Fine-tuning convolutional neural networks to classify marine protists

Anissa Garcia[1], Yulisma Martinez[1], Andrea Gonzalez[1]Alyssia Gonzalez[1], Brooke Wright[1], Eric C. Orenstein[2], Darcy A. A. Taniguchi[1]

[1]Department of Biological Sciences, California State University San Marcos [2]Scripps Institution of Oceanography, University of California, San Diego

Introduction

- Marine protists are single-celled organisms that make up the base of the ocean's food web.
- Marine protist populations respond rapidly to environmental changes.
- *Lingulodinium polyedra* is a type of mixotrophic dinoflagellate that can make bioluminescent blooms on the Pacific coast (Fig.1).
- Ciliates are unicellular consumers and important grazers.
- These organisms are hard to study due to being microscopic and many lack pigments.
- These organisms are also abundant making them time consuming to study; teaching a computer to classify images can automate this process, cutting down on classification time.

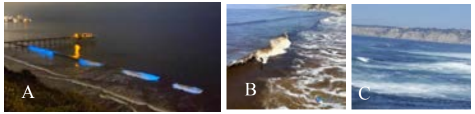

Fig. 1. Panel A displays a *Lingulodinium polyedra* bloom in 2020 at night, by Scripps Memorial Pier. Photo by Andrea Gonzalez. Panel B displays *Lingulodinium polyedra* bloom in 2020 during the day. Photo from From scripps.ucsd.edu. Panel C displays a typical view of La Jolla shores. Photo by Anissa Garcia.

Goal

To train and use a machine learning system that classifies images, a convolutional neural network (CNN), specifically to identify plankton before and during a bloom.

Methods

- A camera-microscope system, the Scripps Plankton Camera System, located off the Scripps Memorial Pier in La Jolla, California, is collecting continuous images of particles and plankton.
- We are using images taken from the Scripps Pier from every Tuesday of 2018 and 2019 to make a training set (Fig.2).
- A total of 258,660 images were classified by humans into 4 categories (Cilate, *L. polyedra,* Questionable, and Other).

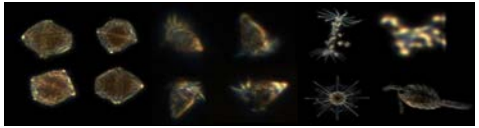

Fig. 2. Example images of plankton from the Scripps Plankton Camera System (spc.ucsd.edu) LEFT: *L. polyedra* MIDDLE: ciliates RIGHT: other

- These images were used in an existing CNN to fine tune it to classify our plankton categories of interest (Fig. 3). We used the InceptionV3 network.
- A second set of images were run through the fine-tuned CNN and classified by humans to see how well the the classifier does with novel data.
- All images in the training and novel data set were quality controlled. This quality control was done by having at least two people classify the same set of images.

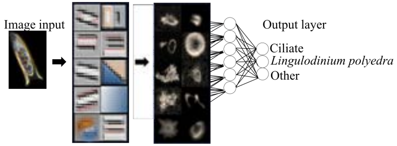

Fig. 3. Schematic of a convolutional neural network (CNN). An image is fed to the trained network. The image will go through different convolutional layers in the network. The network will then identify the image as either a Ciliate, *L. polyedra*, or Other.

Results

Before (02/06/2020)			
	Ciliate	*L. polyedra*	Other
Individual A	2.06%	0.05%	97.89%
Individual B	3.80%	0.18%	96.20%
Trained Network	14.06%	8.40%	76.90%
During (04/17/2020)			
	Ciliate	*L. polyedra*	Other
Individual A	1.38%	9.44%	89.19%
Individual B	4.78%	29.14%	66.08%
Trained Network	16.00%	33.00%	51.00%

Table 1. Comparisons of image labeling between humans and a trained network. Includes images for before a bloom (02/06/2020) and during a bloom (04/17/2020) in La Jolla, CA.

- The trained network is currently 86% accurate.
- When using the novel bloom data, there was a trend classifying *L. polyedra* and other images, but not for ciliates.
- The number of *L. polyedra* increased during a bloom.
- There were few ciliates before and during a bloom.
- There were more other images classified than ciliate and *L. polyedra* combined.

Next steps

- Continue to collect images in our categories to further train the CNN
- Input correction factor to increase accuracy
- Create a time series for Ciliates *L. polyedra*
- Find a detectable threshold for *L.polyedra*
- Add in new categories of plankton, including different groups of dinoflagellates

Acknowledgements:

The Titan Xp used for this research was donated by the NVIDIA Corporation

References:

1. Calbet, A., & Landry, M. R. (2004). Phytoplankton Growth, Microzooplankton Grazing, and Carbon Cycling in Marine Systems. *Limnology and Oceanography, 49*(1), 51-57. doi:10.4319/lo.2004.49.1.0051
2. Orenstein, E. C., & Beijbom, O. (2017). Transfer Learning and Deep Feature Extraction for Planktonic Image Data Sets. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. doi:10.1109/wacv.2017.125
3. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2016.308