CMP3753M Project Proposal

Luke Roberts

University of Lincoln
25722923@students.lincoln.ac.uk

1 Introduction

Understanding the universe has been a pursuit of humanity for thousands of years. Because of advances in observational technology such as the Hubble Space Telescope (HST) and more recently the James Webb Space Telescope (JWST), we can observe very distant galaxies. The time taken for the light of distant galaxies to reach us means that our night sky is a window into the past, allowing astrophysicists to understand the evolution of our universe in more depth. Photographs such as the Hubble Ultra-Deep field show us an ancient universe full of developing galaxies [], and the amount of observed galaxies is only increasing. By identifying how galaxy morphology changes over time we have a clearer picture of the overall changes over the last several billion years. However, it is extremely time consuming for scientists to categorise galaxies in their research. A study in 2016 [1]calculated that there are 2.0×10^{12} galaxies in the observable universe, and it is therefore reasonable to suggest automating galaxy categorisation using computers. One way to achieve this is to train a deep learning model to catagorise images of galaxies.

Deep Learning has recently revolutionised both scientific research and modern life in a profound way. From categorisation of X-rays images in the medical field; machine translation of natural language such as Google Translate and large text generation models such as ChatGPT, deep learning has become the best way to categorise and generate unstructured data. To do many of these tasks which were once thought impossible, machine learning engineers create deep learning models which utilise a dataset that learns patterns about that data. Image classification is a form of deep learning that uses images as input data and is used for a wide variety of applications in scientific research. Like all forms of deep learning, image processing models are trained so that they more accurately categorise new input data []. At the start of training, the model performs poorly when tested, however through analysing how the output fails, the model can tweak its own parameters in order to achieve greater accuracy. The choice of which machine learning model to use and the specific hyperparameters are important for maximising the efficiency of the model. This can be due to the content and quality of the training data being used.

2 Aims and Objectives

2.1 Aims

2.2 Objectives

3 Project Plan and Risk Analysis

3.1 Project Plan

3.2 Risk Analysis

References

1. Castelvecchi, D. (2016). Universe has ten times more galaxies than researchers thought. Nature. https://doi.org/urlhttps://doi.org/10.1038/nature.2016.20809.