Individual Tiger Identification using Transfer Learning

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

Vaibhav Singh
Data Scientist, Valiance Analytics Pvt. Ltd.
Noida, Uttar Pradesh

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

ABSTRACT
This paper describes the methods to classify images of tigers into their respective classes using deep learning models. The classes here represent the individual tigers themselves. With the use of motion activated cameras it has been possible to get a huge number of images of animals in their natural habitat but to make any use of this data in unique animal identification has been a challenge, which is aimed to be solved in this paper. There are techniques available to identify species of animals from the images, but classifying a particular animal specie into its individuals yet remains a tough task. The method used has a pipeline that involves the YOLOv8 model and EfficientNetB3 model with transfer learning to classify an image of tiger while working with images of 98 unique tigers. Out of 192 tigers whose images were available, the paper included only 98 in the model after setting a threshold of availability of 15 images at least for each tiger.

General Terms
Tiger Classification, Animals, Forest, Tiger Reserve, Animal Identification, Pattern Recognition, Wildlife Preservation.

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

1. INTRODUCTION
Wildlife preservation has been a challenge and an act of utmost importance since quite some time now, and when it comes to wild animals in their natural habitat, the task is never easy. People have been using all sorts of techniques to track the animals, their movement and behavior in unaltered environments. Collar tag, strap cameras and what not have helped our cause so far, but we wanted to go even further with motion detection cameras. We're working with a large tiger reserve with more than 100 unique tigers. The pictures are captured by 466 motion-capture cameras, and have images from all hours of the day, different environmental conditions and illumination. This paper aims to identify the tigers individually, which can help enormously in movement tracking of these wild animals as well as understanding their density at any given time and location in their natural habitat. Before feeding the images in the model, it requires proper cleaning and preprocessing and good results can be expected when the whole tiger is in the picture.

2. AIM
To identify individual tigers based on their stripe pattern using deep learning and computer vision, as the current techniques are inefficient and no reliable output can be generated out of them.

3. PROPOSED METHODOLOGY
This research paper proposed the implementation of deep learning based architecture as well as feature-based methods like SIFT for identifying patterns of the unique tigers we're trying to identify.

The 2 types of methods we used as as follows:

A. Extracting SIFT features and comparing them with the ground truth.

B. Transfer learning based model on a CNN model[3] that has been pre-trained on a large sample of diverse images.

The main focus here has been on the application of transfer learning on a pure CovNet algorithm trained on ImageNet dataset. Training the last few layers as per our requirement can be an efficient way to exploit the versatility of such a model. [1]

4. DATA COLLECTION AND CLEANING
The images captured by the cameras are captured in cloud buckets, labeled with their class(the class in this paper is the same as the identity of a unique tiger, e.g., T142-F means a female tiger numbered 142) and other information such as camera number and location from where it was captured. For training purposes, only the class is needed out of the available information.

There are multiple images associated with each tiger, ranging from 3 images for a tiger to 100+ images for another. This research paper sets a threshold of requirement of 15 images of a particular tiger to include it in the pipeline sample, as it has been observed that below this number, the accuracy dips rapidly with the current architecture. Thus 98 out of the total 192 tigers are left which whom the research continues.

First the object detection feature of the YOLO v8 model is used for getting the bounding box around tigers[7]. Although YOLO v8 doesn't have a label of a tiger, it identified the tigers in the images as zebras. Since the bounding boxes provided by YOLO around the animal are pretty accurate, in can be used for the purpose.

Fig 1: Sample original image 1

Fig 2: Sample original image 2

Fig 3: Cropped image from sample 1

Fig 4: Cropped image from sample 2

5. SIFT FEATURES

The scale-invariant feature transform (SIFT) is a computer vision algorithm to detect, describe, and match local features in images[2].

SIFT keypoints of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of keypoints that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches.

The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

5.1 Implementation

An image of a tiger is taken as a reference and the SIFT descriptors are extracted from the algorithm[6] provided by OpenCV. The features extracted here act as the reference or base for that tiger, which can be later used to compare with the features extracted from other images of the same tiger. Then the SIFT features for other images of the same tiger are extracted and are matched, with its reference feature obtained earlier, using BFmatcher (Brute-force matcher). This way, the number of feature descriptors matched is collected.

5.2 SIFT Performance Result

This paper took 3 images of a particular tiger, with the complete body of the tiger available in each of the pictures. After getting the cropped images from Yolo model like shown in Fig 3, we extracted the SIFT features of the images. Out of the three images, one(Tiger 1) is set as reference, and the other 2 (Tiger 2 and Tiger 3) as test samples. It has to be kept in mind that all the 3 images belonged to the same tiger.

The results achieved from this were not very satisfactory, the reason being the wide range of feature descriptor count we got for the different images of the same tiger. As shown in the below images, when features of 2 test images (Tiger 2 and Tiger 3) of the same tiger were compared to that of it's reference image(Tiger 1), it was difficult to conclude anything solid from the matches due to varying number of descriptors. This can be attributed to the presence of varying background in the images, which can produce useless and unrelated SIFT features

Tiger	Descriptors	Matches
Tiger 1 (Reference)	312	102
Tiger 2 (Test)	274	

Fig 5: SIFT feature comparison between T1 and T2

Tiger	Descriptors	Matches
Tiger 1 (Reference)	312	115
Tiger 3 (Test)	1337	

Fig 6: SIFT feature comparison between T1 and T3

Thus, we move forward with Deep learning algorithms to draw concrete conclusions.

6. DATA PREPROCESSING

6.1 Data Augmentation

One challenge while working with deep learning architecture is its hunger for large training dataset[4]. Since the dataset at hand only has a limited number of training images for each tiger, the method of artificial data generation using augmentation[5] was given a go.

In the training dataset, the range of image counts, for different tigers, varies a lot as mentioned earlier. Hence to avoid the challenge of class imbalance, we create augmented copies of the existing images and bring the image count for each tiger to 100.

The techniques used for augmentation are as follows:

A. Rotation : Images were rotated along X and Y axis

B. Horizontal flip : Flipping the image by 180 degrees horizontally

C. Width shift : Shifting the images along Y axis

D. Height shift : Shifting the image along X axis

E. Zooming : Zooming the image in and out

As a final step, each image is scaled down to 200 x 300 pixels resolution.

Fig 7: For the reference image above, we created below augmented + scaled images

Fig 8: Rotated and zoomed in

Fig 9: Horizontal flipped and right shifted

Fig 10: Horizontal flipped, left shifted and rotated

6.2 Data Split

The complete data is split in train(75 %) and test subsets and further the test data into test (12.5%) and validation (12.5%) dataframes. Stratified sampling is employed for this split, which means the proportions of images of a particular tiger were also distributed among the train, test and validation set in the same ratio.

7. DEEP LEARNING

Deep learning is a part of machine learning that has grown in popularity over the past decade. Deep learning is widely used in the field of species classification [3] because of its superior performance over other statistical methods. Common machine learning algorithms require a preprocessing step called feature extraction. Feature extraction is a complex process and requires proper knowledge of the problem.

Deep learning on the other hand does no longer require this manual process of feature extraction. Artificial neural network (ANN) performs feature extraction automatically and on their own at the raw records. This allows the ANN to identify complicated patterns in the image with the assistance 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:

A. Feedforward neural network: These types of neural networks only allow the flow of information in the forward direction.

B. Feedbackward neural network: These types of neural networks allow the flow of information bidirectionally, meaning both forward and backward. They use a

feedback mechanism to improve the learning process.

7.1 Convolutional Neural Network

Convolutional neural networks belong to a family of feedforward neural networks. One of the limitations with a fully connected neural network is the presence of huge amounts of parameters which in turn slows the computation and increases the odds of overfitting. To solve this problem CNN was developed, implementing the idea of kernels while sharing the parameters over the input image matrix with the help of a sliding window mechanism.

Convolutional neural networks are simply a series of different convolutional layers stacked over each other, with the later layers identifying more complex patterns than the initial ones. There are several types of layers in a CNN such as the dense layer, the pooling layer, a convolutional layer and so on[1]. CNN with the help of nonlinear activation functions identifies the complex patterns.

A simplified CNN architectural diagram as shown in figure below consists of a series of convolution and pooling layers followed by the Fully connected layers at the end of the network. Convolutional layers tackle the task of feature extraction with the help of linear functions (convolution operation) followed by the activation functions. Pooling layer assists in reducing the size of the image matrix, thus reducing the number of trainable parameters in the subsequent layers. It also helps in suppressing the effect of noise and distortions in the image.

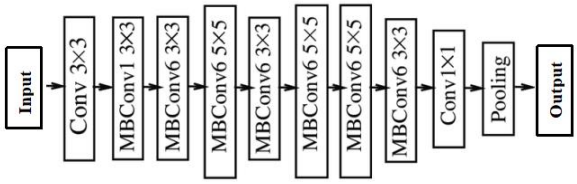

Fig 11: A typical EfficientNet-B3 architecture

7.2 Transfer Learning

The number of resources and data required to train deep learning models is enormous. To tackle this issue Transfer learning[3] applies the learning from one task to another identical task. It is a technique where a model trained on one task is then used for the second task with certain alteration. Transfer learning has been proved to be successful in the field of computer vision and image analysis. One of the major reasons for this is the fact that the initial feature extraction process is almost similar for different types of images which includes identifying edges, corners etc[8].

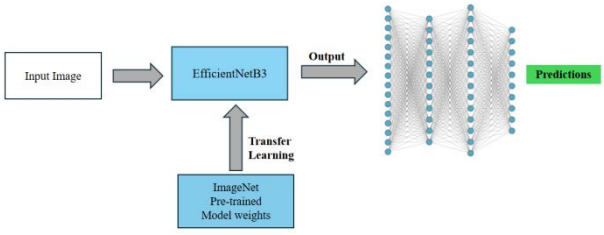

Fig 12: Architecture for Transfer learning with EfficientNetB3

In a general ConvNet Architecture the initial convolution and pooling layers perform the task of feature extraction and this learning can be utilized for any new task while the last fully connected layers do the classification, responsible for predicting various classes and that is subjective to the task at hand.

7.3 Training

This research paper uses the model 'EfficientNetB3' with the weights of 'ImageNet' model and further tune it to the training dataset of labeled tigers on both the original and augmented dataset. A new layer is added to incorporate the newly added information from our dataset. Two convolutional layers have been introduced with 256 and 98 neurons respectively.

7.4 Result

This paper with the help of transfer learning architecture has been able to achieve a validation accuracy of more than 85% while training on a dataset of 98 different tigers. The below images show the graphs of loss and accuracy at various epochs of training. Optimal validation results have been achieved at the end of 8 epochs.

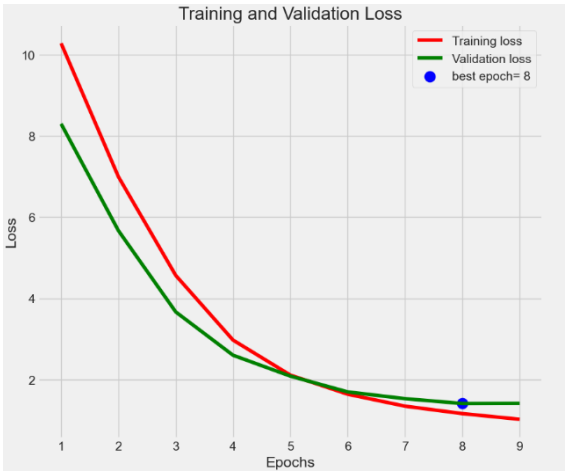

Fig 13: Training and Validation loss

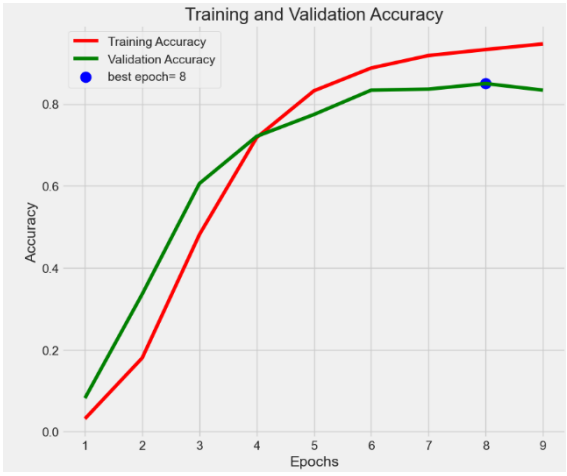

Fig 14: Training and Validation accuracy

Total Test Accuracy= (Sum of accurately predicted Tiger images)/(Sum of total images in test dataset)×100……(1)

On the reserve set of dataset, i.e., test dataset (808 images belonging to 98 different labels, out of 192 total tiger classes available, for which at least 15 unique images were available) which was kept aside for testing, we've achieved 88.49% accuracy. The table below shows classification report for a sample of 27 tigers. As apparent from the table, in most cases

the recall and precision values for a particular tiger class is greater than 0.8, which signifies reliable model performance.

TIGER CLASS	PRECISION	RECALL	F1 - SCORE	SUPPORT
T155 - Cubs	0.60	0.50	0.55	6
T15 - F	0.83	0.63	0.71	8
T134 - M	0.75	0.75	0.75	4
T157 - M	1.00	0.75	0.86	4
T14 - F	0.92	0.80	0.86	15
T158 - F	0.86	0.80	0.83	15
T126 - M	0.95	0.84	0.89	25
T155 - F	0.67	0.86	0.75	14
T32 - M	1.00	0.86	0.92	7
T12 - F	0.82	0.90	0.86	10
T49 - M	1.00	0.91	0.95	11
T5 - F	0.95	0.91	0.93	22
T49 - M	1.00	0.91	0.95	11
T5 - F	0.95	0.91	0.93	22
T130 - F	1.00	0.93	0.96	14
T137 - M	0.82	0.93	0.88	15
T14 - Cubs	0.70	0.93	0.80	15
T127 - F	0.83	1.00	0.91	5
T132 - F	1.00	1.00	1.00	3
T138 - M	0.86	1.00	0.92	12
T142 - M	0.79	1.00	0.88	19
T154 - F	1.00	1.00	1.00	12
T163 - F	1.00	1.00	1.00	12
T164 - M	0.90	1.00	0.95	9
T165 - Cubs	1.00	1.00	1.00	4
T4 - F	0.92	1.00	0.96	11
T51 - M	1.00	1.00	1.00	19

Fig 15: Testing classification report for a sample of 27 tigers

	PRECISION	RECALL	F1-SCORE	SUPPORT
ACCURACY			0.88	808
MACRO AVG.	0.89	0.85	0.86	808
WEIGHTED AVG.	0.89	0.88	0.88	808

Fig 16: Overall testing classification report

We've set up a threshold for the probability of 0.5 at which a tiger class will be predicted, if the highest probability for any class does not reach this threshold, the image will be classified as a '*new tiger*'. With sufficient '*new tigers*' and enough individual images of these tigers, this architecture can be employed to a bigger dataset and the model can be even more robust.

8. CONCLUSION

The data we've used in this paper belongs to a specific tiger reserve, and the results may vary if the geographical location varies for input data. It has been observed that with higher quality and more training data, there is a possibility of obtaining better results using Deep Learning. Future scope of this research can expand with the usage of images from various geographical background and including photos captured from a wide range of cameras so as to make the model more versatile for a better all round performance. It is important to keep in mind the possibilities of a CNN trained on the tiger images only, if we have comprehensive dataset that allows it. Next iterations of this research is going to focus on surpassing the performance with lesser images of individual tigers, which will widen the scope of the end goal of real-time tiger monitoring.

With this paper, we look forward to a better understanding of Tigers in their natural environment, and an improved human-wildlife interactions for a sustainable future.

9. REFERENCES

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

[2] Wikipedia: https://en.wikipedia.org/wiki/Scale-invariant_feature_transform.

[3] Shailendra Singh Kathait, Ashish Kumar, Piyush Dhuliya. Deep Learning-based Model for Wildlife Species Classification, IJCA Volume 186 – No.1, January 2024

[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] Cong Geng and X. Jiang, "SIFT features for face recognition," 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, 2009, pp. 598-602, doi: 10.1109/ICCSIT.2009.5234877..

[7] Brahm Dave, Meet Mori, Anurag Bathani, Parth Goel,Wild Animal Detection using YOLOv8, Procedia Computer Science,Volume 230,2023,Pages 100-111,ISSN 1877-0509

[8] M. H. Salem, Y. Li and Z. Liu, "Transfer Learning on EfficientNet for Maritime Visible Image Classification," 2022 7th International Conference on Signal and Image Processing (ICSIP), Suzhou, China, 2022, pp. 514-520, doi: 10.1109/ICSIP55141.2022.9887050.