Towards Multi-Species Animal Re-Identification

Maik Fruhner
University of Applied Sciences Osnabrueck
Albrechtstrasse 30
49076, Osnabrueck, Germany
m.fruhner@hs-osnabrueck.de

Prof. Dr. Heiko Tapken
University of Applied Sciences Osnabrueck
Albrechtstrasse 30
49076, Osnabrueck, Germany
h.tapken@hs-osnabrueck.de

ABSTRACT

Animal Re-Identification (ReID) is a computer vision task that aims to retrieve a query individual from a gallery of known identities across different camera perspectives. It is closely related to the well-researched topic of Person ReID, but offers a much broader spectrum of features due to the large number of animal species. This raises research questions regarding domain generalization from persons to animals and across multiple animal species. In this paper, we present research on the adaptation of popular deep learning-based person ReID algorithms to the animal domain as well as their ability to generalize across species. We introduce two novel datasets for animal ReID. The first one contains images of 376 different wild common toads. The second dataset consists of various species of zoo animals. Subsequently, we optimize various ReID models on these datasets, as well as on 20 datasets published by others, with the objective of evaluating the performance of the models in a non-person domain. Our findings indicate that the domain generalization capabilities of OSNet AIN extend beyond the person ReID task, despite its comparatively small size. This enables us to investigate real-time animal ReID on live video data.

Keywords

re-identification, deep learning, computer vision, animals

1 INTRODUCTION

Re-identification (ReID) within computer vision pertains to the identification of individuals among various images of different camera angles. The complexity arises from diverse factors like pose, lighting, obstructions and appearance discrepancies, such as alterations in clothes, accessories, hairstyles in humans, or shifts in fur, feather patterns, and skin in the animal domain. To tackle this challenge, modern ReID systems commonly employ deep learning algorithms to extract image features, followed by a similarity measure to determine matches.

The ReID of animals is an active field of research [Rav+20] that faces challenges due to the sheer diversity and different appearance of the various species and the fact that they are often difficult to distinguish within a species by non-experts.

Because of this, current animal ReID literature mostly focuses on a single species with manually crafted features such as skin landmarks, scars, fur patterns and face recognition. Only few papers show feasible results in a cross species setup.

Our work addresses the adaptation and optimization of various person ReID algorithms to the animal domain. We demonstrate the effectiveness of established CNN-based person ReID algorithms on two datasets created by our own as well as several open source datasets. Our new datasets are made public to the research community with download links provided in the summary.

2 RELATED WORK

Image-based re-identification of animals has been an active research topic for many years. Photo identification of animals can be traced back to 1996, when Raj investigated the possibilities of recognizing wild marine animals over several years by hand [Raj98].

Methods for animal ReID based on artificial neural networks were not introduced until years later. Especially the rise of CNNs has brought new ideas and possibilities into the field of re-identification in general. In the following we present related work based on the person and animal ReID tasks. Our research focuses on the domain generalization between persons and animals and not on the development of completely new algorithms for animal ReID, as AI models for human ReID have already proven successful. We therefore try to build upon these findings instead of starting from scratch. Over the past years ReID-specific models have been developed for the person recognition task. Some well performing ones have been implemented in the highly popular Torchreid [Zho+19a] framework, which was used for our study.

Yu et al. [Yu+17] present a ReID model based on ResNet-50, in which not only the high-level features of the output layer are used. Instead, the modified architecture has a parallel branch in the last residual block, which taps the results of the two penultimate layers. According to the authors, these should contain the *mid-level* features. To calculate the overall result, the features of all three final layers are combined before the loss function is applied.

The "Multi-Level Factorization Net" (MLFN), presented by Chang et al. [Cha+18], is based on the idea that more features are needed for a robust ReID than a camera image from a single perspective. The researchers investigated the possibility of automatically learning and finding view-independent discriminative features and combined their results in a new network architecture.

Sun et al. [Sun+18] use an approach that internally divides the image into several areas in order to examine and compare important features at the part level. The part-based convolutional baseline (PCB) network splits an input image into p different fragments, which are stacked vertically to represent different body parts. For $p = 6$ these can be head, shoulders, chest, hips, legs and shoes. These sections are used to compare them with the corresponding parts of other images.

Li et al. [Li+18] investigated the problem that people are not always perfectly aligned within their bounding boxes. The team addressed this phenomenon using the attention mechanism. A novel module for *Harmonious Attention* (HA) is able to learn hard and soft attentions, which are tailored for coarse and detailed features respectively.

Zhou et al. [Zho+19b] state that features are to be found not only on multiple, but on all scaling levels. They therefore define an "omni-scale" approach, which is a hybrid of different homogeneous and heterogeneous scaling features. Based on this approach the authors present the Omni Scale Network (OSNet). A novel deep convolutional network family, which is an order of magnitude smaller than ResNet-50, but at the same time achieves better results in the ReID task. According to the authors, "omni-scale feature learning" also proves to be a useful approach for other computer vision tasks. The OSNet [Zho+19b] and OSNet AIN [Zho+21] model families have shown outstanding results in the person ReID task and multi-dataset domain generalization scenarios.

Like many current approaches, the latest ReID advances are based on transformers. Vision transformers (ViT) [Dos+20] show remarkable results in various computer vision tasks, although they have not been researched as long as CNNs. An early attempt to address person ReID via ViT is TransReID [He+21]. The authors justify this fundamentally new strategy by arguing that ViT has the advantage over CNN approaches of being able to better understand the global context of the image input and also to better recognize fine details. This approach was recently further improved to the SOLIDER architecture [Che+23], which uses the SwinTransformer presented by Microsoft [Liu+21].

Ravoor et al. conducted a survey on animal ReID [Rav+20] and mention several studies that use person ReID models for this topic. They found that (variants of) PCB and ResNet50 were frequently used for feature extraction and as backbones, respectively. However, they conclude that PCB might not be suitable for animal ReID due to its vertical structure intended for analyzing the human upright pose.

Schneider et al. compared the siamese and triplet-loss similarity methodologies based on different CNN architectures [Sch+20] for the animal ReID task. They used one person dataset and four animal datasets and found that the triplet-loss comparisons can outperform human observers for the selected datasets.

A notable development for animal ReID is *MegaDescriptor* presented at the beginning of 2024 by Cermák et al. [Čer+24]. MegaDescriptor is intended to be a foundation model that can solve many computer vision tasks relating to animals, including ReID. The authors show impressive results across 29 public datasets. However, the authors treat animal ReID as a closed world classification problem, where all the animals to be found in the gallery set are already present during training. In the person ReID setting we adopt, training and evaluation sets are disjoint, so no ID specific features can be learned by the model.

3 DATASETS

The difference between the ReID of people and animals lies in the diversity of appearances of different animal species and the method of data acquisition. While the task for humans mostly involves processing pedestrians on surveillance cameras, the development of animal focused algorithms is much more diverse due to many factors.

3.1 Public

An increasing number of animal datasets with annotations on the identity of the individuals can be found online. Due to permissive licenses, they are often also

Dataset Name	# Images	# IDs
AerialCattle2017 [And+17]	46340	23
ATRW [Li+19]	5415	182
BelugaID [Lil22a]	5902	788
Cows2021 [Gao+21]	8670	181
FriesianCattle2015 [Til+16]	377	40
FriesianCattle2017 [And17]	940	89
GiraffeZebraID [Par+17]	6925	2056
HappyWhale [Che+22]	51033	15587
HumpbackWhaleID [How+18]	15697	5004
HyenaID2022 [Lil22b]	3129	256
IPanda50 [Wan+21]	6874	50
LeopardID2022 [Lil22c]	6806	430
NDD20 [Tro+20]	2657	82
NOAARightWhale [Chr15]	4544	447
NyalaData [Dla+20]	1942	237
OpenCows2020 [Wil+20]	4736	46
SealID [Nep+22]	2080	57
SeaTurtleID [Ada+24]	7774	400
StripeSpotter [Lah+11]	820	45
WhaleSharkID [Hol+09]	7693	543
ZindiTurtleRecall [Zin23]	12803	2265

Table 1: Evaluated public datasets

available for further research. In most cases, a dataset contains animals of exactly one species. A database that shows and annotates different species in multiple videos was published by Kuncheva et al. [Kun+22]. They aggregated a dataset on pigs, koi and pigeons with a total of 93 identities.

However, a problem with using and combining many public datasets is that often each research team publishes their data in a non standard format. As a result, a great effort of pre-processing work is required to integrate all the necessary datasets into the training process. This problem was addressed by Cermak et al. with the fairly new Wildlife Toolkit [Čer+24]. The framework bundles various datasets into a unified Python API. This allows researchers to download and use public animal ReID datasets in a streamlined workflow without the need for manual data pre-processing and conversion.

For our work, we selected 21 medium to large-sized datasets showing whole bodies of wild animals, zoo animals and farm animals. Datasets containing very few individuals or only showing animal faces were not considered. The evaluated datasets and their references are listed in table 1.

3.2 Ours

Additionally, we introduce two novel datasets for animal re-identification. The first one, ToadID [Fru+24b], contains images of 376 individual common toads from different camera angles. The second one is named ZooMixID [Fru+24a] and contains images of 180 animals of five different species.

Figure 1: A toad from the ToadID dataset captured from five camera angles

Perspective	# of Images
Front	1513
Left	983
Right	1025
Back	985
Top	2739
Total	**7245**

Table 2: Summary of the ToadID dataset

3.2.1 ToadID

During the spring seasons of 2022 and 2023, a conservation effort in southern Lower Saxony, Germany, led to the rescue of more than 400 toads at a local lake. As there is currently no public dataset about toads available, these animals were recorded on video, before they were released at their natural habitats. Each video was carefully crafted to showcase only one toad at a time from various angles, all under one minute in length. Out of the total videos produced, 376 were deemed suitable for use, providing an equal number of unique toad identities for the research dataset.

Videos were processed to extract frames at a rate of one frame per second. A select subset of these frames received bounding box annotations to facilitate the creation of a preliminary object detection dataset. These annotated frames were used in training a Yolov5m object detector, which was subsequently utilized to extract the animals from the remaining images.

The result of this effort is a comprehensive dataset containing 7,245 unique images, representing 376 distinct identities of common toads. These images are categorized according to five different camera perspectives as listed in table 2: front, back, top, left, and right. Figure 1 gives an example of the dataset by displaying images of a single toad identity captured from all five viewpoints.

3.2.2 ZooMix

The objective of the second dataset is to present a greater ReID challenge by being smaller in size while

Figure 2: Examples of each species from the ZooMix multi domain dataset

Species	**# of Images**	**# of IDs**
Camel	92	5
Goat	144	30
Penguin	149	24
Toad	183	50
Tortoise	272	51
Total	**840**	**160**

Table 3: Summary of the ZooMix dataset

at the same time ranging across multiple animal domains. It serves as the basis for exploring two specific hypotheses. The first hypothesis questions whether a re-identification task remains feasible with a limited amount of training data. The second hypothesis examines whether the inclusion of highly distinct species benefits or hinders the training process, specifically whether it enhances the overall outcome by providing diversity or if it introduces complications that degrade the performance for individual species.

It contains 840 images featuring 160 individual animals of five distinct species: tortoise, camels, penguins, goats, and a selection of toads from the previous dataset. With the exception of the toads, these animals were filmed over several weeks at a local zoo. Identifying animals in a zoo with computer vision might be of interest for the employees to support their daily tasks. However, unlike the scenario with the toads, filming each animal individually was impractical due to the zoo's environment, necessitating the subsequent extraction of individuals through manual annotation. figure 2 presents an example of each species. The dataset's composition is detailed in table 3.

4 EXPERIMENTS

Our experiments on the transfer of personen ReID algorithms to the animal domain were carried out on the university's HPC cluster containing multiple A100 GPUs using the public as well as own datasets described above. To save time and computational resources, not all possible permutations of models and hyper parameters were tested on all datasets. Instead, the ToadID dataset was used in a grid search to generally determine whether there are person ReID models that are suitable for the identification of animals. The best performing model was then also trained on the remaining datasets.

The model architectures examined are (see chapter 2):

- Harmonious Attention CNN (hacnn) [Li+18]
- Multi-level Factorisation Net (mlfn) [Cha+18]
- Omni-Scale Net (osnet) [Zho+19b]
- Omni-Scale Net with Batch Normalization (osnet_ibn) [Zho+19b]
- Omni-Scale Net with Instance Normalization (osnet_ain) [Zho+19b]
- Part-based convolutional baseline (pcb) [Sun+18]
- Resnet50 with Mid-level Representations (resnet50mid) [Yu+17]

The OSNet models come in different scales, later indicated by an x, followed by a scaling factor. The PCB model was used with $p = 4$ and $p = 6$, representing the number of parts used for splitting the inputs.

A crucial step that strongly influences the outcome of an experiment is the organization of the input datasets. In our study, we used Torchreid's Train/Query/Gallery approach. In this scenario, the individuals in the training dataset are disjoint from those in the reference gallery. Therefore, no animal seen during the evaluation was seen while training before. This results in the model learning general features and patterns rather than the details of individual identities.

Each dataset and model combination was trained in three different training/test splits, which are 75/25, 50/50 and 25/75. We expect that the larger the test split, the more difficult the task becomes, as there are not only fewer identities to train on, but also more identities to choose from when testing.

There are also several approaches for distributing the remaining data to the query and gallery datasets. It must be decided whether an identity can appear multiple times in the gallery, which increases the probability of finding a correct match (by chance). This is called a multi-shot gallery, in contrast to the single-shot gallery, which contains only one image per identity. In our experiments, we investigate both scenarios, where each identity in the test set is represented by exactly one image in the query set (i.e., each animal must be found once in the gallery). During evaluation we consider this task to be a closed world scenario, meaning that each individual in the query can be found inside the gallery. An open world task, where unknown identities might appear, will be studied in future experiments.

Due to the nature of the different postures of animals, we have adapted the input layer of all CNNs. Person ReID models usually define a rectangular, portrait-oriented input layer to depict standing persons in a minimal bounding box. For the transfer to the animal domain, we decided to use a uniform, square input layer, as animals might appear in any orientation. The ReID models that already define a fixed input size were adapted accordingly. Tests were carried out with 256x256 pixels input size. Preliminary tests showed that increasing the input size to 512x512 had no positive effect on the results, apart from a huge increase in allocated VRAM and longer training times.

In summary, a total of 14 models were studied in a grid search over 30 epochs each with a square input shape of 256x256 pixels. The following hyper parameter permutations were evaluated resulting in a total of 672 runs for the model search:

- randomly initialized weights vs. weights pre-trained on person datasets
- softmax vs. triplet loss functions
- single-shot vs. multi-shot gallery setting
- sequential vs. random data sampler
- 75/25, 50/50 and 25/75 data splits

For the evaluation metric, we used the ReID ranking system typical for persons. For this metric, feature vectors are calculated from input images by the convolutional neural networks. These can be compared by measuring their distances (Euclidean in our case) in a high-dimensional space. The distances are ranked in ascending order, resulting in a top-k list of predictions. We report the Rank-1/-5/-10 results of our experiments. All training runs were carried out with deterministic calculation modes of all relevant software components in order to make the results comparable between models and datasets as well as reproducible by others.

5 RESULTS

We present our results in two sections. First, we show how the different models performed on our ToadID dataset in order to deduce which models might be generally suitable for animal ReID. Then we highlight the test results of the other datasets on the best model.

5.1 Model Search

The model search revealed some clear insights into the potentials for a domain adaptation between persons and animals. We summarize the results, as not all 672 runs can be displayed here. Firstly, pre-training on the human domain clearly helps the models listed in chapter 4 to recognize animals as well. The top 10 models in the search results all used pre-trained weights, while all but a few of the randomly initialized models occupied the last ranks. No model without pre-training achieved an

Model	mAP	R-1	R-5	R-10
hacnn	79.0	92.2	98.9	98.9
mlfn	65.5	81.1	90.0	93.3
osnet_ain_x0_25	83.0	95.6	98.9	100.0
osnet_ain_x0_5	90.9	97.8	100.0	100.0
osnet_ain_x0_75	93.6	98.9	98.9	100.0
osnet_ain_x1_0	94.4	98.9	100.0	100.0
osnet_ibn_x1_0	74.4	87.8	97.8	97.8
osnet_x0_25	83.8	94.4	100.0	100.0
osnet_x0_5	91.6	97.8	100.0	100.0
osnet_x0_75	92.6	96.7	98.9	98.9
osnet_x1_0	94.8	98.9	100.0	100.0
pcb_p4	87.2	95.6	97.8	98.9
pcb_p6	88.6	95.6	96.7	96.7
resnet50mid	92.5	97.8	100.0	100.0

Table 4: Model search results with a data split of 75/25, pretrained weights and a multi-shot setting

mAP and Rank-1 score greater than 48.6 and 71.3 respectively in the case of a 25/75 data split and a multi-shot setting. We found that the single-shot scenario removed too many gallery images from the task, as each individual is only shown in one image, making the task much simpler. Therefore, we chose to use the multi-shot setting with pre-trained weights for the remaining experiments.

As mentioned in chapter four, the split had an immense impact on the reported model performances. With much training data and few query / gallery IDs, almost all models achieved high ranking scores, as shown in table 4. This also holds true for the 50/50 split shown in table 5. However, the results become more meaningful as soon as the number of training samples is reduced to 25% and the number of possible individuals in the gallery is increased. Table 6 shows the corresponding outcomes of a model search using a random data sampler, triplet loss and pretrained weights for a multi-shot gallery containing 75% of the animal IDs. It can be seen that the OSNet family in particular continues to achieve high scores, while the results of other model architectures seem to fall off.

5.2 Single Dataset

Based on the results of the model search, the generalization to different animal domains can be investigated. In addition to our two datasets, 20 public datasets were used to train OSNet AIN (osnet_ain_x1_0). A dataset split of 20/80 was used in accordance with the training configuration of the WildlifeToolkit authors. The results are listed in table 7. The datasets marked with an asterisk have been modified to make them more suitable for the train/query/gallery evaluation method. Although they show the animals from several camera angles, the viewing angles are so drastically different that a match-

Model	mAP	R-1	R-5	R-10
hacnn_square	63.1	84.3	92.4	95.1
mlfn	52.4	72.4	86.5	93.0
osnet_ain_x0_25	71.1	89.7	96.2	99.5
osnet_ain_x0_5	81.8	94.6	100.0	100.0
osnet_ain_x0_75	84.0	95.7	97.8	99.5
osnet_ain_x1_0	89.1	95.7	98.9	100.0
osnet_ibn_x1_0	57.9	80.0	91.9	94.1
osnet_x0_25	71.9	89.7	95.1	97.8
osnet_x0_5	82.7	94.1	98.4	99.5
osnet_x0_75	84.4	94.6	99.5	100.0
osnet_x1_0	89.2	98.4	98.9	100.0
pcb_p4	77.1	90.3	95.7	97.8
pcb_p6	79.8	92.4	96.2	97.8
resnet50mid	83.5	94.6	98.9	99.5

Table 5: Model search results with a data split of 50/50, pretrained weights and a multi-shot setting

Model	mAP	R-1	R-5	R-10
hacnn	46.8	70.9	84.4	88.3
mlfn	28.6	52.1	69.1	77.0
osnet_ain_x0_25	57.7	83.7	92.6	95.7
osnet_ain_x0_5	63.1	85.5	93.6	96.8
osnet_ain_x0_75	66.7	86.2	94.7	**97.5**
osnet_ain_x1_0	**71.7**	**90.4**	94.3	**97.5**
osnet_ibn_x1_0	35.8	58.9	75.2	81.6
osnet_x0_25	55.9	78.7	93.3	95.4
osnet_x0_5	64.3	83.3	93.3	95.7
osnet_x0_75	67.4	85.8	94.7	96.1
osnet_x1_0	70.2	87.6	**95.0**	97.2
pcb_p4	57.9	77.0	87.6	92.2
pcb_p6	57.3	78.4	91.8	94.7
resnet50mid	65.5	85.8	94.0	95.7

Table 6: Model search results with a data split of 25/75, pretrained weights and a multi-shot setting

ing was not possible. Therefore, only the camera angle with the most images was retained.

Our ToadID dataset again achieved a high rank-1 performance of 85.0%, while the model did not reach more than 50% rank-1 accuracy for any other dataset. Compared to the performance of OSNet AIN on small and large person ReID datasets, it can be said that domain adaptation works. Zhou et al. [Zho+21] report 38.3%, 68.0% and 86.6% rank-1 scores on the small GRID, VIPeR and CUHK01 datasets, respectively. Rank-1 results of 94.8%, 72.3% and 88.7% were obtained for the large datasets Market1501, CUHK03 and Duke, respectively. Considering that many animal ReID datasets contain less than a hundred IDs, some between 200 and 1000, while very few datasets contain more than a thousand different animals, the overall results of the training runs for single datasets show solid performances.

dataset	mAP	r1	r5	r10
ToadID (ours)	62.6	85.0	92.4	95.0
OpenCows2020	54.0	48.6	62.2	62.2
ATRW	52.1	47.2	56.0	60.4
Cows2021	52.0	45.5	57.2	66.2
StripeSpotter	31.8	43.8	59.4	65.6
FriesianCattle2017	49.0	39.1	60.9	67.2
HyenaID2022	19.0	36.5	58.9	66.0
SeaTurtleIDHeads	12.9	35.7	53.6	60.4
GiraffeZebraID	31.0	35.3	41.8	45.4
ZooMix (ours)	43.6	34.9	50.8	65.9
HumpbackWhaleID	32.8	26.1	39.0	45.3
LeopardID2022	15.1	21.8	38.2	44.7
FriesianCattle2015	33.4	18.8	46.9	84.4
ZindiTurtleRecall	7.7	17.1	46.3	54.9
WhaleSharkID	8.9	15.2	29.6	36.3
SealID	15.4	13.3	15.6	20.0
BelugaID*	16.9	11.6	22.2	27.8
NOAARightWhale	10.8	10.0	11.4	12.0
AerialCattle2017	14.1	10.0	20.0	20.0
HappyWhale	11.7	8.3	14.3	17.8
NyalaData	6.1	6.4	18.1	27.1
NDD20*	6.6	3.1	9.4	14.1
IPanda50	3.3	2.5	2.5	5.0

Table 7: Results for single dataset training runs

5.3 Multi Dataset

Interesting effects were observed when different datasets are combined in a training run and jointly influence the learning process of osnet_ain_x1_0. Using multiple datasets from the same species (e.g. cattle) as the training sources results in a significantly larger number of images to learn from. As a result, all tested cattle datasets receive a massive improvement in rank-1 scores when used together, as opposed to when used for training individually. The improved results are displayed in table 8.

However, the combination of seemingly independent datasets can also lead to an improvement in model performance. While blindly joining all datasets does not improve the model's performance, successes were achieved when merging datasets of somewhat visually related species. As shown in table 9 animals living on land and in the sea were combined in two experiments, respectively. In the land inhabitants, the rank-1 results of half of the species were improved, while the other half declined minimally. Three quarters of the results of the evaluated datasets improved for the marine species. We observed that overall, datasets with good recognition scores can help the weaker, usually smaller datasets. We assume that the large increase in training data makes it easier to train the feature extractors and that the learned features are therefore (at least partially) transferable between the domains.

dataset	mAP	r1	r5	r10
OpenCows2020	61.1	59.5	62.2	62.2
FriesianCattle2017	66.8	59.4	73.4	82.8
Cows2021	56.5	51.0	63.4	70.3
AerialCattle2017	45.1	45.0	45.0	45.0
FriesianCattle2015	44.6	31.2	56.2	90.6

Table 8: Results improve when combining datasets from the same domain (cattle).

dataset	mAP	r1	r5	r10
ToadID	62.9	82.7	91.0	93.4
HyenaID2022	21.7	**38.6**	52.3	62.9
StripeSpotter	32.0	34.4	56.2	59.4
GiraffeZebraID	29.1	32.4	40.2	47.1
LeopardID2022	18.3	**24.8**	41.6	46.9
NyalaData	8.5	**10.1**	23.4	38.3
HumpbackWhaleID	30.0	23.6	36.4	42.4
WhaleSharkID	8.7	**16.6**	27.7	35.2
BelugaID	17.8	**12.4**	23.1	27.8
NDD20	11.1	**7.8**	12.5	15.6

Table 9: Results can improve when combining datasets from different domains (Top: *Land*, Bottom: *Marine*). Rank-1 improvements (highlighted in bold) of up to almost 5% can be observed.

6 CONCLUSION

6.1 Summary

In this paper we investigated the transferability of algorithms for person ReID to animals of different species. Using a cross-search through different CNN-based models and hyper parameters, the family of OSNets was found to be suitable. We applied OSNet AIN to over 20 different datasets, two of which we created ourselves.

While small datasets suffer from too few training examples, some larger marine datasets present a major challenge with the task of matching only fins or fins with underwater images. The application of the comparison of two images from different perspectives - as it's defined for person ReID - has only proven successful for some datasets. Best results were achieved when the animal ReID task was closer to the human domain. Our ToadID dataset with high-resolution, pre-cropped images showing feature-rich animal textures achieved the highest rank-1 results in our experiments.

OSNet AIN achieves reasonable results with its 2.2M parameters in the standard configuration without much customization. The rank-1 scores across the several animal datasets are comparable to those across person ReID datasets. In some cases, ReID performance can be improved by combining multiple datasets from different animal species.

6.2 Discussion

Due to the different approaches of individual research teams, such as the creation and splitting of datasets, the use of randomized or deterministic calculations and the way experiments and their results are presented, findings are difficult to compare.

Although many wildlife datasets have been streamlined into a single API framework, they still have very different structures and content. Some datasets contain fully cropped images, others provide full images with bounding box or even segmentation annotations. Therefore, not every dataset can be effectively combined with others to train a common task.

Furthermore, the reported model performances cannot be easily compared with each other, as research teams use different evaluation measures for the ReID of animals. Many see the task as a closed-world classification problem, where the training IDs should be identified in a reference set during testing. However, in person ReID, and thus in our context, the training and test IDs are disjoint, which makes the task much more difficult.

Finally, for many datasets, there is no default split between training and testing data, so other researchers have to create their own split. The WildlifeToolkit, for example, applies an automatically generated split to each dataset for the closed-world scenario. However, this split does not fit our setting in the person ReID transfer context. This forced us to create a different split using the disjoint split method, which naturally leads to a rather unrelated research question.

6.3 Outlook

The potential of omni-scale learning will be further investigated in subsequent experiments. More in-depth hyperparameter searches and investigations into adjustments to the network architecture are just two starting points for further improving animal ReID using OSNets.

A third dataset is being developed that differentiates pigs using top-down recordings. Pig farming is a good example of the relevance of the closed-world scenario examined in this paper, in which no unknown individuals can occur. The wildlife datasets, on the other hand, already indicate their open-world setting in their name, which is also part of our further research. In addition to publishing our datasets, we will contribute to the WildlifeToolkit to integrate the datasets directly so that other researchers can easily use them.

7 REFERENCES

[Ada+24] Lukáš Adam et al. "SeaTurtleID2022: A Long-Span Dataset for Reliable Sea Turtle Re-Identification". In: *Proceedings of* *the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2024, pp. 7146–7156.

[And+17] William Andrew, Colin Greatwood, and Tilo Burghardt. "Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning". In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. 2017, pp. 2850–2859.

[And17] Will Andrew. *FriesianCattle2017*. 2017.

[Čer+24] Vojtěch Čermák et al. "WildlifeDatasets: An Open-Source Toolkit for Animal Re-Identification". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2024, pp. 5953–5963.

[Cha+18] Xiaobin Chang, Timothy M. Hospedales, and Tao Xiang. "Multi-level Factorisation Net for Person Re-identification". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2018, pp. 2109–2118. ISBN: 978-1-5386-6420-9.

[Che+22] Ted Cheeseman et al. *Happywhale - Whale and Dolphin Identification*. 2022. URL: https://kaggle.com/competitions/happy-whale-and-dolphin.

[Che+23] Weihua Chen et al. "Beyond Appearance: A Semantic Controllable Self-Supervised Learning Framework for Human-Centric Visual Tasks". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2023, pp. 15050–15061. ISBN: 979-8-3503-0129-8.

[Chr15] Wendy Kan Christin B. Khan Shashank. *Right Whale Recognition*. 2015. URL: https://kaggle.com/competitions/noaa-right-whale-recognition.

[Dla+20] Nkosikhona Dlamini and Terence L van Zyl. "Automated Identification of Individuals in Wildlife Population Using Siamese Neural Networks". In: *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. 2020, pp. 224–228.

[Dos+20] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *International Conference on Learning Representations* (2020).

[Fru+24a] Maik Fruhner, Heiko Tapken, and Eva Stroetmann. *Images of 180 individual zoo animals*. 2024. URL: https://doi.pangaea.de/10.1594/PANGAEA.967637.

[Fru+24b] Maik Fruhner, Heiko Tapken, and Eva Stroetmann. *Images of 376 individuals of common toads (Bufo bufo) from southern Lower Saxony, Germany*. 2024. URL: https://doi.pangaea.de/10.1594/PANGAEA.967135.

[Gao+21] Jing Gao et al. *Towards Self-Supervision for Video Identification of Individual Holstein-Friesian Cattle: The Cows2021 Dataset*. 2021. arXiv: 2105.01938 [cs.CV].

[He+21] Shuting He et al. "TransReID: Transformer-based Object Re-Identification". In: *2021 IEEE/CVF International Conference on Computer Vision*. Ed. by Eric Mortensen. Piscataway, NJ: IEEE, 2021, pp. 14993–15002. ISBN: 978-1-6654-2812-5.

[Hol+09] Jason Holmberg, Brad Norman, and Z Arzoumanian. "Estimating population size, structure, and residency time for whale sharks Rhincodon typus through collaborative photo-identification". In: *Endangered Species Research* 7 (Apr. 2009), pp. 39–53.

[How+18] Addison Howard, Ken Southerland, and Ted Cheeseman. *Humpback Whale Identification*. 2018. URL: https://kaggle.com/competitions/humpback-whale-identification.

[Kun+22] Ludmila I. Kuncheva et al. "A Benchmark Database for Animal Re-Identification and Tracking". In: *The Fifth IEEE International Image Processing, Applications and Systems Conference (IPAS'22)*. Piscataway, NJ: IEEE, 2022, pp. 1–6. ISBN: 978-1-6654-6219-8.

[Lah+11] Mayank Lahiri et al. "Biometric animal databases from field photographs: identification of individual zebra in the wild". In: *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*. ICMR '11. Trento, Italy: Association for Computing Machinery, 2011. ISBN: 9781450303361. URL: https://doi.org/10.1145/1991996.1992002.

[Li+18] Wei Li, Xiatian Zhu, and Shaogang Gong. "Harmonious Attention Network for Person Re-identification". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2018, pp. 2285–2294. ISBN: 978-1-5386-6420-9.

[Li+19] Shuyuan Li et al. *Amur Tiger Re-identification in the Wild*. June 13, 2019.

[Lil22a] Lilawp. *Beluga ID 2022 - LILA BC*. 2022. URL: https://lila.science/datasets/beluga-id-2022/.

[Lil22b] Lilawp. *Hyena ID 2022 - LILA BC*. 2022. URL: https://lila.science/datasets/hyena-id-2022/.

[Lil22c] Lilawp. *Leopard ID 2022 - LILA BC*. 2022. URL: https://lila.science/datasets/leopard-id-2022/.

[Liu+21] Ze Liu et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows". In: *2021 IEEE/CVF International Conference on Computer Vision*. Ed. by Eric Mortensen. Piscataway, NJ: IEEE, 2021, pp. 9992–10002. ISBN: 978-1-6654-2812-5.

[Nep+22] Ekaterina Nepovinnykh et al. "SealID: Saimaa Ringed Seal Re-Identification Dataset". In: *Sensors* 22.19 (2022). ISSN: 1424-8220.

[Par+17] Jason Parham et al. "Animal population censusing at scale with citizen science and photographic identification". In: AAAI Spring Symposium - Technical Report. United States: AI Access Foundation, 2017, pp. 37–44.

[Raj98] Lesley Raj. "Photo-identification of Stichopus mollis". In: *SPC Beche-de-mer Information Bulletin* 10 (1998), pp. 29–31.

[Rav+20] Prashanth C. Ravoor and Sudarshan T.S.B. "Deep Learning Methods for Multi-Species Animal Re-identification and Tracking – a Survey". In: *Computer Science Review* 38 (2020), p. 100289. ISSN: 15740137.

[Sch+20] Stefan Schneider, Graham W. Taylor, and Stefan C. Kremer. "Similarity Learning Networks for Animal Individual Re-Identification - Beyond the Capabilities of a Human Observer". In: *2020 IEEE Winter Applications of Computer Vision workshops (WACVW)*. Piscataway, NJ: IEEE, 2020, pp. 44–52. ISBN: 978-1-7281-7162-3.

[Sun+18] Yifan Sun et al. "Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline)". In: *15th European Conference on Computer Vision*. Ed. by Vittorio Ferrari et al. 2018, pp. 480–496.

[Til+16] Tilo Burghardt and Will Andrew. *FriesianCattle2015*. 2016.

[Tro+20] Cameron Trotter et al. *NDD20: A large-scale few-shot dolphin dataset for coarse and fine-grained categorisation*. 2020. URL: http://arxiv.org/pdf/2005.13359.pdf.

[Wan+21] Le Wang et al. "Giant Panda Identification". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 2837–2849. URL: https://github.com/iPandaDateset/iPanda-50.

[Wil+20] William Andrew et al. *OpenCows2020*. 2020.

[Yu+17] Qian Yu et al. "The Devil is in the Middle: Exploiting Mid-level Representations for Cross-Domain Instance Matching". In: *arXiv.org* (2017).

[Zho+19a] Kaiyang Zhou and Tao Xiang. "Torchreid: A Library for Deep Learning PersonRe-Identification in Pytorch". In: *arXiv preprint arXiv:1910.10093* (2019). URL: https://arxiv.org/pdf/1910.10093.pdf.

[Zho+19b] Kaiyang Zhou et al. "Omni-Scale Feature Learning for Person Re-Identification". In: *2019 International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2019, pp. 3701–3711. ISBN: 978-1-7281-4803-8.

[Zho+21] Kaiyang Zhou et al. "Learning Generalisable Omni-Scale Representations for Person Re-Identification". In: *IEEE transactions on pattern analysis and machine intelligence* (2021).

[Zin23] Zindi. *Turtle Recall: Conservation Challenge*. 12/13/2023. URL: https://zindi.africa/competitions/turtle-recall-conservation-challenge/data.