Self-Supervised Pretraining for Fine-Grained Plankton Recognition

Supplementary Material

S1. Pretraining dataset class overlap

The class overlaps with different datasets are shown in Table 1. In total, there are not many overlapping classes between the dataset except with DAPlankton and SYKE-Plankton_IFCB_2022, which share most of their classes.

Table 1. Overlapping Classes Between Plankton Datasets

Dataset Pair	Overlapping Classes	Count
DAPlankton [1] & SYKE-Plankton_IFCB_2022 [5]	aphanizomenon-flosaquae, centrales-sp, chaetoceros-sp, chaetoceros-sp-single, chlorococcales, chroococcales ciliata, cryptomonadales, cryptophyceae-teleaulax cyclotella-choctawhatcheeana, dinophyceae, dinophysis-acuminata dolichospermum-anabaenopsis, dolichospermum-anabaenopsis-coiled euglenophyceae, eutreptiella-sp, gymnodiniales, gymnodinium-like, heterocapsa-rotundata, heterocapsa-triquetra heterocyte, katablepharis-remigera, melosira-arctica mesodinium-rubrum, monoraphidium-contortum, nitzschia-paleacea nodularia-spumigena, oocystis-sp peridiniella-catenata (as peridiniella-catenata-single/chain in SYKE) pseudopedinella-sp, pyramimonas-sp, skeletonema-marinoi snowella-woronichinia	33
DAPlankton [1] & PMID2019 [7]	chaetoceros (chaetoceros-sp in DA), skeletonema (skeletonema-marinoi in DA)	2
SYKE-Plankton_IFCB_2022 [5] & PMID2019 [7]	chaetoceros (chaetoceros-sp in SYKE), skeletonema (skeletonema-marinoi in SYKE)	2
PMID2019 [7] & UDE Diatoms [4]	navicula	1
SYKE-Plankton_ZooScan_2024 [3] & Kaggle Plankton [2]	copepod-calanoid, copepod-cyclopoid (multiple cyclopoid in Kaggle)	2
SYKE-Plankton_ZooScan_2024 [3] & Lake Zooplankton [6]	daphnia, synchaeta (daphnia-sp, synchaeta-sp in SYKE)	2
Lake Zooplankton [6] & Kaggle Plankton [2]	diatom-chain (diatom-chain-string/tube in Kaggle)	1

S2. Results

The fine-tuning results for DAPlankton$_{LAB}$ and DAPlankton$_{SEA}$ are shown in Table 2 and Table 3, respectively. For DAPlankton$_{SEA}$, the full fine-tuning results look very similar to DAPlankton$_{LAB}$, and there is no real difference between the different pretraining methods.

Table 2. Mean accuracy and standard deviation (in %) for DAPlankton$_{LAB}$ with limited training labels.

	IFCB			CS			FC		
Model	1%	5%	10%	1%	5%	10%	1%	5%	10%
ViT-L (imagenet)	78.08 ± 3.54	95.23 ± 0.25	97.51 ± 0.26	62.87 ± 7.45	91.63 ± 1.12	95.22 ± 0.76	52.04 ± 4.39	91.72 ± 0.34	94.31 ± 0.35
ViT-L (no-daplankton)	83.83 ± 2.03	95.68 ± 0.35	97.41 ± 0.23	69.90 ± 4.73	90.30 ± 0.80	93.57 ± 0.47	79.58 ± 2.58	92.87 ± 0.33	95.00 ± 0.48
ViT-L (with-daplankton)	96.92 ± 0.37	98.29 ± 0.30	98.75 ± 0.24	91.76 ± 0.89	96.26 ± 0.52	96.97 ± 0.33	83.10 ± 1.44	93.86 ± 0.65	95.72 ± 0.64

Table 3. Mean accuracy and standard deviation (in %) for DAPlankton$_{SEA}$ with limited training labels.

	IFCB				CS			
Model	1%	5%	10%	Full	1%	5%	10%	Full
ViT-L (imagenet)	86.47 ± 1.16	96.82 ± 0.19	97.90 ± 0.13	98.91 ± 0.04	53.92 ± 6.44	85.68 ± 2.17	91.29 ± 0.59	95.80 ± 0.44
ViT-L (no-daplankton)	89.43 ± 0.16	96.35 ± 0.30	97.54 ± 0.17	98.84 ± 0.10	72.65 ± 2.72	87.32 ± 0.54	90.49 ± 0.79	95.32 ± 0.06
ViT-L (with-daplankton)	95.44 ± 0.36	97.79 ± 0.18	98.24 ± 0.14	98.85 ± 0.06	85.89 ± 0.24	91.70 ± 0.54	93.23 ± 0.49	95.57 ± 0.24

References

[1] Daniel Batrakhanov, Tuomas Eerola, Kaisa Kraft, Lumi Haraguchi, Lasse Lensu, Sanna Suikkanen, María Teresa Camarena-Gómez, Jukka Seppälä, and Heikki Kälviäinen. DAPlankton: Benchmark dataset for multi-instrument plankton recognition via fine-grained domain adaptation. In *ICIP*, pages 158–164, 2024. 1

[2] Robert K. Cowen, Su Sponaugle, Kelly L. Robinson, Jessica Luo, and Cedric Guigand. Planktonset 1.0: Plankton imagery data collected from f.g. walton smith in straits of florida from 2014-06-03 to 2014-06-06 and used in the 2015 national data science bowl (ncei accession 0127422). `https://doi.org/10.7289/v5d21vjd`, 2015. 1

[3] Joona Kareinen, Annaliina Skyttä, Tuomas Eerola abd Kaisa Kraft, Lasse Lensu, Sanna Suikkanen, Maiju Lehtiniemi, and Heikki Kälviäinen. SYKE-plankton_ZooScan_2024. `https://doi.org/10.23729/fa115087-2698-4aa5-aedd-11e260b9694d`, 2024. 1

[4] Michael Kloster, Andrea Burfeid-Castellanos, Mimoza Dani, Ntambwe Albert Serge Mayombo, Bank Beszteri, and Danijela Vidaković. UDE Diatoms in the Wild 2024, 2024. 1

[5] Kaisa Kraft, Otso Velhonoja, Jukka Seppälä, Heidi Hällfors, Sanna Suikkanen, Pasi Ylöstalo, Sílvia Anglès, Sami Kielosto, Harri Kuosa, Sirpa Lehtinen, Johanna Oja, and Timo Tamminen. SYKE-plankton_IFCB_2022. `https://b2share.eudat.eu/records/abf913e5a6ad47e6baa273ae0ed6617a`, 2022. 1

[6] Sreenath P Kyathanahally, Thomas Hardeman, Ewa Merz, Thea Bulas, Marta Reyes, Peter Isles, Francesco Pomati, and Marco Baity-Jesi. Deep learning classification of lake zooplankton. *Frontiers in Microbiology*, 12:746297, 2021. 1

[7] Qiong Li, Xin Sun, Junyu Dong, Shuqun Song, Tongtong Zhang, Dan Liu, Han Zhang, and Shuai Han. Developing a microscopic image dataset in support of intelligent phytoplankton detection using deep learning. *ICES Journal of Marine Science*, 77:1427–1439, 2020. 1