Dataset and preprocessing

Optimized Data Downloading**:** downloading files that cannot be used in the neural network is waste both of computation time and storage. The HURSAT database has images of hurricanes from around the world, but the best track data from HURDAT2 only contains wind speeds for hurricanes from the Atlantic and Pacific Oceans. So, before downloading a hurricane’s satellite imagery from HURSAT, the code checks to see whether the best track data has records for that hurricane from besttrack.csv. If best track has no data for that hurricane, the satellite image is not downloaded. This conserves local storage space and cuts down on execution time by training only necessary images.

Cropping Satellite Images: The most valuable information about a hurricane’s intensity is near the center. So, we crop satellite images to remove the outer part of the hurricane from the image. After reading the satellite image from its NetCDF file, turning it into a NumPy array, the code crops the image so that only a 50-by-50-pixel square at the center is remaining. Removing this data reduces the amount of time spent on data augmentation and model training in.

Merging HURSAT and HURDAT2 to Match Satellite Images with their Wind Speed: Each satellite image file provides us with the name of the hurricane, as well as the time and date of the satellite image. However, it does not provide us with the wind speed of the hurricane at that time. We find the wind speed for each satellite image by searching for the hurricane’s name, date, and time in the best track dataset. Once the wind speed is retrieved, both the satellite image and wind speed are appended to NumPy arrays in unison. This effectively labels the satellite image with its wind speed.