Object Detection Using Convolutional Neural Networks

Sanskriti Labroo   
*Coding*  
*RoboManipal*MIT, Manipal  
sanskritilabroo@gmail.com

*Abstract*—This paper is a compiled review of the work of several researchers in the field of object detection using convolutional neural networks (CNN). In recent years, the stellar performance of CNN’s as efficient state-of-the art object detection algorithms has grown exponentially and procured viable results. The objective of this paper is to explore and compare two of the most widely recognized detector models, that is the Single Shot Multi-Box Detector (SSD) and a Faster Region-based Convolutional Neural Network (Faster-RCNN) based on past research and experimental data. These models enable object detection by utilizing principles of transfer learning, selective search and image classification. Results are obtained on the basis of speed and accuracy received from several parameters such as mean average precision and minimum loss.

Keywords—convolutional neural networks, transfer learning, selective search, image classification

# INTRODUCTION

Computer vision is a rapidly evolving field in artificial intelligence that has several applications in robotics. Robots can be equipped with computer vision systems that allow them to navigate complex environments and locate objects with ease. Using the position of these objects, these robots can then carry out other tasks they have been programmed to do. Object detection has been effectively carried out with several algorithms such as background subtraction and frame differencing in the past. However, deep-learning algorithms such as R-CNN and SSD have now proven that machine learning is the most ideal tool presently for real-time object detection and localization. This forms an integral component of computer vision and can have a significant impact on a robot’s performance.

## Transfer Learning

These researchers have also used the concept of transfer learning in their approach to these algorithms wherein pre-existing knowledge gained from training a model on one task can be used to train the same model on a different task, but with improved accuracy and performance. The first task is thus used as a starting point for the second and this allows models to learn faster with limited features and less data, while still achieving high levels of success.

The idea stems from the history of machine learning that shows similarity in features and patterns across several models. At it’s core, it is seen that tasks often have recurring themes, albeit their functionality might be different. In papers like [1] and [2], pre-trained models are used from widely available large image databases such as ImageNet, InceptionV2 and Mobile Net specifically.

1. Transfer learning procedure

Transfer learning is often seen in image classification where machine learning algorithms are used to identify the presence of things within a picture and set them into separate classes.

## Tensorflow

TensorFlow is a free, open-source library used in machine learning and artificial intelligence, particularly deep learning of neural networks. It provides several models and resources that can be used for experimental purposes as done by researchers in the cited papers.

## Selective Search

Selective search is a well-known algorithm used for generating regions of interest in object detection and image segmentation tasks. It is used to group together neighboring pixels based on similarity in color or shape and form regions that may represent or classify as a feature of an object.

This algorithm starts off with identifying large areas that may be objects and then narrows down it’s search selectively to more specific locations for objects, recursively combining them to get exact precision. It follows a continuous process of finding the similarity between any two regions and the following two regions until the whole object is covered in a single region using bounding boxes.

# CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks are a class of artificial neural networks in deep learning that specializes in processing data and is especially focused on being utilized in visual imagery and image recognition.

These networks are designed to inherently learn and train based on input data that contains a variety of features stored in a matrix-like format. They perform similar tasks as regular neural networks, but with more depth and complexity. Essentially, a neural network consists of multiple layers that each perform some type of activation function on the given data to obtain certain required computations each time it is processed. Likewise, a CNN adopts a volumetric form for the layers with divisions for height, width and depth. The input is an image with pixels stored mathematically and often representing it’s three dimensions in a [h][w][d] format.

1. CNN Architecture

Each time it is processed, the convolutional layers are responsible for learning features from the input image by applying a set of kernel filters. Kernel filters are custom matrices that perform a dot product with a sub-region of the input data and get the output as a matrix of dot products. They are used to produce feature maps which can further help with image feature classification. After filtering, pooling is performed to down sample the feature map and only take the most critical and necessary features through a maximum pooling approach. Once a series of filtering and pooling layers are applied, we can get a two-dimensional output image that is known as a fully connected layer, much resembling that of a neural network. This is then classified to locate objects through selective search with the help of object detection models.

Due to it’s multiple layers, CNN’s are able to understand increasingly abstract and specific parts of input images like never before. They have many advantages and continuously deliver superior outputs in the field of deep learning.

Top of Form

# OBJECT DETECTION MODELS

## Faster-RCNN

Introduced in 2014 by Ross Girshick, RCNN is one of the first and largest most successful applications of CNN’s in object detection. Considered as a multi-stage detector, the R-CNN utilizes two major steps in achieving it’s goals and correctly localizing image features.

First, this algorithm identifies and creates regions of interest all across the input image by using the selective search algorithm which leads to consistent and largely accurate predictions of objects within boxes.

Once these regions have been identified, the second stage of the process involves using convolutional neural networks to extract features from these proposed regions by applying layers to each and every box and using the identified features for classification in the following steps using a support vector machine(SVM) which can group these regions as containing an object or not. Finally, a regression model is used to redefine the bounding box and improve it’s outline of the identified object.

However, this approach was found to be exceedingly slow and thus, in the new and developed fast-RCNN algorithm, the stages are performed in reverse so as to first apply the networks and later identify regions of interest. Additionally, the SVM is replaced with solely utilizing an activation function in the output layer to compute all the complexities, thus converting it to a single-stage detector.

Opting for an even better approach, the faster-RCNN algorithm was used wherein region proposals are done using a trained neural network in itself and then the method follows the same logic as the fast-RCNN.

## Single -Shot Detector (SSD)

On the other hand, the single-shot detector is a single-stage detector which finds objects from the input images directly in one pass, without undergoing a region proposal algorithm. In this case, a previously trained neural network is used for feature extraction and instead of predicting bounding boxes, the CNN’s are applied to create different layers which are then extracted into grids. Lastly, a non-maximum suppression algorithm is applied to confine the entire object into one grid region instead of multiple.

The key advantage of this algorithm is its speed and efficiency and the fact that it can be used with images of multiple scales and sizes better.

1. Object Detection

# EXPERIMENTAL RESULTS

In this section, the results of experiments conducted by aforementioned researchers have been discussed. In Galvez’s experiment [1], they chose to use a limited dataset

Fig 4. Losses identified in the models from paper [1]

with person and quadrotor only, training and testing the data in an 80:20 ratio. On applying both models to the data, they observed minimum loss and more accuracy with the Faster-RCNN model, yet the SSD model reached the desired objective of identify the person and quadrotor in a lesser number of steps, seemingly faster. The mean average precision values were also used as a parameter to detect the better localizer among the two and it showcased similar results.

# CONCLUSION

Resultantly, researched data has clearly demonstrated the speed and accuracy of the convolutional neural network models in object detection for greater applications. From the results of several paper, we can tell that the SSD model has much better speed detection, yet the Faster-RCNN is the more accurate object detector.

Thus, this paper gives a general overview of the role played by CNNs in object detection and how it can further be used in robot surveillance and navigation systems, remote sensing imagery and dynamic object detection from videos.

##### References

1. R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 2023-2027, doi: 10.1109/TENCON.2018.8650517.
2. R. Gavrilescu, C. Zet, C. Foșalău, M. Skoczylas and D. Cotovanu, "Faster R-CNN:an Approach to Real-Time Object Detection," 2018 International Conference and Exposition on Electrical And Power Engineering (EPE), Iasi, Romania, 2018, pp. 0165-0168, doi: 10.1109/ICEPE.2018.8559776.
3. Redmon, Joseph, et al. “You Only Look Once: Unified, Real-Time Object Detection.” *Cvr*, 1 Jan. 1970, www.cv-foundation.org/openaccess/content\_cvpr\_2016/html/Redmon\_You\_Only\_Look\_CVPR\_2016\_paper.html.
4. Aktas, Yagmur Cigdem. “Object Detection with Convolutional Neural Networks.” *Medium*, Towards Data Science, 31 Jan. 2022, towardsdatascience.com/object-detection-with-convolutional-neural-networks-c9d729eedc18.
5. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in neural information processing systems, 2012.
6. C. Szegedy, S. Reed, D. Erthan and D. Anguelov: “Scalable, highquality object detection” Published 2014 in ArXiv