

基于多域时序特征挖掘的伪造人脸检测方法

朱春陶^{1,2,3}, 尹承禧^{1,2,3}, 张博林^{1,2,3}, 殷琪林^{1,2,3,4}, 卢伟^{1,2,3}

(1. 中山大学计算机学院, 广东 广州 510006; 2. 广东省信息安全技术重点实验室(中山大学), 广东 广州 510006;
3. 机器智能与先进计算教育部重点实验室(中山大学), 广东 广州 510006; 4. 山东省计算机网络重点实验室, 山东 济南 250014)

摘要: 随着计算机技术在金融服务行业中的不断发展, 金融科技便利了人们的日常生活, 与此同时, 数字金融存在着危害性极大的安全问题。人脸生物信息作为人物身份信息的重要组成部分, 广泛应用于金融行业中的支付系统、账号注册等方面; 伪造人脸技术的出现不断冲击着数字金融安全体系, 给国家资产安全和社会稳定造成了一定的威胁。为了应对伪造人脸带来的安全问题, 提出了一种基于多域时序特征挖掘的伪造人脸检测方法。所提方法从视频在空域和频域中存在的时序特征出发, 基于人脸统计特征数据分布的一致性以及时间上动作趋势的一致性, 对篡改特征进行区分增强。在空域中, 所提方法使用改进的长短记忆网络(LSTM)来挖掘帧间的时序特征; 在频域中, 利用3D卷积层来挖掘不同频段频谱的时序信息, 并与主干网络提取到的篡改特征进行融合, 进而有效地区分伪造人脸和真实人脸。所提方法在主流数据集上表现优越, 证明了所提方法的有效性。

关键词: 人脸身份认证; Deepfake检测; 时序特征; 多域特征

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023044

Forgery face detection method based on multi-domain temporal features mining

ZHU Chuntao^{1,2,3}, YIN Chengxi^{1,2,3}, ZHANG Bolin^{1,2,3}, YIN Qilin^{1,2,3,4}, LU Wei^{1,2,3}

1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China
2. Guangdong Province Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou 510006, China
3. Ministry of Education Key Laboratory of Machine Intelligence and Advanced Computing, Sun Yat-sen University, Guangzhou 510006, China
4. Shandong Key Laboratory of Computer Networks, Shandong 250014, China

Abstract: Financial technology has greatly facilitated people's daily life with the continuous development of computer technology in the financial services industry. However, digital finance is accompanied by security problems that can be extremely harmful. Face biometrics, as an important part of identity information, is widely

收稿日期: 2023-01-09; 修回日期: 2023-04-23

通信作者: 卢伟, luwei3@mail.sysu.edu.cn

基金项目: 国家自然科学基金(U2001202, 62072480)

Foundation Item: The National Natural Science Foundation of China (U2001202, 62072480)

引用格式: 朱春陶, 尹承禧, 张博林, 等. 基于多域时序特征挖掘的伪造人脸检测方法[J]. 网络与信息安全学报, 2023, 9(3): 123-134.

Citation Format: ZHU C T, YIN C X, ZHANG B L, et al. Forgery face detection method based on multi-domain temporal features mining[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 123-134.

used in payment systems, account registration, and many other aspects of the financial industry. The emergence of face forgery technology constantly impacts the digital financial security system, posing a threat to national asset security and social stability. To address the security problems caused by fake faces, a forgery face detection method based on multi-domain temporal features mining was proposed. The tampering features were distinguished and enhanced based on the consistency of statistical feature data distribution and temporal action trend in the temporal features of videos existing in the spatial domain and frequency domain. Temporal information was mined in the spatial domain using an improved LSTM, while in the frequency domain, temporal information existing in different frequency bands of the spectrum was mined using 3D convolution layers. The information was then fused with the tampering features extracted from the backbone network, thus effectively distinguishing forged faces from real ones. The effectiveness of the proposed method was demonstrated on mainstream datasets.

Keywords: face authentication, Deepfake detection, temporal features, multi-domain features

0 引言

随着计算机技术和人工智能技术的不断延伸, 诸多行业的发展进入了一个新的阶段。计算机技术与金融服务行业的融合极大地推进了金融科技的发展。金融数字化、智能化等在便利了人们日常生活的同时, 存在着诸多安全隐患。金融科技自发展以来, 人脸生物信息作为个体的重要身份信息, 在移动支付系统、账号注册、网贷申请等方面经常作为重要凭证。近年来, 由于深度学习技术的不断完善优化, 人脸伪造的方式愈加多样, 人脸伪造技术生成的虚假人脸也愈加真实; 利用人脸伪造技术制作的假人脸或者假视频更是具有欺骗人脸验证系统的能力, 对数字金融体系中依托人脸识别的服务威胁极大。常见的人脸伪造方法有 Deepfakes^[1]、Face2Face^[2]、FaceSwap^[3]等, 在 3D 打印技术的加持下, 生成并制作高清的 3D 人脸面具变得容易, 并且已经有滥用人脸伪造技术进行违法犯罪的事件出现, 如利用伪造人脸突破人脸识别系统来窃取用户支付宝余额等。为了应对伪造人脸带来的安全问题, 研究伪造人脸的检测技术十分必要。

现有的主流人脸伪造方法主要依托于生成对抗网络 (GAN, generative adversarial network)^[4]、变分自编码器 (VAE, variational auto-encoder)^[5], 也被称为深度伪造技术 (Deepfake)。其中生成对抗网络在图像生成、人脸编辑、风格转换等任务中表现优越, 高质量伪造图像的出现使得“眼见不一定为实”。人脸伪造方法整体上可以分为编辑、合成、重现、替换, 其中编辑指的是对人脸

的某种属性进行变化, 如发色、肤色、配饰等; 合成是指人脸的整体生成, 如虚拟角色创建。人脸的编辑和合成在研究领域中十分热门, 但是, 重现和替换带来的问题更引人警惕^[6], 重现是指让目标对象重现源对象的一些行为, 如表情、嘴部动作等, 重现技术的不当使用可能会造成诽谤、错误信息传播或者欺诈、勒索等违法犯罪行为发生, 严重威胁社会安全和信任体系稳定。替换即人脸身份交换, 可以通俗地理解为两张人脸的交换。人脸替换的滥用同样会危害社会公共秩序, 如用来传播不当的政治观点等。由此可见, 深度人脸伪造技术不仅威胁着数字金融安全体系的稳定, 而且对社会公共安全以及个人隐私有着十分不利的影响。

与此同时, 伪造人脸的检测方法也在快速发展。现有的检测方法主要从空域视觉伪影、频域伪造痕迹以及时序不一致性 3 个方面出发。空域视觉伪影指的是伪造人脸在生成过程中存在的一些缺陷, 如眼睛颜色不一致^[7]、牙齿细节丢失、光照不一致引起的局部阴影异常^[8]、在融合边缘处留下的轮廓伪影^[9]等。频域伪造痕迹多是挖掘对抗生成网络在上采样操作时引入的篡改痕迹, 在频谱中会有较明显的差异^[10], 常用的频域变换操作为离散余弦变换 (DCT, discrete cosine transform)、离散傅里叶变换 (DFT, discrete fourier transform) 和离散小波变换。由于频域篡改痕迹相较于空域篡改痕迹更加稳定, 在进行低质量伪造人脸视频检测时, 加入频域特征是一种有效的方式^[11]。时序不一致性主要关注视频帧之间的时序特征, 通过捕获伪造人脸视频帧与真实人脸视

帧之间的时序特征差异,如眨眼频率^[12]、头部对齐与抖动、人脸关键点位置变化^[13]等,来进行伪造人脸的检测判断。

大部分伪造人脸检测方法是单帧的检测方法,通过从空域或者频域中提取篡改特征,进行最终的结果判断;虽然利用帧间时序特征的多帧检测方法已经存在,但这些方法多在卷积操作之后仅从空域中捕获时序特征,忽略了从频域中挖掘时序特征,同时,部分方法需要从长时段的连续视频帧中挖掘出有效信息,这在数据处理上具有一定的压力。基于此,本文从时序特征一致性的角度出发,提出了一种基于多域时序特征挖掘的伪造人脸检测方法。首先,在生成伪造人脸视频时,深度伪造方法会尽可能地拟合真实人脸的数据分布,但由于生成方法在诸多因素的限制下,对光照、噪声等一些固有特征的生成欠缺,伪造人脸的数据分布与真实人脸的数据分布存在一定差异,并且很难保持每帧中的人脸数据分布一致;其次,伪造人脸视频比真实人脸视频存在更多的抖动现象,其人脸动作存在僵硬、不顺滑的问题,即伪造人脸视频在时序上存在统计特征数据分布的不一致性和动作趋势上的不一致性。因此,本文提出了空域时序特征挖掘模块(STFM, spatial domain temporal feature mining module)和频域时序特征挖掘模块(FTFM, frequency domain temporal feature mining module)。STFM对长短期记忆(LSTM, long short-term memory)神经网络^[14]单元进行了改动,受ST-LSTM^[15]启发,本文利用不同的记忆单元有侧重地挖掘人脸纹理细节以及人脸主体数据的时序特征。这样的主要目的是突出视频帧之间微小细节上的变动以及对人脸整体数据分布一致性的捕获。之后将提取到的空域时序特征与主干网络中提取到的低层特征进行融合。频域中的时序信息主要指邻近视频帧频谱信息在数据分布上的一致性,其具有良好的稳定性,不易被压缩等操作破坏。因此,FTFM可以在捕获视频帧频谱数据分布一致性的同时,借由频域信息的稳定性,稳定模型的检测性能。本文所提方法在多个主流人脸伪造数据集上表现优秀,验证了本文所提方法的有效性。

本文所提方法有以下主要贡献。① 设计了一

种基于LSTM的空域时序特征挖掘模块,有侧重地对空域时序特征进行捕获。② 关注频域内视频帧频谱信息的时序性,设计了一种频域内的时序特征挖掘模块,增强了模型的稳定性。

③ 在多个主流人脸伪造数据集上表现优秀,具有有效性。

1 相关工作

1.1 人脸伪造方法

自从GAN^[4]和VAE^[5]被提出并不断完善,深度人脸伪造方法发展迅速,出现了很多即使是非专业人士也能操作的软件工具,并且创造出的虚假人脸具有高真实性。Face2Face^[2]是一种基于图信息的方法,用于目标视频的实时面部再现,通过跟踪源视频和目标视频中的人脸面部表情,将源视频中的人脸进行扭曲变形,进而拟合目标人脸序列的表情动作,合成最终的目标视频流。FaceSwap^[3]同样是一种基于图信息的方法,与Face2Face不同的是,FaceSwap是将人脸区域部分从源视频转移到目标视频,在操作时,首先会通过检测人脸关键点来提取人脸面部区域,然后利用输入图像来最小化映射形状和局部关键点之间的差异,最后对人脸部分和其余的背景部分进行融合,调整颜色、色差等。NeuralTextures^[16]方法在对图像进行重建时,所获取的图像3D信息并不完美,因此,在后续的渲染操作中,利用像素特征补全由于3D信息欠缺所造成的伪造痕迹,从而生成逼真的高质量图像。

近年来,深度学习的发展使得视频合成在更加逼真的同时变得更加容易。Deepfakes^[1]利用编码器学习提取人脸区域特征,并利用解码器再生新的人脸。Liu等^[17]提出了一种基于Coupled GAN的无监督的图像到图像转换框架,这是Deepfakes算法的基础。Karras等^[18]提出了ProGAN用于生成高分辨率图像。StyleGAN^[19]和AdaIN^[20]的提出使得生成器可以生成具有指定特征的人脸图像(如发色、肤色、饰品等)。Li等^[21]提出了两阶段的换脸框架:第一阶段生成高真实性的换脸图像;第二阶段则在第一阶段生成的换脸图像的基础上增加面部遮挡。随着深度学习技术的发展和不断完善,基于GAN结构的人脸伪造方法相继出现,

并且伪造人脸逐渐肉眼难以分辨。除去在照片 AI 修复以及部分娱乐性质场景下的应用, 伪造人脸技术同样给部分不法分子提供了作案工具, 为了维护社会稳定和公共安全秩序, 对伪造人脸检测技术的研究是亟须且必要的。

1.2 伪造人脸检测方法

在人脸伪造技术迅速发展的同时, 伪造人脸检测方法迅速涌现, 这些方法大多数从空域发现提取篡改痕迹特征, 对输入样本进行真假判别。在检测技术发展之初, 检测虚假数字媒体的方法多依赖于人工设计的特征^[22-24]来发现伪造区域或者放大真实图像和虚假图像之间细微的不一致性。在检测伪造人脸时, 对应的人工设计的脸部特征包括眨眼^[12]、不一致的头部动作^[25]、脸部表情变化^[26]等, 但人工设计特征存在耗时的缺点, 也很难应对新的未知伪造方法。近年来, 由于人工智能和深度学习技术的快速发展, 利用深度神经网络在空域中提取图像中层和高层信息来检测伪造人脸的方法表现突出, Xception^[27]网络在构造时引入了深度可分离卷积。MesoNet^[28]则利用两个浅层网络提取图像的介观特征来进行深度伪造检测。Nguyen 等^[29]提出了一种利用胶囊网络来检测人脸篡改的方法。Li 等^[30]通过关注大多数人脸篡改方法中存在的混合操作, 提出了一种新的图像表征 Face X-ray, 大大提升了检测方法的泛化性能。但是, 在空域中挖掘篡改痕迹特征的检测方法很容易受到数据集质量的影响, 其在一些高压缩、低质量的视频中检测性能会下降。

考虑到频域信息在低质量的伪造视频检测中有较好的稳定性, 近些年出现了一批利用频域篡改痕迹的伪造人脸检测方法。Durall 等^[31]提出在利用 DFT 将待检测样本转到频域之后, 平均获得的各频带振幅可以挖掘人脸篡改操作中的异常信息。F³Net^[32]利用不同但互补的频域成分来挖掘频域中的篡改痕迹特征, 进而检测伪造人脸。Liu 等^[10]在关注高层语义特征的同时, 注意到了上采样操作在相位谱中留下的篡改痕迹, 提出的 SPSSL 在跨库检测中取得了优越的性能。频域内篡改特征的利用在增强检测方法的准确性和泛化性等方面发挥了很大的作用。与此同时, 基于时序一致

性的检测方法相继出现, 与之前单帧的方法相比, 为了捕获挖掘时序特征, 该类方法通常基于视频片段进行伪造人脸视频检测。Güera 等^[33]设计了一种从时序上检测伪造视频的方法, 利用卷积神经网络提取帧级特征, 并利用提取到的特征来训练循环神经网络。Hu 等^[34]提出了一种两分支 (two-stream) 卷积网络, 一分支用来提取单帧特征, 另一分支用来提取视频段时序特征, 最后综合两分支的得分来进行视频真假判断。Lu 等^[35]则利用长程注意力机制从时空域中捕获伪造人脸视频存在的空间缺陷和时间缺陷, 进而检测视频真伪情况。考虑到捕获时序特征需要较多的连续视频帧以及频域内的时序特征经常被忽略, 本文提出了一种基于多域时序特征挖掘的伪造人脸检测方法。

2 基于多域时序特征的检测方法

本文关注的时序特征是指人脸图像在数据分布、运动趋势上的一致性。现有的人脸伪造方法在进行人脸生成时, 涉及几何变形、光照模拟等操作, 来保证所生成人脸在整体色调、动作拟合等方面的协调性, 使得伪造人脸看起来更加真实。生成器在人脸伪造时会趋于拟合真实的数据分布, 这个过程中会不可避免地引入一些真实人脸图像中不存在的成分, 如不合常理的阴影、黑点以及非自然噪声等; 另外, 伪造人脸存在细节丢失, 在牙齿、人脸纹理、瞳孔等区域表现模糊, 与真实人脸图像在单帧和帧间内表现的数据分布不一致。在时序上, 未经伪造的真实人脸视频在眨眼、口型、表情等变化上连贯而平滑, 伪造的人脸则存在一些不一致的抖动。基于此, 对时序特征的挖掘捕获可以有效区分真实人脸和伪造人脸。

本文提出的伪造人脸检测方法主要包括空域时序特征挖掘模块、频域时序特征挖掘模块以及主干网络部分。由于是对时序特征进行捕获, 视频帧以成组的方式输入网络, 在空域中利用改进的 LSTM^[14]进行时序特征挖掘和捕获; 同时将视频帧从空域转换到频域, 利用 3D 卷积块挖掘频谱中隐藏的数据分布信息, 最后与主干网络提取到的篡改特征进行融合增强, 从而对伪造视频进行分类。基于多域时序特征挖掘的伪造人脸检测框架如图 1 所示。

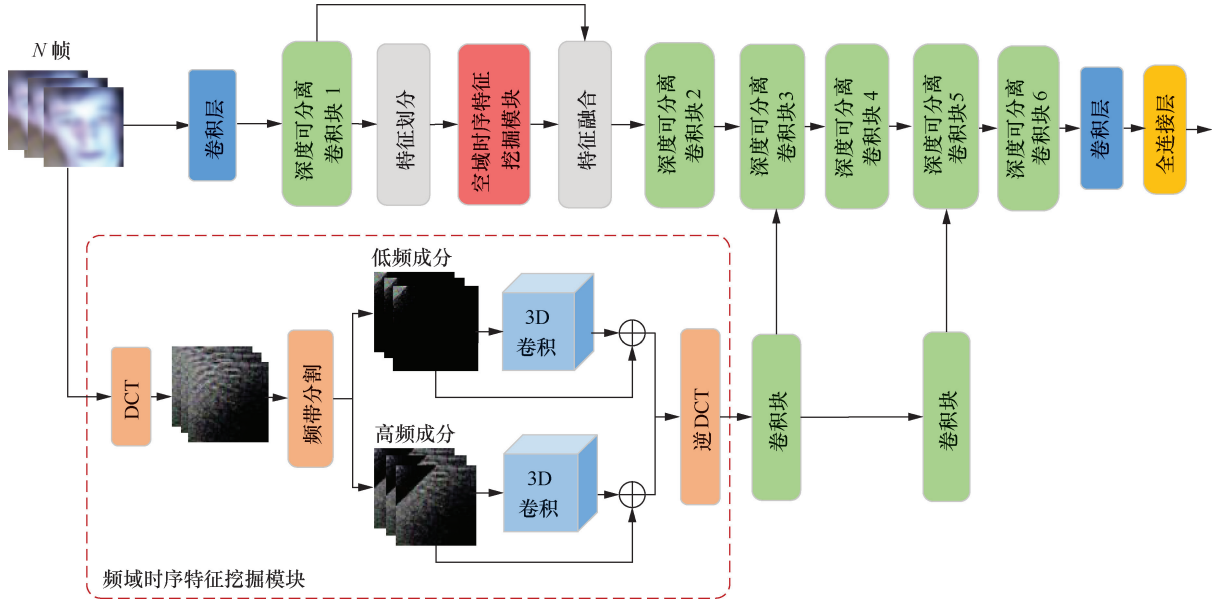


图1 基于多域时序特征挖掘的伪造人脸检测框架

Figure 1 Framework of forgery face detection method based on multi-domain temporal features mining

2.1 空域时序特征挖掘模块

空域时序特征的提取主要针对 RGB 人脸视频帧进行操作，利用卷积 LSTM^[36]实现。LSTM 常被应用于具有明显时序关系约束的领域，如自然语言处理、机器翻译等；在图像处理任务中，传统的 LSTM 并不适用，卷积神经网络更具优势。将卷积操作与 LSTM 结合起来的卷积 LSTM 为图像处理带来了一种新的思路。受 ST-LSTM^[15]的启发，本文设计了在空域中挖掘时序特征的卷积 LSTM 组成单元，空域时序特征挖掘模块的基本组成单元如图 2 所示。

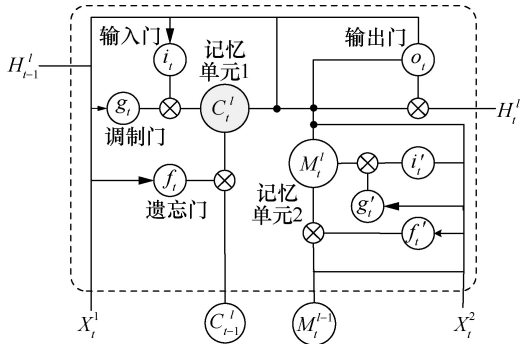


图2 空域时序特征挖掘模块的基本组成单元

Figure 2 The spatial domain temporal feature mining module basic cell

本文所提方法在将视频帧特征输入空域时序特征挖掘模块之前，对视频帧进行了特征划分操作，将视频帧特征划分为可表征纹理等成分的细

节特征以及表征人脸整体情况的主体特征。视频帧特征划分流程如图 3 所示。本文将视频帧在经过卷积操作之后获取的低层特征表示为 F_t ，即当前视频帧组内的第 i 帧特征，这里提供一种进行特征划分的方法，利用 DCT 变换后频谱能量分布的规整性，将高频部分对应的纹理细节特征和低频部分对应的视频帧主体内容特征进行分离，然后把得到的 X_t^1 、 X_t^2 输入空域时序特征挖掘模块。

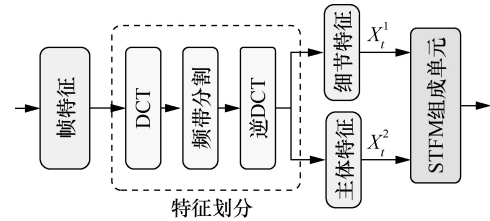


图3 视频帧特征划分流程

Figure 3 Video frames feature division Process

特征分离的操作可以表示为

$$X_t^j = D^{-1}\{D(F_t) \odot [f_b^j + \delta(f_v^j)]\}, j=1, \dots, N \quad (1)$$

其中， D 表示离散余弦变换， D^{-1} 表示离散余弦逆变换， $\odot$ 表示点乘操作， f_b^j 和 f_v^j 的设置引用 F³Net^[32]中的 FAD 头，表示基滤波核和可学习滤波核，可学习滤波核提供一定的冗余性。 δ 表示激活函数， $\delta(x) = (1 - \exp(x)) / (1 + \exp(x))$ ， $N=2$ ， X_t^1 和 X_t^2 分别表示细节特征和主体特征。

特征划分的目的在于增强模块对数据分布一致性和运动趋势一致性的约束,使得空域时序特征挖掘模块在时序捕获时更有目的性,在一定程度上避免由于卷积操作堆叠,模块过于关注视频帧的语义特征而忽略篡改痕迹特征的情况。

视频帧特征处理完之后,将得到的 X_t^1 和 X_t^2 输入空域时序特征挖掘模块,该模块主要由图 2 所示的基本组成单元构造。与传统的卷积 LSTM 基本单元相比,本文在原有卷积操作的基础上,引入了 1×1 卷积层来进行跨通道交互,以关注更多特征信息。之后,受 ST-LSTM 中对记忆单元的利用方式的启发,本文所提方法利用 C_t^l 记忆单元处理视频帧细节特征,挖掘人脸细节在时序上的一致性,利用 M_t^l 记忆单元处理视频帧的主体特征,挖掘人脸特征数据分布上的一致性,并在输出隐状态时进行整合,得到最终的时序特征部分。 C_t^l 记忆单元和 M_t^l 记忆单元内的特征获取和处理可表示为

$$\begin{aligned} g_t &= \tanh(W_{xg} * X_t^1 + W_{hg} * H_{t-1}^l) \\ i_t &= \sigma(W_{xi} * X_t^1 + W_{hi} * H_{t-1}^l) \\ f_t &= \sigma(W_{xf} * X_t^1 + W_{hf} * H_{t-1}^l) \\ C_t^l &= f_t \odot C_{t-1}^l + i_t \odot g_t \\ g_t' &= \tanh(W_{xg}' * X_t^2 + W_{mg}' * M_{t-1}^{l-1}) \\ i_t' &= \sigma(W_{xi}' * X_t^2 + W_{mi}' * M_{t-1}^{l-1}) \\ f_t' &= \sigma(W_{xf}' * X_t^2 + W_{mf}' * M_{t-1}^{l-1}) \\ M_t^l &= f_t' \odot M_{t-1}^l + i_t' \odot g_t' \end{aligned} \quad (2)$$

其中, $*$ 表示卷积操作,每层卷积后新增了 1×1 卷积层, $\odot$ 表示点乘操作, X_t^1 和 X_t^2 表示当前时间步输入的纹理特征和主体特征, H_{t-1}^l 表示上一时间步得到的隐状态特征, C_{t-1}^l 和 M_{t-1}^l 表示上一时间步得到的第 l 层的记忆单元特征, $\tanh$ 和 σ 表示激活函数操作, W_* 和 W_*' (如 W_{xi} 、 W_{xi}' 等) 表示对应的卷积核参数, g_t 和 g_t' 表示调制门特征, i_t 和 i_t' 表示输入门特征, f_t 和 f_t' 表示遗忘门特征。

同时,考虑到一组视频帧内的信息冗余性,为了让模块更多地关注到每一帧的异常部分,在记忆单元内的特征进行整合之前,加入了残差操作。该操作可表示为

$$\begin{aligned} R_t^l &= C_t^l - X_t^1 \\ D_t^l &= M_t^l - X_t^2 \\ o_t &= \sigma(W_{ho} * H_{t-1}^l + W_{co} * R_t^l + W_{mo} * D_t^l) \\ H_t^l &= o_t \odot \tanh(W_{hl} * [C_t^l, M_t^l]) \end{aligned} \quad (3)$$

其中, o_t 表示输出门特征, R_t^l 和 D_t^l 表示对应记忆单元内获取的残差特征, H_t^l 表示当前时间步获取的第 l 层的隐状态特征, W_* (如 W_{ho} 等) 表示对应的卷积核参数, W_{hl} 表示 1×1 卷积操作。

考虑到卷积块堆叠的结构虽然有利于提取更抽象的人脸语义信息,但同时会忽略消减一部分篡改痕迹特征,这对伪造人脸细微篡改痕迹的挖掘是不利的。因此,本文所提方法在使用空域时序特征挖掘模块时,采用浅层的结构进行空域时序特征的捕获。针对特征融合操作,本文主要利用一对并行的空间注意力机制^[37]和通道注意力机制^[38],与主干网络中提取到的特征进行交互融合。

2.2 频域时序特征挖掘模块

频域时序特征挖掘模块的设计首先考虑到以往基于时序特征挖掘的检测方法多是从 RGB 域入手,挖掘伪造人脸面部特征在时序上的不一致性,忽略了对频域信息的利用;其次是频域信息在低质量视频的检测任务中表现更加稳定,更有利于稳定模型的检测性能。频域时序特征挖掘模块所关注的特征为视频帧频谱数据的统计特征,即邻近视频帧的频谱在数据分布上保持一致性。

频域内的时序特征挖掘模块首先利用 DCT 将视频帧从空域转到频域,得到对应视频帧的频谱;考虑到中频频带所包含的篡改痕迹特征不明显,所提模块主要关注频谱的低频频带和高频频带部分,利用与空域时序特征挖掘模块中类似的频带分离操作,即

$$r_i = D(R_i) \odot [f_b^i + \delta(f_v^i)], i = 1, \dots, N \quad (4)$$

其中, R_i 表示当前输入视频帧内的第 i 帧视频帧, D 表示离散余弦变换, $\odot$ 表示点乘操作, f_b^i 以及 f_v^i 表示基滤波核和可学习滤波核, δ 表示激活函数, $\delta(x) = (1 - \exp(x)) / (1 + \exp(x))$, $N = 2$ 。在频带分离完成之后,挖掘对应频带中的时序信息,本文所提方法使用 3D 卷积块对频带信息进行处理,常用的 2D 卷积操作是指卷积核在输入

图像所在的二维空间中进行滑窗运算,3D 卷积操作则是卷积核在三维空间中进行滑窗运算,不仅可以捕获输入频带信息空间上的特征,而且可以捕获其时序上的特点。在频带信息处理完毕之后,将获得的特征与原本的频带信息融合,进行特征增强。

考虑到随着卷积层数的加深,神经网络模型不可避免地存在信息丢失现象,因此,在将频域中提取到的时序特征与主干网络中提取到的篡改痕迹特征进行融合时,选择在主干网络的中层、高层卷积操作部分进行,在稳定网络检测性能的同时,补充主干网络在卷积操作中的信息丢失,进一步提升模型的检测性能。之后,将多域时序特征与主干网络提取到的篡改痕迹特征进行卷积和全连接操作,对输入网络的视频帧组进行真伪判别。

整个网络的损失优化设置以交叉熵损失 l_{CE} 为主,考虑到在空域时序特征挖掘模块中对 LSTM 基本单元的改进,本文希望两个不同的记忆单元所提取到的信息能尽可能不同,即有侧重地学习不同的时序特征,如动作细节方面的时序特征、人脸数据分布上的时序特征,从而尽可能地捕获到视频帧之间的制约关系,来帮助检测工作的进行。因此,本文所提方法引入 ST-LSTM 中所提到的两个记忆单元之间的损失函数 l_{CS} 。使用两个记忆单元内存储的特征在每个时间步上余弦相似度的绝对值作为损失函数的具体数值。因此,网络最终的损失函数设置为

$$L = l_{CE} + \alpha l_{CS} \quad (5)$$

其中, α 为调节损失权重的超参数。

3 实验结果

本节的主要内容是实验结果的对比展示,首先对使用到的数据集进行介绍,之后将本文提出的方法与一些主流的检测方法进行单库和跨库的实验对比,最后进行有关本文所提方法的消融实验,并对实验结果进行分析。

3.1 数据集及实验设置

为了验证所提方法的有效性,本文选取在 3 个主流的伪造人脸数据集上进行实验,选择的数据集为 FaceForensics++^[39]、Celeb-DF^[40] 以及

DFDC^[41]。

FaceForensics++ 包含 Deepfakes^[1]、Face2 Face^[2]、FaceSwap^[3]、NeuralTextures^[16] 这 4 种人脸伪造篡改方法,该数据集对 1 000 个真实视频进行篡改伪造,每种篡改方法制造的视频具有 3 种不同的质量等级,即无压缩视频 C0 (RAW)、低压缩视频 C23 (HQ) 和高压缩视频 C40 (LQ)。随着压缩等级的提高,伪造视频检测的难度相应增大。由于近些年出现的伪造人脸检测方法在 RAW 或者 HQ 视频上的检测结果表现优秀,达到了 98% 以上的准确率,实验主要在检测难度更大的 LQ 视频数据上进行,更考验模型的检测性能。Celeb-DF 数据集是一个大型的通用深度伪造检测数据集,经常用来进行跨库实验测试,检测模型的泛化能力。Celeb-DF 与 FaceForensics++ 相比有更好的视觉效果,对边缘伪影、颜色转换的处理更加精细。DFDC 数据集是目前可获得的最大的人脸交换数据集。

本文所提方法的主干网络基于深度可分离卷积进行搭建,主干网络中每个深度可分离卷积块中有 2 层或 3 层深度可分离卷积层。优化器选择 Adam,初始学习率设置为 $lr = 0.0001$,并随着训练轮次的增加递减,训练时每个轮次的批大小设置为 16,损失 L 中 α 设置为 0.2,在进行网络的训练和测试时,传入网络的人脸图像大小设置为 256×256 ,传入网络的视频帧组,每组中的帧数设置为 6,每两帧之间的间隔为 2。本次实验所采用的评价指标有 ACC (accuracy) 和 AUC (area under curve)。其中,ACC 指标在大多数伪造检测实验中都会使用到,AUC 则可以更好地表征模型的泛化能力。在进行帧级 ACC 和 AUC 计算时,样本总数为视频帧样本数。在进行视频级 ACC 和 AUC 的指标计算时,样本总数为视频数。对于单帧级检测方案,对样本视频内所有视频帧的检测结果进行统计投票,以票数多的检测结果作为样本视频真假的判断结果,进而得出视频级的 ACC 和 AUC 指标。

3.2 对比实验

本小节将本文方法与一些表现突出的伪造人脸检测方法进行结果对比。其中 Xception^[27] 在伪造人脸检测任务中表现突出,经常作为检测方法

的基本对比方法; F³Net^[32]则主要在频域中挖掘伪造篡改特征; RFM^[42]和 Multi-attention^[43]的突出特点在于对注意力机制的利用。以上几种方法主要是单帧的伪造检测方法, 本文方法同样与多帧检测方法(RNN^[33]和 Two-stream^[34])进行了对比, 其中 RNN 主要采用循环神经网络的思想, Two-stream 则进行了双分支方案设计, 一分支用来进行帧级的伪造检测, 另一分支则是对分段的视频流进行时空上的信息提取检测, 在高压压缩率视频检测中性能优越。

对比实验首先在 FaceForensics++ 中的 4 种不同的人脸伪造数据集上进行, 由于高压压缩的视频在检测任务中难度更大, 更能体现模型的优越性, 本文选择在 LQ 视频质量下进行实验。在结果对比时, 分别展示了视频级和帧级的实验结果, 其中, 视频级检测结果如表 1 所示。从表 1 中可以看出, 本文方法在视频级检测中, ACC 和 AUC 指标与对比方法相比结果为最优或次优, 检测结果处于次优时, 也与最优的检测结果很相近, 说明本文提出的方法在视频级的伪造人脸检测任务

中具有很大优势, 所提的多域时序特征挖掘方法能够有效捕获伪造视频帧间的不一致性特征并辅助检测任务的进行。对于帧级检测结果, 如表 2 所示, 所提方法在帧级检测上同样占有优势, 特别是针对 Face2Face 和 FaceSwap 伪造方法, 但对于 Deepfakes 和 NeuralTextures 伪造方法则略有逊色, 分析原因, 本文方法在一定程度上对单帧人脸空域细节的捕获不够充分, 更侧重于捕获视频帧间的时序特征, 而多数单帧检测方法对人脸视频帧中存在的空域的伪造痕迹具备很强的提取能力。从整体实验结果可以看出, 不论是帧级检测还是视频级检测, 本文方法可以有效捕获到视频帧间的时序信息, 进而弥补单帧检测方法对时序特征的忽略; 能够通过挖掘视频帧之间数据分布的一致性以及人脸表情动作的连续性, 来暴露伪造人脸视频由于引入噪声数据以及不协调的动作抖动遗留下的篡改痕迹, 进而更有效地对伪造人脸视频进行判别检测。

本文同样在 Celeb-DF 和 DFDC 数据集上进行了实验。检测结果如表 3 和表 4 所示, 针对视

表 1 在 FaceForensics++ 数据集上, 不同检测方法的视频级检测结果
Table 1 The video-level performance of different detection methods on FaceForensics++

检测方法	Deepfakes		Face2Face		FaceSwap		NeuralTextures	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
Xception	0.967 9	0.976 4	0.917 9	0.931 7	0.960 7	0.970 5	0.817 9	0.823 6
Multi-attention	0.978 6	<u>0.982 0</u>	0.946 4	0.955 4	<u>0.967 9</u>	<u>0.976 4</u>	0.817 9	0.838 9
F ³ Net	0.971 4	0.978 5	0.946 4	<u>0.956 0</u>	0.971 4	0.974 8	0.835 7	0.846 4
Two-stream	0.946 4	0.980 0	0.864 8	0.940 0	0.852 7	0.940 0	0.800 5	0.900 0
RNN	0.817 0	0.860 0	0.708 5	0.770 0	0.668 3	0.760 0	0.691 0	0.780 0
本文方法	<u>0.975 0</u>	0.984 3	<u>0.942 9</u>	0.966 8	0.971 4	0.980 4	<u>0.817 9</u>	<u>0.871 9</u>

注: 加粗指标为最优实验结果, 下划线指标为次优实验结果, 余同。

表 2 在 FaceForensics++ 数据集上, 不同检测方法的帧级检测结果
Table 2 The frame-level performance of different detection methods on FaceForensics++

检测方法	Deepfakes		Face2Face		FaceSwap		NeuralTextures	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
Xception	0.945 8	0.986 8	0.893 5	0.960 4	0.924 2	0.978 0	0.743 5	0.812 4
Multi-attention	<u>0.962 3</u>	<u>0.992 2</u>	0.897 6	0.959 6	0.942 5	0.984 5	<u>0.771 4</u>	<u>0.845 7</u>
RFM	0.952 4	0.992 1	0.894 0	0.956 1	0.928 9	0.980 0	0.748 3	0.822 9
F ³ Net	0.963 3	0.992 9	<u>0.899 3</u>	0.963 0	0.945 6	<u>0.980 7</u>	0.774 8	0.853 0
本文方法	0.954 1	0.989 1	0.905 6	<u>0.961 5</u>	<u>0.943 2</u>	0.978 6	0.769 3	0.839 8

觉效果更好的 Celeb-DF 数据集, 本文方法依旧表现优异, 取得了较好的检测结果。

表3 不同检测方法在 Celeb-DF 数据集上的检测结果
Table 3 The results of different detection methods on Celeb-DF dataset

检测方法	ACC			
	帧级		视频级	
Xception	0.978 0	0.998 1	0.992 3	0.990 1
Multi-attention	<u>0.986 3</u>	0.995 2	0.990 3	0.988 7
RFM	0.980 9	<u>0.997 9</u>	—	—
F ³ Net	0.987 6	0.993 9	0.994 2	<u>0.991 6</u>
本文方法	0.984 5	0.997 4	<u>0.992 3</u>	0.992 7

表4 不同检测方法在 DFDC 数据集上的检测结果
Table 4 The results of different detection methods on DFDC dataset

检测方法	ACC			
	帧级		视频级	
Xception	0.889 1	0.960 1	0.898 4	0.897 8
Multi-attention	<u>0.910 2</u>	<u>0.975 3</u>	<u>0.921 9</u>	0.939 6
RFM	0.900 1	0.967 3	—	—
F ³ Net	0.924 5	0.977 8	0.937 5	0.924 5
本文方法	0.895 0	0.958 1	0.890 6	<u>0.928 9</u>

从不同检测方法在 DFDC 数据集上的检测结果 (见表 4) 可以看出, 本文方法的表现一般, 这是由于 DFDC 数据集中人脸所占的比例较小, 且使用的伪造技术多样, 视觉效果更好, 因此检测时难度有所提升, 时序信息在捕获时难度增大, 加上本文方法在对视频帧进行分组时, 每组中的视频帧数基数较小, 对长时间时序信息的捕获不充分, 因此检测结果偏低, 但本文方法的性能依旧超过了 Xception 检测方法, 表明了本文所提方法的有效性。

接着, 测试模型的泛化性能, 本文在 FaceForensics++ 数据集上对模型进行训练, 然后在 Deepfakes 和 Celeb-DF 数据集上对模型进行测试, 以 AUC 指标来对比模型的泛化性能。首先以 HQ 的 FaceForensics++ 数据集为训练集, 进行了跨库实验, 检测结果如表 5 所示。从结果可以看出, 本文方法在跨库性能上表现突出, 泛化性能处于前列, 表明本文提出的方法具有良好的泛化性能, 具备检测未知人脸伪造方法的能力。为

了测试本文方法在低质量视频检测环境中是否依旧具有稳定的泛化性能、具备一定的抗环境干扰能力, 同样在 LQ 视频环境下设计了跨库实验。检测结果如表 6 所示, 在低质量视频下的跨库实验中, 本文方法依然可以取得 0.728 7 的 AUC 指标性能, 对比方法的性能则随着视频质量的降低受到了较大影响, 表明了本文方法具备一定的稳定性。

表5 不同检测方法的跨库实验检测结果 (HQ)
Table 5 The cross-dataset results of different detection methods on HQ dataset

检测方法 (HQ)	AUC	
	FaceForensics++	Celeb-DF
Xception	0.985 1	0.673 1
Multi-attention	0.994 3	0.647 0
RFM	0.988 2	0.697 3
F ³ Net	0.995 3	0.635 2
本文方法	0.992 6	0.767 0

表6 不同检测方法的跨库实验检测结果 (LQ)
Table 6 The cross-dataset results of different detection methods on LQ dataset

检测方法 (LQ)	AUC	
	FaceForensics++	Celeb-DF
Xception	0.904 2	0.617 1
Multi-attention	0.924 6	0.595 0
RFM	0.910 0	0.635 9
F ³ Net	0.933 5	0.620 4
本文方法	0.927 1	0.728 7

3.3 消融实验

本节在 FaceForensics++ (LQ) 数据集上对所设计的模块, 进行系统的消融实验。首先, 对于输入网络中每组视频帧内的帧数进行消融实验, 观察每组样本内的视频帧数对模型整体性能的影响, 实验结果如表 7 所示, 通过观察, 根据检测结果, 当输入视频帧数为 6 时, 所达到的检测性能较优; 当帧数继续增加时, 检测效果没有明显提升。对结果进行分析, 从 LSTM 本身的结构特点出发, 在较长的时序信息捕获过程中, 对于距当前时间步较远的信息, 会有一定的信息丢失, 因此当增加了帧数后, 对模型本身的效果提升作用有限; 也可能是因为帧数增加之后, 信息冗余

增大,模型更容易关注到人脸本身的语义信息、偶然出现的噪声点等,这对人脸真伪的检测是不利的,进而检测结果没有明显提升。

表 7 针对输入帧数,所提方法在 FaceForensics++ 数据集上的消融实验结果

Table 7 For the frame number of input, the ablation results of the proposed method on FaceForensics++

帧数	ACC			
	Deepfakes	Face2Face	FaceSwap	NeuralTextures
4	0.948 6	0.901 6	0.948 6	0.760 4
6	0.954 1	0.905 6	0.943 2	0.769 3
8	0.952 6	0.903 2	0.942 0	0.765 6

接下来,对所提出的空域时序特征挖掘模块、频域时序特征挖掘模块进行消融实验,实验对照组设置为主干网络、主干网络和空域时序特征挖掘(主干+空域)、主干网络和频域时序特征挖掘(主干+频域)、主干网络和多域时序特征挖掘(主干+多域)。该组对照实验在输入视频帧数为 6 的情况下进行。实验结果如表 8 所示,在空域时序特征挖掘模块和频域时序特征挖掘模块单独使用时,与主干网络相比,检测性能都有一定提升,这表明对空域、频域中的时序信息进行关注是有效的,同时,对空域中的帧间时序信息进行挖掘带来的性能提升更突出,优于单独使用频域内的频谱数据分布特征,但同时将空域和频域内提取到的帧间时序信息与主干网络提取到的特征进行

融合时,可以使模型的性能进一步提升,这说明对频域信息的利用可以补充部分主干网络由于卷积堆叠操作造成的信息丢失,使得模型的检测准确率更高。针对频域信息使得模型性能增幅较小的问题,可能与对频域内信息挖掘不充分有关,这将是未来模型改进的一大方向。两组消融实验从不同的角度展示了所设计方法的有效性。

本文对所提空域时序特征挖掘模块基本单元的有效性进行了消融实验,在主干网络的基础上,与直接使用 ST-LSTM 基本单元进行对比实验。实验结果如表 9 所示,对结果进行分析,本文所提出的空域时序特征挖掘模块基本组成单元更有利于挖掘视频帧间的时序特征,在伪造人脸视频检测任务中表现突出,从而验证了本文所提空域时序特征挖掘模块基本组成单元设计的有效性和优越性。

4 结束语

本文从视频帧间存在的时序特征出发,提出了一种基于多域时序特征挖掘的伪造人脸检测方法,利用改进的循环神经网络基本单元以及 3D 卷积操作,分别在空域和频域内进行时序特征提取操作,并将提取到的特征与主干网络中发现的伪造痕迹特征相融合,以更有效地检测伪造人脸。在进行时序特征提取时,有针对性地对空域特征和频域特征进行了划分,从而使网络模型更有侧

表 8 针对各个模块,所提方法在 FaceForensics++数据集上的消融实验结果

Table 8 For each proposed module, the ablation results of the proposed method on FaceForensics++

对照组	Deepfakes		Face2Face		FaceSwap		NeuralTextures	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
主干	0.943 4	0.983 9	0.893 3	0.954 9	0.932 8	0.973 7	0.750 5	0.826 0
主干+空域	0.950 4	0.990 0	0.904 3	0.964 1	0.941 7	0.980 5	0.759 0	0.836 6
主干+频域	0.945 1	0.990 1	0.896 9	0.962 7	0.940 2	0.978 9	0.750 0	0.829 1
主干+多域	0.954 1	0.989 1	0.905 6	0.961 5	0.943 2	0.978 6	0.769 3	0.839 8

表 9 针对空域时序特征挖掘模块基本组成单元,所提方法在 FaceForensics++数据集上的消融实验结果

Table 9 For STFM basic cell, the ablation results of the proposed method on FaceForensics++

对照组	Deepfakes		Face2Face		FaceSwap		NeuralTextures	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
主干	0.943 4	0.983 9	0.893 3	0.954 9	0.932 8	0.973 7	0.750 5	0.826 0
主干+ST-LSTM	0.941 2	0.987 5	0.898 0	0.958 0	0.936 0	0.975 1	0.754 9	0.841 2
主干+STFM	0.950 4	0.990 0	0.904 3	0.964 1	0.941 7	0.980 5	0.759 0	0.836 6

重点地关注到视频帧间差异。实验结果表明, 本文提出的伪造检测方法可以有效关注视频帧间的时序特征, 并在多个公开的人脸伪造数据集上表现突出, 具有优越性和稳定性。

参考文献:

- [1] DeepFakes GitHub non official project based on original deepfakes thread[EB].
- [2] THIES J, ZOLLHÖFER M, STAMMINGER M, et al. Face2Face: real-time face capture and reenactment of RGB videos[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2387-2395.
- [3] FaceSwap GitHub non official project based on original deepfakes thread[EB].
- [4] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. arXiv Preprint arXiv: 1406.2661. 2014.
- [5] XU W J, KESHMIRI S, WANG G H. Adversarially approximated autoencoder for image generation and manipulation[J]. IEEE Transactions on Multimedia, 2019, 21(9): 2387-2396.
- [6] 王鹏程, 郑海斌, 邹健飞, 等. 面向商用活体检测平台的鲁棒性评估[J]. 网络与信息安全学报, 2022, 8(1): 180-189.
WANG P C, ZHENG H B, ZOU J F, et al. Robustness evaluation of commercial liveness detection platform[J]. Chinese Journal of Network and Information Security, 2022, 8(1): 180-189.
- [7] MATERN F, RIESS C, STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations[C]//Proceedings of 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). 2019: 83-92.
- [8] PENG B, WANG W, DONG J, et al. Optimized 3D lighting environment estimation for image forgery detection[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(2): 479-494.
- [9] NIRKIN Y, WOLF L, KELLER Y, et al. DeepFake detection based on discrepancies between faces and their context[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 6111-6121.
- [10] LIU H G, LI X D, ZHOU W B, et al. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 772-781.
- [11] 林佳滢, 周文柏, 张卫明, 等. 空域频域相结合的唇型篡改检测方法[J]. 网络与信息安全学报, 2022, 8(6): 146-155.
LIN J Y, ZHOU W B, ZHANG W M, et al. Lip forgery detection via spatial-frequency domain combination[J]. Chinese Journal of Network and Information Security, 2022, 8(6): 146-155.
- [12] LI Y Z, CHANG M C, LYU S W. Exposing AI created fake videos by detecting eye blinking[C]//Proceedings of 2018 IEEE International Workshop on Information Forensics and Security (WIFS). 2019: 1-7.
- [13] SUN Z K, HAN Y J, HUA Z Y, et al. Improving the efficiency and robustness of deepfakes detection through precise geometric features[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 3608-3617.
- [14] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [15] WANG Y B, WU H X, ZHANG J J, et al. PredRNN: a recurrent neural network for spatiotemporal predictive learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2208-2225.
- [16] THIES J, ZOLLHÖFER M, NIEBNER M. Deferred neural rendering[J]. ACM Transactions on Graphics, 2019, 38(4): 1-12.
- [17] LIU M, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks[J]. arXiv Preprint arXiv: 1703.00848, 2017.
- [18] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv Preprint arXiv:1710.10196, 2017.
- [19] KARRAS T, LAINE S, AILA T M. A style-based generator architecture for generative adversarial networks[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 4396-4405.
- [20] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). 2017: 1510-1519.
- [21] LI L Z, BAO J M, YANG H, et al. Advancing high fidelity identity swapping for forgery detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 5073-5082.
- [22] PAN X Y, ZHANG X, LYU S W. Exposing image splicing with inconsistent local noise variances[C]//Proceedings of 2012 IEEE International Conference on Computational Photography (ICCP). 2012: 1-10.
- [23] FERRARA P, BIANCHI T, DE ROSA A, et al. Image forgery localization via fine-grained analysis of CFA artifacts[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(5): 1566-1577.
- [24] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.
- [25] YANG X, LI Y Z, LYU S W. Exposing deep fakes using inconsistent head poses[C]//Proceedings of ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019: 8261-8265.
- [26] AGARWAL S, FARID H, GU Y, et al. Protecting world leaders against deep fakes[C]// Proceedings of CVPR Workshops. 2019, 1: 38.
- [27] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 1800-1807.
- [28] AFCHAR D, NOZICK V, YAMAGISHI J, et al. MesoNet: a compact facial video forgery detection network[C]//Proceedings of 2018 IEEE International Workshop on Information Forensics and Security (WIFS). 2019: 1-7.
- [29] NGUYEN H H, YAMAGISHI J, ECHIZEN I. Capsule-forensics: using capsule networks to detect forged images and videos[C]//Proceedings of ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

2019: 2307-2311.

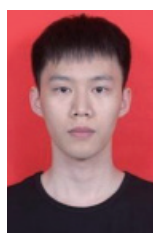
- [30] LI L Z, BAO J M, ZHANG T, et al. Face X-ray for more general face forgery detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 5000-5009.
- [31] DURALL R, KEUPER M, PFREUNDT F J, et al. Unmasking DeepFakes with simple features[J]. arXiv Preprint arXiv: 1911.00686, 2019.
- [32] QIAN Y Y, YIN G J, SHENG L, et al. Thinking in frequency: face forgery detection by mining frequency-aware clues[C]//Proceedings of European Conference on Computer Vision. 2020: 86-103.
- [33] GÜERA D, DELP E J. Deepfake video detection using recurrent neural networks[C]//Proceedings of 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2019: 1-6.
- [34] HU J, LIAO X, WANG W, et al. Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1089-1102.
- [35] LU W, LIU L Y, ZHANG B L, et al. Detection of deepfake videos using long-distance attention[J]. IEEE Transactions on Neural networks and Learning Systems, 2023, (99): 1-14.
- [36] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting[C]//Proceedings of the 28th International Conference on Neural Information Processing System. 2015: 802-810.
- [37] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[J]. arXiv Preprint arXiv: 1506.02025, 2015.
- [38] WANG Q L, WU B G, ZHU P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 11531-11539.
- [39] RÖSSLER A, COZZOLINO D, VERDOLIVA L, et al. FaceForensics: learning to detect manipulated facial images[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2020: 1-11.
- [40] LI Y Z, YANG X, SUN P, et al. Celeb-DF: a large-scale challenging dataset for DeepFake forensics[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 3204-3213.
- [41] DOLHANSKY B, HOWES R, PFLAUM B, et al. The deepfake detection challenge (DFDC) preview dataset[J]. arXiv Preprint arXiv:1910.08854, 2019.
- [42] WANG C R, DENG W H. Representative forgery mining for fake face detection[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 14918-14927.
- [43] ZHAO H Q, WEI T Y, ZHOU W B, et al. Multi-attentional deepfake

detection[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 2185-2194.

[作者简介]



朱春陶(1997-), 女, 河南周口人, 中山大学硕士生, 主要研究方向为深度伪造检测。



尹承禧(1999-), 男, 广东广州人, 中山大学硕士生, 主要研究方向为多媒体信息安全和数字图像取证。



张博林(1999-), 男, 陕西宝鸡人, 中山大学硕士生, 主要研究方向为深度伪造检测。



殷琪林(1995-), 男, 江苏泰州人, 中山大学博士生, 主要研究方向为数字多媒体取证。



卢伟(1979-), 男, 河南南阳人, 中山大学教授、博士生导师, 主要研究方向为多媒体信息安全、人工智能安全与对抗、信息隐藏。