Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

Optimized-YOLO: Algorithm for CPU to Detect Road Traffic Accident and Alert System

Deeksha Gour
Computer Science & Engineering
Oriental Institute of Science and Technology
Bhopal, India

Amit Kanskar
Assistant Professor: Dept. of Computer Science & Engg
Oriental Institute of Science and Technology
Bhopal, India

***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. These systems are able to generate traffic tickets automatically. In this paper we are proposing an artificial intelligence based traffic monitoring system which can detect the occurrence of accidents of vehicles such as 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. This paper focuses on an optimized-Yolo algorithm which is capable of detecting accidents in real time, also can run on central processing unit based devices such as laptops or mobile phones. Laptops and mobile phones are not generally equipped with large graphical processing units. The model is trained on custom dataset achieving a mean average precision(mAP) of of 33.31%. Optimized-yolo is designed for creating smaller and faster detection models apart from its original Yolo V3.**

Index Terms*— Vehicle detection, Deep Learning, Convolutional Neural Network,Wireless communication, Machine Learning, Python, OpenCV, Optimised YOLO, Darknet. CPU Based object detection.

I.INTRODUCTION

Road Accidents is a very serious and high priority public health concern as statistics shows more than 1.25 million[23] people die each year as a result of road crashes. Different risk factors such as Speeding, Drunk driving, 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 lead to the increased severity of the accident. With the advancement in the fields of Artificial Intelligence, Machine learning and Deep learning we are able to make our devices 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.

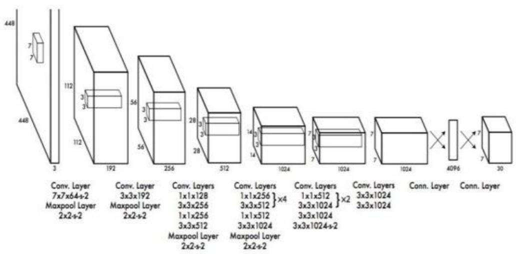
II. PROBLEM STATEMENT

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 damage 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 in 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 the occurrence of an 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. OPTIMIZED YOLO ARCHITECTURE AND IMPLEMENTATION

Optimized 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.

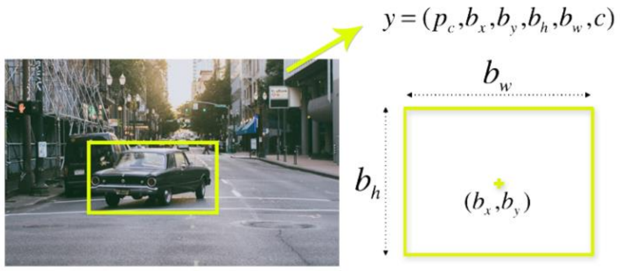


International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

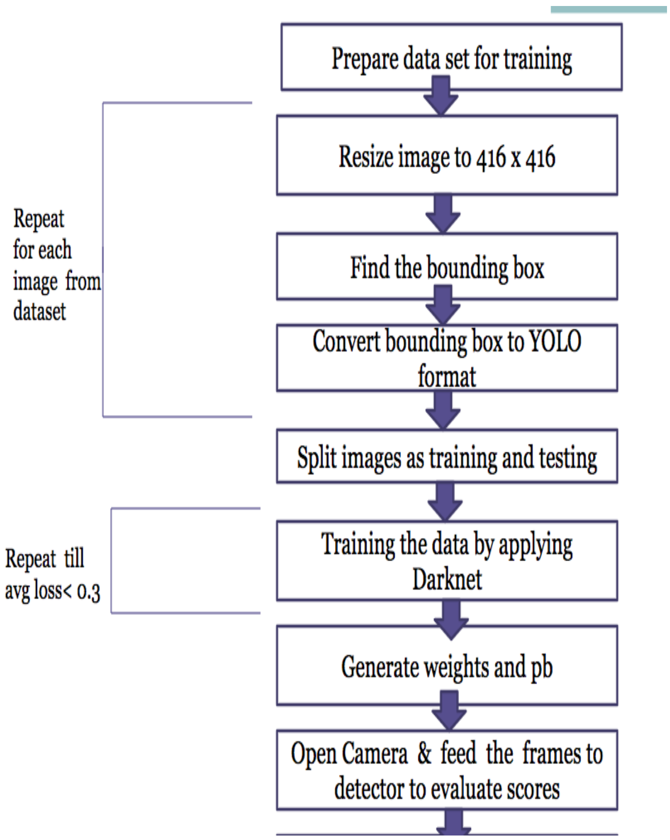
In this study, we are going to apply optimised yolo algorithm for detection of objects through a live feed or an image. The working of this optimised yolo is very simple as yolo is based on regression. Unlike CNN which selects interesting parts in an image, yolo on the other hand predicts the class and 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.



FLOWCHART :



a. ENVIRONMENT SETUP

Darknet framework was used to train and test model. The training was carried on i5 processor with clock speed of 2.5Hz. Testing was carried out by applying tensorflow and models generated by training custom dataset.

b. DATASET

Optimized Yolo is trained on custom dataset. This dataset consists of images of car accidents. Table I shows the details of the custom dataset used for training.

TABLE I
CUSTOM DATASET

Dataset	**Number of Images**	**Number of classes**
custom dataset (car accident)	500	1

c. IMAGE SIZE

Each image was filtered properly and were resized to 416 x 416 px. By resizing the image to this dimensions overheads of resizing done by darknet is reduced there by performance is improved.

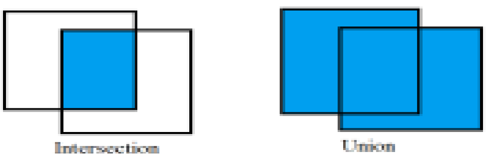
d. KEY INDICATOR

The Confidence/Model scores generated by applying a real time feed to check for accident detection by means of pb models and tensorflow, if the score reaches a threshold of 0.7 or greater then accident is detected and sms is sent.

Another key indicator is the mAP value of the optimized yolo. Mean Average Precision is the mean of average precision of the class. In this paper, only one class is to be detected so mean average precision is also the average precision(AP).

AP= TP/(TP+FP)

where ,
AP: Average Precision
TP : model correctly predicts the positive class.
TN: model correctly predicts the negative class.
FP: model incorrectly predicts the positive class.
Intersection over union (IOU), is used to identify true positive or false positive. IOU is calculated by using two bounding box, Prediction box and Actual annotated ground truth on the same image. By calculating % overlap of these boxes by dividing the intersection area by the union area. Based on this IOU ratio, a threshold is chosen on this IoU to classify object detection as a True Positives or false positive.



Optimized-Yolo used 24 layers of network.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

TABLE II
OPTIMIZED-YOLO ARCHITECTURE

```
layer     filters    size
    0 conv     16  3 x 3 / 1
    1 max          2 x 2 / 2
    2 conv     32  3 x 3 / 1
    3 max          2 x 2 / 2
    4 conv     64  3 x 3 / 1
    5 max          2 x 2 / 2
    6 conv    128  3 x 3 / 1
    7 max          2 x 2 / 2
    8 conv    256  3 x 3 / 1
    9 max          2 x 2 / 2
   10 conv    512  3 x 3 / 1
   11 max          2 x 2 / 1
   12 conv   1024  3 x 3 / 1
   13 conv    256  1 x 1 / 1
   14 conv    512  3 x 3 / 1
   15 conv     18  1 x 1 / 1
   16 yolo
   17 route  13
   18 conv    128  1 x 1 / 1
   19 upsample           2x
   20 route  19 8
   21 conv    256  3 x 3 / 1
   22 conv     18  1 x 1 / 1
   23 yolo
```

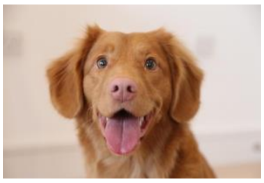
ALGORITHM TERM

x1,y1 : coordinate of left corner of object in concern within the image
x2,y2 : coordinate of bottom right corner of object in concern within the image
<object-class> : integer number of object from 0 to (classes-1)
<x> <y> <width> <height> - float values relative to width and height of image, it can be equal from (0.0 to 1.0]
<x> = <absolute_x> / <image_width> or <height> = <absolute_height> / <image_height>
<x> <y> - are center of rectangle (are not top-left corner)

ALGORITHM

```
for each image do
   resize image to 416 x 416
  generate box labels ( x1,y1,x2,y2) and store in a file
  convert generated labels into yolo format and    store in a
file(<object-class> <x> <y> <width> <height>)
end for
for each batch of 64 images with subdivision of 8 do
    train detector to generate weights
     stop training :
       if( avg loss <0.3)
end for
open camera
for each camera frame do
     Read the camera frame
     convert to byte array
     feed array to the classifier to generate scores
     if(score <= 0.70)
       alertSms(Number One , Number Two , Location)
       press q to stop
     endfor
```

SAMPLE RUN

Input	Output	Accident %
		91%
		97%
		0%

IV. RELATED WORK

*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 the 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 Optimized-YOLO algorithm which can be applied on CPU based devices. In this paper optimised-yolo algorithm is trained on custom datasets of car accident images with the mAP of 33.31% 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 feed 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. FUTURE SCOPE

The proposed system can also be used to detect the severity of the accident, possibly can detect the number plate and if connected to centralized system can also be used to inform the emergency contact associated with the number plate or the insurance agencies.

VII. 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. Cireşan, 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/