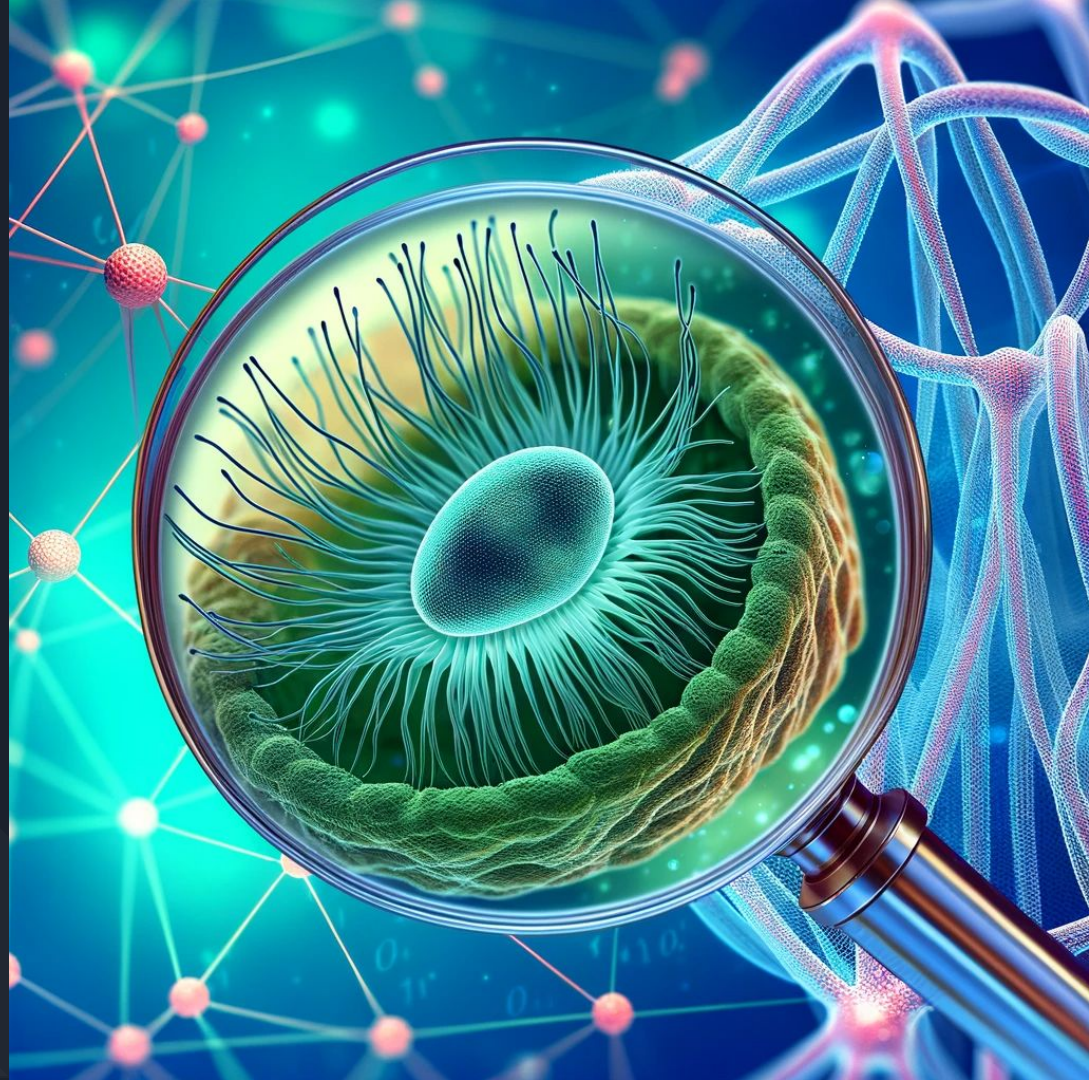
Phytoplankton classification

Philipp Meisinger

27.02.2024



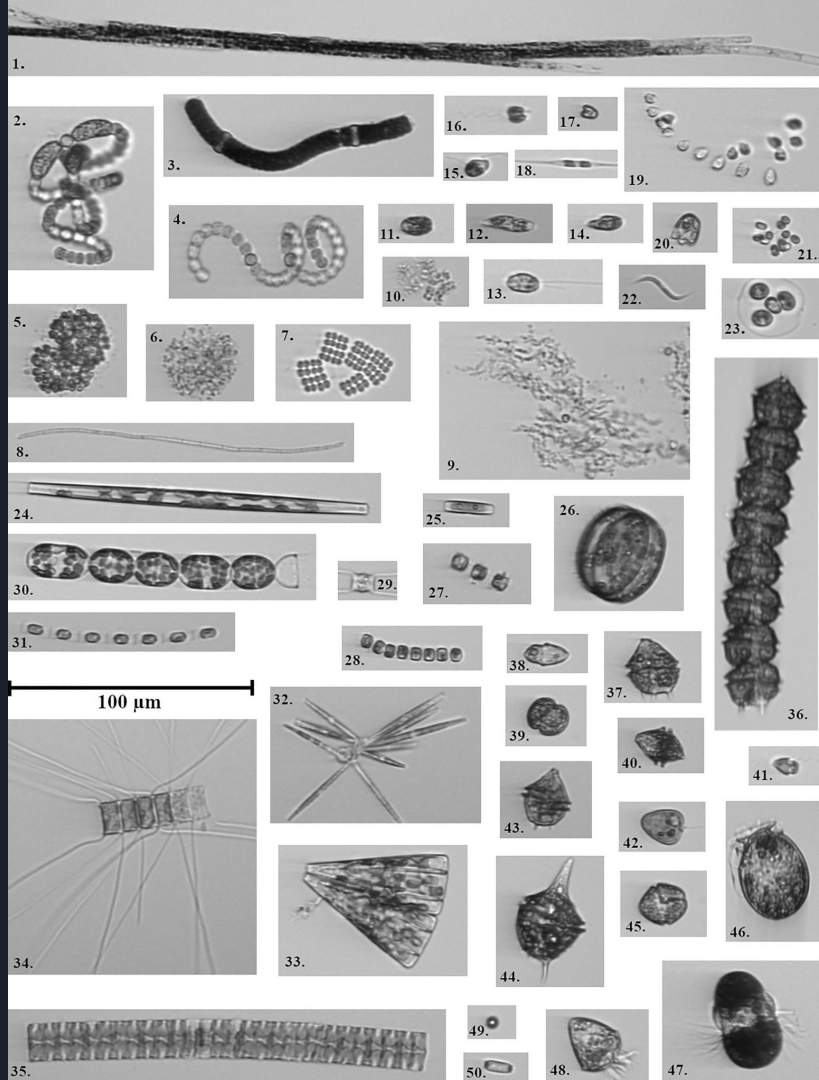
1 The Project

- Multiclass classification problem with 50 classes
- Phytoplankton classification using Convolutional Neural Networks (CNNs).

Kraft et al. 2022
Kraft K, Velhonoja O, Eerola T, Suikkanen S, Tamminen T, Haraguchi L, Ylöstalo P, Kielosto S, Johansson M, Lensu L, Kälviäinen H, Haario H and Seppälä J (2022) Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks. Front. Mar. Sci. 9:867695. doi: 10.3389/fmars.2022.867695

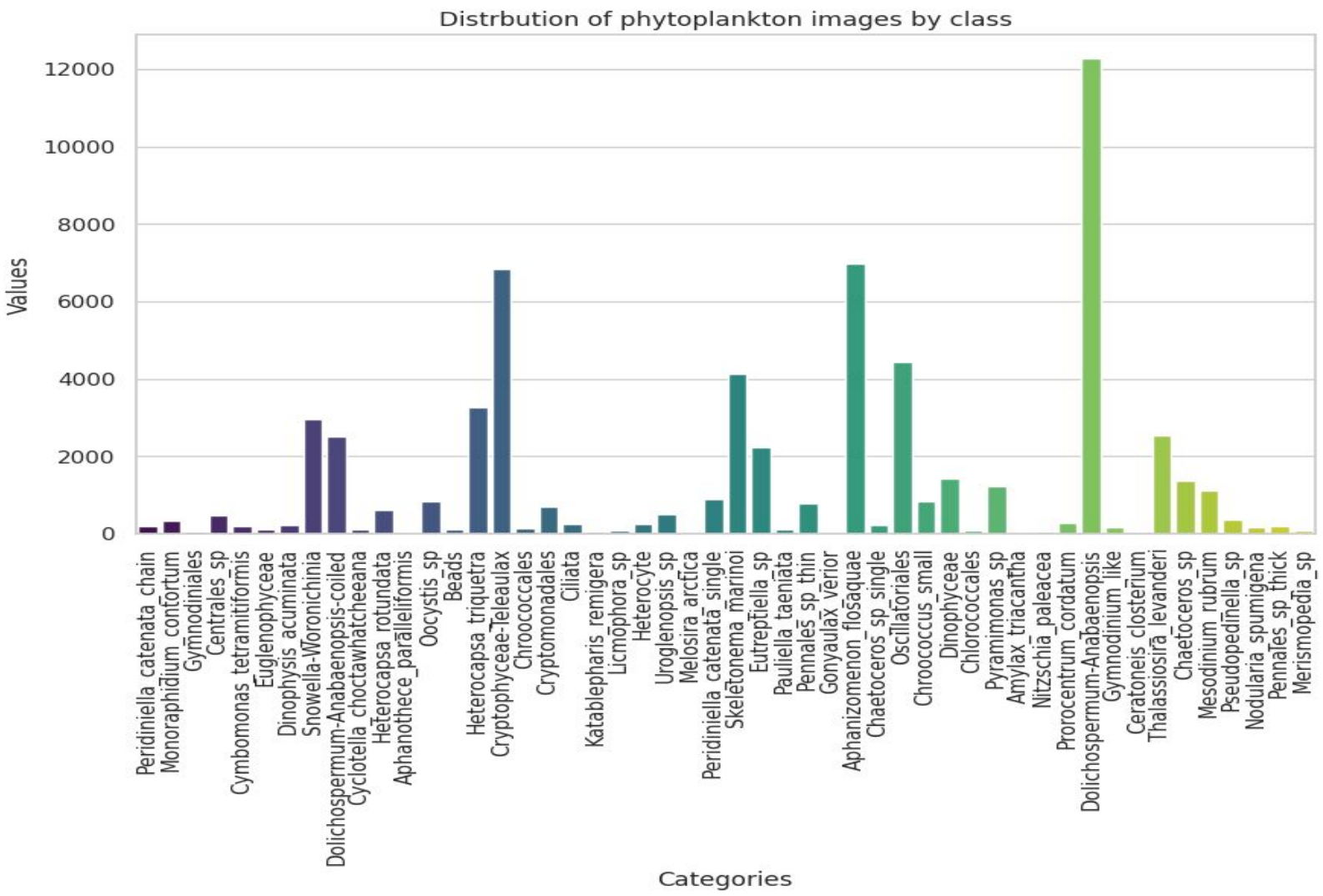
2.1 The Dataset

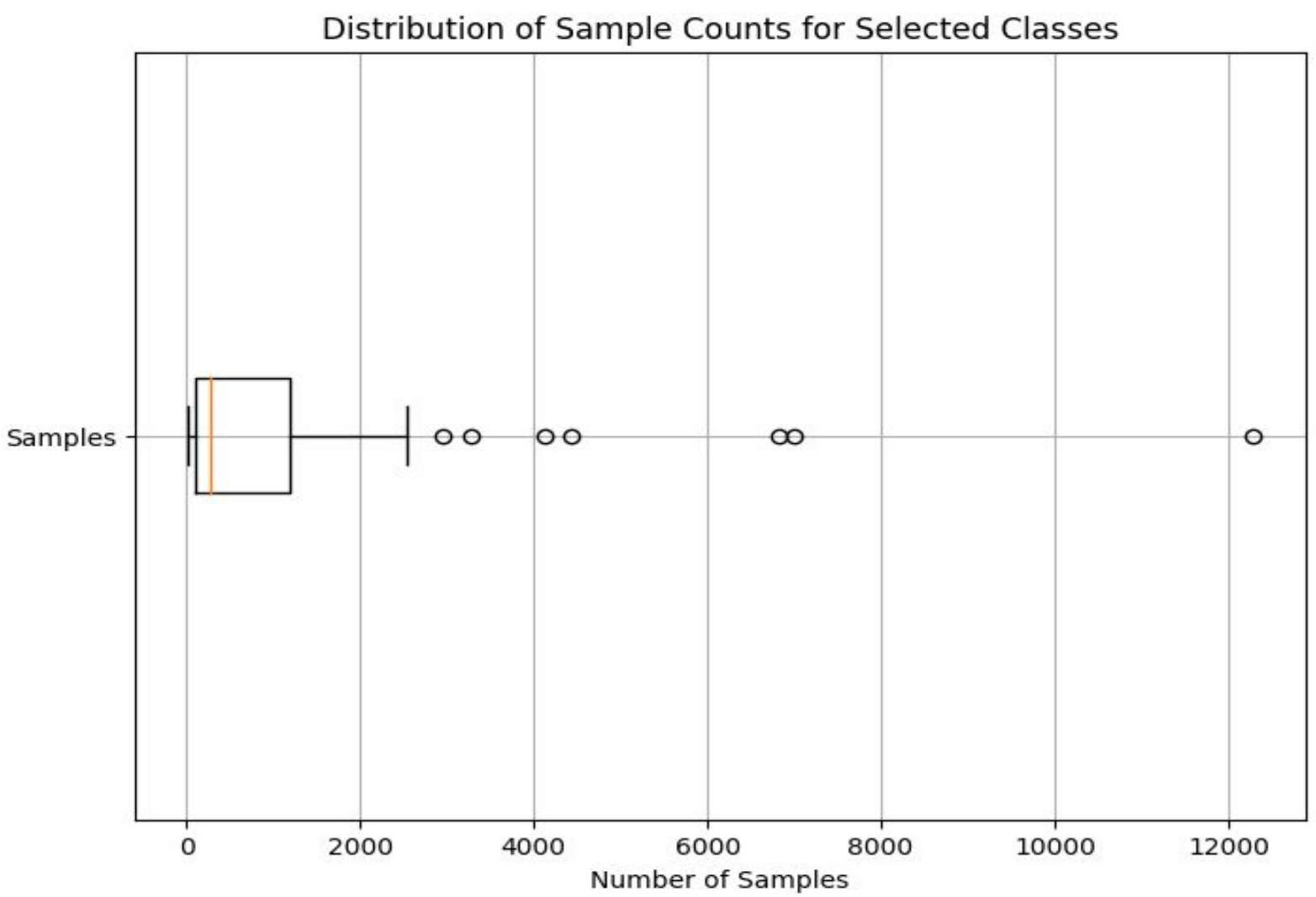
- Dataset consisting of ~63k labeled images across 50 distinct classes.
- Cropped images; non-uniform in size and dimension.
- Similarities across classes and species.
- Highly imbalanced



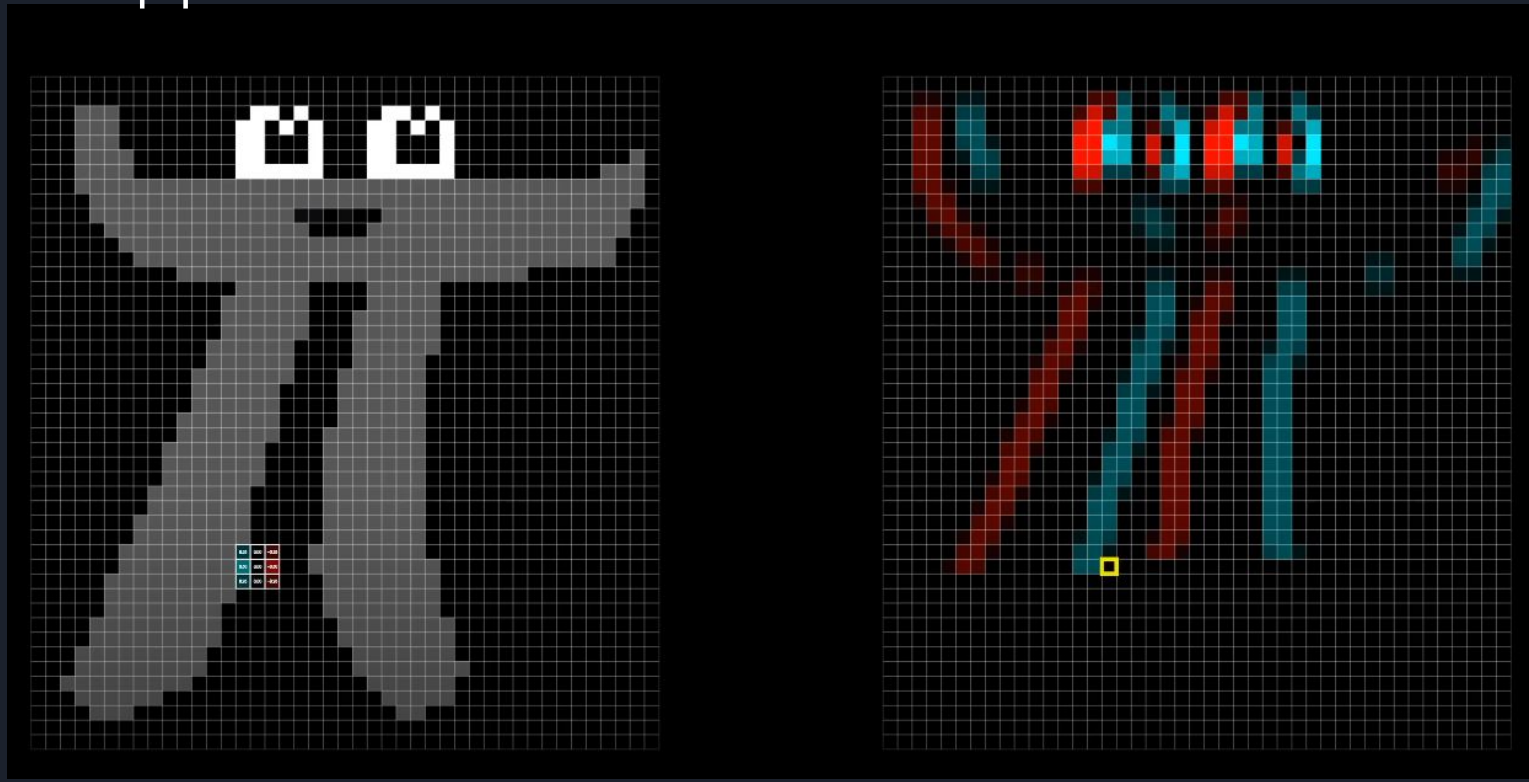
2.2 The Dataset

- Lowest # of samples: **19**
- Highest # of samples: **12280**
- Mean # of samples: **1261.48**
- Median # of samples: **269.5**
- Standard Deviation # samples per class: **2246.96**


Distrbution of phytoplankton images by class
Values
12000
10000
8000
6000
4000
2000
0
Peridiniella_catenata_chain
Monoraphidium_contortum
Gymnodiniales
Centrales_sp
Cymbomonas_tetramitiformis
Euglenophyceae
Dinophysis_acuminata
Snowella-Woronichinia
Dolichospermum-Anabaenopsis-coiled
Cyclotella_choctawhatcheeana
Heterocapsa_rotundata
Aphanothece_parallelliformis
Oocystis_sp
Beads
Heterocapsa_triquetra
Cryptophyceae-Teleaulax
Chroococcales
Cryptomonadales
Ciliata
Katablepharis_remigera
Licmophora_sp
Heterocyte
Uroglenopsis_sp
Melosira_arctica
Peridiniella_catenata_single
Skeletonema_marinoi
Eutreptiella_sp
Pauliella_taeniata
Pennales_sp_thin
Gonyaulax_verior
Aphanizomenon_flosaquae
Chaetoceros_sp_single
Oscillatoriales
Chroococcus_small
Dinophyceae
Chlorococcales
Pyramimonas_sp
Amylax_triacantha
Nitzschia_paleacea
Prorocentrum_cordatum
Dolichospermum-Anabaenopsis
Gymnodinium_like
Ceratoneis_closterium
Thalassiosira_levanderi
Chaetoceros_sp
Mesodinium_rubrum
Pseudopediniella_sp
Nodularia_spumigena
Pennales_sp_thick
Merismopedia_sp
Categories


Distribution of Sample Counts for Selected Classes
Samples
0
2000
4000
6000
8000
10000
12000
Number of Samples

3 The Approach



“**But what is a convolution?**” **by** 3Blue1Brown

4 Model results: v.1

Class range	Macro Average F1 Score	Weighted Average F1 Score
0 - 9	91%	98%
10 - 19	77%	95%
20 - 29	48%	89%
30 - 39	97%	99%
40 - 49	94%	96%
Parent model	85%	86%
Cascading Model	10.80%	10.80%

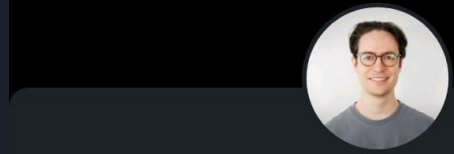
4 Model results: v.2

Class range	Macro Average F1 Score	Weighted Average F1 Score
Parent model	85%	86%
Cascading Model	10.80%	10.80%
30 classes	94%	94%
20 classes	85%	85%

Thank you!

Let’s connect!

Project Link



Philipp Meisinger

Software Engineer @ 8tronix