Iceberg Detection using Sentinel-1 Satellite Images

Matthew Lee

December 21, 2017

1 Domain Background

This project is based on the Statoil/C-CORE Iceberg Classifier Challenge, posed on kaggle.com. The challenge aims to improve the detection of icebergs using satellite imagery. Icebergs present a significant danger to shipping, especially in adverse weather conditions. To mitigate risks caused by icebergs, many operators use shore based techniques and aerial reconnaissance. However, in remote locations or during harsh weather conditions, other approaches must be used. This is where satellite imagery can help. The Sentinel-1 satellites use imaging techniques that are able to penetrate through clouds and can operate at night as they are their own lightsource. However telling ships apart from Icebergs is difficult as the images do not have colour and are have a low resolution of of the images produced.

The radar takes two images, with different polarisations, Horizontal-Horizontal(HH) and Horizontal-Vertical(HV). HH means the image was formed by transmitting light with a horizontal polarization and receives light with a horizontal polarisation. HV means transmitting with horizontal polarisation and receiving on vertical polarisation. The polarisation of light backscattered to the satellite depends on the material, polarisation of transmitted light and the angle of the reflecting surface. This means that by using both polarisations makes classification of ice vs water easier. REFERENCE YU HERE.

I was interested in this project because of my previous experience with polarisation and satellite imagery from my Physics degree. I'm really interested in the application of satellites and am quite suprised that satellite image classification is not already more widely used, given the considerable investment involved to transport it into space and our inability to upgrade existing satellite hardware. This means that work to upgrade the software, to extract more value from these satellites is especially important.

2 Problem Statement

The problem to be solved is one of image classification. Does a given sentinnel-1 image contain an iceberg? The model that solves this problem must have a high accuracy in order to be trusted to handle the important task of guiding ships. The metric used for this is discussed in the Evaluation Metrics section. The problem is repeatable because there is constant shipping across the atlantic, with over 10 billion tonnes being shipped globally each year. [**?**]

3 Datasets and Inputs

The dataset is provided in a JSON list with the following fields for each entry in the list:

- ID: The ID of the image

- band_1: The flattened image data in a list. Images are 75x75 so the flattened list has 5625 elements. band_1 is for the HH polarisation.
- band_2: The same as band_1 except this image is of the HV polarisation.
- inc_angle: The incidence angle that the satellite was at when the images were taken. This can be NA so some preprocessing nay be needed.
- is_iceberg: This field only exists for the training set and indicates whether the object is an iceberg. ...

There are 1604 samples in the training set and 8424 in the test set. The test set includes some machine generated images to prevent hand scoring.

4 Solution Statement

I intend to use a convolutional neural network to solve this problem. I will experiment with transferred models such as TODO: name this right. vgg16 model.

5 Benchmark model

The ¡insert reference¿ paper achieved precision of ¡97¿ TODO check value and if that's a representative benchmark for this problem set. perhaps take the median kaggle solution's result? Or is there a previous paper I can use as a benchmark?

6 Evaluation Metrics

The Kaggle competition measures the performance of a given model in achieving this based off of its log loss. As there are only two classifications for an image; has an iceberg or does not have an iceberg, the classification is binary and the log loss equation is simple. The log loss equation for calculation INSERT REFERENCE FOR EQUATION EVEN THOUGH I SIMPLIFIED IT

$$logloss = -\frac{1}{N}\sum_{i=1}^{N} y_{i1}\ln(p_{i1}) + y_{i2}\ln(p_{i2})$$

Where N is the number of images classified, y_{i1} is 1 if image has an iceberg and 0 otherwise, y_{i2} is 0 if image has an iceberg and 1 otherwise, p_{i1} is the predicted probability of image i having an iceberg, p_{i2} is the predicted probability of image i not containing an iceberg.

This equation sums the natural logarithms of the incorrect predictions and divides by the total number of predictions. This means that the log loss is reduced by improving accuracy. This is interesting as I would have expected the competition to have valued recall most, as a ship wrongly classified as an

iceberg poses no danger to another ship, where as an iceberg wrongly classified as a ship could be extremely dangerous. For the purposes of this project I will use Kaggle's logloss equation as the main metric, however I will still measure recall as it will be an important metric in real world scenarios.

7 Project Design

Your text goes here