基于超分辨率重建的强压缩深度伪造视频检测

孙 磊 张洪蒙* 毛秀青 郭 松 胡永进

(战略支援部队信息工程大学 郑州 450001)

摘 要：经典的深度伪造(DeepFake)视频检测方法一般使用卷积神经网络进行检测，但在强压缩深度伪造换脸视频数据集上表现较差，并会对真实数据做出大量误检测。针对这个问题，该文提出一种基于超分辨率重建的强压缩深度伪造视频检测方法。该方法基于深度神经网络检测模型，通过融入超分辨率重建技术，恢复强压缩视频所损失的空间与时间信息，进而提升对强压缩视频的检测准确率。使用FaceForensics++及DFDC数据集进行实验，针对强压缩的深度伪造视频，该方法较ResNet50提高了单帧以及视频的测试准确率，有效缓解强压缩真实视频的误检测问题。

关键词：深度伪造检测；超分辨率重建；强压缩视频；深度学习

中图分类号：TN911.73; TP309.2 **文献标识码**：A **文章编号**：1009-5896(2021)10-2967-09

DOI: 10.11999/JEIT200531

Super-resolution Reconstruction Detection Method for DeepFake Hard Compressed Videos

SUN Lei ZHANG Hongmeng MAO Xiuqing GUO Song HU Yongjin

(*PLA Strategic Support Force Information Engineering University*, *Zhengzhou* 450001, *China*)

Abstract: The forensics methods of DeepFake video generally use convolution neural networks. However, these methods perform poorly on hard compressed DeepFake datasets and make a large number of false detections on real data. To solve the problem above, a method of hard compressed DeepFake video detection based on deep neural network model is proposed, which improves the detection accuracy of hard compressed video by incorporating super-resolution reconstruction technology and recovering the loss of the spatial and temporal information during hard compression. Experiments are performed with the FaceForensics++ Datasets and DFDC (the DeepFake Detection Challenge) Datasets for hard compressed DeepFake video, which improve the test accuracy of single frame and video compared to ResNet50, and effectively alleviate the problem of false detection of real video with hard compression.

Key words: DeepFake detection; Super resolution reconstruction; Video hard compression; Deep learning

1 引言

深度伪造(DeepFake) 是人工智能技术滥用所产生的问题之一[1]，由“深度学习”(Deep learning)和“伪造”(Fake)二词组合而成，专指基于人工智能尤其是深度学习的人体图像合成技术。这一概念最早出现在2017年底，Reddit网站用户“Deepfakes”在网上发布了一段使用FakeAPP合成某一明星的色情视频[2]，引发各界关注。随后，研究者使用卷积神经网络(Convolutional Neural Network, CNN)和生成对抗网络(Generative Adversarial Network, GAN)等深度网络模型提高视频的换脸效果及生成效率[3–5]。伪造音视频能转变演讲人的表情、身份和演讲内容等，达到以假乱真的程度，对数据隐私和社会安全构成严重危害。因此，各种伪造内容的检测和过滤方法也随之出现。腾讯、阿里、谷歌等国内外各人工智能实验室均提出了深度伪造视频图像检测方案，学术界关于深度伪造视频的检测研究也逐渐从传统检测往机器学习算法检测的方向发展。

DeepFake换脸视频在创建时需要通过仿射人脸变换，例如缩放、旋转和剪切，来匹配源视频的人脸各区域特征。由于扭曲变换的人脸区域和周围环境之间的分辨率不一致，该过程会产生伪影特征，可由CNN模型捕捉，例如VGG16[6], ResNet50,

收稿日期：2020-06-30；改回日期：2020-12-31；网络出版：2021-02-02
*通信作者： 张洪蒙 meng19950929@stu.xjtu.edu.cn
基金项目：国家重点研发计划(2017YFB0801900)
Foundation Item: The National Key R&D Program of China (2017YFB0801900)

ResNet101和ResNet152[7]。Zhou等人[8]提出了双流框架检测深度伪造视频，通过GoogleNet学习伪影特征并行人脸分类，使用三重态流来进行伪造人脸检测。Afchar等人[9]提出的MesoNet网络，并对压缩数据进行测试，准确率达到70.47%，但对特征提取未作充分解释。Matern等人[10]从眼睛等区域提取特征向量，并训练一个小型全连接神经网络对真伪图像分类。文献[11]提出一种计算人脸交并比的新方法，并在3个不同的基础分割网络上实现，显著降低了跨库检测的平均错误率。Rössler等人[12]提出了基于卷积神经网络的检测模型，在压缩数据集上的检测率为81.00%。Li等人[13]提出了一种基于伪影的深度伪造检测方法，通过ResNet50分类。该文章为DeepFake视频的检测提供了新思路，但在强压缩视频检测上表现较差，还有进一步改进的空间，这是由于在强压缩后，数据质量降低，在真实数据集中出现伪造数据集中的伪影特征，导致CNN对其误判。

在平常社交网络上接触到的视频由于带宽约束、存储空间限制等因素通常经过压缩后进行传输。视频压缩会引入下采样、模糊和量化噪声等降质技术降低空间冗余和时间冗余[14]，所以在强压缩格式的真实数据集中，视频中会引入类似于伪造视频帧所独有的伪影特征，使得以上检测方法难以广泛应用到复杂压缩质量格式的数据集上。所以，针对社交网络中传播的视频的真伪判别，需要研究对强压缩视频的深度伪造检测方法。

2 基于超分辨率重建的强压缩视频检测方法

本文考虑到强压缩视频的伪造检测必要性，探究了当前检测方法在强压缩视频数据集的可用性；借鉴双网络模型的视频超分辨率方法，并对网络结构进行改进，提高其在人脸视频超分辨率重建的表现；设计卷积神经检测网络，将视频超分辨率重建技术引入到深度伪造检测方法中，提高了真实视频单帧测试的准确率，在强压缩视频测试中整体表现更好。

本文的检测方法整体框架如图1。检测模型分为训练和测试两个部分，在训练时，使用高分辨率图像以及真实图像经过高斯模糊等技术产生低分辨率图像以及模拟伪造图像，分别对应形成字典对对超分辨率重建的神经网络进行训练。测试时，首先使用训练好的动态滤波网络对视频帧采样重建，由残差生成网络补充重建帧的高频细节，然后叠加成最终的重建帧，最后输入到训练好的卷积神经网络中进行检测并输出判别视频为伪造的概率。

2.1 视频超分辨率重建

视频的超分辨率重建技术指的是将给定的质量不高且分辨率不高的视频，通过信号及图像改善的方法来提升其品质[15]。文献[16]提出双网络模型进行图像超分辨率重建，本文在其神经网络上改进使其适用于人脸视频超分辨率重建场景，降低真实数据集中的伪影特征，又使伪造视频中的伪影特征不被重建消失。本文隐式地利用运动信息来生成动态

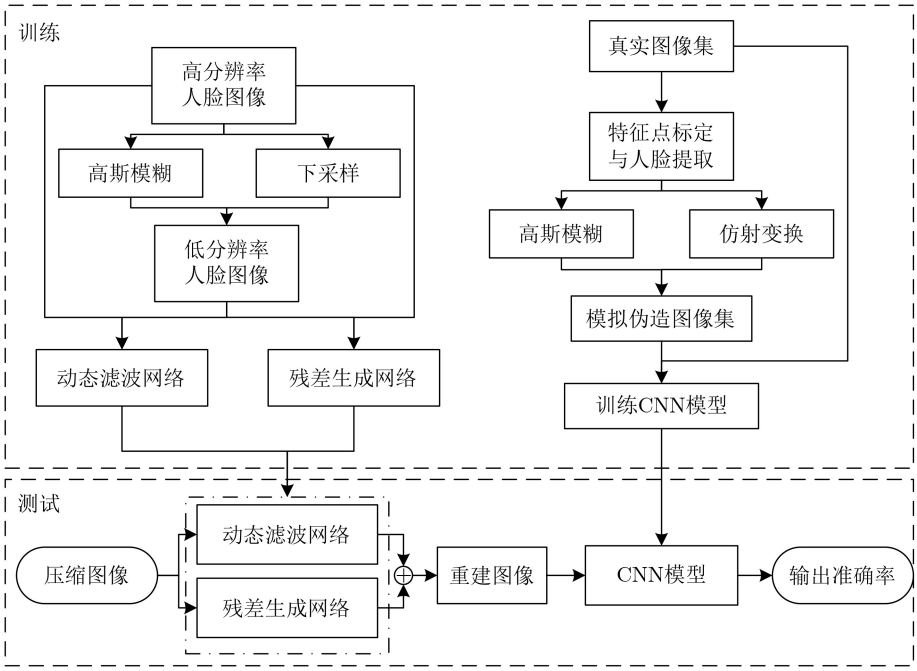


图 1 本文检测模型整体框架

上采样滤波器，利用生成的上采样滤波器，可以通过对输入中心帧进行局部滤波来直接构建高分辨率(High Resolution, HR)帧，并使用残差网络补充高频细节，可以生成时间上一致的更清晰的HR视频。本文进行视频超分辨率重建的框架结构如图2所示。

动态滤波网络由滤波生成单元和动态滤波层组成。滤波生成网络的每个单元由ReLU激活函数以及3D卷积组成，根据输入的视频帧动态生成滤波器，并将其输入至动态滤波层，该层采用生成的滤波器对输入进行动态卷积等操作，本方法将其应用于低分辨率帧低频重建。本方法所采用的动态滤波网络及残差网络中均包含了11个生成单元共22个卷积层，2个上采样卷积层以及1个全连接层。在滤波生成单元后，卷积层采用所生成的动态滤波器，reshape之后结合中心输入帧x_t经过softmax层得到高清帧$\tilde{y}_t$。残差网络用于生成具有高频细节的残差图像，由多个输入帧构成，并且使用动态滤波网络构建的高清帧作为基线，与计算出的残差相加，即将残差r_t添加到上采样结果$\tilde{y}_t$，得到最终的高分辨率输出帧$\hat{y}_t$。其具体算法见算法1。

算法1 基于滤波器与残差生成网络的视频超分辨率重建

输入：低分辨率视频帧x_t(t为视频总帧数)

输出：高分辨率视频帧$\hat{y}_t$

(1) 输入低分辨率视频帧x_t，并将初始低分辨率视频帧x_t划分成采样块；

(2) 通过滤波器生成网络和残差生成网络分别对低分辨率视频帧x_t的低频分量和高频分量进行处理；

(3) 取块$b_i(b_i \subset x_t)$，利用下式，通过滤波器$f_t^{(y,x,v,u)}$构建低频域的高清块$\tilde{b}_l$；

$$\tilde{y}_t(yr+v, xr+u) = \sum_{j=-n}^{n}\sum_{i=-n}^{n} f_t^{(y,x,v,u)} \cdot (j+n, i+n)x_t(y+j, x+i)$$

其中，y和x是低分辨率网格中的坐标，v和u是每个$r \times r$采样块$(0 \le v,u \le r-1)$中的坐标，t表示时间轴，n是所选取序列帧数。

(4) 取块$b_i(b_i \subset x_t)$，通过残差生成网络去除伪高频成分，获得有效高频残差块；

(5) 重复步骤3和步骤4，直至完成所有视频帧块的操作，得到高清帧$\tilde{y}_t$和残差帧r_t；

(6) 使用滤波器生成网络构建的高清帧作为基线，与生成的残差帧r_t叠加得到最终的高分辨率视频帧$\hat{y}_t$；

(7) 重复步骤2—步骤6，直至完成所有视频帧的操作，实现对低分辨率视频的超分辨率重建。

在网络的构建中，动态滤波网络和残差生成网络共享大部分权重以减少开销，其中网络的参数共享是受密集块启发而设计的，并针对人脸图像超分辨率重建问题进行了适当修改。因为3D卷积层比2D卷积层更适用于人脸动作识别和视频数据的时域空域特征提取，故原方法采用2D卷积层从视频数据中学习时空特征。据文献[17]所述，人脸超分辨率重建属于低级视觉特征的捕捉，网络输入和输出的空间分布十分相似，而神经网络中的批量归一化(Batch Normalization, BN)层白化中间特征的方式破坏了原始空间的表征，因此在重建模型中需要

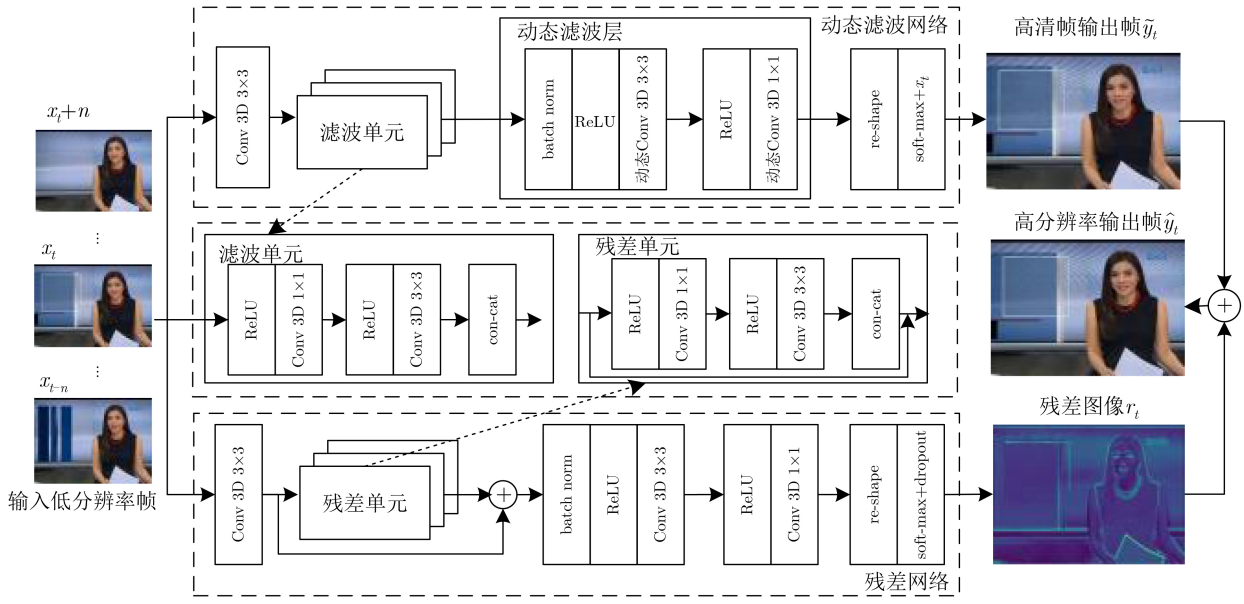


图 2 视频超分辨率重建网络结构

部分层或参数来恢复这种表征，所以同等级的参数量，含有BN层的模型会降低人脸超分辨率重建的效果。因此，本方法首先移除动态滤波网络中的BN层。另外，考虑到残差生成网络用于捕捉重建帧的高频细节，需要模型有较高的泛化能力，同时保持两个网络参数共享以提高网络收敛速度，本方法将残差生成网络中的BN层移除后，在最后增加dropout层解决归一化问题，提高模型拟合能力。

2.2 基于卷积神经网络的强压缩深度伪造视频检测

本文检测算法在ResNet50模型[6]的基础上进行改进构成端到端的神经网络。在训练过程中，使用负样本生成方法，使模型对人脸区域的特征提取更敏感，提高分类准确率。

2.2.1 负样本生成算法

为了增强CNN捕捉伪影特征的能力并简化训练过程，本文通过模拟DeepFake中的仿射翘曲变换生成的分辨率不一致性对负样本数据进行预处理。

首先，如图3(f)所示，创建RoI(Region of Interest)区域，即矩形区域减掉根据眼睛与嘴巴底部的特征点创建的凸多边形所形成的区域。具体来说，使用人脸特征点的坐标来确定RoI，例如$[y_0-\hat{y}_0, x_0-\hat{x}_0, y_1-\hat{y}_1, x_1-\hat{x}_1]$，其中$y_0$, x_0, y_1, x_1表示可以覆盖所有脸部特征点坐标的最小边界框。变量$\hat{y}_0$, $\hat{x}_0$, $\hat{y}_1$, $\hat{x}_1$是$[0, h/5]$和$[0, w/8]$之间的随机值，其中h, w分别是矩形人脸的高度和宽度。其伪代码如表1所示。

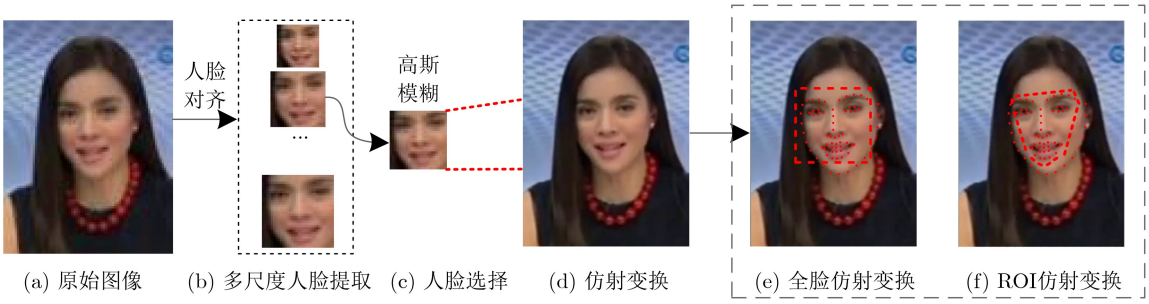


图 3 负样本生成及选择RoI区域过程

表 1 负样本生成的伪代码

输入：图像路径path，图像标签L，真实图像i_r
输出：伪造图像i_f
参数：随机数r_1, r_2，转换矩阵m_t，特征点坐标p_{68}。

```
(1)  begin
(2)    for i_r in path:
(3)      i_r=dlib.align(i_r) //人脸对齐
(4)        if L= 1:
(5)          if r_1 < 0.5
(6)            face = cv2.warpAffine(i_r, m_t* size, (size,
               size)) //仿射变换
(7)            face = cv2.GaussianBlur(face, (5, 5)) //高斯
               模糊
(8)              if r_2 < 0.5
(9)                part_mask = dlib.mask(i_r, p_68) //特征
                   点标定
(10)             i_f =i_r * (1 - part_mask) +i_f *
                 part_mask
(11)               i_r = i_f
(12)             L = 0
(13)     else:
(14)       continue
(15)   return i_r
(16) end
```

2.2.2 卷积神经网络结构

ResNet50模型中包含了49个卷积层1个全连接层。其中，第2至第5阶段中的id block表示的是恒等残差块，即不改变输入输出图像的尺寸，conv block代表的是添加尺度的卷积残差块，每个残差块包含3个卷积层，结构如图4所示。

图4中的阶段1到阶段5表示残差块，conv 是卷积操作的卷积层，batch norm是批量正则化处理，激活函数使用ReLU函数，maxpool表示最大池化操作，avgpool表示全局平均池化层操作。其具体结构参数如表2所示。经过残差块的连续卷积运算后，得到的特征图矩阵的通道数量越来越深，然后通过flatten图层将特征图的的大小更改为 2048×2048。最后一个卷积层的特征图矩阵输入到全连接层fc，对图像检测的概率由 softmax分类器输出。

考虑到本文训练集采用的是CelebA数据集，而测试集则是FaceForensics++数据集，两个数据集存在分布不一致的可能性，故本文在softmax层加入tanh函数，在权值矩阵和偏置量都没有改变的情况下，通过改变激活函数，将原模型的决策边界进行了软化，输入层和隐藏层之间的权值矩阵将不会再局限于稀疏矩阵，而可以是任意矩阵，有效防止模型的过拟合问题。

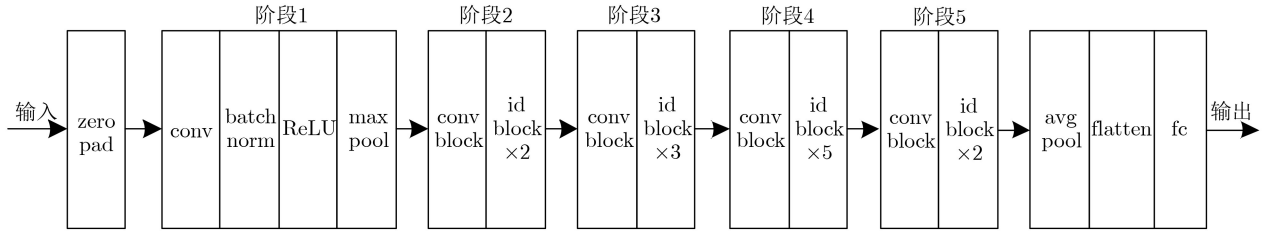


图 4 卷积神经网络结构

表 2　改进的ResNet50结构参数

网络层	conv_1	conv_2	conv_3	conv_4	conv_5	fc
输出大小	112×112	56×56	28×28	14×14	7×7	1×1
改进后的 ResNet50	7×7, 64, stride2	3×3 maxpool,stride2 $\begin{bmatrix}1\times1, & 64\\ 3\times3, & 64\\ 1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\\ 1\times1, & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\\ 1\times1, & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\end{bmatrix}\times3$	average pool, softmax+tanh

3　分析与讨论

3.1　数据集介绍

FaceForensics++数据集由Rössler等人[12]所制作的FaceForensics数据集[18]扩充而来，后又收录谷歌与Jigsaw 联合制作的大型DeepFake数据集[19]，进一步丰富原有伪造与真实视频。目前，FaceForensics++作为标准数据集已广泛使用在深度伪造检测模型的训练和测试中。Rössler等人[12]根据伪造方法的不同将FaceForensics++数据集划分为DF(DeepFake), F2F(Face2Face), FS(FaceSwap)和NT(NeuralTextures)4个子数据集，每个子数据集包含1000个视频，视频具有3种压缩格式，分别为无损压缩、轻压缩、强压缩，总共包含超过180万张伪造图像。

DFDC (the DeepFake Detection Challenge)为Kaggle举办的DeepFake检测挑战赛的正式数据集，共有119196个视频，由演员拍摄，质量较FaceForensics++数据集高。

香港中文大学Liu等人[20]创建了CelebA数据集，即名人人脸属性数据集。该数据集包含10177个名人的图像(人均约有20张)，共计202599张。

3.2　超参数设置

为了提高训练的人脸数据的多样性，本文改变了所有训练样本的图像信息：亮度、对比度、失真度和锐度等，并采用动态的方式来产生负样本。对于每个训练批次，随机选择一半正样本按照前文预处理的方法将其转化为模拟伪造数据，使得训练数据更加多样化。由于改进后的ResNet50 神经网络输入数据大小为224×224×3，所以需要在输入数据之前进行图像预处理，把数据裁剪成指定大小。批训练大小设置为64，学习速率从0.001开始，每1000步衰减95%，并使用随机梯度下降优化方法，训练过程在第20个迭代周期终止，使用困难样本挖掘策略[21]对模型进行微调，最终使用训练好的参数模型对数据进行测试。

4　实验结果分析

4.1　实验设置

本文选择FaceForensics++数据集来评估所提出的检测方法。其中，训练集采用Celeb A数据集；测试集中的伪造人脸数据集为FaceForensics++数据集中DeepFake视频库，真实数据集为对应的真实人脸视频库。实验环境采用的是64位Windows 10 操作系统下的TensorFlow深度学习平台。

4.2　实验结果

为了验证本文算法对视频及视频帧的有效性，分别针对视频及视频帧进行伪造检测，具体步骤是：随机选择FaceForensics++中的真伪视频拆分成帧运行检测算法并分析实验结果，同时测试所选视频并和现有方法的检测结果进行比较和量化分析。本文对比实验部分所参考的评价参数为测试准确率及受试者操作特征(Receiver Operating Characteristic, ROC)曲线，如式(1)及式(2)所示

$$\text{真正类率}=\frac{\text{真正类}}{\text{真正类}+\text{假负类}} \tag{1}$$

$$\text{假正类率}=\frac{\text{假正类}}{\text{假正类}+\text{真负类}} \tag{2}$$

其中，ROC曲线的横轴为假正类率(False Postive Rate, FPR)，代表检测模型预测的真实视频中实际伪造视频占所有伪造视频的比例；纵轴为真正类率(True Postive Rate, TPR)，代表检测模型预测的真实视频中实际真实视频占所有真实视频的比

例。AUC(Area Under roc Curve)是一种用来度量分类模型好坏的标准，其值就是处于ROC 曲线下方的那部分面积的大小。

4.2.1 视频帧的检测效果对比

本节首先验证所提方法在强压缩数据集上对真实视频帧检测准确率的提高，随机选择FaceForensics++中强压缩格式的真实与相对应的伪造的视频进行测试，将视频拆分成共396帧进行测试并输出其为伪造视频帧的概率，对比方法为文献[13]中未重建的ResNet50检测方法。

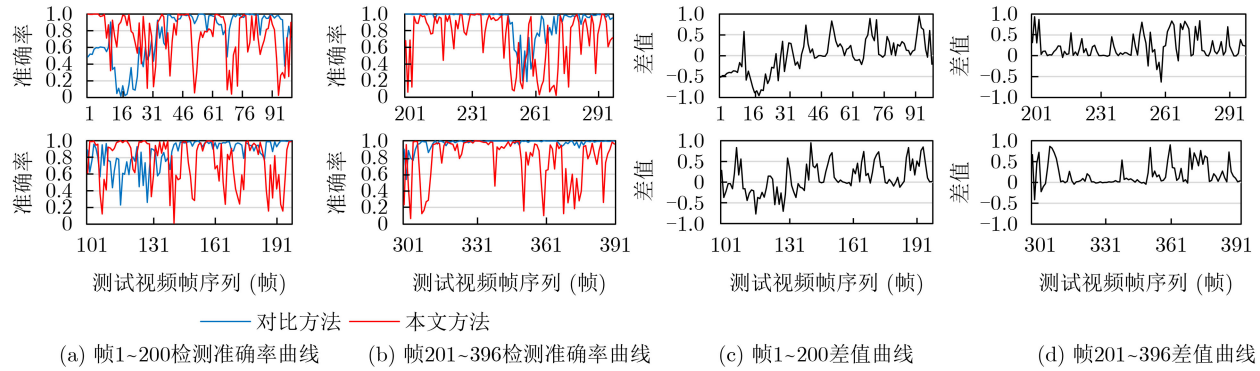
图5(a)及图6(b)为未重建方法与本文方法对强压缩真实视频每一帧检测的准确率曲线，纵轴表示为视频帧检测为伪造的概率，可以看出未重建方法对真实视频帧的误判较为严重。图5(c)及 图5(d)中曲线含义为未重建方法与本文方法所得出的准确率的差值，从图中可以看出差值主要集中在0～0.5之间，证明在对视频进行重建之后，本文方法判断为伪造视频的概率明显降低，即对真实视频帧的检测准确率有所提高。

图6(a)及 图6(b)为使用未重建方法与本文方法对强压缩伪造视频每一帧检测的准确率曲线，图6(c)及图6(d)中曲线含义为未重建方法与本文方法所得出的准确率的差值，从图中可以看出差值主要在–0.1～0.3之间波动，说明本文方法对伪造视频的检测准确率的影响较小。但结合图5中对真实视频帧的检测结果，表明加入超分辨率重建提高了深度伪造视频检测方法在视频帧上的检测准确率。

4.2.2 视频的检测效果对比

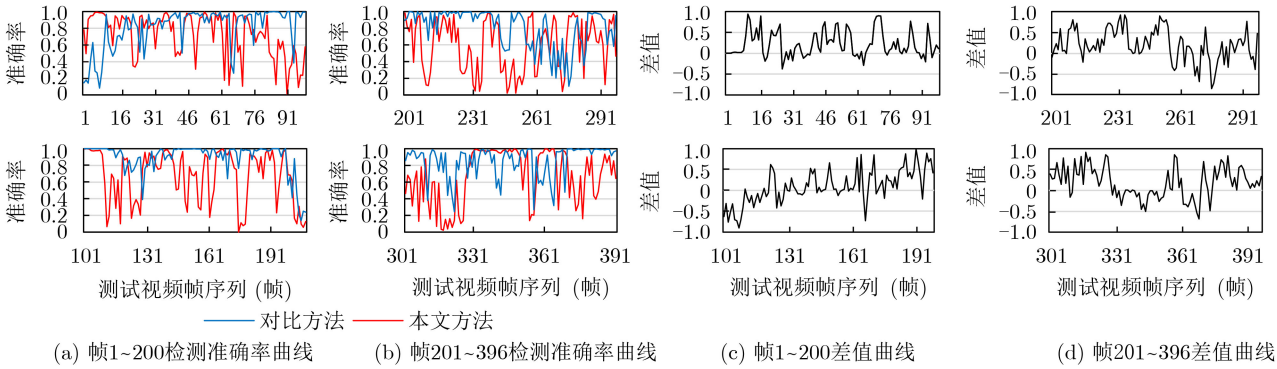
本节验证所提方法在强压缩数据集上对真实视频检测准确率的提高，在FaceForensics++数据集中DeepFake视频库随机选择100个强压缩格式的真实与相对应的伪造的视频进行测试。

图7(a)分别为未重建方法以及超分辨率重建的方法在强压缩真实视频中的检测准确率曲线，横轴为视频编号，纵轴为检测为伪造视频的概率。可以看出，在后段数据中，未重建方法对视频的伪造检测概率上下波动，而在重建之后对视频可以进行精确的判别，前段数据中方法对强压缩的原始视频出现误判，在重建之后提升了其判别概率。在对视频内容统计研究后发现，比如图5(a)中编号14视频，未重建和重建后结果相似，均对视频造成误判，类似视频还有23, 28等。此类视频中人物存在大量夸张表情，同时在强压缩技术的影响下，导致原方法对其误判；而经历超分辨率重建后，虽然对视频消除了部分伪影，但视频中人物表情未有改变，故而使得超分重建方法所起作用较小。而编号20, 24之类的视频，人物距离摄像机较远，从而导致提取到的人脸分辨率更低，在Resize至检测网络中时，相



(a) 帧1~200检测准确率曲线 (b) 帧201~396检测准确率曲线 (c) 帧1~200差值曲线 (d) 帧201~396差值曲线

图 5 两种方法下的真实视频帧检测准确率对比及差值曲线



(a) 帧1~200检测准确率曲线 (b) 帧201~396检测准确率曲线 (c) 帧1~200差值曲线 (d) 帧201~396差值曲线

图 6 两种方法下的伪造视频帧检测准确率对比及差值曲线

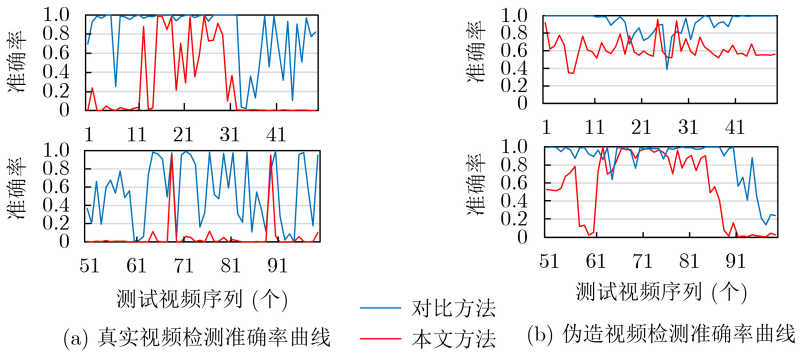


图 7 两种方法下的视频检测准确率对比

比其他网络特征更少，因此检测准确率降低；此类视频在经过超分重建后均有较为明显的提升。

图7(b)分别为未重建方法以及超分辨率重建的方法在强压缩伪造视频中的检测准确率曲线，本文方法对伪造视频的检测概率有所影响，这是由于伪造图像经过超分重建后，平滑了人脸与背景的拼接区域，部分消除了换脸算法所产生的伪影，因而一定程度上降低了模型检测准确率。

本文基于AUC度量对在FaceForensics++数据集中DeepFake视频库随机选择100个强压缩格式的真实与相对应的伪造的视频评估本文检测方法，其中未重建方法在强压缩视频集上的AUC值为63.36%，重建之后为94.86%，见图8。可以看到矩形框选中的区域，此处原始模型呈反向判别的趋势，由纵横坐标可以看出，此处FPR值较高，而相对应的TPR值较低，由FPR和TPR的定义可知，在此处样本点作为阈值的条件下，真实视频中判断正确所占比例很低，即此处表示原始模型对真实视频误判的情况。结果表明，基于超分辨率重建的方法提高了对强压缩视频的检测准确率，可以较好地解决真实数据出现误判的问题。

本文方法及2种检测算法(文献[9]、文献[13])各自在FaceForensics++数据集中4个数据库以及DF-DC公开测试集上的检测结果的AUC值如表3所示，并在MesoNet网络添加超分辨率网络形成

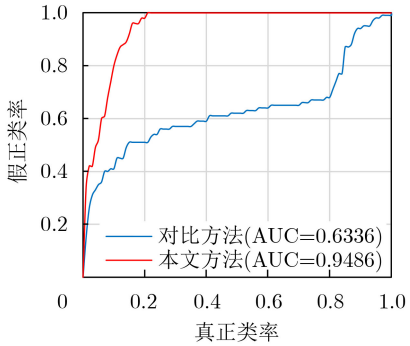


图 8 两种方法下的ROC曲线对比

表 3　各算法强压缩数据集检测结果对比

AUC	DF	F2F	FS	NT	DFDC
MesoNet[10]	81.27	62.20	66.27	56.47	63.51
VSR-MesoNet	81.62	**63.71**	63.84	**58.65**	65.34
ResNet50[13]	63.36	57.48	60.12	51.96	58.37
本文	**94.86**	58.31	**70.62**	57.23	**71.88**

VSR-MesoNet模型，同其余模型在原有测试集上进行测试。从表3可以看出，本文方法的检测效果在AUC值上明显优于文献[13]未重建的对比方法，在DF和FS以及DFDC数据集上优于MesoNet和VSR-MesoNet。但由于F2F和NT所生成视频是基于面部重现的原理，而本文方法基于图像分割掩膜技术所提取拼接特征来进行检测，故在F2F和NT数据集上表现较差。加入超分重建网络的MesoNet即VSR-MesoNet和原始MesoNet相比，在各大数据集上表现各有优劣，其中在F2F及NT数据集上有相对明显的提升，但在DF及DFDC数据集上提升不明显，并在FS数据集上有所下降。按照提出该网络的文献[9]所述，MesoNet网络激活的特征主要为眼睛区域的细节特征，其次是鼻子区域，并非依靠伪影对真伪视频进行分类，而本文网络针对的是去除拼接区域的伪影，故而将超分辨率重建应用在MesoNet网络提升有限。

5　结束语

本文提出一种基于超分辨率重建的强压缩深度伪造视频的检测方法。首先，在对真实数据的人脸区域进行对齐后，施加高斯模糊对人脸拼接区域进行处理，将其转化为负样本，减少负样本生成工作；然后将真实数据和处理数据输入到神经网络中进行训练，由于先前数据的预处理，神经网络对于人脸边缘伪影特征的提取更加敏感，有效提高神经网络的收敛速度；再使用神经网络对测试数据进行超分辨率重建，神经网络采用残差网络在超分辨率重建中对视频帧像素点间差值进行预测，增加重建

细节；最后，利用训练好的深度网络模型对超分辨率重建后的视频进行测试。实验证明，针对强压缩深度伪造视频，所提方法对真实视频的误检率降低，并在视频单帧的测试中精确度更高。另外对无损压缩数据的检测准确率相较于文献[13]方法有所降低，并且在除DF数据集的其他子数据集表现较差，这也是后续需要重点研究的工作。

参 考 文 献

[1] 陈宇飞, 沈超, 王骞, 等. 人工智能系统安全与隐私风险[J]. 计算机研究与发展, 2019, 56(10): 2135–2150. doi: 10.7544/issn1000-1239.2019.20190415.
CHEN Yufei, SHEN Chao, WANG Qian, *et al*. Security and privacy risks in artificial intelligence systems[J]. *Journal of Computer Research and Development*, 2019, 56(10): 2135–2150. doi: 10.7544/issn1000-1239.2019.20190415.

[2] Faceswap: Deepfakes software for all[EB/OL]. https://github.com/deepfakes/faceswap, 2018.

[3] KORSHUNOVA I, SHI Wenzhe, DAMBRE J, *et al*. Fast face-swap using convolutional neural networks[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3697–3705. doi: 10.1109/ICCV.2017.397.

[4] Faceswap-GAN[EB/OL]. https://github.com/shaoanlu/faceswap-GAN, 2019.

[5] Keras-VGGFace: VGGFace implementation with Keras framework[EB/OL]. https://github.com/rcmalli/keras-vggface, 2019.

[6] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. The 3rd International Conference on Learning Representations, San Diego, USA, 2015.

[7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[8] ZHOU Peng, HAN Xintong, MORARIU V I, *et al*. Two-Stream neural networks for tampered face detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 1831–1839. doi: 10.1109/CVPRW.2017.229.

[9] AFCHAR D, NOZICK V, YAMAGISHI J, *et al*. MesoNet: A compact facial video forgery detection network[C]. 2018 IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 2018: 1–7. doi: 10.1109/WIFS.2018.8630761.

[10] MATERN F, RIESS C, and STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations[C]. 2019 IEEE Winter Applications of Computer Vision Workshops, Waikoloa Village, USA, 2019: 83–92. doi: 10.1109/WACVW.2019.00020.

[11] 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. doi: 10.11999/JEIT200077.
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. doi: 10.11999/JEIT200077.

[12] RÖSSLER A, COZZOLINO D, VERDOLIVA L, *et al*. Faceforensics++: Learning to detect manipulated facial images[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 1–11. doi: 10.1109/iccv.2019.00009.

[13] LI Yuezun and LYU Siwei. Exposing deepFake videos by detecting face warping artifacts[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 46–52.

[14] SCHWARZ H, MARPE D, and WIEGAND T. Overview of the scalable video coding extension of the H. 264/AVC standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, 17(9): 1103–1120. doi: 10.1109/TCSVT.2007.905532.

[15] PARK S C, PARK M K, and KANG M G. Super-resolution image reconstruction: a technical overview[J]. *IEEE Signal Processing Magazine*, 2003, 20(3): 21–36. doi: 10.1109/MSP.2003.1203207.

[16] JO Y, Oh S W, KANG J, *et al*. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 3224–3232. doi: 10.1109/CVPR.2018.00340.

[17] NAH S, KIM T H, and LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 257–265. doi: 10.1109/CVPR.2017.35.

[18] RÖSSLER A, COZZOLINO D, VERDOLIVA L, *et al*. FaceForensics: A large-scale video dataset for forgery detection in human faces[EB/OL]. https://arxiv.org/abs/1803.09179, 2018.

[19] NICHOLAS D, ANDREW G, PER K, *et al*. Deepfakes detection dataset by Google & jigsaw[EB/OL]. 2019. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html.

[20] LIU Ziwei, LUO Ping, WANG Xiaogang, *et al*. Deep learning face attributes in the wild[C]. 2015 IEEE International Conference on Computer Vision, Santiago, USA, 2015: 3730–3738. doi: 10.1109/ICCV.2015.425.

[21] SHRIVASTAVA A, GUPTA A, and GIRSHICK R. Training Region-Based object detectors with online hard example mining[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 761–769. doi: 10.1109/CVPR.2016.89.

孙　磊：男，1973年生，教授，主要研究方向为密码与系统安全、机器学习安全.

张洪蒙：男，1995年生，硕士生，研究方向为计算机视觉.

毛秀青：男，1980年生，副教授，主要研究方向为智能信息系统安全.

郭　松：男，1985年生，讲师，主要研究方向为计算机视觉.

胡永进：男，1981年生，讲师，主要研究方向为网络信息防御.

责任编辑：马秀强