REVIEW OF METHODS IN OBJECT RECOGNITION IN MOBILE BASED APPLICATIONS

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

Nurul Najiha binti Hamdan[3]

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] , nnajiha7@graduate.utm.my[3])

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 and reviews its 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

I. 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.

Object recognition is a broad term that refers to a group of related computer vision tasks involving the identification of objects in digital images (Brownlee, 2021). Object recognition techniques are divided into two categories. Object recognition using deep learning and object recognition using machine learning are the terms for these techniques. The system uses Convolutional Neural Network (CNN) to learn about an object's inherent features in order to classify and improve classification efficiency in deep learning. In deep learning, there are two approaches to object recognition. These methods include building a model from the ground up as well as using a pre-trained deep learning model.

As previously mentioned, ImageNet is the current image database that CNN models use to extract features in deep learning methods. The first noteworthy technique we can apply to our object is training the model from scratch by collecting labelled datasets and designing a network architecture that will learn the features in order to construct the model. Another technique is feature extraction using a pre-trained deep learning model. This technique is time efficient, has high level accuracy in recognizing the object and provides faster results as all the features of the model have been trained using the deep learning method (Choe, Choi & Kim, 2020).

Two main deep learning model families that are used in systems that include a pre-trained deep learning model are the R-CNN Model Family and the YOLO Model Family. To locate and identify objects, the R-CNN model is used. It stands for "Regions-Based Convolutional Neural Network." The R-CNN model has three

types of techniques: R-CNN, Fast R-CNN, and Faster R-CNN.

It is a technique that employs selective search to extract only 2000 regions from an input image, which are referred to as region proposals. These 2000 candidate region proposals will be squared and fed into a 4096-dimensional feature vector generated by a convolutional neural network. The CNN serves as a feature extractor, with the extracted features being fed into an SVM to identify the object's existence within the candidate region proposal. In addition to forecasting the presence of an object within the region proposals, the algorithm predicts four offset values to improve the bounding box's accuracy.

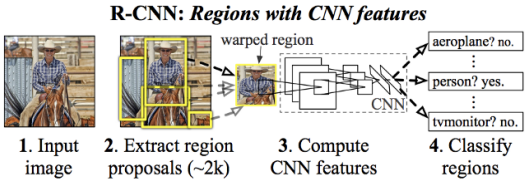

Figure 1.0 : Summary of the architecture of R-CNN model (Girshick, Donahue, Darrell, & Malik, 2014)

Faster R-CNN replaces the selective search algorithm with an object detection algorithm that allows the network to learn region proposals in a less time-consuming process.

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 each bounding box has class labels through a list of colours to differentiate the objects. 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 bounding box's centre falls inside the cell. Based on the x, y coordinates, as well as the distance, height, and trust, each grid cell generates a bounding box. In addition, each cell is used to predict a class.

The YOLO algorithm's flaw is that it has difficulty identifying small objects in

images; for example, a flock of birds will be difficult to detect. This is all due to the algorithm's spatial constraints.

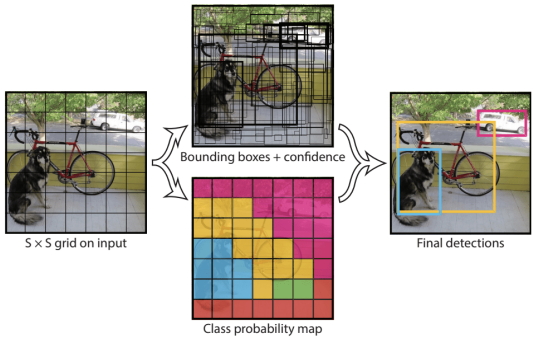

Figure 3.0 : Summary of the predictions by YOLO Model

The performance of YOLO v2 has been enhanced as a result of a range of training and architectural changes, including the use of batch normalisation and high-resolution feedback images. It was trained on two object recognition datasets at the same time and can predict 9,000 different object types, hence the name "YOLO9000."

Minor changes, such as a deeper feature detector network and minor representational changes, were made in YOLO v3.

II. 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 unable to recognize the accurate object due to a lack of data on the object. 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.

III. CONCLUSION

As a final observation, 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 identifies instances of semantic objects in digital images and videos. It is related to artificial intelligence, 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 expanded into the system so that the mobile application system can display a higher percentage of accurate results.

REFERENCES

[1] Brownlee, J. (2021, January 26). A gentle introduction to object recognition with deep learning. Retrieved March 30, 2021, from https://machinelearningmastery.com/object-recognition-with-deep-learning/

[2] Ahn H., Rhee SB. (2015) Research of Object Recognition and Tracking Based on Feature Matching. In: Park J., Stojmenovic I., Jeong H., Yi G. (eds) Computer Science and its Applications. Lecture Notes in Electrical Engineering, vol 330. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45402-2_152

[3] Daegyu Choe, Eunjeong Choi, Dong Keun Kim, "The Real-Time Mobile Application for Classifying of Endangered Parrot Species Using the CNN Models Based on Transfer Learning", *Mobile Information Systems*, vol. 2020, Article ID 1475164, 13 pages, 2020. https://doi.org/10.1155/2020/1475164

[4] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

https://doi.org/10.1109/ICCV.2015.169

[5] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

https://doi.org/10.1109/CVPR.2014.81