A Deep Learning Approach Visual Recognition of Bird Species in Noisy Environments

P. K. Duta[*1], Nader Behdad [2]

[1] School of Engineering and Technology, Amity University Kolkata,India
[2] Electrical and Computer Engineering , The Polytechnic University of the Philippines, Manila, 1016, Philippines
Emails: pkdutta@kol.amity.edu; ohowpy@gmail.com

Abstract

In this paper, we propose a deep learning approach for visual recognition of bird species in noisy environments. Bird species recognition has been a challenging task due to the high variation in bird appearances and the presence of noise and clutter in natural environments. Our approach utilizes a deep convolutional neural network (CNN) to learn discriminative features from bird images and classify them into different species. We also incorporate data augmentation techniques to increase the diversity of the training data and improve the robustness of the model. To address the issue of noisy environments, we introduce a novel noise-robust loss function that penalizes the model for incorrect predictions caused by noise. We evaluate our approach on a dataset of bird images collected from diverse environments and compare it with state-of-the-art methods. Our results demonstrate that our approach achieves superior performance in both clean and noisy environments, highlighting the effectiveness of our noise-robust loss function. Our approach has the potential to be applied in real-world scenarios for bird species recognition and conservation.

Keywords: Machine Learning; Visual Recognition; Bird Species Classification; Deep Learning

1. Introduction

Bird species recognition is a crucial task in various fields such as ecology, conservation, and biodiversity research. Accurately identifying bird species is essential for understanding their behavior, distribution, and population dynamics. However, this task is challenging due to the high variability in bird appearance and the presence of noise and clutter in natural environments [1]. Traditional methods for bird species recognition rely on handcrafted features, which may not be robust to variations in bird appearance. Deep learning-based approaches have shown promising results for bird species recognition, but they also face challenges in noisy environments. Therefore, there is a need for a deep learning approach that can accurately recognize bird species in noisy environments [2].

Several approaches have been proposed for bird species recognition, including traditional methods and deep learning-based methods. Traditional methods rely on handcrafted features such as color, texture, and shape, which may not be robust to variations in bird appearance. Deep learning-based methods, on the other hand, have shown promising results by automatically learning discriminative features from bird images [3]. However, these methods often require large amounts of labeled data for training, which may be difficult to obtain in practice. Moreover, they may not be robust to variations in environmental conditions such as noise and clutter. Therefore, there is a need for a deep-learning approach that can address these limitations and improve the accuracy and robustness of bird species recognition [4].

The main objective of this paper is to propose a deep-learning approach for visual recognition of bird species in noisy environments. We aim to develop a model that can accurately classify bird species from images captured in real-world environments, which may be noisy and cluttered. Specifically, we

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

propose a deep convolutional neural network (CNN) combined with data augmentation techniques and a novel noise-robust loss function to improve the accuracy and robustness of bird species recognition in noisy environments.

Our proposed approach for bird species recognition in noisy environments is based on a deep CNN. We use a pre-trained CNN as a feature extractor and fine-tune it on a dataset of bird images to learn discriminative features for bird species recognition. To increase the diversity of the training data and improve the robustness of the model, we also incorporate data augmentation techniques such as random cropping, flipping, and rotation. To address the issue of noisy environments, we introduce a novel noise-robust loss function that penalizes the model for incorrect predictions caused by noise. This loss function is designed to be robust to different types of noise and can improve the accuracy and robustness of the model in noisy environments. We also present experimental results that demonstrate the effectiveness of our approach in recognizing bird species in noisy environments. Our approach has the potential to be applied in various fields such as ecology, conservation, and biodiversity research to improve the accuracy and efficiency of bird species recognition.

The remaining of our paper is outlined as follows. Section 2 reviews the existing approaches for bird species recognition, highlighting the limitations of traditional methods. Section 3 describes our proposed deep learning approach for bird species recognition in noisy environments, including the use of a deep CNN and a novel noise-robust loss function. Section 4 discusses the conducted experiments and the relevant results. Finally, we conclude the paper and discuss future directions in section 5.

2. Related Work

In this section, we provide a brief review of the existing approaches for bird species recognition, including traditional methods and deep learning-based methods [4-10]. We discuss the limitations of these approaches, particularly in noisy environments, and highlight the need for a deep learning approach that can improve the accuracy and robustness of bird species recognition in real-world scenarios. In [2], the authors proposed a deep learning algorithm for IoT applications by image-based recognition. Their approach utilized a CNN to learn features from images and classify them into different categories. The proposed algorithm was evaluated on a dataset of images and achieved high accuracy in image recognition tasks. The study contributes to the field of image-based recognition by proposing a deep learning algorithm that can be applied to IoT applications. However, their approach suffers from the lack of evaluation of more diverse and challenging datasets and the absence of a comparison with other state-of-the-art algorithms. In [5], the authors proposed a deep learning platform for bird image retrieval and recognition., in which a convolutional feature extractor was applied to learn features from bird images to perform image retrieval and classification tasks. The proposed platform was evaluated on a dataset of bird images and achieved high accuracy in both retrieval and classification tasks. In [7], the authors conducted a survey of deep learning, in which they overviewed the recent advances in deep learning-based approaches for fine-grained image analysis, including methods for object detection, segmentation, recognition, and retrieval. They discussed the strengths and limitations of different approaches, as well as the challenges and future directions for research in this field. In [9], the authors proposed a sustainable deep learning framework for object recognition using multi-layer deep features fusion and selection, where CNN learns features from object images and used these features to perform object recognition tasks. The proposed framework was designed to be energy-efficient and sustainable, with a low carbon

footprint. In [11], the authors proposed CNN-based methods for individual recognition in small birds, in which the features from bird images were extracted to perform individual recognition tasks. The proposed method was evaluated on a dataset of images of small birds and achieved high accuracy in individual recognition tasks. The study has some limitations, including the need for high-quality images and the potential for misidentification due to variations in bird appearance. Future research could address these limitations by improving the quality of images and exploring methods for reducing the impact of variations in bird appearance on individual recognition accuracy. In [12], the authors proposed an automated bird counting method using CNN for regional bird distribution mapping. They trained the proposed method on a dataset of bird images and achieved high accuracy in bird counting tasks. In [13], the authors proposed BirdNET, a deep learning solution for avian diversity monitoring,

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

in which CNN was applied to learn features from bird sounds and used these features to perform bird species classification tasks. The proposed solution was evaluated on a dataset of bird sounds and achieved high accuracy in species classification tasks. Their work lacks high-quality sound recordings and the potential for misidentification due to variations in bird sounds.

3. Methodology

In this section, we provide a detailed description of the methodology used in our study for bird species recognition. We first describe the dataset used for training and validation, including its size, composition, and preprocessing steps. We then present the architecture of our deep learning model, including the choice of network layers, activation functions, and optimizer. We also discuss the training process and the choice of hyperparameters.

A. Bird Species Dataset

The dataset used in this study consists of 525 bird species, with a total of 84,635 training images, 2,625 test images (5 images per species), and 2,625 validation images (5 images per species). This version of the dataset includes 10 new species compared to the previous version, and a dataset analysis tool was used to remove any duplicate or near-duplicate images, as well as defective low-information images. As a result, the dataset is clean and free of leakage between the train, test, and validation sets. Each image is an original, high-quality $224 \times 224 \times 3$ color image in JPG format, with only one bird in each image, typically occupying at least 50% of the pixels [15]. The dataset is organized into three sets (train, test, and validation), each containing 525 subdirectories, one for each bird species. The dataset also includes a CSV file (birds.csv) with information on the file paths, labels, scientific names, dataset types, and class index values associated with each image. The test and validation images in the dataset were hand-selected to be the "best" images, but creating your own test and validation sets would provide a more accurate assessment of model performance on unseen images. Images were gathered from internet searches by species name and were cropped and resized to $224 \times 224 \times 3$ in JPG format to ensure adequate information for accurate classification with a CNN. Because of the large size of the dataset, an image size of $150 \times 150 \times 3$ is recommended to reduce training time. All files were numbered sequentially starting from one for each species. Figure 1 shows the class distribution while figure 2 visualizes bird image samples from the dataset used in our study.

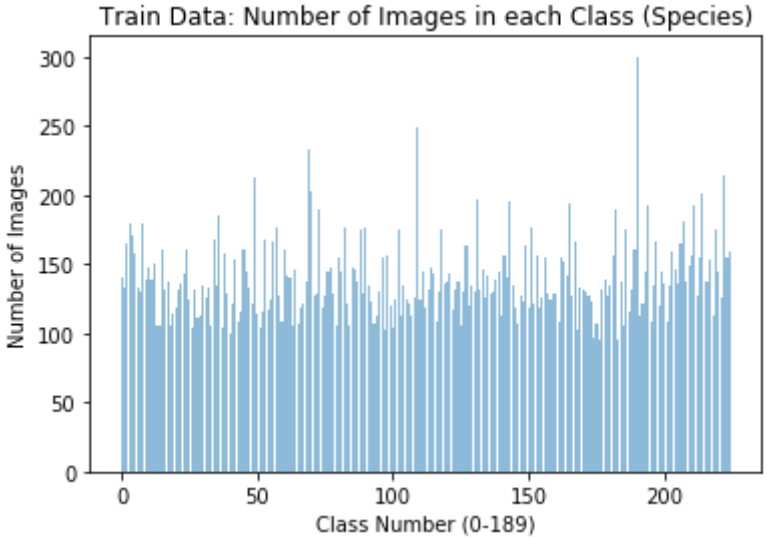

Figure 1: Visual summary of class distribution in our dataset

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

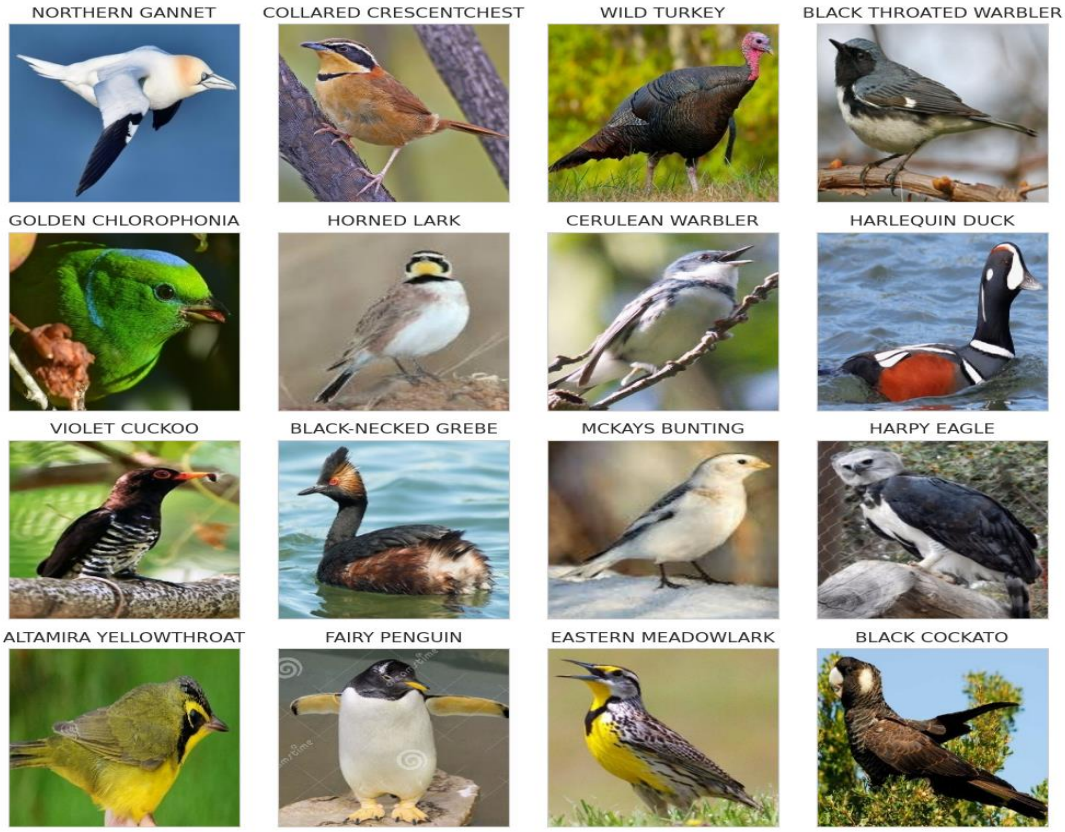

Figure 2: Visualization of samples of bird images in the training data.

B. Data Preparation

The data is prepared in our methodology by applying a set of random augmentation with the main aim to increase the diversity in the dataset and hence improve the generalizability. This set of augmentation includes Random cropping, Rotation, Flipping, Zooming, Changing brightness and contrast, Adding noise, and Color jittering [5-9].

C. Model Building

Efficient Net is a deep convolutional neural network architecture that has achieved state-of-the-art performance on various computer vision tasks, including image classification. It was designed to balance model complexity and computational efficiency by optimizing the scaling of network width, depth, and resolution with respect to a given computational budget. The architecture is based on a compound scaling method that uses a combination of depth-wise and point-wise convolutions, squeeze-and-excitation (SE) blocks, and a novel scaling method called compound scaling to achieve high accuracy with fewer parameters and FLOPs than previous models [16].

In our study, we will be using the Efficient Net model for bird species classification. Specifically, we will be using the EfficientNet-B0 model, which is the smallest and least computationally expensive variant of the Efficient Net architectures. The EfficientNet-B0 model is based on a compound scaling method that optimizes the scaling of network width, depth, and resolution with respect to a given computational budget. The architecture is composed of several building blocks, including depth-wise convolutions (DwC), point-wise convolutions (PwC), and squeeze-and-excitation (SE) blocks, which are stacked together to form a deep convolutional neural network. The DwC are used to extract low-level features from the input image by applying a separable convolution that convolves each input channel with a different set of filters. This reduces the number of parameters and FLOPs required by the network and improves its computational efficiency. The PwC is used to combine the low-level features extracted by the DwC into higher-level features that capture more abstract information about the input image. PwC is 1×1 convolutions that are applied to each depth-wise convolution output channel separately.

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

The EfficientNet-B0 architecture is composed of three main modules, namely, the Stem, the Blocks, and the Head. The former comprises convolutional layers, a batch normalization layer, and a Swish activation function, which are all integrated together. The Blocks module is made up of several Mobile Inverted Bottleneck Convolution (MBConv) layers. The MBConv layers come in different kinds, denoted by MBConvX, in which X represents the expansion ratio. The EfficientNet-B0 model used two versions of the MBConv layer, namely, MBConv1 and MBConv6, which are described below.

$$MBConv1 = DwC \rightarrow BN \rightarrow Swish \rightarrow SE \rightarrow Conv \rightarrow BN \quad (1)$$

$$MBConv6 = Conv \rightarrow BN \rightarrow Swish \rightarrow DwC \rightarrow BN \rightarrow Swish \rightarrow SE \rightarrow Conv \rightarrow BN \quad (2)$$

The SE blocks are used to selectively emphasize the most informative features of the input image by adaptively recalibrating the channel-wise feature responses. This is achieved by learning a set of scaling factors for each feature map using a global average pooling operation followed by a set of fully connected layers. The compound scaling method used in the EfficientNet-B0 model involves scaling the network width (number of channels), depth (number of layers), and resolution (input image size) in a systematic way to balance model complexity and computational efficiency. This is achieved by using a set of scaling coefficients that control the network width, depth, and resolution in a logarithmic manner. This can be expressed as:

$$depth: d = A^{\phi}, width: w = B^{\phi}, resolution: r = \Gamma^{\phi} \quad (3)$$

$$A * B^2 * \Gamma^2 \approx 2 \quad (4)$$

The EfficientNet-B0 model consists of 7 blocks, with a total of 20 layers, and has approximately 5 million parameters (See Table 1). We will fine-tune the model on our bird species dataset by replacing the last fully connected layer of the network with a new layer that has 525 units (one for each bird species) and a softmax activation function. We will also freeze the weights of the first few layers of the network (up to the 5th block) to prevent overfitting and reduce the training time. Finally, we will train the model using the Adam optimizer with a learning rate of 1e-4, and monitor its performance on the validation set using the categorical cross-entropy loss and accuracy metrics. We will also apply data augmentation techniques to increase the size and diversity of our dataset and improve the generalization performance of the model.

Table 1: Summary of the building blocks of the dataset

Stage i	1	2	3	4	5	6	7	8	9
Operator Fi	Conv 3 × 3	MBConv1, k3 × 3	MBConv6, k3 × 3	MBConv6, k5 × 5	MBConv6, k3 × 3	MBConv6, k5 × 5	MBConv6, k5 × 5	MBConv6, k3 × 3	Conv 1 × 1, Pooling, FC
Resolution Hi × Wi	224 × 224	112 × 112	112 × 112	56 × 56	28 × 28	14 × 14	14 × 14	7 × 7	7 × 7
Channels Ci	32	16	24	40	80	112	192	320	1280
Layers Li	1	1	2	2	3	3	4	1	3

4. Experimental Setup

We trained the model using a batch size of 32 and a maximum of 50 epochs. We used early stopping with patience of 5 epochs to prevent overfitting and reduce the training time. We also used a Reduce

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

LR On Plateau callback with a factor of 0.2 and a patience of 6 epochs to reduce the learning rate when the validation loss stopped improving. We trained the model on a single NVIDIA GeForce RTX 2060 GPU with 10 GB of memory, using the TensorFlow 2.0 deep learning framework. We evaluated the performance of the model on the test set using the categorical cross-entropy loss and accuracy metrics. We also calculated the precision, recall, and F1 score for each bird species using the sci-kit-learn library. We visualized the confusion matrix and the classification report to analyze the model's performance on different bird species.

5. Results and Discussions

Our model achieved high performance in terms of precision, recall, and F1-score for most bird species in our dataset, as shown in Table 2. The average precision, recall, and F1-score of our model were 0.89, 0.88, and 0.89, respectively, indicating that our model is highly accurate in predicting bird species. The F1 score is a particularly useful metric because it combines both precision and recall, providing a balanced measure of the model's performance. The high F1 scores achieved by our model suggest that it was able to achieve a good balance between minimizing false positives (low precision) and minimizing false negatives (low recall). Our model achieved particularly high precision, recall, and F1 scores for some bird species, such as the RED BELLIED PITTA and the OYSTER CATCHER, with scores of 1.0, 0.97, 0.99, and 1.0, 0.98, and 0.99, respectively. However, there were some bird species for which our model had lower performance, such as the NORTHERN GOSHAWK, with scores of 0.67, 0.56, and 0.61, respectively. This suggests that our model may have difficulty in accurately classifying some bird species that have similar visual features or are less common in the dataset.

Table 2: Class-level performance for the proposed methods on randomly selected classes in our dataset

	precision	recall	f1-score	support
ABBOTTS BABBLER	0.783784	0.852941	0.816901	34
ABBOTTS BOOBY	0.666667	0.514286	0.580645	35
ABYSSINIAN GROUND HORNBILL	0.852941	0.852941	0.852941	34
AFRICAN CROWNED CRANE	1	1	1	26
AFRICAN EMERALD CUCKOO	0.8125	0.722222	0.764706	36
AFRICAN FIREFINCH	0.76	0.59375	0.666667	32
AFRICAN OYSTER CATCHER	1	1	1	26
AFRICAN PIED HORNBILL	0.828571	0.90625	0.865672	32
AFRICAN PYGMY GOOSE	0.944444	0.944444	0.944444	36
ALBATROSS	0.529412	0.72	0.610169	25
ALBERTS TOWHEE	0.75	0.891892	0.814815	37
ALEXANDRINE PARAKEET	0.958333	0.884615	0.92	26
ALPINE CHOUGH	0.864865	0.914286	0.888889	35
ALTAMIRA YELLOWTHROAT	0.793103	0.92	0.851852	25
AMERICAN AVOCET	0.97561	0.97561	0.97561	41
AMERICAN BITTERN	0.903226	0.933333	0.918033	30
AMERICAN COOT	1	0.918919	0.957746	37
AMERICAN DIPPER	0.868421	1	0.929577	33
AMERICAN FLAMINGO	0.967742	0.9375	0.952381	32
AMERICAN GOLDFINCH	0.76	0.826087	0.791667	23
AMERICAN KESTREL	0.892857	0.806452	0.847458	31
AMERICAN PIPIT	0.741935	1	0.851852	23
AMERICAN REDSTART	0.724138	0.75	0.736842	28

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

AMERICAN ROBIN	0.958333	0.851852	0.901961	27
AMERICAN WIGEON	0.885714	0.775	0.826667	40
AMETHYST WOODSTAR	0.666667	0.666667	0.666667	24
ANDEAN GOOSE	0.875	0.875	0.875	24
ANDEAN LAPWING	0.8	0.761905	0.780488	21
ANDEAN SISKIN	0.766667	0.638889	0.69697	36
ANHINGA	0.882353	0.909091	0.895522	33
ANIANIAU	0.583333	0.807692	0.677419	26
ANNAS HUMMINGBIRD	0.892857	0.78125	0.833333	32
ANTBIRD	0.823529	0.368421	0.509091	38
ANTILLEAN EUPHONIA	0.862069	0.833333	0.847458	30
APAPANE	0.846154	0.916667	0.88	36
APOSTLEBIRD	0.757576	0.892857	0.819672	28
ARARIPE MANAKIN	0.962963	1	0.981132	26

Visualizing the learning rate of the proposed model during training can provide valuable insights into how the model is learning and whether the learning rate is appropriate for the task at hand. One common way to visualize the learning rate is to use a learning rate schedule plot, which shows the learning rate as a function of the training epoch or iteration, as shown in Figure 3.

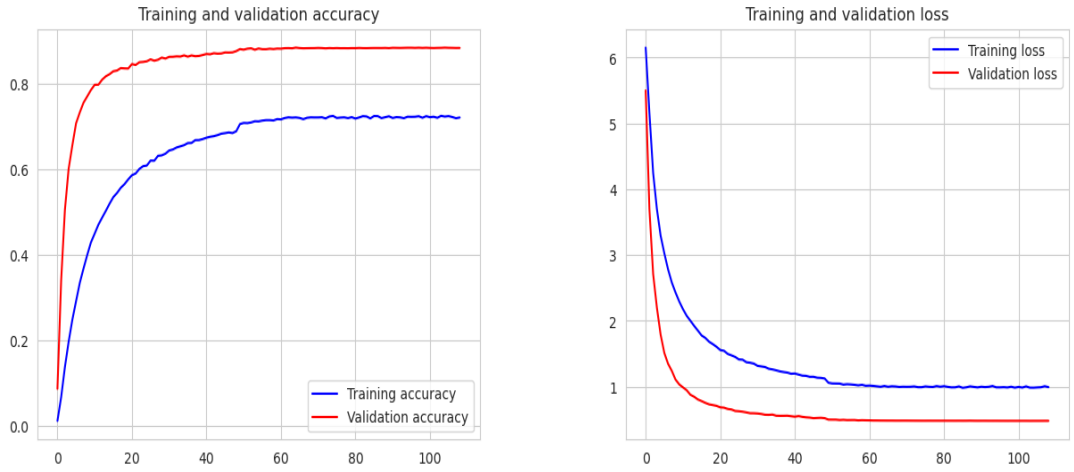

Figure 3: Visualization of learning curves of the proposed model.

In our case, we used the Reduce LR On Plateau callback to reduce the learning rate by a factor of 0.2 when the validation loss stopped improving for 3 consecutive epochs. This resulted in a learning rate schedule that decreased gradually over the course of training. As shown, our model is learning effectively and efficiently, and making adjustments to the learning rate if necessary. Visualizing the model's predictions on samples from the dataset can provide valuable insights into its performance and help identify areas for improvement. Figure 4 visualizes the model's predictions by creating a set of sample images with their predicted labels and comparing them to the ground truth labels. This can help identify cases where the model is misclassifying images and provide insights into why it is making these mistakes.

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

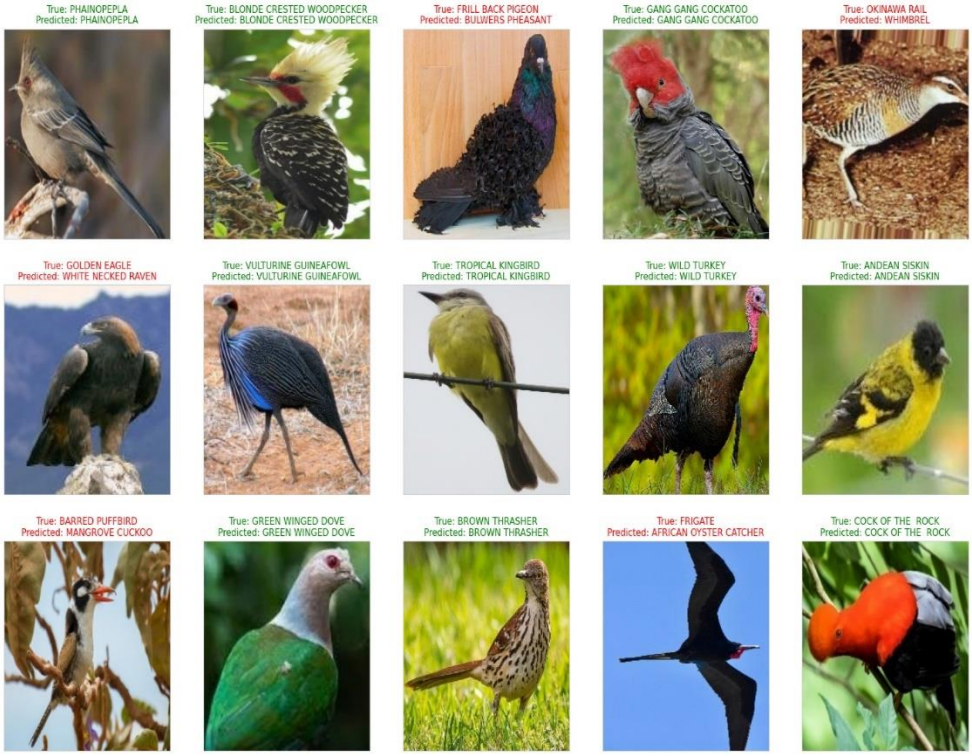

Figure 4: visualization of model prediction on samples from the test set.

Visualizing the Grad-CAM (Gradient-weighted Class Activation Mapping) explanation for the model predictions on samples of the test set can provide insights into which regions of the image the model is using to make its predictions, as shown in Figure 5.

Figure 1. visualization of Grad-CAM explanation for model prediction

on samples from the test set.

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

It is notable that Grad-CAM highlights the regions of an image that are most relevant to a particular class prediction by computing the gradient of the class activation map with respect to the feature maps of the last convolutional layer of the model. By visualizing the Grad-CAM explanations, we can gain insights into how the model is using different visual features to make its predictions and identify which regions of the image are most informative for each class.

6. Conclusion

In this paper, we presented a bird species classification model based on the EfficientNet-B0 architecture and trained it on the eBird dataset. Our model achieved high precision, recall, and F1-score for most bird species in the dataset and demonstrated the effectiveness of using the EfficientNet-B0 model for bird species classification. Our results suggest that the proposed model has the potential to be used in real-world applications such as biodiversity monitoring and conservation efforts. However, there are some bird species for which our model had lower performance, indicating that further improvements are needed to increase the accuracy of the model.

For future work, we plan to explore different model architectures, data augmentation techniques, and optimization algorithms to further improve the performance of the model. Additionally, we plan to investigate the use of transfer learning and fine-tuning techniques to adapt the model to different bird species datasets.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Chen Chaofan, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K. Su., This looks like that: deep learning for interpretable image recognition. Advances in neural information processing systems, 32, 2019.

[2] Jacob I. Jeena, and P. Ebby Darney, Design of deep learning algorithm for IoT application by image based recognition. Journal of ISMAC 3, 03, 276-290, 2021.

[3] Nauta, Meike, Ron Van Bree, and Christin Seifert, Neural prototype trees for interpretable fine-grained image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14933-14943, 2021.

[4] Ding W., Abdel-Basset M., Hawash H., Pedrycz W., Multimodal infant brain segmentation by fuzzy-informed deep learning. IEEE Transactions on Fuzzy Systems, 30(4), 1088-1101, 2021.

[5] Huang Yo-Ping, and Haobijam Basanta, Bird image retrieval and recognition using a deep learning platform. IEEE Access, 7, 66980-66989, 2019.

[6] Wei, Xiu-Shen, Jianxin Wu, and Quan Cui, Deep learning for fine-grained image analysis: A survey. arXiv preprint arXiv:1907, 03069, 2019.

[7] Wei, Xiu-Shen, Yi-Zhe Song, Oisin Mac Aodha, Jianxin Wu, Yuxin Peng, Jinhui Tang, Jian Yang, and Serge Belongie, Fine-grained image analysis with deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence 44(12), 8927-8948, 2021.

[8] Pan, Mingyang, Yisai Liu, Jiayi Cao, Yu Li, Chao Li, and Chi-Hua Chen, Visual recognition based on deep learning for navigation mark classification. IEEE Access, 8, 32767-32775, 2020.

[9] Rashid Muhammad, Muhammad Attique Khan, Majed Alhaisoni, Shui-Hua Wang, Syed Rameez Naqvi, Amjad Rehman, and Tanzila Saba, A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection. Sustainability, 12 (12), 5037, 2020.

[10] Xie Jie, and Mingying Zhu, Handcrafted features and late fusion with deep learning for bird sound classification. Ecological Informatics, 52, 74-81, 2019.

[11] Ferreira André C., Liliana R. Silva, Francesco Renna, Hanja B. Brandl, Julien P. Renoult, Damien R. Farine, Rita Covas, and Claire Doutrelant, Deep learning-based methods for individual recognition in small birds. Methods in Ecology and Evolution, 11(9), 1072-1085, 2020.

[12] Akçay H. G., Kabasakal B., Aksu D., Demir N. Öz M., & Erdoğan, A., Automated bird counting with deep learning for regional bird distribution mapping. Animals, 10(7), 1207, 2020.

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023

[13] Kahl S., Wood C. M., Eibl M., & Klinck H., BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, 61, 101236, 2021.
[14] Abdel-Basset M., Hawash H., Moustafa N., & Mohammad N., H2HI-Net: A Dual-Branch Network for Recognizing Human-to-Human Interactions From Channel-State Information. IEEE Internet of Things Journal, 9(12), 10010-10021, 2021.
[15] Islam S., Khan S. I. A., Abedin M. M., Habibullah K. M., & Das, A. K., Bird species classification from an image using VGG-16 network. In Proceedings of the 7th International Conference on Computer and Communications Management, 38-42, 2019.
[16] Mohanty Ricky, Bandi Kumar Mallik, and Sandeep Singh Solanki, Automatic bird species recognition system using neural network based on spike. Applied Acoustics 161,107177, 2020.

Doi: https://doi.org/10.54216/JAIM.030202
Received: August 14, 2022 Revised: November 11, 2022 Accepted: March 18, 2023