Spiking Neural Networks and Audio Classification

Farah Medjahed, Philippe Devienne, Abou El Hassan Benyamina

Spiking Neural Networks and Audio Classification

Farah Medjahed[1], Phillipe Devienne[2] and Abou El Hassan Benyamina[1]

(medjahed.farah@edu.univ-oran1.dz, Philippe.Devienne@univ-lille.fr, benyamina.hassen@univ-oran1.dz)

1. Computer Science Department, Oran1 University, Ahmed Ben Bella, Oran, Algeria

2. Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL, F-59000 Lille, France

Keyworks : Automatic Sound Classification, Spiking Neural Networks, Natural computing.

I Introduction

For many years, artificial neural networks have attracted the attention of researchers, they are used in various fields: image processing, signal processing, handwriting recognition, facial recognition. Neural networks are widely used in sound recognition which has become a hot topic of great interest. In this article, we will present in the first part the birth of artificial neural networks passing from the biological aspect to the mathematical aspect and we will mention the different types of neural networks and learning rules that exist. In the second part we will elaborate the audio processing, by presenting the different techniques of audio representation and the main methods used in audio classification. Finally, we will conclude by citing the current work which is in progress which is related to audio classification using Spiking Neural Networks.

II Spiking neural network

Different types of neural network exist we can cite: MLP (Multi-Layer Perception) which consists of an input layer, an output layer and at least one intermediate layer called the hidden layer where each neuron of a layer is connected with all the neurons of the next layer (except the neurons of the output layer). CNN (Convolutional Neural Network), mainly used in image processing, consists also of a succession of layers: an input layer, an output layer and hidden layers made up of numerous convolutional layers, Pooling layers, ReLU (Rectified Linear Unit) correction layers and layers fully connected. Unlike Multilayer Perceptron and Convolutional Neural Network process data of fixed size and go through a fixed number of layers and computational steps and give outputs of fixed sizes, the Recurrent Neural Network (RNN) [1] manipulates variable-sized entries. A traditional recurrent neural network consists of an input layer, an output layer, and a recurring layer. We have also SNN (Spiking Neural Network) which is the type of neural network that mimics the brain the most because it applies the actual functioning of the neuron. In a biological neuron, a pulse is generated when the sum of the changes in the potential of the presynaptic membrane exceeds the threshold.

III Audio Classification

An audio signal is a signal that contains information in the audible frequency range [2]. Audio classification is divided into two parts the front end and the back end [3]. The front end is the part that treats the input to extract features from the audio and the back end is responsible for the classification and the prediction of the output. The audio signal contains information which identifies the sound. Audio representation is responsible for the extraction of these information or features which represents the audio signal. Many feature extraction techniques are used we can find: MFCC (Mel-Frequency Cepstral Coefficient) [4] which is the most used feature extraction technique for audio processing tasks. MFCC consists of a succession of operations. ZCR [5] of an audio frame is the rate at which the signal's sign changes during the frame. In other words, it's the number of times the signal's value changes from positive to negative and back, divided by the frame's length. LPC is used to compress audio. It is the most widely used method in speech recognition. The basic idea of linear prediction is the use of a linear combination of the past time-domain samples to predict the current time- domain sample [6]. We have also The Mel spectrogram [7] which is the representation of the audio in the form of time and frequency. For the audio classification we have GMM (Gaussian Mixture Models) [8] a probabilistic model. The basis for using GMM in audio classification is that the distribution of feature vectors extracted from a class can be modeled by a mixture of Gaussian density. SVM (Support Vector Machines) [9] transform data into a high-dimensional space, this converts the classification problem into a simpler one which can use linear discriminant functions. HMM (Hidden Markov Models) are widely used classification models in speech recognition [10]. HMM is a finite set of states, each of which has a probability distribution associated with it. A collection of probabilities known as transition probabilities governs transitions between states. According to the corresponding probability distribution, an outcome or observation can be generated in a specific state. Only the outcome is known and the underlying state sequence is obscured [4]. Several works have used the Convolutional Neural network in the context of sound classification. CNN is commonly used in image classification and it has improved the performance of the classification in this domain. Dealing with audio input, lead to transform a sound classification problem into an image classification problem. The idea is to use CNN for the classification by transforming the audio into spectrograms and to use these later as an input of a Convolutional Neural Network.

IV Work in progress

The natural world is analog and yet most of the sensors, with which we observe and monitor the real-world data, use discrete quantities. To avoid as much as possible approximation and heavy latency due to digital conversion, our challenge is to copy biological systems and neurological processes. The first step was to design a bioinspired analog cochlea at Lille that we are interfacing with embedded AI algorithms (implemented on low power neuroprocessors) to mimics the cochlea and the audio cortex.

We opt for the use of Spiking Neuron Network in the sound detection for its advantages. SNN are well adapted to processing spatio-temporal event-based data from neuromorphic sensors, which are power efficient. SNN are highly computationally and energy-efficient model it can be exploited in a neuromorphic hardware device. Trying to add something new in this domain, we are working on sound detection using spiking neural networks, the audio data will be converted into impulses using the artificial cochlea designed at Lille then this later will be used as inputs into a spiking neural network for sound detection. the following figure summarize our work.

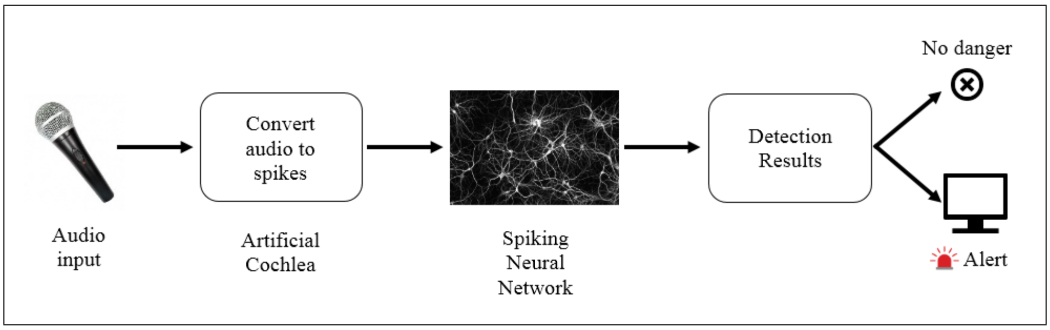

Figure 1: Proposed scheme of sound detection

Our domain of application is sound detection agricultural field related to WATERMED project. The objective of WATERMED 4.0 is to develop and to apply an integrated decision support system based on the Internet of Things, for managing the whole water cycle in agriculture, monitoring water resources (conventional and non-conventional) and water demands including the measure of economic, energy, social and governance factors that influence the water use efficiency in Mediterranean agricultural production areas. The objective of this application is to detect troops of wild boars in order to intercept them or divert them from agricultural fields in time as application results, if there are any wild boars that are detected, an alert is raised and from the strength of the sound, the wild boars can be located.

Training SNNs requires many labeled data that are expensive to obtain in real-world applications, as traditional artificial neural networks (ANNs). In order to address this issue, transfer learning has been proposed and widely used in traditional ANNs, and only very recently to SNNs. We have developed CNN audio neural networks on Nengo which can run SNNs too on different hardware devices (Spinnaker, Loihi, ...). This experimentation is in progress, the first results are encouraging.

V Conclusion

Everyone knows the paradox to build both small and intelligent sensor since the most active research in AI is based on deep convolutional neural networks which are very time-consuming (up to several weeks with GPU servers) and yield enormous energy consumption, despite most recent innovations in parallel digital architectures. However, the brain and biological systems in general, can perform high-performance calculations with much higher efficiency than our most powerful computers, and they do it very quickly and with very low energy consumption.

In parallel, recent works have been published which revolve around audio classification, this subject which has become an interesting subject which pushes researchers to improve classification techniques for much better performance.

In this short paper, we attempted to give an overview of neural networks and the most used classification techniques and give an idea of our work in progress related to audio classification using spiking neural networks. We try to imitate mammal hearing through neuromorphic implementations to design ultra-low-power acoustic and autonomous sensors that could detect very specific events of interest (for instance wild boars) and produce alerts with some associated spatial info. This work is in part supported by the WATERMED 4.0.

References

[1] Alexander Jaffe. Long short-term memory recurrent neural networks for classification of acute hypotensive episodes. 2017.

[2] Christian S. Jensen, Richard T. Snodgrass (auth.), LING LIU, and M. TAMER ÖZSU (eds.). *Encyclopedia of Database Systems*. Springer US, 2009.

[3] Jordi Pons and Xavier Serra. Randomly weighted cnns for (music) audio classification. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 336–340, 2019.

[4] K.S. Rao and Manjunath K.E. Speech recognition using articulatory and excitation source features. *SpringerBriefs in Speech Technology*, 2019.

[5] Theodoros Giannakopoulos and Aggelos Pikrakis. *Introduction to Audio Analysis: A MATLAB Approach*. Academic Press, Inc., USA, 1st edition, 2014.

[6] Li Deng and Douglas O Shaughnessy. *Speech Processing A Dynamic and Optimization Oriented Approach*. CRC Press, 2003.

[7] Mushtaq, Zohaib, Su, and Shun-Feng. Efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques for spectrogram images. *Symmetry*, 12(11), 2020.

[8] P. Dhanalakshmi, Sengottayan Palanivel, and Vivekanandan Ramalingam. Classification of audio signals using aann and gmm. *Applied Soft Computing*, 11:716–723, 01 2011.

[9] Andrey Temko and Climent Nadeu. Classification of acoustic events using svm-based clustering schemes. *Pattern Recogn.*, 39(4):682–694, April 2006.

[10] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.