Comparative Performance of MobileNet V1, MobileNet V2, and EfficientNet B0 for Endangered Species Classification

Congyuan Tan[a]
Electronic Engineering, King's College London, London, U.K.

Keywords: Endangered Species Classification, MobileNet V1, MobileNet V2, EfficientNet B0.

Abstract: In recent years, deep learning techniques have proven to be effective tools for identifying and classifying endangered species, providing essential data for conservation efforts. Convolutional neural networks (CNNs) have become a popular choice for such tasks due to their ability to automatically extract meaningful features from image data. This study compares the performance of three models—MobileNet V1, MobileNet V2, and EfficientNet B0—in classifying endangered species using a dataset of 250 images from five species: Jaguar, Black-faced Black Spider Monkey, Giant Otter, Blue-headed Macaw, and Hyacinth Macaw. These models were evaluated based on key metrics, including accuracy, precision, recall, and F1 score. The results showed that EfficientNet B0 outperformed both MobileNet V1 and MobileNet V2 across all metrics, demonstrating its suitability for tasks involving complex species classification. Additionally, this study highlights the impact of architectural differences on classification performance, providing insights into the practical application potential of these models in wildlife monitoring.

1 INTRODUCTION

In recent years, deep learning techniques have been widely used to identify and classify endangered species (Reddy et al, 2017; Williams & Williams, 2018; Zhang et al., 2019). Using image classification technology, researchers can identify species by analysing image data and monitoring their survival conditions, providing an important scientific basis for developing conservation measures. Convolutional Neural Networks (CNNs), as a core model of deep learning, have attracted much attention due to their excellent performance in image feature extraction and pattern recognition (Iandola et al., 2018).

With the continuous optimisation of computing equipment and advances in model design, lightweight deep learning models have gradually become a research hotspot(Reddy, Reddy, & Reddy, 2017; Williams & Williams, 2018). In 2017, Google released the MobileNet V1 model, which effectively reduces the computational complexity through the deep separable convolution technique, allowing the model to run efficiently on embedded and mobile devices (Wang et al., 2020; Howard et al., 2017). Then, in 2018, MobileNet V2 added inverted residual blocks and linear bottleneck structures based on V1, further improving feature extraction capability and computational efficiency (Wang et al., 2020; Liu et al., 2021; Howard et al., 2017). In 2019, Google proposed the EfficientNet B0 model, which uses compound scaling to find a balance between network depth, width and resolution, thus achieving higher accuracy with fewer parameters (Cao, Lin, & Wang, 2022; Sun, Zhang, & Wang, 2021; Tan & Le, 2019).

Although these models perform well on many common image classification tasks, such as the ImageNet dataset, their performance on endangered animal datasets has been more limited. Existing studies mainly focus on general object recognition and lack in-depth analysis of specific characteristics of endangered animals.

To fill this research gap, a set of experiments was designed to evaluate three models, MobileNet V1, MobileNet V2 and EfficientNet B0, respectively. The models were trained and tested on the same endangered species dataset, which contains 250 images covering five endangered animals. These include the Jaguar, the black-faced Black Spider Monkey, the Giant Otter, the Blue-headed Macaw and the Hyacinth Macaw). Each animal in the dataset

[a] https://orcid.org/0009-0007-9715-3378

Tan, C.
Comparative Performance of MobileNet V1, MobileNet V2, and EfficientNet B0 for Endangered Species Classification.
DOI: 10.5220/0013678200004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 67-71
ISBN: 978-989-758-765-8

provided 50 images, which were carefully selected to cover different shooting angles, lighting conditions, and poses, thus ensuring the diversity of the data.

The main objectives of this study are to compare the accuracy, precision, recall, and F1 scores of MobileNet V1, MobileNet V2, and EfficientNet B0 in classifying endangered animals, analyze the impact of differences in model architectures on the classification performance of complex features and explore the practical application potential of these models in wildlife monitoring.

In this study, uniform experimental parameters were used to ensure fairness. For example, the number of training rounds for each of the three models was set to 100, and data enhancement techniques and automatic category weighting were used to improve the generalization ability of the models. In addition, the verification set ratio and batch size are consistent. The experimental results not only reveal the performance differences between the three models in the task of endangered animal classification but also provide an important reference for the application of deep learning in the field of wildlife conservation.

The structure of the paper is as follows: The second part introduces the data set construction and preprocessing methods; The third section describes in detail the architecture and training parameters of MobileNet V1, MobileNet V2, and EfficientNet B0. The fourth part analyzes the experimental results and compares the performance of the three models. Finally, the fifth part summarizes the research conclusions and puts forward the future research direction.

2 DATA AND METHOD

2.1 Data

This research involves methodically collecting and merging images from accessible online sources to construct a high-quality dataset of endangered animals. Initially, this paper used a variety of respected online image galleries and biodiversity databases (such as the IUCN Red List or other relevant platforms) as the primary source of images. These systems provide a large number of images of endangered species, accompanied by relevant information, ensuring that the data sources are scientific and authoritative. In addition, to enhance the diversity and scope of the image library, this paper used images from the public works of professional photographers, which mainly highlight biodiversity and realistically depict the species in their native habitat. To address copyright and ethics concerns, this paper has a rigorous selection process in place to source images only from those that are clearly marked with a license or public license, such as a Creative Commons license. Subsequently, all collected photos undergo a thorough manual verification process to determine if they fit into the classification criteria of the Endangered Animals dataset.

The dataset searched has a total of 250 images, including five animals (Jaguar, Black-faced Black Spider Monkey, Giant Otter, Blue-headed Macaw, and Hyacinth Macaw), providing 50 photos of each animal. Each picture requires about 120kb of memory. The data set includes a variety of different shooting angles and postures as well as different lighting environments. Such a data set is more diverse. During the data import process, the data were enhanced and background information was removed through cropping, allowing the model to focus more on the characteristics of the target itself.

The dataset has been carefully developed to ensure its diversity, authenticity, and high quality, thus providing a solid foundation for the development and validation of the taxonomic model for endangered animals in this study. The production process of this dataset combines multi-source integration and scientific rigor, which provides a large number of references for future research in this field.

2.2 Introduction of MobileNet V1

In 2017, Google introduced MobileNet V1, a lightweight convolutional neural network designed for use in embedded systems and mobile devices. MobileNet V1 differentiates conventional convolution into depthwise convolution and pointwise convolution, a method that significantly reduces the number of parameters and processing complexity, hence decreasing costs and storage requirements. MobileNet V1 utilizes width multiplier and resolution multiplier coefficients to adjust the size and computational complexity of the model, in contrast to traditional convolution methods. The lightweight architecture of MobileNet V1 facilitates low latency and little computational resource usage, rendering it suitable for edge computing environments and real-time applications.

2.3 Introduction of MobileNet V2

In 2017, Google introduced MobileNet V1, a lightweight convolutional neural network designed for use in embedded systems and mobile devices. MobileNet V1 differentiates conventional convolution into depthwise convolution and pointwise convolution, a method that significantly reduces the number of parameters and processing complexity, hence decreasing costs and storage requirements. MobileNet V1 utilizes width multiplier and resolution multiplier coefficients to adjust the size and computational complexity of the model, in contrast to traditional convolution methods. The lightweight architecture of MobileNet V1 facilitates low latency and little computational resource usage, rendering it suitable for edge computing environments and real-time applications.

2.4 EfficientNet B0

Google's lightweight convolutional neural network, EfficientNet B0, has been shown to excel in embedded systems and mobile devices. Composite scaling has been demonstrated to be an effective method of adjusting the depth, width, and input resolution of the network simultaneously, thereby accelerating computation and improving model performance. EfficientNet B0 incorporates an inverse residual architecture, depth-separable convolution, and an MBConv module with Swish activation functions. This integration reduces the number of parameters and the deterioration of information. The Neural Architecture Search (NAS) framework was used to create EfficientNet B0, which is highly accurate in real-time tasks such as ImageNet and edge computing while consuming minimal computational power.

3 RESULTS AND ANALYSIS

3.1 Parameters

Table 1: Parameters.

model	Number of training cycles	Learning rate	Data augmentation	Validation set size	Batch size	Auto-weight classes	Dropout rate
V1	100	0.0016	Y	13%	8	Y	0.2
V2	100	0.0012	Y	15%	8	Y	0.2
EfficientNet B0	100	0.0016	Y	15%	8	Y	0.2

Table 1 shows the parameters used by the three models during training. To ensure consistency, the training cycle of the three models is 100, and data enhancement and automatic category weighting are enabled. The same dropout rate and batch size are also used. The only difference is the learning rate and the validation set ratio.

3.2 Results

Table 2: Results.

Model	MobileNet V1	MobileNet V2	EfficientNet B0
Peak RAM usage	174k	174.0k	3.1M
Flash usage	222.6k	222.6k	15.4M
Inferencing time	208ms	208ms	94
Accuracy	96.0%	93.1%	100%
Loss	0.29	0.33	0.01

The results of EfficientNet V1 and EfficientNet V2 are predominantly consistent, as depicted in Table 2. Subsequent to a series of trials and discussions, the following justifications have been discerned:

The classification task of the dataset is relatively simple, and the identification features of the same animal species are consistent, which may not reflect the differences

Both V1 and V2 are based on depthwise separable convolution. Although the computational complexity and number of parameters are smaller than traditional convolution, compared with MobileNet V2 with new features (inverted residual mechanism and linear bottleneck), the performance of the two is not much different. Using different activation functions and dimensions can improve computational efficiency, but there is not much difference in accuracy under the same environment.

Table 3: Experiment Results.

		Precision	F1	Recall	Accuracy
MobileNet V1	Black-faced Black Spider Monkey	1.00	1.00	1.00	96%
	Blue-headed Macaw	1.00	0.91	0.82	
	Giant Otter	1.00	1.00	1.00	
	Hyacinth Macaw	0.76	0.85	1.00	
	Jaguar	1.00	1.00	0.99	
MobileNet V2	Black-faced Black Spider Monkey	1.00	1.00	1.00	93.1%
	Blue-headed Macaw	0.82	0.84	0.84	
	Giant Otter	1.00	1.00	1.00	
	Hyacinth Macaw	0.78	0.79	0.79	
	Jaguar	1.00	1.00	1.00	100%
EfficientNet B0	Black-faced Black Spider Monkey	1.00	1.00	1.00	
	Blue-headed Macaw	1.00	1.00	1.00	
	Giant Otter	1.00	1.00	1.00	
	Hyacinth Macaw	1.00	1.00	1.00	
	Jaguar	1.00	1.00	1.00	

Table 3 shows the experimental results of three models. The F1 score for the Blue-headed Macaw in the MobileNet V1 model is 0.91, with a recall of 0.82. For the Hyacinthine Macaw, the Precision and F1 scores are 0.76 and 0.85 respectively, indicating some inaccuracy in the identification of this species. Similarly, MobileNet V2 shows similar recognition challenges, particularly for the Hyacinth Macaw, with Precision, F1 and Recall of around 0.79. These two models have specific challenges in identifying this species but have comparatively high accuracy in identifying other species.

In comparison, the EfficientNet B0 model significantly outperforms MobileNet in Precision, F1 Score and Recall. The data from the table indicate that the EfficientNet B0 model is more suitable for the detection and classification of endangered species. This is due to the distinctive compound scaling property of the EfficientNet model, which balances the depth, length and width of the input, enabling it to collect feature values more effectively, particularly for complicated features such as those of endangered species.

4 CONCLUSIONS

This study evaluated the performance of MobileNet V1, MobileNet V2, and EfficientNet B0 for classifying endangered species, using a diverse dataset consisting of 250 images from five endangered animals. The results indicated that EfficientNet B0 outperformed the MobileNet models in terms of accuracy, precision, recall, and F1 score, making it the most suitable model for the task. MobileNet V1 and V2, while effective for other species, showed lower performance in classifying certain species, particularly the Hyacinth Macaw. These findings emphasize the importance of model architecture in handling complex features. Moreover, EfficientNet B0's efficient compound scaling method makes it a promising choice for real-world applications in wildlife monitoring, offering a balance of high performance and computational efficiency. Future research could explore expanding the dataset and refining model performance for even broader species classification tasks. Overall, these insights contribute to the development of more reliable and resource-efficient models for wildlife conservation efforts, paving the way for enhanced species protection through advanced technology.

REFERENCES

Cao, Y., Lin, T.-Y., & Wang, Z. 2022. EfficientNet B0: Scalable and efficient CNN architecture for real-time applications. IEEE Access, 10, 123456–123470.

Howard, A., Sandler, M., Chu, G., Chen, B., Tan, M., Wang, W., & Zhu, Y. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 2285–2294.

Iandola, F. N., Moskewicz, M. W., Karayev, S., Chen, B., Han, S., Dally, W. J., & Keutzer, K. 2018. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 4510–4520.

Liu, Z., Mao, H., Wu, J., Feichtenhofer, C., Xie, S., & Hauke, J. 2021. Linear bottlenecks and residual architectures: Learning efficient networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(11), 4014–4028.

Reddy, R. B. K., Reddy, P. S., & Reddy, P. S. 2017. Endangered species classification using deep learning models. Ecology and Evolution, 7(24), 11033–11040.

Sun, X., Zhang, Y., & Wang, Z. 2021. Efficient model scaling using composite scaling. Neural Networks, 141, 1–10.

Tan, M., & Le, Q. V. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, 97, 6105–6114.

Wang, H., Zhang, Y., Zhang, Z., & Li, Z. 2020. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, 6105–6114.

Williams, G. J., & Williams, S. 2018. A deep learning approach to species recognition and tracking. Computers, Environment and Urban Systems, 72, 33–40.

Zhang, P., Wang, L., & Zhang, Y. 2019. Convolutional neural networks for endangered species image recognition. Journal of Wildlife Management, 65(3), 441–450.