A. WildlifeDatasets: Supplementary Materials

A.1. Ablation study on model size

To showcase and quantify the performance of different MegaDescriptor flavors, we compare five variants, e.g., **B**ase, **S**mall, **T**iny, and **L**arge-224 and **L**arge-384, originating from corresponding variations of the Swin architecture. All the models were trained and evaluated using the same setting. Naturally, the model performance in terms of accuracy increased with an increasing model size, i.e., the MegaDescriptor-L-384 outperformed smaller flavors by a considerable margin in most cases. Overall, higher model complexity achieved higher performance with few exceptions, where it underperformed by a small margin, e.g., by 2.53%, 0.48%, and 0.08% on FriesianCattle2017, LeopardID2022, and MacaqueFaces respectively. This is more or less statistically insignificant, given the poor quality of the data and the data acquisition.

We visualized the accuracy of all provided MegaDescriptor flavors in Figure 9 and Table 5.

A.2. Online Documentation – Dataset samples and tutorials

We provide extensive documentation to give users a better orientation within the WildlifeDatasets toolkit and available features. It covers a wide range of use cases of the toolkit, including a guide to installation and dataset downloading, tutorials, and how to contribute. Notably, the documentation includes a detailed description of the datasets, including image samples.
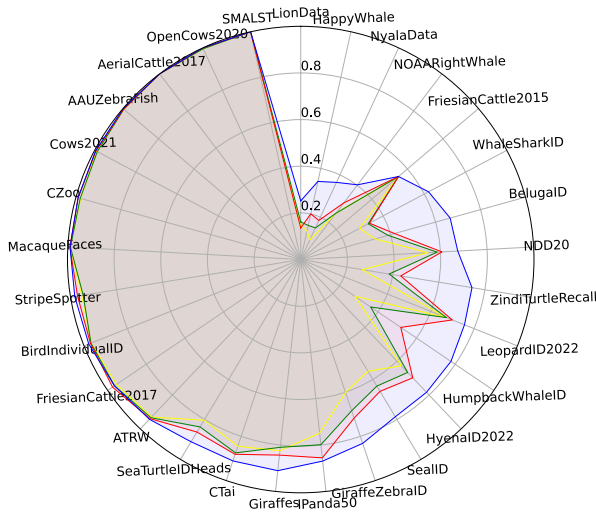
Figure 9. **Ablation study on model size/complexity.** Accuracy of different MegaDescriptor flavors MegaDescriptor-L (Swin-L/p4-w12-384), MegaDescriptor-B (Swin-B/p4-w7-224), MegaDescriptor-S (Swin-S/p4-w7-224), and MegaDescriptor-T (Swin-T/p4-w7-224) on 29 animal re-identification datasets.

	MegaDescriptor-T-224	MegaDescriptor-S-224	MegaDescriptor-B-224	MegaDescriptor-L-224	MegaDescriptor-L-384
AAUZebraFish	99.55	99.55	99.63	99.85	99.93
ATRW	93.02	93.40	93.95	93.67	94.33
AerialCattle2017	100.0	100.0	100.0	100.0	100.0
BelugaID	33.12	38.84	41.74	47.92	66.48
BirdIndividualID	96.73	96.81	97.04	97.21	97.82
CTai	84.46	87.46	88.10	90.68	91.10
CZoo	97.87	98.11	99.05	98.35	99.05
Cows2021	99.13	98.73	99.37	99.37	99.54
FriesianCattle2015	55.00	55.00	55.00	55.00	55.00
FriesianCattle2017	95.45	96.46	97.47	98.99	96.46
GiraffeZebraID	60.15	68.40	71.72	78.04	83.17
Giraffes	82.46	80.97	84.33	87.69	91.04
HappyWhale	12.58	14.98	20.07	25.34	34.30
HumpbackWhaleID	28.12	36.25	51.83	63.54	77.81
HyenaID2022	62.70	66.67	69.84	77.30	78.41
IPanda50	74.84	79.85	85.53	85.45	86.91
LeopardID2022	67.06	67.27	69.92	76.06	75.58
LionData	14.84	16.13	13.55	20.65	25.16
MacaqueFaces	99.04	98.89	99.12	98.96	99.04
NDD20	54.61	58.57	60.64	61.58	67.42
NOAARightWhale	25.16	24.95	30.51	34.69	40.26
NyalaData	9.35	14.95	18.46	21.73	36.45
OpenCows2020	99.58	99.58	100.0	99.79	100.0
SMALST	100.0	100.0	100.0	100.0	100.0
SeaTurtleIDHeads	80.38	83.74	86.31	89.86	91.18
SealID	55.88	63.31	65.95	70.02	78.66
StripeSpotter	95.12	94.51	96.95	97.56	98.17
WhaleSharkID	28.58	32.74	33.31	50.03	62.02
ZindiTurtleRecall	26.77	38.38	43.45	58.14	74.40

Table 5. **Ablation study on model size/complexity**. We compare five MegaDescriptor flavors, e.g., **L**arge, **B**ase, **S**mall, and **T**iny, in terms of accuracy. In general, models with a bigger model size or higher input resolution outperformed their *smaller* variants by a considerable margin. The best-performing model – MegaDescriptor-L-384 – underperformed by 2.53%, 0.48%, and 0.08% on FriesianCattle2017, LeopardID2022, and MacaqueFaces, respectively.