GALAXY MORPHOLOGY CLASSIFICATION
(BEETLEJUICE 2.0)

INTRODUCTION

The Galaxy morphology is really important to understand how the Galaxy was formed and how it has changed over time. When we look at the Galaxy we can see things, like arms and bars and bulges. These things are formed because of the way the Galaxy pulls on itself with gravity. Because of what is happening inside the Galaxy. It takes a long time for these things to happen.

The problem is that it can be very hard to tell the types of Galaxy morphology apart. This is because they can look very similar and there can be things that get in the way of our view like noise and we might not be able to see everything clearly. The Galaxy morphology is what we are trying to understand so we need to be able to see the Galaxy morphology.

Convolutional Neural Networks (CNNs) are well suited for this task, as they can automatically learn hierarchical spatial features from images. In this work, we develop a CNN-based pipeline to classify galaxy morphologies into multiple classes using labeled image data.

OBJECTIVE

The objective of this task is to build a multi-class image classification model that predicts the morphological class of a galaxy from an input image. The model is trained to distinguish between ten galaxy morphology classes, including spiral and barred spiral galaxies, edge-on systems, merging galaxies, and smooth ellipticals.

DATASET AND EXPLORATION

We are working with the Galaxy Morphology Classification dataset that you can find on Hugging Face. This Galaxy Morphology Classification dataset has a lot of RGB galaxy images. These galaxy images are labeled into ten morphological categories. We looked at each galaxy image to make sure it was loading correctly and that the labels were consistent, for the Galaxy Morphology Classification dataset.

Initial exploration showed that galaxy morphologies vary significantly in appearance and that class frequencies are imbalanced, with some classes appearing much less frequently than others. The dataset was split into training and validation subsets for model development and evaluation.

PREPROCESSING

All images were resized to a fixed resolution of 224×224 pixels to ensure compatibility with standard CNN architectures. Pixel values were normalized using ImageNet mean and standard deviation values. No augmentation was applied to the validation set in order to ensure a consistent and fair evaluation.

MODEL ARCHITECTURE AND TRAINING

We used a ResNet-18 network because it works really well and does not need a lot of computer power. The last layer of the ResNet-18 network was changed so it could tell us which of the ten morphology classes something belonged to. We changed the ResNet-18 network to make it better at predicting the morphology classes.

The model was trained using cross-entropy loss and optimized with the Adam optimizer. Training was performed for three epochs while monitoring both training and validation performance. The model was trained for three epochs. Although the final epoch achieved higher training accuracy, the best validation performance was observed at epoch 2. This indicates the onset of overfitting beyond epoch 2, and hence early stopping at epoch 2 would be optimal.

EVALUATION AND RESULTS

The model achieved a peak validation accuracy of approximately **79%** at the second training epoch. Although training accuracy continued to increase beyond this point, validation accuracy saturated and validation loss increased in the third epoch, indicating the onset of overfitting. Consequently, the model corresponding to epoch 2 provides the best generalization performance. Performance was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix.

INTERPRETATION

The confusion matrix shows us that galaxy classes that look similar in shape have the rates of being mixed up. For example barred spiral galaxies are often mistaken for loose spiral systems because the bar is hard to see. When we look at galaxies from the side and they have a bulge in the middle the model gets it right most of the time. This means the model is good at

recognizing features that we can see easily like the shape of the galaxy. The galaxy classes that are similar in shape, like the barred galaxies are still the ones that get mixed up the most.

Several morphology classes do not appear in the validation set, resulting in zero entries in the corresponding rows of the confusion matrix. As a result, performance metrics for these classes cannot be reliably assessed. This behavior reflects class imbalance in the dataset split rather than a limitation of the model itself.

GENERALIZATION AND ROBUSTNESS

To make things better and stronger we can try some things with the data when we are training. We can turn the data around, flip it sideways and make it brighter or darker. These changes help the data keep its meaning and make it work better when things are seen from angles or in different light. Data augmentation techniques like these are really helpful. Data augmentation techniques can really improve the data. Additionally, robustness could be evaluated by introducing mild noise or blur to validation images, simulating observational perturbations commonly encountered in astronomical imaging. Probability calibration techniques such as temperature scaling could further reduce the overconfidence observed at later training epochs.

To see how well the model works we tried something during training. We changed the pictures a bit by flipping them, rotating them a little and making them brighter or darker. These changes do not alter what the galaxy morphologies mean. They do help the model get used to seeing things from different angles and in different lighting, which is what galaxy morphologies and model robustness are all about and that is why we used galaxy morphologies and data augmentation to test model robustness.

The augmented training pipeline was designed to reduce overfitting observed in the baseline model by increasing input diversity. While full convergence was not pursued due to computational constraints, this approach is expected to improve generalization by preventing the model from relying on spurious visual cues. Such augmentation strategies are particularly relevant for astronomical imaging, where observational conditions can vary significantly.

CONCLUSION AND LIMITATIONS

This project used a kind of computer program called a CNN to look at pictures of galaxies and figure out what they look like. The computer program was really good at it. Got the right answer most of the time. When they made mistakes they actually made sense. Told us something about how the program was thinking.. If we let the program keep trying to learn from the pictures for

too long it started to get a little too good at recognizing the pictures it had seen before and not good enough at recognizing new ones. This shows that we need to stop the program from learning after a while and make sure it does not get too carried away. Galaxy morphology classification is what the project was about and CNN was really good at it.

Limitations of this work include class imbalance in the validation set and the lack of robustness testing under severe observational noise. Future work could address these limitations through balanced sampling strategies, data augmentation, and calibration-aware training.