International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 2, Mar-Apr-2020, ISSN (Online): 2395-566X

Fire Detection System using Video Surveillence Based on Facenet

Dr. V. Subedha, G.Akash, N.Lokesh,P.Sasikumar
pithodcse@gmail.com,akashgkv152@gmail.com, nlokeshkumar199@gmail.com,sasikukarperiasamy123@gamil.com

***Abstract* – An accidental fire is a mishap that could be either man-made or natural. Accidental fire occurs frequently and can be controlled but may at times result in severe loss of life and property. Fire detection using hand-crafted features is a tedious and time-consuming method. The accuracy of the existing system using Alex net is 78% to 92%.The Project main idea is to detect the fire as soon as possible .The main concept used in the project is facenet pertained model. It is recognition technique to detect the fire on the surroundings. It uses smaller convolutional kernels and contains no dense, fully connected layers, which helps keep the computational requirements to a minimum. Despite its low computational needs, the experimental results demonstrate that our proposed solution accuracies that are comparable to other, more complex models, mainly due to its increased depth. The embedded processing capabilities of smart cameras have given rise to intelligent CCTV surveillance systems. Fire is the most dangerous abnormal event, as failing to control it at an early stage can result in huge disasters, leading to human, ecological and economic losses. Inspired by the great potential of CNNs, propose a lightweight CNN based on the SqueezeNet architecture for fire detection in CCTV surveillance networks. A Project approach can both localize fire and identify the object under surveillance. The accuracy of the system using Facenet is 98%.**

***Keywords* – Convolutional neural networks (CNNs), deep learning, fire detection, fire disaster, fire localization, image classification, surveillance networks.**

I. INTRODUCTION

The Rapid Growth of Urbanization increase the House, Building and Theme Parks. When there is a fire, it will cause so many damage for property and also loss of human life. To Avoid this type of incident different types of technologies are invented. They are mainly based on sensor. It detect the fire only it come closer range of the sensor. So Sometime fire get enlarge and make a severe damage. we use deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. According to incomplete statistics, there were 312,000 fires in the country in 2016, with 1,582 people killed and 1,065 injured, and a direct property loss of 3.72 billion dollars. Fire detection is vitally important to protecting people’s lives and property. The current detection methods in cities rely on various sensors for detection including smoke alarms, temperature alarms, and infrared ray alarms. Although these alarms can play a role, they have major flaws. First, a certain concentration of particles in the air must be reached to trigger an alarm. When an alarm is triggered, a fire may already be too strong to control, defeating the purpose of early warning. Second, most of the alarms can only be functional in a closed environment, which is ineffective for a wide space, such as outdoors or public spaces. Third, there may be false alarms. When the non-fire particle concentration reaches the alarm concentration, it will automatically sound the alarm. Human beings cannot intervene and get the latest information in time. To prevent fires and hinder their rapid growth, it is necessary to establish a monitoring system that can detect early fires. Establishing a camera-based automatic fire monitoring algorithm and FaceNet model. Greatly reducing the cost increases the economic feasibility of such systems. In the preprocessing module, the frame difference detection operates quickly and does not include complex calculations, has low environmental requirements, and does not need to consider the time of day, weather, and other factors. The camera-based fire monitoring system can monitor the specified area in real time through video processing. When a fire is detected based on the video, it will send a captured alarm image to the administrator. The administrator makes a final confirmation based on the submitted alarm image. For example, when an accident occurs on a highway and causes a fire, based on the image transmitted by the detection algorithm, one can immediately rescue the victims, saving precious time and minimizing damage. We combine motion detection based on frame difference with color detection based on the RGB/HSI model. Color detection is only for regions of motion that the motion detection phase is completed. Our method has improved the precision and reduced redundant calculation. In addition, we have improved the frame difference method. According to the spatial correlation between consecutive image frames, we have improved the traditional methods of detecting fire from one single image frame.

Temporal information is combined with the flame features through a space-time flame centroid stability-based detection method. At the same time, we combine the data obtained during the fire preprocessing phase to reduce computational redundancy and computational complexity. We extracted various flame features, spatial variability, shape variability, and area variability. We used the support vector machine to train, complete the final verification, reduce the false negatives rate and false positives rate, and improve the accuracy.

II. EXISTING SYSTEM

We consider several candidate architectures, with reference to general object recognition performance within [12], to cover varying contemporary CNN design principles [13] that can then form the basis for our reduced complexity CNN approach. AlexNet [14] represents the seminal CNN architecture comprising of 8 layers. Initially, a convolutional layer with a kernel size of 11 is followed by another convolutional layer of kernel size 5. The output of each of these layers is followed by a max pooling layer and local response normalization. Three more convolutional layers then follow, each having a kernel size of 3, and the third is followed by a max pooling layer and local response normalization. Finally, three fully connected layers are stacked to produce the classification output.

VGG-16 [15] is a network architecture based on the principle of prioritizing simplicity and depth over complexity – all convolutional layers have a kernel size of 3, and the network has a depth of 16 layers. This model consists of groups of convolutional layers, and each group is followed by a max pooling layer. The first group consists of two convolutional layers, each with 64 filters, and is followed by a group of two convolutional layers with 128 filters each. Subsequently, a group of three layers with 256 filters each, and another two groups of three layers with 512 filters each feed into three fully connected layers which produce the output. Here we implement the 13-layer variant of this network by removing one layer from each of the final three groups of convolutional layers (denoted VGG-13).

III. PROPOSED SYSTEM.

Our Project idea is to detect the fire and avoid the damages and loses in human life and properties. Fire can be detected using video surveillance in deep learning using Convolution Neural Network (CNN) Algorithm. This Algorithm has four-layer Convolution layer, max pooling layer and ReLU. Convolution layer takes input as matrix representation and then features of the image are extracted within this layer using filter. The output of the convolution layer is passed into Rectified Linear Unit (ReLU) Layer. The ReLU Layer extract positive values portion from matrix based on activation function. The output is passed through the Max pooling function to obtained maximum value for each patch of the feature map. Now Using fully connected layer the output is compared with the fire dataset and the result is produced.

IV. MODULE DESCRIPTION

1. Data Collection

The anomaly detection system applies Convolutional Neural Network (CNN) to classify 5000 data sets and provides major developments in the experimental result. The CNN Model gives the detection result based on features established by training dataset. A Deep Learning is established on event classifier trained through 4000 frames of videos. First, randomly selected 1000 images per event category are a training set and 1000 images are a validation set for 4 categories. The CNN model accomplished 100% anomaly detection accuracy on the validation data set after training.

2. Feature Extraction

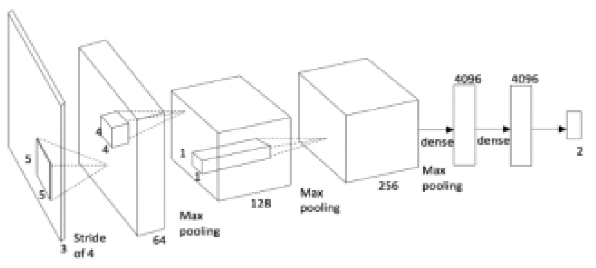
In the proposed system, the convolutional network model is constructed with some crucial parameters. The three convolution layers are implemented by the activation function of layers namely, Relu and max pooling layer. In this layer, it has filters. The kernel size is 2×2. The model is trained for four classes. Since there are four neurons in the output layer. The special activation function of this network for classifying the dataset is categorical-cross entropy.

3. Training

The training phase of this work has 6 epochs and 4000 training samples to implement the model for extracting crucial features and good training. The needed dataset of video frames are stored in a stack array and modified size as 150 x150. The stacked array of the dataset is changed into a batch file and provides data to the CNN model for the training process. The model extracts the features through epochs to detect the anomaly using separate labels (0,1,2,3) for anomalies namely.

Fig 1.1

	C ($\times10^6$)	**A (%)**	**A:C**	**fps**
Alexnet	71.9	91.7	1.3	4.0
FireNet	68.3	91.5	1.3	**17.0**
InceptionV1	6.1	**93.4**	15.4	2.6
InceptionV1-OnFire	**1.2**	**93.4**	**77.9**	8.4
Chenebert et al. [17]	-	-	-	0.16



4. Testing

The final phase of testing in the CNN model detecting the anomaly is in different video events is taken and converted these into frames. 100 datasets of video events are stored in a stack array of a batch file and modified size as 150x150. The event video frames are collected from different events namely Sports, Protest, Temple, etc. From each video, 30 frames are collected and stored for a test container. In that, 10 false datasets are collected from other videos and stored in the test container. From the batch file, the testing data is sent to the trained model. The CNN model finds four categories of the anomaly and shows anomaly name for each category correctly.

V.EVALUATION

For the comparison of the simplified CNN architectures out- lined we consider the True Positive Rate (TPR) and False Pos- itive Rate (FPR) together with the F-score (F), Precision (P) and accuracy (A) statistics in addition to comparison against the state of the art in non-temporal fire detection [6]. We ad- dress two problems for the purposes of evaluation:- (a) full- frame binary fire detection (i.e. fire present in the image as whole - yes/no?) and (b) superpixel based fire region local- ization against ground truth in- frame annotation [8].

Table-I: Statistical performance - full-frame fire detection.

	TPR	FPR	F	P	A
AlexNet	0.91	0.07	0.93	0.95	0.92
InceptionV1	0.96	0.09	0.95	0.94	0.93
VGG-13	0.93	0.11	0.93	0.92	0.91
FireNet	0.92	0.09	0.93	0.93	0.92
InceptionV1-OnFire	0.96	0.10	0.94	0.93	0.93

Table 2. Statistical results - size, accuracy and speed (fps). CNN training and evaluation was performed using fire im- age data compiled from Chenebert et al. [6] (75,683 im- ages) and also the established visual fire detection evalua- tion dataset of Steffens et al. [8] (20593 images) in addition to material from public video sources (youtube.com: 269,426 images) to give a wide variety of environments, fires and non-fire examples (total dataset: 365,702 images). From this dataset a training set of 23,408 images was extracted for training and testing a full-frame binary fire detection problem (70:30 data split) with a secondary validation set of 2931 im- ages used for statistical evaluation. Training is from random initialisation using stochastic gradient descent with a momen- tum of 0.9, a learning rate of 0.001, a batch size of 64 and categorical cross-entropy loss. All networks are trained using a a Nvidia Titan X GPU via TensorFlow (1.1 + TFLearn 0.3).

From the results presented in Table 1, addressing the full-frame binary fire detection problem, we can see that the InceptionV1-OnFire architecture matches the maximal performance of its larger parent network InceptionV1 (0.93 accuracy / 0.96 TPR, within 1% on other metrics). Further- more, we can see a similar performance relationship between the FireNet architecture and its AlexNet parent.

Computational performance at run-time was performed using at average of 100 image frames of 608 360 RGB colour video on a Intel Core i5 2.7GHz CPU and 8GB of RAM. The resulting frames per second (fps) together with a measure of architecture complexity (parameter complexity, C), percentage accuracy (A) and ratio A : C are shown in Table 2. From the results presented in Table 2,we observe significant run-time performance gains for the reduced com- plexity FireNet and InceptionV1-OnFire architectures com- pared to their parent architectures. Whilst FireNet provides a maximal 17 fps throughput, it is notable that InceptionV1- OnFire provides the maximal accuracy to complexity ratio. Whilst the accuracy of FireNet is only slightly worse than that of AlexNet, it can perform a classification 4.2 times faster. Similarly InceptionV1-OnFire matches the accuracy of InceptionV1 but can perform a classification 3.3× faster.

Table-II:Statistical results - size, accuracy and speed (fps).

Detection (full-frame)	TPR	FPR	F	P	A
Chenebert et al. [17]	0.99	0.28	0.92	0.86	0.89
InceptionV1-OnFire	0.92	0.17	0.90	0.88	0.89

Table –III: Statistical results - localization).

Localization (pixel region)	TPR	F	P	S
Chenebert et al. [17]	0.98	0.90	0.83	0.80
InceptionV1-OnFire	0.92	0.88	0.84	0.78

To evaluate within the context of in-frame localization (Section 2.3), we utilise the ground truth annotation available from Steffens et al. [8] to label image superpixels for train- ing, test and validation. The InceptionV1-OnFire architec- ture is trained over a set of

International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 2, Mar-Apr-2020, ISSN (Online): 2395-566X

54,856 fire (positive) and 167,400 non-fire (negative) superpixel examples extracted from 90% of the image frames within [23]. Training is performed as per before with validation against the remaining 10% of frames comprising 1178 fire (positive) and 881 non-fire (negative) examples. The resulting contour from any fire detected su- perpixels is converted to a bounding rectangle and tested for intersection with the ground truth annotation (Similarity, S: correct if union over ground truth>0.5 as per [23]).

From the results presented in Table 3 (lower), we can see that the combined localization approach of superpixel region identification and localized InceptionV1-OnFire CNN classi- fication performs marginally worse than the competing state of the art Chenebert et al. [6] but matching overall full-frame detection (Table 3, upper). However, as can be seen from Ta- ble 2, this prior work [6] has significantly worse computa- tional throughput than any of the CNN approaches proposed here. Example detection and localization are shown in Fig- ures 1 and 4B (fire = green, no-fire = red).

VI. CONCLUSIONS

Overall we show that reduced complexity CNN, experimentally defined from leading architectures in the field, can achieve 0.93 accuracy for the binary classification task of fire detection. This significantly outperforms prior work in the field on non-temporal fire detection [6] at lower complexity than prior CNN based fire detection [22]. Furthermore, reduced complexity FireNet and InceptionV1-OnFire architectures offer classification accuracy within less than 1% of their more complex parent architectures at 3-4_ of the speed (FireNet offering 17 fps). To these ends, we illustrate more generally a architectural reduction strategy for the experimentally driven complexity reduction of leading multi-class CNN architectures towards efficient, yet robust performance on simpler binary classification problems.

REFERENCE

[1]. K. Muhammad et al., “Secure surveillance framework for IoT systems using probabilistic image encryption,” IEEE Trans. Ind. Informat., to be published, doi: 10.1109/TII.2018.2791944.
[2]. T. Qiu, Y. Yan, and G. Lu, “An autoadaptive edge-detection algorithm for flame and fire image processing,” IEEE Trans. Instrum. Meas., vol. 61, no. 5, pp. 1486–1493, May 2012.
[3]. B. C. Ko, S. J. Ham, and J. Y. Nam, “Modeling and formalization of fuzzy finite automata for detection of irregular fire flames,” IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 12, pp. 1903–1912, Dec. 2011.
[4]. P. V. K. Borges and E. Izquierdo, “A probabilistic approach for vision based fire detection in videos,” IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 5, pp. 721–731, May 2010.
[5]. P. Foggia, A. Saggese, and M. Vento, “Real-time fire detection for video surveillance applications using a combination of experts based on color, shape, and motion,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 9, pp. 1545–1556, Sep. 2015.
A. Chenebert, T.P. Breckon, and A. Gaszczak, “A non-temporal texture driven approach to real-time fire detection,” in Proc. International Con- ference on Image Processing. September 2011, pp. 1781–1784, IEEE.
[6]. K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, “Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 2, pp. 339–351, Feb. 2015.
[7]. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” IEEE Trans Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274–2282, 2012.
[8]. Y. Li et al., “No-reference video quality assessment with 3D shearlet transform and convolutional neural networks,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 6, pp. 1044–1057, Jun. 2016.
[9]. K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, “Convolutional neural networks based fire detection in surveillance videos,” IEEE Access, vol. 6, pp. 18174–18183, 2018.
[10]. C.R. Steffens, R.N. Rodrigues, and C.S. da Costa Botelho, “Non- stationary vfd evaluation kit: Dataset and metrics to fuel video-based fire detection development,” pp. 135–151, 2016.
[11]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A.C. Berg, “Imagenet large scale visual recognition challenge,” Int. J. of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
[12]. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol.521, no. 7553, pp. 436, 2015.
[13]. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[14]. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.