Automation using Machine Learning and Object Detection

Prof. Trupti Shah
Department Of Electronics and Tele-Communication(EXTC)
Vidyavardhini's College of Engineering &
Technology(VCET) Mumbai,India

Akshaykumar Pillai
EXTC Vidyavardhini's College of
Engineering and Technology
Mumbai, India

Akash Dhayalkar
EXTC Vidyavardhini's College of
Engineering and Technology
Mumbai, India

Meghan Yesji
EXTC Vidyavardhini's College of
Engineering and Technology
Mumbai, India

Abstract - A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning-based approach, which leads to slow and non-optimal performance. In this paper, a completely deep learning- based approach is used to solve the problem of object detection in an end-to-end fashion. The paper aims to incorporate state-of-the-art technique for detecting the object placed in front of the webcam with the goal of achieving high accuracy with a real-time performance using deep learning. Based on the detected image several preprogrammed robots are used to transport the object in the detected image from the place where humans cannot work flawlessly to the desired location efficiently. This paper comes with the combination of deep learning and robotics which can be used in several areas such as mines, construction sites, steel factories etc where human works in a risky environment. The network is trained on the most publicly available data set, on which an object detection challenge is conducted annually.

Keyword:- Machine Learning, object detection, Single shot detection, automation, robotics.

I. INTRODUCTION

Efficient and accurate object detection has been an important topic in the advancement of computer vision systems. With the advent of deep learning techniques, the accuracy for object detection has increased drastically. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning- based approach, which leads to slow and non-optimal performance. The main aim of object detection is to find the exact location of an object in each picture accurately and mark the object with the appropriate category. To be very clear, the problem that object detection seeks to solve involves determining where and what the object is. In this paper, a completely deep learning-based approach is used to solve the problem of object detection in an end-to-end fashion. Once the image in front of the camera is detected accurately then based on that image a conveyor belt containing the particular item shown in the image will get triggered. This trigger of conveyor belt will make it roll and the item above the belt will move forward and eventually fall on the robot carrier placed underneath the conveyor belt. Once the item is in the carrier robot, the robot will move forward following a line and reach the desired destination. Upon picking up the item from the robot carrier at the destination, the robot will move in reverse direction and will halt at the initial position. This paper comes with the combination of deep learning and robotics which can be used in several areas such as mines, construction sites, steel factories etc where human works in a risky environment. The following challenges have been identified. 1. The need to distinguish between similar objects. 2. Identification of multiple objects in a single frame, where some objects might be only partially visible, and others are overlapping. 3. Collecting and pre-processing of datas for training. The network used here in this paper can be enforced on unified detection YOLO [4] or Single shot detection (SSD) [5]. The network is trained on the most publicly available dataset, on which an object detection challenge is conducted annually. The resulting system will be fast and accurate, thus aiding those applications which require object detection.

II. LITERATURE SURVEY

2.1. R-CNN (Region based convolutional neural network)
[1] :- To find a way around the problem of choosing a vast

number of regions, Ross Girshick et al. papered a method in which he use a discriminating search to dig out just 2000 regions from a image and he called them region proposals. Therefore, now, as an alternative of trying to categorize a huge number of regions, we can just effort with 2000 regions.

2.2. Fast R-CNN [2] :- The similar author of earlier paper (R- CNN) also solved some of drawback of R-CNN to construct the object detection algorithm called fast R-CNN. This algorithm is similar to the previous R-CNN algorithm. But instead feed region proposals to CNN, we can provide the input image to build the convolutional map. This is fast than R-CNN because there is no need to feed proposals of 2000 region to convolutional neural network each time. Instead of this, the convolution operation is made only one time per image and a feature map generates from it.

2.3. Faster R-CNN [3] :- R-CNN & Fast R-CNN finds region proposal using selective search. This is a time-consuming & slow process which affects the performance of network. Due to this problem, Shaoqing Ren et al. built an algorithm of object detection that eliminate the discriminating search algorithm and let the network learn the region proposal. Same as fast R-CNN, the image is provided to a convolutional network as an input that will provide the convolutional feature map. A different network is used to guess the region proposals instead using of selective search algorithm on the feature map to recognize the region proposals. Reshaping of the region proposals which are predicted is done using a RoI pooling layer which classifies the image within the planned region and guess the offset value for bounding boxes.

2.4. YOLO (You Only Look Once) [4] :- All of the earlier object detection algorithms use regions to localize the object inside the image. Apart from the above seen algorithms which are region based, You Only Look Once or YOLO is an algorithm based on object detection which is much different.

The only convolutional network predicts the bounding boxes and the class probabilities for these boxes in YOLO.

2.5. SSD (Single Shot Detection) [5]:-A one single shot can be taken to detect numerous objects inside the image in SSD, while regional proposal network (RPN) approaches such as R- CNN series which requires two shots, one for proposal of generating regions and the other for detection of object of every proposal. Hence, SSD is faster than RPN based approach.

III. PROBLEM STATEMENT

The major challenge in this problem is that of the variable dimension of the output which is caused due to the variable number of objects that can be present in any given input image. Any general deep learning task requires a fixed dimension of input and output for the model to be trained. Another important obstacle for widespread adoption of object detection systems is the requirement of real-time (30fps) while being accurate in detection. The more complex the model is, the more time it requires for inference; and the less complex the

model is, the less is the accuracy. This trade-off between accuracy and performance needs to be chosen as per the

application. Classifications as well as regression are the major problems involved which is leading the model to be learnt simultaneously. This adds to the complexity of the problem. A lot of work is there in object detection by the use of traditional computer vision techniques (sliding windows, deformable part models). However, lack of accuracy of deep learning-based techniques. Among the deep learning-based techniques, two broad class of methods are prevalent: two stage detection (RCNN [1], Fast RCNN [2], Faster RCNN [3]) and unified detection (Yolo [4], SSD [5]. The robot used here follows a line to transport the object from source to destination; the irregularity in the line can make the robot to halt unnecessarily. Moreover, the path surface should be even so that the carrier robot can move back and forth flawlessly.

IV. PROPOSED METHDOLOGY

4.1 SSD

Sliding window detection, as its name suggests, slides a local window across the image and identifies at each location whether the window contains any object of interests or not.

Multi-scale increases the robustness of the detection by considering windows of different sizes. Such a brute force strategy can be unreliable and expensive: successful detection requests the right information being sampled from the image, which usually means a fine-grained resolution to slide the window and testing a large cardinality of local windows at each location. Input and Output: The input to an SSD is an image which is of fixed size, for example, 512x512 image for SSD512. The fixed size constraint is mainly for efficient training with batched data. Being fully convolutional, the network can run inference on images of different sizes. The output of SSD is a prediction map. Each location in this map stores classes confidence and bounding box information as there are indeed

an object of interests in every location. Obviously, there will be a lot of false alarms, so a further process is used to select a list of most likely prediction based on simple heuristics.
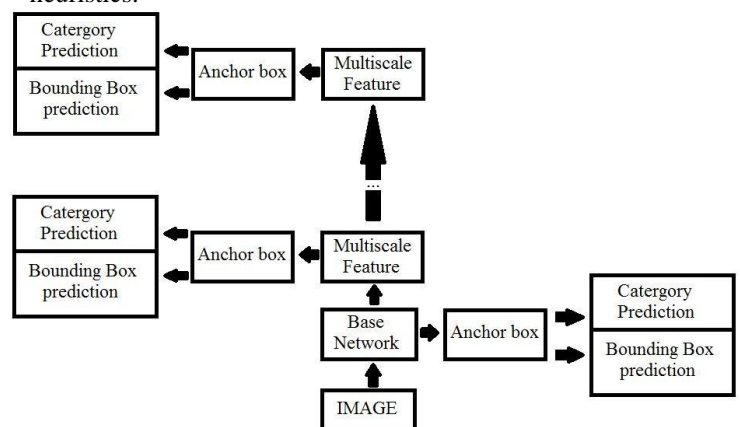

Fig. 4.1 Block diagram of SSD

4.2. Yolo

A only convolutional network is the one which identifies more than one bounding boxes as well as class probabilities for boxes at a same time. YOLO trains the full images and optimizes detection performance. This unique model has numerous benefits when compared with traditional methods of object detection.

Firstly YOLO is very fast. Since we make detection as a reversion problem we do not need a difficult pipeline. We simply run the neural network on a fresh image during testing to predict detections. Secondly, the base network operates at a speed of 45 fps with no batch operating on a Titan X GPU, while a quick version runs at more than 150 fps. It means streaming video can be processed in real-time with about less than 25 ms of latency. When compared with sliding window and region proposal-based techniques, YOLO observes the whole image during training time so it completely encodes contextual information about classes as well as their appearance. Fast R-CNN, a popular detection method, makes error in background patches in the image for the objects because it cannot observe the large context. YOLO makes almost less than half number of errors in background when compared with fast R-CNN. Thirdly YOLO learns generalizable representations of objects. When it is trained on normal images and tested on artwork, YOLO beat top detection methods like DPM and R-CNN by a large margin. Since YOLO is highly generalizable model it can break down if applied to fresh domains or unpredicted inputs.

4.3 Conveyor belt

Once the object in front of the camera is detected and identified with the help of the above mentioned algorithms then the conveyor belt at the distant location holding the detected object will start. This will allow the object to move forward and eventually fall on a carrier robot which is placed underneath the conveyor belt at its end of the length

Fig. 4.3 Conveyor belt

4.4 Line follower robot

Line follower Robot is a machine which follows a black line. Concept of working of line follower is related to light. We use here the behavior of light at black and white surface. When

light fall on a white surface it is almost full reflected and in case of black surface light is completely absorbed. This behavior of light is used in building a line follower robot. In this arduino based line follower robot we have used IR Transmitters and IR receivers also called photo diodes. They are used for sending and receiving light. IR transmits infrared lights. When IR rays falls on white surface, it's reflected back

and caught by photodiodes which generates some voltage

changes. When IR light falls on a black surface, light is absorb by the black surface and no rays are reflected back, thus photo diode does not receive any light or rays.

Here in this arduino line follower robot when sensor senses white surface then arduino gets 1 as input and when senses black line arduino gets 0 as input.

Based on these fundamentals, the robot will reach the destination with the object and will stop and the stop mark. Upon manually picking up the object from the carrier at the stop point of the robot; the robot will move backward to the initial position i.e. under the end length of the corresponding conveyor belt
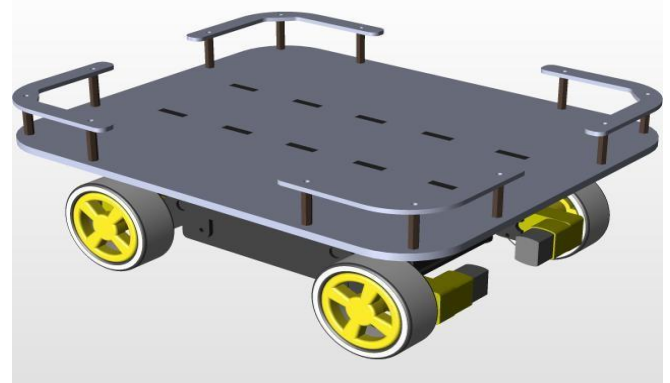

Fig. 4.5 Line follower robot with a carrier Procedure flow

1. Object sample is shown in front of the webcam .
2. The algorithm used will detect and categorize the object.
3. Once the object is detected and categorized accurately a unique signal corresponding to the object will be sent to a controller which controls a conveyor belt placed at a distant location.
4. Thus the corresponding conveyor belt holding the original object whose sample was shown in front of the webcam will get started
5. As a result, the object above the conveyor belt will move forward and eventually fall on the carrier robot.
6. When the object arrives at the carrier the robot will start moving in forward direction until it reaches the stop mark at the desired destination.
6. When the robot reaches the desired destination, a person should manually pick up the object from the carrier.
7. As soon as the object is picked up from the carrier the robot will move backward to its initial position.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

V. RESULT

The proposed system is able to accurately identify the object in front of the camera and with the help of detected object a conveyer belt at a distinct location is triggered successfully. Upon this trigger of the conveyor belt the detected bottle is loaded on to the robot carrier and the robot start to follow the particular predefined line and reach the desired destination.

VI. CONCLUSION

An accurate and efficient object detection system has been developed which achieves comparable metrics with the existing state of a art system. This paper uses recent techniques in the field of computer vision and the deep learning. Custom data set was created using labeling and the evaluation was consistent. An efficient transportation robot is also built to transport an object from a distant point to a desired location

VII. ACKNOWNLEDGMENT

REFERENCES

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[2] Ross Girshick. Fast R-CNN. In International Conference on Computer Vision (ICCV), 2015.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems (NIPS), 2015.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, ChengYang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.