Organization: **ML4SCI: DeepLense**

Project Title: Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Introduction

About Me

Full Name: Diptarko Choudhury

Time Zone: IST (UTC+5:30)

Email: diptarko.choudhury@niser.ac.in

GitHub ID: dc2506

Matrix Handle: dc2506

Whatsapp/Phone: 9593169517

Education

University: National Institute of Science Education and Research (NISER), Bhubaneswar

Program: 5-Year Integrated Masters in Physics, with a Minor in Computer Science

Current Year: Second year (4th semester ongoing)

PERSONAL BACKGROUND

I am an undergraduate 2nd Year student at the National Institute of Science Education and Research (NISER), pursuing my Integrated Masters in Physics. Along with my Integrated Masters, I am also doing a Minor in Computer Science. My main interest lies in the field of computational physics and Machine Learning. I am proficient in C++, Python and Java. I have some experience using Machine Learning Frameworks such as Tensorflow and Pytorch. I enjoy doing Machine Learning and Deep Learning in my free time. I started it as a hobby, but it grew to become a passion with time. I have participated in several Kaggle and private competitions and even secured the First Place in ML4SCI 2021 Higgs Challenge. I am also quite interested in particle and high energy physics since I often like to read about them in my leisure time.

I am also part of many extracurricular clubs in my college and love to take part in events such as debate and extempore. I am also an active member of RTC (Robotech Tech Club) of NISER, where I mentor Computer Vision and Machine Learning projects. I am also working on a pet project of the RTC Club, AMAR(an autonomous rover), where I am currently dealing with Object Detection and Image Segmentation.

Being highly interested in physics and Machine Learning, I came across the ML4SCI 2021 Hackathon. I participated in it, and my interest grew even further to work on a project related to this specific field.

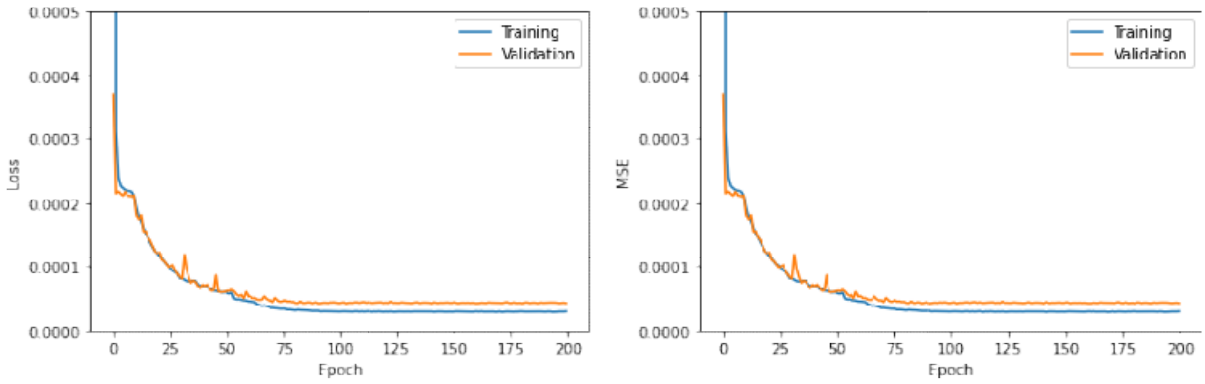
Why this project?

ML4SCI had a great variety in terms of projects to choose from. However, DeepLense was one of those few projects that captured my attention. The main motive behind machine learning or Artificial Intelligence is to replicate the human learning process in computers. Machine Learning tries to make computers have cogitative abilities similar to humans. There are very few examples in machine learning where the machine's ability to do a particular task is greater than that of a general human. The same is true for this project. A simple look at the problem images(simulated) gives us little to no idea about the nature of the sub-structure.

Nevertheless, an algorithm like a CNN can easily beat a human in this task, classify a strong gravitational lensing image, and give us an idea about its substructure. It also quite accurately predicts the mass of the dark matter halo, which we humans can never do by looking at the image. This is what makes me interested in this project, as it can act as a tool for understanding the universe, which otherwise might be quite difficult. Moreover, there is still much experimentation and work to do for this project, and there is quite a lot of room left for improvements.

Insights Gained from the Evaluation Tests

The organization wanted me to complete an evaluation test. I have completed the two evaluation tests required for this project and have submitted them as per the instruction. Since the evaluation tests were highly correlated with my current project/proposal, I gained quite some insight. There were two tasks given, classification and regression. EfficientNet architecture was used for both of them. Since this project mainly focuses on the Regression Task, I will stick to it.



From the above plot, it is clear that the model has quite a good Generalization and needs not be stopped early, showing signs that we can scale up the model without much overfitting. Moreover, when we tried to use the Cyclic Learning Rate policy, we found that the network got trapped in bad local minima. Hence we had to use Reduce Learning Rate on Plateau. Better experimentation and proper research might help us to find a much better learning rate policy to get the network to converge better and faster. Since the images are simulated, we will need a much better and well-researched image pre-processing technique to get the best results.

The Proposal

Abstract:

We are trying to implement a method to use Deep Learning to infer information from the Strong Gravitational Lensing Images. The main motive of this project is to find Dark Matter properties using deep regression techniques and use them to extend the already present DeepLense pipeline.

Description:

The project will try to extend the functionality of the DeepLense pipeline by adding necessary models and regressional techniques to infer Dark Matter properties from Strong Gravitational Lensing Images. Earlier, there were efforts made to make classification models based on the sub-halo structures of these images. These endeavours were highly successful and gave very high AUC and accuracy values. A somewhat similar approach will be taken but for finding required dark matter properties. Conventional Networks and CNNs are generally trained and designed from an image classification point of view; hence necessary modifications are necessary to make them fit for regression tasks. This project will mainly focus on finding and implementing the best architecture for the given task at hand.

Deliverables of my project-

- Robust and fast input pipeline
- Efficient and state of the art Image Augmentation and pre-processing Techniques
- Using state of the art training, optimization and regularisation techniques to achieve the highest validation score.
- Development of a robust and user-friendly output pipeline and integrate it with the pre-existing DeepLense pipeline.

Timeline

CURRENT PROGRESS

I have been working on the Evaluation test of the projects and finished working on them. I gained quite a bit of insight after building the basic architecture and enough experience to build a much more complex and tailor-made architecture to suit my problem the best.

COMMUNITY BONDING

In consultation with the mentors, I will set up a proper channel and weekly schedule for communications and discussions. I will use this time to further my understanding of the problem statement and clarify doubts (if any) with my mentors. I try to mainly understand the problem statement during this time and get a shallow idea of the quantities or properties with which I will be dealing. I will also shortlist some paper after consultation with my mentor to read and understand the details of Gravitational Lensing and Dark Matter.

Week 1-2 (2 Weeks)

I will devote the first two weeks to reviewing the literature. I will start to read about various new architectures used in the field of Computer Vision, which can be used to build my model. I will also read about gravitational lensing and try to get a basic understanding of what I am working with. Also, I will try to get myself exposed to various state of the art techniques(like stochastic depths, Image Augmentation techniques etc.) and their application in the field of computer vision, which might help to increase my model's performance. I will also read the research papers given on the website in much more depth to understand what I am working with.

Week 3

I will start building the basic input pipeline for our neural network, which will be robust enough to accommodate any changes in future without getting bottlenecked.

Week 4

This week I will try to understand the nature of the input data and try to implement ideas like outlier detection. I will try to find which type of image pre-processing works the best for the data. I will use our previous (Evaluation Test) model to benchmark the image processing techniques.
I will even try to implement some new Image processing and Augmentation techniques, which we researched in the first two weeks.

Week 5-6 (2 Weeks)

These two weeks will be entirely spent on experimenting and benchmarking various state of the art neural network architectures. I will be experimenting with various regularisation techniques (like Stochastic depth, block drop, dropout, normalization etc.) to fight overfitting so that I can efficiently scale the model. By the end of the sixth week, I will finalize the final architecture of our model.

Week 7

I will formally implement the final model this week. I will spend this week debugging the model with the help of Tensorboard and other related software. I will also look for any bottlenecks using Tensorflow Profiler. I will modify the input pipeline to prevent any form of bottlenecks. I will make sure that the model is ready to be trained from the next week onwards and free from any bugs or bottlenecks.

Week 8 – 9 (2 Weeks)

I will start to train the model. Since training and fine-tuning is a long and lengthy process, two weeks are allocated for it. I will also decide on the learning rate policy in these two weeks. Along with the learning rate policy, I will also be

experimenting with various new optimizations techniques like AdamW(optimizer), one cycle training, and LR range test to get super convergence.

Week 10-11(2 Weeks)

I will continue to train and fine-tune the model. In these last two weeks, I will extensively focus on hyper-parameter tuning. I will deploy multiple methods such as Bayesian, Evolutionary and Pruning to effectively tune the hyper-parameters and obtain improved scores.

Week 12

I will merge all our code this week and add comments and documentation (any remaining) to various parts of our code for it to be easily understandable to the end-user. I will also make an output pipeline for the end-user to use the network and merge it with the main code.

FUTURE DEVELOPMENT

Although our aim in this project will be to achieve the highest result using CNNs still, there will be much room for improvement remains. To scale up our model further, I would like to use Transformer and Graph neural network architectures in future to get better validation metrics. Also, using larger networks calls for higher computational resources; hence to scale my network, I would like to make use of distributed training strategies to reduce my training time. I would also want to make the model much more deployable in future so that it can evaluate using minimum computational power.