# MSc Project - Reflective Essay

| | |
|---|---|
| Project Title: | Detection and segmentation of robot hands and objects selected for picking |
| Student Name: | Muraleekrishna Harikumar Sreeja |
| Student Number: | 210696370 |
| Supervisor Name: | Lorenzo Jamone |
| Programme of Study: | MSc Artificial Intelligence |

## I. Overview

My objective of object recognition and segmentation of robot hands and the objects that are chosen to pick was accomplished in the project. The entire code was developed in Python using Google Colab Notebook.

The Google Colab link below provides access to the project's source code and results:

https://tinyurl.com/2pdxkyhn

## II. Strengths

- Robot hand sometimes overlapped the objects that they were assigned to pick because the camera is positioned above them. This gave rise to the misguided impression that the object wouldn't be effectively detected. From most of the overlapped images, following sufficient training, the objects were mostly correctly detected. There were only a few occasions where the objects were not detected. This makes it a key strength of the model used here.
- When the model was presented with a single and cluttered set of the same object, the model reliably identified these separately in most instances beyond the expectations.
- In contrast to other object detection models, Mask-RCNN predicts an object mask parallel with the existing branch for bounding box recognition. This allowed us to derive a considerably more precise spatial layout of an object.

## III. Weakness

- Some of the images in the dataset have robot hand overlapping the objects. Even though the objects in these images were visible to the human eye, the model was unable to identify them because of the projection of robot hands over the objects that obscured its field of vision. This is due to the camera's position being above the robot's hand in the dataset. The solution to this problem, however, is to change the camera's location from the top to the front of the item and the robot's hand.
- This model's training was accelerated with the use of a Google Colab notebook. Google Colab is a cloud-based Python IDE that is mostly used for machine learning and data analysis. Because we were using a free version of the Google Collab, there were session constraints and only one GPU could be utilised, along with a RAM limit of 12 GB. So sometimes due to the session constraints, the training gets terminated in the middle, forcing the model to be trained from scratch.
- Since the robot hand was continuously moving in the images of the main dataset, it had many different orientations, but the objects that were chosen for picking

had just 10-12 different orientations each. It was the reason why the robot hand's confidence score was higher than that of the grocery objects (about equal to 1) indicating that the model was properly trained to recognise the robot hand under any scenario but not sufficient in the case of the objects chosen to pick. As these could underperform when under unsupervised settings.

## IV. Relationship Between Theory and Practice

- Faster R-CNN was used to create Mask R-CNN. Faster R-CNN outputs a class label and a bounding-box offset for each candidate object, whereas Mask R-CNN adds a third branch that outputs the object mask. Along with this, the primary benefits of mask rcnn were its performance, adaptability, and simplicity. With an object recognition accuracy of >80% on papers, mask rcnn surpassed most object detection algorithms currently available [1].
- Our experimental findings also revealed that in most cases, the accuracy/confidence of the detections was in the same range. Considering the overall scenario, our implementation of the mask R-CNN-based model worked exactly as per the theory.

## V. Work to conduct with more time

- More training may have been done if there had been more time, leading to more accurate results.
- Similarly, other object detection techniques could have been performed to compare and discover which model is the most suitable for this dataset.
- Using an image polygonal annotation tool like labelme, the dataset could have been properly annotated with the availability of more time.
- Furthermore, the confidence scores of each class could have been taken and plotted to get an overall idea of how efficiently each object is detected.

## VI. Legal, Social and Ethical Issues and Sustainability

- Since no human or animal subjects were used in the study, the ethical concerns are almost exclusively indirect. The amount of computing power needed to carry out this job is a sustainability concern. It took around 30 to 45 minutes to complete each epoch, and it took 7 to 8 hours to complete all 10 epochs. A prompt will display every one to two hours to remind you to keep your browser open; if you don't, the training will stop, and you'll have to repeat it for another seven to eight hours. This included significant energy use and raised issues regarding resource waste and environmental sustainability. There are more issues with other models that require days to train. More energy was also needed to re-train the model if false findings were obtained.
- The MS COCO pictures dataset has a Creative Commons Attribution 4.0 License attached to it. As a result, this license permits one to share, modify, remix, and expand upon the work including for commercial purposes if the original source or creator is given the credit. As a result, this doesn't transgress any laws [2].
- To summarise, no major ethical concerns were present in this model except the sustainability issue regarding carbon footprint.

## VII. Conclusion

I am quite pleased with the project overall. The aim of implementing and testing the object identification and segmentation of the robot hand and the things it is tasked with picking was carried out successfully. The project contains a thorough background and literature evaluation, both of which, in my opinion, are effectively put in place and provide summaries of the overall idea and relevant work. Additionally, I believe that the experiments support the discussion and hypothesis presented in the study.

## Reference


[1] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[2] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014, September. Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer, Cham.