fire

MDPI

Review

Video Fire Detection Methods Based on Deep Learning: Datasets, Methods, and Future Directions

Chengtuo Jin [1], Tao Wang [2,3,*], Naji Alhusaini [2], Shenghui Zhao [2,4], Huilin Liu [1], Kun Xu [1] and Jin Zhang [5]

[1] School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China
[2] School of Computer and Information Engineering, Chuzhou University, Chuzhou 239000, China
[3] Unmanned Emergency Equipment and Digital Reconstruction of Disaster Process Joint Laboratory of Anhui Province, Chuzhou 239000, China
[4] Anhui Engineering Research Center of Intelligent Perception and Elderly Care, Chuzhou 239000, China
[5] School of Economics and Management, Anhui University of Science and Technology, Huainan 232001, China
* Correspondence: wtchzu@163.com

Abstract: Among various calamities, conflagrations stand out as one of the most-prevalent and -menacing adversities, posing significant perils to public safety and societal progress. Traditional fire-detection systems primarily rely on sensor-based detection techniques, which have inherent limitations in accurately and promptly detecting fires, especially in complex environments. In recent years, with the advancement of computer vision technology, video-oriented fire detection techniques, owing to their non-contact sensing, adaptability to diverse environments, and comprehensive information acquisition, have progressively emerged as a novel solution. However, approaches based on handcrafted feature extraction struggle to cope with variations in smoke or flame caused by different combustibles, lighting conditions, and other factors. As a powerful and flexible machine learning framework, deep learning has demonstrated significant advantages in video fire detection. This paper summarizes deep-learning-based video-fire-detection methods, focusing on recent advances in deep learning approaches and commonly used datasets for fire recognition, fire object detection, and fire segmentation. Furthermore, this paper provides a review and outlook on the development prospects of this field.

Keywords: fire recognition; fire object detection; fire segmentation; deep learning

Citation: Jin, C.; Wang, T.; Alhusaini, N.; Zhao, S.; Liu, H.; Xu, K.; Zhang, J. Video Fire Detection Methods Based on Deep Learning: Datasets, Methods, and Future Directions. *Fire* **2023**, *6*, 315. https://doi.org/10.3390/fire6080315

Academic Editor: Tao Chen

Received: 30 June 2023
Revised: 27 July 2023
Accepted: 1 August 2023
Published: 14 August 2023

1. Introduction

Throughout human history, various disasters have constantly affected and threatened humanity and society due to changes in the environmental conditions for human survival and production. When fire is considereda hazard factor, the natural disaster phenomenon caused by the uncontrolled spread of fire in time or space is called a fire disaster. Among various disasters, fire is among the most-common and -significant threats to public safety and social development. Forests and urban buildings serve as carriers that are affected by and damaged by fire disasters and threaten human society. In recent years, due to the continuous development of the social economy, the scale and quantity of urban high-rise buildings have been increasing. According to data from the Fire Rescue Bureau of China's Ministry of Emergency Management, the number of reported fires in mainland China in 2022 was 825,000, a year-on-year increase of 7.8%. The direct property loss caused by fires was CNY 7.16 billion, a year-on-year increase of 1.2% in deaths. The majority of fatal fires occurred in residential areas [1]. Regarding forest fires, due to their strong destructive power, they can cause the extinction of many forest species, cause soil erosion, and pose a catastrophic ecological threat to people's lives and property safety. In 2022, there were 709 forest fires in mainland China, causing 17 deaths [2]. Due to global warming, extreme weather, such as heat waves and droughts, has become more frequent. The data show that

the overall form of fires in China is still severe, there are still hidden dangers in fire safety, and major and extraordinary fires still occur occasionally.

Fire safety is an integral component of the national emergency management system and necessitates continual capability modernization. In recent years, efforts to enhance the pre-fire risk prevention capabilities have been consistently strengthened. Among these, accurate, effective, and timely fire detection plays a pivotal role in initiating prompt fire-fighting measures. In order to minimize potential loss of life and property, it is imperative to identify the source of fire at its nascent stage, fortify early warning systems in affected areas, and promptly implement measures to prevent the spread of flames.

Fire detection generally encompasses flame detection and smoke detection. Current fire-detection approaches for urban buildings primarily rely on conventional sensor perception, including thermal detection (e.g., fixed-temperature detectors), chemical compound smoke detection (e.g., ionization and gas sensors), and optical radiation detection (e.g., ultraviolet and infrared sensors) [3]. When a fire occurs, the characteristic signals of mass flow, such as combustible gases, smoke particles, and aerosols, as well as energy flow signals, such as flame radiation, can be sensed by sensors and transformed into easily processed physical quantities. Through signal-processing methods, alarm operations can be realized. Traditional sensor-based fire-detection systems are ideal for small indoor areas such as homes and offices due to their low requirements for ambient light and high sensitivity. However, these traditional systems are not acceptable for large open environments since they can only detect smoke and flames through ionization-generated particles when the sensors are close to the fire source. This proximity limitation can lead to transmission delays, creating a risk that the fire at the scene could propagate and elude effective control measures. Moreover, these sensors cannot provide information regarding the initial location of the fire, its propagation direction, its scale, smoke spread direction, the growth rate, and other factors essential for monitoring fire evolution [4].

Satellite remote sensing technology is currently the primary means of wildfire monitoring. Its fundamental principle lies in identifying and monitoring fire spots by utilizing the electromagnetic radiation characteristics released during biomass combustion. This technology boasts wide-ranging fire detection, high resolution, and timely response to dynamic changes, particularly playing an increasingly crucial role in the detection of wildfires in landscapes and the prediction of potential hazards [5]. In recent years, the utilization of optical data provided by satellite sensors such as HuanJing (HJ)-1B—WVC/IRMSS, Terra/Aqua-MODIS, and Himawari-8/9—AHI-8 has become a vital decision-making basis for early wildfire warning, emergency resource allocation, and post-disaster dynamic assessment [6,7]. However, the methods based on satellite remote sensing technology for wildfire detection also possess certain limitations. Cloud cover and dense smoke generated by the combustion may obscure the wildfire area, affecting the quality of satellite images and, thus, reducing the accuracy of fire monitoring. Meeting the real-time monitoring demands for wildfires necessitates satisfying the requirements of both temporal and spatial resolution in ground-observation satellites simultaneously.

Over the past few years, the continuous enhancement of smart cities and emergency construction as part of various countries' governmental policies has led to the wide deployment of video surveillance systems in diverse production and living environments. Leveraging existing video surveillance resources for fire detection demonstrates high feasibility and application value. Researchers in this field have made substantial progress in computer-vision-based fire-detection systems to overcome the limitations of sensor-based traditional fire-detection systems. Video-based fire-detection technology offers rapid response, strong anti-interference capabilities, and low costs. Moreover, it provides comprehensive and intuitive feedback on fire scene information, which can be promptly transmitted to monitoring centers through network transmission technology, offering valuable guidance for fire emergency response. Therefore, computer-vision-based fire-detection systems, compared to conventional detection systems, can issue early warning signals more quickly, affording individuals more time for evacuation and fire-extinguishing efforts.

Recently, numerous computer-vision-based fire-detection systems have been developed. These systems mainly fall into three categories: recognition, object detection, and segmentation. Additionally, from an algorithmic standpoint, they can be classified into two types: approaches based on handcrafted feature extraction and deep-learning-based approaches.

Deep-learning-based fire-detection approaches have become mainstream in this field. With the continuous expansion of the data volume, advancements in big data technology, and the remarkable performance of hardware such as GPUs, deep learning has experienced a surge of research interest. It has achieved impressive results in various domains, surpassing traditional pattern-recognition approaches in image classification and object detection. It has been successfully applied in multiple industry sectors, including autonomous driving [8], smart agriculture [9], healthcare [10], the Industrial IoT [11,12], sentiment analysis [13], and conversational systems [14]. Deep learning enables end-to-end learning, eliminating the need for handcrafted feature extraction and reducing reliance on prior knowledge. Deep-learning-based fire-detection approaches require abundant and diverse training samples to train deep neural networks, selecting fire-related information from low-level to high-level features, thereby achieving accurate detection of fire at different granularities. Figure 1 illustrates the mainstream deep-learning-based video-fire-detection framework.

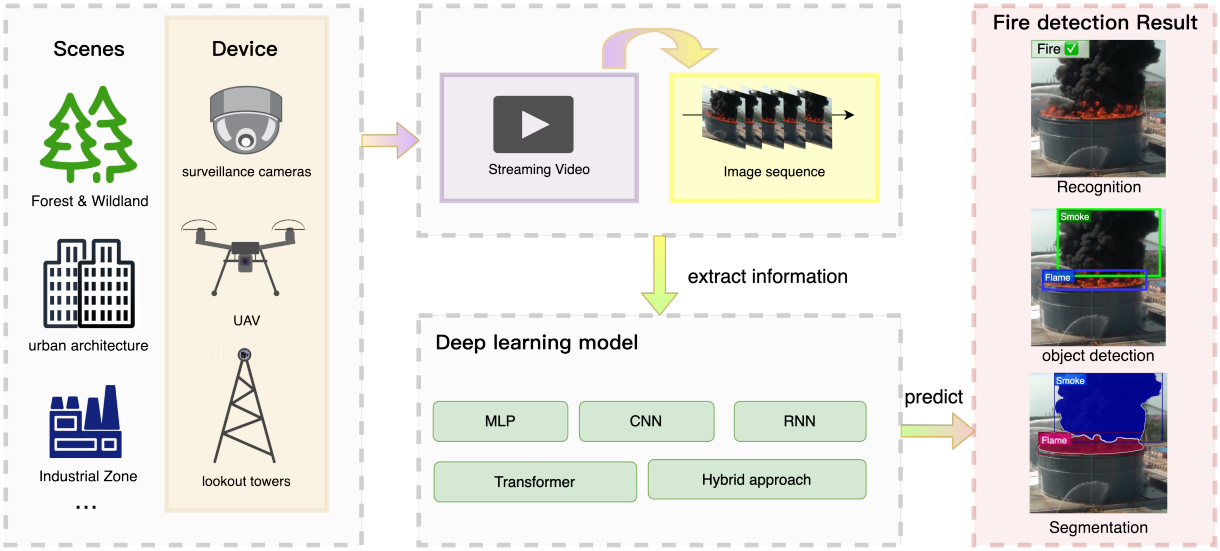

Figure 1. The framework of video-fire-detection methods based on deep learning. The process commences by collecting video sequences from diverse camera devices and extracting pertinent information from these sequences. Subsequently, this extracted information is fed into a deep learning model network for fire detection, ultimately leading to recognition, object detection, or segmentation outcomes.

The main contributions of this paper are as follows:

- We aimed to explore and analyze current advanced approaches used in video-based fire detection and their associated systems. We discuss the challenges and opportunities in designing and developing deep learning approaches for fire detection, focusing on recognition, object detection, and segmentation;
- We present the most-widely used public datasets for the fire-recognition, fire-object-detection, and fire-segmentation tasks;
- We discuss various challenges and potential research directions in this field.

The rest of the paper is organized as follows: Section 2 provides an overview of the two categories of video-based fire-detection approaches: approaches based on handcrafted feature extraction and approaches based on deep learning. Section 3 introduces commonly used datasets and evaluation metrics for these tasks. Section 4 reviews the methods' models based on deep learning for fire recognition, fire object detection, and fire segmentation. Section 5 discusses the main challenges of deep-learning-based video-based fire-detection methods. Finally, in Section 6, we conclude the paper.

2. Background and Related Work

From the perspective of algorithm types, one can typically classify video-fire-detection approaches into two categories: approaches based on handcrafted feature extraction and approaches based on deep learning.

For approaches based on handcrafted feature extraction, research on their features mainly focuses on static and dynamic characteristics. In videos, flame characteristics typically manifest as sustained burning in shades of orange-red, emitting heat and brightness. On the other hand, smoke commonly appears as white, gray, or black feather-like plumes composed of tiny particles of smoke or combustion. Under the influence of rising hot air, smoke swiftly moves within the environment. The shapes, densities, and colors of flames and smoke vary depending on the size of the fire source, the type of combustible materials, and the environmental conditions [15].

Static features of flames include color and appearance, while static features of smoke include color, texture, and blur level. Relevant studies use probability density functions, different color spaces, and texture analysis. For example, Kong et al. [16] subtracted the difference between flame and background colors to obtain candidate flame regions and used features such as area and color along with logistic regression to determine the probability of a region proposal being a flame. Filonenko et al. [17] proposed a probabilistic smoke-detection method using the RGB and HSV color spaces. However, this method requires good video image quality and cannot effectively handle smoke detection at night.

In the other category of approaches, the motion characteristics of fire are utilized. Dynamic features of flames include flickering, shape changes, and area variations, while dynamic features of smoke include motion direction and contour changes. Due to the movement characteristics of smoke and fire, these methods aim to extract the moving components that may contain targets in the image's foreground. Unlike flames, when the environmental visibility is good, the camera can easily capture smoke from a distance in the initial minutes of a fire due to its upward movement [18]. In these methods, motion values, direction, and energy have shown promising results under certain conditions. For example, Ye et al. [19] proposed a method that utilizes motion characteristics and extracts smoke and fire from the current frame of region proposals using adaptive background subtraction. Motion blobs are classified through spatiotemporal wavelet analysis, Weber contrast analysis, and color segmentation, resulting in high smoke- and flame-detection rates. As the fire situation evolves, under the influence of heat, the overall motion direction of smoke tends to be upward and gradually spreads outward, and the high-frequency signal of the smoke gradually decays compared to the background image's edge information. In order to address the issue of high false favorable rates resulting from the use of static features alone, the objective of reducing false positives can be achieved by incorporating both static and dynamic features. Lin et al. [20] proposed a smoke-detection method based on irregular motion region's dynamic texture, which utilizes the inter-frame information of smoke video sequences and describes it as using LBP dynamic texture descriptors. The authors analyzed the sliding window block feature-extraction method, which is greatly affected by factors such as the block size and motion scale coefficient. They designed a dynamic texture feature extraction algorithm for irregular regions, significantly reducing smoke-detection false alarm rates while ensuring a high detection rate.

In general, fire recognition using approaches based on handcrafted feature extraction involves extracting suspected smoke or flame regions from images or videos through

techniques such as sliding windows or segmentation. Then, features such as color and texture are extracted using SIFT, HOG, and LBP and transformed into feature vectors. Finally, classifiers such as SVM and Adaboost are employed to classify the feature vectors and determine the presence of smoke or flame. This approach has the advantage of utilizing handcrafted, designed features to describe the characteristics of smoke or flame, thereby improving detection accuracy. However, approaches based on handcrafted feature extraction heavily rely on the expertise and extensive experimentation of the designer. They struggle to cope with variations in smoke or flame caused by different combustible materials, lighting conditions, and airflow. These methods are time-consuming, prone to false alarms, lack real-time capability, and have limitations in long-range detection [21].

As a new technology in computer vision and supervised learning, deep learning has brought great hope to fire detection. Fire-detection approaches based on deep learning largely avoid the reliance on handcraft processes and can automatically extract high-level features that are difficult to obtain with traditional techniques, enabling accurate recognition and segmentation of fire scenes. These methods primarily utilize convolutional neural networks, recurrent neural networks, or their variants to construct end-to-end fire-detection systems. Due to the rapid development of deep learning models, researchers in the field continue to discover new variants, with each variant specifically targeting certain image features.

In recent years, numerous reviews on fire-detection methods have been published [22–28]. Table 1 provides a summary of the existing reviews in this field over the past few years. Xia et al. [23] pointed out that traditional methods rely on statistical learning approaches to achieve fine-grained tasks, but when the video image resolution is low, failure to effectively combine multiple visual features can result in higher false alarm rates. On the other hand, deep learning approaches have achieved superior efficiency and accuracy in finer-grained smoke-monitoring tasks. Gaur et al. [25] indicated that deep learning could complement approaches based on handcrafted feature extraction. Since directly using convolutional neural networks makes it challenging to detect lower-level features such as color, edges, or textures, it is recommended to employ a hybrid approach that combines handcrafted feature extraction and engineered features from deep learning models or to use 3D convolutional networks to extract features at different scales and fuse them, achieving the coexistence of multi-scale features. Chaturvedi et al. [26] pointed out that smoke detection based on machine learning and deep learning computer vision techniques has shown promising performance. They also provided an outlook on future research focuses in this field, including establishing datasets with complex backgrounds and variations, exploring interpretable deep learning approaches to reduce false alarm rates, achieving early distant smoke detection, and researching lightweight detection models suitable for IoT devices such as unmanned drones.

Based on the task, fire-detection methods can primarily fall into three categories: fire recognition, fire object detection, and fire segmentation:

(1) **Fire recognition**:
Fire recognition refers to determining whether there is the presence of smoke or flames in an image. It is also known as global fire recognition and represents the coarsest-grained recognition task in fire detection.

(2) **Fire object detection**:
Fire object detection is an extension of fire recognition in fire-detection tasks. Its main objective is to detect fire or smoke objects in a given image. The core functionality of this task is to roughly locate fire instances in the image and bounding box estimation. These bounding boxes provide localization information for the targets and serve finer-grained tasks in fire detection.

(3) **Fire segmentation**:
Fire segmentation involves accurately classifying every pixel in the image, separating the fire objects and their detailed boundaries from the image. It represents a comprehensive task encompassing fire classification, localization, and boundary delineation. Image segmentation can effectively identify and track fire events. When

a fire occurs, image segmentation can use surveillance cameras in open areas to capture the distribution of flames and make relatively accurate predictions about the spread of the fire, enabling quick localization of specific areas and appropriate responses. Smoke segmentation typically outputs masks with detailed boundaries involving object classification, localization, and boundary description.

Table 1. Comparison of previous surveys and reviews on fire detection.

Author	Year	Scenes	Key Notes
Gaur et al. [22]	2019	Building	This work discussed the advancements in fire-sensing technology and highlights the disparities between hardware and method development.
Xia et al. [23]	2019	Outdoor	This work comprehensively reviewed recent research results in smoke recognition, detection, and pixelwise smoke segmentation from both traditional and deep learning perspectives.
Bu et al. [24]	2019	Multi-scene environment	This endeavor entailed conducting a thorough examination of the visual-based intelligent fire-detection system, dividing it into two distinct categories: forest fires and all of the environment.
Gaur et al. [25]	2020	Indoor, Outdoor	This work focused on the discussion of the handcrafted rules and classifiers method and the deep learning method used for fire flame and smoke detection.
Chaturvedi et al. [26]	2022	Outdoor	This work primarily discussed the research progress in smoke detection focused on outdoor environmental scenes using visual technology. It comprehensively presented three research directions in smoke detection: classification, segmentation, and bounding box estimation.
Bouguettaya et al. [27]	2022	Forest, wildland	This work primarily focused on the comparative analysis of utilizing unmanned aerial vehicles (UAVs) and remote sensing technology based on deep learning approaches for the early detection of wildfires in forested and barren terrains.
Ghali et al. [28]	2023	Forest, wildland	This work conducted a comprehensive literature review on deep learning approaches for the classification, detection, and segmentation of wildland fires and introduced popular wildfire datasets in this field.

3. Datasets and Evaluation Metrics

3.1. Datasets

It is widely acknowledged that datasets constitute a crucial component of deep learning applications. In the domain of fire detection research, in order to significantly enhance detection effectiveness by enabling the model to extract a plethora of diverse and rich features, a sizable and well-curated dataset consisting of high-dimensional images and video sequences is imperative. Regarding the color aspects of samples, the range of flame colors spans from blue to red, with the specific hue being contingent upon factors such as the burning material and flame temperature. Smoke also plays a pivotal role in fire detection, typically manifesting as shades of gray, white, or black within videos. From the perspective of sample testing facilities, it is necessary to encompass various sectors such as industry, agriculture, infrastructure, households, and forests. In order to accomplish the

distribution of image features across such extensive application domains, the requirement arises for a voluminous and heterogeneous dataset.

Below are some commonly used and publicly available datasets in the field of fire detection:

- VisiFire dataset [29]:
  The VisiFire dataset is a widely used public video dataset for fire and smoke detection. It consists of four categories of videos: flame, smoke, other, and forest smoke. The fire set comprises 13 videos, the smoke set 21 videos, the forest smoke set 21 videos, and the other video set 2 videos. Dharmawan et al. [30] selected 12 commonly used videos from the VisiFire dataset for frame-by-frame segmentation annotation, resulting in 2684 annotated frames.
- BoWFire dataset [31]:
  The BoWFire dataset comprises 226 images of varying resolutions, with 119 images depicting fires and 107 representing non-fire scenes. The fire images encompass different scenarios of urgent fire events, such as building fires, industrial fires, accidents, and riots. The non-fire images include fire-like objects in red or yellow hues and sunset scenes. Additionally, a training set consisting of 240 images with a resolution of 50 $\times$ 50 px is provided, comprising 80 fire images and 160 non-fire images.
- Corsican Fire Database [32]:
  The Corsican Fire Database is a comprehensive dataset containing multi-modal wildfire images and videos. It includes 500 visible images, 100 multi-modal fire images incorporating visible light and near-infrared spectra, and 5 multi-modal sequences depicting fire propagation. The Corsican Fire Database encompasses annotations regarding fire and background attribute categories, describing visual information related to fires, such as flame color, smoke color, fire distance, percentage of smoke obscuring flames, background brightness, vegetation conditions, and weather conditions. Each image in this dataset is accompanied by its corresponding segmentation mask, which can be utilized for fire-segmentation research.
- FESB MLID dataset [33]:
  The FESB MLID dataset comprises 400 natural Mediterranean landscape images and their corresponding semantic segmentation. These images are segmented into 11 semantic categories, including smoke, clouds and fog, sunlight, sky, water surface, and distant views, among others. Additionally, an unknown region category is added, resulting in 12 defined classes. This dataset contains several challenging samples, where many smoke features are small-scale or distant smoke instances.
- Smoke100k [34]:
  Due to the hazy edges and translucent nature of smoke, the manual annotation of smoke objects can be challenging. The Smoke100k dataset provides a large-scale synthetic smoke image dataset for training smoke-detection models. The dataset includes three subsets: Smoke100k-L, Smoke100k-M, and Smoke100k-H, with 33 k, 36 k, and 33 k images, respectively. Each subset comprises synthetic smoke images, background images, smoke masks, and ground-truth bounding box positions. The Smoke100k dataset generates three different smoke masks based on smoke density to simulate the dynamic motion of rising smoke, blending pure smoke images with background images to generate synthetic smoke scene images.
- Video Smoke Detection Dataset [35]:
  The Video Smoke Detection Dataset (VSD) consists of three smoke videos, three non-smoke videos, and four sets of smoke and non-smoke image datasets. The image datasets are referred to as Set 1, Set 2, Set 3, and Set 4. Set 1 comprises 552 smoke images and 831 non-smoke images. Set 2 comprises 668 smoke images and 817 non-smoke images. Set 3 consists of 2201 smoke images and 8511 non-smoke images. Set 4 contains 2254 smoke images and 8363 non-smoke images. The non-smoke images exhibit many similarities to the smoke images in color, shape, and texture.
- FLAME dataset [36]:
  The Fire Luminosity Air-Based Machine Learning Evaluation (FLAME) provides aerial

images and videos of burning piled detritus in the Northern Arizona forests, collected using two unmanned aerial vehicles (UAVs). The dataset includes four photographic modes captured with conventional and thermal imaging cameras: normal, Fusion, WhiteHot, and GreenHot. The fire-recognition task comprises 48,010 RGB aerial images, divided into 30,155 fire images and 17,855 non-fire images, curated explicitly for wildfire recognition. The dataset includes 2003 segmentation masks with pixel-level annotations for the fire-segmentation task. This dataset serves as a valuable resource for fire recognition, segmentation methods, and further development of visual-based fire spread models.

- Flame and Smoke Detection Dataset [37]:
  The Flame and Smoke Detection Dataset (FASDD) is a large-scale dataset containing 100,000-level flame and smoke images from various sources, including surveillance cameras, drones, multi-source remote sensing satellite images, and computer graphics paintings depicting fire scenes. Moreover, the FASDD dataset encompasses a significant number of small-scale flame and smoke objects, posing challenges for deep learning research on small object detection. It consists of two subsets: FASDD_CV, which includes 95,314 samples captured from surveillance cameras, lookout towers, and drones, and FASDD_RS, comprising 5773 remote sensing image samples. Additionally, FASDD provides annotation files in three different formats.
- D-Fire dataset [38]:
  The D-Fire dataset is a collection of fire and smoke images specifically designed for object-detection-method development. Considering the diverse morphology of smoke and flame, the dataset incorporates data from the Internet, fire simulations, surveillance cameras, and artificially synthesized images where artificial smoke is composited with green landscape backgrounds using computer software. The D-Fire dataset consists of 21,527 images annotated with YOLO format labels, amounting to 26,557 bounding boxes. Among these, 1164 images depict fire, 5867 images solely smoke, 4658 images fire and smoke, and 9838 images as negative examples.
- DSDF [39]:
  The dataset for smoke detection in foggy environments (DSDF) is designed for studying smoke detection in foggy conditions. It comprises over 18,413 real-world images collected in both normal and foggy weather conditions. These images are annotated with four distinct categories, namely: non-smoke without fog (nSnF), smoke without fog (SnF), non-smoke with fog (nSF), and smoke with fog (SF). The dataset consists of 6528 images for nSnF, 6907 for SnF, 1518 for nSF, and 3460 for SF. DSDF covers a wide range of smoke variations in terms of color, size, shape, and density. Additionally, the samples in the dataset provide rich background information, which contributes to enhancing the detection model's generalization capability in real-world scenarios.
- DFS [40]:
  The Dataset for Fire and Smoke Detection (DFS) contains 9462 fire images collected from real-world scenes. The images are categorized based on the proportion of the flame area in the image, including Large Flame, Medium Flame, and Small Flame, with 3357, 4722, and 349 images, respectively. In addition to the annotations for "Flame" and "Smoke", the DFS includes a new category called "Other" to label objects such as vehicle lights, streetlights, sunlight, and metal lamps, comprising a total of 1034 images. This "Other" category is included to reduce false positives caused by misclassification.

In the context of deep-learning-based video-fire-detection tasks, addressing challenges such as limited sample quantity and uneven sample distribution across different scenarios is crucial to enhance the performance, generalization capability, and robustness of the models employed. One effective approach involves the targeted utilization of data augmentation techniques. Specifically, for fire images, data augmentation methods such as CutMix, Mosaic, MixUP, GridMask, and Gaussian blur can be employed to augment the diversity of the dataset. Figure 2 shows the Mosaic and MixUp data augmentation methods, with the white boxes representing bounding boxes of smoke or flame objects. To enable the

network model to learn more-comprehensive semantic features, a combination of data augmentation techniques is often employed in various forms, which results in richer and more-complex approaches compared to geometric and color transformations.

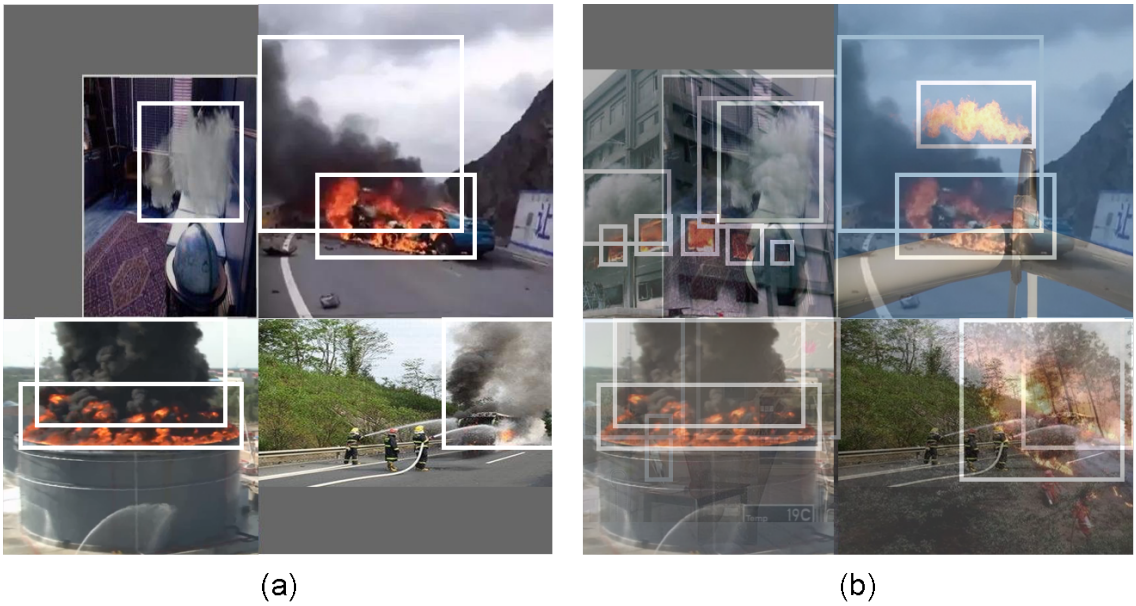

Figure 2. Visual comparison of Mosaic and MixUP. Mosaic involves combining four randomly cropped images to create a new image, thereby enhancing the diversity of the training data and enriching the spatial semantic information, as shown in subfigure (**a**). MixUp employs linear interpolation on input data and labels to generate novel training data, effectively expanding the training dataset. Subfigure (**b**) shows uses a combination of the Mosaic and Mixup data augmentation strategy.

3.2. Evaluation Metrics

3.2.1. Evaluation Metrics for Fire Recognition

Due to the particular nature of fire-detection tasks, it is necessary to evaluate the recognition performance from various perspectives in addition to solely relying on accuracy. From a multi-classification standpoint, objective metrics for recognition tasks include true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). TPs refers to the number of instances correctly classified as fires, FPs the number of instances mistakenly classified as fires, TNs the number of instances correctly classified as non-fires, and FNs the number of instances mistakenly classified as non-fires. When analyzing recognition tasks in multi-classification scenarios, a set of commonly used evaluation metrics includes the Accuracy Rate (AR), Detection Rate (DR), Precision Rate (PR), False Alarm Rate (FAR), and False Negative Rate (FNR).

Accuracy Rate is the probability of correctly classifying both positive and negative samples in fire-detection tasks. It evaluates the overall recognition performance of the model and serves as an essential reflection of the algorithm's overall performance.

$$AR = \frac{TPs + TNs}{TPs + TNs + FPs + FNs} \times 100\% \quad (1)$$

The Detection Rate represents the probability of correctly classifying all fire samples as fires, reflecting the algorithm's accuracy in identifying fire targets.

$$DR = \frac{TPs}{TPs + FNs} \times 100\% \quad (2)$$

The Precision Rate, on the other hand, refers to the proportion of samples identified as fires that are true fire samples.

$$PR = \frac{TPs}{TPs + FPs} \times 100\% \tag{3}$$

Both the False Alarm Rate and the False Negative Rate are vital indicators in fire detection. The False Alarm Rate denotes the proportion of non-fire samples incorrectly identified as fire samples, while the False Negative Rate represents the proportion of fire samples not recognized as such. Given the low probability of fire incidents, real-time surveillance videos usually do not contain actual fires. Excessive false alarms would burden the managerial staff with the task of verification. Hence, it is crucial to control false alarms in fire detection.

$$FAR = \frac{FPs}{FPs + TNs} \times 100\% \tag{4}$$

$$FNR = \frac{FNs}{FPs + TNs} \times 100\% \tag{5}$$

The Fmeasure is a commonly used metric that combines the Recall Rate and Precision Rate, serving as the harmonic mean between these two metrics. A higher Fmeasure, closer to 1, indicates greater accuracy, allowing for effective differentiation of the strengths and weaknesses of the algorithm. The parameter β in F_β denotes the degree of bias towards the Precision Rate or Recall Rate when evaluating the algorithm.

$$F_\beta = \left(1 + \beta^2\right) \times \frac{PR \times DR}{\beta^2 \times PR + DR} \times 100\% \tag{6}$$

$$F_1 = \frac{2 \times PR \times DR}{PR + DR} \times 100\% \tag{7}$$

3.2.2. Evaluation Metrics for Fire Object Detection and Segmentation

In the fire-object-detection and -segmentation task, the overall goal is that the method can quickly identify all smoke and fire objects in the video frame and can assist the relevant personnel in locating the fire area in the video. The fire object detection task can be considered a fire-recognition task extended to the time axis, and therefore, the evaluation metrics related to the fire-recognition task can be used. In general, the performance of the detection model can also be measured by the size of the area where the predicted bounding box overlaps with the true value bounding box.

The Intersection over Union (IoU), also known as the Jaccard Overlap, reflects the degree of overlap between a candidate box and the corresponding ground-truth box, specifically the ratio of their intersection to their union. When this ratio is 1, it signifies complete overlap. It is widely accepted that if the IoU between a candidate box and a fire objectś ground-truth bounding box is greater than 0.5, the detection is considered correct; otherwise, it is deemed erroneous. RoI(y) represents the rectangular region corresponding to the coordinate vector $\hat{y}$, with y and y denoting the coordinates of the candidate box and the ground-truth box, respectively.

$$f_{\text{IoU}} = \frac{|\,\text{RoI}(\hat{y}) \cap \text{RoI}(y)|}{|\,\text{RoI}(\hat{y}) \cup \text{RoI}(y)|} \tag{8}$$

For positive samples, only when all objects in an image are detected can we classify it as a true positive. Even if multiple objects are detected successfully, they are classified as false negatives. The detection requirement is not met if even one fire object is missed. For negative samples, if there is no detection box, they are classified as true negatives; if there is a detection box, they are classified as false positives.

The mean average precision (mAP) is used to evaluate the performance of object detection algorithms. The average precision (AP) for a specific target category is obtained by integrating the recall–precision curve. Here, N represents the number of object categories, and AP_n denotes the precision for the n^{th} object category. A higher AP value and a closer mAP value to 1 indicate a higher overall recognition accuracy of the model. mAP@0.5

refers to the mAP at an IoU threshold of 0.5. mAP@[0.5:0.95] refers to taking the threshold of IoU from 0.5 to 0.95, with a step size of 0.05, and then calculating the mAP under these conditions.

Common segmentation evaluation metrics include the mean Intersection over Union (mIoU), average mean-squared error (mMSE), and Dice coefficient. The mIoU is the standard measure for semantic segmentation, calculating the average ratio of the intersection to the union for all categories. A higher mIoU indicates higher accuracy. In the equation, GT represents the ith ground-truth, PR represents the predicted segmentation map for the ith image, and N represents the number of images in the set.

$$f_{\text{mlou}} = \frac{1}{N}\sum_{i}^{N}\frac{|GT_i \cap PR_i|}{|GT_i \cup PR_i|} \tag{9}$$

The average mean-squared error (mMSE) can be used as another quantitative evaluation metric for segmentation. It is calculated by averaging the mean-squared differences between the predicted results of the model and the ground-truth, across all pixels. Here, x_j^k represents the coordinate of the j^{th} pixel in the k^{th} image, and n is the number of images in the test dataset. A smaller mMSE value indicates better segmentation results for the fire-segmentation model.

$$\text{MSE} = \sum_{k=1}^{N}\sqrt{\frac{1}{M_k}\sum_{j=1}^{M_k}\left(P\left(\boldsymbol{x}_j^k\right) - G\left(\boldsymbol{x}_j^k\right)\right)^2} \tag{10}$$

$$\text{mMSE} = \frac{1}{n}\,\text{MSE} \tag{11}$$

In the pixel-level fire-segmentation task, the evaluation metrics for fire recognition can be extended for use at the pixel level.

Currently, no standardized video fire dataset is available for training and testing fire-recognition, -detection, and -segmentation tasks. Furthermore, the diversity of fire scenes, shooting distances and angles, and imaging devices can result in significant variations among fire videos. As a result, the same algorithm may exhibit different detection performances on different datasets. Therefore, the specific numerical values of the evaluation criteria, such as the Detection Rate and the False Alarm Rate for relevant detection algorithms, have limited reference values. Instead, their relative values are more meaningful in guiding algorithm improvements during the research process.

In addition to the aforementioned metrics, algorithms utilized for fire-recognition, -detection, and -segmentation tasks should also consider performance indicators that affect real-time capability, stability, and operational costs, such as model size, detection speed, and related measures such as the giga floating-point operations per second (GFLOPs), frames per second (FPS), and inference time. These performance indicators are of practical importance when considering the application of such algorithms in engineering contexts.

4. Deep-Learning-Based Approaches for Videos Fire Detection

4.1. Fire Recognition Methods

Fire recognition refers to determining the presence of smoke or flames in an image, representing the coarsest classification level in fire detection. Classic classification networks based on CNNs, such as AlexNet [41], GoogLeNet [42], VGGNet [43], ResNet [44], MobileNet [45], and DenseNet [46], have been employed for fire-recognition tasks, significantly enhancing the robustness of the models in identifying environmental features, while reducing false positives. Muhammad et al. [47] proposed a method that incorporates the lightweight network SqueezeNet as the backbone network and fine-tunes its architecture, including using smaller convolutional kernels and excluding dense fully connected layers for video fire detection. The results demonstrated that this model is more efficient regarding model size and inference speed. Additionally, the authors developed an algorithm to

select fire-sensitive feature maps from the convolutional layers, enabling further analysis of flame propagation speed and determination of the fire's severity or degree of combustion. Yuan et al. [48] proposed a deep multi-scale CNN (DMCNN) for smoke recognition. It incorporates the multi-scale convolutional structure of Inception to achieve scale invariance and adopt multi-scale additive merging layers to reduce computational cost while preserving more dynamic and static smoke features. Khudayberdiev et al. [49] proposed a lightweight fire-detection model, Light-FireNet, inspired by the concept of hard Swish (H-Swish). This network combines a more-lightweight convolution mechanism with a novel architectural design, resulting in a smaller model size while maintaining high detection accuracy. Zheng et al. [50] proposed a dynamic CNN model named DCN_Fire for assessing the risk of forest fires. This approach utilizes principal component analysis (PCA) reconstruction techniques to enhance inter-class discriminability and employs saliency detection to segment flame images into standard sizes for model training. The experimental results demonstrated that DCN_Fire achieved an accuracy of 98.3% on the test set. Majid et al. [51] proposed a CNN model with an attention mechanism for fire detection and employed GRAD-CAM visualization to display the contribution distribution of the model's flame predictions. This model achieved high Recall Rates of 95.4% and 97.61%. Tao et al. [52] proposed an adaptive frame-selection network (AFSNet) for video smoke recognition, as shown in Figure 3. This method automatically selects the most-useful video frames to reduce feature redundancy and introduces enhanced dilated convolution to mitigate the loss of detailed information. The method learns discriminative representations by considering multi-scale, context, and spatiotemporal information. The experimental results indicated that the proposed method achieved an accuracy of 96.73% and an F1-score of 87.22% on the SRSet dataset. However, the proposed method has not been experimentally evaluated for smoke recognition in mobile scenarios. SE-EFFNet, proposed by Khan et al. [53], utilizes EfficientNet-B3 as the backbone network for extracting useful features. It employs stacked autoencoders to achieve effective feature selection. EfficientNet ensures a balance among depth, width, and resolution dimensions while introducing the dense connectivity network from DenseNet to ensure effective fire scene recognition. The experimental results demonstrated that, compared to the baseline model, SE-EFFNet achieved a lower false positive rate, false negative rate, and higher accuracy. However, this architecture may suffer from overfitting in complex environments, and its performance in real-time processing on resource-constrained devices is average when deployed.

Utilizing only the CNN for extracting static features in fire recognition would increase computational complexity, impact detection performance, and lower recognition accuracy due to the subtle nature of early fire features. However, we can gather crucial information for fire recognition by observing dynamic characteristics such as the movement, flickering, and diffusion of smoke. Fusing deep and traditional features through multi-feature fusion methods is an important research direction. Huang et al. [54] combined CNN with traditional spectral analysis techniques for fire detection. They proposed a novel Wavelet-CNN approach for feature extraction, utilizing the 2D Haar transform to extract spectral features from images, which were then input into the FPN network. The experimental results yielded promising outcomes. The authors employed ResNet50 and MobileNet v2 as the backbone networks, and compared to not using the 2D Haar transform, they achieved improved fire detection accuracy and reduced false alarms. Kwak et al. [55] performed preprocessing on flame regions using color- and corner-detection techniques and employed dark channel prior characteristics and optical flow for smoke detection. They proposed a fire-detection method based on deep learning and image filtering. As a result, this method effectively reduced the false detection rates and improved the accuracy, achieving 97% accuracy for flame detection and 94% for smoke detection.

By utilizing a CNN to extract various low-level and high-level features and an RNN to capture dependencies and sequential patterns, new perspectives have emerged for utilizing computer vision in fire detection. Hu et al. [56] attempted to combine deep convolutional long short-term memory (LSTM) networks with optical flow methods for real-time fire

detection. By studying the static and dynamic characteristics of fire detection, they transformed fire images into optical flow images in real-time and utilized deep LSTM with sequence learning capability for training. The experimental results demonstrated that incorporating optical flow as the input improved the detection performance, although the effect was less pronounced for video frames with slow changes in fire or smoke. Ghosh et al. [57] proposed a combined CNN and RNN model, a hybrid deep learning approach, for forest fire detection. By incorporating the RNN, the model can better capture the correlations and dynamic features among adjacent frames in fire videos, thereby improving the accuracy and efficiency of fire detection. However, RNN models have higher computational complexity and may not meet real-time requirements in fire detection, particularly for slowly evolving fire incidents, where their performance is suboptimal.

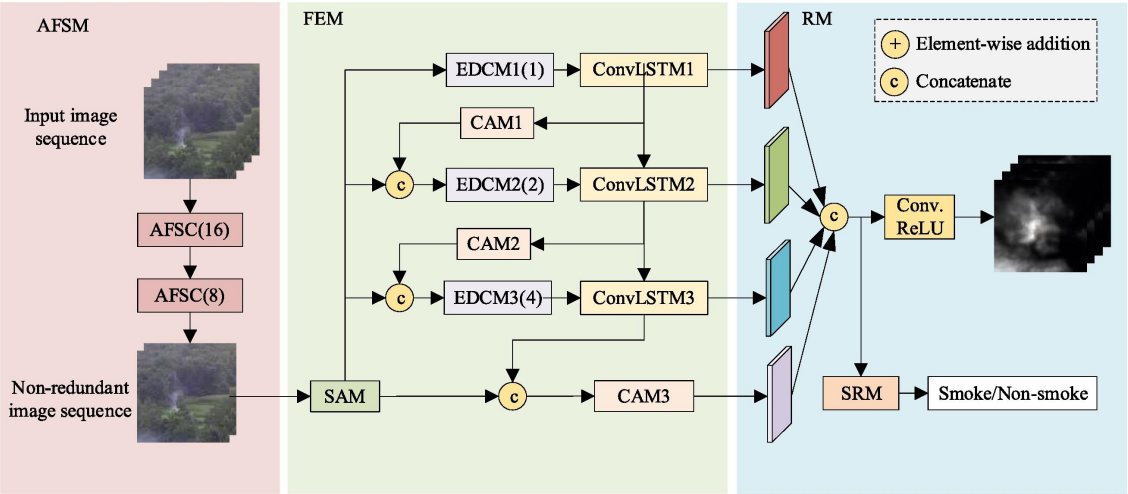

Figure 3. AFSNet's network architecture. The architecture consists of an adaptive frame-selection module, a feature-extraction module, and a recognition module. The figure was borrowed from the original paper [52].

Currently, most fire-detection methods only consider normal weather conditions, and when video images are degraded in adverse weather environments, using traditional image-processing algorithms for dehazing would increase the computational costs. To improve the generalization performance of fire detection in adverse weather conditions, He et al. [58] proposed a method for smoke detection in both normal and hazy conditions. This method introduced an attention mechanism module that combines spatial attention and channel attention to address the detection of small smoke particles. By incorporating lightweight feature-level and decision-level fusion modules, the method enhances the discrimination between smoke, fog, and other similar objects while ensuring the real-time performance of the model. However, the method still faces challenges in detecting small targets due to the high probability of small targets being affected by fog interference. Gong et al. [39] proposed a smoke-detection method that combines the dark channel assisted with mixed attention and feature fusion. They employed a two-stage training strategy to address the issue of data imbalance. The method was evaluated on a smoke-detection dataset containing hazy environments, and the experimental results showed an accuracy of 87.33% and an F1-score of 0.8722 for the proposed method.

For more details on the fire-detection methods, please refer to Table 2.

Table 2. The main methods of fire recognition based on deep learning.

Method	Technique	Application Scenario	Dataset	Evaluate
Muhammad et al. [47]	SqueezeNet, feature map selection	Fire detection in monitoring scenarios	BoWFire Dataset	PR (%) = 86; F-m (%) = 91
Yuan et al. [48]	Deep multi-scale convolutional	Multi-scene smoke detection	The datasets of smoke images	DR (%) = 98.55; AR (%) = 99.14; FAR (%) = 0.36
Khudayberdiev et al. [49]	Hard Swish	Multi-scene fire detection	55,500 images, including fire and non-fire	AR(%) = 97.83; PR (%) = 98.37 F-m (%) = 99.18
Zheng et al. [50]	Dynamic CNN, PCA reconstruction techniques	Forest fire smoke detection	More than 4000 forest fire risk images	AR (%) = 98.3; FNR (%) = 0.13
Majid et al. [51]	EfficientNet-B0, attention mechanism, Grad-CAM	Multi-scene fire detection	7977 images, including fire and non-fire	AR (%) = 95.40; DR (%) = 97.61; FNR (%) = 94.76
Tao et al. [52]	Adaptive frame-selection convolution, dilated convolution	Smoke detection in surveillance video scenes	SRSet	DR (%) = 96.73; FAR (%) = 3.16; F-m (%) = 96.57
Khan et al. [53]	EfficientNet, autoencoder, weights' randomization	Fire detection in surveillance video scenes	Foggia Dataset [59]	AR (%) = 97.20; FAR (%) = 0.042; FNR (%) = 0.034
Huang et al. [54]	Haar wavelet transform, Faster R-CNN	Fire detection in surveillance video scenes	5667 images, including fire and non-fire	PR (%) = 89.0; F-m (%) = 94.0
Kwak et al. [55]	Dark channel prior, Lucas–Kanade method, Inception-V4	Multi-scene fire detection	8000 images, including flame, smoke, and non-fire	AR-flame (%) = 97.0; AR-smoke (%) = 94.0
Hu et al. [56]	Deep LSTM, optical flow method	Open space fire detection	The video dataset includes 100 fire videos and 110 non-fire videos	AR (%) = 93.3; F-m (%) = 90.0
Ghosh et al. [57]	Combination of CNN and RNN networks for feature extraction	Forest fire smoke detection	Mivia Dataset	AR (%) = 99.54; DR (%) = 99.75
He et al. [58]	Spatial and channel attention mechanism, FPN	Smoke detection in fog scenes	Fog smoke dataset for 33,666 images	AR (%) = 92.3088; F-m (%) = 92.3833
Gong et al. [39]	Dark-channel-based mixed attention, two-stage training strategy	Smoke detection in fog scenes	DSDF	AR (%) = 87.33; F-m (%) = 87.22

4.2. Fire-Object-Detection Methods

Currently, deep learning object-detection networks such as Faster RCNN [60–63], YOLO [64–74], and SSD [75,76] have demonstrated outstanding performance in fire-detection applications. By combining traditional fire-recognition methods, Barmpoutis et al. [62] proposed a novel fire-object-detection approach. This method trains a Faster R-CNN model to obtain candidate fire regions in images and then utilizes a multidimensional texture analysis based on higher-order linear dynamical systems to determine whether the region proposals are fire regions. The experimental results indicated that, compared to using VGG16 and ResNet101 as the base networks with the YOLO v3 and SSD methods, this approach achieved a higher F1-score. Chaoxia et al. [63] improved the Faster R-CNN for fire detection by employing a color-guided anchoring strategy and a global

information-guided strategy. However, the performance of this method was compromised due to the predefined anchors. Although two-stage object-detection algorithms exhibit high accuracy and precise localization, they are more complex and slower in detection speed. Addressing the deployment requirements of industries with a high-risk of fires, such as the chemical industry, Wu et al. [73] proposed an intelligent fire-detection method utilizing cameras. This method consists of three steps: motion detection, fire detection, and fire classification. It combines motion detection based on background subtraction, fire object detection based on YOLO, and region classification of fire and fire-like images using Xception. By employing a region-classification model for fire recognition, the system outputs the coordinates of the fire region once a fire is detected. Mukhiddinov et al. [66] developed an automatic fire-detection system based on an improved YOLOv4 model and convolutional block attention module for visually impaired individuals. However, this system suffers from a high false alarm rate. Regarding one-stage object detection algorithms, these models do not require generating region proposals, resulting in faster computational speed. They are suitable for scenarios that demand real-time performance. However, they often exhibit missed detections and false positives for small objects.

Due to the variations in fire color, lighting conditions, flame characteristics, and smoke shapes across different scenarios, traditional visual-based methods require the manual design of fire-detection features. In fire-object-detection methods based on deep learning, the effectiveness of detecting fires at different scales using a single-scale feature map is not sufficiently robust. Therefore, researchers have attempted to incorporate multi-scale feature-fusion techniques to balance the models' accuracy, model size, and inference speed. Li et al. [77] proposed EFDNet, which employs a multiscale feature-extraction mechanism to enhance spatial details in the lower-level feature-extraction stage, thus improving the discriminative ability for fire-like objects. They utilized an implicit deep supervision mechanism through dense skip connections to enhance the interaction between information flows, transforming shallow spatial features into high-level semantic information. Additionally, they employed channel attention mechanisms to selectively emphasize the contributions of different feature maps, capturing richer and more-effective deep semantic information. This approach achieved an accuracy of 95.3% with a compact model size of only 4.80 MB. Huo et al. [72] introduced an optimized multi-scale smoke-detection method based on the YOLOv4 model. They enhanced the features of small targets by incorporating a spatial pyramid pooling (SPP) module and reduced the network parameters using depthwise-separable convolutions. The experimental results demonstrated that their method achieved an accuracy of 98.5% and a detection speed of 32 FPS, exhibiting higher sensitivity to early-stage smoke detection.

To reduce false alarms and achieve improved predictive performance, Venâncio et al. [74] proposed a hybrid fire-detection method based on temporal and spatial patterns. This method consists of two stages: firstly, potential fire events are detected using a target-detection model based on the YOLO architecture, and then, the dynamic changes of these events over time are analyzed, including the duration and spatial expansion of flame and smoke objects. Xie et al. [78] introduced a video-based fire-detection method that utilizes dynamic features based on motion flicker and deep static features. Dynamic features were extracted by analyzing the differences in motion and flicker characteristics between fire and other objects in the video. They also proposed an adaptive lightweight convolutional neural network (AL-CNN) to extract deep static features of fire. This approach aims to reduce computational burden while avoiding the loss of image features caused by fixed-size image inputs.

The utilization of 3D convolutional networks allows for the simultaneous extraction of feature information from both temporal and spatial dimensions, thereby enhancing the efficiency of object detection. Huo et al. [79] proposed an end-to-end 3D convolutional smoke-object-detection network called 3DVSD. The network captures the moving objects within the input video sequence, extracting spatiotemporal features through 3D convolution. It utilizes the temporal variations in static features to perform identification and localization in the time dimension. The authors also investigated the influence of different

time steps and time spans between two video frames on the detection performance. The experimental results demonstrated that 3DVSD achieved an accuracy of 99.54% and a false positive rate of 1.11%. However, it should be noted that this method is only applicable to stationary surveillance cameras. When the camera moves, it becomes challenging to effectively capture smoke features, leading to potential false positives or missed detections.

In recent years, the Transformer framework has become the mainstream architecture in natural language processing due to its robust feature-extraction capabilities. Its applications in the visual domain have also gained widespread attention and utilization [80]. Li et al. [81] incorporated a CNN into the popular DETR network for fire detection, as shown in Figure 4. This network enhanced the detection of small objects by adding a normalization-based attention module [82] during the feature-extraction stage and employing deformable attention mechanisms in the encoder–decoder structure. However, this approach has drawbacks, such as slower processing speed and higher device requirements. Yang et al. [83] proposed a fire-detection network called GLCT, which combines a CNN and Transformer. The model introduces a backbone network called MobileLP based on the MobileViT block [84], enabling feature extraction of both global and local information. By combining SPP with a BiFPN for feature fusion and incorporating YOLO Head, the GLCT network was constructed holistically. The experimental results demonstrated that the GLCT network achieved an mAP of 80.71, striking a balance between speed and accuracy. Yan et al. [71] proposed a fire-detection model based on an enhanced YOLOv5. They enhanced the model's feature-extraction capability by introducing coordinate attention blocks, Swin Transformer blocks, and an adaptive spatial feature fusion model.

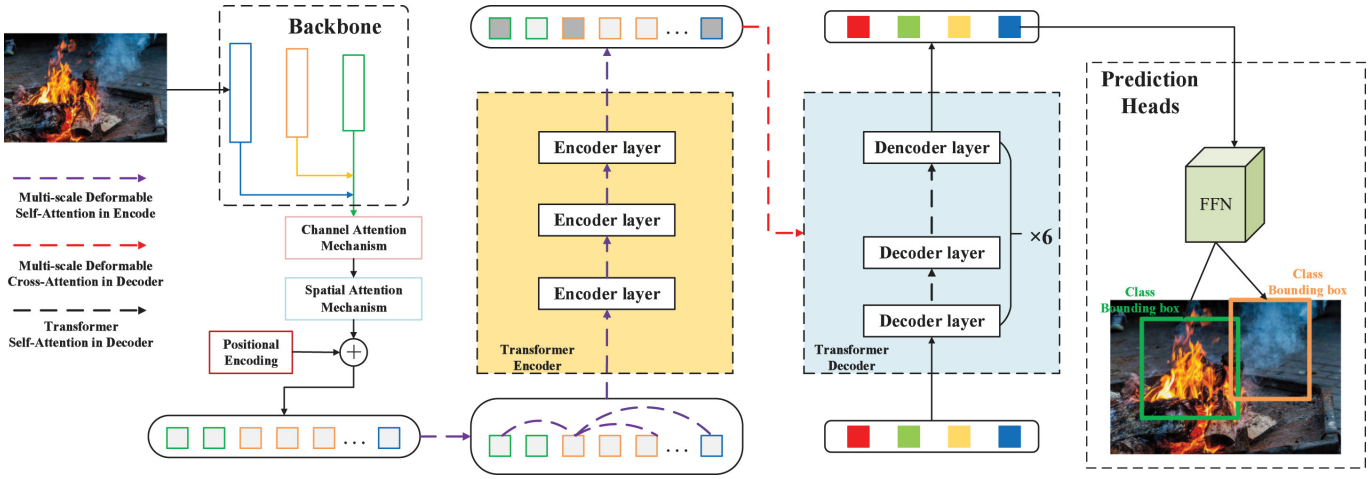

Figure 4. Overview of the proposed framework from the original paper [81]. The network model consists of three parts, feature extraction network, encoder–decoder structure, and prediction head.

For a more-detailed overview of fire-object-detection methods, please refer to Table 3.

Table 3. The main methods of fire object detection based on deep learning.

Method	Technique	Application Scenario	Dataset	Evaluate
Barmpoutis et al. [62]	Faster R-CNN, linear dynamical systems, Grassmannian VLAD encoding	Multi-scene fire detection	Corsican Fire Database	F-m (%) = 99.7
Chaoxia et al. [63]	Faster R-CNN, color-guided anchoring strategy, global information network	Multi-scene fire detection	3719 images, including fire and non-fire	AR (%) = 93.36 F-m (%) = 94
Chen et al. [67]	YOLOv5s, CoT, CA, BiFPN	Multi-scene fire detection	2976 images including BowFire and forest fire	mAP@0.5(%) = 87.7
Yan et al. [71]	YOLOv5,CA,ASFF Swin transformer	Multi-scene fire detection	2059 flame images	mAP@0.5 (%) = 66.8 mAP@[0.5:0.95] (%) = 33.8
Huo et al. [72]	YOLOv4, SPP, Depthwise-separable convolution	Multi-scene fire detection	9270 images, including smoke and non-smoke	AR (%) = 97.8 FAR (%) = 1.7 F-m (%) = 97.9
Wu et al. [73]	YOLO, background subtraction	Multi-scene fire detection	5075 flame images	mAP@0.5 (%) = 60.4
Venâncio et al. [74]	YOLOv5, TPT, AVT	Multi-scene fire detection	D-Fire Dataset	mAP@0.5 (%) = 79.10 ± 0.36 AP_{smoke} (%) = 85.88 ± 0.35 AP_{fire} (%) = 72.32 ± 0.52
Huo et al. [79]	YOLO layer, 3D convolutional, SPP	Multi-scene fire detection	14,700 images, including smoke and non-smoke	AR (%) = 99.54 FAR (%) = 1.11 FNR (%) = 0.14
Li et al. [81]	DETR, NAM	Multi-scene fire detection	26,060 images including fire, smoke and two-object with both smoke and fire	AP_{smoke} (%) = 76.0 AP_{fire} (%) = 81.7
Yang et al. [83]	MobileViT, SPP, BiFPN, YOLO Head	Multi-scene fire detection	3,717 images of the early stages of the fire	mAP@0.5(%) = 80.71

4.3. Fire-Segmentation Methods

Semantic segmentation of fire is the most-granular classification task within fire detection. This task entails pixel-level classification of an image, where the goal is to separate the fire targets and their intricate boundaries from the image itself, enabling fire recognition, localization, and boundary delineation. Image-segmentation techniques prove effective in identifying and tracking fire incidents, facilitating accurate predictions of fire spread patterns, swift localization of specific areas, and subsequent measures [85]. For example, the smoke-segmentation task can be regarded as a dense binary classification problem on a per-pixel basis. Considering that most smoke can be perceived as inconspicuous entities due to the small size and lack of prominence in the early stages of a fire, it often exhibits traits such as translucency and low contrast, rendering smoke region segmentation challenging. Traditional methods of fire semantic segmentation heavily rely on manually designed features. However, with the advent of FCN [86], deep learning has gradually been applied in the field of image semantic segmentation. Numerous novel semantic segmentation network architectures have been developed for fire semantic segmentation, such as U-Net [87–89], and DeepLab [90,91], among others.

Khan et al. [92] presented a deep-learning network for smoke detection and segmentation in hazy environments. The network utilizes an efficient lightweight CNN architecture called EfficientNet and employs the DeepLabv3+ network for per-pixel smoke segmentation. The experimental results demonstrated that the network achieved a global accuracy

of 91.34% and an mIoU of 77.86% on a custom dataset. Yuan et al. [93] proposed the classification-assisted gated recurrent network (CGRNet), which incorporates attention-embedding gated recurrent units (GRUs) to learn the spatial correlations of smoke and long-range feature dependencies. The network employs the Xception network for feature extraction and utilizes a semantic segmentation module composed of four branches. Multiple attention convolutional GRU units are stacked to construct a deep network. The classification results are used to assist the segmentation process and enhance segmentation accuracy. The experimental results demonstrated that this network performs favorably in smoke semantic segmentation tasks. However, despite the improvement in detection accuracy to some extent, the complex algorithmic architecture of the network increases the requirements for device capabilities in practical applications. Due to the nature of fire incidents, network models need to achieve real-time monitoring during the smoke-segmentation process. However, accurate smoke semantic segmentation requires rich spatial information and a larger receptive field. Shahid et al. [94] proposed using a self-attention mechanism to augment spatial characteristics with temporal characteristics for fire detection and segmentation, enabling the network to reduce reliance on spatial factors such as shape or size while leveraging strong spatiotemporal dependencies. This approach consists of two stages: in the first stage, region proposals are extracted using spatial–temporal feature extraction to separate fire region features from the background; in the second stage, each region proposal is classified as fire or non-fire. Yuan et al. [95] proposed the cubic-cross-convolutional attention and count prior embedding network (CCENet) to address the smoke-semantic-segmentation task by combining attention mechanisms and global information from smoke pixel counts. It generated a cubic-cross-convolutional kernel by fusing convolutional results along different axes and introduced a cubic-cross-convolutional attention mechanism to handle long-range dependencies of smoke pixels. The experimental results on synthetic and real smoke datasets confirmed that the proposed module enhanced the segmentation task performance. Li et al. [96] proposed a dual-path real-time smoke segmentation network based on BiSeNet [97]. This network utilizes spatial path encoding to enrich spatial information details and leverages the contextual path structure, lightweight models, and global average pooling to provide a sufficient receptive field for extracting fire smoke features, enhancing the ability to capture global information. The experimental results demonstrated that this method achieved excellent performance while reducing complexity, ensuring real-time segmentation speeds, and offering high practical value.

To meet the requirements of real-time fire segmentation on computationally constrained devices, Song et al. [98] proposed the squeezed fire binary segmentation network (SFBSNet). This model utilizes an encoder–decoder architecture and achieves real-time, efficiency, and high-precision fire segmentation by introducing confusion blocks and depthwise-separable convolutions. The experimental results demonstrated that SFBSNet achieved an IoU of 90.76% on the Corsican Fire Dataset. Furthermore, the proposed method was successfully ported to embedded devices, yielding favorable results. Yuan et al. [99] presented a lightweight real-time smoke-segmentation network, as shown in Figure 5. This network enhances the feature-encoding capability while reducing computation by designing a channel split and shuffle attention module (CSSAM). The experimental results showed this method's excellent performance on synthetic and real smoke datasets, with the network parameters being less than 1 million.

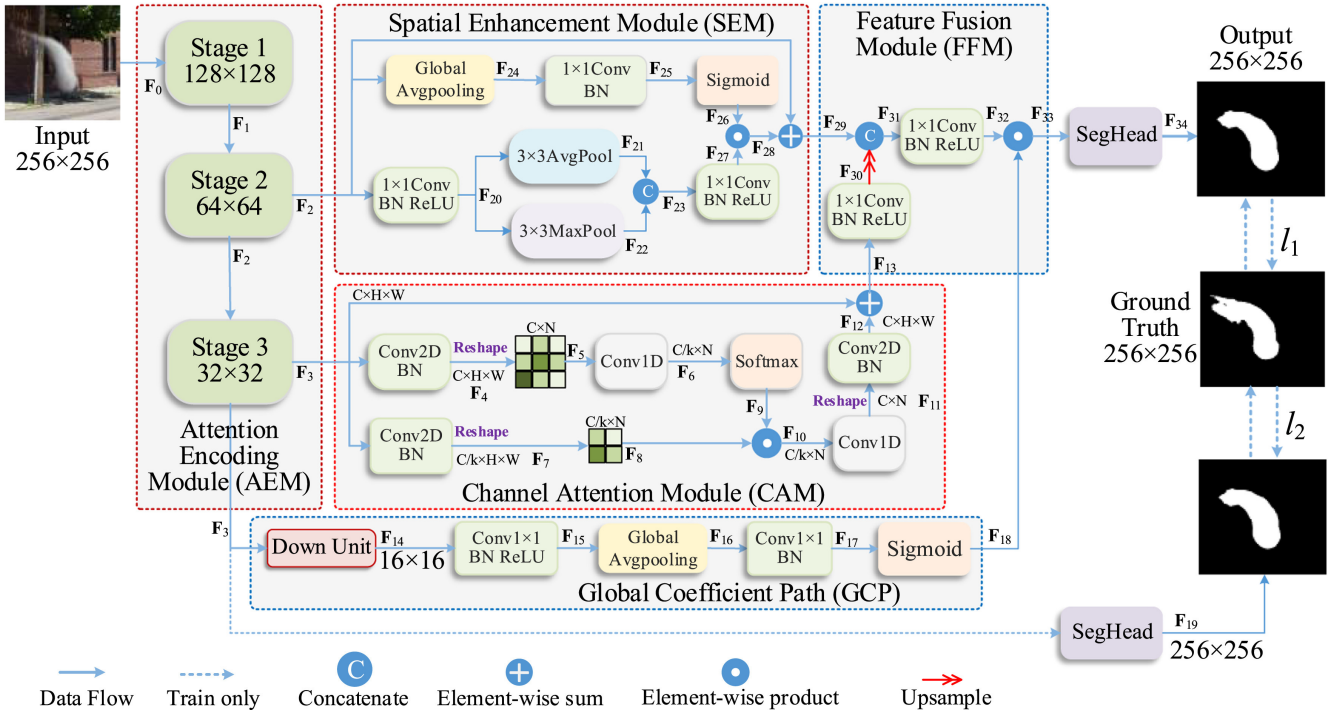

Figure 5. Overview of the proposed architecture from the original paper [99]. It consists of an attentional feature-extraction encoder and a feature-fusion decoder.

For forest fire scenarios, aerial images obtained from UAVs differ from ground images, offering advantages such as comprehensive coverage and high resolution. Utilizing UAV imagery for forest fire detection presents unique advantages and challenges, particularly for early-stage wildfire detection [100]. Barmpoutis et al. [91] introduced a novel remote-sensing system for early-stage fire detection. They employed an RGB 360-degree camera mounted on a UAV for early forest fire detection, utilizing two Deeplab V3+ models to identify candidate fire regions and perform flame and smoke segmentation tasks. Given that the RGB 360-degree camera provides precise horizon segmentation, the authors proposed a novel adaptive approach using the Karcher mean algorithm. This approach compares region proposal blocks with designated blocks of natural objects, such as clouds and sunlight, to reduce false positive detections. The experimental results demonstrated an F1-score of 0.946 for this method. Due to the typically large resolution of aerial images, directly slicing these high-resolution images for segmenting small fire areas can affect real-time detection. Guan et al. [101] proposed an improved instance segmentation model called MaskSU R-CNN for forest fire detection and segmentation in aerial images. By introducing new attention mechanisms and utilizing a U-Net to reconstruct the MaskIoU branch of Mask R-CNN, they achieved an accuracy of 91.85% and an F1-score of 0.903. Perrolas et al. [102] proposed a method based on the quad-tree search to achieve the localization and segmentation of fires at different scales. The quad-tree search approach can adaptively work at different scales, achieving high computational efficiency. The experimental results demonstrated an accuracy of 95.8% for the proposed method, showcasing its ability to segment small fire areas in high-resolution aerial images. Ghali et al. [103] developed an ensemble learning approach combining the EfficientNet-B5 and DenseNet-201 models to detect wildfires in aerial images. The authors utilized two Transformer-based models, TransUNet and TransFire, and a deep-CNN-based model called EfficientSeg for wildfire-segmentation tasks. The experimental results demonstrated 85.12% accuracy and an F1-score of 0.8477 for the recognition task. For the segmentation task, TransUNet achieved 99.90% accuracy and an F1-score of 0.999.

For a more-detailed overview of fire-semantic-segmentation methods, please refer to Table 4.

Table 4. The main methods of fire segmentation based on deep learning.

Method	Technique	Application Scenario	Dataset	Evaluate
Khan et al. [92]	EfficientNet, DeepLabv3+	Smoke detection in fog scenes	Fog smoke dataset for 252 images	mAR (%) = 93.33 mIoU (%) = 77.86 F-m (%) = 50.76
Yuan et al. [93]	Xception, GRU, CCL, PPM	Smoke detection in complex scenes	Synthetic smoke image dataset and a real smoke image dataset	mIoU (%) = 82.18 mMSE = 0.2212
Shahid et al. [94]	3D convolution, UNet++, self-attention	Fire detection in surveillance video scenes	1033 videos of which 559 contain fire and 434 contain normal scenes	F-m (%) = 84.80
Yuan et al. [95]	Cubic-cross-convolution, PPM, CPA	Multi-scene smoke detection	A synthetic smoke dataset consisting of 70,632 images	mIoU (%) = 76.01
Li et al. [96]	BiSeNet, PPM, ECA	Multi-scene smoke detection	8280 actual scenes of smoke images	AR (%) = 98.0 mIoU (%) = 80.9
Song et al. [98]	FusionNet, depthwise-separable convolution	Multi-scene fire detection	Corsican Fire Database	mIoU (%) = 90.76
Yuan et al. [99]	CSSAM, CA, SE	Multi-scene smoke detection	A synthetic smoke dataset consisting of 70,632 images	mIoU (%) = 74.2
Barmpoutis et al. [91]	DeepLab V3+, post-validation adaptive	Forest fire smoke detection	Fire detection 360-degree dataset	mIoU (%) = 77.1 F-m (%) = 94.6
Guan et al. [101]	MS R-CNN, UNet, FPN	Forest fire smoke detection	FLAME	mIoU (%) = 82.31 F-m (%) = 90.30
Perrolas et al. [102]	SqueezeNet, Deeplabv3+, Quadtree search,	Forest fire smoke detection	Corsican Fire Database	F-m (%) = 90.30 mIoU-fire (%) = 88.51
Ghali et al. [103]	EfficientSeg, Transformer	Forest fire smoke detection	FLAME	F-m (%) = 99.9

5. Discussion

In recent years, the domain of video fire detection, based on deep learning, has witnessed rapid advancement. Leveraging its formidable capabilities in representation learning and generalization, it has demonstrated superior performance to conventional methods in enhancing fire-detection accuracy, reducing false alarms, and minimizing instances of missed detection. Nonetheless, it is important to note that these techniques are still in the developmental stage. Within the realm of deep-learning-based video fire detection applications, several pressing issues and challenges demand resolution. Through a thorough analysis, this paper proposes potential recommendations for improving the current methodologies:

(1) **Establishing a high-quality fire dataset**:
In the domain of fire-detection research, the significant improvement of fire-detection models relies on the construction of a large-scale dataset comprising high-dimensional images. Such a dataset enables the models to extract diverse and rich features. However, the field currently faces challenges such as limited samples, sample imbalance, and a lack of diversity in the background, resulting in the absence of an authoritative standard dataset. The limited availability of publicly accessible fire videos and image datasets further restricts the models' generalization capabilities. It is recommended that researchers construct a high-quality dataset encompassing a wide

range of scenes, including public buildings, forests, and industrial areas. This dataset should incorporate various modalities of data, such as visible light and infrared, while considering different environmental conditions such as indoor settings, haze, and nighttime scenarios. To address the aforementioned issues, the utilization of generative adversarial networks (GANs) can aid in generating realistic fire images. Moreover, 3D computer graphics software can simulate highly controlled smoke and flame effects, integrating them with existing background image datasets to create synthetic data. By expanding the dataset, not only can the issue of sample imbalance be effectively alleviated, but it can also enhance the detection performance of fire-detection methods. Hence, it is advisable to prioritize the construction of high-quality fire datasets in video fire detection research, enabling a comprehensive exploration and evaluation of fire detection algorithms' performance and application capabilities.

(2) **Exploring information fusion and utilization with multiple features**:
In the context of fire detection in various scenarios, such as chemical industrial parks, forests, and urban buildings, the morphology of smoke and flames exhibits diversity, accompanied by a wide range of scale variations and significant feature changes. Deep-learning-based video-fire-detection models still have room for improvement in effectively extracting the essential characteristics of fires. Furthermore, in video-based fire detection, limited research addresses the utilization of information between consecutive video frames to capture the correlation between static features and dynamic changes. Therefore, it is recommended that researchers fully exploit the color, texture, flicker, and other characteristics of flames and smoke. Additionally, it is essential to consider the temporal and spatial information within the video sequence to effectively reduce the false negative and false positive rates of fire detection models.

(3) **Building lightweight models for edge computing devices**:
In recent years, deep learning has achieved significant success in fire detection. However, the inference process of deep learning models heavily relies on high-performance computing devices, particularly in complex environments where long-distance data transmission and centralized processing negatively impact efficiency. As a distributed computing architecture, edge computing places computational capabilities closer to the end devices to meet the high computational and low-latency requirements of deep learning [104]. Due to edge computing devices' limited computing and storage capacities, real-time performance is compromised for deep learning models with large network parameters and computational complexity. Thus, immature challenges persist in combining deep-learning-based fire-detection methods with low-power small-scale edge computing devices. To adapt to the resource limitations and real-time requirements of edge computing devices, it is recommended that researchers focus on studying lightweight fire-detection methods to enhance the detection efficiency of the models. Research on model compression primarily focuses on techniques such as quantization, pruning, and knowledge distillation. These methods compress the model's size and computational load by reducing the data precision, parameter compression, and knowledge transfer [105]. Additionally, designing efficient and lightweight backbone network architectures is an important research direction.

(4) **Conducting research on fire scene reconstruction and propagation trends based on video**:
By utilizing surveillance devices installed in the vicinity of fire scenes, such as buildings and lookout towers, we can gather abundant information about the fire. This fire-related information can be utilized to infer the physical parameters of the fire and assess the trends in fire spread. This provides vital auxiliary support for fire management and emergency response, including fire propagation prediction, intelligent graded response, and handling accidents and disasters. However, the current research in this field, specifically deep learning methods based on fire scene video

data, remains inadequate. It is recommended that researchers combine theoretical models from fire dynamics and heat transfer and utilize a vast amount of real fire data from various scenarios for their studies. Through the analysis of fire scene video data, it is possible to infer the physical parameters of the fire, such as the fire size and flame heat release rate (HRR). Furthermore, it is essential to investigate fire situation analysis based on video features and the actual conditions of the hazard-formative environment and the hazard-affected bodies to infer the fire's propagation trends. This will significantly contribute to enabling emergency rescue personnel to conduct rescue operations based on the fire situation. Therefore, it is advisable to carry out research on fire scene reconstruction and propagation trends based on video in order to provide more-effective decision support for fire management and emergency response.

(5) **Research on fire detection methods for unmanned emergency rescue equipment**: In recent years, the development of unmanned emergency rescue equipment has emerged as a prominent focus within the field of emergency response. When confronted with complex and extreme emergency scenarios, the utilization of unmanned rescue equipment enhances the efficiency and safety of fire rescue operations, thereby reducing casualties and property losses. Consequently, the study of fire-detection methods holds paramount importance in researching unmanned emergency rescue equipment, serving as a crucial technology for achieving equipment control and decision autonomy. For instance, the application of unmanned aerial drones in firefighting and rescue operations can encompass a wide range of emergency inspection tasks, thereby facilitating precise fire scene management. Unmanned drones equipped with visible light and infrared sensors can detect potential fire hazards day and night, thus enhancing real-time situational awareness for firefighting and rescue efforts. Furthermore, by integrating fire-detection methods with intelligent firefighting and rescue equipment, coupled with the utilization of unmanned automated firefighting vehicles, it becomes possible to identify areas affected by flames and to automatically respond by implementing appropriate extinguishing measures. Consequently, it is recommended to conduct research on fire-detection methods tailored explicitly for unmanned emergency rescue equipment, thereby promoting the intelligent and integrated development of such equipment and enhancing the efficiency of emergency response operations.

6. Conclusions

Fire safety is a crucial component within the modernization of the national emergency management system and capabilities. Accurate, effective, and timely fire detection plays a vital role in the initial stages of firefighting. Traditional fire-detection methods primarily rely on sensor-based technologies, but have limitations and shortcomings. In recent years, with the advancements in computer vision technology, researchers have widely applied deep learning in fire detection. By utilizing deep learning methods, remarkable results have been achieved in fire detection, surpassing traditional methods under certain conditions. Therefore, this study focused on the research status of video-based fire monitoring methods based on deep learning, explicitly exploring three aspects: recognition, object detection, and segmentation. The article presented an overview of the architectures of fire-detection methods developed in recent years, discussing their advantages and limitations in relevant fire-detection tasks. Commonly used datasets and evaluation metrics for these tasks were also introduced. Finally, the article further discussed the significant challenges and future research directions in applying deep-learning-based fire-detection methods in this field.

Author Contributions: Conceptualization, C.J. and T.W.; methodology, C.J. and T.W.; data curation, K.X. and J.Z.; formal analysis, C.J. and N.A.; writing—original draft preparation, C.J.; writing—review and editing, C.J. and T.W.; funding acquisition, T.W., S.Z. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the Anhui Province Scientific Research Plan Project, under Grant No.2022AH051101; Anhui Provincial Quality Engineering Project Modern Industry College, under Grant No.2021CYXY054; the open Foundation of Anhui Engineering Research Center of Intelligent Perception and Elderly Care, Chuzhou University, under Grant No.2022OPB01; Chuzhou Scientific Research Plan Project, under Grant No.2021ZN007; Reform and Practice of New Engineering Research on Emergency Technology and Management, under Grant No.2021XGK04; the Graduate Student Innovation Fund of the Anhui University of Science and Technology, under Grant No.2022CX2130.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. National Fire and Rescue Administration of China's Ministry of Emergency Management. The National Police Situation and Fire Situation in 2022. 2023. Available online: https://www.119.gov.cn/qmxfxw/xfyw/2023/36210.shtml (accessed on 3 August 2023).
2. Ministry of Emergency Management of the People's Republic of China. Basic Situation of National Natural Disasters in 2022. 2023. Available online: https://www.mem.gov.cn/xw/yjglbgzdt/202301/t20230113_440478.shtml (accessed on 3 August 2023).
3. Quintiere, J.G. *Principles of Fire Behavior*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2016. [CrossRef]
4. Çetin, A.E.; Dimitropoulos, K.; Gouverneur, B.; Grammalidis, N.; Günay, O.; Habiboğlu, Y.H.; Töreyin, B.U.; Verstockt, S. Video fire detection—Review. *Digit. Signal Process.* **2013**, *23*, 1827–1843. [CrossRef]
5. Wooster, M.J.; Roberts, G.J.; Giglio, L.; Roy, D.P.; Freeborn, P.H.; Boschetti, L.; Justice, C.; Ichoku, C.; Schroeder, W.; Davies, D. Satellite remote sensing of active fires: History and current status, applications and future requirements. *Remote Sens. Environ.* **2021**, *267*, 112694. [CrossRef]
6. 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] [PubMed]
7. Kaku, K. Satellite remote sensing for disaster management support: A holistic and staged approach based on case studies in Sentinel Asia. *Int. J. Disaster Risk Reduct.* **2019**, *33*, 417–432. [CrossRef]
8. Wani, J.A.; Sharma, S.; Muzamil, M.; Ahmed, S.; Sharma, S.; Singh, S. Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges. *Arch. Comput. Methods Eng.* **2022**, *29*, 641–677. [CrossRef]
9. Muruganantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sens.* **2022**, *14*, 1990. [CrossRef]
10. Thirunavukarasu, R.; Gnanasambandan, R.; Gopikrishnan, M.; Palanisamy, V. Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review. *Comput. Biol. Med.* **2022**, *149*, 106020. [CrossRef]
11. Khalil, R.A.; Saeed, N.; Masood, M.; Fard, Y.M.; Alouini, M.S.; Al-Naffouri, T.Y. Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet Things J.* **2021**, *8*, 11016–11040. [CrossRef]
12. Younan, M.; Houssein, E.H.; Elhoseny, M.; Ali, A.A. Challenges and recommended technologies for the industrial internet of things: A comprehensive review. *Measurement* **2020**, *151*, 107198. [CrossRef]
13. Peng, S.; Cao, L.; Zhou, Y.; Ouyang, Z.; Yang, A.; Li, X.; Jia, W.; Yu, S. A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Netw.* **2022**, *8*, 745–762. [CrossRef]
14. Ni, J.; Young, T.; Pandelea, V.; Xue, F.; Cambria, E. Recent advances in deep learning based dialogue systems: A systematic survey. *Artif. Intell. Rev.* **2023** , *56*, 3055–3155. [CrossRef]
15. Geetha, S.; Abhishek, C.; Akshayanat, C. Machine vision based fire detection techniques: A survey. *Fire Technol.* **2021**, *57*, 591–623. [CrossRef]
16. Kong, S.G.; Jin, D.; Li, S.; Kim, H. Fast fire flame detection in surveillance video using logistic regression and temporal smoothing. *Fire Saf. J.* **2016**, *79*, 37–43. [CrossRef]
17. Filonenko, A.; Hernandez, D.C.; Wahyono.; Jo, K.H. Smoke detection for surveillance cameras based on color, motion, and shape. In Proceedings of the 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), Poitiers, France, 19–21 July 2016; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]
18. Govil, K.; Welch, M.L.; Ball, J.T.; Pennypacker, C.R. Preliminary results from a wildfire-detection system using deep learning on remote camera images. *Remote Sens.* **2020**, *12*, 166. [CrossRef]
19. Ye, S.; Bai, Z.; Chen, H.; Bohush, R.; Ablameyko, S. An effective algorithm to detect both smoke and flame using color and wavelet analysis. *Pattern Recognit. Image Anal.* **2017**, *27*, 131–138. [CrossRef]
20. Lin, G.; Zhang, Y.; Zhang, Q.; Jia, Y.; Xu, G.; Wang, J. Smoke detection in video sequences based on dynamic texture using volume local binary patterns. *KSII Trans. Internet Inf. Syst.* **2017**, *11*, 5522–5536.

21. Muhammad, K.; Khan, S.; Elhoseny, M.; Hassan Ahmed, S.; Wook Baik, S. Efficient Fire Detection for Uncertain Surveillance Environment. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3113–3122. [CrossRef]
22. 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 Sensors J.* **2019**, *19*, 3191–3202. [CrossRef]
23. Xue, X.; Feiniu, Y.; Lin, Z.; Longzhen, Y.; Jinting, S. From traditional methods to deep ones: Review of visual smoke recognition, detection, and segmentation. *J. Image Graph.* **2019**, *24*, 1627–1647.
24. Bu, F.; Gharajeh, M.S. Intelligent and vision-based fire-detection systems: A survey. *Image Vis. Comput.* **2019**, *91*, 103803. [CrossRef]
25. 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]
26. 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]
27. 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]
28. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Remote Sensing: Classification, Detection, and Segmentation. *Remote Sens.* **2023**, *15*, 1821. [CrossRef]
29. Cetin, A.E. Computer Vision Based Fire Detection Dataset. 2014. Available online: http://signal.ee.bilkent.edu.tr/VisiFire/ (accessed on 3 August 2023).
30. Dharmawan, A.; Harjoko, A.; Adhinata, F.D. Region-based annotation data of fire images for intelligent surveillance system. *Data Brief* **2022**, *41*, 107925.
31. Chino, D.Y.; Avalhais, L.P.; Rodrigues, J.F.; Traina, A.J. 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; IEEE: Piscataway, NJ, USA, 2015; pp. 95–102.
32. 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]
33. Braović, M.; Stipaničev, D.; Krstinić, D. FESB MLID Dataset. 2018. Available online: http://wildfire.fesb.hr/index.php?option=com_content&view=article&id=66%20&Itemid=76 (accessed on 3 August 2023).
34. Cheng, H.Y.; Yin, J.L.; Chen, B.H.; Yu, Z.M. Smoke 100k: A Database for Smoke Detection. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 596–597.
35. Yuan, F. Video Smoke Detection Dataset. 2020. Available online: http://staff.ustc.edu.cn/~yfn/vsd.html (accessed on 3 August 2023).
36. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Comput. Netw.* **2021**, *193*, 108001. [CrossRef]
37. Wang, M.; Jiang, L.; Yue, P.; Yu, D.; Tuo, T. FASDD: An Open-access 100,000-level Flame and Smoke Detection Dataset for Deep Learning in Fire Detection. *Earth Syst. Sci. Data Discuss.* **2023**, published online. . [CrossRef]
38. de Venâncio, P.V.A.; Lisboa, A.C.; Barbosa, A.V. An automatic fire-detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Comput. Appl.* **2022**, *34*, 15349–15368. [CrossRef]
39. Gong, X.; Hu, H.; Wu, Z.; He, L.; Yang, L.; Li, F. Dark-channel based attention and classifier retraining for smoke detection in foggy environments. *Digit. Signal Process.* **2022**, *123*, 103454. [CrossRef]
40. Wu, S.; Zhang, X.; Liu, R.; Li, B. A dataset for fire and smoke object detection. *Multimed. Tools Appl.* **2023**, *82*, 6707–6726. [CrossRef]
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
42. 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, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
45. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
46. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
47. Muhammad, K.; Ahmad, J.; Lv, Z.; Bellavista, P.; Yang, P.; Baik, S.W. Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications. *IEEE Trans. Syst. Man, Cybern. Syst.* **2019**, *49*, 1419–1434. [CrossRef]
48. Yuan, F.; Zhang, L.; Wan, B.; Xia, X.; Shi, J. Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition. *Mach. Vis. Appl.* **2019**, *30*, 345–358. [CrossRef]

49. Khudayberdiev, O.; Zhang, J.; Abdullahi, S.M.; Zhang, S. Light-FireNet: An efficient lightweight network for fire detection in diverse environments. *Multimed. Tools Appl.* **2022**, *81*, 24553–24572. [CrossRef]
50. 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]
51. Majid, S.; Alenezi, F.; Masood, S.; Ahmad, M.; Gündüz, E.S.; Polat, K. Attention based CNN model for fire detection and localization in real-world images. *Expert Syst. Appl.* **2022**, *189*, 116114. [CrossRef]
52. Tao, H.; Duan, Q. An adaptive frame selection network with enhanced dilated convolution for video smoke recognition. *Expert Syst. Appl.* **2023**, *215*, 119371. [CrossRef]
53. Khan, Z.A.; Hussain, T.; Ullah, F.U.M.; Gupta, S.K.; Lee, M.Y.; Baik, S.W. Randomly Initialized CNN with Densely Connected Stacked Autoencoder for Efficient Fire Detection. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105403. [CrossRef]
54. Huang, L.; Liu, G.; Wang, Y.; Yuan, H.; Chen, T. Fire detection in video surveillances using convolutional neural networks and wavelet transform. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104737. [CrossRef]
55. Kwak, D.K.; Ryu, J.K. A Study on Fire Detection Using Deep Learning and Image Filtering Based on Characteristics of Flame and Smoke. *J. Electr. Eng. Technol.* **2023**, published online. . [CrossRef]
56. Hu, C.; Tang, P.; Jin, W.; He, Z.; Li, W. Real-Time Fire Detection Based on Deep Convolutional Long-Recurrent Networks and Optical Flow Method. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]
57. 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]
58. He, L.; Gong, X.; Zhang, S.; Wang, L.; Li, F. Efficient attention based deep fusion CNN for smoke detection in fog environment. *Neurocomputing* **2021**, *434*, 224–238. [CrossRef]
59. 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]
60. Zhang, Q.x.; Lin, G.h.; Zhang, Y.m.; Xu, G.; Wang, J.j. Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images. *Procedia Eng.* **2018**, *211*, 441–446. [CrossRef]
61. 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]
62. Barmpoutis, P.; Dimitropoulos, K.; Kaza, K.; Grammalidis, N. Fire detection from images using faster R-CNN and multidimensional texture analysis. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 8301–8305.
63. Chaoxia, C.; Shang, W.; Zhang, F. Information-guided flame detection based on faster r-cnn. *IEEE Access* **2020**, *8*, 58923–58932. [CrossRef]
64. An, Q.; Chen, X.; Zhang, J.; Shi, R.; Yang, Y.; Huang, W. A robust fire-detection model via convolution neural networks for intelligent robot vision sensing. *Sensors* **2022**, *22*, 2929. [CrossRef]
65. Chen, C.; Yu, J.; Lin, Y.; Lai, F.; Zheng, G.; Lin, Y. Fire detection based on improved PP-YOLO. *Signal Image Video Process.* **2023**, *17*, 1061–1067. [CrossRef]
66. Mukhiddinov, M.; Abdusalomov, A.B.; Cho, J. Automatic Fire Detection and Notification System Based on Improved YOLOv4 for the Blind and Visually Impaired. *Sensors* **2022**, *22*, 3307. [CrossRef] [PubMed]
67. Chen, G.; Zhou, H.; Li, Z.; Gao, Y.; Bai, D.; Xu, R.; Lin, H. Multi-Scale Forest Fire Recognition Model Based on Improved YOLOv5s. *Forests* **2023**, *14*, 315. [CrossRef]
68. Kristiani, E.; Chen, Y.C.; Yang, C.T.; Li, C.H. Flame and smoke recognition on smart edge using deep learning. *J. Supercomput.* **2023**, *79*, 5552–5575. [CrossRef]
69. Wu, Z.; Xue, R.; Li, H. Real-Time Video Fire Detection via Modified YOLOv5 Network Model. *Fire Technol.* **2022**, *58*, 2377–2403. [CrossRef]
70. Yin, H.; Chen, M.; Fan, W.; Jin, Y.; Hassan, S.G.; Liu, S. Efficient Smoke Detection Based on YOLO v5s. *Mathematics* **2022**, *10*, 3493. [CrossRef]
71. Yan, C.; Wang, Q.; Zhao, Y.; Zhang, X. YOLOv5-CSF: An improved deep convolutional neural network for flame detection. *Soft Comput.* **2023**, published online. [CrossRef]
72. Huo, Y.; Zhang, Q.; Jia, Y.; Liu, D.; Guan, J.; Lin, G.; Zhang, Y. A deep separable convolutional neural network for multiscale image-based smoke detection. *Fire Technol.* **2022**, *58*, 1445–1468. [CrossRef]
73. Wu, H.; Wu, D.; Zhao, J. An intelligent fire-detection approach through cameras based on computer vision methods. *Process Saf. Environ. Prot.* **2019**, *127*, 245–256. [CrossRef]
74. de Venâncio, P.V.A.; Campos, R.J.; Rezende, T.M.; Lisboa, A.C.; Barbosa, A.V. A hybrid method for fire detection based on spatial and temporal patterns. *Neural Comput. Appl.* **2023**, *35*, 9349–9361. [CrossRef]
75. Li, C.; Cheng, D.; Li, Y. Research on fire detection algorithm based on deep learning. In Proceedings of the International Conference on Cloud Computing, Performance Computing, and Deep Learning (CCPCDL 2022), Wuhan, China, 11–13 March 2022; SPIE: Cergy, France, 2022; Volume 12287, pp. 510–514.

76. Jandhyala, S.S.; Jalleda, R.R.; Ravuri, D.M. Forest Fire Classification and Detection in Aerial Images using Inception-V3 and SSD Models. In Proceedings of the 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 5–7 January 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 320–325.
77. 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] [PubMed]
78. Xie, Y.; Zhu, J.; Cao, Y.; Zhang, Y.; Feng, D.; Zhang, Y.; Chen, M. Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features. *IEEE Access* **2020**, *8*, 81904–81917. [CrossRef]
79. Huo, Y.; Zhang, Q.; Zhang, Y.; Zhu, J.; Wang, J. 3DVSD: An end-to-end 3D convolutional object detection network for video smoke detection. *Fire Saf. J.* **2022**, *134*, 103690. [CrossRef]
80. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]
81. Li, Y.; Zhang, W.; Liu, Y.; Jing, R.; Liu, C. An efficient fire and smoke detection algorithm based on an end-to-end structured network. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105492. [CrossRef]
82. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based attention module. *arXiv* **2021**, arXiv:2111.12419.
83. Yang, C.; Pan, Y.; Cao, Y.; Lu, X. CNN-Transformer Hybrid Architecture for Early Fire Detection. In Proceedings of the Artificial Neural Networks and Machine Learning—ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, 6–9 September 2022; Part IV; Springer: Berlin, Germany, 2022; pp. 570–581.
84. Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.
85. 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]
86. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3431–3440.
87. Mseddi, W.S.; Ghali, R.; Jmal, M.; Attia, R. Fire detection and segmentation using YOLOv5 and U-net. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; IEEE: Piscatway, NJ, USA, 2021; pp. 741–745.
88. 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]
89. Wang, Z.; Yang, P.; Liang, H.; Zheng, C.; Yin, J.; Tian, Y.; Cui, W. Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery. *Remote Sens.* **2022**, *14*, 45. [CrossRef]
90. 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; SPIE: Cergy, France, 2020; Volume 11533, pp. 134–145.
91. 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]
92. Khan, S.; Muhammad, K.; Hussain, T.; Del Ser, J.; Cuzzolin, F.; Bhattacharyya, S.; Akhtar, Z.; de Albuquerque, V.H.C. Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments. *Expert Syst. Appl.* **2021**, *182*, 115125. [CrossRef]
93. Yuan, F.; Zhang, L.; Xia, X.; Huang, Q.; Li, X. A gated recurrent network with dual classification assistance for smoke semantic segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 4409–4422. [CrossRef] [PubMed]
94. Shahid, M.; Virtusio, J.J.; Wu, Y.H.; Chen, Y.Y.; Tanveer, M.; Muhammad, K.; Hua, K.L. Spatio-Temporal Self-Attention Network for Fire Detection and Segmentation in Video Surveillance. *IEEE Access* **2021**, *10*, 1259–1275. [CrossRef]
95. Yuan, F.; Dong, Z.; Zhang, L.; Xia, X.; Shi, J. Cubic-cross convolutional attention and count prior embedding for smoke segmentation. *Pattern Recognit.* **2022**, *131*, 108902. [CrossRef]
96. Li, Y.; Zhang, W.; Liu, Y.; Shao, X. A lightweight network for real-time smoke semantic segmentation based on dual paths. *Neurocomputing* **2022**, *501*, 258–269. [CrossRef]
97. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
98. 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*, 120. [CrossRef]
99. Yuan, F.; Li, K.; Wang, C.; Fang, Z. A Lightweight Network for Smoke Semantic Segmentation. *Pattern Recognit.* **2023**, *137*, 109289. [CrossRef]
100. Sudhakar, S.; Vijayakumar, V.; Kumar, C.S.; Priya, V.; Ravi, L.; Subramaniyaswamy, V. Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires. *Comput. Commun.* **2020**, *149*, 1–16. [CrossRef]
101. 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]
102. 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] [PubMed]
103. 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]

104. Chen, J.; Ran, X. Deep learning with edge computing: A review. *Proc. IEEE* **2019**, *107*, 1655–1674. [CrossRef]
105. Kamath, V.; Renuka, A. Deep Learning Based Object Detection for Resource Constrained Devices-Systematic Review, Future Trends and Challenges Ahead. *Neurocomputing* **2023**, *531*, 34–60. [CrossRef]