Deep Learning for Wave Height Classification in Satellite Images for Offshore Wind Access

UNIVERSITY OF HULL

R.J.Spick, J.A.Walker

The School of Engineering and Computer Science, University of Hull,
Cottingham Road, Hull, Yorkshire HU6 7RX

Objectives

How can we create a Deep Learning model that can be used to classify areas that may contain little or no information on the state of the sea?

- Create a framework for processing of satellite images.
- Utilise methods to remove and clean satellite images of all unnecessary data (clouds/land etc.).
- Train a Deep Learning model to learn the representation of sea height in satellite images.
- Visualise and test results on unseen data.

Introduction

Satellite images provide a robust and high resolution view of our world. Large amounts of work have been completed regarding the application of Deep Learning to land satellite data such as predicting poverty [1], however very little has been investigated on the uses for offshore applicability.

Supervised Deep Learning requires an input and a corresponding label for that class, the network then attempts to learn a representation of the input data usually as a discrete class. Once a network has been trained the learned representation can be used for future classification problems [4]. Combining this method with a single dense final layer gives the ability to output continuous values, in this case corresponding to the wave height.

The scope of this work is to provide an alternative way of discovering the sea state on a given day, providing higher detailed information for sea that is within a several hundred kilometers of the shore line and away from other measuring tools. Specifically aiming at the access of offshore wind farms. Through several objective steps, that will be discussed in more detail, we have created a tool that receives a 10k resolution Sentinel-2 satellite image; pre-processes the image in to a usable format and, using the pre-trained model, outputs a predicted sea height value for the input image/area.

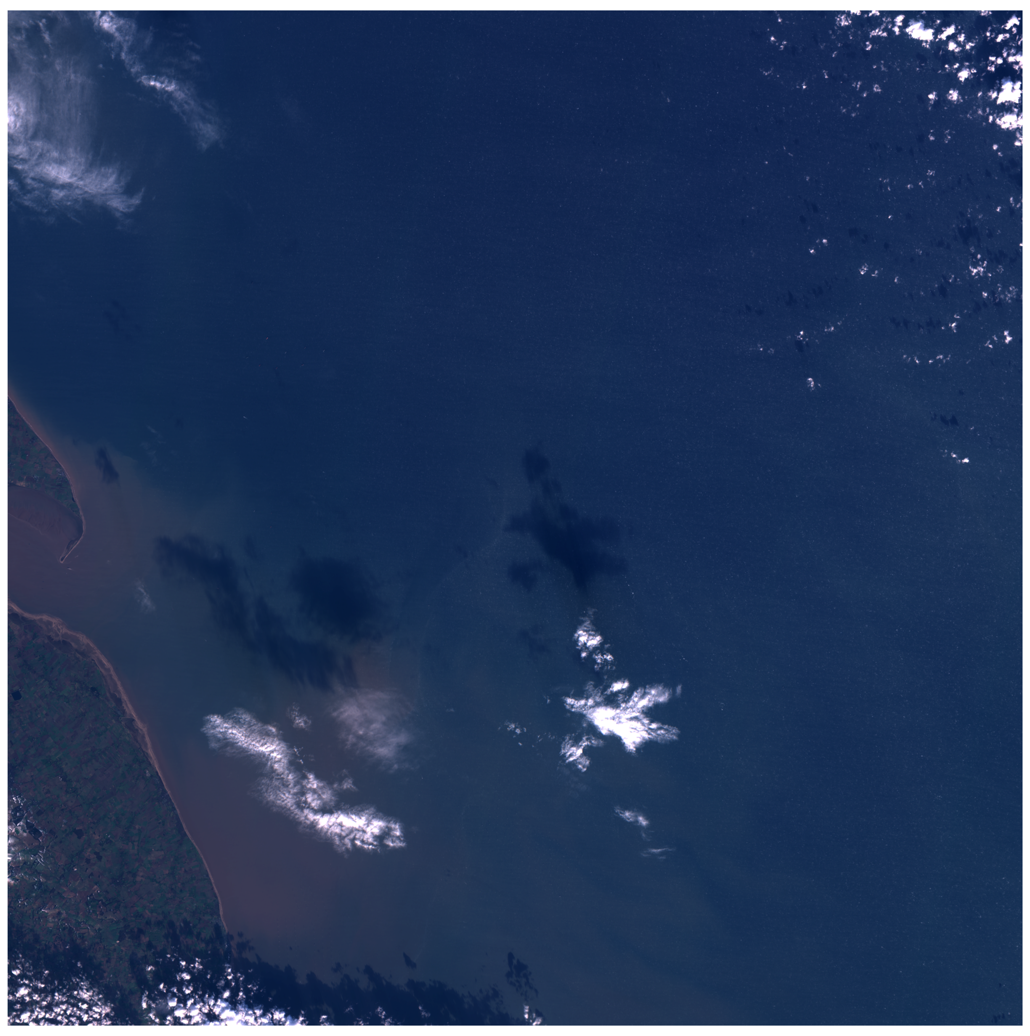


Figure 1: Example Sentinel-2 RGB Image off the Coast of Hull

Early Methodologies

RGB satellite images are very rarely produced in a clean and usable format from source. Figure 1 shows a good-case scenario of a partially cloudy image containing some land. Therefore, before we could proceed with the main portion of the experiment we required a method for the robust removal of any anomalous data.

The framework developed focused on analysing sub-sampled images of the main image, each 10,800 x 10,800 resolution input image was divided equally split in to 2025 samples of 244 x 244 pixel resolution. The sub sampling is required for a more accurate training model, artificially providing more data points to train our learning model on. We also chose to down sample our images for the process of segmenting the image for removal of those divided images that contained cloud or land.

Satellite Image Cleaning

The removal of atmospheric anomalies in satellite imaging is a key part in most pre-processing steps for the use of satellite images. Most methods are not robust enough for the level of accuracy we require or depend on large levels of manual interaction [5], and so we will be constructing a separate Deep Learning model for the segregation of the cloud.

To construct the data set needed we first took 30 input satellite images from various positions around the north sea, each image was divided in to a sub-sampled set leaving a substantial data set to work with of 60,750 images.

For these images we used a manual filtering approach in order to construct a labeled database of two classes, one for images we want the model to remove/throw away and one for the pure sea images that we wish to keep.

A method of adaptive thresholding [2] between neighboring pixel values was chosen, leaving similar neighboring pixel colours white, whereas the massively different cloud and land pixels will be set to black.

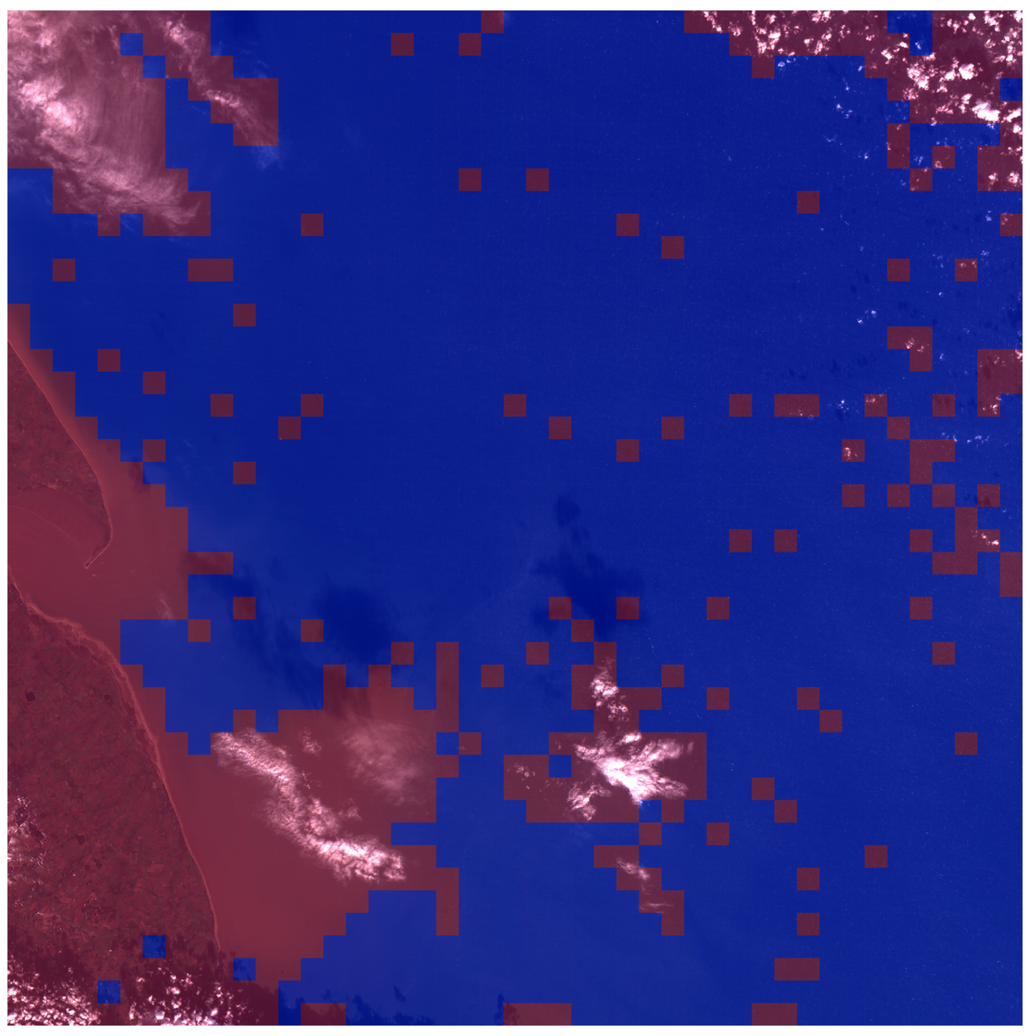


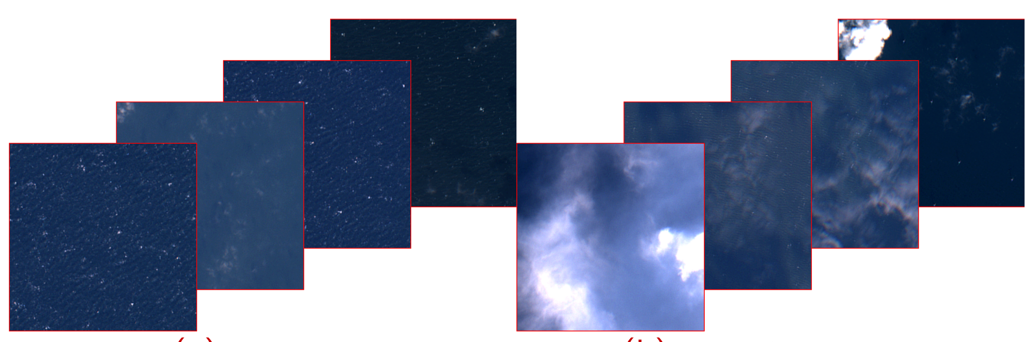
Figure 2: RGB Image that has been classified using our neural network. Red representing anomolous data and Blue representing pure sea data.

We fed these manually split images in to a Deep Convolutional Neural Network [3] which learned a representation of the two classes, this allowed the future use of this model instead of the manual based approach. Figure 2 shows the analysis and output of one trained image.

Learning the Sea Height

Using the previous methods of separating the sea sub-sampled images from the input satellite image we are left with a vast data set of images similar to Figure 3 (Sea Class). To determine the sea height given an input we initially need some label data for the purpose of grouping our satellite images. We specifically selected a region of the ocean that we had access to buoy data for, this ended up being just off the coast of Hornsea, we used the Dowsing-wavenet data. This allowed the pairing of each satellite image in to a class of sea height i.e. (0.8,1.3,1.4 ... 2.1,2.4).

Using a Deep Convolutional Neural Network [3] the model was trained on 15K purely sea images with their representative labels until we reached a loss of 0.07 across the learned model. The network consisted of a pre-trained VGG16 network with the last dense layer fine tuned on the training data [6].



(a) Sea Class (b) Cloud/Land Class

Figure 3: Example of a set of images from the two classes that have been segregated

Test Accuracy & Analysis

The network was tested on unseen satellite input images, these images were picked from random points and time where we had access to buoy data to determine how accurate the model was. We utilised part of the framework preprocessing to clean these input satellite images and split them in to smaller sea images. The smaller images were passed through the network and individually classified with a sea height, at the end of the testing we produced both an average and median sea height value for the entire satellite image (Figure 4).

After testing on 10 satellite images we measured the accuracy of the model as the average percentage difference between the actual sea height from buoy data from the median predicted values. The model achieved an average loss across the test set of 0.17 meters, this can be seen as the averaged loss line on the histogram Figure 5.

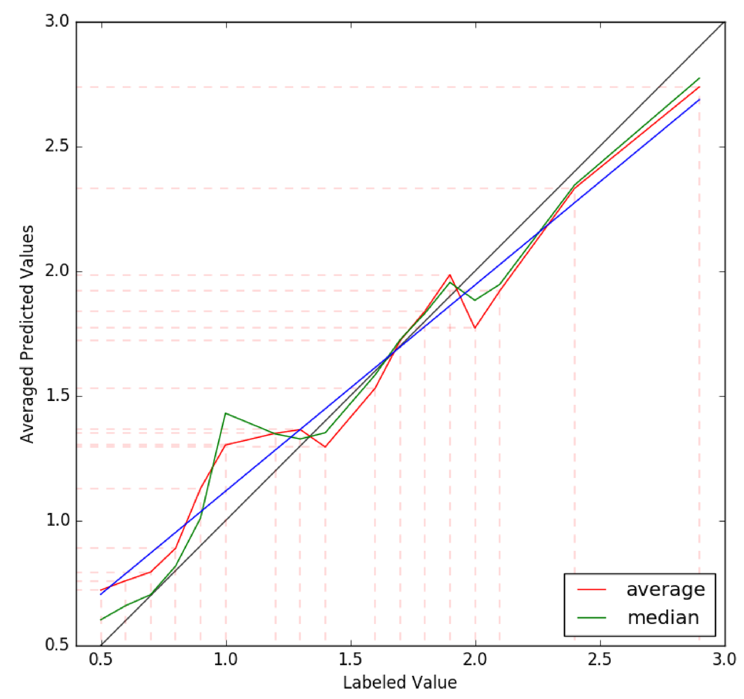


Figure 4: Model's Labeled Inputs vs. Learned Representations

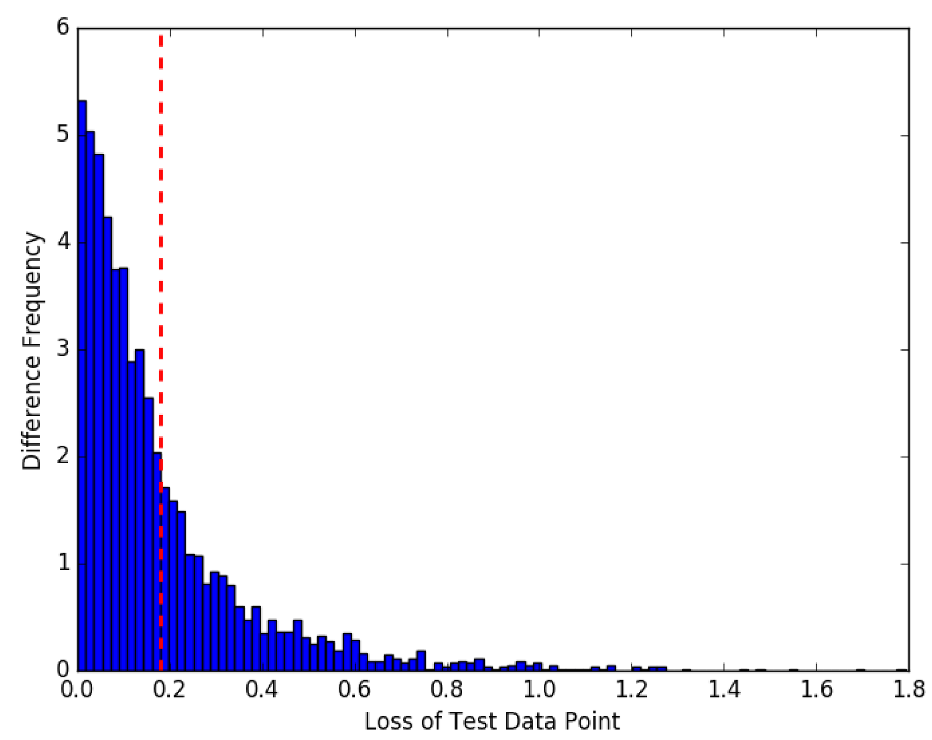


Figure 5: Histogram of Test Data Loss with Average Loss Plotted

Conclusion

The problem of determining sea height for vessel transport in areas that have little access to data can be expensive and tedious. With access to offshore wind farms needed at a moments notice for repairs. We have created a model that can take satellite data and predicted the sea height to an average of 0.17 meter difference for that specific time, this method could be deemed very useful with the upcoming wind farms that space several hundred kilometers (Hornsea Project One, 407km^2).

References

[1] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[2] Xiaoyi Jiang and Daniel Mojon. Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):131–137, 2003.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[5] Chao-Hung Lin, Po-Hung Tsai, Kang-Hua Lai, and Jyun-Yuan Chen. Cloud removal from multitemporal satellite images using information cloning. *IEEE transactions on geoscience and remote sensing*, 51(1):232–241, 2013.

[6] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.