Inspiring : Journey From a small Village in India To UK - Bhusan Chettri

Bhusan Chettri | Bhusan Chettri Research

Ensemble Models for Spoofing Detection in Automatic Speaker Verification

Bhusan Chettri, Daniel Stoller, Veronica Morfi, Marco A. Martínez Ramírez, Emmanouil Benetos, Bob L. Sturm
b.chettri@qmul.ac.uk

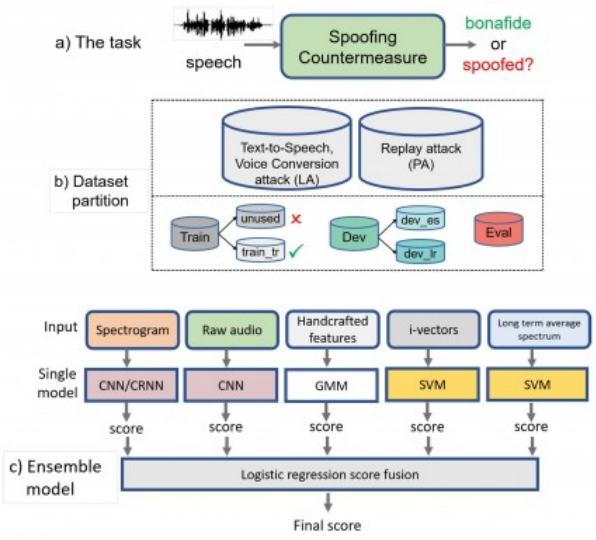
Queen Mary University of London

KTH VETENSKAP OCH KONST

1 Introduction

- We explore ensemble models for spoofing detection on the ASVspoof 2019 logical access (LA) and physical access (PA) datasets [1].
- We find models appear to have improved generalisation when we partition those datasets to ensure disjoint attack conditions [2].
- We examine why some models work so well and find they are using specific irrelevant cues in the recordings.

2 Tasks and Model description



3 Experimental results

- Metric: tandem-DCF (t-DCF) [3] and equal error rate (EER)
- LFCC GMM (B1) and CQCC-GMM (B2) are official baselines

Model	Set	LA attack		PA attack	
		t-DCF	EER%	t-DCF	EER%
B1	Dev	0.0663	2.71	0.2554	11.96
B2	Dev	0.0123	0.43	0.1953	9.87
ensemble	Dev	**0.0**	**0.0**	**0.0354**	**1.33**
B1	Eval	0.2116	8.09	0.3017	13.54
B2	Eval	0.2366	9.57	0.2454	11.04
ensemble	Eval	**0.0755**	**2.64**	**0.1492**	**6.11**

4 What is the CNN exploiting in the PA dataset?

- We find that a CNN performs much better when trained on the last 4 seconds of every recording than on the first 4 seconds.
- We find this comes from silent segments in the spoof recordings.

Intervention I: remove silence from the end at test time.

Model	t-DCF	EER %
B1	$0.2036 \rightarrow 0.2741$	$9.18 \rightarrow 13.27$
B2	$0.1971 \rightarrow 0.2959$	$10.06 \rightarrow 15.59$
CNN	$0.1672 \rightarrow 0.5018$	$5.98 \rightarrow 19.8$

Intervention II: train the models removing silence from the end.

Model	t-DCF	EER %
B1	$0.2036 \rightarrow 0.9528$	$9.18 \rightarrow 54.76$
B2	$0.1971 \rightarrow 0.9463$	$10.06 \rightarrow 57.98$
CNN	$0.1672 \rightarrow 0.2626$	$5.98 \rightarrow 11.20$

Intervention III: remove silence during both training and testing.

Model	t-DCF	EER %
B1	$0.2036 \rightarrow 0.8614$	$9.18 \rightarrow 41.09$
B2	$0.1971 \rightarrow 0.9448$	$10.06 \rightarrow 58.71$
CNN	$0.1672 \rightarrow 0.3129$	$5.98 \rightarrow 12.85$

How about the evaluation set?

- Models show similar behaviour under above interventions.

5 Conclusion

- We find ensemble models are better than the baselines in detecting unseen spoofing attacks, yielding 3^{rd} rank in the LA task.
- We find their performance on the PA task is inflated due to a cue (existence of silence) in the recordings of the dataset [4].
- We propose removing this cue in the PA dataset [5] for more reliable estimate of performance.

[1] Massimiliano et. al. ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection. In *Proc. Interspeech*, 2019.
[2] Partition details: https://github.com/BhusanChettri/ASVspoof2019/.
[3] Kinnunen et. al. t-DCF: a Detection Cost Function for the Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification. In *Proc. Speaker Odyssey*, 2018.
[4] B. L. Sturm. A Simple Method to Determine if a Music Information Retrieval System is a "Horse". In *IEEE Transactions on Multimedia*, 2014.
[5] B. Chettri and B. L. Sturm. A Deeper Look at Gaussian Mixture Model Based Anti-Spoofing Systems. In *IEEE ICASSP*, 2018.

INTERSPEECH 2019 Conference, 15-19 September, Graz, Austria.

Sheffield, England Jul 18, 2023 (Issuewire.com**)** - Pakyong, 25 Aug: The Journey from a small town in South Sikkim district, India to London, UK wasn't easy for Dr. Bhusan Chettri.

Son of Mr. Tilak Chettri and Mrs. Krishna Maya Thapa, Mr. Bhusan has successfully attained a doctorate degree from the prestigious Queen Mary University of London (QMUL), UK where he started his Ph.D. in a fully-funded QMUL principal research scholarship in the year 2016.

His Ph.D. thesis focused on the analysis and design of voice biometric systems security using Machine Learning and AI. He was supervised by Dr. Emmanouil Benetos, Associate Professor at Centre for Digitial Music, QMUL, and Dr. Bob Sturm, Associate Professor at KTH Royal Institute of Technology, Sweden.

With 5 peer-reviewed conference papers and 2 journals of very reputed and highly recognized international acclaim, **Dr.** Bhusan Chettri having worked with the pioneers of the field has gathered many accolades and has managed to carve a niche for himself in the research community globally.

His thesis is publicly available online to read and download through the link below: https://theses.eurasip.org/theses/866/voice-biometric-system-security-design-and/

Prior to that Mr, Bhusan had completed his MSc in Speech and Language Processing from the University of Sheffield, the UK in the year 2014 after which he started working there as a research assistant. He completed his earlier education at Temi Tarku and Rabangla, Sikkim, and then went on to pursue his B.Tech in Computer Science and Engineering from SMIT, Sikkim, India. He also attained an M.Tech and MBA degree from the same university.

Thereafter he started his career at SMIT as an Assistant professor, was, later on, promoted to Associate professor, and continued working there for 7 years before he decided to take the next leap of moving abroad for his higher education.

Currently, Dr. Bhusan Chettri is working as a post-doctoral researcher under Associate Professor, Dr. Tomi Kinumen at the University of Eastern Finland. His next endeavour in life as a Lecturer of Data Analytics, at QMUL, London UK starts on the 1st of September 2020 and we wish him all the best for his future.

The Voice Of Sikkim wishes him all the best for his career and prospects ahead.

RELATED ITEMS: COMPUTER ENGINEERING, DR BHUSHAN CHETTRI, MACHINE LEARNING, AND ARTIFICIAL INTELLIGENCE AI

Publons | LinkedIn

Media Contact

Bhusan Chettri

bhusanchettri5@gmail.com

Source : https://ieeexplore.ieee.org/author/37086452985

See on IssueWire