Fusion-based system for SASV 2022

Xingjia Xie[1], *Lin Li*[1], *Qingyang Hong*[2]

[1]School of Electronic Science and Engineering, Xiamen University, China
[2]School of Informatics, Xiamen University, China
{lilin,qyhong}@xmu.edu.cn

Abstract

This paper describes the system from xmuspeech for Spoofing Aware Speaker Verification Challenge 2022. We propose a fusion strategy based on score-level fusion. For the task, we evaluate our system on ASVspoof 2019 LA development set and evaluation set which greatly improves the performance compared with the Baseline2. Our best submission obtained on the evaluation set SASV-EER 1.155%, while the performance on the development set is SASV-EER 0.723%.

Index Terms: SASV, speaker recognition, fusion-based system

1. System Description

For the challenge, inspired by the system from the team of Y Zhang et al [1], we adopt the similar system settings and the same dataset without any extra data other than score-level fusion method. The initial system sums the scores produced by the separate systems, we replace addition with multiplication and make a great improvment on the evaluation set and development set for SASV-EER. Based on this, we use the BOSARIS toolkit [2] to fuse multiple system's scores. Surprisingly, the three EERs calculated by the fused scores can be improved to some extent.

Table 1: The results of *SASV Challenge.*

Model	DEV			EVAL		
	SASV-EER(%)	**SV-EER(%)**	**SPF-EER(%)**	**SASV-EER(%)**	**SV-EER(%)**	**SPF-EER(%)**
ECAPA-TDNN [2]	17.38	1.88	20.30	23.83	1.63	30.75
Baseline1 [1]	13.07	32.88	0.06	19.31	35.32	0.67
Baseline2 [4]	4.85	12.87	0.13	6.37	11.48	0.78
Our Baseline2	4.78	12.80	0.10	6.33	11.32	0.80
Multi_task	11.66	——	——	11.38	——	——
Baseline1 with multiplication for score-level fusion [1]	2.156	4.178	0.197	2.886	4.283	0.894
pr_s_f [1]	1.094	2.022	0.067	1.527	1.955	0.801
Baseline1_s_i [1]	1.685	2.561	0.067	2.449	3.091	0.764
Fusion of Baseline1_s_i_and_pr_s_f [1]	0.559	0.876	0.067	1.844	2.421	0.931
Fusion of pr_s_f_and_Baseline1_s_i_and_multi_task [1]	0.723	1.387	0.067	1.155	1.485	0.773

2. Conclusions

It is surprising that a simple strategy can improve the SASV performance a lot. We proposed a simple but effective fusion-based method for spoofing aware speaker verification (SASV). The result suggests that the multiplication for score-level fusion has a better discrimination ability.

3. References

[1] Y Zhang, G Zhu, and Z Duan, "A New Fusion Strategy for Spoofing Aware Speaker Verification," arXiv preprint arXiv:2202.05253, 2022.

[2] N Brümmer and E De Villiers, "The BOSARIS toolkit: Theory, algorithms and code for surviving the new DCF," arXiv preprint arXiv:1304.2865 (2013).

[3] RK Das, R Tao, and H Li, "HLT-NUS SUBMISSION FOR 2020 NIST Conversational Telephone Speech SRE," arXiv preprint arXiv:2111.06671 (2021).

[4] J Jung, H Tak, and H Shim, "SASV Challenge 2022: A Spoofing Aware Speaker Verification Challenge Evaluation Plan," arXiv preprint arXiv:2201.10283, 2022.

[5] X Wang, J Yamagishi, and M Todisco, "ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech. Computer Speech & Language," 2020, 64: 101114.

[6] J Jung, H HS, and H Tak, "AASIST: Audio Anti-Spoofing using Integrated Spectro-Temporal Graph Attention Networks," arXiv preprint arXiv:2110.01200 (2021).