OBJECT DETECTION AND CLASSIFICATION USING YOLOV3

Pavithra P M[1], Dr. Bhavani R[2]

[1]*M.E. Student, Computer Science And Engineering, Government College of Technology, Coimbatore, Tamilnadu, India*

[2]*Assistant Professor, Computer science and Engineering, Government College of Technology, Coimbatore, Tamilnadu, India*

Abstract - *Object detection has several advantages in computer vision technologies. It is used in image retrieval, security, observations, etc. The goal of object detection system is object localization and identifying the category to which the object belongs. In this paper, a deep learning algorithm YOLO (You Only Look Once) is used for object detection and classification. This proposed method yields mean average precision (mAP) of 95% for traffic scenario images in identifying traffic lights, car, bus, person and motorcycle.*

Key Words: Object detection, Classification, Deep learning, YOLO, Traffic scene, Convolutional Nueral Network.

1. INTRODUCTION

Object detection is a computer vision technique that is used to identify and locate objects in an image. Specifically, object detection draws bounding boxes around the detected objects, which allows to locate where the objects are in a given image. Object detection takes major role in surveillance and security, traffic checking and activity. It is used in object tracking. For example, tracking individuals in a mall. In autonomous driving vehicle, object detection is needed to decide what to do next like accelerate, apply breaks or turn. It needs to know location of the objects in the road. The proposed system is developed using YOLO for detection and classification of objects in traffic scene images.

2. LITERATURE SURVEY

Unified, Real-Time object detection, paper written by Joseph Redmon. Their prior work is on detecting objects employing a regression algorithm. to urge higher accuracy and good predictions they need proposed YOLO algorithm during this paper. one neural network predicts bounding boxes and sophistication probabilities directly from full images in one evaluation. The mAP value of this paper is 63.4% [10]. Machine Learning Technique named YOLO (You Only Look Once) algorithm using Convolutional Neural Network is employed for the thing Detection. during this work COCO dataset is employed which has the pretrained weight and threshold value is about to 0.3.Object detected successfully in images and videos [1]. An android application is created which detects different types of objects and it returns voice feedback to the user. It says that object detection using YOLO algorithm is faster as compared to other classification algorithms and it makes localization errors but it predicts less false positives in the background. The proposed method provides the average IoU is 83.19%, mAP is 98.14% [7]. Objects detection in shelf images, it can solve many problems in retail sales like monitoring the amount of products on the shelves, completing the missing products and matching the planogram continuously. The experimental study is performed using Coca Cola images obtained from Imagenet and grocery dataset with YOLO algorithm. At 900 iteration the entire loss of single class model is 25.39% and total loss of ten class model is 29.84% [8].YOLOv2 is used for improving the computation and processing speed and at the same time efficiently identify the objects in the video records. The classification algorithm creates a bounding box for every class of objects for which it is trained, and generates an annotation describing the particular class

of object. In this system GPU is used to increase the computational speed. YOLOv2 processes at 40 frames per second. The execution time for a image was close to 0.5 second [5].

Traffic scene perception (TSP) has objective to extract accurate real-time environment information on road, which involves three phases: detection of objects of interest, and tracking of objects in motion. Since recognition and tracking often believe the results from detection, the facility to detect objects of interest effectively plays a crucial role in TSP. This paper, focussed on three classes of objects. These are traffic signs, cars and cyclists. The advantage of using one common framework is that the detection speed is far faster, since all dense features need only to guage once within the testing phase. Average Precision of KITTYCar test set is 87.19% and the run time is 1.5s [4]. In order to detect the thing of interest in a picture with multiple similar objects, a window for feature matching is employed. The window is optimized in size for better performance of the thing detection. As a practical example a service that visualizes the situation of a desired book during a library is taken into account [6]. A research progress in the development of object detection using Deep Learning based on drone camera. Purpose of this research is to deliver important medical aids for patients in emergency situations. This case can be simplified into delivery of an item from start to the global position. In this paper, the combination of MobileNet and the Single Shot Detector (SSD) framework for fast and efficient deep learning based method to object detection. The MobileNet SSD detector was used as object detector with high accuracy detection with average about 14 FPS [2]. The role of deep learning techniques based on convolutional neural network for object detection is elucidated. Deep learning frameworks and services available for object are also enunciated. Deep learning techniques for state-of-the-art object detection systems are used in this paper. The author suggest the system should be extended to cope with real time full motion video generating frames at 30 to 60 per second [11]. A new scheme for detection and tracking of specific objects in a knowledge-based framework is used. This scheme uses a supervised learning method: Support Vector Machines. Problems detection and tracking are solved by a SVM classifier. Objects are tracked along the time by a SVM tracker with complete 6 parameters affine model. The model is applied in a video surveillance application for detection and tracking of frontal view faces [3].

3. PROPOSED SYSTEM

The proposed system was implemented using Darknet-53 which acts as acts a backbone for YOLOv3 to detect and classify the objects. The steps involved in the proposed system are data collection, data modeling, train and test the model and Performance analysis.

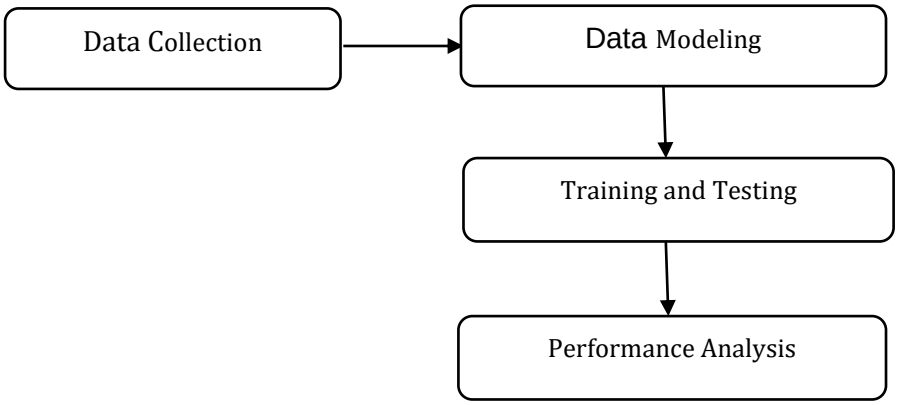

Fig-1: Proposed System

3.1 Data collection

Dataset is prepared from Open Images Dataset v5 using OIDv4 toolkit. The dataset has five class of objects. These are bus, car, person, motorcycle and traffic lights. The images are collected with its labels and bounding box coordinates. Data annotation also done using OIDv4 toolkit.

3.2 Data modeling:

Object detection and classification is implemented using YOLOv3. Darknet-53 [9] has 53 convolutional layers which is stacked with 53 more layers producing 106 layers.

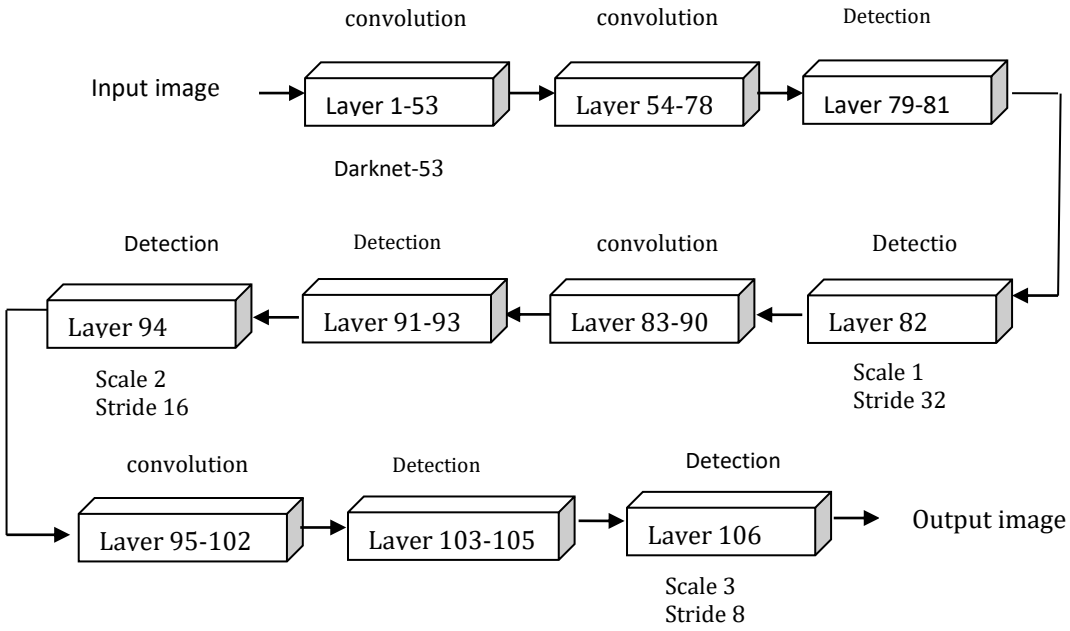

Fig. 2 : YOLOv3 Architecture

The input image is a 416x416 RGB image. In Darknet-53 each convolution layer is followed by a batch normalization layer and LeakyReLU layer. YOLOv3 makes detections at three different scales. Network downsamples input image by straide 32,16 and 8 at layer 82,94 and 106 respectively. The resultant feature maps are 13x13,26x26 and 52x52. Then each detection is made using 1x1 detection kernel which yields the detection feature maps 13x13x255,26x26x255 and 52x52x255. Here, 13x13 is liable for detecting large objects, 26x26 is liable for detecting medium objects and 52x52 is liable for detecting small objects. YOLOv3 predicts 507 bounding boxes at scale 1 (13x13), 2028 bounding boxes at scale 2 (26x26) and 8112 bounding boxes at scale (52x52) for an image. These are filtered using following two methods.

i) Filtering based on the score

A threshold value is set and compared with the confidence score of the boxes. Then the bounding boxes with less threshold value is removed. The box confidence is calculated as

Box confidence = Pr(object) x Intersection Over Union

ii) Non-Max Suppression

The non-max suppression is to select the best bounding for an object and it suppress all other bounding boxes. It considers confidence score of a bounding box and overlap of the bounding boxes.

3.3 Training and testing

The training was done using Google Colaboratory it provide Tesla P4 GPU for faster and efficient training of the network. Training dataset has a batch of 1500 images. The dataset was trained for 10000 iterations as 5(total classes) * 2000 = 10000. The total amount of time required to train the network with the above configurations was approximately 14-16 hours. The weights thus generated after 10000 iterations were used to detect and analyze the performance. In testing process a batch of 100 images has been tested. It detects bounding boxes and class labels with better accuracy.

3.4 Performance analysis

The parameters used for testing are precision, recall, mAP, IoU and f1 score. These performance metrics are calculated using true positive, false positive and false negative values.

Precision is that the number of positive class predictions out of all positive classes

$$\textbf{Precision} = TP / (TP+FP)$$

Recall is the number of positive class predictions made out of all predicted result.

$$\textbf{Recall} = TP / (TP+FN)$$

F1-score is that the weighted average of precision and recall.

$$\textbf{F1-score} = 2 * (precision*recall) / (precision+recall)$$

Intersection over Union is a measure of overlap between actual bounding box and the predicted bounding box.

$$\textbf{IoU} = \frac{Area(A \cap B)}{Area(A \cup B)}$$

Mean Average precision is that the mean of average precision.

4. RESULT AND DISCUSSION

Detected objects with bounding boxes are shown in the following images.

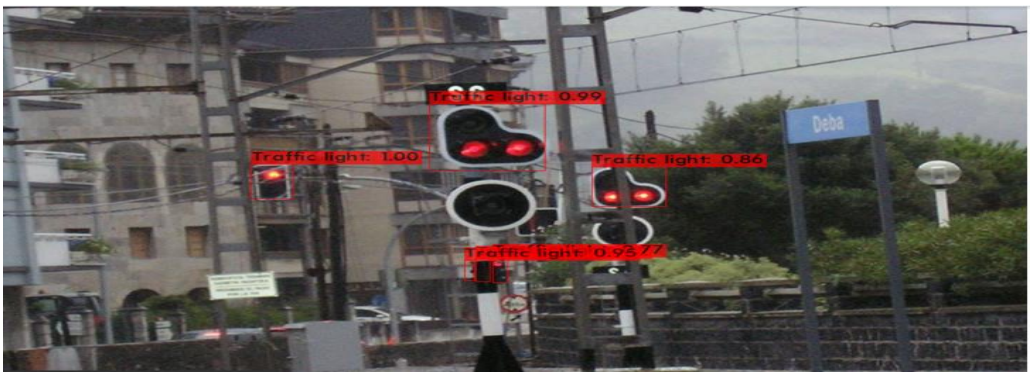

Fig. 3 : Results obtained from images

Class Name	**True positive (TP)**	**False Positive (FP)**	**Average Precision (AP)**
Bus	36	1	99.50 %
Car	60	1	93.57 %
Person	117	28	83.77 %
Traffic light	28	5	74.22 %
Motor cycle	27	1	99.60 %

Table 1 : Average Precision for each class

Table 1 shows the average precision of each class which is used to calculate the mean Average Precision.

Table 2 : Performance metrics for confidence threshold 0.25

TP	**FP**	**FN**	**Precision**	**Recall**	**F1-score**	**Average IoU**	**mAP**
268	36	43	0.88	0.86	0.87	71.82%	90.13%

Table 2 shows the performance metrics of the object detection model for confidence threshold 0.25. The mean Average Precision of Object detection and classification is 90.13% and the run time is 3 sesonds.

5. CONCLUSION AND FUTURE ENHANCEMENT

This work is developed with objective of detecting the objects in traffic scene images. The Bounding boxes are drawn around the detected objects along with the label indicating the class to which the object belongs. The proposed system yields the mAP of 90% and average IoU of 72%. This model can detect five objects in traffic scenes which can be scaled further to detect more number of objects.

REFERENCES

[1] Amin, P., Anushree, B.S., Shetty, B.B., Kavya, K. and Shetty, L., 2019. Object Detection using Machine Learning Technique.

[2] Budiharto, W., Gunawan, A.A., Suroso, J.S., Chowanda, A., Patrik, A. and Utama, G., 2018, April. Fast object detection for quadcopter drone using deep learning. In 2018 3rd International Conference on Computer and Communication Systems (ICCCS) (pp. 192-195). IEEE.

[3] Carminati, L., Benois-Pineau, J. and Jennewein, C., 2006, October. Knowledge-based supervised learning methods in a classical problem of video object tracking. In 2006 International Conference on Image Processing (pp. 2385-2388). IEEE.

[4] Hu, Q., Paisitkriangkrai, S., Shen, C., van den Hengel, A. and Porikli, F., 2015. Fast detection of multiple objects in traffic scenes with a common detection framework. IEEE Transactions on Intelligent Transportation Systems, 17(4), pp.1002-1014.

[5] Jana, A.P. and Biswas, A., 2018, May. YOLO based Detection and Classification of Objects in video records. In *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 2448-2452). IEEE.

[6] Lee, J., Bang, J. and Yang, S.I., 2017, October. Object detection with sliding window in images including multiple similar objects. In 2017 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 803-806). IEEE.

[7] Masurekar, O., Jadhav, O., Kulkarni, P. and Patil, S., 2008. Real Time Object Detection Using YOLOv3.

[8] Melek, C.G., Sonmez, E.B. and Albayrak, S., 2019, July. Object detection in shelf images with YOLO. In *IEEE EUROCON 2019-18th International Conference on Smart Technologies* (pp. 1-5). IEEE.

[9] Redmon, J. and Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

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

[11] Zhou, X., Gong, W., Fu, W. and Du, F., 2017, May. Application of deep learning in object detection. In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) (pp. 631-634). IEEE.