Final Project Proposal – Group1

Problem Statement:

The Project aims at developing the Neural Network using CNN Architecture to classify galaxy images. We have seen in our course that the CNN implementation for image classification problems gives us better accuracy over unseen data in comparison to other deep learning networks. Hence, CNN.

The network will be trained over the Galaxy Zoo data set, which comprises 61,578 images of the galaxies in the training set and 79,975 in the testing set (which will be used as validation set in the project). The data set has 37 labels representing the morphology (or shape) of the galaxy in 37 different categories. We will use a standard CNN deep learning network with max pooling. Frameworks like PyTorch or TensorFlow will be used as background. The trained model will be utilized for checking model's accuracy, training and validation loss, confusion matrix, F1 and Cohen Kappa scores over the unseen testing set. The dataset is big enough and will be uploaded on AWS cloud for access and computation to check the performance of the network. The project work is distributed among all the team members and rough estimated time frame of code completion and group report is April 20th, 2020.

Link to the Dataset:

https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/data

Reference Links:

https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html
https://towardsdatascience.com/an-image-classifier-with-deep-learning-7284af97b36a
https://web.stanford.edu/class/cs231a/prev_projects_2015/C231a_final.pdf
https://towardsdatascience.com/galaxy-zoo-classification-with-keras-219184aff581