GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

# OBJECT RECOGNITION USING DEEP LEARNING

**A Report for the Evaluation 3 of Project 2**


*Submitted by*

**MAYANK MATHUR**

**(1613107029)**


***in partial fulfilment for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION OF BUISNESS ANALYTICS AND OPTIMIZATION**


**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**Under the Supervision of**

**U SAMSON EBENZAR**

**Asst.Professor**


**APRIL / MAY- 2020**

# SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

## BONAFIDE CERTIFICATE

Certified that this project report **"SOME PERFORMANCE ASPECTS OF OBJECT RECOGNITION USING DEEP LEARNING"** is the bonafide work of "**MAYANK MATHUR (1613107029)"** who carried out the project work under my supervision.

**SIGNATURE OF HEAD**

Dr. MUNISH SHABARWAL,

PhD (Management), PhD (CS)

**Professor & Dean,**

**School of Computing Science &**

**SIGNATURE OF SUPERVISOR**

Dr. SANJEEV KUMAR PIPAL, M.Tech., Ph.D.,

**Professor**

**School of Computing Science &**

**Abstract:**

In this ever-growing world there is a direct relationship between video analysis and image identification. This has been a debatable topic in the recent years. As we discuss about the machine learning techniques, we find that deep learning as a part of machine learning has been drastically increasing and demanding over the years. As there is a development in the deep learning methods , impactful tools such as semantic analysis comes into play. This analysis helps us to analyse the given video medium and aggregates it with the object detectors to better depict the object and classify its class. A classifier in the deep leaning promotes and makes this image classification strategy very easy to understand. . These models have evolved differently in network architecture, training strategy and optimization function.

This paper has focused more towards deep learning-based object detection frameworks. The review begins with the deep learning tool namely Convolution Neural Network (CNN). In this survey, we focus on the specific tasks such as object detection and face detection.

## TABLE OF CONTENTS


| CHAPTER NO. | TITLE | PAGE NO. |
|---|---|---|
| 1 . | Introduction | 1 |
| | 1.1 Purpose | |
| | 1.2 Scope | |
| 2 . | Literature Survey | 5 |
| | 2.1 Quantitative Analysis on Reviews<br>2.2 Problem Statement<br>2.3 Proposed Model<br>2.4 Result | |
| 3 . | Conclusion/Future Enhancement | 17 |
| 4 . | References | 19 |

## 1. Introduction:

An object detection is a computer vision system in which the objects can easily be located in images or videos. It classifies semantic objects and helps us to predict the behaviour of the pattern. It uses semantic approach to better visualize, how the group of small cells develop a pattern and based on this, it predicts the object of that class. Ex: Humans ,cars etc.

Object detection system usually finds objects and this task is very difficult for machines as compared to humans who perform OD very effortlessly and instantaneously. We can picturize this scenario by looking into the basic of how we input any image and thus get a Yes/No output from the OD system. This review basically aims to detect the objects that are stored in the input image and processes the object using image classifier to get the desired output.

Here, we distinguish the objects based on the labelling approach. Given that the input image consists of one or more objects. The task of image classifier is to create a set of labels pertaining to a corresponding set of models know by the system. The problem constraint of object detection is determined where objects are detected in a given image i.e. object localization and which category do each object belongs to i.e. object classification.

### 1.1 Purpose of analytics:

As a shopper and enterprise object and face detection solutions proliferate. The OD system is currently gaining support as a potential tool for averting terrorist crimes, object detection is used in many law enforcement areas. This system has also been developed for computer vision and neural network systems. This helps us to increase its demand for user verification and systems.

The performance of the object detection method can be better calculated with the help of precision and recall analytics i.e. Box AP calculation.

Most organization uses this technique for verification and validation purposes. It also involves detecting objects from a particular class. The OD systems constructs a model from a set of trained datasets. Here comes the utility of supervised learning techniques

**1.2 Scope:**

The OD system is a computer application designed for automatically identifying or verifying a person from a digital image or a video frame. One of the ways to do this is by comparing selected facial features from the trained data set . It is typically used in security systems and can be compared to other biometrics like fingerprint scanning or eye iris recognition systems. Some facial recognition algorithms are identified by facial features extraction , from an image of the subject's face. Let us take an example of an algorithm that might analyse the relative position, size, and shape of the eyes, nose, cheekbones, and jaw. These features are used to search for other images with matching features. Other algorithms have normalized and stored a huge number of face images and then compress the face data, only saving the data in the image that is useful for face recognition

Scope in India:

1)To prevent the frauds of ATM in India, it is highly recommended to prepare the database of all ATM customers with the banks in India and deployment of high-resolution camera and face recognition software at all ATMs. So, whenever user will enter in ATM his/her image will be taken to permit the access after it is being matched with stored photo from the database.

2) Duplicate voters have been reported in India. To prevent this cause, a database of all voters, including all the constituencies, is recommended to be prepared. At last, at the time

of voting, a higher resolution camera and face recognition device are equipped of voting site, should accept a subject face and generate the recognition for voting only if the match is found.

3) Passport and visa verification can also be done using face recognition technology, which thereby helps in criminal check.

4)Driving license verification can also be performed with the help of face recognition technology.

## 2. Literature survey:

With the effect to understand the model, we have systematically reviewed, we have studied and presented a systematic analysis of how the OD system works, which include object recognition techniques from various digital libraries (NPTEL, Google Scholar) and many more. We studied the overall description and problem statement of each paper. We have stipulated the most suitable model for the review. Our review is based on moving as well as static objects depending on the state of video stream. We have also formulated the basic research questions which makes this paper more precise.

1) To identify the popular techniques used for object detection?
2) To analyse the issues related to the upcoming techniques of object recognition?
3) To analyse the future scope of research in Object Recognition?

After studying we found that the process of recognition is a huge task for computer vision systems. A few research papers provide that accuracy and speed is the key to the above problem. The possible approach to merely, find a solution is by using the FSVM i.e. Support Vector Machine technique. It deals with the analysis of classifying the images

using classification and regression problems. The two approaches necessary for training data sets are:

1)Learning Phase and

2)Testing Phase

## 2.1 Quantitative Analysis on Reviews:

The review from Zhong and his members have focused and dealt with different sub problems related to clutter and low resolution. However , they have eliminated the problem by use of different degrees of advanced R-CNN. They have proposed a model solution which focuses on general problems related to object detection like object detection pipelines , providing many different tasks in a single system. The three amin object detection techniques used in this paper are:

1. Object Detection techniques and methods.
2. Face detection checker and
3. Pedestrian object detector.

They have also seen to the issue and have understood the object detection landscape to visualize the concepts and evolutionalized the rectangular Box AP to detect the objects.

A review on the topic of Object detection has also been carried out by Prasad. They have together discussed the solution to the problems pertaining to detect objects in real time environment such as real images. They have also studied various aspects such as feature types, learning model, object templates, matching schemes and boosting methods , which considered the problems of object detection in remote sensing objects and have understood and evaluated the concepts of pre-segmentation of object detection in real time systems.

Many research papers and surveys on object detection have been published and recognized over the past few years. The problems linked with theses research papers are:

1. Object localization: It helps to determine and localize the shape of the object, thus creating a single object instance, which is referred to as a static image.
2. Object classification: It helps to determine and analyse that only one object of a given instance is present in an image.
3. Object recognition: It helps to determine whether a particular instance of object is present in the image.
4. Post-Estimation of image: It determines the view of the object and helps create an area of localization of the object.

The topic of object presence classification mentioned earlier could be explained by using object detection methods. On the other hand , various methods like localization of object and measurement of the scale of objects are required with proper precision and accuracy of maintaining and obscuring objects , so that the instance of objects are easily detectable.

Also, object recognition technique could be solved using different techniques which does not require detecting the object in advance.

Thus, by solving OD problems we can provide the solution for image database problems. Also, determining the instance of an image patch, i.e., evaluating the likeliness , even seemed to be a new and additional problem for an image window to contain an object of any class. Most of the OD's are based on statistical classifiers, which have set a ground and solved most of the issues related to the following research in terms of training and evaluation procedures and classification techniques.

Many OD techniques have considered to follow the preliminary approach towards solving any type of problem. In OD, these problems constitute and is commonly called the sliding window problem. In order to check and localise the objects appearing on the screen , we used different scales and locations to perform an exhaustive search.

In order to search for an image, we use classifier , which helps to identify in a given patch of image , a trace of an object. In most of the OD problems, the classifier works at a level to measure a given scale size and patch size. There are many versions of the input image i.e. there are many sources from where an image can be extracted. Thus , we use classifier to classify all possible patches of that particular size , for each image extracted from the source.

## 2.2 Problem Statement:

Object detection is rapidly finding its importance in computer networks. The ongoing studies are keener towards finding facial expressions, walk patterns of humans , cars and many various other objects. The fuzzy support vector machine (FSVM) is an excellent method used in our paper for detection of objects over patches of images. We input the data from a sample set consisting to as a training or sample data. We have used the FSVM classifier to separate the objects and mark the corresponding patterns same in behaviour as one, and groups it with the one having and resembling same characteristics and pattern.

Object Detection Algorithm:

We have trained and labelled a set of sample data associated with a fuzzy membership called the localising function f(x).

$$(a_1 ,b_1 , c_1 ),\ldots (a_i ,b_i , c_i ),\ldots\ldots (a_n ,b_n , c_n )$$

$a_i \in R^d$ ,i=1,2,3,…n belongs to either of two classes given a label $b \in \{1,1\}$ and a fuzzy membership $c_i$ . The fuzzy membership $c_1$ is called the attitude of the corresponding point $a_i$ towards one class.

Let $z = \psi(x)$ corresponds to the feature space vector mapping with the same space vector Z. A solution provided to this problem is done by mapping the input space into a higher dimension feature space . at last, we search for the optimal result of the hyperplane in this feature space.

$\min \frac{1}{2} \|w\|^2 + C \sum_{j=1}^{n} \tau_i \in_I$ ……..(1)

Subject to $b_i(\omega_I * a_i + d) >= 1-\in_I$ j=1,2,3,…n …….(2)

$\in_I >= 0$……..(3)

The decision function becomes:

$f(x)=sgn(\sum_{i=1} \alpha_1^* b_i \, l(x_i, x) + d^*)$……..(4)

where $d^*$ corresponds to the threshold value, which can be obtained by any support vector $a_i$. Since the corresponding $\alpha_i$ is null if $x_i$ is not the support vector, the decision function above is actually only for support vectors.

The problem statement of object detection algorithm goes as below:

The input images extracted from the sample space can be divided into various sub images. We select a handful of images from the sub images and name it under the “object” template.

1. At first, we select those sub images which are directly mapped with the “object” template. We disclose this group of sub images and place the results of the matched “object” samples into one.
2. The object’s which are not detected are selected from the “object” samples and then these “object” samples are aggregated as one and is sent into the FSVM classifier for training the test data available.
3. This FSVM classifier thereby removes the duplicate “object” samples from the training data and compiles the result of the final template matching.
4. Only at last, the FSVM classifier confirms the presence of these object samples and corresponding patches of the images required to be extracted from the video are marked with rectangular sliding windows regarded to as BOX AP. Thus , each sub image is classified by using the template matcher and FSVM classifier.

The three broad steps used in this algorithm are mentioned below:

1. Mapping of templates: It has larger computational power and complexity. This is because , it constitutes larger dataset of images stored in the training data set. Thus, to compute the matching we use pyramid algorithm for better and fast matching of images over other sub images. We can also use multi- scale approach to solve fixation of template scale and linear resampling.
2. Classifier for advanced approach (FSVM): The object samples sustained in the training and test dataset are collected and passed through the FSVM classifier to limit the training space and thus reduce the blindness and keeps the resolution of image higher. It however helps us to improve the classification efficiency and reduces the training complexity.

3. The final step of this algorithm is to mark the matched sample of images with and optimal rectangle called the BOX AP , where objects are visible and the result is based on FSVM classifier.

### 2.3 Proposed Model:

The Key issues in implementing the OD model arises with a set of questions:

1) Whether the whole image have to be scanned for locating the objects?
2) Is accuracy and speed to locate the objects is maintained?
3) Should we combine all the classifiers?
4) How could we select unbiased set of classifiers in SVM?
5) Is performance of the OD maintained?
6) How different positions of an object could be handled and presented as a single entity?

How the OD system works could be better understood by looking at figure1 which shows the steps involved to transform the image into identification. The basic aim of this model is to detect instances of class, present in the images to categorize various objects into respective object classes.

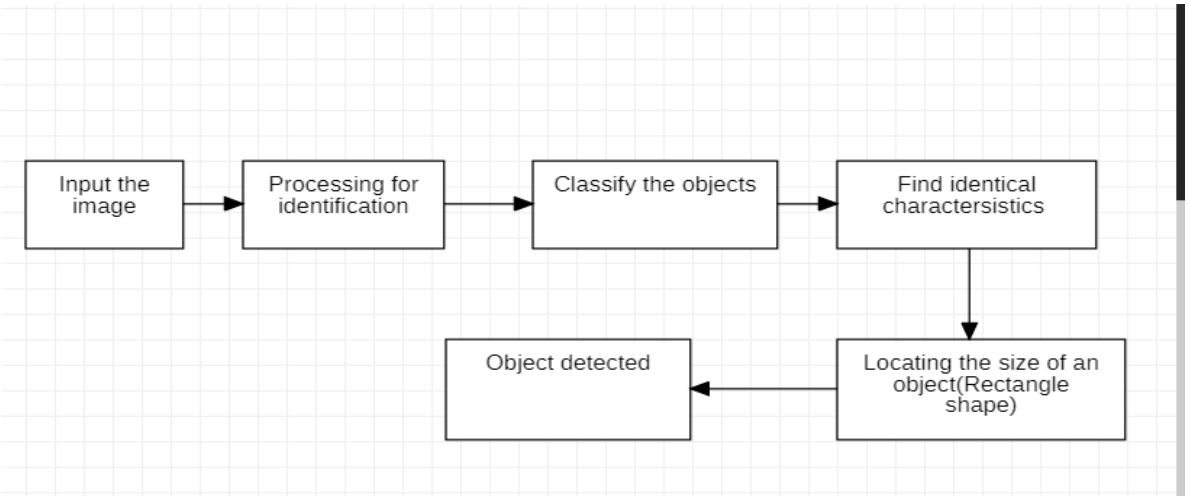



Figure 1

Figure 2 depicts how the OD system works when implemented in real world. It shows that we require trained data sets to study and identify the objects. We thus require an object detector  to compare the behaviour of different objects and to thus categorize them based on the similarity of their patterns.

The object template block then makes use of the learning's that were done previously to represent the objects with various representations like histogram representation, random forest representation, etc. Whereas on the other hand, learning through validation block does not require any sort of training as they are validated beforehand.

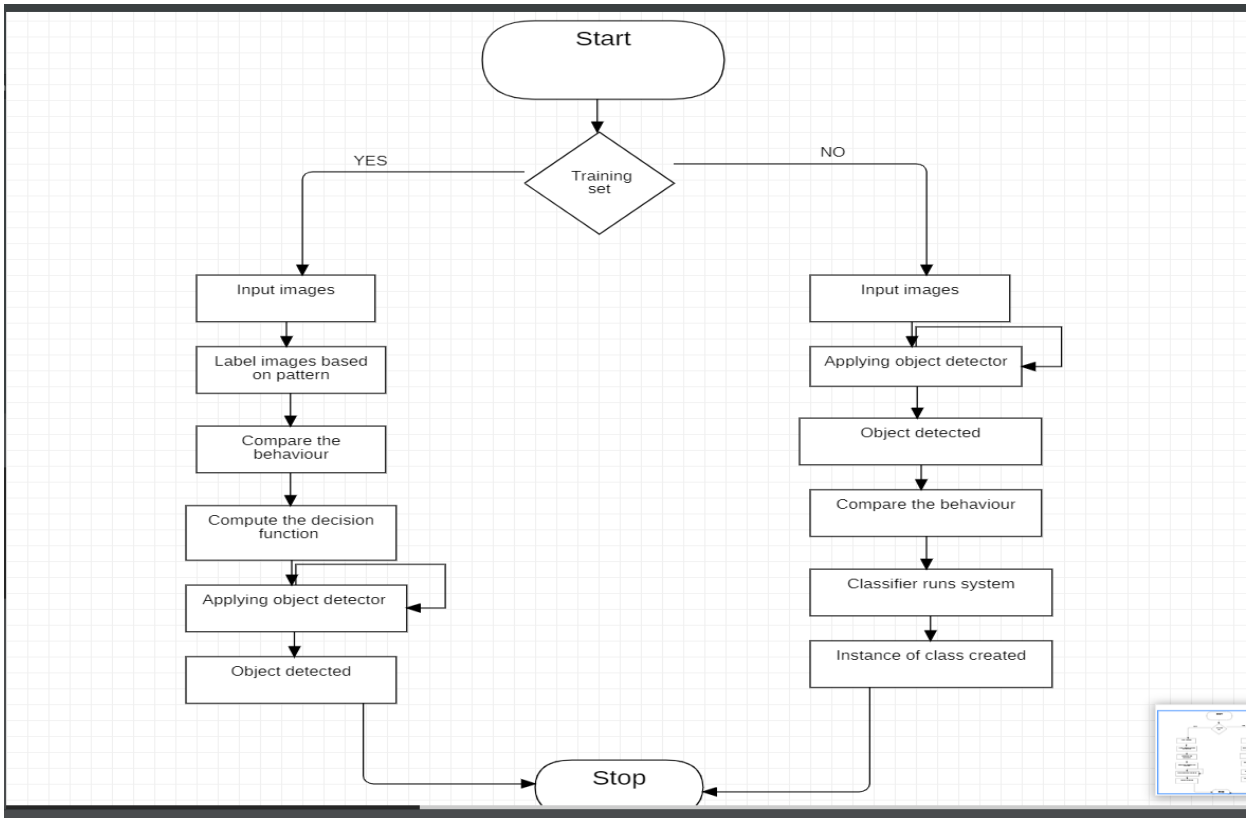



Figure 2

## 2.4 Result:

In the world of ever-growing data , it is difficult to manage a lot set of data as input image. There are situations in which it is difficult for computers to predict and filter the right set of data. The problem arises when we deal with the images of movements in space or the movement of a person roaming about , the images of same object look differently. On the other hand , many devices have been designed that are capable of capturing images significantly.

This literature identifies that accuracy and time of travel ate the main parameters to predict the nature of OD. A few research papers provide a balancing solution between accuracy and time. The possible approach to this solution can be designed by using the approach of finding a minimum possible path to reach the end vertices. However , we can use the shortest path algorithm to solve this problem. Thus, it is preferable to use FSVM algorithm compared to CNN approach because it gives us more efficiency compared to CNN.

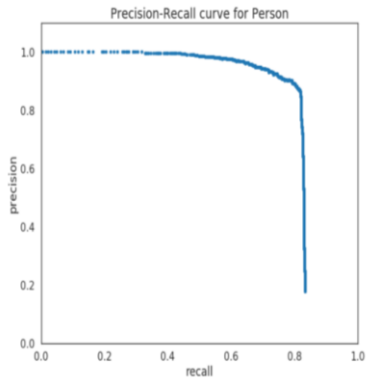
In many applications, there is a need for recognizing object classes, and hence we use multiple binary classifier so that we can detect multiple frames per second and thus increase the efficiency of the period. This approach detects those areas where the probability of occurrence of object of interest is more. Steiner tree-based approach is developed to provide a solution to this problem. However, the Steiner-tree approach can be used to classify images present in a particular image or a set of frames of an image referring to a video. Multi-scale booster detector can also be used to recognize the objects in a multi-scale platform.

During the evaluation phase the classifier uses the set of values that are obtained for a particular image and matches those values with the range of values store in the

database for particular class. When the parameter value obtained for the given image is the same as the fixed value for the parameter, then it follows the path through the node of one level to other, but when the parameter value for the given image does not match the fixed value of the parameter, then the intermediate Steiner node is to be created for this value for the relevant level. For later case, the evaluation of the intermediate Steiner node can be carried out using the Euclidian distance to check the relevance of this intermediate node with the main nodes of that level. Finally, the Steiner tree can be created based on the identified main nodes. Last node of the created Steiner tree is the Steiner node for the class of the object.

We have precision and recall method to analyse the evaluation of object detectors. The two steps followed in precision and recall are:

1. Determining whether an object exists in the image (classification)
2. Determining the location of the object (localization, a regression task).

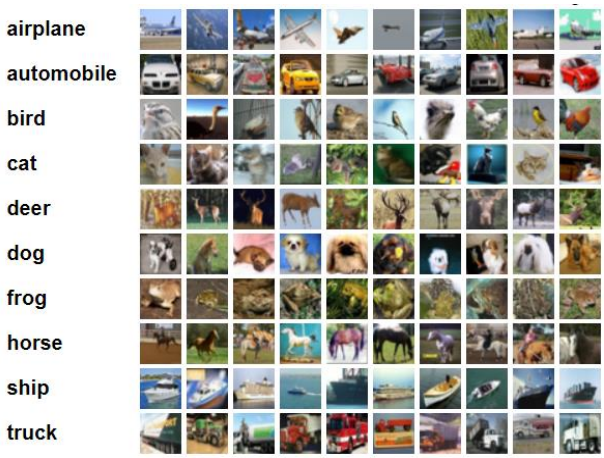



Here we thus calculate AP score which is defined as the mean precision at the set of 11 equally spaced recall values, Recall_i = [0, 0.1, 0.2, …, 1.0]. Thus,

$$AP = \frac{1}{11} \sum_{\text{Recall}_i} \text{Precision}(\text{Recall}_i)$$

**Dataset:**

The dataset is collected in the model's section where there are more than 350 images labelled accordingly. On the contradiction, it depicts the Box AP ratio for the same viewing video. These images in the dataset maps the image and recollects and thereby localises the image being played in the video. The sample dataset of 10 images is shown below in figure (3).




## 3. Conclusion:

This paper obscures the idea , how to review of the various methods for detecting objects in images as well as in videos. The process of Object detection method used in this paper is sliding window approach , basically called as Box Ap.

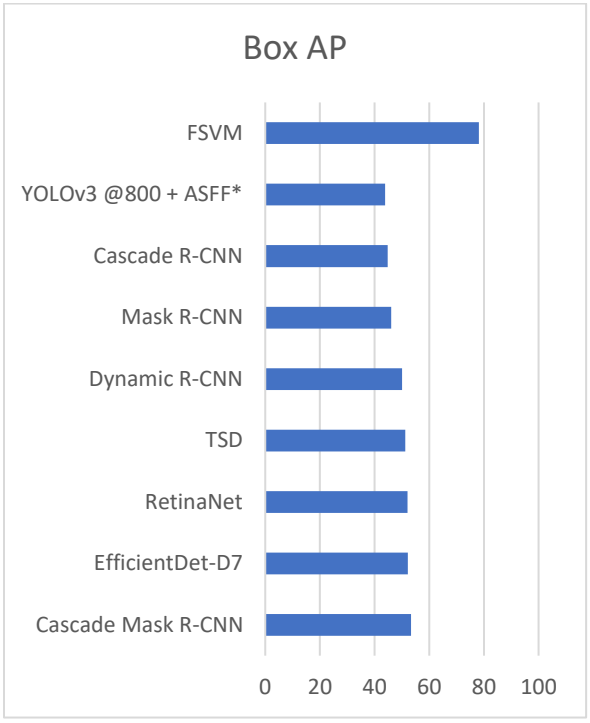
This paper also provides the idea to solve the multi class OD problem based on the Steiner tree.

Comparison with other detection systems:

1. This model is been in comparison with RCNN and other faster CNN techniques. The RCNN method uses selective search mechanism to create the labelled rectangular box known as BOX AP. This method enables to look at an image through windows of different sizes and patches. Then, it extracts the region and then passes this region through CNN labels to generate CNN based features. At last, we classify the image by adding a feature of SVM(support vector machine) to detect the object in the region proposed.
2. The next method of faster CNN uses ROIPOOL (Region of interest Pooling). It runs on the mechanism that only one image is shared to the other computational sub regions. It allows only one per pass of the actual image. It has a BOX AP of about 70. Its drawback is real time performance.

3. As compared to our proposed model with uses the FSVM , the same as SVM specializes the machine to work in a real-life video object detection. It classifies and localises the objects faster than other faster CNN detection techniques. It has a BOX AP recognition value of above 71.2, which makes it functional in the real time environment.

The tabulation of various object detection methods over Box AP shows the efficiency of various methods as under:




Hence on the basis of the above estimation , it is thereby understood that FSVM has a much better evaluation of objects compared to other methods of object detection.

**Report of the output generated:**

The results from COCO dataset installed in the system has the following output generated after run.

## 4. References:

[1] V. Ferrari, T. Tuytelaars, L. V. Gool, Simultaneous Object Recognition and Segmentation, published in 8th European Conference on Computer Vision, Prague, Czech Republic, Vol. 3021, pp. 40-54, 2004.

[2] Y. LeCun , F. J. Huang, L. Bottou, Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting, published in IEEE Conference on Computer Vision and Pattern Recognition, Vol.2, pp. 1-8, 2004.

[3] Object Detection with Deep Learning: A Review Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE 16-04-2019

[4] A review and an approach for object detection in images Kartik Umesh Sharma* and Nileshsingh V. Thakur

[5]Camp, K. and Stiefelhagen, R. (2007) 'Automatic person detection and tracking using fuzzy controlled active cameras', IEEE Conference on Computer Vision and Pattern Recognition.

[6] Chen, C. and Tian, Y. (2010) 'Door detection via signage context-based hierarchical compositional model', Computer Vision and Pattern Recognition Workshop, pp.1–6.

[7] Elbouz, M., Ayman, A. and Brosseau, C. (2011) 'Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance', Optical Engineering, Vol. 50, No. 6, pp.1–13.

[8]Karasulu, B. (2010) 'Review and evaluation of well-known methods for moving object detection and tracking videos', Journal of Aeronautics and Space Technologies, Vol. 4, No. 4, pp.11–22.

[9] Kaur, R. and Dhir, V. (2013) ‘Fuzzy logic based novel approach for face detection’, International Journal of Latest Research in Science and Technology, Vol. 2, No. 1, pp.558–566.

[10] Liang, Z., Chi, Z., Fu, H. and Feng, D. (2012) ‘Salient object detection using context sensitive hyper graph representation and partitioning’, Pattern Recognition, Vol. 45, No. 11, pp.3886–3901