MDX
Music Demixing Workshop

Feature-informed Latent Space Regularization for Music Source Separation

Yun-Ning Hung[1] **and Alexander Lerch**[1]

1 Center for Music Technology, Georgia Institute of Technology, USA

In partnership with

ISMIR2021

Abstract

The integration of additional side information to improve music source separation has been investigated numerous times, e.g., by adding features to the input (Slizovskaia et al., 2021) or by adding learning targets in a multi-task learning scenario (Hung & Lerch, 2020). These approaches, however, require additional annotations such as musical scores, instrument labels, etc. in training and possibly during inference. The available datasets for source separation do not usually provide these additional annotations (Rafii et al., 2017). In this work, we explore self-supervised learning strategies utilizing VGGish features (Hershey et al., 2017) for latent space regularization; these features are known to be a very condensed representation of audio content and have been successfully used in many MIR tasks. We introduce three approaches to incorporate the features with a state-of-the-art source separation model, and investigate their impact on the separation result.

References

[1] Slizovskaia, et al. "Conditioned Source Separation for Musical Instrument Performances." IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 2083-2095.

[2] Hung, et al. "Multitask learning for instrument activation aware music source separation." ISMIR, 2021.

[3] Rafii, et al. "MUSDB18-a corpus for music separation." 2017.

[4] Hershey, et al. "CNN architectures for large-scale audio classification." ICASSP, 2017.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., & others. (2017). CNN architectures for large-scale audio classification. *2017 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 131–135.

Hung, Y.-N., & Lerch, A. (2020). Multitask learning for instrument activation aware music source separation. *ISMIR*.

Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., & Bittner, R. (2017). *MUSDB18-a corpus for music separation*.

Slizovskaia, O., Haro, G., & Gómez, E. (2021). Conditioned source separation for musical instrument performances. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 2083–2095.