JIG
JOURNAL OF IMAGE AND GRAPHICS
主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
07
VOL.25
ISSN1006-8961
CN11-3758/TB

俯视深度头肩序列识别 P1393

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第7期（总第291期）
2020年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

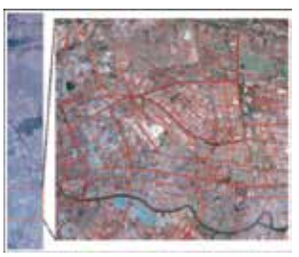
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年7月 第25卷 第7期（总第291期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

非真实感绘制技术研究现状与展望(第1283页)

图卷积网络下牙齿种子点自动选取(第1481页)



迁移学习下高分快视数据道路快速提取(第1501页)

Volume 25, Number 7
Published July 16, 2020

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Research status and progress of non-photorealistic rendering technology(P1283)

Automatic selection of tooth seed point by graph convolutional network(P1481)

Rapid road extraction from quick view imagery of high-resolution satellites combined with transfer learning(P1501)

E-mail: jig@radi.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2020)07-1408-13

论文引用格式:Wang Y H, Song X N and Wu X J. 2020. Two-stream face spoofing detection network combined with hybrid pooling. Journal of Image and Graphics,25(07):1408-1420(汪亚航,宋晓宁,吴小俊. 2020. 结合混合池化的双流人脸活体检测网络. 中国图象图形学报,25(07):1408-1420)[DOI:10.11834/jig.190419]

结合混合池化的双流人脸活体检测网络

汪亚航,宋晓宁,吴小俊

江南大学物联网工程学院, 无锡　214122

摘　要:**目的**　人脸识别技术在很多领域起着重要作用,但大量的欺诈攻击对人脸识别产生了威胁,比如打印攻击和重放攻击。传统的活体检测方法是以手工方式提取特征且缺乏对时间维度的考虑,导致检测效果不佳。针对以上问题,提出一种结合混合池化的双流活体检测网络。**方法**　对数据集提取光流图像并进行面部检测,得到双流网络的两个输入;在双流网络末端加入空间金字塔和全局平均混合池化,利用全连接层对池化后的特征进行分类并进行分数层面的融合;对空间流网络和时间流网络进行融合得到一个最优结果,同时考虑了不同颜色空间对检测性能的影响。**结果**　在CASIA-FASD(CASIA face anti-spoofing database)和replay-attack两个数据集上做了多组对比实验,在CASIA-FASD数据集上,等错误率(equal error rate,EER)为1.701%;在replay-attack数据集上,等错误率和半错误率(half total error rate,HTER)分别为0.091%和0.082%。**结论**　结合混合池化的双流活体检测网络充分考虑时间维度,提出的空间金字塔和全局平均混合池化策略能有效地利用特征。针对包含多种攻击类型、图像质量差异较大的数据集,本文提出的网络模型均能取得较低的错误率。

关键词:活体检测;卷积神经网络;双流网络;光流法;空间金字塔池化;全局平均池化

Two-stream face spoofing detection network combined with hybrid pooling

Wang Yahang, Song Xiaoning, Wu Xiaojun

School of Internet of Things Engineering, Jiangnan University, Wuxi　214122, China

Abstract: **Objective** The convenient and efficient technique of biometric feature identification triggers the wide and profound research on face recognition, which draws extensive concern and has been widely used in various authorization scenarios, including mobile phone unlocking, access control, and face payment. The rise of Internet of Things application terminals including mobile phone cameras make much private facial information of user easily and instantaneously available, which lead to serious threat for the traditional face identification systems. Therefore, more efficient and accurate identifying of face spoofing attacks, including replay and print attacks, reducing threats, and ensuring system security have become an urgent issue to be addressed. Face anti-spoofing mainly ensures that the real entity is displayed in front of the camera, rather than a video or a photo. Many techniques, including convolutional neural network (CNN) methods, have appeared to address this issue and aroused a great concern constantly in recent years. Typically, the hand-crafted features including LBP (local binary pattern), HOG (histogram of oriented gradient) jointly using a kind of classifier such as SVM (support vector machine), is to be the most commonly used technique for face anti-spoofing. However, the diversity of spoofing

收稿日期:2019-08-20;**修回日期**:2019-12-17;**预印本日期**:2019-12-24

基金项目:国家重点研发计划项目(2017YFC1601800);国家自然科学基金项目(61876072);中国博士后科学基金特助项目(2018T110441);江苏省自然科学基金项目(BK20161135);江苏省六大人才高峰项目(XYDXX-012)

Supported by: National Key Research and Development Program of China (2017YFC1601800); National Natural Science Foundation of China (61876072)

means still causes great difficulties in manually extracting effective features. Most CNN methods cannot fully exploit the valuable information in the case of temporal dimensionality and thus result in poor detection. To address this issue, this study presents a two-stream spoofing detection network based on a scheme of hybrid pooling. **Method** Those methods that only use spatial information to solve the face anti-spoofing problem are not generalized. The face anti-spoofing framework proposed in this paper includes spatial stream, temporal stream and fusion modules. The entire design process of experiment is divided into two aspects, one is to use spatial net to achieve spatial information, and another is to apply the temporal net to acquire temporal information. The algorithm implementation is detailed as following. First, optical flow pictures are extracted from the dataset, on which face detection is performed, the optical flow pictures are utilized as the input of the temporal stream to learn temporal information, and the original face pictures are used as the input of the spatial stream to learn spatial information. Second, a shallow network is adopted for the spatial stream network, in which spatial pyramid pooling is combined with global average pooling at the end of the network to obtain hybrid pooling features. We then perform classification using a fully connected layer. For the temporal stream network, we adopt a deep network using residual blocks given that effective temporal features are difficult to extract. Spatial pyramid pooling and global average pooling are also synthesized at the end of the network for the classification. Results obtained from the spatial and temporal stream networks are used for the final score fusion for improved classification. The choice of different color spaces may affect the final result; thus, we compare the effectiveness of different color spaces by experiment and determine the optimal color space of the presented model. **Result** The proposed method is demonstrated on two public benchmark datasets, namely, CASIA-FASD (CASIA face anti-spoofing database) and replay-attack. We obtain the results of 2.141% EER (equal error rate) in the spatial stream, 9.005% EER in the temporal stream, and 1.701% EER in fusion on CASIA-FASD; and 0.071% EER and 0.109% HTER (half total error rate) in the spatial stream, 17.045% EER and 21.781% HTER in the temporal stream, and 0.091% EER and 0.082% HTER in fusion on replay-attack. **Conclusion** The two-stream spoofing detection network combined with the hybrid pooling scheme achieves promising results on different datasets. The proposed method utilizes the effective information in temporal dimensionality. The hybrid of spatial pyramid and global average pooling means can exploit the effective information in multiple scales. We can also relieve the burden of the high dimensionality of datasets. Experimental results obtained from popular datasets demonstrate the merits of the proposed method, especially its robustness to the diversity of spoofing means and the large differences in the quality of image datasets. Our promising results will encourage further work on synthesizing an informative stream and lead to successful solutions for other application domains in the future.

Key words: anti-spoofing; convolutional neural network (CNN); two-stream network; optical flow method; spatial pyramid pooling; global average pooling

0 引　言

卷积神经网络(convolutional neural network, CNN)的流行以及计算机视觉的发展使得高精度的人脸识别成为可能。Taigman 等人(2014)提出的 DeepFace 方法的精度达到 97.35%,错误率比之前表现最佳的方法降低了 27%。与指纹或者虹膜相比,人脸包含更多信息,也更加可区分,并且人脸识别在交互上更加友好。但随着人脸识别的广泛应用,出现了很多欺诈策略。人脸欺诈主要是通过在人脸识别系统前面呈现照片或者视频等假的实体而骗过系统。对于没有加入活体检测的人脸识别算法,通过一个人的照片就能骗过系统(Duc 和 Minh, 2009)。高稳定性、高识别率的欺诈检测算法对人脸识别系统十分重要。

人脸欺诈策略主要有打印攻击(Tan 等,2010)、视频攻击(Chingovska 等,2012)和 3D 面具攻击(Erdogmus 和 Marcel,2014)3 种方式,如图 1 所示。由于 3D 面具攻击需要更多的面部信息以及较高的成本,使用的概率相对较低,所以重点解决打印攻击和视频攻击。不同的攻击方式存在着不同的弱点。打印攻击,由于是打印的照片,所以缺少时间上的信息,如眼动信息或者嘴动信息等;视频重放攻击,由于用显示屏作为呈现设备,因此存在质量降低、产生摩尔噪声等问题。结合混合池化的双流活体检测网络充分利用了二者的特点,对时间和空间分别建模并进行融合。

(a) 打印攻击　(b) 视频攻击　(c) 3D面具攻击

图 1　常见攻击方式

Fig. 1　Common attack methods

((a) print attack; (b) video attack; (c) 3D mask attack)

主流的检测算法大致分为两类:基于传统的方法和基于深度学习的方法。Chingovska 等人(2012)提出使用局部二值模式(local binary pattern, LBP)作为欺诈检测的一个有效特征,并用支持向量机(support vector machine, SVM)作为分类器,取得了不错的结果。还有很多方法都是基于 LBP 进行的研究。除了 LBP 特征之外,方向梯度直方图(histogram of oriented gradient, HOG)(Dalal 和 Triggs, 2005)、Haar(Erdem 等,2011)、SURF(speeded up robust features)(Boulkenafet 等,2017)等传统特征也应用到欺诈检测领域。由于欺诈图像的质量以及纹理都与真正的人脸有所区别,Boulkenafet 等人(2016)利用图像的纹理特征、Wen 等人(2015)利用图像的质量来进行欺诈检测。但上述方法都是针对单张图片,忽略了对应的时间维度信息,于是出现了一些基于运动的方法,Pan 等人(2007)和 Kollreider 等人(2007)分别利用眼动或者嘴动信息,根据正常人每秒的眼动次数或嘴动次数来识别打印攻击。尽管增加时间维度信息后,欺诈检测的错误率不断降低,但在现实中的表现却不尽如人意,原因是数据集无法收集到所有可能出现的欺诈类型,为此,Arashloo 等人(2017)将欺诈检测的问题定义为异常检测,把欺诈样本看做异常点,这样在训练时只需要学习正样本的分布即可,而与欺诈样本无关。类似于上面的方案,Karthik 和 Katika(2017)和 Nikisins 等人(2018)分别利用图像质量和高斯混合模型(Gaussian mixture model, GMM)来学习正样本的分布。虽然传统的手工提取特征方法也能取得不错结果,但是特征设计比较困难且泛化能力有限,比如基于运动的方案对打印攻击有效,对视频重放攻击却无能为力。

随着深度学习的兴起以及卷积神经网络强大的特征表达能力,相关方法引入到欺诈检测领域。Yang 等人(2014)提出使用卷积神经网络解决人脸欺诈问题。相较于传统特征,卷积神经网络具有强大的表征能力,能提取出更具有区分性的特征,并且这个提取特征的过程可以从学习中得到,省去了人工设计的步骤。卷积神经网络的核心在于如何设计监督网络结构,能够使得网络学习的泛化能力更强、提取的特征可区分度更高。Lucena 等人(2017)使用迁移学习方案,有效减少了过拟合和网络训练难度。Li 等人(2016)利用分块将原图像分割成若干小块,然后分块训练。Rehman 等人(2018)采取一些特殊的训练策略来避免网络的过拟合效应。Liu 等人(2018)用 rPPg(remote photo plethysmography)以及面部深度作为监督来引导网络学习到更有效的信息。上述深度学习方法都是基于空间维度信息,没有利用时间维度信息。Xu 等人(2015)引入了 CNN-LSTM(long short term memory)结构,利用 CNN 学习每个帧的空间特征信息,然后用 LSTM 学习多帧之间时间信息。同样,Asim 等人(2017)将 LBP-TOP(local binary patterns from three orthogonal planes)与 CNN 相结合是由于 LBP-TOP 包含时间维度信息,Gan 等人(2017)使用 3D-CNN 也是考虑了时间上的差异。而 Jourabloo 等人(2018)则假设照片或者视频攻击会引入噪声,将欺诈检测问题看做估计噪声函数的过程。Pérez-Cabo 等人(2019)利用度量学习来学习正负样本的分布,Liu 等人(2019)提出的深度树网络(deep tree network, DTN)则对之前提出的异常检测方案的缺点做了进一步的改进。

结合上述方案,可以看出不同的攻击方案各有特点。很多算法对时间维度信息的处理采用 3D-CNN 或者 LBP-TOP 等特征,但这些特征将时间维度信息和空间维度信息混在一起,不能很好表征时间维度的作用。因此,本文提出一种双流网络结构,包含空间流和时间流两部分,能清晰表征时间流的作用。除此之外,受到 Zhang 等人(2019)方案的启发,设计了一种结合空间金字塔和全局平均混合池化的模块。空间金字塔池化可以多尺度地学习特征层的信息,相对于之前的全连接,全局平均池化能够在不破坏特征空间信息的前提下降低特征的维度。

最后将两个池化的特征分别输入到全连接层进行分类，将分类结果融合为最终的结果。为了验证本文方案的有效性，从多个方面进行实验。1）针对很多方案提到的不同颜色空间会对最终结果产生影响，对不同颜色空间进行测试，选出最适合的颜色空间；2）进行空间流、时间流以及融合后的对比；3）进行结合混合池化及不结合混合池化的对比。

本文贡献可以总结如下：1）创新性地使用了空间金字塔和全局平均混合池化；2）提出一个结合空间信息和时间信息的双流网络；3）在空间流上进行了不同颜色空间的测试，并得出最适合的颜色空间。

1　相关工作

深度学习不断发展，逐渐应用于计算机视觉的各个领域。深度学习的卷积核具有可学习性，且较传统手工提取的特征具有更强的泛化能力，因此深度学习广泛应用于图像特征的学习。深度学习的关键是网络的结构设计。卷积神经网络一般包括输入层、卷积层、池化层、激活函数等。本文以输入层和池化层为切入点，提出了一个输入是光流图像和普通图像的混合池化网络。

1.1　光流法

光流由 Arnheim 和 Gibson（1952）首次提出，代表着空间运动物体在成像平面上的像素运动的瞬间速度，是通过图像序列中像素在时间维度上的变化以及相邻帧之间的相关性来找到上一帧和当前帧之间的对应关系，从而计算出相邻帧之间物体的运动信息的一种方案。

假设一帧图像的像素为 $I(x,y,t)$，其中 x，y 代表像素所在位置，t 代表第 t 帧。在时间 $\mathrm{d}t$ 后移动到下一帧图像的 $(x+\mathrm{d}x, y+\mathrm{d}y)$ 位置，根据光流法的在连续两帧图像之间像素的灰度值不改变这一假设，可以得到

$$I(x,y,t) = I(x+\mathrm{d}x, y+\mathrm{d}y, t+\mathrm{d}t) \tag{1}$$

对式（1）右侧进行泰勒级数展开，消去相同项，可得

$$\boldsymbol{I}_x\mathrm{d}x + \boldsymbol{I}_y\mathrm{d}y + \boldsymbol{I}_z\mathrm{d}t = 0 \tag{2}$$

令 $u = \dfrac{\mathrm{d}x}{\mathrm{d}t}$，$v = \dfrac{\mathrm{d}y}{\mathrm{d}t}$，将式（2）两边除以 $\mathrm{d}t$，得到

$$\boldsymbol{I}_x u + \boldsymbol{I}_y v = -\boldsymbol{I}_t \tag{3}$$

式（3）称为光流方程，式中 $\boldsymbol{I}_x$ 和 $\boldsymbol{I}_y$ 是图像梯度，$\boldsymbol{I}_t$ 是时间方向的梯度，只需求出 u，v 即可。u，v 可以通过 Lucas 和 Kanade（1997）的方法计算得到。

图 2 展示了通过 RGB 图像生成的光流图像，可以看出对于打印攻击，由于纸张是长方形，导致光流图像也呈现长方形的轮廓。在活体检测中，利用光流图像代表时间维度上的信息。

图 2　RGB 图像及对应的光流图像

Fig. 2　RGB picture and corresponding optical flow picture

1.2　空间金字塔池化

空间金字塔池化是 He 等人（2014）为了解决目标检测中目标尺度不统一提出的方法，可以避免对图像进行缩放或剪裁，消除普通卷积神经网络需要固定大小输入的要求。空间金字塔池化层结构如图 3 所示，假设输入的是一个 $w \times h \times c$ 的特征图，第 1 级是对每个通道做最大池化，最终得到一个维度为 c 的特征。第 2 级是先将图像分为 4 个小块，然后对每个小块进行最大池化，得到一个 $4 \times c$ 维的特征。第 3 级是先将图像分为 16 个小块，再对每个小块进行最大池化，得到一个 $16 \times c$ 维的特征。最后将 3 级的特征连接在一起，得到一个 $21 \times c$ 维的特征，这个特征就是3级空间金字塔池化的输出结果。从计

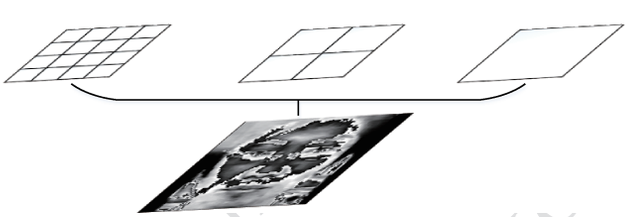


图 3　空间金字塔池化

Fig. 3　Space pyramid pooling

算结果可以明显看出,输出的特征大小与输入的特征大小无关。

对于活体检测而言,很多研究都表明对图像进行缩放会破坏图像的信息。加入空间金字塔池化之后,可以输入任意大小的图像,而不用进行缩放。空间金字塔池化结构能够做到对特征信息进行融合,先处理特征的整体,再处理特征的局部,能够从细密级和粗糙级上理解图像,这样就能够学习到多个尺度上的图像特征。可见,在活体检测方面,空间金字塔池化具有如下优势:1)图像无需进行缩放,减少了高频信息损失;2)可以从多个尺度融合深层特征信息。

1.3 全局平均池化

早期的卷积神经网络,输入一般经过卷积层、池化层以及一个或多个全连接层。这种网络结构的全连接层往往参数过多,极易出现过拟合现象。所以Lin 等人(2014)提出使用全局平均池化层代替全连接层,这样既能够对数据进行降维,还能够极大减少网络的参数。全局平均池化的结构如图 4 所示。可以看出,对于一个输入为 $w \times h \times c$ 的特征图,经过全局平均池化后得到一个维度为 $c \times 1$ 的特征。全局平均池化降维和减少参数的特性使其能对整个网络在结构上做正则化进而防止过拟合。

Zhou 等人(2016)利用全局平均池化特征图的每个通道产生一个特征的特点,最后分类结果中每个特征的权重代表了每个特征图的权重,将特征图的每个通道和权重相乘再相加,最后得到一个类激活图,通过这个类激活图就能够对结果进行一些可视化,清楚地看到是图像的哪个部分起到作用,为以后的研究指明方向。

1.4 不同颜色空间输入

HSV 颜色空间包含色相、饱和度、明度 3 个分量,YC_rC_b 颜色空间包含明亮度和色度,RGB 颜色空间则包含红、绿、蓝 3 个颜色分量。每个颜色空间包含的信息不同,各有特点。RGB 中包含丰富的空间信息,与人类看到的色彩最为相近,而 HSV 和 YC_rC_b 颜色空间则对亮度更加敏感。RGB 颜色空间可以转化为 HSV 和 YC_rC_b,具体计算为

$$\boldsymbol{V} = \frac{1}{3}(\boldsymbol{R} + \boldsymbol{G} + \boldsymbol{B}) \tag{4}$$

$$\boldsymbol{S} = 1 - \frac{3}{\boldsymbol{R} + \boldsymbol{G} + \boldsymbol{B}}[\min(\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B})] \tag{5}$$

$$\boldsymbol{H} = \arccos\left\{\frac{[(\boldsymbol{R} - \boldsymbol{G}) + (\boldsymbol{R} - \boldsymbol{B})]/2}{[(\boldsymbol{R} - \boldsymbol{G})^2 + (\boldsymbol{R} - \boldsymbol{B})(\boldsymbol{G} - \boldsymbol{B})]^{\frac{1}{2}}}\right\} \tag{6}$$

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{C}_r \\ \boldsymbol{C}_b \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{G} \\ \boldsymbol{B} \end{bmatrix} \tag{7}$$

相关研究对不同颜色空间的有效性进行了大量实验,有些只采用单个颜色空间,比如只使用 RGB,也有些采用多个颜色空间堆叠,比如 Boulkenafet 等人(2017)采用 HSV + YC_rC_b 的方案进行学习。本文使用 RGB、HSV、YC_rC_b 以及多个颜色空间堆叠的方案进行实验,进而得出一个有效的颜色空间作为本次实验空间流的输入。

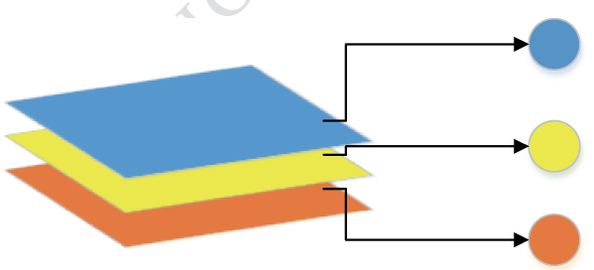


图 4 全局平均池化

Fig. 4 Global average pooling

2 结合混合池化的双流网络

当前欺诈检测的方法多是基于卷积神经网络,很多方案为了避免过拟合,在输入图像层面做了一些工作,如将图像分割为多个小块(Li 等,2016),或利用 LSTM(Xu 等,2015)结构来获取时间维度信息。基于已有的研究,本文提出了一种结合空间金字塔和全局平均混合池化的双流网络结构。在空间金字塔池化中,能够从特征层面进行分块操作。双流网络结构中的时间流能够更加有效地学习到时间维度信息。最后通过一个超参数动态调节时间流和空间流的比例。

2.1 双流网络结构

双流网络结构包括空间流网络和时间流网络。很多方法如 CNN-LSTM 或 3D-CNN,尝试将时间信息考虑进去,它们将时间维度信息和空间信息合成了一种特征,但时间信息带来的不一定就是积极的影响,如对于重放攻击或者活动幅度较小的攻击,时

间信息就会带来负面影响，所以直接使用带时间维度信息的特征并不能灵活地分配空间信息和时间信息的比例。而在双流网络中，空间维度和时间维度是分开的，可以动态地进行两个维度结果的融合。双流网络结构如图 5 所示。其中，BN 为批归一化（batch normalization），fc 为全连接（fully connected）。对于空间流网络（SpatialNet），定义为

$$\boldsymbol{s}_{\mathrm{s}} = SN(\boldsymbol{x}) \tag{8}$$

式中，$\boldsymbol{x}$ 代表 RGB 或者 HSV 等不同颜色空间的图像。从图 5 可以看出，SpatialNet 包含多个卷积层、以及由全局平均池化（global average pooling，GAP）和空间金字塔池化（spatial pyramid pooling，SPP）构成的混合池化层。$\boldsymbol{s}_{\mathrm{s}}$ 代表整个空间流网络的输出分数。对于时间流网络（TemporalNet），定义为

$$\boldsymbol{s}_{\mathrm{t}} = TN(\boldsymbol{x}') \tag{9}$$

式中，$\boldsymbol{x}'$ 代表光流图像，即两幅图像对应像素的位置差，因此包含了时间维度信息。由于光流图像表现出来的特征相对不容易学习，所以对于时间流的网络使用了更深的网络结构。从图 5 可以看出，TemporalNet 包含多个带有残差块的层。通过加入带有残差块的层，能够加深网络的结构，这就使得光流网络能够学到可区分的特征。$\boldsymbol{s}_{\mathrm{t}}$ 代表整个时间流网络的输出分数，最终整个网络的输出为

$$\boldsymbol{s}_f = (1-\alpha)\times\boldsymbol{s}_{\mathrm{s}} + \alpha\times\boldsymbol{s}_{\mathrm{t}} \tag{10}$$

式中，α 是一个超参数，代表空间流网络和时间流网络的比例。在活动比较大的数据集中可以选取较大的 α，在活动比较小的数据集中则选取比较小的 α。相比较过去的融合方法，比如使用最小二乘法（Schwartz 等，2011）进行融合，本文提出的方案也更具有可调节性以及解释性，但由于是手工指定的权重，融合的效果可能会有所下降。

在网络学习的过程中，SpatialNet 的损失为

$$\boldsymbol{l}_{\mathrm{s}} = cn_{\mathrm{S}}(\boldsymbol{s}_{\mathrm{s}}, \boldsymbol{label}) \tag{11}$$

式中，$\boldsymbol{s}_{\mathrm{s}}$ 是空间流网络的输出，$\boldsymbol{label}$ 代表输入 $\boldsymbol{x}$ 的标签，cn_{S} 为交叉熵损失（cross_entropy），空间流则是优化 $\boldsymbol{l}_{\mathrm{s}}$，使其最小即可。

类似的，TemporalNet 的损失为

$$\boldsymbol{l}_{\mathrm{t}} = cn_{\mathrm{T}}(\boldsymbol{s}_{\mathrm{t}}, \boldsymbol{label}) \tag{12}$$

式中，$\boldsymbol{s}_{\mathrm{t}}$ 是时间流网络的输出，时间流则是优化 $\boldsymbol{l}_{\mathrm{t}}$，使其最小即可。

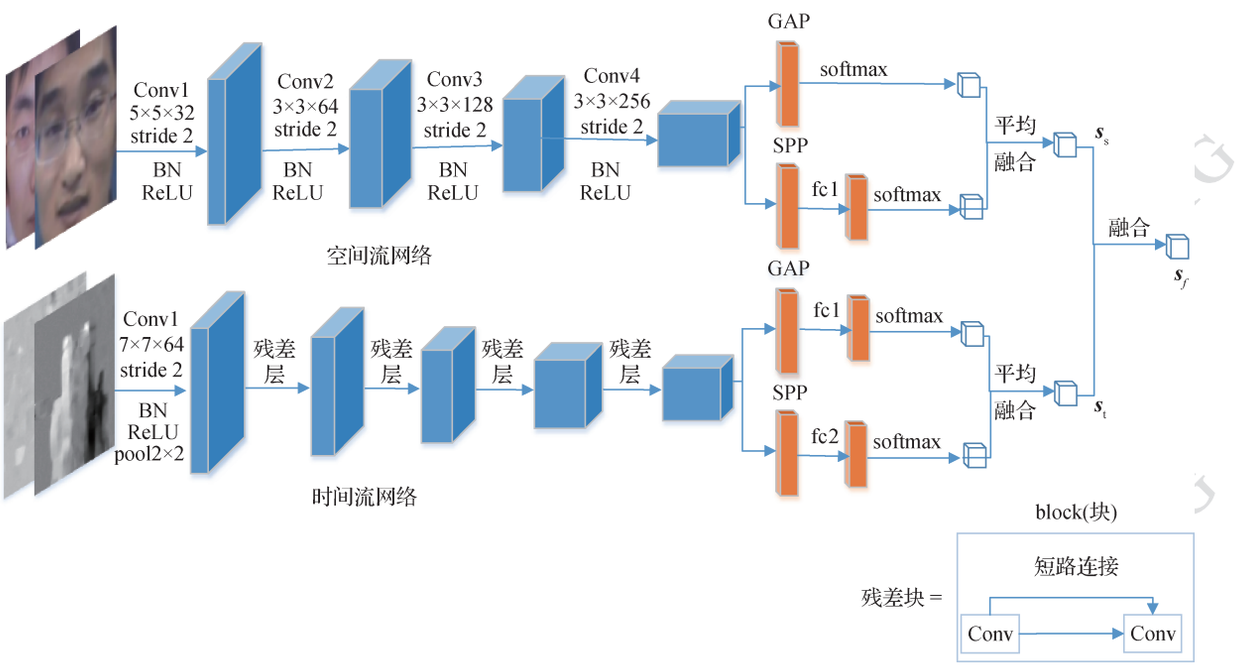


图 5　网络结构（时间流中的残差层由一个残差块和多个卷积层连接而成）

Fig. 5　Network structure (the residual layer in the temporal stream is connected by a residual block and multiple convolution layers)

2.2　空间金字塔和全局平均混合池化

为了学习到更加可区分的特征，很多方法都采用了局部与全局相结合的方案，将原始数据切割成多个小块，然后将每个小块都作为一个输入进行学习。本文创新性地引入了空间金字塔池化（SPP），在多个尺度上对特征图进行分块，取代直接对输入

图像分块。如图 5 所示,输入图像经过前面的多个卷积模块后,能够得出一个特征图,通过空间金字塔池化将特征图划分为多块,得到一个 n 维的特征,最后使用全连接层以及 softmax 对这个 n 维特征进行分类得到一个代表真脸置信度的分数。

在欺诈检测领域,传统卷积神经网络利用最大池化或者平均池化达到下采样的目的,并将池化后的结果直接展开成 1 维的,但是这样会破坏图像空间层面的信息。在目标检测领域,很多方案采用全局平均池化(global average pooling, GAP)的方案,在进行全连接操作之前,先进行全局平均池化可实现减少参数、降低特征图的维度及减少过拟合,同时可保证不破坏特征图的空间特性。在图 5 中,输入图像经过前面的多个卷积模块,能够得出一个特征图,对特征图进行全局平局池化,得到一个 m 维的特征,最后使用 softmax 对这个 m 维特征进行分类得到一个分数。

空间金字塔池化对特征分块,较好地展示了信息的局部性特征。全局平均池化对整个特征图进行池化,较好地展示了信息的全局性特征。将两个池化的分数融合成一个空间金字塔全局平均混合池化,这样就能够学习到更加全面的信息。

3 实验结果与分析

3.1 数据集

3.1.1 CASIA-FASD 数据集

CASIA-FASD 数据集是 Zhang 等人(2012)提出的,包含不同图像质量以及不同攻击方式的视频。共 50 个主题,训练集 20 个,测试集 30 个。每个主题包括 3 个真脸以及 9 个攻击。攻击类型分为弯曲照片攻击、剪洞攻击和视频攻击。图 6 展示了 3 种攻击方式,每种攻击分为高分辨率、低分辨率以及正常分辨率 3 种情况。CASIA-FASD 数据集中定义了 7 种测试方案:1)弯曲照片测试,仅使用弯曲照片和真实照片;2)剪洞测试,仅使用剪洞照片和真实照片;3)视频测试,仅使用视频攻击照片和真实照片;4)低分辨率测试,仅使用低分辨率的图像;5)正常分辨率测试,仅使用正常分辨率的图像;6)高分辨测试,仅使用高分辨率的图像;7)整体测试,测试时将所有数据都进行测试。本文按照上面 7 种测试方案,对 CASIA-FASD 数据集进行测试。

(a) 弯曲照片攻击　(b) 剪洞攻击　(c) 视频攻击

图 6　CASIA-FASD 数据集中的 3 种攻击方式

Fig. 6　Three types of attacks in CASIA-FASD dataset ((a) warped photo attack; (b) cut photo attack; (c) video attack)

3.1.2 replay-attack 数据集

replay-attack 数据集是 Chingovska 等人(2012)提出的,包含不同场景以及不同攻击方式的视频。为了便于评估,数据集分为训练集、验证集和测试集。每个集合包含真脸以及攻击。replay-attack 数据集包含打印攻击和视频重放攻击。攻击分为手持和固定两种形式。由于固定和手持造成的晃动并不大,所以在时间维度上面的信息较少。数据集中收集了不同光照条件下的信息,这也使得测试结果更加接近现实情形。

3.2 实验准备

在开始实验前,需要进行数据预处理。对于空间流,1)对视频进行图像采样,本文采用每隔 3 帧取样一幅图像的方式;2)使用多任务卷积神经网络(multi-task convolutional neural network, MTCNN)(Zhang 等,2016)进行人脸检测,将人脸部分裁剪出来,得到 224 × 224 像素大小的人脸;3)进行数据增强。由于数据集中攻击样本较少,所以扩大攻击样本,并且加入一些随机亮度、随机翻转等。对于时间流,由于两幅空间流才能对应一幅光流的图像,为了统一处理,将视频第 1 帧 $\boldsymbol{t}_1$ 作为启动帧,在训练时不使用,因此空间流的样本是 $\{\boldsymbol{t}_2,\boldsymbol{t}_3,\boldsymbol{t}_4,\cdots\}$,对应光流的照片是 $\{\boldsymbol{t}_2-\boldsymbol{t}_1,\boldsymbol{t}_3-\boldsymbol{t}_2,\boldsymbol{t}_4-\boldsymbol{t}_3,\cdots\}$,从而使空间流和时间流对应起来,方便后面融合。光流的计算采用 OpenCV 中的光流计算方法。受 Yang 等人(2014)的启发,图像的背景信息在欺诈检测中也会起到重要作用,所以对于光流图像是利用带背景信息的图像产生的光流,而不是经过人脸检测后的面部图像。

训练时采用 Adam 优化器对卷积神经网络进行

优化,学习率采用 0.000 1,batch 的大小为 32,对时间流网络和空间流网络分别进行训练。测试时将两个网络的结果代入式(10)得到最终的预测结果。

3.3　评估方法

在 CASIA-FASD 数据集上进行实验,采用等错误率(equal error rate,EER)作为评判标准,等错误率是错误接受率(false acceptance rate,FAR)与错误拒绝率(false rejection rate,FRR)相等时的错误率,也是 ROC(receiver operating characteristic)曲线与对角线的交点。在 replay-attack 数据集上,使用等错误率和半错误率(half total error rate,HTER)做为评判标准,HTER 具体计算为

$$H = \frac{FA(\kappa) + FR(\kappa)}{2} \tag{13}$$

式中,FA 是假阳性率,又称为误诊率,代表本来是正样本但被预测为负样本的百分比,FR 是假阴性率,又称为漏诊率,代表本来是负样本但被预测为正样本的百分比,κ 可以通过验证集确定。

3.4　实验结果与分析

3.4.1　实验结果

为了验证空间金字塔和全局平均混合池化以及时间流网络的作用,在 CASIA-FASD 和 replay-attack 数据集上进行大量的对比实验。

表 1 和表 2 展现了不同方案在 CASIA-FASD 数据集中 7 种场景下的 EER,表 1 没有加入空间金字塔和全局平局混合池化,表 2 加入了空间金字塔和全局平局混合池化。可以看出,增加了空间金字塔和全局平均混合池化后取得了更好的结果,用 RGB 颜色空间作为输入的空间流的 EER 从 2.963% 降到 2.141%。时间流上的 EER 从 11.414% 降到 9.005%。在空间流上,进行了不同通道以及多通道堆叠的实验。可以看出,在本文提出的结构中,RGB 通道表现出了更好的结果,原因是 RGB 到 YC_rC_b 可以通过线性变换得到,而这正是卷积神经网络所擅长的。在时间流上,单个时间流表现欠佳,为 9.005%。但与空间流融合后有所提升,达到 1.701%。通过对 CASIA-FASD 的 7 种不同场景进行测试可以看出,本文方法对于正常质量下的图像具有更好的鉴别能力,而对于不同的欺诈方式表现结果差别不是太大。

表 3 和表 4 展示了本文提出的方案在 replay-attack 数据集中的 EER 和 HTER。可以看出,加了空间金字塔和全局平均混合池化后明显有更好结果,在空间流上能取得 0.071% 的 EER 和 0.109% 的 HTER。而对于颜色空间,在未加混合池化的时候,HSV 颜色空间在 HTER 上也取得了不错结果,而加入混合池化后 RGB 则取得了更好结果,主要原因是 replay-attack 数据集中有很多亮度不稳定的样本,所以在对颜色空间更敏感的 HSV 上取得了微弱的优势,而加入混合池化后,混合池化能够指导网络学习到更加全局以及局部的信息,而且 RGB 颜色空间也可以通过变换得到 HSV,所以 HSV 的微弱优势就被抵消了。从表 3 和表 4 还可以看出,replay-attack 数据集中的时间流网络效果不是很好,原因是 replay-attack 数据集中大多数视频抖动非常小,所以在时间维度产生的信息也相对较少。因此在 replay-attack 数据集的实验中可以调小时间流网络所占的比例,减少造成的负影响。最终空间流和时间流融合使得 EER 和 HTER 分别达到 0.091% 和 0.082%。

表 1　不同方案使用基础网络在 CASIA-FASD 数据集中 7 种场景下的 EER

Table 1　EER in seven scenarios on CASIA-FASD data set with different approaches by using basic network

/%

方案	不同分辨率			不同攻击类型			全部数据
	低分辨率	正常分辨率	高分辨率	弯曲攻击	剪洞攻击	视频攻击	
RGB 通道	4.968	2.325	**1.377**	3.381	2.753	**3.039**	2.963
HSV 通道	6.183	3.394	2.209	4.682	3.135	3.237	3.661
YC_rC_b 通道	6.624	**0.563**	3.177	3.709	3.534	3.212	3.537
多通道	4.451	4.393	2.171	4.682	2.692	4.085	3.741
时间流	8.223	8.036	16.578	9.285	9.196	15.333	11.414
RGB 通道 + 时间流融合	**2.617**	1.194	3.336	**1.603**	**1.891**	3.766	**2.381**

注:加粗字体表示最优结果。

表 2　不同方案使用混合池化网络在 CASIA-FASD 数据集中 7 种场景下的 EER

Table 2　EER in seven scenarios on CASIA-FASD dataset with different approaches using hybrid pooling network

/%

方案	不同分辨率			不同攻击类型			全部数据
	低分辨率	正常分辨率	高分辨率	弯曲攻击	剪洞攻击	视频攻击	
RGB 通道	6.225	**0.125**	**1.801**	2.141	2.075	2.303	2.141
HSV 通道	6.282	2.325	2.209	3.365	3.384	2.865	3.226
YC_rC_b 通道	4.854	0.544	2.224	3.154	2.897	3.173	2.925
多通道	6.282	0.732	2.224	3.613	3.021	3.116	3.191
时间流	9.761	2.891	14.601	7.126	8.117	12.061	9.005
RGB 通道 + 时间流融合	**3.309**	0.225	2.209	**1.644**	**1.395**	**2.141**	**1.701**

注：加粗字体表示最优结果。

表 3　不同方案使用基础网络在 replay-attack 数据集中的 EER 和 HTER

Table 3　EER and HTER on replay-attack dataset with different approaches using basic network

/%

方案	EER			HTER		
	固定	手持	全部数据	固定	手持	全部数据
RGB 通道	**0.011**	0.655	**0.351**	0.841	**1.161**	1.001
HSV 通道	0.051	**0.651**	0.381	**0.098**	0.562	**0.325**
YC_rC_b 通道	0.075	0.967	0.532	1.274	1.571	1.421
多通道	0.453	0.704	0.413	0.401	0.751	0.727
时间流	23.751	23.215	23.496	30.52	30.709	30.614
RGB 通道 + 时间流融合	**0.011**	0.655	**0.351**	0.841	**1.161**	1.001

注：加粗字体表示最优结果。

表 4　不同方案使用混合池化网络在 replay-attack 数据集中的 EER 和 HTER

Table 4　EER and HTER on replay-attack dataset with different approaches using hybrid pooling network

/%

方案	EER			HTER		
	固定	手持	全部数据	固定	手持	全部数据
RGB 通道	**0**	**0.118**	**0.071**	**0.005**	0.223	0.109
HSV 通道	0.011	0.504	0.261	0.009	0.776	0.385
YC_rC_b 通道	0.001	0.513	0.261	0.224	0.822	0.517
多通道	0.011	0.443	0.241	1.844	1.848	1.846
时间流	16.934	17.217	17.045	21.646	21.924	21.781
RGB 通道 + 时间流融合	0.01	0.131	0.091	0.01	**0.141**	**0.082**

注：加粗字体表示最优结果。

表 5 展示了本文方法在 CASIA-FASD 数据集上与主流方法的对比，无论是与结合了时间维度信息的方法（Asim 等，2017）还是与采用了训练策略的方法（Rehman 等，2018）对比，本文方法都取得了更好的结果。

表 5　在 CASIA-FASD 数据集上本文方法与主流方法对比

Table 5　Comparison of our method with others on CASIA-FASD dataset

/%

方法	EER
DoG(Zhang 等,2012)	17
基于图像质量(Galbally 和 Marcel,2014)	32.4
基于 CNN(Yang 等,2014)	5.02
CNN + LSTM(Xu 等,2015)	5.17
多特征(Patel 等,2016)	5.88
LBP-TOP(Asim 等,2017)	8.02
CNN + 训练策略(Rehman 等,2018)	4.59
3DCNN(Gan 等,2017)	5.25
本文	**1.70**

注:加粗字体表示最优结果。

表 6 展示了本文方法在 replay-attack 数据集上与主流方法的对比,无论是与基于图像质量(Galbally 和 Marcel,2014)的方法、结合了时间维度信息的方法(Asim 等,2017)还是与采用了一些训练策略(Rehman 等,2018)的方法对比,本文方法都取得了更好的结果。

表 6　在 replay-attack 数据集上本文方法与主流方法对比

Table 6　Comparison of our method with others on replay-attack dataset

/%

方法	EER	HTER
LBP-TOP(Asim 等,2017)	3.22	4.7
基于分块(Li 等,2016)	2.87	6.11
多特征融合(Patel 等,2016)	–	14.6
CNN(Rehman 等,2018)	–	5.74
图片质量(Galbally 和 Marcel,2014)	–	15.2
小网络做初始化(de Souza 等,2018)	0.33	2.5
LBP(Chingovska 等,2012)	13.9	13.8
本文	**0.091**	**0.082**

注:加粗字体表示最优结果,“–”代表无数据。

3.4.2　实验结果分析

实验结果表明,本文方法取得了有效结果。相较于过去的方法,本文提出的混合池化结构能够在特征层面进行全局和局部信息的学习,空间金字塔和全局平均混合池化能够对特征进行多尺度和大尺度的划分,获得更加丰富的感受野。双流结构中,空间网络用来学习图像的空间特性,能够检测出一些空间上的欺诈特征,比如打印边框、手机边框等。而时间网络用来学习视频的时间特性,能够检测出一些时间上的欺诈特征,比如眼动、嘴动等。双流结构对空间信息和时间信息分别建模,通过参数来调节两个模型的占比,这样能够对空间和时间的有效性进行区分。

3.5　结果可视化

卷积网络一直被视为“黑盒”,因为看不到卷积神经网络的内部结构,而可视化方案很好地解决了这一问题。可视化方案增大了结果的可解释性,通过分析可视化结果,也能为未来的研究提供参考。图 7 展示了真脸、打印攻击和视频攻击 3 个样本及在网络中输出的特征图。从特征图可以看出,打印攻击会造成图像质量降低,相较于真脸,打印攻击的特征图会产生模糊现象;视频攻击会产生镜面反射,相较于真脸,视频攻击的特征图产生了大块的亮斑。

(a) 真脸及特征图　(b) 打印攻击及特征图 (c) 视频攻击及特征图

图 7　特征图的可视化

Fig. 7　Feature visualization((a) true face and feature maps; (b) print attack and feature maps; (c) video attack and feature maps)

图 8 使用类激活图的方式展示了在不同攻击方式下卷积网络的关注点,图中红色代表关注度大的点,可以看出关注点都在边缘区域,原因是边缘部位更容易区分。打印的照片边缘是长方形,视频的边缘有黑边,但是真脸的边缘是光滑曲线,可见卷积神经网络聪明地学习到了关键区域。但是这种方法也有弊端,如果遇见离摄像头比较近的样本,就不会出现边缘,可能会导致误判。集中所有注意力在边缘的学习,也会影响到卷积神经网络学习到更加具有

代表性以及更加本质的特征。

(a) 真脸　(b) 打印攻击　(c) 视频攻击

图 8　使用类激活图进行可视化

Fig. 8　Visualization with class activation maps

((a) true face; (b) print attack; (c) video attack)

4　结　论

本文结合空间维度以及时间维度,提出了结合混合池化的双流活体检测网络结构模型。空间流用来学习图片的空间特征,并且对不同颜色空间图片的空间流结果对比选择出最适合的颜色空间。空间金字塔和全局平均混合池化是通过在网络里面加入混合池化块来进行的,空间金字塔池化可以通过多尺度的学习,获得更加丰富的感受野,全局平均池化能够充分学习全局信息并且结合类激活图能够做一个很好的可视化工作。时间流利用光流图片来获取时间维度的信息,与空间流相互结合达到加强结果的目的。

通过对比仅使用空间流、仅使用时间流以及二者融合的结果,可以看出二者融合后确实能够对结果进一步加强。通过对基础网络和加入混合池化模块网络的对比,可以看出混合池化模块的有效性。实验结果表明,本文方法在错误率以及半错误率上都优于当前的主流方案。

通过对空间流的结果和时间流结果的对比,可以看出当前的工作中,虽然时间流能够起到辅助作用,但是其结果相对空间流还有一定差距。产生该情况的原因主要是光流图片生成的质量可能不高,以及卷积网络对光流这种比较细微的线索不太容易学习。

在今后的研究中,将进一步考虑产生更加高质量的光流图片,并设计出更加合理的并能够对细微线索放大的网络结构。

参考文献(References)

Arashloo S R, Kittler J and Christmas W. 2017. An anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol. IEEE Access, 5: 13868-13882 [DOI: 10.1109/ACCESS.2017.2729161]

Arnheim R and Gibson J J. 1952. The perception of the visual world. Journal of Aesthetics and Art Criticism, 11(2): 172-173 [DOI: 10.2307/426044]

Asim M, Ming Z and Javed M Y. 2017. CNN based spatio-temporal feature extraction for face anti-spoofing//Proceedings of the 2nd International Conference on Image, Vision and Computing (ICIVC). Chengdu, China: IEEE: 234-238 [DOI: 10.1109/ICIVC.2017.7984552]

Boulkenafet Z, Komulainen J and Hadid A. 2016. Face spoofing detection using colour texture analysis. IEEE Transactions on Information Forensics and Security, 11(8): 1818-1830 [DOI: 10.1109/tifs.2016.2555286]

Boulkenafet Z, Komulainen J and Hadid A. 2017. Face anti-spoofing using speeded-up robust features and fisher vector encoding. IEEE Signal Processing Letters, 24(2): 141-145 [DOI: 10.1109/LSP.2016.2630740]

Chingovska I, Anjos A and Marcel S. 2012. On the effectiveness of local binary patterns in face anti-spoofing//Proceedings of 2012 International Conference of Biometrics Special Interest Group (BIOSIG). Darmstadt, Germany: IEEE: 1-7

Dalal N and Triggs B. 2005. Histograms of oriented gradients for human detection//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego, CA, USA: IEEE: 886-893 [DOI: 10.1109/CVPR.2005.177]

de Souza G B, Papa J P and Marana A N. 2018. On the learning of deep local features for robust face spoofing detection[EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1806.07492.pdf

Duc N M and Minh B Q. 2009. Your face is not your password//Black Hat Conference. [EB/OL]. [2009-08-15]. https://www.blackhat.com/presentations/bh-dc-09/Nguyen/BlackHat-DC-09-Nguyen-Face-not-your-password-slides.pdf

Erdem C E, Ulukaya S, Karaali A and Erdem A T. 2011. Combining haar feature and skin color based classifiers for face detection//Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing. Prague, Czech Republic: IEEE: 1497-1500 [DOI: 10.1109/ICASSP.2011.5946777]

Erdogmus N and Marcel S. 2014. Spoofing face recognition with 3D masks. IEEE Transactions on Information Forensics and Security, 9(7): 1084-1097 [DOI: 10.1109/TIFS.2014.2322255]

Galbally J and Marcel S. 2014. Face anti-spoofing based on general image quality assessment//Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE: 1173-1178 [DOI: 10.1109/ICPR.2014.211]

Gan J Y, Li S L, Zhai Y K and Liu C Y. 2017. 3D convolutional neural network based on face anti-spoofing//Proceedings of the 2nd International Conference on Multimedia and Image Processing (ICMIP).

Wuhan, China: IEEE: 1-5 [DOI: 10.1109/ICMIP.2017.9]

He K M, Zhang X Y, Ren S Q and Sun J. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916 [DOI: 10.1109/TPAMI.2015.2389824]

Jourabloo A, Liu Y J and Liu X M. 2018. Face de-spoofing: anti-spoofing via noise modeling//Proceedings of the 15th European Conference Computer Vision ECCV 2018. Munich, Germany: Springer: 297-315 [DOI: 10.1007/978-3-030-01261-8_18]

Karthik K and Katika B R. 2017. Image quality assessment based outlier detection for face anti-spoofing//Proceedings of the 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA). Mumbai, India: IEEE: 72-77 [DOI: 10.1109/CSCITA.2017.8066527]

Kollreider K, Fronthaler H, Faraj M I and Bigun J. 2007. Real-time face detection and motion analysis with application in "Liveness" assessment. IEEE Transactions on Information Forensics and Security, 2(3): 548-558 [DOI: 10.1109/TIFS.2007.902037]

Li L, Feng X Y, Boulkenafet Z, Xia Z Q, Li M M and Hadid A. 2016. An original face anti-spoofing approach using partial convolutional neural network//Proceedings of the 6th International Conference on Image Processing Theory, Tools and Applications (IPTA). Oulu, Finland: IEEE: 1-6 [DOI: 10.1109/IPTA.2016.7821013]

Lin M, Chen Q and Yan S C. 2014. Network in network [EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1312.4400.pdf

Liu Y J, Jourabloo A and Liu X M. 2018. Learning deep models for face anti-spoofing: binary or auxiliary supervision//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 389-398 [DOI: 10.1109/CVPR.2018.00048]

Liu Y J, Stehouwer J, Jourabloo A and Liu X M. 2019. Deep tree learning for zero-shot face anti-spoofing [EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1904.02860.pdf

Lucas B D and Kanade T. 1997. An iterative image registration technique with an application to stereo vision//Proceedings of the 7th International Joint Conference on Artificial Intelligence. San Francisco: ACM: 674-679

Lucena O, Junior A, Moia V, Souza R, Valle E and Lotufo R. 2017. Transfer learning using convolutional neural networks for face anti-spoofing//Proceedings of the 14th International Conference on Image Analysis and Recognition. Montreal, QC, Canada: Springer: 27-34 [DOI: 10.1007/978-3-319-59876-5_4]

Nikisins O, Mohammadi A, Anjos A and Marcel S. 2018. On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing//Proceedings of 2018 International Conference on Biometrics (ICB). Gold Coast, QLD: IEEE: 75-81 [DOI: 10.1109/ICB2018.2018.00022]

Pan G, Sun L, Wu Z H and Lao S H. 2007. Eyeblink-based anti-spoofing in face recognition from a generic webcamera//Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro: Brazil: IEEE: 1-8 [DOI: 10.1109/ICCV.2007.4409068]

Patel K, Han H and Jain A K. 2016. Secure face unlock: spoof detection on smartphones. IEEE Transactions on Information Forensics and Security, 11(10): 2268-2283 [DOI: 10.1109/TIFS.2016.2578288]

Pérez-Cabo D, Jiménez-Cabello D and Costa-Pazo A and López-Sastre R J. 2019. Deep anomaly detection for generalized face anti-spoofing [EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1904.08241.pdf

Rehman Y A U, Po L M and Liu M Y. 2018. LiveNet: improving features generalization for face liveness detection using convolution neural networks. Expert Systems with Applications, 108: 159-169 [DOI: 10.1016/j.eswa.2018.05.004]

Schwartz W R, Rocha A and Pedrini H. 2011. Face spoofing detection through partial least squares and low-level descriptors//Proceedings of 2011 International Joint Conference on Biometrics (IJCB). Washington DC, USA: IEEE: 1-8 [DOI: 10.1109/IJCB.2011.6117592]

Taigman Y, Yang M, Ranzato M A and Wolf L. 2014. DeepFace: closing the gap to human-level performance in face verification//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE: 1701-1708 [DOI: 10.1109/CVPR.2014.220]

Tan X Y, Li Y, Liu J and Jiang L. 2010. Face liveness detection from a single image with sparse low rank bilinear discriminative model//Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece: Springer: 504-517 [DOI: 10.1007/978-3-642-15567-3_37]

Wen D, Han H and Jain A K. 2015. Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security, 10(4): 746-761 [DOI: 10.1109/tifs.2015.2400395]

Xu Z Q, Li S and Deng W H. 2015. Learning temporal features using LSTM-CNN architecture for face anti-spoofing//Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition (ACPR). Kuala Lumpur, Malaysia: IEEE: 141-145 [DOI: 10.1109/ACPR.2015.7486482]

Yang J W, Lei Z and Li S Z. 2014. Learn convolutional neural network for face anti-spoofing [EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1408.5601.pdf

Zhang K P, Zhang Z P, Li Z F and Qiao Y. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10): 1499-1503 [DOI: 10.1109/lsp.2016.2603342]

Zhang P, Zou F H, Wu Z W, Dai N L, Mark S, Fu M, Zhao J and Li K. 2019. FeatherNets: convolutional neural networks as light as feather for face anti-spoofing [EB/OL]. [2019-08-15]. https://arxiv.org/pdf/1904.09290.pdf

Zhang Z W, Yan J J, Liu S F, Lei Z, Yi D and Li S A. 2012. A face antispoofing database with diverse attacks//Proceedings of the 5th IAPR International Conference on Biometrics (ICB). New Delhi, India: IEEE: 26-31 [DOI: 10.1109/ICB.2012.6199754]

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]

作者简介

汪亚航,1994 年生,男,硕士研究生,主要研究方向为图像识别与深度学习。
E-mail: coderwangson@163.com

宋晓宁,通信作者,男,教授,主要研究方向为模式识别、高维图像处理和计算机视觉。
E-mail: x.song@jiangnan.edu.cn

吴小俊,男,教授,博士生导师,主要研究方向为模式识别、计算智能、计算机视觉、信息融合。
E-mail: wu_xiaojun@jiangnan.edu.cn