

Homework # 14

Isaac Ayala Lozano

April 6, 2020

The authors proposed a new strategy to determine the location of a system in a 3D environment employing a Convolutional Neural Network trained with semi-synthetic data and then fine-tuned with real data. A comparison was performed for changes in the training strategy and tested against other known algorithms. Results indicate that the CNN outperformed the rest by a slight margin.

From the article, the following are identified:

Proposal Use of a Convolutional Neural Network (CNN) to estimate the three dimensional position of an acoustic source using the raw audio signal as the input information.

Challenges 1) Estimation of the regression function cannot be obtained analytically due to random noise and distortion. 2) Amount of available data for training of the CNN is limited. 3) Generation of training data must consider signal propagation and acoustic noise conditions of the room, as well as recording process conditions.

Techniques used 1) Use of Deep Learning to obtain a regression function directly from the CNN. 2) Generation of semi-synthetic data using Gaussian white noise as well as Discrete Fourier Transforms, 3) Use of mean-squared error for loss function in CNN training.

Experiments Experiment 1: Use of a single sequence of sounds for fine tuning. Experiment 2: Comparison of fine-tuning with semi-synthetic data for CNN training versus use of raw sound data for CNN training. Experiment 3: Add another fine-tuning sequence to the CNN training. Experiment 4: Use of static speaker sequences in training and refinement.

Results and discussion The resulting Convolutional Neural Network proved to surpass other algorithms designed to determine the 3D location of sound sources based on the test data used in the study like SRP and GMBF. This was proven to be true for tests where the CNN had been training in a two-step process using semi-synthetic data and fine-tuning employing real data. The CNN was also more robust to changes in the origin of the sound, like gender and height.

Extrapolation to the content viewed in class

It is possible to generate synthetic sets of test data employing a few samples of real data and multiple distributions to generate variations in the signals. This permits the creation of new data using the small set of original data, hence making the training set for the CNN much bigger. This is evidenced in the paper, as the authors themselves used a small set of raw real data to create a bigger training set for the CNN using Gaussian white noise.