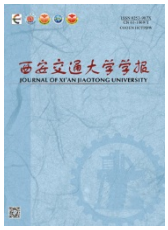


西安交通大学学报
Journal of Xi'an Jiaotong University
ISSN 0253-987X,CN 61-1069/T

《西安交通大学学报》网络首发论文

题目：采用超复数小波生成对抗网络的活体人脸检测算法
作者：李策，李兰，宣树星，杨静，杜少毅
收稿日期：2020-09-28
网络首发日期：2020-12-15
引用格式：李策，李兰，宣树星，杨静，杜少毅．采用超复数小波生成对抗网络的活体人脸检测算法．西安交通大学学报.
https://kns.cnki.net/kcms/detail/61.1069.T.20201215.0923.002.html

西 安 交 通 大 学 学 报
JOURNAL OF XI'AN JIAOTONG UNIVERSITY

采用超复数小波生成对抗网络的活体人脸检测算法

李策[1]，李兰[1]，宣树星[1]，杨静[2]，杜少毅[3]
（1. 兰州理工大学电气工程与信息工程学院，730050，兰州；2. 西安交通大学自动化科学与工程学院，710049，西安；
3. 西安交通大学人工智能学院人工智能与机器人研究所，710049，西安）

摘要：为了提升人脸识别系统判别图像真实性的能力，针对较难检测到未知的活体人脸攻击问题，提出了一种采用超复数小波生成对抗网络的活体人脸检测算法。采用了四个不同类型的数据集，随机选择三个数据集作为训练集，另一个则为测试集，即训练中未知的活体人脸。首先，训练集视为三个源域，输入到超复数小波生成对抗网络中，使一个特征生成器与三个判别器进行对抗，当特征生成器成功欺骗过三个判别器时，形成了具有三个源域共享且区别于三个源域的特征空间，能够检测到不同于源域的人脸特征。同时，在判别器上设置了域间和域内的三元组约束函数来提高判别器的性能，并且将超复数小波的细节子带图与卷积网络联合学习图像多个方向的细节纹理特征，用来提升判别器鉴定活性人脸特征的能力。然后，由于真/假人脸的远程光电体积描记术和深度图都具有较大的差异，所以将其嵌入到特征空间中，增强生成特征空间检测人脸特征的泛化性能，形成通用的特征空间。最后，在该特征空间中使用测试集进行判别分类得到真/假的结果。结果表明，在CASIA-FASD、Idiap Replay-Attack和NUAA数据集上，*C* (area under curve, AUC)分别为84.65、86.06、91.21，*H*(half total error rate, HTER)分别为24.05、21.05、15.01，均高于对比算法。

关键词：活体人脸检测；超复数小波；生成对抗网络
中图分类号：TP751.1　　**文献标识码：**A
DOI：

Face Anti-Spoofing Algorithm Using Generative Adversarial Networks with Hypercomplex Wavelet

LI Ce[1], LI Lan[1], XUAN Shuxing[1], YANG Jing[2], DU Shaoyi[3]
(1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, China; 2. College of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, 710049, China; 3. Institute of Artificial Intelligence and Robotics, College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an, 710049, China)

Abstract: In order to improve the ability of the face recognition system to determine the face anti-spoofing, aiming at the problem that the existing anti-spoofing detection algorithms are difficult in unknown domain attacks, this paper proposes a face anti-spoofing detection algorithm using GAN with hypercomplex wavelet transform. Four different types of datasets are used, three datasets are randomly selected as the training datasets, and the other dataset is used as the test datasets, that is, the face anti-spoofing that is unknown during training. First of all, training datasets are regarded as three source domain datasets, which are input into the generation network to make a feature generator against three discriminators. When that successfully deceives three discriminators, a feature space with three source domains shared features and different from three domains is formed, which can detect the

收稿日期：2020-09-28。　作者简介：李策（1974-），男，教授；杜少毅（通信作者），男，教授。　基金项目：国家自然科学基金资助项目（61971343，61866022）；陕西省重点研发计划高校联合项目(2020GXLH-Y-008)。
网络出版时间：　　网络出版地址：

characteristics of unknown domain data. The triple constraint function inter-class and intra-class are set to improve the performance of the discriminator, and the detail subbands of the hypercomplex wavelet transform and convolution network are combined to learn the detail features of multiple directions of images. Then, the depth map and remote photo plethysmography signal difference are embedded in the feature space to enhance the generalization performance of the generated feature space for the living face features. Finally, use the test dataset for discriminative classification in the feature space to obtain live/fake results. The results show that on the CASIA-FASD, Idiap Replay-Attack and NUAA datasets, the AUC is 84.65, 86.06, 91.21, and the HTER is 24.05, 21.05, 15.01, which are higher than the comparison algorithm.

Keywords: face anti-spoofing detection; hypercomplex wavelet; generative adversaria networks;

随着图像篡改技术的不断提升以及人脸识别系统的自动化和无人监督化的发展[1-2]，近年来篡改者很容易通过照片和视频的手段复制人脸，从而达到攻击人脸识别系统的目的。然而，在智能技术迅速发展的现在，人脸识别系统已在考勤/门禁、安防和金融等领域得到了广泛的应用，因而人们在日常生活中已将“刷脸”办事变成了一种流行的方式。例如，车站安检、小区门禁锁/手机人脸解锁、支付宝刷脸支付和自动取款机刷脸取款等。所以，恶意伪造合法用户的人脸来攻击人脸识别系统，会使得人脸身份认证系统受到严重的威胁[3]。对于人脸的篡改攻击会使得人脸识别系统的安全性无法得到保证，更会让使用者的隐私暴露到大众面前，甚至财产可能会被不法分子盗取。刷脸时代的到来，在给人们生活带来方便的同时，同样也存在着极大的安全隐患。因此，辨别人脸图像的真假性成为一个社会关注的问题，用于检测人脸是否被篡改的活体检测技术也已成为计算机视觉领域的研究热点之一。

活体人脸检测技术主要是利用有生命个体的活体人脸所具有的深度信息、运动信息和纹理信息，与假人脸进行区别，来有效辨别人脸的真假性。常见的伪造技术有打印照片、视频回放和面具攻击三种，如图 2 所示。真实人脸与三种伪造假人脸有着很大的区别：真实人脸是通过摄像头首次拍摄活的人脸来获取，而假的人脸是通过摄像头二次甚至多次采集首次拍摄的照片获得的，多次采集照片会使得图像原来的纹理信息、三维结构和运动信息发生变化。

近年来，针对身份认证的活体人脸检测技术，国内外相关研究团队主要针对打印照片和视频攻击等问题展开研究，并取得了一定的成果。传统的方法一般将人脸活体检测视为一个二分类问题，利用真/假人脸的区别进行分类。Bharadwaj 等人首先在连续帧上通过运动放大面部微小的动作进行预处理图像，然后使用局部二值模式(local binary patterns, LBP)提取图像的特征，最后采用支持向量机(support vector machine, SVM)分类获取结果[4]。后来，其在文献[5]对算法又进行了改进，在进过预处理步骤后，提取了 LBP 和方向光流直方图特征分析图像的特征，不仅获得了区分真/假人脸的纹理特征，而且提供了真实人脸的活性特征，在一定程度提升了算法性能，但是预处理引入的噪声对整体结果造成了一定的影响。Tirunagari 人则首先对图像进行动态模式分解得到最大运动能量的子空间图，然后再进行纹理分析；但是基于面部动作的算法对于打印照片抖动和视频攻击效果不好[6]。Li 等人引入了测量活体人脸心率的方法，通过远程光电体积描记术(Remote Photo Plethysmography, rPPG)来预测心率，由于照片人脸和活体人脸提取的心率信号分布不同，而视频人脸和活体人脸的心率信号近似，所以设计了一个级联的活体检测算法，首先使用脉冲在频域上的不同分布来区分照片人脸攻击，若判断为活体则使用 LBP 纹理分析判断是否为视频攻击，但是心率信号的鲁棒性很低，所以脉冲特征的判别能力不够[7]。

随着深度学习的发展，相关算法在人脸活体检测领域受到了越来越多的关注。Xu 等模拟传统的提取纹理特性并分类的思路设计了一个 CNN-LSTM 算法检测活体人脸，有一定的提升效果[8]。由于视频和照片没有深度图信息，而活体人脸的鼻子、嘴巴和额头之间会有一定的深度，Atoum 等人首次将人脸的深度图作为判别活体与假人脸的判别特征，但是没有与最优的传统算法比较[9]。Liu 等人将二分类问题转换成了特征监督问题，将活体人脸的心率信号和深度图作为判别真假人脸的特性，并且巧妙的设计了一个非刚性注册层来对齐各帧人脸的非刚性运动，突破了传统的方法，取得了很好的效果[10]。Song 等人将活体检测放到人脸检测算法框架里面作为一个类，通过比较背景、真人脸和假人脸的置信度分数来鉴别是否为活体[11]。Zhang 等人提出了

一种轻量级的网络 FeatherNets，并设计了一种新的基于深度图和 IR 数据的多模态融合分类器来提升人脸图像的活体检测性能[12]。Yu 等人将人脸图像的活体检测问题转换为一种图像材质识别的问题，判别人脸的真假[13]。Yu 提出了一种中心差分卷积的方法，可以很好的提取到伪造图像的特征，并且不容易受到图像光照和型号的影响[14]。Jourabloo 等人将图像噪声引入到判别人脸活体检测的问题中，将活体检测问题转换为去欺骗问题[15]。Luo 等人使用卷积网络融合不同尺度的人脸图像特征，在单数据集上取得了一定的效果[16]。Shao 等人利用生成式对抗网络(generative adversarial networks, GAN)网络设计了一个多对抗性判别网络来生成一个具有活体人脸一般特性的生成器，对出现未知攻击的类型有一定的提升效果[17]。Liu 构造了一个含多种攻击类型的 SiW-M 数据集，训练了一个深度决策树网络，可以实现零样本下未知攻击类型数据的鉴别，但是涉及的超参数较多，且没有明确的决策阈值[18]。

综上所述，人脸活体检测算法是通过提取人脸面部纹理、三维结构和面部运动信息三类特征作为真/假人脸的判别依据。虽然这些算法已经取得了一定的结果，但是仍然存在以下缺陷:1) 难以鉴别未知的人脸攻击图像。2) 对于获取一般的真/假人脸特征存在一定的难度。所以，针对以上问题进行研究，本文提出了一种采用超复数小波生成对抗网络的活体人脸检测算法，解决训练阶段未出现的人脸图像检测

针对现有活体人脸检测算法很难检测到训练时未出现的跨域人脸数据特征问题，利用超复数小波(hypercomplex wavelet transform, HWT)[19]可以提取图像丰富的细节信息的特点，运用生成对抗网络来调整训练数据和测试数据之间的特征分布，使得源数据的训练模型可以适应目标数据的特性，提出了一种采用超复数小波生成对抗网络的活体人脸检测算法。

首先，将三种不同类型的数据集视为三个域的源数据，转换到 HSV 空间并分别输入到生成对抗网络中，训练一个特征生成器与三个判别器进行对抗，当特征生成器成功欺骗过三个判别器时，则形成了具有三个源域共享特征且区别于三个域的特征空间，用来预测未知域人脸图像的特征。同时，在判别器上设置了域间和域内的三元组约束函数来提高判别器的性能，并且将特征生成器网络的池化层连接了超复数小波变换的 12 幅子带图，增加人脸图像水平、垂直和对角方向的细节信息。然后，为了让生成的具有区别和共享三个源域的特征空间泛化到训练集中没有的未知目标域特征空间上，将活体人脸和伪造人脸都具有的深度图和 rPPG 信号差异嵌入到特征空间中，形成通用特征空间，提高特征空间判别人脸活性特征的能力。最后，使用源数据中未出现的测试数据进行测试，通过该特征空间进行判别分类，得到真/假人脸的判别结果，算法框架如图 1 所示。

1 本文所提算法

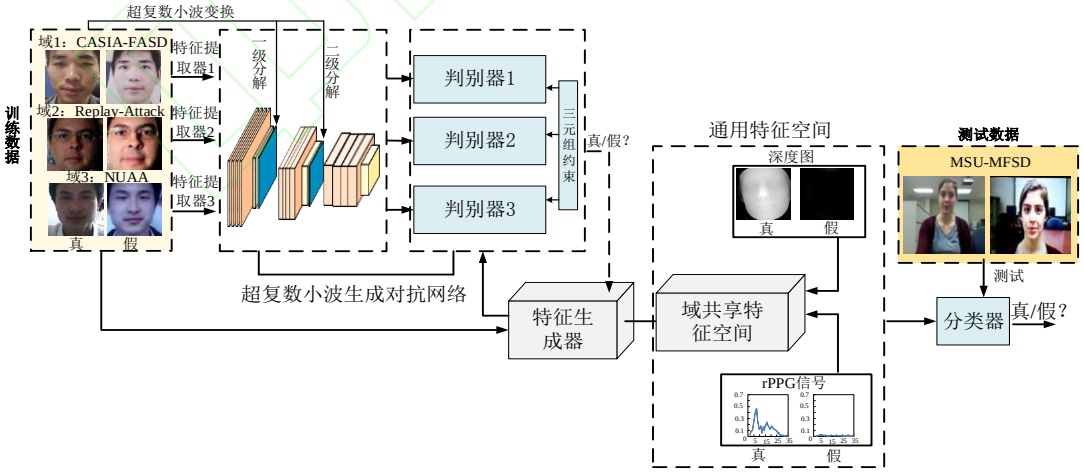


图 1　本文所提算法框架

1.1　通用特征空间

为了增强人脸活体检测算法对于未知的活体人脸图像的判别能力，需要学习到多个域人脸数据集所具有的通用特征空间，因此，利用 GAN 网络的特性，用生成器与多个具有源域特征的判别网络进行对抗，使其能够实现从可见的多源域数据分布中学习一个不可见的目标域数据分布的预测模型[17]。

下面将介绍具体实现过程。

假设有 N 个源域数据集，$X=\{X_1,X_2,...,X_N\}$，对应的标签为 $Y=\{Y_1,Y_2,...,Y_N\}$，每个标签有两种情况。当数据来自于活体人脸，则 $Y=1$，当数据来自于伪造人脸时则 $Y=0$。在获取到源数据集后，对于 N 个源域数据，先分别提取每个源域数据的判别性特征，具体为分别在不同域下使用二分类的交叉熵损失函数训练得到各自的模型，预训练的多个特征提取器表示为 $M_1,M_2,...,M_N$。以一个源域为例，描述二分类的交叉熵损失函数：

$$L_{cls}(X_1,Y_1;M_1,C_1)=-E_{(x_1,y_1)\sim(X_1,Y_1)}\sum_{k=1}^{K}[k=y_1]\log C_1(M_1(x_1)) \tag{1}$$

式中，(x_1,y_1) 从 (X_1,Y_1) 中取样，C_1 是模型分类器，因人脸活体检测中只有真/假两种情况，所以 $K=2$，$k\in(1,2,...,K)$ 为所属类的标签。

不同域下的判别特征空间会偏向于各自的源域数据特征，使得他们较难检测到源域中看不见的伪造人脸攻击，因此，利用生成器与不同域下的判别特征进行对抗，获取不同域下的通用特征空间，不会偏向于任何源域，使其具有活性人脸判别特征的一般性，从而能够很好的检测未知的目标域数据。接下来介绍具体实现原理与过程。

假设有 N 个源域数据，则 N 个特征提取器可以编码 N 个判别器，训练一个特征生成器，使其与 N 个判别器对抗，当特征生成器能够欺骗所有的判别器时，则特征生成器生成了一个 N 个源域共享的特征空间。所提算法中只有三个源域，则对上述生成对抗过程进行建模，目标函数可以用公式(2)描述：

$$L_{DG}(X,X_1,X_2,...,X_N;G,D_1,D_2,...,D_N)=\sum_{i=1}^{N}((E_{x\sim X}[\log(D_i(G(x)))])+E_{x_i\sim X_i}[1-\log(D_i(M_i(x_i)))]) \tag{2}$$

式中，G 表示特征生成器，它形成每个源域判别器都难以区分的通用特征空间，D_i 表示第 i 个域判别器，它将学习到的特征空间与第 i 个源域的特征空间区分开。通过在特征空间中进行多对抗性学习的过程后，可以由特征生成器 G 自动学习并生成通用的人脸活体特征空间。

1.2 超复数小波生成对抗网络

HWT 因其具有近似平移不变性、时频局部化和丰富的相位信息等特性，所以受到了广泛的关注。一幅图像经过 HWT 后，可以得到 16 幅子带图，分别为 4 幅近似图和 12 幅细节子带图，其中 12 幅细节子带图分别为水平、垂直和对角方向的细节信息。HWT 的 12 幅子带图提供了图像多个方向的细节特征，而且 HWT 是建立在四元数的基础上，将彩色图像表示为四元数矩阵的形式，可以更加全面的表示彩色图像的特征。所提算法主要通过学习多个源域数据共享的特征建立通用特征空间，来预测不可见的人脸数据，而 HWT 多个方向的纹理细节特征和颜色特征可以提升特征学习的能力，因此将其与卷积网络联合学习，能够助于卷积网络提取到活体/伪造人脸的多个方向的细节信息，从而为活体人脸检测算法提供有力的判别特征，最终形成用于测试不可见人脸数据的通用特征空间。

活体人脸和伪造人脸在纹理细节信息上存在着明显的差异，因此，在图 1 的判别器特征提取的卷积网络部分中引入了 HWT 的 12 幅细节子带图，重构了判别器特征提取部分，学习图像多个方向上纹理细节特征。由于卷积网络中的池化过程会使得特征图尺寸缩小一半，而 HWT 每一级分解过程也会使得图像尺寸缩小一半，所以 HWT 一级分解的 12 幅子带图与判别器特征提取卷积网络的第一次池化层级联起来，共同输入到下一个卷积网络中进行学习，相应的二级分解子带图与第二次池化连接起来在下一层网络中进行卷积操作，下面介绍具体实现过程。

HWT 建立在四元数的基础上，形式上由实数和复数小波构成，因此在对输入图像进行 HWT 时，将图像的 H、S 和 V 空间表示为四元数的三个复数形式，且另实数部分为 0，则完成了图像到四元数的映射。由于使用 RGB 形式表示图像时，图像三种颜色分量之间的高度相关性以及亮度和色度信息的不完全分离，使得 RGB 在伪造人脸攻击检测方面存在一定的缺陷，而图像的 HSV 颜色空间为基于亮度和色度信息的分离，提供了较好的学习判别线索[9]。因此，将二维图像信号映射到四元数解析信号中进行 HWT 时，利用公式(3):

$$f^q(x,y)=f(x,y)+\mathbf{i}f_{Hx}(x,y)+\mathbf{j}f_{Hy}(x,y)+\mathbf{k}f_{Hxy}(x,y) \tag{3}$$

将所提算法二维图像的 H、S 和 V 映射到了四元数解析信号中，其中，$\mathbf{i}$、$\mathbf{j}$、$\mathbf{k}$ 分别是水平、垂直和对角方向上的向量，令实数部分为 0，则二维图像的四元数解析信号可简化为:

$$f^q(x,y)=\mathbf{i}H+\mathbf{j}S+\mathbf{k}V \tag{4}$$

式中，上标 q 表示四元数的二维信号，H 表示色度通道，S 表示饱和度通道，V 表示亮度通道。

根据超幅数小波变换的工作原理[19]可得 HWT 的一级分解为：

$$\begin{cases} Q_{LL} = LL_{\varphi_h(x)\varphi_h(y)} + \mathbf{i}LH_{\varphi_h(x)\psi_h(y)} + \mathbf{j}HL_{\psi_h(x)\varphi_h(y)} + \\ \qquad \mathbf{k}HH_{\psi_h(x)\psi_h(y)} \\ Q_{LH} = LL_{\varphi_g(x)\varphi_h(y)} + \mathbf{i}LH_{\varphi_g(x)\psi_h(y)} + \mathbf{j}HL_{\psi_g(x)\varphi_h(y)} + \\ \qquad \mathbf{k}HH_{\psi_g(x)\psi_h(y)} \\ Q_{HL} = LL_{\varphi_h(x)\varphi_g(y)} + \mathbf{i}LH_{\varphi_h(x)\psi_g(y)} + \mathbf{j}HL_{\psi_h(x)\varphi_g(y)} + \\ \qquad \mathbf{k}HH_{\psi_h(x)\psi_g(y)} \\ Q_{HH} = LL_{\varphi_g(x)\varphi_g(y)} + \mathbf{i}LH_{\varphi_g(x)\psi_g(y)} + \mathbf{j}HL_{\psi_g(x)\varphi_g(y)} + \\ \qquad \mathbf{k}HH_{\psi_g(x)\psi_g(y)} \end{cases} \tag{5}$$

式中，$\mathbf{i}$、$\mathbf{j}$、$\mathbf{k}$ 三个方向上的虚部分别表示超复数变换后的水平、垂直和对角上的细节特征信息，所提算法中将这四个通道上三个方向的 12 幅细节子带图与判别器特征提取网络的池化层特征图连接起来，输入到下一层卷积网络中学习多个方向上活体/伪造人脸图像的纹理细节特征，并且将图像提高卷积网络提取图像细节特征的能力，从而可以提升每一个源域判别器的鉴别能力，进一步使得生成器生成更加通用的人脸活体特征空间。

1.3 三元组约束损失函数

在判别器鉴定图像是真/假人脸的过程中，对于同一个域中的每一个真实的人脸图像，相同身份的伪造人脸具有相似的面部特征，而不同身份的真实人脸具有不同的面部特征，这一特点降低了判别器鉴定人脸是否为活体的能力，并且在不同域、不同环境和不同攻击类型数据中，这种现象会变得更加严重。判别器的能力会影响生成器生成通用的人脸活体特征空间的性能。所以本文算法借鉴文献[17]的方法，设置了同一域间和不同域之间的三元组约束损失函数来解决这一问题。如图 2 所示，就是录制视频攻击类型中可能出现的典型情况。

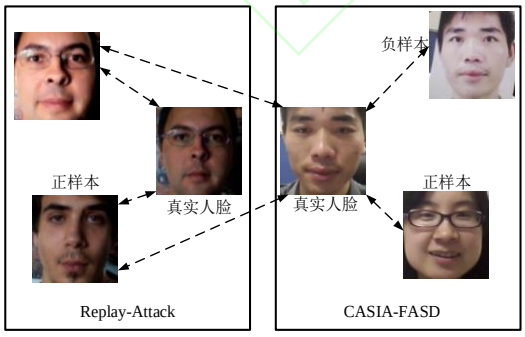


图 2 三元组约束示例[17]

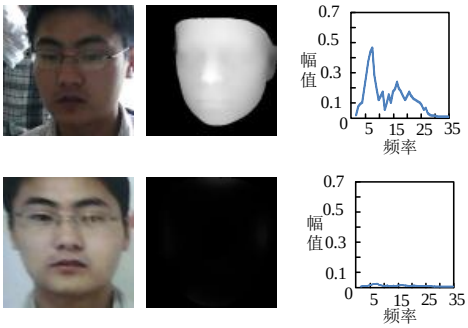
两个约束为：1) 在同一域下，每个真实的人脸到其正样本(所有活体人脸)的距离小于其负样本(所有伪造人脸)的距离。2) 在不同域下，每个活体人脸到其跨域中的正样本的距离小于到其跨域负样本的距离。其损失函数为:

$$\begin{aligned} L_{Trip}(X,Y;G,E) = & \sum [\| E(G(x_i^a)) - E(G(x_j^p)) \|_2^2 \\ & - \| E(G(x_i^a)) - E(G(x_j^n)) \|_2^2 + \alpha_1] \\ & + \gamma \sum [\| E(G(x_i^a)) - E(G(x_k^p)) \|_2^2 \\ & - \| E(G(x_i^a)) - E(G(x_k^n)) \|_2^2 + \alpha_2] \end{aligned} \tag{6}$$

式中，E 表示特征嵌入器，上标 a 和 p 表示同一类，而 a 和 n 是不同的类，下标 i 和 j 表示相同的域，而 i 和 k 是不同的域。1 和 2 分别代表预定义的域内和跨域边距。

1.4 深度图和 rPPG 信号

活体人脸检测技术主要是通过真实人脸和伪造人脸的区别来判断人脸是否为活体的一种技术。活体人脸图像是首次采集的真实人脸，而伪造的人脸是通过二次采集活体人脸图像得到的，所以活体人脸图像具有伪造人脸不具有的活性特征。一方面，因为活体生命个体的心脏跳动引起的脸部血液流动，会使得脸部皮肤发生细微的亮度变化，这种变化可以由远程光电体积描技术 rPPG 测量，相应的活体人脸和伪造人脸的 rPPG 信号分布有着很大的区别，真实和伪造人脸的 rPPG 信号如图 3 第 3 列所示。另一方面，无论是录制的视频还是打印照片，人脸都为平面，而活体人脸的鼻子、嘴巴和额头之间会有一定的深度，所以将人脸图像映射到深度图上时，活体人脸深度图的五官之间有深度值，而伪造人脸则整体呈现为一个平面图，如图 3 第 2 列所示。



人脸图像 深度图 rPPG 信号

图 3 活体人脸和伪造人脸的深度图和 rPPG 信号对比 (第一行为真实人脸，第二行为伪造人脸)[20]

在活体人脸图像数据集中，其活体人脸和伪造人脸图像的 rPPG 信号和深度图都会有这种差异性，因此针对现有算法在特征分类阶段出现训练集中没有的跨域数据集的现象，导致不能很好的获取活体和伪造人脸区分特征的问题，算法将 rPPG 信号和深度图两种真实人脸图像固有的活体特征属性嵌入到特征空间中作为判别依据，增强特征空间对于人脸活体特征的泛化能力，解决未知的人脸数据攻击

问题。

所提算法中，利用人脸对齐网络(position map regression network, PRNET)来估计人脸图像的深度图[21]，在具体的实现过程中，由于伪造人脸图像为一个平面，所以将伪造人脸的深度图值设置为 0，而真实人脸的鼻子、嘴巴和额头之间有一定的深度，所以深度图有一定的数值，伪造/真实人脸深度图的对比如图 3 所示，图 4 的第二行是算法中估计真实人脸深度图的例子。使用深度图信息的损失函数如下所示:

$$L_{Dep}(X;Dep)=\|Dep(G(X)-I)\|_2^2 \tag{7}$$

式中，Dep 是深度图估计器，而 I 是人脸深度图的 GT 图。

rPPG 信号的传统计算方法对于面部表情和光照比较敏感，并且较难区分视频回放攻击类型[22]。因此本章算法使用了文献[10]中的 RNN 网络估计 rPPG 信号，克服了传统方法的缺点。类似于深度图的设定，设伪造人脸的 rPPG 信号为 0，而真实人脸有一定的数值，真实/伪造人脸 rPPG 信号的对比如图 3 所示，图 4 第三行是算法估计真实人脸 rPPG 信号的例子。使用 rPPG 信号的损失函数:

$$L_{rPPG}(X;rPPG)=\|rPPG(G(X)-f)\|_2^2 \tag{8}$$

式中，rPPG 是信号估计器，而 f 是 rPPG 信号的 GT 图。

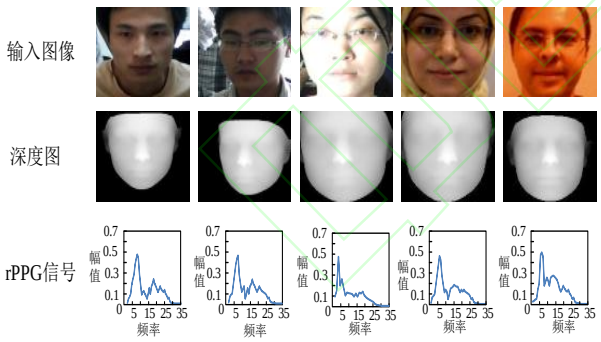


图 4 真实人脸的深度图和 rPPG 信号示例

1.5 深度图和 rPPG 信号

如图 1 所示，在得到通用特征空间的过程中，首先，使用二分类交叉损失函数 L_{cls} 在不同数据域下获取各自的判别特征；接着，为了得到不会偏向任何域且共享的通用特征空间，采用了生成器和判别器对抗损失函数 L_{DG}；然后，为了提升判别器鉴定人脸是否为活体的能力，使用了三元组约束函数 L_{Trip}；最后，为了使得通用空间更具一般性，分别将深度图损失 L_{Dep} 和 rPPG 损失 L_{rPPG} 嵌入到通用特征空间中。因此，算法总损失函数 L 可以表示为:

$$\min_{G,E,C,Dep,rPPG}\ \max_{D_1,D_2,\ldots,D_N} L=L_{cls}+L_{DG}+L_{Trip}+L_{Dep}+L_{rPPG} \tag{9}$$

考虑到算法结构的复杂性，将算法的训练过程分为两个阶段以进行优化。1）利用 L_{cls}、L_{DG} 和 L_{Trip} 损失函数一起训练 G、E、C 和 D_1、D_2、$\ldots D_N$；2）使用 L_{cls}、L_{Dep} 和 L_{rPPG} 损失函数训练 G、Dep 和 $rPPG$；在训练过程中交替重复这两个阶段，直到算法收敛为止，来使得特征生成器 G 得到更加通用的特征空间。

2 实验结果与分析

在实验部分，首先介绍了使用的数据集和评价标准，接着为了验证所提算法的性能，在实验部分分别与现有的基于纹理和深度学习的算法进行了对比，包括 MS LBP[23]、Binary CNN[24]、IDA[25]和 CT[26]的算法进行了对比。此外，为了证明所提算法在类似算法中的性能，与类似于所提算法模型的域生成对抗网络算法进行了对比，包括 MADDG[17]和 MMD-AAE[27]两种算法。然后，为了验证所提算法中 HWT 和 rPPG 信号的有效性，分析了单独添加这两个模块的效果。最后，可视化了超复数小波生成对抗网络的特征提取层，进一步分析了网络提取的特征。

在训练过程中，使用了 Adam 优化器优化训练过程，算法在 pytorch 框架上实现，特征生成器、深度图和 rPPG 信号估计的卷积层的大小为 3×3，判别器的卷积层大小为 4×4，输入图像尺寸大小为 256×256。在具体实现训练时，如 1.5 节内容所示，分两个阶段进行，第一阶段学习率为 1e-5，第二阶段学习率为 1e-4,每个输入数据域的 bachsize 设置为 20，即三个输入数据域的 bachsize 为 60。算法中参数 γ、α_1 和 α_2 分别设定为 0.1、0.2 和 0.5。

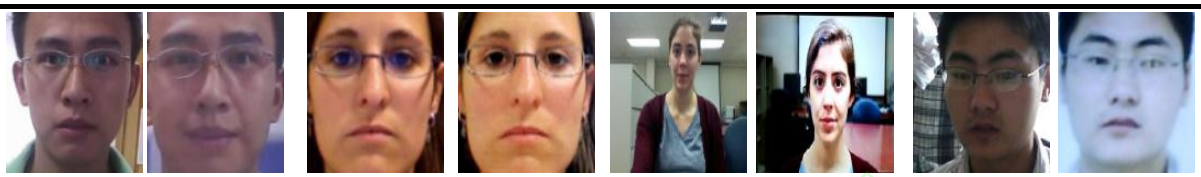
为了对比算法结果的公平性，所有测试都在相同的计算机环境下获得，CPU 测试平台为 Intel(R) Xeon(R) CPU E5-2618L v3，GPU 为英伟达 GTX 1080 11G 显存，使用 Ubuntu x64 操作系统。

2.1 数据集

使用了四个公开的活体人脸检测数据集评估所提算法的性能，四个数据集分别 CASIA-FASD(C)[28]、Idiap Replay-Attack(I)[29]、MSU-MFSD(M)[30]和 NUAA(N)[20]，包含打印照片和视频回放两种攻击类型，表 1 为四种公开数据集的参数信息，图 5 为四个公开数据集的图像示例。从表 1 和图 5 可以看出，四个数据集在采集设备、光照条件、背景复杂和分辨率等方面都不同，因此假设每个数据集为一个域，则四个域之间的图像特征存在着明显的差异性。

表 1 四个公开数据集的参数信息

数据集	光照条件(有/无)	背景复杂(是/否)	采集设备	图像分辨率	真/伪样本数
CASIA-FASD(C)[28]	无	是	iPad 索尼 NEX-5	640×480 1920×1080	150/450
Idiap Replay-Attack(I)[29]	有	是	iPhone 3GS、iPad	320×240	200/100
MSU-MFSD(M)[30]	无	是	佳能 550D iPhone 5S	2048×1536 1130×640	70/210
NUAA(N)[20]	有	否	网络摄像头	640×480	5105/7509



CASIA-FASD[28]　Replay-Attack[29]　MSU-MFSD[30]　NUAA[20]

图 5　真实人脸的深度图和 rPPG 信号示例

2.2　评价标准

评估人脸活体检测算法的性能一般主要从单数据集和跨数据集测试两个方面进行衡量。训练和测试数据都来自于同一数据集即为单数据集测试的性能。训练和测试数据来自于不同的数据集即为跨数据集测试的性能。所提算法主要是针对训练时未出现的伪造人脸攻击人脸识别系统的问题，因此用跨数据集测试衡量了所提算法的性能。人脸活体检测算法的性能评估要综合考虑活体和伪造人脸的识别率。在所提算法中，使用了人脸活体检测算法常用的评价标准，半错误率 *H*(half total error rate, HTER)和 *C*(area under curve, AUC)来评估所提算法的性能。

H 是错误接收率 *A*(false acceptance rate, FAR)和错误拒绝率 *R*(false rejection rate, FRR)总和的一半。*A* 是指算法将本来是伪造的人脸判断为活体人脸的概率，*R* 是指算法将本来是活体的人脸判断为伪造人脸的概率。*A*、*R* 和 *H* 的计算过程如公式(10)-(12)所示：

$$A = \frac{N_{FR}}{N_{Fake}} \tag{10}$$

$$R = \frac{N_{RF}}{N_{Real}} \tag{11}$$

$$H = \frac{A + R}{2} \tag{12}$$

式中，N_{FR} 表示伪造的人脸鉴定为活体人脸的次数，N_{Fake} 表示伪造人脸检测的总次数，N_{RF} 表示活体的人脸鉴定为伪造人脸的次数，N_{Real} 表示活体人脸检测的总次数。分别使用 *R* 与 *A* 便可绘制 *O*(Receiver Operating Characteristic curve, ROC)曲线，*O* 曲线下的面积即为本文使用的评价标准 *C*。

2.3　在三个公开数据集上训练的对比结果

为了验证算法的性能和说明算法所提生成对抗网络的有效性，与非生成对抗网络算法(基于纹理和深度学习算法)进行了对比。在实验过程中，共使用了四个数据集，任意选三个数据集作为训练，另一数据集作为测试来评估算法的性能。所提算法将随机选择的三个训练数据集作为源域获取共享的域特征，并将另一个训练过程中未出现的数据集作为目标域进行测试,四个数据集分别用 C、I、M 和 N 表示。表 2 为所提算法与非生成对抗网络算法的对比结果，其中 C&I&M—N 表示使用 C、I 和 M 数据集进行训练，使用 N 数据集进行测试，所提算法都优于对比算法，最优精度用加粗进行了标注，这是因为对比算法都侧重于从多个源域中学习仅适合源域数据分布的特征空间，而所提算法利用了多个源域特征之间的关系，学习了它们之间具有区分性和共享的特征空间，这个特征空间在源域和看不见的目标域之间共享。因此，它能提取出更通用的区分线索，从而证明了所提使用生成对抗网络进行人脸面部攻击检测的有效性。

为了更好的评估本文所提算法的性能，还与类似的生成对抗网络提取域共享特征的算法进行了对比，如表 3 所示。对于 MMD-AAE[27]算法，它是通过对抗特征学习将多个源域对准任意先验分布来学习通用的特征空间。但是，仅将多个源域与预定义的分布进行对齐仍然较难学习到未知目标域下的特征，所以能够在可见的多个源域和未知的目标域之间提取通用特征的本文算法，优于该算法。而对于 MADDG[17]算法，通用特征空间中缺少了对于图像域间的细节特征信息。而在本文算法中，不仅在特征提取部分利用卷积网络学习了 HWT 的 12 幅细节

子带图，人脸图像细节信息的学习增加了判别器的鉴定能力，从而有利于特征生成器生成域间共享和区分性的特征空间，而且在学习到特征空间后，将深度图和 rPPG 信号嵌入到了特征空间中，形成了一个更加通用的活体人脸判别特征空间，虽然在 MSU-MFSD 数据集上的测试结果低于 MADDG[17]，但在其他三个测试数据集上都优于两个对比算法，而且明显的提升了 NUAA 为测试数据集时的精度，所以，所提算法在一定程度上优于 MADDG，因为 MSU-MFSD 数据集有较大的分辨率，使用 MADDG 算法便已较好的提取到了人脸的活性特征，因此，所提算法的优势并不明显。

表 2 在非生成对抗上的对比结果(%)

算法	C&I&M—N		C&I&N—M		C&M&N—I		M&N&I—C	
	H	*C*	*H*	*C*	*H*	*C*	*H*	*C*
MS LBP[23]	35.12	64.46	32.21	68.12	50.17	51.53	53.42	45.02
Binary CNN[24]	34.14	65.08	29.85	78.12	33.48	66.98	32.98	67.93
IDA[30]	55.46	41.29	63.12	30.05	29.21	80.46	52.36	46.87
CT[26]	30.64	72.67	28.38	81.58	37.49	62.47	29.54	79.98
Ours	**15.01**	**91.21**	**17.52**	**88.45**	**21.98**	**86.06**	**24.05**	**84.65**

表 3 在生成对抗算法上的对比结果(%)

算法	C&I&M—N		C&I&N—M		C&M&N—I		M&N&I—C	
	H	*C*	*H*	*C*	*H*	*C*	*H*	*C*
MMD-AAE[27]	23.32	84.15	27.34	81.47	32.46	75.05	45.89	57.26
MADDG[17]	17.90	88.58	**17.02**	**89.20**	22.54	84.72	24.83	83.78
Ours	**15.01**	**91.21**	17.52	88.45	**21.98**	**86.06**	**24.05**	**84.65**

通过表 2 和表 3 的实验结果，可以看出，所提算法都表现出了较好的效果。由于所提算法的思路主要是通过提取一个通用的特征空间来检测训练中未出现的人脸数据，虽然测试的数据在训练过程中不可见，但是他们仍然与多个源域数据共享某些共有的面部攻击特征，例如，不可见域的打印或者视频回放人脸攻击数据，虽然在环境和材质方面都有不同，但是他们与可见域的训练人脸数据在本质上都是纸张或者屏幕。因此，提取多个源域的人脸攻击共有属性可以提升算法检测训练中不可见人脸攻击的性能。

2.4 在两个公开数据集上训练的对比结果

为了更进一步验证所提算法的有效性，使用两个数据集作为源域进行训练，运用另一数据集测试。具体为将 MSU-MFSD(M)和 Idiap Replay-Attack(I)两个数据集作为训练数据，分别使用 NUAA(N)和 CASIA-FASD(C)作为测试集评估算法的性能。实验对比结果如表 4 所示。发现所提算法即使在两个源域训练的情况下，与其他算法相比，依然表现出了较好的结果，证明了使用生成对抗网络产生通用特征空间的有效性。而且，与表 2 相比，使用三个源域进行训练极大的提升了算法的检测性能，而其他算法的改进能力并不是很大，同时也说明了使用生成对抗网络的优势。

表 4 在两个公开数据集上训练的对比结果(%)

算法	M&I—N		M&I—C	
	H	*C*	*H*	*C*
MS LBP[23]	48.07	54.86	51.6	52.09
IDA[30]	42.17	63.15	45.16	58.8
CT[26]	51.96	52.02	55.17	46.89
Ours	**38.16**	**66.42**	**40.67**	**64.78**

2.5 HWT、深度图和 rPPG 信号模块分析

为了验证在 MADDG[17]算法模型的基础上添加 HWT 和 rPPG 信号的有效性，在算法模型上分别加入 HWT 和 rPPG 信号，使用 CASIA-FASD(C)、

Idiap Replay-Attack(R)和 MSU-MFSD(M)数据集训练，NUAA(N)数据集测试，并在 *H* 和 *C* 上证明了两个模块的作用，实验结果如表 5 所示。发现在基本算法模型的基础上，由于 HWT 能够提供图像的纹理细节信息，rPPG 信号反映真/假人脸图像上的差异性，都能提升了判别器鉴别活体人脸的能力，使得获取的特征提取器更具有一般性，因此当单独加入 HWT 和 rPPG 信号时，对算法的检测精度都有一定程度的提升作用。

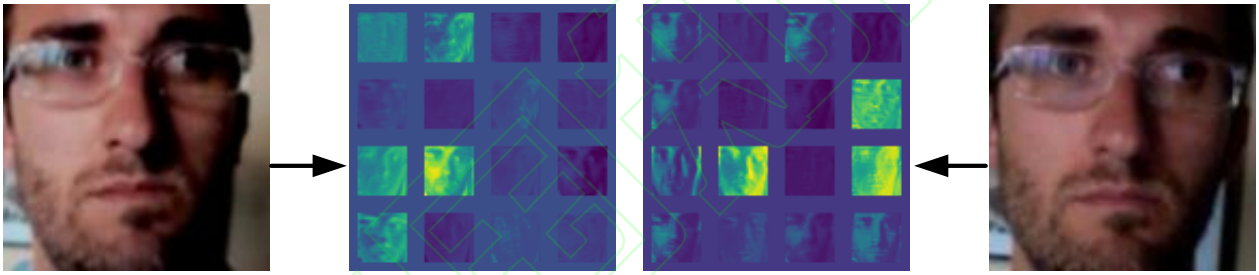
表 5 HWT、深度图和 rPPG 信号模块上的结果(%)

算法	*C*	*H*
MADDG_wo/Dep	71.25	25.32
MADDG[17]	88.58	17.90
Ours1(rPPG)	89.52	16.31
Ours2(HWT)	89.23	16.19
Ours3(rPPG+HWT)	**91.21**	**15.01**

同时，表 5 说明了使用 HWT 和 rPPG 信号的合理性。为了验证 MADDG[17]算法中添加深度图的有效性，使用了没有深度图的算法模型进行了测试，用 MADDG_wo/Dep 表示，如表 5 所示，可以看出，没有深度图时，会使得算法模型的性能大幅下降，从而说明添加深度图作为辅助信息的有效性。类似于添加深度图的方式，添加了 rPPG 辅助信息，和深度图共同来提升算法的性能，使其获得的特征空间具有通用性。

2.6 可视化分析

为了验证本文所提算法在生成对抗网络特征提取卷积层引入 HWT 的 12 幅子带图的有效性，可视化了超复数小波生成对抗网络提取特征的中间过程，一个卷积核对应一个子特征图，特征图对应图像颜色信息以及多个方向的纹理细节特征，仅展示了部分可视化图，如图 6 所示，可以看出，网络不仅可以很好的提取到图像的边缘轮廓信息，而且较好的获取到了人脸图像中的鼻子、眼睛和嘴巴等细节纹理特征。真实图像所获得的特征图在反应人的面部特征时结构更加清晰，而伪造图像则不同，为判别器判别图像真假提供了有效的依据，从而促进形成更加通用的特征空间，提高分类精度。



真实图像　　可视化结果　　可视化结果　　伪造图像

图 6　超复数小波生成对抗网络特征提取可视化结果

3 结 论

本文算法针对现有活体人脸检测算法很难检测到未知的人脸攻击问题，利用 HWT 的细节子带图能够提取图像丰富的细节特征的特点，以及生成对抗网络能够使得源数据的训练模型适应目标数据的特性，提出了一种基于超复数小波生成对抗网络的活体人脸检测算法。首先，将 HSV 空间的三个源域数据输入到生成对抗网络，使得一个特征生成器和三个判别器对抗，形成具有三个源域共享特征且区别于三个域的特征空间，同时，在网络上设置了三元组约束函数来提升判别人脸活体特征的能力，并且将超复数小波变换的 12 幅细节子带图输入卷积网络联合学习图像的细节纹理特征。然后，将活体/伪造人脸都具有的深度图和 rPPG 信号嵌入到特征空间中提高判别人脸活性特征的泛化能力，获取到通用的特征空间。最后，采用训练中未出现的数据进行测试，得到真/假人脸的分类结果。主客观结果表明，所提算法对于活体人脸检测算法有一定的提升作用。未来将在网络和数据占用内存方面进行优化，来适当增加源域数据集的个数，从而更好的提高活体人脸的检测精度。

参考文献:

[1] 谢文化. 人像识别的技术演进路线 [J]. 中国公共安全, 2017(7): 135-138.

[2] 杨艺芳, 王宇平. 一种鉴别稀疏局部保持投影的人脸识别算法 [J]. 西安交通大学学报, 2016, 50(6): 54-60.
YANG Yifang, WANG Yuping. A face recognition algorithm based on discriminant sparse locality and preserving projections [J]. Journal of Xi'an Jiaotong University, 2016, 50(6): 54-60.

[3] 汪亚航, 宋晓宁, 吴小俊. 结合混合池化的双流人脸活体检测网络 [J]. 中国图象图形学报, 2020, 25(7): 1408-1420.
WANG Yahang, SONG Xiaoning, WU Xiaojun.

Two-stream face spoofing detection network combined with hybrid pooling [J]. Journal of Image and Graphics, 2020, 25(7): 1408-1420.

[4] BHARADWAJ S, DHAMECHA T I, VATSA M, et al. Computationally efficient face spoofing detection with motion magnification [C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ, USA: IEEE, 2013: 105-110.

[5] SIDDIQUI T A, BHARADWAJ S, DHAMECHA T I, et al. Face anti-spoofing with multifeature videolet aggregation[C]// Proceedings of the 23rd IEEE International Conference on Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 1035-1040.

[6] TIRUNAGARI S, POH N, WINDRIDGE D, et al. Detection of face spoofing using visual dynamics [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 762-777.

[7] LI X B, KOMULAINEN J, ZHAO G Y, et al. Generalized face anti-spoofing by detecting pulse from face videos[C]// Proceedings of the 23rd IEEE International Conference on Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: IEEE, 2016: 4244-4249.

[8] XU Z Q, Li S, DENG W H. Learning temporal features using LSTM-CNN architecture for face anti-spoofing[C]// Proceedings of the 3rd IEEE Asian Conference on Pattern Recognition. Piscataway, NJ, USA: IEEE, 2015: 141-145.

[9] ATOUM Y, LIU Y J, JOURABLOO A, et al. Face anti-spoofing using patch and depth-based CNNs[C]// Proceedings of the 2017 IEEE International Joint Conference on Biometrics. Piscataway, NJ, USA: IEEE, 2017: 319-328.

[10] LIU Y J, JOURABLOO A, LIU X M. Learning deep models for face anti-spoofing: binary or auxiliary supervision[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 389-398.

[11] SONG Xiao, ZHAO Xu, FANG Liangji, et al. Discriminative representation combinations for accurate face spoofing detection [J]. Pattern Recognition, 2019, 85: 220-231.

[12] ZHNAG P, ZOU F H,Wu Z W, et al. FeatherNets: convolutional neural networks as light as feather for face anti-spoofing[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019:1574-1583.

[13] YU Zitong, LI Xiaobai, NIU Xuesong, et al. Face anti-spoofing with human material perception[C]// Proceedings of the 2020 European Conference on Computer Vision. Cham, Germany: Springer, 2020: 557-575.

[14] YU Z T, ZHAO C X, WANG Z Z, et al. Searching central difference convolutional networks for face anti-spoofing [C]// Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2020: 5294-5304.

[15] JOURABLOO A, LIU Yaojie, LIU Xiaoming. Face de-spoofing: Anti-spoofing via noise modeling[C]// Proceedings of the 2018 European Conference on Computer Vision. Cham, Germany: Springer, 2018: 297-315.

[16] LUO S Y, KAN M, WU S Z, et al. Face anti-spoofing with multi-scale information[C]//Proceedings of the 2018 IEEE Conference on Pattern Recognitio. Piscataway, NJ, USA: IEEE, 2018:3402-3407.

[17] SHAO R, LAN X Y, LI J W, et al. Multi-adversarial discriminative deep domain generalization for face presentation attack detection[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 10023-10031.

[18] LIU Y, STEHOUWER J, JOURABLOO A, et al. Deep tree learning for zero-shot face anti-spoofing[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 4680-4689.

[19] BAYRO-CORROCHANO E. The theory and use of the quaternion wavelet transform [J]. Journal of Mathematical Imaging and Vision, 2006, 24(1): 19-35.

[20] TAN Xiaoyang, LI Yi, LIU Jun, et al. Face liveness detection from a single image with sparse low rank bilinear discriminative model[C]//Proceedings of the 2010 European Conference on Computer Vision. Cham, Germany: Springer, 2010: 504-517.

[21] FENG Yao, WU Fan, SHAO Xiaohu, et al. Joint 3D face reconstruction and dense alignment with position map regression network[C]// Proceedings of the 2018 European Conference on Computer Vision. Cham, Germany: Springer, 2018: 557-574.

[22] DE HAAN G, JEANNE V. Robust pulse rate from chrominance-based rPPG [J]. IEEE Transactions on Bio-Medical Engineering, 2013, 60(10): 2878-2886.

[23] MÄÄTTÄ J, HADID A, PIETIKÄINEN M. Face spoofing detection from single images using micro-texture analysis [C]//Proceedings of the 2011 IEEE International Joint

Conference on Biometrics. Piscataway, NJ, USA: IEEE, 2011: 6117510.

[24] YANG JIANWEI, LEI ZHEN, LI S Z. Learn convolutional neural network for face anti-spoofing [EB/OL]. [2020-08-01]. https://arxiv.org/abs/1408.5601

[25] CHEN Haonan, HU Guosheng, LEI Zhen, et al. Attention-based two-stream convolutional networks for face spoofing detection [J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 578-593.

[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] LI H L, PAN S J, WANG S Q, et al. Domain generalization with adversarial feature learning[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 5400-5409.

[28] ZHANG Z W, YAN J J, LIU S F, et al. A face anti-spoofing database with diverse attacks[C]//Proceedings of the 2012 IEEE International Conference on Biometrics. Piscaraway, NJ, USA: IEEE, 2012: 26-31.

[29] CHINGOVSKA I, ANJOS A, MARCEL S. On the effectiveness of local binary patterns in face anti-spoofing[C]//Proceedings of the IEEE International Conference of Biometrics Special Interest Group. Piscataway, NJ, USA: IEEE, 2012: 6313548.

[30] WEN Di, HAN Hu, JAIN A K. Face spoof detection with image distortion analysis [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 746-761.

（编辑　陶晴）