open.unmix.app - towards audio separation on the edge

Fabian-Robert Stöter
Inria and LIRMM, University of Montpellier, France
fabian-robert.stoter@inria.fr

Maria Clara Machry
Universidade Federal do Rio Grande do Sul
mcmachry@gmail.com

Delton de Andrade Vaz
Universidade Federal do Rio Grande do Sul
delton.vaz@gmail.com

Stefan Uhlich
Sony Europe B.V., Germany
stefan.uhlich@sony.com

Yuki Mitsufuji
Sony Corporation, Japan
Yuhki.Mitsufuji@sony.com

Antoine Liutkus
Inria and LIRMM, University of Montpellier, France
antoine.liutkus@inria.fr

ABSTRACT

Audio separation has recently become a mature technology, with state of the art reaching performance that permits wide audience application. While recent research continues to develop sophisticated methods to make results even better, we present our recent work in bringing the reference open-unmix model to the edge. This results in open.unmix.app, a web-based platform for audio separation. We demonstrate two contributions: First, we report on the difficulties of bringing a state-of-the-art deep learning-based model to javascript using TensorFlow.js including all necessary pre- and postprocessing. Solving them enables performing separation and enhancement of music and speech signals directly in the browser. Second, we present a platform that exploits this web-separation technology and allows users to share separated multitrack audio on the web using a custom, embeddable multi-track player.

1. DESCRIPTION

Audio source separation deals with the problem of unmixing a given audio signal into its original sources. For instance, separating music means recovering vocals, bass, drums, etc. from the stereo mixture. Several separation scenarios may be identified, depending on the number of sources or whether they are of the same kind as in speech separation, or of a different kind, which is typical in music or in denoising. In all cases, audio separation has been an interesting and challenging task for researchers with many contributions, notably in the last two decades. With the advent of deep learning, many new systems were proposed which significantly improved previous state-of-the-art, that was mostly model-based rather than trained on data. For a recent overview of separation methods we refer to [4, 1, 6]. For these new learning-based models, access to training data is crucial, because they have to include the ground truth reference sources which are usually copyrighted.

In the last two years, open-source systems such as [7, 3, 2] were released that immediately caught attention also outside of academic research. These systems are trained by python-based frameworks, which makes client-side inference difficult to non-expert users. In this demo, we showcase our effort to develop a web-based separation platform that brings AI-based audio models to the browser. This platform provides not only separation capabilities, but also a convenient way to share separation results. Remarkably, web-based separation has already been proposed in [5]. However, the separation model there is very simplified and cannot be considered state of the art anymore. Furthermore, our proposed platform additionally offers a natural way to share results through an embeddable multi-track player that combines statically hosted stems under a simple shortened link.

For the first time, our demo allows state-of-the-art models to be be used in the browser, without sacrificing separation quality.

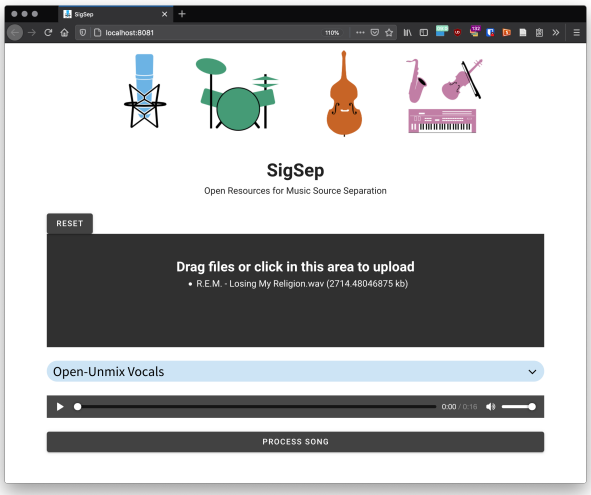

Figure 1: Screen capture of browser based separation interface with the open-unmix model selected.

1.1 Source Separation

In this demo [1], we showcase three audio separation models: two music separation models, and a speech denoising

[1] https://sigsep.github.io/open-unmix-js (temporary link for review)

Web Audio Conference WAC-2021, July 5–7, 2021, Barcelona, Spain.

model. The music separation models are derived from open-unmix [7] and spleeter [2], which are both seen as reference methods within and outside the research community. Furthermore, open-unmix was also trained for speech separation as shown in [8], which is the third model being showcased. While spleeter was trained in tensorflow, open-unmix variants were trained on PyTorch. The open-unmix models were converted to tensorflow via ONNX [2] and on-the-wire weight quantization was applied to reduce the models file size. In all three cases, the models are spectrogram-based, which means that they perform separation in the magnitude short-term-Fourier transform domain ($y = model(|STFT(mixture)|)$) and output separated magnitude spectrogram. Going back to the waveform domain requires manipulating complex numbers, which is still not implemented for all necessary operators (such as "tf.exp") in tensorflow.js. For this reason, we manually converted the python-based pre-and post-processing pipelines from Python to JavaScript. Thus, as part of the release of this demo, we make a fully invertible tensorflow.js-based STFT/ISTFT pipeline available. Which is useful for other web-audio projects operating in the spectral domain, too.

1.2 Sharing Results

With the rise of commercial source separation products, users often want to share their separation results with other users on social media platforms. However, platforms such as Youtube, Soundcloud, Twitter, Facebook only allow single stream audio content which is why users used non-optimal work-arounds such as:

- uploading the separated stems individually, sharing multiple links.
- producing original video content where streams can be soloed or muted, often screencasted from a DAW session.
- host their own website that uses a multi-track audio player such as [9].

While the latter is often the most favorable solution to assess the separation quality, users often go with the first since its the simplest way. In this demo, we present our efforts to help users bridge this gap and offer a multi-track audio player, that is able to mix audio stems on the fly, provided they are already hosted on static cloud storage such as Dropbox. For this purpose, we aggregate the stem URLs in a Firestore database and offer a shareable link through a Vue application.

The share platform can be accessed via [3].

2. TECHNICAL REQUIREMENTS

open.unmix.app can be demonstrated with a low-cost setup and runs fine in a mobile or desktop browser. However, for larger tracks, more RAM is advised. To assess the quality, we advise demo visitors to use headphones.

3. ACKNOWLEDGMENTS

This work was partly supported by the research programme KAMoulox (ANR-15-CE38-0003-01) funded by ANR, the French State agency for research. Thank to KoeKestra.[4]

[2]Souce Code will be made available.

[3]https://sigsep.github.io/share/ (temporarily link for review)

[4]https://koekestra.com/spleeter_js

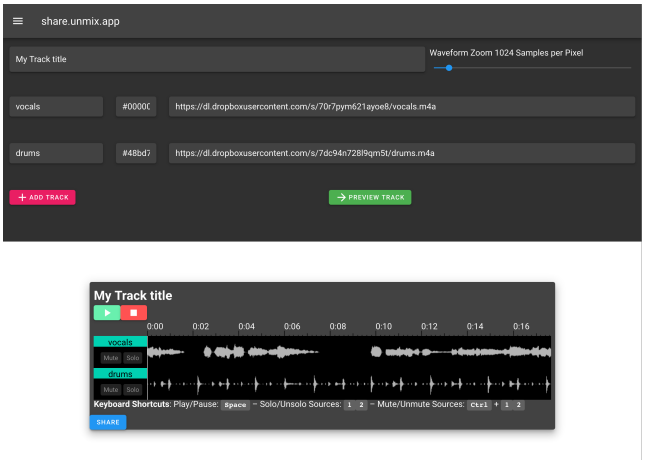

Figure 2: Screenshot of share platform that allows to share multitrack files on the web.

4. REFERENCES

[1] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F. Stöter. Musical source separation: An introduction. *IEEE Signal Processing Magazine*, 36(1):31–40, Jan. 2019.

[2] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam. Spleeter: A fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, 5(50):2154, June 2020.

[3] M. Pariente, S. Cornell, J. Cosentino, S. Sivasankaran, E. Tzinis, J. Heitkaemper, M. Olvera, F.-R. Stöter, M. Hu, J. M. Martín-Doñas, D. Ditter, A. Frank, A. Deleforge, and E. Vincent. Asteroid: The PyTorch-based audio source separation toolkit for researchers. *arXiv:2005.04132 [cs, eess]*, May 2020.

[4] Z. Rafii, A. Liutkus, F. R. Stöter, S. I. Mimilakis, D. FitzGerald, and B. Pardo. An overview of lead and accompaniment separation in music. 26(8):1307–1335, Aug. 2018.

[5] G. Roma, A. J. R. Simpson, E. M. Grais, and M. D. Plumbley. Remixing musical audio on the web using source separation. In *Proceedings of the 2nd Web Audio Conference (WAC-2016)*, Atlanta, Georgia., Apr. 2016.

[6] F.-R. Stöter, A. Liutkus, and N. Ito. The 2018 signal separation evaluation campaign. In *Latent Variable Analysis and Signal Separation: 14th International Conference, LVA/ICA 2018, Surrey, UK*, pages 293–305, 2018.

[7] F.-R. Stöter, S. Uhlich, A. Liutkus, and Y. Mitsufuji. Open-Unmix - A Reference Implementation for Music Source Separation. *Journal of Open Source Software*, 4(41):1667, Sept. 2019.

[8] S. Uhlich and Y. Mitsufuji. Open-unmix for speech enhancement (umx se), May 2020.

[9] N. Werner, S. Balke, F.-R. Stöter, M. Müller, and B. Edler. Trackswitch.js: A Versatile Web-Based Audio Player for Presenting Scientifc Results. In *3rd Web Audio Conference, London, UK*, 2017.