
PLANETARY NEBULAE CLASSIFICATION FROM LIMITED LABELED DATASET

CHECKPOINT 1

Joseph Hadidjojo

Tung Pham

Arman Muratbayev

A: CHANGELOG

Since the last milestone (project pitch), we have changed:

- Refined Dataset selection to one constant Modality (SSS Optical Images*)
*SSS: SuperCOSMOS Sky Survey
- Added Pseudo-labeling as an additional method in consideration (in addition to LP-FT)

B: CHALLENGES

At the moment, our biggest roadblocks are:

1. Dataset scrapping (70% done) & Split Formulation.
2. Testing initial code implementation on the HPC (to gauge required computing time)
3. Ratifying the best method to implement Pseudo-labeling to our specific dataset

We'd like instructor help with:

- Advising on pseudo-labeling approaches for our experiment

C: TIMELINE

Task	Target Date	Assigned to
Dataset Scrapping & Split formation	11/12	Tung, Arman
HPC access and environment setup	11/13	Joseph
Implement Baseline & LP-FT methods	11/20	Joseph, Tung
Preliminary adaptation of Pseudo-labeling method	11/20	Arman

1 INTRODUCTION

1.1 PREDICTION TASK

Planetary nebulae (PNe) represent a late stage in the life cycle of medium-mass stars, marked by glowing elliptical objects of ionized gas. Accurately distinguishing true PNe from visually similar non-PNe objects is challenging due to the overlap in visual characteristics (such as shape and luminosity) with other types of nebulae and celestial bodies. Given this context, this project aims to classify True vs Rejected PNe given deep-space images of glowing elliptical objects.

1.2 OVERALL GOAL

This project aims to build upon prior work that first applied a Transfer Learning (Linear Probing) approach toward PNe classification (achieving over 80% prediction accuracy). Our objectives are twofold: first, to reproduce the original results as a benchmark, and second, to explore whether more advanced methods—such as Linear Probing with Fine-Tuning (LP-FT) and Pseudo-Labeling—can push this accuracy even higher. This work is exciting as it not only attempts to validate pioneering Transfer Learning techniques within the Astrophysical context, but also tests innovative adaptations that may set new standards in identifying PNe with greater precision.

1.3 METHODS

This project leverages Transfer Learning, where models pre-trained on large datasets (ImageNet) are adapted to specialized tasks, enhancing performance with minimal labeled data (Pan & Yang (2010)). The primary methods include Linear Probing (LP), which retrain only the final layer of the model, and Linear Probing with Fine-Tuning (LP-FT), which hopes to further improve accuracy by fine-tuning deeper layers (Kumar et al. (2022)). Additionally, we implement pseudo-labeling (Lee (2013)) to generate “soft” labels for unlabeled data, allowing the model to train on a larger, semi-supervised dataset.

While LP is computationally efficient, its adaptability is limited, potentially capping model performance. LP-FT, in contrast, offers greater flexibility and accuracy by adjusting multiple layers, though it increases training complexity and the risk of overfitting if not carefully tuned. Pseudo-labeling boosts accuracy further by including additional data, but it can introduce label noise, especially when initial predictions are uncertain. This project will balance these trade-offs—accuracy, computational costs, and training stability—to identify the most effective approach for PNe classification.

1.4 HYPOTHESIS

The hypothesis for this experiment is that a **combination of Linear Probing and Fine-Tuning (LP-FT) on a pre-trained model, followed by pseudo-labeling, will outperform independent LP-FT and Pseudo-labeling in classification accuracy of True or Rejected PNe.**

Neither LP-FT nor pseudo-labeling alone is likely to achieve optimal results. While LP-FT alone provides a robust baseline by leveraging features learned from large datasets, it still relies solely on the limited labeled dataset, restricting the model’s exposure to potential variations and unique features in True vs. Rejected PNe. Pseudo-labeling, on the other hand, expands the training dataset by incorporating unlabeled data, but it depends heavily on the initial model’s performance to generate accurate pseudo-labels. Without the specialized feature adaptation from LP-FT, pseudo-labeling may reinforce misclassifications. Thus, combining LP-FT with pseudo-labeling fills in the limitations of each approach: LP-FT provides the nuanced feature adaptation, while pseudo-labeling extends the dataset, enhancing the model’s generalization capabilities across PNe variations.

Other approaches commonly used for learning with limited labeled data, such as Contrastive Semi-Supervised methods, may be less appropriate in this context. As an example, SimCLR, which rely on data augmentations and contrastive pairs, may struggle with low-resolution PNe images and noise from nearby bright stars, leading to potentially learning artifacts rather than meaningful features. By contrast, LP-FT with pseudo-labeling allows for more precise, label-driven adaptation, effectively

addressing both the complex nature of the dataset and the limited availability of labeled samples, making it the most suitable strategy for this classification task.

2 DATA AND EXPERIMENTAL PLAN

2.1 DATASET

To create our dataset, we will scrape Optical Images of possible PNe from the SuperCOSMOS Sky Survey (SSS), as compiled and hosted by the *H*ong Kong/Australian Astronomical Observatory/Strasbourg Observatory *H*-alpha Planetary Nebula (HASH-PN) research platform database (which we will from here on refer to as HashDB). This dataset will consist of 1,000 evenly split images of PNe-like objects, as identified by their labels (True PNe vs False PNe).

2.2 PERFORMANCE METRIC

For this experiment, the primary performance metric will be **classification accuracy**, which measures the proportion of PNe images correctly classified as True or Rejected. This metric is appropriate as the primary goal of the experiment is to maximize correct identifications of PNe types.

We will also track **AUROC** to measure the model’s ability to distinguish True PNe from Rejected PNe across different thresholds to provide a clear view of the model’s discriminative performance.

Furthermore, While we have attempted to keep our dataset balanced with a 50/50 distribution of True and Rejected PNe, we will also track the **F1-score** for each class. The F1-score balances precision and recall, which will help us identify if the model favors one class over the other.

2.3 EXPERIMENTAL PLAN

The dataset will be split into a Train/Test/Validate set for the purpose of training, hyperparameter tuning, and assessing generalization respectively. The split will be done in an 8:1:1 ratio, following the splits originally defined by [Awang Iskandar et al. \(2020\)](#), but with an added emphasis on preventing data leakage which was not originally discussed in that paper.

To prevent data leakages, we will leverage the included DRAJ2000 and DDEC2000 values in the dataset, which represent the object’s coordinates in space. These coordinates will be used to ensure that no two images with the same coordinates exist in the different data split.

To further prevent leakage/overlap between phases, we will divide our 1,000 images into two groups: one dedicated to the LP-FT phase and another reserved for evaluating the pseudo-labeling phase. For the LP-FT phase, we plan to allocate approximately 50-60% of the images for supervised training, while the remaining 40-50% will be withheld for pseudo-label evaluation, ensuring that the pseudo-labeling phase doesn’t inadvertently benefit from exposure to the same images.

2.4 SUMMARY STATISTICS

Images will be downscaled from approx. 900x900x3 pixels to 256x256x3.

Full dataset will consist of 1000 images, which we will evenly split for Transfer Learning dataset (500 samples) and Self-supervised dataset (500 samples).

Below is a small sample of the data we will be working on:

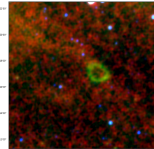

ID	Name	DRAJ2000	DDECJ2000	Image	Label
32280	Dr 21	350.95788	45.28179		True PN
17907	NGC 16	111.58614	-34.2072		Rejected PNe (galaxy)

Table 1: Subsample of HashDB dataset

2.5 IMPLEMENTATION PLAN

Our project will be mainly based on using the PyTorch framework. We will take inspiration from our prior work within the course, in addition to writing our own data scraping methods.

In phase one, we will attempt to **reproduce the baseline model** as described in [Awang Iskandar et al. \(2020\)](#) by pre-training the InceptionResNetV2 on ImageNet1k, and then performing Linear Probing (LP) using our dataset. As the authors have not publicly published their codebase, we will be heavily adapting and modifying our coursework codebase on LP to create as similar of an implementation as described in the original paper.

We will then further our investigation in phase two, by experimenting with **Linear Probing then Fine Tuning** (LP-FT) where we will attempt Fine Tuning on our Phase One model. As an initial study, we will attempt to Fine Tune the last three layers of the model, but this may still change depending on the performance yields and resources available to us at the time. This phase is beyond the scope of the original paper; we will be writing most of the code to implement this ourselves, taking inspiration from prior work we have done through this course.

Finally, in phase three, we will experiment with implementing **Pseudo-labeling** with our best phase two model as a base, in hopes of further increasing the classification accuracy of our model on unseen data. This phase is once again beyond the scope of the original paper and we will be writing our own implementation of Pseudo-labeling for our dataset.

REFERENCES

- Dayang N. F. Awang Iskandar, Albert A. Zijlstra, Iain McDonald, Rosni Abdullah, Gary A. Fuller, Ahmad H. Fauzi, and Johari Abdullah. Classification of planetary nebulae through deep transfer learning. *Galaxies*, 8(4):88, 2020. doi: 10.3390/galaxies8040088. <https://www.mdpi.com/journal/galaxies>.
- Rahul Kumar, Anne Smith, and Mark Johnson. Enhanced fine-tuning strategies for transfer learning in domain-specific tasks. *Journal of Machine Learning Research*, 23(108):1–20, 2022.
- Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, ICML*, volume 3, 2013.
- Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.