# Gravitational Lens Finding for Dark Matter Substructure Pipeline

## Personal Information


- Name : Adityam Ghosh
- Email : adityam.ghosh@gmail.com
- LinkedIn: https://www.linkedin.com/in/adityam-ghosh/
- GitHub: https://github.com/lucifermorningstar1305
- Country of Residence: New Zealand
- Timezone: NZST (UTC+12:00)
- University: University of Auckland
- Major: Data Science


## Project Description

- Name: Gravitational Lens Finding for Dark Matter Substructure Pipeline
- Description: Strong gravitational lensing is a promising probe of the substructure of dark matter to better understand its underlying nature. Deep learning methods have the potential to accurately identify images containing substructure, and differentiate WIMP particle dark matter from other well motivated models, including vortex substructure of dark matter condensates and superfluids.
- Goal:
  - **Extension** : The project will focus on the further development of DeepLense pipeline to incorporate SOTA vision models into the pipeline.
  - **Focus**: The focus of this project is using SOTA vision models along with other Linear Networks to augment the performance of DeepLense algorithms with mock survey data.
  - **Expected Result:** A joint vision + linear network model for DeepLense training and inferencing.

## About Me

I am a highly motivated graduate student currently pursuing a Master's degree in Data Science at the University of Auckland. I have 2.5+ years of working experience as a Machine Learning Engineer. Along with Python, PyTorch, Scikit-Learn, Pandas and other frameworks, I also have experience with ONNX which is a cruicial framework for optimizing models for runtime inferencing.

Previously I haven't participated in any Google Competetions but I have tried my hands on some **Kaggle competetions** and currently I hold a **Kaggle Notebook Expert** rank and slowly moving myself to the Master title. You can see my Kaggle profile in here. I am currently in my final year of my Master now and working on my dissertation that involves latest SOTA models in Natural Language Processing and Natural Language Entailment.

# Homework

- Installed necessary libraries for developing the pipeline.
- Study the latest models tried and tested in the DeepLense pipeline.
- Accomplished the qualification task posted by mentor in here
  - Converted the image to be processable for the lastest SOTA vision model.
  - Implemented a classification algorithm using **Efficient Net** (for the images) and a **Shallow Neural Network** (for the survey data) for classifying dark matter images into substructure and no-substructure.
  - Added different data augmentation into the pipeline to cover possible edge cases and increase training data size.

# Plan

- Given an image $I$, the pipeline will first process the image using the function $f(I) \rightarrow I'$ which will produce the new processed image $I'$. suitable for the vision transformer model $m$.
- The new image $I'$ will undergo several augmentation $A(I)$ to cover almost all edge cases and increasing training data size.
- The final set of images will then be batched into batches $B = [b_1, b_2, \cdots]$ and then passed into the vision model $m(B)$.
- The mock surveys $m_s$ are passed onto a different network $S$.
- The final model $m'$ is trained on the combined representation of mock-surveys and images to classify if the images are from lenses or not.

# Timeline

## Official overview

- **May 4th** - accepted GSoC contributor projects announced
- **May 4th-28th** - Community Bonding Period | GSoC contributors get to know mentors, read documentation, get up to speed to begin working on their projects
- **May 29th - July 13th** - Coding officially begins!
- **July 14th** - Midterm evaluation deadline (standard coding period)
- **July 14th - August 21st** - Work Period | GSoC contributors work on their project with guidance from Mentors

- **August 21st - 28th** - Final week: GSoC contributors submit their final work product and their final mentor evaluation (standard coding period)

## Details

- **Before May 4th**
  - get familiar with the DeepLense pipeline.
  - get familiar with Dark Matter Physics in the angle of Computer Science.
- **May 4th - 28th**
  - Engage with my mentors and community.
  - Find various other methods are suitable for progressing the project.
  - Clarify the DeepLense pipeline in details.
- **May 29th - July 13th**
  - Implement the image processing tools.
  - Look for additional data collection techniques (especially for the minority classes).
  - Add baseline model to the pipeline.
  - Check for data errors and try rectify them.
- **July 14th - August 21st**
  - Implement the standard model in the pipeline.
  - Run different cross-validation technique to obtain unbiased error estimates and compare results with previously published models.
  - Run a different angle on the approach - adapt anomaly detection algorithm on the data and compare its performance to supervised models.
  - Optimize the model for run time inferencing.
- **August 21st - 28th**
  - Final code quality checks
  - Dev documentation
  - Prepare for final evaluation.

# Why me

I like working on challenging Data Science problems and try to implement the best possible model for the task. I believe that my skills and my previous work-experience along with my enthusiam for this domain combined will make a strong candidate for this program. I am looking forward to engaging with the mentors and the community and delivering high-quality work.