计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200100020

基于深度特征增广的跨域小样本人脸欺诈检测算法

孙文赟[1]　金　忠[2]　赵海涛[3]　陈昌盛[1]
1 深圳大学电子与信息工程学院 深圳市媒体信息内容安全重点实验室　广东 深圳 518060
2 南京理工大学计算机科学与工程学院高维信息智能感知与系统教育部重点实验室　南京 210094
3 华东理工大学信息科学与工程学院　上海 200237
(wenyunsun@szu.edu.cn)

摘　要　随着人脸识别技术的发展,人脸欺诈攻击已经成为一项实际的安全问题,人脸欺诈检测算法用于及早发现该类攻击,保护系统安全。文中将一种经典域自适应算法扩展到深度神经网络中,首先定义了基于深度特征增广的域自适应层,提出了一种基于深度特征增广的跨域小样本人脸欺诈检测算法。该算法在已有的基于全卷积神经网络的人脸欺诈检测深度神经网络的中部嵌入域自适应层将卷积特征图增广,来适配源域和目标域的差异,随后根据增广后的特征图进行像素级分类,最后将像素级概率图从空间上融合为帧级决策。文中在CASIA-FASD,Replay-Attack和OULU-NPU 3个数据集和6个常见测评协议(2个CASIA-FASD与Replay-Attack跨库协议和4个OULU-NPU标准协议)下进行实验,验证了算法在不同背景、不同攻击设备、不同相机等跨域情况下的性能。实验表明,基准FCN人脸欺诈检测算法已经能够达到较好的性能,在此基础上,借助小样本目标域数据学习域自适应模型,可进一步显著提升性能,将错误率减半(CASIA-FASD训练+Replay-Attack测试的HTER指标从27.31%降至11.23%,Replay-Attack训练+CASIA-FASD测试的HTER指标从37.33%降至21.83%,OULU-NPU标准协议IV的ACER指标从9.45%降至5.56%),实验结果验证了基于深度特征增广的跨域小样本人脸欺诈检测算法的有效性。

关键词:模式识别;系统安全;人脸图像分析;人脸欺诈检测;深度学习
中图法分类号　TP311

Cross-domain Few-shot Face Spoofing Detection Method Based on Deep Feature Augmentation

SUN Wen-yun[1],JIN Zhong[2],ZHAO Hai-tao[3] and CHEN Chang-sheng[1]
1 Shenzhen Key Laboratory of Media Security,College of Electronics and Information Engineering,Shenzhen University,Shenzhen,Guangdong 518060,China
2 School of Computer Science and Engineering,Nanjing University of Science and Technology,Nanjing 210094,China
3 School of Information Science and Engineering,East China University of Science and Technology,Shanghai 200237,China

Abstract　The face recognition technology is improving rapidly these days. One the other side,the face presentation attack has become a practical security problem. To protect the system,face presentation attack detection methods are employed for detecting such attacks in advance. This paper extends a classic domain adaptation method to the deep neural network scenario,defines a feature augmentation-based domain adaptation layer,proposes a cross-domain few-shot face presentation attack detection method based on deep feature augmentation. This method is based on the existing method based on Fully Convolutional Network and improves the existing method by embedding a domain adaptation layer in the middle of the network. The new layer augments the feature maps,adapts the difference between the source and target domains. Then,a pixel-level probability map is predicted based on the augmented the feature maps. Finally,the prediction map is fused to a frame-level decision. Experiments are conducted on the CASIA-FASD,Replay-Attack and OULU-NPU datasets. Six commonly used protocols including the cross-dataset protocols between CASIA-FASD and Replay-Attack,the standard protocols of the OULU-NPU dataset are followed. The training and test

到稿日期:2020-01-03　返修日期:2020-04-23　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61902250,61702340,61872188);国家重点基础研究发展计划(2014CB349303);中国博士后科学基金面上项目(2018M643183);广东省自然科学基金(2017A030310382);深圳市基础研究自由探索项目(JCYJ20180305124550725,827/000213)
This work was supported by the National Natural Science Foundation of China(61902250,61702340,61872188),National Basic Research Program of China(2014CB349303),China Postdoctoral Science Foundation(2018M643183),Natural Science Foundation of Guangdong Province(2017A030310382) and Shenzhen Basic Research and Free Exploration Project(JCYJ20180305124550725,827/000213).
通信作者:陈昌盛(cschen@szu.edu.cn)

data are cross different backgrounds，presentation attack instruments and cameras. The experiment results show that the baseline method，the Fully Convolutional Networkbased face presentation attack detection method has already achieved state-of-the-art performance. The performance can be further improved by learning the domain adaptation model on small-sample data in the target domain. The proposed method can halve the error rate by introducing domain adaptation (train on CASIA-FASD and test on Replay-Attack：decreased from 27.31% to 11.23%，train on Replay-Attack and test on CASIA-FASD：decreased from 37.33% to 21.83%，OULU-NPU's standard protocol IV：decreased from 9.45% to 5.56%). This confirms the effectiveness of the proposed method.

Keywords Pattern recognition，System security，Facial image analysis，Face spoofing detection，Deep learning

1 引言

人脸图像可由数码相机、网络摄像头、智能手机等设备方便地采集。然而便利性是一把双刃剑，人脸图像在获得广泛使用的同时，也成为了最不可靠的生物模态。随着人脸识别技术的迅猛发展，基于深度学习和大数据的现代人脸识别算法已经超过了人眼的识别能力，但这些算法容易遭受人脸欺诈攻击。攻击者往往不需要专业的技术或昂贵的设备就能够骗过人脸识别算法，人脸欺诈已经成为当前重要的安全问题之一。攻击者常常使用照片、视频、三维面具等工具来模仿真实人脸，欺骗人脸识别算法，从而获得系统访问授权。

防御者一方可采用人脸欺诈检测算法（又称活体检测算法），辅助判断输入人脸图像/视频的真假，发现人脸欺诈攻击，配合人脸识别算法的工作。欺诈人脸中包含摩尔纹、异常LBP直方图特征、压缩失真、二次采集导致的色彩失真等，算法可以根据这些线索拒绝欺诈人脸，防止其进入人脸识别算法。目前，基于深度学习的人脸欺诈算法成为主流，一些卷积神经网络（Convolutional Neural Network，CNN）[1-3]和全卷积神经网络（Fully Convolutional Network，FCN）[4-7]已被用于检测静态图像中的人脸欺诈，一些递归神经网络（Recurrent Neural Network，RNN）[4]则被用于视频级融合，这些深度学习方法往往采用端到端的方式同时学习图像特征和分类器。

本文的主要目的是提升基于FCN的人脸欺诈检测算法的跨域小样本学习性能，主要贡献包括以下几点。

（1）将一种经典域自适应算法扩展到深度神经网络中，定义了基于深度特征增广的域自适应层。

（2）提出了一种基于深度特征增广的跨域小样本人脸欺诈检测算法，使用目标域小样本学习，提升人脸欺诈检测算法的跨域性能。

（3）在CASIA-FASD，Replay-Attack和OULU-NPU数据集上进行实验，验证了改进算法的有效性，根据标准协议将所提算法与众多已有算法[4,6,8-14]进行对比，结果显示，在小样本目标域额外数据的帮助下，所提算法的性能优于对比算法。

2 相关工作

2.1 人脸欺诈攻击

攻击者使用照片、视频、三维面具等工具来模仿指定身份的人脸，从而达到攻击人脸识别系统的目的，该行为被定义为人脸欺诈攻击。故可将人脸欺诈攻击划分为以下3类[15-16]。

（1）照片攻击。攻击可通过真人拍摄、社交网络下载等方式获取指定身份的人脸照片，将照片显示在屏幕上或打印在纸张上。在没有人脸欺诈检测算法的辅助下，人脸识别算法往往难以区分真实人脸和屏幕/纸张上的人脸。照片攻击是最容易发起的攻击方式。

（2）视频攻击。一些检测方法根据几何变化、眨眼等动态线索来区分屏幕/纸张上的人脸和真实人脸，而攻击者采用录制或下载指定身份的人脸视频，并在多种显示设备上重放的方法来通过该类检测。视频攻击比照片攻击复杂，但仍然易于发起。

（3）三维面具攻击。攻击者使用塑料或硅胶材料的面具模仿指定身份的人脸，由于面具是三维的，比平面的屏幕或纸张更具有真实感。三维面具攻击较少见的原因是制作面具的难度和成本相对较高，因此，在一些人脸欺诈检测数据集中并不包含此类攻击。

人脸欺诈检测算法是针对这些攻击而设计的防火墙，避免人脸识别算法遭受这些攻击。

2.2 一般的人脸欺诈检测算法

人脸欺诈检测算法基本可按照输入数据的种类分为静态图像和动态视频两大类。其中，静态图像数据一类可进一步划分为基于频域特征、空域特征、深度特征和迁移学习4类方法。

在照片攻击和视频攻击中，人脸首先被相机采集，然后打印/显示在纸张/屏幕上，最后被另一个相机重新采集。从采集角度看，该类攻击属于更为一般的图像重采集攻击范畴，故已有一大类防御方法。Muammar等[17]研究了LCD为中间媒介的图像重采集的数学模型。由于LCD屏幕使用RGB亚像素原理合成彩色像素，而相机中采用Bayer滤波器采集彩色像素，离散的像素位置和色彩的亚像素偏移造成重采集图像中包含了一种名为摩尔纹的纹理伪影，类似于LCD屏幕的情况，摩尔纹也会出现在基于半色调技术的印刷品上。摩尔纹是一种图像中的周期噪声，在频域中，可使用离散傅里叶变换（Discrete Fourier Transform，DFT）来分析摩尔纹[18-19]；在空域中，LBP和SIFT等常用图像特征亦可用于摩尔纹检测[20-21]。此外，色彩的失真和较低的质量度是检测重采样图像的另一个线索，这类线索可使用彩色纹理直方图特征[10]、彩色矩特征[19]和图像质量评估（Image Quality Measurement，IQM/Image Distortion Analysis，IDA）特征[13-14]表示。最近，CNN[1-3]和FCN[4-7]也被用于直接学习人脸欺诈检测任务特

征和分类器，2.3－2.4 节将详细对比这些基于深度学习的方法。还有一些研究[22-24]关注不同播放设备(纸张/屏幕)、采集设备(相机)、个体和攻击方式之间的域迁移。

如图 1 所示，应用于动态视频数据的方法可进一步划分为基于时空特征、几何特征和深度特征 3 类方法。用于静态图像数据的二维图像特征可直接推广为三维形式，用于处理视频数据。Pereira 等[25]在时空空间的 3 个正交平面上提取 LBP 特征(LBP-TOP)，用于检测人脸欺诈工具。类似地，大多数基于直方图的二维特征均可推广为对应的三维形式(如 HOG-TOP，BSIF-TOP，LPQ-TOP，LDP-TOP)，这些三维时空特征的提取方法与传统二维特征并没有太大的区别。此外，光流特征是一种提取视频中运动信息的有效方法[26-28]。Yin 等[29]基于光流特征寻找人脸欺诈的运动线索。Pinto 等[30]提出了一种基于底层运动特征和中层视觉编码的特征，用于人脸欺诈检测。Marsico 等[31]在人脸特征点周围提取几何不变特征来检测视频重放中的线索。Liu 等[4]使用 RNN 将帧级预测融合为视频级预测，以实现视频数据的人脸欺诈检测。

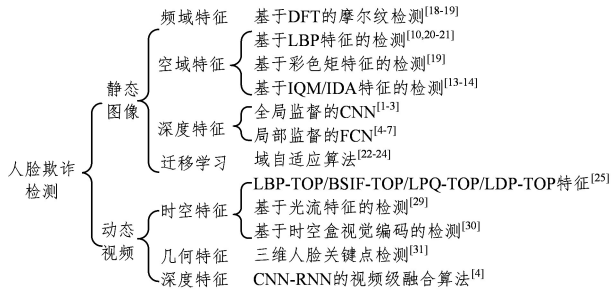


图 1 人脸欺诈检测的相关工作

Fig. 1 Related studies on face presentation attack detection

2.3 基于 CNN 的人脸欺诈检测算法

CNN 在众多计算机视觉领域取得成功，一种直接的方法是将已对齐的人脸图像送入 CNN 分类器。Menotti 等[1]使用超参数搜索方法来寻找合适的用于人脸欺诈检测的 CNN 网络结构，为缩小超参数的搜索范围，被搜索的 CNN 最多包含 3 个卷积层。Rehman 等[2]使用端到端的方式训练了一个 11 层的 VGG 网络及两个变种网络，用于人脸欺诈检测。Nagpal 等[3]探索了更深的基于 ResNet 和 GoogLeNet 的人脸欺诈检测。以上 3 种方法因将整张已对齐的人脸图像送入 CNN 而被称为全局算法。由于人脸欺诈线索(摩尔纹、特定 LBP 特征、压缩失真、色彩失真等)普遍重复地存在于图像的各位置，利用全局算法计算并不高效，故我们可将人脸局部块送入分类器。例如，Atoum 等[5]设计了一个基于局部块的人脸欺诈检测 CNN，该网络仅使用了少数局部块，并没有考虑局部块之外的区域，下文将回顾基于局部监督的 FCN 能够避免数据浪费的问题，且比文献[5]的方法的计算效率更高。

2.4 基于 FCN 的人脸欺诈检测算法

欺诈失真是一种加在干净人脸图像上的高频弱信号，基于 Jourabloo 等[8]的研究发现，欺诈失真信号有两个重要属性：普遍性和重复性。首先，普遍性是指失真在空间中每个位置都存在；其次，重复性是指信号在空间中按照固定的模式重复。基于这两点性质，可使用 FCN 将局部块映射为局部标签，同一张人脸的局部处处相等(真：1，假：0)[6]。Sun 等[7]则给出了一般性的理论分析，证明了基于局部监督的 FCN 比全局监督的 CNN 更适合人脸欺诈检测任务，并采用像素级局部三元标签监督来训练 FCN 并达到领先的性能。此外，Liu 等[4]和 Atoum 等[5]使用深度图作为 FCN 的辅助监督数据，深度图与局部标签方法相似，亦能够获得相同的好处。由于充分利用了欺诈失真信号的普遍性和重复性，基于 FCN 的人脸欺诈检测往往比基于 CNN 的方法效果更好。

然而已有的基于 FCN 人脸欺诈检测算法存在一些局限性。跨域/跨库是目前人脸欺诈检测算法的主要研究方向，而 FCN 并不能消除域偏移。本文在之前有关 FCN 的研究基础上[7]，提出了一种基于深度特征增广的跨域小样本人脸欺诈检测算法。

3 基于深度特征增广的跨域小样本人脸欺诈检测

3.1 基准 FCN 网络

如 2.4 节中的回顾，基于 FCN 的人脸欺诈检测是近两年出现的一种方案，其在人脸欺诈检测任务上比 CNN 更有优势。Jourabloo 等[8]发现了欺诈线索具有普遍性和重复性，即摩尔纹和失真等以重复方式出现在图像中的每个局部区域，该性质是 FCN 方法的科学依据，Liu 等[4]、Atoum 等[5]、George 等[6]和 Sun 等[7]相继研究了基于 FCN 的人脸欺诈检测算法。三元标签监督的 FCN 方法[7]与深度图监督的 FCN 方法[4]共享了相同的 FCN 网络主体结构，为了避免随意创造新的网络，本文将 FCN 作为基准网络，并对其进行改进。文献[4,7]中的 FCN 使用了 RGB-HSV 六通道的输入层，在一些初步尝试之后，我们发现附加的 HSV 并没有明显的效果，故将其移除以简化网络。

图 2 展示了所提出的网络结构，其中，与大多数 CNN/FCN 类似，本文的 FCN 的输入是维度为 $256\times256\times3$ 的 RGB 图像，我们沿用了与基准网络[4,7]一致的卷积层和跳连接的设计，网络总共包含了 13 个 3×3 卷积层和 3 个最大池化层，随着层数加深，特征图的空间尺寸逐渐降低到 32×32 像素。网络的中部设置了两个跳连接，用于鼓励网络特征图的感受野/分辨率的尺度多样性。卷积层后没有设置全连接层，最后一个卷积层直接跟随一个 Sigmoid 层，逐像素地将特征归一化到[0，1]范围以表示概率。训练阶段直接使用像素级监督学习概率图，而测试阶段则将像素级预测平均以获得帧级预测。网络的深度和卷积核大小都接近于 19 层 VGG 网络[32]，但由于网络不包含全连接层，总计仅有 2 243 405 个可训练参数，比常见的包含全连接层的传统 CNN 的参数量少。基于 FCN 的人脸欺诈检测是一项像素级分类任务，属于底层图像任务而非高层的有关语义的任务，目前网络的非线性拟合能力足够满足要求。网络中的两个跳连接为其带来了深度上的灵活性，前向和反向信号可经由跳连接快速传播。监督标签的形式是 FCN 与 CNN 的区别之一。在人脸欺诈检

测任务中，由于有关欺诈的线索（摩尔纹、特定 LBP 特征、压缩失真和色彩失真等）存在于每个局部图像块中，只需观察一个较小的图像块即可充分地做出决策，故 FCN 适用于该类任务。然而，这类性质并不普遍存在于所有图像分类任务中。在人脸识别等任务中，观察一个人脸图像块并不能判断图像块所属的身份，必须观察整张图像才能做出决策，故 FCN 不适合人脸识别等全局任务。

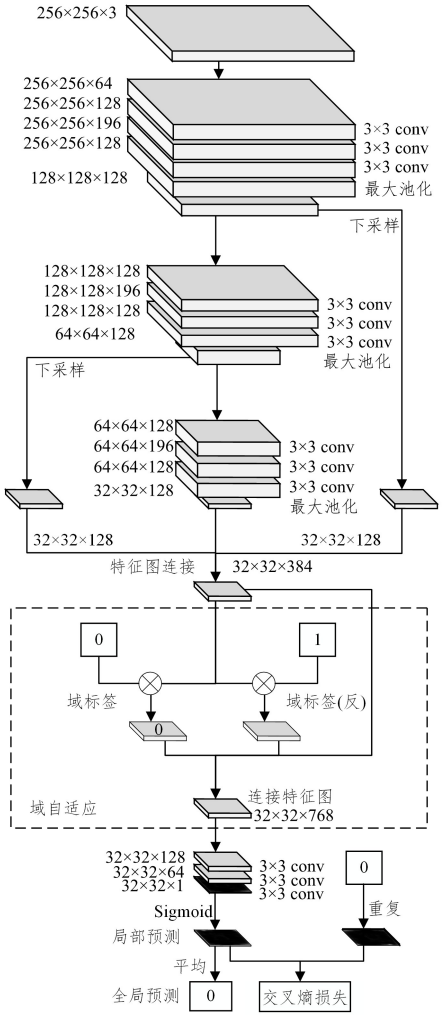


图 2 网络结构

Fig. 2 Network architecture

有两个因素使得 FCN 更适合人脸欺诈检测：在训练中，像素级标签监督比全局图像监督稠密，梯度在稠密监督的作用下更精确，效果类似于将不同图像块分别训练，但更加高效；在测试中，像素级决策对应不同输入图像块，这些决策融合后可提升测试准确率。

3.2 基于特征增广的域自适应层与 FCN 网络

在图 2 中，虚线框中包含的为提出的基于特征增广的域自适应层，该层嵌入在深度神经网络中工作，其设计思路来自于文献[33]，本文将其推广并用于深度神经网络。为将源域 D^s 中的特征 $x\in\mathbb{R}^C$ 适配为目标域 D^t 中的特征 $F(\boldsymbol{x})$，文献[33]定义了一种基于特征增广的映射：

$$F(\boldsymbol{x})=\begin{cases}\langle \boldsymbol{x},\boldsymbol{x},\boldsymbol{0}\rangle, & \boldsymbol{x}\in D^s\\ \langle \boldsymbol{x},\boldsymbol{0},\boldsymbol{x}\rangle, & \boldsymbol{x}\in D^t\end{cases}$$

其中，$\boldsymbol{0}=\langle 0,0,\cdots,0\rangle\in\mathbb{R}^C$，$F(x)\in\mathbb{R}^{3C}$。该增广方式还对应于一个核版本，令 φ 为特征空间到核希尔伯特空间（Reproducing Kernel Hilbert Space，RKHS）的映射，$k(\boldsymbol{x},\boldsymbol{x}')=\langle\varphi(\boldsymbol{x}),\varphi(\boldsymbol{x}')\rangle$ 为核函数，则（无限维）RKHS 空间下的增广为：

$$F(\boldsymbol{x})=\begin{cases}\langle \phi(\boldsymbol{x}),\phi(\boldsymbol{x}),\boldsymbol{0}\rangle, & \boldsymbol{x}\in D^s\\ \langle \phi(\boldsymbol{x}),\boldsymbol{0},\phi(\boldsymbol{x})\rangle, & \boldsymbol{x}\in D^t\end{cases}\tag{2}$$

由于 RKHS 的维度被扩大，增广后的新核函数为：

$$K(\boldsymbol{x},\boldsymbol{x}')=\begin{cases}2k(\boldsymbol{x},\boldsymbol{x}'), & \boldsymbol{x}\boldsymbol{x}'\\ k(\boldsymbol{x},\boldsymbol{x}'), & \boldsymbol{x}\boldsymbol{x}'\end{cases}\tag{3}$$

基于以上思路，本文可以导出一种适用于深度神经网络的特征增广。在全连接网络中，令 f 为输入数据 $\boldsymbol{x}$ 到深度特征 $f(\boldsymbol{x})\in\mathbb{R}^C$ 的非线性映射，则深度特征空间中的特征增广可被定义为：

$$F(\boldsymbol{x})=\begin{cases}\langle f(\boldsymbol{x}),f(\boldsymbol{x}),\boldsymbol{0}\rangle, & \boldsymbol{x}\in D^s\\ \langle f(\boldsymbol{x}),\boldsymbol{0},f(\boldsymbol{x})\rangle, & \boldsymbol{x}\in D^t\end{cases}\tag{4}$$

在 CNN 与 FCN 中，令 $f(\boldsymbol{x})\in\mathbb{R}^{W\times H\times C}$ 为 C 个 $W\times H$ 像素的特征图，通过将 $\langle\cdot,\cdot,\cdot\rangle$ 推广为特征图连接以获得增广后的特征 $F(\boldsymbol{x})\in\mathbb{R}^{W\times H\times 3C}$，相比拉直为向量后再连接，特征图连接能够保持特征的空间结构。式(1)、式(2)与式(4)分别为基于特征增广的域自适应方法的原始、核、深度神经网络版本，它们可方便地从 D^s 和 D^t 两个域的自适应任务推广到 K 个域的情况，若增广前的特征维度为 C，则增广后的特征维度为 $(K+1)C$。

深度神经网络域自适应层由式(4)定义，并插入到深度神经网络的串行结构中，将深度特征/特征图增广为原先维度/个数的 3 倍，域自适应层无参且可导，其的加入不影响传统的端到端训练与测试。值得注意的是，一些神经网络域自适应方法[34]学习的是域不变特征，而本文方法则是学习域不变特征、源域特征和目标域特征的组合，且下一层的权重也可拆分为域不变权重、源域权重和目标域权重。在图 2 的深度神经网络中，虚线框内是域自适应层，它位于网络的中部，其之前的卷积层提取不同分种辨率的特征图，特征图在增广、域自适应后进行像素级分类，最后将像素级概率预测从空间上融合为帧级决策。

4 实验与结果分析

4.1 数据集与预处理

本实验使用了 CASIA Face Anti-Spoofing Dataset（CASIA-FASD）[35]，Idiap replay-attack dataset（Replay-Attack）[36] 和 Oulu-NPU face anti-spoofing dataset（OULU-NPU）[37] 3 个数据集。CASIA-FASD 数据集拥有采集自 50 个个体的 600 个视频，Replay-Attack 数据集拥有采集自 50 个个体的 1 300 个视频，OULU-NPU 数据集拥有采集自 55 个个体的 4 950 个视频，视频评价长度约 6.5 s，每秒 25 帧。

在预处理中，我们使用了一个高性能的基于 HOG 与 SVM 的人脸包围盒检测器，来提取人脸包围盒并裁剪人脸图像。约 0.28%的人脸在检测中被拒，被拒的人脸大多为欺诈人脸，它们对视频级决策和结果评价的影响不大。最终我们

在 CASIA-FASD, Replay-Attack 和 OULU-NPU 数据集上分别获得了 26 824, 112. 500, 130 785 张真脸和 83 961, 233 331, 529 942 张欺诈人脸。表 1 列出了预处理前后的数据量。

表 1 CASIA-FASD, Replay-Attack 与 OULU-NPU 数据集的数据量

Table 1 Data size of CASIA-FASD, Replay-Attack, and OULU-NPU dataset

项目	数量		
	CASIA-FASD	Replay-Attack	OULU-NPU
个体	50	50	55
人脸视频	600	1 300	4 950
人脸帧	111 027	347 498	661 905
检出包围盒-真	26 824	112 500	130 785
检出包围盒-欺	83 961	233 331	529 942
拒绝的帧-真	45	0	1
拒绝的帧-欺	197	1 667	1 177

裁剪后的人脸具有不同的尺寸，而神经网络的输入尺寸是固定的，一种常见方法是将图像缩放以适配，但缩放会导致与欺诈检测相关的高频纹理丢失，故实验采用了延拓＋随机裁剪的方式适配尺寸。当图像长/宽小于神经网络输入层尺寸时，延拓图像直到长/宽大于或等于神经网络输入层尺寸，然后在延拓后的图像中随机采样所需的子图像。在测试阶段，则将任意尺寸图像送入全卷积网络以获得对应尺寸的概率图像，将概率图像平均融合为帧级决策。

4.2 训练与测试

将 CASIA-FASD 与 Replay-Attack 数据集分别记为 C 和 R，并采用一个简化的符号“训练集→测试集”来描述训练与测试的协议，例如，$C\rightarrow R$ 和 $R\rightarrow C$ 表示常见的 CASIA-FASD 与 Replay-Attack 之间的跨库实验。实验采用数据集代码加下标的方式表示该数据集的子集，由于 CASIA-FASD 和 Replay-Attack 数据库各有 50 个个体，我们在其中各选择 2 个个体创建子集 C_2 和 R_2，用于小样本域自适应学习。具体地，域自适应算法采用协议 $C+R_2\rightarrow R_{48}$ 与 $R+C_2\rightarrow C_{48}$，该协议接近但不等同于协议 $C\rightarrow R$ 与 $R\rightarrow C$。

OULU-NPU 数据集自带 4 个标准训练测试协议，其中协议 I、协议 II 与协议 III 分别针对跨背景、跨攻击设备和跨相机的情况，最有挑战性的协议 IV 是以上 3 种情况的组合，协议 IV 几乎接近于跨库情况。OULU-NPU 的测试集包含 20 个个体，我们从中选出 1 个个体用于小样本域自适应。在 4.3 节中，实验将根据 OULU-NPU 的 4 个标准协议与 CASIA-FASD 和 Replay-Attack 之间的两个跨库协议对比所提方法与已有方法的性能。训练采用随机梯度下降法，学习率为 0.001，每批 10 个样本，根据表 1 数据，CASIA-FASD, Replay-Attack 和 OULU-NPU 中的真实人脸和欺诈人脸的比例大致为1∶3，1∶2 和1∶4。为解决数据库中类别的分布不平衡的问题，实验采用了文献[43]中的随机采样方法。首先将训练集打乱，分离为真实和欺诈两组，在随机梯度下降的每个迭代中，分别从两组中按序抽取 5 个真实人脸和 5 个欺诈人脸，组成当前代的训练数据。训练在 400 000 代后结束。

在测试阶段，我们将可训练参数固定，并在训练集上计算帧级概率预测值，根据文献[4,6,8,41]的做法，将帧级别概率预测值在时间维平均融合为视频级概率预测值，进一步提高人脸欺诈检测的准确性。最终，在 CASIA-FASD 与 Replay-Attack 数据集上汇报半总错误率(Half Total Error Rate, HTER)，在 OULU-NPU 数据集上汇报平均分类错误率(Average Classification Error Rate, ACER)、攻击呈现分类错误率(Attack Presentation Classification Error Rate, APCER)和友好呈现分类错误率(Bonafide Presentation Classification Error Rate, BPCER)。由于 ACER 为 APCER 和 BPCER 的均值，其综合考虑了假阳性与假阴性两种错误情况，故遵循 OULU-NPU 数据集的要求[37]，实验对比排序以 ACER 指标为准，单独对比排序各方法的 APCER 和 BPCER 不具有实际意义。

4.3 实验结果与分析

我们将基准 FCN 网及所提出的带有域自适应层的 FCN 网络与 18 种已有方法进行对比，其中包括了深度像素级二元监督法[6]、深度辅助监督法[4]、噪声模型法[8]、ST3ASN[9]、彩色纹理法[10-11]和 IQM/IQA＋SVM[12-13]等，所有对比方法在 2 种 CASIA-FASD/Replay-Attack 跨库协议与 4 种 OULU-NPU 标准协议下进行实验测评，并分别获得了表 2 与表 3 中的结果，由此可以得出以下两点结论。

(1)基准 FCN 网络能够达到与对比方法接近的性能，分别在 $C\rightarrow R$ 与 $R\rightarrow C$ 协议上获得了 1/13 和 3/13 排名，在 OULU-NPU 的 4 个标准协议上分别获得了 1/10, 4/10, 3/10 和 3/10 的名次。

(2)在小样本外部训练数据的帮助下，带有域自适应层的 FCN 网络在 6 个协议下均获得了最佳排名。小样本域自适应方法能够显著提高性能，在 $C\rightarrow R$, $R\rightarrow C$ 协议和 OULU-NPU 协议 IV 中，域偏移较大，引入域自适应层可以将错误率减半($C\rightarrow R$：从 27.31%降低到 11.23%，$R\rightarrow C$：从 37.33%降低到 21.83%，OULU-NPU 协议 IV：从 9.45%降低到 5.56%)；而在域偏移较小的情况下，引入域自适应层仍然可带来一定提升(OULU-NPU 协议 II 的错误率从 2.78%降低到 1.95%，协议 III 的错误率从 2.92%降低到 2.09%)。

表 2 CASIA-FASD 与 Replay-Attack 跨库协议的性能对比

Table 2 Performance comparison under cross-dataset protocol between CASIA-FASD and Replay-Attack datasets

方法	$C\rightarrow R$	$R\rightarrow C$
运动特征法[39]	50.20	47.90
LBP＋SVM[39]	55.90	57.60
LBP-TOP＋SVM[39]	49.70	60.60
运动特征法[40]	50.10	47.00
时空盒视觉编码法[30]	34.40	50.00
CNN[41]	48.50	45.50
IQM＋IQA＋SVM[12-14]	37.35	40.90
彩色纹理特征法[11]	47.00	39.60
彩色纹理特征法[10]	30.30	37.70
深度辅助监督法[4]	27.60	28.40
噪声模型法[8]	28.50	41.10
STASN[9]	31.50	30.90
基准 FCN 网络	27.31	37.33
带有域自适应层的 FCN 网络	**11.23**	**21.83**

表 3 OULU-NPU 标准协议下的性能对比

Table 3 Performance comparison under the standard protocols of OULU-NPU dataset

方法	协议 I			协议 II			协议 III			协议 IV		
	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER
LBP+SVM[6]	32.29	12.92	51.67	25.14	30.00	20.28	25.92±11.25	28.50±23.05	23.33±17.98	48.33±6.07	41.67±27.03	55.00±21.21
IQM+SVM[14]	25.00	19.17	30.83	14.72	12.50	16.94	21.95±8.09	21.94±9.99	21.95±16.79	36.67±12.13	34.17±25.89	39.17±23.35
CPqD[42]	6.90	2.90	10.80	—	—	—	—	—	—	—	—	—
GRADIANT[42]	6.90	1.30	12.50	2.50	3.10	1.90	3.80±2.40	2.06±3.90	5.00±5.30	10.00±5.00	5.00±4.50	15.00±7.10
MixedFASNet[42]	—	—	—	6.10	9.70	2.50	6.50±4.60	5.30±6.70	7.80±5.50	—	—	—
MassyHNU[42]	—	—	—	—	—	—	—	—	—	22.10±17.60	35.80±35.30	8.30±4.10
深度辅助监督法[4]	1.60	1.60	1.60	2.70	2.70	2.70	2.90±1.50	2.70±1.30	3.10±1.70	9.50±6.00	9.30±5.60	10.40±6.00
MILHP[43]	4.60	8.30	0.80	5.40	5.60	5.30	4.00±2.90	1.50±1.20	6.40±6.60	12.00±6.20	15.80±12.80	8.30±15.70
STASN[9]	1.90	1.20	2.50	2.20	4.20	0.30	2.80±1.60	4.70±3.90	0.90±1.20	7.50±4.70	6.70±10.60	8.30±8.40
噪声模型法[8]	1.50	1.20	1.70	4.30	4.20	4.40	3.60±1.60	4.00±1.80	3.80±1.20	5.60±5.70	5.10±6.30	6.10±5.10
深度像素级二元监督法[6]	**0.42**	0.83	0.00	5.97	11.39	0.56	11.11±9.40	11.67±19.57	10.56±14.06	25.00±12.67	36.67±29.67	13.33±16.75
基准 FCN 网络	**0.42**	0.00	0.83	2.78	3.33	2.22	2.92±1.34	3.89±2.93	1.94±1.74	9.45±5.04	11.39±6.90	7.50±6.30
带有域自适应层的 FCN 网络	**0.42**	0.00	0.83	**1.95**	2.22	1.67	**2.09±0.98**	2.78±1.76	1.39±1.12	**5.56±2.52**	6.94±3.95	4.17±3.15

6 种协议尽可能创造了典型的跨域情况（不同背景、不同攻击设备和不同相机等）。总体上，6 种协议上的实验结果表明，基准 FCN 网络与带有域自适应层的 FCN 网络均可在跨域情况下良好工作。

结束语 本文在文献[7]的基础上，在深度神经网络中引入基于特征增广的域自适应层，借助目标域中的小样本扩展训练数据，提升跨域性能，所提方法在 6 种协议下与 18 种已有方法进行对比，获得了较好的实验结果。本文方法具有一定的实际用价值，在应用中，可尝试将目前的主流数据集合并作为适配源域，在特定环境和设备等条件下，采集少量（1～2 人）的多种视频为目标域，如此迁移获得针对特定环境和设备等条件的高性能人脸欺诈检测模型。

本文提出的特征增广的域自适应层仅适用于目标域已知情况下的小样本学习任务，例如，在跨攻击设备、跨相机的实验中，当给定较少的目标域数据时能够显著提升性能（错误率减半）。但本文方法不可用于目标域标签未知的无监督域自适应学习任务和目标域图像与标签未知的零样本学习任务，这两点是后续研究的重要方向。

参 考 文 献

[1] MENOTTI D, CHIACHIA G, PINTO A, et al. Deep representations for iris, face, and fingerprint spoofing detection[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 864-879.

[2] REHMAN Y A U, PO L M, LIU M. Deep learning for face anti-spoofing: An end-to-end approach[C]// Signal Processing: Algorithms, Architectures, Arrangements, and Applications. IEEE, 2017: 195-200.

[3] NAGPAL C, DUBEY S R. A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing[C]// International Joint Conference on Neural Networks. IEEE, 2018.

[4] LIU Y, JOURABLOO A, LIU X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision[C]// Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 389-398.

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

[6] GEORGE A, MARCEL S E B. Deep Pixel-wise Binary Supervision for Face Presentation Attack Detection[C]// International Conference on Biometrics. IEEE, 2019.

[7] SUN W, SONG Y, CHEN C, et al. Face Spoofing Detection based on Local Ternary Label Supervision in Fully Convolutional Network[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3181-3196.

[8] JOURABLOO A, LIU Y, LIU X. Face de-spoofing: Anti-spoofing via noise modeling[C]// European Conference on Computer Vision. IEEE, 2018: 290-306.

[9] YANG X, LUO W, BAO L, et al. Face Anti-Spoofing Model Matters, So Does Data[C]// Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 3507-3516.

[10] 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.

[11] BOULKENAFET Z, KOMULAINEN J, HADID A. Face anti-spoofing based on color texture analysis[C]// Conference on Image Processing. IEEE, 2015: 2636-2640.

[12] NIKISINS O, MOHAMMADI A, ANJOS A E, et al. On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing[C]// International Conference on Biometrics. IEEE, 2018: 75-81.

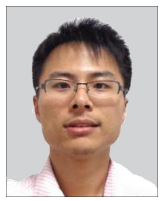
[13] WEN D, HAN H, JAIN A K. Face spoof detection with image distortion analysis[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 746-761.

[14] GALBALLY J, MARCEL S E B, 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.

[15] PATEL K, HAN H, JAIN A K. Secure face unlock: Spoof detection on smartphones[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(10): 2268-2283.

[16] KUMAR S, SINGH S, KUMAR J. A comparative study on face spoofing attacks[C]// International Conference on Computing,

Communication and Automation. IEEE, 2017: 1104-1108.
[17] MUAMMAR H, DRAGOTTI P L. An investigation into aliasing in images recaptured from an LCD monitor using a digital camera[C]// International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 2242-2246.
[18] GARCIA D C, DE QUEIROZ R L. Face-spoofing 2D-detection based on moire-pattern analysis[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 778-786.
[19] NI R, ZHAO Y, ZHAI X. Recaptured Images Forensics Based On Color Moments and DCT Coefficients Features[J]. Journal of Information Hiding and Multimedia Signal Processing, 2015, 6(2): 323-333.
[20] MÄÄTTÄ J, HADID A, PIETIKÄINEN M. Face spoofing detection from single images using micro-texture analysis[C]// International Joint Conference on Biometrics. IEEE, 2011.
[21] PATEL K, HAN H, JAIN A K, et al. Live face video vs. spoof face video: Use of moire patterns to detect replay video attacks [C]// International Conference on Biometrics. IEEE, 2015: 98-105.
[22] LI H, 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.
[23] ARASHLOO S R, KITTLER J, CHRISTMAS W. Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(11): 2396-2407.
[24] YANG J, LEI Z, YI D, et al. Person-specific face antispoofing with subject domain adaptation[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 797-809.
[25] DE FREITAS PEREIRA T, KOMULAINEN J, ANJOS A E, et al. Face liveness detection using dynamic texture[J]. EURASIP Journal on Image and Video Processing, 2014, 2014(1): 2.
[26] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]// International Conference on Neural Information Processing Systems. NeurIPS Foundation, 2014: 568-576.
[27] SUN W, ZHAO H, JIN Z. 3D convolutional neural networks for facial expression classification[C]// Asian Conference on Computer Vision. AFCV, 2016: 528-543.
[28] SUN W, ZHAO H, JIN Z. A facial expression recognition method based on ensemble of 3D convolutional neural networks[J]. Neural Computing and Applications, 2019, 31(7): 2795-2812.
[29] YIN W, MING Y, TIAN L. A face anti-spoofing method based on optical flow field[C]// International Conference on Signal Processing. IEEE, 2016: 1333-1337.
[30] PINTO A, PEDRINI H, SCHWARTZ W R, et al. Face spoofing detection through visual codebooks of spectral temporal cubes [J]. IEEE Transactions on Image Processing, 2015, 24(12): 4726-4740.
[31] DE MARSICO M, NAPPI M, RICCIO D, et al. Moving face spoofing detection via 3D projective invariants[C]// International Conference on Biometrics. IEEE, 2012: 73-78.
[32] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.
[33] DAUMÉ H I. Frustratingly easy domain adaptation[J]. arXiv: 0907.1815, 2009.
[34] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, et al. Domain separation networks[C]// International Conference on Neural Information Processing Systems. NeurIPS Foundation, 2016: 343-351.
[35] ZHANG Z, YAN J, LIU S, et al. A face antispoofing database with diverse attacks[C]// IAPR International Conference on Biometrics. IEEE, 2012: 26-31.
[36] CHINGOVSKA I, ANJOS A E, MARCEL S E B. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing[C]// International Conference of Biometrics. IEEE, 2012.
[37] BOULKENAFET Z, KOMULAINEN J, LI L, et al. OULU-NPU: A mobile face presentation attack database with real-world variations[C]// International Conference on Automatic Face & Gesture Recognition. IEEE, 2017.
[38] HE H, GARCIA E A. Learning from imbalanced data[J]. IEEE Transactions on Knowledge & Data Engineering, 2008(9): 1263-1284.
[39] DE FREITAS PEREIRA T, ANJOS A E, DE MARTINO J E M, et al. Can face anti-spoofing countermeasures work in a real world scenario? [C]// International Conference on Biometrics. IEEE, 2013.
[40] BHARADWAJ S, DHAMECHA T I, VATSA M, et al. Computationally efficient face spoofing detection with motion magnification[C]// Conference on Computer Vision and Pattern Recognition. IEEE, 2013: 105-110.
[41] YANG J, LEI Z, LI S Z. Learn convolutional neural network for face anti-spoofing[J]. arXiv: 1408.5601, 2014.
[42] BOULKENAFET Z, KOMULAINEN J, AKHTAR Z, et al. A competition on generalized software-based face presentation attack detection in mobile scenarios[C]// International Joint Conference on Biometrics. IEEE, 2017: 688-696.
[43] LIN C, LIAO Z, ZHOU P, et al. Live Face Verification with Multiple Instantialized Local Homographic Parameterization [C]// International Joint Conference on Artificial Intelligence. IJCAI Organization, 2018: 814-820.



SUN Wen-yun, born in 1987, Ph.D. His main research interests include deep learning and facial image analysis.

CEHN Chang-sheng, born in 1986, Ph.D, lecturer, postgraduate supervisor, is a member of China Compu- ter Federation. His main research interests include 2D barcode, pattern recognition, machine learning and information security.