Capstone Project Proposal

-Nick Hack

Domain Background: Hurricane damage evaluation is a major part of community recovery. Identification of damaged areas allows federal agencies and insurers the ability to start servicing hardest hit areas. This can be a labor intensive process, FEMA one the worlds largest disaster relief organization hands out excel based surveys for local residents and authorities to fill out damage related information with a 30 day timeline to submit for aide (Preliminary Damage Assessments | FEMA.gov). This process can add weeks onto the process of distributing funds and other disaster relief efforts. Image classification models provide a unique opportunity to quickly identify areas that have hurricane or storm damage allowing relief agencies to focus their outreach efforts to areas in greatest need.

Problem Statement: Manual Hurricane and storm damage assessments are time consuming and increase the time to disbursement of relief to communities in need. Convolution Neural Network models provide an opportunity to classify damaged areas through satellite image classification. This classification process will provide areas to focus on faster than working with local authorities to manually assess, on the order of minutes to hours vs days to weeks.

Datasets and Inputs: Kaggle dataset of satellite images of hurricane damage. Provided in Project proposal submission. The Kaggle data set divided into satellite images showing damaged and undamaged areas of Texas after Hurricane Harvey. Data originally taken from: https://ieee-dataport.org/open-access/detecting-damaged-buildings-post-hurricane-satellite-imagery-based-customized and can be cited with http://dx.doi.org/10.21227/sdad-1e56 and the original paper is here: https://arxiv.org/abs/1807.01688 (Sourced: Satellite Images of Hurricane Damage | Kaggle)

Benchmark model: There have been other models created on this data set. I have chosen to use the pertained ResNet CNN model. This is a model that has been used throughout the Udacity course and is a good baseline model. The Benchmark accuracy for the ResNet model is 70% for this data set based on the work completed by RATNA SAMBHAV who submitted his results to Kaggle for public viewing on the dataset (Univ.Ai AI2 Project | Kaggle).

Evaluation Metrics: The Kaggle dataset is labeled images. Evaluation of the model will be based on accuracy of the model to classify the test dataset against the labeling of damaged vs undamaged. Based on other models that have been created on this data set an accuracy of 70% is the benchmark I will be attempting to achieve for the pretrained ResNet CNN model.

Project Design: Using the methods learned through out the Udacity course, I will first import the image data into Amazon S3 and then complete some simple heuristic analysis of the data. This will help to

identify any bias in the image data or associated information attached to each image. After a pretrained Resnet CNN model will be used to train, test, and validate classification of images by damaged and non-damaged. Once the model has been trained and the best hyperparameters chosen based on highest accuracy of classification, an endpoint and will be setup so images can be submitted for classification.