# 引入全局语义增强的人脸欺诈特征提取研究①

蔡体健, 陈 均, 罗词勇, 刘遵雄, 陈子涵

华东交通大学 信息工程学院 南昌 中国 330013

**摘要** 基于人脸反欺诈的领域知识, 针对人脸活体检测中特征在网络中逐层稀释的问题, 该文提出了基于语义增强和交叉注意力优化的人脸活体检测模型。具体来说, 首先利用活体样本无欺诈噪声的先验知识, 采用活体人脸的半边约束方法来提取欺诈强相关特征; 利用欺诈特征的全局移不变性特点, 结合深度度量学习技术以及异常检测等方法, 该文在 U-Net 瓶颈层添加语义增强模块来增强欺诈特征, 捕获长距离的移不变性特征, 同时对比了三个不同的语言增强模块在模型上的性能, 然后在编码块和解码块之间的跳跃连接后引入交叉自注意力模块, 以进一步增强全局的欺诈信息和重要区域的关注。此外, 该文将 U-Net 模型的解码块中的传统卷积算子替换为中心差分卷积算子, 以提取细粒度的欺诈特征, 并通过计算中心像素与周围像素之间的差异, 去除光照、环境的影响, 以此提高模型的鲁棒性能。经过在四个常用的人脸活体检测数据集 CASIA-MFSD、MSU-MFSD、OULU-NPU、Replay-Attack 上测试与评估, 进行了数据集内实验、跨数据集实验和消融实验等, 对模型进行了复杂度分析以及对部分实验进行了可视化分析, 该文模型能够有效降低人脸分类的错误率。

**关键词** 人脸活体检测; 全局语义增强; 交叉注意力; 中心差分卷积; 深度度量学习

**中图法分类号** TP391.41 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2025.03.09

# Research on Introducing Global Semantic Enhancement for Face Fraud Feature Extraction

CAI Tijian, CHEN Jun, LUO Ciyong, LIU Zunxiong,CHEN Zihan

School of Information Engineering, East China Jiao Tong University, Nanchang 330013, China

**Abstract** Based on the domain knowledge of face anti-fraud, this paper proposes a face liveness detection model based on semantic enhancement and cross-attention optimization to address the problem of feature dilution layer by layer in the network during face liveness detection. Specifically, we first use the prior knowledge that living samples are free of fraud noise, and use the half-edge constraint method of living faces to extract strong correlation features of fraud; we also use the global shift invariance characteristics of fraud features, combined with deep metric learning technology and anomaly detection, etc. Method, this article adds a semantic enhancement module to the U-Net bottleneck layer to enhance fraud features and capture long-distance shift invariance features. At the same time, it compares the performance of three different language enhancement modules on the model, and then in the encoding block and decoding A cross-self-attention module is introduced after the skip connection between blocks to further enhance the global fraud information and focus on important areas. In addition, this paper replaces the traditional convolution operator in the decoding block of the U-Net model with a central difference convolution operator to extract fine-grained fraud features and remove them by calculating the difference between the central pixel and the surrounding pixels. The influence of lighting and environment is used to improve the robust performance of the model. After testing and evaluation on four commonly used face liveness detection data sets CASIA-MFSD, MSU-MFSD, OULU-NPU, and Replay-Attack, intra-dataset experiments, cross-dataset experiments, and ablation experiments were conducted to verify the model. Complexity analysis and visual analysis of some experiments were conducted. The model in this article can effectively reduce the error rate of face classification.

**Key words** face liveness detection; global semantic enhancement; cross-attention; central difference convolution; deep metric learning

## 1 引言

低碳数字经济中, 刷脸支付、刷脸办证、刷脸解锁、刷脸看病等业务的应用需求在不断增加。然而, 针对人脸识别系统的恶意攻击[1-2]也越加频繁, 如图 1 所示, 这些攻击主要包括重放攻击(打印、视频)、3D 面具攻击和 AI 换脸等, 这给人脸识别系统带来了极大的威胁。与其他二元视觉任务不同, 人脸活体检

**通讯作者:** 陈均, 硕士, Email: 1318141847@qq.com。

本课题得到国家自然科学研究基金(No. 62162026 ), 江西省自然科学基金资助项目(No. 20232BAB202055, No. 20242BAB25114)资助。
收稿日期: 2023-07-11; 修改日期: 2023-09-23; 定稿日期: 2025-01-10

测(Face Anti-Spoofing, FAS)是一个自我演变的问题(即攻击与防御迭代发展), 这使得它更具挑战性。此外, 其他二分类视觉任务[3](如人类性别分类)高度依赖明显的基于外观的语义线索(如发型、穿着、脸型); 而 FAS 中的内在特征(如材料和几何结构)通常与内容无关(如与面部属性和身份无关), 因此, FAS 被视为一个结构材料的识别问题, 不同的材质具有微妙且细微的差别, 即使人眼也很难分辨, 然而材质不会随空间位置的改变而改变, 所以材质特征具有全局移不变特点。此外, 传统神经网络所挖掘的欺诈特征实际上是多种“强相关-弱相关-不相关”特征的“纠缠体”, 其中包含不可靠的欺诈线索, 例如光照和人脸结构信息等影响因素, 因此有必要将强相关的欺诈特征从“纠缠体”中分离出来。

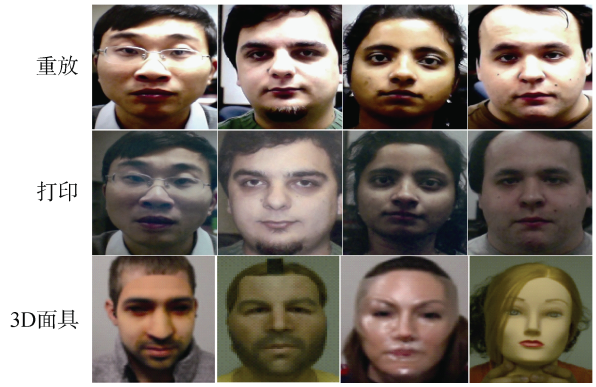


**图 1 人脸欺诈类型示例**

**Figure 1 Examples of face spoofing types**

近年来, 随着深度学习技术的盛行, 一系列的卷积神经网络变体被应用于人脸活体检测并取得了突破性进展。早期的全卷积神经网络[4-7]用卷积层取代了神经网络的全连接层。神经网络为了扩大神经元的感受野, 逐渐缩小特征图的尺度, 最终生成分辨率很低的预测, 导致分类正确率降低。Yang 等人[8]首次提出使用 8 层浅层 CNN 进行特征表示的端到端深度 FAS 方法。Yu 等人[9]提出的中心差分卷积网络能够很好的提取伪图像的特征, 且不易受光照影响。后来, 有研究人员提出基于 U-Net[10]的网络结构, 它采用逐步上采样学习来恢复特征图分辨率, 同时保持神经元相对较大的感受野。同时利用跳跃连接增强浅层和深层特征的融合, 对原有的全连接层进行改进, 有效提高了最终的分类性能。Jourabloo 等人[11]将 FAS 重新表述为一个欺诈噪音建模问题, 并设计了一个编码器-解码器架构, 用像素级监督(例如, 活体人脸的零噪声图)来估计潜在的欺诈模式。有了这种对活体人脸的单边约束, 这些模型就能灵活地挖掘欺诈攻击的欺骗线索。Feng 等人[12]设计了一个欺诈线索生成器, 通过最小化活体样本的欺诈线索, 同时对欺诈样本的欺诈线索不施加显性约束, 以此有效提取欺诈特征。

但是由于卷积操作的固有局部性, 基于 CNN 的方法很难学习明确的全局语义信息和长距离的语义信息交互[13]。一些研究试图通过使用反卷积层[14-15]、自注意机制[16-17]和图像金字塔[18]来解决这个问题。然而, 这些方法在模拟长距离的依赖性方面仍有局限性。最近, Liu 等人[19]提出上下文感知网络 FECANet, 设计了特征增强模块抑制图片噪声和相关重建模块来编码多尺度的全局语义特征; Jiang 等人[20]提出了一种基于特征金字塔结构的多尺度特征嵌入方法, 旨在将高级语义特征与低级丰富的视觉特征相结合; Jia 等人[21]提出一个端到端的单边域泛化框架来提高人脸反欺诈的泛化能力, 有效提高了 FAS 的泛化能力。

综上所述, 传统的人脸活体检测方法和大部分使用 CNN 的方法主要基于局部特征, 这些方法很容易受到攻击者使用高质量照片或视频进行欺诈; 同时随着 3D 打印面具以及合成人脸的出现, 目前许多人脸活体检测方法由于缺乏对全局语义信息[22]的理解, 可能无法准确检测这些新型攻击。因此, 本文设计网络的目的是增强网络中逐层稀释的特征, 利用相应模块提取全局语义特征和局部细节信息, 结合深度度量学习技术完成对活体和欺诈人脸的分类。具体来说, 本文网络模型是基于编码器-解码器的 U-Net 框架, 在编码器中, 一系列卷积层和连续下采样层与语义增强模块相结合可提取具有较大感受野的深度特征。然后, 解码器结合中心差分卷积将提取的深层特征上采样到输入分辨率, 用于像素级的语义预测, 来自编码器的不同尺度的高分辨率特征与交叉注意力融合, 以减轻下采样造成的空间信息损失。

## 2 方法

本文所提出的模型基于 U-net 框架, 在增强全局语义信息和提高系统鲁棒性方面进行了优化。首先在 U-net 的瓶颈层增加了全局语义增强模块(Global Context Module, GCM), 以此来捕获长距离的全局移不变的欺诈特征; 其次, 用多头交叉注意力模块(Mutil-Head Cross Attention, MHCA)来取代跳连接, 通过交叉注意力将编码块特征中非关注的信息给过滤掉, 并增强重要区域的关注; 再者, 用中心差分卷积(Central Difference Convolution, CDC)取代解码层的一般卷积, 以提高模型的鲁棒性。本文通过增加

GCM,MHCA 模块块引入人脸的全局语义信息, 可以捕获全局脸部欺诈信息, 相比于局部特征,全局特征有利于更好地区分真实人脸与欺诈攻击; 利用全局信息可以在一定程度抑制局部细节和变化对模型判断的影响(例如光照, 环境等), 以此提高模型对图像变化的鲁棒性。同时利用全局语义可以加强模型对负样本进行有效学习的能力, 而不局限于局部细节。总体来说, 全局语义信息的引入可从多个方面提高模型的鲁棒性和准确性。另外通过 CDC 计算中心像素与周围像素之间的差异, 可捕获到更多的细节信息, 通过局部特征的微小变化, 可进一步提高模型的鲁棒性; 在实验中通过计算多尺度损失提高模型对于不同人脸尺度的适应能力, 同时模型可更好地泛化到不同场景。因此结合以上模块和技术, 本模型可有效提高检测性能和鲁棒性。

## 2.1 网络框架

本文的网络总架构图如图 2 所示。

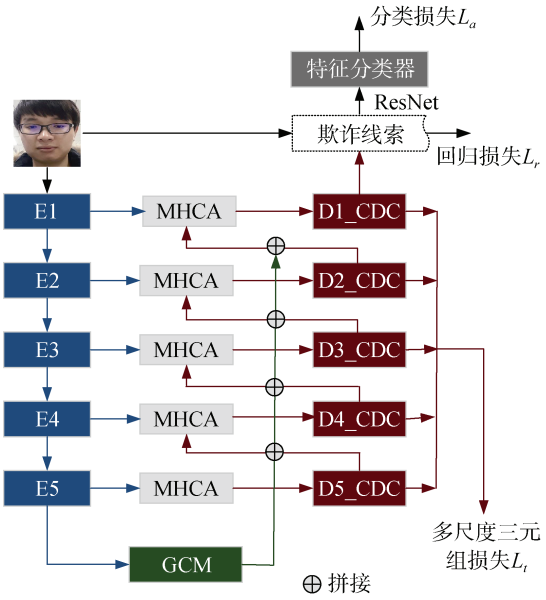


**图 2 模型总框架**

**Figure 2 The overall framework of the model**

本文使用 ResNet18[23]作为网络的编码器, 它总共包含五个编码块, 对应的解码器也有五个解码块, 每个解码块由两个Conv-BN-ReLu组成, 本文将解码块中的传统卷积替换成 CDC 算子, 用来提取更加高级的语义特征图, 全局语义增强模块位于编码器分支的顶部, 它捕获全局上下文信息并密集地连接到解码器路径中每一层的解码块。同时, 在 U-Net 的 encoder 和 decoder 路径之间的每一个跳跃连接之后引入 MHCA 模块, 给每个解码层的每个位置特征列一个不同感受野的局部上下文增强, 同时巧妙地利用前一层的预测置信度作为指导, 迫使当前层关注更难的区域。

## 2.2 语义增强模块

在编码器分支的顶部, 本文考虑加入一个语义增强模块, 以此来捕获欺诈的全局语义特征。本文比较了 GCM[24]、语义特征增强模块[25](Semantic Feature Enhancement Module, SFEM)、空洞空间金字塔池化[26](Atrous Spatial Pyramid Pooling, ASPP)在 Unet 网络上的性能(见 3.6), 根据 3.6 节表 3 的对比, 选择最佳表现的模块作为本文模型的语义增强模块。

(1) 全局语义模块

GCM 包含四个分支, 用于提取不同尺度的语境特征。具体来说, 该模块由一个全局平均池化分支、两个自适应局部平均池化分支组成, 并分别输出空间大小分别为 $1\times1, 3\times3, 5\times5$ 的三个特征图。它还包含一个具有 non-local 操作的身份映射分支, 以捕捉长距离的依赖性, 同时保持原始分辨率, non-local 可以捕获每个位置信息的全局依赖性来增强编码器的输出, 最后对得到的四个特征图进行上采样并拼接之后将本模块的全局语义特征送入每个解码器。结构如图 3 所示。

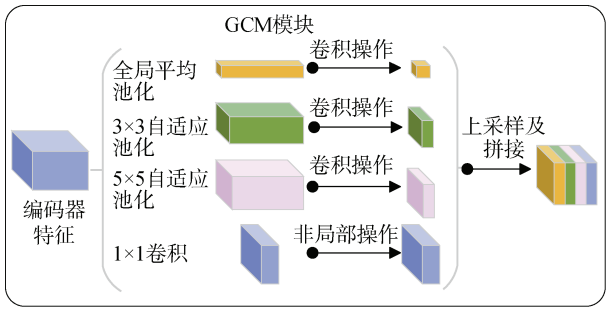


**图 3 全局语义模块**

**Figure 3 Global context module**

(2) 语义特征增强模块

SFEM 由三个平行的分片非局部块组成, 如图 4 所示。它将编码器特征图的输出作为输入, 并将非局部注意力分别应用于特定窗口大小的分片, 而不是应用自适应平均池。第一个分支将图像分成四个大小($\frac{H}{2}\times\frac{W}{2}$)的块, 分别在每个块上应用非局部空间注意力, 并将它们折叠在一起。类似地, 第二分支生成 16 个大小($\frac{H}{4}\times\frac{W}{4}$)的块, 并对每个分片执行与第一分支相同的操作。这三个分支的输出被连接起来, 然后是一个压缩和激励块(Squeeze-Excitation, SE), 用于计算各通道的特征图信息, 然后将 SE 块的结果

发送到所有解码器层。

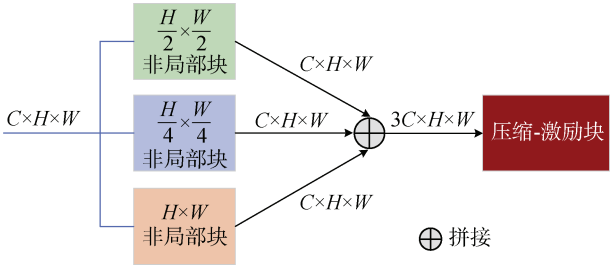


**图 4　语义特征增强模块**

**Fig. 4　Semantic feature enhancement module**

根据本文的编码器-解码器结构, 非局部块可以由以下公式表示:

$$y_i = \frac{1}{C(e_l)} \sum_{\forall j} f\left(e_{il}, e_{jl}\right) g\left(e_{jl}\right) \tag{1}$$

其中, $e_l$ 表示来自编码器的特征, $e_{il}$ 和 $e_{jl}$ 表示来自相同编码器层 $l$ 的特征。

(3) 空洞空间金字塔池化

ASPP 由空洞卷积和空间金字塔池化组成。如图 5 所示, 空间金字塔池化用来捕获多尺度的上下文信息, 结合空洞卷积可以通过修改空洞滤捕获长距离信息。如果想要对图片提取的特征具有较大的感受野, 并且又想让特征图的分辨率不下降太多(分辨率损失太多会丢失许多关于图像边界的细节信息), 这两个是矛盾的, 想要获取较大感受野需要用较大的卷积核或池化时采用较大的步长, 对于前者计算量太大, 后者会损失分辨率。而空洞卷积就是用来解决这个矛盾的。即可让其获得较大感受野, 又可让分辨率不损失太多。具有不同空洞卷积滤率的 ASPP 能够有效地捕获多尺度信息。

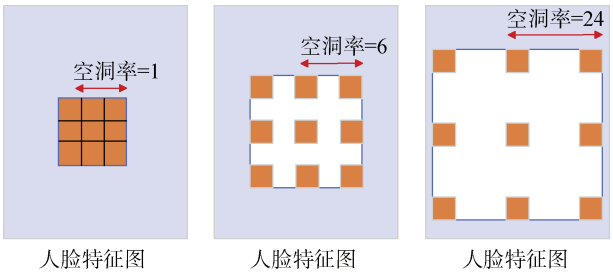


**图 5　空洞空间金字塔池化**

**Figure 5　Atrous spatial pyramid pooling**

空洞卷积最初是为了有效计算小波变换而提出的, 其公式如下:

$$y[i] = \sum_k x[i + rk] w[k] \tag{2}$$

其中, 输入特征图 $x$ 和滤波器 $w$ 的卷积生成输出 $y$, 而空洞卷积率 $r$ 对应于我们对输入信号进行采样的步长。它相当于将输入 $x$ 与上采样的滤波器进行卷积, 通过在每个空间维度的两个连续的滤波器值之间插入 $r$-1 个零来产生(因此被称为空洞卷积)。标准卷积是空洞率 $r$=1 的一个特例, 而空洞卷积允许通过改变 $r$ 值来适应性地修改滤波器的感受野。

综上所述, GCM 模块引入多尺度池化和非局部技术, 多尺度池化有助于模型提取不同尺度的特征, 而非局部技术有助于模型获取图像中的全局欺诈特征。SFEM 模块主要使用了非局部技术, 增强了模块提取全局欺诈特征的能力, 然而其提取多尺度特征能力相对较弱。ASPP 模块使用了空洞卷积和空间金字塔池化技术, 提高了模型捕获多尺度特征的能力, 然而提取全局欺诈特征的能力相对较弱。因此, GCM 模块获得了较好的综合性能。

## 2.3　基于多头交叉注意力的信息加强

多头交叉注意力模块[27](Mutil-Head cross attention, MHCA)的作用类似于一个门控函数, 输入分别是跳跃连接过来的编码器特征图 S 和上一层解码器的特征图 Y 处理后的结果。S 经过卷积和下采样得到 V(值矩阵), 头的数量则和特征图 S 的通道数有关, Y 嵌入后的结果作为 Q(查询矩阵), K(键矩阵), 最后跳跃连接的输出 S 是经过 Y 加权处理后的结果, 计算出来的注意力权重会被缩放到 0-1 之间, 最终交叉注意力的计算结果 Z 作为一个过滤器, 再与 S 做点积, 其中权值较小的元素代表噪声或者不相关的区域, 可以被去除。经过这样精简处理 S, 再将这一精简之后的结果 S 与 Y 做级联。MHCA 如图 6 所示:

其中位置编码通过不同频率的正弦和余弦函数计算得到, 可帮助 MHCA 建立长距离依赖以及为特征增加空间位置信息。最终的输出是一个位置编码矩阵, 其大小与输入张量的大小相同, 但通道维度是原始输入通道数。这样, 就可以将位置编码矩阵与输入张量相加或连接, 以引入位置信息。在 U-Net 的跳跃连接中引入 MHCA 可以将高水平特征图的语义丰富性与来自跳跃连接的高分辨率特征图结合起来, 旨在进一步去除特征的一些无关因素, 通过权重的处理让模型忽略特征中噪声等次要信息, 关注更加细粒度的重要信息。

## 2.4　中心差分卷积信息的提取

由于传统的卷积算子固有的局部性, 是导致人脸活体检测中特征细粒度不高的原因之一, 受到中心差分卷积网络的启发, 本文将特征解码模块中的每个解码块的传统卷积算子改进为中心差分卷积算子, 以此来恢复更多详细的人脸特征。

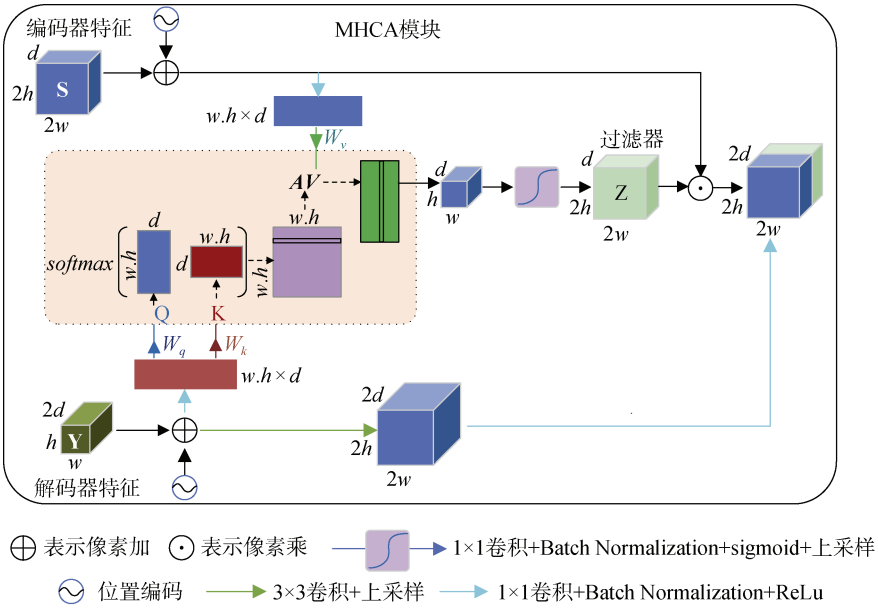


**图 6　多头交叉注意力**

**Figure 6　Muti-Head cross-attention**

因为卷积操作在通道维度上是一致的，为简单起见，以下卷积都用二维表示。传统卷积有两个主要操作，首先对输入特征图 $x$ 的局部感受野区域 $R$ 进行采样，然后通过加权求和对采样值进行聚合。其公式表示如下:

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n) \tag{3}$$

其中，$p_0$ 表示输入和输出特征图的当前位置，$p_n$ 枚举了感受野 $R$ 中特征图的位置。

根据局部二元模式(LBP)[28]的思想，中心差分卷积算子将中心差分操作引入传统卷积算子，以此增强特征表示能力。与传统卷积类似，中心差分卷积也包含两个操作，采样和聚合。采样与传统卷积类似，对于聚合步骤，如图 7 所示，中心差分卷积倾向于聚合采样值中心的中心方向梯度。

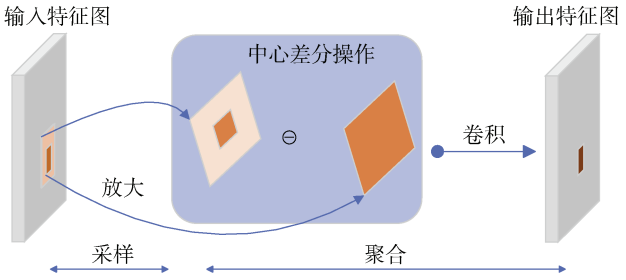


**图 7　中心差分卷积**

**Figure 7　Central difference convolution**

对于人脸反欺骗任务来说，强度级别的语义信息和梯度级别的细节信息对于区分活体和欺骗的人脸都是至关重要的，这表明将传统卷积与中心差分卷积相结合可以提供更强大的建模能力。因此，中心差分卷积公式如下:

$$y(p_0)=\underbrace{\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n)}_{\text{传统卷积}}+\underbrace{\theta\cdot -x(p0)\cdot\sum_{p_n\in R} w(p_n)}_{\text{中心差分卷积}} \tag{4}$$

其中，超参数 $\theta\in[0,1]$对强度级和梯度级信息的贡献进行权衡。$\theta$ 的值越大，意味着中心差分梯度信息越重要。

## 2.5　深度度量学习

深度度量学习(Deep Metric Learning, DML)研究如何在一个特定的任务(如基于异常检测的方式)上学习一个距离函数，使得该距离函数能够帮助这些任务取得较好的性能。本文引入多尺度度量学习技术，来获得更清晰的分类边界。

具体来说，首先提取 E5 层至 D1_CDC 层共 6 层的多尺度特征，使用全局平均池化(Global Average Pooling, GAP)将他们向量化，得到一组特征向量 $\{\boldsymbol{V}\}$，然后计算多尺度特征向量的三元组损失，其目标是促使锚样本和正样本之间的距离尽可能小，而锚样本与负样本之间的距离尽可能大。三元组损失函数的表达式如下:

$$L_t=\frac{1}{T}\sum_{i=1}^{T}\max\left(d(a_i,p_i)-d(a_i,n_i)+m,0\right) \tag{5}$$

$$d(i,j)=\left\|\frac{v_i}{\|v_i\|_2}-\frac{v_j}{\|v_j\|_2}\right\|_2 \tag{6}$$

其中, $T$ 是样本对个数, $a_i$ 代表锚点样本, $p_i$ 代表活体样本, $n_i$ 代表欺诈样本, $m$ 代表预定义的边界常数, $d(i,j)$ 表示两个归一化向量后的欧几里得距离。

### 2.6 回归损失

在 FAS 中, 虽然假定活体样本具有相同的性质, 但由于攻击媒介的多样性, 欺诈样本可能非常多样化。这种多样性使得欺诈样本很难在特征表示空间中形成一个紧凑的区域。受异常检测[29]方法的启发, 本文假设活体样本属于一个封闭集, 而欺诈样本属于一个开放集, 欺诈线索生成器采用无监督的方法学习欺诈线索, 即最小化活体样本的欺诈损失, 而不对欺诈样本做任何约束, 这样促使网络学习到更多的欺诈线索, 从而提高网络模型的泛化能力。

具体来说, 本文将RGB图像I作为输入, 欺诈线索生成器生成一个相同大小的欺诈线索映射 C, 由于活体样本不包含任何欺诈材质, 那么活体样本的欺诈线索映射 $C$ 应趋为零。因此, 欺诈线索生成器的一个优化目标是最小化欺诈线索映射 $C$, 其目标损失函数是像素级的回归损失 $L_r$ , 其公式如下:

$$L_r = \frac{1}{N_l}\sum_{I_i \in \text{live}} \left\| C_i \right\|_1 \tag{7}$$

其中, $N_l$ 是一个批次的活体样本数量, 欺诈样本并不参与该目标函数的优化。

### 2.7 总体损失和测试策略

本文使用特征分类器作为欺诈线索的辅助放大器, 有助于学习更多具有辨识度的欺诈特征。具体来说, 在欺诈线索 $C$ 生成后, 将 $C$ 与原图像 $I$ 叠加, 形成叠加图像 $S$, 将 $S$ 作为特征分类器的输入。根据下文实验所计算的平均分类错误率, 选择 $S$ 作为分类器的输入, 可以提高分类准确率, 并得到更具辨识度的欺诈线索。其分类损失是二元交叉熵损失:

$$L_a = \frac{1}{N}\sum_{i=1}^{N} z_i \log q_i + \left(1 - z_i\right)\log\left(1 - q_i\right) \tag{8}$$

其中, $N$ 为样本数, $z_i$ 为二进制标签, $q_i$ 为网络预测值。

本文所优化的 U-Net 网络损失分为三个部分: 活体样本中欺诈线索的像素级回归损失 $L_r$ 、活体样本和欺诈样本的三元组损失 $L_t$ 以及特征分类器的辅助分类损失 $L_a$。所以训练期间的总损失由 $L_r$ 、$L_t$ 和 $L_a$ 组成, 其总损失 $L$ 如下:

$$L = \alpha_1 L_r + \alpha_2 \sum_{k \in \{E1-D1\_\text{CDC}\}} L_t^k + \alpha_3 L_a \tag{9}$$

其中, $k$ 代表网络中应用三元组函数受惩罚的层, $\alpha_1$ 、$\alpha_2$ 、$\alpha_3$ 代表平衡不同损失函数的影响而分配的权重。

在测试阶段, 本文使用生成的欺诈线索映射作为评估依据而没有用分类器的输出。将欺诈得分定义为欺诈线索映射 $C$ 的均值, 欺诈得分计算方式如下:

$$score = \overline{\left\| C \right\|_1} \tag{10}$$

其中, 计算得出的分数 $score$ 实际上是测试样本为欺诈样本的概率, 值越大, 则样本为欺诈样本的可能性越高。

## 3 实验和结果分析

### 3.1 数据集

实验中主要使用四个数据集, 分别是 OULU-NPU[30]、CASIA-MFSD[31]、Replay-Attack[32]、MSU-MFSD[33]。OULU-NPU 数据集收集了高分辨率的人脸图片, 可以较好地模拟真实场景, 由 4950 个真实和攻击视频组成, 这些视频是用 6 台移动设备的前置摄像头录制的, 共有三种不同的光照条件和背景场景, 本文将其用于数据集内部测试, CASIA-MFSD、Replay-Attack、MSU-MFSD 数据集收集了大量低分辨率视频。本文将其用于数据集之间的交叉测试, 评估模型的泛化性能。

### 3.2 实验设置

实验的硬件环境为 NVDIA GeForce RTX 3080Ti 显卡, 编程语言为 Python3.7, 框架采用 Pytorch。实验前, 将数据集的 RGB 图片大小裁剪为 $224\times224\times3$, 然后随机选择数据, 以保证正负样本比例为 1:1, 最后进行数据增强[34]处理来减少过拟合的影响, 比如将图片水平和垂直翻转, 旋转和缩放。在训练阶段, batch size 设置为 32, 使用 Adam[35]优化器训练网络模型, 学习率初始设置为 5e-4, 训练迭代次数为 2500 个 ephco, 在编码器和解码器中, 使用 Relu 和 sigmoid 作为激活函数, 三元组损失边界常量 m 设为 0.5, 权重 $\alpha_1$、$\alpha_2$、$\alpha_3$ 分别设置为 5、1、5。

### 3.3 评价指标

使用 OULU-NPU 数据集做内部测试时, 本文比较真实人脸分类错误率(Bona-Fide Presentation Classification Error Rate, BPCER), 欺诈人脸分类错误率(Attack Presentation Classification Error Rate, APCER)以及平均分类错误率(Average Classification Error Rate, ACER)。在使用 CASIA-MFSD、Replay-Attack、MSU-MFSD 数据集做交叉实验时, 本文比较半总错误率(Half Total Error Rate, HTER), 它是错误拒绝率(False Reject Rate, FRR)和错误接受率(False Accept Rate, FAR)的均值, 同时使用曲线下面积(Acer Under

Circle, AUC)作为三个数据集内的交叉测试实验。

### 3.4 数据集内部测试

本文首先针对最终的优化模型在OULU-NPU上进行了数据集的内部测试。表1使用了OULU-NPU中开发的四个协议来评估本文模型的性能。与之比较的方法有 Auxiliary[36]、STASN[37]、LDA[38]、FaceDs[11]、LGSC[14]和TTN-T[20]。

**表1 对OULU进行数据集内部测试**

**Table1 Tesing on the OULU-NPU dataset**

| 协议 | 方法 | APCER(%) | BPCER(%) | ACER(%) |
|---|---|---|---|---|
| 1 | STASN | 1.2 | 2.5 | 1.9 |
| | Auxiliary | 1.6 | 1.6 | 1.6 |
| | LGSC | 0.8 | **0.0** | 0.4 |
| | TTN-T | 1.2 | **0.0** | 0.6 |
| | LDA | 1.1 | 0.4 | 0.7 |
| | 本文模型 | **0.03** | 0.6 | **0.3** |
| 2 | FaceDs | 4.2 | 4.4 | 4.3 |
| | Auxiliary | 2.7 | 2.7 | 2.7 |
| | STASN | 4.2 | 0.3 | 2.2 |
| | LGSC | **0.8** | 0.6 | 0.7 |
| | TTN-T | **0.8** | 0.8 | 0.8 |
| | LDA | 1.0 | 2.0 | 1.5 |
| | 本文模型 | 1.0 | **0.2** | **0.6** |
| 3 | STASN | 4.7±3.9 | **0.9±1.2** | 2.8±1.6 |
| | FaceDs | 4.0±1.8 | 3.8±1.2 | 3.6±1.6 |
| | Auxiliary | 2.7±1.3 | 3.1±1.7 | 2.9±1.5 |
| | LGSC | 1.5±1.4 | 1.9±1.9 | 1.7±1.6 |
| | TTN-T | **0.8±0.9** | 1.4±1.8 | 1.9±2.3 |
| | LDA | 1.6±1.2 | 1.7±1.1 | 1.5±1.2 |
| | 本文模型 | 1.0±0.8 | 1.2±1.1 | **1.1±0.7** |
| 4 | FaceDs | 5.1±6.3 | 6.1±5.1 | 5.6±5.7 |
| | Auxiliary | 9.3±5.6 | 10.4±6.0 | 9.5±6.0 |
| | STASN | 6.7±9.6 | 8.3±8.4 | 7.5±4.7 |
| | LGSC | 5.8±4.9 | **1.7±2.6** | 3.7±2.1 |
| | TTN-T | 4.2±2.4 | 3.8±4.0 | 4.0±2.3 |
| | LDA | **2.1±2.2** | 3.9±5.7 | **2.7±3.3** |
| | 本文模型 | 3.5±2.0 | **3.2±1.5** | 3.3±2.1 |

根据表1所示, 本文所优化的U-net模型的综合性能在所有协议上都有着不错的表现。具体来说, 相比于STASN、Auxiliary和LGSC等方法, 协议1中本文模型的APCER仅有0.03, 且协议1与协议2中本文的ACER均为最佳, 协议3与协议4在未知欺诈类型上评估了模型的泛化能力, 并且相比于其他方法本文模型在ACER指标上取得了最佳效果。实验结果表明, 本文模型对未知的环境条件、攻击媒介和相机传感器具有较强的泛化能力。

### 3.5 跨数据测试

为了进一步证明所优化的网络模型的泛化能力, 本文设置了跨数据测试实验。具体来说, 模型在一个数据集上训练, 然后在另一个数据集上测试。跨数据集的评估是具有挑战性的, 因为在不同的数据集之间, 真实样本和欺诈样本的数据分布有很大差异。本文选择在CASIA-MFSD和Replay-Attack数据集上进行跨数据集测试以评估模型的泛化能力, 选择LGSC、PatchNet[39]、LBP、BaseNet-Fusion[19]、STASN、FaceDS作为对比。根据表2, 本文模型在HTER(%)指标下取得了最佳效果。

**表2 跨数据集测试的HTER指标**

**Table2 HTER metrics for crossdataset testing**

| 方法 | 训练 | 测试 | 训练 | 测试 |
|---|---|---|---|---|
| | CASIA | Replay | Replay | CASIA |
| LBP | 47.0 | | 39.6 | |
| STASN | 31.5 | | 30.9 | |
| FaceDS | 28.5 | | 41.1 | |
| Auxiliary | 27.6 | | 28.4 | |
| LGSC | 27.4 | | 23.7 | |
| BaseNet-Fusion | 27.9 | | 38.5 | |
| PatchNet | **9.9** | | 26.2 | |
| 本文模型 | 17.3 | | **25.6** | |

如表2所示, 本文所优化的U-Net网络取得了较好的性能。具体来说, 以综合表现相当的PatchNet举例, 使用CASIA-MFSD做训练集, Replay-Attack做测试集时, 本文模型的HTER指标比PatchNet升高了7%左右; 使用Replay-Attack做训练集, CASIA-MFSD做测试集时, 本文模型的HTER为25.6%, 比PatchNet略低1%。由此可见, 本文模型的泛化性能有待提高。在跨数据集测试中, 我们注意到从高分辨率数据集(CASIA-MFSD)到低分辨率数据集(Replay-Attack)的性能有所下降。在高分辨率图像作为输入的情况下, 所提出的方法利用了丰富的纹理信息, 而这些信息在低分辨率图像上可能是缺失的。相反, 从低分辨率图像中学习到的欺诈线索可以很好的推广到高分辨率图像中。

### 3.6 语义增强模块对比

为了测试不同的语义增强模块对人脸特征增强的效果, 本文比较了GCM、SFEM、ASPP在原始U-Net上的表现, 计算了在OULU-NPU数据集下的ACER指标。如表3所示。

表3表明, 三个语义模块对U-Net都有明显提升, 其中在协议1的测试下, U-Net+GCM的APCER、

**表 3 在 OULU 数据集上进行测试**

**Table3 Testing on te OULU-NPU dataset**

| 协议 | 方法 | APCER(%) | BPCER(%) | ACER(%) |
|---|---|---|---|---|
| 1 | U-Net+GCM | **0.03** | 0.6 | **0.3** |
| | U-Net+SFEM | 0.4 | **0.4** | 0.4 |
| | U-Net+ASPP | 0.5 | 0.7 | 0.6 |
| 2 | U-Net+GCM | 0.8 | **0.6** | **0.7** |
| | U-Net+SFEM | 1.0 | 0.8 | 0.9 |
| | U-Net+ASPP | **0.7** | 1.7 | 1.2 |
| 3 | U-Net+GCM | **2.1±3.3** | **0.6±1.0** | **1.4±2.2** |
| | U-Net+SFEM | 3.0±4.1 | 0.3±4.4 | 1.7±4.3 |
| | U-Net+ASPP | 2.7±3.6 | 1.6±2.9 | 2.2±3.3 |
| 4 | U-Net+GCM | **4.2±3.6** | **1.2±0.8** | **2.7±2.2** |
| | U-Net+SFEM | 5.0±4.1 | 0.7±2.3 | 2.8±3.2 |
| | U-Net+ASPP | 4.5±5.6 | 2.2±3.1 | 3.4±4.3 |

ACER 都达到了最佳效果, 而在其他三个协议下的综合效果也都是最佳, 说明 GCM 模块的嵌入给网络模型带来了较出色的分类精度。具体来说, 在协议 1 下, U-Net+GCM 的 ACER 比 U-Net+ASPP 和 U-Net+SFEM 分别低 0.1 和 0.3; 在协议 2 下, U-Net+GCM 的 ACER 比 U-Net+ASPP 和 U-Net+SFEM 分别低 0.5 和 0.2; 在协议 3 和协议 4 下, U-Net+SFEM 与 U-Net+ASPP 在四个协议下的综合效果也略低于 U-Net+GCM。

然后本文使用 CASIA、Replay-Attack、MSU 在重放和打印攻击之间执行数据集跨类型测试, 其中表 4 比较了 GCM、SFEM、ASPP 结合原始 Unet 网络测试的效果。

如表 4 所示, 本文使用 CASIA-MFSD、Replay-Attack 和 MSU-MFSD 进行重放和打印攻击之间的数据集内交叉类型的测试, 可以看出, 基于 GCM 的方法达到了最佳的整体性能。综上所述, 经过在 OULU-NPU 的四个协议中测试以及在 CASIA-MFSD、MSU-MFSD 和 Replay-Attack 数据集的跨类型测试, 本文将 GCM 作为本文网络的语义增强模块。

**表 4 在三个数据集中跨类型测试的 AUC(%)**

**Table 4 Cross-type testing on the OULU-NPU dataset**

| 方法 | CASIA-MFSD | | Replay-Attack | | MSU-MFSD | |
|---|---|---|---|---|---|---|
| | Video | Photo | Video | Printed Photo | HR Video | Printed video |
| UNet+GCM | **98.2** | **99.92** | **99.99** | **99.43** | **99.99** | 90.97 |
| UNet+SFEM | 98.07 | 96.0 | 99.92 | 99.2 | 98.0 | 92.28 |
| UNet+ASPP | 97.02 | 98.0 | 97.0 | 99.2 | 99.3 | 91.25 |

### 3.7 中心差分卷积 θ 设置

在 Unet 网路中的编码器分支顶部嵌入 GCM 模块后, 继续将解码器的传统卷积算子修改为中心差分卷积算子(Central Difference Convolution, CDC)。根据公式 6, 其中 $\theta$ 控制中心差分的影响, 即控制人脸特征中梯度信息的利用程度, 本文首先利用 OULU-NPU 数据集测试了不同的 $\theta$ 对模型性能的影响, 如图 8 所示, 当 $\theta > 0.3$ 时, 中心差分卷积比传统卷积( $\theta = 0$, ACER=3.8%)表现出更好性能, 表明基于 CDC 的细粒度信息有助于人脸活体检测任务, 因为 $\theta = 0.7$ 时获得最佳性能, 所以本文将此设置用于接下来的实验。

### 3.8 消融实验

为了验证本文所提出模型中语义增强模块、中心差分卷积以及交叉注意力的有效性和必要性, 本文将最终模型与表 5 中的几个变体模型在 OULU-NPU 的第一个协议下进行了比较。具体来说, 首先本文分别对原始 U-Net 框架下加入 GCM、CDC、MHCA 做了单个模块的测试, 然后测试 GCM+CDC、GCM+MHCA、CDC+MHCA 两个模块在原始框架下的性能, 同时比较了 CDC 与空洞卷积[40](Diated Convolution, DC) 和深度可分离卷积[41](Deepwise Seperabel Convolution, DSC) 的性能, 最后将原始 U-Net 和本文模型做对比。

如表 5 所示, 在不加入任何模块时, 模型即为原始 U-Net 网络, 不难看出, 在分别加入 GCM、CDC、MHCA 的情况下, 平均分类错误率都优于原始网络,

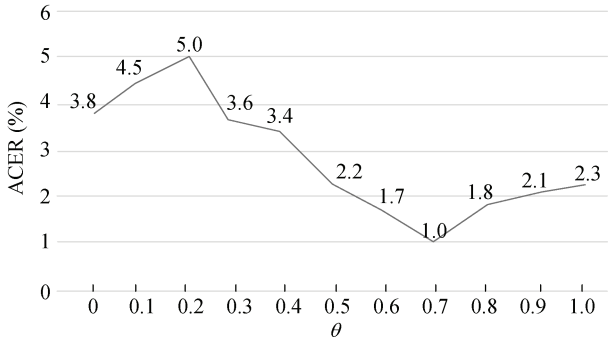


**图 8 $\theta$ 对 CDC 的影响**

**Figure 8 Effect of $\theta$ on CDC**

表 5 在 OULU-NPU 协议 1 下的消融实验测试

Table 5 Ablation testing on protocol 1 of the OULU-NPU dataset

| RL | GCM | CDC | DC | DSC | MHCA | APCER(%) | BPCER(%) | ACER(%) |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 2.1 | 4.3 | 3.2 |
| | ✓ | | | | | 1.8 | 3.3 | 2.6 |
| | | ✓ | | | | 2.0 | 3.6 | 2.8 |
| | | | | | ✓ | 2.2 | 4.0 | 3.1 |
| | ✓ | ✓ | | | | 1.7 | 3.0 | 2.4 |
| | ✓ | | | | ✓ | 1.9 | 3.7 | 2.8 |
| | | ✓ | | | ✓ | 1.9 | 3.5 | 2.7 |
| | ✓ | ✓ | | | ✓ | 1.1 | 0.1 | 0.6 |
| ✓ | ✓ | | ✓ | | ✓ | 1.0 | 0.8 | 0.9 |
| ✓ | ✓ | | | ✓ | ✓ | 1.2 | 0.5 | 0.8 |
| ✓ | ✓ | ✓ | | | ✓ | **0.9** | **0.0** | **0.5** |

表明了三个模块的有效性。且在相同条件下, CDC 对 U-Net 的优化明显优于 DC 和 DSC。根据 ACER, 只加入 GCM, 平均分类错误率降低了 0.6%, CDC 和 GCM 分别降低了 0.4%和 0.1%。加入两个模块时对模型略有提升, 当三个模块一起加入时, ACER 降低至 0.6%。继续将回归损失 RL 添加至网络模型中, BPCER 达到最佳效果, ACER 降低至 0.5%。综述所述, 本文所加入的三个模块都对模型有着较好的提升。

## 3.9 模型复杂度分析与可视化分析

### 3.9.1 模型复杂度分析

将模型中的传统卷积替换为 CDC 时, 在实际计算中, 只会引入额外的差分卷积参数用来计算差分卷积的结果, 并不会增加参数数量。针对 GCM 的加入, 它使用了不同的池化尺寸以及卷积操作, 具有四个相同结构的模块, 因为输出通道数的不同, 每个结构都会不同的参数量增加, 且其中 Non-Local 非局部块同样会造成复杂度的增加, 取决于输入特征图的大小和通道数, 主要的复杂度来源于对应的矩阵乘法操作和卷积操作。最后针对 MHCA 的引入, 其主要复杂度来源于 key,value 以及 query 的计算, 假如输入通道数为 256, 则会增加 $256\times256\times3$ 的参数量, 其中卷积和上采样操作同样会增加部分参数量, 使得复杂度有一定提升。综上所述, 根据输入特征图大小以及输入输出通道数的不同, 本文对于 U-Net 的优化会增加一定的复杂度, 计算量有所提升, 表 6 是对模型复杂度量化的表示。

如表 6 所示, 本文使用浮点运算次数(Floating Point Operations, FLOPs)衡量模型的复杂度, 并选择 LGSC、CDCN++[9]、PatchNet 做为对比。本节侧重模型复杂度的比较, 所以选择相似网络模型进行比较。例如 LGSC 中同样使用了编码器-解码器的网络结构, CDCN++中也使用了中心差分卷积, 可以对比中心差分卷积在不同网络模型中的效果, 而 PatchNet 作为一个轻量化模型, 在实验方面更容易进行比较, 使得比较更具可行性。不难看出, 其中 LGSC 与本文模型都使用了 U-Net 作为骨干网络, 但是本文模型的 FLOPs 相较于 LGSC 有所提升。因为本文对 U-Net 网络做了轻量化处理, 例如减少了编码器和解码器中的卷积块数量, 有利于降低模型的参数量和计算复杂度, 同时适当的减少了每个卷积层的通道数, 有助于减少每个层的通道数量, 从而减少模型的内存消耗。

表 6 模型计算量对比

Table 6 Model computational complexity comparison

| 方法 | FLOPs |
|---|---|
| LGSC | 9.56G |
| CDCN++ | 50.97G |
| PatchNet | 1.82G |
| 本文模型 | 19.12G |

### 3.9.2 样本可视化分布

如图 9 所示是本文通过 t-SNE[42]显示 OULU-NPU 协议 1 上测试的真实人脸特征和欺诈人脸特征的分布情况。其中(a)表示二维分布, (b)表示三维分布。

如图 7 所示, 红色代表活体样本, 蓝色代表欺诈样本, 它们展示了网络中解码层的特征嵌入, 不难看出, 所有活体样本都集中在一个集群当中, 而欺诈样本则远离活体样本, 该情况符合我们认为欺诈样本是活体样本的离群值的假设, 同时也证实了活体样本与活体样本之间的紧凑性和活体样本与欺诈样本的类间可分离性。

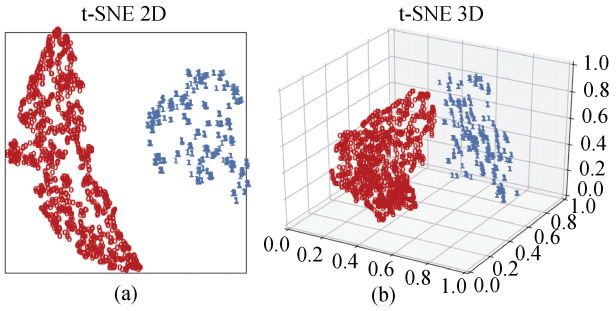


**图 9　样本可视化分布情况**

**Figure 9　visualization distribution**

### 3.9.3　欺诈线索特征映射

为了更加直观地理解欺诈线索，使用本文网络所训练的最佳模型参数生成了 OULU-NPU 数据集中活体人脸、打印人脸和视频重放人脸的欺诈线索示例。如图 10 所示，不难看出，活体人脸的欺诈线索映射图为全 0，而不同攻击人脸的欺诈线索也有较明显的区别。

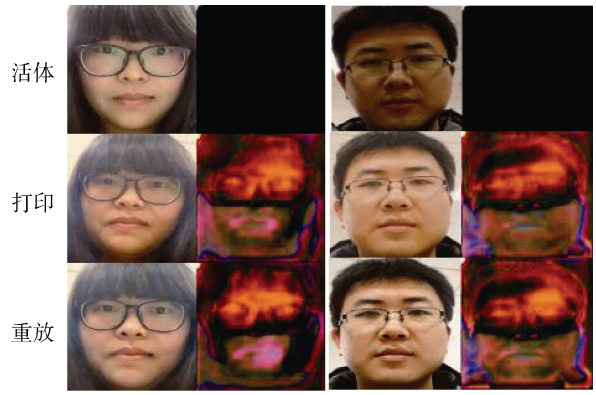


**图 10　欺诈线索图**

**Figure 10　Spoof cue map**

## 4　结束语

本文基于语义增强和中心差分卷积所优化的 U-Net 模型，结合多头交叉注意力，在人脸活体检测领域表现出较高的鲁棒性。在语义增强模块的选取上，GCM 可以提取全局上下文信息，ASPP 可以提取不同尺度的空间信息，SFEM 则是根据注意力机制选择性的增强模型中的重要特征，而根据在人脸数据集上的测试，GCM 在 U-Net 模型中的综合表现效果最佳。此外，为了提取人脸特征强度级别和梯度级别的细粒度信息以及空间信息，本文将中心差分卷积算子替代了解码块中的传统卷积算子，并在跳跃连接后引入 MHCA，形成本文最终的模型。经过数据集内部测试、交叉测试以及消融测试，证明所优化的 U-Net 模型用于人脸活体检测任务可以得到不错的性能。

实验中跨数据集测试的数据表明，本文模型在跨数据集检测的精度还有待提高。目前多分类活体检测可将不同攻击样本按照各自的欺诈特征进行分类，可以更好的学习到不同样本的内在共性特征，接下来的目标是尝试将数据集按类别精细划分，利用细分类活体检测，结合域泛化、度量学习等技术使模型性进一步提升。

## 参考文献

[1] Lu Z Q, Lu Z M, Shen F L, et al. A Survey of Face Anti-Spoofing[J]. *Journal of Cyber Security*, 2020, 5(2): 18-27.
(卢子谦, 陆哲明, 沈冯立, 等. 人脸反欺诈活体检测综述[J]. *信息安全学报*, 2020, 5(2): 18-27.)

[2] Liu S Q, Lan X Y, Yuen P C. Remote Photoplethysmography Correspondence Feature for 3D Mask Face Presentation Attack Detection[C]. *Computer Vision – ECCV 2018*, 2018: 577-594.

[3] Liu Z X, Ma R C. Facial Gender Classification Based on Eigenfaces and LS-SVM Classifiers[J]. *Journal of East China Jiaotong University*, 2007, 24(5): 85-88.
(刘遵雄, 马汝成. 基于特征脸和 LS-SVM 分类器的人脸性别分类[J]. *华东交通大学学报*, 2007, 24(5): 85-88.)

[4] Akbari M, Mohrekesh M, Nasr-Esfahani E, et al. Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network[C]. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2018: 69-72.

[5] Brandao P, Mazomenos E, Ciuti G, et al. Fully Convolutional Neural Networks for Polyp Segmentation in Colonoscopy[C]. *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017.

[6] Li G B, Yu Y Z. Contrast-Oriented Deep Neural Networks for Salient Object Detection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(12): 6038-6051.

[7] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]. *2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.

[8] Yang J W, Lei Z, Li S Z. Learn Convolutional Neural Network for Face Anti-Spoofing[EB/OL]. 2014: 1408.5601. https://arxiv.org/abs/1408.5601v2.

[9] Yu Z T, Zhao C X, Wang Z Z, et al. Searching Central Difference Convolutional Networks for Face Anti-Spoofing[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 5295-5305.

[10] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015: 234-241.

[11] Jourabloo A, Liu Y J, Liu X M. Face De-Spoofing: Anti-Spoofing via Noise Modeling[M]. Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 297-315.

[12] Feng H C, Hong Z B, Yue H X, et al. Learning Generalized Spoof

Cues for Face Anti-Spoofing[EB/OL]. 2020: 2005.03922. https://arxiv.org/abs/2005.03922v1.

[13] Chen J N, Lu Y Y, Yu Q H, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[EB/OL]. 2021: 2102.04306. https://arxiv.org/abs/2102.04306v1.

[14] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFS[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.

[15] Gu Z W, Cheng J, Fu H Z, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2281-2292.

[16] Schlemper J, Oktay O, Schaap M, et al. Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images[J]. *Medical Image Analysis*, 2019, 53: 197-207.

[17] Wang X L, Girshick R, Gupta A, et al. Non-Local Neural Networks[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7794-7803.

[18] Zhao H S, Shi J P, Qi X J, et al. Pyramid Scene Parsing Network[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6230-6239.

[19] Liu H F, Peng P, Chen T, et al. FECANet: Boosting Few-Shot Semantic Segmentation with Feature-Enhanced Context-Aware Network[J]. *IEEE Transactions on Multimedia*, 2023, 25: 8580-8592.

[20] Jiang W, Huang K, Geng J, et al. Multi-Scale Metric Learning for Few-Shot Learning[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(3): 1091-1102.

[21] Jia Y P, Zhang J, Shan S G, et al. Single-Side Domain Generalization for Face Anti-Spoofing[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 8481-8490.

[22] Fu J, Liu J, Wang Y H, et al. Adaptive Context Network for Scene Parsing[C]. *2019 IEEE/CVF International Conference on Computer Vision*, 2019: 6747-6756.

[23] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.

[24] Zhang R F, Li G B, Li Z, et al. Adaptive Context Selection for Polyp Segmentation[C]. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 2020: 253-262.

[25] Patel K, Bur A M, Wang G H. Enhanced U-Net: A Feature Enhancement Network for Polyp Segmentation[J]. *Proceedings of the International Robots & Vision Conference International Robots & Vision Conference*, 2021, 2021: 181-188.

[26] Chen L C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[EB/OL]. 2017: 1706.05587. https://arxiv.org/abs/1706.05587v3.

[27] Petit O, Thome N, Rambour C, et al. U-Net Transformer: Self and Cross Attention for Medical Image Segmentation[EB/OL]. 2021: 2103.06104. https://arxiv.org/abs/2103.06104v2.

[28] Määttä J, Hadid A, Pietikäinen M. Face Spoofing Detection from Single Images Using Micro-Texture Analysis[C]. *2011 International Joint Conference on Biometrics*, 2011: 1-7.

[29] Arashloo S R, Kittler J. Client-Specific Anomaly Detection for Face Presentation Attack Detection[EB/OL]. 2018: 1807.00848. https://arxiv.org/abs/1807.00848v1.

[30] Boulkenafet Z, Komulainen J, Li L, et al. OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations[C]. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition*, 2017: 612-618.

[31] Zhang Z W, Yan J J, Liu S F, et al. A Face Antispoofing Database with Diverse Attacks[C]. *2012 5th IAPR International Conference on Biometrics*, 2012: 26-31.

[32] Chingovska I, Anjos A, Marcel S. On the Effectiveness of Local Binary Patterns in Face Anti-Spoofing[C]. *The International Conference of Biometrics Special Interest Group*, 2012: 1-7.

[33] Wen D, Han H, Jain A K. Face Spoof Detection with Image Distortion Analysis[J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(4): 746-761.

[34] Zhang Y X, Li G, Cao Y, et al. A Method for Detecting Human-Face-Tampered Videos Based on Interframe Difference[J]. *Journal of Cyber Security*, 2020, 5(2): 49-72.
(张怡暄, 李根, 曹纭, 等. 基于帧间差异的人脸篡改视频检测方法[J]. *信息安全学报*, 2020, 5(2): 49-72.)

[35] Kingma D P, Ba J, Hammad M M. Adam: A Method for Stochastic Optimization[EB/OL]. 2014: 1412.6980. https://arxiv.org/abs/1412.6980v9.

[36] Liu Y J, Jourabloo A, Liu X M. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 389-398.

[37] Yang X, Luo W H, Bao L C, et al. Face Anti-Spoofing: Model Matters, so Does Data[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3507-3516.

[38] Sun Q H, Yin Z F, Wu Y C, et al. Latent Distribution Adjusting for Face Anti-Spoofing[EB/OL]. 2023: 2305.09285. https://arxiv.org/abs/2305.09285v1.

[39] Wang C Y, Lu Y D, Yang S T, et al. PatchNet: A Simple Face Anti-Spoofing Framework *via* Fine-Grained Patch Recognition[C]. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 20249-20258.

[40] Wang Y J, Wang G D, Chen C, et al. Multi-Scale Dilated Convolution of Convolutional Neural Network for Image Denoising[J]. *Multimedia Tools and Applications*, 2019, 78(14): 19945-19960.

[41] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1800-1807.

[42] Der Maaten L V, Hinton G E. Visualizing Data Using t-sne[J]. *Journal of machine learning research,*2008,9(11): 2579-2605.

**蔡体健** 于 2016 年在中南大学计算机应用技术专业获博士学位, 现为华东交通大学信息工程学院副教授, 硕士生导师, 研究领域为计算机视觉、深度学习、稀疏表示等。研究兴趣包括人脸活体检测, 图像处理。Email: cai2017@ecjtu.edu.cn

**陈均** 于 2024 年在华东交通大学计算机技术专业获硕士学位,研究领域为深度学习。研究兴趣包括人脸活体检测, 图像处理。Email: 1318141847@qq.com

**罗词勇** 于 2024 年在华东交通大学计算机技术专业获硕士学位,研究领域为深度学习。研究兴趣包括人脸活体检测, 图像处理。Email: 1755318494@qq.com

**刘遵雄** 现为华东交通大学信息工程学院教授, 硕士生导师, 研究领域为机器学习理论算法及应用、数据挖掘技术、模式识别及图像分析理解等。研究兴趣包括肿瘤分割, 医学图像处理。Email: 153010729@qq.com

**陈子涵** 于 2024 年在华东交通大学计算机技术专业获硕士学位, 研究领域为深度学习。研究兴趣包括肿瘤分割, 医学图像处理 Email: 1103151915@qq.com