Kirin Patel
AE 457
October 7th

Assumptions and Constraints

Assumptions

1. Material resources
    a. It will be assumed that I have access to at least two computers, my Macbook and my Windows desktop. I plan on using my Macbook for most of the development of the machine learning model while my Windows desktop will be used for model training if I do not have access to a Kubernetes cluster for use with Kubeflow Pipelines.
        i. After my Internship at Google, I talked with my Host (manager) about using one of his Kubernetes clusters for this project. While he agreed to me using his cluster it should not be assumed that this access will be granted for the entirety of the upcoming school year.
2. Technological resources
    a. Either of the following two approaches, Python3 will be the language of choice, this is due to the low complexity of the language and large support from third party libraries. Additionally, Python2.7 will not be considered due to its deprecation which is occurring soon.
    b. Machine Learning approach
        i. Keras
        ii. Kubeflow Pipelines (if I still have access to a Kubernetes cluster)
        iii. Will likely require an Open CV approach to handle data pre-processing in an efficient manner
    c. Open CV/Feature Recognition approach
        i. There will need to be a broad understanding for the shape and color of meteors within a spectrograph
3. Financial resources
    a. No financial resources are expected to be required. However, should a budget be obtained, money can be used to train a machine learning model on TPUs/GPUs/CPUs within GCP, but this is not required as I have a personal machine at home with a GTX 1070 which can be used to train a machine learning model.
4. Logistical resources
    a. Assuming spectrographs are provided in image (JPG) format
    b. Assuming data will come pre-processed (locations of meteors already provided for each spectrograph)
5. Referential resources (people, research, etc.)
    a. Hervé Lamy

        i. Hervé is the project lead of BRAMS and will be the direct contact for obtaining spectrograph data from BRAMS.
        ii. It is crucial to get in contact with him ASAP so a model/algorithm can be created.
        iii. Should I be unable to obtain data from Hervé, I can scrape the website using Python libraries manually.
    b. Use this paper to implement a machine learning model/algorithm that will account for color saliency
6. Other factors that you identify
    a. Motivation
        i. I often lose motivation and switch focus between multiple projects at a time. It is likely that this will become an issue throughout the development of this project.

Constraints

1. Limitations of the project manager (student and mentor)
    a. Time
        i. If I am unable to obtain pre-processed data from BRAMS and Zooniverse, then I will have to scrape BRAMS manually to download spectrographs. After they are downloaded, I will also have to manually process the data to identify meteors within the spectrographs. This is required because without this information, a machine learning model cannot be trained in any meaningful way.
2. Anticipated barriers to completion of the project
    a. Lack of data
        i. While BRAMS collets multiple spectrographs per minute, that data is not labeled. This means that the data must first be pre-processed before it can be used for training a machine learning model.
    b. Lack of pre-processed data
    c. Low knowledge of Keras/Open CV
        i. I have investigated Keras but I do not have much knowledge on how to use it to generate machine learning models
        ii. I have used Open CV in the past to do real time image processing and feature detection but I do not remember much of what I did previously
3. Anticipated barriers to large-scale implementation of the project
    a. Lack of data
        i. Zooniverse has roughly 12000 classifications, while this is a large number, it is not as large of a data set that is desired for large-scale implementation. Accuracy concerns arise with the size of the dataset as there are likely edge case situations that are being excluded.
    b. Low knowledge of Keras/Open CV

4. Generalizability of the project findings
    a. Project will likely result in a machine learning model that is able to detect vertical features within a spectrograph. Should this be expanded upon to include color salience, then the model will become more refined but also limited in terms of generalizability.
5. Other factors that you identify
    a. If Kubeflow is used as a method for training the machine learning model, then all future trainings will require the usage of Kubeflow as well. Kubeflow runs on Kubernetes, which when run within GCP or other cloud providers can become extremely expensive. There are local solutions such as minikube, which allows for a Kubernetes cluster to be run locally, but this still is not ideal as it would be an extremely bulky process.