iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54039

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Real Time Fire and Smoke Detection System

Dr Shallu Bashambu[1], Anupam Gupta[2], Sarthak Khandelwal[3]
[1]*Assistant Professor,* [2,3]*Department of IT Maharaja Agrasen Institute of Technology Delhi, India*

Abstract: ***This research paper presents a real-time fire and smoke detection system using the YOLOv5 object detection algorithm. The system aims to detect fire and smoke in images and video streams captured by a camera in real-time, without the need for any preprocessing or manual intervention. The proposed system uses the YOLOv5 algorithm to detect the fire and smoke regions in the input images and videos. The YOLOv5 model is trained on a dataset of annotated images to recognize fire and smoke patterns accurately. The proposed system has been tested on different datasets and has achieved high accuracy and precision in detecting fire and smoke in real-time. The experimental results demonstrate that the proposed system is robust and efficient, and it can detect fire and smoke in real-time with high accuracy and low latency. The proposed system can be used in various applications, such as early warning systems, fire safety, and disaster management. It can also be integrated with the CCTV network directly.***

Index Terms: Yolo, Convolutional neural network, Computer vision, image classification, image recognition, image generation, real-time detection

I. INTRODUCTION

Fire has played a significant role in the progress of human civilization. On the contrary, it is one of the major disasters resulting in huge loss of human lives and property all over the planet. Early fire detection can help to alert and prevent these disasters, save both lives and property. Usually, the combustion of objects begins with smoke before catching fire. Thus the presence of smoke can be used as an indicator to detect fire earlier.

Fire incidents have been on the rise for the past few years in India. On average, fire-related accidents killed 35 people every day in the five years between 2016 and 2020, according to a report by Accidental Deaths and Suicides in India (ADSI), maintained by the National Crime Records Bureau. Properties worth millions of rupees are gutted due to such incidents.

To reduce such disasters, detecting fire at an early stage without any false alarms is crucial. As the traditional detection systems proved to be ineffective in outdoor environments, various autonomous fire detection technologies are being developed, and are widely used in real life. With the recent uprising of artificial intelligence and its subsidiaries, several detection systems based on computer vision came into existence. These vision-based detection systems overcome the limitations of size, location and have larger surveillance coverage. They also offer faster detection, fewer false alarms and reduce human intervention. Researchers have invested significant effort to address the issues of system complexity and false detection in terms of computer vision technology.

Hence, we intend to use this advancement to the benefit of the society and develop a simpler Fire and smoke detection system using Convolutional Neural Networks that can inspect an image and give out the decision whether the image contains smoke or fire.Data augmentation has been used to increase the amount of data by adding slightly modified copies of already existing data[Derived from original images with some sort of minor geometric transformations (such as flipping, translation, rotation, or the addition of noise) in order to increase the diversity of the training set.] or creating new synthetic data from existing data. It acts as a regularizer and helps reduce overfitting when training the machine learning model. This system can be used to save manual labour and can be seamlessly integrated with existing surveillance systems without further need of hardware if the model is tweaked a bit to take video as an input as well.

II. RELATED WORKS

Various methods have been developed to detect fires from videos and images. This section describes some of the significant work done in the field of fire detection from his images and videos using computer vision and neural networks. Sebastian Frisi et al. [1] proposes his convolutional neural network (CNN) to detect fires in videos. The convolutional neural network has proven to be very good at object classification. Within the same architecture, the network has the ability to perform feature extraction and classification, and shows great promise for using CNNs to detect fires in videos.

To improve the precision of smoke detection, a novel method based on Deep Convolutional Neural Networks is suggested [2]. It automatically extracts features from photos and may be trained entirely from the initial raw pixel values to the classifier output.

©IJRASET: All Rights are Reserved

Extracted Smoke detection from a visual situation is noteworthy since it offers noteworthy advantages. Due to the enormous range of colours, textures, and shapes, it is still difficult to identify smoke accurately from a visual environment.

The manual process of manually extracting features from input photos and training classifiers for classification used in traditional smoke detection or classification techniques is frequently confusing, laborious, and difficult. Traditional techniques perform noticeably worse, especially for huge image data sets. A novel method built on CNN has been suggested to increase the accuracy of smoke detection since it can be trained from the beginning to the end, from raw pixel values to classifier outputs, and automatically extract features without the need for laborious image pre-processing.

The results of experiments demonstrate that this method outperforms conventional methods on large datasets, reaching the state-of-the-art level on small datasets with low false alarm rates.

In paper [3] adaptive background subtraction to detect foreground moving objects the proposed video fire-detection system is used and then verified by the rule based fir colour model to determine whether the detected foreground object is a fire or not.

Paper [4] focuses on the challenge related to the actual deployment of the vision system. The background subtraction is performed in a windowed way for improved accuracy, an attentive mechanism is employed to focus a computationally expensive frequency analysis of potential fire regions, interaction with a people detection and tracking system is inclined so as to enable model-based false alarms rejection, a new colour-based model of fires appearance as well as a new wavelet-based model of fires frequency signature are proposed. Binary hashes and per-block histograms are described in this paper [5] by Tsung-Han Chan et al. The basic data processing component Main Cascade Component (PCA), PCA is used to train multistage filter banks in this proposed His architecture. This was accomplished with two simple variants of his PCANet called RandNet and LDANet. PCANet shares the same topology as the dataset above, but the cascaded filters are randomly selected or learned from LDA. They also extensively tested these basic networks on a number of visual benchmark datasets for various tasks such as: B. His LFW for facial recognition. MNIST for handwritten digit recognition, multiples, extended Yale B, AR, FERET records for faces. The potential of PCANet to serve as a simple competitive foundation for texture classification and object detection can be demonstrated by additional experiments using other public datasets.

The research [8] suggests a novel approach to interpret video data produced by a typical camera watching a scene in order to identify fire and/or flames in real-time.By analysing the video in the wavelet domain, flames and flame flickering are detected in addition to regular motion and colour cues.Using a temporal wavelet transform, the quasi-periodic behaviour at the frame boundaries is captured.By computing the spatial wavelet transform of the region of moving flame colour, the colour change of the flame region is captured.The inconsistencies at the borders of the zones coloured in fire are another indicator employed in the fire detection method.To reach a conclusion, all of the aforementioned hints are merged.The proposed method is quite effective at detecting fire and/or flame, according to experimental data.Additionally, compared to techniques that rely solely on movement and colour cues, it significantly lowers false alarms caused by ordinary fire-coloured moving objects.

Early warning is essential to reduce loss of life and property due to fire. A fire alarm system based on light detection and analytics is proposed in article [9]. The system uses the HSV and YCbCr colour models under certain conditions to separate orange, yellow, and bright lights from background and ambient lights. Flame growth is analysed and calculated based on frame deltas. Overall accuracy for the experiment was greater than 90%.

III. METHODOLOGY

A. Convolutional Neural Network

Convolutional Neural Networks (commonly referred to as CNNs or ConvNets) are a subclass of neural networks that are particularly well-suited for processing data with grid-like topologies, such as image data.A binary representation of visual data is a digital picture.It has a grid-like arrangement of pixels, and the brightness and colour of each pixel are represented by pixel values.

The moment the human brain sees a picture, it begins processing a vast quantity of data.

The complete visual field is covered by the connections between each neuron, each of which operates in its own receptive field.

Similar to how each neuron in the human visual system responds to stimuli exclusively in a certain region of the visual field known as the receptive field, each neuron in a CNN only processes data in its receptive field.Lines, curves, and other basic patterns are recognised by the first layer, followed by more complicated patterns (surfaces, objects, etc.).

Computers can sight using CNN.

B. Convolutional Neural Network Architecture

CNNs consist of basically three layers: convolutional layers, pooling layers, and fully connected layers.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

C. Convolutional Layers

The foundational units of CNNs are convolutional layers.It carries the majority of the network's computational burden.

The dot product between two matrices is performed by this layer. When the limited area of the receptive field is represented by one matrix, which is the set of parameters that may be learned, or the kernel.

Although more detailed, kernels are spatially smaller than pictures.This implies that if the image is made up of his three (RGB) channels, the depth will cover all three of them while the height and breadth of the core will be spatially smaller.

The kernel slides the picture's height and width during the forward pass to provide an image representation of its receptive region.

An activation map, a two-dimensional representation of her in the image is the result of this. It displays the reaction of the kernel for every pixel in the picture.Stride refers to the kernel float size.

The formula: yields the size of the output volume given an input of size W x W x T, number of kernels Dout, space size F, increment S, and fill P.

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

Formula for Convolution Layer

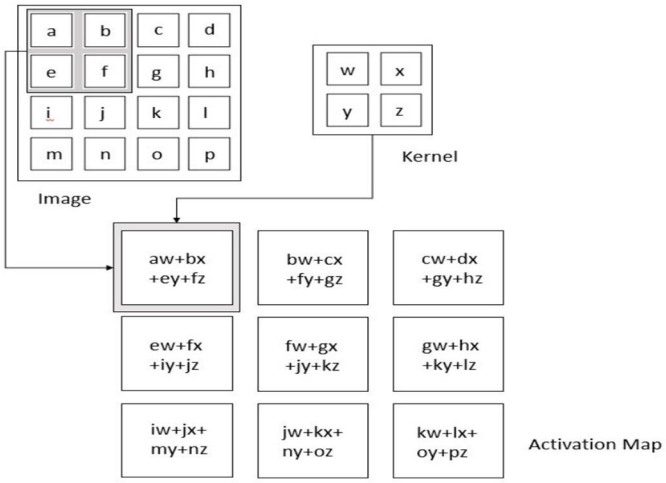


Figure 3: Convolution Operation

D. The Motivation Behind Convolutions

Researchers interested in computer vision have been driven by three fundamental concepts: sparse interactions, parameter sharing, and equivariant representations.

Let's go into further depth about each one.

With a matrix of parameters reflecting interactions between input and output units, the simple neural network layer multiplies numbers using matrix multiplication. This implies that all input and output devices communicate with one another.

Convolutional neural networks, on the other hand, show sparse interactions. Making the kernel smaller than the input does this. When an image is processed in the kernel, we may identify relevant information made up of tens or hundreds of pixels, even though the picture may include millions or thousands of pixels.Less parameters must thus be stored, which increases the statistical effectiveness of the model while simultaneously reducing the memory footprint of the model.

If computing features at one spatial location (x1, y1) is valuable, then computing features at subsequent spatial points ((x2, y2), etc.) should also be useful.

In other words, one 2D slice of neurons.

H.The activation map should be produced using the same set of weights. Convolutional networks share parameters, unlike standard neural networks, which only employ each weight matrix member once.

H. The weight used to one input to obtain the output is the same as the weight applied elsewhere.

Convolutional neural networks have the attribute of layers having identical variance to transformations because to parameter sharing.It claims that when the input is altered, the output is also altered.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

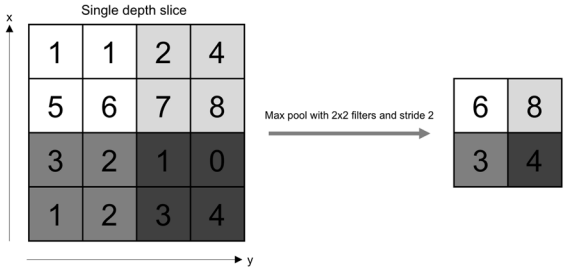


Figure 4: Pooling Operation (Source: O'Reilly Media)

The following formula may be used to calculate the size of the output volume if we have an activation map of size W x W x D, a pooling kernel of spatial size F, and stride S:

$$W_{out} = \frac{W - F}{S} + 1$$

An output volume of size Wout x Wout x D will result from this.
With pooling, an item would always be recognisable regardless of where it is in the screen because there is some translation invariance.

E. Fully Connected Layer

As in a typical FCNN, all of the neurons in this layer are completely linked to every neuron in the layer before it and the one after it. Thus, it may be estimated as per normal by matrix multiplication and subsequent bias effects.
The FC layer facilitates the mapping of input and output representations.

F. Non-Linearity Layers

The nonlinearity level is frequently added right after the convolution level and imparts nonlinearity to the activation map since convolution is a linear operation and the picture is anything but linear.
Nonlinear operations come in a variety of forms, the most popular of which are:

1) Sigmoid

The mathematical formula for the sigmoid nonlinearity is $\sigma(\kappa) = 1/(1\ e^{-}\kappa)$.
Squeeze an actual number into the range from 0 to 1.
However, the fact that the gradient is almost zero when activation occurs at either end of the sigmoid is a highly undesirable characteristic.Backpropagation effectively "kills" the local gradient when it gets very modest.
The output of the Sigmoid will either be all positive or all negative if the input entering the neuron is always positive, leading to a zigzag dynamic of weight gradient updates.

2) Tanh

Tanh reduces the range of real numbers to [-1, 1].Their activity is saturated, just like sigmoids, but their output is zero-centred, unlike sigmoid neurons.

3) ReLU

In recent years, the Rectified Linear Unit (ReLU) has gained a lot of popularity.
Find the value of the function $f(\kappa)$=max (0,κ).Or to put it another way, activation is only a threshold of zero.ReLU converges six times faster and is more dependable than Sigmoid and Tanh.ReLU can, unfortunately, be susceptible during training, which is one disadvantage.The neuron may update to the point where it no longer updates due to a strong gradient running through it.By choosing an acceptable learning rate, it may be accommodated.

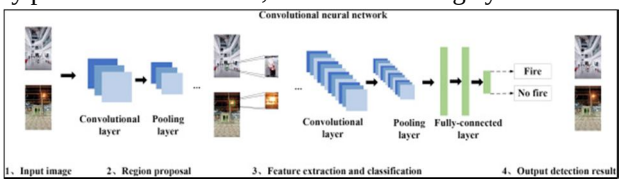
G. APPLICATIONS

Here are a few current uses for convolutional neural networks.

1) With CNN, we now have advanced models like R-CNN, Fast R-CNN, and Faster R-CNN that serve as the main pipeline for many object identification models used in autonomous cars, facial detection, and other applications.

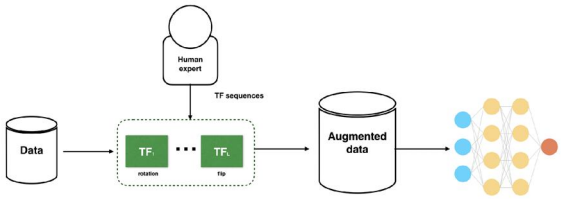
2) Semantic Segmentation: In order to include rich data into an image segmentation model, a team of academics from Hong Kong created a CNN-based Deep Parsing Network in 2015. Additionally, UC Berkeley researchers developed fully convolutional networks that outperformed cutting-edge semantic segmentation.

3) Recurrent neural networks and CNNs are used to provide subtitles for photos and videos. Applications for this include activity detection and providing descriptions of films and pictures for visually impaired people. In order to make sense of the enormous volume of videos that are routinely posted to the network, YouTube has largely utilised it.



H. Data Augmentation

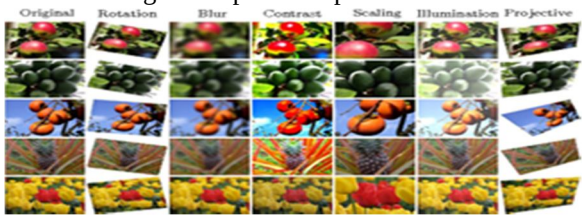
By creating additional data points from existing data, a group of techniques known as "data augmentation" may be used to artificially enhance the amount of data.This involves making minor adjustments to the data or creating new data points using deep learning models.



I. Importance

Deep learning applications in particular continue to vary and grow quickly in number.Data augmentation strategies, for example, are data-centric approaches to model creation that might be a useful tool for overcoming obstacles in the artificial intelligence field.By creating additional and distinct instances for training datasets, data augmentation helps machine learning models perform better and produce better results.A machine learning model performs better and is more accurate if the dataset is large and sufficient.

Data collection and labelling may be time-consuming and expensive operations for machine learning models.

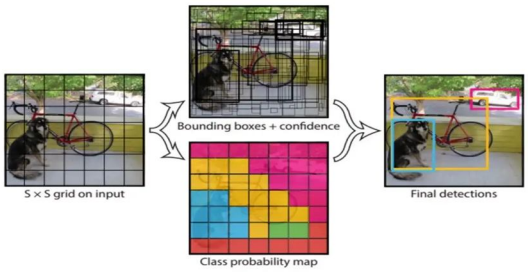


J. YOLO - You only look once

In this model of object detection, a convolutional neural network predicts numerous bounding boxes and probabilities of defined and labelled classes for those boxes. YOLO (version 5 used) trains on complete images and optimises the detection process. This algorithm has quite a few benefits over conventional models of object-detection. Foremost is the increased speed. As we label the images in the dataset, to teach the computer how and what to detect, we do not need the complex layers and algorithms. We just deploy our model on a fresh image when testing to predict the presence of desired objects. The network runs on several frames per second (usually 45 or more) even without batch processing on several common GPUs. Advanced systems can, however, process more than 150 frames per second. This also shows that we can manage object detection in video streams as well as real time object detection with latency of less than 25 milliseconds.

©IJRASET: All Rights are Reserved

Second, when making predictions, YOLO thinks broadly about the image.Because YOLO sees the full image during training and testing, unlike sliding window and region proposal-based approaches, it implicitly stores contextual information about classes in addition to their appearance.Fast R-CNN, a popular approach for object detection, misinterprets background patches in an image as objects because it lacks context awareness.Compared to Fast R-CNN, YOLO creates fewer than half as many background mistakes. Thirdly, this model understands the generalised forms of objects and items. Hence, when trained on real world images and tested on artwork, YOLO outdoes the other detection models such as DPM and R-CNN by a great margin.



K. Network Architecture and Training

It's intriguing to alter loss functions to achieve better outcomes.Two things are noteworthy:

1) Differential weighting for predictions with high confidence from training boxes with and without objects.

2) Calculate the square root of the width and height of the enclosing box to penalise errors in small and large objects differentially.

L. Limitations of YOLO

Since each grid cell can only predict two boxes and only have one class, YOLO places substantial spatial limits on bounding box predictions.The number of nearby objects that our model can predict is constrained by space.Small items that occur in groups, like flocks of birds, are difficult for our model to handle.Our approach has trouble applying to objects with novel or odd aspect ratios or configurations since it learns to estimate bounding boxes from data.Since our design contains numerous down - sampling layers from the input image, our model also uses rather coarse data to predict bounding boxes.Last but not least, even though we train on a loss function that roughly represents detection performance, our loss function handles errors in small bounding boxes and large bounding boxes equally.

IV. RESULT

Hence by using Convolutional Neural Networks, we were able to develop a software to distinguish between fire and smoke images in real-time. By using data augmentation, we increased the size of the dataset so as to train our model better. At the end of our efforts. The screenshots of said results have been presented here. Now after completion we may visit the several avenues of the project's future scope to determine what is achievable presently and work on it.

A. Webcam Detection of Fire and Smoke in Real-Time

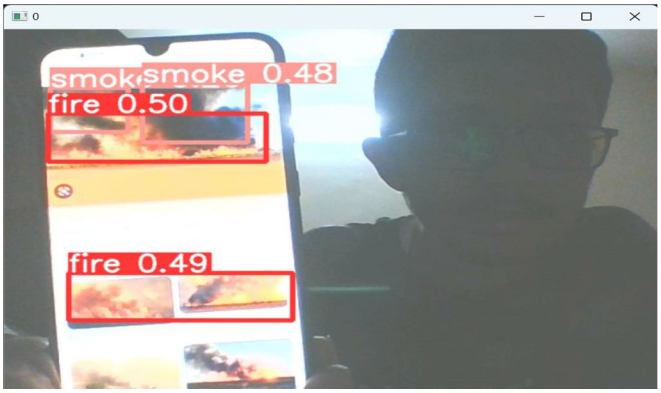


Image - 1

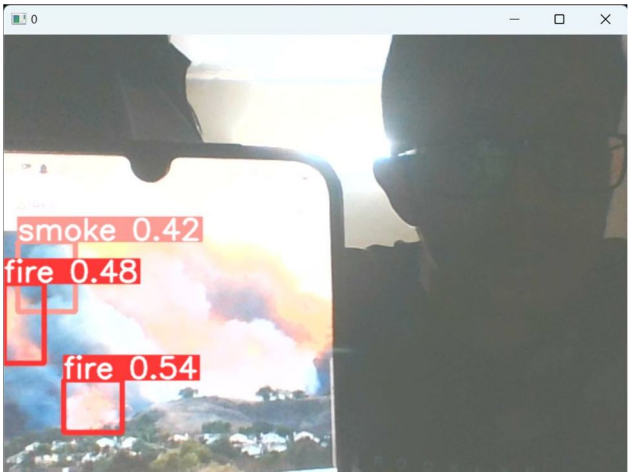


Image - 2

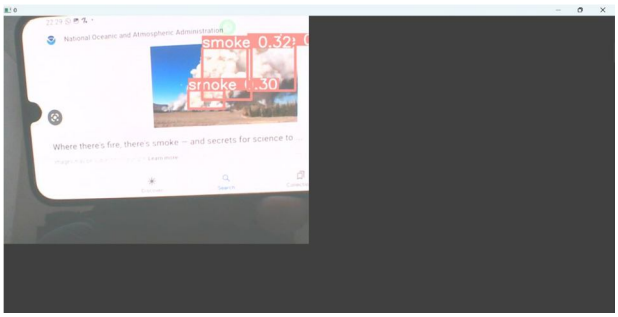


Image - 3

V. FUTURE SCOPE

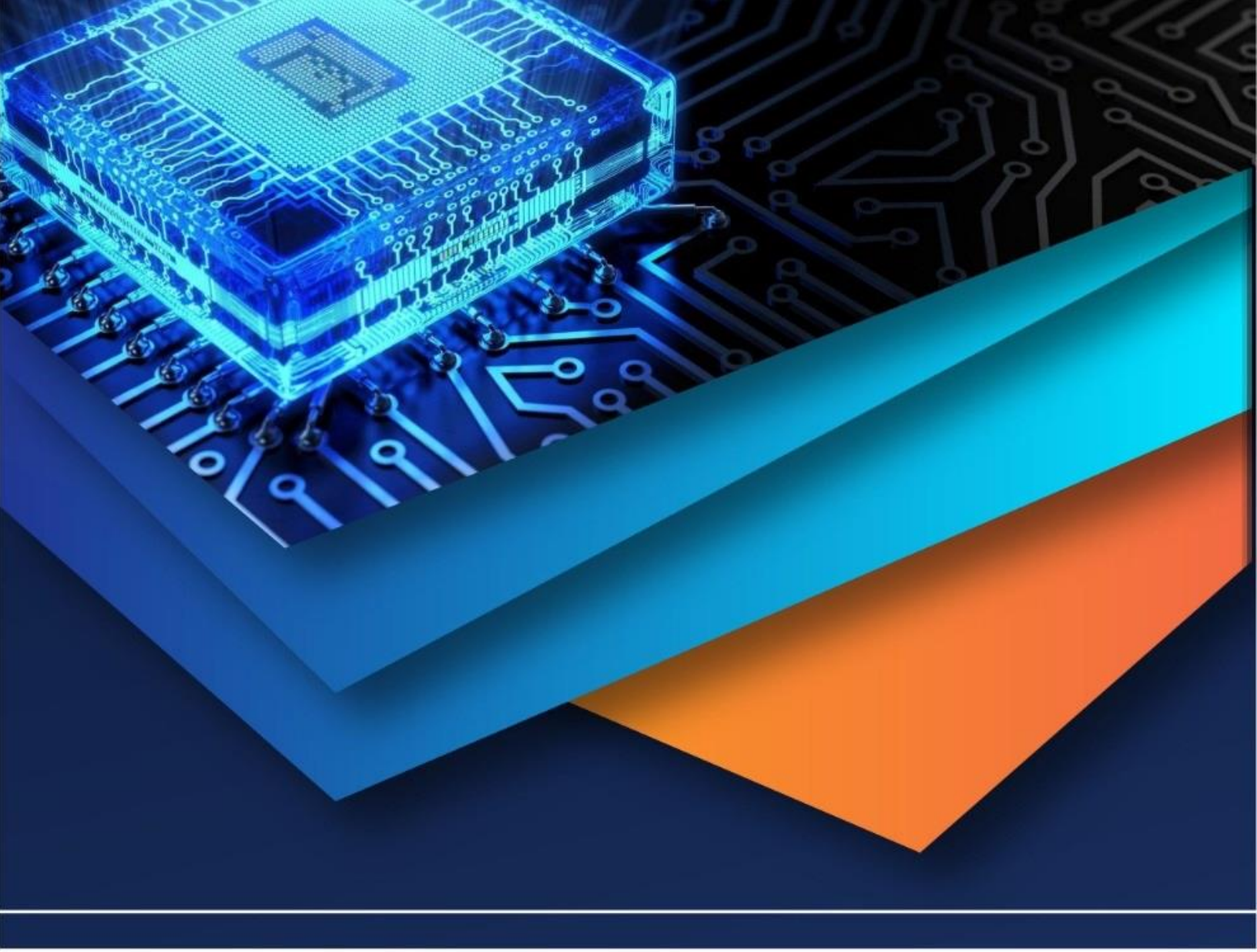
The system being designed shows tremendous scope for further enhancement and development:
The accuracy of image classification can be enhanced by using more customised datasets of a greater size and variety.
Sensors can be integrated with the system to avoid false alarms and improve efficiency.
An alarm system can be amalgamated with the system so that it can call the emergency services as soon as it detects fire and prevent loss of life and property in case nobody is there to report it.

REFERENCES

[1] S. Frizzi, R. Kaabi, M. Bouchouicha, J. M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016.
[2] C. Tao, J. Zhang, and P. Wang, "Smoke detection based on deep convolutional neural networks," IEEE International Conference on Industrial Informatics, 2016.
[3] A. E. Gunawardena, R. M. M. Ruwanthika, and A. G. B. P. Jayasekara, "Computer vision based fire alarming system," IEEE 2016.
[4] P. Santana, P. Gomes, and J. Barata "A vision-based system for early fire detection," 2012 IEEE International Conference on Systems, Man, and Cybernetics, pp. 14- 17, 2012 October.
[5] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, "PCA Net: A simple deep learning baseline for image classification," IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5017-5032, December 2015.
[6] T. Celik, and K. K. Ma, "Computer vision based fire detection in color images," IEEE Conference on Soft Computing in Industrial Applications (SMCia/08), June 2008.
[7] J. Sharma, O. C. Granmo, M. Goodwin, and J. T. Fidje, "Deep convolutional neural networks for fire detection in images," Springer International Publishing, 2017.
[8] B. U. Toreyin, Y. Dedeoglu, U. Gudukbay, and A. E. Cetin "Computer vision based method for real-time fire and flame detection," Elsevier, 2015.
[9] J. Seebamrungsat, S. Praising, and P. Riyamongkol, "Fire detection in the buildings using image processing," 3 rd ITC International IEEE Conference on Student Project, 2014.
[10] T. H. Chen, P. H. Wu, and Y. C. Chiou, "An early fire detection method based on image processing," IEEE International Conference on Image Processing, 2004

[11] https://www.kaggle.com/datasets/kutaykutlu/forest-fire
[12] Zeiler, Matthew D., and Rob Fergus. "Visualising and understanding convolutional networks." European conference on computer vision,pp. 818-833, 2014. doi: 10.1007/978-3-319-10590-1_53
[13] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Delving deep into rectifiers: Surpassing human-level performance imagenet classification." Proceedings of the IEEE international conference on computer vision, pp. 1026-1034. 2015. doi: 10.1109/ICCV.2015.123.
[14] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016. doi: 10.1109/CVPR.2016.90.
[15] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." Cvpr, 2015. doi: 10.1109/CVPR.2015.7298594.
[16] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014
[17] Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition, vol. 1, no. 2, p. 3. July, 2017. doi: 10.1109/CVPR.2017.243
[18] https://arxiv.org/pdf/1506.02640v5.pdf
[19] https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)