

Deep Learning for Ecology report on bird recording at INTAKA island

Emmanuel NYANDU KAGARABI
African Institute for Mathematical Sciences
AIMS SOUTH AFRICA
Cape Town, South Africa
emmanuelnk@aims.ac.za

Abstract—Theory is good, but practice is better. In this report, we present the quintessence of our assignment where we used the Convolutional Neural Networks (CNNs) to study bird sounds recorded at INTAKA island using Raspberry Pi Zero ARU on February 3, 2024.

Index Terms—Deep Learning, ecology, spectrogram, Raspberry Pi Zero ARU, CNN.

I. INTRODUCTION

The latest report from the Intergovernmental Panel on Climate Change ([Groupe d'experts Intergouvernemental sur l'Evolution du Climat \(GIEC\)](#)) shows that the cause of global warming since 1850 is mainly man-made. As a result of our greenhouse gas emissions, we have succeeded in modifying our planet's climate, resulting in a total global surface temperature rise of 1.07°C. GIEC's experts are warning of the consequences of global warming: an increase in the frequency and intensity of extreme phenomena (tropical cyclones, heat waves, droughts, floods), rising sea levels, melting sea ice, glaciers and permafrost. This disruption has a direct impact on terrestrial and ocean ecosystems, threatening their functioning and the biodiversity that supports them, their functioning, as well as the biodiversity of which they are composed [1, p.3]. According to the same report, a global warming of 1.5°C compared to the average pre-modern global temperature, 6 % of insects, 8 % of plants and 4 % of vertebrates would lose more than half their geographic range. Without the capacity to adapt rapidly to global warming, species could be doomed to extinction. So the study of behaviours is an important element in understanding how individuals use their habitat, identify their capacity to adapt to climate change and identify the factors that threaten the equilibrium of ecosystems. Numerous studies have demonstrated the importance of studying behaviours for the conservation of endangered species.

Nevertheless, observing behavior in the wild remains a difficult process, depending on environmental conditions, and the presence of an observer generally has an impact on the behavior of the species studied. It is therefore important to develop methods for studying the behavior of species in their natural environment, while limiting the impact of this observation [1]. These 21st-century

crises highlight the need, more than ever, to study living things in order to better protect them. That's what we're trying to do with the Deep Learning for Ecology lecture, more especially the Bio-acoustics section, which is by definition the exploration of animal vocalizations and natural soundscapes [3], aspects developed in a such study's methodology can assist ecologists, wildlife management teams, and researchers in reducing the amount of time spent analyzing large acoustic datasets obtained from passive acoustic monitoring studies [2, p.2]. Bioacoustics, has proven to be a valuable source of data for both a better understanding of animal behavior and for biodiversity monitoring. Over the past few years, deep learning has revolutionized several research fields such as bioinformatics and medicine by enabling automated processing of large and complex datasets. Considered a branch of machine learning, deep learning refers to algorithms, commonly called deep neural networks, able to automatically detect very complex and highly discriminating patterns in data [2].

The succession of processing layers performing linear and non-linear transformations allows the neural networks to learn representations of data with multiple levels of abstraction. This ability makes deep learning particularly relevant for solving complex problems such as speech recognition, object detection, computer vision, and many other domains. Moreover, the most common approach in deep learning in bioacoustics is to convert the raw acoustic recordings into a succession of spectrograms or mel-spectrograms – a visual representation of the frequency spectrum of sound over time – and treat it as an image classification problem. These spectrograms generally give a fixed image in time of the sounds and deprive them of their context [2]. Having collected bird sounds at INTAKA island, we are using Convolutional Neural Networks in this work to discern the presence or absence of a bird sound in any given audio recording in our dataset.

II. MATERIALS AND METHODS

A. Data collection

It all began on Friday, February 02, 2024, when we were taught to assemble the Raspberry Pi Zero ARU kit in the lecture hall. The Raspberry Pi Zero is an incredibly small

and affordable single-board computer developed by the Raspberry Pi Foundation. It's designed to be extremely compact, while offering sufficient processing power for a variety of applications (For example automatic recording in this report). Here, below (Fig. 1), is the complete kit with its various parts.

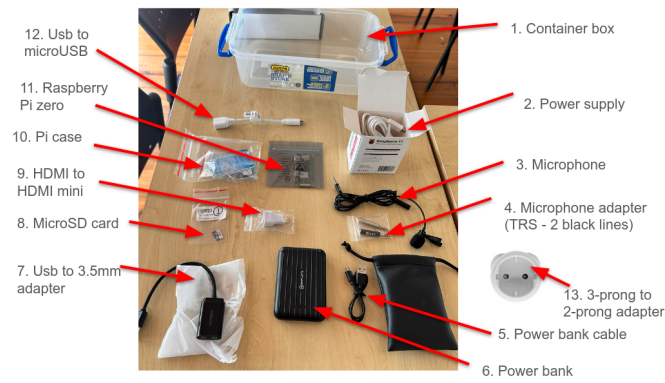


Fig. 1: Raspberry Pi Zero ARU

The next day; Saturday 03/02/2024 was the big day when we left AIMS early for INTAKA island (Fig.2) to collect data. We left AIMS in the morning (at 7am) and arrived at INTAKA at 8am. Once there, in groups of two, each student with a complete recording kit, the good work began around 8:30 am and we were asked to disconnect the power bank every 5 minutes and reconnect it again. This was to optimize the recording because, without it,

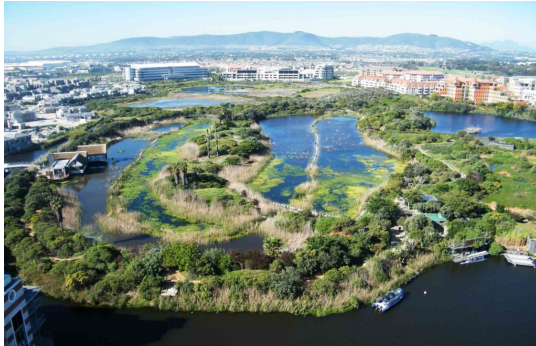


Fig. 2: Intaka Island Nature Reserve, Century City, Cape Town

some people might assume that they were recording when the kit was probably not plugged in properly. At the same time, at the start of each recording, the geographical coordinates of the recording location were determined using the telephone's GPS. Around 11:45 am, it's time to go back at AIMS. Once back, the annotation work had begun. Samples were collected from each audio file. Conventionally, a sample in which the sound of a bird was present had to be 'annotated' with the digit **1**, and with **0** in the opposite case. We illustrate this with a screenshot (Fig. 3) of the

Sonic Visualiser window where the annotations were made.

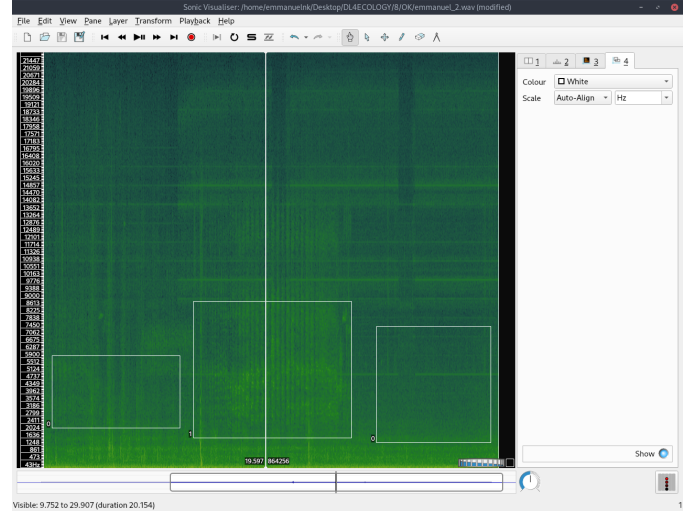


Fig. 3: Sonic Visualiser

B. Data pre-processing

I annotated all my audio files and create a folder named after my ID : 08, within which I should create two sub-folders: **audios** for all my recordings (the .wav files) and **annotations** for the corresponding annotation files (the .svl files). Each of my comrades followed the same procedure, and together we compiled the *AIMS_BIRD_DATASET* folder, which was studied and corrected perfectly by a number of comrades, whom I congratulate in passing. At the end of the process, we all used a uniform database. The desired folder structure is illustrated in the following class template (Fig. 4).



Fig. 4: AIMS_BIRD_DATASET

C. Training and testing the models

In my work, each recording was downsampled to 22,050 Hz and converted into a mel-scale spectrogram to be used as an input image for Convolutional Neural Network (CNN). I performed a Hann analysis window size of 46 ms (1,024 samples) with a hop size of 12 ms (256 samples) and 128 mel frequency bins. I restricted the frequency range of the spectrograms between 150 Hz and $15kHz = 15 \times 1000Hz = 15,000Hz$. Using our manual annotations, the songs were extracted from the spectrograms and divided into equal 3s segments using a sliding window with an overlap of 1s. Each spectrogram was thus an image of size 128×259 . I applied [frequency masking](#) as a data augmentation. You can find easily the details of my model used into my colab notebook. The models were implemented in Python 3 using the **TensorFlow** and **Keras** libraries, and the audio processing and spectrogram construction were performed using Librosa.

III. RESULTS AND DISCUSSION

Considering (128,259,1) as the shape of input data (height=128 pixels, width=259 pixels, and number of channel=1), I wrote a code which defines a Convolutional Neural Network (CNN) using Tensorflow and keras for a classification task (presence or absence of bird sound). It comprises multiple convolutional layers with ReLU activation, followed by **max-pooling** layers for downsampling. A **flattening** layer prepares the data for fully connected layers, and the final output layer uses **softmax** activation for classification. The model is compiled with **categorical cross-entropy** loss, the **Adam** optimizer, and **accuracy** metrics. Applying `model.summary()`, I get a total of 305,698 parameters, which corresponds to approximately 1.17 megabytes of memory. All of these parameters are trainable, meaning they will be updated during the training process based on the input data and the defined loss function. There are no non-trainable parameters in my model. Using the command `model.fit(X_dataset,Y_dataset,epochs=3, validation_split=0.2)`, the model shows improvement in both training and validation metrics with each epoch. The training loss decreases, and the training accuracy increases as the model learns from the training data. The validation loss and accuracy also improve, indicating that the model generalizes well to unseen data. Finally, I wrote a code for generating two plots to summarize the training history of the model.

IV. CONCLUSION

This intense work enabled us to combine theory and practice in the Deep Learning for Ecology course. With this work, we succeeded in implementing an algorithm for detecting the sound of a bird. We didn't use all the data available in our database, as RAM memory was running low and we had no option but to reduce the size of the data. For more information, please refer to our [colab](#).

DATA AVAILABILITY

All code for training and testing the neural networks is available at my own Github repository https://github.com/emmanuelnyandukagarabi/DL4ECOLOGY/blob/main/emmanuelnk_DL4E_Assignement_2.ipynb.

ACKNOWLEDGMENTS

I would like to warmly thank our supervisors Dr Emmanuel Dufourq, Laurene and Matthew for introducing us to this very interesting field. At the same time, I'd like to thank all the AIMS colleagues with whom we share this historically exciting moment in our scientific career.

REFERENCES

- [1] Lorène Jeantet. *Identification automatique des comportements à partir de bio-loggers: une solution pour améliorer les connaissances sur l'écologie des tortues marines?* PhD thesis, Université de Strasbourg, 2021.
- [2] Lorene Jeantet and Emmanuel Dufourq. Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. *Ecological Informatics*, 77:102256, 2023.
- [3] Lorène Jeantet, Victor Planas-Bielsa, Simon Benhamou, Sebastien Geiger, Jordan Martin, Flora Siegwalt, Pierre Lelong, Julie Gresser, Denis Etienne, Gaëlle Hiélaud, et al. Behavioural inference from signal processing using animal-borne multi-sensor loggers: a novel solution to extend the knowledge of sea turtle ecology. *Royal Society Open Science*, 7(5):200139, 2020.