Iceberg and Ship: Image Processing and Classification

Business Problem:

Drifting icebergs present threats to navigation and activities in areas such as offshore of the East Coast of Canada.

Currently, many institutions and companies use aerial reconnaissance and shore-based support to monitor environmental conditions and assess risks from icebergs.

However, in remote areas with particularly harsh weather, these methods are not feasible, and the only viable monitoring option is via satellite.

Objective:

The purpose is to identify if a remotely sensed target is a ship or iceberg and find out which features will be most helpful for doing classification.

I will have to explore a variety of transformation techniques to better understand how ships and icebergs reflect light and use these insights to distinguish the images.

Background:

Synthetic aperture radar (SAR) images are captured from a moving antenna emitting successive pulses of radio waves.

Institutions mount these antennas on unmanned aerial reconnaissance vessels to surveil land areas. The images in this case study were collected from the Sentinel-1 satelite constillation, orbiting 600 kilometers above the earth.

The Sentinel-1 is a side looking radar (SLAR), which means that the radar points in a direction perpendicular to the direction of flight. The Sentinel-1 transmits and receives energy in two channels: a vertical channel and a horizontal channel.

These separate channels will help when analyzing the images because different objects have different polarimetric scattering features.

Data:

The data labels are provided by human experts and geographic knowledge on the target. All the images are 75x75 images with two bands.

The data (train.json, test.json) is presented in json format. The files consist of a list of images, and for each image, you can find the following fields:

id - the id of the image

band_1, band_2 - the flattened image data.

Each band has 75x75 pixel values in the list, so the list has 5625 elements. Note that these values are not the normal non-negative integers in image files since they have physical meanings - these are float numbers with unit being dB. Band 1 and Band 2 are signals characterized by radar backscatter produced from different polarizations at a particular incidence angle. The polarizations correspond to HH (transmit/receive horizontally) and HV (transmit horizontally and receive vertically). More background on the satellite imagery can be found here.

inc_angle - the incidence angle of which the image was taken.

Note that this field has missing data marked as "na", and those images with "na" incidence angles are all in the training data to prevent leakage.

is_iceberg - the target variable, set to 1 if it is an iceberg, and 0 if it is a ship. This field only exists in train.json.