Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021:8453 – 8460

Research Article

Speeding Vehicle Detection in Unconstrained Environment

T Prathima[a], A Govardhan[b], Y Ramadevi[c], S Murari[d]
[a,c,d] Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[b] Jawaharlal Technological University, Hyderabad, Telangana, India

Abstract

In this paper a method is proposed to identify speeding vehicles in two phases, in the initial phase using YOLOV3 vehicles are detected, and in the later phase blur detection is used to identify whether the detected vehicle is speeding. We are able to achieve commendable results and accuracy of our results are directly dependent on the accuracy of YOLO object detection. To detect whether the vehicle is speeding Laplacian kernel is convolved against the gray intensity image and blur variance is computed. If the blur variance is in the order of 10^{-4} and it is further decreasing across the neighbouring frames we can identify that the vehicle as speeding vehicle.

Keywords: *Object Detection, YOLO, RCNN, Blur Detection, Speeding Vehicle*

Introduction

Object detection is a crucial step in many computer vision tasks. It helps in recognising gestures, identifying vehicle type, analysis of surveillance videos etc. In OD objects are identified by inserting a bounding box round the object in the entire image, and there may be several bounding boxes drawn to recognise different objects based on location of object with in the image and aspect ratio. After the bounding boxes are detected the regions will be given to classifier and the object is labelled. The landmark contributions in the evolution of Object Detection (OD) can be thought of as OD based on traditional approaches earlier to year 2012 and contributions after 2021 which are mainly using deep learning (DL) models, with the arrival of Alexnet proposed in [1] and R-CNN in the year 2014 in [2, 3]. OD based on deep learning methods since then has transformed the research orientation. Fast R - CNN [4], Faster R - CNN [5], YOLO [6, 7, 8], SSD [9] are the famous works reported. These DL models can further be classified into 2-stage or 1-stage models [10]. Two stage algorithms basically localises the objects within the image by inserting bounding boxes in the first phase and later labels the objects in the bounding boxes in second phase. Where as in one stage OD algorithms, bounding box prediction and the class probability of the object are carried in single stage.

In this work architecture is proposed to detect two objects: i. speeding vehicles, bikes and ii. glasses and wine bottles. The purpose of taking up this task is to insert statutory warning in the segments of the video where speeding vehicles are found as “over speeding is dangerous” and frames which has glasses and bottles as “alcohol consumption is harmful for health”. These statutory warnings must be inserted in the videos broadcasted for public viewing on television as per the directives of Information and Broadcasting ministry, Govt. of India. The task of automating the process of identifying speeding vehicle is done in two phases, in the first phase vehicle detection is carried out using YOLO and in the second phase the speeding vehicle is detected based on the change in percentage of blur. This paper is organised as follows: evolution of OD algorithms is presented in Section 2, Section 3 details the existing work on detecting speed of the vehicles and blur detection, dataset description is provided in Section 4, results are provided in Section 5, Conclusions and future scope are discussed in Section 6.

Evolution of Object Detection Algorithms

For object detection, the location of the object must be identified within the image, the task of localising the objects in an image can be framed as regression problem. Localising an object using regression approach will work for one object, but localising multiple objects is a complex task and it requires the no of objects available in the image to be estimated prior. An alternative approach is detection based on sliding-window which is in practice for atleast twenty years to detect constrained objects such as hands, faces, pedestrians, etc. the above approach works well when all the objects share a same aspect ratio. To avoid selecting large no. of regions selective search [11] approach is used in RCNN to generate 2000 such regions and these regions are called the region proposals. RCNN during testing generates two thousand region proposals which are class independent for every input image. A fixed length feature vector for every proposed region is generated using a CNN and then regions are classified with linear Support Vector Machines trained to identify specific category of objects. Warping technique uses anisotropic scaling is used to generate a fixed size CNN input from each region proposed, irrespective of the shape of the region. Non uniform scaling means that different scales are applied to every dimension, marking it as anisotropic. On the other hand opposite of anisotropic is isotropic scaling, where same scale is applied for all dimensions. In RCNN, the authors resized any image into fixed size, irrespective of aspect ratio of the image. For example an image with 1280 x 720 size is resized to 224 x 244, then scales are 1280 / 224 and 720 / 224, which results in anisotropic scaling.

Selective Search proceeds as follows: i. Sub-segmentation is initially done and many candidate regions are generated ii. Using greedy algorithm similar regions are recursively combined into bigger regions and lastly iii. From the regions generated candidate region proposals are finalised

The two thousand region proposals are warped as a square and are inputted to CNN that gives a feature vector of 4096 dimensions. The CNN extracts the features and are fed into Support Vector Machine to label the objects present within candidate region proposals. To adjust boundaries of the box which bounds the region proposals, an offset value is given which helps in adjusting the boundaries in the process of predicting the presence of an object within the region proposal. Classifying two thousand region proposals for every image is time consuming and so this approach cannot be applied real time. Selective search algorithm which does the task of generating region proposals is a fixed algorithm and learning is not happening at the algorithm level and this may result in probable region proposals which are not good.

Fast R-CNN

The drawbacks of R-CNN to some extent are addressed in fast RCNN. Initially image is given as input to CNN instead of the region proposals and CNN outputs convolutional feature map, inturn from these feature maps region proposals are identified. Identified regions are warped into squares. Region of Interest Pooling layer will reshape these squares into fixed size and these are given as input to full connected layer. From RoI feature vector the class of the proposed regions and offset are predicted by using the softmax layer. Fast R-CNN is relatively faster than RCNN as the convolution operation is carried out only once for input image and the need to input two thousand region proposals is there by avoided. The process of generating region proposals slows down the process in both RCNN and fast RCNN as the algorithm used to generate region proposals is selective search and selective search consumes more time. Faster RCNN is proposed to eliminate the need for selective search algorithm and a Region Proposal Network (RPN) learns region proposals from the input image instead of selective search.

In the OD algorithms discussed so far object is localised within the input image and the neural net doesn't looks at the full image and only looks in parts. In You Only Look Once (YOLO) one single CNN does the task of identifying bounding boxes and also estimates the class probabilities for these boxes. Input image is split into S x S grid, within each such grid 'm' boxes are considered, and for every such bounding box the net gives a class probability and offset values. Class probabilities that are above threshold are selected and are used for detecting the object in the input image. When compared to other OD algorithms YOLO is faster. One observed limitation of YOLO is it fails to detect small objects in the input image.

Table 1: Performance comparison of OD algorithms on COCO dataset [7]

S.No	No. of Stages	Method	Back Bone	AP_{75}	AP_{50}	AP
a.	Two stage methods	Faster RCNN +++	ResNet-101-C4	37.4	55.7	34.9
b.		Faster RCNN w FPN	ResNet-101-FPN	39.0	59.1	36.2
c.		Faster RCNN G-RMI	Inception – Res Net - V2	36.7	55.5	34.7
d.		Faster RCNN w TDM	Inception - Res Net - v2 -TDM	39.2	57.7	36.8
e.	One Stage methods	YOLO V2	DarkNet-19	19.2	44.0	21.6
f.		SSD 513	ResNet – 101 - SSD	33.3	50.4	31.2
g.		DSSD 513	ResNet – 101 - DSSD	35.2	53.3	33.2
h.		Retina Net	ResNet – 101 - FPN	42.3	59.1	39.1
i.		Retina Net	ResNeXt – 101 - FPN	44.1	61.1	40.8
j.		ROLOv3 608 x 608	Darknet-53	34.4	57.9	33.0

Motivation

In India we have CBFC [12] Central Board of Film Certification is a legal body under MIB, Ministry of Information & Broadcasting, Government of India. Any film has to be certified by CBFC before releasing it to the public under the Cinematograph Act of 1952.

Following are the objectives of Film certification as per CBFC's website:

a. the medium of film remains responsible and sensitive to the values and standards of society
b. artistic expression and creative freedom are not unduly curbed
c. certification is responsible to social changes
d. the medium of film provides clean and healthy entertainment and
e. as far as possible, the film is of aesthetic value and cinematically of a good standard.

To ensure the above, CBFC insists that scenes in videos which contain smoking, alcohol or drugs, speeding vehicles are to be warned through a message inserted on the frames of the video. We are motivated from the objectives of CBFC and directives by I&B ministry, we proposed to automate the process of detecting speeding vehicles.

Blur Detection

To estimate the speed of vehicle, initially it is planned using subtraction of background and detection of Blobs technique, as speed estimation carried out using blob displacement [13, 14,15,16], or in [17, 18] word is dependent on relative motion between static camera and on object that is moving. This approach cannot be used for the video clips that are tested against in the current work, as these snippets are clips from Tollywood and Hollywood movies. As it is common to shoot movies making use of multiple cameras and camera tricks, and usage of these cinematographic techniques results in giving the audience feel that the objects are moving at a faster pace but in reality objects are moving slowly. To counter these visual tricks and to get results blur detection is used.

The reasons for a blurry frame or image may be due to camera shake, due to motion blur or it may also be due to improper focus of lens of the camera. Motion blur is because of the non-sync between shutter speed of the capturing device and the speed of the moving object. To detect blur in the current work it is sufficient to detect motion blur, but as discussed earlier to tackle cinematography tricks motion blur alone will not suffice. To highlight the scene and to grab the attention of the spectator either the background is blurred keeping the object of interest sharp or vice versa. Irrespective of the reason for blur, to identify whether an image or blurry or not especially to handle the speeding vehicle detection scenario the works [19, 20, 21] are referred. Empirically for the chosen dataset i.e., movie clips from both Hollywood and Tollywood Laplacian kernel approach gave good results. Firstly, frames are converted into grayscale, in the second step laplacian kernel is convolved on the gray images. Variance metric is computed from the convolution. Sharp images have high variance and blurred images have low variance. A 3 x 3 Laplacian kernel is used and it is given as [[0, 1, 0], [1, -4, 1] , [0, 1, 0]]. It is observed that the computed variances for images that are blurry are found to be in the order of 10^{-4}.

Datasets

Datasets used for training YOLO are taken from Open images [22]. There are around six versions of datasets publicly available in the dataset with around nine million annotated images with labels, bounding boxes, segmentation masks, relationships between objects and localized narrative. The bounding boxes have been mainly drawn manually for consistency and accuracy. There sixteen million bounding boxes for six hundred different categories of objects, making it the largest among all the available existing datasets. The images also contain numerous objects and lot of complex scenes. Annotations for visual relationships are also made available indicating objects with their relationship. The dataset has 59.9 Million image-level labels across 19,957 classes. The image categories considered for our experimentation are cars, bikes, glasses, bottles and cigarettes. The images from the categories are downloaded with the help of [23].

Total images considered for training and testing YOLO is around 8000 images across four classes of images.

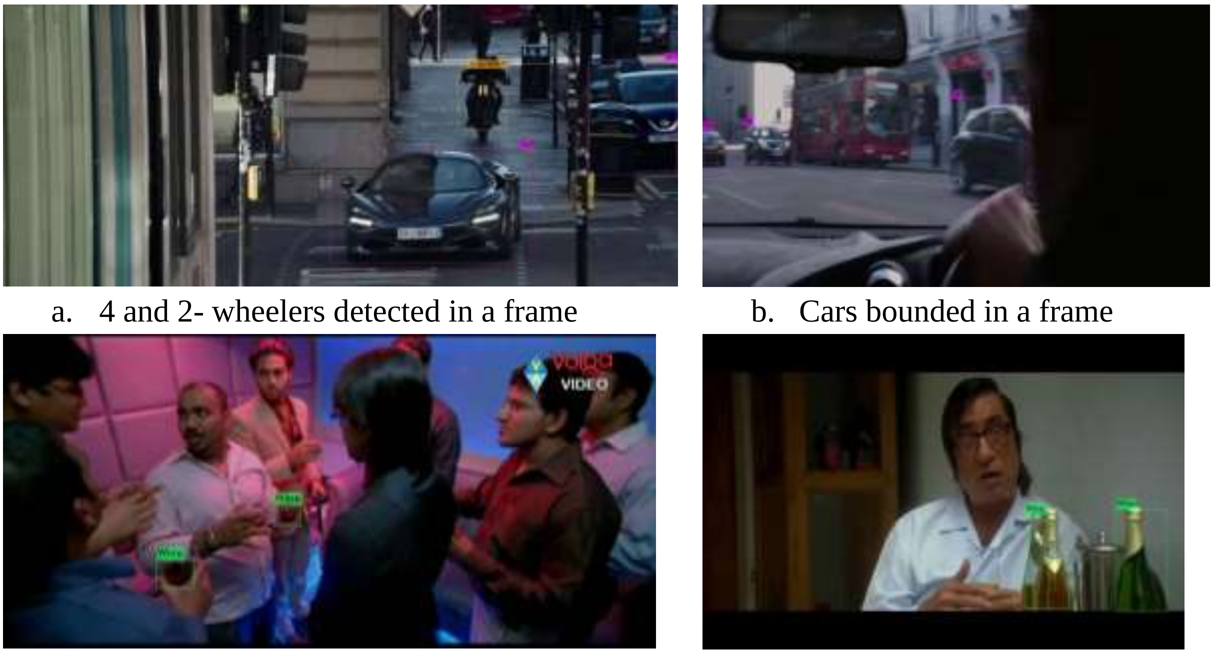
Table 2: Datasets used for training YOLO

S.No	Class	No. of Images
a.	4 Wheelers (Cars)	2000
b.	2 Wheelers (Bikes)	1964
c.	Glasses and Bottles	2036
d.	Cigarettes	2100

After the training is done, YOLO was given movie clips downloaded from YouTube [24]. The movies are from both Hollywood and Indian movies.

Results

The output of the first stage is object detection by YOLOv3, following are the samples



a. 4 and 2- wheelers detected in a frame

b. Cars bounded in a frame

c. Idenified multiple wine glasses in a frame

d. Detected Wine bottles

Fig. 1: Objects and their boundaries detected by YOLO

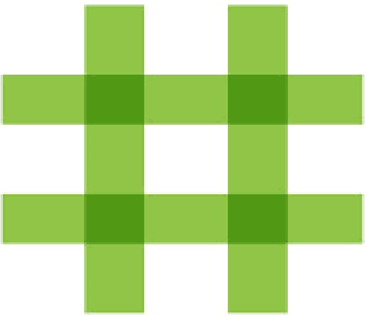
After YOLO was trained, testing is carried out by giving movie clips, video snippets given for testing are taken from YouTube. The results of the test phase by YOLO are detailed in Table 3 below:

Table 3: Object detection results obtained from YOLO

Source	Total frames	Total frames with respective objects	Total frames detected with objects
C1	346	4 Wheeler - 265	212
C2	581	4 Wheeler - 438	430
		2 Wheeler - 155	88
C3	199	4 Wheeler - 71	50
C3		2 Wheeler - 12	6
C4	359	4 Wheeler - 256	161
C5	444	4 Wheeler - 315	253
C6	554	Bottle/Glass - 443	355
C7	568	Bottle/Glass - 340	66
C8	446	4 Wheeler - 405	365

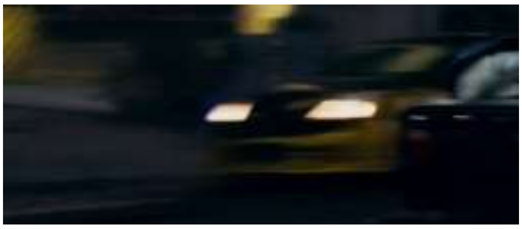
Firstly, frames with vehicles are identified using YOLO. For the detected frames with vehicles bounded by YOLO is given as input to blur detection algorithm. Blur is calculated by convolving Laplacian kernel on the frame that is converted as gray image. If the percentage of blur is increasing over neighbouring frames then the frames are marked as frames with speeding vehicle. Secondly, Bottles and glasses are identified from frames using YOLO and for frames with these objects statutory warning can be inserted as “Drinking is injurious”.

Video clip is given inputted to YOLO, object identification is done by YOLO by bounding the objects with boxes on the frames the video is given as output. It is observed that the frame rate is slightly reduced for the outputted video by YOLO. YOLO also didn’t identify small objects as we tried to train the YOLO to detect cigarettes the results given out were too low. YOLO was again trained by removing cigarette images. The average object detection rate was 76% for four wheelers (cars) and two wheelers (bikes) as 52% for glasses and bottles. The results of blur detection for sample input images are presented below in figure 2:

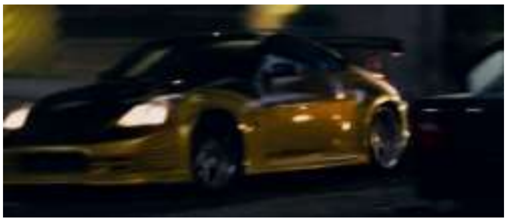


Hash image variance of Blur: 0.0212

Bear image variance of Blur: 0.029



Blur variance: 0.00026



Blur variance: 0.00048

Fig. 2: Varying variance of Blur for different images using Laplacian Kernel Approach

Conclusions and Future Scope

In this work we tried to identify cars, bikes, Wine bottles and glasses from the video clips using YOLO. Speeding cars and bikes are identified in two phases, in the first phase the said objects are identified by OD algorithm YOLO, and from the frames with cars and bikes are given as input to blur detection process, percentage of blur for the consecutive frames with vehicles in them are calculated. The change of blur percentage is observed for consecutive frames and if it is found to be increasing i.e., if the variance is decreasing then a statutory warning may be displayed when those frames are played. YOLO is also trained to detect Cigarettes to alert the spectators with alert message that smoking is injurious, as the object is tiny detection rate was erroneous. This can be addressed by training the OD algorithm on multi resolution images which are small.

Given any kind of video we can further extend the work to automate the process of detecting violence and or horror, which desensitises the audience, scenes which encourages alcohol consumption, drug addiction, tobacco consumption, cigarette smoking, cruelty to animals, etc. And once the segment of video with such content can be identified a statutory warning can be inserted which cautions the audience.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In CVPR, 2014.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region based convolutional networks for accurate object detection and segmentation,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 1, pp. 142–158, 2016.

[4] R. Girshick, “Fast r-cnn,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[5] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, 2015, pp. 91–99.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[7] Joseph Redmon and Ali Farhadi ,”YOLOv3: An Incremental Improvement”, 2018, arXiv

[8] https://pjreddie.com/darknet/yolo/

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in European conference on computer vision. Springer, 2016, pp. 21–37.

[10] Z. Zou, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” arXiv, vol. 1905.05055, 2018.

[11] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders, "Selective Search for Object Recognition", In International Journal of Computer Vision, 2013

[12] https://www.cbfcindia.gov.in/main/guidelines.html

[13] C. Maduro, K. Batista, P. Peixoto, and J. Batista, “Estimation of Vehicle Velocity and Traffic Intensity Using Rectified Images,” IEEE International Conference on Image Processing (ICIP), pp. 777–780, 2008.

[14] H. Zhiwei, L. Yuanyuan, and Y. Xueyi, “Models of Vehicle Speeds Measurement with a Single Camera,” International Conference on Computational Intelligence and Security Workshops (CISW), pp. 283–286, 2007.

[15] H. A. Rahim, U. U. Sheikh, R. B. Ahman, and A. S. M. Zain, “Vehicle Velocity Estimation for Traffic Survillance System,” World Academy of Science, Engineering and Technology (WASET), p. 772, 2010

[16] Schoepflin T.N. and Dailey D.J., “Dynamic Camera Calibration of Roadside Traffic Management Cameras for Vehicle Speed Estimation,” Intelligent Transportation Systems (ITS), 2003.

[17] Huei-Yung Lin, Kun-Jhih Li, Chia-Hong Chang,”Vehicle speed detection from a single motion blurred image”, Image and Vision Computing, Volume 26, Issue 10, 2008, Pages 1327-1337, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2007.04.004.

[18] S. Hua, M. Kapoor and D. C. Anastasiu, "Vehicle Tracking and Speed Estimation from Traffic Videos," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 153-1537, doi: 10.1109/CVPRW.2018.00028.

[19] J. Shi, L. Xu and J. Jia, "Discriminative Blur Detection Features," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2965-2972, doi: 10.1109/CVPR.2014.379.

[20] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," Proceedings. International Conference on Image Processing, 2002, pp. III-III, doi: 10.1109/ICIP.2002.1038902.

[21] Said Pertuz and D. Puig and M. Garcia, “Analysis of focus measure operators for shape-from-focus”, Pattern recognition, 2013, v. 46, pages: 1415-1432

[22] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, A. Kolesnikov, T. Duerig, and V. Ferrari, “The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale.”,IJCV, 2020.

[23] Vittorio, Angelo, “Toolkit to download and visualize single or multiple classes from the huge Open Images v4 dataset”, https://github.com/EscVM/OIDv4_ToolKit, GitHub repository,2018.

[24] https://www.youtube.com/