E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP399　　**文献标识码**: A　　**文章编号**: 1006-8961(2024)07-1834-15

论文引用格式:Li W, Huang T Q, Huang L Q, Zheng A K and Xu C. 2024. Large-scale datasets for facial tampering detection with inpainting techniques. Journal of Image and Graphics, 29(07):1834-1848(李伟, 黄添强, 黄丽清, 郑翱鲲, 徐超. 2024. 面向人脸修复篡改检测的大规模数据集. 中国图象图形学报, 29(07):1834-1848)[DOI:10.11834/jig.230422]

面向人脸修复篡改检测的大规模数据集

李伟[1,2,3],黄添强[1,2,3*],黄丽清[1,2,3],郑翱鲲[1,2],徐超[1,2]

1. 福建师范大学计算机与网络空间安全学院,福州　350117; 2. 福建省公共服务大数据挖掘与应用工程技术研究中心,福州　350117; 3. 数字福建大数据安全技术研究所,福州　350117

摘　要: **目的**　图像合成方法随着计算机视觉的不断发展和深度学习技术的逐渐成熟为人们的生活带来了丰富的体验。然而,用于传播虚假信息的恶意篡改图像可能对社会造成极大危害,使人们对数字内容在图像媒体中的真实性产生怀疑。面部编辑作为一种常用的图像篡改手段,通过修改面部的五官信息来伪造人脸。图像修复技术是面部编辑常用的手段之一,使用其进行面部伪造篡改同样为人们的生活带来了很大干扰。为了对此类篡改检测方法的相关研究提供数据支持,本文制作了面向人脸修复篡改检测的大规模数据集。**方法**　具体来说,本文选用了不同质量的源数据集(高质量的人脸图像数据集CelebA-HQ及低质量的人脸视频数据集FF++),通过图像分割方法将面部五官区域分割,最后使用两种基于深度网络的修复方法CTSDG(image inpainting via conditional texture and structure dual generation)和RFR(recurrent feature reasoning for image inpainting)以及一种传统修复方法SC(struct completion),生成总数量达到60万幅的大规模修复图像数据集。**结果**　实验结果表明,由FF++数据集生成的图像在基准检测网络ResNet-50下的检测精度下降了15%,在Xception-Net网络下检测精度下降了5%。且不同面部部位的检测精度相差较大,其中眼睛部位的检测精度最低,检测精度为0.91。通过泛化性实验表明,同一源数据集生成的数据在不同部位的修复图像间存在一定的泛化性,而不同的源数据制作的数据集间几乎没有泛化性。因此,该数据集也可为修复图像之间的泛化性研究提供研究数据,可以在不同数据集、不同修复方式和不同面部部位生成的图像间进行修复图像的泛化性研究。**结论**　基于图像修复技术的篡改方式在一定程度上可以骗过篡改检测器,对于此类篡改方式的检测方法研究具有现实意义。提供的大型基于修复技术的人脸篡改数据集为该领域的研究提供了新的数据来源,丰富了数据多样性,为深入研究该类型的人脸篡改和检测方法提供了有力的基准。数据集开源地址 https://pan.baidu.com/s/1-9HIBya9X-geNDe5zcJldw?pwd=thli。

关键词: 图像篡改;深度学习;图像修复;数据集;基准

Large-scale datasets for facial tampering detection with inpainting techniques

Li Wei[1,2,3], Huang Tianqiang[1,2,3*], Huang Liqing[1,2,3], Zheng Aokun[1,2], Xu Chao[1,2]

1. College of Computer and Network Space Security, Fujian Normal University, Fuzhou　350117, China;

2. Fujian Provincial Engineering Research Center for Public Service Big Data Mining and Application, Fuzhou　350117, China;

3. Digital Fujian Big Data Security Technology Institute, Fuzhou　350117, China

Abstract: **Objective** DeepFake technology, born with the continuous maturation of deep learning techniques, primarily utilizes neural networks to create non-realistic faces. This method has enriched people's lives as computer vision advances

收稿日期:2023-07-06;**修回日期**:2023-12-19;**预印本日期**:2023-12-26

*通信作者:黄添强　fjhtq@fjnu.edu.cn

基金项目:国家自然科学基金项目(62072106);福建省科技创新平台项目(2023-P-003);福建省科技厅产学研合作项目(2021H6004)

Supported by: National Natural Science Foundation of China (62072106); Fujian Province Science and Technology Innovation Platform Project (2023-P-003)

and deep learning technologies mature. It has revolutionized the film industry by generating astonishing visuals and reducing production costs. Similarly, in the gaming industry, it has facilitated the creation of smooth and realistic animation effects. However, the malicious use of image manipulation to spread false information poses significant risks to society, casting doubt on the authenticity of digital content in visual media. Forgery techniques encompass four main categories: face reenactment, face replacement, face editing, and face synthesis. Face editing, a commonly employed image manipulation method, involves falsifying facial features by modifying the information related to the five facial regions. As one of the commonly employed methods in facial editing, image inpainting technology involves utilizing known content from an image to fill in missing areas, aiming to restore the image in a way that aligns as closely as possible with human perception. In the context of facial forgery, image inpainting is primarily used for identity falsification, wherein facial features are altered to achieve the goal of replacing a face. The use of image inpainting for facial manipulation similarly introduces significant disruption to people's lives. To support research on detection methods for such manipulations, this paper produced a large-scale dataset for face manipulation detection based on inpainting techniques. **Method** This paper specifically focuses on the field of image tampering detection, utilizing two classic datasets: the high-quality CelebA-HQ dataset, comprising 25 000 high-resolution (1 024 × 1 024 pixels) celebrity face images, and the low-quality FF++ dataset, consisting of 15 000 face images extracted from video frames. On the basis of the two datasets, facial feature regions (eyebrows, eyes, nose, mouth, and the entire facial area) are segmented using image segmentation methods. Corresponding mask images are created, and the segmented facial regions are directly obscured on the original image. Two deep neural network-based inpainting methods (image inpainting via conditional texture and structure dual generation (CTSDG) and recurrent feature reasoning for image inpainting (RFR)) along with a traditional inpainting method (struct completion(SC)) were employed. The deep neural network methods require the provision of mask images to indicate the areas for inpainting, while the traditional method could directly perform inpainting on segmented facial feature images. The facial regions were inpainted using these three methods, resulting in a large-scale dataset comprising 600 000 images. This extensive dataset incorporates diverse pre-processing techniques, various inpainting methods, and includes images with different qualities and inpainted facial regions. It serves as a valuable resource for training and testing in related detection tasks, offering a rich dataset for subsequent research in the field, and also establishes a meaningful benchmark dataset for future studies in the domain of face tampering detection. **Result** We present comparative experiments conducted on the generated dataset, revealing notable findings. Experimental results indicate a 15% decrease in detection accuracy for images derived from the FF++ dataset under the ResNet-50 benchmark detection network. Under the Xception-Net network, the detection accuracy experiences a 5% decline. Furthermore, significant variations in detection accuracy are observed among different facial regions, with the lowest accuracy recorded in the eye region at 0.91. Generalization experiments suggest that inpainted images from the same source dataset exhibit a certain degree of generalization across different facial regions. In contrast, minimal generalization is observed among datasets created from different source data. Consequently, this dataset also serves as valuable research data for studying the generalization of inpainted images across different facial regions. Visualization tools demonstrate that the detection network indeed focuses on the inpainted facial features, affirming its attention to the manipulated facial regions. This work provides new research perspectives for methods of detecting image restoration-based manipulations. **Conclusion** The use of image inpainting techniques for tampering introduces a challenging scenario that can deceive conventional tampering detectors to a certain extent. Researching detection methods for this type of tampering is of practical significance. The provided large-scale face tampering dataset, based on inpainting techniques, encompasses high- and low-quality images, employing three distinct inpainting methods and targeting various facial features. This dataset offers a novel source of data for research in this field, enhancing diversity and providing benchmark data for further exploration of image restoration-related forgeries. With the scarcity of relevant datasets in this domain, we propose the utilization of this dataset as a benchmark for the field of image inpainting tampering detection. This dataset not only supports research in detection methodologies but also contributes to studies on the generalization of such methods. It serves as a foundational resource, filling the gap in the available datasets and facilitating advancements in the detection and generalization studies in the domain of image inpainting tampering. This benchmark includes a large-scale inpainting image dataset,

totaling 600 000 images. The dataset's quality is evaluated based on accuracy on manipulation detection networks, generalizability across different inpainting networks and facial regions, and modules such as data visualization.

Key words: image tampering; deep learning; image inpainting; dataset; benchmark

0 引 言

深度伪造(DeepFake)作为一种新兴的人工智能技术,引起了诸多学者和领域爱好者的研究兴趣。深度伪造技术可以在影视行业中创造出令人惊叹的画面,同时节约制作成本;在游戏行业中展现流畅真实的动画效果,提升用户体验。然而,深度伪造技术也是一把双刃剑,除了带来科技创新,也不可避免地引发了许多社会安全问题。深度伪造技术的出现给国家安全领域带来了新的挑战,如在政治抹黑、军事欺骗、经济犯罪甚至恐怖主义行动中的恶意运用,导致国家间的紧张关系和社会混乱。因此,为了应对深度伪造技术带来的社会安全风险,强大且高效的篡改检测方法显得尤为必要。

人脸伪造是指利用传统或深度学习方法制作出非真实人脸的一种技术。随着深度学习的快速发展,高效率和高质量的生成结果使其成为伪造技术的主力军。伪造技术主要包括面部重现(face reenactment)、面部替换(face replace)、面部编辑(face editing)和全脸合成(face synthesis)4大类。面部重现常使用生成对抗网络(generative adversarial network, GAN)来修改面部表情。FF++数据集中的Face2Face和Neural Textures也属于面部重现的方式。面部替换将源人物的面部替换到目标人物的脸上,是有目标的替换。除了GAN之外,还有使用深度学习网络的Face Shifter技术和使用图形学的Face Swap技术。面部编辑主要用于添加、更改或删除目标身份的属性,如衣服、胡子、年龄、体重、颜值、种族、肤色等,常使用StarGAN和StyleGAN等GAN方法来完成。全脸合成则是生成一张完全不存在的人脸,主要用于生成虚假的网络身份。

图像修复技术(image-inpainting)(Zhao等,2021)是一种使用图像中的已知内容去填补缺失区域,使修复后的图像尽可能满足人类感知的一种技术手段。同样,图像修复技术也是一把双刃剑,它的出现极大地丰富了数字图像的应用,如图像编辑、影视特效、动画补全、数字文化遗产保护等,但其不当使用也会招致许多负面影响,例如在网络上将某些不方便露面的人脸图像通过修复技术复原,从而给人物带来困扰等。因此,该技术一直是研究者们重点关注的问题之一。

图像修复在人脸伪造方面主要运用于身份造假,即通过改变人脸五官达到更换人脸的目的,这为社会安全带来了很大的潜在威胁。而现有的人脸篡改数据集并没有针对图像修复这一类型的篡改数据,这为研究相关的检测网络带来了困难。因此,本文制作了一批面向人脸修复篡改检测的大规模数据集。具体来说,使用在图像篡改领域中常用的经典基准数据集FF++和CelebA-HQ数据集,在其基础上应用图像分割算法先分割出人脸的五官区域,再分别使用不同的图像修复方法,最终生成大规模修复图像的篡改数据集。本文提供的数据集为图像篡改检测方法带来新的数据支持,对不断提高篡改检测方法的检测精度和对不同篡改图像检测泛化性的提升具有十分重要的意义。

此外,由于在修复图像篡改伪造领域缺乏伪造检测的标准,本文将生成的大规模数据集拓展为基准数据,以便支持对该领域的研究及数据集后续的扩展。该基准数据集采用高低质量的人脸图像在其不同的面部区域使用不同的修复方式以生成庞大规模的数据。同时,本文分别用人眼分辨及结构相似性(structure similarity index measure, SSIM)和峰值信噪比(peak signal-to-noise ratio, PSNR)指标对生成的数据集的质量进行评估。并且使用基准检测网络ResNet-50、MesoNet(Afchar等,2018)和Xception-Net(Chollet,2017)对数据集的篡改效果进行检测。结果表明,检测精度确实有不同程度的下降,从而证明制作数据集的必要性。

本文主要贡献如下:1)制作了一批面向人脸修复篡改检测的大规模数据集,为该领域的相关研究提供数据基础。2)利用一系列指标对数据集生成质量进行评估,并对使用基准检测网络对图像伪造效果进行了检测,证明了制作数据集的质量。3)提出了一个基于修复图像篡改数据集的基准。

1 相关工作

本文涉及计算机视觉、数字媒体取证和数字图像处理等多个领域的相关内容，下面将简要介绍相关方法。

1.1 面部篡改方法

随着计算机技术的快速发展，深度学习在计算机视觉和数字媒体领域扮演着越来越重要的角色。根据Zollhöfer等人(2018)的研究数据，过去20年中，关于数字媒体中虚拟人脸的研究呈现持续增加的趋势。

早期的伪造检测算法主要依赖编码器—解码器(decoder-encoder)网络来生成篡改图像，通过对图像进行降采样以获取融合的特征数据，然后修改这些特征数据后再进行上采样，从而生成视觉上更加合理的结果。而生成对抗网络(GAN)作为一种不依赖先验知识、生成样本更加逼真的方法，目前已经成为生成图像的主要手段。

Dale等人(2011)通过人脸视频交换提出了最早的人脸互换方法，利用3D几何体将源人物的面部重建到目标人物的面部。Face2Face是由Thies等人(2016)提出的一种实时面部再现系统，结合3D模型重建和图像渲染技术以生成最终的图像。Neural Textures(Thies等，2019)将原始视频数据使用渲染网络优化目标人物的神经纹理，通过结合光度重建损失与对抗损失进行训练以生成更细致的结果。Face Swap(Korshunova等，2017)使用两个具有共享编码器的自动编码器，分别训练以重建目标图像，是DeepFake中最常用的篡改方式。

1.2 图像修复

图像修复技术通过图像中已知区域的内容和结构信息对图像中破损或缺失的区域进行推测并修复。作为计算机图形学和计算机视觉领域的重要分支之一，图像修复的目标是将图像进行合理的修补，以尽可能地满足视觉感知的需求。近年来图像修复技术受到了研究者们的广泛关注，并运用于网络安全、影视文化创作和日常生活等多个领域。

在依托神经网络的深度学习技术广泛运用于计算机视觉领域之前，传统的图像修复算法主要通过像素间的相关性及内容结构上的相似性对图像进行修复。基于几何图像模型的变分和偏微分技术是最早运用的修复技术。如BSCB模型(Bertalmio等，2000)提出使用三阶偏微分方程对平滑传输过程进行模拟。CDD(curvature-driven-diffusions)模型(Chan和Shen，2001)使用曲率扩散强度进行图像修复，该模型使用扩散的思想，根据待修复区域的边缘信息向空白区域的内层进行逐层填补，并且每一层的填补都会参考上一层的修复结果。基于几何图像模型的变分和偏微分技术在小尺度破损图像的修复上有着较好的效果。对于缺失区域较大的图像，则多采用基于样本的方法进行修复。从背景区域中寻找合理的内容填充到缺失区域中使整幅图像形成合理的效果。这种方式同时也可以填充图像的纹理细节。比较有代表性的是由Criminisi等人(2004)提出的基于块的纹理合成Criminisi算法和Cheng等人(2005)使用不同优先权函数改进的Criminisi算法。后续的许多传统修复算法也都是在Criminisi算法的基础上改进得到。

近年来，研究者尝试将深度学习技术引入到图像修复领域，并提出了许多使用深度网络的图像修复算法。context-encoder网络(Pathak等，2016)将编码—解码网络引入图像修复工作中，通过将生成对抗网络GAN与编解码网络结合，使用重构损失和对抗损失约束网络来提高修复效果。Liu等人(2018)提出了部分卷积(partial-convolutions)来代替普通卷积网络。该网络使用一种含有自动掩膜(mask)机制的模型来完成对破损图像的修复。Yu等人(2018b)将注意力机制引入图像修复模型，提出了一种含有内容感知层的双阶段粗细图像修复模型。Xiong等人(2019)利用图像的已知区域得到前景轮廓作为结构先验来指导缺失区域的修补，提出了一种前景感知图像修复模型。生成模型的方法是通过学习图像的离散概率分布来预测二维图像中的像素点，从而进行图像的修复。扩散模型作为当下最火热的图像生成模型，也运用到图像修复的领域中，Lugmayr等人(2022)提出Repaint模型，将破损的图像结合到预训练的扩散模型反向生成的步骤中，以完成图像修复；Repaint模型通过生成方式修复的图像所蕴含的信息更加丰富，能够得到图像背景中不存在的像素信息。

数据集主要使用了3种修复方式修复掩膜(mask)的图像：CTSDG(image inpainting via conditional texture and structure dual generation)(Guo等，2021)、RFR(image inpainting via conditional texture

and structure dual generation)(Li 等, 2020a)和 SC(struct completion)(Huang等,2014)。其中前两种方法用深度学习网络进行图像修复,而第3种方法则使用传统方法进行图像修复。

1.3 检测方法

伪造生成图像逐渐逼真为社会带来了安全隐患,使得近年来更多人开始关注伪造图像的篡改检测,并且提出了许多高效精确的模型。当今主流的检测方法主要可分为基于伪影的篡改图像检测、基于数据驱动的篡改图像检测和基于信息不一致的篡改图像检测。如Amerini等人(2020)提出使用光流法对假脸视频进行篡改检测。Li等人(2019)发现视频中眼睛闪烁具有固定规则这一生物特征,后又提出估算出人脸图像中3D头部摆动姿势,通过结合支持向量机(support vector machines,SVM)分类器对篡改有效进行检测。

当今最流行和最有效的方法仍然基于深度神经网络(deep neural network,DNN)。图像篡改检测作为经典的二分类问题,许多优秀的分类网络在图像篡改检测领域也有着非常高的精度。残差神经网络(ResNet)由微软研究院的He等人(2016)提出,ResNet通过在网络结构中添加"快捷连接(short connection)"来解决深度网络过大导致网络性能发生的退化问题,并且在当年的图像分类任务中取得了冠军。其中ResNet-50网络将49个卷积层和全连接层结合起来,是目前最常使用的ResNet网络之一。Xception-Net网络(Chollet, 2017)由Google团队于2017年提出,是一种基于Inception-Net网络的变体,通过添加深度可分离卷积和残差分支,在保证精度的情况下减少了网络的计算量,也是常用的基准检测网络之一。Cozzolino等人(2017)提出了一种基于残差的局部描述符方式,通过对网络微调能够在小型数据集上获得更好的性能。Zhao等人(2021)提出了频率感知判别特征学习,用于图像面部篡改检测。Nirkin等人(2020)通过改进双流残差结构以提高检测网络的精度。Li等人(2021)将图像的空间域和频域进行融合设计双流网络结构来进行图像篡改检测。

1.4 取证分析数据集

为了支持研究者对图像伪造篡改的研究以及后续相关领域的工作,为众多图像篡改检测方法提供训练和测试数据,近几年有许多公共数据集陆续提出。

UADFV(exposing AI created fake video by detecting eye blinking)数据集于2018年提出,是最早运用于图像篡改领域的公共数据集,包含分辨率为294 × 500像素的真假视频各49个。这些视频都与著名演员尼古拉斯·凯奇的原貌进行了交换。数据集不包含声音信息,且视频质量较低,人眼可观察到部分人脸拼接的痕迹,是多媒体取证领域最早的基准测试数据集。

Rössler等人(2019)提出一个新的面部操作基准数据集FaceForenscis++。该数据集一经发布便成为图像篡改检测领域最流行的数据库,数据集包含4种生成方法生成的视频,并且提供高低压缩率的视频数据,生成伪造图像超过400个。多数的图像篡改检测网络都会采用该数据集进行训练和测试,已成为该领域的基准数据集。数据集主要使用DeepFake、FaceSwap、Face2Face及Neural Texture 4种伪造方法将来自YouTube的1 000个真实视频进行篡改伪造,以生成大规模的篡改视频数据。该数据集的视频还包含3种不同压缩方式:原始视频(RAW、c0)、高压缩率(HQ,c23)以及低压缩率(LQ,c40)。

CelebA是由香港中文大学提供的开放人脸数据集,包含10 177个名人身份的202 599幅图像,以及40种不同的人脸属性,是计算机视觉领域最常用的高质量数据集之一。Celeb-DF(Li, 2020b)使用DeepFake算法合成的包含890个真实视频和5 639个假视频的高质量数据集。CelebA-HQ则是CelebA数据集的高质量版本,每一幅图像的分辨率都达到了1 024 × 1 024像素,总共有30 000幅,用于高质量的图像任务。

DeeperForensics-1.0(DF-1.0)(Jiang 等, 2020)是用于真实世界人脸伪造的大型数据集,使用高保真的人脸交换方法DVAE(DeepFake variational autoendecoder)来解决视频低质量的问题,伪造视频是在FF++数据集基础上合成的。

视频篡改检测数据库(video forgery detection database,VFDD)由华南理工大学电子与信息学院多媒体信息安全检测与智能处理中心开发,于2017年发布1.0版本,包含了12台设备在8种不同场景下拍摄的原始视频505段,对其篡改后所得到的视频135段,共计640段。该数据集于2019年发布了2.0版本,统一了全部视频的命名格式并且对全部视频进行了勘误。

DeepFake TIMIT数据集包含了16对相似人群,

每人生成了10个视频，各包含低分辨率（64 × 64像素）和高分辨率（128 × 128像素）两个版本，各320个视频，共640个视频。数据集带有原始视频的音轨信息，没有对音轨信息进行修改。

FakeAVCeleb（Khalid等，2021）是目前为止所知的首个同时包含伪造视频和伪造音频的篡改数据集，从VoxCeleb2数据集选择了490个真实数据，原始VoxCeleb2包含1 092 009个真实数据，真实数据的种族、男女比例均衡，利用了Faceswap、DeepFaceLab和FSGAN算法生成换脸视频，利用SV2TTS生成伪造的声音，利用Wav2Lip生成唇部的模拟，生成了超过20 000个篡改视频。

2 修复数据集的制作方法

本文的核心贡献在于提供了一个基于修复图像的大规模数据集，对传统篡改图像基准数据集进行了拓展，以在具体应用场景中支持图像修复这一篡改手段的检测。

如图1所示，为了制作数据集，本文选用了两个在图像篡改领域常用的基准数据集作为源数据集，在其基础上使用图像分割网络分割出人脸图像的五官区域并制作掩膜（mask）图像，最后使用3种不同的图像修复算法生成最终的大规模数据集。下面将详细介绍所使用的源数据集、分割方法及不同的修复方法。

2.1 使用的源数据集

本文选择了两个在图像篡改领域最常用的基准数据集：FF++和CelebA-HQ。

FF++数据集包含真实图像和被篡改的图像，是图像篡改检测领域最具代表性的数据集之一，包含1 000个真实视频和4 000个通过DeepFake、Face2Face、FaceSwap和Neural Textures共4种不同方法生成的假视频。虽然数据集中考虑了多种生成技术，但其视觉效果较差，图像合成痕迹明显。

CelebA-HQ是在CelebA基准数据集上生成的高质量图像。CelebA数据集包含超过20万幅名人图像，涵盖了多种变化姿势和杂乱背景，是计算机视觉领域常用的数据集之一。CelebA-HQ通过训练高分辨率GAN生成新的高质量人脸数据集，每幅图像的分辨率都达到了1024 × 1024像素，用于研究高质量图像修复和检测的效果。

为制作样本，首先从FF++数据集选取1 000个未经压缩的真实视频，并将每个视频通过每隔16帧抽取1帧的方式得到对应的图像。随后，使用人脸裁剪算法提取了这些图像中的面部区域，总共得到1.5万幅人脸数据集图像。对于CelebA-HQ数据集，本文选用2.5万幅高质量的人脸图像，且未经任何修改。这一系列数据处理步骤为后续的研究提供了高质量的图像样本。

图1 基于修复技术的人脸数据集制作流程

Fig. 1 The process of creating a inpainted facial image dataset

2.2 数据集中采用的图像分割方法

本文采用了语义分割领域中经典的轻量级网络BiSeNet（bilateral segmentation network for real-time semantic segmentation）（Yu等，2018a）进行人脸数据集的五官分割任务。该模型在计算量、参数量和精度之间取得了较好的平衡，能够实时地分割出所需的语义信息。

首先，对图像分割网络BiSeNet进行训练，使其能准确地分割出人脸的五官区域。对于使用深度学习网络进行图像修复的需求，分别为其制作分割出

的五官掩膜(mask),而对于传统算法修复图像,只需得到分割后的图像即可。

具体而言,本文在两种数据集上制作了具有mask区域的待修复图像及其对应的掩膜,其中包括眼睛、鼻子、嘴巴和眉毛4个区域。

2.3 数据集中采用的图像修复方法

在最重要的图像修复任务中,本文采用了3种使用不同修复原理的图像修复算法对之前预处理的图像进行修复。包含两种使用深度学习网络进行图像修复的算法CTSDG和RFR及一种传统算法SC。下面详细介绍3种不同的修复方法。

2.3.1 基于结构信息约束的图像修复方法

CTSDG(Guo等,2021)是一种使用深度学习网络基于结构信息约束的图像修复方法。深度生成的方法通过引入结构先验的方式在图像修复领域取得了很大的进步,然而由于在结构重建过程中缺乏与图像纹理的适当交互,会使之前的修复方法在处理大面积掩膜(mask)区域时效果不尽人意,从而导致结果失真。而CTSDG是一种新颖的用于图像修复的双流网络,它以耦合的方式对结构约束的纹理合成和纹理边缘引导结构重建进行建模,使它们更好地相互利用,以获得更合理的生成内容。此外,为了增强全局一致性,网络还设计了双向门控特征融合(bi-directional gated feature fusion,Bi-GFF)模块来交换和结合结构及纹理信息,并设计了上下文特征聚合(contextual feature aggregation,CFA)模块,使用区域亲和学习和多尺度特征聚合来细化具体的生成内容。从而使图像生成的效果更加合理,符合人们的认知需求。

2.3.2 基于注意力机制的深度学习网络图像修复方法

RFR-Net(Li等,2020a)是另一种使用基于注意力机制的深度学习网络的图像修复方法,提出了一个作用于特征层面的渐进式图像修复网络——循环特征推理网络,以解决填充大型掩膜(mask)中心区域不连贯的问题。该网络主要由递归特征推理模块和知识一致注意力(knowledge consistent attention,KCA)模块构成,与人们解决问题的方式相似,即首先解决较容易的部分,然后将结果作为解决困难部分的附加信息。RFR模块反复推理卷积特征图的孔洞边界,然后将其作为下一次推断的直接线索。为了保证循环过程中不同特征图之间的一致性,该网络在特征推理过程中特别设计了一种知识一致注意力模块。该模块的当前循环过程中每个像素最终的注意力得分是由前面循环过程中的注意力得分和当前注意力得分加权获得。使用这种作用于特征层面的渐进式修复方法可以取得更精细的修复效果。

2.3.3 基于平面结构指导的传统图像修复方法

SC(Huang等,2014)是一种基于平面结构指导的传统图像修复方法,该方法是一种典型的基于补丁的传统图像修复方法。模型首先估计平面投影参数,将已知区域分割成3个相互垂直的平面。然后通过定义块偏移和变换的先验概率,将该信息转换为低层完成算法的软约束,是目前生成质量较高的传统修复方法。

2.4 本文制作的数据集构成

数据集采用源FF++数据集1.5万幅图像和CelebA-HQ数据集2.5万幅图像作为基础,分别使用了3种不同的修复方式(RFR和CTSDG两种深度学习网络的修复方式和SC传统图像修复方式),并且在5种不同的面部区域进行修复(包括眉毛、眼睛、鼻子、嘴巴以及4种部位同时修复),最终生成了总数为60万幅的基于修复图像检测的大规模数据集。这个大规模数据集使用了不同的预处理方式、不同的修复方法,包含了不同质量、不同面部修复部位的修复图像。为后续的相关领域检测任务提供丰富的训练和测试数据,同时也为该领域的后续研究提供了一份有价值的基准数据。图2展示了制作的数据集总体构成。

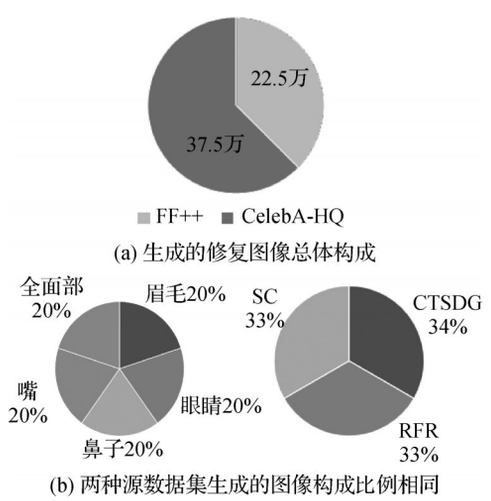


图2 制作的数据集总体构成

Fig. 2 The overall composition of the created dataset
((a) the overall composition of the generated inpainted images; (b) the composition ratio of images generated by the two source datasets was the same)

3 实 验

为了证明本文生成的面向修复技术的人脸篡改大规模数据集的优越性，本文分别对图像修复质量、图像在基准检测网络上的精度、图像在不同修复网络间的泛化性及图像的可视化方面进行了相关对照实验，下面将详细介绍每个实验的内容。本文的代码均通过 Python 来编写，使用 4 张 TeslaV100s 显卡来进行网络的训练与测试。

3.1 图像修复质量

本文的数据集是基于修复技术来进行图像生成的，修复的质量将直接影响到图像检测的精度。如图 3 所示，通过深度网络（CTSDG、RFR）进行修复后

(a) 在CelebA-HQ数据集上进行图像修复的最终结果

(b) 在FF++数据集上进行图像修复的最终结果

图 3 生成的数据集结果展示

Fig. 3 Presentation of the generated dataset results ((a) the final results of image inpainting on the CelebA-HQ dataset; (b) the final results of image inpainting on the FF++ dataset)

得到的图像较为真实。由于缺少图像先验知识，传统修复方法（SC）只能使用待修复图像的周围区域进行填补，从而导致生成效果不太理想。

进一步，本文采用了广泛应用于生成图像评估的两个指标：峰值信噪比（PSNR）和结构相似性（SSIM）。峰值信噪比通过比较变化后的图像与真实图像之间的差剖面（即可视误差）来评价图像的质量，而结构相似性则重新设计了图像的评价标准，更符合人的主观评价标准，是图像生成领域常用的指标之一。

如表1和表2所示，使用深度网络进行修复的图像的SSIM值与PSNR值要略高于使用传统方法修复的图像。SSIM值和PSNR值越高，表示图像质量越高，说明使用深度网络进行图像修复较为合适。由于眉毛部位及眼睛部位占图像比例较小，所以其SSIM值和PSNR值也较高。低质量的数据集整体的分辨率较低，使用FF++数据集制作的图像的评价指标要高于使用CelebA-HQ数据集。

表1　CelebA-HQ数据集修复图像的SSIM和PSNR指标

Table 1　The SSIM and PSNR metrics were used to evaluate the quality of image inpainting using the CelebA-HQ datasets

部位	CTSDG		RFR		SC	
	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB
眉毛	**0.982**	**38.61**	**0.990**	41.05	**0.989**	36.00
眼睛	0.981	38.24	**0.990**	**41.06**	0.988	**37.33**
鼻子	0.972	32.88	0.985	37.65	0.978	30.90
嘴	0.978	34.84	0.987	34.73	0.983	33.87
全面部	0.967	32.12	0.980	34.53	0.965	29.68

注：加粗字体表示各列最优结果。

表2　使用FF++数据集修复图像的SSIM和PSNR指标

Table 2　The SSIM and PSNR metrics were used to evaluate the quality of image inpainting using the FF++ datasets

部位	CTSDG		RFR		SC	
	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB
眉毛	0.997	49.32	0.990	**40.02**	0.982	34.46
眼睛	0.994	44.10	0.991	39.70	**0.983**	**34.68**
鼻子	0.983	33.55	0.978	34.80	0.963	29.30
嘴	**0.998**	**55.14**	**0.995**	45.59	0.979	33.54
全面部	0.976	34.63	0.959	32.42	0.934	27.27

注：加粗字体表示各列最优结果。

3.2　检测效果

图像修复通过填充图像中的缺失部分，使其看起来完整和真实。为了评估修复后图像的质量，本文将修复后的图像视为假（fake）图像，然后通过与真实图像进行比较，进行图像篡改检测任务，使用ACC（accuracy）评价指标来观察两种图像之间的差异。

具体来说，对于FF++数据集，为了与修复图像保持一致，本文同样另选取500个未经压缩的真实视频，将这些视频通过每隔16帧抽取1帧得到对应的图像，之后对这些图像进行水平边缘裁剪后将图像的尺寸调整至256×256像素，最终生成了15 000幅图像，作为真实图像与生成的图像进行比较。对于CelebA-HQ数据集，直接随机选取其中的15 000幅图像做为真实图像即可。对于修复图像，本文在两种不同的源数据集上，使用3种不同的修复方式，修复了包括眉毛、眼睛、鼻子和嘴巴4种面部区域的修复图像各15 000幅。

对于检测网络，本文选取了图像篡改检测领域的经典基准网络ResNet-50、MesoNet和Xception-Net进行精度检测。实验中分别选取10 000幅真实图像与10 000幅修复后的图像作为训练集进行模型训练，训练过程中将网络的batch_size设置为12，学习率为10^{-4}，并使用10^{-5}进行微调，训练轮次为100。其中80%作为训练数据，20%作为验证数据。将剩余的5 000幅真实图像与5 000幅修复图像做为测试数据来测试数据集效果。基准网络在不同数据集上的检测效果（ACC）如表3所示，检测效果差证明制作的数据集质量较高。

最后，将制作好的修复图像数据集与DFDC（Dolhansky等，2019）、Celeb-DF等篡改领域经典数据集一同放入具有高度代表性的图像篡改检测网络中进行分类效果测试，从而评估生成图像的质量。

与其他基准数据集相比，本文的数据集在检测网络上的检测效果均有不同程度的下降，FF++数据集修复的图像在ResNet-50网络上的检测精度下降了15%，在Xception-Net网络上的精度下降了5%。证明图像修复这种篡改方式在一定程度上能骗过现有的检测模型，还有值得研究的空间。

表 3　基准网络在不同数据集上的检测效果(ACC)

Table 3　The detection effect on the benchmark network on different datasets (ACC)

/%

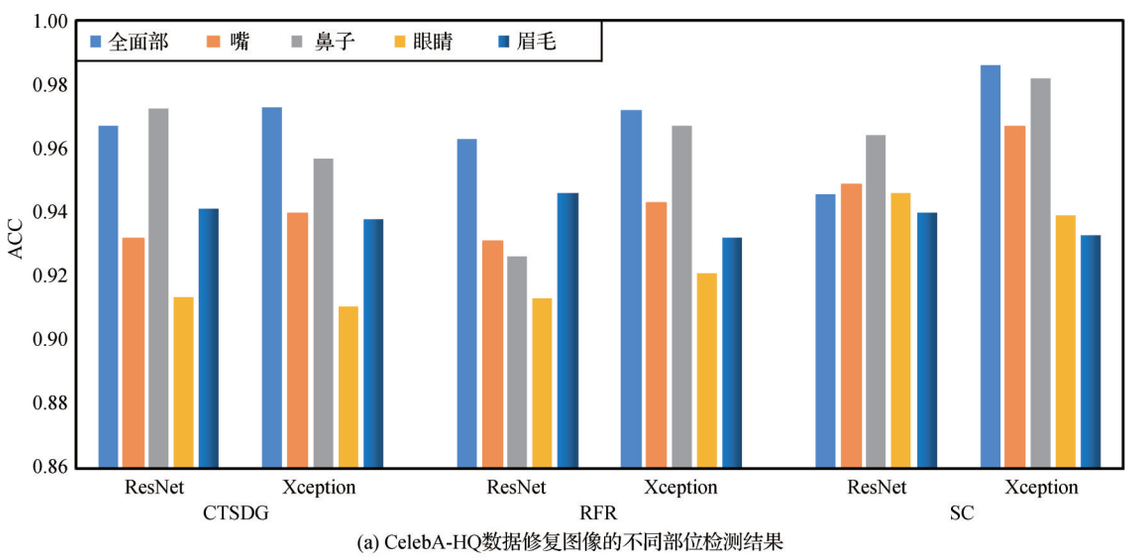
数据集	ResNet-50	MesoNet	Xception
UADFV	97.4	84.3	91.2
DeepFakeTIMIT(LQ)	99.9	87.8	95.9
DeepFakeTIMIT(HQ)	93.2	68.4	94.4
DFDC	89.4	**75.3**	**72.2**
FF++	91.2	76.2	92.1
FF++(inpainting)	**74.1**	79.3	87.5
CelebA-HQ(inpainting)	88.3	86.4	90.7

注：加粗字体表示各列最低检测值。

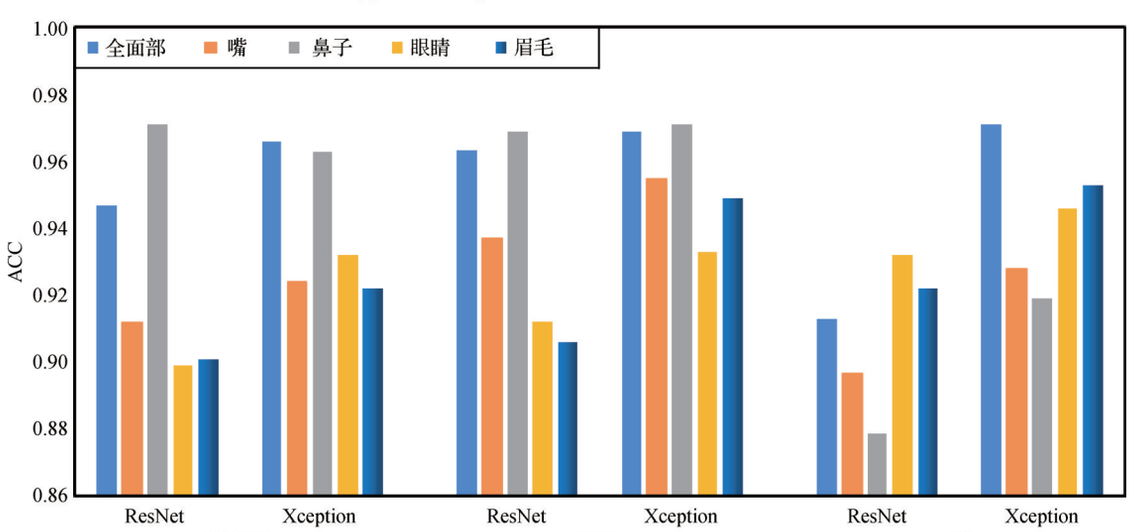
3.3　具体部位分辨效果

为了验证不同修复部位对篡改检测效果的影响，本文进一步对具体面部部位进行了篡改检测实验。同样，本实验将源数据集作为真实图像(real)，修复数据集作为假图像(fake)，并使用基准检测网络 ResNet-50 和 Xception-Net 进行二分类网络的检测。本实验同样使用 ACC 评价指标衡量不同面部区域的差异。

如图 4 所示，CelebA-HQ 和 FF+数据集在不同的五官部位的检测精度有一定的差距，由于眼睛部位在整个图像中所占的比例最小，使得修复的区域最少，导致眼睛部位的检测精度最低。同样，修复区域



(a) CelebA-HQ数据修复图像的不同部位检测结果



(b) FF++数据修复图像的不同部位检测结果

图 4　生成的基于修复技术的大规模数据集在不同部位的篡改检测精度

Fig. 4　The accuracy of tamper detection at different locations in a large-scale dataset generated based on inpainting techniques ((a) ACC of different parts in the inpainted images of the CelebA-HQ dataset; (b) ACC of different parts in the inpainted images of the FF++ dataset)

为整个面部五官的图像由于待修复区域较大，因此检测精度最高；但相较于篡改检测领域来说，图像的检测精度还有待提升的空间。传统修补方法修复质量较低，图像存在伪影，检测器可以轻松地识别出，故检测精度也最高。

3.4 泛化性

该实验旨在研究由同一源数据集（如FF++数据集）中使用一种修复方法（如RFR）修复出某一面部区域（如眉毛区域）的图像在使用其他修复方式（CTSDG和SC）训练的检测网络上不同面部区域（眼睛、鼻子、嘴巴）的检测精度，或者在不同源数据集（CelebA-HQ）上生成的各种部位图像训练的检测网络上的泛化性。在本实验中选取了CelebA-HQ数据集，使用两种深度修复网络（CTSDG和RFR）和一种传统修复网络（SC）上的面部区域位置进行训练。这里使用了Xception-Net网络做为检测模型进行训练。测试了训练的模型在同一数据集的不同修复方法下的相同面部部位上的检测结果，同时也测试了在不同数据集上相同方式的检测精度（使用ACC评价指标来展示）。

如表4所示，不同数据集之间的图像分布不同，所以检测网络在不同的数据集中几乎没有表现出泛化性。在同一数据集中，检测网络在同为深度网络的修补方法之间泛化性较强，而在传统修复方法上泛化性较弱。说明对基于修复技术的图像篡改检测技术的泛化性也是一个值得研究的问题。另外，通过实验数据可以看到，通过传统方法修复的数据训练的检测网络在使用深度学习修复的数据上几乎没有泛化性，这也是一个值得思考的问题。

表4 CelebA-HQ数据集不同部位的数据在不同数据集上不同修复方法上的泛化性（ACC）

Table 4 Generalizability of data from different parts of the CelebA-HQ dataset using different methods on different datasets (ACC)

方法	部位	CelebA-HQ		FF++		
		RFR	SC	CTSDG	RFR	SC
CTSDG	眉毛	0.904	0.821	0.558	0.504	0.502
	眼睛	**0.958**	0.809	**0.569**	**0.573**	**0.513**
	鼻子	0.957	0.670	0.503	0.504	0.505
	嘴	0.931	0.779	0.510	0.507	0.503
	全面部	0.954	**0.863**	0.501	0.502	0.503
		CTSDG	SC	CTSDG	RFR	SC
RFR	眉毛	0.897	0.855	0.497	0.511	0.504
	眼睛	0.912	0.734	**0.556**	0.516	0.513
	鼻子	0.947	0.906	0.501	0.501	0.509
	嘴	0.852	0.826	0.512	**0.547**	0.501
	全面部	**0.953**	**0.949**	0.553	0.494	0.502
		RFR	CTSDG	CTSDG	RFR	SC
SC	眉毛	0.536	0.501	0.506	**0.604**	0.511
	眼睛	0.516	0.514	**0.510**	0.505	0.529
	鼻子	0.646	0.519	0.493	0.507	0.522
	嘴	0.556	0.526	0.509	0.533	**0.627**
	全面部	**0.596**	**0.599**	0.501	0.524	0.599

注：加粗字体表示在不同数据集不同部位的最优检测值。

3.5 结果可视化

这一部分主要展示神经网络的可解释性。积分梯度（integrated gradient）和Grad-CAM是常见用来显示深度神经网络中具体某一层的梯度变化情况，以视图的方式直观地展现网络推理过程的实用性工具。

本小节通过展示检测网络对修复图像数据集检测效果的可视化结果来说明数据集的多样性及特殊性。

3.5.1 Grad-CAM

Grad-CAM（gradient-weighted class activation map）是一种生成类别激活热力图的方法，可以直观地展现神经网络某一层关注的区域。在本实验中使用Grad-CAM来展示分类网络在进行分类时是否关注了图像所修复的区域，从而探讨修复图像数据集的分类效果。如图5所示，ResNet-50网络关注的部位确实为修复的区域。

3.5.2 积分梯度

积分梯度（integrated gradient）作为一种解决传统基于梯度的可解释性方法中梯度饱和缺陷的方法，具有更强的全局可解释性。相较于Grad-CAM，积分梯度能够更完整地展现网络对图像的推理过程，呈现更全面的可解释性效果。如图6所示，3幅图像中的第1列为修复后的图像，第2列和第3列为检测网络不同层的梯度积分反向计算的结果。使用积分梯度方式显示的图像中也能看见ResNet-50检测网络所关注的修复区域，虽然显示效果没有热力图清晰，但积分梯度方式因具有全局可解释性也更具有说服力。

图 5　Grad-CAM 在 ResNet-50 上展示的结果

Fig. 5　Grad-CAM′s results displayed on the ResNet-50

图 6　积分梯度在 ResNet-50 上展示的结果

Fig. 6　The results of integrated gradient displayed on the ResNet-50

4　基准数据集

由于在图像篡改检测领域缺少基于修复技术的大规模图像数据集，本文创建了一个面对修复图像的篡改检测数据集基准。该基准包含了总规模达到 60 万幅的大规模修复图像、对于生成数据集的质量评价、数据集在篡改检测网络上的精度、数据集在不

同修复网络和不同面部区域间的泛化性以及数据可视化等模块。

数据集的基本构成以及数据集在检测网络上的精度在表5中直观展示出来。数据集本身包含大量的篡改伪造(DeepFake)图像,并且涵盖了不同分辨率、不同压缩率的图像,通过多种修复方式对人脸面部五官区域进行修补生成的修复图像数据集。

作为基线,本文在基准数据集上对先前训练的模型进行了测试,包括FF++修复数据集的较低质量版本。除了Xception-Net网络外,同时也对输入的人脸图像进行了预处理操作。分类模型在修复数据集上的相对性能与原始数据集相似。然而,由于图像修复任务是不确定的,生成的图像结果并不唯一,这给生成结果带来了不确定性,从而增加了检测的难度。因此,检测模型在修复数据集上的准确率相较于原始数据集有所下降,总体性能较低。

在高质量数据集(CelebA-HQ)上的研究方法与低质量数据集类似,将高质量图像以同样的预处理方式送入分类网络进行训练。由于CelebA-HQ原始数据集中并没有篡改假脸数据,因此选择Celeb-DF数据集作为原始数据集的DeepFake版本进行网络训练。网络性能与低质量数据集类似,相较于原始数据集,检测网络的性能都有不同程度的下降。然而,由于CelebA数据集的普遍质量较高,相应的识别率相较于FF++数据集有所提升。

表5　修复图像数据集的基本内容

Table 5　The content of the inpainting datasets

项目	内容
源数据集	FF++(15 000),CelebA-HQ(25 000)
修复的面部部位	眉毛、眼睛、鼻子、嘴、全面部
采用的修复网络	CTSDG,RFR,SC
生成的图像总数量	600 000幅
检测网络ResNet-50的ACC/%	74.1(FF++),88.3(CelebA-HQ)
检测网络MesoNet的ACC/%	79.3(FF++),86.4(CelebA-HQ)
检测网络Xception的ACC/%	**87.5**(FF++),**90.4**(CelebA-HQ)

注:加粗字体为数据集在经典检测网络上的最佳检测精度。

本文希望能将制作的基于图像修复技术的大规模人脸篡改数据集作为该细分领域的基准数据集,为修复数据和对应检测网络的研究提供数据参考与支持。

5　结　论

数字媒体取证技术是当今热门研究课题之一,一个高精度且具有泛化性的检测网络对保护网络信息安全具有重要的现实意义。本文制作了一批面向修复技术的人脸篡改检测大规模数据集,为图像修复这类伪造和对应检测网络的进一步研究提供数据支持。同时,本文也对数据集的修复质量进行了评价并对数据集的制作意义进行了实验说明。希望这一基准数据集能为未来修复图像数字媒体取证领域的研究做出贡献,特别是对于图像这种方式的篡改伪造。

本文所提供的数据集也有一定的局限性,如图像的修复部位局限于人脸的五官部位,而伪造也可能发生在人脸的其他区域,或者非人脸图像也会出现图像修复这种方式的伪造。后续将进一步完善该数据集,扩充数据集的类型与数量,以提高对该细分领域的进一步研究。

参考文献(References)

Afchar D, Nozick V, Yamagishi J and Echizen I. 2018. MesoNet: a compact facial video forgery detection network//2018 IEEE International Workshop on Information Forensics and Security (WIFS). Hong Kong, China: IEEE: 1-7 [DOI: 10.1109/WIFS. 2018. 8630761]

Amerini I, Galteri L, Caldelli R and Bimbo A D. 2020. Deepfake video detection through optical flow based CNN//Proceedings of 2019

IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South): IEEE: 1205-1207 [DOI: 10.1109/ICCVW.2019.00152]

Bertalmio M, Sapiro G, Caselles V and Ballester C. 2000. Image inpainting//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA: ACM: 417-424 [DOI: 10.1145/344779.344972]

Chan T F and Shen J H. 2001. Nontexture inpainting by curvature-driven diffusions. Journal of Visual Communication and Image Representation, 12(4): 436-449 [DOI: 10.1006/jvci.2001.0487]

Cheng W H, Hsieh C W, Lin S K, Wang C W and Wu J L. 2005. Robust algorithm for exemplar-based image inpainting//Proceedings of 2005 International Conference on Computer Graphics, Imaging and Visualization. [s.l.]: [s.n.]: 64-69

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Cozzolino D, Poggi G and Verdoliva L. 2017. Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection//Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security. Philadelphia, USA: ACM: 159-164 [DOI: 10.1145/3082031.3083247]

Criminisi A, Pérez P and Toyama K. 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 13(9): 1200-1212 [DOI: 10.1109/TIP.2004.833105]

Dale K, Sunkavalli K, Johnson M K, Vlasic D, Matusik W and Pfister H. 2011. Video face replacement. ACM Transactions on Graphics, 30(6): 1-10 [DOI: 10.1145/2070781.2024164]

Dolhansky B, Howes R, Pflaum B, Baram N and Ferrer C C. 2019. The deepfake detection challenge (DFDC) preview dataset [EB/OL]. [2023-06-20]. https://arxiv.org/pdf/1910.08854.pdf

Guo X F, Yang H Y and Huang D. 2021. Image inpainting via conditional texture and structure dual generation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14134-14143 [DOI: 10.1109/ICCV48922.2021.01387]

He K M, Zhang X Y, Ren S Q and Sun J. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016 [DOI:10.1109/CVPR.2016.90]

Huang J B, Kang S B, Ahuja N and Kopf J. 2014. Image completion using planar structure guidance. ACM Transactions on Graphics, 33(4): #129 [DOI: 10.1145/2601097.2601205]

Jiang L M, Li R, Wu W, Qian C and Loy C C. 2020. Deeperforensics-1.0: a large-scale dataset for real-world face forgery detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2889-2898 [DOI: 10.1109/CVPR42600.2020.00296]

Khalid H, Tariq S, Kim M and Woo S S. 2021. FakeAVCeleb: a novel audio-video multimodal deepfake dataset [EB/OL].[2023-06-20]. https://arxiv.org/pdf/2108.05080]

Kietzmann J, Lee L W, McCarthy I P and Kietzmann T C. 2020. Deepfakes: trick or treat? Business Horizons, 63(2): 135-146 [DOI: 10.1016/j.bushor.2019.11.006]

Korshunova I, Shi W Z, Dambre J and Theis L. 2017. Fast face-swap using convolutional neural networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 3677-3685 [DOI: 10.1109/ICCV.2017.397]

Li H D, Li B, Tan S Q and Huang J W. 2018. Detection of deep network generated images using disparities in color components [EB/OL]. [DOI:10.1016/j.sigpro.2020.107616]

Li J M, Xie H T, Li J H, Wang Z Y and Zhang Y D. 2021. Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 6454-6463 [DOI: 10.1109/CVPR46437.2021.00639]

Li J Y, Wang N, Zhang L F, Du B and Tao D C. 2020a. Recurrent feature reasoning for image inpainting//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 7760-7768 [DOI: 10.1109/CVPR42600.2020.00778]

Li Y Z, Yang X, Sun P, Qi H G and Lyu S W. 2020b. Celeb-DF: a large-scale challenging dataset for deepfake forensics//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 3204-3213 [DOI: 10.1109/CVPR42600.2020.00327]

Liu G L, Reda F A, Shih K J, Wang T C, Tao A and Catanzaro B. 2018. Image inpainting for irregular holes using partial convolutions//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 89-105 [DOI: 10.1007/978-3-030-01252-6]

Lugmayr A, Danelljan M, Romero A, Yu F, Timofte R and Van Gool L. 2022. RePaint: inpainting using denoising diffusion probabilistic models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11461-11471 [DOI: 10.1109/CVPR52688.2022.01117]

Nirkin Y, Wolf L, Keller Y and Hassner T. 2020. DeepFake detection based on discrepancies between faces and their context. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10): 6111-6121 [DOI: 10.1109/TPAMI.2021.3093446]

Pathak D, Krähenbühl P, Donahue J, Donahue J and Efros A A. 2016. Context encoders: feature learning by inpainting//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2536-2544 [DOI: 10.1109/CVPR.2016.278]

Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J and Niessner M. 2019. Faceforensics++: learning to detect manipulated facial images//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1-11 [DOI: 10.1109/ICCV.2019.00009]

Thies J, Zollhöfer M and Nießner M. 2019. Deferred neural rendering: image synthesis using neural textures. ACM Transactions on Graphics, 38(4): #66 [DOI: 10.1145/3306346.3323035]

Thies J, Zollhöfer M, Stamminger M, Theobalt C and Nießner M. 2016. Face2face: real-time face capture and reenactment of RGB videos// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2387-2395 [DOI: 10.1109/CVPR.2016.262]

Van Den Oord A, Kalchbrenner N and Kavukcuoglu K. 2016. Pixel recurrent neural networks//Proceedings of the 33rd International Conference on Machine Learning. New York, USA: JMLR.org: 1747-1756 [DOI: 10.5555/3045390.3045575]

Xiong W, Yu J H, Lin Z, Yang J M, Lu X, Barnes C and Luo J B. 2019. Foreground-aware image inpainting//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5840-5848 [DOI: 10.1109/CVPR.2019.00599]

Yu C Q, Wang J B, Peng C, Gao C X, Yu G and Sang N. 2018a. BiSeNet: bilateral segmentation network for real-time semantic segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 325-341 [DOI: 10.1007/978-3-030-01261-8_20]

Yu J H, Lin Z, Yang J M, Shen X H, Lu X and Huang T S. 2018b. Generative image inpainting with contextual attention//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5505-5514 [DOI: 10.1109/CVPR.2018.00577]

Zhao H Q, Wei T Y, Zhou W B, Zheng W M, Chen D D and Yu N H. 2021. Multi-attentional deepfake detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 2185-2194 [DOI: 10.1109/CVPR46437.2021.00222]

Zollhöfer M, Thies J, Garrido P, Bradley D, Beeler T, Pérez P, Stamminger M, Nießner M and Theobalt C. 2018. State of the art on monocular 3D face reconstruction, tracking, and applications. Computer Graphics Forum, 37(2): 523-550 [DOI: 10.1111/cgf.13382]

作者简介

李伟，男，硕士研究生，主要研究方向为对抗性深度学习与数字媒体取证。E-mail：lw261737@163.com

黄添强，通信作者，男，博士生导师，教授，主要研究方向为机器学习安全和数字媒体取证。E-mail：fjhtq@fjnu.edu.cn

黄丽清，女，讲师，主要研究方向为视频图像超分辨和多媒体内容安全。E-mail：lqhuang@fjnu.edu.cn

郑翱鲲，男，硕士研究生，主要研究方向为对抗性深度学习与数字媒体取证。E-mail：362603727@qq.com

徐超，男，讲师，主要研究方向多媒体内容安全。E-mail：xuchao@fjnu.edu.cn