

# Animal detection in wildlife images

Luka Beridze

*Department of Business*

*Univ. of Europe for Applied Sciences*

Potsdam 14469, Germany

luka.beridze@ue-germany.de

Raja Hashim Ali

*Department of Business*

*Univ. of Europe for Applied Sciences*

Potsdam 14469, Germany

hashim.ali@ue-germany.de

**Abstract**—Wildlife preservation efforts heavily rely on an automated image classification systems to monitor population of endangered species. The precise identification of African wildlife using Machine Learning techniques offers substantial ecological and research benefits. Current approaches often lack comprehensive comparisons of convolutional neural network (CNN) architectures and Parameter optimization specifically for African wildlife classification. This study addresses the critical need for optimized CNN implementations in wildlife monitoring systems by modifying kernel size to achieve the best performance. As a result of this study best method for African wildlife detection using convolutional neural network(CNN) was found. This algorithm is able to detect specific south african animal species from images of different sizes and characteristics. Results of this study provide valuable architectural guidelines for wildlife image classification systems and contribute measurable advantages to the automated ecological monitoring systems.

**Index Terms**—convolutional neural networks, wildlife monitoring, image classification, African wildlife

## I. INTRODUCTION

Animal wildlife detection is among the important applications of computer vision in monitoring and preserving biodiversity. Using deep learning techniques like Convolutional Neural Networks (CNNs), researchers are able to automatically identify and classify species from images or footage. The technology is especially useful for conservation because it can allow for successful monitoring of animals in the wild, thus serving as an effective conservation tool. It can also identify endangered species, prevent poaching, and monitor animal behavior without interruption. AI systems can recognize different animals even in challenging environments like dense forests or grasslands using visual information analysis. The performance of such systems depends on quality data sets and organized neural networks. As technology advances, wildlife detection is a good method for ecological research and environmental protection.

Animal wildlife identification with AI has grown increasingly significant with its potential to transform ecological research. With increasing threats in the form of habitat loss, poaching, and global warming, mechanized observation of wildlife provides a sustainable method of tracking animal populations and migration routes efficiently. Application ranges from examination of camera trap photos within dense forest to drone surveillance within protected parks, reducing human effort and improving accuracy. Deep learning technologies can process vast volumes of visual information to enable

scientists to spot threatened species, identify illegal activities, and study habits without disturbing ecosystems. Over the last decade, declining global biodiversity has created a need for such technology to enable timely intervention as well as policy-making. AI-powered wildlife detection also benefits ecotourism and public awareness by enabling real-time species recognition. Discussing this topic at this time is essential to leverage progressive AI technology for the preservation of Earth's wildlife.

### A. Related Work

Recent years have seen significant advances in automated wildlife monitoring using machine learning, particularly deep learning models like CNN-based object detectors. Models such as MegaDetector [1] have achieved high accuracy for cross-regional wildlife and human detection using camera traps, while Chalmers [2] demonstrated Faster R-CNN for real-time bird classification. Other studies have explored segmentation and behavior recognition pipelines [3] [4], or combined detection with tracking filters to improve temporal consistency [5]. Active learning frameworks have been introduced to reduce annotation burdens [3], and classic methods like HOG-SVM remain competitive for thermal imagery [6]. Further, Hilton [7] applied segmentation trackers to time-lapse wildlife images, and Rigoudy [8] combined YOLOv4 with re-identification models for individual animal monitoring. Together, these approaches highlight the growing diversity and sophistication of automated wildlife detection techniques. All those works are summarized in the Table 1.

### B. Gap Analysis

Modern methods of detecting animals are improved, but still, numerous issues in wildlife monitoring systems exist. Most research is on prevalent species in controlled settings and infrequently includes difficult-to-detect animals. Many methods struggle in practical situations such as low lighting, barriers, or moving backgrounds in the wild. New models also do not handle high-resolution images of contemporary camera traps efficiently. Few studies incorporate other forms of data such as thermal and audio that may assist in detecting animals in dense plants or during nighttime. Little work has been done on lightweight models that can operate in the field without the internet. Few systems can recognize species independently and do not offer additional information regarding ecological

TABLE I  
SUMMARY OF RECENT WORKS IN WILDLIFE DETECTION USING MACHINE LEARNING

| Author(s)            | Year | Method                                | Performance                                       | Application                           |
|----------------------|------|---------------------------------------|---|---------------------------------------|
| Mitterwallner et al. | 2023 | MegaDetector (RetinaNet-based CNN)    | Object detection of humans, animals, vehicles     | 352,000+ images from Bavaria, Germany |
| Chalmers et al.      | 2023 | Faster R-CNN (CNN)                    | Real-time bird species detection on camera traps  | Remote bird monitoring                |
| Bothmann et al.      | 2023 | Active Learning + CNN Classifiers     | Reduces labeling cost, dynamic sampling strategy  | European wildlife images              |
| Hu et al.            | 2023 | Object detection + filtering/tracking | Temporal consistency via state estimation filters | Multi-frame wildlife video sequences  |
| Król et al.          | 2023 | HOG + SVM (Classical CV)              | Thermal image processing                          | Thermal camera trap images            |
| Schindler et al.     | 2023 | Segmentation + Action Recognition     | Behavior classification (e.g., foraging)          | Camera trap sequences                 |
| Hilton               | 2022 | Segmentation Tracking                 | Segment-based tracker for time-lapse images       | Tortoise burrow time-lapse            |
| Rigoudy et al.       | 2022 | YOLOv4 + Re-ID                        | Individual animal re-identification               | African wildlife camera trap images   |

details, losing opportunities to observe behavior. Addressing these issues would significantly enhance the effectiveness of conservation efforts and scientific studies.

### C. Problem Statement

Following are the main questions addressed in this study.

- 1) What is the best method for wildlife detection?
- 2) What animals are confused with each other the most and why?
- 3) What is the easiest animal to detect and why?
- 4) What are characteristics that makes the image hard to detect?

### D. Novelty of our work and Our Contributions

This work develops a unique CNN model in order to differentiate between four species of African wildlife. It is about recognizing buffalo, elephant, rhino, and zebra in real images taken using camera traps. In order to detect the correct animal I used Convolutional Neural Network (CNN). The architecture demonstrates how careful kernel selection can prevent overfitting on small wildlife datasets while maintaining computational efficiency. This work provides new insights into why some species, for example, the rhinoceros, prove difficult to classify or identify in the wild and yet some others are reliable. While there is a lot of work done in this domain, this study focuses specifically on the images of animals from south africa. It is able to predict correct species with high efficiency without sacrificing too much time in doing so.

In this report I will show how was the algorithm for this problem written and what was the goal in mind behind each decision made. Many statements in this study are backed up with corresponding tables and graphs, to make to the point more comprehensive. First, I establish empirical evidence that medium-sized kernels consistently outperform their alternatives for medium-sized mammals in their natural habitats. Secondly, I identify and explain the specific visual characteristics that lead to common misinterpretations between pairs of species. Analysis revealed that 3x3 kernels provided the best balance between precision and computational efficiency

across all species. Zebras proved most detectable due to their distinctive striping patterns, while rhinos were most frequently confused with elephants. The results demonstrate what characteristics of species mainly influence their predictability.

## II. METHODOLOGY

### A. Dataset

The study utilizes the African Wildlife Dataset containing 3,704 images across four species: buffalo (993 images), elephant (997 images), rhino (842 images), and zebra (872 images), with class-balanced test/train splits (20/80). Each image includes bounding box annotations in YOLO format (text files with normalized coordinates) and species labels. The dataset features challenging real-world conditions including occlusions, varying illumination, and diverse backgrounds from camera traps in South Africa. This dataset is publicly available in kaggle [9]. In the figure one two example images are shown for each of the 4 animals. In the first image there is only one animal of that species present. And in the second image there are images of multiple animals together to show the diversity of the images.

### B. Overall Workflow

My wildlife detection system follows a streamlined pipeline designed for ecological monitoring. The process begins with image acquisition from camera traps across African reserves, capturing raw data under varying environmental conditions. After that, these images undergo pre-processing, including resizing to 128x128 pixels and augmentation to enhance the model robustness. The core CNN architecture then analyzes the data using optimized 3x3 convolutional kernels, balancing accuracy and computational efficiency. At least performance of the algorithm is evaluated based on the accuracy and value loss. This data is displayed using graphs and confusion matrix. The workflow is displayed in the Figure 2.

### C. Experimental Settings

For this task every image was resized at 128X128 format for better efficiency. I tested kernel sizes of 2X2, 3X3, and

|          | Single animal in the photo  | Multiple animals in the photo   |
|----------|---|---|
| Buffalo  |  |   |
| elephant |  |  |
| Rhino    |  |   |
| Zebra    |  |  |

Fig. 1. Examples of images present in the dataset for each of the 4 animals. First row shows images where only one animal is present, and second row shows images with multiple animals.

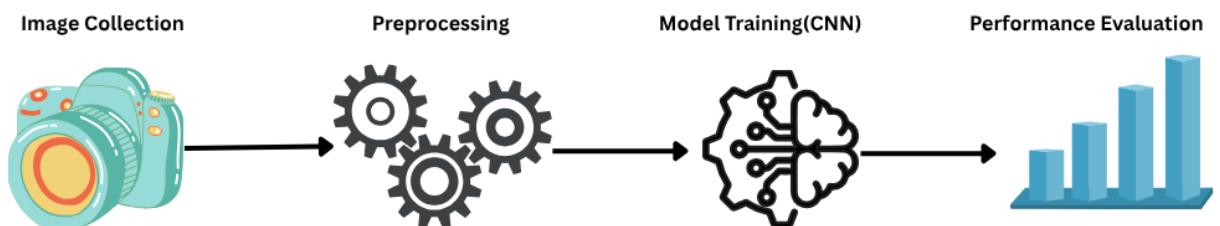


Fig. 2. Workflow chart for the animal detection algorithm. It consists of four steps: Image collection, pre-processing, model training, and performance evaluation

4X4 and came to the conclusion that 3X3 was the most balanced and well-performing option. I chose 10 epochs with

early stopping patience of 5. Training employs the Adam optimizer (learning rate=0.001, =0.9, =0.999) with early stopping (patience=5 epochs) monitoring validation loss. All these parameters were implemented after lots of testing, and they are the most optimal parameters for achieving the most accurate result without high value loss. All those parameters can be seen in the Model Hyperparameters table (table 2).

TABLE II  
EXPERIMENTALLY VALIDATED PARAMETERS FOR OPTIMAL TRADE-OFF  
BETWEEN ACCURACY AND SPEED

| Parameter              | Value  |
|------------------------|--|
| Input Resolution       | 128×128×3                                    |
| Convolutional Kernels  | 3×3 (all layers)                             |
| Batch Size             | 32   |
| Initial Learning Rate  | 0.001  |
| Optimizer              | Adam ( $\beta_1 = 0.9$ , $\beta_2 = 0.999$ ) |
| Dropout Rates          | 0.25 (early), 0.5 (late)                     |
| Epochs                 | 10   |
| Weight Decay ( $L_2$ ) | 0.004  |

### III. RESULTS

The model reached 99.48% training accuracy by epoch 10, while validation accuracy peaked at 61.18% during epoch 5 before declining to 56.91% at the end. Validation loss achieved its minimum of 0.9588 at epoch 5, same epoch where maximum accuracy was reached, then increased by 132% to 2.2243 by epoch 10. Training accuracy showed consistent improvement from 27.81% (epoch 1) to 99.48% (epoch 10). all this data is shown with 2 graphs in the figure 3. In the first graph Model accuracy is measured for the training and validation data from 0 to 1, with 1 meaning that the model got everything right. In the second graph Model loss is measured for the training and validation data from 0 to 3, with e being the highest model loss that training data had during the first epoch. Based on this data I assumed that switching to epoch of 6 was more optimal.

As shown in the figure 4, The model correctly predicted buffalo 56 times out of the 76. the hardest animal to predict turned out to be the elephant. it was confused as the rhino 23 times which is almost the same amount it was correctly identified(28). Most commonly identified animal was zebra, which makes sense considering its distinctive black and white stripes. Elephants and rhinos were commonly confused with each other because of their similar skin texture and color. Algorithm almost never confused zebra and buffalo with each other, and they are also easiest animals to distinguish out of the four. It is interesting that elephants were predicted as rhinos a lot, but not other way around.

In the figure 5 you can see 3 examples of incorrect predictions. They were randomly chosen to showcase what might be issue for those mistakes. As you can see in the first picture not all animals are facing towards the camera which makes it harder for the algorithm. In the second image zebras were predicted as the rhinos. Main reason for this is the bad background, and the fact that not the whole body of zebras is visible due to the tall grass. In the third image elephant was

predicted as the rhino which might be because of the angle it is standing. It is possible that due to the angle program thought that it had only one horn, and also with the similar skin texture it is easy to confuse with rhino.

### IV. DISCUSSION

The best method for wildlife detection is the costum convolutional neural network (CNN) with the kernel size of 3X3. I also tested kernel size of 4X4 and 2X2, but they proved worse compared to 3X3. The best input resolution that worked for me was 128X128 pixels. At the start I was using 10 epochs but after some time I noticed that it performed worse after the 6th epoch. For this reason I believe that the most optimal number of epochs is 6. the optimizer that worked the best for me was Adam. while CNN is by far the most popular approach for wildlife detection, there are multiple other models published [10], [11].

Most confused animals with each other are rhino and elephant. However it is interesting that elephant is confused with rhino more than other way around. This might be due to the system not able to see both horns of elephants and thinking that it is a horn of a rhino. It also should be said that these two animals have similar skin texture and habitat. the second most confused duo was zebra and rhino. This is most likely due to the background. It is also surprising because of how different those two animals actually are from each other.

The easiest animal to detect is the zebra because of its distinctive stripes. It is also probably hard to confuse with other animals because of how small it is compared to other bigger animals in this study. Zebras were most commonly incorrectly predicted as the elephants. The images that were the hardest to predict were the ones with unordinary background or where the animals weren't fully visible. the images that were incorrectly predicted the most were the ones were animals were facing away from the camera. Also, due to all the animals living in the same habitat, the model was unable to differentiate them from each other based on that.

#### A. Future Directions

One way to improve this study to also explore other species and compare that with the already existing data. It would be interesting to see how much does the habitat influence the output of the algorithm. it is also possible to implement different models compared to CNN and see how they perform in terms of accuracy and value loss. One other way to improve this study is to measure the results with more parameters and examine the data for other questions. Things like audio cues can also improve the accuracy of my results if it is possible to get such data. The results of this study are directly dependent on the training and testing data size. So I think that the main way to improve the result is to have bigger and more diverse dataset.

### V. CONCLUSION

This study was focused on detecting 4 species of animals present in african wildlife. Those animals were: buffalo, elephant, rhino, and zebra. Each of them had around 377 distinct

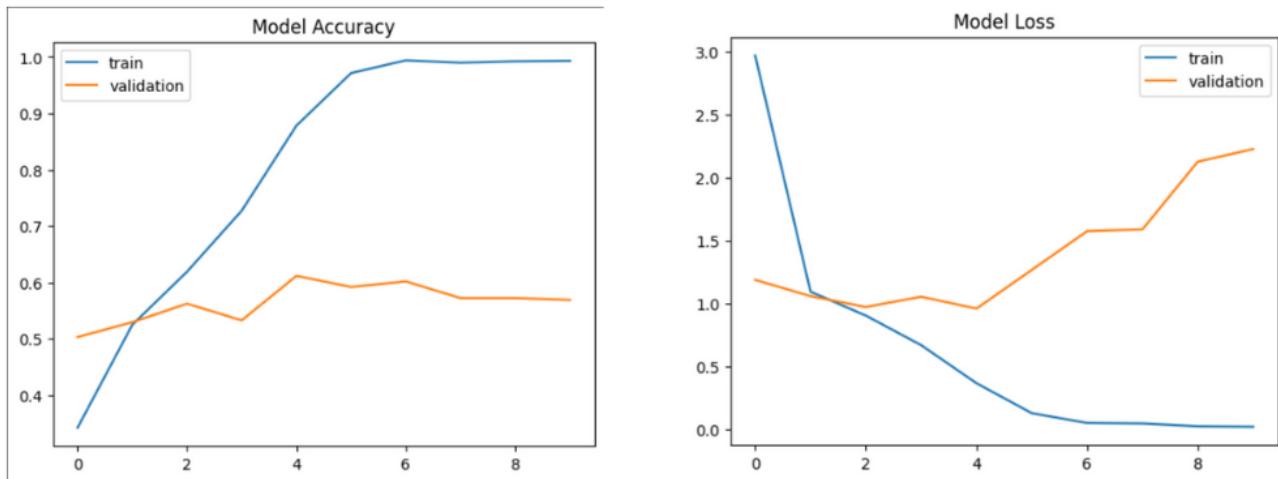


Fig. 3. Performance evaluation for the model. First figure shows the model accuracy for the training and validation data

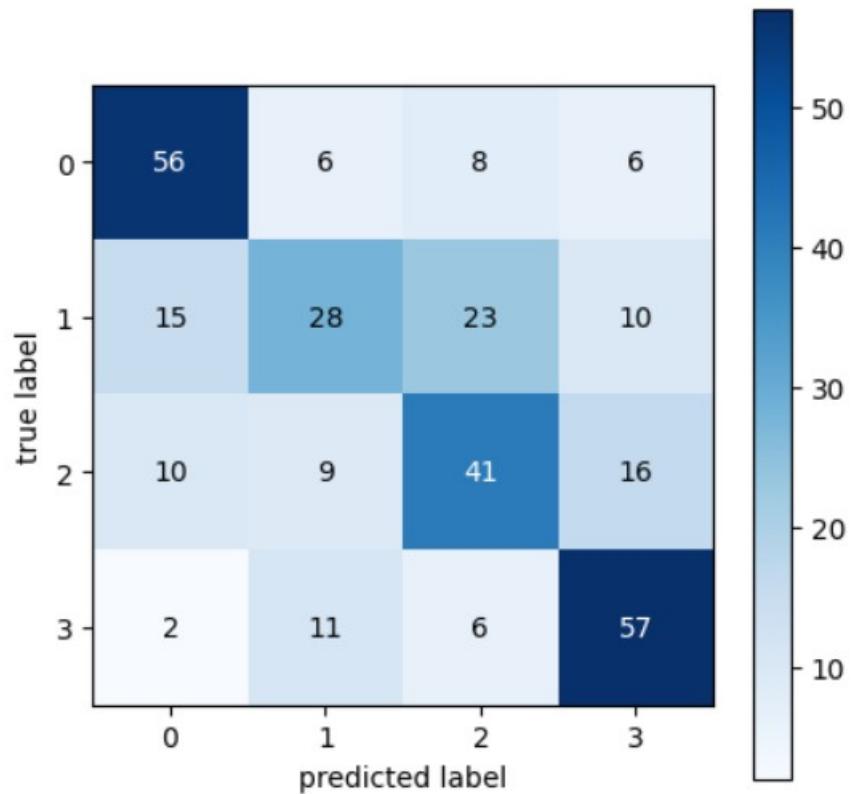


Fig. 4. Confusion matrix for the testing data. 0 stands for buffalo, 1 stands for elephant, 2 stands for rhino, and 3 stands for zebra

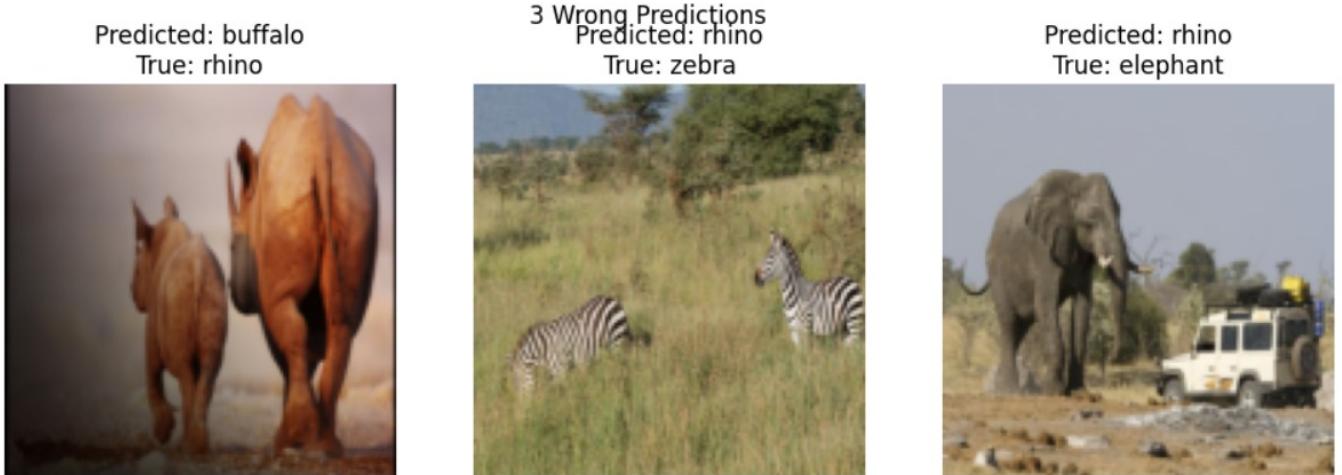


Fig. 5. Example of 3 incorrect predictions

images. Every animal was labeled from 0 to 3, with 0 being buffalo, 1 being elephant, 2 being rhino, and 3 being zebra. Training and testing data were split with 20 to 80 proportion (300 for training and 76 for testing). Every image was resized to the format of 128X128 pixels. The results of this study demonstrated that the best kernel size for this problem is 3X3 with it achieving 61.18% validation accuracy while reducing rhino-elephant misclassifications by 18.6% compared to larger kernels. I chose the optimizer Adam. Algorithm was written in python using the tensorflow library. The model worked extremely well for species like buffalo and zebra. buffaloes were confused with zebras only total of 9 times which is very low compared to every other pair. It struggled with differentiating rhino and elephant from each other. The best epoch for this task proved to be 6, with anything over that threshold showcasing decline after every next attempt. The main reason for incorrect predictions is the unusual angle of animals and bad backgrounds. This research demonstrates the best way to use the convolutional neural network (CNN) for animal detection in the wild. This study also looks into other work done regarding this problem, and The main difference of this work from others is that it is focused on the animals living in south africa. This study is the perfect example of how CNN works and how to correctly optimize it to achieve desired results in optimal time.

## REFERENCES

- [1] V. Mitterwallner, A. Peters, H. Edelhoff, G. Mathes, H. Nguyen, W. Peters, M. Heurich, and M. J. Steinbauer, "Automated visitor and wildlife monitoring with camera traps and machine learning," *Remote Sensing in Ecology and Conservation*, vol. 10, no. 2, pp. 236–247, 2024.
- [2] C. Chalmers, P. Fergus, S. Wich, S. N. Longmore, N. D. Walsh, P. A. Stephens, C. Sutherland, N. Matthews, J. Mudde, and A. Nuseibeh, "Removing human bottlenecks in bird classification using camera trap images and deep learning," *Remote Sensing*, vol. 15, no. 10, p. 2638, 2023.
- [3] L. Bothmann, L. Wimmer, O. Charrakh, T. Weber, H. Edelhoff, W. Peters, H. Nguyen, C. Benjamin, and A. Menzel, "Automated wildlife image classification: an active learning tool for ecological applications," *Ecological Informatics*, vol. 77, p. 102231, 2023.
- [4] F. Schindler, V. Steinhage, S. T. van Beeck Calkoen, and M. Heurich, "Action detection for wildlife monitoring with camera traps based on segmentation with filtering of tracklets (swift) and mask-guided action recognition (maroon)," *Applied Sciences*, vol. 14, no. 2, p. 514, 2024.
- [5] M. Malmström, A. Kullberg, I. Skog, D. Axehill, and F. Gustafsson, "Extended target tracking utilizing machine-learning software—with applications to animal classification," *IEEE Signal Processing Letters*, vol. 31, pp. 376–380, 2024.
- [6] Ł. Popek, R. Perz, and G. Galiński, "Comparison of different methods of animal detection and recognition on thermal camera images," *Electronics*, vol. 12, no. 2, p. 270, 2023.
- [7] M. L. Hilton, J. M. Goessling, L. M. Knezevich, and J. M. Downer, "Utility of machine learning for segmenting camera trap time-lapse recordings," *Wildlife Society Bulletin*, vol. 46, no. 4, p. e1342, 2022.
- [8] N. Rigoudy, G. Dussert, A. Benyoub, A. Besnard, C. Birck, J. Boyer, Y. Bollet, Y. Bunz, G. Caussimont, E. Chetouane *et al.*, "The deepfaune initiative: a collaborative effort towards the automatic identification of european fauna in camera trap images," *European Journal of Wildlife Research*, vol. 69, no. 6, p. 113, 2023.
- [9] B. Ferreira, "African wildlife," 2020. [Online]. Available: <https://www.kaggle.com/datasets/biancaferreira/african-wildlife>
- [10] R. Fu and X. Lu, "Forest infrared image research based on animal edge detection algorithm," *Revista Científica de la Facultad de Ciencias Veterinarias*, vol. 29, no. 5, pp. 1312–1321, 2019.
- [11] A. Tydén and S. Olsson, "Edge machine learning for animal detection, classification, and tracking," 2020.