A Deep Learning-Based Transfer Learning Approach for the Bird Species Classification

Tejalal Choudhary[1(✉)], Shubham Gujar[2], Kruti Panchal[3], Sarvjeet[4], Vipul Mishra[1], and Anurag Goswami[1]

[1] Bennett University, Greater Noida, India
tejalal.choudhary@gmail.com,
{vipul.mishra,anurag.goswami}@bennett.edu.in
[2] Vishwakarma Institute of Information Technology, Pune, Maharashtra, India
gujar303shubham@gmail.com
[3] Devang Patel Institute of Advance Technology and Research, Anand, India
kbpanchal34@gmail.com
[4] PDPM IIITDM, Jabalpur, Madhya Pradesh, India
901sarvjeet@gmail.com

Abstract. Birds are present in various scenarios that appear in different shapes, sizes, and colors. It is estimated that there exist around 10,000 different species all over the world. Watching birds is a common practice but identifying their species requires bird knowledge. Human ability to recognize the birds through images is much easier as compared to audio classification, so bird species classification through images is in more trend and accurate too. Traditional machine learning and audio-based approaches are not suitable for bird classification. In this paper, we proposed a transfer learning-based approach for the classification of the bird dataset of 200 bird species. We performed numerous experiments with different deep learning models and the experimental results suggest that the proposed transfer learning-based approach significantly performs better than other state-of-the-art methods.

Keywords: Deep learning · Convolutional Neural Network · Transfer learning · Image classification

1 Introduction

Nowadays, a number of people visit bird sanctuaries to see birds of different species, to get knowledge of the bird species, to enjoy their variant colors, and characteristics of the birds. But people do not know various kinds of birds due to their varying color, shape, and textures. Normally a person without experience in the field of ornithology has knowledge of approximately 20 species. Moreover, decrease in bird population have become a concern for the world as they respond to the environment changes quickly. Therefore, in such a situation, it will be

D. Garg et al. (Eds.): IACC 2020, CCIS 1368, pp. 43–52, 2021.
https://doi.org/10.1007/978-981-16-0404-1_4

beneficial to have a mechanism that can help to identify bird species so that the government or sanctuaries could preserve them and save them from getting extinct.

Deep learning-based classification methods have become famous in various domains and have shown tremendous results. Convolutional Neural Networks (CNNs) are one of the variants of deep learning architectures. CNN has shown excellent results in image classification [21]. Image classifiers are required to know in which category the bird image belongs. Automatic bird detection can help avoid mortality of birds, bird collisions at wind turbines are becoming a growing concern. Bird species recognition can also help create awareness and maintain social responsibility among individuals which will reduce the hunting of birds and maintain ecological balance. Generally, bird classification can be done either by image or audio. Every bird species has a unique voice so classifying them by their sound will give high accuracy. But if there are some problems such as background noise or birds not singing for a long time then this approach will not work. Classification based on audio is less practical [11], images are more reliable. In last decade, deep learning have shown excellent performances in various classification, object detection, image generation, natural language processing, and segmentation applications. CNNs are one of the popular variants of deep neural networks. CNNs have been used to solve a variety of classification problems and achieved outstanding performances. The identification of birds can be seen as a classifying the bird to one of the classes of the dataset.

In this paper, we have proposed a transfer learning-based approach for the classification of bird species. Our main aim is to classify the bird images to one of the classes available in the dataset. Here, we have used CNNs to classify the image. The prior visual component of the birds is shape, size, wings, color, size and small features of birds like its eyebrows and eyes which creates edges of different shape that also helps to differentiate the bird species [4]. We have used transfer learning and pre-trained CNNs on a 200 bird species dataset. We experimented with pre-trained models such as VGG16, ResNet50, and MobileNetV2. The experiments with various CNN architectures show that our method is superior than other methods in the domain.

This paper has been divided into various sections. Section 2 discusses the important contribution made by other researchers in the same field. In Sect. 3, we have discussed our transfer learning-based approach for the classification of the birds. Experiments and results have been discussed in Sect. 4. Finally, the paper is concluded in Sect. 5.

2 Related Work

In recent years, there are numerous approaches for the classification of birds using images or audio-data. For bird classification using audio data, a spectrogram is generated from a sound file, for signal and noise that is divided into chunks and stored as samples that are trained using convolutional neural networks [19]. Fine-Grained Visual Categorization (FGVG) aims to differentiate visual categories.

Implementing fine-tuning via transfer learning offers an effective solution for domain-specific FGVC. The authors worked with the pre-trained model AlexNet and NASNet that took images of input size 224×224 and 331×331 respectively and achieved an accuracy of 90% on the validation set [3].

In another research [8], the classification of bird species found in Bangladesh is done using the VGG16 model for feature extraction. The last two FC layers were removed and classification algorithms were used instead such as SVM ($n = 5$), KNN ($n = 10$), Naïve Bayes with 4-fold cross-validation. The SVM algorithm has achieved 89% accuracy, and the Random Forest algorithm has achieved 89% accuracy. The authors in [10], proposed an approach with bird ROI detection using Mask R-CNN to localize images and also DeepNet, InceptionResNetV2, and InceptionV3 architectures were performed for multi-stage training with transfer learning using ImageNet pre-trained weights. In a research [13], the authors used the SVM algorithm with color-based feature extraction for 6 bird species classification. [14] proposed a methodology using CNN architecture with 3 convolutional layers with the ReLU activation function. [21] proposed a deep learning-based approach which includes AlexNet, VGG16, and GoogleNet based classification using image localization techniques such as R-CNN, pose normalized net, PS-CNN, and deep LAC.

[2] proposed a methodology that uses motion features combined with a Normal Bayes classifier and SVM classifier. While [15] performed bird-classification based on parts of the birds. In another research, [12] used an approach that eliminated the background elements and uses a color histogram to extract the features of the bird and classify it. [7] also proposed a deep learning-based approach for the classification of the Taiwan bird dataset. The authors experimented with CNN with skip connection, without skip connection and SVM. It is found that CNN with skip connections reveal the highest accuracy than CNN without skip connections and SVM. However, the scale of the dataset is very small. The dataset has only 23 bird classes. In our proposed approach, pre-trained models VGG16, ResNet50, and MobileNetV2 are trained and tested for classifying bird-species into 200 classes.

3 Methodology

CNNs have become popular in recent years and have shown outstanding performances on various computer vision tasks. We proposed a deep learning-based transfer learning approach for bird species classification. In this section, first, we have discussed the proposed transfer learning-based approach, fundamentals of the CNNs, and the architectures used in the experiments. Our methodology of classifying birds is divided into the following parts.

– The use of the existing best performing CNNs.
– Transfer learning to transfer and utilize the knowledge learned by pre-trained models to classify the bird species.
– Finally, the performance of all the CNNs is compared.

3.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) widely used in computer vision applications, such as image recognition, image classification, object detection, and are also attracting interest across a variety of domains apart from robotics and self-driving cars [5]. In CNN, the backpropagation technique is used for learning spatial relationships of features from an input image. Convolutional layers are the fundamental component of CNN and pooling layers both collectively perform feature extraction from the inputs and the dense layer layers map the extracted features into the final output, such as classification. The convolution layers consist of convolutional operations followed by max-pooling and non-linearity. To detect non-linear features for multi-layer networks in CNN activation functions are used. The most used activation functions are sigmoid, tanh, linear, ReLu, and Leaky ReLu. Mathematical operations such as linear operations are done commonly on CNN. For binary classification tasks, the sigmoid function is used as an output layer for CNN. The kernel is a grid or matrix used as a filter for extracting features from input images moved over each image position. The features get added up to the output of the layer and this output is passed on to the next layer hierarchically and progressively which becomes more complex and also helps CNN for great performance. Further, multi-class image classification is done using the Softmax layer. CNN has shown excellent performances in a variety of applications that deal with image data, computer vision, medical image analysis, and also in natural language processing (NLP) with great results [1]. Figure 2 show the architecture of the VGG16 model used in our experiments. The original model has 1000 nodes in final softmax layer, whereas we change it according to the number of classes in the dataset.

3.2 Transfer Learning

In transfer learning the core idea is training the model over a source domain and applying the learned knowledge on a target domain where the both domain belongs to the labeled data type. The problem addressed here belongs to the classification problem where the source domain features are different from the target domain features [18]. Transfer learning is done in two steps. First, using a pre-trained model network for feature extraction, and second, fine-tuning a pre-trained model network with the new data. Pre-trained model networks have shown great results for deep learning applications, large scale datasets have been used for training these models, and this knowledge can be used in new models. Transfer learning is very significant if we have a small dataset. The model starts over-fitting after a few epochs when we have a small dataset. If the first model has been trained on large enough data and is general, the learned features can also be used for the classification of new class images [20].

4 Experiments

In this section, first, we have provided the detail of the dataset used for our experiments followed by the details of each model and its results. We have performed experiments on distinct convolutional architectures. The convolutional architectures used in our experiments are VGG16, ResNet50, and MobileNetV2. The initial experiments are performed with the VGG16 network. After getting motivating results with VGG16, we extended our experiments to ResNet50 and MobileNetV2. All the experiments are performed on a single NVIDIA GPU with 16 GB RAM.

4.1 Dataset

The dataset used for training and testing convolutional models is taken from [9]. The dataset consists images of 200 species of different birds. Following the standard guidelines, the dataset is divided into three parts: training, validation, and testing. The training set consists of 27,506 images, approximately 125–150 per species. The validation and testing set have 1000 images each. The image dataset is needed to be preprocessed before they can be used for training. First, all the images are scaled to the same size. In order to reduce the noise and disturbances in the image, the images are normalized so they are ready to be used for training. Figure 1 shows three images from the 200 bird images dataset. The dimensions of the images are $224 \times 224 \times 3$.

Fig. 1. Three images from the dataset.

4.2 Experiments on VGG16

VGG16 is the most common and widely used CNN architecture. It has 138M learnable parameters and 30.97B FLOP. It has 16 layers, out of which 13 are the convolutional and 3 fully connected layers [17]. The first set of experiments are performed on the VGG16 network. VGG16 model has very consistent architecture, in VGG16 filters of size 3×3 and the same padding is used, and the activation function ReLU is used. After this max pooling layer with a size of 2×2 and a stride of the same dimensions is applied that extracts the important information. The model is evaluated using the test dataset. For experimentation, a single NVIDIA 16 GB GPU is used with Keras framework and TensorFlow at the backend. Adam optimizer with cross-entropy loss is used. Python libraries and matplotlib are used for visualizations. Data augmentation was also done with the shift, flip, rotation, zoom methods using Imagedatagenerator, and ImageNet pre-trained weights were also used. The best result achieved from VGG16 was 96% testing accuracy and 0.235 loss with batch size 32, 50 epochs, and 0.001 learning rate. Table 1 shows the summary of the experiments performed with VGG16 model.

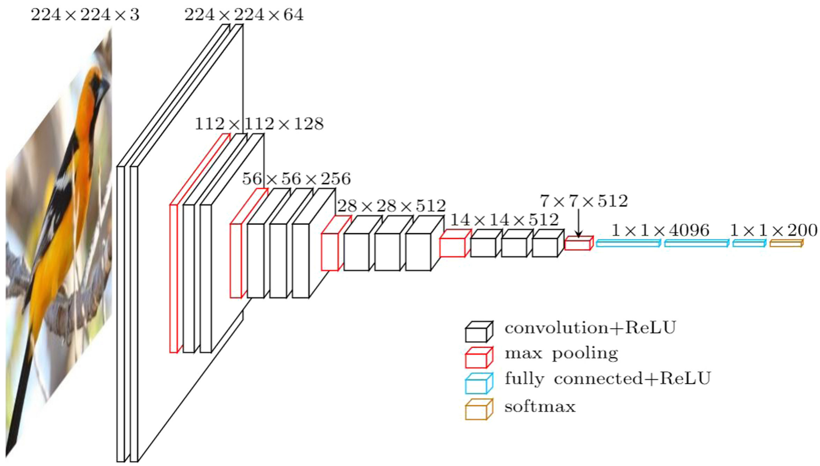

Fig. 2. VGG16 architecture used in our experiments

Table 1. Results of the experiments performed on the VGG16 model.

Exp no	Batch size	Training accuracy(%)	Testing accuracy(%)	Epochs	Learning rate
1	8	95	94	40	0.01
2	16	97	96	50	0.01
3	32	98	96	50	0.001
4	**32**	**99**	**96**	**50**	**0.001**

4.3 Experiments on ResNet50

ResNet50 is a deeper and more complex network than VGG16 and it has 50 layers [6]. ResNet50 has 5 stages with residual blocks after every stage. Each residual block has 3 layers with $1*1$ convolutions and $3*3$ convolutions. These residual blocks help to avoid information loss for deep networks, speed up the training time, and thus boosts up the performance of the model. ResNet50 is different from other sequential networks, here, the output of the one layer is also used as input to another layer that is not the immediate layer. This avoids over-fitting for the training dataset and gives generalized results. ResNet was the winning model for the ImageNet (ILSVRC) competition 2015 and is also used for many object detection applications as well.

The ResNet50 model is implemented using a single NVIDIA Tesla K8 12 GB GPU with TensorFlow backend and Keras framework. Python packages and python is also used as a programming language for experimentation. Adam was used as an optimizer. Transfer learning was used using ImageNet pre-trained weights and the ResNet50 model with more than 23 million parameters was trained. The best model achieved 85% testing accuracy with 32 batch size, 50 epochs, 0.01 learning rate, and cross-entropy loss around 0.70. Table 2 shows the results of the experiments performed on the ResNet50 model.

Table 2. Results of the experiments performed on the ResNet50 model.

Exp no	Batch size	Training accuracy(%)	Testing accuracy(%)	Epochs	Learning rate
1	**32**	**99**	**85**	**50**	**0.01**
2	64	99	83	43	0.01
3	32	99	65	40	0.001
4	32	99	68	80	0.001

4.4 Experiments on MobileNetV2

MobileNetV2 is a comparatively small model in terms of the learnable parameters and FLOPs than VGG16 and ResNet50 [16]. The MobilnetV2 consists of 2 types of blocks: (1) residual block with stride1 (2) Another block with stride 2 for downsizing. For both types of blocks, there are 3 layers. In the first layer, there is a $1 * 1$ convolution with ReLU activation. The second layer is a depthwise convolution. and the last layer is $1 * 1$ convolution without any non-linearity. The architecture contains the initially fully convolutional layers with 32 filters, followed by 19 residual bottleneck layers. As the MobileNetV2 gives the benefit of memory-efficient inference and utilizes standard operations that are present in neural networks. MobilenetV2 improves the wide range of performance points for the ImageNet dataset. We compare the accuracy of three models VGG16, ResNet50, and MobilenetV2 with each other for bird species image classification.

We used NVIDIA GPU to train our models. Python programming language is used for experiments. The models are implemented using python packages and Keras with Tensorflow as backend. For image classification, we used MobileNetV2 architecture, all the images are of 224×224 pixels. Also, data augmentations methods such as shift, flip, rotation, zoom are used on the dataset. Cross entropy and adam are used as loss function and optimizers respectively. Table 3 shows the details of the experiments performed with the MobileNetV2 model.

Table 3. Results of the experiments performed on the MobileNetV2 model.

Exp no	Batch size	Training accuracy(%)	Testing accuracy(%)	Epochs	Learning rate
1	**32**	**99**	**95**	**50**	**0.0001**
2	32	93	89	50	0.01
3	64	87	83	30	0.0001
4	128	85	78	30	0.0001

4.5 Analysis

In our proposed methodology, we experimented with VGG16, ResNet50, and MobileNetV2 convolutional architectures. The maximum testing accuracy achieved with VGG16 is 96%, maximum accuracy with ResNet50 is 85% and MobileNetV2 is 95%. Figure 3 show the training and testing accuracy of all the models. It can be seen from Fig. 3 that out of all the models, VGG16 performed better than MobileNetV2 and ResNet50. Table 4 summarizes the best results achieved with VGG16, ResNet50, and MobileNetV2 models. Figure 4 shows the accuracy and loss information of the VGG16 model for the best trained model. The experiments with VGG16, ResNet50, and MobileNetV2 shows that using transfer learning for bird species classification is an effective way of identifying birds of different species. In future, we will work to classify larger birds datasets.

Table 4. Summary of the best results from various experiments performed with CNN Models.

CNN Model	Batch size	Training accuracy(%)	Testing accuracy(%)	Epochs	Learning rate
VGG16	**32**	**99**	**96**	**50**	**0.0001**
ResNet50	32	99	85	50	0.01
MobileNetV2	32	99	95	50	0.0001

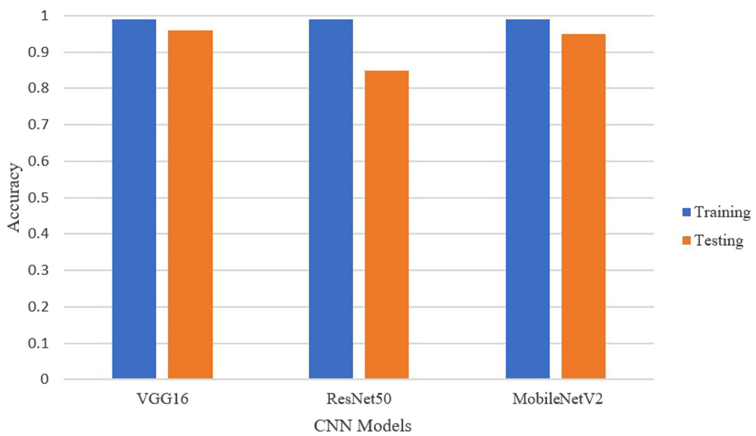

Fig. 3. Accuracy comparison models.

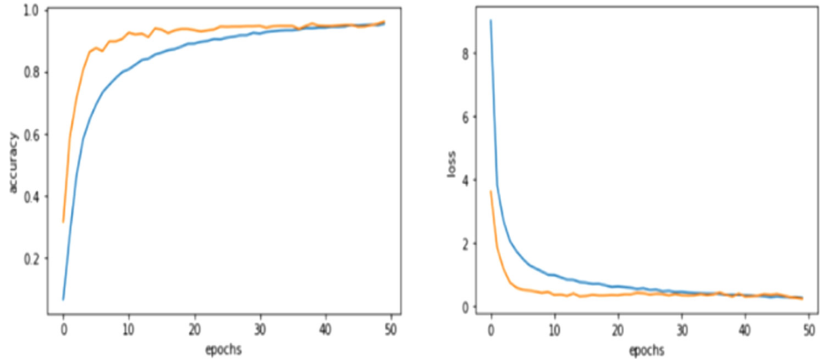

Fig. 4. Accuracy and loss information of the VGG16 model for the best case.

5 Conclusion

In this paper, we proposed a deep learning-based transfer learning method to classify the species of the bird from high-resolution images by using convolutional neural networks. Our proposed approach achieves a maximum accuracy of 96% for the VGG16 model. Furthermore, this model can be used with fine-tuning for various applications including bird detection for wind turbines to avoid an accident from crashing. This model can be coupled with a Web-based app or mobile app for user applications. These apps can work further for creating social awareness about bird species and also to avoid the extinction of these bird species. Thus this application leads to create a more sustainable and friendly environment in the future.

References

1. Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)
2. Atanbori, J., Duan, W., Murray, J., Appiah, K., Dickinson, P.: Automatic classification of flying bird species using computer vision techniques. Pattern Recogn. Lett. **81**, 53–62 (2016)
3. Cui, Y., Song, Y., Sun, C., Howard, A., Belongie, S.: Large scale fine-grained categorization and domain-specific transfer learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
4. Gavali, P., Mhetre, M.P.A., Patil, M.N.C., Bamane, M.N.S., Buva, M.H.D.: Bird species identification using deep learning (2019)
5. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

7. Huang, Y.P., Basanta, H.: Bird image retrieval and recognition using a deep learning platform. IEEE Access **7**, 66980–66989 (2019)
8. 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 2019 7th International Conference on Computer and Communications Management, pp. 38–42 (2019)
9. Kaggle: Bird species dataset (2020). https://www.kaggle.com/gpiosenka/100-bird-species
10. Kumar, A., Das, S.D.: Bird species classification using transfer learning with multistage training. In: Arora, C., Mitra, K. (eds.) WCVA 2018. CCIS, vol. 1019, pp. 28–38. Springer, Singapore (2019). https://doi.org/10.1007/978-981-15-1387-9_3
11. Lopes, M.T., Gioppo, L.L., Higushi, T.T., Kaestner, C.A., Silla Jr, C.N., Koerich, A.L.: Automatic bird species identification for large number of species. In: 2011 IEEE International Symposium on Multimedia, pp. 117–122. IEEE (2011)
12. Marini, A., Facon, J., Koerich, A.L.: Bird species classification based on color features. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4336–4341. IEEE (2013)
13. Nazery, N.A.: Color-based of bird species classification using support vector machine. Ph.D. thesis, Universiti Teknologi MARA (2017)
14. Niemi, J., Tanttu, J.T.: Deep learning-based automatic bird identification system for offshore wind farms. Wind Energy **23**(6), 1394–1407 (2020)
15. Pang, C., Yao, H., Sun, X.: Discriminative features for bird species classification. In: Proceedings of International Conference on Internet Multimedia Computing and Service, pp. 256–260 (2014)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Published as a Conference Paper at ICLR (2015)
18. Singh, R., Ahmed, T., Kumar, A., Singh, A.K., Pandey, A.K., Singh, S.K.:Imbalanced breast cancer classification using transfer learning. IEEE/ACM Trans. Comput. Biol. Bioinf.(2020)
19. Sprengel, E., Jaggi, M., Kilcher, Y., Hofmann, T.: Audio based bird species identification using deep learning techniques. Technical report (2016)
20. Talo, M., Baloglu, U.B., Yıldırım, Ö., Acharya, U.R.: Application of deep transfer learning for automated brain abnormality classification using MR images. Cogn. Syst. Res. **54**, 176–188 (2019)
21. Zhao, B., Feng, J., Wu, X., Yan, S.: A survey on deep learning-based fine-grained object classification and semantic segmentation. Int. J. Autom. Comput. **14**(2), 119–135 (2017). https://doi.org/10.1007/s11633-017-1053-3