计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330251
60(8): 1727−1739, 2023

二次解耦与活体特征渐进式对齐的域自适应人脸反欺诈

封 筠[1] 史屹琛[1] 高宇豪[1] 贺晶晶[1] 余梓彤[2]
[1](石家庄铁道大学信息科学与技术学院 石家庄 050043)
[2](大湾区大学 广东东莞 523000)
(fengjun@stdu.edu.cn)

Domain Adaptation for Face Anti-Spoofing Based on Dual Disentanglement and Liveness Feature Progressive Alignment

Feng Jun[1], Shi Yichen[1], Gao Yuhao[1], He Jingjing[1], and Yu Zitong[2]
[1](*School of Information Science and Technology*, *Shijiazhuang Tiedao University*, *Shijiazhuang* 050043)
[2](*Great Bay University*, *Dongguan*, *Guangdong* 523000)

Abstract Although existing face anti-spoofing methods perform well in intra-domain testing, their performance significantly degrades in cross-domain scenarios. Current cross-domain face anti-spoofing methods based on domain adversarial alignment cannot guarantee that the alignment task directly serves the classification task since the alignment and classification networks are independent of each other. We propose a domain adaptation for face anti-spoofing method based on domain invariant liveness features dual disentanglement and progressive adversarial alignment. Firstly, the source domain features are heuristically disentangled into domain specific features and domain invariant features. Then, the gradient of classifier is used to perform a second disentanglement of the live-related and live-unrelated features in the domain invariant features. To alleviate optimization difficulties during training, a curriculum learning method is adopted to progressively align target domain features and the combination of live-related and live-unrelated features, gradually increasing the proportion of live-related features, and enhancing the correlation between the target domain features and face anti-spoofing task. From a causal perspective, we provide an explanation for the liveness alignment domain adaptation. Experimental results on CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, and OULU-NPU datasets demonstrate that the proposed method achieves the best average *HTER* value of 22.5% compared with ten existing methods and the current state-of-the-art performance on four evaluation protocols. Especially the *HTER* values of I-M and O-M evaluation protocols achieve 12.4% and 12.8%, respectively. The proposed method can significantly reduce the error rates of the model in the target domain and has better cross-domain generalization ability.

Key words face anti-spoofing; domain adaptation; dual disentanglement; domain adversarial progressive alignment; curriculum learning

摘 要 现有的人脸反欺诈(face anti-spoofing, FAS)方法虽然在域内测试表现良好,但在跨域场景下性能会大幅度下降.当前基于域对抗对齐的跨域人脸反欺诈方法,因其对齐网络和分类网络彼此独立,无法保证对齐任务直接服务于分类任务.提出了一种基于二次解耦与活体特征课程学习渐进式对抗对齐的域自适应人脸反欺诈(domain adaptation for face anti-spoofing based on dual disentanglement and liveness feature

收稿日期: 2023-03-31; **修回日期**: 2023-06-16
基金项目: 国家自然科学基金项目(61772070, 61972267); 河北省高等学校科学技术研究重点项目(ZD2021333)
This work was supported by the National Natural Science Foundation of China (61772070, 61972267) and the Key Projects of Science and Technology Research in Colleges and Universities of Hebei Province (ZD2021333).

curriculum learning progressive adversarial alignment, DDCL）方法，首先将源域特征启发式解耦为域相关特征和域无关特征，之后使用分类器的梯度信息将域无关特征中的活体相关和无关特征进行第 2 次解耦. 在训练过程中为减轻优化难度，通过课程学习的方式对目标域特征与活体相关、无关特征的组合进行渐进式对抗对齐，逐步提高活体相关特征的比重，增强目标域特征与活体检测任务的相关性，从因果角度给出活体对齐域自适应的解释. 在 CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD 与 OULU-NPU 公开数据集上的实验结果表明，与现有 10 种方法相比，所提出的方法获得了 22.5% 的最佳平均 *HTER* 值，并在 4 个测评协议上均达到了当前先进水平，尤其是 I-M 和 O-M 测评协议的 *HTER* 值分别达到了 12.4% 和 12.8%，能显著降低模型在目标域上的错误率，具有更好的跨域泛化能力.

关键词 人脸反欺诈；域自适应；二次解耦；域对抗渐进式对齐；课程学习

中图法分类号 TP391

近年来，人脸识别系统被广泛应用于门禁、安防及支付等需要身份验证的场合，其高效、易用的特点备受赞誉. 然而，人脸数据可通过社交媒体、视频网站等途径轻松获取，非法用户常使用恶意的伪造人脸对识别系统进行欺骗攻击. 基于活体检测的人脸反欺诈（face anti-spoofing, FAS）技术作为前置保护措施，可确保人脸识别系统的安全性和可靠性，近年来吸引了国内外研究者的广泛关注.

随着深度学习技术在计算机视觉领域的快速发展，深度神经网络模型被广泛应用于人脸反欺诈任务，其训练需要大量数据，当测试数据与训练数据不服从同一分布时，模型的性能会大幅度下降. 受限于数据采集的高额成本，收集各领域数据并完成标签并不现实. 因此，需要在数据受限的情况下提升模型的泛化能力，即提高在跨域场景下的性能. 为了解决该问题，无监督领域自适应技术被应用于人脸反欺诈任务，使用有标签的源域数据与无标签的目标域数据共同训练得到一个在目标域上性能良好的模型. 其主要思想是将源域数据与目标域数据的分布进行对齐，使源域的标签知识可以被引入无监督的目标域中. 研究者从不同的层面采用相应的对齐策略进行域自适应人脸反欺诈方法研究，目前主流的对齐策略受生成对抗网络启发，使用领域对抗训练的方式对齐源域和目标域特征.

领域对抗神经网络训练（domain-adversarial training of neural networks, DANN）[1] 方法在对齐源域和目标域特征时，将其作为一个整体进行对齐，如图 1（a）所示. 然而，源域提取的特征中有大量与活体检测任务无关的信息，如人脸的轮廓、五官信息等. 由于特征对齐与下游的分类任务并行、独立，所以将目标域的特征与这些无关信息对齐，不仅无法直接服务于活体检测任务，还可能使模型训练向次优方向推进. 本文提出一种基于二次解耦与活体特征课程学习渐进式对抗对齐的域自适应人脸反欺诈（domain adaptation for face anti-spoofing based on dual disentanglement and liveness feature curriculum learning progressive adversarial alignment, DDCL）方法，如图 1（b）所示. 在训练时加强源域和目标域信息的交互，使得对齐任务直接服务于分类任务. 通过领域对抗训练，渐进式地将目标域特征向源域的活体相关特征对齐，在减轻优化难度的同时保证目标域提取到与活体任务更为相关的分类特征.

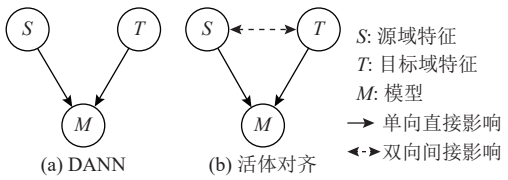


Fig. 1 Comparison diagram of different alignment strategies for domain adaptation

图 1 域自适应不同对齐策略对比示意图

本文的主要贡献包括 4 个方面：

1）提出一种基于启发式建模与分类器梯度的二次解耦方式，首先将源域特征解耦为域相关特征和域无关特征，之后将域无关特征解耦为活体相关特征和活体无关特征，用于后续领域特征对齐；

2）提出一种基于课程学习的领域对抗渐进式特征对齐训练策略，对源域解耦出的活体相关、无关特征进行线性加权组合，将目标域特征与其对齐，即在模型初始训练阶段将目标域特征与源域的活体无关特征进行对齐，之后逐步提高活体相关特征所占比重，最终将目标域与源域活体相关特征进行对齐；

3）从因果推断的角度出发，将本文所提 DDCL 方法与主流的对抗对齐域自适应方法进行比较，不同于之前方法的源域和目标域的对齐和分类彼此独立，DDCL 方法训练时源域和目标域信息交互更为密切，特征对齐可直接服务于活体检测任务；

4)在4个公开数据集上的大量实验结果表明本文所提方法的优越性，可以显著提高无监督域自适应人脸反欺诈性能，与当前先进结果相比具有较强竞争力.

1 相关工作

1.1 人脸反欺诈

人脸反欺诈任务的目标是判断当前待检测人脸是来自于真实人脸还是各种材质的假体攻击. 早期研究者根据专家的先验知识，设计了一系列的手工特征，如纹理特征[2-4]、图像质量[5-6]、生理信号[7-8]、脸部运动[9-11]等. 纹理特征分析方法被广泛应用于人脸反欺诈技术，如LBP[2,12]，SIFT[13]，SURF[14]，HOG[15]等. 虽然手工特征方法对于真假人脸的判别非常重要，但是因其受限于研究者掌握的先验知识，同时需要高分辨率图像数据，导致手工特征尽管在训练数据集上表现很好，但由于图像采集条件和攻击媒介的多样性，使得手工提取特征的方法难以具有高的鲁棒性.

在计算机视觉领域，数据驱动的深度学习方法表现大幅度优于手工提取特征方法，将深度神经网络，如CNN, Transformer等引入人脸反欺诈任务，识别性能通常会有较大提升，是当前研究的重点. Yang等人[16]使用CNN作为特征提取器，分类真实人脸和欺诈样本. 研究发现，纯神经网络往往难以满足判别要求，此后出现一系列辅助信息如深度图[17-19]、反射图[15]、光流信号[20-22]等与深度学习方法相结合，模型设计和优化侧重各有不同. Yu等人[23]巧妙地将手工LBP特征与CNN结合，较普通CNN而言能捕获到更多连续的伪造线索，如晶格伪影；还使用神经架构搜索(neural architecture search, NAS)技术自动探索网络架构最优参数，提高判别效率和精度，相比于现有方法其准确率高，但跨库测试错误率较高，模型泛化能力欠佳.

1.2 无监督域自适应人脸反欺诈

为提升活体检测模型的泛化能力，充分利用全部数据，减小源域和目标域数据因光照、环境等因素产生的领域分布差异，研究者将域自适应技术引入人脸反欺诈. 现有的无监督域自适应人脸反欺诈方法，主要包括数据分布对齐和领域对抗对齐2类方法.

在数据分布对齐方法中，Li等人[24]通过最小化源域和目标域特征空间之间的最大均值差异(maximum mean discrepancy, MMD)[25]，学习到一个泛化性更强的分类器. Tu等人[26]通过减小源域和目标域之间基于核方法的MMD距离来提高模型的泛化性. 然而仅仅通过减小领域之间的MMD距离可能无法充分探索源域之间的有用信息，因此目前使用对抗迁移学习的方式成为研究热点[27].

在领域对抗对齐方法中，Kim等人[28]提出一种风格指导的领域自适应框架，通过风格选择归一化构造推理自适应模型，实现利用特定领域的风格信息指导，自动将模型适配到目标数据. Hamblin等人[29]提出一种新的领域自适应框架，利用多模态数据改善基于可见光的呈现攻击检测(presentation attack detection, PAD)任务. Wang等人[30]采用对抗训练方式由特征提取器获得源域和目标域的共同特征，同时使用三元组损失在特征空间上尽可能分散真实人脸和假体攻击，最后使用K近邻分类. El-Din等人[31]认为只使用对抗训练方式进行领域自适应，会在目标域与源域攻击方式和设备类型不同的情况下无法得到好的结果，所以为保存一些目标域特有的属性，采用深度聚类生成伪标签进行辅助训练.

1.3 课程学习

由易到难的学习策略在人类教育中很常见，研究者将其引入深度学习领域. 课程学习作为一种模仿人类学习方式的深度学习训练范式，其主要思想是模型先从简单数据开始学习，然后逐步增加学习数据的难度，直至学习整个数据集. Yang等人[32]利用课程学习将目标域样本与动态选择的源域样本对齐，以利用源域样本的不同的可迁移性. Shu等人[33]提出从较多的域内数据(类似于目标域)训练到较少的域内数据，指导模型在充分利用源域数据的同时适应目标域. Gong等人[34]将每种特征与教师联系，设计一种多模态课程学习策略以整合来自不同特征模态的信息. Wang等人[35]提出一个统一的动态课程学习框架，自适应地调整每个批次的抽样策略和权重，以提高泛化和辨别能力.

2 本文方法

鉴于当前基于对抗训练的域自适应人脸反欺诈方法，通常无法保证对齐任务直接服务于活体分类任务，模型往往会向着次优的方向训练，本文首先通过双解耦获得域无关活体相关特征，即将由启发式解耦所得到的域无关特征，进一步解耦为活体相关特征和活体无关特征. 由于活体无关特征对齐简单，但对真假人脸分类任务而言，其作用弱于活体相关特征，在充分解耦的理想情况下，活体无关特征对分

类任务没有帮助,所以接着采用基于课程学习的渐进式特征对齐域对抗训练策略,即在训练前期将目标域与源域活体无关部分进行对齐,随着训练的迭代,逐步将目标域特征与源域的活体相关特征对齐,从而提升模型在目标域上的泛化能力,本文所提DDCL方法的整体流程如图2所示.

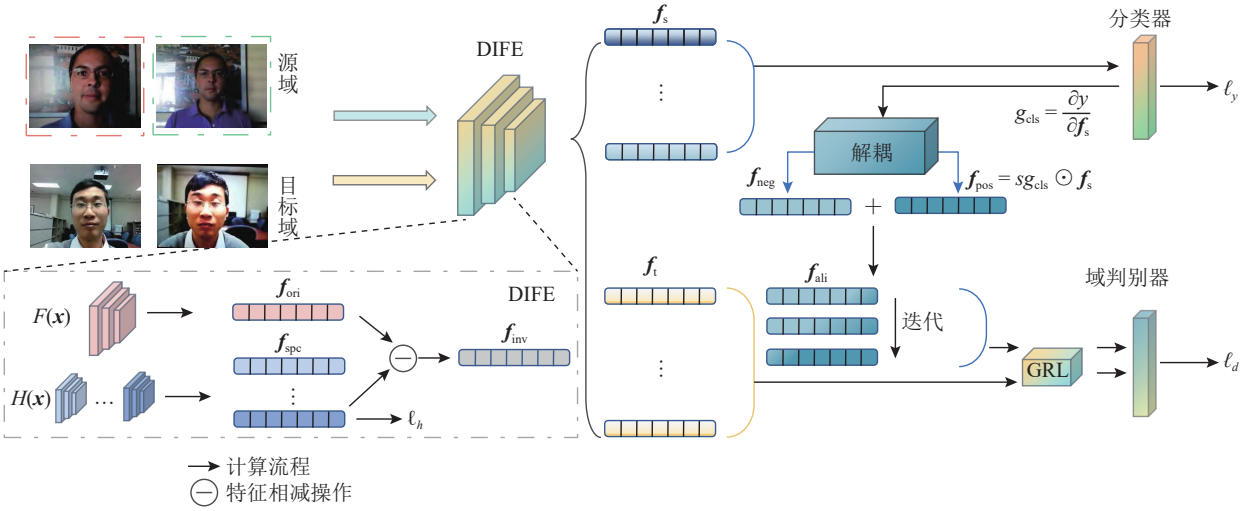


Fig. 2 Pipeline of DDCL method proposed in this paper

图2 本文DDCL方法整体流程

输入模型的数据$\{\boldsymbol{x}_i\}_{i=1}^{N}(\boldsymbol{x}_i\in[0,255]^{3\times H\times W})$包含多个领域的真实人脸和假体攻击,其中$N$是训练集大小,$H\times W$是图像尺寸.整体模型主要由域无关特征提取器(domain invariant feature extractor, DIFE)、域判别器(discriminator)和分类器(classifier)3部分构成.DIFE通过启发式建模提取源域和目标域共有的域无关特征,域判别器用于判断输入的特征来自源域还是目标域,分类器根据提取好的特征进行真实人脸和假体攻击的分类.利用源域数据训练分类器之后,计算标签y对于源域特征的梯度g_{cls},使用梯度对源域特征进行解耦得到活体无关特征$\boldsymbol{f}_{\text{neg}}$和活体相关特征$\boldsymbol{f}_{\text{pos}}$.随着训练轮次的迭代,调整$\boldsymbol{f}_{\text{neg}}$和$\boldsymbol{f}_{\text{pos}}$的加权参数组合为$\boldsymbol{f}_{\text{ali}}$,通过对抗训练将目标域特征$\boldsymbol{f}_{\text{t}}$与$\boldsymbol{f}_{\text{ali}}$对齐.

2.1 形式化定义

现有在源域上训练的人脸反欺诈模型通常不能很好地推广到目标域数据.为了解决该问题,本文的研究重点是如何提升人脸活体在跨域场景下的泛化能力,首先对无监督域自适应人脸反欺诈任务进行形式化定义:

给定有标签的源域$D_{\text{s}}=\{(\boldsymbol{x}_i^{\text{s}},y_i^{\text{s}})\}_{i=1}^{n_{\text{s}}},\boldsymbol{x}_i^{\text{s}}\in\mathbb{R}^d$和无标签的目标域$D_{\text{t}}=\{(\boldsymbol{x}_i^{\text{t}})\}_{i=1}^{n_{\text{t}}},\boldsymbol{x}_i^{\text{t}}\in\mathbb{R}^d$,其中$\boldsymbol{x}$为人脸数据,$y$为对应的标签(真实人脸或假体攻击),源域和目标域的数据分布因存在协变量偏移而不同,即$p_{\text{s}}(\boldsymbol{x}_{\text{s}})\neq p_{\text{t}}(\boldsymbol{x}_{\text{t}})$.目的是使用有标签的源域数据和无标签的目标域数据学习同一个网络,得到活体检测模型$\gamma:\boldsymbol{X}\to\boldsymbol{Y}$,对于无标签的目标域数据$\boldsymbol{x}_{\text{t}}$,可以准确地预测其标签$y$,即

$$y=\gamma(\boldsymbol{x}_{\text{t}}),\boldsymbol{x}_{\text{t}}\in D_{\text{t}}\ . \tag{1}$$

进一步,γ可以被分解为特征提取器ω和分类器ϕ两部分,即$\gamma=\phi\circ\omega$,其中ω负责提取目标域和源域共有且与任务相关的特征$\omega:\boldsymbol{X}\to\boldsymbol{Z}$,$\phi$对提取到的特征进行分类$\phi:\boldsymbol{Z}\to\boldsymbol{Y}$.因此式(1)可改写为

$$y=\phi(\omega(\boldsymbol{x}_{\text{t}})),\boldsymbol{x}_{\text{t}}\in D_{\text{t}}\ . \tag{2}$$

同时引入域判别器$D:\boldsymbol{Z}\to\{0,1\}$,用于减小领域之间的分布差异.

2.2 域无关活体特征二次解耦

为达到域无关活体特征充分解耦目的,本文提出一种基于启发式建模与分类梯度的二次解耦方法.首先利用启发式建模将源域特征解耦为域相关和域无关部分,之后通过分类器梯度,将域无关特征解耦为活体相关部分和活体无关部分.

1)基于启发式建模的域无关特征解耦

在域自适应中,由于源域和目标域数据之间存在领域差异,直接得到域无关特征用于下游任务并不现实,为了减轻源域和目标域特征的对齐难度,通过启发式建模解耦特征.

假设1:假设特征$\boldsymbol{f}_{\text{ori}}$由域相关特征$\boldsymbol{f}_{\text{spc}}$和域无关特征$\boldsymbol{f}_{\text{inv}}$组成,且对于$\boldsymbol{f}_{\text{spc}}$建模的难度要小于对$\boldsymbol{f}_{\text{inv}}$特征建模.

$$\boldsymbol{f}_{\text{ori}}=\boldsymbol{f}_{\text{spc}}+\boldsymbol{f}_{\text{inv}}. \tag{3}$$

该假设被认为是领域自适应的先验假设[36]，$\boldsymbol{f}_{\text{spc}}$的建模难度介于$\boldsymbol{f}_{\text{inv}}$和$\boldsymbol{f}_{\text{ori}}$之间. 为了减轻$\boldsymbol{f}_{\text{inv}}$的建模难度，借鉴启发式搜索的思想，对$\boldsymbol{f}_{\text{spc}}$建模以逼近理想的$\boldsymbol{f}_{\text{inv}}$，本文构建多重子网络提取特征，如图 2 左下所示. 具体来说，使用一个基础神经网络$F(\boldsymbol{x})$提取全局特征，多重子网络$H(\boldsymbol{x})$提取对应的域相关特征$\boldsymbol{f}_{\text{spc}}$，对域无关特征$\boldsymbol{f}_{\text{inv}}$进行辅助表示，理想的域无关特征$\boldsymbol{f}_{\text{inv}}$可以表示为

$$\boldsymbol{f}_{\text{inv}} = \boldsymbol{f}_{\text{ori}} - \sum \boldsymbol{f}_{\text{spc}}, \tag{4}$$

其中，$\sum$用于形式化表达多重子网络. $\boldsymbol{f}_{\text{spc}}$旨在对域相关特征进行建模，且该建模过程有利于得到理想的$\boldsymbol{f}_{\text{inv}}$，即$\boldsymbol{f}_{\text{spc}}$表征的是当前全局特征$\boldsymbol{f}_{\text{ori}}$与理想域无关特征$\boldsymbol{f}_{\text{inv}}$之间的差异. 根据假设 1，提取$\boldsymbol{f}_{\text{spc}}$难度小于$\boldsymbol{f}_{\text{inv}}$，因此可以利用$\boldsymbol{f}_{\text{spc}}$引导$\boldsymbol{f}_{\text{inv}}$的提取. 为了满足以上要求，根据文献 [36−37] 中对于$H(\boldsymbol{x})$和$F(\boldsymbol{x})$的相似性、独立性和终止条件分析，$H(\boldsymbol{x})$和$F(\boldsymbol{x})$参数的余弦相似度应为−1，为了实现方便，将$H(\boldsymbol{x})$和$F(\boldsymbol{x})$的初始化参数设置为互为相反数.

在域自适应中，若$\boldsymbol{f}_{\text{inv}}$训练到理想的收敛状态，$H(\boldsymbol{x})$提取到的特征$\boldsymbol{f}_{\text{spc}}$应逐步收敛到接近于 0，以使得$\boldsymbol{f}_{\text{inv}}$可以有效代表域无关特征. 将$\boldsymbol{f}_{\text{spc}}$的 L1 范数作为正则项，以逐渐减少$\boldsymbol{f}_{\text{inv}}$中的域相关部分，其损失为

$$\ell_h = \sum_{k=1}^{M} \left| \boldsymbol{f}_{\text{spc}}^k \right| , \tag{5}$$

其中，M为域相关特征数量.

2）基于分类器梯度的活体特征解耦

若仅将特征解耦为域相关和域无关并不是最优的，这是由于真实人脸和假体人脸的数据都包含完整、清晰的人脸结构部分，如人脸的五官、轮廓及肤色等，故而域无关的特征中包含大量与活体任务不相关的特征. 将源域和目标域的活体无关特征进行对齐尽管简单，但无法保证模型向最优方向进行优化. 本文提出基于分类器梯度的第 2 次解耦方式，将源域的域无关特征解耦为活体相关与活体无关 2 部分，训练的理想状态是将目标域特征与源域的活体相关特征对齐.

假设 2：假设特征$\boldsymbol{f}_{\text{inv}}$由活体相关特征$\boldsymbol{f}_{\text{pos}}$和活体无关特征$\boldsymbol{f}_{\text{neg}}$组成，$\boldsymbol{f}_{\text{pos}}$较$\boldsymbol{f}_{\text{neg}}$难对齐.

$$\boldsymbol{f}_{\text{inv}} = \boldsymbol{f}_{\text{pos}} + \boldsymbol{f}_{\text{neg}}. \tag{6}$$

Grad-CAM[38-39] 通过图像分类层的最后一层输出权重衡量上一层生成的每个通道的重要性，再对各通道的所有像素点的值加权，得到对于分类结果最重要的像素点. 通过该方式可以识别出对于当前分类任务来说，图像的哪些部分是与任务最相关的. 使用特征提取器ω得到源域特征$\boldsymbol{f}_{\text{s}}$，分类器$\phi$对其进行分类，可以得到对应类别的预测结果$y$对于$\boldsymbol{f}_{\text{s}}$的梯度$g_{\text{cls}}$：

$$g_{\text{cls}} = \frac{\partial y}{\partial \boldsymbol{f}_{\text{s}}} . \tag{7}$$

将g_{cls}和$\boldsymbol{f}_{\text{s}}$做 hadamard 积，得到活体相关的特征信息$\boldsymbol{f}_{\text{pos}}$：

$$\boldsymbol{f}_{\text{pos}} = s g_{\text{cls}} \odot \boldsymbol{f}_{\text{s}} , \tag{8}$$

其中，s为一个非负的自适应缩放系数，目的是保证$\boldsymbol{f}_{\text{pos}}$与$\boldsymbol{f}_{\text{s}}$两者能量大小保持一致，确保$\boldsymbol{f}_{\text{pos}}$在对齐时占据主导地位. 其计算方式为

$$s = \sqrt{\frac{\|\boldsymbol{f}_{\text{s}}\|_2}{\|g_{\text{cls}} \odot \boldsymbol{f}_{\text{s}}\|_2}}. \tag{9}$$

同时根据假设 2 得到活体无关特征$\boldsymbol{f}_{\text{neg}}$：

$$\boldsymbol{f}_{\text{neg}} = \boldsymbol{f}_{\text{s}} - \boldsymbol{f}_{\text{pos}} , \tag{10}$$

之后将目标域特征渐进式地与活体相关特征$\boldsymbol{f}_{\text{pos}}$和活体无关特征$\boldsymbol{f}_{\text{neg}}$的加权组合进行对齐.

2.3 基于课程学习的渐进式特征对齐

受人类认知原理的启发，Bengio 等人[40] 提出了课程学习的概念，即模仿人类课程中有意义的学习顺序，在模型训练时由容易到复杂、逐步进阶地学习样本和知识. 课程学习的核心在于利用人类专家的先验知识设计一个排序函数，据此对每个数据任务给出其学习的优先度.

在领域特征对齐时，利用梯度计算将源域的域无关特征解耦为活体相关$\boldsymbol{f}_{\text{pos}}$和活体无关$\boldsymbol{f}_{\text{neg}}$. 其中活体无关部分在源域和目标域中广泛存在，故对齐容易，但活体相关部分对齐则困难很多. 受课程学习的启发，本文提出一种渐进式特征对齐训练策略，在训练的不同阶段分别对齐活体无关部分和活体相关部分，以使得模型顺利优化，如图 3 所示. 具体来说，通过将目标域特征$\boldsymbol{f}_{\text{t}}$与源域$\boldsymbol{f}_{\text{pos}}$和$\boldsymbol{f}_{\text{neg}}$的线性组合$\boldsymbol{f}_{\text{ali}}$进行对齐实现的，随着训练迭代次数的递进，逐步增加$\boldsymbol{f}_{\text{pos}}$的权重，$\boldsymbol{f}_{\text{ali}}$计算为

$$\boldsymbol{f}_{\text{ali}} = (1-\alpha) \times \boldsymbol{f}_{\text{neg}} + \alpha \times \boldsymbol{f}_{\text{pos}}, \alpha \in [0, 1], \tag{11}$$

其中，权重α随着训练的迭代单调递增. 为了避免复杂的超参数选择，本文给出一种简单的计算方式，即$\alpha = t/T$，t和T分别是当前训练的轮次以及总共的训练轮次.

在训练开始时$\alpha = 0$，此时$\boldsymbol{f}_{\text{t}}$只与$\boldsymbol{f}_{\text{neg}}$对齐，对齐难度低但对齐的特征并不具备理想的真人、假体鉴别能力；逐步增加$\boldsymbol{f}_{\text{pos}}$在$\boldsymbol{f}_{\text{ali}}$所占比重即加大$\alpha$，提升$\boldsymbol{f}_{\text{t}}$与活体任务的相关性；在训练的最终阶段$\alpha = 1$，$\boldsymbol{f}_{\text{t}}$将只与$\boldsymbol{f}_{\text{pos}}$进行对齐，这时所提取到的特征泛化性强且与

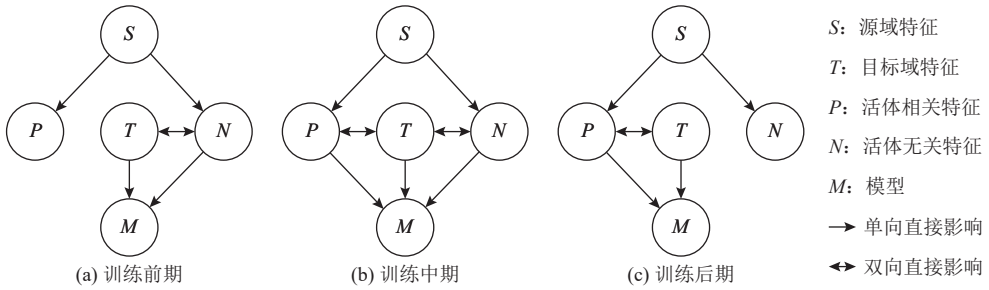


Fig. 3 Progressive feature alignment diagram

图 3 渐进式特征对齐示意图

活体检测相关, 利于后续目标域分类.

2.4 基于对抗训练的域自适应活体检测

借鉴 DANN[1] 方法, 假定领域按数据集划分, 即来自同一数据集的数据属于相同领域, 不同数据集为不同领域, 领域包含的类别为真实人脸和假体攻击. 训练阶段的目标是:

1)训练特征提取器和分类器, 实现源域数据的准确分类;

2)通过对抗方式训练特征提取器, 欺骗域判别器, 以学习领域不变的特征表示.

具体来说, 首先通过最小化分类损失和特征提取器的损失, 优化特征提取器参数$\boldsymbol{\theta}_\omega$和分类器参数$\boldsymbol{\theta}_\phi$:

$$\hat{\boldsymbol{\theta}}_\omega, \hat{\boldsymbol{\theta}}_\phi = \arg\min_{\boldsymbol{\theta}_\omega, \boldsymbol{\theta}_\phi} \ell_y\left(\boldsymbol{\theta}_\omega, \boldsymbol{\theta}_\phi, \boldsymbol{\theta}_d\right), \tag{12}$$

这里, ℓ_y为交叉熵损失, 如式(13)所示.

$$\ell_y = -\sum_{i=1}^{n_s} p\left(\boldsymbol{x}_i^s\right) \log\left(q\left(\boldsymbol{x}_i^s\right)\right), \tag{13}$$

其中, p为真实概率分布, q为预测的概率分布.之后将特征提取器参数$\boldsymbol{\theta}_\omega$和分类器参数$\boldsymbol{\theta}_\phi$固定, 最大化域判别器 d 的损失, 优化参数$\boldsymbol{\theta}_d$:

$$\hat{\boldsymbol{\theta}}_d = \arg\max_{\boldsymbol{\theta}_d} \ell_d\left(\boldsymbol{\theta}_\omega, \boldsymbol{\theta}_\phi, \boldsymbol{\theta}_d\right). \tag{14}$$

损失函数ℓ_d为

$$\begin{aligned}\ell_d(\boldsymbol{x}_s, \boldsymbol{x}_t) = &E_{\boldsymbol{x}_s \sim D_s}\left[\log(d(\omega(\boldsymbol{x}_s)))\right] + \\ &E_{\boldsymbol{x}_t \sim D_t}\left[\log(1-d(\omega(\boldsymbol{x}_t)))\right],\end{aligned} \tag{15}$$

交替执行式(12)和式(14)相应步骤, 直到网络收敛, 在特征提取器和域判别器之间引入梯度反转层(gradient reversal layer, GRL)以方便训练. 前向传播时, GRL 是一个恒等映射; 反向传播时, 通过乘以负的系数将梯度进行反转.

使用特征提取器提取源域人脸特征$\boldsymbol{f}_s$, 将其解耦为$\boldsymbol{f}_{pos}$和$\boldsymbol{f}_{neg}$, 并进行线性加权组合得到源域待对齐特征$\boldsymbol{f}_{ali}$, 之后通过领域对抗训练的方式将待对齐$\boldsymbol{f}_{ali}$特征与目标域特征$\boldsymbol{f}_t$对齐. 即将式(15)简化为

$$\ell_d = E\left[\log(d(\boldsymbol{f}_{ali}))\right] + E\left[\log(1-d(\boldsymbol{f}_t))\right]. \tag{16}$$

由式(5)(13)(16)得到总体损失为

$$\ell_{total} = \ell_y + \ell_d + \ell_h. \tag{17}$$

2.5 从因果角度对活体对齐域自适应的解释

因果图是一个有向无环图$G = \langle N, L\rangle$, 能够用于表示结构因果模型. 其中, 每个变量在节点集N中均有一个对应的节点, 因果链接L可描述这些变量如何相互作用. 图 1(a)可视为通常采用的领域对抗训练方法的因果图, 源域和目标域数据作为因, 训练所得模型作为果, 模型参数由源域和目标域数据共同训练得到. 边 $S\to M$ 和 $T\to M$ 分别表示源域数据和目标域数据对于最终模型参数的影响, 可以理解为源域的分类任务与目标域的对齐任务对于模型的作用. 但在这种训练范式下, 形状为对撞结构的因果图在节点 M 不固定时, 源域数据 S 和目标域数据 T 没有建立联系. 由于源域和目标域提取的特征中有着大量与活体检测任务无关的信息, 领域对抗对齐任务无法直接服务于分类任务, 故其对于模型参数的优化为次优方向.

干预是因果推断中的一项技术, 通过直接操作变量来分析因果关系. 本文通过将活体信息从源域特征中解耦, 并使目标域特征向其对齐, 可视为在源域和目标域之间施加干预操作, 对应的因果图如图 1(b)所示. 通过干预手段, 在节点 S 和节点 T 之间建立联系, 将目标域与源域中活体相关部分进行对齐, 使得对齐任务直接服务于活体检测分类任务. 图 3 详细展示了基于课程学习渐进式对齐的干预过程, 在模型训练的不同阶段施加不同的干预措施. 具体来说, 目标域特征从最初仅与源域活体无关特征对齐, 逐步过渡为与源域活体无关和相关特征组合对齐, 最终渐变为仅与源域活体相关特征对齐. 通过干预的手段, 将先验知识人为地引入到模型的训练过程, 使得模型的因果图不再是对撞结构, 从而避免源域和目标域的训练和对齐任务相互独立, 使得模型的优化

更为高效.

3 实验与结果

3.1 数据集

本文对人脸反欺诈技术中广泛使用的4个公开数据集进行测评：CASIA-MFSD(C), Idiap Replay-Attack(I), MSU-MFSD(M), OULU-NPU(O).

1)CASIA-MFSD[41]. 由50个志愿者参与录制，共计600个视频. 该数据集收集的活体和假体的人脸信息较为丰富，其中每个志愿者录制了3个活体人脸视频和9个假体人脸视频，共计12个视频. 假体攻击包括完整的平展、弯曲彩色照片假体攻击、挖去眼睛的假体攻击以及视频重放假体攻击.

2)Idiap Replay-Attack[42]. 由50个志愿者参与录制，共计1 300个视频. 这些视频是由320×240分辨率的MacBook上的网络摄像头在2种情况下拍摄，即背景单一和光照均匀的固定条件，以及背景颜色丰富和自然光照不利的复杂条件. 使用佳能PowerShot型摄像头拍摄高分辨率的人脸视频，然后使用iPad 1(1 024×768)和iPhone 3GS(480×320)进行回放，并打印在纸上.

3)MSU-MFSD[43]. 由35个志愿者参与录制，共计280个视频. 这些视频分别由分辨率为640×480和720× 480的笔记本电脑摄像头和智能手机摄像头拍摄. 主要有打印照片攻击和视频重放攻击2种不同的假体攻击.

4)OULU-NPU[44]. 由55个志愿者参与录制，共计4 950个视频. 这些视频使用6款移动设备的前置摄像头，在3种不同光照条件和背景场景中拍摄. 假体攻击类型包括打印照片攻击和视频重放攻击，使用2台不同的打印机和2台不同的显示设备进行攻击.

3.2 实现细节

主干网络采用ResNet50, 分类器为单层全连接层，输入输出维度均为(1 024, 2)，判别器使用3层全连接层，输入输出维度分别为(2, 1 024)，(1 024, 1 024)和(1 024, 2). 启发式子网络采用单层全连接层，尺寸为(2, 1 024). Batchsize大小为36, 采用随机梯度下降算法优化模型，初始学习率为1e－3, 衰减系数设置为5e－4. 使用MTCNN人脸检测模型对原始视频数据集进行人脸区域的检测和裁剪，人脸图片的大小为256×256×3. 在PyTorch深度学习框架上进行实验，主要硬件配置为Intel Core i7-7800X CPU和NVIDIA Tesla A100.

3.3 评价指标

使用半错误率*HTER*作为评价指标，其计算公式为

$$HTER=\frac{FAR+FRR}{2}\ . \tag{18}$$

其中，*FAR*是错误接受率，表示将假体攻击判断成活体人脸的比率；*FRR*是错误拒绝率，表示将活体人脸判断成假体攻击的比率. 显然，*HTER*越小，则模型性能越好.

3.4 不同方法对比

为了验证所提方法的有效性和先进性，在C, I, M, O 4个数据集上随机选择2个数据集分别作为源域与目标域，进行域自适应实验，如C-I测评协议表示C为源域且I为目标域. 由表1可见，与现有10种方法相比，本文所提DDCL方法在4个测评协议上均达到了当前先进水平，获得最佳*HTER*结果，尤其是

Table 1 *HTER* Comparison of Different Methods

表1 不同方法的*HTER*对比 %

方法	C-I	C-M	C-O	I-M	I-C	I-O	M-C	M-I	M-O	O-I	O-C	O-M	平均值
KSA[24]	39.3	15.1		33.3	**12.3**		**9.1**	34.9					24.0
Yang 等[45]	49.2	18.1		36.7	39.6		49.6	49.6					40.5
DRCN[46]	44.4	27.6		42.0	48.9		28.9	36.8					38.1
ADDA[47]	41.8	36.6		35.1	49.8		39.0	35.2					39.6
ADA[30]	17.5	9.3	29.1	30.5	41.6	**29.6**	17.7	5.1	31.2	26.8	19.8	31.5	24.1
DR-UDA[48]	**15.6**	**9.0**	28.7	29.0	34.2	38.5	16.8	**3.0**	30.2	**25.4**	**19.5**	27.4	23.1
DupGAN[49]	42.4	33.4		36.2	46.5		27.1	35.4					36.8
De-spoof[50]	28.5				41.1								34.8
Auxiliary[17]	27.6				28.4								28.0
STASN[51]	31.5				30.9								31.2
DDCL	24.6	12.2	**22.7**	**12.4**	36.2	32.0	22.8	20.0	**23.0**	28.3	22.4	**12.8**	**22.5**

注: 加粗数字表示该协议下的最佳结果，下划线数字表示次优结果.

在 I-M, M-O, O-M 这 3 个测评协议上的 *HTER* 值分别为 12.4%, 23.0% 和 12.8%, 性能分别超出 10 种对比方法中的最佳方法 DR-UDA 16.6 个百分点, 7.2 个百分点与 14.6 个百分点. 同时可以看到, 在 M-I 测评协议上的 *HTER* 值高于 DR-UDA 方法 17.0 个百分点, 仍有较大的提升空间. 总体来说, 本文所提 DDCL 方法在 12 个测评协议上获得了 22.5% 的最佳平均 *HTER* 值, 性能略超出 DR-UDA 方法 0.6 个百分点, 取得了与当前先进结果相比更强的竞争力, 能显著降低模型在目标域上的错误率, 具有更好的跨域泛化能力.

3.5 消融实验

本节通过 4 方面消融实验以考察所提方法中各个策略的有效性, 包括启发式解耦、渐进式特征对齐方式、对齐特征加权组合方式以及权重参数计算方法.

3.5.1 启发式解耦的影响

为了验证启发式解耦对实验结果的影响, 通过多重子网络将源域特征解耦成域相关和域无关 2 种特征, 以观察是否有助于模型训练. 表 2 给出是否使用启发式解耦方式的实验结果, 启发式解耦的平均 *HTER* 值较无启发式解耦降低 3.0 个百分点, 可见使用启发式解耦能显著提高模型性能, 在 I-M, I-O, M-C, M-I, M-O, O-I, O-C, O-M 等 8 个测评协议上的结果都优于不使用启发式解耦的结果, 尤其是在 I-M 与 O-C 测评协议上的提升效果最为明显, *HTER* 值分别降低 11.8 个百分点与 13.5 个百分点. 但在 C-I, C-M, C-O 和 I-C 这 4 个测评协议上, 启发式解耦的结果并不如无启发式解耦, 不过两者 *HTER* 差值最高为 5.1 个百分点. 分析其原因, 可能是 C 数据集的图像风格特征与其他数据集相比并不突出, 致使域无关解耦的效果并不是特别明显. 因此, 启发式解耦尤其适用于源域与目标域的图像风格差异明显(即不同域之间存在具有明显差异的域相关特征)的域自适应人脸反欺诈任务.

Table 2 Influence of Heuristic Disentanglement on *HTER*

表 2 启发式解耦对 *HTER* 的影响 %

解耦方式	C-I	C-M	C-O	I-M	I-C	I-O	M-C	M-I	M-O	O-I	O-C	O-M	平均值
有启发	24.6	12.2	22.7	**12.4**	36.2	**32.0**	**22.8**	**20.0**	**23.0**	**28.3**	**22.4**	**12.8**	**22.5**
无启发	**23.9**	**7.1**	**18.9**	24.2	**35.0**	38.6	24.8	22.7	24.2	34.7	35.9	15.4	25.5

注: 加粗数字表示该协议下的最佳结果.

为了进一步说明多重子网络有助于启发式解耦, 在 O-I, O-C 与 O-M 测评协议上实验了不同个数的子网络对结果的提升, 将子网络数量分别设置为 2, 3, 4. 由表 3 可知, 除 O-C 测评协议中使用 4 个子网络相较于不使用启发式解耦(表 3 中子网络数量为 0)测试的 *HTER* 值有所增加外, 其他多重子网络解耦实验结果都优于无启发式解耦, 尤其是当子网络数量为 3 时, 这 3 个测评协议的 *HTER* 值均为最佳, 表明合适数量的多重子网络对于域相关、域无关解耦有显著作用.

此外, 为了进一步验证多重子网络能够有效解耦域相关特征和域无关特征, 分别计算这 2 种特征在源域和目标域之间的最大均值差异, *MMD* 是一个衡量不同分布之间差异的度量方式, *MMD* 值越小则分布之间的差异越小. 图 4 为按源域对 12 个测评协议进行分组, 对于每个测评协议分别展示训练时源域和目标域间的域相关特征的 *MMD* 变化曲线, 以及两者域无关特征的 *MMD* 变化曲线, 图 4 中域相关特征标记为 spc 且以虚线表示, 域无关特征标记为 inv 且以实线表示. 其中, 8 个测评协议上域相关特征的 *MMD* 值约在 2~5 之间, 而 12 个测评协议上针对域无关特征的 *MMD* 值均在 0~1.5 之间, 可知在同一测评协议上域相关特征与域无关特征的 *MMD* 值差异较显著, 同时发现对于同一源域的域无关特征, *MMD* 值相近, 表明多重子网络可以有效地解耦出源域和目标域共有的域无关特征.

Table 3 Influence of Multiple Sub-networks on *HTER*

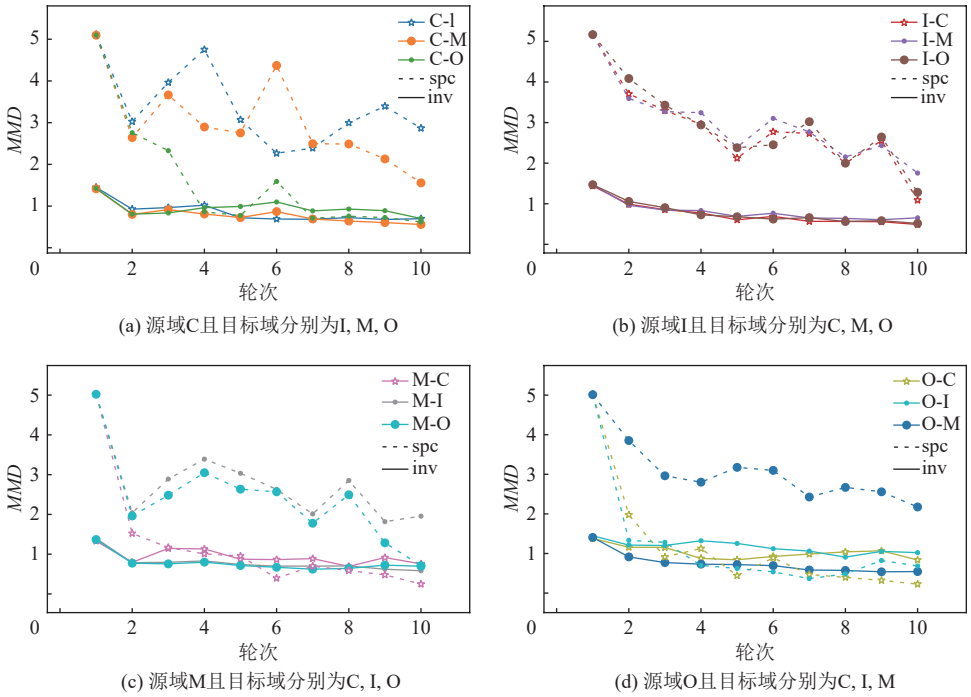
表 3 多重子网络对 *HTER* 的影响 %

子网络数量	O-I	O-C	O-M
2	29.6	29.8	15.0
3	**28.3**	**22.4**	**12.8**
4	33.0	37.1	15.2
0	34.7	35.9	15.4

注: 加粗数字表示该协议下的最佳结果.

3.5.2 课程学习渐进式对齐的影响

表 4 展示了不同对齐方式的影响, 其中 DANN 为不进行活体特征解耦, 其将源域和目标域特征直接通过对抗训练的方式进行对齐; 活体相关对齐为在对抗训练全程目标域特征仅与源域的活体相关特



(a) 源域C且目标域分别为I, M, O　(b) 源域I且目标域分别为C, M, O

(c) 源域M且目标域分别为C, I, O　(d) 源域O且目标域分别为C, I, M

Fig. 4　*MMD* variation curves between source domain and target domain during training

图 4　训练时源域和目标域间的 *MMD* 变化曲线

Table 4　Influence of Feature Alignment Patterns on *HTER*

表 4　特征对齐方式对 *HTER* 的影响　%

对齐方式	C-I	C-M	C-O	I-M	I-C	I-O	M-C	M-I	M-O	O-I	O-C	O-M	平均值
DANN	22.2	17.3	19.8	25.6	56.0	43.5	33.7	26.0	19.8	**25.4**	28.9	23.2	28.5
活体相关	**21.5**	22.7	**19.4**	22.5	52.9	40.3	39.4	22.8	**19.0**	26.9	28.4	28.8	28.7
活体无关	22.2	20.3	21.9	27.8	36.6	**30.7**	33.1	30.0	23.7	36.2	29.1	**11.9**	27.0
DDCL	24.6	**12.2**	22.7	**12.4**	**36.2**	32.0	**22.8**	**20.0**	23.0	28.3	**22.4**	12.8	**22.5**

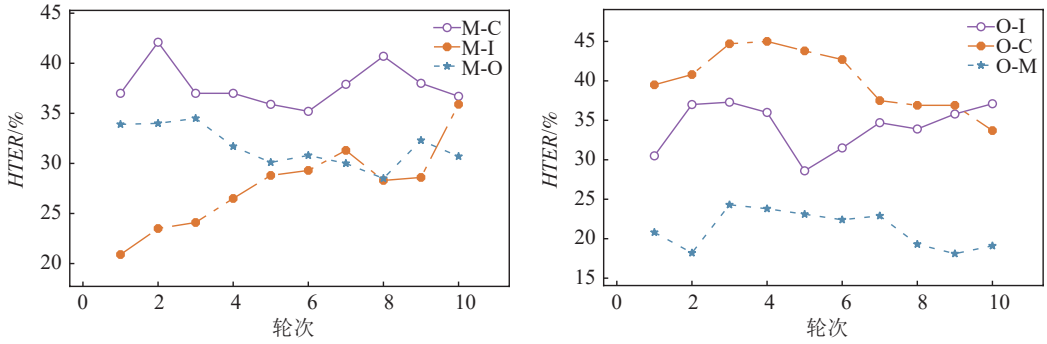
注: 加粗数字表示该协议下的最佳结果.

征对齐; 活体无关对齐则为在对抗训练全程目标域特征仅与源域的活体无关特征进行对齐.

由表 4 可知, 本文所提 DDCL 方法与其他 3 种不采用课程学习的对齐方式相比, 在 C-M, I-M, I-C, M-C, M-I 与 O-C 这 6 个测评协议上达到最佳结果, 尤其是 C-M, I-M 与 M-C 结果较活体相关对齐方式有大幅度改善, 性能提升均超过 10.0 个百分点. 从平均 *HTER* 值来看, DDCL 对齐方式取得最佳结果, 分别较 DANN、活体相关对齐、活体无关对齐这 3 种不采用课程学习的对齐方式性能提升 6.0 个百分点, 6.2 个百分点与 4.5 个百分点. 尽管活体相关对齐在 C-I, C-O 与 M-O 的结果要优于其他方式, 但优势并不十分明显, 较 DDCL 性能提升均低于 4.0 个百分点; 同时活体相关对齐的 C-M, I-C, I-O, M-C 与 O-M 性能下降严重, 尤其是 I-C, M-C 与 O-M 结果较 DDCL 性能下降均超过 16.0 个百分点. 同时发现, 活体无关对齐方式这种理想情况下无用的对齐方式, 虽然在 I-O 和 O-M 上取得了最佳结果, 但与 DDCL 的性能差异不显著. 活体无关对齐方式的 *HTER* 随训练轮次变化曲线如图 5 所示, 可见活体无关对齐大部分在训练的最初阶段取得较好结果, 但随着训练的迭代, *HTER* 波动较大, 存在逐步上升的趋势, 表明活体无关特征与分类任务相关性不强. 此外, 由图 6 的热力图可以看出, 渐进式对齐所关注的人脸区域更多, 且不局限于五官等活体无关部分. 因此总体来说, 与其他 3 种不采用课程学习的对齐方式相比, 引入课程学习进行渐进式特征对齐的有效性较为显著.

3.5.3　对齐特征加权组合方式的影响

这里验证不同加权组合方式对最终结果的影响, 主要目的是探索对齐过程前中后期的活体相关、无关的特征比重对最终模型泛化能力的影响. 表 5 对比了线性加权、二次加权和正弦加权这 3 种加权方式



(a) 源域M且目标域分别为C, I, O (b) 源域O且目标域分别为C, I, M

Fig. 5 *HTER* variation curves of different training epochs in live-unrelated feature alignment

图 5 活体无关特征对齐不同训练轮次的 *HTER* 变化曲线

(a) I-M(DANN) (b) C-M(DANN) (c) I-M(DDCL) (d) C-M(DDCL)

Fig. 6 Heat maps comparison of different feature alignment patterns

图 6 不同特征对齐方式的热力图对比

Table 5 Influence of Weighted Combination Methods on *HTER*

表 5 加权组合方式对 *HTER* 的影响 %

加权方式	I-M	I-O	M-I	M-O	O-I	O-M
线性	12.4	**32.0**	**20.0**	**23.0**	28.3	**12.8**
二次	**6.5**	37.9	**20.0**	31.5	25.0	14.2
正弦	20.7	**32.0**	20.2	30.5	**23.9**	13.3

注: 加粗数字表示该协议下的最佳结果.

在 6 个测评协议上的 *HTER* 值, 其中线性加权计算如式(11)所示, 二次加权与正弦加权分别如式(19)与式(20)所示.

$$\boldsymbol{f}_{\text{ali}} = \left(1-\alpha^2\right)\times\boldsymbol{f}_{\text{neg}} + \alpha^2\times\boldsymbol{f}_{\text{pos}}, \tag{19}$$

$$\boldsymbol{f}_{\text{ali}} = (1-\sin\alpha)\times\boldsymbol{f}_{\text{neg}} + \sin\alpha\times\boldsymbol{f}_{\text{pos}}. \tag{20}$$

由表 5 可知, 3 种加权方式均取得良好结果, 其中线性加权在 I-O, M-I, M-O 与 O-M 这 4 个测评协议上均取得最佳结果, 综合效果最优. 受限于不同数据集具有不同的活体无关和活体相关特征分布比例, 线性加权是一种较为均衡的加权方式, 在 6 个测评协议上的结果都较为准确, 故本文选用线性加权.

3.5.4 权重参数计算方式的影响

表 6 给出特征加权组合式中权重参数α的不同计算方式对结果的影响, 分别使用t/T(当前迭代次数/总迭代次数)和n/N(当前使用的样本量/训练样本总量)这 2 种计算方式进行对比, 前者的结果在 I-M, I-O, M-O 与 O-M 这 4 个测评协议上都优于后者, 综合效果最优.

Table 6 Influence of Weight Parameter Calculation Methods on *HTER*

表 6 权重参数计算方式对 *HTER* 的影响 %

计算方式	I-M	I-O	M-I	M-O	O-I	O-M
t/T	**12.4**	**32.0**	20.0	**23.0**	28.3	**12.8**
n/N	19.1	34.4	**11.8**	32.7	**19.8**	23.5

注: 加粗数字表示该协议下的最佳结果.

4 总 结

本文提出一种基于课程学习活体特征渐进式对齐的无监督域自适应人脸反欺诈方法, 通过启发式与分类器梯度的二次解耦特征, 提取活体相关与无关信息, 使用渐进式域对抗训练策略, 将目标域特征向源域对齐, 可提升无标签目标域特征与人脸反欺诈任务的相关性, 同时减轻模型优化难度. 在 4 个公

开基准数据集上的跨域实验结果表明，本文所提DDCL方法可以有效提升人脸反欺诈模型在跨域场景下的泛化能力，取得与当前先进结果相比更强的竞争力，尤其适用于源域与目标域的图像风格差异明显的域自适应人脸反欺诈任务.

与现有文献的10种方法实验对比可知，DDCL方法尽管获得了最佳平均*HTER*值，但未在所有的跨数据集相关实验中取得最佳结果，在一些测评协议上的*HTER*指标仍有较大的改善空间，后续工作将考虑提升方法的跨模型架构泛化能力，使得目前需要手工设置的一些超参数可以通过网络训练得到，如将解耦与模拟退火等方法相结合，寻找最优的解耦参数，以适应于使用更加先进的网络模型提取特征，进一步提高DDCL方法在所有测评协议上的跨域泛化性能. 未来工作也可将课程学习活体特征渐进式对齐的思路引入域泛化人脸反欺诈模型中，通过对齐多个域之间的活体相关信息，获得更加通用的活体检测特征空间，加强对于未知领域的真实人脸和假体攻击有效区分的能力，进一步提升模型的泛化性与鲁棒性.

作者贡献声明：封筠提供了关键的意见和建议，指导实验并修改和审定论文；史屹琛提出了论文的研究思路和方法，完成了实验设计、数据采集和分析，并撰写了部分论文内容；高宇豪和贺晶晶参与了对论文的修改和完善；余梓彤提供了关键的意见和建议、完善实验方案.

参 考 文 献

[1] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research, 2016, 17(1): 1−35

[2] De F P T, Anjos A, De M J M, et al. LBP-TOP based countermeasure against face spoofing attacks[C] //Proc of the Asian Conf on Computer Vision. Berlin: Springer, 2012: 121−132

[3] Liu Chengjun, Wechsler H. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition[J]. IEEE Transactions on Image Processing, 2002, 11(4): 467−476

[4] Shu Xin, Tang Hui, Yang Xibei, et al. Research on face anti-spoofing algorithm based on DQ_LBP[J]. Journal of Computer Research and Development, 2020, 57(7): 1508−1521 (in Chinese)
(束鑫, 唐慧, 杨习贝, 等. 基于差分量化局部二值模式的人脸反欺诈算法研究[J]. 计算机研究与发展, 2020, 57(7): 1508−1521)

[5] Galbally J, Marcel S, Fierrez J. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition[J]. IEEE Transactions on Image Processing, 2013, 23(2): 710−724

[6] Ikeuchi K, Miyazaki D, Tan R T, et al. Separating reflection components of textured surfaces using a single image[J]. Digitally Archiving Cultural Objects, 2008, 6(3): 353−384

[7] Yu Zitong, Peng Wei, Li Xiaobai, et al. Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement[C] //Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 151−160

[8] Li Xiaobai, Komulainen J, Zhao Guoying, et al. Generalized face anti-spoofing by detecting pulse from face videos[C] //Proc of the Int Conf on Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4244−4249

[9] Pan Gang, Sun Lin, Wu Zhaohui, et al. Eyeblink-based anti-spoofing in face recognition from a generic webcamera[C] //Proc of the IEEE 11th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2007: 1−8

[10] Pan Gang, Wu Zhaohui, Sun Lin. Liveness detection for face recognition[J]. Recent Advances in Face Recognition, 2008, 9(2): 109−124

[11] Chingovska I, Yang Jianwei, Lei Zhen, et al. The 2nd competition on counter measures to 2D face spoofing attacks[C] //Proc of the Int Conf on Biometrics. Piscataway, NJ: IEEE, 2013: 1−6

[12] Boulkenafet Z, Komulainen J, Hadid A. Face anti-spoofing based on color texture analysis[C] //Proc of the IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2015: 2636−2640

[13] Patel K, Han H, Jain A K. Secure face unlock: Spoof detection on smartphones[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(10): 2268−2283

[14] Boulkenafet Z, Komulainen J, Hadid A. Face antispoofing using speeded-up robust features and Fisher vector encoding[J]. IEEE Signal Processing Letters, 2016, 24(2): 141−145

[15] Komulainen J, Hadid A, Pietikäinen M. Context based face anti-spoofing[C] //Proc of the 6th IEEE Int Conf on Biometrics: Theory, Applications and Systems . Piscataway, NJ: IEEE, 2013: 1−8

[16] Yang Jianwei, Lei Zhen, Li S Z. Learn convolutional neural network for face anti-spoofing[J]. arXiv preprint, arXiv: 1408.5601, 2014

[17] Liu Yaojie, Jourabloo A, Liu Xiaoming. Learning deep models for face anti-spoofing: Binary or auxiliary supervision[C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 389−398

[18] Atoum Y, Liu Yaojie, Jourabloo A, et al. Face anti-spoofing using patch and depth-based CNNs[C] //Proc of the IEEE Int Joint Conf on Biometrics. Piscataway, NJ: IEEE, 2017: 319−328

[19] Wang Zezheng, Yu Zitong, Zhao Chenxu, et al. Deep spatial gradient and temporal depth learning for face anti-spoofing[C] //Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5042−5051

[20] Wang Hongfei, Cheng Xin, Zhao Xiangmo, et al. Face liveness detection based on fusional optical flow and texture features[J].

Computer Engineering and Applications, 2022, 58(6): 170−176 (in Chinese)
(王宏飞, 程鑫, 赵祥模, 等. 光流与纹理特征融合的人脸活体检测算法[J]. 计算机工程与应用, 2022, 58(6): 170−176)

[21] Wang Yahang, Song Xiaoning, Wu Xiaojun. Two-stream face spoofing detection network combined with hybrid pooling[J]. Journal of Image and Graphics, 2020, 25(7): 1408−1420 (in Chinese)
(汪亚航, 宋晓宁, 吴小俊. 结合混合池化的双流人脸活体检测网络[J]. 中国图象图形学报, 2020, 25(7): 1408−1420)

[22] Ma Siyuan, Zheng Han, Guo Wen. Deep optical strain feature map for face anti-spoofing[J]. Journal of Image and Graphics, 2020, 25(3): 618−628(in Chinese)
(马思源, 郑涵, 郭文. 应用深度光学应变特征图的人脸活体检测[J]. 中国图象图形学报, 2020, 25(3): 618−628)

[23] Yu Zitong, Zhao Chenxu, Wang Zezheng, et al. Searching central difference convolutional networks for face anti-spoofing[C] //Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5295−5305

[24] Li Haoliang, Li Wei, Cao Hong, et al. Unsupervised domain adaptation for face anti-spoofing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(7): 1794−1809

[25] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test[J]. The Journal of Machine Learning Research, 2012, 13(1): 723−773

[26] Tu Xiaoguang, Zhang Hengsheng, Xie Mei, et al. Deep transfer across domains for face anti-spoofing[J]. Journal of Electronic Imaging, 2019, 28(4): 43001

[27] Weng Zejia, Chen Jingjing, Jiang Yugang. On the generalization of face forgery detection with domain adversarial learning[J]. Journal of Computer Research and Development, 2021, 58(7): 1476−1489 (in Chinese)
(翁泽佳, 陈静静, 姜育刚. 基于域对抗学习的可泛化虚假人脸检测方法研究[J]. 计算机研究与发展, 2021, 58(7): 1476−1489)

[28] Kim Y E, Nam W J, Min K, et al. Style-guided domain adaptation for face presentation attack detection[J]. arXiv preprint, arXiv: 2203.14565, 2022

[29] Hamblin J, Nikhal K, Riggan B S. Understanding cross domain presentation attack detection for visible face recognition[C] //Proc of the 16th IEEE Int Conf on Automatic Face and Gesture Recognition. Piscataway, NJ: IEEE, 2021: 1−8

[30] Wang Guoqing, Han Hu, Shan Shiguang, et al. Improving cross-database face presentation attack detection via adversarial domain adaptation[C] //Proc of the Int Conf on Biometrics. Piscataway, NJ: IEEE, 2019: 1−8

[31] El-Din Y S, Moustafa M N, Mahdi H. Adversarial unsupervised domain adaptation guided with deep clustering for face presentation attack detection[J]. arXiv preprint, arXiv: 2102.06864, 2021

[32] Yang Luyu, Balaji Y, Lim S N, et al. Curriculum manager for source selection in multi-source domain adaptation[C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2020: 608−624

[33] Shu Yang, Cao Zhangjie, Long Mingsheng, et al. Transferable curriculum for weakly-supervised domain adaptation[C] //Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2019, 33(1): 4951−4958

[34] Gong Chen, Tao Dacheng, Maybank S J, et al. Multi-modal curriculum learning for semi-supervised image classification[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3249−3260

[35] Wang Yiru, Gan Weihao, Yang Jie, et al. Dynamic curriculum learning for imbalanced data classification[C] //Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 5017−5026

[36] Cui Shuhao, Jin Xuan, Wang Shuhui, et al. Heuristic domain adaptation[J]. Advances in Neural Information Processing Systems, 2020, 33: 7571−7583

[37] Wei Guoqiang, Lan Culing, Zeng Wenjun, et al. ToAlign: Task-oriented alignment for unsupervised domain adaptation[J]. Advances in Neural Information Processing Systems, 2021, 34: 13834−13846

[38] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C] //Proc of the IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2017: 618−626

[39] Chattopadhay A, Sarkar A, Howlader P, et al. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks[C] //Proc of the IEEE Winter Conf on Applications of Computer Vision . Piscataway, NJ: IEEE, 2018: 839−847

[40] Bengio Y, Louradour J, Collobert R, et al. Curriculum learning[C] //Proc of the 26th Annual Int Conf on Machine Learning. New York: ACM, 2009: 41−48

[41] Zhang Zhiwei, Yan Junjie, Liu Sifei, et al. A face antispoofing database with diverse attacks[C] //Proc of the 5th IAPR Int Conf on Biometrics. Piscataway, NJ: IEEE, 2012: 26−31

[42] Chingovska I, Anjos A, Marcel S. On the effectiveness of local binary patterns in face anti-spoofing[C] //Proc of the Int Conf of Biometrics Special Interest Group. Piscataway, NJ: IEEE, 2012: 1−7

[43] Wen Di, Han Hu, Jain A K. Face spoof detection with image distortion analysis[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 746−761

[44] Boulkenafet Z, Komulainen J, Li Lei, et al. OULU-NPU: A mobile face presentation attack database with real-world variations[C] //Proc of the 12th IEEE Int Conf on Automatic Face & Gesture Recognition. Piscataway, NJ: IEEE, 2017: 612−618

[45] Yang Jianwei, Lei Zhen, Yi Dong, et al. Person-specific face antispoofing with subject domain adaptation[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 797−809

[46] Ghifary M, Kleijn W B, Zhang M, et al. Deep reconstruction classification networks for unsupervised domain adaptation[C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2016: 597−613

[47] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation[C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 7167−7176

[48] Wang Guoqing, Han Hu, Shan Shiguang, et al. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection[J]. IEEE Transactions on Information Forensics and Security, 2020, 16(7): 56−69

[49] Hu Lanqing, Kan Meina, Shan Shiguang, et al. Duplex generative adversarial network for unsupervised domain adaptation[C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 1498−1507

[50] Jourabloo A, Liu Yaojie, Liu Xiaoming. Face de-spoofing: Anti-spoofing via noise modeling[C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2018: 290−306

[51] Yang Xiao, Luo Wenhan, Bao Linchao, et al. Face anti-spoofing: Model matters, so does data[C] //Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 3507−3516

Feng Jun, born in 1971. PhD, professor. Member of CCF. Her main research interests include computer vision, machine learning, and complex network analysis.

封　筠, 1971 年生. 博士, 教授. CCF 会员. 主要研究方向为计算机视觉、机器学习、复杂网络分析.

Shi Yichen, born in 1998. Master. His main research interests include face anti-spoofing and transfer learning.

史屹琛, 1998 年生. 硕士. 主要研究方向为人脸反欺诈、迁移学习.

Gao Yuhao, born in 2000. Master candidate. His main research interest includes face anti-spoofing.

高宇豪, 2000 年生. 硕士研究生. 主要研究方向为人脸反欺诈.

He Jingjing, born in 2000. Master candidate. Her main research interest includes face anti-spoofing.

贺晶晶, 2000 年生. 硕士研究生. 主要研究方向为人脸反欺诈.

Yu Zitong, born in 1992. PhD, assistant professor. His main research interests include biometric recognition and multimedia security.

余梓彤, 1992 年生. 博士, 助理教授, 主要研究方向为生物特征识别和多媒体安全.