VISIONARY DETECTIVES

GROUP - 1 PRESENTS....

“Wildlife Species Classification on Real-world Imagery From Camera Traps”

- Udaiveer Singh, Ramam Agarwal, Mukund Jha

AGENDA

- Problem Statement
- Literature Review
- Dataset and Feature Preprocessing
- Proposed Methodologies
- Performance Metrics and Deployability of the ML Solution

PROBLEM STATEMENT

PROBLEM STATEMENT

- Camera traps are widely used in wildlife conservation to monitor animal movement, behavior patterns, and ecological changes over time.
- However, conservationists face significant challenges in processing the vast number of images these traps generate, especially for identifying endangered or rare species. Manual labeling is time-consuming, error-prone, and often infeasible at scale.
- This creates a critical need for automated, accurate, and unbiased image classification systems that can support biodiversity monitoring, assist in early threat detection (e.g., poaching, invasive species), and ultimately enhance conservation efforts.

APPLICATIONS

- Real-time monitoring of wildlife populations.
- Supporting endangered species tracking in national parks and protected zones.
- Automating biodiversity surveys for ecological research.

POTENTIAL IMPACT

- Enhances conservation efforts with scalable and affordable AI solutions.
- Reduces manual effort, enabling wider deployment even in low-resource settings.
- Ensures fair species recognition, avoiding bias toward frequent species.

CHALLENGES, SOLUTIONS, AND IMPACT

- **Existing Challenges:**
  - Generate vast data, overwhelming manual analysis capacity.
  - Experts must manually sift through large datasets.
  - Many videos contain no animals, making filtering inefficient.

- **Our Solution:**
  - Use computer vision and deep learning for automation.
  - Implement species detection and classification
  - Apply object detection to streamline data processing and improve analysis.

- **Potential Applications:**
  - Wildlife Conservation – Real-time automated species monitoring.
  - Biodiversity Research – Understanding species behavior and ecosystems.
  - Anti-Poaching Efforts – Detecting human presence in restricted zones.

- **Potential Impact:**
  - Reduces human workload and analysis time.
  - Improves species identification accuracy.
  - Enables ecologists to focus on strategic conservation efforts.


LITERATURE
REVIEW

EXISTING RESEARCH ON WILDLIFE SPECIES CLASSIFICATION

1. Limited Classification Scope

Most studies focus on a single species rather than multiple species.

> **Abstract:** This paper presents the evaluation of 36 convolutional neural network (CNN) models, which were trained on the same dataset (ImageNet). The aim of this research was to evaluate the performance of pre-trained models on the binary classification of images in a "real-world" application. The classification of wildlife images was the use case, in particular, those of the Eurasian lynx (lat. "Lynx lynx"), which were collected by camera traps in various locations in Croatia. The collected images varied greatly in terms of image quality, while the dataset itself was highly imbalanced in terms of the percentage of images that depicted lynxes.

Source: Classification efficiency of pre-trained deep CNN models on camera trap images. Journal of Imaging, 8(2), 20 by Stančić, A., Vyroubal, V., & Slijepčević, V. (2022)

EXISTING RESEARCH ON WILDLIFE SPECIES CLASSIFICATION

2. **Unrealistic Datasets**

Some models achieve high accuracy by training on high-definition RGB images, which do not reflect typical camera trap conditions.

Figures: These images are from NACTI (North American Camera Trap Images) dataset introduced in “A deep active learning system for species identification and counting in camera trap images” by Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021)

EXISTING RESEARCH ON WILDLIFE SPECIES CLASSIFICATION

2. Unrealistic Datasets

In some papers, duplicate and unsuitable (blurred/unclear) files were excluded from the image dataset

> In the pre-processing phase, images were extracted from the video files, with a frequency of two images per second. The extracted images were added to the collection of the captured still images. Duplicate and unsuitable (damaged) files were excluded from the image dataset. All collected images were labeled according to the camera's location, designation, and timestamp. The last step in the pre-processing phase was the resizing of

Source: Classification efficiency of pre-trained deep CNN models on camera trap images. Journal of Imaging, 8(2), 20 by Stančić, A., Vyroubal, V., & Slijepčević, V. (2022).

EXISTING RESEARCH ON WILDLIFE SPECIES CLASSIFICATION

3. Computational Complexity

Constructing CNN models from scratch and using high-resolution images can be computationally expensive. Requires significant memory, storage, and processing power, making large-scale deployment difficult.

Abstract

1. Motion-activated cameras ("camera traps") are increasingly used in ecological and management studies for remotely observing wildlife and are amongst the most powerful tools for wildlife research. However, studies involving camera traps result in millions of images that need to be analysed, typically by visually observing each image, in order to extract data that can be used in ecological analyses.
2. We trained machine learning models using convolutional neural networks with the ResNet-18 architecture and 3,367,383 images to automatically classify wildlife species from camera trap images obtained from five states across the United States. We tested

Source: Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution, 10(4), 585-590 by Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., VerCauteren, K. C., Snow, N. P., ... & Miller, R. S. (2019).

HOW TO FIX THESE ISSUES

- **Multi-Species Classification**
  - Develop a model that classifies multiple species instead of focusing on a single species.
- **Train on Realistic Dataset**
  - Use diverse and challenging images (grayscale, blurred, and night-time images) to reproduce real-world applicability.
- **Optimizing Computational Efficiency**
  - Utilize pre-trained models and fine-tuning techniques to enhance accuracy while reducing processing time.

OUR DATASET

DATA COLLECTION

- Dataset is taken from the site drivendata.org
- We have obtained our dataset from Taï National Park, sourced from research conducted by the Wild Chimpanzee Foundation and the Max Planck Institute for Evolutionary Anthropology.
- The dataset consists of camera trap images capturing various species in their natural habitat.
- Our goal is to classify the species appearing in these images using computer vision and deep learning techniques.

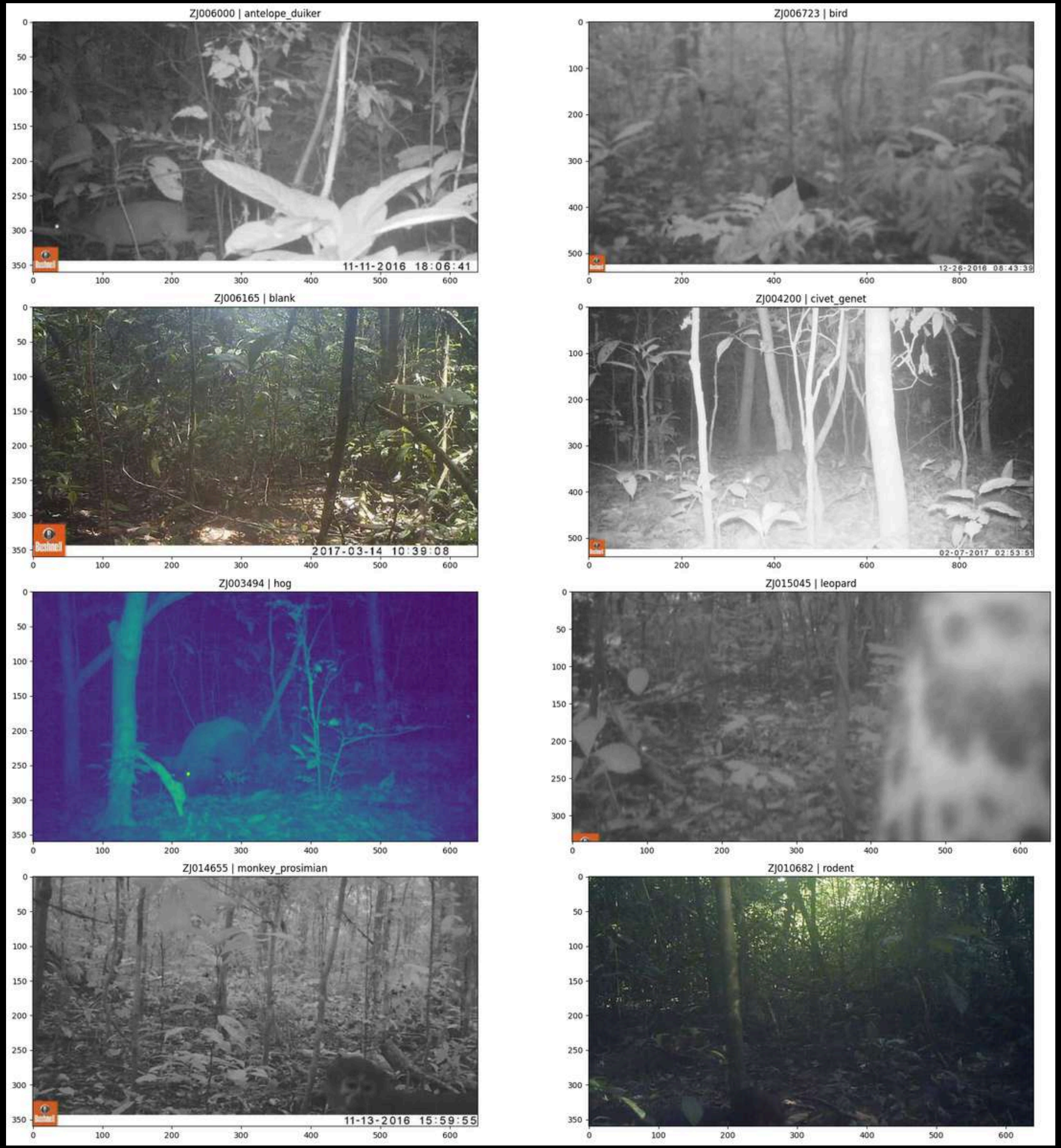


Figure: Samples from our dataset of each species along with image id and label

DATA COMPOSITION

Description of the Dataset

The dataset consists of 16448 camera trap images taken in Taï National Park, with associated metadata including capture location.

Key Features:

- **id** – Unique identifier for each captured image.
- **filepath** – Location where the image is stored.
- **site** – Unique identifier for the image capture location.

The dataset includes seven animal classes:

- Birds, Civets, Duikers, Hogs, Leopards, Monkeys, and Rodents
- It also contains some images without any animals.
- There is no site overlap between training and test datasets, requiring models to have strong generalization capabilities.

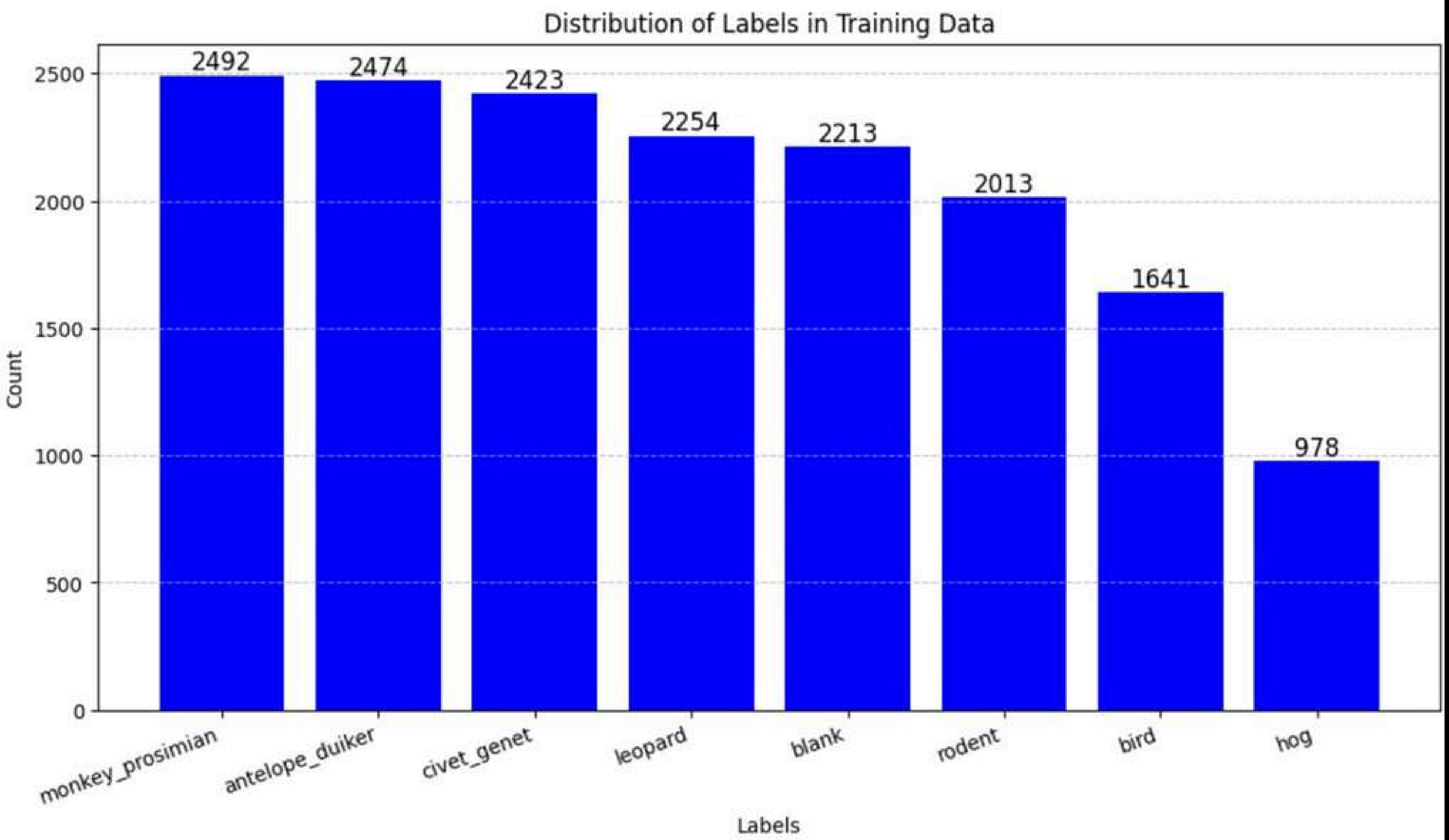


Figure: Bar Plot showing distribution of labels in training data

FEATURE PREPROCESSING

- **Image Size:** 224×224 image size chosen as the default size for optimal performance since it provides the best input resolution for the model.
- **Size Reduction:** Optimized image size for efficiency; EfficientNet-B0 is larger due to a higher FLOPs target.
- **One-hot encoding:** Each image has a single label, as the dataset is already one-hot encoded.

Table 1. **EfficientNet-B0 baseline network** – Each row describes a stage i with $\hat{L}_i$ layers, with input resolution $\langle \hat{H}_i, \hat{W}_i \rangle$ and output channels $\hat{C}_i$. Notations are adopted from equation 2.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	28×28	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Source: Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Data Augmentation – Random transformations are applied to each image during training, introducing variance to help the model generalize better to new data. This way, the model is always seeing something a little different than what it's seen before. This extra variance in the training data is what helps the model on new data.

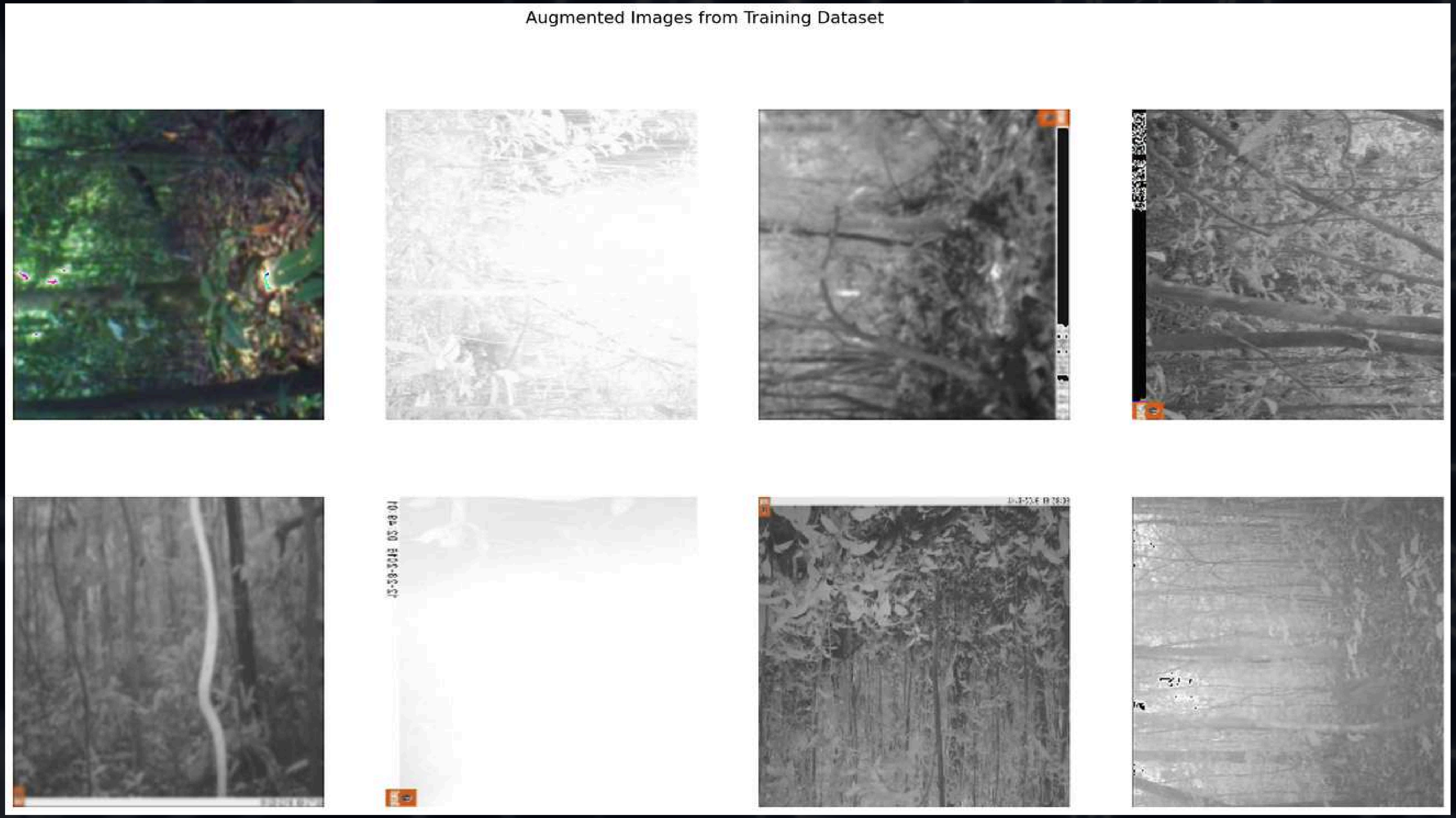


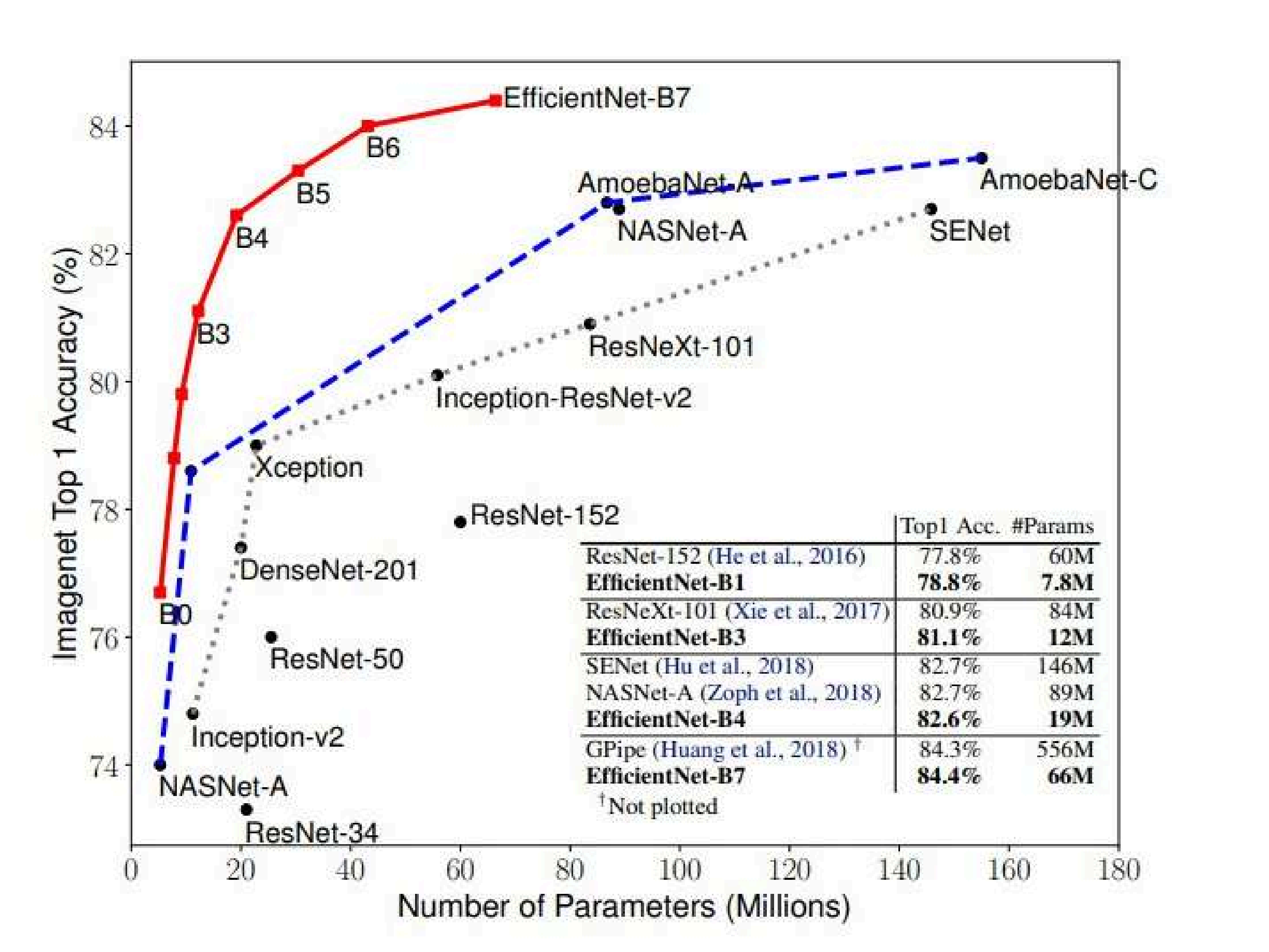
Figure: Sample of images after augmentation from training data


OUR ML
METHODOLOGY

SOME IMPORTANT THINGS WE IMPLEMENTED

Why EfficientNet?

- It is 8.4x smaller and 6.1x faster in inference than traditional ConvNets, making it highly efficient.
- EfficientNet has achieved 84.4% top-1 / 97.1% top-5 accuracy on ImageNet.
- EfficientNet has achieved state-of-the-art accuracy while using far fewer parameters than other deep networks like ResNet and Inception.



Source: Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

SOME IMPORTANT THINGS WE IMPLEMENT

Transfer Learning

- Used a pretrained base (EfficientNet_B0) for feature extraction and attached an untrained dense head for classification.
- The convolutional base extracts important image features, while the dense head made of neural network determines the image class.
- This is a proven method which reduces computational cost and is especially useful when we have limited data.

CHALLENGES

CHALLENGES

Evaluation Metric : Log Loss

- Log Loss is an extremely hard metric to deal with. It is highly sensitive to instances where the model makes very confident, but incorrect predictions.¶

- If the model predicts a high probability for the wrong class, the log loss penalty can be very high, even though the model might be right on many other instances.¶

- Therefore, accuracy was compromised to some extent to make the log loss as low as possible.

RESULTS

PERFORMANCE METRICS AND DEPLOYABLITY OF OUR ML SOLUTION

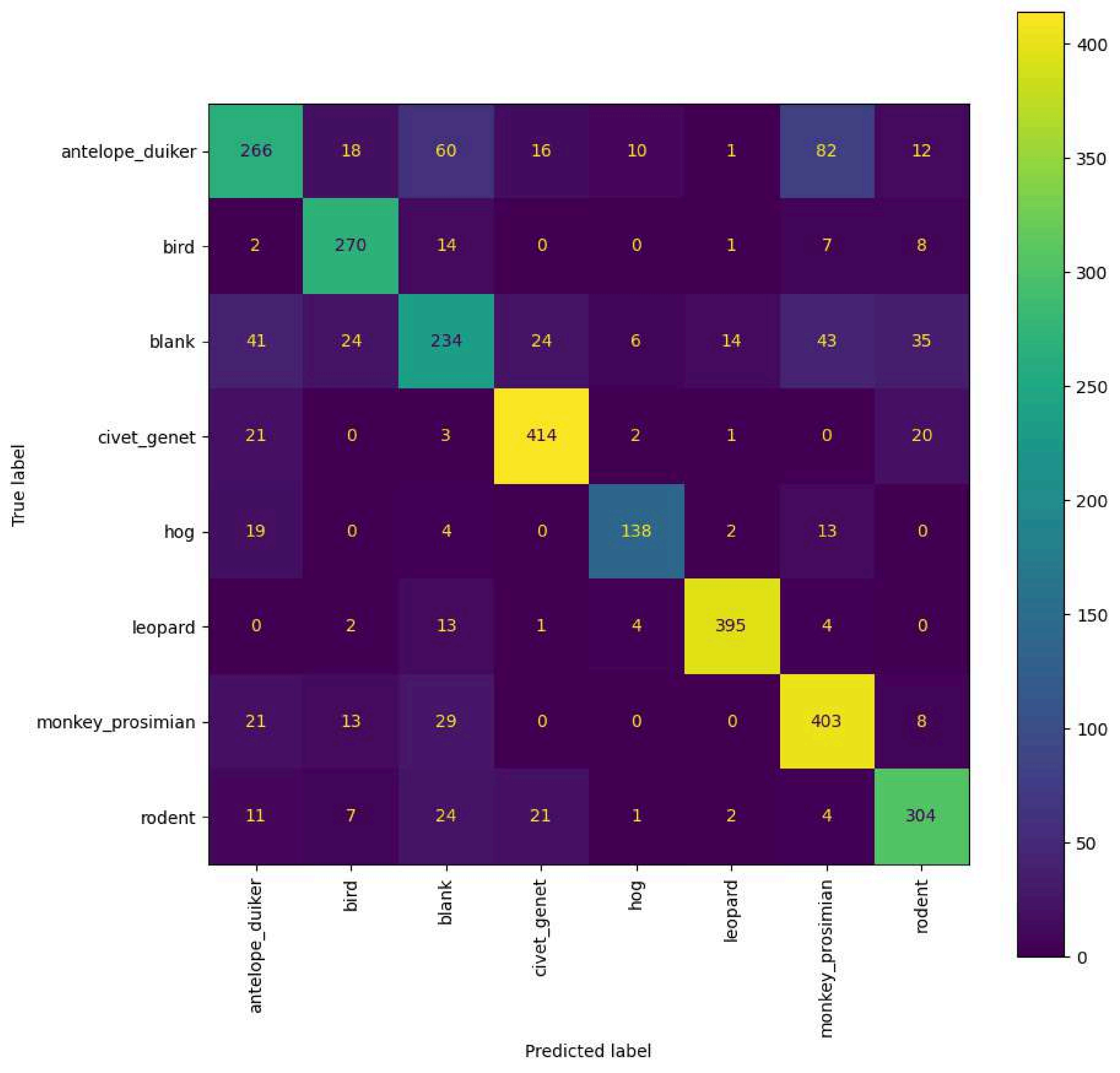
- We achieved Cross entropy loss/ Log loss of 1.2567 on test data which is quite high from the benchmark baseline model having log loss of 1.8210.

- Our competition is an advanced level practice competition. With over 1500 participants joined and 330 leaderboard ranks, we achieved #24 rank which makes us in top 7.2%

#20	ihebchemkhi99 4mo 1w ago · 8 submissions	1.1712
#21	bojesomo 2y 11mo ago · 10 submissions	1.2060
#22	StephenESchell 10mo 3w ago · 174 submissions	1.2098
#23	bob522297210 2mo 3w ago · 30 submissions	1.2327
#24	Visionary Detectives 9h 22min ago · 16 submissions	1.2567
#25	Shinda 12h 10min ago · 6 submissions	1.2711
#26	Albert_rich 9h 48min ago · 9 submissions	1.2785
#27	DeviPrasad 7mo ago · 8 submissions	1.2964

CONFUSION MATRIX OF OUR MODEL

- Despite focusing less on accuracy, we were able to predict most of the classes accurately.
- We achieved an accuracy of 78.3% which could have been easily improved had our metric not been Log Loss.



FEASIBILITY OF SOLUTION IN PLAKSHA

- Our solution cannot be implemented at Plaksha because there are not much animals within our campus to be detected by our model.
- Apart from this, there are no endangered species in our college premises to be identified and classified to primarily serve the purpose of our model.
- However, if our model was trained on images of domestic animals, we could have used the model for classifying the breed of dogs and other insects in our university.

IMPROVEMENTS OVER PREVIOUS RESEARCH

- **Multi-Species Classification** – Classifying 7 different species instead of focusing on a single species.
- **Realistic Dataset** – Includes blurred, poor-quality, and grayscale images to better reflect real-world conditions.
- **Computational Efficiency** – Uses a pretrained EfficientNet base with a dense head instead of building models from scratch, making it more computationaly efficient.
- **Optimized Model Selection** – EfficientNet achieves 84.4% top-1 / 97.1% top-5 accuracy on ImageNet while being 8.4× smaller and 6.1× faster in inference than the best existing ConvNet.

REFERENCES

1. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
2. Simões, F., Bouveyron, C., & Precioso, F. (2023). DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning. Ecological Informatics, 75, 102095.
3. Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., VerCauteren, K. C., Snow, N. P.& Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution, 10(4), 585-590.
4. Stančić, A., Vyroubal, V., & Slijepčević, V. (2022). Classification efficiency of pre-trained deep CNN models on camera trap images. Journal of Imaging, 8(2), 20.


Thank you