Audio Spectrogram Recognition with Neural Networks

Farah Medjahed, Philippe Devienne, Hassan Benyamina, Mohammed Kamel Benhaoua

Proceedings OLA'2022 International Conference on Optimization and Learning

ola 2022

18-20 July 2022, Syracusa, Italia

Organizers

UNIVERSITÀ degli STUDI di CATANIA

Université de Lille

Audio Spectrogram Recognition with Neural Networks

Farah Medjahed[1], Philippe Devienne[2], Abou El Hassan Benyamina[1] and Mohammed Kamel Benhaoua[3]

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

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

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

3. University of Mascara - Department of Computer Science, Algeria

Keyworks : Automatic Sound Classification, Convolutional Neural Network (CNN), Spiking Neural Networks (SNN), Natural Computing.

This work is in part supported by the European Prima S1\WATERMED 4.0 project for agriculture in semi-arid areas and the French ANR-21-CE04-0020 ULP COCHLEA for nano acoustic sensors and biodiversity.

I Introduction

Artificial neural networks interest many researchers for recent years, they are employed in a variety of sectors, including image processing, signal processing, handwriting recognition, and facial recognition. Sound recognition is a popular application of neural networks, which has sparked a lot of interest.

In this article, we will elaborate in the first part the audio classification, by presenting the different techniques of audio classification that are used. In the second part, we will present the Convolutional Neural Network and the Spiking Neural Network. Finally, we will conclude by citing the current work which is in progress related to audio classification using both these two types of neural networks.

II Audio classification

Audio classification is a hot topic that occupies the attention of researchers as it is used in different applications in different fields : telecommunication, robotics, marine and agricultural fields as well as other fields. Audio classification consists of automatically determining the nature of the sound signal. Several works have addressed this subject and different classification techniques have been used. Dhanalakshm and al. [1] have used the GMM (Gaussian Mixture Models) method as a method of classification which is 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 densities. Temko and al. [2] worked on SVMs (Support Vector Machines) which transform

data into a high-dimensional space, this convert the classification problem into a simpler one which can use linear discriminant functions. L.R. Rabiner and al.[3] used the HMM method in their work which is widely used classification models in speech recognition. 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. However, recent works on audio classification use Convolutional Neural Networks [[4],[5], [6], [7]...]. The advantage of this technique is that it can classify a large amount of data and increase the accuracy for image classification.

CNN (Convolutional Neural Network) is a feed-forward neural network which is able to classify images based on feature extraction. CNN is composed of three main layers : convolution and pooling layers, which perform feature extraction, and a fully connected layer, which maps the extracted features into the appropriate class. Several CNN models have been proposed each of them has a specific architecture but their purpose is the same is to increase the classification accuracy.[8]

SNN (Spiking Neural Network) [9] is the type of neural network that mimics the brain the most, it is more biologically plausible than other traditional neural networks 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 Work in progress

First, our work is to do the audio classification using convolutional neural networks by transforming the audio signal into spectrograms and use these later as an input of a Convolutional Neural Network to do the classification. Second, the same classification is made using Spiking Neural Networks based on the CNN-SNN conversion method named Spkeras which was proposed by Dengyu Wu et al. [5] which consists on detecting the activation layer in CNN to create SpikeActivation layer.

The dataset that we use include 20 animals and instruments sound [10]. This dataset is constructed using Animal Sound Data and Instrument Data. Each audio is split into multiple samples. These samples are divided into three parts: Train, Validation and Test sets which are disjoint. The train set is composed of 16,636 samples, the validation set contains 3,249 samples and the test set contains 3,727 samples. The code of SpKeras uses Tensorflow and Keras.

This work consists on training of the train set using a Convolutional Neural Network then, we use SpKeras to convert CNN into SNN after that we evaluate SNN model.

The objective is to prove the utility of using an Spiking Neural network than a Convolutional Neural Network, in addition, SNN are highly computationally and energy-efficient model and it can be exploited in a neuromorphic hardware device such as SpiNNaker. The following table presents the results of the experiments which shows the change in accuracy related to the number of epochs.

Number of epochs	CNN accuracy	SNN accuracy
50	80%	71%
100	77%	73%
200	79%	75%
300	80%	77%

IV Conclusion

In this brief study, we sought to provide an overview of audio classification and the most used classification techniques that exist and introduction of Convolutional and spiking neural network, as well as a glimpse into our ongoing work on audio classification using convolutional and spiking neural networks. As a complementary work, we will use a neuromorphic hardware in order to have an ultra-low power acoustic classifier and for better performance.

References

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

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

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

[4] Heart sound classification based on improved mfcc features and convolutional recurrent neural networks. *Neural Networks*, 130:22–32, 2020.

[5] Xiaowei Huang Dengyu Wu, Xinping Yi. A little energy goes a long way: Energy-efficient, accurate conversion from convolutional neural networks to spiking neural networks. 6 Mar 2021.

[6] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, and Kevin Wilson. Cnn architectures for large-scale audio classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017.

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

[8] Shouheng Peng Zewen Li, Wenjie Yang and Fan Liu. A survey of convolutional neural networks: Analysis, applications, and prospects. 1 Apr 2020.

[9] Hammouda Elbez, Benhaoua Kamel, Philippe Devienne, and Pierre Boulet. Vs2n : Interactive dynamic visualization and analysis tool for spiking neural networks. pages 1–6, 06 2021.

[10] Cheng-Hao Tu Yi-Ming Chan. Sound20. `https://github.com/ivclab/Sound20`, 2018.