e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

FEATHER SCAN: A SURVEY ON COMBINING CNN-BASED AUDIO AND IMAGE MODELS FOR BIRD IDENTIFICATION

Prof. A. M. Todkar[*1], Mr. Vipul Badole[*2], Ms. Prathamesh Gawale[*3], Ms. Pavan Gawande[*4], Ms. Varun Inamdar[*5]

[*1]Prof. Department Of Computer Engineering Sinhgad Academy Of Engineering Kondhwa Pune, India.

[*2,3,4,5]Student, Department Of Computer Engineering Sinhgad Academy Of Engineering Kondhwa Pune, India.

ABSTRACT

Bird species identification is essential for biodiversity monitoring, conservation efforts, and ecological studies. Traditional bird identification methods rely on either visual or auditory cues, limiting their accuracy in complex environments. "FeatherScan" presents a survey of advanced techniques combining Convolutional Neural Network (CNN)-based models for both audio and image data to improve identification accuracy. By leveraging CNN architectures optimized for audio spectrograms and bird imagery, this study explores the integration of dual-modality models to capture distinct species characteristics. We review current advancements in audio and image recognition for bird species, analyze their strengths and limitations, and propose an integrated approach. FeatherScan aims to provide insights into the potential of hybrid models, enhancing the precision of automated bird identification systems and contributing to the development of comprehensive ecological monitoring tools.

Keywords: Bird Identification, CNN, Audio Classification, Image Classification, Multimodal Learning.

I. INTRODUCTION

Bird identification plays a vital role in biodiversity conservation, environmental monitoring, and understanding ecological systems. Traditional methods for identifying bird species often depend on human expertise, using either auditory cues like bird calls or visual traits. However, these methods can be limited by environmental noise, lighting conditions, and human error. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the capabilities of automated species identification using either audio or image data. Audio-based CNN models analyze bird calls through spectrograms, capturing subtle patterns in frequency and duration, while image-based CNNs process bird images to recognize distinctive physical traits.

While single-modality models achieve commendable results, their effectiveness is constrained in challenging environments where one modality alone may be insufficient. FeatherScan explores a multi-modal approach by combining CNN-based models for both audio and image data, aiming to create a more robust bird identification system. This survey provides an overview of the methodologies and architectures involved in dual-modality CNN models, assesses current research, and discusses potential integration strategies. By combining the strengths of audio and image data, FeatherScan seeks to address the limitations of single-modality identification and contribute to the development of accurate and comprehensive bird identification solutions, facilitating more effective conservation efforts and ecological insights.

II. RELATED WORK

Research in automated bird identification has increasingly focused on deep learning, with Convolutional Neural Networks (CNNs) demonstrating strong performance in both image and audio-based tasks. Image-based identification methods typically leverage CNNs trained on bird images, extracting visual features like color, shape, and texture to classify species. Well-known models, including ResNet and VGG, have shown high accuracy in this domain. However, they often struggle in environments with visual occlusion or low lighting, which can obscure key identifying features.

For audio-based identification, CNNs process bird sounds through spectrogram analysis, where audio signals are converted into images representing frequency over time. This method has proven effective in distinguishing species based on unique vocal patterns, even in noisy environments. Studies using architectures like WaveNet and VGGish for audio classification have demonstrated robust performance, particularly when identifying birds

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

with distinct calls. While these single-modality approaches are effective, their limitations highlight the potential of hybrid systems. Some recent studies have explored multi-modal approaches, combining audio and image models to enhance species recognition, though this research remains limited. This survey aims to expand on these initial studies, investigating the benefits of integrating CNNs for both image and audio data to create a comprehensive bird identification tool capable of overcoming the constraints faced by single-modality systems.

III. LITERATURE SURVEY

Image-Based Bird Identification

CNN-based image classification has seen significant advancements in the identification of bird species. Early work utilized CNN architectures like AlexNet and VGG, achieving promising results in large-scale datasets such as CUB-200. Studies employing ResNet and Inception models showed improvements by focusing on deeper and more complex architectures, allowing for the extraction of intricate visual features like feather patterns, beak shapes, and colors. However, these models are sensitive to challenging environmental factors, such as partial occlusions, lighting variations, and background clutter, which can reduce identification accuracy in natural settings.

Audio-Based Bird Identification

Audio-based approaches leverage CNNs applied to spectrograms, where bird calls and songs are analyzed visually. Work by researchers using VGGish and WaveNet has demonstrated that CNNs can capture unique frequency patterns and sound duration characteristics in spectrograms, making them effective for species with distinct calls. Audio-based models have shown resilience in environments with background noise and are particularly useful for identifying birds in low-visibility settings. The BirdCLEF challenge, an annual competition in bioacoustics, has propelled advancements in this field, encouraging the development of increasingly refined audio-CNN models.

Hybrid CNN Models for Multi-Modal Identification

Integrating both audio and image data in hybrid models is an emerging research area. Preliminary studies have shown that combining visual and auditory cues can enhance bird identification accuracy, as multi-modal models can leverage strengths from each modality. Hybrid CNNs, often using parallel or sequential architectures, process audio and image inputs separately and then fuse the results in a final classification layer. Research from the ImageCLEF and LifeCLEF challenges indicates that multi-modal approaches can reduce false positives and improve identification in complex environments, where single-modality inputs might fail.

Challenges and Gaps in Multi-Modal Approaches

Although promising, the hybrid model approach faces challenges, particularly in synchronizing audio and visual data inputs, as birds may be visible but silent or vice versa. Additionally, the computational cost and data requirements for training multi-modal CNNs are high, often requiring large labeled datasets for both images and audio. While recent studies have attempted to address these challenges by optimizing architectures and employing transfer learning, a standardized methodology for hybrid bird identification remains underdeveloped.

IV. METHODOLOGIES

In the FeatherScan project, we combine CNN-based models for audio and image data to create a robust bird identification system. This section outlines the methodologies used for data preprocessing, model architecture selection, feature extraction, and the integration of both audio and image modalities.

A. Data Collection and Preprocessing

- **Image Data**: High-resolution images of bird species are collected from online databases such as CUB-200 and ImageNet, focusing on labeled datasets that cover diverse bird types and backgrounds. Preprocessing includes resizing, normalization, and augmentation (e.g., rotations, flips) to improve model generalization in real-world scenarios.
- **Audio Data**: Bird audio recordings are sourced from bioacoustic repositories like Xeno-canto and BirdCLEF, covering a variety of calls, songs, and environmental conditions. Audio preprocessing includes noise reduction, normalization, and transformation into spectrograms using short-time Fourier transform (STFT), enabling CNNs to process audio as visual data.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

B. Model Architecture for Image and Audio Processing

- **Image CNN Model**: For image processing, we employ CNN architectures such as ResNet and VGG. These architectures are chosen for their depth and ability to capture detailed visual features. Models are pre-trained on large image datasets and fine-tuned on bird images to improve classification accuracy.
- **Audio CNN Model**: For audio spectrograms, we use CNN models like VGGish and WaveNet, known for their effectiveness in sound pattern recognition. These models are also pre-trained and then fine-tuned on bird audio data.
- Spectrograms are treated as image data, allowing audio CNNs to capture frequency and time patterns unique to each bird's vocalization.

C. Multi-Modal Feature Fusion

To integrate both image and audio data, FeatherScan uses a parallel processing architecture. Each modality (image and audio) is processed independently in dedicated CNN models, resulting in feature vectors from both the image and audio pathways.

- **Feature Extraction**: Feature vectors from both modalities are extracted from the final layer of each CNN model.
- **Feature Fusion**: The extracted features are concatenated and passed through a fully connected layer, allowing the model to learn combined patterns across both modalities.
- **Decision Layer**: The fused features are input into a softmax classifier for final species classification, capturing the complementary information from both image and audio inputs.

D. Model Training and Optimization

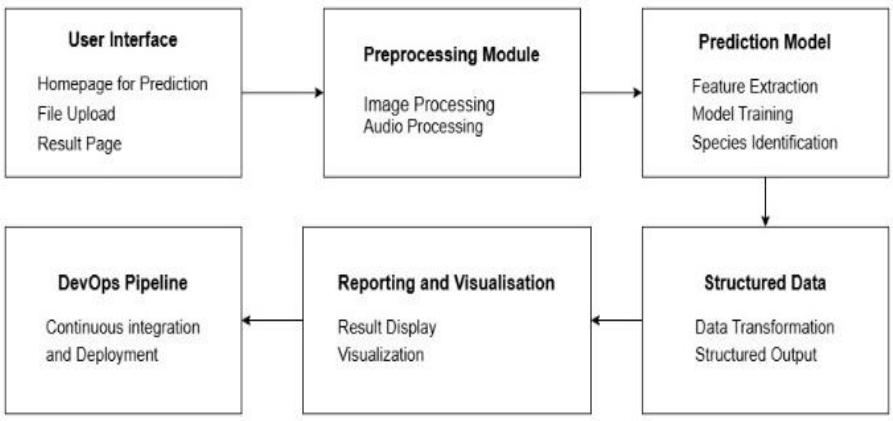
Training involves supervised learning, using labeled image-audio pairs for each species. Techniques such as data augmentation, dropout, and learning rate scheduling are applied to reduce overfitting and enhance model performance. For optimization, FeatherScan uses the Adam optimizer due to its adaptive learning rate, which helps in balancing convergence speed and model accuracy. Transfer learning is employed where feasible to leverage pre-trained weights, further improving efficiency. Convolutional Neural Network (CNN)

E. Evaluation Metrics and Testing

To evaluate model performance, FeatherScan uses standard metrics such as accuracy, precision, recall, and F1-score. Cross-validation ensures that the model's robustness is tested across various subsets of the dataset. Additionally, ablation studies are conducted to assess the contribution of each modality individually, demonstrating the effectiveness of the hybrid approach over single-modality models.images, CNNs save time and improve the precision of medical analysis.

V. PROPOSED SYSTEM

A. Design Concept



The design concept for FeatherScan is structured to provide a seamless user experience from data input through to species identification, utilizing a pipeline that integrates data preprocessing, model prediction, and result visualization. The following components make up the FeatherScan system architecture:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

1. User Interface

Homepage: The homepage serves as the entry point for users, offering an intuitive interface for uploading image and audio files for bird species identification.

Prediction and Result Page: Users can view their prediction results on the result page, which displays the identified species along with any additional details like confidence level and a visualization of features that contributed to the prediction.

2. Preprocessing Module

Image Processing: Uploaded bird images are resized, normalized, and enhanced for better feature recognition.

Audio Processing: Audio files undergo noise reduction and are converted into spectrograms, allowing the system to treat them as image data for input into the CNN model.

3. Prediction Model

Feature Extraction: CNN-based models independently extract features from image and audio inputs, creating feature vectors for each modality.

Model Training: The feature vectors are processed in fully connected layers and trained to classify bird species accurately.

Species Identification: The fused features from both modalities are used to predict the bird species with a softmax classifier.

4. Structured Data

Data Transformation: Prediction results are structured and transformed into a user-friendly format for display.

Structured Output: Outputs are organized, allowing users to see detailed insights and data on identified species.

5. Reporting and Visualization

Result Display: Prediction results are presented to users in a clear format, showing the species name, confidence score, and visual indicators.

Visualization: The system provides additional visualizations like spectrograms and feature maps, helping users understand how the model arrived at its prediction.

6. DevOps Pipeline

Continuous Integration and Deployment: FeatherScan is built with a DevOps pipeline for automatic integration and deployment, ensuring updates are tested and released efficiently. This pipeline helps keep the system up-to-date with the latest improvements in accuracy and user experience.

VI. OBJECTIVES

A. Develop a Multi-Modal Bird Identification System

Combine CNN-based audio and image models to improve bird species identification accuracy, utilizing both visual and auditory cues.

B. Implement Effective Preprocessing

Design preprocessing modules for optimizing image and audio data, ensuring clean, high-quality inputs for reliable model predictions.

C. Enhance Feature Extraction and Prediction Accuracy

Leverage CNN architectures for extracting detailed features from each modality, integrating them to produce robust species identification.

D. Provide a User-Friendly Interface

Create an intuitive user interface with easy file uploads and clear result displays, allowing users to interact smoothly with the system.

E. Ensure Continuous Integration and Deployment

Establish a DevOps pipeline to facilitate regular updates, testing, and efficient deployment of system improvements.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

F. Visualize and Report Results

Deliver structured data output and visualizations to help users interpret prediction results and understand model insights.

VII. FUNCTIONALITIES

1. **File Upload**: Allows users to upload bird image and audio files for identification.
2. **Preprocessing**: Processes uploaded files by resizing images, reducing noise in audio, and converting audio to spectrograms.
3. **Feature Extraction**: Extracts visual and auditory features using CNN models.
4. **Species Prediction**: Identifies bird species by classifying the extracted features through a trained CNN model.
5. **Result Display**: Shows the predicted species, confidence score, and relevant visualizations (e.g., spectrograms, feature maps).
6. **Continuous Deployment**: Automatically updates and deploys new versions of the system through a DevOps pipeline.
7. **Data Transformation**: Converts raw prediction data into a structured output for easy user interpretation.
8. **Visualization**: Provides graphical representations of the prediction results for better user understanding.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project advisor and all faculty members for their invaluable guidance and support throughout the development of this project. We also appreciate the resources and datasets provided by the various open-source repositories, which were essential to our research. Our heartfelt thanks go to our families and friends for their continuous encouragement. Lastly, we acknowledge the contributions of all the researchers whose work laid the foundation for this project.

VIII. CONCLUSION

In conclusion, FeatherScan demonstrates the potential of combining CNN-based audio and image models for accurate bird species identification. By integrating both visual and auditory data, the system enhances identification accuracy, even in challenging environments. The innovative use of multi-modal data, alongside effective preprocessing, feature extraction, and a user-friendly interface, sets a strong foundation for future advancements in bioacoustic and image-based recognition systems. Through continuous integration and deployment, FeatherScan ensures consistent improvements, making it a valuable tool for researchers and bird enthusiasts alike.

VII. REFERENCES

[1] D Stowell, M Wood , H Pamula, H Glotin , " Automatic acoustic detection of birds through deep learning : The First Bird Audio Detection Challenge ",IEEE , July 2018,vol1

[2] Yannis Stylianou , Mike Wood, Dan S, Herve G, Bird Detection in Audio: A Survey and A Challenge, IEEE, September, 2016, vol1,Italy

[3] A. Thakur, V. Abrol, P. Sharma, and P. Rajan, Bird Sound Identification using Deep Learning, 2024

[4] C.-H. Lee, C.-C. Han, and C.-C. Automatic acoustic detection of birds through deep learning: The First Bird Audio Detection Challenge, 2023

[5] Xeno-canto, https://www.xeno-canto.org/, 2018, [Online; accessed 2018- 06- 20]