A Benchmark Database for Animal Re-Identification and Tracking

Ludmila I. Kuncheva
School of CSEE
Bangor University
Bangor, UK
l.kuncheva@bangor.ac.uk

Francis Williams
School of CSEE
Bangor University
Bangor, UK
eeub05@bangor.ac.uk

Samuel L. Hennessey
School of CSEE
Bangor University
Bangor, UK
sml18vly@bangor.ac.uk

Juan J. Rodríguez
Departamento de Ingeniería Informática
Universidad de Burgos
Burgos, Spain
jjrodriguez@ubu.es

***Abstract*—While there are multiple sources of annotated images and videos for human and vehicle re-identification, databases for individual animal recognition are still in demand. We present a database containing five annotated video clips each containing between 9 and 27 identities. The overall number of individual animals is 20,490, and the total number of classes is 93. The database can be used for testing novel methods for animal re-identification, object detection and tracking. The main challenge of the database is that multiple animals are present in the same video frame, leading to problems with occlusion and noisy, cluttered bounding boxes. To set-up a benchmark on individual animal recognition, we trained and tested 26 classification methods for the five videos and three feature representations. We also report results with state-of-the-art deep learning methods for object detection (MMDet) and tracking (Uni-Track).**

***Index Terms*—Animal re-identification, Benchmark database, Classification of images, Object detection and tracking.**

I. Introduction

In Multi-Object Tracking (MOT), multiple objects in a video are identified and tracked over time while keeping a record of their identities. There has been substantial interest in this topic over the years, mostly in the context of tracking vehicles and pedestrians on the road [1]. The Multi Object Tracking initiative https://motchallenge.net/ [2] streamlined the advances in the area by providing a suite of benchmark data sets as well as unified evaluating metrics.

In the current era of global concern about preserving the environment and the animal diversity, animal re-identification from images and video becomes a task of high priority [3]–[5]. Animal re-identification is the task of recognising an individual animal in different images. For example, an animal may have temporarily left the camera view and re-entered later. Re-identification means assigning this animal its correct identity. Conversely, in tracking, the algorithm will likely initiate a new track (new identity) for the reappearing object. Sometimes, re-identification (of people mostly) is meant to identify the same object across footage from different cameras running at the same time. Whichever way re-identification is defined, it is in essence a classification task, where a classifier is trained on annotated video footage or time-lapse image collections to distinguish between the existing identities. Developing animal re-identification algorithms and methodologies is perceived to be just the beginning of a major trend that could stand to revolutionise our approach to animal ecology. [4]

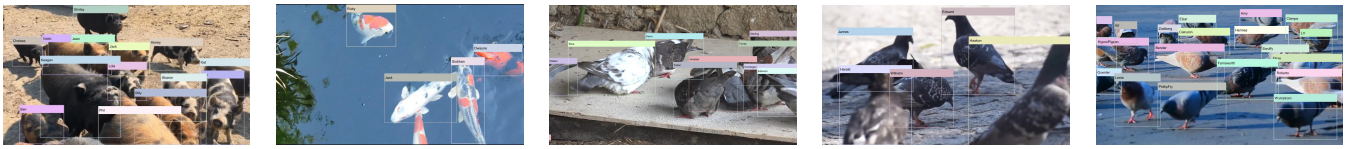
Machine Learning is expected to play a fundamental role in this quest [6]. Unfortunately, animal re-identification and tracking, especially in unrestricted environments, has received much less attention compared to people and vehicle tracking [7]. Attempts at automatic tracking of animals in video can be dated back to 1996 [8]. Much of the work has been dedicated to livestock tracking [9], [10]. While studies on animal tracking *in controlled environments* usually report high rate of success [11]–[13], it has been noted that long term tracking of individual animals (sometimes over days or weeks) is still an open challenge [12].

The current publication culture is that data and code are made publicly available. Unfortunately, there are no widely accepted standards, formats or protocols for creating such databases. This impedes wide cross-evaluation of animal re-identification methods. Table I shows a list of databases for animal re-identification, acknowledging at the same time that our list is by no means exhaustive.

In view of the shortage of unified benchmark resources, Tuia et al. [6] call for curating and publishing well-annotated benchmark datasets. Our study addresses this call by offering a collection of annotated short videos of animals. In addition to the data description, we report the results from a classification experiment (re-identification of the animals in the videos) that can serve as a baseline for future classification. We also provide results with state-of-the-art, off-the-shelf object detector and multi-object tracker.

The rest of the paper is organised as follows. Our proposed database is detailed in Section II. Section III contains the classification experiment, Section IV reports the results from the object detection and the tracking experiment, and Section V gives our conclusion.

This work is supported by the UKRI Centre for Doctoral Training in Artificial Intelligence, Machine Learning and Advanced Computing (AIMLAC), funded by grant EP/S023992/1. This work is also supported by the Junta de Castilla León under project BU055P20 (JCyL/FEDER, UE), the Ministry of Science and Innovation under project PID2020-119894GB-I00 co-financed through European Union FEDER funds, and the Ministry of Universities under mobility grant PRX21/00638.



(a) Pigs (b) Koi fish (c) Pigeons (curb) (d) Pigeons (ground) (e) Pigeons (square)

Fig. 1: Examples of annotated frames from the animal re-identification database. Links to the annotated videos are available at https://github.com/LucyKuncheva/Animal-Identification-from-Video.

TABLE I: Databases for animal re-identification

Publication	Animal	N	c	Comment
Livestock				
[14]	Holstein-Friesian cattle	7,043	46	https://data.bris.ac.uk/data/dataset/10m32xl88x2b61zlkkgz3fml17
Aquatic wildlife				
[15]	Great White Shark	2,456	85	*https://saveourseas.com/
[16]	Bottle-nose Dolphin	10,713	401	*https://sarasotadolphin.org/meet-dolphins
[16]	Humpback Whale	7,173	3,572	*https://www.cascadiaresearch.org/projects/photo-id
[4], [17]	Humpback Whale	9,850	4,251	https://www.kaggle.com/c/humpback-whale-identification
[18]	Killer Whale	86,789	367	*https://baycetology.org
Terrestrial wildlife				
[19]	Gorilla	5,428	7	*https://data.bris.ac.uk/data/dataset/4vnrca7qw1642qlwxjadp87h7
[20]	Chimpanzee	598	24	ChimpZoo † http://www.saisbeco.com/
[20]	Chimpanzee	1,432	71	ChimpTai † http://www.saisbeco.com/
[21]	Sociable weaver	17,500	50	https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13436
[22]	Zebra	N/A	85	https://www.researchgate.net/publication/221318569_Biometric_animal_databases_from_field_photographs_Identification_of_individual_zebra_in_the_wild
[23]	Amur tiger	3,649	92	https://cvwc2019.github.io/challenge.html
[24], [25]	Elephant	2,078	276	http://www.inf-cv.uni-jena.de/Research/Datasets/ELPephants.html
Lab animals				
[26], [27]	Fruit Fly	288,000	60	https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/JP4WDF
This study	Pigeons, Koi fish, Pigs	20,490	93	https://github.com/LucyKuncheva/Animal-Identification-from-Video

Table notes:

- N is the number of images (individual animals), c is the number of identities (classes).
- * – Available upon request
- † – Available upon purchasing a license

II. The Database

A. Overview

A snapshot of the database is shown in Figure 1. Short video clips were sourced from Pixabay https://pixabay.com/ under the Pixabay license. The videos capture the movement of groups of animals within 9-24 seconds. Majority of the animals are present throughout the video clip, some leaving and entering the camera view a several times. Each video was manually annotated with the animal identities. The annotations are presented in our database in a unified format.

The characteristics of the five videos are summarised in Table II. We have a total of 2379 frames, 20,490 clips, and 93 identities, which is in line with the databases we found elsewhere. We also display an imbalance metric for each video, which is calculated as the size of the largest class divided by the size of the smallest class. We note that we are not proposing a database that is 'better'; we are contributing *another* database that can enrich the research on animal identification and monitoring.

TABLE II: Characteristics of the videos

Video	#Frames	Length in s	#Objects	#Classes	Min p/f	Max p/f	Avr p/f	Imbalance
Pigs	500	16	6184	26	4	20	12.4	10.5
Koi fish	536	22	1635	9	1	6	3.1	2.8
Pigeons (curb)	443	17	4700	14	8	13	10.6	3.1
Pigeons (pavement)	600	24	3079	17	3	8	5.1	19.3
Pigeons (square)	300	9	4892	27	1	23	16.3	24.8

Table notes: k is the number of frames; l is the video length in seconds; N is the number of objects (individual animal clips); c is the number of classes (animal identities); Min p/f is the minimum number of animals per frame (image); Max p/f and Avr p/f are respectively the maximum and the average numbers.

Some distinguishing features of the database are listed below:

(1) Multiple animals in an image. All video clips contain multiple animals in a single frame as summarised in Table II. The distribution of the number of animals per frame, across all five videos, is shown in Figure 2.

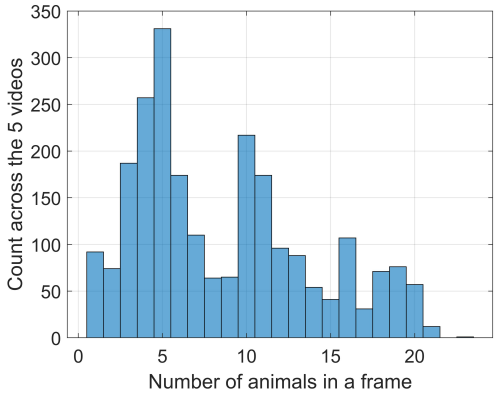


Fig. 2: Distribution of the number of animals per frame/image.

(2) Considerable difficulty.

- *Intra-class variability.* An animal may have very different appearances throughout the video which may make it practically indistinguishable from another animal.
- *Occlusion.* Quite often, the animals occlude one another, which results in one bounding box containing multiple animals of parts thereof. This introduces a large amount of intra-class noise in the data. Figure 3 demonstrates the occlusion problem where a bounding box for one animal contains a large portion of another. The images for an animal may vary dramatically in size, appearance and resolution (due to the different sizes of the bounding boxes when the animal is close to the camera or farther away).
- *String-shaped subclusters.* The images of an animal will be naturally clustered due to the contingency of the video frames. Thus, each class will possibly consist of several string-shaped sub-clusters.
- *Inter-class similarity and label noise.* Finally, the animals in each video are quite similar to one another, which makes labelling as well as re-identification difficult. It has been noted that, in such scenarios, the human annotator is predictably outperformed by the machine learning approaches [27]. Consequently, there may be some label noise in the database.

B. Organisation of the material

a) Availability.: The GitHub repository Animal-Identification-from-Video is available at https://github.com/LucyKuncheva/Animal-Identification-from-Video.

b) Videos.: The videos are accessible through web links. Each video needs to be separated into frames so that the annotations are matched to the respective frame. Examples of the annotated videos are also available through web links (Figure 1).

c) Annotations.: The core part of the database are the comma-delimited (CSV) files containing the annotations. There are three files for each video: an overall file with all annotations, a file with the training annotations and a file with the testing annotations (needed for the classification experiment presented in Section III). All CSV files are formatted as shown in Table III.

III. A CLASSIFICATION EXPERIMENT

Here we carry out a classification experiment with a view to set up a baseline for future comparisons.

A. Training and testing splits

With video data, splitting the set of objects randomly is not a sensible option because of the video frame contingency. If split randomly, some of the near-identical objects will fall in the training set and some will fall in the testing set. This will make the classification task deceptively easy. This is why we split all videos into two halves, based on the number of frames. All objects in the first part of the video comprise the training data and the objects in the second part, the testing data.

As the environments in the videos are not restricted to an enclosed space (such as fish tank or a lab cage), some animals move in and out of camera view. In all videos except the Koi fish video, the testing data contains classes (individual animals) that are not present in the training data. Classifying the new animals using a classifier which has not seen in these classes will only introduce random noise in the estimates of the testing accuracy. This is the reason for providing separate training and testing annotation files. The training and testing data can be constructed from the full CSV files. We have supplied them in the database for convenience.

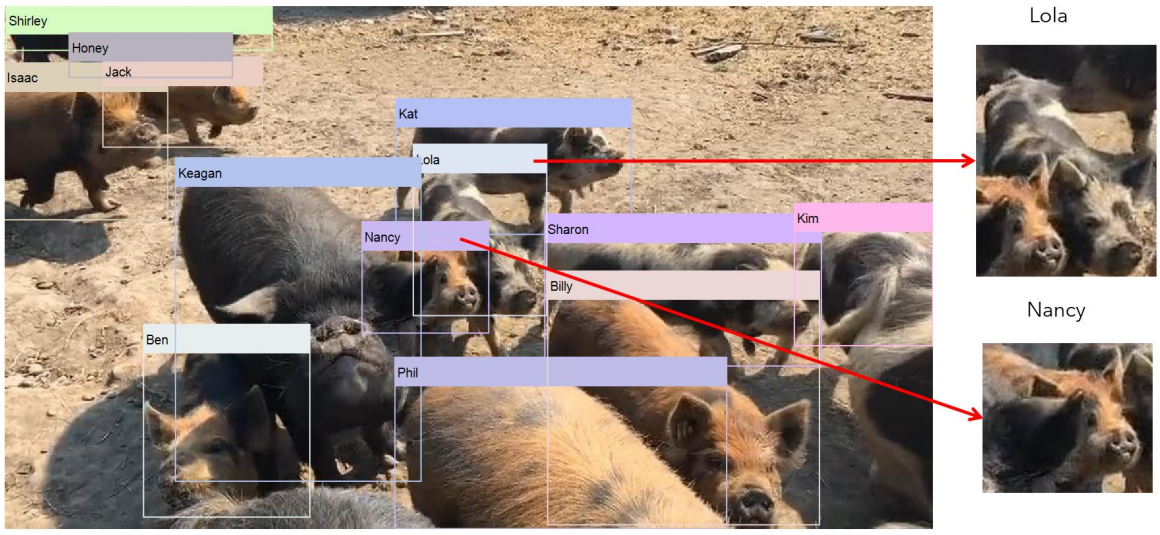


Fig. 3: Example of bounding boxes which include more than one animal. Both Lola and Nancy appear in the box enclosing Lola, while Nancy has a separate box for herself.

TABLE III: An example of the format and content of the CSV files with the video annotations.

ID	x	y	width	height	filename	$\max_x$	$\max_y$
Mahrez	1059	85	221	312	scene00001.jpg	1280	720
Torres	686	174	367	342	scene00001.jpg	1280	720
Sterling	564	132	283	145	scene00001.jpg	1280	720
...	...	...	...	...	...	...	...

Table notes: x and y are the pixel coordinates of the top left corner of the bounding box. The width and the height are given in pixels. The file name corresponds to the video frame with the respective number. $\max_x$ are $\max_y$ the respective image dimension in pixels.

B. Feature representations

We resized all individual images to size 56-by-56 and extracted the following feature representations:

- *Colour-related.* RGB moments: 54 features (3-by-3 blocks, RGB means and RGB standard deviations for each block).
- *Shape-related.* Histogram of Oriented Gradients (HOG) features: 441 features).
- *Texture-related.* Local Binary Patterns (LBP) features: 50 features.

C. Classification approaches

The 26 classification methods were sourced from the scikit-learn [28] Python library. They were applied through *Lazy Predict* [29], covering a variety of classification approaches, representatives of which are listed in the caption of Figure 4. Each classifier was trained and tested on each of the three feature representations (RGB, HOG, and LBP) for each video. The best representative of each classifier

D. Results

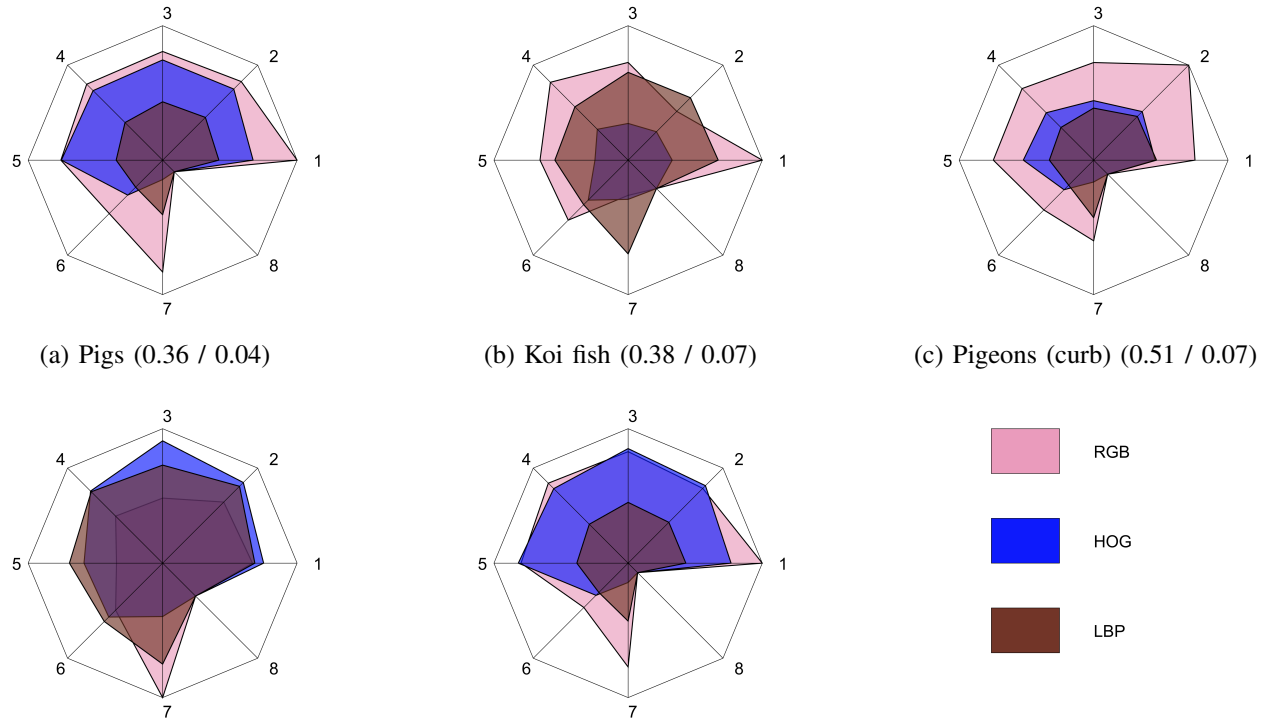
Figure 4 shows a summary of the results. For the sake of clarity and space only eight of the 26 classification methods are shown. The selected methods are the best ones and/or the most representative of each group (e.g., LDA and Logistic regression from the group of linear classifiers, and the Random Forest from the group of ensemble methods). The larger the area of the glyph, the better the feature set. Also, the longer the spoke for a given classification method, the higher the accuracy. Table IV shows the best combination of classifier and feature space for each video. While the glyph plots do not unequivocally favour one feature representation over another, RGB seems to be successful across all videos. This is also reflected in Table IV. The table demonstrates that simple classifiers like LDA or Logistic Regression fare better than more complex methods such as Random Forest or SVM. This result can be attributed to the complexity aspects of the datasets discussed in Section II.

TABLE IV: Best results from the classification experiment.

Video	Classifier	Features	Acc [%]
Pigs	Linear Discriminant Analysis	RGB	35.58
Koi fish	Linear Discriminant Analysis	RGB	37.69
Pigeons (curb)	Logistic Regresion	RGB	50.50
Pigeons (ground)	Quadratic Discriminant Analysis	RGB	21.26
Pigeons (square)	Linear Discriminant Analysis	RGB	53.10

IV. Object detection and tracking experiment

As we offer the database as a possible test-bed for object detection and tracking, we sourced state-of-the-art methods from



(a) Pigs (0.36 / 0.04) (b) Koi fish (0.38 / 0.07) (c) Pigeons (curb) (0.51 / 0.07)

(d) Pigeons (ground) (0.21 / 0.07) (e) Pigeons (square) (0.53 / 0.05) (f) Legend

Fig. 4: Classification accuracy for the three feature spaces, the 5 videos, and the 8 classifiers. The numbers on the glyph plots correspond to the classifier number: 1. LDA; 2. Logistic regression; 3. SVM; 4. Random forest; 5. k-nn; 6. Decision tree; 7. Quadratic; 8. Largest prior. The parentheses enclose (maximum accuracy for the video / and the accuracy of the Largest Prior classifier).

the widely used site Paper-with-Code https://paperswithcode.com/. For object detection, we used the MMDetector [30], and for tracking, we used Uni-Track [31]. Table V reports the standard metrics for these tasks: Average precision @IoU = 0.5 for object detection [32], and HOTA and MOTA for tracking [33], [34]. These values are given only for benchmark reference.

TABLE V: Object detection metrics (Average Precision, AP) and Multi-object tracking metrics (HOTA and MOTA), all in [%]

Video	AP	HOTA	MOTA
Pigs	78.12	38.09	59.17
Koi fish	50.56	27.34	49.48
Pigeons (curb)	55.42	26.27	44.30
Pigeons (ground)	81.05	38.05	66.94
Pigeons (square)	80.38	53.13	65.80

V. Conclusion

Here we present an annotated database for individual animal recognition. Five video clips were annotated with bounding boxes and animal identities, resulting in a total of 20,490 individual clips of 93 identities. The database can be used for classification experiments as demonstrated here. The task is quite challenging owing to the large amount of animals in a single image, which leads to cluttered and overlapping bounding boxes.

Baseline experiments were carried out with 26 classifier models. The results favoured simple feature spaces (RGB) and simple classifiers (linear and quadratic). We also report benchmark results from object detection and tracking using state-of-the-art methods based on deep learning.

Given the dynamic of the videos where animals disappear from camera view and re-appear later, it is interesting to develop classifier models which can abstain from labelling an image. Further on, since the database is extracted from video clips, self-learning models can be developed through tracking, object detection and clustering, requiring minimum user intervention to label lengthy videos.

References

[1] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T. K. Kim, "Multiple object tracking: A literature review," *Artificial Intelligence*, vol. 293, p. 103448, 2021. [Online]. Available: https://doi.org/10.1016/j.artint.2020.103448

[2] P. Dendorfer, A. Ošep, A. Milan, K. Schindler, D. Cremers, I. Reid, S. Roth, and L. Leal-Taixé, "MOTChallenge: A benchmark for single-camera multiple target tracking," *International Journal of Computer Vision*, vol. 129, pp. 845—881, 2021. [Online]. Available: https://doi.org/10.1007/s11263-020-01393-0

[3] R. Kays, M. C. Crofoot, W. Jetz, and M. Wikelski, "Terrestrial animal tracking as an eye on life and planet," *Science*, vol. 348, no. 6240, 2015. [Online]. Available: https://science.sciencemag.org/content/348/6240/aaa2478

[4] S. Schneider, G. W. Taylor, S. Linquist, and S. C. Kremer, "Past, present and future approaches using computer vision for animal re-identification from camera trap data," *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 461–470, jan 2019.

[5] ——, "Similarity learning networks for animal individual re-identification – beyond the capabilities of a human observer," *ArXiV*, 2019. [Online]. Available: http://arxiv.org/pdf/1902.09324v4:PDF

[6] D. Tuia, B. Kellenberger, S. Beery, B. R. Costelloe, S. Zuffi, B. Risse, A. Mathis, M. W. Mathis, F. van Langevelde, T. Burghardt, R. Kays, H. Klinck, M. Wikelski, I. D. Couzin, G. van Horn, M. C. Crofoot, C. V. Stewart, and T. Berger-Wolf, "Seeing biodiversity: perspectives in machine learning for wildlife conservation," 2022. [Online]. Available: https://arxiv.org/pdf/2110.12951.pdf

[7] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. Fisher, "Detecting, tracking and counting fish in low quality unconstrained underwater videos," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2, 2008, pp. 514–519. [Online]. Available: https://www.research.ed.ac.uk/en/publications/detecting-tracking-and-counting-fish-in-low-quality-unconstrained

[8] H. T. Jiang and J. W. Dailey, "A video database system for studying animal behavior," in *Proceedings of the Conference on Multimedia Storage and Archiving Systems*, ser. Proceedings of the Society of Photo-Optical Instrumental Engineers (SPIE), C. C. J. Kuo, Ed., vol. 2916, Boston, MA, USA, 1996, pp. 162–173.

[9] L. Zhang, H. Gray, X. Ye, L. Collins, and N. Allinson, "Automatic individual pig detection and tracking in pig farms," *SENSORS*, vol. 19, no. 5, 2019.

[10] O. Guzhva, H. Ardö, M. Nilsson, A. Herlin, and L. Tufvesson, "Now you see me: Convolutional neural network based tracker for dairy cows," *Frontiers Robotics AI*, vol. 5, no. SEP, pp. 1–9, 2018.

[11] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Managò, D. Scheggia, F. Papaleo, and V. Murino, "Automatic visual tracking and social behaviour analysis with multiple mice," *PLoS ONE*, vol. 8, no. 9, 2013.

[12] P. Marti-Puig, M. Serra-Serra, A. Campos-Candela, R. Reig-Bolano, A. Manjabacas, and M. Palmer, "Quantitatively scoring behavior from video-recorded, long-lasting fish trajectories," *Environmental Modelling & Software*, vol. 106, no. SI, pp. 68–76, 2018.

[13] F. Naiser, M. Šmíd, and J. Matas, "Tracking and re-identification system for multiple laboratory animals," in *Proceedings of the Visual Observation and Analysis of Vertebrate and Insect Behavior Workshop at ICPR'18*, 2018.

[14] J. Gao, T. Burghardt, W. Andrew, A. W. Dowsey, and N. W. Campbell, "Towards self-supervision for video identification of individual holstein-friesian cattle: The cows2021 dataset," 2021. [Online]. Available: https://arxiv.org/abs/2105.01938

[15] B. Hughes and T. Burghardt, "Automated visual fin identification of individual great white sharks," *International Journal of Computer Vision*, vol. 122, no. 3, pp. 542–557, 2017. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s11263-016-0961-y.pdf

[16] H. J. Weideman, Z. M. Jablons, J. Holmberg, K. Flynn, J. Calambokidis, R. B. Tyson, J. B. Allen, R. S. Wells, K. Hupman, K. Urian *et al.*, "Integral curvature representation and matching algorithms for identification of dolphins and whales," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2831–2839. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2017_workshops/papers/w41/Weideman_Integral_Curvature_Representation_ICCV_2017_paper.pdf

[17] "KAGGLE: Humpback whale identification," 2017–2019, data provided by https://happywhale.com/home. [Online]. Available: https://www.kaggle.com/c/humpback-whale-identification

[18] C. Bergler, A. Gebhard, J. R. Towers, L. Butyrev, G. J. Sutton, T. J. Shaw, A. Maier, and E. Nöth, "FIN-PRINT a fully-automated multi-stage deep-learning-based framework for the individual recognition of killer whales," *Scientific reports*, vol. 11, no. 1, pp. 1–16, 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-02506-6

[19] O. Brookes and T. Burghardt, "A dataset and application for facial recognition of individual gorillas in zoo environments," *arXiv preprint arXiv:2012.04689*, 2020. [Online]. Available: https://arxiv.org/abs/2012.04689

[20] A. Loos and A. Ernst, "An automated chimpanzee identification system using face detection and recognition," *EURASIP Journal on Image and Video Processing*, 2013, art. 49. [Online]. Available: https://jivp-eurasipjournals.springeropen.com/track/pdf/10.1186/1687-5281-2013-49.pdf

[21] A. C. Ferreira, L. R. Silva, F. Renna, H. B. Brandl, J. P. Renoult, D. R. Farine, R. Covas, and C. Doutrelant, "Deep learning-based methods for individual recognition in small birds," *Methods in Ecology and Evolution*, vol. 11, no. 9, pp. 1072–1085, 2020. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13436

[22] M. Lahiri, Chayant, R. Warungu, D. I. Rubenstein, and T. Y. Berger-Wolf, "Biometric animal databases from field photographs: Identification of individual zebra in the wild," in *Proceedings of the 1st International Conference on Multimedia Retrieval*, 2011. [Online]. Available: https://www.researchgate.net/publication/221318569_Biometric_animal_databases_from_field_photographs_Identification_of_individual_zebra_in_the_wild

[23] S. Li, J. Li, H. Tang, R. Qian, and W. Lin, "ATRW: A benchmark for Amur tiger re-identification in the wild," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020. [Online]. Available: https://arxiv.org/pdf/1906.05586.pdf

[24] M. Körschens and J. Denzler, "Elpephants: A fine-grained dataset for elephant re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0. [Online]. Available: https://openaccess.thecvf.com/content_ICCVW_2019/papers/CVWC/Korschens_ELPephants_A_Fine-Grained_Dataset_for_Elephant_Re-Identification_ICCVW_2019_paper.pdf

[25] M. Körschens, B. Barz, and J. Denzler, "Towards automatic identification of elephants in the wild," *CoRR*, vol. abs/1812.04418, 2018. [Online]. Available: http://arxiv.org/abs/1812.04418

[26] N. Murali, J. Schneider, J. Levine, and G. Taylor, "Classification and re-identification of fruit fly individuals across days with convolutional neural networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 570–578. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8658974

[27] J. Schneider, N. Murali, G. W. Taylor, and J. D. Levine, "Can drosophila melanogaster tell who's who?" *PloS one*, vol. 13, no. 10, p. e0205043, 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0205043

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[29] S. R. Pandala and B. B. da Silva, "Lazy predict, github repository," https://github.com/shankarpandala/lazypredict, 2021.

[30] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[31] Z. Wang, H. Zhao, Y.-L. Li, S. Wang, P. Torr, and L. Bertinetto, "Do different tracking tasks require different appearance models?" *Thirty-Fifth Conference on Neural Infromation Processing Systems*, 2021.

[32] R. Padilla, S. L. Netto, and E. A. Da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2020, pp. 237–242.

[33] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé, and B. Leibe, "HOTA: A higher order metric for evaluating multi-object tracking," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 548–578, 2021.

[34] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.