GSoC'21 Project Proposal

Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Organization: *ML4SCI*

Mentors: *Micheal Toomey, Stephon Alexander, Sergei Gleyzer, Brandon Ames, Tyler Trupke*

by

Sai Vamsi Alisetti

Contents:

Basic Information:

Name	Alisetti Sai Vamsi
University	Indian Institute of Technology, Palakkad
Field of Study (Major)	Computer Science & Engineering (Batch of 2018)
Email	saivamsi.ds123@gmail.com 111801002@smail.iitpkd.ac.in
GitHub	Vamsi995
LinkedIn	https://www.linkedin.com/in/sai-vamsi-4892a4197/
Phone Number	+91-7995939741
Country/ Region /Timezone	Telangana, India, Indian Standard Time (GMT+5:30)
Resume	Drive Link

About me:

I am a third-year B.Tech undergraduate student pursuing my Computer Science degree at IIT Palakkad. As a computer science student, I've always been interested in problem-solving be it in a social aspect or analytical aspect. And nothing brings me more joy than solving something that is on the cutting edge of research and on the verge of being something new. My philosophy is that we as humans became a civilization because of working together and hence we survived until this day. It might sound bizarre but everyone on this planet is contributing to our survival in one way or the other. In my case, my contribution suffuses in the form of quality code. And with this, I am drawn to the open-source world of programming, where people from different countries come together to collaborate.

I have an immense amount of passion for programming and astronomy. Although most of my astronomy knowledge stems from youtube videos, I have a good amount of understanding of the technical jargon. My love for programming outweighs astronomy, otherwise, I would've aimed to become an astrophysicist. Apart from reading books and programming, I do enjoy playing football, chess, video games, and solving puzzles, and also love to travel.

Project Details:

Overview:

Galaxies in our universe seem to rotate at very high speeds that the gravitational force generated by the matter inside the galaxy is not enough to hold the galaxy together. In theory, they should rip themselves apart because of the immense centrifugal force that counteracts gravity. But the fact that these galaxies are intact gives us a hint that there is some extra mass that is aiding the gravitational force but is invisible. The visible matter in the universe only forms a very small fraction of all the matter in the universe. There is about six times more matter than what we can see, but except it is hidden/invisible. This invisible matter is called dark matter. Dark matter does not interact with electromagnetic fields, i.e it does not absorb, reflect, or emit electromagnetic radiation. Hence it is difficult to detect but its only interaction with the visible matter is through gravity. Strong galaxy-galaxy lensing provides strong evidence of dark matter because most of the light is lensed through the space distorted due to the dark matter halos surrounding the lensing galaxy. The way in which these dark matter halos lens the source galaxies provides us means to study the properties of dark matter and possible ways of decoding the structure inside these dark matter halos.

The volume and complexity of data in astronomy are constantly increasing, making conventional algorithms for data analysis very difficult. The rapid advancement of machine learning and deep learning techniques allows us to solve these problems in a variety of ways.

Regression analysis in machine learning is a tool that is used for predicting a continuous output variable (y) based on the values of one or more dependent variables (x). There are several forms of regression possible out of which I am using polynomial regression, and hence the use of neural networks. These are function approximators in a higher dimension. Hence in the case of a dense neural network (ANN), the number of weight parameters are nothing but the dependent variables that will be adjusted in each backpropagation iteration to predict the correct output.

The focus of this project is on using deep regression techniques for estimating dark matter properties, including population-level quantities (i.e learning the total mass of a halo) and properties of dark matter particle candidates (i.e. axion string density). The evaluation task that is done on the deep regression analysis which included learning the dark matter halo mass from the simulated images, has given me a deep insight as to what the project needs.

About Evaluation Task:

GitHub:https://github.com/Vamsi995/Deep-Regression-Techniques-for-Decoding-Dark-Matter-with-Strong-Gravitational-Lensing

Theoretical research:

The concepts of gravitational lensing and dark matter are relatively new to me and I loved exploring every aspect of them. The evaluation task, especially the first question, made me dig the entire rabbit hole all the way through. To be more specific, after understanding the library and the basics of gravitational lensing, my next challenge was to understand the concept of dark matter. I learned about the possible candidates of dark matter i.e axion, WIMP, MACHOS, etc.., and how dark matter interacts with visible matter only through gravity and hence the use of lensing to observe its substructure.

These two papers Decoding Dark Matter Substructure without Supervision and Deep Learning the Morphology of Dark Matter Substructure helped me understand the dark matter substructure in much more detail. The first paper explains about dark matter substructure and expected signatures from lensing images, and then it talks about the generation of the dataset using the PyAutoLens simulations (Table 1), and then presents two methods (supervised and unsupervised) which classify the data into dark matter halos having no substructure, subhalos as substructure, vortices as substructure (**this is where I got the hint for the first question as in this paper it is mentioned that the vortex substructure is of a superfluid form and on large scales, the density profile of this vortex substructure can be approximated to be a linear mass density**). The supervised method includes using a ResNet18 for image classification (**this is where I drew inspiration to use ResNet18 as a feature extractor**) and the unsupervised method includes the usage of convolutional autoencoders. The novelty of the first paper is the unsupervised classification method which is compared with the supervised method for classification.

The second paper talks about the theory of superfluidity in dark matter and also performs the same supervised analysis as the first paper on the classification of the substructure inside dark matter halos of lensing galaxies from lensing simulations. This paper gives further explanation on the superfluid dark matter as vortices and also a much more detailed description of the parameters used (TABLE 1) in creating the lensing simulations and also has detailed simulation images of the vortex substructure.

The following talk "Decoding Dark Matter Substructure without Supervision" Michael Toomey (Brown) - CFPU SMLI, explains the first paper in brief and helped me understand the bigger picture since all the pieces came together after watching this talk.

From the evaluation test, I have understood how the dataset is structured and have already performed necessary data preprocessing steps to clean the data including data augmentations for improving the performance. The second question, which included training a deep learning model, gave me enough idea on how to transform the data to feed it into any deep learning network. Overall the evaluation test enhanced my skills to perform quality research and gave me an idea as to what the project needs and what is to be done.

Goals / Deliverables:

1. Implementation of popular state-of-the-art deep CNN architectures for feature extraction and performing regression analysis.

2. Creating a custom architecture from the existing CNN architectures with feature fusion techniques.

3. Research on the optimal model for feature extraction and optimal hyperparameter tuning.

4. Implementation of the optimal model and comparison of the regression analysis with the previously implemented models

5. Writing Blog Posts and articles for creating more awareness.

6. Aiming for a paper in this direction by making a custom deep learning architecture.

Approach/Methods:

Note: This is a tentative outline of the final outcome, further refinements will be made during the project.

This project involves a significant research aspect. Therefore I also included some research time in between so that I can discuss with the mentors in detail. The amount of research studies on deep regression on vision tasks is quite limited and hence is a niche area at the moment. The papers that have been published on this topic are either very specific to the task or the dataset or they are very old. One of the only reliable and generalized sources that I could find is this paper A Comprehensive Analysis of Deep Regression which provides a detailed analysis of deep regression on only two selected models ResNet-50 and VGG-16.

Taking inspiration from my evaluation task, I will be dividing the project into two parts. The first part being feature extraction using a Deep CNN network. And the second part is designing the dense ANN which will follow the CNN network. Most of the papers on the innovations in the CNN architectures are based on the image classification results. Hence some of the CNN architectures that I have explored are from image classification papers. It is quite an easy task to convert an image classification model to a regression model, wherein we just have to change the final dense layer to have a single output, and the loss function should be changed to an MSE loss or an MAE loss. But the only catch in doing this is tuning the hyperparameters accordingly.

1. **Implementation of state of the art Deep CNN architectures for feature extraction and performing regression analysis**

   Deep CNN architectures have seen a rise especially in the image classification task of computer vision, where it is used as an effective feature extraction tool. In a similar sense, I plan to use these Deep CNN architectures to extract features from the images and pipe them to a dense network with a single output, for performing regression. There are several innovations in CNN architectures out of which I used ResNet18 in the evaluation test for feature extraction because it was mentioned in the paper, but now I plan to evaluate the dataset on several popular state-of-the-art CNN feature extractors and compare their results.

   I want to evaluate the dataset on the following models:

   1. **ResNet**-34, **ResNet**-50, **ResNet**-110, **ResNet**-164, **ResNeXt** (https://arxiv.org/abs/1512.03385), (https://arxiv.org/abs/1611.05431)

   2. **Inception** v3, **Inception** v4, **Inception**-**ResNet**.

(https://arxiv.org/abs/1512.00567), (https://arxiv.org/abs/1602.07261)

3. **DenseNet**-121**, DenseNet**-161**, DenseNet**-201
   (https://arxiv.org/abs/1608.06993)

4. **VGG**-16**, VGG**-19
   (https://arxiv.org/abs/1409.1556)

In the evaluation test, I have trained the model from the scratch. But now I plan to make two variants of each, i.e

1. Fine Tuning on Pre-trained Model (Imagenet dataset)
2. Training from scratch

I will be following fine-tuning methods which are laid out in detail from the PyTorch library, https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html.

In the fine-tuned variant, I also plan to conduct experiments that include extracting features from the earlier layers of the CNN's for each network and compare results. Furthermore, I will be using SAM (Sharpness-Aware Minimization for Efficiently Improving Generalization) as an optimizer in place of Adam and test the performance too.

2. **Creating a custom architecture from the existing CNN architectures with feature fusion techniques.**

Feature fusion is the integration of multiple different feature information to obtain more prominent feature information. There are different feature fusion techniques like concatenation, summation, weighting, etc ..,. The outline of this approach is to use pre-trained models of multiple CNNs and extract features from the last convolutional layer or the average pooling layer of these models. The feature fusion technique combines the output feature maps from these models and combines them over an aggregate function and feeds this combined feature representation as input to the dense neural regression network. This is relatively new to me, and I would very much like to try this out. The implementation of this would be based on the paper Deep Learning for Feature Extraction in Remote

Sensing: A Case-Study of Aerial Scene Classification . And for more theory on feature fusion I plan to read the following papers:

1. Using Feature Fusion and Parameter Optimization of Dual-input Convolutional Neural Network for Face Gender Recognition

2. Multi-Layers Feature Fusion of Convolutional Neural Network for Scene Classification of Remote Sensing

3. Multilevel Weighted Feature Fusion Using Convolutional Neural Networks for EEG Motor Imagery Classification

There is another method for feature fusion i.e instead of taking features from different CNN architectures, we can use a single CNN and take features from different pooling layers in the CNN and perform feature fusion.

Note: I am not sure if this is the right way to go, but I am open to other ideas too. I want to discuss this approach with the mentors in more detail.

3. **Research on the optimal model for feature extraction and optimal hyperparameter tuning** (**Research**)

This is a research aspect that I want to discuss with the mentors briefly and get their opinions on some of the questions that I have. From the current research that I have done, I have found the state-of-the-art feature extractors, which outperform all the previous popular networks like ResNet, DenseNet, etc,...,.

These are the following architectures that I plan to train on:

1. **EfficientNet** (https://arxiv.org/pdf/1905.11946.pdf)

2. **Big Transfer (BiT)** (https://arxiv.org/pdf/1912.11370.pdf)

3. **Vision Transformer (ViT)** (https://arxiv.org/abs/2010.11929)

4. **NFNet** (https://arxiv.org/pdf/2102.06171v1.pdf)

All these models are on the top of the list in the image classification benchmark, and hence due to this, I plan to use these to extract features. Further research will be done as part of the project, wherein I plan to train it on more models which I explore.

4. **Implementation of the optimal model and comparison of the regression analysis with the previously implemented models**

I have already done some research as to where to find these model architectures in PyTorch. I will be listing down some of the repositories which will be helpful while implementing these models.

1. **EfficientNet:**
    a. https://www.kaggle.com/ateplyuk/pytorch-efficientnet
    b. https://github.com/lukemelas/EfficientNet-PyTorch
    c. https://nni.readthedocs.io/en/latest/TrialExample/EfficientNet.html

2. **BiT:**
    a. https://github.com/google-research/big_transfer

3. **ViT:**
    a. https://github.com/lucidrains/vit-pytorch
    b. https://github.com/jeonsworld/ViT-pytorch
    c. https://towardsdatascience.com/implementing-visualttransformer-in-pytorch-184f9f16f632

4. **NFNet:**
    a. https://github.com/vballoli/nfnets-pytorch
    b. https://github.com/rwightman/pytorch-image-models
    c. https://github.com/benjs/nfnets_pytorch

Hyperparameter tuning is an important issue that needs to be addressed as it provides us with the optimal model and hence is a tedious task if one needs to adjust and verify manually. For this purpose, I plan to use the Ray Tune tool from PyTorch which uses the state of the art algorithms for finding the optimal parameters for our model.

5. **Writing Blog Posts and articles for creating more awareness**.

   During the final phase of the project, I would very much like to document my entire work and also write some blogs relating to this topic, to spread the word out about the advances of deep learning in astronomy. And also I would like to write blogs explaining the science behind lensing and dark matter to the best of my knowledge.

6. **Aiming for a paper in this direction by making a custom deep learning architecture.**

   This is my main goal for this project. I am willing to spend more time even after my project ends with GSoC to develop my own architecture for this problem. As of now, I don't have a solid conclusive idea, but I am fairly confident that after performing extensive research on the literature, I will be able to devise the architecture that I will be developing. I am ready to give a commitment towards this and give my time even after this project.

Future Work:

I will be more than glad to work on this project throughout this year, to complete my main goal, i.e to make my own architecture for this problem. I would want the mentor's help and guidance in this aspect and would love to stay in touch with the mentors even after the project is done and hopefully work on future projects too.

Timeline:

Dates	Task
Pre - GSoC Period	Research on Feature extractors
Community Bonding Period	

May 17 - June 7	Getting to know the community and the mentors. Discussing the project, with mentors and analyzing the problem areas to set up the final project goals and approach. Starting the work on my first goal - analyzing and preparing the data, setting up a proper workflow.
Phase 1 (June 7 - July 16)	
Week 1 - 2 **June 7 - June 21**	Implementation of state-of-the-art Deep CNN architectures for feature extraction and performing regression analysis. Compare and analyze the regression results from each model.
Week 3 **June 21 - June 28**	Research on feature fusion techniques. Reading the mentioned papers, as well as exploring better techniques.
Week 4 - 5 **June 29 - July 11**	Creating a custom architecture from the existing CNN architectures with feature fusion techniques. Comparing the results from different feature fusion methods.
Phase 1 Evaluation (July 12 - July 16)	

Week 6 **July 17 - July 23**	Research on state of the art Deep CNN feature extractors. Writing blog posts, if I get some free time.
Week 7 - 8 **July 23 - August 4**	Implementing the optimal models in PyTorch, and comparing the results. Fine-tuning the models with optimal hyperparameters to get the best model output.
Week 9 **August 4 - August 15**	Working on my own deep learning architecture Implementing my own architecture and performing analysis with the previous models.
Final Evaluation (August 16 - 23)	

Why me?

I think I am a good technical fit for this project because I have a good amount of experience in PyTorch, and have in-depth knowledge in Deep Learning / Machine Learning. I have done Andrew Ng’s course on deep learning, and also took some courses related to it in my institute like Linear Algebra, Probability, Reinforcement Learning, Foundations of Data Science, and Machine Learning. Overall I have more than intermediate knowledge in Deep Learning. Coming to programming in general, I have been programming in python for more than 2 years now, and I am very proficient in the language. I have a good amount of research experience through my internships and my projects. One of the first projects that I was part of was the Pesticide Spraying Robot, which won the Gold Medal in the Inter IIT Tech Meet 2018 conducted in IIT

Bombay. I have also participated in international hackathons like HackMIT and ShellHacks which helped me in improving my time management strategies and leadership skills.

I was recently drawn to open source, wherein I first started by completing the HacktoberFest challenge as a participant. This gave me a huge confidence boost as well as made me more interested in the open-source world. From here I participated in the Winter Of Code program conducted by the DSC NSEC Club, wherein I was awarded as one of the top contributors. This made me a pro in the usage of Git and Github. With this, I made up my mind to commit to open source programming, and hence thought of applying to GSoC this year. This is my first time in GSoC, and I am very excited to see how this progresses.

Astronomy has always excited me, and the amount of time I spend on youtube videos in this aspect is immeasurable. I always wanted to apply my knowledge of programming in these fields but I always ended up assuming that it would be too complex. When I explored it in-depth, I felt very comfortable reading the papers that integrated deep learning with astronomy. I am very much willing to learn more through this project. I want to explore more in this field and I hope I get the chance to do so.

Availability:

I have a summer research internship at IISC, Bangalore on inverse reinforcement learning. Apart from this I might be doing a research project, which is an extended version of a current project that I am a part of. I will be devoting about 20-22 hours per week to the project during GSoC with ML4SCI. If in any case, I happen to fall behind my schedule, I am ready to commit more time to this project by working overtime.

Community Outreach:

Being part of such a flourishing community like ML4SCI is a blessing to me. And I plan to stick around for the long run. Even after the GSoC period is over I would like to still work on more projects or develop this one further. I have joined the necessary communication channels for the same. I will also be providing weekly progress updates, blog posts, articles through the communication channels. And since this is so special to me as my love for programming and astronomy converge at this point, I would like to introduce this concept in my college and possibly start a club in this direction if possible.