Identifying the Presence of Fire Smoke Within Video Sequences Through the Application of Convolutional Neural Networks

Kun WANG
Sichuan Vocational College of Information Technology, Guangyuan 628017, China

Abstract—**Fire detection systems are of great importance due to the rapid spread and high destructive potential of fires. Research on smoke detection has been a significant issue in fire prevention and control because it enables early detection. Conventional models use hand-crafted features based on prior knowledge to determine whether a frame includes smoke. These models are mostly static and sensitive to the fire scene environment, often resulting in false alarms. In this article, convolutional neural networks (CNNs) are used for image recognition to detect smoke in video frames. A common detection model based on 3D CNN and RCNN is developed. An RCNN, using non-maximal suppression, is applied to identify the location of the smoke based on static spatial information. Next, 3D CNN detects smoke by integrating dynamic temporal-spatial features. Extensive experiments have been performed on the Visor and Bilkent datasets. The results of these experiments show that our approach has high performance and accuracy. Additionally, the results indicate that its reliability (in terms of false alarm rate) is better than similar methods. Furthermore, our approach is capable of recognizing black-and-white smoke.**

Keywords—Fire; smoke detection; Convolutional Neural Network (CNN); video frames; spatial-temporal features

I. INTRODUCTION

The early detection of fire basis on detection of the resulting smoke from it with the help of the smart methods and the early containment of the fire for the reduction of the financial damages and the environmental damages (which are caused by it), is considered as one of the important issues in the world of the science and the industry [1]. The traditional methods are based on the caused effects by the fire, such as the heat and the warmth of the combustion gases and the light radiation, which these detectors do not have the necessary efficiency in the large open-roof environments. Therefore, according to the wide use of the CCTV cameras in security systems of the various environments, the tendency to the use from these cameras in the detection of the sudden dangerous events (such as fire) has increased [1, 2, 3, 4, 5, 6]. In addition to proper performance of the detection system of the fire based on video frames in the large open-roof environments, the main advantage of this system, in compared to the other traditional detector systems, is the ability to detect the occurrence of the fire at the early moments. Unlike the other traditional sensors, the camera is not a type of the direct contact. This means that the other types of the used sensors work in a way that the caused effects by the fire (such as the heat, the gas or the warmth) must travel a period of the time, to reach the actual location of the fire (from the location of the sensor) and then, activate the sensor [5, 7, 8, 9]. Another advantage of the use of a camera is the possibility of the investigation of the generated alarm and the confirmation or the disconfirmation of it by a human observer.

Another advantage of the use of the video frames is the presentation of the additional useful information in the direction of the fire mitigation [4, 5, 6, 10]. Mainly, in order to automatically detect, the detection systems basis on the video frames use the features of flame and smoke on fire, such as shape, color and the texture mobility [3, 11, 12]. So far, the various fire detection methods have been presented, which in these methods, a wide range of the different features have been used. In the initial researches in the field of the flame detection based on the video frame processing, they paid more attention to the color component. But, in order to reduce the detection error, they have extracted the other features from the image. In the many researches, the obtained empirical rules by the thresholding on the image in the different color spaces (RGB, HSV and YCbCr), have been used as the color models [2, 10, 13, 14, 15, 9].

Previous researches has primarily utilized convolutional neural networks (CNNs) as feature extraction tools, focusing mainly on images for detecting flames and smoke. However, video sequences offer rich temporal information, with the appearance of smoke being akin to an action. CNNs have also been effectively applied to video analysis, capturing spatiotemporal features. Drawing inspiration from action recognition using 3D CNNs, we have developed a combined detection framework based on Faster RCNN and 3D CNN. The Faster RCNN identifies the location of smoke targets, while the 3D CNN handles smoke recognition. In the current article, the convolutional neural networks are used for the recognition of the image, to detect the smoke on the video frames. A common detection model, which is basis on3D CNN and RCNN, is developed. An RCNN, by the non-maximal convolution, is applied to identify location of the smoke, basis on the static information of the spatial. Next, 3D CNN discovers the smoke detection by integrating the dynamic temporal-spatial information. The remaining of the paper is organized as the below. In the second section of this article, the literature review is done and the presented methods in this scope are briefly discussed. The third section of this article details the proposed model for the detection of the fire basis on its smoke. The fourth section describes details of the used datasets and then, provides the performed tests and their outcomes. Eventually, the fifth section of this article provides conclusions and directions for the next research.

II. Related Work

On scope of the detection of the fire based on smoke, the various methods have been presented, which these approaches are briefly reviewed. In study [16], the authors have used a simple statistical model based on the fuzzy logic in the color space of YCbCr for the description of the fire areas, but since this method only focused on the image, it did not have the expected accuracy. Therefore, in the newer research, instead of the use from the images, the videos have been used, to model a suitable pattern from the fire, by applying the frame-to-frame changes of the fire in several consecutive frames. In [4], the authors have presented a color model, by considering a Gaussian distribution for the flame regions in the space of RGB. Then, by extracting the features such as the variance of the pixel brightness changes, the changes in the area of the fire candidate region on two consecutive frames and the complexity of the investigated regions, they have tried to recognize the incidence of the fire, with help of the Bayesian classification. But the Gaussian assumption of flame regions distribution on space of RGB is not a proven acceptable assumption, which reduces the detection accuracy and increases the computational cost of the method.

In study [17], the Fourier descriptors are used, to interpret the shape of the fire regions. By considering that the Fourier descriptor considers the image signals as a set of the sinusoidal signals and the signals of the fire regions have a random irregular behavior, therefore, it seems that these descriptors do have not the necessary efficiency. In study [18, 12], the mixture scheme of Gaussian is applied, to model the background pixels of images. The use of this method brings computational complexity, which causes the diagnosis to take the much time. In study [18], fuzzy clustering is applied in the colour space of LAB, to identify the flame regions. Due to the nature of the fuzzy clustering method, this method may in some cases face a delay in the quick detection of the fire occurrence. The component of the moving of the flame regions is one of the widely used features for the detection of the flame, which, in addition to the Gaussian mixture distribution method, in many articles such as [4, 16, 19, 20, 17] is used, to extract the moving areas of background.

In study [21], the cumulative geometric independent component analysis model is used for this purpose. But the researches shows that since several consecutive frames are needed for the detection of the moving regions (by using any of the mentioned methods), if this feature is used, then the detection speed will decrease. The derived features from the image analysis contain the important information about the structural status of the different parts of the image and its relationship with the other parts, which is used in study [7, 18]. One of the statistical methods for the analysis of the texture is the methods basis on the co-occurrence matrix, which has been used in [7, 18]. In study [20], with the help of a multilayer network, the candidate neural regions of fire are categorized into the categories of fire and non-fire. Three spatial features and 11 temporal features are extracted by the fire candidate areas on image, and this vector of the features is considered as input of neural network. In [8, 22], the fuzzy methods are applied, to identify the regions of the fire.

In [10, 23], SVM has been applied, to classify regions and to make decisions. Since all the mentioned classification methods are the supervised learning methods, in case of the absence of the valid labeled training data, the use from these methods will be problematic. In [17], the hidden model of Markov is applied, to model the temporal behavior of fire regions. In a fire detection system, the complexity of the used model is directly related to detection time. In [2, 3, 4], the issue of the fire detection time is not discussed. It is expected that in some of the presented systems, the detection time by using the normal processors will reach the several tens of seconds or even minutes, which is a very important challenge, because the identification of the fire in the initial stages of its formation will have a highly influencing role in the process of its control. A few seconds from the start of the fire may result in the irreparable loss of the life and the property. Therefore, the purpose of the article is to provide a system for the detection of the fire with the help of the video image processing, which can recognize the incidence of the fire with optimal accuracy, in a short time (the low calculation volume).

III. The Presented Model

The presented model for smoke detection in video frames is shown on Fig. 1. First, the fast RCNN is used, to create the suspicious boxes from the smoke on the frames, which are selected by the video trails in a static time distance. When a suspicious box from the smoke is recognized on a frame, a clip for this box is exploited by cutting the consecutive frames circa this frame. Next, a 3D CNN extracts the temporal-spatial features for this clip. Eventually, SVM or Softmax is applied, for the training of model. The fire is an event with the low probability. It means which the most recorded video trails do not include the smoke. nevertheless, there are several CNN models, which can locate classify and locate the works together on the videos: S-CNN [25] and R-C3D [24]. The early detection basis on the frame is a reasonable manner for the reduction of the computational complexity. For a video, the sliding time windows with the lengths equal to 16, 32 and 64 and with a sliding step equal to 16 are extracted.

A. Step of Clip Presentation for Region with the Smoke Frame

In study [26], the fast R-CNN is applied, to recognize the smoke of fire in wild forest, basis on the artificial images of the smoke. The artificial images of the smoke are applied, to solve the absence of the data of the image for the smoke and to remove the sample annotations. The experimental outcomes in [26] displays which the fast RCNN, which is trained by the artificial images of the smoke, is the susceptible to the plumes of the smoke, but it is not the susceptible to the lean smoke, according to the absence of the similar examples on the data of the training. The outcomes of the further experiments in study [26] display which the rate of the false alarm is relatively great. Obviously, the many objects without the smoke have received a great score, which indicates that the fast RCNN has a poor ability for the classification. For the improvement of the detection, the use from 3D CNN is considered for the more detection, and the fast RCNN is considered as a proposed model in the presentation of the clip for the smoke frame box. In the models for the detection of the smoke, the color and the motion are mostly applied as the smoke area presentation models, which its aim is the area reduction of the extraction of the feature and the localization of the object.

RPN on the fast RCNN creates 300 suggestions for every target, and several suggestions overlap together. To decrease the redundancy, the fast RCNN takes the non-maximal suppression (NMS) in the exploited boxes, basis on their scores. Nevertheless, NMS significantly decreases number of the suggestions, but it is not proper for the suspicious smoke suggestion. First, there is still a large volume of the overlap, because two suggestions are booked together, if the intersection-over-union is less than 0.3. In addition, the proposed box from NMS is so little for coverage of the entire object from the smoke. Unlike the rigid cases such as the cars and the faces, the smoke has an ambiguous boundary (with its translucent properties), that distracts RPN by the exploited exact boxes. The scope of the smoke is gradually developed on clip, because the smoke spreads during the time. There is a lot of temporal information in the smoke boundary, and it needs to be fed into a 3D CNN for the extraction of the feature. In this aim, the non-maximal annexation (NMA) is designed, to merge 300 boxes (which are exploited from RPN). NMA and NMS are displayed on Algorithm 1. The generated boxes from NMA are the non-overlapping, and every box masks a full object from the smoke. The various outcomes of NMA and NMS are displayed on Fig. 2.

B. Step of the Spatial-Temporal Features Extraction

The 3D CNN for extraction of the used temporal-spatial features in our approach is basis on C3D-v1.0 [28]. In this paper, the overfitting is possible, because the available datasets in this domain are very small. As shown in Fig. 3, a smaller network is used, to avoid the overfitting. This network consists of 5 layers of the 3D convolution, 5 layers of the 3D max-pooling and 3 layers of the fully connected and one layer of the Softmax, for prediction of the labels. As a function of the activation, the ReLU function is located in output of every layer of 3D convolutional and in output of the first two layers of fully-connected. The first 2 layers of fully-connected are followed by the layers of dropout [29], to prevent the overfitting. On common applications of 3D CNNs, input is a clip with 16 frames. Nevertheless, the spread of the smoke is so gradual. This movement is hardly visible in the time intervals less than $1s$. In addition to the clips with the 16 frames, here, the clips with 32 frames and the clips with 64 frames are used as the input of 3D CNN. The size of all frames has been changed to 128×171. Also, the dimensions of the input are equal to $3 \times length \times 128 \times 171$. The random crops with the size equal to $3 \times length \times 112 \times 112$ are used during the training.

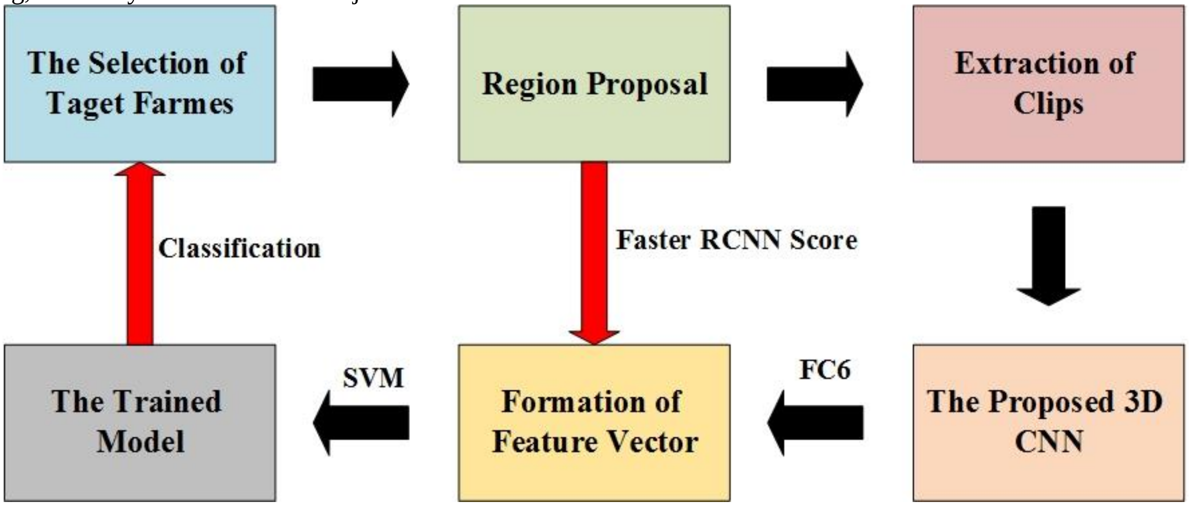


Fig. 1. The general scheme of the presented model for the detection of the smoke in the video frames.

Algorithm 1. Non-Maximum Suppression and Non-Maximum Annexation

```
NMS:
    1. Rank 300 boxes by score
    2. If $\boldsymbol{IoU}(\boldsymbol{boundingbox_i}.\boldsymbol{box_j}) > \mathbf{0.3}$, then delete $\boldsymbol{box_j}$
    3. Else if $\boldsymbol{IoU}(\boldsymbol{boundingbox_i}.\boldsymbol{box_j}) < \mathbf{0.3}$, then save $\boldsymbol{box_j}$ as $\boldsymbol{boundingbox_{i+1}}$
    4. If $\boldsymbol{score}(\boldsymbol{boundingbox_i}) > \mathbf{0.8}$, then alarm
    5. Rank boxes by score
NMA:
    1. If $\boldsymbol{score}(\boldsymbol{box_j}) < \mathbf{0.01}$, then delete $\boldsymbol{box_j}$
    2. Else if $\sum_{\boldsymbol{i=1}}^{\boldsymbol{I}} \boldsymbol{IoU}(\boldsymbol{boundingbox_i}.\boldsymbol{box_j}) == \mathbf{0}$, then save $\boldsymbol{box_j}$ as $\boldsymbol{boundingbox_{i+1}}$
    3. Else if $\boldsymbol{IoU}(\boldsymbol{boundingbox_i}.\boldsymbol{box_j}) < \mathbf{0.6}$ && $\sum_{\boldsymbol{k \neq i}}^{\boldsymbol{I}} \boldsymbol{IoU}(\boldsymbol{boundingbox_k}.\boldsymbol{boundingbox_i} \cup \boldsymbol{box_j}) == \mathbf{0}$, then save $\boldsymbol{boundingbox_i} \cup \boldsymbol{box_j}$ as $\boldsymbol{boundingbox_i}$
    4. Take $\boldsymbol{boundingbox_i}$ $(\boldsymbol{i} = \mathbf{1}.\mathbf{2}. \ldots . \boldsymbol{I})$ as suspected smoke boxes
```

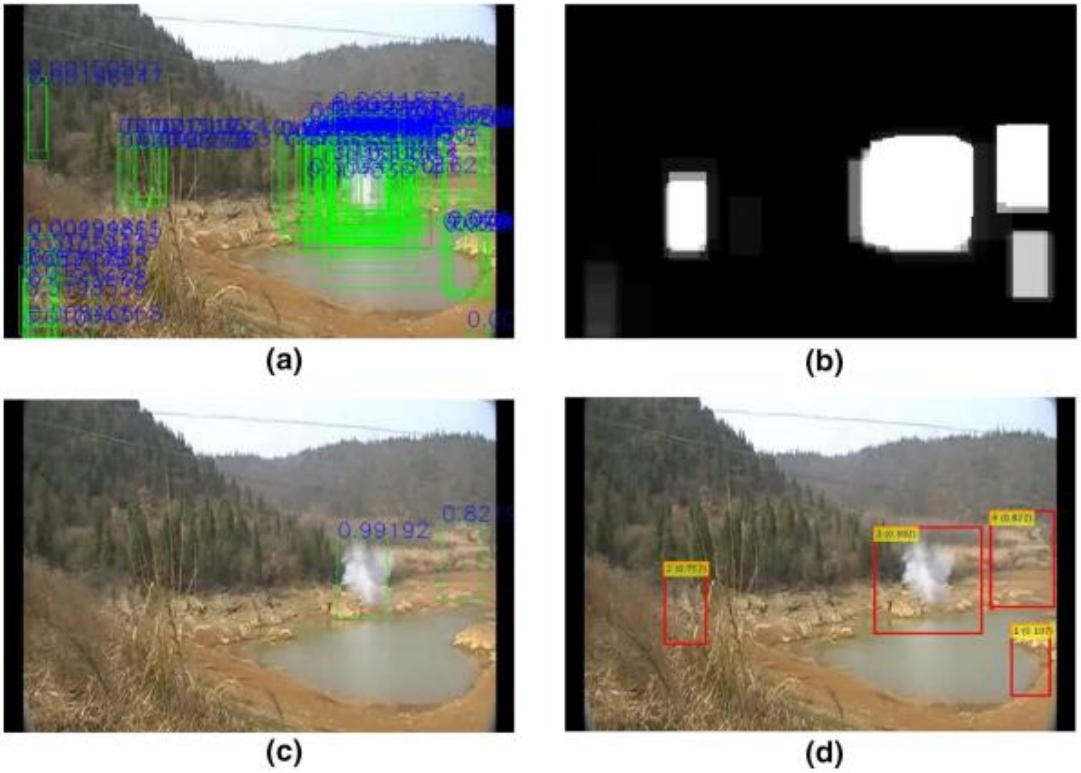


Fig. 2. The disagreement among NMA and NMS. (a) 300 boxes from RPN, (b) the superposition of points for 300 boxes, (c) the boxes from NMS, (d) the boxes from NMA.

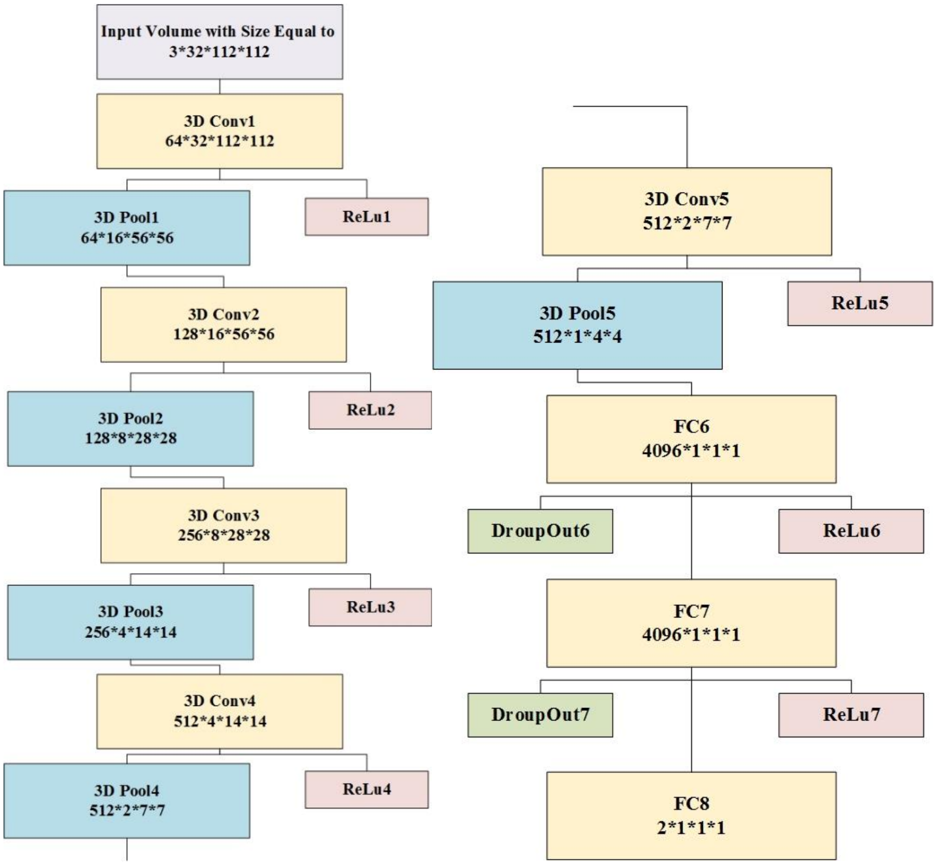


Fig. 3. The structure of the proposed 3D CNN.

C. Data Augmentation

The abundant data with the high quality is the factor for the success of the models of ML. The models that are trained by the low data, often travail from the challenge of overfitting, because they do not distribute well enough to data from the validation set and the testing set. The augmentation of the data is a manner for prevention from the challenge of the data scarcity. There are several ways for the augmentation of the data, like: the noise addition and the applying of the various transformations to the current data. In this paper, the equipped manners for the augmentation of the data are used, to augment the training set. On the layer of the data in 3D CNN, the horizontal rotation and the random cutting are used. Due to the skyward motion of the smoke, the vertical rotation is not used. The imported clips to network are randomly cropped to 112×112, after that they are resized to 128×171. Due to the effects of the daily changes and the weather changes, it is inevitable to change the brightness in the captured videos. To simulate these environmental changes, brightness of clips on the training set is varied by increasing and by decreasing the intensity. The training data is doubled through this operation. The increment and the decrement of brightness in every clip is handled by a possibility number.

In the digital tools for the imaging, the noises of Gaussian and salt-pepper are two popular sources of the noise. To more augment the data of the training, the data is increased as twice size of the main training data, by adding the noises of Gaussian and salt-pepper into clips. Although CNNs can obtain the end-to-end learning, but they act as a black box, because what exactly has been learned, has the ambiguous. The preprocessing in the data of the input may displays the superior efficiency than the initial input. In 2 flow CNNs [27], the flow of the optical is applied as the input on the temporal detection flow, and is greatly superior than the training on the initial frames. As expected, 3D CNN extracts the information of the temporal by the video trails. The pre-processing manners are used, to extract the motion information in the experiments.

IV. Evaluation and Results

Here, details of the datasets and the tests and the outcomes are provided. Python3.7 has been applied, to implement these experiments. Our model is implemented on the computer with RAM 8G and Intel(R) CPU 3.0 GHz Core(TM) i7. CNN is implemented in GPU and the graphics card is GEFORCE 840M for NVIDIA.

A. Datasets

In the conducted experiments, two datasets have been used. Eight clips are chosen from two datasets, which are used for the detection of the smoke (Table I). The first dataset is the Visor1 dataset [30], which is one of the most popular datasets in scope of the detection of the smoke. It contains 14 clips with the various resolutions and the different backgrounds. The clips of VC1-VC5 on Table I are chose by this dataset. VC1-VC2 have the so complicated backgrounds, and quality of these clips is low. The purpose of the choice of these clips is to compute the tolerance of the noise for the presented model and its accuracy for the complicated backgrounds. VC3-VC4 include the non-smoke motion objects and the smoke-like motion objects, with the actual smoke. These clips are taken in outdoors. Thus, they have the various conditions of the lighting. The presence of the wind and the change of the smoke direction are the key reasons for the choice of them. On the other hand, VC5 is recorded in a large hall.

TABLE I. THE FEATURES OF USED CLIPS IN THE EXPERIMENTS

Clip	Name	Dataset	Resolution	Number of Frame	Smoke
VC1	01_ballistic	Visor	320×240	347	Yes
VC2	02_explosion	Visor	320×240	210	Yes
VC3	04_fumogeno1	Visor	320×240	3005	Yes
VC4	05_fumogeno2	Visor	320×240	1835	Yes
VC5	10_hangar	Visor	384×288	2953	Yes
VC6	BehindtheFence	Bilkent	320×240	630	Yes
VC7	BtFence2	Bilkent	320×240	1400	Yes
VC8	ParkingLot	Bilkent	320×240	1726	Yes

The second dataset is the Bilkent dataset [31], which includes the various clips from the scenes that contain the smoke and the fire. The high diversity in the field, the distance from source and the various conditions of the weather are the key features of the dataset. 3 clips are chosen for the tests, which are the clips of VC6-VC8. The hindrances, like the fences in front of camera, make it as the further difficult to detect the smoke in VC6-VC7. Too, the distance from source to camera on VC8 is relatively far, that make the detection as the further difficult. Fig. 4 displays a frame of the chosen clips. The real data has been prepared for all frames of these videos. The provided real background is used in the testing and training.

B. Evaluation Criteria

First, the real background with features of every area is exploited on total frames of the testing clips. That's mean, for total identified candidate areas of the smoke, the labels of the smoke and the non-smoke are tagged by user. 80% of the data is applied for training, and the remnant is applied for testing. The similar to the most works for the detection of the smoke [33, 34, 35, 36, 37], the presented model is analyzed at the level of the frame, so that if there is an area from the smoke in the frame under the processing, then this frame is taken as a frame of the smoke. Otherwise, it is taken as the frame of the non-smoke. By considering the corresponding real background, the output of the presented model for every area can be one of the following cases: True Negative, True Positive, False Negative and False Positive [32]. These cases are applied, to compute $four$ criteria: sensitivity or TPR, specificity or TNR, accuracy and false positive rate or FPR. These criteria are calculated as the follows:

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

$$TNR = \frac{TN}{FP+TN} \tag{1}$$

$$FPR = \frac{FP}{FP+TN} \tag{1}$$

$$Accuracy = \frac{TN+TP}{FN+FP+TN+TP} \tag{4}$$

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 4. The sample frames from the used videos in the experiments. (a) VC1; (b) VC2; (c) VC3; (d) VC4; (e) VC5; (f) VC6; (g) VC7; (h) VC8.

C. Results

Here, the efficiency of our model is evaluated and is compared by the similar methods [33, 34, 35, 36, 37]. The outcomes of total models are quoted on the same clips by the relevant papers. For the clarification of the comparisons, the experimental outcomes for every dataset are reported as separately. VC1-VC5 are in Visor. The outcomes of these $five$ videos are available for the methods in [34, 36, 37]. Table II shows the comparison of outcomes of these videos. According to Table II, it can be seen which in the most situations, the presented model has the superior outcomes than the similar models. VC1 has a very complicated background. Too, the smoke on this clip spreads as so quickly in environment and next, as quickly disappears. Our approach for this clip reaches the accuracy equal to 99.64%, and ranks in 2-th position with just a little difference, in compared to foremost model. According to the rapid fading of smoke on this clip, there is a probability of the misdetection. Nevertheless, our approach for this clip works as very well. On VC2, the similar with VC1, the smoke spreads as so quickly in environment. Nevertheless, our approach yields the foremost performance. VC3 and VC4 are the suitable samples for the analysis of the efficiency of the models in the identification of the actual smoke against the smoke-like cases. On VC3, the method [34] obtains the foremost performance, and the proposed models obtains 2-th position. On VC4, our approach outperforms than the similar models, according to total criteria. Veritably, on this video, our approach significantly decreases the rate of the false detection. As mentioned, VC5 is chosen because of the attendance of smoke on a big space, which makes that smoke is dissipated as quickly. Nevertheless, the presented model in this video achieves an accuracy equal to 99.46%.

Table III shows the comparison of outcomes for VC6-VC8 from Bilkent. In this table, the outcomes of the methods in [33, 34, 36, 37] are available, and their outcomes are reported in these videos. VC6 and VC7 are chosen, due to of the fence in front of the areas. The outcomes display which in VC6, our approach has the superior accuracy than the method in study [37], and is in the first position. The proposed model also performs better than the similar methods on VC7. On VC8, that contains the smoke areas in the far distance, our approach ranks on the 3-th position. nevertheless, its efficiency in this video is so superior than the methods in study [33, 36]. In the following, the ability of our approach in detection of total candidate areas, and investigation of performance of temporal-spatial features on ultimate outcomes are evaluated. For this purpose, efficiency of our approach is evaluated in VC1-VC8 by using temporal-spatial features and without these features. Fig. 5 to Fig. 8 show outcomes. They clearly display which temporal-spatial features increase efficiency of our approach on the most videos. The accuracy improvement on 8 videos is remarkably impressive. Moreover, the values of TPR by our model demonstrate its ability to find total candidate areas of smoke. Too, there is a great reduction in FPR. Moreover, values of TNR confirm significant effect of use from temporal-spatial features in improvement of our model. In most of the examined videos, by adding spatial-temporal features, FPR is greatly decreased. In most situations, its value is near to 0. The extent of reduction in FPR for 8 clips is clearly visible in Fig. 5 to Fig. 8. The effect of use from these features can be clearly viewed on Fig. 9 that displays the visual samples of effectiveness of temporal-spatial features in removal of non-smoke regions.

D. Discussion

Different scenes require specific preprocessing to identify suspected regions in images. Our focus is on comparing the classification performance of various CNN-based fire detection methods. Faster RCNN is considered the initial step in the general fire detection pipeline discussed. Detection results indicate that 3D CNN significantly enhances smoke detection performance by extracting spatial-temporal information. Fine-tuning involves using a pre-trained network on large-scale data to initialize the CNN and making minor adjustments to improve it with new training data. This process saves training time and prevents overfitting in small datasets. However, comparing CNNs with different structures is challenging because the pre-trained network's structure is fixed. Experimental results generally show that 3D CNNs significantly improve smoke detection compared to image-based methods. Considering network design, the models should not be too deep to avoid overfitting with small datasets. Slow Fusion is preferable to Early Fusion, disregarding computational costs. The lack of training data limits the development of video smoke detection, leading to overfitting and poor generalization. While synthetic smoke video is a novel solution, accurately mimicking smoke motion is challenging and often results in poor performance.

TABLE II. THE COMPARISON OF OUTCOMES IN THE DIFFERENT VIDEOS ON VISOR

Clip	Metric	Method in [34]	Method in [36]	Method in [37]	Proposed Method
VC1	TPR	100	-	-	99.34
	TNR	99.70	-	-	99.68
	FPR	-	-	-	0.31
	Accuracy	99.85	96.55	74.86	99.64
VC2	TPR	97.40	-	-	100
	TNR	99.24	-	-	100
	FPR	-	-	-	0
	Accuracy	98.32	98.45	92.51	100
VC3	TPR	99.37	-	-	99.51
	TNR	99.72	-	-	93.24
	FPR	-	-	-	6.85
	Accuracy	99.54	93.20	88.52	97.36
VC4	TPR	99.14	-	-	99.50
	TNR	93.19	-	-	96.07
	FPR	-	-	-	3.92
	Accuracy	96.17	89.16	79.36	98.53
VC5	TPR	-	-	-	100
	TNR	-	-	-	95.94
	FPR	-	-	-	4.05
	Accuracy	-	-	-	99.46

TABLE III. THE COMPARISON OF RESULTS IN THE VARIOUS VIDEOS ON BILKENT

Clip	Metric	Method in [33]	Method in [34]	Method in [36]	Method in [37]	Proposed Method
VC6	TPR	-	-	-	-	96.12
	TNR	-	-	-	-	100
	FPR	-	-	-	-	0
	Accuracy	-	94.92	94.44	96.15	96.37
VC7	TPR	-	-	-	-	99.57
	TNR	-	-	-	-	89.83
	FPR	-	-	-	-	10.16
	Accuracy	90.53	98.70	98.71	96.55	99.37
VC8	TPR	-	-	-	-	93.18
	TNR	-	-	-	-	95.62
	FPR	-	-	-	-	4.37
	Accuracy	78.83	98.47	81.56	100	93.39

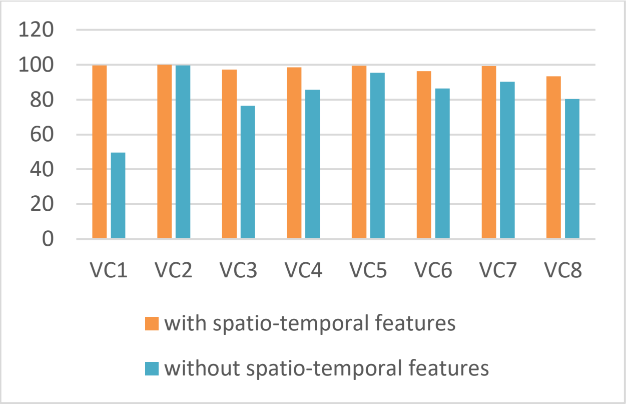


Fig. 5. The comparison of accuracy of our approach without/with the temporal-spatial features.

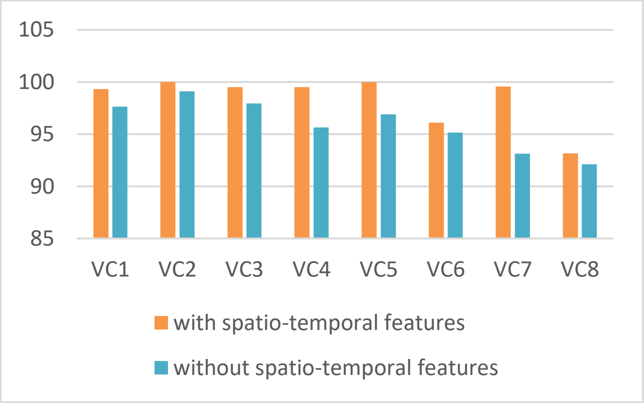


Fig. 6. The comparison of TPR of our approach without/with the temporal-spatial features.

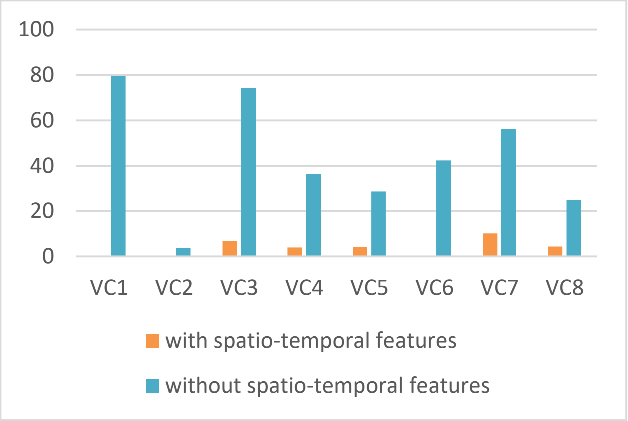


Fig. 7. The comparison of FPR of our approach without/with the temporal-spatial features.

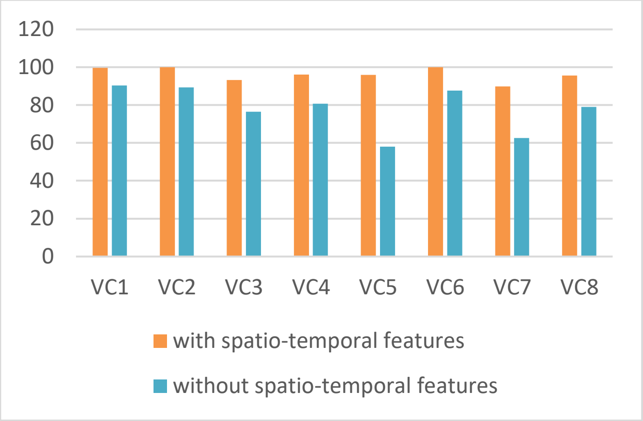
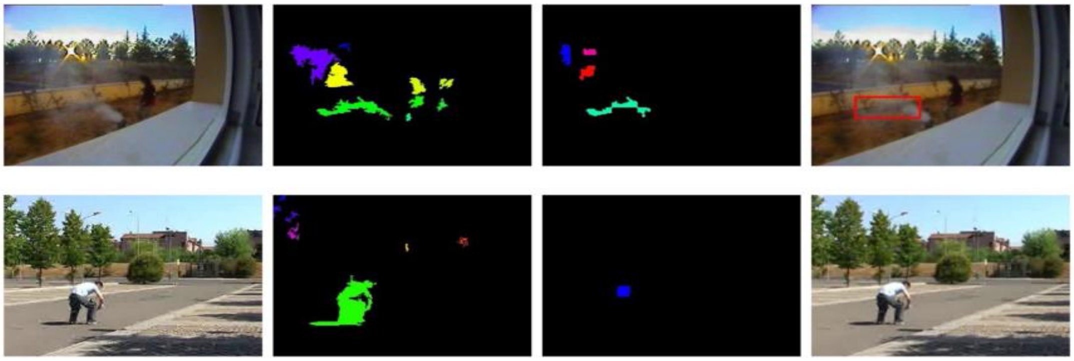


Fig. 8. The comparison of TNR of our approach without/with the temporal-spatial features.



(a) (b) (c) (d)

Fig. 9. The visual comparison of our approach without/with the temporal-spatial features on two images. (a) the input, (b) the exploited background, (c) the output of our model without the temporal-spatial features and (d) the output of our model with the temporal-spatial features.

V. Conclusions and Suggestions

The early detection of fire based on identification of the resulting smoke with the help of smart methods and its early containment for the reduction of the many damages of the financial and the environment by it, is considered one of the important issues in the world of the science and the industry. In this paper, a detection model basis on 3D CNN and the fast RCNN is proposed. The fast RCNN is mostly applied as the method for the presentation of the candidate area of the smoke, which discovers the location of the smoke and the initial detection. 3D CNN is applied, to exploit the temporal-spatial information. To perform the smoke detection, NMA is proposed, to replace NMS in the proposed fast RCNN network. It has the bigger boxes without the overlapping. The outcomes of the tests display which the proposed model greatly betters the detection of the smoke, in compared to the similar approaches. The lack of training data poses a challenge for expanding smoke detection in videos, leading to overfitting and poor generalization. Data augmentation techniques, such as horizontal rotation, random cropping, brightness adjustment, and noise addition, are effective solutions. Future research could focus on identifying and selecting optimal combinations of spatial and temporal-spatial features. Additionally, employing new CNN-based architectures could improve smoke region prediction. Reducing false alarms can also be achieved through frame trail analysis.

REFERENCES

[1] M. Hashemzadeh, B. Adlpour Azar, Retinal blood vessel extraction employing effective image features and combination of supervised and unsupervised machine learning methods, Artif. Intell. Med. 95 (2019) 1–15.

[2] N. Farajzadeh, M. Hashemzadeh, Exemplar-based facial expression recognition, Inform. Sci. 460–461 (2018) 318–330.

[3] M. Hashemzadeh, Hiding information in videos using motion clues of feature points, Comput. Electr. Eng. 68 (2018) 14–25.

[4] N. Farajzadeh, M. Hashemzadeh, A deep neural network based framework for restoring the damaged persian pottery via digital inpainting, J. Comput. Sci. 56 (2021) 101486.

[5] K. Muhammad, J. Ahmad, and S. Wookbaik, "Early fire detection using convolutional neural networks duringnsurveillance for effective disaster management," Neurocomputing, vol. 288, pp. 30-42, 2018.

[6] A. Feizy, “Application of Sparse representation and camera collaboration in visual surveillance systems," Signal and Date Processing Journal, Issuc. 15, No. 3, 2018.

[7] A. E Cetin, et al,"Video fire detection - review," Digital Signal Processes A Reveiw Journal, vol. 23, no.6, pp. 1827-1843, 2013.

[8] H. Chao, and K. Tzu-Hsin, "Real-time video-based fire smoke detection system," International Conference on Advanced Inteliigent Mechatronics, pp.1845-1850, 2009.

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

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

[11] J. Antony, and J. C. Prasad, "Real Time Fire and smoke detection using multi-expert system for video-surveillance," international journal for innovative research in science & technology, vol. 3, 2016.

[12] H. Xian-Feng, et al., "Video fire detection based on gaussian mixture model and multi-color features," Signal, Image and Video Processing, vol. 11, no. 8, pp. 1419-1425, 2017.

[13] J. Seebamrungsat, S. Praising, and P. Riyamongkol, "Fire detection in the buildings using image processing," Third ICT International Student Project Conference, pp. 95-98, 2014.

[14] G. F. Shidik, et al., "Multi color feature background subtraction and time frame selection for fire detection," IEEE International Con- ference on Robotics, Biomimetics, and Inte- lligent Computational Systems (ROBIONETICS), pp. 115-120, 2013.

[15] M.J. Sousa, A. Moutinho, M. Almeida, Wildfire detection using transfer learning on augmented datasets, Expert Syst. Appl. 142 (2020) 112975.

[16] T. Ccilik, H. Ozkaramanli, and H. Demirel, "Fire pixel classification using fuzzy logic and statistical color model," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 1000-1205, 2007.

[17] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, "Flame detection in vidco using hidden Markov models, in IEEE International Conference on Image Processing, vol. 2, pp. 1230-1233, 2005.

[18] T. X. Truong, and J. Kim, "Fire flame detection in video sequences using multi-stage pattern re- cognition techniques," Engineering Applications of Artificial Intelligence, vol. 25, no. 7, pp. 1365- 1372, 2012.

[19] Y. Zhao, and G. Tang, "Fire video recognition based on flame and smoke characteristics, "in The 2nd International Conference on Systems and Informatics (ICSAI 2014), pp. 113- 118, 2014.

[20] K. Dimitropolos, P. Barmpoutis, N. Grammalidis, "Spatio-temporal flame modeling and dynamic texture analysis for automatic vidco-based fire detection," IEEE transactions on circuits and systems for video technology, vol. 25, no. 2, pp.339-351, 2015.

[21] J. Rong, et al.,"Fire flame detection based on GICA and target tracking," Optics & Laser Technology, vol. 47, pp. 283-291, 2013.

[22] H. SunJae. K. Byoungchul, and N. Jae, "Fire- flame detection based on fuzzy finite automation," 20 International Conference on Pattern Recognition(ICPR), pp.3919-3922, 2010.

[23] J. Z, et al., "SVM based forest fire detection using static and dynamic features," Computer Science Information Systems Journal, vol. 8, no. 3, pp. 821-841, 2011.

[24] Xu H, Das A, Saenko K (2017) R-C3D: region convolutional 3D network for temporal activity detection. In: Proceedings of the IEEE international conference on computer vision, pp 5783–5792.

[25] Shou Z, Wang D, Chang SF (2016) Temporal action localization in untrimmed videos via multi-stage CNNs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1049–1058.

[26] Zhang QX, Lin GH, Zhang YM et al (2018) Wildland forest fire smoke detection based on fast R-CNN using synthetic smoke images. Procedia Eng 211:441–446.

[27] Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. Comput Linguist 1(4):568–576.

[28] Du T, Bourdev L, Fergus R et al (2015) Learning spatiotemporal features with 3D convolutional networks. In: IEEE International conference on computer vision. IEEE, pp 4489–4497.

[29] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R (2012) Improving neural networks by preventing coadaptation of feature detectors. Computing Research Repository. http://arxiv.org/abs/1207.0580.

[30] P. Piccinini, S. Calderara, R. Cucchiara, Reliable smoke detection in the domains of image energy and color, in: 2008 15th IEEE International Conference on Image Processing, 2008, pp. 1376–1379.

[31] N. Dedeoglu, B.U. Toreyin, U. Gudukbay, A.E. Cetin, Real-time fire and flame detection in video, in: Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, 2005, pp. ii/669-ii/672 662.

[32] M. Hashemzadeh, A. Zademehdi, Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features, Expert Syst. Appl. 130 (2019) 60–78.

[33] A. Filonenko, D.C. Hernández, K. Jo, Fast smoke detection for video surveillance using CUDA, IEEE Trans. Ind. Inf. 14 (2018) 725–733.

[34] P. Barmpoutis, K. Dimitropoulos, N. Grammalidis, Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition, in: 2014 22nd European Signal Processing Conference, EUSIPCO, 2014, pp. 1078–1082.

[35] D.K. Appana, R. Islam, S.A. Khan, J.-M. Kim, A video-based smoke detection using smoke flow pattern and spatial–temporal energy analyses for alarm systems, Inf. Sci. 418-419 (2017) 91–101.

[36] F. Yuan, Z. Fang, S. Wu, Y. Yang, Y. Fang, Real-time image smoke detection using staircase searching-based dual threshold AdaBoost and dynamic analysis, in: IET Image Processing, Institution of Engineering and Technology, 2015, pp. 849–856.

[37] K. Avgerinakis, A. Briassouli, I. Kompatsiaris, Smoke detection using temporal HOGHOF descriptors and energy colour statistics from video, in: International Workshop on Multi-Sensor Systems and Networks for Fire Detection and Management, 2012.