

Generating Music Medleys via Playing Music Puzzle Games

Yu-Siang Huang^{1,2} Szu-Yu Chou^{1,2} Yi-Hsuan Yang¹

¹Music and Audio Computing (MAC) Lab in CITI, Academia Sinica, Taiwan

²GINM, National Taiwan University, Taiwan

{yshuang, fearofchou, yang}@citi.sinica.edu.tw

Project: https://remyhuang.github.io/music_puzzle_game/

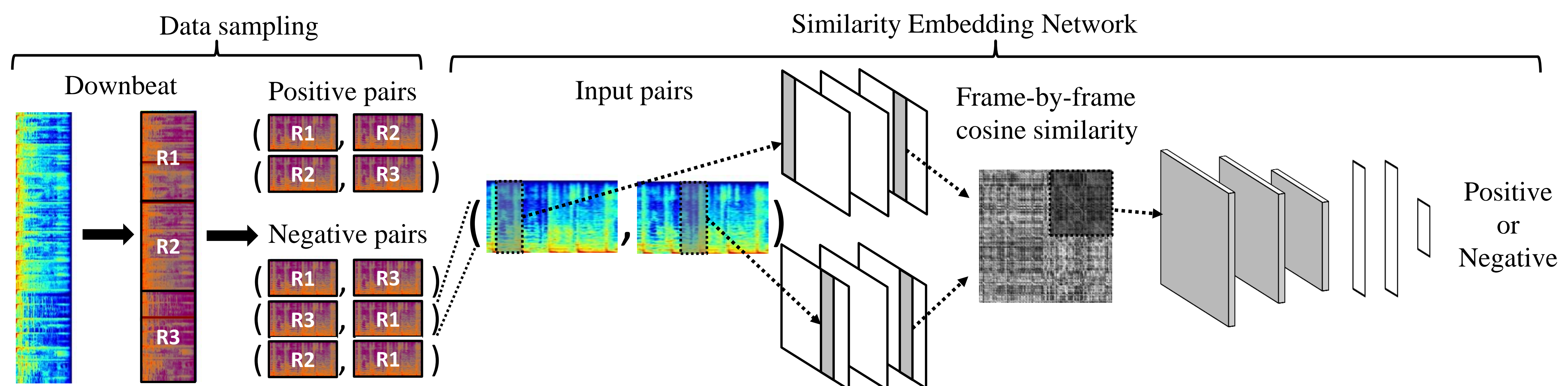
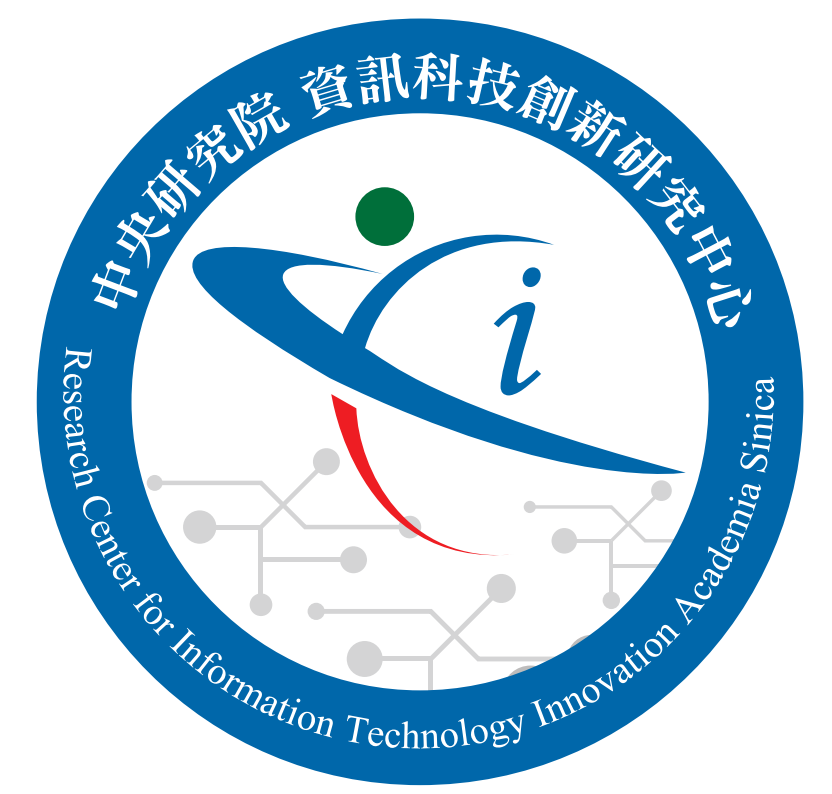


Figure 1: The proposed similarity embedding network for solving music puzzle games and generating music medleys.

Introduction

- Transfer **music medley generation** into **music puzzle games**
- Self-supervised learning task
- Predict whether a given pair is consecutive and in the correct chronological order
- Info from **similarity matrices**
- Significance: the model can be used for general temporal signals

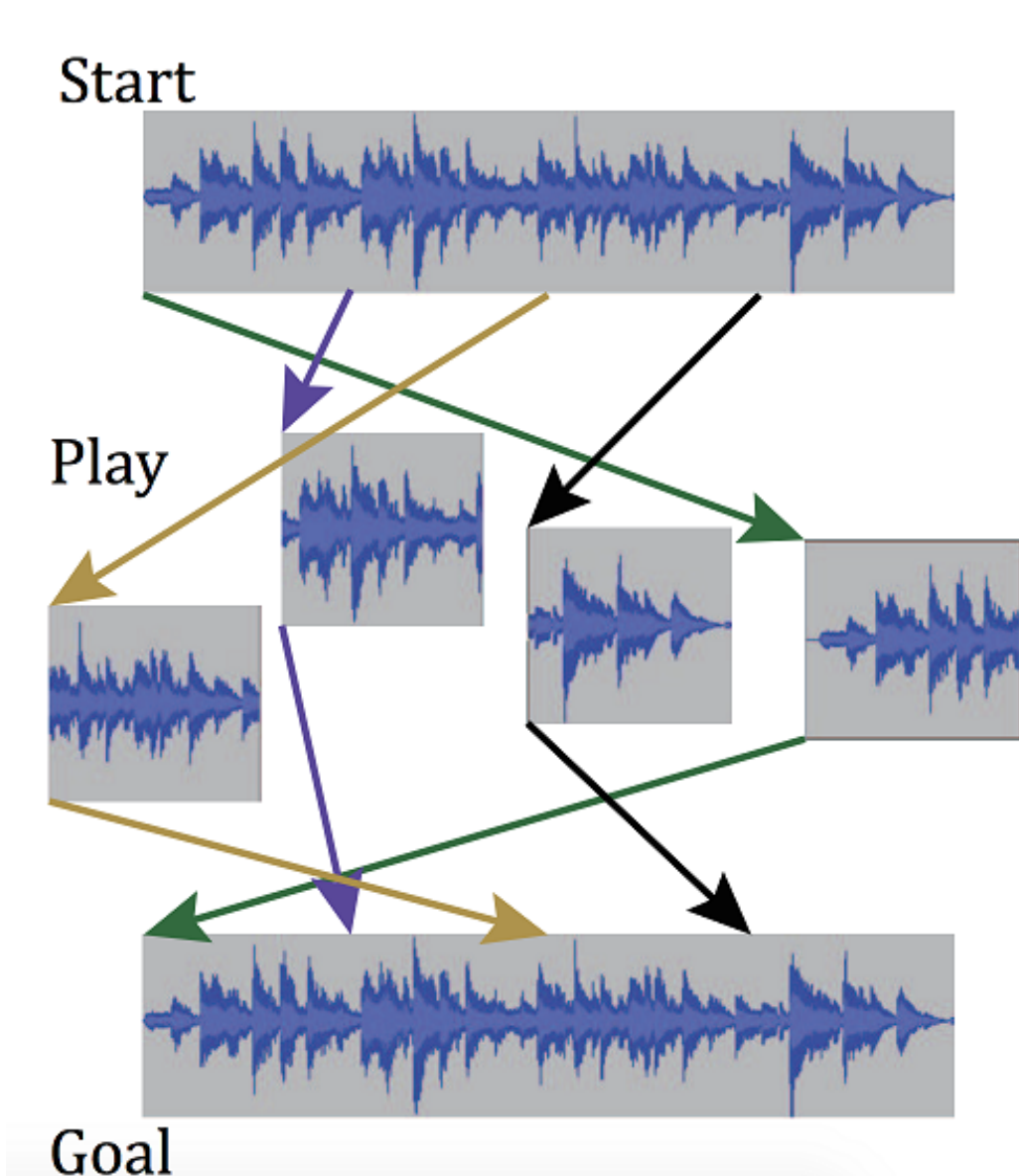


Figure 2: Music puzzle game

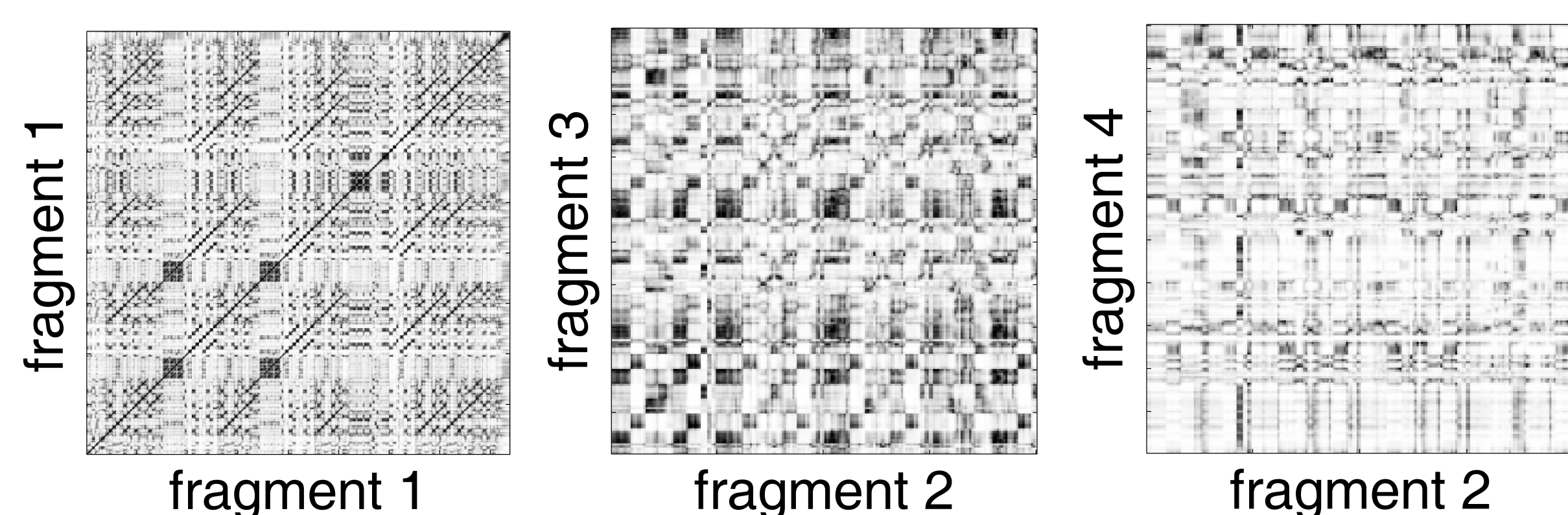


Figure 3: Example visualization of similarity matrix.

Result

- The estimate of each data pair in $[0, 1]$
- Given n fragments, we get this estimate for all the pairs
- Find the overall order with the highest sum of $n-1$ pairs

Method	pairwise accuracy	global accuracy	n-piece	SN	CIN	SEN
SN	0.851	0.825	3	0.692	0.863	0.991
CCSN [1]	0.872	0.840	4	0.472	0.761	0.987
CIN	0.912	0.864	6	0.171	0.499	0.971
TSN [2]	0.911	0.890	8	0.056	0.297	0.961
OPN [3]	0.929	0.916				
SEN (proposed)	0.996	0.994				

Table 1: Result on 3-piece puzzle & n-piece puzzle by different models.

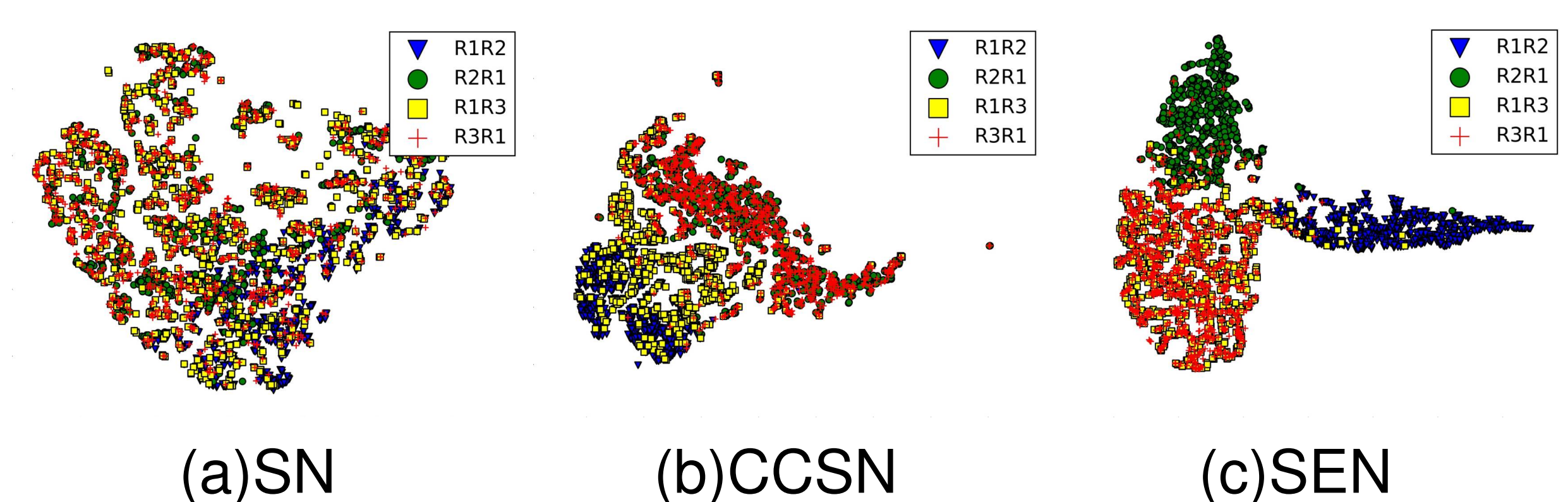


Figure 5: Embedding visualization

Compared model

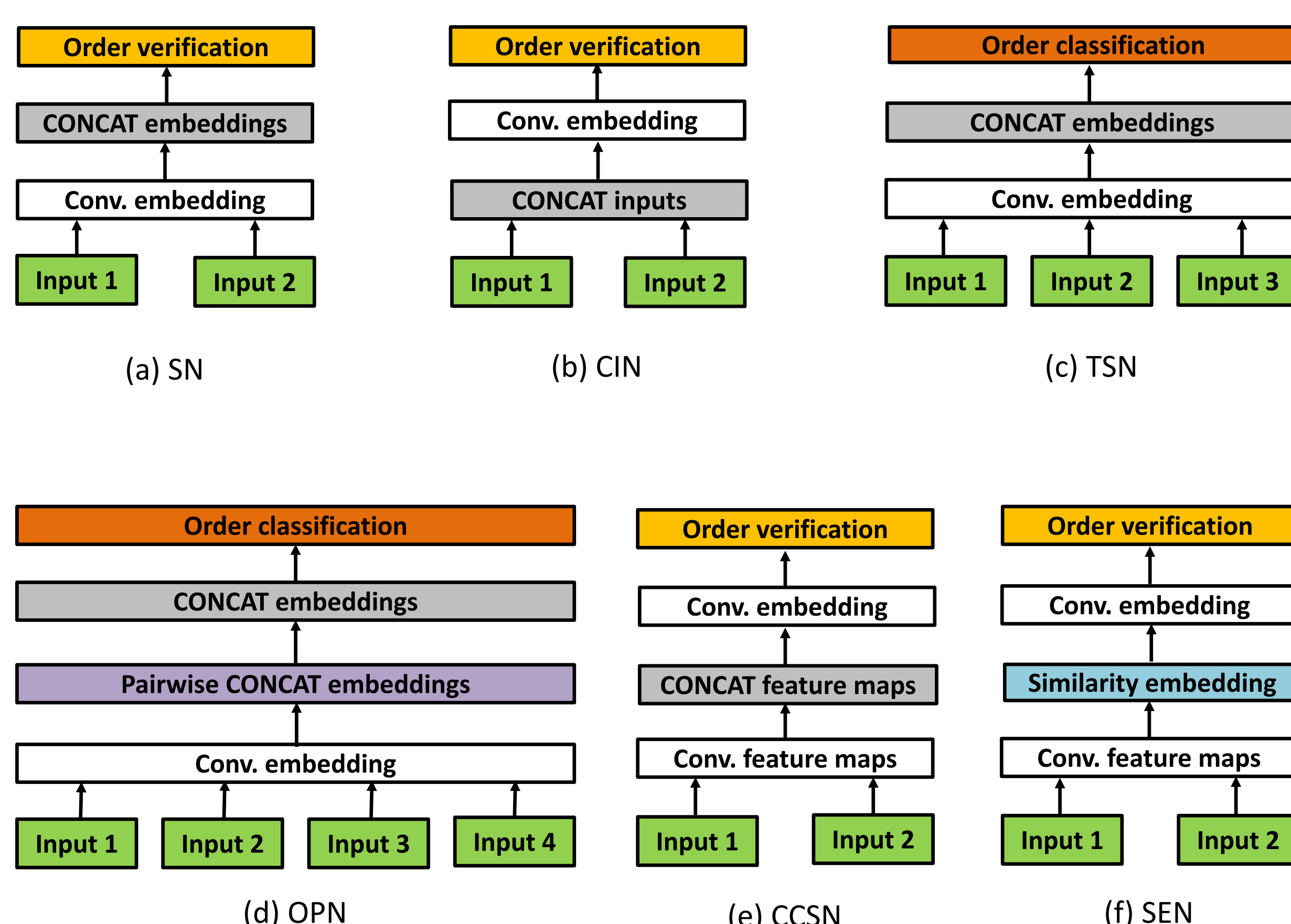
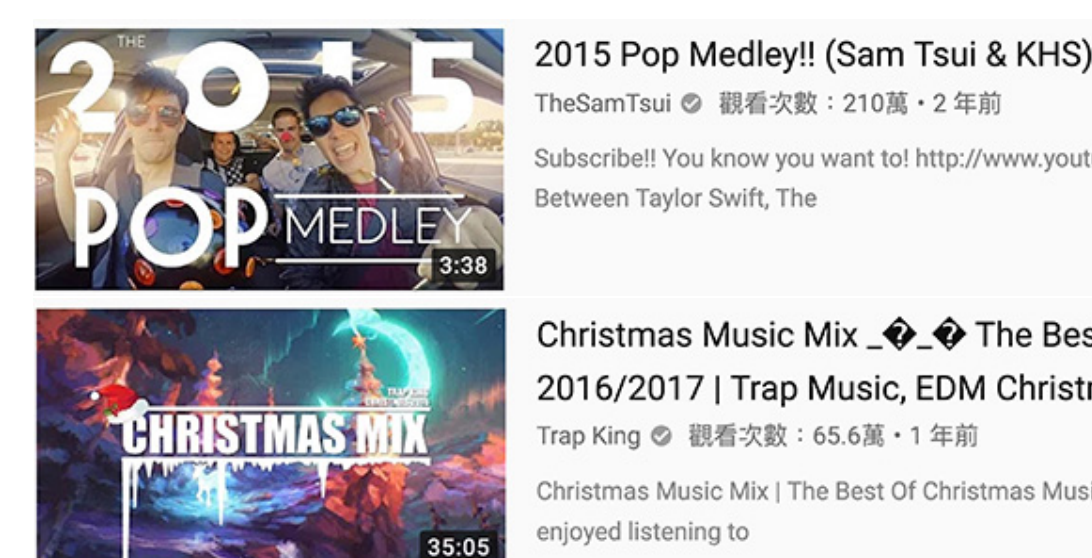


Figure 4: Architecture of all the compared models.

Application & Future works



- Music medley generation
- Cross several genres
- Raw audio signals as inputs
- Investigate musical property
- Add **transition** methods

Reference

- [1] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. Joint learning of single-image and cross-image representations for person re-identification. In *CVPR*, pages 1288–1296, 2016.
- [2] Ishan Misra, C Lawrence Zitnick, Martial Hebert, and . Shuffle and learn: Unsupervised learning using temporal order verification. In *ECCV*, pages 527–544, 2016.
- [3] Hsin-Ying Lee, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Unsupervised representation learning by sorting sequences. *arXiv preprint arXiv:1708.01246*, 2017.