Journal of Advances in Mathematics and Computer Science

Volume 38, Issue 8, Page 24-32, 2023; Article no.JAMCS.101284
ISSN: 2456-9968
(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Real-time Object Detection Using Deep Learning

K. Vaishnavi [a], G. Pranay Reddy [a], T. Balaram Reddy [a], N. Ch. Srimannarayana Iyengar [a] and Subhani Shaik [a*]

[a] Department of IT, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMCS/2023/v38i81787

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/101284

Original Research Article

Received: 03/04/2023
Accepted: 05/06/2023
Published: 15/06/2023

Abstract

As technology improved, object detection, which is connected to video and image analysis, caught researchers' interest. Earlier object recognition techniques are based on hand-crafted features and imprecise architectures and trainable algorithms. One of the main issues with many object detection systems is that they rely on other computer vision methods to support their deep learning-based methodology, which leads to slow and subpar performance. In this article, we present an end-to-end solution to the object detection problem using a deep learning based method. The single shot detector (SSD) technique 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 enhance accuracy of SSD method.

Keywords: Object detection; SSD method; deep learning.

**Corresponding author: Email: shaiksubhani@sreenidhi.edu.in;*

1 Introduction

Image classification, which is defined as figuring out the class of the image, was one of the essential issues. The challenge of image localization, when one item is present in the picture and the system must predict its class and position within the image, is rather challenging (a bounding box around the object). The fact that objects discovery includes both identification and localisation makes it a more challenging challenge. In this instance, an image will be used as the system's input, and the output will be a bounding box that corresponds to every object in the image and specifies the type of object in each box. We built a solution that uses less processing power than the existing techniques while operating at enhanced FPS and fast object detection [1,2]. The SSD mobile net method is used by our object discovery model to identify and celebrate the item in the image. The algorithm in our model analyses appearance existing in an image to pinpoint a specific object.

Object detection is a computer vision technique that helps identify and locate objects in images and movies. With this form of identifying and localizing, detection of objects may be used to count the items in a scenario, locate and identify them precisely, and name them. Have you ever noticed how adeptly Face book can recognise your pals in your photos? In order to tag friends in photographs on Face book, you used to have to click on the friend's profile and enter their names [3-5]. These days, Face book automatically tags everyone in your photos as soon as you upload them. This method is known as face recognition. Face book's algorithms may recognise your friends' faces after just a few times of being tagged. Face book has a facial detection accuracy of 98%, which is comparable to human performance. Faces in picture and video streams on social media and mobile devices may be used to recognise people [6-8].

To be able to update and improve the current attendance system to make it more effective and efficient than before, the main aim is to create a deep learning and facial recognition based model for attendance management especially for education sector. The outmoded method has a lot of uncertainty, which leads to incorrect and unproductive way of recording the presence. Different obstacles arise when the government does not enforce laws under the old system. The innovation will be a face based recognition system. The face is a most used physical trait that may be utilised to precisely identify a person. A face is used to track identity since it is rare that it would diverge or be duplicated. Face databases will be created for this project in order to provide data to the recognizer algorithm [9-11]. After that, during the time allotted for recording attendance, faces will be compared to those in the database to try to identify who they are. A person's attendance is immediately logged when they are recognised, recording the pertinent information onto an excel file.

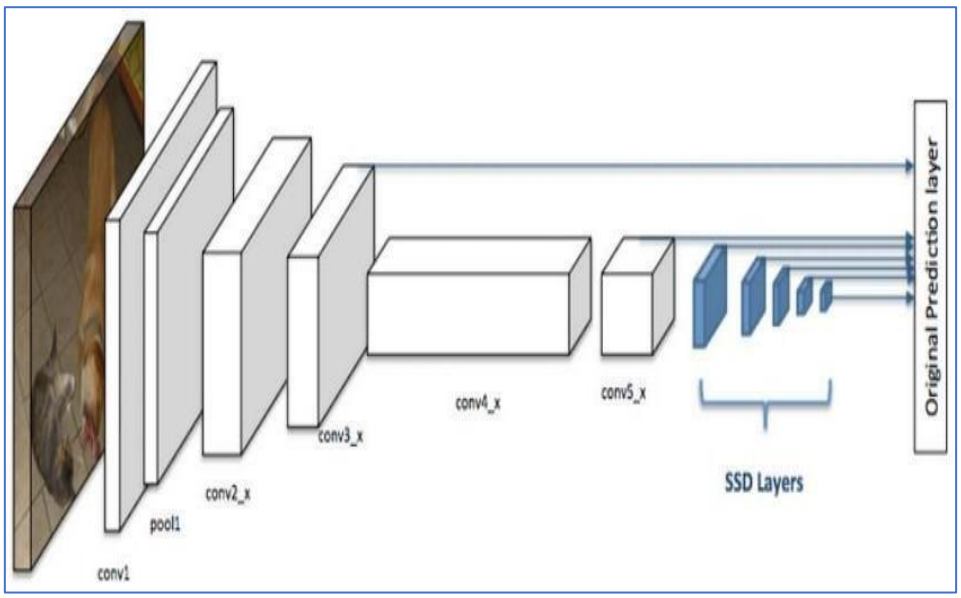

Fig. 1. Figure of the System Architecture

2 Literature Survey

In the 1980s, image recognition technology first became available. Following then, several new technologies in the area of image processing emerged. Several real-world applications, such as picture recovery and video surveillance, heavily rely on object detection. The system - You only look once (YOLO) is designed for instantaneous computing. Previous recognition systems find targets by reusing localizers or classifiers. They apply the model at different locations and sizes on a picture. Image segments with high scores are referred to as detections. We adopt a totally different approach. For processing the full image, we employ a single neural network. This network divides the image into regions and forecasts possibilities as well as box boundaries for each. These bounding boxes are weighted using anticipated probability [12-14].

Compared to classifier-based systems, this approach offers significant advantages. Since it assesses the entire picture while testing, its predictions are informed by the image's overall context. Additionally, it predicts with just one network evaluation as opposed to R-CNN, which needs thousands for a single image. This makes it a hundred and a thousand times quicker than Fast R-CNN and R-CNN, respectively. The input picture is divided into SS cells by the YOLO network. The cell responsible for detecting the object's presence. In addition to their predictions for their respective classes, the B enclosing frame objectless value is predicted for every cell on the grid. The chance that this bounding box includes a certain kind of item is finally determined by combining the bounding box confidence score and the class prediction into a single final score. With little things that emerge in groupings, YOLO v3 struggles.

YOLO V3 is a detector of objects which makes use of features learned by a deep convolutional neural network for detecting object in real time . It consists of 75 convolutional layers with up-sampling layers and skips connections for the complete image one neural network being applied. Regions of the image are made. Later bounding boxes are displayed along with probabilities. The most noticeable feature of YOLO V3 is that the detections at three different scales can be done with the help of it. But the speed has been traded off for boosts in accuracy in YOLO v3, and it does not perform well with small objects that appear in groups.

Faster R-CNN consists of two networks: a framework for object detection based on these concepts and a region proposal network (RPN) for producing zone suggestions. The main distinction between this approach and Fast R-CNN is that it generates region suggestions via selective search. When RPN shares the majority of its computations with the object identification structure, area recommendations are produced in a lot less time than they would in targeted screening. RPN ranks the area boxes, also referred to as anchors, and recommends the ones that are most likely to contain items. Two rapid RCNN algorithms are used by the Region Proposal Network to create regions and identify objects. The first method uses the suggested regions after making recommendations for them. A limitation of faster R-CNN is that it has a difficult training process and a poor processing speed.

3 Methodology

A library or programming package called OpenCV was created primarily to assist programmers in learning about computer vision. OpenCV is an abbreviation for free computer vision software, and the package was developed by Intel Corporation and made accessible to the public between 1999 and 2000. (Library). The most popular, well-known, and well documented library for computer vision. As the programme is open-source, there is no licencing required to use it. As is probably previously known, the bulk of machine learning algorithms require numerical or quantitative inputs. Despite the fact that OpenCV makes it possible for us to apply machine learning techniques to pictures, the raw images are usually need to be processed in order to transform them into features (columns of data). They benefit our machine learning algorithms and are utilised by them

NumPy is a Python package. The name "Numerical Python" refers to a collection of procedures for working with multidimensional array objects and arrays. Jim Hugunin developed Numeric, which was the forerunner to NumPy. There was also the creation of another Num array package with a few new methods.

The quantity of utilities in Dlib has increased significantly since the project's inception in 2002. These include networking, threading, graphical user interfaces, and other tasks that require the use of software today. The development of numerous probabilistic predictive methods has received significant attention in recent study.

Pandas are a rapid, powerful, adaptable, and user-friendly open-source programme for data analysis and manipulation. Using the Python programming language as a foundation.

The Python Imaging Library gives the Python interpreter the capacity to process pictures. This library provides a broad variety of file format compatibility, a helpful internal representation, and rather powerful image processing tools.

The csv module allows classes to receive as well as input structured data in the format of an CSV. Write this data in the format preferred by Excel, developers can instruct without knowing the details of Excel's CSV format. Python's OS module offers ways and resources for working with the operating system.

3.1 Detector for single shots (SSD)

The suggested method makes advantage of an upgraded SSD algorithm for faster real-time detection with increased precision. However, because it ignores the background from outside the boxes, the SSD technique is not suitable for detecting small objects. The suggested technique employs depth-wise separable convolutions and spatial separable convolutions in their convolutional layers to address this problem. In particular, our suggested method combines a multilayer convolutional neural network with a new design. There are two phases to the algorithm. By utilising a resolution multiplier, it first decreases the extraction of spatial dimensions from feature maps. Second, it is built with the use of tiny convolutional filters that apply the best aspect ratio values for object detection. During training, the main goal is to achieve a high confidence score.

A region proposal network is used by faster R-CNN to construct boundary boxes, which are subsequently used to categorise objects. The entire process operates at 7 frames per second, which is significantly below the criteria of real-time processing even though it is thought to be cutting-edge in terms of precision. By eliminating the need for the area proposal network, SSD speeds up the procedure. To compensate for the accuracy decline, SSD adds a few enhancements including default boxes and multi-scale functionality. With the help of these improvements, SSD can now equal the accuracy of the Faster R-CNN while working with pictures of lesser quality, which accelerates the process.

Single Shot Detector is a great deal quicker and accurate than previous approaches. Using feature maps of various dimensions, we generate forecasts on various scales, and then, in order to achieve high accuracy, we separate the forecasts based on ratio of aspect.

High accuracy is achieved even with input photos of low quality because to these properties.

Other algorithms frequently employ the object proposal methodology, in which they devise a mechanism to divide the image into segments and provide suggestions about where those segments may potentially be objects. These algorithms forfeit accuracy. A notion known as "ground truth" separates actual or empirical evidence from assumed evidence. If certain boxes are absent, we cannot just train the algorithm; we must first identify them throughout the training process.

The bounding boxes for each segment will be created by SSD once it has divided the picture into many pieces. After that, it will look through each box on the picture for an object from each class that the network has been trained for. Finally, it will make a comparison between the predicted and actual outcomes. After the comparison, if there is an error, it is back- propagated over the network to help update the weights.

A single shot detector, like YOLO, utilises a multi box with just one shot to find several items in an image. Its object detection technology is quicker and more precise. a quick comparison of the various object detecting

techniques' speed and precision. When employing relatively low-resolution photos, the SSD's fast speed and accuracy are facilitated by the following elements:

The elimination of bounding box suggestions, such those used in RCNNs, is achieved.

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

High detection accuracy for objects is achieved in SSD by employing a large number of boxes or filters with different sizes and aspect ratios. This facilitates detection on various scales.

There are 300 different photographs in the data collection, which comes from the Internet. We will apply the SSD algorithm/model in this project. This will help us identify the object based on its many characteristics (Depends on training).

3.2 Data set description

Among the 300 photos in our collection are depictions of a boat, a bicycle, a cow, a human, a bottle, etc. Our technique is examined using a real-time web camera that records the Items. Following pre-processing, the figure below displays a few examples of pictures.

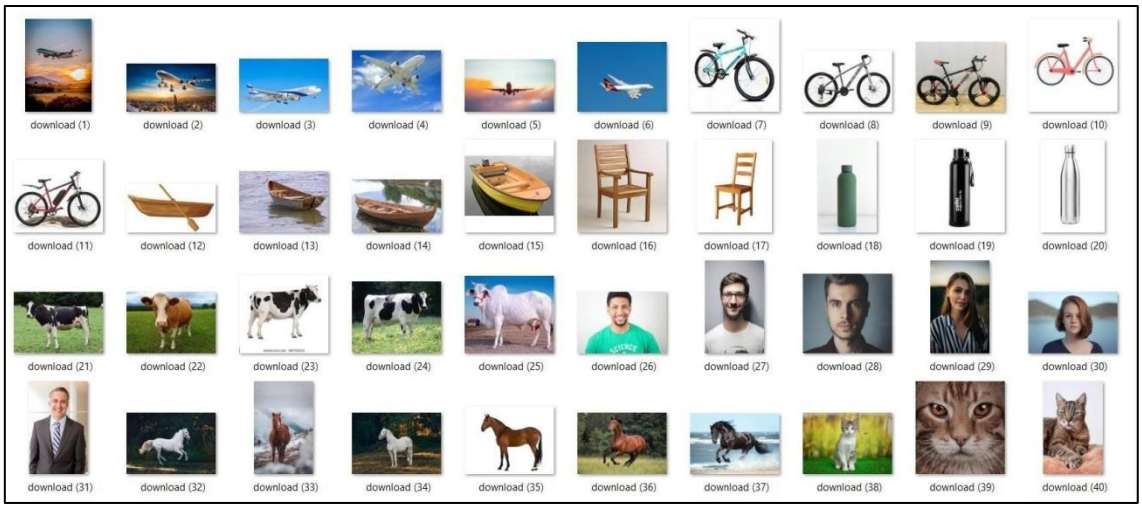

Fig. 2. Pictures in the Dataset

4 Results and Analysis

The following steps are involved in our proposed system.

Step-1. It uses the user's camera to capture the picture as input.

Step-2. It transforms the picture.

Step-3. It takes all the required features out of the picture.

Step-4. To recognise more objects in the image, it divides it into smaller bits.

Step-5. Try to categorise and identify the objects after segmenting them.

Step-6. Then the process of finding things in the image begins.

Step-7. It shows the output 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 Dog

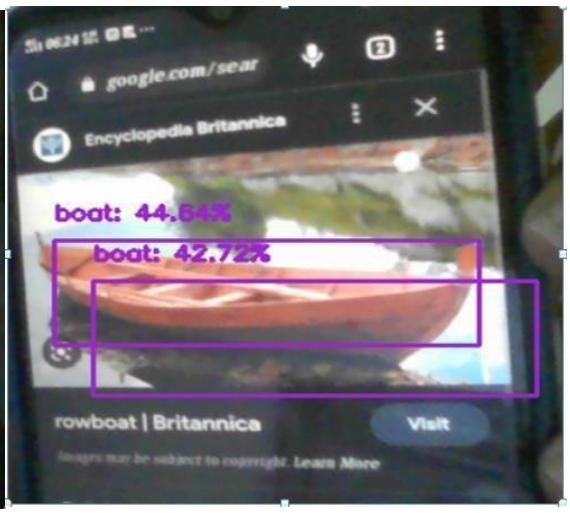

Fig. 5. Product Boats

The object is identified as a dog and the accuracy of object detection in this case is 97.50%.

The object is identified as a boat and the shadow of the boat is also detected.

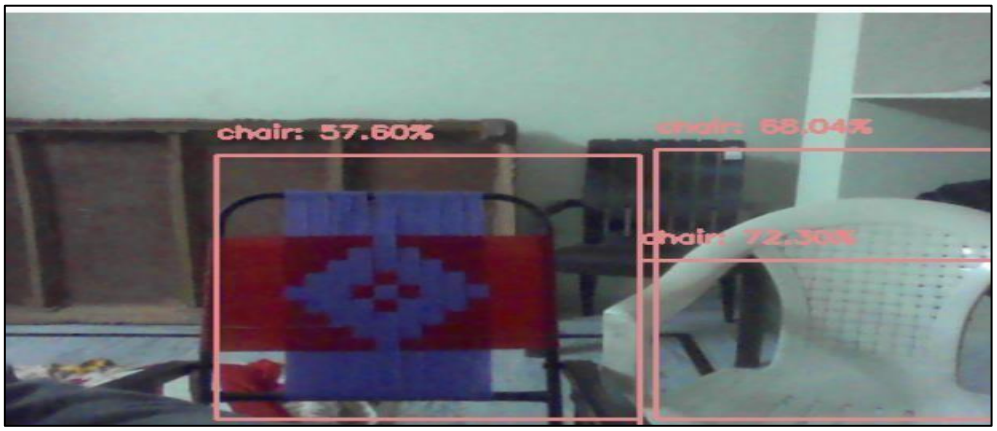

Fig. 6. Chairs in output

This outcome is an example that proves multiple objects can also be detected by this algorithm. In this image three chairs are detected.

Fig. 7. Product bottle

The object is identified as a bottle and the background images are not identified because the bottle is acting as a major product in front of camera.

4.1 Testing types

Testing for accessibility: Making sure your mobile and web apps are functional and usable for both people who have impairments including visual impairment, hearing loss, and other physical or mental difficulties is known as testing for accessibility.

Adoption testing: Acceptance testing makes sure that it is possible to assess whether the software is suitable for delivery by looking at how well end users are able to accomplish the objectives outlined in the industry specifications. Also, it is known as UAT (UAT).

Testing a black box: The term "black box" testing refers to testing a system against which routes and code are hidden.

Complete testing: A technique called end-to-end testing looks at each stage of an application's workflow to make sure everything functions as it should.

Functional evaluation: Software, website, or system's functionality is tested to ensure that it is operating as it should.

Interactive examination: Through interactive testing, also known as manual testing, testers can develop and support testing manually for people who don't use automation and gather data from exterior tests.

Integrity checks: An integrated system's compliance with a set of criteria is ensured through integration testing. To ensure proper system function, it is carried out in an online and offline environment that is integrated.

4.2 Case studies

Case 1: To see if the system can identify, we will test it by observing just one individual at a time.

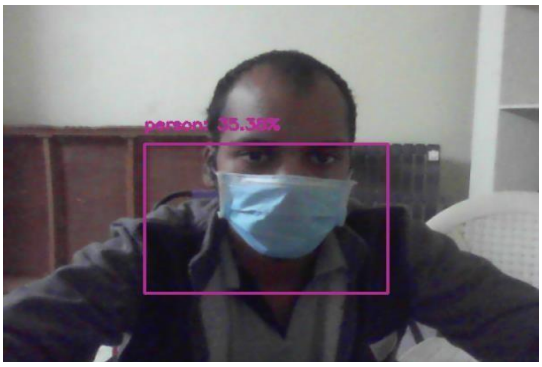

Fig. 8. Identified as a single person, the system will create a border and display the individual

Case 2: To see if the system can recognise many things at once, we will put it to the test.

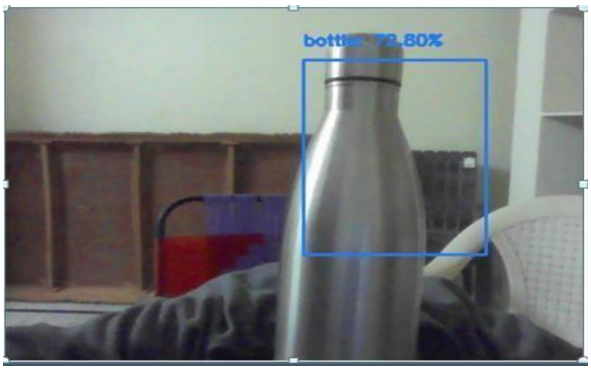

Fig. 9. Several objects were discovered

As a result, the system identifies items that are visible in the camera's field of view.

5 Conclusion and Future Scope

This study creates a deep learning-based item recognizer for identifying objects in images. The study uses an upgraded SSD technique and a multilayer convolution network to recognise things quickly and accurately. Both still images and moving images are handled well by our technology. More than 80% of the predictions made by the proposed model are correct. After removing feature data from the image, convolution neural networks employ feature mapping to get the class label. The major objective of our solution is to improve SSD's object detection process by selecting default boxes with the best feasible aspect ratios.

Object identification technology has the ability to relieve people from regular jobs that robots can carry out more quickly and efficiently, much like the first Industrial Revolution did. This technology is now being tested.

Competing Interests

Authors have declared that no competing interests exist.

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] Towards Data Science.
Available:https://towardsdatascience.com/ssd-single-shot-detector-for-objectdetection-using-multibox-1818603644ca? gi=f02e06e2d636

[7] Available:https://jwcneurasipjournals.springeropen.com/articles/10.1186/s13638-020-01826-x.

[8] Subhani Shaik, Ganesh. Taming an autonomous surface vehicle for path following and collision avoidance using deep reinforcement learning, Dickensian Journal. 2022;22(6).

[9] 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).

[10] 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.

[11] 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.

[12] Ren S, He K, Girshick R, Sun J, Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems. 2015;91-99.

[13] 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.

[14] Jeong J, Park H, Kwak N. Enhancement of SSD by concatenating feature maps for object detection. arXiv preprint arXiv:1705.09587; 2017.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/101284