

A Survey on object recognition using deep learning

¹Shalini Pathak, Arya Institute of Engineering Technology & Management, Jaipur

²Kamlesh Gautam, Arya Institute of Engineering Technology & Management, Jaipur

³Mahesh Regar, Arya Institute of Engineering Technology & Management, Jaipur

⁴Dildar Khan, Arya Institute of Engineering Technology & Management, Jaipur

Abstract

Accurate and accurate discovery has been an important topic in the development of computer viewing systems. With the advent of in-depth learning strategies, the accuracy of object acquisition is greatly increased. This project aims to incorporate state-of-the-art object acquisition techniques with the aim of achieving high accuracy with real-time performance. A major challenge in many object acquisition programs is the reliance on other computer viewing techniques that facilitate an in-depth learning approach supported, leading to slower and less efficient performance. In this project, we use a completely in-depth learning method to solve the acquisition problem in a final way. The network is trained in the most challenging public data database (PASCAL VOC), where the challenge of finding something is done every year. The emerging program is fast and accurate, thus helping those applications that need to find an object.

Keyword: PASCAL VOC, RCNN, Fast RCNN, Faster RCNN

Introduction

Problem Statement

Many of the problems with computer viewing were complemented by their precision before ten years. However, with the advent of deeper learning strategies, the accuracy of these problems has greatly improved. One of the major problems was the image classification, which is defined as guessing the image category. A slightly more complex problem is the design of the scene, where the image contains a single object and the system must predict the location of the object in the image (the binding box around the object). The complex problem (of this project), of the acquisition of an object involves the division and execution of a place. In this case, the input system will be an image, and the output will be a binding box corresponding to all the items in the image, as well as the item category in each box. An overview of all these issues is shown in Figure. 1.

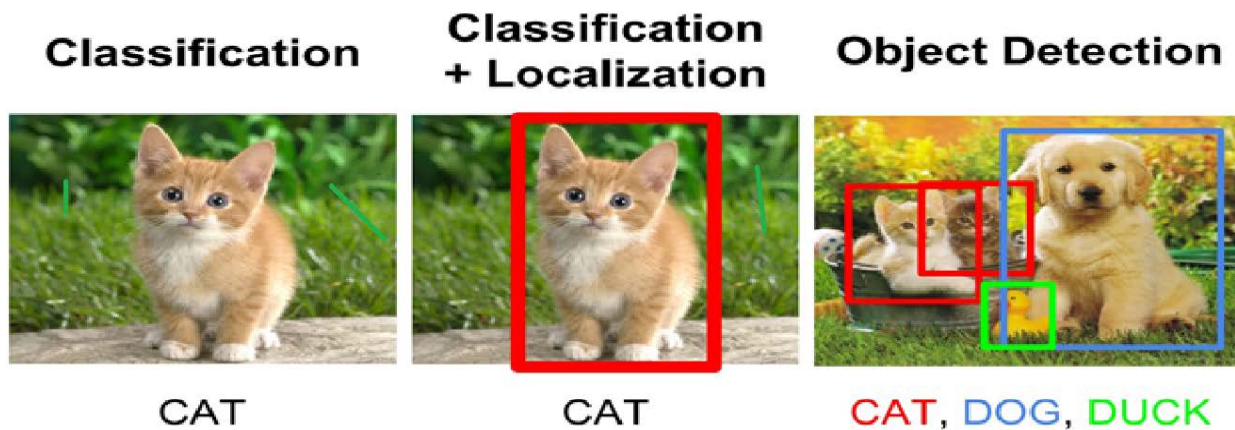
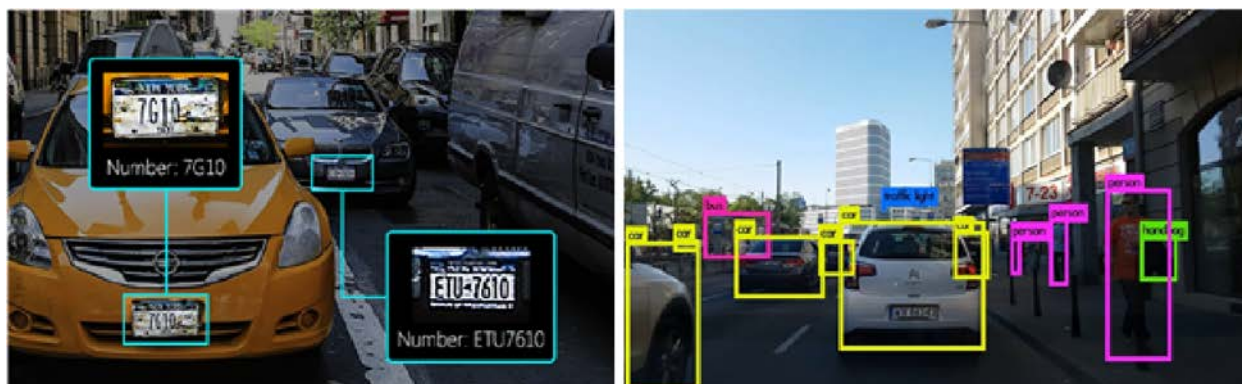


Figure 1: Computer Vision Tasks

Applications

A well-known object detection app for face detection, which is used on almost all mobile cameras. A standard (multi-level) application can be used in private driving where a variety of items need to be found. It also has an important role to play in surveillance programs. These systems can be combined with other functions such as the measurement of the pose where the first phase of the pipe to find the object, and then the second phase will be to measure the position in the obtained region. It can be used to track things and thus can be used in robotics and medical applications. So this problem works with most applications.



(a) Surveillance (b) Autonomous vehicles

Figure 2: Applications of object detections

Challenges

A major challenge to this problem is the different output sizes created due to the number of variables that can be present in any given input image. Any typical machine learning function requires a fixed size to install and output a model to be trained. Another important obstacle to the wide adoption of acquisition programs is the real-time requirement while accurate acquisition. The more complex the model, the more time it needs to be considered; and the more complex the model, the less accurate it is. This trade between accuracy and performance needs to be selected as an application. The

problem involves splitting and reversing, which results in the model being read simultaneously. This adds to the complexity of the problem.

Related Work

There has been a lot of work in finding the object using traditional computer viewing techniques (sliding windows, partial partial models). However, they lack the accuracy of strategies based on in-depth learning. Among the strategies based on in-depth learning, two broad pathways are widespread: dual-stage acquisition (RCNN [1], Fast RCNN [2], Faster RCNN [3]) and integrated acquisition (Yolo [4], SSD [5]). The main concepts involved in these approaches are described below.

Bounding Box

The binding box is a rectangle drawn in the picture that fits snugly with the object in the picture. The matching box is present in all events for everything in the image. In the box, 4 numbers (center x, center y, width, height) are predicted. This can be trained using the distance between the predicted box and the real world. A distance range is a Jaccard distance that includes a union between the predicted true boxes and the ground as shown in Figure. 3.

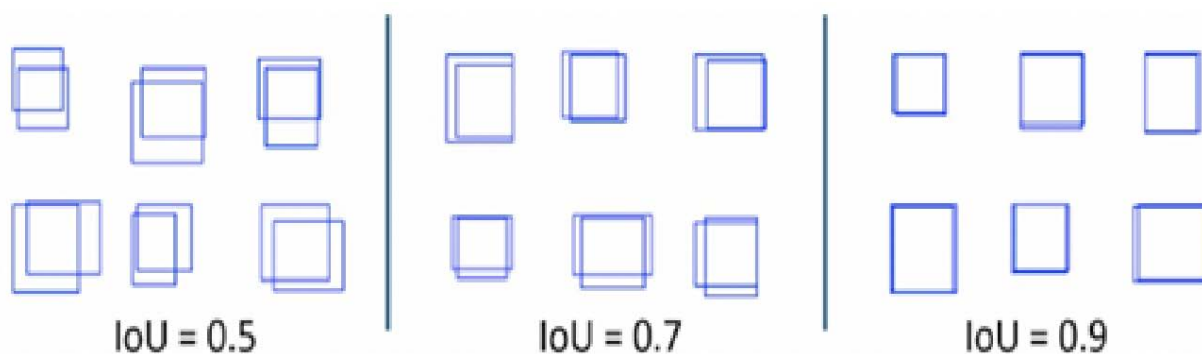


Figure 3: Jaccard distance

Classification + Regression

The binding box is predicted using retreat and the section within the binding box is predicted using a split. A summary of the construction is shown in Figure 4

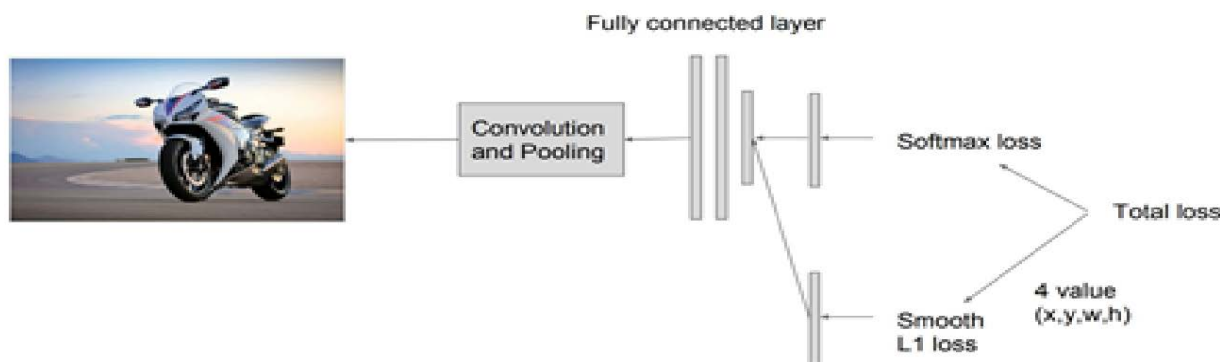


Figure 4: Architecture overview

General methodology

The OD program basically has two main components namely: the training phase and the assessment phase shown in Figure 2 which illustrates the traditional functioning of the OD system. The reading phase is primarily designed to differentiate so that we can see things currently in the image provided as a program installation. The learning phase can be divided into learning by training and by learning by validation. Training by learning includes mainly the training area in which the appropriate learning program is defined, partially supported or patch-operated, etc. presentations such as histogram representation, random forest representation, etc.

While on the other hand, learning with a verification block does not require any kind of training as previously confirmed. So after further enhancing the image, the exact template matching is completed showing the features of the object within the image. The main purpose of the test phase is to determine whether an object is present in a given image by the system such as input and in which case the object is in which category. Here the image is viewed on the object by various search techniques such as the sliding windows process and corresponds to the output of the search method, selected in the item category.

Conclusions

Object discovery can be a key skill in many computer programs and robots. Although great strides have been made in recent years, and a few of the current strategies are part of a larger clientele (e.g., face-to-face acquisition with Smart phones) or integrated auxiliary driving technology, we are still far from achieving human-level performance, especially in terms of open-world learning. It should be noted that the acquisition of the object is not widely used in many areas where it can be very helpful. As mobile robots, with more sophisticated equipment, are beginning to be widely distributed (e.g., quad-copters, drones, and auxiliary robots), the need for object detection systems has become increasingly important. Finally, we would like to think about what we will need for discovery systems for nano-robots or robots that will explore places that have never been seen by humans, such as parts of the ocean or other planets, so Discovery Programs will need to learn new phrases as they encounter them. In such cases, the brain of the open world in real-time will be critical.

Future scope

- Local or terrestrial features used for recognition can be amplified, increasing the efficiency of the object recognition system.
- The geometric properties of the image can be inserted into the vector of the element to be seen. 150 x Unsupervised separator is used instead of a controlled object recognition detector.
- x The proposed object recognition system uses a gray image and discards color details.
- Color details in the image can be used for object recognition. Color-based recognition plays an important role in robots

References

1. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

2. Ross Girshick. Fast R-CNN. In International Conference on Computer Vision (ICCV), 2015.
3. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards realtime object detection with region proposal networks. In Advances in Neural Information Processing Systems (NIPS), 2015.
4. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
5. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, ChengYang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.
6. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.