James Dyson Foundation Undergraduate Bursary 2017/18

Part IIB Project Report: Data Synthesization using Catalogue Images for Classification Learning for Robotic Picking by Michael Cheah (mmxc2)

Figure 1. Robots in Amazon

Figure 2. Human workers packing items in Amazon warehouse.

While robots are heavily relied upon in warehouses such as those owned by Amazon, humans are still necessary for tasks such as picking because robots are still not capable of reliably classifying and grasping randomly placed objects with little training data.

Machine learning techniques are often used to give robots the ability to “learn” to classify objects using only training data. For example, instead of explicitly programming a robot to differentiate an apple from a banana with features defined by the user, e.g. apple is red and round while the banana is yellow and long, a good machine learning algorithm should be able to figure out these features for the robot given enough data to train on.

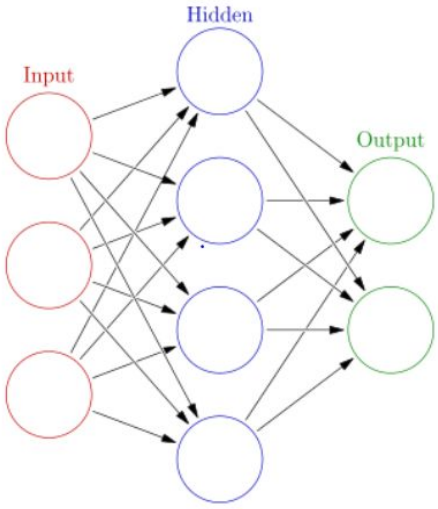

Figure 3. Aritifical Neural Network model

An exciting class of machine learning techniques is deep learning, which uses artificial neural networks. Artificial neural networks are models that are inspired by the biological neural networks of animal brains. They are formed of many connected artificial neurons, which receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation.

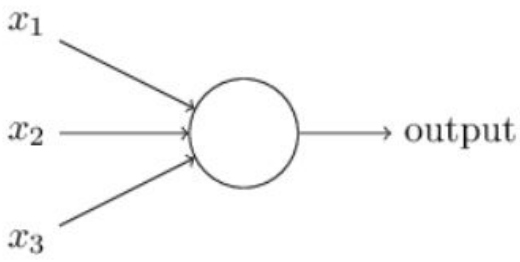

Figure 4. Single neuron example

The inputs of a neuron are weighted i.e. some inputs are more important than others based on their weights. It is these weights that can be “learned” and are used to change the behaviour of the network so that it can perform specific tasks or approximate any function. In this project, the goal is to adjust the weights so that the network learns to classify and locate objects correctly based on visual data (images)

Figure 5. The ImageNet dataset contains over 14 millions labelled images and more than 20 thousand classes.

Figure 6. COCO, a large-scale object detection, segmentation, and captioning dataset with 80 classes in 200k images.

A big drawback to using neural networks is that they require a lot of training data. Many of the best performing neural networks have been trained using extremely large labelled datasets like ImageNet or COCO (Common Objects in Context). However, this is not feasible for many robotic applications. This is especially true for warehouse applications where robots would have to deal with millions of different objects - to manually label all objects would be time-consuming, and not scalable.

Data synthesis can be used to generate training data sets from a very limited set of catalogue images in order to reduce the need for large data sets and extensive manual data labelling. Catalogue images are images of objects from multiple different angles and with clean backgrounds. These type of images are often easily obtained from warehouses databases.

Figure 7. Catalogue images used in project

The process of generating new training images using the catalogue images is shown below

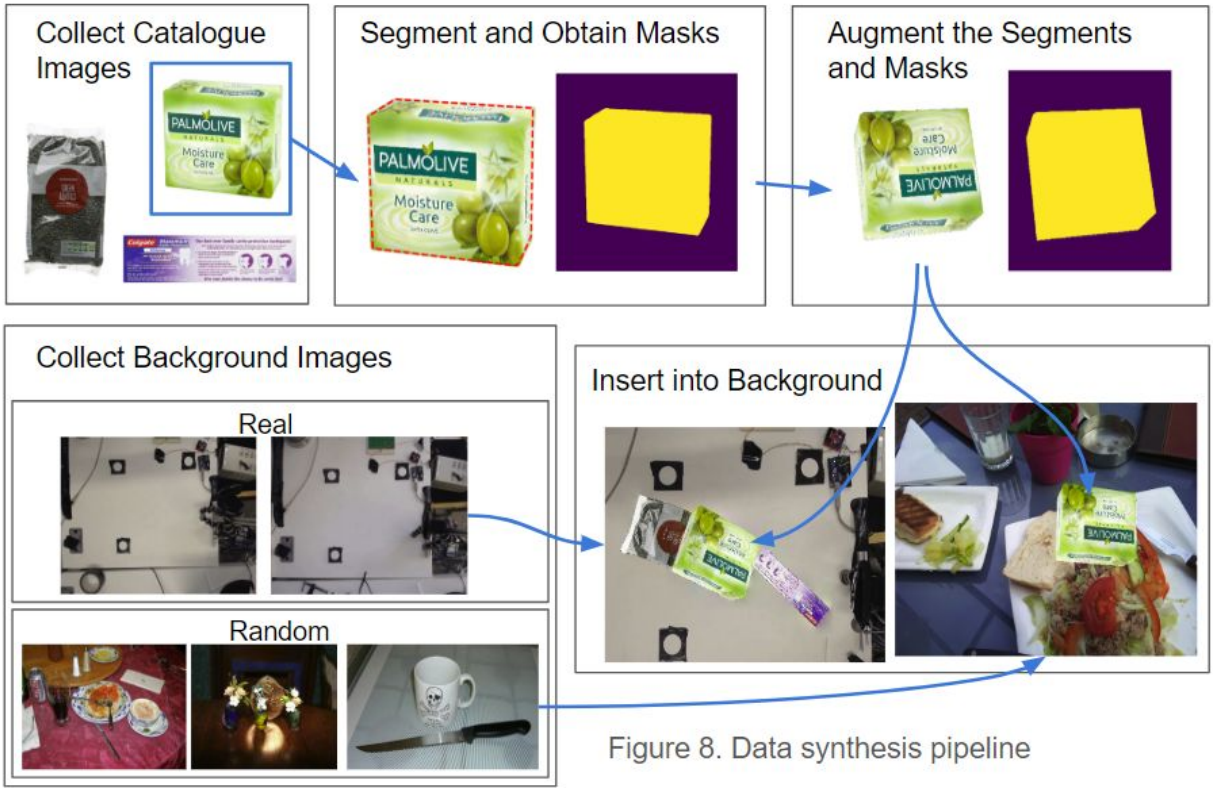

Figure 8. Data synthesis pipeline

This synthetic dataset can then be used to train a object detection network of our choice. In this project, a Mask R-CNN, a network that takes in images as inputs and outputs segmentation masks with labels, is used.

Some examples of the output of the trained network can be seen in Figure 9.

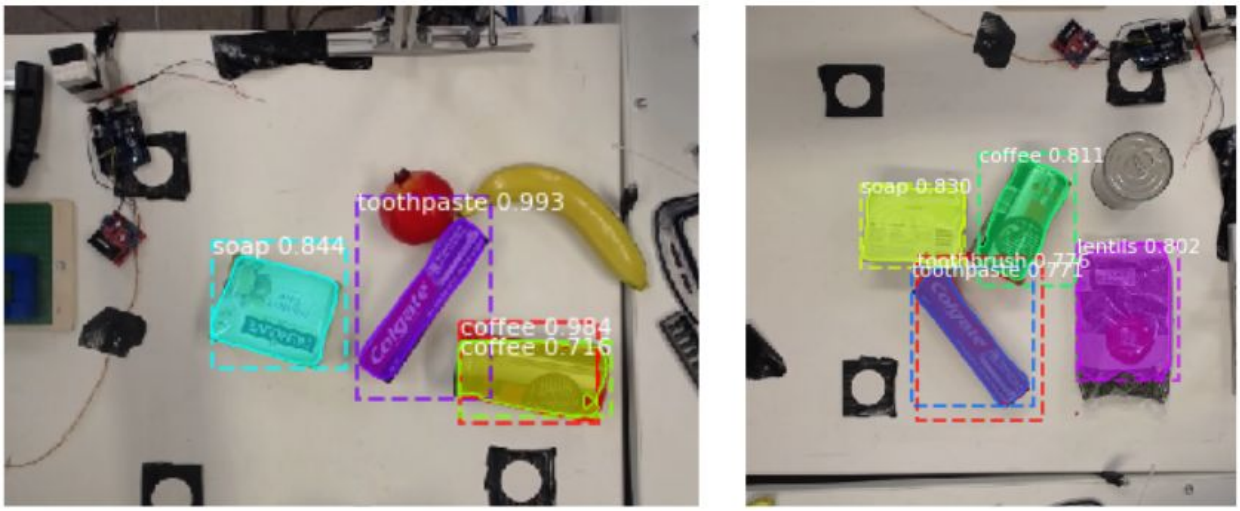

Figure 9. Object detection network output given image

A simple robotic setup is used to test the effectiveness of the trained network using synthetic data. Empirical tests that were carried out found that the trained network and robotic system were able to achieve a grasping success rate of at least 70% for the objects tested.

Only 10 objects were ultimately used in this project directly from Amazon. Future work would involve testing the scalability of this method with more objects. Object segments were also directly cut from catalogue images, thus many poses that are not seen in the catalogue images are missing; more sophisticated methods should be explored to interpolate between those poses so that a more complete representation of the object can be used to synthesize the data.

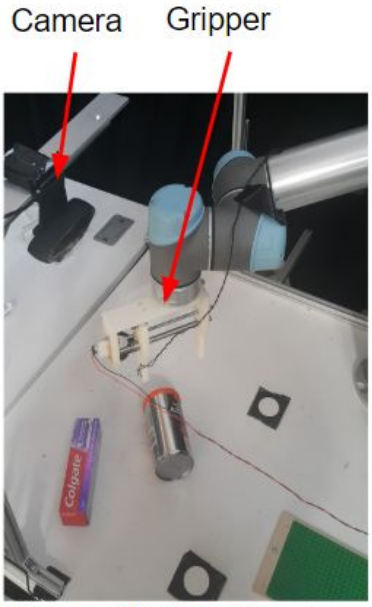

Figure 10. Robotic setup

Image sources:

[1]https://www.youtube.com/watch?v=4sEVX4mPuto
[2]https://qz.com/885425/amazons-massive-fleet-of-robots-hasnt-slowed-down-its-employment-of-humans/
[3] https://en.wikipedia.org/wiki/Artificial_neural_network
[4] http://neuralnetworksanddeeplearning.com/chap1.html
[5] https://medium.com/@mozesr/script-to-get-images-from-image-net-org-7fe8592e6650, www.image-net.org/
[6] http://cocodataset.org/
[7][8][9][10] 4th Year Project report.