

基于面部动作不一致的压缩深度伪造视频检测方法

王玉梅¹, 胡娟¹, 武晓帅¹, 肖靖², 陈雁翔³, 廖鑫¹

(1. 湖南大学 信息科学与工程学院, 长沙 410012; 2. 湖南大学 电气与信息工程学院, 长沙 410082;

3. 合肥工业大学 计算机与信息学院, 合肥 230601)

摘要: 为准确高效地识别社交网络中的压缩深度伪造视频, 通过借鉴表情识别领域的知识, 基于对人脸面部动作的分析, 挖掘伪造视频中五官区域不规则的运动, 提出了一种基于面部动作不一致的压缩深度伪造视频检测方法。鉴于人物的面部运动时, 低维的角度特征能直接反应五官区域的运动。从人脸视频中提取精确的面部特征点, 通过对五官区域建模从而构建五官角度特征, 并从运动和速度 2 种模式对角度特征进行分析, 用长短期记忆网络捕获五官区域的不规则运动。实验结果表明, 所提出的检测方法能够有效辨别压缩视频的真假, 且对不同的压缩因子具有较好的鲁棒性。

关键词: 社交网络压缩视频; 压缩深度伪造视频; 深度伪造视频检测; 面部动作不一致性

中图分类号: TN929; TP309.7

文献标志码: A

文章编号: 1673-825X(2025)03-0445-08

Compressed deepfake video detection method based on inconsistent facial motion

WANG Yumei¹, HU Juan¹, WU Xiaoshuai¹, XIAO Jing², CHEN Yanxiang³, LIAO Xin¹

(1. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410012, P. R. China;

2. College of Electrical and Information Engineering, Hunan University, Changsha 410082, P. R. China;

3. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, P. R. China)

Abstract: To accurately and efficiently detect compressed deepfake videos on social networks, this study proposes a detection method based on facial action inconsistency by leveraging knowledge from the field of facial expression recognition. The method analyzes irregular movements in facial regions of deepfake videos by modeling facial actions. Considering that low-dimensional angular features can directly reflect facial region movements during human expressions, precise facial landmarks are extracted from face videos, and angular features of key facial components are constructed through facial region modeling. These angular features are then analyzed from both motion and velocity perspectives. A long short-term memory (LSTM) network is employed to capture irregular motions in facial regions. Experimental results demonstrate that the proposed method effectively distinguishes real and fake compressed videos and exhibits strong robustness across varying compression factors.

收稿日期: 2024-03-01 修订日期: 2025-03-17 通讯作者: 廖鑫 xinliao@hnu.edu.cn

基金项目: 国家自然科学基金项目(U22A2030, 61972142); 国家重点研发计划(2024YFF0618800, 2022YFB3103500); 湖南省自然科学基金杰出青年基金项目(2024JJ2025); 长沙市科技重大专项经费资助(kh2205033)

Foundation Items: National Natural Science Foundation of China (U22A2030, 61972142); National Key R&D Program of China (2024YFF0618800, 2022YFB3103500); Hunan Provincial Funds for Distinguished Young Scholars (2024JJ2025); Major Scientific and Technological Projects of Changsha (kh2205033)

Keywords: compressed videos in social networks; compressed deepfake videos; deepfake video detection; inconsistent facial motion

0 引言

Deepfake^[1]是人工智能技术进步的产物,源于“deep learning”和“fake”的结合。它将目标人物的脸叠加在原始人脸的相应位置上,形成一个包含目标人物的伪造视频^[2]。随着人工智能技术的发展,Deepfake 技术迅速改进,大多数算法都在网站上开源。这些技术有很多积极的应用场景。例如,Deepfake 技术已经应用于电影、教育、医疗等诸多行业^[1]。然而,如果这些技术被滥用,将严重威胁到国家、社会 and 个人的信誉和安全。因为 Deepfake 技术合成的假视频肉眼难以分辨,在互联网上传播可能会造成公众信任危机或引发战争迷雾^[3]。此外,虚假政治信息的传播和扩散可能导致政治抗议,影响国家和个人的安全。

为了应对 Deepfake 技术的潜在威胁,减轻虚假信息传播带来的负面影响,大量学者已经研发了多种 Deepfake 检测方法^[4-8]。这些方法在检测未压缩视频方面取得优异的性能,大致可以将其分为 2 类:基于手工和基于深度学习的方法。基于手工的方法^[9-15]通过分析伪造过程和挖掘启发式特征来识别伪造视频。随着伪造技术的发展,生成的 Deepfake 视频中没有明显的篡改痕迹,将可能导致这些方法逐渐失效。基于深度学习的方法^[16-33]则依赖于模型的强大学习能力,需要大规模真实和虚假数据来驱动模型以提取可区分的特征。

在社交网络中,由于带宽和存储空间的限制,视频在上传时通常会被压缩。据文献[34]报告,通过 Facebook 和 Youtube 传播后,视频的平均分辨率分别下降了 0.8 和 0.64。同时,为了减少时间和空间冗余,在视频压缩过程中引入了降采样和量化噪声,导致压缩后的视频中出现了阻塞、模糊、振铃和锯齿等压缩伪影^[35],给检测模型提取辨别性特征带来挑战。当视频是通过社交网络传输的压缩视频时,会误导神经网络的学习,导致检测性能较差。文献[31-33]中的方法研究了 Deepfake 视频并考虑了压缩场景。然而,他们试图通过卷积神经网络检测压缩后的 Deepfake 视频,模型的检测性能会被共存的压缩伪影削弱。此外,现有神经网络往往参数量大、特征维度高、模型难以训练、可解释性不足。因此,有效检测压缩的 Deepfake 视频并促进对检测方法

的更深入理解是视频取证中有待解决的关键问题。

本文提出了一种新的基于面部动作不一致性的深度伪造视频检测方法来提高社交网络中深度伪造视频的检测性能。本文通过分析深度伪造视频,发现伪造视频在生成的过程中为了保证视觉上的真实性,会对面部进行扭曲,尤其是五官区域,且在合成的过程中视频是逐帧合成的,未强制执行时间的一致性,这导致伪造视频中存在面部动作不一致性。鉴于角度特征具有一定的鲁棒性,能降低压缩带来的影响,本方法构建角度特征以表征面部动作。该方法仅需从人脸视频中提取面部地标点即可构建角度特征,极大程度提高了模型的识别效率,利于实际场景的部署,更具实用性。

1 压缩视频分析

由于网络带宽、用户存储空间等限制,社交网络中传输的视频多数为压缩视频,本节将分析压缩视频以说明压缩深度伪造人脸视频检测方法的动机。

由于视频在压缩过程中主要利用视频中空间冗余、时间冗余和视觉冗余等信息,视频的文件大小和视觉质量会受到影响,因此,在压缩视频中会产生压缩伪影。本文利用峰值信噪比(peak signal to noise ratio, PSNR)、结构相似性(structural similarity, SSIM)、视觉信息保真度(visual information fidelity, VIF)、相对极边缘相干性(relative polar edge coherence, RECO) 5 个指标来评估压缩视频的质量。PSNR 是评价图像最常用的客观指标之一,其值越大,失真越小。SSIM 是另一个重要的评价指标,主要考虑图像亮度、对比度和结构 3 个关键特征,数值越接近于 1 表明两幅图像之间的相似性越强,图像质量越好。VIF 是一种基于信息论的图像质量指标,其值越显著说明融合后的图像质量越好。此外,本文还使用 RECO 作为衡量视频质量的指标,具体的结果如表 1 所示。未压缩视频和压缩视频的 PSNR 和 SSIM 的值分别为 32.60 dB 和 0.84,这表明压缩引入噪声信息。VIF 和 RECO 的值分别为 0.48 和 0.85,这表明压缩视频存在视觉和边缘噪声。基于上述分析可知,视频压缩后视觉质量受到影响且存在压缩噪声,这意味着压缩深度伪造视频的检测更具挑战性。

表1 FaceForensics++中压缩视频的噪声水平

Tab.1 Noise levels of compressed videos in FaceForensics++

Index	PSNR/dB	SSIM	VIF	RECO
FaceForensics++ LQ	32.60	0.84	0.48	0.85

直观的视觉效果呈现如图1所示,视频伪造的过程中扭曲了人脸区域,使得视频中存在篡改伪影。而视频在压缩的过程会去除视频中的冗余信息,导致视频中也存在压缩伪影。在压缩的伪造视频中,篡改伪影和压缩伪影共存,使得从压缩的伪造视频中提取辨别性特征更具挑战性。为了解决这一问题,本文从几何角度出发检测压缩深度伪造视频。

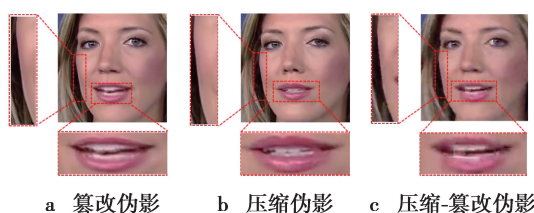


图1 压缩伪影和篡改伪影可视化分析

Fig.1 Visual analysis of compressed artifacts and tampered artifacts

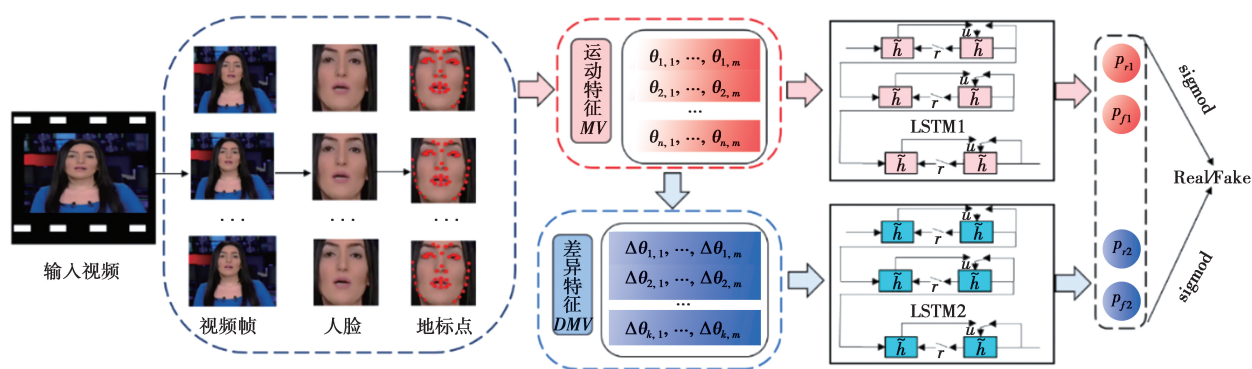


图2 基于面部动作不一致的压缩深度伪造视频检测框架

Fig.2 Compressed Deepfake video detection framework based on the inconsistency of facial movement

2.2 人脸预处理

人脸预处理模块包括视频帧提取、人脸检测、面部地标点提取和面部地标点校正模块。视频帧提取主要是从视频中提取视频帧以便于后续的人脸检测。利用 Dlib 人脸检测器从视频帧中捕获人脸图像,并保存输出的特征点坐标。考虑面部地标点在提取过程中可能会出现抖动的情况,本文引入光流跟踪来检测地标点并引入卡尔曼滤波进一步降噪,以提高地标点的精确性。其中,地标点追踪的过程引入了金字塔型 Lucas-Kanade 操作来预测地标点。

2 基于面部动作不一致的检测框架

2.1 整体框架

该方法主要是捕捉五官区域不规则的运动以检测压缩视频,图2为整体的检测框架。首先,从人脸视频中提取精确的面部特征点。在选取视频片段中,提取连续帧并检测视频中的人脸。检测到人脸后提取人脸特征点坐标,即人脸地标点。其次,对五官区域建模以构建五官角度特征。在面部特征点间,划定眉毛、眼睛、鼻子和嘴巴等4个五官区域。由相关特征点构成的特征向量的夹角来表示各个鲁棒的角度特征。然后,从2种不同模式分析角度特征。在角度特征之上,构建运动特征和差异特征。运动特征由连续视频帧中的所有向量夹角表示,表征视频中面部动作的运动模式。差异特征由相邻视频帧间向量夹角的差值表示,表征视频中面部动作的速度模式。最后,用长短期记忆网络(long short-term memory, LSTM)捕获五官区域不规则的运动。运动特征和差异特征分别输入到结构相同但不共享参数的2个LSTM。利用 Sigmoid 函数处理2个模型的输出并融合得到最终的结果。

在实际应用中,LK 操作会带来一定的噪声,干扰地标点的稳定性。因此,引入了卡尔曼滤波操作来对检测到的信息和实际预测的信息进行降噪处理。

2.3 面部动作特征构建

在表情识别领域,通过提取纹理特征和角度特征来表达不同的表情以识别人物的表情。当出现不同的表情时,角度特征会有明显的区别。借鉴此思想^[36],本文基于深度伪造自身的缺陷,在合成时会对五官区域进行扭曲实现逼真的人脸效果,并将得到的伪造人脸逐帧合成视频,这一过程会使得伪造

视频中人物五官区域存在不规则的运动。基于上述原因,本文对五官区域建模构建几何角度特征以实现深度伪造视频的检测。

针对面部动作几何特征的构建,本文提取眉毛、眼睛、鼻子和嘴巴 4 个区域的角度特征。面部动作特征构造的示意图如图 3 所示。图 3 中带有相同编号向量表示其共同构建一个向量夹角,其中部分夹角由 3 个特征点构成或 4 个特征点构成。

面部运动特征可被表示为

$$MV = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,m} \\ \theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,m} \\ \vdots & \vdots & & \vdots \\ \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,m} \end{bmatrix} \quad (1)$$

式(1)中: MV 用于表示五官区域特征的运动模式; $\theta_{i,j}$ 为特征向量的夹角,表示第 i 帧的第 j 个角度特征。 $n=70$, 为一个视频片段选取的帧数,取值详细分析见 3.4 节。 $m=21$, 为设计的角度特征的总数。基于上述过程,本文构造了五官区域的角度特征以表征视频中面部动作。

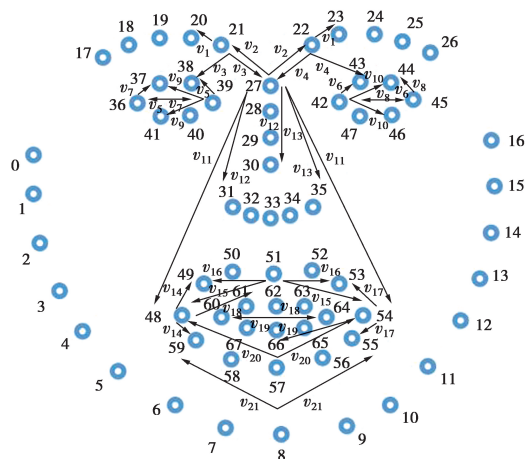


图 3 面部特征点及动作特征向量设计示意图

Fig.3 Design diagram of facial feature points and action feature vectors

除了关注视频中五官的运动,本文还关注五官区域角度特征的速度模式,即相邻帧角度的变化。面部差异特征可被表示为

$$DMV = \begin{bmatrix} \Delta\theta_{1,1} & \Delta\theta_{1,2} & \cdots & \Delta\theta_{1,m} \\ \Delta\theta_{2,1} & \Delta\theta_{2,2} & \cdots & \Delta\theta_{2,m} \\ \vdots & \vdots & & \vdots \\ \Delta\theta_{n-1,1} & \Delta\theta_{n-1,2} & \cdots & \Delta\theta_{n-1,m} \end{bmatrix} \quad (2)$$

式(2)中: DMV 用于表示五官区域特征的速度模式; $\Delta\theta_{i,j} = \theta_{i+1,j} - \theta_{i,j}$, 且 $1 \leq i < n, 1 \leq j \leq m$ 。至此,已完

整的构建五官区域的角度特征并获取了角度特征的速度模式以表征面部动作的规律。构建运动特征和差异特征后,分别单独训练 2 个 LSTM 以从运动和速度 2 个角度捕捉五官区域不规则的运动来检测伪造视频。

2.4 双流融合

2 个模型对输入的视频进行判别后,将分别输出将视频预测为真视频和假视频的概率值。为有效融合 2 个模型的取证知识,通过融合双流结果得到最后的分类结果。鉴于本文的任务是一个二分类任务,针对 2 个模型的输出,本文应用 *sigmoid* 函数将 2 个模型的输出映射到 $[0, 1]$ 的概率分布中,*sigmoid* 函数可被表示为

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

对于 2 个模型的输出可分别表示为

$$p_1 = [p_{r1}, p_{f1}], p_2 = [p_{r2}, p_{f2}] \quad (4)$$

式(4)中: p_{r1}, p_{f1} 分别为运动特征训练的模型将视频预测为真视频的概率和假视频的概率; p_{r2}, p_{f2} 分别为差异特征训练的模型将视频预测为真视频和假视频的概率。通过 *sigmoid* 函数之后将对双流的输出求和得到最终预测结果,表达式为

$$p_r = \text{sig}(p_{r1}) + \text{sig}(p_{r2}), \quad p_f = \text{sig}(p_{f1}) + \text{sig}(p_{f2}) \quad (5)$$

式(5)中: p_r 和 p_f 分别为模型最终将视频预测为真视频和假视频的概率。在我们的分类任务中,当假视频的预测概率值大于真视频时,将该视频判定为假视频,否则将其判定为真视频。通过上述过程,模型将 2 个模型输出的结果融合得到最终的检测结果。

3 实验结果与分析

为进一步验证本文提出的基于面部动作不一致的压缩深度伪造视频检测方法的性能,本节将在相应的伪造数据集上进行相应的实验,验证所提出方法的压缩视频检测性能及其鲁棒性。

3.1 数据集和实验设置

本文实验使用目前公开的具有挑战性的 FaceForensics++ (FF++)^[37] 数据集。FaceForensics++ 数据集包含 1 000 个来自 Youtube 的原始视频和 5 种类型的伪造视频,分别为 DeepFakes (DF)、Face2Face (F2F)、FaceSwap (FS)、NeuralTextures (NT) 和 FaceShifter (FSh)。这些视频包含压缩因子 C23 和 C40

的 2 个压缩版本,即 FaceForensics++ HQ 和 FaceForensics++ LQ。本文关注社交网络中压缩深度伪造视频,实验中,主要使用 FaceForensics++ LQ 中的压缩视频进行实验以评估所提方法的有效性,并在不同压缩因子 C23 和 C40 压缩的视频上,跨压缩率的检测进一步验证了所提方法具有较好的鲁棒性。在具体的实验中,训练集和测试集按照 8:2 的比例划分,即 80%的数据用于模型训练,20%的数据用于模型测试。本文使用的 2 个模型均为 LSTM,2 个模型结构一致,训练时超参数保持一致。具体来说,使用了 LSTM 的双向版本,每个方向包含 64 个 LSTM 单元,以同时处理正向和反向的序列信息。最后一个时间步的正向和反向隐状态在拼接后传递给全连接层,进一步通过 *softmax* 激活函数得到模型的 2 个预测概率值。训练的批量大小为 1 024,最大迭代次数

为 400,优化器为 Adam,学习率为 0.001。

3.2 FaceForensics++ LQ 检测结果

为了验证基于面部动作不一致的方法检测压缩深度伪造视频的性能,本小节在上一节说明了 FaceForensics++ LQ 压缩数据集进行相关实验并与较先进的检测方法进行对比。具体的实验结果分别如表 2 和表 3 所示。

分析上述结果可知,提出的基于面部动作不一致的压缩深度伪造人脸视频检测方法在检测压缩视频时取得较好的检测性能。而由于所提方法融合双流预测结果而不依赖于自适应的分类阈值,准确率 (accuracy, ACC) 虽然整体高于部分方法,但曲线下面积 (area under curve, AUC) 却略有下降。实际上,很难为未知伪造方法预先设定最优的分类阈值。

表 2 FaceForensics++ LQ 压缩深度伪造视频数据集上的检测 ACC (%)
Tab.2 Detection ACC (%) in compressed Deepfake videos of FaceForensics++ LQ

方法	DF	FS	F2F	NT	FSh
FWA[16]	78.50	67.25	66.75	64.75	66.33
Mesonet [17]	82.49	75.25	73.25	63.75	81.75
Capsule[18]	87.50	80.50	81.25	70.50	85.12
SPSL [32]	93.48	92.26	86.02	76.78	—
Re-network[22]	76.44	63.32	67.09	62.06	74.00
FT-two-stream[31]	94.64	85.27	86.68	80.50	77.31
本文方法	85.00	90.00	87.00	82.25	73.50

注:加粗数字为最优值。

表 3 FaceForensics++ LQ 压缩深度伪造视频数据集上的检测 AUC (%)
Tab.3 Detection AUC (%) in compressed Deepfake videos of FaceForensics++ LQ

方法	DF	FS	F2F	NT	FSh
FWA[16]	86.00	68.00	70.00	67.00	69.32
Mesonet [17]	92.90	83.31	72.00	67.00	92.41
Capsule[18]	92.00	78.00	86.00	78.00	92.29
SPSL [32]	98.50	98.10	94.62	80.49	—
Re-network[22]	84.82	67.08	72.65	68.70	82.09
FT-two-stream[31]	98.00	94.00	94.00	90.00	77.49
本文方法	93.73	95.43	93.21	88.84	85.98

注:加粗数字为最优值。

3.3 FaceForensics++ 跨压缩率检测结果

为进一步验证所提出的方法检测压缩深度伪造视频的有效性,本节进行了跨压缩率的实验,具体的检测结果如表 4 所示,其中 C23-C40 表示使用 C23

的数据训练模型,C40 的数据测试。C40-C23 表示使用 C40 的数据训练模型,C23 的数据测试。

分析表 4 的结果可知,所提出的基于面部动作不一致的方法对不同压缩因子具有较好的鲁棒性,

在进行跨压缩率的检测时取得了较好的检测性能。

表 4 FaceForensics++跨压缩率深度伪造视频检测 ACC (%)

Tab.4 Deepfake video detection ACC (%) in FaceForensics++

缩放参数	Capsule[18]		FT-two-stream[31]		本文方法	
	C23-C40	C40-C23	C23-C40	C40-C23	C23-C40	C40-C23
DF	67.75	79.27	88.78	84.20	76.75	89.00
FS	54.00	73.90	77.66	84.19	81.00	94.50
F2F	55.75	80.00	79.95	86.75	82.50	87.72
NT	53.75	63.66	77.88	80.52	77.00	86.25
FSH	—	—	—	—	72.25	87.25

注:加粗数字为最优值。

3.4 消融实验

本文的方法关注时序上的关系,输入帧长度会影响模型在时域建模的能力,于是设置帧长度取值 $n \in \{50, 60, 70, 80, 90\}$, 观察模型在不同帧长度下的检测性能,结果如表 5 所示。

表 5 消融实验-不同帧长度模型的深度伪造视频检测 ACC (%)

Tab.5 Ablation study-Deepfake video detection ACC (%) of different frame length models

n	DF	FS	F2F	NT	FSH	Avg
50	86.50	86.50	85.75	78.25	74.00	82.2
60	86.00	88.25	85.00	75	74.25	81.70
70	85.00	90.00	87.00	82.25	73.50	83.55
80	86.75	87.00	85.50	79.25	75.50	82.80
90	83.25	87.00	87.25	79.00	71.50	81.60

注:加粗数字为最优值。

由表 5 知,当输入的帧长度较短时,模型不能较好地对应时域建模,难以有效捕获时间特征,模型的检测性能受到一定的影响。而当输入的帧长度较长时可能会导致过拟合现象,影响模型的检测性能。当输入的帧长度为 70 时,模型取得最好的检测性能,因此在本文的实验中,输入的帧长度 $n=70$ 。

本文提出的基于面部动作不一致的压缩视频检测方法从运动和速度 2 种模式捕获伪造人脸视频中五官区域不规则的运动。本节进行消融实验以说明从 2 个角度能够更有效地捕获伪造人脸视频中的不规则运动,检测结果如表 6 所示。

分析表 6 可知,单独使用一个模式时候的性能会低于两个模式相结合时候的性能,从 2 个模式能够更有效地捕获五官区域不规则的运动,提高模型

的检测性能。

表 6 消融实验-单流与双流的深度伪造视频检测 ACC (%)

Tab.6 Ablation study-Deepfake video detection ACC (%) of two branches

数据集	DF	FS	F2F	NT	FSH
MV	73.33	58.00	67.48	64.22	64.77
DMV	80.60	80.89	82.40	75.85	71.72
双流融合	85.00	90.00	87.00	82.25	73.50

注:加粗数字为最优值。

4 结束语

本文开展了社交网络中压缩深度伪造视频的检测研究。基于对伪造视频中面部动作的分析,提出了基于面部动作不一致的压缩深度伪造视频检测方法。该方法通过对提取的面部地标点进行降噪处理,并基于提取的地标点构建五官角度特征,以此来直观表示五官区域的运动,进而从运动和速度 2 个模式出发,有效捕获五官区域不规则的运动。对比实验表明,所提出的方法能够有效检测压缩深度伪造视频且对压缩因子具有较好的鲁棒性。此外,本文所设计的特征维度较小、模型的参数较低、易于训练,解决了深度学习模型难以训练的问题。

参考文献:

- [1] WESTERLUND M. The emergence of deepfake technology: A review[J]. Technology Innovation Management Review, 2019, 9(11): 39-49.
- [2] ZHANG P, YANG L, XIE X, et al. Lightweight texture correlation network for pose guided person image generation[J]. IEEE Transactions on Circuits and Systems for

- Video Technology, 2022, 32(7): 4584-4598.
- [3] BİZGÂ A. Consumer Insights [EB/OL]. (2023-10-30) [2025-02-25]. <https://www.bitdefender.com/blog/hot-forsecurity>.
- [4] 暴雨轩, 芦天亮, 杜彦辉. 深度伪造视频检测技术综述 [J]. 计算机科学, 2020, 47(9): 283-292.
- BAO Y X, LU T, DU Y H. Overview of Deepfake video detection technology [J]. Computer Science, 2020, 47(9): 283-292.
- [5] 曹建峰. 深度伪造技术的法律挑战及应对 [J]. 信息安全与通信保密, 2019, 310(10): 35-40.
- CAO J F. Legal challenges and countermeasures of Deepfake technology [J]. Information Security and Communication Confidentiality, 2019, 310(10): 35-40.
- [6] 李旭嵘, 纪守领, 吴春明, 等. 深度伪造与检测技术综述 [J]. 软件学报, 2021, 32(2): 496-518.
- LI X R, JI S L, WU C M, et al. Survey on deepfakes and detection techniques [J]. Journal of Software, 2021, 32(2): 496-518.
- [7] 周文柏, 张卫明, 俞能海, 等. 人脸视频深度伪造与防御技术综述 [J]. 信号处理, 2021, 37(12): 2338-2355.
- ZHOU W B, ZHANG W M, YU N H, et al. An overview of Deepfake forgery and defense techniques [J]. Journal of Signal Processing, 2021, 37(12): 2338-2355.
- [8] 李纪成, 刘琲贝, 胡永健, 等. 基于光照方向一致性的换脸视频检测 [J]. 南京航空航天大学学报, 2020, 52(5): 760-767.
- LI J C, LIU B B, HU Y J, et al. Deepfake video detection based on consistency of illumination direction [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 760-767.
- [9] AGARWAL S, FARID H, GU Y, et al. Protecting world leaders against deep fakes [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. California, USA: IEEE, 2019: 38-45.
- [10] YANG X, LI Y, LYU S. Exposing deep fakes using inconsistent head poses [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Brighton, UK: IEEE, 2019: 8261-8265.
- [11] CİFTÇİ U A, DEMİR İ, YIN L. Fakecatcher: Detection of synthetic portrait videos using biological signals [EB/OL]. (2020-07-15) [2024-02-25]. <https://doi.org/10.1109/TPAMI.2020.3009287>.
- [12] LI Y, CHANG M C, LYU S. Inictu oculi: Exposing ai created fake videos by detecting eye blinking [C]// Proceedings of IEEE International Workshop on Information Forensics and Security. Hong Kong, China: IEEE, 2018: 1-7.
- [13] MATERN F, RIESS C, STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations [C]// Proceedings of IEEE Winter Applications of Computer Vision Workshops. Hawaii, USA: IEEE, 2019: 83-92.
- [14] HOSLER B, SALVI D, MURRAY A, et al. Do deepfakes feel emotions? A semantic approach to detecting deepfakes via emotional inconsistencies [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA: IEEE, 2021: 1013-1022.
- [15] SUN Z, HAN Y, HUA Z, et al. Improving the efficiency and robustness of deepfakes detection through precise geometric features [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021: 3609-3618.
- [16] LI Y, LYU S. Exposing deepfake videos by detecting face warping artifacts [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. California, USA: IEEE, 2019: 46-52.
- [17] AFCHAR D, NOZICK V, YAMAGISHI J, et al. Mesonet: A compact facial video forgery detection network [C]// Proceedings of IEEE International Workshop on Information Forensics and Security. Hong Kong, China: IEEE, 2018: 1-7.
- [18] NGUYEN HH, YAMAGISHI J, ECHIZEN I. Capsule-forensics: Using capsule networks to detect forged images and videos [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Brighton, UK: IEEE, 2019: 2307-2311.
- [19] SUN K, LIU H, YAO T, et al. An information theoretic approach for attention-driven face forgery detection [C]// Proceedings of European Conference Computer Vision. Tel Aviv, Israel: Springer, 2022: 111-127.
- [20] YANG J, XIAO S, LI A, et al. Msta-net: Forgery detection by generating manipulation trace based on multi-scale self-texture attention [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4854-4866.
- [21] CAO J, MA C, YAO T, et al. End-to-end reconstruction-classification learning for face forgery detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 4113-4122.
- [22] GÜERA D, DELP E J. Deepfake video detection using recurrent neural networks [C]// Proceedings of IEEE International Conference on Advanced Video and Signal based Surveillance. Auckland, New Zealand: IEEE,

- 2018; 1-6.
- [23] SABIR E, CHENG J, JAISWAL A, et al. Recurrent convolutional strategies for face manipulation detection in videos[J]. *Interfaces*, 2019, 3(1): 80-87.
- [24] AMERINI I, GALTERI L, CALDELLI R, et al. Deepfake video detection through optical flow basedcnn[C]// *Proceedings of IEEE International Conference on Computer Vision Workshops*. Seoul, Korea: IEEE, 2019: 1-3.
- [25] MASI I, KILLEKAR A, MASCARENHAS R M, et al. Two-branch recurrent network for isolating deepfakes in videos [C]// *Proceedings of European Conference on Computer Vision*. Glasgow, UK: Springer, 2020: 667-684.
- [26] QI H, GUO Q, JUEFEI-XU F, et al. DeepRhythm: Exposing deepfakes with attentional visual heartbeat rhythms [C]// *Proceedings of ACM International Conference on Multimedia*. Seattle, USA: ACM, 2020: 4318-4327.
- [27] HALIASSOS A, VOUGIOUKAS K, PETRIDIS S, et al. Lips don't lie: A generalisable and robust approach to face forgery detection[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE, 2021: 5039-5049.
- [28] WANG H, LIU Z, WANG S. Exploiting complementary dynamic incoherence for deepfake video detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(8): 4027-4040.
- [29] PANG G, ZHANG B, TENG Z, et al. Mre-net: Multi-rate excitation network for deepfake video detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(8): 3663-3676.
- [30] YANG Z, LIANG J, XU Y, et al. Masked relation learning for deepfake detection[J]. *IEEE Transactions on Information Forensics and Security*, 2023, 18(1): 1696-1708.
- [31] 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*, 2021, 32(3): 1089-1102.
- [32] LIU H, LI X, ZHOU W, et al. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE, 2021: 772-781.
- [33] BINH L M, WOO S. Add: Frequency attention and multi-view based knowledge distillation to detect low-quality compressed deepfake images[C]// *Proceedings of AAAI Conference on Artificial Intelligence*. California, USA: AAAI, 2022: 122-130.
- [34] MARCON F, PASQUINI C, BOATO G. Detection of manipulated face videos over social networks: A large-scale study[J]. *Journal of Imaging*, 2021, 7(10): 193-208.
- [35] ZHANG X, WU X. Multi-modality deep restoration of extremely compressed facevideos[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2024-2037.
- [36] 吴珂, 周梦莹, 李高阳, 等. 基于角度几何特征的人脸表情识别[J]. *计算机应用与软件*, 2020, 37(7): 120-124.
- WU K, ZHOU M Y, LI G Y, et al. Facial expression recognition based on geometrical features of angles[J]. *Computer Applications and Software*, 2020, 37(7): 120-124.
- [37] ROSSLER A, COZZOLINO D, VERDOLIVA L, et al. Faceforensics++: Learning to detect manipulated facial images[C]// *Proceedings of IEEE International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019: 1-11.

作者简介:

王玉梅, 硕士, 主要研究方向为多媒体取证。E-mail: yumei-wang@hnu.edu.cn。

胡娟, 博士, 新加坡国立大学博士后, 主要研究方向为生成式 AI 内容取证。E-mail: hujian@hnu.edu.cn。

武晓帅, 博士生, 主要研究方向为信息隐藏和人工智能安全。E-mail: shinewu@hnu.edu.cn。

肖靖, 教授, 硕士生导师, 博士, 主要研究方向为芯片工艺的高性能计算与应用研发等。E-mail: jxiao1985@hnu.edu.cn。

陈雁翔, 教授, 博士生导师, 博士, 主要研究方向为多媒体信息处理、多媒体内容安全等。E-mail: chenyx@hfut.edu.cn。

廖鑫, 教授, 博士生导师, 博士, 主要研究方向为信息内容安全、人工智能安全等。E-mail: xinliao@hnu.edu.cn。

(编辑: 田海江)