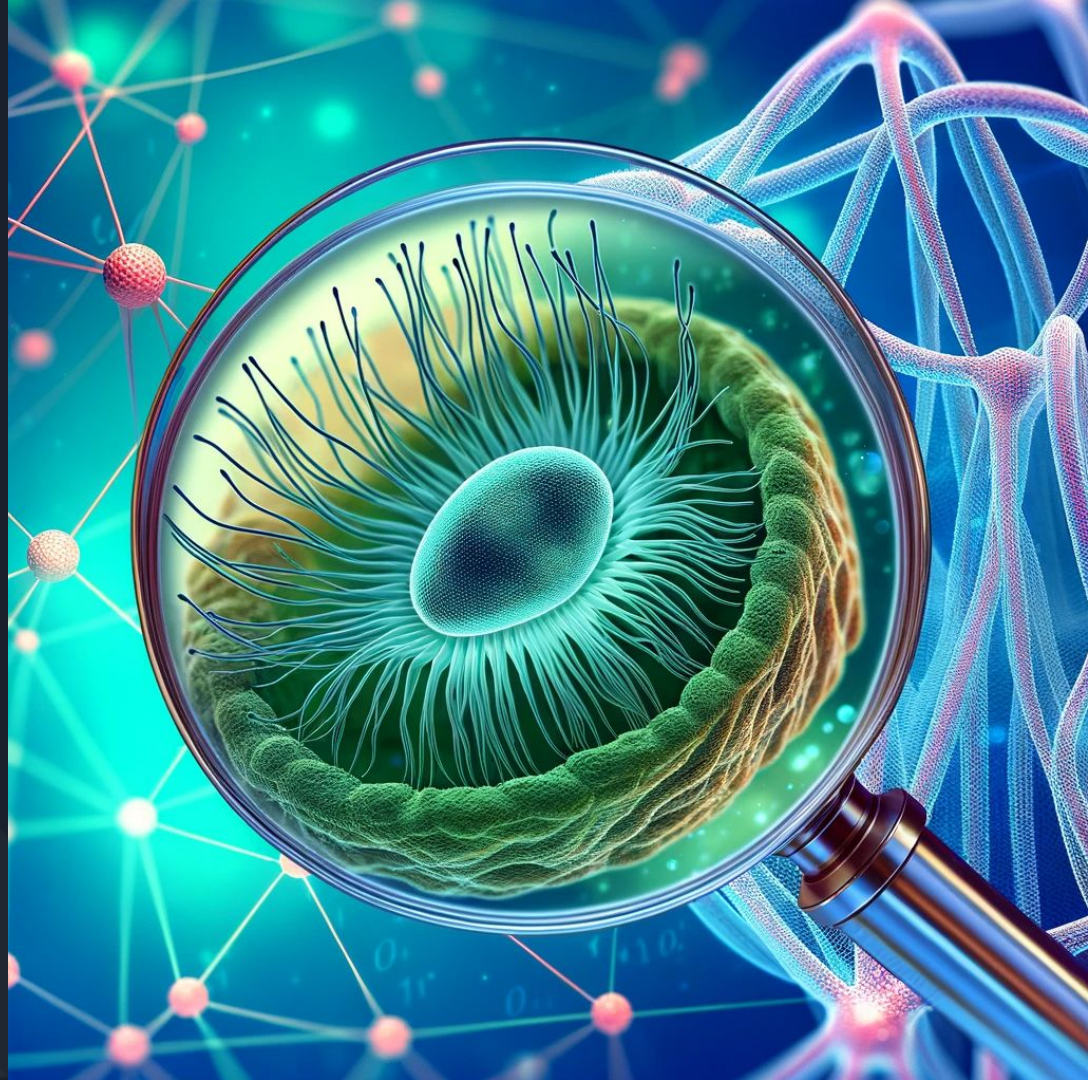# Phytoplankton classification

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.
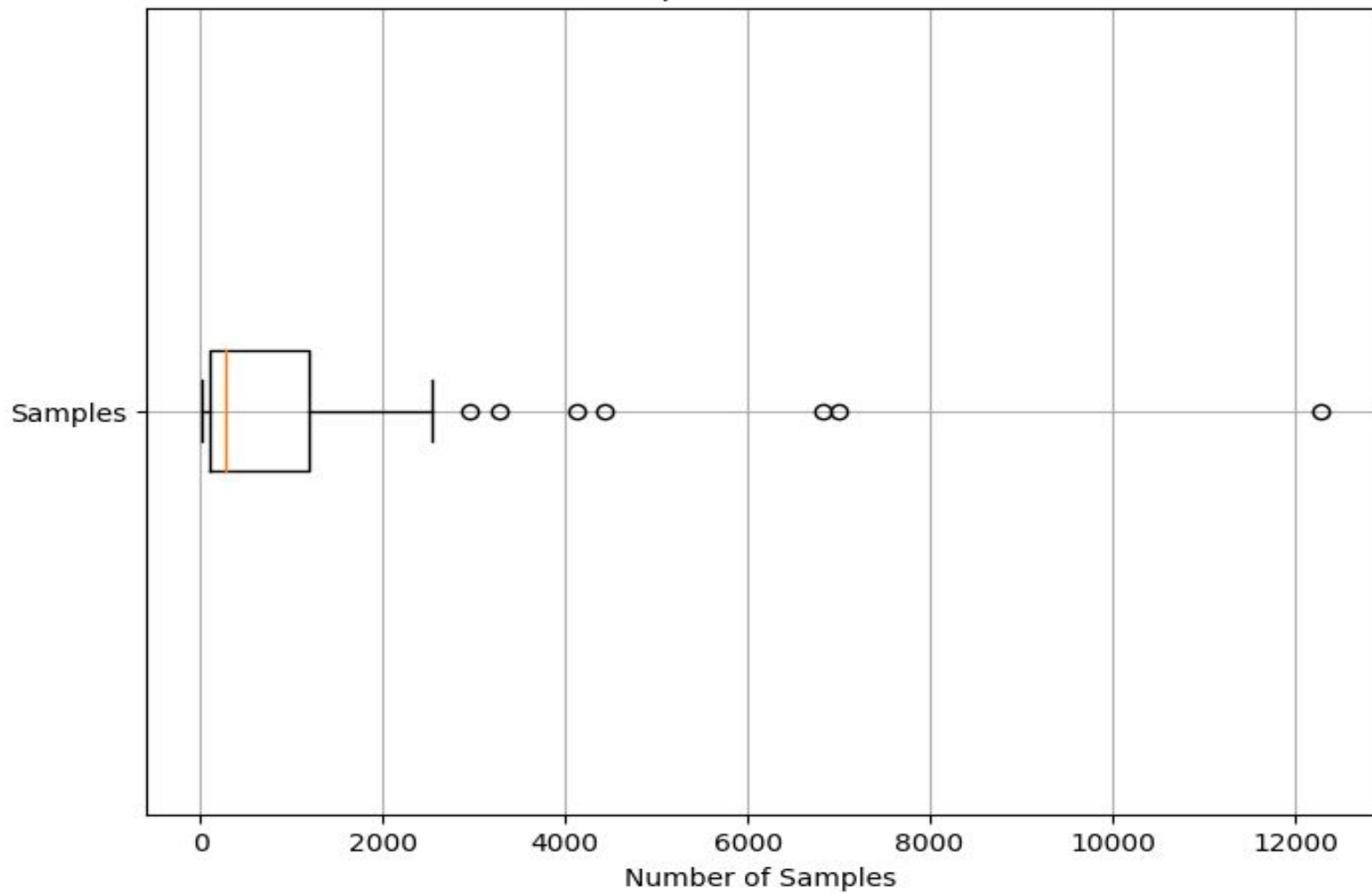- Highly imbalanced
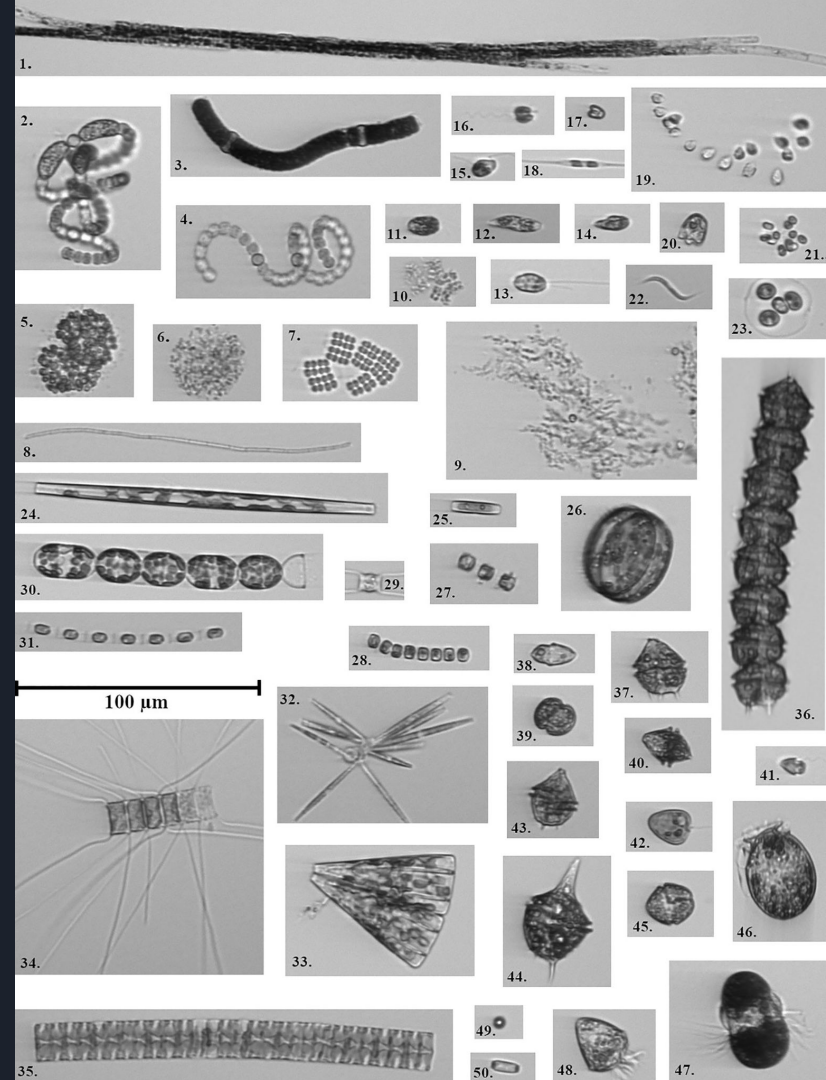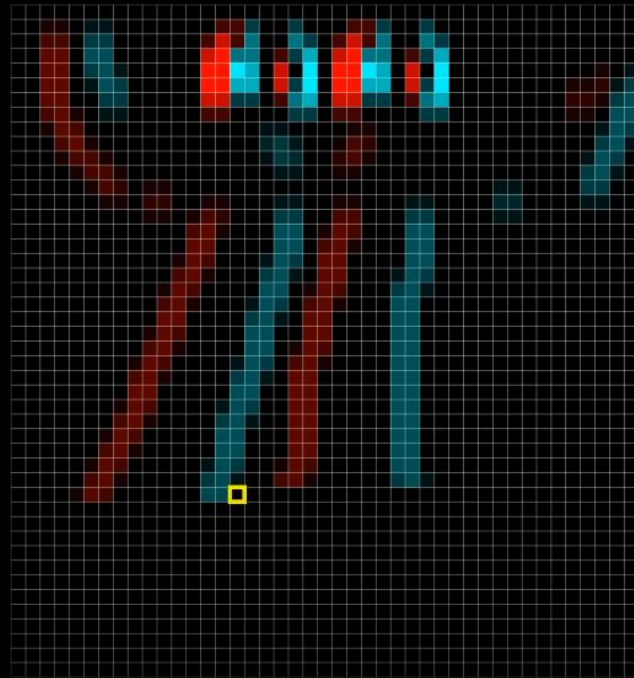
2.2 The Dataset

- 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
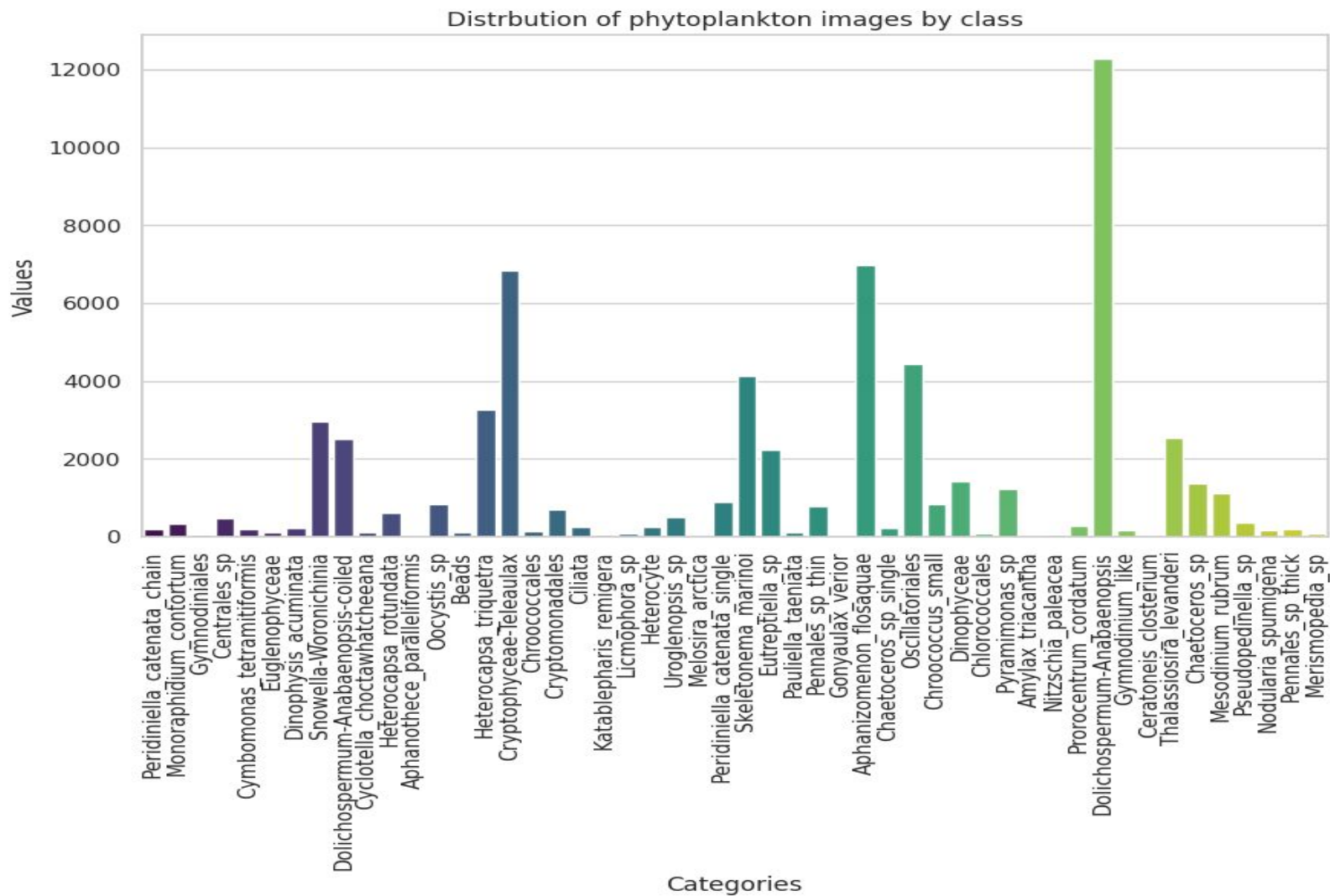
Distribution of Sample Counts for Selected Classes

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!

Project Link

Let's connect!

Philipp Meisinger
Software Engineer @ 8tronix