

Ilaria Manco

Emmanouil Benetos, George Fazekas, Elio Quinton

Bridging audio and language for music understanding


Joint audio-language learning leads to better music representations and enables new music-related tasks.

Are we doing joint multimodal training right?

Multimodal deep learning for MIR


Our experience of music is multimodal: we listen to audio, watch music videos, look at album cover art, read and write reviews, etc. Machine music understanding instead often relies on audio and/or metadata in isolation. My PhD research aims at developing **multimodal representation learning models for music understanding and description**, with a focus on **audio and language**.

MusCaps: Generating Captions for Music Audio



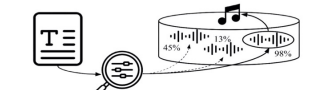
arXiv: [2104.11984](#)
Code: [github.com/ilariamanco/muscaps](#)

Song Describer: a Platform for Collecting Textual Descriptions of Music Recordings



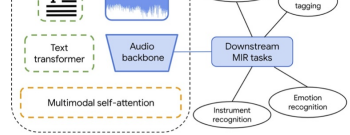
Extended abstract: [\[link\]](#)
Website: [song-describer.streamill.app](#)
Code: [github.com/ilariamanco/song-describer](#)

Contrastive Audio-Language Learning for Music



arXiv: [2208.12208](#)
Code: [github.com/yifeng-he/mancothuusell](#)




Learning Music Audio Representations via Weak Language Supervision



arXiv: [2112.09214](#)
Code: [github.com/ilariamanco/mu2ao](#)

Ilaria Manco

i.manco@qmul.ac.uk
www.ilariamanco.com




Research Interests

- Music information retrieval
- Multimodal representation learning
- Audio-and-language models

Applications

- Music captioning
- Cross-modal retrieval
- Music search, discovery, recommendation


UNIVERSAL MUSIC GROUP