

User-Guided One-Shot Deep Model Adaptation for Music Source Separation

Giorgia Cantisani¹, Alexey Ozerov², Slim Essid¹, and Gaël Richard¹

Arxiv DOI: [hal-03219350](https://arxiv.org/abs/2105.08111)

¹ LTCI, Télécom Paris, Institut Polytechnique de Paris, France ² InterDigital R&D France, France

Abstract

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

In partnership with



Music source separation is the task of isolating individual instruments which are mixed in a musical piece. This task is particularly challenging, and even state-of-the-art models can hardly generalize to unseen test data with significant timbral variation compared to training. Nevertheless, prior knowledge about individual sources can be used to better adapt a generic source separation model to the observed signal. In this work, we propose to exploit a temporal segmentation provided by the user that indicates when each instrument is active to fine-tune a pre-trained deep model for source separation and adapt it to one specific mixture. This approach can be referred to as a one-shot adaptation, as it acts on the target song instance only using the side information available only at test time. Our results are promising and show that state-of-the-art source separation models have large margins of improvement, especially for instruments underrepresented in the training data.