

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, 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4个不同类型的数据集, 随机选择3个作为训练集, 另一个作为测试集, 形成训练时未知的活体人脸。训练集视为3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3个判别器进行对抗, 当特征生成器成功欺骗过3个判别器时, 形成具有3个源域共享且区别于3个源域的特征空间, 能够检测到不同于源域的人脸特征。在判别器上设置了域间和域内的三元组约束函数, 以此提高判别器的性能, 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。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。在该特征空间中使用测试集进行判别分类, 得到真假人脸识别结果。实验结果表明, 在CASIA-FASD、Replay-Attack和NUAA数据集上, 所提算法的接受者操作特性曲线下的面积分别为84.65%、86.06%、91.21%, 半错误率分别为24.05%、21.05%、15.01%, 均高于对比算法的结果。

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 **文献标志码:** A

DOI: 10.7652/xjtub202105013 **文章编号:** 0253-987X(2021)05-0113-10



OSID 码

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

LI Ce¹, LI Lan¹, XUAN Shuxing¹, YANG Jing², DU Shaoyi³

(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. College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Aiming at the problem that the existing anti-spoofing detection algorithms difficultly detect unknown domain attacks, this paper proposes a face anti-spoofing detection algorithm using generative adversarial networks with hypercomplex wavelet transform to improve the ability of face recognition system to determine the face anti-spoofing. Four different types of datasets are adopted, in which three datasets are randomly selected as the training ones, and the remainder as the test dataset to serve for unknown face anti-spoofing detection during training. The 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 this feature generator

successfully deceives the three discriminators, a feature space sharing three source domains and being different from these domains is formed to detect the characteristics of the unknown domain data. The triple constraint functions inter-class and intra-class are set up to improve the performance of the discriminator, and the detailed subbands of the hypercomplex wavelet transform and the convolution network are combined to learn the detailed features in multiple directions of images. Then the depth map and remote photo plethysmography signal are embedded in the feature space to enhance the generalization performance of the generated feature space for the living face features. The test dataset for discriminative classification in the feature space is used to obtain live/fake results. The results show that on the CASIA-FASD, Replay-Attack and NUAA datasets, the proposed algorithm gets the AUC of 84.65%, 86.06% and 91.21%, the HTER of 24.05%, 21.05% and 15.01%, which are higher than those of the comparative algorithms.

Keywords: face anti-spoofing detection; hypercomplex wavelet; generative adversarial 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,日常生活中刷脸办事已经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3种。真实人脸与3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。传统的算法一

般将人脸活体检测视为二分类问题,利用真假人脸的区别进行分类。Bharadwaj等首先在连续帧上通过放大面部微小的动作进行预处理,然后使用局部二值模式(LBP)提取图像的特征,最后采用支持向量机分类获取结果^[4]。后来,其在文献[5]对算法又进行了改进,在经过预处理步骤后,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,在一定程度上提升了算法性能,但是预处理引入的噪声对整体结果造成了一定的影响。Tirunagari等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等引入了测量活体人脸心率的算法,通过远程光电体积描记术(rPPG)来预测心率^[7]。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。

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CNN-LSTM算法检测活体人脸,具有一定的提升效果^[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等将二分类

问题转换成特征监督问题,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,并且巧妙地设计了一个非刚性注册层来对齐各帧人脸的非刚性运动,突破了传统算法,取得了很好的效果^[10]。Song 等将活体检测放到人脸检测算法框架里面,并将其作为一个类,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红外数据的多模态融合分类器,以此提升人脸图像的活体检测性能^[12]。Yu 等将人脸图像的活体检测问题转换为一种图像材质识别问题,判别人脸的真假^[13]。Yu 提出了一种中心差分卷积的算法,可以很好地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 等利用生成式对抗网络(GAN)设计了一个多对抗性判别网络,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 等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共 3 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获取一般的真假人脸特征存

在一定的难度。所以,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。

1 采用超复数小波生成对抗网络的活体人脸检测算法

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(HWT)^[19]可以提取图像丰富细节信息的特点,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 3 种不同类型的数据集视为 3 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 3 个判别器进行对抗,当特征生成器成功欺骗过 3 个判别器时,则形成了具有 3 个源域共享特征且区别于 3 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,以此提高判别器的性能,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 3 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。

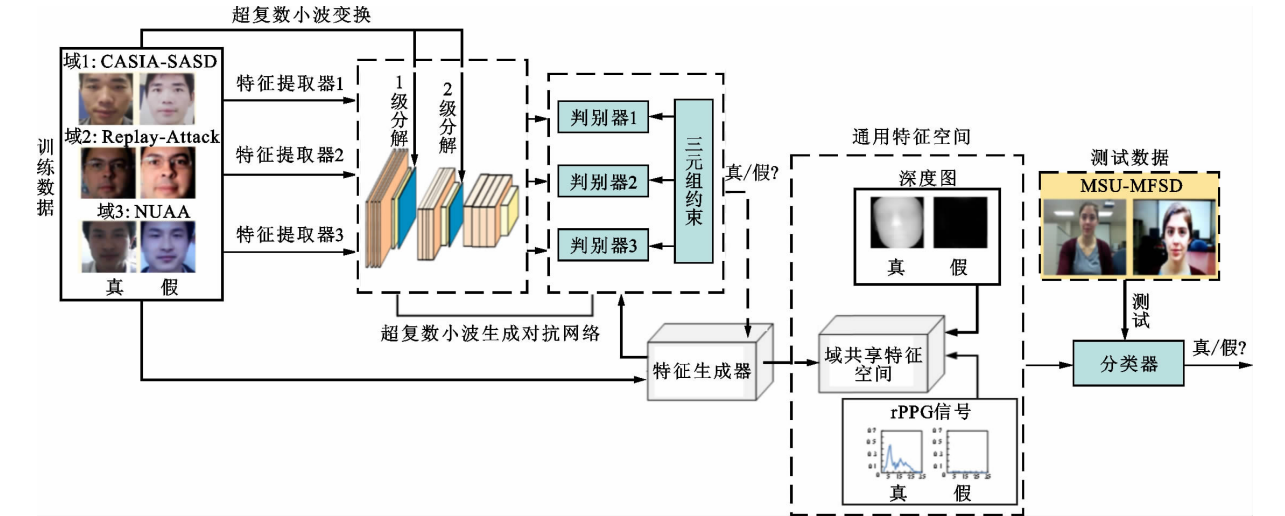


图 1 本文算法框架

Fig. 1 The framework of proposed algorithm

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],实现过程如下。

假设有 N 个源域数据集 $X = \{X_1, X_2, \dots, X_N\}$, 对应的标签为 $Y = \{Y_1, Y_2, \dots, Y_N\}$, 每个标签有两种情况。本文算法 $N=3$ 。当数据来自于活体人脸, 则 $Y_i=1$, 当数据来自于伪造人脸时则 $Y_i=0$ 。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 $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_N$ 。以一个源域为例, 描述二分类的交叉熵损失函数为

$$L_{\text{cls}}(X_1, Y_1; \mathbf{M}_1, \mathbf{C}_1) = -E_{(x_1, y_1) \sim (X_1, Y_1)} \sum_{k=1}^K \log \mathbf{C}_1(\mathbf{M}_1(x_1)) \quad (1)$$

式中: (x_1, y_1) 是 (X_1, Y_1) 的取样; $\mathbf{C}_1$ 是模型分类器; $k \in (1, 2, \dots, K)$ 为所属类的标签, 因为人脸活体检测中只有真和假两种情况, 所以 $K=2$ 。

不同域下的判别特征空间会偏向于各自的源域数据特征, 使得它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, 获取不同域下的不会偏向于任何源域的通用特征空间, 使其具有活体人脸判别特征的一般性, 从而能够很好地检测未知的目标域数据, 实现原理与过程如下。

假设有 N 个源域数据集, 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。对该生成对抗过程进行建模, 目标函数可描述为

$$L_{\text{DG}}(X, X_1, X_2, \dots, X_N; \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_N) = \sum_{i=1}^N \{E_{x \sim X} [\log(\mathbf{D}_i(\mathbf{G}(x)))] + E_{x_i \sim X_i} [1 - \log(\mathbf{D}_i(\mathbf{M}_i(x_i)))]\} \quad (2)$$

式中: $\mathbf{G}$ 表示特征生成器, 用于形成每个源域判别器都难以区分的通用特征空间; $\mathbf{D}_i$ 表示第 i 个域判别器, 用于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 $\mathbf{G}$ 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, 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, 可以更加全面地表示彩色图像的特征。本文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 HWT 与卷积网络联合学习, 有助于卷积网络提取到活体和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, 形成较为通用的特征空间。

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, 重构了判别器的特征提取部分, 学习图像多个方向上的纹理细节特征。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, 所以将 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级池化层级联起来, 共同输入到下一个卷积网络中进行学习。相应地, 二级分解子带图与第二级池化连接起来, 在下一层网络中进行卷积操作, 具体实现过程如下。

HWT 建立在四元数的基础上, 形式上由实数和复数小波构成。使用 RGB 形式表示图像时, 图像 3 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。然而, 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, 在对输入图像进行 HWT 时, 选择将图像的 H、S 和 V 空间表示为四元数的 3 个复数形式, 且令实数部分为 0, 由此完成了图像到四元数的映射, 公式为

$$f^a(x, y) = \mathbf{iH} + \mathbf{jS} + \mathbf{kV} \quad (3)$$

式中: i, j, k 分别是水平、垂直和对角方向上的虚数

单位;上标 q 表示四元数的二维信号; $\mathbf{H}$ 表示色度通道图像; $\mathbf{S}$ 表示饱和度通道图像; $\mathbf{V}$ 表示亮度通道图像。根据超幅数小波变换的工作原理^[19],HWT 的一级分解为

$$\left. \begin{aligned} Q_{LL} &= LL_{\varphi_h(x)\varphi_h(y)} + iLH_{\varphi_h(x)\psi_h(y)} + jHL_{\psi_h(x)\varphi_h(y)} + \\ &\quad kHH_{\psi_h(x)\psi_h(y)} \\ Q_{LH} &= LL_{\varphi_g(x)\varphi_h(y)} + iLH_{\varphi_g(x)\psi_h(y)} + jHL_{\psi_g(x)\varphi_h(y)} + \\ &\quad kHH_{\psi_g(x)\psi_h(y)} \\ Q_{HL} &= LL_{\varphi_h(x)\varphi_g(y)} + iLH_{\varphi_h(x)\psi_g(y)} + jHL_{\psi_h(x)\varphi_g(y)} + \\ &\quad kHH_{\psi_h(x)\psi_g(y)} \\ Q_{HH} &= LL_{\varphi_g(x)\varphi_g(y)} + iLH_{\varphi_g(x)\psi_g(y)} + jHL_{\psi_g(x)\varphi_g(y)} + \\ &\quad kHH_{\psi_g(x)\psi_g(y)} \end{aligned} \right\} \quad (4)$$

式中 i, j, k 这 3 个方向上的数据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。本文算法将这 4 个通道上 3 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,并输入到下一层卷积网络中,以此学习多个方向上活体或伪造人脸图像的纹理细节特征,提高卷积网络提取图像细节特征的能力,提升每一个源域判别器的鉴别能力,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

对于同一个域中的每一个真实人脸图像,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人脸活体特征空间的性能。所以,本文算法借鉴文献^[17]的算法,设置了同一域间和不同域之间的三元组约束损失函数。三元组约束示例如图 2 所示,图

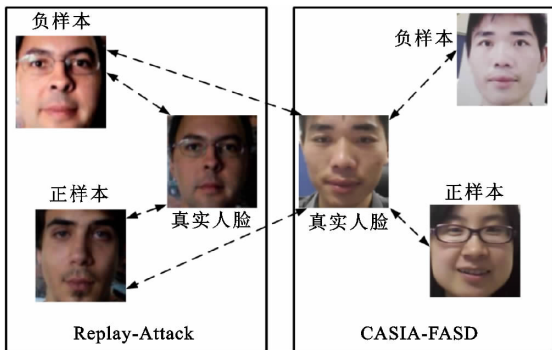


图 2 三元组约束示例^[17]

Fig. 2 The example of the triple constraint^[17]

中虚线表示样本之间的距离。

本文设置两个约束:①在同一域下,每个真实的人脸到其正样本(所有活体人脸)的距离小于到其负样本(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正样本的距离小于到其跨域负样本的距离。损失函数为

$$\begin{aligned} L_{trip}(X_1, X_2, X_3, \dots, X_N; \mathbf{G}, \mathbf{E}) = & \sum_{i=1, j=1}^N [\| \mathbf{E}(\mathbf{G}(x_i^a)) - \mathbf{E}(\mathbf{G}(x_j^p)) \|_2^2 - \\ & \| \mathbf{E}(\mathbf{G}(x_i^a)) - \mathbf{E}(\mathbf{G}(x_j^n)) \|_2^2 + \alpha_1] |_{i=j} + \\ & \gamma \sum_{i=1, k=1}^N [\| \mathbf{E}(\mathbf{G}(x_i^a)) - \mathbf{E}(\mathbf{G}(x_k^p)) \|_2^2 - \\ & \| \mathbf{E}(\mathbf{G}(x_i^a)) - \mathbf{E}(\mathbf{G}(x_k^n)) \|_2^2 + \alpha_2] |_{i \neq k} \quad (5) \end{aligned}$$

式中: $\mathbf{E}$ 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; α_1 和 α_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。活体人脸图像是首次采集的真实人脸,而伪造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,活体生命个体的心脏跳动会引起脸部血液流动,使得脸部皮肤发生细微的亮度变化,这种变化可以由 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。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。活体人脸和伪造人脸的深度图和 rPPG 信号对比如图 3 所示,真实人脸的深度图和 rPPG 信号示例如图 4 所示。

在活体人脸图像数据集中,活体人脸和伪造人脸图像的 rPPG 信号和深度图都会有差异。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,导致不能很好地获取活体和伪造人脸区分特征的问题,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并作为判别依据,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。

本文算法利用人脸对齐网络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像素值设

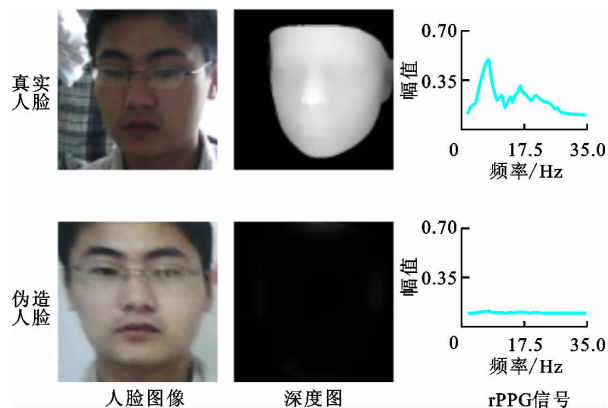


图3 活体人脸和伪造人脸的深度图和rPPG信号对比^[20]

Fig. 3 Comparison of depth map and rPPG between live and fake face^[20]

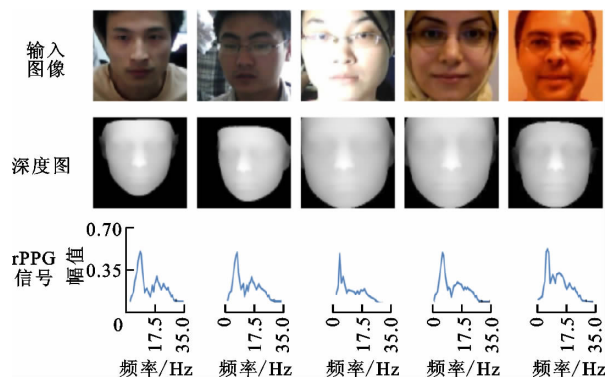


图4 真实人脸的深度图和rPPG信号示例

Fig. 4 The depth map and rPPG of live face

置为0,而真实人脸的鼻子、嘴巴和额头之间有一定的深度,所以深度图像素值非0。使用深度图信息的损失函数为

$$L_D(X; \mathbf{D}) = \|\mathbf{D}(\mathbf{G}(X)) - \mathbf{I}\|_2^2 \quad (6)$$

式中: $\mathbf{D}$ 为深度图估计器; $\mathbf{I}$ 为人脸深度图标定值。

rPPG信号的传统计算算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,本文算法使用了文献[10]中的RNN网络估计rPPG信号,克服了传统算法的缺点。类似于深度图的设定,设伪造人脸的rPPG信号为0,而真实人脸的非0。使用rPPG信号的损失函数为

$$L_r(X; \mathbf{r}) = \|\mathbf{r}(\mathbf{G}(X)) - \mathbf{f}\|_2^2 \quad (7)$$

式中: $\mathbf{r}$ 为信号估计器; $\mathbf{f}$ 为rPPG信号的标定值。

1.5 分类器

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_D 和rPPG损失 L_r 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

$$L = \min(L_{cls} + L_{trip} + L_D + L_r) + \max L_{DG} \quad (8)$$

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 $\mathbf{G}$ 、 $\mathbf{E}$ 、 $\mathbf{C}$ 和 $\mathbf{D}_1$ 、 $\mathbf{D}_2$ 、 $\dots$ 、 $\mathbf{D}_N$;②使用 L_{cls} 、 L_D 和 L_r 损失函数训练 $\mathbf{G}$ 、 $\mathbf{D}$ 和 $\mathbf{r}$ 。在训练过程中交替重复这两个阶段,直到算法收敛为止,使得特征生成器 $\mathbf{G}$ 得到更加通用的特征空间。

2 实验结果与分析

在训练过程中,使用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信号估计的卷积层尺寸为 3×3 像素,判别器的卷积层尺寸为 4×4 像素,输入图像尺寸为 256×256 像素。在具体实现训练时,如1.5小节内容所示,分两个阶段进行,第一阶段学习率为 10^{-5} ,第二阶段学习率为 10^{-4} ,每个输入数据域的batch-size设置为20,即3个输入数据域的batchsize为60。参数 γ 、 α_1 和 α_2 分别设定为0.1、0.2和0.5。

为了公平,所有测试都在相同的计算机环境下获得,CPU为Intel(R) Xeon(R) E5-2618L v3, GPU为英伟达 GTX 1080, 11 GB显存, Ubuntu x64操作系统。

2.1 数据集

使用CASIA-FASD^[23]、Replay-Attack^[24]、MSU-MFSD^[25]和NUAA^[20]共4个公开的活体人脸检测数据集评估本文算法的性能,包含打印照片和视频回放两种攻击类型。4种公开数据集的参数信息和图像示例如图5和表1所示。可以看出,4个数据集在采集设备、光照条件、背景复杂和分辨率等方面



图5 4个公开数据集的图像示例

Fig. 5 Examples of images from four public datasets

表 1 4 个公开数据集的参数信息
Table 1 Parameter information of 4 public datasets

数据集	光照条件	背景复杂	采集设备	图像分辨率/像素	真/伪样本数
CASIA-FASD ^[23]	无	是	iPad 索尼 NEX-5	640×480 1 920×1 080	150/450
Replay-Attack ^[24]	有	是	iPhone 3GS,iPad	320×240	200/100
MSU-MFSD ^[25]	无	是	佳能 550D iPhone 5S	2 048×1 536 1 130×640	70/210
NUAA ^[20]	有	否	网络摄像头	640×480	5 105/7 509

都不同,因此假设每个数据集为一个域,4 个域之间的图像特征存在着明显的差异性。

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。训练和测试数据都来自于同一数据集即为单数据集测试。训练和测试数据来自于不同数据集即为跨数据集测试。本文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。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。本文使用人脸活体检测算法常用的半错误率 H 和接受者操作特性曲线下的面积 C 来评估所提算法的性能。 H 的计算公式为

$$H = (A + R)/2 \tag{9}$$

式中: A 为本来是伪造的人脸判断为活体人脸的概率; R 为本来是活体的人脸判断为伪造人脸的概率。使用 R 与 A 便可绘制接受者操作特性曲线,曲线下的面积即为本文使用的评价标准参数 C 。

2.3 在 3 个公开数据集上训练的对比结果

为了验证本文算法的性能和说明算法所提生成对抗网络的有效性,与基于纹理和深度学习的非生成对抗网络算法 MS LBP^[26]、Binary CNN^[27]、IDA^[25]、CT^[28]进行了对比。在实验过程中,共使用了 4 个数据集,任选 3 个作为训练数据集,另一个作为测试数据集。本文算法将随机选择的 3 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。为方便阅读,将 CASIA-FASD^[23]、Replay-Attack^[24]、MSU-MFSD^[25] 和 NUAA^[20] 4 个数据集简记为 C、I、M、N。用 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。

表 2 为本文算法与非生成对抗网络算法的对比结果。可以看出,本文算法在各数据集下的表现都优于对比算法的,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,而本文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目标域之间共享。由此,本文提出的使用生成对抗网络进行人脸面部攻击检测的有效性得到证明。

为了更好地评估本文算法的性能,还与生成对抗网络提取域共享特征的算法进行了对比,结果如表 3 所示。MMD-AAE 算法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,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^[29]。MADDG^[17] 算法的通用特征空间中缺少了对于图像域间的细节特征信息。本文算法不仅在超复数小波特征提取部分利用卷积网络学习了人脸图像的细节信息,增加了判别器的鉴定能力,有利于提升特征空间的泛化性能,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,获取了可见的多个源域和未知的目标域之间的共享且区分的特征,形成了更加通用的活体人脸判别特征空间。本文算法虽然在 MSU-MFSD 数据集上的测试结果低于 MADDG 算法的,但在其他 3 个测试数据集上都优于两个对比算法的,而且明显提升了 NUAA 为测试数据集时的精度。由于 MSU-MFSD 数据集有较高的分辨率,使用 MADDG 算法已较好地提取到了人脸的活性特征,因此本文算法的优势并不明显。综合分析可知,本文算法在一定程度上优于对比算法。

通过表 2 和表 3 可以看出,本文算法表现出了较好的效果。这是因为本文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。虽然测试数据在训练过程中不可见,但是它们仍然与多个源域数据共享某些面部攻击特征,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,但是它们与可见

表 2 本文算法与非生成对抗网络算法的对比结果

Table 2 Results comparison of the proposed algorithm and the non-generative adversarial network algorithms

%

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 ^[26]	35.12	64.46	32.21	68.12	50.17	51.53	53.42	45.02
Binary CNN ^[27]	34.14	65.08	29.85	78.12	33.48	66.98	32.98	67.93
IDA ^[25]	55.46	41.29	63.12	30.05	29.21	80.46	52.36	46.87
CT ^[28]	30.64	72.67	28.38	81.58	37.49	62.47	29.54	79.98
本文算法	15.01	91.21	17.52	88.45	21.98	86.06	24.05	84.65

注:表中加粗数据为最优值。

表 3 本文算法与生成对抗网络算法的对比结果

Table 3 Results comparison of the proposed algorithm and the generative adversarial network algorithms

%

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 ^[29]	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
本文算法	15.01	91.21	17.52	88.45	21.98	86.06	24.05	84.65

注:表中加粗数据为最优值。

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 2 个公开数据集上训练的对比结果

为了进一步验证本文算法的有效性,使用 M 和 I 两个数据集作为源域进行训练,运用另一个数据集 N 或 C 进行测试,结果如表 4 所示。可以看出,本文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。与表 2 相比可知,使用 3 个源域进行训练极大地提升了算法的检测性能,而其他算法的改进并不是很大,同时也说明了使用生成对抗网络的优势。

表 4 在 2 个公开数据集上训练的对比结果

Table 4 Comparison of training results on 2 public training datasets

%

算法	M&I-N		M&I-C	
	H	C	H	C
MS LBP ^[26]	48.07	54.86	51.6	52.09
IDA ^[25]	42.17	63.15	45.16	58.8
CT ^[28]	51.96	52.02	55.17	46.89
本文算法	38.16	66.42	40.67	64.78

注:表中加粗数据为最优值。

2.5 HWT、深度图和 rPPG 信号模块分析

使用没有深度图的 MADDG^[17] 算法模型验证

深度图的有效性。实验均使用 C、R 和 M 数据集训练,N 数据集测试,结果如表 5 所示。可以看出:没有深度图时,算法的性能大幅下降;在 MADDG 模型上单独加入 HWT 或 rPPG 信号时,算法的检测精度都有一定程度的提升。这是因为 HWT 能够提供图像的纹理细节信息,rPPG 信号能够反映真假人脸图像上的差异,都能提升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。

表 5 HWT、深度图和 rPPG 信号模块上的结果

Table 5 Results on HWT,depth map and rPPG

%

算法	C	H
无深度图的 MADDG ^[17]	71.25	25.32
MADDG ^[17]	88.58	17.90
MADDG ^[17] + rPPG	89.52	16.31
MADDG ^[17] + HWT	89.23	16.19
本文算法	91.21	15.01

注:表中加粗数据为最优值。

2.6 可视化分析

为了验证本文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。图 6 展示了部分可视化图,可以看出,网络不仅可以很好地提取到

图像的边缘轮廓信息,而且较好地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映人的面部特征时结构更加清晰,而伪造图像则不同。这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。

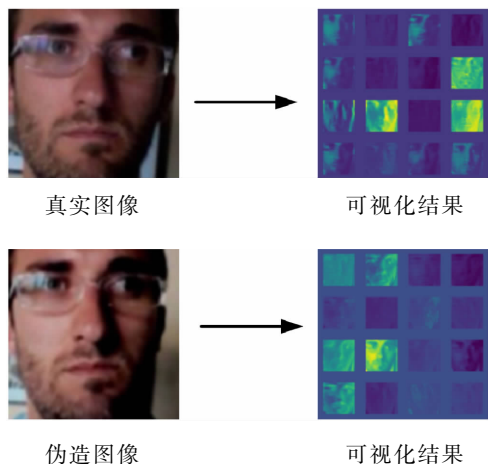


图6 超复数小波生成对抗网络可视化结果

Fig. 6 Visualization results of generative adversarial networks with hypercomplex wavelet

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 3 个源域数据输入到生成对抗网络,使得一个特征生成器和 3 个判别器对抗,形成具有 3 个源域共享特征且区别于 3 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,将活体和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。实验结果表明,在 CASIA-FASD、Replay-Attack 和 NUAA 数据集上,本文算法的接受者操作特性曲线下的面积分别为 84.65%、86.06%、91.21%。

未来将在网络和数据占用内存方面进行优化,适当增加源域数据集的数量,从而更好地提高活体

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: 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 Peng, ZOU Fuhao, WU Zhiwen, 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 Zitong, ZHAO Chenxu, WANG Zezheng, 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 Recognition. 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 J, 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] 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. Piscataway, NJ, USA: IEEE, 2012: 26-31.
- [24] 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.
- [25] 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.
- [26] MÄÄTTÄ J, HADID A, PIETIKÄINEN M. Face spoofing detection from single images using microtexture analysis [C]//Proceedings of the 2011 IEEE International Joint Conference on Biometrics. Piscataway, NJ, USA: IEEE, 2011: 6117510.
- [27] YANG J W, LEI Z, LI S Z. Learn convolutional neural network for face anti-spoofing [EB/OL]. [2020-08-01]. <https://arxiv.org/abs/1408.5601>.
- [28] 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.
- [29] 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.

(编辑 陶晴)