ISSN No:-2456-2165

Semantic Video Mining for Accident Detection

Rohith G
Dept. of Computer Science
Sahrdaya College of Engineering and Technology
Thrissur, India

Twinkle Roy
Dept. of Computer Science
Sahrdaya College of Engineering and Technology
Thrissur, India

Vishnu Narayan V
Dept. of Computer Science
Sahrdaya College of Engineering and Technology
Thrissur, India

Shery Shaju
Dept. of Computer Science
Sahrdaya College of Engineering and Technology
Thrissur, India

Ann Rija Paul
Assistant Prof. (Dept. of Computer Science)
Sahrdaya College of Engineering and Technology Thrissur, India

Abstract:- This paper depicts the efficient use of CCTV for traffic monitoring and accident detection. The system which is designed has the capability to classify the accident and can give alerts when necessary. Nowadays we have CCTVs on most of the roads, but its capabilities are being underused. There also doesn't exist an efficient system to detect and classify accidents in real time. So many deaths occur because of undetected accidents. It is difficult to detect accidents in remote places and at night. The proposed system can identify and classify accidents as major and minor. It can automatically alert the authorities if it deals with a major accident. Using this system the response time on accident can be decreased by processing the visuals of CCTV.

In this system different image processing and machine learning techniques are used. The dataset for training is extracted from the visuals of already occurred accidents. Accidents mainly occur because of careless driving, alcohol consumption and over speeding. Another main cause of death due to accidents are the delay in reporting accidents since there doesn't exist any automated systems. Accidents are mainly reported by the public or by traffic authorities. We can save many lives by detecting and reporting the accident quickly. In this system live video is captured from the CCTV's and it is processed to detect accidents. In this system the YOLOV3 algorithm is used for object detection. Nowadays traffic monitoring has a greater significance. CCTV's can be used to detect accidents since it is present in most of the roads. It is only used for traffic monitoring. Normally accidents can be classified as two classes major and minor. The proposed system is able to classify the accident as major or minor by object detection and tracking methodologies. Every accident doesn't need emergency support. Only major accidents must be handled quickly. The proposed system captures the video and undergo object detection algorithms to identify the different objects like vehicles and people. After the detection phase the system will try to extract the features of the vehicles. The features like length, width and centroid are extracted to classify the vehicle accordingly. The vehicle count is also detected, which can be used for traffic congestion control.

Keywords:- *YOLO V3, SSD , Faster RCNN , RCNN.*

I. INTRODUCTION

Population is increasing day by day. Along with the increase in population, the number of vehicles are also increasing. It is known that the present traffic management system is not efficient. Millions of people die in road accidents every year. This is not only because of the increase in the number of vehicles. There doesn't exist any proper system to detect accidents and to alert the authorities. The higher response time for the arrival of the emergency system causes many precious lives. Normally the road accidents are reported by the people near the accident. Many of the cases the people who witness the accident are not willing to alert the authorities and instead they are busy taking selfies. These types of negligence are causing precious lives. Also we have CCTVs installed on most of the roads. But the CCTV's are not used efficiently. In the modern era where the technology is growing faster we are still dependent on human power for traffic monitoring.Since the number of traffic authorities is low and the number of vehicle users is high it is difficult to control them. Many people lose their life because of undetected accidents. It is difficult to monitor vehicles all the time for humans. But it is easy and possible by using CCTV's. The proposed system uses CCTV for traffic monitoring and accident detection with less human interventions.

The system captures live video from CCTV and processes it to detect accidents in real time. Surveillance cameras are installed in most of the roads. This is mounted on a pole which can give clear vision of vehicles on the road. Present system uses these visuals to monitor and control the traffic manually.

There will be control rooms to monitor the CCTV visuals. The number of traffic authorities is less and one person should have to monitor multiple CCTV visuals which is very difficult and not efficient. After the occurrence of an every minute is very critical, every extra minute that it takes for emergency services to arrive can cost a life. So it is necessary to implement a system that can initially detect and track vehicle accidents automatically so that it can be reported to the concerned authorities quickly so that the emergency services can arrive faster to save many lives. The proposed system can detect and classify accidents in real time.

The video which is captured by the surveillance camera undergo a set of image processing and deep learning techniques to create an efficient traffic surveillance system. Frames are detected from the captured video and transformations like foreground and background extraction are carried out to detect the vehicle. After vehicle detection features such as length, width and centroid are extracted for the classification of vehicles. The vehicles are classified as light, medium and heavy weighted vehicles. The count of the vehicles are also detected based on a reference line which can be made use for traffic monitoring. Using this system vehicle attributes such as variation in acceleration, change in position, variation in area and the variation rate of inclination. These attributes are vital for the detection and classification of accidents. We are getting tremendous data from the surveillance camera. It is difficult to store these huge amounts of data for a long time. Hence the data obtained from surveillance cameras must be summarized to efficiently store it for a long time. In this system after detecting accidents we are sending alerts to the authorities with a timestamp. The video will be summarized and only the part of the accident will be saved for future reference. The main merit of this system is that it doesn't include any physical detectors which should be maintained periodically. No additional installments are required on road for the implementation of this system.

II. RELATED WORKS

There exist different systems for traffic monitoring and accident detection. The latest works are related to apps which can detect car accidents using the internal sensors of the smartphone and sends an emergency notification with the location to pre-selected emergency contacts. This is done using the accelerometer and GPS sensors present in the smartphone. By using this system we can send help as soon as possible. But the main disadvantage of this system is that classification of the accident is not possible. The accident cannot be confirmed by only looking at the sensors inside the smartphone. Alerting authorities on minor accidents can cause problems. Even though these apps are useful it is not efficient. Research is going on for increasing the efficiency of these kinds of apps to detect accidents using smartphone sensors. These types of apps are not suitable for all types of vehicles. In the case of traffic surveillance, another outstanding issue is the speed control on the road. There are different technologies to carry out speed calculations of different types of vehicles. Doppler radar in down-the-road (DTR) configuration is one of the widely used techniques for speed control on public roads. In DTR configuration, there will be a line of travel of the target vehicle where the beam of the antenna is directed along. Target identification is the main shortcoming of DTR Doppler radar. Doppler radar devices have a wide beam width since they are not target selective, and the speed measurement is erroneous when there occurs the presence of two or more targets in the radar beam.

There are several drawbacks in Doppler radar in down-the-road configuration. To minimize this drawback Doppler radar in across-the-road (ATR) configuration is introduced. Vehicle speed is detected by directing the microwave beam across instead of down or along, the road in this system. When compared with DTR the main advantage over ATR configuration is that the operational area of its beam is reduced. This can drastically reduce the target identification problem. These assumptions become a failure when traffic volume is very denser. The added complexity of an ATR radar is the cosine effect. Radial velocity of the target vehicle can be measured using the Doppler radar. It means that the value which is been measured is the product of the speed of the vehicle and the cosine of the angle between the beam and the direction of motion. The angle should be known so that the radar device will be able to correct the cosine effect. In most of the cases calibration is needed because cosine effect can induce uncertainty to the measurement process.

In order to solve the problem of target identification, laser traffic radars are used in DTR configuration. The narrow beam width of these devices will help to select individual vehicles. The technique behind the measuring of speed using this system is that it uses delay history of a burst of laser pulses. One of the drawbacks of this system is that it has to concentrate on a unique target at a time. It means that different lasers must be used for every road line so that it can control the whole road. In this system the laser pulses would all strike on a flat surface and it would be perpendicular to the path of the light wave. In general the waves strike the front or rear of the vehicle and these surfaces will be irregular. The output obtained here is the reflected laser pulse which is dispersed in time. This is due to the different transit times for different portions of the reflected beam which is also dispersed as angle in some cases. So there occurs a confusion in the measurement because of the multiple reflection from the neighboring objects. The ATR laser system can employ two laser beams which can be operated on the time-distance principle which is not possible in DTR laser devices. There will be horizontal bars to which the lasers are mounted which transmit parallel light beams that are separated by a known distance. The beams are directed across the road which will be perpendicular to the direction of traffic flow and it will be done by the equipment. A vehicle is detected by the beam by sensing the changes in the intensity of light reflected. One of the major limitations of this system is that it can only measure the speed of one target at a specific time and the system fails if two vehicles in adjacent lanes overlap each other.

In order to solve the problem of several vehicles in the radar beam a system called range Doppler is introduced. These systems work on the basis of echoes of vehicles located at different distances to the radar and it can discriminate against them. This is the main advantage of this system over the capabilities of current Doppler or laser traffic radars. They have to concentrate on a unique target at a time. This system is also in development so that the association of each echo or speed measurement to each specific target has yet to be solved. Suppose two targets are located at different road lanes and it could be detected by using the same range bin by the radar. Nowadays camera based systems are used to measure the mean speed. The camera can be used to register the license plate of each and every vehicle that is driven in a stretch which can save transit time. Another camera can be used to repeat the same process. By calculating the distance between both cameras and both transit times it is possible for the system to compute the average velocity of every vehicle. The main disadvantage of this system is that it can only compute the average speed. The top speed of the vehicle cannot be identified. Another alternative solution discovered to solve these problems are collision warning, collision avoidance, and adaptive mono pulse Doppler radar. In this Doppler radar system a mono pulse antenna scheme is introduced in order to track multiple targets. It can generate range, velocity, and azimuth angle output data for each vehicle.

III. PROPOSED SYSTEM

We propose a novel accident detection and an accident classification model which could detect the accidents and classify the accidents as major or minor and report the accidents based on their priority. The accident detection is based on object detection and tracking methodologies and major and minor accident detection as classification methodology. The alert thus generated will contain the time stamp and camera ID which helps the appropriate authorities to track the location to take the necessary steps to help the ones affected from the accident. Our proposed accident detection model could be easily explained as modules. The Overall proposed system contains four modules. They are object detection and tracking module, feature extraction module, Accident classification module and alert system.

A. *Object Detection and Tracking*

In order to track the object from the video and extract features from the same, we have to detect the object in each frame of the video and its location in the subsequent frames should be tracked. The bounding box is usually drawn if required around this object detected. It helps the user to visualize the object tracking on the screen and could identify if the object tracking mechanism identifies the object correctly and marks the same. The features like speed, acceleration of the vehicle could be identified with the same.

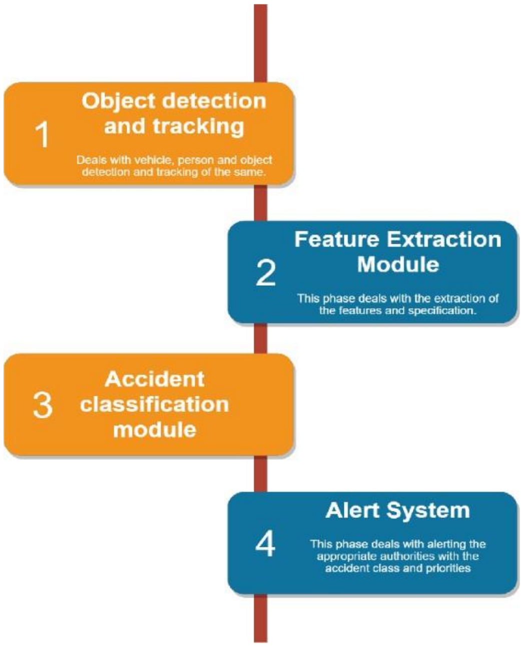
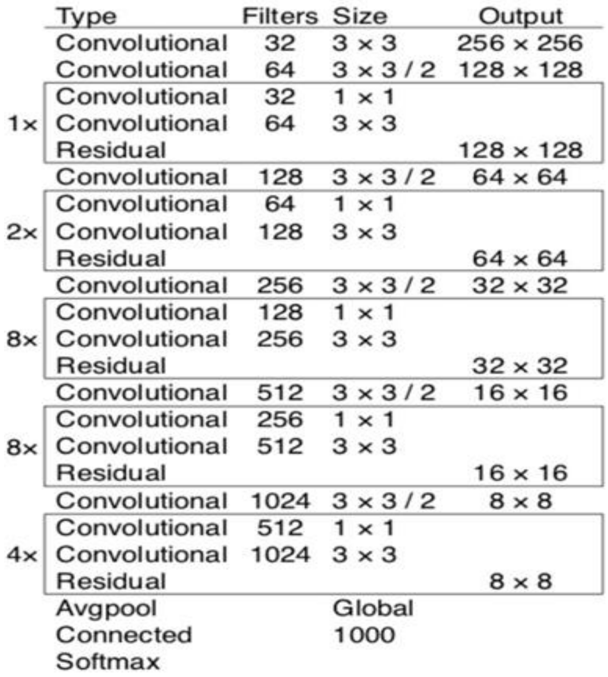

1
Object detection and tracking
Deals with vehicle, person and object detection and tracking of the same.
2
Feature Extraction Module
This phase deals with the extraction of the features and specification.
3
Accident classification module
4
Alert System
This phase deals with alerting the appropriate authorities with the accident class and priorities

Fig. 1:- Phases of accident detection and alert



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 2:- Darknet 53 [35]

There are many object detection algorithms based on machine learning as well as deep learning. Machine learning based algorithms are usually based on SVM (Support Vector Ma- chines) and deep learning performs much faster when compared to machine learning models. Convolution Neural networks (CNN) under the deep learning models to detect objects performs faster than the other object detection algorithms. In this project, we propose YOLO V3 CNN based on darknet. It does object detection in real time. When comparing the performance with other object detection algorithms it stood high in the case of speed but only an average in the case of accuracy. Instance segmentation performs better than YOLO V3 in terms of accuracy but is slower with average hardware availability. As the high end hardware is still not technically viable for implementing along the existing systems, we stick on to YOLOV3 for object detection.YOLO is a combination of object locator as well as object recognizer. First the object from the given images are located and then the objects located are recognized to determine what object is the same and will group into a class of objects or will be left unidentified. YOLO first divides an image into 13 x 13 small images. The size of these 169 cells will depend on the size of the input given to the YOLO model. Each of these divided cells are responsible for the recognizing of the objects correctly. Each of these divided boxes will predict a confidence value for certain objects. Combining these confidence values together and grouping the boxes back will identify the object and will mark a bounding box based on the coordinates predicted along with an overall confidence value for the predicted object under a class. The technique followed here is non-maximum suppression.

Steps Involved in Training the YOLO V3 model Labeling of dataset (images) Train Test file generation Anchor generation Train model to obtain weight

1) *Labelling of dataset:* : The dataset is a collection of images of the vehicles as well as accidents, damaged vehicles etc. Each of the images have to be labelled under the different class we have created. The labelling is done using a labelling program written in python. We have to mark the bounding box on the objects correctly under the correct class. The number of images under each class should be at least 1000, more the number of labelled images given as input more the accuracy. After generating the images, the python program will create a file for each image which will contain the bounding box of the objects under different classes. These files along with the images will be used in the upcoming steps. The files could be in XML PASCAL format or may be in normal text files. YOLO V3 supports both the file types and processing could be done based on any of the two file types.

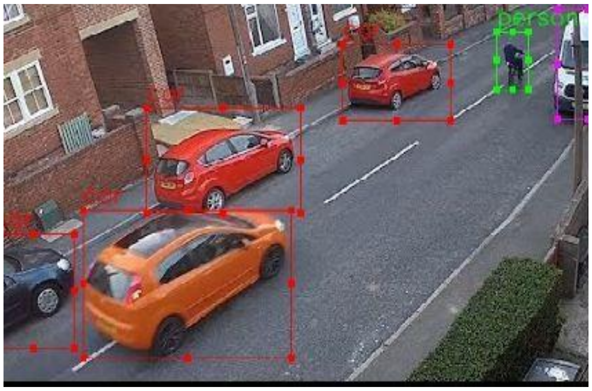


Fig. 3:- Labelling of vehicles

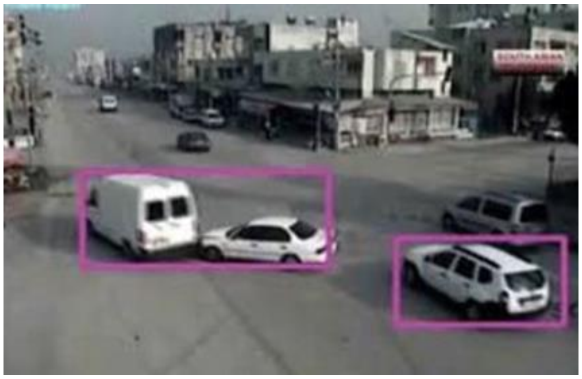


Fig. 4:- Labelling of accident

2) *Train Test Split:* The labelled data should be split into training and testing sets. Training set is used to train the model and the testing set will be used to check the accuracy of the model that we have trained. The testing and training split value could be in any way. In our case we set 80 percent as the training data and the rest 20 percent as the testing dataset. This will create two folders with each type of images along with its corresponding label text or Pascal file in the folders.

3) *Anchor generation:* Anchor generation is done based on the clustering of all the width and height of the input images that we have given and it will cluster all the images into a certain width and height ratio. So instead of predicting a wide range of width and height ratio YOLOV3 will limit the the object detection to the clustered classes of width height ratio. The labelled images along with the labelled data files are given as input in this step and the output will be 2 aggregated final values. These values will be used in the training phase.

4) *Training Model :* The configuration file is changed with the anchor, filters and class. The number of classes is known.

B. already. Filters are calculated from an equation and the anchor values are already obtained in the above step. After editing the configuration file the dataset is ready for training. We train our model using google colab using the GPU of the specifications 1xTesla K80 , compute 3.7, having 2496 CUDA cores , 12GB GDDR5 VRAM. After training of the model the output is a weight file. We will use this weight file generated in the upcoming steps in order to perform object detection.

ISSN No:-2456-2165

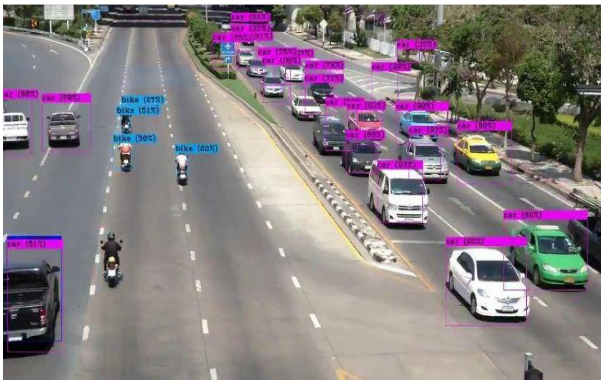
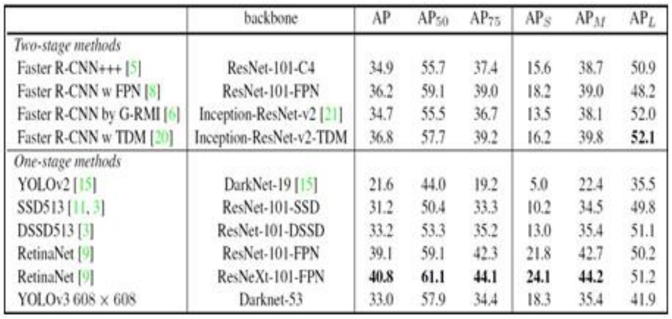


Fig. 5:- training of dataset

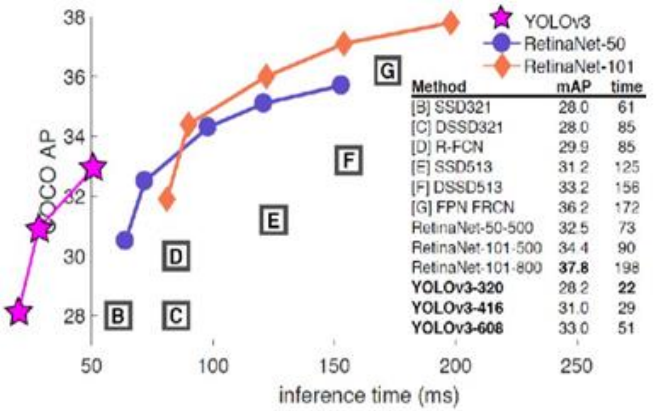
The initial weight generation training phase will consume time. It may last for hours but once the final weight is trained the object detection could be done in real time. The live stream videos of 30 fps (frames per second) could be given as input and the same could be processed in real-time and the objects could be detected. The bounding box prediction is the prediction of the center and the x and y values which will generate the same. The object is tracked in each of the frames one after the other. This is done by tracking the centroid of the object detected and marking the movement of the centroid of the objects. Using this technique the path of the movement of the object could be easily detected. The speed also could be easily calculated from the distance travelled by the centroid of the object and the time taken for the same. Thus various parameters could be obtained from the same. YOLOV3 convolutional neural network is thus a combination of the object locator and classifier.



	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++ [5]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [8]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [6]	Inception-ResNet-v2 [21]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [20]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	**52.1**
One-stage methods							
YOLOv2 [15]	DarkNet-19 [15]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [11, 3]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [3]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [9]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [9]	ResNeXt-101-FPN	**40.8**	**61.1**	**44.1**	**24.1**	**44.2**	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

Table 1:- Comparison of Yolov3 With Other State-Of-The-Art-Models [35]

YOLOV3 will predict the objects faster than SSD but SSD stands a bit higher in terms of accuracy. When comparing with other algorithms YOLOV3 will rank higher in terms of speed. A comparison between the performance of different algorithms on the COCO dataset is shown in the below figure.



Method	mAP	time
[B] SSD321	28.0	61
[C] DSSD321	28.0	85
[D] R-FCN	29.9	85
[E] SSD513	31.2	125
[F] DSSD513	33.2	156
[G] FPN FRCN	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	**37.8**	198
YOLOv3-320	28.2	**22**
YOLOv3-416	31.0	29
YOLOv3-608	33.0	51

Fig. 6:- Algorithm performance comparison on the COCO dataset [34]

YOLOV3 Could identify the objects, classify the same and could give their moving coordinates also if they do but if two objects of the same class come into the frame YOLOV3 face issues in tracking the same. In our case, we need to identify each vehicles uniquely and its parameters have to extracted and stored. We cannot attain this from the direct output of the YOLOV3, thus we need to track the objects uniquely and an ID is to be given for the same.

The case of vehicles on the road, thousands of vehicles pass under the same surveillance camera. These same vehicles might pass through many other cameras. Thus in order to provide a unique ID for the same, it’s better to give a camera ID, Class ID and time stamp for creating a unique ID. The final ID will look similar to Eg: CAM23 Car3 13:16:17:10 03:04:2020 , thus with the help of the camera ID the location of the camera could be identified, The vehicle if more than one enters to the frame each vehicle under the same class will be numbered. The time stamp will contain the time in 24 hour format and the date of the tracked object. Thus we could uniquely identify the vehicles. If we need to track the same vehicles under different traffic surveillance videos the same could be done by aggregating the values together and finally attaining a result. Thus after tracking the objects uniquely now we have to extract features from the same.

C. *Feature Extraction Module*

In the feature extraction module the required features are extracted from the set of features. The informations such as overlapping between vehicles, stopping, velocity, differential motion vector of vehicles, and direction of each vehicle are mainly considered. if more than one vehicle are detected in a frame, we cannot consider single object features. In this case each pair of objects are considered and above mentioned factors are examined and calculate the probability of accidents.

1) *The basic information of object:* Initially all the basic information about the vehicle which is our object has to be extracted. These are obtained when tracking using YOLO V3 model and also by comparing the current frame and previous frame. The information obtained includes object ids, left, top coordinates, right, bottom coordinates, and center(x, y) coordinates.
2) *The overlap of objects:* In accident detection the overlapping feature is important. This can be obtained by comparing the left, right, top, and bottom coordinates of objects. If the left and right coordinates of a vehicle are in between the left or right of the other vehicle, then they are overlapped.
3) *The stop of objects:* The vehicles usually stop after an accident. So stopping vehicles can be considered as a factor. This feature can be estimated by comparing the difference between current coordinates and previous coordinates.
4) *Velocity :* The variation in velocity before and after the occurrence of accidents can be used. The average speed of a vehicle at a general intersection is 60 90km/h. Based on this we classify the velocity into three states such as fast, normal, and slow.
5) *Direction:* The direction between vehicles before and after helps to classify the collisions into broadside collision, a head-on collision, and a rear-end collision. This is calculated by finding the difference of center of the vehicle detected before and after the collision.
6) *Differential motion vector :* This facility helps in identifying the risks at the crossroads. Speed and direction of vehicle are identified before and after accidents. This is because both speed and direction change due to external force during accidents. The trio of features have three discrete values, such as up-down, up and rest, which represent the difference velocity vectors between past and present time. Upstate means differential motion vectors compared to the previous period. An up-down state means reducing the differential motion vectors after the upstate and the rest state. Fig. 2 compares the contrast speed and hazard of the two objects. The differential motion vector and the crash express the dotted line and line, respectively. Spontaneous upward movement means the variation and risk of a moving vector. We can examine the differential speed vector before and after the crash. This feature was chosen as the most important.

D. Accident classification module

The various parameters obtained from feature extraction, like change in velocities, change in direction, change in coordinates etc. before and after accidents, the analysis has been carried out to evaluate the plausibility of collected data. Based on these data collected, cross examination is done by the following methods:

Conservation of Momentum: For all impacts where the road or the wheel forces can be considered as negligible, the variation of momentum for a vehicle is equal but opposite to that of other vehicles. Hence the primary check can be done for momentum relationship. The change in momentum can be calculated and checked whether it crosses the threshold value. Here we consider the threshold as 10 km/h.

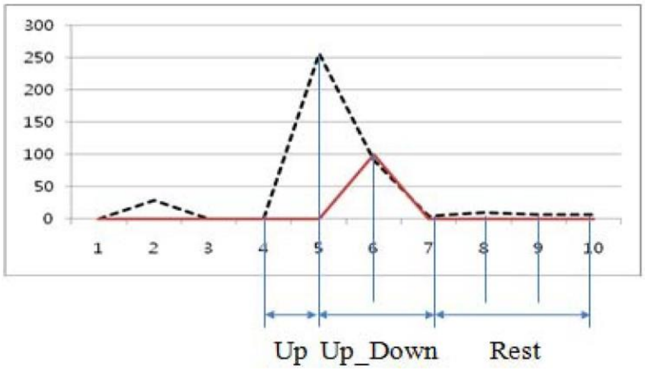


Fig. 7:- Differential motion vector [33]

1) *Velocity Triangles:* If the law of momentum conserva- tion is satisfied, a subsequent check on data can be made considering the velocity triangles. The vector sum of initial velocity and change in velocity must be equal to the post impact velocity.
2) *Energy Loss:* After both the Conservation of Momen- tum check and Velocity Triangles checks are done, law of energy conservation or by the related expression with Energy Equivalent Speed Equations.

E. Alert System

When all the modules are working, there is an active relationship between them. The accident detection module is connected to the server and the server is connected to the hospital.

The feature extraction module monitors for sudden changes in the position of the vehicle. When there is a sudden change in the parameters such as direction, velocity and momentum of the vehicle the accident detection module thinks the crash has occurred and a warning is generated in the graphical user interface. A warning message will be sent to notify the server about the hospital.

Message provides the user to cancel the notification. If the user regains consciousness at any time, he will be able to inform the hospital about it. To avoid unnecessary overhead of uniform and emergency services when conditions are not required. The user has 3 options, the user can cancel the message if the alert is present and the risk is not very severe, or if present. The user is alert, but still wants to notify the server, the message will be sent if the user does not cancel the message. If the user is alert and wants to cancel the notification to the user, buffer time is given to the server to cancel the notification.

The buffer time depends on the user's choice.

If the user does not cancel the notification on the server, the user's location, coordinates and timestamp during the crash is sent to the server. The server accepts these coordinates, assuming the user is unconscious. The server can be in any remote location and provide the appropriate service. It must be available at all times. Since the amount of data sent and received at any given time is very small, no additional costs should be applied to ensure availability at all times.

The server contains a database of all hospital IP addresses. Under the operation of the system and mechanism to identify the nearest hospital based on the coordinates obtained. After identifying the nearest hospital, the server notifies the hospital about the user's coordinates (location).

The Hospital receives a notification from the server about the user's location. The hospital uses a graphical user interface to display the user's location coordinates. Hospital operators can easily map coordinates on a map. This way the victim can provide medical emergency services in a short time. This can reduce time and mortality rate.

IV. EXPERIMENTAL EVALUATION

We propose a model that detects accidents from the video footage and inform the authorities about the accident. Here we are extracting the accident images from the cctv footage. The extraction of images comes under the field of computer vision along with image processing. Mainly object detection is used to focus an object based on its size, coordinates and categorize them to various fields. By object detection we can get two dimensional images which will provide us more details about space, size, orientation etc.

Object detection we can use CNN, R-CNN, Fast RCNN, YOLO and SSD. Early days we use CNN for object detection like face identification, voice identification, etc. CNN is a convolutional neural network which is under the category of feed forward network. It can extract certain characteristics or features from an image. There are three layers in CNN: the convolution layer, pooling layer and the fully connected layer. Convolution layers function is to focus on input requirements. Pooling layer is in between the convolution layer and it is used to increase the effectiveness of feature extraction. Fully connected layers represent the output layer in a convolution neural network. -CNN consists of three modules, one module is dependent on categorical production. Second module is concerned with extraction features. And the last module is SVM(Support Vector Machine).

Fast R-CNN is a network where the whole image can be taken as input. Here a convolution feature map is created which consists of a full image with different convolution layers. The regional detection proposals inputting is different in R-CNN and Fast R-CNN. In R-CNN proposals are inputted as pixels and in Fast R-CNN it is inputted as feature maps.

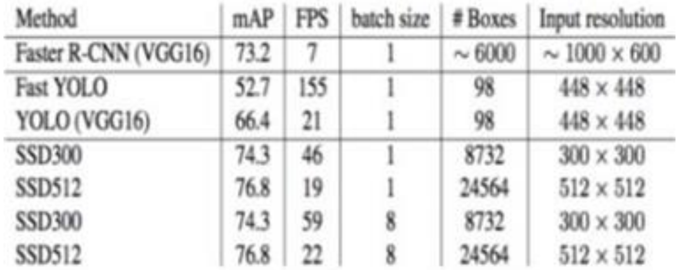
Faster R-CNN is a network where object detection is faster. Region proposal problems solution is found out by this network. Compared to all models its computation speed is very less. So the image resolution is less than the input original image.

YOLO-You Only Look Once is another object detection. In an image what are the objects and where the images are present can be detected by only one look at the image. Instead of classification YOLO uses regression. And it can separate the bounding boxes and class probabilities for every part in an image. Within a single analysis it can predict the bounding boxes with class probabilities, only a single network is used for this process. A single CNN can predict multiple bounded boxes and also weights are given for these bounding boxes.

YOLOv3 is different as it is using logistic regression to find the object within bounding boxes. If the ground truth is overlapped with the bounding box which is greater than any other bounding box then object score must be one. If a bounding box overlaps the ground truth by a threshold which is not best, that type of predictions can be disregarded.

A. CNN vs YOLO

Both Faster R-CNN and YOLO consider its core as Convo- lutional Neural Network. YOLO partitions the image before using CNN for processing the image whereas RCNN keeps the whole image as such and only division of proposals take place later. In YOLO the image is partitioned into grids.



Method	mAP	FPS	batch size	# Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	1	~ 6000	~ 1000 × 600
Fast YOLO	52.7	155	1	98	448 × 448
YOLO (VGG16)	66.4	21	1	98	448 × 448
SSD300	74.3	46	1	8732	300 × 300
SSD512	76.8	19	1	24564	512 × 512
SSD300	74.3	59	8	8732	300 × 300
SSD512	76.8	22	8	24564	512 × 512

Table 2:- SPEED AND PERFORMANCE COMPARISON [36]

Table given above contains the comparison of speed and performance of detectors. Fast YOLO is fastest but YOLO has the highest precision compared to Fast YOLO. So it is used for detection of objects. Fast YOLO is the fastest detector; it is two times precise as any other detectors within 52.7 percent map. YOLO can increase its map to 66.4 percent in its real time performance. When we consider the accuracy FasterRCNN is the most suitable algorithm. Super Fast YOLO can be chosen if accuracy is not given much importance. When we consider the error formation YOLO is having most of the localization errors. While Fast R-CNN has many background errors and only limited no of localization errors.

Thus we choose YOLO as the best object detection algorithm for detecting accidents and classifying it into major and minor accidents.

V. CONCLUSION AND FUTURE WORK

This paper is all about video extraction based accident detection from road traffic surveillance videos. We are extracting the image of accidents from crash videos and thus detect it as an accident. For this detection we use the YOLOv3 neural network which is more precise than any other neural network. The detected accidents are reported to the authorities to reduce the human interventions and to get immediate care for the human lives. Thus it is tested for various stages of road accidents. It is also tested for different types of collisions with different types of vehicles. The results show that this is a precise model for detecting accidents for traffic surveillance and alerting the authorities.

As a future work we can include the classification of the detected accidents into major and minor accidents. Thus major accidents can be reported to the nearby hospitals and minor accidents to the relatives. And also an extension to the current project the road events can also be detected and any road traffic violations can be found out.

ACKNOWLEDGEMENTS

The paper is prepared by taking assistance from various reference papers, we are thankful to them. We also express our gratitude to our professors and guides for helping us throughout the work.

REFERENCES

[1]. **Road Crash Statistics. (Sep. 2016).** [Online]. Available: https://asirt.org/ initiatives/informing-road-users/road-safety-facts/road- crash-statistics.

[2]. **C.-M. Tsai, L.-W. Kang, C.-W. Lin, and W. Lin,** "Scene-based movie summarization via role-community networks," .", *IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 11, pp. 1927–1940, Nov. 2013.*

[3]. **M. Tavassolipour, M. Karimian, and S. Kasaei,** "Event detection and summarization in soccer videos using Bayesian network and Copula,"

[4]. .", *IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 2, pp. 291–304, Feb. 2014.*

[5]. **S. Parthasarathy and T. Hasan,** "Automatic broadcast news sum- marization via rank classifiers and crowdsourced annotation," in Proc. IEEE ICASSP, Apr. 2015, pp. 5256–5260.

[6]. **M. Cote, F. Jean, A. B. Albu, and D. Capson**, "Video summarization for remote invigilation of online exams," .", *in Proc. IEEE WACV, Mar. 2016, pp. 1–9.*

[7]. G. J. Simon, P. J. Caraballo, T. M. Therneau, S. S. Cha, M.

[8]. **R. Castro**, and P. W. Li, "Extending association rule summarization techniques to assess risk of diabetes mellitus," .", *IEEE Trans. Knowl. Data Eng., vol. 27, no. 1, pp. 130–141, Jan. 2015.*

[9]. **T. Yao, T. Mei, and Y. Rui**, "Highlight detection with pairwise deep ranking for first-person video summarization," in Proc. IEEE CVPR, Jun. 2016, pp. 982–990.

[10]. **S. S. Thomas, S. Gupta, and K. S. Venkatesh,** "Perceptual video summarization—A new framework for video summarization,".", *IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 8, pp. 1790–1802, Aug. 2016.*

[11]. **M. Ajmal, M. H. Ashraf, M. Shakir, Y. Abbas, and F. A. Shah,** "Video summarization: Techniques and classification," .", *in Computer Vision and Graphics (Lecture Notes in Computer Science), vol. 7594. Springer, 2012, pp. 1–13. [Online].*

[12]. **S. Zhang, Y. Zhu, and A. K. Roy-Chowdhury,** "Context-aware surveillance video summarization,".", *IEEE Trans. Image Process., vol. 25, no. 11, pp. 5469–5478, Nov. 2016.*

[13]. **L. Zhang, L. Sun, W. Wang, and Y. Tian, "KaaS**: A standard framework proposal on video skimming," .", *IEEE Internet Comput., vol. 20, no. 4, pp. 54–59, Jul./Aug. 2016.*

[14]. **L. Itti, C. Koch, and E. Niebur**, "A model of saliency-based visual attention for rapid scene analysis,".", *IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.*

[15]. **Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang**, "A generic framework of user attention model and its application in video summarization,".", *IEEE Trans. Multimedia, vol. 7, no. 5, pp. 907–919, Oct. 2005.*

[16]. **Y. Pritch, A. Rav-Acha, and S. Peleg**, "Nonchronological video synopsis and indexing,".", *IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1971–1984, Nov. 2008.*

[17]. **Y. He, Z. Qu, C. Gao, and N. Sang**, "Fast online video synopsis based on potential collision graph," .", *IEEE Signal Process. Lett., vol. 24, no. 1, pp. 22–26, Jan. 2017.*

[18]. **S. Chakraborty, O. Tickoo, and R. Iyer**, "Adaptive keyframe selec- tion for video summarization,".", *in Proc. IEEE WACV, Jan. 2015, pp. 702–709.*

[19]. **J. Xu, L. Mukherjee, Y. Li, J. Warner, J. M. Rehg, and V. Singh**, "Gaze-enabled egocentric video summarization via constrained submodular maximization,".", *in Proc. IEEE CVPR, Jun. 2015, pp. 2235–2244.*

[20]. **F. Hussein, S. Awwad, and M. Piccardi**, "Joint action recognition and summarization by sub-modular inference," .", *in Proc. IEEE ICASSP, Mar. 2016, pp. 2697–2701.*

ISSN No:-2456-2165

[21]. **E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni**, "Real- time detection of traffic from Twitter stream analysis,".", *IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 2269–2283, Aug. 2015.*

[22]. **H. Zhao, C. Wang, Y. Lin, F. Guillemard, S. Geronimi, and F. Aioun**, "On-road vehicle trajectory collection and scene-based lane change analysis: Part I," .", *IEEE Trans. Intell. Transp. Syst., vol. 18, no. 1, pp. 192–205, Jan. 2017.*

[23]. **W. Yao et al.**, "On-road vehicle trajectory collection and scene-based lane change analysis: Part II,".", *IEEE Trans. Intell. Transp. Syst., vol. 18, no. 1, pp. 206–220, Jan. 2017.*

[24]. **S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur**, "A survey of vision-based traffic monitoring of road intersections," .", *IEEE Trans. Intell. Transp. Syst., vol. 17, no. 10, pp. 2681–2698, Oct. 2016.*

[25]. **S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi**, "Traffic monitoring and accident detection at intersections," .", *IEEE Trans. Intell. Transp. Syst., vol. 1, no. 2, pp. 108–118, Jun. 2000.*

[26]. **H. Veeraraghavan, O. Masoud, and N. P. Papanikolopoulos**, "Com- puter vision algorithms for intersection monitoring,".", *IEEE Trans. Intell. Transp. Syst., vol. 4, no. 2, pp. 78–89, Jun. 2003.*

[27]. **S. Atev, H. Arumugam, O. Masoud, R. Janardan**, and N. P. Papanikolopoulos, "A vision-based approach to collision prediction at traffic intersections," .", *IEEE Trans. Intell. Transp. Syst., vol. 6, no. 4, pp. 416–423, Dec. 2005.*

[28]. **Y. K. Ki and D. Y. Lee**,"A traffic accident recording and reporting model at intersections," .", *IEEE Trans. Intell. Transp. Syst., vol. 8, no. 2, pp. 188–194, Jun. 2007.*

[29]. **Ö . Aköz and M. E. Karsligil**,"Video-based traffic accident analysis at intersections using partial vehicle trajectories," ", *in Proc. IEEE ICIP, Sep. 2010, pp. 499–502.*

[30]. **K. Yun, H. Jeong, K. M. Yi, S. W. Kim, and J. Y. Choi**, "Motion interaction field for accident detection in traffic surveillance video," .", *in Proc. IEEE ICPR, Aug. 2014, pp. 3062–3067.*

[31]. **H.-S. Song, S.-N. Lu, X. Ma, Y. Yang, X.-Q. Liu, and P. Zhang**, "Vehicle behavior analysis using target motion trajectories," .", *IEEE Trans. Veh. Technol., vol. 63, no. 8, pp. 3580–3591, Oct. 2014.*

[32]. **H. T. Nguyen, S.-W. Jung, and C. S. Won**, "Order-preserving condensation of moving objects in surveillance videos,".", *IEEE Trans. Intell. Transp. Syst., vol. 17, no. 9, pp. 2408–2418, Sep. 2016.*

[33]. **S. S. Thomas, S. Gupta, and V. K. Subramanian**, "Perceptual synoptic view of pixel, object and semantic based attributes of video,".",
J. Vis. Commun. Image Represent., vol. 38, pp. 367–377, Jul. 2016.

[34]. **P. Dollár, R. Appel, S. Belongie, and P. Perona**, "Fast feature pyramids for object detection," .", *IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 8, pp. 1532–1545, Aug. 2014.*

[35]. **Ju-Won Hwang, Young-Seol Lee, and Sung-Bae Cho**, "Hierarchical Probabilistic Network-based System for Traffic Accident Detection at Intersections," .", *Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, 2010.*

[36]. **J. Du** "Understanding of Object Detection Based on CNN Family and YOLO," .", *J. Phys. Conf. Ser., pp. 1–8, 2018.*

[37]. **Sik-Ho Tsang** "Review: Faster R-CNN (Object Detection)," .", https://towardsdatascience.com/review-faster-r-cnn-object- detection- f5685cb30202. [Accessed: 10-April-2019]. *Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/review-faster-r-cnn-object- detection- f5685cb30202. [Accessed: 10-April-2019].*

[38]. ***G. Song, Y. Liu, M. Jiang, Y. Wang, J. Yan, and B. Leng,*** *"Beyond Trade-Off: Accelerate FCN-Based Face Detector with Higher Accuracy," .", IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7756– 7764.*