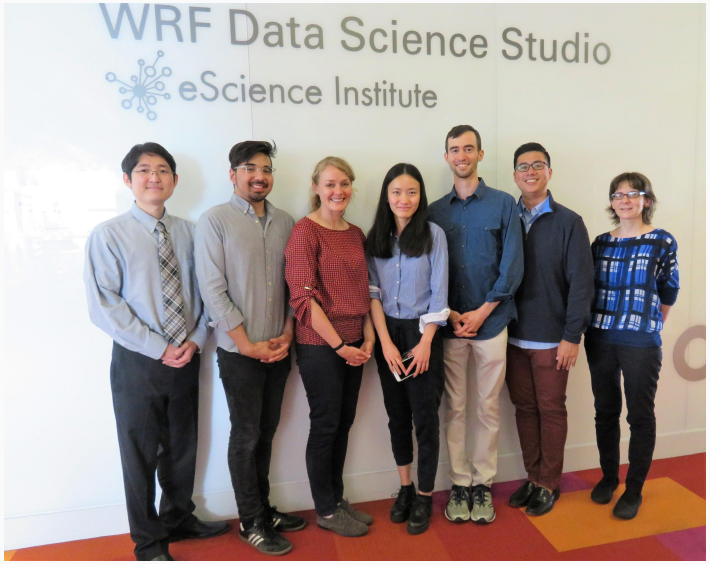
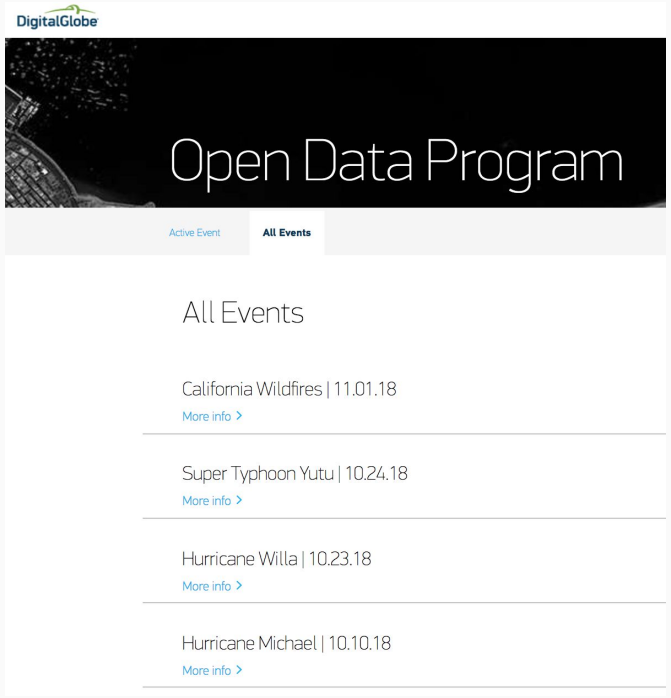
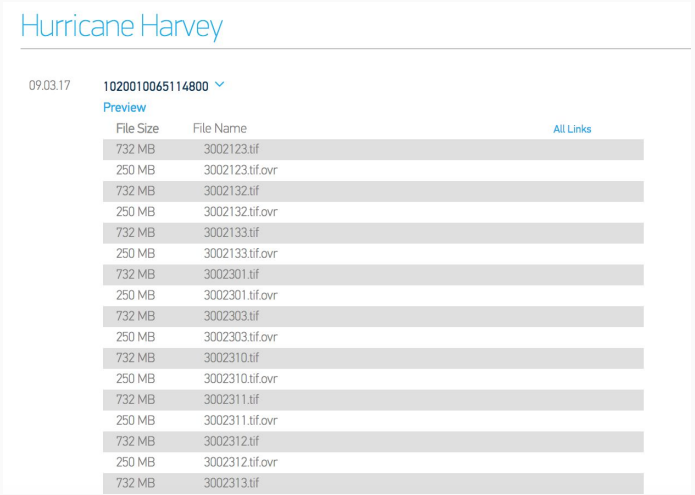
From Open Satellite Data To Emergency Response

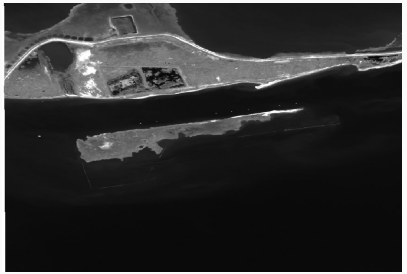
Damage Detection in Post-Hurricane Images
(summer DSSG 2018 project jointly with Disaster Data Science Lab)



Digital Globe Open Data

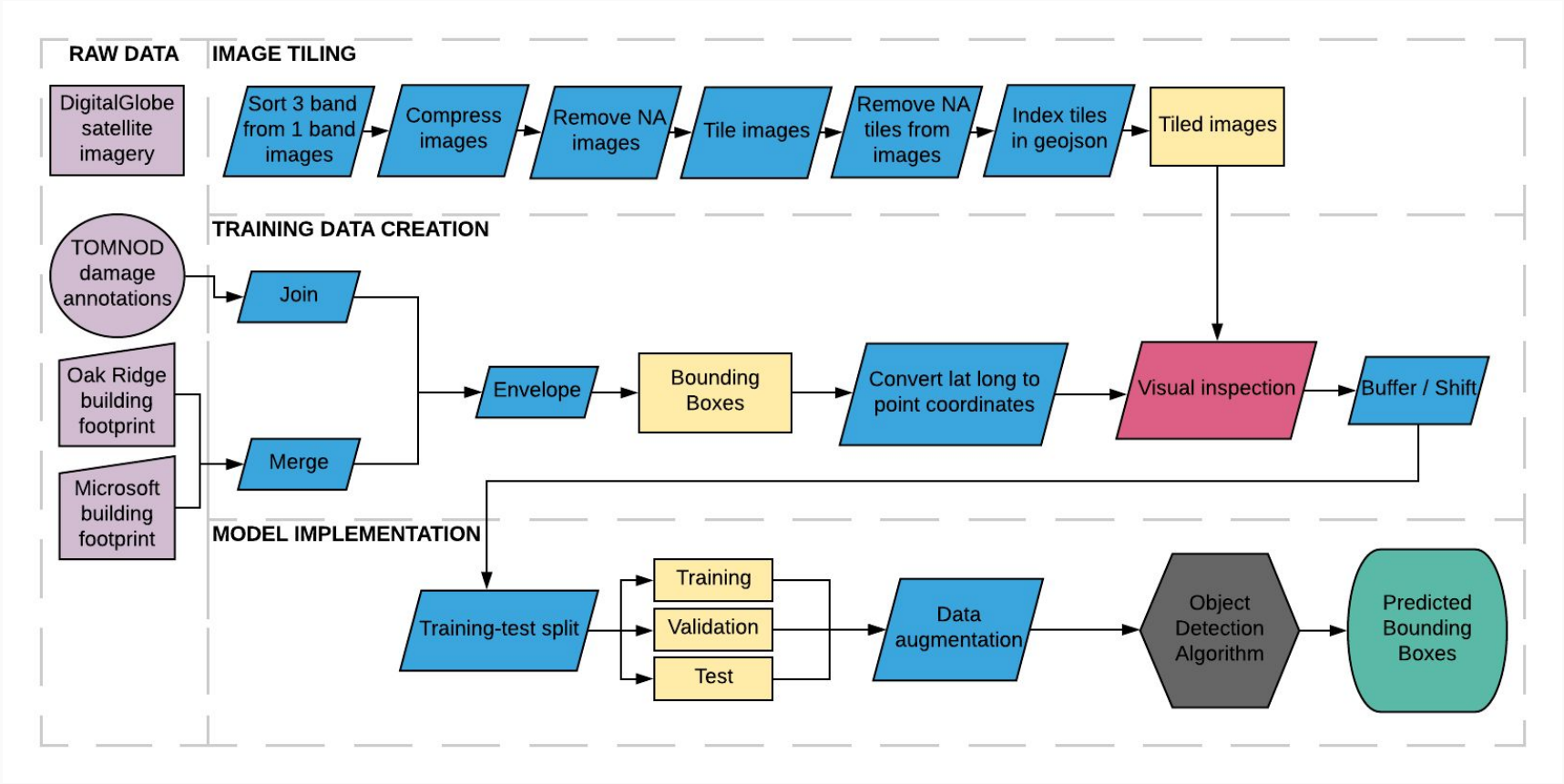






- 3 TB of image data
- Crowdsourced manual annotations in JSON

Data Processing Pipeline



Computational Resources

Hyak: Disaster Data Science Lab's Account: storage + compute nodes (bought-in esources)

Pros:	**Cons:**
● easy to experiment as not charged for every action	● no root access: ○ best to install Python packages through conda ○ Some geospatial libraries conda distributions don't have full functionality ○ no docker support

Amazon Web Services: UW eScience Account (pay per use)

Pros:	**Cons:**
● can use pre-built images: great for deep learning ● can save snapshots of all the work ● can use GPUs without dealing with hardware and drivers ● can use managed databases	● everybody needs to learn about security management ● uploading data is free, but exporting and GPU computations are expensive

Students' Laptops:

Pros:	**Cons:**
● easy to get things working	● not reproducible

Tools

- **Downloading, compressing and tiling:** Hyak => 40GB (mainly gdal)
- **Geospatial processing** (aligning building footprints, extracting bounding boxes): QGIS + Python
  - when started struggling with RAM => set up PostGIS server and used SQL queries

  => subsetted dataset => around 10GB
- **Visual inspection:** patience ...
- **Deep Learning**
  - moved training dataset to AWS ec2 GPU Instance
    - Used DLAMI AWS image with all deep learning packages installed
    - $0.9 per hour
    - set up spot instance: $0.26 per hour
  - Python Tensorflow
    - SSD Object Detection algorithm built-in

Outputs

Datasets:

- Compressed and tiled dataset
- Training Dataset
- PostGIS SQL database with geospatial data
- Pickled trained models

Cloud Backup:

- AWS S3 bucket
- Snapshots for instances + database

Dataset Upload to IEEE Dataport:
https://ieee-dataport.org/submit-dataset

- Free Open Access until end of the year!
- Up to 2TB + 10 year support

Code on GitHub:
https://github.com/DDS-Lab/

Website:
https://dds-lab.github.io/disaster-damage-detection/

IEEE DataPort™

	STANDARD DATASET	OPEN ACCESS DATASET	DATA COMPETITION
Description	Submit your dataset and related files at no cost to you and make your dataset accessible. Standard datasets may be downloaded or accessed "in the cloud" by any IEEE DataPort Subscriber.	Make your datasets freely available to the public! This option is designed for those who need to meet Open Access requirements. When you load an Open Access dataset, your dataset will be accessible to ALL logged in users of IEEE DataPort.	Initiate and host a Data Competition on IEEE DataPort by uploading a dataset and your competition instructions. Establish the competition duration, manage participation, implement submission deadlines, directly receive competition entries on IEEE DataPort, and update the competition as needed.
	Submit a Standard Dataset	Submit an Open Access Dataset	Submit a Data Competition
Cost	FREE	$1,950 (one-time fee) For a limited time, use Promo Code **OPENACCESS1** at checkout to upload an Open Access dataset at no cost.	FREE
Access	IEEE DataPort Paid Subscribers	All IEEE DataPort Users	Access Controlled by Data Competition Administrator

What did we not do?

- Cloud Removal:
  - experimented with Windows-based software - hit or miss

- Did not do fancy parallel processing: just let code run overnight :)
- Planned to set up Geoserver Visualization: have not decided where to host it yet
- We did not test the pipeline on a new dataset
- Did not request direct access to the data through Digital Globe’s S3 bucket