RawNET

Machine Learning for Speaker Recognition
2018/2019

1 Project Description

The aim of the project is to reproduce the results obtained in the paper [1]. This system is an end-to-end speaker recognition tool that is trained directly from raw audio signals. An implementation of this network is provided in Keras (TensorFlow) and the goal of this project is to analyse this code and provide a Pytorch documented implementation of this network.

This network is composed of different types of blocks including

— Convolutional layers
— Gated recurrent unit layers
— Fully connected feed-forward layers
— Residual blocks

The implementation should be done step-by-step, starting with a simplified version and comparing the convergence and behavior of the network after adding each component. In order to start, the loss function will be just considering a cross-entropy function before including

2 Data Description

The dataset is Voxceleb 1, a free corpus of audio signal collected from the web `http://www.robots.ox.ac.uk/~vgg/data/voxceleb/vox1.html` It contains recording from 1,251 speakers divided into a training (1,211 speakers) and testing (40 speakers) set. Pre-processing of the audio signal is described in [1] and the source code is provided to prepare the data before training.

3 Evaluation

In na first step the system will only be evaluated regarding the convergence of the network and the accuracy of the cross-entropy loss function. If time is enough to complete the implementation, then the system will be evaluated on speaker verification task.

4 Project Roadmap

1. Read the RawNET description in the paper
2. Clone the GIT repository containing KERAS source code
3. Convert KERAS source code into Pytorch and validate step by step
4. Run the complete model on the Voxceleb speaker verification task
5. Prepare the final defense

Références

[1] Jee-weon Jung, Hee-Soo Heo, Ju-ho Kim, Hye-jin Shim, and Ha-Jin Yu. Rawnet : Advanced end-to-end deep neural network using raw waveforms for text-independent speaker verification. *INTERSPEECH*, 2019.