This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

AnimalClue: Recognizing Animals by their Traces

Risa Shinoda[1,2,4], Nakamasa Inoue[3,4], Iro Laina[5], Christian Rupprecht[5], Hirokatsu Kataoka[4,5]
[1]The University of Osaka [2]Kyoto University [3]Institute of Science Tokyo
[4]National Institute of Advanced Industrial Science and Technology (AIST)
[5]Visual Geometry Group, University of Oxford

Figure 1. **Example images of AnimalClue.** We present AnimalClue, a dataset designed for identifying animal species based on their traces. Our dataset includes footprints, feces, eggs, bones, and feathers, totaling 159,605 bounding boxes from 968 animal species. We also annotate 22 traits such as habitat, diet, and activity pattern. We establish four benchmarks for evaluating models: classification, detection, instance segmentation, and traits classification. AnimalClue has the potential to advance research in animal tracking, an area that remains underexplored due to the limited availability of publicly accessible datasets.

Abstract

Wildlife observation plays an important role in biodiversity conservation, necessitating robust methodologies for monitoring wildlife populations and interspecies interactions. Recent advances in computer vision have significantly contributed to automating fundamental wildlife observation tasks, such as animal detection and species identification. However, accurately identifying species from indirect evidence like footprints and feces remains relatively underexplored, despite its importance in contributing to wildlife monitoring. To bridge this gap, we introduce ***AnimalClue****, the first large-scale dataset for species identification from images of indirect evidence. Our dataset consists of 159,605 bounding boxes encompassing five categories of indirect clues: footprints, feces, eggs, bones, and feathers. It covers 968 species, 200 families, and 65 orders. Each image is annotated with species-level labels, bounding boxes or segmentation masks, and fine-grained trait information, including activity patterns and habitat preferences. Unlike existing datasets primarily focused on direct visual features (e.g., animal appearances), AnimalClue presents unique challenges for classification, detection, and instance segmentation tasks due to the need for recognizing more detailed and subtle visual features. In our experiments, we extensively evaluate representative vision models and identify key challenges in animal identification from their traces. Our dataset and code are available at* https://dahlian00.github.io/AnimalCluePage/

1. Introduction

Conserving wildlife is essential for sustaining ecological balance and maintaining healthy ecosystems. However, numerous human activities, such as habitat destruction, chemical pollution, and the depletion of food resources, pose severe threats to wildlife species. To mitigate these impacts, computer vision technologies are increasingly expected to play a significant role by automating comprehensive and continuous observation of wildlife populations. For example, animal detection models implemented with unobtrusive

cameras can remotely monitor animals without disturbing their natural behaviors. This enables researchers to detect changes in behavioral patterns.

Nevertheless, identifying a wide range of animal species remains challenging, as most models overlook nocturnal animals or exhibit camouflage behaviors. Wildlife monitoring can leverage indirect biological evidence, such as footprints, bone fragments, and feathers to address this challenge. This approach is widely employed in ecological surveys; however, its scalability is limited due to heavy reliance on manual human inspection. Consequently, computer vision models capable of accurately identifying animal species from such indirect evidence are in high demand, as they could significantly expand the scale and effectiveness of wildlife monitoring efforts.

Comprehensive datasets have contributed to the progress of animal identification using computer vision [9, 15, 24, 35, 36, 40, 42, 46, 48]. However, the majority of computer vision datasets for animal studies have focused on direct visual observation tasks such as animal identifications [4, 23, 28], activity recognition [9, 11], image-text alignment [12, 47], and pose estimation [8, 51–53]. Several datasets involve indirect animal clues, but their focus is primarily limited to classification tasks with a small number of species [10, 10, 30, 33]. Given that classification, detection, and segmentation are established foundational tasks in the field of computer vision, developing a new indirect-evidence dataset that incorporates these tasks across a diverse range of species has significant potential to advance research.

Motivated by this, we introduce **AnimalClue**, a dataset that includes five major types of animal traces: footprints, feces, eggs, bones, and feathers, with sophisticated bounding boxes and segmentation masks. It encompasses 968 species with a total of 159,605 bounding boxes and 141,314 annotation masks. Furthermore, we annotate 22 species-specific traits, including habitat, activity pattern, and diet, to facilitate the research for animal tracking. Analyzing animal traits depending on their animal traces has not been previously explored.

We establish four benchmarks: classification, detection, instance segmentation, and habitat prediction for animal traces. Our findings indicate that models trained on our dataset develop the ability to recognize animal traces, which differ significantly from the visual features of the animals themselves. However, our dataset remains challenging: (a) Learning rare species is challenging, making generalization to rare species difficult. (b) Detection and instance segmentation tasks are particularly challenging, with models struggling to achieve high mAP. The highest mAP for order detection is 0.57, and the best order instance segmentation mAP is 0.48. We will make the dataset publicly available for research purposes.

2. Related Work

Animal species identification. Automating the identification of animal species is important for effective wildlife monitoring. Species labels can be obtained from various sources, including images [15, 24, 36, 42, 46, 48], videos [9, 35, 40], 3D models [50], and pose annotations [7, 53]. These datasets and benchmarks contribute to developing models and applying computer vision techniques in the animal domain. While direct animal identification has been well-explored, there remains room to investigate methods that identify animals indirectly, such as through animal traces left behind. In the rest of this section, we discuss related work on the identification of these traces.

Footprint identification. Recognizing animals from their footprints has been the subject of previous research, as humans also leave traces of their footprints, which often need to be identified [6, 22]. Similarly, identifying animal footprints is crucial for wildlife monitoring and conservation efforts. Previous works attempt individual identification [3, 18], species identification [3], and sex identification [26]. While previous research has explored the utility of footprints as animal clues, the associated images are not openly available. Recently, OpenAnimalTracks [41] was introduced as a classification and detection dataset for animal footprints. However, our AnimalClue differs from OpenAnimalTracks in several key aspects: (i) species distribution: AnimalClue includes 117 species compared to OpenAnimalTracks' 18 species; (ii) AnimalClue contains approximately five times more footprint bounding boxes, and additionally includes segmentation masks; and (iii) our dataset exclusively contains images licensed under Creative Commons.

Feces identification. Feces are a key indicator of health. In the medical field, several studies have explored automated feces analysis for human health [13, 16, 56], focusing primarily on feces detection and health classification. Recently, there has been research on automating the classification of chicken feces to assess health abnormalities into six categories [34]. For dogs, the DFML dataset [30] comprises 1,623 photos of dog feces, yet the recognition of feces across multiple species remains relatively unexplored. Considering the limited research efforts applying computer vision to animal feces compared to the human domain, AnimalClue presents a challenging dataset with the potential to advance animal feces recognition through computer vision.

Egg identification. Few previous studies have explored bird egg detection [19, 45, 49], and those that do typically focus on a single species due to their primary interest in agriculture rather than species identification for animal tracking. Moreover, these studies use data that are not publicly available. As for openly available datasets, the Automated Egg Classification dataset [10] includes 3k images

Dataset	Track Type	#Species	#Bbox	Task	#Traits
OpenAnimalTracks [41]	Footprint	18	3,579	CLS, DET	0
DFML [30]	Feces	1	1,623	CLS	1
Automated Egg Classification [10]	Egg	2	2,943	CLS	0
Skull2Animal dataset [33]	Bone	4	4,962	CLS	0
FeathersV1 [5]	Feather	595	28,272	CLS	0
AnimalClue (Ours)	All 5 Types	968	159,605	CLS, DET, SEG	22

Table 1. **Comparison with previous animal tracking datasets.** CLS, DET, and SEG indicate classification, detection, and instance segmentation, respectively. Our AnimalClue contains diverse species and a larger number of bounding boxes.

of chicken and duck eggs. However, species identification from eggs remains largely unexplored, and our dataset aims to address this gap.

Bone identification. Bone recognition has gained significant attention in the medical field, particularly for human bones, with several datasets contributing to this area [2, 29, 38]. For animals, the Skull2Animal dataset [33] includes skull images of four species (e.g., dogs, cats, leopards, and foxes), comprising 5k images. This dataset contributes to the prediction of skull images for animal species identification. For the purpose of animal identification, there are not enough publicly available datasets.

Feather identification. Feathers are key traces of birds, with their appearance varying depending on the bird species and the part of the body. During molting seasons, feathers can be found in abundance. FeathersV1 [5] is a dataset designed for fine-grained bird feather categorization, offering high-quality images to facilitate detailed classification. Our dataset differs from FeathersV1 in two significant ways: (i) most of our data is captured in the wild, not in controlled environments, making it more suitable and challenging for animal tracking, and (ii) we provide both bounding box and fine-grained pixel-level annotations.

3. AnimalClue

In this section, we introduce AnimalClue, for the purpose of wildlife conservation from multiple animal traces. During dataset collection, we carefully sourced images from the Internet under the Creative Commons license. For annotations, labels were verified by multiple citizen scientists, and we assigned precise bounding boxes and segmentation labels.

3.1. Trace Types and Animal Species

In tracing the types of animal species, we have collected five clues from animals: footprints, feces, bones, eggs, and feathers. These types of animal clues are also commonly used to recognize animal species in wildlife conservation. Figure 2 illustrates examples of the correspondence between animal traces and their respective species. For a complete species list, please refer to the supplementary material.

3.2. Data Collection

We have collected images from iNaturalist [17], a web platform for citizen scientists focusing on wildlife. We chose iNaturalist for two key reasons: i) Quality: The iNaturalist community comprises citizen scientists who ensure higher-quality data than random social media platforms. Images are reviewed by the iNaturalist community, ensuring reliable labels. We chose research-grade images, confirmed by other citizen scientists for the image and animal species pair. ii) Licensing: Users can select their images' licenses.

We have selected several iNaturalist projects that focus on uploading images of animal traces. From these, we downloaded the images under the Creative Commons licenses, along with their corresponding animal species labels. We have removed any ambiguous or unclear images from AnimalClue to improve data quality. Additionally, we exclude images containing human faces to protect privacy. Please refer to the supplementary material for examples of removed images. Importantly, we collect only research-grade images, where labels are verified by multiple researchers to ensure the quality of AnimalClue.

3.3. Data Annotation

Annotation type. Once we obtain the animal trace images and corresponding species labels, we proceed with bounding boxes and segmentation masks for object detection and instance segmentation tasks. Figure 2 illustrates the example images and their annotations on AnimalClue. Different annotation types are applied flexibly depending on the type of animal trace. For animal footprints, we annotate each footprint with a bounding box. In contrast, we use segmentation masks for feces and feathers to capture more detailed features. We limit footprint annotations to bounding boxes because footprints are traces rather than physical objects, and they are often ambiguous in some parts. In contrast, feces, bones, eggs, and feathers can be annotated with fine-grained, pixel-level segmentation masks. Defining clear bounding box groups for feces can be challenging, and in this sense, instance segmentation is helpful in deter-

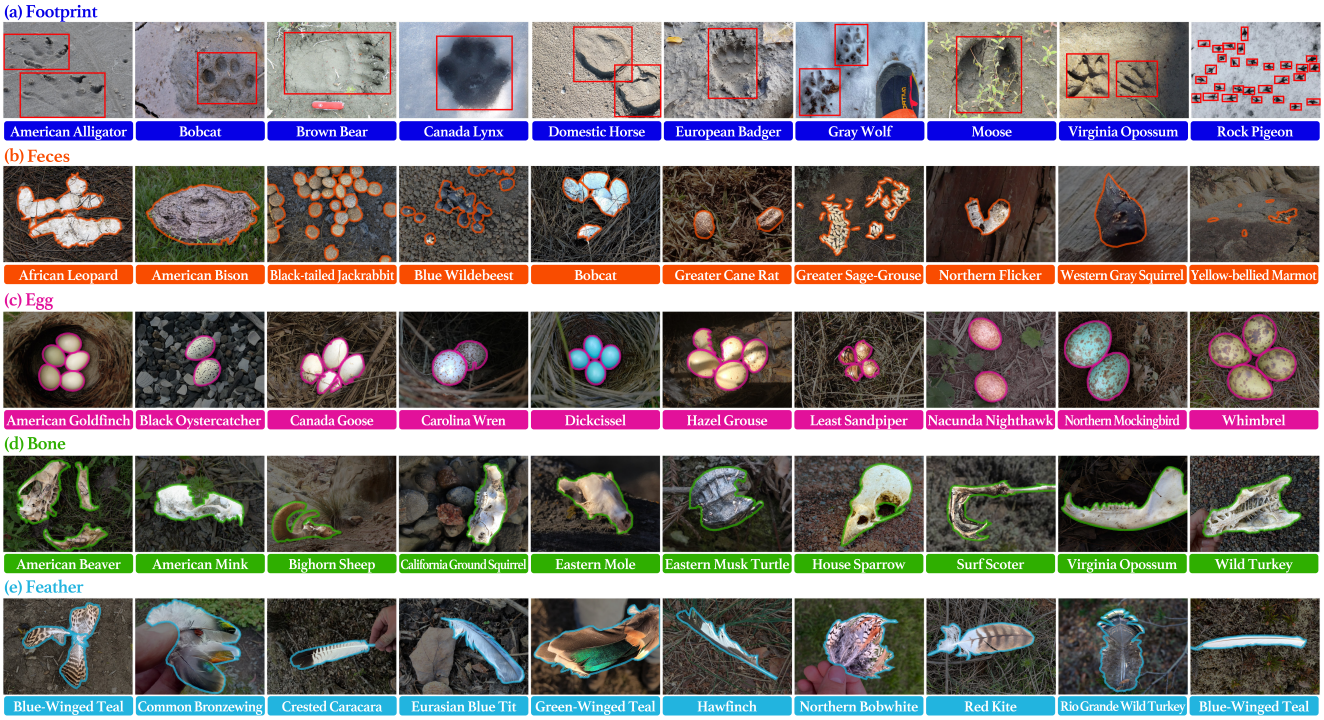

Figure 2. **Example images and annotations on AnimalClue.** The figures illustrate the five different animal clues, (a) footprints, (b) feces, (c) eggs, (d) bones, and (e) feathers, which are observed indirectly. There exist segmentation labels and bounding boxes for all clues except for footprints.

mining boundaries.

Annotation process. For half of the footprint annotations, we use a third-party annotation service to assign bounding boxes, and authors cross-check all bounding boxes to ensure they are precise enough for model training. For the remaining footprint, feces, egg, bone, and feather annotations, the authors conducted the annotations.

For footprints, feces, and bones, we manually annotate from scratch. For eggs and feathers, we use the Segment Anything Model as an initial annotation, which is then reviewed by the authors. If an annotated mask is incorrect, we manually correct it.

Fine-grained trait annotations. We annotated each species with multiple ecological and behavioral attributes to facilitate a comprehensive analysis of species traits. Specifically, we assigned the following attributes:

- Taxonomic Classification (Categorical): Each species was labeled with its Order and Family.
- Diet Type (Categorical): Species were classified as herbivorous, carnivorous, omnivorous, or other specialized feeding types.
- Activity Pattern (Categorical): Species were labeled as diurnal, nocturnal, crepuscular, or cathemeral based on their primary active periods.
- Locomotion and Habitat Usage (True / False): We assigned Arboreal, Aquatic, Terrestrial, Fossorial, and Aerial to indicate primary movement and habitat types. Locomotion was further categorized into Quadrupedal or Bipedal.
- Habitat Preferences (True / False): Each species was marked for its presence in Forest, Grassland, Desert, Wetland, Mountain, or Urban environments.
- Climatic Distribution (True / False): Species were categorized based on their occurrence in Tropical, Subtropical, Temperate, Boreal, or Polar regions.
- Social and Predatory Behavior (True / False): Herding indicates if the species lives in groups, while Predator specifies if it primarily hunts other animals.
- Migratory Behavior (True / False): Species were marked as Migratory if they undergo seasonal long-distance movements.

Dataset split. To ensure dataset quality, we do not separate images from the same iNaturalist entry into different dataset splits. Since a single entry can contain multiple images of the same subject, often taken from different angles, splitting them across training, validation, and test sets could lead to data leakage. Therefore, all images within the same submission are assigned to the same split. We divide the dataset into training, validation, and test sets using a 7:1:2 ratio.

Frequency categorization. To gain additional insights during the evaluation of models trained on AnimalClue, we categorized the dataset based on the training splits into three groups: frequent, intermediate, and rare categories. The

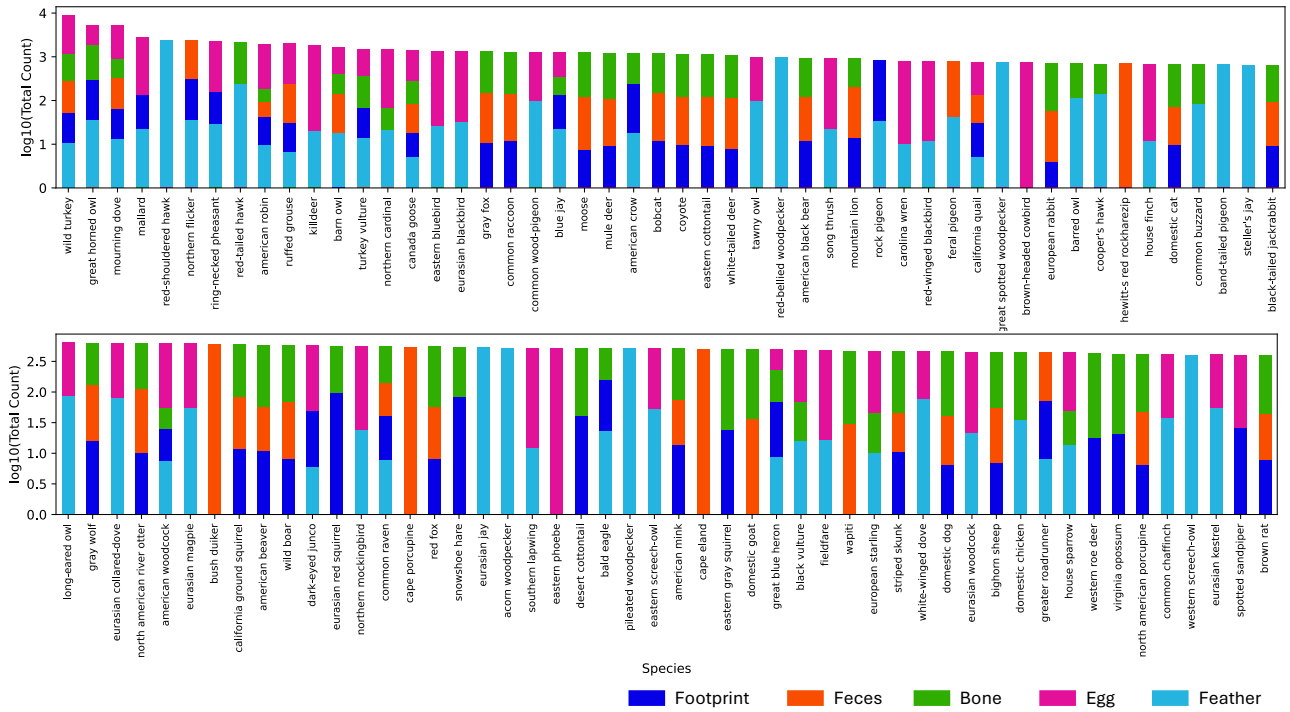

Figure 3. **Species distribution in our AnimalClue.** We illustrate the top 100 most frequent species, with the vertical axis representing log frequency.

taxonomy-based split was applied separately to footprints, feces, bones, eggs, and feathers. The top 20% of categories were classified as frequent, the next 60% as intermediate, and the bottom 20% as rare.

3.4. Statistics of AnimalClue

AnimalClue consists of 968 species, 200 families, and 65 orders. The dataset includes a total of 159,605 bounding boxes across five trace types:

- Footprints: 18,291 bounding boxes from 7,581 images, covering 117 species, 46 families, and 20 orders
- Feces: 18,932 bounding boxes from 6,433 images, covering 101 species, 46 families, and 21 orders
- Bones: 16,553 bounding boxes from 12,908 images, covering 269 species, 112 families, and 45 orders
- Eggs: 29,434 bounding boxes from 9,394 images, covering 283 species, 67 families, and 20 orders
- Feathers: 76,395 bounding boxes from 60,491 images, covering 555 species, 89 families, and 30 orders

The total number of bounding boxes matches that of the classification dataset, while the number of images aligns with the detection and segmentation datasets. Additionally, we provide detailed trait annotations, including activity patterns, habitat types, climatic conditions, and behavioral characteristics, to facilitate animal tracking from multiple perspectives. The species distribution across these trace sources is illustrated in Figure 3. We present the distribution of the top 100 species.

4. Experiments

We conduct experiments to validate and establish benchmarks on the proposed AnimalClue. Here, we show the image classification, detection, instance segmentation, and traits classification results.

4.1. Image Classification

Settings. As mentioned in Section 3.3, we cropped all animal traces using their assigned bounding boxes to minimize background effects as much as possible.

Baselines. We adopt five baseline models: VGGNet-16 (VGG-16) [31], ResNet-50 [14], EfficientNet-B1 (EFNet-B1) [44], Vision Transformer-Base (ViT-B) [25], and SwinTransformer-Base (Swin-B) [32]. To allow convergence, we train the listed recognition models for 100 epochs on footprints, feces, bones, and eggs, and 50 epochs on feathers.

The implementation details are listed in the supplementary materials.

Metrics. We report the top-1 accuracy that measures the percentage of times the model's highest-probability prediction matches the correct label.

Results. We show the top-1 accuracy (%) for different taxonomies and models in Table 2. The results indicate

Model	Species					Family					Order				
	Footprint	Feces	Egg	Bone	Feather	Footprint	Feces	Egg	Bone	Feather	Footprint	Feces	Egg	Bone	Feather
All Categories															
VGG-16 [31]	28.8	29.6	45.2	14.7	56.7	45.6	46.6	61.1	31.0	66.1	62.1	65.1	81.2	54.2	78.7
ResNet-50 [14]	23.7	29.4	41.1	18.3	59.7	41.7	48.6	59.9	33.6	70.1	58.9	64.8	80.8	53.0	81.8
EFNet-B1 [44]	25.9	30.5	41.0	15.0	55.9	42.2	48.4	56.3	29.3	64.4	56.5	61.3	74.5	45.4	77.1
ViT-B [25]	29.2	32.2	46.7	15.0	55.9	47.0	51.8	63.7	28.8	69.9	61.8	69.3	83.1	50.9	81.2
Swin-B [32]	32.3	38.6	49.4	20.5	65.3	49.3	56.8	65.1	37.6	72.4	66.1	70.4	84.0	56.6	77.7
Frequent Categories															
ResNet-50 [14]	24.8	30.6	49.4	26.7	68.5	44.0	51.7	65.5	42.8	74.2	62.9	70.6	83.9	62.7	80.4
Swin-B [32]	33.7	40.3	58.5	25.6	65.3	52.4	60.6	70.0	47.8	72.4	69.2	75.1	86.2	64.4	80.8
Rare Categories															
ResNet50 [14]	7.4	25.0	12.1	1.00	2.16	27.0	25.8	22.4	10.6	16.9	33.3	28.6	63.0	15.4	26.1
SwinB [32]	14.2	28.1	14.1	4.91	2.52	40.4	22.7	26.3	15.4	22.1	44.4	31.0	66.7	11.5	30.4

Table 2. **Classification accuracy for all, frequent, and rare categories of animal species**. Throughout the species, family, and order categorization, Swin-B model tends to have higher accuracies on AnimalClue.

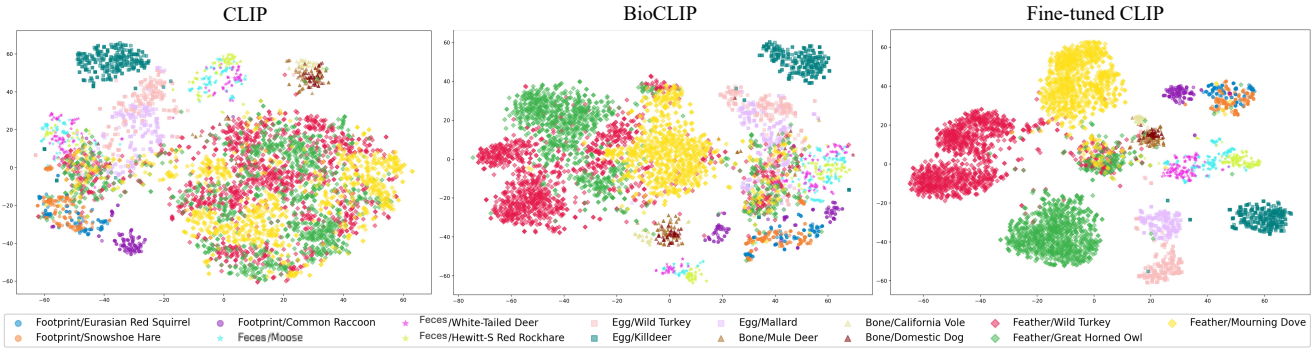

Figure 4. **Visualization of t-SNE.** By using a labeled dataset specialized for observing indirect animal clues, the separability among categories has been improved. When visualized in the feature space, the categories are better distinguished.

that Swin-B [32] tends to achieve the highest top-1 accuracy among the baseline models. Among the trace types, feathers exhibit the highest accuracy across all taxonomies, despite having the highest number of species. Following feathers, eggs achieve the second-highest accuracy, which also corresponds to the second-largest number of species. As shown in Figure 2, both feathers and eggs often feature distinctive colors and patterns, which may contribute to improved identification accuracy. In contrast, bones exhibit lower accuracy. According to Figure 2, bone appearance varies by body part, making it difficult to achieve high accuracy. Additionally, we present the results for both frequent and rare categories of animal species, as described in Sec. 3.3. Swin-B also achieves higher scores on AnimalClue.

4.2. Feature Space Analysis

Baselines. We evaluate CLIP-based models and analyze their feature extraction ability by visualizing t-SNE. The models used in our evaluation include: a) ViT-B-32, the standard CLIP model [37] pretrained by OpenAI. b) BioCLIP [43], a domain-specific CLIP model designed for biological image understanding. c) Fine-tuned CLIP, a CLIP model trained on AnimalClue using contrastive loss to improve performance on our specific dataset.

Metrics. For t-SNE visualization, we use a perplexity value of 30 and run 1000 iterations to ensure stable clustering.

Results. Figure 4 shows that our fine-tuned CLIP model achieves the best feature separation across categories. BioCLIP demonstrates better differentiation compared to the standard CLIP model, particularly in distinguishing between feathers and eggs. However, species-level feature separation remains challenging, especially for footprints and eggs. This indicates that for animal trace identification, models need to be trained specifically on trace data, even if they have been pre-trained on a large dataset of animal images.

4.3. Object Detection

Baselines. We adopt five baseline models: YOLOv8 [21], YOLOv11 [20], Faster-RCNN [39], DINO [54], RT-DETR [55]. We train the models for 100 epochs for

Model	Species					Family					Order				
	Footprint	Feces	Egg	Bone	Feather	Footprint	Feces	Egg	Bone	Feather	Footprint	Feces	Egg	Bone	Feather
All Categories															
YOLOv8 [21]	0.10	0.11	0.13	0.08	0.25	0.17	0.16	0.33	0.14	0.19	0.22	0.20	0.50	0.16	0.43
YOLOv11 [20]	0.10	0.12	0.14	0.07	0.25	0.17	0.16	0.36	0.13	0.35	0.24	0.21	0.47	0.16	0.43
Faster-RCNN [39]	0.04	0.06	0.12	0.07	0.08	0.07	0.09	0.17	0.0	0.12	0.08	0.13	0.26	0.10	0.22
DINO [54]	0.08	0.12	0.20	0.07	0.15	0.12	0.17	0.32	0.17	0.23	0.14	0.22	0.52	0.22	0.34
RT-DETR [55]	0.10	0.17	0.04	0.01	0.17	0.21	0.25	0.42	0.17	0.40	0.31	0.28	0.57	0.21	0.50
Frequent Categories															
YOLOv11 [20]	0.17	0.18	0.38	0.11	0.56	0.29	0.26	0.63	0.29	0.70	0.33	0.40	0.73	0.44	0.81
RT-DETR [55]	0.19	0.20	0.21	0.06	0.53	0.33	0.31	0.62	0.39	0.76	0.35	0.46	0.73	0.48	0.84
Rare Categories															
YOLOv11 [20]	0.04	0.05	0.04	0.003	0.14	0.07	0.08	0.07	0.05	0.05	0.10	0.02	0.26	0.007	0.005
RT-DETR [55]	0.006	0.09	0.00	0.00	0.00	0.11	0.21	0.08	0.00	0.00	0.34	0.10	0.38	0.10	0.00

Table 3. **Detection results on footprints, feces, eggs, bones, and feathers datasets (mAP@50-95).**

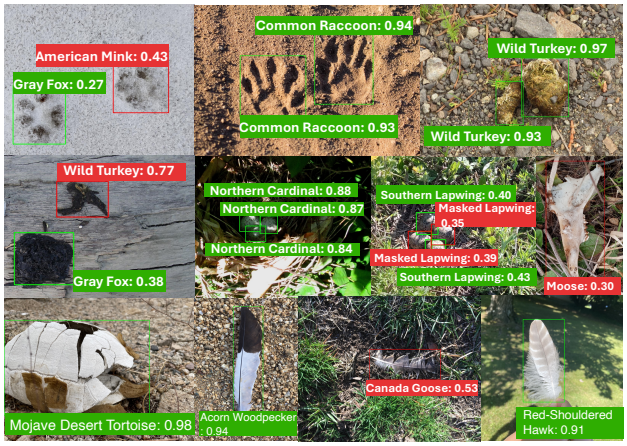

Figure 5. **Visualization of YOLOv11 detection results.** The green bounding box denotes the correct detection, and the red bounding box denotes the wrong detection.

YOLOv8 and YOLOv11, and 50 epochs for DINO, Faster-RCNN, and RT-DETR. The implementation details are found in the supplementary materials.

Metrics. We report the mean average precision (mAP) over classes. mAP is computed by the mean of AP on different thresholds of intersection of interest (IoU), *i.e.*, 0.5, 0.55, 0.60, ..., 0.95.

Result. We present the bounding box mAP in Table 3. Overall, RT-DETR achieves the best results across all categories. However, when examining rare categories, RT-DETR fails to detect them effectively, while its detection performance remains strong for frequent categories. The low mAP for rare categories contributes to the overall lower score across all categories. Among the animal trace types, feathers achieved the best results in both major categories and overall. However, detecting minor categories remains challenging. This may be attributed to the wide category range of feathers. In Figure 5, we show the visualization of object detection on AnimalClue using YOLOv11. The

Figure 6. **Visualization of YOLOv11 instance segmentation results.** The green mask denotes the correct mask, and the red mask denotes the wrong mask.

model correctly recognizes traces, including small ones, which are accurately identified at the species level. However, for faint footprints or traces in less visible areas, while they are recognized as traces, species misidentification is more common. Since our dataset focuses on images in the wild, it presents a challenging benchmark.

4.4. Instance Segmentation

Baselines. We adopt four baseline models: YOLOv8 [21], YOLOv11 [20], Mask-RCNN [1], and MaskDINO [27]. For training, YOLOv8 and YOLOv11 run for 100 epochs, whereas Mask R-CNN and Mask DINO are trained for 50 epochs. For implementation details, please refer to the supplementary materials.

Metrics. We report the mean average precision (mAP) for segmentation across classes. mAP is calculated as the mean

Model	Species				Family				Order			
	Feces	Egg	Bone	Feather	Feces	Egg	Bone	Feather	Feces	Egg	Bone	Feather
All Categories												
YOLOv8 [21]	0.11	0.11	0.07	0.24	0.14	0.29	0.13	0.33	0.20	0.44	0.15	0.41
YOLOv11 [20]	0.11	0.12	0.06	0.24	0.15	0.32	0.13	0.34	0.20	0.45	0.15	0.41
Mask-RCNN [1]	0.08	0.16	0.05	0.08	0.10	0.22	0.08	0.17	0.13	0.35	0.10	0.24
MaskDINO [27]	0.13	0.25	0.07	0.18	0.18	0.32	0.11	0.27	0.23	0.48	0.16	0.37
Frequent Categories												
YOLOv8 [21]	0.14	0.29	0.11	0.48	0.22	0.54	0.29	0.66	0.34	0.65	0.43	0.79
YOLOv11 [20]	0.16	0.33	0.10	0.49	0.24	0.57	0.28	0.68	0.37	0.64	0.45	0.80
Rare Categories												
YOLOv8 [21]	0.07	0.02	0.006	0.12	0.04	0.03	0.06	0.06	0.03	0.26	0.05	0.02
YOLOv11 [20]	0.06	0.03	0.005	0.11	0.07	0.06	0.08	0.05	0.06	0.27	0.05	0.002

Table 4. **Segmentation results on fecess, eggs, bones, and feathers datasets (Mask mAP@50-95).**

	Diet Type	Activity Pattern	Aquatic	Urban	Tropical	Polar	Herding	Predator
Footprint	57.3 / 47.0	69.9 / 65.8	84.4 / 70.2	95.2 / 60.0	91.3 / 63.5	62.2 / 61.8	84.2 / 73.3	80.7 / 94.3
Feces	76.0 / 64.7	72.9 / 72.1	93.1 / 75.1	98.6 / 59.3	82.7 / 70.0	69.1 / 45.6	74.7 / 53.4	88.4 / 85.7
Egg	68.7 / 48.2	95.4 / 73.0	88.4 / 83.4	76.8 / 75.2	90.6 / 69.7	93.8 / 75.4	82.5 / 81.5	96.9 / 73.4
Bone	64.8 / 59.0	60.6 / 60.7	81.5 / 73.3	75.2 / 61.4	72.6 / 64.9	87.5 / 62.7	73.9 / 73.1	76.9 / 72.1
Feather	81.3 / 80.0	89.7 / 63.7	92.1 / 81.6	83.0 / 81.5	91.5 / 69.2	94.7 / 61.9	86.3 / 81.7	91.8 / 90.7

Table 5. **Traits classification results on footprints, feces, eggs, bones, and feathers datasets (Acc/F1 score).**

of AP at different intersection over union (IoU) thresholds, *i.e.*, 0.5, 0.55, 0.60, ..., 0.95.

Result. YOLOv8, YOLOv11 and MaskDINO achieve competitive results across three different taxonomic levels: species, family, and order. Overall, the trend is similar to the results from object detection as shown in Table 4. Among animal traces, feathers tend to show high mAP despite having the highest number of species. This might be attributed to their distinct color appearance. Note that 'footprint' does not contain labels for instance segmentation. Therefore, we exclude it from this table. In Figure 6, we show the visual results of instance segmentation on AnimalClue using YOLOv11. The model tends to mislabel visually similar species (e.g., Masked Lapwing and Northern Lapwing, Gray Wolf and Coyote).

4.5. Traits Classification

Next, we conduct experiments on trait attribute classification using animal traces. Among the 22 traits, we present classification results for Diet Type (six categories) and Activity Pattern (three categories), as well as Aquatic, Urban, Tropical, Polar, Herding, and Predator, which are binary (True/False) classifications.

Baselines. Here, we chose Swin Transformer [32] for the baseline model, which achieved the best results in classification tasks.

Metrics. We report top-1 accuracy as a measure of overall classification performance and F1 score.

Results. As shown in the Table 5, feathers achieved the best results. For feces, the aquatic and predator traits showed high accuracy. This may be attributed to differences in feces characteristics between aquatic and terrestrial animals and the fact that higher hierarchical animals, such as predators, tend to have larger feces. Inferring an animal's traits from its traces may depend on the type of trace, with certain traces being more suitable for identifying specific traits.

5. Conclusion

We introduced AnimalClue, a dataset for identifying animal species based on their traces, including footprints, feces, eggs, bones, and feathers. Our comprehensive annotations for classification, detection, and pixel-level segmentation, along with 22 traits and taxonomy annotations, bridge the gap between animal tracking and computer vision. Our experiments show that there is still room for improvement, especially in rare category detection and segmentation. As animal tracking plays a crucial role in understanding animal movements in a given region and is a non-invasive approach, we hope our dataset contributes to advancing research in animal tracking and wildlife conservation.

6. Acknowledgment

This research was supported by the AIST policy-based budget project "R&D on Generative AI Foundation Models for the Physical Domain" and JST FOREST Grant Number JPMJFR206F. We used ABCI 3.0 provided by AIST and AIST Solutions.

References

[1] Waleed Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017. 7, 8

[2] Iftekharul Abedeen, Md. Ashiqur Rahman, Fatema Zohra Prottyasha, Tasnim Ahmed, Tareque Mohmud Chowdhury, and Swakkhar Shatabda. FracAtlas: A dataset for fracture classification, localization and segmentation of musculoskeletal radiographs. *Scientific Data*, 10(1), 2023. 3

[3] Sky Alibhai et al. A footprint technique to identify white rhino ceratotherium simum at individual and species levels. *Endangered Species Research*, 2008. 2

[4] William Andrew, Colin Greatwood, and Tilo Burghardt. Visual localisation and individual identification of holstein friesian cattle via deep learning. In *IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2850–2859, 2017. 2

[5] Alina Belko, Konstantin Dobratulin, and Andrey Kuznetsov. Feathers dataset for fine-grained visual categorization, 2020. 3

[6] Hanhua Cao, Huanping Zhang, Zhendan Liu, and Jianhuang Lai. Footprint recognition and feature extraction method based on artificial intelligence. In *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pages 1302–1305, 2020. 2

[7] Jinkun Cao, Hongyang Tang, Hao-Shu Fang, Xiaoyong Shen, Cewu Lu, and Yu-Wing Tai. Cross-domain adaptation for animal pose estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[8] Jinkun Cao, Hongyang Tang, Hao-Shu Fang, Xiaoyong Shen, Yu-Wing Tai, and Cewu Lu. Cross-domain adaptation for animal pose estimation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9497–9506, 2019. 2

[9] Jun Chen, Ming Hu, Darren J. Coker, Michael L. Berumen, Blair Costelloe, Sara Beery, Anna Rohrbach, and Mohamed Elhoseiny. Mammalnet: A large-scale video benchmark for mammal recognition and behavior understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13052–13061, 2023. 2

[10] Gaurav Dutta. Automated egg classification dataset. Kaggle, 2023. 2, 3

[11] Liqi Feng, Yaqin Zhao, Yichao Sun, Wenxuan Zhao, and Jiaxi Tang. Action recognition using a spatial-temporal network for wild felines. *Animals*, 11(2), 2021. 2

[12] Valentin Gabeff, Marc Russwurm, Devis Tuia, and Alexander Mathis. Wildclip: Scene and animal attribute retrieval from camera trap data with domain-adapted vision-language models. *bioRxiv*, pages 2023–12, 2023. 2

[13] David Hachuel, Akshay Jha, Deborah Estrin, Alfonso Martinez, Kyle Staller, and Christopher Velez. Augmenting gastrointestinal health: A deep learning approach to human stool recognition and characterization in macroscopic images, 2019. 2

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*. IEEE Computer Society, 2016. 5, 6

[15] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie. The inaturalist species classification and detection dataset. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8769–8778, Los Alamitos, CA, USA, 2018. IEEE Computer Society. 2

[16] Sae Hwang, JungHwan Oh, Wallapak Tavanapong, Johnny Wong, and Piet C. de Groen. Stool detection in colonoscopy videos. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3004–3007, 2008. 2

[17] iNaturalist. iNaturalist: Connecting People with Nature. Accessed: 2025-02-16. 3

[18] Zoe C Jewell, Sky K. Alibhai, Florian Johannes Weise, Stuart J. Munro, Marlice van Vuuren, and Rudie J. van Vuuren. Spotting cheetahs: Identifying individuals by their footprints. *JoVE*, 2016. 2

[19] Dengfei Jie, Jun Wang, Hao Wang, Huifang Lv, Jincheng He, and Xuan Wei. Real-time recognition research for an automated egg-picking robot in free-range duck sheds. *Journal of Real-Time Image Processing*, 22, 2025. 2

[20] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. 6, 7, 8

[21] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. 6, 7, 8

[22] Tanapon Keatsamarn and Chuchart Pintavirooj. Footprint identification using deep learning. In *2018 11th Biomedical Engineering International Conference (BMEiCON)*, pages 1–4, 2018. 2

[23] Faizan Farooq Khan, Xiang Li, Andrew J. Temple, and Mohamed Elhoseiny. Fishnet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20496–20506, 2023. 2

[24] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011. 2

[25] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 5, 6

[26] Binbin V. Li, Sky Alibhai, Zoe Jewell, Desheng Li, and Hemin Zhang. Using footprints to identify and sex giant pandas. *Biological Conservation*, 2018. 2

[27] Feng Li, Hao Zhang, Huaizhe xu, Shilong Liu, Lei Zhang, Lionel M. Ni, and Heung-Yeung Shum. Mask dino: Towards a unified transformer-based framework for object detection and segmentation, 2022. 7, 8

[28] Shuyuan Li, Jianguo Li, Hanlin Tang, Rui Qian, and Weiyao Lin. Atrw: A benchmark for amur tiger re-identification in the wild. In *Proceedings of the 28th ACM International Conference on Multimedia*, page 2590–2598, New York, NY, USA, 2020. Association for Computing Machinery. 2

[29] Zhangyong Li, Wang Chen, Yang Ju, Yong Chen, Zhengjun Hou, Xinwei Li, and Yuhao Jiang. Bone age assessment based on deep neural networks with annotation-free cascaded critical bone region extraction. *Frontiers in Artificial Intelligence*, 6, 2023. 3

[30] Jinyu Liang, Weiwei Cai, Zhuonong Xu, Guoxiong Zhou, Johnny Li, and Zuofu Xiang. A fine-grained image classification approach for dog feces using mc-scmnet under complex backgrounds. *Animals*, 13(10), 2023. 2, 3

[31] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, 2015. 5, 6

[32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5, 6, 8

[33] Alexander Martin, Haitian Zheng, Jie An, and Jiebo Luo. Jurassic world remake: Bringing ancient fossils back to life via zero-shot long image-to-image translation. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 9320–9328, New York, NY, USA, 2023. Association for Computing Machinery. 2, 3

[34] Arnas Nakrosis, Agne Paulauskaite-Taraseviciene, Vidas Raudonis, Ignas Narusis, Valentas Gruzauskas, Romas Gruzauskas, and Ingrida Lagzdinyte-Budnike. Towards early poultry health prediction through non-invasive and computer vision-based dropping classification. *Animals*, 13(19), 2023. 2

[35] Xun Long Ng, Kian Eng Ong, Qichen Zheng, Yun Ni, Si Yong Yeo, and Jun Liu. Animal kingdom: A large and diverse dataset for animal behavior understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19023–19034, 2022. 2

[36] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2

[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 6

[38] Pranav Rajpurkar, Jeremy Irvin, Aarti Bagul, Daisy Ding, Tony Duan, Hershel Mehta, Brandon Yang, Kaylie Zhu, Dillon Laird, Robyn L Ball, et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*, 2017. 3

[39] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *TPAMI*, 2017. 6, 7

[40] Tomoaki Saito, Asako Kanezaki, and Tatsuya Harada. Ibc127: Video dataset for fine-grained bird classification. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016. 2

[41] Risa Shinoda and Kaede Shiohara. Openanimaltracks: A dataset for animal track recognition. *arXiv preprint arXiv:2406.09647*, 2024. 2, 3

[42] Risa Shinoda and Kaede Shiohara. Petface: A large-scale dataset and benchmark for animal identification. *arXiv preprint arXiv:2407.13555*, 2024. 2

[43] Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun Chao, and Yu Su. BioCLIP: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19412–19424, 2024. 6

[44] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *ICML*, 2019. 5, 6

[45] Muammer Turkoglu. Defective egg detection based on deep features and bidirectional long-short-term-memory. *Computers and Electronics in Agriculture*, 185:106152, 2021. 2

[46] Grant Van Horn, Steve Branson, Ryan Farrell, Scott Haber, Jessie Barry, Panos Ipeirotis, Pietro Perona, and Serge Belongie. Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 595–604, 2015. 2

[47] Edward Vendrow, Omiros Pantazis, Alexander Shepard, Gabriel Brostow, Kate E. Jones, Oisin Mac Aodha, Sara Beery, and Grant Van Horn. Inquire: a natural world text-to-image retrieval benchmark. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2025. Curran Associates Inc. 2

[48] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. Caltech-ucsd birds-200-2011 (cub-200-2011). Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[49] Haijun Wang, Jianhua Mao, Jianyi Zhang, Huanyu Jiang, and Jianping Wang. Acoustic feature extraction and optimization of crack detection for eggshell. *Journal of Food Engineering*, 171:240–247, 2016. 2

[50] Jiacong Xu, Yi Zhang, Jiawei Peng, Wufei Ma, Artur Jesslen, Pengliang Ji, Qixin Hu, Jiehua Zhang, Qihao Liu, Jiahao Wang, Wei Ji, Chen Wang, Xiaoding Yuan, Prakhar Kaushik, Guofeng Zhang, Jie Liu, Yushan Xie, Yawen Cui, Alan Yuille, and Adam Kortylewski. Animal3d: A comprehensive dataset of 3d animal pose and shape. In *Proceedings of*

the IEEE/CVF International Conference on Computer Vision (ICCV), pages 9099–9109, 2023. 2

[51] Yuxiang Yang, Junjie Yang, Yufei Xu, Jing Zhang, Long Lan, and Dacheng Tao. Apt-36k: A large-scale benchmark for animal pose estimation and tracking. *Advances in Neural Information Processing Systems*, 35:17301–17313, 2022. 2

[52] Shaokai Ye, Anastasiia Filippova, Jessy Lauer, Steffen Schneider, Maxime Vidal, Tian Qiu, Alexander Mathis, and Mackenzie Weygandt Mathis. Superanimal pretrained pose estimation models for behavioral analysis. 2023.

[53] Hang Yu, Yufei Xu, Jing Zhang, Wei Zhao, Ziyu Guan, and Dacheng Tao. Ap-10k: A benchmark for animal pose estimation in the wild. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. 2

[54] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M. Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection, 2022. 6, 7

[55] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection, 2023. 6, 7

[56] Jin Zhou, Nick DeCapite, Jackson McNabb, Jose R. Ruiz, Deborah A. Fisher, Sonia Grego, and Krishnendu Chakrabarty. Stool image analysis for precision health monitoring by smart toilets. In *Proceedings of the 6th Machine Learning for Healthcare Conference*, pages 709–729. PMLR, 2021. 2