Zurich University of Applied Sciences

Department Life Sciences and Facility Management

Institute of Natural Resource Sciences

TERM PAPER 1

---

ecoacoustics

Author:
Julian Kraft[1]

Tutor:
Dr. Matthias Nyfeler[2]

Affiliations:
[1]Institute of Natural Resource Sciences
[2]Institute of Computational Life Sciences

Imprint

Project type: Term Paper 1
Title: ecoacoustics
Date: May 24, 2024
Keywords: ecoacoustics, deeplearning, machinelearning, signalprocessing, audioanalysis

Author: Julian Kraft[1] (kraftjul@students.zhaw.ch)
Tutor: Dr. Matthias Nyfeler[2] (nife@zhaw.ch)
Affiliations: [1]Institute of Natural Resource Sciences
[2]Institute of Computational Life Sciences

Abstract

Contents

1 Introduction

1.1 Background

The question about biodiversity and its importance has been a topic of interest for many years. The term biodiversity is a contraction of biological diversity, which refers to the variety and variability of life forms on Earth. In recent years, a massive decline in biodiversity has been observed, which is mainly due to human activities. The loss of biodiversity is a major concern because it can have a significant impact on the ecosystem and the services it provides (Brondízio et al. 2019). In order to quantify biodiversity and monitor its changes, it is essential to have a reliable and efficient method for measuring biodiversity. Traditional methods for measuring biodiversity are time-consuming and expensive, and they are not suitable for large-scale monitoring. But what if there was a non invasive method that could be used to monitor biodiversity in a fast and efficient way? Ecoacoustics might deliver a promising solution to this problem even tough there remain some issues to be fixed (Scarpelli et al. 2021). Passive acoustic monitoring (PAM) like the using of sound recordings to monitor biodiversity are currently widely researched and developed and even combined with modern artificial intelligence (AI) methods (Deng 2023).

The focus of this study is to reproduce the results of the paper (Faiss 2022) and to create a model that can classify the insect sounds with a high accuracy. Furthermore the model will be tested and evaluated for its performance and accuracy. The results will be discussed and compared to the results of the original paper. The goal is to prof that this technology could be accessible for everyone with the knowledge and a regular gaming computer with a graphic processing unit (GPU).

1.2 Insects of Interest

There are countless species of insects in the world, and many of them produce sounds for various reasons. In this study, we are interested in the sounds produced by two groups of insects: Orthoptera and Cicadidae. Orthoptera is an order of insects that includes grasshoppers, crickets, and katydids. (Capinera 2008) Cicadidae, a members of the superfamily Cicadoidea Westwood are four-winged insects with sucking mouthparts that possess three ocelli and a rostrum that arises from the base of the head. (Sanborn 2008)

1.3 Dataset

The dataset used in this study is the InsectSet32 dataset (Faiss 2022). Containing this description:

This dataset contains recordings of 32 sound producing insect species with a total 335 files and a length of 57 minutes. The dataset was compiled for training neural networks to automatically identify insect species while comparing adaptive, waveform-based frontends to conventional mel-spectrogram methods for audio feature extraction. This work will be submitted for publication in the future and the dataset can be used to replicate the results or for similar research. Roughly half of the recordings (147) are of nine species belonging to the order Orthoptera. These recordings stem from a dataset that was originally compiled by Baudewijn Odé (unpublished). The remaining recordings (188) are of 23 species in the family Cicadidae. These recordings were selected from the Global Cicada Sound Collection hosted on Bioacoustica (Baker et al. 2015a), including recordings published in (Baker et al. 2015b; Popple 2017). Many recordings from this collection included speech annotations in the beginning of the recordings, therefore the last ten seconds of audio were extracted and used in this dataset. All files were manually inspected and files with strong noise interference or with sounds of multiple species were removed. Between species, the number of files ranges from four to 22 files and the length from 40 seconds to almost nine minutes of audio material for a single species. The files range in length from less than one second to several minutes. All original files were available with sample rates of at least 44.1 kHz or higher but were resampled to 44.1 kHz mono WAV files for consistency.

The files are split into training, validation and test sets. And there are two .csv files containing the labels and the filenames of the recordings.

2 Materials and Methods

2.1 Programming Language and Frameworks

To build and train the deep learning model, the programming language Python was used. The Frameworks PyTorch, Lightning are very popular and powerful tools for building deep learning models.

2.2 Data Processing

To load and process the the data on the fly, a custom data loader was implemented. The data loader reads the audio files and their corresponding labels from the dataset and applies the necessary transformations to the audio files.

2.2.1 Sample Size

The audio files are of different lengths. In order to avoid the model being biased towards the length of the audio files, the audio files are sampled to a fixed length. Since there is files below the fixed length, the audio files are padded with zeros, it seems not enough to just sample the files and pad them if needed. This would allow bias to be introduced because of the padding. To avoid this, the audio files are sampled to a random length between 1 and 5 seconds and then padded with zeros to the fixed length of 5 seconds.

2.2.2 Transformation

Before the audio files are fed into the model, they are transformed into a mel-spectrogram - short for melody spectrogram. A mel-spectrogram is a visual representation of the audio signal aiming to mimic the human perception of sound and is commonly used in audio processing tasks like speech recognition and music genre classification. It does however provide certain advantages for audio classification in general and can therefore be used in the field of ecoacoustics as well (Stowell 2022, p. 7). The mel-spectrogram is a 2D array that represents the frequency content of the audio signal over time.

2.2.3 Deep Learning Model

2.2.4 Training

2.2.5 Evaluation

3 Results

4 Discussion

References

Brondízio, E. S., J. Settele, S. Díaz, and H. T. Ngo, eds. (2019). *The Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Bonn: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). ISBN: 978-3-947851-20-1.

Scarpelli, M. D. A., B. Liquet, D. Tucker, S. Fuller, and P. Roe (Dec. 17, 2021). "Multi-Index Ecoacoustics Analysis for Terrestrial Soundscapes: A New Semi-Automated Approach Using Time-Series Motif Discovery and Random Forest Classification". In: *Frontiers in Ecology and Evolution* 9, p. 738537. ISSN: 2296-701X. DOI: 10.3389/fevo.2021.738537. URL: https://www.frontiersin.org/articles/10.3389/fevo.2021.738537/full (visited on 05/06/2024).

Deng, I. (2023). "Harnessing the Power of Sound and AI to Track Global Biodiversity Framework (GBF) Targets". In.

Faiss, M. (Sept. 12, 2022). *InsectSet32: Dataset for Automatic Acoustic Identification of Insects (Orthoptera and Cicadidae)*. Version 0.1. Zenodo. DOI: 10.5281/zenodo.7072196. URL: https://zenodo.org/records/7072196 (visited on 04/18/2024).

"Orthoptera" (2008). In: *Encyclopedia of Entomology*. Ed. by J. L. Capinera. Dordrecht: Springer Netherlands, pp. 2695–2695. ISBN: 978-1-4020-6359-6. DOI: 10.1007/978-1-4020-6359-6_1892. URL: https://doi.org/10.1007/978-1-4020-6359-6_1892 (visited on 04/28/2024).

Sanborn, A. (2008). "Cicadas (Hemiptera: Cicadoidea)". In: *Encyclopedia of Entomology*. Ed. by J. L. Capinera. Dordrecht: Springer Netherlands, pp. 874–877. ISBN: 978-1-4020-6359-6. DOI: 10.1007/978-1-4020-6359-6_666. URL: https://doi.org/10.1007/978-1-4020-6359-6_666 (visited on 04/28/2024).

Baker, E., B. Price, S. Rycroft, J. Hill, and V. S. Smith (Jan. 1, 2015a). "BioAcoustica: A Free and Open Repository and Analysis Platform for Bioacoustics". In: *Database* 2015, bav054. ISSN: 1758-0463. DOI: 10.1093/database/bav054. URL: https://doi.org/10.1093/database/bav054 (visited on 04/28/2024).

Baker, E., B. Price, S. Rycroft, and M. Villet (Sept. 2, 2015b). "Global Cicada Sound Collection I: Recordings from South Africa and Malawi by B. W. Price & M. H. Villet and Harvesting of BioAcoustica Data by GBIF". In: *Biodiversity Data Journal* 3, e5792. ISSN: 1314-2828. DOI: 10.3897/BDJ.3.e5792. URL: https://bdj.pensoft.net/article/5792/ (visited on 04/28/2024).

Popple, L. W. (Oct. 27, 2017). "A Revision of the Myopsalta Crucifera (Ashton) Species Group (Hemiptera: Cicadidae: Cicadettini) with 14 New Species from Mainland Australia". In: *Zootaxa* (Vol. 4340 No. 1). DOI: 10.11646/zootaxa.4340.1. URL: https://doi.org/10.11646/zootaxa.4340.1 (visited on 04/28/2024).

Stowell, D. (Mar. 21, 2022). "Computational Bioacoustics with Deep Learning: A Review and Roadmap". In: *PeerJ* 10, e13152. ISSN: 2167-8359. DOI: 10.7717/peerj.13152. arXiv: 2112.06725 [cs, eess, q-bio]. URL: http://arxiv.org/abs/2112.06725 (visited on 03/05/2024).