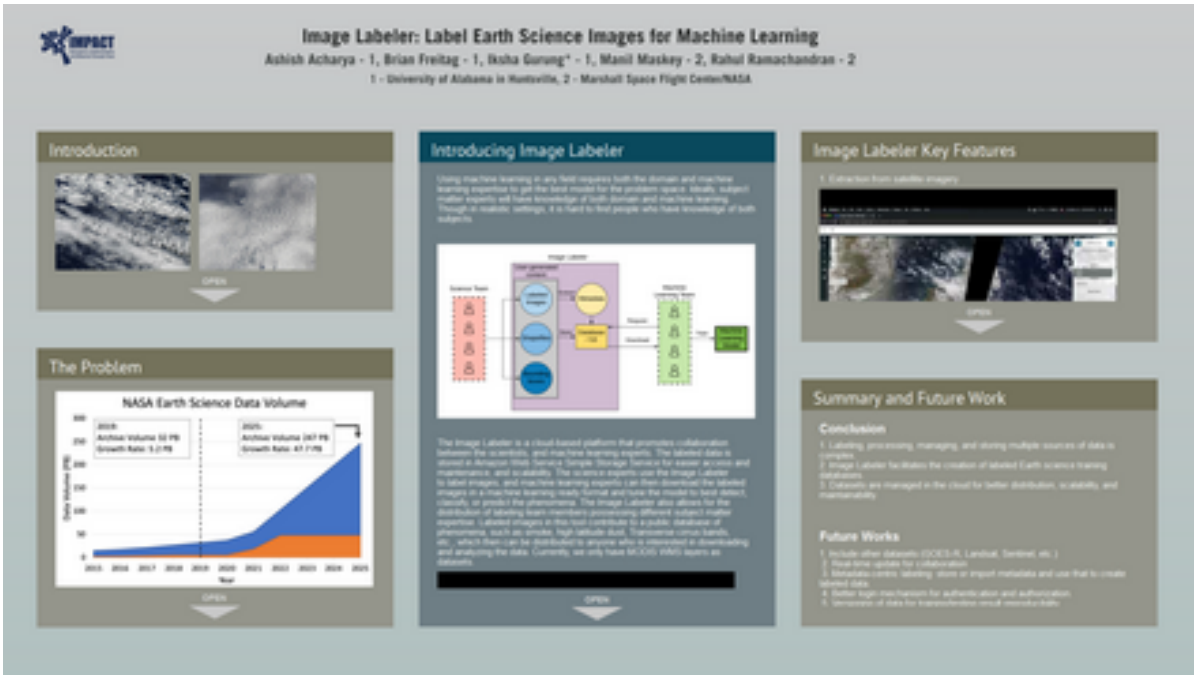
Image Labeler: Label Earth Science Images for Machine Learning

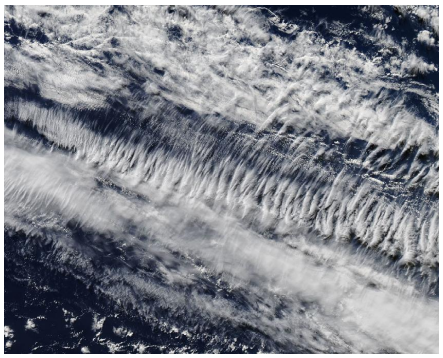


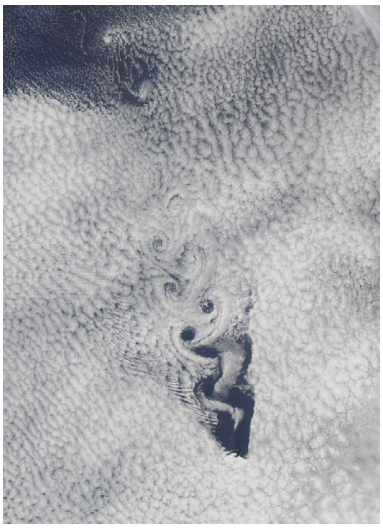
Ashish Acharya - 1, Brian Freitag - 1, Iksha Gurung* - 1, Manil Maskey - 2, Rahul Ramachandran - 2

1 - University of Alabama in Huntsville, 2 - Marshall Space Flight Center/NASA

PRESENTED AT:

INTRODUCTION





Advances in machine learning (ML) have made it possible to automatically detect Earth science phenomena from satellite imagery. While useful, ML algorithms typically require an extensive dataset containing labeled images for training. Developing a systematic approach to labeling and management of the resulting training datasets is quite cumbersome. For example, multiple people might be working on the same dataset, sharing it across for consistency becomes an issue as the number of people involved increases. With this in mind, we present the Image Labeler. The Image Labeler is a fast and scalable cloud-based tool that facilitates the rapid development of Earth science event databases, to aid automated ML-based image classification.

THE PROBLEM

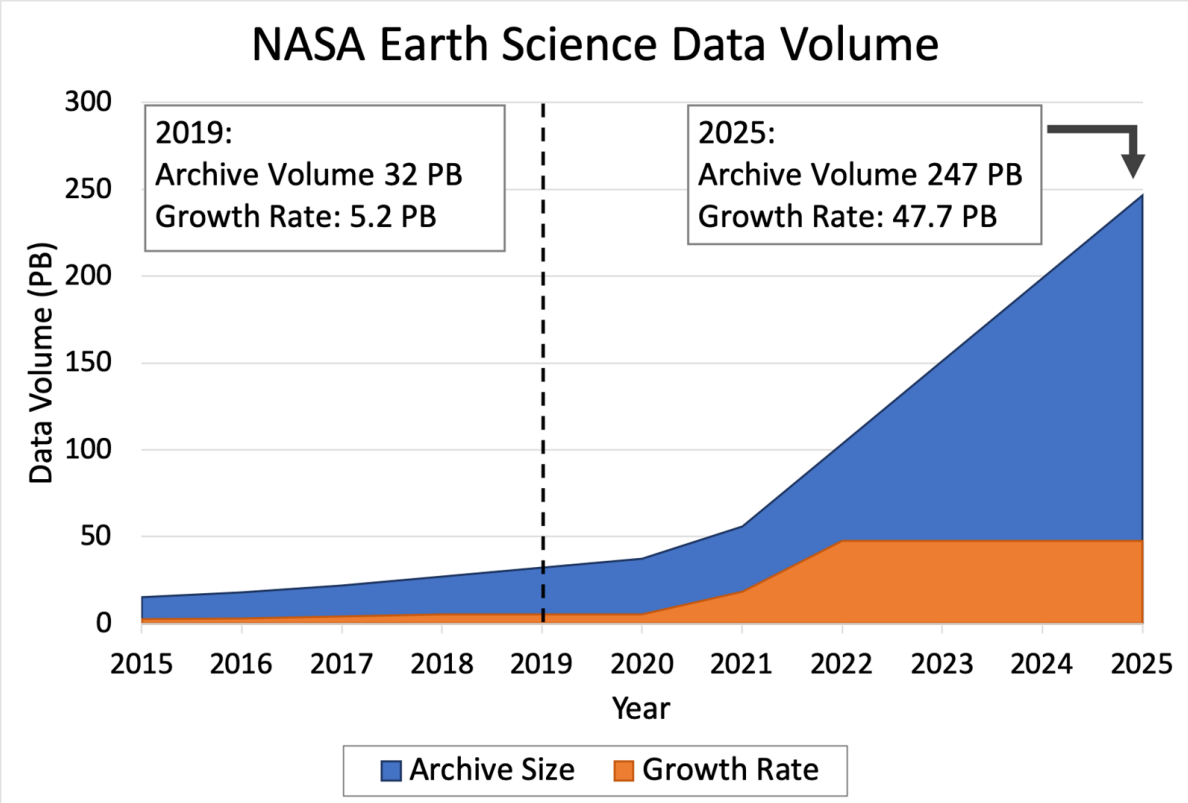


Figure courtesy https://earthdata.nasa.gov/esds/nasa-earth-science-data-systems-program-highlights-2018 (https://earthdata.nasa.gov/esds/nasa-earth-science-data-systems-program-highlights-2018)

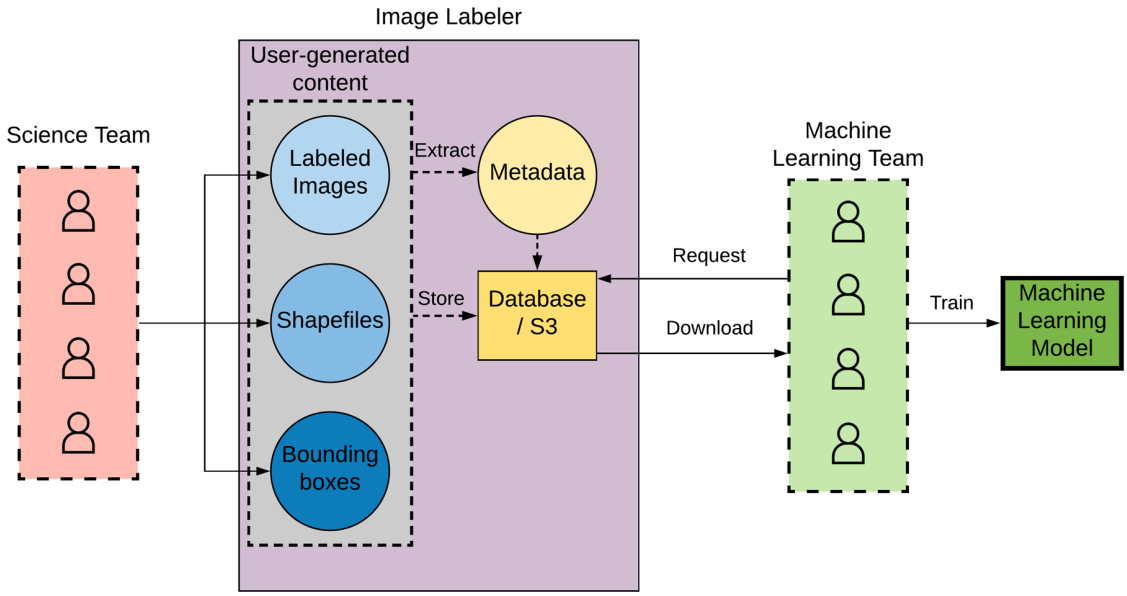
Machine learning algorithms are data-driven; more data leads to better models. The volume of NASA Earth science data is increasing exponentially making machine learning a valuable approach to making use of the large data quantity. However, there is a lack of labeled data limiting the application of machine learning algorithms in Earth science.

In recent years, the scientific community has started applying machine learning to Earth science data. While there are multiple efforts, most of the labeled data generation is done manually and spread across multiple machines. After the experiments, labeled data generated remains stagnant and it is hard to share it with interested people.

Labeled data in Earth science, in essence, is a phenomena database containing a list of phenomena with well-curated data labels representing phenomena. There is also a lack of publicly available phenomena databases that can be leveraged by Earth science experts for any analysis.

INTRODUCING IMAGE LABELER

Using machine learning in any field requires both the domain and machine learning expertise to get the best model for the problem space. Ideally, subject matter experts will have knowledge of both domain and machine learning. Though in realistic settings, it is hard to find people who have knowledge of both subjects.



The Image Labeler is a cloud-based platform that promotes collaboration between the scientists, and machine learning experts. The labeled data is stored in Amazon Web Service Simple Storage Service for easier access and maintenance, and scalability. The science experts use the Image Labeler to label images, and machine learning experts can then download the labeled images in a machine learning ready format and tune the model to best detect, classify, or predict the phenomena. The Image Labeler also allows for the distribution of labeling team members possessing different subject matter expertise. Labeled images in this tool contribute to a public database of phenomena, such as smoke, high latitude dust, Transverse cirrus bands, etc., which then can be distributed to anyone who is interested in downloading and analyzing the data. Currently, we only have MODIS WMS layers as datasets.

[VIDEO] https://www.youtube.com/embed/wD6uDLYERtA?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

https://labeler.nasa-impact.net

IMAGE LABELER KEY FEATURES

1. Extraction from satellite imagery

[VIDEO] https://www.youtube.com/embed/V7ptvoBdy6k?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

2. Tagging/Untagging images with an example earth science event

[VIDEO] https://www.youtube.com/embed/iVlbaLKm1Gw?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

3. Team-based collaborative labeling

[VIDEO] https://www.youtube.com/embed/qHb3La0vMDg?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

4. Support for geotiffs and shapefiles

[VIDEO] https://www.youtube.com/embed/UTeDGOWfZ-U?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

5. Extract Bounding boxes over images.

[VIDEO] https://www.youtube.com/embed/S-m0mhpmltQ?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0

SUMMARY AND FUTURE WORK

Conclusion

1. Labeling, processing, managing, and storing multiple sources of data is complex.
2. Image Labeler facilitates the creation of labeled Earth science training databases.
3. Datasets are managed in the cloud for better distribution, scalability, and maintainability.

Future Works

1. Include other datasets (GOES-R, Landsat, Sentinel, etc.)
2. Real-time update for collaboration
3. Metadata-centric labeling: store or import metadata and use that to create labeled data.
4. Better login mechanism for authentication and authorization.
5. Versioning of data for training/testing result reproducibility

Sorry but time is up!

ABSTRACT

The application of machine learning for image-based classification of earth science phenomena, such as hurricanes, is relatively new. While extremely useful, the techniques used for image-based phenomena classification require storing and managing an abundant supply of labeled images in order to produce meaningful results. Existing methods for dataset management and labeling include maintaining categorized folders on a local machine, a process that can be cumbersome and not scalable.
Image Labeler is a fast and scalable web-based tool that facilitates the rapid development of image-based earth science phenomena datasets, in order to aid deep learning application and automated image classification/detection. Image Labeler is built with modern web technologies to maximize the scalability and availability of the platform. It has a user-friendly interface that allows tagging multiple images relatively quickly. Essentially, Image Labeler improves upon existing techniques by providing researchers with a shareable source of tagged earth science images for all their machine learning needs. Here, we demonstrate Image Labeler's current image extraction and labeling capabilities including supported data sources, spatiotemporal subsetting capabilities, individual project management and team collaboration for large scale projects.

SWITCH TEMPLATE