IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 12, December 2020

DOI 10.17148/IJARCCE.2020.91218

Systematic Survey on Object Detection and Recognition using Machine Learning Techniques

Aparna Bodke[1], Asjadurrahman Ansari[2], Rohan Sirsulwar[3], Tehsina Shaikh[4] , Prof. K.S.Mulani[5]

Student, IT, Sinhgad Institute of Technology, Lonavala, India[1-4]

Professor, IT, Sinhgad Institute of Technology, Lonavala, India[5]

Abstract: In this project, we use a completely deep learning based approach to solve the problem of object detection in an end-to-end fashion. The network is trained on the most challenging publicly available dataset MS COCO like (SSD, RCNN, Faster RCNN, YOLO v3, 4 etc.), on which object detection challenge is conducted annually. The objects are detected in boxes by this dataset where objects like car, bike, person, etc.

Keywords: Object detection, convolution neural network, scoring system, selective search, deep learning, MS COCO, SSD, RCNN, YOLO, OD model.

I. INTRODUCTION

All Object Recognition has two parts: Category Recognition and Detection. Category Detection deals with distinguishing the object from background. Moreover, Category Recognition deals with classifying the object into one of the preferred categories. It is identifying a process of a specific object in digital image or video. Generally , Object Detection algorithms rely on machine learning or pattern recognition algorithms using appearance-based or feature-based techniques. For example, it is used to find instances of real life objects like fruits, chairs, cups, animals, etc.

II. LITERATURE SURVEY

In various fields, there is a necessity to detect and also track them effectively while handling occlusions and other include complexities. Many researchers (Almeida and Gutting 2004, Aure lie Bugeau and Nicolas Papadakis 2010, Hsiao-Ping Tsai 2011) attempted for various approaches in object tracking. The nature of techniques largely depends on the application domain. Some of these search works which made the evolution to proposed work in the field of object tracking.

Object detection is an important task, yet a challenging vision task. It had already been applied in many computer vision fields, such as artificial intelligence, smart video surveillance and robot navigation, military guidance, medical scenarios, safety detection, etc. It is a critical part of many applications such as image auto-automation, image search, and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision and machine learning.

In recent years, a number of successful single-object tracking systems appeared, but in the presence of several objects, object detection becomes difficult and when objects are fully or partially occluded, they are obstructed from the human vision which further increases the problem of detection. Decreasing illumination and acquisition angle. The proposed MLP based object tracking system is made robust by an optimum selection of unique feature sandal so by implementing the Adaboost strong classification method[1].

III. SYSTEM ARCHITECTURE

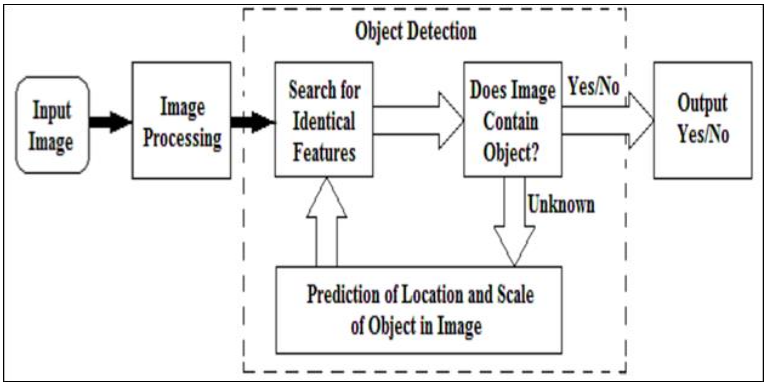

Fig.1. OD Model

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 12, December 2020

DOI 10.17148/IJARCCE.2020.91218

These three computer vision tasks:

Image Classification: Predict the type or class of an object in an input image or video.

- Input: An image or video with a single or multiple objects, such as a photograph or recordings.
- Output: A class label (e.g. class labels are being mapped by two or more integers).

Object Localization: Locate the presence of objects in an image or a video and indicate their location with a bounding box.

- Input: An image or video with a single or multiple objects, such as a photograph or recordings.
- Output: One or more bounding boxes (e.g. defined by a point, width, and height).

Object Detection: Locate the presence of objects with a bounding box and types or classes of the located objects in an input image or a video.

- Input: An image with one or more objects, such as a photograph.
- Output: One or more bounding boxes (e.g. defined by a height, width, point), and a class label for each bounding box.

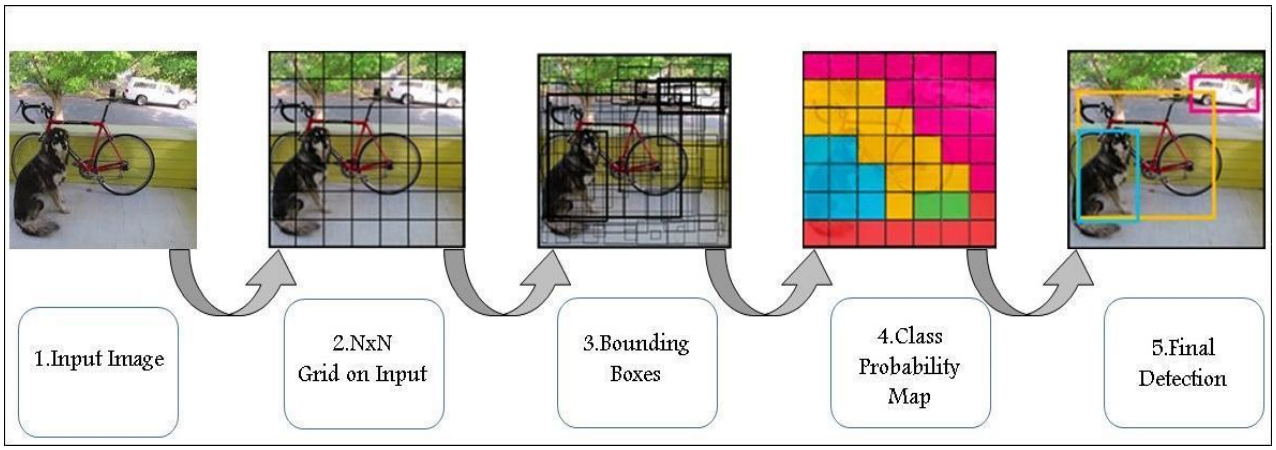

Fig. 2. Image processing

This system divides the input image into an S x S grid. The grid cell is responsible for detecting that object if the centre of an object falls into a grid cell. The B bounding boxes and confidence scores for those boxes are being predicted by each grid cell . The confidence score shows how confident the model is that the box contains an object and also how accurate it thinks the box is that it predicts. Formally we define confidence as Pr(Object)* IOU.
The confidence scores should be zero, if no object exists in that cell. Otherwise the confidence score should be equal to the intersection over union (IOU) between the predicted box and the ground truth[9].

Each bounding box consists of 5 predictions: p, q, r, s and confidence. The centre of the box relative to the bounds of the grid cell is represented by the (p, q) coordinates. The height and width are predicted relative to the whole input image. Finally the confidence prediction represents the IOU between the predicted boc and any ground truth box.[Shraddha Mane (ICICCS 2019)] Each grid cell also predicts C conditional class probabilities , Pr(Class i|Object).
Such are some probabilities which then conditioned on the grid cell containing an object. Regardless of the number of boxes B,we only predict one set of class probabilities per grid cell [6].

At the time of testing we multiply the probabilities of conditional classes and confidence predictions of the individual box which gives us class-specific confidence scores for each box. These scores encode both the probability of that class appearing in the box and how many well the predicted box fits the object. We split the image into (s x s) grid. Grid cell is responsible for detecting an object. The cell is responsible for detecting the object only if the centre of an object falls into a grid cell. Each cell predicts the B number of bounding boxes.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 12, December 2020

DOI 10.17148/IJARCCE.2020.91218

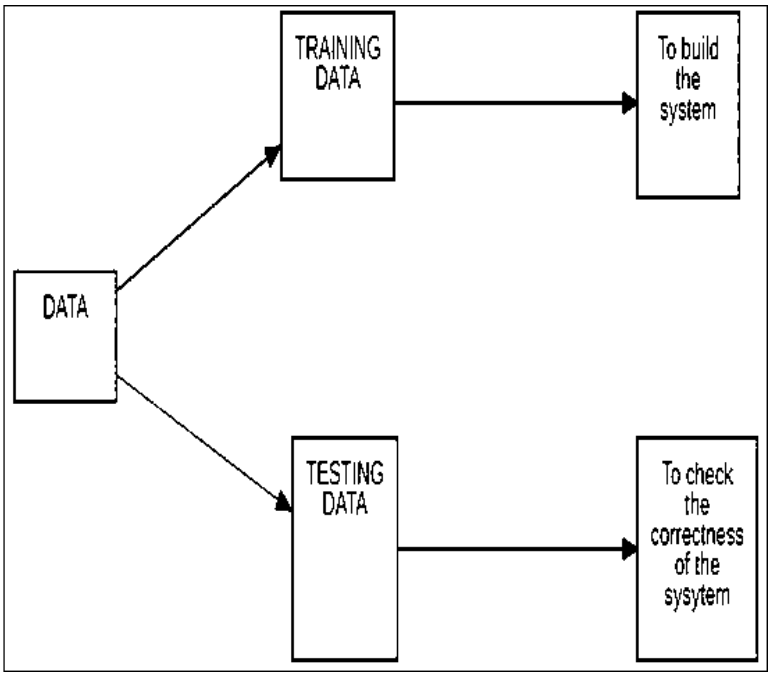

Fig. 3 Two types of Data (Training Data and Testing Data)

In machine learning, a common task is there to study and build the algorithms that can learn from and make predictions on data. Such algorithms function by making data-driven predictions or decisions, through building a mathematical model from input data.[1][10][7]

The model is initially a suitable fit for a training dataset, which is a set of examples used to fit the parameters (e.g. weights of connections between neurons in ANN) of the model. The data used to build the final model usually comes from multiple datasets. In particular, for the creation of this model three datasets are used in different stages.

A supervised learning method is used to train the model (e.g. a naive Bayes classifier) on the training dataset using, for example using optimization methods such as gradient descent or stochastic gradient descent.[8]

In practice, when we are operating on the training dataset it often consists of an input vector(or scalar), in which the answer key is commonly denoted as the target or label. After running the current model on a training dataset it produces a result, which then comes into comparison with the target for each input vector in the training dataset. The result of the comparison between the target and input vector is used as a basis and then a specific learning algorithm is used, the parameters of the model are adjusted. The model fitting can include both variable selection and parameter estimation[1].

Successively, then the most fitted model is used to predict the responses for the observations in a second dataset called the validation dataset. Unbiased evaluation of a model provided by the validation dataset fits on the training dataset while tuning the model's hyperparameters (e.g. the number of hidden units i.e. layers in a neural network).

For regularization with the validation dataset the method introduced of stopping training when the error on the validation dataset increases, as this is a sign of overfitting to the training dataset which is called “early stopping”.
The fact that the validation dataset’s error may fluctuate during training by producing multiple local minima which makes simple procedure complicated in practice. These complications create many ad-hoc rules for deciding when overfitting has truly begun.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 12, December 2020

DOI 10.17148/IJARCCE.2020.91218

IV. FLOW CHART

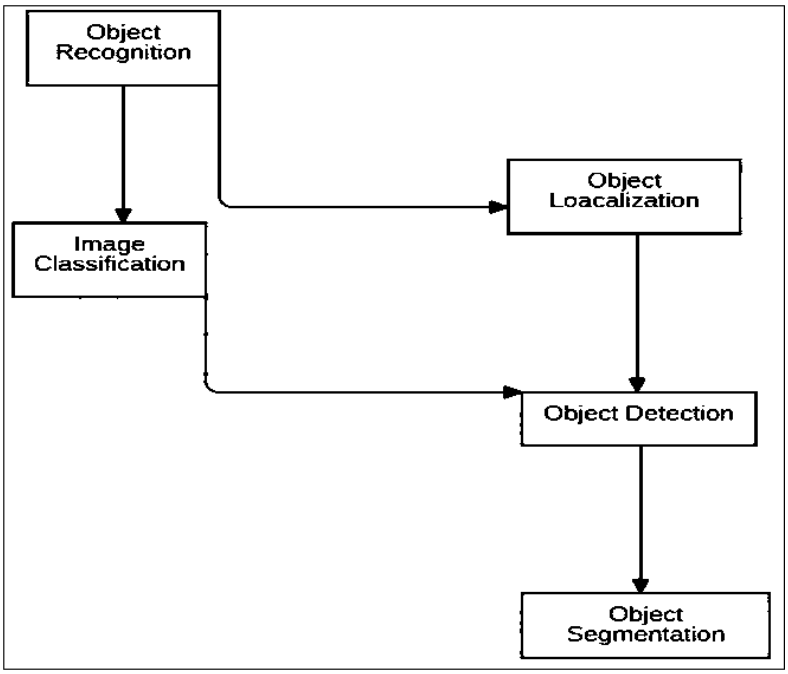

Fig. 4. Flowchart of OD Model

Basically an OD system can be described easily by seeing Fig. 4 which shows the basic stages that are involved in the process of OD. The basic input to the OD system can be an image or scene in case of videos. The basic aim of this system is to detect objects that are present in the image or scene or simply in other words the system needs to categorise the various objects into respective object classes[1].

In computer vision object segmentation is another breakdown to this which is also called "object instance segmentation" or "semantic segmentation," where instances of recognized objects are indicated by highlighting the specific pixels of the object instead of a coarse bounding box[1].

The OD problem can be defined as a labelling problem based on models of known objects. Given an image containing one or more objects of interest and a set of labels corresponding to a set of models known to the system, the work of assigning correct labels to regions in the image is expected by this system. The OD problem cannot be solved until the image is segmented and without at least a partial detection, segmentation process cannot be applied. The term detection has been used to refer to many different visual abilities including identification, categorization and discrimination[1].

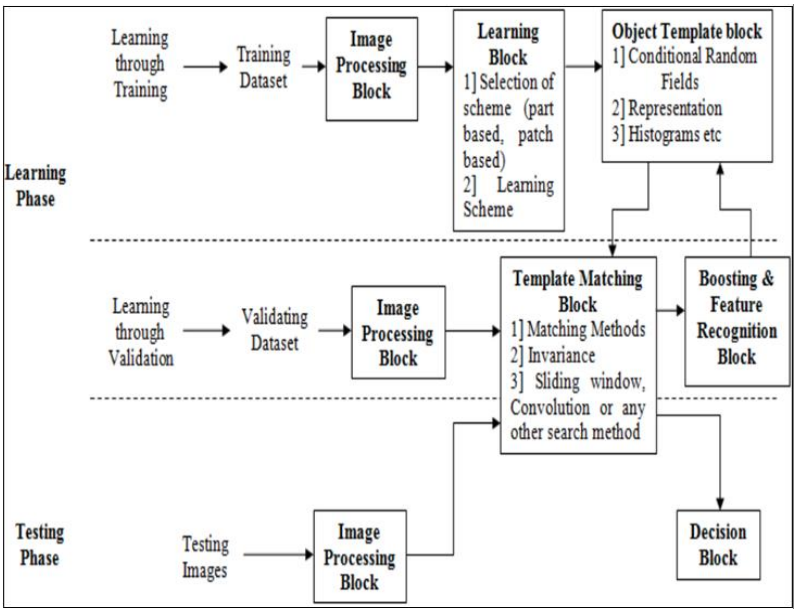

Fig. 5. Basic OD Model[1][3]

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 12, December 2020

DOI 10.17148/IJARCCE.2020.91218

The OD problem can be defined as a labelling problem based on models of known objects. Given an image or a video containing one or more objects of interest and a set of labels corresponding to a set of models known to the system, the system is expected to assign correct labels to regions in the image or video[1].

The OD problem cannot be solved until the image is segmented and without at least a partial detection, the segmentation process cannot be applied. The term detection has been used to refer to many different visual abilities including identification, categorization, and discrimination[1].

V. CONCLUSION

This paper uses recent techniques in the field of computer vision and deep learning. Custom dataset was created using labelling and the evaluation was consistent.

This can be used in many custom and real-time applications, which require object detection for pre-processing in their pipeline. This paper also provides experimental results on different methods for object detection and identification and compares each method for their efficiencies.

REFERENCES

[1]. Kartik Umesh Sharma and , Niles Singh V. Thakur,(IEEE – 2020) – A Review and an Approach for Object Detection in Images.
[2]. Ning Wang1 , Yang Gao , Hao Chen , Peng Wang , Zhi Tian , Chunhua Shen , Yanning Zhang(IEEE Paper - 2020) - NAS-FCOS: Fast Neural Architecture Search for Object Detection.
[3]. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao (IEEE Paper - 2020) - YOLOv4: Optimal Speed and Accuracy of Object Detection.
[4]. Seijoon Kim, Sungsik Park, Byunggook Na, Sungroh Yoon (IEEE Paper -2019) - Spiking-YOLO: Spiking Neural Network for Energy-Efficient Object Detection.
[5]. Jung Uk Kim and Yong Man Ro (IEEE paper- 2019) - Attentive Layer Separation for object classification and Object Localization in Object Detection.
[6]. Shraddha Mane , Prof. Suraya Mangale (ICICCS 2019) - Moving object detection and tracking Using Convolutional Neural Networks.
[7]. Liyan Yu, Xianqiao Chen, Sansan Zhou (IEEE Paper - 2019) - Research of Image Main Objects Detection Algorithm Based on Deep Learning.
[8]. Ahmed Fawzy Elaraby, Prof. Ayman Hamdy, Dr. Mohamed Rehan (IEEE paper - 2019) - A Kinetic- Based 3D Object Detection and Recognition System with Enhanced Depth Estimation Algorithm.
[9]. Kanimozhi S, Gayathri G, Mala T (IEEE Paper - 2018) - Multiple Real-time object identification using Single shot Multibox detection.
[10]. Christian Szegedy, Alexander Toshev, Dumitru Erhan (IEEE Paper-2018) - Deep Neural Networks for Object Detection.