International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Object Detection for Relative Positioning

Anish George[1], Drishya P Suresh[1], M Shama[1], Ayush Kumar Sharma[1], Anuradha M. G[2], Bharathkumar Hegde[3]

[1]Dept. of Electronics and Communication Engineering, JSS Academy Of Technical Education, Bengaluru, India
[2]Asst. Professor, Dept. of Electronics and Communication Engineering, JSS Academy Of Technical Education, Bengaluru, India
[3]Director, Autoyos Private Limited, Bengaluru, India

Abstract - *Object detection in computer vision has been advanced vastly with the culmination of Deep Neural Networks. Object detection is a computer vision technique for locating instances of objects in images or videos. With the increase of training data and the improvement of machine performance, the object detection methods based on convolutional neural networks (CNNs) have become the mainstream algorithm in field of the current object detection. Popular object detection algorithms based on neural networks include R-CNN, YOLO and SSD. The object detection model can be used to track multiple objects of different classes in real time. The object detection models can be further explored to incorporate real time positioning. These systems can be vastly exploited in automation of tasks such as automobile self-parking, industrial automation and robotics.*

Key Words: **Object Detection, Positioning, Relative Positioning, FPS, mAP, YOLO**

1.INTRODUCTION

Computer vision is a field of image processing that helps computers to obtain a higher level of understanding from the content of digital images. It is used to mimic the ability of human vision and can be used to automate tasks that the human visual system can do. Object detection in computer vision started in the early 2000's with an efficient algorithm known as the "Viola and Jones Algorithm" for facial detection. The algorithm soon made its way to be used in Python's OpenCV library. The facial features were hand-coded and as a result, the algorithm was not efficient in detecting faces with different configurations. This was then followed by HOG (Histograms of Oriented Gradients) algorithm. It tried to identify darker pixels surrounding the pixel under consideration and an arrow was drawn in the direction where the image was getting darker. This was repeated till a simpler representation of the image was made and it captured the essence of its features. There was still the drawback of hand-coding the feature map.

Computer vision technology had improved significantly with Deep Learning. This is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. Deep Learning uses convolutional neural networks that helps in accurate detection of objects and its classification. Classification in this case was an iterative process wherein, object detection was done in a small area of pixels and the pixel area continuously increased till the entire image was covered.

Faster R-CNN is an algorithm with excellent performance both in detection accuracy and in detection speed. Faster R-CNN replaces the selective search method with region proposal network (RPN) which makes the algorithm much faster. Faster RCNN which combined the RPN network and the Fast R-CNN network is one of the best ways for object detection of R-CNN series based on deep learning.

The CNN models have a higher accuracy but are very slow as the image is passed through the neural network for each proposed object region. The SSD model offers good accuracy of predictions at a good frame rate as the object localization and classification is done in a single forward pass of the network, using bounding box regression technique. The YOLO algorithm was also developed to combat the brute force approach of iterative search. YOLO expanded as "You Only Look Once" looks at the image once, instead of iterations. YOLO is a fast and accurate method for real-time object detection. These object detection algorithms can be trained on custom datasets to perform tasks requiring real-time positioning.

2. LITERATURE REVIEW

In 2001, Viola and Jones proposed a new real-time object detection framework using Haar Features [1]. This was mainly focused on the problem of face detection. In 2005, Dalal and Triggs proposed the Histogram of Oriented Gradients (HOG) feature descriptor for use in computer vision and image processing [2]. It identified the darker pixels surrounding the pixel under consideration and an arrow was drawn in the direction where the image got darker. This was repeated till a simpler representation of the image was made, capturing the essence of its features. These algorithms have been implemented in the OpenCV library for programming facial detection modules. The drawback with these

algorithms was that the feature map had to be hand-coded and thus were hard to adapt.

In 2014, J. Donahue et al. proposed the Region-CNN (RCNN) model that uses the concept of selective search for region proposals to hypothesize object locations [3]. While it had a good accuracy of around 66% mAP, it was computationally expensive and slow (47 seconds per test image) as the image is passed through the neural network for each proposed object region. In 2015, Ross Girshick proposed the Fast R-CNN model with 66.9% mAP that took 2 seconds per test image [4]. The Faster R-CNN model proposed by R. Girshick et al. in 2016 took 0.2 seconds per test image, with a 66.9% mAP [5]. However, the CNN models cannot be used for real-time object detection. CNNs also take a longer amount of time to train the network due to the large number of region proposals per image.

Single Shot MultiBox Detector (SSD) was a single shot object detection method proposed by W. Liu et al. [6] for real-time object detection. Single Shot Multibox Detector or SSD is a single shot object detection algorithm designed for real-time applications. Two stage networks like Faster R-CNN use a region proposal network to create boundary boxes and utilize those boxes to classify objects. While they are highly accurate, it has a very low processing speed of 7 frames per second, which makes them unsuitable for real time detection. SSD speeds up the process by eliminating the need of the region proposal network. To recover the drop in accuracy, SSD applies a few improvements like multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN's accuracy using lower resolution images, which further pushes the speed higher. SSD only takes one single shot to detect multiple objects within the image, while regional proposal network based approaches need two shots - one for generating region proposals and one for detecting the object of each proposal. Thus, SSD is able to achieve real time processing speed with accuracy comparable to that of Faster R-CNN.

A newer model called YOLO (You Only Look Once) was developed to combat the brute force approach of iterative search as seen in CNNs. It was proposed in 2016 by Joseph Redmon et al. as a faster light-weight model for real-time object detection [7]. YOLO has a processing speed of 45 frames per second. A smaller version of the network, Fast YOLO, has a processing speed of 155 frames per second while achieving double the mAP of other real-time detectors. In 2017, Redmon and Farhadi proposed YOLO9000 (or YOLOv2) as an improvement on YOLO [8]. It has a 76.8 mAP at 67 FPS, and 78.6 mAP at 40 FPS. It can detect over 9000 object categories.

YOLO divides the image into a group of 13 by 13 cells to identify different classes of images. Each of these cells are responsible for finding five bounding boxes. A bounding box describes the rectangle enclosing the object. YOLO outputs a confidence score that shows the certainty of the bounding box enclosing an object. This certainty is defined by the thickness of the boundary of the bounding box. The content within the box is also classified. The confidence score and class prediction are combined into one final score that gives the probability of an object being detected in the box. As there are 169 grids cells that predicts 5 bounding boxes each, a total of 845 bounding boxes are obtained on the image. Based on the confidence threshold most of these boxes are omitted, retaining boxes with higher confidence values and object is detected.

Accuracy is measured as the mean average precision mAP: the precision of the predictions. SSD300 achieves 74.3% mAP at 59 FPS while SSD500 achieves 76.9% mAP at 22 FPS, which outperforms Faster R-CNN which has 73.2% mAP at 7 FPS and YOLOv1 that achieves 63.4% mAP at 45 FPS. Fast YOLO achieves a mAP of 52.7% at 155 FPS.

3. METHODOLOGY

The various object detection algorithms like SSD and YOLO were implemented for accurate real-time object detection. The neural network was trained on a dataset of around 1200 images to detect custom objects in a real time video feed. All these methods implementing CNNs were proved to be very accurate for the purpose of detecting objects. The speed and accuracy of these algorithms were compared. The detected objects were to be positioned to some predetermined reference. In order to achieve this, the centers of the bounding box and reference box were located. A line was drawn to join these centers. If the length of the line was below a certain threshold, the object was considered to be aligned. Otherwise, the object was not aligned. The video feed was taken from a webcam connected to the computer. The relative position is determined by two parameters Δx and Δy, which represent the relative change in the x and y coordinates required for alignment.

4. RESULTS

Accurate and rapid object detection is a crucial requirement for the autonomy of any system implementing computer vision for real-time positioning. The object detection was done using SSD and YOLO models that were trained on a dataset of around 1200 annotated images. The R-CNN model was not implemented owing to slow detection speeds, thus it was not suitable for real-time purposes.

Figure 1 shows the output of relative positioning. The white spot is the custom object to be detected. As seen in the figure, the YOLOv2 algorithm detects it with 96.4% certainty. The red box denotes the bounding box of the object and the blue box is the predetermined reference to which object has to be aligned. The green dotted lines for Δx and Δy are shown for reference purposes. Alignment needs to be done such that Δx and Δy are minimum. This happens when the length of the

line joining the two centers, shown in yellow in the figure, is below a certain threshold. Once below the threshold, the object is aligned.

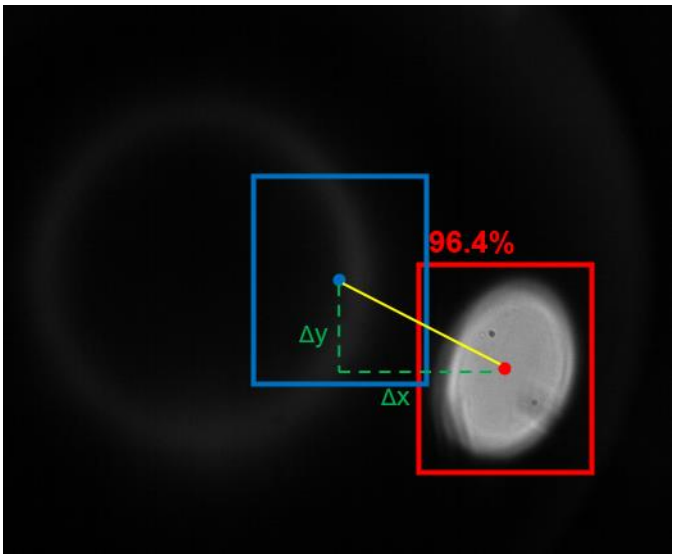

Fig -1: Relative positioning output

The YOLO model had a frame rate of 6 FPS on an Intel i5-8250 CPU and a frame rate of 22 FPS on an NVIDIA GTX 1070 GPU. The SSD model had a frame rate of 8 FPS on an Intel i5-8250 CPU. These models had a good accuracy in detecting the objects. However due slower frame rates especially in lower specification hardware such as mobile processors, these models cannot be used for real-time applications in such devices. Performance can be enhanced with the help of peripheral hardware accelerators.

5. CONCLUSIONS

The field of computer vision is rapidly flourishing, holding vast potential in many fields such as automation on a global scale. With further research and refinement, computer vision technology will be easier to train and be able to recognise images with even more accuracy. Deep learning techniques are able to do end-to end object detection without specifically defining features, and are typically based on CNNs. It has many practical applications such as autonomous driving systems, surveillance applications and industrial product inspection It is also widely used in applications such as image annotation, activity recognition, face detection and video object tracking.

There are multiple object detection algorithms such as R-CNN, YOLO, and SSD. While R-CNN is the most accurate, it is also the slowest one out of the three. YOLO is an object detection algorithm that sees the complete input image at once as opposed to looking at it multiple times via region proposals as seen in R-CNN. It also runs the image through CNN only once instead of multiple times as region proposals. This makes it super-fast and as a result, it is the fastest object detection algorithm that can be used in real time applications. One limitation for YOLO is that it only predicts one type of class in one grid hence, it struggles with very small objects. But this is significantly reduced in YOLOv3. With multiple improvements, YOLO-v3 can be seen as the most ideal choice of algorithm in objection for real time applications. SSD is a balanced object detection algorithm with good accuracy as well as real-time execution. Further improvements in accuracy and speed for detection in SSD algorithms can help improve the field of computer vision and hence human convenience.

Relative positioning of objects detected by these algorithms can find vast applications especially in autonomous systems. These systems can be used in automated product manufacturing, for maintaining a safe distance between a vehicle and any other objects such as pedestrians or another vehicle and in robotics.

ACKNOWLEDGEMENT

The project was done under the wing of Autoyos Private Limited. We deeply appreciate the support provided by Dr. Bharathkumar Hegde, and are grateful for the facilities and resources provided by Autoyos.

REFERENCES

[1] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.

[2] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, pp. 886-893. IEEE, 2005.

[3] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

[4] Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.

[5] Ren, Shaoqing, 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, pp. 91-99. 2015.

[6] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016.

[7] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

[8] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271. 2017.