Automated System for Bird Species Identification Using CNN

Dharmapuri Siri[1*], Sashank *Desu*[1] ,Karthik *Alladi*[1] ,*Duggi* Swathi[1], Swati *Singh*[2] ,*Vellanki* Srilakshmi[1]

[1]Department of Computer Science, Gokaraju Rangaraju Institute of Engineering and Technology, , Hyderabad, Telangana, India
[2]Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, 248007, India

Abstract. There are around 11000 different bird species in the globe. Rarely are certain bird species encountered. 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, the convolutional neural networks(CNN) is utilized for an automated bird species identification system. 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. This will lead to quick identification of the bird species using an automated process.

1 . Introduction

Many individuals like learning about the environment and viewing wildlife, especially birds. By watching bird behavior and migratory patterns, bird-watchers 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 recognize and categorize the many kinds of birds because of their unique characteristics, including their color, size, and viewing angle. Bird species identification a difficultundertaking the frequently gives in label that are unclear. Perhaps skilled bird watchers sometimes disagree when givena picture of a bird's species. It is a difficult problem that putsa 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 bird-watchers in identifying species. In this study, convolutional neural networks[1-7] are used to

* Corresponding author: siri1686@grietcollege.com

extract informationfrom photos in order to investigate the utility of deep learning for bird identification.

Fig. 1. Images of different species of birds

2 . Literature Review

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. Thesestudies employed datasets of bird voices for species classification and a dataset of right whale sounds for whale detection [1].

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 [2].

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 [3].

3 Methodology

A user uploads an image.

3.1 Upload Image

The image is then stored in a temporary database

3.2 Feed Image

The images are then sent to CNN and combined with the training data. In order to ascertain the truth as much as possible, many arrangements and features of the bird such as its head, body, colour, beak, shape and general must be taken into account. Deep convolutional networks are used to combine each element to remove features from multiple layers of the network. The images are then classified using CNN and an unsupervised technique called deep learning.

3.3 Pre-processing

Grayscale indicates that just the information about the light's intensity is represented by each pixel's value. In such snapshots, just the contrast between the sharpest black and the clearest white is seen. In other words, the image is made up of three colors: black, white, and grey, which comes in a range of tints. Grayscale is a range of apparent-colorless grey colors. Using the grey scale method, the image is classified pixel by pixel. The characteristics are then aggregated and delivered to the classifier. A classifier is a particular kind of deep learning algorithm used to categorize 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 unlabeled images and output classification labels for each image. Classifier algorithms [8-9] employ sophisticated mathematical and statistical methods to forecast the likelihood that a data input will be categorized in a specific way. An image's likelihood to be a parrot, woodpecker, or another categorization that the classifier has been trained to recognize is predicted statistically by the classifier.

3.4 Apply CNN

3.4.1 Convolutional layers, pooling layers, and all layers are the three layers that make up a CNN[10-12]. You can extract auxiliary features from images using convolutional methods. Pooling is a technique used in preprocessing to reduce the number of neurons while saving important data. Layers allow values to compress values into multiple values. A reference layer connects all neurons in one layer to all neurons in another layer. CNNs provide higher accuracy due to the detailed information about each neuron. CNN consists of two parts: extraction and classification.

3.4.2 Feature Extraction

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

3.4.2.1 Convolutional layer

Convolutional layers, pooling layers, and fully connected layers are the three main types of convolutional neural network layers. Convolutional layers perform operations on their inputs before passing the result to the next layer. Convolutions resemble individual neurons' responses to sensory input. Fully connected feedforward neural networks can classify inputs and learn features, but this scheme cannot be used for image processing. Shallow designs, as opposed to deep structures, still require a very large number of neurons due to the relatively large input data associated with the image where each pixel is a large variable. For example, the second layer of the fully connected layer contains 10,000 weights for each neuron in a (small) 100 x 100 image. problem. Image size is independent of image size, and for 5 * 5 tiles with equal total weights, only 25 training parameters are required. Backpropagation solves the problem of vanishing or exploding gradients when training traditional multilayer neural networks. Multiple feature maps make up a convolutional layer. Unique features from different locations in the previous layer are extracted using each neuron in the same feature map. However, for one neuron, local features of similar regions are extracted from the previous separate feature maps. To create a new feature, we first convolved the input feature map with the learned kernel and then fed the result into a nonlinear activation function.

3.4.2.2 Pooling Layer

Local or global pooling layers [13] of convolutional networks integrate the output of groups of neurons in one layer into individual neurons in the next layer. For example, the best value for each group of neurons in the previous layer is used for maximum pooling. Another example is integration, which uses the combined results of each neuron in the upper layer. An example of the fuzzy filtering method is the sampling method. It also has a secondary elimination effect. The process is placed between two convolutional layers.

3.4.2.3 Fully Connected Layer

Absolutely connected layers allow every neuron in one layer to further communicate with every neuron in every other layer. This is basically the same as the old Multilayer Perceptron Neural Communities (MLP). The smoothed matrix classifies the photo by passing it through a fully connected layer. Fully connected layers do not retain spatial information. The output layer follows the last fully connected layer. SoftMax regression [15] is regularly used in type packages because it produces results with surprising probability distributions.

3.4.3 Classification

Putting anything or someone right into a certain organization or system primarily based on unique standards is the definition of classifying. Google's open-supply device mastering framework for dataflow programming is as TensorFlow, and it could be used for ramification of purposes. An autograph is created, which includes a chain of nodes that eventually forms a community, and the dataset is retrained to improve popularity accuracy.

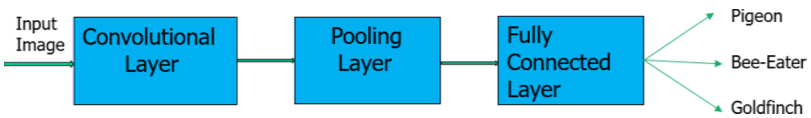

Fig. 2. Illustration of the component layers in the deep learning pipeline

3.5 Front-End

3.5.1 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.

3.5.2 CSS

CSS is a language used to create style sheets that explain how markup-primarily based documents must be organized and appear. It offers HTML extra skills. It frequently works at the side of HTML to trade the layout and visible attraction of web pages and consumer interfaces. It can be used with any XML record type, consisting of undeniable XML, SVG, and XUL. Most websites utilize HTML, CSS, and JavaScript to create the user interfaces for a variety of mobile programs in addition to on-line apps.

3.5.3 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.

3.6 Back-End

3.6.1 Python

Python [16] is a dynamically semantic, excessive-level, interpreted programming language. It's far best for usage in fast application development in addition to a scripting or glue language to link existing additives because of its excessive-degree built-in statistics structures, dynamic typing, and dynamic binding. The cost of retaining a code is decreased by using Python syntax. Modularity and code reuse are supported by Python's support for modules and packages.

4 . Results

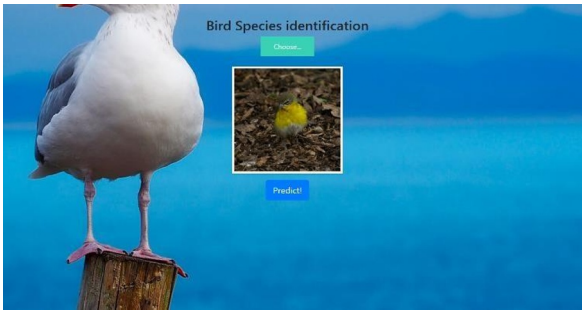

Fig. 3. Homepage of the web application

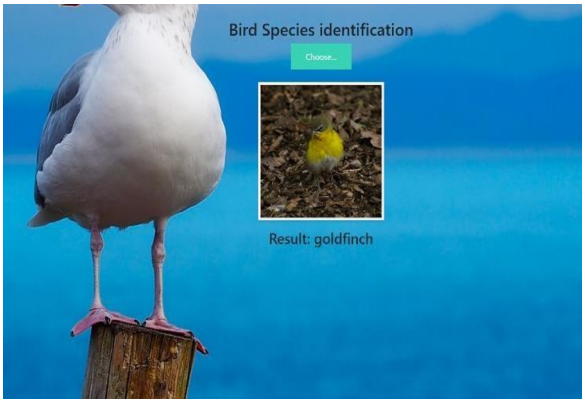

Fig. 4. Homepage of the web application after the prediction results

5 . Conclusion

The main purpose of this website is to raise public awareness of bird watching, bird identification and especially the identification of birds native to India. It also enhances observation by meeting new identification process optimization requirements[17-20]. Convolutional Neural Networks are a technique used in test scenarios (CNNs). Using feature extraction, he finds the photo. Appropriate approaches are used to extract pictures of characteristics and categories. The main goal of this project is to classify numerous bird species based on user provided photos. We choose CNNs because they are suitable for complex algorithms and provide a high level of numerical accuracy. Based on the results obtained, the approach showed 80% accuracy in predicting the discovery of new bird species. The automated bird species identification system will lead to quick identification of the bird species.

6 . Future Scope

A system can predict the bird by taking bird fossils as an input. Create an app rather than a website for consumer convenience. The cloud, which offers strong processing and can store a lot of data for comparison, may be used to bring a system into action.

References

1. Gundavarapu, M.R., Saginala, R., Varma, M.A., ...Bodduluri, A.S., *Deep Learning Framework for Liver CT Image Segmentation and Risk Prediction,* LNNS,189–201, (2023)**.**
2. Dusa, D., Gundavarapu, M.R.8th *International Conference on Advanced Computing and Communication Systems*, ICACCS 2022, 1023–1028, (2022).
3. Tejaswini Priyanka, Y. Reshma Reddy, D. Vajja, G. Ramesh and S. Gomathy (2023). *A Novel Emotion based Music Recommendation System using CNN. 2023 7th International Conference on Intelligent Computing and Control Systems* (ICICCS), Madurai, India, 592-596, doi: 10.1109/ICICCS56967.2023.10142330.
4. Ramesh, G., Reddy, K.S.S., Ramu, G., Reddy, Y.C.A.P., Somasekar, J. (2023). *An Empirical Study on Discovering Software Bugs Using Machine Learning Techniques.*
5. In: Buyya, R., Hernandez, S.M., Kovvur, R.M.R., Sarma, T.H. (eds) *Computational Intelligence and Data Analytics. Lecture Notes on Data Engineering and Communications Technologies,* vol **142**. Springer, Singapore. https://doi.org/10.1007/978-981-19-3391-2_14.
6. Chandrika Lingala, and Karanam Madhavi, "*A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms ",* E3S Web of Conferences 309, 01043 (2021), ICMED 2021. https://doi.org/10.1051/) e3sconf/202130901043 SCOPUS.
7. Chandra Sekhar Reddy P , Sakthidharan G, Kanimozhi Suguna S, Mannar Mannan J, Varaprasada Rao P, IJEAT. **8,** (2019).
8. Chandu, Akash Munikoti, Chaitra Nagaraj, Ganesh Murthy, *International Conference on Artificial Intelligence and Signal Processing*,(2020).
9. Anisha Singh, Akarshita Jain, Bipin Kumar Rai, IJREISS Journal,(2020).
10. T. S. Brandes, "*Automated sound recording and analysis techniques for bird surveys and conservation",* Bird Conservation International Conference, (2008).
11. Fagerlund S., EURASIP Journal on Apl Sgl Prcs, (2007).
12. Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Nature Methods, (2018).
13. Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. EPFL. (2016).
14. Anisha Singh, Akarshita Jain, Bipin Kumar Rai, IJREISS Journal,(2020).
15. T. S. Brandes, "*Automated sound recording and analysis techniques for bird surveys and conservation",* Bird Conservation International conference, (2008).
16. Fagerlund S. EURASIP Journal on Apld Sgnl Pro (2007).
17. Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Nature Methods. (2018).
18. Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. EPFL. (2016).
19. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. ACM. (2012).
20. Welinder, Peter et al. 2010. Technical Report CNS-TR-2010-001, California Institute of Technology, (2010).

21. ,Rafael L. Aguiar, Yandre M.G. Costa, Loris Nanni,"*In Bird and whale species identification using sound images",* Published in (2019)
22. Padmanabhan Rajan, Anshul Thakur, Vinayak Abrol, "*In Audio hashing for bird species classification* ", Published in (2018)
23. Pralhad Gavali, Ms. Neha Chandrakhant Patil, Ms.Prachi Abhijeet Mhetre*," Bird Species Identification using Deep Learning"*,Year of Published in (2017)