引文格式：胡永健，佘惠敏，刘琲贝，等.利用人脸3DMM重构信息检测深度伪脸视频[J].武汉大学学报(信息科学版)，2024，49(2)：190-196.DOI：10.13203/j.whugis20210427

Citation：HU Yongjian，SHE Huimin，LIU Beibei，et al.Deepfake Video Detection Using 3DMM Facial Reconstruction Information[J].Geomatics and Information Science of Wuhan University，2024，49(2)：190-196.DOI：10.13203/j.whugis20210427

利用人脸3DMM重构信息检测深度伪脸视频

胡永健[1]　佘惠敏[1]　刘琲贝[1]　陈香全[1]　刘光尧[2]

1　华南理工大学电子与信息学院，广东　广州，510641

2　公安部物证鉴定中心，北京，100038

摘　要：提出一种基于人脸三维形变模型(3D morphable model，3DMM)的深度伪脸视频检测算法，利用3DMM强大的人脸形状、纹理、表情和姿态参数估算能力，逐帧获取鉴伪基本信息。设计面部行为特征计算模块和静态外貌特征提取模块，以滑动窗为单位，在时间轴上分别从表情和姿态参数提取人物的面部行为特征，从形状和纹理参数计算人物的静态外貌特征。鉴伪过程利用人物外貌特征与面部行为特征的一致性来完成。所提出的算法人物针对性强，可解释性好。该方法与同类算法相比，半总错误率更低，抗视频压缩能力更好，计算更加简便。

关键词：深度伪脸视频检测；三维形变模型；人脸外貌特征；面部行为特征；深度神经网络；面部生物特征

中图分类号：P237；TN911.73　　**文献标识码**：A　　**收稿日期**：2022-02-24

DOI：10.13203/j.whugis20210427　　**文章编号**：1671-8860(2024)02-0190-07

Deepfake Video Detection Using 3DMM Facial Reconstruction Information

HU Yongjian[1]　*SHE Huimin*[1]　*LIU Beibei*[1]　*CHEN Xiangquan*[1]　*LIU Guangyao*[2]

1　School of Electronic and Information Engineering，South China University of Technology，Guangzhou 510641，China

2　Institute of Forensic Science，Ministry of Public Security，Beijing 100038，China

Abstract：**Objectives**：The emergence of deepfake technique leads to a worldwide information security problem. Deepfake videos are used to manipulate and mislead the public. Though there have been a variety of deepfake detection methods，the features extracted generally suffer from poor interpretability. To solve this problem，a deepfake video detection method using 3D morphable model (3DMM) of face is proposed. **Methods**：The 3DMM is employed to estimate parameters of shape，texture，expression，and gesture of the face frame by frame，constituting the basic information of deepfake detection. The facial behavior feature extraction module and the static face appearance feature extraction module are designed for the construction of feature vectors on a sliding window basis. The facial behavior feature vector is derived from the expression and gesture parameters while the appearance feature vector is calculated with the shape and texture parameters. The consistency measured by cosine distance between the appearance feature vector and the behavior feature vector is the criterion for authentication of the face for each sliding window across the video. **Results**：The effectiveness of the proposed method is evaluated with three public datasets. The overall half total error rates (HTER) obtained on FF＋＋，DFD and Celeb-DF dataset are 1.33%，4.93% and 3.92% respectively. For the severely compressed videos，C40 of DFD，the HTER is 7.09%，showing a good robustness against video compression. The model complexity is around 1/4 of that of the most related work. **Conclusions**：The proposed algorithm has good person pertinence and clear interpretability. Compared with state-of-the-art methods in literature，the proposed algorithm demonstrates lower half total error rates，better resistance to video compression and less computational cost.

Key words：deepfake detection；3D morphable model；appearance feature；facial behavior feature；convolutional neural network；facial biometric

基金项目：国家重点研发计划(2019QY2202)；广州开发区国际合作项目(2019GH16)；中新国际联合研究院项目(206-A018001)。

第一作者：胡永健，博士，教授，主要从事多媒体信息安全、图像处理及人工智能研究。eeyjhu@scut.edu.cn

通讯作者：刘琲贝，博士。eebbliu@scut.edu.cn

深度伪造是指利用深度学习网络生成并篡改视频中的人物面部，通过换脸实现身份更改或偷换的技术。为追求更逼真的伪造效果，还会融合计算机图形学方法对人脸建模[1]。现有深度伪造假脸视频检测器主要通过检测伪造痕迹来判别视频真假。如文献[2]提出唇读取证的方法，利用伪造视频中口腔运动高级语义上的不规则性作为检测线索来区分真假唇形；文献[3]利用换脸视频中合成人脸与目标人脸面部特征点所估计的头部姿势的不一致性来检测伪造视频；文献[4]根据换脸视频在协调眼球运动的时域一致性方面存在局限性，利用眼动特征判断视频是否存在换脸；文献[5]根据换脸过程存在人脸标志点偏移这一局限性，基于人脸标志点构造时空域特征角特征来检测换脸视频；文献[6]利用视频伪造过程中假脸图片放缩留下的痕迹；文献[7]将视频换脸视为特殊的拼接篡改问题，利用神经分割网络预测篡改区域，再利用预测区域与先验人脸区域的交并比来判断真假；文献[8]将篡改区域定位任务与分类任务相结合，设计可训练的注意力模块，使用掩膜标签计算损失函数引导网络训练，改善检测模型的泛化性能。

上述文献方法采用了不同的检测思路。文献[2-4]依据真假人脸的生理部位（包括唇、头、眼）特征差异；文献[5-6]依据换脸视频制作过程中的整体画面缺陷或痕迹。可以看到，这两类方法都具有较清晰的物理解释，其原因是所使用的特征基于有明确物理定义的空间或变换域算子，最后的决策由机器学习的模式分类器进行。文献[7-8]依据深度神经网络从训练数据集中学习特征来进行真假人脸视频的分类，库内检测准确率也明显高于前面两类方法，由于依赖深度卷积神经网络提取鉴伪特征，物理意义不够清晰。

上述 3 类方法存在如下几个方面的问题：(1)检测性能易受图像篡改方式、换脸生成方法、视频压缩率等因素的影响；(2)手工特征的可解释性好但准确率不够高，深度网络检测器的准确率高但可解释性不够好；(3)属于泛对象检测，即对所有检测对象都同等对待，缺少针对特定人物对象的检测方法。

生物特征是人类最本质的特征，不易受视频压缩等常见视频处理操作的影响。软生物特征指在语义层面上对人物身份信息进行描述进行描述，例如，一个人说话时面部表情和头部姿态会展现出独特的动态模式。

图 1 所示为 Celeb-DF 中同一个人物身份对应的 A、B、C 不同视频段的一段动作采样示意图。

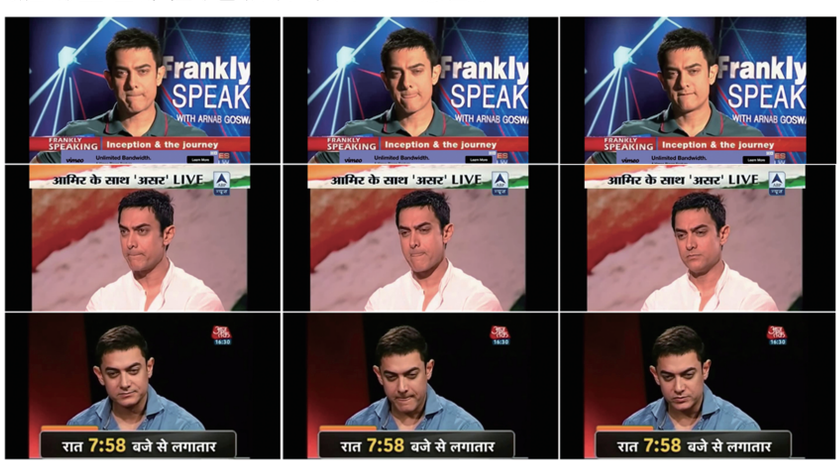

图 1　DFD 数据库人脸面部软生物特征示意图

Fig. 1　Soft Biological Features of Human Face in DFD Database

图 1 中人物在说话时会呈现出一个习惯性的头部动作和面部表情。对比 A、B、C 视频可以发现，在第 1 列的视频帧中，人物抿紧嘴唇，睁大双眼，头部微微倾斜；第 2 列视频帧中，人物眯眼，做出抿嘴吐舌的动作，同时微微低头；第 3 列视频帧中，人物抿嘴，眉间微蹙。这一系列面部行为特点在该人物的其他视频中也反复出现，故这一软生物特征可与人物身份关联。

深度身份伪造技术将目标人物身份（即目标人脸）篡改为源人物身份（即源人脸），但保留了目标人物的面部动作和表情。因此，伪造人物在说话时，虽具有目标人物的面部动作，但外貌形象却是源人物的，存在面部行为特征与外貌特征不一致的情况。受此启发，加州大学伯克利分校 Agarwal 等[9-10]提出一种解决方案，设计了一个带度量学习目标函数的卷积神经网络：在训练阶段，从参考视频中逐段学习人脸外貌和行为特征，构造各人物的人脸外貌特征和行为特征参考

集;在检测阶段,通过逐段对比待测视频相应的特征是否存在于上述两个参考集中,且人脸外貌特征和行为特征是否匹配来鉴定真伪。

为更好地表达人脸外貌特征,笔者认为还需引入人脸的三维信息。为此,本文提出一种基于人脸三维形变模型(3D morphable model,3DMM)图像重构信息的深度伪造假脸视频检测算法,利用3DMM模型强大的人脸形状、纹理、表情和姿态参数估算能力,获取鉴伪判断所需要的基本信息。将上述参数分别送入自行设计的外貌特征计算模块和面部行为特征提取模块,以滑动窗为单位刻画人脸的外貌特征和与之对应的行为特征[11]。逐窗根据外貌特征和行为特征的一致性来判断视频是否发生换脸操作。实验结果表明,本文算法比文献[10]具有更低的半总错误率和更好的抗视频压缩能力,且计算复杂度更低。同时,与其他泛对象假脸视频检测算法相比,亦具有独特优势。除了文献[9-10],尚未见到基于外貌特征和行为特征一致性的假脸视频检测工作。

1 本文算法

本文算法整体流程如图2所示。ID为IDentity缩写,功能模块包括3DMM图像重构信息输出模块、面部行为特征提取模块、面部外貌特征计算模块及真假判断模块4部分。

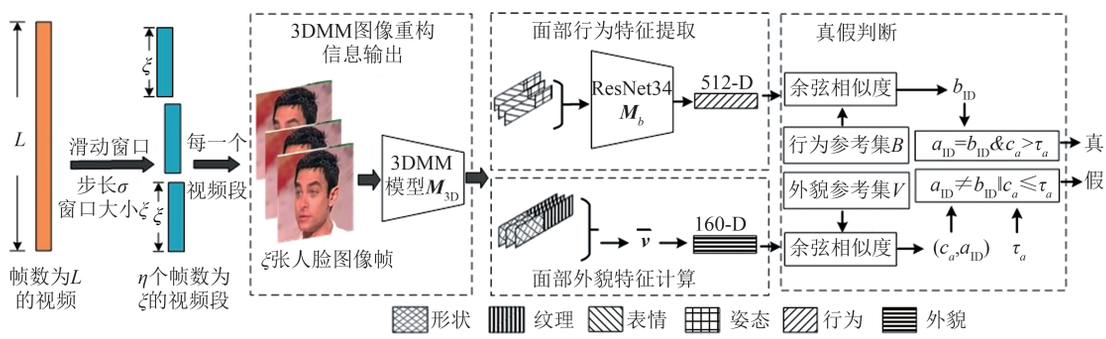

图2 本文算法流程及其主要功能模块

Fig. 2 Our Algorithm Flowchart and Its Major Function Modules

1.1 3DMM图像重构信息输出模块

用于篡改检测的人脸基本信息包括形状、纹理、表情和姿态。其中,形状和纹理用于刻画面部外貌(即物理特征),而表情与头部姿势用于刻画人物说话时的面部行为(即软生物特征)。文献[12]利用3D扫描人脸数据库统计出3D人脸的规律,构造出一个可形变的人脸模型,用于3D人脸重建。文献[13]提出一个3DMM模型用于抗姿态和光照变化的人脸识别。文献[14]提出一个用于可视化计算的3D人脸表情数据库。

从数学上来说,假设任意三维人脸可由线性空间中的m个人脸向量进行加权组合得到,则每张人脸包含的形状向量和纹理向量分别表示为:

$$\boldsymbol{S}=\bar{\boldsymbol{S}}+\sum_{i=1}^{m-1}a_i\boldsymbol{s}_i+\sum_{i=1}^{n-1}\gamma_i\boldsymbol{e}_i,\boldsymbol{T}=\bar{\boldsymbol{T}}+\sum_{i=1}^{m-1}b_i\boldsymbol{t}_i \quad (1)$$

式中,$\bar{\boldsymbol{S}}$与$\bar{\boldsymbol{T}}$分别为BFM(basel face model)数据集[13]平均人脸形状向量和平均人脸纹理向量;a_i为形状参数;b_i为纹理参数;$\boldsymbol{s}_i$和$\boldsymbol{t}_i$分别代表BFM数据集中的形状主成分和纹理主成分;$\boldsymbol{e}_i$为FaceWarehouse数据集[14]中的人脸表情主成分;γ_i为人脸表情参数;m与n为模型基向量数量。通过改变人脸形状参数a_i可生成不同人物身份但具有相同表情的三维人脸,通过改变人脸表情参数γ_i可生成相同身份但具有不同表情的三维人脸,如悲伤、喜悦和惊讶等表情。三维重建的标准人脸是正向人脸,可借助旋转矩阵和平移向量来表示人物头部的三维姿态信息。

本文借用上述3DMM模型一次性获取形状、纹理、表情和姿态4类信息,即形状参数向量$\boldsymbol{\alpha}_{ID}=(a_1,a_2,\cdots,a_i)^{\mathrm{T}}\in\mathbb{R}^i$,纹理参数向量$\boldsymbol{\alpha}_{tex}=(b_1,b_2,\cdots,b_t)^{\mathrm{T}}\in\mathbb{R}^t$,表情参数向量$\boldsymbol{\alpha}_{exp}=(\gamma_1,\gamma_2,\cdots,\gamma_e)^{\mathrm{T}}\in\mathbb{R}^e$和姿态参数矩阵。具体实现时,采用文献[15]中基于弱监督学习的3D人脸重建算法,其主干网络为ResNet50,通过连接257个神经元组成的全连接层,输出257维向量,如图3所示。其中形状参数向量$\boldsymbol{\alpha}_{ID}\in\mathbb{R}^{80}$,纹理参数向量$\boldsymbol{\alpha}_{tex}\in\mathbb{R}^{80}$以及表情参数向量$\boldsymbol{\alpha}_{exp}\in\mathbb{R}^{64}$,并将姿态参数简化为三维向量$\boldsymbol{p}\in\mathbb{R}^3$,表示人物头部3个维度上的姿态角。

1.2 面部行为特征与外貌特征提取模块

本文在文献[16]中的ResNet34基础上构建面部行为特征提取模块,提取时间轴上滑动窗中全体帧图像的表情和姿态来表征与滑动窗视频

段对应的面部行为特征。为了学习映射 $f(\cdot)$，提出利用多相似性度量学习损失函数（multi-similarity loss，MS-Loss）[17]来训练 ResNet34，得到维度 $d=512$ 的行为特征。

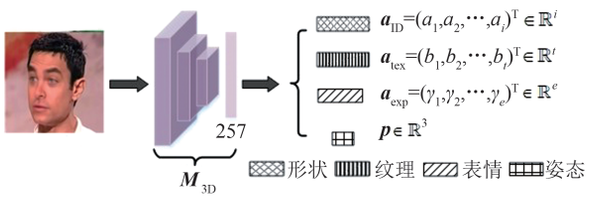

图 3　3DMM 图像重构信息输出模块

Fig. 3　3DMM Image Reconstruction Information Output Module

具体而言，对长度为 ξ 帧的视频段，逐帧提取表情参数向量 $\boldsymbol{\alpha}_{\mathrm{exp}}\in\mathbb{R}^{64}$ 和姿态参数向量 $\boldsymbol{p}\in\mathbb{R}^3$，串接得到 $\chi=\left[\boldsymbol{\alpha}_{\mathrm{exp}},\boldsymbol{p}\right]\in\mathbb{R}^{67}$，再对 ξ 帧图像的 χ 进行堆叠，得到矩阵 $\chi=\left[\chi_1^{\mathrm{T}},\chi_2^{\mathrm{T}},\cdots,\chi_\xi^{\mathrm{T}}\right]\in\mathbb{R}^{67\times\xi}$。将 χ 输入至面部行为特征提取网络，利用卷积神经网络学习得到映射 $f(\cdot):\mathbb{R}^{67\times\xi}\rightarrow\mathbb{R}^d$，网络的输出即可用于表征人物的面部行为特征。

把单帧图像提取得到的形状参数向量 $\boldsymbol{\alpha}_{\mathrm{ID}}\in\mathbb{R}^{80}$ 和纹理参数向量 $\boldsymbol{\alpha}_{\mathrm{tex}}\in\mathbb{R}^{80}$ 进行拼接得到单幅人脸的外貌特征 $\boldsymbol{\varphi}=\left[\boldsymbol{\alpha}_{\mathrm{ID}},\boldsymbol{\alpha}_{\mathrm{tex}}\right]\in\mathbb{R}^{160}$。以滑动窗中 ξ 帧图像的 $\boldsymbol{\varphi}$ 在时间维度上的平均值来表达与滑动窗视频段对应的面部外貌特征 $\boldsymbol{v}=\sum_{i=1}^{\xi}\boldsymbol{\varphi}_i/\xi\in\mathbb{R}^{160}$。

1.3　参考集的构建、检测流程和判断决策

无论是在训练还是测试阶段，视频的预处理方式均相同。以视频的滑动窗为基本检测单元，将长度为 L 的视频采用窗口大小为 ξ、步长为 σ（$\sigma\leqslant\xi$）的方式划分成 η 段。在训练阶段，用给定全体人物的视频构造参考集。对于每个人物 ID 的视频，首先按上述方式进行预处理；然后逐滑动窗得到 512 维的面部行为特征 $\boldsymbol{b}$ 与 160 维的面部外貌特征 $\boldsymbol{v}$，组成有人物 ID 对应关系的面部行为特征参考集 $\boldsymbol{B}$ 和外貌特征参考集 $\boldsymbol{V}$。

在测试阶段，将待测视频预处理后得到滑动窗视频段，设为 t，得到 512 维的面部行为特征 $\boldsymbol{b}_t$ 与 160 维的面部外貌特征 $\boldsymbol{v}_t$，分别与参考集 $\boldsymbol{B}$ 和参考集 $\boldsymbol{V}$ 里的特征向量进行余弦相似度匹配，得到外貌特征相似度最大时所对应的人物标签 a_{ID} 和面部行为特征相似度最大时所对应的人物标签 b_{ID}，以及最大的外貌相似度 c_a，如下所示：

$$\begin{cases}a_{\mathrm{ID}}=\arg\max\limits_{\mathrm{ID}}\left\{\max\left(\boldsymbol{v}_i\cdot\boldsymbol{v}_t\right)\right\}\\ b_{\mathrm{ID}}=\arg\max\limits_{\mathrm{ID}}\left\{\max\left(\boldsymbol{b}_j\cdot\boldsymbol{b}_t\right)\right\}\\ c_a=\max\left(\boldsymbol{v}_i\cdot\boldsymbol{v}_t\right)\end{cases}\tag{2}$$

式中，$\boldsymbol{v}_i$ 表示外貌特征参考集 $\boldsymbol{V}$ 中第 i 个特征向量；$\boldsymbol{b}_j$ 表示面部行为特征参考集 $\boldsymbol{B}$ 中第 j 个特征向量。若 $a_{\mathrm{ID}}\neq b_{\mathrm{ID}}$，则判定该窗测试视频段为假。若 $a_{\mathrm{ID}}=b_{\mathrm{ID}}$，再进一步检查 c_a 是否大于设定的外貌特征相似度阈值 τ_a：若 $c_a>\tau_a$，则判定该窗测试视频段为真；否则为假。

2　实验结果与分析

深度伪造视频检测可看成是真假二分类任务，将真实视频看作阳性样本，阳性的正确预测数目记为 TP（true positive），错误预测数目为 FN（false negative）；将深度伪造视频看作阴性样本，阴性的正确预测数目记为 TN（true negative），错误预测数目为 FP（false positive）。真阳率（TP rate，TPR）=TP/（TP+FN），假阳率（FP rate，FPR）=FP/（TN+FP），真阴率（TN rate，TNR）=TN/（TN+FP），假阴率（FN rate，FNR）=FN/（TP+FN）。半总错误率（half total error rate，HTER）是将 FPR 和 FNR 综合起来的一种评测手段，计算公式为：HTER=（FPR+FNR）/2。ROC（receiver operating characteristic curve）曲线又称接受者操作特征曲线，是根据不同阈值画出来的一条曲线，横坐标与纵坐标分别为 FPR 和 TPR。AUC（area under curve）则表示 ROC 曲线下的面积。

2.1　实验环境与数据集

算法主要基于框架 Pytorch-1.1.0 实现，GPU 卡为 TITAN XP，系统为 Ubuntu16.04，CPU 为 Intel（R）Xeon（R）CPU E5-2683 v3 @ 2.00 GHz，CUDA 版本为 9.0.0，CUDNN 版本为 7.1.4。利用 OpenCV 的 VideoCapture 类将视频解码成图像序列，对序列中每帧图像利用 RetinaFace[18]人脸检测算法检测并裁剪出人脸区域，利用人脸标志点进行对齐，再用双线性插值方法将其统一调整成大小为 224×224 像素的图像。

面部行为特征提取网络需要预先训练。笔者采用 VoxCeleb2 视频数据集[19]作为训练集。VoxCeleb2 是牛津大学制作的视听人脸数据集，全部为真实场景下的人物说话面部视频，本文随机选取了 354 位人物，共 27 207 段视频。设置最

大训练迭代次数为20 000,训练批尺寸为128,采用SGD优化器作为训练优化器,初始学习率为0.01,采用学习率随训练迭代次数衰减策略,避免模型训练后期在最小值附近徘徊。

算法的性能验证是在近两年深度伪造研究中使用的最为广泛的3个数据集上进行的,包括FF++(FaceForensics++)[20]、DFD(deepfakes detection)[21]和Celeb-DF[11]。按文献[10]的做法,DFD库只选择了有说话的人物个体视频段:从真实视频的363段中选185段,从假脸视频的3 068段中选1 577段。表1列出了上述所使用的4个数据集的相关统计数据和属性。

表1 数据集基本情况

Tab. 1 Information of Datasets

项目	数据集的数据和属性			
	VoxCeleb2	FF++	DFD	Celeb-DF
真视频数	27 207	1 000	185	590
假视频数	0	1 000	1 577	5 639
人物	354	1 000	28	59

划分数据集。本文针对每个人物ID统一随机选择80%的真实视频段作为参考集,将剩余20%的真实视频段和所有的篡改视频段作为测试集。

2.2 外貌特征相似度阈值τ_a的选取

τ_a的选取会影响算法的性能,只有待检测外貌特征向量与外貌参考集中某向量的余弦相似度大于设定阈值时,才被认为是有效的匹配;否则,认为待检测视频是伪造视频或是不相关视频(即待测视频的人物不在参考集中)。图4给出了FF++、DFD和Celeb-DF这3个数据集的真实样本和伪造样本中阈值τ_a与准确率的关系曲线。

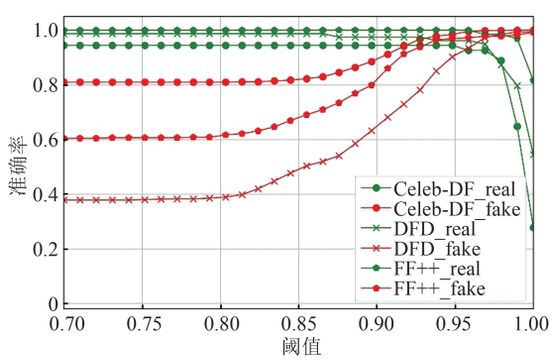

图4 3个数据集外貌相似度阈值曲线

Fig. 4 Threshold Curves of Appearance Similarity on the 3 Datasets

图4显示,在3个数据集上,曲线都有相似的趋势:τ_a在一定范围内,绿线(即TPR)和红线(即TNR)基本保持不变;当$\tau_a>0.8$时,TNR开始上升,FPR开始降低;τ_a超过0.98后,TPR开始急速下降,表明FNR增加。综合分析,取$\tau_a=0.95$在不同数据集上均能有效平衡真实视频与伪造视频检测准确率,故本文实验设置$\tau_a=0.95$。

2.3 视频滑动窗尺寸ξ对算法性能的影响分析

滑动窗尺寸ξ是影响算法性能的因素之一,ξ太小则窗内视频太短,无法准确刻画面部行为的时域特征;ξ太大则窗内视频太长,包含过多不同的面部行为,会相互干扰,同时会增加计算复杂度。本文取不同的ξ值来讨论对算法性能的影响,以DFD数据集为例,$\xi=50$、100、200帧这3种情况的实验结果见表2。

表2 视频序列长度对算法性能的影响/%

Tab. 2 Influence of Video Lengths on the Detection Performance/%

指标	滑动窗尺寸ξ/帧		
	50	100	200
TPR	95.18	**96.20**	89.47
TNR	**94.30**	93.94	92.65
HTER	5.26	**4.93**	8.94

当$\xi=50$帧时,HTER和TPR值分别居中,说明帧长可能过短,无法捕获一个完整的面部动作;$\xi=200$帧时,HTER最大,TPR值最小,说明帧长可能过长,包含了多个不同的行为,行为之间的干扰降低了所提取特征的可区分性;当$\xi=100$帧时,HTER值最小,TPR最高,综合检测性能最好。本文选取100帧作为输入视频滑动窗大小,为避免各个视频段行为特征的重复提取,令滑动步长与窗口大小相等,即$\sigma=\xi=100$帧,相邻滑动窗无重叠视频帧。

2.4 算法的有效性、稳健性和计算复杂度

本文选取同样基于人脸外貌和行为特征的文献[10](简称Agarwal算法)作为对比算法。由于Agarwal算法没有公开源代码,本文按其所述进行仿真。在FF++、DFD以及Celeb-DF数据集上的实验结果如表3和表4所示,模型复杂度的对比结果如表5所示。

有效性方面:TPR和TNR的值越大越好,HTER的值越小越好。表3显示,在FF++数据集中,本文算法在TNR指标上略逊于Agarwal算法,但在TPR指标上高出Agarwal算法约22%,在HTER指标上有接近10%的性能提升;在DFD数据集中,本文TPR高于Agarwal算法近6%,HTER有近2.3%的性能提升。在Celeb-DF数据库中,本文TPR、TNR及HTER均优于Agarwal算法。在所测试的3个数据集上,本文算法在HTER指标上分别为1.33%、4.93%和3.92%,全部优于Agarwal算法。

表 3　本文算法与 Agarwal 算法的检测性能比较/%

Tab. 3　Comparison of Detection Performance of Our Proposed Algorithm with Agarwal Algorithm/%

算法	指标								
	数据集 FF++			数据集 DFD			数据集 Celeb-DF		
	TPR	TNR	HTER	TPR	TNR	HTER	TPR	TNR	HTER
Agarwal[10]	77.63	**99.92**	11.23	90.26	**95.22**	7.26	69.87	95.48	16.32
本文算法	**100.00**	97.35	**1.33**	**96.20**	93.42	**4.93**	**92.59**	**99.56**	**3.92**

视频传输容易受到网络带宽的影响,因此抗压缩性能是算法稳健性的重要指标。为验证本文算法的抗视频压缩稳健性,分别在 FF++和 DFD 数据集的 C0(无损)、C23(高质量)和 C40(低质量)这 3 种不同压缩率的视频上进行实验,并与 Agarwal 算法相比,实验结果见表 4。可以看到,在不同压缩率下,2 个算法的 HTER 值均有所增加,变化趋势一致,但无论在何种压缩率下,本文算法均比 Agarwal 算法的 HTER 值小,且增长的幅度更小,说明本文算法的稳健性更好。

表 4　抗视频压缩能力对比测试/%

Tab. 4　Comparison of Robustness Against Video Compression/%

算法	数据集 FF++			数据集 DFD		
	C0	C23	C40	C0	C23	C40
Agarwal[10]	11.23	13.59	20.21	7.26	9.59	16.25
本文算法	**1.33**	**1.40**	**2.28**	**4.93**	**5.39**	**7.09**

表 5 显示,在计算量方面,与 Agarwal 算法相比,本文算法在模型参数、计算量及计算时间上均有明显优势。总模型的参数量约降低了 3/4,计算量约降低了 2/3,平均时长约降低了 5/6。

表 5　本文算法与 Agarwal 算法的模型复杂度比较

Tab. 5　Comparison of Model Complexity of Our Proposed Algorithm with Agarwal Algorithm

算法	面部外貌模型指标			面部行为模型指标			总模型指标		
	参数大小/MB	FLOPs/GB	平均时长/(s•100 帧$^{-1}$)	参数大小/MB	FLOPs/GB	平均时长/(s•100 帧$^{-1}$)	参数大小/MB	FLOPs/GB	平均时长/(s•100 帧$^{-1}$)
Agarwal[10]	134.26	15.48	17.60	48.43	4.92	0.70	182.69	20.40	18.30
本文算法	**25.56**	**4.12**	**2.70**	**21.54**	**2.02**	**0.01**	**47.10**	**6.14**	**2.71**

2.5　与其他算法进行比较

考虑到文献[10]中 Agarwal 算法与 FaceWarping 等算法[6]进行了对比,本文进行了类似的实验,选取 FaceWarping 算法[6]和 FFD 算法[8]作为对比算法。为了确保准确,统一设置实验场景:所有算法均在 DFD 数据库上训练,在 DFD、FF++以及 Celeb-DF 数据集上测试,结果见表 6。

FaceWarping 有两种情况:(1)在 DFD 真实样本上训练得到的模型;(2)用 FaceWarping 算法作者提供的公开权重得到的模型。FFD_VGG16 和 FFD_Xception 是指 FFD 算法分别以 VGG16 和 Xception 网络为骨干得到的模型,本文算法按 Agarwal 算法的测试方式构造参考集。

表 6　DFD 数据库训练模型跨库测试结果/%

Tab. 6　Cross-Database Test Results of Training Model in DFD Database/%

算法	数据集 DFD		数据集 FF++		数据集 Celeb-DF	
	HTER	AUC	HTER	AUC	HTER	AUC
FaceWarping[6]	39.51	70.79	38.50	82.33	50.30	47.73
FaceWarping(公共权重)[6]	35.92	71.19	21.81	88.41	45.91	56.19
FFD_VGG16[8]	3.21	**99.80**	24.09	84.69	41.79	62.62
FFD_Xception[8]	**1.42**	99.79	29.72	88.29	46.21	63.43
Agarwal[10]	7.26	95.92	11.23	95.31	16.32	92.11
本文算法	4.93	97.50	**1.33**	**99.92**	**3.92**	**98.30**

在 DFD 库检测中,FFD_VGG16 算法和 FFD_Xception 算法的性能,比 Agarwal 算法和本文算法都好,但在 FF++和 Celeb-DF 库中的表现远远低于 Agarwal 算法和本文算法,即库内性能好但跨库性能退化,这是目前泛对象换脸视频检测算法普遍存在的问题。Agarwal 算法和本文算法不管在哪个库中性能均比较稳定,而本文算法的 HTER 和 AUC 值全面优于 Agarwal 算法。

3 结 语

本文提出了一种基于人脸3DMM图像重构信息的深度伪造假脸视频检测算法,从人脸的软生物特征和物理特征两个方面给出了判断视频中人脸真伪的依据,利用面部行为特征与人物外貌特征是否一致判断是否发生换脸操作。由于3D人脸重构研究历史长,模型较为成熟,本文利用3DMM模型可一次性准确获得人脸的4类基本特征信息,所构造的鉴伪特征质量高,时间开销低,实现方便。本文算法尤适合于对检测结果有强可解释性需求的应用场景,例如,司法鉴定、政治或公众人物的短视频鉴伪等场景。

参 考 文 献

[1] Huang Ruobing, Jia Yonghong. Face Swapping Using Convolutional Neural Network and Tiny Facet Primitive [J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(3): 335-340. (黄若冰, 贾永红. 利用卷积神经网络和小面元进行人脸图像替换[J]. 武汉大学学报(信息科学版), 2021, 46(3): 335-340.)

[2] Haliassos A, Vougioukas K, Petridis S, et al. Lips Don't Lie: A Generalisable and Robust Approach to Face Forgery Detection[C]//Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021.

[3] Yang X, Li Y Z, Lyu S W. Exposing Deep Fakes Using Inconsistent Head Poses[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 2019.

[4] Li M, Liu B, Hu Y, et al. Exposing Deepfake Videos by Tracking Eye Movements[C]//International Conference on Pattern Recognition, Milan, Italy, 2021.

[5] Li M, Liu B, Hu Y, et al. Deepfake Detection Using Robust Spatial and Temporal Features from Facial Landmarks [C]//IEEE International Workshop on Biometrics and Forensics, Rome, Italy, 2021.

[6] Li Y Z, Lyu S W. Exposing DeepFake Videos by Detecting Face Warping Artifacts[EB/OL]. [2021-02-20]. https://arxiv.org/abs/1811.00656.

[7] Hu Yongjian, Gao Yifei, Liu Beibei, et al. DeepFake Videos Detection Based on Image Segmentation with Deep Neural Networks [J]. *Journal of Electronics & Information Technology*, 2021, 43(1): 162-170. (胡永健, 高逸飞, 刘琲贝, 等. 基于图像分割网络的深度假脸视频篡改检测[J]. 电子与信息学报, 2021, 43(1): 162-170.)

[8] Dang H, Liu F, Stehouwer J, et al. On the Detection of Digital Face Manipulation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020.

[9] Agarwal S, Farid H, Gu Y, et al. Protecting World Leaders Against Deep Fakes[C]//IEEE Conference on Computer Vision & Pattern Recognition Workshops, Long Beach, USA, 2019.

[10] Agarwal S, Farid H, El-Gaaly T, et al. Detecting Deep-Fake Videos from Appearance and Behavior [C]// IEEE International Workshop on Information Forensics and Security, New York, USA, 2020.

[11] Li Y Z, Yang X, Sun P, et al. Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020.

[12] Blanz V, Vetter T. A Morphable Model for the Synthesis of 3D Faces [C]//The 26th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, USA, 1999.

[13] Paysan P, Knothe R, Amberg B, et al. A 3D Face Model for Pose and Illumination Invariant Face Recognition[C]//The 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2009.

[14] Cao C, Weng Y L, Zhou S, et al. FaceWarehouse: A 3D Facial Expression Database for Visual Computing[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(3): 413-425.

[15] Deng Y, Yang J, Xu S C, et al. Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019.

[16] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016.

[17] Wang X, Han X, Huang W, et al. Multi-similarity Loss with General Pair Weighting for Deep Metric Learning[C]//Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019.

[18] Deng J K, Guo J, Ververas E, et al. RetinaFace: Single-Shot Multi-level Face Localisation in the Wild[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020.

[19] Chung J S, Nagrani A, Zisserman A. VoxCeleb2: Deep Speaker Recognition [EB/OL]. [2021-04-20]. https://arxiv.org/abs/1806.05622.

[20] Rössler A, Cozzolino D, Verdoliva L, et al. FaceForensics: Learning to Detect Manipulated Facial Images [C]//IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019.

[21] DeepFakes Detection Dataset[EB/OL]. [2021-04-20]. https://github.com/ondyari/FaceForensics.