计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200300128

基于注意力的热点块和显著像素卷积神经网络的人脸防伪方法

吴晓丽　胡　伟
北京化工大学信息科学与技术学院　北京 100029
(xlwu@mail.buct.edu.cn)

摘　要　人脸防伪用于验证被测试者是否为真实活体,是计算机视觉领域的一个研究热点。攻击手段的多样性以及人脸识别主要在嵌入式、移动式等不具备高计算能力的设备上应用,使得快速有效的人脸防伪计算成为具有挑战性的任务。针对该问题,文中提出了一种基于注意力的热点块和显著像素卷积神经网络的方法。其中,热点块机制以对5个热点块的判别来取代对整张人脸的判别,显著降低了计算量,迫使网络模型集中关注更具有鉴别信息的热点块,提高了网络模型的准确率;显著像素方法对输入的人脸图像进行显著像素预测,通过判断显著预测图是否符合人脸的深度特性来鉴别活体与攻击。该方法将热点块与显著像素的结果进行融合,充分发挥了局部特征和全局特征的作用,进一步提升了人脸防伪的效果。与现有方法相比,所提方法在CASIA-MFSD、Replay-Attack以及SiW数据集上都达到了很好的效果。

关键词:人脸防伪;活体检测;注意力机制;热点块;显著像素;卷积神经网络
中图法分类号　TP183

Attention-based Hot Block and Saliency Pixel Convolutional Neural Network Method for Face Anti-spoofing

WU Xiao-li and HU Wei
College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Abstract　Face anti-spoofing is used to verify whether the testee is a real person. The diversity of attack methods and the application of face recognition on various embedded and mobile devices with low computing capabilities have made face anti-spoofing a very challenging task. Aiming at face anti-spoofing, an attention-based hot block and saliency pixel convolutional neural network method is proposed. The hot block method replaces the discrimination of the entire face with the determination of 5 hot blocks, which not only reduces the amount of calculation, but also forces the network to focus on hot spots with more discerning information, so as to improve the accuracy of the network. On the other hand, the saliency pixel method performs saliency pixel prediction on the input face image to determine whether the saliency prediction map meets depth characteristics of the face to identify the liveness and the attack. This method fuses the results of hot blocks and saliency pixels to give full play to the role of local features and global features, and further enhances the effect of face anti-spoofing. Compared with existing methods, the proposed method has achieved good results on CASIA-MFSD, Replay-Attack and SiW datasets.

Keywords　Face anti-spoofing, Liveness detection, Attention mechanism, Hot block, Saliency pixel, Convolutional neural network

1　引言

随着科技的发展,人脸识别系统因其便利性、快捷性等优点渐渐取代了传统的指纹识别系统。近年来兴起的支付宝“刷脸支付”、苹果面容ID解锁、人脸闸机等人脸识别系统已成为我们日常工作和生活的重要组成部分,承担着甄别欺诈行为、保障用户利益的责任。

实际上,验证被测试者是否为真实活体被称为活体检测或人脸防伪,是独立于人脸识别的研究课题。在实际身份验证场景中,活体检测一般嵌套在人脸检测与人脸识别的模块中,是人脸识别前的关键一步,可有效抵御照片、屏幕翻拍、面具等常见的攻击手段,从而保障用户利益。然而,攻击手段丰富多样,除了上述常见的攻击手段外,还不断出现了如3D面具、化妆、换脸等新型攻击手段。攻击手段的多样性使得活体检测本身具有一定的复杂性,加之人脸识别主要应用在各类嵌入式、移动式等不具备高计算能力的设备上,这意味着活体检测面临着更严峻的挑战。

现有的人脸防伪方法主要分为两个方向。1)早期的人工设计特征方法。此类方法通过人眼观察活体与非活体的差异来进行特征的设计,再用分类器判别提取的特征。活体与非活体的差异主要表现在以下几个方面:颜色、纹理等材料方面的特性(如皮肤与纸张就有显著的材料特性差异);表情变化;图像或视频质量;非刚性运动变形。鉴于这些差异,早期的方法可以分为3类:基于纹理的方法;基于运动的方法;基于图

到稿日期:2020-03-23　返修日期:2020-07-04　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。
通信作者:胡伟(huwei@mail.buct.edu.cn)

像质量和反射的方法。基于纹理的方法通过发掘各种攻击媒介特有的纹理特征来实现，其缺点在于难以提取具有鲁棒性的纹理特征；基于运动的方法通过检测面部微表情（如眨眼、嘴部微动作）来达到对人脸视频分类的目标，其缺点在于只适用于静态攻击而不适用于动态攻击（如重放攻击、面具攻击）；基于图像质量和反射的方法通过设计特征来捕获攻击图片中叠加的照明和噪声信息，以鉴别活体和非活体，随着现在攻击媒介的分辨率的提高，该类噪声减少，对这类设计特征形成了挑战。2）基于深度学习的方法。研究人员在强大的神经网络的基础上进行了各种尝试，如加入其他信息进行辅助监督，用个人数据集对网络进行预训练或微调，探索将攻击样本分离出真人＋攻击噪声的方法以及提出新型深度网络结构，皆是最新的研究成果。虽然有不少基于深度学习的方法在效果上已经超越了人工设计特征的方法，但其计算效率仍有待提高。加之人脸识别的广泛应用，使得在实际应用中人们会期望实际模型运算得越快越好。此外，虽然用大型数据集训练的卷积神经网络经过时间的推移可以自发聚焦于人脸图像的重要区域，但对于小型数据集，在数据量不足的情况下，则需要引入注意力机制对脸部重要区域和重要像素进行强化，以提升网络模型对人脸真伪判断的效果。

基于以上观点，本文提出了基于注意力的热点块和显著像素卷积神经网络的人脸防伪方法。基于注意力的热点块方法将注意力集中在 5 个重要的脸部区域小块上，以对 5 个热点块的判别来取代对整张人脸的判别，显著降低了计算量，迫使网络模型集中关注更具鉴别信息的热点块，从而提高了网络的准确率。基于注意力的显著像素方法对人脸的显著像素点进行预测，通过判断显著预测图是否具有人脸的深度特性来鉴别活体和攻击。所提方法将热点块方法与显著像素的结果进行融合，充分发挥了局部特征和全局特征的作用，进一步提升了效果。本文方法使得网络聚焦于脸部的重要区域，不仅在小型数据集上证明了两个注意力机制的有效性，还在大型数据集上取得了很好的结果。本文的贡献如下。

（1）提出了一种基于注意力机制的热点块人脸防伪方法。针对因数据量不够导致模型无法精准聚焦人脸重要区域的问题，通过强化网络对脸部重要区域的关注，在小型数据集上对准确率进行有效提升。

（2）人为设计的热点块是先行假定规则形状和固定大小的，但实际上存在注意力区域为不规则形状和不固定范围的可能性。针对这一问题，本文提出了显著像素人脸防伪方法。显著像素方法通过估计每个像素点的兴趣强度来实现分类，这与综合了位置信息和局部 RGB 特征的热点块方法相比，有着明显的不同。该方法与热点块方法进行融合后，效果比使用单一方法更好。

（3）为了验证本文方法的有效性，本文分别在 CASIA-MFSD，Replay-Attack 以及 SiW 数据集上进行了对比实验。实验结果表明，与现有方法相比，无论是在 CASIA-MFSD 和 Replay-Attack 这两个小型数据集上，还是在大规模的 SiW 数据集上，本文方法都取得了非常具有竞争力的结果。

2 相关工作

早期的人脸防伪方法，首先通过直接观察活体与非活体的差异，然后根据差异设计特征，最后将特征送入分类器进行决策。经典的人工设计特征包括 LBP[1-3]，HOG[4-5]，SIFT[6] 和 SURF[7]。同时，研究人员还探寻不同输入域的解决方案，如 HSV，YCbCr 色彩空间[8-9]和傅里叶光谱[10]。除了上述单帧的方法外，基于运动的方法通过检测连续帧的眨眼[11-12]和嘴唇运动[13]微表情来鉴别活体和非活体。

在卷积神经网络发展初期，研究人员尝试部分使用卷积神经网络[14-17]，把神经网络仅当作特征提取器或是特征分类器。在 Patel 等[15]和 Li 等[16]的工作中，预先训练的 CaffeNet 和 VGG-face 模型被用作特征提取器。在 Li 等[16]和 Feng 等[17]的工作中，设计不同的人脸图像输入形式，将多尺度的人脸图像或人工特征输入到卷积神经网络中，对活体与非活体进行直接分类。

后续对深度学习方法的研究，按照输入形式划分，可分为单帧的方法、多帧的方法和多模态的方法（输入为 RGB 图像、近红外图像、深度图像的组合）。

（1）单帧方法。Atoum 等[18]首先考虑把人脸深度图作为活体与非活体的差异特征，用随机人脸块以及人脸预测深度图的融合结果来鉴别活体和非活体。本文所提方法在思想上与该方法类似，该方法使用随机块机制来提取局部纹理特征并利用全连接卷积神经网络（Fully Convolutional Network，FCN）进行深度图像的预测。本文所提方法的优势在于热点块机制具备更充分的理论和实验支持以及选择了速度更快的网络结构进行显著预测。Jourabloo 等[19]提出了一种新思路，将攻击图像分离出真人图像＋伪造噪声，把分离出的攻击噪声作为分类辅助信息。此外，Liu 等[20]设计了一个用于检测未知欺骗攻击的深度树网络，并定义了一个新型人脸防伪指标 ZSFA（Zero-Shot Face Anti-spoofing）以检测模型对从未见过的攻击类型的应对能力。

（2）多帧方法。在 Xu 等[21]、Yang 等[22]和 Liu 等[23]的工作中，利用长短期记忆卷积神经网络（Long Short-Term Memory Convolutional Neural Networks，LSTM-CNN）结构，将视频的连续帧作为输入，充分发挥时序特征的作用。Liu 等[23]利用远程光体积描记法（rPPG、心脏脉冲信号）和深度信息来对网络进行辅助监督。特别地，Yang 等[22]还加入了区域注意力机制（Region Attention Module，RAM）学习人脸的重要区域。

（3）多模态方法。Zhang 等[24]将多种模态的人脸图像（RGB 图像、NIR 图像、深度图像）作为输入，提出了适用于多模态的融合方法。

同时，Yang 等[22]认为人脸防伪任务不仅取决于模型的好坏，还取决于数据集。早期的 CASIA-MFSD 以及 Replay-Attack 数据集的采集设备早已过时，这些数据集规模太小，不适用于深度学习。近两年发布了 3 个人脸防伪数据集，包括 Liu 等[23]提出的 SiW 数据集，其采集设备更新，且更大型；Liu 等[20]提出的 SiW-M 数据集，其包含更多的攻击类型，适配 ZSFA 指标；Zhang 等[24]提出的 CASIA-SURF 数据集，适用于多模态训练，证明了用该多模态数据集训练后再用目标数据集进行微调比只用目标数据集训练的效果更好。

本文所提方法结合了区域分块和辅助信息的思想并引入

了注意力机制，并通过将注意力聚焦于人脸的重要区域块和重要像素来实现。

3 方法

3.1 总体框架概述

如图 1 所示，本文提出的方法包括两个处理流程：基于注意力的热点块和显著像素方法。基于注意力的热点块方法通过对人脸的 68 个区域的预训练，选取表现最好的 5 个区域作为热点块送入 ShuffleNet[25]，计算 5 个热点块的活体分数的平均值，该值即为热点块方法的得分。基于注意力的显著像素方法首先利用训练好的 PoolNet[26] 对输入图像进行显著像素预测，再将显著预测图送入一个 5 层的 CNN 进行分类，即得到显著像素方法的得分。最后，将上述两种方法得到的结果按照合适的权重进行融合，得到最终得分以判断输入是否为真实人脸。

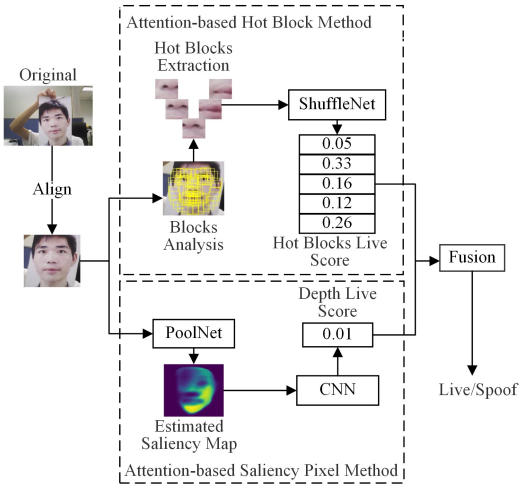


图 1 本文提出的人脸防伪方法的框架图

Fig. 1 Framework of the proposed face anti-spoofing approach

3.2 注意力机制

人眼会自发地聚焦于图片上的关注区域而忽视不值得被关注的区域，同样地，对于人脸防伪，并非所有的人脸区域都值得被网络关注。热点块方法的注意力机制通过对人脸的关键区域进行强化，从而实现网络效果的提升。显著像素方法的注意力机制通过对人脸的显著像素点进行预测，来判断显著预测图是否符合人脸深度特性，从而发挥全局特征的作用。由热点块的分析结果可知，鼻子和嘴部包含了大量的区分性信息。由显著像素的预测图的分析结果可知，下颌区域更值得被关注。

3.3 基于注意力的热点块方法

3.3.1 具体流程

基于注意力的热点块方法，首先计算输入人脸图像的 68 个关键点，并以这些关键点为中心进行固定大小的块截取；然后将其送入网络中进行预训练，并记录每个 epoch 的平均准确率以及各个块的准确率；最后基于记录的准确率，应用热点块选取算法选取表现最佳的 5 个热点块送入网络训练，将 5 个热点块的平均得分作为输入人脸的最终分数。

针对热点块的选取，本文进行了块尺寸实验，分别用 96×96，48×48，24×24 的块进行实验，发现尺寸 24×24 的效果最好，且计算量较小。

3.3.2 注意力机制在热点块方法中的应用

注意力机制主要表现在热点块的选取上。热点块的选取步骤如下：1）剔除平均准确率未超过 0.9 的 epoch 信息；2）将超过当前 epoch 准确率的块的权重设置为 1.1，未超过当前 epoch 准确率但排名前 7 的块的权重设置为 1，余下的块的权重设置为 0.9；3）计算每个块的加权平均分数以及方差；4）以加权平均分数为主，方差为辅，选出排名 Top5 的块作为热点块。

3.3.3 不同数据集的热点块分析

基于上述热点块的选取算法，在不同数据集上进行热力图的绘制并描出人脸 68 个关键点的相应位置，如图 2 所示。部分面颊区域不具有热度的原因在于本文热点块的大小固定为 24×24，以关键点为中心进行截取，导致部分面颊区域没有被覆盖到。

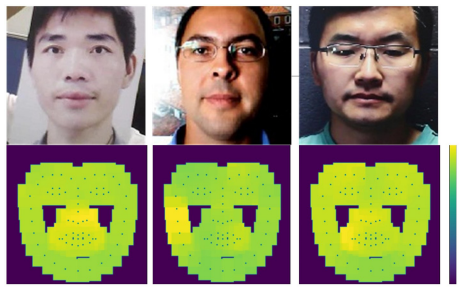


图 2 CASIA-MFSD(左列)、Replay-Attack(中间列)以及 SiW(右列)数据集上的样例图和热力图

Fig. 2 Samples and heatmaps on CASIA-MFSD(left col), Replay-Attack(middle col) and SiW(right col) databases

由图 2 可知，CASIA-MFSD 数据集[27] 的热点区域集中于鼻子以及嘴部；SiW 数据集[23] 的热点区域集中于鼻子以及右嘴角；Replay-Attack 数据集[28] 的热点区域集中于人脸的右边缘以及左嘴角。CASIA-MFSD 数据集在采集时无额外光照、无表情变化、无摆动；SiW 数据集在采集时有额外光照、有表情变化、有一定摆动；Replay-Attack 数据集在采集时有额外光照(分为可控光和逆光)、无表情变化、无摆动。CASIA-MFSD 和 SiW 数据集的热点区域集中于鼻子和嘴巴，这说明鼻口区域在人脸防伪任务中是更值得关注的区域。Replay-Attack 数据集的热点区域与 CASIA-MFSD 和 SiW 的热点区域不同的原因在于：采集光照的特殊性导致在邻近脸部的右边缘的背景处出现了活体与攻击的明显差异，该差异的主要表现取决于该区域背景是否反光。

本文基于同一个模型在 CASIA-MFSD 和 Replay-Attack 数据集上进行了 3 组对比实验：5 个随机块、五官块、Top5 热点块。实验效果与预期一致，准确率从高到低依次为 Top5 热点块＞五官块＞5 个随机块，这充分证明了热点块机制的有效性。

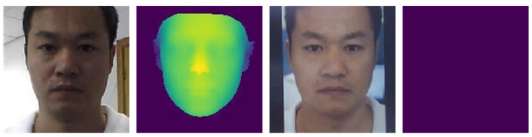
在 CASIA-MFSD，Replay-Attack，SiW 数据集上使用各自的热点块，帧的准确率分别达到了 96.95%，99.05%，97.91%。特别地，在 Replay-Attack 上使用鼻口区域的热点块进行实验时，准确率达到了 97.43%，这证明了鼻口区域针

对人脸防伪任务具有一定的通用性。

3.4 基于注意力的显著像素方法

3.4.1 具体流程

假定活体人脸的显著像素分布具有人脸的深度特性，攻击人脸的显著像素表现为全平面。因此，本文使用 PRNet[29] 生成活体图像的显著标签，对攻击图像直接生成扁平标签，如图 3 所示。将人脸图像和相应的显著标签成对地投入到 PoolNet 中进行训练，将训练好的模型作为显著预测器，对测试的人脸图进行显著像素预测。通过对显著预测图进行分类，得到显著方法的得分。



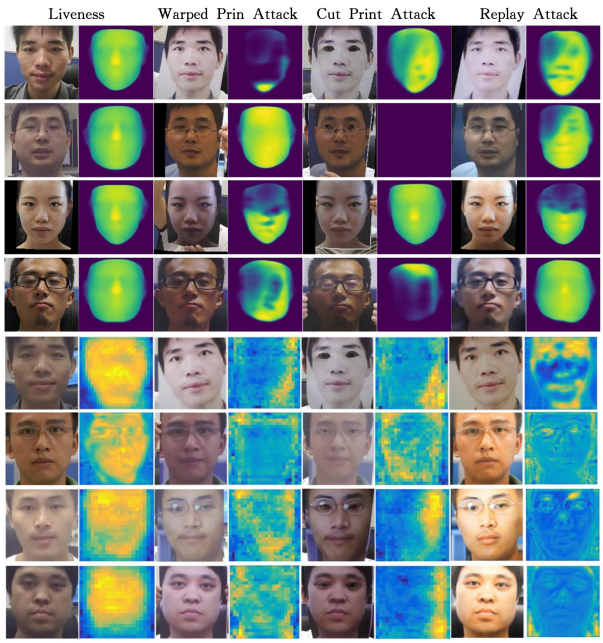
注：前两个图像是活体图像以及使用 PRNet 生成的相应的显著标签。剩下两个图像是攻击图像以及相应的扁平显著标签

图 3 CASIA-MFSD 训练集上的显著标签

Fig. 3 Saliency labels on CASIA-MFSD training dataset

3.4.2 注意力机制在显著像素方法中的应用

注意力机制主要表现在预测显著像素图上，这部分工作借助网络完成。由图 4 前 4 行给出的使用 PoolNet 生成的显著预测图可知，活体的显著预测图具有清晰的脸部深度信息和边缘，攻击的显著预测图大多不具备正常的人脸形状且更加平坦。预测网络对活体与非活体一视同仁地进行了显著性像素的预测，使得显著预测图与图 3 所示的显著标签具有一定的差距。攻击图像的显著预测图与活体图像的显著预测图之间存在着明显的差异，该差异为级联分类器鉴别活体与非活体提供了依据。



注：第 1、第 2 列是活体人脸图像及其对应的显著预测图，其余 6 列是 3 种不同类型的攻击图像（扭曲的打印、剪切的打印以及重放攻击）及其对应的显著预测图。第 1 行－第 4 行是本文的显著预测结果，后 4 行是 Atoum 等[18]的深度预测结果

图 4 在 CASIA-MFSD 测试集上的显著预测图

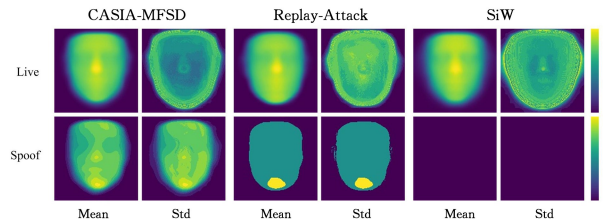
Fig. 4 Saliency estimation on CASIA-MFSD testing dataset

图 4 后 4 行为 Atoum 等[18]使用 FCN 生成的深度预测图，这些深度预测图有大片的马赛克，因此本文探寻一种具有更好预测效果的网络。PoolNet 在显著目标检测领域取得了成功，考虑深度的预测与显著目标的检测具有一定的共通之处，本文预计 PoolNet 在人脸显著预测上也能做到像素级的精准预测。

相比 FCN，PoolNet 不仅在预测图上具有更清晰的形状表现，还在 4.3.2 节的实验中取得了更好的结果。

3.4.3 关于预测显著图的分析

图 5 给出了基于 Replay-Attack，CASIA-MFSD，SiW 测试集统计的活体人脸和攻击人脸的显著预测图的均值以及标准差。活体与非活体的显著预测图在均值和标准差上的差异证明了显著预测图具备区分活体与非活体的能力。在 CASIA-MFSD，Replay-Attack，SiW 数据集上，显著像素预测方法在帧的准确率上分别达到了 97.30%，98.42%，99.03%，这充分证明了显著像素预测方法的有效性。从 3 个数据集的均值、标准差以及准确率来看，显著像素预测方法更适用于大型数据集。



注：第 1 列－第 2 列、中间 2 列、最后 2 列分别是 CASIA-MFSD、Replay-Attack、SiW 数据集的所有的测试样本的显著预测图的统计。第 1 行是活体人脸的显著预测图的均值和标准差，第 2 行是攻击人脸的显著预测图的均值和标准差

图 5 活体和攻击人脸的显著预测图的均值和标准差

Fig. 5 Mean and standard deviation of the estimated saliency maps of live and spoof faces

3.4.4 对显著预测图进行分类的网络

本文使用 5 层卷积神经网络对显著预测图进行分类。该网络包含 4 个卷积层和 1 个全连接层；每个卷积层后接批处理归一化、ReLU 和池化层。将该网络的损失定义为正则化损失和交叉熵损失之和。

4 实验验证

4.1 实验数据

为了验证本文方法的有效性，分别在 CASIA-MFSD，Replay-Attack，SiW 数据集上进行实验。

(1)CASIA-MFSD 数据集。该数据集是 2012 年发布的视频数据集，包括主题数 50 个，场景 3 个，总计 150 个活体视频以及 450 个攻击视频，全部为正脸，无表情变化，无额外光照。显示设备有且仅有 iPad，攻击手段按照媒介划分，可分为 1 种打印攻击和 1 种重放攻击。

(2)Replay-Attack 数据集。该数据集是 2012 年发布的视频数据集，包括主题数 50 个，场景 1 个，总计 200 个活体视频以及 1000 个攻击视频，全部为正脸，无表情变化，有额外光照。显示设备有 iPhone 3GS，iPad，攻击手段按照媒介划分，可分为 1 种打印攻击和 2 种重放攻击。

(3)SiW 数据集。该数据集是 2018 年发布的视频数据集,包括主题数 165 个,场景 4 个,总计 1 320 个活体视频以及 3 300 个攻击视频,人脸摆动范围为[−90°,90°],有表情变化,有额外光照。显示设备有 iPad Pro,iPhone 7,Galaxy S8,Asus MB168B,攻击手段按照显示设备划分,可分为 2 种打印攻击和 4 种重放攻击。

4.2 参数设置

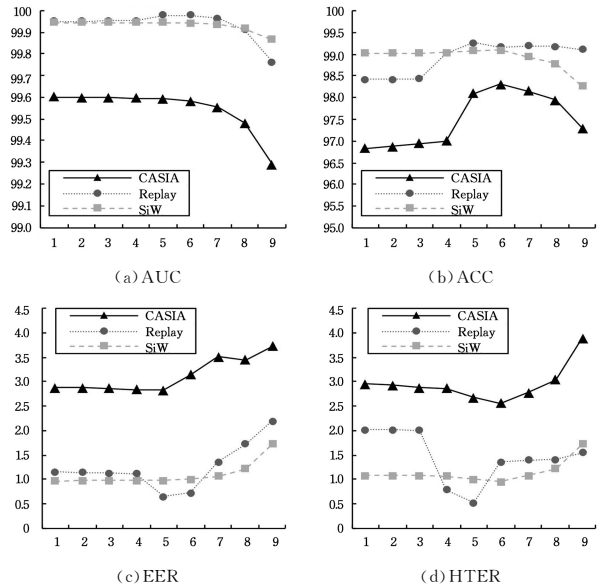
4.2.1 模型参数

热点块使用的网络采用 Tensorflow 实现框架,使用批处理,批处理大小为 50,学习率为 0.001,epoch 数目为 100。显著像素方法中用于预测显著图的网络采用 Pytorch 实现框架,学习率为 0.00005,在第 15 个 epoch 之后的衰减率为0.1,epoch 数目为 30;显著像素方法中用于分类的 5 层 CNN 采用 Tensorflow 实现框架,使用批处理,批处理大小为 50,学习率为 0.001,epoch 数目为 100。

无论是基于热点块的方法还是基于显著像素的方法,输入特征皆为对齐后的人脸,对齐方法由 Wu 等[30]提出,并且仅使用 RGB 色彩空间。热点块的大小固定设置为 24×24。

4.2.2 融合权重

为了确定合适的融合权重,根据 AUC(Area Under Curve)、准确率(ACC)、等错误率(EER)、半错误率(Half Total Error Rate,HTER)这 4 个指标在 CASIA-MFSD,Replay-Attack,SiW 数据集上分别进行了关于权重的内部实验。热点块方法和显著像素方法的融合权重分别为(1:9),(2:8),(3:7),(4:6),(5:5),(6:4),(7:3),(8:2),(9:1),依次编号为第 1 组—第 9 组,实验结果如图 6 所示。



注:第 1 组—第 9 组,热点块方法和显著像素方法的融合权重依次为(1:9),(2:8),(3:7),(4:6),(5:5),(6:4),(7:3),(8:2),(9:1)。评价指标包括 AUC,ACC,EER 和 HTER

图 6 不同融合权重的结果比较

Fig. 6 Comparison of the results of different fusion weights

由图 6 可知,针对 AUC、ACC、EER 以及 HTER 这 4 个指标,第 5 组和第 6 组的融合权重的效果最为显著。在第 5 组和第 6 组中,本文选择了 EER 更低的第 5 组(5:5)作为最终的融合比例。这是因为在这 4 个指标中,EER 是使错误接受率和错误拒绝率相对平衡的阈值点,是生物识别安全系统中最常用的一种评价标准。EER 低意味着被测系统既注重安全性又注重便利性。因此,EER 越低,被测系统越具有实际应用的价值,故本文最终选用的融合权重为 1:1。

4.3 方法有效性分析

4.3.1 内部阶梯实验

为了验证本文方法的有效性,根据 ACC 和 HTER 这两个指标进行了基于帧的阶梯实验:1)整脸输入本地复现的 Atoum 提出的块网络[18]作为基准;2)整脸输入本文使用的块网络;3)在本文使用的块网络上应用热点块机制;4)在热点块实验结果的基础上融合显著像素结果。

表 1 和表 2 列出了在 CASIA-MFSD 和 Replay-Attack 数据集上的内部比较结果。可以看出,通过逐步加入热点块以及融合显著像素的结果,ACC 呈阶梯式增长,HTER 呈阶梯式降低,经过融合后效果达到最佳。因为在数据预处理时对视频的每一帧进行截取,帧之间存在高度的相似性,所以热点块的使用可以削弱过度拟合,与整脸训练相比,准确率得到了有效提升。在融入显著像素的结果后,效果又进一步得到了提升,这证明了本文所提方法的有效性。

表 1 在 CASIA-MFSD 数据集上基于帧的内部实验的准确率与半错误率的比较

Table 1 ACC and HTER of frame-based internal experiments on CASIA-MFSD dataset

(单位:%)

Method	ACC	HTER
Atoum's patch CNN	96.19	4.17
ShuffleNet	96.16	4.75
ShuffleNet+Hot blocks	96.86	4.44
ShuffleNet+Hot blocks+Saliency	**98.10**	**2.68**

表 2 在 Replay-Attack 数据集上基于帧的内部实验的准确率与半错误率的比较

Table 2 ACC and HTER of frame-based internal experiments on Replay-Attack dataset

(单位:%)

Method	ACC	HTER
Atoum's patch CNN	97.43	1.70
ShuffleNet	97.19	1.87
ShuffleNet+Hot blocks	99.05	1.66
ShuffleNet+Hot blocks+Saliency	**99.27**	**0.52**

4.3.2 显著像素方法有效性实验

为了进一步验证显著像素方法的有效性,我们进行了一组基于视频的 EER 和 HTER 的比较实验。

如表 3 所列,本文的显著预测方法相比 Atoum 等[18]的深度方法取得了更好的效果,证明了显著像素方法的有效性。同时,本文的融合方法也取得了更好的效果。

表 3 在 CASIA-MFSD 数据集上的显著像素方法有效性实验

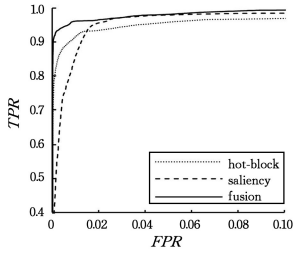
Table 3 Validity experiments of saliency pixel method on CASIA-MFSD dataset

(单位:%)

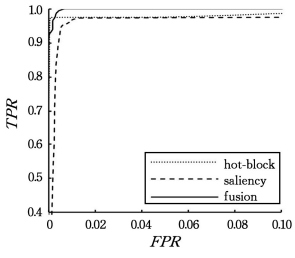
Method	EER	HTER
Atoum's depth[18]	2.85	2.52
Our Saliency	**1.90**	**1.51**
Atoum's fusion[18]	2.67	2.27
Our Fusion	**1.52**	**1.88**

4.3.3 融合分析

此外，本文还通过 ROC(Receiver Operating Characteristic)图，对热点块、显著像素以及融合方法进行了对比分析。由图 7 可知，融合方法在 CASIA-MFSD 和 Replay-Attack 数据集上的效果比热点块和显著像素方法都更好，融合方法的 AUC 面积大于热点块和显著像素方法，证明了融合方法的有效性。



(a)CASIA-MFSD



(b)Replay-Attack

图 7 不同数据集上的基于帧的 ROC 融合比较

Fig. 7 Frame-based ROC fusion comparison on different datasets

4.4 与其他方法的比较实验

本文在 CASIA-MFSD、Replay-Attack 以及 SiW 数据集上将所提方法与其他方法进行对比实验。

本文将所提方法与以下方法进行比较(按时间顺序排列)，并简要阐述以下方法在这 3 个数据集上的实验情况。

(1)2017 年的单帧方法。Atoum 等[18]提出使用随机块机制＋FCN 来进行深度预测的方法。该方法在 CASIA-MFSD 以及 Replay-Attack 两个数据集上分别进行了内部实验，未在 SiW 数据集上进行实验(SiW 数据集还未发布)。

(2)2018 年的多帧方法。Liu 等[23]在 LSTM-CNN 的基础上，利用远程光体积描记法以及深度信息来进行辅助监督。该方法在 CASIA-MFSD 以及 Replay-Attack 两个数据集上进行了相互的交叉实验，发布了 SiW 数据集并公布了相应的内部测试协议和实验结果。

(3)2018 年的单帧方法。Jourabloo 等[19]提出将攻击图分离出真人＋伪造噪声的方法。该方法在 CASIA-MFSD 和 Replay-Attack 数据集上进行了相互的交叉实验，未使用 SiW 数据集进行实验。

(4)2019 年的单帧方法。Liu 等[20]设计了一个用于检测未知欺骗攻击的深度树网络，并定义了一个新型人脸防伪指标 ZSFA，以检测模型对从未见过的攻击类型的应对能力。该方法在 CASIA-MFSD 以及 Replay-Attack 进行了 ZSFA 实验，未在 SiW 数据集上进行实验。

(5)2019 年的多帧方法。Yang 等[22]在 LSTM-CNN 的基础上加入了 RAM 以学习偏差，进而学习人脸的重要区域。该方法在 CASIA-MFSD 以及 Replay-Attack 两个数据集上进行了相互的交叉实验，并在 SiW 数据集上进行了内部实验。

(6)2019 年的多模态方法。Zhang 等[24]将多种模态的人脸图像(RGB 图像、NIR 图像、深度图像)作为输入，提出了 CASIA-SURF 多模态数据集和适用多模态的融合方法。该方法在 CASIA-MFSD 数据集上进行了以下比较：FAS-TD-SF[31]方法＋SiW 训练＋CASIA-MFSD 测试、FAS-TD-SF[31]方法＋CASIA-SURF 训练＋CASIA-MFSD 测试、其他 5 种方法＋Replay-Attack 训练＋CASIA-MFSD 测试，证明了使用 CASIA-SURF 数据集训练的模型在 CASIA-MFSD 测试集上表现最好，凸显了 CASIA-SURF 数据集的贡献。该方法还在 SiW 数据集上进行了以下比较：FAS-TD-SF[31]方法＋SiW 训练＋SiW 测试、FAS-TD-SF[31]方法＋CASIA-SURF 训练＋SiW 测试、Liu 等[23]发布的 SiW 数据集内部测试，结果再次证明使用 CASIA-SURF 训练的结果最好。因此，本文方法只与该文展示的 FAS-TD-SF[31]方法＋SiW 训练＋SiW 测试的结果进行比较。

鉴于以上情况，本文所提方法主要进行了以下几组比较实验。

(1) CASIA-MFSD，Replay-Attack 各自内部实验与 Atoum 等[18]的比较(同为单帧方法)。

(2)CASIA-MFSD，Replay-Attack 各自内部 ZSFA 实验与 Liu 等[20]的比较(同为单帧方法)。

(3)SiW 内部实验与 Liu 等[23]的、Yang 等[22]的、FAS-TD-SF[31]方法的比较(除了本文方法外，其余皆为多帧方法)

(4) CASIA-MFSD、Replay-Attack 的交叉实验与 Jourabloo 等[19]的比较(同为单帧方法)。因为 Liu 等[23]和 Yang 等[22]的方法为多帧方法，故不在该交叉实验中进行比较。

4.4.1 CASIA-MFSD，Replay-Attack 内部实验

实验评价指标为 EER 和 HTER。

由表 4 和表 5 可知，本文方法在 Replay-Attack 数据集上的 EER 和 HTER 和在 CASIA-MFSD 数据集上的 HTER 都优于其他方法。这说明本文的方法更有效。

表 4 CASIA-MFSD 数据集上不同方法的等错误率、半错误率比较

Table 4 EER and HTER comparison on CASIA-MFSD dataset of different methods

(单位：%)

Method	EER	HTER
Fine-tune VGG-Face[16] *	5.20	—
DPCNN[16] *	4.50	—
Yang et al.[14]	4.92	—
LSTM-CNN[21] *	5.17	5.93
Boulkenafet et al.[9]	6.20	—
Siddiqui et al.[32]	3.14	—
Boulkenafet et al.[7]	2.80	—
Haralick features[33]	—	1.10
Moire pattern[34]	—	0
Atoum et al.[18] *	2.67	2.27
Ours *	**1.52**	1.88

注：* 标记了基于 CNN 的方法

表 5 Replay-Attack 数据集上不同方法的等错误率、半错误率比较

Table 5 EER and HTER comparison on Replay-Attack dataset of different methods

（单位：%）

Method	EER	HTER
Fine-tune VGG-Face[16]*	8.40	4.30
DPCNN[16]*	2.90	6.10
Yang et al.[14]*	2.14	—
Boulkenafet et al.[9]	0.40	2.90
Boulkenafet et al.[7]	0.10	2.20
Moire pattern[34]	—	3.30
Atoum et al.[18]*	0.79	0.72
Ours*	**0**	**0.62**

注：* 标记了基于 CNN 的方法

4.4.2 CASIA-MFSD，Replay-Attack 内部 ZSFA（Zero-Shot of Face Anti-spoofing）实验

实验的评价指标为 AUC。ZSFA 用于检验模型对从未见过的攻击的应对能力，如 Video 列，表示训练过程使用其余两种攻击进行训练，测试时用 Video 攻击进行测试。其他情况以此类推。

由表 6 和表 7 可知，本文方法在任一值的比较上都超过了 2019 年的 Deep Tree[20] 方法，仅在 CASIA-MFSD 的 Video 结果上败给了其他方法。本文方法在这两个数据集上应对从未见过的攻击的效果是非常好的。

表 6 在 CASIS-MFSD 数据集上进行 ZSFA 测试的 AUC

Table 6 AUC of ZSFA testing on CASIA-MFSD dataset

（单位：%）

Methods	AUC		
	Video	Cut Photo	Warped Photo
$OC\text{-}SVM_{RBF}$+BSIF[35]	70.7	60.7	95.9
SVM_{RBF}+LBP[36]	91.5	91.7	84.5
NN+LBP[37]	**94.2**	88.4	79.9
Deep Tree[20]	90.0	97.3	97.5
Our hot-block method	90.53	98.94	97.57
Our saliency method	73.76	98.04	92.41
Our fusion	90.69	**98.96**	**97.91**

表 7 在 Replay-Attack 数据集上进行 ZSFA 测试的 AUC

Table 7 AUC of ZSFA testing on Replay-Attack dataset

（单位：%）

Method	AUC		
	Video	Digital Photo	Printed Photo
$OC\text{-}SVM_{RBF}$+BSIF[35]	84.3	88.1	73.7
SVM_{RBF}+LBP[36]	99.1	98.2	87.3
NN+LBP[37]	99.8	95.2	78.9
Deep Tree[20]	99.9	99.9	99.6
Our hot-block method	99.98	99.37	99.23
Our saliency method	99.72	98.71	96.48
Our fusion	**99.99**	**99.98**	**99.72**

4.4.3 SiW 内部实验

为证明我们的方法在新的数据集上同样适用，本文还在 2018 年发布的 SiW 数据集上进行了实验，并与目前最好的方法进行了比较。除了本文方法以外，其他方法都是多帧方法且都是最新的成果。

本文遵循 SiW 数据集的测试协议[23]，用训练集的前 60 帧训练＋测试集的所有帧进行测试。评价指标为：APCER（Attack Presentation Classification Error Rate）、BPCER（Bona Fide Presentation Classification Error Rate）以及 ACER（APCER 与 BPCER 的平均值）[38]。

由表 8 可知，本文方法在 ACER 指标上取得了最好的结果，作为单帧方法能在该指标上优于多帧方法，再一次证实了本文方法的有效性。

表 8 在 SiW 数据集上不同方法的 APCER，BPCER 和 ACER 比较

Table 8 APCER，BPCER and ACER comparison of different methods on SiW dataset

（单位：%）

Method	APCER	BPCER	ACER
Auxiliary[23]	3.58	3.58	3.58
FAS-TD-SF[31]	1.27	**0.83**	1.05
STANS[22]	—	—	1.00
Our hot-block method	2.207	1.935	2.071
Our saliency method	**0.209**	1.948	1.078
Our fusion	0.316	1.682	**0.999**

4.4.4 CASIA-MFSD，Replay-Attack 交叉实验

由表 9 可知，本文方法在这两个数据集上的交叉测试结果不是很好。由上文可知，两个数据集在热点块以及显著预测图上存在明显的差异，从而导致了两个数据集之间的交叉实验结果不理想。

表 9 不同方法就 HTER 在 CASIA-MFSD 与 Replay-Attack 数据集上交叉实验的比较

Table 9 Cross testing comparison on CASIA-MFSD dataset versus Replay-Attack dataset in terms of HTER

（单位：%）

Method	HTER			
	Train CASIA	Test Replay	Train Replay	Test CASIA
FaceDS[19]	**28.5**		**41.1**	
Our hot-block	33.4		47.7	
Our Saliency	50.5		67.1	
Our Fusion	45.9		65.2	

4.4.5 速度比较实验

另外，本文还进行了速度比较实验，与本地复现的 Atoum 等[18] 提出的随机块机制＋FCN 进行深度预测的方法进行比较，评价指标为 FPS（Frames per second）。测试环境为 2× Intel Xeon E5-2620v4，内存 256 GB，GPU GTX1 080 8GB。所有实验的 batch size 统一设置为 50 张，在 CASIA-MFSD 测试集的 66 450 张图片上进行，对每个方法进行 5 组测速实验，以均值±标准差表示。

由表 10 可知，本文的热点块方法的 FPS 比 Atoum 等[18] 的随机块方法的 FPS 高出 56.98%。由表 11 可知，本文的显著像素预测网络的预测（阶段 1）用时是 Atoum 等[18] 的 FCN 深度预测网络的 65.9%；本文使用的显著预测图分类网络是 5 层的卷积神经网络，而 Atoum 等[18] 使用的是 SVM，SVM 的分类时间是本文的 5 层卷积神经网络的 325 倍。SVM 分类耗时长的原因包括：1）Atoum 等[18] 选用的 RBF（径向基函数）核需要将输入数据映射到高维空间，这将付出较大的计算代价；2）SVM 无法使用 GPU 加速运算。综上所述，无论是与 Atoum 等[18] 的 patch 方法相比较，还是与他们的 depth 预测方法相比较，本文方法的 FPS 都显著超越了他们的。

表 10 与 random-patch 方法的 FPS 比较

Table 10 FPS comparison with random-patch method

Method	FPS/(f/s)
Atoum's Random-patch	1 351.026±41.706
Our Hot-block	**2 120.816±39.714**

表 11 与 depth 方法的 FPS 比较

Table 11 FPS comparison with depth method

Method	Stage1/s	Stage2/s	FPS/(f/s)
Atoum's Depth	853.815±7.644	6579.190±78.296	8.941±0.086
Our Saliency	**563.076±0.577**	**20.213±0.529**	**113.923±0.099**

注：Stage1 表示的是预测深度所花费的时间，Stage2 表示的是对预测图进行分类所花费的时间

结束语 本文提出基于注意力机制的热点块和显著像素的卷积神经网络方法。基于注意力的热点块卷积神经网络方法将注意力集中在 5 个重要的脸部区域小块上，以对 5 个热点块的判别来取代对整张人脸的判别；基于注意力的显著像素方法用于预测人脸的显著像素点，通过判别显著预测图是否具有人脸的深度信息来鉴别活体和攻击。本文方法将上述所得的两个结果进行融合，进一步提升了效果。为验证方法的有效性，本文分别在 CASIA-MFSD，Replay-Attack 以及 SiW 数据集上进行验证。实验结果表明，我们的方法具有良好的适应性，能够迫使网络集中注意力于更具鉴别性信息的部分，以规避网络学习其他无关信息的可能性，进而提高网络的正确率。在未来的工作中，我们可以把该方法应用到实际的工程应用中，也可以尝试改进该方法的模型，使用大数据集训练来提升其泛化能力，从而使其具有更好的鲁棒性。

参 考 文 献

[1] MÄÄTTÄ J，HADID A，PIETIKÄINEN M. Face spoofing detection from single images using micro-texture analysis[C]// 2011 International Joint Conference on Biometrics(IJCB). 2011：1-7.

[2] DE FREITAS PEREIRA T，ANJOS A，DE MARTINO J M，et al. LBP-TOP Based Countermeasure against Face Spoofing Attacks[C]// Computer Vision - ACCV 2012 Workshops. Berlin，Heidelberg：Springer，2013：121-132.

[3] DE FREITAS PEREIRA T，ANJOS A，DE MARTINO J M，et al. Can face anti-spoofing countermeasures work in a real world scenario? [C]// 2013 International Conference on Biometrics (ICB). 2013：1-8.

[4] KOMULAINEN J，HADID A，PIETIKÄINEN M. Context based face anti-spoofing[C]// 2013 IEEE Sixth International Conference on Biometrics：Theory，Applications and Systems (BTAS). 2013：1-8.

[5] YANG J W，LEI Z，LIAO S H，et al. Face liveness detection with component dependent descriptor[C]// 2013 International Conference on Biometrics(ICB). 2013：1-6.

[6] 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.

[7] BOULKENAFET Z，KOMULAINEN J，HADID A. Face Antispoofing Using Speeded-Up Robust Features and Fisher Vector Encoding[J]. IEEE Signal Processing Letters，2017，24(2)：141-145.

[8] 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.

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

[10] LI J W，WANG Y H，TAN T N，et al. Live face detection based on the analysis of Fourier spectra[C]// Biometric Technology for Human Identification. Proceedings of SPIE - The International Society for Optical Engineering，2004，5404：296-303.

[11] PAN G，SUN L，WU Z H，et al. Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera[C]// 2007 IEEE 11th International Conference on Computer Vision. 2007：1-8.

[12] SUN L，PAN G，WU Z H，et al. Blinking-Based Live Face Detection Using Conditional Random Fields[C]// Advances in Biometrics. Berlin，Heidelberg：Springer，2007：252-260.

[13] KOLLREIDER K，FRONTHALER H，FARAJ M I，et al. Real-Time Face Detection and Motion Analysis With Application in "Liveness" Assessment[J]. IEEE Transactions on Information Forensics and Security，2007，2(3)：548-558.

[14] YANG J，LEI Z，LI S Z. Learn Convolutional Neural Network for Face Anti-Spoofing[J/OL]. CoRR，2014，http：//arxiv. org/abs/1408.5601.

[15] PATEL K，HAN H，JAIN A K. Cross-Database Face Antispoofing with Robust Feature Representation[C]// Biometric Recognition. Cham：Springer International Publishing，2016：611-619.

[16] LI L，FENG X Y，BOULKENAFET Z，et al. An original face anti-spoofing approach using partial convolutional neural network[C]// 2016 Sixth International Conference on Image Processing Theory，Tools and Applications(IPTA). 2016：1-6.

[17] FENG L T，PO L M，LI Y M，et al. Integration of image quality and motion cues for face anti-spoofing：A neural network approach[J]. Journal of Visual Communication and Image Representation，2016，38：451-460.

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

[19] JOURABLOO A，LIU Y J，LIU X M. Face De-spoofing：Anti-spoofing via Noise Modeling[C]// Computer Vision - ECCV 2018. Cham：Springer International Publishing，2018：297-315.

[20] LIU Y J，STEHOUWER J，JOURABLOO A，et al. Deep Tree Learning for Zero-Shot Face Anti-Spoofing[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019：4675-4684.

[21] XU Z Q，LI S，DENG W D. Learning temporal features using

LSTM-CNN architecture for face anti-spoofing[C]// 2015 3rd IAPR Asian Conference on Pattern Recognition(ACPR). 2015: 141-145.

[22] YANG X, LUO W H, BAO L C, et al. Face Anti-Spoofing: Model Matters, so Does Data[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2019: 3502-3511.

[23] LIU Y J, JOURABLOO A, LIU X M. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 389-398.

[24] ZHANG S F, WANG X B, LIU A J, et al. A Dataset and Benchmark for Large-Scale Multi-Modal Face Anti-Spoofing[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2019: 919-928.

[25] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.

[26] LIU J J, HOU Q B, CHENG M M, et al. A Simple Pooling-Based Design for Real-Time Salient Object Detection[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2019: 3912-3921.

[27] ZHANG Z W, YAN J J, LIU S F, et al. A face anti-spoofing database with diverse attacks[C]// 2012 5th IAPR International Conference on Biometrics(ICB). 2012: 26-31.

[28] CHINGOVSKA I, ANJOS A, MARCEL S. On the effectiveness of local binary patterns in face anti-spoofing[C]// 2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group(BIOSIG). 2012: 1-7.

[29] FENG Y, WU F, SHAO X H, et al. Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network[C]// Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 557-574.

[30] WU X, HE R, SUN Z N, et al. A Light CNN for Deep Face Representation With Noisy Labels[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2884-2896.

[31] WANG Z Z, ZHAO C X, QIN Y X, et al. Exploiting temporal and depth information for multi-frame face anti-spoofing [J/OL]. CoRR, 2018, http://arxiv.org/abs/1811.05118.

[32] SIDDIQUI T A, BHARADWAJ S, DHAMECHA T I, et al. Face anti-spoofing with multifeature videolet aggregation[C]// 2016 23rd International Conference on Pattern Recognition (ICPR). 2016: 1035-1040.

[33] AGARWAL A, SINGH R, VATSA M. Face anti-spoofing using Haralick features[C]// 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems(BTAS). 2016: 1-6.

[34] PATEL K, HAN H, JAIN A K, et al. Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks [C]// 2015 International Conference on Biometrics(ICB). 2015: 98-105.

[35] ARASHLOO S R, KITTLER J, CHRISTMAS W. An Anomaly Detection Approach to Face Spoofing Detection: A New Formulation and Evaluation Protocol[J]. IEEE Access, 2017, 5: 13868-13882.

[36] BOULKENAFET Z, KOMULAINEN J, LI L, et al. OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations[C]// 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2017). 2017: 612-618.

[37] XIONG F, ABDALMAGEED W. Unknown Presentation Attack Detection with Face RGB Images[C]// 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems(BTAS). 2018: 1-9.

[38] ISO/IEC JTC 1/SC 37. Information technology-Biometric presentation attack detection: ISO/IEC 30107-1[S]. Online Browsing Platform, 2016.

WU Xiao-li, born in 1995, postgraduate. Her main research interests include face anti-spoofing and deep learning.

HU Wei, born in 1979, Ph.D, associate professor. His main research interests include face recognition, real-time global illumination rendering, image editing, image recognition and multiple-projector based tiled display.