Deep Learning-based Model for Wildlife Species Classification

Shailendra Singh Kathait
Co-Founder and Chief Data Scientist, Valiance Analytics Pvt. Ltd.
Noida, Uttar Pradesh

Ashish Kumar
Principal Data Scientist, Valiance Analytics Pvt. Ltd.
Noida, Uttar Pradesh

Piyush Dhuliya
Valiance Analytics Pvt. Ltd.
Noida, Uttar Pradesh

Ikshu Chauhan
Research Scholar
Doon University

ABSTRACT
In this paper, we describe that motion-activated cameras are nowadays widely used in Ecological parks as well as wildlife sanctuaries. These cameras capture the images whenever any motion is observed by the sensors. They are also capable of capturing infrared images therefore providing millions of images, which was earlier a very expensive as well practically infeasible task. However, extracting useful information from these images about any wildlife species is still a time-consuming and labour-intensive task. We demonstrate that deep learning models can be used for extracting this information near human-level accuracy. We trained VGG16 ConvNet architecture using transfer learning on a dataset of 33,511 images of 19 species from the Ladakh region of India and achieved training and testing accuracy of 89.12% overall. The pipeline developed here has wide application in wildlife monitoring across different national parks.

General Terms
Wild-Life Species Classification, Species classification, Animals, Forest, Animal Identification

Keywords
Transfer learning, Data Augmentation, Computer vision, OpenCV, Image Processing, Machine learning, Deep Learning, Convolutional Neural Networks.

1. INTRODUCTION
In the last few decades, many wildlife species have become extinct, and some are on the verge of extinction. In the year 2022 alone 7 species have become extinct [1]. If these trends continue in the future, soon the most common species of animal will be on the verge of extinction. Therefore, proper wildlife monitoring of rare species of animals is required which will help us to develop natural habitats and environment for these species to thrive according to their needs. Motion sensor cameras were an attempt to capture these species, study their behavior and extract useful information. They are used for understanding the distribution and population sizes of various species in addition to the use of GPS trackers, and radio tracking devices.

The main motive of these was to reduce human intervention in their natural habitat and reduce the risks to human life along with saving tons of laborious work. However, this method comes with its own set of challenges. The images are captured as soon as some motion is detected resulting in continuous shots of images. Images are captured in a very dynamic environment having continuously changing weather, improper illumination, and angles that sometimes block the view of the camera. This collection of images requires proper cleaning before they can be published on any open-source platform. In Fact, many of these captured images are empty (don't have any animal within them), and it happens when the camera is triggered by some external factor like winds.

2. PROPOSED METHODOLOGY
This research work proposed the use of deep learning-based algorithms coupled with some image processing techniques to detect species present in the images [2].

There were broadly 3 different types of training that could have been employed here:

A. Training CNN from scratch.
B. Unsupervised learning followed by Supervised fine Tuning.
C. Transfer Learning i.e., Fine-tuning a CNN model which is pre-trained on a large dataset containing Millions of different images.

This research work is mainly focused on the use of transfer learning on pre-trained CNN models. Models trained on the ImageNet Dataset were being used for the further fine-tuning process. The basic Ideology behind this lies in the fact that the initial layers of the ConvNet Architecture identify preliminary features such as edges, and corner points which are common to any sort of classification problem.[3]

2.1 DATA ACQUISITION AND CLEANING
Images trapped by the cameras were being stored on the AWS S3 bucket. This data was being pulled into the working instance. The Images were classified manually according to the species to which they belong. Images having more than one species were segregated into all the corresponding species which are named as 1) woolyhare, 2) pica, 3) domestic, 4) fox, 5) snowleopard, 6) marmot, 7) chukar, 8) human, 9) blue sheep, 10) urial, 11) horse, 12) ibex, 13) snowcock, 14) wolf, 15) stone matin, 16) kiang, 17) dogs, 18) pallas cat, 19) lynx. There was a total of 33,511 images belonging to 19 different classes, with rare classes having a smaller number of images as compared to other classes.

Data cleaning was also performed manually where images containing no species (outliers) were removed. This is a necessary step before training the model since these outliers

will affect our model weights which in turn will affect the overall accuracy of the model.

2.2 DATA AUGMENTATION

One of the biggest drawbacks of the deep learning models is their hunger for large amounts of Diverse Training Data. [4]. The amount of training data adversely affects the accuracy of the model. For Detecting different wildlife species with similar kind of background, the amount of training data required per class increases dramatically.

There were also large amounts of infrared images which were captured during the night-time which makes the learning process of the model more difficult with such a small amount of training images.

To overcome this problem, we adopted a very common method called Data Augmentation. Data Augmentation is helpful in places where increasing the training set will help in Generalizing the model over the new unseen images thus preventing overfitting during training. [5]

Data Augmentation helps to increase the number of training images by generating new images from the existing one by using various Image processing techniques.

Some of the techniques that were used in this research are:

A. Shearing: Images were distorted along X and Y axes to a certain extent to create a perception of how humans see things from different angles.
B. Brightness: Changing brightness will create a natural effect of dynamic lighting conditions that will enhance the learning process of the model.
C. Flipping: Images were flipped horizontally and vertically.
D. Rotation: Images were rotated by a certain angle.
E. Mean: Mean subtraction was performed.

An example of the Data Augmentation techniques has been shown in below figures (a), (b), (c), (d) and (e).

(a) Original Image

(b) Brightness Increased Image

(c) Flipped Image

(d) Rotated Image

(e) Mean Difference Image

2.3 DEEP LEARNING

Deep learning is a subset of machine learning which has gained popularity over the last decade. Deep learning is widely applied in the field of image classification because of the edges it has over machine learning algorithms. A typical Machine learning algorithm requires a pre-Processing step called feature extraction. Feature extraction is a complex process that requires proper knowledge of the Problem.[6]

Deep learning on the other hand does not require this manual process of feature extraction. Artificial neural network (ANN) performs feature extraction directly and on their own on the raw data. This allows the ANN to identify complex patterns in the images with the help of non-linear activation functions stacked over the deep layers.

The working of Artificial neural networks is inspired from the working of neurons in the brain but in a simplified way. ANN are generally subdivided into 2 categories:

Feedforward neural network: These types of neural networks allow the information to flow only in the forward direction. They don't use any feedback mechanism.

FeedBackward neural network: These types of neural network allow the flow of information bidirectionally i.e., both forward and backward. They use a feedback mechanism to enhance the learning process.[7]

CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are a type of feedforward neural network. One of the biggest problems with the fully connected neural network was the presence of large amounts of parameters which in turn requires extensive computational power and increases the chance of overfitting. To solve this problem CNN was developed. They use a concept of kernels in which they share the parameters over the input image matrix with the help of a sliding window mechanism.

Convolutional neural networks are nothing, but a series of different convolutional layers stacked over each other. With the later layers identifying more complex patterns than the initial ones. There are various types of layers such as the Convolutional layer, the Pooling layer, a fully connected layer that is used in CNN architecture[2]. CNN uses nonlinear activation functions to identify the complex patterns.

A typical CNN block consists of a series of Convolution and pooling layers followed by the Fully connected layers at the end of the network. Convolutional layers perform feature extraction by typically using linear operation (convolution operation) followed by the activation function. The Pooling layer is responsible for reducing the size of the image matrix to reduce the number of trainable parameters in the subsequent layers. It also helps to reduce the effect of noise and distortions in the image.[8]

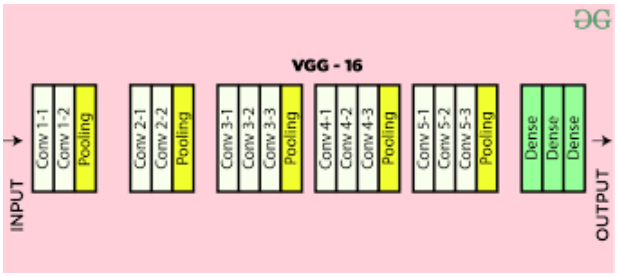

Fig 1. VGG-16 Architecture

TRANSFER LEARNING

The number of resources and data required to train deep learning models is enormous. To solve this problem Transfer learning applies the learning from one task to another similar task. It is a technique where a model trained on one task is then used with certain modification for the second task. Transfer learning had shown promising results in the field of computer vision and image analysis. One of the major reasons being the fact that initial feature extraction process is more or less similar for different types of images which includes detecting edges, corners etc. [9]

In a typical ConvNet Architecture The initial convolution and pooling layers acts as a feature extractor and this learning can be used for any new task while the last fully connected layers act as classification layers responsible for predicting various classes and that is a problem to problem dependent.

TRAINING OF MODEL ON DATASETS

VGG16 ConvNet Architecture was used as in Fig.1 for the training. Freezing and unfreezing of layers was utilized for the training of data set utilized for this purpose. Freezing and unfreezing mechanism helps in using the layers of pre trained model. A new layer can be incorporated to the pretrained model to enhance its output or according to requirements the no. of layers can be freeze and unfreeze so that unnecessary time to train the model from scratch can be avoided [10]. In this we have skipped the last three layers of VGG16 ConvNet Architecture and added four dense layers with 1024, 256,128, 64 no. of neurons successively in each layer.

3. RESULTS

33,511 images of 19 species of Ladakh region in India were utilized. 80% were used for training and 20% for validation. Thereafter, 3309 unseen images were tested on the trained model. This test set of 3309 images led to a confusion matrix given in Table 1. The confusion matrix gives us the data of rightly classified species e.g. out of 302 test images of birds, the trained model could rightly identify 284 images of birds. The right classification is bolded in the test set confusion matrix provided in Table 1.

Table 1. Confusion Matrix on Test Dataset

	BIRDS	blue sheep	chukar	dogs	domestic	fox	horse	human	ibex	kiang	lynx	marmot	pallas cat	pica	snowcock	snowleopard	stone mation	urial	wolf	woolyhare
BIRDS	284	0	1	1	0	0	0	4	0	0	0	3	0	0	2	2	0	0	4	1
blue sheep	1	195	0	1	6	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0
chukar	70	0	154	0	1	0	0	5	0	0	0	0	0	3	2	7	0	0	0	0
dogs	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
domestic	1	0	0	2	304	0	0	6	0	0	0	0	0	0	0	1	0	0	0	0

fox	0	0	0	2	1	240	0	6	0	0	0	0	0	10	0	10	0	1	2	19
horse	0	0	0	3	44	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0
human	2	0	0	0	0	0	0	235	0	0	0	2	0	0	0	1	2	0	0	0
ibex	1	0	1	1	3	0	0	4	56	0	0	0	0	0	0	0	0	0	0	0
kiang	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	1	0
lynx	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
marmot	3	0	0	0	0	0	0	0	0	0	0	253	0	0	1	0	0	0	0	0
pallas cat	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0
pica	5	1	0	0	0	0	0	3	0	0	0	0	0	296	0	8	0	0	0	2
snowcock	8	0	0	0	0	0	0	0	0	0	0	1	0	0	52	1	0	0	0	0
snowleopard	4	0	1	0	1	1	0	4	0	0	0	1	0	5	0	248	0	0	2	4
stone matin	0	0	0	0	1	0	0	0	0	0	0	0	0	7	0	2	24	0	0	1
urial	2	1	0	6	2	0	0	0	0	0	0	0	0	0	0	4	0	119	2	0
wolf	2	0	0	3	1	0	0	0	0	0	0	1	0	1	0	3	0	0	48	2
woolyhare	2	0	0	0	0	2	0	0	0	0	0	0	0	7	0	1	0	0	0	329

(X-Axis: True Class, Y-Axis: Predicted Class)

Based on the test-set confusion matrix another matrix was determined which gave us the accuracy of individual species as given in Table.2.

Overall test accuracy :89.12058023572075

Species	Test Accuracy	Train Count	Validation count	Test Count	Total Count
pallas cat	100%	53	10	16	79
marmot	98.44%	1934	342	257	2533
human	97.11%	2629	464	242	3335
kiang	96.97%	110	20	33	163
domestic	96.82%	2531	447	314	3292
woolyhare	96.48%	2277	402	341	3020
dogs	94.44%	706	125	18	849
BIRDS	94.04%	1023	181	302	1506
pica	93.97%	2238	395	315	2948
blue sheep	91.98%	2476	437	212	3125
snowleopard	91.51%	2037	360	271	2668
urial	87.5%	1720	304	136	2160
ibex	84.85%	1635	289	66	1990
snowcock	83.87%	1144	202	62	1408
fox	82.47%	2167	383	291	2841
wolf	78.69	1267	224	61	1552
stone matin	68.57	310	55	35	400
chukar	63.64	1841	326	242	2409
lynx	50	67	12	2	81
horse	49.46	312	56	93	461

Table.2

The test accuracy matrix delivers us the accuracy of classification of the 19 species we have considered in this experiment. The accuracy level varies from about 49% to about 100% for individual species.

The graph shown in fig.2. shows us the individual accuracy of various species.

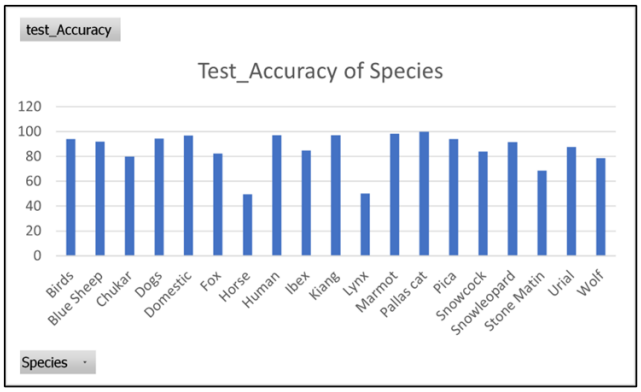

fig.2. Test Accuracy in every Class

The overall test accuracy can be calculated using the following formula:

Overall Test Accuracy= (Sum of accurately identified species from confusion Matrix)/(Sum of total test count species)×100…….(1)

The overall test accuracy obtained is 89.12%.

4. CONCLUSION

The dataset utilized used in this paper consists of species of Ladakh region in India. The model utilized over here has geographical limitations as we might not get the same output for other areas. The overall test accuracy can be further enhanced by using a more pronounced training data. As the training data will increase so will the efficiency of the model.

5. FUTURE WORK

As part of Future research, wildlife species classification models need to be built on regional contexts, like Himalayan,

North Eastern, Western Ghat, and Central regions in India, which is imperative for increased overall accuracy of the model. Incorporating advanced techniques such as transfer learning and attention mechanisms holds promise for enhancing classification accuracy. Future research should collaborate with conservation organizations for comprehensive field data collection. Additional real-life data scenarios ensure model generalization and practical applicability in real-world scenarios, advancing our collective commitment to wildlife preservation.

6. REFERENCES

[1] John R Platt, “The book of the dead- the species declared extinct in 2022”, The Revelator:. Jan 19, 2023.

[2] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021).

[3] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[4] Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M. et al. Deep learning applications and challenges in big data analytics. Journal of Big Data 2, 1 (2015).

[5] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019).

[6] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 2319-2323.

[7] D. K. Mishra, R. Bhati, S. Jain and D. Bhati, "A Comparative Analysis of Different Neural Networks for Face Recognition Using Principal Component Analysis and Efficient Variable Learning Rate," 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, Kota Kinabalu, Malaysia, 2010, pp. 354-359, doi: 10.1109/AMS.2010.78.

[8] AS. Kumar and E. Sherly, "A convolutional neural network for visual object recognition in marine sector," 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, 2017, pp. 304-307, doi: 10.1109/I2CT.2017.8226141.

[9] Tammina, Srikanth. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. 10.29322/IJSRP.9.10. 2019.p9420.

[10] Debanshu Banerjee, Taylor D. Sparks, comparing transfer learning to feature optimization in microstructure classification, iScience, Volume 25, Issue 2, 2022, 103774, ISSN 2589-0042, https://doi.org/10.1016/j.isci.2022.103774.