文章编号：1003-0530(2021)12-2371-19

数字视频伪造被动取证技术研究综述

丁湘陵[1,3,5] 杨高波[2] 赵险峰[3,4] 谷 庆[1] 熊义毛[1]

(1. 湖南科技大学计算机科学与工程学院物联网工程系，湖南湘潭 411201；2. 湖南大学信息科学与工程学院通信工程系，湖南长沙 410082；3. 中国科学院信息工程研究所信息安全国家重点实验室，北京 100093；4. 中国科学院大学网络空间安全学院，北京 100093；5. 广东省信息安全技术重点实验室，广东广州 510000)

摘 要：数字视频伪造被动取证技术直接依据已获得的视频数据本身来判别其真实性，具有更好的适应性，逐渐成为取证研究领域的主流。为了从整体上梳理与描述数字视频伪造被动取证技术，分析了常见的视频伪造操作的特点和它们遗留的痕迹以及对视频被动取证的影响，从取证手段和采用技术 2 个角度，归纳与总结了基于数字视频来源、基于视频篡改遗留痕迹、基于深度学习框架和基于原始视频特征表征等视频被动取证的典型方法，并详细地探讨了视频伪造被动取证领域面临的挑战和未来的发展趋势。

关键词：数字视频取证；被动取证；真实性鉴别；深度学习

中图分类号：TP309 **文献标识码：**A **DOI：**10.16798/j.issn.1003-0530.2021.12.009

引用格式：丁湘陵，杨高波，赵险峰，等. 数字视频伪造被动取证技术研究综述[J]. 信号处理，2021，37(12)：2371-2389. DOI：10.16798/j.issn.1003-0530.2021.12.009.

Reference format：DING Xiangling，YANG Gaobo，ZHAO Xianfeng，et al. A survey of digital video passive forensics techniques[J]. Journal of Signal Processing，2021，37(12)：2371-2389. DOI：10.16798/j.issn.1003-0530.2021.12.009.

A Survey of Digital Video Passive Forensics Techniques

DING Xiangling[1,3,5] YANG Gaobo[2] ZHAO Xianfeng[3,4] GU Qing[1] XIONG Yimao[1]

(1. School of Computer Science and Engineering，Hunan University of Science and Technology，Xiangtan，Hunan 411201，China；2. College of Computer and Communication，Hunan University，Changsha，Hunan 410082，China；3. State Key Laboratory of Information Security，Institute of Information Engineering，Chinese Academy of Sciences，Beijing 100093，China；4. School of Cyber Security，University of Chinese Academy of Sciences，Beijing 100093，China；5. Guangdong Provincial Key Laboratory of Information Security Technology，Guangzhou，Guangdong 510000，China)

Abstract：The video forgery passive forensics can directly distinguish the authenticity and integrity of the obtained video by utilizing the tampering traces without the aid of any prior axillary information such as digital watermark or perceptual hash signature. Thus，it has better adaptability and has become an important topic in the field of information security. In this survey，the characteristics of universal video forgery operations，their left tampered traces and influences to video passive forensics were in-depth analyzed. From two aspects of forensics strategy and techniques used，we discuss typical algorithms and methods in passive video forensics，including video source identification based，forgeries trails based，deep learning based，original video feature representation based techniques. Each of these categories of video forensics is summarized in detail，along with a critical analysis of the state of the art. The potential research directions and challenges of video passive forensics field are also investigated in detail.

Key words：digital video forensics；passive forensics；authentication verification；deep learning

收稿日期：2021-06-09；修回日期：2021-09-05

基金项目：国家重点研发计划(2019QY2202，2020AAA0140000，2019QY(Y)0207)；湖南省自然科学基金面上项目(2020JJ4029)；信息安全国家重点实验室开放课题(2021-ZD-07)；广东省信息安全技术重点实验室开放基金(2020B1212060078)

1 引言

随着智能手机、平板电脑等便携式视频采集设备的普及,人们借助抖音和 YouTube 等视频分享网站、微信和微博等社交媒体,可以快捷地传播与分享视频。视频内容生动直观,人们逐渐习惯利用视频记录和分享生活的点滴,并且通过观看公共视频获取信息。这样,数字视频带给我们丰富的信息和愉悦。同时, 随着 Adobe Premiere、Monkey 和会声会影等视频编辑工具软件的快速发展,视频编辑与操作变得越来越容易和方便。此外,随着深度学习等人工智能技术的不断发展,2019 年以来深度视频合成取得了前所未有的视觉效果,并且很多的相关代码打包后开源,催生出大量的深度视频伪造工具,包括 ZAO、DAIN、Deepnude 等。在这样的背景下,借助社交网络等快速传播的视频良莠不齐,改变了人们"眼见为实"的传统认识。实际上,"几乎在每一次的重大事件中,社交媒体上都会出现误导性视频和照片"[1-2]。

防范和打击网络音视频信息的非法使用,可以同时采取管理和技术两种手段。一方面,政府需要及时应对新形势,加强信息监管。为此,我国国家互联网信息办公室、文化和旅游部、国家广播电视总局在 2019 年 11 月底联合印发了《网络音视频信息服务管理规定》。同年 12 月初,中国网络空间管理局颁布了《网络信息内容生态治理规定》,要求从 2020 年开始,对于利用虚拟现实和人工智能技术创建的视频进行显式标记,以避免此类内容破坏社会秩序,防范可能带来的风险。另一方面,迫切需要发展数字视频取证技术,有效地辨识包括深度视频篡改在内的各类虚假视频。视频被动取证是指直接依据已获得的视频数据本身,挖掘视频采集设备相关的异常或者视频篡改操作遗留的痕迹,判别视频的真实性和完整性。相对于数字水印和感知哈希等[3-4]主动取证技术,视频被动取证具有更好的适应性,逐渐成为取证研究领域的主流。因此,研究数字视频伪造的被动取证技术,对于以网络视频辟谣和司法取证为代表的诸多应用来说,具有重要的理论与现实意义。

当前,视频被动取证研究已经取得了一些阶段性的进展,但它们大多数是围绕视频帧间(帧插入、帧删除、帧复制和帧率上转换等)和视频帧内(视频对象复制-粘贴、基于样本合成的视频对象修复)等传统视频篡改方式展开的[3]。近几年,随着深度学习技术走向深入,涌现了深度视频篡改伪造,它们不再采取简单地搜索视频空域或者时域的相似块/片去填充待篡改区域/视频帧的传统方式,而是通过数据驱动方式,训练学习视频数据集的潜在规律或潜在特征分布,实现视频的区域/帧篡改。现有研究表明,深度视频篡改削弱或克服了传统视频篡改的取证痕迹,显著弱化了现有被动取证方法的性能,甚至使之失效。针对深度视频篡改的被动取证更加具有挑战性,其研究尚处于初步探索阶段,也是现阶段视频被动取证亟待解决的新问题。

本文针对数字视频伪造被动取证技术展开综述,依据数字视频自身固有的特性和视频伪造遗留的痕迹,从视频篡改伪造手段的技术特点和存在问题的角度,分析和总结现有的视频篡改取证方法,并且对未来本领域的发展进行展望。

2 数字视频伪造操作特点及遗留的篡改痕迹

视频是静止图像在时间维度的延伸。视频被动取证是从借鉴静止图像被动取证开始的。然而,数字视频具有空域和时域的双重特征,蕴含的视觉信息和数据量远大于相同空间分辨率的静止图像。视频被动取证可以借鉴,但不能简单地沿用图像被动取证的思想和方法,而是应该充分地考虑视频自身固有的时域特性,通过分析各种视频篡改伪造操作的特点,挖掘其可能遗留的痕迹,发展有针对性的专用取证方法。

相对于图像篡改,视频篡改伪造更为复杂和耗时[2],原因在于:首先,数字视频具有时域特点,在篡改过程中除了保持较高的空域保真度,还必须保持时域一致性,避免鬼影和抖动等问题;其次,视频数据量大,通常都是经过压缩编码后存储和传输,而视频编码标准繁多,各种编码标准的特征工具和码流语法存在显著的差异;再次,视频篡改伪造手段除了涵盖绝大部分图像伪造手段外,还有一些特有的手段,包括帧级操作(帧插入、帧删除、帧重排序和帧率上转换)和对象级操作(对象移除、复制-移动-粘贴、对象修复和放大剪切)等,如图 1(a) ~ (f)

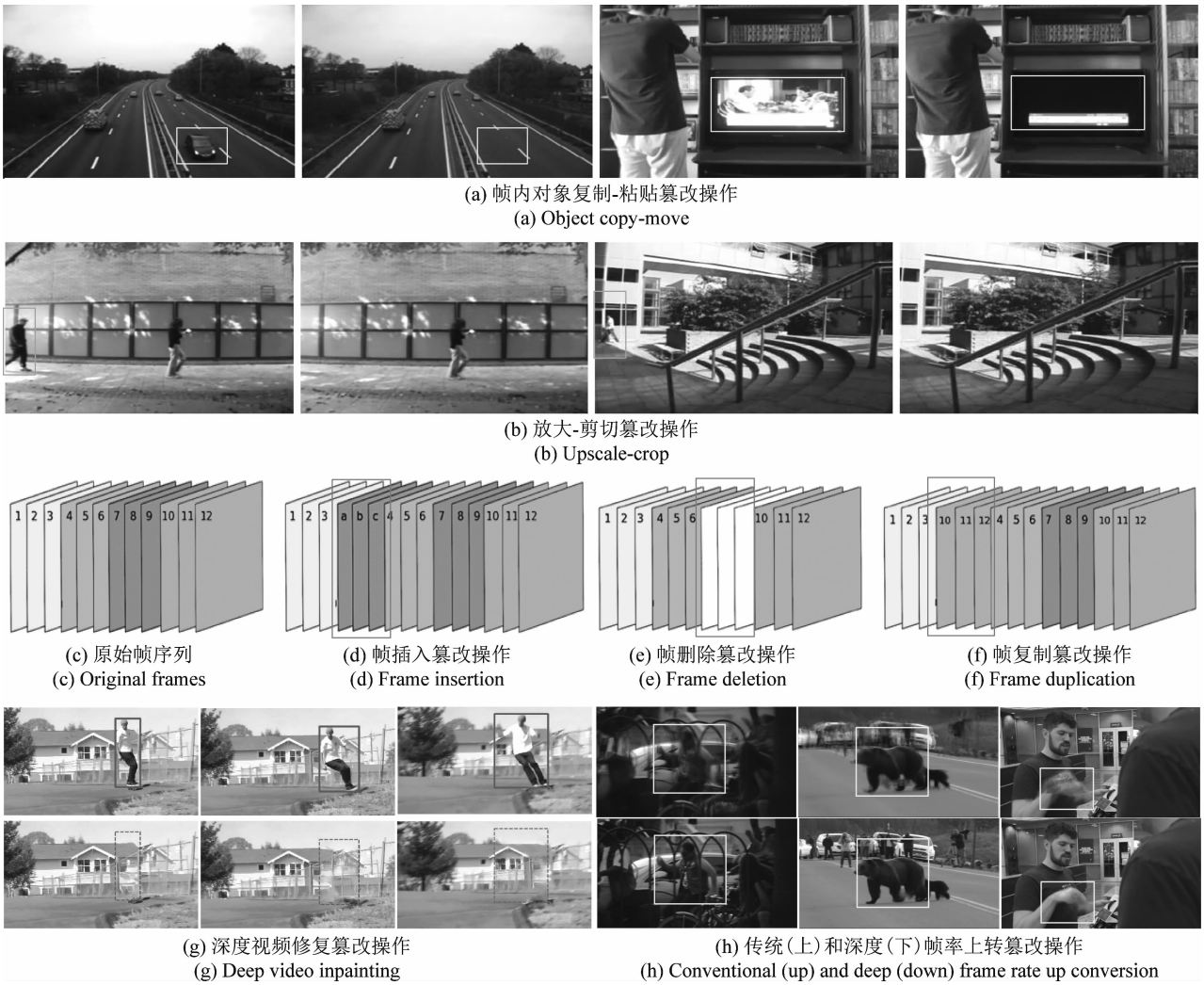

图 1　视频篡改操作示例

Fig. 1　The examples of video forgery operations

所示。最后,还出现了一些结合视频特点的深度伪造技术,主要集中在深度视频修复和深度视频帧率上转换,如图 1(g) ~(h)所示。因此,现在的数字媒体取证主要集中在图像被动取证[5],视频被动取证在研究的广度和深度都明显不足。实际上,相对于静止图像压缩编码,视频压缩编码的压缩比更高,视频篡改痕迹可能与压缩编码的效应混淆在一起,增加了取证算法设计的难度。当然,数字视频也具有其独特的特性,由于视频具有时间维度,它实际上也在帧间连续性和帧间累积特性方面提供了更多的取证线索。

视频篡改伪造本身是一个信号处理过程,并且通常是不可逆的。因此,视频篡改不可避免地遗留一些痕迹,并且痕迹通常与具体的视频篡改手段和技术方式有关。视频篡改伪造从技术角度可分为传统篡改和基于深度学习的篡改两类。其中,传统视频篡改遗留的痕迹主要体现在:① 视频帧内篡改本质上类似于图像篡改,其取证可以从成像传感器模式噪声、颜色插值模式、光照一致性、对象边缘效应等图像取证通常利用的痕迹角度进行。一般来说,篡改区域的统计特性不同于其他区域,而自然视频的统计特性在帧内呈现一致性。② 对于视频帧间篡改,插入/删除帧的光流、运动矢量场、运动补偿后的残差等可能存在时域的突变,并且这种突变可能存在周期性。③视频篡改过程为了满足空时域的一致性约束,可能会强制空域的篡改,从而在空域遗留伪影、人工拼接痕迹、对象融合边界效应(例如轻微模糊)等细微痕迹。

深度视频伪造通常采用编码器-解码器结构。从信号处理的角度看,它也可以视为一个复杂的信号处理系统。类似于自然图像的相机内部信号处理过程遗留传感器模式噪声和颜色插值模式,深度视频伪造也会遗留一些与深度模型的信号处理过程有关的痕迹,主要体现在:①采用的深度网络模型在保证收敛所设计的损失函数,例如像素值间的 L_1 损失、风格和样式间的感知损失、帧间时域一致性损失等,某种程度上是通过能量最大化进行的,与主观视觉并不完全吻合;②在特征提取过程中采用不同的卷积核设计,旨在尽可能多地保持更多的信息,使得学习到的特征能够表征目标,实现重构时能够有效地修复目标区域,这也与能量密切相关;③在目标对象合成过程中,都采用了上采样层,尽管上采样方式包括转置卷积、最邻近插值和双线性插值等,已有文献研究表明它们都不同程度地引入了棋盘效应,体现为合成对象的能量奇异性。上述这些原因导致视频篡改伪造痕迹在视觉上表现为细微的局部纹理和结构失真、轻微模糊和时域运动轨迹不够平滑等,与原始视频相比,能量存在一定的奇异性。

3 视频伪造被动取证技术

依据取证的目的和技术手段,视频被动取证技术可以大致分为四大类:第一类是数字视频来源取证;第二类针对视频内容伪造,依据遗留的痕迹人为设计特征实现检测;第三类基于 CNN 等深度学习框架,在样本驱动下自动学习辨别性强的识别特征,完成端到端的篡改检测和识别;第四类基于原始视频特征表示的单类(One-Class, OC)检测方法,从原始视频或特定人物自然视频中提取语义或统计特征,有效检测未知伪造方法合成的视频。图 2 是视频被动取证的详细分类。

3.1 视频来源被动取证技术

视频采集过程与图像采集类似,但是视频编码标准与静止图像编码具有较大的差异。因此,图像来源被动取证可以为视频来源取证提供有益的检测思路,但无法直接扩展到视频来源取证。依据视频采集、处理与传播的过程,现有的视频来源取证方法可以分为三类,分别为基于视频成像设备和文件格式一致性的被动取证技术、基于视频压缩编码器的被动取证技术和面向网络传输技术的被动取证技术,其特性和优缺点如表 1 所示。

1) 基于成像设备和文件格式一致性的被动取证技术

由于自然视频在采集过程中, 受成像原理和传感器物理特性的影响, 会遗留一些与设备有关的特有信号。最终采集到的视频以不同的文件格式存储, 从而具有该文件格式的存储特性。利用特有的这些信号或特性作为取证线索, 就可以对视频的篡改进行检测。常用的特征包括光子响应非均匀[6]、电网频率[7]和视频容器结构[8]。已有的研究结果表明: 这类方法具有较高的检测准确率, 但都假设待检测视频的来源或格式已知。

2)基于视频压缩编码器的被动取证技术

现有的视频资源大多数都是以压缩格式存储,

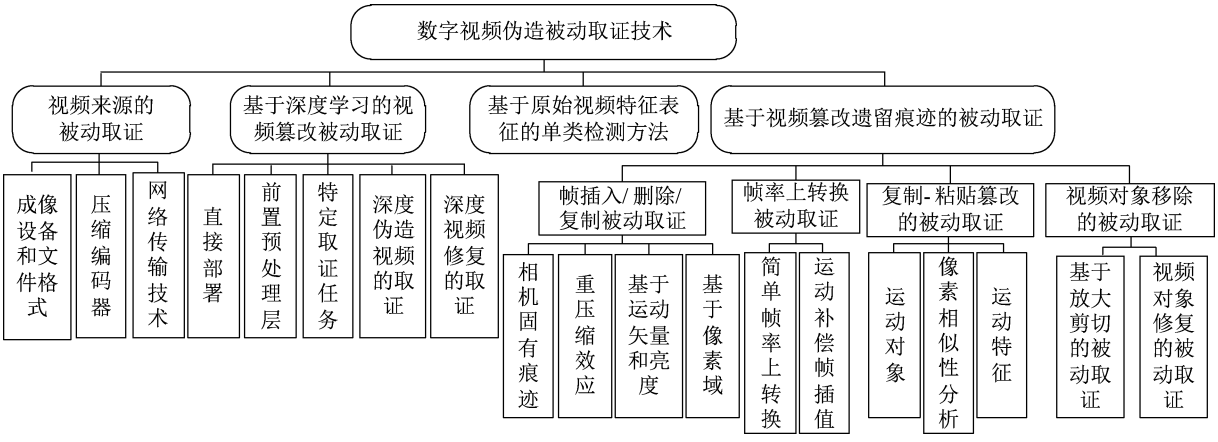

图 2 数字视频伪造被动取证技术分类

Fig. 2 The classification of video passive forensics techniques

表 1　视频来源被动取证技术总结

Tab. 1　Summary of video source-based forensics techniques

类型	采用的特征和参考文献	优缺点
基于成像设备和文件格式一致性的被动取证	光子响应非均匀[6]	具有较高的检测准确率,但都假设待检测视频的来源或格式已知
	电网频率[7]	
	视频容器结构[8]	
基于视频压缩编码器的被动取证	编码器[9-10]	在指定的编码器上具有 90% 以上的检测率,但是针对新的编码器性能退化严重
	QP 和运动矢量[11-12]	针对 H.264/AVC 进行检测,其他编码器条件如 HEVC 下性能未知
面向网络传输技术的被动取证	质量评价指标[13]	算法成熟度不够,网络快速发展以及其复杂的体系和参数导致无法直接判断视频的网络来源

且第一次压缩编码过程由摄像机内置的专用编码芯片完成,这也就提供了视频来源鉴别的另一种方法,把鉴别摄像机的任务简化成鉴别视频编码器。在视频编码体系中,某些编码参数的选择是编码标准未规范的部分,由具体的编码算法与待编码信号特征共同决定。在视频编码过程中,可供选择的编码参数就特别多,而且不同的设备制造商会采用不同的算法来控制这些编码参数,即利用编码框架中的开放模块,根据需求开发新方案,使得不同品牌摄像机的内置视频编码器不同,因此,取证分析者在鉴别编码器时,可以利用这些开放部分来验证视频的来源。现有的方法主要识别视频采用的编码器如 H.264/AVC、MPEG2 等[9-10],或压缩参数如量化参数、运动矢量等[11-12]。但是,这类方法都是假设摄像机采用的压缩参数来进行来源鉴别。

3)面向网络传输技术的被动取证技术

光纤入户解决了视频实时传输的带宽需求,而不同的网络通道由于丢包、误差等不同,会影响接收端的重建视频内容。如果从重建帧中反向推导出通道损失模式,如误差概率、突发度或其他误差分布统计特征,便可鉴别传输协议或码流传输设备。现有的鉴别网络传输痕迹的算法主要使用传输统计特性估计通道失真。A. R. Reibman 等[13]使用质量评价指标来估计重建视频中的丢包损失,进一步还利用网络统计特性、视频的时空域损失以及误差传播来计算通道损失[14]。但是,这些算法都不成熟,将其应用于判断视频的网络来源还有很多的技术障碍。

3.2　基于视频篡改遗留痕迹的被动取证技术

这类方法是视频被动取证研究的重点,并且绝大多数成果都集中在帧间篡改和帧内篡改两类视频篡改方式上。帧间篡改是指时域上的篡改,包括帧插入、帧删除、帧复制和帧率上转换等,容易导致帧间统计特性的异常,其取证通常以时域不一致性为线索进行。帧内篡改是指空间域上的篡改,但在篡改过程中需要考虑帧间运动和帧内内容的时域一致性,通常产生包括混淆伪影、模糊边界和块/片在时空域的异常相似性等操作痕迹,包括对象复制-移动-粘贴和对象移除等。这类取证技术采用的特征和优缺点总结如表 2 所示。

3.2.1　针对帧插入/删除/复制的被动取证技术

这类篡改主要涉及同源视频帧的插入、异源视频帧的插入、视频帧的连续删除以及同源视频帧的复制。由于处理的是多个连续视频帧的操作,那么在相机固有痕迹、重压缩效应、运动矢量和亮度、像素域等方面都会遗留下痕迹。下面从遗留痕迹角度阐述除帧率上转换被动取证技术外的帧间篡改的被动取证技术。

1)基于相机固有痕迹的被动取证技术

数字视频在拍摄过程中由于采用特定的设备,必然会遗留下与相机相关的固有的特定痕迹。这类固有痕迹,例如光响应不均匀性,最开始用于相机来源识别,现在也被研究者用于来检测视频帧间伪造。这类取证方法主要用于异源视频帧的插入检测。Wang W[15]最先利用单帧场间的运动与相邻帧场间的运动间的相等性来判断篡改视频的运动异常,采用差分算法进行运动估计,实现交织(interlaced)和解交织(De-interlaced)视频的检测。它不仅获得较好的取证结果,而且对压缩及帧率变换均不敏感。Kobayashi M 等[16]利用异源视频帧插入导致的光子散粒噪声(Photon Shot Noise,PSN)的时域不一致性实现帧间篡改的定位检测。该算法在未压缩状态下能获得 97% 的定位精度,但是,一旦发生压缩,性能退化严重。这类取证技术具有与视频成像设备和文件格式一致性的取证技术一样的缺陷,且要求插入的视频帧异源。

表 2 基于视频篡改遗留痕迹的被动取证技术总结

Tab. 2 Summary of video forgeries trails-based forensics techniques

类型	子类	采用的特征和参考文献	优缺点
针对帧插入/删除/复制的被动取证	基于相机固有痕迹	单帧场间与相邻帧场间的运动异常性[15]	能获得较好的检测结果,且对压缩和帧率变化不敏感,但是仅能针对交织和解交织视频
		光子散粒噪声的时域不一致性[16]	尽在未压缩状态下能获得 97% 的定位精度,压缩、噪声或模糊情况下,性能退化严重
	基于重压缩效应	双量化效应和运动估计误差[18]	有效检测 MPEG-1 和 MPEG-2 变码率情况下的重压缩,但不适用定码率情况下的视频重压缩
		广义 Benford 法则[19]	仅在 MPEG-1 定码率和变码率编码模式下能有效检测视频重压
		Markov 特征[20-21]	仅对与 JPEG 压缩类似的 MPEG-4 视频编码标准有效
		量化 DCT 系数直方图[22]	仅对 H.264 编码有效
		帧内、帧间宏块变化[23-24]	分别对 H.264 和 HEVC 在相同编码参数下的重压缩检测
		in-loop 过滤和 PU 变化[25]	仅对 HEVC 的转码检测有效
	基于运动矢量和亮度	运动补偿边缘差分[27-28]	专门针对 MPEG 提出 GOP 估计,在其他编码标准下性能退化
		预测误差[29]	在 H.264 仅 I,P 帧编码的情况下具有较高的检测精度,但是具有 B 帧情况下未深入讨论,且阈值判定根据经验设置
		帧间光流/重力场的连续性[30-33]	实现帧删除, 帧复制或帧插入的篡改检测,但需要较多的计算资源且容易受到噪声的影响
	基于像素域	块亮度变化描述器[34]	仅对 25 帧以上帧插入具有 86% 的检测精度,但是在帧数较少、噪声、模糊环境下,检测性能下降
		切片相似性[35]	仅在 4 个视频上测试,且误判率较高
		灰度值相关系数[36]	对帧插入和帧删除具有高于 95% 的检测精度,但是在强压缩、噪声、模糊状态下,检测精度下降
		Zernike 矩和局部二值模式[37-38]	在未压缩条件下获得高达 90% 以上的检测效果,但是在强压缩状态下,精度普遍下降
针对帧率上转换的被动取证	针对简单帧率上转换	结构相似度[41]	样本全面,置信度高,但主要针对采用帧复制的商业软件 ImToo,AVS
	针对运动补偿帧插值	预测误差[43],运动效应[44]和噪声变化[45-46]	考虑帧率上转和下转情况,但当帧率 24 帧/秒上转到 25 帧/秒,检测准确率陡降,且只能进行帧率估计,不能实现插值帧定位
		边界强度[47]、平均纹理变化[48]和运动效应[50]	可检测帧复制、帧平均和运动补偿帧插值,但检测算子选择需要先验知识,待检测视频具有场景变化时,容易出现误判
		残差信号[51]	对 11 种不同的运动补偿帧插值算法进行分类,但是,检测算法复杂度较高
		效应指示图和切比雪夫矩[49]	在强压缩、噪声、模糊状态下具有较高的鲁棒性,但是,不能有效检测帧复制和帧平均,参数不能自适应设置
针对复制-移动-粘贴篡改的被动取证	基于像素相似性分析	空时像素相关性[52]	仅对固定背景视频有效
		时域噪声残差[53]	仅对不同来源视频帧有效
		旋转不变[54]	在强压缩、噪声、模糊状态下,性能退化严重
		指数-傅里叶矩[55]	仅对镜像篡改有效
		隐写特征[56]	仅对视频背景填充移除对象有效,但是移除对象大小未进行深入分析且不能实现篡改区域的定位
	基于运动对象	鬼影效应[57],能量可疑度[58]和能量差值[59]	主要针对运动对象从固定背景视频移除的情况
		小波基[60]	仅对移动对象移除且边界轮廓存在模糊或鬼影效应有效
		运动轨迹[61]	针对自由飞行物体的伪造,但要求建立运动对象的轨迹模型,且参数计算量大,普适性不强
	基于运动特征	篡改区域光流异常[61]	在 REWIND 数据集获得平均 89% 以上的定位精度,但是,对于变化的 GOP 结构失效

续表 2

类型	子类	采用的特征和参考文献	优缺点
针对视频对象移除的被动取证	视频对象修复	模糊的时域不规则变化[62]	主要针对静止背景对象修复,真实视频误判率较高
		光流不连续性[63]和海森矩阵[64]	计算复杂度较高,仅能实现区域级篡改定位
		空时域 LBP[65]	实现区域级篡改检测和定位,但是在强压缩、噪声、模糊状态下,性能退化
		随机过程参数变化[66]	仅能实现视频序列级和片级检测和定位
	针对放大-剪切	传感器噪声[67-68]	在不同缩放因子情况下具有平均精度 98%,但是,方法的性能与视频内容以及判决阈值密切相关

2)基于重压缩效应的被动取证技术

视频压缩编码过程虽然在一定程度上阻碍了图像取证方法向视频领域进行扩展,但是,由于它会在生成视频中留下压缩痕迹,而篡改过程不可避免地需要进行二次压缩,也就是先解压,再执行篡改,最后重新编码篡改视频。因此,可以通过验证压缩痕迹的一致性进行视频的真实性鉴别[17]。Wang W[18]构建 I 帧量化 DCT 系数直方图中周期性的双量化效应的静态特征和基于运动估计的误差的时域特征,利用傅里叶变换等工具实现 MPEG-1 和 MPEG-2 的重压缩检测。虽然该方法能够有效检测变码率情况下的视频重压缩过程,但其不适用于定码率情况下的视频重压缩检测,这是因为在定码率编码模式下,码率控制策略会为每个宏块选定一个量化参数,不同的量化参数将会混叠双量化效应。随后,Chen W 等[19]基于广义 Benford 法则提出了一个针对 MPEG-1 视频的重压缩检测方法。通过分析 MPEG-1 视频的 I、P 和 B 帧中非零量化 DCT 系数的第一位数分布,发现它们均满足广义 Benford 法则,而视频重压缩过程会破坏该概率分布特性。实验结果表明,在定码率和变码率编码模式下,该方法对 MPEG-1 视频的重压缩检测均能取得理想的效果。根据 Chen C 等[20]的实验结果,Markov 特征能够有效区分原始和经过重压缩的 JPEG 图片和 MPEG-4 视频压缩编码标准中预测残差也采用类似 JPEG 图片压缩编码的处理方式,Jiang X 等[21]提出将 Markov 特征用于 MPEG-4 视频的重压缩检测,实验验证 Markov 特征则能够有效区分原始和经过重压缩的 MPEG-4 视频。Liao D 等[22]首先提出了针对 H.264 视频的重压缩检测方法,他们在实验中发现重压缩会向 H.264 视频的量化 DCT 系数分布直方图中引入视觉上可察觉的扰动,且扰动程度和前后两次压缩所用量化参数的差值成正比。以此为取证线索,作者以量化 DCT 系数的直方图分布数值作为特征,训练 SVM 分类器进行 H.264 视频的重压缩检测。最近,蒋兴浩团队根据视频重压质量退化理论,构建帧内、帧间宏块变化特性对 H.264 和 HEVC 分别进行了在相同编码参数下的重压缩检测,实验表明,提出的特征能有效检测视频重压缩[23-24]。此后,Sun 等人利用视频转码工具转码为 HEVC 压缩视频过程中,in-loop 过滤和 PU 划分的变化引起的转码视频的质量变化,提出取证特征实现了 HEVC 的转码检测[25]。这类方法都是专门针对特定的视频编码标准提出由针对性的取证特征实现重压缩检测,缺乏通用性能力。此外,这类方法虽然能有效的检测重压效应,但是它只能表明该视频经历了重压,且不是所有经过重压的视频都会发生篡改操作,还需融合特定伪造操作的取证特征才能验证是否发生篡改。

3)基于运动矢量和亮度的被动取证技术

视频的帧间操作通常会导致帧间统计特征,尤其是运动矢量和亮度的异常,因此,通过提取帧间运动矢量或亮度信息的统计特征,并在时域窗口滑动实现帧间篡改的检测。冯春晖等[26]考虑到视频序列中具有不同的运动强度、内容和抖动变化等的干扰,从运动残差提取稳健的波动特征,利用滑动窗口识别帧删除点。实验结果表明,该方法能有效的消除重编码 I 帧、异常亮度变化、聚焦变化和抖动干扰,并达到 90% 的真正率。苏育挺[27],董琼[28]利用运动补偿边缘残差分别提出帧删除检测算法,并确定视频的 MPEG 原始 GOP 结构。Stamm M C 等[29]建模视频帧删除/添加前后 P 帧的预测误差,从原始帧与连续帧均值差提取统计特征,阈值化计算实现帧篡改检测。Chao J[30]、Wang W[31]、Wu Y[32]和 Jia S[33]

等研究者都利用帧操作时相邻帧间光流矢量或重力场的连续性,实现帧删除、帧复制或帧插入的篡改检测,取得令人满意的结果。但是,运动信息的提取需要较多的计算资源且容易受到噪声的影响,因此,提高算法的识别精度,同时减少算法的复杂度是本类算法需要解决的核心问题。Zheng L 等[34]提出利用时域亮度的变化提取块亮度变化描述器实现帧插入和删除。但是,该算法仅对25帧以上的帧插入有86.3%的检测精度和79.4%的定位精度,在插帧数目较少、噪声、模糊等环境下,检测精度下降严重。

4)基于像素域的被动取证技术

Lin G S 等[35]利用视频切片间的相似性检测帧复制篡改,然而,该方法仅在4个视频上进行测试。Wang Q[36]等提出利用灰度值相关系数的帧间一致性来检测帧间篡改,对于帧插入和帧删除具有高于95%的检测精度。Liu Y[37]等采用Zernike矩提取反色度信息分析帧间篡改。张珍珍等人[38]首先利用局部二值模式编码视频帧,随后,计算局部二值模式编码的帧间相关系数,最后使用切比雪夫不等式实现帧插入和删除点的定位。虽然这类方法在未压缩条件下能获得高达90%以上的检测或定位效果,但是在强压缩状态下,检测精度都普遍下降严重。

3.2.2 针对帧率上转换的被动取证技术

帧率上转换依据运动对象的运动轨迹,在两个连续帧间插入一个或多个中间视频修复帧,它属于视频特有的操作手段。它最开始应用在电影和电视节目的制作过程,用于生成慢运动效果和视频预测的外插帧(extrapolation frame)。然而,它也可被伪造者用于更改视频原始语义或属性的恶意目的。例如,不同相机拍摄的视频可能具有不同的帧率,就需要在低帧率视频中插入合成的视频帧来保证拼接的视频具有一致的播放速度;由于能消除帧间篡改引起的跳跃(jump-cut)效应,可以用作反取证手段,使得基于光流连续性的帧间篡改取证工具失效[39]等。根据帧率上转换的划分,其取证技术可以分为基于简单帧率上转换的被动取证和基于运动补偿帧插值的被动取证。下面将从帧率上转换划分角度阐述该类被动取证研究现状。

1)针对简单帧率上转换的被动取证技术

简单帧率上转直接采用帧复制和帧平均来实现两帧间插入新帧,这样的操作必然导致插入帧与前一帧具有高度的相似性或与前后两帧具有比例关系[40]。研究人员根据插帧原理研究简单帧率上转换的篡改检测并根据插帧周期实现原始帧率的估计。2007年,Wang W 等[15]最先提出利用运动自适应算法检测交织和去交织视频的帧平均插值的篡改检测。2014年,边山等[41]利用结构相似度处理由商业编辑软件ImToo,AVS等生成的高帧率视频,发现相邻两帧间存在峰值相似度,从而利用傅里叶变化估计原始帧率。随后,林晶等[42]从光流角度检测了帧复制的篡改。虽然这些算法能获得令人满意的结果,但是,这些算法不能检测运动补偿帧插值生成的高帧率视频,也不能定位插值帧位置。

2)针对运动补偿帧插值的被动取证技术

运动补偿帧插值根据前、后视频帧间的运动矢量,在两者间插入的视频帧中合成运动对象,填充由对象运动导致的孔洞(hole)区域。它在帧内根据运动对象的运动轨迹合成新的对象区域;在帧间合成新的视频帧,不同于简单的帧复制,属于视频特有的帧间操作方式。因此,它的关键操作是运动对象轨迹的精确估计和像素合成。其中,像素合成主要涉及运动对象在修复帧中的合成和由于运动对象移动引起的孔洞区域的填充,也就是运动对象引起的遮挡和被遮挡区域的处理。然而,受运动估计精度、遮挡和被遮挡处理以及融合精度等因素的制约,修复的视频帧在运动对象以及其边缘的视觉质量、对象的运动轨迹和修复帧的周期性等都可能与真实视频存在细微的差异,从而提供了的取证线索。本节中的被动取证技术主要根据这些取证线索提出有针对性的取证方法。

Bestagini[43], Dae-Jin Jung[44]和李然[45-46]分别利用预测误差、运动效应和噪声变化,检测采用运动补偿帧插值技术的视频帧修复操作,并且能够估计原始帧率;湖南大学杨高波教授团队围绕这类篡改方式,开展了一系列的研究:分别利用边界强度[47]、平均纹理变化[48]、模糊效应[49]和运动效应[50]的周期变化进行检测,并且能够定位修复帧的位置;湖南科技大学的丁湘陵等深入分析视频运动补偿帧插值原理,提出从残差信号(Residual Signal)的角度进行取证[51]。其中,残差信号是指生成的插值帧和缺失的原始帧之差。通过理论建模和实验

观察，发现不同的视频帧插值方法所遗留的残差信号存在差异，从而将不同视频帧插值方法的辨别问题转化为区分不同程度的残差信号问题。它对于同时经历了帧插值和高效压缩的视频，既能判别帧插值操作，还能进一步揭示具体的类型（包括帧平均、帧重复、运动补偿帧插值和多种公开的帧率上转换软件工具等），实现了更深层次的取证目标。此外，该作者还利用 artifacts 区域和高残差能量间的强相关性，设计效应指示图自适应标记候选区域，并结合切比雪夫矩建模时域不一致性，进行插值帧的鲁棒检测和定位[49]。

3.2.3　针对复制-移动-粘贴篡改的被动取证

通过复制视频中某帧的背景区域或对象完成某些重要目标或运动对象的掩盖或添加是一种常见的视频篡改伪造。虽然这类篡改操作都会进行后处理操作来保证粘贴边缘与背景间的一致性，但是，复制区域和相对应的粘贴来源区域能基本相似。因此，此类篡改的取证工作主要通过查找高度相似的区域或对象来进行复制-移动-粘贴的区域定位。根据取证线索，可以分为基于像素相似性分析的被动取证、基于运动对象的被动取证和基于运动特征的被动取证。

1）基于像素相似性分析的被动取证

这类检测方法直接来源于篡改操作原理，通过穷举搜索视频连续帧内相同区域或视频帧内不同区域间的像素相似性分析来实现篡改取证。针对固定背景视频，Wang W 等[52]计算视频空时像素间的相关系数实现视频复制-粘贴的定位。Hsu C C 等[53]利用时域噪声残差间的相关性实现对象篡改区域的定位。D' Amiano L 等[54]利用 PatchMatch 的视频版本实现旋转不变复制粘贴篡改取证。Su L 等[55]利用指数-傅里叶矩进行具有镜像的篡改区域的定位检测。Chen 等[56]利用隐写特征建模视频运动残差，实现视频对象移除篡改帧的定位。

2）基于运动对象的被动取证

这类取证主要利用运动对象添加/移除引起的与背景间的过滤效应角度考虑。针对运动对象从固定背景视频移除的情况，现有的工作主要从遗留的鬼影效应[57]、背景能量可疑度[58]、时域能量差值[59]角度识别篡改视频；陈日超等[60]也提出一种利用小波基对添加对象轮廓提取统计特征实现对象篡改检测方法。这些方法具有一个共同缺点在于：要求固定背景视频，且移除/添加的对象是运动的，应用受到很大的局限。近年来，还出现一些特定条件下的视频对象篡改的取证方法。针对对象运动轨迹的篡改，Conotter V 等[61]对自由飞行物体的三维抛物线运动轨迹进行建模，估计出三维空间的运动，并与正常轨迹对比，从而判定视频的真假。该算法从运动几何出发，性能不受视频压缩和视频质量的影响，能够取得理想的检测效果。显然，这类方法要求建立运动对象的轨迹模型，且模型参数计算量大，普适性不强。

3）基于运动特征的被动取证

视频对象的复制-移动-粘贴要求保证视频空时域的一致性，而实际的操作中，很难保证视频的运动连贯性，一般都假设为直线运动，那么对象篡改区域必然存在运动特征的异常。现有的方法提取运动特征主要从光流和连续帧差角度提取。代表性的工作有：陈盛达等[56]利用隐写分析中的冲突操作计算连续八帧中的中值计算待分析帧的运动残差，接着利用图像隐写分析特征建模该运动残差，使用最大投票机制实现原始帧、对象移除帧和双压帧的区别。但是，它不能实现篡改区域的定位。Bidokhti A 等[61]利用篡改对象区域光流的异常，计算光流变化因子，再根据它峰值的周期性和自相关性定位篡改区域。该方法能检测 REWIND 数据集的视频序列，获得平均 89.4% 的定位精度。但是，该方法对于变化的 GOP 结构失效。

3.2.4　针对视频对象移除的被动取证

视频对象移除主要移除不期望出现的对象，根据操作方式，可以分为视频对象修复和放大剪切移除两种典型方式。视频对象修复最初的目的是用于陈旧影像的划痕恢复。实际上，它也用于恶意的语义对象移除。从本质上看，视频对象修复是一个病态问题，原因在于：对象移除容易引起部分信息的丢失，难以依据周围区域唯一正确地恢复。当然，视频具有时域特性，使得移除区域可能在相邻的视频帧中找到更多的线索。视频帧内对象移除的另一种简单操作就是通过将视频帧分辨率放大，通过裁剪的方式将呈现犯罪事实的视频边缘对象移除。

1）视频对象修复的被动取证

现阶段，大多数视频修复篡改的被动取证都是

围绕传统的视频修复篡改方法，依据其产生的篡改痕迹开展的。其中，视频对象修复篡改取证主要针对基于扩散的视频对象修复和基于样本合成的视频对象修复。基于扩散的视频对象修复仅能修复狭小的区域，例如陈旧影像的划痕。该方法在扩散方向上呈现像素一致性，并且当修复区域较大的时候，会产生明显的模糊现象。基于样本合成的修复利用邻近空时域的相似样本进行修复，虽然能填充较大区域，但是缺乏恢复非重复或复杂纹理区域的能力，尤其是在视频空时域找不到匹配样本的时候。现有的取证方法就是依据这些线索展开的。代表性的工作有：Lin C S 等从不同维度视频切片角度，分析由于模糊导致的不规则变化，实现区域级检测和定位[62]；Saxena S 等发现视频修复容易导致光流出现不连续现象，通过对光流强度计算转移概率矩阵，实现视频修改的检测和篡改区域的定位[63]；Mustapha Aminu Bagiwa 等采用海森矩阵的统计相关进行检测和定位[64]；北京交通大学赵耀和倪蓉蓉教授团队提出了一种先进行帧对齐，再利用空间 LBP 检测可能的篡改区域，并且结合时域 LBP 剔除误判区域的方法，可以实现区域级的视频修复篡改检测和定位[65]。Aloraini M 等人[66]从序列和片分析的角度实现视频对象篡改的检测和定位。通过将视频序列建模为随机过程，分析过程参数的变化检测视频篡改。建模视频序列为正常片和异常片的混合模型来进行片级分析，通过识别每个片的分布来实现篡改片和原始片的区分。再根据异常片的时域运动定位视频篡改区域。

2）针对放大-剪切的被动取证

这类操作通过放大和剪切保持与原有视频一致的空间分辨率，那么这些被操作的视频帧必然经历重采样过程，因此，这类篡改的取证线索来源于重采样痕迹。代表作有：Hyun D K 等[67]使用传感器噪声作为取证特征，分析参考传感器噪声与放大帧间的相关性实现放大剪切的检测。Singh R D 等[68]也利用传感器噪声，分析像素相关峰值和局部传感器噪声异常变化捕捉重采样痕迹。它在各种测试环境下都获得了满意的结果，尤其在不同缩放因子情况下平均精度高达 98%。但是，该类方法的性能与视频内容以及判决阈值密切相关。

3.3 基于深度学习的视频篡改检测技术

传统视频篡改取证方法都普遍采用“篡改痕迹提取+取证特征设计+分类器选择”的框架。其中，取证特征是依据提取的篡改痕迹而专门设计的，分类器则通常采用支持向量机。由于设计的取证特征针对性过强，往往存在泛化能力不足的缺陷，一旦迁移到其他篡改操作的取证时会出现性能退化的问题。此外，由于数字视频篡改可能在伪造过程中遗留下众多语义或统计异常，而精确表征这些异常难度较大。因此，以卷积神经网络（Convolutional Neural Networks，CNN）为代表的深度学习被引入到视频被动取证领域，自主学习潜在的特征表征，出现了一些基于深度学习的视频被动取证工作。根据网络模型构建方式的不同，基于深度学习的视频篡改取证方法可分为三种：第一种直接部署已经存在的深度网络模型；第二种是通过在深度网络模型前添加前置预处理层的方式提取视频篡改特定痕迹实现视频篡改检测；第三种则通过修改 CNN 架构和损失函数，设计特定视频取证任务的视频篡改检测技术。以上三种主要针对的是视频帧内和帧间篡改取证，还有专门针对深度视频伪造技术（例如 Deepfake，DeepApp 等）的检测方法以及针对深度视频修复的视频取证技术。这类取证技术采用的网络模型/特征和优缺点总结如表 3 所示。

1）直接部署已存在的深度网络模型

这类视频篡改取证方法直接在已有的深度网络模型（例如 Xception 和 C3D）上部署，主要集中在帧间篡改，使用 CNN 提取特征，根据特征在时间域上的变化识别篡改。Long C J 等[69]直接利用 C3D 网络模型建模 16 帧序列，并检测中间帧是否为帧删除点，此外，还引入置信值峰值和时域窗口尺度抑制噪声的干扰。实验结果表明，在具有快速相机运动和放大变化的视频数据集中，该算法不仅能很好的识别帧删除操作，还能准确的检测出帧删除点的位置，但是，对于前后 16 帧进行的帧删除不能有效检测，此外，未对压缩视频进行测试。随后，该作者还提出一个由粗到精的深度卷积网络检测和定位帧复制[70]。该网络直接采用 I3D 网络模型获得最可能的候选复制帧序列和选择序列，随后采用基于 ResNet 的孪生网络识别复制帧序列和对应的选择序列，最后，应用不连续检测器微调，在 I3D 网络实

表 3　基于深度学习的视频篡改检测技术总结

Tab. 3　Summary of deep learning-based forensics techniques

类型	采用的网络模型/特征和参考文献	优缺点
直接部署已存在的深度网络模型	C3D 网络模型[69]	在具有快速变化和放大变化的视频中能有效识别帧删除操作，但是，对于前后 16 帧进行的帧删除不能有效检测，此外，未对压缩视频进行测试
	I 3D+ResNet 网络模型[70]	实现复制帧检测和来源确定，但是，复制的帧必须在 15 帧以上，且没有在压缩状态下进行验证
	3D-CNN [71]	在较广的视频质量下获得平均 97% 的精度，但是仅在 H. 264 固定 GOP 和定码率下进行测试
前置预处理层	高通滤波器+递归神经网络[72]	能实现蓝屏抠像和社交视频的真实性。但是，这类方法没有考虑时域信息，且采用逐帧操作，容易收到视频编码的影响
	离散余弦，视频质量误差+CNN [73]	
特定取证任务的网络结构	编码器信息和视频质量信息[74]	该方法在 CIF、4CIF、PAL 和 720p 分辨率下的视频，对于 MPEG2、MPEG4、H. 264 和 H. 265 具有满意的检测结果，但在小篡改区域的检测精度有待提高
	高频组件+平均池化和全局平均池化+1×1 卷积核[75]	在不同编码参数下，该模型获得满意的 H. 264 重定位 I 帧检测，但是针对仅第一帧为 I 帧其他为 P 或 B 帧编码的情况下无法进行检测
	帧内编码景深设计专用的网络[76]	实现双压和比特率转码的检测，但是主要针对 HEVC 编码的视频
	空间富裕模型和 C3D 的时空三路网络模型[78]	实现时域和空域篡改区域的定位，但是计算复杂度较高，需要大量训练数据和长的训练周期
	基于遮挡图、连续视频帧的 CNN [79-80]	实现视频帧率上转换的检测和插入帧的定位，但是不能实现帧率下转的检测
针对深度伪造视频的被动取证	轻量级 3D 卷积神经网络[82]和多级卷积神经网络[83]	这类取证方法具有很强的领域特性，如果将此类方法直接逐帧应用或迁移到其他深度视频伪造取证，由于没有充分利用时域信息或修复视频中并无人脸的领域特性，导致检测率并不高
	MesoNet+Flow-CNN	
	变形效应	
针对深度视频修复伪造的被动取证	高通预处理模块、4 个级联的 Resnet 块和 2 个上采样块的网络模型[90]	实现像素级的深度图像修复定位，不同的图像修复方法都会遗留下部分相似痕迹。但是没有充分利用视频特有的时域信息，也没有考虑训练样本与测试样本之间的失配问题，不能直接推广到深度视频对象修复篡改的取证
	神经架构搜索和注意力机制[91]	
	空时高通过滤层+精化子网络[92]	能有效定位篡改区域，但是，高质量的修复视频中定位精度不高
	RGB 帧和误差级分析帧+卷积 LSTM [93]	在未压缩状态具有 90% 以上的检测精度，但是没有测试在视频压缩状态下的检测效果

现复制帧序列和来源帧序列。但是，复制的帧必须在 15 帧以上，且没有在压缩状态下进行验证。Bakas J 等[71]利用 3D-CNN 检测视频帧间篡改，在该网络中引入了用来提取时域信息，尤其是帧间的异常突变的差分层。该方法在较广的视频质量下获得平均 97% 的精度，但是仅在 H. 264 固定 GOP 和定码率下进行测试。

2）前置预处理层

依据视频伪造方法以及其遗留痕迹等作为先验知识，在 CNN 前设计预处理层，过滤视频内容，突出奇异信号与视频原有信号间的差异，送入 CNN 模型进行篡改判别，有代表性的网络结构有自编码器、胶囊网络和 MesoNet。D'Avino, D. 等[72]首先设计高通滤波器计算残差，在量化残差后使用共生矩阵直方图提取特征，再将特征送入具有递归神经网络的自编码器中，这样篡改对象不能很好的拟合学习的模型而具有较大的编码误差，从而被判断为异常对象，实现蓝屏抠像对象的检测；Zampoglou M 等[73]结合两个用于手动验证的离散余弦变化和视频质量化误差作为前置预处理层，再使用用于图像分类的深度卷积神经网络识别社交媒体视频的真实性。但是，此类方法的预处理层并没有考虑时域信息，且采用逐帧操作，容易受到视频编码的影响。

3）特定视频取证任务的 CNN 网络结构

这类取证方法主要结合取证的实际问题，设计特定视频取证任务的 CNN 网络结构，主要针对视频编码器，重压缩，相机模型，复制-粘贴和帧率上转换等。这类方法考虑时域方向内容的连续性，普遍采用 RNN，GRU，LSTM 和 ConvLSTM 等时序网络，具有较好的检测性能。代表作有：Verde, S 等[74]训练

一个12层的网络提取视频编码器信息,另一个7层的网络提取视频质量信息,结合这两个网络,实现具有不同编码策略或编码参数的视频拼接并实现拼接点定位。该方法在CIF、4CIF、PAL和720p四种分辨率下的视频,对于四类常用编码器(MPEG2、MPEG4、H.264和H.265具有满意的检测结果。但是,在较小篡改区域的检测精度有待提高。何沛松等[75]提出双压H.264视频的重定位I帧的逐帧检测模型。在这个模型中,使用三个连续的视频帧作为短视频序列作为输入样本,利用高频组件作为预处理层消除多样化的视频内容影响,平均池化和全局平均池化在网络中被考虑用来阻止过拟合,1×1卷积核在较深层中提起更多的内在模式。在公开获得的YUV序列的不同编码参数下,该模型获得满意的重定位I帧检测。但是针对仅第一帧为I帧其他为P或B帧编码的情况下无法进行检测。此外,作者还根据H.264和HEVC的帧内编码景深设计专用的网络,并结合残差的直方图构成混合取证特征,能有效的实现双压和比特率转码的检测[76]。Bondi L等[77]利用基于相机模型的CNN特征聚族实现篡改检测和定位。姚烨等提出基于空间富裕模型和C3D的时空三路网络模型实现时域和空域篡改区域的定位[78]。对于时域检测器,三路3D卷积神经网络作为编码器,双向长短期记忆模型作为解码器;对于空域定位器,基于ResNet12的C3D三路网络被设计作为编码器,区域候选网络(region proposal networks,RPNs)被用作解码器;损失函数采用focal和GIoU的联合优化。实验结果表明,该空时检测和定位算法具有99%以上的帧分类和96%的区域定位。但是计算复杂度较高,需要大量训练数据和长的训练周期。Minseok Yoon[79]、Ding[80]等研究人员从遮挡图、连续视频帧的堆叠角度设计端到端的卷积神经网络,实现视频帧率上转换的检测和插入帧的定位,但是不能实现帧率下转的检测。

4)针对深度伪造视频的被动取证技术

一些专门针对基于GAN的人脸视频合成的取证算法。这些方法主要利用眼、鼻和口等面部属性或面部表情的自然变化规律的异常痕迹,通过眨眼、眼球对称和心率变化等实现虚假人脸图像/视频的检测,获得了较好的效果[81-83]。在2020年由Facebook联合微软主办的国际“DeepFake检测竞赛”(DeepFake Detection Challenge,DFDC)中,白俄罗斯的Selim Seferbekov等人和中国科学技术大学俞能海、张卫明科研团队分别取得第一名和第二名的成绩。前者采用经典的深度神经网络二分类训练流程,对多种不同生成方法下的换脸视频进行检测。该方法采用新型复合扩展网络作为深度特征提取器,引入了多种不同层次的数据增广技术类型,并对训练数据集进行了巨大的增广扩充,最后的预测结果采用了启发式的概率值预测算法进行判别。后者采用弱监督数据增广网络实现模型训练与数据增广同步,从多个角度突出伪造人脸图像不同区域以及弱化非核心区域。该方法能够在一定程度上实现对多种换脸视频的良好检测能力。此外,Liu J等和Xu Zh等人分别利用轻量级3D卷积神经网络[82]和多级卷积神经网络[83]实现Deepfake视频的检测;Gu Y等人和Li G等人在Deepfake视频中分别发现声音-视频的不一致性[84]和合成人脸区域对称性的不一致性[85]揭露Deepfake伪造。针对Face2Face生成的虚假人脸视频,Afchar D和Amerini I等分别提出了紧凑的人脸视频伪造检测网络MesoNet和基于光流不连续的Flow-CNN,分别取得了95%和81.61%的检测率[86-87]。针对Deepfake合成的虚假人脸视频,Li Y等利用合成的人脸视频遗留的变形效应(warping artifacts)作为取证线索逐帧识别真假人脸[86-87]。这类取证方法具有很强的领域特性,如果将此类方法直接逐帧应用或迁移到其他深度视频伪造取证,由于没有充分利用时域信息或修复视频中并无人脸的领域特性,导致检测率并不高。关于这类篡改更丰富的取证方法,读者可以阅读梁瑞刚[86]和Mirsky Y[87]的综述论文。

5)针对深度视频修复伪造的被动取证技术

为了克服传统视频对象修复方法的缺陷,近几年涌现了一些深度视频对象修复方法[88-89],它们不仅能推理视频帧结构和时域运动结构,获得更多的细节信息,还能创建不可见对象,使得修复的视频区域与未篡改区域在空时域保持较高的感知一致性,从而弱化了传统视频对象修复篡改遗留的取证线索。正因如此,前述的视频对象修复取证方法对于深度视频对象修复的检测存在严重的性能退化。到目前为止,仅有深圳大学黄继武教授团队[90]、澳门大学周建涛[91]、湖南科技大学丁湘陵[92]、马里兰

大学的 Zhou Peng[93] 针对深度修复篡改进行了一些有益的尝试。Li H[90] 利用深度图像修复与原始图像间的高频信息的差异，提出高通预处理模块、4 个级联的 Resnet 块和 2 个上采样块的网络模型，实现像素级的深度图像修复定位。Wu H 等[91] 利用神经架构搜索和注意力机制实现深度图像修复的篡改定位。该网络由三个子网络组成：通过特殊的卷积模块以提取图片中隐藏的篡改痕迹的增强子网络；通过神经架构搜索设计具有更好泛化性的网络框架的增强子网络；通过新提出的全局、局部感知模块来更好地帮助整体网络进行决策的决策子网络。实验结果不仅表明本文提出的网络可更加精确地对图像修复区域进行定位，并且还说明不同的图像修复方法都会遗留下部分相似痕迹。前者[90-91] 为深度视频修复被动取证的开展提供了思路，但是不能直接推广到深度视频对象修复篡改的取证，原因在于：没有充分利用视频特有的时域信息，也没有考虑训练样本与测试样本之间的失配问题，可能导致检测率不高。因此，Ding 等在文献[90]的基础上进行扩展，引入了空时高通过滤层，以及针对定位轮廓不准确，增加精化子网络，实现深度视频修复篡改的定位检测[92]，但是，高质量的修复视频中定位精度不高。Zhou P 等[93] 利用 RGB 帧和误差级分析帧作为输入，结合卷积 LSTM 预测修复区域，但是，该方法没有测试在视频压缩状态下的检测效果。

3.4　基于原始视频特征表征的单类检测方法

过于依赖已知伪造方法或伪造视频遗留痕迹的先验知识是当前视频内容伪造被动取证的局限，因为在很多实际环境下，依据待测视频很难估计所采用的篡改手段或分析其遗留的痕迹，尤其是，在针对新出现以及未来的视频伪造手段时。针对这个问题，有学者提出基于原始视频特征表征的单类检测方法，部分解决了这类情况下的检测，这类方法从原始视频或特定人物自然视频中提取语义或统计特征，由于篡改视频与自然视频存在本质的差异，从而在检测中，通过比对待测视频与自然视频间的特征差异进行真假判决。代表性的工作有：Ding 等[50] 在对采用运动补偿帧插值技术伪造的高帧率视频和原始视频分别执行视频帧间运动补偿帧差，提取帧差残差信号，计算它们的低阶统计特征：方差、倾斜度和峰度等，发现存在本质的统计差异，使用一类分类器，通过对原始视频训练其帧差残差信号的低阶统计特征，获得决策超平面，能有效检测出新的视频帧修复方法合成的修复帧。针对在较少 deepfake 视频，现有检测器性能退化的问题，Khalid H 等先使用一类变分自编码器训练真实人脸图像，再用训练的模型检测类似 Deepfake 的非真实图像，将其判定为异常者。实验表明该方法在没有任何虚假图像的情况下，能获得 97.5% 的检测精度[94]。Li Q 等[95] 基于单高斯分布模型，采用一类分类器和仅目标类视频实现双压检测。作者首先处理解压视频帧，提取 SPAM（subtractive pixel adjacency matrix）特征检测遗留在双压过程中的痕迹。随后，因为原始视频帧提取的 SPAM 特征近似高斯分布，所以采用基于高斯密度的一类分类器。因为视频内容的多样性，采用 ensemble 策略和投票机制提升检测的健壮性。实验结果表明，该方法能在仅有原始视频序列的情况下，有效检测双压视频，同时也优于完全监督学习方法。这类方法存在的主要问题在于原始视频的选择，如果原始视频和测试视频存在较大的运动模式和内容差异度，将导致误判率增加。

4　存在的挑战和未来的研究展望

视频具有类型多样、结构和内容复杂、对象运动非线性等特点。当前，主流的视频被动取证主要是针对传统视频伪造手段，包括基于扩散/样本合成的视频修复、基于 MCFI/FRUC 的视频帧修复和视频帧插入/删除等。近几年，视频篡改伪造更多地采用了深度学习的手段，摆脱了传统视频篡改伪造手段对于领域相关知识的依赖，而是通过数据驱动的方式，取得了前所未有的快速发展。尤其是 2019 年以来，深度视频修复、深度视频帧率上转换、深度视频风格迁移以及深度视频抠像等深度视频伪造手段不断出现，它们取得了“以假乱真”的视觉效果，而且用相应的伪造工具软件，例如“ZAO”、“DeepNude”和“DAIN”等，支持一键生成，显著降低了篡改伪造的难度。恶意篡改者将借助先进的深度视频篡改技术，以改变视频的原始语义或属性，并且更好地掩盖操作痕迹。因此，在人工智能背景下，视频被动取证衍生了一些新的困难和挑战。针对深度视频伪造，研究切实有效的检测和防御手

段,是视频被动取证领域的未来研究方向。我们认为,以下3个问题值得深入的研究:

1)深度视频篡改的深度取证技术

目前,国内外针对深度视频篡改伪造的被动取证相对较少,已有的研究工作主要集中在深度人脸视频合成,即DeepFake的取证。深度视频篡改涉及到各种深度网络模型,它们弱化或完全克服了传统视频篡改手段遗留的痕迹,从而遗留的痕迹更加细微。以深度视频对象修复为例,通过推理视频帧结构和时域运动结构,能够获得更多的细节信息,还能创建不可见对象,使得视频修复区域与未篡改区域在空时域更好地保持感知一致性。为此,需要研究各类深度视频修复的工作原理,即采用的深度网络模型。虽然深度视频修复网络模型能够很好地推理修复视频,但是其本身是一个多层级联卷积过滤的复杂处理系统,必然会遗留源生成器属性(source generator attribution)。深度视频修复的被动取证可以采用白盒和黑盒测试手段,从深度网络模型技术原理的角度推演视频合成过程可能产生的异常现象,尤其是在合成对象或移除对象过程中在篡改区域及其与背景区域间的过渡区域周围,挖掘细微的篡改痕迹。此外,由于这些新的深度视频篡改网络模型不断出现,使得视频被动取证很可能面临开放集场景的挑战。

2)研究健壮、开放、可解释的数字视频伪造取证技术

视频数据量大,通常编码后存储和传输。视频被动取证需要重点考虑其对于视频压缩的鲁棒性。同时,视频可以结合语音,成为音视频,即具有语音和视觉的多模态结合体。为此,可以研究多模态的特征融合机制。从音频的角度提取/学习语速、频率分布和音素等特征,从视频的角度提取/学习像素域、压缩域和光流域等特点,并且结合深度融合机制,以提升取证准确率,保证取证的健壮性。还可以考虑引入单分类、元学习、小样本和零样本学习等机器学习的新技术,从已有的大量原始视频数据得到辨别性强的特征,摆脱对于训练/测试数据集的过度依赖,实现开放集条件下的视频被动取证;此外,将深度模型的可解释性技术引入到基于深度学习的视频被动取证,将篡改痕迹和可视特征引入取证分析中,在评价模型的脆弱性的同时促进现有取证技术的性能改进。

3)构建大规模深度篡改视频数据集

训练深度网络模型的视频数据集对于提升深度网络的取证性能至关重要。现有的视频篡改数据集通常是利用公开的原始视频集合,例如YouTube-8M,DAVIS,Xiph.org Video Test Media等,采用深度视频篡改技术,自行构建的视频数据集。此外,还有一些公开的小型视频数据集,包括SULFA和REWIND等。这些视频数据集的规模过小,不能满足基于深度学习的深度视频篡改的取证需求。最新的较大规模的视频数据集,包括FaceForensics++和DeeperForensics-1.0等,都是结合最新的深度人脸视频篡改技术构建的,不能迁移到其他深度视频伪造手段的模型训练中,因此,构建大规模深度篡改视频数据集迫在眉睫。

5 结论

数字视频被动取证研究是在不需要预先嵌入任何先验信息的情况下,直接依据已获得的视频数据本身来鉴别其真实性和来源。本文分析了视频伪造行为的特点、传统视频篡改和深度视频篡改遗留的痕迹以及对视频取证的影响,从伪造行为、篡改痕迹和取证特征3个角度归纳了视频被动取证领域的研究进展,并随着人工智能领域的成熟,对未来针对深度视频伪造篡改取证的发展趋势进行了探讨。可以预计,在新一轮人工智能浪潮中,基于深度学习的视频被动取证和针对深度视频伪造行为的被动取证将是数字视频被动取证的研究热点,也仍然是数字媒体内容安全领域的研究前沿,在相关理论,尤其是取证结果可解释性领域的完善、被动取证算法健壮性的提高以及大规模视频伪造数据集和相关评价标准的建立等方面展开深入研究,可以获得更多引领前沿的研究成果。

参考文献

[1] 杨锐, 骆伟祺, 黄继武. 多媒体取证[J]. 中国科学:信息科学, 2013, 43(12): 1654-1672.
YANG Rui, LUO Weiqi, HUANG Jiwu. Multimedia forensics[J]. Scientia Sinica(Informationis), 2013, 43(12): 1654-1672. (in Chinese)

[2] 陈威兵, 杨高波, 陈日超, 等. 数字视频真实性和来源

的被动取证[J]. 通信学报, 2011, 32(6): 177-183.
CHEN Weibin, Yang Gaobo, CHEN Richao, et al. Digital video passive forensics for its authenticity and source [J]. Journal on Communications, 2011, 32(6): 177-183. (in Chinese)

[3] SINGH R D, AGGARWAL N. Video content authentication techniques: a comprehensive survey[J]. Multimedia Systems, 2018, 24(2): 211-240.

[4] SITARA K, MEHTRE B M. Digital video tampering detection: an overview of passive techniques[J]. Digital Investigation, 2016, 18: 8-22.

[5] KAUR H, JINDAL N. Image and video forensics: a critical survey[J]. Wireless Personal Communications, 2020: 1-22.

[6] JUNG D J, HYUN D K, RYU S J, et al. Detecting recaptured videos using shot-based photo response non-uniformity[C]// International Workshop on Digital Watermarking (IWDW). Springer, Berlin, Heidelberg, 2011: 281-291.

[7] 崔三帅, 毛毛雨, 林晓丹, 等. 图像视频中 ENF 信号的分析及应用综述[J]. 应用科学学报, 2019, 37(5): 573-589.
CUI Sanshuai, MAO Maoyu, LIN Xiaodan, et al. Survey on analysis and applications of electric network frequency (ENF) signals in image and video[J]. Journal of Applied Sciences, 2019,37(5): 573-589. (in Chinese)

[8] IULIANI M, SHULLANI D, FONTANI M, et al. A video forensic framework for the unsupervised analysis of MP4-like file container[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(3): 635-645.

[9] SU Yuting, ZHANG Jing, JI Zhong. A source video identification algorithm based on features in video stream[C]// 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing. IEEE, 2008, 1: 719-723.

[10] BESTAGINI P, MILANI S, TAGLIASACCHI M, et al. Codec and GOP identification in double compressed videos[J]. IEEE Transactions on Image Processing, 2016, 25(5): 2298-2310.

[11] TAGLIASACCHI M, TUBARO S. Blind estimation of the QP parameter in H.264/AVC decoded video[C]// 11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS). IEEE, 2010: 1-4.

[12] VALENZISE G, TAGLIASACCHI M, TUBARO S. Estimating QP and motion vectors in H.264/AVC video from decoded pixels[C]// Proceedings of the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence (MFSI), 2010: 89-92.

[13] REIBMAN A R, VAISHAMPAYAN V A, SERMADEVI Y. Quality monitoring of video over a packet network[J]. IEEE Transactions on Multimedia, 2004, 6(2): 327-334.

[14] REIBMAN A R, POOLE D. Characterizing packet-loss impairments in compressed video[C]//2007 IEEE International Conference on Image Processing (ICIP). IEEE, 2007, 5: V-77-V-80.

[15] WANG Weihong, FARID H. Exposing digital forgeries in interlaced and deinterlaced video[J]. IEEE Transactions on Information Forensics and Security, 2007, 2(3): 438-449.

[16] KOBAYASHI M, OKABE T, SATO Y. Detecting forgery from static-scene video based on inconsistency in noise level functions[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(4): 883-892.

[17] 谢丰, 蒋兴浩, 孙锬锋. 视频双压缩检测技术综述[J]. 通信技术, 2017,50(3): 393-399.
XIE Feng, JIANG Xinghao, SUN Tanfeng. Overview of video double compression detection[J]. Communications Technology, 2017,50(3): 393-399. (in Chinese)

[18] WANG Weihong, FARID H. Exposing digital forgeries in video by detecting double MPEG compression[C]// Proceedings of the 8th workshop on Multimedia and security, 2006: 37-47.

[19] CHEN Wen, SHI Yunqing. Detection of double MPEG compression based on first digit statistics[C]// International Workshop on Digital Watermarking (IWDW). Springer, Berlin, Heidelberg, 2008: 16-30.

[20] CHEN Chunhua, SHI Yunqing, SU Wei. A machine learning based scheme for double JPEG compression detection[C]//2008 19th International Conference on Pattern Recognition (ICPR). IEEE, 2008: 1-4.

[21] JIANG Xinghao, WANG Wan, SUN Tanfeng, et al. Detection of double compression in MPEG-4 videos based on Markov statistics[J]. IEEE Signal Processing Letters, 2013, 20(5): 447-450.

[22] LIAO Dandan, YANG Rui, LIU Hongmei, et al. Double H.264/AVC compression detection using quantized nonzero AC coefficients[C]// Media Watermarking, Security, and Forensics III. International Society for Optics and Photonics, 2011, 7880: 78800Q.

[23] JIANG Xinghao, HE Peisong, SUN Tanfeng, et al. De-

tection of double compression with the same coding parameters based on quality degradation mechanism analysis [J]. IEEE Transactions on Information Forensics and Security, 2017, 13(1): 170-185.

[24] JIANG Xinghao, XU Qiang, SUN Tanfeng, et al. Detection of HEVC double compression with the same coding parameters based on analysis of intra coding quality degradation process[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 250-263.

[25] XU Qiang, JIANG Xinghao, SUN Tanfeng, et al. Detection of transcoded HEVC videos based on in-loop filtering and PU partitioning analyses[J]. Signal Processing: Image Communication, 2021, 92: 116109.

[26] FENG Chunhui, XU Zhengquan, JIA Shan, et al. Motion-adaptive frame deletion detection for digital video forensics[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(12): 2543-2554.

[27] SU Yuting, ZHANG Jing, LIU Jie. Exposing digital video forgery by detecting motion-compensated edge artifact[C]// 2009 International Conference on Computational Intelligence and Software Engineering (ICCISE). IEEE, 2009: 1-4.

[28] DONG Qiong, YANG Gaobo, ZHU Ningbo. A MCEA based passive forensics scheme for detecting frame-based video tampering[J]. Digital Investigation, 2012, 9(2): 151-159.

[29] STAMM M C, LIN W S, LIU K J R. Temporal forensics and anti-forensics for motion compensated video [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(4): 1315-1329.

[30] CHAO Juan, JIANG Xinghao, SUN Tanfeng. A novel video inter-frame forgery model detection scheme based on optical flow consistency[C]// International Workshop on Digital Watermarking (IWDW). Springer, Berlin, Heidelberg, 2012: 267-281.

[31] WANG Wan, JIANG Xinghao, WANG Shilin, et al. Identifying video forgery process using optical flow[C]//International Workshop on Digital Watermarking (IWDW). Springer, Berlin, Heidelberg, 2013: 244-257.

[32] WU Yuxing, JIANG Xinghao, SUN Tanfeng, et al. Exposing video inter-frame forgery based on velocity field consistency[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014: 2674-2678.

[33] JIA Shan, XU Zhengquan, WANG Hao, et al. Coarse-to-fine copy-move forgery detection for video forensics[J]. IEEE Access, 2018, 6: 25323-25335.

[34] ZHENG Lu, SUN Tanfeng, SHI Yunqing. Inter-frame video forgery detection based on block-wise brightness variance descriptor[C]// International Workshop on Digital Watermarking (IWDW). Springer, Cham, 2014: 18-30.

[35] LIN Guoshiang, CHANG Jiefan, CHUANG Chenghung. Detecting frame duplication based on spatial and temporal analyses[C]// 2011 6th International Conference on Computer Science & Education (ICCSE). IEEE, 2011: 1396-1399.

[36] WANG Qi, LI Zhaohong, ZHANG Zhenzhen, et al. Video inter-frame forgery identification based on consistency of correlation coefficients of gray values[J]. Journal of Computer and Communications, 2014, 2(4): 51.

[37] LIU Yuqing, HUANG Tianqiang. Exposing video inter-frame forgery by zernike opponent chromaticity moments and coarseness analysis[J]. Multimedia Systems, 2017, 23(2): 223-238.

[38] ZHANG Zhenzhen, HOU Jianjun, MA Qinglong, et al. Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames[J]. Security and Communication Networks, 2015, 8(2): 311-320.

[39] YAO Haichao, NI Rongrong, ZHAO Yao. An approach to detect video frame deletion under anti-forensics[J]. Journal of Real-Time Image Processing, 2019, 16(3): 751-764.

[40] 何天琦, 蒋兴浩, 孙锬锋. 视频帧率上转换检测技术综述[J]. 网络与信息安全学报, 2018, 4(10): 1-11.
HE Tianqi, JIANG Xinghao, SUN Tanfeng. Review of video frame rate up conversion detection[J]. Chinese Journal of Network and Information Security, 2018, 4(10): 1-11. (in Chinese)

[41] BIAN Shan, LUO Weiqi, HUANG Jiwu. Detecting video frame-rate up-conversion based on periodic properties of inter-frame similarity[J]. Multimedia Tools and Applications, 2014, 72(1): 437-451.

[42] 林晶, 黄添强, 李小琛, 等. 基于光流周期特性的视频帧率上转篡改检测[J]. 计算机系统应用, 2017, 26(6): 131-136.
LIN Jing, HUANG Tianqiang, LI Xiaochen, et al. Detection of video frame-rate up-conversion using periodic properties of optical flow[J]. Computer Systems & Applications, 2017, 26(6): 131-136. (in Chinese)

[43] BESTAGINI P, BATTAGLIA S, MILANI S, et al. Detection of temporal interpolation in video sequences[C]// 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2013: 3033-3037.

[44] JUNG D J, LEE H K. Frame-rate conversion detection based on periodicity of motion artifact[J]. Multimedia Tools and Applications, 2018, 77(5): 6095-6116.

[45] LI Ran, LIU Zhenghui, ZHANG Yu, et al. Noise-level estimation based detection of motion-compensated frame interpolation in video sequences[J]. Multimedia Tools and Applications, 2018, 77(1): 663-688.

[46] 李然, 梅腊腊, 邬长安, 等. 针对视频运动补偿帧率提升篡改的主动混噪取证算法[J]. 电子与信息学报, 2018, 40(3): 713-720.
LI Ran, MEI Lala, WU Changan, et al. Active noised-mixed forensics algorithm for tampering of video motion-compensated frame rate up-conversion[J]. Journal of Electronics & Information Technology, 2018, 40(3): 713-720. (in Chinese)

[47] YAO Yuxuan, YANG Gaobo, SUN Xinming, et al. Detecting video frame-rate up-conversion based on periodic properties of edge-intensity[J]. Journal of Information Security and Applications, 2016, 26: 39-50.

[48] XIA Min, YANG Gaobo, LI Leida, et al. Detecting video frame rate up-conversion based on frame-level analysis of average texture variation[J]. Multimedia Tools and Applications, 2017, 76(6): 8399-8421.

[49] DING Xiangling, ZHU Ningbo, LI Leida, et al. Robust localization of interpolated frames by motion-compensated frame interpolation based on an artifact indicated map and tchebichef moments[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(7): 1893-1906.

[50] DING Xiangling, LI Yue, XIA Min, et al. Detection of motion compensated frame interpolation via motion-aligned temporal difference[J]. Multimedia Tools and Applications, 2019, 78(6): 7453-7477.

[51] DING Xiangling, YANG Gaobo, LI Ran, et al. Identification of motion-compensated frame rate up-conversion based on residual signals[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(7): 1497-1512.

[52] WANG Weihong, Farid H. Exposing digital forgeries in video by detecting duplication[C]// Proceedings of the 9th Workshop on Multimedia & Security, 2007: 35-42.

[53] HSU C C, HUNG T Y, LIN C W, et al. Video forgery detection using correlation of noise residue[C]// 2008 IEEE 10th Workshop on Multimedia Signal Processing. IEEE, 2008: 170-174.

[54] D'AMIANO L, COZZOLINO D, POGGI G, et al. A patchmatch-based dense-field algorithm for video copy-move detection and localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(3): 669-682.

[55] SU Lichao, LI Cuihua, LAI Yuecong, et al. A fast forgery detection algorithm based on exponential-Fourier moments for video region duplication[J]. IEEE Transactions on Multimedia, 2017, 20(4): 825-840.

[56] CHEN Shengda, TAN Shunquan, LI Bin, et al. Automatic detection of object-based forgery in advanced video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 26(11): 2138-2151.

[57] ZHANG Jing, SU Yuting, ZHANG Mingyu. Exposing digital video forgery by ghost shadow artifact[C]// Proceedings of the First ACM Workshop on Multimedia in Forensics, 2009: 49-54.

[58] 刘雨青, 黄添强. 基于时空域能量可疑度的视频篡改检测与篡改区域定位[J]. 南京大学学报:自然科学版, 2014, 50(1):61-71.
LIU Yuqing, HUANG Tianqiang. Digital video forgeries detection and tamper areas location based on temporal and spatial energy suspicious degree[J]. Journal of Nanjing University(Natural Science), 2014, 50(1):61-71. (in Chinese)

[59] 王嘉萱, 苏育挺. 基于时序特征聚类的对象删除篡改检测[J]. 电子测量技术,2012, 35(11):49-52.
WANG Jiaxuan, SU Yuting. Object deleted video detection based on temporal characteristics clustering[J]. Electronic Measurement Technology, 2012, 35(11):49-52. (in Chinese)

[60] CHEN Richao,YANG Gaobo, ZHU Ningbo. Detection of object-based manipulation by the statistical features of object contour[J]. Forensic Science International, 2014, 236: 164-169.

[61] BIDOKHTI A, GHAEMMAGHAMI S. Detection of regional copy/move forgery in MPEG videos using optical flow[C]//2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP). IEEE, 2015: 13-17.

[62] LIN C S, TSAY J J. A passive approach for effective detection and localization of region-level video forgery with

spatio-temporal coherence analysis[J]. Digital Investigation, 2014, 11(2): 120-140.

[63] SAXENA S, SUBRAMANYAM A V, RAVI H. Video inpainting detection and localization using inconsistencies in optical flow[C]//2016 IEEE Region 10 Conference (TENCON). IEEE, 2016: 1361-1365.

[64] BAGIWA M A, WAHAB A W A, IDRIS M Y I, et al. Digital video inpainting detection using correlation of hessian matrix[J]. Malaysian Journal of Computer Science, 2016, 29(3): 179-195.

[65] BAI Shanshan, YAO Haichao, NI Rongrong, et al. Detection and localization of video object removal by spatio-temporal lbp coherence analysis[C]//International Conference on Image and Graphics (ICIG). Springer, Cham, 2019: 244-254.

[66] ALORAINI M, SHARIFZADEH M, SCHONFELD D. Sequential and patch analyses for object removal video forgery detection and localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(3): 917-930.

[67] HYUN D K, RYU S J, LEE H Y, et al. Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise[J]. Sensors, 2013, 13(9): 12605-12631.

[68] SINGH R D, AGGARWAL N. Detection of upscale-crop and splicing for digital video authentication[J]. Digital Investigation, 2017, 21: 31-52.

[69] LONG Chengjiang, SMITH E, BASHARAT A, et al. A c3d-based convolutional neural network for frame dropping detection in a single video shot[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017: 1898-1906.

[70] LONG Chengjiang, BASHARAT A, HOOGS A, et al. A coarse-to-fine deep convolutional neural network framework for frame duplication detection and localization in forged videos[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2019: 1-10.

[71] BAKAS J, NASKAR R. A digital forensic technique for inter-frame video forgery detection based on 3D CNN[C]// International Conference on Information Systems Security (ICISS). Springer, Cham, 2018: 304-317.

[72] D'AVINO D, COZZOLINO D, POGGI G, et al. Autoencoder with recurrent neural networks for video forgery detection[J]. Electronic Imaging, 2017, 2017(7): 92-99.

[73] ZAMPOGLOU M, MARKATOPOULOU F, MERCIER G, et al. Detecting tampered videos with multimedia forensics and deep learning[C]//International Conference on Multimedia Modeling (ICMM). Springer, Cham, 2019: 374-386.

[74] VERDE S, BONDI L, BESTAGINI P, et al. Video codec forensics based on convolutional neural networks[C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 530-534.

[75] HE Peisong, JIANG Xinghao, SUN Tanfeng, et al. Frame-wise detection of relocated I-frames in double compressed H.264 videos based on convolutional neural network[J]. Journal of Visual Communication and Image Representation, 2017, 48: 149-158.

[76] HE Peisong, LI Haoliang, WANG Hongxia, et al. Frame-wise detection of double HEVC compression by learning deep spatio-temporal representations in compression domain[J]. IEEE Transactions on Multimedia, 2020.

[77] BONDI L, LAMERI S, GUERA D, et al. Tampering detection and localization through clustering of camera-based CNN features[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017: 1855-1864.

[78] YANG Quanxin, YU Dongjin, ZHANG Zhuxi, et al. Spatiotemporal trident networks: detection and localization of object removal tampering in video passive forensics[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020.

[79] YOON M, NAM S H, YU I J, et al. Frame-rate up-conversion detection based on convolutional neural network for learning spatiotemporal features[J]. arXiv preprint arXiv:2103.13674, 2021.

[80] DING Xiangling, HUANG Yanming. Identification of frame-rate up-conversion based on spatial-temporal edge and occlusion with convolutional neural network[C]// 2020 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2020: 1-5.

[81] TOLOSANA R, VERA-RODRIGUEZ R, FIERREZ J, et al. Deepfakes and beyond: a survey of face manipulation and fake detection[J]. Information Fusion, 2020, 64: 131-148.

[82] LIU Jiarui, ZHU Kaiman, LU Wei, et al. A lightweight 3D convolutional neural network for deepfake detection[J]. International Journal of Intelligent Systems, 2021: 1-15.

[83] XU Zhaopeng, LIU Jiarui, LU Wei, et al. Detecting facial manipulated videos based on set convolutional neural networks[J]. Journal of Visual Communication and Image Representation, 2021, 77: 103119.

[84] GU Yewei, ZHAO Xianfeng, GONG Chen, et al. Deepfake video detection using audio-visual consistency[C]//International Workshop on Digital Watermarking (IWDW). Springer, Cham, 2020: 168-180.

[85] LI Gen, CAO Yun, ZHAO Xianfeng. Exploiting facial symmetry to expose deepfakes[C]//2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021, to appear.

[86] 梁瑞刚, 吕培卓, 赵月, 等. 视听觉深度伪造检测技术研究综述[J]. 信息安全学报, 2020, 5(2): 1-17.
LIANG Ruigang, LV Peizhuo, ZHAO Yue, et al. A survey of audiovisual deepfake detection techniques[J]. Journal of Cyber Security, 2020, 5(2): 1-17. (in Chinese)

[87] MIRSKY Y, LEE W. The creation and detection of deepfakes: a survey[J]. ACM Computing Surveys (CSUR), 2021, 54(1): 1-41.

[88] KIM D, WOO S, LEE J Y, et al. Deep video inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 5792-5801.

[89] CHANG Yaliang, ZHE Yuliu, HSU Winston. Vornet: spatio-temporally consistent video inpainting for object removal[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019: 0-0.

[90] LI Haoliang, HUANG Jiwu. Localization of deep inpainting using high-pass fully convolutional network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 8301-8310.

[91] WU Haiwei, ZHOU Jiantao. IID-Net: Image inpainting detection network via neural architecture search and attention[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021.

[92] DING Xiangling, PAN Yifeng, LUO Kui, et al. Localization of deep video inpainting based on spatiotemporal convolution and refinement network[C]//2021 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2021: 1-5.

[93] ZHOU Peng, YU Ning, WU Zuxuan, et al. Deep video inpainting detection[J]. arXiv preprint arXiv: 2101.11080, 2021.

[94] KHALID H, WOO S S. OC-FakeDect: classifying deepfakes using one-class variational autoencoder[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020: 656-657.

[95] LI Qiushi, CHEN Shengda, TAN Shunquan, et al. One-class double compression detection of advanced videos based on simple gaussian distribution model[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021.

作者简介

丁湘陵　男, 1981 年生, 湖南株洲人。湖南科技大学副教授, 博士, 硕士生导师, 主要研究方向为多媒体内容安全、图像/视频处理等。
E-mail: xianglingding@hnust.edu.cn

杨高波　男, 1974 年生, 湖南岳阳人。湖南大学教授, 博士, 博士生导师, 主要研究方向图像/视频信号处理和多媒体通信, 包括图像/视频信息安全、图像/视频篡改取证、压缩域视频水印、高效视频编码及其优化与大数据权属保护技术。
E-mail: yanggaobo@hnu.edu.cn

赵险峰　男, 1969 年生, 安徽宿州人。中国科学院信息工程研究所研究员, 中国科学院大学教授, 博士, 博士生导师, 主要研究方向为多媒体安全与智能分析的理论与技术, 包括多媒体处理与智能分析、多媒体信息隐藏与检测、多媒体伪造取证与防护、多媒体内容智能生成、数字水印与多媒体版权保护等。
E-mail: zhaoxianfeng@iie.ac.cn

谷　庆　男, 1996 年生, 安徽人。湖南科技大学在读硕士研究生, 主要研究方向为多媒体内容安全。
E-mail: 3334143570@qq.com

熊义毛　女, 1996 年生, 湖南人, 湖南科技大学在读硕士研究生, 主要研究方向为多媒体内容安全。
E-mail: 1203661414@qq.com