WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real-Time Object Detection Using Deep Learning

Sushmitha KC[1], Asst Prof. Uma Mageswari[2]

[1,2] Dept. of Computer Applications, MCA, Presidency College (Autonomous) Bangalore, India
sushma6148@gmail.com, uma.mageswari@presidency.edu.in

ABSTRACT:

Researchers became interested in object detection as technology advanced because of its connection to video and image analysis. Previous methods of object recognition rely on manually created features, imperfect structures, and trainable algorithms. Many object identification systems suffer from slow and poor performance since they rely on other computer vision techniques to support their deep learning-based approach. This is one of the key problems with these systems.. In this paper, we propose a deep learning based end-to-end solution to the object detection problem. The fastest approach for object detection from an image utilizing a single layer of a convolution network is the single shot detector (SSD) technique. Improving the accuracy of SSD technique is the main objective of our research.

Keywords: Object detection; SSD method; Deep learning.

1.INTRODUCTION:

One of the key problems was image classification, which is the process of determining the class of the image. When there is only one object in the image and the system has to guess its class and location inside the image, image localization presents a difficult problem. It is a more difficult task since item discovery involves both identification and localization. In this case, the system will take an image as input and produce a bounding box that matches each object in the image and indicates what kind of thing is inside each box. We developed a method that operates at higher frames per second

(FPS) and faster object recognition while consuming less computing power than the current methods. Our object discovery model recognizes the object in the picture and celebrates it using the SSD mobile net method. Our model's algorithm identifies a particular object by analyzing its appearance in an image.

A computer vision approach called object detection aids in the identification and location of things in pictures and videos. This method of localizing and identifying allows for the counting of objects in a scenario, their accurate location and identification, and their naming. Ever notice how well Facebook recognizes your friends in your pictures? Previously, you had to click on the friend's profile and enter their names in order to tag friends in photos on Facebook. Facebook now automatically tags all of your friends on images as soon as you upload them. Face recognition is the term for this technique. After just a few instances of being tagged, Facebook's algorithms might be able to identify the faces of your friends. Facebook's 98% facial recognition accuracy is on par with human ability. People can be identified by their faces in photo and video broadcasts on social media and mobile devices.

The main goal is to develop a deep learning and facial recognition based model for attendance management specifically for the education sector in order to be able to update and improve the current attendance system to make it more effective and efficient than previously. There is a great deal of uncertainty in the outdated method, which makes it difficult and ineffective to record presence. When the government does not enforce the laws under the previous system, many challenges occur. A recognition system that recognizes faces will be the innovation. One of the most common physical characteristics that may be used to accurately identify a person is their face. Since it is uncommon for a face to diverge or duplicate, it is used to track identification. For this research, face databases will be constructed to supply the recognizer algorithm with data [9–11]. Following that, faces will be matched to those in the database to attempt to determine who they are within the period allocated for registering attendance. As soon as someone is identified, their attendance is promptly recorded, and the relevant data is entered into an excel file.

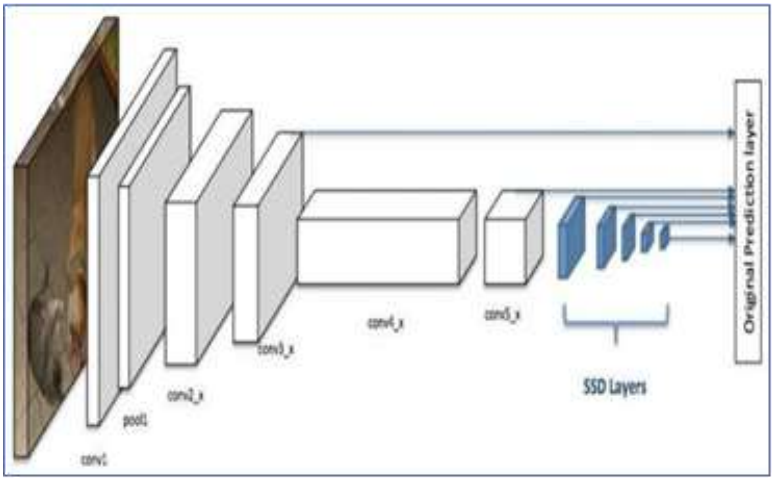

Fig-1. Figure of the System Architecture

2. LITERATURE SURVEY

The first picture recognition technology was made available in the 1980s. In the years that followed, a number of novel image processing technologies were developed. Object detection is crucial to many real-world applications, including image recovery and video surveillance. For instantaneous computing, the You Only Look Once (YOLO) system was created. Earlier recognition systems use classifiers or localizers again and again to find targets. On an image, they apply the model in various sizes and positions. High-scoring image segments are called detections. We take an entirely different tack. We use a single neural network to process the entire image. This network separates the image into regions, projects potential outcomes for each, and draws box boundaries. With the help of expected probability, these bounding boxes are weighted.

There are several benefits to this method over classifier-based systems. Its predictions are influenced by the total context of the image because it evaluates the complete image during testing. Moreover, it makes predictions using a single network assessment, unlike R-CNN, which requires numerous evaluations for a single image. As a result, it is 100 times faster than R-CNN and 1,000 times faster than Fast R-CNN. The YOLO network divides the input image into SS cells. the cell that is in charge of identifying the object. The B enclosing frame objectless value is projected for each grid cell in addition to their forecasts for their individual classes. The bounding box confidence score and the class prediction are combined into a single final score to determine the likelihood that this bounding box contains a specific type of item. YOLO v3 problems with minor items that appear in groups.

YOLO V3 is an object detector that uses features that a deep convolutional neural network has learned to recognize objects in real time. It uses a single neural network to process the entire image and has 75 convolutional layers with upsampling layers and connections skipped. Sections of the picture are created. Probabilities are shown with later boundary boxes. The main characteristic of YOLO V3 that stands out is its ability to do detections at three distinct scales. However, with YOLO v3, speed has been sacrificed for increased accuracy, and it struggles when dealing with little items that emerge in groups.

Faster R-CNN is composed of two networks: a region proposal network (RPN) that generates zone proposals and a framework for object detection based on these ideas. This method's primary difference from Fast R-CNN is that it uses selective search to produce region suggestions. Area recommendations are generated far faster by RPN than by focused screening when the majority of its computations are shared with the object identification structure. The area boxes, also called anchors, are ranked by RPN, which suggests the ones that have the highest likelihood of containing things. The Region Proposal Network creates regions and identifies objects using two fast RCNN algorithms. The first approach makes recommendations for the proposed regions and then uses them. Faster R-CNN's processing speed and challenging training procedure are two of its drawbacks.

3. METHODOLOGY

The main purpose of the OpenCV library or programming package is to help programmers learn computer vision. The term "OpenCV" refers to a set of free computer vision software that was created by Intel Corporation and released to the public in 1999 and 2000. (A library). The most widely used, renowned, and extensively documented computer vision library. Since the software is open-source, using it does not require a license. Most machine learning algorithms require numerical or quantitative inputs, as you are likely already aware. Even though OpenCV enables us to apply machine learning techniques to images, raw images typically require processing in order to convert them into features, which are columns of data. They are employed by and beneficial to our machine learning algorithms.

NumPy is a module for Python. The term "Numerical Python" describes a group of operations for manipulating arrays and multidimensional array objects. Numeric was created by Jim Hugunin and was the precursor to NumPy. Additionally, a new Num array package including a few new methods was created.

Since the project started in 2002, Drib has seen a considerable increase in the number of utilities. These include graphical user interfaces, networking, threading, and other modern software-intensive jobs. Recent research has focused a great deal of attention on the development of various probabilistic prediction techniques.

Pandas is an open-source tool that is quick, strong, flexible, and easy to use for analyzing and manipulating data. utilizing Python as the primary programming language.

The Python Imaging Library allows the Python interpreter to process images. This library provides a wide range of file format compatibility, a helpful internal representation, and quite powerful image processing features.

Classes can input and receive structured data in CSV format using the csv module. Developers do not need to be familiar with Excel's CSV format to provide instructions on how to write this data in the format that Excel prefers. The OS module in Python provides tools and methods for interacting with the operating system.

3.1. Detector for single shots (SSD)

The proposed approach takes advantage of an improved SSD algorithm to achieve more accurate and faster real-time detection. However, the SSD technique is not appropriate for small object detection because it ignores the backdrop from outside the boxes. To solve this issue, the proposed method uses both spatial and depth-wise separable convolutions in their convolutional layers. Specifically, our proposed solution blends a new design with a multilayer convolutional neural network. The algorithm consists of two stages. First, it reduces the extraction of spatial dimensions from feature maps by applying a resolution multiplier. Second, the optimal aspect ratio values for object detection are applied by the use of small convolutional filters in its construction. The major objective of training is to obtain a high confidence score.

Faster R-CNN builds boundary boxes using a region proposal network, which are then used to classify objects. Though considered state-of-the-art in terms of precision, the entire process runs at 7 frames per second, much below the requirements of real-time processing. SSD expedites the process by removing the requirement for the area proposal network. Multi-scale functionality and default boxes are two improvements added by SSD to offset the accuracy reduction. These enhancements allow SSD to operate with lower-quality images and still match the accuracy of the Faster R-CNN, which speeds up the process.

Comparing Single Shot Detector to earlier methods, it is far faster and more precise. We construct forecasts on different scales using feature maps of different dimensions, and then we split the forecasts based on aspect ratio to obtain high accuracy

These features allow for high accuracy to be attained even with low quality input photographs.

The object proposal methodology is widely used by other algorithms. Essentially, these algorithms create a way to segment an image and then make recommendations about possible item locations within those segments. Accuracy is sacrificed by these algorithms. The concept of "ground truth" is used to distinguish between actual or empirical evidence and conjectured evidence. We cannot simply train the algorithm if some boxes are missing; we must first detect them throughout the training process.

SSD will build the bounding boxes for each segment after it has divided the image into multiple segments. Next, it will search every box in the image for an object belonging to every class that the network has been trained to recognize. Lastly, a comparison between the anticipated and actual results will be made. If an error occurs after the comparison, it is back-propagated throughout the network to assist in updating the weights.

Similar to YOLO, a single shot detector uses a multi box and a single shot to find several things in an image. Its technique for detecting objects is more accurate and faster. a brief comparison of the accuracy and speed of the different object detection methods. Utilizing comparatively low-resolution images, the SSD's rapid speed and precision are enabled by the subsequent components:

Bounding box suggestions like the ones employed in RCNNs are eliminated.

A convolution filter with progressive loss is employed to take item classes and offsets in bounding box positions into account.

SSD achieves high object detection accuracy by using a large number of boxes or filters with various aspect ratios and sizes. This makes detection easier on different scales.

The data collection includes 300 distinct photos that were sourced from the Internet. In this project, the SSD model and algorithm will be used. Based on the object's numerous attributes, this will assist us in identifying it.

3.2. Data set description

Our collection of 300 images includes pictures of people, boats, bicycles, cows, bottles, and other objects. A webcam that records the Items in real time is used to evaluate our method. After pre-processing, some sample images are shown in the figure below.

Fig-2. Pictures in the Dataset

4. RESULTS AND ANALYSIS

Our suggested system involves the following steps.

Step 1: To take the photo for input, it uses the user's camera.

Step 2: The image is altered.

Step 3: Every necessary feature is removed from the image.

Step 4: It splits the image into smaller pieces so that more items may be recognized.

Step 5: After the objects have been segmented, attempt to classify and identify them.

Step 6: After that, the task of identifying objects inside the picture starts.

Step 7: The user is shown the output.

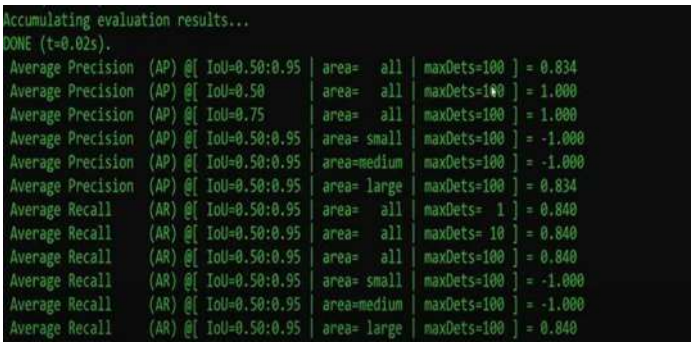

Fig-3. Accurate results

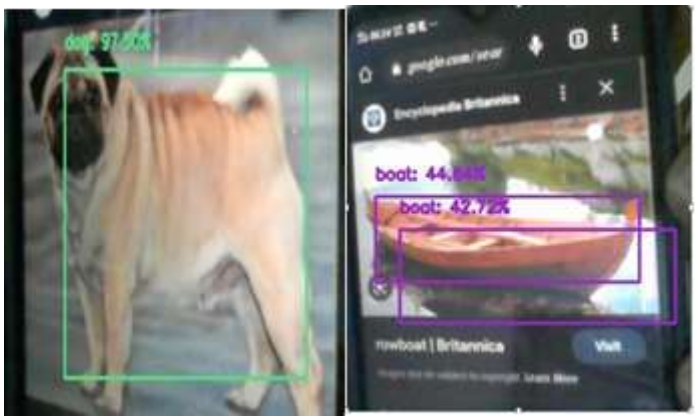

Fig-4. Results

In this instance, the object is recognized as a dog, and the object detection accuracy is 97.50%.

The object is recognized as a boat, and the boat's shadow is also discerned.

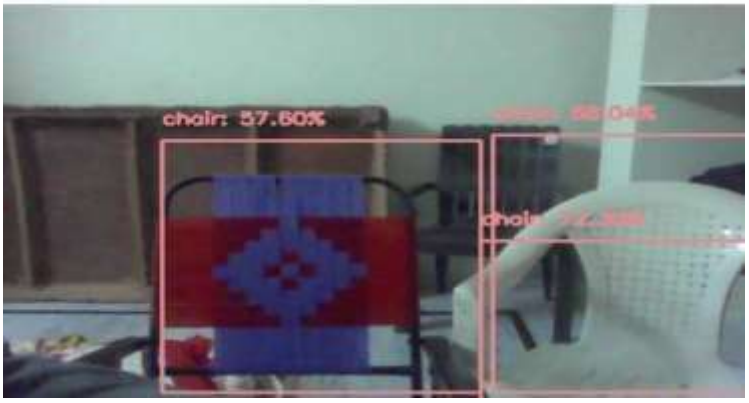

Fig-5. Chairs in output

This result demonstrates that this method is also capable of detecting several items. There are three chairs visible in this picture.

Fig-6. Bottle in output

Because the bottle is posing as a prominent product in front of the camera, the object is recognized as a bottle, and the background images are not identified.

4..1Testing types

Testing for accessibility: includes ensuring that your mobile and online apps work and are useful for users with a variety of disabilities, such as hearing loss, vision impairment, or other physical or mental challenges.

Adoption testing: Acceptance testing verifies that end users' ability to meet the goals specified in industry requirements can be used to determine whether software is appropriate for delivery. It goes by the name UAT (UAT) as well.

Testing black box: Testing a system with secret paths and code is referred to as "black box" testing.

Complete testing: End-to-end testing is a process that examines every phase of an application's workflow to ensure that everything works as it should.

Functional evaluation: The functioning of any software, website, or system is tested to make sure everything is working as it should.

Interactive examination: Testers can create and assist manual testing for individuals who do not use automation by using interactive testing, sometimes referred to as manual testing, which collects data from external tests.

Integrity checks: Integration testing makes ensuring an integrated system complies with a set of requirements. It is conducted in an integrated online and offline environment to guarantee optimal system performance.

4.2.Case studies

Case 1: We will test the system by watching one person at a time to see if it can identify.

Fig-8. Identified as a single person, the system will create a border and display the individual

Case 2: We will test the system to determine if it can recognize multiple objects at once.

Fig-8. Several objects were discovered

Consequently, objects that are visible within the camera's range of vision are recognized by the system.

5. CONCLUSION AND FUTURE SCOPE

The goal of this work is to develop an item recognizer for photos using deep learning. To recognize objects fast and precisely, the study makes use of a multilayer convolution network and an improved SSD approach. Our system manages both moving and stationary photos with ease. The suggested model produces more than 80% accurate predictions. Convolution neural networks use feature mapping to obtain the class label once feature data from the image has been removed. Our solution's main goal is to enhance SSD's object detecting capabilities by choosing default boxes with the most advantageous aspect ratios.

Similar to what happened during the first Industrial Revolution, object recognition technology has the potential to free humans from routine tasks that robots can complete more rapidly and effectively.

REFERENCES

1. Subhani Shaik, Ida Fann. Performance indicator using machine learning techniques, Dickensian Journal. 2022;22(6).
2. Vijaya Kumar Reddy R, Subhani Shaik B, Srinivasa Rao. Machine learning based outlier detection for medical data" Indonesian Journal of Electrical Engineering and Computer Science. 2021:24(1).
3. Dong J, Li H, Guo T, Gao Y. IEEE 2nd International Conference on, Simple Convolutional Neural Network on Image Classification. Conf. Using Big Data. 10.1109/ICBDA.2017. 8078730, p. 721–724 in ICBDA; 2017.
4. 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
5. Item Detection and Recognition in Pictures, Sandeep Kumar, Aman Balyan, and Manvi Chawla, IJEDR. 2017;1-6.
6. Available:https://jwcneurasipjournals.springeropen.com/articles/10.1186/s13638-020-01826-x.
7. Subhani Shaik, Ganesh. Taming an autonomous surface vehicle for path following and collision avoidance using deep reinforcement learning, Dickensian Journal. 2022;22(6).
8. Vijaya Kumar Reddy R, Shaik Subhani, Rajesh Chandra G, Srinivasa Rao B. Breast Cancer Prediction using Classification Techniques, International Journal of Emerging Trends in Engineering Research. 2020;8(9).

9. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014;580-587.

10. Dai J, Li Y, He K, Sun J, R-fcn: Object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems. 2016:379-387.