OpenAnimals: Revisiting Person Re-Identification for Animals Towards Better Generalization

Supplementary Material

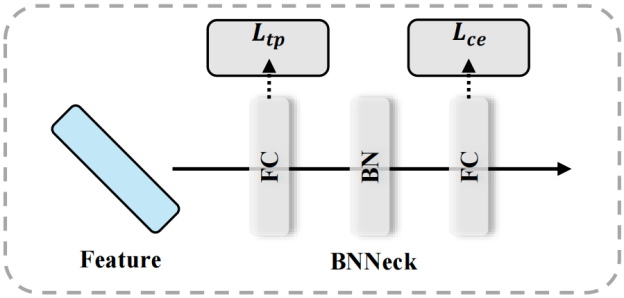

Figure 4. Illustration of BNNeck in BoT [33]. *FC* and *BN* for *Fully-Connected Layer* and *Batch Normalization*, L_{tp} and L_{ce} for *Triplet* and *Cross-Entropy Loss*.

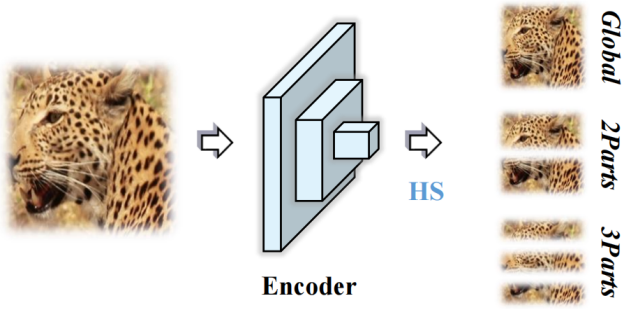

Figure 5. Illustration of Multi-Branch Architectures in MGN [52]. *HS* for *Horizontal Split*.

6. Appendix

6.1. Methodology

6.1.1. Illustration of BNNeck and Multi-Branch

BNNeck [33] and *Multi-Branch Architectures* [52] are two fundamental modules that have demonstrated significant effectiveness in person re-identification and have been incorporated into ARBase for animal re-identification. More specifically, BNNeck introduces a batch normalization layer that decouples the features generated by the backbone into two distinct spaces, which are then utilized for the independent computation of triplet loss and cross-entropy loss, respectively. On the other hand, Multi-Branch Architectures [52] employ a horizontal partitioning approach to divide the extracted features, producing fine-grained, part-level feature representations.

In the main paper, we provide a concise textual description of these modules. For further clarity, the structures and functionalities of these modules are visually illustrated in Figures 4 and 5, respectively, to assist the readers in better understanding their operations.

Table 6. Dataset statistics. *#ID* and *#Img* for number of identities and images in each subset.

Datasets	Train Set		Probe Set		Gallery Set	
	#ID	#Img	#ID	#Img	#ID	#Img
HyenaID [1]	145	1535	110	220	111	1374
LeopardID [2]	260	3058	122	244	170	3504
SeaTurtleID [3]	224	3790	172	344	176	3448
WhaleSharkID [19]	320	3847	197	394	223	3452

Table 7. Ablation study on *Data* and *Backbone*. *IBN* for *Instance-Batch Normalization*, *MB* for *Multi-Branch Architecture*.

Method	Input Resolution	LeopardID		SeaTurtleID	
		R1	mAP	R1	mAP
BoT [33]	[256,128]	54.92	27.65	84.01	41.92
AGW [56]	[256,128]	54.10	28.67	85.17	46.18
SBS [16]	[384,128]	51.23	26.54	84.01	44.63
MGN [52]	[384,128]	53.69	28.21	86.05	46.67
BoT [33]	[384,384]	62.70	34.64	86.05	49.18
AGW [56]	[384,384]	60.66	34.55	88.08	53.80
SBS [16]	[384,384]	59.43	33.12	86.92	52.40
MGN [52]	[384,384]	61.48	33.64	88.66	53.52
ARBase(*w/o IBN*)	[384,384]	**64.34**	36.82	88.37	54.57
ARBase(*w/o MB*)	[384,384]	63.11	34.99	**88.95**	55.39
ARBase(**Ours**)	[384,384]	**64.34**	**37.08**	86.92	**55.99**

6.2. Experiments

6.2.1. Setup

Datasets In Section 3.2, we provided a brief introduction to the datasets used in our experiments, namely HyenaID [1], LeopardID [2], SeaTurtleID [3], and WhaleSharkID [19]. The detailed statistics for these datasets are presented in Table 6. It is important to note that the identities used for training and testing do not overlap.

Implementation Details For person re-identification methods, we adhere to their original configurations. For ARBase, the batch size is set to $[4, 16]$ (4 identities and 16 samples per identity). We use $m = 0.3$ in Eq (1) for the triplet loss and $\epsilon = 0.1$ in Eq (2) for the cross-entropy loss. The initial learning rate is set to 0.00035, and the training lasts for 120 epochs.

6.2.2. More Ablation Studies

In Table 7 and Table 9, we present the ablation studies of ARBase (*i.e.*, *Data*, *Backbone*, *Head*, *Loss*, and *Training & Testing*) on two additional benchmarks, namely LeopardID [2] and SeaTurtleID [3]. The experimental settings remain consistent with those used in the main paper.

Table 8. Performance comparison to ResNet50 equipped with either triplet loss or ArcFace loss under the *open-set* setting, trained separately on each dataset. *R1* for *Rank-1 Accuracy*, *mAP* for *mean Average Precision*.

Method	Input Resolution	HyenaID [1]		LeopardID [2]		SeaTurtleID [3]		WhaleSharkID [19]	
		R1	mAP	R1	mAP	R1	mAP	R1	mAP
ResNet50 + Triplet	[256, 128]	48.18	31.03	47.54	24.99	70.93	33.05	34.01	13.54
ResNet50 + ArcFace	[256, 128]	50.45	26.95	44.67	21.09	72.97	32.29	42.39	14.97
ResNet50 + Triplet	[384, 384]	56.36	37.01	47.95	26.83	75.87	41.90	41.12	16.89
ResNet50 + ArcFace	[384, 384]	53.64	29.91	48.36	20.92	77.03	34.14	48.98	17.86
ARBase (**Ours**)	[384, 384]	**73.18**	**44.87**	**64.34**	**37.08**	**86.92**	**55.99**	**62.44**	**29.45**

Table 9. Ablation study on *Head*, *Loss* and *Training*.

Method	LeopardID		SeaTurtleID	
	R1	mAP	R1	mAP
ARBase (*w/o BNNeck*)	56.97	31.43	76.45	43.44
ARBase (*w/o Label Smoothing*)	**64.34**	37.28	86.92	52.29
ARBase (*w/o Cosine Annealing*)	63.93	**37.78**	**89.53**	**56.52**
ARBase(**Ours**)	**64.34**	37.08	86.92	55.99

Table 10. Performance comparison to existing animal methods [28, 57] using their protocols.

Method	ELPephants		ATRW		
	R1	mAP	R1(s)	R1(c)	mmAP
PGCFL [28]	33.4	18.5	90.8	86.3	66.9
UPBFA [57]	38.7	24.3	92.0	84.6	68.6
ARBase	**60.5**	**43.1**	**96.3**	**86.5**	**70.4**

It is important to highlight that our primary objective is to develop a robust base model that generalizes well across various species, rather than optimizing for the highest performance on a single benchmark. For instance, the highest performance on SeaTurtleID is achieved by ARBase without the use of cosine annealing. To this end, *we adopt the designs for ARBase that have demonstrated effectiveness across at least three different benchmarks.* Despite this focus on generalization, ARBase achieves state-of-the-art performance on all benchmarks, significantly outperforming the baseline models. Notably, ARBase improves the mAP by 9.32% on SeaTurtleID compared to the best-performing baseline, as shown in Table 3.

6.2.3. More Performance Comparison

Comparison to MegaDescriptor [6]. As aforementioned, WildlifeDatasets [6] has made significant efforts to consolidate publicly available datasets for animal re-identification into a unified framework. Moreover, the repository introduces a MegaDescriptor model, against which we offer a detailed comparison with our research.

First, the MegaDescriptor model is trained on 29 publicly available datasets under a *closed-set* setting, indicating that the training data is relatively large in scale, with all identities present during the training phase. In contrast, our ARBase model is trained separately on each dataset using an *open-set* setting, where testing identities are not included in the training data. In our opinion, the open-set setting not only presents greater challenges but also promotes the development of innovative algorithms, as it assesses the model's ability to generalize to unseen identities. Besides, our primary focus is to revisit person re-identification techniques in the context of animal re-identification, aiming to improve generalization by adopting settings similar to those in person re-identification studies [56].

Second, the MegaDescriptor model employs a standard classification backbone with either triplet loss or ArcFace loss[6] for re-identification tasks. However, as demonstrated in person re-identification studies [56], this paradigm may not sufficiently capture the complexities inherent in re-identification tasks. For experimental comparisons, we adopt a ResNet50 backbone within our framework[7], integrating either triplet loss or ArcFace loss under the *open-set* setting, with models trained separately on each dataset. Furthermore, we assess the impact of varying input resolutions on model performance. The experimental results, shown in Table 8, indicate that the performance is significantly inferior to that of our ARBase model.

Comparison to PGCFL [28] and UPBFA [57]. To further validate ARBase, we compare it against two animal methods *using their evaluation protocols* [28, 57] in Table 10. PGCFL [28] proposes a pose-guided complementary feature learning method for tiger re-identification, which enhances feature diversity by guiding network branches to focus on different body regions. UPBFA [57] develops an unsupervised feature alignment method with background removal to address background bias and pose variation in animal re-identification. As shown in Table 10, without any modification, ARBase achieves state-of-the-art results (R1: +21.8%, mAP: +18.8% on ELPephants), confirming its strong generalization ability.

[6] ArcFace loss is a variant of cross-entropy loss that introduces an angular margin. In our experiments, the hyperparameters for this loss are set based on the grid search strategy described in [6].

[7] ResNet50 serves as the backbone in ARBase as well as our revisiting experiments.

Table 11. Statistic analysis. The average and standard deviations are obtained over three runs. *R1* for *Rank-1 Accuracy*, *mAP* for *mean Average Precision*.

Method	HyenaID [1]		LeopardID [2]		SeaTurtleID [3]		WhaleSharkID [19]	
	R1	mAP	R1	mAP	R1	mAP	R1	mAP
BoT [33]	58.18 ± 0.37	34.42 ± 0.80	54.92 ± 0.01	27.89 ± 0.23	83.53 ± 0.36	42.06 ± 0.70	52.71 ± 0.63	20.86 ± 0.01
AGW [56]	56.36 ± 0.37	33.00 ± 0.54	54.24 ± 0.19	28.76 ± 0.22	85.37 ± 1.67	45.65 ± 0.64	52.29 ± 1.08	20.82 ± 0.28
SBS [16]	52.27 ± 0.37	30.29 ± 0.21	52.60 ± 1.93	27.13 ± 0.57	83.63 ± 1.68	44.69 ± 0.11	46.45 ± 0.75	18.64 ± 0.40
MGN [52]	55.45 ± 0.37	31.12 ± 0.19	54.10 ± 0.89	28.02 ± 0.50	86.15 ± 0.14	46.68 ± 0.27	50.51 ± 0.55	21.24 ± 0.23
ARBase(**Ours**)	72.73 ± 0.64	44.34 ± 0.38	64.48 ± 0.19	37.09 ± 0.38	87.31 ± 0.76	55.67 ± 0.42	61.68 ± 0.90	29.41 ± 0.31

Table 12. Comparison of model complexity in terms of parameter count and FLOPs.

Method	Param.	FLOPs	Method	Param.	FLOPs
BoT [33]	23.5M	18.3G	SBS [16]	23.6M	18.4G
AGW [56]	23.6M	18.4G	MGN [52]	68.8M	42.0G
ARBase (*w/o MB*)	23.5M	18.3G	ARBase (*w/ MB*)	64.6M	42.0G

Table 13. Effect of training data scale. The results are averaged over three runs.

Method	Ratio of Training IDs	HyenaID		WhaleSharkID	
		R1	mAP	R1	mAP
ARBase	25%	58.94	33.75	51.19	19.69
	50%	64.55	36.68	57.61	25.12
	75%	68.33	41.63	59.81	27.17
	100%	73.18	44.87	62.44	29.45

6.2.4. Extended Model Analyses

Statistical Analysis. To ensure fair comparisons, we fixed the random seeds across all experiments mentioned above, maintaining consistent dataloader behavior across all methods. Furthermore, we repeated the comparison presented in Table 3 using three different random seeds. The resulting averages and standard deviations, reported in Table 11, further confirm the robustness and consistent performance improvements of ARBase.

Comparison of Model Complexity. Table 12 compares ARBase with baseline models in terms of parameter count and FLOPs. The complexity of ARBase without *Multiple Branch* is similar to that of BoT/AGW/SBS, while ARBase with *Multiple Branch* is comparable to MGN. Notably, even without *Multiple Branch*, ARBase still significantly outperforms these baselines including MGN, as demonstrated in Table 4 and Table 7. This confirms that ARBase's performance gains are not due to increased model capacity.

Effect of Training Data Scale. Animal ReID datasets are generally smaller than those in human ReID due to data collection difficulties. In Table 13, we conduct an ablation study on training data scale using the same evaluation protocol. Specifically, we randomly sample a fraction of training identities for model optimization and report average results over three runs. The performance comparison shows that current datasets remain insufficient.

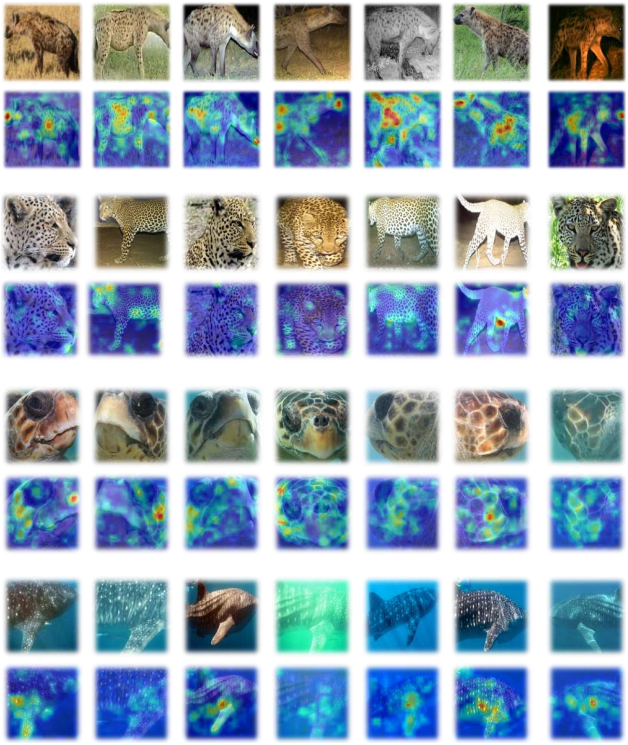

Figure 6. Visualization of heatmaps. These heatmaps are obtained by applying max pooling along the channel dimension of the features extracted by the backbone.

Feature Visualization. In Figure 6, we present the heatmaps generated by the backbone of ARBase trained on each benchmark, highlighting the key regions used for the re-identification of each species. From these visualizations, it is evident that animal re-identification predominantly relies on body or head texture features. This observation is intuitive, as individual animals exhibit unique texture patterns, much like human fingerprints. Interestingly, this finding is consistent with conclusions from biological research [18], further validating the critical role of texture features in distinguishing individual animals across species.

6.3. Discussion

6.3.1. Discussion on Novelty

Novelty (and value) come in many forms in papers[8]. In our opinion, the value of a paper should be judged based on *whether it advances the field and inspires future work*. When we began addressing animal ReID, we faced three significant challenges outlined in Section 3: (1) ***Lack of a Flexible Codebase***, (2) ***Unclear Generalization from Person ReID***, and (3) ***Necessity of a Strong Base Model***. To tackle these issues, we devoted substantial efforts to building OpenAnimals, revisiting person ReID paradigms, and developing ARBase. We believe these contributions can *serve as valuable resources for researchers in this field, sparing them significant time and effort while fostering progress*.

6.3.2. Discussion on Writing

Our manuscript is not organized in the same manner as most regular CVPR papers. The organization draws inspiration from impactful works [10, 33] which adopt similar approaches to structuring their contributions. Animal ReID is an emerging field, and it remains unclear which techniques and methodologies from person ReID can be effectively generalized to this domain. This lack of clarity poses significant challenges to the development of animal ReID and directly motivates our work.

6.3.3. Discussion on Future Work

Our study makes a meaningful contribution towards advancing animal re-identification, yet this task warrants further exploration. Here, we identify some promising directions for future research:

(a) **Attribute-assisted Animal Re-Identification**: Semantic attributes, such as gender and age, are useful auxiliary tools for person re-identification. For animal re-identification, summarizing long-term attributes could enhance identity recognition.

(b) **Video-based Animal Re-Identification**: Due to the challenges associated with data collection and annotation, current benchmarks for animal re-identification are primarily image-based. Videos, however, provide richer information and could be more promising for accurate animal re-identification.

(c) **Generalizable Animal Re-Identification**: Animal re-identification involves various species, making it valuable to develop a generalized model capable of recognizing multiple species. This is particularly feasible with the emergence of Large Language Models (LLMs) that encapsulate rich knowledge across different species.

[8] https://medium.com/@black_51980/novelty-in-science-8f1fd1a0a143 by Michael Black, Director at the Max Planck Institute for Intelligent Systems.