Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing

摘要

本文提出了一种新的多模态面部反欺骗 (FAS) 方法，旨在提高面部识别系统对呈现攻击的安全性。随着传感器制造和多模态学习技术的进步，多模态 FAS 方法不断涌现，但它们在泛化到未知攻击和部署条件时面临挑战。这些挑战主要来自于两个方面：模态不可靠性和模态不平衡。模态不可靠性指的是某些模态传感器，如深度和红外传感器，在不同环境中会发生显著的领域偏移，导致跨模态特征融合时传播不可靠的信息。模态不平衡则是指训练过程中过度依赖主导模态，阻碍了其他模态的收敛，降低了对那些仅使用主导模态无法区分的攻击类型的有效性。为了解决模态不可靠性问题，本文提出了不确定性引导的交叉适配器 (U-Adapter)，用于识别每个模态中不可信检测区域，并抑制这些区域对其他模态的影响。针对模态不平衡问题，本文提出了重新平衡模态梯度调制 (ReGrad) 策略，通过自适应调整梯度来重新平衡所有模态的收敛速度。此外，本文还提供了第一个大规模基准测试，用于评估多模态 FAS 在领域泛化场景下的性能。

关键词： 多模态面部反欺骗; 不确定性估计

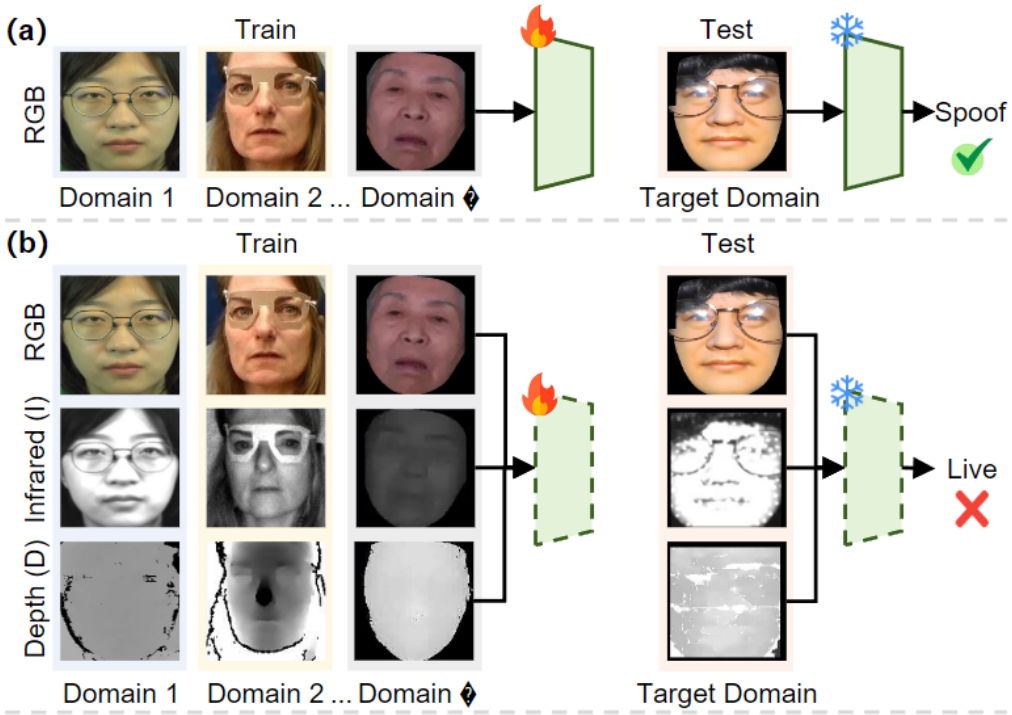
1 引言

在面部识别技术迅猛发展的今天，其安全性问题也日益凸显，尤其是面部反欺骗（FAS）领域。该文章针对面部识别系统中存在的安全漏洞，尤其是在监控和移动支付等关键应用中，提出了一种创新的多模态 FAS 方法。这些系统容易受到打印照片、视频重放和 3D 面具等面部呈现攻击的威胁，这些攻击严重挑战了 FR 系统的安全性，并限制了其在更广泛场景中的应用。

文章中提出的多模态 FAS 方法，旨在通过结合 RGB、深度和红外等多种模态的数据，来提高系统的泛化能力和鲁棒性。然而如图1所示，现有的多模态 FAS 方法在面对未知部署条件时，其性能并不理想。这主要归因于两个问题：模态不可靠性和模态不平衡。模态不可靠性是指在不同环境或使用不同型号的传感器时，某些模态可能会经历显著的领域偏移，导致特征提取的不可靠性。模态不平衡则是指模型在多模态学习过程中可能会过度依赖于某个收敛速度较快的主导模态，而忽视了其他收敛速度较慢的模态。

为了解决这些问题，该文章提出了一种新的框架，包括不确定性引导的交叉适配器（U-Adapter）和重新平衡模态梯度调制（ReGrad）策略。U-Adapter 通过利用每个模态的不确定性信息，减少不可靠区域对其他模态的影响。ReGrad 策略则通过动态调整梯度，平衡各模态的收敛速度，确保所有模态都能在抵抗未知攻击中发挥作用。

该文章的研究具有重要的实际应用价值，为提高面部识别系统的安全性提供了一种有效的解决方案。文章中的方法在多个基准数据集上进行了广泛的实验验证，显示出优越的性能和泛化能力。此外，文章还提供了开源代码，为复现工作提供了便利。通过复现该文章的工作，深入理解多模态 FAS 的原理和实现，这不仅有助于我掌握相关技术，也有助于我在未来的研究中应用这些知识。通过复现，我期待能够验证文章中提出的方法，并探索其在不同场景下的应用潜力。



(c)

Method	Modality	HTER (%) ↓	AUC (%) ↑
SSDG	(a): RGB	**24.07**	**83.69**
SSDG*	(b): RGB+D+I	26.09 (+2.02)	82.03 (-1.66)

图 1. (a) 单模态和 (b) 多模态情况下的 DG 场景说明。(c) SSDG 的 DG 性能。虽然 SSDG 可以处理更多模态，但与单模态相比，它在多模态场景中的表现更差。* 表示文章中重新实现的多模态版本。

2 相关工作

2.1 域泛化

面部反欺骗（FAS）中的域泛化（Domain Generalization）专注于训练模型，使其能够在多个源域上学习并有效应对未见的目标域 [6, 18]。研究已证明对抗性训练、非对称三元组损失和争议学习在构建跨域共享特征空间方面的有效性 [8, 9, 13, 15, 16, 21, 24]。此外，通过最小化风格特征差异来学习领域不变特征的方法，以及基于元学习的策略，都在模拟领域偏移和学习鲁棒特征空间方面显示出潜力 [2, 4, 7, 22, 25, 26]。最新的研究强调了参数高效的迁移学习，这使得预训练的视觉变换器（ViT）能够更好地适应新领域并减少过拟合 [1, 3, 6, 20]。然而，这

些主要针对单模态（如 RGB）FAS 的方法在多模态场景下存在局限性，尤其是在处理模态不可靠性方面。

2.2 多模态 FAS

多模态 FAS 通过结合 RGB、深度和红外等多种光谱来检测活体和欺骗痕迹，利用不同模态间的互补信息来提高检测的准确性 [14]。早期方法通过通道级联或独立分支进行特征提取和晚期融合来整合多模态信息 [5,10,11,17,19,23]。近期研究进一步引入了基于注意力的特征融合技术和自适应跨模态损失函数，以增强模态间的信息互补性 [2,12]。跨模态转换技术也被提出，用以缩小不同模态间的语义差距，从而提升 FAS 系统的整体性能。这些进展展示了多模态 FAS 在提高面部识别安全性方面的潜力。

3 本文方法

3.1 U-Adapters

如图2所示，U-Adapters 是为解决多模态 FAS 中模态不可靠性问题而提出的。在多模态 FAS 领域，尤其是在领域泛化场景下，显著的领域偏移可能导致某些模态输入提取出不可靠的活体/欺骗特征。这些不可靠的特征在跨模态融合过程中传播，对其他模态产生不利影响。为了解决这一问题，U-Adapters 通过不确定性估计来识别每个模态内不可靠的局部特征，并在基于注意力的跨模态特征融合中降低这些特征的权重，从而减少它们的负面影响。

U-Adapters 采用蒙特卡洛采样来估计不确定性，并基于此不确定性进行跨模态特征融合。这种方法受到参数高效迁移学习（PETL）的启发，通过在预训练的 ViT 中引入适配器来进行微调。U-Adapters 的设计允许模型在跨模态融合过程中识别并抑制不可靠信息的传播，从而提高模型对不同模态的适应性和平衡性。在微调过程中，U-Adapters 通过不确定性估计模块（UEM）来识别不可靠标记，并在融合过程中抑制这些标记的交互，确保融合特征的可靠性。

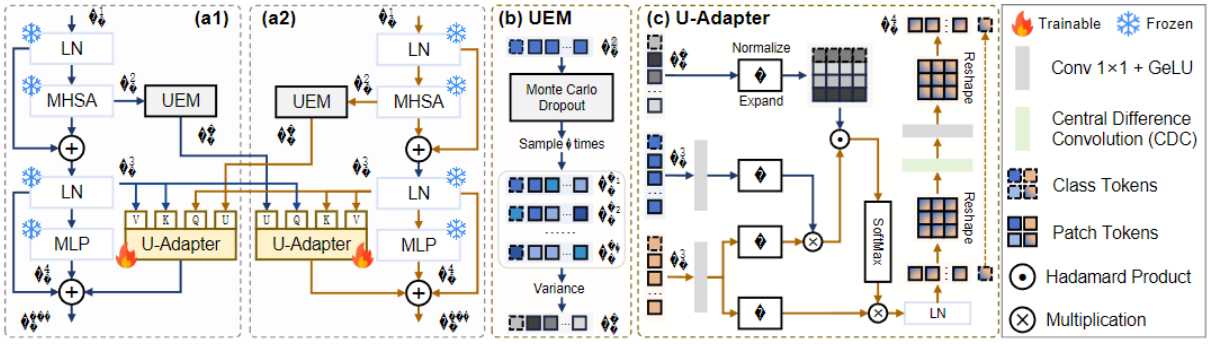


图 2. (a1)-(a2) 利用提出的 U-Adapters 对 ViT 进行微调的图示，展示了 RGB (R) 和深度 (D) 模式之间的互动。请注意，只有 U-Adapters 的参数是可训练的。(b) 用于识别不可靠标记的不确定性估计模块（UEM）。(c) U-Adapter 的详细结构，它采用跨模态融合技术，可抑制不可靠标记对其他模态的干扰。融合后，再整合中心差分信息，以实现细粒度的欺骗表示。

3.2 ReGrad

如图3所示，ReGrad 策略是为了解决多模态 FAS 中的模态不平衡问题而设计的。在面对各种未知攻击和部署条件时，ReGrad 旨在确保每个模态都能充分收敛，以充分利用它们的辨别能力，而不是过度依赖在源域中表现出色的特定模态。ReGrad 通过自适应地调制所有模态的梯度来实现这一点，这些梯度同时作用于每个可训练层。

ReGrad 策略考虑了模态之间的冲突和收敛程度，通过分解梯度成分来调整模态的学习速度。如果两个模态的梯度不冲突，ReGrad 会抑制收敛较快模态的梯度，以加速较慢模态的学习。如果梯度冲突，ReGrad 会去除与较慢模态梯度方向相反的部分，以防止较快模态的梯度严重减慢较慢模态的收敛速度。此外，ReGrad 还采用了基于不确定性的抑制方法，以防止不可靠模态误导其他模态。

为了进一步平衡模态，ReGrad 引入了单边原型损失（SSP）来监督模态分支并监控它们的收敛速度。SSP 损失鼓励每个模态的特征被吸引到相应域的原型，从而通过内部动力优化每个模态分支。最终的损失函数结合了交叉熵损失和 SSP 损失，以确保模型性能并增强泛化能力。

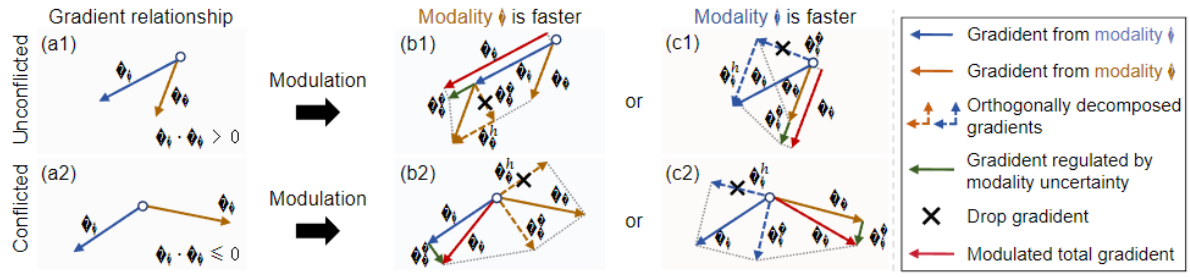


图 3. 通过提出的的 ReGrad 在不同情况下进行梯度调制的示意图：(第 1 行）非冲突（a1）和更快的模式 j（b1）或 i（c1）。(第 2 行）冲突（a2）和更快的模式 j (b2) 或 i (c2)。

4 复现细节

4.1 与已有开源代码对比

原文代码是开源的，但只开源了关键部分，本次复现便针对开源代码进行复现实验，探索可行的参数与训练设置，复现结果与原文结果大体相当。

4.2 实验环境搭建

编程语言使用 Python。使用 Pytorch 框架。

4.3 创新点

对 ReGrad 的梯度收集与计算部分代码进行了优化，使得代码更加精简和易读，同时保留其原效果，并且运行速度上有所提升。

Method	$CPS \rightarrow W$	$CPW \rightarrow S$	$CSW \rightarrow P$	$PSW \rightarrow C$	Average
ViT (Baseline)	20.88	44.05	33.58	42.15	35.16
MMDG (原文结果)	12.79	15.32	18.95	29.93	19.25
MMDG (复现结果)	14.02	18.44	15.76	33.24	27.86

表 1. 复现协议一的 HTER 结果指标

Method	Missing D	Missing I	Missing D & I	Average
ViT (Baseline)	40.04	36.77	36.20	37.67
MMDG (原文结果)	24.89	23.39	25.26	24.51
MMDG (复现结果)	25.77	25.84	28.33	26.64

表 2. 复现协议二的 HTER 结果指标

5 实验结果分析

复现结果如表1、2、3所示，复现结果与原结果大体相当。

6 总结与展望

本次计算机前沿技术论文复现中，复现了论文的模型，并获得了预期的效果和指标。同时通过此次复现，了解关于该领域多模态学习中的一些关键问题和解决思路，并且加深了对不确定性理论和多模态梯度理论的理解。由于该工作各模块的工作原理绑定得比较紧密，所以并没成功实现替换或改进部分模块来进一步提升性能，这是本次复现的不足之处。未来可以考虑将这些理论应用到 CLIP、LLM 等基础架构中来进一步实现更好的效果。

Method	$CW \rightarrow PS$	$PS \rightarrow CW$
ViT (Baseline)	42.66	42.75
MMDG (原文结果)	20.12	36.60
MMDG (复现结果)	23.81	39.46

表 3. 复现协议三的 HTER 结果指标

参考文献

[1] Rizhao Cai, Yawen Cui, Zhi Li, Zitong Yu, Haoliang Li, Yongjian Hu, and Alex Kot. Rehearsal-free domain continual face anti-spoofing: Generalize more and forget less. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8037–8048, 2023.

[2] Rizhao Cai, Zhi Li, Renjie Wan, Haoliang Li, Yongjian Hu, and Alex C Kot. Learning meta pattern for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 17:1201–1213, 2022.

[3] Rizhao Cai, Zitong Yu, Chenqi Kong, Haoliang Li, Changsheng Chen, Yongjian Hu, and Alex C Kot. S-adapter: Generalizing vision transformer for face anti-spoofing with statistical tokens. *IEEE Transactions on Information Forensics and Security*, 2024.

[4] Zhekai Du, Jingjing Li, Lin Zuo, Lei Zhu, and Ke Lu. Energy-based domain generalization for face anti-spoofing. In *Proceedings of the 30th ACM international conference on multimedia*, pages 1749–1757, 2022.

[5] Anjith George and Sébastien Marcel. Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks. *IEEE Transactions on Information Forensics and Security*, 16:361–375, 2020.

[6] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *European conference on computer vision*, pages 37–54. Springer, 2022.

[7] Yunpei Jia, Jie Zhang, and Shiguang Shan. Dual-branch meta-learning network with distribution alignment for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 17:138–151, 2021.

[8] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020.

[9] Fangling Jiang, Qi Li, Pengcheng Liu, Xiang-Dong Zhou, and Zhenan Sun. Adversarial learning domain-invariant conditional features for robust face anti-spoofing. *International Journal of Computer Vision*, 131(7):1680–1703, 2023.

[10] Chenqi Kong, Kexin Zheng, Yibing Liu, Shiqi Wang, Anderson Rocha, and Haoliang Li. M ^{3} fas: An accurate and robust multimodal mobile face anti-spoofing system. *IEEE Transactions on Dependable and Secure Computing*, 2024.

[11] Chenqi Kong, Kexin Zheng, Shiqi Wang, Anderson Rocha, and Haoliang Li. Beyond the pixel world: A novel acoustic-based face anti-spoofing system for smartphones. *IEEE Transactions on Information Forensics and Security*, 17:3238–3253, 2022.

[12] Kaicheng Li, Hongyu Yang, Binghui Chen, Pengyu Li, Biao Wang, and Di Huang. Learning polysemantic spoof trace: A multi-modal disentanglement network for face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 1351–1359, 2023.

[13] Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen. Domain invariant vision transformer learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6098–6107, 2023.

[14] Ajian Liu, Zichang Tan, Zitong Yu, Chenxu Zhao, Jun Wan, Yanyan Liang, Zhen Lei, Du Zhang, Stan Z Li, and Guodong Guo. Fm-vit: Flexible modal vision transformers for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 18:4775–4786, 2023.

[15] Yaojie Liu and Xiaoming Liu. Spoof trace disentanglement for generic face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3813–3830, 2022.

[16] Yuchen Liu, Yabo Chen, Mengran Gou, Chun-Ting Huang, Yaoming Wang, Wenrui Dai, and Hongkai Xiong. Towards unsupervised domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20654–20664, 2023.

[17] Olegs Nikisins, Anjith George, and Sébastien Marcel. Domain adaptation in multi-channel autoencoder based features for robust face anti-spoofing. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

[18] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition, pages 10023–10031, 2019.

[19] Tao Shen, Yuyu Huang, and Zhijun Tong. Facebagnet: Bag-of-local-features model for multi-modal face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[20] Koushik Srivatsan, Muzammal Naseer, and Karthik Nandakumar. Flip: Cross-domain face anti-spoofing with language guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19685–19696, 2023.

[21] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24563–24574, 2023.

[22] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4123–4133, 2022.

[23] Zitong Yu, Yunxiao Qin, Xiaobai Li, Zezheng Wang, Chenxu Zhao, Zhen Lei, and Guoying Zhao. Multi-modal face anti-spoofing based on central difference networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 650–651, 2020.

[24] Haixiao Yue, Keyao Wang, Guosheng Zhang, Haocheng Feng, Junyu Han, Errui Ding, and Jingdong Wang. Cyclically disentangled feature translation for face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3358–3366, 2023.

[25] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20453–20463, 2023.

[26] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma. Generative domain adaptation for face anti-spoofing. In *European Conference on Computer Vision*, pages 335–356. Springer, 2022.