INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 04 | APRIL - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

Bird Species Identification Using CNN

Dr. D Siri
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
Siri1686@grietcollege.com

J. Snehith Reddy
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
snehith700@gmail.com

G. Maneesh
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
maneeshhanddilip@gmail.com

V. Ajay Kumar Reddy
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
ajayvelamuri777@gmail.com

G. Jagadeshwara Sai
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
jaggunisetty@gmail.com

P. Vijay Kumar
Computer Science and Engineering Department
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
Vijayparagati123@gmail.com

***Abstract*— In our world, there are above 11000 bird species. Some bird species are being found rarely. Bird identification is a challenging task that usually leads to unclear labelling. When presented a picture of a bird, even professional bird watchers differ on the species. Despite having the same basic components across all bird species, form and appearance can vary greatly. Intraclass variance is substantial due to variations in lighting and backdrop, as well as a wide range of instances. Additionally, visual recognition of birds by humans is more comprehensible than audible recognition of birds. Consequently, we utilised convolutional neural networks (CNN). CNNs are a powerful Deep Learning ensemble that have shown to be effective in image processing. The dataset is used for both training and testing of a CNN system that classifies bird species. Everyone can quickly determine the name of the bird they wish to know by following this strategy.**

Keywords— Bird species, Image Processing, Convolutional Neural Networks.

I. INTRODUCTION

Many individuals like learning about the environment and viewing wildlife, especially birds. By watching bird behavior and migratory patterns, birdwatchers aid in the preservation of biodiversity. Because of the close proximity of the bird features and the background of the photos, as well as the typical inexperience of observers, it is still challenging for bird watchers to identify birds from photographs.. Humans are unable to recognise and categorise the many kinds of birds because of their unique characteristics, including their colour, size, and viewing angle. Bird species identification a difficult undertaking the frequently gives in label that are unclear. Perhaps skilled bird watchers sometimes disagree when given a picture of a bird's species. It is a difficult problem that puts a load on both the visual capacities of humans and computers. Diverse bird species can have dramatically different forms and looks yet sharing the same essential ingredients. Therefore, computer-based photos are required to aid birdwatchers in identifying species. In this study, convolutional neural networks are used to extract information from photos in order to investigate the utility of deep learning for bird identification.

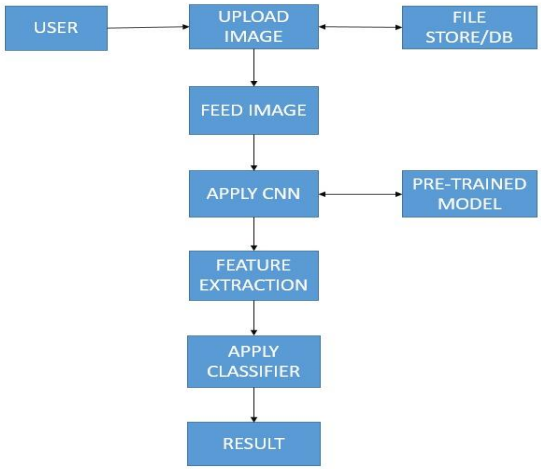
II. LITERATURE REVIEW

Survey Paper-1
Authors: Rafael L. Aguiar , Yandre M.G. Costa, Loris Nanni
Title: Bird and whale species identification using sound images
Year of Publication: 2019
Approach: The authors of this study offer a novel and effective technique for automating the identification of birds and whales by utilizing some of the most powerful texture descriptors currently available in the computer vision field. Starting with the audio file, pictures made from various spectrograms as well as from harmonic and percussion images are used to build the visual features of sounds. These studies employed datasets of bird voices for species classification and a dataset of right whale sounds for whale detection.

Survey Paper-2
Authors: Padmanabhan Rajan, Anshul Thakur,Vinayak Abrol
Title**:** Audio hashing for bird species classification
Year of Publication**:** 2018
Approach: For the purpose of classifying different bird species, convex representation-based audio hashing method is used. Convex-sparse models of a bird vocal communication are created using the recommended method, which combines matrix factorization with archetypal analysis. The convex models are scrambled using Bloom filters and noncryptographic hash algorithms to produce fully connected compact binary codes. The class-specific k-medoids clustering technique is used to group the retrieved conv-codes from of the training samples, with the Jaccard coefficient serving as the similarity measure. The hash functions and slots in a hash table are utilized as pointers towards the species identification information, while the cluster centers serve as the table's keys.

Survey Paper-3
Authors: Pralhad Gavali, Ms. Neha Chandrakhant Patil, Ms. Prachi Abhijeet Mhetre
Title: Bird Species Identification using Deep Learning
Year of Publication: 2017
Approach: The Caltech-UCSD Birds 200 [CUB-200-2011] dataset is utilized by this approach for the testing and training. A photo is converted into a greyscale image using the tensor flow approach and convolutional neural network (CNN) algorithm to produce an autograph. A score sheet is created as a consequence of comparing the testing dataset to these different nodes. The needed bird species may be predicted using the highest score once the score sheet has been analyzed.

III. METHODOLOGY



Upload Image:

A user uploads an image. The image is then stored in a temporary database.

Feed Image:

The image is then sent to CNN, which is used in conjunction with a training dataset. For classifying to be as accurate as possible, a variety of alignments and features are taken into account, including the bird's head, body, colour, beak, form, and overall appearance. Each alignment is delivered through a deep convolutional network in order to extract features from the network's various levels. The classification of the image is then performed using CNN and an unsupervised process known as deep learning.

Apply CNN:

Convolutional layer, pooling layer, and fully linked layer are the three layers that make up CNN. Small-scale visual feature extraction from an image is possible using convolutional layers. Pooling is a technique used to save the crucial data while reducing the number of neurons from the preceding convolutional layer. A function that condenses values into a range is applied by the activation layer to values. A layer that is totally connected connects every neurone in one layer to every neurone in every other layer. CNN provides more precision due to its detailed categorisation of each neurone. There are two parts to CNN:

A. Feature Extraction
B. Classification

Feature Extraction:
Features are found whenever the network performs a sequence of convolutional and pooling procedures.

A. Convolutional layer:

Convolutional layer, pooling layer, and fully-connected layer are the three major types of convolutional neural network layers. Before the results are passed on to the next layer, convolutional layers perform an action on the input. Convolution resembles the response of a single neuron to sensory input. Fully linked feed forward neurall networks may categorise input and learn features, it is impractical to employ this design for image processing. Due of the relatively large input sizes associated with images, where each pixel is a major variable, a shallow design, which is the reverse of deep architecture, would still require a very high number of neurones. As an illustration, the second layer of a completelly connected layer includes 10,000 weights for each neurone in a (small) picture with a dimension of 100 by 100. This issue is resolved by convolution, which lowers the number of free parameters and enables the network to be deeper with fewer parameters. For example, tiling parts of size 5 × 5 with the same shared weights only requires 25 learnable parameters, regardless of the size of the image. By employing back propagation, it is possible to overcome the issue of disappearing or exploding gradients while training conventional multi-layer neural networks. Several feature maps make up the convolutional layer. The unique characteristics of multiple places in the preceding layer are extracted using each neurone of the same feature map; however, for a single neurone, the local characteristics of similar sites in the preceding separate feature map are extracted. The input feature maps are first convolved with a learnt kernel to form a new feature, and the outcome is then fed into a nonlinear activation function.

B. Pooling layer:

Convolutional networks' local or global pooling layers combine the outputs of neurone clusters in one layer into a single neurone in the subsequent layer. The greatest value from each set of neurones in the layer before is used, for instance, in max pooling. Another example is average pooling, which uses the average value from each group of neurones in the layer above. A fuzzy filtering approach analogue is the sampling process. It also has the impact of secondary feature extraction. The typical placement is in the middle of two convolutional layers.

C. Fully Connected layer:

Through completely connected layers, every neuron in one layer may communicate with every neuron in every other layer. It is equivalent to the standard multi-layer perceptron neural network in principle (MLP). The flattened matrix runs through a fully connected layer to categorize the images. Fully linked layers do not retain any spatial information. An output layer follows the last layer that is completely linked. SoftMax regression is frequently used for classification applications because it produces results with a good probability distribution.

Pre-processing:

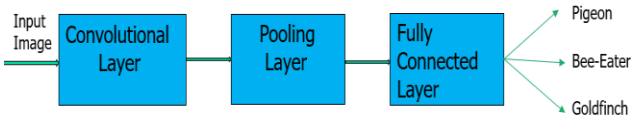
Grayscale indicates that just the information about the light's intensity is represented by each pixel's value. Typically, just the contrast between the sharpest black and the clearest white is evident in such images. In other words, the image simply consists of three colours : black, white, and grey, the latter of which comes in a variety of shades. Grayscale is a spectrum of apparent-colourless hues of grey. The image is categorised pixel by pixel using the grey scale approach. After that, the characteristics are combined and sent to the classifier. A classifier is a particular kind of deep learning algorithm used to categorise data input. Labelled data is used to train classifier algorithms; in the case of image recognition, for instance, the classifier is given training data that includes labels for the pictures. When the classifier has had enough training, it may take in unlabelled images and output classification labels for each image. Classifier algorithms employ sophisticated mathematical and statistical methods to forecast the likelihood that a data input will be categorised in a specific way. An image's likelihood to be a parrot, woodpecker, or another categorisation that the classifier has been trained to recognise is predicted statistically by the classifier.

Classification:

Putting anything or someone into a certain group or system based on specific criteria is the definition of classifying. Google's open-source machine learning framework for dataflow programming is as TensorFlow, and it may be used for a variety of purposes. In this project TensorFlow is used for classification. An autograph is created during classification, consisting of a series of nodes that eventually forms a network, and the dataset is retrained to improve recognition accuracy.

Output:

To provide potential outcomes, the input will be checked to the training dataset. The best output will be created with the aid of the score sheet, which is prepared based on this network.



IV. FRONT END

A. *HTML*

The preferred markup language for building Web pages is HTML. A Web page's structure is described in HTML. The fundamental idea behind HTML is to add formatting instructions to some organized material, which is typically a combination of text and graphics. This formatting data is used by the web browser to properly handle the content. HTML 4.01 was replaced by HTML5, which added new features and capabilities to the language while also enhancing or deleting some of the functionality that was previously there. Programming languages like JavaScript and technologies like Cascading Style Sheets (CSS) can be useful.

B. *CSS*

CSS is a language used to create style sheets that specify how markup-based documents should be organized and appear. It gives HTML extra capabilities. It often works in conjunction with HTML to modify the style and appearance of web pages and user interfaces. Any type of XML document, including plain XML, SVG, and XUL, may be used with it. The majority of websites employ CSS, HTML, and JavaScript to build the user interfaces for numerous mobile applications as well as online apps.

C. *JavaScript*

Live Script is where JavaScript got its start. The goal was to find a language that was less complex than Java and could be used to create client-side, in-browser applications. The language is most frequently used to modify the components of the document object model and works best for short-term activities. JavaScript is an accessible programming language with first-class capabilities (JS). You may interpret it or have it just-in-time compiled. Node.js, Apache CouchDB, and Adobe Acrobat are just a few of the applications that employ JavaScript, despite it being best known as the scripting language for Web pages.

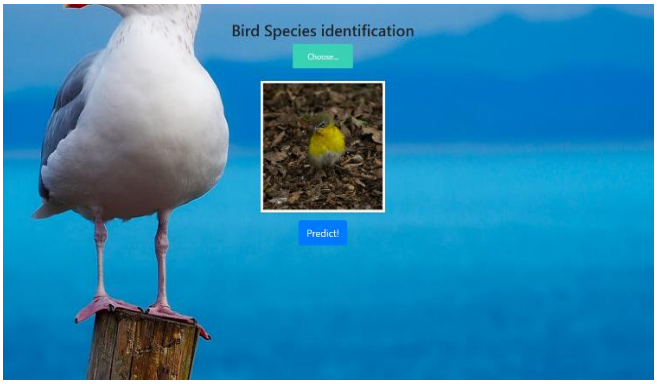
Backend

Python:

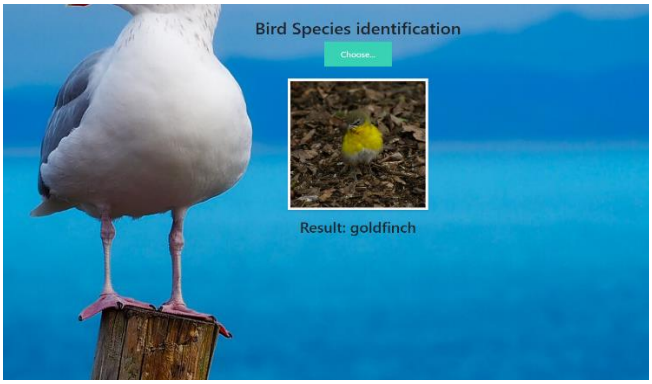
Python is a high-level, dynamically semantic, object-oriented, interpreted programming language. Due to its high-level built-in data structures, dynamic typing, and dynamic binding, it is well suited for use in Rapid Application Development as well as for utilisation as a scripting or glue language to connect existing components. Python's syntax reduces the expense of maintaining programmes. Programming code modularity and reuse are encouraged by Python's support for modules and packages.

V. RESULTS

This is the webpage to choose the image of the bird.



After click on predict button, the system will predict the name of the bird.



VI. CONCLUSION

The website's major goal is to raise public awareness of birdwatching, identification, and in particular, the identification of birds found in India. Additionally, it complies with the demand for streamlining the process of bird identification, which enhances bird viewing. Convolutional Neural Networks are the technology utilised in the test scenario (CNN). Utilizing feature extraction, it finds photos. To extract characteristics and categorise photos, the approach utilised is adequate. The main goal of the project is to classify the various bird species from a photograph that the user submits. We chose CNN because it is appropriate for using complex algorithms and provides strong numerical precise accuracy. The method predicted the discovery of bird species with an accuracy of 80% based on the findings that were generated.

VII. FUTURE SCOPE

1. A system can predict the bird by taking bird fossils as an input.
2. Create an app rather than a website for consumer convenience.
3. A system may be put into operation using the cloud, which can store a lot of information for comparison and offer powerful processing.

REFERENCES

[1] Chandu, Akash Munikoti, Chaitra Nagaraj, Ganesh Murthy," Birds Species Identification using Audio.", IEEE Journal,2020.

[2] Mario Lasseck," Audio-based Birds Species Identification with Deep CNN", ResearchGate Journal,2018.

[3] Anisha Singh, Akarshita Jain, Bipin Kumar Rai," Image-based Birds Species Identification", IJREISS Journal,2020.

[4] T. S. Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conservation International, 2008.

[5] Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2007

[6] Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Deep learning. Nature Methods, 2018

[7] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. Audio BasedBird Species Identification using Deep Learning Techniques. 2016.

[8] Mario Lasseck," Audio-based Birds Species Identification with Deep CNN", ResearchGate Journal,2018.

[9] Anisha Singh, Akarshita Jain, Bipin Kumar Rai," Image-based Birds Species Identification", IJREISS Journal,2020.

[10] T. S. Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conservation International, 2008.

[11] Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2007

[12] Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Deep learning. Nature Methods, 2018

[13] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. Audio BasedBird Species Identification using Deep Learning Techniques. 2016.

[14] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks."

[15] Welinder, Peter et al. 2010. "Caltech-UCSD Birds 200." Technical ReportCNS-TR-2010-001, California Institute of Technology, 2010.