied DeepLab v3+ with an Xception model as a backbone to detect fire on RGB and Infrared (IR) images. Three loss functions (Dice, Cross entropy, and Tversky loss), three learning rates (10^{-1}, 10^{-2}, and 10^{-3}), and the CorsicanFire dataset (1135 RGB images and 640 IR images) were used to train and evaluate this model. Experiment results showed a promising performance (accuracy of 98.48%) for the deployment of this architecture in wildland fire segmentation. Harket et al. [131] also adopted DeepLab v3+ to segment fire on aerial images. Using multimodal data (RGB, IR, Near-Infrared) learning data, Dice loss, and ResNet50 as a backbone, DeepLab v3+ achieved an accuracy of 98.70%. It showed a high capacity for detecting and localizing fire zones in aerial images. Barmpoutis et al. [129] proposed a novel fire detection remote sensing system, which uses two DeepLab v3+ models with a novel validation method to improve wildfire segmentation on RGB images collected from a 360-degree camera mounted on a UAV. DeepLab v3+ models were employed to detect the fire areas, while the validation method was adopted to reject the false-positive regions of smoke or flame. Using 150 images of urban and forest areas (real and synthetic fire images), this system reduced the false-positive rate and showed great potential for wildfire segmentation tasks. Pan et al. [133] proposed a method, named FireDGWF, which consists of Faster R-CNN and weakly supervised segmentation method (WSFS) for forest fire detection tasks. FireDGWF consists of two steps: the detection process and the grading process. In the detection process, three methods (ShuffleNet [141], WSFS as a region segmentation method, and Faster R-CNN as a region detection method) were used

to identify, segment, and locate wildfire areas in input images. In the grading process, a fuzzy evaluation mechanism was adopted to evaluate the severity level of wildfires based on the three outputs of the detection process. FireDGWF reached a competitive result (an accuracy of 99.6%) compared to existing methods such as U-Net and DeepLab v3. Wang et al. [134] presented a comparative analysis of wildland fire image segmentation methods using aerial images. Four semantic segmentation methods (U-Net, DeepLab v3+, FCN [142], and PSPNet [143]) were evaluated with two backbones (VGG16 and ResNet50). Using 4200 aerial images, U-Net with ResNet 50 showed the best accuracy with 99.91%. Perrolas et al. [132] proposed a multi-task method for detecting fire regions based on the quad-tree search method. The input images were divided into fixed sizes of patches. For each patch, a classification model (SqueezeNet) was employed to determine which patch includes a fire instance. If a fire was detected, a segmentation model (U-Net) was used to segment the fire areas. If the segmented fire areas are small, the patch was split into four portions, and each portion was zoomed in to locate the precise shape of the fire. The patches were finally combined to create the complete fire mask. The proposed method achieved an accuracy of 95.8%, demonstrating its ability to locate small fire areas. Guan et al. [74] developed a wildfire detection segmentation (called MaskSU R-CNN) based on MS R-CNN method [144], which is an improved version of Mask R-CNN model [145]. A higher F1-score of 90.30% was obtained, surpassing the popular semantic segmentation methods such as SegNet [146], U-Net, PSPNet, and DeepLab v3. Li et al. [138] introduced a real-time forest fire segmentation method based on an improved version of DeepLab v3+. The MobileNet v3 model was adopted as the encoder to reduce the inference time and model parameters. The improved DeepLab v3+ showed an accuracy of 92.46% higher than the original DeepLab v3+ and an inference time of 59 frames per second (FPS), reducing the time of DeepLab v3+ by 35 FPS. Encoder-decoder models showed impressive performance on various tasks, in particular fire detection and segmentation task. They are based on a convolution layer, which extracts only local features. However, they are still limited by high computational costs and global feature modeling.

Recently, transformers were developed to address encoder-decoder limitations. They model long-range dependencies between patches as input using the attention module, which removes irrelevant information while focusing on some relevant information. Transformers were first used in natural language processing (NLP), where they performed very well owing to their generalization and simplicity [147,148]. Then, they were adopted in computer vision tasks such as image super-resolution [149] and object segmentation [150]. The obtained results are very impressive compared to convolutional models. Vision transformers are also used in fire segmentation. For instance, Ghali et al. [60,137] studied the potential of transformers in the context of forest fire segmentation using ground and aerial images. Two vision transformers (TransUNet [151] and Medical Transformer (MedT) [152]) were explored in segmenting and detecting wildfire areas in ground images. Various settings (varying backbone and input size) were employed to evaluate these models. TransUNet and MedT revealed an F1-score value of 97.70% and 96.00% outperforming deep CNNs models (EfficientSeg, U^2-Net, and U-Net) [137] and traditional methods (fusion color space method [125]). TransUNet and Transfire, which are modified MedT, were also adopted for fire segmentation tasks using aerial images [60]. TransUNet and TransFire achieved an F1-score value of 99.90% and 99.82%, respectively. They also proved their reliability in extracting the finer details of forest fire on aerial images and in overcoming many challenges such as small wildfire areas and background complexity [60]. Zhang et al. [135] also proposed an encoder-decoder U-shape method with an attention module, called ATT Squeeze U-Net (Attention U-Net and SqueezeNet) to detect wildfire on RGB images. The encoder uses SqueezeNet with eight Fire modules as a backbone to extract wildfire features. Each Fire module includes a 1×1 convolutional layer, a 3×3 depthwise convolutional layer, and channel Shuffle operations. The decoder employs 3×3 and 1×1 convolutional layers, a ReLU activation function, and three DeFire modules, each one consisting of ReLU functions and 1×1 and 3×3 convolutional layers. Three attention modules were also

adopted in skip connections to rely on the encoder and the decoder. ATT Squeeze U-Net achieved an accuracy of 90.67% better than popular methods such as Attention U-Net and U-Net using the CorsicanFire dataset. Niknejad and Bernardino [136] also developed an encoder-decoder segmentation method with an attention mechanism to identify and segment wildfire regions. First, the encoder of DeepLab v3+ was employed as a backbone to extract wildfire characteristics. Then, two attention modules, the decoder of DeepLab v3+, and a Sigmoid function were used as a decoder, before generating the binary mask as output using a convolutional layer. This proposed network improved the accuracy of the original DeepLab v3+ and U-Net models by 0.84% and 1.14%, respectively.

5. Datasets

Similarly to all problems based on deep learning models, a large learning dataset is required. In the literature, many large databases have been put available for researchers in order to make a comparison to other techniques dealing with the same problem easier. However, this is not the case for fire detection problems which makes their evaluation a bit challenging. The number of available fire datasets is very limited. In this section, we present the popular datasets used for wildfire recognition, detection, and segmentation tasks (see Table 4).

Table 4. Wildland fire datasets overview. RGB (visible spectrum) covers a wavelength range of 0.4 to 0.75 μm. NIR (Near Infrared) ranges from 0.7 to 1.4 μm. LWIR (Long Wave Infrared)) includes a wavelength range of 8 to 15 μm.

Ref.	Data Name	RGB/IR	Image Type	Fire Area	Number of Images/Videos	Labeling Type
[153,154]	BowFire	RGB	Terrestrial	Urban/Forest	226 images: 119 fire images and 107 non-fire images 226 binary mask	Classification Segmentation
[47,155]	FLAME	RGB/LWIR	Aerial	Forest	48,010 images: 17,855 fire images and 30,155 non-fire images 2003 binary mask	Classification Segmentation
[45,156]	CorsicanFire	RGB/NIR	Terrestrial	Forest	1135 images and their corresponding binary mask	Segmentation
[157,158]	FD-dataset	RGB	Terrestrial	Urban/Forest	31 videos: 14 fire videos and 17 non-fire videos 50,000 images: 25,000 fire images and 25,000 non-fire images	Classification
[159]	ForestryImages	RGB	Terrestrial	Forest	317,921 images	Classifcation
[160]	VisiFire	RGB	Terrestrial	Urban/Forest	12 videos	Classification
[161]	Firesense	RGB	Terrestrial	Urban/Forest	29 videos: 11 fire videos, 13 smoke videos, and 25 non-fire/smoke videos	Classification
[68]	MIVIA	RGB	Terrestrial	Urban/Forest	31 videos: 17 fire videos and 14 non-fire videos	Classification
[118]	FiSmo	RGB	Terrestrial	Urban/Forest	9448 images and 158 videos	Classification
[64,162]	DeepFire	RGB	Terrestrial	Forest	1900 images: 950 fire images and 950 non-fire images	Classification
[69]	FIRE	RGB	Terrestrial	Forest	999 images: 755 fire images and 244 non-fire images	Classification
[72,73]	FLAME2	RGB/LWIR	Aerial	Forest	53,451 images: 25,434 fire images, 14,317 fire/non-smoke images, and 13,700 non-fire	Classification

- **BowFire** (Best of both worlds Fire detection) [153,154] dataset is a public data of fire. It consists of 226 images (119 fire images and 107 non-fire images) with different resolutions, as shown in Figure 4. Fire images represent emergencies with various fire situations (forest, burning buildings, car accidents, industrial fires, etc.) and fire-like objects such as yellow or red objects and sunsets. It includes both forest and non-

forest images. Nevertheless, it is important to note that the non-forest images were filtered out to ensure the performance and reliability of the trained wildfire models. BowFire also contains the corresponding masks of fire/non-fire images for the fire segmentation task, as shown in Figure 5.

Figure 4. Examples from the BowFire dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

Figure 5. Samples of the BowFire dataset. (**Top**): RGB images; (**Bottom**): their corresponding binary masks.

- **FLAME** (Fire Luminosity Airborne-Based Machine Learning Evaluation) dataset [47,155] consists of aerial images and raw heat-map footage collected by thermal cameras and visible spectrum onboard two drones (Phantom 3 Professional and Matrice 200). It contains four types of videos that are a green-hot palette, normal spectrum, fusion, and white-hot. It includes 48,010 RGB aerial images (with a resolution of 254 × 254 pix.), which are divided into 17,855 images without fire and 30,155 images with fire for the wildfire classification task, as illustrated in Figure 6. It also comprises 2003 RGB images

with a resolution of 3480 × 2160 pix. and their corresponding masks for fire segmentation task, as depicted in Figure 7.

Figure 6. Examples from the FLAME dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

Figure 7. Samples of the FLAME dataset. (**Top**): RGB aerial images; (**Bottom**): their corresponding binary masks.

- **CorsicanFire** dataset [45,156] consists of NIR (near infrared) and RGB images. The NIR images are collected with a longer exposure/integration time. CorsicanFire includes a larger number of fire images with many resolutions (1135 RGB images and their corresponding masks) that are widely used in the context of fire segmentation. It describes the visual information of the fire such as color (orange, white-yellow, and red), fire distance, brightness, smoke presence, and different weather conditions. Figure 8 shows CorsicanFire dataset samples and their corresponding binary masks.

Figure 8. Examples from the CorsicanFire dataset. (**Top**): RGB images; (**Bottom**): their corresponding masks.

- **FD-dataset** [157,158] is composed of two datasets, BowFire and dataset-1 [9], which contains 31 videos (14 fire videos and 17 non-fire videos) and fire/non-fire images collected from the internet. It contains 50,000 images with numerous resolutions (25,000 images with fire and 25,000 images without fire) describing various fire incidents such as red elements, burning clouds, and glare lights. It also includes fire-like objects such as sunset and sunrise, as illustrated in Figure 9. This dataset consists of both forest and non-forest images, but it is important to mention that only the forest images were selected for training the forest fire models in order to improve their performance.

Figure 9. Examples from the FD-dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

- **ForestryImages** [159] is a public dataset proposed by the University of Georgia's Center for Invasive Species and Ecosystem Health. It contains a large number of images (317,921 images with numerous resolutions) covering different image categories such as forest fire (44,606 images), forest pests (57,844 images), insects (103,472 images), diseases (30,858 images), trees (45,921 images), plants (149,806 images), wildlife (18,298 images), etc. as shown in Figure 10.

Figure 10. Examples from the ForestryImages dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

- **VisiFire** [160] is one of the widely used datasets. It contains twelve videos of different fire scenes collected by a frame grabber and an ordinary surveillance camera. Figure 11 shows VisiFire dataset samples.

Figure 11. Examples from the VisiFire dataset.

- **Firesense** dataset [161] is a public dataset developed within the "FIRESENSE - Fire Detection and Management through a Multi-Sensor Network for the Protection of Cultural Heritage Areas from the Risk of Fire and Extreme Weather (FP7-ENV-244088)" project to train and test smoke/fire detection algorithms. It contains eleven fire videos, thirteen smoke videos, and twenty-five non-fire/smoke videos. Figure 12 depicts Firesense dataset samples.

Figure 12. Examples from the Firesense dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

- **MIVIA** dataset [68] is a collection of videos used for various tasks. It contains 31 videos (17 fire videos and 14 non-fire videos) collected in a real environment and from the Firesense dataset. Figure 13 depicts MIVIA fire detection dataset samples.

Figure 13. Examples from the MIVIA dataset. (**Top**): Fire images; (**Bottom**) line: non-fire images.

- **FiSmo** is public data for fire detection developed by Cazzolato et al. [118] in 2017. It contains images and video data with their annotation. It contains 9448 images with multiple resolutions and 158 videos acquired from the web. Each video data presents three labels that are fire, non-fire, and ignore. The image data is collected from four datasets: Flickr-FireSmoke [163] (5556 images: 527 fire/smoke images, 1077 fire images, 369 smoke images, and 3583 non-fire/smoke images), Flickr-Fire [163] (2000 images: 1000 fire images and 1000 non-fire images), BowFire, and SmokeBlock [164,165] (1666 images: 832 smoke images and 834 non-smoke images). Figure 14 presents FiSmo fire detection dataset samples. FiSmo is comprised of forest and non-forest images, but

it should be noted that the non-forest images are generally removed to improve the efficiency of the wildfire classification models.

Figure 14. Examples from the FiSmo dataset. (**Top**): Fire images; (**Bottom**): non-fire images.

- **DeepFire** dataset [64,162] was developed to address the problem of wildland fire recognition. It comprises RGB aerial images with a resolution of 250 × 250 pix. downloaded from various research sites using many keywords such as forest, forest fires, mountain, and mountain fires, as depicted in Figure 15. It includes a total of 1900 images, where 950 images belong to the fire incident and 950 images remain to the non-fire incident.

Figure 15. DeepFire dataset example. (**Top**): Fire images; (**Bottom**): non-fire images.

- The **FIRE** dataset is a public dataset developed by Saeid et al. [69] during the NASA Space Apps Challenge in 2018 for the fire recognition task. It comprises two folders (fireimages and non-fireimages). The first folder consists of 755 fire images with

various resolutions, some of which include dense smoke. The second consists of 244 non-fire images such as animals, trees, waterfalls, rivers, grasses, people, roads, lakes and forests. Figure 16 presents some examples of FIRE dataset.

Figure 16. FIRE dataset example. (**Top**): Fire images; (**Bottom**): non-fire images.

- **FLAME2** dataset [72,73] represents public wildfire data collected in November 2021 during a prescribed fire in an open canopy pine forest in Northern Arizona. It contains IR/RGB images and videos recorded with a Mavic 2 Enterprise Advanced dual RGB/IR camera. It is labeled by two human experts. It contains 53,451 RGB images (25,434 Fire/Smoke images, 14,317 Fire/non-smoke images, and 13,700 non-fire/non-smoke images) with a resolution of 254 × 254 pix. extracted from seven pairs of RGB videos with a resolution of 1920 × 1080 pix. or 3840 × 2160 pix. It also includes seven IR videos with a resolution of 640 × 512 pix. Figure 17 shows some examples of FLAME2 dataset.
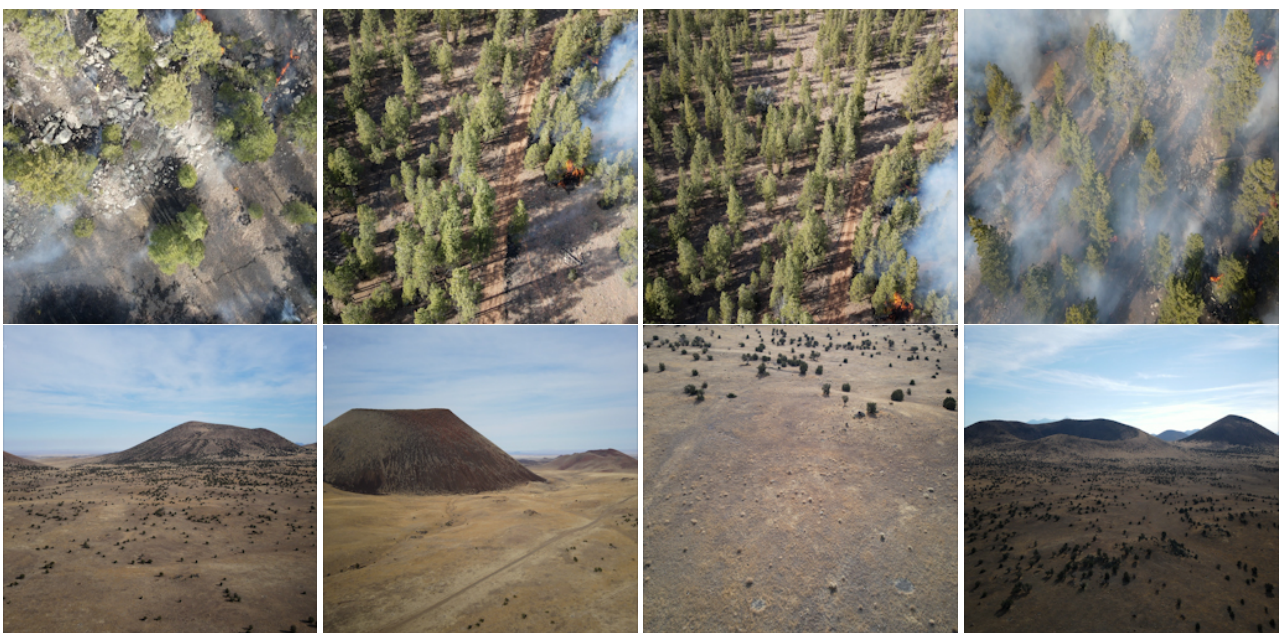
Figure 17. FLAME2 dataset example. (**Top**): Fire images; (**Bottom**): non-fire images.

6. Discussion

In this section, we discuss two main parts related specifically to wildland fire (not forest smoke) classification, detection, and segmentation tasks: data preprocessing and model results. The first section deals with the problem of collecting and preprocessing wildfire data, in which we discuss the challenges faced by researchers. The second section is related to the interpretation of the performance achieved by DL including vision transformers.

6.1. Data Collection and Preprocessing

Large datasets are required to train deep learning algorithms. Several datasets are available to researchers to facilitate the technical evaluation of deep learning models. For forest fire recognition, detection, and segmentation problems, the lack of public datasets makes their training and their evaluation a challenging task. Table 4 presents the popularly used datasets for wildland fire classification, detection, and segmentation tasks. In [45,47,72], authors collected images and videos to build FLAME [47], FLAME2 [72], and CorsicanFire [45] datasets, which are designed specifically for forest fires. CorsicanFire contains 1135 images with their binary mask representing the segmented wildfire areas. FLAME and FLAME2 contain aerial images and videos collected via drones. FLAME consists of 48,010 images for the wildfire classification task. It also includes 2003 images with their corresponding fire mask. FLAME2 includes 53,451 images for the forest fire classification task. To address the lack of public wildland fire datasets, authors in [38,43,50,70,77,81,85,88,135] used images very close to real wildland fires as well as images downloaded from the web or combined multiple datasets, including examples of fires unrelated to the forest context such as fires in urban scenarios, industrial fires, and fires in an indoor environment to train their proposed models.

On the other hand, data augmentation techniques are widely used to increase the diversity of the learning data and improve the performance of deep learning models. It allows applying transformations on images such as mosaic data augmentation, image occlusion methods (cutout, grid mask, and mix-up), photometric transformation (contrast, brightness, and shear), and geometric transformations (flip, rotation, and cropping). For the forest fire problem, the adopted data augmentation methods are chosen based on the characteristics of the fire images. For example, color space-based data augmentation methods are not suitable for the wildfire problem because they may result in the loss of some wildfire information. Table 5 presents an overview of data augmentation techniques employed in the wildfire classification, segmentation, and detection tasks. Geometric transformations (cropping, horizontal/vertical flipping, rotation, translation, scaling, symmetry, horizontal/vertical reflection, left/right reflection, and zooming) are the most commonly used in these tasks. For instance, Ghosh and Kumar [67] used three data augmentation techniques (horizontal flipping, vertical flipping, and zooming) to augment the training data, resulting in three times more images than the original learning image set. Zhang et al. [63] used different data augmentation techniques (mix-up, rotation, and flip) for generating 184,499 new images. Park et al. [53] utilized horizontal flipping, rotation, zoom and brightness, as well as CycleGAN to generate new wildfire data while respecting its diversity and particularities. CycleGAN involves the automatic learning of image-to-image translation models without requiring paired samples. It applies cycle coherence, where generative networks learn to translate images between domains, ensuring that the translated images still keep coherence with the original images. These techniques resulted in a total of 3585 new generated images.

Table 5. Overview of data augmentation techniques

Task	Ref	Data Augmentation Techniques
Wildfire Classification	[38]	Crop, horizontal/vertical flip
	[49]	Crop, rotation
	[51]	Crop
	[53]	Horizontal flip, rotation, zoom rotation, brightness, CycleGAN
	[56]	Shift, rotation, flip, blur, varying illumination intensity
	[58]	Crop, horizontal flip
	[60]	Rotation, shear, zoom, shift
	[62]	Horizontal flip, rotation
	[63]	Mix-up, rotation, flip
	[66]	Rotation,horizontal/vertical mirroring, Gaussian blur, pixel level augmentation
	[67]	Horizontal/vertical flip, zoom
Wildfire Detection	[84]	Translation, image scale, mosaic, mix-up, horizontal flip
Wildfire segmentation	[60,122,127,137]	Horizontal flip, rotation
	[116,126]	Left/right symmetry
	[131]	Translation, rotation, horizontal/vertical reflection, left/right reflection
	[134]	Flip, rotation, crop, noise

6.2. Model Results Discussion

The performance of deep learning models evaluates the ability of these models to correctly classify, detect, or segment objects. In a wildland fire classification task, the performance of the model can be evaluated based on its ability to classify wildfires. In a wildland fire segmentation task, the accuracy of the DL models can be evaluated on their effectiveness in identifying and detecting the visible surface of fire in the input images. In the wildland fire detection task, the performance of the wildfire detection models can be discussed based on their potential to detect and localize wildfires in the input images. Different metrics and datasets are used to evaluate wildfire models, making their comparison a bit challenging. Vison-based ML models are among the first methods used for the recognition, detection, and segmentation of forest fires. They obtained a great performance. However, some limitations still exist, such as slow time response and false alarms, especially when the distance to the wildfire is large or the area of the wildfire is small. To overcome these drawbacks, deep learning models are proposed. They automatically learn high and low-level features. These models showed higher performances than ML methods in wildfire classification, detection, and segmentation tasks on aerial and terrestrial images. As an example, MobileNet v2 [56], InceptionResNet v2 [71], and ResNet50 [66] showed excellent performance in wildland fire classification, reaching an accuracy of 99.70%, 99.90%, and 99.94%, respectively. In the wildland fire detection task, excellent results were obtained with SSD [78], Yolo v4 [82], and ensemble learning [84], which integrates Yolo v5 and U-Net by achieving an accuracy of 99.88%, 99.35%, 99.60%, respectively. Vision Transformer,TransUNet, [60,137] obtained excellent performance in the wildland fire segmentation task, achieving F1 scores of 97.70% and 99.90% using ground and aerial images, respectively. However, a large number of labeled images are required for their training. Transfer learning technique is used to repurpose a pretrained model on very large data, such as the COCO dataset, to solve a new task (forest fires in our case). The main idea is to finetune the parameters of the pretrained model to avoid overfitting caused by the small amount of training data. These pretrained DL models showed excellent potential in wildfire tasks. For example, Lee et al. [38] used pretrained AlexNet, GoogleNet, and VGG13 to identify forest fires on aerial images, achieving an interesting accuracy of 99%. Yandouzi et al. [66] presented a forest fire classification benchmark using eleven pretrained models (VGG16, ResNet50, MobileNet, VGG19, NASNetMobile, InceptionResNet v2, Xception, Inception v3, ResNet50 v2, DenseNet, and MobileNet v2). Experimental results showed that ResNet50 and VGG16 are more suitable for wildfire recognition, with accuracies higher than 99.9%.

Wildfire classification, detection, and segmentation models based on the convolution operator showed higher performance. However, they remain limited in modeling the global context. Several studies applied the attention mechanism, which determines global dependencies to solve the forest fires problem. As an example, Lin et al. [89] proposed a forest fire detection method, STPM_SAHI, which incorporates the attention module. STPM_SAHI achieved an average accuracy of 89.4%, outperforming existing one-stage detectors (Yolo v5 and EfficientDet). Zhang et al. [135] incorporated the attention gate module into the Wildfire semantic segmentation method, ATT Squeeze U-Net, to remove irrelevant and noisy characteristics transmitted by skip connections. ATT Squeeze U-Net improved the accuracy of U-Net by 2.15%. Guan et al. [74] designed a novel method, DSA-ResNet, by adding an attention module to the ResNet method for identifying wildfire on aerial images. Testing results showed an interesting accuracy of 93.65% improving the performance of popular CNNs, VGGNet, GoogLeNet, ResNet, and SE-ResNet by 8.79%, 5.42%, 2.37%, and 1.19%, respectively. In [60,137], the vision Transformer, TransUNet, which adopted the self-attention module is used in segmenting forest fire regions on aerial and ground images. It achieved excellent performance, surpassing existing wildfire segmentation models such as EfficientSeg [60,137].

Early forest fire detection is very important to detect wildfires and reduce their damage. The inference time of fire detection is also employed to receive real-time information about the occurrence of fires and reduce the area burned. Table 6 presents the overview of the inference time used in the wildfire classification, detection, and segmentation tasks. Different GPUs are used to calculate the inference time, which makes their comparison difficult. As an example, the one-stage detector shows fast detection time for wildland fire detection. Yolo v4 with MobileNet V3 showed a faster inference time of 19.76 FPS, which is suitable for real-time wildfire detection [82]. In [77], a new forest fire detection method based on Yolo v3 achieved high accuracy and a detection rate of 3.2 FPS, confirming its ability and feasibility for fire monitoring tasks in a UAV platform. In [135], ATT Squeeze U-Net also obtained a reliable segmentation time of 0.65 FPS slightly slower than the binary segmentation model, U-Net due to the addition of an attention module.

Table 6. Overview of inference time

Task	Ref	Methodolgy	Configuration	Time (FPS)
Wildfire Classification	[38]	GoogLeNet	3 NVIDIA GTX Titan X GPUs	24.79
	[60]	EfficientNet-B5, DenseNet201	NVIDIA Geforce RTX 2080Ti GPU	55.55
	[63]	FT-ResNet50	NVIDIA GeForce RTX 2080Ti GPU	18.10
Wildfire Detection	[77]	Modified Yolo v3	Drone with NVIDIA 4-Plus-1 ARM Cortex-A15	3.20
	[81]	Yolo v5, EfficientDet, EfficientNet	NVIDIA GTX 2080Ti GPU	14.97
	[82]	Yolo v4 with MobileNet v3	NVIDIA Jetson Xavier NX GPU	19.76
	[86]	Yolo v5, CBAM, BiFPN, SPPFP	NVIDIA GeForce GTX 1070 GPU	44.10
	[87]	FCDM	NVIDIA GeForce RTX 3060 GPU	64.00
	[89]	STPM_SAHI	NVIDIA RTX 3050Ti GPU	19.22
Wildfire Segmentation	[60]	TransUNet	NVIDIA V100-SXM2 GPU	1.96
		TransFire		1.00
	[137]	TransUNet	NVIDIA Geforce RTX 2080Ti GPU	0.83
		MedT		0.37
	[112]	SFEwAN-SD	NVIDIA GTX 970 MSI GPU	25.64
	[121]	wUUNet	NVIDIA RTX 2070 GPU	63.00
	[127]	Deep-RegSeg	NVIDIA Tesla T4 GPU	6.25
	[131]	DeepLab v3+	NVIDIA GeForce RTX 3090 GPU	0.98
	[133]	FireDGWF	2 NVIDIA GTX 1080Ti GPUs	6.62
	[134]	U-Net	NVIDIA GeForce RTX 2080Ti GPU	1.22
		DeepLab v3+		1.47
		FCN		2.33
		PSPNet		2.04
	[135]	ATT Squeeze U-Net	NVIDIA GeForce GTX 1070 GPU	0.65
	[138]	Improved DeepLab v3+ with MobileNet v3	NVIDIA RTX 2080 Ti GPU	24.00

7. Conclusions

In this paper, a comprehensive literature review only of wildland fire (not wildland smoke) classification, detection, and segmentation techniques based on recent deep learning models including vision transformers has been presented. The latter proved their potential in recognizing and detecting wildfires, especially small fire areas, as well as in segmenting fire regions overcoming challenging problems such as the presence of smoke and the background complexity using aerial and ground images. We also presented the popularly used datasets for these tasks. Finally, We discussed the data preprocessing and model interpretation ability related to wildfire recognition, detection, and segmentation tasks. Our discussion highlights how deep learning approaches including vision transformers perform better than traditional methods, showing their reliability and potential in these tasks, as well as their main research limitations such as data collection and labeling.

Deep learning including vision transformers models showed excellent performance in the classification, detection, and segmentation of wildfire areas, even in the presence of smoke and wildfire-like objects. However, some challenges remain for future research, such as real-time monitoring and the scarcity of labeled datasets. Using a virtual simulation platform to process the extent of wildfires and predict their spread can be an excellent solution to the challenge of limited labeled data. In addition, using drones with lightweight DL models as well as integrating distance and telemetry sensors into ground fire camera systems can provide real-time information to monitor wildfires, by giving the location and distance of the fire.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicles
DL	Deep Learning
ML	Machine Learning
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit
PReLU	Parametric ReLU
LReLU	Leaky ReLU
DCNN	Deep Convolutional Neural Network
LBP	Local Binary Patterns
CycleGAN	Cycle-consistent Generative Adversarial Network
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
NCA	Neighborhood Component Analysis
RNN	Recurrent Neural Network
Yolo	You only look once
AP	Average Precision
SSD	Single Shot MultiBox Detector
mAP	mean Average Precision
PANet	Path Aggregation Network

FPN	Feature Pyramid Network
CSPNet	Cross Stage Partial Network
BiFPN	Bi-directional Feature Pyramid Network
RPN	Region Proposal Network
FP	False Positive rate
ASPP	Atrous Spatial Pyramid Pooling
IR	Infrared
FPS	Frames per second
MedT	Medical Transformer
BowFire	Best of both worlds Fire detection
FLAME	Fire Luminosity Airborne-based Machine learning Evaluation
NIR	Near infrared

References

1. 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]
2. Çelik, T.; Demirel, H. Fire detection in video sequences using a generic color model. *Fire Saf. J.* **2009**, *44*, 147–158. [CrossRef]
3. Toulouse, T.; Rossi, L.; Celik, T.; Akhloufi, M. Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods. *Signal Image Video Process.* **2016**, *10*, 647–654. [CrossRef]
4. Rossi, L.; Molinier, T.; Akhloufi, M.; Tison, Y.; Pieri, A. A 3D vision system for the measurement of the rate of spread and the height of fire fronts. *Meas. Sci. Technol.* **2010**, *21*, 105501. [CrossRef]
5. Rossi, L.; Akhloufi, M. Dynamic Fire 3D Modeling Using a Real-Time Stereovision System. In Proceedings of the Technological Developments in Education and Automation, Barcelona, Spain, 5–7 July 2010; pp. 33–38.
6. Cruz, H.; Eckert, M.; Meneses, J.; Martínez, J.F. Efficient Forest Fire Detection Index for Application in Unmanned Aerial Systems (UASs). *Sensors* **2016**, *16*, 893. [CrossRef] [PubMed]
7. Mueller, M.; Karasev, P.; Kolesov, I.; Tannenbaum, A. Optical Flow Estimation for Flame Detection in Videos. *IEEE Trans. Image Process.* **2013**, *22*, 2786–2797. [CrossRef]
8. Dimitropoulos, K.; Barmpoutis, P.; Grammalidis, N. Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 339–351. [CrossRef]
9. Foggia, P.; Saggese, A.; Vento, M. Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 1545–1556. [CrossRef]
10. 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), Hammamet, Tunisia, 20–22 December 2018; Volume 1, pp. 332–340.
11. 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]
12. Mahmoud, M.A.I.; Ren, H. Forest fire detection and identification using image processing and SVM. *J. Inf. Process. Syst.* **2019**, *15*, 159–168. [CrossRef]
13. Van Hamme, D.; Veelaert, P.; Philips, W.; Teelen, K. Fire detection in color images using Markov random fields. In Proceedings of the Advanced Concepts for Intelligent Vision Systems, Sydney, Australia, 13–16 December 2010; pp. 88–97.
14. Bedo, M.V.N.; de Oliveira, W.D.; Cazzolato, M.T.; Costa, A.F.; Blanco, G.; Rodrigues, J.F.; Traina, A.J.; Traina, C. Fire detection from social media images by means of instance-based learning. In Proceedings of the Enterprise Information Systems, Barcelona, Spain, 7–9 October 2015; pp. 23–44.
15. Ko, B.; Cheong, K.H.; Nam, J.Y. Early fire detection algorithm based on irregular patterns of flames and hierarchical Bayesian Networks. *Fire Saf. J.* **2010**, *45*, 262–270. [CrossRef]
16. Ren, B. Neural Network Machine Translation Model Based on Deep Learning Technology. In Proceedings of the Application of Intelligent Systems in Multi-Modal Information Analytics, Online, 23 April 2022; pp. 643–649.
17. McCoy, J.; Rawal, A.; Rawat, D.B.; Sadler, B.M. Ensemble Deep Learning for Sustainable Multimodal UAV Classification. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–10. [CrossRef]
18. Zhang, Y.; Kwong, S.; Xu, L.; Zhao, T. Advances in Deep-Learning-Based Sensing, Imaging, and Video Processing. *Sensors* **2022**, *22*, 6192. [CrossRef] [PubMed]
19. Hazra, A.; Choudhary, P.; Sheetal Singh, M. Recent Advances in Deep Learning Techniques and Its Applications: An Overview. In Proceedings of the Advances in Biomedical Engineering and Technology, Werdanyeh, Lebanon, 7–9 October 2021; pp. 103–122.
20. Seo, P.H.; Nagrani, A.; Arnab, A.; Schmid, C. End-to-End Generative Pretraining for Multimodal Video Captioning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 17959–17968.
21. Wang, Y.; Yue, Y.; Lin, Y.; Jiang, H.; Lai, Z.; Kulikov, V.; Orlov, N.; Shi, H.; Huang, G. AdaFocus V2: End-to-End Training of Spatial Dynamic Networks for Video Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 24–28 June 2022; pp. 20062–20072.

22. Cui, J.; Qiu, H.; Chen, D.; Stone, P.; Zhu, Y. Coopernaut: End-to-End Driving With Cooperative Perception for Networked Vehicles. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–20 June 2022; pp. 17252–17262.

23. Ait Nasser, A.; Akhloufi, M.A. A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography. *Diagnostics* **2023**, *13*, 159. [CrossRef]

24. Mahoro, E.; Akhloufi, M.A. Applying Deep Learning for Breast Cancer Detection in Radiology. *Curr. Oncol.* **2022**, *29*, 8767–8793. [CrossRef]

25. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Net. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef]

26. 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]

27. Akhloufi, M.A.; Couturier, A.; Castro, N.A. Unmanned Aerial Vehicles for Wildland Fires: Sensing, Perception, Cooperation and Assistance. *Drones* **2021**, *5*, 15. [CrossRef]

28. 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]

29. Geetha, S.; Abhishek, C.; Akshayanat, C. Machine vision based fire detection techniques: A survey. *Fire Technol.* **2021**, *57*, 591–623. [CrossRef]

30. Bot, K.; Borges, J.G. A Systematic Review of Applications of Machine Learning Techniques for Wildfire Management Decision Support. *Inventions* **2022**, *7*, 15. [CrossRef]

31. Cruz, H.; Gualotuña, T.; Pinillos, M.; Marcillo, D.; Jácome, S.; Fonseca C., E.R. Machine Learning and Color Treatment for the Forest Fire and Smoke Detection Systems and Algorithms, a Recent Literature Review. In Proceedings of the Artificial Intelligence, Computer and Software Engineering Advances, Quito, Ecuador, 26–30 October 2020; pp. 109–120.

32. 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]

33. Liu, Y.H. Feature Extraction and Image Recognition with Convolutional Neural Networks. *J. Phys. Conf. Ser.* **2018**, *1087*, 062032. [CrossRef]

34. Hara, K.; Saito, D.; Shouno, H. Analysis of function of rectified linear unit used in deep learning. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 1–8.

35. 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.

36. Jin, X.; Xu, C.; Feng, J.; Wei, Y.; Xiong, J.; Yan, S. Deep Learning with S-Shaped Rectified Linear Activation Units. *AAAI Conf. Artif. Intell.* **2016**, *30*, 1737–1743. [CrossRef]

37. Boureau, Y.L.; Bach, F.; LeCun, Y.; Ponce, J. Learning mid-level features for recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2559–2566.

38. 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), Berlin, Germany, 3–6 September 2017; pp. 252–253.

39. 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.

40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 5–10 December 2012; pp. 1097–1105.

41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.

42. Zhao, Y.; Ma, J.; Li, X.; Zhang, J. Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery. *Sensors* **2018**, *18*, 712. [CrossRef]

43. Srinivas, K.; Dua, M. Fog Computing and Deep CNN Based Efficient Approach to Early Forest Fire Detection with Unmanned Aerial Vehicles. In *Inventive Computation Technologies*; Smys, S., Bestak, R., Rocha, Á., Eds.; Springer International Publishing: Coimbatore, India, 29–30 August 2019; pp. 646–652.

44. Wang, Y.; Dang, L.; Ren, J. Forest fire image recognition based on convolutional neural network. *J. Algorithms Comput. Technol.* **2019**, *13*, 1748302619887689. [CrossRef]

45. Toulouse, T.; Rossi, L.; Campana, A.; Celik, T.; Akhloufi, M.A. Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Saf. J.* **2017**, *92*, 188–194. [CrossRef]

46. 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.

47. 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. Net.* **2021**, *193*, 108001. [CrossRef]

48. 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.

49. Arteaga, B.; Diaz, M.; Jojoa, M. Deep Learning Applied to Forest Fire Detection. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 9–11 December 2020; pp. 1–6.

50. Rahul, M.; Shiva Saketh, K.; Sanjeet, A.; Srinivas Naik, N. Early Detection of Forest Fire using Deep Learning. In Proceedings of the IEEE region 10 conference (TENCON), Osaka, Japan, 16–19 November 2020; pp. 1136–1140.

51. Sousa, M.J.; Moutinho, A.; Almeida, M. Wildfire detection using transfer learning on augmented datasets. *Expert Syst. Appl.* **2020**, *142*, 112975. [CrossRef]

52. 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.

53. Park, M.; Tran, D.Q.; Jung, D.; Park, S. Wildfire-Detection Method Using DenseNet and CycleGAN Data Augmentation-Based Remote Camera Imagery. *Remote Sens.* **2020**, *12*, 3715. [CrossRef]

54. 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.

55. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232.

56. 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.

57. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

58. Tang, Y.; Feng, H.; Chen, J.; Chen, Y. ForestResNet: A Deep Learning Algorithm for Forest Image Classification. *J. Phys. Conf. Ser.* **2021**, *2024*, 012053. [CrossRef]

59. Dutta, S.; Ghosh, S. Forest Fire Detection Using Combined Architecture of Separable Convolution and Image Processing. In Proceedings of the 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 36–41.

60. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. *Sensors* **2022**, *22*, 1977. [CrossRef] [PubMed]

61. 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.

62. 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, Xi'an, China, 12–14 March 2021; pp. 6–7.

63. Zhang, L.; Wang, M.; Fu, Y.; Ding, Y. A Forest Fire Recognition Method Using UAV Images Based on Transfer Learning. *Forests* **2022**, *13*, 975. [CrossRef]

64. Khan, A.; Hassan, B.; Khan, S.; Ahmed, R.; Abuassba, A. DeepFire: A Novel Dataset and Deep Transfer Learning Benchmark for Forest Fire Detection. *Mob. Inf. Syst.* **2022**, *2022*, 5358359. [CrossRef]

65. Dogan, S.; Datta Barua, P.; Kutlu, H.; Baygin, M.; Fujita, H.; Tuncer, T.; Acharya, U. Automated accurate fire detection system using ensemble pretrained residual network. *Expert Syst. Appl.* **2022**, *203*, 117407. [CrossRef]

66. Yandouzi, M.; Grari, M.; Idrissi, I.; Boukabous, M.; Moussaoui, O.; Ghoumid, K.; Elmiad, A.K.E. Forest Fires Detection using Deep Transfer Learning. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2022**, *13*, 268–275. [CrossRef]

67. Ghosh, R.; Kumar, A. A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire. *Multimed. Tools Appl.* **2022**, *81*, 38643–38660. [CrossRef]

68. Vento, M.; Foggia, P.; Tortorella, F.; Percannella, G.; Ritrovato, P.; Saggese, A.; Greco, L.; Carletti, V.; Greco, A.; Vigilante, V.; et al. MIVIA Fire/Smoke Detection Dataset. Available online: https://mivia.unisa.it/datasets/video-analysis-datasets/ (accessed on 5 January 2023).

69. Saied, A. Fire Dataset. Available online: https://www.kaggle.com/datasets/phylake1337/fire-dataset?select=fire_dataset%2C+06.11.2021 (accessed on 5 January 2023).

70. Zheng, S.; Gao, P.; Wang, W.; Zou, X. A Highly Accurate Forest Fire Prediction Model Based on an Improved Dynamic Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 6721. [CrossRef]

71. K. Mohammed, R. A real-time forest fire and smoke detection system using deep learning. *Int. J. Nonlinear Anal. Appl.* **2022**, *13*, 2053–2063. [CrossRef]

72. 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]

73. Hopkins, B.; O'Neill, L.; Afghah, F.; Razi, A.; Rowell, E.; Watts, A.; Fule, P.; Coen, J. FLAME2 Dataset. Available online: https://dx.doi.org/10.21227/swyw-6j78 (accessed on 5 January 2023).

74. Guan, Z.; Miao, X.; Mu, Y.; Sun, Q.; Ye, Q.; Gao, D. Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model. *Remote Sens.* **2022**, *14*, 3159. [CrossRef]

75. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

76. 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]

77. 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 Yolo v3. In Proceedings of the 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 22–26 July 2019; pp. 1–5.

78. Wu, S.; Zhang, L. Using Popular Object Detection Methods for Real Time Forest Fire Detection. In Proceedings of the 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 8–9 December 2018; pp. 280–284.

79. Jiao, Z.; Zhang, Y.; Mu, L.; Xin, J.; Jiao, S.; Liu, H.; Liu, D. A Yolo v3-based Learning Strategy for Real-time UAV-based Forest Fire Detection. In Proceedings of the Chinese Control Furthermore, Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 4963–4967.

80. 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*, 166–179. [CrossRef]

81. Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A Forest Fire Detection System Based on Ensemble Learning. *Forests* **2021**, *12*, 217. [CrossRef]

82. Wang, S.; Zhao, J.; Ta, N.; Zhao, X.; Xiao, M.; Wei, H. A real-time deep learning forest fire monitoring algorithm based on an improved Pruned+ KD model. *J. Real-Time Image Process.* **2021**, *18*, 2319–2329. [CrossRef]

83. Kasyap, V.L.; Sumathi, D.; Alluri, K.; Reddy Ch, P.; Thilakarathne, N.; Shafi, R.M. Early detection of forest fire using mixed learning techniques and UAV. *Comput. Intell. Neurosci.* **2022**, *2022*, 3170244. [CrossRef]

84. Mseddi, W.S.; Ghali, R.; Jmal, M.; Attia, R. Fire Detection and Segmentation using Yolo v5 and U-Net. In Proceedings of the 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 741–745.

85. Zhao, L.; Zhi, L.; Zhao, C.; Zheng, W. Fire-YOLO: A Small Target Object Detection Method for Fire Inspection. *Sustainability* **2022**, *14*, 4930. [CrossRef]

86. Xue, Z.; Lin, H.; Wang, F. A Small Target Forest Fire Detection Model Based on Yolo v5 Improvement. *Forests* **2022**, *13*, 1332. [CrossRef]

87. Xue, Q.; Lin, H.; Wang, F. FCDM: An Improved Forest Fire Classification and Detection Model Based on Yolo v5. *Forests* **2022**, *13*, 2129. [CrossRef]

88. Barmpoutis, P.; Dimitropoulos, K.; Kaza, K.; Grammalidis, N. Fire Detection from Images Using Faster R-CNN and Multidimensional Texture Analysis. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8301–8305.

89. Lin, J.; Lin, H.; Wang, F. STPM_SAHI: A Small-Target Forest Fire Detection Model Based on Swin Transformer and Slicing Aided Hyper Inference. *Forests* **2022**, *13*, 1603. [CrossRef]

90. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

91. 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.

92. Redmon, J.; Farhadi, A. Yolo v3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.

93. Bochkovskiy, A.; Wang, C.; Liao, H.M. Yolo v4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.

94. Jocher, G.; Stoken, A.; Chaurasia, A.; Borovec, J.; Chanvichet, V.; Kwon, Y.; TaoXie, S.; Changyu, L.; Abhiram, V.; Skalski, P.; et al. Yolo v5. Available online: https://github.com/ultralytics/yolov5 (accessed on 5 January 2023).

95. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.

96. 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, 11–14 October 2016; pp. 21–37.

97. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimed. Tools Appl.* **2022**, *82*, 1–33. . [CrossRef]

98. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

99. 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, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.

100. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.

101. Pincott, J.; Tien, P.W.; Wei, S.; Kaiser Calautit, J. Development and evaluation of a vision-based transfer learning approach for indoor fire and smoke detection. *Build. Serv. Eng. Res. Technol.* **2022**, *43*, 319–332. [CrossRef]

102. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 390–391.

103. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. PANet: Few-Shot Image Semantic Segmentation With Prototype Alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9197–9206.

104. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]

105. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI, Munich, Germany, 5–9 October 2015; pp. 234–241.

106. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

107. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

108. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *arXiv* **2020**, arXiv:2010.04159.

109. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

110. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523–3542. [CrossRef]

111. Toulouse, T.; Rossi, L.; Akhloufi, M.; Celik, T.; Maldague, X. Benchmarking of wildland fire colour segmentation algorithms. *Iet Image Process.* **2015**, *9*, 1064–1072. [CrossRef]

112. Gonzalez, A.; Zuniga, M.D.; Nikulin, C.; Carvajal, G.; Cardenas, D.G.; Pedraza, M.A.; Fernandez, C.A.; Munoz, R.I.; Castro, N.A.; Rosales, B.F.; et al. Accurate fire detection through fully convolutional network. In Proceedings of the 7th Latin American Conference on Networked and Electronic Media (LACNEM), Valparaiso, Chile, 6–7 November 2017; pp. 1–6.

113. Frizzi, S.; Kaabi, R.; Bouchouicha, M.; Ginoux, J.M.; Moreau, E.; Fnaiech, F. Convolutional neural network for video fire and smoke detection. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 24–27 October 2016; pp. 877–882.

114. Wang, G.; Zhang, Y.; Qu, Y.; Chen, Y.; Maqsood, H. Early Forest Fire Region Segmentation Based on Deep Learning. In Proceedings of the Chinese Control Furthermore, Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 6237–6241.

115. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *arXiv* **2016**, arXiv:1602.07360.

116. Choi, H.S.; Jeon, M.; Song, K.; Kang, M. Semantic fire segmentation model based on convolutional neural network for outdoor image. *Fire Technol.* **2021**, *57*, 3005–3019. [CrossRef]

117. Quan, T.M.; Hildebrand, D.G.C.; Jeong, W.K. FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics. *Front. Comput. Sci.* **2021**, *3*, 34. [CrossRef]

118. Cazzolato, M.T.; Avalhais, L.P.S.; Chino, D.Y.T.; Ramos, J.S.; Souza, J.A.; Rodrigues-Jr, J.F.; Traina, A.J.M. FiSmo: A Compilation of Datasets from Emergency Situations for Fire and Smoke Analysis. In Proceedings of the SBBD2017—SBBD Satellite Events of the 32nd Brazilian Symposium on Databases—DSW (Dataset Showcase Workshop), Uberlandia, MG, Brazil, 2–5 October 2017; pp. 213–223.

119. Akhloufi, M.A.; Tokime, R.B.; Elassady, H. Wildland fires detection and segmentation using deep learning. In Proceedings of the Pattern Recognition and Tracking XXIX, Orlando, FL, USA, 18–19 April 2018; p. 106490B.

120. 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 Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Québec City, QC, Canada, 14 September 2017; pp. 240–248.

121. Bochkov, V.S.; Kataeva, L.Y. wUUNet: Advanced Fully Convolutional Neural Network for Multiclass Fire Segmentation. *Symmetry* **2021**, *13*, 98. [CrossRef]

122. 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.

123. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]

124. Yesilkaynak, V.B.; Sahin, Y.H.; Unal, G.B. EfficientSeg: An Efficient Semantic Segmentation Network. *arXiv* **2020**, arXiv:2009.06469.

125. Dzigal, D.; Akagic, A.; Buza, E.; Brdjanin, A.; Dardagan, N. Forest Fire Detection based on Color Spaces Combination. In Proceedings of the 11th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 28–30 November 2019; pp. 595–599.

126. Song, K.; Choi, H.S.; Kang, M. Squeezed fire binary segmentation model using convolutional neural network for outdoor images on embedded device. *Mach. Vis. Appl.* **2021**, *32*, 1–12. [CrossRef]

127. Ghali, R.; Akhloufi, M.A.; Souidene Mseddi, W.; Jmal, M. Wildfire Segmentation Using Deep-RegSeg Semantic Segmentation Architecture. In Proceedings of the 19th International Conference on Content-Based Multimedia Indexing, New York, NY, USA, 14–16 September 2022; pp. 149–154.

128. Harkat, H.; Nascimento, J.; Bernardino, A. Fire segmentation using a DeepLabv3+ architecture. In Proceedings of the Image and Signal Processing for Remote Sensing XXVI, Online, 21–25 September 2020; pp. 134–145.

129. 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]

130. Harkat, H.; Nascimento, J.M.; Bernardino, A. Fire Detection using Residual Deeplabv3+ Model. In Proceedings of the 2021 Telecoms Conference (ConfTELE), Leiria, Portugal, 11–12 February 2021; pp. 1–6.

131. Harkat, H.; Nascimento, J.M.P.; Bernardino, A.; Thariq Ahmed, H.F. Assessing the Impact of the Loss Function and Encoder Architecture for Fire Aerial Images Segmentation Using Deeplabv3+. *Remote Sens.* **2022**, *14*, 2023. [CrossRef]

132. 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]

133. Pan, J.; Ou, X.; Xu, L. A Collaborative Region Detection and Grading Framework for Forest Fire Smoke Using Weakly Supervised Fine Segmentation and Lightweight Faster-RCNN. *Forests* **2021**, *12*, 768. [CrossRef]

134. Wang, Z.; Peng, T.; Lu, Z. Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks. *Forests* **2022**, *13*, 1133. [CrossRef]

135. Zhang, J.; Zhu, H.; Wang, P.; Ling, X. ATT Squeeze U-Net: A Lightweight Network for Forest Fire Detection and Recognition. *IEEE Access* **2021**, *9*, 10858–10870. [CrossRef]

136. Niknejad, M.; Bernardino, A. Attention on Classification for Fire Segmentation. In Proceedings of the 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Virtually, 13–15 December 2021, pp. 616–621.

137. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Wildfire Segmentation Using Deep Vision Transformers. *Remote Sens.* **2021**, *13*, 3527. [CrossRef]

138. Li, M.; Zhang, Y.; Mu, L.; Xin, J.; Yu, Z.; Jiao, S.; Liu, H.; Xie, G.; Yingmin, Y. A Real-time Fire Segmentation Method Based on A Deep Learning Approach. *IFAC-PapersOnLine* **2022**, *55*, 145–150. [CrossRef]

139. Chen, L.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.

140. 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.

141. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.

142. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

143. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.

144. Huang, Z.; Huang, L.; Gong, Y.; Huang, C.; Wang, X. Mask Scoring R-CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6409–6418.

145. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.

146. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]

147. Gillioz, A.; Casas, J.; Mugellini, E.; Khaled, O.A. Overview of the Transformer-based Models for NLP Tasks. In Proceedings of the 15th Conference on Computer Science and Information Systems (FedCSIS), Sofia, Bulgaria, 6–9 September 2020; pp. 179–183.

148. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]

149. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning Texture Transformer Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5791–5800.

150. Ye, L.; Rochan, M.; Liu, Z.; Wang, Y. Cross-Modal Self-Attention Network for Referring Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 10502–10511.

151. 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.

152. 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.

153. Chino, D.Y.T.; Avalhais, L.P.S.; Rodrigues, J.F.; Traina, A.J.M. BoWFire: Detection of Fire in Still Images by Integrating Pixel Color and Texture Analysis. In Proceedings of the 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Brazil, 26–29 August 2015; pp. 95–102.

154. de Oliveira, W.D. BowFire Dataset. Available online: https://bitbucket.org/gbdi/bowfire-dataset/src/master/ (accessed on 5 January 2023).

155. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. FLAME Dataset. Available online: https://ieee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-UAVs (accessed on 5 January 2023).

156. Toulouse, T.; Rossi, L.; Campana, A.; Celik, T.; Akhloufi, M.A. CorsicanFire Dataset. Available online: https://feuxdeforet.universita.corsica/article.php?id_art=2133&id_rub=572&id_menu=0&id_cat=0&id_site=33&lang=en (accessed on 5 January 2023).

157. Li, S.; Yan, Q.; Liu, P. An Efficient Fire Detection Method Based on Multiscale Feature Extraction, Implicit Deep Supervision and Channel Attention Mechanism. *IEEE Trans. Image Process.* **2020**, *29*, 8467–8475. [CrossRef]

158. Li, S.; Yan, Q.; Liu, P. FD-Dataset. Available online: http://www.nnmtl.cn/EFDNet/ (accessed on 5 January 2023).

159. The University of Georgia's Center for Invasive Species and Ecosystem Health. ForestryImages Dataset. Available online: https://www.forestryimages.org/ (accessed on 5 January 2023).

160. Cetin, E. VisiFire Dataset. Available online: http://signal.ee.bilkent.edu.tr/VisiFire// (accessed on 5 January 2023).

161. Grammalidis, N.; Dimitropoulos, E.C.k. Firesense Dataset. Available online: https://zenodo.org/record/836749#.YumkVL2ZPIU/(accessed on 5 January 2023).

162. Khan, A.; Hassan, B.; Khan, S.; Ahmed, R.; Abuassba, A. DeepFire Dataset. Available online: https://www.kaggle.com/datasets/alik05/forest-fire-dataset (accessed on 5 January 2023).

163. Flickr Team. Flickr-FireSmoke and Flickr-Fire Datasets. Available online: https://www.flickr.com/ (accessed on 5 January 2023).

164. Cazzolato, M.T.; Bedo, M.V.N.; Costa, A.F.; de Souza, J.A.; Traina, C.; Rodrigues, J.F.; Traina, A.J.M. Unveiling Smoke in Social Images with the SmokeBlock Approach. In Proceedings of the 31st Annual ACM Symposium on Applied Computing, Pisa, Italy, 4–8 April 2016; pp. 49–54.

165. Cazzolato, M.T.; Bedo, M.V.N.; Costa, A.F.; de Souza, J.A.; Traina, C.; Rodrigues, J.F.; Traina, A.J.M. SmokeBlock Dataset. Available online: https://goo.gl/uW7LxW/ (accessed on 5 January 2023).