

Wild Animal Detection using Computer Vision

Sahil Faizal^a, Sanjay Sundaresan^b

Department of Computer Science, Vellore Institute of Technology, Chennai, Tamil Nadu, India

^asahil.faizal2019@vitstudent.ac.in

^bsanjay.work03@gmail.com

Abstract—Detection and identification of wild animals for tracking and protection purposes has become increasingly important with the deterioration of the environment, and technology is the agent of change which augments this process with novel solutions. Computer vision is one such technology which uses the abilities of artificial intelligence and machine learning models on visual inputs. Convolution neural networks (CNNs) have multiple layers which have different weights for the purpose of prediction of a particular input. The precedent for classification, however, is set by the image processing techniques which provide nearly ideal input images that result in optimal CNN performance. Image segmentation is one such widely used image processing method which provides a clear demarcation of the areas of interest in the image, be it regions or objects. The efficacy of the CNN is directly related to the efficacy of the segmentation techniques. Further, it is a well-established fact that heterogeneity in image sources is detrimental to the performance of CNNs. Thus, the added functionality of heterogeneity elimination is performed by the image processing techniques, introducing a level of consistency that sets the tone for the input layers of the CNN.

Keywords— Object Detection, computer vision, Artificial Intelligence, CNNs, Image segmentation, data augmentation

I. INTRODUCTION

The expansion of urban areas in modern times has resulted in widespread displacement of habitats in forested areas. As a result, wild animals are forced to venture into the human settlements that often infringe on their routine activities. More often than not, food is the primary motivator for such peregrinations. It is at this point that there is tangible danger to any humans that inadvertently cross the path of these animals when they are at their most ferocious predispositions. Hence, a need arises for the detection of wild animals at the border of human settlements close to wild habitats. A robust, reliable and effective preemptive warning mechanism would drastically eliminate risk of fatal human-animal conflict, both in the interest of protecting human lives and avoiding loss of endangered animals. Moreover, such a system would also be useful in wildlife sanctuaries and biosphere reserves to monitor the movement of animals at the border areas of such establishments which have often proved difficult to control. The usage of technology and robust cameras is not an alien concept in most major biosphere reserves and national parks around the world. Although there has been a considerable amount of progress, software-based tools have not been explored to a satisfactory extent in these use cases. Computer vision has the ability to transform the tracking and monitoring process with the accuracy that its components and supporting techniques provide. The automation-augmented reduction of man-hours invested in searching for and tracking wild animals is perhaps the biggest potential boon that computer

vision can provide. The pre-processing involved in the application of computer vision algorithms is often under-documented although it plays a key role in the success of the algorithm. A deep understanding of the nature of the inputs is necessary to make appropriate changes at crucial junctures of processing to meet the often-convoluted criteria required by complicated deep learning algorithms. Transforming the images is invariably necessitated due to the erratic nature of real-world data feeds. The absence of an artificial synthesis element in the generation of inputs via raw camera stills adds to the intricacies involved in the image processing component.

II. LITERATURE SURVEY

Wildlife protection has had a significant number of approaches which integrate myriad technology stacks to solve niche issues.

[1] is a system that combines deep learning classification with dynamic background modelling to evolve a swift and precise method for human and animal detection from highly cluttered camera trap pictures. Background modelling helps generate region proposals for foreground objects, which are then classified using the DCNN, resulting in improved efficiency and increased accuracy. The proposed system achieves 82% accuracy in segmenting images into human, animal and background patches.

[2] uses a network of cameras, connected to PIR motion sensors, so that image capture is triggered only when some movement is detected. The images captured through these cameras are processed to detect the presence of wild animals using YOLO, and if an animal is found, identify the species. Once identified, the animals are tracked for a suitable time using CSRT in order to determine their intent – such as to find whether they are moving across the village, or into it. In the latter case, alerts are generated and local authorities are notified through proper channels. The models at present can detect 5 types of animals namely (elephant, zebra, giraffe, lion and cheetah).

[3] is a comparative study on 4 different algorithms based on deep neural networks. Two variants of single shot multi-box detector (SSD) and two variants of faster region-based CNN (Faster R-CNN) have been compared. Two different activation functions are also compared. The SSD variants outperform all faster R-CNN variants and provide more precise detection compared to the latter.

[4] explored non-intrusive monitoring of animals. They generated huge volumes of data as the made use of multiple camera trapping networks. To analyze the data, they used very deep CNN framework and chose 26 species from the Snapshot Serengeti (SSe) dataset. The proposed model achieved an accuracy of 88.9%. A comparison with other techniques was also carried out which showcased that their model outperformed previous approaches.

[5] proposes an automated wildlife monitoring system which leverages state-of-the-art deep CNN architecture. The model achieved 90.4% accuracy for 3 animal classes. A single labeled dataset was used for training purposes, and a focus is placed on filtering animal images.

[6] is a comparative study between the bag of visual words and deep learning CNN techniques for wild animal recognition. The comparison is done for gray level as well as color information. The features extracted by the BOW models were combined with a regularized L2 support vector machine for classification. This study suggests that there is a clear performance gulf between CNN methods and BOW.

[7] Uses YOLOv3, a CNN architecture as a pre-trained model through transfer learning technique. Fine tuning was subsequently performed using an amalgamation of self-shot and crowdsourced images. The model locates the object detected and adds a bounding box upon it.

III. PROPOSED WORK

The intrinsic nature of any classification technique is the ability to accurately identify the major features of the target that it aims to predict. It is a well-documented fact that Deep Learning models and frameworks provide enhanced accuracy when the inputs have decreased source-induced heterogeneity. Since raw real-world feeds do not generate ideal or optimal images for classification, the onus is on the application of image processing techniques to act as the liaison between the classifier and the inputs. Some of the most influential challenges to classifier performance with regards to light intensity, unavailability of high-quality night-vision training images, noisy or element-rich backgrounds of captured images and luminance problems because of shadow effect need to be mitigated. The goal is to enable the classifier to extract features optimally from the images in order to assign a class to them with minimal loss and maximum accuracy. Ergo, the preprocessing needs to be performed specifically to ensure that the features are as distinctly visible as possible. The proposed sequence of operations on input images involves:

1. K-Means based segmentation via OpenCV, with a K-Value of 3
2. Noise removal in segmented images in order to increase the contrast between regions, ensuring the highest level of distinction.
3. Data Augmentation by flipping and rotating the noise-free segmented images to increase the diversity of the training set with respect to orientation.

4. Feeding the processed images to the InceptionResNetV2 classifier which predicts the class of the animal.
5. Using the convolutional layers of pre-trained neural network with existing pre-trained values which were obtained from highly diverse bulk training samples.
6. Completely customizing the fully connected layers, keeping the entire model open for training and updating weights.

The usage of K value of 3 is motivated primarily by the need to minimize induced noise in the input images and subsequently avoid adding to the existing noise. A higher K-value would theoretically be useful for classification, but it was observed to result in incoherent boundaries and significant overlap between regions. The necessity for noise removal is once again dictated by the need to optimize the images for feature extraction. Since InceptionResNetV2 classifier is CNN-based with 164 layers, a larger and more imperatively diverse training set is ideal. Diversity with respect to number of samples distributed across classes exists in standardized datasets, but orientation-based diversity has to be artificially introduced, thus propagating the need for data augmentation. At this juncture in the processing pipeline, fine tuning is introduced via transfer learning technique, primarily to boost the performance of the classifier and bridge the gap to its optimal accuracy. As the pre-trained weights of InceptionResNetV2 are proven to be effective in image-classification use-cases, the convolutional layers of InceptionResNetV2 are set with pre-trained weights. In the interest of flexibility, however, the fully connected layers are built according to the specifics of the training set. The usage of this model-building approach is the ideal balance between accuracy and flexibility. The proposed approach also differs from those documented in section II with regards to the number of animal classes under consideration. The 10 classes of animals are as listed:

```
{'Cheetah': 0,
'Chimpanzee': 1,
'Elephant': 2,
'Fox': 3,
'Jaguars': 4,
'Lion': 5,
'Orangutan': 6,
'Panda': 7,
'Panthers': 8,
'Rhino': 9}
```

Figure 1: Animal Classes under consideration

Additionally, class imbalance is also eliminated by the data augmentation performed in step 3, where low sample sizes for classes are mitigated if present.

IV. SYSTEM DESIGN

The system can be divided into four parts - data collection, preprocessing, re-sampling and model building.

Data collection: -

The input images were sources from a curated dataset consisting of 6484 images, based primarily on the IUCN Red List of Threatened Species.

Pre-Processing: -

It was found that these images had a lot of distortion and noise in them so removing them was a necessity for the CNN algorithm to work well. After experimenting with various noise removal and enhancement techniques it was found that K-Means segmentation was successful enough to bring in a differentiating factor between the images as it was able to remove the background of the images leaving behind the animals in the images.

Re-sampling: -

Once this stage was done, it was noticed that significant class imbalance existed between the classes, hence some data samples were synthesized from original ones through rotation, flipping and zooming to increase sample count.

Model Building: -

Once the images were deemed fit for undergoing training, different models were built including VGG19, VGG16, InceptionV3, MobileNetV2, MobileNetV3, InceptionResNetV2 to gauge the potential of fine-tuning technique to classify images into multiple classes based on its past experience of undergoing training with ImageNet dataset having 1 million samples. Eventually it was found that InceptionResNetV2 had the best results after training for 40 epochs. Data Augmentation was used before feeding data into the model to bring in diversity in the dataset. The technique of Fine Tuning involves the usage of initial weights used in CNN layers as the best ones from the ImageNet training. The fully connected layers are custom-built as per our need and the entire model is subjected to weight changes by backpropagation during the trading phase. In the fully connected layer first and second layers have 512, 256 neurons while the third(final) has 10 with softmax activation, Adam optimizer and Categorical cross entropy loss. When overfitting occurred, between the first and second layer, a dropout layer was added to cut off 25% of the connections in between them. The evaluation of the model performance is done on the basis of accuracy obtained after the building process.

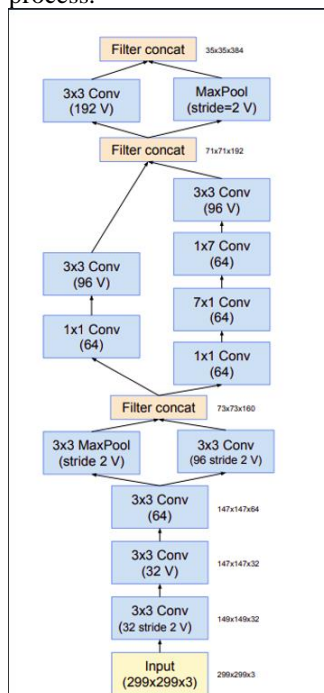


Figure 2: Model Architecture

V. OUTPUT



Figure 3: Sample Input image



Figure 4: Corresponding Segmented Image (which forms model input)

The graphs below depict the accuracy of the model upon testing with a standardized image set which is a randomly generated subset of the source dataset.

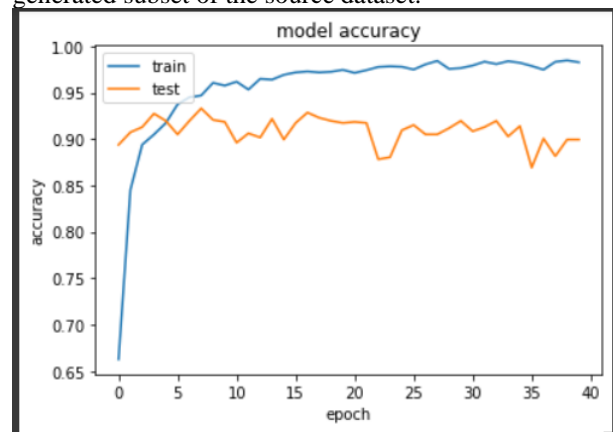


Figure 5: Model accuracy plot for 40 epochs

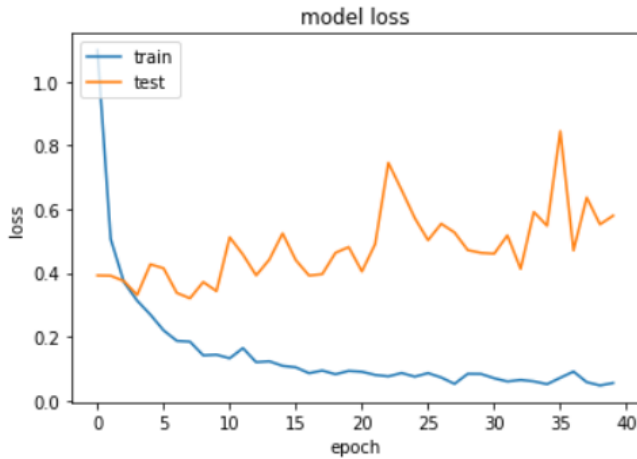


Figure 6: Corresponding model loss

VI. FUTURE WORK

Although the obtained accuracy is satisfactory for real world use cases, robust detection of wild animals exclusively at night time when there is no natural light, is potentially the most challenging yet impactful expansion. The constraints involved in processing images from night vision cameras are highly detrimental to the application of computer vision models, as the distinction between features plays the central role in the working of the model. However, it is possible to process the images generated during night hours in order to incorporate some contrast and distinction between regions of interest in the image. The availability of standardized night-vision wildlife camera datasets with a large number of animals would be the ideal catalyst for this potential expansion.

VII. ACKNOWLEDGEMENT

We wish to place on record our sincere gratitude to **Dr. Geetha S**, our course faculty, for encouraging us to be at our innovative best by guiding us towards a unique project that seeks to be different from the work done hitherto. We also take this opportunity to thank our friends, family, seniors and peers for providing the appropriate constructive criticism at different times during the development of this project, thus

enabling us to maximize the efficiency of our concept application to fulfil our potential to the best of our abilities.

REFERENCES

1. [H. Yousif, J. Yuan, R. Kays and Z. He, "Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification," 2017 IEEE International Symposium on Circuits and Systems \(ISCAS\), 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050762.](#)
2. [Sayagavi, A.V., Sudarshan, T.S.B., Ravoor, P.C. \(2021\). Deep Learning Methods for Animal Recognition and Tracking to Detect Intrusions. In: Senjyu, T., Mahalle, P.N., Perumal, T., Joshi, A. \(eds\) Information and Communication Technology for Intelligent Systems. ICTIS 2020. Smart Innovation, Systems and Technologies, vol 196. Springer, Singapore.](#)
3. [Okafor, E., Berendsen, G., Schomaker, L., Wiering, M. \(2018\). Detection and Recognition of Badgers Using Deep Learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. \(eds\) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science\(\), vol 11141. Springer, Cham.](#)
4. [Alexander Gomez Villa, Augusto Salazar, Francisco Vargas, Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks, Ecological Informatics, Volume 1, 2017, Pages 24-32, ISSN 1574-9541.](#)
5. [H. Nguyen et al., "Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring," 2017 IEEE International Conference on Data Science and Advanced Analytics \(DSAA\), 2017, pp. 40-49, doi: 10.1109/DSAA.2017.31.](#)
6. [E. Okafor et al., "Comparative study between deep learning and bag of visual words for wild-animal recognition," 2016 IEEE Symposium Series on Computational Intelligence \(SSCI\), 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850111.](#)
7. [Yuanqin Dai "Wildlife recognition from camera trap data using computer vision algorithms", Proc. SPIE 12155, International Conference on Computer Vision, Application, and Design \(CVAD 2021\), 1215503 \(20 December 2021\); https://doi.org/10.1117/12.2626540](#)