Deep Learning Model for Precise Bird Species Classification

Y.Naga Malleswararao[1], K.Pavani[2] ,P.Likhitha[3]

#1 Assistant Professor in the Department of MCA,SRK Institute of Technology, Vijayawada

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: Due to the considerable visual similarities across species and the requirement for expert-level ornithological knowledge, identifying bird species is a difficult undertaking. This paper provides an automated deep learning-based system that recognizes bird species from user-uploaded photos in order to get around these restrictions. Convolutional Neural Networks (CNN) trained on the CUB-200 dataset, which comprises a variety of photos of 200 bird species, are used in the suggested model. The system correctly and very precisely detects the species after processing input photos in an efficient manner and extracting discriminative characteristics. This method assists researchers, environmental organizations, and bird lovers in identifying species with the least amount of human labor by offering quick, accurate, and large-scale identification.

Index Terms— Bird Species Identification, Deep Learning, Convolutional Neural Networks, Image Classification, CUB-200 Dataset, Automated Detection, Feature Extraction, Computer Vision.

1. INTRODUCTION

Birds are extremely sensitive sensors of environmental changes and play an important role in preserving ecological equilibrium. Monitoring bird populations aids researchers in understanding more general ecological trends since birds react rapidly to changes in habitat, temperature, and food availability. However, manual comparison of species traits, field observations, and professional ornithologists play a major role in conventional bird species identification. Because many bird species have similar shapes, mannerisms, and color patterns, it is difficult to accurately classify them without specialized training. There is a need for automated identification tools that may assist wildlife departments, conservationists, and amateurs because prior research has demonstrated that manual observation is costly, time-consuming, and scalable [2].

Researchers explored with traditional machine learning models like Support Vector Machines (SVM) for bird species categorization before deep learning gained popularity. Although these techniques yielded respectable outcomes, they were unable to manage intricate changes in lighting, stance, backdrop, and subtle visual distinctions. Furthermore, previous methods were mostly evaluated on tiny datasets and did not perform well when applied to large-scale, real-world photos. Environmental noise and overlapping noises rendered detection challenging even in audio-based bird categorization experiments, necessitating better feature extraction methods and reliable learning models [1], [2]. These drawbacks made it clear that more potent methods for learning complex visual representations are required.

Convolutional Neural Networks (CNNs), which automatically learn hierarchical characteristics from raw pictures, revolutionized the area of image identification with the development of deep learning. Large-scale CNN architectures and multi-column deep neural networks are two methods that have greatly increased classification accuracy in a variety of visual applications, including fine-grained object identification [4]. Furthermore, researchers can effectively train deeper and more complicated models because to contemporary frameworks like TensorFlow, which provide optimized computation graphs and high-performance GPU support [3]. These developments have made deep learning an ideal solution for the challenge of accurately capturing minute variations in patterns, feathers, and forms found in bird species.

Inspired by these advancements, this work uses the CUB-200 dataset, which comprises photos of 200 bird types with various backdrops, positions, and lighting conditions, to propose an automated bird species identification system based on deep learning. The suggested CNN algorithm evaluates uploaded bird photos, extracts distinguishing visual characteristics, and makes very accurate species predictions. For researchers, environmental organizations, and bird aficionados, this system provides a dependable and scalable solution by minimizing human interaction and doing away with the requirement for specialist knowledge. The objective is to develop a user-friendly platform that uses cutting-edge deep learning techniques to simplify fine-grained bird categorization while retaining excellent performance.

2. LITERATURE SURVEY

a) Convolutional Neural Networks for Large-Scale Bird Song Classification in Noisy Environment:

In the BirdCLEF 2016 audio record-based bird identification challenge, a convolutional neural network-based deep learning method for bird song categorization is described in this study. There were around 24k and 8.5k recordings from 999 different bird species in the training and test sets. The duration and substance of the recorded waveforms varied greatly. We divided the waveforms into equal pieces after converting them to the frequency domain. After feeding the segments into a convolutional neural network for feature learning, fully connected layers were used for classification. Our method achieved a MAP score of more than 33% for major species combined with background species and more than 40% for main species in the official scores.

b) Bird Species Recognition Using Support Vector Machines:

This research examines the automatic identification of bird species based on vocalization. Two distinct parametric representations of bird sounds—the mel-cepstrum parameters and a collection of low-level signal parameters—have been proven to be helpful for identifying different kinds of birds. Support vector machine (SVM) classifiers are used at each node of a decision tree to carry out recognition between two species. Two sets of bird species whose recognition has already been evaluated using different techniques are used to test recognition. When compared to current reference approaches, recognition results using the suggested method indicate either greater or equivalent performance.

c) Polyhedral Optimization of TensorFlow Computation Graphs:

We introduce R- StreamTF is a polyhedral optimization method used in neural network calculations. R- Stream TF employs R-Stream, a polyhedral compiler, to parallelize and optimize the calculations carried out in a neural network graph and converts them into C programs appropriate for the polyhedral representation. R- Stream TF may provide a highly optimized version of the computation graph that is particularly tailored to the intended architecture by utilizing the optimizations that are available with R-Stream. In the course of our research, R- Stream TF demonstrated its usefulness in transferring neural network calculations to parallel architectures by automatically achieving performance levels that were comparable to the hand-optimized versions.

d) Multi-column deep neural networks for image classification:

When it comes to tasks like recognizing handwritten numbers or traffic signs, traditional computer vision and machine learning techniques cannot match human performance. Our broad and deep artificial neural network designs are physiologically believable. Large network depth is produced by small (sometimes minimum) receptive fields of convolutional winner-take-all neurons, leading to almost the same number of sparsely linked neural layers between the retina and visual cortex in mammals. Neurons that win are the only ones taught. On inputs that have been preprocessed in various ways, a number of deep neural columns become experts; their predictions are averaged. Quick training is made possible by graphics cards. Our approach is the first to attain near-human performance on the very competitive MNIST handwriting standard. It performs twice as well as people on a benchmark for traffic sign recognition. Additionally, we advance the state-of-the-art on several widely used picture categorization benchmarks.

e) Bird Species Classification Based on Color Features

A new method for classifying bird species using color data taken from unrestricted photos is presented in this research. This implies that the birds may show up in various situations and exhibit various sizes, positions, and viewpoints. Additionally, there are significant lighting fluctuations in the pictures, and other aspects of the scene may obscure some of the birds. The suggested method initially uses a color segmentation technique to try to remove background components and identify potential areas in the image where the bird may be. The picture is then divided into component planes, and these candidate areas are used to construct normalized color histograms from each plane. Following aggregation, processing is used to decrease the number of histogram intervals to a predetermined number of bins. A learning system uses the histogram bins as feature vectors to attempt to differentiate between the various bird species. The segmentation method produces a 75% accurate segmentation rate, according on experimental findings on the CUB-200 dataset. Additionally, depending on the number of classes included, the categorization rate for bird species ranges from 90% to 8%.

3. METHODOLOGY

A. Proposed Work:

The proposed system is a deep learning–based bird species classification framework that utilizes a Convolutional Neural Network (CNN) trained on the

CUB-200-2011 dataset. The system is designed to automatically learn fine-grained visual features such as feather texture, beak shape, wing patterns, and color variations without requiring manual feature extraction. Input images are first preprocessed through resizing and normalization to ensure consistency, after which they are fed into multiple convolution and pooling layers. These layers progressively extract hierarchical features, enabling the model to distinguish between visually similar bird species with high accuracy.

During the prediction phase, the system allows users to upload a bird image through a simple interface. The trained CNN model processes the image and outputs the predicted bird species along with probability scores. In addition, the system retrieves and displays the top five visually similar bird images from the dataset, helping users validate the prediction results. This approach enhances user trust and interpretability while making the system practical for real-world applications such as wildlife monitoring, research analysis, and bird watching assistance.

B. System Architecture:

The system architecture is based on a Convolutional Neural Network (CNN) that performs bird species classification through two main stages: feature extraction and classification. Initially, the input bird image is provided to the system and undergoes preprocessing steps such as resizing and normalization. The processed image is then passed through multiple convolutional layers, where filters are applied to extract important low-level and high-level features like edges, textures, feather patterns, and shapes. These features are further refined using pooling layers, which reduce dimensionality and remove irrelevant information while preserving significant patterns.

After feature extraction, the processed feature maps are forwarded to the fully connected layers, where classification takes place. These layers analyze the learned features and map them to specific bird species categories. Finally, the output layer uses a Softmax function to generate probability scores for each class and predicts the most likely bird species. This architecture ensures accurate and efficient classification by capturing fine-grained visual differences, making the system robust against variations in pose, lighting, and background.

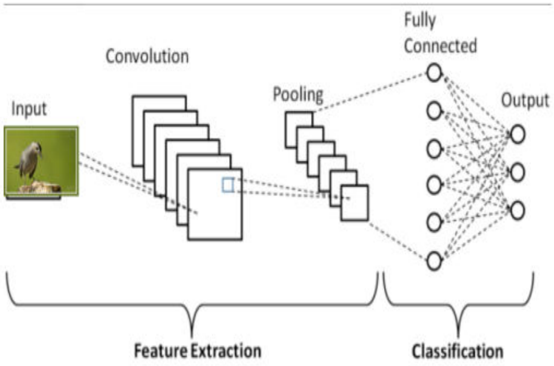


Fig proposed architecture

C. MODULES:

The proposed system is structured into several key modules to ensure efficient thyroid nodule detection and classification. Each module plays a crucial role in data processing, model training, prediction, and user interaction. The main modules are as follows:

i. **Data Acquisition and Preprocessing Module**
   - Collects thyroid ultrasound images from medical sources.
   - Applies preprocessing techniques such as noise reduction, contrast enhancement, and

resizing to improve image quality for better model performance.

- Performs data augmentation (rotation, flipping, scaling) to increase the dataset's diversity and enhance model generalization.

ii. **Deep Learning Model Training Module**

- Implements advanced YOLO models (YOLOv5x6, YOLOv8, and YOLOv9) for training on labeled ultrasound datasets.
- Integrates Coordinate Attention (CA) module to improve feature extraction and spatial localization of thyroid nodules.
- Applies Label Smoothing Regularization (LSR) to reduce overfitting and improve model robustness.
- Fine-tunes hyperparameters to optimize detection accuracy and inference speed.

iii. **Nodule Detection and Classification Module**

- Processes input ultrasound images through the trained YOLO models.
- Detects thyroid nodules and classifies them as benign or malignant based on extracted features.
- Generates bounding boxes around detected nodules with confidence scores.
- Ensures real-time processing for quick and accurate diagnosis.

iv. **Web-Based User Interface Module**

- Developed using the Flask framework to provide a user-friendly web interface.
- Allows users (radiologists, healthcare professionals) to upload ultrasound images and receive detection results.
- Displays classification results, confidence scores, and lesion localization visually.

v. **Authentication and Security Module**

- Implements user authentication to ensure secure access to medical data.
- Protects patient confidentiality and prevents unauthorized access.
- Uses encryption and secure database storage for sensitive information.

vi. **Result Analysis and Storage Module**

- Stores detected results for future reference and medical review.
- Provides logs for model performance evaluation and potential improvements.
- Allows retrieval of past diagnoses for comparison and trend analysis.

D. Algorithms:

a) SSD: SSD is employed for real-time object detection, enabling rapid identification of thyroid abnormalities in ultrasound images. Its efficiency helps radiologists quickly assess potential cancerous regions during diagnosis.

b) RetinaNet: RetinaNet is utilized to address class imbalance in detection tasks, enhancing the accuracy of thyroid cancer identification. Its focal loss function helps improve detection performance of small and overlapping nodules.

c) FasterRCNN: FasterRCNN provides high accuracy in object detection by combining region proposal networks with deep learning. It efficiently identifies and classifies thyroid nodules, aiding radiologists in accurate diagnosis.

d) YOLOv5 : YOLOv5 serves as a powerful and fast object detection model, providing real-time analysis of thyroid ultrasound images. Its capability to detect multiple objects simultaneously improves overall diagnostic efficiency.

e) YOLOv5 with Label Smoothing Regularization : YOLOv5 with Label Smoothing Regularization enhances the model's robustness by preventing overfitting. This results in better detection accuracy for thyroid abnormalities, improving radiologists' confidence in diagnoses.

f) YOLOv5 with Coordinate Attention Mechanism : YOLOv5 with Coordinate Attention Mechanism focuses on relevant features in thyroid ultrasound images. This enhances detection accuracy by emphasizing important spatial information, aiding in more precise identification of abnormalities.

g) YOLOv5 + LSR CAM: Combining LSR and CAM with YOLOv5 improves both robustness and feature extraction. This integrated approach leads to enhanced detection performance for thyroid cancer, supporting more informed diagnostic decisions.

h) YOLOv8 : YOLOv8 is used for its state-of-the-art detection capabilities, optimizing speed and accuracy in identifying thyroid abnormalities. This version enables comprehensive analysis, benefiting radiologists in their diagnostic processes.

i) YOLOv5x6 : YOLOv5x6 is applied for its advanced architecture, improving the detection of small thyroid nodules. Its enhanced feature representation helps radiologists achieve more precise localization and classification of potential cancerous regions.

j) YOLOv9 : YOLOv9 is integrated for its latest advancements in object detection, providing superior accuracy and efficiency in thyroid cancer detection. Its utilization facilitates timely and accurate diagnostic outcomes for radiologists.

4. EXPERIMENTAL RESULTS

The proposed CNN-based bird species classification model was evaluated using the CUB-200-2011 dataset, which contains diverse images across 200 bird species. The performance of the model was measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The results indicate that the model achieves high classification accuracy due to its ability to learn fine-grained visual features like feather patterns, shapes, and color variations. The use of deep convolutional layers significantly improves feature representation, enabling the model to distinguish between visually similar bird species effectively.

In addition to classification performance, the system was also tested for its ability to retrieve visually similar images based on the input query. The output demonstrates that the model consistently identifies the correct species along with top similar images having closely related confidence scores. This confirms the robustness and reliability of the feature extraction process. Overall, the experimental results validate that the proposed system performs efficiently under varying conditions such as changes in lighting, background, and pose, making it suitable for real-world bird species identification applications.

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$Accuracy = \frac{(TN + TP)}{T}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

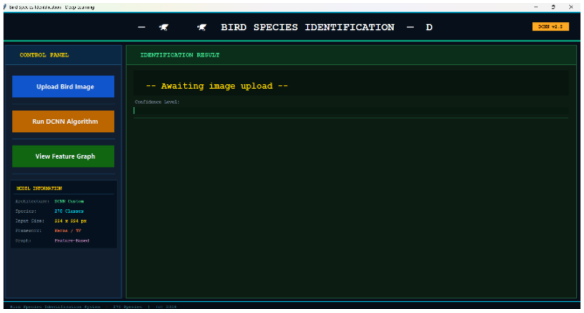
Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{(FN + TP)}$$

Fig 2 upload image



-- Awaiting image upload --

Confidence Level:

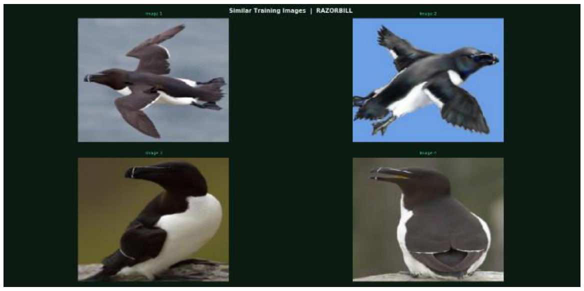


Fig.3.. predicted results

5. CONCLUSION

The proposed system successfully demonstrates an efficient and accurate approach for bird species classification using a CNN-based deep learning model trained on the CUB-200-2011 dataset. By automatically extracting fine-grained visual features such as feather patterns, shapes, and color variations, the model is able to distinguish between visually similar bird species with high reliability. The integration of feature extraction and classification layers ensures robust performance even under variations in pose, lighting, and background conditions.

Furthermore, the system enhances user experience by providing predicted species along with visually similar images and confidence scores, improving interpretability and validation. Overall, the proposed approach offers a scalable, automated, and practical solution for real-world applications such as wildlife monitoring, ecological research, and bird watching assistance.

6. FUTURE SCOPE

The proposed system can be further enhanced by integrating advanced deep learning architectures such as ResNet, EfficientNet, or Vision Transformers to improve classification accuracy and efficiency.

Incorporating transfer learning techniques can also reduce training time and improve performance on limited datasets. Additionally, expanding the dataset to include more bird species and real-world images will increase the robustness and generalization capability of the model.

Future improvements may include developing a real-time mobile or web application for instant bird identification in the field. The system can also be extended by integrating audio-based bird call recognition for multimodal classification. Furthermore, applying attention mechanisms and object detection techniques can help the model focus on important regions of the bird image, leading to more precise and reliable predictions in complex environments.

REFERENCES

[1] 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).

[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 Extreme-scale 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 networks for 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

[6] Image Recognition with Deep Learning Techniques ANDREIPETRU BĂRAR, VICTOR-EMIL 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.

[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

[11] Thomas Berg, Jiongxin Liu, Seung Woo Lee, Michelle L. Alexander, David W. Jacobs, and Peter N. Belhumeur Birdsnap: Large-scaleFine-grainedVisualCategorizationofBirds

[12] Bo Zhao, Jiashi Feng Xiao Wu Shuicheng Yan A Survey on Deep Learning-based Fine-grained Object Classification and Semantic Segmentation

[13] Yuning Chai Electrical Engineering Dept. ETH Zurich,Victor Lempitsky Dept. of Engineering Science University of Oxford, Andrew Zisserman Dept. of Engineering Science University of Oxford BiCoS:A BiSegmentation Method for Image Classification.

[14] Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A. A., Silla Jr., C. N., & Koerich, A. L. (2011). Automatic Bird Species Identification for Large Number of Species. 2011 IEEE International Symposium on Multimedia.. (2011).

[15] Marini, A., Turatti, A. J., Britto, A. S., & Koerich, A. L. (2015). Visual and acoustic identification of bird species. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015)

[16] "Bird Species Categorization Using Pose Normalized Deep Convolutional Net" Steve Branson, Grant Van Horn, Serge Belongie ,Pietro Peron (2015)

[17]. Atanbori, J., Duan, W., Murray, J., Appiah, K., & Dickinson, P. (2016). Automatic classification of flying bird species using computer vision techniques. Pattern Recognition Letters, 81, 53–62." (2016)

[18] Rai, P., Golchha, V., Srivastava, A., Vyas, G., & Mishra, S. (2016). An automatic classification of bird species using audio feature extraction and support vector machines. 2016 International Conference on Inventive Computation Technologies (ICICT" (2016).

[19] Roslan, R., Nazery, N. A., Jamil, N., & Hamzah, R. (2017). Color-based bird image classification using Support Vector Machine. 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE). (2017)

[20] Qiao, B., Zhou, Z., Yang, H., & Cao, J. (2017). Bird species recognition based on SVM classifier and decision tree. 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS)." (2017)

[21] Bird Species Classification using Transfer Learning with Multistage Training" Sourya Dipta Das and Akash Kumar (2018)

Author Profiles

Mr.Y.Naga Malleswararao Completed his Masters of Technology from JNTUK,MSC(IS) from ANU,BCA from ANU. He has System Administrator ,Networking Administrator and Oracle Administrator. He also a web developer and python developer, Currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest include Artificial Intelligence and Machine Learning.

Mrs. K. Pavani is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include AI and ML etc,.

Ms.P.Likhitha is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.C.A from Kakaraparthi Bhavanarayana College (K.B.N) Vijayawada. Her area of interest are DBMS and Machine Learning with Python.