三维深度点云监督和置信度修正的人脸欺诈检测算法

胡永健[1,3],蔡楚鑫[1,3],刘琲贝[1],王宇飞[2],廖广军[2]
(1. 华南理工大学电子与信息学院,广东广州 510641; 2. 广东警官学院刑事技术系,广东广州 510440;
3. 中新国际联合研究院,广东广州 510700)

摘 要: 基于深度学习的人脸身份认证由于使用便捷和用户体验好,成为我国当今最受欢迎的人工智能技术应用之一. 人脸识别和认证系统必须确保所比对的人脸是真实人脸,否则输出的结果没有任何商业价值. 位于系统前端的人脸欺诈检测也称活体检测是保障人脸识别和认证系统有效输出的关键. 现有人脸欺诈检测算法虽然库内性能尚佳,但由于实验室训练环境无法完全模拟真实应用场景,造成源域和目标域的数据在分布上存在差异,导致跨库检测性能明显下降. 尽管通过增加检测特征的种类和个数可以改善算法性能,但会导致检测网络构造复杂,模型变大,计算复杂度增加. 为了改善算法的跨库检测性能并降低计算的复杂度,本文提出一种基于三维(3D)深度点云监督和置信度修正机制的人脸欺诈检测算法. 主要贡献包括:设计了DenseBlockNet,仅用较浅层的DenseBlockNet网络即可提取真假人脸之间具有很好区分度的深度信息特征,模型小;将DenseBlockNet输出的二维深度图与采样点位置进行关联,构造三维深度点云,采用倒角损失函数监督预测的深度点云与实际点云标签之间的三维空间距离,同时还采用图二元交叉熵损失监督预测的深度图与深度图标签之间的差异;在3D深度点云预测模块中引入置信度修正机制,修正二分类误差,同时避免库内过拟合,提高算法的泛化能力. 所提出方法与包括2种最新文献的8种典型算法在Replay-attack、CASIA-FASD、MSU-MFSD、Rose-Youtu、OULU-NPU等5个主流人脸欺诈检测数据库上进行了充分的对比实验,实验结果表明,所提出的算法在库内和跨库检测中均能保持半总错误率最低或次低,且模型最小,参数量最少,计算复杂度最低.

关键词: 人脸欺诈检测; 三维深度点云; 3D深度点云监督; 置信度修正; 深度学习; 泛化能力

基金项目: 国家重点研发计划(No.2019QY2202);广州黄埔开发区国际合作项目(No.2019GH16);2021年度广东省重点建设学科科研能力提升项目(No.2021ZDJS047);教育部科技部司法鉴定技术应用与社会治理学科创新引智基地项目(No.B20077)

中图分类号: TP391.4;TP309.1 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3282-12

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220949

3D Depth Point Cloud Supervision and Confidence Correction for Face Spoofing Detection

HU Yong-jian[1,3], CAI Chu-xin[1,3], LIU Bei-bei[1], WANG Yu-fei[2], LIAO Guang-jun[2]
(1. *School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong* 510641, *China*;
2. *Department of Criminal Science and Technology, Guangdong Police College, Guangzhou, Guangdong* 510440, *China*;
3. *China-Singapore International Joint Research Institute, Guangzhou, Guangdong* 510700, *China*)

Abstract: Due to feasibility and friendly user interaction, deep learning-based face recognition and identity authentication becomes one of the most popular artificial intelligence technologies in China. The face recognition and identity authentication system should secure that the captured face for verification is a living face rather than a fake face or called spoofing face. Otherwise, the output of the system is useless for business. Face spoofing detection or called living face detection mechanism is set in the front part of the system, and plays a key role in distinguishing a fake face from the input faces. Most current face anti-spoofing algorithms perform well in intra-dataset. However, the model training in lab is unable to simulate all aspects in the real-world application scenarios. As a result, the data distribution in source domain is not always similar to the data distribution in target domain, which causes the lab-trained algorithms barely perform as well as in lab. Al-

收稿日期:2022-08-10;修回日期:2022-11-06;责任编辑:赵克

though we can mitigate the performance degradation with the increase of detection feature types and dimensions, it tends to make the detection network very complex in structure and large in model size. In order to improve the generalization ability of model without resorting to large model, we design a face spoofing detection network using the 3D depth point cloud supervision and confidence correction scheme. The proposed approach consists of three major contributions. First, we design a shallow convolutional neural network called DenseBlockNet. It can well extract distinctive depth features between real faces and spoofing ones and has a small model size. Second, we establish the relationship between the 2D depth map produced by DenseBlockNet and the coordinates of sampling points, and thus create a 3D depth point cloud. We adopt the Chamfer loss to minimize the distance between the learned 3D depth point cloud and the ground truth 3D depth point cloud label, and use the binary cross entropy loss to supervise the difference between the learned 2D depth map and the ground truth 2D depth map label. Third, we introduce a prediction confidence map to correct the error of the learned 3D depth point cloud, so that it can avoid overfitting in intra-datasets and obtain good generalization ability in inter-datasets. Extensive experiments are conducted on 5 popular presentation attack databases, namely Reply-attack, CASIA-FASD, MSU-MFSD, Rose-Youtu, and OULU-NPU. Compared with 8 representative methods including 2 SOTA methods, the proposed method can achieve the least or second least half-total-error-rates in either intra-dataset or inter-dataset tests. Besides, it has the smallest model, the least amount of model parameters and the lowest computational complexity.

Key words: face spoofing detection; 3D depth point cloud; 3D point cloud supervision; confidence correction; deep learning; generalization ability

Foundation Item(s): National Key Research and Development Project (No.2019QY2202); Science and Technology Foundation of Guangzhou Huangpu Development District (No.2019GH16); 2021 Scientific Research Capability Improvement Program for Key Discipline Construction of Guangdong Province (No.2021ZDJS047); Forensic Sciences and Social Governance Disciplinary Innovation Base (No.B20077)

1　引言

人脸识别具有重要的商业应用潜力,早在20多年以前就已经吸引学界的关注(例如文献[1]). 当今,人脸识别已被广泛应用于要求身份认证的人类生活各个层面(例如手机解锁、门禁和刷脸支付). 与此同时,不法分子也试图利用非法获取的目标者照片或视频,伪造身份,对人脸识别系统进行攻击. 因此,只有确保进入人脸识别系统比对环节的人脸是真实活体人脸而非照片/纸张打印或手机视频播放/回放等形成的虚假人脸/欺诈人脸/假脸,众多采用了身份认证的现代人工智能技术才能获得用户的信赖,得以推广应用. 典型易于实现且成本低廉的攻击方式包括照片打印、视频回放等二维呈现攻击. 随着手机、相机及打印机等硬件质量的不断改进,纸张打印和视频播放的质量不断提高,检测攻击的难度不断加深,潜在的威胁愈来愈大. 因此,针对防范这类常见攻击的研究具有重要的学术价值和极大的商业意义,一直以来都是学界和工业界研究的热点和难点.

人脸欺诈检测从特征提取的角度可粗略分为基于传统手工特征和基于深度学习特征两大类. 根据技术动机又可细分为7类[2]:(1)利用真实人脸的皮肤具有纸张打印、视频播放等假脸所不具备的图像纹理和颜色等视觉特征[3~5];(2)利用假脸所没有的真实活体人脸体温、面部血管脉搏等生物特征[6~8];(3)考虑到真实人脸面孔由于器官有凸凹,与摄像头有远近距离的不同,而纸张打印、视频播放等欺诈人脸来源于纸张或手机屏幕,是二维平面,没有这种距离差,故可利用深度、梯度或运动特征等辅助信息进行区别[9~12];(4)借助多个硬件检测设备,不同成像原理的摄像头对活体人脸和假体人脸的响应不同,例如,热敏传感器只能对有特定温度的人脸活体成像,而对纸张、视频屏幕等假脸无法成像,故可利用不同模态的图像信息[13~15];(5)利用半监督、自监督/无监督学习网络,提取真实人脸和假脸之间一般性的有区分度特征[16~19];(6)针对假脸材质以及在二次或多次成像过程中带入的成像噪声进行建模,利用解耦合/解纠缠学习网络,分离出假脸由于打印设备色域局限性引起的颜色失真或由于手机播放时屏幕呈现的摩尔纹等特有噪声,作为鉴别真假人脸的依据[20~22];(7)借助工业检测中将真实人脸看成是一类数据(即正品),而将所有非真实人脸均看成是异常数据(即次品)的思路[23~25]. 总的来看,上述7类方法在库内均能获得较好的效果,但在跨库检测中性能差异较大,普遍存在准确率急剧下降,模型泛化性能不足的问题. 直观上来看,通过增加特征种类和个数可以改善检测能力,但会导致硬件成本增高和计算复杂度增大. 因此,如何确保在实验室有限训练样本条件下所设计的算法能够适应真实世界中不同应用场景的未知检测对象自始至终都是人脸欺诈检测技术面临的最大挑战.

本文从真实人脸和欺诈人脸在深度信息上的本质差异出发,设计深度特征提取网络,构造3D点云形式的三维深度特征图,用较小的网络模型和较低的计算复杂度,有效地降低了算法的库内和跨库检测错误率.

主要贡献包括：(1)针对欺诈噪声大多来源于假脸材质以及多次成像留下的管道噪声和环境噪声，设计了轻量化的浅层特征提取网络DenseBlockNet，提取具有区分度的深度信息特征；(2)构造3D点云深度图将深度特征与采样点位置坐标进行关联，并利用倒角损失函数监督预测的点云与实际点云标签之间的三维空间距离，同时还利用图二元交叉熵损失(Map Binary Cross-Entropy Loss，MBCE)监督预测的深度图与深度图标签之间的差异；(3)在3D点云预测模块中引入了置信度修正机制，避免库内过拟合，以提高算法对不同域的泛化能力．在流行的5个人脸欺诈检测数据库上与8种典型的算法进行了实验对比，结果表明，本文算法无论在库内还是跨库，错误率保持最低或次低，且模型参数量和模型尺寸最小，浮点运算次数(FLoating Point Operations，FLOPs)最低，综合性能最优．

2 相关工作

利用软件估计深度等辅助信息来改善算法的泛化性能由于无需额外新增硬件而受到学界和工业界的广泛关注．文献[9]首次将深度图引入人脸欺诈检测，其双分支检测网络中的一条用卷积神经网络(Convolutional Neural Network，CNN)从待测图像的多个随机分块中提取特征，逐块打分，并对当前待测图像输出一个判定为假脸的平均分数；另一条分支是基于三维形变模型(3D Morphable Mode，3DMM)深度标签训练后得到的全卷积网络(Fully Convolutional Network，FCN)，可对待测图像输出一个深度图矩阵，由其构造特征向量，送入支持向量机(Support Vector Machine，SVM)分类器得到一个判定为假脸的分数．两条支路的分数加权后求平均即为当前图像的最终判定结果．

文献[9]需借助3DMM深度标签来训练FCN网络获取深度图，这种方法不直接，为此文献[10]进行了改进，提出了一个端到端的深度图输出网络DenseBlock2Net，在真实人脸深度图为全1，假脸深度图为全0的假设条件下，通过逐像素监督训练得到深度图．文献[10]算法的整体框架如图1所示，具体做法是使用DenseNet169前8层作为特征提取网络，包括2个DenseBlock和2个Transition Block．前8层输出384通道的14×14特征图，再沿通道经1层1×1卷积，利用sigmoid激活函数获得预测的14×14深度图，对假脸和真脸分别使用14×14的全0和全1按图二元交叉熵损失L-MBCE进行逐像素监督，另一分支将预测的14×14深度图展平并通过全连接层使用二元交叉熵损失L-BCE进行二元监督．测试时使用预测的0-1深度图的均值作为判断真脸和假脸的依据．文献[10]的不足是依赖单帧图像来估计深度图，未能考虑相邻帧深度图及相邻帧运动模式之间的相互影响．

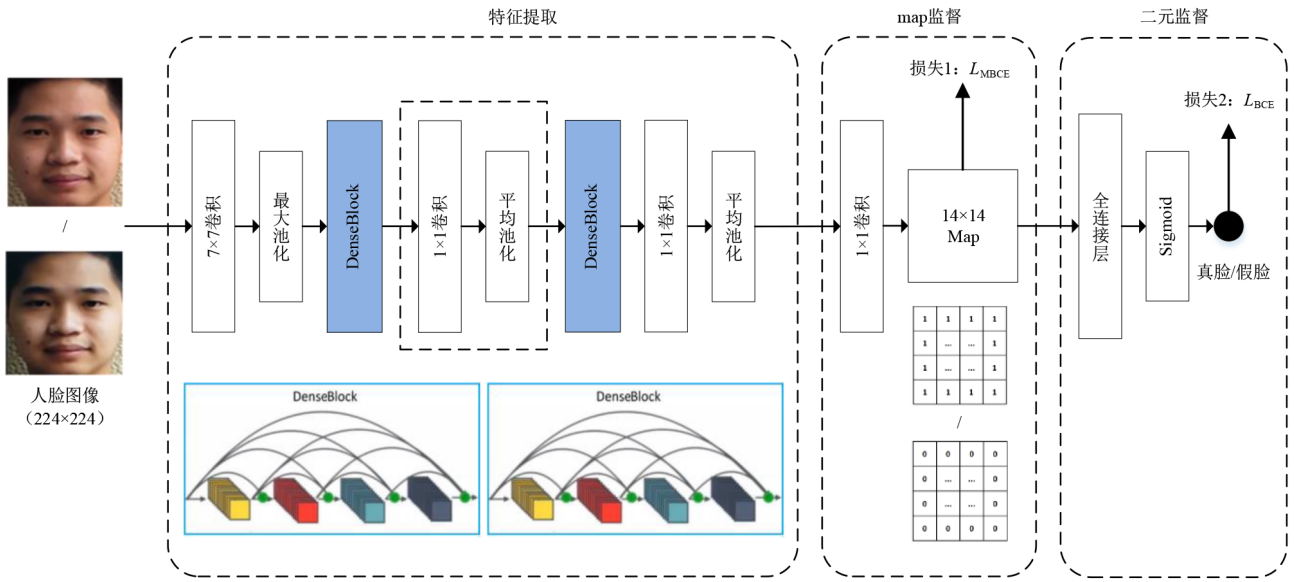


图1 文献[10]算法的整体框架图

文献[11]对此提出了一种称为DepthMap的改进算法，先用空间梯度残差模块(Residual Spatial Gradient Block，RSGB)分别提取各帧人脸图像的低阶、中阶和高阶细粒度空间特征并加以融合，预测各帧人脸的粗糙深度图；然后设计了时空传播模块(Spatio-Temporal Propagation Module，STPM)，利用短时空时模块(Short-Term Spatio-Temporal Block，STSTB)获取相邻帧人脸图像之间的短时空时域信息，并利用卷积门控循环单元(Convolutional Gated Recurrent Unit，ConvGRU)对其进行融合，获取表征多帧的长时空时域信

息,再对粗糙深度图进行细化,获得精细且有区分度的深度图. 该方法的不足是所得到的深度图没有与人脸面部的具体采样点位置进行对应,且时空模块的计算复杂度高.

文献[12]进一步进行了改进,提出了一种3D点云(3DPC-Net)的编解码网络用于真假脸鉴别. 其深度图有三个维度,称为三维点云,其中水平和垂直方向的位置坐标分别为x和y,深度值为z. 对于真实人脸,每个点(x,y)的z值非0,即有深度;而对于假脸,z值恒为0,即无深度. 该文采用3DDFA[26]首先生成原始的3D点云标签,重构的稠密人脸点云共有53 215个采样点. 兼顾计算的复杂度和特征的多样性后,从总共53 215个点中随机采样10 000个点,训练时再从中随机采样L_2距离最大的2 500个点作为最终的3D点云标签. 3DPC-Net网络的目的是对于给定的RGB(红、绿、蓝)人脸图像,能够生成带有3D人脸结构特征的3D深度图. 其整体网络结构如图2所示. 首先将224×224×3的人脸图像输入到ResNet-18前5层构建的编码器中,提取1×256的全局深度特征,接着复制全局特征向量,尺寸变为2 500×256,与选取2 500个点的x和y坐标拼接,尺寸变为2 500×258,然后送入由两层1D卷积构建的解码器中,获得2 500×3的预测3D点云,并利用倒角损失(Chamfer Loss)监督,减小预测3D点云和真实3D点云标签的差异. 测试时根据预测3D点云的深度通道的均值判别真脸和假脸. 该方法的不足有2点:一是提取全局深度特征时需要较多的采样点;二是直接把256维全局特征复制2 500份与2 500个采样点进行对应,难以准确反映每个位置的真实深度特征.

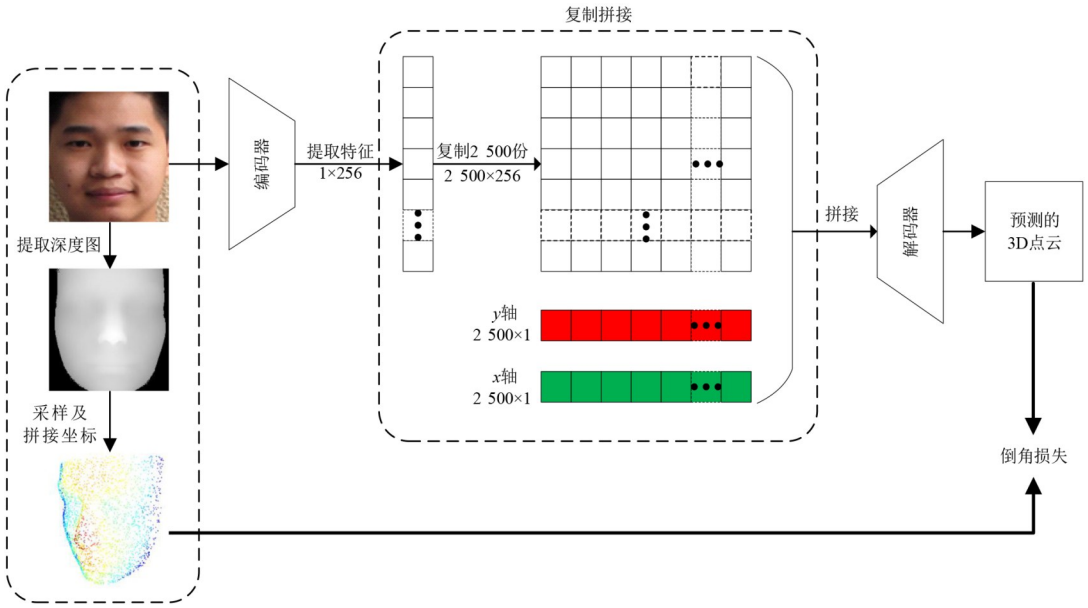


图2　文献[12]算法的整体框架图

针对文献[12]的不足,并受文献[10~12]的启发,我们提出了一种带有深度位置信息的3D点云监督人脸欺诈检测算法,为避免库内过拟合,进一步利用了置信度修正策略,提高算法的泛化性能. 在计算的复杂度方面,为满足工业上的实时性要求,设计了一种轻量化的特征提取网络. 所提出的网络与原始文献相比,各方面性能均有明显改善.

3　本文算法整体框架与关键环节

本文算法整体框架如图3所示,包括数据预处理、深度特征提取、3D点云重建与监督以及置信度预测与修正4个模块. 训练阶段,数据预处理模块抠取人脸、提取人脸深度图并对其采样,然后与x、y坐标位置信息拼接得到3D点云标签,该标签将作为3D点云重建网络的监督信息. 用DenseBlockNet提取人脸深度信息特征,与采样点坐标(x,y)拼接构成包含位置信息的深度特征. 所得深度特征分别送入2个网络,通过3D点云重建网络获得预测3D点云,通过置信度预测网络获得预测置信度图,后者对前者进行逐点修正. 最后,利用修正后的预测3D点云进行二元分类. 测试阶段则无需提取3D点云监督信息. 下面对关键环节进行描述.

3.1　数据预处理模块和深度特征提取模块

为获取人脸3D点云标签,数据预处理模块首先抠取人脸图像,提取人脸深度图并归一化到[0,1]. 令真实人脸深度图为D,则假脸深度图为1-D. 参考文献[12]的做法,分别对抠取的真脸和假脸进行采样,真实人脸的z非0,假脸z为0,与x坐标、y坐标拼接,得到3D点云标签.

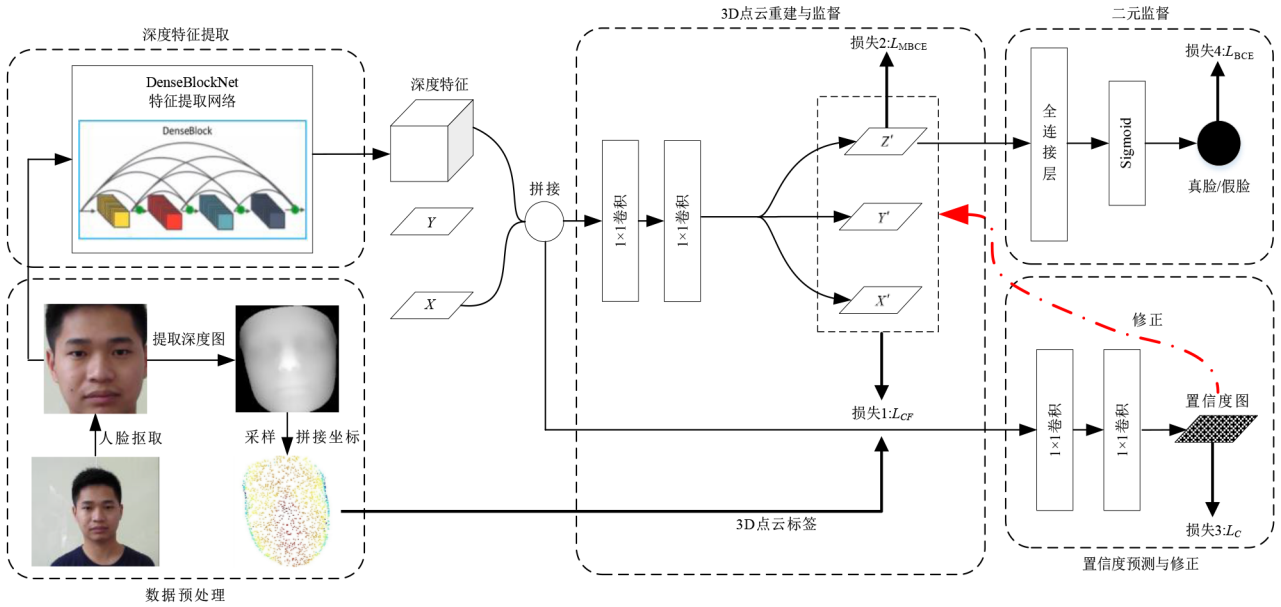


图3 本文算法整体框架图

考虑到假脸样本中的伪影痕迹,如模糊、条带效应、摩尔纹等更适合被表征为表象信息[6],且深度属于与局部特征相关的表象信息,本文设计了浅层网络DenseBlockNet提取深度特征. CNN网络的特点是网络层数越深,感受野越大,特征越抽象. 浅层网络主要用于提取局部且精细的表象信息,深层网络主要用于提取全局且抽象的语义信息[27].

我们参考文献[10]的做法,将DenseNet169前8层作为特征提取网络,包含2个DenseBlock和2个TransitionBlock. 同时,对网络做了进一步轻量化,只保留一个DenseBlock和一个TransitionBlock,将原第2个DenseBlock替换为卷积核数为256的3×3卷积层,将原第2个TransitionBlock替换为平均池化. DenseBlockNet的网络结构如图4所示. 输入为224×224×3的人脸图像,最终提取得到14×14×256的深度特征. 其中,DenseBlock包含6个由1×1卷积和3×3卷积构成的DenseLayer层,每层的输出以前馈方式拼接到后面各层的输入,可减少梯度消失并很好地保留各层特征;TransitionBlock由1×1卷积和平均池化层构成,用于标准化和下采样以减少特征尺寸;每个卷积层后面跟着批次归一化层(Batch Normalization layer,BN),并使用ReLU进行非线性激活. DenseBlockNet的详细结构如表1所示.

3.2 3D点云重建和监督

如前指出,文献[12]直接将全局深度特征通过复制生成3D点云不够精确,且直接使用稠密3D点云进行监督的计算复杂度过高. 为提高深度特征与位置的匹配精度,同时降低计算量,本文设计了3D深度点云重建网络,采用了采样点级的位置信息匹配策略,网络结构如图5所示. 先将DenseBlockNet提取的尺寸为14×14×256的深度特征与14×14=196采样点对应的x坐标、

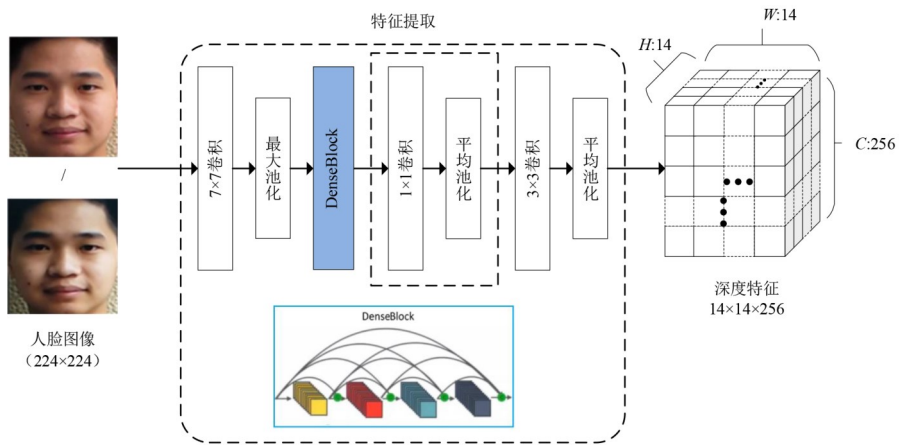


图4 轻量化深度特征提取网络DenseBlockNet结构图

表1　DenseBlockNet详细结构

结构	网络层	输入通道数/个	步长/像素	填充值/像素	输出尺寸/像素	输出通道数/个
	输入	3	—	—	224×224	3
初步提取	7×7卷积	3	2	3	112×112	64
	最大池化	64	2	1	56×56	64
DenseLayer_1	1×1卷积	64	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
DenseLayer_2	1×1卷积	96	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
DenseLayer_3	1×1卷积	128	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
DenseLayer_4	1×1卷积	160	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
DenseLayer_5	1×1卷积	192	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
DenseLayer_6	1×1卷积	224	1	0	56×56	128
	3×3卷积	128	1	1	56×56	32
TransitionBlock	1×1卷积	256	1	0	56×56	128
	平均池化	128	2	0	28×28	128
深度特征输出	3×3卷积	128	1	1	28×28	256
	平均池化	256	2	0	14×14	256

y坐标拼接成尺寸为14×14×258的特征图,然后送入卷积核数目分别为128和3、步长为1的2个1×1卷积层,利用1×1卷积与空间无关的计算方式将采样点深度特征与对应位置信息结合,最后输出尺寸为14×14×3、取值区间为[0,1]的预测3D点云. 其中,2个1×1卷积的激活函数分别为ReLU和Sigmoid.

网络训练时,设深度阈值T_d,从样本符合阈值条件的点中等间隔采样196个样本,构成14×14×2的x坐标、y坐标向量图. 阈值条件:真脸样本的深度大于T_d,假脸样本的深度小于$1-T_d$. 测试时,同样等间隔选取196个坐标作为x坐标、y坐标向量图,以预测对应位置的深度值.

3.3　3D点云监督重建损失函数的设计

本文方法基于采样点深度特征进行3D点云重建,深度特征与位置信息的匹配至关重要,为此引入2个损失函数衡量3D点云重建质量,以监督网络更好学习深度特征. 具体地,采用倒角损失函数L_{CF}衡量预测点云和点云标签之间的点与点三维空间距离,采用图二元交叉熵损失L_{MBCE}衡量预测深度图与深度图标签的差异.

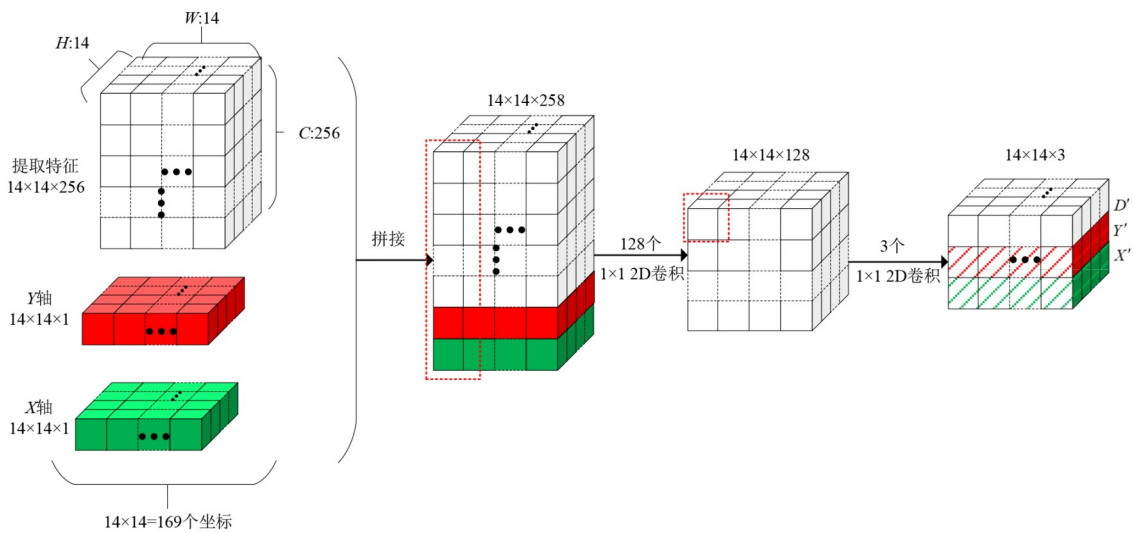


图5　利用采样点深度特征的3D点云重建网络结构示意图

倒角损失函数L_{CF}定义为

$$L_{\mathrm{CF}}\left(P_c, P_c'\right)=\frac{1}{\left|P_c\right|}\sum_{p\in P_c}\min_{p'\in P_c'}\left\|p-p'\right\|_2^2+\frac{1}{\left|P_c'\right|}\sum_{p'\in P_c'}\min_{p\in P_c}\left\|p'-p\right\|_2^2 \tag{1}$$

其中，P_c表示3D点云标签，P_c'表示预测3D点云．p和p'分别表示P_c和P_c'中一个点的三维坐标，$|\cdot|$表示3D点云的总点数，$\|\cdot\|_2$表示L_2距离．损失函数的第一项用于求取P_c中每个点到P_c'的最小距离的均值，第2项用于求取P_c'中每个点到P_c的最小距离的均值，两者综合衡量3D点云P_c与P_c'的差异．

图二元交叉熵损失L_{MBCE}定义为

$$L_{\mathrm{MBCE}}=-\left(Z\log\left(Z'\right)+\left(1-Z\right)\log\left(1-Z'\right)\right) \tag{2}$$

其中，Z和Z'分别表示深度图标签和预测深度图．

重建模块输出的预测深度图被展平为196维特征向量，送入输出神经元数目为1的全连接层和Sigmoid激活层，采用下式定义的二元交叉熵损失L_{BCE}进行二元监督：

$$L_{\mathrm{BCE}}=-\left(y\log\left(y'\right)+\left(1-y\right)\log\left(1-y'\right)\right) \tag{3}$$

其中，y和y'分别表示分类分数标签和预测分类分数．

3.4 置信度预测与修正

深度神经网络具有优良的预测性能，但无法察觉其预测结果的错误，可能造成分类分数非常高的错误预测[28]，这对二分类人脸欺诈检测任务的影响尤其严重．本文方法引入置信度预测与修正模块来提升人脸欺诈检测算法泛化性能．由于没有可用于置信度估计的真实标签，直接预测置信度较困难．本文借鉴文献[29]直接预测置信度的方法，在训练过程通过预测的置信度对分类结果进行加权修正，同时使用置信度损失进行惩罚，利用两者的博弈来逐步提高置信度预测的合理性．

设置信度预测网络输出的置信度为$c\in[0,1]$，训练阶段对预测3D点云P_c'和预测深度图Z'的修正如式(4)和式(5)所示：

$$P_c''=c\times P_c'+\left(1-c\right)\times P_c \tag{4}$$

$$Z''=c\times Z'+\left(1-c\right)\times Z \tag{5}$$

修正后的3D点云P''和深度图Z''引入了3D点云标签P_c和深度图标签Z的先验知识．由于$1-c$表示引入的先验知识占比，置信度预测网络会倾向于输出数值较小的c以便引入更多先验知识．为此，引入式(6)的置信度损失L_C进行惩罚，防止置信度c的预测值过小，通过两者的博弈以输出合理的置信度．

$$L_C=-\log\left(c\right) \tag{6}$$

在测试阶段，由于没有真实标签作为先验知识，所以无法按式(4)和式(5)的方式进行置信度修正．为此，我们提出一种无先验知识修正方法，根据验证集确定分类分数阈值T，再按式(7)定义对预测深度图进行置信度修正．

$$\begin{cases}z''=z'\times c, & \text{if } z'>T\\ z''=z'/c, & \text{if } z'<T\end{cases} \tag{7}$$

需要说明的是，根据预测置信度应用的范围不同，有全局修正和逐采样点修正两种方案．全局修正方案假设深度图每个点的置信度是一样的，预测的全局置信度c为标量．实验表明全局修正方案的检测性能不是最优，因此本文采用逐采样点修正方案，即假设3D点云每个点的置信度是有差异的．逐采样点置信度预测与修正的示意图如图6所示．深度特征与位置信息拼接后被送入由卷积核数分别为128和1，步长为1的2个1×1卷积层，利用1×1卷积与空间无关的计算方式预测3D点云对应点的置信度，2个卷积层的激活函数分别为ReLU和Sigmoid，最后输出14×14×1的预测置信度图．测试时先使用预测置信度图对预测深度图进行修正，再计算深度均值，作为最终预测结果．

3.5 总损失函数

综合上述各模块，本文方法构建的总损失函数为

$$L=\lambda_1 L_{\mathrm{CF}}+\lambda_2 L_{\mathrm{MBCE}}+\lambda_3 L_C+\lambda_4 L_{\mathrm{BCE}} \tag{8}$$

其中，λ_1、λ_2、λ_3、λ_4分别表示倒角损失、图二元交叉熵损失、置信度损失和二元交叉熵损失的占比权重．需注意，引入置信度修正模块后，式(1)和式(2)定义的L_{CF}和L_{MBCE}中，P_c'、p'和Z'应分别替换为修正后的P_c''、p''和Z''.

4 本文算法的实验评估

4.1 人脸欺骗检测数据库

本文在Replay-Attack[30]、CASIA-MFSD[31]、MSU_MFSD[32]、Rose_Youtu[33]和OULU-NPU[34]5个流行的人脸欺诈检测数据库上进行了实验，分别简称为R库、C库、M库、Y库和OULU库．其中，OULU库常用于衡量库内跨域性能，包含4种跨域策略．P1跨场景策略使用训练集中Session1和Session2的数据训练，在测试集的Session3数据上测试；P2跨攻击媒介策略使用训练集中Print1和Display1的数据训练，在测试集的Print2和Display2数据上测试．OULU库包含6个手机设备拍摄的视频，P3跨设备策略使用其中5个设备的数据训练，用余下的1个设备的数据测试，需要进行6个子实验，最终结果使用均值±标准差的形式表示．P4策略结合了P1、P2和P3三种策略，也需要进行6个子实验．

通常，测试集的分类阈值由库内验证集达到等错误率(Equal Error Rate，EER)时对应的阈值来决定．EER指验证集错误接受率(False Acceptance Rate，FAR)和错误拒绝率(False Rejection Rate，FRR)相等时的错误率．由于C库、M库和Y库没有划分验证集，本

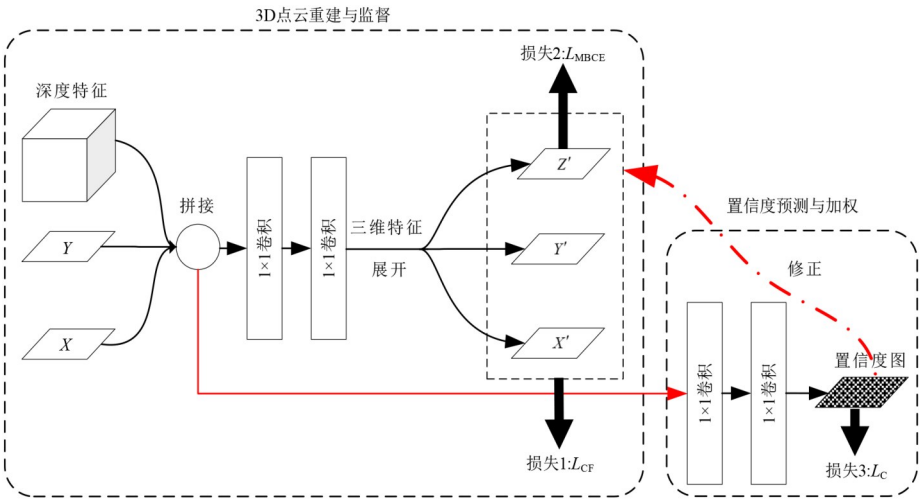


图6　逐采样点置信度预测与修正示意图

文使用其测试集来计算分类阈值.

4.2　算法评价指标

为全面评估算法性能,实验分为库内实验和跨库实验两部分. 对于跨库实验,使用跨库测试集的半总错误率(Half Total Error Rate, HTER)作为性能指标,即FAR和FRR的均值. 对于R库的库内实验,以验证集的EER和测试集的HTER作为性能指标. 对于无验证集的C库、M库和Y库的库内实验,以测试集的EER为性能指标. 特别地,遵循数据库的官方指引[35],对于OULU库分别计算攻击呈现分类错误率(Attack Presentation Classification Error Rate, APCER)、真实呈现分类错误率(Bona Fide Presentation Classification Error Rate, BPCER)和平均分类错误率(Average Classification Error Rate, ACER)作为衡量指标,分别对应于按照攻击媒介单独计算的最差FAR、FRR和前两者的均值.

此外,使用参数量、模型大小和浮点运算次数FLOPs等指标衡量算法模型在存储大小、计算复杂度方面的性能.

4.3　实验环境及参数设置

实验在Ubuntu14.04.6操作系统上进行,使用深度学习框架Pytorch1.7.1实现. 实验中,先使用Dlib工具库的人脸检测模型抠取人脸,并修改尺寸为224×224×3作为网络输入,使用3DDFA算法[26]提取人脸深度图,在训练时对输入的人脸图像使用通用的数据增强,包括随机水平翻转、尺寸为16的Cutout以及亮度、对比度和饱和度的随机抖动.

关于模型训练中的超参数设置,为了合理且有足够多的候选点,经过实验测试,将深度阈值T_d取值为0.7. 按照与深度图的相关性,在通过实验测试后,将总损失函数中4个损失L_{CF}、L_{MBCE}、L_C、L_{BCE}的占比权重λ_1、λ_2、λ_3、λ_4分别设为1.5,1.0,0.5和0.1.

训练时使用的小批量为32,使用SGD优化器最小化总损失函数,初始学习率和学习率衰减参数分别为1×10^{-3}和0,并采用步进学习率调整策略,每连续10个训练周期损失未下降则学习率衰减为原来的0.8. 最多训练300个训练周期,保存验证损失最小的模型为最终模型. 测试时,对每一个样本进行测试,使用修正后的预测深度图的均值作为最终的预测分类分数.

综上,实验中主要参数的设置如表2所示.

表2　实验主要参数取值

实验参数	取值
网络输入尺寸	224×224×3
深度阈值T_d	0.7
损失权重λ_1	1.5
损失权重λ_2	1.0
损失权重λ_3	0.5
损失权重λ_4	0.1
批量尺寸Batch Size	32
SGD优化初始学习率	1×10^{-3}
学习率衰减触发周期	10个
学习率衰减倍数	0.8
最大训练周期	300个

5　实验结果与分析

为验证所提出算法的性能,我们仿真了包括Colour Texture[3]、MSR-MobileNet[4]、DRL-FAS[18]、DeepPixBis[10]、3DPC-Net[12]、DeSpoofing[20]、OCDA[35]和DSDG[36]8种近期典型的人脸欺诈检测算法,在环境相同的条件下进行了库内和跨库对比实验,下列表格中加粗数据表示最优实验结果.

5.1　库内实验

如表3所示,相较于对比算法,本文方法在R库、M库和Y库的库内测试中均取得最低EER,分别为

表3 库内结果对比 单位:%

训练库→测试库	C→C	R→R	M→M	Y→Y
指标	EER	EER	EER	EER
Colour Texture[3]	2.28	2.10	4.93	7.07
MSR-MobileNet[4]	3.66	0.43	3.03	5.48
DRL-FAS[18]	**0.20**	**0.00**	0.26	1.80
DeepPixBis[10]	1.46	0.01	0.33	1.40
3DPC-Net[12]	1.77	0.04	0.98	2.70
OCDA[35]	4.27	0.47	0.68	3.96
DSDG[36]	1.22	0.44	0.42	1.62
本文方法	0.55	**0.00**	**0.01**	**1.21**

0.00%、0.01%和1.21%,在C库上EER为0.55%,是次优结果,略高于使用强化学习的DRL-FAS方法,但以较大幅度低于其他算法. 在OULU库四个测试策略上的实验结果如表4所示,本文算法在P1、P3和P4测试策略中均取得最优结果,平均分类错误率ACER分别为0.23%、0.88%±0.65%和1.38%±1.03%. 在P2测试策略中,本文算法ACER为1.41%,仅略高于最新的DSDG方法的1.37%,为次优结果. 上述实验结果充分证明本文方法具有良好的库内检测以及库内跨域检测性能.

表4 OULU库4个测试策略的结果对比 单位:%

测试策略	P1	P2	P3	P4
指标	ACER	ACER	ACER	ACER
MSR-MobileNet[4]	6.70	6.30	6.30±2.20	11.30±3.90
DRL-FAS[18]	4.70	1.90	3.00±1.50	7.20±3.90
DeepPixBis[10]	0.42	5.97	11.11±9.40	25.00±12.67
DepthMap[11]	1.00	1.90	2.70±0.60	5.00±2.20
3DPC-Net[12]	1.20	3.00	2.80±0.50	3.50±5.40
DeSpoofing[20]	1.50	4.30	3.60±1.60	5.60±5.70
OCDA[35]	4.51	5.57	4.95±1.41	12.40±5.44
DSDG[36]	0.37	**1.37**	1.40±1.50	2.30±2.30
本文方法	**0.23**	1.41	**0.88±0.65**	**1.38±1.03**

5.2 跨库实验

如表5所示. 在R→C和R→M这2个跨库测试中,本文算法取得了最优检测结果,半总错误率HTER分别为20.34%和13.56%. 在C→R、M→R和Y→R测试中,本文方法效果为次优,HTER分别为12.58%、6.36%和9.17%,仅次于应用域自适应框架的OCDA方法,而与其他对比算法比较则展现出较为明显的优势. 在Y→C测试中,本文方法仍为次优,性能略低于使用强化学习的DRL-FAS方法,HTER为10.80%,相较于包括最新的OCDA方法和DSDG方法在内的其它对比算法仍展现出明显的性能提升. 上述结果显示本文方法具有良好的跨库检测性能.

表5 与其他算法的跨库测试对比 单位:%

训练库→测试库	C→R	R→C	M→R	R→M	Y→C	Y→R
指标	HTER	HTER	HTER	HTER	HTER	HTER
Colour Texture[3]	29.80	38.16	35.20	33.47	41.35	28.11
MSR-MobileNet[4]	30.25	33.53	26.48	38.55	32.45	20.86
DRL-FAS[18]	28.40	33.20	29.70	15.60	**8.10**	20.00
DeepPixBis[10]	25.71	35.22	27.42	38.54	25.72	22.00
3DPC-Net[12]	22.26	27.73	27.60	20.91	17.41	21.16
DeSpoofing[20]	28.50	41.10	33.20	27.80	37.20	38.50
OCDA[35]	**3.50**	31.90	**2.90**	20.80	21.70	**3.00**
DSDG[36]	15.10	26.70	21.90	17.43	12.51	17.43
本文方法	12.58	**20.34**	6.36	**13.56**	10.80	9.17

5.3 消融实验

本节通过消融实验讨论3D点云监督、置信度修正及浅层特征提取网络的作用.

5.3.1 轻量化深度特征提取网络的作用

如3.1节所述,本文采用的深度特征提取网络DenseBlockNet是对文献[10]中的特征提取网络进行轻量化改进得来,其出发点是认为浅层网络更适于提取局部且精细的表象信息. 表6给出了在图3所示的算法框架下,对其中的深度特征提取模块分别采用文献[10]原特征提取网络(记为DeepPixBis-Dense)和本文轻量提取网络(DenseBlockNet)所得到的性能对比. 可以看出,采用轻量化特征提取网络得到的HTER明显下降,尤其是在M→R的跨库实验中,HTER由12.72%下降到7.61%,下降了5个百分点. 表7给出在OULU库上的对比实验结果,DenseBlockNet与DeepPixBis-Dense相比,在四个测试策略中的ACER均有下降. 实验表明浅层特征提取网络对跨域性能有较好的提升作用.

5.3.2 3D点云重建与监督的作用

为验证本文算法中3D点云重建与监督模块的作用,表8对比了是否采用该模块对检测性能的影响. 去除3D点云重建与监督模块时,为与算法其他模块保持兼容,将深度特征提取网络DenseBlockNet的输出通过一个1×1卷积层,得到一个14×14×1的特征图,再进行二元分类. 从表可见,使用3D点云重建与监督模块后,

表6　深度特征提取网络DenseBlockNet的消融实验结果HETR　　单位:%

训练库	C库		R库			M库		Y库		
测试库	C(库内)	R(跨库)	R(库内)	C(跨库)	M(跨库)	M(库内)	R(跨库)	Y(库内)	C(跨库)	R(跨库)
DeepPixBis-Dense	0.75	17.11	0.02	22.71	17.16	0.17	12.72	1.32	11.46	10.80
DenseBlockNet	**0.63**	**15.24**	**0.00**	**21.68**	**15.30**	**0.01**	**7.61**	**1.27**	**11.28**	**10.19**

表7　深度特征提取网络DenseBlockNet的消融实验在OULU库的测试结果　　单位:%

测试策略	P1	P2	P3	P4
指标	ACER	ACER	ACER	ACER
DeepPixBis-Dense	0.69	1.92	2.31±1.04	4.13±2.53
DenseBlockNet	**0.35**	**1.73**	**1.21±0.68**	**2.80±1.84**

表8　3D点云重建与监督作用的消融实验结果HTER　　单位:%

训练库	C库		R库			M库		Y库		
测试库	C(库内)	R(跨库)	R(库内)	C(跨库)	M(跨库)	M(库内)	R(跨库)	Y(库内)	C(跨库)	R(跨库)
无3D点云重建模块	1.46	25.71	0.17	35.22	38.54	0.34	27.42	**1.40**	25.72	22.00
有3D点云重建模块	**1.24**	**18.01**	**0.00**	**24.48**	**20.55**	**0.04**	**17.11**	1.44	**12.39**	**11.63**

在库内和跨库实验上HTER均有不同程度的下降(除了在Y库的库内测试中有小幅度上升),在跨库实验中的性能改进较为明显,尤其是在R→M库上,HTER下降了近18个百分点,表明3D点云重建与监督模块的有效性.

5.3.3　逐采样点置信度修正的作用

为验证逐采样点置信度修正模块的作用,表9和表10对比了是否采用该模块对检测性能的影响. 可以看出,引入了逐采样点置信度修正后,库内和跨库实验中得到的HTER均有一定程度的下降;在OULU库四个策略实验中得到的ACER也均有所下降. 这表明引入逐采样点置信度修正有利于提高检测算法性能.

表9　逐采样点置信度修正作用的消融实验结果HTER　　单位:%

训练库	C库		R库			M库		Y库		
测试库	C(库内)	R(跨库)	R(库内)	C(跨库)	M(跨库)	M(库内)	R(跨库)	Y(库内)	C(跨库)	R(跨库)
无置信度修正模块	0.63	15.24	**0.00**	21.68	15.30	**0.01**	7.61	1.27	11.28	10.19
有置信度修正模块	**0.55**	**12.58**	**0.00**	**20.34**	**13.56**	**0.01**	**6.36**	**1.21**	**10.80**	**9.17**

表10　逐采样点置信度修正作用的消融实验在OULU库的测试结果　　单位:%

测试策略	P1	P2	P3	P4
指标	ACER	ACER	ACER	ACER
无置信度修正模块	0.35	1.73	1.21±0.68	2.80±1.84
有置信度修正模块	**0.23**	**1.41**	**0.88±0.65**	**1.38±1.03**

5.4　模型大小和复杂度分析

如表11所示,与一些基本网络结构和其他算法相比,本文提出的检测网络参数量最少、模型最小、计算量最小,说明本文算法在部署代价和计算效率方面具有一定的优势.

表11　各种网络结构的计算复杂度指标对比表

网络类型	网络结构	参数量/M	FLOPs/G	模型大小/MB
基本网络	ResNet-18	11.68	2.38	44.67
	Inception-V3	27.16	2.85	103.94
	DenseNet169	14.15	3.42	54.96
图监督	DeepPixBis[10]	1.43	2.11	5.60
3D点云监督	3DPC-Net[12]	11.34	1.91	43.34
	本文方法	**0.74**	**1.53**	**2.89**

6　结束语

本文针对现有人脸欺诈检测算法普遍存在的跨库检测错误率高、泛化性能不好的问题进行了研究,提出了一种基于3D深度点云监督和置信度修正的人脸欺诈检测算法. 通过逐采样点的3D深度点云监督,引导网络学习深度特征与采样点位置的精确对应关系,并利用置信度对预测的3D深度点云进行了修正,避免网络过于依赖库内特征而导致跨库性能下降,泛化性能不好的问题. 我们利用对比文献原作者公开的程序代码或自己编写的程序代码实际仿真了8种近期典型的人脸欺诈检测算法,在5个公认度最高的人脸欺诈检测

数据库上进行了大量实验. 实验结果表明,本文算法在保证库内检测错误率最低或次低的同时获得了跨库检测最低或次低的好性能,显著改善了跨库检测性能,且网络模型更小,计算复杂度更低,适合实际应用场景的部署. 将来的改进方向包括引入梯度、运动等多种辅助信息监督,更全面反映真实人脸和假脸的特征差异.

参考文献

[1] 丁嵘, 苏光大, 林行刚. 使用关键点信息改进弹性匹配人脸识别算法[J]. 电子学报, 2002, 30(9): 1292-1294.
DING R, SU G D, LIN X G. Using key points to improve elastic matching in face recognition[J]. Acta Electronica Sinica, 2002, 30(9): 1292-1294. (in Chinese)

[2] 胡永健, 王宇飞, 刘琲贝, 等. 人脸欺诈检测最新进展及典型方法[J]. 信号处理, 2021, 37(12): 2261-2277.
HU Y J, WANG Y F, LIU B B, et al. A survey on the latest development and typical methods of face anti-spoofing [J]. Journal of Signal Processing, 2021, 37(12): 2261-2277. (in Chinese)

[3] BOULKENAFET Z, KOMULAINEN J, HADID A. Face spoofing detection using colour texture analysis[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(8): 1818-1830.

[4] CHEN H N, HU G S, LEI Z, et al. Attention-based two-stream convolutional networks for face spoofing detection [J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 578-593.

[5] 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 (CVPR). Piscataway: IEEE, 2020: 5294-5304.

[6] PINTO A, GOLDENSTEIN S, FERREIRA A, et al. Leveraging shape, reflectance and albedo from shading for face presentation attack detection[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3347-3358.

[7] YU Z T, LI X B, NIU X S, et al. Face anti-spoofing with human material perception[C]//European Conference on Computer Vision. Cham: Springer, 2020: 557-575.

[8] LIU S Q, LAN X Y, YUEN P C. Multi-channel remote photoplethysmography correspondence feature for 3D mask face presentation attack detection[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2683-2696.

[9] ATOUM Y, LIU Y, JOURABLOO A, et al. Face anti-spoofing using patch and depth-based CNNs[C]//2017 IEEE International Joint Conference on Biometrics (IJCB). Piscataway: IEEE, 2017: 319-328.

[10] GEORGE A, MARCEL S. Deep pixel-wise binary supervision for face presentation attack detection[C]//2019 International Conference on Biometrics (ICB). Piscataway: IEEE, 2019: 1-8.

[11] WANG Z Z, YU Z T, ZHAO C X, et al. Deep spatial gradient and temporal depth learning for face anti-spoofing [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 5041-5050.

[12] LI X, WAN J, JIN Y, et al. 3DPC-net: 3D point cloud network for face anti-spoofing[C]//2020 IEEE International Joint Conference on Biometrics. Piscataway: IEEE, 2020: 1-8.

[13] LIU W H, WEI X K, LEI T, et al. Data fusion based two-stage cascade framework for multi-modality face anti-spoofing[J]. IEEE Transactions on Cognitive and Developmental Systems, 2022, 14(2): 672-683.

[14] LIU A J, TAN Z C, WAN J, et al. Face anti-spoofing via adversarial cross-modality translation[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2759-2772.

[15] SIGMUND D, KERCKHOFF F, MAGDALENO J Y, et al. Face presentation attack detection in ultraviolet spectrum via local and global features[C]//2020 International Conference of the Biometrics Special Interest Group (BIOSIG). Piscataway: IEEE, 2020: 1-5.

[16] QUAN R J, WU Y, YU X, et al. Progressive transfer learning for face anti-spoofing[J]. IEEE Transactions on Image Processing, 2021, 30: 3946-3955.

[17] QIN Y X, ZHAO C X, ZHU X Y, et al. Learning meta model for zero-and few-shot face anti-spoofing[C]//The 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 11916-11923.

[18] CAI R Z, LI H L, WANG S Q, et al. DRL-FAS: A novel framework based on deep reinforcement learning for face anti-spoofing[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 937-951.

[19] CAI R Z, LI Z, WAN R J, et al. Learning meta pattern for face anti-spoofing[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1201-1213.

[20] JOURABLOO A, LIU Y J, LIU X M. Face de-spoofing: Anti-spoofing via noise modeling[C]//European Conference on Computer Vision. Cham: Springer, 2018: 297-315.

[21] ZHANG K Y, YAO T P, ZHANG J, et al. Face anti-spoofing via disentangled representation learning[C]//European Conference on Computer Vision. Cham: Springer, 2020: 641-657.

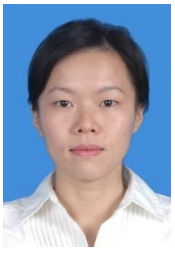
[22] WANG G Q, HAN H, SHAN S G, et al. Cross-domain

face presentation attack detection via multi-domain disentangled representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 6678-6687.
[23] QIN Y X, ZHANG W G, SHI J P, et al. One-class adaptation face anti-spoofing with loss function search[J]. Neurocomputing, 2020, 417: 384-395.
[24] FATEMIFAR S, ARASHLOO S R, AWAIS M, et al. Client-specific anomaly detection for face presentation attack detection[J]. Pattern Recognition, 2021, 112: 107696.
[25] LI Z, LI H L, LAM K Y, et al. Unseen face presentation attack detection with hypersphere loss[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2020: 2852-2856.
[26] ZHU X Y, LIU X M, LEI Z, et al. Face alignment in full pose range: A 3D total solution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 78-92.
[27] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
[28] Guo C, Pleiss G, Sun Y, et al. On calibration of modern neural networks[C]//International Conference on Machine Learning. Cambridge MA: JMLR, 2017: 1321-1330.
[29] KENDALL A, GAL Y. What uncertainties do we need in Bayesian deep learning for computer vision?[C]//The 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2017: 5580-5590.
[30] CHINGOVSKA I, ANJOS A, MARCEL S. On the effectiveness of local binary patterns in face anti-spoofing[C]// 2012 International Conference of Biometrics Special Interest Group (BIOSIG). Piscataway: IEEE, 2012: 1-7.
[31] ZHANG Z W, YAN J J, LIU S F, et al. A face antispoofing database with diverse attacks[C]//The 5th IAPR International Conference on Biometrics (ICB). Piscataway: IEEE, 2012: 26-31.
[32] 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.
[33] LI H L, LI W, CAO H, et al. Unsupervised domain adaptation for face anti-spoofing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(7): 1794-1809.
[34] BOULKENAFET Z, KOMULAINEN J, LI L, et al. OULU-NPU: A mobile face presentation attack database with real-world variations[C]//The 12th IEEE International Conference on Automatic Face & Gesture Recognition. Piscataway: IEEE, 2017: 612-618.
[35] LI Z, CAI R Z, LI H L, et al. One-class knowledge distillation for face presentation attack detection[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 2137-2150.
[36] WU H T, ZENG D, HU Y B, et al. Dual spoof disentanglement generation for face anti-spoofing with depth uncertainty learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4626-4638.

作者简介

胡永健　男,1962年出生,湖北武汉人. 华南理工大学电子与信息学院教授. 主要研究方向为多媒体信息安全、图像处理、人工智能及其应用. 中国电子学会会员编号:E190005206S.
E-mail: eeyjhu@scut.edu.cn

蔡楚鑫　男,1996年出生,广东揭阳人. 华南理工大学电子与信息学院硕士研究生. 主要研究方向为多媒体信息安全、图像处理、人工智能及其应用.
E-mail: eechuxincai@mail.scut.edu.cn



刘琲贝(通讯作者)　女,1980年出生,广东广州人. 华南理工大学电子与信息学院讲师. 主要研究方向为多媒体信息安全、图像处理、人工智能及其应用.
E-mail: eebbliu@scut.edu.cn

王宇飞　男,1987年出生,海南海口人. 广东警官学院刑事技术系讲师. 主要研究方向为多媒体信息安全、图像处理、人工智能及其应用.
E-mail: 20220710@gdppla.edu.cn

廖广军　男,1981年出生,四川渠县人. 广东警官学院刑事技术系教授. 主要研究方向为多媒体信息安全.
E-mail: 20060266@gdppla.edu.cn