

中图法分类号: TP18; TP391.41 文献标识码: A 文章编号: 1006-8961(2026)01-0138-16

论文引用格式: Wang C Y, Sun X Y, Li L, Zhao G Z, He Z F and Sun Z N. 2026. Iris presentation attack detection based on spatial and frequency feature fusion. Journal of Image and Graphics, 31(1):0138-0153(王财勇, 孙娴蕴, 李林, 赵光哲, 何召锋, 孙哲南. 2026. 基于空间域与频域特征融合的虹膜呈现攻击检测. 中国图象图形学报, 31(1):0138-0153)[DOI:10.11834/jig.240783]

基于空间域与频域特征融合的虹膜呈现攻击检测

王财勇^{1*}, 孙娴蕴¹, 李林¹, 赵光哲¹, 何召锋², 孙哲南³

1. 北京建筑大学智能科学与技术学院, 北京 102616; 2. 北京邮电大学人工智能学院, 北京 100876;

3. 中国科学院自动化研究所多模态人工智能系统全国重点实验室, 北京 100190

摘要: 目的 虹膜呈现攻击检测(presentation attack detection, PAD)是确保虹膜识别系统安全可靠的关键技术之一。针对现有检测方法在跨域场景下泛化性不足、难以有效检测合成虹膜的问题,提出一种基于空间域与频域特征融合的虹膜PAD模型。利用频域信息对环境变化的鲁棒性和合成/真实图像的频域差异,促使模型结合虹膜图像在空间域和频域中的互补特征,提升其检测精度和合成虹膜识别能力。**方法** 首先对输入图像使用局部二值模式(local binary pattern, LBP)算法进行处理,生成空间域图像,接着使用离散余弦变换(discrete cosine transform, DCT)将空间域图像转换为频域图像;然后利用两个EfficientNet_b0作为骨干网络,分别对空间域和频域图像进行特征提取;进一步,通过两个空间通道融合注意力模块分别融合骨干网络所提取的中、高层特征;最后,在使用二元交叉熵损失函数的基础上,引入三元组损失进行联合训练,以提升模型所提特征的区分度。此外,为了让模型重点关注图像中的虹膜区域,引入了低频语义信息引导的注意力模块。**结果** 在最新的虹膜PAD比赛LivDet-Iris 2023数据集上,将本文方法与比赛中的冠军方法及多种现有方法进行比较。相比于冠军方法,本文方法的综合检测错误率ACER1和ACER2分别下降9.32%和3.71%,其中合成虹膜的平均检测错误率APCER下降22.78%,打印虹膜的检测错误率APCER下降6.32%。相比于其他先进的现有方法,本文方法的综合检测错误率ACER1和ACER2分别下降8.91%和3.58%。此外,在LivDet-Iris 2023数据集上进行消融实验以验证所提模块的有效性,结果显示加入频域信息有效提升了模型分类真假虹膜的准确率。**结论** 本文提出的虹膜呈现攻击检测模型有效利用图像的空间域和频域信息,提高了模型检测各类型虹膜呈现攻击的能力,且表现出更好的泛化性。本文模型和代码在<https://github.com/xianyunsun/fre-iris-pad>开放。

关键词: 呈现攻击检测(PAD);虹膜识别;多域特征融合;注意力机制;合成虹膜

Iris presentation attack detection based on spatial and frequency feature fusion

Wang Caiyong^{1*}, Sun Xianyun¹, Li Lin¹, Zhao Guangzhe¹, He Zhaofeng², Sun Zhenan³

1. School of Intelligence Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 102616, China;

2. School of Artificial Intelligence, Beijing University of Posts and Telecommunication, Beijing 100876, China; 3. State Key Laboratory

of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Objective Iris recognition is a popular biometric recognition technology known for its uniqueness, stability,

收稿日期: 2025-01-02; 修回日期: 2025-04-26; 预印本日期: 2025-05-03

* 通信作者: 王财勇 wangcaiyong@bucea.edu.cn

基金项目: 北京市自然科学基金项目(4242018); 国家自然科学基金项目(62106015, 62176025, U23B2054, 62276263); 北京建筑大学“建大英才”培养工程项目(JDYC20220819)

Supported by: Beijing Municipal Natural Science Foundation (4242018); National Natural Science Foundation of China (62106015, 62176025, U23B2054, 62276263); BUCEA Pyramid Talent Training Project (JDYC20220819)

non-contact nature, and accuracy. By leveraging the complex and unique texture patterns of the human iris, iris recognition systems can accurately identify individuals and are widely used in critical sensors such as border control, prisons, banks, and coal mines. However, with the wide application of iris recognition systems, the threat of presentation attacks (PAs) has become increasingly prominent. Attackers attempt to deceive these systems using fake irises, such as printed irises, textured contact lenses, or artificial eyes, thereby enabling impersonation of others or evasion of identification. Aiming to ensure the security and reliability of iris recognition systems, iris presentation attack detection (PAD) is typically implemented as a preliminary step to distinguish between bona fide and fake inputs. Despite notable progress in iris PAD methodologies, current solutions face persistent challenges. One major issue is their sensitivity to environmental variations during image acquisition—factors such as lighting conditions, camera types, and subject-specific attributes can substantially influence detection accuracy. Another major limitation is the difficulty of effectively detecting synthetic irises, which are becoming increasingly sophisticated. This study introduces an innovative framework for iris PAD, designed to overcome these limitations through the integration of spatial-domain and frequency-domain features. Spatial-domain features capture local textures and structural information, key elements commonly exploited in PAs, while frequency-domain features exhibit improved robustness to environmental variability and can magnify differences between real and synthetic images. By leveraging the complementary advantages of both domains, the proposed method aims to enhance the detection accuracy and generalization capability. **Method** The proposed framework adopts a multifaceted approach to leverage spatial and frequency information from iris images. Initially, spatial-domain representations are generated using the local binary pattern (LBP) algorithm, which encodes texture characteristics by analyzing pixel intensity variations. The application of LBP enhances the fine-grained texture details in the spatial-domain images, emphasizing critical patterns while suppressing irrelevant noise, thereby facilitating effective feature extraction in subsequent stages. This preprocessing step ensures that the subsequent network can effectively capture distinguishing characteristics between bona fide and fake samples. Concurrently, the spatial-domain images are transformed into the frequency domain using the discrete cosine transform (DCT). The spatial and frequency domain representations are processed through two independent EfficientNet-B0 backbone networks for feature extraction. Aiming to integrate the spatial and frequency features, the framework employs spatial-channel fusion attention (SCFA) modules, which comprise parallel spatial attention and channel attention branches. These modules optimize feature fusion. SCFA modules integrate mid- and high-level representations from both domains, ensuring that their complementary information contributes effectively to the final classification. A low-frequency semantic guided attention (LSGA) module further refines the model's focus, leveraging low-frequency information to emphasize the iris region while suppressing irrelevant background details, thereby improving discrimination capability. The model is trained using a combination of triplet loss and binary cross-entropy loss (BCE loss) functions. While BCE loss helps the model learn a separation boundary in the feature space, triplet loss increases inter-class distance by encouraging the network to maximize the distance between bona fide and fake samples. This dual-loss strategy strengthens the model's capability to distinguish between the samples effectively. Overall, these components create a holistic framework designed to deliver robust performance across diverse PAD scenarios. **Result** The effectiveness of the proposed method was evaluated on the LivDet-Iris 2023 dataset, a benchmark dataset introduced by the latest LivDet-Iris 2023 competition, which features diverse and challenging PAs, including synthetic iris images. Comparative experiments were conducted against the competition's winner method and several state-of-the-art (SOTA) approaches. The proposed method resulted in a notable reduction in the classification error rate ACER1 and ACER2 by 9.32% and 3.71%, respectively, demonstrating superior performance in distinguishing bona fide and fake samples. Notably, for synthetic iris attacks, the attack presentation classification error rate (APCER) was substantially reduced by 22.78%. Compared with a recent SOTA method which only utilizes spatial information, the proposed method achieved a reduction of 10.48% and 6.22% in terms of ACER1 and ACER2, respectively, and a reduction of 24.78% on the APCER during detection of synthetic iris samples. Ablation studies were performed on the LivDet-Iris 2023 dataset to validate the contributions of individual components. Results from these experiments confirmed the effectiveness of integrating frequency-domain information, which consistently improved model accuracy in classifying bona fide and fake iris images. The SCFA and LSGA modules were also shown to further boost model performance. **Conclusion** This study introduces a novel iris PAD framework that effectively combines spatial-domain and frequency-domain

information. By leveraging the complementary strengths of the two domains, the proposed method enhances the robustness and accuracy of iris PAD models. The integration of attention mechanisms, specifically spatial-channel fusion and low-frequency semantic guidance, ensures that the model focuses on relevant regions, thereby improving its capacity to distinguish between bona fide and fake samples. Experimental results on the LivDet-Iris 2023 dataset demonstrate the superior performance of the proposed approach. This study highlights the importance of incorporating frequency-domain features into PAD models to mitigate the effects of environmental variations and improve generalizability. The code is linked at: <https://github.com/xianyunsun/fre-iris-pad>.

Key words: presentation attack detection (PAD); iris recognition; multi-domain feature fusion; attention mechanism; synthetic iris

0 引言

虹膜识别技术因其唯一性、稳定性和准确性等特性,广泛应用于现实生活中。然而,呈现攻击时常威胁着虹膜识别系统的安全性和可靠性。它通过向传感器提供假体虹膜样本,引导识别系统做出错误决策,以达到攻击者隐藏自身身份或冒用他人身份的目的。为了降低呈现攻击对识别系统的影响,开发高精度的虹膜呈现攻击检测(presentation attack detection, PAD)技术用于准确分类真实和攻击样本,已成为虹膜识别领域的关键瓶颈问题。

现有的虹膜PAD技术往往存在泛化性不足的问题。此处泛化性主要指模型在面对不同采样环境、不同采样设备等不同外界条件下的泛化性。现有的虹膜PAD模型往往只针对特定采样环境,当光照、采集设备等与PAD无关的外界条件发生变化、进而影响了虹膜样本的纹理特征时,模型往往无法应对,检测精度下降,严重影响了模型在现实场景中的部署(王财勇等,2024;Verma等,2023)。

此外,现有虹膜PAD技术还存在无法有效检测合成虹膜的问题。现有的虹膜PAD模型往往仅针对传统物理攻击方法开发,如打印虹膜、屏显虹膜、纹理隐形眼镜和义眼虹膜等。但近年来,随着生成式人工智能的不断发展,计算机合成的虹膜图像也达到了以假乱真的程度(Yadav和Ross,2023;Wang等,2022),形成了在传统物理攻击方式之外的一种新的数字攻击手段。目前鲜有针对数字攻击设计的专用检测方法,而大部分针对物理攻击设计的PAD方法直接用于检测合成虹膜时难以取得令人满意的检测精度。

在与虹膜PAD类似的人脸PAD领域中,已有研

究表明,向模型中加入频域信息能同时提高模型检测物理攻击(Fang等,2022;李盼盼等,2021)和数字攻击(Tan等,2024;Tian等,2023)的能力,启发我们在设计虹膜PAD方法时,不仅要考虑利用包含更多纹理细节的空间域特征,也应该考虑结合频域特征。因此,针对虹膜PAD技术在上述两方面中存在的不足,本文提出引入频域信息,并使用深度学习方法对频域和空间域信息进行融合。一方面,光照等外界环境变化主要影响全局亮度等图像的低频部分,而进行PAD所需的线索主要来自细节纹理等图像的中高频部分,使用频域信息能一定程度上对二者进行解耦,让模型更容易区分并利用图像中与PAD相关的分量,减少无关分量对检测的影响,有助于提高模型对环境变化的鲁棒性。另一方面,对于数字攻击样本,虽然合成虹膜在空间域中与真实虹膜非常接近,但在高频部分往往存在与生成模型相关的特殊纹理,加入频域信息有助于检测合成图像。图1展示了真实虹膜、合成虹膜和打印虹膜及其各自在经过局部二值模式(local binary pattern, LBP)提取特征后的特征图在经过离散余弦变换(discrete cosine transform, DCT)及快速傅里叶变换(fast Fourier transform, FFT)后的频谱图像,虽然这些样本在空间域比较相似,但在频域上却有明显差异,说明频域图像可以为虹膜PAD提供有效线索。

除了检测方法以外,算法的评估也是制约虹膜PAD发展的一个瓶颈。目前主流虹膜PAD数据集均存在规模较小、攻击方法种类有限的问题,难以有效反映虹膜PAD方法在真实场景中的应用潜力。如LivDet-Iris 2017比赛数据集中(Yambay等,2017)的IIITD-WVU子集和Notre Dame子集仅包含真实样本和戴美瞳的攻击样本。CASIA-Iris-Fake数据集(Sun等,2014)虽然同时包含打印、塑料义眼、美瞳

和合成虹膜,但其中的合成虹膜图像不包含眼周部分,与其他样本差异较大;而另外3种攻击方法每种样本数均不足1 000幅。此外,已有数据集大多采集时间较早,而近些年,打印虹膜、屏显虹膜等假体攻击的设备也在更新,在生成假体样本的质量上明显改进,加大了检测难度。因此,在评估虹膜PAD模型时,有必要选择近几年采集、使用较新的攻击设备且攻击样本中包含高质量合成虹膜的数据集,以尽可能真实全面地反映模型在实际应用场景下的表现。

本文为解决现有虹膜PAD方法面对未知环境和未知攻击类型,特别是合成虹膜时泛化性不足的问题,提出一种基于空间域与频域特征融合的虹膜PAD模型,并进行了全面而有效的性能评估。具

体贡献如下:1)提出一种融合空间域特征和频域特征信息的虹膜PAD模型,使用深度神经网络提取及融合图像特征,并在二分类问题常用的交叉熵损失基础上引入三元组损失辅助训练,有效提高模型检测的准确性和泛化性,并显著提高模型检测合成虹膜的能力。模型和代码在 <https://github.com/XianyunSun/fre-iris-pad> 开放。2)考虑到空间域和频域图像的差异性,在模型中引入适配虹膜PAD的空间通道融合注意力模块,实现对两种特征的有效融合。3)在模型中引入低频语义信息引导的注意力模块,基于输入图像的低频分量进行注意力计算,引导模型自动关注图像中的虹膜区域,从而免除了虹膜定位、归一化等步骤,减少计算量,提高模型的轻量性和检测的实时性。

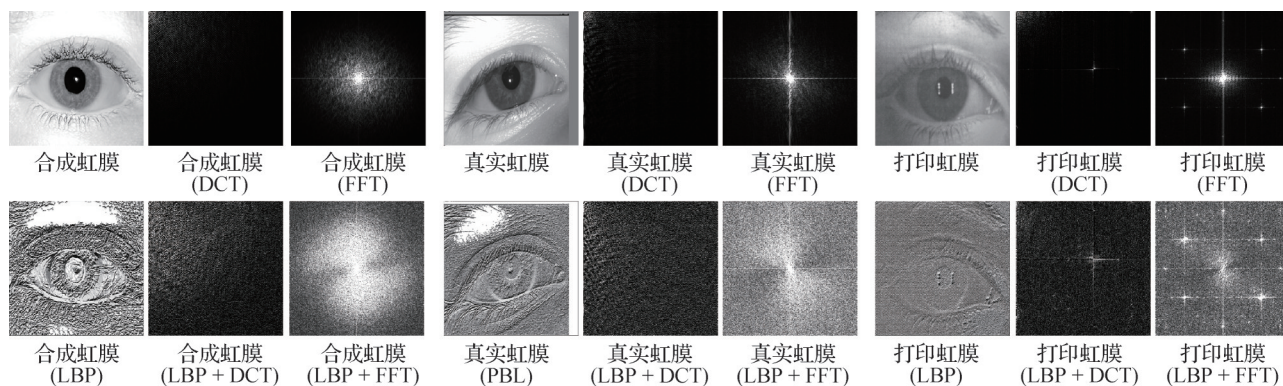


图1 合成虹膜、真实虹膜和打印虹膜在进行LBP增强及DCT变换和FFT变换后的对比图

Fig. 1 Comparison of the synthetic, live and printed iris images after LBP, DCT or FFT processing

1 相关工作

1.1 传统虹膜呈现攻击检测方法

传统虹膜PAD算法通常使用手工设计的特征描述子提取虹膜的细节纹理特征,并采用支持向量机(support vector machine, SVM)等算法进行真假分类。常用的特征描述子包括LBP(Ojala等,1994)、二值化统计图像特征(binanzed statistical image features, BSIF)(Kannala和Rahtu,2012)等。此外,Khade等人(2023)提出使用Haar变换提取虹膜图像中的频域信息,并使用SVM训练分类模型。传统检测算法的优势在于其对算力需求低,但一旦采集设备和环境光照等发生变化,或出现了未知的攻击方法,其准确率往往会大幅下降,缺乏泛化性。

1.2 基于深度学习的虹膜呈现攻击检测方法

随着神经网络的发展,深度学习方法也逐渐应用于虹膜活体检测之中。最早的基于深度学习的虹膜PAD模型是由Silva等人(2015)提出的CLD-net,该方法使用一个简单的两层神经网络,以端到端的方式对虹膜样本进行真假分类。随着算力发展和数据集规模的增长,将更复杂的神经网络引入到这一领域,特别是将一些经典的图像分类骨干网络微调以适应虹膜PAD任务,如Trokielewicz等人(2018)提出使用VGG-16(Visual Geometry Group 16-layer network)模型对尸体虹膜进行分类;Yadav等人(2019)及Sharma和Ross(2020)分别提出基于DenseNet的DensePAD和D-netPAD。也有更新的研究(Tapia等,2025)探索视觉基础模型(DinoV2(Oquab等,2024)和VisualOpenClip(Cherti等,2023))在虹膜PAD任务上微调的表现。虽然

DinoV2在微调后的检测精度高于DenseNet等骨干网络,但庞大的参数量和计算开销限制了其在实际场景中的部署。

为了加强模型提取有效纹理特征的能力,一些方法引入了注意力机制。如Chen和Ross(2021)在深度模型最后加入了并联的空间注意力模块和通道注意力模块,引导模型更关注图像中的虹膜区域;吕梦凌等人(2022)提出使用循环注意力模拟人眼的视觉特性,加强模型对于纹理隐形眼镜的检测能力;Jaswal等人(2024)提出稠密特征校准注意力辅助网络DFCANet模型,通过采用特征校准卷积和残差学习,使模型不仅能关注到局部细节纹理,还能关注到图像的全局信息。除注意力机制外,刘明康等人(2020)提出使用残差网络将原始图像映射到可分离的灰度图像空间,以此增大真假虹膜的特征差异,提高分类的准确性。

此外,还有一些方法结合了多源信息进行虹膜活体检测,通过结合它们的互补优势,提升了模型的泛化性。如Choudhary等人(2022)首先训练Densenet和多种手工特征算子分别提取特征,并用SVM对这些特征进行分类,然后通过分数层融合及统计学分析等方法中选出一组最合适的专用特征,显著提高模型检测纹理隐形眼镜的能力。Kuehlkamp等人(2019)首先对图像使用多个不同的BSIF算子提取特征,并使用卷积神经网络分别进行分类,再对多个预测结果进行分数层融合。Agarwal等人(2022)则融合了两种不同的输入图像:原始图像和经过对比度受限的自适应直方图均衡化(contrast-limited adaptive histogram equalization, CLAHE)增强后的图像,将二者分别作为输入进行分类,并对预测分数进行加权平均,有效提升了模型在检测纹理隐形眼镜时的泛化性。Li等人(2022)和另一些方法则同时关注虹膜区域和其他眼部区域,如Graganiello等人(2016)认为可以通过引入巩膜区域的特征来更准确地区分佩戴纹理隐形眼镜、佩戴透明隐形眼镜和不戴隐形眼镜的虹膜图像;Hoffman等人(2019)则融合了虹膜和眼周区域的特征,提升模型的泛化性。

在结合频域信息方面,Li等人(2022)将未知环境下的虹膜PAD视为域适应问题,将目标域图像的低频部分和源域图像的高频部分相结合构造训练样本,以增加模型的泛化性。但这种做法需要至少两

个不同域上的数据进行训练,且模型输入仍为空间域图像,没有显式地利用频域信息。

与上述多源信息融合的方法不同,本文所提方法不仅在空间域对输入图像进行真假分析,也发掘了更鲁棒的频域特征进行补充判别,并使用深度学习提取两个域的信息,所提特征相比传统方法更有表征能力。此外,本文还将传统LBP特征结合到深度学习模型中,最终联合增强模型分类虹膜样本真假时的准确性和泛化性。

2 本文方法

如图2所示,本文所提模型包括提取空间域和频域特征的双流网络(黄色部分)、空间通道融合注意力模块(蓝色部分)、低频信息引导的注意力模块(紫色部分)及三元组损失和二元交叉熵损失的联合使用(红色部分)。接下来,本节将详细介绍其中各个模块的模型结构。

2.1 空间域和频域特征提取双流网络

图1中的例子展示出真实虹膜和攻击样本之间在频域图像上的差异,其中DCT变换图像与FFT变换图像的右下1/4部分对应。由图1可以看出,真实虹膜与合成虹膜虽然在空间域上比较相似,但经过LBP和DCT处理后则呈现出明显差异:真实虹膜在高频段细节丰富,而合成虹膜则在高频段较为平滑;打印虹膜则在高频区域出现了其他图像没有的特殊纹理。这表明频域图像可以为虹膜真假分类提供有判别力的信息,且LBP可以显著放大真实样本与攻击样本之间的差异,使模型更容易分类。

如图2中黄色方框及箭头所示,本文所提模型主要包含空间域和频域两个特征提取分支。原始输入图像首先通过LBP算法进行增强处理,所得结果作为空间域特征提取分支的输入,接着使用DCT算法进一步将其转换为频域图像,作为频域特征提取分支的输入。

进一步地,考虑到虹膜PAD模型往往需要部署到嵌入式设备上,能调用的算力和存储空间有限,因此选用参数量较小的EfficientNet_b0网络(Tan和Le,2019)作为骨干网络,分别提取空间域图像和频域图像的特征。此外,由于虹膜PAD往往以图像细节纹理信息作为分类依据,所以,在直接使用骨干网络最终输出的特征之外,本文模型还使用骨干网络

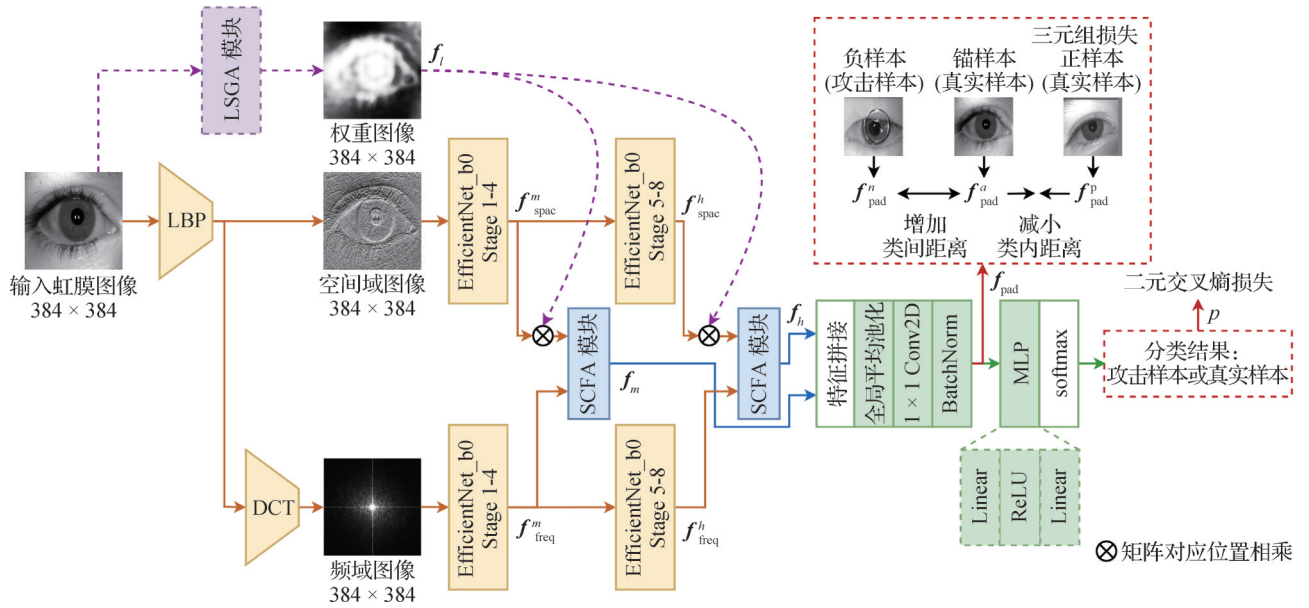


图2 本文方法的模型结构总体框架图

Fig. 2 The general workflow of the proposed method

提取的中层特征作为对图像细节信息的补充。

2.2 空间通道融合注意力(SCFA)模块

空间域和频域特征在语义上存在差异,直接将网络两个分支提取的特征简单拼接无法达到最佳效果。为了解决这一问题,本文模型引入空间通道融合注意力(spatial-channel fusion attention, SCFA)模块,以更好地进行特征融合。如图3所示,此模块由空间注意力分支和通道注意力分支并联而成,其中通道注意力调整频域和空间域特征在通道维度的表示,引导模型关注不同通道的相对重要性;空间注意力则改变二者特征图在空间位置上的分布,引导模型聚焦于对呈现攻击检测重要的空间区域。

SCFA模块在处理双流网络提取的中层和高层特征时计算流程相同,此处以中层特征为例详细描述SCFA模块的计算过程。将空间域分支提取的中层特征记为 f^m_{spac} ,频域分支提取的中层特征记为 f^m_{freq} ,二者大小均为 $C/2 \times H \times W$,其中, C, H, W 分别为特征的通道数、高和宽。二者在通道维拼接得到的特征 $f \in \mathbf{R}^{C \times H \times W}$ 为SCFA模块的输入。具体为

$$f = [f^m_{\text{spac}}, f^m_{\text{freq}}] \quad (1)$$

在空间注意力分支,首先分别对 f 中来自空间域的特征和来自频域的特征沿着通道维度进行平均池化(AveragePooling)和最大池化(MaxPooling),得到共计4个 $1 \times H \times W$ 大小的特征图,分别记为 $f^{\text{avg}}_{\text{spac}}$ 、 $f^{\text{max}}_{\text{spac}}$ 、 $f^{\text{avg}}_{\text{freq}}$ 和 $f^{\text{max}}_{\text{freq}}$,这一操作能在尽可能保持原空间域

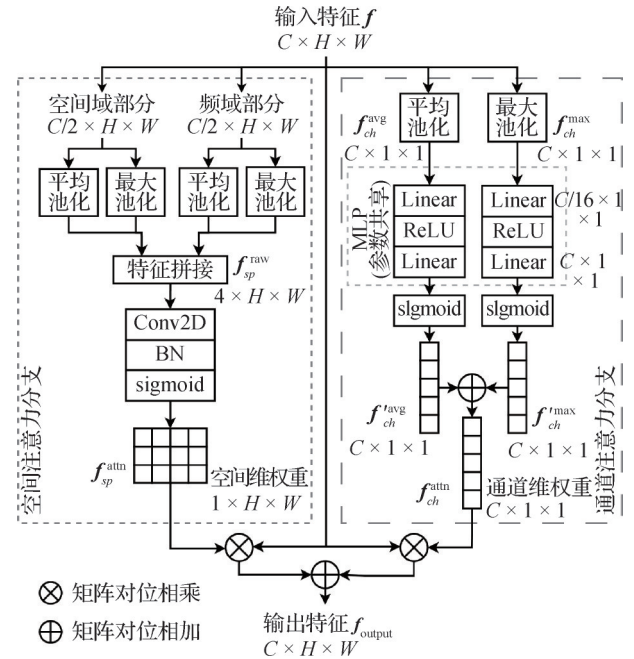


图3 所提SCFA模块结构框架图

Fig. 3 Workflow of the proposed SCFA module

特征中各个通道的关键信息的同时降低特征矩阵维度,减少计算量。具体为

$$f^{\text{max}}_{\text{spac}} = \text{MaxPooling1D}(f^m_{\text{spac}}) \quad (2)$$

$$f^{\text{avg}}_{\text{spac}} = \text{AveragePooling1D}(f^m_{\text{spac}}) \quad (3)$$

$$f^{\text{max}}_{\text{freq}} = \text{MaxPooling1D}(f^m_{\text{freq}}) \quad (4)$$

$$f^{\text{avg}}_{\text{freq}} = \text{AveragePooling1D}(f^m_{\text{freq}}) \quad (5)$$

接着,将这些特征图在通道维进行拼接,得到

$f^{\text{raw}}_{\text{sp}} \in \mathbf{R}^{4 \times H \times W}$,具体为

$$f_{sp}^{raw} = [f_{spac}^{max}, f_{spac}^{avg}, f_{freq}^{max}, f_{freq}^{avg}] \quad (6)$$

为了进一步提取特征图在空间上的关系,将 f_{sp}^{raw} 通过一个卷积模块进行特征细化,通道数映射为1。该模块由1个卷积层(Conv2D)后接1个批量归一化(batch norm, BN)层组成。之后通过sigmoid函数将细化后的特征进行归一化,生成空间注意力权重映射 f_{sp}^{attn} ,具体为

$$f_{sp}^{attn} = \text{sigmoid}(\text{BN}(\text{Conv2D}(f_{sp}^{raw}))) \quad (7)$$

在通道注意力分支,首先对 f 在空间维度上进行平均池化和最大池化,得到2个 $C \times 1 \times 1$ 大小的特征图 f_{ch}^{avg} 和 f_{ch}^{max} ,分别代表各个通道在空间维度的平均信息和最大信息,具体为

$$f_{ch}^{max} = \text{MaxPooling2D}(f) \quad (8)$$

$$f_{ch}^{avg} = \text{AveragePooling2D}(f) \quad (9)$$

通过一个参数共享的多层感知机(multilayer perceptron, MLP)分别学习 f_{ch}^{max} 和 f_{ch}^{avg} 中通道之间的全局关系,并通过sigmoid函数进行权重归一化,从而为每个通道生成一个对应的注意力权重,分别得到 f'_{ch}^{max} 和 f'_{ch}^{avg} ,具体为

$$f'_{ch}^{max} = \text{sigmoid}(\text{MLP}(f_{ch}^{max})) \quad (10)$$

$$f'_{ch}^{avg} = \text{sigmoid}(\text{MLP}(f_{ch}^{avg})) \quad (11)$$

式中,MLP包含1个“挤压”线性层,其将特征图的通道数映射为 $C/16$ 、1个ReLU(线性修正单元,rectified linear unit)激活层和1个“扩张”线性层,其将特征图的通道数映射回 C 。

进一步地,取上述两个通道注意力权重向量的平均值作为最终的通道注意力权重向量 f_{ch}^{attn} ,具体为

$$f_{ch}^{attn} = 0.5 \times f'_{ch}^{max} + 0.5 \times f'_{ch}^{avg} \quad (12)$$

在获取了空间注意力权重和通道注意力权重之后,将输入特征 f 分别与之按位相乘(表示为 $\otimes$),并进一步地取其结果的平均值作为注意力机制筛选后的特征 f_{output} ,具体为

$$f_{output} = 0.5 \times f \otimes f_{sp}^{attn} + 0.5 \times f \otimes f_{ch}^{attn} \quad (13)$$

2.3 低频语义信息引导的注意力(LSGA)模块

对于虹膜PAD而言,呈现攻击主要发生在虹膜区域,但是原始采集的虹膜图像也可能会包含一些干扰的信息,如眼周特征或图像边框等。因此,如果直接将原始图像用于模型训练,模型可能会关注到这些无关信息,导致泛化性下降。针对此问题,部分现有方法进行了显式的虹膜定位或归一化操作,驱使模型在虹膜区域进行真假判别,然而这带来了额外的时间消耗,不利于模型的轻量化和检测的实时性。为了解决这一问题,本文提出由低频语义信息引导的注意力(low-frequency semantic guided attention, LSGA)模块,使得模型可以隐式地关注到虹膜区域。

与进行呈现攻击检测时需要细节纹理不同,区分图像中的虹膜和非虹膜区域需要全局信息,过多的细节反而可能造成干扰。因此首先对图像进行低频滤波,事先排除一部分高频细节。具体地,如图4所示,首先对输入图像进行DCT变换,从而获取频域图像;接着使用低通滤波器对之进行掩膜操作,获取低频特征图,并随之进行DCT反变换,得到输入图像的低频分量。进一步地,对图像的低频分量进行连续卷积操作,从而得到了关于虹膜区域的注意力权重映射 f_1 。由图4中的“权重图像”可以看到,在无需额外监督的情况下,模型自动学到了为虹膜区域赋予更接近1的权重(图中白色部分),而对眼周等其他无关区域则赋予更接近0的权重(图中黑色部分),有效区分了图像中的虹膜和非虹膜区域。

显然,得到的注意力权重仅与空间域图像相关,因此如图1中紫色箭头所示,将 f_1 下采样后分别与空间域分支提取的中层特征 f_{spac}^m 和高层特征 f_{spac}^h 对位相乘,得到更新后的空间域特征,该特征聚焦于虹膜区域信息。

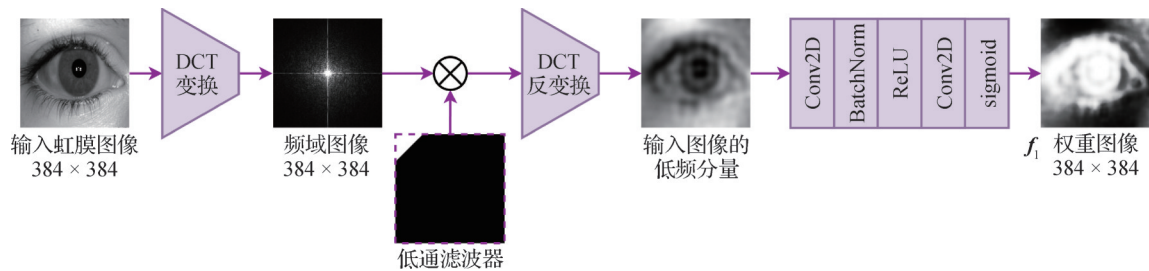


图4 所提LSGA模块结构框架图

Fig. 4 Workflow of the proposed LSGA module

2.4 分类器及其损失函数

在获取了空间域和频域融合后的中层和高层特征映射后,将二者沿通道维拼接,并通过全局平均池化进行特征聚合,从而得到初始的多域、多粒度融合特征向量,而后通过 1×1 卷积层进一步进行通道维的融合和压缩,得到降维后的特征向量 $f_{\text{pad}} \in \mathbf{R}^{224}$ 。在此之后,将 f_{pad} 输入到包含一个隐藏层的 MLP 分类器中进行虹膜真假分类。MLP 的具体结构如图 2 所示,其第 1 个线性层的输出通道数为 112,第 2 个线性层的输出通道数为 2。进一步地,MLP 的输出结果通过 softmax 函数进行归一化,从而得到真假虹膜的预测分数,其中真实虹膜的分数表示为 $p \in [0, 1]$ 。

虹膜 PAD 属于二分类问题,因此需要使用二元交叉熵(binary cross-entropy, BCE)损失进行训练,确定分类边界。BCE 损失 L_{bce} 的计算方法为

$$L_{\text{bce}} = -[y_{\text{gt}} \log p + (1 - y_{\text{gt}}) \log (1 - p)] \quad (14)$$

式中, y_{gt} 表示样本的真实标签, $y_{\text{gt}} = 0$ 代表攻击样本, $y_{\text{gt}} = 1$ 代表真实样本。

此外,在常规 BCE 损失的基础上,本文还引入了三元组损失辅助训练。这是因为一部分攻击样本非常逼真,其特征靠近真假虹膜的分类边界,导致模型不够鲁棒,容易产生误检。而三元组损失可以显式地增大类间距离,同时减少类内差异,使特征远离分类边界、更有区分度。三元组损失 L_{tri} 的计算方法为

$$L_{\text{tri}}(f_{\text{pad}}^a, f_{\text{pad}}^p, f_{\text{pad}}^n) = \max \left\{ \left\| f_{\text{pad}}^a - f_{\text{pad}}^p \right\|_2 - \left\| f_{\text{pad}}^a - f_{\text{pad}}^n \right\|_2 + m, 0 \right\} \quad (15)$$

式中, $f_{\text{pad}}^a, f_{\text{pad}}^p, f_{\text{pad}}^n$ 分别表示一个三元组中锚样本、正样本和负样本的特征。在构建三元组时,首先从训练数据中随机选取一个样本作为锚样本,之后选取与锚样本类别标签相同的样本作为正样本,选取与锚样本类别标签不同的样本作为负样本。如锚样本为真实样本,则正样本从其他真实样本中随机选择,负样本从攻击样本中随机选择。三元组损失分别计算锚样本—正样本及锚样本—负样本之间的欧氏距离,用 $\| \cdot \|_2$ 表示。 m 为超参数,用于控制锚样本—正样本和锚样本—负样本之间距离的期望间隔,即三元组损失要求模型提取的不同类样本特征之间的距离至少比同类样本特征之间的距离大 m ,保证了特征均能远离分类边界,便于模型进行分类。

图 5 展示了加入三元组损失对模型所提特征分布的影响,其中蓝色和橙色圆点分别代表训练集中的真实样本和攻击样本的特征,使用 t-SNE (t-distributed stochastic neighbor embedding) 进行特征降维及可视化:图 5(a) 为只使用 BCE 损失训练时的结果,真实样本特征与攻击样本特征距离较近,甚至发生了一部分混合,模型难以准确分类这些混合在一起的特征;而图 5(b) 为使用 BCE 损失和三元组损失联合训练的结果,两类样本特征之间出现了清晰的分界,有助于模型进行正确分类。

模型的总损失 L 为上述两个损失函数的加权平均,表示为

$$L = L_{\text{bce}} + \lambda L_{\text{tri}} \quad (16)$$

式中, λ 为权重,用来度量不同损失函数的重要性。

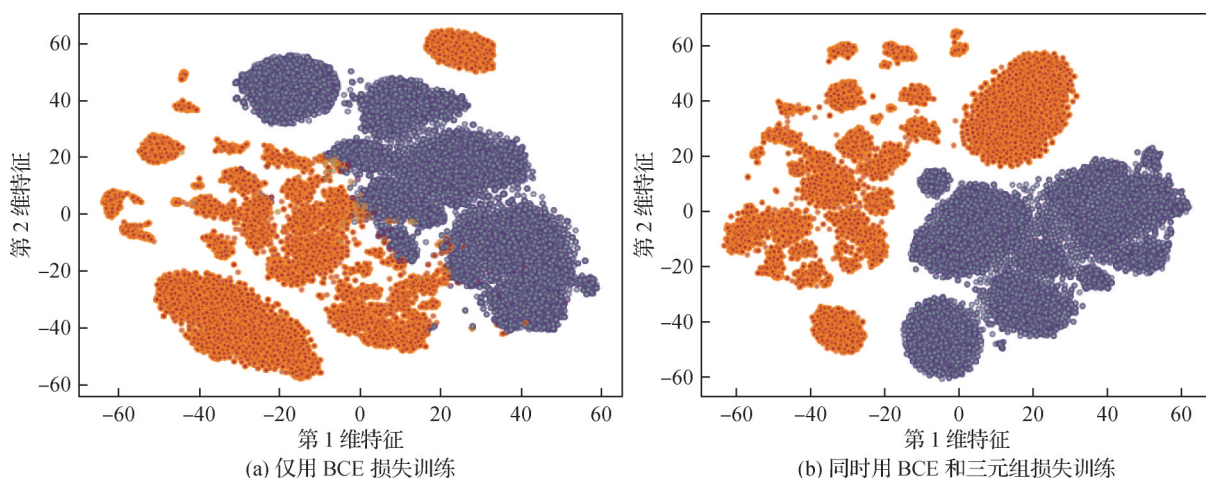


图 5 训练集中模型所提虹膜特征的 t-SNE 可视化结果

Fig. 5 t-SNE visualization of the iris features extracted by the proposed model on the training set
(a) trained with only BCE loss; (b) trained with both BCE and triplet loss

3 实验设计与结果分析

为了尽可能真实全面地反映所提方法在真实场景下的表现,本文选择在 LivDet-Iris 2023 比赛测试协议(Tinsley 等, 2023)下与现有方法进行比较,并在此协议下进行消融实验,以验证所提方法中各个模块的有效性。本节将对实验部分进行详细阐述。

3.1 实验协议

本文沿用 LivDet-Iris 2023 比赛协议对所提方法进行性能评估。LivDet-Iris 2023 比赛允许使用任意数据集进行模型训练,而后在比赛组织者提供的同一数据集下进行测试。比赛中使用的评价指标如下:

呈现攻击误判率(attack presentation classification error rate, APCER)和真实呈现误判率(bona fide presentation classification error rate, BPCER)作为评价模型准确性的指标。APCER 代表模型将攻击样本误分类为真实样本的比例,BPCER 代表模型将真实样本误分类为攻击样本的比例。更进一步地,比赛按照用于呈现攻击的工具(presentation attack instruments, PAI)分别计算其呈现攻击误判率 f_{APCER}^{PAI} ,以便于全面分析模型对不同攻击方式的检测性能。具体为

$$f_{APCER}^{PAI_j} = \frac{1}{N_{PAI_j}} \sum_{i=1}^{N_{PAI_j}} pred_i \quad (17)$$

$$f_{BPCER} = \frac{1}{N_{BF}} \sum_{i=1}^{N_{BF}} (1 - pred_i) \quad (18)$$

式中, $f_{APCER}^{PAI_j}$ 表示第 j 种攻击方式对应的呈现攻击误判率, N_{PAI_j} 代表第 j 种攻击方式的样本数, N_{BF} 为真实样本的数量。 $pred_i$ 表示模型对第 i 个样本的预测标签(0或1), 1为真实样本, 0为攻击样本。

此外,为了对参赛方法进行排名,比赛还计算了两种平均呈现误判率 $ACER1$ 和 $ACER2$, 具体为

$$f_{ACER1} = \frac{1}{2} \sum_{j=1}^{Num_{PAI}} \frac{N_{PAI_j}}{N_{PA}} \times f_{APCER}^{PAI_j} + \frac{1}{2} f_{BPCER} \quad (19)$$

$$f_{ACER2} = \frac{1}{2} \times \frac{1}{Num_{PAI}} \sum_{j=1}^{Num_{PAI}} f_{APCER}^{PAI_j} + \frac{1}{2} f_{BPCER} \quad (20)$$

式中, Num_{PAI} 表示攻击种类的总数,即为7种, N_{PA} 表示攻击样本的总数。由式(19)和式(20)可知, f_{ACER1} 考虑了数据集中不同攻击种类样本的数量占比,因

此,占比较大的攻击类型的检测结果将对这一指标产生更大的影响。而 f_{ACER2} 同等地对待不同的攻击类型,因此结果不受攻击类型样本数量的影响。

3.2 数据集

3.2.1 训练集

LivDet-Iris 2023 比赛规则不限制训练集,因此,本文汇总了多个主流数据集用做训练集。这一新的大规模结合数据集(combined dataset)的具体组成及样本数量详见表1。

3.2.2 测试集

本文使用 LivDet-Iris 2023 比赛中提出的数据集作为测试集。该数据集包含4种物理攻击类型和3种不同质量的数字攻击类型,共13 332个样本。数据集中的物理攻击类型包括由5个不同品牌打印机打印的虹膜图像(printed eyes, PE)、将隐形眼镜放在假眼或打印虹膜上之后采集的图像(fake or printed eyes with add-ons, F/P)、戴纹理隐形眼镜的虹膜图像(textured contact lenses, CL),以及通过 Kindle 电子屏幕上显示的虹膜图像(eyes displayed on Kindle e-Ink, ED)。与其他主流数据集相比,此数据集中的物理攻击样本在采集时使用了近几年生产的更先进的打印和屏幕显示设备,攻击样本更为逼真,更具有挑战性,也更贴近实际场景。合成虹膜样本主要使用 StyleGAN2(Karras 等, 2020)和 StyleGAN3(Karras 等, 2021)生成,包含低、中、高3个质量类别,分别来自生成模型训练的早期、中期和后期。此数据集是近几年内唯一包含大规模合成虹膜的 PAD 数据集。表2展示了此数据集中各个类型的样本数量,图6展示了此数据集中的一些样本。

3.3 实验设置

在训练模型时,使用在 ImageNet 上预训练的权重对 EfficientNet_b0 骨干网络进行初始化,其他模块则随机初始化。采用随机剪裁和翻转、随机 Gamma 增强和随机模糊的方法对数据集进行扩充,以缓解模型的过拟合问题。此外,为了更好地保留虹膜图像细节,提高检测的准确性,选用分辨率更高的 384×384 像素取代常规的 224×224 像素作为模型的输入尺寸。为此,原始输入图像需要缩放至 384×384 像素。训练时,批次大小为32,且保证每批次中的真实样本和攻击样本数量相同。使用随机梯度下降算法(stochastic gradient descent, SGD)优化模型参数,动量设为0.9,初始学习率为0.0001,之后每5轮乘

表 1 本文所提结合数据集的样本构成
Table 1 The construction of the combined dataset

图像种类	呈现攻击类型	数据集名称	图像数量/幅	图像数量合计/幅
真实虹膜图像	-	IITD Contact Lens Iris (Kohli 等, 2013)	4 316	35 312
		LivDet-Iris 2017 Notre Dame subset (Yambay 等, 2017)	2 400	
		IITD Combined Spoofing Database (Kohli 等, 2016)	6 022	
		Notre Dame Photometric Stereo Iris (Czajka 等, 2019)	3 132	
		NDIris3D (Fang 等, 2021)	3 442	
		CASIA-Iris-Fake (Sun 等, 2014)	6 000	
		ND Cross Sensor Iris 2013 (Doyle 等, 2013)	10 000	
假体虹膜图像	义眼	CASIA-Iris-Fake (Sun 等, 2014)	640	35 459
	戴纹理隐形眼镜的虹膜	IITD Contact Lens Iris (Kohli 等, 2013)	2 267	
		LivDet-Iris 2017 Notre Dame subset (Yambay 等, 2017)	1 701	
		LivDet-Iris 2017 Notre Dame subset (Yambay 等, 2017)	2 400	
		Notre Dame Photometric Stereo Iris (Czajka 等, 2019)	2 664	
		NDIris3D (Fang 等, 2021)	3 408	
	打印的戴纹理隐形眼镜的虹膜	LivDet-Iris 2017 IITD-WVU subset (Yambay 等, 2017)	2 902	
		Iris printed CL1 (Tapia 等, 2022)	1 500	
	病变虹膜	Warsaw-BioBase-Disease-Iris v2.1 (Trokielewicz 等, 2015)	1 099	
	尸体虹膜	Warsaw-BioBase-Post-Mortem-Iris v1-1 (Trokielewicz 等, 2016)	978	
	打印虹膜	LivDet-Iris 2017 IITD-WVU subset (Yambay 等, 2017)	2 904	
	合成虹膜	CASIA-Iris-Syn (Wei 等, 2008)	10 000	
		StyleGan-Iris-PADv1 (Maureira 等, 2022)	2 996	
总计			70 771	

注:“-”表示无。

表 2 LivDet-Iris 2023 测试集样本构成
Table 2 The construction of the LivDet-Iris 2023 testset

图像类型	图像数量/幅
真实图像	6 500
打印虹膜图像(PE)	522
隐形眼镜放在假体上的图像(F/P)	266
戴纹理隐形眼镜的虹膜图像(CL)	4
屏显虹膜图像(ED)	40
低质量生成虹膜图像(LQ)	2 000
中质量生成虹膜图像(MQ)	2 000
高质量生成虹膜图像(HQ)	2 000
总计	13 332

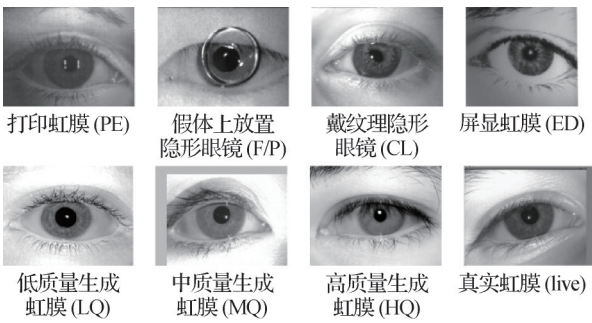


图 6 LivDet-Iris 2023 测试集中的样本示例

Fig. 6 Examples from the LivDet-Iris 2023 test dataset

3.4 性能比较

3.4.1 对比方法

以 0.7, 共训练 60 轮。在三元组损失中, m 设置为 0.5, λ 设置为 1.0。

表 3 中 Winner-LivDet-Iris 2023 为 LivDet-Iris 2023 比赛中在 ACER2 指标上取得最优的方法。D-netPAD (raw) 为 Sharma 和 Ross (2020) 提出的方

法。LMFD-PAD(Fang等,2022)为结合频域信息的人脸PAD方法,该方法提出多层次频域分解(multi-level frequency decomposition,MFD)模块,分别提取输入图像在低、中、高频段分量,对原图和频域分解之后的图像分别提取特征并进行特征融合。DFCANet(Jaswal等,2024)为近期提出的结合自校准卷积,只在空间域对图像进行处理的虹膜PAD方法。为了保证公平性,上述所有方法均在3.2节中描述的训练集和测试集上进行训练和测试,超参数等均采用3.3节中的设置。此外,在后续的实验中,出于简单考虑,均取0.5作为阈值进行虹膜真假判别。若预测真实样本的概率 $p > 0.5$,则样本判定为真实虹膜,否则为呈现攻击。

3.4.2 与参赛方法相比

表3展示了在LivDet-Iris 2023测试协议下,本文方法与Winner-LivDet-Iris 2023方法的结果比较。Winner-LivDet-Iris 2023方法虽然能较为准确地识别出真实样本,但也将绝大部分合成虹膜误判为真实样本。与之相比,本文方法在保证可以准确判别真实虹膜和物理攻击样本的基础上,显著提高了合成虹膜的检测准确率。特别地,本文方法仅通过改进模型,将低质量生成虹膜图像的检测错误率从95.55%大幅降低为38.60%,对中、高质量生成虹膜的检测错误率亦有所下降。此外,在保持BPCER较低的前提下,本文方法相比冠军方法,对于打印虹膜的检测准确率也有所提升。这可能是因为打印虹膜样本在频域与真实样本有显著差异,而本文方法充分利用了这一信息。总的来看,相比于Winner-

LivDet-Iris-2023方法,本文方法在综合检测错误率指标ACER1和ACER2上,分别下降了9.32%和3.71%,其中合成虹膜的平均检测错误率指标APCER下降22.78%,打印虹膜的检测错误率指标APCER下降6.32%。

3.4.3 与其他现有方法相比

与结合频域信息的人脸PAD方法LMFD-PAD(Fang等,2022)相比,本文方法显式地输入频域信息,大幅提升了模型检测合成虹膜的准确率,综合检测结果也得到了显著提升,这说明人脸和虹膜图像存在差异,人脸PAD模型不适合直接用于虹膜PAD。与其他没有使用频域信息的方法相比,结合注意力机制的DFCANet(Jaswal等,2024)几乎无法识别合成虹膜,检测打印虹膜的准确率也显著低于本文方法。而D-netPAD(Sharma和Ross,2020)的准确性则接近随机猜测,无法有效分类活体样本和攻击样本。总的来看,相比于先进的现有方法LMFD-PAD,本文方法在综合检测错误率指标ACER1和ACER2上,分别下降了8.91%和3.58%。

因此,从综合角度看,本文方法在两种评价指标ACER1和ACER2上都取得了优于所有现有方法的精度,充分说明本文方法能较好地检测出各种类型的呈现攻击,且具有更好的准确性和泛化性。

3.5 消融实验

为了验证本文所提各个模块的必要性,在LivDet-Iris 2023比赛协议下,从定量和定性两个角度进行消融实验,并对结果进行分析。

表3 在LivDet-Iris 2023比赛测试协议上的检测性能对比
Table 3 Performance comparison under the LivDet-Iris 2023 competition protocol

											/%
方法	APCER ↓								BPCER ↓	ACER ↓	
	PE	F/P	CL	ED	Syn(LQ)	Syn(MQ)	Syn(HQ)	Syn(平均)		ACER1	ACER2
Winner-LivDet-Iris 2023 (Tinsley等,2023)	22.80	1.50	0.00	0.00	95.55	94.40	94.85	94.93	0.14	42.65	22.15
D-netPAD(raw) (Sharma和Ross,2020)	45.40	50.75	75.00	52.50	44.50	45.35	45.35	45.07	55.92	50.65	53.59
LMFD-PAD(Fang等,2022)	13.98	2.63	0.00	5.00	90.35	95.60	96.40	94.12	0.62	42.24	22.02
DFCANet(Jaswal等,2024)	22.61	6.77	0.00	22.50	94.45	98.20	98.15	96.93	0.37	43.81	24.66
本文	16.48	1.88	0.00	10.00	38.60	88.45	89.40	72.15	1.91	33.33	18.44

注:加粗字体表示ACER1和ACER2两列中的最优结果。“↓”表示值越低越好。

3.5.1 定量角度

在这一部分实验中,替换或移除本文模型中的部分模块进行训练和测试,观察模型分类能力的变化。实验设置及结果详见表4。实验时,除“消融模块”一栏描述改动外,其余模型结构、超参数设置等均与原模型相同。其中, EfficientNet(224)和 EfficientNet(384)分别表示采用输入尺寸为 224×224 像素和 384×384 像素的原始图像训练骨干网络; EfficientNet(LBP)表示采用 LBP 增强后的图像训练单流骨干网络,即只使用本文模型中的空间域特征提取分支; EfficientNet(DCT)表示将 LBP 增强后的图像通过 DCT 变换到频域后训练骨干网络,即只使用本文模型中的频域特征提取分支。分辨率设为 224 指输入大小设为 224×224 像素;使用 MFD 模块表示将 DCT 变换替换为 LMFD-PAD(Fang 等,2022)中的 MFD 模块;-SCFA 模块指将双流模型提取的中层和高层特征直接拼接并送入分类器。

此外,为了验证所提思路在 Transformer 结构上

的有效性,测试了直接使用 Deit3 模型(Touvron 等,2022)进行二分类和使用 Deit3 作为骨干网络时的分类效果。实验中使用 Deit3_small,输入图像尺寸仍为 384×384 像素,划分图像块尺寸为 16×16 像素,其他超参数设置保持不变。

从表4中的结果可以得到以下结论:1)仅使用空间域信息时,输入大小为 224×224 像素的 EfficientNet 几乎完全无法检测出合成虹膜。尽管将输入大小扩大到 384×384 像素改善了这一现象,但却导致模型在检测打印虹膜时准确率下降,无法彼此兼顾。此外,使用 LBP 增强可以一定程度上改善骨干网络无法有效检测合成虹膜的问题,但效果有限。2)仅使用频域信息时,虽然模型检测合成虹膜时准确性有所提升,但却无法有效检测出屏显和纹理隐形眼镜等物理攻击样本。这一结果说明融合空间域和频域特征有效提升了模型检测不同种类攻击样本时的泛化性,二者缺一不可。3)同时使用空间域和频域信息时,加入 LBP 增强有助于提升模型的总体

表 4 在 LivDet-Iris 2023 比赛测试协议上的消融实验结果

Table 4 Results of the ablation studies under the LivDet-Iris 2023 competition protocol

消融模块	APCER ↓					BPCER ↓	ACER ↓	
	PE	F/P	CL	ED	Syn(平均)		ACER1	ACER2
△ EfficientNet(224)	17.05	0.75	0.00	5.00	95.93	0.71	43.16	22.54
△ EfficientNet(384)	23.95	2.63	0.00	17.5	78.22	0.88	35.80	20.35
△ EfficientNet(LBP)	22.80	1.88	0.00	10.00	85.90	0.85	39.08	21.31
△ EfficientNet(DCT)	20.50	24.81	75.00	52.50	76.68	3.22	36.72	30.38
*分辨率设为 224	38.51	5.26	0.00	22.50	86.32	0.98	40.30	23.72
*使用 MFD 模块	13.41	0.38	0.00	5.00	95.97	0.54	42.94	22.18
*三元组损失权重 $\lambda = 0.1$	14.37	3.01	0.00	5.00	81.90	1.62	37.39	19.96
*三元组损失权重 $\lambda = 10.0$	38.70	5.64	0.00	25.00	84.77	0.94	39.62	23.71
-三元组损失	22.03	1.88	0.00	0.00	77.95	3.82	37.01	20.32
-LBP 增强	10.50	0.38	0.00	5.00	94.93	1.35	43.06	22.49
-SCFA 模块	19.16	3.01	0.00	15.00	80.75	0.85	36.72	20.38
-LSGA 模块	20.50	4.51	0.00	10.00	79.37	0.89	36.20	19.95
△ Deit3_small	9.39	0.00	0.00	17.50	93.97	1.57	42.46	22.84
*使用 Deit3 作为骨干网络	22.41	24.81	75.00	5.00	73.60	2.57	35.11	29.36
本文	16.48	1.88	0.00	10.00	72.15	1.91	33.33	18.44

注:加粗字体分别代表 ACER1 和 ACER2 两列中的最优结果。“△”表示仅使用单流的骨干网络用于特征提取,“*”表示替换本文模型中的部分设置,“-”表示移除本文模型中的部分模块。“↓”表示值越低越好。

检测准确性,特别是有助于准确检测合成虹膜。但直接加入频域信息而不进行任何特征融合(即移除SCFA模块)反而会对模型产生负面影响,印证了空间域和频域图像差异性较大,不能简单拼接处理,表明了SCFA模块的有效性和必要性。4)使用MFD模块对图像中的不同频域分量提取特征也无法取得很好效果,表明了显式使用频域图像的必要性。5)加入LSGA模块后模型在检测各种攻击样本时准确率均有所提升,表明使用图像中的低频语义信息引导模型关注虹膜区域有助于提升模型性能。6)在使用BCE损失的基础上,加入三元组损失可以有效加大真实样本和攻击样本在特征空间中的差异,使模型提取到更有区分度的特征,从而提高分类准确性。此外,应合理选取总损失中二者的比重:三元组损失权重过小时难以起到辅助训练的作用;权重过大时则可能因为其优化目标与BCE损失不同反而对模型训练有害。7)对于Transformer结构,直接使用Deit3模型对空间域图像提取特征同样存在难以准

确分类合成虹膜的问题。如果将Deit3作为骨干网络替换本文所提模型中的EfficientNet,模型虽然在检测合成虹膜时的准确率有小幅提升,但在检测其他物理攻击时的准确率则大幅下降,导致总体模型性能下降。这可能是因为Transformer结构更适合提取图像中的全局特征和图像不同部分之间的关系,而虹膜PAD主要依赖图像中的细节纹理,使得Transformer结构难以在这一问题上发挥其优势。

3.5.2 定性角度

在这一部分中,选取4个有代表性的样本进行可视化分析,如图7所示。其中,图像左下角数字代表模型预测分数,小于0.5判断为攻击样本,大于0.5判断为真实样本。数字的绿色底色代表模型分类正确,红色底色代表分类错误。为了便于观察频域图像的差异,图中同时给出输入图像经过LBP处理后的DCT和FFT图像。

根据图7进行如下分析:1)由前3列可以看出,仅使用空间域特征(第2列)或频域特征(第3列)无

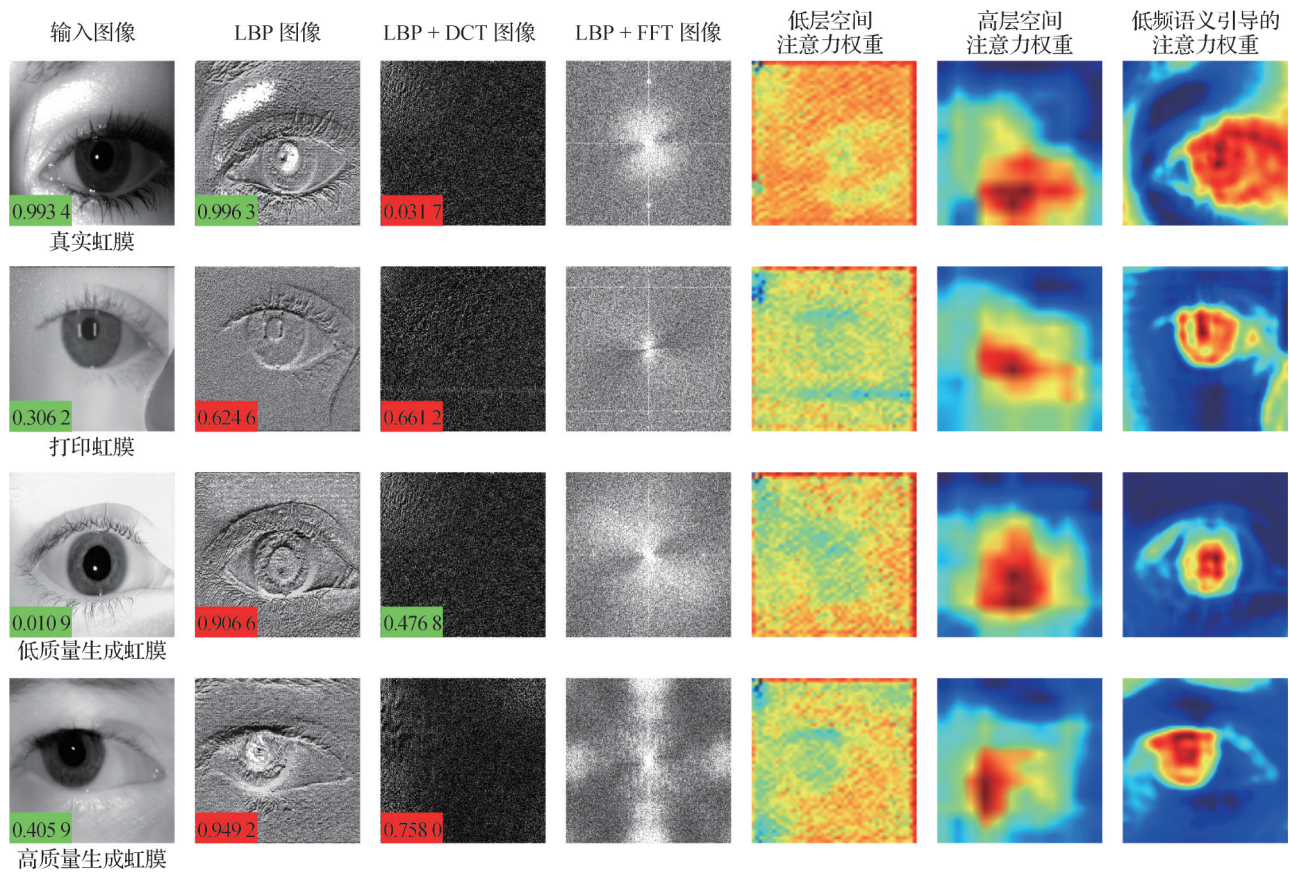


图7 使用融合特征、仅空间域特征和仅频域特征的分类结果及使用融合特征时的注意力权重的可视化示意图
Fig. 7 Illustration of classification results utilizing fused features , only spatial features , only frequency features , and the corresponding attention weights when using fused features , respectively

法保证模型能正确地进行虹膜真假分类,而使用空间域和频域融合的特征(第1列)则有效避免了错误的分类结果,从而验证了多域特征融合的有效性。2)第5列—7列展示了模型中注意力层的输出结果,其中深蓝色代表数值接近0,红色代表数值接近1。第5、6列分别展示了模型中低层和高层特征对应的两个SCFA模块中空间注意力分支输出的注意力权重图。在低层注意力中,注意力权重对一部分区域进行了抑制,对其他区域则进行了增强。被抑制的区域往往出现在图像左上角,即频域图中较亮的低频部分,或睫毛、眼睛轮廓等空间域图像中的突出部分。推测空间注意力为这些区域产生了接近0的权重可能是为了避免这些区域过于突出,干扰模型提取频谱图的高频区域纹理等细节信息。3)高层注意力则可能是因为在此前已经进行了空间域和频域的信息融合,两个分支中互相包含了对方的信息,导致注意力集中的位置与空间域图像不再有明确的对应关系。4)LSGA模块输出的权重图能使模型更关注虹膜及眼部等与PAD相关的区域,同时对无关区域赋予接近0的注意力权重,避免模型过拟合到这些区域。

4 结 论

针对现有虹膜呈现攻击检测模型泛化性不足、合成虹膜检测准确率较低的问题,本文提出一种融合频域和空间域特征的双流虹膜PAD模型。为了凸显虹膜图像中的纹理细节,本文方法首先将输入图像进行LBP增强,并使用DCT变换到频域。然后使用轻量级的EfficientNet_b0分别对空间域和频域图像进行特征提取,并引入空间通道特征融合注意力模块进行特征融合。为了让模型更好地关注虹膜区域,本文还引入了低频语义信息引导的注意力模块。在最新的虹膜呈现攻击检测比赛LivDet-Iris 2023数据集上的实验结果表明,本文方法的检测准确性和泛化性均超过了已有方法,且在检测低质量生成虹膜图像上具有一定的优势。这些结果证明了引入频域信息并结合深度学习、特征融合的方法有助于提升虹膜PAD模型的性能。

然而,本文方法仍存在局限性。如在检测中、高质量生成虹膜图像时的准确性仍然有较大的提升空间、双流EfficientNet_b0网络结构参数量仍存在压缩

的空间。此外,现有合成虹膜的种类和数量较少,也影响了虹膜PAD模型的开发。因此,未来的主要工作将聚焦于寻找更高效、更轻量化的特征提取骨干网络及特征融合方法,进一步提升虹膜PAD模型检测合成虹膜的精度,并使用更多方法生成虹膜样本,丰富训练数据。

参考文献(References)

- Agarwal A, Noore A, Vatsa M and Singh R. 2022. Generalized contact lens iris presentation attack detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3): 373-385 [DOI: 10.1109/TBIOM.2022.3177669]
- Chen C J and Ross A. 2021. An explainable attention-guided iris presentation attack detector//*Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*. Waikoloa, USA: IEEE: 97-106 [DOI: 10.1109/WACVW52041.2021.00015]
- Cherti M, Beaumont R, Wightman R, Wortsman M, Ilharco G, Gordon C, et al. 2023. Reproducible scaling laws for contrastive language-image learning//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: IEEE: 2818-2829 [DOI: 10.1109/CVPR52729.2023.00276]
- Choudhary M, Tiwari V and Venkanna U. 2022. Iris liveness detection using fusion of domain-specific multiple BSIF and DenseNet features. *IEEE Transactions on Cybernetics*, 52(4): 2370-2381 [DOI: 10.1109/TCYB.2020.3005089]
- Czajka A, Fang Z Y and Bowyer K. 2019. Iris presentation attack detection based on photometric stereo features//*Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, USA: IEEE: 877-885 [DOI: 10.1109/WACV.2019.00098]
- Doyle J S, Bowyer K W and Flynn P J. 2013. Variation in accuracy of textured contact lens detection based on sensor and lens pattern//*Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems*. Arlington, USA: IEEE: 1-7 [DOI: 10.1109/BTAS.2013.6712745]
- Fang M L, Damer N, Kirchbuchner F and Kuijper A. 2022. Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection//*Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, USA: IEEE: 1131-1140 [DOI: 10.1109/WACV51458.2022.00120]
- Fang Z Y, Czajka A and Bowyer K W. 2021. Robust iris presentation attack detection fusing 2D and 3D information. *IEEE Transactions on Information Forensics and Security*, 16: 510-520 [DOI: 10.

- 1109/TIFS.2020.3015547]
- Gragnaniello D, Poggi G, Sansone C and Verdoliva L. 2016. Using iris and sclera for detection and classification of contact lenses. *Pattern Recognition Letters*, 82: 251-257 [DOI: 10.1016/j.patrec.2015.10.009]
- Hoffman S, Sharma R and Ross A. 2019. Iris + ocular: generalized iris presentation attack detection using multiple convolutional neural networks//*Proceedings of 2019 International Conference on Biometrics (ICB)*. Crete, Greece: IEEE: 1-8 [DOI: 10.1109/ICB45273.2019.8987261]
- Jaswal G, Verma A, Roy S D and Ramachandra R. 2024. Learning joint local-global iris representations via spatial calibration for generalized presentation attack detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 6(2): 195-208 [DOI: 10.1109/TBIOM.2024.3355136]
- Kannala J and Rahtu E. 2012. BSIF: binarized statistical image features//*Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*. Tsukuba, Japan: IEEE: 1363-1366
- Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J and Aila T. 2020. Training generative adversarial networks with limited data//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 12104-12114
- Karras T, Aittala M, Laine S, Härkönen E, Hellsten J, Lehtinen J, et al. 2021. Alias-free generative adversarial networks//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Virtual Event: Curran Associates Inc.: 852-863
- Khade S, Gite S, Thepade S D, Pradhan B and Alamri A. 2023. Iris liveness detection using fragmental energy of haar transformed iris images using ensemble of machine learning classifiers. *Computer Modeling in Engineering and Sciences*, 136(1): 323-345 [DOI: 10.32604/cmescs.2023.023674]
- Kohli N, Yadav D, Vatsa M and Singh R. 2013. Revisiting iris recognition with color cosmetic contact lenses//*Proceedings of 2013 International Conference on Biometrics (ICB)*. Madrid, Spain: IEEE: 1-7 [DOI: 10.1109/ICB.2013.6613021]
- Kohli N, Yadav D, Vatsa M, Singh R and Noore A. 2016. Detecting medley of iris spoofing attacks using desist//*Proceedings of the 8th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*. Niagara Falls, USA: IEEE: 1-6 [DOI: 10.1109/BTAS.2016.7791168]
- Kuehlkamp A, Pinto A, Rocha A, Bowyer K W and Czajka A. 2019. Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 14(6): 1419-1431 [DOI: 10.1109/TIFS.2018.2878542]
- Li P P, Wang C L and Sun Z Q. 2021. Face liveness detection based on multi-feature fusion with an attention mechanism. *Information and Control*, 50(5): 631-640 (李盼盼, 王朝立, 孙占全. 2021. 基于注意力机制的多特征融合人脸活体检测. *信息与控制*, 50(5): 631-640) [DOI: 10.13976/j.cnki.xk.2021.0581]
- Li Y C, Lian Y, Wang J J, Chen Y H, Wang C M and Pu S L. 2022. Few-shot one-class domain adaptation based on frequency for iris presentation attack detection//*Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore, Singapore: IEEE: 2480-2484 [DOI: 10.1109/ICASSP43922.2022.9746635]
- Liu M K, Wang H M, Li Q and Sun Z N. 2020. Enhanced gray-level image space for iris liveness detection. *Journal of Image and Graphics*, 25(7): 1421-1435 (刘明康, 王宏民, 李琦, 孙哲南. 2020. 增强型灰度图像空间实现虹膜活体检测. *中国图象图形学报*, 25(7): 1421-1435) [DOI: 10.11834/jig.190503]
- Lyu M L, He Y Q, Yang J K, Jin W Q and Zhang L J. 2022. Anti-spoofing detection method for contact lens irises based on recurrent attention mechanism. *Acta Optica Sinica*, 42(23): #2315001 (吕梦凌, 何玉青, 杨峻凯, 金伟其, 张丽君. 2022. 基于循环注意力机制的隐形眼镜虹膜防伪检测方法. *光学学报*, 42(23): #2315001) [DOI: 10.3788/AOS202242.2315001]
- Maureira J, Tapia J E, Arellano C and Busch C. 2022. Analysis of the synthetic periocular iris images for robust presentation attacks detection algorithms. *IET Biometrics*, 11(4): 343-354 [DOI: 10.1049/bme2.12084]
- Ojala T, Pietikainen M and Harwood D. 1994. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions//*Proceedings of the 12th International Conference on Pattern Recognition (ICPR)*. Jerusalem, Israel: IEEE: 582-585 [DOI: 10.1109/ICPR.1994.576366]
- Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. 2024. DINOv2: learning robust visual features without supervision [EB/OL]. [2024-02-02]. <https://arxiv.org/pdf/2304.07193.pdf>
- Sharma R and Ross A. 2020. D-NetPAD: an explainable and interpretable iris presentation attack detector//*Proceedings of 2020 IEEE International Joint Conference on Biometrics (IJCB)*. Houston, USA: IEEE: 1-10 [DOI: 10.1109/IJCB48548.2020.9304880]
- Silva P, Luz E, Baeta R, Pedrini H, Falcao A X and Menotti D. 2015. An approach to iris contact lens detection based on deep image representations//*Proceedings of the 28th SIBGRAPI Conference on Graphics, Patterns and Images*. Salvador, Brazil: IEEE: 157-164 [DOI: 10.1109/SIBGRAPI.2015.16]
- Sun Z N, Zhang H, Tan T N and Wang J Y. 2014. Iris image classification based on hierarchical visual codebook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6): 1120-1133 [DOI: 10.1109/TPAMI.2013.234]
- Tan C C, Zhao Y, Wei S K, Gu G H, Liu P and Wei Y C. 2024. Frequency-aware deepfake detection: improving generalizability through frequency space domain learning//*Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*. Vancouver,

- Canada: AAAI: 5052-5060 [DOI: 10.1609/aaai.v38i5.28310]
- Tan M X and Le Q V. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR: 6105-6114
- Tapia J E, Gonzalez S and Busch C. 2022. Iris liveness detection using a cascade of dedicated deep learning networks. IEEE Transactions on Information Forensics and Security, 17: 42-52 [DOI: 10.1109/TIFS.2021.3132582]
- Tapia J E, González-Soler L J and Busch C. 2025. Towards iris presentation attack detection with foundation models [EB/OL]. [2025-03-25]. <https://arxiv.org/pdf/2501.06312.pdf>
- Tian C, Luo Z M, Shi G M and Li S Z. 2023. Frequency-aware attentional feature fusion for deepfake detection//Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10094654]
- Tinsley P, Purnapatra S, Mitcheff M, Boyd A, Crum C, Bowyer K, et al. 2023. Iris liveness detection competition (LivDet-Iris) — the 2023 edition//Proceedings of 2023 IEEE International Joint Conference on Biometrics (IJCB). Ljubljana, Slovenia: IEEE: 1-10 [DOI: 10.1109/IJCB57857.2023.10448637]
- Touvron H, Cord M and Jégou H. 2022. DeiT III: revenge of the ViT//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 516-533 [DOI: 10.1007/978-3-031-20053-3_30]
- Trokielewicz M, Czajka A and Maciejewicz P. 2015. Assessment of iris recognition reliability for eyes affected by ocular pathologies//Proceedings of the 7th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS). Arlington, USA: IEEE: 1-6 [DOI: 10.1109/BTAS.2015.7358747]
- Trokielewicz M, Czajka A and Maciejewicz P. 2016. Human iris recognition in post-mortem subjects: study and database//Proceedings of the 8th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS). Niagara Falls, USA: IEEE: 1-6 [DOI: 10.1109/BTAS.2016.7791175]
- Trokielewicz M, Czajka A and Maciejewicz P. 2018. Presentation attack detection for cadaver iris//Proceedings of the 9th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS). Redondo Beach, USA: IEEE: 1-10 [DOI: 10.1109/BTAS.2018.8698542]
- Verma P, Selwal A and Sharma D. 2023. A survey on data-driven iris spoof detectors: state-of-the-art, open issues and future perspectives. Multimedia Tools and Applications, 82(13): 19745-19792 [DOI: 10.1007/s11042-022-14014-4]
- Wang C, He Z F, Wang C Y and Tian Q. 2022. Generating intra- and inter-class iris images by identity contrast//Proceedings of 2022 IEEE International Joint Conference on Biometrics (IJCB). Abu Dhabi, United Arab Emirates: IEEE: 1-7 [DOI: 10.1109/IJCB54206.2022.10007974]
- Wang C Y, Liu X Y, Fang M L, Zhao G Z, He Z F and Sun Z N. 2024. A survey on iris presentation attack detection. Acta Automatica Sinica, 50(2): 241-281 (王财勇, 刘星雨, 房美玲, 赵光哲, 何召锋, 孙哲南. 2024. 虹膜呈现攻击检测综述. 自动化学报, 50(2): 241-281) [DOI: 10.16383/j.aas.c230109]
- Wei Z S, Tan T N and Sun Z N. 2008. Synthesis of large realistic iris databases using patch-based sampling//Proceedings of the 19th International Conference on Pattern Recognition (ICPR). Tampa, USA: IEEE: 1-4 [DOI: 10.1109/ICPR.2008.4761674]
- Yadav D, Kohli N, Vatsa M, Singh R and Noore A. 2019. Detecting textured contact lens in uncontrolled environment using DensePAD//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, USA: IEEE: 2336-2344 [DOI: 10.1109/CVPRW.2019.00287]
- Yadav S and Ross A. 2023. iWarpGAN: disentangling identity and style to generate synthetic iris images//Proceedings of 2023 IEEE International Joint Conference on Biometrics (IJCB). Ljubljana, Slovenia: IEEE: 1-10 [DOI: 10.1109/IJCB57857.2023.10449250]
- Yambay D, Becker B, Kohli N, Yadav D, Czajka A, Bowyer K W, et al. 2017. LivDet iris 2017—Iris liveness detection competition 2017//Proceedings of 2017 IEEE International Joint Conference on Biometrics (IJCB). Denver, USA: IEEE: 733-741 [DOI: 10.1109/BTAS.2017.8272763]

作者简介

王财勇,男,副教授,主要研究方向为生物特征识别、计算机视觉与模式识别。E-mail: wangcaiyong@bucea.edu.cn

孙娴蕴,女,硕士研究生,主要研究方向为生物特征识别和计算机视觉。E-mail: sunxianyun@stu.bucea.edu.cn

李林,女,硕士研究生,主要研究方向为生物特征识别和计算机视觉。E-mail: lilin@stu.bucea.edu.cn

赵光哲,男,教授,主要研究方向为计算机视觉与图像处理、模式识别和人工智能。E-mail: zhaoguangzhe@bucea.edu.cn

何召锋,男,教授,主要研究方向为生物特征识别、视觉计算、智能博弈决策和AI+IC协同优化。

E-mail: zhaofenghe@bupt.edu.cn

孙哲南,男,研究员,主要研究方向为生物特征识别、模式识别和计算机视觉。E-mail: znsun@nlpr.ia.ac.cn