

## 基于改进高斯滤波网络的深度伪造检测方法

瞿远近<sup>a</sup>, 吴起<sup>b</sup>

安徽理工大学 a. 计算机科学与工程学院; b. 人工智能学院, 安徽淮南 232001

**摘要:**针对虚假视频的检测中,特征噪声多、数据量大和检测准确率低的问题,提出一种改进高斯滤波对地标点进行去噪和改进深度网络模型来提升精度的算法(IGFNet);算法将高斯滤波扩展到空域和值域上,在滤除低频噪声的同时尽可能保留高频噪声,使得在后续特征数据的处理中,标记点的精度得以提升;采用特征化的数据代替原始数据以减少数据量,使得送入网络的数据量大大缩小,有效缩短了训练时长和减少了网络参数量,在实际应用中可以增加鉴别虚假视频的效率;并且针对人脸属性特征点的差异采用不同深度的双流神经网络,从而更加有效地学习如何鉴别虚假视频。实验表明:改进高斯滤波网络算法(IGFNet)有效地增加了真假脸检测的准确率,在与当前较为优秀的方法如 Meso4, Xception, LRNet 等的对比下,IGFNet 的准确率均有不同程度的提升,在跨数据集的测试上提升尤为明显;在压缩过的虚假视频测试中,IGFNet 的泛化下降程度最低,显示了较强的鲁棒性;通过改进网络加入梯度热力图以直观判断出 IGFNet 对于深度伪造图片的鉴别能力。

**关键词:**深度学习;深度图像伪造;改进高斯滤波;循环神经网络

**中图分类号:**TP391.41;TP183 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2023.0004.006

### Depth Forgery Detection Method Based on Improved Gaussian Filter Network

QU Yuanjin<sup>a</sup>, WU Qi<sup>b</sup>

a. School of Computer Science and Engineering; b. School of Artificial Intelligence, Anhui University of Science and Technology, Anhui Huainan 232001, China

**Abstract:** Aiming at the problems of excessive feature noise, a large amount of data and low detection accuracy in the detection of fake videos, an algorithm was proposed to improve the Gaussian filter to denoise the landmark points and improve the deep network model, so as to improve the accuracy. The Gaussian filter is extended to the spatial domain and the value domain, and the high-frequency noise is retained as much as possible while filtering out the low-frequency noise so that the accuracy of the marker points can be improved in the subsequent processing of the feature data. The characteristic data is used instead of the original data to reduce the amount of data, which greatly reduces the amount of data sent to the network, effectively shortens the training time and reduces the number of network parameters, and can quickly identify fake videos in practical applications. In addition, according to the difference of face attribute feature points, dual-flow neural networks with different depths are used to learn how to identify fake videos more effectively. The experimental results show that the improved Gaussian Filter network algorithm (IGFNet) effectively increases the accuracy of real and fake face detection. Compared with the current excellent methods such as Meso4, Xception and LRNet, the accuracy of IGFNet has been improved to different degrees, especially in the test across datasets. In the test of compressed fake videos, IGFNet has the lowest degree of generalization degradation, showing strong robustness. The ability of IGFNet

**收稿日期:**2022-05-12 **修回日期:**2022-07-18 **文章编号:**1672-058X(2023)04-0041-07

**基金项目:**安徽省自然科学基金项目(2008085MF220);安徽省高校自然科学基金项目(KJ2019A0109);安徽省重大科技专项基金项目(18030901025);安徽理工大学研究生创新基金项目(2021CX2102)。

**作者简介:**瞿远近(1996—),男,安徽芜湖人,硕士研究生,从事计算机视觉研究。

**引用格式:**瞿远近,吴起.基于改进高斯滤波网络的深度伪造检测方法[J].重庆工商大学学报(自然科学版),2023,40(4):41—47.

QU Yuanjin, WU Qi. Depth forgery detection method based on improved Gaussian filter network [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(4): 41—47.

to discriminate deep forgery images is intuitively judged by improving the network and adding the gradient heat map.

**Keywords:** deep learning; deep image forgery; improved Gaussian filtering; recurrent neural networks

## 1 引言

深度伪造一词出自于深度学习和伪造的组合,是采用深度学习的方法对视频或者图片中的人脸进行替换或者编辑的一种方法。早在 2017 年,DeepFake 算法的出现就引发了热议,而在之后的 2018 年,BuzzFeed 发表了一段巴拉克奥巴马的造假演讲视频,这毫无疑问带来了极大的安全问题。当不受控制的虚假视频出现在公众视野中时,无疑会混淆视听,造成信息混乱,甚至影响国家形象与安全,而一系列相关的便利应用如 FakeApp,FaceSwap,Zao 等的出现又大大拉低了这一技术的使用难度。这使得对深度伪造视频的检测变得极为迫切。

文献[1]将空洞卷积和注意力机制用于深度伪造检测,提高了检测的准确性,在不同人脸的截取比例下表现较好,但是对噪声较多的图片检测时表现不佳。而 Face X-ray<sup>[2]</sup>则通过卷积神经网络比较图片内是否存在不同的噪声,来判断是否为换脸图像,并且通过这些噪声的差异来画出融合的边缘区域,当伪造者对伪造的图像进行二次加噪的时候,就较为容易躲过方法的检测,并且立体的面部信息采用二维卷积难以捕获完全,所以文献[3]提出 3D 卷积神经网络来对深度伪造视频进行训练,以加强检测准确性。这些方法一定程度上提高了检测准确率,但是在含噪声较多的图片上变得难以有效处理,而且这些方法大都需要增加网络参数,以至增大训练复杂度。

面对训练复杂度增大的情况,SSTNET 算法<sup>[4]</sup>简化 Xception 流,并且使用长短时记忆网络来分析时空特征进而检测被篡改的面孔,这样的做法大大减小了网络参数量,但是检测精度不够。将颜色域和频率域结合起来可以提高检测精度,为了实现这一目标,Two-Branch 算法<sup>[5]</sup>采用不同尺度下的拉普拉斯高斯算子,提高了模型的泛化性能,但是这种方法反而又增大了网络结构。应对这个问题,LRNet 算法<sup>[6]</sup>利用 Landmark 算法的预测和校准进行特征增强,并且采用循环神经网络作为主干网络,做到了在已经校准数据下的高精度检测的效果,但是在跨域数据集上表现不佳。文献[7]使用循环神经网络将卷积长短时神经网络引入深度伪造的检测,从而进一步加强了伪造视频检测的精度,但这方法也存在泛化性差的问题。

这些检测方法大都对特定问题提出了相对应的解决方案,但这些方法在去噪检测、较小训练数据量和提高准确率上难以同时满足。在以上学者工作的基础上做了进一步改进,通过文本化图片信息,用地标点数据代替原始数据减少了数据量和计算量;通过将改进高斯滤波应用到文本化的信息达到了良好的数据去噪效果;采用双流循环神经网络对去噪后的文本信息进行训练进一步提升了深度伪造视频的检测精度。

## 2 技术与危害

### 2.1 基本伪造理论框架

对于深度伪造视频检测方法的研究,需要先探究理论与方法。Deepfake 生成伪造视频的框架通常由一个编码器和两个解码器组成。针对目标人脸和原人脸采用不同的解码器,而这两个解码器共享一个编码器。不同的编码器解码器的组合用于不同的人脸重构,共享的公共编码器和其中一个解码器可以将目标人脸替换到该解码器的所对应的人脸上,整体网络结构如图 1 所示。



图 1 换脸方法

Fig. 1 Face-swapping method

### 2.2 伪造方法与检测重要性

目前的深度伪造大致有假脸生成、面部编辑、人脸替换 3 种主要方法。假脸生成采用生成对抗网络的方法,通过无监督学习原有数据的分布以生成出没有出现过的人脸,由于这种方法生成的图像现实中不存在,所以在安全方面的威胁比其他 2 种要小;面部编辑更多倾向于将面部空间进行时域上的转换,在隐空间内将模型的输入参数进行调整,以期待发现更多的连续

特征;人脸替换指将人脸替换到原有视频或者图片的目标人脸上,这种方法现在发展较为成熟,方法也较多,这类方法产生的伪造视频常具有迷惑性。另一种替换方案替换的是表情,与人脸整个替换不同的是,只将指定的人脸表情进行了替换,从而达到伪造情绪的作用。这一技术搭配声音和动作的迁移,可以产生以假乱真的名人政客的演讲发言视频等,造成极大的社会不良影响,这让深度伪造视频的检测成为迫在眉睫的问题。

### 3 模型设计

所提出的深度伪造检测算法(IGFNet)框架由预处理、去噪、特征嵌入和神经网络分类等部分组成。在预处理时,为了更好地处理数据,且保留连续帧特性,对视频不进行采样取帧,而是截取所有帧,以获得更全面的信息。

在人脸预处理时,主要有人脸检测、标志点检测与标志点对齐 3 个方面。这部分首先将提取出来的每一帧执行检测程序,从中找出感兴趣区域并裁剪,如图 2 所示。将这些裁剪出来的人脸进行标志点检测,特征提取采用 Landmark 检测器,一个人脸需要检测 68 个标志点,由这些标志点可以勾勒出明显的人脸属性轮廓。将这些标记点对齐到预设位置后就完成了人脸预处理部分。Landmark 检测器在单帧上提取的数据精度可以满足需求,但是在连续帧中,人脸各部分边缘的动态模糊,导致对齐的标记点在时序上产生晃动,这些噪声导致训练精度难以提高,还需要对数据进行去噪。现采用改进的高斯滤波方法对此过程进行改进,从而产生稳定可靠的标志点序列帧。

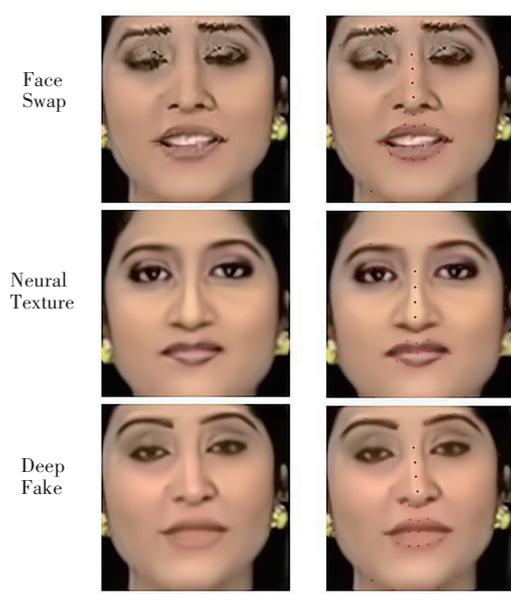
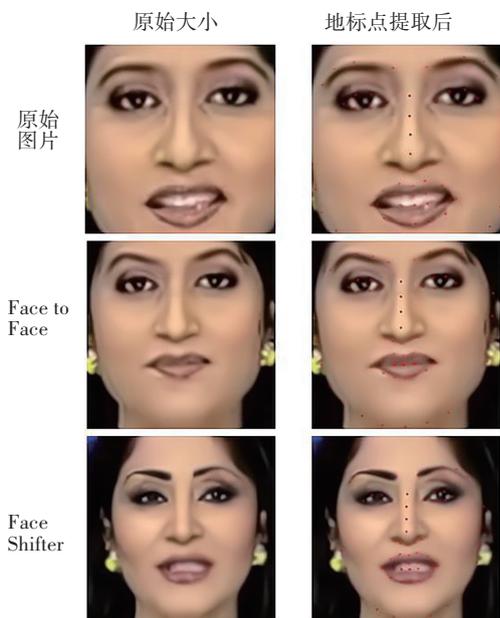


图 2 对 5 种不同换脸方法的人脸 Landmark 标定

Fig.2 Landmark calibration of the face for five different methods

#### 3.1 高斯滤波

在对人脸地标点特征数据提取的过程中,通常会引入正态分布噪声,这种噪声又称高斯噪声,这种噪声的出现让深度检测模型的精度难以进一步提升。为了消除这种噪声所带来的影响,一般用高斯滤波的方式对这类数据进行去噪。高斯滤波可以对提取出来的面部特征点进行平滑处理。滤波对于每个图片帧的特征点集的邻域进行采样,而后做平滑处理,这是一种线性平滑滤波,可以很大程度上保留原有数据特征的分布。一般的二元高斯分布如式(1)所示:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

其中,σ 是正态分布的标准偏差,该偏差决定了高斯分布的宽窄。

高斯滤波的计算方式,在实际运算中,直接计算是复杂的,利用傅里叶变换可以将这一复杂的计算简化,因为高斯函数的傅里叶变换具有函数性质不变形的特点。傅里叶变换可以将任意函数分解成正弦函数的线性表示其一般公式如式(2)所示:

$$F(\omega) = \int f(t) e^{-i\omega t} dt \quad (2)$$

其中,f(t)是周期函数,其周期为 t,而 F(ω)是 f(t)的象函数。这种方式可以满足大部分高斯噪声的滤除,但是这种方法难以消除长序列帧下人脸地标点对应的低频信息,而这些信息是影响模型精度的因子之一,所以针对问题,对去噪模块进行了改进,引入了改进高斯方法。

#### 3.2 改进高斯滤波

改进的滤除方法在滤除低频噪声的同时尽可能保留高频噪声。高斯滤波对单位信号噪声去除力度不大,而改进的高斯滤波能弥补这一缺点。改进的高斯滤波方

法,采用了高斯核对目标进行去噪,同时在式(1)中引入了高斯方差。对高斯滤波的具体改进点:受到  $\alpha$  滤波的启发,将仅为空域上定义的高斯滤波拓展到空域和值域上。将高斯滤波的指数部分单独来看,可得式(3):

$$t(x, y, a, b) = \exp\left(-\frac{(x-a)^2 + (y-b)^2}{2\sigma_d^2}\right) \quad (3)$$

式(3)是定义在定义域核上的函数,在实际地标检测点的过滤中,由于缺乏空间信息,地标点之间的相对位置信息无法有效获取,这里引入空间信息表示,如式(4)所示:

$$p(x, y, a, b) = \exp\left(-\frac{\|f(x, y) - f(a, b)\|^2}{2\sigma_r^2}\right) \quad (4)$$

这是定义在值域上的核,将式(3)和式(4)作乘积可得式(5):

$$r(x, y, a, b) = \exp\left(-\frac{(x-a)^2 + (y-b)^2}{2\sigma_d^2} - \frac{\|f(x, y) - f(a, b)\|^2}{2\sigma_r^2}\right) \quad (5)$$

式(5)中,  $(x, y)$  表示滑动窗口中心值,  $(a, b)$  表示滑动窗口内的坐标值。式中乘积表示可以同时引入时域和空域的信息。利用这种方法可以有效将数据中不满足模型需求的噪声进行剔除。

将计算得出的改进高斯滤波应用到面部地标点的去噪中,过程如图 3 所示。不同于传统的去噪流程,方法去噪的目标是地标点数据,而不是整张图片,这样可以大大减少计算量;同时由于高斯去噪的稳定性需求,将目标图片在面部地标点提取前进行边缘轮廓提取,然后对轮廓进行地标提取,将原始地标点和轮廓地标点加权平均后再进行改进高斯去噪,以获得更加精确的地标点数据。

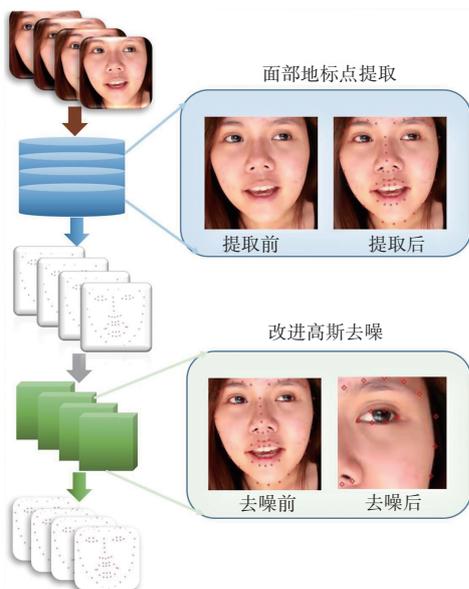


图 3 地标点提取与去噪过程图

Fig. 3 Process diagrams of landmark extraction and denoising

### 3.3 网络结构

提出的 IGFNet 网络结构如图 4 所示,在对人脸数

据分割之后,将面部地标提取出来形成文本文件,之后对文本文件进行校准和去噪,去噪方法采用本文提出的改进高斯方法,而后分别送入不同的向量序列为训练做准备。对文本的训练部分采用循环神经网络,由于单一循环神经网络无法有效获取全局特征,这里采用双流循环神经网络,一个循环神经网络用于寻找面部运动状态,另一个寻找数据点的时间不连续性。最后将循环神经网络的结果加权求和得出最终输出。

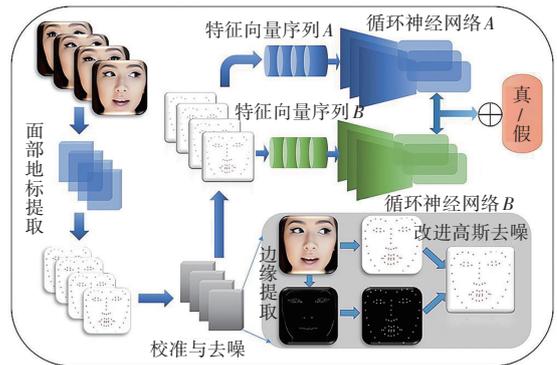


图 4 IGFNet 网络框架结构图

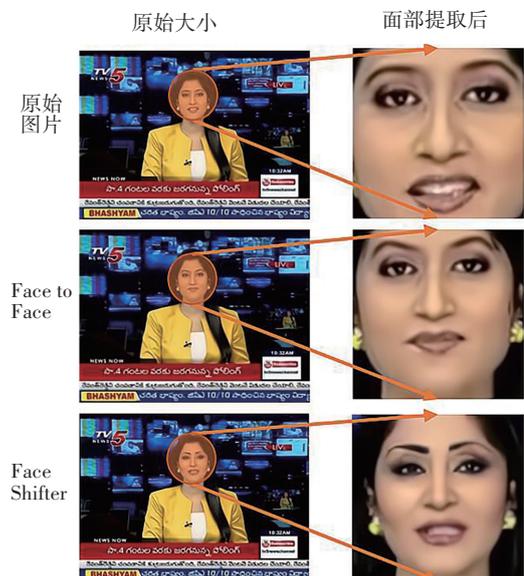
Fig. 4 IGFNet network framework structure

对于图 4 框架中用到的循环神经网络,为了更好地捕捉深层和浅层信息,IGFNet 设置了两种不同的网络结构,应对不同信息学习难度的大小,在网络结构上做了不同的跳层处理,在实验部分对不同深度的网络的结果做了实验并展示。

## 4 实验

### 4.1 相关数据集

在深度伪造方面,当前主要用到数据集有 Face Forensics++<sup>[8]</sup>, UADFV<sup>[9]</sup>, Celeb DF<sup>[10]</sup> 等。FaceForensics++<sup>[8]</sup> 的真实数据和伪造数据比例为 1:4, 真实数据由 1 000 个来自 YouTube 的视频组成,而伪造数据则分别采用了图 5 所示的不同方法产生。



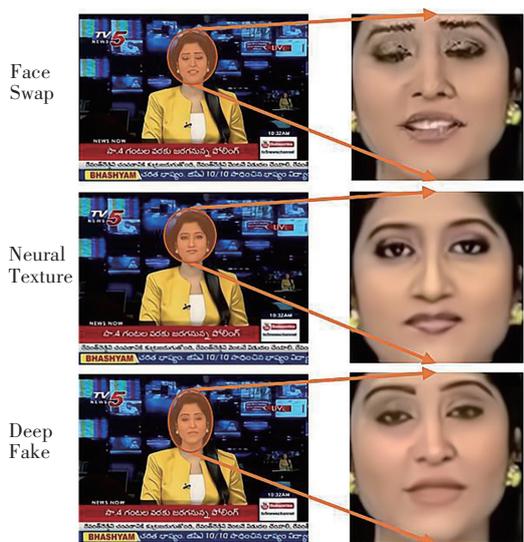


图 5 5 种换脸方法效果比较图

Fig. 5 Comparison of the results of five face-swapping methods

UADFV<sup>[9]</sup>由 49 个真实视频和 49 个伪造视频组成,作为早期伪造视频数据集的代表,很多经典模型中都使用了这个数据集。

Celeb DF<sup>[10]</sup>是一个高质量数据集。Celeb DF 包含 5 639 个假视频和 540 个真实视频,每项工作还提供了一个基准,便于进行评估。

### 4.2 实验预设

这部分介绍对预处理、去噪、特征嵌入和循环神经网络 4 个部分的实施细节以及参数设置。对于预处理,需要先将不同的图片进行面部提取,如图 6 所示,之后采用 Dlib 库对人脸进行检测和对地表进行提取,在去噪部分,采用改进的高斯方法对数据进行去噪;特征嵌入和神经网络分类采用了循环神经网络。



图 6 预处理流程示例图

Fig. 6 Diagrams of pretreatment process

参数设置:在模型中用到的是双流循环神经网络(RNN),该 RNN 由门控循环单元(GRU)组成,输出单元设置为  $k=64$ ,两个全连接层的单元数分别为 64 和 2。为了公平起见,采用 LRNet<sup>[6]</sup>的设置,对数据集采用 8 : 2 分割,即 80%用于培训,20%用于测试。每个视频被分割成固定长度为 60 的片段,当  $fps$  为 30 时,总计为 2 s。优化器用的是 Adam,批量大小设置为是 256,学习步长设置为 0.001,所有数据在该分类模型下都训练 800 轮。

### 4.3 评估指标

关于对模型的评估,采用 Celeb DF<sup>[10]</sup>基准,同时也遵循了他的设置,在 FF++上训练模型,在其他数据集上进行测试。评估指标为 AUC(Area Under Curve),表示的是 ROC 曲线下方的面积大小,这个指标的取值介于 0.5~1 之间,这个指标最大为 1,越大表示效果越好。

### 4.4 定量分析

表 1 展示了不同测试数据集下 AUC 的性能评估结果。方法(IGFNet)在训练数据集 FF++<sup>[8]</sup>上获得了媲美 LRNet<sup>[6]</sup>的 AUC 分数,且在其他数据集上具有比 LRNet<sup>[6]</sup>更优秀的表现,比如在 UADFV 和 Celeb-DF 数据集上分别达到了 99.5 和 63.7,均高于此前的方法。表明方法(IGFNet)可以有效地捕捉人脸地标点的异常抖动。此外,它在推广到压缩数据集时具有一定鲁棒性。如表 2 所示。

表 1 不同测试数据集下 AUC 的性能评估

方法	数据集		
	UADFV	FF++	Celeb-DF
Meso4 <sup>[11]</sup>	84.3	84.7	54.8
Xception <sup>[12]</sup>	80.4	99.7	48.2
Capsule <sup>[13]</sup>	61.3	96.6	57.5
CNN+RNN <sup>[14]</sup>	70.9	98.3	61.5
LRNet <sup>[6]</sup>	98.5	99.9	56.5
IGFNet	99.5	99.8	63.7

表 2 压缩视频下的鲁棒测试

方法	FF++			下降率/%
	raw	c23	c40	
Xception <sup>[12]</sup>	99.7	93.3	86.5	6.4/13.2
X-Ray <sup>[2]</sup>	99.1	87.3	61.6	11.8/37.5
LRNet <sup>[6]</sup>	99.9	97.3	95.7	2.6/4.2
IGFNet	99.8	97.7	96.3	2.1/3.5

为了验证方法对视频压缩的鲁棒性,在 FF++<sup>[8]</sup>数据集上,通过实验比较了当前最佳分类器的基准:Xception<sup>[12]</sup>、新提出的 LRNet<sup>[6]</sup>和先进的 X-Ray<sup>[2]</sup>,结

果见表 2。每个检测器在原始视频 (*raw*) 上进行训练,并在 3 个不同压缩率的视频版本上进行测试。在 Celeb DF<sup>[10]</sup> 上使用了它的基准设置, Xception<sup>[12]</sup> 使用 FF++ (*c23*)<sup>[8]</sup> 进行训练, LRNet<sup>[6]</sup> 和本文的方法直接在 FF++ (*raw*)<sup>[8]</sup> 上进行训练。如从表 2 所示的结果中可以得出,提出的方法 (IGFNet) 的性能相对来说对视频压缩更具不变性,方法在 *c23*, *c40* 数据集上的测试下降率均最低,在泛化性能的表现上优于前者所提出的方法。

#### 4.5 定性分析

双流网络定性分析,对比了 3 种不同的网络结构: DoubleTinyRNN, SingleDeepRNN 和 DoubleDeepRNN, 这 3 种网络分别代表了双流浅层循环神经网络,单流神经网络和双流深层神经网络,且这 3 种网络都采用了改进高斯滤波方式。

采用双流循环神经网络不同于卷积神经网络,会在训练到一定轮数的时候产生震荡,如图 7 所示,这种震荡产生的原因是因为在循环神经网络的时间序列中,信息重复利用率高,导致其函数空间是崎岖的,当学习率较大的时候,容易跳出当前优化方向,进而重新优化产生折线,但当学习率设置过小的时候,又会极大提升训练时间,并且影响实验结果,在多次实验中选定学习率为 0.001,从而达到训练速度和实验结果之间的平衡。通过对比图 7 中不同网络结构的准确率折线图可以得知,当双流网络都较深的时候,效果不明显,且大步长下容易梯度消失;当双流网络都较浅的时候,准确率无法达到最高;当双流网络中仅学习面部运动的网络较深的时候,准确率最高,效果最好。

从图 7 与图 8 对比之中不难发现,准确率会伴随着训练损失一起震荡,且幅度相近。当双流网络都较浅的时候,训练过程中,损失的数值下降最快,但是无法达到最小;当双流网络都较深的时候,损失的数值下降最满,且无法完成训练;当双流网络中仅学习面部运动的网络较深的时候,对应的损失能持续降低到相对最小。

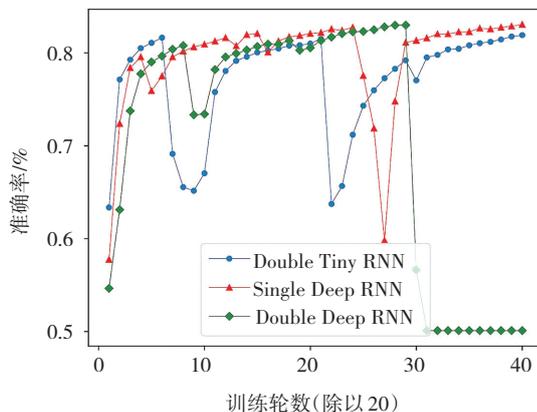


图 7 不同网络结构下的准确率

Fig. 7 Accuracy rate under different network structures

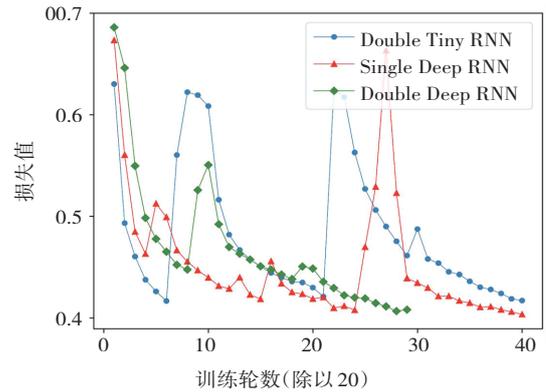


图 8 不同网络结构对应的学习损失曲线

Fig. 8 Learning loss curves corresponding to different network structures

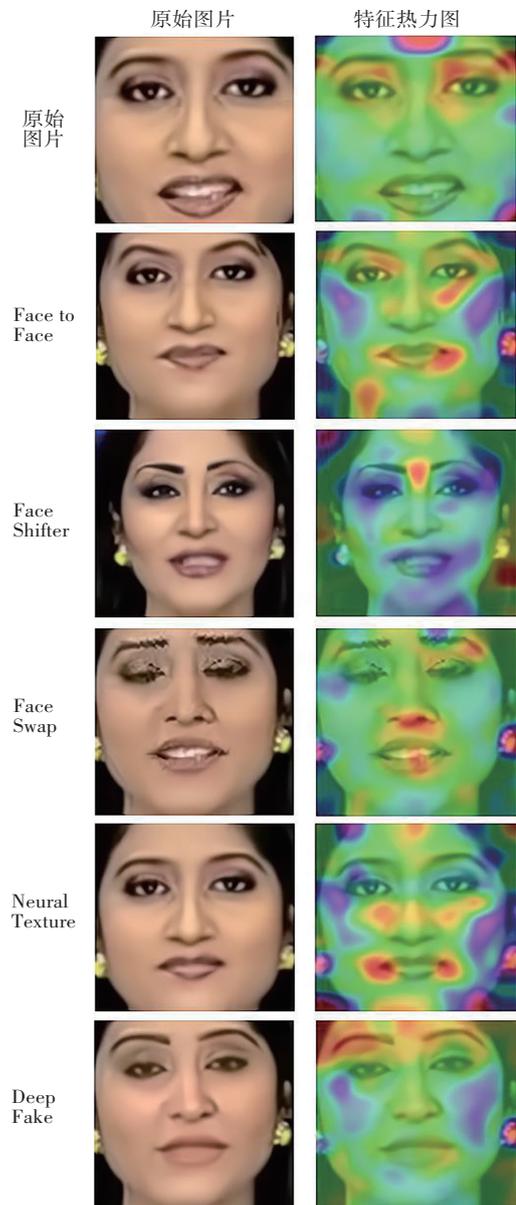


图 9 5 种换脸方法特征梯度热力图

Fig. 9 Gradient thermodynamic diagrams of five face-swapping methods

由于RNN的方法无法直观可视化算法效果,通过抽取中间特征层,生成梯度热力图,进行可视化分析。将双流网络的一支改为CNN直接对原始数据进行处理,另一分支保持原样。图9展示了该单流网络的热力图分析。如图9所示,原始的图像面部特征梯度较为均匀,其他的换脸图像在不同程度上展现了异常的热力值分布,而且这些部分通常是换脸视频中容易产生细微抖动的地方,这显示了方法在双流网络中能更有效地加强这一部分的判断,从而提高检测准确率。

## 5 结束语

利用改进的高斯滤波人脸标记点数据进行去噪处理,文本化的特征数据可以大大减小数据量和缩短训练时长,不同深度的双流网络能有效鉴别出不同的人脸属性。实验证明:所提出的工作(IGFNet)可以有效提高伪造视频的检测准确率,同时在压缩视频上检测准确率的降低也不明显,展现了较好的鲁棒性。在未来,所提出的方法可以通过分析挖掘跨数据库人脸标记点特征,从而推进跨域检测这一深度伪造视频检测的难题。

## 参考文献(References):

- [1] 张时润,彭勃,王伟,等.基于空洞卷积和注意力机制的深度伪造检测[J].现代电子技术,2022,45(5):42—48.  
ZHANG Shi-run, PENG Bo, WANG Wei, et al. Deep forgery detection based on hole convolution and attention mechanism [J]. Modern Electronic Technology, 2022, 45(5): 42—48.
- [2] LI L, BAO J, ZHANG T, et al. Face x-ray for more general face forgery detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5001—5010.
- [3] 邢豪,李明.基于3D CNNs的深度伪造视频篡改检测[J].计算机科学,2021,48(7):86—92.  
XING Hao, LI Ming. Deep forgery video tamper detection based on 3D CNNs [J]. Computer Science, 2021, 48(7): 86—92.
- [4] WU X, XIE Z, GAO Y, et al. SSTNet: detecting manipulated faces through spatial, steganalysis and temporal features [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. 2020: 2952—2956.
- [5] MASI I, KILLEKAR A, MASCARENHAS R M, et al. Two-branch recurrent network for isolating deepfakes in videos [C]//European Conference on Computer Vision. 2020: 667—684.
- [6] SUN Z, HAN Y, HUA Z, et al. Improving the efficiency and robustness of deepfakes detection through precise geometric features [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3609—3618.
- [7] 郑博文,夏华威,陈睿东,等.基于卷积长短期记忆网络的换脸视频检测[J].激光与光电子学进展,2021,58(24):309—317.  
ZHENG Bo-wen, XIA Hua-wei, CHEN Rui-dong, et al. Face-changing video detection based on convolutional long short-term memory network [J]. Advances in Laser and Optoelectronics, 2021, 58(24): 309—317.
- [8] ROSSLER A, COZZOLINO D, VERDOLIVA L, et al. Faceforensics++: learning to detect manipulated facial images [C]//IEEE/CVF International Conference on Computer Vision. 2019: 1—11.
- [9] LI Y, CHANG M-C, LYU S. In ictu oculi: exposing AI generated fake face videos by detecting eye blinking [J]. ArXiv preprint arXiv: 180602877, 2018.
- [10] LI Y, 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. 2020: 3207—3216.
- [11] AFCHAR D, NOZICK V, YAMAGISHI J, et al. Mesonet: a compact facial video forgery detection network [C]//IEEE International Workshop on Information Forensics and Security (WIFS). 2018: 1—7.
- [12] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251—1258.
- [13] NGUYEN H H, YAMAGISHI J, ECHIZEN I. Capsule-forensics: using capsule networks to detect forged images and videos [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. 2019: 2307—2311.
- [14] SABIR E, CHENG J, JAISWAL A, et al. Recurrent convolutional strategies for face manipulation detection in videos [J]. Interfaces, 2019, 3(1): 80—87.