Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

CLASSIFICATION OF BIRD IMAGES BY CNN

D. Vijay kumar[1], J. Harish[2], D. Gowtham[3], A. Upendra[4].
[1, 2, 3, 4] Department of Computer Science and Engineering at Raghu Engineering College, Visakhapatnam, India

Mr. P P N G Phani kumar
Assistant Professor
Department of Computer Science and Engineering at Raghu Engineering College, Visakhapatnam, India

Abstract — **People juggle numerous responsibilities and activities in today's fast-paced world. Bird watching has become a well-liked hobby among those looking for relaxation. To appreciate the beauty of numerous bird species, many lovers visit bird sanctuaries. We created a Deep Learning model that can detect 60 different bird species to help birders identify birds in their natural habitat. To analyse bird picture data, our model uses the Convolutional Neural Network (CNN) technique. We used the Microsoft Bing Image Search API version 7 to create our dataset and randomly divided the data into an 80:20 ratio. The CNN model's accuracy rate throughout the training phase was 93.19%. The model's accuracy rate during testing was 84.91%. On a Windows 10 computer, our complete experimental research was carried out.**

Deep Learning, CNN Model, Classification and Prediction, TensorFlow, and Keras are some related terms.

I. INTRODUCTION

A branch of ML, which is a branch of AI is Deep Learning. In essence, neural networks and algorithms inspired by the human brain can learn from vast volumes of data using a framework called "deep learning." Computers can solve complex issues using this method even when the input is varied, unstructured, and related. Deep Learning algorithms' performance greatly increases as they gain new knowledge.

In the current era, it has become difficult and confusing to identify different species of birds. Due to their remarkable sensitivity to environmental changes, birds can assist us in understanding the ecosystem. However, it takes a lot of human labour to gather and analyse bird data. While many people go to bird sanctuaries to watch birds, they frequently have trouble telling different species apart and their distinctive traits.

We can learn more about birds, their surroundings, and their biodiversity by being aware of these distinctions. Due to observer limitations like location, distance, and equipment, it can be difficult to precisely classify birds based on their unique traits. This makes it difficult to identify birds with the human eye. Even ornithologists struggle with bird species identification since it necessitates a thorough comprehension of their.

Fig 1: A Venn diagram identifying deep learning as a branch of artificial intelligence that is really a branch of machine learning.

Images, audio, and/or video can all be used for bird identification. Birds can be found using audio and video processing algorithms that analyse the auditory and visual information. Processing this information, however, may be challenging because of the mingled noises, such as insects, and the presence of other things in the frame. People often do better at identifying birds from photos than from audio or video, though.Therefore, compared to speech or video, utilising photographs to identify birds is frequently a simpler method.

Using the Convolutional Neural Network (CNN) method, we created an interface to forecast birds in their natural environments. The procedure entailed gathering and localising a sizable dataset of bird species, creating a CNN architecture akin to the VGG Net Network, then training the model using the bird dataset. To identify a target object, the trained and categorised data was then placed on disc. An end-user uploads a sample bird image to the client-server architecture, which retrieves data from the qualified model stored on the disc to identify the species of bird from the image. With the use of this technique, birds can be autonomously identified from collected photos, revealing important information about different bird species.

II. RELATED WORKS

A. *Bird Species Identification Using Support Vector Machines [1]*

The primary focus of this study is the automated identification of bird species using their vocalisations. The recognition procedure makes use of a decision tree with a Support Vector Machine (SVM) classifier that can discriminate between two species at each node. The effectiveness of the recommended approach was evaluated using two collections of bird species that had previously been studied using diverse methods. The outcomes demonstrated that the suggested strategy produced better or equal results in comparison to the current reference models.

B. Color-Based Classification of Bird Species [2]

Extraction of colour features from unrestricted photos is the foundation of the suggested method for classifying bird species. By eliminating the background components, a colour segmentation approach is utilised to pinpoint the possible locations for the bird. Standardised colour histograms are generated for each plane from these potential sites once the picture has been divided into component planes. The aggregation process is then used to reduce the quantity of histogram intervals to a certain amount of bins. A learning system uses these histogram bins as feature vectors to distinguish between various bird species.

C. Deep Learning Techniques for Image Recognition [3]

Deep learning neural networks like CNN and Deep Belief Network (DBN) were specifically utilised in this research study to recognise pictures. These models were developed and evaluated using the Caltech101 dataset. The Caltech101 database issuer used the SVM-KNN approach as the benchmark model in order to contrast their performance.. Before putting the suggested strategy into practise, many dataset preparation approaches were used, resulting in a 67.23% accurate recognition score. This result represented a 1% increase over the benchmark algorithm's recognition score.

D. Identification of bird species using an SVM classifier and decision tree [4]

In this study, a novel method for classifying various bird species was put forth. It involved comparing the ratio of the eye's distance from the beak's root to the beak's breadth. The suggested method used these additional traits with an SVM and multi-scale decision tree structure to produce the desired recognition outcome. The suggested method produced a remarkable correct classification rate of about 84% in the studies, which used a sizable collection of bird photos. This study proved that the suggested traits can offer helpful information for identifying different bird species and that adding these elements to the identification system can greatly increase recognition accuracy.

E. Deep Convolutional Nets with Pose Normalised Categorization of Bird Species [5]

In this work, the use of architecture for fine-grained visual classification was suggested. The method was successful in categorising bird species according to human experts. In order to identify the qualities of neighbouring photographs, the architecture first makes an estimate of the object's posture. These qualities are also employed for categorisation. Deep convolutional networks are used to analyse areas of the picture that the pose locates and normalises in order to discover the features.

A thorough analysis of feature learning optimisation for fine-grained categorization and contemporary deep convolution technology implementations was conducted. The trials enhance the state-of-the-art success in classifying bird species, with proper categorization rates significantly higher (75% vs. 55-65%) than with the previous methods.

F. Bird identification using image recognition [6]

The major objective of this study was to identify several bird species using a user-submitted photograph of a bird. Transfer learning is a technique for optimising an AlexNet-pretrained model. SVMs (Support Vector Machines), a kind of supervised machine learning, are used for classification.

Because it can carry out complex computations and has outstanding precision in numerical precision, MATLAB was chosen. Currently, the developer's accuracy ranges from 80% to 85%. This initiative has a broad scope and succeeds in its objective. In order to monitor and research wildlife, camera traps can be deployed to record animal habits and movements in specific locations.

G. Automatic Flying Bird Species Classification Using Computer Vision Technique [7]

The goal of this research was to create an accurate, automated method for recognising certain bird species using video data captured while the birds were in flight. This article provided a brand-new, comprehensive set of appearance characteristics for video categorization. They incorporated motion characteristics like wing beat frequency and curvature. The dataset was made up of seven species. The experimental evaluations of the appearance and motion characteristics were combined with the Regular Bayes and Support Vector Machine classifiers. The Normal Bayes and SVM classifiers each achieved a classification rate of 92 and 89 percent, respectively.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

III. PROPOSED MODEL FOR DEEP LEARNING.

By assigning various components of the input image different weights and distinctions, Deep learning algorithms known as convolutional neural networks (CNNs) are capable of differentiating between distinct images. CNN needs a lot less pre-processing than other classification algorithms do. Filters were normally hand-engineered in conventional ways, but when given enough training cases, CNN can learn these filters on its own. CNN's design was influenced by the way neurons are connected in the human brain, where each neuron only responds to inputs that fall within its receptive field. These receptive fields encompass the whole visual field when taken as a whole. Understanding the essential elements that control how Convolutional Neural Networks function is crucial.

- CNN; input image
- Output Label (Class of Images)

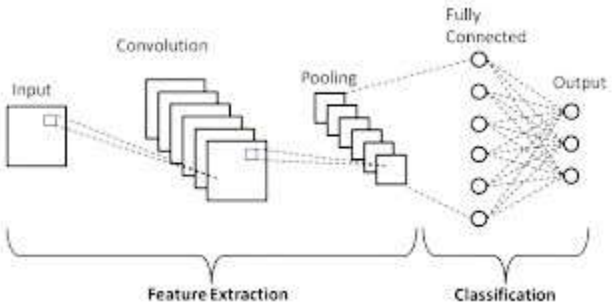

Fig 2: A diagram depicting interaction of the elements

Convolutional Neural Networks are created by first **applying the Rectifier Function**, then **pooling**, **flattening**, and finally creating the **Full Connection**.

Convolution, then ReLU:

The initial stage of the process involves convolution. A third function that depicts how the shapes of the two functions interact is created via convolution, a mathematical action on two functions.
Convolution is mathematically expressed as:

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau \qquad (1)$$

The input picture, the feature detector, and the function map are the three key components of the convolution procedure. Convolution is an vital step since it helps to shrink the size of the input picture. The feature detector scans the input image to extract the most important components, then builds the map of feature while ignoring the remainder. Convolutional neural networks generate convolutional layers with diverse feature maps by combining several feature detectors. Convolutional neural networks are particularly helpful because during training, the network discovers and decides the elements that are crucial for scanning and precisely recognising pictures, some of which may be invisible to the human eye.

The ReLU, an extra step, is added to the convolution process to improve the pictures' nonlinearity. The rectifier function is used to make up for any linearity that may have been introduced during the convolution procedure because pictures are inherently nonlinear. The arrangement of the colours is the main distinction between the original and corrected versions of the image.

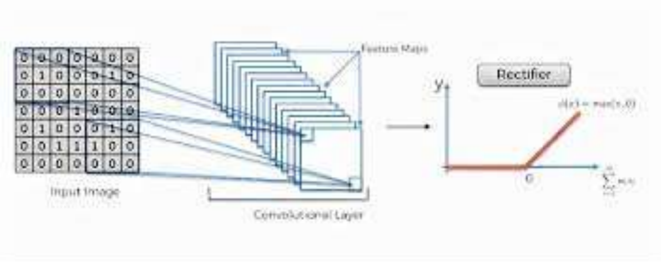

Fig 3: Diagram showing the use of the Rectifier Function after convolution

Pooling:

When features are placed during prediction, the network should not be impacted by their placement in the training set of images, according to the principle of spatial variation., is a crucial component of a convolutional neural network.If the features are differently tilted, the function has a different texture, or the features are significantly closer or farther away, none of these things should have an impact on the network.

Pooling achieves the flexibility that the neural network should have to properly recognise even distorted features. The spatial variance capabilities of the CNN is provided by pooling. Additionally, pooling reduces the number of parameters and the size of the picture through dimensionality reduction, which reduces the CPU power required to analyse the data.

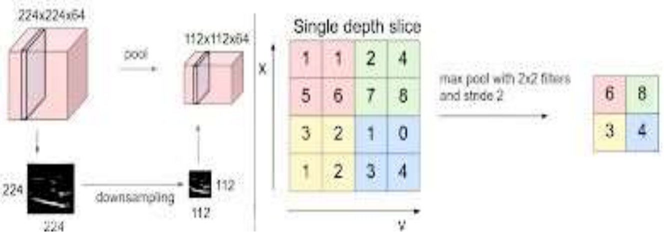

Fig. 4: A graphic illustrating Max Pooling

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Flattening:

During the flattening stage, the Pooled Feature Map is reorganised into a column of values. The flattened result's long input vector is forwarded to the artificial neural network for extra processing.

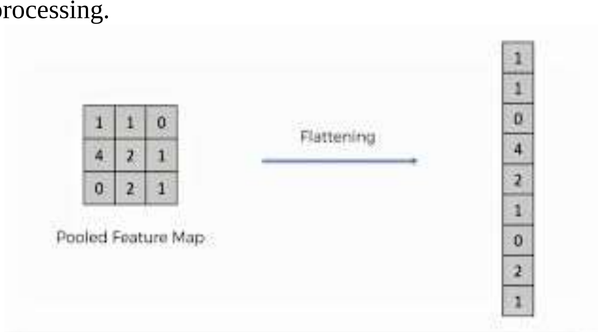

Fig 5: An illustration of the flattening of merged feature maps

Full Connection:

Integrating artificial neural networks with the convolutional neural network is the last stage in the process of building a CNN.
The Step of Full Connection has three layers:
• Output Layer • Input Layer • Fully Connected Layer

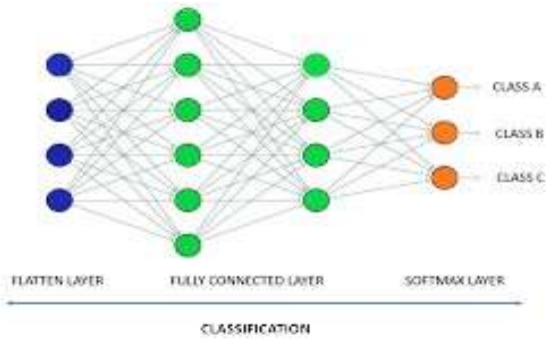

Fig 6: Fully Connected Layer

Data from the flattening process are contained in a data vector in the convolutional neural network's input layer. After accepting this information, the artificial neural network combines all of the characteristics into a larger set of features, which improves the convolutional network's capacity to effectively identify pictures. At the network's conclusion are the output layers that generate predictions..
This is how the entire connecting procedure operates:

- The receptor neuron recognises a certain characteristic in a layer that is fully linked and preserves its value.
- This value is communicated to all the classes by the neuron.
- The classes control and determine if the value of the feature is applicable to them.

Once the Convolutional Neural Network has been developed from scratch, the final step is to evaluate the network's effectiveness by conducting tests.

IV. METHODOLOGY

algorithm of Deep Learning, which is Convolutional Neural Network (CNN).

Requirements:

The Django web framework is used to deploy the complete system, which is constructed using Python 3 in Atom Editor.

Hardware requirements include an
8th generation Intel Core i7 processor,
8 GB or more of RAM,
a 500 GB or more hard drive,
and an NVIDIA GTX 960 graphics card.

Software prerequisites:
• Windows 10 as the operating system
Python v3.8.2;
Atom Editor;
Keras,
TensorFlow,
NumPy,
Scikit-Learn, and OpenCV;
Frontend HTML5, CSS3, and JavaScript; IDE:

• Compiling and georeferencing the bird dataset.

To create our Deep Learning image dataset, we leveraged Microsoft's Bing Image Search API v7, which is a part of Microsoft's Cognitive Services. The dataset primarily focuses on birds found in the Asian sub-continent and contains 8218 images of 60 different bird species. The dataset was built using Python3 in the Atom Editor and deployed using the Django web framework.

- **Putting the CNN architecture into practise.**

The VGGNet network is to be scaled down and made more portable using the CNN design [9].
The following describe the VGGNet architectures:
•A series of 3X3 convolutional layers with varying depths.
•Max Pooling to reduce the amount of parameters and picture size.
Before a softmax classifier, fully connected layers are present at the end of the network.

The identification of bird photos and their categorization into different species are the project's main goals. This project has been developed on the mainstream

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

- ***Developing CNN Model.***

Deep learning commonly divides the dataset into training and testing sets in order to evaluate the performance of the trained model. The train_test_split function was used in this project to generate a random 80:20 split of the dataset, with 80% of the data being used for training and the remaining 20% being used for testing. Data augmentation methods were applied to the ImageDataGenerator class to increase the diversity of the training data and prevent overfitting. The Keras framework and Adam optimizer were used to train the model. The Matplotlib backend was employed during training to store data in the background.

- ***Evaluating the Trained Model's Efficacy.***

In order to identify the species of a bird in a sample image that the user has supplied, the classification script makes use of the trained CNN model. The user submits the image using a web portal, and the server-side testing software gets data from the stored model and label binarizer files to produce the prediction. This is accomplished through a client-server architecture.

The user is subsequently presented with the projected bird species via the online portal. By using fresh, previously undiscovered photos, this approach enables the identification and classification of bird species.

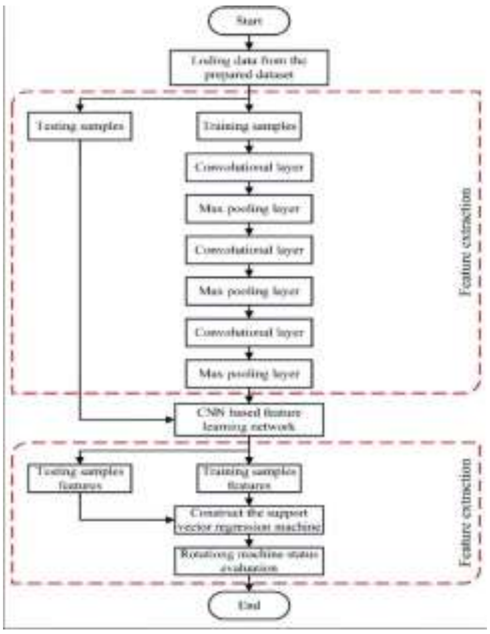

Fig 7: Flow of System

V. RESULTS OF EXPERIMENTAL AND ANALYSIS

These are impressive results for a CNN trained on a dataset of 8218 pictures including 60 bird species. A 93.19% accuracy rate on the training set and an 84.91% accuracy rate on the testing set show that the model is adapting effectively to new data. The TensorFlow backend, the Adam optimizer, and the 80:20 split of training and testing data all contributed to the model's success. The batch size of 32 and the initial learning rate of 1e-3 are notable changes to the model's parameters that may have improved the model's performance. This is a positive result overall for the use of deep learning in picture categorization tasks.

The metrics that were observed were plotted on a graph, which is shown in the

Fig 8: Training Loss and Accuracy Plot

Through the client-server architecture, the prediction script receives the sample picture from the end user as an input. The trained model and label binarizer file are then retrieved from memory after the script loads and preprocesses the picture. The output picture is then presented in a new window with the prediction superimposed on it using the cv2.

Take the example of Fig. 9 below, which served as the input image for the system's prediction.

Fig 9: Input Image

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

After accounting for the model evaluation, the system generates a new window with the expected outcome and accuracy metrics imprinted in it.

Fig 10: A prediction-related output terminal

VI. CONCLUSION AND FUTURE ENHANCEMENTS

This study uses a popular deep learning algorithm called the convolutional neural network to provide a way for identifying the species of birds in photographs. With a 93.19% accuracy rate on the training set and an 84.91% accuracy rate on the testing set, the CNN Model was built from the ground up, trained, and assessed.

The following options are available for improving the application:

- Based on the design of our system, 60 species of birds are currently predicted. The number of species is always expandable.
- We currently have an accuracy of about 93%. To boost accuracy, more images might be used for each class.
- The whereabouts of birds may be shown using the Google Maps API.
- Information on the projected bird, either manually entered or taken from online sources, can be shown.
- A large database that tracks users and their actions may be created.

REFERENCES

[1] Fagerlund, Seppo. "Bird species recognition using support vector machines." EURASIP Journal on Advances in Signal Processing 2007, no. 1 (2007): 038637.
[2] Marini, Andréia, Jacques Facon, and Alessandro L. Koerich. "Bird species classification based on color features." In 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4336-4341. IEEE, 2013.
[3] Barar, Andrei Petru, Victor Neagoe and Nicu Sebe. "Image Recognition with Deep Learning Techniques." Recent Advances in Image, Audio and Signal Processing: Budapest, Hungary, December 10-2 (2013).
[4] Qiao, Baowen, Zuofeng Zhou, Hongtao Yang, and Jianzhong Cao. "Bird species recognition based on SVM classifier and decision tree." First International Conference on Electronics Instrumentation & Information Systems (EIIS), pp. 1-4, 2017.
[5] Branson, Steve, Grant Van Horn, Serge Belongie, and Pietro Perona. "Bird species categorization using pose normalized deep convolutional nets." arXiv preprint arXiv: 1406.2952 (2014).
[6] Madhuri A. Tayal, Atharva Mangrulkar, Purvashree Waldey and Chitra Dangra. “Bird Identification by Image Recognition.” Helix Vol. 8(6): 4349- 4352
[7] Atanbori, John, Wenting Duan, John Murray, Kofi Appiah, and Patrick Dickinson. "Automatic classification of flying bird species using computer vision techniques." Pattern Recognition Letters (2016): 53-62.
[8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv: 1409.1556 (2014).