International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

CNN MODEL FOR REAL-TIME VIDEO FIRE-SMOKE DETECTION

MITHUNKUMAR M
Electronics and Communication Engineering
Bannari Amman Institute of Technology
Erode,India
mithunkumar.ec19@bitsathy.ac.in

GOKULRAJU S
Electronics and Communication Engineering
Bannari Amman Institute of Technology
Erode,India
gokulraju.ec19@bitsathy.ac.in

KRISHNAKANTH M
Electronics and Communication Engineering
Bannari Amman Institute of Technology
Erode, India
krishnakanth.ec19@bitsathy.ac.in

***Abstract*—Convolutional neural networks are being used by vision-based systems to detect fire during surveillance thanks to recent developments in embedded processing (CNNs). However, the application of such algorithms in surveillance networks is constrained by the fact that they typically need more processing time and memory. In this study, we suggest a CNN architecture for surveillance movies that can efficiently identify fire. Given its fair computing complexity and adaptability for the desired purpose in comparison to other computationally expensive networks like AlexNet, the GoogleNet architecture served as the model's inspiration. The model is adjusted taking into account the nature of the target problem and fire data in order to strike a compromise between efficiency and accuracy. The efficiency of the proposed framework is demonstrated by experimental findings on benchmark fire datasets, which also confirm its appropriateness for fire detection in CCTV surveillance systems.**

I. INTRODUCTION (*HEADING 1*)

Smarter surveillance has been made possible by the increased embedded processing power of smart devices, opening up a variety of practical applications in various fields like e-health, autonomous driving, and event monitoring [1]. During surveillance, a variety of anomalous events, such as fire, accidents, disasters, medical emergencies, fights, and floods, may occur, making it crucial to obtain early information. This can significantly reduce the likelihood of major disasters and control an anomalous event promptly with relatively little potential loss. One of these often occurring anomalous events is fire, whose early detection during monitoring might prevent house fires and other fire disasters [2]. Physical impairment is the second-ranked cause of home fire deaths after other variables, affecting 15% of cases[3]. The US experienced 1345500 fires in total in 2015, according to the NFPA data. These fires caused $14.3 billion in losses, 15700 civilian fire injuries, and 3280 civilian fire fatalities. Additionally, there were civilian fire injuries and fatalities every 33.5 and 160 minutes, respectively. 78% of fire fatalities were solely attributable to residential fires [4]. One of the main causes is that disabled persons must wait longer to leave because conventional fire warning systems require strong fires or near proximity and do not promptly sound an alarm for such people. This calls for the installation of reliable fire alarming systems for monitoring. Given their low cost and ease of installation, vision sensors have been the foundation for the majority of fire alarming systems created to date. As a result, the majority of the study is focused on employing cameras to detect fires.

Since the 1990s, there has been research in this area. The literature has a number of video-based fire and flame detection methods. The bulk of these algorithms [1], [2], [3], [4], [5], [6] and [7] concentrate on the colour and shape properties in combination with the temporal behaviour of smoke and flames. Following that, the objective is to construct a rule-based algorithm or a multidimensional feature vector that is used as an input to a traditional classification algorithm, such as SVM, Neural Network, Adaboost, etc. In order to create appropriate feature extractors, traditional video fire detection algorithms rely on expert knowledge. The rule-based models and the discriminative characteristics must be developed by experts.

II. RELEVANT WORK

There are an increasing number of publications discussing video fire detection in the literature. The creation of efficient video fire detection systems has been significantly aided by a number of researchers [6]. A multi-sensor fire detector that merges visual and non-visual flame properties from moving objects was proposed by Verstockt [1]. He used standard video as well as thermal long wave infrared (LWIR) photos. He first extracts moving elements using a dynamic backdrop subtraction. Additionally, LWIR moving objects are filtered using hot object segmentation with histograms. With an emphasis on the specific geometric, temporal, and spatial disorder properties of flame regions, a set of flame features analyses these moving objects.

The probability of the bounding box disorder, primary orientation disorder, and histogram roughness of the hot moving objects in LWIR are then combined to create a LWIR flame probability. Similar to this, the same math is used to regular video to determine the likelihood of a video flame. In order to provide a conclusion regarding the presence of flames, he finally integrates the LWIR and the video flame probability. A four-step video-based detection technique is used by Toreyin [2]. By first performing a three-frame differencing operation to identify regions of valid motion, followed by adaptive background subtraction to extract the full moving region, he first approximated moving pixels and regions. Second, he employed an RGB colour space Gaussian mixture model to identify fire-coloured pixels.

The distribution of fire colours is derived from examples of photos that feature fire zones. A temporal wavelet transform is used in the third stage to analyse the flame flicker. In order to assess colour fluctuations in pixel values, a spatial wavelet analysis of moving regions comprising fire mask pixels is carried out. Prerequisite for significant spatial differences is a fire region. Two models were created by Celik [3], one for smoke detection and the other for fire detection. The old heuristic rules were replaced with a fuzzy logic model with rules. Due to this decision, the classification was better able to distinguish between actual fire and coloured things that resemble fire. A statistical analysis was conducted for smoke detection based on the hypothesis that smoke exhibits a greyish colour under various lighting conditions.

III. Convolutional Neural Network

A deep learning network architecture that learns directly from data is a convolutional neural network (CNN or ConvNet). CNNs are very helpful for recognising objects, classes, and categories in photos by looking for patterns in the images. They can be quite useful for categorising signal, time-series, and audio data.

Multiple layers make up a convolutional neural network.

A. Convolutional layers:

It serves as the foundation of CNN. These layers are made up of a grid of rectangular neurons with a narrow receptive field that extends over the entire depth of the input. As a result, the convolutional layer is simply the preceding layer's image convolution, with the convolution filter determined by the weights.

B. Pooling layers:

There might be a pooling layer after each convolutional layer. Subsampling is used in pooling layers' input. This pooling can be done in a variety of ways, such by taking the mean, maximum, or a learned linear combination of the neurons in the block. For instance, the maximum pooling for a 2*2 window is shown

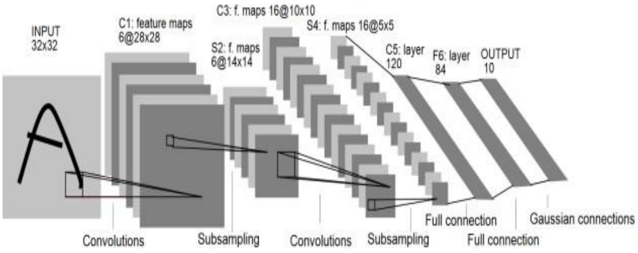


Fig .1 . LeNet-5, a Convolutional Neural Network for digits recognition

C. Fully Connected Network:

After a number of convolutional and max pooling layers, fully connected layers are used in the neural network to perform the high-level reasoning.

Every layer in convolutional neural networks serves as a detection filter to check for the presence of particular characteristics or patterns in the raw input. The initial layers of a CNN find traits that are reasonably simple to identify and analyse. More abstract traits are being detected by later layers. By combining all of the distinctive qualities in the input data that were identified by the earlier layers of the CNN, the final layer is able to classify data in an incredibly precise manner. The proposed CNN architecture for video fire and smoke detection is provided in the following section.

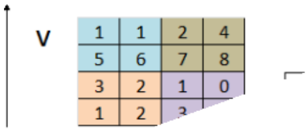


Fig.2. Max pooling in CNN

IV. CNN for Video Fire and Smoke Detection

A. Structure:

Our classification architecture combines convolution with Max pooling in a traditional convolutional neural network fashion. However, we select a tiny network to obtain a quick classification.

A 3x3 convolutional kernel is used in two convolutional processes that are applied sequentially to an RGB colour image. Layer three is followed by layer two of the same structure. The convolutional layer two and five are followed by a Max pooling 3x3 with stride 2. There are 16 feature maps in layers one through four. There is only one feature map for layers five and six. Layers 7 and 8 are totally interconnected. A three-way Softmax receives the output of the final fully connected layer and generates a distribution over three class labels.

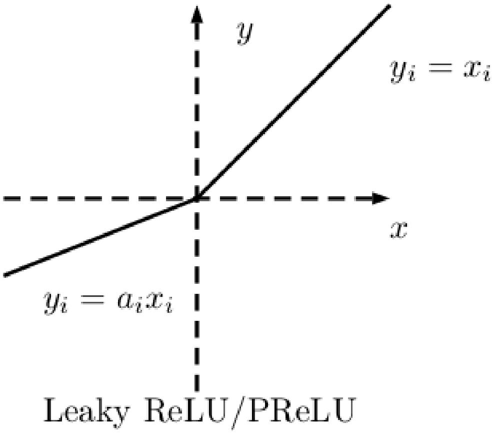


Fig .3. Leaky ReLu

B. Training:

Our classification is to determine whether a picture contains smoke or fire. The classifier is trained using a set of labelled photos to address this issue. Additionally, we want to pinpoint where the fire and smoke are in a picture or a video. The training set consists of 27919 64x64 pixel RGB tagged images. 7257 for fire, 11752 negative, and 8915 for smoke (no fire or smoke). We divide the photos into 3 subsets: training 60%, validation 20%, and test 20%. The training data was created using a PC with an Intel Xeon microprocessor (frequency CPU 3,1GHz, RAM 16Go), and a GTX 980 Ti graphics card (2816 cores, 6 GB memories).

We employed a stochastic gradient descent (SGD) method with 100-miniature batches. The network's weight is initialised at random. Momentum is 0.9 and the starting learning rate is 0.01. Every five epochs, the learning rate declines by a factor of 0.95. On the other hand, the momentum rises to 0.9999. Several trials were conducted to find the settings with the best accuracy.

V. RESULTS

On the test set, the classification accuracy is 97.9%. The test set includes Fig. 4, 1758 smoke images, and 1427 fire photographs. 2399 negative photographs from CNN architecture, or 5584 images. The confusion matrix for each class is shown in Tables 1 to 3. The false negatives and false positives on the fire confusion matrix do not have smoke images. In a similar vein, the fire picture for false negatives and false positives is not present in the smoke confusion matrix. We can draw the conclusion that the parameters of our CNN model enable accurate classification of fire and smoke.

Fire	True Class		
		True	False
Hypothesis Class	True	1400	3[a]
	False	27[a]	4154

TABLE .1. FIRE CONFUSION MATRIX

Smoke	True Class		
		True	False
Hypothesis class	True	1698	3[a]
	False	60[a]	3800

TABLE .2. SMOKE CONFUSION MATRIX

No Smoke/Fire	True Class		
		True	False
Hypothesis class	True	2370	87[a]
	False	29[b]	3098

TABLE.3. NEGATIVE CONFUSION MATRIX

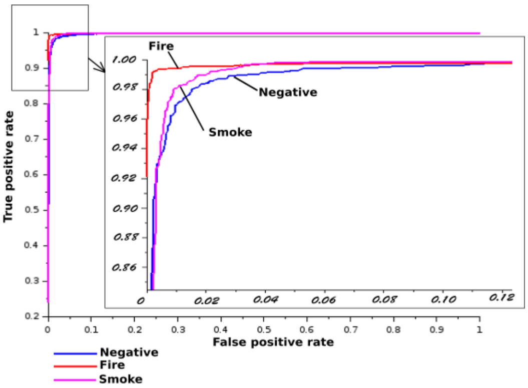


Fig.4. ROC Curves for Smoke, Fire and Negative

Our goal is to identify a fire on a video or identify a fire that has already started. The accuracy of the detection depends significantly on the processing time. As a result, we opt to employ the "light structure" shown in Fig. 4. In practise, sliding windows are used to identify and categorise objects on either the original or altered images. To classify these windows, a convolutional neural network and fully linked layers are used. The window location must be changed and the convolutional neural network must be run once again in order to assess the complete image of a video frame. Instead of moving a 64x64 pixel window around the image to find the fire and the smoke, we use a very different strategy and choose to focus on the last feature map.

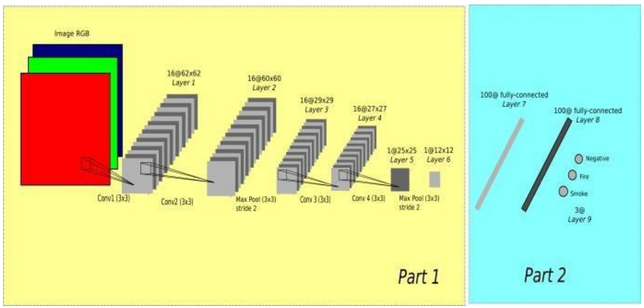
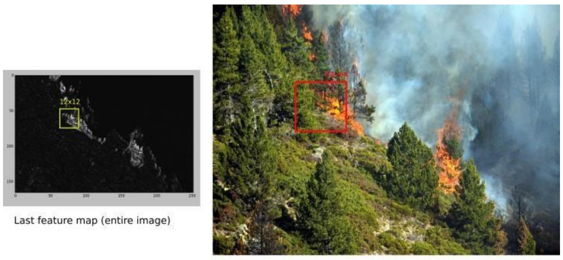


Fig .5. Two Parts of CNN

Part 1: 6 Layers Part 2: Two fully Connected Layers and the Output Layers

We assess the final feature map (layer 6) of the entire image using the first segment of the network. We are aware that a sliding window with a size of 64x64 pixels in the RGB image equates to a window size of 12x12 pixels in the final feature map based on the CNN's network structure.



Last feature map (entire image)

Input image RGB

Fig .6. Sliding Window

We apply a sliding window of size 12x12 pixels to the final feature map in order to detect the fire and smoke in a frame video (see Fig. 8). Utilizing the GPU of the graphic card, we realise a tensor 12x12x1xN (N: number of windows) from the most recent feature map to accelerate the prediction for each 12x12 window. With this approach, the precision appears to be unaffected, and the speed of detection and prediction rises in accordance with the original size image and the anticipated number of windows, asdemonstrated.

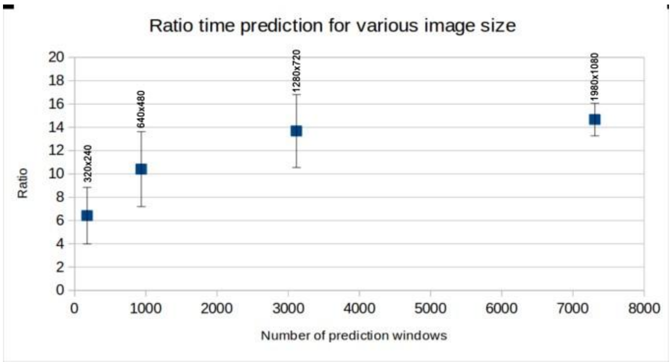


Fig.7. Ratio Time Prediction for various Image

Ratio is (time prediction for the complete original image) / (time prediction on the last feature map plus time to create the last feature map). sliding windows on the final feature map with a step of 4 pixels and 16 pixels on the original image. a video's more than 200 frames were processed.

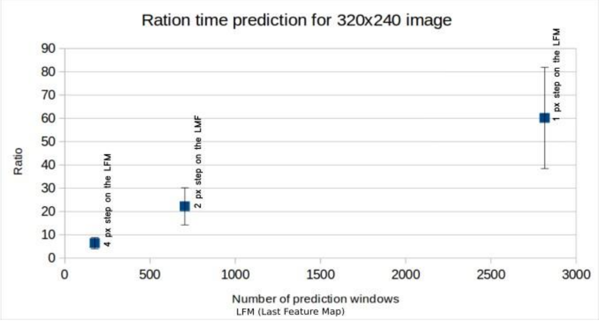


Fig.8. Ratio time prediction for various sliding stages and 320x240 image sizes.

The feature map was moved with windows measuring 12x12 pixels and a step of 2 pixels to produce the classification and localization results shown in Figs. 11.1 and 11.2. The detection and localization mask is shown in Figs. 9.1(c) and 9.2(c). Green colour indicates smoke detection, whereas red colour indicates a fire. The likelihood of detecting smoke or fire affects how intensely the colours red and green are. It is possible to project the locations of the fire and smoke onto the RGB original image by knowing their localisation in the feature map .

VI. CONCLUSION

This research presented a vision-based fire and smoke detection technique. The suggested algorithm employs a convolutional neural network-based deep learning strategy (CNN). The confusion matrix and ROC curves show that the detection stage's overall accuracy is very high. We demonstrated that scanning the feature map immediately during the detection test as opposed to the entire source frame could reduce the time cost by a factor of 6 to 60.

We hope to enhance the strategy in upcoming work by utilising a 3D convolutional neural network. In fact, CNN can only currently accept 2D inputs, which forces us to only process video input frame-by-frame. In contrast, 3D CNN uses 3D convolutions to extract characteristics from both the spatial and temporal dimensions. As a result, it was possible to encode the motion information of fire and smoke, which significantly reduced the time cost. Additionally, we must enhance our training set in order to maximise the identification and localisation of smoke and fire on a video. Because of his shape and texture, smoke is more challenging to locate and detect. Our model can only detect red fires; in order to detect other fire colours, we must expand our training set to include additional fire colours like blue, etc. We also intend to compare our algorithm to more established techniques using a larger range of video fire images with various materials, sources, andventilations.

REFERENCES

[1] S. Verstockt, A. Vanoosthuyse, S. Van Hoecke, P. Lambert, and R. Van de Walle, Multi-sensor fire detection by fusing visual and non-visual flame features, In Proceedings of International Conference on Image and Signal Processing, June 2010, pp. 333 –341.

[2] B. U. Toreyin, Y. Dedeoglu,U. Gudukbay, A. E. Cetin, ,Computer vision based method for real-time fire and flame detection,Pattern recognition letters, 2006, 27,1,pp. 49-58.

[3] Barmpoutis, P., Dimitropoulos, K., Grammalidis, N.: Real time video fire detection using spatio-temporal consistency energy. In: 10th IEEE international conference on advanced video and signal based surveillance, pp. 365–370 (2013)..

[4] K. Borges, P. Vinicius, J. Mayer and E. Izquierdo, Efficient visual fire detection applied for video retrieval, Signal Processing Conference, 2008 16th European, IEEE, 2008.

[5] Girshick, R., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE CVPR, pp. 580–587 (2014)

[6] P. Gomes, P. Santana and J. Barata, A vision-based approach to fire detection, International Journal of Advanced Robotic Systems, 09-2014.

[7] E. Çetin et al, Video fire detection – Review, Digital Signal Processing, Volume 23, Issue 6, December 2013, pp. 1827-1843

[8] Hossain, S., Lee, D.J.: Deep learning-based real-time multiple object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. Sensors **19**(15), 3371 (2019)

[9] D. H. Hubel and T. N. Wiesel, Ferrier lecture: Functional architecture of macaque monkey visual cortex, Proceedings of the Royal Society of London, Series B, Biological Sciences, 1977, 198(1130):pp. 1–59.

[10] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, in Proceedings of the IEEE, vol. 86,no.11,pp.2278-2324,N