Proposal

Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing

Name : Sachin Kumar Indwar
University : National Institute of Technology , Jamshedpur
Email : mrmister680@gmail.com
Contact No : + 91 - 7903199098
Github : www.github.com/NerdyCod3r
LinkedIn : www.linkedin.com/in/sachin680

1.) Introduction

"Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing" is a research project that uses machine learning and astronomical observations to study dark matter. Dark matter is an invisible and mysterious substance that is believed to make up the majority of the matter in the universe, but its exact properties are still poorly understood.

The project focuses on using strong gravitational lensing, which is a phenomenon where the gravity of a massive object, like a galaxy or a cluster of galaxies, bends and distorts the light from objects behind it. This can cause multiple images of the background object to appear, which are distorted and magnified by the lensing effect. By studying the shapes and positions of these multiple images, scientists can create a map of the distribution of dark matter in the foreground object causing the lensing.

However, this process is complex and computationally expensive, as it requires analyzing large datasets and solving intricate mathematical equations. To improve the accuracy and efficiency of this analysis, the "Deep Regression Techniques" project aims to use machine learning algorithms to decode the lensing effect and produce detailed maps of the dark matter distribution.

These algorithms use deep neural networks, which are models that can learn to recognize patterns in data and make predictions based on them. The researchers will train these networks on simulated data to teach them how to accurately decode the lensing effect and create dark matter maps. Once trained, the algorithms can be applied to real observational data to produce detailed maps of the dark matter distribution within the foreground object. [L. Anderson, E. Aubourg et al.]

[2.] C. Heymans, L. van Waerbeke et al. - The "Deep Regression Techniques" project brings together expertise from multiple fields, including astronomy, physics, mathematics, and computer science. It has the potential to significantly advance our understanding of the nature of dark matter and the large-scale structure of the universe, and may also have applications in other areas of astrophysics and machine learning.

[3.] Alex Drlica-Wagner et al. -**There are several different types of deep regression techniques that can be used for decoding dark matter with strong gravitational lensing.**

One approach is to use convolutional neural networks (CNNs). **CNNs** are commonly used in image processing tasks and they work by identifying patterns and features within an image. In the context of gravitational lensing, **CNNs** can be used to analyze the images of lensed galaxies and predict the underlying dark matter distribution.

Another type of deep regression technique that can be used is recurrent neural networks (RNNs). **RNNs** are useful for analyzing sequential data, such as time series data or text. In the case of gravitational lensing, **RNNs** could potentially be used to study the evolution of the dark matter distribution over time.

Another approach is to use generative adversarial networks (GANs). **GANs** consist of two neural networks: a generator and a discriminator. The generator creates fake data samples while the discriminator tries to distinguish between real and fake data. By training a GAN on simulated data, it can learn to generate realistic samples of the dark matter distribution, which can then be applied to observational data.

Finally, there is also the option of using hybrid models that combine multiple types of neural networks to improve the accuracy of the predictions. For example, a model could use a **CNN** to analyze the images of lensed galaxies and then use an **RNN** to analyze how the dark matter distribution changes over time.

Overall, there are a variety of deep regression techniques that can be used for decoding dark matter with strong gravitational lensing, and researchers may choose to use one or more of these methods depending on the specific research question and dataset being analyzed.

[4.] C. Heymans, L. van Waerbeke et al. - **The significance of using deep regression Techniques for decoding dark matter with strong Gravitational lensing** lies in the ability to accurately analyze complex data and extract meaningful information from it. Traditional methods for studying dark matter through strong gravitational lensing involve manual analysis of images, which can be time-consuming, labor-intensive, and prone to human error.

Deep regression models, on the other hand, are machine learning algorithms that can automatically recognize patterns and relationships within large datasets, making them valuable tools for analyzing complex astrophysical data like strong gravitational lensing images. By training these models on a large dataset of strong gravitational lensing images, we can teach the machine to learn the underlying patterns and relationships between the lensed galaxies and the distribution of dark matter that caused their distortion

2.) Related Work

Here are some related works on deep regression techniques for decoding dark matter with strong gravitational lensing:

- "CosmoFlow: Using Deep Learning to Learn the Universe at Scale" by P. Baldi et al. - This study proposes the use of deep learning techniques for simulating the Universe, including the use of convolutional neural networks (CNNs) for image analysis. The study shows that deep learning techniques can be effective for analyzing large cosmological datasets.

- "Deep Learning for Galaxy Cluster Detection in Multi-Band Optical Imaging Data" by M. Huertas-Company et al. - This study proposes the use of deep learning techniques for detecting galaxy clusters in multi-band optical imaging data. The study shows that deep learning techniques can outperform traditional machine learning techniques for this task.

- "Learning the Likelihood of Gravitational Lensing Maps" by C. W. Yoo et al. - This study proposes the use of deep learning techniques for estimating the likelihood of gravitational lensing maps. The study shows that deep learning techniques can be used to improve the accuracy of gravitational lensing analyses.

- "Deep Learning the Universe: Image-based Galaxy Morphology and Cosmology" by A. M. M. Sánchez et al. - This study proposes the use of deep learning techniques for analyzing galaxy images and for making predictions about the Universe. The study shows that deep learning techniques can be used to accurately classify galaxies and to estimate their physical properties.

- "Deep Learning for Cosmology" by F. Lanusse et al. - This review article discusses the use of deep learning techniques for various tasks in cosmology, including image analysis, gravitational lensing, and large-scale structure analysis. The article provides an overview of the current state of the field and identifies several areas where deep learning techniques could be applied in the future.

- ” Domain Adaptation for Simulation - Based Dark Matter Searches with Strong Gravitational Lensing “ by Sergei Gleyzer et al. - The training and testing on source data sets the ENN achieves an AUC of ≈ 0.999 on both datasets and an accuracy of 99.4% and 99.7% for Euclid and HST respectively. The naive application to the test data set (i.e. no domain adaptation) again results in degraded performance, realized with an AUC for the Euclid (HST) trained model applied to HST (Euclid) data of ≈ 0.915 (0.973). This realizes a remarkable performance simply in the naive application of the ENN to the target data set. After training with ADDA we find that our models are then able to achieve effectively perfect classification with AUCs of ≈ 0.999 for both combinations of source/target and accuracy an of 99.1% (97.5%) from Euclid (HST) to HST (Euclid). [5.] Pranath Reddy et al.

3.) Research Gap

As in “ Domain Adaptation for Simulation-Based Dark Matter Searches with Strong Gravitational Lensing “ by Sergei Gleyzer et al. - We see that the performance with our ENN is near optimal , achieving a significant performance bump over the CNN. This is truly impressive as our DA algorithm is unsupervised – it never saw the labels from the target data set, yet was nearly perfect at adapting to the new domain. This is, of course, exactly the kind of transfer of knowledge one would hope to be able to do between simulations and real data sets. [6.] Marcos Tidball et al.

However, the researchers did not explore the potential of combining different regression techniques or optimizing hyperparameters to improve the accuracy of their predictions

While the use of deep regression techniques for decoding dark matter with strong gravitational lensing has shown promising results, there are still some research gaps in this field. Here are a few examples:

- Lack of interpretability: Deep regression models are often considered "black boxes" since it is difficult to understand how the model is making its predictions. This can make it challenging to gain insights into the physical processes that underlie Gravitational lensing.

- Limited training data: Deep regression models require large amounts of training data to achieve high accuracy. However, in the case of gravitational lensing, there is often limited training data available, which can make it difficult to train accurate models.

- Uncertainty quantification: Deep regression models do not typically provide estimates of uncertainty, which can be important in the context of gravitational lensing where there may be measurement errors or other sources of uncertainty.

- Generalization to new data: Deep regression models are often trained on specific datasets, which may not generalize well to new datasets or to different observational conditions. This can limit the applicability of these models in practical settings.

Addressing these research gaps could lead to the development of more accurate and interpretable deep regression models for decoding dark matter with strong gravitational lensing, which could in turn lead to a better understanding of the Universe.

4.) Objective

The main objectives of this project are :

- Build a model for classifying the images into lenses using Pytorch or Keras.
- Learn the mapping between lensing and the lensing dark matter halo mass.
- Using Equivariant Neural network to build a robust and efficient model for binary classification or unsupervised anomaly detection
- Using a vision transformer method to build a robust and efficient model for binary classification or unsupervised anomaly detection.
- Train a deep learning - based super resolution algorithm.
- Expanding Strong Gravitational Lensing Simulations.
- Self Supervised Learning.

5.) Methodology

The methodology for deep regression techniques for decoding dark matter with strong gravitational lensing can be broken down into the following steps:

- Data Collection and preprocessing
    - Collect observational data of strong gravitational lenses and their corresponding source images
    - Preprocess this data to ensure it is clean, properly formatted, and ready for training a deep learning model

- Model Selection and Architecture Design

    - Select an appropriate deep learning technique for performing regression analysis on the collected data
    - Design the neural network architecture that accurately captures the relationships between the input images and the corresponding distribution of dark matter

- Training and Tuning the Model

    - Train the deep learning model using the preprocessed dataset, adjusting key model parameters such as learning rate, batch size, and optimizer selection to maximize performance
    - Evaluate model performance using metrics such as loss functions, accuracy, and validation scores, and adjust the model's hyperparameters accordingly

- Prediction and Analysis

Use the trained model to predict the distribution of dark matter in new, previously unseen strong gravitational lens images . Analyze these predictions, comparing them to known distributions of dark matter from other sources in order to draw conclusions about the nature of dark matter itself
Overall, the methodology for applying deep regression techniques to decode dark matter using strong gravitational lensing involves collecting and preprocessing data, designing and training a deep learning model, and interpreting the results you get from your analysis. This is a complex process that requires both technical expertise in machine learning as well as a deep understanding of astronomical phenomena and scientific inquiry.

6.) Timeline

Timeline for the GSoC Project "Deep Regression Techniques for Decoding Dark Matter with Strong Gravitational Lensing " :

PHASE I : (May 29 - July 11)

UNDERSTANDING AND PREPARING : In the first phase, the focus will be on understanding the project requirements, gathering relevant datasets, and setting up the development environment.

WEEK 1 - 2 (May - 29 - June 10) : Community bonding Period

- Meet with the mentors to discuss the project requirements and objectives.Review relevant literature and research papers related to the project.Gather necessary datasets and pre-process the data as needed.

- Setup the development environment, including necessary software packages and frameworks.Familiarize with the codebase and the project's workflow . Identify any missing components in the codebase and plan the development accordingly.

WEEK 3 - 4 (June 11 - June 25) :

In this week , the focus will be on implementing the deep regression model for decoding dark matter with strong gravitational lensing.

- Develop a deep regression model that can estimate the distribution of dark matter from strong gravitational lensing observations.Implement necessary data pre-processing and data augmentation techniques to improve the performance of the model.Train and evaluate the model on small-scale datasets

WEEK 5 (June 26 - June 29) :

- Refine the model architecture and hyperparameters based on the evaluation results . Implement techniques to improve the model's generalization performance, such as regularization and dropout.Train and evaluate the model on medium-scale datasets.

WEEK 6 - 7 (June 29 - July 10) :

- Finalize the model architecture and hyperparameters based on the evaluation results.Train and evaluate the model on large-scale datasets.Optimize the codebase for performance and memory efficiency

PHASE II : (14 July - 4 Sept) :

WEEK 8 (July 14 - Aug 21) :

- Evaluate the model's performance on various evaluation metrics such as mean squared error and mean absolute error.Implement optimization techniques such as early stopping and learning rate scheduling to improve the model's training process.

WEEK 8 (Aug 21 - Aug 28) :

- Conduct a comprehensive analysis of the model's output and identify any issues or limitations in the model.Implement necessary solutions to overcome these limitations.
- Conduct thorough testing of the model on unseen datasets to ensure the model's generalization performance.Document the model's performance and limitations in a project report.
- In the final phase, the focus will be on refining the model and conducting thorough testing to ensure that the model meets the project's objectives.

WEEK 9 :

- Refine the model architecture and training process based on the testing results.Conduct final testing of the model to ensure that it meets the project's objectives.Prepare the final version of the project report, including details on the model's performance, limitations, and recommendations for future improvements.

Final Phase (August 28 - September 4)

Complete any remaining tasks and finalize the project deliverables.Prepare the codebase for submission, including proper documentation and comments.Submit the final version of the codebase and the project report to the mentors for review and feedback . By the end of week 12, the project should be complete, and the final deliverables should be submitted. These deliverables include a functional deep regression model for decoding dark matter with strong gravitational lensing,
a detailed project report, and well-documented code. Additionally, the project's success will be evaluated based on the quality of the model and how well it meets the project objectives.

7.) Projects

1. REGRESSION MODELS

I have done a project on House Price prediction in Python. As I have trained the model to determine the Continuous values, so i used three Regression Models

- SVM - Support Vector Machine
- Random Forest Regressor
- Linear Regressor

Conclusion : Out of above regression models , SVM model gave me better accuracy , further to get much better results I ensemble learning Techniques like Bagging and Boosting.

2. OPTMIZATION TECHNIQUES

Also I have done Project on Optimization problem like :

- Genetic Algorithm
- Particle Swarm Optimization
- Ant Colony Optimization

8.) Open Source

This is my first attempt at open source development.From a personal standpoint, I am interested in Open Source Development because it provides an opportunity for me to continuously learn and improve.The collaborative nature of Open Source Development means that I can benefit from the knowledge and expertise of a global community of developers, leading to more effective and innovative solutions.

9.) Qualifications

University : National Institute of Technology ,Jamshedpur, India
Major : Computer Science
Course : Masters in Computer Application(MCA)
Currently Pursuing : 2nd year
Expected Graduation: 2024
Languages : English **,** Hindi

Technical Skills :

- **Programming Languages** : C/C++ , Java , Python , JavaScript
- **Web Technologies** : HTML/CSS , JavaScript , XML , Node.Js , Express
- **Databases** : MySQL
- **Machine Learning and Data Science** : Scikit-Learn
- **Operating System** : Windows and Linux
- **Version control systems** : Git

Along with these skills I have basic knowledge of PHP and CyberSecurity . I have started deep learning in which algorithms can ingest and process unstructured data , like text and images , and it automates Feature extraction and detection.

10.) Summer Plans

I have no other commitments this Summer. So I'll be able to give 40 hours or more per week. I
I am ready to commit extra time if needed in order to finish up the goals of the project. My
Summer break starts from 22nd May 2023, so I can start working full time from that day on. I'll not be taking any vacations. My classes start on around 16th July but I will be able to commit enough
time for the project as there are no exams during the period.

11.) GSoC Experience

No,I have not participated in previous GSoC projects nor I have not applied for any other Organization.I am totally focused on Machine Learning .

I decided to apply for the Google Summer of Code because the kind of exposure and experience a platform like this would provide me is a huge reason for me to want to be a part of it. I have developed a lot of projects during my college days, but I always wanted to contribute that really help people at a global scale. I find nothing more exciting than working for a company like Google with a goal as impactful as this one. I sincerely love the goal of ML4SCI brings together researchers from Universities and Scientific laboratories with motivated students to join existing Scientific collaborations and contribute to cutting edge Science projects across a wide variety of disciplines and would be more than happy to work towards it. It would be really great for me to apply my skills and contribute to such an organization.

I would like to follow up on this project even after the GSoC program is over. I'll be happy to maintain this part of the project after the GSoC period.

12.) Final Words

I will make the necessary changes post-submission in my proposal, if so required by the mentors. I truly hope to see this project happen. I have basic knowledge in the field of MACHINE LEARNING , I have been trying to improve upon that by contributing in open source communities.

14.) Broader Impact

This work serves to augment the understanding and application of machine learning in cosmology - which is still very much in its initial stages. This work serves to increase the accessibility to those interested in applications of machine learning for strong lensing applications around the globe as our simulation data set and analysis pipeline is open sourced. Given the computational requirements of our implementation, those who have limited access to computing power may be at a disadvantage [Sergei Gleyzer et al. University of Alabama et. al]

References

1. Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. A&A, 594(A13):63, 2016. arXiv.
2. L. Anderson, E. Aubourg et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples . MNRAS, 441(1):24– 62, 2014. MNRAS.
3. C. Heymans, L. van Waerbeke et al. CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey . MNRAS, 427(1):146–166, 2012. MNRAS.

4. Matthew R. Buckley and Annika H. G. Peter. Gravitational probes of dark matter physics. Phys. Rept., 761:1–60, 2018.
5. Alex Drlica-Wagner et al. Probing the Fundamental Nature of Dark Matter with the Large Synoptic Survey Telescope. arXiv e-prints, page arXiv:1902.01055, February 2019.
6. Joshua Simon et al. Testing the Nature of Dark Matter with Extremely Large Telescopes. Bull. Am. Astron. Soc, 51(3):153, May 2019.
7. Stephon Alexander, Sergei Gleyzer, Evan McDonough, Michael W. Toomey, and Emanuele Usai. Deep Learning the Morphology of Dark Matter Substructure. Astrophys. J., 893:15, 2020.
8. Ana Diaz Rivero and Cora Dvorkin. Direct Detection of Dark Matter Substructure in Strong Lens Images with Convolutional Neural Networks. Phys. Rev. D, 101(2):023515, 2020.
9. Sreedevi Varma, Malcolm Fairbairn, and Julio Figueroa. Dark Matter Subhalos, Strong Lensing and Machine Learning. arXiv e-prints, page arXiv:2005.05353, May 2020.
10. Johann Brehmer, Siddharth Mishra-Sharma, Joeri Hermans, Gilles Louppe, and Kyle Cranmer. Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. Astrophys. J., 886(1):49, Nov 2019.