Bird Classification using Deep Learning

Piyush Bidwai, Vaibhav Mahalle[2], Eshan Gandhi[3], Sharda Dhavale[4]

[1,2,3]*Student, B.E. Electronics & Telecommunication, Pimpri Chinchwad College of Engineering, Pune*
[4]*Assistant Professor, Dept. of Electronics & Telecommunication, Pimpri Chinchwad College of Engineering, Pune*

---***---

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 may be in different sizes, shapes, colors and angles from human perspective. In 21st century the world is moving towards digitalization and effective monitoring system in every sector. Most of the population uses mobile phones, so it is possible that anyone can capture the bird's image. By using Convolutional neural networks (CNN) algorithm that image is converted into a grey scale format to generate autograph by using Pytorch model, 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 analyzing the score sheet It can predict the required bird species.*

Key Words: *Classification, Convolutional neural network (CNN), Pytorch , dataset, grey scale format.*

1. INTRODUCTION

In today's scenario, bird behavior and population has become an important issue. Birds help to detect other organisms in the environment. Basically bird species identification from their sound is an important and challenging problem. There are also some different methods through which we can monitor bird species. As many birds migrate according to the environmental changes, so the use of automated methods for bird species identification is an effective way to evaluate the quantity and diversity of the birds which appear in the region.

Artificial intelligence and machine learning sounded like a science fiction prophecy of a technological feature. Image recognition is one of the most accessible application of it. Machine learning embedded in consumer websites and applications is changing the way visual data is organized and processed. Image recognition and identification has grown so effectively because it uses deep learning algorithm. This is the machine learning method designed to resemble the way a human brain functions. Thus the computers are taught to recognize visual elements within an image.

By noticing emerging patterns and relying on large databases, algorithms can make sense of images and formulate relevant categories and tags. Classification of birds is generally based on its classes within one category. Classifying birds pose is an extra challenge over categories, because of the similarity between the classes. So, the bird's species identification plays an important role in identifying that a particular image of bird belongs to which species. Bird species identification means predicting the bird species belongs to which category by using an image.

1.1 Existing System

From ancient times, bird identification has been a difficult task for many Ornithologists. They are required to study all the details of birds such as their existence in environment, their biology, their distribution etc. Bird classification is usually done by ornithology experts based on classification proposed by Linnaeus: Kingdom, Phylum, Class, Order, Family and Species.

1.2 Methodology

Deep learning is a branch of Machine learning based on a set of algorithms that attempt to model high level abstraction in data by using a deep graph with multiple processing layers, composed of multiple linear and nonlinear transformation. Widely used algorithm of deep learning in image classification are convolutional neural network (CNN). Therefore, CNN allows classification based on Tenser Flow, Pytorch Models.

So in this system, the software contributes a major part that utilizes python language, Pytorch model and Raspberry pi to perform this bird classification. The input image is taken from electronic device and is converted into grey scale format. Deep learning models used to find vast number of neurons. These algorithms learns more about image as it goes through several neural networks. Following figure is used classify the image for feature extraction.
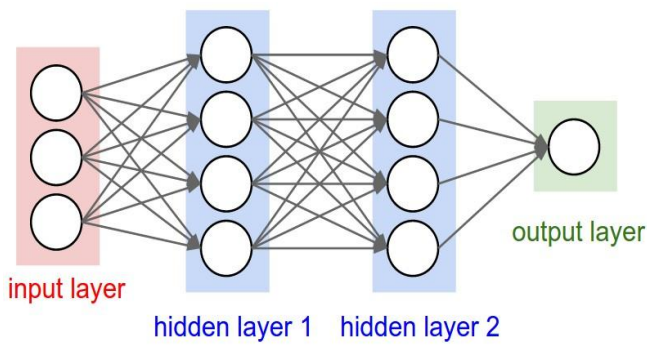
Fig 1 block diagram

Above diagram shows the three layers of neural network.

1.3. Algorithm

In this system, deep learning algorithm has been used for developing the system, because the inputted image is not known. CNN is a class of deep neural network mostly used for analyzing 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 a classifier.

CNN consist 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.
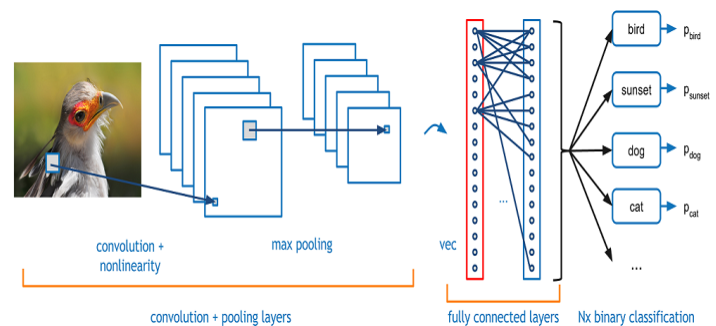

Image classification: image classification in machine learning is commonly done in two ways

1. Grey scale

2. Using RGB values.

Normally all the data is mostly converted into grey scale. In grey scale 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:

Pytorch model is majorly used in this system. Pytorch is open source software library which is developed by Facebook's AI Research lab (FAIR). It is an open source machine learning library based on Torch library used for application such as computer vision and natural language processing. The heart of pytorch is python language. Pytorch is released under modified BSD license.

Pytorch provides two features:

1. Tensor computing (like NumPy) with strong acceleration via graphics processing unit (GPU).

2. Deep neural networks built on a tape based autodiff system.

Dataset:

Dataset is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question
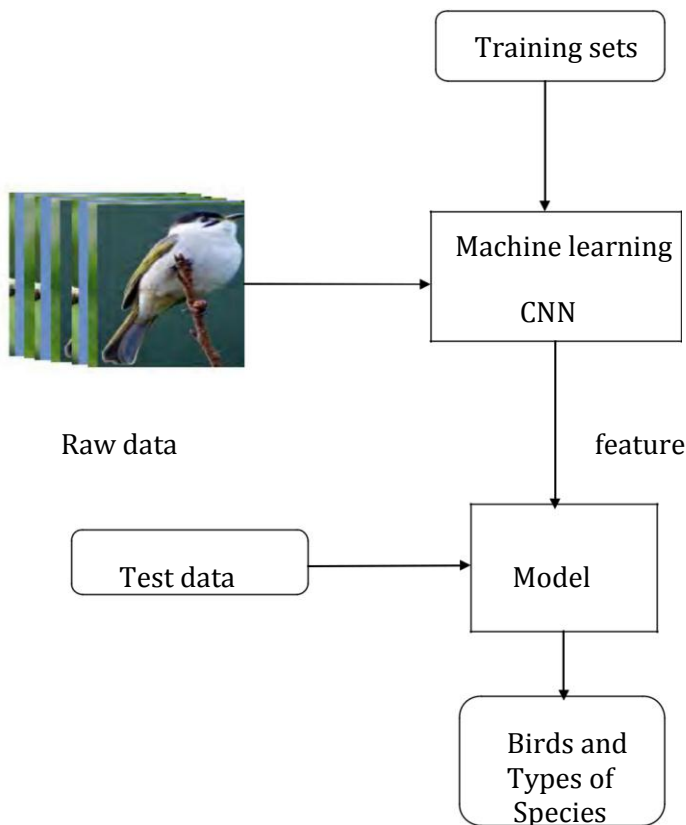
3. Proposed approach:

Fig -2: Proposed System Block Diagram

3.1 Block diagram Explanation:

1. Raw Data:

   It is the data in unstructured form. We cannot predict some relevant information from it. It represents a single, implicitly structured data item in the table.

2. Training set:

   The training dataset comprises of raw data samples that were incorporated into the training model to determine specific feature parameters, Perform co-relational task.

3. Deep learning CNN:

   It module for extracting unique features of birds with the CNN and predicting the most classified labels for the input images. The model of CNN conjuration for bird identification utilized a stack of convolutional layers comprising of input layer, two FC layers and one final output layer.

4. Test Data:

   The test dataset will used to test the classifier parameters and access the performance of actual prediction of the network model. Once the features will be extracted from the raw data, the trained prediction model will be deployed to classify new input images.

5. Feature Extraction:

   Firstly, extracting features from raw input images is our primary task where extracting relevant and descriptive information for fined grained object recognition. However, because of semantic and intra class variance, feature extraction will be challenging. We are going to separately extract the features in relevant positions for each part of an image and subsequently learning the parts of the model features that were mapped directly to the corresponding parts.

6. Predictive model:

   The proposed model can predict the uploaded image of the bird as a bird. The proposed system will predict and differentiate various birds' images.

4. Software Implementation:

1. Python and Pytorch:

   Every once in a while, a python library is developed that has the potential of changing the landscape in the field of deep learning. PyTorch is one such library. Among the various deep learning libraries – PyTorch has been the most flexible and effortless of them all. This fits right into the python programming methodology, as we don't have to wait for the whole code to be written before getting to know if it works or not. We can easily run a part of the code and inspect it in real time. PyTorch is a python based library built to provide flexibility as a deep learning development platform.

2. Dynamic computation graphs:

   Instead of predefined graphs with specific functionalities, PyTorch provides a framework for us

   To build computational graphs as we go, and even change them during runtime. This is valuable for situations where we don't know how much memory is going to be required for creating a neural network.

3. CNN:

In deep learning, Convolutional neural network (CNN) is a class of deep neural network mostly used for analyzing 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. 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 volume considered are width, height and depth. The methodology proposed of achieving this is by using a convolutional neural network.
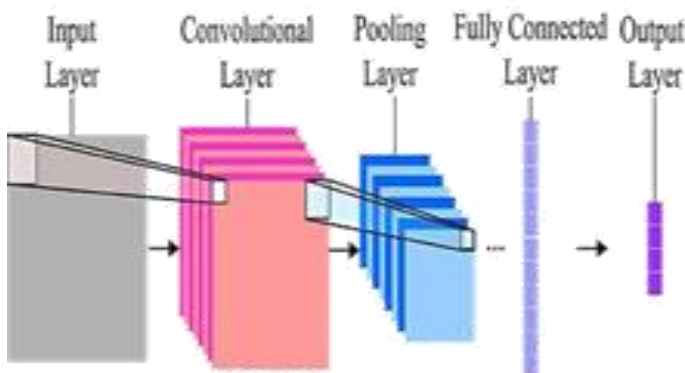

Basic architecture of CNN

The input layer takes image pixel inputs with three color channels R, G, B and passes it to convolutional layer. In CNN's the neurons are constructed in 3D dimensions which the depth means the activation volumes instead of 2D as in simple neural network. In convolutional layer, the main purpose of this layer is to extract image features and preserve the spatial connections between pixels from input. It computes the output of neurons which means it transform the image

Pixel values into output volume or final class scores. This convolution operation is done by striding filter on input image. Every image can be represented as a matrix of pixel values in CNNs, and the filter here in another name called kernel which is a matrix as well which is the concept from image processing. The operation is striding the filter on feature map each time by one pixel for every position and computing the

convolved output by adding up the output of multiplication of the matrixes. Each convolutional layer not only computes the final feature map but also there is activation function for taking input volume from previous layers and parameters of neuron such as bias and weights.
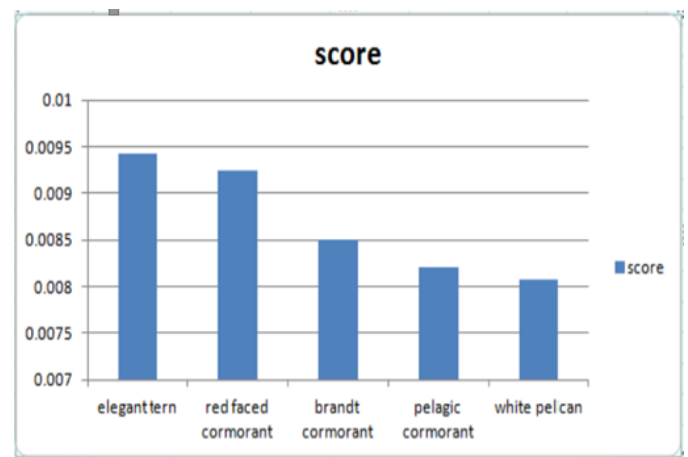
In the ReLU layer, feature map is passed to this layer to change all negative pixel value to zero. Most of images from real world are non-linear, but CNNs operation is a linear process. To make the CNNs learning the non-linear data, ReLU function is implemented to introduce the non-linearity to CNNs.
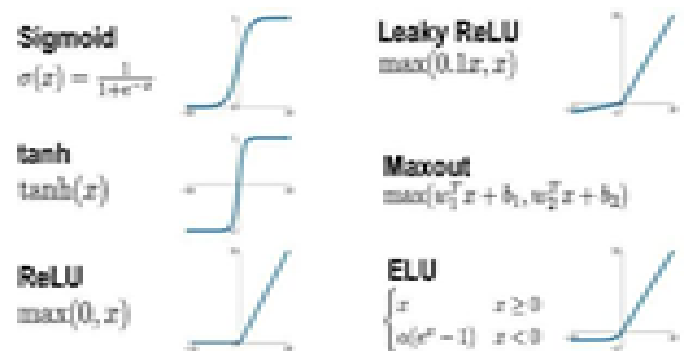
The pooling function reduces the size of feature map, but it still preserves the most important spatial details of the feature map. The purpose of pooling function is to reduce Parameters and computation in the network; thus, it controls over fitting issue. There are many types of pooling functions: max, sum, average etc. In this paper, max type of pooling will be used in pooling layer. In max pooling operation, you define a spatial neighborhood first like [2x2] and you stride it by 2 to the rectified feature map to select the largest pixel value in each area.
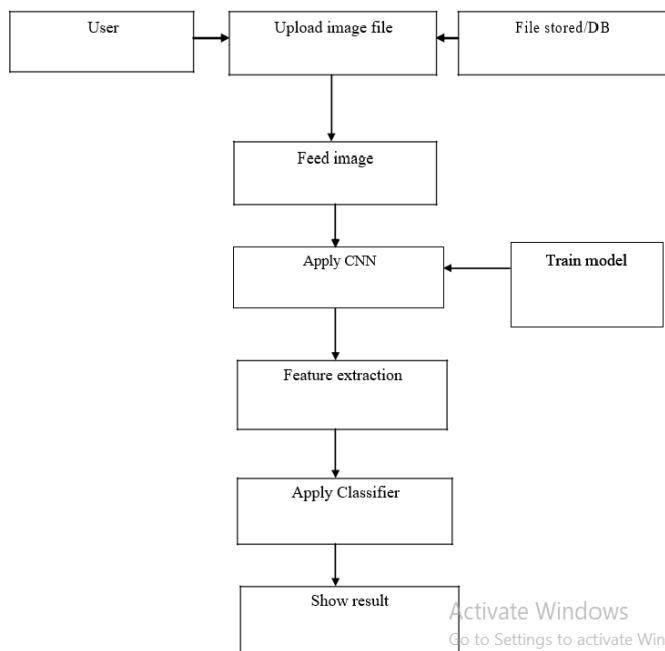
Flow chart for bird classification system

5. SIMULATIONS/RESULTS:

This is the sample photo taken from the internet. By using Deep learning convolutional neural network technique, the image gets scanned through various parameters as shown in figure.
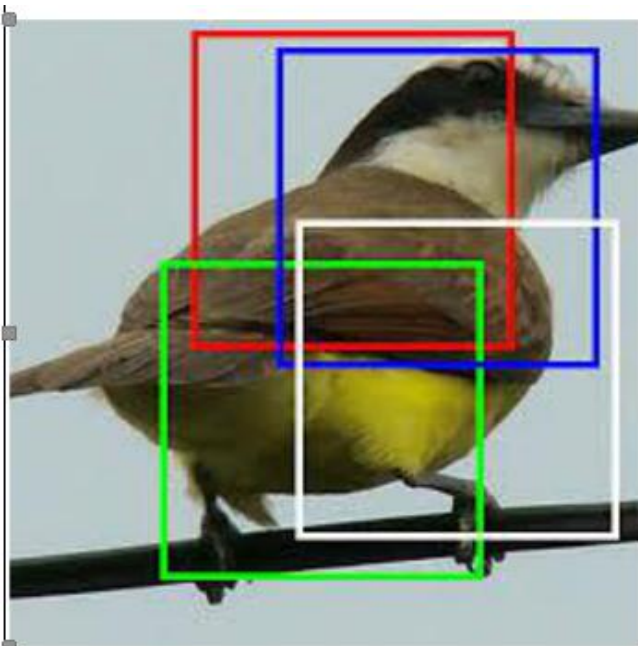

Results

After analyzing 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 elegant tern, red faced cormorant, Brandt cormorant and many more.

6. Conclusion:

The study of classification investigates a method to identify the bird species using deep learning algorithm on the dataset for classification of image. The system will be connected with a user friendly system where user will upload photo for identification purpose and it gives the desired output. The proposed system will works on the principle based on detection of a part and extracting CNN features from multiple convolutional layers. These features will be given to the classifier for classification purpose. On basis of the results the system will try to achieve maximum accuracy in prediction of bird species. The system will conduct a series of experiments in a dataset composed of several image to achieve maximum efficiency.

ACKNOWLEDGEMENT

REFERENCES:

[1] XIE, Z., A. SINGH, J. UANG, K. S. NARAYAN and P. ABBEEL. Multimodal blending for high-accuracy instance recognition. In: 2013 IEEE/RSJ International Conference on Intel-ligent Robots and Systems. Tokyo:

IEEE, 2013, pp. 2214–2221. ISBN 978-14673-6356-3.DOI: 10.1109/IROS.2013.6696666.

[2] EITEL, A., J. T. SPRINGENBERG, L. D. SPINELLO, M. RIEDMILLER and W. BUR-GARD. Multimodal Deep Learning for Ro-bust RGB-D Object Recognition. In: 2015IEEE/RSJ, International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE, 2015, pp.681–687. ISBN 978-1-4799-9994-1.DOI: 10.1109/IROS.2015.7353446.

[3] RUSSAKOVSKY, O., J. DENG, H. SU,J. KRAUSE, S. SATHEESH, S. MA, Z. HUANG,A. KARPATHY, A. KHOSLA, M. BERNSTEIN,A. C. BERG and L. FEI-FEI. Image Net Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV). 2015, vol. 115, no. 3, pp. 211–252. ISSN 1573-1405. DOI: 10.1007/s11263-015-0816-y.

[4 KRIZHEVSKY, A., I. SUTSKEVER and G. E. HINTON. Image Net classification with deep convolutional neural networks. Annual Conference on Neural Information Processing Systems (NIPS). Harrah's Lake Tahoe: Curran Associates, 2012, pp. 1097– 1105. ISBN 978-1627480031

[5] 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).