Bison Identification Using Deep Convolutional Ensembles: A Comparative Study

Abir Zaman
Department of Computer Science
University of Siegen
Email: Abir.Zaman@student.uni-siegen.de

***Abstract*—This paper investigates the effectiveness of modern deep learning architectures in identifying individual bison from a small, imbalanced dataset comprising 74 images across 9 classes. We evaluate several pre-trained convolutional neural networks (CNNs), each fine-tuned with data augmentation, and compare their performance to a weighted ensemble model. To optimize the ensemble, we employ a grid search strategy to select weights. Our results show that the ensemble model significantly outperforms all individual architectures—achieving a validation accuracy of 91.00%. These findings highlight the promise of ensemble learning in low-resource, fine-grained visual classification tasks.**

***Index Terms*—Image Classification, Ensemble Learning, CNN, Bison Identification, Transfer Learning, Grid Search**

I. Introduction

Individual animal identification plays a crucial role in ecological studies, behavior analysis, and wildlife monitoring. This research addresses the challenge of fine-grained visual classification of bisons in a highly imbalanced and low-data scenario using deep convolutional neural networks (CNNs) and ensemble strategies.

II. Dataset

The dataset comprises 74 labeled images of bison captured at various angles. Each subfolder corresponds to a unique bison (class). Due to class imbalance, the class *Quanna* was excluded from training and validation because it contained only a single image, which can affect model generalization and stratification.

III. Model Selection and Training

We trained and evaluated the following architectures using 5-fold stratified cross-validation with image augmentation:

- **EfficientNet-B0, B3:** Optimized for parameter efficiency.
- **ResNet-18, ResNet-50:** Incorporate residual connections for stable training.
- **ViT-B16:** Transformer-based architecture for visual representation.
- **InceptionV3:** Deeper CNN using factorized convolutions and auxiliary classifiers.
- **DenseNet-121:** Dense connections enable feature reuse across layers.

All models were initialized with ImageNet-pretrained weights and input images were resized to 224×224.

A. Ensemble Strategy

An ensemble model was constructed via weighted soft voting across all seven architectures. The ensemble prediction was defined as:

$$\text{Prediction} = \sum_{i=1}^{N} w_i \cdot P_i$$

where P_i is the prediction vector from model i and w_i is its corresponding weight.

B. Weight Optimization with Grid Search

To optimize the ensemble weights, grid search was performed with the following setup:

- **Weight Range:** For each model, weights were sampled from a continuous space using `np.linspace(0.1, 0.5, 5)`.
- **Search Space:** All possible combinations were generated using `itertools.product`.
- **Evaluation:** Each combination was evaluated using ensemble predictions on the validation set.
- **Selection:** The weight set yielding the highest validation accuracy was chosen as the final configuration.

IV. Results and Discussion

TABLE I: Comparison of Model Performance (K-Fold Validation)

Model	Train Loss	Val Loss	Train Acc.	Val Acc.
EfficientNet-B3	1.4036	1.6236	64.18%	54.00%
EfficientNet-B0	1.1302	1.4346	82.76%	58.00%
ResNet50	0.6948	1.2038	92.23%	60.67%
ResNet18	0.1757	1.1905	98.98%	68.95%
ViT-B16	0.0541	1.5493	98.99%	63.43%
InceptionV3	0.3806	1.1685	97.64%	60.76%
DenseNet-121	0.3836	1.0740	99.33%	72.95%

The grid-searched ensemble model achieved an accuracy of 91.00%, a significant improvement over the highest-performing individual model (DenseNet-121 at 72.95%). The method effectively reduced misclassifications, especially in underrepresented classes. The class *Quanna*, only one image, remained part of the label space but contributed minimally to training and validation due to its limited representation.

TABLE II: Classification Report – Precision, Recall, F1-Score Per Class

Class	Precision	Recall	F1-Score	Support
Elsa	1.00	0.71	0.83	7
Faye	0.85	1.00	0.92	17
Quaida	0.83	0.71	0.77	7
Quandra	1.00	0.83	0.91	6
Quanna	1.00	1.00	1.00	1
Quara	1.00	0.93	0.96	14
Quino	0.90	0.90	0.90	10
Quirinius	1.00	1.00	1.00	8
Quirly	0.67	1.00	0.80	4
Accuracy		0.91		74
Macro Avg.	0.92	0.90	0.90	74
Weighted Avg.	0.92	0.91	0.91	74

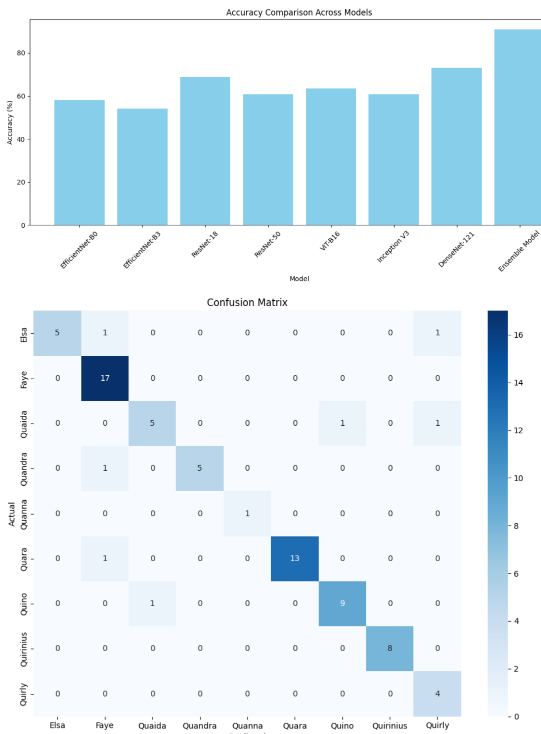

Fig. 1: Accuracy and Confusion Matrix (Ensemble Predictions)

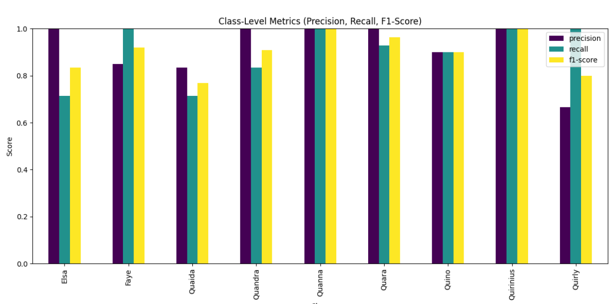

Fig. 2: Precision, Recall, F1-Score by Class

V. Conclusion

This study demonstrates the effectiveness of deep learning ensembles enhanced with grid search for fine-grained animal identification in low-data, imbalanced settings. The proposed strategy combines model diversity with optimized weighting to outperform all individual classifiers. Future work will explore meta-ensembling strategies and few-shot learning frameworks for even better generalization under data-scarce conditions.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

[3] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.