Neurocomputing 596 (2024) 127975

Contents lists available at ScienceDirect

Neurocomputing

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

Survey Paper

Visual fire detection using deep learning: A survey

Guangtao Cheng [a,*], Xue Chen [b], Chenyi Wang [a], Xiaobo Li [c], Baoyi Xian [a], Hao Yu [d]

[a] School of Information Engineering, Tianjin University of Commerce, Tianjin, China
[b] Law School, Tianjin University, Tianjin, China
[c] School of Marine Science and Technology, Tianjin University, Tianjin, China
[d] Huixiang Intelligent Technology (Tianjin) Co., Ltd, Tianjin, China

ARTICLE INFO

Communicated by J. Yu

Keywords:
Fire detection
Deep learning
Fire classification
Fire location
Fire segmentation

ABSTRACT

Visual Fire Detection (VFD), through the rapid and accurate identification of smoke and flame in images and videos, is crucial for early fire warning and reducing fire hazards. In recent years, the introduction of deep learning has significantly advanced this field, especially in the automatic extraction of discriminative features necessary for VFD. This paper provides a comprehensive review of the latest technological advancements in fire detection using deep learning, offering a broad perspective. Initially, it details the publicly available benchmark datasets widely used in VFD research and the corresponding evaluation metrics, providing a basis for researchers to assess the performance of various algorithms. Subsequently, we propose a systematic categorization framework, dividing VFD tasks into three key directions: fire classification, fire localization, and fire segmentation. For these directions, we thoroughly review the innovative improvements in deep learning models tailored for image and video inputs and discusses how these advancements enhance the accuracy and efficiency of fire detection. Finally, we highlight the challenges in the field and explore future research directions, intending to inspire and guide both newcomers and seasoned researchers in this area.

1. Introduction

Fire poses a significant threat to human safety, health, and property worldwide, highlighting the indispensability of fire detection techniques in modern society and daily life. Traditional fire detection methods primarily rely on point electronic sensors, which suffer from significant response delays as they require the concentration of smoke particles or the temperature at the fire source to reach a predetermined alarm threshold before triggering an alarm [1]. Moreover, these systems are limited in their ability to cope with environmental changes, being mostly suitable for stable indoor environments but impractical in large, open outdoor spaces. Point sensors also fall short in providing comprehensive information of fire circumstances, such as the initial fire location, fire size, fire propagation direction, and fire growth rate. These parameters are crucial for safety analysis, firefighting strategies, and risk assessment. The shortcomings of traditional point sensors-based methods underscore need for innovative fire detection techniques that can overcome these limitations and offer enhanced capabilities for both indoor and outdoor environments.

Visual Fire Detection (VFD) [2], utilizing vision sensors, has become an effective supplement or alternative to traditional fire detection methods. When smoke or flame appear within the camera's field of view, VFD can immediately detect the fire, significantly reducing the inherent alarm delays of traditional point sensors and demonstrating its great potential in early fire detection. Additionally, vision sensors provide critical information about the development of a fire, such as the speed and direction of fire spread, which are crucial for assessing fire evolution and minimizing damage. In recent years, with increasing demands for rapid response, non-contact detection, and systems suitable for various environments, VFD technology has gained broader attention. This technology extensively employs image processing (IP) and machine learning (ML) techniques to analyze images and video data collected by surveillance cameras, effectively issuing alerts at the early stages of a fire.

IP methods capture handcrafted static and dynamic feature representations from smoke or flame to generate fire candidate regions, which are subsequently verified by ML algorithms. Following this pipeline, significant advancements have been achieved in fire detection field. Some survey papers have reviewed these achievements. Çetin et al. [3] presented a survey on video processing techniques for fire location, analyzing uncontrolled fire behavior in visible and infrared (IR) spectral range cameras. The study discussed traditional IP methods utilized for fire location in two distinct scenarios: short-range (< 100 m) and wild environments (> 100 m). Ojo et al. [4] conducted a comprehensive survey of video smoke location methods based on IP and

* Corresponding author.
E-mail address: chenggt@tjcu.edu.cn (G. Cheng).

https://doi.org/10.1016/j.neucom.2024.127975
Received 1 March 2024; Received in revised form 15 May 2024; Accepted 29 May 2024
Available online 31 May 2024
0925-2312/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Table 1
Summary of the published reviews on fire detection.

Networks	Year	Task			Technology			Target	
		Classification	Location	Segmentation	IP	ML	DL	Smoke	Flame
Çetin et al. [3]	2013		✓		✓			✓	✓
Ojo et al. [4]	2014		✓		✓	✓		✓	
Matlani et al. [5]	2018		✓		✓	✓		✓	
Bu et al. [6]	2019	✓	✓		✓	✓	✓	✓	✓
Dhall et al. [7]	2020	✓	✓		✓	✓	✓	✓	✓
Geetha et al. [11]	2021	✓	✓		✓	✓	✓	✓	✓
Chaturvedi et al. [12]	2022	✓	✓	✓	✓	✓	✓	✓	
Our paper	2024	✓	✓	✓			✓	✓	✓

ML methods, reviewing their advantages and disadvantages. Matlani et al. [5] provided a review of various methods for locating smoke through videos, emphasizing the extraction of visual features and the identification of candidate fire regions. Bu et al. [6] reviewed intelligent and vision-based fire location systems based on environmental contexts, discussing applications in forest fires and other environmental scenarios. Dhall et al. [7] conducted a review of existing methods for forest fire disaster management, covering aspects such as fire incident localization, fire behavior analysis, victim health diagnosis, and fire rescue procedures. As depicted in Table 1, studies concerning fire detection utilizing traditional IP and ML methods predominantly concentrate on fire localization, with comparatively less emphasis placed on fire classification and fire segmentation.

The emergence of deep learning (DL) techniques, which autono mously extract robust features from raw image data and optimize feature representations, has catalyzed substantial performance enhancements across various computer vision tasks [8–10]. Since 2017, researchers in the field of fire detection have adopted a range of DL methods to address fire detection problems, ranging from convolutional neural networks (CNNs) to contemporary Transformer based models. These advancements have begun to be received attention in some published surreys, with respective mentions of 5, 2, 19, and 53 papers on DL-based fire detection introduced in [6,7,11], and [12], respectively. Currently, DL methods dominate the state-of-the-art in VFD technology. Given the over 100 publicly available academic papers published in this field, there exists a urgent need for a systematic literature survey to comprehensively identify, categorize, and evaluate these advanced techniques.

In this paper, we are dedicated to consolidating the advancements made in fire detection using DL techniques, reviewing 130 papers, and discussing open challenges and promising directions to promote future studies on fire detection. The primary highlights of the survey are as follows:

- We present a comprehensive review of fire detection techniques based on DL, covering benchmark datasets, performance metrics, various DL methods for fire detection, as well as challenges and research directions in the field.
- In order to provide readers with a quick and clear overview of the recent developments in fire detection, we introduce a systematic taxonomy categorizing fire detection techniques into three branches: fire classification, fire location, and fire segmentation. For each branch, we conduct an in-depth analysis of representative techniques along with their characteristics.
- We discuss several open challenges encountered in fire detection using DL techniques. Additionally, we identify the promising research directions to give insightful guidance for the fire detection community and foster a deeper understanding of the current landscape in fire detection with DL.

The remainder of this survey is organized as follows: Section 2 presents the research taxonomy. Section 3 provides a review of the most widely used fire detection datasets. Section 4 enumerates popular evaluation metrics for fire detection models. Section 5 reviews the fire classification networks based on DL. Section 6 offers a overview of existing DL methods for fire location. Section 7 describes existing DL models for fire segmentation. Section 8 discusses the main challenges and opportunities of fire detection techniques. Finally, Section 9 concludes the paper.

2. Taxonomy

The occurrence of a fire is often determined by the presence of smoke and flame. However, due to the distinct behaviors of smoke and flame, research on fire detection typically involves the investigation of smoke detection, flame detection, and the concurrent detection of both flame and smoke.

In this survey, we take a advanced DL perspective to systematically and comprehensively analyze the recent advancements of fire detection techniques. To facilitate prospective research on different topics, we categorize fire detection techniques into three primary categories: fire classification, fire location, and fire segmentation. The fire classification task involves determining the presence or absence of fire in an image, thereby classifying the entire image into fire and non-fire categories. Fire localization extends beyond mere classification to pinpoint the location of fire within an image through bounding boxes. Fire segmentation, a litter different, involves a dense classification where each pixel in an image is categorized as fire or non-fire. Within each primary category, we further subdivide approaches into image-based and video-based subcategories. The research distribution of 130 fire detection papers selected for the proposed survey is depicted in Fig. 1, and the taxonomy of existing DL techniques for fire detection is systematically summarized in Fig. 2.

As illustrated in Fig. 1, this paper provides a detailed analysis of 130 relevant articles published in recent years. Based on the categorization of research tasks, researchers have shown a significant emphasis on fire localization compared to fire classification and segmentation. Specifically, there are 62 publications focusing on fire localization, constituting 48% of the total articles, indicating a high level of attention and activity in this area. In contrast, both fire classification and segmentation each have 34 articles, accounting for 26% each, demonstrating a comparatively lower emphasis compared to fire localization. In terms of data types, image data predominates in fire detection research. A significant majority, 100 articles, focus on image-based fire detection, representing a high proportion of 77%. In contrast, research on video-based fire detection is relatively limited, with 30 articles published, accounting for 23%. This may reflect the advantages of image data in processing and analysis, as well as the complexity involved in acquiring and handling video data. Regarding the examination of fire patterns, smoke detection has emerged as a primary focus for researchers. A total of 72 articles are dedicated to smoke detection, comprising 55% of the total articles, likely due to the critical role of smoke detection in early fire warning systems. Flame detection has received attention with 27 articles published, accounting for 21%, albeit to a lesser extent compared to smoke detection. Additionally, 31 articles explore the simultaneous detection of smoke and flame, comprising 24%, indicating efforts by some researchers to enhance the accuracy and reliability of fire detection by concurrently identifying both patterns.

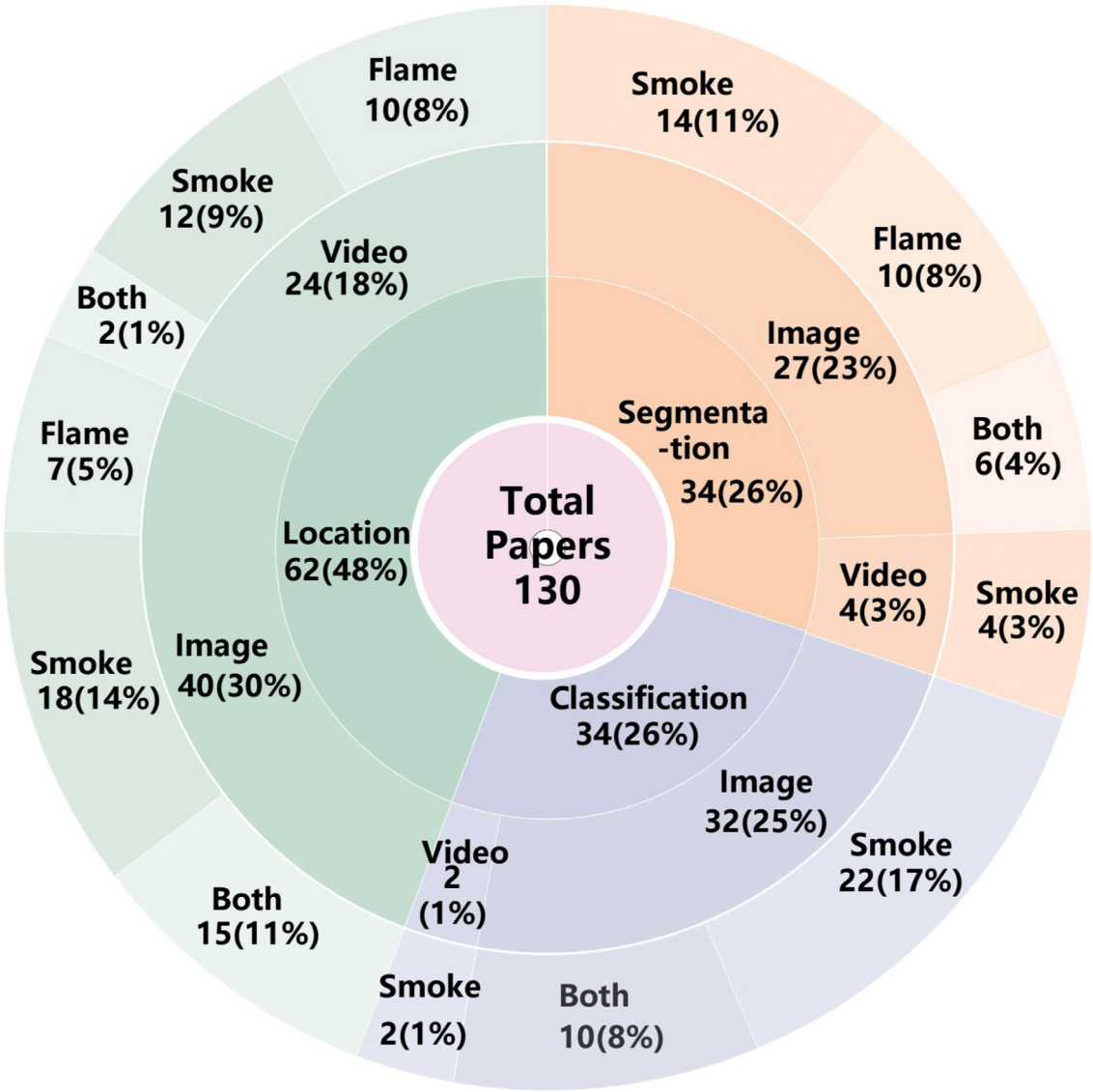


Fig. 1. Statistics of the number of papers on fire detection based on DL.

3. Datasets

DL techniques are highly dependent on the size and quality of the training data. To enable researchers in developing more accurate and reliable fire detection models, numerous state-of-the-art datasets have been publicly released. This advancement aids in enhancing the early identification of fire incidents, contributing significantly to the safeguarding of life and property.

We investigate a total of 11 commonly used datasets, comprising 7 image datasets and 4 video datasets. Representative images from these datasets can be depicted in Fig. 3. Table 2 provides a summary of distinguishing characteristics among these datasets, encompassing data type, fire target, capacity, and usage. These datasets serve as vital contributors to the substantial progress in the fire detection field. They not only provide a common ground for measuring and comparing the performance of various approaches but also drive the field towards addressing increasingly complex, practical, and challenging problems. A detailed summary of the datasets commonly used in VFD is given below.

KMU Fire and Smoke Dataset (CVPR) [13]. The dataset contains 38 videos categorized into four groups: short distance flame, short distance smoke, distant wildfire smoke, and other disturbance objects similar to smoke or flame.

VisiFire Dataset [14]. The dataset comprises of 13 video clips of flames captured from real fire incidents, 21 video clips of smoke collected from various environments, 4 smoke video clips recorded from forest scenes, and 2 other clips.

MIVIA Fire Detection Dataset [15]. The fire dataset collects 31 fire videos acquired from real environments and downloaded from the Internet, including 14 fire videos and 17 non-fire videos. The smoke dataset contains 149 videos with smoke and non-smoke, each spanning 15 min.

Video Smoke Detection Dataset (VSD) [16]. This dataset is released by the research group led by Yuan and holds significant prominence in smoke classification. It comprises of a total of 24,198 images, divided into four subsets. Set 1 consists of 552 smoke and 831 non-smoke images, Set 2 includes 668 smoke and 817 non-smoke images, Set 3 contains 2201 smoke and 8511 non-smoke images, and Set 4 comprises of 2254 smoke and 8363 non-smoke images. The database has been widely used for training and evaluating smoke classification models.

Corsican Dataset [17]. It contains nearly 2000 wildfire images captured through various camera configurations. The images were

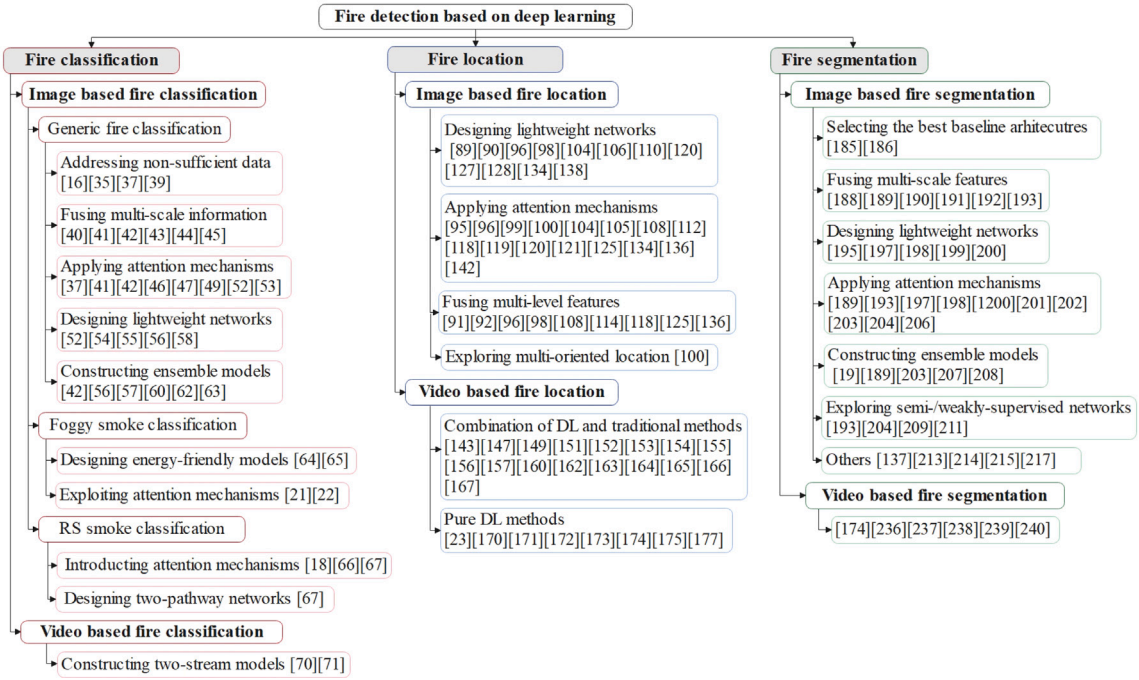


Fig. 2. Structured taxonomy of the existing DL techniques for fire detection.

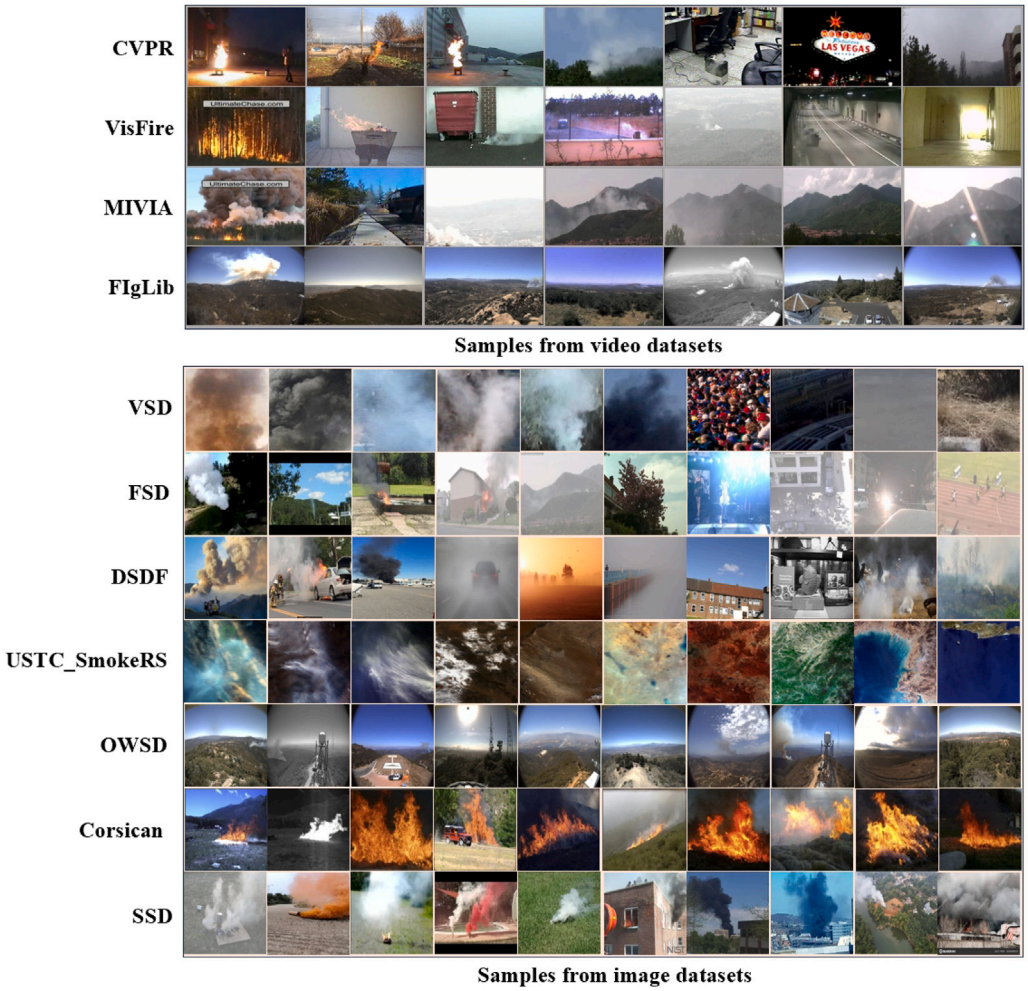


Fig. 3. Representative samples from the publicly available fire detection datasets.

Table 2
A summary of fire detection datasets.

Dataset	Year	Date type	Fire target	Capacity	Usage
CVPR [13]	2012	Video	Flame, smoke	38	Fire location
VisiFire [14]	2014	Video	Flame, smoke	38	Fire location
MIVIA [15]	2015	Video	Flame, smoke	180	Fire location
VSD [16]	2017	Image	Smoke	24 198	Generic smoke classification
Corsican [17]	2017	Image	Flame	2000	Flame segmentation
USTC_SmokeRS [18]	2019	Image	Smoke	6225	Remote sensing smoke classification
SSD [19]	2019	Image	Smoke	70 632	Smoke segmentation
OWSD [20]	2020	Image	Smoke	2192	Smoke detection
FSD [21]	2021	Image	Smoke	33 666	Foggy smoke classification
DSDF [22]	2022	Image	Smoke	18 412	Foggy smoke classification
FlgLib [23]	2022	Video	Smoke	315	Fire location

taken in the visible and near-infrared spectral ranges at a resolution of 1024 × 768 pixels. These image sequences are of significant value for researching multispectral fusion based fire segmentation techniques.

USTC_SmokeRS Dataset [18]. The dataset consists of a total of 6225 satellite images across six scenes: smoke, dust, cloud, land, haze, and seaside. Smoke is considered the target category for forest fire detection. Haze and dust Dust and haze exhibit significant similarities in color and texture to smoke. Cloud consistently appears in satellite images and is prone to being misidentified as smoke during forest fire detection. Land and seaside scenes are considered backgrounds.

Smoke Segmentation Dataset (SSD) [19]. The dataset contains 70,632 smoke images with a resolution of 256 × 256 pixels. These images are generated by blending virtual smoke images with real background images, specifically designed for the smoke segmentation task. The dataset is partitioned into training and validation sets in an 8:2 ratio for training deep smoke segmentation models. Additionally, three virtual test datasets (DS01, DS02, and DS03) are created for performance evaluation. Each test dataset consists of 1000 synthetic smoke images.

Open Wildfire Smoke Dataset (OWSD) [20]. The dataset has been meticulously curated for the development of vision based wildfire detection models. It encompasses 2192 smoke images along with their corresponding annotations stored in XML format.

Foggy Smoke Dataset (FSD) [21]. The dataset is established through a combination of online and offline shooting, specifically designed for smoke classification in a fog scenario. It consists of 33,666 images derived from both real fog and synthetic sources, extracted from 560 videos. This dataset is categorized into four classes: smoke, smoke with fog, nonsmoke, nonsmoke with fog. It covers various smoke patterns with different colors, sizes, dispersion, and concentrations.

DSDF [22]. The dataset contains an extensive collection of 18413 real-work images associated with smoke and fog. The captured samples are divided into four different categories: smoke without fog (SnF), smoke with fog (SF), nonsmoke without fog (nSnF), nonsmoke with fog (nSF). Notably, all samples in the dataset are derived from real-world scenarios, distinguishing it from FSD [21]. This database serves as a benchmark dataset for evaluating the performance of smoke classification algorithms in the foggy environment.

FIgLib Dataset [23]. FIgLib captures sequences of wildfire videos observed by fixed cameras stationed on mountaintops in Southern California, USA. The dataset includes 315 wildfire videos obtained from 101 cameras at 30 stations over the past five years. Each video is sampled at approximately 60-second intervals, resulting in a total of 81 images per wildfire video.

In general, the effectiveness of training deep neural networks is often contingent on the availability of large datasets, both in terms of sample numbers and parameter distribution. However, fire-related incidents are relatively infrequent and unpredictable, posing a challenge in training robust DL models for fire detection. Moreover, the creation of datasets often relies on early published datasets where image quality is markedly poor, resulting in substantial data duplication across different datasets. Consequently, the urgent research imperative in the development of the fire detection field is the collection of large and diverse datasets capable of effectively training DL models.

4. Evaluation metrics

Various performance evaluation metrics are utilized to analyze and compare fire detection models, thereby driving the rapid development of fire detection techniques. In this section, we will provide a brief description of commonly used performance metrics in fire detection methods using DL.

4.1. Metrics for fire classification

Fire classification results can be quantitatively evaluated in terms of true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN). TP signifies a true positive case where both the prediction and actual observation indicate fire; TN refers to an instance where both the prediction and the actual observation denote non-fire; FP occurs when a non-fire sample is predicted as fire; FN indicates an example where the prediction is non-fire while the actual observation is fire. Metrics such as accuracy, precision, recall, F-measure, and false alarm rate, derived from TP, FP, FN, and TN, are widely adopted for the evaluation of fire classification techniques. They are defined as follows.

Accuracy (Acc). This metric represents the proportion of correct predictions among all predictions made.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{1}$$

Precision (P). It is calculated as the ratio of true positives to the total number of samples predicted as positives, reflecting the accuracy of positive predictions.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

Recall (R). It represents the ratio between true positives and the total number of actual positive samples, indicating the model's ability to identify positive instances accurately.

$$R = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

F-measure (F-m). It provides a balanced evaluation of precision and recall metrics.

$$F - m = 2 \times \frac{P \times R}{P + R} \times 100\% \tag{4}$$

False alarm rate (FAR). It quantifies the proportion of falsely classified negative instances (false positives) relative to the total number of negative samples, highlighting the rate of misclassification for negative cases.

$$FAR = \frac{FP}{FP + TN} \times 100\% \tag{5}$$

Computation efficiency. It is an important consideration for real applications of deep neural networks. The number of learnable parameters, the total number of float-point operations (FLOPs), and the number of images processed per second (FPS) are widely applied to quantify the computation efficiency.

4.2. Metrics for fire location

DL based fire localization methods involve first classifying input images into fire and non-fire categories, and then accurately locating fire targets within the images classified as fire. The mean average precision (mAP) is the most frequently selected performance metric for evaluating fire localization models based on DL. It is calculated individually for each class, and then averaged to derive an overall score.

In terms of quantitative localization measures, the intersection over union (IoU) is computed between labeled objects and the model's predictions.

$$IoU(b^p, b^g) = \frac{Area(b^p \cap b^g)}{Area(b^p \cup b^g)} \tag{6}$$

where b^g is the bounding box of the annotated ground truth, and b^p signifies the predicted bounding box from an fire location model. When IoU between the location frame and the annotated frame surpasses a specified threshold, it is deemed as a positive sample; otherwise, it is categorized as a negative sample. Various IoU threshold values, such as 0.25, 0.5, and 0.75, are commonly adopted for fire location model evaluation.

Average precision (AP) indicates the area under the precision–recall curve, and mAP represents the mean AP across all classes and IoU thresholds. The calculation formula for mAP is defined as follows:

$$mAP = \frac{1}{c}\sum_{i=1}^{c} AP_i \tag{7}$$

where c is the total number of classes, and AP_i is the average precision of bounding box prediction for ith class across the considered IoU thresholds.

4.3. Metrics for fire segmentation

Fire segmentation is a task that involves densely classifying each pixel of the input image as either fire or non-fire. To evaluate the performance of models designed for this task, researchers commonly apply the following defined metrics.

Pixel accuracy (PA). It is the proportion of pixels that are correctly classified across all classes to the total number of pixels. PA for c foreground classes is defined as follows:

$$PA = \frac{\sum_{i=0}^{c} p_{ii}}{\sum_{i=0}^{c}\sum_{j=0}^{c} p_{ij}} \times 100\% \tag{8}$$

where p_{ij} is the number of pixels of class i predicted as class j.

Mean pixel accuracy (mPA) extends PA, calculating the proportion of correctly classified pixels for each category separately and then averaging these proportions across all categories:

$$mPA = \frac{1}{c+1}\sum_{i=0}^{c} \frac{p_{ii}}{\sum_{j=0}^{c} p_{ij}} \times 100\% \tag{9}$$

Mean intersection over union (mIoU). It is used to measure the overlap between the predicted results and the ground truth labels. It calculates the average of the intersection-over-union ratios for each class:

$$mIoU = \frac{1}{c}\sum_{i=1}^{c} mIoU_i \tag{10}$$

where c is the total number of classes, and $mIoU_i$ represents IoU for ith class.

Average mean squared error (mMSE). For the ith test image with a resolution of $n_i \times m_i$ pixels, MSE_i is calculated as the average per-pixel squared difference between the prediction result P_i and its corresponding ground truth G_i, and mMSE for N images is defined as:

$$mMSE = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{n_i \times m_i} \frac{\left|P_i^k - G_i^k\right|}{n_i \times m_i} \tag{11}$$

5. Fire classification

In this section, fire classification techniques are commonly categorized into two types: image based fire classification and video based fire classification. Image based classification focuses on static fire features extracted from an individual frame or image. In contrast, video based classification incorporates both static and dynamic fire features by analyzing sequences of frames or videos.

5.1. Image based fire classification

Recently, various CNNs have been widely applied in the field of computer vision. Popular deep networks such as AlexNet [24], VGGNet [25], GoogleNet [26], ResNet [27], DenseNet [28], Xception [29], MobileNetV3 [30], SqueezeNet [31], RepVGG [32], and SEResNet [33] have achieved excellent performance across numerous tasks. Inspired by these models, many researchers have developed CNN techniques for fire classification, further driving the advancement of fire detection technology. This section will comprehensively review DL based fire classification methods.

Based on the type of data, we categorize image based fire classification into three branches: (1) Generic fire classification, where the input image is classified into two categories: fire or non-fire; (2) Foggy smoke classification, which processes the input image through deep models, resulting in classification into four categories: non-smoke, smoke, non-smoke with fog, and smoke with fog; (3) Remote sensing (RS) smoke classification, where the input image is classified into six categories: cloud, dust, haze, land, seaside, and smoke. We provides a detailed introduction to the research progress within each of these three branches.

5.1.1. Generic fire classification

CNN techniques have garnered significant attention from researchers [34], driving the rapid development of methods for visual fire image classification. We will analyze these methods from five key aspects, which are summarized in Fig. 4.

Addressing non-sufficient data. The availability of abundant fire image samples for training and testing poses a significant challenge. Many publicly available datasets addressing this problem are relatively small, hindering the development of reliable fire classification models. To overcome this limitation, several techniques, including data augmentation, transfer learning, and generative adversarial networks (GANs), have been employed to enhance the performance of CNNs when faced with a non-sufficient fire dataset. The pioneering work [16] presented the first approach using CNN for smoke classification. They introduced a deep normalization and convolutional neural network (DNCNN) architecture with 14 layers. To address the scarcity of training samples, data augmentation strategies such as horizontal flip, vertical flip, and random rotation were applied, generating a more extensive training set to mitigate overfitting caused by imbalanced and insufficient samples. Yin et al. [35] combined a deep convolutional generative adversarial network (DCGAN) [36] with a CNN to extract smoke features. DCGAN was used to generate various smoke patterns and numerous false positive objects based on real fire scenes to expand the existing smoke datasets. The subsequent CNN leveraged this augmented dataset for effective smoke feature extraction and classification. Majid et al. [37] applied transfer learning to enhance the feature representation ability of CNN for small-scale smoke datasets. Their study revealed that EfficientNetB0 [38] emerges as the most suitable network choice for fire image classification. Sathishkumar et al. [39] leveraged pre-trained CNN models including VGG16 [25], InceptionV3 [26], and Xception [29] to automatically recognize fire images. They also applied transfer learning to preserve the network's prior capabilities while training it for a new task.

Fusing multi-scale information. The utilization of diverse filters in convolutional operations proves advantageous in extracting different aspects or features from input images, enhancing the multi-scale

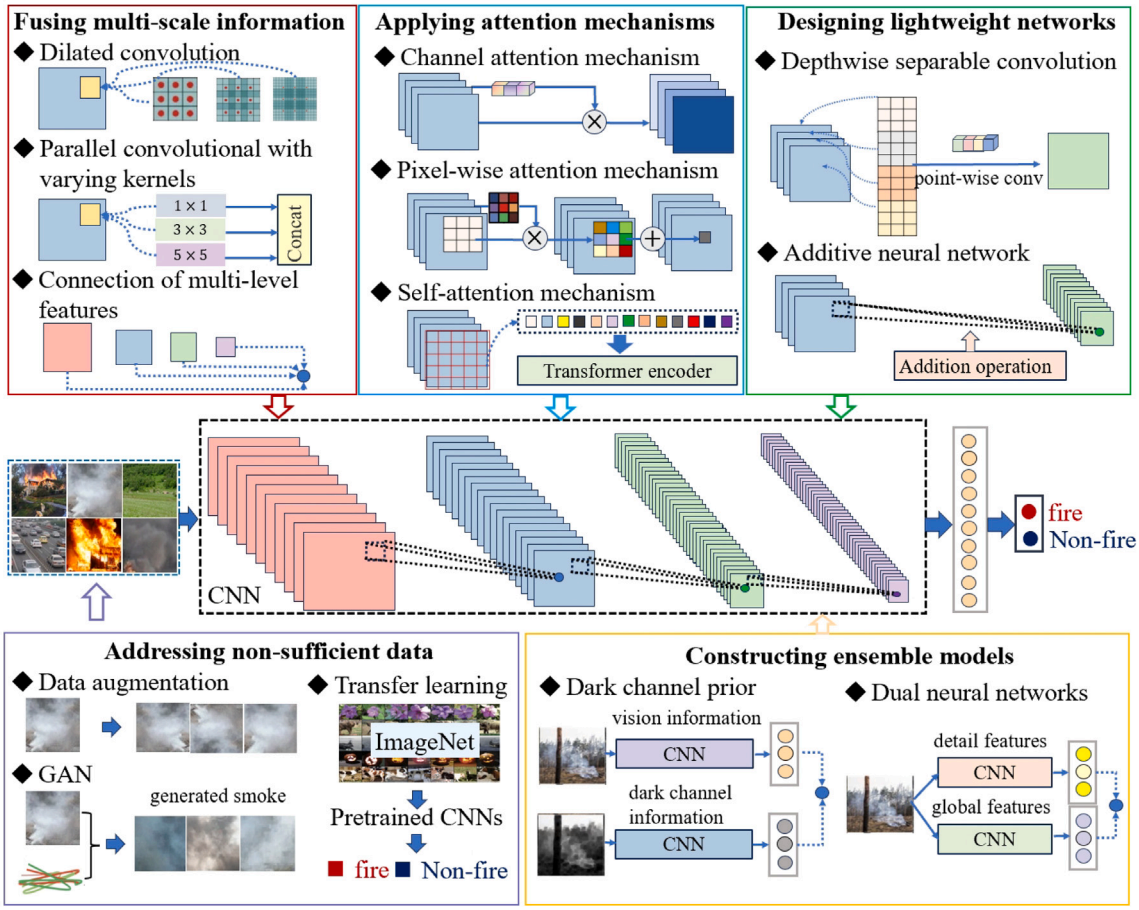


Fig. 4. Overview of generic fire classification methods using DL.

context aggregation of fire targets. Dilated convolution operations are often employed to expand the receptive field, obtaining arbitrary scale channel relevance features by adjusting the dilated rate. Yakhyokhuja et al. [40] employed dilated convolutional layers in constructing a network architecture for flame or smoke recognition. This approach allowed for the avoidance of excessive depth in learning, particularly beneficial for the straightforward binary flame and smoke classification task. Wang et al. [41] introduced a new dilated efficient channel attention module (DECA) for fire image recognition. It extracted fine-grained features through global max pooling (GMP) and global average pooling (GAP), fusing them with various scale features. Li et al. [42] devised a densely dilated convolutional network (DDCN) by selecting dilated convolution instead of common convolutions. This choice optimized the receptive fields for extracting deep and abstract features from smoke images. Yuan et al. [43] proposed a deep multi-scale convolutional neural network (DMCNN) specifically designed for classifying smoke images. This network consisted of stacked basic blocks, each containing parallel convolutional layers with varying kernel sizes but an equal number of filters, enabling the extraction of multi-scale features from smoke images. Jeon et al. [44] proposed a multi-scale prediction framework leveraging feature maps from deeply stacked convolutional layers for fire image recognition. The feature-squeeze (FS) block was designed to squeeze the feature maps spatially for extracting multi-scale information. Zheng et al. [45] developed a modified deep CNN model (MDCNN) for forest fire surveillance. This model integrated feature maps from various resolutions of the VGG network [25], facilitating the effective recognition and analysis of multi-scale fire objects.

Applying attention mechanisms. The integration of attention mechanisms in fire image classification has proven pivotal in directing the model's focus to small and subtle fire targets within an image, effectively reducing background disturbances and extracting more informative features, resulting in significant improvements in fire classification. Li et al. [46] introduced a fire detection technique that integrates multi-scale feature extraction (MFE), implicit deep supervision, and a channel attention mechanism. The MFE module captured richer spatial details, and implicit deep supervision improved information flow through dense skip connections. Additionally, the channel attention mechanism selectively emphasized contributions across diverse feature maps. Cheng et al. [47] proposed a pixel-specific modification to the standard convolution operation, leading to the development of a straightforward and efficient model (PACNN) equipped with pixel-aware attention for smoke image classification. This model dynamically learned pixel-conditioned weights, allowing for the adaptive recalibration of pixel features across feature channels to achieve greater feature diversity and enlarge the feature representation space. Majid et al. [37] incorporated an attention mechanism into the EfficientnNetB0 network [38] to draw model's focus to the relevant sections of fire image, resulting in a significant performance improvement. Wang et al. [41] combined dilated convolution with efficient channel attention (ECA) to classify fires. They utilized multiple convolution layers with diverse kernel sizes to extract channel relevance across various scales, enabling the model to effectively capture features of different sizes and complexities. Li et al. [42] introduced a multiscale-channel attention module (MS-CAM) [48] that incorporates both local and global features to improve its ability to detect fires accurately. Guo et al. [49] applied the channel and spatial attention mechanism to capture informative smoke and fire characteristics within an image, thereby enhancing the accuracy of fire classification.

Compared with CNN focusing on local information, Transformer network [50,51] can capture long-distance dependencies through a self-attention mechanism, deriving global information. Zhao et al. [52] designed an accurate and lightweight neural network that combines CNN and self-attention (CSANet) for smoke image classification. The network utilized multiscale CSA blocks, combining depthwise convolution and asymmetric convolutions to reduce model complexity. The self-attention operation in the Transformer architecture captured the dependencies between spatial information and extracted correlations between channel information. Cheng et al. [53] integrated the beneficial aspects of CNNs into the vision Transformer framework, and presented a novel convolution-enhanced vision Transformer network (CViTNet) specifically designed for smoke classification. They devised a convolutional token embedding to establish a cascaded multi-scale pyramid network structure, combining the advantages of CNN for capturing local features and Transformer for extracting global features.

Designing lightweight networks. Lightweight CNNs with depthwise separable convolution have emerged as a solution to the challenges of model size and speed. Such networks prove beneficial for a wide range of applications, especially on edge or embedded devices with limited resources and computational capabilities. Yang et al. [54] proposed a non-temporal lightweight network using depthwise separable convolution for fire classification. The squeeze and excitation module was further used to enhance the representation of the global features of fire. Yar et al. [55] introduced an innovative fire classification framework centered on a lightweight CNN for resource-constrained devices. Similar to the VGG network architecture [25], the proposed classification model applied small convolutional kernels to extract fine-grained information of fire targets in images. As the network deepens, the number of channels in feature maps increased with the reduction of feature map resolution. Zhang et al. [56] introduced the depthwise separable convolution into the DCNN [57] to improve the efficiency of smoke classification. Yang et al. [52] applied depthwise convolution and asymmetric convolution with large kernels to reduce computational demands. They designed a three-pass fusion method to compensate for the limitations of depthwise convolution in extracting channel information. Pan et al. [58] presented an innovative approach for wildfire smoke classification, introducing the computationally efficient additive deep neural network (AddNet). AddNet is characterized by a unique vector operator that exclusively employs addition and sign manipulation operations [59]. This design choice eliminates multiplicative operations, making AddNet particularly suitable for implementation on standard microprocessors and digital signal processors. The simplicity and computational efficiency of AddNet offer a promising avenue for enhancing wildfire smoke classification capabilities without imposing a heavy computational burden.

Constructing ensemble models. Constructing ensemble models involving multiple complementary CNN architectures has proven effective in enhancing visual feature representation for fire targets in images. In these models, the final convolutional layers of distinct architectures are concatenated to obtain robust predictions. Liu et al. [60] introduced a dual convolution network (DarkC-DCN) based on dark channel prior [61] for smoke classification. This network extracted features from two different channels of the input image, obtaining both generalized and precise features for a more comprehensive representation of smoke features. Zhang et al. [62] developed a dual-channel network (DC-CNN) to enhance the accuracy of smoke image classification. They applied a common AlexNet [24] subnetwork for generalized feature extraction and a dedicated CNN for specific and detailed features. The fusion of these subnetworks' features improved the overall robustness of the network. Gu et al. [57] proposed an end-to-end deep dual-channel neural network (DCNN) consisting of dual subnetworks. The first subnetwork excelled at extracting detailed features of smoke, while the second subnetwork incorporated skip connections and global average pooling to capture base features. The concatenation of these features enhanced their representation ability, complementing each other. Zhang et al. [56] addressed the efficiency of DCNN [57] by using the depthwise separable convolution. Liu et al. [63] introduced ensemble deep convolutional neural networks (EDCNN) for robust smoke recognition. Different subnetwork architectures were independently trained and fused using an ensemble method to capture diverse aspects of smoke. Li et al. [42] proposed an adversarial fusion network (AFN) for wildfire smoke classification. AFN consisted of a subnetwork based on dilated convolutions and another subnetwork based on attention mechanisms. It reduced the gap between the source domain and the target domain, improving the model's generalization ability.

Performance evaluation. Several representative models for generic smoke image classification were evaluated on the frequently utilized VSD dataset [16]. The larger Set3 was typically augmented to create a balanced training dataset, while Set 4 served as the validation set. The smaller Set 1 and Set 2 were commonly employed as the test sets. In the training procedural of DCNN [16] and IDCNN [56] networks, validation Set4 was also augmented to mimic Set3 and extended to the balanced extension. Recall, accuracy, FAR, number of parameters, and FLOPs indicators were selected for comparison. The initial CNN models were used as benchmark architectures for smoke image classification. The performance scores were gathered from the respective original articles and listed in Tabel 3. Notably, CViTNet [53] effectively combines the advantages of CNN and Transformer network architectures to achieve the best performance in terms of Recall index on Set1 and Set2. Additionally, AFN [42], relying on ensemble models, attains an impressive accuracy of 99.78% on Set 1, performing other models. Meanwhile, smoke classification networks constructed solely using CNN, such as DNCNN [16] and PACNN [47], demonstrate advantages in model complexity, with only 43M FLOPs and 0.3M learnable parameters, respectively. It is discernible from Table 3 that performance on the VSD dataset has approached saturation, indicating the maturity of these models.

5.1.2. Foggy smoke classification

Image data captured under harsh weather conditions often suffers from issues such as low contrast, quality degradation, and blurred details, rendering smoke classification networks designed for normal weather conditions ineffective in addressing these challenges. Currently, researchers have begun to focus on smoke classification in foggy weather environments. In contrast to generic fire image classification, DL models designed for foggy smoke image classification categorize images into four classes: smoke, smoke with fog, non-smoke, and non-smoke with fog. Researchers are striving to address smoke classification in foggy weather by designing energy-friendly models and applying attention mechanism strategies.

Designing energy-friendly models. An energy-efficient lightweight CNN architecture alleviates the computational burden of hardware, making it suitable for the efficient fire analysis based on embedded applications. Khan et al. [64] conducted an investigation into various state-of-the-art CNN models, and designed an energy-efficient system for early smoke classification in surveillance videos captured in both normal and foggy environments. Muhammad et al. [65] contributed to the development of a lightweight CNN architecture tailored for foggy smoke classification. This approach took into account essential requirements such as accuracy, running time, and deployment feasibility, making it well-suited for smoke classification in industrial settings.

Exploiting attention mechanisms. Attention mechanism proves effective in assigning higher weights to regions of fire, mitigating interference from harsh environments, and addressing the challenge of detecting small fire targets. He et al. [21] proposed a smoke classification method applicable to both normal and foggy weather conditions. Their approach introduced an channel-spatial attention mechanism module to address the challenges of classifying small smoke instances effectively. Additionally, they devised a lightweight feature pyramid network (FPN) that not only improves discriminant feature representations between different objects but also reduces the number of model

Table 3
Performance comparison of generic smoke classification models on the public VSD dataset [16].

Networks	Set1			Set2			#Parameters (M)	FLOPs (M)
	Recall	Accuracy	FAR	Recall	Accuracy	FAR		
AlexNet [24]	94.20%	97.54%	0.24%	95.06%	97.54%	0.37%	56.9	721
VGG-16 [25]	96.19%	97.48%	0.60%	95.05%	97.48%	0.48%	120.0	13 680
RepVGG-A1 [32]	94.38%	97.32%	0.72%	94.33%	97.14%	0.49%	12.8	2637
Inception-V4 [26]	94.57%	97.11%	1.20%	96.80%	98.34%	0.37%	13.5	13 386
ResNet-34 [27]	95.29%	97.76%	0.60%	94.33%	97.14%	0.49%	21.3	3671
SEResNet-34 [33]	95.47%	97.32%	1.44%	96.51%	98.14%	0.49%	21.4	3672
DenseNet-121 [28]	95.83%	97.90%	0.72%	96.95%	98.54%	**0.12%**	7.0	2865
DNCNN [16]	95.28%	97.83%	0.48%	96.36%	98.08%	0.48%	20.0	**43**
DMCNN [43]	97.28%	98.63%	0.48%	98.55%	99.14%	0.36%	1.4	–
PACNN [47]	98.01%	98.91%	0.48%	97.67%	98.67%	0.49%	**0.3**	90
CViTNet [53]	**99.82%**	99.20%	1.20%	**99.27%**	99.20%	0.86%	4.5	346
DCNN [57]	97.50%	99.70%	**0.12%**	99.00%	**99.40%**	0.24%	2.7	–
IDCNN [56]	97.80%	98.50%	1.08%	98.80%	99.00%	0.86%	0.6	–
AFN [42]	99.64%	**99.78%**	**0.12%**	–	–	–	–	–
HAM-ResNet [49]	99.40%	99.50%	0.17%	99.10%	**99.40%**	0.24%	–	–

parameters, ensuring real-time performance. In a similar vein, Gong et al. [22] integrated dark-channel-assisted mixed attention and feature fusion techniques into smoke classification model. They introduced dark channel prior information to enhance the model's early feature extraction. This was achieved through the design of a new mask branch, which is responsible for extracting discriminative features from locally focused regions of the image. A channel-spatial attention mechanism was also introduced into the later feature extraction of the model.

Performance evaluation. We present a performance summary of the representative algorithms, utilizing the publicly available FSD [21] and DSDF [22] as benchmark datasets for evaluation. Metrics such as accuracy, precision, and F-measure were applied to quantify the performance of smoke classification algorithms in foggy environments. The values directly borrowed from the corresponding papers were shown in Fig. 5. Unsurprisingly, attention mechanism-based algorithms proposed by He et al. [21] and Gong et al. [22] demonstrate the best performance on FSD and DSDF dataset, respectively. They exhibit robustness and the potential to capture finer representations of smoke targets. However, it is noteworthy that state-of-the-art algorithms achieves accuracy, precision, and F-measure below 88% on the DSDF dataset, which comprises of real-world images. This indicates that smoke classification in harsh environments remains a challenging problem, warranting further exploration.

5.1.3. Remote sensing smoke classification

Remote sensing (RS) images are extensively employed in disaster monitoring due to their broad coverage, timely acquisition, and cost-effectiveness. Wildfires, as a destructive disaster, pose a significant threat to human life and production activities. As advancements in disaster mitigation and prevention technologies continue, researchers are increasingly considering the use of RS images for detecting wildfire. Smoke serves as a crucial indicator of wildfire occurrence, making it a vital signal for early detection. Recent developments in smoke image classification based on remote sensing categorize images into six classes: cloud, dust, haze, land, seaside, and smoke. Attention mechanisms and two-pathway networks are consistently regarded as strategies that enhance model capabilities in identifying and analyzing RS images, thereby contributing significantly to the effectiveness of wildfire monitoring and management.

Introducing attention mechanisms. The integration of attention mechanisms has introduced a new impetus for extracting the most informative features from smoke in images. Ba R et al. [18] underscored the significance of attention mechanisms and presented SmokeNet, a novel CNN model for RS smoke classification. Channel attention [33], spatial attention, and residual attention (RA) modules [67] were seamlessly incorporated into the CNN architecture, enabling the model to concentrate on label-specific information crucial for accurate image classification. Chen et al. [66] took a parallel approach in designing channel attention and spatial attention to extract salient features of smoke signals while eliminating irrelevant information. Leveraging the strengths of both CNN and visual Transformer [51], Zheng et al. [68] proposed a novel forest fire smoke detection model (SR-Net) capable of recognizing smoke or cloud in satellite remote sensing images. The adoption of depthwise separable convolution helped in reducing computational complexity.

Constructing two-pathway networks. To address the critical need for both global information and local features in RS smoke classification, a two-pathway network has been devised to amalgamate multi-level features for conclusive classification. Chen et al. [66] directed their focus towards global information and salient features within RS images, introducing a self-adaptive feature aggregation (SAFA) network with dual pathways for smoke classification, as shown in Fig. 6. The global information extraction pathway (GIEP) enhanced information exchange among multi-level features by establishing connections between low-level and high-level features. Simultaneously, the salient feature extraction pathway (SFEP) integrated channel-spatial attention modules into ResBlocks [27] to emphasize salient and discriminative features. Finally, the outputs of the two subnetworks were weighted and aggregated to obtain the final classification result.

Performance evaluation. We present the quantitative results for several initial CNNs [25,27,28,33,69] and representative RS smoke classification algorithms [18,66] discussed in the preceding sections. The extensively utilized USTC_SmokeRS dataset was chosen as the benchmark dataset for performance evaluation, and the results were depicted in Fig. 7. Leveraging the ensemble models strategy, SAFA [66] attains the highest performance with an accuracy of 96.22%.

5.2. Video based fire classification

Fire classification methods based on the single image often fall short in capturing the dynamic and evolving nature of fire. To address this limitation, a paradigm shift towards exploiting temporal features from video data has been embraced. In this context, two-stream networks have emerged as a powerful architecture, comprising both a spatial stream subnetwork and a temporal stream subnetwork. As illustrated in Fig. 8, these subnetworks collectively extract static and dynamic features, respectively, which are then fused to facilitate robust fire image recognition. Hu et al. [70] introduced a two-stream convolutional neural network for video smoke classification. The spatial stream subnetwork extracted intra-frame appearance features from an individual video frame, while the temporal stream subnetwork captured inter-frame motion information through optical flow analysis between adjacent frames. The enhanced architecture processed concatenated frames, employing a multitask learning strategy to concurrently recognize smoke and compute optical flow. This approach enables the simultaneous capture of both intra-frame and inter-frame features.

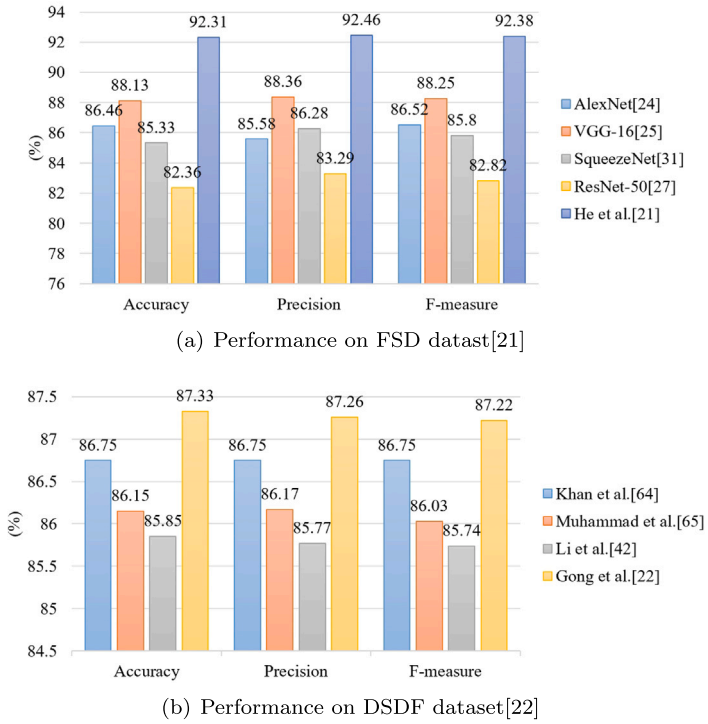


Fig. 5. Performance comparison of foggy smoke classification methods on the public foggy smoke datasets: (a) Performance evaluation on FSD dataset [21]; (b) Performance evaluation on DSDF dataset [22].

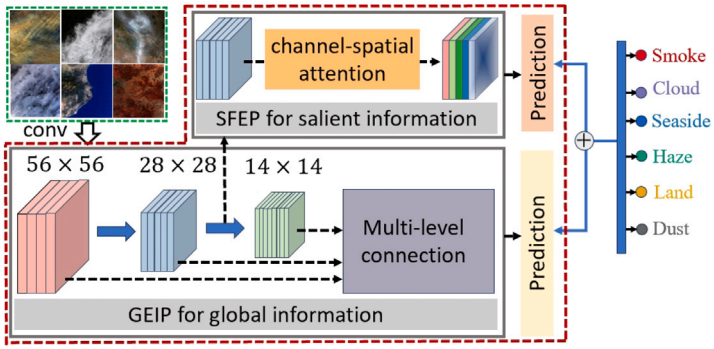


Fig. 6. SAFA structure [66] for RS smoke classification.

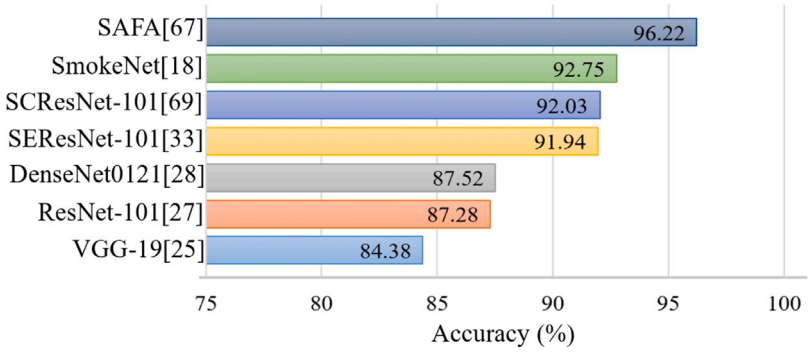


Fig. 7. Performance comparison of RS smoke classification methods on the public USTC_SmokeRS dataset.

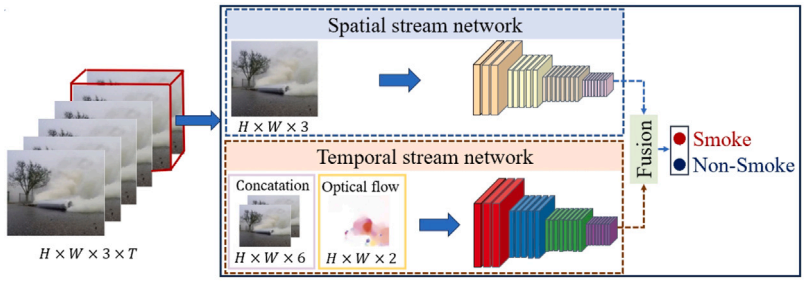


Fig. 8. Illustration of fire classification based on video.

Similarly, Yin et al. [71] proposed a video-based smoke classification method incorporating both spatial and motion networks. The spatial network learned spatial features from a single images, while the motion network focuses on two consecutive video frames to gather motion texture information. These representative features from both subnetworks were fused and fed into a recurrent neural network (RNN) [72] to discern discriminative information across the entire video sequence. Their approaches showcase the advantages of spatial–temporal stream networks on self-created smoke video classification datasets. The video-based smoke classification methods extract both static and dynamic features of fire smoke to enhance the accuracy of smoke classification. However, it requires considerable computational costs. Each single frame needs to be processed by a spatial-stream CNN. Additionally, consecutive frame pairs require computations involving a temporal CNN and optical flow extraction.

6. Fire location

The problem of fire location involves detecting the precise fire position in an image and generating a bounding box around it. The two categories of fire location methods, one based on spatial features from a single image and the other based on the spatial–temporal information from videos, are described as follows.

6.1. Image based fire location

Object detection techniques (OD) [73] in DL have gained widespread attention and have been widely applied in various fields such as autonomous driving, robotics, intelligent surveillance. Fire location based on an image selects popular object detection frameworks as the baselines and improves them in terms of lightweight structures, attention mechanisms, multi-level features, and multi-oriented location. In comparison to CNNs used for fire classification, object detection frameworks are more complex, typically composed of multiple components. Consequently, researchers often introduce a series of modification strategies into existing object detection frameworks to enhance fire localization performance. In contrast to the description of fire classification techniques, we firstly provide an overview of improvement strategies based on different object detection frameworks, and then summarize them.

6.1.1. Fire location based on YOLO series

Among numerous deep object detection frameworks, YOLO (You Only Look Once) [74] has achieved an exceptional balance between computational speed and accuracy, making real-time processing feasible in commercial product applications. Therefore, the YOLO series of object detection frameworks have undergone multiple iterations, continuously addressing various issues in object detection and improving overall performance [75].

The efficacy of advanced object detection models for fire location have been assessed, highlighting the superiority of the YOLO network in terms of both accuracy and speed. Li et al. [76] conducted tests with Faster-RCNN [77], R-FCN [78], SSD [79], and YOLOv3 [80] for flame and smoke location, revealing that YOLOv3 achieved higher accuracy and robustness. Abdusalomov et al. [81] evaluated the default YOLOv3, YOLOv4 [82], and their tiny versions for flame location, selecting YOLOv3 as the baseline due to its superior performance in addressing flame-related challenges. Zheng et al. [83] explored the feasibility of using popular object detection networks such as EfficientDet [84], Faster R-CNN [78], YOLOv3 [80], and SSD [79] to locate forest fire smoke. Results indicated that EfficientDet achieved the highest mAP, while YOLOv3 exhibited the fastest detection speed. Smadi et al. [85] compared the accuracy and speed of different YOLO models, including YOLOv3, YOLOv5 [86], and YOLOv7 [87], for wildfire smoke location. Experimental findings revealed that YOLOv5x outperformed other object detection networks in terms of mAP, albeit with a slower detection rate. Consequently, YOLO variants have become widely adopted for fire location, contributing to enhance accuracy and efficiency.

YOLOv2 [88] incorporated several enhancements over the original YOLOv1 [74], aiming to improve its performance while maintaining the same speed and increasing its capability to detect 9000 categories. With the goal of developing intelligent IoT devices for fire and smoke location in both indoor and outdoor environments, Saponara et al. [89] introduced a real-time and embedded implementation of a fire and smoke locating technique using the YOLOv2 object detection network. The proposed detector was trained on a dataset consisting of 400 images in Matlab and was subsequently deployed on a single-board embedded system, the NVIDIA Jetson Nano, functioning as a standalone application.

YOLOv3 [80] proposed Darknet53 as the backbone network and designed a feature pyramid network (FPN) to fuse information from multi-stage feature maps, enhancing the detection capability of multi-scale objects in an image. Building upon the YOLOv3 network, Qin et al. [90] introduced a fire location method that combines a classification model and a target detection model using deep learning. The depthwise separable convolution was employed for classifying fire images, while the YOLOv3 target regression function was utilized to output the fire location information for images classified as fire. Masoom et al. [91] enhanced the original YOLOv3 network by introducing an additional detection scale to improve its ability to represent small smoke instances in the wild. Principal component analysis (PCA) was applied for feature selection, reducing the dimensional complexity of the improved YOLOv3. Li et al. [92] proposed an adaptive linear feature-reuse network (ALFRNet) to locate smoke with various sizes and shapes. The feature extraction process was refined by the depthwise convolution (ADC) module, adjusting the weight of the convolutional layer under different convolution kernels adaptively based on the geometric distortion of different smoke instances. The double linear feature-reuse (DLFR) module fused feature maps through linear operations, and the hybrid attention-guided (HAG) module incorporated the ADC module between two DLFR modules before CBAM [93]. Additionally, cluster NMS (CNMS) [94] was employed to mitigate large deviations in the detection box and enhance the overall detection performance. Sun et al. [95] integrated channel attention and spatial attention operations [93] into branches in the neck network of YOLOv3 to enhance the discrimination of smoke, flame, and similar objects.

An anchor-free mechanism was introduced to reduce model complexity and improve generalization ability.

YOLOv4 [82] aimed to achieve an optimal balance through numerous modifications categorized as bag-of-freebies and bag-of-specials. Building upon the YOLOv4 object detection framework, Wang et al. [96] introduced a lightweight fire location network aiming for a balance between efficiency and accuracy. The CSPDarknet53 backbone network was replaced with the lightweight MobileNetv3 [30], and the path aggregation structure was transformed into a bidirectional feature pyramid network (BiFPN) [84]. The depthwise separable attention module (DASM) was proposed, incorporating depthwise separable convolution and coordinate attention network (CAN) [97]. Huo et al. [98] presented an improved YOLOv4-based smoke location network. The CSPDarknet53 backbone network of YOLOv4 was expanded by adding a convolution path to enhance the effective extraction of image features. A spatial pyramid pooling (SPP) module was introduced into the early and middle stages of the backbone network to enhance the feature representation capability of multi-scale objects. Additionally, depthwise separable convolutions were applied again to reduce network parameters.

YOLOv5 [86] incorporated several improvements from the YOLOv4 [82] framework but was developed in PyTorch instead of Darknet. Wang et al. [99] selected YOLOv5 m as the baseline model, and proposed a smoke location model by introducing the dynamic anchor box mechanism and an attention mechanism consisting of channel attention and spatial attention in series. Hu et al. [100] addressed the rotation angle of smoke and introduced a multi-oriented forest fire smoke location network (MVMNet) based on the YOLOv5 structure. They replaced the SPP module in the feature extraction network of YOLOv5 with softpool spatial pyramid pooling (Soft-SPP) [101] and incorporated the value conversion-attention mechanism (VAM) module based on a joint weight allocation strategy from the horizontal and vertical directions. Additionally, they proposed the mixed-NMS strategy, combining Skew-NMS [102] and DIoU-NMS [103], considering angle information and center point distance information. Yin et al. [104] introduced an improved lightweight YOLOv5s network to address the challenge of small and thin smoke in the early fire stage. An attention mechanism combining spatial and channel attention was integrated into the original YOLOv5 to reduce scene interference, and the upsampling operation in the feature fusion network was refined to increase the receptive field of the smoke feature map. Xu et al. [105] introduced an enhanced YOLOv5n model for flame and smoke location. They incorporated channel and spatial attention mechanisms into the detection head of the YOLOv5n framework to enhance feature extraction capabilities for fire targets. Furthermore, they replaced the C3 convolutional module in the original network structure with a C2f convolutional module to facilitate information flow. Li et al. [106] conducted an analysis of flame characteristics within the YCbCr color space and developed a shallow-guide-deep (SGD) module for flame localization on an embedded edge computing platform. They opted for ShuffleNetv2 [107], a network with reduced parameters, as the backbone for flame feature extraction. Additionally, they integrated an SE attention module [33] into the SGD to enhance the model's ability to concentrate on flame targets within images. Ma et al. [108] proposed a flame and smoke location algorithm based on YOLOv5. They improved the backbone network by introducing the ODConvBS module, consisting of omni-dimensional dynamic convolution (ODConv) [109], batch normalization, and the SiLU activation function. The neck network of YOLOv5 was upgraded with the Gnconv-FPN pyramid structure and the shuffle attention module, effectively combining channel and spatial information. Hong et al. [110] designed a lightweight network for smoke location by replacing standard convolution with depthwise separable convolution and proposing a local cross-channel interaction strategy (ECA module) for straw-burning smoke at a remote distance. The trained model was deployed on the Jetson Xavier NX embedded platform. Bahhar et al. [111] proposed a fire location architecture by combining the fire classification branch network based on CNN and the fire location branch with a YOLOv5 detector. The results from the two branches were fused to give the final decision. Qian et al. [112] proposed a high-accuracy forest fire smoke location model following the structure of YOLOv5. They inserted the omni-dimensional dynamic convolution (ODConv) [109] and self-attention in Transformer into the backbone network. The SimAM attention mechanism, considering both channel and spatial dimensions, was integrated into the neck network to extract smoke features. The decoupling head was used to further improve the accuracy and efficiency of the model.

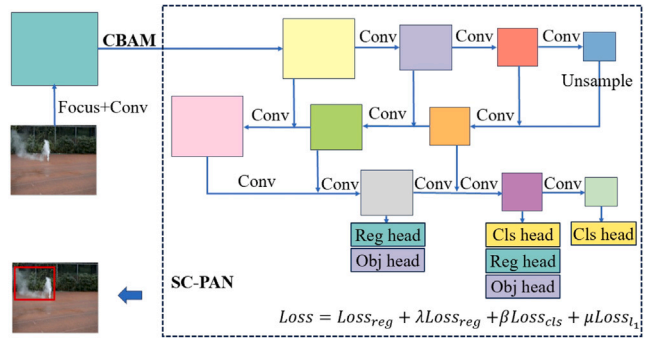


Fig. 9. SASC-YOLOX structure [119] for smoke location.

PPYOLO [113] incorporated ten existing tricks to enhance the accuracy of the detector while maintaining its speed. Building upon the PPYOLO benchmark network, Zhan et al. [114] introduced an adjacent layer composite network (ALCN) for the UAV aerial photography of forest fire smoke location. It comprised of two identical ResNet50-vd units and a SoftPool module [101] to boost the feature extraction capability. A recursive feature pyramid with deconvolution [115] and dilated convolution [116] (RDDFPN) was introduced to integrate shallow visual features with deep semantic information extracted by ALCN. To reduce missed and false detections, they suggested a global optimal non-maximal suppression (GO-NMS) as an alternative to matrix NMS in post-processing.

YOLOX [117] introduced five key changes: an anchor-free architecture, multiple positives, a decoupled head, advanced label assignment, and strong augmentations. Building upon YOLOX-L as the baseline, Zhan et al. [118] proposed a forest smoke location model based on the parallel spatial domain attention mechanism and small-scale transformer feature pyramid network (PDAM-STPNNet). PDAM included a parallel local and global attention mechanism to comprehensively extract texture features of smoke. Additionally, STPN was integrated into the neck network to capture rich global and contextual information. Wang et al. [119] introduced a self-attention and self-cooperation YOLOX (SASC-YOLOX) network for smoke location. As illustrated in Fig. 9, extending the efficient YOLOX structure, three lightweight strategies were implemented: (1) a self-cooperation PAN (SC-PAN) designed as the head of the network to refine features and remove redundancy, (2) the lightweight attention module (CBAM) [93] inserted into the backbone to emphasize meaningful features, and (3) modification of the loss function to improve the feature representation.

YOLOv7 [87] introduced several architectural modifications and a series of bag-of-freebies techniques, resulting in improved accuracy without impacting inference speed. Jin et al. [120] developed an improved YOLOv7 network for real-time smoke location. Depthwise separable convolution was used to instead of the traditional full convolution module for feature extraction. Additionally, they strengthened the multi-level feature fusion network by adjusting weights and incorporating cross-layer connections to improve feature representation. Furthermore, they devised a channel-spatial mechanism to extract contextual information, enabling the model to better understand the spatial

relationships between features and enhance smoke location accuracy. Xu et al. [121] introduced a CNTCB-YOLOv7 network for forest fire location, incorporating ConvNeXtV2 [122] and Conv2Former [123] into the YOLOv7 architecture. They inserted the designed efficient layer aggregation network (ELAN) structure into the backbone network, and leveraged attention mechanisms to improve fire localization performance by guiding the network's focus on critical information pertaining to forest fire targets.

YOLOv8 [124], developed by Ultralytics, the same company responsible for YOLOv5, was launched in January 2023. YOLOv8 is designed to serve as a rapid single-stage object detection algorithm, offering five distinct scale versions: YOLOv8n (nano), YOLOv8s (small), YOLOv8 m (medium), YOLOv8l (large), and YOLOv8x (extra-large). This algorithm is versatile, supporting various tasks including image classification, object detection, and instance segmentation. It is compatible with a wide range of hardware platforms, spanning from central processing units (CPUs) to graphics processing units (GPUs). In terms of detection methodology, YOLOv8 shares similarities with YOLOv5 [26] and YOLOv7 [27], comprising four primary components: input, backbone, neck, and head. Guo et al. [125] proposed an efficient lightweight forest smoke location model based on YOLOv8. Firstly, a simple BiFPN module [84] was integrated into the neck of YOLOv8 to enhance feature fusion capability in forest smoke location, significantly optimizing both parameter count and model computational costs. Subsequently, a simple and parameter-free attention mechanism (SimAM) [126] was introduced to address the complex background of forest smoke images, suppressing feature information from non-smoke targets, thus improving accuracy without introducing additional parameters to the model. Wang et al. [127] improved upon the YOLOv8 framework and presented the MEG-YOLOv8 forest smoke location network. Three novel modules, multiscale feature extraction, global feature enhancement, and edge feature enhancement, were integrated into YOLOv8. The multiscale feature extraction module and edge feature enhancement module collectively enhanced the network's feature representation capabilities, thereby improving smoke detection accuracy and reducing false alarms. Additionally, the global feature enhancement module extracted structural smoke information, further aiding the network in accurately identifying smoke regions and minimizing false alerts. Shamta et al. [128] deployed a real-time forest fire location method using YOLOv8 on an embedded artificial intelligence computer, and demonstrated the performance of YOLOv8 through a comprehensive analysis. Yuan et al. [129] presented a lightweight forest fire location model (FFYOLO) based on YOLOv8. Firstly, a channel prior dilatation attention (CPDA) and a mixed-classification head (MCDH) were proposed to improve the feature extraction performance of fire and smoke. Subsequently, the lightweight GSConv module [130] was applied to the neck part of YOLOv8, which effectively reduces model parameters without sacrificing accuracy. Finally, a knowledge distillation strategy was adopted during training to enhance the model's generalization ability and reduce false detections. Yang et al. [131] proposed an enhanced multi-scale forest fire location method based on YOLOv8. This method adjusted the C2f structure of YOLOv8 by integrating deformable convolutions [132] and spatial channel convolution (SCConv) [133] modules to better adapt to the complexity of forest fires. Additionally, a coordinate attention mechanism was introduced into the location network to more effectively capture feature information and improve model accuracy. Zhang et al. [134] presented Ship-Fire Net, a lightweight ship fire location algorithm built upon YOLOv8n. GhostnetV2-C2F was leveraged to refine the backbone network of the YOLOv8n framework, thereby improving its capability to capture long-range information. Furthermore, spatial and channel reconfiguration convolution (SCConv) [133] was applied to mitigate redundant features, leading to a substantial reduction in both complexity and computational costs. Additionally, omni-dimensional dynamic convolution was integrated into the neck section of the YOLOv8n network, improving accuracy while reducing network parameters.

6.1.2. Fire location based on other OD structures

Chaoxia et al. [135] improved Faster R-CNN [77] by introducing a color-guided anchoring strategy, restricting anchor locations, and employing a global information-guided strategy to mitigate false alarms. Building upon the backbone network of EfficientDet [84], Jiang et al. [136] proposed a novel smoke location network. This model incorporated a self-attention mechanism in Transformer and established a cross-level dataflow of multi-level features. Guan et al. [137] introduced a feature entropy-guided model for fire location. Their approach utilized a cross entropy loss function guided by feature entropy to balance training sample differences in entropy and complexity. Li et al. [138] derived a fast and efficient flame location model from FoveaBox [139] by incorporating the lightweight MobiliNetV3 [30] and an anchor-free structure. The model applied a balanced feature pyramid and an SE-Block [33] for feature extraction and relationship calibration, respectively. Huang et al. [140] proposed an early forest fire smoke localization network based on deformable DETR [141]. The network improved its feature extraction capability for distant forest fire smoke targets by enhancing multi-scale contextual feature extraction and refining the pyramid pooling module. Jin et al. [142] developed a novel lightweight deep learning algorithm for forest fire localization. This algorithm integrated the reparameterization vision Transformer (RepViT) with the traditional CNN to enhance feature extraction. To address the challenge of detecting small fire targets within an image, it employed the simple parameter-free attention module (SimAM) [126], which directly computes multidimensional weights for feature maps. Furthermore, it introduced a bidirectional fusion network that combines top-down and bottom-up approaches. This network incorporated the lightweight ghost shuffle convolution (GSConv) technique [130] to reduce the number of model parameters.

The investigations into smoke localization through single-image analysis predominantly center on improving established general object detection frameworks. The YOLO series of networks are frequently employed as the foundational frameworks for such studies. The prevalent improvements can be succinctly summarized in Fig. 10.

Designing lightweight structures. Depthwise separable convolution, a type of decomposition convolution that utilizes convolution kernels for each input channel, is instrumental in reducing the number of parameters without altering the network structure. This efficiency-enhancing technique has found application in various models, particularly those designed for lightweight structures. Several models, as discussed earlier, adopted depthwise separable convolution to replace standard convolution within existing object detection networks [90,96,98,104,106,120,134,138]. This strategic substitution aimed to enhance the efficiency of fire location models, making them well-suited for deployment on embedded drone devices [89,110,127,128].

Applying attention mechanisms. The attention mechanism serves as a pivotal tool in refining the focus of a network on the fire target while mitigating the influence of non-fire background information in an image. This concerted effort leads to a substantial improvement in the accuracy of early fire location. Two prevalent types of attention mechanisms, namely channel attention and spatial attention, play crucial roles in this process. The channel attention mechanism dynamically assigns weights based on the significance of each channel, allowing selective emphasis on the contributions of different feature maps. Simultaneously, spatial attention aims to bolster the representation of essential regions within an image while dampening the impact of irrelevant background areas. In architectures dedicated to fire location based on single images, it was common practice to fuse channel attention and spatial attention. This fusion ensured a comprehensive evaluation of the interesting regions from both spatial and channel perspectives [95,99,104,105,108,112,119,121]. Traditional channel and spatial attention mechanisms can only refine features in one dimension, either the channel or spatial dimension. In contrast to these traditional mechanisms, SimAM is a straightforward and efficient lightweight attention module. It can directly generate 3D weights for feature maps, enabling feature

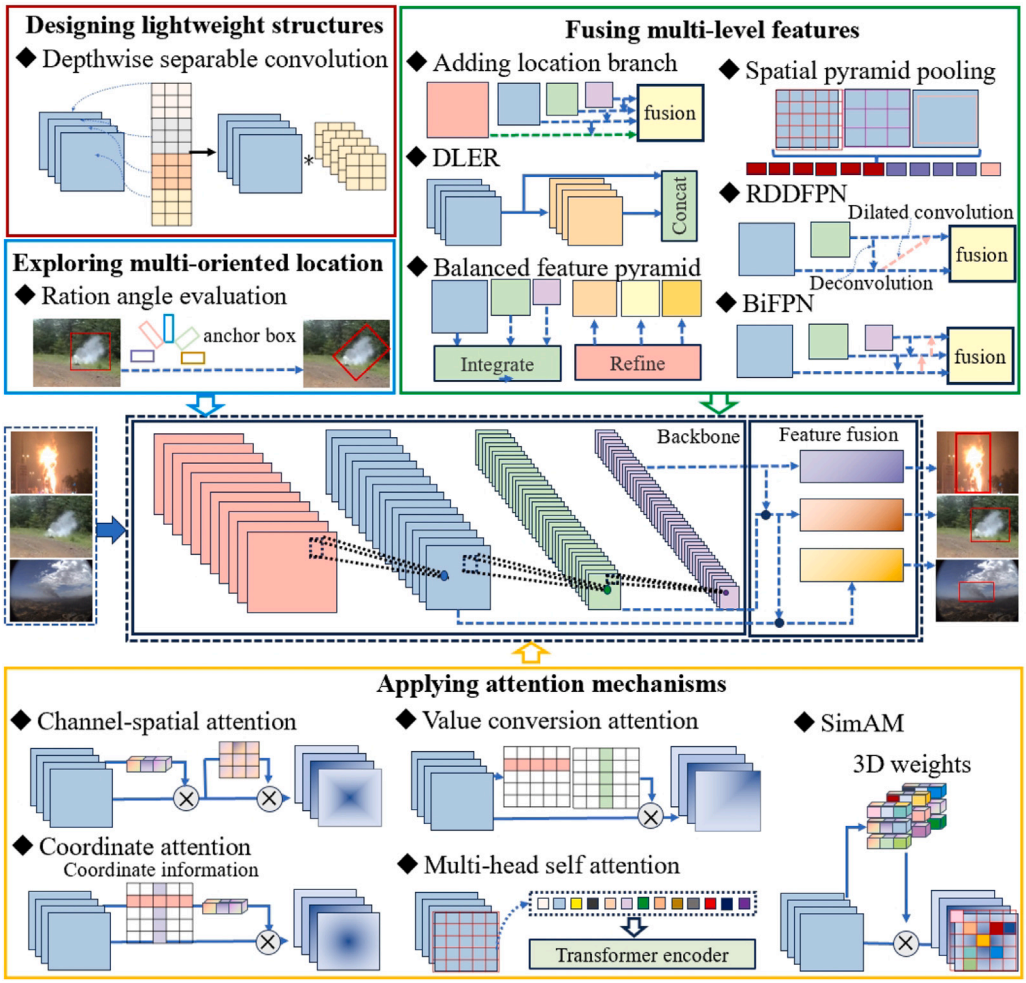


Fig. 10. Overviews of image based fire location methods using DL.

calibration across multiple dimensions without the requirements for extra parameters. As a result, SimAM has been recently integrated into the exist object detection frameworks to enhance the network's capability in extracting smoke features, consequently improving the model's detection accuracy [125,142]. To address irregular fire patterns and complex environments, the coordinate attention network (CAN) incorporated position information into the channel attention, placing emphasis on the spatial structure of the fire target [96,134]. Additionally, the value conversion-attention mechanism module (VAM) enhanced the extraction performance of the texture feature extraction network [100]. VAM achieved this by leveraging color and texture feature information from smoke images and refining the weight distribution of texture information through a joint weight assignment strategy in both horizontal and vertical directions. Furthermore, the integration of self-attention mechanisms within the Transformer architecture [51] significantly contributed to the fusion of global and local features [112,118,120,136]. This fusion improved the feature representation capability specific to fire targets, ultimately elevating the accuracy of early fire location. The collaborative effect of these attention mechanisms underscores their pivotal role in refining the network's ability to discern and locate fires amidst complex visual environments.

Fusing multi-level features. In the realm of object detection models, the integration of a feature pyramid network (FPN) has been a common strategy to fuse feature maps of varying scales, addressing the challenge of detecting targets with multiple features and different scales in real scenes. However, recent advancements have introduced novel modules and approaches to further enhance the capabilities of these models. A dedicated scale location branch was added to existing object detection architectures to specifically recognize small fire instances. This augmentation improved the model's ability to handle fire targets of varying sizes, especially in complex scenes [91]. The DLFR module focused on reusing information from original smoke features. By expanding feature diversity and allowing the neural network to utilize smoke features of different depths and attributes, the DLFR module contributed to improve location performance [92]. BiFPN introduced a top-down pathway, bottom-up pathway, and cross-scale connection to enhance the efficiency of multi-scale feature fusion. This modification not only improved location accuracy but also introduced a cross-level dataflow to ensure that sufficient contextual information and details are available during feature extraction [96,125,136]. The spatial pyramid pooling was always utilized to facilitate the fusion of local multi-scale features, enhancing the feature extraction ability for small fire targets. This improvement ensured that the model can capture relevant contextual information [98]. The DPPM, an expansion of the pyramid pooling module, was introduced to extract more abstract information from concatenated multi-scale feature maps [118]. The Gnconv operation, introduced into FPN, achieved large-range and high-order spatial interactions [108]. This addition enhanced the feature pyramid structure, making the model more adaptable to rich feature extraction for small object detection in aerial scenes. RDDFPN further improved the feature fusion capability of the feature pyramid structure, making the model suitable for both rich feature extraction and small object detection in aerial scenes [114]. Unlike traditional FPN methods that used side connections, the balanced feature pyramid focused on strengthening

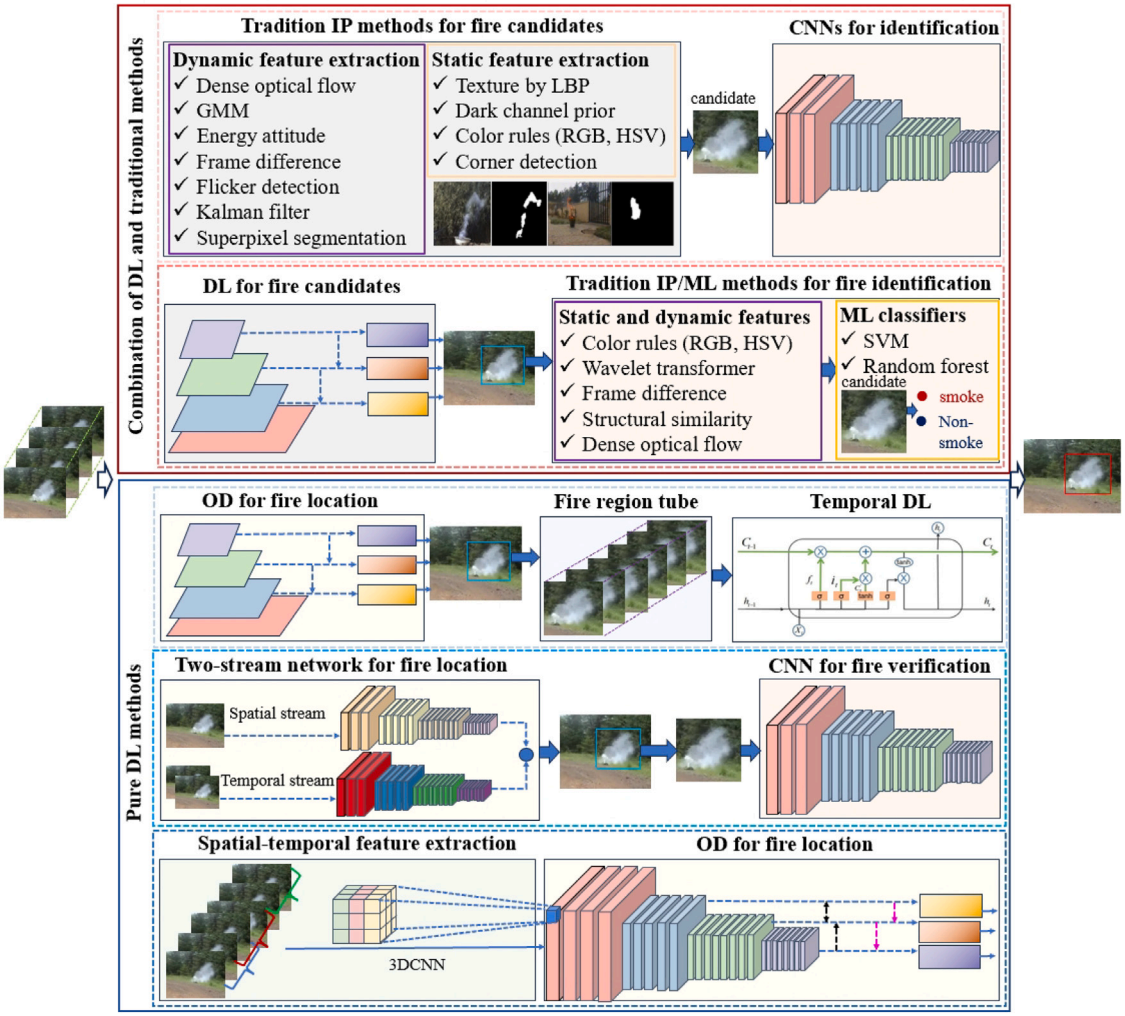


Fig. 11. Overview of video based fire location methods using DL.

multi-level features of flame images using deeply integrated balanced semantic features [138].

Exploring multi-oriented location. To identify the source of fires and estimate fire development trends, exploring multi-oriented location becomes imperative in the realm of fire detection. By introducing angle parameters and designing multi-oriented anchor boxes during the training process of the smoke localization network, smoke localization results with rotation angles can be achieved [100]. This can provide essential information for fire rescue efforts and is encouraged for further research.

6.2. Video based fire location

Current image based fire location methods frequently encounter challenges related to the absence of dynamic information, while video based methods offer a promising avenue to address this limitation. They typically adhere to a two-stage process comprising fire candidate regions extraction and subsequent fire verification. The first stage narrows down the scope and pinpoints the suspected fire regions, reducing the computational load and focusing the subsequent analysis on areas with a higher likelihood of containing fire information. The second stage identifies the extracted fire candidates. By doing so, the algorithms optimally extracts features from the narrowed-down areas, enhancing the precision and efficiency of fire location. Based on the different techniques employed in each stage, video based fire location methods can be categorized into two types: the combination of DL and traditional methods, and pure DL methods. In the former, traditional IP and ML techniques are applied to one stage, while the other stage is addressed using DL techniques. Conversely, pure DL methods exclusively employ DL techniques in both stages. The overview of fire location techniques based on videos is described in Fig. 11.

6.2.1. Combination of DL and traditional methods

Traditional IP methods are typically employed to extract static and dynamic features of fires, generating fire candidate regions. CNNs are then utilized to identify these candidates. Shi et al. [143] introduced an efficient video smoke location network tailored for embedded applications. This method combined texture, shape, and motion information extracted by local binary patterns (LBP) [144] and a dense optical flow estimator [145] (LBP-motion) to generate fire regions. A lightweight MobileNetV2 framework [146] was utilized to locate smoke within the regions identified by LBP-motion. Li et al. [147] employed the dark channel information prior [61] and Gaussian mixture model (GMM) [148] to extract smoke candidates. Subsequently, a novel lightweight score clustering CNN (SCCNN) was proposed to identify these candidates efficiently. Kabbi et al. [149] developed a preprocessing step incorporating smoke motion, color, and energy attitude. Extracted features were fed into a deep brief network (DBN) [150] for smoke classification. Zhang et al. [151] proposed real-time flame location framework DeepFireNet, utilizing static features based on color rules and dynamic features based on frame difference. Convolutional networks were employed to detect suspected fire areas. Xie et al. [152]

generated the suspected fire regions by dynamic features extracted by background subtraction and flicker evaluation, and then performed fire identification by a new lightweight CNN method. Alessio et al. [153] applied Kalman filtering motion detection and color information to select the automatic candidate regions, which then were verified by a CNN. Jia et al. [154] utilized GMM [148] to remove static background and segment suspected smoke regions. A trained deep neural network, through transfer learning, recognized the suspected regions. Wu et al. [155] presented a two-stage smoke location structure based on dense optical flow [145] and ResNet [27]. Cheng et al. [156] introduced an efficient video flame location fusion algorithm, combining low-level visual operations with multi-scale CNN and spatial pyramid pooling. Ryu et al. [157] applied HSV color conversion and Harris corner detection [158] for region of interest extraction, and determined flame presence using an Inception-V3 network [159]. Nguyen et al. [160] introduced a multistage fire detection method that integrates CNNs and a long short-term memory (LSTM) network [161]. The approach began by detecting potential fire candidates through the analysis of salient features such as color, flickering frequency, and brightness. Following this initial stage, the visual features of each identified candidate were extracted from sequence images using a pretrained CNN model. These visual features were then input into the LSTM network, facilitating more nuanced flame prediction. Wang et al. [162] proposed a novel framework for wildfire location using a superpixel segmentation method for extraction of fire candidates and ESE-ShuffleNet for real-time identification. Kwak et al. [163] utilized dark channel prior [61] and Lucas–Kanade optical flow technique [145] for smoke location, followed by Inception-V3 network [159] for accurate determination. Chen et al. [164] introduced an improved GMM [148] and YOLOv5 [86] smoke location algorithm.

In contrast to the aforementioned research process, some researchers opt for DL techniques to extract fire candidate regions, while traditional IP and ML techniques are employed for fire verification. Liu et al. [165] introduced a smoke location method involving the division of an image into sub-blocks. Within each sub-block, the researchers applied a DNCNN [16] for the initial extraction of suspected smoke regions. Subsequently, the visual change of these suspected smoke regions was computed, and a support vector machine (SVM) was employed to enhance the precision of smoke location and mitigate false alarms. This two-step process, combining DNCNN and SVM, contributed to a more robust smoke location framework, capable of discerning real smoke regions while minimizing false positives. Lee et al. [166] presented an innovative method for locating fire using spatial and temporal information within a deep learning framework. The Faster RCNN [78] was adapted to extract initial candidate regions for smoke and flame. Then spatial and temporal features were employed to refine and verify these candidates. Park et al. [167] developed an innovative method for night-time flame location. An ELASTIC block [168] was integrated with the existing YOLOv3 network [80] to enhance the location accuracy. To capture the dynamic characteristics of night-time flames, a temporal fire-tube was formed by accumulating continuous frames. Additionally, a histogram of the optical flow emanating from the flame was computed and converted into a bag-of-features (BoF) histogram. Subsequently, a random forest classifier [169] was employed to verify potential fire candidates.

6.2.2. Pure DL methods

Different from the aforementioned video based fire location techniques, some researchers completely abandon traditional IP and ML methods, opting for pure DL techniques to analyze the static features and dynamic behaviors of fires in video data. Among them, object detection networks, temporal deep networks, two-stream convolutional neural networks, and 3D convolutional neural networks have gained favor from researchers.

Object detection methods are initially employed to generate fire candidate regions from input images. Subsequently, a temporal fire-tube is created by accumulating continuous frames, and a temporal network is employed for fire identification. Jeong et al. [170] proposed an approach combining a YOLO detector [80] and a long short-term memory (LSTM) [161] network to improve wildfire smoke location performance. Keyframes reflecting the slow motion of smoke were first detected, and YOLOV3 [80] was employed to search for smoke candidate regions. Shallow LSTM, guided by a teach-student framework, improved the representation of temporal characteristics of wildfire smoke. Kim et al. [171] utilized Faster R-CNN [78] to extract fire candidate regions, incorporated LSTM [161] to accumulate local features within bounding boxes for short-term decisions, and employed majority voting to ensure reliable long-term decisions. Dewangan et al. [23] presented a spatial–temporal gridded image classification approach. It employed a ResNet34 network [27] to extract pixel representations from each tile independently, combining temporal information with LSTM, and feeding temporally-combined tiles into a vision Transformer [51] to enhance smoke prediction.

Unlike the fire candidate region generation based on object detection techniques, dual-stream convolutional neural networks with temporal and spatial branches simultaneously introduce spatial and temporal information for extraction of fire candidates. Therefore, in the fire verification stage, only the spatial information of the candidate regions needs to be considered. Shahid et al. [172] proposed a two-stage cascaded architecture to detect flames exhibiting irregularities in color, size, form, texture, and flickering frequency. In the initial stage, a spatial–temporal network located flame candidates by combining shape and motion flicker. The second stage applied a ResNet [27] classification model to distinguish whether the identified region is a flame. Cao et al. [173] presented the quasi video smoke detector (QuasiVSD), an efficient solution for capturing motion representations of smoke from only two frames. The model utilized a dual-frame motion-aware backbone (DMB) with appearance and motion pathways to extract smoke candidates. Additionally, a weakly guided attention module was proposed to refine the feature representation for the extracted candidates. Tanveer et al. [174] put forth a two-stage architecture for locating early fire in videos. In the initial stage, fire region proposals were generated by a spatial–temporal network employing a self-attention module. The subsequent stage involved classifying each region proposal as either a flame or non-flame, using a pretrained state-of-the-art classifier. Qiang et al. [175] proposed a novel two-stream forest fire smoke location method. The static pathway extracted local features using a VGG16 network [25], while the dynamic pathway highlighted smoke movement through a Bi-long short-term memory (BLSTM) network [176].

3D convolutional networks (3DCNN) are commonly used in video processing tasks as they can extract robust spatial and temporal features simultaneously. Huo et al. [177] proposed an end-to-end video smoke localization network (3DVSD) based on 3DCNN. As illustrated in Fig. 12, this network replaced 2DCNN in the backbone of existing object detection frameworks with 3DCNN, allowing for the simultaneous utilization of spatial and temporal features of smoke in video sequences for accurate smoke recognition and localization. 2D deep separable convolutions were applied into the neck and head sections to reduce the number of parameters.

7. Fire segmentation

Unlike fire location, which focuses on identifying the presence and location of fire objects, fire segmentation serves as both a successor and extension of fire location, requiring the accurate separation of fire components from the intricate background scene at the pixel level. This task involves generating a detailed mask with precise edges, encompassing aspects of object classification, localization, and boundary delineation.

With the introduction of fully convolutional networks (FCNs) [178], encoder–decoder style deep network architectures have been designed for image semantic segmentation tasks [179]. FCN replaces the fully connected layers at the end of traditional CNNs with convolutional

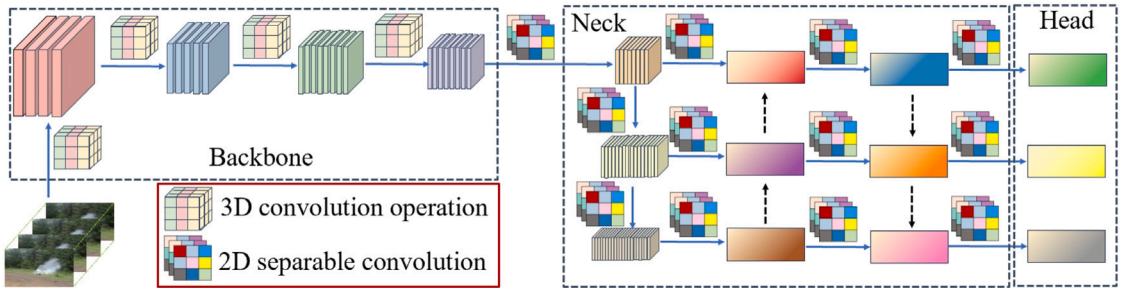


Fig. 12. 3DVSD framework [177] for video based smoke location.

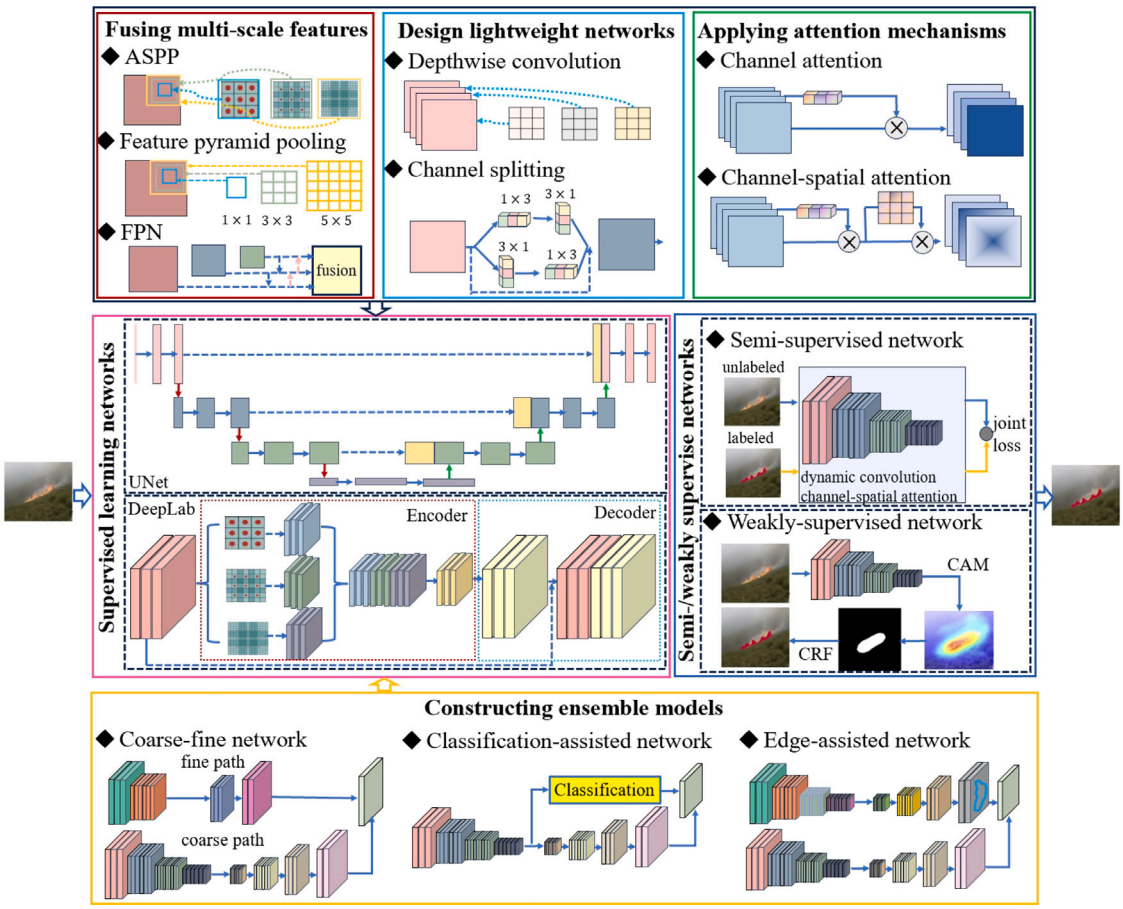


Fig. 13. Overview of image based fire segmentation methods using DL.

layers for image feature extraction, and then utilizes deconvolutional layers for feature map upsampling, achieving end-to-end semantic segmentation of objects in images.

The Deeplab networks [180–182] utilize dilated convolutions to expand the receptive field, thereby capturing contextual information across a wider spatial range and enhancing the accuracy of semantic segmentation. Recent DeepLabv3+ [183] uses the DeepLabv3 [182] as the encoder, with modifications to the Xception backbone [29] using depthwise and pointwise convolutions.

The UNet architecture [184] begins by employing a network similar to FCN architecture to extract contextual information from input images, and then utilizes deconvolution operations to upsample the feature maps, increasing the spatial resolution of the output. Finally, it incorporates skip connections to effectively fuse low-level features from early layers with high-level features from later layers. These skip connections facilitate the seamless integration of detailed spatial information with high-level semantic context, enabling the network to generate accurate and detailed segmentation masks.

Based on the aforementioned object segmentation frameworks, significant progress has been made in fire segmentation fields. Below, we will review image based fire segmentation and video based fire segmentation separately.

7.1. Image based fire segmentation

Fire segmentation techniques based on a single image are summarized in Fig. 13.

Selecting the best baseline architecture. Wang et al. [185] conducted a comparative study of semantic segmentation methods applied to forest fire segmentation tasks. The authors created a dataset consisting of remote-sensing forest fire images captured by drones under various conditions. Theoretical and experimental analysis were

performed on four classic semantic segmentation architectures and two backbone networks. The results indicated that U-Net [184] and DeepLabV3+ [182] achieves better experimental results on the mentioned dataset. Harkat et al. [186] performed a detailed analysis of the impact of different loss functions and encoder architectures on various types of aerial images for the fire segmentation task, using the DeepLabv3+ network [183]. The model trained with a ResNet-50 backbone [27] and Dice loss [187] demonstrated satisfactory performance.

Fusing multi-scale features. In fire segmentation tasks, the dilated convolution operation [180] is often employed to capture multi-scale information of fire targets. Khan et al. [188] conducted smoke segmentation in images using the DeepLabv3+ architectureDeepLabv3+ [183]. In their approach, the encoder employed a atrous spatial pyramid pooling (ASPP) module to extract multi-scale smoke features by using different dilation rates in atrous convolutions. Following the encoder stage, the decoder increased the spatial resolution of the feature maps by bilinear upsampling, and fused low-level features and high-level semantic information to refine the segmentation results and improve the model's overall performance in smoke segmentation tasks. Yuan et al. [189] proposed a smoke segmentation method based on the multi-scale context contrastive local (MCCL) model. This approach utilized multi-scale features obtained through atrous convolutions to capture valuable contextual information. Hou et al. [190] introduced an efficient improved DeepLabv3+ [183] for flame and smoke segmentation. The encoder network utilized a DCNN [56] backbone with atrous convolutions to extract multi-scale features, while the decoder network integrated low-level features from the backbone output and high-level features processed by ASPP+ to enhance the accuracy of smoke boundary segmentation. Wen et al. [191] proposed a dense multi-scale context and asymmetric pooling embedding network (DMAENet) to segment smoke. The dense multi-scale context module (DMCM) captured information from sub-regions at different scales, enhancing feature representation capacity with asymmetric convolution. The asymmetric pooling enhancement module (APEM) utilized asymmetric pooling to improve feature extraction for irregular objects. The feature fusion module (FFM) deeply merged low-level spatial information with high-level semantic information, improving feature robustness in an attentional manner. Yuan et al. [192] proposed a multi-scale residual module (MRM) for accurate smoke segmentation, which significantly increased the width and depth of the residual path. Deconvolution and bilinear interpolation were combined to refine the multi-scale feature representation of smoke. Sun et al. [193] incorporated the path aggregation feature pyramid network (PAFPN) into the feature fusion component of the SOLOv2 framework [194]. By integrating bottom-up connections, PAFPN effectively shortened the information propagation path, thereby maximizing the utilization of low-level features to enhance smoke localization.

Design lightweight networks. Li et al. [195] introduced an improved version of DeepLabv3+ [183] for real-time fire segmentation. The lightweight MobileNetV3 [30] was introduced in the encoder to improve segmentation speed. Two different shallow features were incorporated into the decoder network to ensure the network contained rich fire feature information. Building upon UNet++ [196], Wang et al. [197] opted for a lightweight network, MobileNetV2 [146], as the convolution unit to reduce the number of parameters required for training. They designed a smoke segmentation network model specifically for optical satellite video. Hu et al. [198] integrated depthwise separable convolution with atrous convolution, thereby reducing computational complexity while preserving robust representational capabilities for smoke segmentation in remote images. This operation enabled the module to effectively extract multiscale features. Muhammad et al. [199] introduced an efficient early fire detection and segmentation network tailored for Internet of Things (IoT) systems. They utilized depthwise separable convolution, pointwise group convolution, and channel shuffling strategies to effectively reduce model parameters and computational costs. Yuan et al. [200] aimed at segmenting smoke on computation-limited devices and proposed a lightweight real-time smoke segmentation network. They utilized the channel splitting strategy to reduce network parameters and memory consumption for faster inference speed.

Applying attention mechanisms. Channel attention and spatial attention mechanisms are commonly integrated into fire segmentation network architectures to enhance the ability to extract salient features of fire targets from input images. Zheng et al. [201] introduced a channel-spatial attention mechanism into UNet structure [184] for smoke segmentation. In addition, they applied Transformer to capture the global information of smoke and improved the segmentation accuracy of thin smoke. Wang et al. [202] proposed an improved UNet architecture [184] utilizing attention mechanisms and residual blocks for smoke segmentation in satellite images. The residual block with a skip connection structure fused coarse semantic and local appearance information, and the SEBlock module [33] was applied to emphasize smoke information and reduce background interference. Wang et al. [197] introduced the CBAM module [93] into the UNet++ network [196] to enhance smoke target features and suppress irrelevant ground objects and noise information. Yuan et al. [189] proposed a convolutional GRU channel attention for a dual classification assisted smoke segmentation network to mine inherent semantic relationships of features and obtain long-range, context-dependent information feature extraction. Zhang et al. [203] introduced self-attention, edge attention, and channel attention mechanisms into a smoke segmentation network to improve the network's ability to learn distinguishable features from the spatial and channel perspectives. Yuan et al. [200] proposed a lightweight network to segment network, incorporating channel split and shuffle attention modules (CSSAM) in the encoder and channel attention module (CAM)in the decoder. Sun et al. [193] introduced CBAM attention [93] into the Resnet-50 [27] backbone feature extraction network for flame and smoke segmentation. Wang et al. [204] focused on fire targets in complex forest environments by utilizing channel and spatial attention mechanisms, thereby improving the segmentation accuracy of forest fires. Hu et al. [198] utilized SCSE [205] to recalibrate channel-wise features and dynamically adjust the importance of spatial regions within an image. This assisted the model in focusing on critical smoke-related information. Wang et al. [206] presented an adaptive multi-scale attention module to enhance smoke segmentation precision. This module was incorporated at the head of each branch of the atrous spatial pyramid pooling (ASPP) to enhance the model's ability to handle multi-scale information.

Constructing ensemble models. Yuan et al. [19] proposed an end-to-end dual-path smoke segmentation network. The coarse path constructed an encoder–decoder architecture with skip connections to extract global contextual information of fire smoke, while the fine path built a shallow encoder–decoder framework to capture detailed smoke information. Yuan et al. [189] introduced a dual classification assisted gated recurrent neural network for smoke segmentation. One classification branch was responsible for extracting global information to implicitly differentiate between smoke and non-smoke targets, while another classification branch directly outputted the classification result, explicitly eliminating misclassified targets. Zhang et al. [203] designed a dual-path network for both smoke segmentation and smoke edge detection. The smoke semantic edge features served as effective supplementary information to the semantic segmentation features, enhancing the accuracy of segmentation results. Li et al. [207] proposed a dual-path energy-efficient smoke segmentation network. The spatial path extended the receptive field and encoded rich multi-scale information through pyramid pooling modules, while the contextual path introduced channel attention modules, leveraging lightweight models and global average pooling to reduce complexity while ensuring effective model performance. Li et al. [208] presented a multi-path edge information interaction network, refining flame segmentation results by explicitly extracting edge information.

Table 4
Performance comparison of smoke segmentation networks on the public SSD dataset.

Networks	DS01		DS02		DS03		#Parameters (M)
	mIoU	mMSE	mIoU	mMSE	mIoU	mMSE	
FCN-8s [178]	64.03%	0.3221	63.28%	0.3353	64.38%	0.3202	–
SegNet [212]	56.94%	0.3976	56.77%	0.4173	57.18%	0.4047	24.9
SMD [213]	62.88%	0.3209	61.50%	0.3379	62.09%	0.3255	–
TBFCN [214]	66.67%	0.3021	65.85%	0.3196	66.20%	0.3070	–
Deeplab v1[180]	68.41%	0.2981	68.97%	0.3030	68.71%	0.3010	–
LRN [214]	66.43%	0.2745	67.71%	0.3078	67.46%	0.3041	–
HG-Net2 [215]	64.27%	0.3186	63.06%	0.3380	64.18%	0.3273	–
HG-Net8 [215]	62.10%	0.3187	61.89%	0.3301	62.45%	0.3215	–
LKM [216]	75.82%	0.2658	74.93%	0.2799	75.39%	0.2748	–
RefineNet [217]	77.16%	0.2486	76.75%	0.2590	77.52%	0.2525	–
PSPNet [218]	78.71%	0.2366	78.01%	0.2480	78.39%	0.2430	42.5
CCL [219]	78.87%	0.2349	77.95%	0.2498	78.55%	0.2429	–
DFN [220]	80.87%	0.2269	79.90%	0.2411	80.60%	0.2332	–
W-Net [221]	73.06%	0.2688	73.97%	0.2548	73.36%	0.2596	31.1
DANet [222]	82.91%	0.2110	81.68%	0.2276	82.48%	0.2188	47.6
ERFNet [223]	69.9%	–	67.9%	–	68.7%	–	2.06
LEDNet [224]	69.0%	–	67.8%	–	68.5%	–	0.91
DFANet [225]	63.2%	–	59.4%	–	61.8%	–	2.18
CGNet [226]	68.6%	–	65.5%	–	67.2%	–	**0.49**
Unet [184]	72.7%	–	69.8%	–	70.7%	–	–
Deeplab v2 [181]	77.7%	–	76.5%	–	77.9%	–	23.8
DMNet [227]	78.6%	–	76.9%	–	77.8%	–	56.2
DSS [19]	71.04%	0.2745	70.01%	0.2894	69.81%	0.2861	29.9
LWNet [200]	74.2%	–	72.5%	–	72.8%	–	0.88
CCENet [228]	74.67%	–	74.57%	–	75.17%	–	–
CGRNet [189]	82.61%	0.2138	81.67%	0.2280	82.18%	0.2212	–
EANet [203]	**83.67%**	**0.2045**	**82.45%**	**0.2220**	**83.10%**	**0.2145**	–
DMAENet [191]	79.4%	–	78.8%	–	79.2%	–	32.0
MRM [192]	79.03%	0.2432	78.30%	0.2553	78.65%	0.2505	–

Exploring semi-/weakly-supervised learning. Obtaining a significant volume of annotated fire data is frequently a time-consuming and labor-intensive process, and the efficacy of training DL models heavily relies on the quality of these annotations. To address these challenges, semi-supervised learning and weakly-supervised learning have emerged as effective solutions, requiring only a limited amount of labeled data. Sun et al. [193] proposed a semi-supervised method for flame and smoke segmentation based on the improved SOLOv2 algorithm [194]. Deformable convolution replaced conventional convolutional kernels, CBAM attention [93] was integrated into the backbone network, and a bidirectional multiscale fusion APFPN substituted the FPN. Introducing color and morphological features of flames and smoke facilitated the determination and filtering of pseudo-labels generated in semi-supervised learning, reducing errors resulting from incorrect labels. Wang et al. [204] introduced a semi-supervised approach for segmenting forest fires in UAV-acquired images, requiring only a small amount of pixel-level annotated images for training. They used dynamic convolutions instead of regular convolutions to enhance the feature representation capability of various fire patterns. Attention mechanisms were also applied to emphasize the salient features of forest fire targets and mitigate the influence of complex backgrounds. In the context of weakly supervised learning, Marto et al. [209] proposed an active learning (AL) methodology to segment flame and smoke pixels in images. The method involved an AL classification stage, followed by a segmentation stage with class activation mapping (CAM) [210] on learned classifiers, thereby reducing labeling effort through the integration of weakly supervised learning with active learning. Amara et al. [211] developed a weakly-supervised method to segment flame and smoke in wildfire scenarios. A VGG network [25] detected the presence of flame/smoke in the entire image, using the class activation map method [210] to create a probabilistic heatmap indicating regions where flame or smoke is likely present. Subsequently, binarization and the conditional random field (CRF) method [181] were applied to create a segmentation mask at the pixel level.

Others. Inspired by the UNet structure [184], Guan et al. [137] reconstructed the MaskIoU branch of mask scoring R-CNN (MS R-CNN) [229] using a U-shaped network. This adaptation aimed to accurately distinguish individual fire entities and pixel information, thereby reducing segmentation errors. Wen et al. [230] employed a modified UNet neural network architecture to segment smoke plumes in geostationary satellite imagery. The adapted network maintained the same number of layers in the encoding and upsampling blocks but reduced the number of convolutional filters. Additionally, the ReLU activation function was replaced with PreLU activations. Frizzi et al. [231] designed a CNN-based encoder–decoder network for segmenting smoke and flames. They replaced convolutional layers in VGG architecture [25] with fully connected layers, and captured contextual information of fire targets by merging features from both the encoding and decoding stages. Perrolas et al. [232] proposed a computationally efficient method based on the quad-tree search algorithm [233] for flame and smoke segmentation in high-resolution images. The classification network determined whether the image patch contained instances of fire or smoke, and the segmentation network computed the initial segmentation of fire if fire or smoke is detected. Martins et al. [234] presented an instance segmentation network to capture the shape, color, and spectral features of smoke. Subsequently, ML techniques were utilized for further identification of smoke objects.

Performance evaluation. The synthetic smoke dataset (SSD) [19] is always used to compare smoke segmentation networks with several outstanding semantic segmentation methods based on DL, including FCN-8s [178], SegNet [212], SMD [213], TBFCN [235], LRN [214], Deeplab v1 [180], LRN [214], HG-Net [215], LKM [216], RefineNet [217], PSPNet [218], CCL [219], DFN [220], W-Net [221], DANet [222], ERFNet [223], LEDNet [224], DFANet [225], CGNet [226], Unet [184], Deeplab v2 [181], DMNet [227]. We directly borrow the values form the corresponding papers. Table 4 reports quantitative results on the three synthetic test datasets. Among the compared methods, EANet [203] consistently outperforms others in terms of mIoU and mMSE metrics across all three test datasets. The visualized segmentation results of real fire images are shown in Fig. 14.

7.2. Video based fire segmentation

Video based fire segmentation integrates spatial and temporal information within a video sequence to enhance the network's adaptability

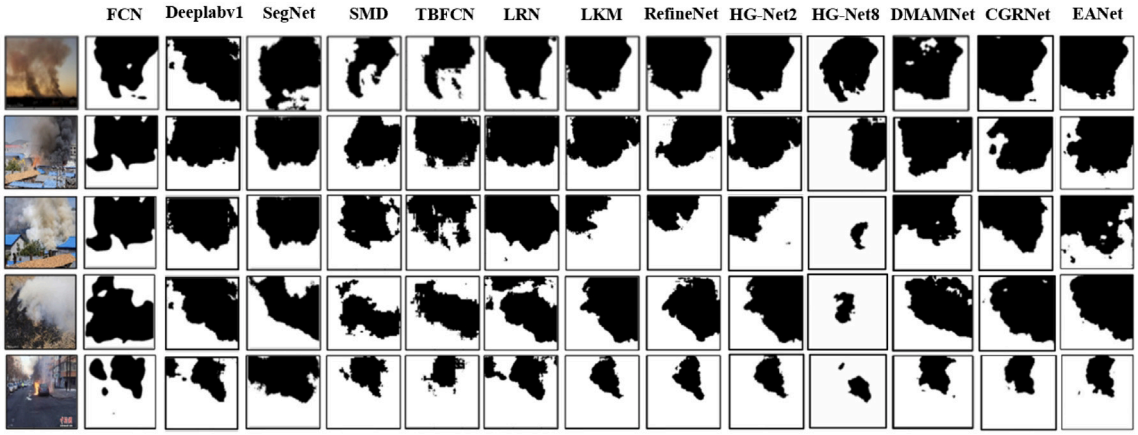


Fig. 14. Smoke segmentation results of real fire images.

to complex natural environments and slow-moving fire targets. Tanveer et al. [174] deigned a semi-supervised network to segment fire region from the background in a video based on a given keyframe. A spatial–temporal network that adopts self-attention was used to combine spatial and temporal features for extracting high-quality segmentation maps. Li et al. [236] introduced a 3D parallel fully convolutional network to enhance the intelligence of smoke segmentation. This model facilitated rapid and accurate segmentation by mitigating interference from the complicated backgrounds. However, it was susceptible to false negatives in cases where smoke moves slowly. Gao et al. [237] proposed a new video smoke segmentation algorithm based on deep saliency networks. This method extracted informative pixel-level and object-level smoke saliency features, and combined them with deep features to segment smoke regions in images. Jia et al. [238] introduced a video based smoke segmentation network that combines a conditional generative adversarial network with Unet [184], showing effectiveness in smoke texture analysis. Ding et al. [239] proposed a spatial–temporal dual-stream smoke segmentation network (STCNNsmoke). The spatial stream sub-network extracted the spatial features of smoke, while the temporal stream sub-network represented the dynamic features of smoke using dense optical flow. Zhu et al. [240] presented an encoder–decoder 3DCNN network for wildfire smoke segmentation. The model utilized 3D residual blocks in the encoder to extract spatial–temporal features, aiding in more accurate segmentation and tracking of dynamic changes in smoke.

8. Challenges and further directions

Due to its superior performance and significant potential compared to traditional methods, deep learning (DL) has become an important research topic in the field of visual fire detection. However, deep learning still faces the following challenges in fire detection.

8.1. Open challenges

Scarcity of public fire datasets. The primary challenge hindering the development of fire detection methods is the lack of public fire datasets that can be used as benchmarks for training models and evaluating performance. Although some researchers have utilized their own created datasets, only a few have chosen to make these datasets publicly available. The scarcity of publicly accessible datasets makes it challenging for researchers to compare their methods with others, therefore impeding the fast development of fire detection techniques.

Distant fire detection. Despite DL-based fire detection techniques have shown enhanced performance, effectively localizing small fire objects remains a challenging task. The visual complexity of distant fire objects, combined with the distance between the camera and the fire source, poses difficulties in their identification. The apparent size of the fire region appears small, and its movement appears slower than its actual speed, complicating accurate detection. Additionally, the translucent nature of distant fire smoke and the complexity of the background further contribute to the challenge of fire detection. Existing research highlights the increased difficulty of locating fires during the initial phase, particularly when images or videos are captured from a distance.

Intraclass variation. Intraclass variation refers to the diversity observed among different images belonging to the same class. The behavior of smoke and flames in uncontrolled fires is inherently irregular, exhibiting variations in shape, size, color, and texture. Such variability can be attributed to multiple factors, including the materials involved, prevailing weather conditions, and illumination variations. Additionally, cameras, not being color or spectral measurement devices, employ distinct sensors and color/illumination balancing methods. As a result, they may generate varied images and videos for the same scene due to their internal settings. Addressing these challenges requires innovative approaches to enhance the robustness and reliability of fire detection techniques.

Conflict between accuracy and complexity. In the realm of published papers on DL based fire detection, a majority (77%) rely solely on single-image based methods, with only a minority (23%) harnessing video data to exploit the dynamic characteristics of fire for detection. Relying solely on static features from a single image can make fire detection vulnerable to interference from other similar objects, resulting in frequent false alarms during practical applications and adversely impacting user experience. The continuous frames present in video data facilitate the capture of dynamic fire behaviors, thereby enhancing fire detection accuracy. However, typical operations employed for processing video data, such as optical flow extraction, 3D convolutional neural networks, and multi-stream convolutional neural network architectures, often impose significant computational burdens and increase hardware costs.

8.2. Further directions

Despite rapid evolution of fire detection technologies, there are still many problems where research needs to be done. This section discusses several study directions, which are compiled in Fig. 15.

Diverse fire datasets. Fire detection is a broad research area with valuable applications in safety. However, the scarcity of open-source datasets for fire detection research poses a significant challenge. Moreover, the existing datasets are often outdated, with suboptimal image/video resolutions. Consequently, there is an urgent need to generate datasets of higher quality and resolution, addressing the collective interest of the research community. In the computer vision community,

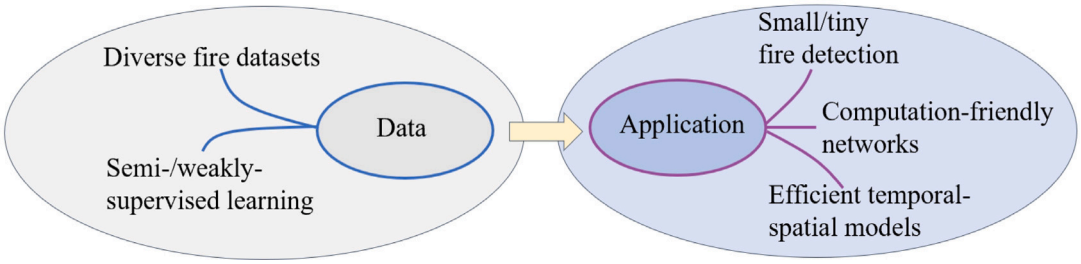


Fig. 15. Further directions of visual fire detection.

diffusion models [241] have been extensively employed for generating realistic synthetic data across various fields. Additionally, deep domain adaptation [242] excels in acquiring transferable representations across both a limited data source domain and the target domain of practical application. This process involves the construction of practical fire data, thereby enhancing the generalization capabilities of fire detection models. Leveraging these models for the creation of fire datasets offers the flexibility to introduce variations in fire patterns and environmental conditions. This approach holds promise in enhancing the quality of datasets, therefore empowering the research community to advance fire detection capabilities.

Semi-/weakly-supervised learning. In fire detection research, the amount of the annotated training samples is always limited. In addition, the number of fire samples is usually much smaller than that of non-fire samples for binary classes. These limitations pose a challenge, particularly for fully supervised learning approaches, where a large amount of accurately labeled data is required to train models effectively. The semi-/weakly-supervised learning framework [243,244] presents a promising alternative to the challenge of limited annotated training samples in fire detection research. Semi-/weakly-supervised learning methods can infer valuable insights and patterns from the available annotations, effectively learning to detect fires despite the limited training samples. They not only alleviates the need for extensive manual annotation but also offers flexibility in dealing with diverse and dynamic fire detection scenarios.

Small/tiny fire detection. As indicated by reported results, deeper CNNs may fail to effectively represent features of small objects. In deep CNN architectures, pooling and subsampling operations generate feature maps with gradually reduced resolutions [245,246]. Consequently, small/tiny fire objects lose spatial information as the network deepens, making accurate detection challenging. Additionally, the presence of small/tiny fires in images leads to highly imbalanced data distribution between fire and background samples during the training procedure, which leads to models favoring the non-fire category. To enhance the detection of small/tiny fire objects, it is necessary to effectively fuse low-level spatial features and high-level semantic features in DL architectures. Additionally, strategies for balancing fire and non-fire examples during training are crucial. These strategies may include data based methods [247,248] and loss function based methods [249] to address the category imbalance of small objects. By incorporating these strategies into the training process of DL architectures for fire detection, it is possible to enhance the model's capability to accurately detect small/ tiny fire objects, thereby improving overall performance and reliability in fire detection applications.

Computation-friendly networks. Unmanned aerial vehicles (UAVs) equipped with computer vision-based remote sensing systems present a pragmatic and increasingly viable option for swift, mobile, and cost-effective alternatives in forest fire monitoring, detection, and control. However, contemporary DL methods for fire detection often entail the design of numerous convolutional layers and thousands of feature channels, resulting in billions of floating point operations (FLOPs). Although they achieve satisfactory accuracy, it comes at the cost of increased computation time and memory requirement, rendering such models impractical for deployment on computation-limited devices. This limitation impedes the application of DL models in UAV and Internet of Things (IoT) based solutions for wildfire detection. Despite a wealth of review papers [7,250–253] on UAV-based fire detection solutions, it becomes apparent that there is a shortage of computation-friendly models suitable for deployment on UAVs and IoT devices, ensuring rapid fire detection. In tandem with developments in other domains, where there is a growing interest in crafting computation-friendly models, a similar exploration can be pursued in the domain of fire detection.

Efficient temporal–spatial models. Intelligent fire prevention and fire rescue have not benefited greatly from video based fire technologies, which now concentrate mostly on early fire detection and alarm activities. One the one hand, video data might help prevent fires by identifying risky activities before they happen, such smoking in places that are off-limits, using fires without permission in grasslands and forests, and blocking fire escapes. On the one hand, video data can offer valuable insights into the dynamic behavioral processes of fires at the early to mid-stage of development. These insights are essential for accurate fire detection and situational analysis, which in turn leads to a satisfactory recognition accuracy and improves the integration of automatic firefighting systems with video-based fire detection products. Therefore, video data analysis is essential to both fire prevention and fire rescue. Nevertheless, this benefit entails undue computational demands. In order to improve the dependability of fire technologies in real-world scenarios, urgent research is needed on methods for creating effective deep temporal–spatial networks based on video data, such as structure search, model pruning, knowledge distillation, and parameter quantization. This is crucial for the development of automatic fire prevention, fire detection and fire rescue systems.

9. Conclusion

Fire detection is crucial for mitigating the devastating impacts of fire disasters. This survey paper delivers a comprehensive overview of the research progress in fire detection leveraging advanced deep learning (DL) techniques. To foster a holistic understanding of the state-of-the-art in this field, we propose a systematic taxonomy that organizes existing fire detection approaches into three main branches: fire classification, fire location, and fire segmentation. By structuring the literature according to this taxonomy, we provide an in-depth review of representative DL methods for each branch, underscoring the efficacy of attention mechanisms, multi-level information fusion, lightweight networks, and deep ensemble models in enhancing the accuracy and efficiency of fire detection systems. Additionally, we review the commonly used datasets, evaluation metrics, and the corresponding performance results reported in the related literature. We also discuss the current challenges and pinpoint promising future research directions in fire detection. In conclusion, while there have been significant advances in fire detection owing to the remarkable developments in DL techniques, considerable challenges remain. This survey is intended to serve as a valuable resource to guide researchers in their future efforts to further enhance fire detection technologies.

CRediT authorship contribution statement

Guangtao Cheng: Writing – review & editing, Writing – original draft, Methodology. **Xue Chen:** Methodology. **Chenyi Wang:** Visualization. **Xiaobo Li:** Writing – review & editing. **Baoyi Xian:** Methodology. **Hao Yu:** Resources.

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

Data will be made available on request.

Acknowledgments

This work was supported by Tianjin City High School Science and Technology Fund Planing Project under Grant 2022SK040.

References

[1] A. Gaur, A. Singh, A. Kumar, K.S. Kulkarni, S. Lala, K. Kapoor, V. Srivastava, A. Kumar, S.C. Mukhopadhyay, Fire sensing technologies: A review, IEEE Sens. J. 19 (9) (2019) 3191–3202.
[2] B.U. Töreyin, Y. Dedeoğlu, U. Güdükbay, A.E. Cetin, Computer vision based method for real-time fire and flame detection, Pattern Recognit. Lett. 27 (1) (2006) 49–58.
[3] A.E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y.H. Habiboğlu, B.U. Töreyin, S. Verstockt, Video fire detection–review, Digit. Signal Process. 23 (6) (2013) 1827–1843.
[4] J.A. Ojo, J.A. Oladosu, Video-Based Smoke Detection Algorithms: A Chronological Survey, International Institute for Science, Technology and Education (IISTE), ISSN, 2014, pp. 38–50.
[5] P. Matlani, M. Shrivastava, A survey on video smoke detection, in: Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2016, vol. 1, Springer, 2018, pp. 211–222.
[6] F. Bu, M.S. Gharajeh, Intelligent and vision-based fire detection systems: A survey, Image Vis. Comput. 91 (2019) 103803.
[7] A. Dhall, A. Dhasade, A. Nalwade, V. Mohan Raj, V. Kulkarni, A survey on systematic approaches in managing forest fires, Appl. Geogr. 121 (2020) 102266.
[8] S. Dong, P. Wang, K. Abbas, A survey on deep learning and its applications, Comp. Sci. Rev. 40 (2021) 100379.
[9] P.-H. Conze, G. Andrade-Miranda, V.K. Singh, V. Jaouen, D. Visvikis, Current and emerging trends in medical image segmentation with deep learning, IEEE Trans. Radiat. Plasma Med. Sci. (2023).
[10] J. Yu, D. Tao, M. Wang, Y. Rui, Learning to rank using user clicks and visual features for image retrieval, IEEE Trans. Cybern. 45 (4) (2015) 767–779.
[11] S. Geetha, C. Abhishek, C. Akshayanat, Machine vision based fire detection techniques: A survey, Fire Technol. 57 (2021) 591–623.
[12] S. Chaturvedi, P. Khanna, A. Ojha, A survey on vision-based outdoor smoke detection techniques for environmental safety, ISPRS J. Photogramm. Remote Sens. 185 (2022) 158–187.
[13] B. Ko, J.-Y. Kwak, J.-Y. Nam, Wildfire smoke detection using temporospatial features and random forest classifiers, Opt. Eng. 51 (1) (2012) 017208.
[14] A. Cetin, 2014, Available: http://signal.ee.bilkent.edu.tr/VisiFire/.
[15] P. Foggia, A. Saggese, M. Vento, Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion, IEEE Trans. Circuits Syst. Video Technol. 25 (9) (2015) 1545–1556.
[16] Z. Yin, B. Wang, F. Yuan, X. Xia, J. Shi, A deep normalization and convolutional neural network for image smoke detection, IEEE Access 5 (2017) 18429–18438.
[17] T. Toulouse, L. Rossi, A. Campana, T. Celik, M.A. Akhloufi, Computer vision for wildfire research: An evolving image dataset for processing and analysis, Fire Saf. J. 92 (2017) 188–194.
[18] R. Ba, C. Chen, J. Yuan, W. Song, S. Lo, SmokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention, Remote Sens. 11 (14) (2019) 1702.
[19] F. Yuan, L. Zhang, X. Xia, B. Wan, Q. Huang, X. Li, Deep smoke segmentation, Neurocomputing 357 (2019) 248–260.
[20] A.I. For Mankind and HPWREN, 2020, Available: https://github.com/aiformankind/wildfire-smoke-dataset.
[21] L. He, X. Gong, S. Zhang, L. Wang, F. Li, Efficient attention based deep fusion CNN for smoke detection in fog environment, Neurocomputing 434 (2021) 224–238.
[22] X. Gong, H. Hu, Z. Wu, L. He, L. Yang, F. Li, Dark-channel based attention and classifier retraining for smoke detection in foggy environments, Digit. Signal Process. 123 (2022) 103454.
[23] A. Dewangan, Y. Pande, H.-W. Braun, F. Vernon, I. Perez, I. Altintas, G.W. Cottrell, M.H. Nguyen, FIgLib & SmokeyNet: Dataset and deep learning model for real-time wildland fire smoke detection, Remote Sens. 14 (4) (2022) 1007.
[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[25] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[28] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[29] C. Fran, et al., Deep learning with depth wise separable convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017.
[30] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al., Searching for mobilenetv3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314–1324.
[31] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size, 2016, arXiv preprint arXiv:1602.07360.
[32] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, J. Sun, Repvgg: Making vgg-style convnets great again, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13733–13742.
[33] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[34] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Netw. Learn. Syst. (2021).
[35] H. Yin, Y. Wei, H. Liu, S. Liu, C. Liu, Y. Gao, Deep convolutional generative adversarial network and convolutional neural network for smoke detection, Complexity 2020, 1–12.
[36] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint arXiv:1511.06434.
[37] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E.S. Gündüz, K. Polat, Attention based CNN model for fire detection and localization in real-world images, Expert Syst. Appl. 189 (2022) 116114.
[38] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[39] V.E. Sathishkumar, J. Cho, M. Subramanian, O.S. Naren, Forest fire and smoke detection using deep learning-based learning without forgetting, Fire Ecol. 19 (1) (2023) 1–17.
[40] Y. Valikhujaev, A. Abdusalomov, Y.I. Cho, Automatic fire and smoke detection method for surveillance systems based on dilated CNNs, Atmosphere 11 (11) (2020) 1241.
[41] J. Wang, J. Yu, Z. He, DECA: a novel multi-scale efficient channel attention module for object detection in real-life fire images, Appl. Intell. (2022) 1–14.
[42] T. Li, C. Zhang, H. Zhu, J. Zhang, Adversarial fusion network for forest fire smoke detection, Forests 13 (3) (2022) 366.
[43] F. Yuan, L. Zhang, B. Wan, X. Xia, J. Shi, Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition, Mach. Vis. Appl. 30 (2019) 345–358.
[44] M. Jeon, H.-S. Choi, J. Lee, M. Kang, Multi-scale prediction for fire detection using convolutional neural network, Fire Technol. 57 (5) (2021) 2533–2551.
[45] S. Zheng, X. Zou, P. Gao, Q. Zhang, F. Hu, Y. Zhou, Z. Wu, W. Wang, S. Chen, A forest fire recognition method based on modified deep CNN model, Forests 15 (1) (2024) 111.
[46] S. Li, Q. Yan, P. Liu, An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism, IEEE Trans. Image Process. 29 (2020) 8467–8475.
[47] G. Cheng, X. Chen, J. Gong, Deep convolutional network with pixel-aware attention for smoke recognition, Fire Technol. 58 (4) (2022) 1839–1862.
[48] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, K. Barnard, Attentional feature fusion, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 3560–3569.
[49] N. Guo, J. Liu, K. Di, K. Gu, J. Qiao, A hybrid attention model based on first-order statistical features for smoke recognition, Sci. China Technol. Sci. (2024) 1–14.

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[51] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16 × 16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[52] Y. Zhao, Y. Wang, H.-K. Jung, Y. Jin, D. Hua, S. Xu, Lightweight smoke recognition based on deep convolution and self-attention, Math. Probl. Eng. 2022 (2022).

[53] G. Cheng, Y. Zhou, S. Gao, Y. Li, H. Yu, Convolution-enhanced vision transformer network for smoke recognition, Fire Technol. 59 (2) (2023) 925–948.

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

[55] H. Yar, T. Hussain, Z.A. Khan, D. Koundal, M.Y. Lee, S.W. Baik, et al., Vision sensor-based real-time fire detection in resource-constrained IoT environments, Comput. Intell. Neurosci. 2021 (2021).

[56] J. Zhang, W. Xie, H. Liu, W. Dang, A. Yu, D. Liu, Compressed dual-channel neural network with application to image-based smoke detection, IET Image Process. 16 (4) (2022) 1036–1043.

[57] K. Gu, Z. Xia, J. Qiao, W. Lin, Deep dual-channel neural network for image-based smoke detection, IEEE Trans. Multimed. 22 (2) (2019) 311–323.

[58] H. Pan, D. Badawi, X. Zhang, A.E. Cetin, Additive neural network for forest fire detection, Signal Image Video Process. 14 (2020) 675–682.

[59] A. Afrasiyabi, B. Nasir, O. Yildiz, F.T.Y. Vural, A.E. Cetin, An energy efficient additive neural network, in: 2017 25th Signal Processing and Communications Applications Conference, SIU, IEEE, 2017, pp. 1–4.

[60] Y. Liu, W. Qin, K. Liu, F. Zhang, Z. Xiao, A dual convolution network using dark channel prior for image smoke classification, IEEE Access 7 (2019) 60697–60706.

[61] K. He, J. Sun, X. Tang, Single imahe2010singlege haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.

[62] F. Zhang, W. Qin, Y. Liu, Z. Xiao, J. Liu, Q. Wang, K. Liu, A dual-channel convolution neural network for image smoke detection, Multimedia Tools Appl. 79 (2020) 34587–34603.

[63] H. Liu, F. Lei, C. Tong, C. Cui, L. Wu, Visual smoke detection based on ensemble deep cnns, Displays 69 (2021) 102020.

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

[65] K. Muhammad, S. Khan, V. Palade, I. Mehmood, V.H.C. De Albuquerque, Edge intelligence-assisted smoke detection in foggy surveillance environments, IEEE Trans. Ind. Inform. 16 (2) (2019) 1067–1075.

[66] S. Chen, Y. Cao, X. Feng, X. Lu, Global2Salient: Self-adaptive feature aggregation for remote sensing smoke detection, Neurocomputing 466 (2021) 202–220.

[67] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.

[68] Y. Zheng, G. Zhang, S. Tan, Z. Yang, D. Wen, H. Xiao, A forest fire smoke detection model combining convolutional neural network and vision transformer, Front. For. Glob. Change 6 (2023) 1136969.

[69] J.-J. Liu, Q. Hou, M.-M. Cheng, C. Wang, J. Feng, Improving convolutional networks with self-calibrated convolutions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10096–10105.

[70] Y. Hu, X. Lu, Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features, Multimedia Tools Appl. 77 (2018) 29283–29301.

[71] M. Yin, C. Lang, Z. Li, S. Feng, T. Wang, Recurrent convolutional network for video-based smoke detection, Multimedia Tools Appl. 78 (2019) 237–256.

[72] L.R. Medsker, L. Jain, Recurrent neural networks, Des. Appl. 5 (64–67) (2001) 2.

[73] R. Kaur, S. Singh, A comprehensive review of object detection with deep learning, Digit. Signal Process. 132 (2023) 103812.

[74] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[75] T. Diwan, G. Anirudh, J.V. Tembhurne, Object detection using YOLO: Challenges, architectural successors, datasets and applications, Multimedia Tools Appl. 82 (6) (2023) 9243–9275.

[76] P. Li, W. Zhao, Image fire detection algorithms based on convolutional neural networks, Case Stud. Therm. Eng. 19 (2020) 100625.

[77] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. 28 (2015).

[78] J. Dai, Y. Li, K. He, J. Sun, R-fcn: Object detection via region-based fully convolutional networks, Adv. Neural Inf. Process. Syst 29 (2016).

[79] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 21–37.

[80] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, 2018, arXiv preprint arXiv:1804.02767.

[81] A. Abdusalomov, N. Baratov, A. Kutlimuratov, T.K. Whangbo, An improvement of the fire detection and classification method using YOLOv3 for surveillance systems, Sensors 21 (19) (2021) 6519.

[82] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection, 2020, arXiv preprint arXiv:2004.10934.

[83] X. Zheng, F. Chen, L. Lou, P. Cheng, Y. Huang, Real-time detection of full-scale forest fire smoke based on deep convolution neural network, Remote Sens. 14 (3) (2022) 536.

[84] M. Tan, R. Pang, Q.V. Le, Efficientdet: Scalable and efficient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10781–10790.

[85] Y. Al-Smadi, M. Alauthman, A. Al-Qerem, A. Aldweesh, R. Quaddoura, F. Aburub, K. Mansour, T. Alhmiedat, Early wildfire smoke detection using different YOLO models, Machines 11 (2) (2023) 246.

[86] G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, L. Diaconu, J. Poznanski, L. Yu, P. Rai, R. Ferriday, et al., Ultralytics/yolov5: v3. 0, Zenodo (2020).

[87] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475.

[88] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271.

[89] S. Saponara, A. Elhanashi, A. Gagliardi, Real-time video fire/smoke detection based on CNN in antifire surveillance systems, J. Real-Time Image Process. 18 (2021) 889–900.

[90] Y.-Y. Qin, J.-T. Cao, X.-F. Ji, Fire detection method based on depthwise separable convolution and yolov3, Int. J. Autom. Comput. 18 (2021) 300–310.

[91] M. Masoom S, Q. Zhang, P. Dai, Y. Jia, Y. Zhang, J. Zhu, J. Wang, Early smoke detection based on improved YOLO-PCA network, Fire 5 (2) (2022) 40.

[92] J. Li, G. Zhou, A. Chen, Y. Wang, J. Jiang, Y. Hu, C. Lu, Adaptive linear feature-reuse network for rapid forest fire smoke detection model, Ecol. Inform. 68 (2022) 101584.

[93] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[94] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, W. Zuo, Enhancing geometric factors in model learning and inference for object detection and instance segmentation, IEEE Trans. Cybern. 52 (8) (2021) 8574–8586.

[95] Y. Sun, J. Feng, Fire and smoke precise detection method based on the attention mechanism and anchor-free mechanism, Complex Intell. Syst. (2023) 1–14.

[96] Y. Wang, C. Hua, W. Ding, R. Wu, Real-time detection of flame and smoke using an improved YOLOv4 network, Signal Image Video Process. 16 (4) (2022) 1109–1116.

[97] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13713–13722.

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

[99] Z. Wang, L. Wu, T. Li, P. Shi, A smoke detection model based on improved YOLOv5, Mathematics 10 (7) (2022) 1190.

[100] Y. Hu, J. Zhan, G. Zhou, A. Chen, W. Cai, K. Guo, Y. Hu, L. Li, Fast forest fire smoke detection using MVMNet, Knowl.-Based Syst. 241 (2022) 108219.

[101] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[102] X. Yang, J. Yang, J. Yan, Y. Zhang, T. Zhang, Z. Guo, X. Sun, K. Fu, Scrdet: Towards more robust detection for small, cluttered and rotated objects, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8232–8241.

[103] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, X. Xue, Arbitrary-oriented scene text detection via rotation proposals, IEEE Trans. Multimedia 20 (11) (2018) 3111–3122.

[104] H. Yin, M. Chen, W. Fan, Y. Jin, S.G. Hassan, S. Liu, Efficient smoke detection based on YOLOv5s, Mathematics 10 (19) (2022) 3493.

[105] F. Xu, X. Zhang, T. Deng, W. Xu, An image-based fire monitoring algorithm resistant to fire-like objects, Fire 7 (1) (2024) 3.

[106] B. Li, F. Xu, X. Li, C. Yu, X. Zhang, Early stage fire detection system based on shallow guide deep network, Fire Technol. (2024) 1–19.

[107] N. Ma, X. Zhang, H.-T. Zheng, J. Sun, Shufflenet v2: Practical guidelines for efficient cnn architecture design, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 116–131.

[108] J. Ma, Z. Zhang, W. Xiao, X. Zhang, S. Xiao, Flame and smoke detection algorithm based on ODConvBS-YOLOv5s, IEEE Access 11 (2023) 34005–34014.
[109] C. Li, A. Zhou, A. Yao, Omni-dimensional dynamic convolution, 2022, arXiv preprint arXiv:2209.07947.
[110] R. Hong, X. Wang, Y. Fang, H. Wang, C. Wang, H. Wang, Yolo-light: Remote straw-burning smoke detection based on depthwise separable convolution and channel attention mechanisms, Appl. Sci. 13 (9) (2023) 5690.
[111] C. Bahhar, A. Ksibi, M. Ayadi, M.M. Jamjoom, Z. Ullah, B.O. Soufiene, H. Sakli, Wildfire and smoke detection using staged YOLO model and ensemble CNN, Electronics 12 (1) (2023) 228.
[112] J. Qian, J. Lin, D. Bai, R. Xu, H. Lin, Omni-dimensional dynamic convolution meets bottleneck transformer: A novel improved high accuracy forest fire smoke detection model, Forests 14 (4) (2023) 838.
[113] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding, et al., PP-YOLO: An effective and efficient implementation of object detector, 2020, arXiv preprint arXiv:2007.12099.
[114] J. Zhan, Y. Hu, G. Zhou, Y. Wang, W. Cai, L. Li, A high-precision forest fire smoke detection approach based on ARGNet, Comput. Electron. Agric. 196 (2022) 106874.
[115] A. Odena, V. Dumoulin, C. Olah, Deconvolution and checkerboard artifacts, Distill 1 (10) (2016) e3.
[116] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, 2015, arXiv preprint arXiv:1511.07122.
[117] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, Yolox: Exceeding yolo series in 2021, 2021, arXiv preprint arXiv:2107.08430.
[118] J. Zhan, Y. Hu, W. Cai, G. Zhou, L. Li, PDAM–stpnnet: A small target detection approach for wildland fire smoke through remote sensing images, Symmetry 13 (12) (2021) 2260.
[119] J. Wang, X. Zhang, K. Jing, C. Zhang, Learning precise feature via self-attention and self-cooperation YOLOX for smoke detection, Expert Syst. Appl. 228 (2023) 120330.
[120] C. Jin, A. Zheng, Z. Wu, C. Tong, Real-time fire smoke detection method combining a self-attention mechanism and radial multi-scale feature connection, Sensors 23 (6) (2023) 3358.
[121] Y. Xu, J. Li, L. Zhang, H. Liu, F. Zhang, CNTCB-YOLOv7: An effective forest fire detection model based on ConvNeXtV2 and CBAM, Fire 7 (2) (2024) 54.
[122] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I.S. Kweon, S. Xie, Convnext v2: Co-designing and scaling convnets with masked autoencoders, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 16133–16142.
[123] Q. Hou, C.-Z. Lu, M.-M. Cheng, J. Feng, Conv2former: A simple transformer-style convnet for visual recognition, 2022, arXiv preprint arXiv:2211.11943.
[124] G. Jocher, A. Chaurasia, J. Qiu, YOLO by ultralytics, 2023, Available: https://github.com/ultralytics/ultralytics.
[125] X. Guo, Y. Cao, T. Hu, An efficient and lightweight detection model for forest smoke recognition, Forests 15 (1) (2024) 210.
[126] L. Yang, R.-Y. Zhang, L. Li, X. Xie, Simam: A simple, parameter-free attention module for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2021, pp. 11863–11874.
[127] Y. Wang, Y. Piao, H. Wang, H. Zhang, B. Li, An improved forest smoke detection model based on YOLOv8, Forests 15 (3) (2024) 409.
[128] I. Shamta, B.E. Demir, Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV, PLoS One 19 (3) (2024) e0299058.
[129] B. Yun, Y. Zheng, Z. Lin, T. Li, FFYOLO: A lightweight forest fire detection model based on YOLOv8, Fire 7 (3) (2024) 93.
[130] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, Q. Ren, Slim-neck by gsconv: A better design paradigm of detector architectures for autonomous vehicles, 2022, arXiv preprint arXiv:2206.02424.
[131] Z. Yang, Y. Shao, Y. Wei, J. Li, Precision-boosted forest fire target detection via enhanced YOLOv8 model, Appl. Sci. 14 (6) (2024) 2413.
[132] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 764–773.
[133] J. Li, Y. Wen, L. He, Scconv: spatial and channel reconstruction convolution for feature redundancy, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 6153–6162.
[134] Z. Zhang, L. Tan, R.L.K. Tiong, Ship-fire net: an improved YOLOv8 algorithm for ship fire detection, Sensors 24 (3) (2024) 727.
[135] C. Chaoxia, W. Shang, F. Zhang, Information-guided flame detection based on faster R-CNN, IEEE Access 8 (2020) 58923–58932.
[136] M. Jiang, Y. Zhao, F. Yu, C. Zhou, T. Peng, A self-attention network for smoke detection, Fire Saf. J. 129 (2022) 103547.
[137] Z. Guan, F. Min, W. He, W. Fang, T. Lu, Forest fire detection via feature entropy guided neural network, Entropy 24 (1) (2022) 128.
[138] Y. Li, W. Zhang, Y. Liu, Y. Jin, A visualized fire detection method based on convolutional neural network beyond anchor, Appl. Intell. 52 (11) (2022) 13280–13295.
[139] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, J. Shi, Foveabox: Beyound anchor-based object detection, IEEE Trans. Image Process. 29 (2020) 7389–7398.
[140] J. Huang, J. Zhou, H. Yang, Y. Liu, H. Liu, A small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection, Forests 14 (1) (2023) 162.
[141] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai, Deformable detr: Deformable transformers for end-to-end object detection, 2020, arXiv preprint arXiv:2010.04159.
[142] L. Jin, Y. Yu, J. Zhou, D. Bai, H. Lin, H. Zhou, SWVR: A lightweight deep learning algorithm for forest fire detection and recognition, Forests 15 (1) (2024) 204.
[143] J. Shi, W. Wang, Y. Gao, N. Yu, Optimal placement and intelligent smoke detection algorithm for wildfire-monitoring cameras, IEEE Access 8 (2020) 72326–72339.
[144] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[145] B.K. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1–3) (1981) 185–203.
[146] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
[147] C. Li, B. Yang, H. Ding, H. Shi, X. Jiang, J. Sun, Real-time video-based smoke detection with high accuracy and efficiency, Fire Saf. J. 117 (2020) 103184.
[148] Z. Zivkovic, Improved adaptive Gaussian mixture model for background subtraction, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004, ICPR 2004, vol. 2, IEEE, 2004, pp. 28–31.
[149] R. Kaabi, M. Bouchouicha, A. Mouelhi, M. Sayadi, E. Moreau, An efficient smoke detection algorithm based on deep belief network classifier using energy and intensity features, Electronics 9 (9) (2020) 1390.
[150] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[151] B. Zhang, L. Sun, Y. Song, W. Shao, Y. Guo, F. Yuan, DeepFireNet: A real-time video fire detection method based on multi-feature fusion, Math. Biosci. Eng. 17 (6) (2020) 7804–7818.
[152] Y. Xie, J. Zhu, Y. Cao, Y. Zhang, D. Feng, Y. Zhang, M. Chen, Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features, IEEE Access 8 (2020) 81904–81917.
[153] A. Gagliardi, F. de Gioia, S. Saponara, A real-time video smoke detection algorithm based on Kalman filter and CNN, J. Real-Time Image Process. (2021) 1–11.
[154] Y. Jia, W. Chen, M. Yang, L. Wang, D. Liu, Q. Zhang, Video smoke detection with domain knowledge and transfer learning from deep convolutional neural networks, Optik 240 (2021) 166947.
[155] Y. Wu, M. Chen, Y. Wo, G. Han, Video smoke detection base on dense optical flow and convolutional neural network, Multimedia Tools Appl. 80 (2021) 35887–35901.
[156] Y. Cheng, K. Chen, H. Bai, C. Mou, Y. Zhang, K. Yang, Y. Gao, Y. Liu, An efficient fire detection algorithm based on multi-scale convolutional neural network, Fire Mater. 46 (7) (2022) 981–992.
[157] J. Ryu, D. Kwak, Flame detection using appearance-based pre-processing and convolutional neural network, Appl. Sci. 11 (11) (2021) 5138.
[158] J. Chen, L.-h. Zou, J. Zhang, L.-h. Dou, The comparison and application of corner detection algorithms, J. Multimedia 4 (6) (2009).
[159] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[160] M.D. Nguyen, H.N. Vu, D.C. Pham, B. Choi, S. Ro, Multistage real-time fire detection using convolutional neural networks and long short-term memory networks, IEEE Access 9 (2021) 146667–146679.
[161] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625–2634.
[162] P. Wang, J. Zhang, H. Zhu, Fire detection in video surveillance using superpixel-based region proposal and ESE-ShuffleNet, Multimedia Tools Appl. (2021) 1–28.
[163] D.-K. Kwak, J.-K. Ryu, A study on the dynamic image-based dark channel prior and smoke detection using deep learning, J. Electr. Eng. Technol. (2021) 1–9.
[164] X. Chen, Y. Xue, Y. Zhu, R. Ma, A novel smoke detection algorithm based on improved mixed Gaussian and YOLOv5 for textile workshop environments, IET Image Process. 17 (7) (2023) 1991–2004.
[165] T. Liu, J. Cheng, Z. Yuan, H. Hua, K. Zhao, Video smoke detection with block DNCNN and visual change image, KSII Trans. Internet Inf. Syst. 14 (9) (2020).
[166] Y. Lee, J. Shim, False positive decremented research for fire and smoke detection in surveillance camera using spatial and temporal features based on deep learning, Electronics 8 (10) (2019) 1167.
[167] M. Park, B.C. Ko, Two-step real-time night-time fire detection in an urban environment using static ELASTIC-YOLOv3 and temporal fire-tube, Sensors 20 (8) (2020) 2202.

[168] H. Wang, A. Kembhavi, A. Farhadi, A.L. Yuille, M. Rastegari, Elastic: Improving cnns with dynamic scaling policies, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2258–2267.
[169] O. Kim, D.-J. Kang, Fire detection system using random forest classification for image sequences of complex background, Opt. Eng. 52 (6) (2013) 067202.
[170] M. Jeong, M. Park, J. Nam, B.C. Ko, Light-weight student LSTM for real-time wildfire smoke detection, Sensors 20 (19) (2020) 5508.
[171] B. Kim, J. Lee, A Bayesian network-based information fusion combined with DNNs for robust video fire detection, Appl. Sci. 11 (16) (2021) 7624.
[172] M. Shahid, I.-F. Chien, W. Sarapugdi, L. Miao, K.-L. Hua, Deep spatial-temporal networks for flame detection, Multimedia Tools Appl. 80 (2021) 35297–35318.
[173] Y. Cao, Q. Tang, S. Xu, F. Li, X. Lu, Quasivsd: efficient dual-frame smoke detection, Neural Comput. Appl. 34 (11) (2022) 8539–8550.
[174] M. Shahid, J.J. Virtusio, Y.-H. Wu, Y.-Y. Chen, M. Tanveer, K. Muhammad, K.-L. Hua, Spatio-temporal self-attention network for fire detection and segmentation in video surveillance, IEEE Access 10 (2021) 1259–1275.
[175] X. Qiang, G. Zhou, A. Chen, X. Zhang, W. Zhang, Forest fire smoke detection under complex backgrounds using TRPCA and TSVB, Int. J. Wildland Fire 30 (5) (2021) 329–350.
[176] W. Du, Y. Wang, Y. Qiao, Recurrent spatial-temporal attention network for action recognition in videos, IEEE Trans. Image Process. 27 (3) (2017) 1347–1360.
[177] Y. Huo, Q. Zhang, Y. Zhang, J. Zhu, J. Wang, 3DVSD: An end-to-end 3D convolutional object detection network for video smoke detection, Fire Saf. J. 134 (2022) 103690.
[178] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[179] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2021) 3523–3542.
[180] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, 2014, arXiv preprint arXiv:1412.7062.
[181] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.
[182] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587.
[183] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[184] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[185] Z. Wang, T. Peng, Z. Lu, Comparative research on forest fire image segmentation algorithms based on fully convolutional neural networks, Forests 13 (7) (2022) 1133.
[186] H. Harkat, J.M. Nascimento, A. Bernardino, H.F. Thariq Ahmed, Assessing the impact of the loss function and encoder architecture for fire aerial images segmentation using deeplabv3+, Remote Sens. 14 (9) (2022) 2023.
[187] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, QuÉBec City, QC, Canada, September 14, Proceedings 3, Springer, 2017, pp. 240–248.
[188] S. Khan, K. Muhammad, T. Hussain, J. Del Ser, F. Cuzzolin, S. Bhattacharyya, Z. Akhtar, V.H.C. de Albuquerque, Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments, Expert Syst. Appl. 182 (2021) 115125.
[189] F. Yuan, L. Zhang, X. Xia, Q. Huang, X. Li, A gated recurrent network with dual classification assistance for smoke semantic segmentation, IEEE Trans. Image Process. 30 (2021) 4409–4422.
[190] F. Hou, X. Rui, Y. Chen, X. Fan, Flame and smoke semantic dataset: Indoor fire detection with deep semantic segmentation model, Electronics 12 (18) (2023) 3778.
[191] G. Wen, F. Zhou, Y. Ma, H. Pan, H. Geng, J. Cao, K. Li, F. Yuan, A dense multi-scale context and asymmetric pooling embedding network for smoke segmentation, IET Comput. Vis. (2023).
[192] F. Yuan, L. Zhang, X. Xia, Smoke semantic segmentation with multi-scale residual paths and weighted middle surveillances, Multimedia Tools Appl. (2023) 1–26.
[193] G. Sun, Y. Wen, Y. Li, Instance segmentation using semi-supervised learning for fire recognition, Heliyon 8 (12) (2022).
[194] X. Wang, R. Zhang, T. Kong, L. Li, C. Shen, Solov2: Dynamic and fast instance segmentation, Adv. Neural Inf. Process. Syst. 33 (2020) 17721–17732.
[195] M. Li, Y. Zhang, L. Mu, J. Xin, Z. Yu, S. Jiao, H. Liu, G. Xie, Y. Yingmin, A real-time fire segmentation method based on a deep learning approach, IFAC-PapersOnLine 55 (6) (2022) 145–150.
[196] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018, pp. 3–11.
[197] T. Wang, J. Hong, Y. Han, G. Zhang, S. Chen, T. Dong, Y. Yang, H. Ruan, AOSVSSNet: Attention-guided optical satellite video smoke segmentation network, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 8552–8566.
[198] X. Hu, F. Jiang, X. Qin, S. Huang, X. Yang, F. Meng, An optimized smoke segmentation method for forest and grassland fire based on the unet framework, Fire 7 (3) (2024) 68.
[199] K. Muhammad, H. Ullah, S. Khan, M. Hijji, J. Lloret, Efficient fire segmentation for internet-of-things-assisted intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. (2022).
[200] F. Yuan, K. Li, C. Wang, Z. Fang, A lightweight network for smoke semantic segmentation, Pattern Recognit. 137 (2023) 109289.
[201] Y. Zheng, Z. Wang, B. Xu, Y. Niu, Multi-scale semantic segmentation for fire smoke image based on global information and U-net, Electronics 11 (17) (2022) 2718.
[202] Z. Wang, P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian, W. Cui, Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery, Remote Sens. 14 (1) (2021) 45.
[203] L. Zhang, F. Yuan, X. Xia, Edge-reinforced attention network for smoke semantic segmentation, Multimedia Tools Appl. (2023) 1–26.
[204] J. Wang, X. Fan, X. Yang, T. Tjahjadi, Y. Wang, Semi-supervised learning for forest fire segmentation using UAV imagery, Forests 13 (10) (2022) 1573.
[205] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I, Springer, 2018, pp. 421–429.
[206] G. Wang, F. Wang, H. Zhou, H. Lin, Fire in focus: Advancing wildfire image segmentation by focusing on fire edges, Forests 15 (1) (2024) 217.
[207] Y. Li, W. Zhang, Y. Liu, X. Shao, A lightweight network for real-time smoke semantic segmentation based on dual paths, Neurocomputing 501 (2022) 258–269.
[208] S. Li, P. Liu, Q. Yan, R. Qian, Optimized deep learning model for fire semantic segmentation, Comput. Mater. Continua 72 (3) (2022).
[209] T. Marto, A. Bernardino, G. Cruz, Fire and smoke segmentation using active learning methods, Remote Sens. 15 (17) (2023) 4136.
[210] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[211] B. Amaral, M. Niknejad, C. Barata, A. Bernardino, Weakly supervised fire and smoke segmentation in forest images with CAM and CRF, in: 2022 26th International Conference on Pattern Recognition, ICPR, IEEE, 2022, pp. 442–448.
[212] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[213] W. Wang, J. Shen, L. Shao, Video salient object detection via fully convolutional networks, IEEE Trans. Image Process. 27 (1) (2017) 38–49.
[214] M.A. Islam, S. Naha, M. Rochan, N. Bruce, Y. Wang, Label refinement network for coarse-to-fine semantic segmentation, 2017, arXiv preprint arXiv:1703.00551.
[215] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VIII 14, Springer, 2016, pp. 483–499.
[216] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters–improve semantic segmentation by global convolutional network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4353–4361.
[217] G. Lin, A. Milan, C. Shen, I. Reid, Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1925–1934.
[218] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.
[219] H. Ding, X. Jiang, B. Shuai, A.Q. Liu, G. Wang, Context contrasted feature and gated multi-scale aggregation for scene segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2393–2402.
[220] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Learning a discriminative feature network for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1857–1866.

[221] F. Yuan, L. Zhang, X. Xia, Q. Huang, X. Li, A wave-shaped deep neural network for smoke density estimation, IEEE Trans. Image Process. 29 (2019) 2301–2313.
[222] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
[223] E. Romera, J.M. Alvarez, L.M. Bergasa, R. Arroyo, Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, IEEE Trans. Intell. Transp. Syst. 19 (1) (2017) 263–272.
[224] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, L.J. Latecki, Lednet: A lightweight encoder-decoder network for real-time semantic segmentation, in: 2019 IEEE International Conference on Image Processing, ICIP, IEEE, 2019, pp. 1860–1864.
[225] H. Li, P. Xiong, H. Fan, J. Sun, Dfanet: Deep feature aggregation for real-time semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9522–9531.
[226] T. Wu, S. Tang, R. Zhang, J. Cao, Y. Zhang, Cgnet: A light-weight context guided network for semantic segmentation, IEEE Trans. Image Process. 30 (2020) 1169–1179.
[227] J. He, Z. Deng, Y. Qiao, Dynamic multi-scale filters for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3562–3572.
[228] F. Yuan, Z. Dong, L. Zhang, X. Xia, J. Shi, Cubic-cross convolutional attention and count prior embedding for smoke segmentation, Pattern Recognit. 131 (2022) 108902.
[229] Z. Huang, L. Huang, Y. Gong, C. Huang, X. Wang, Mask scoring r-cnn, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6409–6418.
[230] J. Wen, M. Burke, Wildfire smoke plume segmentation using geostationary satellite imagery, 2021, arXiv preprint arXiv:2109.01637.
[231] S. Frizzi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, M. Sayadi, Convolutional neural network for smoke and fire semantic segmentation, IET Image Process. 15 (3) (2021) 634–647.
[232] G. Perrolas, M. Niknejad, R. Ribeiro, A. Bernardino, Scalable fire and smoke segmentation from aerial images using convolutional neural networks and quad-tree search, Sensors 22 (5) (2022) 1701.
[233] R.A. Finkel, J.L. Bentley, Quad trees a data structure for retrieval on composite keys, Acta Inform. 4 (1974) 1–9.
[234] L. Martins, F. Guede-Fernández, R. Valente de Almeida, H. Gamboa, P. Vieira, Real-time integration of segmentation techniques for reduction of false positive rates in fire plume detection systems during forest fires, Remote Sens. 14 (11) (2022) 2701.
[235] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, X. Bai, Multi-oriented text detection with fully convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4159–4167.
[236] X. Li, Z. Chen, Q.J. Wu, C. Liu, 3D parallel fully convolutional networks for real-time video wildfire smoke detection, IEEE Trans. Circuits Syst. Video Technol. 30 (1) (2018) 89–103.
[237] G. Xu, Y. Zhang, Q. Zhang, G. Lin, Z. Wang, Y. Jia, J. Wang, Video smoke detection based on deep saliency network, Fire Saf. J. 105 (2019) 277–285.
[238] Y. Jia, H. Du, H. Wang, R. Yu, L. Fan, G. Xu, Q. Zhang, Automatic early smoke segmentation based on conditional generative adversarial networks, Optik 193 (2019) 162879.
[239] Z. Ding, Y. Zhao, A. Li, Z. Zheng, Spatial–temporal attention two-stream convolution neural network for smoke region detection, Fire 4 (4) (2021) 66.
[240] G. Zhu, Z. Chen, C. Liu, X. Rong, W. He, 3D video semantic segmentation for wildfire smoke, Mach. Vis. Appl. 31 (2020) 1–10.
[241] F.-A. Croitoru, V. Hondru, R.T. Ionescu, M. Shah, Diffusion models in vision: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2023).
[242] M. Wang, W. Deng, Deep visual domain adaptation: A survey, Neurocomputing 312 (2018) 135–153.
[243] D. Zhang, J. Han, G. Cheng, M.-H. Yang, Weakly supervised object localization and detection: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2021) 5866–5885.
[244] Y. Chen, M. Mancini, X. Zhu, Z. Akata, Semi-supervised and unsupervised deep visual learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2022).
[245] Y. Liu, P. Sun, N. Wergeles, Y. Shang, A survey and performance evaluation of deep learning methods for small object detection, Expert Syst. Appl. 172 (2021) 114602.
[246] G. Cheng, X. Yuan, X. Yao, K. Yan, Q. Zeng, X. Xie, J. Han, Towards large-scale small object detection: Survey and benchmarks, IEEE Trans. Pattern Anal. Mach. Intell. (2023).
[247] A. Shrivastava, A. Gupta, R. Girshick, Training region-based object detectors with online hard example mining, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 761–769.
[248] Y. Cao, K. Chen, C.C. Loy, D. Lin, Prime sample attention in object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11583–11591.
[249] Q. Qian, L. Chen, H. Li, R. Jin, Dr loss: Improving object detection by distributional ranking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12164–12172.
[250] A. Bouguettaya, H. Zarzour, A.M. Taberkit, A. Kechida, A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms, Signal Process. 190 (2022) 108309.
[251] F. Carta, C. Zidda, M. Putzu, D. Loru, M. Anedda, D. Giusto, Advancements in forest fire prevention: A comprehensive survey, Sensors 23 (14) (2023) 6635.
[252] A. Rahman, S. Sakif, N. Sikder, M. Masud, H. Aljuaid, A.K. Bairagi, Unmanned aerial vehicle assisted forest fire detection using deep convolutional neural network, Intell. Autom. Soft Comput. 35 (3) (2023).
[253] J. Su, X. Zhu, S. Li, W.-H. Chen, AI meets UAVs: A survey on AI empowered UAV perception systems for precision agriculture, Neurocomputing 518 (2023) 242–270.

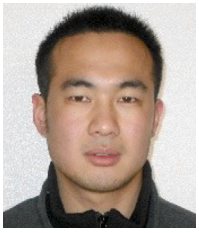
Guangtao Cheng received the B.S. degree in information and computing science from Northeast Petroleum University, Heilongjiang, China, in 2007, the M.S. degree in computational mathematics from Northeast University, Shenyang, China, in 2009, and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2015. He is currently a senior engineer with the School of Information Engineering, Tianjin University of Commerce. His current main areas of interests include image processing, computer vision, fire detection, and deep learning.

Xue Chen received her Ph.D. degree from the School of Computer Science and Technology, Tianjin University, Tianjin, China. She is currently a lecturer at the law school, Tianjin University, China. Her research interests include complex network, large-scale data mining, and network embedding.

Chenyi Wang is currently pursuing the B.S. degree in intelligent science and technology with the School of Information Engineering, Tianjin University of Commerce, Tianjin, China. Her current research interests include fire detection, deep learning.

Xiaobo Li received his B.S. degree in mathematics and applied mathematics from and the Ph.D. degree in optical engineering from Tianjin University, Tianjin, China, in 2014, and 2019, respectively. He worked as a Post-Doctoral Researcher with the Chinese University of Hong Kong, Hong Kong, China, from 2020–2022. He is currently an Associate Professor with the School of Marine Science and Technology, Tianjin University, Tianjin, China. His main research interests include ocean optics, polarization imaging, and marine metrology.

Baoyi Xian is currently pursuing the B.S. degree in intelligent science and technology with the School of Information Engineering, Tianjin University of Commerce, Tianjin, China. His current research interests include fire detection, deep learning.



Hao Yu received the M.S. degree in probability theory and mathematical statistics from Tianjin University, Tianjin, China, in 2009. He is currently a senior engineer with Huixiang Intelligent Technology (Tianjin) Co., Ltd, Tianjin, China. His main research interests are machine vision and intelligent detection, theoretical research and engineering applications of deep learning algorithms.