第 60 卷 第 10 期/2023 年 5 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于真实人脸类别对抗机制的人脸活体检测算法

张磊[1,2]，盖绍彦[1,2*]，达飞鹏[1,2,3]

[1]东南大学自动化学院，江苏 南京 210096；

[2]东南大学复杂工程系统测量与控制教育部重点实验室，江苏 南京 210096；

[3]东南大学深圳研究院，广东 深圳 518063

摘要 针对现有人脸活体检测算法在单一数据集内表现良好而在多个数据集间泛化能力较差的问题，提出一种聚焦于真实人脸的活体检测方法。在数据输入阶段，每轮训练会向网络输入所有源域的真实人脸的同时只随机输入其中一个源域的虚假人脸。在特征学习阶段，使用Resnet18网络作为主干网络，对不同残差块的输出特征进行基于注意力机制的加权融合。利用三元组损失和对抗损失对融合后的真实人脸特征进行领域内和领域间的聚合，利用三元组损失对融合后的虚假人脸特征只进行领域内的聚合。在分类阶段，利用交叉熵损失对所有源域的真实人脸和虚假人脸进行分类。所提方法在4个人脸活体检测数据集中进行了实验，实验结果表明所提方法相比其他方法具有更低的识别错误率和更高的鲁棒性。

关键词 图像处理；模式识别；人脸活体检测；三元组损失；生成对抗机制；多尺度注意力融合机制

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP220649

Face Liveness Detection Algorithm Based on Real Face Category Adversarial Mechanism

Zhang Lei[1,2], Gai Shaoyan[1,2*], Da Feipeng[1,2,3]

[1]*School of Automation, Southeast University, Nanjing* 210096, *Jiangsu, China*;

[2]*Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing* 210096, *Jiangsu, China*;

[3]*Shenzhen Research Institute, Southeast University, Shenzhen* 518063, *Guangdong, China*

Abstract Given that existing face liveness detection algorithms perform well in a single data set but have poor generalization ability in cross multiple data sets; therefore, this study proposes a liveness detection method centering on real faces. During the data input stage, each round of training will input the real faces of multiple source domains into the network, while only randomly input false faces of one source domain. During the feature learning stage, Resnet18 serves as the backbone network to weight fuse the output features of different residual blocks based on the attention mechanism. Triple loss and adversarial loss are used to aggregate the fused real face features within each domain and cross domains, while triplet loss is used to aggregate the fused fake face features within each domain. During the classification stage, cross-entropy loss is used to classify real and false faces in all source domains. The proposed method was tested on four live face detection data sets, and the experimental results reveal that the proposed method has a lower recognition error rate and higher robustness than other methods.

Key words image processing; pattern recognition; face liveness detection; triplet loss; generative adversarial mechanism; multi-scale attention fusion mechanism

1 引 言

近年来，随着计算机视觉技术的不断发展，人脸识别技术获得广泛应用，小到办公室打卡、小区门禁，大到移动支付、火车站身份核验等。然而，大多数的人脸识别技术仅仅考虑当前人脸的语义特征，即该人脸属

收稿日期：2022-01-27；**修回日期**：2022-01-30；**录用日期**：2022-02-16；**网络首发日期**：2022-02-26

基金项目：国家自然科学基金(51475092)、江苏省前沿引领技术基础研究专项(BK20192004C)、深圳市科技创新委员会(JCYJ20180306174455080)

通信作者：*qxxymm@163.com

于哪个个体，并不会考虑该人脸是真实人脸还是虚假人脸。真实人脸指的是相机直接拍摄到的人脸，虚假人脸则是通过攻击手段得到的人脸。常见的攻击手段有打印攻击、视频重放攻击、三维面具攻击等，这些攻击手段给人脸识别系统带来了巨大挑战。为了解决这个问题，很多学者投入到人脸活体检测算法的研究中。人脸活体检测算法大致可以分为基于纹理和基于时域的两大类方法。前者通过人工设计的算子[1]或者深度学习方法抽取出分类特征，如颜色[2]、形变[3-4]、深度[5]等特征，进行真假脸的分类。后者通过抽取连续视频帧中的时域信息进行分类，如rPPG方法[6-7]利用真实人脸面部的血液流动性，而虚假人脸没有这种流动性。

尽管现有检测方法能够在单个数据集内取得较好的识别精度，但是将网络模型应用于其他数据集时识别率会大幅下降，即模型的泛化能力较差。当在单个数据集上训练模型时，模型学习的分类特征具有数据集偏向性[8]，即模型没有考虑各个数据集之间的内在联系。为了解决这个问题，许多方法使用领域自适应技术借助不带标签的目标域数据来减小源域和目标域之间的分布差异。然而在实际应用中，目标域数据的获取成本往往很高，有些情况下我们对目标域数据甚至一无所知[9]，因此有些学者使用领域泛化技术来应对该问题。领域泛化指在训练阶段使用多个源域数据集，同时不接触任何目标域数据的含义。本文中领域(domain)等价于数据集，源域(source domain)即训练数据集，目标域(target domain)即测试数据集。传统的领域泛化致力于通过对齐不同源域的数据学习到一个泛化的特征空间，并且假设模型从目标域中抽取的特征能够靠近该泛化的特征空间，从而保证模型的泛化性能。文献[10]中的方法利用多重对抗机制学习泛化的特征空间；文献[11]中的方法对样本之间的相对重要性进行重新加权，以提升模型的泛化性能；文献[12]中的方法利用对抗机制降低欺诈无关性因素对泛化性能的影响。这三个方法对真脸和假脸都采用了相同的处理方式，即模型学习到的真实人脸和虚假人脸空间各自都是紧凑的。而在人脸活体检测任务中，真脸和假脸是完全不同的两个类别。对于真脸而言，其大都是相机直接拍摄的人脸。对于假脸而言，其收集方式多种多样，如纸张打印、视频重放、三维面具等。真假脸收集方式的差异使得不同源域的真脸特征相对于假脸特征来说更容易聚集起来，如果对真脸和假脸采用相同的聚集方式会导致分类精度的下降[9]。文献[9]中的方法对不同源域的假脸进行发散，而对不同源域的真脸进行聚集。在训练集较小、测试集较大的实验中，由于文献[9]中的方法的假脸发散机制，本就不容易聚集的假脸被迫与其他源域的假脸进行进一步发散，从而靠近甚至跨过分类边界，造成分类精度的下降。

针对以上问题，本文借鉴文献[9]中对真假脸采用不同聚集方式的思路。具体来说，本文在文献[9]的基础上各自提升真脸和假脸的聚集度，即采用进一步聚集真脸而不发散假脸的方式，从而保持真假脸之间聚集度的相对差距。对于假脸，仅利用三元组损失进行领域内的聚集，而不进行领域间的聚集或发散。对于真脸，利用对抗损失和三元组损失进行领域内和领域间的聚集，从而增强真实人脸的领域不变性。如此一来，真脸聚集度依旧比假脸聚集度高，有助于学习到泛化能力更强的分类边界。

2 基本原理

图1的右图为所提方法的动机图。图中的四，五、六、八角星分别表示不同数据集的虚假人脸，圆形和不同方向的椭圆表示不同数据集的真实人脸。其中六角星表示测试数据集的假脸，可以看出左图中六角星跨过分类边界。水平方向的椭圆表示测试数据集的真脸。其他情况均为训练数据集的真脸和假脸。曲线为分类边界。

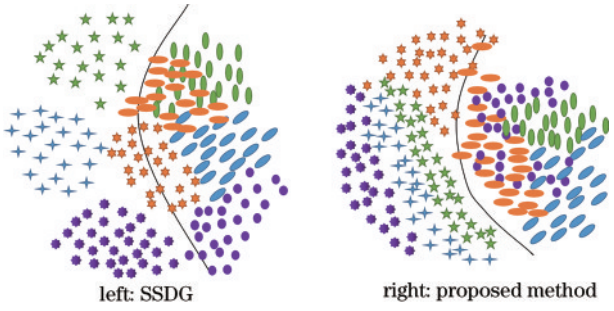

图1 不同方法的动机

Fig. 1 Motivation of different methods

如图1左图所示，不同数据集的假脸各自发散，其中测试数据集的假脸跨过分类边界，说明模型预测出现误判；而在右图中，不同数据集的真脸混合在一起，表明所提模型中真脸的领域不变性相较于文献[9]的SSDG模型更强，同时不同数据集的假脸虽然聚集在一起，但没有混合。从图1左图和右图的比较中可以得出结论：在保持真假脸聚集度差异的前提下，进一步聚集真脸的同时不发散假脸能够得到泛化能力更强的分类边界。

虽然真假脸之间聚集度的相对差距得以保持，但是不适当的领域不变性会影响分类精度[13-14]。为了更好地平衡领域不变性和分类任务，本文采用多尺度注意力融合机制对主干网络不同残差块的输出特征进行融合，得到融合特征，再对该融合特征进行上述的真假脸聚集操作，如图2中fusion feature所示。同时分类器会对该融合特征进行真脸和假脸的二分类。

图2为所构建的人脸活体检测的主体框架。由于现有的人脸活体检测方法[15-17]证实使用Resnet18网络

作为主干网络可以获取不错的效果，同时该网络相比于其他Resnet分支如Resnet50[18]等含有更少的参数量，有利于缓解过拟合问题。因此将Resnet18作为主干网络，对应图2中的$Conv_1$、$Conv_2$、$Conv_3$、$Conv_4$、$Conv_5$。

如图2中source fake和source real所示，主干网络的输入为三个源域的真脸及其中随机一个源域的假脸。通过多尺度注意力融合模块得到主干网络中4个残差块$Conv_2$、$Conv_3$、$Conv_4$、$Conv_5$的输出的融合权重，利用该权重与原残差块的输出进行加权融合，得到融合特征，该融合特征随后经L2 normalization进行归一化。对于归一化后的融合特征，首先只有其中的真实人脸特征进入特征判别器分支，如图2虚线箭头所示，特征生成器和特征判别器利用对抗损失进行真实人脸类别对抗，使得特征生成器学习到的真实人脸特征是领域无关且类别无关的。接着利用三元组损失对融合特征中的真实人脸和虚假人脸特征进行度量学习，对真实人脸特征进行领域间和领域内的聚集，对虚假人脸特征只进行领域内的聚集。最后分类器利用交叉熵损失对融合特征进行二分类，即判断对应的输入图像是真脸还是假脸。

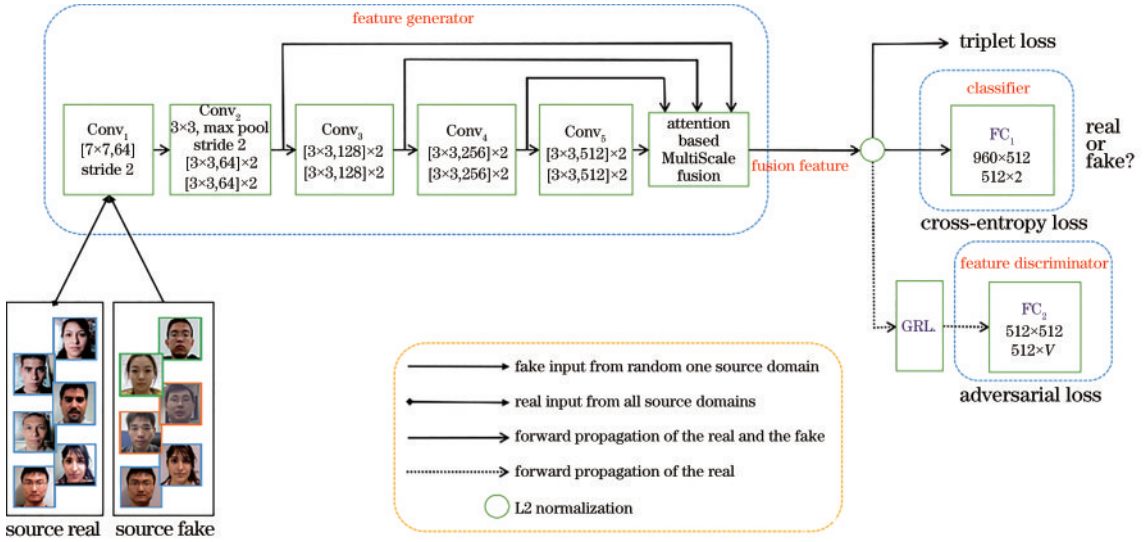

图2　模型总体架构

Fig. 2　Overall architecture of the proposed model

2.1　多尺度注意力特征融合

网络的浅层特征包含更多的轮廓信息[19]，而深层特征包含判断真假的语义相关的信息。不同数据集下的真实人脸和虚假人脸的轮廓信息大同小异，因此浅层特征相对于深层特征而言在不同数据集间的差异较小，适用于对领域不变性的学习，而深层特征由于包含了复杂的真假类别信息，因而更适用于分类任务。鉴于此，对网络中不同层的特征进行融合，从而充分利用各层特征的特点平衡领域不变性和分类任务的学习。相比于直接拼接不同层特征的方式，基于注意力的加权特征融合能够充分挖掘不同层特征的重要性，同时也能平衡两个任务[20]。

如图2和图3所示，$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$分别为主干网络中不同残差块$Conv_2$、$Conv_3$、$Conv_4$、$Conv_5$的输出，经过全局平均池化层各自特征图的宽度w和高度h降至一维，同时通道数目保持不变。随后经过特征拼接层对各组特征图从通道维度进行拼接，得到特征图$\boldsymbol{X}_{\mathrm{con}}\in\mathbf{R}^{1\times1\times c_{\mathrm{con}}}$，通道数$c_{\mathrm{con}}=\sum_{i=1}^{4}c_i$，其中$c_1$、$c_2$、$c_3$、$c_4$分别为$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$的通道数。$\boldsymbol{X}_{\mathrm{con}}$的计算式为

$$\boldsymbol{X}_{\mathrm{con}}=f_{\mathrm{cat}}(\boldsymbol{X}_i'),\tag{1}$$

$$\boldsymbol{X}_i'=\mathrm{GlobalAvgPool}(\boldsymbol{X}_i),\tag{2}$$

式中：$1\leqslant i\leqslant4$；GlobalAvgPool(·)为全局平均池化操作；$\boldsymbol{X}_i$和$\boldsymbol{X}_i'$分别为池化前和池化后的特征图；f_{cat}表示特征拼接。在得到特征图$\boldsymbol{X}_{\mathrm{con}}$后，经过两个卷积层得到通道数为4的输出向量，接着通过复制操作将该向量每个通道的数目分别扩展为c_1、c_2、c_3、c_4，从而得到不同特征的融合权重$\boldsymbol{w}_i$，其表达式为

$$\boldsymbol{w}_i=f_{\mathrm{epd}}\left\{\mathrm{SoftMax}\left\{f_{\mathrm{Conv_1}}^{1\times1}\left\{\mathrm{ReLU}\left[f_{\mathrm{Conv_2}}^{1\times1}(\boldsymbol{X}_{\mathrm{con}})\right]\right\}\right\}\right\},\tag{3}$$

式中：ReLU为激活函数，增加非线性；SoftMax函数对输出的权重特征进行归一化，即四个权重之和为1，从而起到更好的平衡作用；f_{epd}为特征展开层，将权重特征$\boldsymbol{w}_i$的形状扩充成与池化后特征$\boldsymbol{X}_i'$一致的形状，方便逐元素加权计算。最后，对权重特征与池化后的特征进行逐元素相乘得到各自加权后的特征。该加权特征再进入特征拼接层，得到最终的融合特征$\boldsymbol{X}_{\mathrm{fus}}$，其表达式为

$$\boldsymbol{X}_{\mathrm{fus}}=f_{\mathrm{cat}}\left(\boldsymbol{w}_i\otimes\boldsymbol{X}_i'\right),\tag{4}$$

式中：⊗表示特征图逐元素相乘。

2.2　真实人脸类别对抗

不同数据集的真脸采集途径大都是通过摄像头

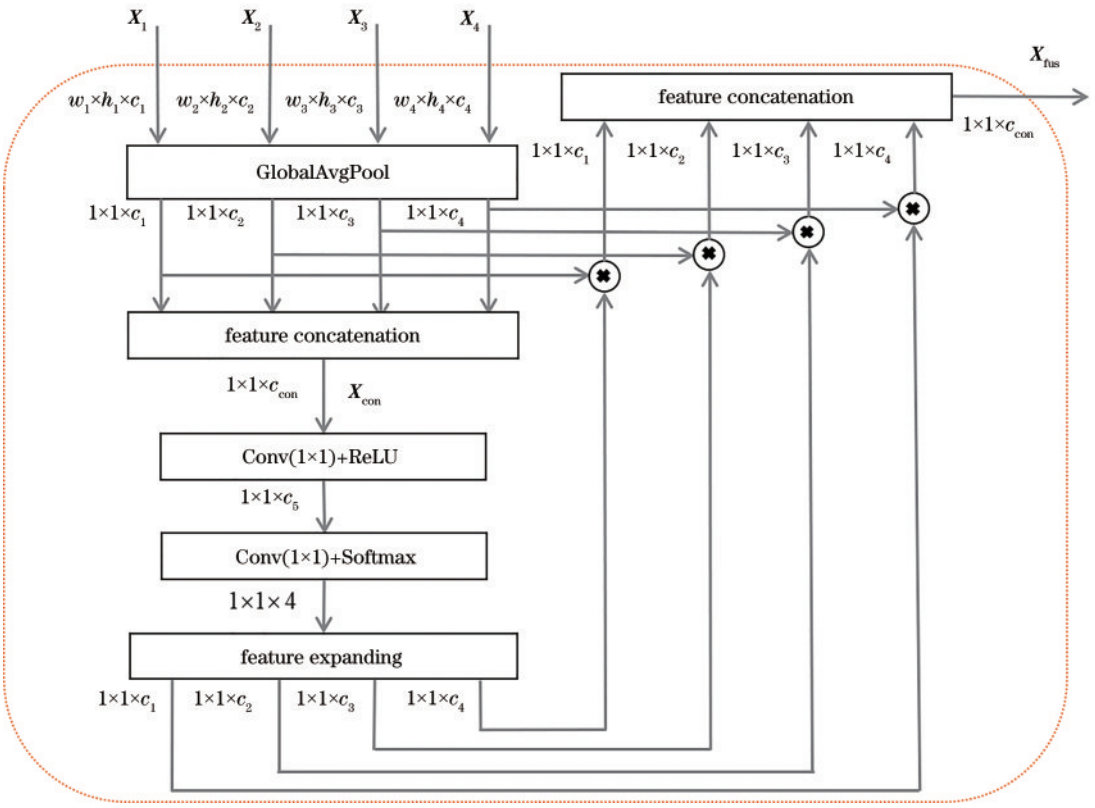

图 3　多尺度注意力融合模块

Fig. 3　Multi-scale attention fusion module

的，然后再组合不同分辨率的摄像头和光照等条件。对于假脸来说，在真实人脸收集条件的基础上新增了很多获取方式，比如纸张打印、视频回放、纸张切割、人脸面具等。因此相对于真脸来说，对齐不同领域的假脸特征会更加困难，如果对真实人脸和虚假人脸采用相同的对齐方式，会对模型的分类精度产生影响[9]。文献[9]中的方法分离不同领域的虚假人脸，同时对齐不同领域的真实人脸。然而由于虚假人脸的攻击方式层出不穷，特别在小型数据集上训练时模型接触的攻击方式较少，当在大型测试集上进行测试时由于攻击方式的多样性，目标域的虚假人脸的分布远离源域中虚假人脸的特征空间，转而靠近甚至跨过模型的分类边界，从而影响分类精度。

鉴于此，在保持真脸和假脸聚集度相对差异的基础上，将目光聚焦于真实人脸的对齐，通过真实人脸类别对抗机制进一步聚集不同领域的真实人脸。如图 2 所示，特征判别器在判断各个真脸特征属于何种数据集的基础上，还要判断该特征属于数据集中的何种类别，从更细粒度的角度进一步聚集不同领域的真实人脸。特征生成器与特征判别器进行对抗，从而学习到的是领域无关且类别无关的真实人脸特征。对于虚假人脸，不进行这种对抗学习。

假设现在有 N 个源域，每个源域中的真实人脸类别为 c_{i_1}，其中 $1\leqslant i_1\leqslant N$。总的类别数为 $N_{\text{total}}=\sum_{i_1=1}^{N}c_{i_1}$。每个源域中都包含真实人脸 $\boldsymbol{X}_{\text{r}}$ 和虚假人脸 $\boldsymbol{X}_{\text{f}}$，如图 2 所示，不同源域的真实人脸和虚假人脸通过 Resnet18 主干网络，其中四个残差块的输出作为输入通过多尺度注意力融合模块，得到融合特征。该特征由真实人脸和虚假人脸各自融合后的特征 $\boldsymbol{X}_{\text{r}}'$ 和 $\boldsymbol{X}_{\text{f}}'$ 组成，二者计算式为

$$\boldsymbol{X}_{\text{r}}'=G(\boldsymbol{X}_{\text{r}}),\boldsymbol{X}_{\text{f}}'=G(\boldsymbol{X}_{\text{f}}),\tag{5}$$

式中：G 为特征生成器，由 Resnet18 主干网络和多尺度注意力融合模块组成。

如图 2 的虚线分支所示，融合特征中只有真脸融合特征 $\boldsymbol{X}_{\text{r}}'$ 进入判别器分支。判别器 D 需要判断 $\boldsymbol{X}_{\text{r}}'$ 的领域标签和类别标签，特征生成器 G 需要欺骗特征判别器 D，通过对抗训练生成器学习到的真实人脸特征是领域无关且类别无关的。在训练的过程中，通过最大化判别器的损失来优化生成器，通过最小化判别器的损失来优化判别器。由于实验中包含多个源域以及每个源域中包含多个真实人脸类别，因此使用交叉熵损失来优化生成器和判别器，优化计算式为

$$\min_{D}\max_{G}L_{\text{adv}}(G,D)=$$
$$-E_{\boldsymbol{x},\boldsymbol{y}\sim\boldsymbol{X}_{\text{r}},\boldsymbol{Y}_{\text{D}}}\sum_{n=1}^{N_{\text{total}}}\boldsymbol{1}_{[n=y]}\log D\left[G(\boldsymbol{x})\right],\tag{6}$$

式中：$\boldsymbol{X}_{\text{r}}$ 表示训练集中真实人脸的集合；$\boldsymbol{Y}_{\text{D}}$ 表示 $\boldsymbol{X}_{\text{r}}$ 对应的领域标签和类别标签的组合标签集合，集合中的元素个数对应图 2 中 FC_2 的输出节点数 V；$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别表示当前真实人脸样本及其组合标签；n 用于遍历组

合标签集合；$\mathbf{1}_{[n=y]}$为1时表示模型正确预测当前真实人脸样本的组合标签，为0时则表示预测错误。

为了同时优化生成器和判别器，沿用文献[9]中的梯度反转层(GRL)，如图2中GRL所示。通过真脸类别对抗的学习，能够获取到一个真实人脸的泛化空间，在该泛化空间的基础上，可以继续进行分类任务的学习。

2.3 三元组损失

针对不同领域的虚假人脸特征难以对齐的问题，文献[9]采用的是不聚集转而发散的方法。具体来说，给不同源域的假脸设置不同的标签，利用三元组损失进行不同源域假脸的发散；同时给不同源域的真脸设置相同的标签，利用三元组损失进行不同源域真脸的聚集。然而这种方式在训练集数据小而测试集数据大的情况下会导致分类精度的下降。原因在于假脸攻击方式本就众多，如果在训练的时候使用的是小型数据集，而测试的时候使用大型数据集，训练集和测试集中攻击方式数目的差距会导致源域的假脸特征和目标域的假脸特征难以对齐，而该方法又对各个源域的假脸进行发散，这会导致目标域中假脸分布被迫远离源域的假脸特征空间，从而靠近甚至跨过模型的分类边界，造成分类精度下降。

鉴于此，本文沿用三元组损失，不同于文献[9]将不同源域的假脸看成不同的类别，本文每轮训练中只随机输入一个源域的假脸，从而避免了针对不同源域的假脸是聚集还是发散的选择问题，因为无论是聚集还是发散都会导致分类精度的下降，同时也能将网络的注意力更多聚焦于不同源域的真脸。每轮训练都会输入不同源域的真脸，从而能够聚集不同源域的真脸，生成更加紧凑的真脸特征空间，学习到泛化能力更强的分类边界。

具体来说，假设有三个源域，在每轮训练的时候都从三个源域中抽取真实人脸输入网络，同时只随机抽取其中一个源域的虚假人脸输入网络。如图4所示，三角形、四边形、五边形分别表示源域1、源域2、源域3的真实人脸。类圆形表示三个源域中任意一个的虚假人脸，这里选择源域3进行演示。图4中边框加粗的图形表示是三元组中的锚点(anchor)，与其相连的图形是正样本(positive)和负样本(negative)，分别表示锚点的同类和异类。三元组损失能在拉近锚点和正样本的同时分开锚点和负样本，所提方法将三个源域的真实人脸看成一类同时将源域3中的虚假人脸看成另一类。结合三元组损失，可以达到三种目标：对齐领域内和领域间的真脸；对齐领域内的假脸；分离真脸和假脸。

如图4所示，不同领域的真脸聚集到一起，但是并没有混合在一起。三元组损失结合真实人脸类别对抗机制，能够得到如图1右图所示更加聚集的真实人脸空间，即各个领域的真实人脸混合在一起。不同领域的假脸则只进行各自领域内的聚集，而没有进行领域间的聚集，这样有助于提升模型的泛化能力。三元组损失用来优化生成器，计算式为

$$\min_{G} L_{\mathrm{tri}}(G)=\sum_{\boldsymbol{x}_i^{\mathrm{a}},\boldsymbol{x}_i^{\mathrm{p}},\boldsymbol{x}_i^{\mathrm{n}}}\left[\left\|f\left(\boldsymbol{x}_i^{\mathrm{a}}\right)-f\left(\boldsymbol{x}_i^{\mathrm{p}}\right)\right\|_2^2-\left\|f\left(\boldsymbol{x}_i^{\mathrm{a}}\right)-f\left(\boldsymbol{x}_i^{\mathrm{n}}\right)\right\|_2^2+\alpha\right]_+ \tag{7}$$

式中：$\boldsymbol{x}_i^{\mathrm{a}}$表示随机选择的样本，$\boldsymbol{x}_i^{\mathrm{p}}$表示与$\boldsymbol{x}_i^{\mathrm{a}}$领域标签相同的样本，$\boldsymbol{x}_i^{\mathrm{n}}$表示与$\boldsymbol{x}_i^{\mathrm{a}}$领域标签不同的样本；$\alpha$是预定义的阈值参数。

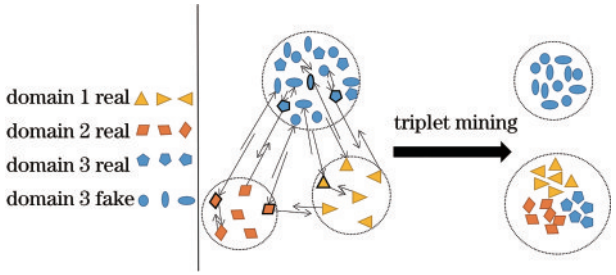

图4 三元组损失图

Fig. 4 Illustration of triplet loss

2.4 总体损失

人脸活体检测的目标是判断输入的图像是真实人脸还是虚假人脸，因此在生成器之后添加一个分类器进行真假人脸的判定。使用交叉熵损失进行生成器和分类器的优化，因此整个网络的总体损失为

$$L_{\mathrm{total}}=L_{\mathrm{cls}}+\lambda_1 L_{\mathrm{tri}}+\lambda_2 L_{\mathrm{adv}}, \tag{8}$$

式中：λ_1和λ_2为平衡参数。

3 实验结果与分析

数据集选择。为了验证所提模型的有效性，在4个公开人脸活体检测数据集上进行对比实验。数据集包括OULU-NPU、CASIA-FASD、Idiap Replay-Attack、MSU-MFSD，分别简称为O、C、I、M。主要实验有4组，每组以其中一个数据集为测试集，另外三个数据集为训练集，得到的组合为O&C&I to M、I&C&M to O、O&C&M to I、O&M&I to C。4个数据集的人物、视频、真假脸类别数目不尽相同，详细数据如表1所示。

实验条件。操作系统为Ubuntu18.04，深度学习框架为PyTorch 1.2.0，CPU为i5 10400 F，内存为16 GB，GPU为NVIDIA GeForce GTX 1080Ti。

网络结构。主干网络采用Resnet18，保留其前17层，第17层使用全局平均池化代替原来的平均池化，随后添加2个全连接层作为分类器，其输入和输出维度分别为(512,512)和(512,2)。多尺度注意力融合模块的输入来自4个残差块的输出向量，该模块的输出经过降维得到仅剩下batchsize和channel两个维度的特征向

表 1 四个数据集的详细信息

Table 1 Details of four datasets

Dataset	Number of identities	Number of videos	Number of real categories	Number of fake categories
MSU-MFSD	35	280	2	6
CASIA-FASD	50	600	3	9
Idiap Replay-Attack	50	1200	4	20
OULU-NPU	55	4950	18	72

量，该向量经过分类器，得到最终的真假分数。判别器为一个全连接层，输入的维度为 512，输出的维度为当前实验中的领域标签和真脸类别标签的组合数。

实现细节。利用 MTCNN 人脸检测算法对原始视频帧进行人脸区域的截取，人脸大小为 256×256×3。为了平衡各个数据集之间的容量差异，对于 O、I、C、M 的每个视频，分别随机抽取 1、5、6、12 帧用于训练。根据收集方式，将 O、I、C、M 中真实人脸类别数分别设置为 6、4、3、2。由于每轮训练时只随机选取一个源域的假脸并且选取所有源域的真脸，考虑到真假脸二分类任务的数据平衡，将 O&C&I to M、I&C&M to O、O&C&M to I、O&M&I to C 四组实验的 batchsize 分别设置为 13、9、11、12，batchsize 表示假脸和真脸的 batchsize 比值，将四组实验中真脸的 batchsize 统一设置为 5。采用随机梯度下降算法优化模型，初始学习率为 0.01，衰减系数设置为 5×10^{-4}，总共训练 400 轮。α 设置为 0.1，四组实验中的 λ_1 和 λ_2 的组合分别设置为(1.0,0.5)、(1.0,0.4)、(1.0,0.5)、(1.0,0.5)，方便与基准方法进行对比。c_1、c_2、c_3、c_4、c_5 分别为 64、128、256、512、60，同时沿用了文献[9]中的归一化方法，如图 2 的 L2 normalization 所示。

评价指标。与文献[9]中的方法相同，使用半错误率(HTER)和曲线下面积(AUC)作为评价指标，同时使用 t-SNE[21]可视化特征分布。

3.1 消融实验

为了验证所提方法中各个组件的有效性，在 4 组实验中均进行了消融实验，分别移除多尺度注意力机制模块、真实人脸类别对抗、三元组损失中的其中一个，结果如表 2 所示。可以看出：任何一个模块的缺失都会导致网络性能的降低，表明了所提方法中的各个组件的有效性，各个组件组合在一起能够取得最佳性能。

表 2 所提方法不同组件的评估结果

Table 2 Evaluation results of different components of the proposed method

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%
w /o att	5.27	95.20	12.11	94.31	10.07	95.99	13.54	93.22
w /o ad	5.71	96.10	10.77	94.51	14.92	93.05	12.30	94.97
w /o tri	7.14	96.53	10.66	95.19	21.42	80.51	22.06	86.37
all	4.52	97.24	9.88	95.46	9.21	96.97	11.45	95.32

3.2 基准方法比较与可视化

基准方法中的判别器仅仅判别当前真实人脸的领域归属。所提方法在此基础上新增真实人脸的类别归属，能够进一步聚集不同领域的真实人脸，使得整个真实人脸特征空间更加紧凑，有利于形成泛化能力更好的分类边界。基准方法使用残差网络最后一个残差块的输出特征进行多个任务的学习，这可能影响各个任务的平衡。所提方法使用多尺度注意力机制将残差网络的多个残差块的输出特征输入一个全新的子网络，得到加权权重，再利用该权重对多个残差块的输出特征进行加权融合，得到融合特征，利用该特征能够更好地平衡各个任务的学习。基准方法使用非对称三元组损失将来自所有源域的真实人脸看成同一类，不同源域的虚假人脸看成不同的类别，从而使得不同源域的虚假人脸的特征分布各自发散，然而这种发散机制会导致分类精度下降。针对该问题，所提方法在每轮训练中只随机输入一个源域的假脸，从而避免了对不同源域的虚假人脸是发散还是聚集的选择，能够让网络将注意力集中在真实人脸的聚集上。

为了从评价指标的角度验证所提方法的有效性，在 4 个数据集上与基准方法 SSDG 进行对比实验。对比实验总共有 4 组，每组以 O、C、I、M 四个数据集中的一个数据集为测试集，另外三个数据集为训练集，得到的组合为 O&C&I to M、I&C&M to O、O&C&M to I、O&M&I to C。实验结果如表 3 所示，所提方法在 4 组实验中相比基准方法错误率更低且鲁棒性更强，特别是在 I&C&M to O 实验中。该组实验是 4 组实验中最难的一组，因为测试数据集 O 的样本容量接近 I、C、M 三个训练数据集容量的 2 倍，该实验中的模型在测试时需要面临更多的未知性，因此基准方法表现较差。如表 3 所示，所提方法在 I&C&M to O 实验中性能提升明显，HTER 降低了 26.6%，AUC 提升了 4.1%。

表 3　所提方法与基准方法的比较结果

Table 3　Comparison results between the proposed method and the corresponding baseline method

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%
SSDG	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
Proposed method	4.52	97.24	9.88	95.46	9.21	96.97	11.45	95.32

为了从可视化的角度验证所提方法的有效性，分别在基准方法 SSDG 和所提方法中的 I&C&M to O 实验中从 4 个数据集中的真脸和假脸均随机抽取 300 个样本进行 t-SNE 可视化。图 5 中 src1_real 和 src1_fake 表示训练数据集 I 中的真脸和假脸样本，tgt_real 和 tgt_fake 表示测试数据集 O 中的真脸和假脸样本。如图 5 左图所示，由于基准方法选择对不同领域的假脸进行发散，当模型在测试集上进行测试时，虚假人脸被迫靠近甚至跨过分类边界，导致分类精度下降。如图 5 右图所示，所提方法选择对真实人脸进行进一步的聚集，能够得到一个更紧凑的真实人脸空间，不同领域的真脸混合在一起，表明模型学习到的真实人脸领域无关性更强；对不同领域的虚假人脸没有进行发散，从而避免了基准方法中的问题，同时保持真脸和假脸聚集度的相对差异，能够得到一个泛化能力更强的分类边界。

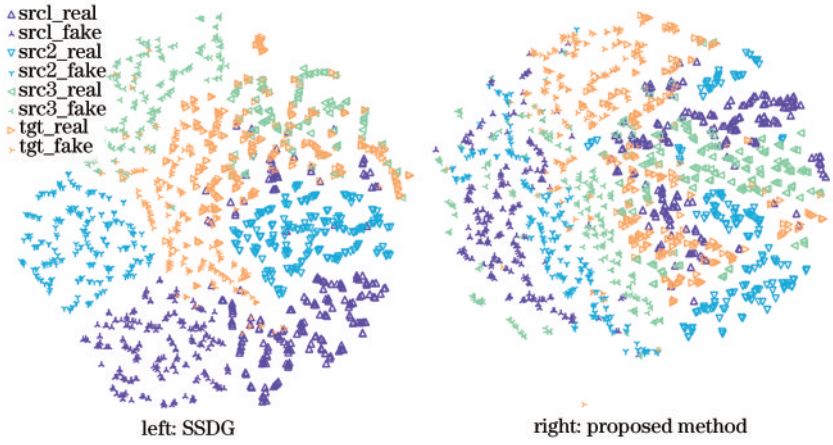

图 5　I&C&M to O 实验中对基准方法和所提方法所得分类特征的 t-SNE 可视化

Fig. 5　t-SNE visualization for the classification features obtained by baseline method and proposed method under the I&C&M to O testing task

为了验证需要区分对待真脸和假脸这一思路的正确性，利用所提方法得到的模型结合 Grad-CAM[22] 进行模型关注区域的演示。如图 6 所示，对于第一行中不同数据集的真实人脸，模型关注的区域大都是整个面部区域，而对于第二行中相对应的虚假人脸，由于攻击类型的多样性，模型关注的区域不尽相同，这验证了要将真实人脸和虚假人脸区分对待的结论，即不能采用相同的聚集方式。

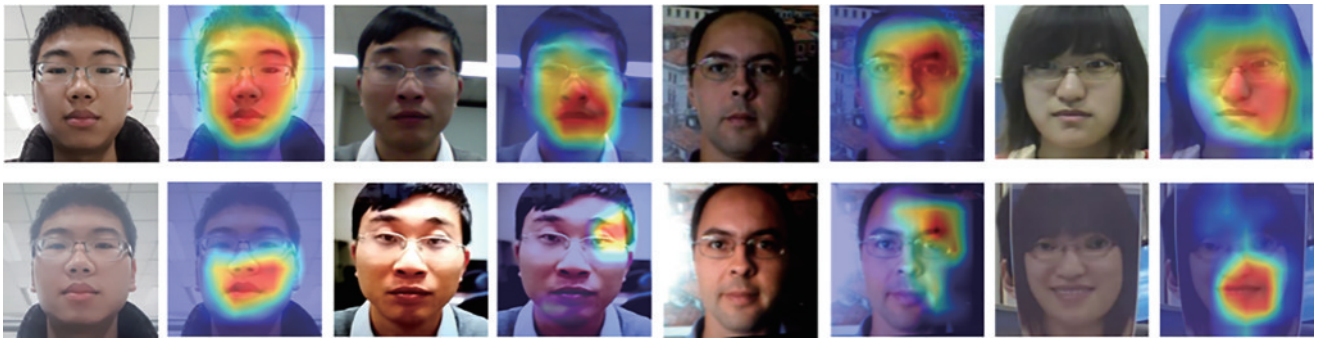

图 6　I&C&M to O 实验中的 Grad-CAM 可视化

Fig. 6　Grad-CAM visualization under the I&C&M to O testing task

3.3　有限源域比较

在有限源域情况下对所提方法进行评估。具体来说，选择 M 和 I 作为源域数据集，剩下的 O 和 C 分别作为目标域数据集。如表 4 所示，所提方法的指标能够

接近或者超过其他方法。尽管训练时只用到两个源域数据集，所提方法依旧能够充分聚集两个源域中的真实人脸，从而得到泛化能力更强的分类边界。

表 4 有限源域情况下不同领域泛化方法的比较结果

Table 4 Comparison result of different domain generalization methods in the case of limited source domains

Method	M&I to C		M&I to O	
	HTER /%	AUC /%	HTER /%	AUC /%
MS-LBP[23]	51.16	52.09	43.63	58.07
IDA[4]	45.16	58.80	54.52	42.17
LBP-TOP[24]	45.27	54.88	47.26	50.21
MADDG[10]	41.02	64.33	39.35	65.10
SSDG-M[9]	31.89	71.29	36.01	66.88
DRDG[11]	31.28	71.50	33.35	69.14
DASN[12]	21.48	83.41	21.74	80.87
Proposed method	17.88	89.91	22.70	86.17

3.4 不同主干网络比较

为了验证不同主干网络对所提方法的影响，分别使用 Resnet18、Resnet34、Resnet50 三个网络作为主干网络进行 4 组实验。如表 5 所示，其中 Avg HTER 和 Avg AUC 分别表示 4 组实验中 HTER 和 AUC 的平均值；Speed 表示测试环境中 GPU 为 NVIDIA GeForce GTX 1080Ti，模型输入尺寸为 256 × 256 × 3 时整个网络的前向推理速度；FLOPs 和 Params 分别表示整个模型的计算量和参数总数。可以看出：主干网络为 Resnet18 时的 5 个指标均优于其他两个主干网络，这是因为 Resnet34 和 Resnet50 相对于 Resnet18 含有更多的网络参数，不仅计算量增加，而且容易过拟合。与此同时，相信所提方法在更有效的主干网络上能取得更佳的表现。

表 5 不同主干网络的比较结果

Table 5 Comparison result of different backbone networks

Backbone	FLOPs /10^9	Params /10^6	Speed /(frame·s^{-1})	Avg HTER /%	Avg AUC /%
Resnet50	5.37	25.60	46.57	12.05	93.51
Resnet34	4.79	21.81	52.65	12.31	93.87
Resnet18	2.37	11.70	100.75	8.76	96.29

3.5 与其他方法比较

如表 6 所示，所提方法在 4 组实验中的表现均要优于其他方法，表明了所提方法的有效性和先进性。IDA[4]、MS-LBP[23]、Binary CNN[25]、Color Texture[26]、LBP-TOP[24]、Auxiliary[27]等方法没有考虑不同领域之间的内在联系，因此在进行跨数据集测试时性能大幅下降。文献[10-12]中的方法对真脸和假脸采用相同的处理方式，模型学习到的泛化特征空间中真实人脸和虚假人脸的分布各自都是紧凑的。文献[9]中的方法通过对比实验证明不能对真实人脸和虚假人脸采用相同对齐方式的结论，因此其选择对不同领域的假脸进行发散，对不同领域的真脸进行聚集，然而当训练集较小、测试集较大时会导致测试集的虚假人脸靠近甚至跨过分类边界，从而影响分类精度。考虑到上述的人脸活体检测算法均为二分类任务，因此在文献[9]的基础上，所提方法分别提升真实人脸和虚假人脸的分布聚集度，这样真实人脸的聚集度依旧是大于虚假人脸的。如表 3 所示，所提方法在 4 组实验中的表现均有

表 6 所提方法与使用领域泛化进行人脸活体检测的其他方法的比较结果

Table 6 Comparison result between the proposed method and other methods for domain generalization on face anti-spoofing

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%	HTER /%	AUC /%
MS-LBP[23]	29.76	78.50	54.28	44.98	50.30	51.64	50.29	49.31
Binary CNN[25]	29.25	82.87	34.88	71.94	34.47	65.88	29.61	77.54
IDA[4]	66.67	27.86	55.17	39.05	28.35	78.25	54.20	44.59
Color Texture[26]	28.09	78.47	30.58	76.89	40.40	62.78	63.59	32.71
LBP-TOP[24]	36.90	70.80	33.52	73.15	29.14	71.69	30.17	77.61
Auxiliary(Depth)	22.72	85.88	33.52	73.15	29.14	71.69	30.17	77.61
Auxiliary[27]	—	—	28.40	—	27.60	—	—	—
MADDG[10]	17.69	88.06	24.50	84.51	22.19	84.99	27.89	80.02
SSDG[9]	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
DRDG[11]	12.43	95.81	19.05	88.79	15.56	91.79	15.63	91.75
DASN[12]	8.33	96.31	12.04	95.33	13.38	86.63	11.77	94.65
Proposed method	4.52	97.24	9.88	95.46	9.21	96.97	11.45	95.32

提升，特别是在I&C&M to O实验中，该组实验的训练集大小要远小于测试集大小。由于实际应用中所能获取的数据集是有限的，如果在有限的数据集上进行训练的模型能够在从未接触的更大数据集中取得优异的表现，就更符合实际需求，进一步说明所提方法的实用性。

4 结　论

针对人脸活体检测中跨数据集泛化能力较差的问题，构建基于真实人脸类别对抗机制的人脸活体检测网络。该网络利用三元组损失对各个源域中的假脸只进行领域内的聚集，而没有进行领域间的发散或聚集；同时对各个源域中的真实人脸的类别进行细分，通过真实人脸类别对抗机制和三元组损失进行领域内和领域间的对齐，学习到一个更紧凑的真实人脸特征空间，同时能够保持真脸和假脸特征空间聚集度的相对差异。通过多尺度注意力融合机制，所提方法能够更好地平衡分类任务和领域不变性任务，得到泛化能力更强的分类边界。所提方法在4组实验中的表现均优于其他方法，表明所提方法的有效性和先进性。下一步将通过生成对抗网络来扩充数据集以达到更好的泛化效果。

参 考 文 献

[1] 孔月萍，刘霞，谢心谦，等. 基于梯度方向直方图的人脸活体检测方法[J]. 激光与光电子学进展，2018，55(3)：031009.
Kong Y P, Liu X, Xie X Q, et al. Face liveness detection method based on histogram of oriented gradient [J]. Laser & Optoelectronics Progress, 2018, 55(3): 031009.

[2] Boulkenafet Z, Komulainen J, Hadid A. Face antispoofing using speeded-up robust features and fisher vector encoding[J]. IEEE Signal Processing Letters, 2017, 24(2): 141-145.

[3] Galbally J, Marcel S, Fierrez J. Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition[J]. IEEE Transactions on Image Processing, 2014, 23(2): 710-724.

[4] 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.

[5] Atoum Y, Liu Y J, Jourabloo A, et al. Face anti-spoofing using patch and depth-based CNNs[C]//2017 IEEE International Joint Conference on Biometrics, October 1-4, 2017, Denver, CO, USA. New York: IEEE Press, 2017: 319-328.

[6] Liu S Q, Yuen P C, Zhang S P, et al. 3D mask face anti-spoofing with remote photoplethysmography[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9911: 85-100.

[7] Liu S Q, Lan X Y, Yuen P C. Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11220: 577-594.

[8] Torralba A, Efros A A. Unbiased look at dataset bias [C]//2011 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 1521-1528.

[9] Jia Y P, Zhang J F, Shan S G, et al. Single-side domain generalization for face anti-spoofing[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 8481-8490.

[10] Shao R, Lan X Y, Li J W, et al. Multi-adversarial discriminative deep domain generalization for face presentation attack detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 10015-10023.

[11] Liu S B, Zhang K Y, Yao T P, et al. Dual reweighting domain generalization for face presentation attack detection[EB/OL]. (2021-06-30) [2021-12-25]. https://arxiv.org/abs/2106.16128.

[12] Kim T, Kim Y. Suppressing spoof-irrelevant factors for domain-agnostic face anti-spoofing[J]. IEEE Access, 2021, 9: 86966-86974.

[13] Akuzawa K, Iwasawa Y, Matsuo Y. Adversarial invariant feature learning with accuracy constraint for domain generalization[EB/OL]. (2019-04-29) [2021-12-25]. https://arxiv.org/abs/1904.12543.

[14] Xie Q Z, Dai Z H, Du Y L, et al. Controllable invariance through adversarial feature learning[EB/OL]. (2017-03-31)[2021-12-25]. https://arxiv.org/abs/1705.11122.

[15] 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.

[16] 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.

[17] Shi L, Zhou Z, Guo Z H. Face anti-spoofing using spatial pyramid pooling[C]//2020 25th International Conference on Pattern Recognition (ICPR), January 10-15, 2021, Milan, Italy. New York: IEEE Press, 2021: 2126-2133.

[18] Jia S, Li X, Hu C B, et al. 3D face anti-spoofing with factorized bilinear coding[EB/OL]. (2020-12-13) [2021-12-25]. https://arxiv.org/abs/2005.06514.

[19] Xie J H, Luo C, Zhu X P, et al. Online refinement of low-level feature based activation map for weakly supervised object localization[EB/OL]. (2021-10-12) [2021-12-25]. https://arxiv.org/abs/2110.05741.

[20] Chen B L, Yang W H, Wang S Q. Generalized face antispoofing by learning to fuse features from high- and

low-frequency domains[J]. IEEE MultiMedia, 2021, 28(1): 56-64.

[21] van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9: 2579-2625.

[22] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision, October 22-19, 2017, Venice, Italy. New York: IEEE Press, 2017: 618-626.

[23] Määttä J, Hadid A, Pietikäinen M. Face spoofing detection from single images using micro-texture analysis [C]//2011 International Joint Conference on Biometrics (IJCB), October 11-13, 2011, Washington, DC, USA. New York: IEEE Press, 2011.

[24] de Freitas Pereira T, Komulainen J, Anjos A, et al. Face liveness detection using dynamic texture[J]. EURASIP Journal on Image and Video Processing, 2014, 2014(1): 2.

[25] Yang J W, Lei Z, Li S Z. Learn convolutional neural network for face anti-spoofing[EB/OL]. (2014-08-24) [2021-12-25]. https://arxiv.org/abs/1408.5601.

[26] 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.

[27] 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, June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 389-398.