

Article

Improving Computer Vision-Based Wildfire Smoke Detection by Combining SE-ResNet with SVM

Xin Wang ¹, Jinxin Wang ^{2,*} , Linlin Chen ³ and Yinan Zhang ²¹ Xuzhou Fu'an Technology Co., Ltd., Xuzhou 221008, China; dianaxinwang@163.com² School of Safety Engineering, China University of Mining and Technology, Xuzhou 221116, China; zhangyinan@cumt.edu.cn³ Gengcun Coal Mine, Henan Dayou Energy Company Limited, Yima 472400, China

* Correspondence: wangjinxin@cumt.edu.cn

Abstract: Wildfire is one of the most critical natural disasters that poses a serious threat to human lives as well as ecosystems. One issue hindering a high accuracy of computer vision-based wildfire detection is the potential for water mists and clouds to be marked as wildfire smoke due to the similar appearance in images, leading to an unacceptable high false alarm rate in real-world wildfire early warning cases. This paper proposes a novel hybrid wildfire smoke detection approach by combining the multi-layer ResNet architecture with SVM to extract the smoke image dynamic and static characteristics, respectively. The ResNet model is improved via the SE attention mechanism and fully convolutional network as SE-ResNet. A fusion decision procedure is proposed for wildfire early warning. The proposed detection method was tested on open datasets and achieved an accuracy of 98.99%. The comparisons with AlexNet, VGG-16, GoogleNet, SE-ResNet-50 and SVM further illustrate the improvements.

Keywords: wildfire early warning; smoke detection; deep learning; support vector machine; computer vision



Citation: Wang, X.; Wang, J.; Chen, L.; Zhang, Y. Improving Computer Vision-Based Wildfire Smoke Detection by Combining SE-ResNet with SVM. *Processes* **2024**, *12*, 747. <https://doi.org/10.3390/pr12040747>

Academic Editors: Yuanyuan Pu, Sitao Zhu and Xinglong Zhao

Received: 4 March 2024

Revised: 30 March 2024

Accepted: 5 April 2024

Published: 7 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wildfires, caused by natural factors and various improper human activities, e.g., bonfires, burning of debris and littering lit cigarettes, pose a serious threat to human lives, vegetation canopy and national economy every year, especially for the countries and regions with large forestland area and dry climates [1,2]. The wildfire that happened in Australia in 2020 caused 23 fatalities and burnt vegetation over more than 14 million acres [3]. In Canada, wildfires have burned approximately 2 million hectares of forest annually over the last several decades [4]. With the deepening impact of climate change, as well as the ever-growing areas of wildland–urban interface and human activities, the frequency and destructiveness of large wildfires keep increasing, which makes the escalating wildfire situation a global issue. In the last 10 years, the number of wildfires and affected forest area increased by 400% and 600%, respectively, in the western United States [5]. This pressure motivates fire management agencies and researchers to pursue efficient real-time wildfire early discovery and decision-making technologies to minimize the loss.

Smoke rises and becomes visible before fire flames, which provides an effective indicator for wildfire early warning [6]. Conventional smoke sensors detect smoke particles according to the micro-current or absorbed light intensity changes of the transmitter, and report a fire alarm if it is present. Although being widely used in indoor fire detection, these detectors are installed in a fixed location and usually ask for a stable background condition, which therefore have limitations on small coverage area, delay in fire detection, and high false alarm rate caused by excessive environment humidity. These disadvantages prohibit smoke sensors from being widely using in outdoor, large, open space fire alarm systems, i.e., wildfire detection. With the continued development of machine learning, wildfire

detection is working toward online intelligent monitoring products and systems. Under such a trend, computer vision-based wildfire detection has recently been suggested for wildfire early discovery in order to overcome the aforementioned disadvantages. Different from sensor-based fire detection, these methods analyze and extract smoke characteristics from fire/smoke images captured by ground-, unmanned aerial vehicle- or satellite-based online monitoring systems, and detect the presence of smoke using various machine learning models. Computer vision-based approaches get rid of the limitations of environment condition, have a wide coverage area and are thought to be a promising wildfire online monitoring technology nowadays.

Existing computer vision-based wildfire detection approaches fall into two categories: conventional machine learning- and deep learning-based methods. Conventional machine learning methods firstly make use of various advanced image processing techniques to characterize wildfire smoke based on various manually extracted features, e.g., colors [7–10], textures [11–13], energy [14], irregular shape [15,16] and HOG [17,18], and then construct a smoke identification model for wildfire early warning via mining the corresponding relations between image features and object classifications. Although being well understood and fundamentally sound, the detection accuracy is highly related to the extracted image features and a false or missing warning may occur if the feature extraction is incomplete. Instead of extracting features manually, the deep learning approach automatically obtains smoke features from the training datasets by exploiting a series of convolution and pooling operations, which avoids the professionalism, limitations and singularity of manual feature extraction, and thus can get a complete description of wildfire smoke. Currently, researchers attempt to investigate and detect wildfire smoke using various deep learning network architecture [19–25]. Yin et al. [26] proposes a novel deep normalization network (DCDNN) for smoke image recognition by combining normalization methods with convolutional neural networks. Yuan et al. [27] present a deep multi-scale network for smoke image recognition. In their approach, multi-scale convolution was introduced to obtain scale invariance avoiding image features shifting at various scales. Sun et al. [28] make use of kernel principal component analysis to simplify the computational pressure of CNN, and design optimization strategies for multi-convolution kernels and batch normalization to improve loss functions. Celik et al. [29] propose a method which combines improved static segmentation and connectivity analysis-based dynamic segmentation. They established a multi-channel CNN by utilizing the characteristics of color, shape and area changes. Yuan carried out a series of studies on wildfire smoke detection using a deep convolutional neural network and its extension [30–32]. Recently, he presented a comprehensive review of the deep learning-based approach for wildfire detection and early warning [33].

Despite a number of deep learning-based approaches that have been developed and applied to improve wildfire smoke detection, one issue hindering high accuracy is the potential for water mists and clouds to be marked as wildfire smoke due to the similar appearance in images, leading to an unacceptably high false alarm rate in real-world wildfire early warning cases. Some researchers are devoted to improving the identification accuracy between wildfire smoke and water mists/clouds. Existing solutions mainly focus on extracting multiple and complete image features for wildfire smoke characterization [34–36], or, constructing a robust smoke detection model via using improved classical deep neural networks architectures, e.g., AlexNet [37], VGG [38], GoogleNet [39,40] and DenseNet [41]. Furthermore, some works also explore how to establish high-depth neural networks architectures to extract the smoke image characteristics, where shallow convolutional layers with small receptive fields focus on extracting local image features, e.g., edges, colors and textures, while deep convolutional layers have a large receptive field and extract rich semantic information on a large scale [42–44]. However, recent evidence reveals that the detection accuracy rate will not increase after reaching a certain level even for a “very deep” model [45]. The residual neural network (ResNet) mitigates this problem by amending the underlying mapping with the residual one and improves the detection accuracy to new levels. However, as for wildfire smoke detection and location, the false warning rate is still

hard to accept for real-world practice, which limits its applicability. Existing methods are not effective in distinguishing between wildfires and clouds due to the incomplete image feature extraction and vanishing gradient.

In this paper, we propose a novel hybrid wildfire smoke detection approach by combining deep learning architectures and conventional machine learning, i.e., SE-ResNet and SVM. The SE attention mechanism is introduced to quantify the relationships between multiple extracted features so as to improve the performance and generalization ability of the ResNet model. The image micromorphology and semantic features are extracted using ResNet and SVM, respectively, to characterize the wildfire smoke from different perspectives. In order to locate and segment the smoke, a fully convolutional network is utilized to obtain the classification of each pixel. A fusion decision procedure is proposed for wildfire early warning. The proposed detection method is tested on open datasets and achieves an F1-score of 99%. The comparisons with AlexNet, VGG-16, GoogleNet, SE-ResNet-50 and SVM further illustrate the improvements.

The remainder of the paper is organized in the following manner. Section 2 presents the whole framework and early warning procedure of the proposed hybrid approach. Section 3 establishes an improved ResNet-based wildfire smoke detection model by taking advantage of the SE attention mechanism and fully convolutional network. Section 4 explains the histogram of oriented gradient (HOG) feature extraction and smoke identification using SVM. A real case study is presented in Section 5, and the advantages are illustrated via comparison with existing popular models. Finally, Section 6 summarizes the paper.

2. Framework of Proposed Approach

The framework of the proposed hybrid smoke detection approach is presented as Figure 1. The wildfire smoke as well as water mists and clouds videos are captured through various visual inspection systems. Typical inspection systems take the form of ground-, unmanned aerial vehicle (UAV)-, or satellite-based visual monitoring systems. While UAVs have a flexible monitoring path, satellite and fixed fire monitoring camera in ground-based systems detect the wildland with a certain range. The image frame is then extracted from the captured video at regular time intervals. The time intervals should be well assigned to fully describe the evolution and development of smoke at all stages. Pictures of water mists and clouds are also collected and organized for model establishment and robustness verification. Some image-processing algorithms, i.e., Richardson-Lucy and gamma correction, are exploited to remove camera shake and uneven lighting effects so as to improve the picture quality. In order to avoid model overfitting caused by a small training sample, the collected real pictures are enhanced using some data augmentation methods, including random scaling, flipping and color enhancement.

The ResNet model and SVM were utilized to construct a hybrid smoke detection model from the collected pictures. In this paper, a multi-layered ResNet architecture, i.e., ResNet-50, was applied to extract the dynamic and static characteristics from different depths. The selection of ResNet-50 was motivated by the fact that ResNet can handle gradient vanishing in deep neural networks and an excessively deep network is not necessary. As previously mentioned, the SE attention mechanism was introduced to each down-sampling block of ResNet-50 to evaluate the importance of multiple convolutional features. In order to segment the smoke region in a picture, the fully connected layer in ResNet-50 was replaced with a full convolutional layer to output the class of each pixel. In addition to the convolutional features, a histogram of oriented gradient was also extracted to establish the SVM model, given that HOG is insensitive to light changes and can effectively eliminate interference from back-lighting and tree leaves, which is used for secondary detection.

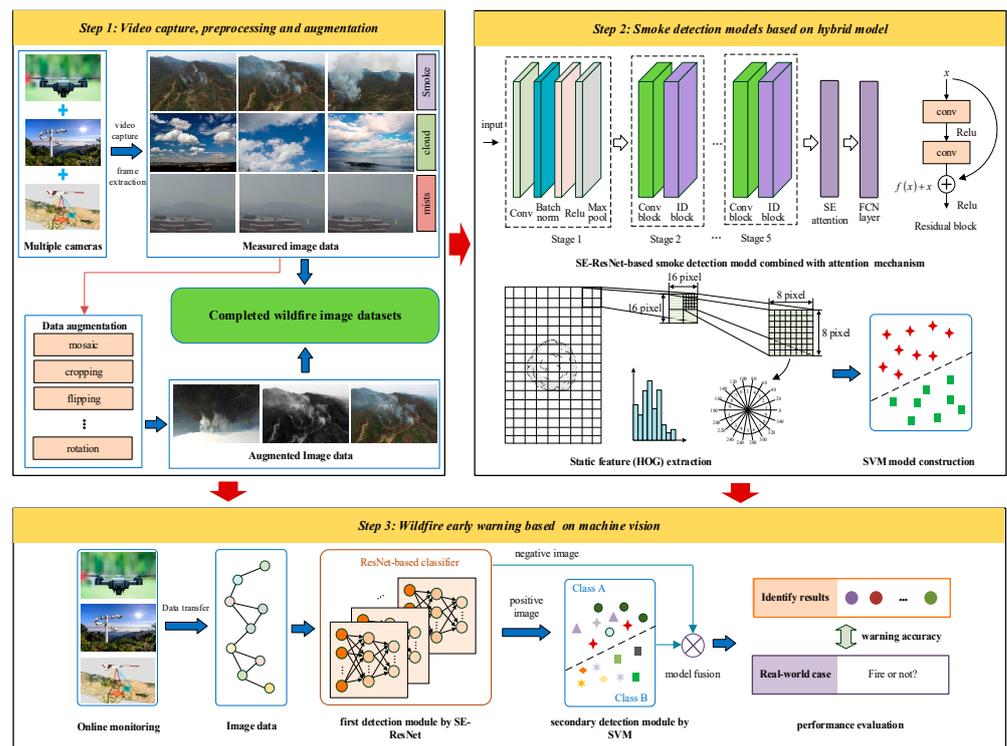


Figure 1. Framework of proposed approach.

In the online monitoring stage, the SE-ResNet and SVM hybrid model are jointly used for smoke detection. Specifically, the SE-ResNet model is firstly called to detect if there is smoke and to segment the smoke region if present. Since the image segmentation-based smoke detection model usually performs well on recall value but has a low score on precision value, a predicted positive image was further inputted to SVM for secondary detection. A fire alarm was reported only if the positive label was output by the hybrid model. The procedure of wildfire smoke detection using the hybrid model is presented as Algorithm 1.

Algorithm 1: Detecting wildfire smoke using a hybrid machine learning model

Input: online wildfire monitoring video

Output: class label of each frame and early warning results

Steps:

1. Extract the image frames according to time intervals t .
 2. For each image frame, call SE-ResNet model for smoke detection.
 3. If a smoke region is labeled, then input this image frame to SVM model for secondary detection; otherwise, return the classification label as non-smoke.
 4. If smoke is detected by SVM, return the classification label as smoke.
 5. Issue the wildfire early warning results according to the steps above.
 6. Repeat until all image frames are detected.
-

3. Wildfire Smoke Segmentation Using SE-ResNet Combined with FCN

In this paper, we exploit ResNet as the backbone of a first-order detector in a hybrid wildfire smoke detection model, given that ResNet can construct a high-depth network architecture to fully integrate the wildfire smoke low/mid/high level image without network degradation and gradient diffusion. ResNet trains a deeper network by adding the idea of residual learning to the traditional convolutional neural network via the use of shortcut connections and by fitting residual representations. The adjacent convolutional layers of Resnet are connected by shortcuts to form residual blocks while multiple residual block structures overlap to compose the residual neural network. The structure of the

residual block is shown in Figure 2, where x denotes the input vector and $F(x)$ represents recast mapping in a residual block.

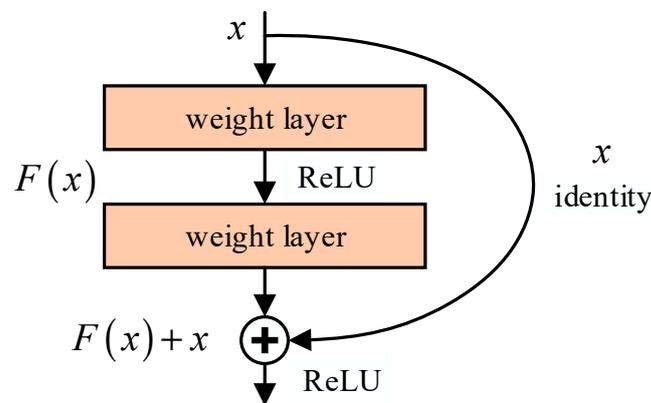


Figure 2. Residual learning in ResNet.

Instead of making a network model that directly fits a desired underlying mapping $H(x)$, a residual block in a ResNet model is used to try to find a residual mapping $F(x) + x$, where $F(x) = H(x) - x$. The recast of underlying mapping is realized via the shortcut connection represented as the curve in Figure 2. Thus, the output of a residual block is given as

$$x_{l+1} = x_l + F(x_l, W_l) \quad (1)$$

A linear mapping is necessary for the cases in which the identity mapping of x in the shortcut connection has different dimensionality with $F(x)$, i.e.,

$$x_{l+1} = W_s x_l + F(x_l, W_l) \quad (2)$$

The output feature X_s in s -th network layer can be deduced recursively from the output of a shallow layer X_l .

$$X_s = X_l + \sum_{i=1}^{s-1} F(x_i, W_i) \quad (3)$$

In the back propagation stage, during model training, the error gradient is derived as

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_s} \frac{\partial x_s}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_s} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{s-1} F(x_i, W_i) \right) \quad (4)$$

The error gradient in Equation (4) contains two multipliers, in which $\frac{\partial \varepsilon}{\partial x_s}$ transmits the training error directly to the shallow convolutional layer without passing through the weight layer, and $\frac{\partial \varepsilon}{\partial x_s} \cdot \frac{\partial}{\partial x_l} \sum_{i=1}^{s-1} F(x_i, W_i)$ updates the convolution kernels of the current weight layer according to the learning error. The special structure in ResNet, i.e., residual block, ensures the training error can skip the intermediate layer and feed back to the shallow convolutional layer, which solves the gradient degradation for a deep network.

A ResNet model is the multi-layer combination of a series of residual blocks. Various classical ResNet models are proposed with different numbers of weighted layers. The network architecture of ResNet-50 is shown as Figure 3. A ResNet-50 has fifty weight layers and consists of one convolutional block, four residual blocks and one fully connected output layer. The convolutional block is composed of a convolutional layer with a filter size of 3×3 , a batch normalization layer, an activation layer (ReLU) and a 3×3 max pooling layer. Residual blocks are the core structures of ResNet-50. The residual blocks in a ResNet-50 model have two different bottlenecks: convolution residual blocks and identity blocks. All convolution residual blocks and identity blocks are composed of three convolutional layers

with filter sizes of 1×1 , 3×3 and 1×1 , respectively, and the difference lies in whether Equation (2) is activated. The input and output features in convolution residual block have different dimensions; therefore, an additional convolution operation is needed in the shortcut path. According to the combination of convolution residual block and identity block, ResNet-50 can be divided into four stages; each stage has one convolution residual block and the number of identity blocks is 2, 3, 5 and 2, respectively. The feature maps extracted from the four stages are then flattened and linked with a fully connected layer to output the class label of the research object.

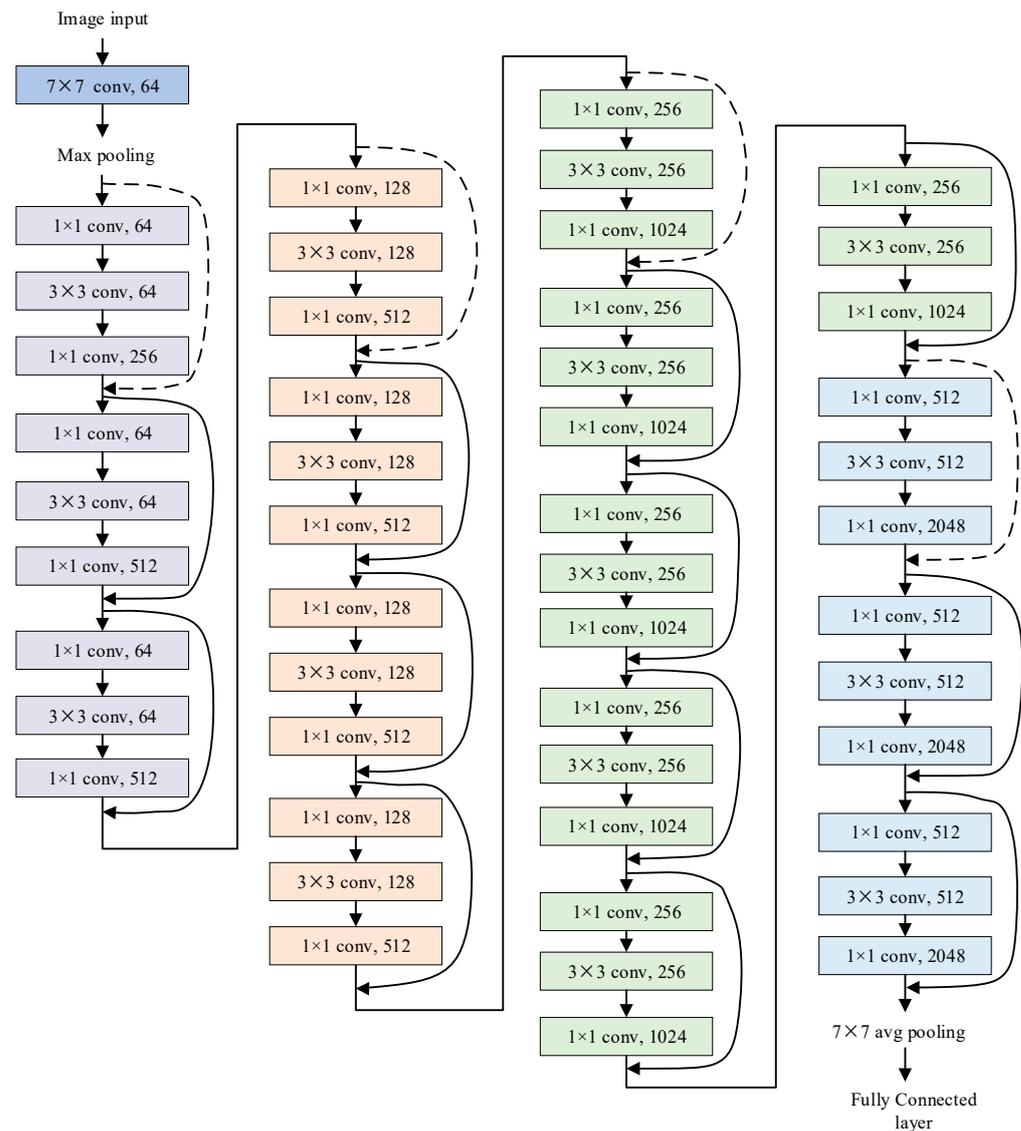


Figure 3. Network architecture of ResNet-50.

The last convolutional layer of ResNet-50 outputs 2048 feature maps to describe the original image. Conventional ResNet-50 uses a fully connected network as an output layer to map the features with class label and all feature maps are assigned with the same weight, which may cause the vital features to be drowned out by the more inessential ones. This paper introduces the SE channel attention mechanism to ResNet-50 down-sampling blocks to weight the feature maps of each stage. The multi-scale feature pyramid network of the wildfire smoke detection model after introducing the SE block is presented as Figure 4. The SE block consists of a squeeze operation and an excitation operation, summarizing the overall information about each feature map and scaling the importance of each feature

map. Convolutional layers in ResNet-50 learn the local spatial connection patterns in the corresponding receptive field using multiple filters. The local spatial connection patterns are then weighted and scaled to the input size via the SE block. The semantic features at all scales are used to describe the input image and determine the class label, which therefore is more accurate compared with conventional ResNet-50. Meanwhile, the output layer of ResNet-50 is replaced by a fully convolutional network to obtain the class label of every pixel for smoke segmentation. All pixels are identified as smoke or non-smoke objects according to the extracted image features, and the whole image is consequently classified if the smoke pixels exceed a threshold.

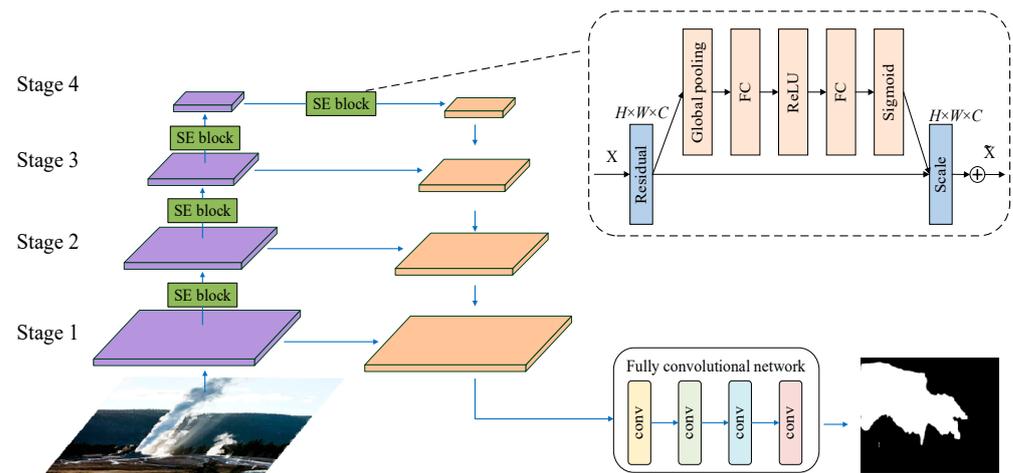


Figure 4. Feature pyramid network for wildfire smoke detection.

4. Secondary Smoke Detection Based on SVM Model

4.1. HOG for Wildfire Smoke Description

Wildfire smoke normally has clear edges in both vertical and horizontal directions, which provides a clue for hand-crafted features extraction. A set of features may be considered as various options to describe smoke image edges. However, two main problems should be taken into consideration. Some visual inspection systems, especially in terms of unmanned aerial vehicles, have a flexible monitoring path and variable shooting angle and distance, which lead to various image distortions, such as barrel, pincushion and wave distortions, to name a few. Although some image data-processing methods can be used to correct these distortions, the smoke edge features may be masked during the processing. Further, the uneven and dim lighting conditions also bring considerable difficulties for wildfire smoke edge description. Given the situations above, the HOG, which is widely used in pedestrian identification, is introduced to depict the wildfire smoke.

A histogram of oriented gradients is a classical feature descriptor that computes and counts histograms of gradient directions in local regions of an image to represent image features for computer vision-based object identification. The HOG describes the appearance and edges of objects in an image through the intensity distribution of gradient and contours direction, which therefore can effectively minimize the effects of translation, rotation and lighting variations. The procedure of the HOG algorithm is shown in Figure 5. The digital image is firstly transformed into a grey-scale map, and the horizontal and vertical gradients G_x , G_y of each pixel $f(x, y)$ are then calculated via partial derivative as

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \quad (5)$$

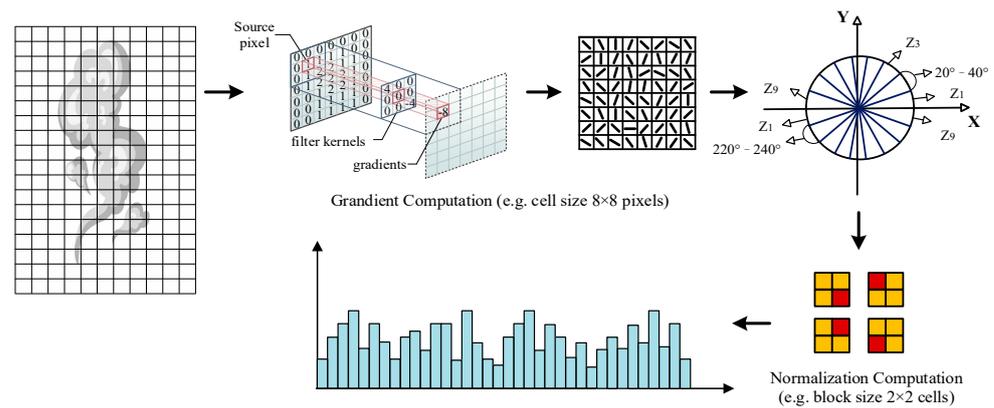


Figure 5. Procedure for HOG features vector extraction.

Let the first difference be calculated using the first difference and the filter kernels of $[-1, 0, 1]$ and $[-1, 0, 1]^T$, respectively, which can be written as a more computable form

$$\begin{cases} G_x = f(x + 1, y) - f(x - 1, y) \\ G_y = f(x, y + 1) - f(x, y - 1) \end{cases} \quad (6)$$

The gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ are calculated as

$$\begin{cases} m(x, y) = \sqrt{G_x^2 + G_y^2} \\ \theta(x, y) = \arctan\left(\frac{G_y}{G_x}\right) \end{cases} \quad (7)$$

The orientation histograms, consisting of several bins with even intervals within $0 - 180^\circ$, are used to describe the gradient distribution in a local image region called cells. A cell is a square area in an image with the same pixels in the x -axis and y -axis. The gradient orientation $\theta(x, y)$ of each pixel determines the bins this pixel contributes, and the weight is decided via gradient magnitude $m(x, y)$. Considering a pixel belonging to bin t , the contributions of the pixel to v_t and $v_{t\pm 1}$ are calculated via

$$v_t = (1 - \alpha)m(x, y), \quad v_{t\pm 1} = \alpha \cdot m(x, y) \quad (8)$$

where α can be calculated via

$$\alpha = (t + 0.5) - \frac{n \cdot \theta(x, y)}{\pi} \quad (9)$$

where n denotes the number of bins.

The score of each bin is organized in series to form a cell vector. To compensate for the changes in lighting, the normalization of each orientation histogram is applied on a certain area in an image called a block. The block is made of several cells with same size in the x -axis and y -axis. Connect the cell vectors in series to form a block vector, and suppose the orientation histograms are normalized using the L2-norm; then the normalized value of the block vector is

$$v'_i = \frac{v_i}{\sqrt{\|v\|_2^2 + \varepsilon^2}} \quad (10)$$

where i is the i -th element in a block vector; v_i and v'_i denote the i -th element before and after normalization; ε represents a small constant to avoid division by zero; and $\|v\|_2$ is the L2-norm of a block vector which is calculated as Equation (11), in which s denotes the length of the block vector.

$$\|v\|_2 = v_1^2 + v_2^2 + \dots + v_s^2 \quad (11)$$

The block vectors are then combined in the manner above to construct an HOG feature vector of a whole image. The feature vector is a row vector which contains edge information for all objects in an image. A classifier can be established accordingly for wildfire smoke detection via learning the feature vectors of the training image datasets.

4.2. SVM Model for Smoke Detection

In this section, we make use of the support vector machine (SVM) to construct a wildfire smoke classifier from the feature vectors extracted above. The support vector machine is a supervised machine learning algorithm that realizes binary classification by finding the optimal hyperplane to maximize the margin between different classes. Letting $\mathbf{x}_i \in \mathbf{R}^n$ ($i = 1, \dots, M$) be the m -dimensional inputs, and $y_i \in \{+1, -1\}$ be the corresponding binary outputs, the training data can be represented as pairs of input–output pairs, i.e., $T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_k, y_k)\}$. As for the task of wildfire smoke detection, input $\mathbf{x} \in \mathbf{R}^n$ represents the extracted feature vector in Section 4.1, where n is the vector dimension; binary output y denotes the labels of the image, i.e., smoke and non-smoke; T depicts the training datasets derived from the historical monitoring images.

The mapping function between \mathbf{x} and y is defined as

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (12)$$

where \mathbf{w} is a normal vector with dimension of m , which separates the two classes on both sides, and b represents bias.

There exists an infinite hyperplane when the training datasets are linearly separable. The SVM maximally distinguishes pairs of classes by finding the unique optimal hyperplane $f(\mathbf{x}) = 0$, that is

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq 1, & \forall y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1, & \forall y_i = -1 \end{cases} \quad (13)$$

The optimal problem can be reformulated as Equation (14) after introducing a regularization term $\|\mathbf{w}\|^2/2$.

$$\begin{cases} \min_{\mathbf{w}, b} \frac{\|\mathbf{w}\|^2}{2} \\ s.t. y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, k \end{cases} \quad (14)$$

Equation (14) is essentially a convex quadratic programming problem; the optimal hyperplane is determined via the solution \mathbf{w}^* and b^* . The Lagrangian function can be introduced to search the optimum. After considering the Karush-Kuhn-Tucker conditions, the optimal solutions of \mathbf{w}^* and b^* can be derived as

$$\begin{cases} \mathbf{w}^* = \sum_{i=1}^k \lambda_i^* y_i \mathbf{x}_i \\ b^* = y_j - \sum_{i=1}^k \lambda_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}_j) \end{cases} \quad (15)$$

where λ_i^* is the Lagrangian multiplier reduced via the training data; $\mathbf{x}_i \cdot \mathbf{x}_j$ represents the inner product operator.

The decision function is therefore given by

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^* \cdot \mathbf{x} + b^*) \quad (16)$$

The conventional SVM is a linear classifier. In order to deal with nonlinear classification, kernel transformation is introduced to map inputs from the linear space to the feature

space, which extends SVM to high-order nonlinear problems. Let $\phi(\mathbf{x})$ be a mapping from input space X to feature space H , for $\forall \mathbf{x}, \mathbf{z} \in \mathbf{R}^n$, with the kernel function $K(\mathbf{x}, \mathbf{z})$ defined as

$$K(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) \quad (17)$$

The explicit expression of $\phi(\mathbf{x})$ is normally unnecessary. Replace the inner product $\mathbf{x}_i \cdot \mathbf{x}_j$ with the kernel function $K(\mathbf{x}, \mathbf{z})$, and the nonlinear support vector machine can be written as

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^k \lambda_i^* y_i K(\mathbf{x}, \mathbf{x}_i) + b^* \right) \quad (18)$$

Kernel function $K(\mathbf{x}, \mathbf{z})$ usually takes different forms, typically Gaussian, radial basis function, Sigmoid and polynomial. Radial basis function can map the original data space to an infinite dimensional feature space, making it more widely used in practical applications. Supposing the radial basis kernel function is used in Equation (17), the decision function can be written as

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^k \lambda_i^* y_i \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) + b^* \right) \quad (19)$$

where γ represents an adjustable constant term, where a small γ maps the data into a low-dimensional space, whereas a large γ corresponds to a high-dimensional space. The preferred value for the general classification problem is 0.1.

5. Case Study

5.1. Datasets

Currently, the wildfire smoke detection community has no standard database with appropriate scale, rich content, and unified format which is authoritative enough to evaluate any smoke detection model. Researchers usually construct their respective datasets from web or fire agent surveillance video, or from burning tests in the laboratory, to illustrate their model's performance. Some fire smoke datasets have been proved to be well-constructed for model validation in various applications and are open access. The datasets used in our research were sourced from the State Key Laboratory of Fire Science at the University of Science and Technology of China (USTC) and have been publicly released for research usage (<http://smoke.ustc.edu.cn>, accessed on 22 October 2023.). As for wildfire smoke images, unmanned aerial vehicles configured with high-definition cameras are used to capture real videos of wildfires under various scenarios, and the images are then extracted from the videos every 5 to 25 frames and unified to a size of 960×540 . In order to extend the image samples' number and diversity, some synthetic image technologies are used to generate certain smoke images under given scenarios and combustion conditions.

Image data in the sub-datasets Video Smoke Detection Base on Deep Saliency Network and Smoke Detection Based on Scene Parsing and Saliency are used to demonstrate our hybrid model. The first sub-dataset contains 5700 images, where the training set has 1401 smoke and 1499 non-smoke images, while the test dataset consists of 1399 smoke and 1401 non-smoke images. The second sub-datasets has 4695 images, consisting of 2695 images for training and 2000 images for testing. Since our research focuses on wildfire smoke detection, the smoke images of building fires, indoor fires, car fires and indoor simulated combustion testing were removed from the datasets. In order to fully verify the proposed hybrid model on smoke and smoke-like objects identification, a substantial number of water mist, cloud and dust images were selected both from the USTC datasets and from other open fire detection datasets as interference information. Specifically, the datasets in our research contain six classes of images, i.e., wildfire smoke, water mist, cloud, dust, wildland without fire and background scenery. Some sample images of the rebuilt datasets in our research are shown in Figure 6. The datasets were randomly separated into a training set and testing set, both for SE-ResNet and SVM model construction.



Figure 6. Image samples: (a) real wildfire smoke; (b) water mists; (c) clouds.

5.2. Results

Some online data augmentation technologies were firstly applied in this paper to avoid the overfitting of the SE-ResNet-based smoke detection model and improve the generalization ability. The activation probabilities are presented as Table 1. As for random cropping, we randomly generated an aspect ratio within a certain range, and set the clipping area accordingly in the captured picture. The cropped pictures were then scaled up to the normal size after interpolation. Given that the lighting conditions in wildland are usually complex and varied, the color and brightness of training pictures in this paper were also randomly changed with a probability of 0.5 to reduce the sensitivity of the SE-ResNet network to color and brightness. In order to ensure a smoke detection model with translation and rotation invariance, flipping and random rotation were introduced in the data augmentation step with the probabilities of 0.2 and 0.3, respectively. The lost pixels during these operations were filled with zeroes to maintain an unchanged picture size.

Table 1. Activation probabilities of data augmentation operations.

Data Augmentation Operations	Horizontal Flip	Vertical Flip	Cropping	Color Dithering	Random Rotation	Gaussian Noise
Activation probabilities	0.2	0.2	0.3	0.5	0.3	0.4

The HOG feature extraction corresponding parameters in the SVM model and the hyper parameters of SE-ResNet were set as in Table 2. Since the HOG features of wildfire smoke images are usually non-linear and extremely large in quantity, the RBF kernel function was used in SVM construction to search the optimal hyperplane under the curse of dimensionality. As for the SE-ResNet model, we made use of the cross-entropy loss function to guide the deep model learning process since wildfire smoke detection is essentially a binary classification problem. Given that the quantity of smoke pictures is relatively smaller than the quantity of non-smoke pictures, and given that this class imbalance may lead to a small contribution to the loss function value in the case of a positive label, this paper introduced a weight variable to adjust the loss function value of sample learning. The weighted cross-entropy loss function is presented as Equation (20).

$$Loss = -\beta y \log \hat{y} - (1 - y) \log(1 - \hat{y}) \quad (20)$$

where β represents the weight of the positive label; y depicts the sample label; and \hat{y} represents the output value of the positive sample.

Table 2. Parameters of HOG extraction and SE-ResNet model.

HOG Parameter	Value	Hyper Parameters	Value
gradient filter	sobel	initial learning rate	0.01
cell size	8×8	epochs	50
block size	16×16	batch size	4
bin number	9	momentum	0.9
normalization	L2-Norm		

A momentum-based gradient descent algorithm was used to improve possible oscillation phenomena during SE-ResNet model parameter updates and the momentum was assigned as 0.9. The testing datasets were enhanced using the data augmentation tricks mentioned above to validate the effectiveness and robustness of the proposed approach. All testing pictures were classified into smoke and non-smoke and the smoke detection results were evaluated via accuracy, precision, recall and F1-score, which are derived from the confusion matrix. The indices are calculated as Equations (21)–(24).

$$Accuracy = \frac{TP + TN}{N} \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

$$F1 - score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \quad (24)$$

where TP represents the number of positive samples correctly classified; TN means the correctly classified negative samples; FP describes the negative samples incorrectly marked as positive; FN depicts the positive samples incorrectly marked as negative; and N represents the total number of testing samples.

The smoke detection results of the proposed hybrid model are shown in Table 3 row two. In order to illustrate the advantages of the hybrid model, the performance of classical deep learning architectures on the testing datasets are also presented in this paper and the key indices are listed in Table 3. All calculations were performed on a computer workstation with Ubuntu18.04, GeForce RTX 2080 Ti, CUDA 10.2 and GPU driver of NVIDIA 450.36.06.

Table 3. Performance of hybrid model and classical deep learning architectures.

Model	Accuracy	Precision	F1-Score	Running Time
Hybrid Model	98.99%	98.04%	0.99	3.2779
AlexNet	87.88%	80.65%	0.89	0.0997
VGG-16	90.91%	85.96%	0.92	0.3132
GoogleNet	94.95%	90.91%	0.95	0.1515
SE-ResNet	87.88%	80.65%	0.89	2.1495
SVM	90.91%	84.75%	0.92	1.1480

Conventional deep learning methods exhibit unsatisfactory performance and relatively higher false alarm rates on wildfire smoke detection under the interference of smoke-like objects, where the largest false warning rate was 19.35% as for the AlexNet-based model, and even for the GoogleNet model, which is thought of as one of the most promising deep learning architectures, the false warning rate was still 9.09%. This high false warning rate clearly prohibits the computer vision-based wildfire early warning approach from being widely used in the real world. On the contrary, the hybrid model shows a good performance in natural environment wildfire smoke detection even with the interference of clouds and water mist, with an accuracy, precision and F1-score of 98.99%, 98.04% and 0.99, respectively. Compared with AlexNet, VGG-16 and GoogleNet, the accuracy of the hybrid

model increased by 11.11%, 8.08% and 4.04%, respectively. Further analyzing the model performance by combining multiple indices, it can be noted that the improvements of the hybrid model mainly contributed to the superior performance on the index of precision, which was increased by 17.39% compared to the AlexNet-based model. This improvement shows that the proposed hybrid model has favorable performance on correctly identifying the wildfire smoke against smoke-like objects interference by making use of smoke static and dynamic features, which obviously improves the application potential of computer vision-based wildfire early warning in practice. In addition, column 5 in Table 3 presents the running times for each detection. The AlexNet model has the fastest monitoring speed (i.e., 0.0997 s for each picture) due to the simple model architecture. On the other hand, the proposed hybrid model has two classifiers, i.e., SE-ResNet as a primary classifier and SVM as a secondary one, so it takes more time for wildfire detection compared with other models, i.e., 3.2779 s for each picture. Although the model takes more time, it is acceptable in real-world practice.

In order to illustrate the necessity and advantages of combining the SE-ResNet and SVM models, we also tested the SE-ResNet and SVM models separately on the USTC datasets and the results are presented in Table 3, row 6 and row 7. The performances of both the SE-ResNet and SVM models were unsatisfactory, especially for SE-ResNet with an accuracy of 87.88%. This is because the SE-ResNet identifies the whole image based on the class label of every pixel. Figure 7 shows the image segmentation results on wildfire smoke and water mists and clouds. It can be noted that although the SE-ResNet model can accurately identify and locate smoke regions in a whole image, some pixels of water mists and clouds may be mistakenly recognized as smoke due to the similar image features and therefore causes a false warning. We could assign a relatively high threshold to avoid the false warning; this strategy, however, would make smoke from a small fire unrecognizable, which thus leads to a time delay in reporting fire incidents and deviates from the original purpose of wildfire early warning. The SVM was used as secondary detector in the hybrid model, which avoids the false warning given by SE-ResNet and comprehensively improves the performance for smoke detection and anti-interference capacity.



Figure 7. Image segmentation using SE-ResNet: (a) wildfire smoke and (b) false segmentation of water mists and clouds.

5.3. Discussion

We propose a novel hybrid wildfire smoke detection approach combining deep learning architectures and conventional machine learning, i.e., SE-ResNet and SVM. The hybrid model shows excellent performance on the datasets. However, this experiment still has potential drawbacks or areas for further improvement, mainly focusing on the accuracy and speed of the detection model. The potential future improvements can be listed as follows.

- (1) The hybrid model will have more accuracy if a self-attention mechanism is added, which can enable the model to capture dependencies at a distance in an image by assigning different weights to features.
- (2) Wildfire detection requires smaller models with faster monitoring for real-time monitoring. Methods such as quantization, weight sharing, knowledge distillation, etc., can reduce the size of the model to increase the speed of the model. This is vital for wildfire detection.
- (3) While some datasets are well-constructed and publicly available, they may lack the scale, richness, and unified format required for comprehensive model evaluation. This lack of standardized datasets can limit the comparability and reliability of different detection models' performance evaluations. It is necessary to establish a standard database for smoke detection models.

6. Conclusions

Wildfire is one of the most serious worldwide natural disasters which leads to huge life and economy loss every year. Computer vision technology, with its characteristics of non-contact and rapid response, facilitates the real-time accurate early warning of wildfire. The primary purpose of this paper was to develop a practice-oriented intelligent identification approach for wildfire smoke and smoke-like objects, to decrease the false warning rate in real-world practice. For this purpose, we investigated the characteristics of deep learning- and conventional machine learning-based wildfire early warning approaches, and proposed a hybrid wildfire smoke detection methodology by combining ResNet with an SVM model. The ResNet model was improved by an SE block and fully convolutional network to quantify the relationships between multiple extracted features. The advantages of the hybrid model are illustrated via open wildfire datasets, for which the accuracy was improved by 11.11%, 8.08% and 4.04%, respectively, compared with Alexnet, VGG-16 and GoogleNet. The comparisons show the proposed approach can effectively address the low accuracy of existing models in recognizing smoke-like objects, i.e., cloud and water mists, which presents the practicability in real-world cases.

Furthermore, the accuracy of the hybrid model was increased to 98.99% from 87.88% and 90.91% compared with the single models of SE-ResNet-50 and SVM. The improvements reveal that the hybrid model comprehensively captures the dynamic and static features of wildfire smoke, avoiding the limitations of incomplete feature extraction in the case of single deep learning models.

Author Contributions: Conceptualization, X.W. and J.W.; methodology, X.W.; software, X.W. and Y.Z.; validation, X.W., L.C. and Y.Z.; formal analysis, X.W. and L.C.; investigation, X.W. and J.W.; resources, Y.Z.; data curation, X.W. and L.C.; writing—original draft preparation, X.W.; writing—review and editing, X.W. and J.W.; visualization, X.W.; supervision, J.W.; project administration, J.W. and Y.Z.; funding acquisition, J.W. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Jiangsu Province, China (Grant No.: BK20221117), the Basic Research Project of Xuzhou City, China (Grant No.: KC22001), and The Jiangsu Funding program for Excellent Postdoctoral Talent (Grant No.: 2023ZB386).

Data Availability Statement: The datasets used in this research can be found at <http://smoke.ustc.edu.cn>, accessed on 22 October 2023.

Conflicts of Interest: Author Xin Wang was employed by the Xuzhou Fu'an Technology Co., Ltd., author Linlin Chen was employed by the Gengcun Coal Mine, Henan Dayou Energy Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chaturvedi, S.; Khanna, P.; Ojha, A. A survey on vision-based outdoor smoke detection techniques for environmental safety. *ISPRS J. Photogramm.* **2022**, *185*, 158–187. [[CrossRef](#)]
2. Bouguettaya, A.; Zarzour, H.; Taberkit, A.M.; Kechid, A. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Process.* **2022**, *190*, 108309. [[CrossRef](#)]
3. Boylan, J.L.; Lawrence, C. The development and validation of the bushfire psychological preparedness scale. *Int. J. Disast. Risk Re.* **2020**, *47*, 101530. [[CrossRef](#)]
4. Oliver, J.A.; Pivot, F.C.; Tan, Q.; Cantin, A.S.; Wooster, M.J.; Johnston, J.M. A machine learning approach to waterbody segmentation in thermal infrared imagery in support of tactical wildfire mapping. *Remote Sens.* **2022**, *14*, 2262. [[CrossRef](#)]
5. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors* **2022**, *22*, 1977. [[CrossRef](#)] [[PubMed](#)]
6. Toulouse, T.; Rossi, L.; Campana, A.; Celik, T.; Akhloufi, M.A. Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Saf. J.* **2017**, *92*, 188–194. [[CrossRef](#)]
7. Zhao, Y.Q.; Tang, G.Z.; Xu, M.M. Hierarchical detection of wildfire flame video from pixel level to semantic level. *Expert Syst. Appl.* **2015**, *42*, 4097–4104. [[CrossRef](#)]
8. Ko, B.; Park, J.; Nam, J.Y. Spatiotemporal bag-of-features for early wildfire smoke detection. *Image Vis. Comput.* **2013**, *31*, 786–795. [[CrossRef](#)]
9. Almeida, J.S.; Huang, C.X.; Nogueira, F.G.; Bhatia, S.; De Albuquerque, V.H.C. EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire-Smoke Detection. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7889–7898. [[CrossRef](#)]
10. Pundir, A.S.; Raman, B. Deep belief network for smoke detection. *Fire Technol.* **2017**, *53*, 1943–1960. [[CrossRef](#)]
11. Jakovcevic, T.; Stipanicev, D.; Krstinic, D. Visual spatial-context based wildfire smoke sensor. *Mach. Vis. Appl.* **2013**, *24*, 707–719. [[CrossRef](#)]
12. Pundir, A.S.; Raman, B. Dual deep learning model for image based smoke detection. *Fire Technol.* **2019**, *55*, 2419–2442. [[CrossRef](#)]
13. De Venâncio, P.V.A.B.; Campos, R.J.; Rezende, T.M.; Lisboa, A.C.; Barbosa, A.V. A hybrid method for fire detection based on spatial and temporal patterns. *Neural Comput. Appl.* **2023**, *35*, 9349–9361. [[CrossRef](#)]
14. Prema, C.E.; Vinsley, S.S.; Suresh, S. Multi feature analysis of smoke in YUV color space for early forest fire detection. *Fire Technol.* **2016**, *52*, 1319–1342. [[CrossRef](#)]
15. Peng, Y.S.; Wang, Y. Real-time forest smoke detection using hand-designed features and deep learning. *Comput. Electron. Agric.* **2019**, *167*, 105029. [[CrossRef](#)]
16. Luo, L.; Yan, C.W.; Wu, K.L.; Zheng, J.Y. Smoke detection based on condensed image. *Fire Saf. J.* **2015**, *75*, 23–35. [[CrossRef](#)]
17. Wang, Y.; Li, M.; Zhang, C.X.; Chen, H.; Lu, Y.M. Weighted-fusion feature of MB-LBPUH and HOG for facial expression recognition. *Soft Comput.* **2020**, *24*, 5859–5875. [[CrossRef](#)]
18. Ko, B.; Kwak, J.Y.; Nam, J.Y. Wildfire smoke detection using temporospatial features and random forest classifiers. *Opt. Eng.* **2012**, *51*, 017208. [[CrossRef](#)]
19. Li, X.Q.; Chen, Z.X.; Wu, Q.M.J.; Liu, C.Y. 3D parallel fully convolutional networks for real-time video wildfire smoke detection. *IEEE Trans. Circ. Syst. Vid.* **2020**, *30*, 89–103. [[CrossRef](#)]
20. Li, J.Y.; Zhou, G.X.; Chen, A.B.; Lu, C.; Li, L.J. BCMNet: Cross-layer extraction structure and multiscale downsampling network with bidirectional transpose FPN for fast detection of wildfire smoke. *IEEE Syst. J.* **2023**, *17*, 1235–1246. [[CrossRef](#)]
21. Wang, X.T.; Pan, Z.J.; Gao, H.; He, N.X.; Gao, T.G. An efficient model for real-time wildfire detection in complex scenarios based on multi-head attention mechanism. *J. Real Time Image Process* **2023**, *20*, 66. [[CrossRef](#)]
22. Labati, R.D.; Genovese, A.; Piuri, V.; Scotti, F. Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation. *IEEE Trans. Syst. Man Cybern. Syst.* **2013**, *43*, 1003–1012. [[CrossRef](#)]
23. Gunay, O.; Toreyin, B.U.; Kose, K.; Cetin, A.E. Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video. *IEEE Trans. Image Process.* **2012**, *21*, 2853–2865. [[CrossRef](#)] [[PubMed](#)]
24. Bugaric, M.; Jakovcevic, T.; Stipanicev, D. Adaptive estimation of visual smoke detection parameters based on spatial data and fire risk index. *Comput. Vis. Image Underst.* **2014**, *118*, 184–196. [[CrossRef](#)]
25. Fernandes, A.M.; Utkin, A.B.; Chaves, P. Automatic early detection of wildfire smoke with visible-light cameras and EfficientDet. *J. Fire Sci.* **2023**, *41*, 122–135. [[CrossRef](#)]
26. Yin, Z.J.; Wan, B.Y.; Yuan, F.N.; Xia, X.; Shi, J.T. A deep normalization and convolutional neural network for image smoke detection. *IEEE Access* **2017**, *5*, 18429–18438. [[CrossRef](#)]
27. Yuan, F.N.; Zhang, L.; Wan, B.Y.; Xia, X.; Shi, J.T. Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition. *Mach. Vis. Appl.* **2019**, *30*, 345–358. [[CrossRef](#)]
28. Sun, X.F.; Sun, L.P.; Huang, Y.L. Forest fire smoke recognition based on convolutional neural network. *J. For. Res.* **2021**, *32*, 1921–1927. [[CrossRef](#)]
29. Celik, T.; Demirel, H. Fire detection in video sequences using a generic color model. *Fire Saf. J.* **2009**, *44*, 147–158. [[CrossRef](#)]
30. Yuan, F.N.; Zhang, L.; Xia, X.; Huang, Q.H.; Li, X.L. A gated recurrent network with dual classification assistance for smoke semantic segmentation. *IEEE Trans. Image Process* **2021**, *30*, 4409–4422. [[CrossRef](#)]
31. Yuan, F.N.; Zhang, L.; Xia, X.; Huang, Q.H.; Li, X.L. A wave-shaped deep neural network for smoke density estimation. *IEEE Trans. Image Process* **2020**, *29*, 2301–2313. [[CrossRef](#)] [[PubMed](#)]

32. Yuan, F.N. A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with Adaboost for video smoke detection. *Pattern Recogn.* **2012**, *45*, 4326–4336. [[CrossRef](#)]
33. Xia, X.; Yuan, F.N.; Zhang, L.; Yang, L.Z.; Shi, J.T. From traditional methods to deep ones: Review of visual smoke recognition, detection, and segmentation. *J. Image Graph.* **2019**, *24*, 1627–1647. [[CrossRef](#)]
34. Zhao, Y.Q.; Li, Q.J.; Gu, Z. Early smoke detection of forest fire video using CS Adaboost algorithm. *Optik* **2015**, *126*, 2121–2124. [[CrossRef](#)]
35. Calderara, S.; Piccinini, P.; Cucchiara, R. Vision based smoke detection system using image energy and color information. *Mach. Vis. Appl.* **2011**, *22*, 705–719. [[CrossRef](#)]
36. Zhang, X.; Xie, J.B.; Yan, W.; Zhong, Q.Y.; Liu, T. An Algorithm for smoke ROF Detection Based on Surveillance Video. *J. Circuit. Syst. Comp.* **2013**, *22*, 1350010. [[CrossRef](#)]
37. Zhang, F.; Qin, W.; Liu, Y.B.; Xiao, Z.T.; Liu, J.X.; Wang, Q.; Liu, K.H. A dual-channel convolution neural network for image smoke detection. *Multimed. Tools Appl.* **2020**, *79*, 34587–34603. [[CrossRef](#)]
38. Qiang, X.H.; Zhou, G.X.; Chen, A.B.; Zhang, X.; Zhang, W.Z. Forest fire smoke detection under complex backgrounds using TRPCA and TSVB. *Int. J. Wildland Fire* **2021**, *30*, 329–350. [[CrossRef](#)]
39. Khan, S.; Muhammad, K.; Mumtaz, S.; Baik, S.W.; De Albuquerque, V.H.C. Energy-efficient deep CNN for smoke detection in foggy IoT environment. *IEEE Internet Things* **2019**, *6*, 9237–9245. [[CrossRef](#)]
40. Muhammad, K.; Khan, S.; Palade, V.; Mehmood, I.; De Albuquerque, V.H.C. Edge intelligence-assisted smoke detection in foggy surveillance environments. *IEEE Trans. Ind. Inform.* **2020**, *16*, 1067–1075. [[CrossRef](#)]
41. Li, T.T.; Zhao, E.T.; Zhang, J.G.; Hu, C.H. Detection of wildfire smoke images based on a densely dilated convolutional network. *Electronics* **2019**, *8*, 1131. [[CrossRef](#)]
42. Cheng, G.T.; Chen, X.; Gong, J.C. Deep convolutional network with pixel-aware attention for smoke recognition. *Fire Technol.* **2022**, *58*, 1839–1862. [[CrossRef](#)]
43. Wang, Z.L.; Zhang, T.H.; Wu, X.Q.; Huang, X.Y. Predicting transient building fire based on external smoke images and deep learning. *J. Build. Eng.* **2022**, *47*, 103823. [[CrossRef](#)]
44. Xu, G.; Zhang, Y.M.; Zhang, Q.X.; Lin, G.H.; Wang, J.J. Deep domain adaptation based video smoke detection using synthetic smoke images. *Fire Saf. J.* **2017**, *93*, 53–59. [[CrossRef](#)]
45. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. Available online: <https://arxiv.org/pdf/1512.03385> (accessed on 22 October 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.