Object Detection in Deep Learning

|  |
| --- |
| Sujeet Suresh Kumbhar |
| *Department of MCA* |
| *KLS Gogte Institute of Technology* |
| *Belgaum, Karnataka* |
| [2gi20mc076@students.git.edu](mailto:2gi20mc076@students.git.edu)  Sujeetkumbhar03@gmail.com |

**Abstract**:

Due to object discovery’s affinity with program reasoning and image understanding, it has gained much research consideration in recent age. Traditional object discovery systems are built on not foreign looks and shallow educable architectures. Their acting easily stagnates by building complex collections that combine diversified reduced-level image countenance accompanying high-level framework from object detectors and setting classifiers. With the fast development in deep knowledge, more effective tools, that are smart to learn pertaining to syntax, high-ranking, deeper appearance, are introduced to address the questions existent in traditional architectures. These models function otherwise in network architecture, preparation game plan and addition function, etc. In this paper, we provide a review on deep education located object detection foundations. Our review starts with a brief establishment on the annals of deep knowledge and its representative form, that is to say Convolutional Neural Network (CNN). Then we focus on conventional general object detection architectures in addition to few modifications and valuable tricks to improve discovery efficiency further. As distinct particular discovery tasks exhibit different traits, we still concisely survey several particular tasks, containing salient object discovery, face discovery and pedestrian detection. Experimental studies are further determined to compare miscellaneous patterns and draw some significant ends. Finally, several hopeful guidance and tasks are determined to serve as directions for future introduce both object discovery and appropriate neural network located education schemes.

**Introduction:**

Deep learning technology has become popular nowadays due to the state-of-the-art obtained in image classification, object detection, natural language processing. Object detection is the

process of identifying the instance of the class to which the object belongs and localize the location of the object by the bounding box around the object. For object detection, deep CNNs have been widely used. CNN is a sort of

feed-forward neural network that operates on the weight-sharing concept. Convolution is a blend of two functions multiplied, and it is an integration that shows how one function

overlaps with another. The layered architecture of CNN for object recognition is shown in Figure 1. detection. To create feature maps, the image is convolved with the activation function. to minimise the spatial complexity to generate abstracted feature maps, network feature maps are handled with pooling layers. This procedure is done for the second time. The desired number of filters are applied, and feature maps are formed as a result. These feature maps are eventually processed.

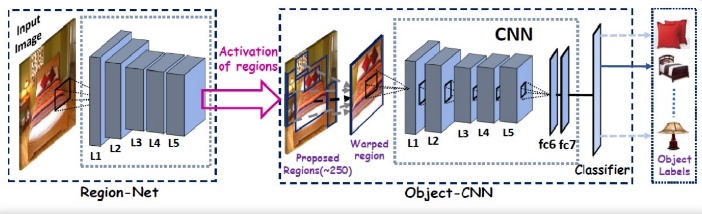


Figure 1: Use of CNN for object detection.

Convolutional neural networks work in a similar way to classic supervised learning approaches in that they take in input images, detect their features, and then move a grader over them. Features, on the other hand, are learned automatically! All of the arduous work of extracting and describing features is done by the CNN itself: during the training phase, the classification error is minimised in order to optimise the classifier's parameters as well as the features.

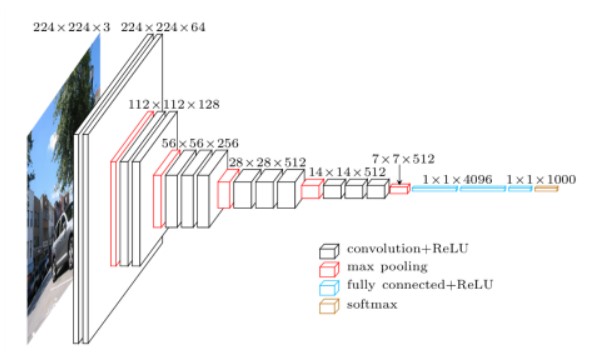


Figure 2: Architecture of CNN

There are multiple layers present in the convolution neural network as they help to detect object more accurately and filters the image to detect objects with more accuracy.

To identify any objects there are several types of image annotation methodologies available:

**Bounding boxes**: The most prevalent type of annotation in computer vision is bounding boxes. Bounding boxes are rectangular boxes that define where the target object is located. The x and y axis coordinates in the upper-left corner and the x and y axis coordinates in the lower-right corner of the rectangle can be used to determine them. Object recognition and localization tasks frequently employ bounding boxes.



Figure 3: Bounding Box

Polygonal: Objects don't necessarily have to be rectangles. Polygonal segmentations, based on this concept, are a sort of data annotation in which complicated polygons are employed instead of rectangles to characterise the shape and location of an object with greater precision.



Figure 4: Polygonal labelling

3D cuboids: Similar to bounding boxes, 3D cuboids contain additional depth information about the item. As a result, 3D cuboids can be used to generate a 3D representation of an item, allowing systems to identify properties such as volume and position in 3D space.



Figure 5: Cuboid Shape

**Different Methodologies:**

There are some algorithms available to detect any object.

* [Fast R-CNN](https://analyticsindiamag.com/top-8-algorithms-for-object-detection/#h-1-fast-r-cnn)
* [Faster R-CNN](https://analyticsindiamag.com/top-8-algorithms-for-object-detection/#h-2-faster-r-cnn)
* [YOLO (You Only Look Once)](https://analyticsindiamag.com/top-8-algorithms-for-object-detection/#h-8-yolo-you-only-look-once)
* SSD (Single Shot Detector)

1.FastR-CNN: Fast Region-Based Convolutional Network Method, or Fast R-CNN, is a training approach for object detection written in Python and C++ (Caffe). This algorithm primarily addresses the shortcomings of R-CNN and SPPnet while also increasing their speed and accuracy.

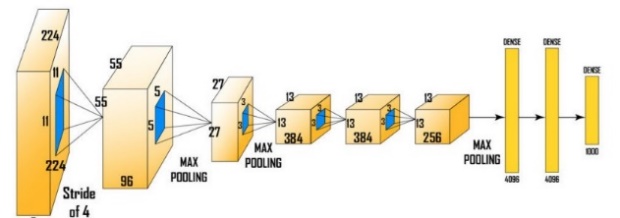


Figure 6: Architecture of R-CNN

2.Faster R-CNN: Faster R-CNN is a similar object detection algorithm to R-CNN. This approach uses the Region Proposal Network (RPN), which is more cost-effective than R-CNN and Fast R-CNN at sharing full-image convolutional features with the detection network. A Region Proposal Network is a fully convolutional network that predicts object limits and object scores at each position of the object and is trained end-to-end to generate high-quality region proposals, which are then employed by Fast R-CNN for object detection.

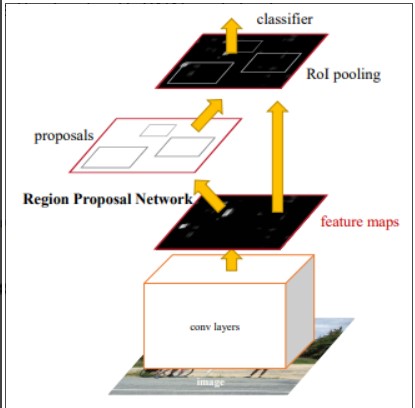


Figure 7: Architecture of Faster R-CNN

3. YOLO: The YOLO algorithm takes a different method. A single neural network is applied to the entire image using the method. The network then divides the image into areas, generating bounding boxes and predicting probabilities for each. The projected probabilities are used to weight the generated bounding boxes.

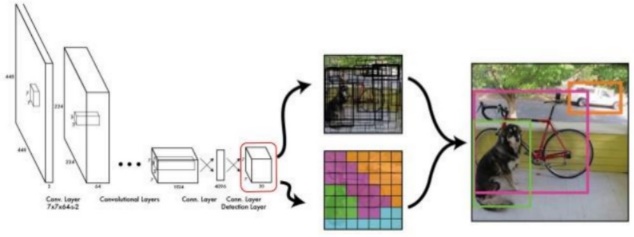


Figure 8: YOLO Object Detection

The YOLO neural network is a convolutional neural network. There are 24 convolutional layers in total, followed by two fully linked layers. Each layer is significant in its own right, and the layers are distinguished by their functions

4. SSD: SSD is intended for real-time object detection. Faster R-CNN creates boundary boxes using a region proposal network and then it uses those boxes to classify things. The entire process operates at 7 frames per second, which is regarded state-of-the-art in terms of precision. Far less than what real-time processing needed. By removing the requirement for the region proposal network, SSD speeds up the procedure. SSD implements a number of improvements, including multi-scale features and default boxes, to make up for the reduction in accuracy. These improvements allow SSD to match the accuracy of the Faster

R-CNN utilising lower resolution images, increasing the speed even further. It achieves real-time processing speed and even surpasses the accuracy of the Faster R-CNN, according to the following comparison.

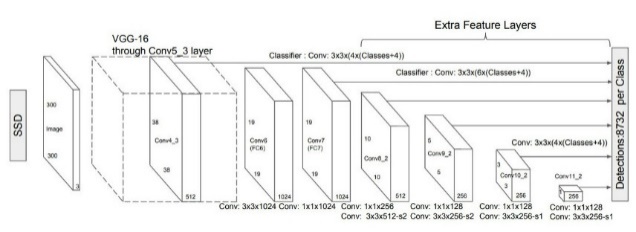


Figure 9: Single Shot Multi Box Detector

The SSD object detection composes of 2 parts:

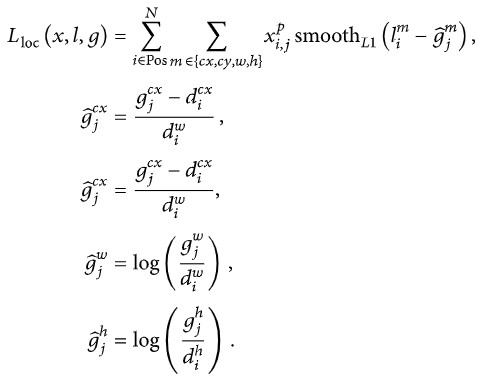
* Extract feature maps, and
* Apply convolution filters to detect objects.

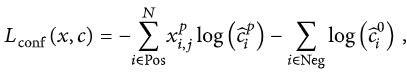
The SSD training method is derived from the Multi Box approach, but it can handle a variety of object classes. Assume that there is an indicator for matching the i-th default box to the j-th ground truth box of class. We can write using the pairing method described above. The weighted sum of the localization loss (loc) and the confidence loss (conf) is the long-term loss function:



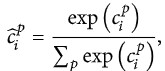
where *N* is the number of paired default boxes. If *N* = 0, set the loss to 0.

Small item detection relies on higher-resolution feature maps. The first object detection layer, conv4 3, has a spatial dimension of 38 38, which is a significant reduction from the input image. As a result, as compared to other detection algorithms, SSD typically performs poorly for small objects. If this is a problem, we can alleviate it by using higher-resolution images.

The loss of confidence is the soft max on multi classes confidence (*c*):

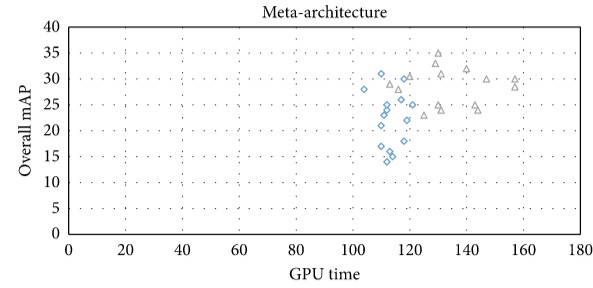


Where,



**Time and Accuracy:**

The average MAP of each meta–architecture is represented in the scatter plot (Figure 6). The picture depicts the average running time of each image, which ranges from a few hundred milliseconds to 200 milliseconds. In general, SSD meta-architectures are faster but have lower accuracy, while faster R-CNN meta-architectures are more accurate but take at least 130 milliseconds to process one image.

Figure 10: time and accuracy

The Above factors, richer and correct representations of an object can promote detection accuracy remarkably. Brain-inspired  
mechanism is a powerful way to further improve detection performance.

**Application Domains of Object Detection:**

Object detection has applications in a wide range of fields, including defence(surveillance), human-computer interaction (HCI), robotics, transportation, and retrieval. In just a few hours, sensors used for long-term surveillance generate petabytes of picture data. These data are analysed and combined with other data to provide a clear picture of the current situation. scenario. Object detection is used in this procedure to track entities such as people, cars, and suspicious objects. The data from the raw imagery. Detecting and spotting wild animals in sterile zones such as industrial zones. Object detection has other applications, such as detecting vehicles parked in restricted locations.

**Application and Branches:**

Object detection has been widely employed in a variety of industries to assist individuals in completing activities, including security, military, transportation, medical, and life.

In this paper, we go over the most common and latest methodologies used in these disciplines.

1) FIELD OF SECURITY:

Face detection, pedestrian detection, and fingerprint identification are some of the most well-known security applications. Anomaly detection, fraud detection, and so on.

Face detection is the process of identifying people's faces in a crowd. Because of the extreme positions, illumination is required Face detection is still difficult due to resolution differences and other factors. a challenging mission Many studies concentrate on the development of an accurate detector.

2) ANAMOLY DETECTION:

Anomaly detection is a type of object detection that is best taught with real-world examples from various industries. A custom object recognition model in agriculture, for example, might properly identify and pinpoint probable cases of plant disease, allowing farmers to discover dangers to their crop yields that would otherwise go undetected by the naked eye. Object detection could also be utilised in health care to help treat illnesses with specific and unique clinical lesions. Skin care and the

treatment of acne are two examples of this. An object detection model might locate and identify cases of acne in seconds.

3)SELF-DRIVING CARS:

The success of autonomous vehicle systems is dependent on real-time car detection models. In order to travel around the world safely and effectively, these systems must be able to detect, find, and track items surrounding them. Object detection is a foundational task that underpins current work on making self-driving cars a reality. While tasks like image segmentation can (and often are) applied to autonomous vehicles, object detection remains a foundational task that supports current work on making self-driving cars a reality.

**Conclusion:**

The first phase in the implementation of self-driving cars and robotics is object detection. We decoded the role of deep learning algorithms based on CNN for object detection in this work. The paper also discusses the various deep learning frameworks and services available for object detection. Because of its vast range of applications, correctly detecting an item in a surveillance video is one of the most important research areas in computer vision. Processing an image received from a surveillance camera is difficult due to the following factors: poor resolution, lighting volatility, active objects in the background, and minor changes in the background such as leaf waving. An overview of current advances in object detection algorithms has been presented. The detection process occurs in background modelling, object detection, and object classification, and all available object detection techniques are categorized into background subtraction, optical flow, and spatiotemporal filter methods, with the benefits and drawbacks of each method discussed in relation to different types of datasets.

**References:**

[1] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. (2017). “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” IEEE transactions on pattern analysis and machine intelligence,

[2] J. Jeong, H. Park, and N. Kwak, ``Enhancement of SSD by concatenating

feature maps for object detection,'' 2017, *arXiv:1705.09587*. [Online].

Available: <https://arxiv.org/abs/1705.09587>

[3]Application of Deep Learning for Object Detection/sciencedirect/

[4] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan,

P. Dollár, and C. L. Zitnick, ``Microsoft COCO: Common objects in

context,'' in *Computer Vision\_ECCV*, D. Fleet, T. Pajdla, B. Schiele, and

T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740\_755.

[5] S. Ren, K. He, R. Girshick, and J. Sun, ``Faster R-CNN: Towards real-time

object detection with region proposal networks,'' *IEEE Trans. Pattern*

*Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137\_1149, Jun. 2017.

[6]<https://docs.microsoft.com/en-us/cognitive-toolkit/>

[7]<https://towardsdatascience.com/image-data-labelling-and-annotation-everything-you-need-to-know-86ede6c684b1>

[8] <https://www.analyticsvidhya.com/>

[9]<https://www.geeksforgeeks.org/introduction-deeplearning>

[10] https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06