GCN-ID: A Benchmark Dataset for Great Crested Newt Re-Identification Using AI Foundation Models

Maria Joao Rosa[1]
maria.duarte-rosa@aru.ac.uk
Mohammed Abu Baker[1]
MYA120@aru.ac.uk
Chung Man Tang[1]
chung-man.tang@aru.ac.uk
Jim Littlemore[2]
james.littlemore@aru.ac.uk

[1] School of Computing and Information Science
Anglia Ruskin University
Cambridge, UK

[2] School of Life Sciences
Anglia Ruskin University
Cambridge, UK

Abstract

Great crested newts (*Triturus cristatus*) are a long-lived species of Class Amphibia and are protected by wildlife legislation across Europe. A key bioindicator of the health of the environment, great crested newts (GCN) have experienced significant population declines. Individual GCNs can be differentiated based on their unique orange-yellow underbelly markings with black blotching patternisation, but manually identification is time-consuming and requires domain knowledge. We collect and publicly release a comprehensive dataset of images (n = 1232) and videos (n = 1233) of GCNs (n = 206 different individuals) that can be used to develop AI models to aid in the autonomous identification of individuals of this species. Our data set comes with bounding boxes, segmentation masks, and multiple query-database splitting strategies. In addition, we propose two novel splitting methods that can be used to estimate performance inflation in the context of animal re-identification by: i) exploiting different biases in re-identification models; and ii) retrieving outlier samples with maximal difference in capture conditions. Using such splits, we show that the performance of State-of-the-Art (SoTA) foundation models for animal re-identification drops significantly (up to 36 percentage points for top-1 accuracy) when compared with random test splits. Nonetheless, we obtain a top-1 accuracy of 63% (top-5 of 83%) for the MegaDescriptor and a top-1 accuracy of 73% (top-5 of 93%) for the MiewID using the hardest bias-exploiting split with minimal data processing. This benchmarking exercise offers awareness of the effects of biases on accuracy estimation, but also encouraging baselines for future research on GCN re-identification based on AI foundation models.

1 Introduction

The field of automated image analysis for biodiversity knowledge is rapidly evolving, particularly with recent Artificial Intelligence (AI)-based approaches [□], helping to study imperiled species and combat the global biodiversity crisis. In the UK, the abundance of wildlife has declined on average by 19% since 1970 [5] and globally by approximately 70% [15]. Encouragingly, there is growing recognition of the value AI-based image analysis brings to biodiversity monitoring, especially its ability to accurately re-identify (re-id) individual animals, which is key to understanding population dynamics, behaviour, and ecosystem function. [18].

Great crested newts (GCNs) are a widespread but protected species found in Europe. Dependent upon freshwater ecosystems, in the UK this species has experienced long-term declines in populations attributed to the disappearance of ponds, the stocking of fish, water pollution, land use changes, habitat fragmentation, and urban development [10, 12]. Aligned with requirements for wildlife licensing in the UK, population numbers of GCN are typically estimated by conducting repeated visits to ponds by licensed personnel each spring, using torchlight surveys and live funnel-trapping methods. GCNs possess unique orange-yellow underbelly markings with black blotching patternisation, whereby specific individuals can be re-identified over time in passive 'capture-recapture' studies [9], but the sorting of photographic images is error-strewn and time-consuming.

It is therefore clear that for GCNs and other species, the application of computer-based tools for automatic animal re-identification could improve the accuracy of current population census methods and facilitate studies of population demography for conservation management [20]. Current methods to identify GCNs from photographs include manual inspection and matching, software algorithms that match local image-based features [7, 8, 14] and, more recently, AI-based models [4]. Manual inspection by eye is time-consuming and error-prone, with the accuracy usually dropping as the photo database gets larger. Local feature extractors can provide an alternative method, but do not scale well and may require substantial preprocessing and data engineering to extract meaningful features from images [20]. AI-based methods, in particular deep learning models, on the other hand, promise to achieve comparing or improved performances, whilst requiring much less manual work and explicit feature engineering [21].

To develop highly accurate deep learning (AI) models, developers usually need large amounts of data and significant computational resources. Fortunately, several pretrained (also known as foundation) models now exist that can be used specifically for animal re-id tasks. These models have been originally trained on hundreds of different animal re-identification datasets, and therefore do not need to be trained from scratch to be used on a new (and smaller) dataset. The most popular of these foundation models are the MegaDescriptor [17] and MiewID [16] models. The MegaDescriptor is a Swin (shifted window) Transformer based model that has been trained on 33 different animal species. The MiewID is a convolutional neural network (CNN) model based on EfficientNetVX that has been trained on 120 thousand samples and 15 thousand different individual animals. Both models have been trained to learn an effective image embedding space where images belonging to the same individual animal are clustered together in this space, whilst images from different individuals should be well-separated.

These models are open-source and relatively easy-to-use, requiring minimal AI experience and data processing. They show SoTA performance on numerous re-identification datasets [16, 17], but to the best of our knowledge, they have not been trained or applied to GCN re-identification. In addition, there seems to be no other deep learning model available in the literature that has been specifically trained on GCNs. One likely reason for this gap may be the fact that there are no publicly available datasets specifically released to advance re-id tasks on this species.

In this paper, we provide the first open access dataset for GCN re-id studies, including a large amount of image ($n = 1232$) and video ($n = 1233$) data (total number of files $n = 2465$) for a total of 206 individual animals. The image dataset also incorporates bounding boxes and segmentation masks to facilitate local feature extraction and AI-based model development. To the best of our knowledge, this is the first public release of an expert-curated dataset specific to GCNs at this scale.

In typical object re-identification tasks, the data is split into query and database (also called 'gallery' or 'reference') sets. One query sample comprises the image of one individual animal to be identified, by retrieving the images in the database most similar to the query image (e.g., the Top-1 or Top-5 most similar). To estimate the model's re-identification performance on the entire query set, we count how many times the correct individual was retrieved from the database, either in the Top-1 or Top-5 most similar images (Top-1 and Top-5 accuracies, respectively). The query-database split is usually done randomly from the pool of available data. However, random query-database splits may inflate model performance and not be faithful to what one would expect in the field. The main reason for this includes the fact that images of the same individual may share several biases such as the same background, posture alignment and lighting conditions, including reflections, camera, angle and image resolution, that the model can associate with that particular individual but do not correspond to true identification of the animal.

To mitigate this issue, Adam *et al.* (2024a) [1] designed a time-aware splitting method for sea turtles in which individuals that were encountered more than once had all the samples from one of their encounters placed in the query split, and the rest going into the database. As capture conditions of the query and database samples are different, it is highly likely that the query-database samples do not share many environmental biases. When compared to random splits, Adam *et al.* (2024a) [1] found that re-id foundation models significantly dropped their accuracies with time-aware splits.

These time-aware splits are indispensable when the data is time-stamped, but not all datasets present this feature. In another study, Adam *et al.* (2024b) [2] designed a splitting strategy to tackle this limitation, known as 'similarity-aware' splits. This method of splitting the data makes use of the fact that images from the same individual in the same encounter (taken in close temporal proximity) have high proximity in feature (embedding) space. By clustering the embedded images, each cluster is treated as an observation, and all of its images are assigned to either query or database sets. The authors used this splitting strategy, combined with the time-aware splits, to create the Wildlife-10K dataset [2]. By having default splits that minimise environmental biases, it is hoped that this dataset acts as a benchmark for the field of animal re-identification, but does not contain GCNs.

Unfortunately, as can be common in the field, our dataset is not timestamped nor does it contain true multiple encounters from the same individual, making the discussed splitting strategies not directly applicable to our data. Nevertheless, we utilise the same concepts to design two novel strategies for splitting our dataset into query-database sets: a) "Least-similar" strategy: given a query image, the "least-similar" split looks for samples from the same individual animal in the database, selecting only from those that exhibit the greatest separation from their nearest neighbours in feature space belonging to that same animal; b) "Hard/Bias-exploiting" strategy: the "hard" split works in a similar way to the "least-similar" splitting strategy, but in combination with selecting the least similar images for each individual, we simultaneously search in feature space for samples from different individuals that are most similar to the query image. These two strategies try to make the problem of re-identification harder for the foundation models, in order to assess how their performance is compromised by different environmental biases.

In summary, this study makes the following contributions:

- A new open-access dataset composed of more than one thousand images and videos of GCNs corresponding to more than two hundred individual animals, together with bounding boxes and segmentation masks.
- Three different query-database splitting strategies that allow researchers to assess and exploit the environmental biases in the dataset.
- Benchmarks for the dataset using two AI foundation models for animal re-id (the MegaDescriptor and MiewID), establishing baseline results (including top-1 and top-5 accuracies) for future GCN re-id studies, utilising different query-database splitting strategies.

2 Methodology

2.1 Data Collection

A total of eight ponds known to historically hold populations of GCN were surveyed on multiple occasions between 9th May to 13th June 2024 at a field site on the outskirts of Cambridge, UK. The precise location of the third-party owned ponds have been anonymised to comply with best practice guidance for handling sensitive biological records. Each survey visit included an early evening funnel trap deployment exercise for live capture (Figure 1) performed in accordance with the standardised methods outlined by Natural England (2001). The following morning, live capture funnel traps were emptied and captured GCN individuals were photographed. All work was supervised by an ecologist complying with welfare and ethical requirements.

Figure 1: (a) Early evening funnel trap deployment; (b) Early morning trap check with captured live GCN; (c) GCN with characteristic underbelly pigmentation.

Field images were obtained from stills and live photo bursts captured mostly in daylight hours using an iPhone SE second generation smartphone running iOS 17. For image capture, amphibians were humanely placed in a small transparent plastic container $15.5x11.5x4.5cm$ and gently held in place using a moist sponge. An additional glass container was also used. The box was manually inverted to enable photographs to be taken of the underbelly 'region of interest'. Several images of the animals were taken at different angles and lighting levels. The photographing setup was not deliberately contrived, but it represents a realistic dataset typical of photographs collected by ecologists using handheld devices in the field. The image and video components of the live photos were then separated into distinct files. A total of $n = 1232$ images and $n = 1233$ videos were captured for $n = 206$ live GCN captures. Figure 2 illustrates the count of files for each identity, where most individuals captured were photographed 4-6 times. Data annotation (i.e. GCN identity) was done and recorded as the data were collected. No re-captures of the same newt were recorded across different survey dates.

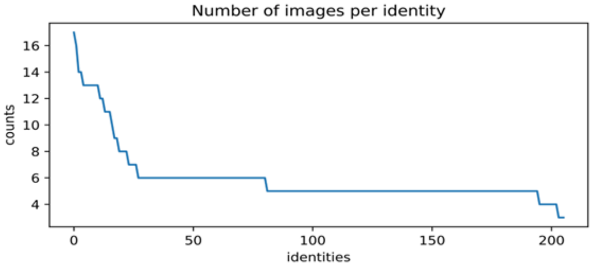

Figure 2: The count of GCN files for each identity.

2.2 Data Processing

For this study, we used raw or minimally processed data. The processing steps aimed to be as unsupervised as possible so that the workflow is dataset-agnostic. This processing consisted of obtaining bounding boxes and segmentation masks from the image data. The video data, although present in our dataset, was not processed in this benchmark study. The boxes and masks were obtained using first a zero-shot open-set object detection GroundingDINO model [13] and then a SAM-2 unified promptable segmentation model [19]. GroundingDINO takes a text prompt and an image and returns one or more bounding boxes with certainty scores. We manually experimented with different text prompts and the highest confidence scores were obtained with `"newt amphibian reptile"`. SAM-2 can take as input image or video data and returns segmentation masks of objects guided by points or bounding boxes. The bounding box around the GCN that was obtained from the previous step was ported to SAM-2, along with its raw image, and the resulting segmentation mask was saved to the dataset's CSV file in Run Length Encoding (RLE) format.

GroundingDINO succeeded in drawing bounding boxes around 97% of samples. To assess correctness, 100 random samples were manually inspected, and all were confirmed to be satisfactory. The remaining samples (3%) were not recognised by GroundingDINO. For these samples, bounding boxes were manually created. SAM-2 successfully created segmentation masks for all samples that had bounding boxes. 100 samples were randomly selected and manually checked for quality assurance. Most samples were accurate, with only the newt being included in the mask. However, a handful of the checked samples had the mask covering other areas of the image beside the GCN. For these images, we added negative points around the bounding box edges and re-ran SAM-2 to obtain a better segmentation output. For the unchecked images, it is possible that a small number of these images might contain parts of the background that were not corrected.

(a) Original

(b) Rotated

(c) Background removed

(d) Background removed and rotated

Figure 3: An example of a raw, rotated, segmented (background removed) and rotated + background removed sample from our dataset.

2.3 Data Splits

To evaluate model performance in animal re-identification tasks, how the dataset is divided into query and database sets is critical. The query set contains images of the animal that the model needs to identify, and the database contains a relatively larger set of images that we compare the query with, in order to retrieve the correct match for each query image. This split simulates real-world scenarios, in which ecologists try to identify a new image of an animal by matching it to known individuals in the database.

The decision on how to allocate images to the query and database sets can be made randomly or be based on specific criteria. Randomly allocating individuals has been commonly used, but can lead to performance overestimation, given that different samples from the same encounter taken under identical conditions (e.g. background, lighting, posture) may end up on both query and database sets. This problem is called ‘bias’ in re-identification tasks. In this study, we estimate the effect of this bias in GCN re-identification and provide ways to split the data that try to mitigate the effects of this bias.

For this effect our dataset can be divided using the following data splits (for all splits the default query size is 30 images, the rest of the dataset corresponds to the database):

- Random: a random split, where in our case 30 randomly chosen images are assigned to the query set and the rest to the database. Usually this is the first choice for many re-id studies. We have evaluated both models on a random split to compare against other splitting strategies.

- Least-similar: For each individual, this strategy selects samples that exhibit the greatest distance from their nearest neighbours in model embedding space. To create such a split, the whole dataset was run through both MiewID and MegaDescriptor to obtain embeddings for each image. Then the cosine similarity between all samples was calculated. For each sample and individual, the highest cosine similarity of the same identity was calculated. Samples with the minimum value for each individual were selected for the query set.

- Hard/Bias-exploiting: This works in a similar way to the least-similar splitting strategy, but whilst trying to find the least similar image for each individual, we simultaneously search for samples that were most similar to incorrect individuals. All samples were again processed using the MiewID and MegaDescriptor to obtain embedded images. The highest correct (same individual) similarity score and incorrect (different individual) similarity scores were calculated. These were then subtracted and averaged across models. The query samples that possess the highest average score are chosen for the query set (Figure 4).

2.4 Model Baselines

We assess the performance of the MegaDescriptor and MiewID animal re-id foundation models using the splitting strategies described above. In addition, we also assess the performance of these models when we try to eliminate environmental biases in the data using the following techniques:

- Background removal: The most significant bias in re-identification datasets is the background. As many re-id datasets do not contain re-encounters, or the images are extracted from videos, the background remains largely static across images of the same individual. To remove the background bias, we used the segmentation masks (described above) to crop out the individual newt from the image.

- Rotation: Animal posture and alignment can also be a confound in identification tasks. Therefore we also rotated all images to a 90 degree angle from their original orientation, except for the query image.

For all models, splitting strategies, and bias-removal methods, we use the following metrics to obtain baseline results: top-1 or top-5 accuracies check whether the most similar image (top-1) or one of the 5 most similar images (top-5) in the database belongs to the same identity of the query image; mAP is a standard metric that evaluates the ranking quality of retrieved results. For each query, it calculates the precision at each rank where the correct match appears, it then averages the precision across ranks (AP). To obtain the final mAP values, the average precisions (AP) are averaged across all queries.

3 Results

Both foundation models reach the highest top-1 and top-5 accuracies on the dataset when the data is split randomly and no debiasing technique has been performed on the model (Table 1). Moreover, the MegaDescriptor achieves top-1 and top-5 accuracies of 93% and 100%, respectively, while MiewID achieves 93% and 96.7%. This result aligns with our understanding that model performance is inflated when environmental biases are not taken into account (either in the splitting strategy or in the processing of the data).

The splitting strategies introduced for addressing the bias in the data affect the model performance significantly even when the images have no further processing. For instance, the MegaDescriptor top-1 score drop from 93.3% on random sets, to 63.3% on bias-exploiting sets, and finally to 56.6% when the least-similar strategy is used. The MiewID appears to be more robust to different splitting strategies than the MegaDescriptor, with the largest drop being the bias-exploiting split with a 73.3% top-1 accuracy.

Debiasing techniques affect scores even in the absence of any sophisticated splitting strategy, for instance on random sets. Moreover, by removing the background on the random split image samples, top-1 accuracies for MegaDescriptor drops by 13.3% and that of MiewID drops by 6.6%. When the match database is also rotated, scores decrease significantly more, even on the random split whereby the 90% top-1 accuracies drops to 23.3% for

	MegaDescriptor			MiewID		
	Top-1	Top-5	mAP	Top-1	Top-5	mAP
Random						
Original	93.3%	100.0%	77.1%	93.3%	96.7%	85.0%
Background Removed	80.0%	93.3%	65.8%	86.7%	96.7%	72.0%
Rotated	16.7%	73.3%	22.4%	3.3%	36.7%	9.5%
Background Removed + Rotated	23.3%	56.7%	21.1%	6.7%	23.3%	9.4%
Least similar						
Original	56.7%	90.0%	49.8%	76.7%	90.0%	57.1%
Background Removed	36.7%	60.0%	29.0%	36.7%	60.0%	19.6%
Rotated	13.3%	60.0%	21.7%	13.3%	36.7%	13.4%
Background Removed + Rotated	16.7%	40.0%	12.3%	3.3%	20.0%	6.3%
Hard						
Original	63.3%	83.3%	46.0%	73.3%	93.3%	49.0%
Background Removed	46.7%	60.0%	30.4%	46.7%	66.7%	28.0%
Rotated	16.7%	56.7%	18.1%	3.3%	26.7%	7.5%
Background Removed + Rotated	6.7%	33.3%	11.4%	0.0%	6.7%	4.2%

Table 1: Accuracies of different deep models for under different splitting strategies and sample de-biasing techniques.

MegaDescriptor and to 6.7% for MiewID, which is a 70% and 86% drop from the original images, respectively.

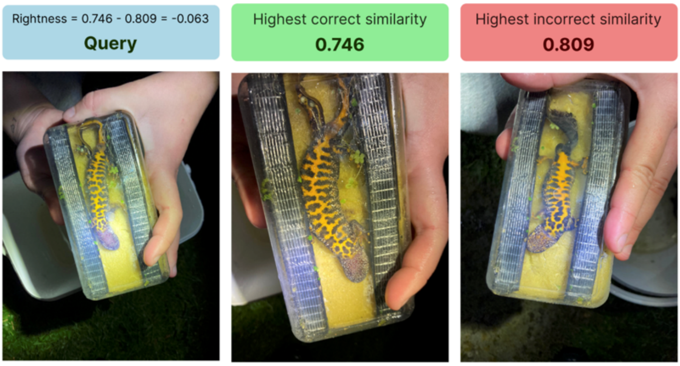

Figure 4: Example query image of numbered GCN 13 sharing more of bias factors with the wrong identity (GCN17) than with the highest scoring image for the same individual.

Our baseline results show that both foundation models achieve very high scores on random sets and biased images, but their susceptibility to debiasing and splitting is different. The MegaDescriptor appears to be more susceptible to the introduced splitting strategies but slightly more robust to debiasing, whilst for the MiewID the opposite seems to be the case.

Although we show that foundation models for animal re-id can be very susceptible to environmental biases and results obtained on randomly divided data can be highly inflated, accuracies for the hard/bias-exploiting set are still encouraging for GCN re-id: top-1 accuracy of 63% (top-5 of 83%) for the MegaDescriptor and top-1 accuracy of 73% (top-5 of 93%) for the MiewID on the raw data; and top-1 accuracy of 47% (top-5 of 60%) for the MegaDescriptor and top-1 accuracy of 47% (top-5 of 67%) for the MiewID on the segmented data. Although not yet sufficient for accurate and efficient re-id in the field, this benchmarking exercise offers encouraging baselines for future research on GCN re-identification using AI models.

4 Conclusions

We collect and make public a curated image ($n = 1232$) and video ($n = 1233$) dataset of a population of 206 great crested newt individuals along with AI-targeting features such as bounding boxes and segmentation masks. We also include multiple subsets based on different query-database splitting strategies, including two novel strategies, that allow model performance evaluation exploiting different bias conditions. We tested two popular SoTA foundation models (the MiewID and MegaDescriptor) for animal re-identification on the GCN dataset, utilising these different splits and bias addressing techniques. We show that the performance of these models drops significantly (up to 36 percentage points for top-1 accuracy) when compared with random test splits. Nevertheless, we obtain a top-1 accuracy of 63% (top-5 of 83%) for the MegaDescriptor and a top-1 accuracy of 73% (top-5 of 93%) for the MiewID using the hardest bias-exploiting split with minimal data processing. This benchmark results and the released dataset offer awareness of the effects of biases on performance estimation, hopefully inviting and facilitating further AI model development for animal re-id.

Limitations: the most significant limitation of the GCN dataset is its lack of annotated images from the same individual collected across different surveys and/or close ponds (recapture events). A study by Drechsler *et al*. (2015) [9] found recapture rates of GCN around 9.83% using a local features-based software tool and $n = 1648$ images. Since the main focus of our study was on performance inflation due to environmental biases, we did not explicitly find any instances of re-captures. But we cannot confirm that they do not exist in our dataset. Further manual checking of similarity rankings is needed to confirm this.

Like other amphibians, GCN may undergo slight ventral pigmentation changes on their underbelly as they age and these patterns are not fixed [3]. Future model development would need to assimilate changes in patternisation in the region of interest. These changes have not been addressed in this study.

One natural avenue for future work is to include more recently developed models, such as [6], and to fine-tune the foundation models on our dataset. To date, none of these models have been trained on GCN. The next step would be to re-train these models using a training set from our data and possibly include different training strategies again to mitigate the effect of environmental biases. One other issue we have not addressed in this work is the imbalance in the identities of the animals, as seen in Figure 2. To obtain more reliable accuracies the effect of this imbalance should be addressed in future studies, for example by reporting average accuracies per identity.

To better evaluate the ecological relevance of our GCN dataset and the re-id models benchmarked in this study, a new round of image/video collection from the same ponds is planned. This extended dataset will potentially allow us to identify true capture-recapture events and refine the model baselines to better reflect real-world field applications.

5 Acknowledgment

This work was supported by Anglia Ruskin University (ARU), through the HEIF and QR Commercial Engagement Funding for the project AI for Autonomous Biodiversity Assessment for GCNs, and by Knowledge Transfer Partnership (KTP) grants from the Innovation Vouchers Fund. The authors would like to thank the Cambridge City Council, People's Trust for Endangered Species and Chevron Green Services for their collaboration and support.

References

[1] Lukáš Adam, Vojtěch Čermák, Kostas Papafitsoros, and Lukas Picek. Seaturtleid2022: A long-span dataset for reliable sea turtle re-identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 7146–7156, 2024.

[2] Lukáš Adam, Vojtěch Čermák, Kostas Papafitsoros, and Lukas Picek. Wildlifereid-10k: Wildlife re-identification dataset with 10k individual animals. *arXiv preprint arXiv:2406.09211*, 2024.

[3] JW Arntzen and SFM Teunis. A six year study on the population dynamics of the crested newt (triturus cristatus) following the colonization of a newly created pond. *Herpetological Journal*, 3(3):99–110, 1993.

[4] LARS Briggs, RIINU Rannap, PIRET Pappel, FLORIAN Bibelriether, and ALEKSI Päivärinta. Monitoring methods for the great crested newt triturus cristatus. *Protection of Triturus cristatus in the Eastern Baltic region LIFE2004NAT/EE/000070 Action A*, 2, 2006.

[5] Fiona Burns, Simone Mordue, Nida Al-Fulaij, Philipp Boersh-Supan, Katherine L Boughey, Philip Briggs, Mark A Eaton, Colin Harrower, Angus C Jackson, Steve Langton, et al. State of nature 2023 terrestrial and freshwater animal dataset for the united kingdom and its constituent countries. *Data in Brief*, page 111646, 2025.

[6] Vojtěch Cermak, Lukas Picek, Lukáš Adam, Lukáš Neumann, and Jiří Matas. Wildfusion: Individual animal identification with calibrated similarity fusion. In *European Conference on Computer Vision*, pages 18–36. Springer, 2024.

[7] Jonathan P Crall, Charles V Stewart, Tanya Y Berger-Wolf, Daniel I Rubenstein, and Siva R Sundaresan. Hotspotter—patterned species instance recognition. In *2013 IEEE workshop on applications of computer vision (WACV)*, pages 230–237. IEEE, 2013.

[8] Paul M de Zeeuw, Elena Ranguelova, and Eric J Pauwels. Towards an online image-based tree taxonomy. In *Industrial Conference on Data Mining*, pages 296–306. Springer, 2007.

[9] Axel Drechsler, Tobias Helling, and Sebastian Steinfartz. Genetic fingerprinting proves cross-correlated automatic photo-identification of individuals as highly efficient in large capture–mark–recapture studies. *Ecology and Evolution*, 5(1):141–151, 2015.

[10] Tony Gent. The conservation of the great crested newt triturus cristatus in the uk. *Rana*, 4:295–305, 2001.

[11] Virginia Gewin. Ai carves out a niche in ecology and conservation research. *Nature Methods*, pages 1–6, 2025.

[12] Richard A Griffiths. Great crested newts (triturus cristatus) in europe. *Species conservation and management. Oxford University Press, Oxford*, pages 281–291, 2004.

[13] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European conference on computer vision*, pages 38–55. Springer, 2024.

[14] Maximilian Matthé, Marco Sannolo, Kristopher Winiarski, Annemarieke Spitzen-van der Sluijs, Daniel Goedbloed, Sebastian Steinfartz, and Ulrich Stachow. Comparison of photo-matching algorithms commonly used for photographic capture–recapture studies. *Ecology and evolution*, 7(15):5861–5872, 2017.

[15] Emirates Nature-WWF. Annual report 2021 a conservation diary. 2022.

[16] Lasha Otarashvili, Tamilselvan Subramanian, Jason Holmberg, JJ Levenson, and Charles V Stewart. Multispecies animal re-id using a large community-curated dataset. *arXiv preprint arXiv:2412.05602*, 2024.

[17] K Papafitsoros et al. Wildlifedatasets: An open-source toolkit for animal re-identification. 2023.

[18] Laura J Pollock, Justin Kitzes, Sara Beery, Kaitlyn M Gaynor, Marta A Jarzyna, Oisin Mac Aodha, Bernd Meyer, David Rolnick, Graham W Taylor, Devis Tuia, et al. Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity*, pages 1–17, 2025.

[19] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

[20] Stefan Schneider, Graham W Taylor, Stefan Linquist, and Stefan C Kremer. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, 10(4):461–470, 2019.

[21] Stefan Schneider, Graham W Taylor, and Stefan C Kremer. Similarity learning networks for animal individual re-identification: an ecological perspective. *Mammalian Biology*, 102(3):899–914, 2022.