Personal Information

Name：Zhongchao Guan

English Name: Andrew Guan

Telephone: +8617861401509

Email: gg884691896@gmail.com

University: Beijing University of Posts and Telecommunications

Major: Machine Learning and Multi-Modal Machine Learning

Github: AllenWrong

Location: Beijing, China

Time Zone: UTC+8

Mentors

- Michael Toomey (Brown University)
- Stephon Alexander (Brown University)
- Brandon Ames (University of Alabama)
- Sanaz Kiyadeh (University of Alabama)
- Yurii Halychanskyi(University of Washington)

About me

I am a graduate student in my first year at Beijing University of Posts and Telecommunications and studying machine learning and multi-modal machine learning. With Andrew Ng's course, I have learned machine learning and deep learning. Recently, I am very interested in energy-based model. I have done many images classification projects with the VGG, ResNet, Xception and so on. I am proficient in python, java and have the ability to write English documents. Besides, I am familiar with PyTorch, NumPy, pandas and Matplotlib. I think I am competent for this work. I have built professional experience through my past coursework which help me to grow and develop knowledge in machine learning. I am willing to contribute in the project with my passion and expertise.

Why this project

I am excited to see the application of deep learning in the area of science. I can’t wait to participate in this community and use deep learning to solve some problem in other nature science area. Opensource project is a great opportunity for me to gain more practical experience.

I have done many classification tasks. In my opinion, there exists much similarity between the task of regression and classification task, the classification tasks are much easier compared with

regression task. This project is about using SOTA deep learning to solve a regression task and which align with my goal to explore regression task.

Project description

"This project will focus on further development of the DeepLense pipeline that combines state-of-the art of deep learning models with strong lensing simulations based on lenstronomy." This project focus on using deep regression techniques for estimating dark matter properties. This project requires contributor to use deep regression models to expand the DeepLense.

The idea is simple. But there exist many points needed to be thought carefully.

- **Data**. In the deep learning task, although the quality of the data is not so important as the traditional machine learning task, it also has an influence on the performance of deep models. Some appropriate transforms of the data are useful for the deep model to recognize the pattern.
- **Model**. Selecting the appropriate model is also important. Some model is too simple that it is not so powerful to recognize the pattern in the data. Some model is so complicate that it is so powerful that it can overfit the data easily. Selecting model according to the current SOTA model or experiments is good idea.
- **The order of magnitude of label**. According my experience of the evaluation test, I found that the order of magnitude of label is critical. For example: Suppose that the values: mse is 1e-2, label of sample a is 10.23, label of sample b is 0.03. What I mean is that the current mse is pretty well for the sample a but pretty bad for sample b. I need to reconsider this problem. Maybe some normalization tricks is useful.
- **Tricks of training.** Regression task has its own model training trick. This should be considered carefully.

Commitments

If I am selected, I am willing to work with mentors and other coworkers to achieve the common goal. I will be available and able to work at May 20.

My summer vacation starts by the July 11, it is feasible for me to start coding officially on June 13. But at the phase (June 13 - July 11), I can commit average 5 hours a day and about 35 hours a week. After July 11, I can commit about 6-8 hours a day and about 40 hours or more a week. I can put in more effort if the project requires after July 11.

Summary:

Phase	Do	Time And Effort
May 20 -	Communicate with mentors，read documentations and begin working	about 5 hours a day / 35 hours a week

June 13 - July 25	Officially coding and research papers	about 5 hours a day / 35 hours a week
July 26 - September 1	Involving the wholeness of myself to this work.	about 6-8 hours a day / 40 hours a week
September 2 - end	Repair some detail.	about 5 hours a day / 35 hours a week

Timeline

May 20 - June 13

During this period, I will communicate with mentors, read documentations and begin working.

Specifically:

- Understand the detail about the project, the detailed requirement of this project.
- Understand what the documentations should I write.
- Understand what I need to report every week and the form of report.
- Begin working.

June 14 - June 15

Explore the dataset.

Specifically:

- Evaluate the quality of the data simply
- If the data is very large, config a better hardware computer.
- Show the dataset images. Feel it first through my eyes.
- Explore the distribution of regression labels then decide if I need to do some preprocess
- Read some papers to investigate some common data preprocessing methods.

June 16 - June 17

Build a baseline using PyTorch.

June 18 - June 19

Buffer day. I think the buffer day is necessary in plan. Buffer day can be used to cope with some uncertainty.

June 20 – July 1

Research papers, brainstorm and form some experiment plan.

Specifically:

- Research the recently SOTA deep models and design how to use them for this task.
- Research the regression cases and summary the success experience of them.

- Research the application experience of the deep model in Dark Matter. Some experience can give me inspiration.
- Form literature research report.
- Design experiment plan (I need to try many plan). (Consult the mentors to check the feasibility.)

July 2 - July 24

Experiment with the plan.

Specifically:

- Execute the experiment plan.
- Record some inspiration, idea and new findings and ideas. Talk to mentors about the new found and new idea.
- Choose 2-4 better plan which will be further tuned.

July 25

Submit phase 1 evaluation.

July 26 - July 29

Phase 1 evaluation.

Specifically:

- Talk to mentors about the progress of this project.
- Ask for some advice for the next phase.

July 30 - September 1

At this phase, I will do experiment selected from the previous stage with the new idea and new found continuously.

Specifically:

- Iterative experiment. (experiment → record new idea, new found, (or talk with mentors) → literature research → design experiment → experiment)
- Determine the final results.

September 2 - end

Finalize some detail. Documenting. Submitting.