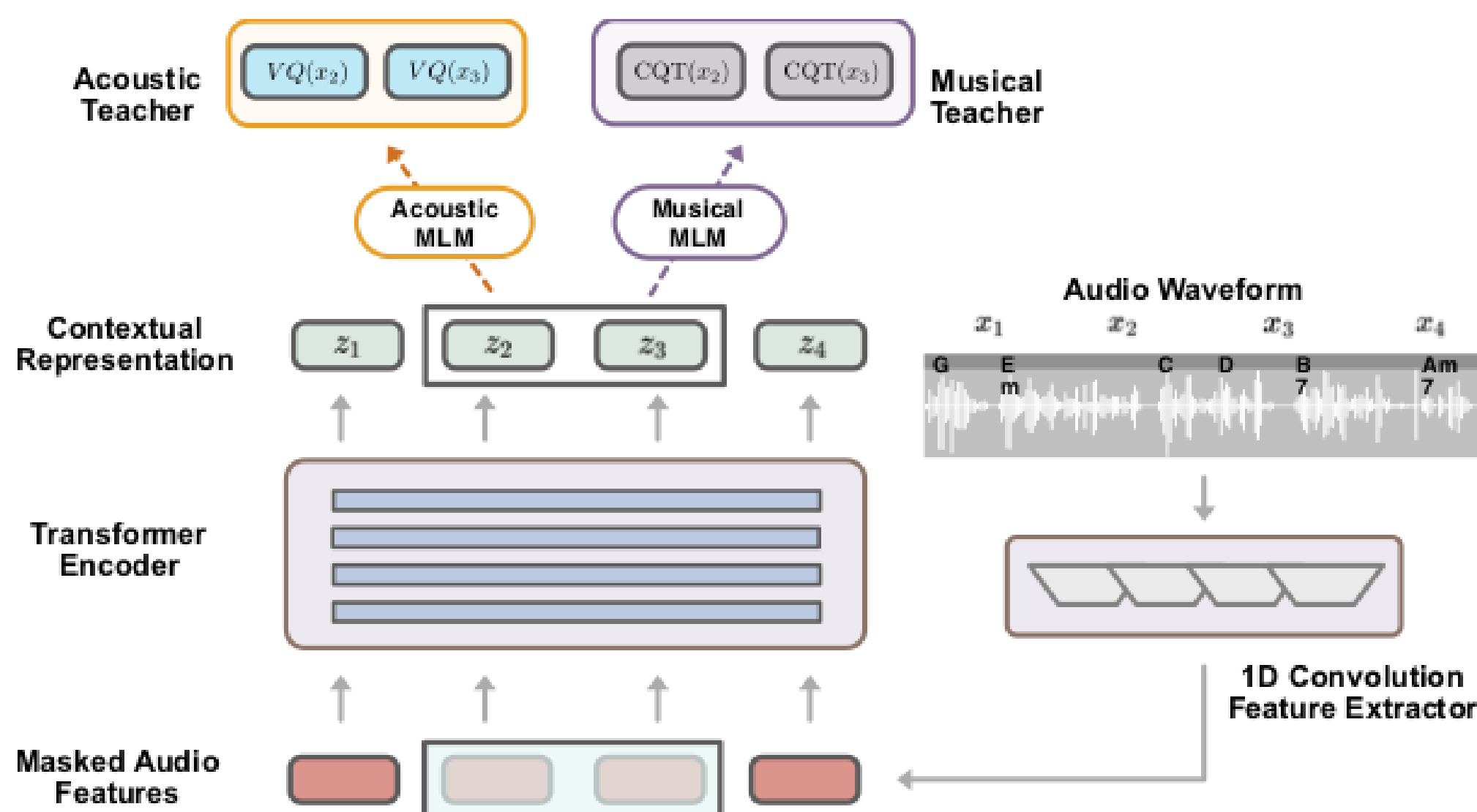# Self-supervised Learning for Music Information Retrieval

Yinghao Ma

2022

Finding

We propose a Music undERstanding model with large-scale self-supervised Training (MERT), which uses a masked language modelling style for pre-training. We identified a superior combination of EnCodec feature prediction and CQT reconstruction. Results indicate that our BERT-style transformer encoder performs well on 14 MIR tasks, attaining SOTA overall score.

Question

Training a SSL model requires many music recording. One alternative is to use important high-quality data (e.g. English Wikipedia for NLP). Which kind of datasets is important for music? What is the MIR datasets for world music (Chinese music, Indian music etc.), which have some distribution shift from the training dataset (typically Western pop music)?

Supervisor(s): Emmanouil Benetos, Chris Donahue