AUTHOR GUIDELINES FOR MLSP PROCEEDINGS MANUSCRIPTS

Author(s) Name(s) omitted for double blind review

Author Affiliation(s) omitted for double blind review

ABSTRACT

In this paper we propose a convolutive and recurrent neural network based extensions to autoencoders for source separation.

***Index Terms*—** One, two, three, four, five

1. INTRODUCTION

Talk about what we are doing and why it is interesting.

Give some background.

2. AUTOENCODERS

2.1. Feed-forward Autoencoder

Define standard autoencoder. Provide a toy example to illustrate the shortcomings of this.

2.2. CNN-CNN Autoencoder

The approximation $\hat{X}$ for a given spectrogram X is computed as follows:

$$\begin{aligned}\hat{H}(k,t) =& \sigma_1 \left(\sum_{f,t'} X(f,t-t') F_e(f,t',k) \right) \\ \hat{X}(f,t) =& \sigma_2 \left(\sum_{k} \sum_{t'} \hat{H}(k,t-t') F_d(f,t',k) \right) \end{aligned} \tag{1}$$

2.3. Multilayer of CNN-CNN Autoencoder

This loses the interpretability, but maybe good for accuracy?

2.4. RNN-CNN Autoencoder

This is the same as CNN-CNN case, except the computation of the activations H. In the CNN encoder, each filter was of finite length. With RNN-CNN version, we are attempting to use an infinite length filter. The computation of $\hat{H}$ is as follows:

$$\begin{aligned} Z(:,k,t) =& \sigma(WZ(:,k,t-1) + UX(:,t)) \\ \hat{H}(k,t) =& \sum_f Z(f,k,t) \end{aligned} \tag{2}$$

Give a toy example which shows what this model can do that CNN-CNN can not.

3. EXPERIMENTS

Try each model in a given K range in speech-speech source separation task.

4. REFERENCES

Thanks to XYZ agency for funding.