electronics

MDPI

Article

MammalClub: An Annotated Wild Mammal Dataset for Species Recognition, Individual Identification, and Behavior Recognition

Wenbo Lu, Yaqin Zhao *, Jin Wang, Zhaoxiang Zheng, Liqi Feng and Jiaxi Tang

College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China; luwenbo@njfu.edu.cn (W.L.); 200360316wj@njfu.edu.cn (J.W.); zhengzhaoxiang@njfu.edu.cn (Z.Z.); dream6182@163.com (L.F.); tangjiaxi@njfu.edu.cn (J.T.)
* Correspondence: zhaoyaqin@njfu.edu.cn

Abstract: Mammals play an important role in conserving species diversity and maintaining ecological balance, so research on mammal species composition, individual identification, and behavioral analysis is of great significance for optimizing the ecological environment. Due to their great capabilities for feature extraction, deep learning networks have gradually been applied to wildlife monitoring. However, training a network requires a large number of animal image samples. Although a few wildlife datasets contain many mammals, most mammal images in these datasets are not annotated. In particular, selecting mammalian images from vast and comprehensive datasets is still a time-consuming task. Therefore, there is currently a lack of specialized datasets of images of wild mammals. To address these limitations, this article created a mammal image dataset (named MammalClub), which contains three sub-datasets (i.e., a species recognition sub-dataset, an individual identification sub-dataset, and a behavior recognition sub-dataset). This study labeled the bounding boxes of the images used for species recognition and the coordinates of the mammals' skeletal joints for behavior recognition. This study also captured images of each individual from different points of view for individual mammal identification. This study explored novel intelligent animal recognition models and compared and analyzed them with the mainstream models in order to test the dataset.

Keywords: mammal dataset; species recognition; individual identification; behavior recognition

check for updates

Citation: Lu, W.; Zhao, Y.; Wang, J.; Zheng, Z.; Feng, L.; Tang, J. MammalClub: An Annotated Wild Mammal Dataset for Species Recognition, Individual Identification, and Behavior Recognition. *Electronics* **2023**, *12*, 4506. https://doi.org/10.3390/electronics12214506

Academic Editor: KC Santosh

Received: 30 September 2023
Revised: 27 October 2023
Accepted: 1 November 2023
Published: 2 November 2023

1. Introduction

As the construction of ecological civilization and the protection of biodiversity are being paid greater attention, the importance of wildlife diversity monitoring has become increasingly prominent. Scientific management and continuous monitoring are prerequisites for wildlife research, protection, and management. However, as they have been affected by various factors, such as climate change, habitat fragmentation, and loss of habitats, the numbers of many species of wild animals have sharply decreased, and the distribution of wildlife populations has significantly shrunk [1]. Thus, it is urgently necessary to rapidly track and assess population dynamics by using wildlife diversity monitoring. In particular, the species identification, individual identification, and behavioral analysis of animals are not only the premise of research on wildlife diversity monitoring [2,3], but they also help animal management staff further understand the health and requirements of animals, thus contributing to the improvement of animal welfare [4]. Mammals play an important role in conserving species diversity and maintaining ecological balance. Medium and large mammals, such as tigers and lions, which are at the top of the food chain, are especially endangered due to habitat fragmentation. Therefore, it is particularly important to monitor and study the diversity of mammals.

With the widespread application of infrared cameras in wildlife monitoring, intelligent recognition methods for wildlife based on machine vision technology provide scientific data for wildlife protection and management. However, training the intelligent recognition models requires a large number of annotated samples. In particular, due to the huge workload involved in manually labeling skeletal data for the analysis of animal behavior, data on animal skeleton labeling are particularly lacking, which largely restricts in-depth and extensive research on intelligent mammalian recognition. Therefore, this article created a publicly available, more comprehensive dataset for the intelligent recognition of mammals (named MammalClub) and developed intelligent recognition models for the analysis and testing of the dataset. The main contributions of this study are as follows.

- This study constructed a publicly available, relatively comprehensive dataset of medium to large terrestrial mammals for the development of stable deep learning models. The dataset contains three sub-datasets: those for species identification, individual identification, and behavioral analysis. We acquired a large number of images by filming and downloading documentaries.
- Our species dataset is labeled for each image with the location of the mammalian target, i.e., bounding boxes, without delineating the species, as in existing species datasets. This annotation allows the dataset to be used directly for animal target detection, which is a prerequisite for individual identification and behavioral analysis.
- For the application of the animal behavior recognition model, this study spent much time collecting and organizing videos of mammalian behavior and used the DeepLabCut software (v2.1.9) to extract the animal skeleton in each frame and label the information of the point coordinates of the animal's joints, which can be easily used as input for the daily animal behavior recognition model.
- Since wild animals are usually shy and their behavior is not under human control, it is difficult to obtain individual animal datasets. We took images from many different sides of animals' bodies and collected internet documentary videos to construct the individual identification datasets for 15 species.

In this article, we introduced a challenging dataset, and three sub-datasets are included in this dataset for the species recognition, individual identification, and behavior recognition of mammals. To test the practicality of the dataset, we explored novel intelligent recognition models and compared them with mainstream models. Our dataset can provide important data support for the training and testing of wild mammal species recognition models, individual identification models, and behavioral recognition models with camera-trap images or surveillance videos.

2. Related Works

With the rapid development of machine vision in various fields [5–7], it has also been used in wildlife diversity monitoring to provide important data support for the research, conservation, and management of wildlife [8–11]. The images captured in various wildlife reserves are accumulating in millions of units, and these images contain a large amount of key information about wild species and populations [12], individual activities [13], illness and injury [14], etc. Currently, these data mainly rely on artificial visual screening, which is far behind the speed of image accumulation and seriously restricts the effective application of data in research, protection, and management work. Therefore, many researchers have explored methods for the intelligent recognition of wild animals by using image-processing technologies [15–19]. Especially in the last decade, deep learning has rapidly evolved and has been widely used in various fields, such as farm produce detection [20,21], equipment fault diagnosis [22,23], animal monitoring [24,25], etc. In wildlife monitoring, most research based on deep learning has focused on species recognition [26–28]. A few researchers have reported their research findings on animal individual identification [29–33] and behavior recognition [34–37].

In recent years, many studies have used deep learning models for the intelligent recognition of wildlife images captured by camera traps [33,38,39]. Compared to the use of computers with high-performance hardware configurations, the construction and labeling of datasets are more important for model training, and they are also the prerequisite and foundation for deep-learning-based animal recognition. Therefore, the training of intelligent recognition models based on deep learning requires a large number of animal image samples [40–42]. Previous studies [43,44] on animal datasets have played important roles in the development of robust deep learning models (e.g., iNaturalist [45], Animals with Attributes [46], Caltech UCSD Birds [31], and Florida Wildlife Animal Trap [47]). The creators of Snapshot Serengeti [48] collected standardized data on many reserves in Africa, including about 2.65 million camera trap image sequences, for a total of 7.1 million images. At the species level, 61 categories were labeled, but about 76% of the images were marked as empty.

Although researchers have constructed some wildlife datasets, there are few datasets for the intelligent identification of mammals. Moreover, the existing datasets have limitations in multiple aspects: (1) Most animal datasets delineate the species of animals, and they also contain some mammal images. For example, the PASCAL VOC dataset only contains three kinds of livestock (cows, horses, and sheep), and the COCO dataset includes a limited number of mammals (elephants, bears, zebras, giraffes, etc.). Although the Snapshot Serengeti dataset contains a large number of mammalian images, it is not annotated with animal targets. (2) In practice, animal research and conservation are more concerned with the daily behavior of animals, such as feeding, resting, and hunting. There are few datasets dedicated to the identification of daily animal behavior. The creators of only a few datasets have collected and labeled information on mammalian posture, such as the joint coordinates of an animal's skeleton. However, the intelligent division benchmark for these datasets is pose estimation, so it can only be used in recognition models for the simple poses (e.g., standing or running) of animals. For example, AP-10K [49] is the first large-scale benchmark for mammalian pose estimation; it includes 10,015 images that were collected and filtered from 23 animal families and 54 species. Animal Kingdom [50] is a large and diverse dataset for understanding animal behavior, and it contains 850 animal species in a variety of scenarios, with the animal skeletons being annotated for some of the videos. The VOC2011 dataset [43] contains a subset for animal posture estimation, though it only contains five mammals. The action recognition dataset in [51] focused on seven action categories. (3) Individual recognition datasets are scarce, and only a few are currently available for specific animals. For example, ATRW [52] was first proposed to study the individual identification of Amur tigers and monitor their number and distribution in the wild; it includes over 8000 video clips from 92 Amur tigers with bounding boxes, pose keypoints, and tiger identity annotations. Tri-AI [53] contains individual animals' faces for six species. Pictures of meerkats, lions, and red pandas were captured during the daytime, whereas those of golden monkeys and tigers were captured at night.

3. Data Description

A large-scale dataset of wild mammals called MammalClub was constructed. The dataset was mainly sourced from wildlife documentaries (such as https://www.05jl.com/ (accessed on 6 March 2022), https://tv.cctv.com/lm/ryzr/ (accessed on 11 August 2022), https://www.bilibili.com/bangumi/play/ep661675 (accessed on 14 March 2023), etc.). This study also recorded some mammal images in zoos. The dataset contains 24 mammal species and has a total of 118,595 images. Many challenging images were taken from different natural environments, such as rock cliffs, late autumn leaves, grassland forests, and the Gobi desert, as well as during the late night and early morning, in rain and snow. The dataset includes three sub-datasets, i.e., a species recognition sub-dataset (SRSD), an individual identification sub-dataset (IISD), and a behavior recognition sub-dataset (BRSD). The three sub-datasets are detailed in Table 1. As shown in Table 1, the SRSD contains 12 mammal species. The IISD includes 11 mammal species, and the individual numbers of each species range from 15 to 31. For each ID, the study selected images from different sides of the individual mammals. The BRSD contains five kinds of behaviors (i.e., chasing, walking, eating, watching, and resting), adding up to 18 species. It is composed of 2058 labeled video clips, of which 435 clips are for animal Resting, 283 clips are for animal Chasing, 373 clips are for animal Eating, 560 clips are for animal Walking, and 407 clips are for animal Watching. In fact, although different species of animals have fur with different colors, for mammals, the skeletal joint movements involved in the same behaviors, such as chasing and resting, are similar.

Table 1. Composition of the MammalClub dataset.

SRSD		**IISD**			**BRSD**	
Class Name	**Number of Images**	**Species**	**Ids**	**Number of Images**	**Behavior Category**	**Number of Images**
Brown Bear	415	Brown Bear	20	5493	Chasing	10,783
Cow	351	Elephant	18	2371	Eating	11,577
Leopard	471	Giraffe	20	2222	Resting	7210
Deer	404	Horse	19	5295	Walking	34,415
Elephant	404	Kangaroo	20	4323	Watching	16,114
Giraffe	438	Lion	20	4078	**Species in BRSD**	
Horse	400	Giant Panda	15	2891	Antelope, Guanaco, Lion, Zebra, Hyena, Wild Boar, Leopard, Elephant, Tiger, Fox, Brown Bear, Polar Bear, Gnu, Wolf, Monkey, Deer, Cow	
Kangaroo	367	Polar bears	20	2088		
Koala	438	Red Panda	16	614		
Lion	397	Tiger	31	1093		
Tiger	399	Zebra	19	3144		
Zebra	400					

When deep learning algorithms are used for species recognition and behavior recognition in wild mammals, it is necessary to annotate the images in the datasets. Therefore, the study used the Labelme 4.5.13 software to label the location coordinates of the bounding boxes for the mammal objects in the images. This study also obtained the coordinates of 18 key joints of each mammal by using the DeepLabCut pose estimation software. The joint coordinates were used as input for behavior recognition models. Of course, the mammal individual identification sub-dataset did not need to be annotated. The study detailed the image annotations for species identification and behavior identification as follows.

3.1. Species Recognition Sub-Dataset (SRSD)

3.1.1. Composition of the Species Recognition Sub-Dataset

The species recognition sub-dataset consists of 12 common wild mammal species, with a total of 4884 images. Table 1 shows the details of the sub-dataset. There are about 400 images in each species, so the dataset is balanced, which is beneficial when training species recognition models. All of the images in the dataset are two-dimensional RGB color images with a single resolution of 1280 × 720. As shown in Figure 1, the sub-dataset contains not only whole-body images of wild mammals but also images of their body parts, such as the limbs of leopards or the back of a zebra. In addition, some images with camouflaged animals whose coat colors are similar to the background of the natural landscape are included. These challenging images increase the difficulty of species recognition.

3.1.2. Bounding Box Annotation

High similarity between consecutive frames leads to data redundancy; thus, the study used the histogram method to calculate the similarity between images and eliminated similar images in the same video clip [54]. The gray-scale histogram counts the number of pixels at a certain grayscale level, as seen in Figure 2. Then, this study labeled the locations of animal targets in the images and calculated the coordinates of the bounding boxes by using the labeling software Labelme. The labeled results were saved as a file in the .json format. The bounding boxes were visualized with the labeling software Labelme. The study divided the annotated dataset into two parts, the training set and the testing set, at a ratio of 0.85.

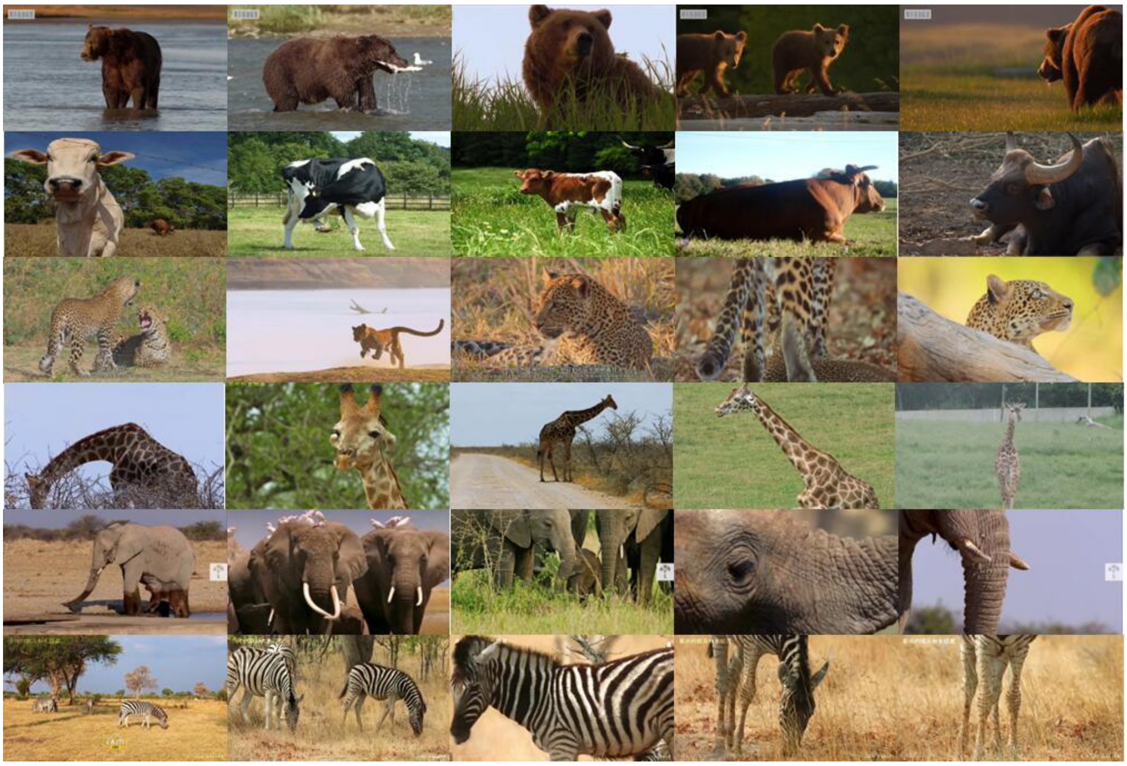


Figure 1. Examples of the images in the SRSD.

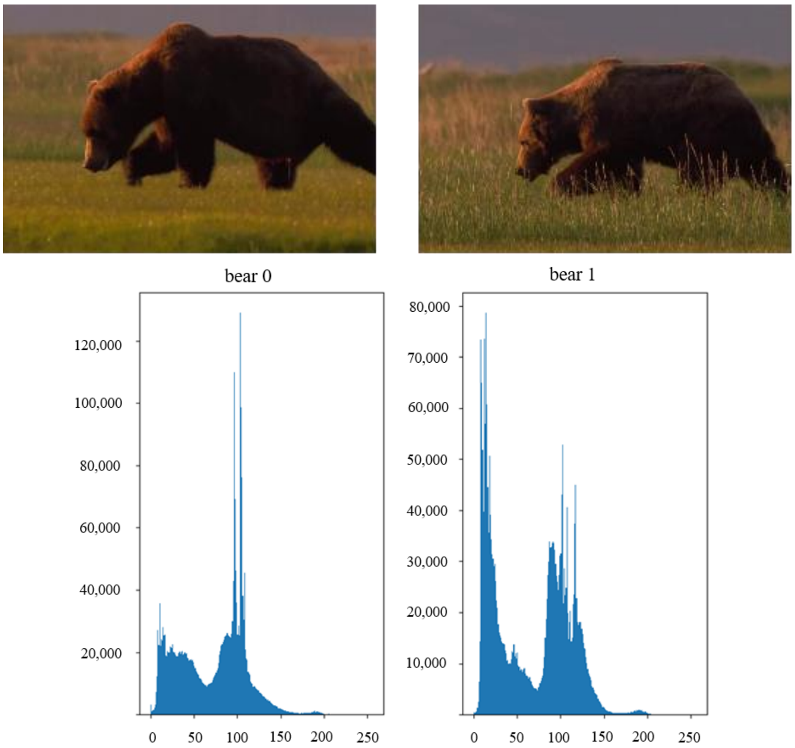


Figure 2. Two consecutive frames and their histograms.

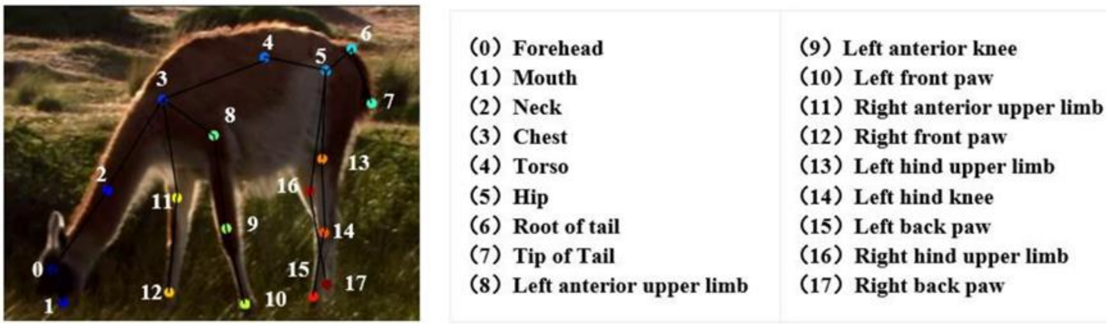
3.2. Behavior Recognition Sub-Dataset (BRSD)

3.2.1. Composition of the Behavior Recognition Sub-Dataset

There are five categories of behaviors in the behavior recognition sub-dataset, namely, chasing, eating, resting, walking, and watching. The same behaviors of different species of mammals have similar skeletal motion characteristics. Therefore, each category of behaviors in the BRSD contained different mammal species, and the same species were also classified in different behavioral categories. The sub-dataset is detailed in Table 1.

3.2.2. Joint Annotation

Though different species of mammals have different coats, the characteristics of their skeletal movement when performing the same behavior, such as running, are similar. Therefore, an animal posture algorithm based on DeepLabCut was adopted in order to build a skeleton-based mammal behavior recognition dataset. This study only calibrated the two-dimensional coordinates of the animal joints on the unobstructed side of their body parts. The labeled joints are shown in Figure 3. The input of the behavior recognition model is the coordinates of the marked joint points. When the animal's body is obstructed, the joint points cannot be detected correctly, so the animal in the dataset for behavior recognition is not obstructed. Due to the different sizes and positions of the animal skeletons that were labeled with DeepLabCut, the study conducted a normalization operation on the original skeletons. First, it took the center of the animal's body (i.e., the trunk) as the coordinate origin and re-constructed the coordinate system in order to map the original skeleton to a new coordinate system. Then, it normalized the new coordinates of the joints to within the range [−1, 1], as shown in Figure 4. The labels were used to generate the coordinates and confidence levels for each joint; therefore, during the labeling process, this study removed all of the joint coordinates of the images, including the labeled joints, with low confidence. Although the animal joints in some images had high confidence values, sometimes, they were not correctly labeled. So, all of the joint coordinates in the images needed to be removed. In fact, in a given video of a behavior, removing very few images did not have much impact on mammalian behavior recognition due to the redundant frames. Therefore, in addition to the original videos, the study also released an Excel file that only contained the joint coordinates of the reserved images, which could be directly used for the recognition of the behavior of wild mammals. The calibration results for some example images are shown in Figure 5.



(0) Forehead	(9) Left anterior knee
(1) Mouth	(10) Left front paw
(2) Neck	(11) Right anterior upper limb
(3) Chest	(12) Right front paw
(4) Torso	(13) Left hind upper limb
(5) Hip	(14) Left hind knee
(6) Root of tail	(15) Left back paw
(7) Tip of Tail	(16) Right hind upper limb
(8) Left anterior upper limb	(17) Right back paw

Figure 3. The corresponding descriptions of the skeletal joints.

3.3. Wild Mammal Individual Identification Sub-Dataset (IISD)

Individual recognition methods identify different individuals of a certain species based on surface features of the animal body, which is an important way to obtain the quantity of animal population. The wild mammal individual identification sub-dataset contains 11 common mammal species, and each species is represented by 15–21 individuals, as shown in Table 1. The number of the images belonging to each ID ranges from 40 to 280. The images were taken from the front, left, right, and rear of the individual animals in order to better evaluate the robustness of the individual identification models. Due to the small differences in body appearance characteristics such as fur color and markings among different individuals, the animals in the individual recognition dataset are not obscured. Figure 6 shows some example images in the individual identification dataset.

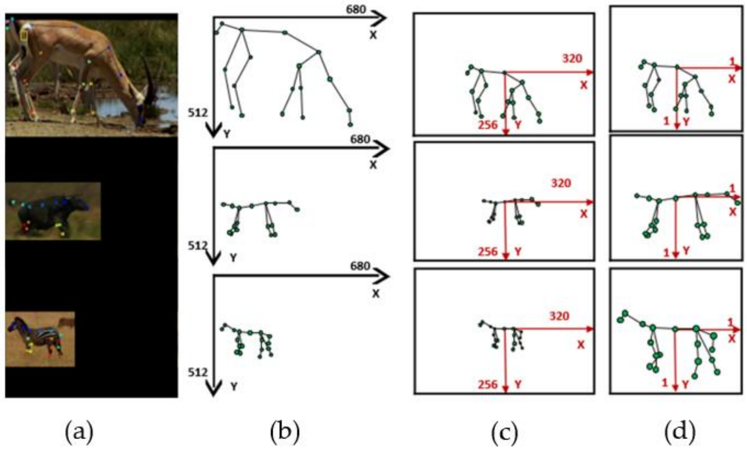


Figure 4. The normalization of the skeletal joints. (**a**) DeepLabCut input. (**b**) DeepLabCut output. (**c**) Centralization. (**d**) Normalization.

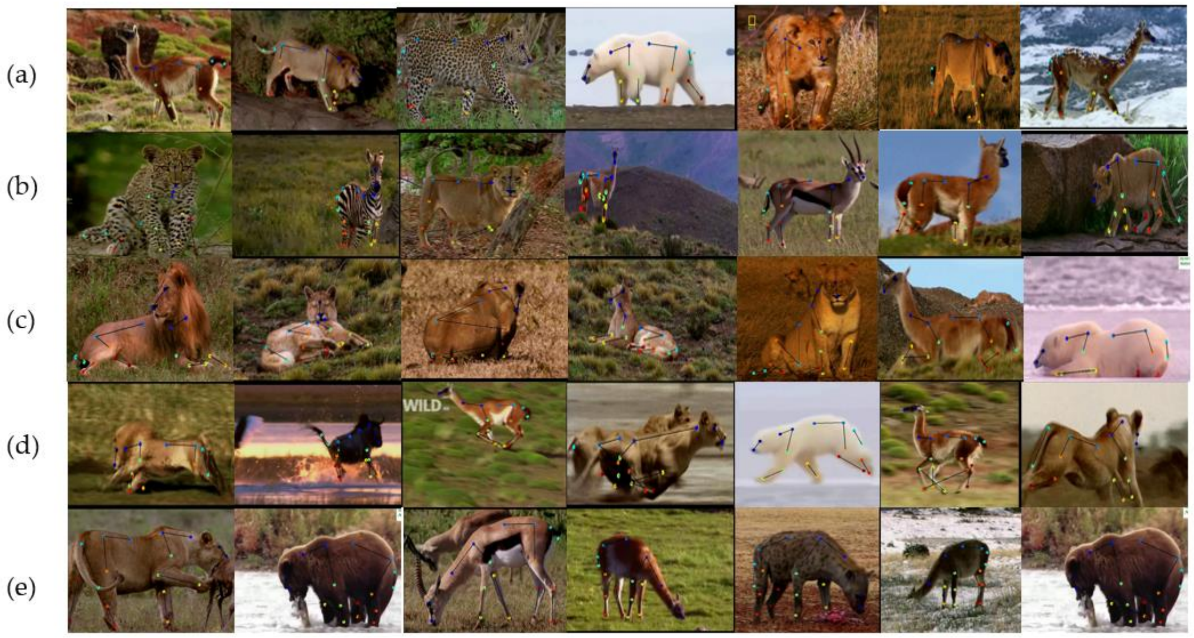


Figure 5. The calibration results for some example images. The walking behavior of mammals is shown as (**a**). The watching behavior of mammals is shown as (**b**). The resting behavior of mammals is shown as (**c**). The chasing behavior of mammals is shown as (**d**). The eating behavior of mammals is shown as (**e**).

3.4. Comparison of Our Dataset with Notable Animal Datasets

This study has summarized some of the notable animal datasets containing mammals in Table 2. Compared with the existing datasets, the main advantage of our dataset is that it includes the species recognition, behavior recognition, and individual identification of mammals, while other datasets only include one of these. Moreover, these existing datasets have a great number of images, and mammals occupy only a small portion of them; thus, it is necessary to spend time and energy collecting the mammal images. In particular, the images in each sub-dataset were also annotated, which provided great convenience for the training of intelligent mammal recognition models based on deep learning.

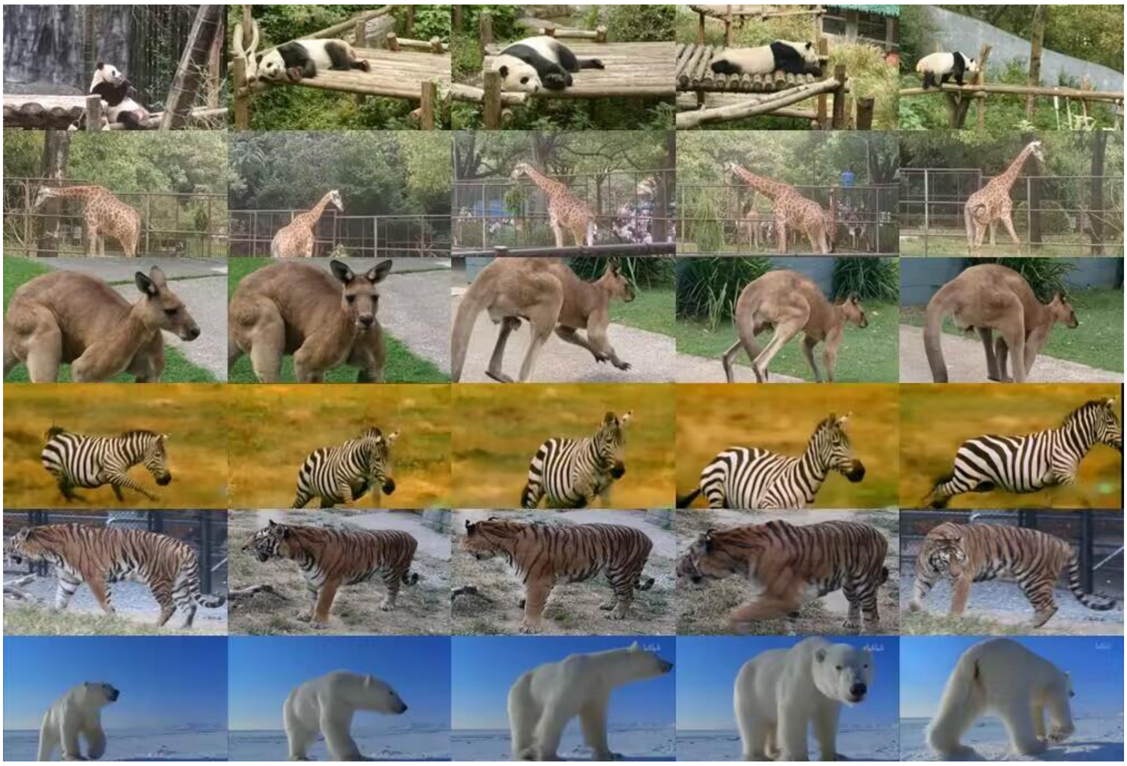


Figure 6. Presentation of some example images in the IISD.

Table 2. Comparison of our dataset with notable animal datasets.

Dataset	**Publicly Available**	**Species Recognition**	**Action Recognition**		**Individual Identification**	
		No. of Mammals	**No. of Annotated**	**No. of Annotated Action Classes**	**No. of IDs**	**No. of Images or Clips**
PASCAL VOC [55]	√	5	x	x	×	×
COCO [56]	√	9	×	×	×	×
Pig Tail-biting [57]	×	1	4396	2	×	×
Wildlife Action [51]	×	11	4000	5	×	×
Dogs [58]	√	1	13	4	×	×
Animal Kingdom [50]	√	207	NA	NA	×	×
ATRW [52]	√	1	×	×	92	8076
C-ZOO [59]	√	1	×	×	24	2109
C-Tai [59]	√	1	×	×	78	5078
MammalClub (Ours)	√	**24**	**2256**	**5**	**218**	**33,612**

4. Experiments

We proposed novel methods based on deep learning for species recognition, individual recognition, and action recognition, respectively [54,60,61]. We conducted experiments on our dataset with the proposed methods and other state-of-the-art methods for a comparison under the same experimental configuration.

4.1. Experimental Results for Species Recognition

In practice, training the intelligent mammal recognition models based on convolutional neural networks requires a large number of samples. However, only a few wild mammal images taken by camera traps are valid and useful because such cameras are triggered irregularly. Similarly, it is quite difficult to capture close-up images of wild mammals in artificial ways because their activities are not controlled by humans. Thus, in practical applications, mainstream classification models are not suitable for classifying wildlife species, especially for endangered wild mammals.

In [54], we adopted ResNet and RPN for feature extraction and region proposals, respectively. We allocated the training and evaluation set at a ratio of 0.85, and the details are shown in [54]. In the training stage, we introduced the few-shot training strategy in a convolutional neural network for FSOD (few-shot object detection) in order to recognize new mammal species that had a small number of samples in the training set. The network training was divided into two stages. In the first stage, we trained the learning abilities of the network with abundant mammal images, and the trained parameters and weights were transplanted directly into novel classes. In the second stage, we fixed them and fine-tuned the weights of the box predictors.

We tested the performance of the proposed method based on the AP, AP50, and AP75 metrics and compared it with that of some state-of-the-art methods. As seen in Table 3, the proposed method had higher values of AP, AP50, and AP75. Our SRSD included multiple small targets, occluded or overlapped animal bodies, camouflaged backgrounds, and incomplete bodies, which increased the difficulty of species identification. Our proposed species recognition method showed superiority in the face of small sample sizes with challenging images.

Table 3. Comparison of the results for the AP, AP50, and AP75 metrics on the SRSD.

Methods	Backbone	AP	AP50	AP75
Sparse R-CNN [62]	ResNet-50	58.7	78.3	64.6
YOLOv8 [63]	CSPDarkNet-53	56.8	79.6	64.7
FCOS [64]	ResNet-50	34.9	56.1	37.1
CenterNet [65]	ResNet-101	40.1	63.4	42.4
Ours	ResNet-101	57.6	85.1	65.0

4.2. Experimental Results for Action Recognition

Not only did we construct the first skeleton-based sub-dataset, but we also proposed a GCN (graph convolutional network)-based mammal behavior recognition method [60]. We randomly divided the animal behavior dataset into 10 subsets and selected nine subsets as the training set and the remaining one as the validation set. First, we divided the animal's body into five parts, namely, the trunk, left forelimb, right forelimb, left hind limb, and right hind limb, so as to convert joint-based graphs into part-based graphs, as shown in Figure 7. On this basis, we built part-wise attention mechanisms that included the part attention bottleneck, part share attention bottleneck, and part convolutional attention bottleneck. The three part-based attention mechanisms were similar, except that their pooling layers were different. After that, we constructed a graph-based search space that contained multiple graph function operations, such as joint-based graph operations and part-based graph operations. Then, we used a neural architecture search (NAS) to search for the best spatial–temporal graph convolutional network. We conducted the experiments on our mammal recognition sub-datasets.

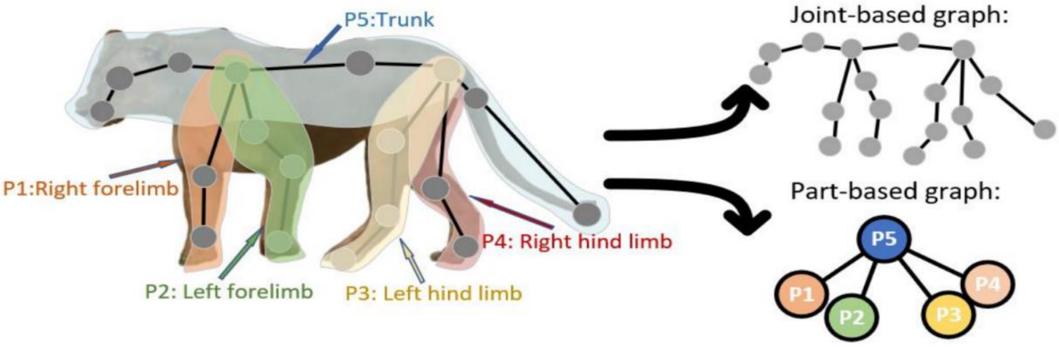


Figure 7. An illustration of five manually designed animal body parts.

The results are shown in Table 4. Three modes of skeletal data, namely, those of the joints, bones, and motion, were used as multi-modal inputs. We evaluated the proposed method with state-of-the-art methods and proved that the proposed method could achieve SOTA performance with fewer parameters.

Table 4. Comparison of multi-modal input algorithms.

Architecture	Stream	Input Modality	Parameter	Accuracy
2s-AGCN	One-stream	Joint	3.45 M	83.91%
	Two-stream	Joint and Bone	6.90 M	84.27%
Shift-GCN	One-stream	Joint	0.69 M	75.16%
	Two-stream	Joint and Bone	1.38 M	77.21%
MS-G3D	One-stream	Joint	2.99 M	84.25%
	Two-stream	Joint and Bone	5.98 M	85.01%
ResGCN	One-stream	Joint	0.73 M	82.97%
	Two-stream	Joint and Motion	0.78 M	83.11%
	Three-stream	Joint, Bone, and Motion	0.89 M	83.42%
Animal-Nas_s1 (ours)	One-stream	Joint	0.56 M	85.65%
		Bone	0.56 M	85.15%
		Motion	0.56 M	85.95%
	Two-stream	Joint and Bone	0.71 M	86.05%
		Joint and Motion	0.71 M	86.30%
		Joint and Motion	0.71 M	86.25%
	Three-stream	Joint, Bone, and Motion	0.82 M	86.98%

4.3. Experimental Results for Individual Identification

Due to their similar fur and the complexity of field environments, it is a challenging task to accurately identify the individual animals of a mammalian species. In order to better extract and fuse global and local information on wild mammals, we first applied the transformer network structure to the individual identification of wildlife [61]. In [61], we introduced the locally aware transformer (LA transformer) and replaced its self-attention module with a cross-attention mechanism block (CAB). The CAB was used to capture the differences in the local features and global information of an animal's appearance by using inner-patch self-attention and cross-patch self-attention. Then, we utilized the locally aware network of the LA transformer to fuse the local features and global features; next, the hierarchical structure of the locally aware network was redesigned according to the distribution of animal body parts in the standing posture.

We evaluated the proposed method and compared its performance with that of two methods, Top-DB-Net [66] and CAL + ResNet [67], on our individual identification sub-dataset. We allocated the training and evaluation set at a ratio of 0.7, and the details are shown in [61]. To reduce the impact of background on recognition accuracy, object detection was applied to intercept animal regions before individual recognition [66,67]. The results are shown in Table 5. Through the experiments on our datasets, we proved that our dataset could be applied well to various individual identification tasks with wild mammals. Specifically, after an image was divided into small blocks, some of them contained background pixels, so the background of the image had a certain impact on individual recognition. As seen in Figure 6 and Table 5, the different sites of a panda's activity resulted in different backgrounds, so the methods used for comparison, i.e., Top-DB-Net and CAL + ResNet, had lower accuracy in identifying the individual pandas. Compared with the two other methods, our individual identification method, the CATLA transformer, improved upon the transformer by introducing a cross-attention mechanism to capture the local features and the global information of the animal appearance; thus, the proposed method was superior to the existing state-of-the-art methods.

Table 5. Experimental results with the different methods.

Methods / Species	CAL + ResNet		Top-DB-Net		Our Method: CATLA Transformer	
	Rank 1	**mAP**	**Rank 1**	**mAP**	**Rank 1**	**mAP**
Elephant	100	87.2	100	93.1	100	**100**
Bear	100	89.1	100	94.9	100	**100**
Giraffe	100	89.5	100	91.7	100	**100**
Horse	100	97.5	100	98.9	100	**99.97**
Kangaroo	100	89.8	100	98.8	100	**100**
Lion	100	100	100	100	100	**100**
Panda	98.9	81.4	100	89.2	100	**100**
Polar bears	100	99.6	100	93.8	100	**100**
Red Panda	100	82.7	100	96.5	100	**99.09**
Tiger	100	93.2	99.3	95.6	99.33	**99.25**
Zebra	100	99.7	100	99.8	100	**99.95**

5. Conclusions

We first introduced a challenging dataset for species recognition, individual identification, and behavior recognition for mammals. Three sub-datasets are included in this dataset. In particular, in addition to annotating the dataset, we created labeling documents for mammalian species recognition and behavior recognition that can be directly applied to train models. We explored novel intelligent animal recognition models and compared and analyzed them with mainstream models in order to test the dataset. Our dataset provides important data support for the training and testing of wild mammal species recognition models, individual identification models, and behavioral recognition models with camera-trap images or surveillance videos.

In future work, we will further expand our datasets in terms of the number of species and the number of individuals in each species. We will also increase the number of types of daily behaviors of the mammals to expand the behavior recognition dataset. The application of simulated data (from computer game environments) is becoming much more popular [68,69]. We will try to utilize the photo-realistic graphical engine of the video game to generate wild animal images based on a specific ecological environment and then construct a virtual animal dataset, which will be used to train and test the generalization ability of the model.

Author Contributions: Conceptualization, W.L. and Y.Z.; methodology, Y.Z., L.F., Z.Z. and J.T.; software, L.F., Z.Z. and J.T.; validation, W.L., J.W. and Z.Z.; investigation, W.L. and J.W.; resources, W.L. and J.W.; data curation, Y.Z.; writing—original draft preparation, W.L.; writing—review and editing, Y.Z. and J.W.; visualization, W.L. and J.W.; Supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 32371583. This research was funded by the Student Practice Innovation and Training Program of Jiangsu Province, grant number 202210298018Z.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available. When the paper is accepted, we will make our dataset and its sources publicly available online. The datasets are available at https://github.com/WJ-0425/MammalClub (accessed on 31 October 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Viani, A.; Orusa, T.; Borgogno-Mondino, E.; Orusa, R. Snow Metrics as Proxy to Assess Sarcoptic Mange in Wild Boar: Preliminary Results in Aosta Valley (Italy). *Life* **2023**, *13*, 987. [CrossRef]

2. Feng, L.; Zhao, Y.; Sun, Y.; Zhao, W.; Tang, J. Action Recognition Using a Spatial-Temporal Network for Wild Felines. *Animals* **2021**, *11*, 485. [CrossRef]
3. Singh, A.; Pietrasik, M.; Natha, G.; Ghouaiel, N.; Brizel, K.; Ray, N. Animal Detection in Man-made Environments. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 1–5 March 2020.
4. Nguyen, H.; Maclagan, S.J.; Nguyen, T.D.; Nguyen, T.; Flemons, P.; Andrews, K.; Ritchie, E.G.; Phung, D. Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 19–21 October 2017.
5. Jia, J.; Fang, Y.; Li, X.; Song, K.; Xie, W.; Bu, C.; Sun, Y. Temporal Activity Patterns of Sympatric Species in the Temperate Coniferous Forests of the Eastern Qinghai-Tibet Plateau. *Animals* **2023**, *13*, 1129. [CrossRef]
6. Zhang, X.; Huo, L.; Liu, Y.; Zhuang, Z.; Yang, Y.; Gou, B. Research on 3D Phenotypic Reconstruction and Micro-Defect Detection of Green Plum Based on Multi-View Images. *Forests* **2023**, *14*, 218. [CrossRef]
7. Dai, X.; Xu, Y.; Song, H.; Zheng, J. Using image-based machine learning and numerical simulation to predict pesticide inline mixing uniformity. *J. Sci. Food Agric.* **2023**, *103*, 705–719. [CrossRef]
8. Vinitpornsawan, S.; Fuller, T.K. A Camera-Trap Survey of Mammals in Thung Yai Naresuan (East) Wildlife Sanctuary in Western Thailand. *Animals* **2023**, *13*, 1286. [CrossRef]
9. Zhong, Y.; Li, X.; Xie, J.; Zhang, J. A Lightweight Automatic Wildlife Recognition Model Design Method Mitigating Shortcut Learning. *Animals* **2023**, *13*, 838. [CrossRef]
10. Kays, R.; Lasky, M.; Allen, M.L.; Dowler, R.C.; Hawkins, M.T.; Hope, A.G.; Kohli, B.A.; Mathis, V.L.; McLean, B.; Olson, L.E.; et al. Which mammals can be identified from camera traps and crowdsourced photographs? *J. Mammal.* **2022**, *103*, 767–775. [CrossRef]
11. Figueroa, K.; Camarena-Ibarrola, A.; García, J.; Villela, H.T. Fast Automatic Detection of Wildlife in Images from Trap Cameras. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 2nd ed.; Bayro-Corrochano, E., Hancock, E., Eds.; Springer Nature: Cham, Switzerland, 2014; Volume 8827, pp. 940–947.
12. Alexander, G.V.; Augusto, S.; Francisco, V. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Inform.* **2017**, *41*, 24–32.
13. Janzen, M.; Ritter, A.; Walker, P.D.; Visscher, D.R. EventFinder: A program for screening remotely captured images. *Environ. Monit Assess* **2019**, *191*, 406. [CrossRef]
14. Orusa, T.; Borgogno Mondino, E. Exploring Short-Term Climate Change Effects on Rangelands and Broad-Leaved Forests by Free Satellite Data in Aosta Valley (Northwest Italy). *Climate* **2021**, *9*, 47. [CrossRef]
15. ENETWILD-consortium; Guerrasio, T.; Pelayo Acevedo, P.; Apollonio, M.; Arnon, A.; Barroqueiro, C.; Belova, O.; Berdión, O.; Blanco-Aguiar, J.A.; Bijl, H.; et al. Wild ungulate density data generated by camera trapping in 37 European areas: First output of the European Observatory of Wildlife (EOW). *EFSA Support. Publ.* **2023**, *20*, 7892E.
16. Enetwild, C.; Blanco-Aguiar, J.A.; Acevedo, P.; Apollonio, M.; Carniato, D.; Casaer, J.; Ferroglio, E.; Guerrasio, T.; Gómez-Molina, A.; Jansen, P.; et al. Development of an app for processing data on wildlife density in the field. *EFSA Support. Publ.* **2022**, *19*, 7709E. [CrossRef]
17. Falzon, G.; Lawson, C.; Cheung, K.-W.; Vernes, K.; Ballard, G.A.; Fleming, P.J.S.; Glen, A.S.; Milne, H.; Mather-Zardain, A.; Meek, P.D. *ClassifyMe*: A Field-Scouting Software for the Identification of Wildlife in Camera Trap Images. *Animals* **2020**, *10*, 58. [CrossRef]
18. Marcella, J.K. Computer-Aided Photograph Matching in Studies Using Individual Identification: An Example from Serengeti Cheetahs. *J. Mammal.* **2001**, *82*, 440–449.
19. Lyra-Jorge, M.C.; Ciocheti, G.; Pivello, V.R.; Meirelles, S.T. Comparing methods for sampling large- and medium-sized mammals: Camera traps and track plots. *Eur. J. Wildl. Res.* **2008**, *54*, 739–744. [CrossRef]
20. Nan, Y.; Zhang, H.; Zeng, Y.; Zheng, J.; Ge, Y. Intelligent detection of Multi-Class pitaya fruits in target picking row based on WGB-YOLO network. *Comput. Electron. Agric.* **2023**, *208*, 107780. [CrossRef]
21. Jin, X.; Liu, T.; Chen, Y.; Yu, J. Deep Learning-Based Weed Detection in Turf: A Review. *Agronomy* **2022**, *12*, 3051. [CrossRef]
22. Yan, X.; She, D.; Xu, Y. Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions. *Expert Syst. Appl.* **2023**, *216*, 119479. [CrossRef]
23. Saufi, S.R.; Ahmad, Z.A.B.; Leong, M.S.; Lim, M.H. Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review. *IEEE Access* **2019**, *7*, 122644–122662. [CrossRef]
24. Graving, J.M.; Chae, D.; Naik, H.; Li, L.; Koger, B.; Costelloe, B.R.; Couzin, I.D. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* **2019**, *8*, e47994. [CrossRef]
25. Bala, P.C.; Eisenreich, B.R.; Yoo, S.B.M.; Hayden, B.Y.; Park, H.S.; Zimmermann, J. Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nat. Commun.* **2020**, *11*, 4560. [CrossRef]
26. Yu, X.; Wang, J.; Kays, R.; Jansen, P.A.; Wang, T.; Huang, T. Automated identification of animal species in camera trap images. *EURASIP J. Image Video Process.* **2013**, *2013*, 52. [CrossRef]
27. Rey, N.; Volpi, M.; Joost, S.; Tuia, D. Detecting animals in African Savanna with UA Vs and the crowds. *Remote Sens. Environ.* **2017**, *200*, 341–351. [CrossRef]
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

29. Schneider, S.; Taylor, G.W.; Linquist, S.S.; Kremer, S.C. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol. Evol.* **2018**, *10*, 461–470. [CrossRef]
30. Steven, J.B.C.; Ian, P.B.; Jessica, J.M.; Peter, P.G. Automated marine turtle photograph identification using artificial neural networks, with application to green turtles. *J. Exp. Mar. Biol. Ecol.* **2014**, *452*, 105–110.
31. Nepovinnykh, E.; Eerola, T.; Kälviäinen, H. Siamese Network Based Pelage Pattern Matching for Ringed Seal Re-identification. In Proceedings of the 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW), Snowmass, CO, USA, 1–5 March 2020.
32. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [CrossRef]
33. Carl, C.; Schönfeld, F.; Profft, I.; Klamm, A.; Landgraf, D. Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *Eur. J. Wildl. Res.* **2020**, *66*, 62. [CrossRef]
34. Cheng, K.; Zhang, Y.; He, X.; Chen, W.; Cheng, J.; Lu, H. Skeleton-Based Action Recognition with Shift Graph Convolutional Network. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
35. Song, Y.F.; Zhang, Z.; Shan, C.; Wang, L. Richly Activated Graph Convolutional Network for Robust Skeleton-Based Action Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 1915–1925. [CrossRef]
36. Zhu, Q.; Deng, H. Spatial adaptive graph convolutional network for skeleton-based action recognition. *Appl. Intell.* **2023**, *53*, 17796–17808. [CrossRef]
37. Song, Y.F.; Zhang, Z.; Shan, C.; Wang, L. Constructing Stronger and Faster Baselines for Skeleton-Based Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 1474–1488. [CrossRef]
38. Hsing, P.; Hill, R.A.; Smith, G.C.; Bradley, S.; Green, S.E.; Kent, V.T.; Mason, S.S.; Rees, J.; Whittingham, M.J.; Cokill, J.; et al. Large-scale mammal monitoring: The potential of a citizen science camera-trapping project in the United Kingdom. *Ecol. Solut. Evid.* **2022**, *3*, 12180. [CrossRef]
39. McCallum, J. Changing use of camera traps in mammalian field research: Habitats, taxa and study types. *Mammal Rev.* **2013**, *43*, 196–206. [CrossRef]
40. David, J.A.; Pietro, P. Toward a science of computational ethology. *Neuron* **2014**, *84*, 18–31.
41. Beery, S.; Agarwal, A.; Cole, E.; Birodkar, V. The iWildCam 2021 Competition Dataset. *arXiv* **2021**, arXiv:2105.03494.
42. Ziegler, L.V.; Sturman, O.; Bohacek, J. Big behavior: Challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **2021**, *46*, 33–44. [CrossRef]
43. Cao, J.; Tang, H.; Fang, H.-S.; Shen, X.; Tai, Y.-W.; Lu, C. Cross-Domain Adaptation for Animal Pose Estimation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
44. Horn, G.V.; Mac Aodha, O.; Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; Belongie, S.J. The iNaturalist Species Classification and Detection Dataset. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; 2017; pp. 8769–8778.
45. Xian, Y.; Lampert, C.H.; Schiele, B.; Akata, Z. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2251–2265. [CrossRef]
46. Wah, C.; Branson, S.; Welinder, P.; Perona, P.; Belongie, S.J. The Caltech-UCSD Birds-200-2011 Dataset. 2011. Available online: https://www.vision.caltech.edu/datasets/cub_200_2011/ (accessed on 6 March 2022).
47. Gagne, C.; Kini, J.; Smith, D.; Shah, M. Florida Wildlife Camera Trap Dataset. *arXiv* **2021**, arXiv:2106.12628.
48. Swanson, A.; Kosmala, M.; Lintott, C.; Simpson, R.; Smith, A.; Packer, C. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Sci. Data* **2015**, *2*, 150026. [CrossRef]
49. Yu, H.; Xu, Y.; Zhang, J.; Zhao, W.; Guan, Z.; Tao, D. Ap-10k: A benchmark for animal pose estimation in the wild. *arXiv* **2021**, arXiv:2108.12617.
50. Ng, X.L.; Ong, K.E.; Zheng, Q.; Ni, Y.; Yeo, S.Y.; Liu, J. Animal Kingdom: A Large and Diverse Dataset for Animal Behavior Understanding. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 19001–19012.
51. Li, W.; Swetha, S.; Shah, D.M. Wildlife Action Recognition Using Deep Learning. 2020. Available online: https://www.semanticscholar.org/paper/Wildlife-Action-Recognition-using-Deep-Learning-Li-Swetha/3edcce3dd3d85da60115da988cf30253e3b59f19 (accessed on 6 March 2022).
52. Li, S.; Li, J.; Tang, H.; Qian, R.; Lin, W. ATRW: A benchmark for Amur tiger re-identification in the wild. *arXiv* **2019**, arXiv:1906.05586.
53. Guo, S.; Xu, P.; Miao, Q.; Shao, G.; Chapman, C.A.; Chen, X.; He, G.; Fang, D.; Zhang, H.; Sun, Y.; et al. Automatic Identification of Individual Primates with Deep Learning Techniques. *iScience* **2020**, *23*, 101412. [CrossRef]
54. Tang, J.; Zhao, Y.; Feng, L.; Zhao, W. Contour-Based Wild Animal Instance Segmentation Using a Few-Shot Detector. *Animals* **2022**, *12*, 1980. [CrossRef]
55. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

56. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision*, 2nd ed.; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; Volume 8693, pp. 740–755.
57. Liu, D.; Oczak, M.; Maschat, K.; Baumgartner, J.; Pletzer, B.; He, D.; Norton, T. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. *Biosyst. Eng.* **2020**, *195*, 27–41. [CrossRef]
58. Pistocchi, S.; Calderara, S.; Barnard, S.; Ferri, N.; Cucchiara, R. Kernelized Structural Classification for 3D Dogs Body Parts Detection. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014.
59. Freytag, A.; Rodner, E.; Simon, M.; Loos, A.; Kühl, H.S.; Denzler, J. Chimpanzee faces in the wild: Log-euclidean CNNs for predicting identities and attributes of primates. In Proceedings of the German Conference on Pattern Recognition, Hannover, Germany, 12–15 September 2016; pp. 51–63.
60. Zhao, Y.Q.; Feng, L.Q.; Tang, J.X.; Zhao, W.X.; Ding, Z.P.; Li, A.; Zheng, Z.Z. Automatically recognizing four-legged animal behaviors to enhance welfare using spatial temporal graph convolutional networks. *Appl. Anim. Behav. Sci.* **2022**, *249*, 105594. [CrossRef]
61. Zheng, Z.X.; Zhao, Y.Q.; Li, A.; Yu, Q.P. Wild Terrestrial Animal Re-Identification Based on an Improved Locally Aware Transformer with a Cross-Attention Mechanism. *Animals* **2022**, *12*, 3503. [CrossRef]
62. Sun, P.; Zhang, R.; Jiang, Y.; Kong, T.; Xu, C.; Zhan, W.; Tomizuka, M.; Li, L.; Yuan, Z.; Wang, C.; et al. Sparse r-cnn: End-toend object detection with learnable proposals. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
63. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
64. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
65. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
66. Quispe, R.; Pedrini, H. Top-DB-Net: Top DropBlock for Activation Enhancement in Person Re-Identification. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.
67. Rao, Y.; Chen, G.; Lu, J.; Zhou, J. Counterfactual Attention Learning for Fine-Grained Visual Categorization and Re-identification. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
68. Staniszewski, M.; Foszner, P.; Kostorz, K.; Michalczuk, A.; Wereszczyński, K.; Cogiel, M.; Golba, D.; Wojciechowski, K.; Polański, A. Application of Crowd Simulations in the Evaluation of Tracking Algorithms. *Sensors* **2020**, *20*, 4960. [CrossRef]
69. Ciampi, L.; Messina, N.; Falchi, F.; Gennaro, C.; Amato, G. Virtual to Real Adaptation of Pedestrian Detectors. *Sensors* **2020**, *20*, 5250. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.