Project Proposal

DSE313 - Artificial Intelligence

Submitted to - Dr Vaibhav Kumar

Team Members	
Mehul Paithane	18142
Yatharth Pal	19340
Yuvraj Sharma	19288

Motivation

With exponential growth in the datasets available for deep space imagery, manual classification of galaxies into categories continues to become less feasible. Efficient and automated classification of galaxies into categories based on their morphologies is required.

The aim of this project is to classify any galaxy morphologically based on its image by training a CNN model.

Research

1. Radio Galaxy Zoo: Compact and extended radio source classification with deep learning V. Lukic, M. Brüggen, J.K. Banfield, O.I. Wong, L. Rudnick, R.P. Norris, B. Simmons https://arxiv.org/abs/1801.0486
2. Galaxy Zoo classification with Keras James https://towardsdatascience.com/galaxy-zoo-classification-with-keras-219184aff581
3. Dataset https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Problem Definition

The main problem is to classify the given image of any galaxy in the test data by their morphologies into different given categories.

Contributions

- This method of classifying galaxies has increased the speed of discoveries and testing theories by heavy margins, as compared to earlier when Humans used to personally observe and classify them.
- These galaxies after classification can be further used for ontologies related to Galaxy classification or Space objects, stored and provides ease-of-access to information related to such galaxies.
- Later such models for classification might be used in the telescopes themselves, where they themselves spot objects and make a better navigator for objects in the night sky.
- Telescopes like the James Webb Telescope, which are being sent into deep space for future missions, can hold these models and analyze celestial objects themselves.

Objectives

A model will be trained to identify the classification of the morphological types of galaxies.

Proposed Methodology

There are around 60000 images in the training dataset which we will use to train the CNN model and another 90000 images to test this model.

The first step would be to divide this huge training dataset into small batches and tackle them one after another as the GPU memory would not be able to handle these many images at once.

The second step would be to make a CNN model, as CNNs have proven to be highly efficient and effective in classifying objects in image data.

The third step would be to test this model and see how accurate our model is.

Possible Results

The possible expected result would be to find the probability for any given galaxy to belong to a certain class of galaxies