Automatic identification of ships on satellite images

Final Project (DRAFT)

**Zhaoxiong Chen**

Keyword: Imagine processing, machine learning, deep learning

**Abstract**

**Introduction**

**Background**

With the development of economy and commercial, there can witness a dramatic growth on the transportation, especially on the international shipping traffic. Indeed, this is beneficial to the cargo industry, but this brings series of challenge to this field. The increased number of ships may directly lead to a burgeon of infraction manner at public sea including ship accidents that causing environmental contamination, piracy in Gulf of Aden and illegal fishing of whales. These matters do great harm to the order of the transportation. Although, more and more satellite has been assigned to monitoring, unfortunately, the data are not fully exploited. There is still huge number of knowledges to be discover.

Goal Description

The goals of this project are:

* Develop a system that identifies the status of the ship through the satellite image;
* Generate some statistics figure that review the traffic status from a clearer view using visualization techniques;
* Make some comparison between different data science method.

**Design and Implementation**

**Data Used**

Dataset used in this project are provided by Airbus, the giant aircraft manufacture. They are published at kaggle.com, an online community of data scientists and machine learners owned by Google. Apart from these datasets, no more data will be collected for this project.

The dataset has 4 parts: training set “train\_v2”, testing set “test\_v2”, pixel encoding file of ship edge “train\_ship\_segmentations\_v2.csv” and files for submission “sample\_submission\_v2.csv”. There are 192,556 items in training set while 15,606 items in testing set. In the encoding file we could see two columns “ImageId” and “EncodedPixels”.

By using Python image processing library “scikit-image”, attributes of the training set and testing set are collected and logged in two CSV files “train\_scikit-image.csv” and “test\_scikit-image.csv” respectively. Explored attributes for each image are shown in Table 1.

Table 1 Analyzed attributes of dataset

|  |  |  |
| --- | --- | --- |
| Attribute (Label in CSV file) | Description | Value |
| Width | Width of image | 768 |
| Height | Height of image | 768 |
| #Channel | Number of channels | 3 |
| #Pixel | Number of pixels | 1769472 |
| MaxPixel | Maximum pixel value | Varies from 0 to 255 |
| MinPixel | Minimum pixel value | Varies from 0 to 255 |
| AveragePixel | Average pixel value | Varies from 0 to 255 |
| Type | Number of bits | 8-bit unsigned integer |
| Extension | Type of image | .jpg |
| Size | Image size (in Byte) | Varies from each image |

**Method Discussion**

**Principles**

**Results and Discussion**

**Conclusion and Further Work**