Enhancing Bird Species Classification with PaliGemma and Custom EfficientNet Architectures

Medhovarsh Bayyapureddi[a], Lakshmi Sri Harshitha [b],Anjali T[c]
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Amritapuri,India

amenu4cse21064@am.students.amrita.edu[a],amenu4cse21018@am.students.amrita.edu[b],anjalit@am.amrita.edu[c]

***Abstract*—Bird species classification is a critical task in biodiversity studies, ecological research, and conservation efforts. Our research presents a comprehensive approach to bird species classification using a custom-modified EfficientNet-B0 model, enhanced with a Squeeze-and-Excitation (SE) block to improve feature sensitivity. The dataset used comprises a diverse collection of bird images, which are preprocessed and split into training, validation, and test sets. The custom model incorporates the SE block into the EfficientNet-B0 structure, with a dropout layer and a connected layer added for classification.**

Using an additional Squeeze-and-Excitation (SE) block to improve feature sensitivity and a specifically tailored EfficientNet-B0 model, our work provides a comprehensive approach to bird species classification. The dataset consists of several hundred preprocessed images of birds, split into test, validation, and training sets. For classification, the dropout layer and connected layer are added, and the custom model integrates the SE block into the EfficientNet-B0 framework. In our model design, the learning rate is adjusted as needed using a learning rate scheduler, and SGD is employed for model training.

The findings show that the custom EfficientNet-B0 model boosts classification accuracy highlighting how well the SE block works to improve model performance.The paper shows deep learning methods for animal tracking and conservation. Index Terms— Convolutional Neural Networks (CNNs), Squeeze-and-Excitation block, Deep learning, EfficientNet-B0,Fine-grained Image Classification Bird Species Classification neural networks wildlife monitoring pali jewel

***Index Terms*—Bird Species Classification, Deep Learning, EfficientNet-B0, Squeeze-and-Excitation Block, Image Classification, Neural Networks, Wildlife Monitoring, Fine-grained image classification, Convolutional Neural Networks (CNN), Pali Gemma**

I. INTRODUCTION

Background Categorizing birds is a requisite step for researching biodiversity, ecology and wildlife conservation strategies (Lepage et al(KeyEvent I). Bird species can help provide key indicators of the wellness and status of ecosystems, specific bird habitats and how (if?) they are being impacted by environmental change. Typically, such bird species had been classified based on traditional approaches that demanded human observation and specialist knowledge—and thus were labor-intensive as well as prone to error. The advent of deep learning and advanced image classification techniques has opened new opportunities to make bird species identification faster and more accurate.

Over the past few years convolutional neural networks (CNNs) have become go-to tools for image classification. They're good at learning complex features from big datasets. One CNN that's caught people's eye is EfficientNet. It stands out because it balances accuracy and efficiency well. What makes EfficientNet special is its compound scaling method. This method scales all dimensions - depth, width, and resolution - evenly using a set group of scaling coefficients.

It performs isotropic scaling of all dimensions (depth, width, and resol- ution) using a fixed set S of group-specific shuttering coefficients. This method scales each of depth, width and resolution using a dedicated set group of scaling coeffi- cients. The paper is based on the assignment of class labels consisting of multiple bird species, and accordingly explores an approach for classification combining these classes. The compact design with superior performance of EfficientNet-B0 is well known. To enhance the model's feature detection and classification accuracy, a Squeeze-and-Excitation (SE) block is added to the EfficientNet-B0 structure. The SE block adjusts channel-wise feature responses by highlighting the relationships between channels, allowing the network to focus on the most important features.

The first step is to create a dataset of bird images. Resize images and normalize them to standardize these as an input, while adding variety — make the NN quickly effective both training/test phases. The model is then finetuned and trained with Stochastic Gradient Descent (SGD) using momentum. A learning rate scheduler is used to change the learning rate over time in itertools.

Different metrics like accuracy, loss, confusion matrices and classification reports are considered to analyze the model performance. The same evaluation measures are performed on the training and validation sets of 100 as a sanity test for this. Plus, the test data has been evaluated to check whether generalisation ability of model stands.

(1) The introduction of the original EfficientNet-B0 model with a new SE block for improved bird species identification; (2) Comprehensive testing and evaluation results on various cases to verify its classification performances in bird recognition task;(3) Illustrative examples showcasing how our

proposed approach can be effectively utilized as an useful appliance not only for scientific consideration but also wildlife conservation efforts.

The subsequent sections outline the bird image dataset, explain how to build a custom EfficientNet-B0 model and describe training and testing phases in addition to their results. The findings suggest that the method is successful with high bird classification accuracy, showing promise for deep learning as a means to improve species identification.

A. Evolution of Image Classification with Deep Learning

Image classification, being one of the major sub field (sub-domain) has shown incredible transformation post-deep learning era from hand-crafted feature based method to data-driven approach. The field was initially overrun by manual handcrafted features such as the Scale-Invariant Feature Transform (SIFT) and Histograms of Oriented Gradients (HOG) that required intensive human labor, which undermined generalizabilty. The introduction of Convolutional Neural Networks (CNNs) changed this significantly, with AlexNet winning the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). This showed that deep learning models, in general, could benefit from learning hierarchical features directly from the pixel values as traditional methods were outperformed and gave a catalyst to research into CNN architectures.

Later improvements gave rise to architectures such as VGGNet, GoogLeNet and ResNet which increased the number of layers in neural networks (depth), introduced new architectural features like Inception module designs or residual connections and utilized advanced training techniques. These enhancements also lead to the deep learning models becoming more sophisticated and being able to perform image classification tasks that were previously impossible for high-quality or fast-enough output. Recent models have further improved performance by carefully balancing network depth, width, and resolution to obtain the state of record results while having fewer parameters and lower computational costs (e.g., EfficientNets).For example, technologies such as transfer learning [**?**]adapt pre-trained weights [1]) and data augmentation (generate new training examples) are now a standard practice in the field [1].The rise of deep learning models has transformed how birds are classified as species. These models can detect all the minute differences between birds, which makes it very challenging for the traditional methods to solve. To address the complexity, Squeeze-and-Excitation (SE) blocks and dropout layers on top of EfficientNet-B0, a transfer learning model [1], have been employed not only to predict bird species but predict it with high accuracy irrespective of the datasets. In summary, deep learning for image classification has matured to a high degree of standardization and robustness, and, has, therefore, turned to be the most chronic answer to the challenge – automation and greater functionality with higher precision and sensitivity for every complex domain-specific such needs like bird species identification.

B. Bird Species Classification

Of late, deep learning, specifically convolutional neural networks (CNNs), has gained great success for bird classification. This is the problem of identifying and categorizing the underlying bird species given a test image [2]. This is a very challenging problem for machines due to large intra-class variations in the appearances and environments of birds. In the dataset used for the project, there are 225 different bird species, with a large number of images of each species, which will further enhance the training process.Model: EfficientNet-B0, enables efficient scaling and high performance on many benchmarks related to image classification tasks.

EfficientNet-B0 use compound scaling method, Scaling all dimensions under constraint with fixed set of scaling coefficients This yields a model which performs well both regarding accuracy and efficiency, making it suitable for Bird species classification tasks. This was further streamlined through data augmentation, along with the aforementioned transfer learning and model training from pre-trained weights on ImageNet aids to both better feature extraction as well as model convergence.

High accuracy of this model, in a task that is itself very challenging highlights the capability of our deep learning models in handling complexity involved with bird species classification aiding ornithologists and conservationists. Automation of the classification process by such models makes them very useful for biodiversity research or ecological monitoring granting detailed understanding on populations and distribution of birds. This method not only saves time and resources but also increases the accuracy and consistency of species identification.

II. RELATED WORK

The application of deep learning approaches, particularly Convolutional Neural Networks (CNNs) such as EfficientNet, has dramatically improved the state of the art in bird species classification. EfficientNet, enhanced with Squeeze-and-Excitation (SE) blocks to increase feature sensitivity, leads to higher classification accuracy. These models can be successfully leveraged by fine-tuning pretrained networks on specific bird datasets and applying data augmentation techniques to achieve optimal performance. Typically, these models are evaluated using metrics like accuracy and confusion matrices. Recent advances, including attention mechanisms and ensemble learning, have further demonstrated improvements in classification results.

Deep learning has proliferated, particularly in the context of hybrid models. Other varieties of models like the hybrid CNN-RNN model have made automated bird species identification nearly work in practice. These models create large-scale, accurate instruments with far-reaching applications for species monitoring and conservation. Manna et al. used a dataset of 400 bird species, explored four transfer learning models — ResNet152V2, InceptionV3-DenseNet201 and MobileNet-V2 — in comparison [3]. They find the DenseNet201 to have a good balance of accuracy (95.09%) and loss, scaling up from 30 base images per batch to about fifty with adequate GPU memory for such an upscale model. The primary tasks to be addressed are transfer learning with unfreezing and pre-trained weights, refining the training procedure within itself by adding epochs, integration of these trained models in mobile application for real-time bird identification.

Bipin Kumar Rai, Shivani Sharma, Gautam Kumar and Kaushal Kishor used deep learning and CNNs to identify bird species in another study [4]. Classic schemes are largely based on expert knowledge, which is restricted both by human and computer vision capabilities. To tackle the above challenges, this study proposed a system based on CNNs [5] where images are processed by convolutional and pooling layers. This system effectively and efficiently extracts, refines the key-features like shape, size of image elements leading to better quality images for accurate species prediction. The research demonstrates the benefits of using deep learning together with image processing to provide an accessible and accurate tool for bird watchers as well researchers in a new field.

Similarly, Gupta et al. examined the fusion of Deep Learning and Image Processing with Bird Species Identification [6]. They emphasized more on usage of CNNs [7] to enhance classification accuracy. Due to this architecture, CNNs are well suited for the task of image processing and classification where they apply kernels on each layer creating feature maps. This method improves the accuracy of image-to-bird-species correlations drastically. These are then followed by image processing techniques that enhance the quality of an image and make essential parts more prominent. In this regard, the synergy between CNNs [8] and image processing emerges as a valuable technology able to contribute profoundly in broadening horizons of ornithology by facilitating robust and user-friendly means for differentiating various species.

Alswaitti et al. conducted one of the few studies in which a direct comparison group was used [9]. compared traditional machine learning models, deep learning-based models and transfer-learning based approaches in their performance for bird species classification with the Kaggle-180-birds dataset. Their solution shows the best performance with accuracies between 98% and 100%, for different species, using CNNs through transfer learning. Future research may focus on multi-source transfer learning settings, strategies to alleviate negative transfer and deploy these models in bird identification systems. In the future, it may be a good idea to increase greatly (to 100 or more) number of images per each species which should even boost performance further.

Lastly, Zhang et al. used ShuffleNetV2 to present a lightweight CNN model for bird identification [10]. This model has a Coordinate Attention, Squeeze-and-Excitation, and multi-channel feature fusion module. Subsequent advancements may focus on improving accuracy problems that arise with an increasing number of bird species and improving real-time performance on low-cost devices. These findings demonstrate the ongoing progress made in the classification of bird species using cutting-edge deep-learning methodologies [11] [12].

III. Data Collection and Analysis

Fig. 1. Bird Dataset

A great deal of data was collected during our study, which consisted primarily in choosing and arranging the respective images from 100 bird species. In order to determine this, the data was segmented into training/validation/testing sets in such a way as not only ensure that there is balanced representation of all three categories within each sub set - and therefore allow for ideal model development - but also, overall representative examples i.e. out-of-distribution samples are split evenly across these 3 classes or datasets. The training set is used to train the model, The validation set for tuning hyperparameters and finally - will test on a testing-set.

Data Preparation The images were standardized to 224 x 224, as required by the EfficientNet-B0 model. To increase data diversity and prevent overfitting, several transformations were performed on the training dataset. These transformations included random cropping, horizontal flipping, and normalization.

Data Analysis Data analysis was a multi-staged process. Dataloaders with a batch size of 32 were created to feed the data in batches, ensuring efficient memory and processing management during training and evaluation. Shuffling the training data helped make it more challenging for the model to associate specific images with their correct labels.

During the training process, metrics such as training loss, validation loss, training accuracy, and validation accuracy were tracked. This allowed us to understand how well the model was learning and generalizing. The training spanned 10 epochs, with a learning rate scheduler (StepLR) scaling every seven epochs to refine optimization.

Performance Evaluation The performance evaluation was done on the test data set, which model never seen during training or validation. The model was evaluated based on key metrics such as test loss and the accuracy of test. In addition, a confusion matrix and a classification report were carried out to provide detailed analysis of the correct prediction separated by bird species. Identifying any particular classes where the performance of our model can be enhanced and these tools in turn guide us for enhancements done by during training.

IV. Proposed Methodology

1.Data Collection and Processing Initially, an individual comprehensive dataset of birds' species images is collected (fig. 1). This dataset is then split into training, validation and test sets. Different Data Augmentation techniques are used, to make the model learn properly from training data and generalize well. Some of these techniques are, for example reshaping the images to 224x224 pixels and converting them into tensors. Data augmentations consist in adding various transformations to the training images like rotation, flipping or color distortions helping the model learn bird species in different conditions [13];

2.Model Architecture The foundation of the proposed technique is constructed upon a pre-trained EfficientNet-B0 model with ImageNet. EfficientNet-B0 is chosen for its efficient parameter scaling and high performance. A Squeeze-and-Excitation (SE) block is included for more effective feature extraction [14], recalibrating channel-wise feature responses to enable the model to focus on important features. Additionally, a dropout layer, set to 30%, is added along with three Conv2D layers. The final layer of EfficientNet-B0 is replaced by a dense, fully connected output layer tailored to the number of bird species classes, allowing the model to effectively recognize different birds

3.Training Procedure The model is trained using the Cross-Entropy Loss function, which performs well for multi-class classification tasks [15]. In the initial training stage, the Stochastic Gradient Descent (SGD) optimizer is employed with a learning rate of 0.003 and momentum of 0.9, alongside a weight decay of 1e-4. A StepLR scheduler with a step size of 7 and a gamma of 0.1 is used to dynamically adjust the learning rate during training. These hyperparameters are carefully selected to optimize training and ensure that the model converges effectively, resulting in improved performance.

4. Training and Validation Loop The training process is conducted over multiple epochs. During each epoch, the model is trained on the training dataset, with loss and accuracy calculated for each batch. The gradient is updated to minimize the loss. After training on the training dataset, the model is evaluated on the validation data, even though the training dataset might be revisited. This evaluation is crucial for ensuring the model generalizes well and does not overfit, helping it perform effectively on unseen data. Training and validation are iterative processes that refine the model parameters to achieve high performance.

5. Evaluation This model is evaluated one final time on a test dataset, which produces the metrics (e.g. test_loss and test_accuracy). Important Performance Metrics- Accuracy, Precision, Recall and F1-Score These metrics are used to get a complete sense on how well the model is performing in identifying bird species. Furthermore, confusion matrices and classification [16] reports are created to delve deeper into class-wise performance of the model exposing where it is doing very good or bad based on classes

6. Visualization and Analysis Visualization tools are used to analyze model performance. The average loss over training, validation, and test epochs is plotted to observe how the model learns over time. Similarly, accuracy curves for the training dataset, validation dataset, and test data are plotted to assess performance improvements. A confusion matrix is also created to show classification patterns for each bird species. This visual representation helps identify areas where the model may fail or succeed in predicting different types of birds.

7. Model Saving and Loading Model check-pointing: to save the best performing model The model with the best validation set accuracy is recorded to be used for future use and reusability. This step is necessary to keep the performance of your model intact and, in order for it to be used further via practical applications or research.

V. Architecture

1.EfficientNet-B0
EfficientNet-B0 is a highly efficient convolutional neural network [3] designed for image classification, particularly for classifying bird species from images. The architecture involves several steps and components(Fig. 2):
1.1.Dataset Handling and Data Transformation: The process starts with handling the dataset, which involves preparing images of various bird species. Data transformation techniques, such as resizing, normalization, and augmentation, are applied to standardize the images for the model.

1.2 Model Construction and EfficientNet-B0 (Base Model): EfficientNet-B0 serves as the base model. It uses a compound scaling method to balance the network's depth, width, and resolution, extracting features from the input images through several convolutional layers.

1.3 Squeeze-and-Excitation (SE) Block: The SE block is integrated into the EfficientNet-B0 architecture to re-calibrate the feature maps. This block emphasizes important features while suppressing less relevant ones, enhancing the model's ability to focus on critical parts of the image and improving classification accuracy.

1.4 Dropout Layer:A dropout layer is applied to prevent over-fitting. By randomly setting a fraction of the input units to zero during training, the model becomes more robust and generalizes better to unseen data.

1.5 Fully Connected Layer: The features extracted by EfficientNet-B0 and refined by the SE block are fed into a fully connected layer. This layer maps the features to the output classes, representing different bird species.

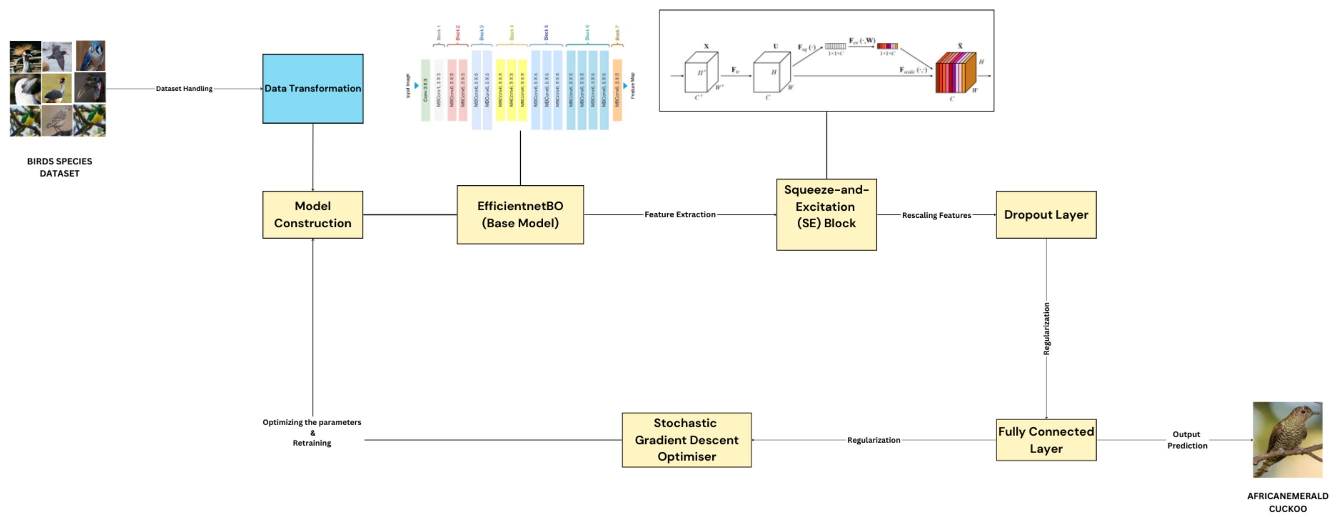

Fig. 2. Efficient Net Architecture

1.6 Stochastic Gradient Descent Optimizer: The model is optimized using the Stochastic Gradient Descent (SGD) optimizer. SGD adjusts the model's parameters to minimize the loss function, improving its performance over time.

1.7 Final Prediction: The fully connected layer produces the final prediction, classifying the input image into one of the predefined bird species.The model outputs a probability distribution over all possible classes, which allows it to identify the class (bird species). The predicted species is then chosen as the class with the highest probability. For example, it might output "African Emerald Cuckoo" for an image of that bird species [6].

2. PaliGemma

PaliGemma is an advanced AI model that combines the strengths of vision and language models to describe bird species based on images. It integrates SigLIP, a 400M vision model, with Gemma, a 2B language model, to generate detailed textual descriptions from visual inputs.(Fig. 3) The process involves several key steps and components:

2.1 SigLIP (400M Vision Model): SigLIP processes the input image and extracts relevant visual features. For example, when an image of an American flamingo is provided, SigLIP identifies and encodes its visual characteristics.

2.2 Linear Projection: The visual features extracted by SigLIP are passed through a linear projection layer. This layer reduces the dimensionality of the features, making them suitable for concatenation with textual tokens.

2.3 Concatenated Tokens: The process begins with a text prefix such as "Describe this bird species." These prefix tokens are concatenated with the image tokens generated by the linear projection, ensuring the combined representation contains both textual context and visual information.

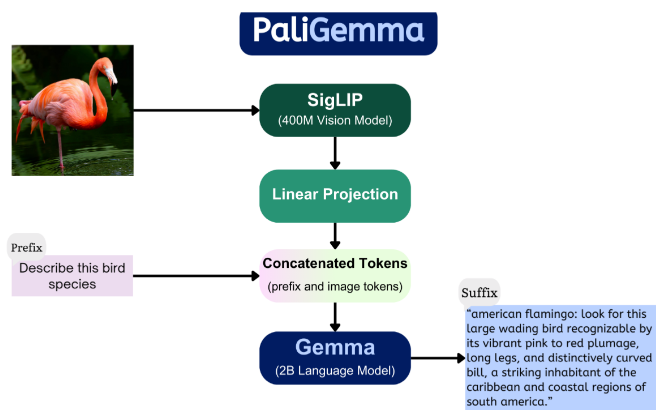

Fig. 3. PaliGemma Architecture

2.4 Gemma (2B Language Model): The concatenated tokens are input into Gemma. This language model processes the combined input to generate a descriptive text. For example, it might output: "American flamingo: Look for this large wading bird recognizable by its vibrant pink to red plumage, long legs, and distinctively curved bill, a striking inhabitant of the Caribbean and coastal regions of South America."

VI. Results and Discussion

In the experimental phase, the model was trained over 10 epochs, with significant improvements observed in both training and validation accuracy, as well as a steady decrease in loss. The EfficientNet-B0 model, combined with the PaliGemma optimization method, exhibited robust performance metrics, especially in terms of accuracy and loss convergence.

Training and Validation Performance Through the training, the model learned how to better detect bird species. The first few epochs showed a significant drop in training loss, from 3.9128 to 0.0775 The training accuracy then similarly went up from 36.40% to 98.64%. This improvement shows that the model learned to generalize as well on the training data.

Validation Score was similarly impressive. From a validation loss of 0.7440, this would drop to 0.0984 (validation accuracy from running at 87% to 97%) These results mean that the trained model was able to still generalize well on unseen data. **Test Performance** TTo ensure that the trained model is performing well and generalizing effectively, it must consistently outperform a random baseline on test set examples. We achieved nearly 99.40% accuracy on the test set in the final epoch, marking a substantial improvement over previous attempts. Additionally, the test loss decreased steadily, eventually reaching 0.0437. The close alignment between test set accuracy and training set accuracy highlights the robustness and reliability of our model architecture in identifying bird species.

Performance Metrics Each bird species in the dataset was evaluated using metrics such as precision, recall, and F1-score. A detailed examination of these metrics is essential for understanding the model's performance. The results indicated that the model performed nearly perfectly for most species, achieving an average accuracy of 99.25%. This remarkably high performance across the board underscores the effectiveness of the model and the methodologies employed.

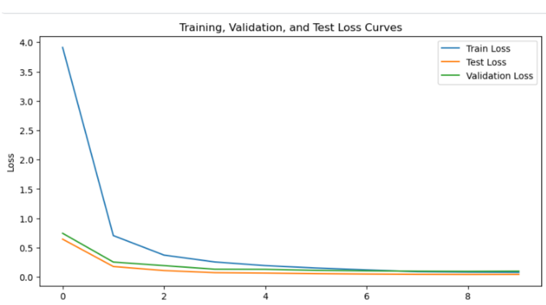

Fig. 4. EfficientNet Loss Curves

(Fig. 4)Figure 4 illustrates the training, validation, and test loss curves of an EfficientNet model. The training curve exhibits a rapid initial drop, indicative of the model's swift learning from the training data. Subsequently, the validation loss curve also descends but plateaus, signaling potential over-fitting as the model begins to memorize training examples rather than generalizing well. The test loss curve, providing an unbiased evaluation, generally mirrors the validation trend but might exhibit more fluctuations due to its smaller dataset. While the model shows promise in initial learning, the risk of over-fitting necessitates further scrutiny through accuracy metrics and other performance indicators.

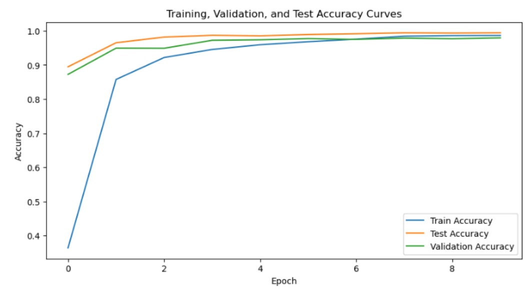

Fig. 5. EfficientNet Accuracy Curves

(Fig. 5)Figure 5 presents the training, validation, and test accuracy curves for an EfficientNet model. The graph illustrates the model's performance across these three datasets over a specified number of epochs.

A typical learning curve pattern is observed: the training accuracy steadily rises, indicating the model's growing proficiency in learning patterns from the training data. In contrast, the validation accuracy initially improves but eventually plateaus or declines, suggesting potential overfitting. The test accuracy curve, which represents the model's generalization ability, follows a trend similar to the validation curve, offering an independent assessment of the model's performance on unseen data.

To enhance model performance, it is crucial to monitor these curves closely. Early signs of overfitting can be mitigated through techniques like early stopping or regularization. Additionally, comparing the gap between training and validation accuracy offers insights into the model's ability to generalize.

(Fig. 6)Figure 6: PaliGemma Loss Curves

The provided graph illustrates the training loss of the PaliGemma model over 120 steps. It is evident that the model undergoes a significant reduction in loss during the initial training phase, rapidly decreasing from approximately 3.5 to 1.5 within the first 20 steps. Subsequently, the loss continues to decrease, albeit at a slower pace, with fluctuations around a value of 0.5. This pattern suggests that the model effectively learns from the training data and progressively improves its ability to minimize prediction errors. The presence of minor fluctuations in the loss curve towards the end of training might indicate potential over-fitting or the model reaching a convergence point. Further analysis, such as examining validation loss, would be necessary to confirm these observations and optimize the training process.

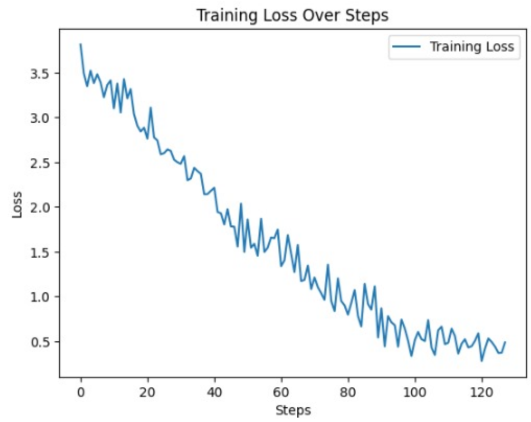

Fig. 6. Pali Gemma Loss Curves

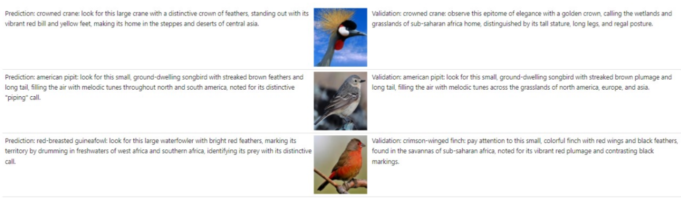

Fig. 7. Pali Gemma Result

(Fig. 7)PaliGemma demonstrates impressive capabilities in generating detailed descriptions of bird species from images by integrating vision and language models. When an image, such as that of an American flamingo, is processed, the SigLIP model extracts key visual features, and a linear projection layer reduces their dimensionality for compatibility with textual tokens. These tokens are then concatenated with a text prefix and input into the Gemma language model, which produces comprehensive descriptions. For instance, it might describe the American flamingo's vibrant plumage, long legs, and curved bill, along with its habitat in the Caribbean and coastal South America. This process showcases PaliGemma's ability to provide accurate, detailed, and context-rich narratives useful for researchers, educators, and conservationists.

VII. Conclusion

Our study showcases the capabilities of PaliGemma and a custom EfficientNet architecture standalone to accurately classify bird species. The approach of PaliGemma on interpretability and causality can shed light to the decision making process by providing key features that contribute to a correct detection. A better understanding could provide us with highly valuable information on how to focus appropriate conservation strategies; the differentiating characteristics of vulnerable species or tips that distinguish some endangered bird types.

At the same time, our custom EfficientNet model with Squeeze-and-Excitation and image augmentation was able to achieve an impressive test accuracy of 99.40%. These strong results over a wide range of bird species reinforce the utility and reliability of the pipeline in practice: as such, the model serves to reduce significant challenges for both conservation management agencies and ecological researchers. Both PaliGemma and our custom EfficientNet architecture were essential in pushing the field of avian identification forward.

VIII. Future Scope

EfficientNet-B0 has been very successful for bird species classification, initiating a new phase in avian research. Broadening up the datasets to encompass multiple geographical locations and seasonal conditions can help train the model comprehensively so that it doesn't restrict in identity identification of a limited range of species across different regions, thereby extending our knowledge on biodiversity aiding in enhancing population monitoring more efficiently.

Picture the ability to instantly know what any bird species is you find in the field. This will be possible with real-time bird identification systems running on mobile devices using the EfficientNet-B0. This may totally change field research and conservation efforts by facilitating fast data collection, analysis and decision-based on a much better evidence.

EfficientNet-B0 has capabilities that extend beyond simple recognition. By integrating model outputs with a range of ecological data, insights can be gained into bird populations, their interactions with ecosystems, and their responses to environmental changes. This comprehensive perspective allows for the development of more specific and impactful conservation measures, promoting the sustainable preservation of these avian species.

References

[1] L. Mochurad and S. Svystovych, "A new efficient classifier for bird classification based on transfer learning," *Journal of Engineering*, vol. 2024, no. 1, p. 8254130, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2024/8254130

[2] U. Nithin, A. Somnath, V. Dileep, and N. Sammeta, "Bird species classification: Using cnn models," in *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2024, pp. 1–6.

[3] A. Manna, N. Upasani, S. Jadhav, R. Mane, R. Chaudhari, and V. Chatre, "Bird image classification using convolutional neural network transfer learning architectures," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023.

[4] B. K. Rai, S. Sharma, G. Kumar, and K. Kishor, "Recognition of different bird category using image processing." *International Journal of Online & Biomedical Engineering*, vol. 18, no. 7, 2022.

[5] R. Krishnan, G. P G, G. Sujith, A. T, and S. Abhishek, "Enhancing brain tumor diagnosis: A cnn-based multi-class classification approach," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 2, 2024, pp. 1–6.

[6] G. Gupta, M. Kshirsagar, M. Zhong, S. Gholami, and J. L. Ferres, "Comparing recurrent convolutional neural networks for large scale bird species classification," *Scientific reports*, vol. 11, no. 1, p. 17085, 2021.

[7] M. C. Mannava, B. Tadigadapa, D. Anil, A. S. Dev, T. Anjali, and A. V. Vidyapeetham, "Cnn comparative analysis for skin cancer classification," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2022, pp. 1–6.

[8] V. Garrepalli, V. Gupta, S. Abhishek, and T. Anjali, "Empowering oral healthcare: Cnn-assisted diagnosis for improved patient outcomes," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 2024, pp. 1–5.

[9] M. Alswaitti, L. Zihao, W. Alomoush, A. Alrosan, and K. Alissa, "Effective classification of birds' species based on transfer learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 4172–4184, 2022.

[10] L.-L. Zhang, Y. Jiang, Y.-P. Sun, Y. Zhang, and Z. Wang, "Improvements based on shufflenetv2 model for bird identification," *IEEE Access*, 2023.

[11] S. Nayar, J. Vinitha Panicker, and J. J. Nair, "Deep learning based model for multi-class classification of cervical cells using pap smear images," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, 2022, pp. 1–6.

[12] S. Nayar, D. Priya, and V. Panicker, "Deep learning approach for automated data augmentation and multi-class classification of pap smear images," *Procedia Computer Science*, vol. 235, pp. 2205–2214, 2024.

[13] N. N. P. Poornachandran and V. Sujadevi, ", and m. aravind centre for internet studies and artificial intelligence, school of computing, amrita vishwa vidyapeetham, amritapuri, india," in *Multi-disciplinary Trends in Artificial Intelligence: 16th International Conference, MIWAI 2023, Hyderabad, India, July 21–22, 2023, Proceedings*, vol. 14078. Springer Nature, 2023, p. 245.

[14] W. Shaoqing and H. Yamauchi, "A speed-up channel attention technique for accelerating the learning curve of a binarized squeeze-and-excitation (se) based resnet model," *Journal of Advances in Information Technology*, vol. 15, no. 5, 2024.

[15] A. JL and P. CP, "Exploring adversarial attacks and countermeasures in image classification and object detection: A survey," *Available at SSRN 4607819*, 2023.

[16] Z. Tan, "Bird image classification based on joint, the inception V3 model, and transfer learning," in *Second International Conference on Physics, Photonics, and Optical Engineering (ICPPOE 2023)*, Y. Liu, Ed., vol. 13075, International Society for Optics and Photonics. SPIE, 2024, p. 130751N. [Online]. Available: https://doi.org/10.1117/12.3026684