DS 5500—Phase 1 Project Report

Mantaray Classification And Localization

Team: Pavan Choudhari, Devanshi Sanghvi, Rishabh Shanbhag

Motivation

Scientists predict that more than 1 million species are on track for extinction in the coming decades. We are in dire need of economic funding (around $100 billion dollars) to help save 50% of endangered species by the end of 2050[1]. Global campaigns and governments are actively trying their best to conserve our biodiversity by introducing international laws and conservation strategies. Our goal is to aid this effort, even so slightly, by assisting researchers and biologists involved in studying and conserving the manta ray species of fish. We will be working on a manta ray dataset to classify and localize the fishes. Existing techniques do not work well with poor image quality. Usually they only help in classification, but don't do a great job at highlighting the location (localizing)[2] .

Dataset

The Manta Ray dataset contains 5085 manta ray images. The training set has about a hundred noisy, and irrelevant images, which the team decided to remove manually from the training set. The dataset provides annotated bounding boxes, which are in a COCO (Common Objects in Context) format[3].

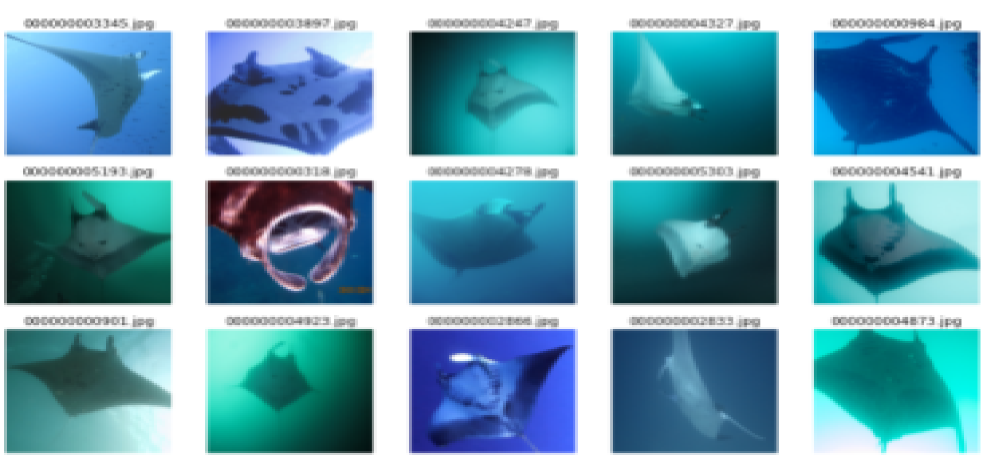

Fig1: A peek at the dataset.

Methodology

The method that we have explored covers both supervised and unsupervised approaches. We have worked on Foreground/Background Separation, Image Segmentation, Manta Ray Classification, and Automatic Background Detection. We will explore these topics further below, discussing more about the risks and the achievements.

Low Risk Approach and Result

For the low risk task, the team worked on multitudes of tasks discussed further. The team implemented different data generators that apply image processing best practices to prepare the data for the deep learning models. This method helped us to deal with restricted memory and GPU usage problems. Since we have a single class of images, the team worked on exploring different methods to generate the second class of images. The use of Photoshop was first explored, and despite the excellent results, it is not a feasible approach as it takes half an hour to process a single image. Since time was of the essence, the team used an image cropping script to split the training data into multiple images, and these images would in turn be the second class. Finally, we experimented with different image processing techniques to remove the background in the images. To automatically remove the background from an underwater image, which had occlusions and weak illumination, is a difficult task, since the approach has to automatically distinguish between the foreground and background images. Our experiments suggest that traditional image processing techniques require human expertise to adjust the different cutoff metrics for each image, and these metrics do not transfer across all images. Whereas, DeepLabV3, a deep learning model, trained on millions of images, is able to infer the background and foreground in an image, with a 0.93 IOU (Intersection Over Union Metric).

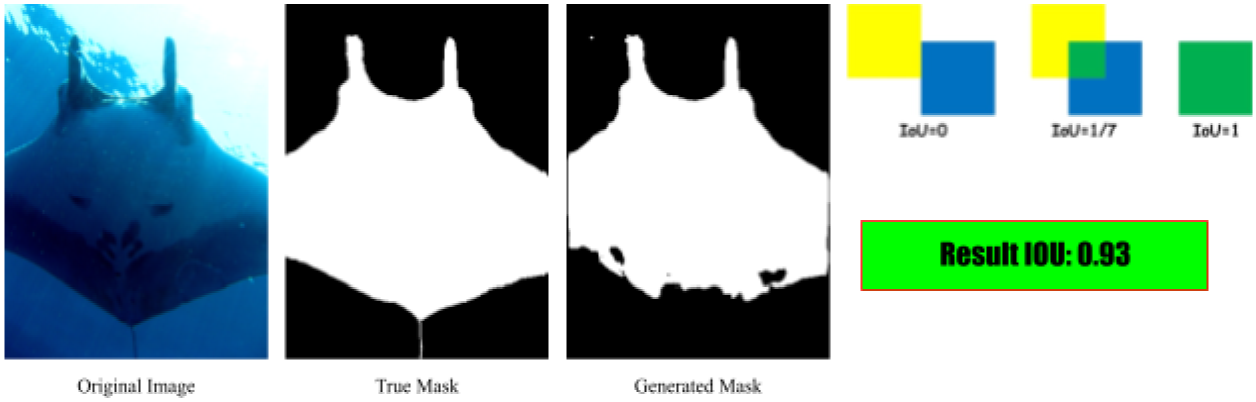

Fig.2 Deep Lab Model IOU Metric Evaluation.

Medium Risk Approach and Result

As a medium risk, we focused on:

- Unsupervised Image Clustering
- VGG16 Model for Manta Ray Classification

In the unsupervised clustering methods, we explored KMeans, and ResNet + KMeans to cluster the manta ray images. KMeans uses all the pixel values in an image to cluster, therefore it takes more time. But, using ResNet as a feature extractor, reduces the input size, and increases the speed of the computation. The best cluster size is 5, based on the elbow method, refer to fig 8.

Next, we trained a VGG-16 classifier[4] using the transfer learning approach. We are fine-tuning the model to identify a manta ray in an image. The model achieves a 67% accuracy at 7 epochs (2 hours train time). The model is not impressive, and we are working on improving the model by changing the model architecture, and training the model for a longer time.

High Risk Approach and Result

Our high risk goal was initially to localize the position of the manta ray in the image. We are exploring the use of YOLO V3 to train the model on a localized bounding box, but the task has proven to be difficult due to limited availability of GPU resources. Therefore, we challenged ourselves with another high risk goal that is to build a Siamese Network to compare and contrast similar images from different ones. Refer to fig 7 to understand the data preprocessing method for Siamese Network. We built a Fully Connected Neural Network as the base model[5], with a custom loss function called contrastive loss. This loss function tries to minimize the distance between similar images, and increase the distance between dissimilar images.

Results

We achieve a training accuracy of 94.08% and 79.32% accuracy on the test set. From the confusion matrix in fig 6. The test F1 score is 0.79. Next, we focus on finding similar images when given an anchor image. The model uses euclidean distance output to find similar images[6].

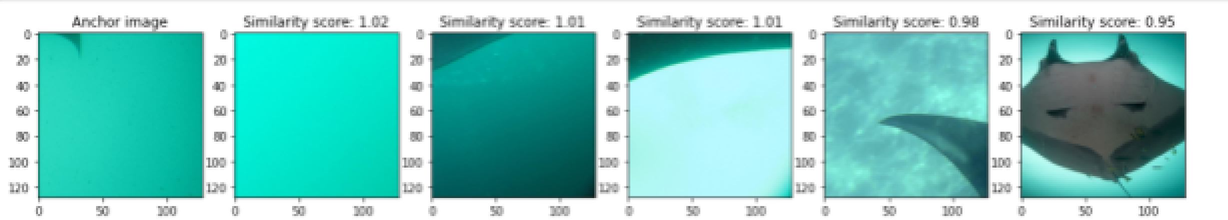

Fig 3: Top-5 similar directory images compared to background anchor image

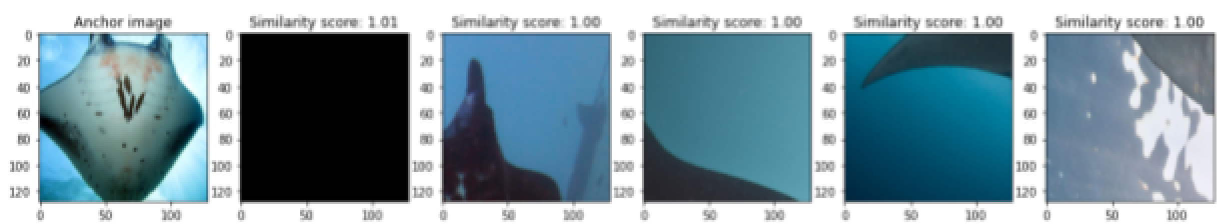

Fig 4: Top-5 similar directory images compared to mantaray anchor image

Conclusion

Biodiversity extinction is an important issue that communities around the globe are working together to abate, and new efforts are being introduced to conserve the endangered species. We are working on improving our classifier, and exploring the use of YOLO V3 to build a better model that focuses on a bounded area rather than the whole image.

References

[1] Halting the extinction crisis. Halting the Extinction Crisis. (n.d.). Retrieved February 7, 2022, from [https://www.biologicaldiversity.org/programs/biodiversity/elements_of_biodiversity/extinction_crisi s/index.html](https://www.biologicaldiversity.org/programs/biodiversity/elements_of_biodiversity/extinction_crisis/index.html)

[2] Moskvyak, Olga, et al. "Robust re-identification of manta rays from natural markings by learning pose invariant embeddings." 2021 Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2019.

[3] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.

[4] Moskvyak, Olga, et al. "Robust re-identification of manta rays from natural markings by learning pose invariant embeddings." *2021 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2019.

[5] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." *ICML deep learning workshop*. Vol. 2. 2015.

[6] Github Project Code File: https://github.com/pavan-web-dev/project

Appendix

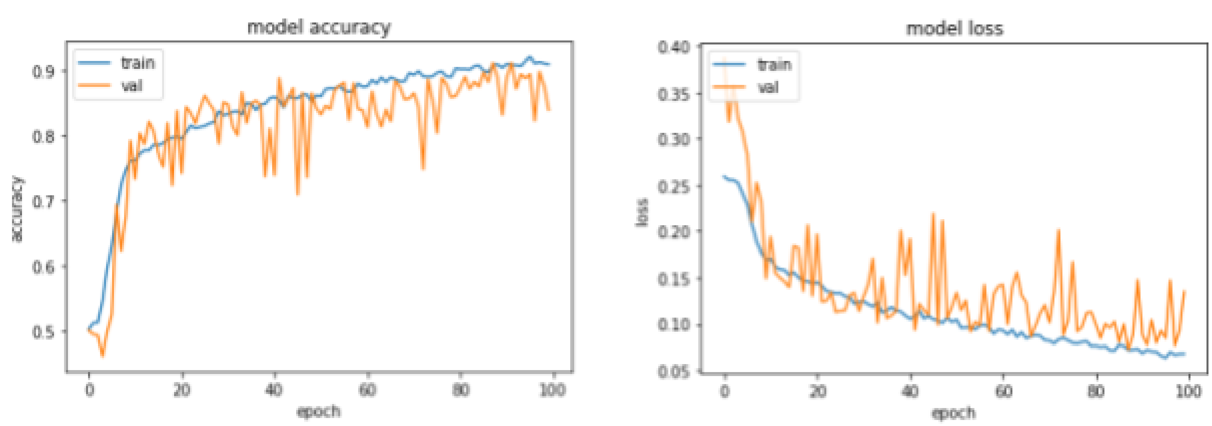

Fig 5: Siamese Network Model Accuracy & Loss

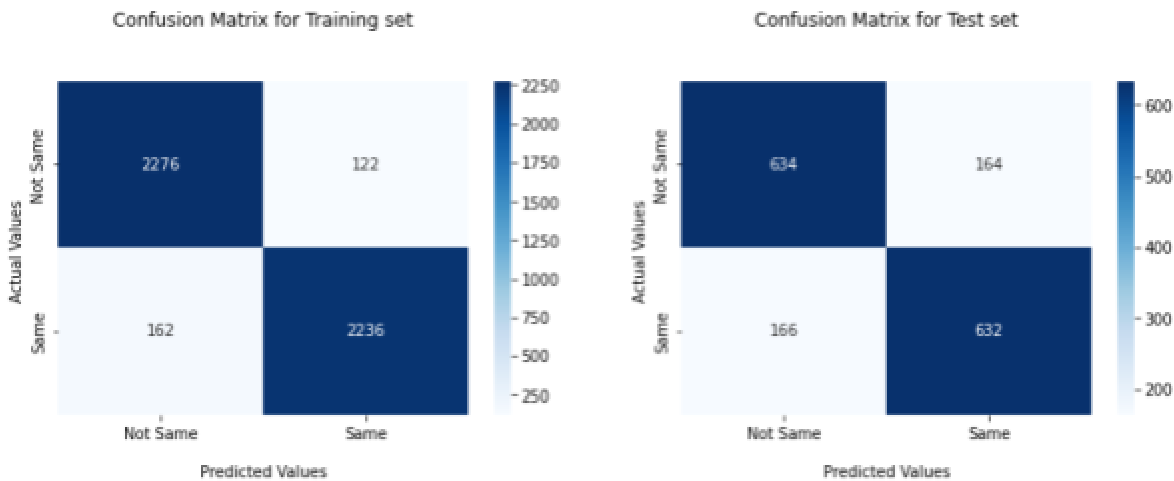

Fig 6: Confusion Matrix for Siamese Network (Training and Testing)

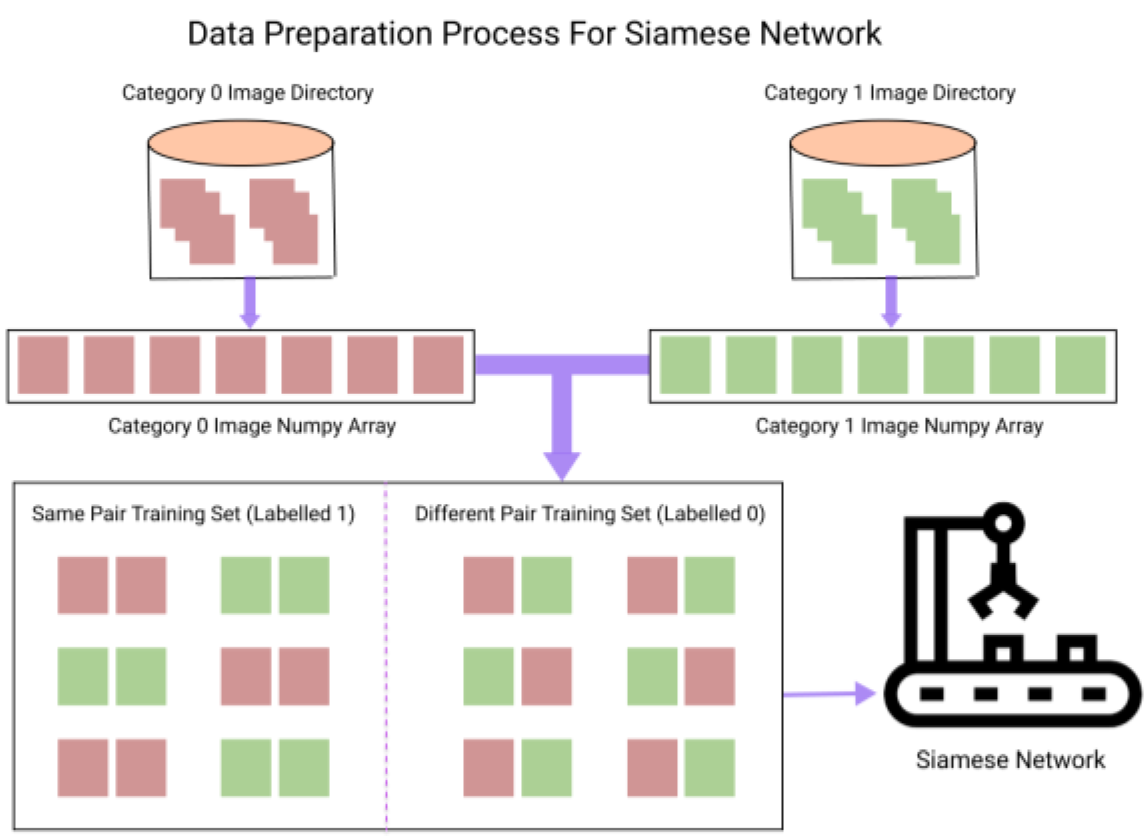

Fig 7: Siamese Model Data Preparation

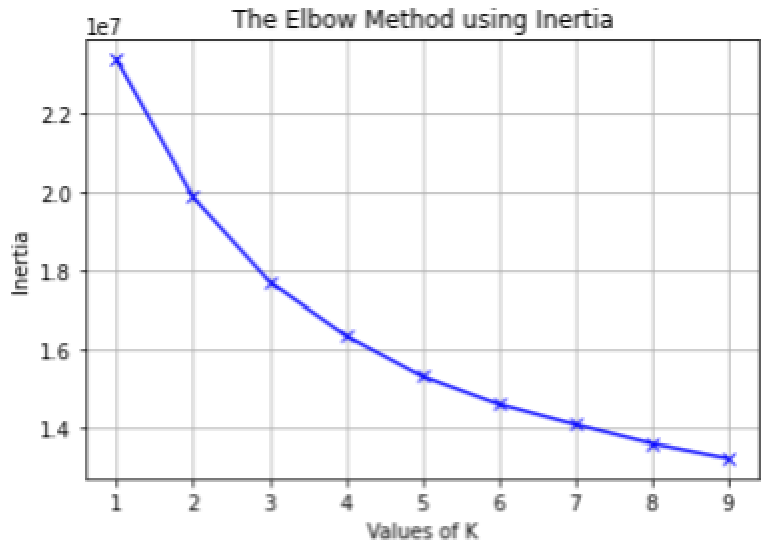

Fig 8: ResNet + KMeans find best k by elbow method