Automated AI Based Road Traffic Accident Alert System: YOLO Algorithm

Deeksha Gour; Amit Kanskar

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 able to generate traffic tickets automatically. 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

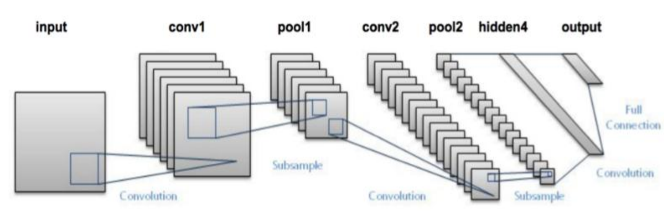
Road Accidents is a very serious and high priority public health concern as the statics shows more than 1.25 million[23] people die each year as a result of road crashes. Different risk factors such as Speeding, Drunk drive, No safety equipments, Distracted driving, Unsafe Vehicle, Law enforcement and 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.

II. PROBLEM STATEMENT

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 set of data, followed by designing convolutional neural network followed by configuring training options followed by training the underlying object detector model which is followed by evaluation of the detector based on the input so as to improve the performance. in this section we will be discussing above mentioned stages, along with conventional and faster R-CNN methods. 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.



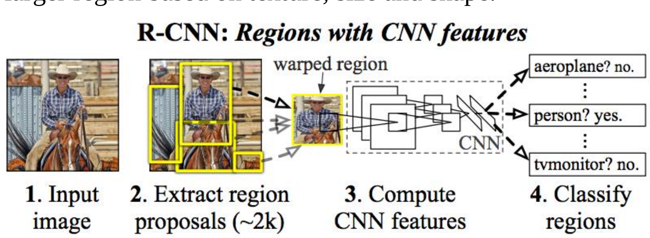
The images are passed as input into the designed network which goes through various convolution layers and pooling layers to which results in the formation of object class..

R-CNN (Regions with Convolutional Neural Network Features)

Two basic concepts are combined and applied in the R-CNN. The first concept is to apply an efficient convolutional neural networks from bottom to up region proposals to locate and dismember objects. The second concept is to apply supervised training for field-specific tuning task when insufficient training images in entered into the system, resulting in significant improvement of performance. The method is named RCNN (CNN-enabled regions) because Regional proposals are combined with CNNs. The whole object detection system can be combined in three maine modules. Firstly, it produces region proposals which are independent,categorically. These categorical regions defines the candidate detection set which is used by our detector. The second module includes a convolutional neural network, in this modules we produce an attribute vector which are of constant length, from every region proposals. The final module, includes a cluster of support vector machine of linear nature which are specific to the class used for evaluation and 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 image window's objectness [1] , Recognition of object by selective Search [3] , object proposals based on category independence [4] , segmentation of object by using min-cut parametric constraints [5] , grouping of Multiscale combinatorial [6] and so on. The above methods implements convolution neural network with square cuts resulting in establishment of cells. Although R-CNN is not based on the specific zone proposal method, selective search methods are used for applying R-CNN to provide comparison with the predetermined work. The RCNN is used to extract region by applying selective search. These extracted regions extracted by selective search in turn forms 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 is, 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. For the calculation of an attribute in a region proposal the input image is converted to a CNN compatible form. Then, we select the simplest transformation of random-shaped region. In this random-shaped region,the pixels in a bounding box are resolved into the required size around the candidate area, without being concerned of the aspect ratio or size. 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,to expand the localization performance simple bounding box regression is used within the application.[13]. This is shown in the following equation (1). The details of this equation can be seen in [8].

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

Classify Regions

Selective search is performed on test images to obtain region proposals. Each region proposal are resolved and passed through Convolutional Neural Network for the attribute calculation. Then using a trained Support Vector Machine, for every class, these attribute vectors are evaluated and scored. Based on the scored regions, a non-maximum greedy suppression is applied independently when there is high-intersection 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 set of auxiliary data by using image-level additional tags. CNN was previously trained on data set (ImageNET ILSVRC2012 [9]) by using additional tags. Caffe Deep Learning framework was used to carry out this training.

Domain-Specific Fine-Tuning

In order to use CNN to a different domain and perform different task, SGD training was performed by using region proposal which were warped, to the function parameters. Convolutional Neural Networks ImageNet specific 1000 way classification layer has been modified over with the N+1 way classification layer. 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

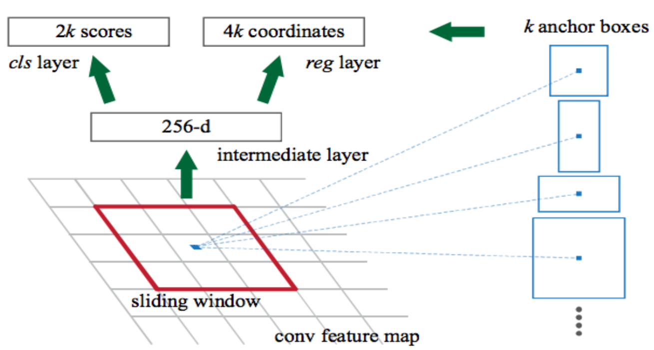
Here, binary classifier training was used to perceive cars. It is a positive example of an image area in which a car is tightly enclosed. In a similar way, a background region that is not interested in cars is a negative example. Areas below this threshold value are identified as negative and those above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

Faster R-CNN

The Faster R-CNN is composed of two main component. The first component used to propose zones viz Regional Proposal Networks is a conventional network. The second component is a detector which utilize these region proposals. This detector is a Faster R-CNN detector. The complete object detection system comes from a single composite network used for detection of object. 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 box regression and box-classification layers. In this mini network, all the spatial locations are shared with all the fully connected layers.This framework 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. These networks are trained by applying "image-centric sampling" strategies. All the batches of the images come from the images involving negative anchor and image involving anchors which are positive. The negative sample anchors are identified and realized by applying a missing function which in turn optimize the anchors. For this reason, a random sample of 256 anchors is shown in an image instead. According to this, if there are less than 128 positive samples, it is filled with negative examples. or else , it is filled with positive examples.This also results in the objective function to be reduced to the lowest by following a multitask loss in case of Fast R-CNN. This loss function is shown in the following equation (2) , the details of the Equation also can be seen in [10].

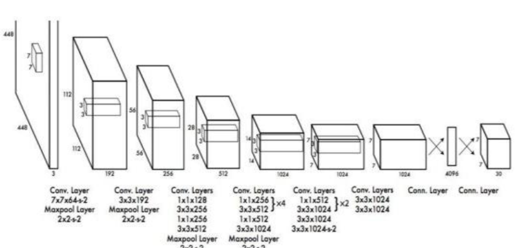
$$L(\{pi\},\{ti\})\ 1\ Ncls\ \sum L\ cls(P\dot{I}, Pi*) + \lambda\ 1\ Nreg\ \sum\ pi\ L\ reg(ti, t\ i*)\ i\ i\ (2)$$

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, accuracy and speeds up the detection of objects.

YOLO: you only look once

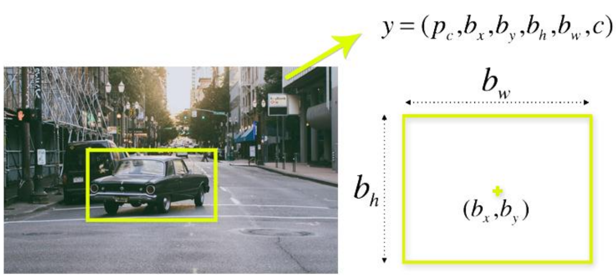
Yolo algorithm achieves its result by applying a neural network on an image..The image is divided in SxS grid and comes up with bounding box[21].This algorithm has 24 convolutional layers which in turn has two fully connected layers. The reduction in feature space is done by Alternating 1x1 convolutional layers from preceding layers. The object identification problem is considered to be a regression problem with the objective of spatially bounding box separation along with the probability of associated classes in the bounding boxes. A single neural network can predict the bounded boxes and class probabilities directly from the input images in just 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 likely to be 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 name of the identified object

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 and intensity 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 convolutional layers, max-pooling layers and three fully connected layers. There are five convolutional layer some of them are followed by other two layers.By using a very efficient and powerful GPU-implementation and non-saturating neurons, 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 the detection of fall which was captured by surveillance camera data processing. The Architecture of CNN AlexNet which used as a initiating 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 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] a cost function was formulated which ensured a good representation and content variation locally of key candidate frames. To extract key frames from the input video relevant dynamic algorithms were applied programmatically on the cost function. The object in the key frames was recognized using the trained model on the existing database (i.e. training images) and use these labelled recognized objects to refine knowledge database. The better the representativeness of hint information the variation between testing and training images will be significantly better and thereby it improves the object recognizing performance.

V. CONCLUSION

In this study, the proposed accident detection system can be trained by using regression based algorithm called YOLO(you only look once) algorithm on the sample vehicle datasets and the vehicle detection process has been successfully performed by the trained model 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: Automatic 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)

[23] https://www.asirt.org/safe-travel/road-safety-facts/