Computer Vision for Wildlife Monitoring and Research

A literature review on computer vision systems to monitor, research, and protect wildlife species

Haider, Tobias Abraham

Vienna University of Technology

Interdisciplinary Project in Data Science

June 2025

Introduction

While biodiversity is decreasing at a rapid pace, the rise of specific species, be they invasive or predatory, concerns societies around the world. As a consequence, researchers and conservationists are interested in continuously monitoring wildlife populations in terms of their geographical distribution, size, and behavior. One tool that is employed with success is camera traps. These self-sufficient camera systems can take photographs of passing animals without disturbing them (Trolliet et al., 2014, 446-447). The photos are manually collected from the traps and annotated with the name of the species present in the image (Tulasi et al., 2023, 2). This literature review examines the scientific efforts to automate wildlife monitoring systems using computer vision and machine learning.

Use Cases for Wildlife Images

Wildlife monitoring has many purposes. The primary use cases include the conservation and health monitoring of endangered wildlife populations, as well as the protection of humans, livestock, and crops from wild animals (Neupane et al., 2022, p. 287). As an example, cameras attached to single-board computers are set up near fauna road crossings in Australia to monitor the movements of koala populations and prevent road fatalities (Trevathan et al., 2025, 101474). Another example is the reidentification of problem-causing Ezo Sika deer populations to protect farmland in Japan (Harie et al., 2021, 199). Furthermore, camera traps can provide valuable data for research on animal behavior. Unlike other methods, such as tagging, branding, marking, and mutilation, camera traps enable non-invasive and cost-effective tracking of individual animals (Zemanova, 2020, 2-3).

Wildlife Computer Vision Tasks

To support these use cases, computer vision is employed to process images and solve very specific tasks or compute metrics that are extremely time-consuming when done by hand. The accuracy of estimating such a metric determines the quality of a wildlife monitoring system. Also, computer vision algorithms are often compared to each other by how well they can estimate these metrics (Neupane et al., 2022, 286).

Identity

Identification refers to the detection of an animal species or individual. Species-level identification is a classification task for which a broad set of model types can be trained. Individual-level identification requires similarity measures between pictures to determine the likelihood that two images depict the same individual. Individual-level identification is also referred to as reidentification. Animal identification models are employed to select relevant samples from large sets of images, and they are a crucial tool for automating animal conservation efforts and research (Weinstein, 2018, 539-540).

Occupancy

In the domain of animal conservation, occupancy states whether a species inhabits a location or not. With it, regional distributions of species are defined. A measurement taken to model occupancy is called presence-absence. Presence-absence simply distinguishes between the presence of an animal signal in a sample (e.g., image) and its absence. The issue with presence-absence is the bias of assuming absence at a location when no signal is detected. (MacKenzie, 2009, 849). When using camera traps for determining occupancy, it might be necessary to sample over an extended time to determine absence with a higher confidence (Kim et al., 2025, 5).

Density

Density refers to population density. That is the number of individuals of one species that inhabit some area. For computer vision, density is a counting task that yields the number of animals present in a sample (video or image) (Weinstein, 2018, 537).

Relative Abundance

The ratio of one sighted species to any other regarded species is known as relative abundance. It is calculated to help compare population sizes in the area where the sampling is performed. Typically, researchers manually sample by counting animal signals correlated to population sizes, such as bird nests or scats, along trails or regions of special interest. When researchers choose camera traps as a sampling method, they deploy them for a certain period and group the collected images by species. Any measurement of relative abundance is biased by its sampling method and chosen animal signals. Still, a good measure of relative abundance over time can give valuable insights into the development of populations over time (Martin-Garcia et al., 2022, 108778).

Behavior and Health

The most complex task discussed in the field of wildlife monitoring is the detection of behavior or the evaluation of the health status of individual animals. In some papers, these are also called description tasks (Weinstein, 2018, 534). In contrast to the other goals, it is the hardest to achieve, the most diverse, and the most difficult to compare. Behavior analysis is very useful for video data, which researchers need to annotate with the action an animal is performing. The basis of behavior and health analysis are the other mentioned tasks, like identity and presence-absence. It is essential to locate the animal and its characteristic features in an image or video frame (using boxes) before applying the models to predict the actions (Vogg et al., 2025, 1154-1156). As of now, research on behavior and health analysis is focused on farm animals. Computer vision systems detecting early signals of illness, pests, and harmful behavior provide great value to large-scale agriculture operations (Bhuiyan & Wree, 2023, 126602-126603). However, this does not mean that the same methods can't provide value to wildlife conservation programs.

Multi-Species Models

In the academic community, there is a significant imbalance in interest regarding the species of interest for computer vision tasks. As a result, there exist numerous models designed for mammals and birds, especially farmland and domestic animals, and far fewer models for reptiles and amphibians. Additionally, the models typically are heavily specialized for a small set of species or a family of species. (Burton et al., 2015, 678-679) One reason for this is limited training data. It is challenging to generalize models applicable to wildlife species if few or no images are available.

Algorithms for Wildlife Image Processing

Wildlife computer vision systems are assemblies of many algorithms executed in a pipeline to improve the desired outcome. Depending on the setup, it can include hardware interactions like image or video capturing performed manually or based on passive infrared triggers. Even when excluding the capture process, processing images consists of multiple steps that transform images/frames into a reduced feature space (no longer a matrix of pixels) that is suitable as input for machine learning models. The following methods represent common processing steps for automated wildlife research (Parham et al., 2018, 1075-1076).

Normalization

Biases in the collected data pose a great risk for model training. To reduce the chance of a model making predictions based on characteristics linked to the image-gathering method and environmental influences like lighting conditions, data is normalized and cleaned. Images are resampled to have the same size, mapped to the same color channels (e.g., RGB, grayscale), and pixel intensities scaled according to a global mean. Additionally, information can be removed on purpose by blurring the image, selecting subregions, or even modifying random parts. These steps force the following algorithms to learn based on the important characteristics of the animal and not on noise and other factors. For imbalanced datasets, instances of oversampled classes can be removed, or instances of undersampled classes can be reused (Li et al., 2021, 13-14).

Image Segmentation

The first real ML processing step is image segmentation. These models can detect regions of interest and already take into account patterns and structures present in the image.

Background Removal

Background removal or subtraction defines the process of removing all pixels that do not belong to an animal. This is the vegetation, sky, clouds, and shade. An effective method for static camera setups is subtracting an image that is believed to show only the background from the captured image. As a result, all background pixels are black, and only the animal is visible. Non-static setups require other methods like convolutional neural networks trained specifically for image segmentation (Weinstein, 2018, 537-538). Sometimes simple methods like the Canny algorithm and Otsu segmentation can even lead to great results, separating an image into foreground and background based on the histogram of the image (Li et al., 2021, 15).

Object Detection

Object detection is an alternative method that aims to separate the region with the animal from the rest. Specialized families of deep neural networks, among the most prominent ones are Region Proposal Networks (RPN) and YOLO (You Look Only Once) networks, find boxes within an image that have a high likelihood of containing an object or animal (Schneider et al., 2018, 322).

Feature Extraction

Up to this point, the processing uses pixels as both input and output. Since ML algorithms like classifiers or clustering algorithms do not work on the pixel level, meaningful features are extracted first. Convolutional neural networks (e.g., AlexNet, VGG, GoogLeNet, and ResNet) are incredibly popular for feature extraction, and the variety of layer architectures proposed in papers is great. (Schneider et al., 2018, 322) These networks can extract edges, shapes, and structures of the image, reducing the feature space drastically. After extracting the features, the visual representation of the image is lost and compressed into a small set of characteristics (Valletta et al., 2017, 205).

Wildlife Model Training and Inference

The final step in an animal computer vision pipeline is the model training. This stage is specifically tailored to achieve the overall goal of the entire system. It also allows for the evaluation of its performance.

Classification

Classification in the context of wildlife images can be many different things. The classic task is species classification, but as mentioned above, classifying behavior, health status, and even emotions is possible (Valletta et al., 2017, 204). With the previously extracted features, simple fully connected neural networks or traditional classifiers (SVM, Random Forest, etc.) are employed to discriminate between a predefined set of classes (Alharbi et al., 2019, 02033).

Clustering

In contrast to classification, which estimates the likelihood of a sample being of one class, clustering uses the distances of the samples in the feature space to build groups of images. These groups could be individuals, species, behavioral states of animals, and much more. The clustering algorithm can be tuned in terms of distance measures, number of clusters to create, and methods for clustering, but they are, per definition, unsupervised. Because of this, the meaning of the clusters is not as clear as the output of a supervised machine learning model (Valletta et al., 2017, 205).

Evaluation of Animal Identification Systems

To compare proposed systems for wildlife computer vision, it is imperative to pick studies dealing with the same task. Here are three studies that aim to automatically identify animal species in camera trap images. The studies use different preprocessing, algorithms, and models, demonstrating the diversity of these systems.

Study 1: Manohar et al. (2016)

Manohar et al. compare two proposed systems, one supervised and one unsupervised. Rather than just trying to create the best model for their data, they put a lot of effort into describing the design differences of unsupervised and supervised machine learning models for animal computer vision.

Data

Sadly, the data is not a standard dataset but a custom one created by the researcher. The dataset contains 2000 images from the researchers and associated organizations as well as public images from the internet. There are 20 different animal species present in the data. The authors don't list the categories, but the figures suggest they're all mammals.

Model

Both their models require the images to have a removed background (image segmentation). The images are then transformed into Gabor features, which are frequency-domain features that describe patterns in the form of wavelets.

The supervised model utilizes features gathered from linear discriminant analysis (LDA). The ruleset/model for the inference is a symbolic classifier.

The unsupervised model works with the principal components of the Gabor features (by applying PCA on them). These components are clustered using the k-means algorithm.

Result

With a train-test split of 70–30 percent and applying the corresponding dimensionality reduction, the authors report a test accuracy of 79.54% for the supervised model and 75.46% for the unsupervised model.

Study 2: Nguyen et al. (2017)

Nguyen et al. use a supervised learning approach trained with large amounts of hand-annotated pictures published by wildlife conservation projects. They demonstrate how the tedious work of annotating camera trap images can be supported by machine learning models.

Data

The Wildlife Spotter dataset contains about 100,000 images of mammals, birds, and reptiles from south-central Victoria, Australia. After some removal of ambiguously labeled images, the researchers end up with a slightly smaller dataset that contains 18 classes/species.

Model

The paper mentions the need for multiple models with different goals: determining presence-absence and species identification. For both tasks, they employ popular deep convolutional neural networks (Lite AlexNet, VGG16, and ResNet-50).

Result

The proposed species identification models have high accuracies for all variants of the convolutional neural networks. Lite AlexNet achieves 87.80%, VGG16 88.03%, and ResNet-50 87.97% test accuracy.

Study 3: Carl et al. (2020)

Carl et al. investigate the capabilities of pretrained deep convolutional neural networks publicly available on the internet when dealing with small amounts of images of European wildlife.

Data

One type of camera trap in the Hainich National Park in Germany provided the images. The researchers constructed their testing data by randomly selecting 10 pictures for each species and adding 10 pictures with no animal present. The total size of the dataset is only 110 pictures.

Model

Unlike the other studies, Carl et al. don’t do any training or model development themselves. They feed the test images into a publicly available FasterRCNN+InceptionResNet V2 model developed by Google. FasterRCNN is performing the object detection, which improves the accuracy of the InceptionResNet V2 model. Both components are convolutional neural networks.

Result

For the given data, an accuracy of 93% is achieved. Due to the small dataset, the accuracy for individual classes varies significantly. Squirrels are detected with an accuracy of 50%, and many other species with an accuracy of 100%.

Summary

The review of many papers and detailed comparison of three individual studies showed the success of convolutional neural networks in the field of wildlife computer vision.

Models like FasterRCNN are widely available and reach accuracies above 90% in many use cases. Transformers are gaining interest and have the potential to enhance the accuracy of existing methods even more (Schneider et al., 2024, 1179). As a consequence, many use cases do not require the training of custom models anymore because publicly available models have excellent object detection and classification capabilities for a large number of animals. A key issue in the reviewed studies is the small number of species (20, 18, and 10, respectively) investigated. Most models are static in terms of species they can discriminate, which makes it difficult to include species with little training data available.

Future Work

The vast majority of research uses supervised machine learning with cleaned and annotated datasets. There is no clear solution for automated and non-invasive wildlife monitoring of rare species and even whole classes of animals like amphibians and reptiles. Unsupervised machine learning could be one option to allow faster labeling of camera trap images. A human-in-the-loop system could facilitate the processing of the image, classifying whole clusters of similar images rather than one by one. Such a system would enable research on more species at risk of extinction.

Still, these models alone, performing the identification, counting, and behavioral analysis tasks, represent only one part of the efforts required to monitor wildlife in real time. Further work on IoT architectures integrating smart camera traps with wildlife monitoring platforms is necessary to support the daily work of conservationists and public authorities.

References

Alharbi, A., Alharbi, A., & Kamioka, E. (2019). Animal species classification using machine learning techniques. *MATEC Web Conf.*, *277*, 02033. 10.1051/matecconf/201927702033

Bhuiyan, M. R., & Wree, P. (2023). Animal Behavior for Chicken Identification and Monitoring the Health Condition Using Computer Vision: A Systematic Review. *IEEE Access*, *11*, 126601-126610. 10.1109/ACCESS.2023.3331092

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675-685. 10.1111/1365-2664.12432

Carl, C., Schönfeld, F., Profft, I., Klamm, A., & Landgraf, D. (2020, 07 14). Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *European Journal of Wildlife Research*, *66*(4), 62-66. 10.1007/s10344-020-01404-y

Harie, Y., Neupane, S. B., Gautam, B. P., & Norio, S. (2021). Augmented Triplet Network for Individual Organism and Unique Object Classification for Reliable Monitoring of Ezoshika Deer. In *2021 Ninth International Symposium on Computing and Networking Workshops (CANDARW)* (pp. 197-200). IEEE. 10.1109/CANDARW53999.2021.00039

Kim, S.-H., Dhakal, T., Cho, K. H., Kim, T., Woo, S., Kim, J.-Y., Lee, D., & Jang, G.-S. (2025, 05). Occupancy of roe deer, water deer, and wild boar in wind farm-integrated forest ecosystems: A case study in Korea. *Ecosphere*, *16*. 10.1002/ecs2.70258

Li, G., Huang, Y., Chen, Z., Chesser, G. D., Purswell, J. L., Linhoss, J., & Zhao, Y. (2021). Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: A Review. *Sensors*, *21*(4). 10.3390/s21041492

MacKenzie, D. (2009, 09). What are the issues with Presence-Absence data for wildlife managers? *Journal of Wildlife Management*, *69*, 849-860. 10.2193/0022-541X(2005)069[0849:WATIWP]2.0.CO;2

Manohar, N., Kumar, Y. H. S., & Kumar, G. H. (2016). Supervised and unsupervised learning in animal classification. *2016 International Conference on Advances in Computing, Communications and Informatics*, 156-161. 10.1109/ICACCI.2016.7732040

Martin-Garcia, S., Rodríguez-Recio, M., Peragón, I., Bueno, I., & Virgós, E. (2022). Comparing relative abundance models from different indices, a study case on the red fox. *Ecological Indicators*, *137*, 108778. https://doi.org/10.1016/j.ecolind.2022.108778

Neupane, S., Sato, K., & Gautam, B. (2022, 11). A LITERATURE REVIEW OF COMPUTER VISION TECHNIQUES IN WILDLIFE MONITORING. *International Journal of Scientific and Research Publications (IJSRP)*, *16*, 282-295. https://www.researchgate.net/publication/366005635_A_LITERATURE_REVIEW_OF_COMPUTER_VISION_TECHNIQUES_IN_WILDLIFE_MONITORING

Nguyen, H., Maclagan, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E. G., & Phung, D. (2017). Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 40-49. 10.1109/DSAA.2017.31

Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., & Berger-Wolf, T. (2018). An Animal Detection Pipeline for Identification. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1075-1083. 10.1109/WACV.2018.00123

Schneider, D., Lindner, K., Vogelbacher, M., Bellafkir, H., Farwig, N., & Freisleben, B. (2024). Recognition of European mammals and birds in camera trap images using deep neural networks. *IET Computer Vision*, *18*(8), 1162-1192. 10.1049/cvi2.12294

Schneider, S., Taylor, G. W., & Kremer, S. (2018). Deep Learning Object Detection Methods for Ecological Camera Trap Data. *2018 15th Conference on Computer and Robot Vision (CRV)*, 321-328. 10.1109/CRV.2018.00052

Trevathan, J., Tan, W. L., Xing, W., Holzner, D., Kerlin, D., Zhou, J., & Castley, G. (2025). A computer vision enhanced IoT system for koala monitoring and recognition. *Internet of Things*, *29*, 101474. 10.1016/j.iot.2024.101474

Trolliet, F., Huynen, M.-C., Vermeulen, C., & Hambuckers, A. (2014, 01). Use of camera traps for wildlife studies. A review. *Biology Agriculture Science Environnement*, *18*, 446-454.

https://www.researchgate.net/publication/266381944_Use_of_camera_traps_for_wildlife_studies_A_review

Tulasi, D., Granados, A., Gunawardane, P., Kashyap, A., McDonald, Z., & Thulasidasan, S. (2023). Smart Camera Traps: Enabling Energy-Efficient Edge-AI for Remote Monitoring of Wildlife. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on AI-driven Spatio-temporal Data Analysis for Wildlife Conservation* (pp. 9–16). Association for Computing Machinery. 10.1145/3615893.3628760

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203-220. 10.1016/j.anbehav.2016.12.005

Vogg, R., Lüddecke, T., Henrich, J., Dey, S., & Nuske, M. (2025, 06 01). Computer vision for primate behavior analysis in the wild. *Nature Methods*, *22*(6), 1154-1166. 10.1038/s41592-025-02653-y

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*(3), 533-545. 10.1111/1365-2656.12780

Zemanova, M. A. (2020). Towards more compassionate wildlife research through the 3Rs principles: moving from invasive to non-invasive methods. *Wildlife Biology*, *2020*(1), wlb.00607. https://doi.org/10.2981/wlb.00607

Appendix

Usage of Generative AI

This report was written independently without the use of generative AI. QuillBot was used for grammar checking and rephrasing.