Proposal for GSoC 2021

Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Personal Details

- Name: Utcarsh Agrawal
- Email: utcarsh_dd2009@yahoo.in
- Phone: +91 8130049340
- Country: India
- Timezone: Asia/Kolkata (UTC+05:30)
- Links: Github LinkedIn
- Language: English

I am a 3rd-year student pursuing B.Tech. in Electrical and Electronics Engineering with great delight in coding. If I am selected, I will commit my full time to this project and can spend at least 40 hours per week or more if the work requires.

Technical Knowledge

I have good experience in Python, Machine Learning, Deep Learning, PyTorch, and also have the basic knowledge of C, HTML, and CSS.

I have done many courses on machine learning and deep learning offered by IBM and Deeplearning.ai respectively on Coursera.

Also, I have worked on many different machine learning-based projects and some of them are listed below: -

- I have created a hand gesture detection-based project. It is a Deep Learning-Based Project which is used to reduce contact with surfaces. We have seen lockers, electronic locks, and elevators which has a keypad on them but in this pandemic situation when no one wants to make contact with the surface or to minimize the contact I have designed such a system that the person has to just show hand gestures and the camera will detect the numbers and accordingly it will work.

- I have also worked on a water quality prediction project. In this project, I have made a water quality prediction method using pyspark and deep learning where the person has to enter some water quality parameters and the quality of the water will be known.

- I have worked on many NLP tasks such as disaster tweet prediction in which the tweet of a person is given and the system will tell whether the tweet is about some disaster or not. Also, I have worked on sentiment analysis prediction wherein we are given a tweet of a person and the system has to predict whether it is a positive or negative tweet.

Apart from that, I have also taken part in many Kaggle competitions where I have applied my machine learning and deep learning skills.

Project

Project Title

Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Project Abstract

Deep learning algorithms have shown excellent results especially in the field of computer vision and have solved many problems with an increased level of difficulty. In this project, using deep learning we have to build the optimal model for deep regression for estimating dark matter properties, including population-level quantities and properties of dark matter particle candidates (i.e., axion string density).

Technical Details

- Understanding the nature of Dark Matter (DM) is one of the most fundamental problems of particle physics and cosmology. Although much researches have been put forth over the last years, but still the true identity of dark matter remains elusive. To solve this issue, the role of machine learning algorithms comes into action. Machine learning (or more specifically deep learning) can help physicists gain information much more smoothly and precisely.
- In this project, we need to estimate dark matter properties, including population-level quantities and properties of dark matter particle candidates and for that first, we need to have a large dataset. But there is a very small dataset available of real galaxy-galaxy lensing images. So, for that PyAutoLens can be used for simulating strong lensing images. This way we can have a vast dataset that will be useful for us to train the model.

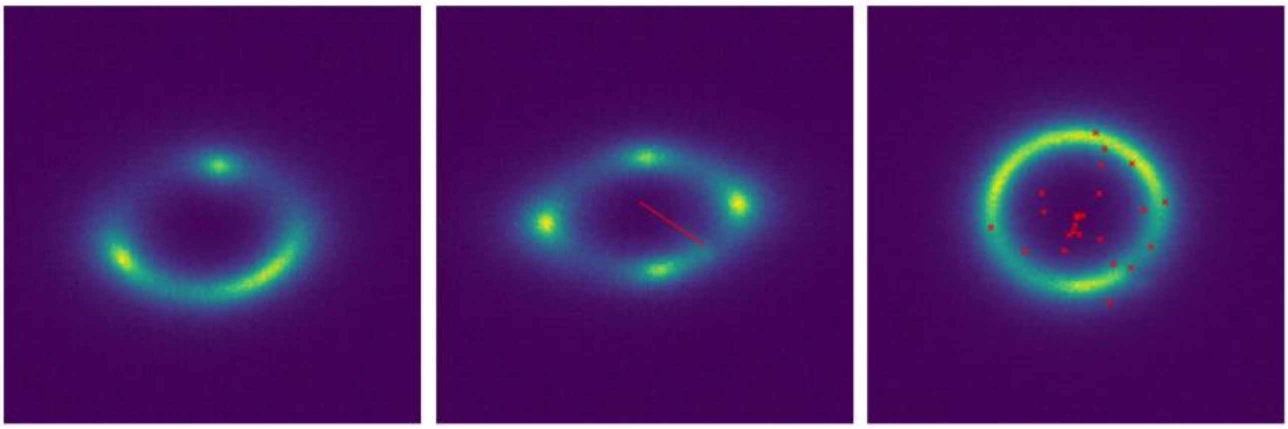
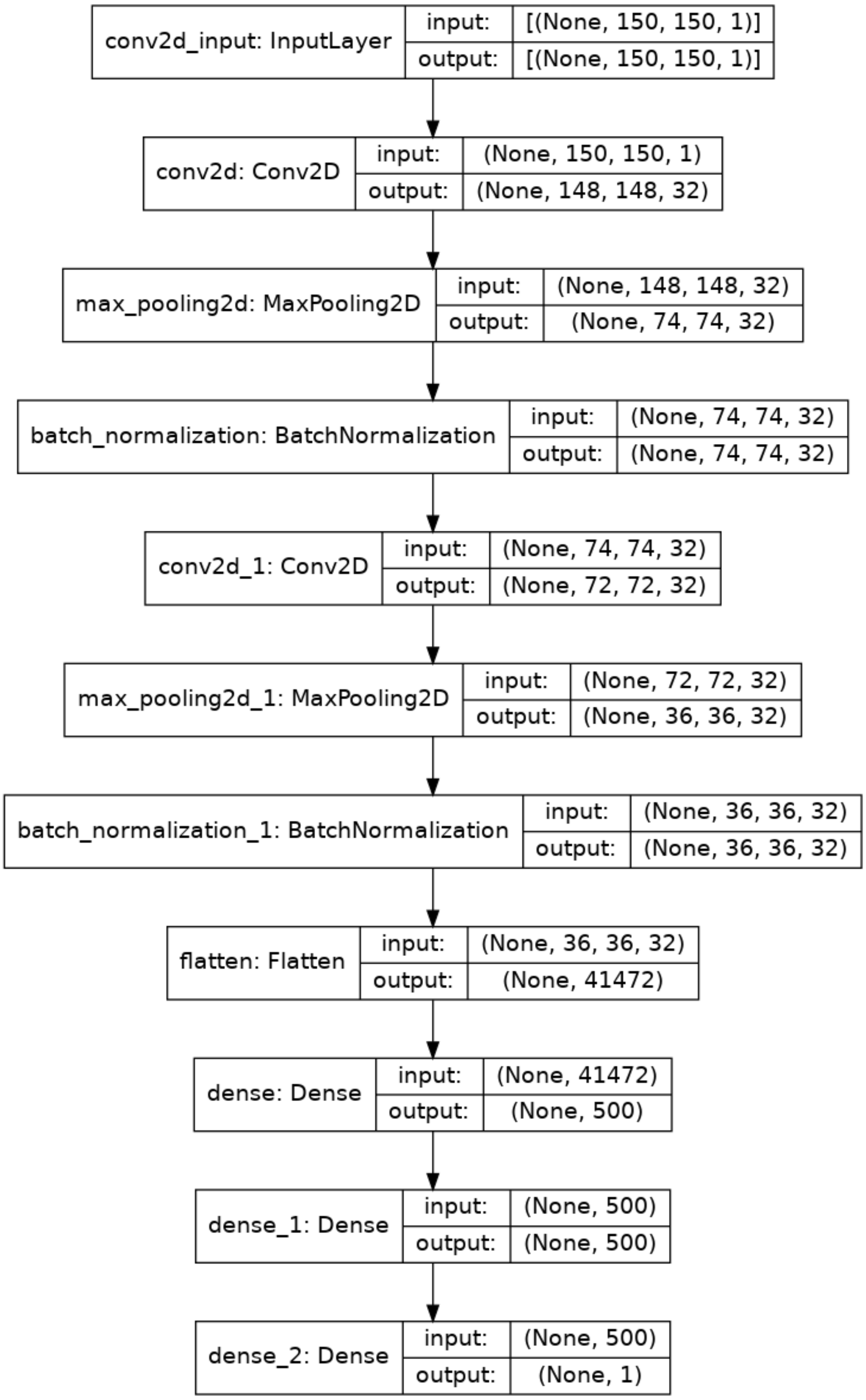
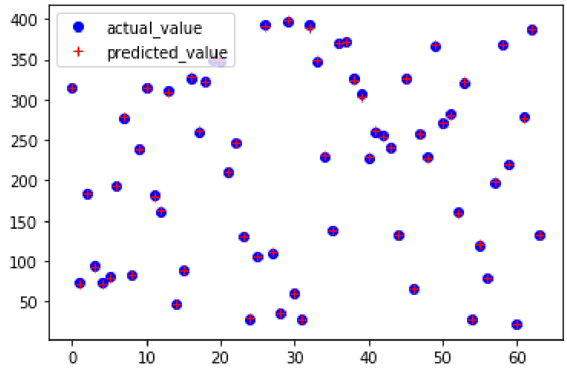


Figure - Simulated lensing images with different substructure. Examples with none (left), vortex (middle), and subhalo (right) with labeled locations of substructure. Image credits to the authors of the paper 'Decoding Dark Matter Substructure without Supervision '.

- Next, a model has to be built from scratch. As we are working with image data, we will work with a Convolution Neural Network. We will first build our own model, train it and then tune it for better results. For eg., referring to the evaluation task where we were supposed to learn the mapping between lensing images and the lensing dark matter halo mass, the architecture of my model which I built is shown below: -



It is a basic convolution model with an output layer as a dense layer with 1 output class as it was a deep regression task. I trained this model for 30 epochs and while training the mean squared error loss decreased from 2580 to 3.43. After that, I used the trained model and made predictions on the test set, and it predicted almost all the values perfectly. Similarly, we can have a model like this with some modifications depending upon the datasets.



Evaluation Task Model Predictions on Test Set

- Subsequently, we can take some custom pre-built model like ResNet, VGG16, AlexNet, etc. and will train it and will then compare with our own model.
- We will finally tune the model accordingly and will then select the best model. As it is a deep regression model, our final layer will be a linear layer with 1 output class which will return the predicted value.
- During training, we will make use of the mean squared error (MSE),

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values

as our loss function. The models are to be implemented using the PyTorch package. We will be using mean squared error (MSE) as our evaluation metric also.

Timeline

Before Community Bonding Period

- Introduced myself and expressed my desire to work with the organization through e-mail.
- Received evaluation tasks and worked on them and submitted them on time.
- I also got familiarized with the PyAutoLens by going through the documentation and the tutorials.
- Also joined the organization's Gitter page.

Community Bonding Period (May 17, 2021 - June 7, 2021)

Introduce myself and get familiar with other members and the mentors. Discuss with the team about the minute details of the whole project like what kind of dataset will be, model architecture recommendations, some doubts regarding PyAutoLens, etc.

Coding Period (June 7, 2021 - August 16, 2021)

Week 1 (June 7 – June 13)

- Analyze the dataset.
- Apply some feature engineering techniques to the dataset.

Week 2-3 (June 14 – June 27)

- Build a model from scratch.
- Train the model.
- Apply the trained model on the test set and check how the model is performing.

Week 4-5 (June 28 – July 11)

- Tune the hyperparameters of the model for better results.

*** First evaluation period (June 11 - 15)**

- Submit the trained model and its performance result.

Week 6-7 (July 12 – July 25)

- Use some famous model architectures like ResNet, VGG16, etc.
- Compare the results with our own model.
- Tune the models accordingly to get the best result.

Week 8-9 (July 26 – August 8)

- Take feedback from the community and make changes accordingly.
- Create documents, blogs, or videos so that others can smoothly understand the project.

Week 10 (August 9 – August 16)

- Buffer time for the unexpected delay.

Important: - Due to covid-19, our semesters have started late and our exams have not yet been scheduled. I will inform the mentors as soon as the information is released. So, my timeline might get disturbed if the exams occur during the coding period. But that will not stop me from completing my project as I have no other commitments other than GSoC and I will definitely manage my time such that I can work on the project and complete it on time.

After GSoC

I would like to keep contributing to ML4SCI after GSoC and will be available to solve issues. I will help people and participate in the discussions and keep contributing and will do whatever I can to help the organization grow.

References

Some of the content for this document has been taken from the paper '[Decoding Dark Matter Substructure without Supervision]'.