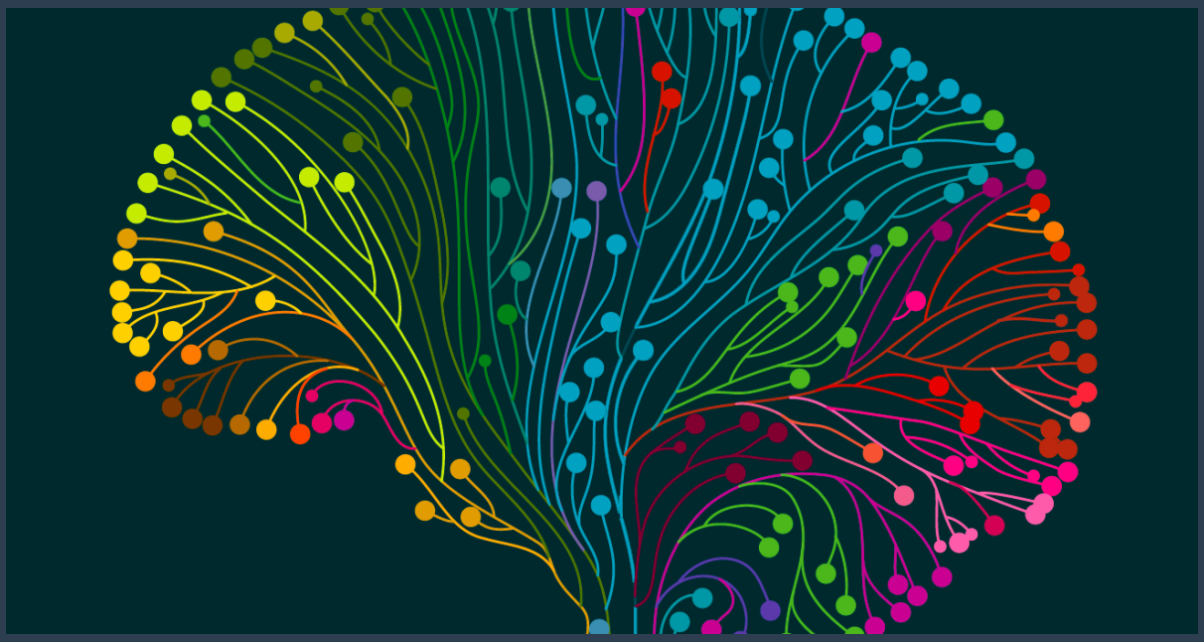
First Report

Sound Detection and Classification
using Spiking Neural Networks



COURREGE Téo
GANDEEL Lo'aï

Date: December 22, 2023

Contents

List of Figures

1 Introduction

Our project addresses the challenge of applying spiking neural networks (SNNs) to audio classification in the field of spiking neural network (SNN) research. This report provides an overview of our initial progress in this area. Our project specifically addresses the problem of audio classification within the broader context of SNNs.

Before delving into the details, we outline key aspects including preprocessing, data manipulation/augmentation, initial model implementations, and a look at preliminary results.

Using audio data primarily from the Google AudioSet database, our work involves preprocessing, which includes conversion of signals to image representations, feature extraction, and consideration of encoding schemes suitable for SNNs. Challenges related to data quality, context, and labeling complexity prompted the exploration of data augmentation strategies to improve model robustness.

In the following sections, we will elaborate on the organizational structure of our project, the technical environment utilized, a detailed account of the work accomplished, challenges faced and solutions implemented, concluding with a reflection on our achievements and future perspectives.

Chat:

In the following sections, this report will delve into the organization that underpin our project. A detailed examination of the technical environment, including the computational tools, software frameworks, programming languages, and relevant documentation, will shed light on the methodological underpinnings of our research.

This is followed by an account of the work performed. The technical and theoretical advances made during the project are highlighted. The challenges faced, ranging from technical hurdles to conceptual complexities, will be discussed in detail, along with the strategic solutions developed to overcome these obstacles. The section will also reflect on the usefulness of the feedback received during the initial stages of the project presentation.

Finally, the report will conclude with a discussion of our achievements, highlighting the areas where significant progress has been made. We will outline our future perspectives, providing a roadmap for the coming weeks and offering insights into the trajectory of our research efforts.

2 The project

2.1 Description of the project - Spiking Neural Networks

2.1.1 Reminder of the Dow

Inspired by the neural signaling patterns of the human brain, SNNs introduce a temporal element into artificial neural networks. This temporal characteristic positions SNNs as promising candidates for real-time processing and pattern recognition tasks.

A Spiking Neural Network is a variant of artificial neural networks designed to more accurately mimic biological neural networks. Unlike traditional artificial neural networks (ANNs) that work with continuously changing time values, SNNs operate with discrete events occurring at defined times. They take a set of spike values as input and produce a set of spike values as output.

The spiking behavior of a neuron in an SNN is modeled by a membrane potential equation. For instance, in a leaky integrate-and-fire (LIF) neuron model, the membrane potential equation is defined by a set of parameters including a time constant (τ), resting potential (u_{r1}), reset potential (u_{r2}), synaptic weights (w_j), and a firing threshold (u_th). The output spike (s) is determined based on the membrane potential (u) and various conditions. This discrete event-based approach distinguishes SNNs from other types of neural networks.[1]

$$\begin{cases} \tau \frac{d\,u(t)}{dt} = -[u(t) - u_{r_1}] + \sum_j w_j \sum_{t_j^k \in S_i^{T_w}} K(t - t_j^k) \\ \begin{cases} s(t) = 1 \quad u(t) = u_{r_2} \text{ if } u(t) \geq u_{th} \\ s(t) = 0 \quad \text{otherwise} \end{cases} \end{cases} \tag{1}$$

This equation was firstly formulated as :

$$\tau \frac{d\,u(t)}{dt} = -[u(t) - u_r] + RI \tag{2}$$

From a mathematical perspective, Equation (2) represents a linear differential equation. Alternatively, an electrical engineer may recognize it as the equation of a leaky integrator or RC-circuit with parallel resistor (R) and capacitor (C). In the realm of neuroscience, this equation is termed the equation of a passive membrane. [2]

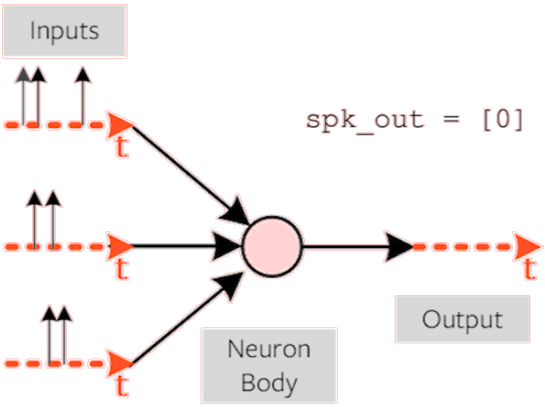


Figure 1: SNN input

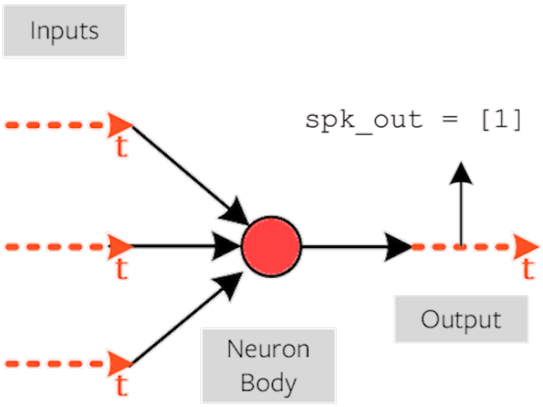


Figure 2: SNN output

2.1.2 Audio classification task

Audio classification is a fundamental problem in the field of audio processing. It involves assigning a label to an audio clip based on its content. The audio classification task can be further divided into two subtasks: sound event detection (SED) and sound event classification (SEC). SED involves detecting the onset and offset times of sound events in an audio clip, while SEC involves assigning a label to each detected sound event.

In the context of audio classification, training an SNN involves working on image representations of audio data and encoding schemes suitable for SNNs.

2.1.3 Objectives of the project

The primary goal of our project is to exploit the temporal processing capabilities of SNNs for audio classification tasks. Specifically, we want to develop models capable of classifying (and possibly detecting) sound events from audio data.

Furthermore, knowing that SNNs consume less power than traditional Artificial Neural Networks (ANNs), but have lower overall accuracy, we want to perform a performance comparison of SNNs with ANNs.

The fullfillment of these objectives would allow us to determine the potential of SNNs for audio classification tasks and to identify the advantages and disadvantages of SNNs compared to other neural networks. Moreover, it is a great way for us to learn more about SNNs and audio classification.

aide

• Décrire le sujet et résumer votre travail • Bien décrire vos contributions et l'intérêt des résultats obtenus

- **Title:** Sound Detection and Classification using Spiking Neural Networks
- **Keywords:** Spiking Neural Networks, Audio Classification, AudioSet, Sound Event Detection, Sound Event Classification
- What is a spiking neural network?
- How it works ?
- Advantages and disadvantages of SNNs compared to other neural networks
- What is the goal of the project?
- Interrest of the project?

3 Organizing the project

3.1 Main tasks

During the last full time period, we worked on:

- **A preprocessing pipeline** that allows us to dowload, format and segment the audio part of the Youtube videos composing the Google Audioset audio files into images. In order to be efficient, the pipeline needed to be parallelized.

  After downloading these, it became also necessary to perform some verification on the data, which includes checking the audio file properties (sample rate, number of channels, etc.) and the labels related to the audio files.

- **Finding some correct data augmentation techniques** that can be used to improve the performance of the SNNs.

- **Finding a way to encode the audio data into spikes** that can be used as input for the SNNs.

- **Implementing the SNNs** that will be used for the audio classification task.

  Training the SNNs on the audio data.

- **Implementing the ANNs** that we would compare to the SNNs.

- **Comparing the performance of the SNNs and the ANNs** on the audio classification task.

3.2 Planning and team organization

3.2.1 Previous work - full time period

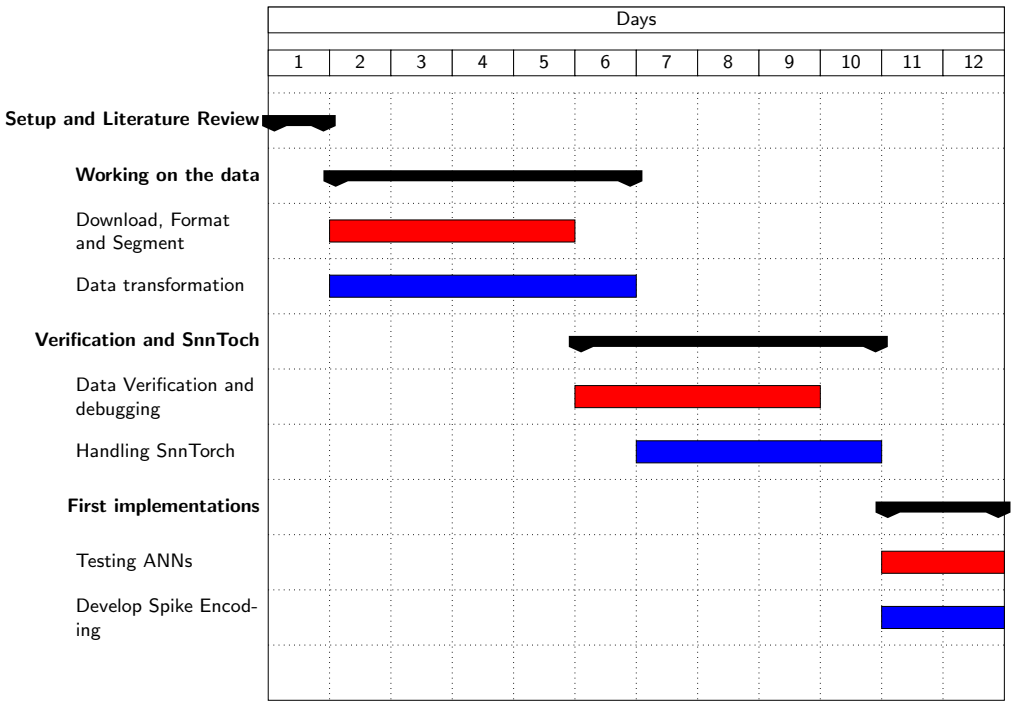


Figure 3: Planning of the full time period

- Téo : blue
- Loaï : red

So far, we have been able to encode and decode the audio data into spikes.

Since the first tasks we had to perform would not be connected until they were finished, we decided to work on them in parallel. This allowed us to be more efficient and, most importantly, to save time.

3.3 Changes in the organization

In the initial planning, there was no task related to the verification of the data nor to the debugging of part of the code. Each subpart of the preprocessing pipeline part takes a lot of time to be implemented and tested. We had to spend some time debugging the code and verifying the data.

aide

• Quelles sont les principales tâches à réaliser ? • Décrire brièvement le planning du projet. • Décrire comment les membres de l'équipe s'organisent pour faire avancer le PER. • Des changements d'organisation ont-ils été nécessaires ?

- Tasks
- Méthode de travail
- Répartition des tâches
- Planning
- Gestion du projet (au fur et à mesure des difficultés rencontrées)

4 Technical environment

4.1 Computational tools

We worked on our personal computers and we used the following computational tools:

Softwares	Visual Studio Code Jupyter Notebook Git Github Google Colab Anaconda Overleaf
Programming languages	Python Latex
Libraries	Pytorch Librosa Numpy Pandas Matplotlib sox Youtube-dl
Frameworks	SnnTorch

Figure 4: Computational tools

(Our Github repository)

4.2 Documentation

4.3 Changing

5 Description du travail réalisé

• Qu'avez-vous produit depuis la fin de la première période à temps plein ? • Quels sont vos avancées techniques et/ou théoriques ? • Quels sont vos résultats ?

5.1 Audio samples

Figure 5: Original Audio

Figure 6: MFCC Reconstruction

Figure 7: Latency Reconstruction

Figure 8: Rate Reconstruction

5.2 Spectrograms, Mel spectrograms, MFCC

- Spectrograms

  Spectrograms are essential graphic tools in audio analysis. They offer a visual representation of the frequency spectrum of a sound signal as a function of time, providing detailed information on the frequency composition and temporal dynamics of an audio signal. This report explores the use of spectrograms in various contexts and highlights their importance in the analysis and understanding of audio signals.

  A spectrogram is generated by applying a short-time Fourier transform (STFT) to an audio signal. This technique divides the signal into small time windows, then applies a Fourier transform to each window to obtain the frequency distribution at that particular moment. The results of these transformations are then represented graphically, using colors to indicate the intensity of frequencies at different periods.

  IMAGE/ FOURIER TRANSFORM EQUATION

  Spectrograms enable in-depth analysis of the temporal characteristics of audio signals. Events such as transients, attacks and decays can be clearly identified, which is essential for understanding dynamic variations in music, speech and other forms of sound.

  By examining the color and intensity of areas in a spectrogram, it's easy to identify the dominant frequencies present in an audio signal. This is particularly useful for detecting anomalies, characterizing musical instruments and separating sound sources.

  Spectrograms are widely used in fields such as professional audio, music and linguistics. In particular, we are interested in Mel spectrograms, as well as MFCCs as input data for our neural networks.

- Mel Spectrograms

  While spectrograms offer a detailed view of the frequency spectrum of an audio signal, a significant evolution in audio analysis occurred with the introduction of Mel Spectrograms.

  These represent an adaptation of traditional spectrograms, using a frequency scale based on the Mel scale, which is more in line with human auditory perception. This transition to Mel Spectrograms has broadened the possibilities of analysis, offering a more faithful representation of the auditory characteristics perceived by the human ear. Let's take a closer look at this innovation and its impact on modern audio analysis.

  While spectrograms offer a detailed view of the frequency spectrum of an audio signal, a significant evolution in audio analysis occurred with the introduction of Mel Spectrograms. These represent an adaptation of traditional spectrograms, using a frequency scale based on the Mel scale, which is more in line with human auditory perception.

  IMAGE

  While Mel Spectrograms have greatly enhanced our ability to represent the frequency spectrum of an audio signal in a way that is more consistent with human perception, Mel Frequency Cepstral Coefficients (MFCC) introduce an additional dimension to audio analysis. MFCCs are an even more advanced transformation, exploiting frequency- and time-domain properties to effectively capture the discriminating characteristics of audio signals.This section explores MFCCs and their central role in the advancement of audio analysis.

- MFCC

  MFCCs are derived directly from Mel Spectrograms and are calculated by applying a discrete cosine transform to the log-power of Mel Spectrograms. This approach captures information

specific to human auditory characteristics while reducing data redundancy.MFCCs thus encapsulate frequency variations over time in a compact way, creating a set of cepstral coefficients that are widely used for automatic speech recognition and other audio signal processing tasks.

Two of the main advantages of MFCCs are compactness and information discrimination. MFCCs condense information while retaining the signal's distinctive characteristics. Compact representation facilitates the storage, transmission and processing of large amounts of data By focusing on perceptual features rather than raw frequency, MFCCs are less sensitive to pitch variations, which improves the robustness of sound recognition. The calculation of MFCCs involves several steps, including Mel scale transformation, calculation of the logarithm of spectral powers, discrete cosine transformation and selection of relevant coefficients.

To summarize, in order to obtain an MFCC from an audio signal, we need to :

- Take the Fourier transform (STFT) of the signal.

- Map the powers of the resulting spectrum onto the mel scale, multiplying it by overlapping window functions (triangular or cosinusoidal).

- Take the logarithm of the amplitudes at each mel frequency.

- Take the discrete cosine transform (DCT) of the list of logarithmic powers of the mel frequencies, as if it were a signal. The MFCCs are the amplitudes of the resulting spectrum.

IMAGE MFCC/ EQUATION DCT/ IMAGE ET EQUATION FONCTIONS FENETRE (EN ANGLAIS WINDOW FUNCTION

5.3 Signal reconstruction

6 Difficultés rencontrées

• Description des difficultés principales rencontrées durant le projet et des solutions mises en place (si cela était possible). • Les commentaires lors de la première soutenance ont-ils été utiles ?

7 Conclusion et perspectives

• Mettre en valeur votre travail et les domaines où vous avez progressé. • Souligner en quoi la formation suivie à Polytech Nice Sophia (ou ailleurs) vous a aidé. • Que comptez-vous faire dans les prochaines semaines ?

References

[1] L. Deng, Y. Wu, X. Hu, L. Liang, Y. Ding, G. Li, G. Zhao, P. Li, and Y. Xie, *Rethinking the performance comparison between SNNS and ANNS*. https://www.sciencedirect.com/science/article/pii/S0893608019302667?ref=pdf_download&fr=RR-2&rr=829f60cc982b9a09, 2020.

[2] R. N. Wulfram Gerstner, Werner M. Kistler and L. Paninski, *Neuronal Dynamics, from single neurons to networks and models of cognition*. https://neuronaldynamics.epfl.ch/index.html, 2014.