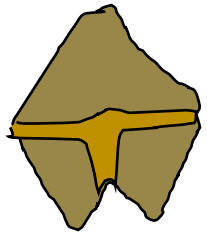
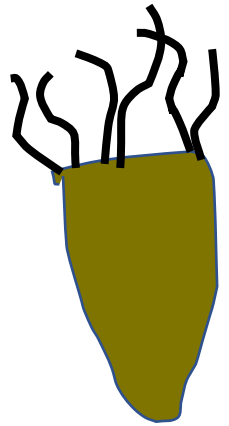


Classifying coastal unicellular plankton using machine learning

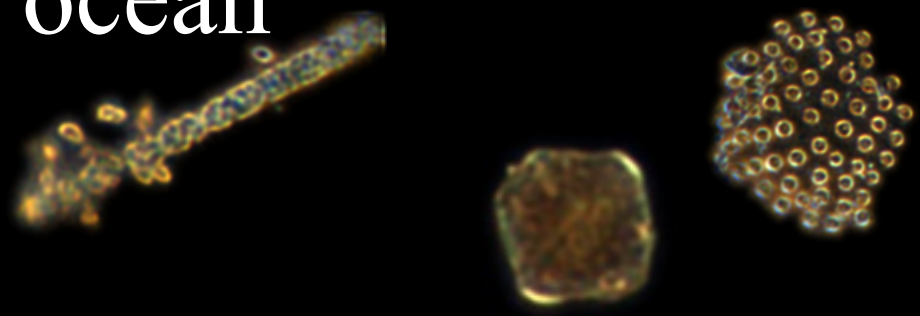


Darcy Taniguchi*, Yulisma Martinez Martinez*, Anissa Garcia*, Andrea Gonzalez*, Eric Orenstein[+]

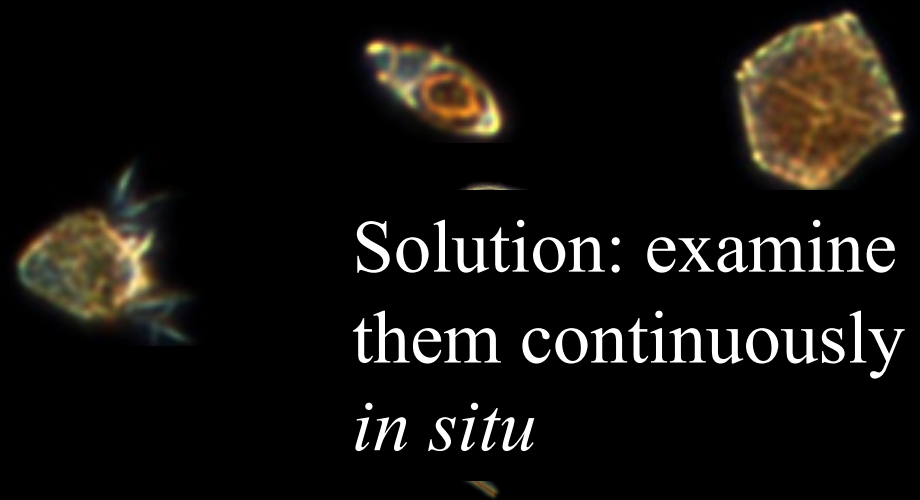
*Biology Department, California State University San Marcos
[+]Scripps Institution of Oceanography, University of California, San Diego

Aim is to examine changes in community of marine plankton, drifters of the ocean

- Plankton produce about half of the oxygen you breathe
- Plankton support the food web in the ocean



- Understanding how and why their populations change is challenging
- They are small (10-500 μm; 1 human red blood cell is ~8 um)
- Traditional sampling and analysis techniques can ruin cells and are just a snapshot in time



Solution: examine them continuously *in situ*

spc.ucsd.edu

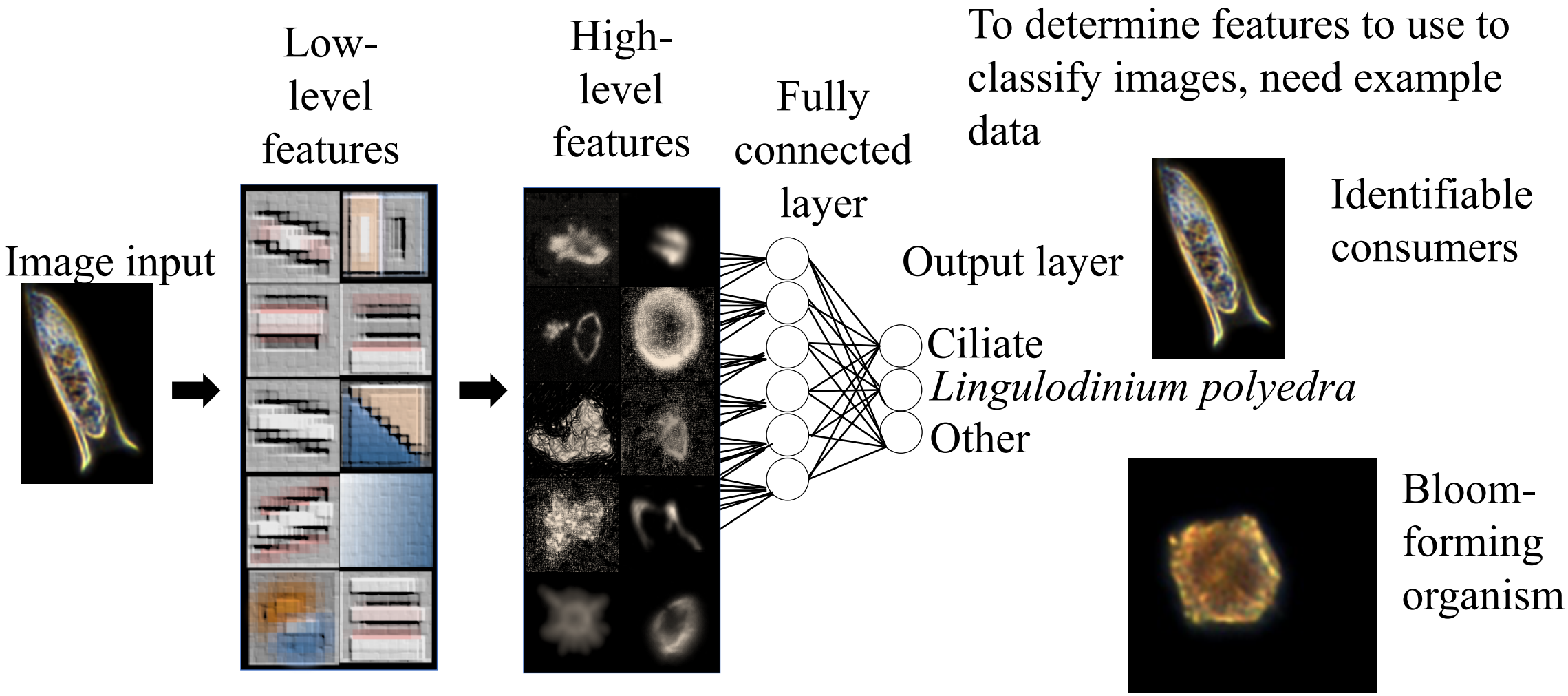
Scripps Plankton Camera System collects images of plankton and particles *in situ* off the Ellen Browning Scripps Memorial Pier in La Jolla, CA

Microscope camera system

- Particles and plankton remain unmanipulated
- Sampling is continuous
- Many images are produced that cannot be efficiently analyzed…by a person

Mkeipper,https://commons.wikimedia.org/wiki/File:Scripps_Pier_La_Jolla_CA.jpg

An artificial neural network, specifically a convolutional neural network, can be trained to automatically classify images

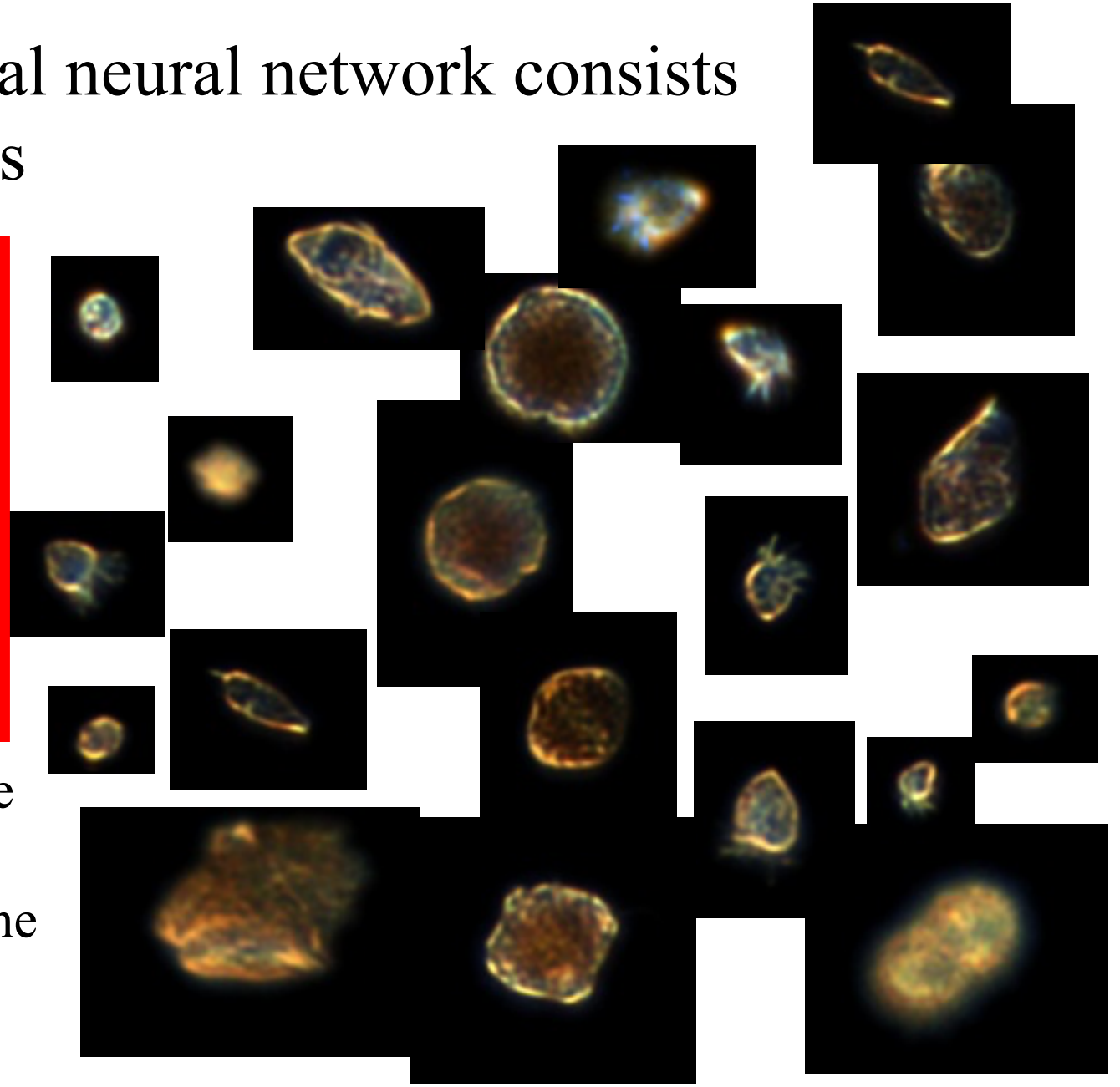


Example data to train the artificial neural network consists of images categorized by humans

- We looked through 129,330 images to classify them as ciliates, *L. poleydra,* or other

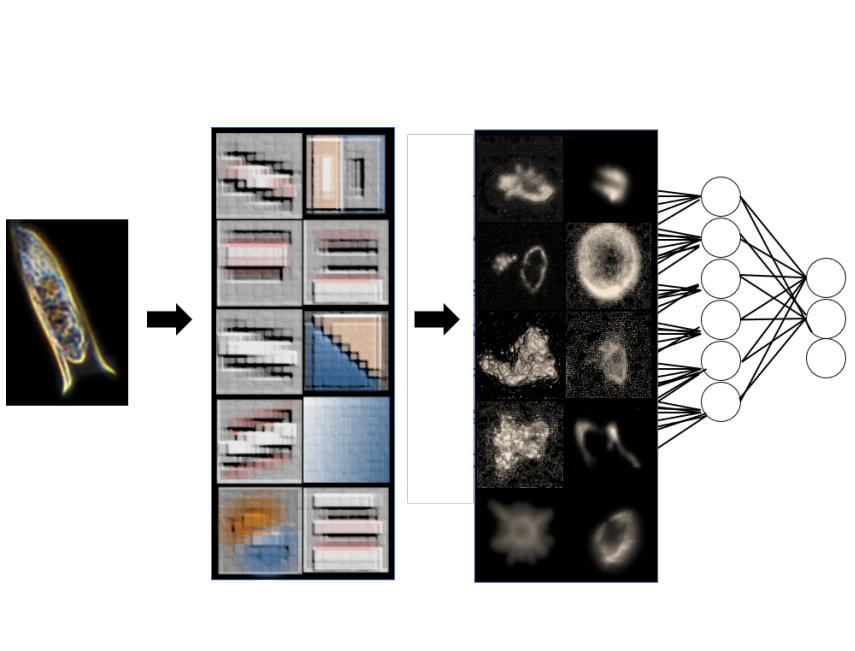
- For quality control, all images were examined by at least two people (≥258,660 images)

- This is one of the biggest areas where the lab needs help
- How many people classified the image the same way? Which images are different?
- It really matters because we need a training set to train the CNN

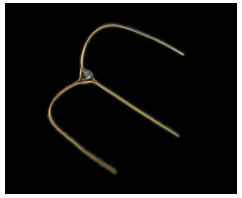
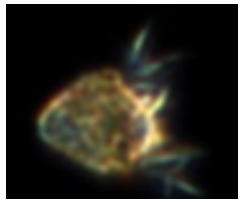
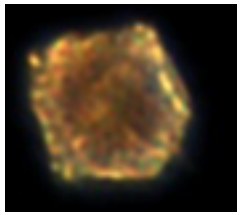


Training data of 450 images in each a category used to fine-tune an existing CNN InceptionV3

- Without any input from a person, trained network classified ~86% of validation images correctly
- Large *L. polyedra* bloom in April-May of 2020 off La Jolla
- Trained CNN classified novel data from February 2020 (no bloom) and April 2020 (bloom)



Percent of images in each category

	Before bloom	During bloom
	76.9%	51.0%
	14.6%	16.0%
	8.4%	33.0%

Next steps

- Continue gathering more training data
- Input correction factor to increase accuracy
- Compare machine classification with human classification for novel data
- Create a time series of these organisms
- Add in new categories of plankton

- This is another area where the lab needs help
- How many images in each category were classified the same by the machine and the human? Are there differences among categories?

Thanks to past and present
Taniguchi Lab members
Keomony Diep
Brooke Wright
Alyssia Gonzalez