Kirin Patel

AE 457

November 11th

Requirement Specifications

There is no way to quickly and effectively identify meteors within spectrographs with a computer. This is problematic because people are required to do the tedious work of manually identifying meteors within spectrographs. While this may take less than a minute per spectrograph, there is a constant stream of spectrographs from BRAMS (Belgian Radio Meteor Stations). To further complicate matters, there are "... [at least] 30 receiving stations are spread on the Belgian territory..." (Gamby, n.d.). This means that BRAMS alone can produce 30+ spectrographs per minute across all of their stations, thus highlighting the major drawback for a person based meteor identification approach. It is both impractical and expensive to employ people to manually identify meteors.

BRAMS currently relies on the Radio Meteor Zoo (RMZ), hosted by Zooniverse is used to crowdsource people to identify meteors within spectrographs (Zooniverse, n.d.). While this alleviates the monetary cost and strain involved with meteor identification, the RMZ still requires the interest and free time of people to be effective. For an effective solution to be presented to BRAMS, a reliance on people and time must be eliminated. By doing so, the interactions between the Earth, meteors, and space can be better inspected if additional time is obtained through more efficient methods of processing data.

This is where a machine based approach is crucial. A solution that can run on low end and/or existing hardware, easily identify meteors within spectrographs, and be continually refined to increase accuracy and robustness provides a clear answer for implementation. Machine learning should be used to generate a feature recognition model that can identify meteors within spectrographs. By using Machine learning to generate a model, the model can be continually improved through future training as more data is accumulated. This allows the machine learning model to become more robust and eliminate edge cases as it learns from new data sets. For additional clarification, the data set will be comprised of spectrographs that are converted to JPG images. This will allow for color salience to be used alongside feature detection. Color salience assigns importance to the colors identified within an image. By combining shape salience which is generally used for feature detection with color salience, the accuracy and effectiveness of the model can be greatly improved when comparing it to a feature detection model that only recognizes shape salience (Weiger, 2006, p. 100).

For the machine learning model and final system implementation, there is a list of requirements that must be met so that the effectiveness of the overall approach can be determined. The requirements are as follows:

Item Name	Minimum Expected State	Desired Expected State	Future State of Project After Senior Design
Machine Learning Model	Model code exists - Written with the Keras framework - Is integratable with Kubeflow Pipelines and/or a standalone training environment	Model is generated - Is able to identify some meteors within a spectrograph - Utilizes color salience to aid in feature detection	Model is generated - Model has 80% accuracy of identifying meteors within a spectrograph - Model is setup within Kubeflow Pipelines for CI/CD

			- Model is continually trained to improve robustness
Data Set	A data set of at least 500 spectrographs exists - Data is already in a JPG format - Data has meteors identified so that training can occur	A data set of at least 5000 spectrographs exist - Same requirements as minimum expected state Data is obtained from RMZ	Have a system that automatically converts a spectrograph into a JPG upon data collection within a BRAMS operated location
Documentation	Have design doc created before implementation begins - Details requirements - Details implementation phases - Details timeline Have documentation that details how to use model	Have documentation for how to initialize project for other users Take screenshots of model and other portions of system to further enhance documentation	Create wiki where documentation can be easily searched
Financial Resources	N/A, project will be conducted using resources accessible for free	N/A, project will be conducted using resources accessible for free	N/A, project will be conducted using resources accessible for free

These requirements should be assumed to be fluid. While they exist as detailed above, they may change in the future as the feasibility of the project and the detailed requirements is not fully known.

Alongside these requirements, there are some assumptions and constraints that must be stated and understood. Unfortunately, at the time of writing, BRAMS has not responded to a request for data access to the data collected via the RMZ. This is the first major constraint. All data must be collected and pre-processed manually. This will greatly limit the data set that can be used to train the machine learning model. With this limited data set, the accuracy of the generated model will be negatively affected as larger data sets allow for increased accuracy and robustness. Alongside this, there is a time limitation. The initial implementation of this project will occur over the course of 16 weeks. However, this project will be worked on alongside three other college courses and a part-time job, limiting the amount of time that can be allocated to this project. There are two additional assumptions and constraints. The first is that access to a Kubernetes cluster to run Kubeflow Pipelines may not be granted. While the project currently has approval from an engineer at Google to be ran on a Kubernetes cluster under their GCP account, this access can and may be revoked at any time. This would then result in a fallback to minikube or a non Kubeflow Pipelines based approach for the training of the machine learning model. Finally, there is my lack of knowledge surrounding Keras and machine learning in general. This will likely introduce a delay in implementation due to the time spent learning frameworks and systems that are being used to create and train the machine learning model.

In summary, I am proposing that a machine learning model that uses shape and color salience for feature detection be created and used by BRAMS to increase the speed and efficiency of meteor identification within collected spectrographs. The goal of this project is to reduce time spent on data processing so that time can be reallocated towards data analysis. The machine learning model will be built with Keras which is a TensorFlow framework, Kubeflow

Pipelines, and data collected from BRAMS. Ideally, the model will allow BRAMS to eliminate the necessity for human interaction when identifying meteors within spectrographs. However, the minimum requirements for success are that a model exists and that documentation exists for the creation of the machine learning model and how to use the model.

Work Cited

Gamby, E. (n.d.). BRAMS. Retrieved from https://brams.aeronomie.be/.

Weijer, J. V. D., Gevers, T., & Bagdanov, A. (2006). Boosting color saliency in image feature detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(1), 150–156. doi: 10.1109/tpami.2006.3

Zooniverse. (n.d.). Retrieved from https://www.zooniverse.org/projects/zooniverse/radio-meteor-zoo.