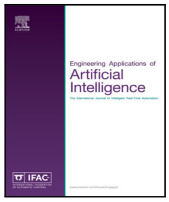
Engineering Applications of Artificial Intelligence 110 (2022) 104737

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Fire detection in video surveillances using convolutional neural networks and wavelet transform

Lida Huang [a], Gang Liu [a], Yan Wang [b], Hongyong Yuan [a], Tao Chen [a,*]

[a] Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing, 100084, China
[b] Institute for AI Industry Research, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Fire detection
Computer vision
Convolutional neural networks
Wavelet analysis

ABSTRACT

Fire is one of the most frequent and common emergencies threatening public safety and social development. Recently, intelligent fire detection technologies represented by convolutional neural networks (CNNs) have been widely concerned by academia and industry, substantially improving detection accuracy. However, CNN-based fire detection systems are still subject to the interference of false alarms and the limitation of computing power. In this paper, taking advantage of traditional spectral analysis in fire image detection technology, a novel Wavelet-CNN method is proposed, which applies the 2D Haar transform to extract spectral features of the image and input them into CNNs at different layer stages. Two classic backbone networks, ResNet50 and MobileNet v2 (MV2) are used to test our method, and experimental results on a benchmark fire dataset and a video dataset show that the method improves fire detection accuracy and reduces false alarms, especially for the light-weight MV2. Despite the low computational needs, the Wavelet-MV2 achieves accuracy that is comparable to state-of-the-art methods.

1. Introduction

Fire disasters often endanger the safety of human life and property. To minimize fire losses, effective fire detection at the early stage and an autonomous response are important and helpful. In ordinary buildings, detectors based on physical signals, like smoke sensors, heat-release infrared flame detectors, ultraviolet flame detectors, and so on, are widely used for fire alarms. However, these traditional physical sensors require proximity to fire sources so that they cannot work in large space buildings and open spaces like plants and ports, and they fail to provide fire details such as the fire location, size, and the degree of burning (Frizzi et al., 2016). To get over such limitations, fire detection systems based on visual sensors have been presented.

Visual fire detection systems have the following advantages: (1) low cost relying on more and more existing surveillance cameras, (2) large monitoring regions, (3) comparatively fast response time without waiting for fire diffusion, (4) fire confirmation without visiting the fire site, and (5) the availability of fire details. Thus, visual fire detection methods have attracted particular attention during the last decade. Traditional visual fire detection methods use hand-crafted features, such as color, texture, shape, edge, and motion. Color information is a key factor in fire detection, and existing mature color models include RGB, HIS, and YCbCr. Chen et al. (2004) proposed a method using the red channel threshold. binti Zaidi et al. (2015) performed fire detection based on RGB and YCbCr features. Li et al. (2018) proposed a fire detection framework based on the color, dynamics, and flickering properties of flames. Schultze et al. (2006) proposed to obtain flame features using spectrograms and sonograms based on the characteristic that flames flicker and move upwards for detection. Töreyin et al. (2006) and Toreyin and Cetin (2007) represented the boundary of flames in the wavelet domain and used the high-frequency natures of the boundaries of fire regions to model flame flicker. Foggia et al. (2015) used an expert system to build a rule set based on fire color, shape, and motion features. Dimitropoulos et al. (2015) built an SVM classifier for fire detection based on motion, texture, flicker, and color probability features. Wang et al. (2019) extracted multiple features for forest fire recognition, including color, texture, area, and shape features. The above researchers built their extractors to improve the accuracy of fire detection. Such hand-crafted features have promoted the development of visual fire detection. However, because of the high complexity of fire scenes in videos, artificially designed features are highly redundant.

Recently, to achieve higher efficiency and better generalization ability, some convolutional neural network (CNN) based visual fire classification and detection methods have been proposed. Muhammad et al. used the pre-trained CNNs like AlexNet (Muhammad et al., 2018a), GoogleNet (Muhammad et al., 2018b), and MobileNetV2 (Muhammad et al., 2019b) as baseline architectures and fine-tuned the fully-connected layers on a small fire dataset, achieving a performance boost.

* Corresponding author.
E-mail address: chentao.b@tsinghua.edu.cn (T. Chen).

https://doi.org/10.1016/j.engappai.2022.104737
Received 16 August 2021; Received in revised form 25 January 2022; Accepted 2 February 2022
Available online 19 February 2022

Frizzi et al. (2016) proposed a nine-layer CNN for identifying whether there is a fire or not. With the rise of object detection approaches, fire detection is no longer satisfied with determining fire, but rather locating and extracting exact areas. Several variations of generic object detection methods have been proposed for fire detection tasks. Sharma et al. (2017) combined the pre-trained VGG16 and ResNet50 to develop a fire detection system. Wu and Zhang (2018) tested Faster R-CNN, YOLO, and SSD for fire detection, and adjusted the YOLO's tiny-YOLO-VOC structure to improve accuracy. Yang et al. (2019) proposed a CNN inspired by MobileNet for fire detection. Wang et al. (2017) replaced the fully connected layer in a lightweight CNN with SVM to get better performance in fire identifying. Muhammad et al. (2019a) fine-tuned SqueezeNet for fire detection and localization. Zhang et al. (2021) proposed the ATT Squeeze U-Net which incorporates SqueezeNet structure into Attention U-Net architecture for fire detection.

Despite all the studies above, there are still some challenges in practical application. It is hard to eliminate false fire alarms as these methods may misclassify natural objects like red clothes, sunset, and light reflections. Once put into use on a large scale, high false alarm rates may greatly reduce the fire detection efficiency and even lead to the paralysis of the fire alarm system. Hence, the difficulty of visual fire detection lies in discriminating the fire-like objects and actual fire. To solve this problem, some researchers combined deep features with traditional fire features like motion and texture. Xie et al. (2020) proposed a video fire detection method that exploits both the deep statics features extracted by CNN and the motion-flicker-based dynamic features extracted by background subtraction and flicker detection to improve accuracy. Wu et al. (2019) presented an intelligent fire detection approach combining motion detection based on background subtraction, fire regions detection based on YOLO, and region classification between fire images and fire-like images using CNN. Barmpoutis et al. (2019) trained a Faster R-CNN model to obtain the candidate flame regions in the image and then used multidimensional texture analysis based on higher-order linear dynamical systems (Dimitropoulos et al., 2017) to determine whether each candidate region is a flame region. The above researches decrease fire detection errors to certain extents, while they increase the calculation and affect the detection efficiency. In this paper, we focus on the combination of deep learning and spectral analysis in one model. Spectral analysis is an effective and low-calculation method for fire texture feature extraction. Generally, CNNs take advantage of spatial characteristics of images like the local neighborhood and feature equivariance, while spectral analysis processes images in the frequency domain. More recently, some studies have demonstrated that the combination of CNN and spectral analysis achieved better or competitive classification accuracy in the general task of texture classification (Fujieda et al., 2017; Lu et al., 2018; Oyallon et al., 2018; Ulicny et al., 2018). Texture classification is the basis of image segmentation, object recognition and other visual tasks, which refers to assigning a predefined texture category (such as banded, cobwebbed, freckled, knitted, and zigzagged) to the image or image region based on its content (Cimpoi et al., 2014; Hayman et al., 2004). However, classification and detection are quite different, and texture classification is general-purpose, not specific for fire texture. The applicability of spectral analysis in fire detection is worth of further study.

In this paper, we introduce the combined method of CNN and spectral analysis into early fire detection. Specifically, the wavelet transform is applied to extract spectral features of the image, and then such features are input to CNNs at different layer stages. We choose the simplest wavelet, 2D Haar, for it is enough to depict different-frequency flame information, but our method is not restricted to Haar. The key idea is that the convolution layer and the pooling layer in CNNs can be regarded as a limited form of spectral analysis. Therefore, these two layers can be generalized by the 2D Haar transform to realize spectral analysis. To evaluate the efficiency of the proposed method, we use images from multiple sources containing a large number of images of fire and fire-like colors. Our key original contributions can be summarized as follows.

(1) It dominates state-of-the-art visual fire detection methods in terms of accuracy and the rate of false alarms by combining spatial characteristics based on CNN and spectral characteristics based on wavelet transform.
(2) Our method significantly improves the performance of lightweight CNN, balancing accuracy and computational complexity. This favors adaptation in surveillance networks with constrained resources in general.
(3) A diverse and balanced fire dataset containing images from multiple sources is introduced. Our dataset consists of images from the Corsican Fire Database (CFDB)(Toulouse et al., 2017), a few fire and non-fire images sampled and augmented from the Foggia's and Sharma's dataset (Foggia et al., 2015), and fire and non-fire images with fire like objects in the background from the internet.

2. Proposed method

The overall framework of our joint fire detection approach is based on the faster R-CNN. As shown in Fig. 1, it has three steps, feature extraction, region proposals generation, and classification and regression. This process of faster R-CNN makes it can be viewed as a form of divide and conquer strategy. The divide and conquer strategy implementing by block-based modular networks has some advantages. On one hand, it provides more interpretability for the current dominant end-to-end methodology. On the other hand, it has been proved to be more effective and efficient to split the task into sub-tasks and apply sub-network modules to find feasible solutions for very complex tasks like landslide or image signal processing (ISP, converting raw image to RGB image) which contain a serial of processes or variables (Abbaszadeh Shahri and Maghsoudi Moud, 2021; Zou et al., 2018).

The specific process of faster R-CNN is as follows. First, the image is input to the pre-trained CNN layers to get the feature pyramid network (FPN). FPN is commonly used in faster R-CNN, and its structure is shown in the illustration on the right of Fig. 1. It is a top-down architecture with lateral connections is developed for building high-level semantic feature maps at all scales (Lin et al., 2017). To better analyze the spectral features of fire, we adopt the wavelet convolutional neural network (Wavelet-CNN) instead of the conventional CNN. Then, using the extracted feature maps, the Region Proposal Network (RPN) can propose a certain number of ROIs (region of interests). At last, the ROIs and feature maps are pooled by the pooling layer and then input to the ROI-Head (consisting of fully connected layers and softmax layers) to determine the classes of these ROIs and fine-tune their positions. Our study focuses on designing a better backbone network for fire detection rather than FPN or detection head design, and the top-down enrichment and later connections are ignored.

2.1. Feature extraction with wavelet-CNN

Faster R-CNN uses conventional CNN layers to extract features. However, CNN models usually include skip-connections, making them known to be universal approximators (Hornik, 1991). It is not clear whether CNN models can learn to perform spectral analyses in practice with available datasets. Therefore, we directly integrate spectral approaches into CNN models based on multiresolution analysis using wavelet transform. Here we do not consider the problems of the shift sensitivity, directionality, and phase information, as they can be solved to a certain extent through the training data learning of CNN models.

We adopt the 2D Haar transform, which performs low-pass and high-pass filtering from horizontal and vertical directions. Given the image of size $M\times N = 2^m\times 2^n$, the decomposition outputs at the ith level are calculated by Wang and José (2010):

$$y_{hh}^{(i)}(u,v) = \sum_{l=1}^{2^{n-i+1}} \left[\sum_{k=1}^{2^{m-i+1}} h(k-2u)\, y_{hh}^{(i-1)}(k,l) \right] h(l-2v) \quad (1)$$

$$y_{hg}^{(i)}(u,v) = \sum_{l=1}^{2^{n-i+1}} \left[\sum_{k=1}^{2^{m-i+1}} h(k-2u)\, y_{hh}^{(i-1)}(k,l) \right] g(l-2v) \quad (2)$$

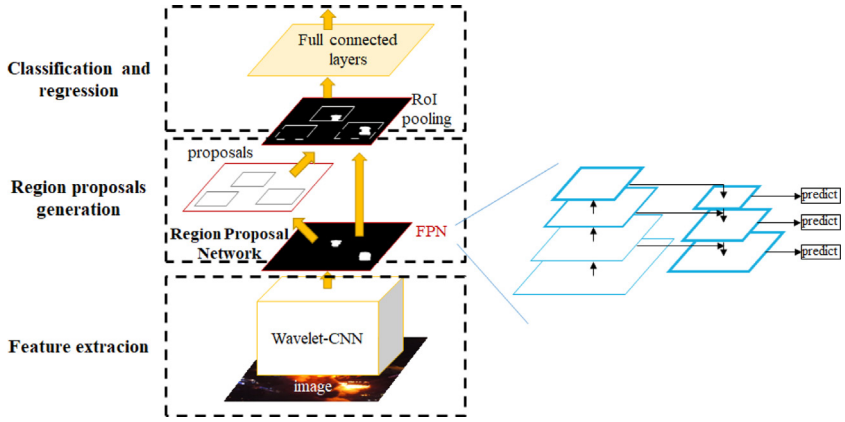


Fig. 1. Overall framework of our joint fire detection approach using faster R-CNN and wavelet transform.

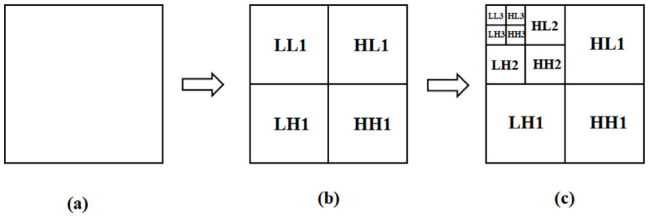


Fig. 2. Results of the Haar wavelet transform.

$$y_{gh}^{(i)}(u,v) = \sum_{l=1}^{2^{n-i+1}} \left[\sum_{k=1}^{2^{m-i+1}} g(k-2u)\, y_{hh}^{(i-1)}(k,l) \right] h(l-2v) \quad (3)$$

$$y_{gg}^{(i)}(u,v) = \sum_{l=1}^{2^{n-i+1}} \left[\sum_{k=1}^{2^{m-i+1}} g(k-2u)\, y_{hh}^{(i-1)}(k,l) \right] g(l-2v) \quad (4)$$

$u \in \{1, 2, \ldots, 2^{m-i}\}$ and $v \in \{1, 2, \ldots, 2^{n-i}\}$. $y_{hh}^{(i)}$ represents components that contain the approximation coefficient of the original image, and $y_{hh}^{(0)}$ = original image. $y_{hg}^{(i)}$, $y_{gh}^{(i)}$ and $y_{gg}^{(i)}$ respectively represent horizontal details, vertical details, and diagonal details coefficients. h and g represent the low-pass filter and the high-pass filter, respectively. We adopt the common form with $h = \{1/\sqrt{2}, 1/\sqrt{2}\}$ and $g = \{1/\sqrt{2}, -1/\sqrt{2}\}$. The effect of the Haar wavelet transform is shown in Fig. 2. Fig. 2(a) is the original image. Fig. 2(b) is the result of the first level wavelet transform, in which $LL1$ shows the 2-stride down-sampling image, $HL1$ shows the horizontal details, $LH1$ shows the vertical details, and $HH1$ shows the diagonal details. Fig. 2(c) shows the results of the second and third-level wavelet transform. It can be seen that after the Haar transform, the dimension of the image is quadrupled and the resolution is halved.

The key idea of the Wavelet-CNN model is concatenating wavelet layers with CNN layers. In this paper, we conduct Haar transforms for triple times. We test two classic backbone networks; one is the high-precision and heavy-weight ResNet50, and the other is the light-weight MobileNet v2. They are representative architecture for server-side and mobile applications respectively. Both of them are frequently used in computer vision and fire detection literature as baselines (Barmpoutis et al., 2019; Muhammad et al., 2019b; Yang et al., 2019). The overview of Wavelet-CNN models is shown in Fig. 3, where (a) is Wavelet-ResNet50 and (b) is Wavelet-MV2. For simplicity, we use an input image of 3 × 224×224 to illustrate the Wavelet-CNN architecture. The blue cubes in Fig. 3 represent convolution feature maps of ResNet50 and MV2, and the orange cubes represent wavelet features. In general, there exists a tradeoff between computation cost of FPN and the ability of detecting small objects. If we reuse more high-resolution feature maps in FPN, we are able to detect smaller objects. In practice, people need to balance the level number of FPN and the detection of small objects. In our experiments, we find the FPN connections given in Fig. 3 can provide satisfactory results on our dataset with relatively small fire.

The first Haar wavelet transform is performed on the original image, and we can get four wavelet features, $LL1$, $HL1$, $LH1$, and $HH1$, each of which has 3 channels with size 112 × 112. These wavelet features are decomposed using fixed parameters without significantly increasing the computational complexity. Then, we concatenate these 12-channel features with convolution features of the same size. Here to keep the parameters of the next convolution layer unchanged, we remove 12 channels from the original convolution features. The second Haar wavelet transform is performed on $LL1$, and we get $LL2$, $HL2$, $LH2$, and $HH2$, with a size 56 × 56. Then we concatenate these wavelet features with convolution features of the same size. Similarly, the third Haar wavelet transform is performed.

2.2. Region proposals generation with RPN

The purpose of this stage is to propose possible locations of objects, which are also called bounding boxes or anchors. To generate region proposals, one of the state-of-art deep learning approaches, the region-based CNN (R-CNN) (Girshick et al., 2014), uses a selective search approach. However, the selective search procedure is very slow. This procedure, rather than the CNN layers, takes up most of the time of object detection. To overcome such drawbacks, an improved variant model, faster R-CNN is proposed (Ren et al., 2017).

Faster R-CNN replaces the selective search procedure with RPN. The structure of RPN is shown in Fig. 4. The feature maps extracted from the Wavelet-CNN are input to the RPN module to simultaneously learn the class of the object as well as the associated bounding box. The outputs are a set of candidate bounding boxes, each with an objectness score, representing the probability of the object belonging to a class. With this end-to-end training procedure, the overall computational complexity is significantly reduced, while the performance is improved. Improving computational efficiency is still an open problem and several new architectures like anchor-free and detection transformer are developed later to achieve more efficient or more direct object detection without this anchor proposal stage.

RPN generates 1000 proposals for each image, and some proposals overlap with each other. To reduce redundancy, the common method is the non-maximum suppression (NMS) algorithm. Denote the list of proposal 1000 boxes as B and the list of filtered proposals D (which is initially empty). The process of NMS is as follows. First, select the proposal box with the highest confidence score, remove it from B and add it to D. Then, calculate the IOU (Intersection over Union) of this proposal with every other proposal. If the IOU is greater than the threshold N, remove that proposal from B. Again take the proposal with the highest confidence from the remaining proposals in B and remove it from B and add it to D. Once again calculate the IOU of this proposal with all the proposals in B and eliminate the boxes which have high IOU than the threshold. This process is repeated until there are no more proposals left in B.

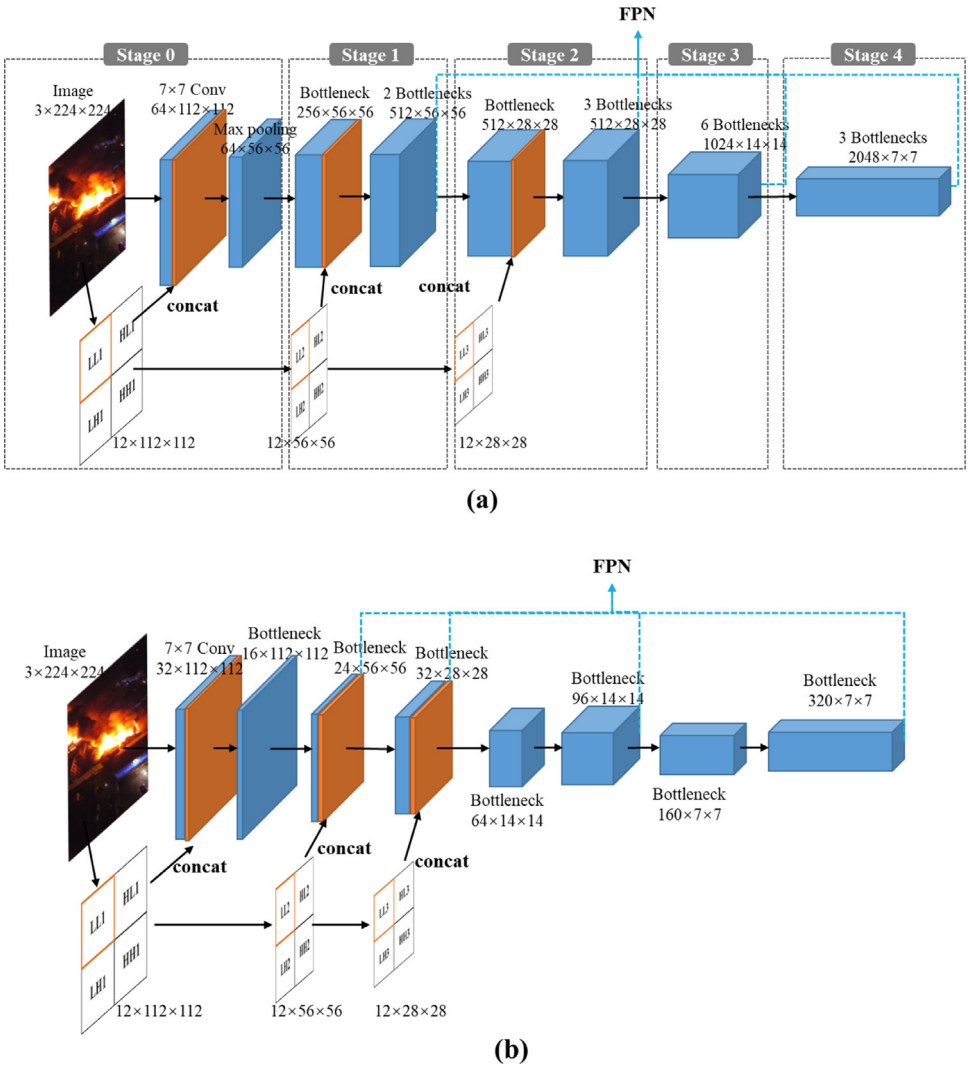


Fig. 3. Wavelet-CNN with a 3-level decomposition of the input image. (a) Wavelet-ResNet50. (b) Wavelet-MV2.

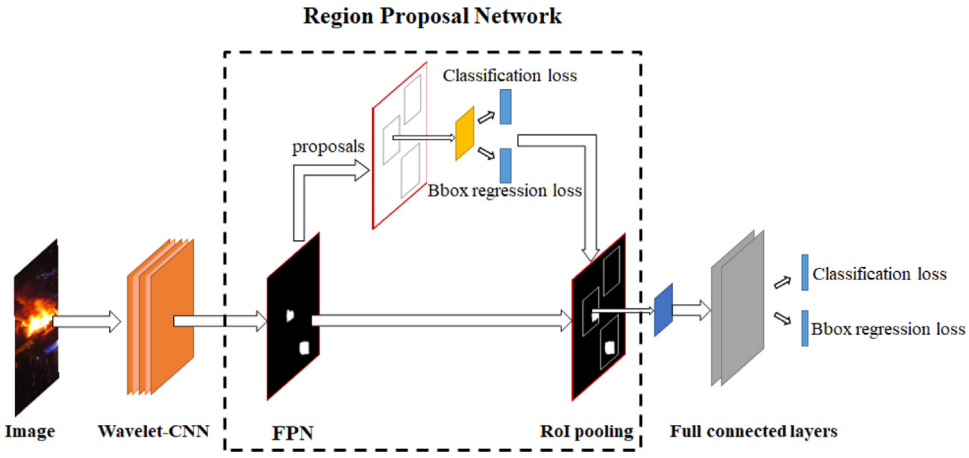


Fig. 4. Structure of RPN.

Here IOU calculation is used to measure the overlap between two proposals, as seen in Fig. 5(a). However, this method is not suitable for suspected fire proposals. A large number of overlaps still exist, and some boxes proposed are too small to represent the fire object. Unlike rigid objects like faces and cars, fire is ruleless fluid with a blurred boundary. Small sparks may fly around, confusing RPN with the proposed precise boxes. To eliminate the interference of small flying sparks and highlight the combustion flame, we propose IOS instead of IOU in the NMS algorithm. IOS equals the area of intersection over the area of the smaller box, as in Fig. 5(b).

The different effects of IOU and IOS on generating bounding boxes are shown in Fig. 6 (the threshold for NMS is set to 0.5). It can be seen

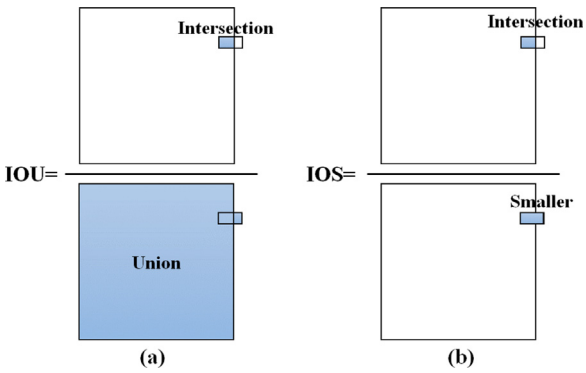


Fig. 5. Schematic of IOU and IOS.

that the bounding boxes generated by IOS do not overlap with each other, and almost every bounding box can cover a whole fire object.

2.3. Classification and regression with ROI-head

Through the ROI Pooling layer in RPN, we can obtain the feature vector of each candidate proposal, which represents the probability of the object belonging to a class. However, the specific class and accurate position of the region proposal are still unknown. To solve this issue, these feature vectors are input to the ROI-Head, in which fully connected layers and softmax layers are performed to determine which class the proposal belongs to and calculate its objectness score. Meanwhile, the bounding-box regression is used to obtain the prediction value of the offset of each region proposal relative to the ground-truth box, using which the region proposal can be modified and its position can be fine-tuned.

3. Experiment results and discussions

In this section, detailed experiments are conducted to evaluate and compare the performance of our method with other state-of-the-art methods. First, the datasets used for experiments are described in detail. Then, to show that our proposed approach improves the fire detection effect, we compare the recognition rates of single CNNs with the combination models of wavelet transform and CNNs of different architectures. To demonstrate the superiority of the proposed approach, we also compare our results with recently published related methods in a benchmark dataset. At last, for the scene of surveillance videos, we add the majority voting mechanism for video frames and test it with some fire and non-fire videos.

3.1. Dataset description

For experimentations, we use two image datasets (ImgDS1 and ImgDS2) and one video dataset (VDS3). ImgDS1 is used for both training and testing. It contains 1135 fire images from the Corsican Fire Database (CFDB) (Toulouse et al., 2017), a few fire and non-fire images sampled and augmented from the Foggia's and Sharma's dataset (Foggia et al., 2015), and some fire and non-fire images from the internet (Google and Baidu). Such non-fire images contain a few images that are hard to distinguish from fire images like a bright red room with high illumination, sunset, red-colored houses and vehicles, bright lights with different shades of yellow and red, etc. ImgDS2 is collected by Chino et al. (2015), consisting of 119 fire images and 107 fire-like images. Here ImgDS2 is used as the benchmark dataset for testing and comparing with other published methods. We use 80% images of ImgDS1 for training and the rest images for testing. With this setting, our model is trained with 2190 fire images and 2215 non-fire images. The statistics of training and testing data are given in Table 1.

Table 1
Statistics of training and testing images in this paper.

	Dataset source	Total images	Fire images	Non-fire images
Training	ImgDS1	4405	2190	2215
Testing	ImgDS1	1036	514	522
	ImgDS2	226	119	107

A few representative images from ImgDS1 and ImgDS2 are shown in Fig. 7.

The video dataset VDS3 is also used for testing. VDS3 consists of 8 fire videos and 12 non-fire videos, where some of them are collected from the reference (Gong et al., 2019), and others are obtained from our experiments.[1] These fire videos contain fires from indoor facilities like large space factories and warehouses and outdoor places like motorways, parks, and gas stations. They also contain house fires, electrical fires, spill fires, and different fire development periods from ignition, development, exuberance to extinction. Sample images from this dataset are shown in Fig. 8, and the video details are shown in Table 2.

3.2. Experiments with images

First, we use ImgDS1 to compare the performance of our proposed model with conventional CNNs. We use ImageNet to pre-train the original CNN and Wavelet-CNN models and fine-tune them with our dataset by combining them with FPN as illustrated in Fig. 3. In practice, suitable pretraining using large-scale public available datasets like ImageNet can help the training of specific tasks with limited training data. We train all the models using the V100 GUP by stochastic gradient descent (SGD), with the batch size as 8 and the learning rate as 0.01.[2] Standard data augmentation methods including flipping, rotating, and cropping are applied, to make the input images to a size of 224×224.

To ensure the reliability of our trained model, we draw out the loss curves during the training process, as shown in Fig. 9. The convergence behavior is very similar between the original models and the wavelet-adapted counterparts.

Here we do not need to care about the training speed unless it is unacceptably slow. We need to care more about the inference speed, which is critical for the real-world deployment of CNN models. The inference speed will be tested later in this section.

In a pattern recognition task, we need to define the confusion matrix (Table 3). Here we consider the confusion matrix of images and the confusion matrix of boxes. For images, the fire image that is correctly detected is a "true positive (TP)"; if not, it is a "false negative (FN)". The non-fire image that is correctly identified as non-fire is a "true negative (TN)"; otherwise, it is a "false positive (FP)". A score threshold of 0.3 is used to display these images. For the bounding boxes, if IOU $>$ 0.2 between the predicted box and the ground truth box, the predicted box is a "true positive (TP)"; otherwise, it is a "false positive

[1] VDS3 has been upload on our Google Drive, and the download link is https://drive.google.com/drive/folders/1ldsHg0M9oU9hIpPtJzADn1NL331MuNw9?usp=sharing.

[2] Our implementation is based on https://github.com/open-mmlab/mmdetection To ensure a fair comparison, all the hyper parameters are set the same as: mmdetection/configs/fast_rcnn_r50_fpn 1x.py in commit f96e57d6 except that the number of object classes is set to 2, the detection threshold is set to 0.3 and the image rescale size is set to (1000, 600). Hyper parameters should be dependent on the specific model. However, in the modern computer vision community, people seldom conduct this analysis. We simply follow this common paradigm, where different models are compared under the same hyper parameter setting. It should be noted that the performance of deep learning methods may be sensitive to various hyperparameters (Asheghi et al., 2020), so it is meaningful to investigate automatic methods for determining the best hyperparameter for specific applications or scenes, although this is beyond the scope of this paper.

Fig. 6. Different effects of IOU and IOS on generating bounding boxes. (a) Bounding boxes generated by IOU. (b) Bounding boxes generated by IOS.

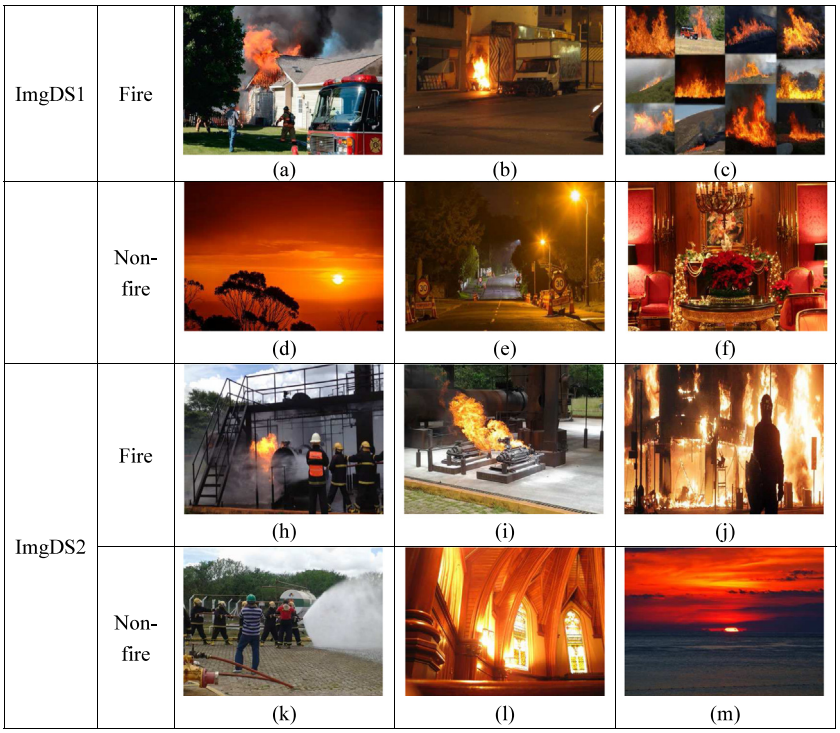


Fig. 7. Representative images of fire and non-fire from ImgDS1 and ImgDS2.

(FP)". Since every part of the image where we do not detect the object is considered negative, it is futile to measure "true negative number (TN)". Therefore, we only measure "false negative (FN)" as the missing of the model. Tables 4 and 5 show the detailed confusion matrix of different models.

To directly compare the performance of these models, we measure the false positive rate (FPR), the false negative rate (FNR), accuracy, precision, recall, and F-measure (F), their definitions are as follows. FPR and FNP are two indicators reflecting the system abnormality (shown in Figs. 10 and 11). Accuracy refers to the proportion of images that is correctly predicted. Precision is the number of true positives divided by the total number of images predicted and labeled as positive. Recall is the number of true positives divided by the total number of images that actually belong to the positive class. There is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other. To address this issue, F-measure is also be used, which refers to the harmonic mean of precision and recall.

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

$$FNR = \frac{FN}{TP + FN} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

We compare the CNN models with wavelet layers and those without (shown in Figs. 10 and 11). It can be seen that no matter ResNet50 or MV2, the wavelet transform makes the false positive rate and the false negative rate reduced, while the accuracy, precision, recall, and F-measure increased. For MV2, the false positive rate of images is reduced by 8.9%, the accuracy is increased by 4.7% and the precision is increased by 6.3%, which is a significant improvement. The performance measured by boxes is somewhat low, which may be caused by the poor consistency of image labeling due to the blurred flame boundary. Even so, it is noticeable that the performance of models with the wavelet transform is improved. These results demonstrate the effectiveness

Fig. 8. Sample images from video dataset.

Table 2
Video details.

Video name	Frames	Frame rate	Resolution	Description
video1	9,500	25	2048*1536	Fire in a large space factory building
video2	8,950	25	1280*720	Fire with incomplete shape
video3	293	10	352*288	Huge fire in a carriage
video4	1,419	30	1920*1080	Fire in a park
video5	9,018	30	1280*720	Electrical fire in a warehouse
video6	5,667	30	1280*720	Spill fire of a truck on the motorway
video7	2,878	30	640*360	Fire in a gas station with low illumination
video8	7,314	25	2048*1536	Fire at night with over exposure
video9	85	25	320*240	No fire, sunset
video10	645	25	640*480	No fire, sunshine
video11	541	30	640*368	No fire, flashing colored lights
video12	155	10	320*240	No fire, headlights at night
video13	5,829	25	1280*720	No fire, headlights with over exposure
video14	723	25	352*288	No fire, moving car
video15	251	25	352*288	No fire, fluttering reddish leaves
video16	196	19	1600*1200	No fire, fluttering red T-shirts
video17	246	25	360*288	No fire, moving red paper bag
video18	6,313	25	1280*720	No fire, red drill platform
video19	41,323	24	1920*1080	No fire, red drill platform with workers
video20	7,478	25	1280*720	No fire, red tank and workers

Table 3
Confusion matrix.

		Actual Class	
		Fire	Non-fire
Predicted Class	Fire	True Positive (TP)	False Positive (FP)
	Non-fire	False Negative (FN)	True Negative (TN)

of our proposed method. That is, concatenating wavelet layers can improve the texture recognition ability of the CNN model especially the light-weight CNN model, so as to improve the performance of fire detection.

The signal-noise-ratio (SNR) affects the image quality in practical applications. We add Gaussian noise to simulate disturbed images, where the mean value is set to zero and the standard deviation is set to 0, 10, 20, 30, 40, and 50, representing different noise levels. The detection results are shown in Fig. 12. We can see that all the models perform worse as the noise increases. Wavelet models are more robust to high noise levels compared with the original ones.

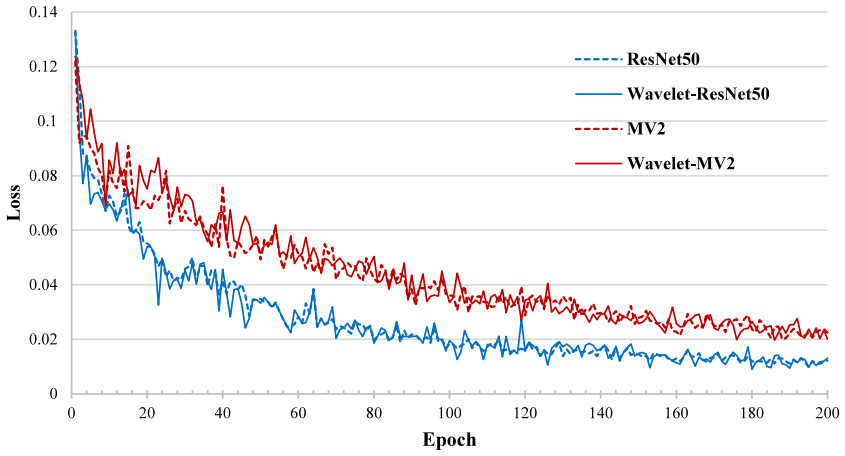


Fig. 9. Training and validation curves for model loss.

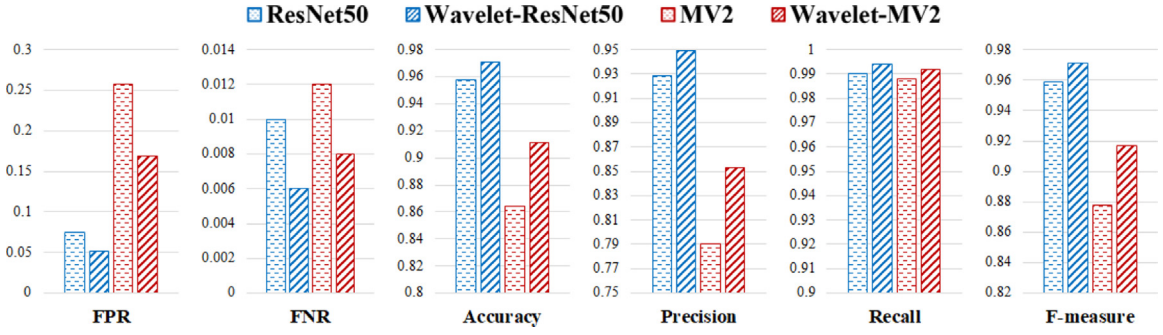


Fig. 10. Performance of different models measured by images.

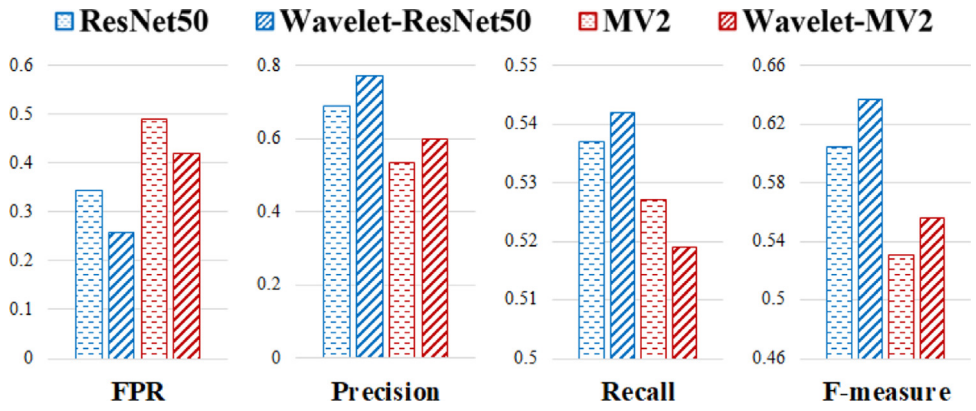


Fig. 11. Performance of different models measured by boxes.

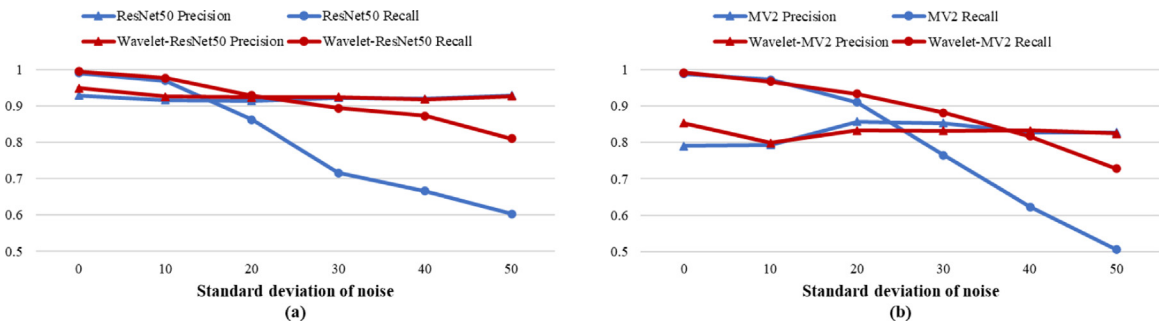


Fig. 12. Performance of different models under different SNRs.

Table 4
Confusion matrix of different models measured by images.

		ResNet50		Wavelet-ResNet50	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	509	39	511	27
	Non-fire	5	483	3	495
		MV2		Wavelet-MV2	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	508	135	510	88
	Non-fire	6	387	4	434

Table 5
Confusion matrix of different models measured by boxes.

		ResNet50		Wavelet-ResNet50	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	796	359	803	237
	Non-fire	685	/	678	/
		MV2		Wavelet-MV2	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	780	676	769	515
	Non-fire	701	/	712	/

Table 6
FPS of different models.

	ResNet50	Wavelet-ResNet50	MV2	Wavelet-MV2
FPS	2.90(0.3448)	3.03(0.3298)	14.35(0.0697)	20.45(0.0489)

We also measure the inference speed of different models with FPS, which means how many frames the model can process at one second (as listed in Table 6). The FPS of the wavelet-CNN model is increased compared with that of the basic CNN model. Especially for the wavelet-MV2 model, its FPS is increased by 42.5%. The results demonstrate that using wavelet layers to replace partial convolution layers can reduce the calculation of the model. Besides, FPS can be increased by a serial of model compression methods for neural networks, including model quantization, pruning, and neural architecture search. We believe these techniques can be applied to speed up the inference of our model.

To demonstrate the superiority of our proposed approach, we compare the performance of our proposed method with the existing fire detection methods in reference using ImgDS2. Although ImgDS2 is not very large, it is quite diverse and challenging with a lot of confusing images. We compare our methods with 10 representative methods, including 4 methods based on CNN models and 5 methods based on hand-crafted features like color, motion, and shape characteristics of the fire. We do not compare with general-purpose detection methods, since detecting fire is very different from detection in COCO or other general-purpose datasets. Using the evaluation metrics of precision, recall, and F-measure, the comparative results are given in Table 7, where the last four are hand-crafted feature models. First off, it has to be noted that ImgDS2 is not used in the training process of all the CNN based models, including our proposed methods and the four basic CNN models. However, although the hand-crafted feature models use ImgDS2 in training, they perform worst in testing. CNN based models have an overall qualitative improvement in performance. Comparing our methods with other CNN models, it can be seen that for the precision and recall rate, these methods have different characteristics. But in terms of F-measure, we can see that our methods are better overall. Moreover, the recall rate of our methods reaches 1, which means no fire image has been missed. That is very important to practical application. We also check the false-positive images of our methods, in which Wavelet-ResNet50 produces 15 false-positive images and Wavelet-MV2 produces 21. Some representative false-positive images are shown in Fig. 13. We can see that orange-red lightings are the main source of false alarms. In the latter study, more negative lighting samples can be added to the training set to solve this problem. This also indicates that

Table 7
Comparison with different existing fire detection methods on ImgDS2.

Method	Precision	Recall	F-measure
Wavelet-ResNet50	0.89	1	0.94
Wavelet-MV2	0.85	1	0.92
MobileNetV2 (Muhammad et al., 2019b)	0.90	0.93	0.92
SqueezeNet (Muhammad et al., 2019a)	0.83	0.97	0.90
GoogleNet (Muhammad et al., 2018b)	0.79	0.93	0.85
AlexNet (Muhammad et al., 2018a)	0.80	0.98	0.88
BowFire (Chino et al., 2015)	0.51	0.65	0.57
Rudz et al. (2013)	0.63	0.45	0.52
Rossi et al. (2011)	0.39	0.22	0.28
Çelik and Demirel (2009)	0.55	0.54	0.54
Thou-Ho (Chen et al., 2004)	0.75	0.15	0.25

reliability and uncertainty analysis (Abbaszadeh Shahri et al., 2021) can be considered to better understand the performance over different scenes.

For real-world surveillance scenarios, it is important that the fire detection system is robust against attacks. We test the effect on the performance of our method against different attacks such as noise, blockage, and rotation. We consider two test images, one is a fire image and the other is a non-fire image. The original fire image is given in Fig. 14(a), which is correctly detected by all models. In Fig. 14(b), the fire region in the image is disturbed by noise and the models still detect it. The detection confidence of the wavelet adapted CNN is higher than that of the corresponding CNN. In Fig. 14(c), the fire region is rotated and partially occluded, and our methods successfully detect the fire object. In Fig. 14(d), the fire region is completely occluded and our methods predict it as no fire. To show the effect on performance against fire-like images, we consider an image in Fig. 14(e), which is predicted as no fire by our methods. The fire-like images with noise and red-colored pattern are given in Fig. 14(f) and (g). We find that the wavelet adapted models still predict them correctly as no fire. To confirm that our method can detect fire with small size, we place a small fire image on Fig. 14(h). The wavelet models detect them correctly with higher confidence. These tests indicate that our detection method can detect fire even if the video frames are affected by noise or the size of the fire is small, which verifies its better performance.

3.3. Experiments with videos

In this section, we use VDS3 to demonstrate the effectiveness of our proposed method for fire detection from real surveillance video. We randomly extract five frames per second from a video and input them into our proposed model. Then the majority mechanism is conducted, which means if three or more of the five frames per second are detected as fire images, a fire alarm will be given; otherwise, no alarm. The computation efficiency can still be expressed in FPS. To make it more intuitive, we take video 1 as an example for further elaboration. The duration of video 1 is 6 min and 20 s. It takes 6 min and 21 s for MV2 to process it, 6 min and 20 s for Wavelet-MV2, 10 min and 56 s for ResNet50 and 10 min and 27 s for Wavelet-Resnet50. In general, under the test environment of this paper, Wavelet-MV2 and MV2 can realize real-time processing. To measure the performance, we calculate the confusion matrix, precision, recall, and F-measure per second, as shown in Table 8.

In general, the performance of the wavelet adapted models is higher than that of the corresponding original models, especially the FPR is 0, which means there is no false alarm. This is particularly important for fire detection. The traditional smoke fire detector often has a high false alarm rate due to the interference of dust and water vapor, which may lead to the alarm waterfall and paralysis of the fire alarm system. Reducing the false alarm rate can improve efficiency, which is essential to the construction of IoT-based Smart Cities (Mohammad et al., 2019). The false alarms of ResNet50 and MV2 occur in video11 and video13, which are caused by orange-red lights, as shown in Fig. 15(b). Besides,

Fig. 13. Representative false-positive images of ImgDS2.

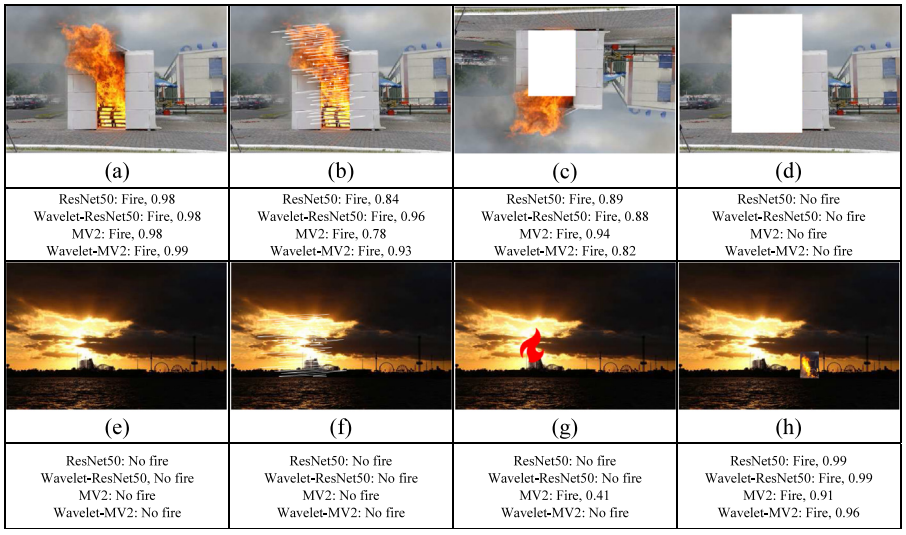


Fig. 14. Robustness analysis using noise attacks for the wavelet adapted models and other CNN models.

there are still some false negatives of wavelet models. We verify that these false negatives appear in video2, video5, and video8. In video2, Wavelet-MV2 has 3 s of missing alarms. This is because some frames in video2 only capture very few flame edges, and Wavelet-MV2 fails to detect them (as shown in Fig. 15(a)). This problem can be improved by modifying the simple alarm logic of minority subordinate to majority. In video5, Wavelet-ResNet50 has 12 s of missing alarms and Wavelet-MV2 has 13 s of missing alarms, and in video8, Wavelet-ResNet50 produces 38 s of missing alarms and Wavelet-MV2 produces 39 s of missing alarms. The fire in video5 is an electrical fire, where the sparks from the wires dropped to the ground and gradually triggered the fire. There is no such fire in our training samples, so it is not detected in the early stage. In video8, the surveillance camera automatically turns on the black and white mode when the illumination is very low, which is common for most cameras; the fire broke out at night, the illumination was not enough just after the fire started, and the camera was still in black and white mode, so our models fail to detect the fire. These missing alarms can be improved by supplementing some electrical fire images and fire images in the black and white mode into the training dataset in the succeeding work.

Table 8
Details of the detection performance of VDS3.

Confusion matrix					
		ResNet50		Wavelet-ResNet50	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	1204	29	1206	0
	Non-fire	49	3212	47	3241
		MV2		Wavelet-MV2	
	Actual Class	Fire	Non-fire	Fire	Non-fire
Predicted Class	Fire	1204	30	1198	0
	Non-fire	49	3211	55	3241

Detection performance						
Method	FPR	FNR	Accuracy	Precision	Recall	F-measure
ResNet50	0.0089	0.0391	0.9826	0.9765	0.9609	0.9686
Wavelet-ResNet50	0	0.0375	0.9895	1	0.9625	0.9809
MV2	0.0093	0.0391	0.9824	0.9757	0.9609	0.9682
Wavelet-MV2	0	0.0439	0.9878	1	0.9561	0.9776

4. Conclusions

This paper presents a combined method of CNN and spectral analysis into early fire detection. We apply the 2D Haar transform to extract spectral features of the image and then input them into CNNs at different layers stages. Two classic backbone networks are used to test our method, the high-precision and heavy-weight ResNet50, and the light-weight MV2. Results show that no matter what kind of network, the introduction of the wavelet layer can reduce the false positive rate, the false negative rate, and the computational complexity, and increase the accuracy, precision, recall, and F-measure. For the lightweight MV2, the performance improvement of the above indicators is more obvious. That is, the combination of wavelet transform can improve the fire identification ability of CNNs, especially light-weight CNNs. The test on real surveillance videos further demonstrates that our proposed model can meet the needs of real-time fire detection on precision and speed.

Our proposed approach can be used in chemical factories and other high-fire-risk industries. The accuracy and speed of our approach can meet the requirements of real-time fire detection. Its industrial deployment will help detect fires at the very early stage, promote emergency management, and thus contribute to loss prevention.

There are still some shortcomings in this study that can be improved. First, our model cannot eliminate all false positives. More

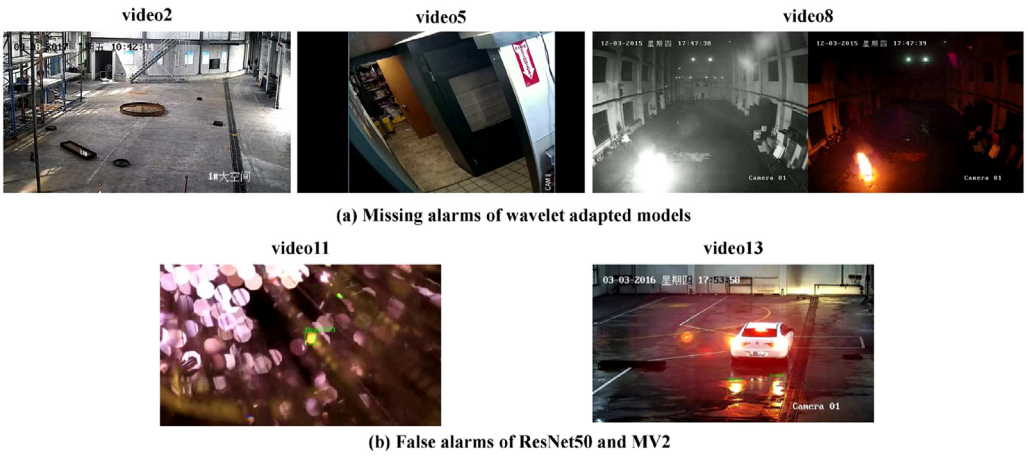


Fig. 15. Representative false negatives of VDS3.

```python
def forward(self, x):
    # output features of MV2 to FPN
    outs = []

    # iterate over all the layers of mv2
    for i, layer in enumerate(self.features):
        # if wavelet transform is added to current layer
        if self.wavelet_indices is not None and i in self.wavelet_indices:
            # if this is the first layer, the input to wavelet transform is the original image
            if i == 0:
                _x, x_LL = self.Wavelet_Layer(x)
                x = layer(x)
                x = torch.cat((x, _x), 1)
            # if this is not the first layer, the input to wavelet transform is the previous LL part
            else:
                x = layer(x)
                _x, x_LL = self.Wavelet_Layer(x_LL)
                x = torch.cat((x, _x), 1)
        else:
            x = layer(x)
        # add selected features to FPN
        if i in self.out_indices:
            outs.append(x)
    return tuple(outs)
```

orange-red lighting images, some electrical fire images, and fire images in the black and white mode can be added to the training set to solve this problem. Second, to verify the universality of the combination of wavelet analysis and the CNN model for fire detection, more types of CNN networks, like vision transformers can be tested. In addition, when applied to the video stream analysis, the model should be combined with reasoning theories to improve detection accuracy.

CRediT authorship contribution statement

Lida Huang: Conceptualization, Methodology, Software, Writing – original draft. **Gang Liu:** Data curation, Validation. **Yan Wang:** Software, Writing – review & editing. **Hongyong Yuan:** Supervision. **Tao Chen:** Resources, Project administration, Writing – review & editing.

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.

Acknowledgments

This research has been supported by the National Key R&D Program of China (Grant No. 2021YFC1523500) and the National Natural Science Foundation of China (Grant No. 72104123).

Appendix

We provide the implementation for the Wavelet-MV2 in Fig. 3 to further illustrate how to combine wavelet transform and CNNs.

Our implementation is based on https://github.com/open-mmlab/mmdetection, all the hyper parameters are set the same as: mmdetection/configs/fast_rcnn_r50_fpn_1x.py in commit f96e57d6 except that the number of object classes is set to 2, the detection threshold is set to 0.3 and the image rescale size is set to (1000, 600).

Training the RPM is very time-consuming in which the obtained features should be saved by CNN in the training process and this take a large amount of memory space. However, we do not consider memory management since it is acceptable on modern computers or servers.

References

Abbaszadeh Shahri, A., Maghsoudi Moud, F, 2021. Landslide susceptibility mapping using hybridized block modular intelligence model. Bull. Eng. Geol. Environ. 80, 267–284. http://dx.doi.org/10.1007/s10064-020-01922-8.

Abbaszadeh Shahri, A., Pashamohammadi, F., Asheghi, R., Abbaszadeh Shahri, H., 2021. Automated intelligent hybrid computing schemes to predict blasting induced ground vibration. Eng. Comput. http://dx.doi.org/10.1007/s00366-021-01444-1.

Asheghi, R., Hosseini, S.A., Saneie, M., Shahri, A.A., 2020. Updating the neural network sediment load models using different sensitivity analysis methods: A regional application. J. Hydroinform. 22, 562–577. http://dx.doi.org/10.2166/hydro.2020.098.

Barmpoutis, P., Dimitropoulos, K., Kaza, K., Grammalidis, N., 2019. Fire detection from images using faster R-CNN and multidimensional texture analysis. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, Brighton, Great Britain, pp. 8301–8305. http://dx.doi.org/10.1109/ICASSP.2019.8682647.

binti Zaidi, N.I., binti Lokman, N.A.A., bin Daud, M.R., Achmad, H., Chia, K.A., 2015. Fire recognition using RGB and ycbcr color space. ARPN J. Eng. Appl. Sci. 10, 9786–9790, http://www.arpnjournals.org/jeas/research_papers/rp_2015/jeas_1115_2983.pdf.

Çelik, T., Demirel, H., 2009. Fire detection in video sequences using a generic color model. Fire Saf. J. 44, 147–158. http://dx.doi.org/10.1016/j.firesaf.2008.05.005.

Chen, T., Wu, P., Chiou, Y., 2004. An early fire-detection method based on image processing. In: ICIP 2004. 170, IEEE, pp. 1707–1710, http://ieeexplore.ieee.org/document/1421401/.

Chino, D.Y.T., Avalhais, L.P.S., Rodrigues, J.F., Traina, A.J.M., 2015. BoWFire: Detection of fire in still images by integrating pixel color and texture analysis. In: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, pp. 9–102. http://dx.doi.org/10.1109/SIBGRAPI.2015.19.

Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A., 2014. Describing textures in the wild. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3606–3613. http://dx.doi.org/10.1109/CVPR.2014.461.

Dimitropoulos, K., Barmpoutis, P., Grammalidis, N., 2015. Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. IEEE Trans. Circuits Syst. Video Technol. 25, 339–351. http://dx.doi.org/10.1109/TCSVT.2014.2339592.

Dimitropoulos, K., Barmpoutis, P., Grammalidis, N., 2017. Higher order linear dynamical systems for smoke detection in video surveillance applications. IEEE Trans. Circuits Syst. Video Technol. 27, 1143–1154. http://dx.doi.org/10.1109/TCSVT.2016.2527340.

Foggia, P., Saggese, A., Vento, M., 2015. Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. IEEE Trans. Circuits Syst. Video Technol. 25, 1545–1556. http://dx.doi.org/10.1109/TCSVT.2015.2392531.

Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux, J.-M., Moreau, E., Fnaiech, F., 2016. Convolutional neural network for video fire and smoke detection. In: IECON 2016. IEEE, Florence, Italy, pp. 877–882. http://dx.doi.org/10.1109/IECON.2016.7793196.

Fujieda, S., Takayama, K., Hachisuka, T., 2017. Wavelet convolutional neural networks for texture classification. http://arxiv.org/abs/1707.07394.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 580–587. http://dx.doi.org/10.1109/CVPR.2014.81.

Gong, F., Li, C., Gong, W., Li, X., Yuan, X., Ma, Y., Song, T., 2019. A real-time fire detection method from video with multifeature fusion. Comput. Intell. Neurosci. 2019, 1–17. http://dx.doi.org/10.1155/2019/1939171.

Hayman, E., Caputo, B., Fritz, M., Eklundh, J.-O., 2004. On the significance of real-world conditions for material classification. pp. 253–266. http://dx.doi.org/10.1007/978-3-540-24673-2_21.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4, 251–257. http://dx.doi.org/10.1016/0893-6080(91)90009-T.

Li, Z., Mihaylova, L.S., Isupova, O., Rossi, L., 2018. Autonomous flame detection in videos with a Dirichlet process Gaussian mixture color model. IEEE Trans. Ind. Inform. 14, 1146–1154. http://dx.doi.org/10.1109/TII.2017.2768530.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: CVPR 2017. Honolulu, HI, USA, pp. 936–944. http://dx.doi.org/10.1109/CVPR.2017.106.

Lu, H., Wang, H., Zhang, Q., Won, D., Yoon, S.W., 2018. A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation. In: 2018 IEEE International Conference on Healthcare Informatics. ICHI, IEEE, New York, NY, USA, pp. 191–198. http://dx.doi.org/10.1109/ICHI.2018.00029.

Mohammad, N., Muhammad, S., Bashar, A., Khan, M.A., 2019. Formal analysis of human-assisted smart city emergency services. IEEE Access 7, 60376–60388. http://dx.doi.org/10.1109/ACCESS.2019.2913784.

Muhammad, K., Ahmad, J., Baik, S.W., 2018a. Early fire detection using convolutional neural networks during surveillance for effective disaster management. Neurocomputing 288, 30–42. http://dx.doi.org/10.1016/j.neucom.2017.04.083.

Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., Baik, S.W., 2019a. Efficient deep CNN-based fire detection and localization in video surveillance applications. IEEE Trans. Syst. Man, Cybern. Syst. 49, 1419–1434. http://dx.doi.org/10.1109/TSMC.2018.2830099.

Muhammad, K., Ahmad, J., Mehmood, I., Rho, S., Baik, S.W., 2018b. Convolutional neural networks based fire detection in surveillance videos. IEEE Access 6, 18174–18183. http://dx.doi.org/10.1109/ACCESS.2018.2812835.

Muhammad, K., Khan, S., Elhoseny, M., Hassan Ahmed, S., Wook Baik, S., 2019b. Efficient fire detection for uncertain surveillance environment. IEEE Trans. Ind. Inform. 15, 3113–3122. http://dx.doi.org/10.1109/TII.2019.2897594.

Oyallon, E., Belilovsky, E., Zagoruyko, S., Valko, M., 2018. Compressing the input for CNNs with the first-order scattering transform. In: ECCV 2018. Munich, Germany, pp. 305–320. http://dx.doi.org/10.1007/978-3-030-01240-3_19.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. http://dx.doi.org/10.1109/TPAMI.2016.2577031.

Rossi, L., Akhloufi, M., Tison, Y., 2011. On the use of stereovision to develop a novel instrumentation system to extract geometric fire fronts characteristics. Fire Saf. J. 46, 9–20. http://dx.doi.org/10.1016/j.firesaf.2010.03.001.

Rudz, S., Chetehouna, K., Hafiane, A., Laurent, H., Séro-Guillaume, O., 2013. Investigation of a novel image segmentation method dedicated to forest fire applications. Meas. Sci. Technol. 24, 075403. http://dx.doi.org/10.1088/0957-0233/24/7/075403.

Schultze, T., Kempka, T., Willms, I., 2006. Audio–video fire-detection of open fires. Fire Saf. J. 41, 311–314. http://dx.doi.org/10.1016/j.firesaf.2006.01.002.

Sharma, J., Granmo, O.-C., Goodwin, M., Fidje, J.T., 2017. Deep convolutional neural networks for fire detection in images. In: International Conference on Engineering Applications of Neural Networks, Communications in Computer and Information Science. Springer, Cham, Athens, Greece, pp. 183–193. http://dx.doi.org/10.1007/978-3-319-65172-9_16.

Toreyin, B.U., Cetin, A.E., 2007. Online detection of fire in video. In: Cvpr 2007. IEEE, pp. 1–5. http://dx.doi.org/10.1109/CVPR.2007.383442.

Töreyin, B.U., Dedeoğlu, Y., Güdükbay, U., Çetin, A.E., 2006. Computer vision based method for real-time fire and flame detection. Pattern Recognit. Lett. 27, 49–58. http://dx.doi.org/10.1016/j.patrec.2005.06.015.

Toulouse, T., Rossi, L., Campana, A., Celik, T., Akhloufi, M.A., 2017. Computer vision for wildfire research: An evolving image dataset for processing and analysis. Fire Saf. J. 92, 188–194. http://dx.doi.org/10.1016/j.firesaf.2017.06.012.

Ulicny, M., Krylov, V.A., Dahyot, R., 2018. Harmonic networks: Integrating spectral information into CNNs. http://arxiv.org/abs/1812.03205.

Wang, Y., Dang, L., Ren, J., 2019. Forest fire image recognition based on convolutional neural network. J. Algorithm. Comput. Technol. 13, 174830261988768. http://dx.doi.org/10.1177/1748302619887689.

Wang, H., José, V., 2010. 2-D wavelet transforms in the form of matrices and application in compressed sensing. In: 2010 8th World Congress on Intelligent Control and Automation. IEEE, pp. 35–39. http://dx.doi.org/10.1109/WCICA.2010.5553961.

Wang, Zhicheng, Wang, Zhiheng, Zhang, H., Guo, X., 2017. A novel fire detection approach based on CNN-svm using tensorflow. In: Proc. Intelligent Computing Methodologies. Liverpool, United Kingdom, pp. 682–693. http://dx.doi.org/10.1007/978-3-319-63315-2_60.

Wu, H., Wu, D., Zhao, J., 2019. An intelligent fire detection approach through cameras based on computer vision methods. Process Saf. Environ. Prot. 127, 245–256. http://dx.doi.org/10.1016/j.psep.2019.05.016.

Wu, S., Zhang, L., 2018. Using popular object detection methods for real time forest fire detection. In: 2018 11th International Symposium on Computational Intelligence and Design. ISCID, IEEE, pp. 280–284. http://dx.doi.org/10.1109/ISCID.2018.00070.

Xie, Y., Zhu, J., Cao, Y., Zhang, Yunhao, Feng, D., Zhang, Yuchun, Chen, M, 2020. Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features. IEEE Access 8, 81904–81917. http://dx.doi.org/10.1109/ACCESS.2020.2991338.

Yang, H., Jang, H., Kim, T., Lee, B., 2019. Non-temporal lightweight fire detection network for intelligent surveillance systems. IEEE Access 7, 169257–169266. http://dx.doi.org/10.1109/ACCESS.2019.2953558.

Zhang, J., Zhu, H., Wang, P., Ling, X., 2021. ATT squeeze U-Net: A lightweight network for forest fire detection and recognition. IEEE Access 9, 10858–10870. http://dx.doi.org/10.1109/ACCESS.2021.3050628.

Zou, B.-J., Guo, Y.-D., He, Q., Ouyang, P.-B., Liu, K., Chen, Z.-L., 2018. 3D filtering by block matching and convolutional neural network for image denoising. J. Comput. Sci. Technol. 33, 838–848. http://dx.doi.org/10.1007/s11390-018-1859-7.