arXiv:2602.01722v1 [eess.AS] 2 Feb 2026

ASV 与 CM 任务的联合优化：BTUEF 团队参加 WildSpoof 挑战赛的提交方案

O˘古兹汗·库尔纳兹[1], *Jagabandhu Mishra*[2],*Tomi Kinnunen*[2],杰马尔·哈尼尔奇,伊[1]

[1] 布尔萨技术大学电气与电子工程系[2] 东芬兰大学计算机学院

摘要

欺骗感知说话人验证（SASV）将自动说话人验证与反欺骗对策结合起来，以提高抵御对抗攻击的鲁棒性。本文研究了我们近期提出的模块化联合说话人验证与反欺骗框架。该框架通过非线性融合实现公开可用的自动说话人验证和反欺骗对策系统的有效复用，明确对两者的交互进行建模，并采用依赖工作条件的可训练 a-DCF 损失进行优化。我们使用 ECAPA-TDNN 和 ReDimNet 作为自动说话人验证嵌入提取器，并使用 SSL-AASIST 作为欺骗对抗模型进行评估；实验分别在 WildSpoof 欺骗感知说话人验证训练数据上进行微调和不进行微调的情况下开展。结果表明，将基于 ReDimNet 的自动说话人验证嵌入与微调后的 SSL-AASIST 表征相结合可获得最佳性能，在进度评估集和最终评估集上的 a-DCF 分别为 0.0515 和 0.2163。

索引词— 欺骗感知说话人验证，WildSpoof，反欺骗对策

1. 引言

自动说话人验证（ASV） [1] 系统广泛应用于安全关键型应用，但仍容易受到重放攻击 [2], 、文本转语音和语音转换等欺骗攻击 [3]。尽管专用的反欺骗对策（CM）系统能够检测伪造语音，但它们无法验证说话人身份，因此在对抗条件下，独立的自动说话人验证或欺骗对抗系统并不足够 [4, 5]。这促使研究者提出欺骗感知说话人验证（SASV），以联合处理说话人验证和欺骗检测。近期的评测活动，包括 SASV2022 挑战赛 [6],、ASVspoof 5（赛道 2）挑战赛 [7], 以及 WildSpoof 挑战赛 [8], ，进一步凸显了对灵活且稳健的欺骗感知说话人验证架构的需求，使其能够适应非受限欺骗场景。

在本研究中，我们扩展了近期提出的统一的欺骗感知说话人验证框架 [9] ，并将其作为受控实验测试平台，系统分析欺骗感知说话人验证中的嵌入级设计选择。具体而言，我们定量评估不同自动说话人验证与反欺骗对策嵌入提取器在冻结与微调两种训练机制下的影响。所有配置均在经校准的非线性分数级融合框架 [10]内运行，该框架使用由用户指定运行条件 [11]定义的联合说话人反欺骗验证目标函数进行优化。

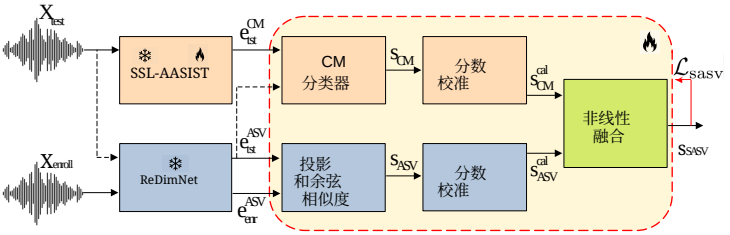


图1。所提出的模块化说话人反欺骗验证架构采用可训练的分数级集成。基于预训练 ReDimNet 的自动说话人验证提取器被冻结；在 SSL-AASIST 分支中，仅训练 AASIST 组件，同时冻结自监督学习编码器。自动说话人验证和反欺骗对策嵌入被转换为校准后的分数，红色虚线框内的所有模块均可训练，并通过非线性融合生成最终的欺骗感知说话人验证分数。

2. 所提方法

如图1所示，所提系统由三个主要组件构成——自动说话人验证分支、反欺骗分支和分数融合模块。

自动说话人验证分支： 给定注册语音和测试语音，$\mathbf{X}_{\text{enr}}$ 和$\mathbf{X}_{\text{tst}}$，固定的自动说话人验证编码器提取说话人嵌入$\mathbf{e}_{\text{enr}}^{\text{asv}}$ 和 $\mathbf{e}_{\text{tst}}^{\text{asv}}$。使用加权余弦相似度计算说话人相似度，其中可学习向量$\mathbf{w}_{\text{asv}}$对嵌入维度重新加权，得到$\mathbf{e}_1 = \mathbf{w}_{\text{asv}} \odot \mathbf{e}_{\text{enr}}^{\text{asv}}$ 和$\mathbf{e}_2 = \mathbf{w}_{\text{asv}} \odot \mathbf{e}_{\text{tst}}^{\text{asv}}$。随后对自动说话人验证分数进行仿射校准，得到 $s_{\text{asv}}^{\text{cal}} = w_0^{\text{asv}} + w_1^{\text{asv}} s_{\text{asv}}$，从而生成适用于融合且易于解释的对数似然比（LLR）。

反欺骗分支： 为引入欺骗感知，使用冻结的反欺骗编码器从测试语句中提取反欺骗嵌入 $\mathbf{e}_{\text{tst}}^{\text{cm}}$ 。该嵌入与自动说话人验证测试嵌入拼接，形成联合表示 $\mathbf{e}_{\text{fused}} = [\mathbf{e}_{\text{tst}}^{\text{asv}}; \mathbf{e}_{\text{tst}}^{\text{cm}}]$，并将其输入基于 MLP 的反欺骗分类器中。

分类器输出的反欺骗分数经过仿射校准，得到 $s_{\text{cm}}^{\text{cal}} = w_0^{\text{cm}} + w_1^{\text{cm}} s_{\text{cm}}$，确保其与自动说话人验证对数似然比兼容。

分数融合： 校准后的自动说话人验证与反欺骗对数似然比通过非线性分数级融合进行组合，从而显式建模二者之间的交互，

$$s_{\text{sasv}} = -\log\left[(1-\tilde{\rho})e^{-s_{\text{asv}}^{\text{cal}}} + \tilde{\rho}e^{-s_{\text{cm}}^{\text{cal}}}\right], \qquad (1)$$

其中 $\tilde{\rho} \in [0,1]$ 控制说话人证据与欺骗证据的相对贡献。这种方法生成单一的联合说话人验证与反欺骗分数，反映对目标身份和真实语音的联合置信度。

联合优化： 所有可训练组件——嵌入重加权、校准层、CM分类器和融合模块——均在联合说话人验证与反欺骗决策层进行端到端优化。只有当试验同时符合目标说话人和真实语音条件时，才将其标记为正例。训练采用二元交叉熵与架构无关检测代价函数（a-DCF） [12], 的加权组合，在保持系统模块化设计的同时，直接与联合说话人验证与反欺骗评估标准对齐。

3. 实验设置

在本次竞赛中，首先使用WildSpoof数据集训练组织方正式提供的基线模型——即**ECAPA-TDNN**、**MFA-Conformer**和**SKA-TDNN**[1]——以获得参考结果。

在这些基线结果的基础上，我们将最先进的自动说话人验证模型整合到所提出的统一欺骗感知说话人验证框架中。对于说话人嵌入提取，我们评估了**ECAPA-TDNN**[2] 和**ReDimNet**[3] 架构在三种自动说话人验证配置下的表现：(i) 预训练且冻结的*ECAPA-TDNN* ，(ii) 在*WildSpoof* 数据集上微调的 *ECAPA-TDNN* 模型，以及 (iii) 预训练且冻结的*ReDimNet* 。我们使用每种配置从 WildSpoof 数据集中提取说话人嵌入，从而在统一的欺骗感知说话人验证框架下进行受控的比较评估。

对于反欺骗分支，我们采用**SSL-AASIST**[4] 模型。在此设置中，仅使用 WildSpoof 数据集的训练集微调基于 AASIST 的欺骗检测组件，而自监督前端保持冻结。随后，在所提出的说话人反欺骗验证架构中使用得到的反欺骗嵌入。

所提出的联合说话人验证与反欺骗后端采用模块化网络实现，由两个全连接隐藏层组成，分别包含 384 和 160 个神经元。整个欺骗感知说话人验证模型训练 300 个周期，批次大小为 192，并使用 0.005 的学习率。式 1 中 $\tilde{\rho}$ 的值设为 0.5。为确保不同自动说话人验证嵌入骨干网络和融合配置之间的比较公平，所有实验均固定使用这些超参数。

4. 结果

ECAPA-TDNN、多尺度特征聚合卷积器（MFA-Conformer）和基于选择性核注意力的 TDNN（SKA-TDNN）模型对应于组织方提供的官方基线系统，并严格按照其原始配置进行训练，分别得到开发集 a-DCF 参考结果 0.5113、0.4453 和 0.3118，如表 1 所示。在这些基线系统中，SKA-TDNN 在开发集上取得了最佳性能，其评估集 a-DCF 随后计算为 0.3821。随后，对组织方提供的 ECAPA-TDNN 进行训练，并将其与在 ASVspoof 5 上训练的预训练 SSL-AASIST 模型结合，得到开发集 a-DCF 0.1193；当对 SSL-AASIST 组件进行微调后，该值进一步降至 0.0985。使用经过微调的 SSL-AASIST 嵌入，并分别搭配预训练的 ECAPA-TDNN 或预训练的 ReDimNet 骨干网络进行的额外实验，得到的开发集 a-DCF 分别为 0.1500 和 0.0470。最佳系统基于预训练的 ReDimNet 和经过微调的 SSL-AASIST，在评估集和最终评估集上的 a-DCF 分别达到 0.0515 和 0.2163。

表 1。开发集、评估集和最终评估集的综合 a-DCF 结果。上方部分展示基线系统，下方部分列出提出的欺骗感知说话人验证配置。T：从头训练；PT：预训练；FT：微调。

模型	Dev	Eval	最终评估集
基线系统			
ECAPA-TDNN	0.5113	N/A	N/A
MFA-Conformer	0.4453	N/A	N/A
SKA-TDNN	0.3118	0.3821	N/A
提出的欺骗感知说话人验证			
T-ECAPA + PT-SSL-AASIST	0.1193	N/A	N/A
T-ECAPA + FT-SSL-AASIST	0.0985	N/A	N/A
PT-ECAPA + FT-SSL-AASIST	0.1500	N/A	N/A
PT-ReDimNet + FT-SSL-AASIST	**0.0470**	**0.0515**	**0.2163**

5. 结论

我们研究了基线系统，发现 SKA-TDNN 在其中表现最佳。采用预训练 ECAPA-TDNN 和 SSL-AASIST 的所提系统优于基线系统，通过微调还能进一步提升性能。将 ECAPA-TDNN 替换为 ReDimNet 作为自动说话人验证嵌入提取器，则可取得额外提升。最终评估集上的表现明显低于评估集，这可能是由于难度更高或存在未见过的条件；待最终评估集标签发布后，我们将进一步分析。

[1]基线模型 GitHub 仓库：https://github.com/wildspoof/SASV_基线 [2]ECAPA-TDNN GitHub 仓库：https://github.com/ TaoRuijie/ECAPA-TDNN[3]ReDimNet GitHub 仓库：https://github.com/IDRnD/ redimnet[4]SSL-AASIST GitHub 仓库：https://github.com/ TakHemlata/SSL_反欺骗

6. 参考文献

[1] Douglas A. Reynolds，“使用高斯混合说话人模型进行说话人识别与验证”，载于自动说话人识别、识别与验证 *ESCA* 研讨会论文集，1994年，第 27–30 页。

[2] Nicolas M¨uller、Piotr Kawa、Wei-Herng Choong、Adri-ana Stan、Aditya Tirumala Bukkapatnam、Karla Pizzi、Alexander Wagner 和 Philip Sperl，“针对音频深度伪造检测的重放攻击”，载于*Proc. Inter-speech2025*，第 2245–2249 页。

[3] Rohan Kumar Das、Xiaohai Tian、Tomi Kinnunen 和 Haizhou Li，“自动说话人验证中的攻击者视角：综述”，载于 *Interspeech 2020* 论文集，第 4213–4217 页。[4] O˘guzhan Kurnaz 、 Selim Can Demirtas¸、Aykut B¨uker、Jagabandhu Mishra 和 Cemal Hanilc¸i，“采用并行嵌入融合的抗欺骗说话人验证：BTU语音组的 ASVspoof5挑战赛方案”，载于自动说话人验证欺骗对策研讨会（*ASVspoof 2024*），第 138–143 页。[5] Aykut B¨uker、O˘guzhan Kurnaz、S¸ule Bekiryazıcı、Selim Can Demirtas¸ 和 Cemal Hanilc¸i，“评估欺骗感知说话人验证中的参数共享：基于 ASVspoof 5数据集的案例研究”，载于*Interspeech 2025* 论文集，第 4573–4577 页。[6]Jee weon Jung、Hemlata Tak、Hye jin Shim、Hee-Soo Heo、Bong-Jin Lee、Soo-Whan Chung、Ha-Jin Yu、Nicholas Evans 和 Tomi Kinnunen，“SASV 2022：首届欺骗感知说话人验证挑战赛”，载于 *Interspeech 2022* 论文集，第 2893–2897 页。[7] Xin Wang、H´ector Delgado、Hemlata Tak、Jee weon Jung、Hye jin Shim、Massimiliano Todisco、Ivan Kukanov、Xuechen Liu、Md Sahidullah、Tomi H. Kinnunen、Nicholas Evans、Kong Aik Lee 和 Junichi Yamagishi，“ASVspoof 5：大规模众包语音数据、深度伪造与对抗攻击”，载于 自动说话人验证欺骗对策研讨会论文集（*ASVspoof 2024*），第 1–8 页。[8] Yihan Wu、Jee-weon Jung、Hye-jin Shim、Xin Cheng 和 Xin Wang，“Wildspoof 挑战赛评测计划”，*arXiv* 预印本 *arXiv:2508.16858*，2025。[9] O˘guzhan Kurnaz、Jagabandhu Mishra、Tomi H Kinnunen 和 Cemal Hanilc¸i，“联合优化说话人检测器与欺骗检测器，实现抗欺骗自动说话人验证”，*arXiv* 预印本 *arXiv:2510.01818*，2025。

[10]Xin Wang、Tomi Kinnunen、Kong Aik Lee、Paul-Gauthier No´e 和 Junichi Yamagishi，“重新审视并改进基于成分数据分析的欺骗感知说话人验证评分融合”，载于*Interspeech 2024* 论文集，第 1110–1114 页。[11] O˘guzhan Kurnaz、Jagabandhu Mishra、Tomi H. Kinnunen 和 Cemal Hanilc¸i，“优化抗欺骗说话人验证的 a-DCF”，*IEEE* 信号处理快报，第 32 卷，第 1081–1085 页，2025 年。

[12] Hye jin Shim、Jee weon Jung、Tomi Kinnunen、Nicholas Evans、Jean-Franc¸ois Bonastre 和 Itshak Lapidot，“a-DCF：一种架构无关度量及其在抗欺骗说话人验证中的应用”，载于*Proc.Odyssey2024*，第 158–164 页。