Classification of Algal Blooms

Darian Madere

Background

- **Phytoplankton** are single-celled algae that live in freshwater, saltwater, and everywhere in between
- **Algal blooms**, big explosions of **phytoplankton growth**, happen naturally every late spring and early fall all around the world
- Sometimes algal blooms grow out of control. This is called a **Harmful Algal Bloom**.

Cylindrotheca closterium, a marine diatom phytoplankton

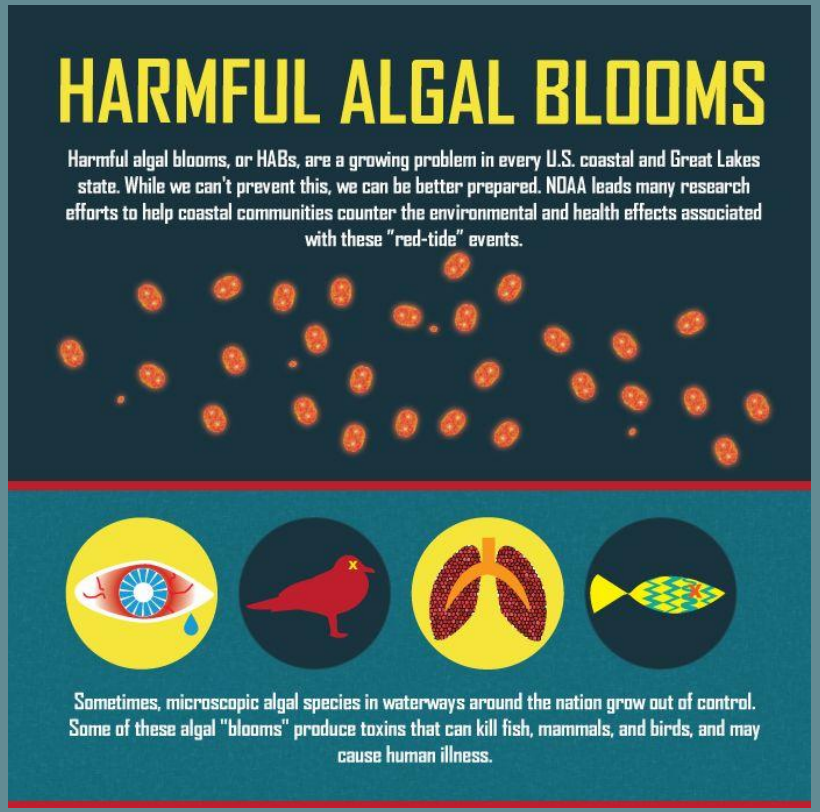
Harmful Algal Blooms

- Harmful Algal Blooms can happen under a combination of circumstances:
  - Excess nutrient from land based runoff
  - Influx of freshwater
  - Higher than normal temperatures
  - Low turbidity
- Harmful Algal Blooms pose a threat to human health and ecosystem health

Red tide, a type of harmful algal bloom

Why Bloom Monitoring Matters

- Early detection of algal blooms helps scientists and local officials make public and ecological health decisions before they become a problem with minimal human interventions
- Scientists can sample blooms as soon as possible to determine if they are toxic, and make important decisions regarding public health:
    - beach closures
    - fisherman warnings
    - seafood warnings



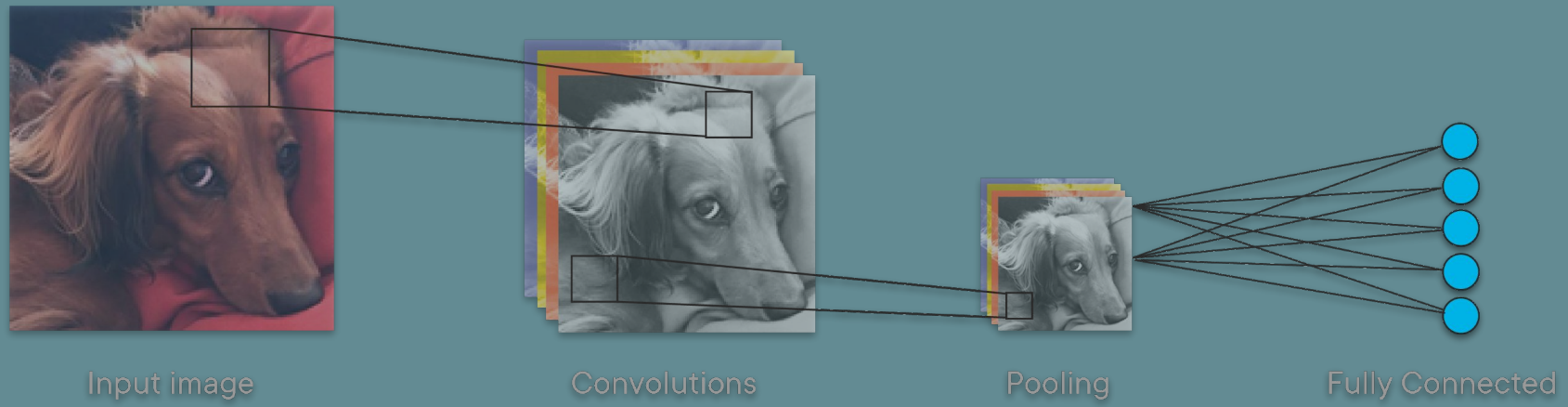
Goal

My goal is to create a model that classifies algal blooms using publicly available imagery obtained from Google

Algal blooms come in all different shapes, sizes and colors, and this highly varied dataset will hopefully lead to a model that is able to classify algal blooms in all forms

Convolutional Neural Networks

- Convolutional Neural Networks (CNN) are a kind of **deep learning** algorithm consisting of alternating **convolution** and **pooling** layers resulting in a **fully connected layer**
- The **convolution layers** are what allows a CNN to detect patterns in images



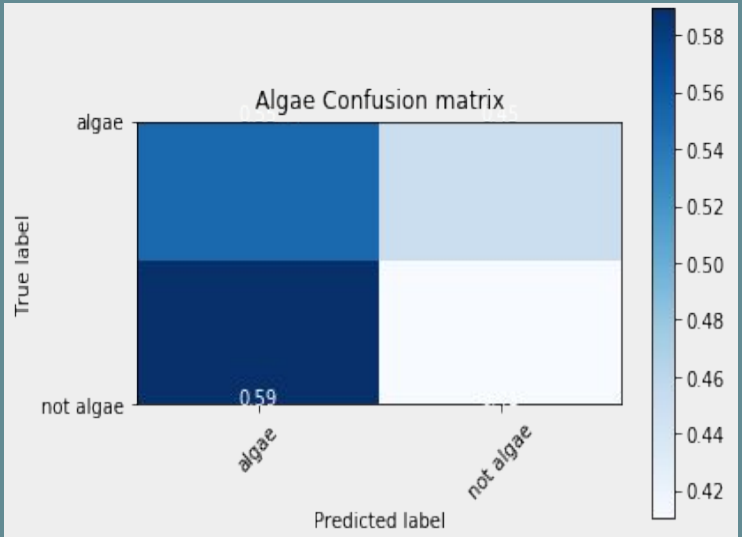
Model Details

- ~1400 images
- Split 80/20/20 into train/test/validation image sets
- 2 classes: Algae and Not algae
- Scaled to 140x140
- Augmentations
  - 90 degree random rotation
  - Random brightness adjustment
  - Random horizontal flipping

Algae

Not Algae

Model Performance



Confusion Matrix

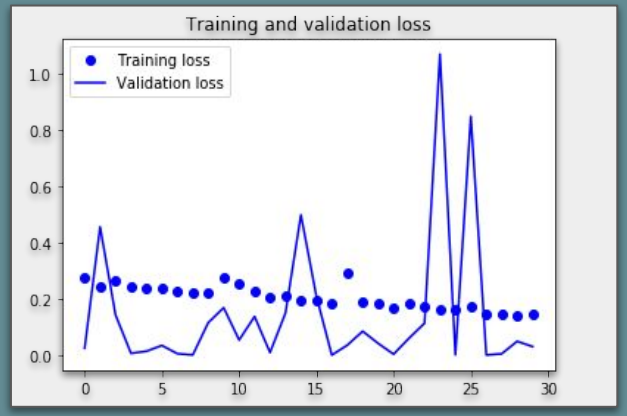
Model Accuracy: 86%

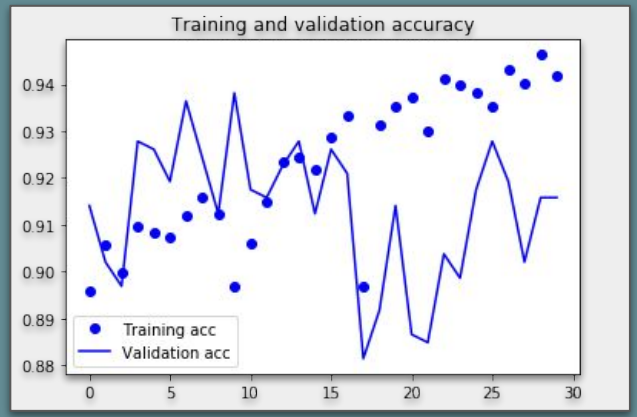
Model Loss: 1.0412065982818604 (yikes!)

- Accuracy is determined after a model is done learning - it tells you how accurate your model's prediction is compared to the true data
- Loss is a summation of the uncertainty of each prediction made for each example in training or validation sets

Model Improvements

- Add more data
  - Manual cropping
  - Saturation adjustment
  - Add more types of augmentation
- Adjust model weights to decrease loss
- Decrease epochs in an effort to decrease loss






Email: dalayne95@gmail.com
Github: www.github.com/dalayne95