Deep Quaternion Neural Networks for 3D Sound Source Localization and Detection

Project of Neural Network Course

Sveva Pepe 1743997
Marco Pennese 1749223
Claudia Medaglia 1758095

Abstract

We work with 3D audio sounds, in particular we analyze the sound event localization and detection (SELD).
We were provided with a dataset containing the sounds that were recorded with the first-order Ambisonics microphone. These sounds are then represented using spherical harmonics decomposition in the quaternion domain, to then be passed to the neural network which will be towed in order to obtain the best possible results in output.
The neural network is quaternion convolutional, with the addition of some recurrent layers. The aim of the project is to detect the temporal activities of a known set of sound event classes and to further locate them in space by using quaternion-valued data processing.

1 Introduction

Talk about SELD, which is made up of DOA and SED. Say what I am in a general way, don't go into too much detail, don't go too far [my advice]. (they are well written in the last paper sent by our friend cominiello). To say that with quaternions the performances are better than without. Then do you. At the end of the introduction, say what the next sections are made up of.

2 Quaternion domain

What are quaternions, and their connection with 3D audio recorded with Ambisonic. Enter the mathematical formula on the composition (that of the real part + imaginary part). A minimum of considerations on active and reactive intensity, in particular the role of active and reactive intensity in DOA.
Formulas of the two "input features" with quaternions.

3 Network Structure

Explain the architecture, then start with the 3 QCNN, what is a convolutional network with quaternions (weights, Hamilton product etc.), which does BatchNormalization (to be specified that this is not based on quaternions), the activation functions, what is MaxPooling (just explain what it does) and conclude with the Dropout (technique that serves blah blah blah).
Then explain the 2 QRNN recurrent layers and then what a recurrent layer is and the relationship with quaternions, very short.
Finally say that the output of the network is doubled, making it possible to do both detection and localization. Explain the outputs with both two fully-connected layers (remember to specify the activation functions to identify that the two "outputs" are different).
I would put image of the network.
We can also specify the dimensions of the fetures after the QCNN, the QRNN and the FC.

4 Dataset

The dataset, what it is made of, and how it was divided by us. How many files each subfolder is made up of.

5   Metrics

The metrics used, therefore the average, SELD SCORE, the confidence interval and I do not know whether to consider the F1 score as well. Obviously with formulas

6   Experiments

Our experiments, then the changes we made, the results that come to us, the various graphs. Let's see what comes out of it, and in case I would make comparisons with the results that have been obtained from the various papers that have provided us.

7   Conclusion

The conclusions regarding the project carried out and you have had results based on the metrics that we used.