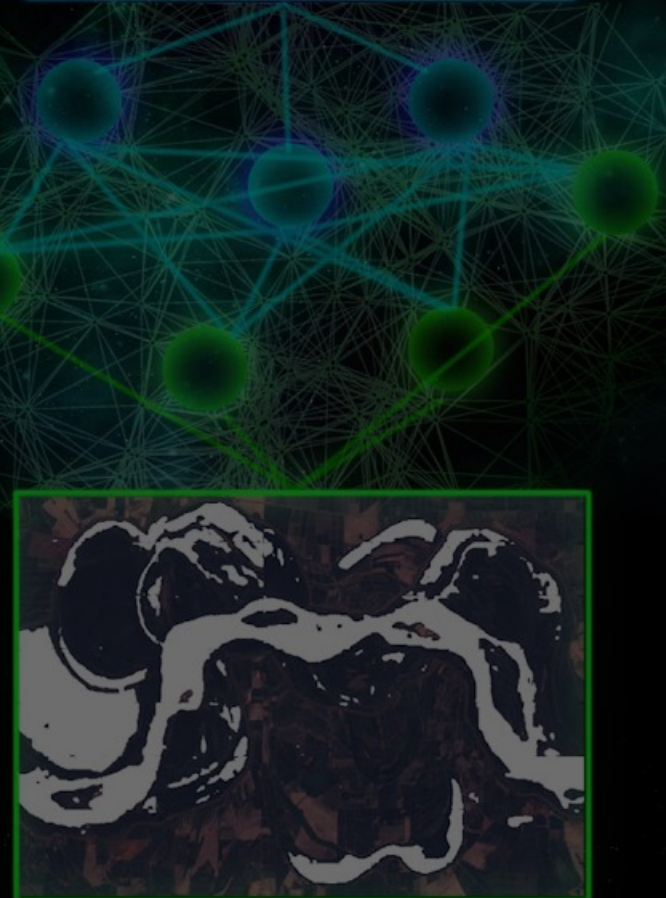


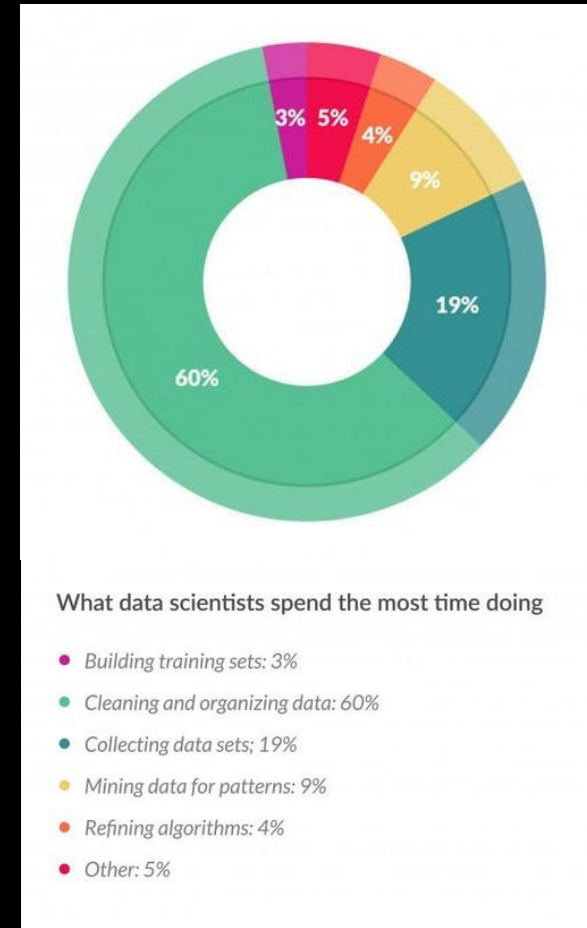
DELTA: An Open-Source Framework to Simplify Machine Learning with Satellite Imagery

Michael von Pohle, Brian Coltin,
Scott McMichael



DELTA Helps Address Two Big Problems

- Labor and time intensive preprocessing, postprocessing, and boilerplate code to turn satellite imagery into an ML-ready dataset
- Difficulty in applying trained models to satellite imagery. Especially in the earth science community among those with limited ML experience



A Few Notes on Satellite Imagery

- Single images are often 1GB+ - 10k x 10k pixels or more
- A dataset of imagery can easily be several hundred GBs in size
- Often hyperspectral – more than RGB
- Can include metadata layers alongside data layers
- May not be rectangular

DELTA Features

- Works with any geospatial imagery/data that GDAL supports
- Has support for unpacking and preprocessing a growing list of common raw satellite imagery sources (Worldview, Sentinel, Landsat)
- Has a collection of common image preprocessing functions
- Performs tiling of imagery, training/inference, and restitching of predictions with adjustable overlap
- Works with h5 and SavedModel model formats
- Uses human readable YAML config files
- Has single node multi-GPU support
- MLFlow/TensorBoard integration
- Modular and extensible for adding custom image formats/layers/losses/metrics/callbacks/augmentations/preprocessing functionality
- Can import ImageryDataset class for use in your own framework to leverage DELTA's imagery handling anywhere you used Tensorflow Dataset

Example: Mapping clouds with Landsat-8 images

- Using USGS SPARCS dataset – Landsat 8
- Demonstrate:
 - Ease of ingesting imagery and labels for ML setup
 - Ease of applying trained models to imagery
 - Ease and customizability for training and inference both for ML scientists and less experienced domain scientists
- Jupyterlab demo

DELTA

- <https://github.com/nasa/delta>
- Michael von Pohle – michael.vonpohle@nasa.gov