

Denoising and Improving Speech Recording using Deep Learning

Hannes Stärk

October 9, 2019

Abstract

Hello.

Contents

1	Introduction	3
1.1	Loss Functions	3

1 Introduction

A paper that offers an overview of the supervised separation of speech from background noise (Wang et al. 2017).

Speech Enhancement GAN (SEGAN) (Pascual et al. 2017) is probably a good approach.

1.1 Loss Functions

MSE makes little sense since a shift in phase should not be regarded as difference. A better idea may be differentiable dynamic time warping distance (Arjovsky et al. 2017), or earth mover’s distance (Cuturi et al. 2017) according to Purwins et al. (2019).

Literature

- Arjovsky, Martin et al. (June 2017). “Wasserstein Generative Adversarial Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup et al. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, pp. 214–223.
- Cuturi, Marco et al. (June 2017). “Soft-DTW: a Differentiable Loss Function for Time-Series”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup et al. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, pp. 894–903.
- Pascual, Santiago et al. (2017). “SEGAN: Speech Enhancement Generative Adversarial Network”. In: *CoRR* abs/1703.09452. arXiv: 1703.09452.
- Purwins, Hendrik et al. (2019). “Deep Learning for Audio Signal Processing”. In: *CoRR* abs/1905.00078. arXiv: 1905.00078.
- Wang, DeLiang et al. (2017). “Supervised Speech Separation Based on Deep Learning: An Overview”. In: *CoRR* abs/1708.07524. arXiv: 1708.07524.