

# DeepPlastic: A Novel Approach to Detecting Epipelagic Bound Plastic Using Deep Visual Models

Gautam Tata<sup>1</sup>  
CSU Monterey Bay

Sarah-Jeanne Royer<sup>2</sup>  
The Ocean Cleanup

Olivier Poirion<sup>3</sup>  
UC San Diego

Jay Lowe<sup>3</sup>  
CSU Monterey Bay

**Abstract**—The quantification of positively buoyant marine plastic debris is critical to understanding how concentrations of trash form across the world’s ocean and identifying high concentration garbage hotspots in dire need of trash removal. Currently, the most common monitoring method to quantify floating plastic requires the use of a manta trawl. Techniques requiring manta trawls (or similar surface collection devices) utilize physical removal of marine plastic debris as the first step and then analyze collected samples as a second step. The need for physical removal before analysis incurs high costs and requires intensive labor—preventing scalable deployment of a real-time marine plastic monitoring service across the entirety of Earth’s ocean bodies. Without better monitoring and sampling methods, the total impact of plastic pollution on the environment as a whole, and details of impact within specific oceanic regions, will remain unknown. This study presents a highly scalable workflow that utilizes images captured within the epipelagic layer of the ocean as an input. It produces real-time quantification of marine plastic debris captured in the video for accurate quantification and physical removal. The workflow includes creating and preprocessing a domain-specific dataset, building an object detection model utilizing a deep neural network, and evaluating the model’s performance. YOLOv5-S was the best performing model, which operates at a Mean Average Precision (mAP) of 0.851 and an F1-Score of 0.89 while maintaining near-real-time speed.

<sup>1</sup>

## I. INTRODUCTION

Plastic pollution poses an imminent threat to the marine environment, food safety, human health, eco-tourism and contributes to climate change [1]. Global plastic production has exceeded 500 million tonnes of plastic and current waste estimations indicate that 30% of all produced plastic will end up in the oceans[2] [3]. Researchers have documented a 5x increase in plastic debris within the Central Pacific Gyre and have shown that plastic pieces now outnumber the native plankton 6:1[4]. Marine plastic debris is capable of killing marine life and affects at least 267 species worldwide—including 87% of all sea turtles and 44% of all seabird species. Wildlife impacts include ingestion, entanglement, starvation, suffocation, infection, and drowning [5]. Marine plastic debris also restrains access to the entire food web for wildlife such as marine mammals, pelagic fish species, sea turtles, and seabirds.

While short-term effects on humans warrant immediate concern, long-term consequences such as environmental remain a mostly unknown problem [6].

A significant amount of marine plastic (about 80%) originates

from land-based sources: Most commonly food containers, such as plastic bags and bottles, and packaging materials. The other 20% stems from shipping vessel discharges and discarded commercial fishing gear. Illegal dumping and accidental chemical leakage create notable plastic waste capable of eradicating marine ecosystems in extremely short periods of time.

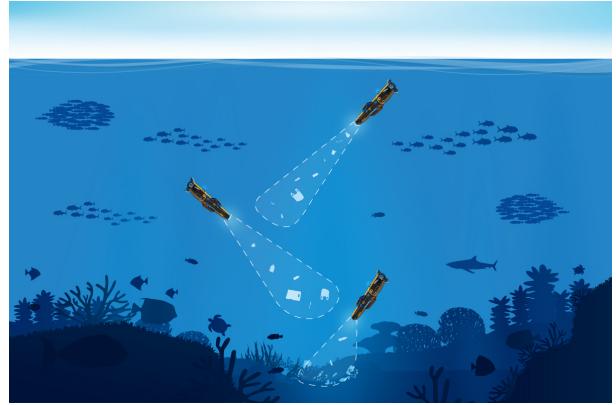


Fig. 1. Concept of real-time plastic detection via AUV’s equipped with cameras and DeepTrash vision

To understand the spatiotemporal distribution of plastic, we require better mitigation strategies. Various in situ approaches to ocean plastic monitoring have been proposed. These in situ methods include using SONAR/LIDAR to map plastic debris[7], human counting via visual methods [8], and debris sampling using fishing nets [9]. However, these methods incur high financial and labor costs.

Buoyant surface plastic becomes denser due to biofouling and then sinks beneath the ocean surface [10]—drastically increasing quantification difficulty. While traditional methods can help provide site-specific data, they can not be used at different locations without incurring relocation costs and necessitate further sampling requirements. These limitations create an opportunity for an alternative method to map marine debris plastic distribution across the world’s oceans.

By applying computer vision and modern deep learning methods—quantification of marine plastic debris can be performed without physical removal. Studies have shown that removing plastic from the oceans will exponentially benefit the ecosystems. Examples include: preventing the movement of invasive species between regions [11], preventing the degradation of plastic into micro-plastics [12], and lessening

<sup>1</sup>Code Availability at <https://github.com/gautamtata/DeepPlastic>

greenhouse gases (thereby decelerating climate change) [13].

In this study, we developed an analytical pipeline that uses images/videos from the epipelagic layer of the ocean and identifies the plastic debris from the ocean's natural background. Unlike other recently proposed algorithms that specialize in monitoring either floating marine plastic [14] or deep-sea marine plastic [15] debris: Our method focuses on marine plastic within the epipelagic layer and uses state-of-the-art deep learning models to produce significantly more promising results capable of quantification in a variety of marine environments.

Our model has been field-tested, and the results indicate that it could be deployed worldwide, in a wide range of water conditions. The dataset was built from images taken across three sites throughout California (South Lake Tahoe, Bodega Bay, San Francisco Bay) along with a compendium of images hosted by research institutions on the internet to increase the representation of marine debris plastics in different locations. The primary source of internet images were underwater photos taken by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) [16]. The training dataset consists of 3200 total images.

In this study, we tested two state-of-the-art deep-learning models YOLOv4-Tiny and YOLOv5-S and reported their performances to infer marine plastic debris in real-time. The main results will be described as follows: 1) the model's precision and accuracy to feasibly identify plastic debris at a mean average precision (mAP) of 85%, 2) insurability that this method can successfully distinguish marine plastic debris from similar-looking non-plastic objects, and 3) A generalized model capable of detecting marine plastic in any oceanic environment. The results show that deep learning models can identify plastic with significant accuracy while operating at a rate that supports real-time applications such as autonomous underwater vehicles (AUVs) for at-scale marine-plastic quantification and monitoring.

**To the best of our knowledge, this study is the first to propose and evaluate the use of deep learning and computer vision methods to quantify marine debris plastic located within the epipelagic layer at the time of this writing.**

## II. RELATED WORK

Increasing demand for identifying and removing plastic from the world's waterways has led to a surge of research in computer vision and AUV solutions. A team of researchers at the University of Minnesota robotics lab recently experimented with AUV deployments for identifying deep ocean marine plastic debris [15]. Another growing trend has been to utilize deep learning and computer vision to automatically identify floating marine plastic on river and ocean surfaces [17].

Additionally, AUV's have been used as a means for environmental surveillance [18], mapping [19], and localization of marine plastic debris [20]. Underwater vision technology has been pushed forward thanks to work done by Ge et al.[21] with LIDAR technology to localize and map marine-plastic

debris on coastal beaches. Further research into implementing LIDAR in conjunction with forward-facing SONAR image models trained by deep convolutional neural networks was conducted by Howell et al. [22], and Valdenegro-Toro et al. [7] which resulted in a model capable of detecting underwater debris with 80% accuracy. Unfortunately, these methods incur high expenses due to retrofitting sonar and an in-house water tank for evaluation.

The University of Minnesota robotics lab [15] annotated and published a dataset of images collected by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) [16]. JAMSTEC released the J-EDI (JAMSTEC E-Library of Deep-Sea Images), which contains marine plastic debris dating back to 1982 and provides data in the form of images and videos. The work presented in this research paper has benefited from the University of Minnesota team, which released close to 3000 annotated images from the JAMSTEC J-EDI dataset. These datasets were used to train our convolutional neural networks (CNNs) to identify features of plastic debris.

Photography, especially video-cameras, have found common application as environmental monitoring systems [23] [24] [25]. Underwater cameras provide a globally accessible and low-cost quantification aid. Combining object detection models with underwater cameras equipped on automobiles such as AUV's makes it possible to observe and monitor sub-surface plastics in known hotspots worldwide [15]. By mounting video cameras to AUV's, buoys, and other submersibles, institutions could feasibly quantify macro-plastics, which constitute 90% of the total plastic mass in the oceans.

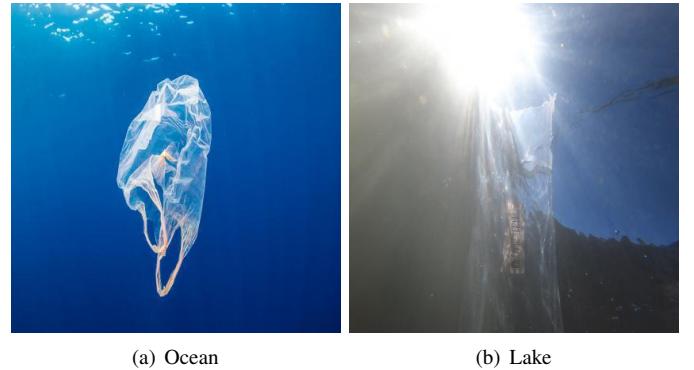


Fig. 2. Example images of marine plastic debris from the DeepTrash dataset in different marine environments

## III. NETWORK ARCHITECTURE

Two of the top performing deep learning architectures commonly used for object detection were selected for this project. Each architecture has different benefits and drawbacks, with the main trade off being speed for accuracy.

- *YOLOv5-S* Unlike the official release of YOLOv4, YOLOv5 currently exists in active development. Therefore, all YOLOv5 related code and models may be subject to modification or deletion without notice. YOLOv5-S has *7.5 million parameters, 140 layers, and operates at*

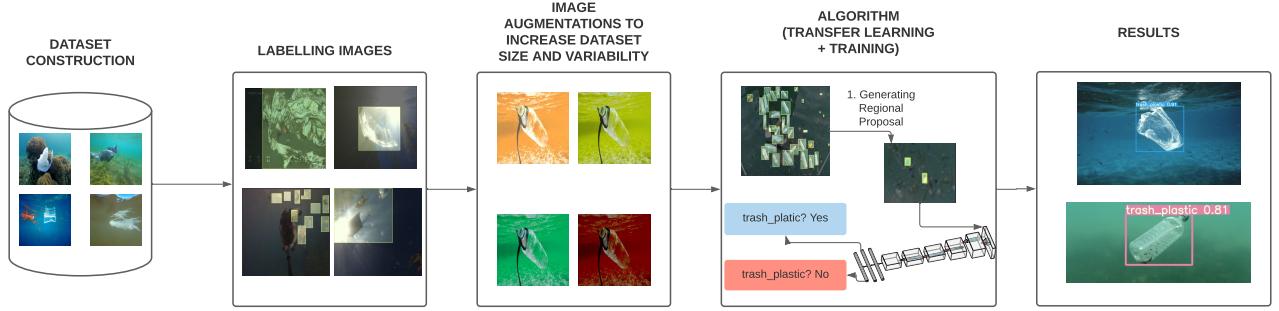


Fig. 3. Methodology for Marine Plastic Detection

a lightweight 7MB (14MB for weights pre-trained on COCO). This architecture uses the *Cross Stage Partial Network* (CSP) [30] as the processing backbone and was trained on MSCOCO to extract rich/informative features from an input image. YOLOv5 also uses a *PANet* [31] for the model-neck to generate feature pyramids and the computational friendly LeakyReLU and Sigmoid activation function. The model uses *SGD* as a default learning rate but these tests were performed with the ADAM adaptive learning rate enabled [32].

- *YOLOv4-Tiny* Inference speeds on YOLOv4-Tiny can reach upwards of *400 frames/second* when using a 1080Ti GPU with accuracy, precision, and recall that meet the demands of a production-ready robotics platform. YOLOv4-Tiny uses a *CSPDarknet53-Tiny* neural network as opposed to the regular *SPDarknet53* network. To simplify the computation process, the YOLOv4-Tiny model uses the *LeakyReLU* as an activation function.

#### IV. METHODOLOGY

##### A. Dataset Construction

The dataset was curated by collecting videos of marine plastic from the field in California and sourcing images from the J-EDI dataset. The videos vary significantly in quality, depth, and visibility to better represent the harshness of marine environments. Each video was recorded in 5 frames per second to produce still images.

After recording, manual identification of marine plastic captured in the still images was performed—with an emphasis on choosing images containing difficult object detection scenarios such as overgrowth and occlusion. Each image would then get annotated to prepare them for the deep learning models. This curation approach ensured that the dataset of images would closely conform to real-world conditions.

Ultimately, over 50,000 base images from which to pull good examples of marine plastic for further annotation was produced in this manner. Annotations were performed using the free tool *supervise.ly* [26] to create the final dataset containing 3200 total images.

##### B. Enhancements of Custom Dataset

The following procedures were implemented for the deep learning models to detect marine plastic:

- Dataset Formatting* The input data, constituted of images and annotation labels for bounding boxes, were converted into either a PyTorch (YOLOv5-S) or a Darknet format (YOLOv4) so that each respective models could process them. The bounding boxes delimited each image's regions of interest based on 2D coordinates located in the respective annotation file [40].
- Image Pre-processing* To ensure that learning occurs on the same image properties, auto orient was applied to strip images of their exchangeable Image file format (EXIF) data [27] so that the models interpret images regardless of image format. Finally, the input images get resized and bounding boxes adjusted to 416x416 pixels.
- Data Augmentation* To mitigate the effects of the model generalizing towards undesired features and to replicate underwater conditions such as variable illumination, occlusion, and color—the dataset was further enhanced by randomly changing the brightness and saturation of the images via PyTorch's built-in Transforms augmentation. These modified images were then added back into the dataset, effectively tripling the size of our dataset.

##### C. Object Detection

We used two state-of-the-art neural network architectures YOLOv5-S and YOLOv4, downloaded from their respective repositories [28] [29]. The following software versions were used: PyTorch v1.8.1, Darknet, OpenCV version 3.2.0, and CUDA 11.2.

1) *Fine Tuning Hyperparameters*: This object detection model uses ADAM [36] as the adaptive learning rate, which utilizes a decaying learning rate for a set number of epochs. The final layer of the network uses Softmax and reflects the usage of a single class.

2) *GPU Hardware*: An NVIDIA Tesla V100@GPU (version 460.32.03) was chosen for due to its proven parallel computing capability. This GPU also has wide accessibility as a pre-enabled GPU available through services such as Google Colab®.

**Algorithm 1** Process for training neural network to detect marine plastic

**Input:** A dataset containing images of required classes, a label map, the pre-trained weights of the transfer learning model, and a configuration file for the pre-trained model.

**Output:** A fully trained object detection algorithm with a file containing the optimized weights of the final model.

- 1: Repeat the process below until model exceeds desired threshold mAP
- 2: Prepare an annotated dataset and split it into training, validation and testing dataset
- 3: Convert the dataset annotations into the appropriate input format (i.e a .yaml file for YOLOv5)
- 4: Fine tune the hyperparameters of the neural network.
- 5: Use SGD or ADAM as adaptive learning rates to fine-tune the weights of the model.
- 6: Monitor the training loss and mean average precision on validation dataset
- 7: If mAP graph converges, stop training to observe and record final validation mAP
- 8: If final model mAP does not exceed threshold mAP, return to step 1
- 9: Obtain the weights of the optimized network
- 10: Deploy the model into production

3) *Training:* After every 1000 epochs (iterations) of training, the model would be evaluated on the validation dataset to calculate precision, recall, and mean average precision (mAP). This means stopping training to check for the following:

- When accuracy stops increasing, the model no longer needs additional training to prevent overfitting.
- Depending on performance, hyperparameters should receive adjustments to optimize for evaluation metrics.

4) *Evaluation Metrics:* After the model has finished training, use the testing and validation datasets containing images mutually exclusive from the training dataset as an input to evaluate the performance of the network.

The model draws a bounding box around successfully detected objects with a confidence score of .50 or higher. The number of true positive bounding boxes drawn around marine plastic debris and true negatives provides the basis of evaluation. The following performance metrics were utilized to produce results:

- **True positive and True negative values:** True positive values represent an outcome in which the models correctly predict a positive class, and conversely, a true negative represents when the model correctly predicts the negative class.
- **Precision and Recall** – represents if the model successfully detected plastic in an image.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

- **Mean Average Precision** – Evaluates how often the network can recognize plastic in a group of images. After collecting the values for true and false positives, generate a precision-recall curve using the Intersection over Union (IoU) formula:

$$IoU = \frac{BBox_{predicted} \cap BBox_{groundTruth}}{BBox_{predicted} \cup BBox_{groundTruth}}$$

Where  $BBox_{predicted}$  and  $BBox_{groundTruth}$  are the areas under the curve for predicted and ground truth bounding boxes, respectively. To ensure accuracy, a high threshold for confidence and IoU must be set-with a correct detection represented by the threshold being exceeded.

The mAP can then be obtained by integrating the precision-recall curve [33]:

$$mAP = \int_0^1 p(x)dx$$

- **F1-Score** – Evaluates the balance between precision and recall values.
- **GPU Speed (ms/IMG)** – Represents how fast the network can infer marine plastic debris contained within an input image.

5) *Visualizing results:* For each processed image, the network populates arrays containing the following data:

- Scores – Confidence scores for the predicted boxes.
- Classes – Labels for each prediction.
- Number of detections – The total number of detections made per image.

A final array containing all bounding boxes which have a confidence score of higher than 50% gets filtered out and used for the output.

The following equation converts the normalized coordinates into image coordinates for rendering bounding boxes on top of images:

$$imgCoord_k = BoxScore_i^j \cdot Width \quad (1)$$

where  $k \in \{\text{left}, \text{right}, \text{top}, \text{bottom}\}$ ,  $i$  is an index of boxes,  $j \in \{0, 1, 2, 3\}$ , and  $Width$  is a width of the image. These image coordinates were used to visualize the results of predicted bounding boxes in Figure 6.

## V. RESULTS

All results expressed in Table I were produced from the validation dataset presented in the methodology section. Since the images used in the training dataset were not isolated laboratory creations, but instead real-world images directly from the field, the general object detection has a more accurate representation of marine plastic debris. This approach comes with a set of trade-offs:

TABLE I  
EVALUATING THE ACCURACY (TP, TN, mAP, AP, AND PRECISION) AND INFERENCE SPEED (GPU SPEED) OF THE MODEL

Model	Dataset	TP	FP	mAP	Recall	Precision	F1	Inference (ms/img)
YOLOv4-Tiny	Deep-Trash(Custom)	584	23	0.84	0.69	0.96	0.80	1.5
YOLOv5-S	Deep-Trash(Custom)	734	48	0.85	0.85	0.93	0.89	2.1
YOLOv4-Tiny	JAMSTEC JEDI <sub>plastic</sub>	219	14	0.93	0.80	0.94	0.86	1.8
YOLOv5-S	JAMSTEC JEDI <sub>plastic</sub>	273	12	0.98	0.98	0.95	0.96	1.4

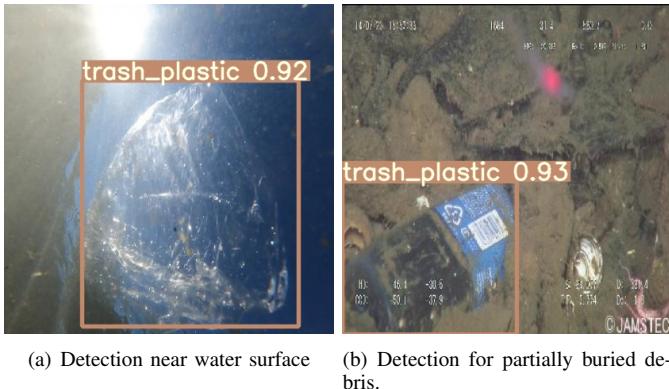


Fig. 4. Results generated by the model with bounding boxes and confidence scores rendered over marine plastic debris.

- The model performs stronger in real-world deployments, and therefore, the evaluation results in Table II do not significantly differ from near-real-time measurements taken from the field.
- Peak performance of the object detection model in a perfectly controlled environment could not be measured, and the highest possible benchmark of a single detection remains unknown.
- These trade-offs indicate the results of this paper better approximate long-term performance across a wider variety of marine environments—leading to a more substantial evaluation of the object detection model’s performance in the field.

#### A. Quantitative Results

The results captured in Table I demonstrate that near-real-time object detection of marine plastic debris in the epipelagic layer of the ocean is both feasible and close to real-world execution. The tested models demonstrate high average precision, mAP, and F1 scores relative to their inference speed. Repeated testing of the model produced a results variance of 2%.

Usually, evaluation results between models showcase a clear relationship between models, such as trading off significant inference speed for increased accuracy, but the results presented in this paper showcase that both YOLOv4-Tiny and YOLOv5-S produce high debris localization metrics when it comes to identifying epipelagic plastic in near-real time.

**YOLOv5-S provides a significantly higher F1 score in**

**exchange for a slight dip in inference performance.**

Reducing the number of classes to 1, i.e., "trash\_plastic," ensures even distribution of class examples within the training dataset. The singular nature of this object detection model may reduce the total number of use cases the model can be utilized for—but guarantees strong performance on use cases within the domain of the model. A single classification also builds upon the performance of the pre-trained weights utilized during transfer learning, as it meant less skewing towards unrelated classifications.

#### B. Evaluation Results

**1) Object Detection:** The mAP values obtained from the object detection models on the validation dataset have been expressed in Table I. Both models demonstrate high accuracy in plastic localization. **It also reveals that the YOLOv5-S model has a higher mAP than the YOLOv4-Tiny model.**

**2) Inference Speed:** These speeds were dictated by the GPU (NVIDIA V100 using a batch size of 32) and includes image pre-processing. **The YOLOv4-S model provided the highest inference speed-to-mAP performance ratio for the provided dataset.**

#### C. Qualitative Results

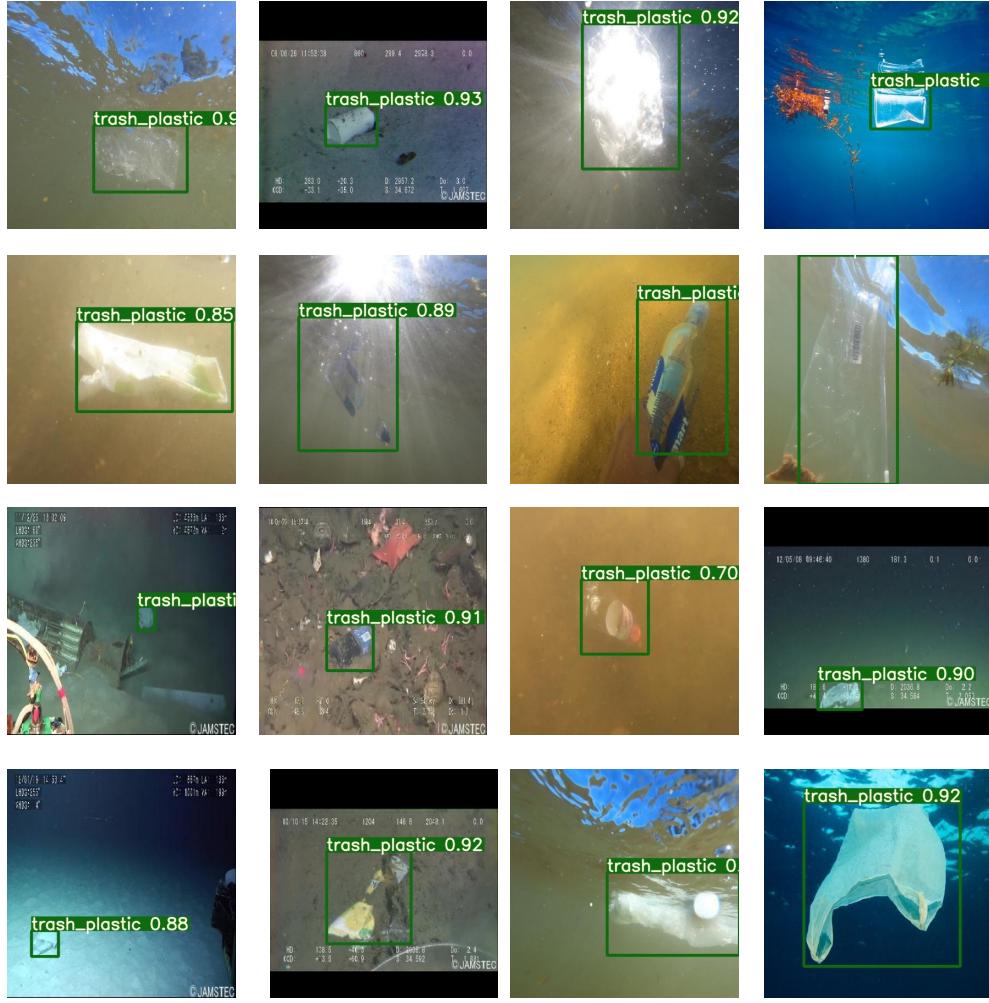
This study focused on determining the feasibility of detecting marine plastic debris for near-real-time monitoring/quantification purposes. To that end, the results in Table I demonstrate that general object detection models can fill this much-needed role. Since a relatively high level of performance can be maintained with such fast inference speeds—we believe that models such as the one presented in this paper can be applied to AUV’s and other tools for real-world solutions. Equally important is that these solutions now have a near-future timeline of implementation and have been proven to be low cost.

## VI. DISCUSSION

In this study, we built a computer vision model that detects marine plastic debris with high precision, visualizes the detections with bounding boxes, and operates at near real-time speeds. These conditions match the requirements for robotic platforms such as AUVs or buoys.

As one of the first object detection models specialized for the epipelagic layer, direct comparison of results can not be easily performed. However, relative performance comparisons

Fig. 5. Inference run on images with the final weights obtained from the Deep Learning model



between DeepPlastic and object detection models geared towards plastic detection in deep-sea and river plastic environments reveal DeepPlastics' state-of-the-art performances.

A research team at the University of Minnesota developed a computer vision model specialized for marine plastic detection in deep-sea environments that achieved a mAP of 67.4%[?]. DeepPlastic achieves a mAP of 85% on plastic in similar conditions.

An article in Earth and Space Science[14] describes a model capable of detecting plastic floating on rivers with a precision rate of 68%. DeepPlastic achieves a precision rate of 93% when detecting marine plastic debris submerged in the ocean.

The datasets used by the two models above are either not public or utilize datasets outside of the domain of DeepPlastic (i.e. the dataset images are not underwater), therefore comparing performances via dataset is not an option for this study.

#### A. Points of Improvement

This model can efficiently monitor and quantify marine plastic. Improvements can be made in the following areas:

1) *Data Augmentation Improvements*: While grayscale, saturation, and vertical/horizontal flipping have been proven data augmentation techniques—emerging techniques such as AutoAugment [?] could be explored to improve the model's variability in the future once ready for adaptation. Other methods such as shear and the cutout regularization technique would be great to utilize after integration technologies improve.

2) *Object Detection Algorithm Improvements*: This improvement concerns the Convolutional Neural Networks (CNNs) of the algorithm. Efforts were focused on high inference speeds, therefore the YOLOv4-Tiny and YOLOv5-S were selected as they were built for real-time object detection. Models with slower inference speed, such as YOLOv5-X, EfficientDet-D7, and YOLOv4, have much better accuracy/precision but do not reach the near real-time speeds required by real-time monitoring systems. With new developments in deep learning, future models could combine the higher performance of the large models while maintaining the speed of the faster models.

Objects that have a similar structure to plastic sometimes

result in mis-classifications. Our model occasionally classifies jellyfish in certain lighting conditions as plastic due to the similar transparency and structural properties found in both plastic and jellyfish. An improved object detection algorithm would be able to more accurately identify the unique latent features of plastic and prevent mis-classifications of this type.

3) *Dataset Improvements*: The data set used in this study is unique and one of the first of its kind. For the data set, we see three main improvements that can be made to enhance the deep learning model:

- Adding more images from different locations
- Using more images from other types of water conditions
- Finally, acquiring a more extensive set of underwater plastic images

As more plastic images from different locations and oceanic conditions become available, they will increase marine plastic debris representation—providing a more comprehensive dataset for model training. We believe this will improve the mAP and overall robustness of the object detection model.

4) *Camera Improvements*: Readily available off-the-shelf cameras have come a long way but still suffer from certain limitations. The first and most substantial limitation revolves around the fact that most underwater cameras will only work during the daytime. If we want to continue the monitoring process during the night-time, better night-vision underwater sensors need to be developed. The second limitation stems from the common H.265 video compression techniques [34] underwater cameras utilize to induce encoding artifacts. This impedes real-time detection by deteriorating the image quality. Developments in end-to-end deep learning video compression techniques [34] could lead to solutions for this limitation once ready for implementation.

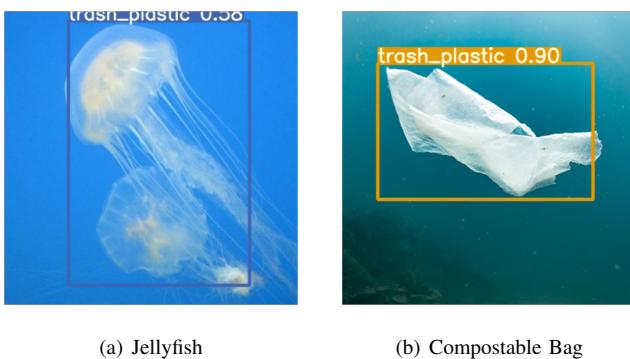


Fig. 6. Misclassification of objects with similar structure to marine plastic.

## VII. CODE AND DATASET AVAILABILITY

All code and instructions to build and utilize the DeepPlastic object detection model can be found [online via GitHub](#).

The DeepTrash dataset can be found in this [publicly available Google Drive folder](#).

## VIII. CONCLUSION

This work's objective was to develop a deep learning vision model capable of consistently identifying and quantifying marine plastic in near real-time. To attain this objective, a pair of general object detection models were constructed using two state-of-the-art deep learning models built for inference speed to measure which performed best.

This study concludes that a marine plastic debris detection system based on the YOLOv5-S model would be fast, accurate, and robust enough to enable real-time marine plastic debris detection. Furthermore, this study shows that effective object detection models can be constructed using readily available, pre-enabled GPUs for reasonable costs.

Furthermore, the dataset created for and utilized by this general detection model demonstrates that massive, highly curated datasets can be used in conjunction with samples relative to the domain of object detection and web scraping to produce promising results.

*This computer vision system enables multiple deployment methods to detect/monitor marine plastic and allows researchers to quantify marine plastic debris without physical removal.*

## IX. FUTURE WORK

Improvement of the dataset would have the highest impact on performance, but collecting additional images would require human labor in fieldwork or preprocessing. A technology capable of producing synthetic images containing marine plastic debris in an ocean environment could provide an automated solution to dataset creation. This could be accomplished with a two-stage autoencoder[35]. Object detection models trained on identifying jellyfish (or other objects similar to marine plastic debris) paired with a DeepTrash object detection model could lead to a decrease in false positives.

Inference speed could be improved through specialized GPU technology or tailoring models towards specific higher power GPUs than used in this study.

An end-to-end video compression technique explicitly developed for near real-time object detection could lead to a better ratio of true positives to true negatives and improved range on object detection.

Tailoring this object detection model for vision-equipped AUVs could result in automated identification and plastic removal devices capable of scalable deployment across large bodies of water, as shown in figure 1. Further optimizations could build in support for stationary monitoring devices such as buoys as well. We hope that such a system will facilitate scalable adoption by researchers and civilians to detect and clean up marine plastic.

## X. ACKNOWLEDGEMENTS

We gratefully acknowledge the help and support of Nikhil Deshmudre for his efforts and help with the deployment of this computer vision system.

The authors would also like to thank Joseph Nelson, Co-Founder of Roboflow.com, for providing us with Roboflow Pro

free of charge, making it easier to iterate on the deep learning models. Some of the images in this dataset were sourced from the TrashCan dataset, where the researchers hand-annotated and open-sourced over 5000 images from the JAMSTEC-JEDI dataset.

The authors would like to thank the researchers from the University of Minnesota, Robotics LAB, and the Japan Agency for Marine-Earth Science and Technology for open sourcing this data to contribute to the advancement of science. Finally, we would like to thank Rae Rose Lowe for her support throughout this process.

## REFERENCES

- [1] C. Schmidt, T. Krauth, and S. Wagner, "Export of plastic debris by rivers into the sea," *Environmental science & technology*, vol. 51, no. 21, pp. 12246–12253, 2017.
- [2] A. Nollkaemper, "Land-based discharges of marine debris: from local to global regulation," *Marine Pollution Bulletin*, vol. 28, no. 11, pp. 649–652, 1994.
- [3] U. EPA, "Municipal solid waste generation, recycling, and disposal in the united states tables and figures for 2012;," *Office of Resource Conservation and Recovery*, 2014.
- [4] J. Clapp, "The rising tide against plastic waste: unpacking industry attempts to influence the debate," *Histories of the dustheap: Waste, material cultures, social justice*, pp. 199–225, 2012.
- [5] D. W. Laist, "Impacts of marine debris: entanglement of marine life in marine debris including a comprehensive list of species with entanglement and ingestion records," in *Marine debris*, pp. 99–139, Springer, 1997.
- [6] H. M. Nevins, S. R. Benson, E. M. Phillips, J. d. de Marignac, A. De Vogelaere, J. A. Ames, and J. T. Harvey, "Coastal ocean mammal and bird education and research surveys (beachcombers), 1997-2007: ten years of monitoring beached marine birds and mammals in the monterey bay national marine sanctuary," 2011.
- [7] M. Valdenegro-Toro, "Deep neural networks for marine debris detection in sonar images," 2019.
- [8] T. van Emmerik, T.-C. Kieu-Le, M. Loozen, K. van Oeveren, E. Strady, X.-T. Bui, M. Egger, J. Gasperi, L. Lebreton, P.-D. Nguyen, *et al.*, "A methodology to characterize riverine macroplastic emission into the ocean," *Frontiers in Marine Science*, vol. 5, p. 372, 2018.
- [9] S. Rech, V. Macaya-Caquilpán, J. Pantoja, M. Rivadeneira, D. J. Madariaga, and M. Thiel, "Rivers as a source of marine litter—a study from the se pacific," *Marine pollution bulletin*, vol. 82, no. 1-2, pp. 66–75, 2014.
- [10] H. Zhang, "Transport of microplastics in coastal seas," *Estuarine, Coastal and Shelf Science*, vol. 199, pp. 74–86, 2017.
- [11] J. T. Carlton, J. W. Chapman, J. B. Geller, J. A. Miller, D. A. Carlton, M. I. McCuller, N. C. Treneman, B. P. Steves, and G. M. Ruiz, "Tsunami-driven rafting: Transoceanic species dispersal and implications for marine biogeography," *Science*, vol. 357, no. 6358, pp. 1402–1406, 2017.
- [12] A. L. Andrade, "Microplastics in the marine environment," *Marine pollution bulletin*, vol. 62, no. 8, pp. 1596–1605, 2011.
- [13] S.-J. Royer, S. Ferrón, S. T. Wilson, and D. M. Karl, "Production of methane and ethylene from plastic in the environment," *PLoS One*, vol. 13, no. 8, p. e0200574, 2018.
- [14] C. van Lieshout, K. van Oeveren, T. van Emmerik, and E. Postma, "Automated river plastic monitoring using deep learning and cameras," *Earth and Space Science*, vol. 7, no. 8, p. e2019EA000960, 2020. e2019EA000960 10.1029/2019EA000960.
- [15] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," 2018.
- [16] J. A. for Marine-Earth Science and Technology, "Deep-sea debris database," 2012.
- [17] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, S. Sharma, and I. H. Sarker, "Aquavision: Automating the detection of waste in water bodies using deep transfer learning," *Case Studies in Chemical and Environmental Engineering*, p. 100026, 2020.
- [18] S. B. Williams, O. Pizarro, M. V. Jakuba, I. Mahon, S. D. Ling, and C. R. Johnson, "Repeated auv surveying of urchin barrens in north eastern tasmania," in *2010 IEEE International Conference on Robotics and Automation*, pp. 293–299, 2010.
- [19] N. Barrett, J. Seiler, T. Anderson, S. Williams, S. Nichol, and S. Nicole Hill, "Autonomous underwater vehicle (auv) for mapping marine biodiversity in coastal and shelf waters: Implications for marine management," in *OCEANS'10 IEEE SYDNEY*, pp. 1–6, 2010.
- [20] L. Chen, S. Wang, K. McDonald-Maier, and H. Hu, "Towards autonomous localization and mapping of auvs: a survey," *International Journal of Intelligent Unmanned Systems*, vol. 1, p. 97–120, May 2013.
- [21] Z. Ge, H. Shi, X. Mei, Z. Dai, and D. Li, "Semi-automatic recognition of marine debris on beaches," *Scientific Reports*, vol. 6, May 2016.
- [22] T. Kurz, S. Buckley, J. Howell, and D. Schneider, "Close range hyperspectral and lidar data integration for geological outcrop analysis," in *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE, Aug 2009.
- [23] M. Mock and E. B. Voelml, "Underwater monitoring and communication system," Feb. 28 1995. US Patent 5,392,771.
- [24] R.-Q. Jan, Y.-T. Shao, F.-P. Lin, T.-Y. Fan, Y.-Y. Tu, H.-S. Tsai, and K.-T. Shao, "An underwater camera system for real-time coral reef fish monitoring," *The Raffles Bulletin of Zoology*, vol. 14, pp. 273–279, 2007.
- [25] S. Premkumardeepak and M. M. Krishnan, "Intelligent sensor based monitoring system for underwater pollution," in *2017 International Conference on IoT and Application (ICIOT)*, pp. 1–4, IEEE, 2017.
- [26] D. Drogzov.
- [27] N. L. Iacobici, M. Frigura-Iliasa, H. E. Filipescu, M. Nen, F. M. Frigura-Iliasa, and M. Iorga, "Digital imaging processing and reconstruction for general applications," in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 231–234, 2020.
- [28] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.
- [29] G. Jocher, A. Stoken, J. Borovec, NanoCode012, ChristopherSTAN, L. Changyu, Laughing, Tkianai, A. Hogan, Lorenzomammana, and et al., *ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements*. Zenodo, Oct 2020.
- [30] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "CspNet: A new backbone that can enhance learning capability of cnn," 2019.
- [31] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," 2018.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
- [33] K. Boyd, K. H. Eng, and C. D. Page, "Area under the precision-recall curve: Point estimates and confidence intervals," in *Proceedings of the 2013th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part III*, ECMLPKDD'13, (Berlin, Heidelberg), p. 451–466, Springer-Verlag, 2013.
- [34] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "Dvc: An end-to-end deep video compression framework," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [35] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.