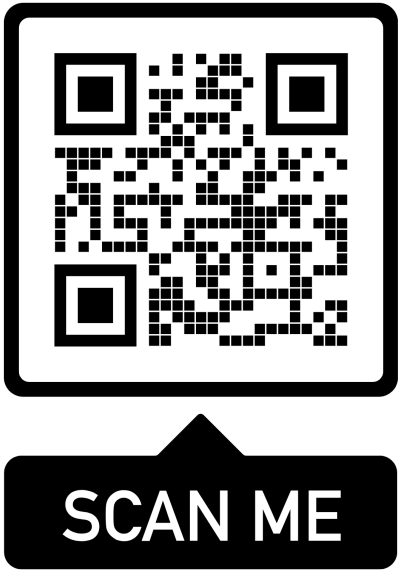
Improving Generalization Ability for Audio Deepfake Detection
提高音频鉴伪的泛化能力

You Zhang 张优，University of Rochester

Dec 28, 2023

Self-introduction

- PhD candidate at University of Rochester, working with Prof. Zhiyao Duan
- B.Eng from UESTC, Exchange studies at UC Berkeley
- Research interests:
  - Computer Audition
    - Audio Deepfake Detection
    - Spatial Audio Personalization
    - Audio-Visual Rendering and Analysis

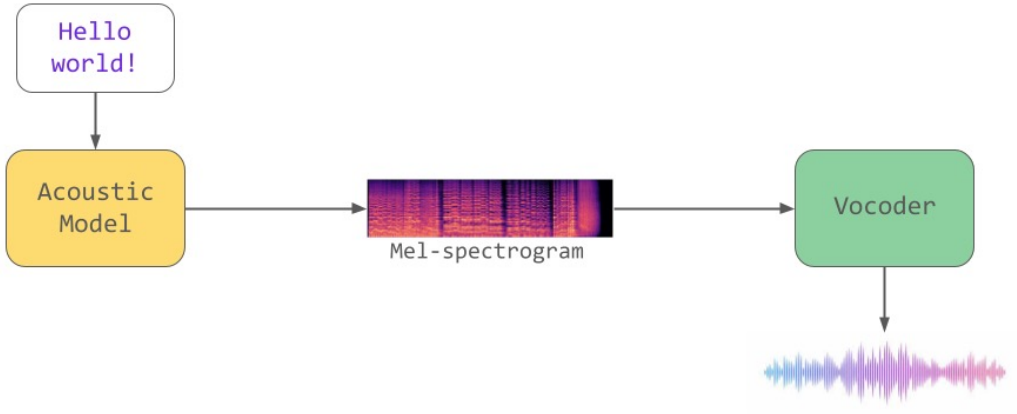


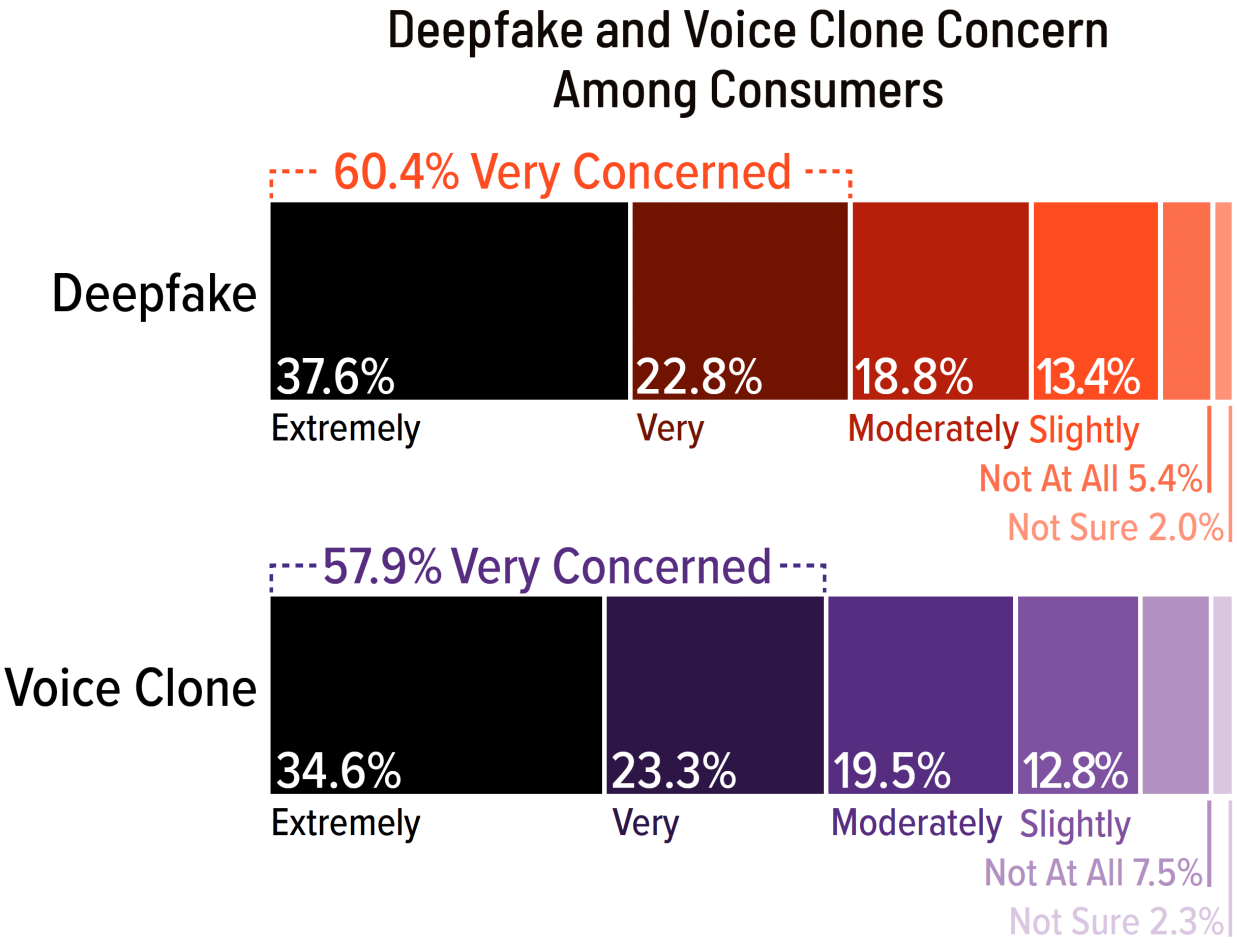
Outline

- Introduction to speech anti-spoofing
- Generalization ability to unseen synthetic attacks
  - One-class learning: OC-Softmax;
  - Multi-center one-class learning: SAMO
- Beyond speech anti-spoofing: **Singing voice** deepfake detection
- Future directions

Demo of Speech Deepfakes

- Tacotron2
- Fastpitch
- BigVGAN

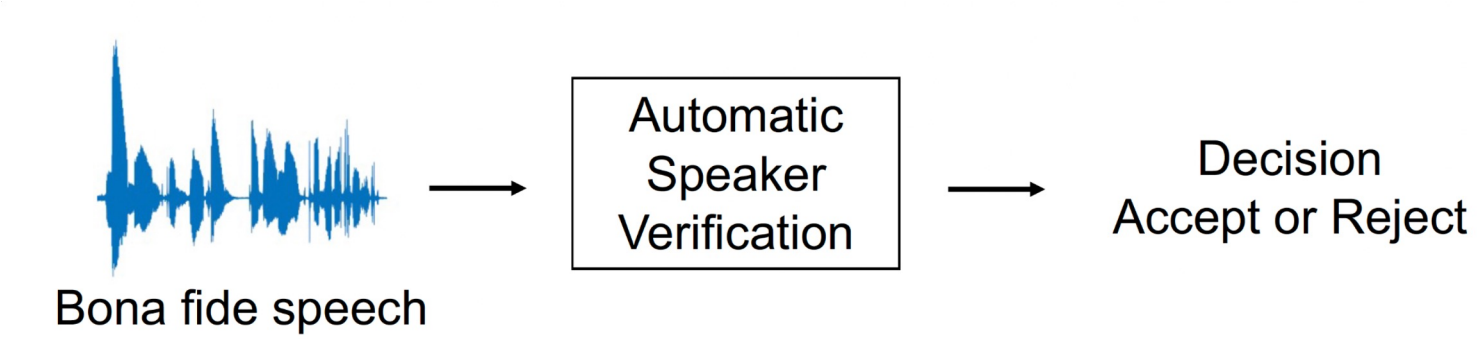




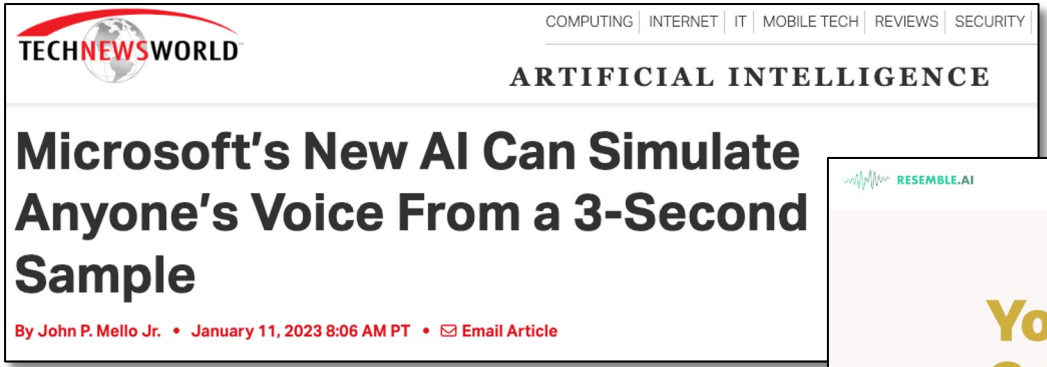
Source: Voicebot

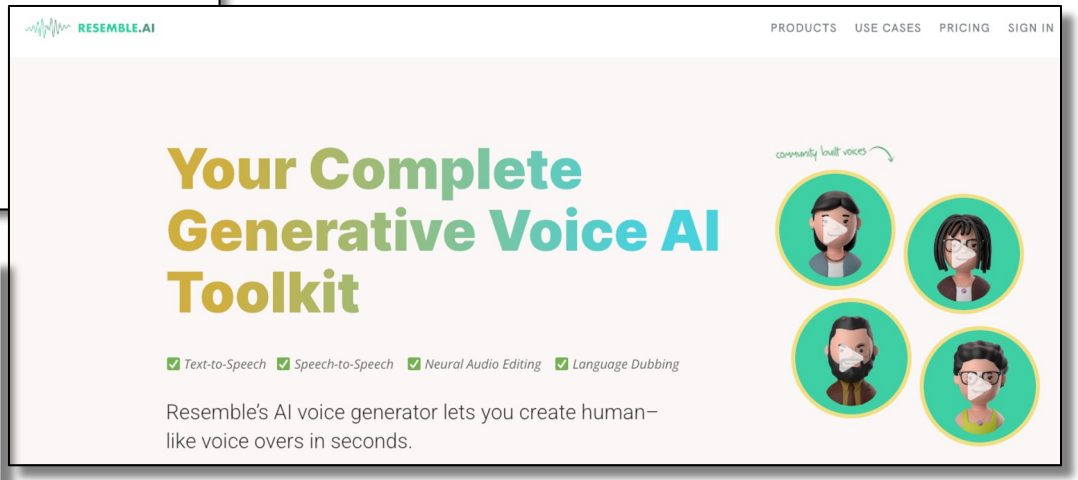
Voice Biometrics

Verify the identity of a speaker



We expect that the input (bona fide speech) is from a real person.





top left
top right
bottom left
bottom right

Spoofing attacks

Impersonation

-- twins and professional mimics

Replay

-- reuse pre-recorded audio, most accessible

Text-to-speech (TTS)

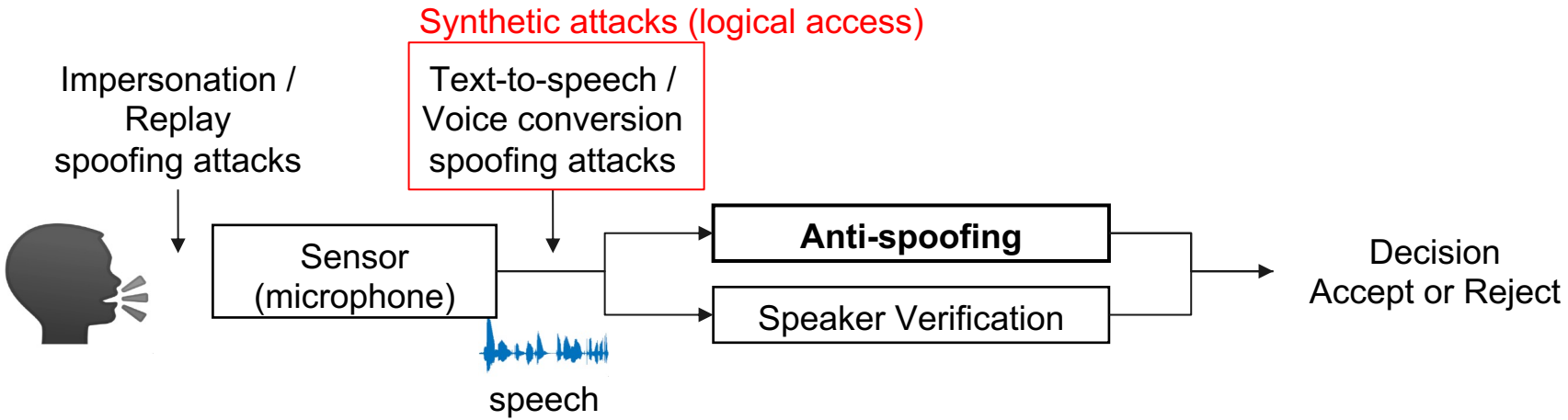
-- convert written text into spoken words with speech synthesis

Voice conversion (VC)

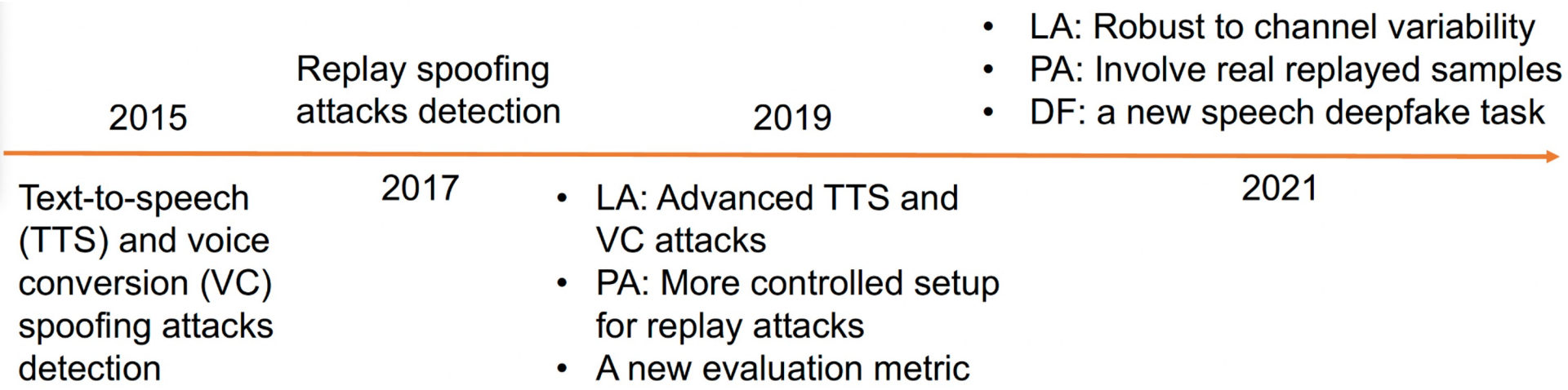
-- convert speech from source speaker to target speaker’s voice

Speech Anti-Spoofing (Speech Deepfake Detection)

A voice anti-spoofing system is desired to distinguish synthetic **speech** from **bona fide speech**.



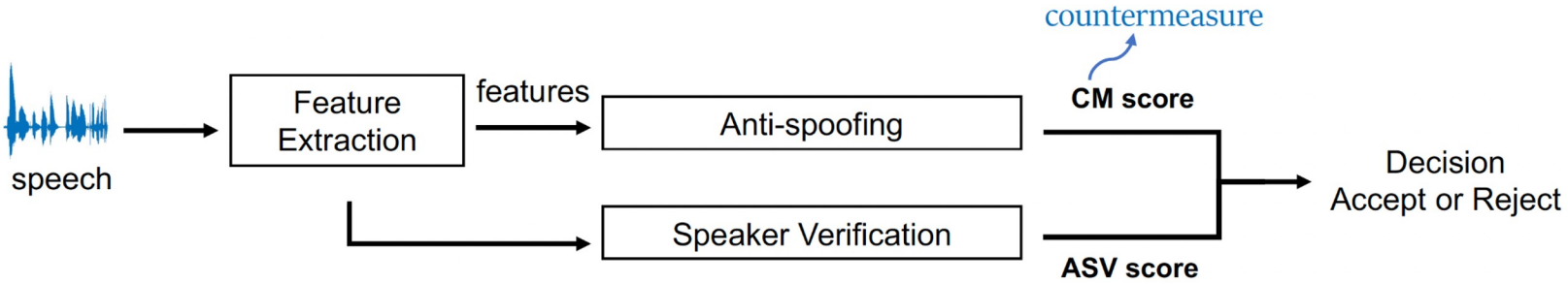
ASVspoof challenge series



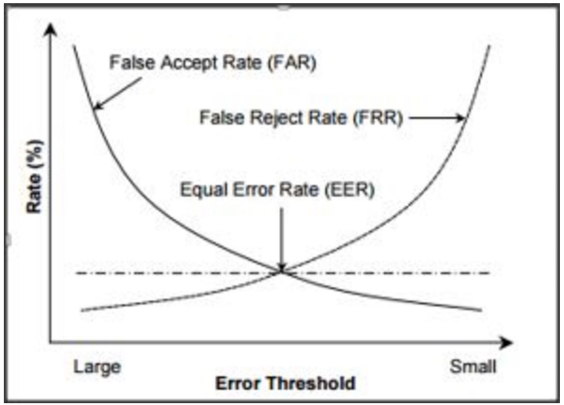
Logical Access (LA): algorithm-related artifacts

Physical Access (PA): device-related artifacts

Evaluation metric



- EER (Equal Error rate)



$$P_{fa}(\theta) = \frac{\#\{\text{spoof trials with score} > \theta\}}{\#\{\text{total spoof trials}\}},$$

$$P_{miss}(\theta) = \frac{\#\{\text{human trials with score} \leq \theta\}}{\#\{\text{total human trials}\}}$$

$$P_{fa}(\theta_{EER}) = P_{miss}(\theta_{EER})$$

Dataset

ASVspoof 2019 Logical Access (TTS + VC)

- Bona fide speech (VCTK dataset)
- 6 known attacks (appear in training set)
- 11 unknown attacks (only appear in eval set)
- 2 attacks (use known algorithms but trained with more data)

	Bona fide	Spoofed	
	# utterance	# utterance	attacks
Training	2,580	22,800	A01 - A06
Development	2,548	22,296	A01 - A06
Evaluation	7,355	63,882	A07 - A19

Research question

Motivation:

- The fast development of speech synthesis are posing increasingly more threat.
- The **distribution mismatch** between the training set and test set for the spoofing attacks class.

How can the anti-spoofing system defend against **unseen** spoofing attacks?

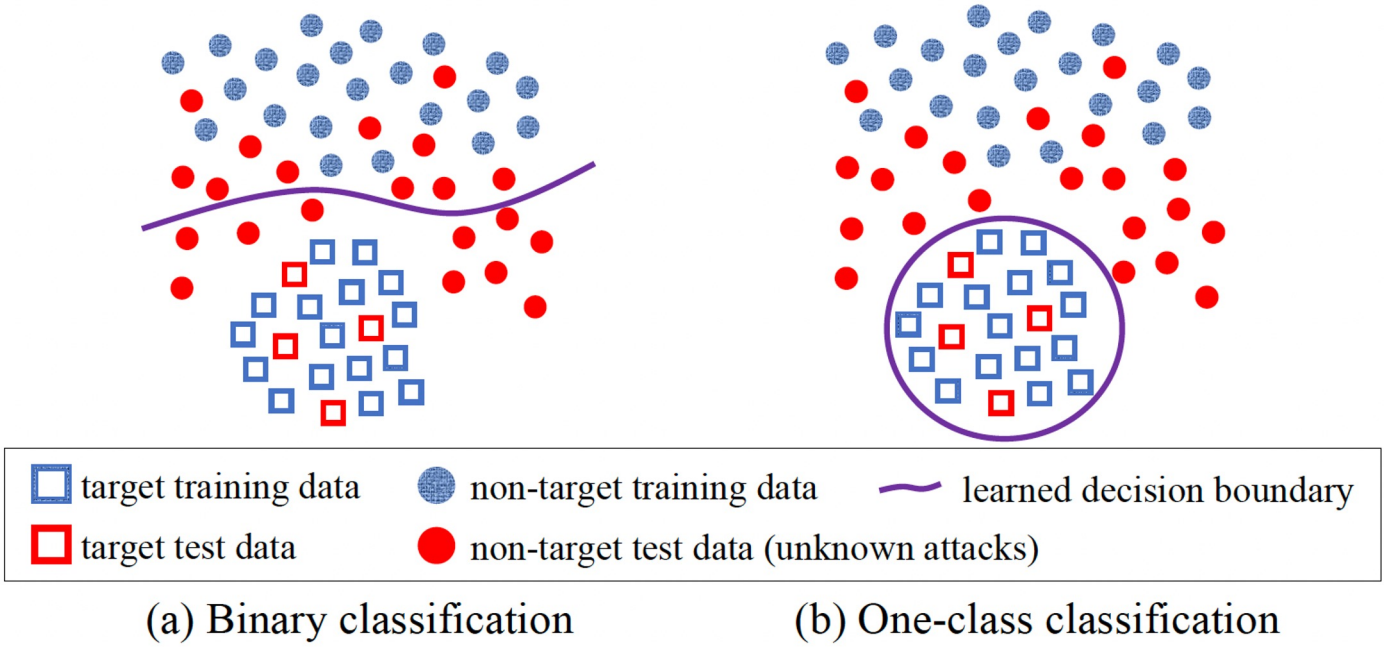
Generalization ability!

Definition of one-class classification

- "One of the classes (referred to as the positive class or target class)
  - is well characterized by instances in the training data.
- For the other class (nontarget),
  - it has either no instances at all,
  - very few of them,
  - or they do **not form a statistically-representative** sample of the negative concept."

Khan, S. S., & Madden, M. G. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review*, *29*(3), 345-374.

Illustration of comparison



You Zhang, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan. “Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks and Channel Variation”, *Handbook of Biometric Anti-spoofing (3rd edition)*, Springer, 2023.

One-class learning

- Compact the bona fide speech representation
- Isolate the spoofing attacks

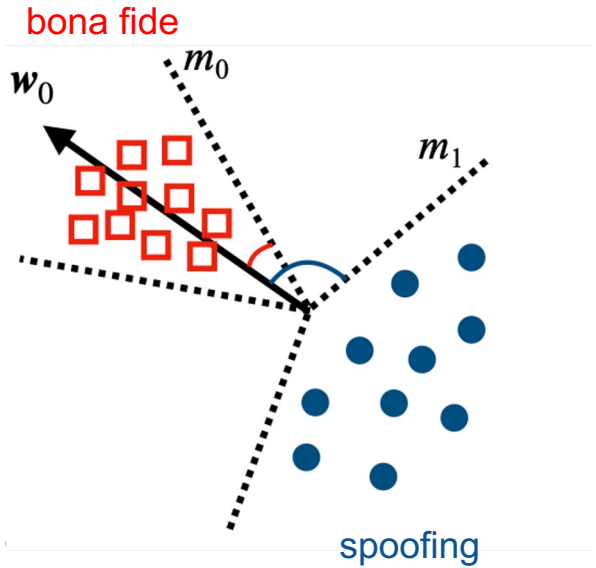
Training: OC-Softmax loss (Proposed)

$$\mathcal{L}_{OCS} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{\alpha(m_{y_i}-\hat{\boldsymbol{w}}_0\hat{\boldsymbol{x}}_i)(-1)^{y_i}}\right).$$

scale factor, center vector, label, # samples, margin, embedding

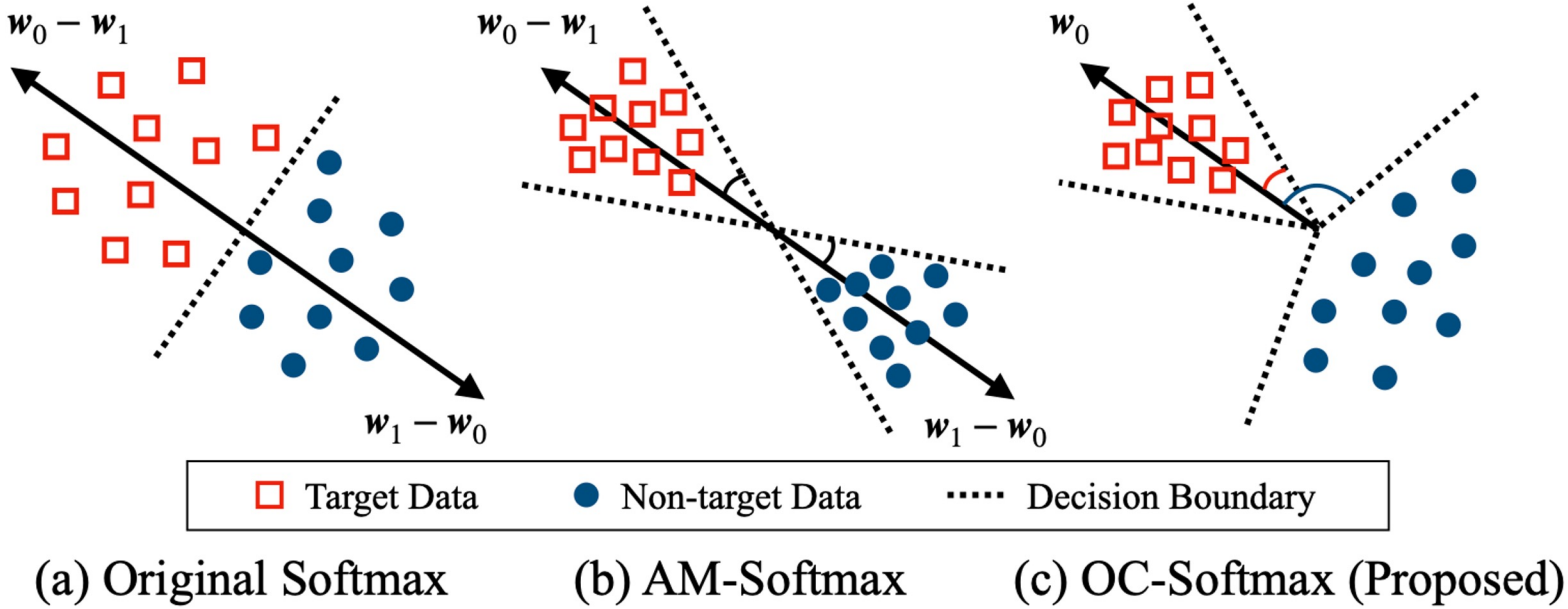
Inference: cosine similarity

$$\mathcal{S}_{OCS} = \hat{\boldsymbol{w}}_0\hat{\boldsymbol{x}}_i.$$



You Zhang, Fei Jiang, and Zhiyao Duan, "One-class Learning Towards Synthetic Voice Spoofing Detection", *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021.

Comparing OC-Softmax with binary classification



You Zhang, Fei Jiang, and Zhiyao Duan, "One-class Learning Towards Synthetic Voice Spoofing Detection", *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021.

Comparing OC-Softmax with binary classification

Softmax:

$$\mathcal{L}_S = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\boldsymbol{w}_{y_i}^T\boldsymbol{x}_i}}{e^{\boldsymbol{w}_{y_i}^T\boldsymbol{x}_i}+e^{\boldsymbol{w}_{1-y_i}^T\boldsymbol{x}_i}} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{(\boldsymbol{w}_{1-y_i}-\boldsymbol{w}_{y_i})^T\boldsymbol{x}_i}\right),$$

AM-Softmax:

$$\mathcal{L}_{AMS} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\alpha(\hat{\boldsymbol{w}}_{y_i}^T\hat{\boldsymbol{x}}_i-m)}}{e^{\alpha(\hat{\boldsymbol{w}}_{y_i}^T\hat{\boldsymbol{x}}_i-m)}+e^{\alpha\hat{\boldsymbol{w}}_{1-y_i}^T\hat{\boldsymbol{x}}_i}} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{\alpha\left(m-(\hat{\boldsymbol{w}}_{y_i}-\hat{\boldsymbol{w}}_{1-y_i})^T\hat{\boldsymbol{x}}_i\right)}\right),$$

OC-Softmax:

$$\mathcal{L}_{OCS} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{\alpha(m_{y_i}-\hat{\boldsymbol{w}}_0\hat{\boldsymbol{x}}_i)(-1)^{y_i}}\right).$$

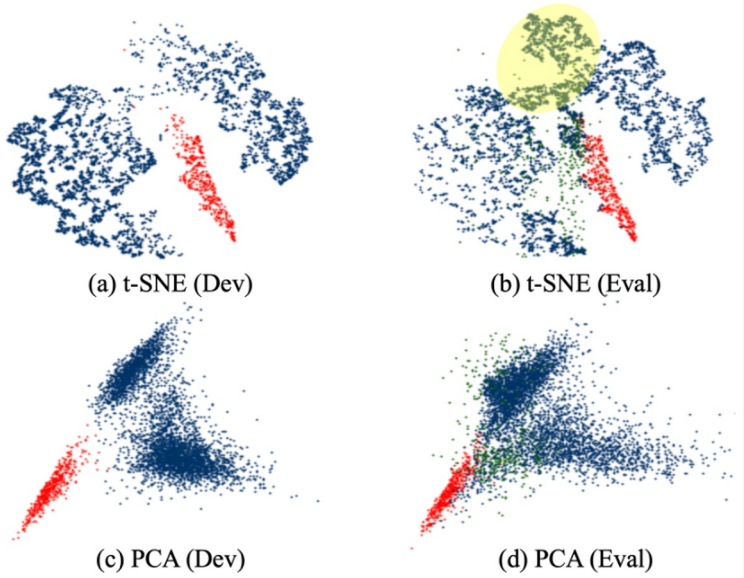
You Zhang, Fei Jiang, and Zhiyao Duan, “One-class Learning Towards Synthetic Voice Spoofing Detection”, *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021.

Evaluation of OC-Softmax

Results on the development and evaluation sets of ASVspoof 2019 LA using different losses

Loss	Dev Set		Eval Set	
	EER (%)	min t-DCF	EER (%)	min t-DCF
Softmax	0.35	0.010	4.69	0.125
AM-Softmax	0.43	0.013	3.26	0.082
OC-Softmax	0.20	0.006	**2.19**	**0.059**

- OC-Softmax performs the best on unseen attacks.
- Achieved the state-of-the-art single-system performance.



Feature Embedding Visualization (red: bona fide, green: A17 attack, blue: spoofing attacks)

You Zhang, Fei Jiang, and Zhiyao Duan, “One-class Learning Towards Synthetic Voice Spoofing Detection”, *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021.

Comparison with single systems

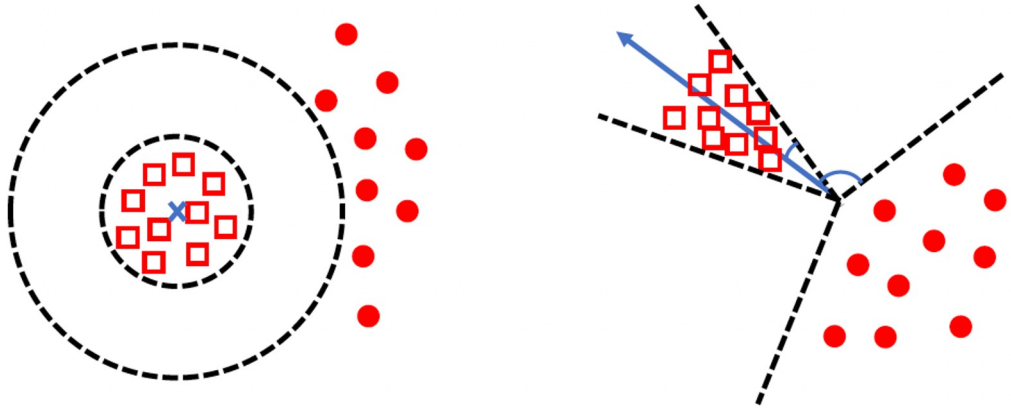
System	EER (%)	min t-DCF
CQCC + GMM [3]	9.57	0.237
LFCC + GMM [3]	8.09	0.212
Chettri et al. [22]	7.66	0.179
Monterio et al. [14]	6.38	0.142
Gomez-Alanis et al. [16]	6.28	-
Aravind et al. [18]	5.32	0.151
Lavrentyeva et al. [21]	4.53	0.103
ResNet + OC-SVM	4.44	0.115
Wu et al. [17]	4.07	0.102
Tak et al. [19]	3.50	0.090
Chen et al. [15]	3.49	0.092
Proposed	**2.19**	**0.059**

You Zhang, Fei Jiang, and Zhiyao Duan, “One-class Learning Towards Synthetic Voice Spoofing Detection”, *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021.

Other one-class loss functions

Euclidean distance-based one-class loss (isolate loss, single-center loss)

Cosine similarity-based one-class loss (OC-Softmax, angular isolate loss)



(a) Euclidean distance-based (b) Cosine similarity-based

You Zhang, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan. “Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks and Channel Variation”, *Handbook of Biometric Anti-spoofing (3rd edition)*, Springer, 2023.

Other one-class loss functions

Isolate loss:

$$L_{ISO} = \frac{1}{|\Omega|}\sum_{\Omega}\max\left(0, \|\boldsymbol{x}_i - \boldsymbol{c}\| - r_0\right) + \frac{1}{|\overline{\Omega}|}\sum_{\overline{\Omega}}\max\left(0, r_1 - \|\boldsymbol{x}_i - \boldsymbol{c}\|\right),$$

Single-center loss:

$$L_{SCL} = \frac{1}{|\Omega|}\sum_{\Omega}\| \boldsymbol{x}_i - \boldsymbol{c} \| + \max\left(0, \frac{1}{|\Omega|}\sum_{\Omega}\| \boldsymbol{x}_i - \boldsymbol{c} \| - \frac{1}{|\overline{\Omega}|}\sum_{\overline{\Omega}}\| \boldsymbol{x}_i - \boldsymbol{c} \| + \beta\sqrt{D}\right)$$

Angular isolate loss:

$$L_{AISO} = \frac{1}{|\Omega|}\sum_{\Omega}\log\left(1 + e^{\alpha\left(m_0 - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i\right)}\right) + \frac{1}{|\overline{\Omega}|}\sum_{\overline{\Omega}}\log\left(1 + e^{\alpha\left(\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i - m_1\right)}\right)$$

You Zhang, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan. “Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks and Channel Variation”, *Handbook of Biometric Anti-spoofing (3rd edition)*, Springer, 2023.

Follow-up works

ONE-CLASS KNOWLEDGE DISTILLATION FOR SPOOFIN

Jingze Lu[1,2], *Yuxiang Zhang*[1,2], *Wenchao Wang*[1], *Zengqiang Sha*

[1]Key Laboratory of Speech Acoustics and Content Understandi
Chinese Academy of Sciences, Beijing, Cl
[2]University of Chinese Academy of Sciences, Bei

{ Interspeech 2022
18-22 September 2022, Incheon, Korea

2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)
2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

Improved One-class Learning for Voice Spoofing Detection

Lixiang Li* † §, Xiaopeng Xue§ †, Haipeng Peng§ †, Yeqing Ren§ † and Mengmeng Zhao§ † ‡
State Key Laboratory of Networking and Switching Technology,
of Posts and Telecommunications, Beijing, China
ering Laboratory for Disaster Backup and Recovery,
of Posts and Telecommunications, Beijing, China
cience and Engineering, Zaozhuang University, Zaozhuang, China
onding author. Email: lixiang@bupt.edu.cn

A Deep One-Class Learning Method for Replay Attack Detection

Yijie Lou, Shiliang Pu, Jianfeng Zhou, Xin Qi, Qinbo Dong, Hongwei Zhou

Hikvision Research Institute, Hangzhou
louyijie@hikvision.com, pushiliang.hri@hi

IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 31, 2023

Generalized Voice Spoofing Detection via Integral Knowledge Amalgamation

Yeqing Ren, Haipeng Peng, Lixiang Li, Xiaopeng Xue, Yang Lan, and Yixian Yang

Research question

Motivation:

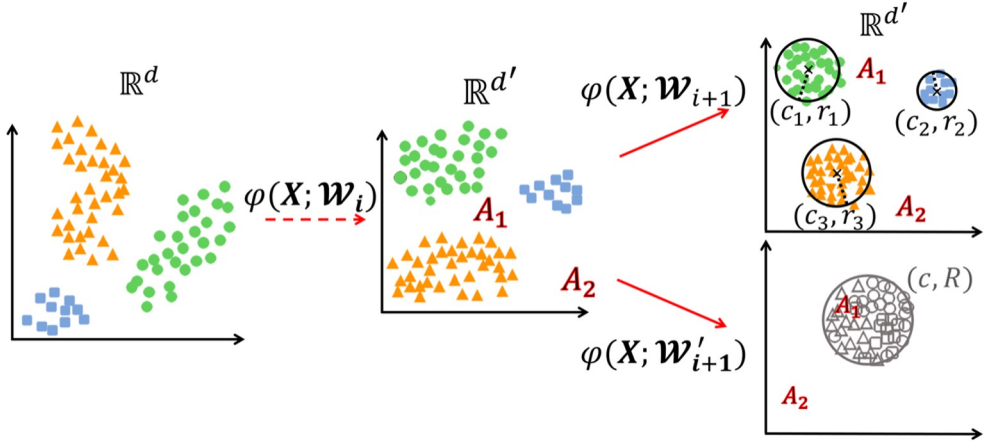
- In our previous work, we compact the embedding space of the bona fide speech into one cluster.
- However, due to **the variety of timbre and speaking traits of different speakers**, the bona fide speech of different speakers naturally forms multiple clusters in the embedding space.

How to improve the **generalization** ability while **maintaining the variation** of bona fide speech?

Inspiring work

Deep Multi-sphere Support Vector (SDM'20)

If data is naturally multi-cluster, merging them into one cluster could be harmful for detecting anomaly.



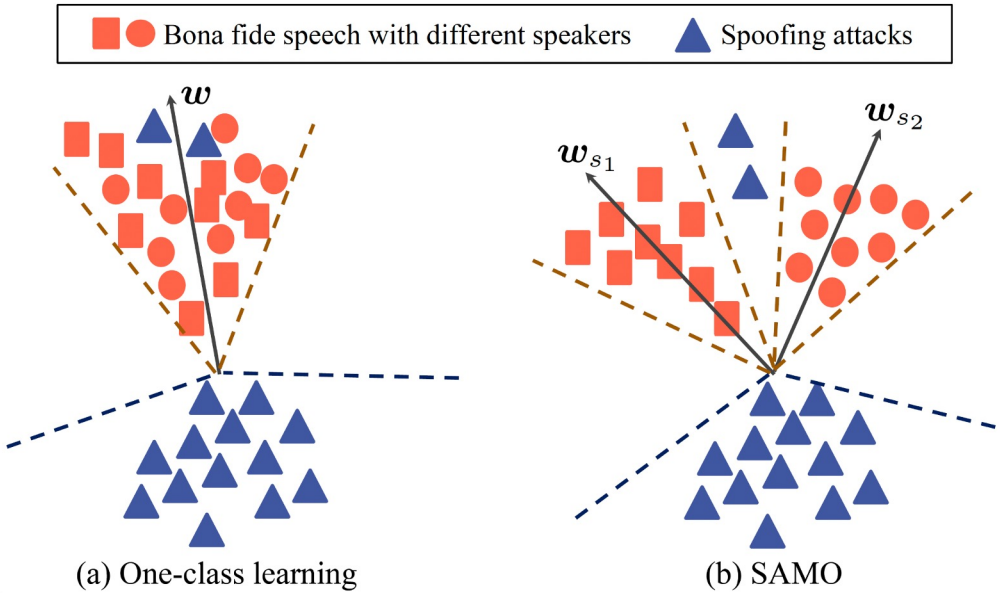
(Figure 2 in Ghafoori et al.)

Ghafoori, Z., & Leckie, C. (2020). Deep multi-sphere support vector data description. In *Proceedings of the 2020 SIAM International Conference on Data Mining* (pp. 109-117). Society for Industrial and Applied Mathematics.

Speaker attractor multi-center one-class learning

Model speaker diversity while maintaining the generalization ability brought by one-class learning

- Discriminate bona fide vs. spoofing attacks
- Cluster bona fide speech according to speakers



Siwen Ding, You Zhang, and Zhiyao Duan. "SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Speaker attractors

Define: a speaker-specific anchor in the embedding space

Compute: average the embeddings of each speaker's bona fide speech

- **Training:** attract bona fide speech embeddings of the same speaker
- **Inference:** cosine similarity between test utterance and enrolled utterance or attractors of training speakers

speaker attractor

embedding

speaker label of *i*-th utterance

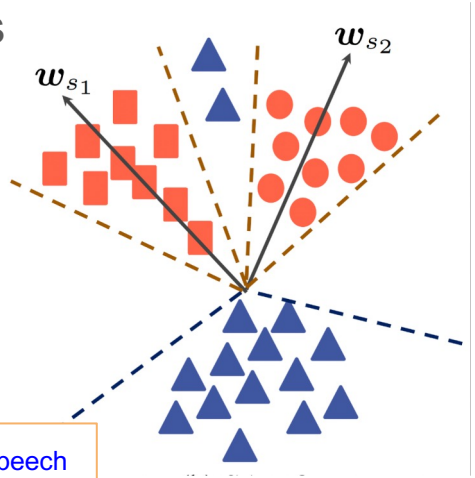
$$S_{SAMO} = \begin{cases} \hat{\boldsymbol{w}}_{s_i} \hat{\boldsymbol{x}}_i & \text{if } s_i \text{ is enrolled} \\ \max_{s} (\hat{\boldsymbol{w}}_s \hat{\boldsymbol{x}}_i), s \in \mathcal{S}_{train} & \text{otherwise} \end{cases},$$

speakers in the training set

Siwen Ding, You Zhang, and Zhiyao Duan. "SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Loss function for multi-center one-class learning

- Compact the bona fide speech representation belonging to the same speaker
- Push away the spoofing attacks from all speaker attractors



scale factor

label

$$\mathcal{L}_{SAMO} = \frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{\alpha(m_{y_i}-d_i)(-1)^{y_i}}\right),$$

samples

margin

where d_i is calculated by

$$d_i = \begin{cases} \hat{\boldsymbol{w}}_{\boldsymbol{s_i}}\hat{\boldsymbol{x}}_{\boldsymbol{i}} & \text{if } y_i = 0 \\ \max\limits_{s}(\hat{\boldsymbol{w}}_s\hat{\boldsymbol{x}}_i),\ s \in \mathcal{S}_{train} & \text{if } y_i = 1 \end{cases}$$

bona fide speech

spoofing attacks

Siwen Ding, You Zhang, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

SAMO training algorithm

- Compact the bona fide utterances spoken by the same speaker
- Push away spoofing utterances from all speaker attractors

Algorithm 1: SAMO Training Algorithm

Require: T: Total number of epochs
M: speaker attractor update interval (# epochs)

Initialize network F with random weights
Initialize speaker attractors $\boldsymbol{w}_s$ as one-hot vectors
for $i \leftarrow 1$ **to** T **do**
  if $i \bmod M = 0$ **then**
    Update $\boldsymbol{w}_s$ as the average bona fide embedding for each speaker $s \in \mathcal{S}_{train}$
  end if
  Update F by $\mathcal{L}_{SAMO}$ with mini-batches ▷ Eq. (3)
end for
return Optimized network F and speaker attractors $\boldsymbol{w}_s$

Siwen Ding, You Zhang, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Comparison with state-of-the-art methods

Table 2. Comparison of our proposed SAMO with Softmax and OC-Softmax on the target-only portion of the ASVspoof2019 LA evaluation set. All the systems use AASIST [9] backbone. The average (best) results across 3 trials with random training seeds are shown.

Method	EER(%)	min t-DCF
Softmax	1.74 (1.25)	0.0583 (0.0425)
OC-Softmax	1.25 (1.17)	0.0415 (0.0393)
SAMO (test w/o enrollment)	1.09 (0.91)	0.0363 (0.0306)
SAMO (test w/ enrollment)	**1.08 (0.88)**	**0.0356 (0.0291)**

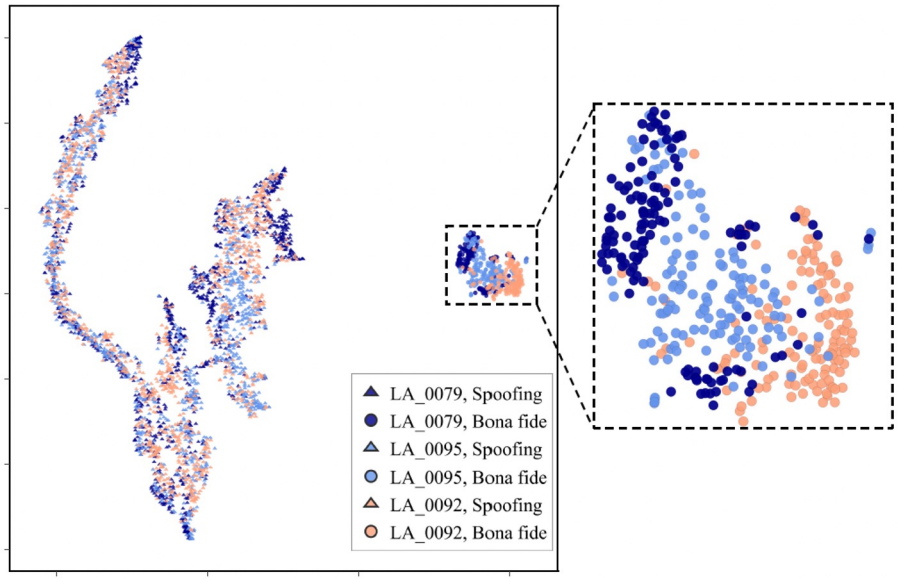
SAMO further improves the performance, indicating the advantage brought by the multi-center modeling of bona fide speech.

Siwen Ding, You Zhang, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Embedding visualization

2D t-SNE visualization of SAMO feature embeddings of bona fide and spoofed speech of three speakers

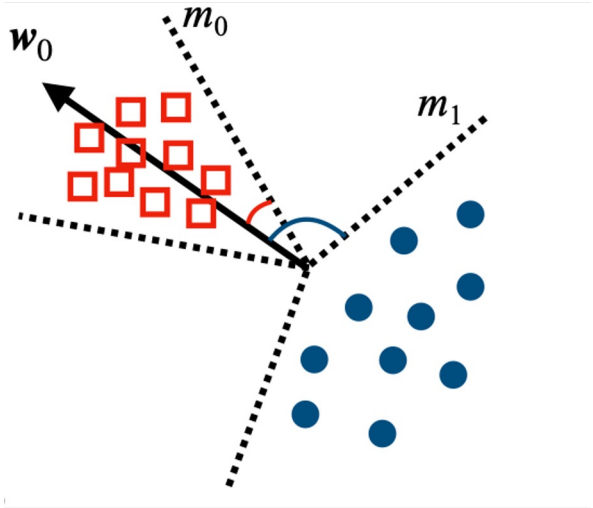
- Bona fide utterances are grouped in a small region.
- Utterances of the three speakers are generally clustered according to speaker identity.



Siwen Ding, You Zhang, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Takeaways

- One-class learning aims to **compact the target** class representation in the embedding space, and **push away non-target**.
- The proposed OC-Softmax and SAMO could improve the **generalization ability** of anti-spoofing system against **unseen spoofing attacks**.



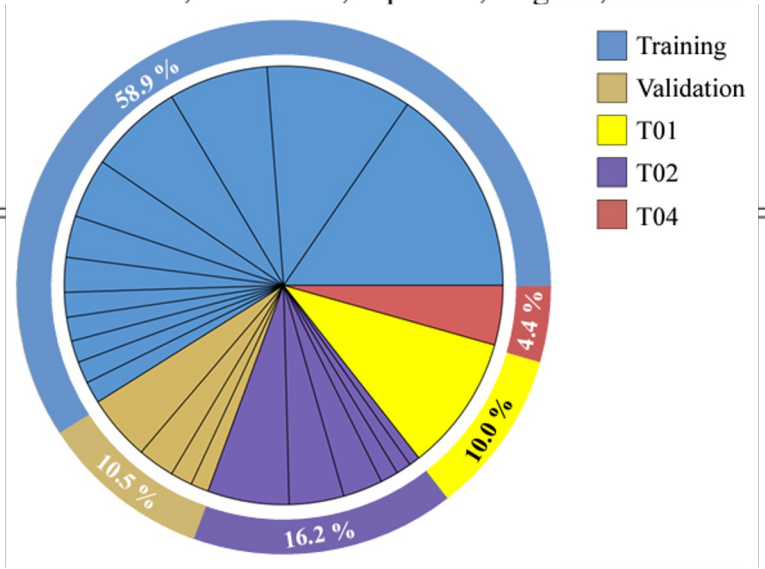
Demo of singing voice deepfakes

https://www.bilibili.com/video/BV1Sb4y1V76A

SingFake dataset: singing voice deepfake detection (SVDD)

Table 1. SingFake statistics for each split.

Splits	Description	# Singers	Languages (Sorted by percentages in the splits)	# Clips (Real / Fake)
Train	Training set	12	Mandarin, Cantonese, Japanese, English, Others	5251 / 4519
Val	Validation set (unseen singers)	4		1089 / 543
T01	Test set for seen singer Stefanie Sun	1		370 / 1208
T02	Test set for unseen singers	6		1685 / 1006
T03	T02 over 4 communication codecs	6		6740 / 4024
T04	Test set for Persian musical context	17		353 / 166

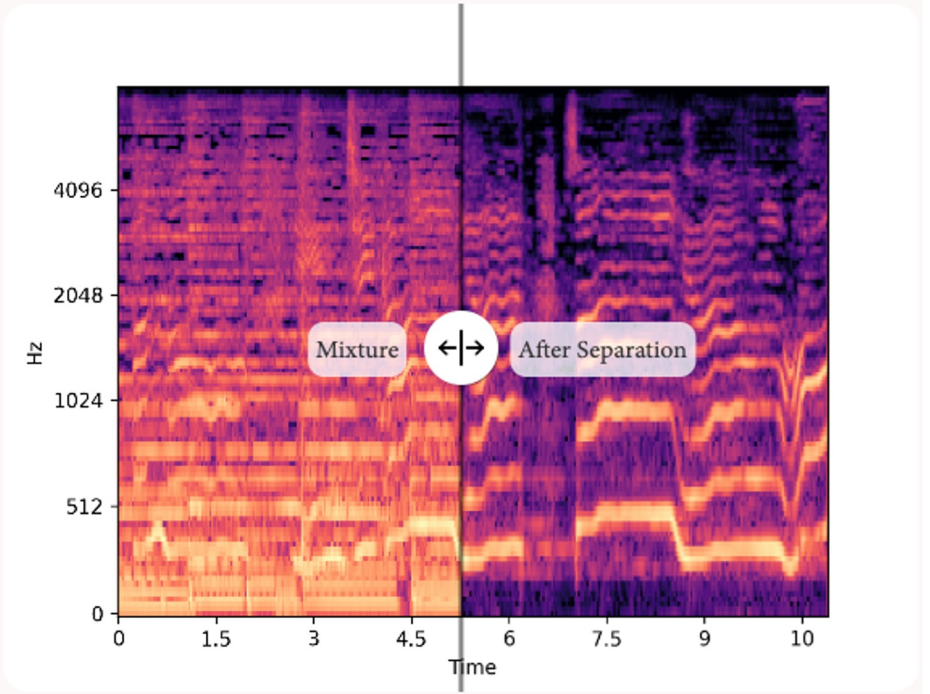


Yongyi Zang*, You Zhang*, Mojtaba Heydari, and Zhiyao Duan. “SingFake: Singing Voice Deepfake Detection”, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. (* equal contribution)

Speech anti-spoofing heavily degrades on SVDD task

Table 2. Test results on speech and singing voice with CM systems trained on speech utterance from ASVspoof2019LA (EER (%)).

Method	ASVspoof2019 LA - Eval	SingFake-T02 Mixture	SingFake-T02 Vocals
AASIST	0.83	58.12	37.91
Spectrogram+ResNet	4.57	51.87	37.65
LFCC+ResNet	2.41	45.12	54.88
Wav2Vec2+AASIST	7.03	56.75	57.26



Yongyi Zang*, You Zhang*, Mojtaba Heydari, and Zhiyao Duan. “SingFake: Singing Voice Deepfake Detection”, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. (* equal contribution)

Performance of training on the SingFake data

Table 3. Evaluation results for SVDD systems on all testing conditions in our SingFake dataset (EER (%))

Method	Setting	Train	T01	T02	T03	T04
AASIST	Mixture	4.10	7.29	11.54	17.29	**38.54**
	Vocals	3.39	8.37	10.65	13.07	43.94
Spectrogram+ResNet	Mixture	4.97	14.88	22.59	24.15	48.76
	Vocals	5.31	11.86	19.69	21.54	43.94
LFCC+ResNet	Mixture	10.55	21.35	32.40	31.85	50.07
	Vocals	2.90	15.88	22.56	23.62	39.27
Wav2Vec2+AASIST (Joint-finetune)	Mixture	**1.57**	**4.62**	**8.23**	13.62	42.77
	Vocals	1.70	5.39	9.10	**10.03**	42.19

Training on singing voices improves SVDD performance

SVDD systems show limited robustness to unseen scenarios

Yongyi Zang*, You Zhang*, Mojtaba Heydari, and Zhiyao Duan. “SingFake: Singing Voice Deepfake Detection”, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. (* equal contribution)

Singing voice deepfake detection (SVDD) challenge

- Controlled setting
  - Clean vocals generated by state-of-the-art singing voice synthesis (SVS) and singing voice conversion (SVC) systems based on open source pop song datasets.
- In-the-wild setting
  - Extended SingFake dataset.

Future directions

Generalizing to diversified spoofing attacks

-- Replay + TTS + VC + Adversarial + PartialSpoof

Robustness

-- Additive noise, channel variation, quality of TTS/VC systems (In-the-wild)

Explainability

-- The artifacts or the cues that distinguish bona fide from spoofed speech

Visually-informed speech anti-spoofing

-- Audio-visual deepfake detection

Takeaways

Introduction to speech anti-spoofing

Generalization ability to unseen synthetic attacks

- One-class learning: OC-Softmax,
- Multi-center one-class learning: SAMO

Beyond speech anti-spoofing: **Singing voice** deepfake detection

Thank you! Questions?

References

[1] **You Zhang**, Fei Jiang, and Zhiyao Duan, ``One-class Learning Towards Synthetic Voice Spoofing Detection", *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021. [link][code][video]

[2] **You Zhang**, Ge Zhu, Fei Jiang, and Zhiyao Duan, ``An Empirical Study on Channel Effects for Synthetic Voice Spoofing Countermeasure Systems", in *Proc. Interspeech*, pp. 4309-4313, 2021. [link][code][video]

[3] Xinhui Chen*, **You Zhang***, Ge Zhu*, and Zhiyao Duan, ``UR Channel-Robust Synthetic Speech Detection System for ASVspoof 2021", in *Proc. ASVspoof 2021 Workshop*, pp. 75-82, 2021. (* equal contribution) [link][code][video]

[4] **You Zhang**, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan, ``Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks and Channel Variation", *Handbook of Biometric Anti-spoofing (3rd Ed.)*, Springer, 2023. [link][code]

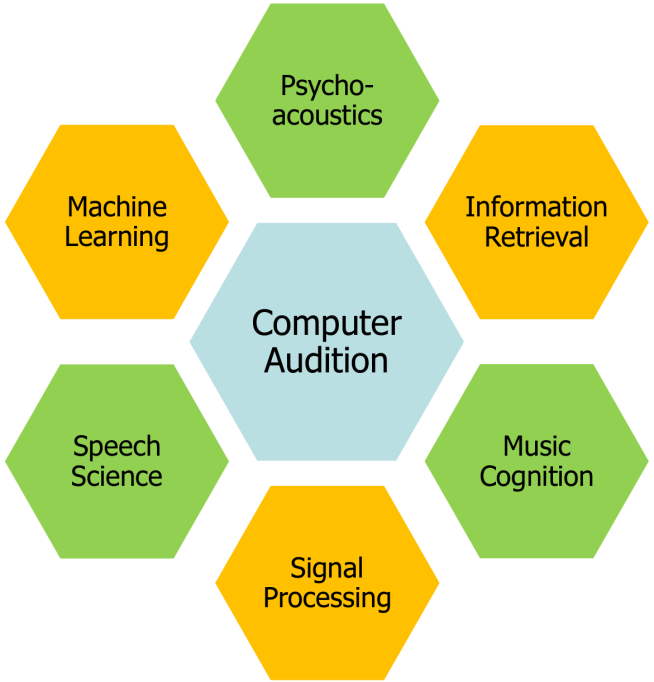
[5] **You Zhang**, Ge Zhu, and Zhiyao Duan, ``A Probabilistic Fusion Framework for Spoofing Aware Speaker Verification", in *Proc. Odyssey*, 2022. [link][code]

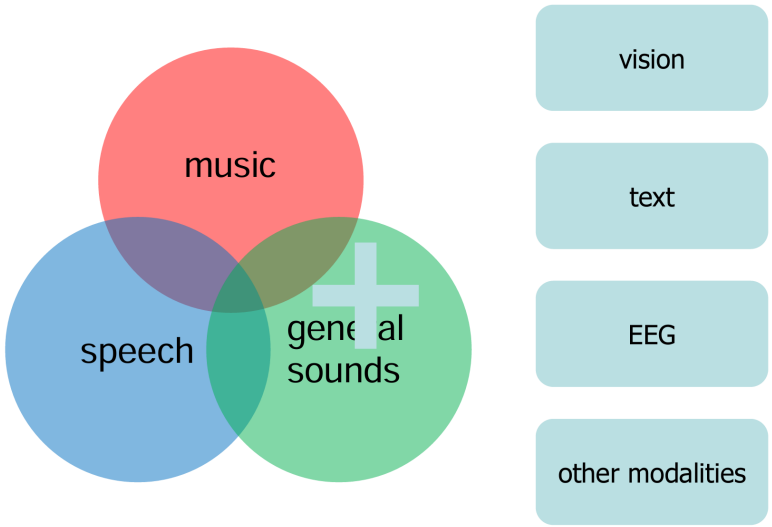
[6] Siwen Ding, **You Zhang**, and Zhiyao Duan. ``SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing'', in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023. [link][code]

[7] Yongyi Zang, **You Zhang**, and Zhiyao Duan. ``Phase Perturbation Improves Channel Robustness for Speech Spoofing Countermeasures", in *Proc. Interspeech*, pp. 3162-3166, 2023. [link][code]

[8] Yongyi Zang*, **You Zhang***, Mojtaba Heydari, and Zhiyao Duan. ``SingFake: Singing Voice Deepfake Detection'', in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. (* equal contribution) [link][code][webpage]

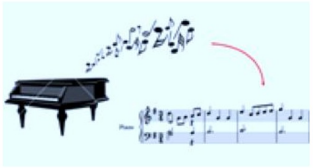
Computer Audition Research Areas





Audio Information Research (AIR) Lab

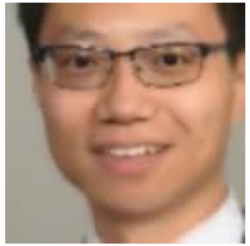
Machine Understanding of Sounds



MUSIC INFORMATION RETRIEVAL
- Music transcription, alignment
- Source separation
- Generation
- Interactive performance

SPEECH PROCESSING
- Separation and enhancement
- Verification and anti-spoofing
- Emotion analysis
- Diarization
- Text-to-speech
- Voice conversion

ENVIRONMENTAL SOUND UNDERSTANDING
- Sound search by vocal imitation
- Sound event detection
- Source localization
- HRTF modeling
- Smart acoustics



AUDIO-VISUAL PROCESSING
- Talking face generation
- Music performance analysis and generation
- Audio-visual source separation