Automated AI Based Road Traffic Accident Alert System: YOLO Algorithm

Deeksha Gour
Computer Science & Engineering
Oriental Institute of Science and Technology
Bhopal, India
deekshagour021@gmail.com

Amit Kanskar
Assistant Professor: Dept. of Computer Science &Engg
Oriental Institute of Science and Technology
Bhopal, India
amit.kanskar@gmail.com

***Abstract*—Road Accidents, a very common reason of tragic deaths and many times the victim dies due to non-reporting of such accidents to the proper authority. Since the accident was not reported the lack of emergency medical care results in death. We live in an era of technology where we are moving towards making the city, A Smart City. A smart city with smart AI based traffic monitoring and reporting mechanism can help providing medical emergencies in real time and this would result in saving lots of life. Traditional Traffic systems are equipped with IP cameras and sensors, and are already installed in most part of the city to monitor and control traffic. These systems are capable of generating traffic tickets. In this paper we are proposing a more advanced traffic monitoring system which can identify and detect moving objects such are cars, bikes etc in live camera feeds and detect collision of these moving objects and immediately send emergency alerts to the nearby authority for them to take necessary actions.**

Index Terms*— *Vehicle detection, Deep Learning, Convolutional Neural Network,Wireless communication, Machine Learning, Python, OpenCV, YOLO

I. INTRODUCTION

Road Accidents, a very common reason of tragic deaths and many times the victim dies due to non-reporting of such accidents to the proper authority. Since the accident was not reported the lack of emergency medical care results in death. We live in an era of technology where we are moving towards making the city, A Smart City. A smart city with smart AI based traffic monitoring and reporting mechanism can help providing medical emergencies in real time and this would result in saving lots of life.

Traditional Traffic systems are equipped with IP cameras and sensors, and are already installed in most part of the city to monitor and control traffic. These systems are capable of generating traffic tickets. In this paper we are proposing a more advance traffic monitoring system which can identify and detect moving objects such are cars, bikes etc in live camera feeds and detect collision of these moving objects and immediately send emergency alerts to the nearby authority for them to take necessary actions.

Road Accidents is a very serious and high priority public health concern as the statics shows more than 1.25 million people die each year as a result of road crashes. Different risk factors such as speeding, drunk drive, No safety equipments, distracted driving, unsafe Vehicle, Law enforcement and more importantly inadequate post-crash emergency care. Any delay in detecting and providing emergency care can leads to the increase severity of the accident. With the advancement in the fields of Artificial Intelligence, Machine learning and Deep learning we are able to make our device smarter and smarter. Traffic surveillance cameras are already installed in almost every part of the city. This paper is motivated with the idea of implementing statistical method of machine learning to detect any kind of collision in a live feed with the application of convolution neural network.

The overall system can be viewed as follows:

Step 1 : Video Surveillance /Mobile Device.

These devices will be powered by supervised machine learning algorithms to identify cars, bike, humans etc. in live camera feeds and analyze collision of these objects in the feeds

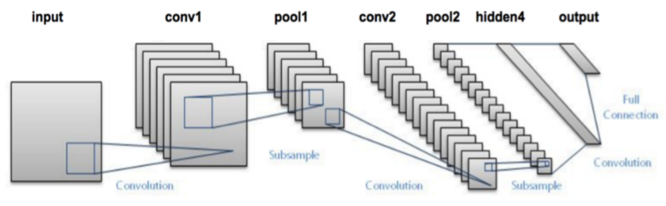
Step 2: A centralized Information/Alert delivery Unit
At the occurrence of collision detected in previous process and alert system will be activated and concerned nearby authority will be notified in real-time

Traditional traffic monitoring system in designed only to monitor traffic or to control the traffic, but it does not provide any solution to decrease the fatal accidental human damages rate which occur due to lack of medical aid in real time. Consider a scenario where an accident occurred but no one was there to report this accident, the victim is critical and every second counts, any delay can result in disability or death. We cannot root out accidents totally but we can improve in providing post accidental care just-in-time.

There are lots of sensor based systems available in the market as well but that require vehicle owners to install those sensors in their vehicles. The working of these systems is based on any damage being sensed by the sensors installed; these signals from the sensors will trigger a system that will alert nearby medical assistance or an emergency contact number. But what if the accident happened of a vehicle which is not equipped with such sensor based system. We need an advance Artificial intelligence based surveillance system which not only can detect occurrence of accident but also can alert to nearby hospitals/ambulance or Traffic policemen in real-time. Our system is based on Neural Network and Deep Learning of object detection along computer vision technology and several methods and algorithms. Our approach will work on still images, recorded-videos, real-time live videos and will detect, classify, track and compute moving object velocity and direction using convolution neural network.

III. VEHICLE AND OTHER OBJECTS IDENTIFICATION

The working method consists of six main stages. These are respectively; loading the data set, the design of the convolutional neural network, configuration of training options, training of the object detector, evaluation of trained detector. These stages and conventional and the faster R-CNN methods will be discussed in this section. A faster method in deep learning is by applying yolo: you only look once[21], This object is capable of detecting objects in real time.



The images as an input are passed to the designed network and then it is sent through various convolutions and pooling layers to form the object class.

R-CNN (Regions with Convolutional Neural Network Features)

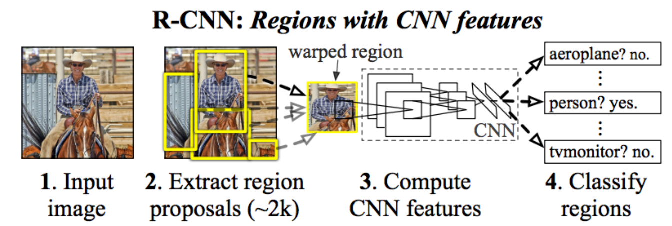
The R-CNN approach combines two basic concepts. From these, the first is to carry out efficient convolutional neural networks from bottom to up region proposals to locate and dismember objects. Next, when the label training data is insufficient, it is followed by a supervised training for a field-specific fine tuning task, which provides significant performance improvement. The method is named RCNN (CNN-enabled regions) because Regional proposals are combined with CNNs.

The working object detection system composed of three modules. Firstly, it categorically produces independent region proposals. These essentially describe the candidate detection set that can be used by the detector. The second module includes a convolutional neural network, producing an attribute vector of constant length from every region. The third module, on the other hand, includes a cluster of linear SVMs that are specific to the class used for assortment of regions [8].

Region Proposal

Various recent studies have provided methods to produce categorical independent zone recommendations. These methods have examples such as the objectness of image windows [1] , selective Search for Object Recognition [3] , category independent object proposals [4] , object segmentation using constrained parametric min-cuts [5] , Multiscale combinatorial grouping [6] and so on [7]. These methods establish cells by implementing convolution neural network with square cuts. Although R-CNN is not based on the specific zone proposal method, R-CNN performs its operations using selective search methods to provide comparison with the predetermined work.

The RCNN is carried out by applying selective search to extract regions. These selective searches are regions which form the objects. These regions are based on texture, colours, varying scales and enclosures. The selective search identifies these regions in the images. The step that is followed by, first it takes input in the form of an image, and then to have multiple regions sub-segmentations are generated from the input image, next smaller similar regions are combines to form a larger region based on texture, size and shape.



CNN (Convolutional Neural Network) for Feature extraction

To identify object with CNN for feature extraction, a feature vector of size 4096 were extracted from each region proposal with Caffe deep learning framework. Features were extracted by forwarding the average output 227x227 red-green-blue image with five convolution layers and two completely connected layers.

To calculate an attribute in a region proposal, the image data is first converted to a form compatible with CNN. Then, the most simple of the possible transformations of the random-shaped regions was selected. Here, all the pixels in a tight bounding box around the candidate area are resolved unto the required size, regardless of the size or aspect ratio. Before dissolving, the tight bounding box was expanded to provide w pixels skewed picture content around the box at the skewed dimension (w = 16 was used). In addition, a simple bounding box regression was used to expand the localization performance within the application [13]. This is shown in the following equation (1). The details of this equation can be seen in [8].

$$w = argmin\ wx \sum(t \quad i - w \quad T \quad 5(P\ \dot{I}\ ))\ 2 + \lambda\|W \quad \|\ 2\ N$$
$i(1)$

Classify Regions

Selective search is performed on test images to obtain region proposals. Each proposal are resolved and advanced through CNN for the calculation of attributes. Then, for each class, each produced attribute vector is scored using the trained support vector machine (SVM). Considering the scored regions, greedy non-maximum suppression is applied independently when there is high-intersection (IoU) overlap with the selected zone with a higher rating over a learned threshold for a rejected region.

R-CNN Training

Supervised Pre-training CNN was previously trained on a large auxiliary data set using only image-level additional tags. CNN was previously trained on data set (ImageNET ILSVRC2012 [9]) using only additional tags. The training was carried out using Caffe Deep Learning framework.

Domain-Specific Fine-Tuning

In order to adjust Convolutional Neural Network to new task and domain name , SGD training was performed to function parameters using only warped region proposals. Convolutional Neural Networks ImageNetspecific 1000 way classification layer has been changed over with the N+1 way classification layer. Convolutional Neural Network framework has not been changed here. (N = 20 for VOC and N = 200 for ILSVRC2013).

All region proposals, which are equal to or greater than 0.5 iou overlap value, were accepted as positive for the box class and others were accepted as negative. In each SGD iteration, 32 positive windows and 96 background windows are properly sampled to create a mini stack of 128 sizes.

Object Category Classifiers

Here, binary classifier training was used to perceive cars. It is a positive example of an image area in which a car is tightly enclosed. In a similar way, a background region that is not interested in cars is a negative example. It is unclear how a partially overlapping region of the car should be labeled. the unclear state is solved by specifying an IoU overlap threshold value. Areas below this threshold value are identified as negative and those above the threshold value as positive.The overlap threshold "0.3" was chosen by conducting a grid search on the verification set. Once the features are removed and the training tags are applied, SVM is applied optimally to all classes.

Faster R-CNN

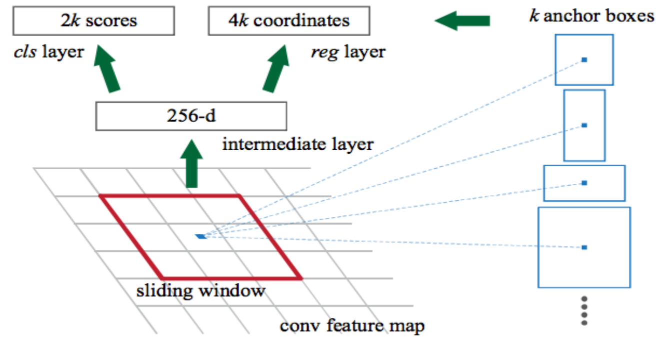
The Faster R-CNN composed of two component. The first component is a conventional network used to propose zones called RPN, and The second component is the Faster R-CNN detector which utilize the region proposals. The whole system comes from a single composite network created for object detection [10].

The first component is a conventional network used to propose regions called RPN (Region Proposal Network)

Region Proposal Networks (RPN)

RPN receives as input image and produces a set of rectangle object tender which all have objectivity score. The RPN is designed with a fully convoluted network. Since calculations are shared with a Faster R-CNN object detection network, it is assumed that both networks share a common set of layers of convolution.

A mini network is moving on the exit of the convolution property map by the last shared convolution layer to produce region proposals. As an input, it takes the space window of the convolutional property map n × n(used as n=3). All sliding windows in work are matched to low dimensional property. This feature composed of two sister fully bound boxregression and box-classification layers. In this mini network, all the fully connected layers are shared in all spatial locations.This framewoek is carried out by the convolution Layer and following the two brothers 1 x 1 convolution layer.



Training RPNs (Region Proposal Network)

RPNs are trained end-to-end with back propagation and SGD. In order to train this network, "image-centric sampling" strategy is applied. All the batches come from the images involving negative and positive sample anchors.

When the missing functions of all the anchors are optimized here, the orientation of the negative examples is realized. For this reason, a random sample of 256 anchors is shown in an image instead. According to this, if there are more than 128 positive samples, it is filled with stacked samples. Otherwise, it is filled with negative examples. In addition, following the multitasking loss for Fast R-CNN, the objective function has been reduced to a minimum. This loss function is shown in the following equation (2) , the details of the Equation also can be seen in [10].

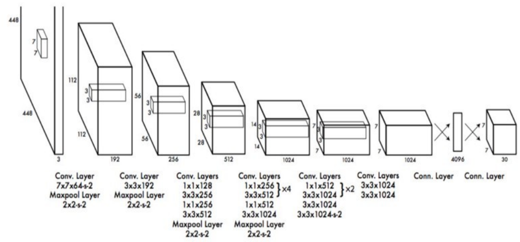
L({pi },{ti }) 1 Ncls ∑L cls(Pİ ,Pi)+λ 1 Nreg ∑ pi L reg(ti ,t i) i i (2)

Comparison of Faster R-CNN and R-CNN Methods

Today, the most sophisticated object detection networks are based on region proposal algorithms for the description and identification of object locations. Faster R-CNN has put forward the regional proposal calculation as a bottleneck by reducing the working time of these detection networks in R-CNN. In the Faster R-CNN, a Region Proposition Network (RPN) is implemented that shares full image convolution characteristics using the detection network, so that almost free region proposals can be made. Faster R-CNN, along with the improved RPN, do not require external zone recommendations, unlike R-CNN. In addition, the RPN improves the quality of the district proposal and thus improves the overall accuracy and speed of object detection.

YOLO: you only look once

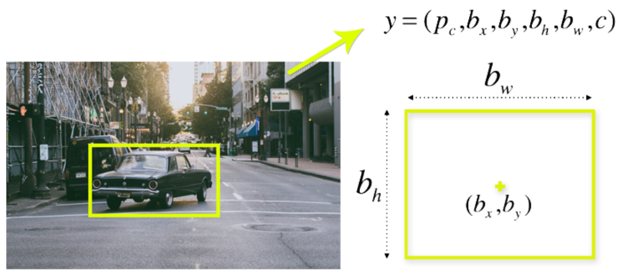
Yolo algorithm applies a neural network to an entire image.The image is divided in SxS grid and comes up with bounding box[21].This algorithm has 24 convolutional layers followed by 2 fully connected layers. Alternating 1x1 convolutional layers reduce the features space from preceding layers. The object identification problem is considered to be a regression problem to spatially separated bounding boxes and associated classes probability. A single neural network can predict the bounding boxes and class probabilities directly from full images in one evaluation which can be optimized end-to-end.



In this study, we are going to apply yolo algorithm for detection of objects through a live feed or an image. The working of yolo is quite simple as yolo is based on regression. Unlike CNN which selects interesting parts in an image, yolo on the other hand predicts the class and bounding boxes for the whole image in one run of the algorithm. To apply this algorithm we need to know what we are going to predict i.e. the objects we are interested in so that we can train our algorithm to look for classes of the objects and the bounding box specifying the object location. The bounding box are described using these four descriptions

- Center of bounding box (b_x, b_y)
- Width (b_w)
- Height (b_h)
- C: class of the object which is identified

Pc is the probability of objects in the bounding box.



IV. LITERATURE REVIEW

*14) **BONGJIN OH, JUNHYEOK LEE PROPOSED***

Two Convolution Neural Network (CNN) can be ensemble to train and recognize or extract scene images and different objects in the images can be identified and stored according to the scene classes. This hybrid CNN outperforms the Places365-ResNet for both top -5 accuracy by 3%.

15) *SHRISTI SONAL AND SAUMYA SUMAN PROPOSED*

Data mining and machine learning techniques were applied on the road traffic data and is analyzed for finding out the key factors for the severity of an accident. Although the characterization of humanity and behavior is an important factor in occurrence of accidents but the spatial feature and infrastructure plays a contributing role in the accident.

16) *A . KRIZHEVSKY, I. SUTSKEVER, AND G. HINTON PROPOSED*

The neural network which has 60 millions parameters and 650,000 neurons consists of five convolutional layers, some of which are followed by max-pooling layers and three fully connected layers. By using non-saturating neurons and a very efficient and powerful GPU-implementation training can be made faster. Regularization method "dropout", were employed to reduce overfitting in the fully-connected layer.

17) *LESYA ANISHCHENKO PROPOSED*

That deep learning and transfers learning techniques can be applied in fall detection by means of surveillance camera data processing. The Architecture of CNN AlexNet which used as a starting point classifier was adopted to detect falling person problem. The cohen's kappa of .93 and .60 was achieved for fall and non-fall respectively for known and unknown classifier surrounding conditions.

18) *CAROLINE ROUGIER, JEAN MEUNIER, ALAN ST-ARNAUD AND JACQUELINE ROUSSEAU PROPOSED*

A computer vision system which can analyze people behaviour and detect unusual events, the approach of this system [18] was based on the motion history and human shape variations. The whole idea of the system was to detect large motion of the person on the video sequence using motion history image and then when a motion is detected shape of the human is then analyzed. Change in human shape is discriminated as normal when person sits or walks and abnormal when person falls.

19) *LIAN PENG, YIMIN YANG,XIAOJUN QI AND HAOHONG WANG PROPOSED*

That a hint information based object identification can be made to improve the object identification accuracy of the conventional object identification system. In this paper [7] they formulate a novel cost function to ensure good local representation and good content variation coverage of candidate key frames. Dynamic programming was applied on this cost function to extract key frames from video input to summarize the whole video. The object in the key frames was recognized using the learned model on the conventional knowledge database (i.e. training images) and use these labelled recognized objects as hint information to refine knowledge database. The better the representativeness of hint information the variation between testing and training images will be significantly better and thus improves the object recognizing performance.

V. CONCLUSION

In this study, the proposed accident detection system can be trained by using regression based algorithm YOLO: you only look once algorithm on the sample vehicle datasets and the vehicle detection process has been successfully performed by the trained vehicle detector being tested on the test data set with the live video feeds from the webcam. The proposed system is faster than other object detection methods and predicts the object better other object detection algorithm such as Faster-CNN or Fast CNN. The input can also be optimized and give better results. Further the system alerts via a wireless communication devices to nearby emergency vehicles.

VI. REFERENCES

[1] B. Alexe, T. Deselaers, V. Ferrari, "Measuring the objectness of image windows", TPAMI, 2012.

[2] Guzel, MS, "Versatile Vehicle Tracking and Counting Application",KaraElmas Science and Eng Journal,7(2),622-626,2017

[3] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M.Smeulders, "Selective Search for Object Recognition,"International Journal of Computer Vision, Cilt. 104, s. 154–171,2013.

[4] I. Endres, D. Hoiem, "Category independent object proposals", ECCV, 2010.

[5] J. Carreira, C. Sminchisescu, "CPMC: Au-tomatic object segmentation using constrained parametric min-cuts", IEEE Transactions on Pattern Analysis and Machine Intelligence, Cilt.34, s. 1312–1328, 2012.

[6] P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping", CVPR, 2014.

[7] D. Cires ̧an, A. Giusti, L. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks", MICCAI, 2013

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, " Rich feature hierarchies for accurate object detection and semantic segmentation.", IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014.

[9] ImageNET Classes Date Set Avaliable at: http://imagenet.org/

[10] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN:Towards real-time object detection with region proposal networks, NIPS, 2015.

[11] Vehicle Detection Data Set, Matlab Official Web Site Avaliable at: “https://www.mathworks. com/”, 2017.

[12] Standford Vehicle Data Set:Avaliable at: http://ai.stanford.edu/~jkrause/cars/car_dataset.Html, 2018.

[13] J. Donahue, Transferrable Represenations for Visual Recognition , PhD Thesis, University of California, Berkeley,2017,

[14] Bongjin Oh, Junhyeok Lee, A case study on scene recognition using an ensemble convolution neutral network, in 2018 20th International Conference on Advance Communication Technology (ICACT), 2018.

[15] Shristi Sonal and Saumya Suman, A Framework for Analysis Of Road Accidents, 2018 International Conference of Emerging Trends And Innovations in Engineering And Technological Research(ICETIETR).

[16] A . Krizhevsky, I. Sutskever, and G. Hinton, ImageNet Classification with Deep Convolution Neural Networks, in Advances in Neural information Processing Systems 22,pp.1106-1114,2012

[17] Lesya Anishchenko,Machine Learning in Video Surveillance for Fall Detection in Ural Symposium of Biomedical Engineering, Radio electronics and Information Technology(USBEREIT)

[18] Fall Detection from human shape and Motion History using video surveillance, in 21st International Conference on Advance Information Networking and Application Workshops (AINAW’07),2007.

[19] Lian Peng, Yimin Yang,Xiaojun Qi and Haohong Wang, Highly accurate video object identification utilizing hint information, in 2014 International Conference on Computing Networking and Communications (ICNC).

[20] P.A. Dhulekar, S.T. Gandhe, Anjali Shewale, Sayali Sonawane, Varsha Yelmame, Motion Estimation for human Activity Surveillance, in 2017 International Conference of Emerging Trends and Innovation in ICT(ICEI)

[21] Joseph Redmon , Santosh Divvala , Ross Girshich, Ali Farhadi, University of Washington, You only look once: Unified Real-time Object Detection,2016

[22] Guanqing Li, Zhiyong Song, Qiang Fu, A New Method Of Object Detection For Small Datasets Under The Framework of YOLO Network, 2018 IEEEE 3rd Advane Information Technology, Electronic and Automation Conference (IAEAC 2018)