Capstone 3 Project Proposal: Post-Disaster Flood Damage Assessment
Allan Kapoor

Problem statement formation

In order to support effective natural disaster response and recovery, how can we leverage remote sensing imagery to quickly and efficiently produce accurate damage assessments after extreme storms?

Context

Natural disasters, especially hurricanes and extreme storms, cause widespread destruction and loss of life and cost the U.S billions of dollars each year. Of the 258 U.S. weather disasters since 1980, tropical cyclones (hurricanes) have caused the most damage: $945.9 billion total, with an average cost of almost $21.5 billion per event[1]. Extreme storms have caused widespread damage to major cities such as Hurricane Katrina (New Orleans, 2005), Superstorm Sandy (New York City, 2012), and Hurricane Harvey (Houston, 2017). More extreme weather due to climate change has increased frequency and intensity of extreme storms - 2020 was the sixth consecutive year in which 10 or more billion-dollar disaster events occurred in the U.S[2].

Effective response after a natural disaster strikes is critical to reducing harm and facilitating long term recovery. Damage assessments in the hours/days after a disaster can help efficiently allocate immediate response resources and determine spatial patterns of damage. Damage assessments also help calculate economic impact estimates which are necessary to secure and approve recovery funding.

However, completing fast and accurate damage assessments immediately after a disaster faces several challenges. Access to damaged areas may be physically cut off due to impacts to the transportation network. An extreme storm may impact hundreds of thousands of structures across hundreds of square miles, requiring significant resources if the assessment is done manually. Especially immediately after an event, limited resources may be tied up with immediate rescue and triage. A model that can perform automated damage assessment based on remote sensing imagery quickly captured by a few planes flying over a disaster site would greatly reduce the effort and increase speed in which useful data (location and extent of damaged structures) could be put in the hands of responders.

Criteria for success

The primary goal of this project is to develop a model that can classify structures as damaged or not damaged based on aerial imagery. The timeline for model development is completion by Jan 31 2021. Deliverables will include:

- A Github repo containing notebooks documenting each step of the process and the final model
- Project Report
- Project Presentation

---

[1] https://coast.noaa.gov/states/fast-facts/hurricane-costs.html
[2] https://coast.noaa.gov/states/fast-facts/hurricane-costs.html

Scope of solution space
The model will focus on classifying images of structures post-flooding into damage categories. As the training dataset consists of geospatial bounding boxes for the damaged buildings along with their damage labels, this is an image classification problem (not object detection). This assumes that a prior process (or algorithm) identifies the location of each building. This is a safe assumption as highly accurate building detection algorithms already exist. In fact, an existing one was used to identify structures in the training dataset. Microsoft has already released geospatial data of building footprints for the entire U.S./Canada. For other countries, similar datasets are in development - and OpenStreetMap already provides good coverage for most of the world.

Constraints
Imagery in the training data only has 3 bands (red, blue, green) and does not include multispectral bands collected by modern remote sensing satellites which could potentially be useful for detecting damage (especially flooding). The training dataset only consists of post-disaster imagery at a single point in time. Other approaches to damage assessment rely on change detection on before/after photos.

As the training data is from post-Hurricane Harvey Houston the trained model will likely only be accurate for areas that have a similar development pattern, building types, and vegetation patterns. For example, the trained model likely would not be accurate for assessing earthquake damage, in an arid landscape, or for a denser urban area with many large/tall buildings as the structure images would differ substantially in content from what the model was trained on.

Finally, while the full training dataset includes 500,000 images, the model will not be trained on the entire dataset to keep training time/costs manageable.

Stakeholders
In the real world, the primary stakeholders for this project would likely be:
- Federal Emergency Management Agency (FEMA)
- Office of Emergency Services (there is a similar
- Local first responders (
- Department of Defence, National Guard, etc

For this capstone project I will be primarily working with my mentor, but also seeking advice and input from coworkers who work in remote sensing and natural disaster response.

Data sources
Labels: Benchmark Dataset for Automatic Damaged Building Detection from Post-Hurricane Remotely Sensed Imagery
https://ieee-dataport.org/open-access/benchmark-dataset-automatic-damaged-building-detection-post-hurricane-remotely-sensed#files

Summary: This dataset contains shapefiles of bounding boxes of buildings labeled as either damaged or non-damaged. Those labeled as damaged also have four degrees of damage from minor to catastrophic. Each bounding box is also indexed to one of the images in the NOAA post-Hurricane Harvey imagery dataset allowing users to match the bounding boxes with the correct imagery for training the algorithm:

Images: NOAA Hurricane Harvey Imagery
https://storms.ngs.noaa.gov/storms/harvey/index.html#9/29.4324/-95.7925

This primary dataset may be supplemented with other datasets if I determine that it is necessary/feasible to incorporate additional data into the model beyond aerial imagery. Other datasets might include:

- Elevation (National Elevation Dataset)
- Soil infiltration capacity (USDA SSURGO)
- Permeability (National Land Cover Database)

Preliminary Approach

I plan to complete this project by carrying out the following steps:

- Data acquisition
    - Download imagery/bounding boxes and then run processing scripts to generate training data
    - Select a subset of the full dataset (equal number of damaged/non-damaged examples)
- EDA
    - Visual inspection of images
    - Spatial distribution of flooded/non-flooded structures
    - Comparison to environmental variables: elevation, FEMA flood zones, soil types, etc.
    - Histograms/summary statistics of image composition compared to labels
- Data Wrangling/Preprocessing
    - Normalize/standardize images
    - Consider reducing image resolution
    - Train/test/validation split
- Modeling
    - Determine evaluation metric (Accuracy?)
    - Dummy model
    - Baseline model (logistic regression?)
    - Simple CNN model
    - Iterate on CNN model
    - Model based on transfer learning (potentially)
- Final model
    - Train best model on google cloud with full training dataset
    - Look at patterns for misclassified images (visual comparison, spatial relationship, etc)

- Documentation
    - Final report, presentation
    - READ.ME
    - Clean/organize notebooks and Github repo