International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

IMAGE BASED BIRD SPECIES IDENTIFICATION USING DCNN ARCHITECTURE

Chaithra R[1], Prof. Vasudev T[2], Prof. Amos R[3]
[1]PG Student, Department of MCA, MITM
[2]Professor, Department of MCA, MITM [3]Assistant Professor, Department of MCA, MITM

Abstract - Now a day, the hobby of bird watching is increasing day by day in people. They have become a recreational thing for the people to relive out stress and tension out of their lives and refresh their minds. Bird watching means, it is an art of observing, studying and researching on the different characteristic features of the birds that belongs to different classes. In India, there is no much scope for this area of interest. Also there are no proper tools and technologies available to make use of in carrying out this task. This have become a hindrance to those who wish to pursue their research in this field. And hence we've developed a deep learning image processing model that can identify the name of the bird by processing the images given to them through the datasets. The model achieved an accuracy of 92.73% which is the accuracy produced by VGG architecture.

Key Words: VGG-16, CNN, Keras, Image Data Generator, Epochs.

I. INTRODUCTION

Bird species identification means making the correct prediction by stating that the bird species belong to which class of birds based on its physical features. As per the survey of 2022, according to the experts at the Princeton University, the estimated count is that there are about 9700 species of bird in the world. But still there are some conflicts that whether this count is correct or not. And also according to BirdLife.org, there are about 11,000 species of birds throughout the world. Hence we can say that there are about 10,000 bird species on an average around the globe. Identifying each of these bird species manually or through the knowledge obtained by books, will not be an easy task; even for the expert birders. Because, the traditional way of classification and identification of bird species is based on the morphology of birds. Morphology is a branch of biology that deals with the study of physical feature like shape, size, and structure of animals, plants and microorganisms. And identifying the birds by taking into count all of these features and analyzing these features manually is a tedious job.

The deep learning model belongs to the category of *supervised learning model.* The meaning of the word "supervised learning" refers to as learning in which the machines are trained using the labelled data called as "training dataset". In our current project we've trained the model with 275 classes of birds, where each class of bird approximately contained 200 images of the same bird from different angles.

To carry out the task of model building and model training, we have used the popular image processing algorithm called CNN (Convolutional Neural Networks) along with the pre-trained Deep-CNN architecture VGG (Visual Geometry Groups) which can go up to 19 layers deep. But we're using VGG with 16 layers. VGG is able to classify 1000 images belonging to 1000 different classes with the accuracy of 92.7%. It is one of the popular methods for image classification kinds of projects and it is also easy to use with transfer learning. This VGG-16 has been considered as one of the best models for computer vision technology till date.

For those who have keen interest in birding, this project will serve as a much helpful tool. Because it have been built in such a way that it is so easy to use, where the user can just upload the image of the bird and identify its name.

II. RELATED WORKS

1. **Bird Species Identification from an Image – A project report on machine learning course at Stanford University -**

   The methodology used here is Scikit Learn Library with different algorithms like Naïve Bayes, SVM, KNN, LDA, Decision Trees, Random Forest and One Versus Rest Classifiers with logistic regression. Even though the training accuracy of decision trees was 99.83% and that of Random Forests was 99.39%, the testing accuracies are only 24.35% and 33.58% respectively. And also the Scikit Learn method is not suitable for in-depth learning.

2. **Image Based Bird Species Identification - International Journal of Research in Engineering, IT and Social Sciences, Volume 10 Issue 04, April 2020 -**

   Implementation was done using OpenCV with CNN. It was developed for computational efficiency with focus on real-time applications. The model works fast because it is written in C/C++ and also it is open source. The disadvantages includes factors like, illumination, background, pose, and mainly the privacy issue. Everything is stored somewhere online which is owned by corporations or freely visible for everyone. The testing accuracy achieved was only 80%.

3. **Bird Identification by Image Recognition – Helix Vol. 8(6), Published: 31st October, 2018**

   The methodology used for the intended purpose is the pre-trained model AlexNet along with CNN. The architecture consists of eight layers out of which five was convolution layers and three fully-connected layers. The accuracy achieved was 100% but, the pit of AlexNet model is not high and hence it struggles to grasp the features of the image. And also, it takes additional time to achieve higher accuracy rates. The AlexNet has over 60 million parameters but a major issue is in terms of over fitting. And also the use of large convolution filters (5*5) is not encouraged.

4. **Image Based Bird Species Identification using Deep Learning – IJCRT Vol. 9, Issue 7 July, 2021 -**

   The methodology used was CNN with Residual Network (ResNet). ResNet have 34 layer network in minimum and ResNet 152V2 is used. The accuracy achieved was 95%. But deeper network usually requires weeks for training. Using this technology, if we want to implement a project as an application, then it is practically infeasible.

5. **Image Based Bird Species Identification using Convolutional Neural Network – IJIRT Vol. 8 Issue 11, April 2022 –**

   This implementation was also done using CNN with OpenCV. As discussed in the previous paper (index no.2) the OpenCV technology has the drawbacks like, illumination issue, background, pose and the important issue which is privacy. The accuracy obtained was 83%.

6. ***Image Based Bird Species Identification using Convolutional Neural Network – IJIRT Vol. 9 Issue 6, June 2020:*** Used CNN algorithm along with VGG, Keras and Adam Optimizer. The accuracy achieved on training set is 93% and on testing set is 84%. The backend technology used was Tensorflow and layers of Convolution (Conv) and Rectified Linear Units (ReLU) were added together.

7. **Bird Species Identification Using Modified Deep Learning (CNN) Model by Integrating Cascaded Softmax layer – JETIR May 2021, Vol. 8 Issue 5 -**

   The methodology used was modified deep CNN by modifying it in two ways – one was by replacing top fc layers by h-fully connected layers. Another one was by introducing the GLM Loss. By implementing this technology, the accuracy achieved was 98% but replacing top fc layers by convolution layer almost produces the exact behaviour or output. And GLM has the drawbacks like – it is sensitive to outliers, low predictive power, predicted variables need to be uncorrelated and it does not select features.

8. ***Bird Species Identification using Image Mining & CNN Algorithm – IRJET Vol. 7 Issue 2, Feb 2020:*** The methodology used was image mining technique. Feature extraction is the prominent step, as the features of an image is the visual property of it. The prominent features considered here are colour, texture and shape characteristics. The colour attribute is the most important feature used to search an image. But colour extraction feature has the following disadvantages:

Feature	Cons
Conventional color histogram	No spatial information
Color moment	Has low precision
Fuzzy color histogram	Needs more computation
Color Structure Descriptor	No spatial information
Color Co-occurrence Feature	Codebook needed
Color coherence vector	High complexity

Table 2.1. Cons of Feature Extraction methods

Features	Cons
Tamura Feature	Highly complex
Wavelet filter	Has low precision
Gabor filter	Expensive computation
Gabor moment	Expensive computation
LBP, LTP, LTRP	Only suitable for gray scale image

Table 2.2. Cons of Texture Extraction methods

And therefore, all these disadvantages have led us to use the technology called ***VGG16*** which can go up to 19 layers deep, but in our proposed system we have made us of VGG which is 16 layers deep. The current technology overcome most of the disadvantages of the above mentioned technologies with an accuracy of 92.7%. The main advantage of using this technology along with CNN is that it works for large amount of data. And hence, the proposed model have made use of this advantage of VGG and the model have been trained with more than 39,000 images belonging to 275 classes of birds and achieved the accuracy of 92.73%.

III. TOOLS AND TECHNOLOGIES USED

A brief description about the usage of tools and technologies are important with respect to project development. The software tools which are required for the project will also serve as a base for carrying out the feasibility study. The below mentioned are the most prominent technologies which are being used in our bird species identification project.

i. **CNN** – Stands for Convolutional Neural Network. It is a deep learning algorithm which takes input which is nothing but an image and assign weights and biases for them in numerous aspects for smooth functioning of differentiating the images from one another. It is the best suited algorithm for image processing kind of projects. The algorithm is mostly used in image recognition because of its high accuracy yielding nature.

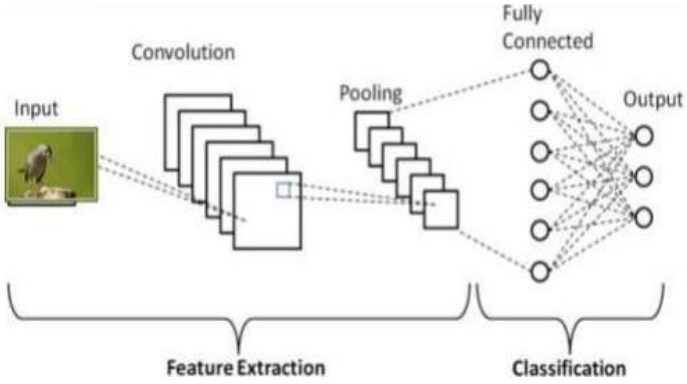

Fig. 3.1. Layers of CNN

The CNN algorithm contains four main layers through which the image processing is done. The four layers of the algorithm are: Convolution Layers, Pooling Layer, Activation Layer and the Fully Connected Layer.

- *Convolution Layer* – This is the first layer in the CNN architecture and this is the layer which is mainly associated with the extraction of features from the image data. Convolution refers to a mathematical operation performed on two functions, say, **f** and **g** which then give rise a resultant third function. This operation is performed on the image data and the filter with a specific size of MxM. The output obtained from this layer is called as feature map which gives the particulars about the input image such as corners and edges.

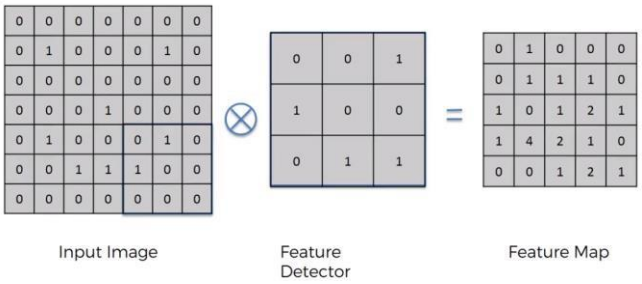

Fig. 3.2. Feature Map obtained from Convolution Layer

- *Pooling Layer* – The convolution layer is preceded by this layer and this is the second layer of the algorithm. The main aim of this layer is to reduce the size of the feature map obtained from the first layer in order to reduce the computational costs. The process of this reduction is done by lessening the links between the layers and individually functioning on each of the generated feature maps. There are several types of pooling techniques available based on the methodology used. The different types of pooling techniques available are Max Pooling, Average Pooling and Sum Pooling.

In the Max Pooling technique, the largest element from each section in the feature map is obtained to form a convoluted feature map. In the case of Average Pooling, the mean of the each section in the feature map is taken and in the Sum Pooling, the sum of the particular predefined section is obtained to bring down the size of the feature map. In this current project, Max Pooling technology is used.

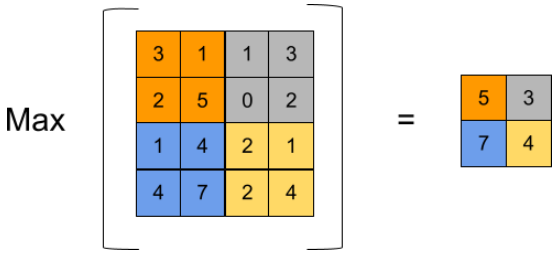

Fig. 3.3. Convoluted Feature Map obtained by Pooling Layer

- *Activation Layer* – The activation layer is the most important layer in the CNN architecture and it is usually located at the end of the neuron. This layers will deliver the outputs based on the given input. Meaning, the output will be a smaller size, if the size of the input is smaller and a larger value is outputted if the input value exceeds the given threshold value. But, the activation function gets fired if and only if the input is large enough to process. Otherwise, it does nothing. The activation layer is the one which adds non-linearity to the network, which is always beneficial in deep learning. Non-linearity means, the neural network has activation functions like ReLU, Sigmoid, Tanh or Softmax in any of its layers or even in more than one layer.

- *Fully Connected Layer* – The fully connected layers usually called as FC layers, consists of weights and biases along with the neurons and this layer is used to ally neurons between two different layers. These FC layers are usually placed before the output layers. The actual process of classification begins to take place in this layer where some mathematical operations are performed on the images that is obtained from the previous activation layer, where, images from this layer will be flattened and then fed to FC layers for further processing.

ii. **VGG –** Stands for Visual Geometry Groups, is a pre-trained deep CNN architecture which was introduced in 2014. The main reason for developing this architectural tool is to see how the depth of the network will have an effect on the accuracy of the model.

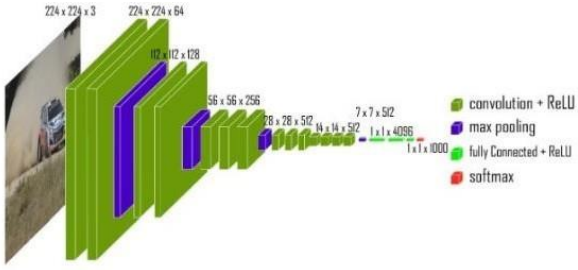

Fig. 3.4. Detailed Architecture of VGG-16

The earlier technology used one convolution layer with one large weighty kernel. But this technology introduced the concept of assembling multiple convolution layers with smaller kernel size as a substitute of using the older technique. This made the number of outputs to reduce at the end of the output layer. One more advantage of this technology is that the number of epochs required in the model fitting process has significantly reduced, say around 10-15 epochs are more than enough to achieve the desired accuracy. Whereas, the older techniques used epochs till 100 which then caused over-fitting problem. Since, VGG is a deep learning architecture, the introduction of more layers has been advantageous over here as it helps us to understand the features of the image in depth. One more advantage of the VGG tool is that it has brought a significant improvement in the processing speed and accuracy. Since we have introduced the small sized kernels, there is also an increase in the non-linearity because of this.

Fig. 3.5. Simplified architecture of VGG-16

From the above figure it can be seen that there are thirteen convolution layers, five pooling layers and three dense layers which makes up to 21 layers. But the learnable layers or weights parameters are only 16 layers. A weight is the parameter within a neural network which transfigure the input data inside the network's hidden layers. And hence it is called as VGG-16. From the figure 2.3.4, it can be seen that the convolution layers Conv1, Conv2, Conv3 and Conv4 have 64, 128, 256 and 512 filters respectively and each of these filters are of the size 3x3. The filters are such things which detects the spatial patterns such as the edges and corners in an image. The higher the quantity of filters, the higher is the amount of parameters that the network is able to extract from the inputted image data and work on accuracy.

iii. **Keras – Image Data Generator –** The Keras library in Python is a high-grade, deep learning application programming interface (API) which was developed by Google for the implementation of the neural networks and to make the work easier. This library, for the beginners, is extremely user-friendly and easy to use because provides a python front end with high level abstraction. But also, this module allows a user to toggle between different back end tools such as Tensorflow, Theano, PlaidML, MXNet and CNTK. Among the above five back-end frameworks, the Tensorflow technique has accepted Keras as its official high-level API and hence, Keras is implanted in Tensorflow. Importing one will automatically load the other. Apart from all these, the other reasons for using Keras is that:

- It's an API that was developed in order to make it as user-friendly and simple.
- The prototyping time in Keras is less.
- Since Keras runs on topmost layer of Tensorflow, it is relatively fast.
- This module will runs smoothly on both CPU and GPU.

The Image Data Generator library of Keras will let us augment the image data by applying random transformations on the image data. This process will not only make the developed model vigorous but will also save up on the overhead memory. The image augmentation is a process in which some kinds of transformations are applied on the original image which will result in the transformed multiple copies of the same image where each copy is different from one another. This method will increase the size of the dataset. And thereby, with just few lines of code we can generate large amount of similar images instantly without the hassle of downloading/collecting new data. Some of the augmentation techniques with Image Data Generator class are: Random Rotations – the image will be rotated clockwise between 0 and 360 degrees, Random Shifts, Random Flips where the flips may be horizontal or vertical, Random Brightness – means randomly darkening and lightening the images, Random Zoom etc.

IV. RESULTS AND CONCLUSION

Fig. 4.1. Bird species Indian Roller – Predicted Correctly

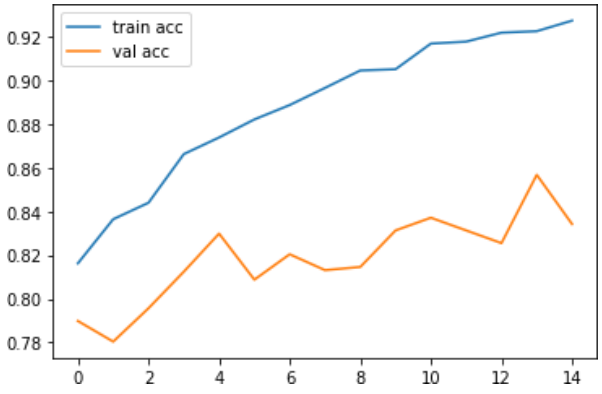

Fig. 4.2. Graph of value accuracy plotted from the values obtained from model fitting

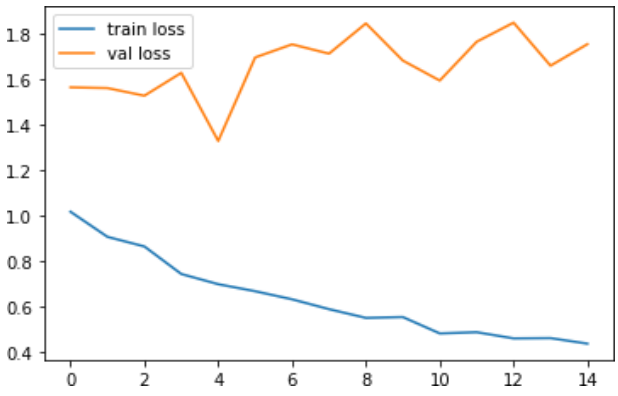

Fig. 4.3. Graph of value loss plotted from the values obtained from model fitting

For the given input image "Indian Roller" the model have accurately predicted the name. And also the graphs for "Value Loss" and "Value Accuracy" are plotted based on the result of the model fitting step. From the graphs it is observed that the loss takes much more time to converge to minima. But when compared to already existing Bird Species Identification Models, some of them have achieved higher accuracy but with less training data and

some models have achieved less accuracy due to the limitations of the tools and techniques used. But our model has been trained with more than 39,000 images with 16 layered architecture and achieved a good accuracy of 92.73%.

To conclude this project, the achieved accuracy is approximately equal to 92.7% which is same as the accuracy which the VGG-16 architecture is capable of achieving. For this accuracy rate, the model is working fine by predicting the name of the bird species accurately. All possible tests were done on the model to test for the effectiveness and robustness and it is observed that the model is robust and does the intended job as per the requirement without creating any problems for the user.

V. FUTURE ENHANCEMENTS

The project has a very vast scope in the future days if it is deployed as a web application or mobile application. To deploy this model as an application many frameworks are available to use with the python language like, *Streamlit, Pytorch, Scikit Learn, Flask, Tensorflow* etc.

While achieving the accuracy, we have used the Adam Optimizer. There are many optimizers available like, *SGD, Adadelta, Adamax, Adagrad, Nadam, Ftrl, RMSprop* etc. which will give better accuracy.

From using VGG-16 we achieved the accuracy of only 92.73%. This is because of some of the disadvantages of the VGG-16 architecture. In this architecture, it is observed that the loss will take more time to converge to minima and also constant learning and re-learning is a problem in VGG-16 which leads to the explosion of gradients and also the model experiences the vanishing gradient problem. To solve these problems and a newer ResNet technique can be used to achieve even better accuracy.

REFERENCES

[1] K Amaravathi, N Lokeshwari, "Bird Species Identification Using Modified Deep Learning (CNN) Model by Integrating Cascaded Softmax Layer", JETIR May 2021, Vol. 8, Issue 5

[2] B V Satya Sai, S Apuroop, S Mounika, D Sai Kumar, "Identification of Bird Species using Convolutional Neural Networks", a report submitted to Anil Neerukonda Institute Of Technology & Sciences (autonomous)

[3] Saundarya Junjur, Punam Avhad, Deepika Tendulkar," Bird Species Identification using Image Mining and CNN algorithm", IRJET, Feb 2020, Volume 07, Issue 02

[4] Aditya Bhandari, Ameya Joshi, Rohit Patki, "Bird Species Identification from an Image", A report submitted to Institute of Computational Mathematics and Engineering, Stanford University.

[5] Anisha Singh, Akarshita Jain, Bipin Kumar Rai, "Image Based Bird Species Identification", International Journal of Research in Engineering, IT and Social Sciences, April 2020, Volume 10, Issue 04, Page 17-24.

[6] Madhuri A. Tayal, Atharva Mangrulkar, Purvashree Waldey, Chitra Dangra, "Bird Identification by Image Recognition", Helix vol. 8(6), Published: 31st October 2021.

[7] A V Siva Krishna Reddy, Dr. M A Srinuvasu, K Manibabu, Ch B V Sai Krishna, D Jhansi, "Image Based Bird Species Identification using Deep Learning", IJCRT, Volume 9, Issue 7, July 2021.

[8] Surendhiran Tamilalagan, "Image Based Bird Species Identification using Convolutional Neural Network", IJIRT, Volume 8 Issue 11, April 2022

[9] Manar Adbulkareem Aj-Abaji, "A Review of Content Based Image Mining System", Journal of University of Babylon for Pure and Applied Sciences, Vol. 27, No. 4, 2019.

[10] Satyam Raj, Saiaditya Garyali, Sanu Kumar, B.E. Scholar, "Image Based Bird Species Identification using Convolutional Neural Network", IJERT, Vol. 9, Issue 06, June 2020.