End-to-end Machine Learning Applications Framework for Earth Science

Brian Freitag[1], Ashish Acharya[1], Muthukumaran Ramasubramanian[1], George Priftis[1], Aaron Kaulfus[1], Iksha Gurung[1], Manil Maskey[2], Rahul Ramachandran[2], Drew Bollinger[3]

1 – University of Alabama in Huntsville; 2 – NASA Marshall Space Flight Center; 3 – DevelopmentSeed

IMPACT Interagency Implementation and Advanced Concepts Team

Introduction

An end-to-end machine learning pipeline from data collection to model deployment has been developed to provide innovative solutions to analyze growing data volumes. Two web applications – Image Labeler and Phenomena Portal – developed by the NASA Interagency Implementation and Advanced Concepts Team (IMPACT) are used to support the rapid development, deployment, and visualization of machine learning (ML) models for Earth Science application.

NASA Data Archive

The growth rate of data into the NASA Earth science data archive will increase from 5 PB to 48 PB per year (Figure 1). Innovative solutions are required to promote the usability of this large data volume to the Earth science community.

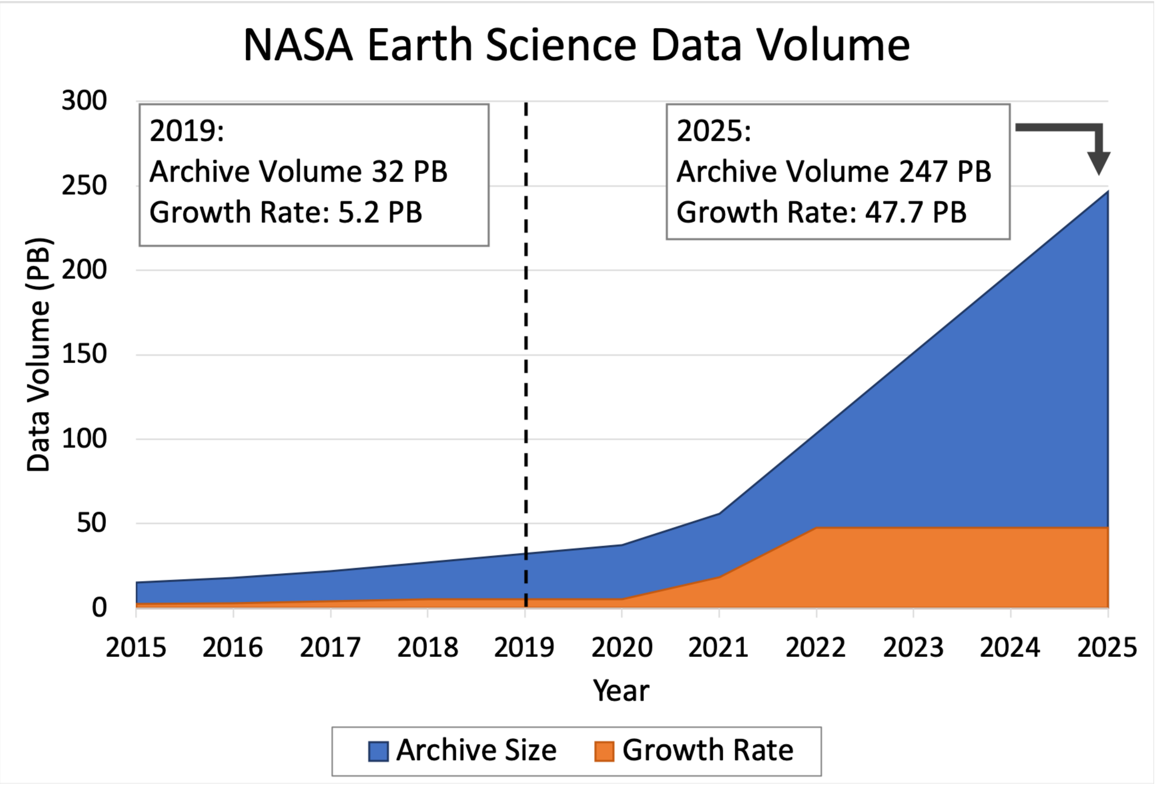

Figure 1. Projected archive volume for NASA Earth science data

ML Approach to Data Problem

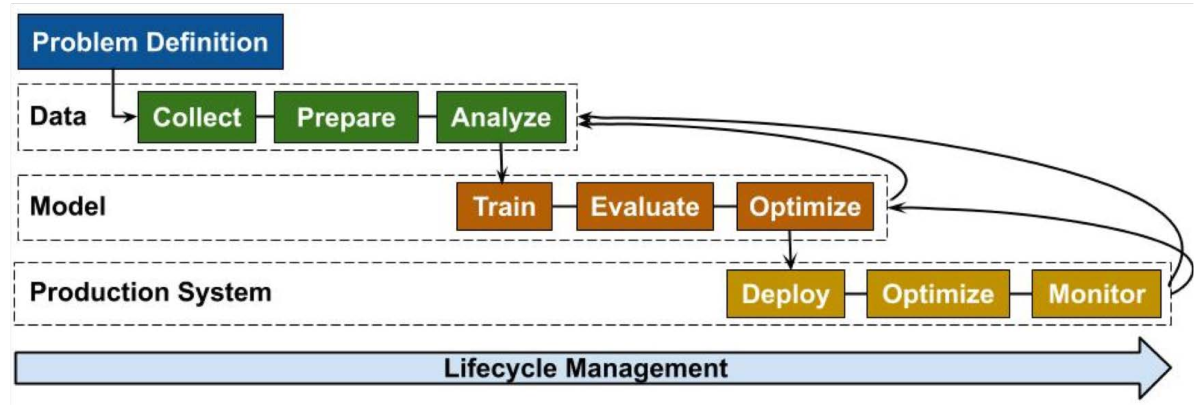

Figure 2. Machine learning lifecycle

Feature detection using machine learning is one approach to address a growing data volume (Figure 2). Using satellite imagery, training datasets are used to train ML models to detect atmospheric phenomena in images. This approach allows for the development of a database of atmospheric phenomena with spatiotemporal information useful for subsetting/analyzing large datasets. Here we demonstrate the machine learning lifecycle using a ML model to detect high latitude dust (HLD).

Data – Collect, Prepare, Analyze

The Image Labeler is a web application designed for the rapid development of image-based ML-ready training data that can be used to train ML models (**labeler.nasa-impact.net**).

Image Labeler Features Include:

1. Team collaboration	4. Bounding box for AOI
2. Image segmentation	5. Cloud Storage
3. Supports user-generated content (images, geotiffs, etc.)	6. Multiple image layers from GIBS for image extraction

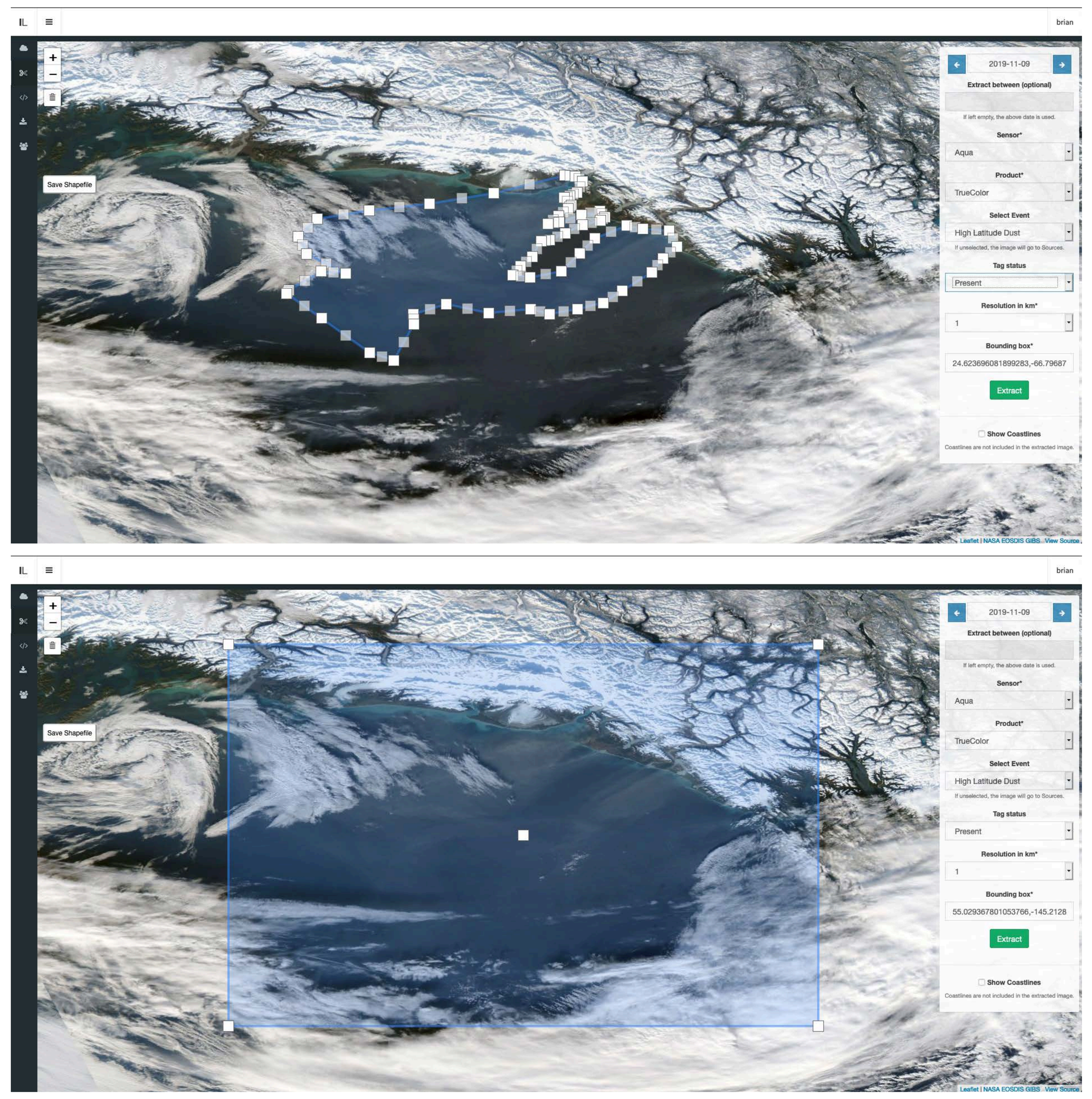

Figure 3. Screenshots from HLD event in Image Labeler

Model – Train, Evaluate, Optimize

More than 150 training samples are generated using the bounding box and image segmentation features in Image Labeler. Training samples are then used to train a pixel-level ML model to detect HLD events in MODIS true color imagery. Model performance metrics are used to evaluate and optimize the ML model configuration. Output from the model is the probability of dust for a given pixel.

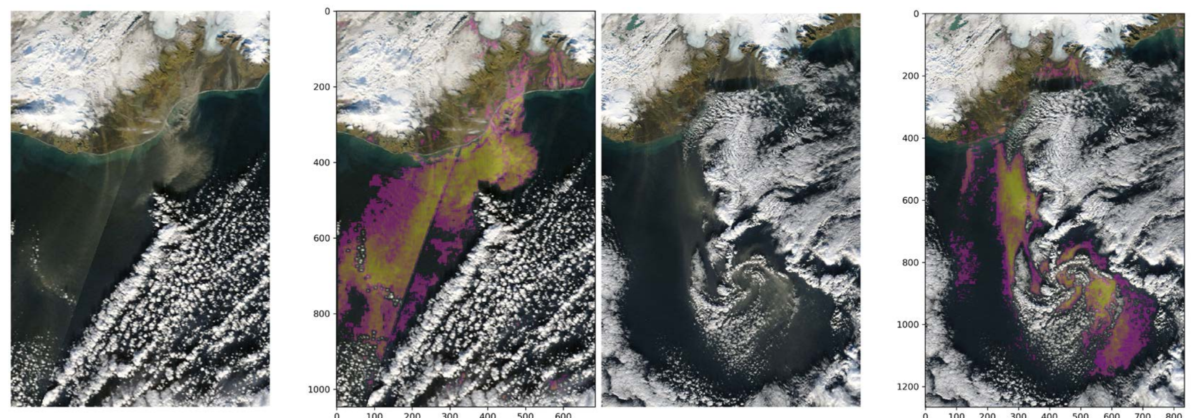

Figure 4. Pixel-based detection of HLD with dust probability shaded.

Production System

Output from ML models are displayed in the Phenomena Portal which serves as a near real-time event detection application with a searchable event database. The user interface allows for feedback on confirmed detections which may be used to augment training datasets for future model versions. Spatiotemporal information from detected events can be passed into ancillary datasets for more detailed scientific study (**phenomena.surge.sh**).

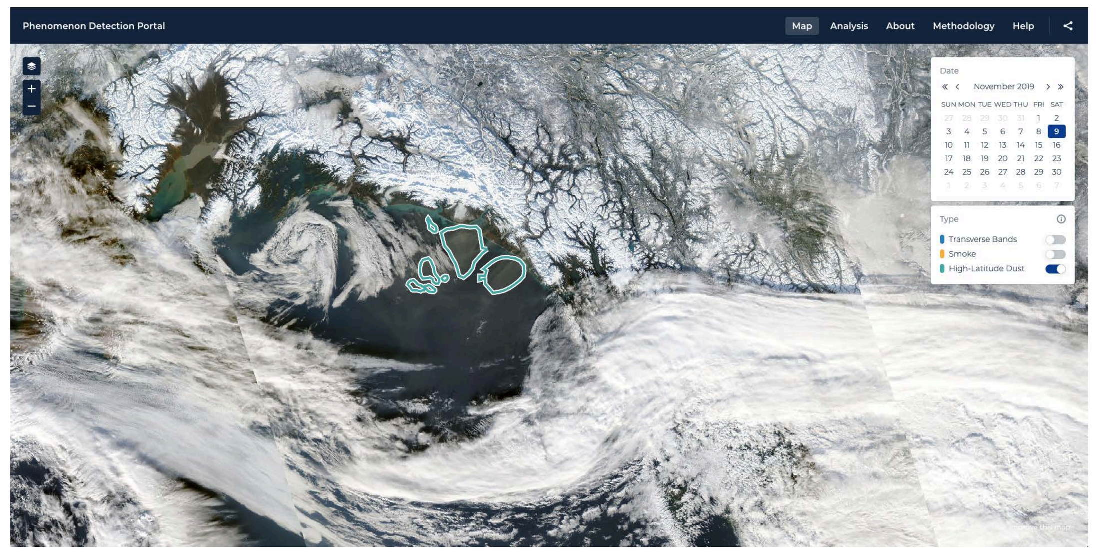

Figure 5. HLD detections in the IMPACT Phenomena Portal

Conclusion

Two web applications were developed to support the rapid deployment of ML models for Earth science application. The web applications are open source and leverage crowd sourcing for labeling training data and validation.

Contact: brian.freitag@nsstc.uah.edu

developmentSEED   THE UNIVERSITY OF ALABAMA IN HUNTSVILLE   NASA