ISSN No:-2456-2165

Abnormal Vehicle Behavior Detection using Deep Learning and Computer Vision

[1]Susmitha John
Computer Science and Engineering (of KTU),
St.Joseph's College Of Engineering and Technology (of KTU), Kottayam,India

[2]Bino Thomas
Computer Science and Engineering (of KTU),
St.Joseph's College Of Engineering and Technology (of KTU), Kottayam, India

Abstract:- In the modern era, usage of video surveillance has increased which in fact increase the size of data. Video surveillance is widely using in both public and private areas for improving the security and safety of human being. Hence, it is important to identify and analyse the video in different angle so as to extract the most important information from the video. The video may contain both usual or unusual event, mostly the users need to find out the unusual event from the video that may affect their security. To differentiate both the events separately, here we are considering a special scenario related with vehicle. The vehicles on road can move in different ways, where they can follow or violate traffic rules, illegal U-turns, accidents etc. In this paper, the unusual event considered is the accidents on the road. The technology used is deep learning and computer vision. The neural network selected is the DenseNet. The DenseNet is a convolutional neural network. The peculiarity of a DenseNet architecture is that each layer in a network is connected to every other layer. For each layer, the feature maps of all the preceding layers are used as inputs, and its own feature maps are used as input for each subsequent layer. The deployment of DenseNet along with computer vision increases the accuracy of the system.

Keywords: *Deep Learning, Computer Vision, Segmentation, Tracking.*

I. INTRODUCTION

The increase in the population rate also increases the need of safety and security of human beings in public and private areas. The usage of video surveillance has become a vast concern of everyday life. As a consequence of these the deployment of cameras has done almost everywhere. Video surveillance are widely used in smart cities, smart offices, etc. Such videos are analyzed and studied through different technologies for extracting important information. And, it is currently a well-researched area and has mainly applications. The most attractive areas include activityrecognition from the video surveillance system. The main focus is on understanding the activities involved for the detection and classification of the targets of interest and analyzing the activities included in the data. The detection and reporting of situations of special interests from a video is vital step, where unexpected things may happen. In such cases, the video surveillance system which can easily interpret the scenes and recognize the abnormal behaviors automatically can an important role in data analytics. The system would then notify operators or users accordingly. This technology includes detection, tracking and counting all the movable objects from video and analyzing their behaviors, and reply to them accordingly. Most challenging part is detection of abnormal events from a video and informing it to responsible authority. The abnormal behavior is difficult to explain, but can be easily notified when it happens. Abnormal behavior is a psychological term for defining actions that are different from what is considered as normal in a particular society or culture or in any other environment. This abnormal behavior definition is functional and useful for many purposes. However, most definitions of abnormal behavior also take into account that from a psychological point of view, mental illness, pain, and stress often play a major role in behavioral patterns. Abnormal events include the situations which are unnecessary or unpredicted events like road accidents, traffic violations, etc.

The monitoring of video from surveillance system can be analyzed and detected for object from video which have several applications. The enhancement in video surveillance system also allows several other editing and storing of videos in more efficient way. The processing and analysis of such video is of great importance. It contains many valuable information that can be used for finding out different activities from the video. The current video surveillance can use many interesting technologies like computer vision and deep learning.

➢ *Objective and Scope:*

The capturing of video and processing such video for further analysis to extract important feature is a challenging task. According to the area of interest we need to process the data because there is no need of the whole data. We have to simplify and change the representation of an image into something that is more meaningful and easier to analyze. An abnormal behavior detection framework based on deep learning algorithm is used. The objectives of this proposed system are:

- Developing a system for detecting the abnormal vehicle behavior.
- Detection is done using the specialized framework where both neural network and computer vision technologies are used.

- Trying to acquire high range of accuracy with less complex structure.

The normal movement of vehicle is considered as normal event where as any accident situation is considered as abnormal event. Abnormal events are always something that is unexpected from the normal situation. The use of neural network will help us to differentiate between these two cases.

The scope of the system includes the deployment in different areas like smart cities etc. Also, the usage of computer vision helps the computers to gain high level of understanding about the video and images.

II. LITERATURE SURVEY

➢ *Unsupervised Anomaly Detection*

The significance of anomaly identification in traffic footage for intelligent transportation systems has recently attracted more attention. They introduced a quick unsupervised anomaly detection system in their paper [2], which consists of three modules: pre-processing, candidate selection, and backtracking anomaly detection. Outputs from the pre-processing module include stationary objects found in videos. The candidate selection module then employs K-means clustering to discover probable anomalous regions after removing the incorrectly classified stationary items using a nearest neighbour method. The backtracking anomaly detection algorithm then determines the onset time of the anomaly and computes a similarity statistic.

➢ *Temporal Segmentation*

They introduce a temporal segmentation and a keyframe selection techniques for user-generated video in this paper [3]. A user generated video temporal segmentation technique has been suggested that creates a partition-based video on a categorization of camera motion because user generated video is rarely arranged in shots and user interests are typically exposed through camera movements. It has been proposed that a Hierarchical Hidden Markov Model (HHMM) which generates a user-meaningful user generated video temporal segmentation be fed motion-related mid-level information. A keyframe selection approach that chooses a key frame for camera motion patterns with fixed content, like zoom, still, or shake, and a group of keyframes for the translation of patterns with dynamic content has also been suggested.

➢ *YOLO*

The implementation of intelligent real-time systems that can identify unusual vehicle actions may notify law enforcement and transportation organisations of potential offenders and help prevent traffic accidents. By creating an application for the identification of anomalous driving behaviour utilising traffic video content, they address this issue in this study [4]. Real videos from traffic cameras are used for evaluation in order to find halted cars and other potential anomalies in driving behaviour. The following steps make up the suggested algorithm for detecting aberrant vehicle behaviour:

- Step 1: Use real-time traffic video sources to identify vehicles.
- Step 2: Using identified cars to extract features of cars.
- Step 3: Utilizing the vehicles tracked, the traffic anomalies are detected.

➢ *Deep Spatio-Temporal Representation*

A novel approach for the automatic detection of traffic accidents in video surveillance was put forth by the author [7]. Instead of using conventional hand-crafted features, the proposed approach method automatically learns the feature representation with the help of spatiotemporal patterns of basic pixel intensity values. They define the vehicle crash as an exceptional occurrence or event. The suggested system uses denoising the autoencoders that have been trained on videos of typical traffic to extract deep representation. The probability of the deep description and the reconstruction error are used to calculate the chances of an accident. A one class support vector machine is used to train an unsupervised model for the probability of the deep representation. Additionally, the intersecting locations of the vehicle's trajectories are employed to lower the rate of false alarms and boost the overall system reliability.

➢ *Adaptive Video-Based Algorithm*

On highways and expressways, a unique vision-based method for detecting the traffic accidents is presented [8]. By utilizing Farneback Optical Flow for motion detection and a statistic heuristic approach for accident identification, this approach is based on an adaptive traffic motion flow modelling technique. On a collection of videos of traffic and accidents on highways, the algorithm was used. The outcomes demonstrate the effectiveness and applicability of the suggested approach when just 240 frames are used to describe traffic movements. This approach avoids using a sizable database in the absence of suitable and widespread accidents videos benchmarks.

➢ *SVM*

In this paper [9], they have used the important statistics for regulators and policy makers which is proposed in an automated fashion. These statistics contain lane usage monitoring, vehicle counting, vehicle speed estimation from video and classification of vehicle type. The vital part of such a proposed system is to detect and classify the vehicles in traffic videos. For this purpose, they implement two models- first is a Mixture of Gaussian with SVM system and the second one is based on Faster RCNN, which is a recently developed popular deep learning architecture for the detection of objects in images. In the experiments, the Faster RCNN performs better than Mixture of Gaussian in the detection of vehicles that may be static, overlapping or in other situations like night-time conditions. Faster RCNN also have better performance than the SVM in the classification task of vehicle types based on appearances.

➢ *Cooperative Vehicle Infrastructure System*

They proposed a methodology for the detection of accidents caused by an automatic car which is based on the Cooperative Vehicle Infrastructure Systems (CVIS) and machine vision [10]. Firstly, the CAD-CVIS dataset is

established which is a novel image, with an intention to increase the accuracy of detecting accidents based on smart roadside devices in CVIS. Particularly, the CAD-CVIS consisted of variety of accident types, weather conditions and accident location, which indeed helps in improving the self-adaptability of the accident detection methods among different traffic situations. Secondly, a deep neural network model YOLO-CA based on CAD-CVIS and deep learning algorithms to detect accident are also developed. For enhancing the performance of the detection of small objects, they use the Multi-Scale Feature Fusion (MSFF) and loss function with dynamic weights in this model.

- *Trajectory Tracking Based Method*

Here, the abnormal vehicle behavior detection is done by tracking the trajectories effectively, the complete procedures are divided mainly into three steps: the target detection and vehicle tracking, analysis of vehicle trajectories, and vehicle behavior analysis [11]. Firstly, a three-frame differencing method is used to achieve the initial target location and proposedan improved tracking algorithm which is based on the Kalman predictor; then, an adaptive segmented linear fitting algorithm is proposed to achieve vehicle trajectory fitting. To establish the vehicle abnormal behavior detection model, two parameters containing the velocity variation rate and direction variation rate are used.

- *Deep Learning Based Methods*

For detecting the salient regions in videos, a deep learning-based method is proposed [15]. It mainly addresses two important issues- First, the deep video saliency model is trained with the absence of adequately huge and pixel-wise video data which is annotated one; and second, training and detection with fast video saliency. The proposed system mainly consists of two modules, one for capturing the spatial and other for temporal saliency information. The dynamic saliency model, explicitly incorporating saliency estimates from the static saliency model, directly produces spatiotemporal saliency inference without time-consuming optical flow computation. For simulating the training video data from existing annotated image datasets, a novel data augmentation approach is enabled in this network preventing the overfitting with the limited number of training videos and to learn the diverse saliency information. The deep video saliency model efficiently learn both the spatial and temporal saliency motions, thus producing an accurate spatiotemporal saliency estimate motivating the synthetic video data and real videos.

The abnormal vehicle behavior analysis is a challenging field in surveillance videos, it is mainly due to the huge variations in different anomaly cases and the high complexities in video surveillance. In this study [12], they proposed a novel smart vehicle behavior analysis framework which is based on a digital twin. The first step is to implement the detection of vehicles based on the deep learning, and then for tracking vehicles both Kalman filtering and feature matching are used. After that, the mapping of tracked vehicle is done to a digital twin virtual scene. According to the modified detection conditions which is set up in the scene, the behavior of vehicle is tested. This process is repeated and the stored behavior data can be used further for the reconstruction of the scene again for a secondary analysis.

In this work [1], they make the assumption that visual elements occurring in a temporal sequence reflect the occurrences of traffic accidents. They had divided the video segmentation two categories- spatial and temporal. To locate the objects spatially in different frames, the spatial segmentation classifies the interested objects from the video. The segmentation helps in identifying and tracking the objects from the video. The model architecture proposed by author extract the visual features followed by the identification of temporal patterns. The public dataset is used for in the training phase where the visual and temporal features are learned using the convolution and recurrent network.

In computer vision, the deep architectures with convolutional structures have been found vastly efficient and frequently used. For deep learning algorithms, Graphics Processing Units (GPUs) are found to be more effective because of its high processing power. Also, the availability of large amount of data has also made it possible to train the deep neural networks efficiently without any delay. The main aim of this paper is to perform a systematic study, in order to explore the prevailing research about the implementations of computer vision approaches based on the deep learning algorithms and Convolutional Neural Networks (CNN) [13]. They selected a total of 119 papers, which were classified according to field of interest, network type, learning paradigm, research and contribution type. This study reveals that this field is a promising and trending area for research. In this research, to explore the computer vision task they choose human pose estimation in video frames. After the study, they proposed three different research direction related to- improving the existing CNN implementations, using the Recurrent Neural Networks (RNNs) for the estimation of human pose and finally depend on unsupervised learning model to train neural networks.

By utilizing both regular and anomalous video, they suggested a technique for learning anomalies [5]. It was also suggested to understand anomaly through the deep multiple-instance ranking framework by utilizing weakly labelled training videos, meaning that the training labels (anomalous or normal) are at the video level rather than the clip-level, in order to prevent annotating the anomalous segments or clips in training videos, which is quite time-consuming. In this method, they use multiple instance learning (MIL) to automatically develop a deep anomaly ranking model that forecasts higher anomaly scores for anomalous video segments by treating normal and anomalous videos as packets and video segments as instances. In order to more accurately localize anomaly during training, they also integrate sparsity and temporal smoothness requirements into the ranking loss function.

Their contribution is mainly a three-fold contribution [14]. Firstly, they proposed a two-stream ConvNet

architecture which includes spatial and temporal networks. Secondly, they demonstrate that a ConvNet is able to achieve very high performance in spite of any limited training data when trained using the multi-frame dense-optical flow. Finally, to increase the amount of training data and improve the performance multitask learning is applied to two different action classification datasets. Using the standard video actions benchmarks of UCF-101 and HMDB-51, this architecture is trained and evaluated. The usage of convolutional network shows good performance in video classification.

In this research [6], they developed an integrated two-stream convolutional network framework that can identify vehicular traffic in video surveillance data in real-time and track them even while identifying serious accidents. A spatial stream network for item detection and a temporal stream network that uses motion features for several object tracking includes two paradigms. By combining motion and appearance features from these two networks, they can detect near-accidents. Furthermore, they show on a range of videos gathered from fisheye and overhead cameras that their methods can be used in real-time and even at a frame rate that is faster than that of the video frame rate.

They present a residual learning framework [16] to make the training of networks easier that are considerably deeper than those used in previous papers. Instead of learning unreferenced functions, they explicitly reformulate the layers as the learning residual functions with the reference to the layer inputs. They provide complete empirical evidence showing that the optimization of these residual networks is easier, and can gain high accuracy with considerably high depth. They evaluate residual network on ImageNet dataset with a depth up to 152 layers which is 8 times deeper than the VGG nets but still having a drawbackthat is lower complexity.

They proposed Residual Attention Network [17], a convolutional neural network using attention mechanism which can incorporate with state-of-art feed forward network architecture in an end-to-end training fashion. The Attention Modules that produce attention-aware features are stacked to create the Residual Attention Network. As the layers get deeper, the attention-aware features from various modules adjust. The feedforward and feedback attention processes are combined into a single feedforward process inside of each Attention Module using a bottom-up top-down feedforward structure. To train extremely deep Residual Attention Networks that are easily scalable up to hundreds of layers, they crucially proposed attention residual learning.

III. METHODOLOGY

➢ *Proposed System*

The proposed methodology is mainly based on computer vision and deep learning technique used in video analytics. For this purpose, video data containing the normal cases and abnormal cases of vehicles. The captured video is segmented into frames using the computer vision technique where OpenCV is used. For further data processing, deep learning neural network is used. The usage of deep learning technique helps in identifying the features by itself making the detection more effective and at ease. Within the deep learning technique, the convolutional neural network which is a type of artificial neural network is widely used for object classification and recognition. Recent study has shown that the convolutional network will be more accurate and efficient to train if network contains less number of layers. This in turn give rise to the new neural network that is the Dense Convolutional Network (DenseNet). For the abnormal vehicle detection, the DenseNet is used in this proposed system. The frames obtained from the videos are used for training the neural network. The dataset is classified into three sets- training set, testing test and validation set. The data is passed over different layers of the network for feature extraction and classification. The integration of computer vision along with neural network increase the system performance.

➢ *Abnormal Vehicle Behavior Detection Framework*

The process of abnormal behavior detection includes several processes. The abnormal vehicle behavior starts with video sequence obtained from video surveillance. The neural network cannot directly handle video data so before forwarding the input data to the DenseNet, the first step considered is the conversion of video sequences into corresponding frames. The usage of computer vision techniques supports the pre-processing of data using the libraries in OpenCV. It includes functions that can be used for pre-processing data. The pre-processing includes: resizing, removing noises etc. The pre-processed image is given as the input to the network. During the training phase, the model is trained using the training data which is labelled as accident and no accident. In this training phase the system will extract all the relevant feature by itself. In the testing phase, some data is given as input to the trained model to predict the output. If it detects the changes correctly, we can justify that the system is accurate and efficient.

- *Computer Vision*

Computer vision enables the system to study and understand images and can derive important features from them. Images can be further identified and processed using computer vision. OpenCV is an image processing library which contain programming functions. OpenCV stands for Open-Source Computer Vision Library which facilitates the research in the computer vision domain provide strong support for the advanced CPU-based projects. It is freely accessible for both commercial as well as academic purpose. The programming languages like Python, C++ and Java as well as the commonly used operating systems like Linux, Windows, iOS, Mac OS, and Android are supported by OpenCV. In real time applications, the computational efficiency is an important factor and OpenCV was designed mainly for this purpose. The multi-core processing concept is used in OpenCV. OpenCV also supports a wide range of deep learning libraries like PyTorch which provide easy implementation of neural networks where the data processing is supported well. The video can be pre-processed using OpenCV library where many function for different purpose

can be used. And our main aim is to convert the video sequence into frames with the help of OpenCV.

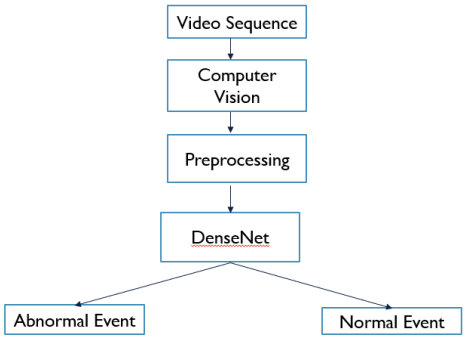

Fig 1 Proposed System Framework

- *Pre-processing*

The pre-processing step includes normalization technique. Normalization refers to normalizing the data dimensions so that they are of approximately the same scale. For Image data there are two common ways of achieving this normalization. One is to divide each dimension by its standard deviation, once it has been zero-centered.

The pre-processing of images can de done using normalization technique, which normalizes each dimension to a specified one so that the min and max value along the dimension will be in a particular range. The purpose of pre-processing step is to make the different values of input image to similar one. If the input data have similar values no pre-processing is required, but they should be of approximately equal importance to the learning algorithm. The relative scales of pixels will be already approximately equal a pixel value which range from 0 to 255, so it is not strictly necessary to perform this additional pre-processing step.

Convolutional Neural Network is a deep learning algorithm, it takes input as image and assigns weight and bias to the different object aspects to differentiate one from other. Convolutional Neural Network is a type of artificial neural network that is used in image processing and recognition which is designed for process the image in pixel data. The Convolutional Neural Networks are used in image classification where the valuable features are recognized by the Convolutional Neural Network by identifying different objects from the images. Less pre-processing steps are only required in Convolutional neural network.

The vanishing gradient problem in the traditional convolutional neural network occur as the layer get deeper which is considered as a problem to overcome. As a solution to this, the Dense Convolutional Network (DenseNet) is developed, where each layer is connected to every other layers in a feed-forward fashion. But in case of the traditional convolutional networks, it contains L layers having L connections that is one layer between each layer and its succeeding layer. In DenseNet there is a total of L(L+1)/2 direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. Concatenation is used for this. Each layer is getting a collective knowledge from all the preceding layers.

The neural network can be thinner and more compact, because each layer receives feature maps from all preceding layers, that is number of channels can be fewer. The growth rate k is the additional number of channels for each layer. So, it will have higher computational efficiency and memory efficiency. The several compelling advantages of DenseNets are- they reduce the vanishing-gradient problem, the reusage of features are encouraged, substantially reduce the number of parameters and the feature propagation is strengthened.

In the image below, consider a convolutional network in which the single image x0 is given as the input. The network comprises of L layers, each of the layers implements a non-linear transformation Hℓ (.), where Hℓ (.) is the composite function of three operations such as BN- Batch Normalization, ReLU -rectified linear units, Conv-Convolution. xℓ is denoted as the output of the ℓ th layer. The components of the DenseNet includes [19]:Connectivity, DenseBlocks, Growth Rate, Bottleneck layers.

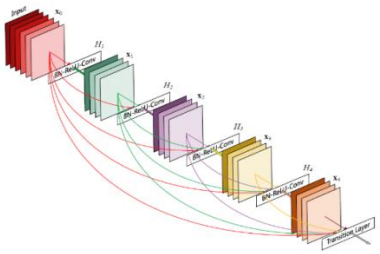

Fig 2 Dense Block

✓ *Connectivity*

The feature maps from all the preceding layers undergo concatenation operation instead of the summing operation as used in normal convolutional layers and the output of that concatenation operation is used as inputs in each layer. Therefore, only fewer parameters are necessary for the DenseNet when compared with the traditional Convolutional Neural Network, and this in turn reduce or discard all the redundant feature, thus allowing feature reuse. Therefore, the feature-maps from all the preceding layers, x0,...,x ℓ -1, the ℓ th layer receives the input as:

$$x\ell = H\ell([x0,x1,...,x\ \ell\ -1]\ ) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where, [x0,x1,...,x ℓ -1] represents the feature-maps concatenation, that is the output obtained in all the preceding layers ℓ (0,..., ℓ -1). The concatenation of Hℓ is done to transform it into a single tensor to make the implementation easy and the multiple inputs of Hℓ is used for concatenation.

✓ *Dense Blocks*

When the size of feature maps changes, the usage of the concatenation operation is not possible in such cases. To obtain higher computational speed, down-sampling must be done in layers which help in reducing the size of the feature

maps by reduction in dimensionality, which is considered as an essential step of Convolutional Neural Network.

To enable this functionality, the DenseNet is divided into DenseBlocks, within each DenseBlocks the dimensionality of feature maps is constant, but the number of filters used will change. Transition Layers are the layers in between blocks which helps in reducing the number of channels to half of that of the existing channels.

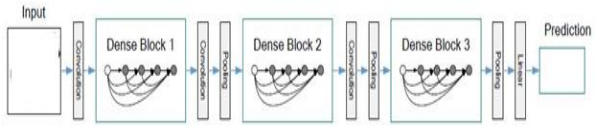

Fig 3 Propagation of Input through DenseNet

In the above image, three dense blocks with a deep DenseNet is shown. Through the convolution and pooling operations down-sampling (i.e. feature-maps size is changed) is performed in the transition layers that is the layers between two adjacent blocks. To enable feature concatenation the size of feature map is kept same within the dense block which is considered as an advantage of this neural network.

The first step of extracting the useful or important information from images is the convolutional layer. Using the small squares of input data the image features are learned for conserving the relationship between the pixels of the frames or images with the help of convolution. By taking the two inputs- matrix and kernel, it is implemented mathematically using the operations. The matrix is the part of the image.

When the given image is too large, the number of parameters are reduced using pooling layers, which is considered as the main job of pooling layers. The spatial pooling which is also termed as the down-sampling or sub-sampling, helps in maintaining the most relevant information by diminishing the dimensionality of each Feature Map.

✓ Growth Rate

The features can be considered as a global state of the neural network. After the propagation through each dense layer by adding ' ƙ ' features on top of the existing features with each layer, the feature map size increases. The growth rate of the network is referred as ' ƙ '. This parameter ' ƙ ' can control the amount of information added in each layer of the neural network. If k feature maps are produced by each Hℓ function, then the ℓ th layer has

$$ƙℓ=ƙ0+ƙ*(ℓ-1)\ldots\ldots\ldots\ldots\ldots (2)$$

input feature-maps where, k0 is defined as the number of channels in the input layer. DenseNet have very thin layers when compared with the existing neural network architectures

✓ *Bottleneck Layers*

In case of more layers, the number of inputs can also be quite high, even though each layer produces only k output feature-maps. Therefore, a bottleneck layer is a 1x1 convolution layer which is introduced before each 3x3 convolution which can improve the speed of computations and efficiency of the network.

The deeper layers use only the extracted feature by spreading the weights of all layers within the dense block and the transition layer[20]. Since the output from the transition layers contain many redundant features, second and third dense block layers assign the least weights as the output of the transition layers. As the model become deeper, more high-level and relevant features are generated and it seems to have high concentration towards the end of the feature maps while using the entire dense block weights by the final layers.

The DenseNet used in this experiment has the dense blocks that each has an equal number of layers. Before entering the first dense block, a convolution with 16 (or twice the growth rate for DenseNet) output channels is performed on the input images. For convolutional layers with kernel size 3x3, each side of the inputs is zero-padded by one pixel to keep the feature-map size fixed. Here 1x1 convolution followed by 2x2 average pooling as transition layers between two contiguous dense blocks is used. At the end of the last dense block, a global average pooling is performed and then a softmax classifier is attached. The feature-map sizes in the three dense blocks are 32x32, 16x16, and 8x8, respectively.

A DenseNet structure with 4 dense blocks on 224x224 input images is used. The initial convolution layer comprises 2k convolutions of size 7x7 with stride 2; the number of feature-maps in all other layers also follow from setting k.

The main advantages of using DenseNet includes:

- Parameter efficiency – In DenseNet only limited number of parameters are added in each layers that is only 12 kernels are learned per layers.
- Implicit deep supervision – The gradient flow is improved through the network that is the feature maps in each layers have direct access to the loss function and its gradient.

➢ *Dependencies*

- *Anaconda*

Anaconda is an open-source software and environment management system used for data analytics, data processing, etc. Anaconda runs on Linux, Windows and MacOS. Anaconda can be used for running, installing and updating the packages easily. It can switch between the local environment on the computer.

- *OpenCV*

OpenCV is an open-source library. OpenCV is mainly used for image processing, computer vision, and machine learning tasks. It plays an important role in the real-time operation with data which have great impact in today's systems. By using OpenCV, the image and video data can be

processed to identify for various applications like recognition of human handwriting, objects or faces. When it is integrated with various libraries, such as NumPy, python become more capable of processing with the array structure of OpenCV for analysis.

To Identify the image pattern and its corresponding various features, vector space is used and perform some mathematical operations on these features. The main purpose of computer vision is to understand the content of the images given. It extracts the necessary description from the pictures, which may be an object, three-dimension model and a text description and so on.

- *Visual Studio Code*

Visual Studio Code is a powerful source code editor that can run on desktop and it is a lightweight code editor. It is available for Linux, macOS and Windows. It contains built-in support for Node.js, JavaScript and TypeScript. It has a ridiculous ecosystem of the extensions for other programming languages. Visual Studio (VS) Code is used to correct and restore coding errors which is cloud and web-based applications. VS code is an open-source code editor.

- *PyTorch*

PyTorch is used for python programming which is an open-source library developed using Torch that can be used in machine learning library. It is a free open-source library and was developed by the AI Research lab of Facebook. It mainly focuses on natural language processing, computer vision and deep learning, and several other applications. Using PyTorch, a programmer can easily build any complex neural network, since it has a core data structure- Tensor and multi-dimensional array such as Numpy arrays.

The features like flexibility, speed, and ease of use makes PyTorch to be used frequently in the most current industries and in the research areas. PyTorch can run project in a fast manner which makes the PyTorch one of the top deep learning tools. PyTorch is one of the best open-sourcelibraryfor image classification, object detection and many other applications. The version of PyTorch used in this work is PyTorch 1.0.1. Using PyTorch, a programmer can process images and videos to develop a highly accurate and precise computer vision model.

IV. RESULTS

The proposed methodology for detecting abnormal vehicle behavior can process the data in an efficient way using deep learning and computer vision. The result proves the efficiency of the system. Comparing with the existing system, the usage of DenseNet makes the framework more effective since it reduces the parameters considered. To verify the robustness of the proposed system, here the video with different situations from the Internet is used. This system correctly identified the unusual event and usual event in an efficient way using the DenseNet with less parameters and classified the events into accident and no accident. The detection framework classifies each frame from a video with an accuracy of 97 percentage. The obtained output is shown below. The Fig. 4 shows the abnormal behavior of the vehicle where the accident has occurred. In Fig. 5, the normal behavior of the vehicle is detected.

Fig 4 Abnormal Behavior

Fig 5 Normal Behavior

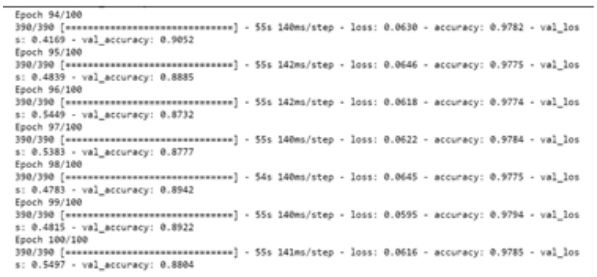

Fig 6 Testing Result

V. CONCLUSION

In smart security field, the abnormal behavior detection from videos is a trending and vast research area. Variety of definitions can be given to abnormal behavior which can be done based on the different surveillance video objects and surveillance scenes. Among different abnormal behavior, the research area mainly focuses on abnormal behaviors detection among vehicles. The main focus of this research is on the detection of the abnormal behaviors. For the abnormal behavior detection, deep learning algorithm-based framework is used. First, the preprocessing of input video is done using the OpenCV library available in computer vision. When the preprocessed data is loaded to DenseNet, it will process this input through different layers. The network is trained using the dataset and it will detect whether the frames are abnormal or normal. The number of parameters get reduced with the help of DenseNet which in turn increase the performance of the system. The result predicts the accuracy has reached at a better level.

➢ *Scope for Further Work*

The proposed system can be implemented in smart cities and intelligent traffic system. The implementation of

such system will help in identifying the accidents fast and to take immediate actions so as to reduce the death rate and can increase human security. In future the system can be trained on large amount of data thereby increasing the system accuracy to next level. By training with variety of data, the system will able to classify wide ranges of data into correct class.

REFERENCES

[1]. Robles-Serrano, Sergio, German Sanchez-Torres, and John Branch-Bedoya. 2021. "Automatic Detection of Traffic Accidents from Video Using Deep Learning Techniques Computers 10", no. 11: 148. https://doi.org/10.3390/computers10110148.

[2]. Doshi, Keval and Y. Yilmaz. "Fast Unsupervised Anomaly Detection in Traffic Videos". 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020): 2658-2664.

[3]. Iván González-Díaz, Tomás Martínez-Cortés, Ascensión Gallardo-Antolín, Fernando Díaz-de-María, "Temporal segmentation and keyframe selection methods for user-generated video search-based annotation", Expert Systems with Applications, Volume 42, Issue 1, 2015, Pages 488-502, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2014.08.001.

[4]. Wang, C., Musaev, A., Sheinidashtegol, P., Atkison, T. (2019). "Towards Detection of Abnormal Vehicle Behavior Using Traffic Cameras". In: Chen, K., Seshadri, S., Zhang, LJ. (eds) Big Data – BigData 2019. BIGDATA 2019. Lecture Notes in Computer Science(), vol 11514. Springer, Cham. https://doi.org/10.1007/978-3-030-23551-2_9.

[5]. W. Sultani, C. Chen and M. Shah, "Real-World Anomaly Detection in Surveillance Videos",2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479-6488, doi: 10.1109/CVPR.2018.00678.

[6]. Huang, Xiaohui & He, Pan & Rangarajan, Anand & Ranka, Sanjay. (2019). Intelligent Intersection: "Two-Stream Convolutional Networks for Real-time Near Accident Detection in Traffic Video".

[7]. D. Singh and C. K. Mohan, "Deep Spatio-Temporal Representation for Detection of Road Accidents Using Stacked Autoencoder", in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp. 879-887, March 2019, doi: 10.1109/TITS.2018.2835308.

[8]. Maaloul, B.; Taleb-Ahmed, A.; Niar, S.; Harb, N.; Valderrama, C. "Adaptive video-based algorithm for accident detection on highways". In Proceedings of the 2017 12th IEEE International Symposium on Industrial Embedded Systems, SIES 2017, Toulouse, France, 14–16 June 2017.

[9]. Arinaldi, A.; Pradana, J.A.; Gurusinga, A.A. "Detection and Classification of Vehicles for Traffic Video Analytics". Procedia Comput. Sci. 2018, 144, 259–268.

[10]. D. Tian, C. Zhang, X. Duan and X. Wang, "An Automatic Car Accident Detection Method Based on Cooperative Vehicle Infrastructure Systems," in IEEE Access, vol. 7, pp. 127453-127463, 2019, doi: 10.1109/ACCESS.2019.2939532.

[11]. Enyuan Jiang, Xuejun Wang. 2015. "Abnormal Vehicle behavior detection based on tracking trajectory". In Journal of Computer and Communications. http://dx.doi.org/10.4236/jcc.2015.

[12]. Li, L., Hu, Z. & Yang, X. "Intelligent Analysis of Abnormal Vehicle Behavior Based on a Digital Twin". J. Shanghai Jiaotong Univ. (Sci.) 26, 587–597 (2021). https://doi.org/10.1007/s12204-021-2348-7

[13]. Nishani, E.; Cico, B." Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation". In Proceedings of the 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, Montenegro, 11–15 June 2017; pp. 11–14.

[14]. Qiao, Han & Liu, Shuang & Xu, Qingzhen & Liu, Shouqiang & Yang, Wanggan. (2021). "Two-Stream Convolutional Neural Network for Video Action Recognition". KSII Transactions on Internet and Information Systems. Vol. 15. 3668-3683. 10.3837/tiis.2021.10.011.

[15]. W. Wang, J. Shen and L. Shao, "Video Salient Object Detection via Fully Convolutional Networks", in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 38-49, Jan. 2018, doi: 10.1109/TIP.2017.2754941.

[16]. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[17]. Wang, Fei & Jiang, Mengqing & Qian, Chen & Yang, Shuo & Li, Cheng & Zhang, Honggang & Wang, Xiaogang & Tang, Xiaoou. (2017). Residual Attention Network for Image Classification. 6450-6458. 10.1109/CVPR.2017.683.

[18]. Gao Huang, Zhuang Liu and Laurens van der Maaten. Jan 2018 Densely Connected Convolutional Networks.

[19]. OpenGenus IQ: Computing Expertise & Legacy-Architecture of DenseNet-121[Online] https://iq.opengenus.org/architecture-of-densenet121.

[20]. Noh, Kyoung & Choi, Jiho & Hong, Jin & Park, Kang. (2020). "Finger-Vein Recognition Based on Densely Connected Convolutional Network Using Score-Level Fusion With Shape and Texture Images". IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2996646.