Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Organization: Machine Learning for Science (ML4SCI) Umbrella

About me:

- **Name:** Mehdi Mabrouki
- **University:** National Institute of Applied Sciences and Technology
- **Email:** mabrouki.mehdi25@gmail.com
- **LinkedIn:** https://www.linkedin.com/in/mehdimabrouki/
- **Kaggle profile** (Kaggle Competitions Expert Top 1% Worldwide)**:** https://www.kaggle.com/mehdimabrouki
- **GitHub:** https://github.com/CallmeMehdi
- **Portfolio:** https://portfolio-callmemehdi.vercel.app/
- **Time Zone:** Central European Standard Time (GMT +1)

Qualifications for the project and Open-Source contributions:

- I am software engineering student that practises Artificial Intelligence often. Although, I do not have a physics background, it is my passion, and I've always

wanted to use my computer skills to solve problems related to this field. I am pretty experienced in Computer Vision. For instance, I have worked before on real-time detection using CCTV cameras. This demonstrates my experience in the Computer Vision field.

- I have read both the “Decoding Dark Matter Substructure without Supervision” paper and the “Decoding Dark Matter Substructure without Supervision” paper and, although I don’t have a physics background, I have some ideas on how to approach the problem at hand and some things we can apply in these cases.

- I have been a Major League Hacking Fellow for 3 months which allowed me to contribute to the popular machine learning library BentoML, I had the chance to work under the educational mentorship of professional software engineers from Google and Dev.to and the project maintainers, I worked on monitoring the YataiService and adding Prometheus endpoints for different functionalities.

- I am a TensorFlow Certified Developer, and I have experience setting pipelines for real-time object detection systems, so, I’m familiar with the tools we will be using in this project.

- I am a Kaggle Competitions Expert (Top 1% Worldwide) so I’m familiar with fine-tuning models and applying different ensemble techniques to achieve state-of-the-art results and improve the model’s performance, so I will be comfortable with the tasks in this project.

Project:

Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Abstract:

Using deep learning methods to implement a Computer Vision model that has the potential to accurately identify different properties in images containing substructures. The focus of this project is to apply regression techniques to estimate dark matter properties. This can be achieved by using different types and architectures of regression models, or a combination of them. The use of Convolutional Neural Networks with a time-series-oriented approach preceded with a feature representational embedding state, can be explicative of the population-level quantities and properties of dark matter particle candidates.
Another good aspect of this project is that we can use the output of the intermediate, feature-representational layers of the first model, pass them to a non-supervised learning clustering model, to create a representation of **additional hidden** (not tackled in the first model) **properties of dark matter.**

Proposal Timeline:

- **Before June 7th:** Setup**:** Getting more information on what types of models I'm going to be training, doing research on the resources needed, and creating a workflow, to optimize my time to train and test multiple models at the same time. Creating training, testing, fine-tuning, serving, and on-device training pipelines for the model, so as to make the implementation easier in further phases.

- **June 7th - June 13th:** Introduction phase where I will be learning more details about the project from the ML4SCI team. I will also be focused on doing some research about the tasks and trying to prepare pipelines and workflow depending on the needs.
- **June 14th - June 27h:** This 1st phase will consist of an introduction of training models (depending on the need) and testing them. This will include a discussion with the project owners on their vision of the models, training, evaluation, fine-tuning, and trying different approaches. For example, including Recurrent Neural Networks, transforming the problem into a time series to add accuracy for the regression output. This will be the first evaluation for me and a way to cope with the team.
- **June 28th - July 4th:** In this phase, I will be serving the model and trying to add it to the DeepLense functionality. This will help get a better overview of the project and make it easier to develop detection applications at later stages. This phase consists of testing different tools for serving the model and testing it on different real use-cases, so as to find the most optimized solution.
- **July 5th - July 11th:** First personal evaluation of what I worked on so far. Testing out the performance of the pipeline on real-time use cases. Measuring out metrics like the mean Average Precision, mean Average Recall, Performance, Latency Time will help us get a better look at the evaluation of the current model. This will enable us to do some tweaks to it and improve its accuracy.
- **July 12th - July 16th:** Getting the mentors' evaluation of my performance in the first phase, improving the pipelines I used for training, trying to add more fine-tuning for improving the accuracy, and getting better results.

- **July 17th - July 20th:** Implementing some ideas I have prepared about improving the architecture of the CNN while keeping it simple, changing the loss functions, and applying hyper-parameter tuning to reach maximum accuracy for the proposed architecture.
- **July 21st - August 4th:** At this phase, I will focus on two things. First of all, tuning the model so that it can have a really good precision and recall. And after that I will be trying to search for ways to include more dark matter properties. This is because in deep learning, the model can sometimes detect hidden features in data, and by taking the output of the intermediate, feature-representational layers of the model and passing them to a clustering model, for example, we can produce clusters of new dark matter properties, that were not noticed in the first approach.
- **August 5th - August 12th:** Testing and Documenting the model created by using examples and code snippets and other documentation elements required.
- **August 13th - August 15th:** Personal evaluation of my work, and execution of some tweaks in the model.
- **August 16th - August 23rd:** Submission of the final work product and evaluation from the mentors.
- **August 23rd - August 30th:** Mentors submitting final evaluation for the work.

Goals:

My set of goals for this project is to reach state-of-the-art results in this task while keeping an optimized response time for the application. This will require paying attention to both the performance and the accuracy simultaneously with a focus on producing high-quality documentation for the project.

I intend, also, to collaborate with my mentors by using my creativity and innovation to contribute to the project ideas and suggest different architectures/approaches.

Post GSoC:

_____Contributing to open-source GitHub projects in the Major League Hacking fellowship has brought me closer to the MLH community, so, I intend to do the same with both, the GSoC community and the ML4SCI team. As a software engineering student and an Artificial Intelligence enthusiast, I'm really interested to be integrated with these communities and get to know people from different backgrounds and cultures.

Also, my contribution to open-source projects doesn't end with the program, so, I will try to use the experience and skills that I improved during Google Summer of Code to contribute more to the community.

Other commitments:

_____I finish my school year in June, so I will be completely free for the Google Summer of Code Program. That's why I will be working **full-time** on this project. This being said, I can easily give around 40 hours/week for the project.