# Using Data Science and Remote Sensing to Understand Ship Traffic with Application to Detecting Forced Labor (Human Trafficking)

# Draft Report by Thomas Keeley & Harry Newton

# Change Log (Weekly Activity)

This Week

* Obtained Joint AIS & GIS data from merging AIS data from Global Fishing Watch (they had to adjust the permissions for the dataset) and Sentinel 2 data (same data as previously, but using an index and source from Google that has more of the images available).
* Completed MVP2 in Google BigQuery to select co-incident signals for each cloud-free image in December 2018.
* Tested Object detection on the Joint data for an image with many co-incident signals.

Next Week:

* For DV model, meet with expert on Human Trafficking
* Finish Report & two video(s)

# Abstract

By using Machine Learning (ML) on readily available imagery and signals collected by satellites, the presence of ships and their activity can be partially understood. In this report, we first showcase our work to improve an ML algorithm that operates on imagery based to identify the location of ships. These results beat the winning submission for a Kaggle competition to do so. We tune this algorithm with training data based on assuming the ground truth by the observation of a navigation signal detected at the same location and time. This navigation signal is the Automatic Identification System (AIS) that most ships emit continuously to comply with international agreements designed to avoid collisions at sea or in port. The imagery and signals data that we used are described in Table 1. We describe this first model as our **Ship Imagery Model** because once tuned, it only requires imagery to identify ships.

We develop a corresponding Ship Signals Model to summarize voyages that can be discerned by AIS data alone, then compare the results for “scenes” where we have corresponding imagery and signals at nearly the same time. We use this Differences Model to find where the location of ships in time and space does not align in the data.

We conclude with an application of the models to the Human Trafficking problem of forced labor crews and propose a risk model and data visualization to highlight ships that have long voyages and gaps in their signals data. This application, like several similar ones identified in the literature review below, showcases the value of combining data from multiple sources.

# Problem Statement

1. Develop a Machine Learning (ML) model to identify ships in imagery data  
2. Develop a pattern analysis Data Visualization (DV) based on navigation signals data from Spire and Global Fishing Watch  
3. Combine these models and data to improve the ML model    
4. For an application of these models and data, propose a risk assessment framework for Forced Labor sometimes present in the fishing industry.

Models like the ones we’ve developed can provide insights into the activity of ships conducting fishing, transport, mining, etc. There are efforts to use data science for each of these areas. <Need to refer to lit review to back this up or avoid the statement>. Of these, we chose to do further work to show how our models could be applied to the fishing industry, particularly to identify practices that could indicate risk factors associated with a ship that may have a crew of forced labor (a type of human trafficking). < Need a much smoother treatment of this once.>

# Literature Review

The research presented in this paper builds on recent Big Data analysis undertaken on data on ship registrations, satellite data, and national fishing activity databases.

The application at the end of the paper to Human Trafficking and Force Labor builds on several efforts by international organizations to combat this human rights violation. Among those is the Associate Press which published a series of articles under the heading “Seafood from Slaves” for which they won the Pulitzer Prize for Public Service on April 18, 2016. The articles and actions generated by them freed 2,000 slaves, according to the AP Press Release that day. Of these articles, McDowell, Mendoza, and Mason [2015] reports on work by Kroodsma, Miler, and Roan [2017] to use AIS data to deduce likely rondevdous of smaller fishing vessels (which may have slave crews) with transshipment ships. Their work generated a global map showing these likely transhipments and analytic insights on the countries of registration for both the fishing vessel and the transhipment ship, as well as the next port visited to presumably offload the cargo. Many of these transhipments were in or near areas with fishing restrictions in place; therefore are widely viewed as likely violations of the fishing laws.

A few of the fishing ships uncovered in this research have been proven to Human Trafficking violations and their crews have been freed. The series of AP article series mentioned above chronicles these wonderful examples in intervention, as well as paints the stark truth that this addresses a very small fraction of the alleged cases. These articles also describe the difficulty that nations have enforcing their laws and show cases where if one country cracks down, the offending ships simply change their sea of operations or their registration, and so elude the authority’s jurisdiction.

This research was sponsored by the Global Fishing Watch (GFW). GFW is an independent, international non-profit organization. Charter members include Google which provides their analytic platform and SkyTruth, who provides the satellite data.

# Outline of the Paper

1. Improved results for Kaggle competition with Ship Imagery Model.
2. Ship Signals Model
3. Combined Model
4. Application to Forced Labor detection.

# Methodology

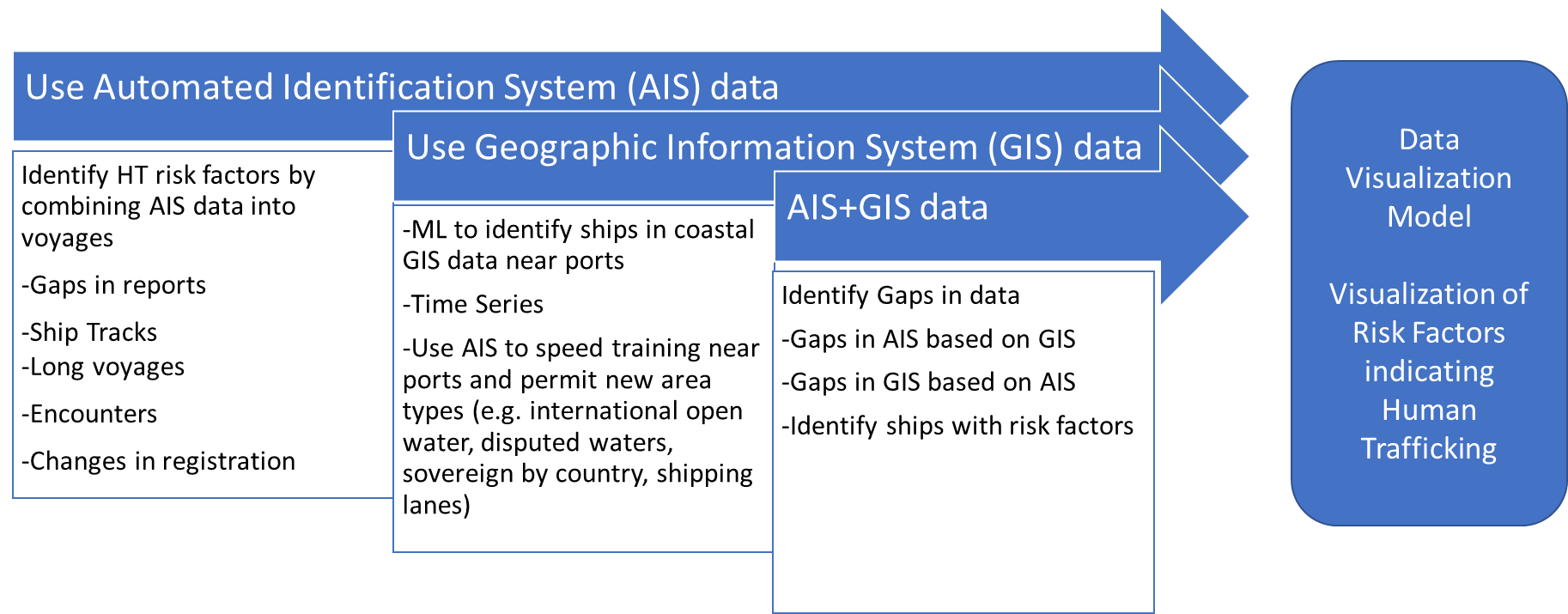
The research presented in this paper is based on Machine Learning to identify ships based on satellite imagines and Big Data Analysis to operate on satellite signals data. We employ both of these sensor modalities to improve these techniques iteratively then also used them in combination to discover insights about the ship traffic. We conclude with a linear model to analyze the data against risk behaviors and present the results using Data Visualization in the form of a Sankey Chart. This methodology is summarized below.

Figure 1 Overall Methodology

# Data Sources (replace!)

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Improved results for Kaggle competition with Ship Imagery Model

The application of object detection in high resolution satellite imagery has been explored extensively over recent years. The targets of this type of analysis varies across domains but generally aims at enhancing the ability to process large volumes of satellite imagery data to locate objects of interest. Target objects typically involve vehicles, roads, buildings and vessels. The integration of Computer Vision and Deep Learning frameworks with satellite imagery analysis has yielded innovative results in automatically detecting objects with great accuracy. The object detection application presented in this report focuses on leveraging Geographic Information System (GIS) software tools and Deep Learning models to develop a framework workflow capable of detecting vessels in satellite imagery.

The innovation of vessel detection in high resolution satellite imagery over recent years has produced very high performing pre-trained models that can be deployed within a user’s environment and applied to a personal use case. Though these high performing models produce benchmark results, they also require a significant amount of computing resources. The application presented in this paper will instead provide the capability to develop and deploy a simpler, more lightweight model that still produces accurate results. The user of this application will be able to either produce their own training data or import from another source, develop and train a Deep Learning model using the Keras framework, deploy the model to run predictions on desired imagery, and finally produce an object detection output with geographic attribution. This type of capability within GIS frameworks has been developed in proprietary software such as ArcGIS. However, the ability to conduct object detection is currently limited in open source GIS software such as QGIS. This application presents the capability to perform this type of analysis on an open source platform with minimal Deep Learning understanding using limited computing resources.

# **Ship Signals Model**

Model 1 & measures.

# **Combined Model**

<add later>

# Application to Forced Labor Risk Factors

<Harry to update…this version is based on original direction which has shifted>

According to Liberty Shared, data on Forced Labor both domestically and internationally is practically non-existent. This is not for a lack of concern or motivated NGO’s who are willing to collect it, but rather the lack of an appropriate way to collect and curate the data. A mobile-based data collection/analysis tool could be created to complement the tool that Liberty Shared currently offers for victim case management. This tool could take the form of an add-in to OpenStreetMap or QGIS and would interface with their Victim Case Management System, which is based on Sales Force. If the idea of this mobile data collection add-in is acceptable to GWU, then we’ll talk with his Chief Technology Officer next week. We’d start with a list of requirements, including multi-language support and ease-of-use. Duncan noted that this would be extremely useful for collecting data across the globe on Forced Labor which is currently under-served, and while they have volunteers to collect this data, the data quality and organization does not provide a legitimate data resource. He has a current effort in Kenya that could help with usability testing.

Traditional wisdom is that Forced Labor is generally aligned with more traditional commerce, so at the end of the supply chain, the goods are sold using traditional currencies to customers and companies that operate in the open and are willing to stop buying a product or service once a legitimate claim is made about forced labor. This stands in stark contrast to sex labor which is often has ties to organized crime and crypto currencies. A reference on this is the recent report by the Financial Threats Council of the Intelligence and National Security Alliance (published May 2020) entitled “Using Intelligence To Combat Trade-Based Money Laundering.”

# Schedule

|  |  |  |
| --- | --- | --- |
| Task | Duration | Due |
| 1. Compare | 2 weeks | 6/30 |

# Analysis of Results

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Future Direction

Include any deficits in our work, such as bias to where we can find the data from both sources…

# Rubric – <This was for the proposal. Replace with one for Final Report>

|  |  |
| --- | --- |
| • What problem did you select and why did you select it? | See problem statement above |
| • What database/dataset will you use? | See table of data sources above. |
| • What data science technique will you use to solve the problem? | Machine Learning, Data Analysis using Pandas, Data Visualization (Sankey) |
| • What framework will you use to implement your work? Why? | Google? |
| • What reference materials will you use to obtain sufficient background on applying the chosen method to the specific problem that you selected? | Lit Review of data science projects and provided by subject matter experts on HT. |
| • How will you judge the performance of your work? What metrics will you use? | Against UT-SA results on Polaris. By request for Peer review from CINA at GMU.. |
| • Provide a rough schedule for completing the project. | See schedule table above. |

Save for possible bias of this research for under-reporting.

# References

### Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

<https://arxiv.org/pdf/1506.01497v3.pdf>

Object Detection with Deep Learning on Aerial Imagery

<https://medium.com/data-from-the-trenches/object-detection-with-deep-learning-on-aerial-imagery-2465078db8a9>

### A systematic study of the class imbalance problem in convolutional neural networks

<https://arxiv.org/pdf/1710.05381.pdf>

### Automated vehicle detection in satellite images using deep learning

<https://iopscience.iop.org/article/10.1088/1757-899X/610/1/012027/pdf>

### Measuring Human and Economic Activity from Satellite Imagery to Support City-Scale Decision-Making during COVID-19 Pandemic

<https://arxiv.org/pdf/2004.07438.pdf>

### Automatic Target Detection in Satellite Images using Deep Learning

<http://www.ist.edu.pk/downloads/jst/previous-issues/july-2017/automatic-target-detection-in-satellite-images-using-deep-learning.pdf>

Kroodsma, D.A., N.A. Miller, and A. Roan 2017. “The Global View of Transshipment:

Revised Preliminary Findings.” Global Fishing Watch and SkyTruth, July 2017. Available online at <http://globalfishingwatch.org>.

“An AP investigation helps free slaves in the 21st century,” a series of articles over 18 months for which they won the Pulitzer Prize for Public Service on April 18, 2016.

McDowell, R., M.Mendoza, and M. Mason. 2015. AP tracks slave boats to Papua New Guinea

https://www.ap.org/explore/seafood-from-slaves/ap-tracks-slave-boats-to-papua-new-guinea.html

Greenpeace. 2016. Turn the tide. Human Rights Abuses and Illegal Fishing in Thailand’s Overseas

Fishing Industry. <http://www.greenpeace.org/australia/Global/australia/reports/Turn-The-Tide.pdf>

[21] Cutlip, K. 2017. Skytruth Blog. http://skytruth.org/2017/01/satellites-leave-no-place-to-hide-forrogue-thai-fishing-fleet/