

STAT 679 Final Project Proposal

Xiaoyang Wang

Ziang Zeng

1 Astronomical Challenge

Our research focuses on classifying celestial objects into stars, galaxies or quasars using their spectral characteristics. With the advancement of astronomical technology, we can obtain a large amount of data, including images and spectral information, from telescopes and large-scale photometry.

The central question of our research is: “How can we effectively use image and spectral data to accurately classify different types of stellar objects?”

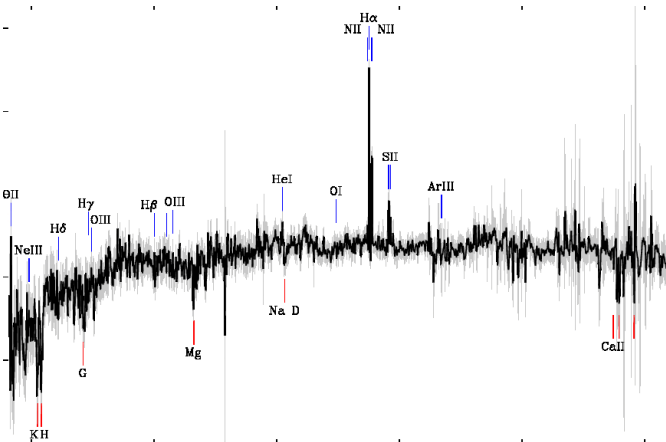
2 Data

We plan to work with the astronomy dataset containing three types of data:

1. Image of the celestial objects:



2. Image of the spectra of the celestial objects.



3. Metadata of the celestial objects, whose variables are as follows:

vars	explanations
objid	Object Identifier
ra	Right Ascension angle (at J2000 epoch)
dec	Declination angle (at J2000 epoch)
u	Ultraviolet filter
g	Green filter
r	Red filter
i	Near Infrared filter
z	Infrared filter
run	Run Number
rerun	Rerun Number
camcol	Camera column
field	Field number
specobjid	Unique ID used for optical spectroscopic objects
class	Object class
redshift	Redshift value based on the increase in wavelength
plate	Plate
mjd	Modified Julian Date
fiberid	fiber ID
plateid	Plate ID

They can be found in this [link](#).

3 Methods

We plan to employ both Deep Neural Networks (DNNs) and statistical models for object classification within the Sloan Digital Sky Survey (SDSS). For image classification, we have selected several DNN candidates based on their proven performance: ResNet, which supports very deep networks through its residual structure; VGG, which utilizes small-sized convolutional kernels (3x3) and pooling layers; and ResNeSt, which enhances the ResNet architecture by introducing a split-attention mechanism. Regarding statistical models, we intend to use either classification trees or logistic regression models, incorporating ensemble learning methods such as bagging or boosting to improve performance. Our ultimate goal is to develop a hybrid model that combines image classifiers with meta-classifiers, aiming for superior accuracy.

4 Contribution

The workload for this project will be equally distributed among us. We will collaboratively preprocess the data by merging and visualizing it. As outlined, one team member will be tasked with developing the statistical models, while another will focus on constructing and fine-tuning the DNN models. Following the training of these separate models, we will integrate their outcomes using techniques such as weighted averaging or regression to create a final model. This combined model is expected to achieve enhanced accuracy, leveraging the strengths of both deep learning and statistical approaches.