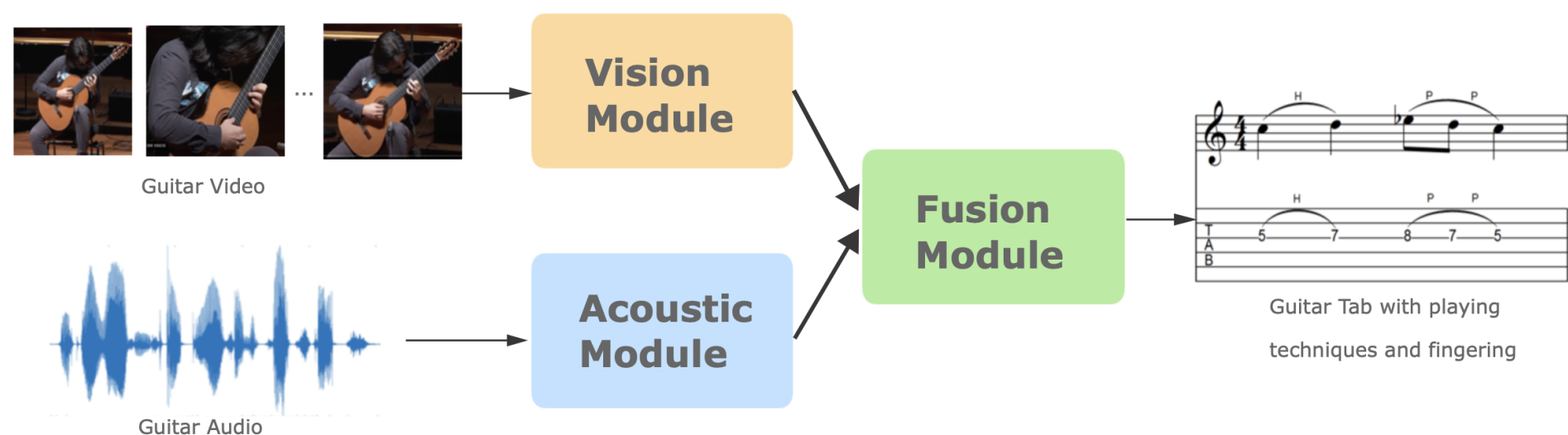


Multimodal Learning for Automatic Music Transcription

Zixun Guo

2023



Finding

There is a significant lack of open-source guitar dataset. Together with my collaborators, we have recently annotated 18 hours of guitar audio from Youtube. Our new dataset, named GAPS, has enabled us to achieve state-of-the-art performance in automatic guitar transcription (audio to midi).

Question

In the future, I aim to fuse the video modality into the transcription network. This mimics how guitarists transcribe guitar audio: first figure out what notes to play from audio, then find out where and how to play these notes on the fretboard. As a result, we will be able to achieve tablature transcription.