IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.12 ⋇ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10211

# BIRD SPECIES RECOGNITION BASED ON LIGHTWEIGHT MODEL

**Mr. Harshavardhan JR[1], Sai Shirisha SB[2], Shivani GK[3], Shreya R Kiran[4], Tanushree R[5]**

Professor, CSE, KSIT, Bengaluru, India[1]

Student, CSE, KSIT, Bengaluru, India[2-5]

**Abstract**: Bird sounds recognition is of great significance in bird protection. With appropriate sound classification, research can automatically predict the quality of life in the area. Nowadays, the deep learning model is used to classify bird sound data with high classification accuracy. However, the generalization ability of most existing bird sound recognition models is poor, and the complicated algorithm is applied to extract bird sound features. To address these problems, a large data set containing 264 kinds of birds is constructed in this paper to enhance the generalization ability of the model, and then a lightweight bird sound recognition model is proposed to build a lightweight feature extraction and recognition network with MobileNetV3 as the backbone. By adjusting the depthwise separable convolution in the model, the recognition ability of the model is improved.

A multi-scale feature fusion structure is designed, and the Pyramid Split Attention (PSA) module is added to the multi-scale feature fusion structure to improve the adaptability of the network to scale extraction of spatial information and channel information. To improve the refinement ability of the model towards the global information, the channel attention mechanism and ordinary convolution are introduced into Bneck module which makes the Bneck module become the Bnecks module. The experimental results show that the accuracy of Top-1 and Top-5 of the model in identifying 264 kinds of birds on the self-built data set is 95.12% and 100%, which are higher than that of MobileNetV1, MobileNetV2, MobileNetV3 respectively. Although the accuracy is lower than ResNet50, the number of parameters and floating-point operations (FLOPs) of the model is only 2.6M and 127M respectively. The accuracy is only reduced by 2.25% while saving costs.

**Keywords**: Attention mechanism, bird sound recognition, deep learning, lightweight, multi-scale feature fusion.

## I. INTRODUCTION

More than 10,000 species of birds are found in almost every environment, from unspoiled rainforests to suburbs and even cities . Nowadays bird species all over the world are extinct to varying degrees. For example, Hawaii, as the extinction capital of the world, has lost 68% of bird species, which may destroy the entire food chain and thus the ecological environment of Hawaii. Using population monitoring, researchers can understand how local birds respond to changes in the environment and conservation efforts. Being able to monitor bird movements in real-time is the first step in this work . At present, many professionals begin to observe birds for a long time to conserve their species . However, most of the monitoring tasks are manual by professionals. As birds fly fast and are difficult to observe, and when they live on land, they are easily frightened by human activities and cannot be recorded by the camera quickly. Therefore, using image recognition to recognize birds in real-time is both difficult and expensive . What's more, many birds are isolated in inaccessible high-altitude habitats. Due to the difficulties in physical monitoring, more and more professionals generally recognize the bird species by hearing and recording.

This method, called bioacoustics monitoring, can provide a passive and cost-effective strategy for the study of endangered bird populations. Nevertheless , if a manual surveillance program is performed, this monitoring process is time-consuming and laborious, and real-time monitoring of birds in areas such as ecological protection zones can't be carried out. Most people in related fields tend to use Internet of things devices to remotely online monitor bird populations. Since most of the bird protection habitats are in the wild, it is difficult for the online monitoring system to transmit the sound of birds back to the server for data processing, recognition and feedback under good network conditions. If the off-line monitoring is carried out in the bird reserve, the low-cost embedded equipment cannot carry the high complexity sound feature extraction algorithm and high-precision sound recognition algorithm. Therefore, aiming at this point, this paper wants to design a lightweight bird voice recognition algorithm, which can not only achieve high accuracy by using simple and single features, but also make the model small enough to run in low-cost embedded devices

DOI: 10.17148/IARJSET.2023.10211

## II. LITERATURE REVIEW

From pristine rainforests to suburbs and cities, there are more than 10,000 different species of birds. In the modern era, some bird species are extinct worldwide. For instance, Hawaii, the world's capital of extinction, has lost 68% of its bird species, which threatens to devastate Hawaii's biological ecosystem and the entire food chain. Using population monitoring, scientists can learn how nearby birds react to environmental changes and conservation initiatives. The first stage in this endeavor is being able to track bird movements in real-time.

In order to preserve their species, many professionals now start watching birds for extended periods of time. However, the majority of monitoring duties Birds fly quickly and are challenging to observe, so when they land they are easily startled by human activity and cannot be immediately recorded by the camera. Therefore, it is difficult and expensive to use image recognition to identify birds in real time. Because it might be challenging to physically monitor birds, more and more experts are learning to identify different bird species through recording and listening. This technique, known as bioacoustics monitoring, can offer a passive and economical approach for the investigation of endangered bird populations.

[1] Over the past several decades, many bird populations in North America have seen population reductions. However, the effectiveness of protected areas in protecting bird populations within their borders from detrimental effects to populations in nearby unprotected areas has rarely been evaluated. The establishment of protected areas has been used as a conservation action to maintain or aid in the recovery of these populations. In California's San Francisco Bay Area, we conducted our study to assess how well land birds fare in protected regions. They evaluated the population growth rates for 14 species over a 23-year period by point count surveys along riparian corridors in coastal Marin County's protected areas, primarily national parks. These growth rates were contrasted with estimates from the North American Breeding Bird Survey, which represent larger, regional populations, from the Coastal California and Northern Pacific Rainforest Bird Conservation Regions. For 9 of the 14 species, there was evidence of a protective effect. They also anticipated that resident species populations could gain more than migrating species populations, but discovered robust evidence for a protective effect for both groups. The potential for protected areas to not only maintain populations despite declines outside of their boundaries but also to help them recover from recent and prior losses is shown by species with increasing growth rates in coastal Marin County protected areas compared to regionally stable or decreasing populations .

[3] In protected tropical forests, bird groups have shown temporary alterations in taxonomy. It is unknown, nevertheless, whether these modifications endanger the preservation of functional diversity. Here, over the course of ten years, we observed the bird populations in a protected area of the Atlantic Forest in southern Brazil. To determine the mechanisms that drive community assembly, we analyse temporal patterns in taxonomic and functional diversity metrics and contrast the observed and expected functional diversity values. Additionally, we confirmed the presence of its components and the temporal beta diversity. When examining the entire community, passerines, non-passerines, and specialised insectivorous passerines, we discovered changes in species richness through time. In these groupings' temporal beta diversity, taxonomic turnover predominated. However, no appreciable changes in functional richness were found, suggesting functional redundancy among the adapting species. The group of specialised insectivorous passerines, however, exhibited functional nestedness more frequently, indicating functional erosion in this group. Additionally, there were rises in community-wide dispersion and functional divergence in non-passerines. These changes appear to be a result of the occurrence of unusual species with distinct functional features and the optimization of resource utilisation. Our findings imply that the communities' functional architecture were arbitrary.

[4] Monitoring and enhancing biodiversity protection involve classifying automatic bird sounds. A unique method for continually monitoring birds has been made possible by recent developments in deep learning and acoustic sensor networks. Various deep learning-based categorization frameworks for identifying and categorising birds have been developed in earlier studies. In order to further enhance bird sound classification ability, we compare various classification models in this work and fuse some of them. In particular, we use two alternative deep learning architectures for building the fused model in addition to using the same deep learning architecture with various inputs. To characterise the various acoustic components of birds, three distinct time-frequency representations (TFRs) of bird sounds are investigated: the Mel-spectrogram, the harmonic-component based spectrogram, and the percussive component based spectrogram. Aside from several TFRs, To categorise bird noises, a separate deep learning architecture called SubSpectralNet is used. The categorization performance can be significantly improved by fusing certain deep learning models, according to experimental results on 43 different bird species.

[5] Automatic classification of animal vocalizations has great potential to enhance the monitoring of species' movements and behaviors.This is particularly true for monitoring nocturnal bird migration, where automated

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁙ Impact Factor 7.12 ⁙ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10211

classification of migrants' flight calls could yield new biological insights and conservation applications for birds that vocalize during migration. In this paper, we investigate the automatic classification of bird species from flight calls and in particular the relationship between two different problem formulations commonly found in the literature: classifying a short clip containing one of a fixed set of known species (N-class problem) and the continuous monitoring problem, the latter of which is relevant to migration monitoring. We implemented a stateof-the-art audio classification model based on unsupervised feature learning and evaluated it on three novel datasets, one for studying the N-class problem including over 5000 flight calls from 43 different species, and two realistic datasets for studying the monitoring scenario comprising hundreds of thousands of audio clips that were compiled by means of remote acoustic sensors deployed in the field during two migration seasons. We show that the model achieves high accuracy when classifying a clip to one of N known species, even for a large number of species. In contrast, the model does not perform as well in the continuous monitoring case. Through a detailed error analysis (that included a full expert review of false positives and negatives) we show the model is confounded by varying background noise conditions and previously unseen vocalizations. We also show that the model needs to be parameterized and benchmarked differently for the continuous monitoring scenario. Finally, we show that despite the reduced performance, given the right conditions, the model can still characterize the migration pattern of a specific species. The paper concludes with directions for future research.

[6] Increasing human activities have contributed to global climate change and thus resulted in a downward trend in the number of species and population sizes of migratory birds. This trend is closely related to a reduction in habitat size and lower habitat quality. The Poyang Lake wetland in China constitutes one of Asia's largest overwintering habitats for migratory birds. Over the past 10 years, restoration projects have improved the habitat ecology of these wetlands. In this study, we assessed the changes in habitat quality for overwintering migratory birds from 2000 to 2012 near two villages in the Poyang Lake wetland using the InVEST model. The average habitat quality for migratory birds has improved by 18.8% and 47.7%. Differences in the degrees of habitat improvement can be attributed to differences in the change of habitat size and in the impact of threat sources that resulted from the combined effects of restoration and human activities in these two villages. © 2015, Science Press, Institute of Mountain Hazards and Environment, CAS and Springer-Verlag Berlin Heidelberg.

[7] Increasing human activities have contributed to global climate change and thus resulted in a downward trend in the number of species and population sizes of migratory birds. This trend is closely related to a reduction in habitat size and lower habitat quality. The Poyang Lake wetland in China constitutes one of Asia's largest overwintering habitats for migratory birds. Over the past 10 years, restoration projects have improved the habitat ecology of these wetlands. In this study, we assessed the changes in habitat quality for overwintering migratory birds from 2000 to 2012 near two villages in the Poyang Lake wetland using the InVEST model. The average habitat quality for migratory birds has improved by 18.8% and 47.7%. Differences in the degrees of habitat improvement can be attributed to differences in the change of habitat size and in the impact of threat sources that resulted from the combined effects of restoration and human activities in these two villages. © 2015, Science Press, Institute of Mountain Hazards and Environment, CAS and Springer-Verlag Berlin Heidelberg.

[8] This paper describes a system for automatic bird species classification based on features taken from the textural content of spectrogram images. The texture features are extracted using three of the most common texture operators described in the Digital Image Processing literature: Local Binary Pattern (LBP), Local Phase Quantization (LPQ) and Gabor Filters. Aiming to perform more fare comparisons, the experiments were performed over a database already used in other works presented in the literature. In the classification step, the SVM classifier was used and the final results were taken using 10-fold cross-validation. The experiments were performed over a challenger dataset composed of 46 classes, and the best accuracy rate obtained is about 77.65%.

[9] Automated classification of organisms into species based on their vocalizations would contribute tremendously to abilities to monitor biodiversity, with a wide range of applications in the field of ecology. In particular, automated classification of migrating birds' flight calls could yield new biological insights and conservation applications for birds that vocalize during migration. This paper explores state-of-the-art classification techniques for large-vocabulary bird species classification from flight calls. In particular, we contrast a 'shallow learning' approach based on unsupervised dictionary learning with a deep convolutional neural network combined with data augmentation. We show that the two models perform comparably on a dataset of 5428 flight calls spanning 43 different species, with both significantly outperforming an MFCC baseline. Finally, we show that by combining the models using a simple late-fusion approach we can further improve the results, obtaining a state-of-the-art classification accuracy of 0.96.

DOI: 10.17148/IARJSET.2023.10211

[10] Bioacoustics plays a major role in the field of ornithology, ecology, behavior study, habitat monitoring, species conservation, and design of the deterrent system. This work focuses on the design and implementation and performance evaluation of an automatic, more efficient, and flexible bird sound-based recognition system for classifying eight species of popular Eurasian birds in the standard online annotated databases. A dedicated virtual instrument tool with effective GUI is developed that acquires, preprocesses the sound samples and generates a statistically evaluated short-term Fourier transform spectrogram-based feature matrix, suited for the characterization of vocalization patterns of bird species. Using the well-labeled feature data of sound records, a multilayer perceptron artificial neural network (MLP-NN) classifier model is designed, trained, tested, and optimized using a feedforward-backpropagation supervised learning algorithm. Various experiments, following a systematic approach, are conducted to optimize the structure of MLP with respect to the number of neurons in the hidden layer, epochs, and learning rate for attaining enhanced recognition accuracy (96.1%), recall (82.6%) and precision (84.5%). The performance of the optimal model is also analyzed in terms of the recognition capabilities of individual bird species that indicate promising results. A few of the birds are recognized accurately and precisely when present as compared to the others. The tendency of the model to wrongly identify or miss bird species also remained low. The model performance over the unseen dataset also remained satisfactory with a cross- validation classification accuracy of 81.4%. The system being scalable, can easily be reused in the future to retrain the model over a large set of sound samples from real-world recordings with improved acoustic features for achieving very high classification accuracy and reliability.

In this paper, the design of a MediBox has been introduced. This MediBox contains separate portions that can be programmed for different user's needs. MediBox helps the users or take care person by specifying the required medicine quantity. MediBox takes control on alerting patient to take the medicines by giving voice alert and patients can take their medications at the right time.

[11] The Covid-19 pandemic represents one of the greatest global health emergencies of the last few decades with indelible consequences for all societies throughout the world. The cost in terms of human lives lost is devastating on account of the high contagiousness and mortality rate of the virus. Millions of people have been infected, frequently requiring continuous assistance and monitoring. Smart healthcare technologies and Artificial Intelligence algorithms constitute promising solutions useful not only for the monitoring of patient care but also in order to support the early diagnosis, prevention and evaluation of Covid-19 in a faster and more accurate way.

On the other hand, the necessity to realise reliable and precise smart healthcare solutions, able to acquire and process voice signals by means of appropriate Internet of Things devices in real-time, requires the identification of algorithms able to discriminate accurately between pathological and healthy subjects. In this paper, we explore and compare the performance of the main machine learning techniques in terms of their ability to correctly detect Covid-19 disorders through voice analysis. Several studies report, in fact, significant effects of this virus on voice production due to the considerable impairment of the respiratory apparatus. Vocal folds oscillations that are more asynchronous, asymmetrical and restricted are observed during phonation in Covid19 patients. Voice sounds selected by the Coswara database, an available crowd-sourced database, have been e analysed and processed to evaluate the capacity of the main ML techniques to distinguish between healthy and pathological voices. All the analyses have been evaluated in terms of accuracy, sensitivity, specificity, F1-score and Receiver Operating Characteristic area. These show the reliability of the Support Vector Machine algorithm to detect the Covid-19 infections, achieving an accuracy equal to about 97%.

[12] Driver action recognition aims to distinguish normal driver action and some abnormal driver actions such as leaving the wheel, talking on the phone, diving with smoking, etc. For the purpose of traffic safety, studies on the computer vision technologies for driver action recognition have become especially meaningful. However, this issue is far from being solved, mainly due to the subtle variations between different driver action classes. In this paper, we present a new video- based driver action recognition approach based on the hybrid spatial–temporal deep learning framework. Specifically, we first design an encoder–decoder spatial–temporal convolutional neural network (EDSTCNN) to capture short-term spatial–temporal representation of driver actions jointly with optical flow prediction. Second, we exploit the feature refinement Network (FRN) to refine the short-term driver action feature. Then, a convolutional long short-term memory network (ConvLSTM) is employed for long-term spatial–temporal fusion. Finally, the fully connected neural network (FCNN) is used for final driver action recognition. In our experiment, we validate the performance of the proposed framework on our self-created datasets, including a simulated driving dataset and a real driving dataset. Extensive experimental results illustrate that the proposed hybrid spatial–temporal deep learning framework obtains the highest accuracy in multiple driver action recognition datasets (98.9% on the SEU-DAR-V1 dataset and 97.0% on the SEU- DAR-V2 dataset).

DOI: 10.17148/IARJSET.2023.10211

[13] Face recognition is one of the essential applications in computer vision, while current face recognition technology is mainly based on 2D images without depth information, which are easily affected by illumination and facial expressions. This paper presents a fast face recognition algorithm combining 3D point cloud face data with deep learning, focusing on key part of face for recognition with an attention mechanism, and reducing the coding space by the sparse loss function. First, an attention mechanism-based convolutional neural network was constructed to extract facial features to avoid expressions and illumination interference. Second, a Siamese network was trained with a sparse loss function to minimize the face coding space and enhance the separability of the face features. With the FRGC face dataset, the experimental results show that the proposed method could achieve the recognition accuracy of 95.33%.

[14] Inspired by that bird sound has various frequency distributions and continuous time-varying properties, a novel method is proposed for the classification of bird sound based on continuous frame sequence and spectrogram-frame linear network (SFLN). In order to form a continuous frame sequence as the standard input for SFLN, a sliding window algorithm of short frame length is suitable for differentiate the Mel-spectrogram of bird sound. The vertical 3D filter in the linear layer moves linearly along the continuous frame and cover its full frequency band. The weight is initialized to a Gaussian distribution to attenuate the high-and low-frequency noise, thereby extracting the long-and short-term features of the continuous frame of the bird sound. Finally, the GRU network is connected and used as a classifier to directly output the prediction results. Four kinds of bird sound from the xeno-canto website are tested to evaluate the influences of different parameters of sliding window on the effect of SFLN-based classification. In the comparison experiment, the mean average precision (MAP) achieves the highest value of 0.97

[15] This paper presents deep learning techniques for acoustic bird detection. Deep Convolutional Neural Networks (DCNNs), originally designed for image classification, are adapted and fine-tuned to detect the presence of birds in audio recordings. Various data augmentation techniques are applied to increase model performance and improve generalization to unknown recording conditions and new habitats. The proposed approach is evaluated in the Bird Audio Detection task which is part of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE) 2018. It provides the best system for the task and surpasses previous state-of-the-art achieving an area under the curve (AUC) above 95 % on the public challenge leaderboard.

## III. CONCLUSION

In this paper, a lightweight bird song recognition algorithm model is proposed. The classification accuracy of this model can reach 95.12%. Compared with other lightweight networks, the model proposed in this paper has a higher recognition rate. Compared with other depth models, the accuracy of the model of this paper is slightly different, and the number of parameters and computations is reduced. From the analysis of ablation experiments, it can be seen that the improvement proposed in this paper can improve the accuracy of model classification and make the model have a good generalization ability. The future work of this paper includes: 1. Applying the model to embedded devices to realize real-time bird monitoring in nature reserves; 2. Collecting more bird sound data and constructing large bird datasets; 3. Simplifying birds Sound feature extraction, reducing the steps and processes of feature extraction.

## REFERENCES

[1] M. D. Dettling, K. E. Dybala, D. L. Humple, and T. Gardali, ''Protected areas safeguard landbird populations in central coastal California: Evidence from long-term population trends,'' Ornithol. Appl., vol. 123, no. 4, Oct. 2021, Art. no. duab035.

[2] G. D. Duckworth and R. Altwegg, ''Why a landscape view is important: Nearby urban and agricultural land affects bird abundances in protected areas,'' PeerJ, vol. 9, Jul. 2021, Art. no. e10719.

[3] H. S. Oliveira and L. dos Anjos, ''Silent changes in functionally stable bird communities of a large protected tropical forest monitored over 10 years,'' Biol. Conservation, vol. 265, Jan. 2022, Art. no. 109407.

[4] J. Xie, K. Hu, M. Zhu, J. Yu, and Q. Zhu, ''Investigation of different CNNbased models for improved bird sound classification,'' IEEE Access, vol. 7, pp. 175353–175361, 2019.

[5] J. Salamon, J. P. Bello, A. Farnsworth, M. Robbins, S. Keen, H. Klinck, and S. Kelling, ''Towards the automatic classification of avian flight calls for bioacoustic monitoring,'' PLoS ONE, vol. 11, no. 11, Nov. 2016, Art. no. e0166866.

[6] C.-Z. Sun, L. Zhen, C. Wang, B.-Y. Yan, X.-C. Cao, and R.-Z. Wu, ''Impacts of ecological restoration and human activities on habitat of overwintering migratory birds in the wetland of Poyang Lake, Jiangxi Province, China,'' J. Mountain Sci., vol. 12, no. 5, pp. 1302–1314, Sep. 2015.

[7] A. D. P. Ramirez, J. I. de la Rosa Vargas, R. R. Valdez, and A. Becerra, ''A comparative between mel frequency cepstral coefficients (MFCC) and inverse mel frequency cepstral coefficients (IMFCC) features for an automatic bird

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.12 ⋇ Vol. 10, Issue 2, February 2023

DOI: 10.17148/IARJSET.2023.10211

species recognition system,'' in Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI), Nov. 2018, pp. 1–4.

[8] J. Salamon, J. P. Bello, A. Farnsworth, and S. Kelling, ''Fusing shallow and deep learning for bioacoustic bird species classification,'' in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2017, pp. 141–145.

[9] R. Pahuja and A. Kumar, ''Sound-spectrogram based automatic bird species recognition using MLP classifier,'' Appl. Acoust., vol. 180, Sep. 2021, Art. no. 108077.

[10] L. Verde, G. De Pietro, A. Ghoneim, M. Alrashoud, K. N. Al-Mutib, and G. Sannino, ''Exploring the use of artificial intelligence techniques to detect the presence of coronavirus COVID-19 through speech and voice analysis,'' IEEE Access, vol. 9, pp. 65750–65757, 2021.

[11] Y. Hu, M. Lu, C. Xie, and X. Lu, ''Video-based driver action recognition via hybrid spatial–temporal deep learning framework,'' Multimedia Syst., vol. 27, no. 3, pp. 483–501, Jan. 2021.

[12] H. Zou and X. Sun, ''3D face recognition based on an attention mechanism and sparse loss function,'' Electronics, vol. 10, no. 20, p. 2539, Oct. 2021.

[13] X. Zhang, A. Chen, G. Zhou, Z. Zhang, X. Huang, and X. Qiang, ''Spectrogram-frame linear network and continuous frame sequence for bird sound classification,'' Ecol. Informat., vol. 54, Nov. 2019, Art. no. 101009..

[14] M. Lasseck, ''Acoustic bird detection with deep convolutional neural networks,'' in Proc. Detection Classification Acoustic Scenes Events Workshop (DCASE), 2018, pp. 143–147.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.

[16] . X. Sheng, J. Kun, and X. Jie, ''Birdsong recognition algorithm based on multi-dimensional neural network depth feature fusion,'' Signal Process., vol. 38, no. 4, pp. 844–853, 2022.