REVIEW OF DATA GATHERING IN OBJECT RECOGNITION APPLICATIONS USING VARIOUS OBJECT RECOGNITION METHODS

Muhammad Amirul Fahmi Noor Anim [1] and
Aqilah Hanim Binti Mohd Taufik[2]

Bachelor of Computer Science(Computer Graphic Software),
Faculty of Engineering, School of Computing,
Faculty of Computing, University Technology Malaysia, Johor Bahru, Malaysia
(E-mail: mafahmi9@graduate.utm.my[1], ahanim8@graduate.utm.my[2])

ABSTRACT

The technique of recognising the object present in photographs and videos is known as Object Recognition. It is one of the most significant machine learning and deep learning applications. The aim of this field is to teach machines to recognise (understand) the content of a picture in the same way that humans do. There are various kinds of methods used in applications published in the mobile application stores that implement object recognition. This paper introduces a comparison of object recognition mobile based applications that uses various image recognition methods. The advantages, disadvantages and also the proper ways to implement each of the methods used will be discussed in this paper.

Key words: Object recognition, Object recognition application, Method

INTRODUCTION

The existing object recognition mobile applications such as “Aipoly Vision”, “CalorieMama” and “CamFind” are operated by using the mobile phone’s camera to detect the objects in front of the users. Although all the mobile applications use object recognition systems, all of them are used for different purposes. As an example, the application “Aipoly Vision" is mainly focused on helping their target users who are blind, visually impaired, and color blind to understand their surroundings. The application is used to recognize objects and colours for their target users. “CalorieMama” on the other hand focuses on displaying the calorie information of the food that is detected through the application. Last but not least, “CamFind” allows users to identify any item just by taking a picture with their smartphone, providing a range of information including related images, local shopping results, price comparisons and web results.

The process of applying object recognition into the mobile applications system is rather a direct implementation of just using the camera and pointing it to the objects in order to detect the objects. The most accurate information regarding the object will then be displayed for the users to see.

There are two different types of object recognition techniques. These techniques are known as object recognition using deep learning and object recognition using machine learning. In deep learning, the system uses Convolutional Neural Network (CNN) to learn about an object's inherent features in order to identify the certain object. There are two approaches to perform object recognition in deep learning. These approaches include training a model from scratch and also using a pre-trained deep learning model.

In a system where it uses a pre-trained deep learning model, there are two big model families for deep learning models which are R-CNN Model Family and YOLO Model Family. R-CNN models stands for "Regions-Based Convolutional Neural Network" which is designed for object localization and object recognition. There are three types of techniques in the R-CNN model which are R-CNN, Fast R-CNN, and Faster R-CNN.

R-CNN technique consists of three modules which are Region Proposal, Feature Extractor, and Classifier. It is a method where it uses selective search to extract just 2000 regions from an input image which are called region proposals. These 2000 candidate region proposals will be squared and fed into a convolutional neural network, which outputs a 4096-dimensional feature vector. The CNN acts as a feature extractor, with the extracted features being fed into an Support Vector Machine (SVM) to classify the presence of the object within that candidate region proposal. The algorithm predicts four values that are offset values to increase the accuracy of the bounding box, in addition to forecasting the presence of an object within the region proposals.

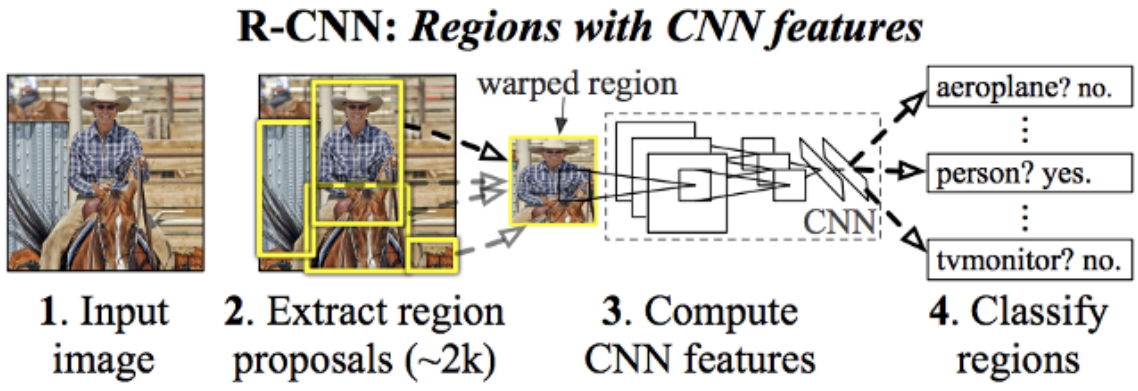

Figure 1.0 : Summary of the architecture of R-CNN model

Fast R-CNN is a slightly upgraded version of R-CNN where it is built with a fast object detection algorithm. In Fast R-CNN, the input image is directly fed to CNN that will generate a convolutional feature map instead of fed with region proposals. This increases the performance of Fast R-CNN during testing time.

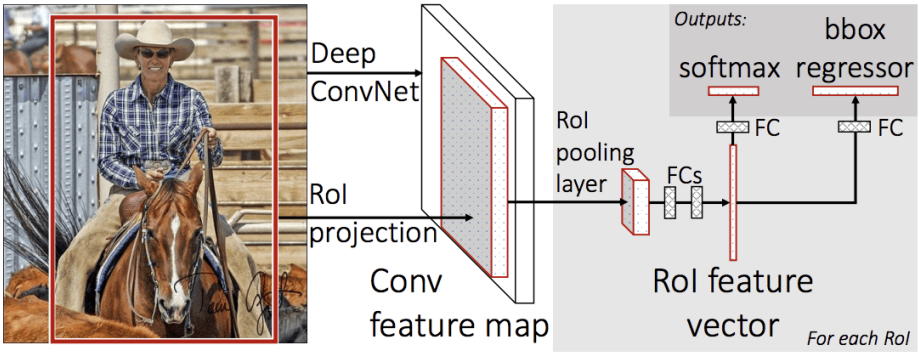

Figure 2.0 : Summary of the architecture of Fast R-CNN model

For Faster R-CNN, the model is upgraded with a less time-consuming process where it was introduced with an object detection algorithm that replaces the selective search algorithm and allows the network to learn region proposals.

YOLO or "You Only Look Once" is a model that is much faster than R-CNN models in achieving object detection in real-time. The model consists of several techniques which are YOLO, YOLOv2 (YOLO9000), and YOLOv3.

YOLO is an approach that uses a single end-to-end trained neural network that takes an image as input and directly predicts bounding boxes and class labels for each bounding box. Although the technique runs at 45 frames per second and up to 155 frames per second for a speed-optimized version of the model, it has lower predictive accuracy (e.g., more localization errors).

The model divides the input image into a grid of cells, with each cell responsible for predicting a bounding box if the centre of a bounding box falls inside the cell. Each grid cell generates a bounding box based on the x, y coordinates, as well as the width, height, and trust. Each cell is also used to predict a class.

The YOLO algorithm's drawback is that it has trouble detecting small objects in images; for example, it would have trouble detecting a flock of birds. This is due to the algorithm's spatial constraints.

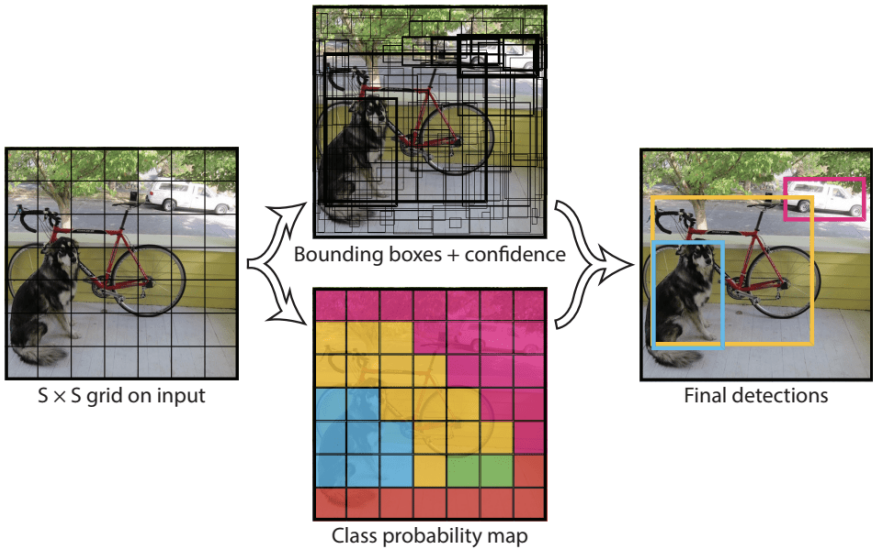

Figure 3.0 : Summary of the predictions by YOLO Model

For YOLO v2, the model's performance has been improved as the model has undergone a number of training and architectural modifications, including the use of batch normalisation and high-resolution feedback images. It was trained simultaneously on two object recognition datasets and is capable of predicting 9,000 object classes, hence the name "YOLO9000."

For YOLO v3,minor improvements were introduced including a deeper feature detector network and minor representational changes.

RESULT

We have conducted a simple evaluation that has been done on an android-based device. We installed and had a quick run through all the three mobile applications in our android phones. There are many pros and cons that we have found in all three mobile application systems. On "Aipoly Vision", although the camera could detect the objects well, it was not able to recognize the accurate object as the data on the object is not that many. As for "CalorieMama", the object recognition system works well most of the time for most foods, but when there is a wrong food shown, the users are unable to edit the right information in the database. Last but not least, we actually did not find any pros on "CamFind" as the mobile application system was only about 10% accurate. This mobile application system only recognizes the colour of the object and has failed to find users the results they needed. There were also many bad reviews on the application in the review section of the Application Store.

CONCLUSION

As the technology in our world is rapidly upgrading day by day, we can see that object recognition softwares and applications like “Google Lens” are being used on a daily basis. Object recognition is a computer technology that deals with identifying instances of semantic objects of a certain class in digital images and videos. It is related to computer vision and image processing. For the mobile application “Aipoly Vision” to be able to be used by their target user, a more thorough addition of data needs to be added into the system so that the mobile application system can display a higher percentage of accurate results.