MULTI-WAVELENGTH GALAXY STRUCTURE DECOMPOSITION USING SUPERVISED MACHINE LEARNING

Deralaza RAFIEFERANTSOA

ABSTRACT

This poster summarizes the current progress of a project developing a supervised machine learning (ML) framework to decompose galaxy structures using multi-wavelength data. A dataset from the 50 Mpc Galaxy Catalog (50MGC) has been assembled, and standardized g, r, i, z data-cube FITS images have been downloaded and generated. A full preprocessing pipeline covering star removal, segmentation, masking, and band normalization has been completed. A convolutional neural network (CNN) architecture has been selected for morphological classification and structural analysis, and initial model training is underway. Ongoing and future efforts will focus on model validation and multi-component decomposition of bulges, disks, and bars to support studies of galaxy evolution.

INTRODUCTION

Galaxy morphology is fundamental for understanding galaxy formation and evolution. Traditional structural decomposition techniques rely on manual or semi-automated parametric fitting, which becomes impractical for the large datasets produced by modern astronomical surveys. Machine learning offers scalable solutions for automating galaxy classification and structural analysis.

OBJECTIVES

- Build a multi-wavelength galaxy images dataset
- Develop a supervised machine models for galaxy classification
- Automate structural decomposition (bulge – disk -bar)
- Apply the framework to study galaxy evolution and physical properties

DATA

- Source: Multi-band FITS image cubes **(g,r,i,z)** from the **DESI Legacy Imaging Surveys**
- Current dataset: 10,157 galaxy images from **50 Mpc Galaxy Catalog (50 MGC)**

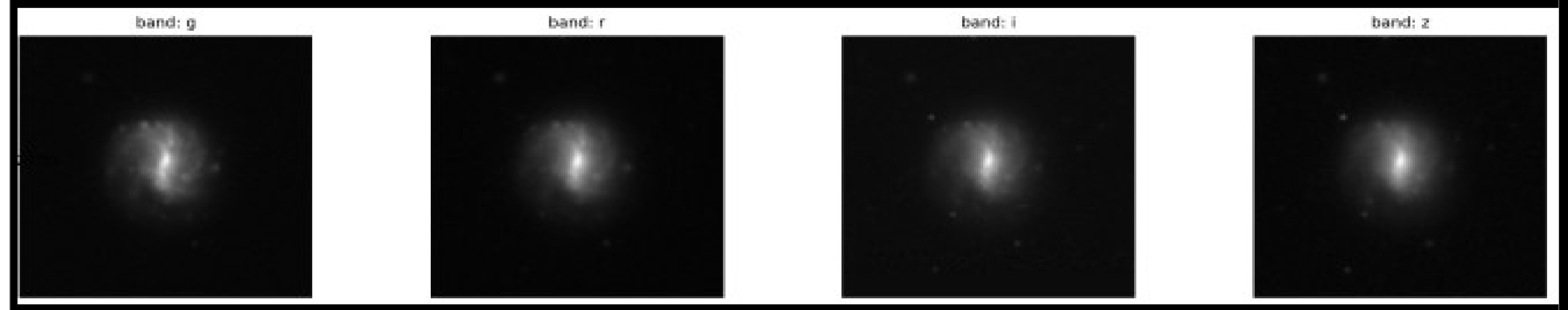

Figure 1: Representative multi-band FITS image cubes in the g,r,i, and z bands

IMAGE PRE-PROCESSING AND DATA PREPARATION

- Applied inter-band normalization and image alignment to ensure consistency across different wavebands.
- Standardized image pre-processing pipeline: foreground star removal, source detection, segmentation, and masking, utilizing Python packages such as **photutils** and **galmask**.

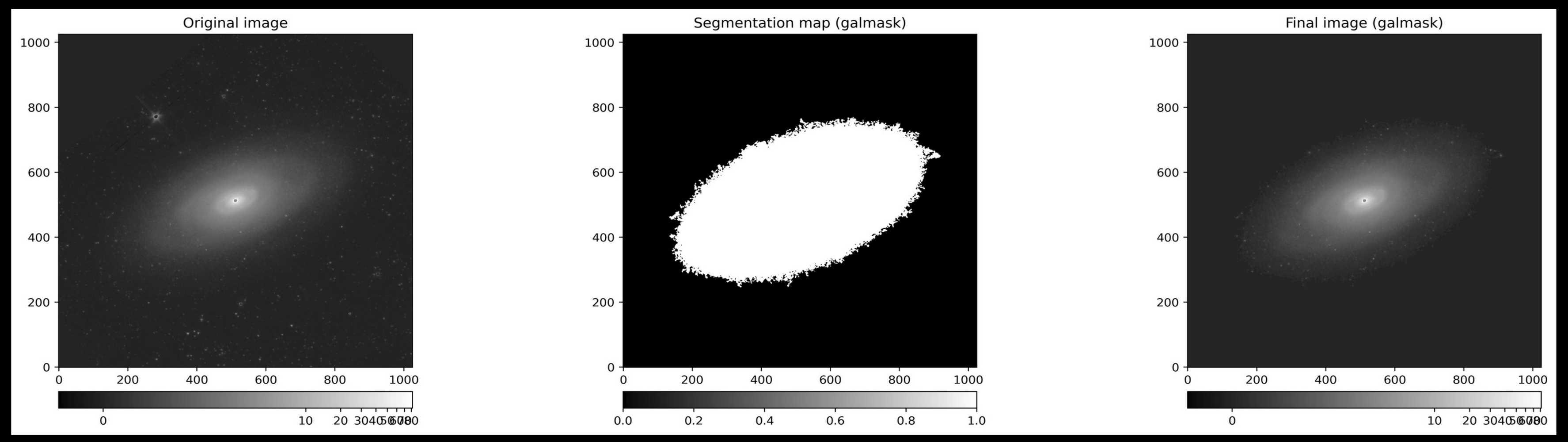

Figure 2: Standardized image pre-processing workflow, showing the original image, the generated segmentation map (galmask), and the final masked image used for training .

MACHINE LEARNING FRAMEWORK DESIGN

- Designed a supervised learning framework based on CNN architectures suitable for multi-channel (multi-wavelength) galaxy images.
- Established the training environment using TensorFlow, with experiments conducted on Google Colab for stability and computational efficiency.

PRELIMINARY RESULTS

Initial experiments were performed using a subset of ~380 galaxies across four morphological classes (Elliptical, Lenticular, Spiral, Irregular). The pipeline runs stably, though classification accuracy is currently limited due to small sample size and ambiguous galaxy morphologies.

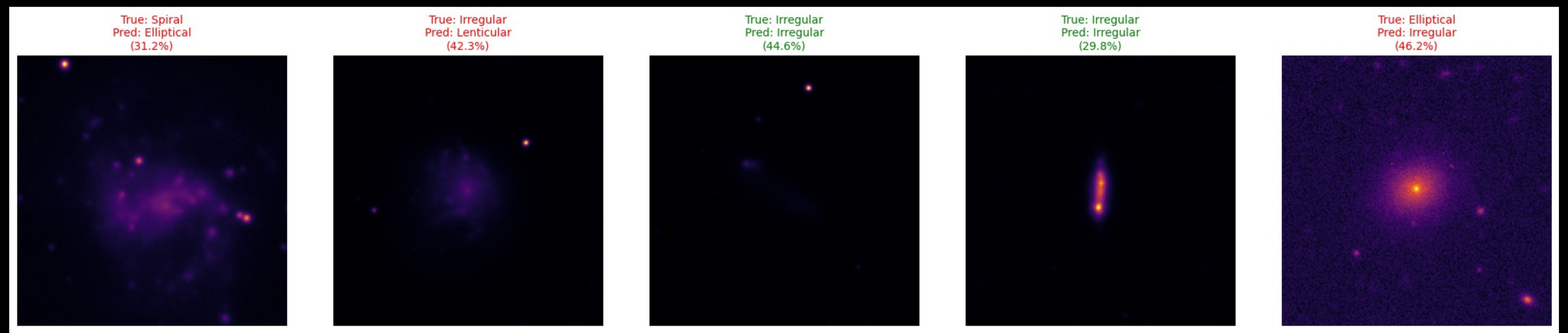

Figure 3: Preliminary classification outputs from the CNN

NEXT STEPS

- Refine the training sample using reliable **T-type classifications**
- Compare **raw vs cleaned images**
- Automated structural decomposition
- Model interpretability using saliency maps

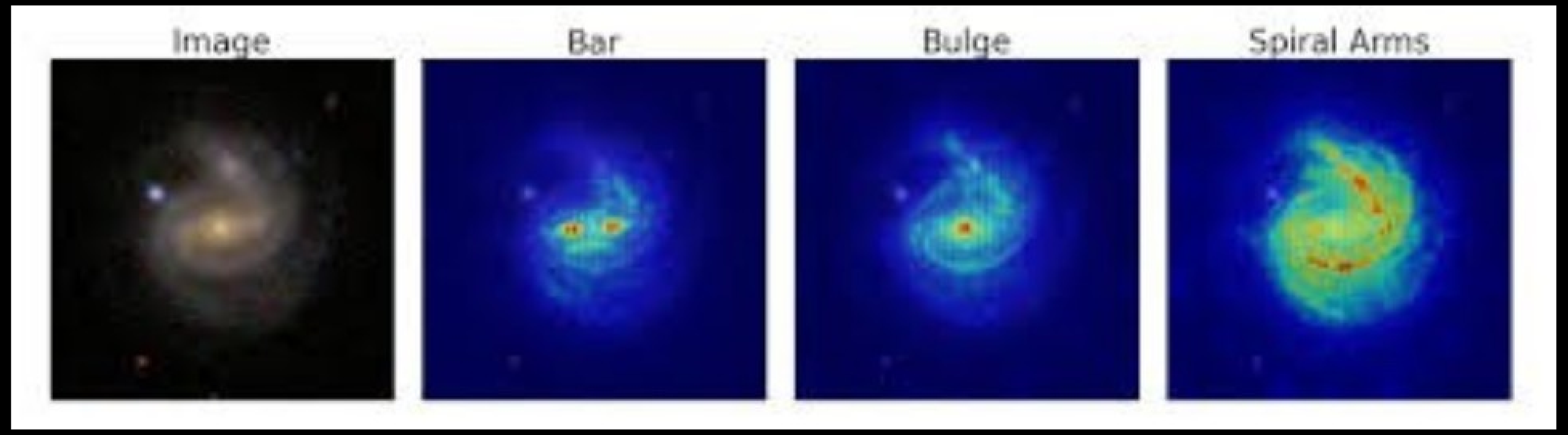

Figure 4: Visualization of planned structural decomposition (Bhambra et al., 2022)

- Extract **1D surface brightness profiles** to improve physical interpretation and to validate predictions

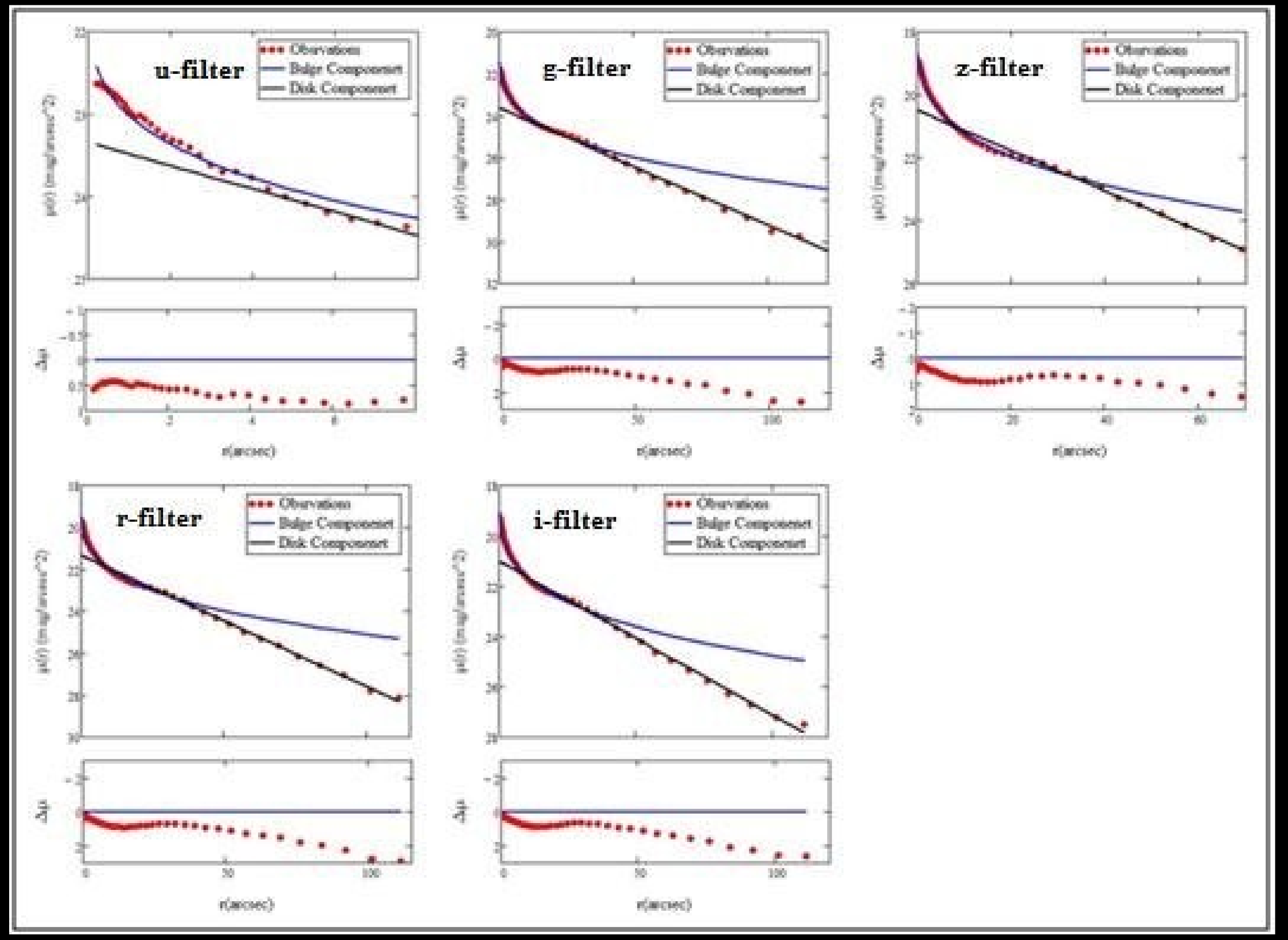

Figure 5: Surface Brightness Profile Decomposition
(Ahmed et al., 2016)

- Generate **synthetic galaxy images with controlled parameters** to augment the training dataset and to improve robustness

REFERENCES:

- Bhambra, P. et al. (2022). MNRAS, 511(4), 5032–5041
- Ahmed et al. (2016), Iraqi J. Sci., 57(3A), 1860–1866