Deep Convolutional Neural Networks for Robotic Grasp Detection

CS39440 Major Project Report

Author: Oliver Thomas (olt13@aber.ac.uk)

Supervisor: Dr Patricia Shaw (phs@aber.ac.uk)

30th March 2021

Version 1.0 (Draft)

This report is submitted as partial fulfilment of a MEng degree in  
Robotics and Embedded Systems Engineering (132C)

Department of Computer Science

Aberystwyth University

Aberystwyth

Ceredigion

SY23 3DB

Wales, UK

Declaration of originality

I confirm that:

* This submission is my own work, except where clearly indicated.
* I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
* I have read the regulations on Unacceptable Academic Practice from the University’s Academic Registry (AR) and the relevant sections of the current Student Handbook of the Department of Computer Science.
* In submitting this work, I understand and agree to abide by the University’s regulations governing these issues.

Name Oliver Thomas

Date ……………………………………………

Consent to share this work

By including my name below, I hereby agree to this project's report and technical work being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Name Oliver Thomas

Date ……………………………………………

Acknowledgements

I am grateful to my supervisor, Patricia, for guiding and advising me while completing this interesting project.

I would also like to thank my parents and my girlfriend Katie for their incredible ongoing support.

Abstract

Include an abstract for your project. This should be approximately 300 words.

The abstract is an overview of the work you have done. Highlight the purpose of the work and the key outcomes of the work.

This paper explores the complicated topic of robotic grasping, following a deep learning approach to ‘learn’ a generalised view of grasping to be applied to novel objects.

Contents

1 Project Background, Analysis & Process 1

1.1 Project Description 1

1.2 Background & Motivation 1

1.3 Analysis 1

1.3.1 General Analysis 1

1.3.2 Grasp Dataset 2

1.3.3 Simulation Environment 2

1.3.4 Convolutional Neural Network 2

1.4 Process 3

1.4.1 Feature-Driven Development 3

1.4.2 Feature Table 4

2 Convolutional Neural Networks 5

2.1 Background 5

2.2 Technical Workings 5

3 Experiment Methods 6

3.1 Experiment Hypothesis 6

3.2 Experiment Design 6

3.2.1 Experiment Outline 6

3.2.2 Ground Truth Evaluation 7

3.2.3 Trained Neural Network Evaluation 7

3.2.4 Experiment Results 8

4 Software Design, Implementation and Testing 9

4.1 Design 9

4.1.1 Overall Architecture 9

4.1.2 Simulation Environment 9

4.1.3 Dataset 9

4.1.4 Applied Technologies 10

4.1.5 Requirements Table 10

4.2 Implementation 10

4.3 Testing 10

4.3.1 Overall Approach to Testing 10

4.3.2 Automated Testing 10

4.3.3 Integration Testing 10

4.3.4 User Testing 10

5 Results and Conclusions 12

5.1 Ground Truth Success 12

5.2 Model Loss and Accuracy 12

5.3 Model Success in Simulation 12

6 Critical Evaluation 13

Annotated Bibliography 14

Appendices 15

A. Third-Party Code and Libraries 16

B. Ethics Submission 17

C. Feature Table 18

D. Code Samples 19

Table of Figures

[Figure 1: The feature table for the project, as seen in week 11. 4](https://prifysgolaber-my.sharepoint.com/personal/olt13_aber_ac_uk/Documents/Work/Major%20Project/mmp_grasp_detect/Reports/Final%20Report.docx#_Toc70808607)

[Figure 2: Initial Gantt chart to plan a rough timeline. 4](https://prifysgolaber-my.sharepoint.com/personal/olt13_aber_ac_uk/Documents/Work/Major%20Project/mmp_grasp_detect/Reports/Final%20Report.docx#_Toc70808608)

[Figure 3: The pipeline for evaluating the trained neural network in simulation. 8](https://prifysgolaber-my.sharepoint.com/personal/olt13_aber_ac_uk/Documents/Work/Major%20Project/mmp_grasp_detect/Reports/Final%20Report.docx#_Toc70808609)

# 

# Project Background, Analysis & Process

## Project Description

The aim of this project is to research and apply (in simulation) a deep convolutional neural network for the purpose of detecting successful grasp poses of objects. The project utilises many interesting technologies, from machine learning libraries to robotic middleware. This report’s main job is to explore these technologies in more depth, but it will also cover both the experimental and software engineering processes involved in completing the project.

## Background & Motivation

Over the last five to ten years, much work and research has gone into the field of robotics. Perhaps most significant is the work in the field of robotic manipulation and grasping. Despite this, robotic grasping is still a challenging topic due to the many environmental inputs that effect how a grasp can be accomplished. One company at the forefront of robot research is Boston Dynamics [1], who have produced many new and exciting robots; including a dog-like robot that can open doors. However, though this may appear like the robot has successfully tackled the grasping problem, it is still very complex. Robots that are designed to perform grasp still typically use a database that contains information on how to grip specific objects that the robot will come across. However, this method does not work well in a constantly changing environment, in which a robot may find many novel objects.

This project was inspired by reading the Review of Deep Learning Methods in Robotics Grasp Detection [2]. This paper discusses the “current state-of-the-art in regards to the application of deep learning methods to generalised robotic grasping”, which fascinated me. Further reading of the review led to the decision to research more into the topic of deep learning in connection with robotic grasp. This was followed by reading many other papers discussing the topic, such as … [], to ascertain the more details about what the project would entail. They also gave more detailed information about certain design issues that will be discussed in *Chapter 4.1*. For instance … [] paper provided details on a basic Convolutional Neural Network (CNN) structure to build off of for this project.

## Analysis

### General Analysis

The nature of this project is research based; therefore, research questions that the project aims to answer need to be defined before the problem can be properly analysed. The main question to be researched is – is it possible for a CNN to learn grasp patterns for specific objects? It is clear from the extensive research that has been done into this area that this should be possible; however, there may be technical limitations in place that prevent this project to produce the most accurate model. The limitations to the implementation of this project will be discussed further in *Chapters 4.2 and* *6*. This question will be evaluated by using metrics during the model training pipeline, with results coming from the model loss function.

The next question to be answered is – will the trained model produced be able to be applied to novel objects? In other words, has the model been able to get a generalised view of grasping? The answer to this question will be found through simulation – the results of which will be used to evaluate the efficacy of the model and the success of the research. The design of the experiments will be shown in *Chapter 3* along with a statement of the project hypothesis.

Upon analysis of this project, it became clear from the beginning that there would be three major component elements to conduct the experiments. These components for the basis for the technical objectives and are as follows:

1. Generating an image – grasp dataset.
2. Training a CNN to predict grasps from images.
3. Building a simulation environment that will be used to run experiments.

The next few sections will analyse each of these objectives, breaking down the possible technologies that could be used, plus a brief description of possible challenges that may arise (though this will be discussed in more detail in the *Design* and *Implementation* sections).

### Grasp Dataset

By reading Caldera’s review into the subject of deep learning for robotic grasping and other papers [], it became readily apparent that there was a main grasp dataset that was standard. This was known as the Cornell Grasp Dataset [] and seemed to include everything that was required for the purposes of this projects research – a dataset of images with labelled grasps. However, unfortunately when trying to obtain the data from the referenced links, the data seemed to have been taken down from the public domain. Hence, more research was required to find data suitable for the project. During this research, two other datasets were found that seemed to meet the requirements – the Jacquard dataset [], and the ACRONYM dataset []. After more analysis of what was on offer from the two options, the ACRONYM dataset was decided upon, mainly due to the fact it is linked to the ShapeNetSem object database, which could easily be used within the simulation environment. In addition to this, the ACORNYM dataset is provided with access to Python scripts to manipulate the data. This task may be the most important part of the project, as the data used in machine learning is fundamental to the usefulness of the output.

### Simulation Environment

Due to previous experience, the simulation side of the project is based in ROS; however, research was required for which simulator was going to be used. The obvious choice was Gazebo [], but other options included the OpenRave and the GraspIt! simulators. After analysing the strengths and weaknesses of each option, Gazebo was agreed on. This is mainly due to its ease of use, experience with the software, and compatibility with the ShapeNetSem models.

Another consideration to make when designing the environment, is to make the scene within the camera as similar to that produced by the grasp dataset. This will help when integrating the trained network into the simulation.

### Convolutional Neural Network

Convolutional Neural Networks (CNNs) are tricky to implement efficiently; however, libraries and platforms such as, Tensorflow [] and PyTorch [], enable the implementation of machine learning algorithms such as CNNs easy. They also support customisation (such as with loss metrics or training loops) and optimisation. Tensorflow is the library of choice for this project, due to some previous experience and the interest in developing these skills further. Initial analysis of the problem shows that the project may need to implement a custom loss function to improve the network for this purpose; however, the standard libraries will likely be used to begin with, in order to create a working prototype.

## Process

### Feature-Driven Development

During the initial analysis of the project, several different software development methodologies were researched. It was decided that this project will use an agile Feature-Driven Development (FDD) methodology for planning and time management. The FDD process followed over the course of the project has been adapted to make it suitable for an individual developer. The main change is the removal of team roles and feature teams, as it is a solo project. The FDD process can be considered a plan-driven approach that incorporates agile elements. It was designed to follow five fundamental development steps that build around the features of the intended program over several one-to-two-week iterations. The FDD workflow steps are:

1. Develop an overall model of the software and write the project outline. This model is the basis for the initial planning of the project onto which each iteration’s designs are built upon. This can be done through updating existing diagrams or adding new diagrams that give more detail about specific features.
2. Build a feature list of all the requirements by breaking up the target functionality into smaller features. Due to the agile nature of FDD, this list is not necessarily final; however, the small nature of the project makes it likely that the list shown in *Figure 1* will be.
3. Plan the development of the features – determine the order in which the features will be developed and who will oversee developing them. A rough timeline for developing the features is shown in *Figure 2*.
4. Design by feature – features are selected for the next iteration and assigned to different teams (though not in this case). The features are designed and added to the overall model of the system.
5. Build by feature – once the designs are inspected, the feature is then implemented to that design. Usually, unit and acceptance testing are applied at this stage in the project process. However, due to the time constraints placed on the project, most of the testing was left to the end of the process.

Steps 4 and 5 are repeated until all the features have been designed, implemented, and tested. Once that is completed, the software produced should comprehensively fulfil the functional requirements set out as the features at the beginning.

### Feature Table and Initial Timeline

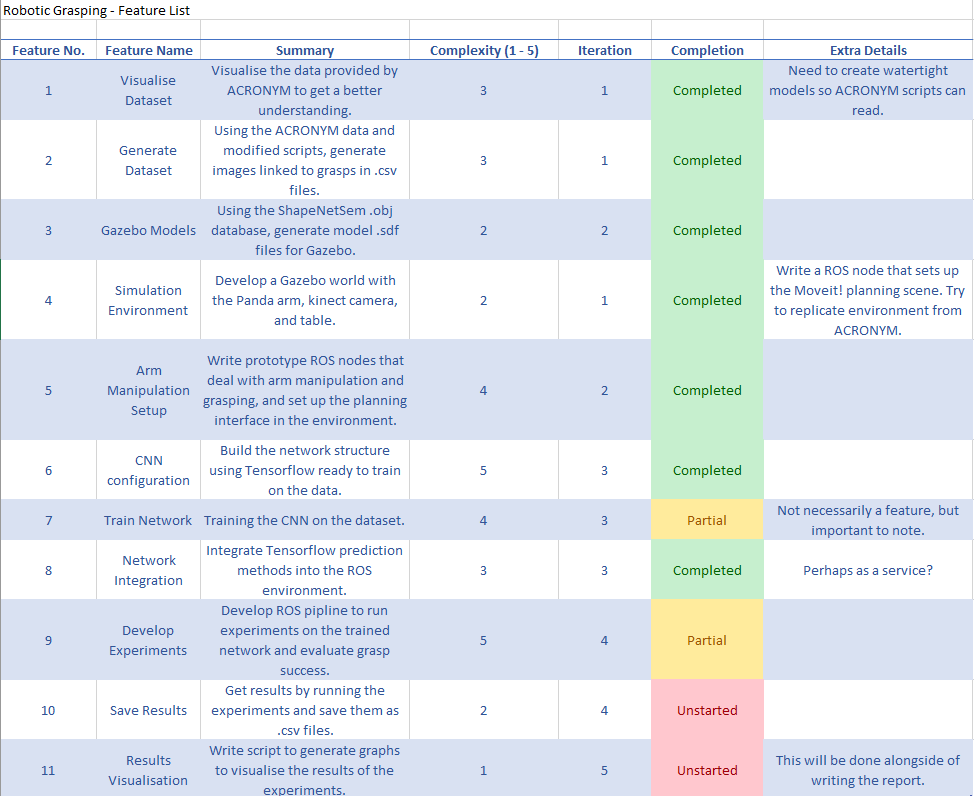


Figure 1: The feature table for the project, as seen in week 11.

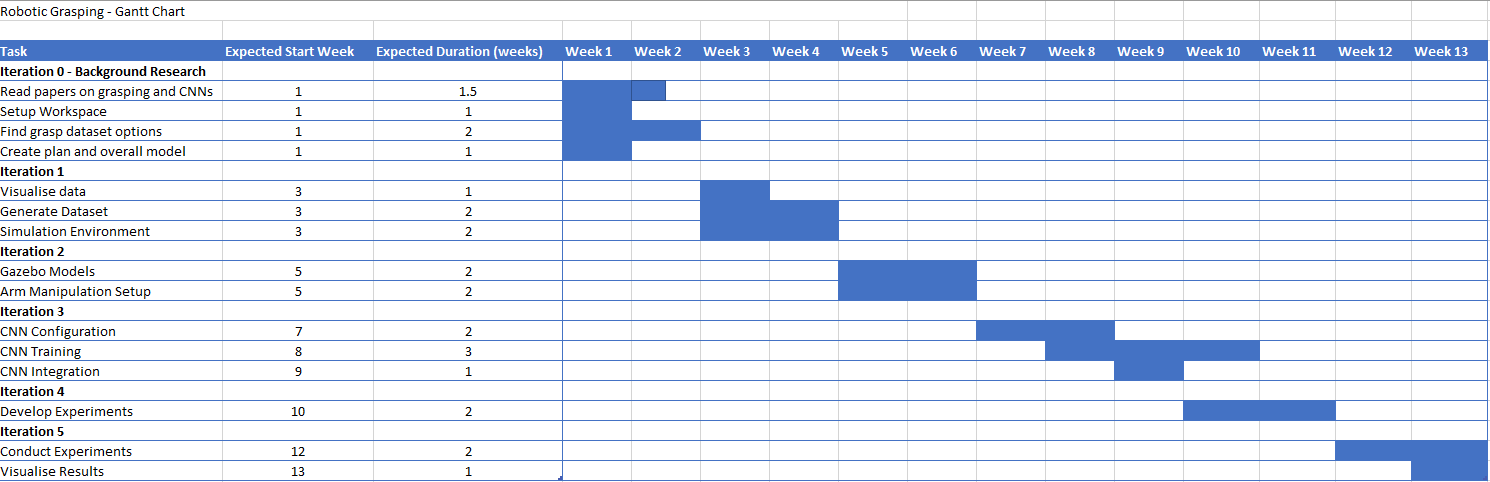


Figure 2: Initial Gantt chart to plan a rough timeline.

# Convolutional Neural Networks

## Background

## Technical Workings

# Experiment Methods

This project poses the question: can robotic grasps be successfully learnt by using a deep learning algorithm? *Chapter 3* will outline the original hypothesis made at the start of the process, and the experiments that were designed to evaluate the aforementioned hypothesis and the success of the project. The supporting software design will be covered in *Chapter 4*.

## Experiment Hypothesis

* Can a robot learn to grasp objects just from input images?
* This approach has been tried before – as mentioned in intro. Quote papers and success. Mention how this differs slightly as only in simulation and using different from norm data.

The hypothesis being investigated in this project is that if a deep convolutional neural network can learn to predict grasps from images, then the trained network can be used to grasp novel objects in simulation. This is functionality is trying to mimic the human intuition and experience that enables us as a species to pick up objects that we may have never encountered. This hypothesis has been tested in other papers, as mentioned in *Chapter 1*, with some reports quoting success rates of up to … This project is not necessarily searching for the most accurate neural network or the most successful grasps; due to time and technical limitations imposed due to the nature of the project (discussed further in *Chapter 4*). However, the expected outcome is a proof of concept that could be further refined to produce better results.

## Experiment Design

* Define what ground truth is
* What is a successful grasp?
* Two main measurements will be taken:
  + Original grasp success in simulation on the grasp dataset
    - The grasp dataset is seen as a ground truth for grasp success
  + Grasp success rate when using the CNN model to calculate the grasps
  + Add Model loss/accuracy during training?
* Using the original data to create a ground truth success will enable a better evaluation of the CNN performance.
  + A CNN trained on bad data is likely not going to perform as well as one trained on good data.
  + All machine learning model capabilities are limited by the quality of the ground truth.
* The experiment will also evaluate how the model performs on different types of objects. E.g., mugs, balls, chairs, etc. Whether certain objects are easier to learn to grasp for.

### Experiment Outline

The aim of this project is to test this hypothesis through experimentation in simulation. This will involve writing software and handling different data, further discussed in *Chapter 4*. The software and data will be used to test the hypothesis by taking the following two measurements:

* Evaluation of the ground truth.
* Evaluation of the trained network.

Before describing the experiments to take these two measurements, I must define two things within the scope of this project:

* What is a successful grasp?
* What is a ground truth?

Perhaps most importantly, for the purpose of this project, I will be defining a successful grasp as a combination of two factors. Furthermore, I will also be recording semi-successful grasps that come close to picking up the object (e.g., if the object is thrown to the side). The two factors that are going to be considered for grasp success are: gripper closure, and object movement within the camera frame. This will be further explained in the Design section of *Chapter 4*.

A ground truth in the field of machine learning is the ‘reality’ that the model is attempting to predict, in this case the ground truths are successful grasp patterns for specific objects. The ground truth is dictated by the input data for a model; hence, the capability and efficacy of all machine learning models is directly linked to the quality of the ground truth. This project is making use of the ACRONYM dataset, as discussed in *Chapter 1.3*, which contains grasp and object information in relation to the ShapeNetSem object database. This grasp dataset is then considered to be the ground truth for these experiments.

As stated before, the training data for a neural network directly impacts the trained output; therefore, it is necessary to determine the quality of the dataset that is being used. The ACRONYM grasp dataset does contain information on the success of each grasp from their physics simulator; however, this does not necessarily mean the labelled grasps will work in this project’s simulation. Therefore, it is essential to conduct an experiment to ascertain the accuracy of the grasps stated in the dataset within the project simulation environment, evaluating the ground truth of the scenario, thus allowing a comparison to the accuracy of the neural network.

### Ground Truth Evaluation

The ground truth experiment will be simple. The objects named in the dataset (a description of how this is produced is in *Chapter 4*) will be spawned into the Gazebo environment and each grasp labelled for that object will be attempted three times. Repeating the test enables an average success rate to be taken for both the different objects and the dataset as a whole.

* Extrapolating from a tested subset due to lack of time to test all grasps?

### Trained Neural Network Evaluation

The experiments to evaluate the efficacy of the trained neural network model will be similar to that of the ground truth data. The pipeline for this evaluation is shown in F*igure 1*. Firstly, a random model from the ShapeNetSem database is spawned into the Gazebo simulation. An image of the object will be generated from the camera in the environment and fed into the trained model, outputting the predicted grasp. This prediction will define a transform from the object, which will then be used to attempt the grasp. The grasp success will be monitored and recorded; looking for partial gripper closure and the object to be moving up in the camera frame for a fully successful grasp. This pipeline, as in the ground truth evaluation, will be repeated three times to enable the taking of an average, thus decreasing the likelihood of outliers in the results.

Spawn a random model

Generate grasp pose from image

Attempt the grasp

Evaluate the successfulness of the grasp

Figure 3: The pipeline for evaluating the trained neural network in simulation.

### Experiment Results

There will be two ways to evaluate the results of the project’s experiments:

* Looking at the average success rate of all objects.
* Looking at the category of object the model performs best on.

The first of these two options best evaluates the initial hypothesis laid out originally, as it focusses on the generalisation of the model and its ability to predict grasps for all kinds of objects. A high average success rate would show that the model (or a similar one) could be robust enough to be applied in many scenarios. On the other hand, a low success rate could be the result of different factors, such as: a low-quality ground truth, an incorrectly setup simulation environment, or a poorly built convolutional neural network.

The second option could perhaps be more interesting in the situation where the grasp success rate is mid-ranged (i.e., 40% - 60%). This is because in this situation the model appears to be working well to some degree; however, we might assume that some types of objects are holding it back. Looking at the success rate of the different category of objects could enable this to be fixed in future iterations of the project. There could be many reasons why some objects have a higher grasping rate than others, for example, the object is bigger or more complicated; therefore, will naturally be harder and more awkward to pick up. The object could even be too large for the gripper on the robot, though this can be mostly mitigated during this project due to its simulated nature.

By using both of these evaluation methods, we can generate a good assessment of how successful the model is performing, how it can possibly be improved, and how correct the original experiment hypothesis was.

# 

# Software Design, Implementation and Testing

* Describe what this chapter is about:
  + The design of different aspects
  + The implementation of the main aspects and the issues that arose during this process
  + How the implementation is tested, both observationally and automated unit testing

This chapter covers the three different stages for completing the supporting software for the experiments. Firstly, the design stage for assessing the requirements and designing the different elements. Secondly, the implementation of those features, also mentioning the problems faced while writing the software. Finally, the testing stage to try to ensure that the software and the data are correct so that the experiments run smoothly.

## Design

### Overall Architecture

* Two main design areas for the project:
  + The simulation and evaluation environment in ROS/Gazebo
    - The environment and robot manipulator
    - Important/specific algorithms
    - ROS node graph / interaction
    - Evaluation pipeline – sequence diagram?
  + The Tensorflow DCNN
    - CNN model structure
    - Integration and use in simulation
    - Preparation of dataset / pipeline of data
* Requirements table

As stated in the project analysis (*Chapter 1.3*), the project has been split into three fundamental tasks:

1. Generating an image – grasp dataset.
2. Training a CNN to predict grasps from images.
3. Building a simulation environment that will be used to run experiments.

By analysing these tasks and the experiments outlined in *Chapter 3*, a list of basic requirements or features can be established. These features can be viewed in more detail in Appendix …

### Simulation Environment

#### Neural Network

### Dataset

### Applied Technologies

Technologies used:

* Languages:
  + Python 2 & 3 for ROS and tensorflow respectively
  + C++ also for ROS
  + BASH script for SCW
* Tensorflow for the CNN
* ROS / Gazebo for simulation
  + OpenCv for image manipulation + depth detection
  + ROS services / actions

### Requirements Table

* Appendix maybe???

## Implementation

Issues:

* Panda robot didn’t want to grip things so changed to fetch robot which I knew worked in my environment

## Testing

### Overall Approach to Testing

### Automated Testing

#### Unit Tests

#### User Interface Testing

#### Stress Testing

#### Other Types of Testing

### Integration Testing

### User Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test No. | Description | Expected Outcome | Actual Outcome | Pass |
| T1 |  |  |  |  |
| T2 |  |  |  |  |
| T3 |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Results and Conclusions

## Ground Truth Success

## Model Loss and Accuracy

* Show graph and discuss both implementations

## Model Success in Simulation

* Discuss affect of different objects for detecting grasps
  + Which are easier?

# 

# Critical Evaluation

Examiners expect to find a section addressing questions such as:

* Were the requirements correctly identified?
* Were the design decisions correct?
* Could a more suitable set of tools have been chosen?
* How well did the software meet the needs of those who were expecting to use it?
* How well were any other project aims achieved?
* If you were starting again, what would you do differently?

Other questions can be addressed as appropriate for a project.

The questions are an indication of issues you should consider. They are not intended as a specification of a list of sections.

The evaluation is regarded as an important part of the project report; it should demonstrate that you are capable not only of carrying out a piece of work but also of thinking critically about how you did it and how you might have done it better. This is seen as an important part of an honours degree.

There will be good things in the work and aspects of the work that could be improved. As you write this section, identify and discuss the parts of the work that went well and also consider ways in which the work could be improved.

In the latter stages of the module, we will discuss the evaluation. That will probably be around week 9, although that differs each year.

# 

# Annotated Bibliography

This final section should list all relevant resources that you have consulted in researching your project. Each reference should also include a brief annotation.

1. Sylvia Duckworth. A picture of a kitten at Hellifield Peel. <http://www.geograph.org.uk/photo/640959>, 2007. Copyright Sylvia Duckworth and licensed for reuse under a Creative Commons Attribution-Share Alike 2.0 Generic Licence. Accessed August 2011.

This is my annotation. I should add in a description here.

1. Mark Neal, Jan Feyereisl, Rosario Rascunà, and Xiaolei Wang. Don’t touch me, I’m fine: Robot autonomy using an artificial innate immune system. In *Proceedings of the 5th International Conference on Artificial Immune Systems*, pages 349–361. Springer, 2006.

This paper…

1. W.H. Press et al. *Numerical recipes in C*. Cambridge University Press Cambridge, 1992.

This is my annotation. I can add in comments that are in **bold** and *italics*and then further content.

1. Various. Fail blog. <http://www.failblog.org/>, August 2011. Accessed August 2011.  
     
   This is my annotation. I should add in a description here.
2. Apache Software Foundation (2014) “*Apache POI - the Java API for Microsoft Documents*” (Online) Available at: <http://poi.apache.org> Accessed: 14th March 2014.

This is my annotation. I should add in a description here.

1. Apache Software Foundation (2004) “Apache License, Version 2.0” (Online) Available at: <http://www.apache.org/licenses/LICENSE-2.0> Accessed: 14th March 2014.

This is my annotation. I should add in a description here.

1. Neil Taylor, “MMP\_S08 Project Report and Technical Work”, 2019 (Online) Available at: <http://blackboard.aber.ac.uk/> Accessed 19th February 2019.

A document that outlines information about the marking guide for the Project Report and Technical Work. This is published in the Resources folder on Blackboard.

# 

# Appendices

The appendices are for additional content that is useful to support the discussion in the report. It is material that is not necessarily needed in the body of the report, but its inclusion in the appendices makes it easy to access.

For example, if you have developed a Design Specification document as part of a plan-driven approach for the project, then it would be appropriate to include that document as an appendix. In the body of your report you would highlight the most interesting aspects of the design, referring your reader to the full specification for further detail.

If you have taken an agile approach to developing the project, then you may be less likely to have developed a full requirements specification. Perhaps you use stories to keep track of the functionality and the ’future conversations’. It might not be relevant to include all of those in the body of your report. Instead, you might include those in an appendix.

There is a balance to be struck between what is relevant to include in the body of your report and whether additional supporting evidence is appropriate in the appendices. Speak to your supervisor or the module coordinator if you have questions about this.

* 1. Third-Party Code and Libraries

If you have made use of any third-party code or software libraries, i.e. any code that you have not designed and written yourself, then you must include this appendix.

As has been said in lectures, it is acceptable and likely that you will make use of third-party code and software libraries. If third-party code or libraries are used, your work will build on that to produce notable new work. The key requirement is that we understand what your original work is and what work is based on that of other people.

Therefore, you need to clearly state what you have used and where the original material can be found. Also, if you have made any changes to the original versions, you must explain what you have changed.

The following is an example of what you might say.

**Apache POI library** – The project has been used to read and write Microsoft Excel files (XLS) as part of the interaction with the client’s existing system for processing data. Version 3.10-FINAL was used. The library is open source and it is available from the Apache Software Foundation [5]. The library is released using the Apache License [6]. This library was used without modification.

Include as many declarations as appropriate for your work. The specific wording is less important than the fact that you are declaring the relevant work.

* 1. Ethics Submission

This appendix includes a copy of the ethics submission for the project. After you have completed your Ethics submission, you will receive a PDF with a summary of the comments. That document should be embedded in this report, either as images, an embedded PDF or as copied text. The content should also include the Ethics Application Number that you receive.

* 1. Feature Table

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Task** | **Description** |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* 1. Code Samples

This is an example appendix. Include as many appendices as you need. The appendices do not count towards the overall word count for the report.

For some projects, it might be relevant to include some code extracts in an appendix. You are not expected to put all of your code here - the correct place for all of your code is in the technical submission that is made in addition to the Project Report. However, if there are some notable aspects of the code that you discuss, including that in an appendix might be useful to make it easier for your readers to access.

As a general guide, if you are discussing short extracts of code then you are advised to include such code in the body of the report. If there is a longer extract that is relevant, then you might include it as shown in the following section.

Only include code in the appendix if that code is discussed and referred to in the body of the report.

Random Number Generator

The Bayes Durham Shuffle ensures that the pseudo random numbers used in the simulation are further shuffled, ensuring minimal correlation between subsequent random outputs.

// Some example code here…

Custom tensorflow code?