International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 2, Mar-Apr-2022, ISSN (Online): 2395-566X

# Real Time Object detection and Tracking Using Open-CV

**Zaid Bin Shafi**
Department of Computer Engineering
Chandigarh University
Mohali,Punjab.

**Abstract-Computer vision is a very progressive and modern part of computer science. From the scientific point of view, theoretical aspects of computer vision algorithms prevail in many papers and publications. The underlying theory is really important, but on the other hand, the final implementation of an algorithm significantly affects its performance and robustness. For this reason, this paper tries to compare the real implementation of tracking algorithms (one part of the computer vision problem), which can be found in the very popular library OpenCV. Moreover, the possibilities of optimizations are discussed. An object Tracking System is used to track the motion trajectory of an object in a video. First, I use the OpenCV's function, select ROI, to select an object on a frame and track its motion using a built-in-tracker. Next, Instead of using select ROI, I use YOLO to detect an object in each frame and track them by object centroid and size comparison. Then I combine YOLO detection with the OpenCV's built-in tracker by detecting the object in the first frame using YOLO and tracking them using select ROI. Video tracking is widely used for multiple purposes such as human-computer interaction, security and surveillance, traffic control, medical imaging, and so on.**

**Keywords- OpenCV, YOLO, object tracking, centroid tracking, Frame Differencing, Single shot detector, Background subtraction.**

## I. INTRODUCTION

To gain a complete image understanding, we should not only concentrate on classifying different images but also try to precisely estimate the concepts and locations of objects contained in each image. This task is referred to as object detection, which usually consists of different subtasks such as face detection pedestrian detection, and skeleton detection. As one of the fundamental computer vision problems, object detection can provide valuable information for semantic understanding of images and videos and is related to many applications, including image classification, human behaviour analysis, face recognition, and autonomous driving.

Meanwhile, inherited from neural networks and related learning systems, the progress in these fields will develop neural network algorithms, and will also have great impacts on object detection techniques which can be considered as learning systems. In an artificial vision, the neural convolution networks are distinguished in the classification of images.

The problem definition of object detection is to see where objects are set in an exceedingly given image (object localization) and which class every object belongs to (object classification). so the pipeline of Old object detection models are often primarily divided into 3 stages: informative region selection, feature extraction and classification.

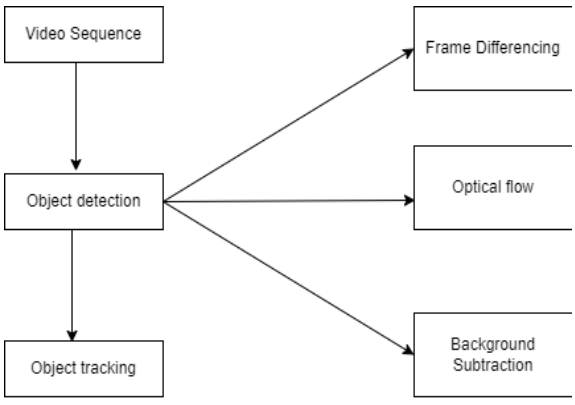


Fig 1.Basic block diagram of object detection

Fig. 1 shows the essential diagram of detection and pursuit. In this paper, SSD and Mobile Nets-based algorithms are enforced for detection and tracking in a python setting. Object detection involves detecting the region of interest of an object from a given class of image. Completely different strategies are –Frame differencing, Optical flow, Background subtraction. This is a technique of detecting and locating an object that is in motion with the assistance of a camera.

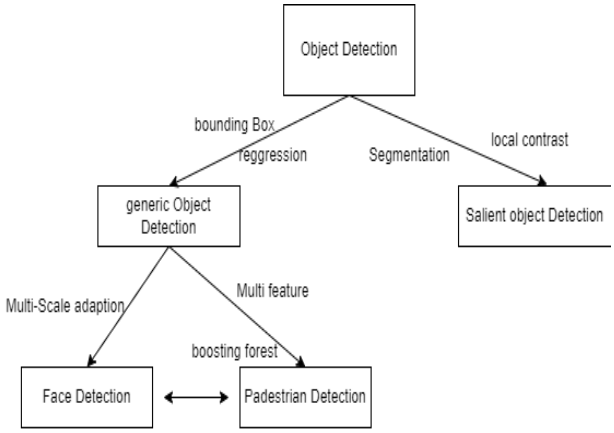


Fig2.The Application domains of object detection.

In this paper, a structured review is provided to summarize representative models and their completely different characteristics in many application domains, as well as generic object detection, salient object detection, face detection, and pedestrian detection Their relationships are pictured in FIG 2.

Based on basic CNN architectures, generic object detection is achieved with bounding box regression, whereas salient object detection is accomplished with local distinction improvement and pixel-level segmentation. Face detection and pedestrian detection are closely associated with generic object detection and are primarily accomplished with multi-scale adaption and multi-feature fusion/boosting forest, respectively.

The dotted lines indicate that the corresponding domains are related to one another under certain conditions. It should be noticed that the lined domains are diversified. Pedestrian and face pictures have regular structures, whereas general objects and scene pictures have more complicated variations in geometric structures and layouts. Therefore, completely different deep models are needed by various pictures.

## II. OVERVIEW OF OPENCV AND OBJECT TRACKING

**1. OpenCV**

The Open-source computer Vision (OpenCV) library is an open source cross-platform computer vision and machine learning software library. It absolutely was originally developed by Intel to advance CPU-intensive applications in 2000. a number of its initial goals still hold nowadays, like providing optimized code for basic computer vision infrastructure, spreading data to create applications quicker, and providing portable, performance-optimized code.The library is authorized with open-source BSD license and may be used for tutorial and business [1] applications. nowadays it's around 2,500 optimized algorithms used to discover and acknowledge human faces, determine varied objects, classify human actions in video, track moving objects, extract 3D models of objects, etc. the most recent stable unleash was revealed in april 2020 as 4.3.0 version.

**2. Object Tracking:**

The goal of object tracking is to estimate the state of the chosen object within the subsequentframes [2]. The thing being tracked is sometimes marked using a rectangle to point its location within the starting1 frame. Once there aren't any changes within the surroundings, object tracking isn't too complex;however this can be seldom the case. Numerous disturbances are a normal occurrence within the universe. These disturbances may embody occlusion, variations in illumination, amendment of viewpoint, rotation, blurring because of motion, etc. The task of designing a robust and efficient hunter is thought to be a really difficult task.

**3. Object tracking algorithms in OpenCV:**

The OpenCV library includes eight algorithms for object tracking, which are available through OpenCV tracking API. Table I provides some information about the available algorithms in the OpenCV library with their publication years and reference to research papers detailing their implementation.

In general, tracking an object in the video involves steps such as:

- Choosing the tracker,
- Selecting the object (target) from the starting frame with the bounding box,
- Initializing the tracker with information about the frame and bounding box, and
- Reading the remaining frames and finding the new bounding box of the object. The last step is usually implemented in the loop.

An OpenCV tracker consists of three main components, which also coincide with the components in a typical tracking algorithm [3]:

**3.1 Tracker Feature Set:**(The model of the target object's visual appearance): used to represent objects of interests. In OpenCV, possible features can be extracted with HAAR, HOG, LBP, Feature2D, etc. Many other global and local features exist in literature.

**3.2 Tracker Sampler Algorithm:**(The mechanism for matching model parts to image regions at each frame): computes the patches over the frame based on the last target location.

**3.3 TrackerModel:**(The mechanism for continuously relearning or updating models of targets which change their appearance over time): internal representation of the target. It stores all state candidates and computes the trajectory. TrackerFeatureSet and Tracker Sampler Algorithm are the visual representation of the target,

while the Tracker Model represents the statistical model.

## III. REVIEW OF LITERATURE

Object tracking has more technical unified areas within the background subtraction methods. There are some works mentioned for frame differencing that use the pixel-wise differences between two frame pictures that are background subtraction for detecting moving regions and background model for object detection by a Gaussian mixture model.

**Lipton et al. [4]** have proposed frame variations for using pixel-wise variations to get the motion objects. In another work, Stauffer and Grimson et al. [5] has projected a Gaussian mixture model on the idea of a background model to spot the thing.

**Liu et al. [6], [7]** have projected background subtraction to detect the motion of objects in a picture by getting the distinction between current and reference background pictures in a pixel-by-pixel.

**Desa and Salih et al. [8]** has projected and impermanenteach background subtraction and frame distinction.

**Sungandi et al. [9]** projected and introduced object detection using frame distinction in low-resolution pictures.

**Jacques et al. [10]** have projected a brand new background model and shadow detection in grayscale video clips.

**Satoh et al. [11]** projected a brand new concept for object tracking using the PISC image-based block matching rule.[12]

**Sugandi et al.**projected tracking techniques for moving individuals using a camera peripheral signalling contact image. Authors projected in stereo camera-based object tracking, used Kalam filter to predict the position and speed of objects within the x-2 dimension that proposes application of extended Kalam filter to calculate the 3D path of an object from 2nd motion.

## IV. DESIGN AND DEVELOPMENT

The most inventive and challenging section of the life cycle is system and style. The term style describes a final system and the method by that it's developed. [13] It refers to the technical specifications that may be applied in implementation the candidate system. The planning could also be outlined as "the method of applying varied techniques and principles for the aim of defining a device, a method or a system in spare details to allow its physical realization".

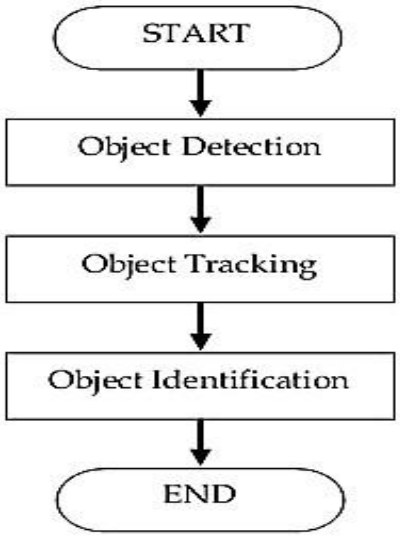


Fig 3. Flow of Procedure.

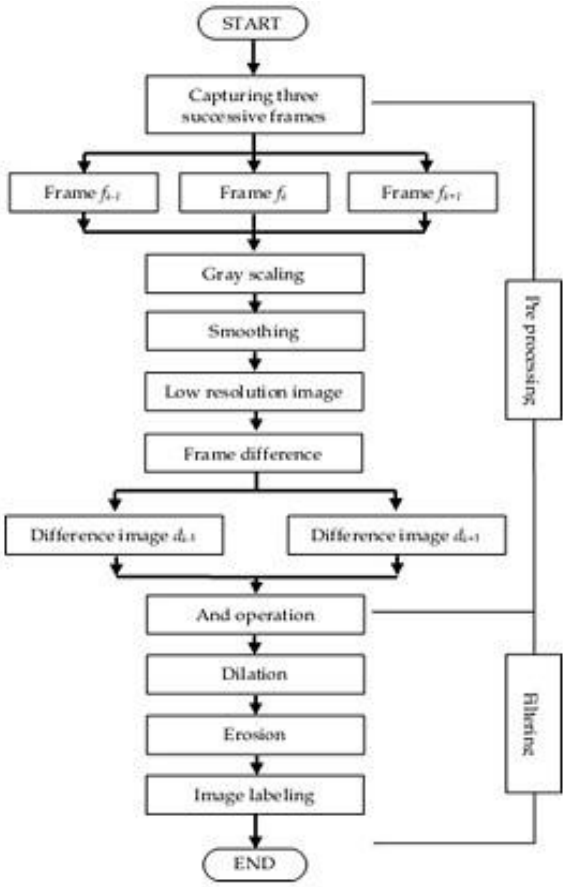


Fig 4. Flow of Object Detection.

## V. MODULES

Implementation is that the stage of the project once the theoretical style is clad into a operating system. Therefore it is thought of to be the foremost important stage in

achieving a successful new system and in giving the user, confidence that the new system can work and be effective. The implementation stage involves careful planning, investigation of the prevailing system and it's constraints on implementation, designing of strategies to achieve changeover and analysis of changeover strategies.

## VI.OBJECTIVE

This article aims to boost the performance of object detection and tracking by increasing the speed of detecting objects and therefore the accuracy of moving objects. At present, it detects only proper objects present in a picture, if the objects are smaller or hidden, then it's troublesome to notice.

To overcome this drawback, we decide another technique referred to as faster R-CNN that identifies every and each object even if it's smaller;however its accuracy of identifying is slow compared to YOLO. Though, it doesn't acknowledge hidden objects, it acknowledges all types of object and also, has to improve the segmentation. If speed is totally paramount, then we use YOLO.

The major objectives are:

- Analysing grid images at a higher speed.
- Analysing tracking methods for detecting hidden or smaller objects in an image.

## VII. RESEARCH PROBLEM

The problem with computer vision, image processing, andmachinevisionis that it determines whether or not the image data has a specific object, feature, or function.One of the major problems was the image classification. Image classification involves labelling an image based on the content of the image.

For example, the objects in a particular image such as tray, fork, spoon, etc. It is only detected by fork and spoon, but not by the tray because inside the data set, tray keyword is not inserted. Therefore, the image classification will not be classified.

## VIII. SOLUTION METHODOLOGIES

There are many ways to detect objects. The best way for detecting object methods are convolutional neural network, RCNN, Fast-RCNN, Faster-RCNN, YOLO, OpenCV, etc. In this project, we use YOLO and OpenCV method for object detection which detects each and every object clearly. The last step is to have the boundary boxes and labeled images. [14]

It is easy to understand and consumes less time to detect the object. In table below it says about how the objects are detected and check the process goes through to detect an object.There are four steps to follow to get a proper object detected image. [15], [16] the steps involved in this method are as follows:

| Steps | Description |
|---|---|
| Step I | Consider an image and we need to create a grid that will give us the features of an object. |
| Step II | In this step, we make using of OpenCV which will read the input image and data points and specify the file path to an image in a Numpy array. |
| Step III | Detecting an image in a grid view after the process of reading image by OpenCV and Numpy and converting grid to rectangular boxes. |
| Step IV | The final step consists of displaying the image with the rectangular box along with the caption on the window. This is done using YOLO and COCO dataset. |

## IX. LIMITATIONS OF EXISTING SYSTEM

Here we discuss the limitations of some of the existing models:

**1. R-CNN (2014):**
It takes more time to train the network as we have to classify 2000 region proposals per image [17]. It cannot be implemented in real-time because it takes around 47 seconds for each and every test image. R-CNN was introduced in the year of 2014 combining region proposals with a CNN. Major drawbacks are that it was slow, hard to train, and consumes large memory. This method uses a selective search to generate regions and to detect the object.

**2. ResNet (2015):**
It is the most powerful deep neural networks which have achieved state-of-the-art performance on the ILSVRC 2015 classification challenge [18]. The first implemented was the VGG network. Suppose the input size is given as 300 and 320, even though the ResNet-101 layer is deeper than the VGG-16 layer, the main disadvantage is that it decreases the accuracy. The main thing is that it has the capacity to undergo a deeper layer. Deshpande et al. / Int. J. Res. Ind. Eng 9(1) (2020) 46-64 52

**3. Fast R-CNN (2015):**
Fast R-CNN uses a single method that extracts features from the regions, then divides them into different classes, and returns boundary boxes for the identified classes. It takes time to concentrate on increasing accuracy and decreasing time. This method was introduced in the year of 2015 [19]. It has high accuracy compared to the previous method and detects the object in a faster way.

**4. Faster R-CNN (2015):**
In the above methods, a selective search is used to detect the region proposals, which is time-consuming and slow in the process. To overcome these problems, a new method was introduced, i.e.RoI Pool layer which is used to classify the image within the proposed region and can find the values for the boundary boxes. This method also consumes time and detects smaller or hidden objects that introduced in the year of 2015 [20].

**5. YOLO (2015):**
YOLO is a single-stage detector. The first breakthrough was in 2015 by Redmon et al. [21], it detects the real-time object and is very fast, better, and stronger compared to other methods. Its accuracy is very high. It has a COCO dataset to store the data of images and videos and has excellent results on the COCO dataset. YOLO helps to detect moving objects, recognizes and helps to display in a rectangular bounding box with a provided caption.[22] The major advantage is that the fast-moving objects are captured very quickly compared to the rest of the methods. This method is mainly used for speed. It is faster compared to any other method.

## X. RESULT AND DISSCUSSION

Fig 5. Before detection [61].

In the above-inserted image to the algorithm, we expect to detect the objects and label them.

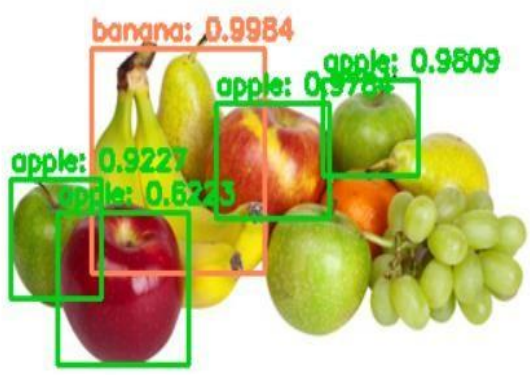


Fig 6. After detection.

Table 1. Comparative analysis based on time.

| Methods | Fast R-CNN | Faster R-CNN | YOLO |
|---|---|---|---|
| Time (Sec) | 41.64326 | 38.53267 | 1.929929 |

Discussion based on Figures (5)-(6), YOLO not only detects bags and person but also detects fruits. In the above image, the detected object is banana and apple. As expected, we have an output with the labeled object. Finally, the objects are correctly detected. This takes less time to detect a particular object.

Because of a number of objects are less and the size of the object as well, so, it has detected in a lesser time compared to the two methods mentioned in Table 1. Detecting this kind of images helps children to recognize fruits along with names, also helps to learn in a faster way.

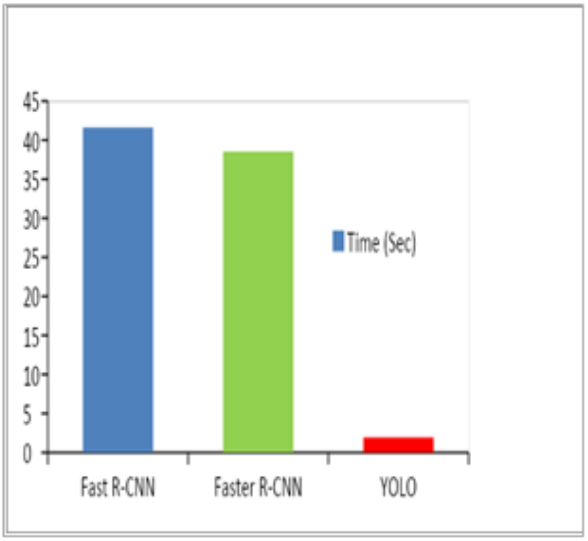


Fig 7. Time is taken to detect.

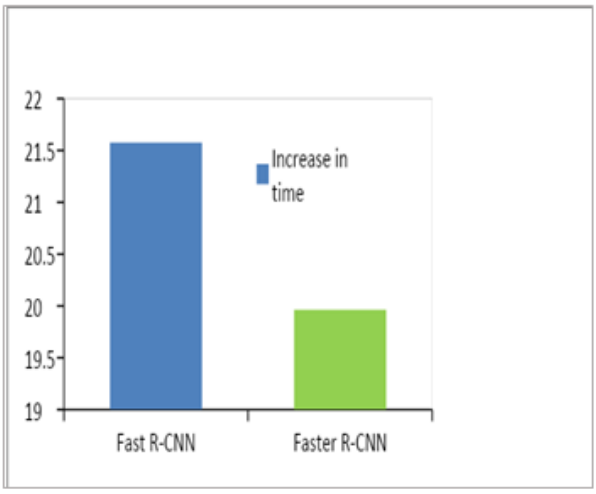


Fig 8. Comparison of time.

As we see in the graph (Figure 7), the Fast R-CNN is 21 times slower than YOLO and Faster R-CNN is 19 times slower than YOLO and another observation is the Faster R-CNN is better than Fast R-CNN but not better than YOLO (see Figure 8).

Fig 9. Before detection [62].

The above image inserted into the algorithm. We expect the algorithm to detect, identify, and label them according to the class assigned to it.

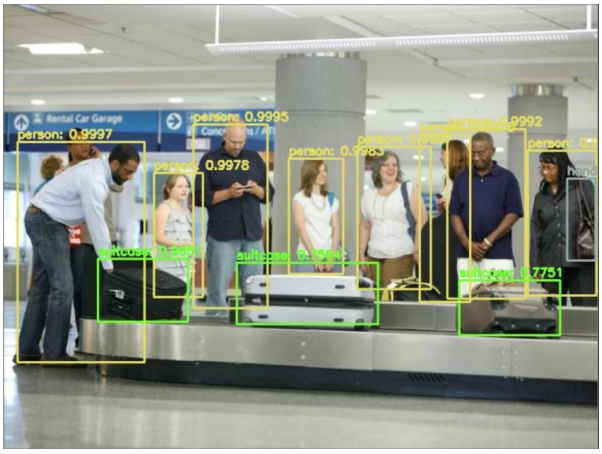


Fig 10. After detection.

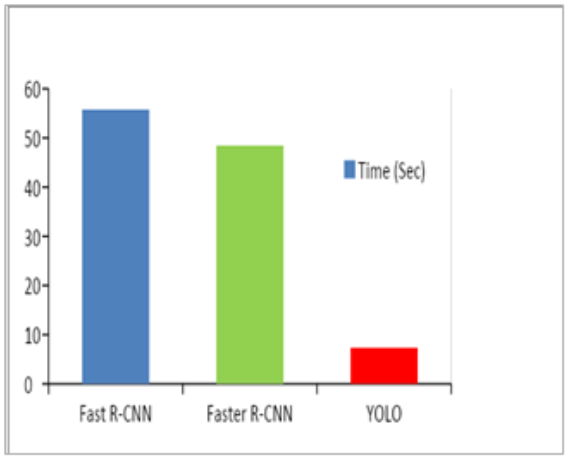


Fig 11. Time is taken to detect.

Table 2. Comparative analysis based on time (example 2).

| Methods | Fast R-CNN | Faster R-CNN | YOLO |
|---|---|---|---|
| Time (Sec) | 55.769754 | 48.45637 | 7.366887 |

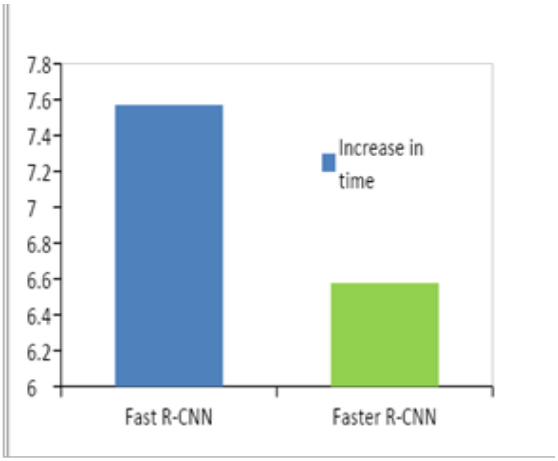


Fig 12. Comparison of time.

Discussion based on Figures (9)-(10), YOLO with OpenCV has detected person and suitcase. In deeper observation, we can also see there's another object detected at the right corner of an image, i.e. handbag. It has consumed more time compared to Figure 6, because the size of the image and the number of objects are more so, it took more time compared to the other two methods as shown in Table 2. This kind of detection helps people to find their lost bags or kids in an airport or any other places and also was used to obstacle avoidance.

As we see in the graph (Figure 11), YOLO is 7 times faster than Fast R-CNN and 6 times faster than Faster R-CNN and also Faster R-CNN is faster than Fast R-CNN but not faster than YOLO (see Figure 12).

Fig 13. Before detection [63].

Above image is a live Bangalore traffic signal which consists of all kinds of vehicle, we expect the algorithm to detect each and every vehicle and label them.

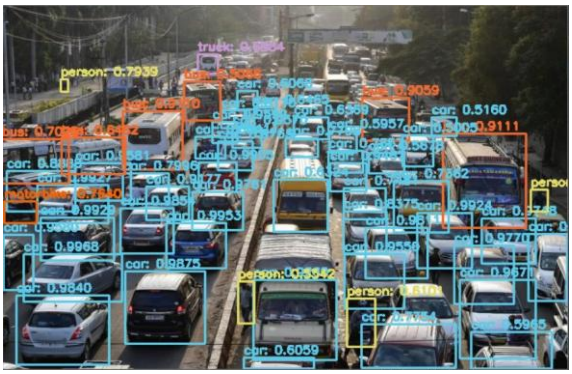

Fig 14. After detection.

Table 3. Comparative analysis based on time.

| Methods | Fast R-CNN | Faster R-CNN | YOLO |
|---|---|---|---|
| Time (Sec) | 83.81045 | 73.88907 | 19.389161 |

Discussion based on Figures (13)-(14), YOLO is able to correctly detect each car, bus, motorbike, truck, and person shown in an image as expected. We can notice that person at the right corner detected is slightly blurred and partially obscured that's a positive point of this method. Also in some parts of the image, persons are not detected because it's very small to detect this YOLO method. It struggles to detect small objects, which need to be improved.

This method helps detecting an object at high speed compared to the other two methods mentioned in Table 3. The use of detecting this inserted image helps to control the traffic signal. This image has consumed more time compared to Figure (6)-(10). Because of there are many objects seen on image and also it has to detect smaller objects, it takes time to understand and computes to give accurate results.

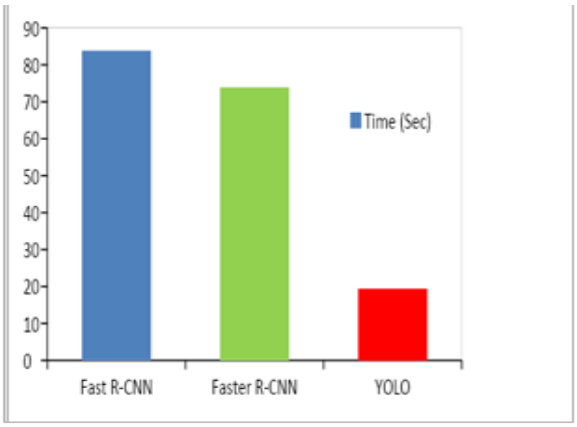


Fig 15. Time is taken to detect.

As we see in the graph (Figure 15), YOLO is 4 times faster than Fast R-CNN and 3 times faster than Faster R-CNN and also Faster R-CNN is better than Fast R-CNN but not better than YOLO and (see Figure 16).

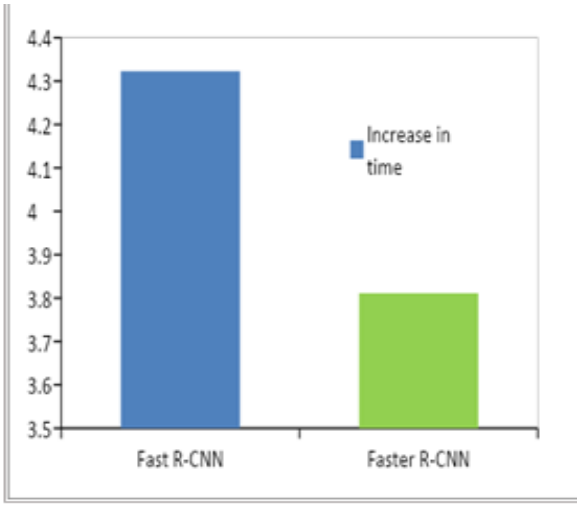


Fig 16. Comparison of time.

The final decision is, as the size of the image and number of objects increases, the time taken also increases and vice versa. Therefore, the number of objects is directly proportional to the time taken in detecting those objects.

## XI. CONCLUSION

Based on test results, the object can be detected more accurately and individually identified with the exact location of an object in the image along with the x and y-axis. This paper provided experimental results on different methods for object detection and identification and compared each method for their efficiencies, also performed efficient object detection while not compromising on the performance.

Objects are detected using SSD algorithm in real time scenarios. Additionally, SSD have shown results with considerable confidence level. Main Objective ofSSD algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance.

This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise. Further extending the work to detect ammunition and guns in order to trigger alarm in case of terrorist attacks. The model can be deployed in CCTVs, drones and other surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted

## REFERENCES

[1] "A brief introduction to OpenCV | Request PDF." https://www.researchgate.net/publication/261424692_A_brief_introduction_to_OpenCV (accessed Apr. 05, 2022).

[2] Y. Wu, J. Lim, and M.-H. Yang, "Object Tracking Benchmark", doi: 10.1109/TPAMI.2014.2388226.

[3] A. Brdjanin, N. Dardagan, D. Dzigal, and A. Akagic, "Single Object Trackers in OpenCV: A Benchmark," INISTA 2020 - 2020 International Conference on INnovations in Intelligent SysTems and Applications, Proceedings, Aug. 2020, doi: 10.1109/INISTA49547.2020.9194647.

[4] E. Su, L. Chen, Y. Xu, and B. Hu, "Target detection in NAO robot golfing," Journal of Physics: Conference Series, vol. 1828, no. 1, Mar. 2021, doi: 10.1088/1742-6596/1828/1/012171.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246–252, 1999, doi: 10.1109/CVPR.1999.784637.

[6] H. Deshpande, A. Singh, and H. Herunde, "Comparative analysis on YOLO object detection with OpenCV," International Journal of Research in Industrial Engineering, vol. 9, no. 1, pp. 46–64, Mar. 2020, doi: 10.22105/RIEJ.2020.226863.1130.

[7] Y. Liu, H. Ai, and G. Xu, "<title>Moving object detection and tracking based on background subtraction</title>," Object Detection, Classification, and Tracking Technologies, vol. 4554, pp. 62–66, Sep. 2001, doi: 10.1117/12.441618.

[8] S. M. Desa and Q. A. Salih, "Image subtraction for real time moving object extraction," Proceedings - International Conference on Computer Graphics, Imaging and Visuualization, CGIV 2004, pp. 41–45, 2004, doi: 10.1109/CGIV.2004.1323958.

[9] B. Sugandi, H. Kim, J. K. Tan, and S. Ishikawa, "Real Time Tracking and Identification of Moving Persons by Using a Camera in Outdoor Environment," International Journal of Innovative Computing, Information and Control ICIC International c, vol. x, 2008.

[10] "Temporal Feature Warping for Video Shadow Detection | DeepAI." https://deepai.org/public ation/temporal-feature-warping-for-video-shadow-detection (accessed Apr. 05, 2022).

[11] Y. Satoh, S. Kaneko, and S. Igarashi, "Robust object detection and segmentation by peripheral increment sign correlation image," Systems and Computers in Japan, vol. 35, no. 9, pp. 70–80, Aug. 2004, doi: 10.1002/SCJ.10241.

[12] H. Deshpande, A. Singh, and H. Herunde, "Comparative Analysis on YOLO Object Detection with OpenCV", doi: 10.22105/riej.2020.226863.1130.

[13] Z. Tang, K. Liu, Z. Yang, Z. Pei, and Z. Zhang, "Object Tracking System for Video Recording based Qt and OpenCV," 2016.

[14] A. Brdjanin, N. Dardagan, D. Dzigal, and A. Akagic, "Single Object Trackers in OpenCV: A Benchmark," Aug. 2020. doi: 10.1109/INISTA49547.2020.9194647.

[15] P. Janku, K. Koplik, T. Dulik, and I. Szabo, "Comparison of tracking algorithms implemented in OpenCV", doi: 10.1051/04.

[16] S. P. Singh, A. Mittal, M. Gupta, S. Ghosh, and A. Lakhanpal, "Comparing Various Tracking Algorithms In OpenCV," 2021.

[17] "CVPR 2014 Open Access Repository." https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html (accessed Apr. 05, 2022).

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", Accessed: Apr. 05, 2022. [Online]. Available: http://www.image-net.org.

[19] "CVPR 2014 Open Access Repository." https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html (accessed Apr. 05, 2022).

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Accessed: Apr. 05, 2022. [Online]. Available: https://github.com/

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 779–788, Dec. 2016, doi: 10.1109/CVPR.2016.91.

[22] H. Vadlamudi, "Evaluation of Object Tracking System using Open-CV in Python." [Online]. Available: www.ijert.org