DJnet: A Dream for Making An Automatic DJ

Yu-Siang Huang[1], **Szu-Yu Chou**[1,2] **and Yi-Hsuan Yang**[1]
[1]Music and Audio Computing (MAC) Lab in CITI, Academia Sinica, Taiwan
[2]National Taiwan University, Taiwan
{yshuang, fearofchou, yang}@citi.sinica.edu.tw

Introduction

We present a research project called **DJnet**, whose goal is to make an fully-automatic DJ using deep learning. We want to let computers learn the following DJ skills:

- **Music Medley** (compose a song from parts of existing pieces)
- **Music Mashup** (blend two or more pre-recorded songs)
- **Music Thumbnailing** (extract a short snippet that represents a whole song, as Figure 1 exemplifies)
- **Electronic Dance Music Genre Classification** (differentiate sub-genres of EDM)
- **Electronic Dance Music Generation**

We demonstrate the research results of our two recent works Music Thumbnailing [1] and Music Medley [2]. We invite you to appreciate and dig deeper in DJ music!

Project website: https://remyhuang.github.io/DJnet/

Music Thumbnailing

Main ideas of our approach:

- Chorus is usually the most memorable and emotional
- Without annotations of the chorus sections of any song, we extract a music snippet of a song that happens to correspond to the song's chorus section by learning from emotion labels
- The key is to apply **attention mechanism** to a convolutional neural network (CNN), as Figure 2 shows
- Not only learn to predict music emotion, but also know where the novel parts are

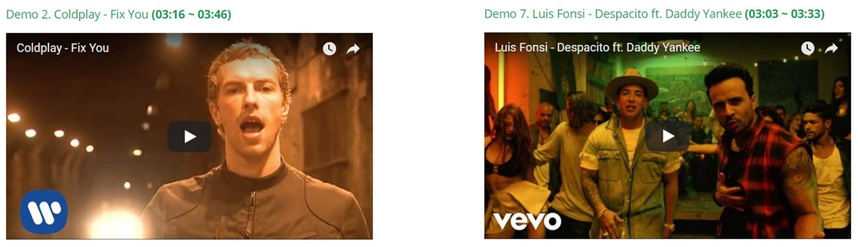

Figure 1: Example 30-second thumbnails

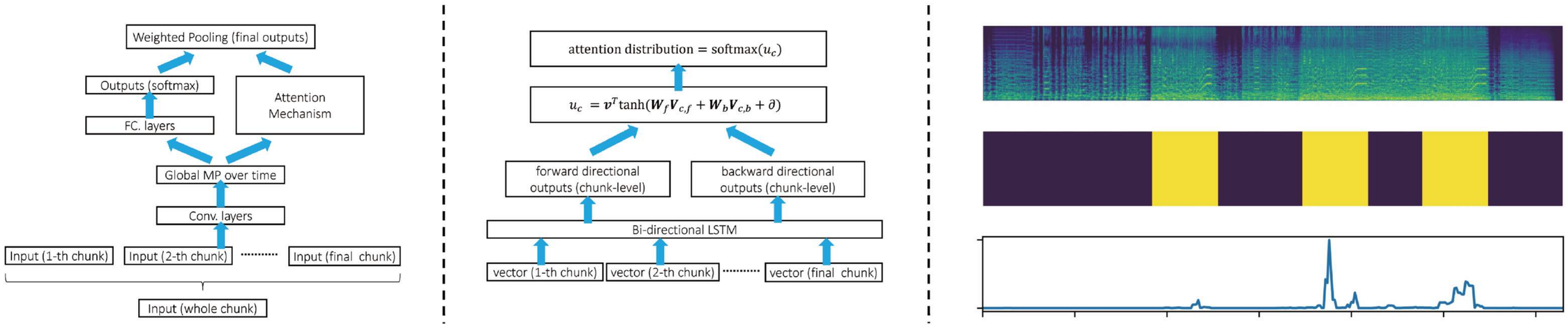

Figure 2: (Left & Middle) The proposed attention-based CNN model for music thumbnailing [1]. (Right) An example result. The first row is the mel-spectrogram, the second row marks the ground truth chorus sections (yellow regions), and the third row shows the attention scores estimated by our model. The peak falls within a chorus section.

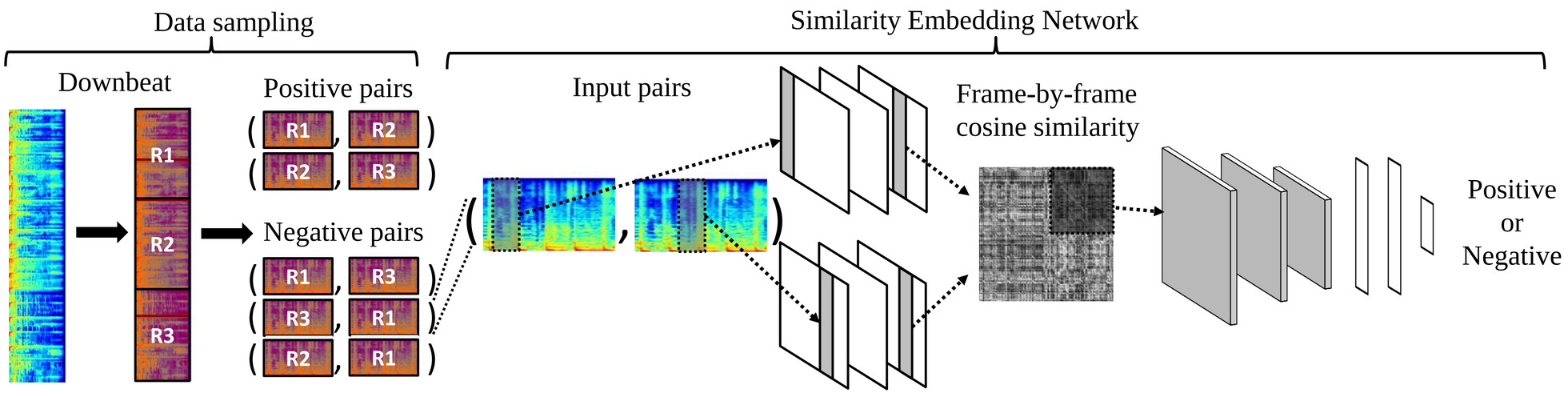

Figure 3: The proposed similarity embedding network for solving music puzzle games and generating music medleys [2].

Music Medley

- **Music puzzle games:** we formulate the task of assembling multiple non-overlapping music fragments in proper order
- **Similarity embedding network:** learn patterns from the similarity matrix of a pair of music fragments, as Figure 3 shows
- **Unsupervised learning:** any music collection can be used
- **Fitness:** pick the ordering with the highest fitness score

Reference

[1] Yu-Siang Huang, Szu-Yu Chou, and Yi-Hsuan Yang. Music thumbnailing via neural attention modeling of music emotion. *Proc. Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, page to appear, 2017.

[2] Yu-Siang Huang, Szu-Yu Chou, and Yi-Hsuan Yang. Similarity embedding network for unsupervised sequential pattern learning by playing music puzzle games. *arXiv preprint arXiv:1709.04384*, 2017.