IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Real-time Object Detection Using Deep Learning

Nikhil Anil Deshmukh

Dept. of C.S, Shah and Anchor Kutchii Engineering College,
Chembur, Mumbai-400088

Abstract

As technology developed, object detection attracted the attention of researchers due to its link to image and video analysis. Earlier object identification techniques relied on trainable algorithms, flawed structures, and manually generated features. Since many object recognition systems depend on other computer vision techniques to bolster their deep learning-based methodology, they frequently operate slowly and poorly. One of the main issues with these systems is this. In this research, we provide a comprehensive solution to the object detection problem using deep learning. The single shot detector (SSD) methodology is the quickest method for object detection from an image using a single layer of a convolution network. Our research's primary goal is to increase the SSD technique's accuracy.

Keywords: Object detection; SSD method; deep learning.

1 Introduction

The method of identifying an image's class, or image classification, was one of the main issues. picture localization is a challenging task when there is just one item in the picture and the system has to estimate its class and location inside the image (a bounding box around the object). Since object discovery requires both identification and location, it is a more challenging process. In this instance, the system will receive an image as input, match each object in the picture to a bounding box, and identify the type of object that is inside each box. Compared to the existing ways, we have created a system that runs at greater frames per second (FPS) while using less computational resources. Moreover rapid identification of objects [1,2]. Using the SSD mobile net technique, our object discovery model detects the item in the image and celebrates it. The algorithm of our model analyzes an object's look in a photograph to determine which one it is.

Object detection is a computer vision technique that helps identify and locate objects in images and movies. The counting of items in a scenario, their precise placement and identification, and their labeling are all made possible by this localization and identification technique. Have you ever noticed how adeptly Facebook can identify your pals from your photos? Before, to tag friends in images on Facebook, you had to go on their profile and input their names.

[3-5]. As soon as you upload an image to Facebook, it now instantly tags every friend in the picture. This method is called face recognition. Facebook's algorithms may be able to recognize your friends' faces after just a few times of being tagged. Facebook's 98% accuracy rate for facial recognition is comparable to that of humans. In social media and mobile video broadcasts, people can be recognized by their faces. [6–8].

In order to update and enhance the current attendance system and make it more effective and efficient than before, the primary objective is to design a deep learning and facial recognition based model for attendance management particularly for the education sector. A great deal of uncertainty surrounds the outdated approach, which leads to to a faulty and inefficient attendance recording system. Numerous problems arise when the laws of the former system are not upheld by the government. The invention will be a face-recognition recognition system. A person's face is among the most typical physical traits that may be used to positively identify them. It is used to monitor identity since faces seldom diverge or duplicate. To provide the recognizer algorithm with data for this study, face databases will be built [9– 11]. After that, during the time allotted for reporting attendance, faces will be compared to those in the database in an effort to identify who they are. Upon identification, attendance is instantly monitored, utilizing the pertinent data being recorded onto Excel spreadsheet.

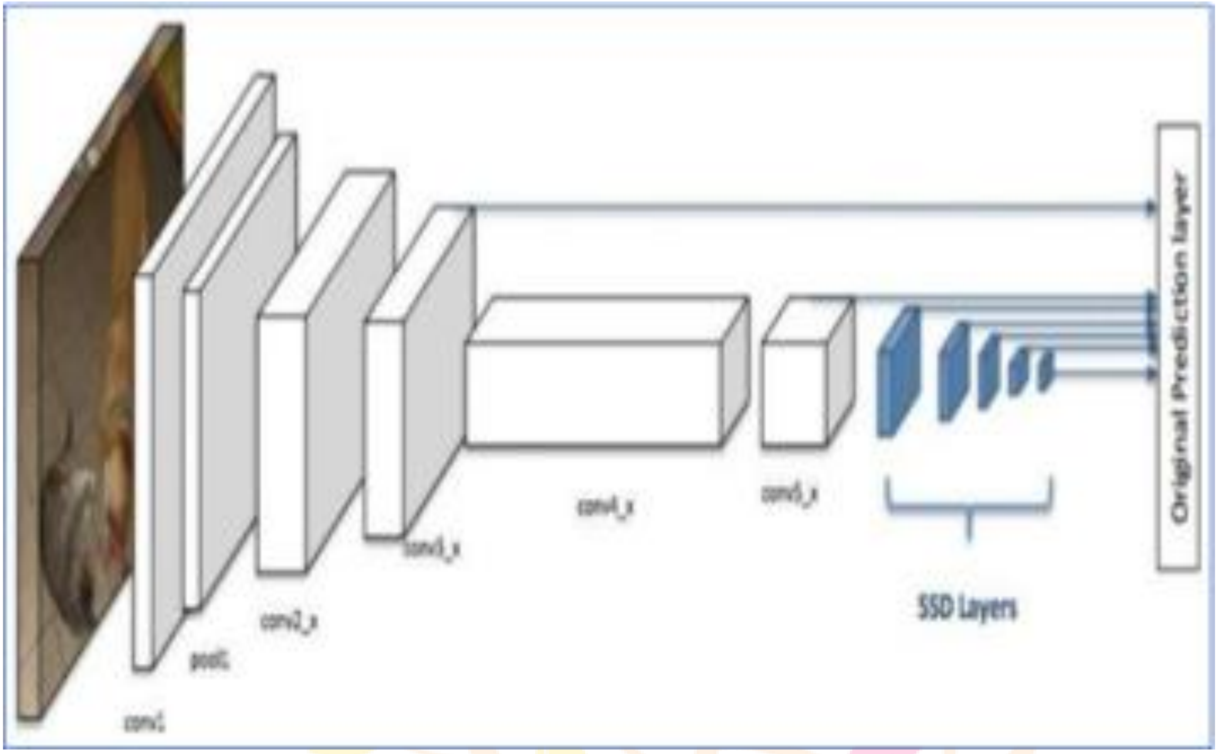


Figure 1: System Architecture Figure

2 Literature Survey

In the 1980s, the first image recognition technologies became available. Many innovative techniques for image processing were created in the years that followed. For many real-world applications, such as picture recovery and video monitoring, object detection is essential. We developed the You Only Look Once (YOLO) technology for immediate computing. Previous recognition systems repeatedly find targets using classifiers or localizers.
They apply the model in different sizes and places on a picture. Detections are portions of high-scoring images. We adopt a very different approach. We process the entire image using a single neural network. This network creates box boundaries, divides the image into regions, and projects possible outcomes for each. These bounding boxes are weighted based on predicted likelihood [12–14].

This approach has a number of advantages over classifier-based systems. Since it assesses the entire image during testing, the entire context of the image has an impact on its predictions. Furthermore, it uses a single network assessment to provide predictions, in contrast to R-CNN, which needs many evaluations for a single picture. It is therefore 1,000 times quicker than Fast R-CNN and 100 times faster than R-CNN. The input picture is split up into SS cells by the YOLO network, which is responsible for recognizing each object.

In addition to their forecasts for their respective classes, each grid cell projects the objectless value of the B surrounding frame. Lastly, the bounding box confidence is added together to get the probability that this bounding box has a specific type of object.summing together the final outcome and the class prediction into a single figure. YOLO v3 issues with little objects that show up in clusters.

YOLO V3 is an object detector that recognizes objects in real time by using features that a deep convolutional neural network has learned. It processes the entire image using a single neural network that consists of 75 convolutional layers, upsampling layers, and skipped connections. There are several sections in the image. Later boundary boxes are used to display probabilities. The most notable aspect of YOLO V3 is its capacity to do detections at three different scales. But in order to achieve more accuracy, YOLO v3 has lost speed, and it has trouble handling little objects that appear in groups.

Two networks make up Faster R-CNN: a region proposal network (RPN) that creates zone suggestions and an object identification framework that utilizes these concepts. The main distinction between this approach and Fast R-CNN is that this method generates region recommendations using selective search. When RPN shares the majority of its computations with the object identification structure, it generates area suggestions much quicker than focused screening. RPN assigns a number to each area box, also known as an anchor, indicating which ones are most likely to hold items. Two quick RCNN algorithms are used by the Region Proposal Network to identify objects and build regions. The first method uses the recommendations made for the suggested regions. A prohibition on faster The two main drawbacks of R-CNN are its slow processing speed and difficult training process.

3 Methodology

The OpenCV programming package or library's primary goal is to aid programmers in their understanding of computer vision. A collection of open-source computer vision programs known as "OpenCV" were developed by Intel Corporation and made available to the general public in 1999 and 2000. (A repository). The most popular, well-known, and well documented computer vision collection. Using the program doesn't require a license because it is open-source. As you probably already know, the majority of machine learning algorithms need numerical or quantitative inputs. Although OpenCV allows us to process pictures using machine learning techniques, raw images usually need to be processed to turn them into features, which are columns of data.
They improve our machine learning techniques that they utilize.

Python has a package called NumPy. A class of operations for working with arrays and multidimensional array objects is called "Numerical Python". Jim Hugunin invented Numeric, which was the forerunner to NumPy.
Furthermore, several additional methods were added to a new Num array package.

There are a significantly greater number of utilities in Dlib now than there were when the project started in 2002. These include threading, networking, graphical user interfaces, and other contemporary software-intensive tasks. The development of several probabilistic forecasting algorithms has sparked a lot of interest in current research.

Pandas is an open-source program for data analysis and manipulation that is rapid, robust, adaptable, and simple to use. using Python as the main language for programming.

The Python Imaging Library allows the Python interpreter to work with pictures. This library provides a wide range of file format compatibility, a helpful internal representation, and some quite powerful image processing tools.

The csv module allows classes to send and receive structured data in CSV format. It is not necessary for developers to be conversant with Excel's CSV format in order to give directions on how to write this data in the way that Excel prefers. Python's OS module offers techniques and tools for working with the operating system.

3.1 Detector for single shots (SSD)

The suggested method makes use of an enhanced SSD algorithm to deliver real-time detection that is both faster and more accurate. But since the SSD method ignores the background that exists outside of the boxes, it is inappropriate for microscopic object recognition. The suggested approach's convolutional layers make use of both depth-wise and spatially separable convolutions to address this problem. In particular, our suggested approach combines a multilayer convolutional neural network with an innovative architecture. There are two steps to the algorithm. By using a resolution multiplier, it first minimizes the extraction of spatial dimensions from feature maps. Second, tiny convolutional filters are used in the building process to apply the ideal aspect ratio values for object detection. The main goal of training is to increase one’s confidence level.

Using a region proposal network, Faster R-CNN creates boundary boxes that are then used to categorize things. The entire process operates at 7 frames per second, far less than what is needed for real-time processing, even if the precision is thought to be state-of-the-art. The SSD speeds up the procedure by doing away with the need for the area proposal network. SSD has improved default boxes and multi-scale capability to counteract the accuracy decrease. These improvements speed up the process by enabling SSD to function with pictures of lesser quality while still matching the accuracy of the Faster R-CNN.

Single Shot Detector is significantly quicker and more accurate when compared to older techniques. To achieve high accuracy, we use feature maps of various dimensions to build predictions on various scales, and then divide the forecasts based on aspect ratio.

High accuracy may be achieved even with low-quality input photos because to these capabilities.

Other algorithms make extensive use of the object proposal technique. In essence, these algorithms provide a method for segmenting an image and then suggest potential placements for items inside those segments. These algorithms forfeit accuracy. To differentiate between genuine or empirical evidence and conjectured evidence, the idea of "ground truth" is employed. If some boxes are missing, we can't merely train the algorithm—we have to find those missing boxes first, during the training phase.

After dividing the picture into many pieces, SSD will construct the bounding boxes for each segment. After that, it will go through each box in the picture to find an object from each class that the network has been trained to identify. Finally, a comparison will be done between the expected and actual outcomes. In order to help update the weights, if an error arises after the comparison, it is back-propagated across the network.

A single shot detector, like YOLO, utilizes a multi box and a single shot to find several objects in an image. Its method of object detection is quicker and more precise. a quick review of the various object detection The speed and precision of the methods. The SSD's rapid speed and precision when using relatively low-resolution photographs are made possible by the following factors:

Bounding box recommendations (such as those used in RCNNs) are removed.

To account for offsets in bounding box locations and item classes, a convolution filter with progressive loss is used.

SSD uses a lot of boxes or filters with different aspect ratios and sizes to achieve excellent object recognition accuracy. This facilitates detection on several levels.

Three hundred different images that were taken from the Internet are included in the data collection. The SSD idea and algorithm will be applied in this project. This will help us identify the item based on its many properties (Depends on training).

3.2 Data set description

Photographs of people, boats, bicycles, cows, bottles, and other items are among the 300 photographs in our collection. We test our technique using a camera that records the Items in real time. The graphic below displays a number of example photographs following pre-processing.

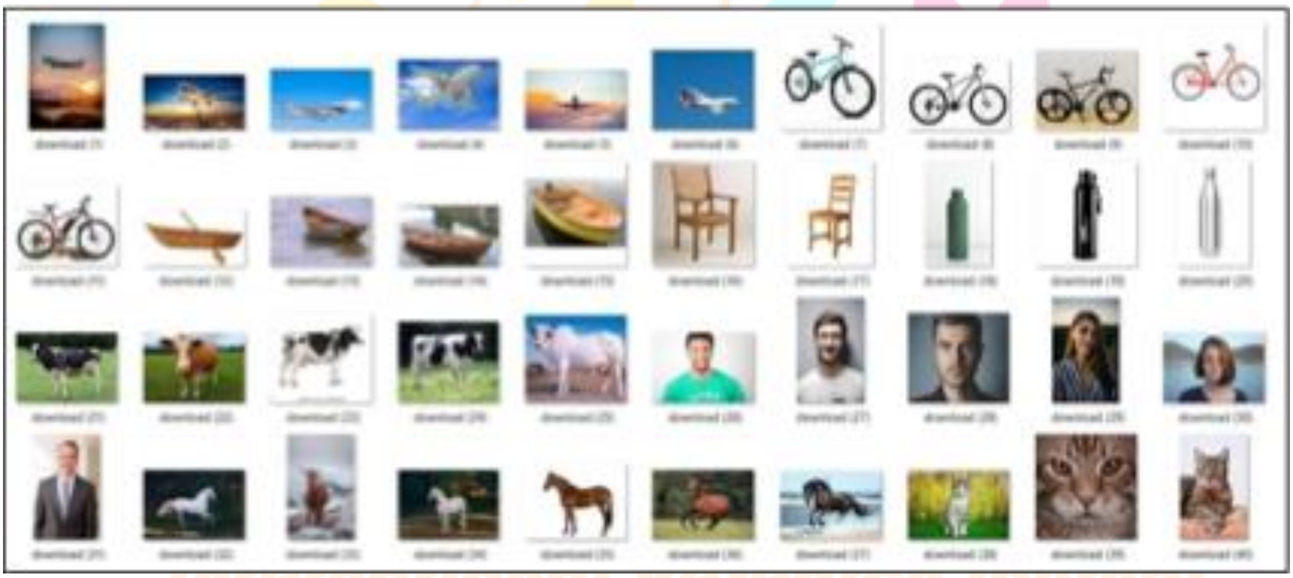


Fig. 2. Pictures in the Dataset

4 Results and Analysis

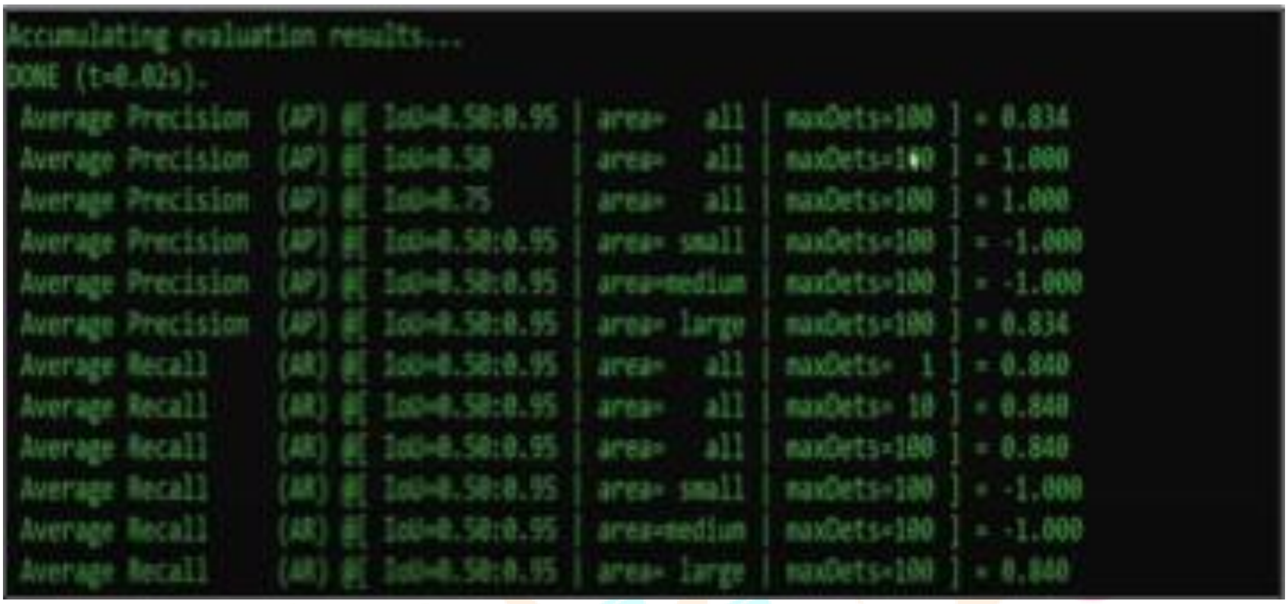
The following steps are included in our proposed system. Step 1: The user's camera is used to take the input photo. Step2: An picture alteration is made.

Step 3: The picture has all required features removed.

Step 4: To enable the recognition of additional elements, the picture is divided into smaller segments.

Step 5: Try to categorize and identify the things once they have been segmented.

Step 6: The next step is to identify the things in the photo. Step 7: The result is displayed to the user.



```
Accumulating evaluation results...
DONE (t=0.02s).
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.834
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 1.000
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 1.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = -1.000
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.834
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.840
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.840
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.840
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = -1.000
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.840
```

Fig. 3. Accurate results

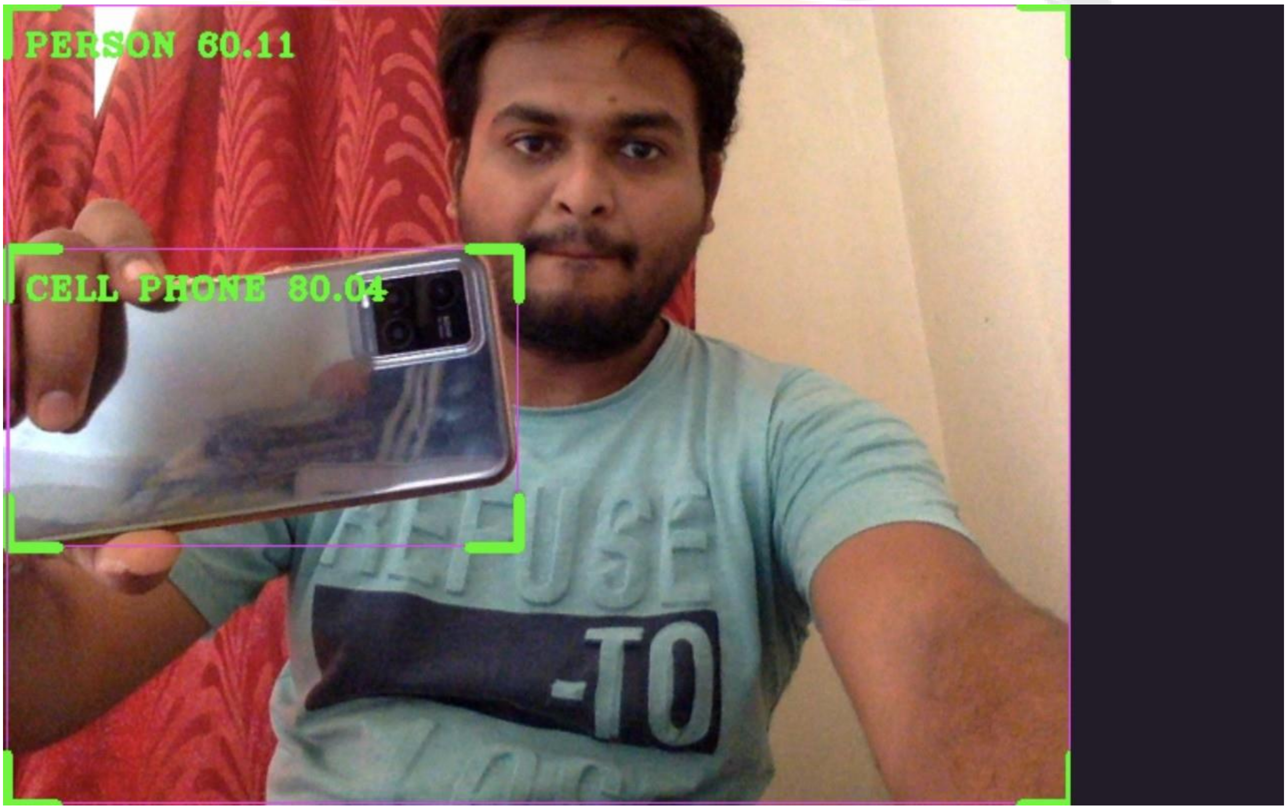


Fig. 4. Results Cell phone

It has been determined that the object in issue is a cell phone and Person; in this case, the object detection accuracy is 80.04%

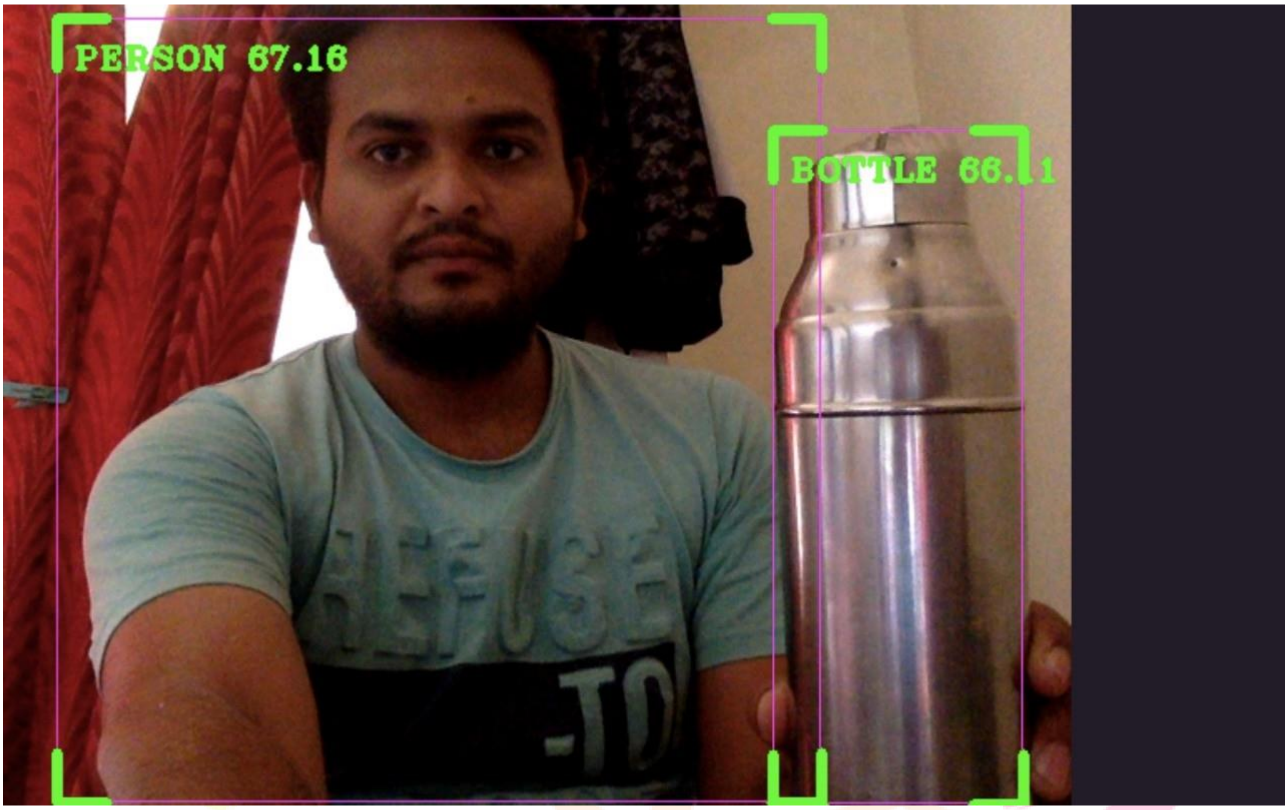


Fig. 5.Results Person and Bottle.

This outcome shows that more than one item may be detected using this approach. This image appears to show one bottle 66.11% and one person 67.16%.

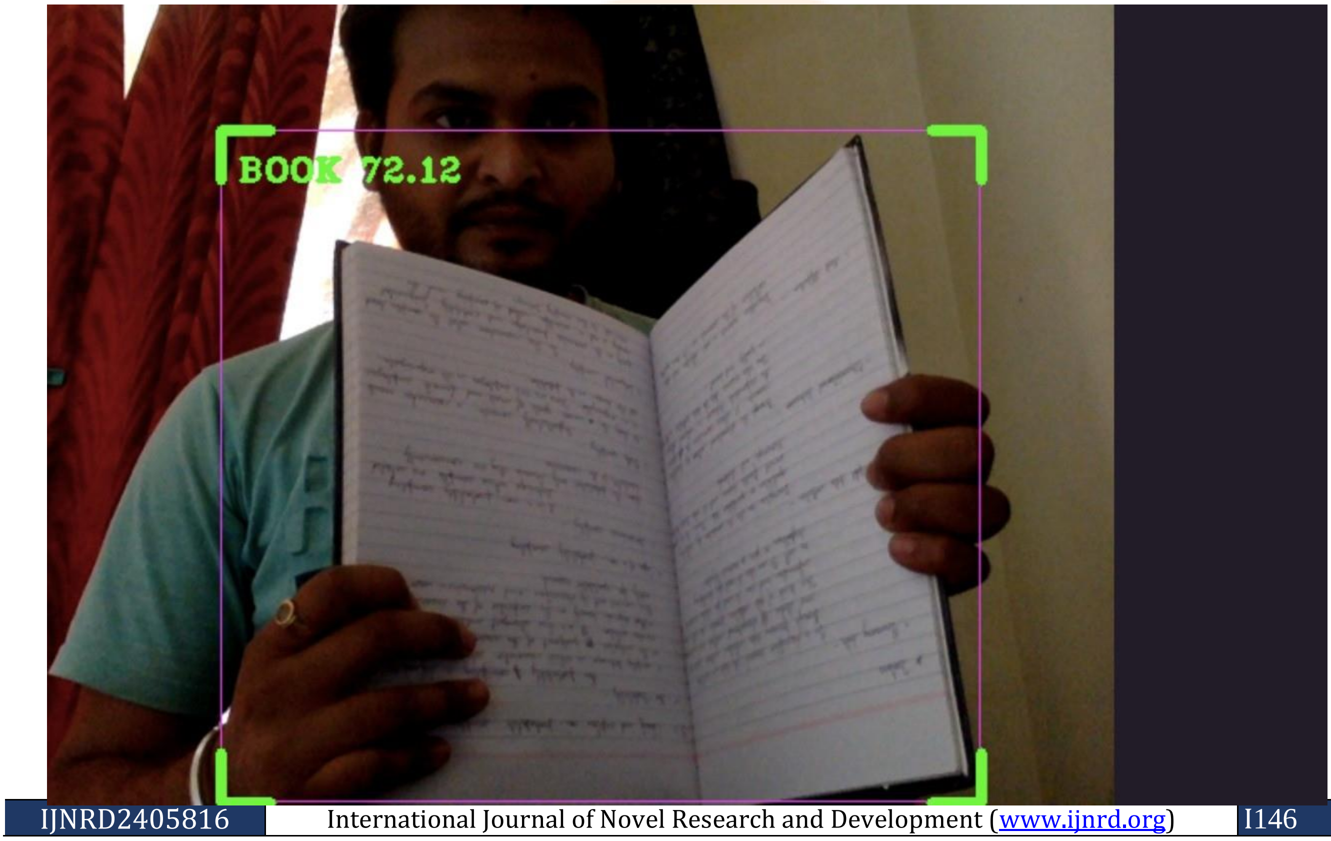


Fig. 6. Product Book

The object is categorized as a book since the book is pretending to be a noteworthy product in front of the camera; the backdrop pictures are still unknown.

4.1 Testing types

Making that your mobile and web apps function and benefit people with different kinds of impairments, such visual impairment, hearing loss, or other physical or mental difficulties, is part of the accessibility testing process.

Adoption testing:- This type of testing confirms that software can be delivered based on whether end users can satisfy the objectives outlined in industry specifications. It also goes by the moniker UAT (UAT).

Testing a "black box":- "Black box" testing is the process of testing a system that has code and hidden pathways.

Comprehensive testing:- To ensure that everything works as it should, a technique called end-to-end testing looks at every stage of an application's workflow.

Functional evaluation:- A software, website, or system's functionality is examined to make sure it is operating as it should.

Examining interactively:- Also called manual testing, interactive testing enables testers to generate and support manual testing for individuals who don't use automation by obtaining data from external tests.

Integrity checks:- Integration testing confirms that a system's integrated components meet a set of specifications. In order to ensure optimal system performance, it is carried out in an integrated online and offline environment.

5 Conclusion and Future Scope

The aim of this effort is to use deep learning to construct an object recognizer for photographs. The study uses an enhanced SSD method together with a multilayer convolution network to detect items quickly and accurately. Both fixed and moving photographs are easily handled by our technology. More than 80% of the predictions made by the proposed model are true. After removing feature data from the image, convolution neural networks employ feature mapping to get the class label. The primary objective of our approach is to improve SSD's object detection performance by selecting default boxes with the best aspect ratios.

Object recognition technology has the ability to liberate people from repetitive activities that robots can accomplish more quickly and efficiently, much to what happened during the first Industrial Revolution.They are presently putting this method to the test.

Competing Interests

There are no competing interests, according to the authors.

References

[1] Du J. Object Detection Comprehension Based on CNN Family and YOLO, J. Phys. Conf. S. 2018;1004(1). DOI: 10.1088/1742- 6596/1004/1/012029

[2] Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks. In 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE. 2015;1316-1322

[3] Item Detection and Recognition in Pictures, Sandeep Kumar, Aman Balyan, and Manvi Chawla, IJEDR. 2017;1-6.