Water-D: Detecting Waterborne Debris with Sim2Real and Randomization

Jie Fu[*,1,2,a] Ritchie Ng[*,3,b] Mirgahney Mohamed[4] Yi Tay[5] Kris Sankaran[1,6] Shangbang Long[7] Alfredo Canziani[8] Chris Pal[1,2] Moustapha Cisse[9]
[1]Mila, [2] Polytechnique Montréal, [3]National University of Singapore, [4]The African Institute for Mathematical Sciences, [5]Nanyang Technological University [6]Université de Montréal
[7]Peking University [8]New York University [9]Google AI
[*]Equal contribution [a]jie.fu@polymtl.ca [b]ritchieng@u.nus.edu

Challenge

From palpable marine debris to microplastics, marine debris pollution has been a perennial problem. This pollution has negative consequences for both ecosystems and human health. Large-scale cleanup efforts are making their way around the world [1]. However, the human resources required to accomplish this goal are limited and the afflicted area is vast. Unmanned vehicles that are capable of automatically detecting and removing small-sized debris would be a great complementary approach to existing large-scale garbage collectors.

Context

There are two traditional approaches to supporting pollution reduction efforts with automation technologies,

- Passive collection: Unmanned trash collection vehicles that collect trash they happen to encounter. This is sufficient in areas with large, densely packed debris.
- Sorting: Classification of trash types in recycling centers. This controlled environment makes robots unnecessary.

Effectively classifying marine debris in an active way remains a challenge, primarily due to the lack of sufficient training data.
In the ML community, data augmentation based on realistic game-based simulations have emerged as a promising candidate for building effective models in resource constrained settings [2–4]

Proposal

Inspired by recent successes in training deep models with synthetic data [2, 3] and domain randomization [4, 5], we propose to train a debris detector based on a mixture of real and synthetic images. The synthetic images are rendered by Unreal Engine 4 [6], and they are further augmented by domain randomization [4, 5]. We aim to deploy a debris-searching unmanned vehicle in polluted waterways in Singapore.

0.1 Setup

We use a YOLOv3 object detector [7] pre-trained on COCO dataset [8] and a small-scale trash dataset [9].

0.2 Augmentation Scheme

We propose `DebrisWorld` in Unreal Engine 4 to generate example small marine debris images. The original three-dimensional environment has a simple water and sky base, this is populated using model marine debris: bottles, cans, paper.
To improve generalization, we propose applying domain randomization to the original water and sky background, generated independently of the marine debris of interest.

System Overview

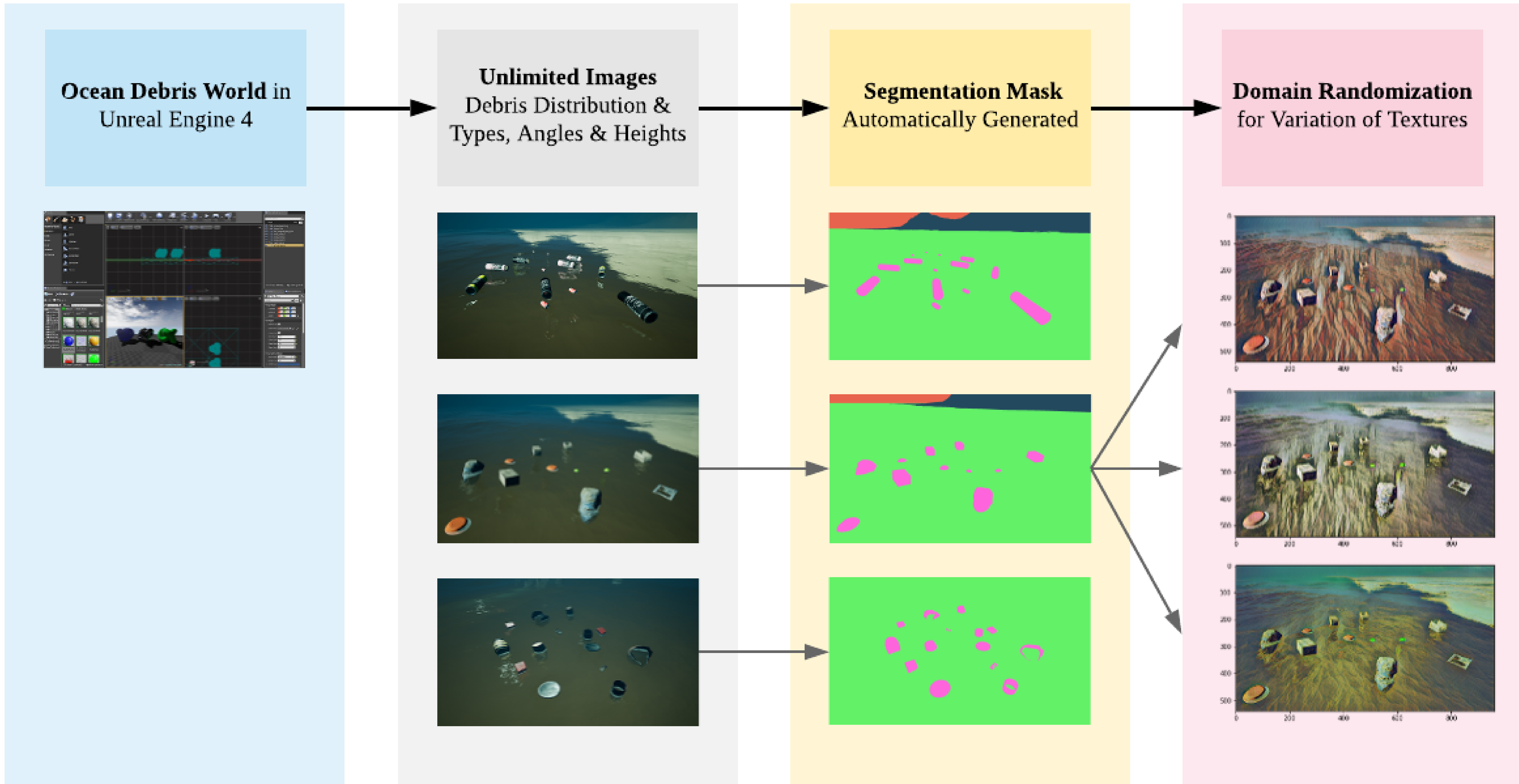

Figure 1: Workflow to generate synthetic ocean debris images using randomization.

Preliminary Work

We have developed a data augmentation scheme for marine debris detection, summarized in Figure 1. This includes the `DebrisWorld` simulator and domain randomization scheme. The randomization-induced heterogeneity in background texture and color is intended to increase the utility of simulated samples across unfamiliar real data environments.
This simulation scheme allows us to generate an infinite number of example images along with segmentation masks, marking what is debris, water, and sky. An initial experiment detecting debris using an unmanned vehicle in Singapore waterways has demonstrated that restricting training to only real labeled data yields unsatisfactory results.

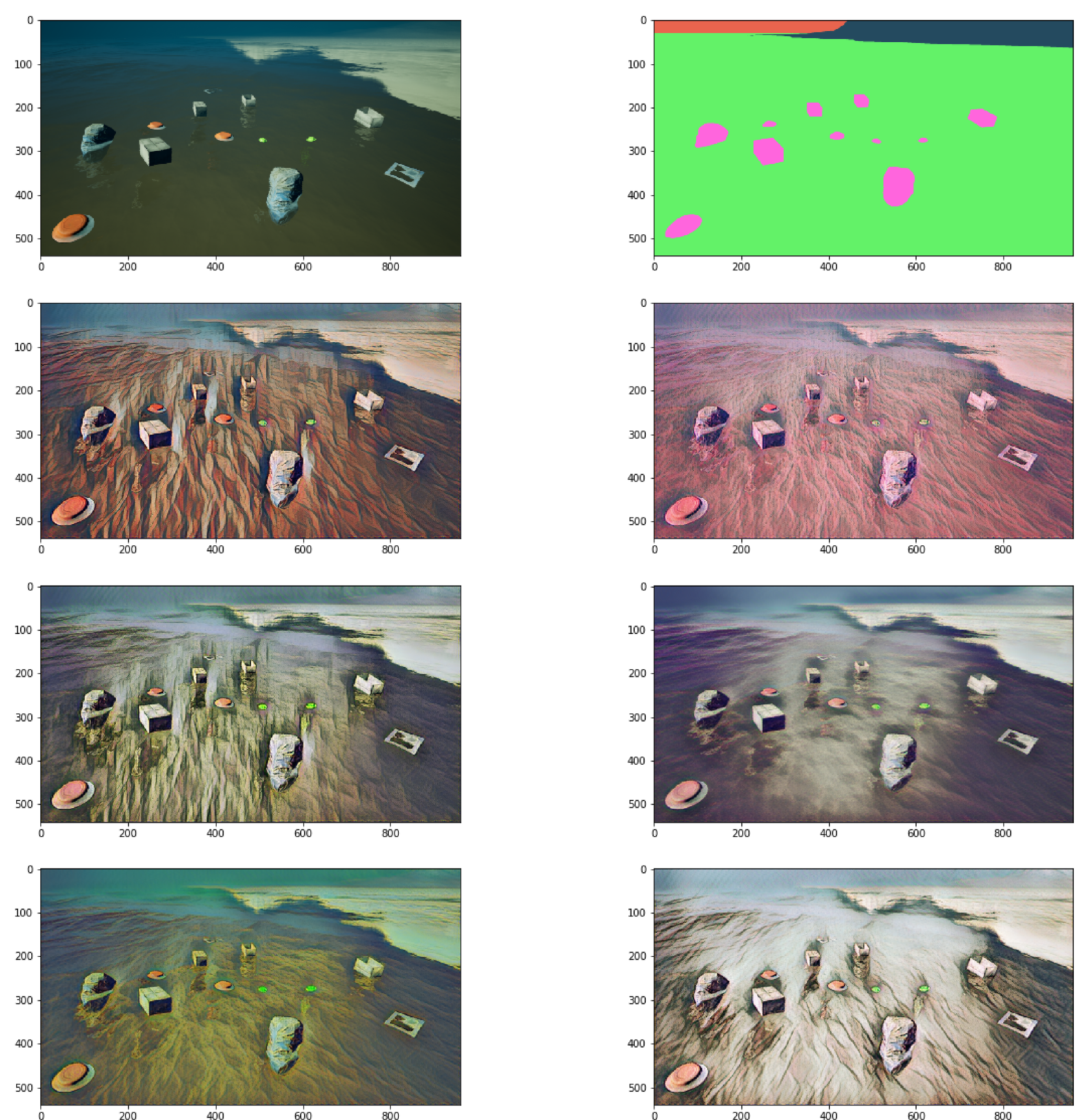

Figure 2: The top left image is the synthetic image of debris in the sea. The top right image is the segmentation mask. The remaining 6 images are domain randomization permutations generated from the original synthetic image.

Conclusion

We focus on the problem of building an effective small-debris detector, to be used in unmanned garbage collection vehicles. We adapt game-based simulation data augmentation strategies in to overcome constraints on collecting and labeling images of small-scale waterborne debris.
Expanding the richness of our simulation and randomization schemes, so that they capture enough variation to allow successful deployment, remains an area of current study. We are also considering the use of anomaly detection methods to flag debris that cannot be easily simulated. All our datasets and model implementations will be made publicly available.

[1] Ocean Cleanup. The ocean cleanup, May 2019.
[2] Tremblay Jonathan et. al. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. *arXiv preprint arXiv:1804.06516*, 2018.
[3] Sara Beery et. al. Synthetic examples improve generalization for rare classes, 2019.
[4] Tobin Josh et. al. Domain randomization for transferring deep neural networks from simulation to the real world. In *Intelligent Robots and Systems (IROS)*, pages 23–30, 2017.
[5] Wenhan Luo et al. End-to-end active object tracking and its real-world deployment via reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, 2018.
[6] Weichao Qiu and Alan Loddon Yuille. Unrealcv: Connecting computer vision to unreal engine. In *ECCV Workshops*, 2016.
[7] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
[8] Tsung-Yi Lin et. al. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
[9] Gary Thung and Mingxiang Yang. Classification of trash for recyclability status. 2016.