UROP Final Report

By Christopher Morse

I. Summary

i. *Introduction*

In recent years, Autonomous Underwater Vehicles (AUVs) have been in the forefront of oceanic research and exploration. Since some of these AUVs depend on visual input to move, interact with the environment, and communicate, it is important for them to make accurate generalizations about their visual observations. To better equip these AUVs to perform well, it is necessary for them to have an object detection architecture (ODA) that is sufficiently trained to safely operate in the real world. Unfortunately, collecting underwater imagery to train the ODA is often time consuming, expensive, and hazardous for both the humans and robots involved. Inspired by this problem, this research sought to explore a more accessible method of data augmentation, through the intersection of deep learning and synthetic data generation, to improve underwater robot-to-robot detection.

ii. *Concept and Data Collection*

As a way to improve robot detection, if a conversion can be made from images of a robot operating in a pool to images that mimic real ocean images, existing ocean datasets could be expanded to provide ODAs with more training data. To do this, it was necessary to compile images of a robot in a pool and in the ocean. The Interactive Robotics and Vision (IRV) Laboratory at the University of Minnesota had several videos of their Aqua robot ("Minnebot") operating in these environments. These videos were then sliced into individual frames and separated into two image datasets.

iii. *Synthetic Data Generation*

CycleGAN, a generative adversarial network which provides unpaired image-to-image translation between domains, was then trained on the pool and ocean datasets as its source and target domains, respectively. The most fundamental features of CycleGAN's architecture are the generator and discriminator networks, which compete with each other throughout the training process. The generator network will try to translate images from one domain (e.g. pool images) to look like images from the other (e.g. ocean images), and the discriminator network will attempt to determine whether or not the image was generated or if it is a real member of the target domain [1]. After training, CycleGAN successfully generated 'fake' ocean images from the pool images. These images were compiled into a new dataset (see Figure 1).

Figure 1. (Left to right) Real pool image, real ocean image, resulting CycleGAN-produced fake ocean image.

iv. *Object Detection*

RetinaNet, an ODA, was used to determine whether or not these synthetic images have any effect on robot detection. RetinaNet consists of a Feature Pyramid Network (FPN) built on a Residual Network (ResNet), a classification subnetwork, and a regression subnetwork. Inside of the ResNet FPN backbone, each training image is passed through convolutional layers that output feature maps. After repeated convolutional feature extraction and reconstruction, the classification subnetwork determines which class(es) are present in each region of the image while the regression subnetwork determines the bounding box dimensions of the detected class(es) in these regions [2].

II. Results

Three detection models were trained using three separate datasets: a set of 511 real ocean images, a set of 511 'fake' ocean images, and a set of the real and fake images combined. After training RetinaNet for 75 epochs on each dataset, an independent testing set of 102 real ocean images was used for evaluation. To evaluate each model's performance over the testing set, precision and recall scores were computed, along with the average precision (AP) across various detection thresholds. Precision is calculated by finding the proportion of the total number of correct detections to the total number of guesses that were made by the ODA, whereas recall is the proportion of the total number of correct detections made by the ODA to the number of all possible instances for detection in the testing set. The AP is calculated in the following way, where P_n and R_n correspond to the precision and recall values at the nth detection threshold, respectively [3]:

$$\mathrm{AP} = \sum_n (R_n - R_{n-1}) P_n$$

As shown in the results in Table 1, we can determine that synthetic data generation, for the purpose of data augmentation, can be an effective method to improve the performance of an ODA. The significant improvement in the AP score shows that the model trained on the combined datasets performed much better than the model trained exclusively on fewer real ocean images (see Figure 2).

	Real Ocean Images	Fake Ocean Images	All Images
Precision	0.6211	0.4897	0.9059
Recall	0.4836	0.2956	0.6481
AP	0.5273	0.2774	0.9573

Table 1. RetinaNet evaluation metrics for each of the three models.

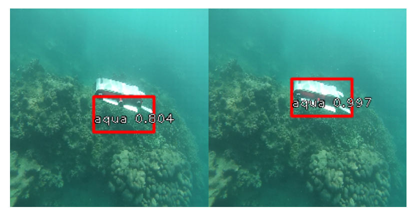

Figure 2. RetinaNet bounding box predictions pre-augmentation (left) and post-augmentation (right).

III. Discussion of UROP Objectives

The majority of my work this summer was spent exploring several alternative methods to generate realistic synthetic image data. In my proposal, I described a plan to run CycleGAN exclusively on ocean backgrounds, and later overlay and color correct images of the robot. This initial plan didn't last very long, since extracting the robot from each image proved to be a much more difficult of a task than I had originally anticipated. I also wanted to see whether or not I could use some 'fake' ocean images in CycleGAN's target domain to depend even less on existing ocean image data. This didn't work out either, as CycleGAN was resistant to retain the foregrounds in the synthetic images. After I was able to generate decently realistic images of robots with CycleGAN, I tried a similar method to produce fake images of divers. However, because of the mobility of human arms and legs, there isn't a consistent shape to each diver. For this reason, CycleGAN struggled to generate fake diver images. Since the synthetic generation of robot imagery worked the best, it became the focus of this project.

IV. Reflection

Overall, I had a very positive experience with the UROP program this summer. It was difficult at times, particularly due to the COVID-19 virus, but I was fortunate to have been able to work on my project from home. It was very enjoyable to work on a separate topic that was independent enough for personal exploration, but still relates to the projects that I'm fortunate to be helping out with in the IRV Laboratory.

References

[1] https://arxiv.org/abs/1703.10593

[2] https://blog.zenggyu.com/en/post/2018-12-05/retinanet-explained-and-demystified/

[3] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html