ABSTRACT

Vision based fire detection framework has lately picked up popularity when contrasted with customary fire recognition framework dependent on sensors. The need of video perception at private, Modern, business regions and woods areas has expanded the use of vision-based fire acknowledgment system. Recently lots of fire related accidents has occurred due to improper Surveillance or unable to cover those uncertain regions like restricted areas in forest or any factory buildings. In order to overcome such accidents , method using Convolutional neural networks (CNN) to detect fire. The critical issue with CNN-based fire disclosure structures is their execution progressively frameworks, in view of their high memory and computational necessities for usage. A unique, well disposed, and computationally effective CNN design, using You only look once (YOLOv3) at an image to predict what objects are present using a single convolution network. which reduces the computational time and cost and improves the accuracy and reduces false alarm.

LIST OF ABBREVIATIONS

CNN – CONVOLUTIONAL NEURAL NETWORK

RGB – RED GREEN BLUE

R-CNN - REGION BASED CONVOLUTIONAL NEURAL NETWORKS

1. INTRODUCTION

1.1 FIRE DETECTION

Disaster management, as a hybrid research area, has attracted the attention of many research communities such as business, computer science, health sciences, and environmental sciences. According to federal emergency management agency policy, there are two main categories of disaster: (1) Technological such as emergencies related to hazardous materials, terrorism, and nuclear power plants etc., and (2) Natural such as floods, earth quakes, and forest fires etc. Regardless of the nature of the disaster, certain characteristics are necessary for effective management of almost all of them. These features include prevention, advance warning, early detection, early notification to the public and concerned authorities, response mobilization, damage containment, and providing medical care as well as relief to affected citizens.

Fire has become a severe threat because of its high frequency and destructive nature. Fire spreads quickly and is difficult to control in a short time, especially in places where combustibles are densely packed, such as residential areas, airports, and forests. To avoid large-scale disasters caused by fire, timely and accurate fire detection is crucial.

For many years, people have been working on traditional contact sensors for fire detection, such as smoke sensors, temperature sensors and particle sensors. They are less expensive and easy to deploy. However, contact sensor-based systems are only suitable for small spaces and have significant limitations for large scenes. Contact sensors need to be triggered directly by flame temperature or smoke, which may result in the loss of optimal fire extinguishing time. Compared with sensor-based methods, vision-based fire detection has many advantages, such as fast response, wide coverage and environmental robustness. Therefore, vision-based methods have received increasing attention

1.2 FIRE DETECTION IN SURVEILLANCE VIDEO

A novel descriptor based on a bag-of -words approach has been proposed for representing motion.The proposed method has been tested on a very large dataset of fire videos acquired both in real environments and from the web. The obtained results confirm a consistent reduction in the number of false positives, without paying in terms of accuracy of renouncing the possibility to run the system on embedded platforms.

Automated fire detection is an active research topic in computer vision. In this paper, we propose and analyze a new method for identifying fire in videos. Computer vision-based fire detection algorithms are usually applied in closed circuit television surveillance scenarios with controlled background. In contrast, the proposed method can be applied not only to surveillance but also to automatic video classification for retrieval of fire catastrophes in databases of newscast content. In the latter case, there are large variations In fire and background characteristics depending on the video instance.

In frame-to-frame changes of specific low level features describing potential fire regions. These features are color, area size, surface coarseness, boundary roughness and skewness within estimated fire regions. Because of flickering and random characteristics of fire, these features are powerful discriminants. The behavioral change of each one of these features is evaluated, and the results are then combined according to the bayes classifier for robust fire recognition.

The performance of fire detection mainly depends on the quality of manually designed features. However, it is very difficult to design a universal feature for the following reasons: 1) different combustibles may cause different flame colors, 2) different lighting conditions may have different effects on the flame, and 3) the shape of the flame is not fixed due to airflow. For these methods, maintaining a good tradeoff between accuracy and false alarms remains a challenge.

1.3 IMAGE PROCESSING SYSTEM

1.3.1 Digitizer

A digitizer converts an image into numerical representation suitable for input into a digital computer. Some common digitizers are

- Microdensitometer

- Flying spot scanner
- Image dissector
- Videocon Camera
- Photosensitive solid-state arrays.

1.3.2 Image Processor

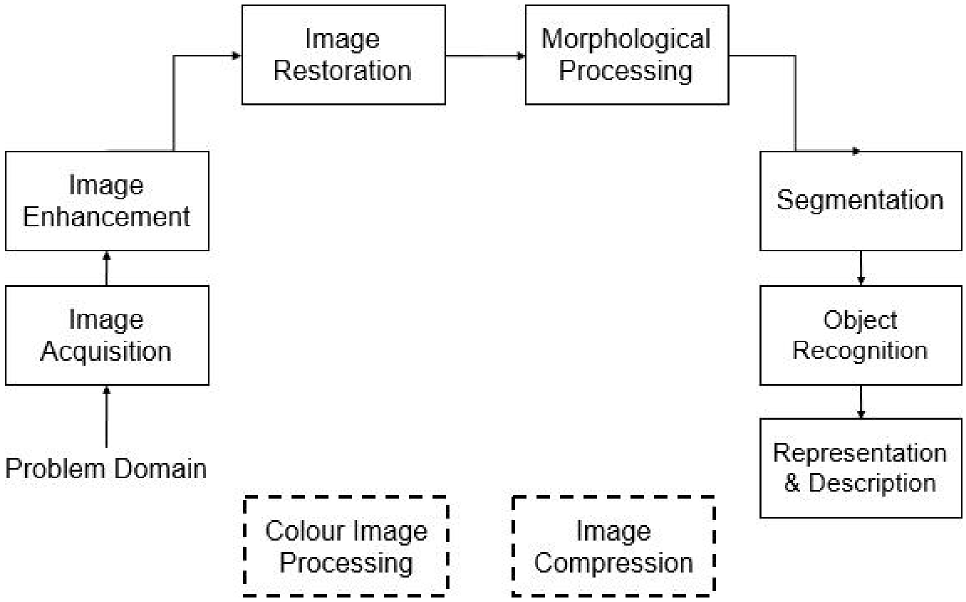


Fig 1.1 Block Diagram Of Fundamental Sequence In An Image Processing System

Image Processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is an image, like video frame or photograph and output, may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two-dimensional signals while applying already set signal processing methods to them.

Digital Computer

Mathematical processing of the digitized image such as convolution,averaging,addition,etc. are done by the computer.

Mass storage

The secondary storage devices used are floppy disks,CD ROMs,etc

Hard copy device

The hard copy device is used to produce a permanent copy of the image and the storage of the software involved.

Operator Console

The operator console consists of equipment and arrangements for verification of intermediate results and for alterations in the software as and when require.

1.4 IMAGE PROCESSING FUNDAMENTAL

Digital Image processing refers processing of the image in digital form.Modern cameras may directly take the image in digital form but genereally images are originated in optical form. They are captured by video cameras and digitalized. The digitalization process includes sampling,quantization.Then these images are processed by the five fundamental process,at least any one of them,not necessarily all of them.

1.4.1 Image Processing Techniques

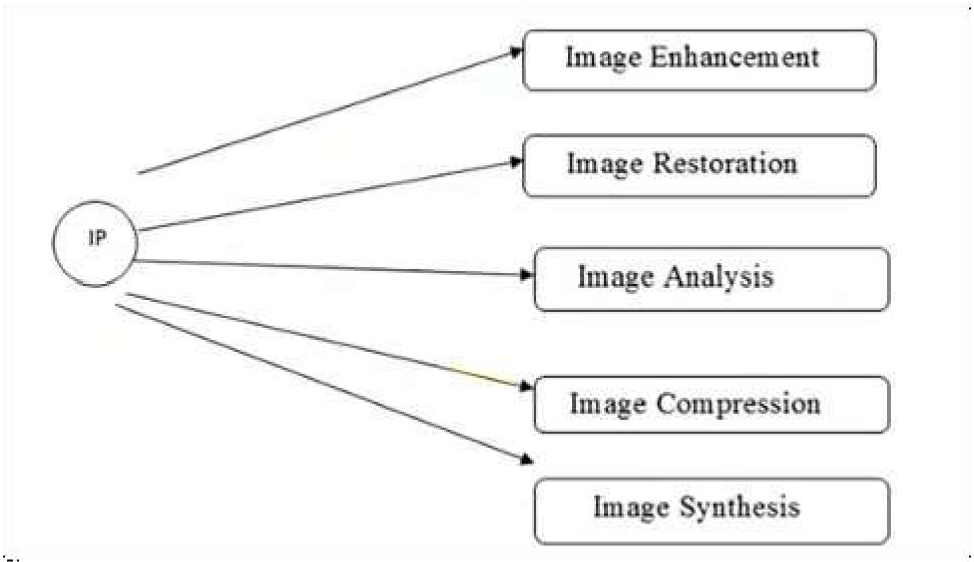


Fig 1.2 Image Processing Techniques

Image Enhancement

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. Such as, changing brightness & contrast etc.

Image Restoration

Image restoration is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.

Image Analysis

Image analysis involves processing an image into fundamental components to extract meaningful information. Image analysis can include tasks such as finding shapes, detecting edges, removing noise, counting objects, and calculating statistics for texture analysis or image quality.

Image Compression

Compression deals with techniques for reducing the storage required to save an image or the bandwidth to transmit it. Particularly in the uses of internet it is very much necessary to compress data. It involves in developing some functions to perform this operation. It mainly deals with image size or resolution.

Image Synthesis

Image synthesis operations create images from other images or non-image data. Image synthesis operations generally create images that are either physically impossible or impractical to acquire.

There are several ways of encoding the information in an image.

1. Binary image

2. Grayscale image
3. Indexed image
4. True color or RGB image

Binary Image

The binary image as it name states, contain only two pixel values.0 and 1.In our previous tutorial of bits per pixel, we have explained this in detail about the representation of pixel values to their respective colors.

Grayscale Image

Grayscale images are monochrome images, Means they have only one color. Grayscale images do not contain any information about color. Each pixel determines available different grey levels. A normal grayscale image contains 8 bits/pixel data, which has 256 different grey levels. In medical images and astronomy, 12 or 16 bits/pixel images are used.

Indexed Image

An indexed image consists of an array, called X in this documentation, and a colormap matrix, called map. The pixel values in the array are direct indices into a colormap

True color or RGM Image

Color images are three band monochrome images in which, each band contains a different color and the actual information is stored in the digital image. The color images contain gray level information in each spectral band. The images are represented as red, green and blue (RGB images). And each color image has 24 bits/pixel means 8 bits for each of the three color band(RGB).

.

2. LITERATURE SURVEY

2.1 LITERATURE SURVEY

1. A.Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S. W. Biak, “Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features”, IEEE Access, vol. PP, pp. 1-1, 2017.

Recurrent neural network (RNN) and long short term memory (LSTM) have achieved great success in processing sequential multimedia data and yielded the state-of –the-art results in speech recognition, digital signal processing, video processing, and text data analysis. In this paper, we propose a novel action recognition method by processing the video data using convolutional neural network (CNN) and deep bidirectional LSTM (DB-LSTM) network. First, deep features are extracted from every sixth frame of the videos, which helps reduce the redundancy and complexity. Next, the sequential information among frame features is learnt using DB-LSTM network, where multiple layers are stacked together in both forward pass and backward pass of DB-LSTM to increase its depth. The proposed method is capable of learning long term sequences and can process lengthy videos by analyzing features for a certain time interval. Experimental results show significant improvements in action recognition using the proposed method on three benchmark data sets including UCF-101, YouTube 11 Actions, and HMDB51 compared with the state-of-the-art action recognition methods.

2. K. Muhammad, M. Sajjad, M. Y. Lee and S. W. Biak, “Efficient visual attention driven framework for key frames extraction from hysteroscopy videos”, Biomedical signal Processing and Control, vol. 33, pp. 161-168, 2017.

Recent years have shown enthusiastic research interests in diagnostic hysteroscopy (DH), where various regions of the female reproductive system are visualized for diagnosing uterine disorders. Currently, the hysteroscopy videos produced during various sessions of patients are stored in medical libraries, which are usually browsed by medical specialists Gynecologists to visualize previous videos of a patient or to study similar cases. However, the abundant redundancy of frames in DH videos make this searching relatively more difficult for gynecologists, wasting their time browsing such large libraries. In this context, video summarization can be used to reduce this redundancy by extracting key frames, thus making the process of browsing and indexing DH videos more efficient. In this letter, we propose an efficient domain-specific visual attention-driven framework for summarizing DH videos. For key frames extraction, multi-scale contrast, texture, curvature, and motion based saliency features are computed for each frame using integral image, which are then fused by a linear weighted fusion scheme to acquire a final saliency map. Experimental results in comparison with other related state-of-the-art schemes confirm the effectiveness and efficiency of the proposed framework.

3. J. Choi and J. Y. Choi, “Patch based fire detection with online outlier learning,” in Advanced Video and signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on, 2015, pp. 1-6.

Fire detection is one of the most interesting issues for surveillance. The existing approaches for the fire detection suffer from a high false positive ratio. To solve the problems, we present a patch-based fire detection algorithm with online outlier learning. In the proposed algorithm, the candidates of fire are obtained in the form of patch, while the classical candidates have been based on pixels or blobs. Because the patches of fire have more distinctive shape than the entire fire, the shape classifier can recognize the candidates correctly from fire-like outliers. In addition, we propose an online outlier learning scheme which handles the irregularity of fire based on the repeatability of shape in time. The proposed algorithm is experimented with new challenging dataset, consisting of 50 positive videos with fire and 44 negative ones with fire-like outliers. By evaluating on the dataset, we validate the performance of our algorithm qualitatively and quantitatively.

4. Bhawna goyal, Sunil grawal. “From Multi-Scale Decomposition to Non-Multi-Scale Decomposition Methods: A Comprehensive Survey of Image Fusion

Techniques and Its Applications” 2011 3rd International Conference on, 2011, pp. 262-265.

Image fusion is a well-recognized and a conventional field of image processing. Image fusion provides an efficient way of enhancing and combining pixel-level data resulting in highly informative data for human perception as compared with individual input source data. In this paper, we have demonstrated a comprehensive survey of multi-scale and non-multi-scale decomposition-based image fusion methods in detail. The reference-based and non-reference-based image quality evaluation metrics are summarized together with recent trends in image fusion. Several image fusion applications in various fields have also been reported. It has been stated that though a lot of singular fusion techniques seemed to have given optimum results, the focus of researchers is shifting toward amalgamated or hybrid fusion techniques, which could harness the attributes of both multi-scale and non-multi-scale decomposition methods. Toward the end, the review is concluded with various open challenges for researchers. Thus, the descriptive study in this paper would form basis for stimulating and nurturing advanced research ideas in image fusion

5. Pasquale Foggia, Alessia Saggese and Mario Vento, “Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion,” IEEE TRANSACTIONS on circuits and systems for video technology, vol 25, pp. 1545-1556, 2015.

In this paper, we propose a method that is able to detect fires by analyzing videos acquired by surveillance cameras. Two main novelties have been introduced. First, complementary information, based on color, shape variation, and motion analysis, is combined by a multiexpert system. The main advantage deriving from this approach lies in the fact that the overall performance of the system significantly increases with a relatively small effort made by the designer. Second, a novel descriptor based on a bag-of-words approach has been proposed for representing motion. The proposed method has been tested on a very large dataset of fire videos acquired both in real environments and from the web. The obtained results confirm a consistent reduction in the number of false positives, without paying in terms of accuracy or renouncing the possibility to run the system on embedded platforms

6. Guanglin Li, Mario Vento “Electromyography pattern Recognition Based control of Powered multifunctional Upper-Limp Prostheses” vol 25. pp.1359-1371, 2014.

The human history has been accompanied by accidental trauma, war, and congenital anomalies. Consequently, amputation and deformity have been dealt with, one way or another, throughout the ages. More than one million individuals in the United States today are living with limb amputations in which there are approximately 100,000 patients with an upper limb amputation. The wars in Iraq and Afghanistan have added to this number. According to the survey results of the Second China National Sample Survey on Disables (SCNSSD 2006) led by the National Statistics Bureau in 2006, approximately 8% of physical disables, or 2.26 million people, live with limb amputations in China alone. Natural disasters and accidents have been making this number increase. The massive earthquakes that occurred in May 2008, Sichuan Province, China, recently increased about 20 thousand of new limb amputees. Expectations for control of upper limb prostheses have always been high because of the standard established by able-bodied dexterity. Most commercially available upper limb prostheses are either body-powered or electrical motor powered. The body-powered prostheses are operated by certain movements of the amputees’ body through a system of cables, harnesses, and sometimes, manual control. In order to operate a body-powered prosthesis, the upper limb amputees have to possess significant strength and control over various body parts, including the shoulders, chest, and residual limb which must have sufficient residual limb length, musculature, and range of motion.

3. AIM AND SCOPE

3.1 AIM

This project A comparison of the proposed and current algorithms reveals that the accuracy of fire detection algorithms based on object classification CNNs is higher than other algorithms.

3.2 SCOPE

Majority of the research since the last decade is focused on traditional features extraction methods for flame detection. The major issues with such methods is their time consuming process of features engineering and their low performance for flame detection. Such methods also generate high number of false alarms especially in surveillance with shadows, varying lightings, and fire-colored objects. To cope with such issues, we extensively studied and explored deep learning architectures for early flame detection. Motivated by the recent improvements in embedded processing capabilities and potential of deep features, we investigated numerous CNNs to improve the flame detection accuracy and minimize the false warnings rate.

3.3 EXISTING SYSTEM

An IOT based fire alarming and authentication system for workhouse using raspberry pi has implemented. Here the system is installed with fire sensor, gas sensor and PIR sensor for the purpose of detection of fire. When the sensor detects the flame the water sprinkler motor gets on automatically. For the purpose of people alert buzzer is used. The measured condition about the sensor monitored through IOT.

3.3.1 Existing Block Diagram

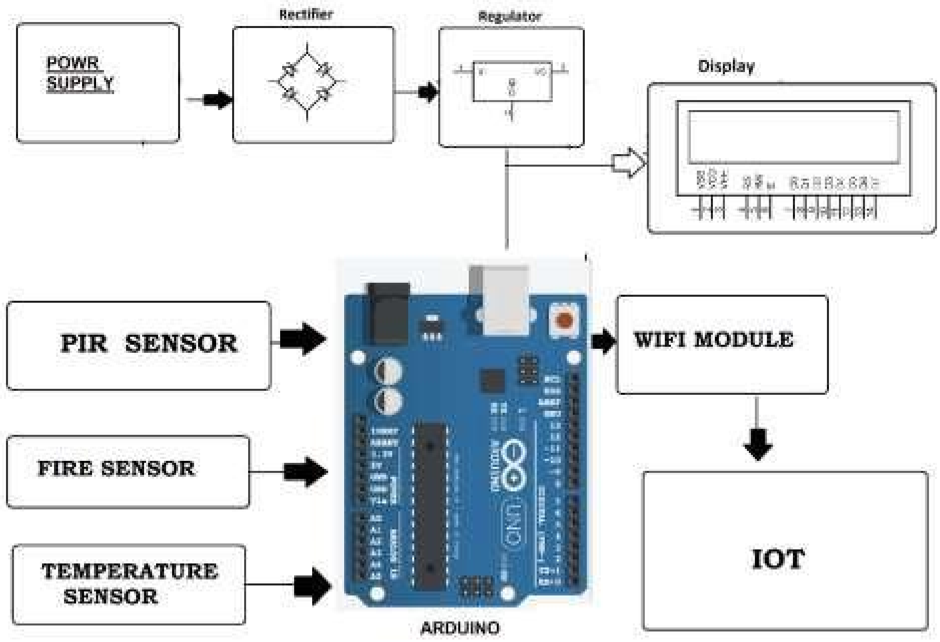


Fig 3.1 Existing Block Diagram

3.3.2 Disadvantage

- Sensor based technology has implemented so it supports for particular distance
- Less Efficiency.
- Time consuming.

4. PROPOSED FIRE DETECTION SYSTEM

4.1 PROPOSED SYSTEM

Deep learning (DL) architectures for this problem and propose a cost effective CNN framework for flame detection is surveillance videos. Our framework avoids the tedious and time consuming process of feature engineering and automatically learns rich features from raw fire data. Inspired from transfer learning strategies trained and fine-tuned a model with architecture similar to Google Net for fire detection, which successfully dominated traditional fire detection schemes. The proposed framework balances the fire detection accuracy and compared to state-of-the-art fire detection schemes based on CNN fire detection information is serially transmitted to controller using UART.when,controller receives fire detected information, sends the alert message using GSM, Buzzer is used for fire indication purpose.

4.1.1 Block Diagram

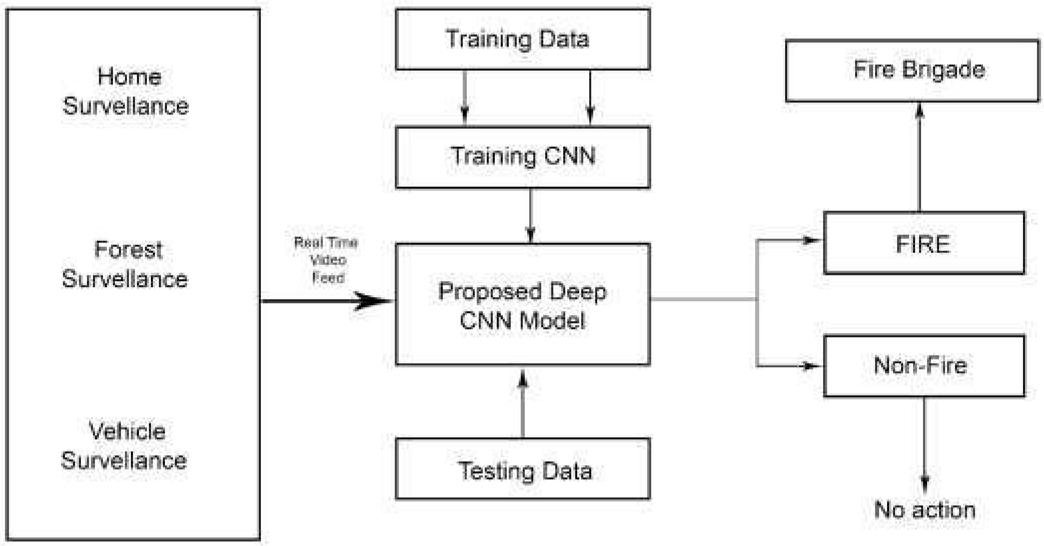


Fig 4.1 Proposed System Block Diagram

In this system the surveillance video converted as N number of frames which is given as input for detecting the fire. Then it involves the image processing technique like preprocessing ,segmentation, feature extraction and classification. Based on Convolutional neural network algorithm it detects fire.

4.2 DATA COLLECTION

The first and foremost step in deep learning is data collection. We have used two datasets to apply the proposed model. One of which is the fire images dataset from GitHub. The second dataset is the normal dataset which consists some non fire areas.

4.2.1 Segregation Of Images

In the dataset, the images which has fire and other images are there and we separated those dataset in separate file has thousands of images which has fire in it

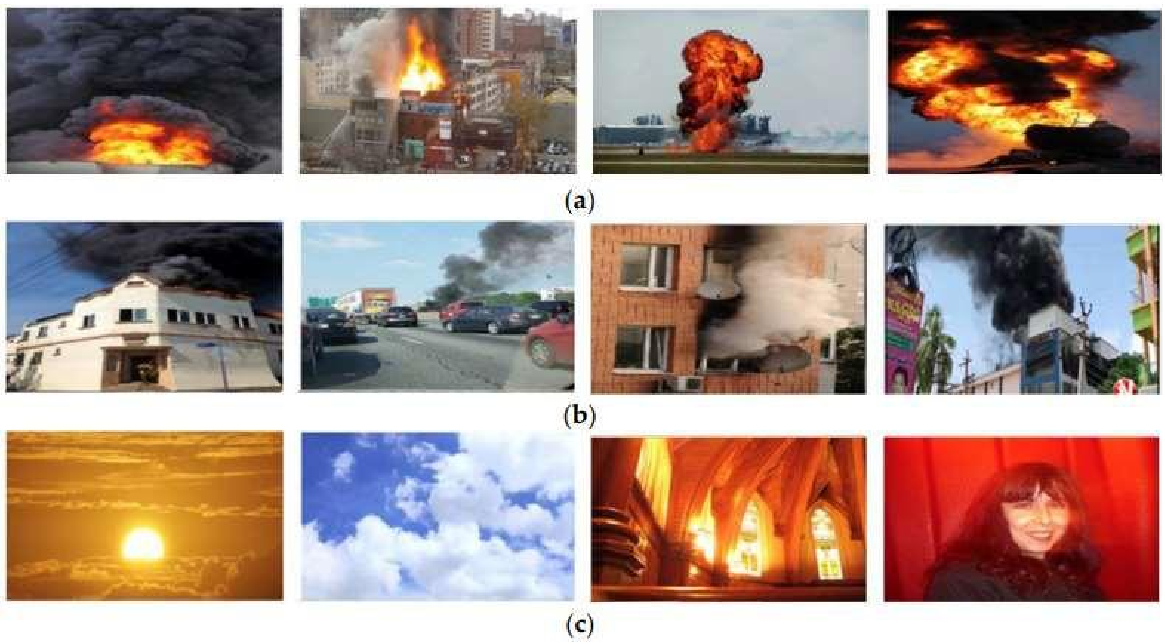


Fig 4.2 Dataset Images

4.3 DATA PREPROCESSING

After splitting datasets into fire and non-fire categories, we convert them into using convert in different color command in opencv. Then the converted images are resized into a common size of 100*100. We resize the images to 100*100 to speed up the process. Larger size of images makes the process to take high amount of time.

After modifying them, we store the modified data into a NumPy array as npy. We have two NumPy arrays for the images and labels. We set the label as 0 for no fire and 1 for fire. The images are stored in data. npy array and labels are stored in target array. We normalize all the images by dividing the entire data array by 255 to get a range from 0 to 1. Then we reshape the data array into 4 dimensional in order to train the neural network. Then the target array is converted into categorical form.

4.3.1 Categorical Form

Categorical Data is the data that generally takes a limited number of possible values. Also, the data in the category need not be numerical, it can be textual in nature. All machine learning models are some kind of mathematical model that needs numbers to work with. This is one of the primary reasons we need to pre-process the categorical data before we can feed it to machine learning models.

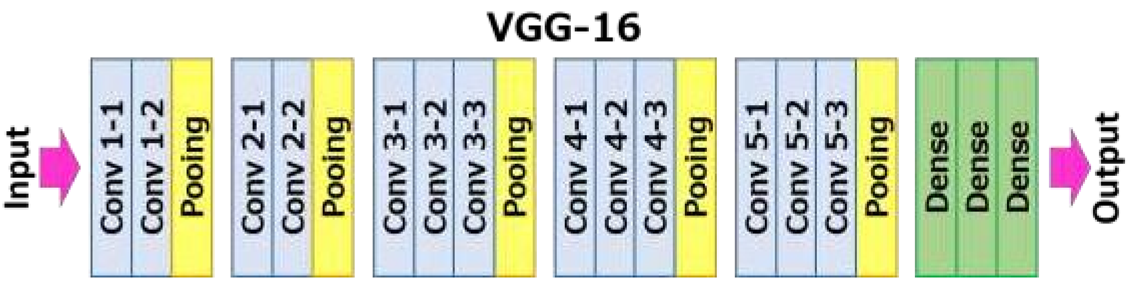
4. 4 TRAINING THE CNN MODEL

The training phase is where the deep learning algorithms are used. We have tried 3 different algorithms for training the model. The best algorithm with more accuracy rate and less validation loss in chosen for predicting the result. The three algorithms are following:

- VGG 16 model
- Mobilenet model
- Parallel CNN

4.4.1 VGG 16 Network model

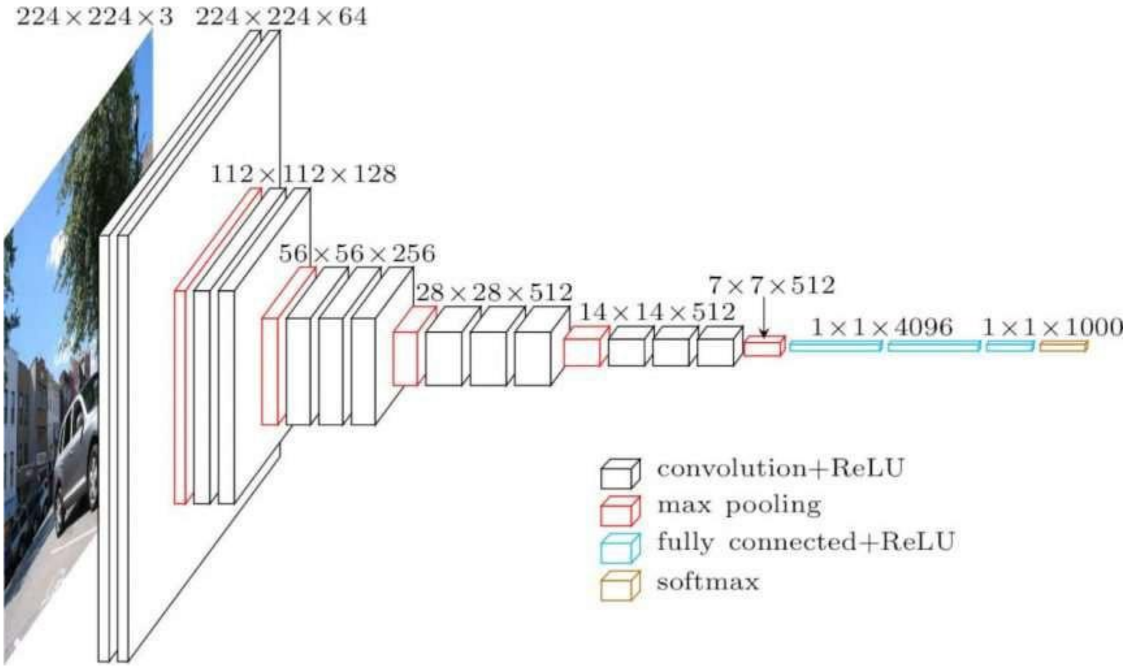
VGG16 is a Convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second Convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU’s.



4.3 VGG-16

4.4.1.1 ARCHITECTURE OF VGG 16

The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of Convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.



4.4 Architecture Of Vgg 16

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

All hidden layers are equipped with the rectification (ReLU) non-linearity. It is also noted that none of the networks (except for one) contain Local Response Normalization (LRN), such normalization does not improve the performance on the ILSVRC dataset, but leads to increased memory consumption and computation time.

4.4.1.2 ***FEATURES OF VGG 16***

- It is also called the Oxford Net model, named after the Visual Geometry Group from Oxford.
- Number 16 refers that it has a total of 16 layers that has some weights.
- It only has Conv and pooling layers in it.

- It always uses a 3 x 3 Kernel for convolution.
- 2x2 size of the max pool.
- It has a total of about 138 million parameters.
- Trained on ImageNet data
- It has an accuracy of 92.7%.
- It has one more version of it Vgg 19, a total of 19 layers with weights.

4.4.1.3 ***DISADVANTAGES OF VGG 16***

Unfortunately, there are two major drawbacks with VGGNet:

- ❖ It is painfully slow to train.
- ❖ The network architecture weights themselves are quite large (concerning disk/bandwidth).

4.4.2 MobileNet

As the name applied, the MobileNet model is designed to be used in mobile applications, and it is TensorFlow's first mobile computer vision model.

MobileNet uses depth wise separable convolutions. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets. This results in lightweight deep neural networks.

A depth wise separable convolution is made from two operations.

- ❖ Depth wise convolution.
- ❖ Pointwise convolution.

4.4.2.1 MobileNet Architecture

First layer is a full convolutional layer. All layers are followed by batch normalization and ReLU non-linearity. However, final layer is a fully connected layer without any non-linearity and feeds to the softmax for classification. For down sampling, strided convolution is used for both depth wise convolution as well as for first fully convolutional layer. The total number of layers for mobilenet is 28 considering depth wise and point wise convolution as separate layers.

4.5 Figure Layers In Mobilenet

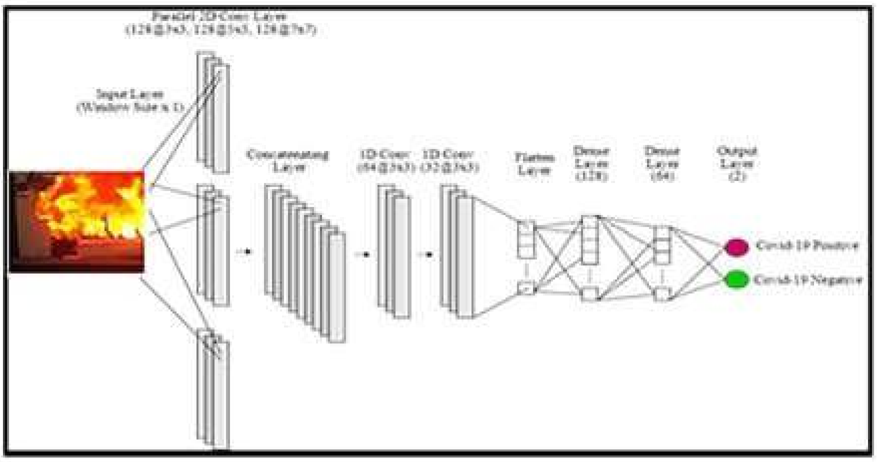
Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

4.6 Figure Architecture Of Mobilenet

4.4.3 Parallel CNN

For the proposed model, i.e., the CNN model, after data pre-processing, we train the dataset with Convolutional Neural Network. We use the NumPy array data from data pre-processing to train the model. We train a sequential model with the parallel convolutional layer. We apply a parallel 2D convolutional layer to the input layer with 128 filters to the kernels of sizes 3, 5, and 7. We concatenate all the conv2D layers into a single layer with padding as same to get the convolution output size same as the input then reduce spatial dimensions using max pooling. Then we apply a convolutional layer with 64 and 32 filters to kernel size 3x3. For the activation parameter, we have used Relu to get either 0 or 1, depending on whether the input is negative or not. Then reduce spatial dimensions using max pooling. We flatten the output of the convolutional layers into a single-dimensional array. This is connected to the 128 and 64 neurons dense layer, also called a fully-connected layer. To prevent the overfitting of the training data, we've added dropout layers with a dropout rate of 0.5. We've grouped the dataset into 2 batches by giving batch_size=2 since the amount of data is large and requires more storage space. We've given a total of 20 epochs for the training process.

The output layer will be of two neurons for the positive and negative results of COVID-19. So once the model is fully trained, the best model with the least loss and high accuracy is saved by setting save_best_only=True. The best model will be saved with an extension of '.h5'. This saved model will be loaded into the web app to predict Covid-19.



4.7 System Architecture

Model: "sequential"

Layer (type)	Output Shape	Param #
model (Functional)	(None, 100, 100, 384)	11008
conv2d_3 (Conv2D)	(None, 98, 98, 64)	221248
activation (Activation)	(None, 98, 98, 64)	0
max_pooling2d (MaxPooling2D)	(None, 49, 49, 64)	0
conv2d_4 (Conv2D)	(None, 47, 47, 32)	18464
activation_1 (Activation)	(None, 47, 47, 32)	0
max_pooling2d_1 (MaxPooling2	(None, 23, 23, 32)	0
flatten (Flatten)	(None, 16928)	0
dropout (Dropout)	(None, 16928)	0
dense (Dense)	(None, 128)	2166912
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dropout_2 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 2)	130

Total params: 2,426,018
Trainable params: 2,426,018
Non-trainable params: 0

4.8 Model Summary

4.4.4 Faster-RCNN

There are two stages in Faster-RCNN setting .In the first stage, feature maps of the original image are generated by feature extraction networks (VGG, ResNet, Inception, Inception Resnet -v2, etc.). And the feature map from some selected intermediate convolutional layer is used to predict proposal regions with objectness scores and locations by Region Proposal Network (RPN). This stage just output scores that estimate the probability of object or not object and box regression for each proposal by a two-class softmax layer and the robust loss function (Smooth L1). In the second stage, the locations of the proposal regions are used to crop features from the same intermediate feature map by ROI pooling. And the regional feature map for each proposal region is fed to the remainder of the network to predict class-specific scores and refine box locations. This network achieves sharing part of the computation by cropping proposals from the feature map generated by the same intermediate convolutional layer in the first stage. This method avoids inputting each proposal region into the front-end CNN to compute the regional feature map. However, per proposal region must be input into the remainder of the network to compute separately. Thus, the detection speed depends on the number of proposal regions from the RPN.

5. RESULTS AND DISCUSSION

5.1 RESULT

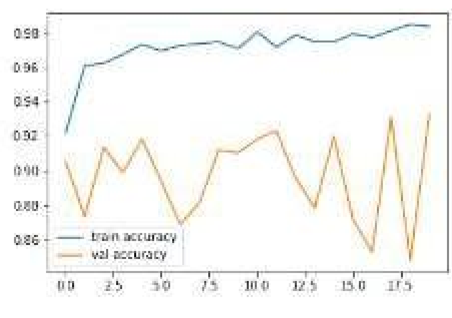
The important goal of this proposed model is to build an effective deep learning model which detects fire with higher accuracy. Here, in this proposed model, we've collected a lot of sample dataset from various sources to train the CNN model. The websites GitHub and Kaggle.com were used for retrieving images.

Fire

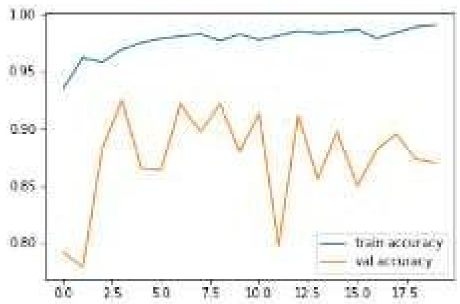
No fire

5.2 ACCURACY RATE

We have used three different CNN models. We compared two pre-trained models with our proposed CNN model with a parallelization strategy. The two pre-trained models are VGG-16 and MobileNet. VGG-16 consists of 16 layers with millions of parameters. We imported these two pre-trained models and used them for training purposes separately. In the below images, the first two images are the accuracy with epoch graph of the pre-trained VGG-16 and MobileNet models. The image at the bottom is the proposed parallel CNN model. This clearly shows that the parallel CNN model has shown a great result with a higher accuracy rate of 98.3%, whereas the best validation accuracy of VGG-16 and MobileNet is 93% and 92% respectively.



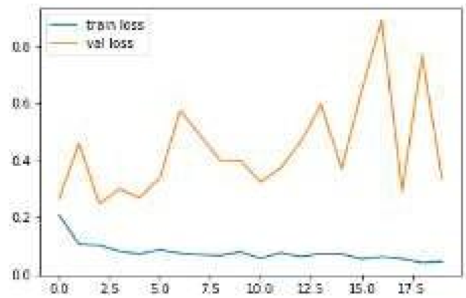
Accuracy of CNN with Vgg 16



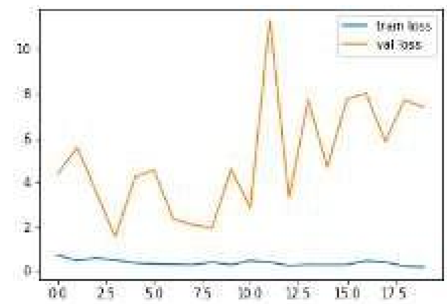
Accuracy of CNN with mobilenet

4.3 VALIDATION LOSS

In the below images, the first two images are the loss to the epoch graph of the two pre-trained VGG-16 and MobileNet models. The image at the bottom is the loss graph of our proposed parallel CNN model. This shows that the loss rate of our proposed model is less than the loss rate of the other two pre-trained models. The loss rate of the best epoch in parallel CNN is 0.04 whereas the loss rate of VGG-16 and MobileNet is 0.25 and 1.53 respectively.



VGG 16



Mobilenet

5.4 FASTER-RCNN

Algorithm	AP(%)		mAP(%)	Detection speed(FPS)	Accuracy
	smoke	Fire			
Faster - RCNN	79.7	84.9	82.3	3	99.43
R-FCN	78.5	83.3	80.9	5	99.20
SSD	72.8	82.8	77.8	16	92.83

We use the best model which is saved to predict the fire (i.e) Faster-RCNN algorithm gives more accuracy than other conventional algorithms and selection of right algorithm is used to detect fire with minimum time and high efficiency of detecting fire.

6.CONCLUSION

6.1 CONCLUSION

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. Fire accidents can be detected using the cameras. So that, here we proposed a CNN approach for fire detection using cameras. Our approach can identify the fire under the camera surveillance. Furthermore, our proposed system balances the accuracy of fire detection and the size of the model using fine-tuning of datasets. We have obtained an accuracy of 99% using Faster-RCNN algorithm. Also the F-measure value is 0.95. These values shows that the model gives a better prediction. We conduct experiments using datasets collected from recording of fire and verified it to our proposed system. In view of the CNN model’s reasonable accuracy for fire detection, its size, and the rate of false alarms, the system can be helpful to disaster management teams in controlling fire disasters in a short time. Thus, avoiding huge losses. This work mainly focuses on the detection of fire scenes under observation. Future studies may focus on deploying the model into raspberry pi and using necessary support packages to detect the real time fire by making challenging and specific scene understanding datasets for fire detection methods and detailed experiments.

5.2 OUTPUT SCREENSHOTS

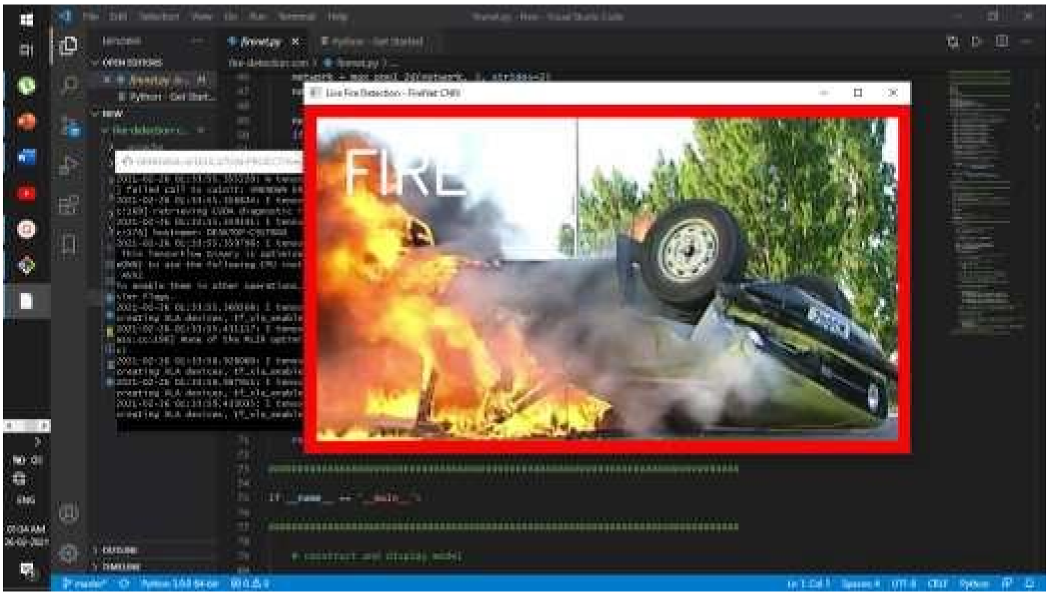


Fig 5.1 Fire

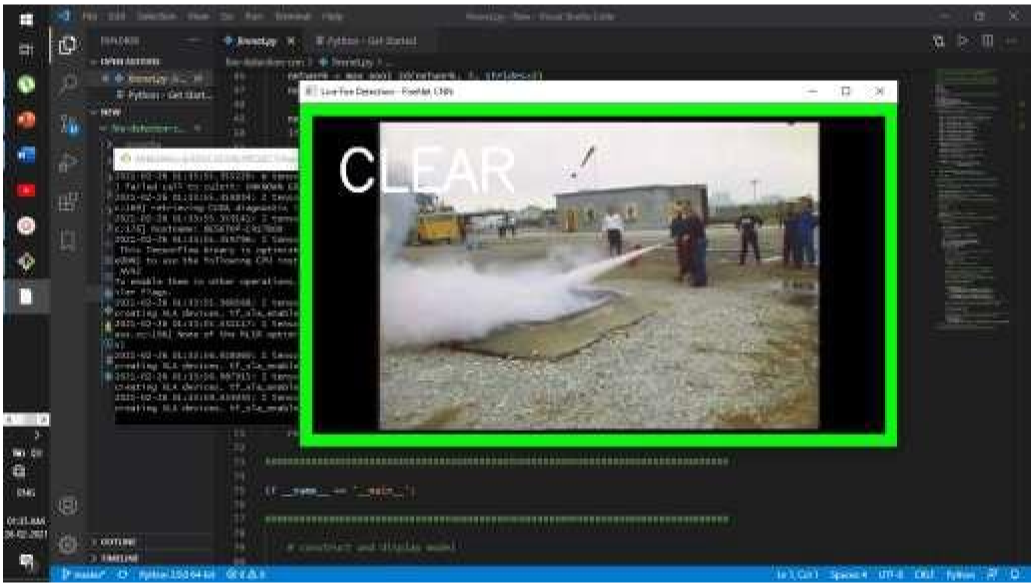


Fig 5.2 No Fire

REFERENCE:

1. Anwar S, Hwang K, Sung W (2015) Fixed point optimization of deep convolutional neural networks for object recognition, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 1131-1135.

2. Bhattacharjee S, Roy P, Ghosh S, Misra S, Obaidat MS (2012) Wireless sensor network-based fire detection, alarming, monitoring and prevention system for Bord-and-Pillar coal mines. J Syst Softw 85(3):571–581

3. Celik T, Demirel H (2009) Fire detection in video sequences using a generic color model. Fire Saf J 44(2):147–158

4. Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y (2015) PCANet: a simple deep learning baseline for image classification? IEEE Trans Image Process 24:5017–5032

5. Chen TH, Wu PH, Chiou YC (2004) An early fire-detection method based on image processing. InImage Processing, 2004. ICIP'04. 2004 International Conference on 2004 Oct 24 (Vol. 3, pp. 1707-1710). IEEE “Chest X-rays can give COVID-19 test results in just 10 minutes: Expert”, The New Indian Express.

6. Chen T-H, Wu P-H, Chiou Y-C (2004) An early fire-detection method based on im- age processing, in: Proceedings of International Conference on Image Process- ing, ICIP’04, 2004, pp. 1707–1710

7. Chino DY, Avalhais LP, Rodrigues JF, Traina AJ (2015) BoWFire: detection of fire in still images by integrating pixel colour and texture analysis, Proceedings of the 28th 2015 Conference on Graphics, Patterns and Images, SIBGRAPI pp. 95-102

8. Foggia P, Saggese A, Vento M (2015) Real-time fire detection for video-surveillance applications using a combination of experts based on colour, shape, and motion. IEEE Trans Circuits Syst Video Technol 25:1545–1556

9. Freund Y (1995) Boosting a weak learning algorithm by majority. Inf Comput 121(2):256–285

10. Habiboğlu YH, Günay O, Çetin AE (2012) Covariance matrix-based fire and flame detection method in video. Mach Vis Appl 23(6):1103–1113

11. Jiang B, Yang J, Lv Z, Tian K, Meng Q, Yan Y (2017) Internet cross-media retrieval based on deep learning. J Vis Commun Image Represent 48:356–366

12. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Proces Syst 25:1097–1105.

13. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: Deep filter pairing neural network for person reidentification, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 152-159.

14. Liu CB, Ahuja N (2004) Vision based fire detection. In Pattern Recognition. ICPR 2004. Proceedings of the 17th International Conference on 2004 Aug 23 (Vol. 4, pp. 134-137). IEEE