An Overview of the Machine Learning Algorithms Used for Music Source Separation

Aidas Žygas, Gražina Korvel

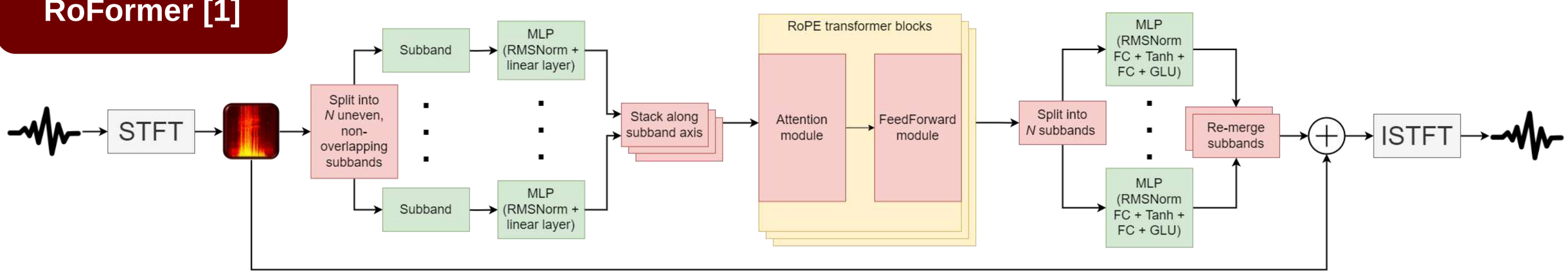
Vilnius university, Faculty of Mathematics and Informatics

Introduction

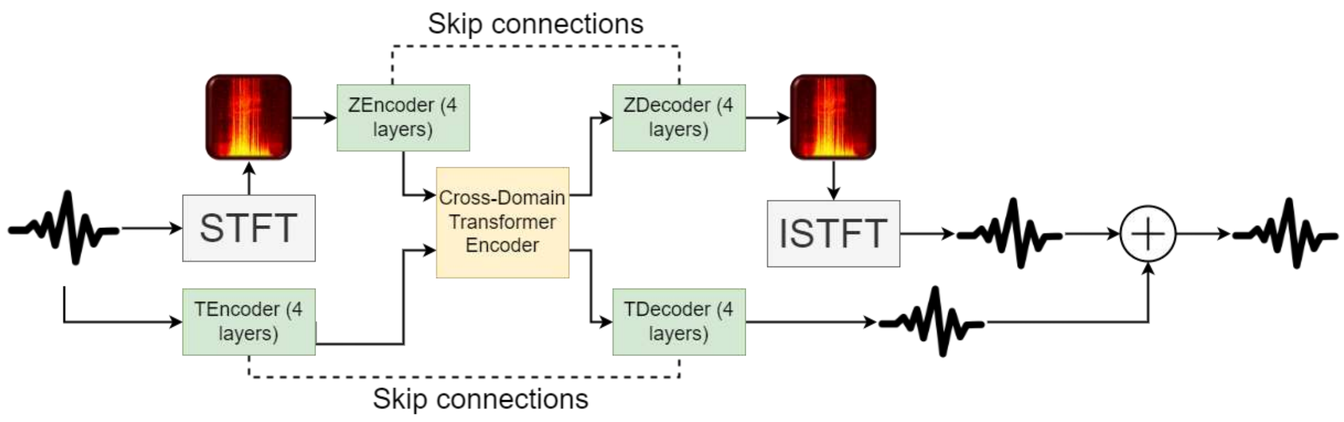
Music source separation involves isolating individual audio sources from a mixed signal, enabling enhanced manipulation and analysis of musical elements. Although there are countless different music source separation models available, most of them utilize modified versions of machine learning algorithms, including 1) Convolutional Neural Networks (CNN), 2) Recurrent Neural Networks (RNN), and 3) Attention-Based Transformers. This work presents and discusses the results of a survey on music source separation approaches. The main purpose of this analysis is to understand the underlying logic of each algorithm and how they differ from each other. Second, it is crucial to compare the performance of the models that adapt machine learning algorithms using a quantitative metric – SDR (Signal to Distortion Ratio) in order to see which machine learning algorithm is the most efficient for separating stems in a musical recording.

Leading model architectures

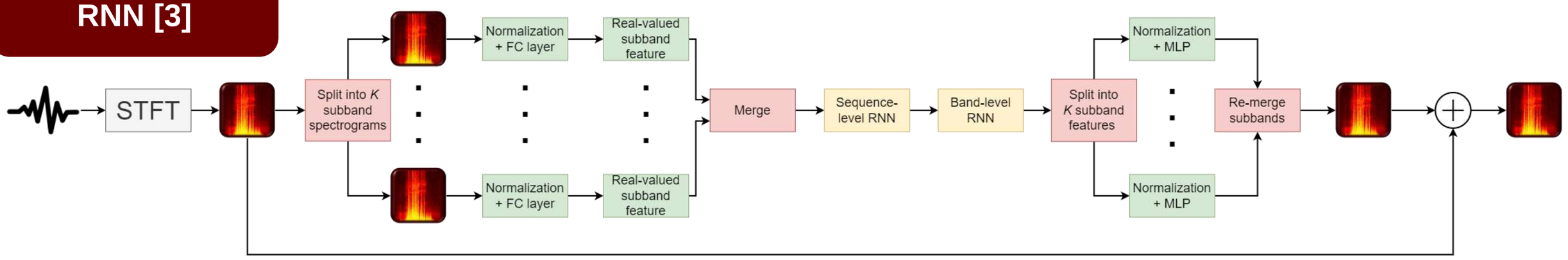
Band-split RoFormer [1]



HT Demucs [2]



Band-split RNN [3]



MUSDB18 dataset

Musdb18 is a dataset consisting of 150 full-length audio recordings of 10 different musical genres, totaling about 10 hours. It also includes isolated versions of the sound sources, which consist of drums, bass, vocals, and other sources.

Model comparison (MUSDB18)

Model	Algorithm	Domain	Extra training data	SDR
Band-split Roformer [1]	Transformer	Frequency	Yes (+450 songs)	9,8
HT Demucs [2]	Transformer /CNN	Time and Frequency	Yes (+3500 songs)	9,2
Band-split RNN [3]	RNN	Frequency	No	8,97
Stripe-Transformer [4]	Transformer	Frequency	No	6,71
D3Net [5]	CNN	Frequency	Yes (+1500 songs)	6,68
Meta-TasNet [6]	CNN	Time	No	5,52
Demucs [7]	CNN	Time	No	5,41
TF-Attention-Net [8]	Transformer	Frequency	Yes (+50 songs)	4,85

References

[1] Lu, Wei-Tsung, et al. "Music source separation with band-split rope transformer“, ICASSP, 2024

[2] Rouard, Simon, et al. "Hybrid transformers for music source separation." ICASSP, 2023

[3] Luo, Yi, et al. "Music source separation with band-split RNN." IEEE/ACM Transactions on Audio, Speech, and Language Processing 31, 2023

[4] Qian, Jiale, et al. "Stripe-Transformer: deep stripe feature learning for music source separation." EURASIP Journal on Audio, Speech, and Music Processing 2023

[5] Takahashi, Naoya, et al. "D3net: Densely connected multidilated densenet for music source separation." arXiv preprint 2020

[6] Samuel, David, et al. "Meta-learning extractors for music source separation." ICASSP 2020

[7] Défossez, Alexandre, et al. "Demucs: Deep extractor for music sources with extra unlabeled data remixed." arXiv preprint, 2019

[8] Li, Tingle, et al. "TF-Attention-Net: An End To End Neural Network For Singing Voice Separation." CoRR, 2019

Acknowledgment

The conference participation is funded by EPAM.