Bird Species Identification Using Machine Learning

Prof. Prerana Chaithra[1], Avinabh[2], Ashwin M[3], Rehan Inamdar[4], Pratik Kumar[5]

[1]Assosciate Professor, Dept of ISE
[2, 3, 4, 5]Dept of ISE
[1, 2, 3, 4, 5] Sapthagiri College of Engineering, VTU
Bengaluru, Karnataka, India

Abstract- *Now a day some bird species are being found rarely and if found classification of bird species prediction is difficult. Naturally, birds present in various scenarios appear in different sizes, shapes, colours, and angles from human perspective. Besides, the images present strong variations to identify the bird species more than audio classification. Also, human ability to recognize the birds through the images is more understandable. So this method uses the Caltech-UCSD Birds 200 [CUB-200-2011] dataset for training as well as testing purpose. By using deep convolutional neural network (DCNN) algorithm an image converted into grayscale format to generate autograph by using tensor flow, where the multiple nodes of comparison are generated. These different nodes are compared with the testing dataset and score sheet is obtained from it. After analysing the score sheet it can predicate the required bird species by using highest score. Experimental analysis on dataset (i.e. Caltech-UCSD Birds 200 [CUB-200-2011]) shows that algorithm achieves an accuracy of bird identification between 80% and 90%. The experimental study is done with the Ubuntu 16.04 operating system using a Tensor flow library.*

Keywords- Autograph; Caltech-UCSD; DCNN; Grayscale pixels; Tensor flow

I. INTRODUCTION

BIRD behaviour and population trends have become an important issue nowadays. Birds help us to detect other organisms in the environment (e.g. insects they feed on) easily as they respond quickly to the environmental changes [2]. But, gathering and collecting information about birds requires huge human effort as well as becomes a very costlier method. In such case, a reliable system that will provide large scale processing of information about birds and will serve as a valuable tool for researchers, governmental agencies, etc. is required. So, bird species identification plays an important role in identifying that a particular image of bird belongs to which bird species belongs to which category by using an image.

The identification can be done through image, audio or video. An audio processing technique makes it possible to identify by capturing the audio signal of birds. But, due to the mixed sounds in environment such as insects, objects from real world, etc. processing of such information becomes more complicated. Usually, human beings find images more effective than audios or videos.

So, an approach to classify bird using an image over audio [8] or video is preferred. Bird species identification is a challenging task to humans as well as to computational algorithms that carries out such a task in an automatic fashion.

Since many decades, ornithologists are facing problems in bird species identification. Ornithologists require studying all the details of birds such as their existence in environment, their biology, their distribution, their ecological impact, etc. Bird identification is usually done by ornithology experts based on classification proposed by Linnaeus: Kingdom, Phylum, Class, Order, Family, and Species [1].

As image based classification systems are improving the task of classifying, objects is moving into datasets with far more categories such as Caltech-UCSD. Recent work has seen much success in this area. CaltechUCSD Birds 200(CUB-200-2011) is a well-known dataset for bird images with photos of 200 categories [4]. The dataset contains birds that are mostly found in Northern America. Caltech-UCSD Birds 200 consists of 11,788 images and annotations like 15 Part Locations, 312 Binary Attributes, 1 Bounding Box. The identification can be done through image, audio or video. An audio processing technique makes it possible to identify by capturing the audio signal of birds. But, due to the mixed sounds in environment such as insects, objects from real world, etc. processing of such information becomes more complicated.A vocabulary of 28 attribute groupings and 312 binary attributes (e.g., the attribute group "bellycolor" contains 15 different color choices) was selected based on an online tool for bird species identification. All attributes are visual in nature, with most pertaining to a color, pattern, or shape of a particular part.

In this paper, instead of recognizing a large number of disparate categories, the problem of recognizing a large number of classes within one category is investigated that of birds. Classifying birds pose an extra challenge over categories, because of the large similarity between classes. In addition, birds are non-rigid objects that can deform in many

ways, and consequently there is also a large variation within classes. Previous work on bird classification has deal with a small number of classes, or through voice.
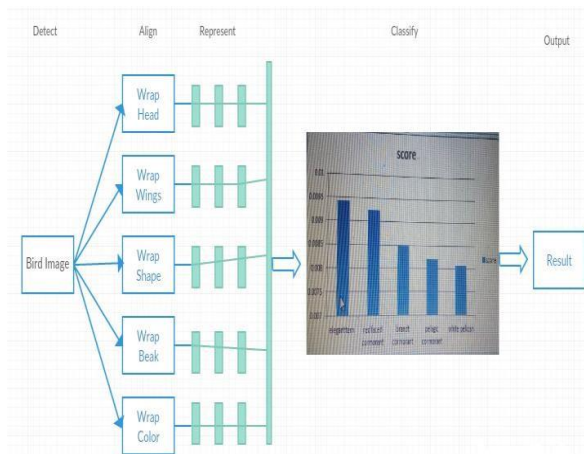

Fig 1: Process of classification

The figure 1 represents the process of detecting the bird from image. The image is getting upload first then from that image the various alignments will be considered such as head, body, color, beak and entire image. Further, each alignment is given through deep convocational network to extract features out from multiple layers of network [3]. After that representation of the image will get consider. Then on the basis of it the classifying result will get generated (i.e. features are aggregated to transfer it to classifier) and the bird species will get found.

This paper is assembled in format: Section II covers the parameters one can consider while identifying a bird visually. Section III contains methodologies used for developing the proposed system. Section IV represents overall flow of the system in detail.

II. BACKGROUND

Basically bird identification is done visually or acoustically. The main visual components comprise of bird's shape, its wings, size, pose, color, etc. However, while considering the parameters time of year must be taken into consideration because bird's wings changes according to their growth. The acoustics components comprise the songs and call that birds make [7]. The marks that distinguish one bird from another are also useful, such as breast spots, wing bars which are described as thin lines along the wings, eye rings, crowns, eyebrows. The shape of the beak is often an important aspect as a bird can recognized uniquely. The characteristics of bird such as shape and posture are the mostly used to identify birds.

Mostly experts can identify a bird from its silhouette because this characteristic is difficult to change. A bird can also be differentiated using its tail. The tail can be recognized in many ways such as notched, long and pointed, or rounded. Sometimes legs are also used for recognizing an image in format long, or short [10].

By considering a single parameter will not yield an accurate result. So, multiple parameters are to be considered in order to get appropriate output. The size of a bird in an image varies depending upon factors such as the resolution, distance between the birds and the capturing device, and the focal distance of the lens. Therefore, based on a practical observation for large number of images, images are differentiated on the basis of color which consists of various pixel. In depth it is found that greater the image quality greater is its accuracy.

The automatic bird species identification for bird images project present a series of comparison conducted in a CUB200 dataset composed of more than 6,000 images with 200 different categories [6]. In this paper, they have considered two different color spaces, RGB and HSV, and a different number of species to be classified. If the image consists of more than 70% of the pixels, the accuracy of output was ranging from 8.82% to 0.43% [1].

III. METHODOLOGY

For developing the system certain methodologies have been used. They are as follows: Dataset (Caltech-UCSD Birds 200), Deep Convolutional Neural Network, Unsupervised learning algorithm, etc.

Algorithm:

In this experiment, unsupervised learning algorithm has been used for developing the system, because the inputted image defined is not known. Also, the data which is given to unsupervised learning algorithm are not labelled, i.e. only the input variables(X) are given with no corresponding output variables. In unsupervised learning, algorithms discover interesting structures in the data themselves. In detail, clustering is used for dividing the data into several groups [4].

In depth, deep learning models used to find vast number of neurons. Deep learning algorithms learn more about the image as it goes through each neural network layer. For classifying **Neural Network** is used. Figure 2 represents layers of neural networks for feature extraction. The neural network is a framework for many machine learning

algorithms. Neural networks consist of vector of weights (W) and the bias (B).
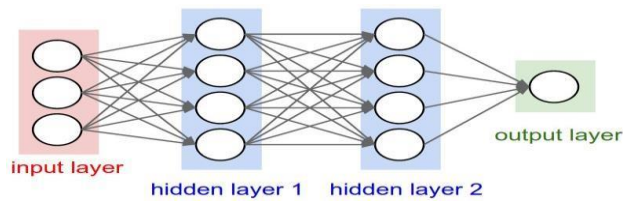

Fig 2: Three layers of Neural Network

In deep learning, convolutional neural network (CNN) is a class of deep neural network mostly used for analysing visual images. It consists of an input layer and output layer as well as multiple hidden layers. Every layer is made up of group of neurons and each layer is fully connected to all neurons of its previous layer. The output layer is responsible for prediction of output. The convolutional layer takes an image as input, and produces a set of feature maps as output [2]. The input image can contain multiple channels such as color, wings, eyes, beak of birds which means that the convolutional layer perform a mapping from 3D volume to another 3D volume. 3D volumes considered are **width, height, depth.** The CNN have two components:

1)   Feature extraction part: features are detected when network performs a series of convolutional and pooling operation.
2)   Classification part: extracted features are given to fully connected layer which acts as classifier.
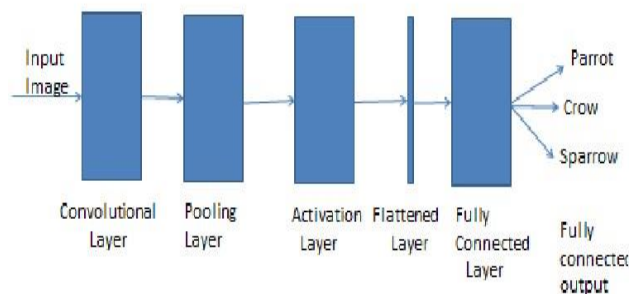

Fig 3: Convolutional Neural Network Layers

CNN consists of four layers: convolutional layer, activation layer, pooling layer and fully connected. Convolutional layer allows extracting visual features from an image in small amounts. Pooling is used to reduce the number of neurons from previous convolutional layer but maintaining the important information. Activation layer passes a value through a function which compresses values into range. Fully connected layer connects a neuron from one layer to every

neuron in another layer. As CNN classifies each neuron in depth, so it provides more accuracy.

Image classification: image classification in machine learning is commonly done in two ways:

1)   Grayscale
2)   Using RGB values

Normally all the data is mostly converted into grayscale. In grayscale algorithm, computer will assign values to each pixel based on how the value of the pixel is it**.** All the pixel values are put into an array and the computer will perform operation on that array to classify the data.

Library:

Tensorflow is open source software library which is created by Google. It gives developers to control each neuron known as a "node", so that the parameters can be adjusted to achieve desired performance. Tensorflow has many built-in libraries for image classification [3]. Tensorflow is responsible for creating an autograph which consists series of processing nodes. Each processing node in the graph represents an operation such as mathematical operation and connection or edge between nodes. With the help of python language Tensorflow provides programmer to perform these operations.

Dataset:

A dataset is a collection of data**.** For performing action related to birds a dataset named **Caltech-UCSD Birds 200** (CUB-200-2011)is used**.** It is an extended version of theCUB-200 dataset, with roughly double the number of images per class and also has new part location annotations for higher accuracy [8]. The detailed information about the dataset is as follows: Number of categories: 200, Number of images: 11,788, Annotations per image: 15 Part Locations, 312 Binary Attributes, 1 Bounding Box

Fig 4: Caltech-UCSD Dataset

Whenever a user will upload an input file on website, the image is temporarily stored in database. This input file is then feed to system and given to CNN where CNN is coupled with trained dataset. A CNN consists of various convolutional layers. Various alignments/features such as head, body, color, beak, shape, entire image of bird are considered for classification to yield maximum accuracy. Each alignment is given through deep convocational network to extract features out from multiple layers of network. Then an unsupervised algorithm called deep learning using CNN is used to classify that image.

The below figure no. 5 represents the actual flow of the proposed system. To develop such system a trained dataset is required to classify an image. Trained dataset consists of two parts trained result and test result. The dataset has to be retrained to achieve higher accuracy in identification using retrain.py in Google Collab. The training dataset is made using 50000 steps taking into consideration that higher the number of steps higher is its accuracy. The accuracy of training dataset is 93%. The testing dataset consists of nearly 1000 images with an accuracy of 80%. Further, dataset is validated with an accuracy of 75% to increase the performance of system.
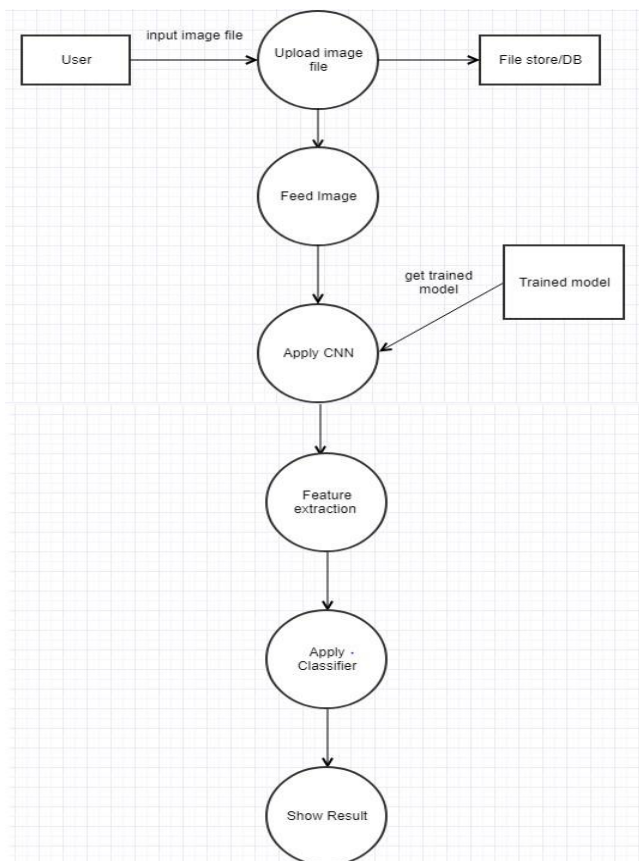
IV. PROPOSEDAPPROACH

Fig 5: Flow of System

Further, a grayscale method is used to classify the image pixel by pixel. These features are then aggregated and forwarded to classifier. Here, the input will be compared against the trained dataset to generate possible results. During classification, an autograph is generated which consist of nodes that ultimately forms a network. On basis of this network, a score sheet is generated and with the help of score sheet output will be produced.

V. EXPERIMENTALANALYSIS

The evaluation of the proposed approach for bird species classification by considering color features and parameters such as size, shape, etc. of the bird on the Caltech-UCSD Birds 200 (CUB-200-2011) dataset. This is an image dataset annotated with 200 bird species which includes 11,788 annotated images of birds where each image is annotated with a rough segmentation, a bounding box, and binary attribute annotations. In this the training of dataset is done by using Google-Collab, which is a platform to train dataset by uploading the images from your local machine or from the Google drive.

After training labelled dataset is ready for classifiers for image processing. There are probably average 200 sample images per species are included in dataset of 5 species which are directly captured in their natural habitat hence also include the environmental parameters in picture such as grass, trees and other factors. Here bird can identify in their any type of position as main focus is on the size, shape and color parameter. Firstly, these factors are considered for segmentation where RGB and grayscale methods are used for histogram. That is the image converted into number of pixels by using grayscale method, where value for each pixel is created and value based nodes are formed which also referred as neurons. These neurons relatively defined the structure of matched pixels is simply like graph of connected nodes.

According to the nodes formed the autograph is generated which understandable by Tensorflow to classify the image. This autograph is then taken by classifiers and image is compared with the pre trained dataset images of Caltech UCSD and the score sheet is generated. The score sheet is a result which contains top 5 match results by which the highest matching value of score sheet is the result of bird species. Here a trial has made to implement 80% accuracy in result by training the Caltech UCSD.

For example, consider below Figure No.6 as input image given to the system for classification of bird which belongs to Northern America. Let's see how it is being evaluated.

Fig 6: Input Image

The system generates following scoresheet after classification which tells us the possibilities that above selected bird belongs to which species.

Table No. 1: Score Sheet

Sr.No	Species	Score Obtained
1	Elegant tern	0.00943
2	Red faced cormorant	0.00924
3	Brant cormorant	0.0085
4	Pelagic cormorant	0.0082
5	White pelican	0.00808

The table no.1 shows the scoresheet based on the result generated by the system. After analysis of these result it has observe that, the species those are having the highest score has been predicted as a required species. this result can be shown in the following graph.
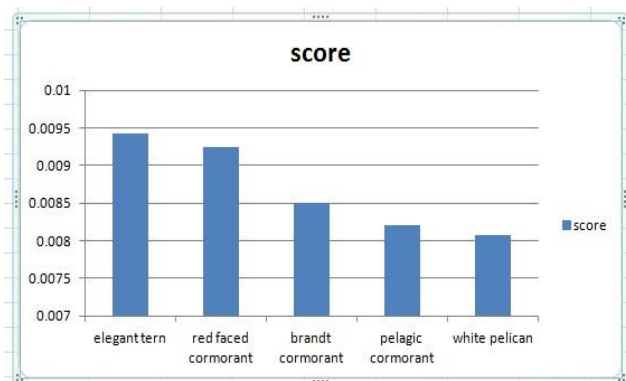

Fig 7: Graph format of Score sheet

Comparison:

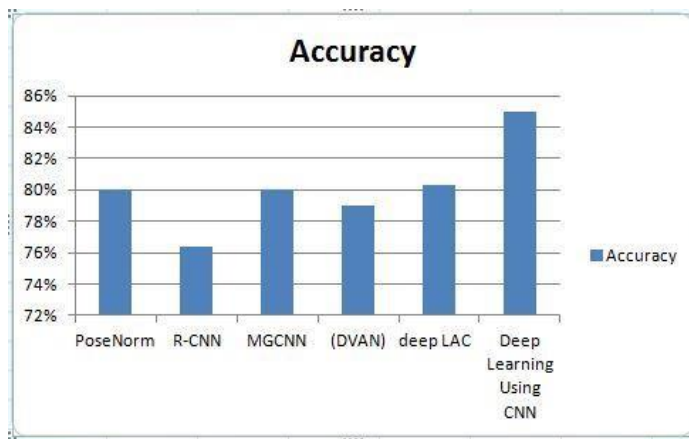| Sr. No	Method	Accuracy
1)	Pose Norm	80%
2)	Part-based R-CNN	76.4%
3)	Multiple granularity CNN	80%
4)	Diversified visual attention network (DVAN)	79%
5)	The deep LAC localization, alignment, and classification	80.3%
6)	Deep Learning Using CNN	85%

Fig 8: Graph format of Accuracy

After analysing the data, it is observed that if a single parameter is used the accuracy obtained is lesser. But, if a combined method is used that is by considering parameters such as pose, wings, color, beak, legs, etc. the accuracy of the project get increase.

VI. CONCLUSION

The present study investigated a method to identify the bird species using Deep learning algorithm (Unsupervised Learning) on the dataset (Caltech-UCSD Birds 200) for classification of image. It consists of 200 categories or 11,788 photos. The generated system is connected with a user-friendly website where user will upload photo for identification purpose and it gives the desired output. The proposed system works on the principle based on detection of a part and extracting CNN features from multiple convolutional layers. These features are aggregated and then given to the classifier for classification purpose. On basis of the results which has been produced, the system has provided the 80% accuracy in prediction of finding bird species.

VII. FUTURESCOPE

1) Create an android/ios app instead of website which will be more convenient to user.
2) System can be implemented using cloud which can store large amount of data for comparison and provide high

computing power for processing (in case of Neural Networks).

REFERENCES

[1] Tóth, B.P. and Czeba, B., 2016, September. Convolutional Neural Networks for Large-Scale Bird Song Classification in Noisy Environment. In CLEF (Working Notes) (pp. 560-568).

[2] Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2007(1), pp.64-64.

[3] Pradelle, B., Meister, B., Baskaran, M., Springer, J. and Lethin, R., 2017, November. Polyhedral Optimization of TensorFlow Computation Graphs. In 6th Workshop on Extreme-scale Programming Tools (ESPT-2017) at The International Conference for High Performance Computing, Networking, Storage and Analysis (SC17).

[4] Cireşan, D., Meier, U. and Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745.

[5] Andr´eia Marini, Jacques Facon and Alessandro L. Koerich Postgraduate Program in Computer Science (PPGIa) Pontifical Catholic University of Paran´a (PUCPR) Curitiba PR, Brazil 80215–901 Bird Species Classification Based on Color Features

[6] Image Recognition with Deep Learning Techniques ANDREI- PETRU BĂRAR, VICTOR-EMIL NEAGOE, NICU SEBE Faculty of Electronics, Telecommunications & Information Technology Polytechnic University of Bucharest.

[7] Xception: Deep Learning with Depthwise Separable Convolutions François Chollet Google, Inc.

[8] Zagoruyko, S. and Komodakis, N., 2016. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928.

[9] Inception-v4,Inception-ResNetand the Impact of Residual Connectionson Learning Christian Szegedy, Sergey Ioffe,VincentVanhoucke ,AlexandrrA.Alemi

[10] Stefan Kahl, Thomas Wilhelm-Stein, Hussein Hussein, Holger Klinck, Danny Kowerko, Marc Ritter, and Maximilian Eibl Large-Scale Bird Sound Classification using Convolutional Neural Networks

[11] Thomas Berg, Jiongxin Liu, Seung Woo Lee, Michelle L. Alexander, David W. Jacobs, and Peter N. BelhumeurBirdsnap: Large-scaleFine-grainedVisualCategorization