Computer Vision for Detection of Illegal Mining Barges in the Rio Madeira

Julian Lee[1*] , **Eric Lin**[1*] , **Maggie Wang**[1*] and **Sayak Maity**[1*]
[1] Harvard College
[*] Equal Contribution
{slee5, eric_lin, maggiewang, smaity}@college.harvard.edu

Abstract

In recent years, computer vision analysis of satellite imagery has been used to identify large ships in both offshore and inland settings, as well as to combat deforestation in the Amazon Forest. However, the presence of mining barges conducting ecologically destructive gold dredging operations in inland rivers remains a major problem relatively unaddressed by artificial intelligence. Challenges in this domain include the small size of the barges in question, their rapid movement patterns, and inconsistent satellite image availability, leading to high difficulty in detection. In our paper, we present three key contributions: (i) a self-curated medium-resolution satellite image dataset ($n = 296$) of small river watercraft, (ii) a temporal image differencing algorithm using radar data to detect watercraft movement under cloud coverage, and (iii) a convolutional neural network model trained using weighted cross-entropy loss that achieves 90.0% balanced accuracy and 100.0% TPR on a small held-out portion of our dataset. To the best of our knowledge, our work is the first application of artificial intelligence to identify small inland watercraft, and our dataset is also the first of its kind.

1 Introduction

Areas in the Amazon River Basin such as the state of Rôndonia in Brazil are rich in gold [1]. Gold prices have increased over 300% in the past two decades, and have recently gone over $1,700 per ounce. This has caused gold mining to be both lucrative and widespread.

Dredging of the Amazon River by illegal gold mining barges causes significant ecological damage through contaminating rivers with mercury, which negatively impacts local wildlife. Although there are occasional police patrols in the area, the Brazilian government's stance is that the large geographical region makes it too difficult to police the illegal miners. This serves as the motivation of our project, as identifying illegal mining operations from satellite imagery can narrow down search areas and accelerate policing.

Curated, publicly available satellite imagery datasets currently do not exist for inland rivers and barges; existing artificial intelligence applications focus on classification of large ships in open water. We thus present our solution to the problem of detecting the presence of an illegal gold mining barge in the Amazon River based on satellite imagery. This will provide the World Wildlife Foundation (WWF) with data-driven insights into the extent of the problem, and enable more targeted interventions from police and local communities against illegal mining.

Our main contributions are the following:

- We curate the first-of-its-kind publicly available inland watercraft dataset [2] for future work in detecting barges from optical data. In previous work, these datasets only contained large boats in offshore open water.
- We utilize deep learning models on optical data, as well as the MASATI dataset for transfer learning, to create a novel barge identification system that can be used for local enforcement.[1]
- We use spatio-temporal differencing of radar data to complement our optical approach and reliably detect the movement patterns of watercraft in the Amazon river under variable weather conditions.

Figure 1: An example of a barge in the Amazon river. Gold mining in the river is harmful to the ecosystem because of the mercury contamination it produces.

[1] Please view our code at our GitHub Repo

2 Related Work

Advances in imagery and object detection models have led to increased research on using satellite imagery for watercraft identification. Current literature largely focuses on using either synthetic aperture radar (SAR) or optical remote sensing images for ship detection. However, these techniques are not generalizable to trickier environments, such as smaller watercraft in inland river areas.

SAR images have been preferred in past research due to their capabilities of seeing through clouds and independence from sunlight, making them work in all weather conditions at any time of the day. Previously, high-quality SAR images were hard to curate, leading to techniques working with simulated images [3]. An increase in the number of satellites capable of SAR imagery led to more widely available SAR datasets and more advanced models [4]. Then, ship detection was accelerated with the advent of deep learning [5]. Kang et al. showed in [6] how modifying the Faster R-CNN architecture with the constant false alarm rate (CFAR) algorithm led to increased accuracy for smaller-sized targets. This was improved further by [7] using YOLOv2 models.

Limitations on SAR image quality led to shortcomings in more complex marine areas with shores, islands, and uneven surfaces. To address this, optical remote sensing images, with much higher quality compared to SAR, were employed. Earlier algorithms using 2D object detection [8] were improved by threshold segmentation [9] and supplanted by the same Faster R-CNN models used for SAR imagery [10]. Optical imagery does lack the same benefits as SAR imagery in that clouds and night time greatly degrade the use of images.

However, this previous work has focused on ship detection in open seas or large river ports, which is very different from the small barges and inland rivers the WWF and other conservation groups work with around the globe. Analysis of satellite imagery has previously been used to target illegal gold mining in the Amazon rainforest to curb deforestation [11]. Illegal sand, gravel, and gold mining [12] occur at alarming rates around the world, yet there is no known prior research in helping authorities detect these practices using image detection.

Our work addresses this problem by presenting novel datasets and solutions for identifying small parked barges in the river, rather than large ships in the open ocean or analyses of land-based operations. By using both SAR and optical imagery, we utilize the benefits of both approaches. In addition, our collaboration with the WWF was important in shaping the specifications for this project.

3 Data & Challenges

We encountered many challenges relating to the procurement of data. First and foremost, we had to curate our own training and testing datasets from satellite image databases, because existing datasets of satellite images containing watercraft are focused on large cargo ships and not small inland barges. This already imposed a practical constraint on our dataset size, as due to time and labor constraints as well as the geographical sparsity of barges, the number of images that we could realistically source, process, and label was extremely limited.

In the process of data curation, we were further constrained by the competing considerations that our satellite image sources had to be high-resolution enough to display mining barges, which range from 50m to just 10m in length, but also available at low-cost for an unlimited length of time so that our solution would be sustainably usable by monitoring groups like the World Wildlife Fund and indigenous peoples. Moreover, we required a high temporal frequency of imaging in order to differentiate between moving watercraft, which are more likely to be innocuous, and watercraft parked in one location over multiple days, which are much more likely to be conducting illegal mining operations. Robustness to cloud coverage was a final major issue, as cloud coverage can render several consecutive days of images unusable for analysis.

Our proposed solution uses three different sources of data to cumulatively address all of these challenges. An overview of the approaches we used can be found in Figure 2.

3.1 Planet Labs Optical Data

Planet Labs is a commercial satellite image provider that provides images of 3m / px resolution using its PlanetScope satellites - we were able to access these images for free under its Education and Research program. This represented an appropriate balance of low cost, acceptable resolution, and daily revisit time to our Area of Interest (AOI).

Using our access to the Planet Labs database [13], we curated a dataset [2] of 296 square image tiles of the Madeira River in Rondônia State, ranging from 224x224px to approximately 300x300px. This dataset consists of 100 samples manually labeled as ‘positive’ - meaning the tile contains a barge - and 196 samples labeled as ‘negative’. Collecting positive samples was challenging because of the geographical sparsity of barges. Downloading large swathes of river imagery and then labeling afterward was infeasible, as this would have led to a class imbalance skewed towards the negative by hundreds or thousands of times.

Instead, we manually searched through the image database for positive samples via visual inspection, individually downloading, square-cropping, and processing each one. We took the same approach for negative samples, which were considerably more abundant. For our dataset of negative samples, we represented the wide range of non-boat-containing images which the model might encounter in the real world, including sediment deposits and clouds which could be mistaken for boats.

3.2 MASATI Pre-Training Data

The MASATI (MAritime SATellite Imagery) dataset [14] consists of optical images of maritime scenes, with image class divisions including with/without ships, and with/without coastline. This data is originally obtained from Microsoft® Bing Maps. The typical image in the dataset has a resolution of approximately 512x512 pixels.

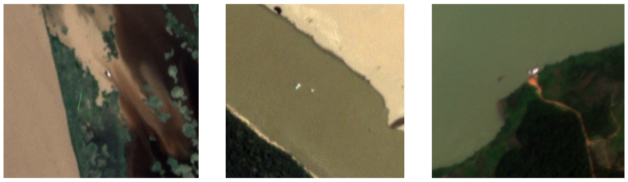


Figure 3: Selected Positive Samples from Planet Labs Dataset

Figure 4: Selected Negative Samples from Planet Labs Dataset

3.3 Sentinel-1 Radar Data

Sentinel-1 [15] is a constellation of two satellites launched by the European Space Agency. It collects radar data at a resolution of 10m / px, and the frequency of coverage over our Area of Interest in Brazil is approximately once every 2.8 days. The data is freely available via Sentinel Hub. Although the resolution of the radar data is lower than that of the Planet Labs optical data, radar data has some salient advantages that address the earlier outlined challenges. First, the data is never obscured by cloud coverage, which provides significantly increased reliability. Second, unlike optical images, radar data collected at night is equally informative, which is extremely useful for this application because illegal gold mining barges often operate under cover of dark. We thus use the consistent data inflow to provided by Sentinel-1 to track the movements, or lack thereof, of suspicious watercraft identified using higher-resolution optical imaging.

4 Evaluation

4.1 Radar Approaches

We use a single polarization (VV), which offers high contrast between the ships and water. We then mask the image using the a GeoJSON of the Rio Madeira outline, which allows us to find temporal differences between two images only in the river regions we are interested in, as shown in Figure 5c.

Using the masked radar image from two different dates, we take a temporal difference between the recent date we are interested in and a later date. The temporal difference allows ships and barges in the more recent date to have a higher pixel value. We deploy a binary threshold in OpenCV to extract only the pixels with high value where the ships and barges are most likely to be.

We then use a Gaussian blur and blob detector to filter the detections by area, as shown in Figure 6a. We get the keypoint location and radius from the blob and draw the keypoint on the initial radar image, as shown in Figure 6c.

We found that there were no false negatives found in the time interval of interest, but we saw multiple false positives due to an imperfect GeoJSON outline of the river. These false positives can be seen in Figure 7. Since the WWF teams in Brazil will be verifying the final output of our system, we designed our system to be more lenient towards false positives and strict towards false negatives.

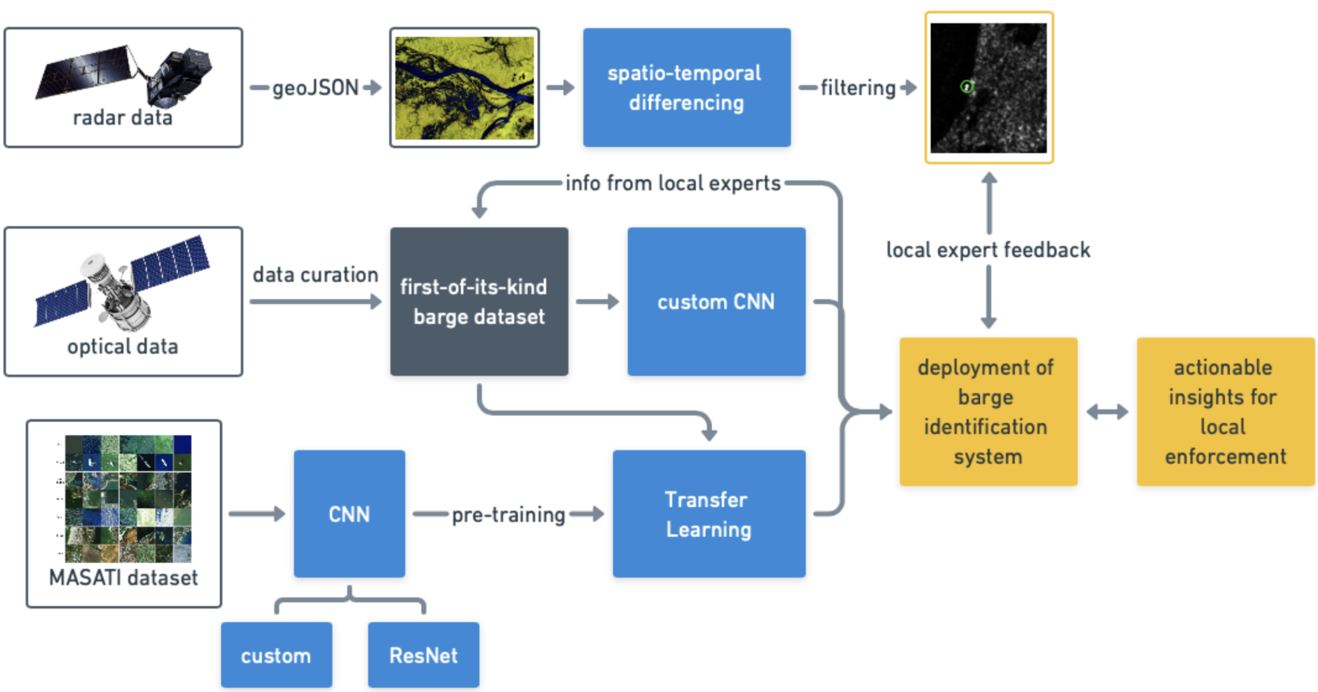
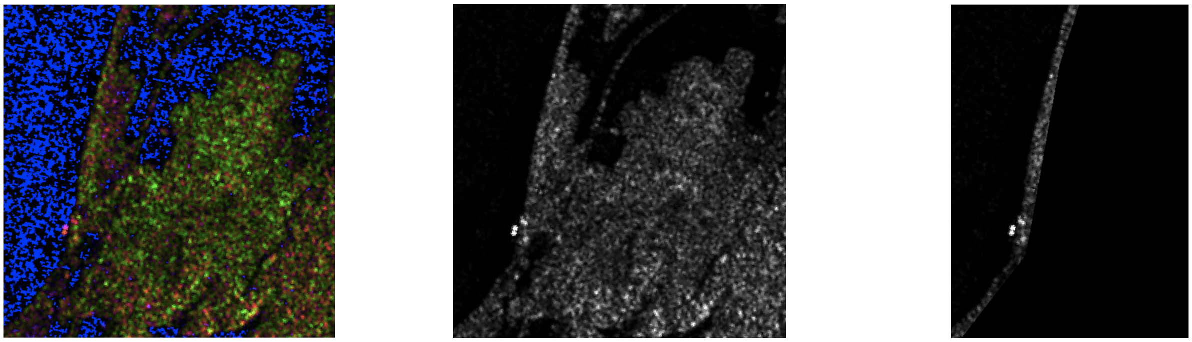


Figure 2: Overview of our approach.

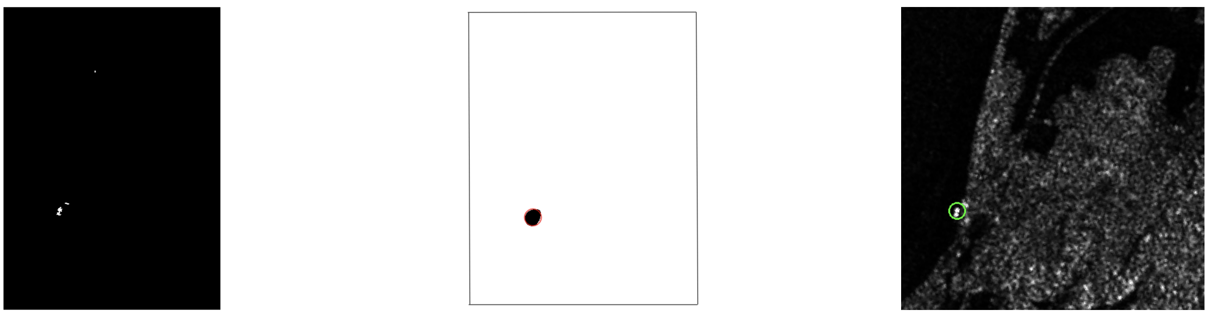


(a) The full Sentinel-1 image with four bands.

(b) Sentinel-1 radar image with a single band (VV).

(c) Sentinel-1 radar image (VV) with masking using the GeoJSON of desired locations of the Amazon River.

Figure 5: The Sentinel-1 data processing pipeline, where we start from a single image band and use a GeoJSON mask to focus only on regions of interest.



(a) We take a temporal image difference between the recent time period we are interested in and an earlier time period. This gives us higher pixel values where there are emerging objects of interest in the recent time period. We then use a binary threshold to remove extraneous values.

(b) We use a Gaussian blur along with a simple blob detector in OpenCV to filter the objects of interest by area. This allows us to filter for barges and ships in the river.

(c) After we find the keypoint location and radius of the blob, we overlay the final detection point on the initial Sentinel radar image to verify the location of the ship or barge.

Figure 6: Temporal difference and filtering steps in the Sentinel-1 pipeline.

(a) False positive found by the spatial-temporal difference method. This false positive occurred because the GeoJSON does not completely cut out where the land is, and we occasionally get spurious detections on land.

(b) The circle in the middle created by the spatial-temporal difference method correctly identifies a barge. The two circles near the bottom of the image are spurious detections due to an imperfect GeoJSON outline of the Amazon river.

Figure 7: False positives found by the Sentinel-1 radar spatial-temporal difference technique. We designed our system to be more lenient towards false positives and strict towards false negatives, since we will have a person verifying the detection output of our system.

4.2 Optical Approaches

In order to work with the small size of our curated dataset, we compared a transfer learning approach to a from-scratch training approach, and surprisingly found the from-scratch training, with certain strategic modifications, to perform better than the transfer learning approach by up to 6.5%. Transfer learning was conducted by pre-training models on 4,980 large ships and coastline images of the MASATI dataset (from Bing Maps). In all experiments, we used 5-fold cross-validation on the Planet Labs curated dataset, with one fold of 60 images being held out until after model selection in order to estimate generalization capacity, and the other four folds of 59 images each being used once as a validation fold for model and hyperparameter selection.

The average of all four validation accuracies was used to evaluate hyperparameter and model settings, an approach which was especially necessary given our small dataset - we found that single train and validation cycles had extremely high variance and were very unstable depending on the train / val / test split used. In order to simulate an inflation of the dataset size, training images were augmented with random horizontal and vertical flips, a random 224x224 crop, and standard color jitters.

Weighted Cross-Entropy Loss

We incorporated weighted cross-entropy loss in order to adapt models to this domain. Weighted cross-entropy loss was originally developed to handle significant class imbalances in datasets, preventing models from neglecting small classes by disproportionately penalizing the misclassification of samples from underrepresented classes during training. In the binary classification case, weighted cross entropy loss is parameterized by a single hyperparameter, w_{pos}, indicating the loss incurred by the misclassification of a positive sample proportionate to the loss incurred by an equally confident misclassification of a negative sample. This is expressed as

$$L(y, p) = -(w_{pos} y \log(p) + (1 - y) \log(1 - p))$$

where p is the probability assigned to the positive class by the model, usually derived through applying the softmax function over the model's output vector. Standard cross-entropy loss is equivalent to the case where $w_{pos} = 1$; if $w_{pos} > 1$, we consider correct classification of positive samples to be more important than correct classification of negative samples, and if $w_{pos} < 1$, the opposite holds true.

Weighted cross-entropy was especially important in our modeling because of the low penalty incurred by false positives in the real world application of our research. In the human-in-the-loop system, a human could easily verify and dismiss tens of false positives, while a false negative would represent a significant missed opportunity to prevent illegal mining. While balanced accuracy remains an objective, a high True Positive Rate (TPR) is more important to decision-makers than a high True Negative Rate (TNR), and so our experimentation focused on tuning the w_{pos} hyperparameter towards this goal.

Custom CNN

We first trained a custom CNN architecture on our curated Planet Lab dataset. With RGB input square images of 3x224x224, the sequential model consisted of a convolutional block with 3 convolutional kernels of size 32x32, followed by 7 convolutional blocks of 3 kernels of size 128x128. The final layer is a fully-connected layer taking the outputs of the convolutional blocks and outputting a binary label. We designed this custom CNN to be small and lightweight, since we anticipated that without pre-training on MASATI, the tasks of learning binary barge classification wouldn't need as many layers as ResNet18. In Table 1, we can see experimental results for the custom model without pre-training. It achieved moderate 5-fold validation accuracy with $w_{pos} = 1.5$.

	balanced acc.	TPR	TNR
Custom, no pre-training	0.75	0.78	0.71
ResNet18, no pre-training	0.83	0.84	0.82
ResNet18, pre-trained	0.77	0.88	0.65

Table 1: Averaged 5-fold cross validation accuracies for custom and ResNet18 models with no pre-training ($w_{pos} = 1.5$) and pre-trained ResNet18 ($w_{pos} = 2.0$).

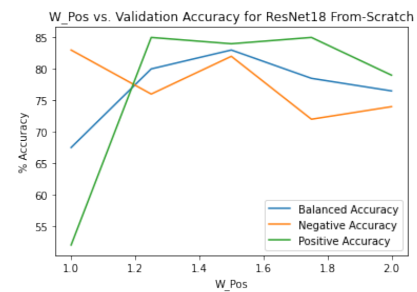


Figure 8: w_{pos} vs. validation accuracy for ResNet-18 from scratch

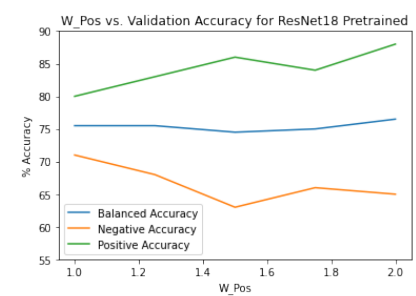


Figure 9: w_{pos} vs. validation accuracy for ResNet-18 pretrained

Hyperparameter Optimization

In speaking to WWF experts, we learned it is crucial to limit the number of false negatives. In Figures 8 & 9, we compare the effects of tuning w_{pos} for a ResNet-18 initialized randomly and with weights learned from pre-training on the MASATI dataset. We observe that tuning w_{pos} has a more significant positive effect on the randomly initialized than on the pre-trained network. With standard cross-entropy loss

($w_{pos} = 1.0$), the transfer learning approach significantly outperforms the from-scratch training, with a 75.5% vs. a 67.5% balanced accuracy. Notably, the from-scratch training with $w_{pos} = 1.0$ had a TPR of just 52%, which was equivalent to random guessing, compared to the 80% achieved by the pre-trained network.

Increasing w_{pos} appears to have no effect on balanced accuracy for the pretrained ResNet18 - as expected, we see incremental gains in TPR as w_{pos} increases, but in order to achieve these gains the model makes similarly sized sacrifices in TNR, so that balanced accuracy is almost unchanged. On the from-scratch network, however, we observe that the gains in TPR as w_{pos} increases are significant enough to outweigh the sacrifice in TNR. In particular, the TPR jumps by over 40% when w_{pos} is tuned from 1.0 to 1.25. We chose $w_{pos} = 1.5$ on the from-scratch model, which trained to 84% validation TPR and 82% validation TNR as our final setting, retrain over all four train/validation folds, and test on the holdout fold to estimate generalization capacity. The testing yielded a TPR of 100% and a TNR of 80%.

We note that the impressive testing result should be taken with caution. The small size of the test set make accuracies susceptible to variance, and the fact that test accuracy outperforms validation accuracy suggests that the test set contained a particularly 'easy' subset of samples. Moreover, although we chose training from-scratch as our final model setting in this paper because it outperformed the transfer learning approach for the majority of w_{pos} thresholds, pretraining did improve stabilization of model performance, as seen in Figure 9, where balanced accuracy remains stable over all w_{pos} settings and positive accuracy increases in a more linear and consistent fashion. Given that the improvement in validation accuracy of approximately 5% exhibited by training from-scratch only accounts for a difference of 16 images labeled correctly across the 5-fold cross validation, we would not discount the transfer learning approach either, and would encourage further exploration of transfer learning for this application, either with enlarged datasets or more model optimization.

4.3 Proposed Integration of Radar and Optical Approaches

In practice, we propose to integrate the temporal differencing and deep learning models by regularly checking the Planet Labs API for cloud-free images over our AOI. After running the images through our optical machine learning pipeline and identifying candidate tiles containing small watercraft, we can then use the georeferencing of the tile to determine whether the craft was located within the Protected Areas designated by the Brazilian government and the government of Rondônia. Finally, we would utilize our temporal differencing radar approach to verify the presence of the barge, guarding against the high false positive rate, and determine the length of time that it has been stationary. Optically identified barges that were pictured in a Protected Area and were shown to be stationary over a period of multiple days would then be flagged to partners at the World Wildlife Fund and within indigenous groups for in-person intervention. For real-world usage by the WWF researchers and local police, these tools will be built into a mobile app, since that is the most common and accessible option.

5 Ethics & Broader Impact

With the creation of new technology tools, it is also important to consider the ethics and broader impact of using the technology. First, we would want to ensure that the local governments are using the tools as we intended. Misuse in general can be mitigated through user training when onboarding to the platform. For concerns regarding surveillance, a geofence can be set in the app to only reveal satellite imagery in the relevant river areas and protected regions. In addition, the satellite imagery itself, having a maximum resolution of 3m / px, is not sufficiently high resolution to pose a threat to individual privacy. Another ethical concern is how it will affect the livelihood of the gold miners. Legally or not, miners live and work on these barges and depend on gold mining for income, perhaps to support their families and children. The disruption of their lives is a relevant social cost that should be considered as this technology is further refined and deployed. Some of the broader impacts of this project include a positive environmental impact. Less illegal gold mining in the Amazon River will translate into better ecosystem health and less mercury contamination for wildlife. Notably, since humans consume fish from the Amazon river, successful reduction of illegal gold mining would ultimately improve human health.

6 Conclusions and Future Work

Through this research, we have made two primary contributions. The first is that we have created a novel dataset for barge detection in rivers. The task is distinctly more challenging than previous work done on ship detection due to constraints including low resolution and small size of the barges. In addition, ambiguity in detection arises if barges are close to the shoreline, and if other visual artifacts are present. The second primary contribution is proving the feasibility of using computer vision techniques and convolutional neural networks to successfully detect barges with a sufficiently high level of accuracy for field use.

In the future, this model could be improved through collection of a larger inland watercraft dataset, which would also increase confidence in the results that we reported. We would like to see continued exploration of the transfer learning approach to see whether it performs better in a setting with more data. Other techniques such as contrastive learning or few-shot detection may increase accuracy with limited datasets. Work is also greatly needed in addressing similar illegal sand and gravel mining around the world. We would like to see more research conducted with local experts as a human in the loop in these AI detection systems.

7 Acknowledgements

Thank you to Professor Milind Tambe, Doria Spiegel, Boriana Gjura, and WWF Senior Analyst Felipe Spina Avino.

References

[1] J. J. Swenson, C. E. Carter, J.-C. Domec, and C. I. Delgado, "Gold mining in the peruvian amazon: Global prices, deforestation, and mercury imports," *PLOS ONE*, vol. 6, no. 4, pp. 1–7, 04 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0018875

[2] E. L. M. W. Julian Lee, Sayak Maity, "Amazon Barge Detection Dataset," 2021. [Online]. Available: https://github.com/sayakmaity/planet-data-288

[3] G. Margarit, J. J. Mallorqui, J. M. Rius, and J. Sanz-Marcos, "On the usage of grecosar, an orbital polarimetric sar simulator of complex targets, to vessel classification studies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3517–3526, 2006.

[4] M. Tello, C. López-Martínez, and J. J. Mallorqui, "A novel algorithm for ship detection in sar imagery based on the wavelet transform," *IEEE Geoscience and remote sensing letters*, vol. 2, no. 2, pp. 201–205, 2005.

[5] M. Ma, J. Chen, W. Liu, and W. Yang, "Ship classification and detection based on cnn using gf-3 sar images," *Remote Sensing*, vol. 10, no. 12, p. 2043, 2018.

[6] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster r-cnn based on cfar algorithm for sar ship detection," in *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*. IEEE, 2017, pp. 1–4.

[7] Y.-L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on yolov2 for sar imagery," *Remote Sensing*, vol. 11, no. 7, p. 786, 2019.

[8] S. Li, Z. Zhou, B. Wang, and F. Wu, "A novel inshore ship detection via ship head classification and body boundary determination," *IEEE geoscience and remote sensing letters*, vol. 13, no. 12, pp. 1920–1924, 2016.

[9] F. Yang, Q. Xu, and B. Li, "Ship detection from optical satellite images based on saliency segmentation and structure-lbp feature," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 602–606, 2017.

[10] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun, "R-cnn-based ship detection from high resolution remote sensing imagery," *Remote Sensing*, vol. 11, no. 6, p. 631, 2019.

[11] D. O. Fuller, "Tropical forest monitoring and remote sensing: A new era of transparency in forest governance?" *Singapore Journal of Tropical Geography*, vol. 27, no. 1, pp. 15–29, 2006.

[12] M. A. Ashraf, M. J. Maah, I. Yusoff, A. Wajid, and K. Mahmood, "Sand mining effects, causes and concerns: A case study from bestari jaya, selangor, peninsular malaysia," *Scientific Research and Essays*, vol. 6, no. 6, pp. 1216–1231, 2011.

[13] P. Team, "Planet Application Program Interface: In Space for Life on Earth," Planet, 2017–. [Online]. Available: https://api.planet.com

[14] A. P. Antonio-Javier Gallego and P. Gil, "Automatic ship classification from optical aerial images with convolutional neural networks," *Remote Sensing*, vol. 10, no. 4, 2018.

[15] "Modified Copernicus Sentinel data/Sentinel Hub," European Space Agency, 2021. [Online]. Available: https://www.sentinel-hub.com/