软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(9):4388−4402 [doi: 10.13328/j.cnki.jos.007285] [CSTR: 32375.14.jos.007285]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

视觉注意力和域特征融合的人脸活体检测方法*

朱建秋, 华 阳, 宋晓宁

(江南大学 人工智能与计算机学院, 江苏 无锡 214122)

通信作者: 宋晓宁, E-mail: x.song@jiangnan.edu.cn

摘 要: 人脸活体检测是人脸识别技术安全落地的有力保障. 但活体攻击方式的不断变换, 给现有检测方法带来了极大的挑战. 为应对层出不穷的未知场景和攻击方式, 提出一种基于视觉注意力和域特征融合的双流人脸活体检测模型. 首先, 提出基于视觉注意力的特征提取模块, 增强模型提取基于全局信息的内容特征的能力. 接着, 构建一种新型的风格特征融合模块, 将内容特征和浅层纹理表达的风格特征相融合来优化样本的特征表示. 此外, 设计基于孪生网络的特征映射策略并修正对比损失函数, 分别强化模型的鲁棒性和规避训练过程中梯度易振荡的问题. 还采用对抗训练来降低模型对样本数据域之间分歧的敏感性, 进一步增强其泛化性. 多项实验结果表明, 所提方法在主流数据集上跨域表现均优于现有模型, 验证其泛化性和强鲁棒性.

关键词: 人脸活体检测; 域泛化; 特征融合

中图法分类号: TP391

中文引用格式: 朱建秋, 华阳, 宋晓宁. 视觉注意力和域特征融合的人脸活体检测方法. 软件学报, 2025, 36(9): 4388–4402. http://www.jos.org.cn/1000-9825/7285.htm

英文引用格式: Zhu JQ, Hua Y, Song XN. Face Anti-spoofing Method Based on Visual Attention and Domain Feature Fusion. Ruan Jian Xue Bao/Journal of Software, 2025, 36(9): 4388–4402 (in Chinese). http://www.jos.org.cn/1000-9825/7285.htm

Face Anti-spoofing Method Based on Visual Attention and Domain Feature Fusion

ZHU Jian-Qiu, HUA Yang, SONG Xiao-Ning

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

Abstract: Face anti-spoofing is a powerful guarantee for the practical security of facial recognition technology. However, the constant evolution of live attack methods poses significant challenges to existing detection methods. To address the increasing number of unknown scenarios and attack methods, a two-stream face anti-spoofing model based on visual attention and domain feature fusion is proposed. First, a visual attention-based feature extraction module is proposed to strengthen the model's capacity to extract content features based on global information. Second, a novel style feature fusion module is designed to optimize the feature representation of the sample by fusing content features with low-level textural style features. Third, a feature mapping strategy based on the Siamese network is developed and the contrast loss function is modified to improve the model robustness and avoid easy gradient oscillation during training, respectively. Furthermore, domain adversarial training (DAT) is used to reduce the sensitivity of the model to differences between sample data domains and further improve its generalization. Extensive experimental results verify the generality and strong robustness of the proposed method, demonstrating that it outperforms existing models in cross-domain performance on mainstream datasets.

Key words: face anti-spoofing; domain generalization; feature fusion

人脸识别技术凭借其独特优势被广泛应用于各个领域. 而人脸活体检测作为其核心技术, 专用于精准辨别真实人脸与伪造人脸, 以此防御欺诈攻击, 保障人脸识别系统安全运行. 常见的攻击方式分为图片打印攻击、视频回

* 基金项目: 国家重点研发计划 (2023YFF1105102, 2023YFF1105105); 国家社科基金重大项目 (21&ZD166); 江苏省自然科学基金 (BK20221535); 江苏省研究生研究与实践创新计划 (KYCX23_2438)
收稿时间: 2023-06-20; 修改时间: 2023-12-25, 2024-07-30; 采用时间: 2024-08-29; jos 在线出版时间: 2025-01-08
CNKI 网络首发时间: 2025-01-16

放攻击和 3D 面具攻击这 3 种. 随着活体攻击技术的不断更迭, 场景多元化趋势明显, 亟需提升检测算法的鲁棒性和泛化能力以对抗新型欺诈手段. 当前端到端学习方法虽取得一定成果, 但在面对未知环境和未曾训练过的攻击类型, 模型的反欺诈检测性能仍显不足. 针对此问题, 研究人员依仗先进的特征提取骨干网络和更全面的域泛化技术来提升人脸活体检测算法的泛化性, 以此来缓解上述问题[1].

具体来说, 在使用不同特征提取骨干网络方面, 许多工作都是以 ResNet[2]作为特征提取网络, 基于人脸深度图辅助监督策略的方法更多使用 DepthNet[3]对人脸特征进行提取. 此外, Deb 等人[4]提出使用 FCN 从人脸图像中学习局部判别性特征; 在 Huang 等人[5]的工作中, 采用 ViT[6]作为骨干网络来提升模型的综合性能. 上述网络框架在特征的提取和精炼都有了较大的改进. 在设计域泛化技术方面, 域泛化可以使模型直接从多个源域学习到通用的特征表达, 且不需要使用目标域中的数据, 这对于模型的实际部署更加实用. 基于域泛化的人脸活体检测算法假设在多个源域和未知的目标域之间存在通用的人脸特征表达, 从源域训练得到的模型可以很好地泛化到新的目标域上. Shao 等人[7]通过提出一种多重对抗判别的域泛化模型首次将域泛化技术引入到人脸活体检测领域. 在此基础上, Wang 等人[8]提出了一种特征重组的域泛化人脸活体检测算法, 该研究基于人脸图像内容特征和风格特征在统计特性上的差异, 通过使用一个双流网络提取图像的内容特征和风格特征, 并对不同的内容特征和风格特征进行组合, 在重组后的特征空间上使用对比学习的策略, 以此获取更具泛化性的特征表达.

以上方法基于各样的特征提取骨干网络和域泛化技术在处理人脸活体检测任务上均取得了较好的实验结果, 但仍有不足. 首先, 基于 CNN 的各样骨干网络, 主要利用卷积层对图像进行特征提取. 尽管多层卷积网络可以有效提取图像的局部特征, 将复杂的像素信息映射到易于区分的特征空间中, 但由于缺乏对图像中长距离像素间的关系特征的提取, 导致丢失了图像内部的全局关联信息以及降低了对图像间整体结构异同的判断能力. 其次, Wang 等人[8]在融合样本的过程中, 双边样本特征有着同等重要的作用, 但该方法会侧重于某一边的样本特征从而导致另一部分信息的丢失, 造成融合后的特征表达能力不足, 模型依旧缺乏跨域反欺诈检测的能力. 此外, 上述方法为了增强与活体相关的信息并抑制特定域的风格信息, 使用多种域泛化策略来减缓模型受域间分歧的影响. 不过其使用的对比损失函数由于不够平稳, 在进行多任务联合学习的过程中均易于导致模型震荡, 亟待调整.

为了解决上述问题, 本文设计了一种新型基于视觉注意力和域特征融合的双流人脸活体检测模型. 具体来说, 本文首先构建了基于视觉注意力网络 (visual attention network, VAN)[9]的人脸特征提取骨干网络来提取样本的内容特征. VAN 采用大核注意力模块 (large-kernel attention, LKA), 结合了卷积网络和自注意力模块的优点, 可同时提取图像近距离的局部特征和远距离特征间的依赖, 以更低算力和时间消耗中提取更为全面的人脸特征信息. 接着, 本文设计了一种新型的风格迁移网络来融合两个分支提取的特征. 需注意, 本文将提取的风格特征处理为顺序和乱序两种, 并分别与顺序的内容特征相融合得到两种融合特征, 分别为完整样本特征和乱序辅助特征. 为了增强模型的泛化性, 本文在融合过程中, 对特征堆叠的高响应区域进行加权, 以降低所提特征对数据域之间差异的敏感性. 此外, 本文还通过对内容特征进行域对抗训练 (domain adversarial training, DAT)[10]以减少域间分歧. 最后, 本文将两种融合特征一同送入基于孪生网络的特征映射模块中, 并对以往的对比损失函数也做进一步的修正, 使模型在多任务联合学习的过程中, 规避了训练易振荡的问题. 综上, 本文贡献如下.

(1) 提出基于 VAN 的内容特征提取模块, 结合卷积网络和自注意力模块的优点, 增强了模型对内容特征的提取能力.

(2) 构建新型的风格迁移融合模块, 以有效且无损地融合活体内容特征和风格特征, 提高了融合特征的表示能力.

(3) 设计基于孪生网络的特征映射策略并对以往的对比损失函数也做进一步的修正, 以规避训练易振荡的问题, 同时采用域对抗训练强化内容特征的活体信息并抑制域间分歧.

1 相关工作

早期基于深度学习的人脸活体检测研究将该问题视为一个简单的二分类任务, 通过二值监督训练出一个端到

端的模型去判定输入人脸的真假. 2014 年, Yang 等人[11]首次将 CNN 引入活体检测任务, 先用 CNN 提取原始图像的特征, 再送入 SVM 分类器中做分类, 显著提升了人脸活体检测模型的性能. 为了缓解由训练数据不足导致的模型过拟合问题, Li 等人[12]将在 ImageNet 上预训练的 CNN 模型迁移到人脸活体检测任务中. 而为了提升模型提取样本特征的能力, Feng 等人[13]利用多重信息作为 CNN 的输入, 包括基于 Shearlet 特征表示的图像信息和通过光流表示的脸部/全局动作信息, 将 3 种特征结合起来用 CNN 检测是否为欺诈人脸. 此外, Xu 等人[14]增加了对时序信息的考虑, 在 CNN 中加入长短期记忆单元 (long short term memory, LSTM), 提出了 CNN-LSTM 网络结构, 通过 LSTM 获取多帧之间的时序动态信息, 从而提高模型的性能. 在丰富的攻击数据下, 基于深度学习的检测算法可以提取到更全面的人脸特征. 为了让模型学习到更精细的人脸特征, 研究人员对添加辅助监督信息的检测方法进行了深入探索. 在图片打印和视频回放欺诈方式下的人脸是不包含面部深度信息的, 即深度一致. 而真实人脸是立体的, 因此面部区域的深度有所不同. 基于上述差异, Atoum 等人[15]首次引入人脸深度图的概念, 并提出了双分支 CNN 的方法, 使用人脸深度图作为辅助监督, 通过将面部的表观信息与人脸深度图信息相结合, 从而能够更好地区分真实人脸和欺诈人脸. 此外, Liu 等人[3]引入远程光电容积描计法 (remote photolethysmography, rPPG) 到人脸活体检测, 并提出了一种更为复杂的 CNN-RNN 结构, 使用人脸 rPPG 信号和人脸深度信息相结合的方式, 从时序和空间两个方面对模型进行监督学习. 而 Zhang 等人[16]利用特征解耦的思想, 将特征分解为活体特征和与活体特征无关的内容特征, 只利用活体特征判别真假人脸. 基于以上思想铺垫, Yu 等人[17]先提出双边卷积网络 (bilateral convolutional network, BCN), 利用双边滤波与 CNN 相结合的方式来提取更多人脸的内在材质属性特征. 接着, 他们还提出一种基于中心差分卷积 (central difference convolution, CDC)[18]的人脸活体检测算法, 有效增强模型对不同环境下细粒度特征的表示能力, 取得了优于其他算法的检测性能. 总的来说, 添加辅助监督信息的人脸活体检测算法的性能有着显著提升, 有效缓解了二值监督下模型易发生过拟合的问题.

尽管基于端到端的深度学习算法可应用于大多数场景, 但在未知数据域和未知攻击类型上的泛化性能仍不够理想, 在安全性要求较高的实际应用中并不稳定. 基于此, 越来越多的研究人员关注如何提升深度人脸活体检测算法的泛化能力. 在人脸活体检测中, 考虑到不同数据集中的攻击数据 (或真实人脸数据) 之间存在一定的相关性, 可以利用迁移学习 (transfer learning) 将在已有数据上学习到的知识 (即辨别真伪人脸的方式) 迁移到新的检测任务或数据中, 以达到提升模型泛化能力的目的. 然而, 往往源域 (source domain) 和目标域 (target domain) 的数据之间存在着分布偏移, 影响模型的泛化性能. 基于域泛化 (domain generalization, DG) 的方法的提出可有效应对这一问题. 基于域泛化的算法是假定源域和未知域之间存在泛化统一的特征空间, 即模型经过训练可以从多个源域中学习到不含域特征或者与域无关的通用特征表示, 使模型在未知的数据域上也能进行较好地适配. 基于域泛化的人脸活体检测的算法在未知数据域和未知攻击方式上的泛化能力得到一定的提升, 是近年来的研究热点. Shao 等人[7]首次将域泛化的思想引入人脸活体检测领域, 提出了一种多重对抗判别的域泛化模型来学习多个源域共享的通用特征表达. Wang 等人[19]提出了基于特征解耦的跨域人脸活体检测算法, 将特征解耦为与判别真伪人脸有关的特征以及测试人员的个体特征. Jia 等人[20]利用单边对抗学习的域泛化框架, 即训练一个特征生成器, 使其只对不同源域中的真实样本提取具有泛化性的特征. 随着对比学习[21,22]在自/半监督领域的兴起, Zhang 等人[10]提出了基于对比学习的域泛化方法, 该方法不受限于对源域数量的需求, 可以广泛应用于各类域泛化的场景中. 然而, 基于域泛化的算法学到的通用特征可能会包含于欺诈攻击无关的信息, 例如测试人员的个体特征和传感器噪声等, 从而导致特征的判别性不佳. 因此, 如何提升算法的泛化性仍是人脸活体检测任务中需要深入探索和研究的重要问题.

2 模型构建

为减小数据源的差异对模型性能的影响, 提高模型在未知攻击方式和未知场景下的检测效果, 本文提出基于视觉注意力和域特征融合的人脸活体检测模型. 如图 1 所示, 模型整体架构包含内容特征和风格特征两个分支. 其中, 内容特征主要包含的图像的真伪标签语义信息和全局特性, 本文使用基于 VAN 的内容特征提取块获取相关特征, 并以域对抗训练来降低模型对数据域层面特征信息的敏感度. 对于风格特征, 其主要侧重样本的浅层纹理信

息和以及一定量的数据域风格信息. 本文使用基于 CNN 的风格特征提取模块提取风格特征并以正序和乱序两种方式与内容特征进行融合, 分别构成了完整样本特征以及用于对比训练的乱序辅助特征. 接着, 完整特征表示用于样本的分类任务以及与乱序辅助特征协同以对比学习的方式来强化与活体相关的特征信息, 减缓域间风格分歧导致的泛化性下降. 最后, 将上述任务中的对抗训练损失、分类损失和对比损失作为联合损失函数对模型进行监督学习. 接下来, 本文将详细描述所提方法的 3 个创新, 包括基于 VAN 的复合特征提取模块、风格特征融合与孪生映射模块以及面向跨域检测的新型对比损失函数.

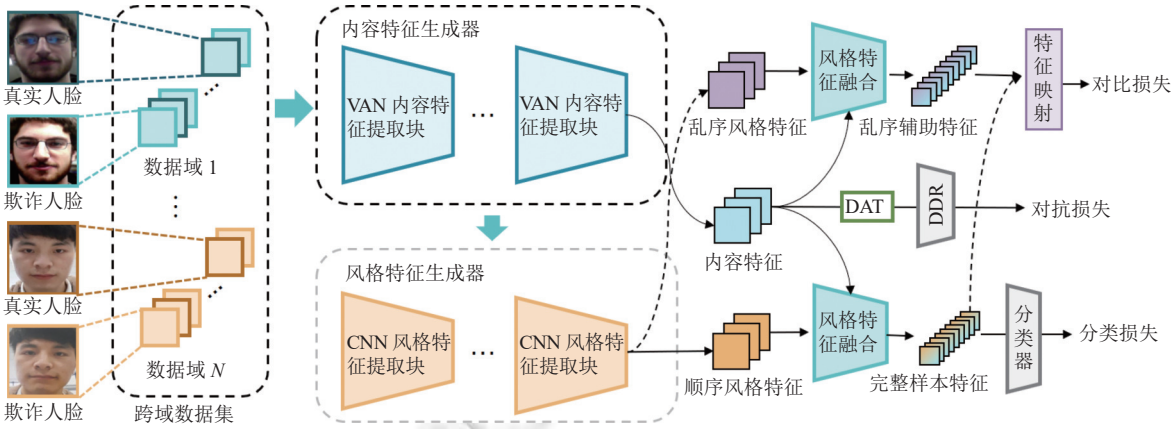


图 1　基于视觉注意力和域特征融合的人脸活体检测方法框架

2.1　基于 VAN 的复合特征提取模块

不同数据域中的样本均包含人脸区域, 且拥有相似的语义特征空间. 但受数据采集设备、场景和光照等差异的影响, 数据域之间的风格差异相差较大. 另外, 无论是真实人脸还是欺诈人脸, 形态和所占整幅图像的空间往往是接近的. 这一系列因素使得在跨域数据中, 相比域间差异, 样本的真伪语义特征差异极小, 进而导致人脸活体检测模型无法进行有效的欺诈检测. 为了增强模型对真伪语义特征的捕获能力, 本文采用生成对抗学习的方式, 设计了如图 2 所示的基于 VAN 的内容特征生成器和基于 CNN 的风格特征生成器, 分别提取样本的两项特征.

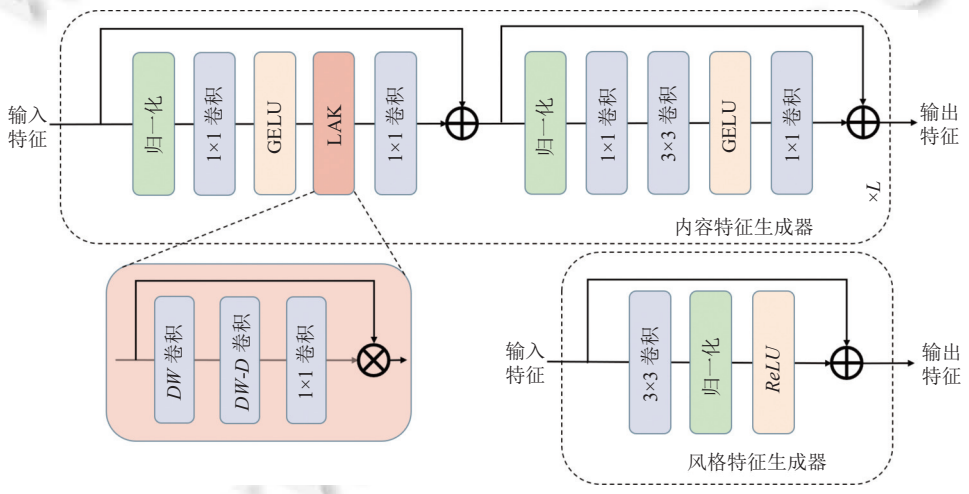


图 2　基于 VAN 的人脸特征提取骨干网络

本文采用基于 VAN 的骨干网络来提取内容特征, VAN 使用大核卷积可以捕捉长距离依赖同时, 兼顾普通卷积提取局部特征的能力. 此外, 为降低大核卷积 (large kernel, LAK) 参数量过大的问题, VAN 将其分解成深度

(depth-wise, *DW*) 卷积, 深度膨胀 (depth-wise dilation, *DW-D*) 卷积以及一个 1×1 的通道卷积, 并按照公式 (1) 和公式 (2) 来提取样本的内容特征.

$$Fc_{i+1} = Attention \otimes Fc_i \tag{1}$$

$$Attention = Conv_{1\times1}(DW\text{-}D(DW(Fc_i))) \tag{2}$$

其中, Fc_i 表示第 i 层内容特征, $Conv_{1\times1}$ 表示 1×1 卷积操作, $\otimes$ 表示逐元素相乘. VAN 因其可同时提取样本的局部特征和长距离依赖而被应用于增强模型捕获内容特征的能力, 而风格特征与内容特征的需求侧重的并不相同, 更局限于纹理、边缘和棱角等浅层特征的提取, 因此本文采用基于卷积神经网络以多尺度聚合的方式构建风格特征生成器, 其计算公式为:

$$Fs_{\text{out}} = \sum_{i=0}^{n} BiUS(Fs_i) \tag{3}$$

$$Fs_{i+1} = ReLU(\text{IN}(\text{CNN}(Fc_i))) + Fc_i \tag{4}$$

其中, Fs_i 和 Fc_i 分别表示第 i 层的风格特征和内容特征, IN 是实例归一化 (instance normalization), *ReLU* 和 CNN 分别是激活函数层和卷积层, *BiUS* 是基于双线性插值的上采样. 基于此, 模型实现了对样本的风格特征提取. 接下来, 本文将介绍内容特征和风格特征的融合方法, 以及用于对比学习的孪生映射模块.

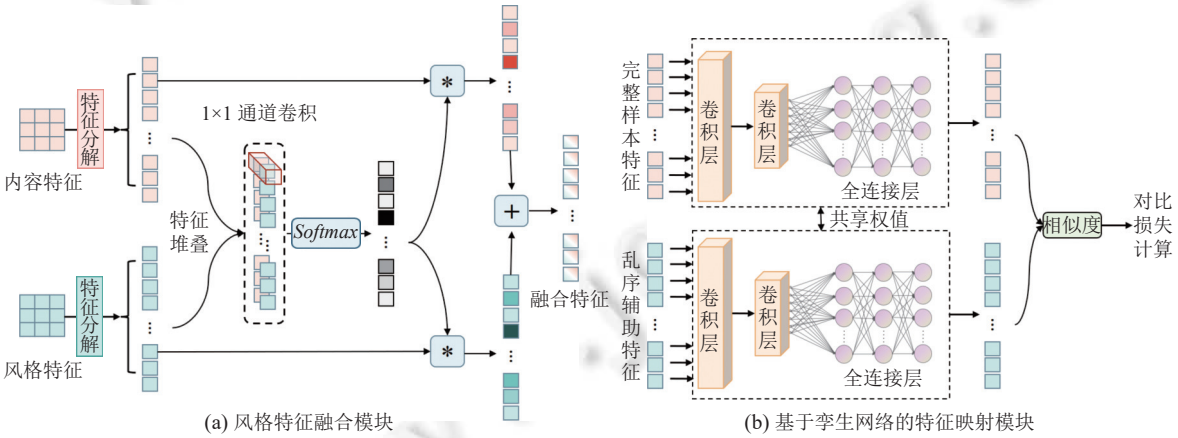
2.2 风格特征融合模块与孪生映射模块

风格特征融合旨在将样本的内容特征和风格信息有机融合, 以便于更好地进行欺诈检测和对比学习. 而两项下游任务均会对特征梯度反穿从而进行监督学习. 因此, 为了更好地将语义信息用于指导模型的学习, 本文采用如图 3(a) 所示的方式将原特征矩阵进行特征分解, 并将分解的序列特征用于特征融合与对比学习. 其中, 本文采用的特征分解是基于滑窗卷积设计的, 公式为: $L_x = flatten(Conv(M_x))$, 其中 L_x 和 M_x 分别是对应的序列特征和矩阵特征. 接着, 本文按照图 3(a) 所示的方式, 将内容特征和风格特征相融合, 对应的融合公式如下:

$$L_f = L_c * W_{\text{att}} + L_s * W_{\text{att}} \tag{5}$$

$$W_{\text{att}} = Softmax(Conv(Concatenate(L_c, L_s))) \tag{6}$$

其中, L_f、L_c 和 L_s 分别是融合特征、内容特征和风格特征, W_{att} 注意力权重矩阵, *Conv* 和 *Concatenate* 分别是卷积计算和级联计算, $*$ 是元素乘.



(a) 风格特征融合模块　　(b) 基于孪生网络的特征映射模块

图 3　风格特征融合与孪生映射

需要注意的是, 本文将内容特征分别与正序和乱序两种方式的风格进行融合, 分别构成了完整样本特征和乱序辅助特征. 具体来讲, 给定一个批处理中长度为 N 的输入序列, x_i 表示输入样本, 其中 $i \in \{1,2,\ldots,N\}$, 其内容特征可以表示为 $Fc(x_i)$, 风格特征如 $Fs(x_i)$. 因此, 对应融合后的完整特征 $Fcs(x_i, x_i)$ 可以表示为:

$$Fcs(x_i,x_i)=Fuse(Fc(x_i),Fs(x_i)) \tag{7}$$

其中, $Fuse()$ 是风格特征融合函数. 而乱序辅助特征 $Fcs(x_i,x_j)$ 则通过原始的 $Fc(x_i)$ 和随机打乱后的 $Fs(x_j)$ 进行融合而成, 公式为:

$$Fcs(x_i,x_j)=Fuse(Fc(x_i),Fs(x_j)) \tag{8}$$

其中, $i\in\{1,2,\ldots,N\}$, $j\in random\ shuffle$. 通过这种方式, 可以得到如图 4(a) 所示的 8 类乱序辅助特征, 根据内容特征和风格特征的真伪标签信息和数据域信息可以分出同域同标签融合特征 (同为真实和同为欺诈), 同域异标签融合特征 (内容为真实、风格为欺诈和相反情况), 异域同标签融合特征 (同为真实和同为欺诈), 异域异标签融合特征 (内容为真实、风格为欺诈和相反情况). 接着, 如图 3(b) 所示, 完整特征与融合特征经过基于孪生网络的特征映射模块 (由权值共享的 1D 卷积和全连接网络组成) 得到对应的映射结果, 并通过余弦相似度来计算两类特征的相似度. 其公式为: $sim_{\cos}=\dfrac{L'_o\cdot L'_d}{\|L'_o\|\|L'_d\|}$, 式中, L'_o 和 L'_d 分别是完整样本特征与乱序辅助特征. 值得注意的是, 对比学习主要目的是促使模型的主干网络可以更侧重于学习样本的语义信息并忽视其域信息. 而本文不希望对比学习的梯度反传过多用于不同的映射模型的参数训练, 但同时也需要映射模块来协助全局模型的稳定收敛, 因此采用孪生网络完成融合特征的映射. 接下来, 本文将详细介绍用于域泛化的对比学习的训练策略以及相应损失函数的修正设计.

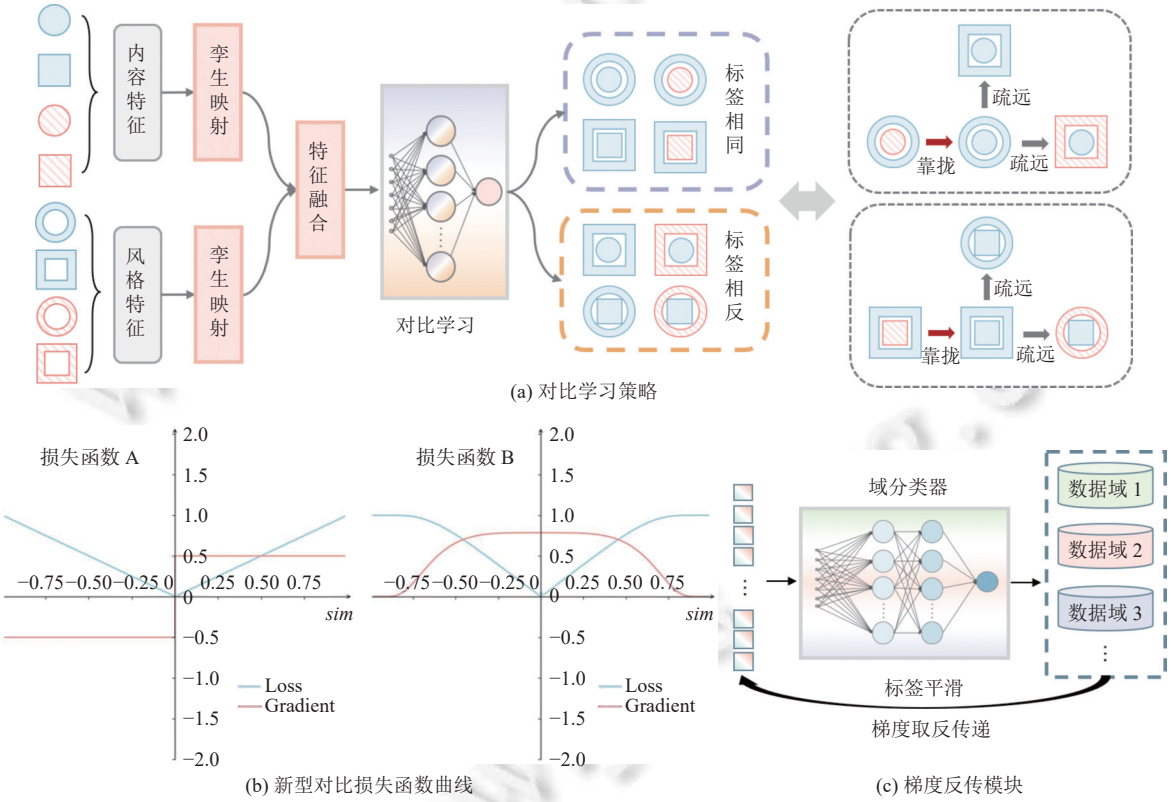


图 4 对比学习及对抗训练

2.3 面向跨域的新型对比损失及对抗训练

域泛化指通过训练策略促使模型在提取样本的高级语义特征时忽视域特征而重视类别特征. 因此, 在处理完整样本特征和乱序辅助特征 (两者的内容特征一致) 时, 本文采用图 4(a) 所示的对比学习策略. 其中, 形状表示的真伪标签, 颜色表示数据域信息. 过程中, 侧重将同标签的融合特征记为同类 ($l_c=1$), 异标签融合特征记为异类

($l_c=-1$), 对数据域差异不做任何处理. 模型先通过余弦相似度来计算完整样本特征和乱序辅助特征的相似性 *sim*, 并基于相似性设计损失函数训练模型, 当乱序辅助特征和完整样本特征属于同类的情况时, 使 *sim* 趋向于 1 (乱序辅助特征向完整样本特征靠拢), 相反则趋向于−1 (乱序辅助特征疏远完整样本特征). 以此来促使模型提取的高级语义特征集中关注于样本的真伪标签信息, 忽视域信息.

标签作为损失函数, 其损失值随相似性 *sim* 的变化如图 4(b) 中的损失函数 A 所示. 这么处理有两个较大的问题影响了模型的训练, 其一, 与常规任务不同的是, 基于相似性的对比学习目的是使得同类的相似性趋于 1, 异类趋于−1, 而损失函数 A 在相似性等于 0 时, 是不可导点, 存有梯度震荡的风险, 影响了模型的整体稳定性. 其二, 损失函数 A 的梯度一直处于一个固定值, 容易导致模型多个迭代下参数反复波动而无法拟合的问题. 为了解决这一问题, 本文提出了一种新型基于对比损失函数 L_{contra}, 其计算公式为:

$$L_{\text{contra}}=\tanh\left(\tan\left(-\frac{sim\cdot l_c\cdot\pi}{2}\right)\right) \tag{9}$$

其中, *sim* 和 l_c 分别时完整样本特征和乱序辅助特征的相似性以及类别归属. 函数图像如图 4(b) 中损失函数 B 所示. 从损失曲线可以看出, 在 $sim\in(-1,1)$ 时函数没有不可导点, 避免了损失函数 A 的第 1 个问题; 而从梯度变化曲线来看, 模型会在 *sim* 处于中间态时收敛的速度要比在邻接处更快, 不仅在训练前期提高了模型的训练速度, 在训练后期也不会因为梯度太大导致参数波动不收敛的问题, 有效地解决了损失函数 A 的第 2 个问题, 具体的实验表现本文将消融实验中给出. 除了对比损失函数, 在训练模型时本文还使用了交叉熵损失函数作为训练模型检测欺诈人脸的二值监督的损失 L_{cls}, 损失函数计算公式为:

$$L_{\text{cls}}=\frac{1}{N}\sum_i-[y_i\cdot\log(p_i)+(1-y_i)\cdot\log(1-p_i)] \tag{10}$$

其中, p_i 是第 i 个样本的预测类别, y_i 是第 i 个样本的标签, N 是批处理的样本量. 此外, 本文采用对抗损失函数对生成的内容特征进行域对抗训练, 计算公式为:

$$\min_D\max_G L_{\text{adv}}(G,D)=-\mathbb{E}_{(x,y)\ (X,Y_D)}\sum_{i=1}^{M}[i=(y\pm\varepsilon)]\log D(G(x)) \tag{11}$$

其中, Y_D 为域标签的集合, M 为不同数据域的数量. G 和 D 分别表示内容特征生成器和域判别器, ε 是标签平滑值. 为了同时优化 G 和 D, 使用如图 4(c) 所示的梯度反向层在反向传播过程中通过将其乘以一个负标量来反转梯度. 本文是通过多任务联合学习的方式来训练模型, 因此最终用于训练的损失函数为:

$$L=L_{\text{cls}}+\lambda_1L_{\text{adv}}+\lambda_2L_{\text{contra}} \tag{12}$$

其中, λ_1 和 λ_2 为损失的权重系数. 根据损失的实际数值和重要性, 权重系数设置均为 1.

3 实验分析

3.1 数据集与评估方法

本文采用 OULU-NPU[23]、CASIA-MFSD[24]和 Replay-Attack[25]和 MSU-MFSD[26]数据集评估所提模型的实验效果. OULU-NPU 数据集包含 5 940 段真实和欺诈视频, 视频长度限制为 5 s, 帧率为 30 Hz, 分辨率为 1920×1080 像素. 实验共包含 4 个测试协议, 协议 1 测试模型在不同数据采集环境下的性能, 协议 2 评测不同的攻击媒介对模型性能的影响, 协议 3 测试模型在不同数据采集设备下的性能, 协议 4 考虑了上述所有因素, 评测模型的综合泛化能力. CASIA-MFSD 数据集由 600 段视频组成, 其中包含 50 位测试者, 每位测试者在 3 种摄像头下获取分辨率为 480×640 像素、640×480 像素、720×1080 像素的 12 段视频, 其中真实人脸 3 段, 3 种攻击方式各 3 段, 攻击方式包括卷曲图像攻击、剪切图像攻击和视频回放攻击. Replay-Attack 数据集包含 1 300 段人脸视频样本, 帧率为 25 Hz, 分辨率大小为 320×240 像素, 由 50 位测试者在光照均匀和光照不均匀两种场景下录制, 攻击方式分为图像攻击和视频回放攻击, 设备支持条件分为手持和固定两种方式. MSU-MFSD 数据集包含 55 位测试者在不同条件下的 440 段人脸视频序列, 目前只公开了 35 名测试者的数据, 其中 15 名作为训练数据, 其余 20 名作为测试数据. 视频

长度为 10 s, 帧率为 30 Hz, 分辨率大小为 720×480. 攻击方式包括照片打印攻击和视频回放攻击, 在照片打印攻击中, 使用分辨率为 5184×3456 的高清晰度照片作为攻击媒介, 视频回放攻击分为高清晰度和低清晰度重放视频攻击两种方式.

另外, 本文引用活体检测领域常用评估指标来评估所提模型的性能, 主要包括半错误率 (half total error rate, *HTER*) 和受试者工作特征曲线下面积 (area under curve, *AUC*). 错误拒绝率 (false rejection rate, *FRR*) 和错误接收率 (false acceptance rate, *FAR*) 是相关评估指标, 分别表示将正样本预测为负样本的概率和将负样本预测为正样本的概率, *FRR* 和 *FAR* 通常成反比关系, 为了综合评价算法的好坏, 引入了 *HTER* 和 *EER* 作为衡量指标. *HTER* 为 *FAR* 和 *FRR* 的均值, 计算公式为:

$$HTER = \frac{FAR + FRR}{2} \tag{13}$$

受试者工作特征曲线 (receiver operating characteristic, *ROC*) 是一种用于分析分类模型性能的工具, 将假阳率 (false positive rate, *FPR*) 作为 X 轴, 将真阳率 (true positive rate, *TPR*) 作为 Y 轴, 其中 *FPR* 表示被错误预测为正样本的负样本数目占所有负样本数量的比率, *TPR* 表示被正确分类为正样本的正样本数目占所有正样本数量的比率. 评估指标 *AUC* 则为 *ROC* 曲线下与坐标轴围成的面积, 取值范围 $[0, 1]$, *AUC* 越接近 1, 检测方法真实性越高.

3.2 实验设置

为了验证所提方法的有效性, 本文实验在 CASIA-FASD、Replay-Attack、OULU-NPU 和 MSU-MFSD 数据集上进行了测试, 采用 *HTER* 和 *AUC* 作为评估标准. 对于视频数据, 本文以特定的时间间隔进行取帧, 获取图像格式的数据后, 采用 MTCNN 进行人脸检测, 然后将人脸进行裁剪和调整为 256×256 作为输入. 本文实验在 NVIDIA RTX 3090 显卡上进行, 网络结构利用 PyTorch 框架进行实现, 损失的优化器是 Adam. 实验具体的超参数设置如表 1 所示.

表 1 参数设置

超参数	值
Learning rate	0.000 1
Learning rate decay	0.5
Batch size	64
Step size	50
Number of VAN blocks	4
Number of CNN blocks	3
Weight decay	0.000 05

在本文的实验中, 将一个数据集视为一个数据域, 共采用 3 种跨数据测试的评估策略. 第 1 种, 该策略使用了 4 个数据集: OULU-NPU (记为 O)、CASLA-FASD (记为 C)、Replay-Attack (记为 I) 和 MSU-MFSD (记为 M), 具体来说, 实验中随机选择 1 个数据库作为测试集, 其余 3 个作为训练集, 得到 4 个协议: O&C&I to M、O&M&I to C、O&C&M to I 和 I&C&M to O; 第 2 种, 为了验证本文方法在有限数据域上仍具竞争力, 本文选择 MSU-MFSD 和 Replay-Attack 作为训练的源域, 其余两个数据集 (即 CASIA-FASD 和 OULU-NPU) 分别作为测试的目标域, 具体实验协议为 M&I to C 和 M&I to O; 第 3 种, 在 CASIA-FASD 与 Replay-Attack 两个数据集之间跨数据库测试, 具体实验协议为 C to I 和 I to C.

3.3 消融实验

为了验证本文所提模块及优化后的损失函数的有效性, 本节在第 1 种评估策略上进行消融实验, 为了分析本文提出方法中每个组成部分的效果, 本节在基线方法 SSAN-M[8]的基础上逐步替换所提模块以获得 5 个方法, 实验使用 *HTER* 作为评估指标, 实验结果如表 2 所示. 表中方法 0 代表基线方法, 该方法使用 DepthNet 作为特征提取主干网络, 特征融合模块采用 AdaIN 算法[27], 损失函数采用第 2.3 节所介绍的损失函数 A. 基线方法在第 1 种评估策略的 4 种协议上分别取得了不错的结果.

方法 1 和方法 2 分别测试了所提出的基于 VAN 的复合特征提取模块和结合特征注意力的风格融合模块的有效性. 从表 2 中可以看到, 除了方法 2 在 O&C&M to I 协议上得到的结果外, 两个模型都明显改善了基准方法在基准数据集上的性能. 这验证了本文的论点, 高质量的特征提取骨干网络和良好的特征融合模块是成功解决人脸活体检测任务的关键. 基于 VAN 的复合特征提取模块能够有效提取图像局部特征的同时也能有效地捕捉远距离依赖性和适应性, 从而加强对人脸图像的语义信息和物理属性特征的提取. 使用结合特征注意力的风格融合模块替换 AdaIN 融合方法, 有效地避免了单侧样本特征缺失的问题. 然而, 方法 2 在 O&C&M to I 上的表现不如基线方法, 分析原因为在样本分布差异大的数据集中, 使用 DepthNet 提取的人脸特征信息不够丰富. 为了验证这一猜想, 本文在方法 3 中同时使用了 VAN 特征提取模块和风格融合模块. 可以看到, 这个模型的性能得到了进一步的提高, 优于方法 1 和方法 2 中的模型. 同时也验证了内容特征与风格特征的充分融合可以进一步提升性能.

表 2　OULU-NPU、CASIA-FASD、Replay-Attack 和 MSU-MFSD 上的消融实验结果 (%)

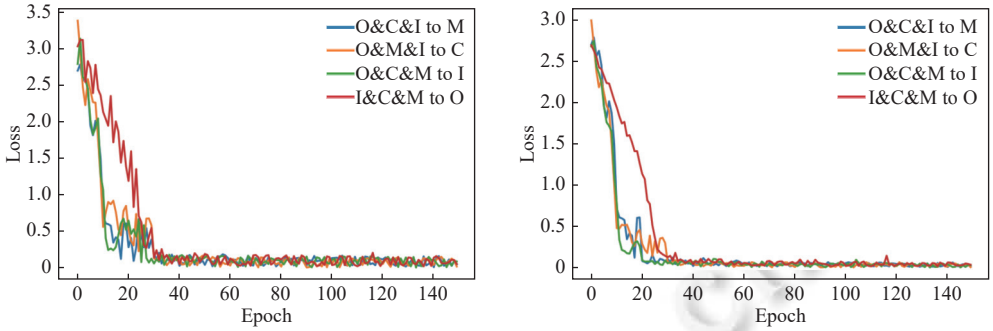
方法	组合			评估结果 (*HTER*)			
	VAN特征提取模块	风格融合模块	新型对比策略	O&C&I to M	O&M&I to C	O&C&M to I	I&C&M to O
0	—	—	—	10.42	16.47	14.00	19.51
1	√	—	—	9.58	15.74	12.75	18.92
2	—	√	—	9.29	13.52	14.87	18.68
3	√	√	—	9.17	12.41	11.00	14.41
4	—	—	√	10.83	15.56	11.75	18.85
5	—	√	√	9.10	13.39	11.27	15.29
6	√	—	√	9.87	12.41	10.45	12.10
7	√	√	√	**6.67**	**10.56**	**8.21**	**10.58**

为验证本文所提基于孪生映射的对比学习的有效性, 本节实验中首先简单地使用孪生映射模块和损失函数 B 替换基线方法中的原始的损失函数 A, 从而得到了方法 4. 然而, 这只小幅度提升了 O&M&I to C、O&C&M to I 和 I&C&M to O 上的性能. 该现象的主要原因是模型需要充分地融合丰富的内容特征的风格特征以更好地判别真伪人脸. 方法 5 和方法 6 则将新型对比策略分别于 VAN 特征提取模块和风格融合模块相组合, 可以看出融合使用的效果更好, 由此可以看出模块之间的具有很好的适配性. 当将 3 个改进组合作为本文所提方法时, 与方法 4 相比, 方法 7 在 4 种协议上的 *HTER* 分别降低了 4.16%、5.00%、3.54% 和 8.27%, 模型性能提升明显. 值得注意的是, 在 O&C&I to M 协议上, 方法 4 与方法 6 相对于基线和方法 1 的表现有所下降. 这种下降主要归因于新型对比学习策略虽然优化了模型对样本信息的学习方式, 但基线方法 AdaIN 融合模块无法平衡内容特征与风格特性, 导致对比学习难以有效抑制风格特征的影响. 该问题在风格单一的 MSU-MFSD 数据集上的测试表现尤为显著, 因此设计一种损失函数平滑的学习机制反而会导致模型性能受到抑制. 针对于此, 本文所提出的风格融合模块可通过无损的融合方式有效解决这一问题, 使得模型在该协议上的整体性能显著提升, 方法 5 和方法 7 的结果也证实了其可靠性. 为更清晰展示不同损失函数的收敛效果, 本文绘制了方法 3 和方法 7 对应的训练损失收敛曲线. 如后文图 5 所示, 可以看到方法 7 的收敛曲线的梯度波动更小, 从而也直观地验证了模型使用损失函数 B 后整体的稳定性更强.

3.4　对比实验

为了整体评估本文模型的性能, 本节在 3 种跨域评估策略上与其他主流方法进行对比实验. 首先, 本节采用第 1 种评估策略, 选择 3 个数据集进行训练, 其余的一个用于测试, 实验结果如表 3 和表 4 所示. 从中可以看到, 除了在 O&C&I to M 协议的 AUC 指标和 O&M&I to C 协议的 *HTER* 指标上, 本文方法稍逊于其他方法, 在策略 1 中的其余子协议上本文方法都取得了最优. 以往的人脸活体检测方法如 LBP-TOP[28]、Color Texture[29]和 Auxiliary[3]等, 都专注于从多个源域学习特征, 而这些特征往往只适合于源域中的数据. 相比之下, 本文所提出的基于视觉注意力和域特征融合的方法充分地利用了多源域特征之间的关系, 并学习它们之间的通用特征表达和真伪高级语义信息. 本文方法在训练集和测试集中的数据分布有很大的差异时, 仍能保持稳定的检测性能. 相较于基于域泛化技术的方法中表现良好的 SSDG-R[20]和 SSAN-M[8]等, 本文方法仍具有一定的优势. 本文设计的特征提取与融合网络

充分地利用了人脸图像的内容和风格特征, 并且通过合理的监督策略缩小了源域和目标域的分布差异, 从而显著提升了模型在目标域上的检测能力.



(a) 方法 3 训练损失收敛曲线 (b) 方法 7 训练损失收敛曲线

图 5 不同损失函数下的特征分布

表 3 OULU-NPU、CASIA-FASD、Replay-Attack 和 MSU-MFSD 上的跨域测试结果 1 (%)

方法	O&C&I to M		O&M&I to C	
	HTER	*AUC*	*HTER*	*AUC*
LBP-TOP[28]	36.90	70.80	42.60	61.05
Color Texture[29]	28.09	78.47	30.58	76.89
Auxiliary (depth only)[3]	22.72	85.88	33.52	73.15
MADDG[7]	17.69	88.06	24.50	84.51
SSDG-M[20]	16.67	90.47	23.11	85.45
SDA[30]	15.40	91.80	24.50	84.40
DAM[31]	12.70	95.66	20.98	85.58
DRDG[32]	12.43	95.81	19.05	88.79
ANRL[33]	10.83	96.75	17.83	89.26
AADF-AS[34]	10.83	96.25	13.59	92.75
FM-ViT[35]	9.58	95.45	11.65	94.35
SSAN-M[8]	10.42	94.76	16.47	90.81
SSDG-R[20]	7.38	97.17	**10.44**	95.94
本文方法	**6.67**	**97.32**	10.56	**96.75**

表 4 OULU-NPU、CASIA-FASD、Replay-Attack 和 MSU-MFSD 上的跨域测试结果 2 (%)

方法	O&C&M to I		I&C&M to O	
	HTER	*AUC*	*HTER*	*AUC*
LBP-TOP[28]	49.45	49.54	53.15	44.09
Color Texture[29]	40.40	62.78	63.59	32.71
Auxiliary (depth only)[3]	29.14	71.69	30.17	77.61
MADDG[7]	22.19	84.99	27.98	80.02
SSDG-M[20]	18.21	94.61	25.17	81.83
SDA[30]	15.60	90.10	23.10	84.30
DAM[31]	15.43	91.22	15.27	90.87
DRDG[32]	15.56	91.79	15.63	91.75
ANRL[33]	16.03	91.04	15.67	91.90
AADF-AS[34]	14.25	93.46	20.87	92.58
FM-ViT[35]	12.83	95.35	14.75	93.35
SSAN-M[8]	14.00	94.58	19.51	88.17
SSDG-R[10]	11.71	96.59	15.61	91.54
本文方法	**8.21**	**97.89**	**10.58**	**96.25**

为了评估本文方法在源域数据集 (即只有两个源数据集) 有限时的泛化能力, 本节采用第 2 种评估策略. 由于 MSU-MFSD 和 Replay-Attack 数据集有着显著的域间差异, 故选择这两个数据集作为源域, 将 OULU-NPU 和 CASIA-FASD 数据集分别作为目标域来进行测试. 实验结果如表 5 所示, 本文所提的方法取得了明显优于其他方法的 *HTER* 和 *AUC*, 很好地验证本文方法即使在具有挑战性的情况下也更加有效. 此外, 将表 5 与表 3、表 4 的实验结果进行对比, 不难发现当源域的数量增加时, 本文方法有着比大多数方法更高的同比提升. 这意味着当有更多的源域时, 本文方法更能够利用域间共享信息和区分真假人脸的判别性信息来学习更广义的表征, 从而提升模型对真伪人脸的鉴别能力. 因此, 本文方法可以更好地利用域泛化的优势.

此外, 本文还采用了第 3 种评估策略, 在 CASIA-FASD 与 Replay-Attack 数据上进行跨域实验, 实验结果如表 6 所示. 表中展示了跨数据库测试的结果, 在使用 CASIA-FASD 数据集作为训练集, Replay-Attack 数据集作为测试集的情况下 (C to I), 本文的方法取得了 12.3% 的 *HTER*, 在使用 Replay-Attack 作为训练集, CASIA-FASD 作为测试集的情况下 (I to C), 本文的方法取得了 22.5% 的 *HTER*, 相比于其他最优方法, 本文所提方法在 *HTER* 上分别降

低了 2.7% 和 3.1%. 同时, 与其他主流方法相比, 本文方法基于域泛化技术的同时引入了新型的特征提取骨干网络与特征融合方法, 实验结果有着显著的提升, 再次有力地验证了本文所提方法的有效性.

表 5 有限源域上的测试结果 (%)

方法	M&I to C		M&I to O	
	HTER	*AUC*	*HTER*	*AUC*
MS-LBP[36]	51.16	52.09	43.63	58.07
IDA[26]	45.16	58.80	54.52	42.17
LBP-TOP[28]	45.27	54.88	47.26	50.21
MADDG[7]	41.02	64.33	39.35	65.10
SSDG-M[20]	31.89	71.29	36.01	66.88
DR-MD-Net[19]	31.67	75.23	34.02	72.65
ANRL[33]	31.06	72.12	30.73	74.10
SSAN-M[8]	30.00	76.20	29.44	76.62
本文方法	**22.45**	**86.03**	**21.84**	**83.29**

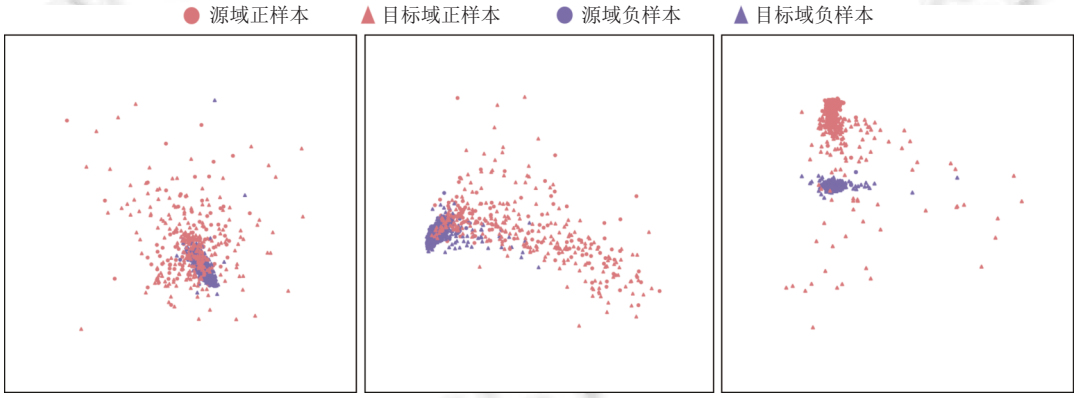
表 6 CASIA-FASD 与 Replay-Attack 上的跨域实验结果 (%)

方法	评估结果 (*HTER*)	
	C to I	I to C
STASN[37]	31.5	30.9
Color Texture[29]	30.3	37.7
Attention[38]	30.0	33.4
De-Spoof[39]	28.5	41.1
Auxiliary[3]	27.6	28.4
ASMN[40]	27.4	28.1
Identity-DS[41]	27.1	31.4
DRL[16]	22.4	30.3
GFA-CNN[42]	21.4	34.3
SSAN-M[8]	15.0	25.6
本文方法	**12.3**	**22.5**

综上所述, 从 4 个主流数据集上的各个对比实验可以看出, 本文方法通过优化特征提取骨干网络与特征融合模块, 成功得到更为丰富的特征, 通过新型对比策略使得模型训练更加鲁棒. 模型在不同策略下的跨域测试中, 都表现得比同类主流方法更好.

3.5 结果讨论与可视化

本文方法使用二值监督损失、对抗损失和对比损失函数对模型进行监督, 为了分析本文所采用的多任务联合监督方式对模型的作用, 本节对使用不同损失函数下的特征分布进行了可视化操作, 如图 6 所示. 图 6(a)、(b) 和 (c) 分别表示使用 L_{cls}、$L_{cls}+L_{adv}$、$L_{cls}+L_{adv}+L_{contra}$ 在 O&C&I to M 协议上的特征分布. 其中, 粉色代表正样本 (即真实人脸), 紫色代表负样本 (即欺诈人脸), 圆形和三角形分别为源域和目标域的样本.



(a) 仅使用 L_{cls} (b) 使用 $L_{cls}+L_{adv}$ (c) 使用 $L_{cls}+L_{adv}+L_{contra}$

图 6 不同损失函数下的特征分布

从图 6 中可以看出, 在仅使用分类损失的情况下, 模型一定程度上可对真伪人脸进行区分, 但对于目标域中的样本模型的区分能力不够理想, 表现在图中仍存在目标域中的正样本与负样本交叠的情况, 分析原因为仅使用二值监督的方式往往只对于已知的数据域具有较好的区分效果, 对于未知的数据域, 域间数据的差异会给模型带来极大的干扰, 从而影响模型的最终性能. 对于图 6(b), 其在前者的基础上加入对模型进行约束, 较好地提升了模型

在区分真伪人脸上的表现, 增加了对抗损失的监督使得模型可以提取更为丰富的人脸特征信息, 有效地缓解了数据域之间的风格差异相差较大带来的干扰, 但真实人脸样本和欺诈人脸样本之间存在着一个较为粗糙的边界, 模型的跨域泛化性能仍有提升空间. 图 6(c) 为本文所采用的最终损失函数下的特征分布, 图中的正样本和负样本各自聚合, 且域间的差异明显缩小. 可视化结果表明, 即使遇到一个未知分布的目标域数据, 本文方法仍然可以很好地适用于目标域.

此外, 为了进一步验证基于风格融合的域泛化的有效性, 本文对上述 4 个基准数据集样本提取的内容特征和融合特征进行重组并做了热力图可视化. 本文将内容特征与重组后的融合特征基于像素区域真实置信度的高低通过 OpenCV 色度图 (COLORMAP_JET) 绘制出对应的热力图, 接着在于原图相融合得到如图 7 所示的结果. 图中热力度 (偏红) 越高指的该区域的真实度越高, 相反则欺诈概率更高. 从图中不难看出, 内容特征整体相对弥散, 尽管真实欺诈样本之间具备一定的区分度, 但仍存在区分难度. 而风格特征则偏向于细节区域, 真伪区分度相对较差, 更侧重于图像的纹理和边缘细节的提取. 相比而言, 融合特征无论是在真实样本还欺诈样本中, 热力基本成大片状分布, 真实人脸与欺诈样本区分度极高, 从侧面论证了风格迁移的域泛化可显著提升模型的检测性能.

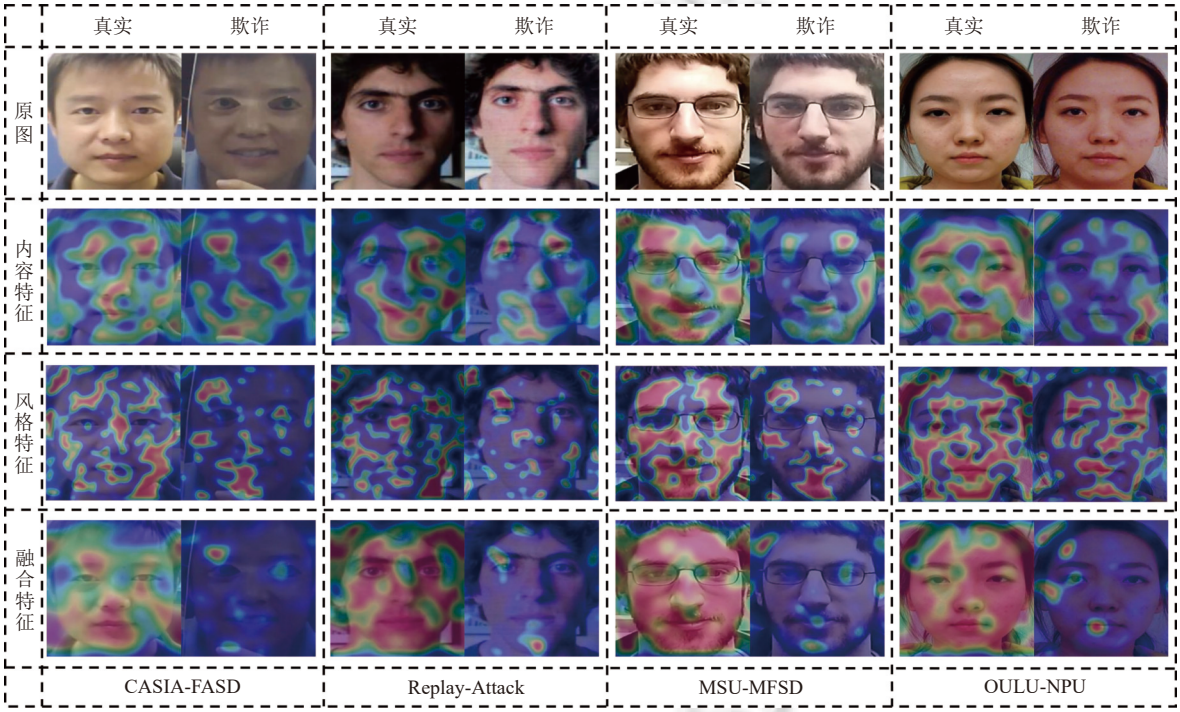


图 7 基准数据集的人脸特征热力图

4 总 结

本文利用域泛化技术, 在不使用目标域数据的情况下学习通用的特征表达, 提出一种新的基于视觉注意力和域特征融合的双流人脸活体检测模型. 首先, 本文构建了基于 VAN 的人脸特征提取骨干网络, 用于获取更为丰富的人脸信息; 其次, 本文构建了一种新型的风格特征融合模块, 将内容特征和浅层纹理表达的风格特征相融合来优化样本的特征表示. 此外, 本文对以往的对比损失函数也做进一步的修正, 结合余弦相似度和双曲正切函数来平稳对比损失梯度, 以规避训练过程中梯度易振荡的问题. 还采用对抗训练来降低模型对样本数据域之间分歧的敏感性. 本文在 4 个数据集上进行 3 种策略的实验, 验证了所提模型的实际效能. 从实验结果来看, 相比于当前主流方法, 本文所提方法在源域和目标域存在显著分布差异和有限源域两种情况下的性能都极具竞争力, 展示了所提模

型的高泛化性能. 不过, 本文所提方法仍有改进的空间, 例如以更小的参数量来提升模型的鲁棒性等. 未来, 将基于此进一步优化模型, 不断为人脸活体检测工作提供新的理论方案.

References:

[1] Zhang F, Zhao SK, Yuan C, Chen W, Liu XL, Zhao HJ. Research progress of face recognition anti-spoofing. Ruan Jian Xue Bao/Journal of Software, 2022, 33(7): 2411–2446 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6590.htm [doi: 10.13328/j.cnki.jos.006590]

[2] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770−778. [doi: 10.1109/CVPR.2016.90]

[3] Liu YJ, Jourabloo A, Liu XM. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 389−398. [doi: 10.1109/CVPR.2018.00048]

[4] Deb D, Jain AK. Look locally infer globally: A generalizable face anti-spoofing approach. IEEE Trans. on Information Forensics and Security, 2020, 16: 1143–1157. [doi: 10.1109/TIFS.2020.3029879]

[5] Huang YH, Hsieh JW, Chang MC, Ke LP, Lyu S, Santra AS. Multi-teacher single-student visual Transformer with multi-level attention for face spoofing detection. In: Proc. of the 32nd British Machine Vision Conf. BMVC, 2021. 22–25.

[6] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai XH, Unterthiner T, Dehghani M, Minderer M, Heigol G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. In: Proc. of the 9th Int'l Conf. on Learning Representations. OpenReview.net, 2021. 1–21.

[7] Shao R, Lan XY, Li JW, Yuen PC. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 10015−10023. [doi: 10.1109/CVPR.2019.01026]

[8] Wang Z, Wang ZZ, Yu ZT, Deng WH, Li JH, Gao TT, Wang ZY. Domain generalization via shuffled style assembly for face anti-spoofing. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 4113–4123. [doi: 10.1109/CVPR52688.2022.00409]

[9] Guo MH, Lu CZ, Liu ZN, Cheng MM, Hu SM. Visual attention network. Computational Visual Media, 2023, 9(4): 733–752. [doi: 10.1007/s41095-023-0364-2]

[10] Zhang YF, Wang X, Liang J, Zhang Z, Wang L, Jin R, Tan TN. Free lunch for domain adversarial training: Environment label smoothing. arXiv:2302.00194, 2023.

[11] Yang JW, Lei Z, Li SZ. Learn convolutional neural network for face anti-spoofing. arXiv:1408.5601, 2014.

[12] Li L, Feng XY, Boulkenafet Z, Xia ZQ, Li MM, Hadid A. An original face anti-spoofing approach using partial convolutional neural network. In: Proc. of the 6th Int'l Conf. on Image Processing Theory, Tools and Applications. Oulu: IEEE, 2016. 1−6. [doi: 10.1109/IPTA.2016.7821013]

[13] Feng LT, Po LM, Li YM, Xu XY, Yuan F, Cheung TCH, Cheung KW. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. Journal of Visual Communication and Image Representation, 2016, 38: 451–460. [doi: 10.1016/j.jvcir.2016.03.019]

[14] Xu ZQ, Li S, Deng WH. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In: Proc. of the 3rd IAPR Asian Conf. on Pattern Recognition. Kuala Lumpur: IEEE, 2015. 141−145. [doi: 10.1109/ACPR.2015.7486482]

[15] Atoum Y, Liu YJ, Jourabloo A, Liu XM. Face anti-spoofing using patch and depth-based CNNs. In: Proc. of the 2017 IEEE Int'l Joint Conf. on Biometrics (IJCB). Denver: IEEE, 2017. 319−328. [doi: 10.1109/BTAS.2017.8272713]

[16] Zhang KY, Yao TP, Zhang J, Tai Y, Ding SH, Li JL, Huang FY, Song HC, Ma LZ. Face anti-spoofing via disentangled representation learning. In: Proc. of the 16th European Conf. on Computer Vision (ECCV 2020). Glasgow: Springer, 2020. 641−657. [doi: 10.1007/978-3-030-58529-7_38]

[17] Yu ZT, Li XB, Niu XS, Shi JG, Zhao GY. Face anti-spoofing with human material perception. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 557−575. [doi: 10.1007/978-3-030-58571-6_33]

[18] Yu ZT, Zhao CX, Wang ZZ, Qin YX, Su Z, Li XB, Zhou F, Zhao GY. Searching central difference convolutional networks for face anti-spoofing. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5294−5304. [doi: 10.1109/CVPR42600.2020.00534]

[19] Wang GQ, Han H, Shan SG, Chen XL. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 6677−6686. [doi: 10.

1109/CVPR42600.2020.00671]

[20] Jia YP, Zhang J, Shan SG, Chen XL. Single-side domain generalization for face anti-spoofing. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 8481−8490. [doi: 10.1109/CVPR42600.2020.00851]

[21] Tian YL, Sun C, Poole B, Krishnan D, Schmid C, Isola P. What makes for good views for contrastive learning? In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 6827–6839.

[22] Cui JQ, Zhong ZS, Tian ZT, Liu S, Yu B, Jia JY. Generalized parametric contrastive learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2024, 46(12): 7463–7474. [doi: 10.1109/TPAMI.2023.3278694]

[23] Boulkenafet Z, Komulainen J, Li L, Feng XY, Hadid A. OULU-NPU: A mobile face presentation attack database with real-world variations. In: Proc. of the 12th IEEE Int'l Conf. on Automatic Face & Gesture Recognition. Washington: IEEE, 2017. 612−618. [doi: 10.1109/FG.2017.77]

[24] Zhang ZW, Yan JJ, Liu SF, Lei Z, Yi D, Li SZ. A face antispoofing database with diverse attacks. In: Proc. of the 5th IAPR Int'l Conf. on Biometrics (ICB). New Delhi: IEEE, 2012. 26−31. [doi: 10.1109/ICB.2012.6199754]

[25] Chingovska I, Anjos A, Marcel S. On the effectiveness of local binary patterns in face anti-spoofing. In: Proc. of the 2012 Int'l Conf. of Biometrics Special Interest Group. Darmstadt: IEEE, 2012. 1−7.

[26] Di W, Hu H, Jain AK. Face spoof detection with image distortion analysis. IEEE Trans. on Information Forensics and Security, 2015, 10(4): 746–761. [doi: 10.1109/TIFS.2015.2400395]

[27] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). Venice: IEEE, 2017. 1510–1519. [doi: 10.1109/ICCV.2017.167]

[28] De Freitas Pereira T, Komulainen J, Anjos A, De Martino JM, Hadid A, Pietikäinen M, Marcel S. Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing, 2014, 2014(1): 2. [doi: 10.1186/1687-5281-2014-2]

[29] Boulkenafet Z, Komulainen J, Hadid A. Face spoofing detection using colour texture analysis. IEEE Trans. on Information Forensics and Security, 2016, 11(8): 1818–1830. [doi: 10.1109/TIFS.2016.2555286]

[30] Wang JJ, Zhang JY, Bian Y, Cai YY, Wang CM, Pu SL. Self-domain adaptation for face anti-spoofing. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 2746−2754. [doi: 10.1609/aaai.v35i4.16379]

[31] Chen ZH, Yao TP, Sheng KK, Ding SH, Tai Y, Li JL, Huang FY, Jin XY. Generalizable representation learning for mixture domain face anti-spoofing. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 1132–1139. [doi: 10.1609/aaai.v35i2.16199]

[32] Liu SB, Zhang KY, Yao TP, Sheng KK, Ding SH, Tai Y, Li JL, Xie Y, Ma LZ. Dual reweighting domain generalization for face presentation attack detection. In: Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence. Montreal: Morgan Kaufmann Press, 2021. 867–873. [doi: 10.24963/ijcai.2021/120]

[33] Liu SB, Zhang KY, Yao TP, Bi MW, Ding SH, Li JL, Huang FY, Ma LZ. Adaptive normalized representation learning for generalizable face anti-spoofing. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 1469–1477. [doi: 10.1145/3474085.3475279]

[34] Liu AJ, Tan ZC, Liang YY, Wan J. Attack-agnostic deep face anti-spoofing. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Vancouver: IEEE, 2023. 6336–6345. [doi: 10.1109/CVPRW59228.2023.00674]

[35] Liu AJ, Tan ZC, Yu ZT, Zhao CX, Wan J, Liang YY, Lei Z, Zhang D, Li SZ, Guo GD. FM-ViT: Flexible modal vision Transformers for face anti-spoofing. IEEE Trans. on Information Forensics and Security, 2023, 18: 4775–5786. [doi: 10.1109/TIFS.2023.3296330]

[36] Määttä J, Hadid A, Pietikäinen M. Face spoofing detection from single images using micro-texture analysis. In: Proc. of the 2011 Int'l Joint Conf. on Biometrics (IJCB). Washington: IEEE, 2011. 1−7. [doi: 10.1109/IJCB.2011.6117510]

[37] Yang X, Luo WH, Bao LC, Gao Y, Gong DH, Zheng SB, Li ZF, Liu W. Face anti-spoofing: Model matters, so does data. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3502−3511. [doi: 10.1109/CVPR.2019.00362]

[38] Chen HN, Hu GS, Lei Z, Chen YW, Robertson NM, Li SZ. Attention-based two-stream convolutional networks for face spoofing detection. IEEE Trans. on Information Forensics and Security, 2020, 15: 578–593. [doi: 10.1109/TIFS.2019.2922241]

[39] Jourabloo A, Liu YJ, Liu XM. Face de-spoofing: Anti-spoofing via noise modeling. In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 297−315. [doi: 10.1007/978-3-030-01261-8_18]

[40] Zheng W, Yue MY, Zhao SH, Liu SQ. Attention-based spatial-temporal multi-scale network for face anti-spoofing. IEEE Trans. on Biometrics, Behavior, and Identity Science, 2021, 3(3): 296–307. [doi: 10.1109/TBIOM.2021.3066983]

[41] Xu YW, Wu LF, Jian M, Zheng WS, Ma YK, Wang ZM. Identity-constrained noise modeling with metric learning for face anti-spoofing. Neurocomputing, 2021, 434: 149–164. [doi: 10.1016/j.neucom.2020.12.095]

[42] Tu XG, Zhao J, Xie M, Du GD, Zhang HS, Li JS, Ma Z, Feng JS. Learning generalizable and identity-discriminative representations for face anti-spoofing. ACM Trans. on Intelligent Systems and Technology (TIST), 2020, 11(5): 60. [doi: 10.1145/3402446]

附中文参考文献:

[1] 张帆, 赵世坤, 袁操, 陈伟, 刘小丽, 赵涵捷. 人脸识别反欺诈研究进展. 软件学报, 2022, 33(7): 2411–2446. http://www.jos.org.cn/1000-9825/6590.htm [doi: 10.13328/j.cnki.jos.006590]

朱建秋(1998－), 女, 硕士生, 主要研究领域为计算机视觉, 人脸活体检测.

宋晓宁(1975－), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为模式识别, 计算机视觉, 生物信息学, 机器学习.

华阳(1997－), 男, 博士生, 主要研究领域为生物信息学, 计算机视觉, 深度学习.