# A Literature Survey on Wildlife Camera Trap Image processing using Machine Learning Techniques

Shreyas Sreedhar
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
shreyas.sreedhar@gmail.com

Sandesh S
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
hssandesh111@gmail.com

Prarthana P
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
prarthanaprasad001@gmail.com

N Karthik Pranav
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
karthikpranav1999@gmail.com

Dr. Prabhanjan S
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
hod.cse@ jyothyit.ac.in

Srinidhi K
Department of Computer Science and Engineering, Jyothy Institute of Technology, Bengaluru, India,
srinidhi.kulkarni@jyothyit.ac.in

***Abstract: Motion Triggered Wildlife Camera traps are rapidly being used to remotely track animals and help perform different ecological studies across the globe. The system captures animal visuals that enable the forest department of the respective country to keep track of critically endangered species, record their actions, research environmental changes in order to generate methods This piece of equipment is typically deployed within the forest area in large numbers, resulting in millions of recorded images and videos. It normally takes days, if not months, to go through the dataset completely and, identify the captured animals. In this paper, we study some classifiers of the fauna image that use the convolution neural network to process and identify the wildlife captured by these camera traps.***

***Keywords: Wildlife Image Camera Traps; Convolution Neural Network; Object Detection; Image Classification; Machine Learning***

## I. INTRODUCTION

The first-time wildlife camera traps were used for behavioural and ecological studies were in the early 1990s to monitor Tigers in Nagarhole National Park, Karnataka, India by K. Ullas Karanth [1]. Since then, wildlife camera traps have been widely used to objectively estimate parameters such as size, density, the survival of endangered species and other secretive animals' species [1]. The security of endangered species needs constant monitoring and up to date information about their presence in the habitat locations and change in their behaviours with minimal human intervention.

At present, the common constraint of this method of monitoring the wildlife is the accumulation of huge amounts of data from the camera trap which is usually sorted out by a manual observer [2]. The usage of manual labour to classify the images is a very tedious process and typically takes months to gather to complete leading to data management issues [3].In order to solve this, machine learning species recognition approach is adopted to reduce the manual labour process and make the workflow efficient.

Inspired by the rise of convolution neural networks in the field of machine learning, in this paper, we investigate the possibilities of using various machine learning based algorithms and CNN models that make the process efficient and we compare the advantages and disadvantages of each model. Second, to facilitate the use of this machine learning model approach for classification of data which is often limited to computer scientists than Ecologists and researchers, we review some of user-friendly applications that runs locally or on cloud-based provider which help the user to detect and classify animals. Furthermore, we propose a solution to automate the process of filtering and sorting of image captured through the wildlife camera traps.

This paper is organized as follows. Section II discusses about various Convolution Neural Networks that are being used. Section III tables the existing animal detection and classification models listing its advantages and disadvantages. Section V concludes this paper.

## II. BACKGROUND

Neural Networks gained popularity after three researchers [4][5][6] discovered very effective techniques based on the early model Neo cognition model by Fukushima [7] in 1985 and 1986. These 3 discoveries used backpropagation to train the neural network but failed to satisfy the performance compared to other machine learning algorithms, making late 1990's a dry period for the growth of Neural Networks. As the introduction of GPU's in the early 2000's enabled researchers to tap into the resources to make the Neural Networks run faster.

The first time the term Convolution Neural Network was used was in 1990 where Lecun et al.[8] inspired by a study on monkeys' brain by Hubel and Weist[9] in 1968 which showed that visual cortex cells of a monkey are spatially close and responsive only to a subset of cells in the retina. This type of organization of the cells proved to show that the visual of any object is highly local despite the changes in the peripheral visual, the object remains the same and recognizable by the brain. Lecun et al. [8] used this method to develop their neural network for hand written recognition which consisted for a 16X16px images as a direct input to the network.

Convolution Neural networks are a subset of Artificial Neural Networks where a machine learning algorithm analyses large amount of data. CNNs can be used for any type of cognitive tests, image processing, video processing, natural language processing to name a few. CNNs are comprised of multi layered neural networks which consists of one or multiple blocks of convolutional and pooling layers with at least one or more fully connected layers and one output layer. These layers use mathematical models to convolute and send the data to the succeeding layer.

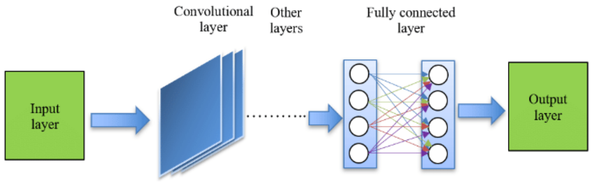

Fig 1. Generic Architecture of CNN Models.

## A. *Object Detection Networks*

The contents in this section discusses on the object detection algorithms, namely – R-CNN, Fast RCNN, Faster CNN and YOLO.

*RCNN*: In 2013 Girshick et al. [10] showed that the object detection algorithm based on a neural network performed better than the existing systems in the period between 2010 and 2012 which primarily used low-level features like SIFT [11] and HOG[12]. R-CNN saw a 30% increase in results compared to the previous best result [13].

The major benefit of is the fact that the first step is executed only once for all the classes and the parameters in the CNN are shared to all the categories. The evaluation is reduced to the first step and done only one per region as shown in Fig. 1. This enables lower memory usage and the computation required for the dot operation in the last step.

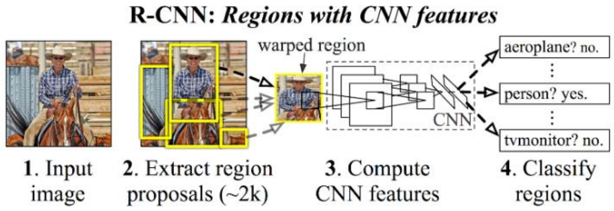

Fig 2. R-CNN: Regions with CNN features.

The drawback of this is the at the training is slow. In the paper, 5000 images were trained which took up to 2.5GPU days. The detection is also slow taking 47 seconds per image on a desktop. This showed the major drawback of this is it usually needs to run one full CNN forward propagation for each region and there are thousands of regions in an image.

Fast R-CNN: In 2015 Ross Girshick [14] produced the Fast RCNN which increased the detection performance in mAP from 62% to 66% on VOC 2012[15] and the testing speeds were 213 faster compared to RCNN. The training speeds also jumped by a factor of 9. Even though the Fast RCNN looked promising and performed better than the RCNN it lacked in detection time as it relied on selective searching for the initial generation of object proposals.

*Faster R-CNN:* Shortly after the release of Fast RCNN, the major bottleneck from the FAST RCNN was solved using region proposal computation [16] which increase the map from 77% to 70.4% on VOC 2012[15]. They also introduced a Region Proposal Network which shared features with the detection framework which had a positive effect on the speed of the detection pipeline thereby the effective running time reduced to just 10milli seconds.

When measured on a K40 GPU, the complete pipeline achieved 17fps with a slight reduction in map on a the Zfnet dataset [17] and achieve 5fps on a very deep VGG16[18] dataset.

*YOLO: You ONLY LOOK ONCE:* In 2015, Redmon et al. [19] published a new approach that previous existent systems didn't possess. This new system made predictions based on the entire image through a single neural evaluation as shown in Fig. 2 which enabled the YOLO model to be trained END to END directly for detection/ It proved to me 100X faster than Fast R-CNN, 3X faster than Faster R-CNN and achieved detection frame rate of 45 frames per second. The standard YOLO network architecture is very deep with 2 fully connected layers and 24 convolutional layers.

Apart from being very fast this architecture enabled YOLO to see the larger context of image than only specific regions for classifications. Another fascinating discovery of YOLO is that it models the size and shape of objects instead of the low-level cues like textures, lighting, background as it looks at the whole image during detection.

YOLO also offers a smaller network called Fast YOLO which uses 9 convolution layers instead of 24 in the standard YOLO. This model compromises on accuracy for lower memory consumption and higher speeds as it is capable of reaching 155 frames per second.

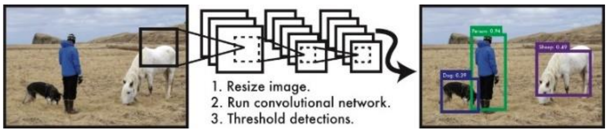

Fig 3. Working of YOLO on an image to perform detection

## III. LITERATURE REVIEW

In this section, we tabulate the data of 20 relevant research articles on Object Detection and Image classification of Animals using Machine Learning Algorithms and by using Convolution neural Networks.

The proposed method from Yu et al [20] showed an 82% accuracy when tested over a database from 2 different field cites consisting of over 7000 camera trap images of 18 animal species. The accuracy was achieved by deploying a species recognition algorithm based on the sparse coding spatial pyramid matching (ScSPM) by converting the images in the dataset into grayscale and using a combination of both SIFT and cLBP descriptors. The model proved to be working well in recognising the species of animals where the average accuracy of using both SIFT + cLBP but the author had to manually go through 10000+ photos, select images where an animal was present and crop the images without hindering the original ratio which reduced the dataset to about 7000 images.

The model used in Jennifer L. Price Tack et al. [21] called developed AnimalFinder which is used to detect animal presence in the wildlife camera trap images by comparing the individual photos with all the images that exist within the dataset. Around 65291 images were collected and then they were classified into 1557 images as deer, 590 as wild pigs and 2108 as racoons. When the threshold value was set to 0.95 the increase in the number of threshold value the AnimalFinder increased the total number of images Flagged which varied from 2174 images. When the threshold is set to 0.005 it is classified into deer images, wild pig images, racoon images with 45 percent, 23 percent and 18 percent respectively. When the threshold is set to 0.95x the percentages increased to 95 percent 97 percent, 94 percent for deer images, wild pig images and racoon images respectively.

Tabak, MA, Norouzzadeh, MS, Wolfson, DW, et al. [22] Shows that they trained their models by using CNN with 3,367,383 images to classify images of wildlife species which are obtained from camera traps automatically. This model performed very well by identifying the correct species with 97.6 percent accuracy. And they also checked the model to separate the empty images from those with animals in other samples of data set. They identified the species accurately with 98 percent accuracy. When they found 85 percent of, we're correctly classified as empty and 94 percent as images contained the animal then 95 percent of recall can be achieved. Their model was used to classify more than 2000 images.

Tabak, MA, Norouzzadeh, MS, Wolfson, DW, et al. [23] uses a shiny application that is available as a package in R called 'MLWIC2: Machine Learning for Wildlife Image Classification in R'. In this model, Motion-activated wildlife cameras are used to observe the animals remotely and noninvasively. But the drawback of this model is that the training can be done on a specific species and of a particular location and the model is not accurate enough when trained with the same species but from different location. The accuracy species was found to be 96.8 percent for the species and 97.3 for the empty animal model. They used a confusion matrix which gives idea of how all the images are classified by the species model. When they evaluated on the obtained sample images the accuracy ranged from 36.3 percent to 91.3 percent.

Banupriya at al,.[24] concentrates on identifying, counting and describing animals using Deep Learning Algorithm, they have trained CNN to know the behaviour of 48 species in about 3.2 million images of Serengeti dataset, this CNN automatically identifies animals with almost 93.8% accuracy their system automates animal detection for 99.3 of the images in the data set with accuracy 96.6% which save a lot of human efforts. The Algorithm performs 5 processes to detect the animal. This uses perceptron's which is a ml unit algorithm. Input holds raw pixel values with 3 channels R, G, B next features of the image is extracted the relationship between pixels are preserved it uses image matrix and filter. Pooling layers reduces the parameter when the size of the image is more. Flattening is used to converts 2d array into single vector the fully connected layers that is the hidden layers of the CNN is used to combine features and attributes to predict the output more accurately.

Benjamin Kellenberger et al., [25] present a called Annotation interface for data-driven ecology (AIDE). It is a framework which performs the task of image annotation for surveys. The AIDE helps in connecting users and machine learning models into a feedback loop in an easy manner. AIDE is a labelling tool which is very versatile as it offers high degree of customisability and provides support to many users. This is the first platform which provides the Machine Learning model to assist the annotation platforms. The annotation interface in Data-driven ecology is in active development and in the coming releases it is ready to be expanded in functionalities.

## IV. CONCLUSION.

The use of Convolution Neural Networks to detect wild animals shows that the detection and classification of animals can be done with good efficiency and accuracy. The applications mentioned in [26][27] proves using machine learning for classification is the way to go for the future and the solution needn't come at a cost as most of the applications are open source. As more companies start to invest towards saving the world, we feel that this limitation of the wildlife camera traps can be made less of a burden and be made more streamlined and easier to use by everyone.

## REFERENCES

[1] K. Ullas Karanth, Estimating tiger Panthera tigris populations from camera-trap data using capture—recapture models, Biological Conservation, Volume 71, Issue 3, 1995, Pages 333-338, ISSN 0006-3207, https://doi.org/10.1016/0006-3207(94)00057-W. (https://www.sciencedirect.com/science/article/pii/000632079400057W)

[2] Newey, S., Davidson, P., Nazir, S. et al. Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective. Ambio 44, 624–635 (2015). https://doi.org/10.1007/s13280-015-0713-1

[3] Newey, S., Davidson, P., Nazir, S. et al. Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective. Ambio 44, 624–635 (2015). https://doi.org/10.1007/s13280-015-0713-1

[4] Parker, D.B. (1985) Learning-Logic: Casting the Cortex of the Human Brain in Silicon. Technical Report Tr-47, Center for Computational Research in Economics and Management Science. MIT Cambridge, MA.

[5] Lecun, Y. (1985). Une procedure d'apprentissage pour reseau a seuil asymmetrique (A learning scheme for asymmetric threshold networks). In Proceedings of Cognitiva 85, Paris, France (pp. 599-604)

[6] Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature 323, 533–536 (1986). https://doi.org/10.1038/323533a0

[7] Kunihiko Fukushima, Sei Miyake, Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position,

[8] Pattern Recognition, Volume 15, Issue 6, 1982, Pages 455-469, ISSN 0031-3203, https://doi.org/10.1016/0031-3203(82)90024-3. (https://www.sciencedirect.com/science/article/pii/0031320382900243)

[9] Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, andL. D. Jackel. Handwritten Digit Recognition with a Back-Propagation Network.In Advances in Neural Information Processing Systems, pages 396–404. Morgan Kaufmann, 1990...

[10] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology, 195(1):215–243, March 1968.

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:1311.2524[cs], November 2013.

[12] D.G. Lowe. Object recognition from local scale-invariant features. In The Proceedings of the Seventh IEEE International Conference on Computer Vision,1999, volume 2, pages 1150–1157 vol.2, 1999.

[13] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, volume 1, pages 886–893 vol. 1, June 2005.

[14] Mark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams,John Winn, and Andrew Zisserman. The Pascal Visual Object Classes Challenge:A Retrospective. International Journal of Computer Vision, 111(1):98–136, June 2014.

[15] Ross Girshick. Fast R-CNN. arXiv:1504.08083 [cs], April 2015.

[16] FMark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision, 111(1):98–136, June 2014.

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: TowardsReal-Time Object Detection with Region Proposal Networks. arXiv:1506.01497[cs], June 2015.

[18] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. arXiv:1311.2901 [cs], November 2013.

[19] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs], September 2014.

[20] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640 [cs], June 2015.

[21] Yu, Xiaoyuan & Jiangping, Wang & Kays, Roland & Jansen, Patrick & Wang, Tianjiang & Huang, Thomas. (2013). Automated identification of animal species in camera trap images. EURASIP Journal on Image and Video Processing. 1. 10.1186/1687-5281-2013-52.).

[22] Price Tack, Jennifer & West, Brian & McGowan, Conor & Ditchkoff, Stephen & Reeves, Stanley & Keever, Allison & Grand, James. (2016). AnimalFinder: A semi-automated system for animal detection in time-lapse camera trap images. Ecological Informatics. 36. 10.1016/j.ecoinf.2016.11.003.

[23] Tabak, MA, Norouzzadeh, MS, Wolfson, DW, et al. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods Ecol Evol. 2019; 10: 585– 590. https://doi.org/10.1111/2041-210X.13120

[24] Tabak, MA, Norouzzadeh, MS, Wolfson, DW, et al. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. Ecol Evol. 2020; 10: 10374– 10383. https://doi.org/10.1002/ece3.6692

[25] BANUPRIYA, N., S. SARANYA, RASHMI SWAMINATHAN, SANCHITHAA HARIKUMAR, and SUKITHA PALANISAMY. "ANIMAL DETECTION USING DEEP LEARNING ALGORITHM." Journal of Critical Reviews 7.1 (2020), 434-439. Print. doi:10.31838/jcr.07.01.85

[26] Kellenberger, B, Tuia, D, Morris, D. AIDE: Accelerating image-based ecological surveys with interactive machine learning. Methods Ecol Evol. 2020; 11: 1716– 1727. https://doi.org/10.1111/2041-210X.13489

[27] https://www.zambacloud.com/

[28] https://www.wildme.org