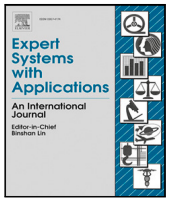
Expert Systems With Applications 255 (2024) 124783

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Review

Fire and smoke detection from videos: A literature review under a novel taxonomy

Diego Gragnaniello [a,*,1], Antonio Greco [a,1], Carlo Sansone [b,1], Bruno Vento [b,1]

[a] Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy
[b] Department of Electrical Engineering and Information Technology (DIETI), University of Napoli Federico II, Italy

ARTICLE INFO

Dataset link: https://github.com/MiviaLab/FireDetectionSurvey2023/blob/main/MIVIA_FIRE_DETECTION_VIDEO_ANNOTATIONS.xlsx

Keywords:
Fire detection
Smoke
Flame
Image
Video
Camera
Survey
Review

ABSTRACT

The recent development of deep learning based fire detection techniques and the availability of smart cameras able to execute these algorithms on the edge paved the way for sophisticated and efficient video-based firefighting systems. However, the limited available data to train these algorithms cast shadows on their robustness and generalization capability. In this survey, we review 153 papers published in the literature and 17 publicly available fire detection datasets with the aim of identifying application scenarios that better describe real-world fire detection challenges. In the proposed taxonomy, these are characterized by two features: i) the fire size in the framed scene that depends on several parameters, foremost the distance from the fire but also the camera optic; ii) the background activity, due to the presence of moving objects that may mislead the detector. On this basis, we analyzed the existing methods under a common scheme according to this new taxonomy and matched the solutions with the needs of specific application scenarios. Similarly, for 9 interesting video datasets acquired from cameras, we labeled 536 videos according to the proposed taxonomy and shared these annotations with the community. The aim of this fire detection review is two-fold: on one hand, we classify the existing scientific works according to the real application scenarios, determining the features that are promising in specific operative conditions; on the other hand, we provide a detailed analysis and annotation of available datasets to promote the development of more reliable validation protocols and the collection of data from missing scenarios.

Contents

1. Introduction 2
2. Recent survey works 3
3. Architecture of methods 4
   3.1. Fire region proposal 5
   3.2. Fire recognition 5
4. Scenario based taxonomy 7
   4.1. Short range with low activity scenario 8
   4.2. Long range with low activity scenario 10
   4.3. Short range with high activity scenario 12
   4.4. Long range with high activity scenario 13
5. Datasets and performance 14
   5.1. Datasets 14
   5.2. Performance comparison 14
      5.2.1. Static camera datasets 15
      5.2.2. Moving camera datasets 15
6. Conclusions 16
7. Future works 16

* Corresponding author.
E-mail addresses: digragnaniello@unisa.it (D. Gragnaniello), agreco@unisa.it (A. Greco), carlo.sansone@unina.it (C. Sansone), bruno.vento@unina.it (B. Vento).

[1] All the authors contributed equally to the work.

https://doi.org/10.1016/j.eswa.2024.124783
Received 15 March 2024; Received in revised form 6 July 2024; Accepted 9 July 2024
Available online 14 July 2024

7.1. Dataset collection 16
7.2. Experimental protocols 16
7.3. Methodological design choices 16
Declaration of competing interest 17
Data availability 17
References 17

1. Introduction

Nowadays, the wide spread of low cost surveillance cameras and, in particular, smart cameras capable of running lightweight processes (Chang, Liu, Xiong, Cai, & Tu, 2021), make them suitable to be employed in several important tasks, from public safety (Zhang, Sun, Wu, & Zhong, 2019) to anomaly detection (Santhosh, Dogra, & Roy, 2020). Among the latter class of applications, fire detection is one of the most important (Di Lascio, Greco, Saggese, & Vento, 2014). For its frequency and danger, fire is among the most devastating and costly adverse phenomena for both human life and natural resources (Akdis & Nadeau, 2022). This is even getting worse due to climate change that impacts a large part of our planet (Nolan et al., 2021). More and more regions are facing for the first time very dry seasons that favor the spread of fire (Halofsky, Peterson, & Harvey, 2020).

The literature about artificial vision applications for fire detection considers two scenarios: *wildfire* and *urban* (Çetin et al., 2013; Gaur, Singh, Kumar, Kumar, & Kapoor, 2020). However, they are too general and do not represent well real world applications. In the wildfire scenario, the scene is acquired either by far cameras framing a wide wooded area from a top hill or from a close view, like the rural area surrounding a building to monitor. The scene is rather static, or small movements cannot be perceived due to the acquisition distance. This has been the main focus of the scientific literature for several years (Abid, 2021; Bouguettaya, Zarzour, Taberkit, & Kechida, 2022; Komarasamy, Gokuldhev, Hermina, Gokulapriya, & Manju, 2020), with plenty of data acquired and made available to the community. Some examples are shown in Fig. 1 (top). The other scenario considered in the literature so far is fire detection in urban areas (Park & Ko, 2020), some examples of which are depicted in Fig. 1 (bottom). It is worth pointing out that in urban scenario the scene may be acquired either at mid-to-short distance, of which indoor acquisition can be considered as a limit case, or exploiting a wider view, for example with cameras installed on top of a building. The urban scenario, with respect to the wildfire scenario, is characterized by a higher activity due to moving people and/or vehicles and more frequent occlusions or moving fire-like objects (e.g., light reflections, car fumes) that can impair the performance of the detector. Even if these two scenarios define the environment in which fire detection is carried out, they are too broad to fully characterize the needs and challenges of real world applications (Gragnaniello, Greco, Sansone, & Vento, 2023). As an example, the same wildfire scene can appear dissimilar when framed at a different distance or with a different camera optic. At the same time, particular application settings result in similar characteristics in different environmental scenarios Çetin et al. (2013). Rural areas surrounding man made infrastructures can exhibit moving objects similar to an urban scenario. On the contrary, the latter can have barely any or few moving objects in some residential areas or day hours. These two scenarios are of little help even when defining Key Performance Indicators (KPIs). For example, the fire detection delay is very important to measure the early fire detection capability of the system. A KPI cannot be defined for any of the two scenarios, since it indirectly depends on the size of the fire. While at short distances or with zoom on the specific area of interest flames are visible since the ignition stage, at a long distance or with a large field of view fire cannot be detected before a visible smoke column is produced. For the same reason, the two scenarios even fail to help in selecting which of the fire traces the detector should exploit, either the flame or smoke. In the same scenario, with a sufficient fire size, both flames and smoke can be detected; meanwhile, at far distances or with large field of view, only the smoke is visible (i.e., flames are perceived at a later stage when fire detection is trivial).

Grounded in these motivations, we propose to reconsider the fire detection scenarios from an application point of view that characterizes the factors having a greater impact on real world fire detection performance. Our aim is to propose a novel scenario taxonomy to promote the development of methods specific to each real world application. Our taxonomy is based on two main features that better characterize most of the real world fire detection applications: the *size* of the fire and the *activity* of the scene. Given a specific camera optic, the pixel size of visible flames and smoke depends on the distance from the fire and on the resolution; the greater the distance and the smaller the resolution, the smaller the pixel size and, thus, the more challenging the fire detection. Regarding the activity, the more movement from people, vehicles, and other objects is in the scene, the higher the probability that false positives are raised, or false negatives are generated due to occlusions. This taxonomy aims to identify, for each method, the features that best fit a specific application scenario and to compare methods on specific fire detection tasks that pose challenges of incomparable levels of difficulty.

To foster the development of methods ready for real world applications, we also carefully analyzed the available datasets and annotated the samples according to the proposed scenario taxonomy. We focused on video datasets, which allow us to analyze both the static and the dynamic information, considering datasets acquired by both a static or moving camera. Even if the latter group is not representative of a surveillance system, moving cameras are widely used in the literature and can help to deal with the scarcity of data affecting the research about fire detection. Firstly, this analysis highlighted the need for fire detection datasets in scenarios with high background activity, which has been only marginally studied in previous works despite being strongly required by the market (e.g., for fire detection systems in tunnels, landfills, production lines, and so on). Secondly, the description of available datasets, reflecting the original wildfire/urban scenarios, highlights their limited size and heterogeneous nature, which prevents to validate methods in specific challenging situations, such as the detection of fires occupying a small portion of the image or in the presence of a strong background activity, separately. This critical aspect is emphasized by common practices, like merging together small datasets that increase the number of samples to the detriment of the interpretation of the obtained results. To optimize the performance on such heterogeneous datasets, more general fire detector methods emerged. These approaches are robust against variations in the scenario characteristics, which is mostly useless for a real world application. Instead, given the critical fire detection task, a reliable method working in a real world application must be designed for the specific scenario in which it is deployed, thus exploiting its pros and trying to prevent or address its challenges. Moreover, even if these general purpose methods excel in a specific real condition, this cannot be verified due to the lack of a dedicated validation protocol.

To summarize the impact of the novel taxonomy on both the description of the methods and the usage of the datasets, the contributions of our work can be summarized as follows:

- We propose a novel taxonomy of application scenarios that highlights the different challenging levels encountered, thus allowing a more fair comparison of the methods.

Fig. 1. Examples of fire images available in datasets for wildfire and urban scenarios.

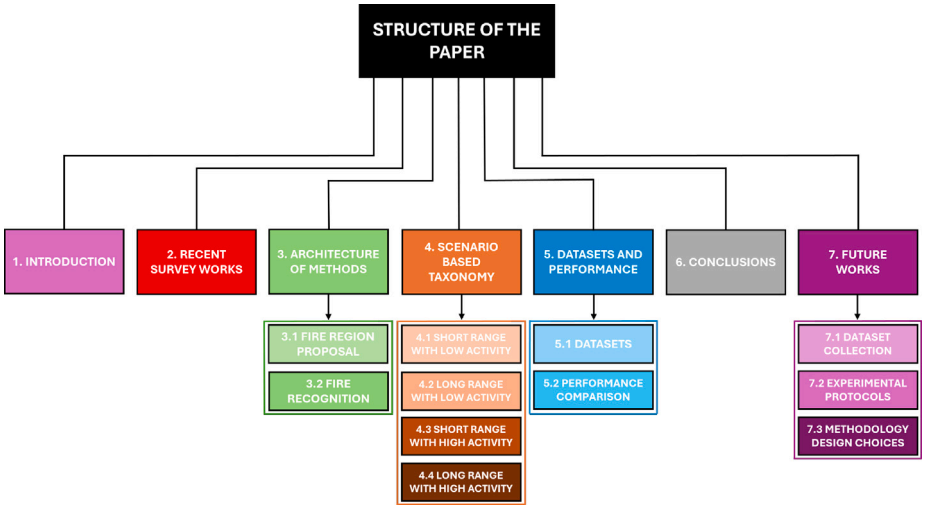


Fig. 2. Organization of the paper in chapters and paragraphs.

- We provide a common scheme to analyze fire detection methods and point out the features that best match the needs of each of the identified application scenarios.
- We provide video-level annotations for available datasets according to the novel taxonomy to highlight the limits of the available datasets and promote the definition of more reliable application oriented validation protocols.

The remainder of the manuscript is organized as shown in Fig. 2. In Section 2 we describe the recent survey works about fire detection and highlight the differences with our review. In Section 3 we describe the architecture of fire detection methods, giving details on the typically adopted region proposal and recognition algorithms. In Section 4 we present our application oriented scenario based taxonomy and its difference with respect to the one available in the literature; we also separately analyze the considered scenarios and discuss the methods that best fit the specific needs of each scenario. In Section 5 we revise the available datasets, pigeonholing them in the proposed taxonomy, and compare the performance achieved by top-scoring techniques. Finally, we draw conclusions and possible future directions in Section 6 and Section 7, respectively.

2. Recent survey works

The last few years have seen the emergence of several studies that examine fire detection methods from various perspectives (Abid, 2021; Bouguettaya et al., 2022; Bu & Gharajeh, 2019; Çetin et al., 2013; Gaur et al., 2019, 2020; Geetha, Abhishek, & Akshayanat, 2021; Jin et al., 2023; Komarasamy et al., 2020). In particular, Gaur et al. (2019) describe the analog sensors that can be used for flame and smoke detection: optical dispersion sensors, ion sensors, heat sensors, gas sensors. These sensors are very useful when detecting indoor fires, but can rarely be employed in outdoor applications, which are the focus of most of the recent literature works. Among them, methods that analyze visual data collected from digital sensors (cameras) focus their attention on wildfire detection (Abid, 2021; Bouguettaya et al., 2022; Komarasamy et al., 2020), while others limit the analysis to the ground video surveillance (Bu & Gharajeh, 2019; Çetin et al., 2013; Gaur et al., 2020; Geetha et al., 2021; Jin et al., 2023) regardless of the environmental scenario. In the first group, the recent work (Bouguettaya et al., 2022) is devoted to a very specific application scenario, that is wildfire detection from drones or unmanned aerial vehicles Harkat, Nascimento, Bernardino, and Ahmed (2023) and Yang et al. (2023). Most of the techniques designed to discover early fire traces in aerial videos are characterized by two processing stages: the first one is carried out on board the vehicle and the second one is performed at the ground base station. Instead, in Komarasamy et al. (2020) the authors focus on wildfire detection, providing a taxonomy of all the technologies that can be involved in this task. Together with image based methods, which are grouped into classical approaches, machine learning based ones, and the more recent deep learning methods, the survey also presents those works exploiting sensor networks. Similarly, in Abid (2021) a wide review of machine learning methods for the detection of forest fires is presented. This survey gives the reader an articulated view of the literature methods, by presenting approaches working on different inputs. Some of them perform fire detection in videos acquired by surveillance cameras, while others exploit thermal cameras or even satellite sensors. Also, some approaches fuse visual

information with weather data to improve fire prediction performance. Even if these surveys are useful for a rapid and broad view of the wildfire detection literature, the comparison among such very different approaches allows us to retrieve general information only, without delving into the details that make a technique more suitable for one scenario or another.

In our work, we are interested in ground video surveillance fire detection techniques, which already consist of a wide and heterogeneous set of methods. This is motivated by the wide spread of surveillance cameras and, in particular, smart cameras with onboard computation systems that may process the acquired video in real time. One of the earlier survey work has been presented in Çetin et al. (2013). The authors separately discuss the methods designed for fire detection from those dedicated to wildfires. In each group, methods using infrared technology are told apart from those using the visible spectrum only. Since this work was before the spread of deep learning, all the methods surveyed propose handcrafted features to enhance or detect the flames or the smoke in the images. Although interesting, the detectors built upon these features cannot get rid of the variations in appearance due to environmental conditions, different combustibles, or even acquisition configurations. Thus, they have been significantly surpassed by recent ones mostly based on deep learning. It is worth noting that the authors of Çetin et al. (2013) discuss the impact of the distance from the camera to the fire and the minimum fire size detectable for each distance. In our work, we revisit this factor making it a key feature characterizing the application scenarios of the proposed taxonomy. More recent similar survey works (Bu & Gharajeh, 2019; Gaur et al., 2020; Geetha et al., 2021; Jin et al., 2023) reviewed these methods focusing their attention on methodological aspects rather than the specific application scenario. In Gaur et al. (2020) the authors present a wide analysis of video-based fire and smoke detection methods. Their work highlights the improvements brought by deep learning methods when compared with classical handcrafted feature approaches. At the same time, the last part of the work presents hybrid approaches. By putting the human experience in the design of descriptive features used by deep neural networks, these techniques partially exploit the representation learning capabilities of deep neural networks while reducing the need for huge amounts of training data. Following a similar path, in a more recent survey (Geetha et al., 2021) the authors focus their attention on the methodological comparison among approaches based on handcrafted features and deep learning. The first group of methods includes approaches able to detect flame or smoke with ad hoc features designed for each specific task. Meanwhile, the second group of methods is composed of more generic image classifiers and object detectors trained using fire detection datasets. In Bu and Gharajeh (2019) several methods proposed for wildfire detection or other scenarios are separately discussed. For each of these scenarios, the authors report the performance achieved by rule based approaches, expert systems, and deep learning based methods. The separate study of methods based either on flame and smoke, as well as handcrafted and deep learning approaches, is interesting from a scientific perspective, but it is of limited utility to design real world applications. Instead, the authors of Jin et al. (2023) mostly focused on deep learning based techniques, which are grouped by the addressed task, that is fire recognition, fire object detection, and fire segmentation. The authors include in the discussion methods designed to deal with points of view different from that of a surveillance camera, too, like taken from drones or lookout towers, regardless of the environment where the acquisition takes place, e.g., urban, industrial, or rural. Finally, in Chaturvedi, Khanna, and Ojha (2022) the authors focus on real-time smoke detection by using image analysis, machine learning, and deep learning approaches. The survey provides a comprehensive review of smoke detection systems and the related datasets. In this work, we analyze both flame and smoke detectors, and tell which are more suitable for each application scenario.

Starting from the above mentioned surveys, in this work we conduct a thorough analysis of video based methods, including those based on advanced deep learning approaches, like recurrent networks to extract dynamic information from the video-based approaches to fully exploit the spatial contextual information. Differently from previous works, our analysis starts from the characteristics of application scenarios to define a novel taxonomy that points out the challenges and needs of each scenario. We aim to help the reader to better identify those methods that are more suitable to address the challenges posed by each of the application scenarios under consideration. In addition, we analyze the characteristics of available video datasets to provide the annotations that allow to build more specific and reliable validation protocols.

In summary, jointly reviewing the scenarios, the methods, and the datasets under a novel application oriented taxonomy can help to highlight the shortcomings of the present research direction and promote the development of approaches and benchmarks that may allow to fill the gap with real world applications.

3. Architecture of methods

In the last years, several rule based and data driven works have been presented to address fire detection. In this section, we review the more recent ones, especially those based on deep learning that set the new state of the art, adopting the scheme presented in Table 1. In more detail, two phases are distinguished, namely fire region proposal and recognition.

Fire region proposal is an optional phase that aims to locate fire candidates in the scene thus simplifying the subsequent fire recognition phase, e.g., by removing the background or segmenting/detecting the image regions that possibly contain fires. We include in this phase the ad hoc preprocessing operations or handcrafted local image descriptors proposed to enhance the flame or smoke. Fire recognition is the actual classification phase, carried out either on the whole image or thereof part previously collected, exploiting static information coming from a single frame or extracting the dynamic information of consecutive frames of the video.

These phases are impacted by two main design features of the methods, the *working resolution* and the respective *trace*, whether it be flames or smoke, they seek. Working resolution is a crucial design parameter to spot fires, particularly when the fire size in pixels is small since the monitored area is framed far away or the field of view is large. Due to hardware limitations during the inference phase, most of the early works resize the images to reduce the computational load. The working resolutions of the methods may vary a lot, starting from 48×48 (Gu, Xia, Qiao, & Lin, 2019; Shi, Lu, & Cui, 2019; Yuan, Zhang, Wan, Xia and Shi, 2019) or 64×64 (Jadon, Omama, Varshney, Ansari, & Sharma, 2019; Khudayberdiev, Zhang, Abdullahi & Zhang, 2022), until over 600×600 (An et al., 2022; Choi, Jeon, Song, & Kang, 2021; de Venâncio, Campos, Rezende, Lisboa, & Barbosa, 2023; Dunnings & Breckon, 2018; Frizzi, Bouchouicha, Ginoux, Moreau, & Sayadi, 2021; Huang, He, Guan, & Zhang, 2023; Huo, Zhang, Jia et al., 2022; Jiang, Zhao, Yu, Zhou, & Peng, 2022; Wu, Xue, & Li, 2022). On average, an intermediate resolution is adopted, i.e. between 200×200 and 300×300 Harkat et al. (2023). The information loss due to resizing, which may prevent to spot fires at an early stage that appear too small in the resized images, has been avoided by carrying out the fire region proposal phase at a higher resolution, then crop and resize the suspected regions to be classified (Aktas, Bayramcavus, & Akgun, 2019; Cao, Tang, Wu, & Lu, 2021; Cao, Yang, Tang, & Lu, 2019; Luo, Zhao, Liu, & Huang, 2018; Nguyen, Vu, Pham, Choi, & Ro, 2021; Shahid, Chien et al., 2021; Wang, Zhang, & Zhu, 2021; Xie et al., 2020). More recently, transformer-based approaches able at handling full resolution videos have been presented (Dewangan et al., 2022; Yazdi, Qin, Jordan, Yang, & Yan, 2022). However, these techniques are still resource demanding, and their application on smart cameras may require non trivial engineering.

Table 1
Methodologies for fire region proposal and recognition.

Processing phase	Approaches
Fire region proposal	Rule based fire segmentation based on color or movement features Background subtraction Learning based (object detection, semantic segmentation)
Fire recognition	Handcrafted features with rule based classifiers or expert systems 2D Convolutional Neural Networks (2D CNNs) Vision Transformers Spatio-temporal handcrafted features 3D Convolutional Neural Networks (3D CNNs) Recurrent Neural Networks (RNNs)

The other method feature regards the particular fire trace, either the flames or the smoke, seek in the scene. Some recent machine learning techniques are general and can be trained to spot both traces. This approach requires more training samples and theoretically leads to optimal performance. Other methods, either based on handcrafted features or not, customize one or both fire region proposal and recognition phases to spot either the flames or the smoke. Flames and smoke features greatly differ, thus making one approach designed for one of them unable to detect the other; this is true especially for methods based on handcrafted features (Celik & Demirel, 2009; Celik, Demirel, Ozkaramanli, & Uyguroglu, 2007; Dimitropoulos, Barmpoutis, & Grammalidis, 2014; Harkat et al., 2023; Liu & Ahuja, 2004; Prema, Suresh, Krishnan, & Leema, 2022; Pundir & Raman, 2019; Sheng, Deng, & Xiang, 2021; Shi, Wang, Gao, & Yu, 2020; Torabian, Pourghassem, & Mahdavi-Nasab, 2021; Töreyin, Dedeoğlu, Güdükbay, & Cetin, 2006; Xie et al., 2020; Yang et al., 2023; Zhong, Wang, Shi, & Gao, 2018). Among them, some methods characterize the fire, e.g., by detecting the flame flickering (Xie et al., 2020) or jointly representing the smoke appearance and movements (Shi et al., 2020) using the Local Binary Pattern and the Optical Flow, respectively. With the advent of deep learning, some works have trained neural networks to localize flames or smoke (Shahid, Chien et al., 2021; Yuan, Zhang, Xia et al., 2019; Zhang, Zhu, Wang, & Ling, 2021). Even when focusing on a single trace, these techniques can be easily re-trained to recognize a different trace. Most of them, however, let the network learn which trace to focus on, thus optimizing performance. However, due to the dataset's scarce representativeness, the network may focus on the most evident trace, e.g. flame, thus performing poorly on smoke-only images. The opposite happens when most of the samples are acquired at a long distance or with a large field of view since flames are not visible, thus the network automatically focuses on smoke Harkat et al. (2023) and Yang et al. (2023). These bias of deep learning methods are often unwanted and prevent to control the network behavior. On the contrary, other works customized either the network input (Abdusalomov, Islam, Nasimov, Mukhiddinov, & Whangbo, 2023) or by color weighting the network output by a custom loss function (Zhang, Zhang, Liu, Li, & Zhao, 2022) to detect a specific trace. Considering the limits of available datasets, these approaches can help to better control the network behavior on new data.

3.1. Fire region proposal

The fire region proposal goal is to collect fire candidates. Since the following phase relies on the outcome of this first step, its poor performance may yield to miss detection. This point is less critical for those approaches not providing a short list of fire candidates, but rather an enhancement version of the image (Li, Zhao, Zhang, & Hu, 2019; Pundir & Raman, 2019; Sheng et al., 2021; Shi et al., 2019, 2020) or, more recently, an attention mask (Cao, Tang, Xu, Li and Lu, 2022; Gong et al., 2022; Jiang et al., 2022; Khudayberdiev, Zhang, Elkhalil and Balde, 2022; Li, Yan, & Liu, 2020; Majid et al., 2022; Shahid & Hua, 2021; Shakhnoza, Sabina, Sevara, & Cho, 2022; Tao, Lu, Hu, Xin and Wang, 2022), including those based on self attention transformer mechanism (Dewangan et al., 2022; Khudayberdiev, Zhang, Elkhalil et al., 2022; Li, Zhang, Liu, Jing and Liu, 2022; Mardani, Vretos, & Daras, 2023; Shahid & Hua, 2021; Yang, Pan, Cao, & Lu, 2022; Yazdi et al., 2022). This soft fire region proposal guides the classification of the whole image carried out in the next recognition phase.

Conversely, most of the works employ explicit fire candidate selection. These can be grouped into three clusters, namely rule based segmentation, background subtraction, and data driven selection. Early proposed methods typically implement rule based fire segmentation techniques (Celik & Demirel, 2009; Chen, An, Yu, & Ban, 2021; Chen, Wu, & Chiou, 2004; Horng, Peng, & Chen, 2005; Liu & Ahuja, 2004; Marbach, Loepfe, & Brupbacher, 2006; Nguyen et al., 2021; Prema et al., 2022; Steffens, Botelho, & Rodrigues, 2016; Verstockt et al., 2012; Xu, Wanguo, Xinrui, Bin and Yuan, 2019; Zhong et al., 2018) like thresholding the image in the color space to isolate flames or detecting the smoke movement using ad hoc features. A complementary path pursued by several approaches (Cao et al., 2021, 2019; Celik et al., 2007; Dimitropoulos et al., 2014; Filonenko, Hernández and Jo, 2017; Foggia, Saggese, & Vento, 2015; Luo et al., 2018; Torabian et al., 2021; Töreyin et al., 2006; Wang et al., 2021; Xie et al., 2020, 2022) is to employ background subtraction techniques that distinguish static and moving regions in the scene by looking at a number of consecutive video frames. These techniques are more reliable and can better adapt to the particular application scenario, subject to wisely tuning the parameters of the background modeling technique. Recently, advanced techniques learn how to collect fire candidates from data. For example, fire segmentation neural networks are adopted for this purpose (Choi et al., 2021; Frizzi et al., 2021; Shahid, Chien et al., 2021; Shahid, Virtusio et al., 2021; Yuan, Zhang, Xia, Huang, & Li, 2021; Yuan, Zhang, Xia et al., 2019; Zhang et al., 2021). On the other hand, object detection architectures can be grouped into those adopting Region Proposal Networks inspired by R-CNN models (Abdusalomov et al., 2023; Cao, Tang & Lu, 2022; Chaoxia, Shang, & Zhang, 2020; Huang, Liu, Wang, Yuan, & Chen, 2022; Kim & Lee, 2019, 2019; Li, Mihaylova, & Yang, 2021; Lin, Zhang, Xu, & Zhang, 2019; Zeng, Lin, Qi, Zhao, & Wang, 2018), and those adopting single shot detectors, mainly using one of the YOLO versions (An et al., 2022; de Venâncio et al., 2023; de Venâncio, Lisboa, & Barbosa, 2022; Hogan et al., 2021; Hu et al., 2022; Huang et al., 2023; Huo, Zhang, Jia et al., 2022; Huo, Zhang, Zhang, Zhu & Wang, 2022; Liu et al., 2016; Park & Ko, 2020; Qian, Shi, Chen, Ma, & Huang, 2022; Saponara, Elhanashi, & Gagliardi, 2021; Wu et al., 2022; Zhang et al., 2022). In addition, a method inspired by the salient object detector (Liu & Han, 2016) has been presented in Xu, Zhang, Zhang et al. (2019). The main drawback of these approaches is the need for region proposal information during training, which takes a lot of time to collect and annotate large and representative datasets. Notwithstanding this, among all the fire region proposal approaches, the latter seems the most promising. This is motivated by the chance to train neural networks to jointly perform fire region proposal and recognition, thus optimizing the whole process, as described in the next section.

3.2. Fire recognition

Eventually exploiting the outcome of the fire region proposal phase, the fire recognition techniques extract spatial features from the single

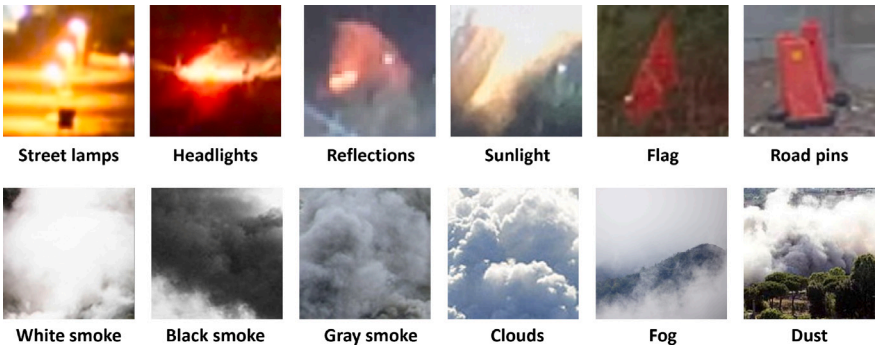


Fig. 3. Examples of challenges in flame and smoke recognition.

frame or spatiotemporal features from multiple frames to distinguish the fire from other elements of the scene. This processing step is anything but trivial, as is evident by observing the examples in Fig. 3. Flames can be confused with street lamps, headlights, reflections, sunlight, flags, road pins, and other objects whose color and appearance are similar to that of the flame. This difficulty is accentuated when working with small fires and in scenarios with high activity. Smoke is even more complex to recognize, as it can have different colors (i.e., white, black, gray) depending on the fuel, it is similar to various atmospheric elements (e.g., clouds, fog), and can be confused with dust raised by wind or moving objects and vehicles. Therefore, this processing step is complex and crucial. To this purpose, it is important to describe in detail the classifiers and the spatial or spatiotemporal features adopted by the methods to make the prediction.

Among the single frame approaches, the fire classifiers proposed in the scientific literature so far ranges from rule-based techniques and expert systems presented in the early works to the recently proposed classifiers based on visual transformers. Different flames and smoke deterministic models have been proposed, mostly among early works to recognize fire (Celik & Demirel, 2009; Celik et al., 2007; Chen et al., 2021, 2004; Foggia et al., 2015; Harkat et al., 2023; Horng et al., 2005; Liu & Ahuja, 2004; Marbach et al., 2006; Töreyin et al., 2006; Verstockt et al., 2012; Yang et al., 2023). On the same path, a recent work (Xie et al., 2022) addressed fire detection in the presence of occlusions exploiting light reflections through a multiexpert system. Except for this one, recent works exploit machine learning models for their superior performance and reliability. In Prema et al. (2022), smoke candidates are first collected utilizing a set of rules based on the smoke's appearance and movement. Then, the co-occurrence of the LBP descriptor is computed and an Extreme Learning Machine classifier is trained. Handcrafted features characterizing the object color, texture or capturing flickering oscillations and object size variation have been proposed in Dimitropoulos et al. (2014), Torabian et al. (2021) and Töreyin et al. (2006) and then classified by rule-based systems or using a binary Support Vector Machine (SVM). Similarly, texture and shape features are used by the authors of Steffens et al. (2016) to train a Random Forest classifier.

The vast majority of the approaches employed CNN models for image classification. We group them in heavyweight (Cao, Tang, Lu, 2022; Cao et al., 2021, 2019; Dunnings & Breckon, 2018; Filonenko, Kurnianggoro & Jo, 2017; Gong et al., 2022; Huo, Zhang, Zhang et al., 2022; Khan et al., 2022; Khan, Muhammad, Mumtaz, Baik, & de Albuquerque, 2019; Li et al., 2019; Muhammad, Ahmad, Mehmood, Rho and Baik, 2018; Pundir & Raman, 2019; Shahid, Chien et al., 2021; Shahid, Virtusio et al., 2021; Sharma, Granmo, Goodwin, & Fidje, 2017; Shi et al., 2019; Tao & Duan, 2023; Xu, Zhang, Liu et al., 2019; Zhang et al., 2020; Zhao, Zhang, & Man, 2020) and lightweight (Aktas et al., 2019; Cao, Tang, Xu et al., 2022; Jain & Srivastava, 2021; Majid et al., 2022; Muhammad, Ahmad, Lv et al., 2018; Muhammad, Khan, Elhoseny, Ahmed, & Baik, 2019; Nguyen et al., 2021; Oh, Ghyme, Jung, & Kim, 2020; Shi et al., 2020; Wang et al., 2021; Yang, Jang, Kim, & Lee, 2019; Yuan, Zhang, Wan et al., 2019) architectures, mostly borrowed by classical computer vision applications and adapted or fine tuned for the fire detection task. In other cases, a custom shallow CNN has been proposed (Almeida, Huang, Nogueira, Bhatia, & de Albuquerque, 2022; Ayala et al., 2020; Ayala, Fernandes, Cruz, Macêdo, & Zanchettin, 2022; Ghosh & Kumar, 2022; Gu et al., 2019; Hosseini, Hashemzadeh, & Farajzadeh, 2022; Jadon et al., 2019; Khudayberdiev, Zhang, Abdullahi et al., 2022; Li et al., 2020; Luo et al., 2018; Muhammad, Ahmad and Baik, 2018; Shakhnoza et al., 2022; Sheng et al., 2021; Xie et al., 2020; Xu, Wanguo et al., 2019; Xu, Zhang, Zhang et al., 2019; Yin, Lang, Li, Feng, & Wang, 2019; Yin, Wan, Yuan, Xia, & Shi, 2017; Yuan et al., 2021; Zhang et al., 2021; Zhong et al., 2018), which optimizes the network consumption when deployed on devices with limited resources. When trained on small datasets, all of these models processing the bare image with no fire region proposal are more prone to training biases. As an example, deep learning models typically perform a global average pooling on top of the convolutional stages, thus relating the final feature vector to classify with the fire relative size (in pixels) in the training samples. To avoid this, different pooling strategies should be considered (Körschens, Bodesheim, & Denzler, 2022). Other works addressed the whole fire detection pipeline through object detection neural networks that jointly locate and classify the object of interest. The techniques can be grouped into anchor based (Abdusalomov et al., 2023; Chaoxia et al., 2020; Dogan et al., 2022; Kim & Lee, 2019; Li et al., 2021; Lin et al., 2019; Zeng et al., 2018) or anchor free (de Venâncio et al., 2023; Hu et al., 2022; Huo, Zhang, Jia et al., 2022; Park & Ko, 2020; Saponara et al., 2021; Wu et al., 2022) approaches. Also for these methods, particular attention has been posed to the reduction of the complexity of such architectures. This is achieved by adopting lightweight backbones (Hogan et al., 2021; Huang et al., 2023, 2022; Jiang et al., 2022; Li, Zhang, Liu & Jin, 2022; Zhang et al., 2022) or pruning the YOLO architecture (An et al., 2022; de Venâncio et al., 2022; Qian et al., 2022). Very recently, transformers have been employed (Dewangan et al., 2022; Khudayberdiev, Zhang, Elkhalil et al., 2022; Li, Zhang, Liu, Jing et al., 2022; Mardani et al., 2023; Shahid & Hua, 2021; Yazdi et al., 2022) to extract high level and long range contextual information from the image, aiming to better distinguish between fire-like objects and fires.

Temporal analysis can help to improve recognition accuracy in challenging operative conditions. As an example, small crops around fire-like or smoke-like objects (e.g., the sunset lights or the fog and the clouds, respectively) may confuse even the human operator. Meanwhile, looking at a few consecutive frames may solve these ambiguities. Apart from majority voting schemes that can be built on top of frame wise detections, the simplest temporal analysis involves two consecutive frames that are compared to extract dynamic handcrafted features (Chen et al., 2004; Filonenko, Hernández et al., 2017; Foggia et al., 2015; Steffens et al., 2016; Torabian et al., 2021). In Jain and Srivastava (2021) the average difference between the current frame and a previous one is used to confirm or reject the classifier prediction when its confidence is low. Especially for smoke detection, other

Table 2
Characteristics of short range, long range, low activity and high activity scenarios and their impact on the design choices.

Scenario	Characteristics
Short range (Fire > 50 PPM)	Flames and smoke are both visible in the early ignition stages. Input resolution is not relevant, due to the sufficient size of the fire. Occlusions must be considered. Fire region proposal may be not necessary.
Long range (Fire ≤ 50 PPM)	Smoke detection may be necessary since flames are not early visible. Input resolution is relevant, due to the small size of the fire. Smoke could be confused with clouds, fog, and dust. Fire region proposal is required to notify the precise position of the fire.
Low activity	Background almost static, low probability to see fire-like objects. Image classification or lightweight region proposal methods can be adopted. Fire recognition can be performed with single frame approaches. Temporal analysis is not mandatory, but useful to reduce false alarm rate.
High activity	High probability to have fire-like moving objects in the scene. Advanced region proposal methods may be necessary to reduce false alarms. Advanced fire recognition is required to distinguish fire from fire-like objects. Temporal analysis may be required for robustness and to reduce false alarms.

works encoded the small movements between consecutive frames by computing the optical flow, which is then fed to machine learning models (Pundir & Raman, 2019; Shi et al., 2020). Similarly, Hu and Lu (2018) proposes a multitask neural network to detect the fire and estimate the optical flow at the same time. This training procedure is a good trade off between (i) feeding the network with pre-computed handcrafted features, which limits the data representation to learn, and, (ii) training a classifier on a couple of raw frames (Cao, Tang, Xu et al., 2022), which let the model learn an unconstrained representation that may require more training data. Longer time analysis is conducted in Chen et al. (2021), Dimitropoulos et al. (2014), Marbach et al. (2006), Verstockt et al. (2012) and Xie et al. (2022), where the segmented flame/smoke region is monitored for seconds or even minutes to predict its development employing deterministic models. More advanced approaches employ machine learning models to fuse spatial and temporal discriminative features. As an example, a two stage approach is proposed in Park and Ko (2020). Firstly, a frame wise YOLO detector is employed to extract a static representation of the fire candidates from 50 consecutive frames, then their embeddings are stacked together and classified by means of a Random Forest classifier. Contrary, 3D convolutional neural networks (Cao, Tang, Lu, 2022; Cao et al., 2021; de Venâncio et al., 2023; Hsu et al., 2021; Huo, Zhang, Zhang et al., 2022; Lin et al., 2019) or recurrent neural networks (Cao et al., 2019; Dewangan et al., 2022; Filonenko, Kurnianggoro et al., 2017; Kim & Lee, 2019; Nguyen et al., 2021; Tao, Lu et al., 2022; Tao, Xie, Wang & Xin, 2022; Yin et al., 2019; Zhao et al., 2020) have been successfully employed to jointly analyze consecutive video frames. In addition, a prior selection of non redundant frames is conducted by the technique presented in Tao and Duan (2023). In the last years, deep learning approaches set the new state of the art in video-based fire detection like in most Computer Vision applications. This promising research path, however, implies additional issues due to the scarcity of data; specific validation protocols are necessary to test the generalization capabilities of these methods, carry out hard sample mining, and fairly compare methods working in the same application scenario.

4. Scenario based taxonomy

To better cover the real conditions in which the fire detection methods operate, we introduce a different taxonomy that allows us to characterize the peculiarities of the application scenarios. On one hand, we distinguish the scenarios based on the size of the fire in the image, which depends on its actual size, the distance from the camera, the field of view, and the resolution of the image. However, we define a rule on the size of the fire independent of all these variables. In this regard, we consider that real video analysis systems for fire detection require as a precautionary constraint that flame or smoke are at least 25 pixels wide; furthermore, to detect a fire in the early stages of ignition, it is possible to assume that it is 0.5 m wide. Therefore, we can conclude that if the fire is at least 50 pixels per meter (PPM), its size is favorable for effective detection; in case it is smaller in the image, the recognition may be more difficult due to the distance from the camera, the field of view and/or the resolution. According to this observation, we define two operative scenarios: (i) short range, in which the acquisition conditions guarantee that the fire is bigger than 50 PPM, and (ii) long range, when the distance from the fire, the field of view or the image resolution prevent to achieve the minimum size constraint. On the other hand, we discuss the activity level of different scenarios and the challenging situations that especially a high activity level may create. We summarize the characteristics of the scenarios in Table 2.

Short range fire detection is the application scenario of smart cities or remote monitoring of man made infrastructures, or also a roadway/highway/railway trait, especially in tunnels. In this scenario, either the camera lacks a wide view or the area of interest surrounds the infrastructure, like an isolated energy plant. Recently, there is also a strong interest for fire detection systems mounted in company warehouses, that fall into this scenario. At a short range, both the flames and the smoke are visible in the scene, except for occlusions, as evident in Fig. 4. Therefore, the size of the fire makes fire region proposal and input resolution not so relevant.

Long range fire detection is carried out through a surveillance camera that frames a far or wide area, especially for reducing the number of cameras and, thus, the overall cost of the system. This scenario well represents the monitoring of wide natural or industrial areas, as shown in Fig. 4. The techniques suitable for this scenario should tackle fire detection by relying on, either implicitly or explicitly, smoke detection. This is because at a long range flames are eventually visible only when the fire is fully developed. Consider that at a certain distance, which depends on the height of the camera, the earth's curvature obstructs the sight at ground level (Çetin et al., 2013). On the other hand, the smoke column is visible from the first moments in favorable weather conditions. Since early detection is based on tiny traces, the working image resolution is another crucial feature of the system, since increasing it is possible to increase proportionally the size in pixel of the fire (and, consequently, the PPM). In this scenario, the far or wide view may include clouds or fog. These weather conditions may affect performance by increasing the false alarm rate; indeed, even a human operator could confuse clouds, dust, or fog for fire smoke looking at a single still frame of the video (Fig. 4, bottom). Another desirable feature of long range monitoring systems is the fire region proposal capability, that allows to send a more precise notification of the fire position; in this case, smoke can be a more distinguishing trace, especially when a few pixels of the fire are visible.

Fig. 4. Examples of fire detection tasks at a short (top) and long (bottom) range. While in the first case flames and smoke are visible except in case of strong occlusions, in the second case it is crucial to perform smoke detection at a sufficient resolution for early fire recognition and to localize the fire for giving precise information on the fire site.

In low activity scenario, the camera points towards a scene in which moving objects are rare or the movement is not perceived. As an example, in a forest or a rural area, the moving objects are very few and the background varies very slowly (e.g., with the sunlight). Detectors designed to work in this scenario may employ simple background subtraction methods. At the same time, the fire classification could rarely be addressed looking at still frames. Since rarely new objects appear in the scene, the fire-like objects are even rarer. These approaches may rely on the appearance only to detect fires, ignoring the dynamic that characterizes flames or smoke. This makes the technique much more simple and able to be executed on the edge (e.g., smart cameras). At most, to make the system stronger against false alarms, the fire alarm can be raised after a majority vote on the decision taken on some consecutive frames.

Contrarily, high activity scenarios may include vehicles passing through the scene (fire detection in tunnels or in landfills), people standing or walking, and eventually carrying objects (fire detection in rural fields, in warehouses or in production lines). By itself, this already poses a great challenge to fire detection, since in many cases a rapid change of scene can mislead the detector. In addition, false alarms may occur due to particular yellowish or reddish objects that casually appear on the scene or due to the smoke/gases/dust produced by machinery or vehicles. For both rule and learning based approaches, it is challenging to take into account all the events that may occur. Moreover, the complex and varying scene can lead to obstructions, thus the fire can be only partially visible from the point of view of the surveillance camera, or visible for a short period, yielding potential miss predictions. For these reasons, this is probably the most challenging and least investigated scenario.

In summary, range and activity may have an impact on different design choices for fire detection methods. To take into account these different peculiarities, we define four scenarios that combine range (short and long) and activity (low and high). An outline of the distribution of the considered methods at varying of the main features characterizing the scenarios defined in our taxonomy is depicted in 5. The graphs report the number of techniques grouped whether they use region proposal mechanisms, divided by the fire traces that they can recognize, the suitable fire range and scene activity they can handle.

Another important information we provide regards the datasets used to design or validate the methods. Table 3 shows a legend of the most used public datasets and web repositories. Also, we report, for each scenario, the number of works using these data. While a thorough description of the considered datasets is presented in Section 5.1, here we point out that most works use private datasets that prevent a direct performance comparison. This is further highlighted in Fig. 6, which depicts the distribution of literature works adopting private or public datasets.

In the following sections, on the basis of the characteristics of the existing methods, we identify the approaches that could be suitable for each proposed scenario.

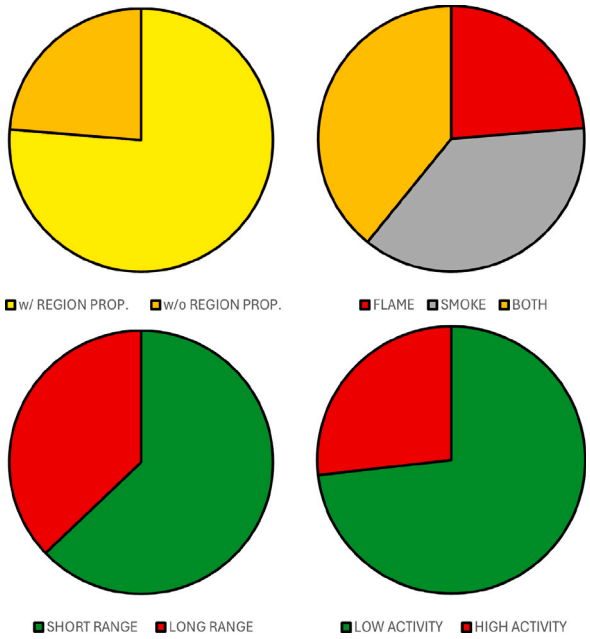


Fig. 5. Graphs depicting the distribution of works implementing region proposal mechanisms (top-left), designed to detect flames, smoke or both (top-right), to work at short or long range (bottom-left), or in a scene with low or high activity (bottom-right).

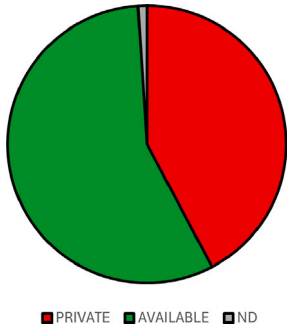


Fig. 6. Distribution of works using public or private datasets. ND means that the dataset information was not provided.

4.1. Short range with low activity scenario

This scenario is the simplest of the four considered, and most of the existing methods are fairly suited for it. These approaches are summarized in Table 4. Short range fire detection methods can rely on spotting flames and/or smoke, which are both visible after a while from the fire

Table 3
Public datasets or web repositories, and their usage in each scenario by varying Range (R) and Activity (A) level.

Reference	ID	Short R Low A	Long R Low A	Short R High A	Long R High A	Total
Cetin (0000)	D01	1	1	2	0	4
Bu and Gharajeh (2019)	D02	1	0	0	0	1
Cazzolato et al. (2017)	D03	1	0	0	0	1
Chino, Avalhais, Rodrigues, and Traina (2015)	D04	2	2	1	0	5
de Venâncio, Rezende, Lisboa, and Barbosa (2021)	D05	0	1	0	1	2
Foggia et al. (2015)	D06	8	2	1	0	11
Gong et al. (2022)	D07	1	0	0	0	1
Hsu et al. (2021)	D08	0	0	5	0	5
Jadon et al. (2019)	D09	2	0	0	0	2
Kose et al. (2010)	D10	1	0	1	0	2
Steffens, Rodrigues, and da Costa Botelho (2015)	D11	1	1	0	0	2
Tuna, Onaran, and Cetin (2009)	D12	1	0	0	0	1
Beneduce, Hill, and Schelle (0000)	D13	1	0	0	0	1
USTC-Smoke	D14	1	0	0	0	1
USTC	D15	1	1	0	0	2
DeepQuestAI	D16	1	0	0	0	1
FIgLib	D17	0	2	0	0	2
Total		23	10	10	1	44

ignition without requiring a high resolution. In addition, low activity fire detection can be carried out with efficient detection methods, also based on background subtraction, performing the classification using one frame, even if more frames may be useful to reduce the false alarms.

Therefore, early approaches for fire localization based on hand-crafted features, characterizing the appearance of the fire, may be enough. These methods can be fully interpretable and provide pixel-level information about fire presence. In Chen et al. (2004), flame and smoke pixels are located in a still frame using a model based on the fire chromatic and morphological characteristics in the RGB color space. Later, in Horng et al. (2005) the HSI color space is exploited to segment fire-like objects first and then delete spurious regions. Similarly, in Celik and Demirel (2009) the flame pixels are better isolated in the YCbCr color space.

More recently, learning-based approaches have proven to achieve superior performance. On this path, the authors of Zhang et al. (2021) tackle fire image classification at short range training a deep learning model for flame segmentation. Similarly, in Xu, Zhang, Zhang et al. (2019) pixel-wise saliency detection is carried out by means of a dedicated CNN. Then, the internal representation is used to classify the image as fire or not. A smoke segmentation approach based on a two-path auto-encoder is proposed in Yuan, Zhang, Xia et al. (2019). The first path is devoted to a rough global segmentation, while the second one is employed to refine the prediction. The same authors proposed a novel approach in Yuan et al. (2021). The technique consists of a multi-resolution smoke segmentation network, which includes a spatial self-attention module, and an image classification branch, whose aim is to assist the segmentation.

A different class of approaches aims to classify the whole image as fire or normal, thus exploiting no localization information during the training procedure. In Muhammad, Ahmad, Lv et al. (2018) and Muhammad et al. (2019), the authors propose to exploit recent CNN architectures, namely SqueezeNet (Iandola et al., 2016) and MobileNet (Howard et al., 2017) respectively, for their good performance and low amount of training parameters. A similar strategy is adopted in Li et al. (2020) to train a shallow 10-layered CNN. In order to better identify the fire in the whole image, the proposed network architecture includes both a channel-attention mechanism and an inception-like block for multi-resolution features extraction. In Li et al. (2020), Muhammad, Ahmad, Lv et al. (2018) and Muhammad et al. (2019) strided layers are employed to limit the network capabilities of carving small traces in the image. An even shallower CNN is proposed in Ayala et al. (2020), where two different variants are presented in order to reduce the number of trainable parameters. The first, inspired by MobileNet (Howard et al., 2017), makes use of separable convolutions and inverted residual connection. The second, which borrows the octave convolution (Chen et al., 2019), splits the input into low and high frequencies. However, the input shape, which is set to 84×84 pixels, bounds this technique to the detection of fully developed fires only. In Yin et al. (2017) the authors propose a 14-layered CNN that included batch normalization to deal with overfitting and speed up the convergence. Gu et al. (2019) defined a dual-path CNN architecture to improve the smoke detection performance. Together with batch normalization layers, skip connections were adopted to reduce the vanishing gradient effect in this deeper architecture. In Sheng et al. (2021), the authors propose to cluster the pixels of the input image through super-pixel segmentation (SLIC Achanta et al., 2012) followed by an iterative clustering algorithm (DBSCAN Shen et al., 2016), to simplify the patterns of both the flames and smoke, whose pixel value variability is reduced. After that, a three layer CNN is trained using the DeepQuest (0000) dataset. Also, by explicitly looking for both flames and its smoke column makes the approach much more robust to occlusions since it is unlike that both are not visible at the same time. In Jain and Srivastava (2021) an indoor fire detection system is presented. In order to preserve the occupants' privacy and improve the fire detection accuracy, the neural network is trained on images acquired by infrared cameras only.

A third class of methods tries to get the best of fire segmentation and full image classification approaches. Early works pursue this goal by employing well established background subtraction methods. Since they spot foreground objects by movement, they are particularly suitable for this scenario. As an example, in Foggia et al. (2015) a background subtraction method is adopted (Conte, Foggia, Petretta, Tufano, & Vento, 2005), then a combination of expert systems is proposed to tell apart flames from other moving objects. More recently, object detector neural networks tackle candidate regions localization and classification at the same time. Even if these approaches are more convenient at a long range, they have also been employed for short range fire detection by Saponara et al. (2021). The limited input resolution of this implementation, namely 128×128, prevent its usability in a different application scenario. In Shahid, Virtusio et al. (2021) a two stage network is proposed to automatically spot fire candidates and to classify them. The first stage is implemented by using two different networks: a 2D spatial network extracting static features and a 3D temporal network characterizing the fire dynamic.

Table 4
Methods suitable for the short range and low activity scenario.

Reference	F/S	Input size	Region proposal	Fire recognition	Dataset
Chen et al. (2004) Horng et al. (2005)	Both	–	Handcr. segmentation	Rule-based	ND
Celik and Demirel (2009)	Flame	256 × 256	Handcr. segmentation	Rule-based	Private
Foggia et al. (2015)	Flame	–	Background sub.	Rule-based	D06
Sharma et al. (2017)	Both	–	None	Resnet50	Private
Ayala et al. (2020) Ayala et al. (2022) Yin et al. (2017)	Both	84 × 84	None	Shallow CNN	D09 D03
Muhammad, Ahmad, Lv et al. (2018)	Flame	224 × 224	Sel. activation map	SqueezeNet	Mix
Zhong et al. (2018)	Flame	227 × 227	Handcr. segm./det.	Shallow CNN	Private
Zeng et al. (2018)	Smoke	299 × 299	RPN	Faster R-CNN	Private
Muhammad, Ahmad, Mehmood et al. (2018)	Both	224 × 224	None	IncevptionV1	Mix
Muhammad, Ahmad, Baik (2018) Hosseini et al. (2022) Almeida et al. (2022)	Both	224 × 224	None	Shallow CNN	Mix
Hu and Lu (2018)	Both	227 × 227	None	Multi-task shallow CNN	Private
Dunnings and Breckon (2018)	Both	608 × 360	None	InceptionV1 based	Mix
Muhammad et al. (2019) Yang et al. (2019)	Flame	224 × 224	None	MobileNet	Private
Gu et al. (2019)	Smoke	48 × 48	None	Two-stream CNN	Private
Shi et al. (2019)	Smoke	48 × 48	Dark channel preproc.	VGG16 based	Mix
Khan et al. (2019)	Smoke	224 × 224	None	VGG16	Mix
Xu, Zhang, Zhang et al. (2019)	Smoke	224 × 224	DHSnet inspired	Shallow CNN	D14
Li et al. (2019)	Smoke	224 × 224	Color based preproc.	DenseNet based	Private
Yuan, Zhang, Xia et al. (2019)	Smoke	256 × 256	U-Net segmentation	Pixel-level only	Private
Yuan, Zhang, Wan et al. (2019) Jadon et al. (2019) Khudayberdiev, Zhang, Abdullahi et al. (2022)	Both	84 × 84	None	Shallow CNN	Mix D09
Li et al. (2020)	Flame	224 × 224	Spatial attention	Shallow CNN	Private
Shi et al. (2020)	Smoke	240 × 180	Optical Flow preproc.	MobileNetV2 based	D13
Zhang et al. (2020)	Smoke	227 × 227	None	Two-stream AlexNet	D15
Oh et al. (2020)	Both	224 × 224	None	EfficientNetB0	Private
Jain and Srivastava (2021)	Flame	112 × 112	None	SqueezeNet based	Private
Zhang et al. (2021)	Flame	224 × 224	U-Net like seg.	Shallow CNN	Mix
Torabian et al. (2021)	Flame	256 × 400	Background sub.	Handcr. feat.+SVM	D01 D10
Sheng et al. (2021)	Smoke	–	SLIC-DBSCAN preprocessing	Shallow CNN	D16
Yuan et al. (2021)	Smoke	256 × 256	CNN-GRU segment.	Shallow CNN	Private
Saponara et al. (2021)	Both	128 × 128	YOLOv2	YOLOv2	D06 D09
Shahid and Hua (2021)	Both	224 × 224	Self-attention	ViT	D06 D04
Shahid, Virtusio et al. (2021)	Both	256 × 256	Spatio-Temporal CNN	DenseNet based	D06
Hogan et al. (2021)	Both	320 × 320	None	YOLOv5	Mix,D11
Gong et al. (2022)	Smoke	224 × 224	Spatial attention	VGG16 based	D07
Ghosh and Kumar (2022)	Both	128 × 128	None	Shallow CNN+LSTM	D06
Khan et al. (2022)	Both	150 × 150	None	EfficientNetB3-based	D06 D04
Huang et al. (2022)	Both	224 × 224	RPN	Faster R-CNN	Mix,D06
Majid et al. (2022)	Both	224 × 224	Spatial attention	EfficientNetB0-based	Private D06 D12
Li, Zhang, Liu, Jin (2022)	Both	224 × 224	Anchor free detector	MobileNetV3	Private
Dogan et al. (2022)	Both	224 × 224	None	Ensemble ResNets	Mix
Khudayberdiev, Zhang, Elkhalil et al. (2022)	Both	224 × 224	Self-attention	Swin Transf.	D02

In conclusion, this scenario does not present really critical challenges. Therefore, methods in the literature based on handcrafted features, background subtraction, or lightweight deep neural networks demonstrated to achieve remarkable performance. At light of the need to run on edge devices, these approaches that require low computational resources are welcome. More complex approaches, which will be analyzed in the next sections, exhibit additional features that are mostly superfluous in this scenario that is characterized by limited variability; so, their adoption would only add computational burden without a significant performance improvement.

4.2. Long range with low activity scenario

In this scenario, the low activity of the scene allows to select fire localization and recognition methods characterized by a medium complexity. The long range requires additional constraints on the design

Table 5
Methods suitable for the long range and low activity scenario. "Mix" means that various datasets have been merged.

Reference	F/S	Input size	Region proposal	Fire recognition	Dataset
Steffens et al. (2016)	Flame	–	Random Forests Pixel classification	Handcr. feat. + Random Forests	D11
Filonenko, Hernández et al. (2017)	Smoke	1920 × 1080	Background sub.	Pixel-level only	Private
Luo et al. (2018)	Smoke	Crop@227 × 227	Background sub.	Shallow CNN	D15
Xu, Zhang, Liu et al. (2019)	Smoke	500 × 500	Single Shoot Detector	VGG16	Private
Pundir and Raman (2019)	Smoke	Crop (various)	SLIC+Optical Flow preproc.	2-stream AlexNet	Mix
Xu, Wanguo et al. (2019)	Both	Crop (various)	Handcr. segmentation	Shallow CNN	Private
Aktas et al. (2019)	Both	Slide@227 × 227	None	MIL SqueezeNet	Mix
Chaoxia et al. (2020)	Flame	600 × 600	Color based RPN	Faster R-CNN	D04
Xie et al. (2020)	Flame	Crop@224 × 224	Background sub. +flicker detection	Shallow CNN	D06 D04
Li et al. (2021)	Flame	600 × 600	Color/motion RPN	R-CNN	Mix
Frizzi et al. (2021) Choi et al. (2021)	Both	640 × 640	CNN based segmentation	Pixel-level only	Private
Wang et al. (2021)	Both	Crop@224 × 224	SuperPixel segmentation +LBP foreground extr.	ShuffleNet	Mix,D06
Zhang et al. (2022)	Flame	–	Anchor free det.	FPN-based	Mix,D01
Prema et al. (2022)	Smoke	360 × 240	Rule-based (color, motion)	LBP+ELM	Private
Cao, Tang, Xu et al. (2022)	Smoke	512 × 512	Weakly guided attention	MobileNetV3	Private
Hu et al. (2022)	Smoke	512 × 512	YOLOv5-based	YOLOv5-based	Private
Huo, Zhang, Jia et al. (2022)	Smoke	608 × 608	YOLOv4-based	YOLOv4-based	Private
Jiang et al. (2022)	Smoke	768 × 768	Self-attention +Anchor free detection	Light CNN	Private
Dewangan et al. (2022)	Smoke	1382 × 1843	Self-attention	ResNet34+ LSTM+ViT	D17
Yazdi et al. (2022)	Smoke	3072 × 2048	Self-attention	DETR	D17
de Venâncio et al. (2022)	Both	416 × 416	Pruned YOLOv4	Pruned YOLOv4	D05
Qian et al. (2022)	Both	416 × 416	Pruned YOLOv3	Pruned YOLOv3	Private
Shakhnoza et al. (2022)	Both	512 × 512	Spatial attention	Shallow CNN	Private
An et al. (2022)	Both	608 × 608	Pruned YOLOv5	Pruned YOLOv5	D01
Yang et al. (2022)	Both	608 × 608	Self-attention	Transformer	Private
Wu et al. (2022)	Both	640 × 640	YOLOv5-based	YOLOv5-based	Mix,D01
Li, Zhang, Liu, Jing et al. (2022)	Both	800 × 800	Self-attention	DETR-based	D06
Mardani et al. (2023)	Both	800 × 800	Self-attention	DETR-based	D06
Abdusalomov et al. (2023)	Both	416 × 416	Color-based preproc.	Mask-RCNN	Private
Huang et al. (2023)	Both	640 × 640	De-fog pretrain	YOLOX	Private

choices. The methods better suited for this scenario should detect smoke, avoiding the confusion with other similar environmental elements (clouds, fog, dust). In this regard, the resolution becomes a relevant aspect for the localization of the fire in the early stages of ignition; if the resolution is sufficiently high, also flames may be detected. According to these considerations, Table 5 summarizes the approaches that may be adopted in long range with low activity scenario.

Most of these methods implement a prior region proposal or fire segmentation. In Frizzi et al. (2021), a CNN based semantic segmentation approach is proposed. The input 640 × 480 image is processed to tell apart the fire and smoke pixels. However, an image-level classification algorithm is not proposed, thus preventing its direct usage in real-world applications. Thanks to the quasi-static background, fire localization through foreground extraction can be reliably addressed at long range. In Filonenko, Hernández et al. (2017), Luo et al. (2018) and Prema et al. (2022), background subtraction methods are adopted, while the method described in Filonenko, Hernández et al. (2017) performs smoke segmentation through color analysis and connected components based morphological operations. In Prema et al. (2022) candidate smoke regions are collected based on both color and movement analysis, then the classification phase is based on LBP co-occurrences. It can be noted that such approaches are prone to errors due to illumination changes, meanwhile deep learning methods have largely improved the smoke segmentation performance. In Luo et al. (2018) a crop around the candidate region is performed, then a CNN is used to tell apart fire smoke patches from normal ones. The same image patch dimension is processed by the technique presented in Aktas et al. (2019). In this case, SqueezeNet is trained in a Multiple Instance Learning fashion to classify the full-resolution image as fire even if only one of the processed patches show smoke.

Recent approaches adopted attention to deal with fire detection in high resolution images. In Jiang et al. (2022) and Shakhnoza et al. (2022) a self-attention based custom CNN is trained to detect smoke frame-wise, with the latter proposing a very shallow CNN. Meanwhile in Cao, Tang, Xu et al. (2022) a similar approach is employed to classify two consecutive frames. This aim to train the network to represent the smoke short dynamic, not to deal with moving objects for which only two frames are not sufficient. On the same path, a complex approach has been presented in Dewangan et al. (2022). First, a custom CNN is used as spatial feature extraction. Then, a LSTM is employed to combine the embeddings extracted by two consecutive frames. On top of this, a Visual Transformer (ViT) (Dosovitskiy et al., 2020) is adopted to classify a sequence of frames as either fire or not. Besides the image-level binary cross-entropy loss computed on the ViT output, the same loss is applied at tile-level (both spatially and temporally) on the features extracted by both the CNN, the LSTM and the ViT to leverage localization information during training. To reduce false alarms due to clouds, the top of the image is discarded since, in this particular scenario, it often depicts the sky.

When dealing with a wide scene, object detector networks can fully exploit their embed region proposal branch. The Mask R-CNN architecture has been fine-tuned in Abdusalomov et al. (2023). The input

Table 6
Methods suitable for the short range and high activity scenario. "Mix" means that various datasets have been merged.

Reference	F/S	Input size	Region proposal	Fire recognition	Dataset
Liu and Ahuja (2004)	Flame	Various	Handcr. segmentation	Rule-based	Private
Marbach et al. (2006)	Flame	Various	Handcr. segmentation	Rule-based	Private
Töreyin et al. (2006)	Flame	Various	Foreground extr. (using flickering)	Rule-based	Private
Celik et al. (2007)	Flame	176×144	Background sub.	Rule-based	Private
Verstockt et al. (2012)	Smoke	384×288	LWIR-camera thresholding	Rule-based	Private
Dimitropoulos et al. (2014)	Flame	320×240	Background sub.	Handcr. feat.+SVM	D01 D10
Filonenko, Kurnianggoro et al. (2017)	Smoke	299×299	None	InceptionV4+GRU	Private
Li, Chen, Wu, and Liu (2018)	Smoke	256×256	3D CNN segmentation	Pixel-level only	Private
Yin et al. (2019)	Smoke	128×64	None	CNN+RNN	D01
Lin et al. (2019)	Smoke	112×112	Faster R-CNN based	C3D-v1.0	Mix
Kim and Lee (2019)	Both	224×224	RPN	Faster R-CNN (ResNet101)+LSTM	Private
Zhao et al. (2020)	Both	224×224	None	VGG16 + LSTM	Private
Chen et al. (2021)	Flame	176×144	Region growing segmentation	Rule-based	Mix
Shahid, Chien et al. (2021)	Flame	336×336 Crop@32×32	U-Net segmentation	ResNet34 based	D06 D04
Hsu et al. (2021)	Smoke	180×180	None	Timeception	D08
Xie et al. (2022)	Flame	Various	Background sub.	Multiexpert	Mix
Tao, Xie et al. (2022)	Smoke	128×128	None	ConvLSTM-based	D08
Tao, Lu et al. (2022)	Smoke	128×128	Spatio-temporal attention	ConvLSTM-based	D08
Cao, Tang, Lu (2022)	Smoke	224×224	FPN	FPN+SE-ResNext-50	D08
Tao and Duan (2023)	Smoke	128×128	Spatio-temporal attention	Frame selection +CNN+LSTM	D08

resolution, equals to 416×416, is just enough to spot small traces at long range, after they have been emphasized by the proposed brightness and contrast enhancement operations. Different YOLO versions have been employed for fire detection (An et al., 2022; de Venâncio et al., 2022; Hu et al., 2022; Huang et al., 2023; Huo, Zhang, Jia et al., 2022; Qian et al., 2022; Wu et al., 2022). Powered by these advanced architectures, most of these approaches exhibit near real-time processing capabilities on input images whose resolution ranges from 416×416 to 640×640. Moreover, to further reduce the necessary resources and make these networks suitable to be executed on the edge, pruning strategies have been adopted in An et al. (2022), de Venâncio et al. (2022) and Qian et al. (2022). A transformer based detector, namely DETR (Carion et al., 2020), has been employed in Yazdi et al. (2022). At the cost of an increased complexity, this method tackles end-to-end fire detection on very high resolution images proving to be able to detect even very distant fires from a barely visible smoke column. Other approaches relying on a similar transformer architecture but working at a mid-resolution are Li, Zhang, Liu, Jing et al. (2022) and Mardani et al. (2023). Instead, a hybrid CNN-transformer architecture is proposed in Yang et al. (2022). The authors tailor the attention mechanism to detect small fires at an early stage.

In summary, this scenario shares the same challenges of the previous one, but requires more advanced techniques specifically tailored for detecting fires at long range. Obviously, long range implies that only few pixels are available to detect smoke or even fewer for flames, especially in presence of partial occlusions. Moreover, in this application context, as images take wider areas, fog and clouds are often present creating confusion with smoke. Given the low background activity, methods trading off a higher processing frame rate for a higher working resolution, are desirable. For these reasons, methods that adopt smoke detection techniques based on high resolution segmentation, deep region proposals, and attention mechanisms could be more suited for this scenario.

4.3. Short range with high activity scenario

Although the short range allows to detect both flames and smoke in the early stages of ignition even at low resolution, the methods applicable in this scenario should comply with the requirements imposed by the high activity in the scene. In particular, the approaches better suited to short range with high activity scenarios should include localization algorithms robust to false positives and occlusions, with recognition methods, typically multi-frame, able to distinguish fire-like objects from flames and smoke Most of the methods discussed in the short range with low activity scenario work on single video frames, thus neglecting the movement information of the scene. Instead, in a real-world application, to get rid of the challenges posed by a dynamic scenario, time analysis is crucial to tell apart actual fires from fire-like objects. The approaches that may be better suited for this scenario are summarized in Table 6.

One of the earliest works performing a time analysis of segmented flames has been proposed in Liu and Ahuja (2004). After segmenting the flame color, a novel temporal feature is expressed as the variation of the Fourier coefficient representing the flame shape in each video frame. A similar approach is pursued by the technique proposed in Marbach et al. (2006) that performs a time analysis on the luminance channel. In Verstockt et al. (2012) a smoke segmentation and classification system has been presented. The candidate region selection is based on infrared camera thresholding over 45 frames, then a rule-based classifier may detect the presence of fire. In Celik et al. (2007), a flame color model is proposed. Each connected component representing a candidate flame region, segmented after background subtraction, is monitored over time. Fire regions are distinguished from fire-like ones based on changes in their location and size. Similarly, in Chen et al. (2021) the authors propose to obtain a rough flame localization in the favorable YCbCr color space, thus they propose an ad-hoc region-growing algorithm to refine the flame segmentation; fire is detected by analyzing the variations in size, position, and shape

Table 7
Methods suitable for the long range and high activity scenario. "Mix" means that various datasets have been merged.

Reference	F/S	Input size	Region proposal	Fire recognition	Dataset
Cao et al. (2019)	Smoke	Crop@299 × 299	Background sub.	InceptionNetV3+Bi-LSTM	Private
Park and Ko (2020)	Both	640 × 480	YOLOv3	YOLOv3 based + RF	Private
Nguyen et al. (2021)	Flame	Crop (various)	Handcrafted seg.	ResNet18+Bi-LSTM	Private
Cao et al. (2021)	Smoke	Crop@224 × 224	Background sub.	Self-attention CNN	Private
Huo, Zhang, Zhang et al. (2022)	Smoke	416 × 416	YOLO-based	YOLO-based 3D-CNN	Mix
de Venâncio et al. (2023)	Both	640 × 640	YOLOv5	YOLOv5+temporal analysis	D05

of the frame-wise flame segmentation outputs. The Wavelet transform adopted in Töreyin et al. (2006) represents spatial color variation and temporal flame flickering. Together with the irregular flame shape, these clues are used to detect short range fires. Instead, in Xie et al. (2022) a multiexpert system is designed to detect the presence of flames after background subtraction. Using 10 consecutive frames, the system is able to spot firelight reflections, thus partially overcoming the problem of occlusions.

Deep learning based approaches have improved both the segmentation and the detection phases. A two-phase method has been presented in Shahid, Chien et al. (2021). In a first stage, a U-Net like segmentation network is trained to segment flames, then around each candidate a 32×32 patch is cropped and classified by means of a ResNet34. In Li et al. (2018) a 3D convolutional auto-encoder with multi-resolution skip-connections is employed to segment smoke in space and time, jointly. The authors exploit the smoke dynamic as well as its color appearance in order to improve the localization performance. The limited resolution adopted is partially counteracted by the fact that the custom CNN architecture has unitary-strided first convolutional layer.

Other works exploit recurrent neural networks on consecutive video frames to characterize the flame or smoke dynamic. The authors of Yin et al. (2019) proposed a CNN+RNN network tested on low resolution videos. Two different CNNs are trained to extract single frame features and to characterize the differences in a couple of consecutive frames. Then, the features are combined by the recurrent network over 80 frames. A similar architecture is adopted in Filonenko, Kurnianggoro et al. (2017), but using a deeper InceptionV4 convolutional neural network and gated recurrent units (GRU) to combine up to 42 frames. In both the latter methods, no prior segmentation or candidate region selection is performed. Other approaches, instead, try to combine the localization capabilities of object detectors with the dynamic analysis, often conducted through 3D CNNs or recurrent networks. In Kim and Lee (2019), the authors adopt a Faster R-CNN architecture for object detection and spatial feature extraction; then a LSTM is employed to combine features extracted by 2 or 3 s of the video. A localization phase based on the same detector architecture is presented in Lin et al. (2019) and trained to spot smoke candidates. Thus, the C3D (Tran, Bourdev, Fergus, Torresani, & Paluri, 2015) 3D CNN is employed to optimize the spatio-temporal feature extraction and classification. Short-time temporal analysis is carried out in Cao, Tang, Lu (2022) to better characterize the smoke dynamics; in particular, the feature extracted by a two-stream convolutional neural network is fused to combine spatial and temporal information.

At a glance, existing methods better suited for this scenario can address specific situations that may occur in this application context, for instance by carrying out a temporal analysis aimed to reduce the impact of the background activity on false positive rate. However, the high variability of this scenario poses a real challenge that should be adequately taken into account in future research works; in fact, focusing on algorithms that allow to distinguish fire-like objects in tunnels, landfills, and production lines would be beneficial to avoid confusion of vehicles (especially those with flashing lights), people, dusty garbage being processed, and parts to be assembled with flames and smoke. To this concern, it is worth noting that existing methods are not specifically designed and/or tested to overcome at the same time all the above mentioned different situations that may cause false alarms when deployed in real fire detection systems. Of course, the challenge here is to address all these issues while preserving a high true positive rate.

4.4. Long range with high activity scenario

This scenario is perhaps the most challenging one among those considered, since the requirements for high activity in the scene are added to the constraints imposed by the long range. In fact, the presence of moving objects is combined with the tiny size of smoke or flames due to the long range. Distinguishing between fire and fire-like moving objects in this scenario is even more challenging, since a few pixels are available to describe them; methods adopting multi-frame high resolution smoke localization and recognition algorithms should be applied to deal with these challenges. Due to the complexity of the operative conditions, only a few existing methods, summarized in Table 7, could be applied to this specific scenario

In Huo, Zhang, Zhang et al. (2022) a 3D CNN is proposed to classify frame sequences after a prior smoke detection based on a YOLO-like architecture. In this case, the input resolution, that is equal to 416×416, is somehow limited by the strided convolutions implemented in the first three layers. 12 video frames are sub-sampled to cover 4 s of the input video. The authors of Park and Ko (2020) propose a similar approach, using Elastic-YOLOv3 with 640 × 480 resolution and performing the temporal analysis over 50 frames through a random forest applied on a bag of features histogram of the optical flow of the fire. In de Venâncio et al. (2023) fire detection is performed with a YOLOv5-based algorithm and a temporal analysis over 30 frames; the verification of the fire persistence allows to substantially reduce the false positive rate. A LSTM based approach is presented in Cao et al. (2019). The method makes use of a pretrained InceptionNet as frame-wise spatial feature extraction to train a bi-directional LSTM. Additionally, an attention module is proposed to get rid of the information coming from long video sequences. The authors employ ViBe (Barnich & Van Droogenbroeck, 2010) to detect foreground objects and, for each of them, a 299 × 299 patch centered on the detected region is cropped and fed to the CNN+LSTM network.

To provide more useful information than the simple fire alarm, recently some works are focusing on fire source prediction. In Cao et al. (2021), a multitask network is trained to jointly recognize the smoke and locate the smoke source. Firstly, a self-attention mechanism is employed to extract foreground features from each video frame. After that, information extracted by a sequence of consecutive frames is jointly processed by two different network heads to predict the smoke position in terms of bounding box, and the fire source position as a pixel-level probability map. Instead, in Nguyen et al. (2021) the authors propose a method that relies on simple handcrafted features to collect fire candidate regions where the flame is visible. Limiting the candidates to flame-like objects impair the performance at long range, where only the smoke is visible; each region is then observed for 16 consecutive frames and processed through a CNN+BiLSTM architecture to decide whether it is a fire or non-fire object.

It is evident that the methods analyzed in this section, suited for managing the situations of this scenario, are only few. The intrinsic complexity of this application context, so as the lack of appropriate datasets collected in these conditions, represent a barrier for the design of effective methods. This is the most challenging scenario among all four considered, essentially because it is easier to misclassify small moving objects with flames or smoke. Promising methods are those that are able to effectively combine the appearance information extracted from

Table 8
Datasets acquired with static (top) and moving (bottom) cameras, characterized in terms of number of videos, minimum and maximum resolution, presence of flame and/or smoke, acquisition environment (indoor/outdoor), range (short/long) and activity (low/high).

Mode	Dataset	# videos	Resolution		Flame	Smoke	Environment		Range		Activity	
			Min	Max			Indoor	Outdoor	Short	Long	Low	High
Static camera	Bilkent (Cetin, 0000)	38	320 × 240	720 × 576	15	31	6	32	22	16	33	5
	D-Fire (de Venâncio et al., 2021)	100	1280 × 720	1280 × 720	0	50	0	100	0	100	100	0
	KMU (Ko et al., 2011)	38	320 × 240	320 × 240	22	21	10	28	16	22	37	1
	Mivia FIRE (Foggia et al., 2015)	31	320 × 240	800 × 600	15	28	4	27	17	14	29	2
	Mivia SMOKE (Foggia et al., 2015)	149	292 × 240	292 × 240	0	76	0	149	0	149	149	0
	Total	356			52	206	20	336	55	301	348	8
Moving camera	FireNet (Jadon et al., 2019)	62	352 × 222	1920 × 1080	43	27	30	32	59	3	55	7
	FireSense (Kose et al., 2010)	49	300 × 240	1600 × 1200	16	19	10	39	44	5	33	16
	FiSmo (Cazzolato et al., 2017)	88	320 × 240	1920 × 1080	80	37	0	88	81	7	33	55
	FURG (Steffens et al., 2015)	22	420 × 240	1920 × 1080	16	14	3	19	22	0	11	11
	Total	221			155	97	43	178	206	15	132	89

a high resolution version of the video with features representing the smoke movement computed from a time series of consecutive frames. To this aim, either background subtraction or multi-frame learning based approaches (RNNs, 3D CNNs, Transformers) may be combined to improve fire localization and recognition sensitivity and specificity. Although the proposed methods are effective in specific situations, there is the possibility of improving their generalization capabilities. Supported by data acquisition campaigns, researchers would be able to develop advanced deep learning methods to fill this gap.

5. Datasets and performance

5.1. Datasets

Although there are various fire image dataset published in the literature (Cazzolato et al., 2016; Chino et al., 2015; Villela et al., 2018), in this survey we consider only video datasets, since they allow for assessing the impact of the background activity on the fire detector performance. The publicly available fire detection video datasets are: Bilkent (Cetin, 0000), D-Fire (de Venâncio et al., 2021), KMU (Ko, Ham, & Nam, 2011), Mivia FIRE (Foggia et al., 2015), Mivia SMOKE (Foggia et al., 2015), FireNet (Jadon et al., 2019), FireSense (Kose et al., 2010), FiSmo (Cazzolato et al., 2017) and FURG (Steffens et al., 2015). It is important to point out that those datasets were not collected with reference to a specific application scenario, but simply including videos coming from disparate scenarios.

For this reason, we have analyzed all the videos present in those datasets with the aim of labeling them with the information necessary to characterize the specific scenario they are suited for; in particular, we have annotated according to: resolution, presence of flames and/or smoke, environment (indoor/outdoor), range (short/long), and activity (low/high). The annotations produced for each video have been made publicly available,[2] in order to allow other researchers to extract a subset of videos having similar characteristics. The annotation introduced in this paper, which takes into account granular information on the application context, will certainly be useful for differentiated experimental comparisons based on the level of activity, the type of trace and size of the fire in the image. We report in Table 8 a summary of the videos available in the considered datasets, that allows to perform a quantitative and qualitative analysis according to the above mentioned parameters.

Overall, 356 videos acquired with static cameras are publicly available (most of them from D-Fire and Mivia SMOKE), mostly stored with a low resolution (except for D-Fire, that however is collected always in the same scenario); this can represent a limitation for methods that must detect fire at long range, especially in the early stages of ignition. There is a majority of videos where smoke is visible (206) and a dearth of clips with flames (52). Almost all the videos are recorded outdoor (336), at long range (301), and, above all, with low activity (348); this deficiency leads to the probable inability to effectively evaluate performance in indoor scenarios (20 videos), at short range (55 videos) and with high activity (only 8 videos). Among the insights deductible from the statistics, the lack of datasets with high activity is the biggest drawback, as this represents the most demanding challenge of fire detection methods, still too underestimated in the literature. As for a qualitative analysis of the available data, it is worth noting the scarcity of negative video samples, in terms of number and variability. In fact, as in D-Fire and Mivia SMOKE, they are always recorded in the same scenario; moreover, they often do not depict fire-like objects that may confuse the algorithms, so it is not possible to adequately test the capability of the methods to reduce the false positive rate. In addition, it is important to clarify that most of the positive videos already start with the fire in progress; this means that the algorithms are trained and tested in most cases on the detection of fires that have already broken out and not of smoke/flames visible in the early stages of ignition. Therefore, it is almost never possible to measure the alarm notification delay, which represents a crucial parameter for evaluating the promptness of fire detection algorithms.

On the other hand, the publicly available videos recorded with moving cameras are 221 exhibiting a wide range of resolutions available, between 320 × 240 and 1920 × 1080. However, high-resolution videos are of little use because most of the videos are recorded at a short range (206). Compared to videos collected with static cameras, there is a better balance in terms of flame/smoke only numerosity, environment type, and activity level; however, the moving camera setup is rather unusual and not typical of ground video surveillance systems. As regards the qualitative content, both positive and negative samples suffer from limited representativeness as those acquired with static cameras and discussed in the previous section.

In the positive video samples, the fire is, apart from very few cases, always clearly visible starting from the first frame. Therefore, also in this case the quality and the representativeness of the data should be substantially improved.

5.2. Performance comparison

In this section, we discuss the adopted validation protocols and some of the top performing techniques for each video dataset, separately. In particular, we selected the methods that at the same time achieve the best performance with the lower complexity. For practical reasons, some methods were not included in the analysis since their performance was achieved by the authors using a proprietary dataset not publicly available. According to this section organization, we separately analyze the performance achieved on data acquired by static and moving cameras.

[2] https://github.com/MiviaLab/FireDetectionSurvey2023/blob/main/MIVIA_FIRE_DETECTION_VIDEO_ANNOTATIONS.xlsx.

5.2.1. Static camera datasets

Among the static camera, datasets reported in Table 8 (top), the D-Fire (de Venâncio et al., 2021) and the Mivia FIRE (Foggia et al., 2015) datasets have been presented together with a similar validation protocol consisting in a single train/test split. In this regard, deep learning methods should always pay attention to not putting frames extracted by the same video in both the training and the test set, thus biasing the performance. Instead, the Bilkent (Cetin, 0000) and the KMU (Ko et al., 2011) datasets have fewer samples, seldom used as test benchmarks for techniques trained on different datasets. Finally, the MIVIA smoke dataset (Foggia et al., 2015) lacks an assessed validation protocol, making the performance comparison not possible. To provide hints on smoke detection performance, we analyzed the methods tested on the RISE dataset (Hsu et al., 2021), which represent a similar task, i.e. industrial smoke emission detection.

The D-Fire (de Venâncio et al., 2021) evaluation protocol, that is to randomly split the set of images in 80%−20% for training and testing, was adopted by de Venâncio et al. (2023, 2022) and de Venâncio et al. (2021) to evaluate both the object detection and video classification tasks. Indeed, all these techniques have been proposed by the same research group and are all based on the YOLO object detector family. As regards the object detection performance, the best result has been obtained in de Venâncio et al. (2023) employing the YOLOv5l architecture and achieving a mAP@0.5 equals 79.10% averaged over the 5-fold cross-validation. In de Venâncio et al. (2023) and de Venâncio et al. (2021), a temporal analysis exploiting the frame wise YOLO detection results is implemented and evaluated on the subset of 100 videos of the D-Fire dataset. The best performance is obtained again by de Venâncio et al. (2023), but using a lighter detector, the YOLOv5s architecture, and carrying out a temporal analysis of the position and size of the detected bounding boxes. By doing this, an accuracy of 72.86% and a F_1−score equal to 0.78 is achieved. This outperformed the approach previously proposed (de Venâncio et al., 2021) that employed the YOLOv4 detector and similar temporal analysis. The need to detect small fires at a long range in the D-Fire dataset has been addressed by means of powerful object detector architecture working at mid-resolution. The dataset is acquired in a low activity scenario, thus the methods achieve good performance even working frame wise or implementing simple rule based temporal classifiers.

The Mivia FIRE dataset (Foggia et al., 2015) is seldom split into train and test samples, with the latter covering 80% of the total videos. Positive videos include long range fires while negative samples include fire-like reddish objects, thus making this small set of test samples challenging. Notwithstanding this, the method proposed in Xie et al. (2020), which firstly subtracts the background and then classifies active image regions through a shallow CNN, reaches a frame level accuracy of 97.98%, with a very low FNR equals 0.84% and only 2.33% FPR. Even better results were obtained in Shahid, Virtusio et al. (2021) by using all the 31 dataset videos for testing and achieving TPR equals 98.9%, Precision equals 99.5%, and thus F_1−score equals 99.2%. The approach employs two deep networks, one for fire candidate collections and the other for their classification. In Wang et al. (2021), a shallower CNN is adopted achieving a slightly better accuracy, equals 99.61%, and an F_1-score equals 99.63%. Preliminary fire candidates, based on handcrafted features, allowed to increase the detection rate while controlling the false alarms. The presence of fire-like and moving objects in this dataset made deep learning approaches reach state of the art performance.

In most of the works, a specific subset of the Bilkent (Cetin, 0000) videos is selected for performance comparison of methods trained on different datasets. Seven test videos (4 positive and 3 negative samples) are selected for direct performance comparison in Li, Zhang, Liu, Jin (2022). Implementing an anchor free CNN detector, the proposed method outperformed the competitors, showing good TPR, above 95% for all videos, and very good TNR, with a few false alarms in only one out of three videos. Related approaches (An et al., 2022; Zhang et al., 2022) confirmed this performance. The technique proposed in Zhang et al. (2022) has been tested on 5 videos only, reaching near perfect performance in terms of both TPR and TNR. On average, these results are confirmed by those presented in An et al. (2022), even if this work does not report per video performance for direct comparison.

Similar to the Bilkent dataset, also from the KMU dataset (Ko et al., 2011) a subset of videos is seldom selected to test the generalization capabilities of models trained or developed using different datasets. In particular, 16 videos are considered, half of which depict fire, while the remaining are negative samples. However, these seem not very challenging. Most of the techniques, including the baseline (Ko et al., 2011), are able to achieve a very low FPR. At the same time, near perfect TPR, equals 99.29%, has been achieved (Dimitropoulos et al., 2014) by means of a model based approach modeling the flame dynamics. Similar performances were later obtained by Torabian et al. (2021) on a larger set of videos by extracting features from consecutive frames and training a binary SVM. In both works, the conducted temporal analyses aim to correctly classify the few videos acquired in a high activity scenario with rather simple machine learning models.

The authors of the RISE dataset (Hsu et al., 2021) split it into train/test subsets in 6 different ways on the basis of the acquisition characteristics, like the camera view. The best baseline presented, based on the Timeception (Hussein, Gavves, & Smeulders, 2019) achieved an average F_1−score over the 6 splits equals 0.823. This has been recently surpassed in Cao et al. (2021), whose EFFNet reached 0.837. With respect to the baseline, the latter approach has twice the number of training parameters, but the overall computational complexity is similar since only 8 frames are processed instead of 36. The RISE dataset is adopted by the authors of Tao, Lu et al. (2022) and Tao, Xie et al. (2022) to validate their methods, achieving a TPR equal to 88.95% and 89.15%, and an FPR equals 8.73% and 10.17%, respectively. Unfortunately, their performance cannot be directly compared with those of the other techniques since a different train/test split has been carried out. All these top performing methods had to fully exploit the powerful spatio-temporal data representation learning of recent architectures to recognize industrial smoke emissions.

Overall, most of the available datasets seem not challenging enough to estimate the methods' performance in real world applications. In particular, the scarcity of data acquired in high activity scenarios prevents shedding light on the real performance that these techniques may reach in smart cities. The major challenges posed by these datasets are the positive samples where fire occupies a small fraction of the scene, and negative samples containing moving fire-like objects.

5.2.2. Moving camera datasets

Datasets acquired by moving cameras do not represent the typical video surveillance application, however, some of them have been used to validate the performance of several video based fire detectors. Here we briefly discuss the performance achieved on the FireNet (Jadon et al., 2019), the FireSense (Kose et al., 2010), the FiSmo (Ayala et al., 2020), and the FURG (Steffens et al., 2015) datasets.

The FireNet dataset (Jadon et al., 2019) is provided together with a baseline. This shallow CNN achieves an accuracy of 93.91% and an F_1-measure equals 95%. The generalization capability of this network has been confirmed by testing its performance on the MIVIA FIRE dataset (Foggia et al., 2015) and achieving an accuracy of 96.53% and F_1-measure equals 96.49% In a more recent work (Saponara et al., 2021), the proposed YOLO architecture, finetuned using Kaggle data, improved the baseline performance reaching an Accuracy equal to 96.58% and an F_1-measure equal to 95.4%.

The FireSense dataset, presented in Kose et al. (2010), is mainly composed of short range videos, with a good percentage of high activity backgrounds. To the best of our knowledge, the only published work (Zhao et al., 2020) tested on the FireSense dataset, exploiting the LSTM architecture, achieved a good TPR equal to 96.18% with no false negatives.

In Ayala et al. (2020), the FiSmo still images are used to validate the proposed KutralNet method. This technique achieved a validation accuracy equal to 90.72%, which is below the ResNet50 performance but using a lightweight CNN and a very small input resolution. This technique seems to not generalize well on the FiSmo test set, for which the accuracy drops to 77.43%. This drawback was partially alleviated by KutralNet+ (Ayala et al., 2022), which reached 80.62%. At the same time, the F_1−score was increased from 0.8094 to 0.8407. The FiSmo dataset is used to train flame segmentation methods, too. As an example, the technique presented in Choi et al. (2021) and based on FusionNet (Quan, Hildebrand, & Jeong, 2021), was employed to perform flame segmentation on the positive samples of the FiSmo dataset. The deep architecture and the high resolution adopted, i.e., 640×640, yields very good pixel level performance, with an accuracy equal to 99.19% and a F_1−score equal to 0.8491. However, on top of this fire segmentation technique, no video level classification approach is proposed.

The FURG dataset (Steffens et al., 2015) baseline performance was improved by the method presented in Steffens et al. (2016), whose frame level accuracy on the test set is 91.9%, with a F_1−score equals 0.937. The transformer-based method presented in Mardani et al. (2023) further improved these results, achieving an accuracy of 97% and a F_1−score of 0.98. The FURG dataset was used in other works, too, like in Hogan et al. (2021) and Yang et al. (2019). However, the adopted protocol, which merges data from different sources, does not allow a direct performance comparison with previous works. In Hogan et al. (2021), the proposed technique reached an accuracy equal to 95.44%. Instead, in Hogan et al. (2021) the EfficientDet-D3 architecture was employed for fire object detection, achieving an AP equal to 78.3%, an AP_{50} equal to 94.0%, and an AP_{75} equal to 89.2%.

Moving camera datasets are more challenging due to their moving background. However, this is counterbalanced by the closest acquisition range, which often makes the fire flames clearly visible. Like the static camera datasets, the quality of the available datasets should be improved, for example including more challenging scenarios like video acquired by high speed trains or under tunnels.

6. Conclusions

In this survey work, we carried out a review of the literature about fire detection systems from ground video surveillance cameras in an application oriented fashion. The survey is based on the analysis of 153 papers published in the literature and 17 publicly available fire detection datasets. The underlying rationale guiding this analysis is to revisit the methods and datasets under a novel taxonomy of the application scenarios built starting from the real world challenges rather than the existing datasets. We defined new scenarios based on the size of the fire and the environmental activity level, thus we highlighted the needs and challenges of each situation in which a system may operate. Consequently, we analyzed to which extent recently proposed methods fit the needs of each scenario. To validate the proposed taxonomy and to give other researchers the possibility to test their methods on the available datasets in the suggested scenarios, we annotated 536 videos collected from 9 relevant video datasets according to this taxonomy and shared these annotations with the community.

7. Future works

Thanks to the comprehensive analysis of the literature, we can identify possible future research directions to improve fire detection effectiveness and robustness; in particular, the focus should be on dataset collection, experimental protocols, and methodological design choices.

7.1. Dataset collection

The literature review made clear the need for more datasets faithfully representing specific application scenarios. Static cameras should be preferred, since standard ground video surveillance systems for fire detection do not have moving cameras. An effort to improve the high activity scenarios fire dataset should be made since they are the less represented yet the more challenging ones. Also, they are the scenarios in which most real world applications operate. For example, to the best of our knowledge, there is no data to evaluate fire detection performance in tunnels with moving vehicles, landfills and warehouses with working machinery, production lines with active manipulators, or similar applications. Dedicated benchmarks should be built through data acquisition campaigns by setting controlled fires. Another major drawback of the available datasets is the lack of sufficient samples depicting the fire ignition. Since most of the videos start with a fully developed fire, they prevent verifying the early detection capability of the system and the detection delay. At the same time, methods adopting background subtraction cannot work properly. Finally, negative samples should receive the same attention as positive ones. All datasets should include fire-like objects, e.g., depending on the specific scenario, headlights, reflections, sunlight, flags, clouds, fog, or dust. These are crucial to reduce the false positive rate of the system, thus making it more reliable.

7.2. Experimental protocols

These should be redesigned to meet the requirements of the application scenarios. The methods should be tested at least in the scenarios for which they are proposed, using videos containing the typical challenges of those scenarios; if the authors claim that an approach is effective in all the real use cases, they need to demonstrate comparable performance across all scenarios. The datasets, collected according to the guidelines listed above, must be organized in scenarios to allow these experiments; moreover, standards must be defined a priori for the division into training, validation and test sets, to guarantee the repeatability of experiments and results, often overlooked in existing experimental analyses. In this survey, we have done a first step in this direction by annotating all the videos in the existing datasets with the information regarding the scenario and making these labels available (see Section 5.1); we are confident that this direction will also be followed by other researchers in the future.

7.3. Methodological design choices

The fire region proposal and recognition algorithms should be defined according to the requirements of the specific application scenario considered for the method. The size in pixels of the flame or smoke in the image should guide the design choices in terms of smoke and/or flame detection, image resolution and necessity to perform fire region proposal. The environmental activity should drive the choice of fire region proposal and recognition methods, both in terms of classification algorithm and temporal analysis. While some existing methods are probably effective enough for scenarios with low activity, independently on the size of the fire, there is certainly room for improvement on the approaches suitable for high activity scenarios, especially with small fires. Promising algorithms based on multi-frame region proposal and recognition (RNNs, 3D CNNs and Vision Transformers) may be able to combine spatial and temporal information to deal with the challenges posed by high activity scenarios. Another possibility to explore is to have methods that can be configured according to the application scenario. The video analysis modules could be activated or deactivated according to the complexity of the scenario, to dynamically adapt them to the operating conditions (e.g., work hours/days or vacancy). In short range with low activity scenario, the algorithm may process low resolution images to recognize smoke or flames in real time; in the same scenario, by only increasing the range, high resolution images could be acquired thus trading higher resolution for a higher time complexity and, perhaps, real time response, and a smoke region proposal module may be activated. In these two scenarios, if there is high activity, In both these cases, when high background activity is present, more complex region proposal and recognition modules could be enabled,

with lower or higher resolution according to the size of the fire. With this modular approach, it is possible to reduce the computational complexity of methods applied in simpler application scenarios and dynamically adapt the solutions to the operating conditions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link https://github.com/MiviaLab/FireDetectionSurvey2023/blob/main/MIVIA_FIRE_DETECTION_VIDEO_ANNOTATIONS.xlsx.

References

Abdusalomov, A. B., Islam, B. M. S., Nasimov, R., Mukhiddinov, M., & Whangbo, T. K. (2023). An improved forest fire detection method based on the detectron2 model and a deep learning approach. *Sensors*, *23*(3), 1512.

Abid, F. (2021). A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technology*, *57*(2), 559–590.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11), 2274–2282.

Akdis, C. A., & Nadeau, K. C. (2022). Human and planetary health on fire. *Nature Reviews Immunology*, *22*(11), 651–652.

Aktas, M., Bayramcavus, A., & Akgun, T. (2019). Multiple instance learning for CNN based fire detection and localization. In *16th IEEE international conference on advanced video and signal based surveillance* (pp. 1–8).

Almeida, J. S., Huang, C., Nogueira, F. G., Bhatia, S., & de Albuquerque, V. H. C. (2022). EdgeFireSmoke: A novel lightweight CNN model for real-time video fire–smoke detection. *IEEE Transactions on Industrial Informatics*, *18*(11), 7889–7898.

An, Q., Chen, X., Zhang, J., Shi, R., Yang, Y., & Huang, W. (2022). A robust fire detection model via convolution neural networks for intelligent robot vision sensing. *Sensors*, *22*(8), 2929.

Ayala, A., Fernandes, B., Cruz, F., Macêdo, D., Oliveira, A. L., & Zanchettin, C. (2020). Kutralnet: A portable deep learning model for fire recognition. In *International joint conference on neural networks* (pp. 1–8).

Ayala, A., Fernandes, B. J. T., Cruz, F., Macêdo, D., & Zanchettin, C. (2022). Convolution optimization in fire classification. *IEEE Access*, *10*, 23642–23658.

Barnich, O., & Van Droogenbroeck, M. (2010). Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, *20*(6), 1709–1724.

Beneduce, R., Hill, R., & Schelle, C. Alert wildfire (group 6).

Bouguettaya, A., Zarzour, H., Taberkit, A. M., & Kechida, A. (2022). A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Processing*, *190*, Article 108309.

Bu, F., & Gharajeh, M. S. (2019). Intelligent and vision-based fire detection systems: A survey. *Image and Vision Computing*, *91*, Article 103803.

Cao, Y., Tang, Q., & Lu, X. (2022). STCNet: spatiotemporal cross network for industrial smoke detection. *Multimedia Tools and Applications*, *81*(7), 10261–10277.

Cao, Y., Tang, Q., Wu, X., & Lu, X. (2021). EFFNet: Enhanced feature foreground network for video smoke source prediction and detection. *IEEE Transactions on Circuits and Systems for Video Technology*.

Cao, Y., Tang, Q., Xu, S., Li, F., & Lu, X. (2022). QuasiVSD: efficient dual-frame smoke detection. *Neural Computing and Applications*, *34*(11), 8539–8550.

Cao, Y., Yang, F., Tang, Q., & Lu, X. (2019). An attention enhanced bidirectional LSTM for early forest fire smoke recognition. *IEEE Access*, *7*, 154732–154742.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *Computer vision–ECCV 2020: 16th European conference* (pp. 213–229).

Cazzolato, M. T., Avalhais, L., Chino, D., Ramos, J. S., de Souza, J. A., Rodrigues, J. F., Jr., et al. (2017). Fismo: A compilation of datasets from emergency situations for fire and smoke analysis. In *Brazilian symposium on databases-SBBD* (pp. 213–223).

Cazzolato, M. T., Bedo, M. V., Costa, A. F., de Souza, J. A., Traina, C., Jr., Rodrigues, J. F., Jr., et al. (2016). Unveiling smoke in social images with the SmokeBlock approach. In *Proceedings of the 31st annual ACM symposium on applied computing* (pp. 49–54).

Celik, T., & Demirel, H. (2009). Fire detection in video sequences using a generic color model. *Fire Safety Journal*, *44*(2), 147–158.

Celik, T., Demirel, H., Ozkaramanli, H., & Uyguroglu, M. (2007). Fire detection using statistical color model in video sequences. *Journal of Visual Communication and Image Representation*, *18*(2), 176–185.

Cetin, A. E. The bilkent VisFire dataset. URL: http://signal.ee.bilkent.edu.tr/VisiFire/index.html.

Çetin, A. E., Dimitropoulos, K., Gouverneur, B., Grammalidis, N., Günay, O., Habiboğlu, Y. H., et al. (2013). Video fire detection–review. *Digital Signal Processing*, *23*(6), 1827–1843.

Chang, Z., Liu, S., Xiong, X., Cai, Z., & Tu, G. (2021). A survey of recent advances in edge-computing-powered artificial intelligence of things. *IEEE Internet of Things Journal*, *8*(18), 13849–13875.

Chaoxia, C., Shang, W., & Zhang, F. (2020). Information-guided flame detection based on faster r-cnn. *IEEE Access*, *8*, 58923–58932.

Chaturvedi, S., Khanna, P., & Ojha, A. (2022). A survey on vision-based outdoor smoke detection techniques for environmental safety. *ISPRS Journal of Photogrammetry and Remote Sensing*, *185*, 158–187.

Chen, X., An, Q., Yu, K., & Ban, Y. (2021). A novel fire identification algorithm based on improved color segmentation and enhanced feature data. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–15.

Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., et al. (2019). Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3435–3444).

Chen, T.-H., Wu, P.-H., & Chiou, Y.-C. (2004). An early fire-detection method based on image processing. *Vol. 3*, In *International conference on image processing* (pp. 1707–1710).

Chino, D. Y., Avalhais, L. P., Rodrigues, J. F., & Traina, A. J. (2015). Bowfire: detection of fire in still images by integrating pixel color and texture analysis. In *SIBGRAPI conference on graphics, patterns and images* (pp. 95–102).

Choi, H.-S., Jeon, M., Song, K., & Kang, M. (2021). Semantic fire segmentation model based on convolutional neural network for outdoor image. *Fire Technology*, 1–15.

Conte, D., Foggia, P., Petretta, M., Tufano, F., & Vento, M. (2005). Meeting the application requirements of intelligent video surveillance systems in moving object detection. In *Pattern recognition and image analysis: third international conference on advances in pattern recognition* (pp. 653–662).

de Venâncio, P. V. A., Campos, R. J., Rezende, T. M., Lisboa, A. C., & Barbosa, A. V. (2023). A hybrid method for fire detection based on spatial and temporal patterns. *Neural Computing and Applications*, *35*(13), 9349–9361.

de Venâncio, P. V. A., Lisboa, A. C., & Barbosa, A. V. (2022). An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Computing and Applications*, *34*(18), 15349–15368.

de Venâncio, P. V. A., Rezende, T. M., Lisboa, A. C., & Barbosa, A. V. (2021). Fire detection based on a two-dimensional convolutional neural network and temporal analysis. In *IEEE latin American conference on computational intelligence* (pp. 1–6).

DeepQuest Deep quest AI fire and smoke dataset. URL: https://github.com/DeepQuestAI/Fire-Smoke-Dataset.

Dewangan, A., Pande, Y., Braun, H.-W., Vernon, F., Perez, I., Altintas, I., et al. (2022). FIgLib & SmokeyNet: Dataset and deep learning model for real-time wildland fire smoke detection. *Remote Sensing*, *14*(4), 1007.

Di Lascio, R., Greco, A., Saggese, A., & Vento, M. (2014). Improving fire detection reliability by a combination of videoanalytics. In *International conference image analysis and recognition* (pp. 477–484).

Dimitropoulos, K., Barmpoutis, P., & Grammalidis, N. (2014). Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(2), 339–351.

Dogan, S., Barua, P. D., Kutlu, H., Baygin, M., Fujita, H., Tuncer, T., et al. (2022). Automated accurate fire detection system using ensemble pretrained residual network. *Expert Systems with Applications*, *203*, Article 117407.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Dunnings, A., & Breckon, T. (2018). Experimentally defined convolutional neural network architecture varians for non-temporal real-time fire detection. In *Proc. international conference on image processing* (pp. 1558–1562).

Filonenko, A., Hernández, D. C., & Jo, K.-H. (2017). Fast smoke detection for video surveillance using CUDA. *IEEE Transactions on Industrial Informatics*, *14*(2), 725–733.

Filonenko, A., Kurnianggoro, L., & Jo, K.-H. (2017). Smoke detection on video sequences using convolutional and recurrent neural networks. In *Computational collective intelligence: 9th international conference* (pp. 558–566).

Foggia, P., Saggese, A., & Vento, M. (2015). Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(9), 1545–1556.

Frizzi, S., Bouchouicha, M., Ginoux, J.-M., Moreau, E., & Sayadi, M. (2021). Convolutional neural network for smoke and fire semantic segmentation. *IET Image Processing*, *15*(3), 634–647.

Gaur, A., Singh, A., Kumar, A., Kulkarni, K. S., Lala, S., Kapoor, K., et al. (2019). Fire sensing technologies: A review. *IEEE Sensors Journal*, *19*(9), 3191–3202.

Gaur, A., Singh, A., Kumar, A., Kumar, A., & Kapoor, K. (2020). Video flame and smoke based fire detection algorithms: A literature review. *Fire Technology*, *56*(5), 1943–1980.

Geetha, S., Abhishek, C., & Akshayanat, C. (2021). Machine vision based fire detection techniques: a survey. *Fire Technology*, *57*(2), 591–623.

Ghosh, R., & Kumar, A. (2022). A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire. *Multimedia Tools and Applications*, *81*(27), 38643–38660.

Gong, X., Hu, H., Wu, Z., He, L., Yang, L., & Li, F. (2022). Dark-channel based attention and classifier retraining for smoke detection in foggy environments. *Digital Signal Processing*, *123*, Article 103454.

Gragnaniello, D., Greco, A., Sansone, C., & Vento, B. (2023). Onfire contest 2023: real-time fire detection on the edge. In *International conference on image analysis and processing* (pp. 273–281). Springer.

Gu, K., Xia, Z., Qiao, J., & Lin, W. (2019). Deep dual-channel neural network for image-based smoke detection. *IEEE Transactions on Multimedia*, *22*(2), 311–323.

Halofsky, J. E., Peterson, D. L., & Harvey, B. J. (2020). Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific northwest, USA. *Fire Ecology*, *16*(1), 1–26.

Harkat, H., Nascimento, J. M., Bernardino, A., & Ahmed, H. F. T. (2023). Fire images classification based on a handcraft approach. *Expert Systems with Applications*, *212*, Article 118594.

Hogan, I., Qiao, D., Luo, R., Moattari, M., Carthy, A., Zulkernine, F., et al. (2021). FireWarn: Fire hazards detection using deep learning models. In *IEEE international conference on cognitive machine intelligence* (pp. 1–10).

Horng, W.-B., Peng, J.-W., & Chen, C.-Y. (2005). A new image-based real-time flame detection method using color analysis. In *IEEE networking, sensing and control* (pp. 100–105).

Hosseini, A., Hashemzadeh, M., & Farajzadeh, N. (2022). UFS-net: A unified flame and smoke detection method for early detection of fire in video surveillance applications using CNNs. *Journal of Computer Science*, *61*, Article 101638.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hsu, Y.-C., Huang, T.-H. K., Hu, T.-Y., Dille, P., Prendi, S., Hoffman, R., et al. (2021). Project RISE: Recognizing industrial smoke emissions. *Vol. 35*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 14813–14821).

Hu, Y., & Lu, X. (2018). Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features. *Multimedia Tools and Applications*, *77*, 29283–29301.

Hu, Y., Zhan, J., Zhou, G., Chen, A., Cai, W., Guo, K., et al. (2022). Fast forest fire smoke detection using MVMNet. *Knowledge-Based Systems*, *241*, Article 108219.

Huang, J., He, Z., Guan, Y., & Zhang, H. (2023). Real-time forest fire detection by ensemble lightweight YOLOX-L and defogging method. *Sensors*, *23*(4), 1894.

Huang, L., Liu, G., Wang, Y., Yuan, H., & Chen, T. (2022). Fire detection in video surveillances using convolutional neural networks and wavelet transform. *Engineering Applications of Artificial Intelligence*, *110*, Article 104737.

Huo, Y., Zhang, Q., Jia, Y., Liu, D., Guan, J., Lin, G., et al. (2022). A deep separable convolutional neural network for multiscale image-based smoke detection. *Fire Technology*, 1–24.

Huo, Y., Zhang, Q., Zhang, Y., Zhu, J., & Wang, J. (2022). 3DVSD: An end-to-end 3D convolutional object detection network for video smoke detection. *Fire Safety Journal*, *134*, Article 103690.

Hussein, N., Gavves, E., & Smeulders, A. W. (2019). Timeception for complex action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 254–263).

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360.

Jadon, A., Omama, M., Varshney, A., Ansari, M. S., & Sharma, R. (2019). FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications. arXiv preprint arXiv:1905.11922.

Jain, A., & Srivastava, A. (2021). Privacy-preserving efficient fire detection system for indoor surveillance. *IEEE Transactions on Industrial Informatics*, *18*(5), 3043–3054.

Jiang, M., Zhao, Y., Yu, F., Zhou, C., & Peng, T. (2022). A self-attention network for smoke detection. *Fire Safety Journal*, *129*, Article 103547.

Jin, C., Wang, T., Alhusaini, N., Zhao, S., Liu, H., Xu, K., et al. (2023). Video fire detection methods based on deep learning: Datasets, methods, and future directions. *Fire*, *6*(8), 315.

Khan, Z. A., Hussain, T., Ullah, F. U. M., Gupta, S. K., Lee, M. Y., & Baik, S. W. (2022). Randomly initialized CNN with densely connected stacked autoencoder for efficient fire detection. *Engineering Applications of Artificial Intelligence*, *116*, Article 105403.

Khan, S., Muhammad, K., Mumtaz, S., Baik, S. W., & de Albuquerque, V. H. C. (2019). Energy-efficient deep CNN for smoke detection in foggy IoT environment. *IEEE Internet of Things Journal*, *6*(6), 9237–9245.

Khudayberdiev, O., Zhang, J., Abdullahi, S. M., & Zhang, S. (2022). Light-FireNet: an efficient lightweight network for fire detection in diverse environments. *Multimedia Tools and Applications*, *81*(17), 24553–24572.

Khudayberdiev, O., Zhang, J., Elkhalil, A., & Balde, L. (2022). Fire detection approach based on vision transformer. In *Artificial intelligence and security: 8th international conference, ICAIS 2022, Qinghai, China, July 15–20, 2022, proceedings, part i* (pp. 41–53).

Kim, B., & Lee, J. (2019). A video-based fire detection using deep learning models. *Applied Sciences*, *9*(14), 2862.

Ko, B. C., Ham, S. J., & Nam, J. Y. (2011). Modeling and formalization of fuzzy finite automata for detection of irregular fire flames. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(12), 1903–1912, URL: https://cvpr.kmu.ac.kr/.

Komarasamy, D., Gokuldhev, M., Hermina, J. J., Gokulapriya, M., & Manju, M. (2020). Review for detecting smoke and fire in forest using different technologies. *Vol. 993*, In *IOP conference series: materials science and engineering*. Article 012056.

Körschens, M., Bodesheim, P., & Denzler, J. (2022). Beyond global average pooling: Alternative feature aggregations for weakly supervised localization. In *VISIGRAPP (4: VISAPP)* (pp. 180–191).

Kose, K., Tsalakanidou, F., Besbes, H., Tlili, F., Governeur, B., Pauwels, E., et al. (2010). FireSense: fire detection and managment through a multi-sensor network for protection of cultural heritage areas from the risk of fire and extreme weather conditions. *Framework Programmes for Research and Technological Development*.

Li, X., Chen, Z., Wu, Q. J., & Liu, C. (2018). 3D parallel fully convolutional networks for real-time video wildfire smoke detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(1), 89–103.

Li, Z., Mihaylova, L., & Yang, L. (2021). A deep learning framework for autonomous flame detection. *Neurocomputing*, *448*, 205–216.

Li, S., Yan, Q., & Liu, P. (2020). An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism. *IEEE Transactions on Image Processing*, *29*, 8467–8475.

Li, Y., Zhang, W., Liu, Y., & Jin, Y. (2022). A visualized fire detection method based on convolutional neural network beyond anchor. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *52*(11), 13280–13295.

Li, Y., Zhang, W., Liu, Y., Jing, R., & Liu, C. (2022). An efficient fire and smoke detection algorithm based on an end-to-end structured network. *Engineering Applications of Artificial Intelligence*, *116*, Article 105492.

Li, T., Zhao, E., Zhang, J., & Hu, C. (2019). Detection of wildfire smoke images based on a densely dilated convolutional network. *Electronics*, *8*(10), 1131.

Lin, G., Zhang, Y., Xu, G., & Zhang, Q. (2019). Smoke detection on video sequences using 3D convolutional neural networks. *Fire Technology*, *55*, 1827–1847.

Liu, C.-B., & Ahuja, N. (2004). Vision based fire detection. *Vol. 4*, In *Proceedings of the 17th international conference on pattern recognition, 2004* (pp. 134–137). IEEE.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). Ssd: Single shot multibox detector. In *Computer vision–ECCV 2016: 14th European conference* (pp. 21–37).

Liu, N., & Han, J. (2016). Dhsnet: Deep hierarchical saliency network for salient object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 678–686).

Luo, Y., Zhao, L., Liu, P., & Huang, D. (2018). Fire smoke detection algorithm based on motion characteristic and convolutional neural networks. *Multimedia Tools and Applications*, *77*, 15075–15092.

Majid, S., Alenezi, F., Masood, S., Ahmad, M., Gündüz, E. S., & Polat, K. (2022). Attention based CNN model for fire detection and localization in real-world images. *Expert Systems with Applications*, *189*, Article 116114.

Marbach, G., Loepfe, M., & Brupbacher, T. (2006). An image processing technique for fire detection in video images. *Fire Safety Journal*, *41*(4), 285–289.

Mardani, K., Vretos, N., & Daras, P. (2023). Transformer-based fire detection in videos. *Sensors*, *23*(6), 3035.

Muhammad, K., Ahmad, J., & Baik, S. W. (2018). Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, *288*, 30–42.

Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., & Baik, S. W. (2018). Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *49*(7), 1419–1434.

Muhammad, K., Ahmad, J., Mehmood, I., Rho, S., & Baik, S. W. (2018). Convolutional neural networks based fire detection in surveillance videos. *IEEE Access*, *6*, 18174–18183.

Muhammad, K., Khan, S., Elhoseny, M., Ahmed, S. H., & Baik, S. W. (2019). Efficient fire detection for uncertain surveillance environment. *IEEE Transactions on Industrial Informatics*, *15*(5), 3113–3122.

Nguyen, M. D., Vu, H. N., Pham, D. C., Choi, B., & Ro, S. (2021). Multistage real-time fire detection using convolutional neural networks and long short-term memory networks. *IEEE Access*, *9*, 146667–146679.

Nolan, R. H., Collins, L., Leigh, A., Ooi, M. K., Curran, T. J., Fairman, T. A., et al. (2021). Limits to post-fire vegetation recovery under climate change. *Plant, Cell & Environment*, *44*(11), 3471–3489.

Oh, S. H., Ghyme, S. W., Jung, S. K., & Kim, G.-W. (2020). Early wildfire detection using convolutional neural network. In *Frontiers of computer vision: 26th international workshop* (pp. 18–30).

Park, M., & Ko, B. C. (2020). Two-step real-time night-time fire detection in an urban environment using static ELASTIC-YOLOv3 and temporal fire-tube. *Sensors*, *20*(8), 2202.

Prema, C. E., Suresh, S., Krishnan, M. N., & Leema, N. (2022). A novel efficient video smoke detection algorithm using co-occurrence of local binary pattern variants. *Fire Technology*, *58*(5), 3139–3165.

Pundir, A. S., & Raman, B. (2019). Dual deep learning model for image based smoke detection. *Fire Technology*, *55*(6), 2419–2442.

Qian, H., Shi, F., Chen, W., Ma, Y., & Huang, M. (2022). A fire monitoring and alarm system based on channel-wise pruned YOLOv3. *Multimedia Tools and Applications*, 1–19.

Quan, T. M., Hildebrand, D. G. C., & Jeong, W.-K. (2021). Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics. *Frontiers in Computer Science*, *3*, Article 613981.

Santhosh, K. K., Dogra, D. P., & Roy, P. P. (2020). Anomaly detection in road traffic using visual surveillance: A survey. *ACM Computing Surveys*, *53*(6), 1–26.

Saponara, S., Elhanashi, A., & Gagliardi, A. (2021). Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *Journal of Real-Time Image Processing*, *18*(3), 889–900.

Shahid, M., Chien, I., Sarapugdi, W., Miao, L., Hua, K.-L., et al. (2021). Deep spatial-temporal networks for flame detection. *Multimedia Tools and Applications*, *80*(28), 35297–35318.

Shahid, M., & Hua, K.-l. (2021). Fire detection using transformer network. In *Proceedings of the international conference on multimedia retrieval* (pp. 627–630).

Shahid, M., Virtusio, J. J., Wu, Y.-H., Chen, Y.-Y., Tanveer, M., Muhammad, K., et al. (2021). Spatio-temporal self-attention network for fire detection and segmentation in video surveillance. *IEEE Access*, *10*, 1259–1275.

Shakhnoza, M., Sabina, U., Sevara, M., & Cho, Y.-I. (2022). Novel video surveillance-based fire and smoke classification using attentional feature map in capsule networks. *Sensors*, *22*(1), 98.

Sharma, J., Granmo, O.-C., Goodwin, M., & Fidje, J. T. (2017). Deep convolutional neural networks for fire detection in images. In *Engineering applications of neural networks: 18th international conference* (pp. 183–193).

Shen, J., Hao, X., Liang, Z., Liu, Y., Wang, W., & Shao, L. (2016). Real-time superpixel segmentation by DBSCAN clustering algorithm. *IEEE Transactions on Image Processing*, *25*(12), 5933–5942.

Sheng, D., Deng, J., & Xiang, J. (2021). Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access*, *9*, 63933–63942.

Shi, X., Lu, N., & Cui, Z. (2019). Smoke detection based on dark channel and convolutional neural networks. In *5th international conference on big data and information analytics* (pp. 23–28).

Shi, J., Wang, W., Gao, Y., & Yu, N. (2020). Optimal placement and intelligent smoke detection algorithm for wildfire-monitoring cameras. *IEEE Access*, *8*, 72326–72339.

Steffens, C. R., Botelho, S. S. D. C., & Rodrigues, R. N. (2016). A texture driven approach for visible spectrum fire detection on mobile robots. In *Latin American robotics symposium and IV Brazilian robotics symposium* (pp. 257–262).

Steffens, C. R., Rodrigues, R. N., & da Costa Botelho, S. S. (2015). An unconstrained dataset for non-stationary video based fire detection. In *Robotics symposium (LARS) and 2015 3rd Brazilian symposium on robotics (LARS-SBR), 2015 12th Latin American* (pp. 25–30).

Tao, H., & Duan, Q. (2023). An adaptive frame selection network with enhanced dilated convolution for video smoke recognition. *Expert Systems with Applications*, *215*, Article 119371.

Tao, H., Lu, M., Hu, Z., Xin, Z., & Wang, J. (2022). Attention-aggregated attribute-aware network with redundancy reduction convolution for video-based industrial smoke emission recognition. *IEEE Transactions on Industrial Informatics*, *18*(11), 7653–7664.

Tao, H., Xie, C., Wang, J., & Xin, Z. (2022). CENet: A channel-enhanced spatiotemporal network with sufficient supervision information for recognizing industrial smoke emissions. *IEEE Internet of Things Journal*, *9*(19), 18749–18759.

Torabian, M., Pourghassem, H., & Mahdavi-Nasab, H. (2021). Fire detection based on fractal analysis and spatio-temporal features. *Fire Technology*, *57*(5), 2583–2614.

Töreyin, B. U., Dedeoğlu, Y., Güdükbay, U., & Cetin, A. E. (2006). Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters*, *27*(1), 49–58.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 4489–4497).

Tuna, H., Onaran, I., & Cetin, A. E. (2009). Image description using a multiplier-less operator. *IEEE Signal Processing Letters*, *16*(9), 751–753.

Verstockt, S., Poppe, C., Van Hoecke, S., Hollemeersch, C., Merci, B., Sette, B., et al. (2012). Silhouette-based multi-sensor smoke detection: coverage analysis of moving object silhouettes in thermal and visual registered images. *Machine Vision and Applications*, *23*, 1243–1262.

Villela, K., Nass, C., Novais, R., Simões, P., Traina, A., Rodrigues, J., et al. (2018). Reliable and smart decision support system for emergency management based on crowdsourcing information. *Exploring Intelligent Decision Support Systems: Current State and New Trends*, 177–198.

Wang, P., Zhang, J., & Zhu, H. (2021). Fire detection in video surveillance using superpixel-based region proposal and ESE-ShuffleNet. *Multimedia Tools and Applications*, 1–28.

Wu, Z., Xue, R., & Li, H. (2022). Real-time video fire detection via modified YOLOv5 network model. *Fire Technology*, *58*(4), 2377–2403.

Xie, Y., Zhu, J., Cao, Y., Zhang, Y., Feng, D., Zhang, Y., et al. (2020). Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features. *IEEE Access*, *8*, 81904–81917.

Xie, Y., Zhu, J., Guo, Y., You, J., Feng, D., & Cao, Y. (2022). Early indoor occluded fire detection based on firelight reflection characteristics. *Fire Safety Journal*, *128*, Article 103542.

Xu, Z., Wanguo, W., Xinrui, L., Bin, L., & Yuan, T. (2019). Flame and smoke detection in substation based on wavelet analysis and convolution neural network. In *3rd international conference on innovation in artificial intelligence* (pp. 248–252).

Xu, G., Zhang, Q., Liu, D., Lin, G., Wang, J., & Zhang, Y. (2019). Adversarial adaptation from synthesis to reality in fast detector for smoke detection. *IEEE Access*, *7*, 29471–29483.

Xu, G., Zhang, Y., Zhang, Q., Lin, G., Wang, Z., Jia, Y., et al. (2019). Video smoke detection based on deep saliency network. *Fire Safety Journal*, *105*, 277–285.

Yang, X., Hua, Z., Zhang, L., Fan, X., Zhang, F., Ye, Q., et al. (2023). Preferred vector machine for forest fire detection. *Pattern Recognition*, *143*, Article 109722.

Yang, H., Jang, H., Kim, T., & Lee, B. (2019). Non-temporal lightweight fire detection network for intelligent surveillance systems. *IEEE Access*, *7*, 169257–169266.

Yang, C., Pan, Y., Cao, Y., & Lu, X. (2022). CNN-transformer hybrid architecture for early fire detection. In *International conference on artificial neural networks* (pp. 570–581). Springer.

Yazdi, A., Qin, H., Jordan, C. B., Yang, L., & Yan, F. (2022). Nemo: An open-source transformer-supercharged benchmark for fine-grained wildfire smoke detection. *Remote Sensing*, *14*(16), 3979.

Yin, M., Lang, C., Li, Z., Feng, S., & Wang, T. (2019). Recurrent convolutional network for video-based smoke detection. *Multimedia Tools and Applications*, *78*, 237–256.

Yin, Z., Wan, B., Yuan, F., Xia, X., & Shi, J. (2017). A deep normalization and convolutional neural network for image smoke detection. *IEEE Access*, *5*, 18429–18438.

Yuan, F., Zhang, L., Wan, B., Xia, X., & Shi, J. (2019). Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition. *Machine Vision and Applications*, *30*, 345–358.

Yuan, F., Zhang, L., Xia, X., Huang, Q., & Li, X. (2021). A gated recurrent network with dual classification assistance for smoke semantic segmentation. *IEEE Transactions on Image Processing*, *30*, 4409–4422.

Yuan, F., Zhang, L., Xia, X., Wan, B., Huang, Q., & Li, X. (2019). Deep smoke segmentation. *Neurocomputing*, *357*, 248–260.

Zeng, J., Lin, Z., Qi, C., Zhao, X., & Wang, F. (2018). An improved object detection method based on deep convolution neural network for smoke detection. *Vol. 1*, In *International conference on machine learning and cybernetics* (pp. 184–189).

Zhang, F., Qin, W., Liu, Y., Xiao, Z., Liu, J., Wang, Q., et al. (2020). A dual-channel convolution neural network for image smoke detection. *Multimedia Tools and Applications*, *79*, 34587–34603.

Zhang, Q., Sun, H., Wu, X., & Zhong, H. (2019). Edge video analytics for public safety: A review. *Proceedings of the IEEE*, *107*(8), 1675–1696.

Zhang, R., Zhang, W., Liu, Y., Li, P., & Zhao, J. (2022). An efficient deep neural network with color-weighted loss for fire detection. *Multimedia Tools and Applications*, *81*(27), 39695–39713.

Zhang, J., Zhu, H., Wang, P., & Ling, X. (2021). ATT squeeze U-Net: a lightweight network for forest fire detection and recognition. *IEEE Access*, *9*, 10858–10870.

Zhao, Y., Zhang, J., & Man, K. L. (2020). Lstm-based model for unforeseeable event detection from video data.

Zhong, Z., Wang, M., Shi, Y., & Gao, W. (2018). A convolutional neural network-based flame detection method in video sequence. *Signal, Image and Video Processing*, *12*, 1619–1627.