Supplementary Material DeepLOKI

1 GRAPHICAL USER INTERFACE- LABELTOOL

We have developed a graphical user interface (GUI) for presorting and manual sorting purposes, which includes a user-friendly image viewer. With this GUI proper label can be generated for training, validation and testing.

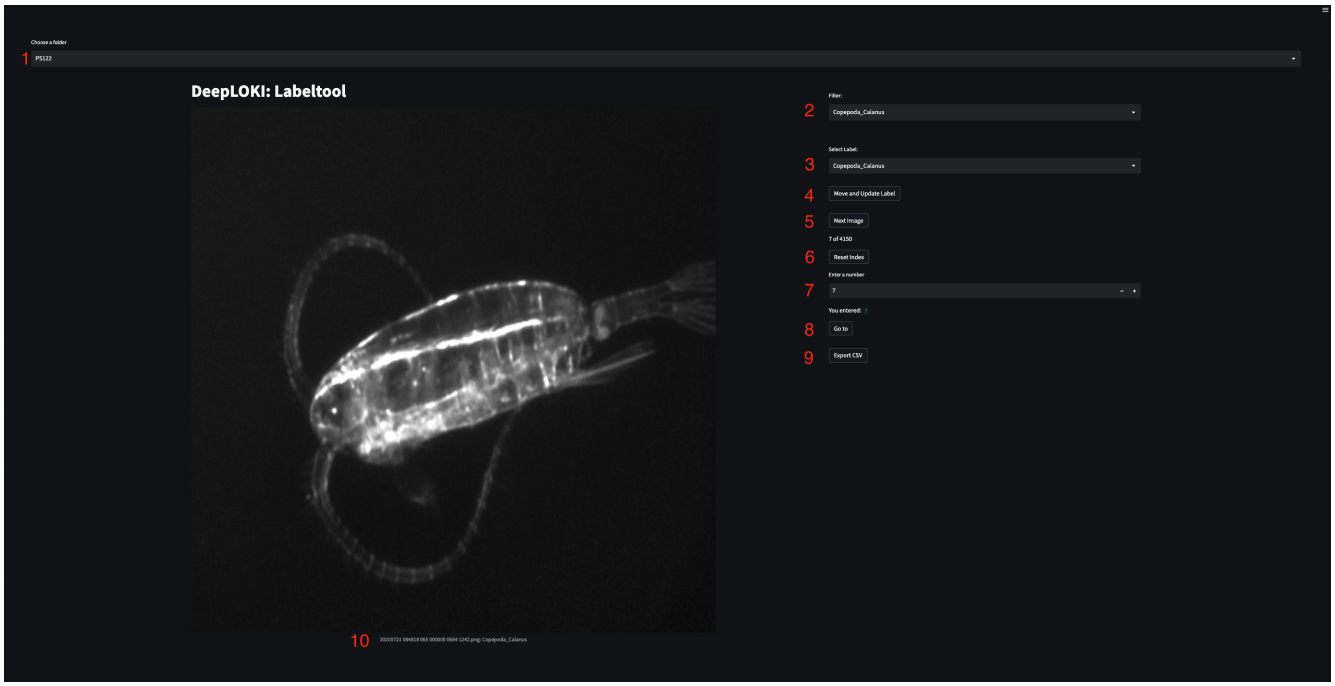

Figure S1. Graphical User Interface- Labeltool: 1. A drop down menu to select the folder with the image, that should be labelled. **2.** A second drop down menu to filter the image by category. **3.** A drop down menu that show actual label, a new label can be selected. **4.** A Button to confirm the label selected in **3**. **5.**Button to show the next image. **6.** A Button to reset the index and show the first image. **7.** A input filed to select a specific image. **8.** A Button to jump the selected image **7**. **9.** A Button to generate a CSV file with the image names, paths and labels. **10.** A text, display the image filename and current label.

2 DINO - STUDENT AND TEACHER APPROACH

In the next figure the Student and teacher approach is shown.

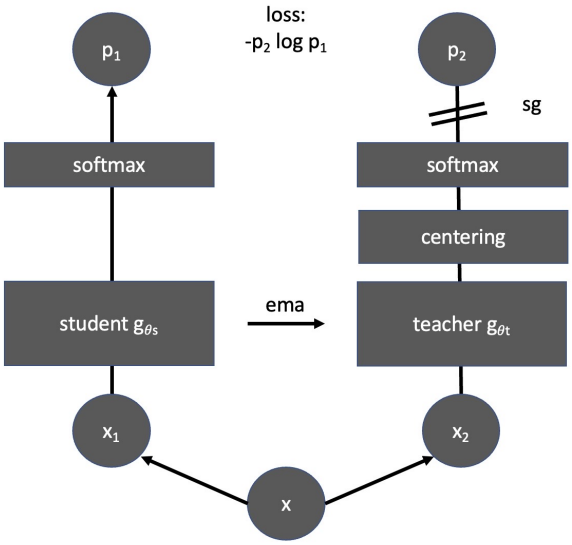

Figure S2. Student and teacher approach copied from dino (Caron et al., 2021). A Student ResNet18 learns to classify global features in an image from local patches supervised by the cross entropy loss from a momentum Teacher ResNet18 embeddings while doing centering and sharpening to prevent mode collapse.

3 DEEP TRANSFER LEARNING

Here we give a Definition of Deep Transfer Learning:

Definition: Deep Transfer Learning A task defined by $\{\mathcal{D}_s, \mathcal{T}_s, \mathcal{D}_t, \mathcal{T}_t, f_T(\cdot)\}$. is called a Deep Transfer Learning task, where $f_T(\cdot)$ is a non-linear function that reflects a Deep Neural Network (Tan et al., 2018; Kronberg, 2022).

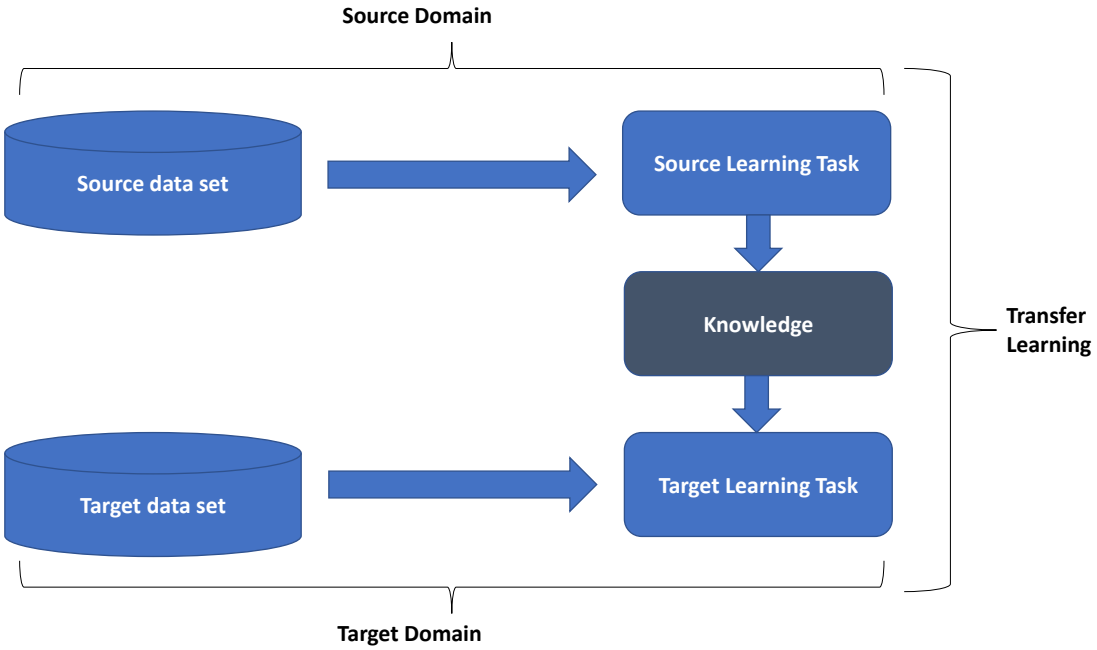

Figure S3. Learning process of Transfer Learning, adapted from (Tan et al., 2018; Kronberg, 2022)

The focus of this study is on network-based deep transfer learning, which involves the reuse of weights from a pre-trained deep neural network in the source domain, and their transfer to the deep neural network in the target domain. Specifically, the weights from the pre-trained network are copied to the deep neural network in the target domain, which shares the same architecture as the source domain network. By utilizing this approach, we aim to improve the efficiency and effectiveness of deep learning in the target domain (Tan et al., 2018; Kronberg, 2022).

4 LOSS FUNCTION AND OPTIMIZER

For both DeepLOKI classifiers we used the Cross-entropy loss as loss function and the Adam Optimizer.

Definition: Cross-entropy loss
The Cross-entropy loss (CEL) is given by

$$L(M(x), y) = -\sum_{i}^{K} y_i \log(M(x_i)),$$

where x is the input, K the number of classes, M the Deep Neural Network and y contains a onehot encoding of the target label (Kronberg, 2022).

Definition: Adam The parameter update by the Adam optimizer is defined by

$$\begin{aligned}
w_t &= w_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \\
\hat{v} &= \frac{v_t}{(1 - \beta_2^t)}, \\
\hat{m} &= \frac{m_t}{(1 - \beta_1^t)}, \\
v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\
m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\
g_t &= \frac{\partial L_t(M, w_{t-1})}{\partial w_t},
\end{aligned}$$

where g is the gradient w.r.t the loss function L_t, m the biased first moment estimate, v the biased second raw moment estimate, $\hat{m}$ the bias-corrected first moment estimate, $\hat{v}$ the bias-corrected second raw moment estimate and α the learning rate.(Kingma and Ba, 2014; Kronberg, 2022)

5 SOME STATISTICS OVER THE DIFFERENT CRUISES

We visualized the zooplankton species distribution of the five cruises Figure S4. This counts are later important for the analysis the bio-diversity. The largest classes of PS99.2 are Copepoda Calanus, Detritus, Copepoda Group Oncaea, Copepoda Microcalanus and Copepoda Metridia longa. The largest classes of PS106.2 are Copepoda Calanus, Copepoda Metridia longa,Copepoda Group Oncaea, Bubble and Detritus. Here, we see that Artefacts/Noise like Bubble and Detritus has a high contribution to the images over all. The largest classes of PS107 are Copepoda Calanus, Copepoda Group Oncaea, Copepoda Microcalanus , Copepoda Metridia longa and Copepoda Calanoida. The largest classes of PS114 are Copepoda Group Oncaea, Detritus, Copepoda Microcalanus, Copepoda Calanus and Copepoda Calanoida. The largest classes of PS122 are Copepoda Calanus, Copepoda Group Oncaea, Detritus, Copepoda Metridia longa and Copepoda Microcalanus. In sum, all cruises was dominated by Copepoda Calanus, Copepoda Group Oncaea and Copepoda Microcalanus. However, a great portion multiplies, detritus and other artefacts or non interesting classes Figure S4.

Calculation the correlation between the normalized count vectors, where non detected species was represented by zeros in the vector.

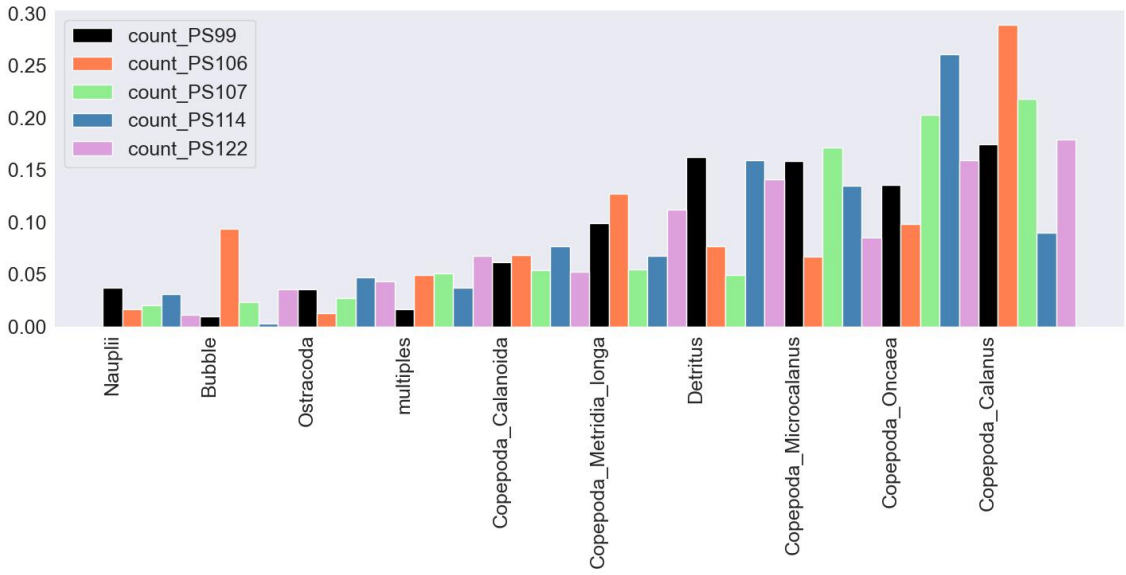

Figure S4. Comparison of Distribution of the Zooplankton over the five cruises over the top 10 classes. The top ten classes are Copepoda_Calanus, Copepoda_Group_Onceaea, Detritus, Copepoda_Microcalanus, Copepoda_Metridia_longa, Copepoda_Calanoida, multiples, Ostracoda, Bubble and Nauplii. The colored bars indicate the number of images taken of the class by LOKI System on the cruise. The cruises are PS99 (black), PS106 (orange), PS107 (green), PS114 (blue) and PS122 (purple).

6 LEARNED LATENT-SPACE EMBEDDING ANALYSIS

In our study we visualized the latent-space Embedding, we observed a pattern. This pattern can be tracked back to difference in parameters. Here we tested, for difference using the t-test and calculate statistics like mean and variance.

Table S1. An Table give a overview about the images parameter differences for the Category Copepoda Calanus using a t-test pairwise, when dividing the latent space umap projection at x = 5 for the test dataset (PS99.2).

Parameter	Difference	P-Value
x_tsne	Yes	0.0
y_tsne	Yes	0.0
x_umap	Yes	0.0
y_umap	Yes	1.037783777273046-110
Unnamed: 0_y	No	0.31876942160123206
new_index	Yes	0.0
img_rank	No	
object_lat	Yes	0.0
object_lon	Yes	1.2926646480204804e-22
object_date	Yes	1.0667464894866881e-134
object_time	Yes	0.0
object_link	No	
object_depth_min	Yes	0.0
object_depth_max	No	
object_annotation_date	No	
object_annotation_time	No	
object_annotation_category_id	Yes	0.0001144381819238334 5
object_index	Yes	3.7886382081375443e-39
object_haul	Yes	0.0
object_bottom_depth	No	
object_pressure	Yes	1.228022300630215e-118
object_temperature	No	
object_salinity	Yes	0.0001221490302737464 5
object_conductivity	No	0.35632112562601237
object_oxygen_concentration	Yes	7.675120848042546e-46
object_temperature_oxsens	No	
object_oxygen_saturation	Yes	1.5487641417002434e-54
object_dr._haardt_fluorescence_channel_a	Yes	0.005757310659651 58
object_area_px	Yes	8.036970509587048e-26
object_form	Yes	0.001618268887689836 3
object_area	Yes	0.0
object_length	No	
object_width	No	0.11707595622189075
object_convexity	Yes	1.8907706699190674e-09
object_structure	Yes	6.914991753442495e-23
object_graymean	Yes	9.177412459225949e-05
object_kurtosis	No	0.6260313309925571
object_skewness	Yes	0.000837827954778929 2
object_hu_moment_1	Yes	3.5443478745462 68e-05
object_hu_moment_2	Yes	3.5443478745462 68e-05
object_hu_moment_3	Yes	3.5443478745462 68e-05
object_hu_moment_4	Yes	3.5443478745462 68e-05
object_hu_moment_5	Yes	3.5443478745462 68e-05
object_hu_moment_6	Yes	3.5443478745462 68e-05
object_hu_moment_7	Yes	3.5443478745462 68e-05
object_fourier_descriptor_01	Yes	6.026281881773818e-17
object_fourier_descriptor_02	Yes	1.2354747581547124e-18
object_fourier_descriptor_03	Yes	1.0365639235645 21e-14
object_fourier_descriptor_04	Yes	1.3955980132954984e-05
object_fourier_descriptor_05	Yes	2.4459408697224 76e-10
object_fourier_descriptor_06	Yes	1.9708507011187 78e-07
object_fourier_descriptor_07	Yes	2.1958688111526 86e-06
object_fourier_descriptor_08	Yes	9.819275577572176-07
object_fourier_descriptor_09	Yes	1.7748005916643043e-07
object_fourier_descriptor_10	Yes	5.6776537060943075e-08
object_posx	No	
object_posy	No	
object_milliseconds	No	
object_timestamp	No	
object_lenght	No	
object_zoomie_group_id	No	
object_mc_width	Yes	1.536257233199282 2e-27
object_mc_height	Yes	1.3552726312009295e-14
object_mc_bx	No	
object_mc_by	No	
object_mc_circ.	Yes	2.030379882012763 7e-08
object_mc_area_exc	Yes	0.025846667352946 82
object_mc_area	No	0.11752926651893286
object_mc_%area	Yes	2.919705923494464 6e-45
object_mc_major	Yes	0.018590802852726185
object_mc_minor	Yes	3.7186614992066 85e-51
object_mc_y	Yes	5.1249286969859 78e-13
object_mc_x	Yes	5.5435676354435 65e-16
object_mc_convex_area	Yes	1.1523472014335107e-28
object_mc_min	Yes	0.0
object_mc_max	Yes	9.1507025693709 97e-29
object_mc_mean	Yes	7.9743683485521 53e-220
object_mc_intden	Yes	2.8856581190312 34e-37
object_mc_perim.	Yes	1.3631156646414717e-15
object_mc_elongation	Yes	3.7851962188905397e-56
object_mc_range	Yes	3.2052951108945 86e-35
object_mc_perimareaexc	No	0.5782255154213629
object_mc_perimmajor	Yes	1.4187782441474808e-18
object_mc_circex	Yes	3.1514902592764 95e-07
object_mc_angle	No	0.10639815452602579
object_mc_bounding_box_area	Yes	2.1897253795278 86e-33
object_mc_eccentricity	Yes	6.9702837965818 21e-32
object_mc_equivalent_diameter	Yes	3.3481263945629 44e-07
object_mc_euler_number	Yes	8.4186179020170 74e-71
object_mc_extent	Yes	5.5377525830490 56e-97
object_mc_local_centroid_col	Yes	5.5435676354438 65e-16
object_mc_local_centroid_row	Yes	5.1249286969859 78e-13
object_mc_solidity	Yes	1.4184259315775915e-36
sample_dataportal_descriptor	No	
acq_instrument	No	

Table S2. An Table give a overview about the images parameter differences for the Category Copepoda Calanus using statistics like mean, median, variance, standard derivation, maximum, minimum, skewness and kurtosis pairwise, when dividing the latent space umap projection at x = 5 for the test dataset (PS99.2).The suffix l and r correspond to the left or right site of the x =5 separation line.

Parameter	mean_r	mean_l	var_r	var_l	min_r	min_l	max_r	max_l	std_r	std_l	median_r	median_l	skew_r	skew_l	kurtosis_r	kurtosis_l
object_annotation_date	20190295.72	20190277.29	5576213.18	442338.27	20171113.0	20190124.0	20200309.0	20200709.0	2361.4	665.09	20190206.0	20190301.0	-2.32	14.85	39.66	222.75
object_date	20160702.73	20160678.19	10.48	1383.59	20160630.0	20160630.0	20160707.0	20160707.0	3.24	37.2	20160703.0	20160707.0	-22.06	-0.52	493.83	-1.73
object_mc_intden	1228034.03	1731426.82	1080469987962.77	1464459699115.74	32332.0	93033.89	8532570.52	15966020.11	1039456.58	1210148.63	959692.51	1537891.01	3.31	4.59	14.02	33.32
object_fourier_descriptor_01	465871.41	377557.1	127831876540.18	65854798285.66	30362.74	38681.93	3983267.0	4898661.66	357535.84	256621.9	378954.28	337377.46	3.4	6.42	17.36	79.87
object_mc_bounding_box_area	192000.62	162856.62	6982570602.5	3148577941.47	44280.0	48555.0	855900.0	677730.0	83561.78	56112.19	175227.0	154638.0	2.73	2.47	11.54	13.54
object_annotation_time	124565.09	132721.03	453216413.87	203615265.93	73402.0	92617.0	213237.0	181225.0	21288.88	14269.38	124525.0	132330.0	0.98	-0.31	3.22	0.11
object_mc_convex_area	97341.27	120671.7	4702472101.48	2845408469.8	8634.0	18712.0	591557.0	609678.0	68574.57	53342.37	81256.0	116931.0	2.73	2.31	11.05	13.37
object_fourier_descriptor_02	157803.14	119727.95	22143867310.63	10401400159.72	6096.57	4447.67	1839578.05	1965171.86	148808.16	101987.25	116455.29	101316.48	3.77	7.33	21.8	99.19
Unnamed: 0_y	110372.92	112592.66	4158856538.02	4252139904.19	176.0	52.0	224333.0	224522.0	64489.2	65208.43	109636.5	115014.0	0.01	0.0	-1.18	-1.24
object_structure	134624.93	101067.2	14117014221.21	6131019865.7	5854.07	3428.93	1440162.94	988558.69	118815.04	78300.83	102843.2	85382.96	3.46	4.02	19.37	30.64
object_annotation_category_id	90169.05	90739.22	21198145.37	16324478.59	80126.0	80126.0	92309.0	92309.0	4604.14	4040.36	92275.0	92275.0	-1.73	-2.25	0.98	3.06
object_time	207706.64	82413.7	3601876281.2	5290253141.71	23056.0	22145.0	233641.0	233619.0	60015.63	72734.13	232921.5	25045.0	-1.97	0.5	1.92	-1.64
object_fourier_descriptor_03	37628.8	27659.0	1968158389.55	933961979.33	200.12	130.61	467322.23	339925.45	44363.93	30560.79	23952.35	19307.11	3.3	4.03	15.22	26.14
object_mc_area	24770.16	25684.57	378483303.18	217344332.33	1252.0	2088.0	141810.0	225522.0	19454.65	14742.6	20101.0	23665.0	3.02	4.67	10.45	37.45
object_mc_area_exc	23829.62	25092.57	355327401.66	205041906.56	1252.0	2087.0	141264.0	219623.0	18850.13	14319.28	19169.0	23156.0	3.03	4.63	10.51	37.28
object_area_px	25970.42	19802.0	421564953.48	179022781.25	1442.0	1184.0	170654.0	270022.0	20532.05	13379.94	21346.0	18419.0	3.2	6.51	12.28	82.55
new_index	7323.09	18380.63	8588226075	2722421668	57.0	8288.0	30156.0	25679.0	9267.27	5217.68	3714.0	17554.0	1.9	-0.08	1.78	-1.34
object_fourier_descriptor_04	20875.29	16196.05	1024715567.85	948303503.26	133.98	215.78	587978.1	725434.92	32011.18	30794.54	11140.14	10119.19	6.92	14.13	84.32	285.03
Unnamed: 0_x	10169.43	10374.87	35294138.9	36073840.42	12.0	2.0	20654.0	20670.0	5940.89	6006.15	10091.0	10579.0	0.01	0.0	-1.19	-1.24
object_fourier_descriptor_05	12656.48	9381.7	282631271.44	179718413.96	144.36	23.87	223690.91	230909.81	16811.64	13405.91	7851.76	6101.75	4.7	8.48	35.27	113.02
object_fourier_descriptor_06	8805.39	6781.57	136340433.05	121296485.74	45.31	23.21	125747.8	324343.72	11676.49	11013.47	5087.67	4459.84	3.97	15.8	22.51	398.04
object_zoomie_group_id	4574.44	6318.5	4676816.12	1093852.03	73.0	5248.0	9259.0	8079.0	2162.59	1045.87	4205.0	5883.5	0.15	0.88	-1.01	-1.04
object_fourier_descriptor_07	6579.57	5209.51	75492286.21	67147896.25	90.28	14.45	152917.89	212694.7	8688.63	8194.38	3946.92	3467.91	5.65	12.9	63.14	268.25
object_index	5850.75	4592.32	5454340.22	9457742.71	127.0	10.0	11083.0	12708.0	2335.45	3075.34	6193.5	3261.0	-0.33	0.54	-0.5	-0.97
object_fourier_descriptor_08	5393.72	4179.26	53500823.8	50955762.96	29.83	4.66	123606.52	184457.66	7314.43	7138.33	3199.45	2742.13	5.4	14.49	55.65	311.87
object_fourier_descriptor_09	4236.52	3236.58	37165190.54	26014496.64	30.2	19.2	88147.49	105676.31	6096.33	5100.44	2474.1	2124.8	5.93	11.2	57.93	187.79
object_mc_perim.	2563.19	3000.97	1551534.11	3324599.38	265.0	537.24	12115.64	25877.85	1245.61	1823.35	2401.42	2658.31	2.1	4.22	8.45	29.28
object_fourier_descriptor_10	3506.66	2608.44	34925711.62	13536119.93	19.66	14.37	153611.03	91931.62	5909.8	3679.15	2044.82	1791.11	12.4	11.21	276.73	221.52
object_pressure	1026.08	2145.26	409256.01	2996705.48	-13.5	23.94	8962.61	9782.45	639.73	1731.1	1027.64	1533.36	5.66	1.28	47.79	1.22
object_posx	1050.99	831.83	370419.98	319730.2	0.0	0.0	2067.0	2052.0	608.62	565.45	1110.0	793.5	-0.22	0.44	-1.11	-0.46
object_posy	644.18	806.42	244181.76	165785.56	0.0	0.0	1587.0	1557.0	494.15	407.17	624.0	829.5	0.18	-0.31	-1.34	-0.55
object_mc_width	459.49	418.13	14829.14	9831.11	171.0	182.0	1080.0	1152.0	121.77	99.15	444.0	405.0	1.22	1.22	2.91	3.64
object_mc_height	417.54	389.83	14406.04	8041.2	165.0	192.0	1065.0	885.0	120.03	89.67	396.0	378.0	1.49	1.35	3.54	3.97
object_oxygen_concentration	448.7	371.99	52328.23	895.26	354.51	350.7	1279.04	518.38	228.75	29.92	359.53	360.59	2.68	2.81	5.75	7.44
object_mc_euler_number	-13.76	359.65	2321.72	647302.05	-296.0	-1841.0	268.0	8662.0	48.18	804.55	-7.0	26.0	-1.29	3.12	4.53	14.43
object_mc_major	334.02	342.43	14488.3	7844.03	80.99	133.53	1058.63	870.79	120.37	88.57	310.77	333.62	2.26	1.94	6.86	7.05
object_mc_max	249.92	255.0	387.1	0.0	69.0	255.0	255.0	255.0	19.67	0.0	255.0	255.0	-4.7	0.0	24.46	0.0
object_mc_range	236.04	229.98	383.04	48.39	56.0	207.0	243.0	241.0	19.57	6.96	242.0	232.0	-4.64	-0.82	24.07	-0.19
object_depth_min	880.65	229.51	112191.36	41087.1	-1.34	2.37	1019.92	1019.23	334.95	202.7	1019.45	158.58	-2.06	1.71	2.35	3.28
object_mc_x	230.28	211.16	5682.87	3872.02	23.23	30.51	628.38	692.18	75.38	62.23	223.92	203.01	0.5	0.91	2.27	3.31
object_mc_local_centroid_col	230.28	211.16	5682.87	3872.02	23.23	30.51	628.38	692.18	75.38	62.23	223.92	203.01	0.5	0.91	2.27	3.31
object_mc_y	207.37	193.56	4027.8	2313.69	32.26	42.99	535.24	462.35	63.46	48.1	199.61	187.17	1.24	0.87	3.8	3.0
object_mc_local_centroid_row	207.37	193.56	4027.8	2313.69	32.26	42.99	535.24	462.35	63.46	48.1	199.61	187.17	1.24	0.87	3.8	3.0
object_mc_equivalent_diameter	165.64	173.94	2904.58	1693.55	39.93	51.55	424.1	528.8	53.89	41.15	156.23	171.71	1.93	1.92	4.79	10.05
object_mc_minor	135.06	165.0	3053.03	3419.65	39.92	56.2	532.89	668.16	55.25	58.48	121.97	155.63	1.72	1.78	4.64	7.01
object_mc_angle	94.29	91.01	3586.72	3408.23	0.03	0.02	180.0	179.98	59.89	58.38	96.78	92.17	-0.13	-0.04	-1.41	-1.34
object_oxygen_saturation	104.0	86.83	2185.24	29.82	77.33	76.37	273.16	117.6	46.75	5.46	87.29	86.25	2.68	2.98	5.66	10.76
object_lat	79.76	79.03	0.11	0.04	78.82	78.82	79.9	79.9	0.33	0.19	79.9	79.13	-1.96	1.37	1.84	5.36
object_convexity	87.2	75.13	3707.59	3169.44	0.0	0.0	553.66	597.19	60.89	56.3	69.54	59.48	1.64	2.11	4.79	8.82
object_mc_mean	49.19	66.61	157.27	274.19	22.53	31.61	98.7	148.29	12.54	16.56	47.14	66.47	0.83	0.21	0.68	-0.16
y_tsne	9.87	42.8	77.97	203.26	-36.3	-7.02	46.76	69.02	8.83	14.26	10.8	40.9	-0.64	-0.23	0.75	0.12
object_salinity	33.63	33.51	1.27	0.47	29.6	29.6	34.52	34.59	1.13	0.68	34.16	33.76	-2.08	-1.56	3.64	3.27
object_conductivity	30.84	30.79	4.14	2.48	25.9	26.46	32.44	32.19	2.03	1.57	31.74	31.51	-1.53	-1.58	0.8	1.35
object_graymean	27.16	26.03	82.7	61.73	10.0	11.0	77.0	91.0	9.09	7.86	26.0	25.0	1.08	1.07	1.92	3.39
object_mc_min	13.88	25.02	3.54	48.39	12.0	14.0	46.0	48.0	1.88	6.96	13.0	23.0	6.9	0.82	94.7	-0.19
object_kurtosis	14.07	13.94	61.08	50.95	0.96	-0.37	58.28	45.45	7.82	7.14	12.56	12.52	1.08	1.15	1.62	1.49
x_tsne	54.74	10.95	152.81	152.28	-45.26	-48.72	72.89	30.83	12.36	12.34	57.28	10.61	-2.01	-1.3	7.63	3.53
y_umap	11.55	10.0	2.07	5.24	0.89	-1.7	13.31	12.99	1.44	2.29	11.84	10.59	-4.28	-1.93	26.8	5.16
object_form	9.04	9.59	25.4	27.66	1.44	1.44	41.19	33.13	5.04	5.26	7.61	8.18	1.62	1.33	3.71	1.88
x_umap	-0.11	9.59	1.2	2.34	-3.06	7.13	4.09	13.44	1.09	1.53	-0.05	9.72	0.14	0.16	-0.28	-1.29
object_mc_perimmajor	7.65	8.58	5.36	12.74	2.9	2.73	22.89	48.58	2.31	3.57	7.54	8.22	0.49	3.63	0.97	26.43
object_length	0.01	0.06	0.03	0.2	0.0	0.0	2.56	5.04	0.17	0.44	0.0	0.0	15.26	7.66	233.0	63.32
object_haul	1.31	5.2	0.62	8.67	1.0	1.0	9.0	9.0	0.79	2.94	1.0	3.0	3.21	0.49	17.73	-1.7
object_area	430.92	4.69	119996.76	311.59	0.56	0.15	3840.0	396.0	346.41	17.65	440.0	2.44	3.26	11.75	21.63	185.28
object_skewness	3.56	3.66	0.86	0.72	1.35	0.69	7.35	6.54	0.93	0.85	3.48	3.56	0.39	0.46	0.01	0.24
object_width	0.01	0.02	0.09	0.02	0.0	0.0	11.7	1.71	0.3	0.15	0.0	0.0	38.9	7.78	1522.07	64.02
object_temperature_oxsens	-0.01	2.36	1.8	2.22	-0.96	-1.68	3.66	3.79	1.34	1.49	-0.8	3.01	1.31	-1.45	0.22	1.18
object_mc_elongation	2.66	2.23	0.79	0.47	1.06	1.01	10.2	5.67	0.89	0.69	2.57	2.12	1.35	0.88	6.17	1.05
object_temperature	3.07	1.89	1.28	3.9	-0.85	-1.68	4.08	3.87	1.13	1.98	3.45	3.08	-1.85	-0.9	2.8	-0.96
object_mc_extent	0.12	0.15	0.0	0.0	0.02	0.03	0.38	0.37	0.05	0.05	0.11	0.14	1.69	1.03	3.94	1.73
object_dr_haardt_fluorescence_channel_a	0.08	0.11	0.03	0.07	0.02	0.02	2.0	2.97	0.16	0.26	0.03	0.03	5.15	5.3	36.3	38.81
object_mc_perimareaexc	0.13	0.13	0.0	0.0	0.02	0.03	0.33	0.61	0.05	0.05	0.12	0.12	0.49	1.98	0.33	9.85
count_y	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
count_x	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
object_hu_moment_6	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_3	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_1	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_7	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_2	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_5	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_hu_moment_4	0.34	0.33	0.01	0.01	0.16	0.17	1.41	1.37	0.1	0.09	0.33	0.32	3.32	3.21	21.63	21.63
object_mc_eccentricity	0.89	0.85	0.01	0.01	0.32	0.14	1.0	0.98	0.09	0.11	0.92	0.88	-2.32	-2.04	6.66	5.92
object_mc_circex	0.06	0.05	0.0	0.0	0.01	0.0	0.5	0.5	0.04	0.04	0.04	0.04	3.02	2.87	15.26	16.03
object_mc_circ.	0.06	0.05	0.0	0.0	0.01	0.0	0.5	0.5	0.04	0.04	0.05	0.04	3.01	2.83	15.19	15.72
object_mc_solidity	0.26	0.22	0.01	0.01	0.04	0.05	0.83	0.81	0.12	0.09	0.24	0.2	1.48	2.18	3.01	7.67
object_lon	1.72	0.77	8.71	7.29	-5.39	-2.73	2.95	2.95	2.95	2.7	2.95	2.85	-1.99	-0.52	1.98	-1.73
object_mc_%area	0.04	0.02	0.0	0.0	0.0	0.0	0.26	0.34	0.04	0.03	0.03	0.01	1.75	3.55	3.47	18.98
object_lenght	0.02	0.0	0.78	0.0	0.0	0.0	31.95	0.0	0.88	0.0	0.0	0.0	36.21	0.0	1311.0	0.0
img_rank	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
object_mc_bx	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
object_mc_by	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

7 TRAIN, VALIDATION AND TEST DATASET

Here we compare the similarity of distribution vectors of the different cruises and the selection of the train, validation and test dataset. From the count perspective, train/val and test are very correlated.

Table S3. An Table give a overview about the images used.

category	**train_val**	**test**
Amphipoda	1169	64.0
Antenna	322	24.0
Bubble	7742	209.0
Chaetognata_head	1451	229.0
Chaetognata_middle	1063	184.0
Chaetognata_tail	559	47.0
Cnidaria	496	91.0
Copepoda_Calanoida	11294	1279.0
Copepoda_Calanus	43620	3614.0
Copepoda_Gaetanus	347	74.0
Copepoda_Heterorhabdus	532	148.0
Copepoda_Metridia_longa	15141	2056.0
Copepoda_Microcalanus	26649	3283.0
Copepoda_Oithona	3285	411.0
Copepoda_Oncaea	34463	2805.0
Copepoda_Paraeuchaeta	626	118.0
Copepoda_Pseudocalanus	2248	100.0
Copepoda_Scaphocalanus	1065	78.0
Copepoda_Scolecithricella	134	22.0
Copepoda_Spinocalanus	385	10.0
Copepoda_dead	248	
Crustacea	150	28.0
Detritus	13730	3363.0
Eggs	3015	251.0
Euphausiacea	185	4.0
Feces	881	79.0
Foraminifera	121	5.0
Nauplii	3730	780.0
Ostracoda	5285	748.0
Polychaeta	302	33.0
Rhizaria	3982	199.0
Trochophora	144	
multiples	10115	347.0
sum	194479	20683

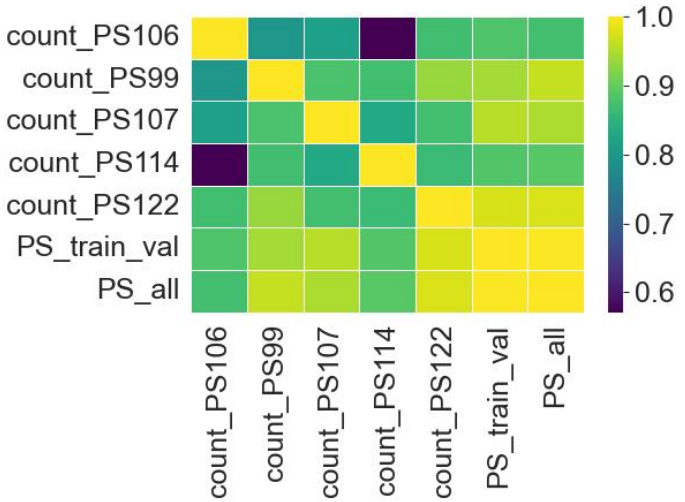

Figure S5. Comparission of Distribution of the Zooplankton over the five cruises heatmap using pairwise Correlation of the 35 classes vector with label numbers

8 CLASSIFICATION REPORTS

We provided Classification Reports for all three approaches and explain a confusion matrix.

Confusion Matrix A common graphic to show the True Positive, True Negative, False Positive and False Negative is the confusion matrix see Figure S6.

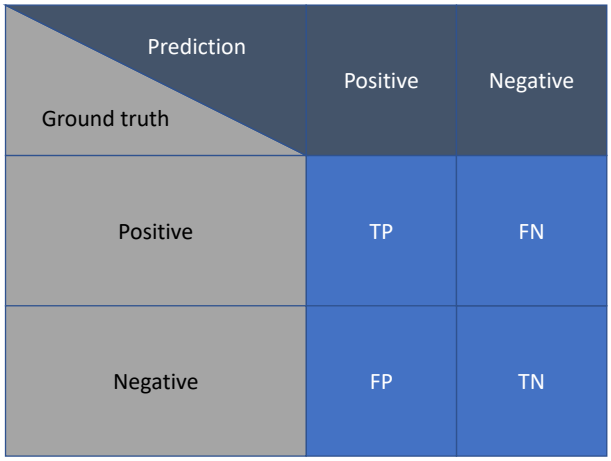

Figure S6. Confusion Matrix, True Positives, True Negatives, False Positives and False Negatives for the binary case. Adapted from (Kronberg, 2022)

Table S4. Comparison of the Classification Report of the EcoTaxa and the Dino ResNet18 - Classifier. For each metric, precision, recall and f1-score, the EcoTaxa score, the Dino ResNet18 score and the difference (delta: Dino ResNet18 - EcoTaxa) is shown. The classes are sorted by their support (number of sample in the reference dataset PS99.2)

category	precision	precision (our)	delta precision	recall	recall (our)	delta recall	f1-score	f1-score (our)	delta f1-score	support
Amphipoda	0.315789474	0.1875	0.235294118	0.577777778	0.8125	0.675324675	0.65625	0.65625	0.65625	64
Antenna	0.044217687	0.541666667	0.081761006	0.318181818	0.291666667	0.304347826	0.5	0.25	0.333333333	24
Bubble	0.613718412	0.813397129	0.699588477	0.943127962	0.95215311	0.947619048	0.95959596	0.909090909	0.933660934	209
Chaetognata_head	0.370860927	0.244541485	0.294736842	0.758893281	0.838427948	0.796680498	0.722807018	0.899563319	0.80155642	229
Chaetognata_middle	0.641350211	0.826086957	0.722090261	0.894409938	0.782608696	0.834782609	0.884146341	0.788043478	0.833333333	184
Chaetognata_tail	0.130952381	0.234042553	0.167938931	0.322580645	0.212765957	0.256410256	0.425	0.361702128	0.390804598	47
Cnidaria	0.401639344	0.538461538	0.460093897	0.786885246	0.527472527	0.631578947	0.76344086	0.78021978	0.77173913	91
Copepoda_Calanoida	0.175438596	0.007818608	0.01497006	0.49670812	0.530883503	0.513227513	0.486603284	0.440187647	0.462233169	1279
Copepoda_Calanus	0.736498258	0.467902601	0.572250423	0.918108108	0.939955728	0.928903473	0.944348824	0.910902048	0.927323944	3614
Copepoda_Gaetanus	0.086956522	0.027027027	0.041237113	0.482758621	0.189189189	0.27184466	0.529411765	0.121621622	0.197802198	74
Copepoda_Heterorhabdus	0	0	0	0.666666667	0.040540541	0.076433121	0.166666667	0.006756757	0.012987013	148
Copepoda_Metridia_longa	0.390955924	0.664396887	0.492252252	0.888079125	0.829766537	0.857933115	0.835413777	0.878891051	0.85660109	2056
Copepoda_Microcalanus	0.677701437	0.330490405	0.444307944	0.908681672	0.860798051	0.884091976	0.853213665	0.897654584	0.87487012	3283
Copepoda_Oithona	0.118426678	0.673965937	0.201454545	0.755555556	0.661800487	0.705577173	0.760115607	0.639902676	0.694848085	411
Copepoda_Oncaea	0.472184532	0.248128342	0.325309652	0.894492132	0.891622103	0.893054812	0.899472892	0.851693405	0.874931331	2805
Copepoda_Paraeuchaeta	0.291666667	0.118644068	0.168674699	0.404255319	0.161016949	0.23030303	0.447368421	0.288135593	0.350515464	118
Copepoda_Pseudocalanus	0.221374046	0.29	0.251082251	0.395683453	0.55	0.460251046	0.47761194	0.64	0.547008547	100
Copepoda_Scaphocalanus	0.050505051	0.448717949	0.09079118	0.354545455	0.5	0.414893617	0.264573991	0.756410256	0.392026578	78
Copepoda_Scolecithricella	0	0	0	0	0	0	0	0	0	22
Copepoda_Spinocalanus	0	0	0	0	0	0	0	0	0	10
Copepoda_dead	0	0	0	0	0	0	0	0	0	0
Crustacea	0	0	0	0	0	0	0	0	0	28
Detritus	0.644868302	0.844484092	0.731299086	0.875566214	0.977103776	0.923552558	0.83997968	0.983348201	0.906027397	3363
Eggs	0.811111111	0.290836653	0.428152493	0.894366197	0.505976096	0.646310433	0.898876404	0.637450199	0.745920746	251
Euphausiacea	0	0	0	0	0	0	0	0	0	4
Feces	0.304878049	0.316455696	0.310559006	0.5	0.088607595	0.150537634	0.6	0.113924051	0.191489362	79
Foraminifera	0	0	0	0	0	0	0	0	0	5
Nauplii	0.136336692	0.294871795	0.186461289	0.748837209	0.825641026	0.785365854	0.831775701	0.684615385	0.751054852	780
Ostracoda	0.662207358	0.264705882	0.378223496	0.948207171	0.954545455	0.951365756	0.951417004	0.942513369	0.946944258	748
Polychaeta	0.142857143	0.03030303	0.05	0.785714286	0.333333333	0.468085106	0.8	0.242424242	0.372093023	33
Rhizaria	0.185314685	0.532663317	0.274967575	0.885057471	0.773869347	0.825737265	0.77092511	0.879396985	0.821596244	199
Trochophora	0	0	0	0	0	0	0	0	0	0
multiples	0.666666667	0.011527378	0.02266289	0.451553931	0.711815562	0.552572707	0.556195965	0.556195965	0.556195965	347
accuracy	0.442150558	0.442150558	0.442150558	0.839143258	0.839143258	0.839143258	0.831407436	0.831407436	0.831407436	20683
macro avg	0.29045238	0.289019875	0.238942484	0.54111192	0.477092733	0.484448021	0.540157905	0.488390717	0.491004459	20683
weighted avg	0.532853689	0.442150558	0.440251837	0.839278132	0.839143258	0.833619198	0.825690234	0.831407436	0.823949051	20683

Table S5. Classification Report for EcoTaxa.

category	precision	recall	f1-score	support
Amphipoda	0.315789474	0.1875	0.235294118	64
Antenna	0.044217687	0.541666667	0.081761006	24
Bubble	0.613718412	0.813397129	0.699588477	209
Chaetognata_head	0.370860927	0.244541485	0.294736842	229
Chaetognata_middle	0.641350211	0.826086957	0.722090261	184
Chaetognata_tail	0.130952381	0.234042553	0.167938931	47
Cnidaria	0.401639344	0.538461538	0.460093897	91
Copepoda_Calanoida	0.175438596	0.007818608	0.01497006	1279
Copepoda_Calanus	0.736498258	0.467902601	0.572250423	3614
Copepoda_Gaetanus	0.086956522	0.027027027	0.041237113	74
Copepoda_Heterorhabdus	0	0	0	148
Copepoda_Metridia_longa	0.390955924	0.664396887	0.492252252	2056
Copepoda_Microcalanus	0.677701437	0.330490405	0.444307944	3283
Copepoda_Oithona	0.118426678	0.673965937	0.201454545	411
Copepoda_Oncaea	0.472184532	0.248128342	0.325309652	2805
Copepoda_Paraeuchaeta	0.291666667	0.118644068	0.168674699	118
Copepoda_Pseudocalanus	0.221374046	0.29	0.251082251	100
Copepoda_Scaphocalanus	0.050505051	0.448717949	0.09079118	78
Copepoda_Scolecithricella	0	0	0	22
Copepoda_Spinocalanus	0	0	0	10
Copepoda_dead	0	0	0	0
Crustacea	0	0	0	28
Detritus	0.644868302	0.844484092	0.731299086	3363
Eggs	0.811111111	0.290836653	0.428152493	251
Euphausiacea	0	0	0	4
Feces	0.304878049	0.316455696	0.310559006	79
Foraminifera	0	0	0	5
Nauplii	0.136336692	0.294871795	0.186461289	780
Ostracoda	0.662207358	0.264705882	0.378223496	748
Polychaeta	0.142857143	0.03030303	0.05	33
Rhizaria	0.185314685	0.532663317	0.274967575	199
Trochophora	0	0	0	0
multiples	0.666666667	0.011527378	0.02266289	347
accuracy	0.442150558	0.442150558	0.442150558	20683
macro avg	0.29045238	0.289019875	0.238942484	20683
weighted avg	0.532853689	0.442150558	0.440251837	20683

Table S6. Classification Report for Resnet pretrained on ImageNet.

category	precision	recall	f1-score	support
Amphipoda	0.65625	0.65625	0.65625	64.0
Antenna	0.5	0.25	0.3333333333333333	24.0
Bubble	0.9595959595959596	0.9090909090909091	0.9336609336609336	209.0
Chaetognata_head	0.7228070175438597	0.8995633187772926	0.8015564202334631	229.0
Chaetognata_middle	0.8841463414634146	0.7880434782608695	0.8333333333333334	184.0
Chaetognata_tail	0.425	0.3617021276595745	0.39080459770114945	47.0
Cnidaria	0.7634408602150538	0.7802197802197802	0.7717391304347825	91.0
Copepoda_Calanoida	0.48660328435609335	0.4401876465989054	0.46223316912972084	1279.0
Copepoda_Calanus	0.9443488238668961	0.9109020475926951	0.9273239436619718	3614.0
Copepoda_Gaetanus	0.5294117647058824	0.12162162162162163	0.19780219780219785	74.0
Copepoda_Heterorhabdus	0.16666666666666666	0.006756756756756757	0.012987012987012986	148.0
Copepoda_Metridia_longa	0.83541377716135	0.8788910505836576	0.8566010903057597	2056.0
Copepoda_Microcalanus	0.8532136653155762	0.8976545842217484	0.874870120231557	3283.0
Copepoda_Oithona	0.7601156069364162	0.6399026763990268	0.6948480845442536	411.0
Copepoda_Oncaea	0.8994728915662651	0.8516934046345811	0.8749313312580114	2805.0
Copepoda_Paraeuchaeta	0.4473684210526316	0.288135593220339	0.35051546391752586	118.0
Copepoda_Pseudocalanus	0.47761194029850745	0.64	0.5470085470085471	100.0
Copepoda_Scaphocalanus	0.2645739910313901	0.7564102564102564	0.39202657807308966	78.0
Copepoda_Scolecithricella	0.0	0.0	0.0	22.0
Copepoda_Spinocalanus	0.0	0.0	0.0	10.0
Copepoda_dead	0.0	0.0	0.0	0.0
Crustacea	0.0	0.0	0.0	28.0
Detritus	0.83997967995936	0.9833482010110021	0.906027397260274	3363.0
Eggs	0.898876404494382	0.6374501992031872	0.745920745920746	251.0
Euphausiacea	0.0	0.0	0.0	4.0
Feces	0.6	0.11392405063291139	0.19148936170212763	79.0
Foraminifera	0.0	0.0	0.0	5.0
Nauplii	0.8317757009345794	0.6846153846153846	0.751054852320675	780.0
Ostracoda	0.951417004048583	0.9425133689839572	0.9469442578912022	748.0
Polychaeta	0.8	0.24242424242424243	0.372093023255814	33.0
Rhizaria	0.7709251101321586	0.8793969849246231	0.8215962441314554	199.0
Trochophora	0.0	0.0	0.0	0.0
multiples	0.5561959654178674	0.5561959654178674	0.5561959654178674	347.0
accuracy	0.8314074360585989	0.8314074360585989	0.8314074360585989	20683.0
macro avg	0.5401579053564513	0.48839071664427847	0.49100445865202436	20683.0
weighted avg	0.825690233639791	0.8314074360585989	0.8239490510869542	20683.0

Table S7. Classification Report for Resnet pretrained with dino.

category	precision	recall	f1-score	support
Amphipoda	0.5777777777777777	0.8125	0.6753246753246752	64.0
Antenna	0.3181818181818182	0.2916666666666667	0.30434782608695654	24.0
Bubble	0.943127962085308	0.9521531100478469	0.9476190476190477	209.0
Chaetognata_head	0.758893280632411	0.8384279475982532	0.7966804979253113	229.0
Chaetognata_middle	0.8944099378881988	0.782608695652174	0.8347826086956522	184.0
Chaetognata_tail	0.3225806451612903	0.2127659574468085	0.2564102564102564	47.0
Cnidaria	0.7868852459016393	0.5274725274725275	0.6315789473684211	91.0
Copepoda_Calanoida	0.4967081199707388	0.5308835027365129	0.5132275132275131	1279.0
Copepoda_Calanus	0.9181081081081081	0.939955727725512	0.9289034727919059	3614.0
Copepoda_Gaetanus	0.4827586206896552	0.1891891891891892	0.27184466019417475	74.0
Copepoda_Heterorhabdus	0.6666666666666666	0.04054054054054054	0.07643312101910829	148.0
Copepoda_Metridia_longa	0.888079125455492	0.8297665369649806	0.8579331154136284	2056.0
Copepoda_Microcalanus	0.9086816720257235	0.860798050563509	0.8840919755983107	3283.0
Copepoda_Oithona	0.7555555555555555	0.6618004866180048	0.7055771725032425	411.0
Copepoda_Oncaea	0.8944921316165951	0.8916221033868093	0.8930548116407784	2805.0
Copepoda_Paraeuchaeta	0.40425531914893614	0.16101694915254236	0.23030303030303031	118.0
Copepoda_Pseudocalanus	0.39568345323741005	0.55	0.4602510460251046	100.0
Copepoda_Scaphocalanus	0.35454545454545455	0.5	0.41489361702127664	78.0
Copepoda_Scolecithricella	0.0	0.0	0.0	22.0
Copepoda_Spinocalanus	0.0	0.0	0.0	10.0
Copepoda_dead	0.0	0.0	0.0	0.0
Crustacea	0.0	0.0	0.0	28.0
Detritus	0.8755662136957101	0.9771037763901279	0.9235525576166386	3363.0
Eggs	0.8943661971830986	0.5059760956175299	0.6463104325699746	251.0
Euphausiacea	0.0	0.0	0.0	4.0
Feces	0.5	0.08860759493670886	0.15053763440860216	79.0
Foraminifera	0.0	0.0	0.0	5.0
Nauplii	0.7488372093023256	0.8256410256410256	0.7853658536585366	780.0
Ostracoda	0.9482071713147411	0.9545454545454546	0.9513657561625584	748.0
Polychaeta	0.7857142857142857	0.3333333333333333	0.4680851063829786	33.0
Rhizaria	0.8850574712643678	0.7738693467336684	0.8257372654155496	199.0
Trochophora	0.0	0.0	0.0	0.0
multiples	0.4515539305301645	0.7118155619596542	0.5525727069351231	347.0
accuracy	0.8391432577479089	0.8391432577479089	0.8391432577479089	0.8391432577479089
macro avg	0.5411119204137417	0.47709273275513275	0.48444802146419264	20683.0
weighted avg	0.8392781322620767	0.8391432577479089	0.8336191980482558	20683.0

REFERENCES

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., et al. (2021). Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9650–9660

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*

Kronberg, R. M. (2022). *Applications of Supervised Deep (Transfer) Learning for Medical Image Classification*. Ph.D. thesis, HHU

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning – ICANN 2018*, eds. V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis (Cham: Springer International Publishing), 270–279