Project Proposal

Bird Singing Detection

1. Project Description

Bird sound detection and identification (e.g. calls and songs) is a well known challenge in audio analysis and there are already some commercial solutions available.[1] However, those solutions are either for very specific areas (limited amount of species / specific sounds) or they are prone to errors and misidentification. Additionally those proprietary solutions do not share their algorithms and source code so that there is no opportunity to evaluate and use them on a large scale.
This project aims to develop a solution to the problem of bird sound identification in cooperation with the University of Natural Resources and Life Sciences Vienna that can be used on a larger scale with already recorded data. The goal is to process already recorded audio files and identify the exact time and duration of species-specific bird calls and songs that occur in these recordings.

2. Dataset

The dataset consists of 1.440 audio recordings. One data sample is a 5 minute audio file with two channels recorded using an acoustic recording unit called AudioMoth.[2] The sample rate is 32 kHz resulting in a digital representation of the file as an array with shape (9600000, 2).
Out of the total 1.440 audio files, information on bird species detected are available for 801 files.. However, additional information on the exact time and duration of the different species singing/calling in the already labelled files is not available (weak labels). A total of 63 different species occur in the labeled files. On average there are between 8 and 9 different species in a 5 minute recording, with at least 1 and at most 19.
The species common blackbird (*Turdus merula*), chaffinch (*Fringilla coelebs*) and great tit (*Parus major*) occur in over 75% of the labeled files, whereas 10 species only occur in 1 audio file.
The audio files were recorded in two different locations which results in different location specific noise.

3. Methodology

3.1. Data Analysis and research

The first step is to get a good overview of the dataset and the characteristics of the audio files and label distribution. Furthermore, initial research has to be conducted on state of the art mechanisms for nose reduction, bird identification, usage of weak labels and additional data sources. For bird identification there exists several challenges on *Kaggle* and other

[1] eg: http://www.songsleuth.com/ or http://www.songsleuth.com/
[2] https://www.openacousticdevices.info/audiomoth

platforms.[3] The problem of noise reduction is also a large research field with different areas of application and well working algorithms and heuristics, which can, to some extent, be applied to audio files as well.
Outcome: Notebook with insights into the dataset structure and audio files.

3.2. Noise Reduction

After initial research and data analysis the first step is to improve the audio quality of the samples as much as possible without losing any relevant information. There already exist several methods and techniques in order to achieve this goal because of the wide range of applications (telefon calls, music, …). From rather simple frequency bases filters (bandpass filter and its extensions[4]) to more dedicated generative neural networks (special GANs which produce a completely new audio stream with only the relevant information[5]), all kinds of methods can be used and combined to get the best result.
This part of the project tries to identify fitting methods and find a good pipeline to produce clean audio files. A combination of generic methods and algorithms tailored to a specific location seems most promising.
Outcome: Audio files with reduced noise and noise reduction pipeline for streams produced by an AudioMoth.

3.3. Bird singing detection and classification

Using this noise reduced audio files the last step is to identify and classify bird sound (song, call). This can be done several ways. Either identifying and classifying bird soundin one step (e.g. using templates of typical bird songs/calls) or splitting up the process. First identifying when and for how long bird singing is present and then classifying the species of the identified song/call. For both parts different machine learning algorithms can be applied (eg. Recurrent Neural Network, Convolutional Neural Networks on Spectrograms, …)
Outcome: Statement about the functionality of different approaches for classifying bird songs/calls. If quality criterion is met, a dataset with strong labels for all 1.440 audio files as well as a pipeline for training and classifying audio files for streams produced by an AudioMoth .

4. Timeline

01.11.2021 - 15.11.2021:	Data Analysis
15.11.2021 - 15.12.2021:	Noise Reduction
15.12.2021 - 31.01.2022:	Bird singing detection and classification
31.01.2022:	Submission of results and report

[3] https://www.kaggle.com/c/birdsong-recognition or https://www.kaggle.com/c/birdclef-2021
[4] https://arxiv.org/pdf/1608.04069.pdf
[5] https://arxiv.org/pdf/1703.09452.pdf

5. Preliminary references

5.1. Noise Reduction Overview:

http://www.iraj.in/journal/journal_file/journal_pdf/6-162-1440572779132-135.pdf (Review)
https://arxiv.org/pdf/1904.12069.pdf (Noisy2Noisy Signal Mapping)

5.2. Bird Singing Identification:

https://www.researchgate.net/publication/328836649_Bird_Sound_Recognition_Using_a_Convolutional_Neural_Network (CNN)
https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13357 (Wavelet Filter)
https://www.sciencedirect.com/science/article/pii/S1470160X2030546X (Ensamble Approach)