**Bird Species Identification from an Image**

Dr.K.Radhika, Manideep Chenna, Jagadeesh Reddy

Chaitanya Bharathi Institute of Technology

Hyderabad, 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, colors, and angles from the human perspective. Besides, the images present substantial variations to identify the bird species more than audio classification. Also, the 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.

Keywords: Deep Networks · Transfer Learning · Object Detection · Bird species classification · Flask

**Introduction:**

Identification of bird species is a challenging task often resulting in ambiguous labels. Even professional bird watchers sometimes disagree on the species given an image of a bird. It is a difficult problem that pushes the limits of the visual abilities for both humans and computers. Although different bird species share the same basic set of parts, different bird species can vary dramatically in shape and appearance. Intraclass variance is high due to variation in lighting and background and extreme variation in pose (e.g., flying birds,swimming birds, and perched birds that are partially occluded by branches).Our project aims to employ the power of machine learning to help amateur bird watchers identify bird species from the images they capture. Many people across countries are getting into this interest of bird-watching as a hobby or extracurricular activity. In modern world, it acts as a great stress buster and a cheap way of getting connected with nature. Another benefit of birdwatching is awareness about nature conservation by observing behavior, migratory pattern, population, and conservation status of bird species. From conservation point of view, it is important that more and more number of people turn towards bird-watching and help collect data that can be used to study birds. Sometimes, bird identification can be difficult for beginners as well.

Bird species are recognized as useful biodiversity indicators. They are responsive to changes in sensitive ecosystems, whilst populations-level changes in behavior are both visible and quantifiable. Suffered from great species variation, it is difficult for non-professionals to identify the sub-category of a bird only by its appearance. However, it is exhausting to annotate all the images by human beings with expert knowledge. Thus, an automatic classification system for bird species are needed, which will be great convenience for many practical applications. For researchers working outdoors, shoot photos can be classified and analyzed immediately by the system, illustration books are no more needed. For the public, the system could provide much fun when combined with culture information like poems and legends. It will arouse peoples interest in birds and could benefit the protections of birds. Apart from that, Classifying bird species is an interesting problem for Fine-grained categorization, also known as subcategory recognition, which is a subfield in object recognition. In recent years, fine-grained classification stood out from basic-level classification, bringing promising applications and new challenges to computer vision society.

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 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.

**Literature Survey:**

Title:Bird Species Identification using Deep Learning

Authors: Prof. Pralhad Gavali, Ms. Prachi Abhijeet Mhetre, Ms. Neha Chandrakhant Patil, Ms. Nikita Suresh Bamane, Ms. Harshal Dipak Buva

Year of Publication: 2019

Title: Audio hashing for bird species classification

Authors: Anshul Thakur , Pulkit Sharma , Vinayak Abrol , Padmanabhan Rajan

Year of Publication: 2019

Title: Bird and whale species identification using sound images.

Authors: Loris Nanni , Rafael L. Aguiar , Yandre M.G. Costa , Sheryl Brahnam ,Carlos N. Silla ,Ricky L. Brattin .

Year of Publication: 2019

**Methodology:**

CONVOLUTION NEURAL NETWORK: 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.

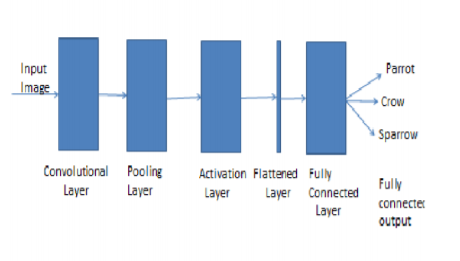


Fig 1: CNN Architecture

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.

Feature Extraction Features are detected when network performs a series of convolutional and pooling operation.

Convolutional Layer: Convolutional neural network layer types mainly include three types, namely Convo- lutional layer, pooling layer and fully-connected layer Convolutional layers apply a convolution operation to the input, passing the result to the next layer. The convolution emulates the response of an individual neuron to visual stimuli. Each convolutional neuron processes data only for its receptive field. Although fully connected feed forward neural networks can be used to learn features as well as classify data, it is not practical to apply this architecture to images. A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10000 weights for each neuron in the second layer. The convolution operation brings a solution to this problem as it reduces the number of free parameters, allowing the network to be deeper with fewer parameters. For instance, regardless of image size, tiling regions of size 5 x 5, each with the same shared weights, requires only 25 learnable parameters. In this way, it resolves the vanishing or exploding gradients problem in training traditional multi- layer neural networks with many layers by using back propagation. The aim of Convolutional layer is to learn feature representations of the inputs. As shown in above, Convolutional layer is consists of several feature maps. Each neuron of the same feature map is used to extract local characteristics of different positions in the former layer, but for single neurons, its extraction is local characteristics of same positions in former different feature map. In order to obtain a new feature, the input feature maps are first convolved with a learned kernel and then the results are passed into a nonlinear activation function.

Pooling Layers: Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters at one layer into a single neuron in the next layer.For example, max pooling uses the maximum value from each of a cluster of neurons at the prior layer Another example is average pooling, which uses the average value from each of a cluster of neurons at the prior layer. The sampling process is equivalent to fuzzy filtering. The pooling layer has the effect of the secondary feature extraction, it can reduce the dimensions of the feature maps and increase the robustness of feature extraction. It is usually placed between two Convolutional layers. The size of feature maps in pooling layer is determined according to the moving step of kernels. The typical pooling operations are average pooling and max pooling. We can extract the high level characteristics of inputs by stacking several Convolutional layer and pooling layer.

Fully connected Fully connected layers connect every neuron in one layer to every neuron in another layer. It isin principle the same as the traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images. In general, the classifier of Convolutional neural network is one or more fully-connected layers. They takeall neurons in the previous layer and connect them to every single neuron of current layer. There is no spatial information preserved in fully-connected layers. The last fully-connected layer is followed by an output layer. For classification tasks, softmax regression is commonly used because of it generating a well-performed probability distribution of the outputs. Another commonly used method is SVM, which can be combined with CNNs to solve different classification tasks.

Receptive field: In neural networks, each neuron receives input from some number of locations in the previous layer. In a fully connected layer, each neuron receives input from every element of the previous layer. In a convolutional layer, neurons receive input from only a restricted subarea of the previous layer. Typically the subarea is of a square shape (e.g., size 5 by 5). The input area of a neuron is called its receptive field. So, in a fully connected layer, the receptive field is the entire previous layer. In a convolutional layer, the receptive area is smaller than the entire previous layer.

Weights: Each neuron in a neural network computes an output value by applying some function to the input values coming from the receptive field in the previous layer. The function that is 21 applied to the input values is specified by a vector of weights and a bias (typically real numbers). Learning in a neural network progresses by making incremental adjustments to the biases and weights. The vector of weights and the bias are called a filter and represents some feature of the input. A distinguishing feature of CNNs is that many neurons share the same filter. This reduces memory footprint because a single bias and a single vector of weights is used across all receptive fields sharing that filter, rather than each receptive field having its own bias and vector of weights.

DATASET: A dataset is a collection of data. For performing action related to birds a dataset named CaltechUCSD Birds 200 (CUB-200-2011) is used. It is an extended version of the CUB-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 dataset is validated with an accuracy of 75% to increase the performance of system.

WEB TECHNOLOGIES: Flask: Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for objectrelational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Extensions are updated far more frequently than the core Flask program. Nowadays, the web frameworks provide routing technique so that user can remember the URLs. It is useful to access the web page directly without navigating from the Home page. It is done through the following route() decorator, to bind the URL to a function.

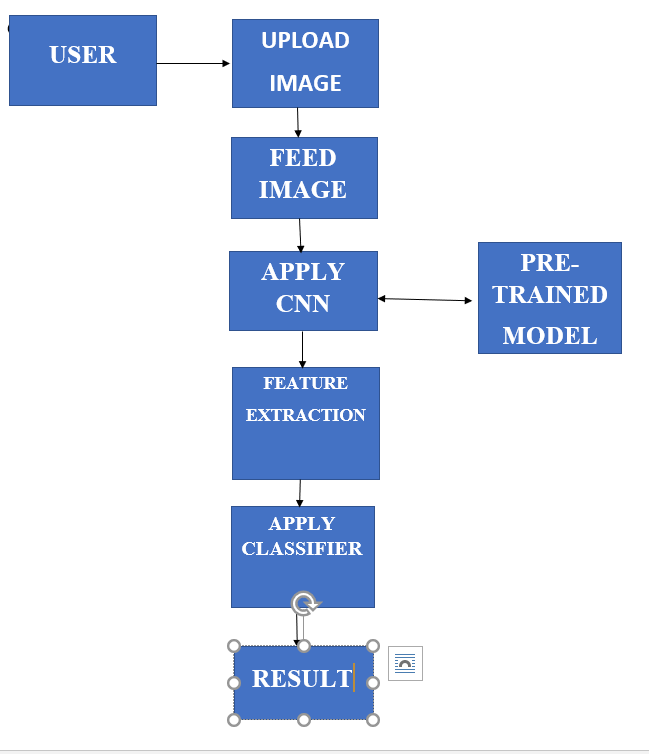
**Implementation:** 

Fig 2: Proposed Architecture

The above figure no.2 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.

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 Model is Deployed in web using Flask.

**Results:**

Model Results:

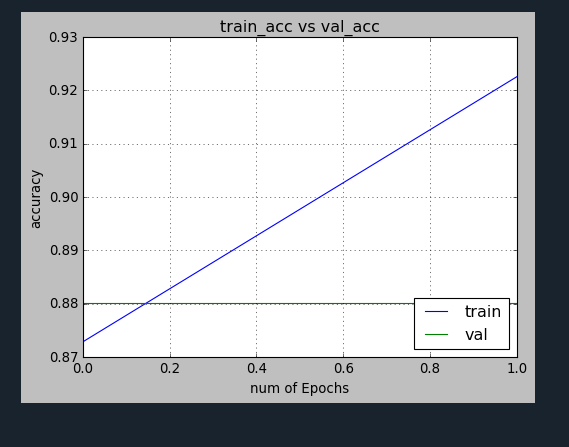
The Accuracy of Training dataset is 92.093% and Testing accuracy of 88%.

Fig 3:Training Accuracy vs Validation Accuracy

Application Results:

After saving the Model in .h5 file it is Deployed in web for Practical application of Model.

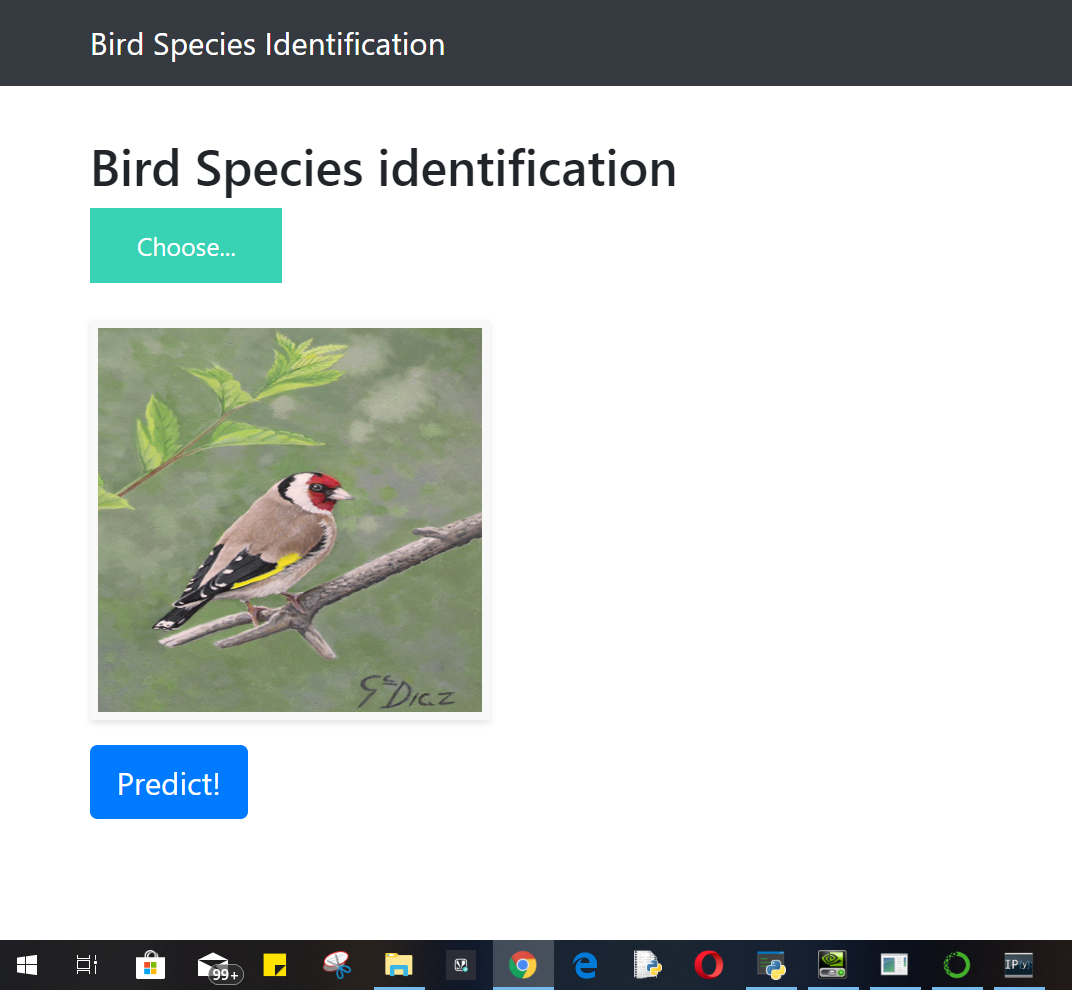


Fig 4: Model Deployed in Web.

**Conclusion and Future Scope:**

The generated system is connected with a user-friendly website where user will upload photo for identification purpose and it gives the desired output.

It Should be Implemented in cloud for fast computation and better results.

**References:**

[1] Tóth, B.P. and Czeba, B., 2016, September. Convolutional Neural Networks for LargeScale 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 Extremescale 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 networksfor 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 4

[6] Image Recognition with Deep Learning Techniques ANDREIPETRU BĂRAR, VICTOREMIL 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. 42

[9] Inception-v4,Inception-ResNetand the Impact of Residual Connectionson Learning Christian Szegedy, Sergey Ioffe,Vincent Vanhoucke ,Alexandrr A.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.