深度伪装与超越:人脸操纵与伪装检测综述

Ruben Tolosana、Ruben Vera-Rodriguez、Julian Fierrez、艾塔米·莫拉莱斯和Javier Ortega-Garcia生物识别和数据模式分析-西班牙马德里自治大学BiDA实验室

*{鲁本.托洛萨纳，鲁本.维拉，朱利安.菲尔雷斯，艾塔米.莫拉莱斯，哈维尔.奥尔特加}@uam.es*

***摘要——对大规模公共数据库的免费访问，以及深度学习技术的快速发展，特别是生成式对抗网络，导致了非常真实的虚假内容的产生，以及在这个假新闻时代对社会的相应影响。***

[arXiv:2001.00179v3](#_bookmark0)【cs。CV]6月18日[2020](#_bookmark1)

**这项调查提供了一个彻底的审查技术，操纵面部图像，包括深假的方法，以及方法来检测这种操纵。特别地，回顾了四种类型的面部操纵:I)全脸合成，ii)身份交换(DeepFakes)，iii)属性操纵，和iv)表情交换。对于每个操纵组，我们都提供了有关操纵技术、现有公共数据库和假货检测方法技术评估的关键基准的详细信息，包括这些评估结果的摘要。在调查中讨论的所有方面中，我们特别关注最新一代的DeepFakes，突出了它对假货检测的改进和挑战。**

**除了调查信息，我们还讨论了开放的问题和未来的趋势，应考虑在该领域的进展。**

***索引术语—假新闻、深度假货、媒体取证、人脸处理、人脸识别、生物识别、数据库、基准***

1. 介绍

**英语字母表中第六个字母**

AKE图像和视频，包括通过数字操作生成的面部信息，特别是使用DeepFake方法[[1],](#_bookmark51)最近已经成为公众非常关注的问题[[2],](#_bookmark52) [[3].](#_bookmark53) 非常流行的术语“DeepFake”是指一种基于深度学习的技术，能够通过用另一个人的脸交换一个人的脸来创建假视频。这个术语起源于一位名叫“deepfakes”的Reddit用户在2017年底声称开发了一种机器学习算法，帮助他将名人的面孔转换成色情视频[[4].](#_bookmark54)除了虚假色情，这些虚假内容的一些更有害的用法包括假新闻、恶作剧和金融欺诈。因此，传统上致力于一般媒体取证的研究领域[[5]–](#_bookmark55)[[11],](#_bookmark58)现在正致力于在图像和视频中检测面部操纵[[12].](#_bookmark59)假脸检测方面的这些新努力的一部分是围绕过去在生物识别反欺骗方面的研究建立的[[13]–[15]](#_bookmark61) 和现代数据驱动的深度学习[[16],](#_bookmark62)[[17].](#_bookmark63)在顶级会议中，越来越多的研讨会表明了人们对假脸检测的兴趣[[18]–](#_bookmark64)[[22],](#_bookmark65)国际项目，如由国防高级研究计划局(DARPA)资助的MediFor，以及竞赛，如最近的媒体取证挑战

(MFC2018)[1](#_bookmark0) 以及深度造假检测挑战赛(DFDC)[2](#_bookmark1)分别由国家标准和技术研究所(NIST)和脸书发起。

传统上，由于缺乏复杂的编辑工具、所需的领域专业知识以及所涉及的复杂且耗时的过程，面部操作的数量和真实性受到了限制。例如，这个主题的早期作品[[23]](#_bookmark66) 通过在音轨的声音和受试者的面部形状之间建立联系，我能够修改使用不同音轨说话的人的嘴唇运动。然而，从这些早期作品到现在，许多东西在最近几年里迅速发展。如今，自动合成不存在的人脸或操纵图像/视频中一个人的真实人脸变得越来越容易，这要归功于:I)对大规模公共数据的可访问性，以及ii)深度学习技术的发展，该技术消除了许多手动编辑步骤，如自动编码器(AE)和生成式对抗网络(GAN)[[24],](#_bookmark67) [[25].](#_bookmark68)因此，ZAO等开放软件和移动应用[3](#_bookmark2)和FaceApp[4](#_bookmark3)已经发布，为任何人创造虚假的图像和视频打开了大门，而不需要任何该领域的经验。

为了响应那些日益复杂和真实的被操纵内容，研究团体正在进行大量努力来设计用于面部操纵检测的改进方法。媒体取证中的传统伪造检测方法通常基于:I)相机内指纹，由相机设备(硬件和软件，例如光学镜头)引入的固有指纹的分析[[27],](#_bookmark70) 彩色滤光片阵列和插值[[28],](#_bookmark71)[[29],](#_bookmark72)和压缩[[30],](#_bookmark73)[[31],](#_bookmark74)以及ii)相机外指纹，对由编辑软件引入的外部指纹的分析，例如复制粘贴或复制移动图像的不同元素[[32],](#_bookmark75)[[33],](#_bookmark76)降低视频中的帧速率[[34]–[36],](#_bookmark78)等等。然而，在传统的假货检测方法中考虑的大多数特征高度依赖于特定的训练场景，因此对于看不见的情况并不鲁棒[[6],](#_bookmark56)[[8],](#_bookmark57)[[16].](#_bookmark62)这在我们生活的时代尤为重要，因为大多数媒体虚假内容通常在社交网络上共享，其平台会自动修改原始图像/视频，例如，通过压缩和调整大小操作[[12].](#_bookmark59)

数字一（one）<https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018>

2<https://deepfakedetectionchallenge.ai/>

3<https://apps.apple.com/cn/app/id1465199127>

4<https://apps.apple.com/gb/app/faceapp-ai-face-editor/id1180884341>

图一。各面部操作组的真假例子。对于整个面部合成，真实图像是从<http://www.whichfaceisreal.com/>和伪造的图像[https://thispersondoesnotexist.com。](https://thispersondoesnotexist.com/)对于身份交换，人脸图像从Celeb-DF数据库中提取[[26].](#_bookmark69) 对于属性操作，真实图像是从<http://www.whichfaceisreal.com/>而假图像是用FaceApp生成的。最后，对于表情交换，从FaceForensics++中提取图像[[12].](#_bookmark59)



表情互换



身份互换

**骗子**

**骗子**

这项调查对应用于面部内容的数字操纵技术进行了深入的审查，因为这种技术有大量可能的有害应用，例如，产生假新闻，在政治选举和安全威胁中提供错误信息[[37],](#_bookmark79) [[38].](#_bookmark80) 具体来说，我们涵盖了四种类型的操作:I)全脸合成，ii)身份交换，iii)属性操作，iv)表情交换。这四种主要类型的面部操作已经被研究团体很好地确立，在过去的几年中受到了最多的关注。除此之外，我们还回顾了其他一些具有挑战性和危险性的人脸操作技术，这些技术并不流行，但像人脸变形。



属性操作



全脸合成

**真实的**

来源

目标

**骗子**

**真实的**

**真实的**

目标

来源

**骗子**

**真实的**

最后，为了完整起见，我们想强调该领域最近的其他调查。在…里[[39],](#_bookmark81)作者从一般的角度讨论了DeepFakes的主题，提出了

R.管理深度欺诈风险的E.A.L框架。此外，Verdoliva最近还调查了[[40]](#_bookmark82)一般认为传统的操纵和伪造检测方法

媒体取证，还有最新的深度学习技术。本调查补充了[[39]](#_bookmark81)和[[40]](#_bookmark82)对每个面部操作组进行了更详细的审查，包括操作技术、现有的公共数据库和伪造检测方法技术评估的关键基准，包括这些评估结果的总结。此外，我们特别关注最新一代的DeepFakes，突出其对假货检测的改进和挑战。

文章的其余部分组织如下。我们首先提供秒。[II](#_bookmark5) 不同类型的面部操作的一般描述。然后，从秒。[III](#_bookmark12)敬Sec。[VI](#_bookmark37)我们描述了每种面部操作的关键方面，包括公共研究数据库，检测方法和基准结果。秘书[VII](#_bookmark41)重点介绍其他有趣的面部处理技术，这些技术在前面的章节中没有涉及。最后，我们提供秒。[VIII](#_bookmark49)我们的结束语，强调了未决问题和未来趋势。

1. 面部操作的类型

根据操作水平，面部操作可分为四大类。图。[1](#_bookmark4)图形总结了每个面部操作组。下面是对它们的描述

表一

全脸合成:公开可用的数据库。

**数据库真实图像虚假图像**

[[41]](#_bookmark83)

10万张生成图像(2019) - 10万张(StyleGAN)

降低操作水平:

* 整张脸的合成:这种操作创造了整个

10万张面孔(2019年)

[[47]](#_bookmark89)

DFFD (2020)

- 100，000英镑(斯泰勒根)

100，000(斯泰勒根)

不存在的人脸图像，通常通过强大的GAN，

[[17]](#_bookmark63) -

200 000(Prog an)

例如，通过最近在[[41].](#_bookmark83)这些技术取得了惊人的效果，生成了具有高度真实感的高质量面部图像。图。[1](#_bookmark4)显示了整个面部的一些例子

-

iFakeFaceDB (2020年)

[[16]](#_bookmark62)

250，000(斯泰勒根)

80，000 (ProGAN)

使用StyleGAN生成的合成[5](#_bookmark7).这种操纵可能有利于许多不同的行业，如视频游戏和3D建模行业，但它也可能被用于有害的应用程序，如在社交网络中创建非常逼真的虚假个人资料，以产生错误信息。

* 身份交换:这种操作包括用另一个人的脸替换视频中一个人的脸。通常考虑两种不同的方法:

*I)经典的基于计算机图形的技术，例如FaceSwap*[6](#_bookmark8)以及ii)被称为DeepFakes的新颖深度学习技术[7](#_bookmark9)例如最近的ZAO移动应用。在Youtube上可以看到这种操纵的非常真实的视频[8](#_bookmark10).这种操纵可以让许多不同的行业受益，尤其是电影业。然而，另一方面，它也可能被用于不良目的，如制作名人色情视频、恶作剧和金融欺诈等。

* 属性操作:这种操作也称为面部编辑或面部修饰，包括修改面部的一些属性，如头发或皮肤的颜色、性别、年龄、添加眼镜等[[42].](#_bookmark84)这种操作过程通常通过GAN进行，例如在[[43].](#_bookmark85)这种操作的一个例子是流行的FaceApp移动应用程序。消费者可以使用这项技术在虚拟环境中试穿各种产品，如化妆品、眼镜或发型。
* 表情互换:这种操作，也称为面部重现，包括修改人的面部表情。尽管在文献中提出了不同的操作技术，例如，通过流行的GAN架构在图像级别[[44],](#_bookmark86)在这个小组中，我们将重点放在最流行的技术上[[45],](#_bookmark87) [[46],](#_bookmark88) 其用另一个人的面部表情代替视频中一个人的面部表情。这种类型的操纵可能会带来严重的后果，例如，马克·扎克伯格说了他从未说过的话的流行视频[9](#_bookmark11).

5[https://thispersondoesnotexist.com](https://thispersondoesnotexist.com/)

6<https://github.com/MarekKowalski/FaceSwap>7<https://github.com/deepfakes/faceswap>

8<https://www.youtube.com/watch?v=UlvoEW7l5rs>9<https://www.bbc.com/news/technology-48607673>

1. 全脸合成
2. *操作技术和公共数据库*

这种操作创建了整个不存在的面部图像。桌子[I](#_bookmark6)总结了主要的公开可用的数据库的研究检测图像处理技术依赖于整个脸的合成。四个不同的数据库与此相关，它们都基于相同的GAN架构:ProGAN[[48]](#_bookmark90)还有StyleGAN[[41].](#_bookmark83)有趣的是，注意到每个伪图像可以由特定的GAN指纹来表征，就像自然图像由基于设备的指纹(即PRNU)来识别一样。事实上，这些指纹似乎不仅依赖于GAN架构，还依赖于它的不同实例[[49]–[51].](#_bookmark92)此外，如表所示[I,](#_bookmark6)值得注意的是，上述四个数据库仅包含使用所讨论的GAN架构生成的伪图像。为了对这个操纵组进行假货检测实验，研究人员需要从CelebA等其他公共数据库中获取真实人脸图像[[52],](#_bookmark93)FFHQ[[41],](#_bookmark83)卡西娅-

WebFace[[53],](#_bookmark94) 和VGGFace2[[54],](#_bookmark95) 除其他外。

接下来我们提供每个公共数据库的描述。在…里[[41],](#_bookmark83)Karras等人发布了一组100，000张合成人脸图像，名为100K-Generated-Images[10](#_bookmark13).这个数据库是使用他们提出的StyleGAN架构生成的，该架构是使用FFHQ数据集训练的[[41].](#_bookmark83)StyleGAN是他们以前流行的方法ProGAN的改进版本，该方法引入了一种基于逐步改进生成器和鉴别器的新训练方法。StyleGAN提出了一种替代的生成器架构，该架构导致自动学习、无监督地分离高级属性(例如，在人脸上训练时的姿势和身份)和生成的图像中的随机变化(例如，雀斑、头发)，并且它实现了对合成的直观的、特定于比例的控制。

另一个公共数据库是100K-Faces[[47].](#_bookmark89) 该数据库包含使用Style- GAN生成的100，000幅合成图像。在这个数据库中，与100K生成的图像数据库相反，使用来自69个不同模型的大约29，000张照片来训练StyleGAN网络，考虑来自更可控场景(例如，具有平坦背景)的面部图像。因此，在图像的背景中不包括由StyleGAN创建的奇怪的人造物。

10<https://github.com/NVlabs/stylegan>



* 1. 伪造(GANprintR后的伪造

图二。在使用GANprintR移除GAN指纹信息后，使用StyleGAN及其改进版本创建的假图像示例[[16].](#_bookmark62)

最近，Dang等人在[[17]](#_bookmark63)一个新的数据库命名为多样化的假脸数据集(DFFD)。关于整个人脸合成操作，作者分别通过预训练的ProGAN和StyleGAN模型创建了10万个和20万个假图像。

最后，内维斯等人在[[16]](#_bookmark62)iFakeFaceDB数据库。该数据库包括分别用StyleGAN和ProGAN创建的250，000和80，000张合成人脸图像。与以前的数据库相比，作为一个附加功能，并且为了防止假检测器，在该数据库中，GAN架构产生的指纹通过一种称为GANprintR (GAN指纹移除)的方法移除，同时保持非常逼真的外观。图。[2](#_bookmark14)显示了一个使用StyleGAN直接生成的假图像示例及其在删除GAN指纹信息后的改进版本。作为GANprintR步骤的结果，与其他数据库相比，iFakeFaceDB对高级伪造检测器提出了更高的挑战。

1. *操纵检测*

最近，不同的研究评估了检测人脸是真实的还是人工生成的难度。桌子[II](#_bookmark16) 显示了这方面最相关的方法的比较。对于每项研究，我们包括与方法、分类器、最佳性能和考虑的数据库相关的信息。我们用粗体突出显示每个公共数据库获得的最佳结果。值得注意的是，在某些情况下，会考虑不同的评估指标，例如曲线下面积(AUC)或等误差率(EER)，这使得研究之间的比较变得复杂。

一些作者建议分析内部GAN管道，以便检测真实和伪造图像之间的不同伪像。在…里[[55],](#_bookmark96)作者假设真实相机图像和假合成图像之间的颜色明显不同。他们提出了一种基于颜色特征和线性支持向量机(SVM)的检测系统用于最终分类，当使用NIST MFC2018数据集进行评估时，实现了70.0%的最终AUC以获得最佳性能[[62].](#_bookmark102)

这方面的另一个有趣的方法是在[[56].](#_bookmark97)Wang等人推测，监控神经元行为也可以作为检测假脸的一项资产，因为逐层神经元激活模式可以捕获对面部操纵检测系统很重要的更微妙的特征。他们提出的方法，命名为

FakeSpoter，从深度人脸识别系统(即VGG人脸)中提取的真实和虚假人脸的神经元覆盖行为作为特征[[63],](#_bookmark104) OpenFace[[64],](#_bookmark105) 和FaceNet[[65]),](#_bookmark106) 然后为最终分类训练一个SVM。作者使用来自CelebA-HQ的真实人脸测试了他们提出的方法[[48]](#_bookmark90)和FFHQ[[41]](#_bookmark83) 通过界面生成的数据库和合成人脸[[66]](#_bookmark107) 还有StyleGAN[[41],](#_bookmark83) 使用FaceNet模型实现最佳性能的最终84.7%的虚假检测准确度。

更好的结果最近在[[57].](#_bookmark98)作者提出了一种基于卷积轨迹分析的伪检测系统。使用期望最大化算法提取特征[[67].](#_bookmark108)流行的分类器如k-最近邻(k-NN)、SVM和线性判别分析(LDA)用于最终检测。他们提出的方法用AttGAN生成的假图像进行了测试[[68],](#_bookmark109)GDWCT[[69],](#_bookmark110)斯塔根[[43],](#_bookmark85)StyleGAN和StyleGAN2[[70],](#_bookmark111)获得最终99.81%的Acc。为了最好的表现。

受隐写分析启发的伪造检测系统也已经被研究。纳塔拉杰等人在1999年提出[[58]](#_bookmark99)一种基于像素共生矩阵和卷积神经网络(CNN)组合的检测系统。他们提出的方法最初是通过CycleGAN创建的各种对象和场景的数据库进行测试的[[71].](#_bookmark112)此外，作者进行了有趣的分析，以查看所提出的方法对通过不同GAN架构(CycleGAN vs. StarGAN)创建的伪图像的鲁棒性，具有良好的概括结果。这种检测方法后来在[[16]](#_bookmark62) 考虑来自100K-Faces数据库的图像，实现了12.3%的EER以获得最佳的假货检测性能。这个结果在表中用斜体表示[II](#_bookmark16) 以表明原始论文中未提供。

许多研究还集中在使用纯深度学习方法检测由GAN架构插入的特殊指纹。于等人在1995年提出[[59]](#_bookmark100) 一种将输入图像映射到其对应的指纹图像的属性网络结构。因此，他们学习了每个源(每个GAN实例加上真实世界)的模型指纹，使得一个图像指纹和每个模型指纹之间的相关指数用作用于分类的softmax logit。他们提出的方法用CelebA数据库中的真实人脸进行了测试[[52]](#_bookmark93) 和通过不同GAN方法产生的合成面(ProGAN[[48],](#_bookmark90) SNGAN[[72],](#_bookmark113) 克莱姆甘[[73],](#_bookmark114) 和MMDGAN[[74]),](#_bookmark115) 实现最佳性能的最终99.5%的假检测准确度。然而，这种方法对于看不见的简单图像扰动攻击，如噪声、模糊、裁剪或压缩，似乎不是非常鲁棒，除非模型再次被重新训练。

与刚刚评论的看不见的条件相关，马拉等人在[[60]](#_bookmark101)这是一项有趣的研究，旨在检测未知类型的伪造生成数据。具体而言，他们提出了一种多任务增量学习检测方法，以便检测和分类新类型的GAN生成的图像，而不会恶化先前图像的性能。基于成功的算法iCaRL，针对分类器的位置提出了两种不同的解决方案

表二

不同检测方法的比较。每个公共数据库获得的最佳结果用粗体标注。斜体的结果表示它们不是在原始工作中提供的。

AUC =曲线下面积，ACC。=精度，EER =等误差率。

|  |  |  |  |
| --- | --- | --- | --- |
| **学习方法** | **分类器** | **最佳性能** | **数据库(一代)** |
| McCloskey和Albright (2018) GAN管道特征 | SVM | AUC = 70.0% | NIST MFC2018 |
| 王等人(2019) GAN管道特征 | SVM | Acc。= 84.7% | 自己的  (界面甘，风格甘) |
| 葛奈拉等人(2020年)  [[57]](#_bookmark98) GAN流水线特性 | *SVM昆士兰大学* | Acc。= 99.81% | 所有者(AttGAN，GDWCT， |
|  |  |  | StarGAN、StyleGAN、StyleGAN2) |
| Nataraj等人(2019年)隐写分析特征 | 美国有线电视新闻网 | *能效比= 12.3%*[*[16]*](#_bookmark62) | *100K面(StyleGAN)* |
| 于等(2019)  [[59]](#_bookmark100) 深度学习功能 | 美国有线电视新闻网 | Acc。= 99.5% | Own (ProGAN，SNGAN， |
|  |  |  | 克拉姆根 |
| 马拉等人(2019年)  [[60]](#_bookmark101) 深度学习功能 | CNN +增量学习 | Acc。= 99.3% | 拥有(CycleGAN，ProGAN， |
|  |  |  | Glow、StarGAN、StyleGAN) |

[[55]](#_bookmark96)

[[56]](#_bookmark97)

[[58]](#_bookmark99)

Dang等人(2020年)[[17]](#_bookmark63)

深度学习特征CNN +注意机制AUC = 100%

**EER = 0.1%**

**DFFD (ProGAN，StyleGAN)**

Neves等人(2020年)[[16]](#_bookmark62)

Hulzebosch等人(2020年)

深度学习功能CNN EER = 0.3% 100K面(StyleGAN)

**EER = 4.5% ifaefacedb**

自己的

[[61]](#_bookmark103) 深度学习功能CNN，AE Acc。= 99.8%

(StarGAN，Glow，ProGAN，StyleGAN)

增量学习[[75]:](#_bookmark116) *I)多任务多分类器(MT- MC)，以及ii)多任务单分类器(MT-SC)。关于实验框架，在研究中考虑了五种不同的GAN方法，CycleGAN*[[71],](#_bookmark112) ProGAN[[48],](#_bookmark90)洋溢[[76],](#_bookmark117) 斯塔根[[43],](#_bookmark85) 还有StyleGAN[[41].](#_bookmark83) 他们提出的基于XceptionNet模型的检测方法取得了令人鼓舞的结果，能够正确检测新的GAN生成的图像。

注意力机制也被应用于进一步改进检测系统的训练过程。Dang等人在2005年进行了[[17]](#_bookmark63)不同类型面部操作的完整分析。他们提出使用注意力机制和流行的CNN模型，如Xception- Net和VGG16。对于整个人脸合成操作，考虑到来自CelebA的真实人脸，作者实现了最终的100% AUC和大约0.1% EER[[52],](#_bookmark93)FFHQ[[41],](#_bookmark83)和FaceForensics++[[12]](#_bookmark59) 通过ProGAN创建的数据库和假图像[[48]](#_bookmark90)还有StyleGAN[[41]](#_bookmark83)方法。令人印象深刻的结果显示了新的注意机制的重要性[[77].](#_bookmark118)

内维斯等人在1999年演出[[16]](#_bookmark62)对这种类型的面部操作进行深入的实验评估，考虑不同的最先进的检测系统和实验条件，即受控和野外场景。考虑了四个不同的假脸数据库:I)在线收集的150，000张假脸[11](#_bookmark17)以及基于StyleGAN架构，ii)100K-faces公共数据库，iii)使用ProGAN生成的80，000个合成人脸，以及iFakeFaceDB数据库，这是先前伪造数据库的改进版本，其中GAN指纹信息已使用GANprintR方法移除。在受控场景中，他们实现了

11[https://thispersondoesnotexist.com](https://thispersondoesnotexist.com/)

与之前最好的研究结果相似(EER = 0.02%)。然而，在更具挑战性的场景中，其中图像(真实的和伪造的)来自不同的源(数据集的不匹配)，伪造检测性能的高度退化被观察到。最后，在他们的公共iFakeFaceDB数据库上获得的结果表明，即使对于最先进的操纵检测方法，iFakeFaceDB也具有很大的挑战性。与这种增强的伪造内容相关，Cozzolino等人在[[78]](#_bookmark119)一种基于GAN的类似方法，将相机痕迹注入合成图像，以欺骗最先进的假探测器。

与…类似[[16],](#_bookmark62) Hulzebosch等人最近在[[61]](#_bookmark103) 对这种人脸处理的深入分析考虑了不同的场景，例如跨模型、跨数据和后处理。假货检测方法是基于流行的例外网络和ForensicTransfer[[79],](#_bookmark120)这是一种自动编码器方法。一般来说，糟糕的概括结果是在看不见的场景下获得的，类似于[[16].](#_bookmark62)最后，为了完整性，我们还包括一些重要的参考文献，以其他最近的研究为重点，检测一般的基于GAN的图像处理，而不是面部处理。在…里

具体而言，我们请读者参考[[80],](#_bookmark121) [[81].](#_bookmark122)

1. 身份互换
2. *操作技术和公共数据库*

由于公众对DeepFakes的极大关注，这是当今最受欢迎的人脸操纵研究路线之一[[2],](#_bookmark52)[[3].](#_bookmark53)它包括用另一个人的脸替换视频中一个人的脸。不同于在图像级别执行的整个面部合成操作，身份交换的目标是生成逼真的假视频。

因为公开可用的假数据库，如UADFV数据库[[82],](#_bookmark123)直到最近的名人DF和DFDC数据库[[26],](#_bookmark69)[[83],](#_bookmark124)进行了很多视觉上的改进，增加了假视频的真实感。因此，身份交换数据库可以分为两种不同的

表三

身份交换:公共可用的数据库。

**第一代**

**数据库真实视频假视频**

几代人。桌子[III](#_bookmark18)总结了每个公共数据库的主要细节，按代分组。正如可以看到的，

UADFV (2018年)

[[82]](#_bookmark123)

deepfakelimit(2018年)

49 (Youtube) 49 (FakeApp)

- 620 (faceswap-GAN)

在这种类型的面部操作中，真的和假的视频

[[1](#_bookmark51)]

通常包含在数据库中。

在这一节中，我们首先提供每个数据库的主要细节，最后在更高的层次上总结关键

face forensics ++ 2019

[[12]](#_bookmark59)

1，000 (FaceSwap)

1000(深假)

1000(Youtube)

**第二代**

两代人之间的差异。

**数据库真实视频假视频**

三个不同的数据库被归入第一代。

UADFV是最早的公共数据库之一[[82].](#_bookmark123)这个数据库包括来自Youtube的49个真实视频，这些视频被用来通过FakeApp移动应用程序创建49个虚假视频[12](#_bookmark19)把他们所有人原来的脸换成了尼古拉斯·凯奇的脸。所以所有假视频都只考虑一个身份。每个视频代表一个

DeepFakeDetection (2019)

[[86]](#_bookmark127)

名人DF (2019年)

[[26]](#_bookmark69)

DFDC预告(2019)

[[83]](#_bookmark124)

363(演员)3068(深假)

890(Youtube)5639(deep fake)

1，131人(演员)4，119人(未知)

个人，典型分辨率为294 500像素，以及

*×*

平均11.14秒。

Korshunov和Marcel在1999年介绍了[[1]](#_bookmark51)Deepfake- TIMIT数据库。该数据库包括来自VidTIMIT数据库的32个对象的620个假视频[[84].](#_bookmark125)虚假视频是使用基于GAN的人脸交换算法创建的[13](#_bookmark20).在这种方法中，生成网络采用了CycleGAN的方法[[71],](#_bookmark112) 使用FaceNet的权重[[65].](#_bookmark106) 多任务级联卷积网络的方法用于更稳定的检测和可靠的人脸对准[[85].](#_bookmark126)此外，卡尔曼滤波器还被用来平滑帧间的边界框位置，并消除交换面上的抖动。关于在DeepfakeTIMIT中考虑的场景，考虑了两种不同的质量:I)具有64×64像素的图像的低质量(LQ ),以及ii)具有128×128像素的图像的高质量(HQ)。此外，根据质量水平，不同的混合技术被应用于假视频。

*×*

*×*

在这种类型的面部操作中，最流行的数据库之一是FaceForensics++数据库[[12].](#_bookmark59)该数据库于2019年初推出，作为原始面部取证数据库的扩展[[87],](#_bookmark128) 只关注表情互换。FaceForensics++包含了从Youtube上摘录的1000个真实视频。关于身份交换假视频，它们是使用计算机图形和DeepFake方法(即学习方法)生成的。对于计算机图形集成方法，作者考虑了公开可用的面交换算法[14](#_bookmark21)而对于DeepFake方法，虚假视频是通过DeepFake FaceSwap GitHub实现创建的[15](#_bookmark22).人脸交换方法包括人脸对齐、高斯牛顿优化和图像混合，以将源人的人脸交换到目标人。DeepFake方法，如[[12],](#_bookmark59) 基于具有共享编码器的两个自动编码器，这两个自动编码器被训练来重建源和目标的训练图像

12<https://www.malavida.com/en/soft/fakeapp/>13<https://github.com/shaoanlu/faceswap-GAN>14<https://github.com/MarekKowalski/FaceSwap>15<https://github.com/deepfakes/faceswap>

分别面对。人脸检测器用于裁剪和

对齐图像。为了创建假图像，将源人脸的经过训练的编码器和解码器应用于目标人脸。然后，使用泊松图像编辑将自动编码器输出与图像的其余部分混合[[88].](#_bookmark129)关于FaceForensics++数据库的数字，每个方法都生成了1000个假视频。后来，在Google的支持下，一个名为DeepFakeDetection的新数据集(由于其更高的真实性而被归入第二代)被包含在FaceForensics++框架中[[86].](#_bookmark127)该数据集包括来自16个不同场景的28个付费演员的363个真实视频。此外，基于DeepFake FaceSwap GitHub实现的数据集中包含了3068个虚假视频。重要的是要注意，对于FaceForensics++和DeepFakeDetection数据库，考虑不同级别的视频质量，特别是:i) RAW(原始质量)，ii) HQ(恒定速率量化参数等于23)，以及iii) LQ(恒定速率量化参数等于40)。这方面模拟了社交网络中通常应用的视频处理技术。

关于第二代包含的数据库，我们强调最近在2019年底发布的Celeb-DF和DFDC数据库。李等提出[[26]](#_bookmark69)名人DF数据库。这个数据库旨在提供视觉质量更好的假视频，类似于在互联网上共享的流行视频[16](#_bookmark23)相比之下，以前的数据库表现出低视觉质量，有许多可见的伪像。Celeb-DF由从Youtube上提取的890个真实视频和5639个虚假视频组成，这些视频是通过公共DeepFake生成算法的改进版本创建的，改善了合成人脸的低分辨率和颜色不一致等方面。

脸书与微软、亚马逊和麻省理工学院等其他公司和学术机构合作，在2019年底发起了一项名为深度造假检测挑战(DFDC)的新挑战[[83].](#_bookmark124) 他们首先发布了一个预览数据集，由66个付费网站的1131个真实视频组成

16[https://www.youtube.com/channel/UCKpH0CKltc73e4wh0 pgL3g](https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g)



**身份互换:第二代**

**身份互换:第一代**

|  |  |  |
| --- | --- | --- |
|  | 弱点限制了自然性，并有利于伪造检测 |  |
| 假面具中的低质量合成人脸颜色对比假面具中的可见边界  来自原始视频的可见元素帧之间的奇怪伪像 | | |

|  |  |  |
| --- | --- | --- |
|  | **增强自然度和阻止伪造检测的改进** |  |
| 场景:室内和室外光照条件:白天、夜晚等。  离摄像机的距离高姿态变化 | | |

图3。第一代身份交换数据库中存在的缺点和第二代中进行的改进的图形表示，不仅在视觉层面，而且在可变性方面(野外场景)。伪图提取自:UADFV和FaceForensics++第1代[[12],](#_bookmark59) [[82];](#_bookmark123) 名人-DF和DFDC(第二代)[[26],](#_bookmark69) [[83].](#_bookmark124)



演员，还有4119个假视频。假视频是用两种不同的未知方法生成的。完整的DFDC数据集随后发布，包含超过470 GB的内容(真实的和伪造的)[17](#_bookmark26).

最后，作为本节的总结，我们将在更高的层次上讨论第一代和第二代伪数据库之间的主要差异。通常，来自第一代的假视频的特征在于:I)低质量的合成人脸，ii)合成的假面具和原始人脸的皮肤之间不同的颜色对比度，iii)假面具的可见边界，

*iv)来自原始视频的可见面部元素，v)低姿态变化，以及vi)连续帧中的奇怪伪像。此外，他们通常会考虑相机位置和光线条件方面的受控场景。在第二代数据库中，这些方面中的许多方面都得到了成功的改进，不仅在视觉层面，而且在可变性方面(野外场景)也是如此。例如，最近的DFDC数据库考虑了不同的采集场景(即室内和室外)、光照条件(即白天、夜晚等)。)、从人到相机的距离以及姿势变化等等。图。*[3](#_bookmark24) 以图形方式总结了第一代身份交换数据库中存在的缺点以及第二代中进行的改进。最后，有趣的是，第二代数据库中包含了大量的虚假视频。

1. *操纵检测*

检测身份交换操作的新方法的开发正在不断发展。桌子[IV](#_bookmark28)提供了该领域最相关的检测方法的比较。对于每项研究，我们都包括与研究方法、分类器、最佳性能和数据库相关的信息。我们用粗体突出显示每个公共数据库获得的最佳结果。值得注意的是，在某些情况下，会考虑不同的评估指标(如AUC和EER)，这使得研究之间的比较变得复杂。最后，用斜体突出显示的结果表明检测系统对不同的未知数据库的泛化能力，即这些数据库不被考虑用于训练。这些结果摘自[[26]](#_bookmark69) 并且没有包括在原始出版物中。

这方面的第一个研究集中在第一代假视频中存在的视听伪像。Korshunov和Marcel在1999年评估了[[1]](#_bookmark51)基线方法基于嘴唇运动和音频语音之间的不一致性，以及生物识别中常用的基于图像的系统的几种变体。对于第一种情况，他们认为梅尔频率倒谱系数(MFCCs)作为音频特征，嘴部标志之间的距离作为视觉特征。然后使用主成分分析(PCA)来降低特征块的维数，最后使用基于长短期记忆(LSTM)的递归神经网络(RNNs)来检测真假视频(基于[[101]).](#_bookmark142)对于第二种情况，他们评估了基于以下方面的检测方法:I)作为特征的原始人脸，以及ii)图像质量度量(IQM)[[102].](#_bookmark143)特别是，他们用了一套

17<https://www.kaggle.com/c/deepfake-detection-challenge>

129个特征中与信噪比、镜面度、模糊度等度量相关的特征。考虑用主成分分析和LDA或SVM进行最终分类。他们提出的基于IQM+SVM的检测方法提供了最好的结果，对于DeepfakeTIMIT数据库的LQ和HQ场景，最终的EER分别为3.3%和8.9%。

按照这种思路，Matern等人在[[89]](#_bookmark130)假货检测系统基于相对简单的视觉方面，例如眼睛颜色、缺少反射以及眼睛和牙齿区域缺少细节。在该分析中考虑了两种不同的分类器:I)逻辑回归模型，和ii)多层感知器(MLP)[[103].](#_bookmark144) 他们提出的方法用一个私人数据库进行了测试，在MLP系统中获得了85.1%的AUC。

文献中还提出了基于面部表情和头部运动的假货检测系统。杨等人观察到[[90]](#_bookmark131)一些深度假像是通过将合成的面部区域拼接到原始图像中而创建的，并且在这样做的过程中，引入了当从面部图像中估计3D头部姿态时会暴露的误差。因此，他们基于使用全套面部标志(从DLib中提取的68个)估计的头部姿态之间的差异进行了研究[[104])](#_bookmark145) 以及在中央面部区域中的那些，以将深度假像与真实视频区分开。一旦这些特征被提取和标准化(平均值和标准偏差)，SVM被考虑用于最终分类。他们提出的方法最初是用UADFV数据库评估的，最终达到89.0%的AUC。然而，这个预先训练的模型(使用UADFV数据库)似乎不能很好地推广到表中描述的其他数据库[IV.](#_bookmark28)Agarwal和Farid在1995年提出了另一个有趣的方法[[91].](#_bookmark132)他们提出了一种基于面部表情和头部运动的检测系统。对于特征提取，考虑使用OpenFace2工具包[[105],](#_bookmark146)获得与面部肌肉运动相关的18个不同面部动作单元的强度和发生率，所述面部肌肉运动例如脸颊抬高、鼻子起皱、嘴伸展等。此外，还考虑了与头部运动相关的四个特征。结果，使用皮尔逊相关性来测量特征之间的线性度，将每个10秒的视频剪辑减少到维数为190的特征向量。最后，作者考虑将SVM用于最终分类。关于实验框架，作者基于从YouTube上下载的感兴趣的人在正式场合的谈话视频，例如每周讲话、新闻采访和公开演讲，建立了自己的数据库。在大多数视频中，人主要面向摄像机。关于DeepFake视频，作者基于faceswap-GAN每人训练了一个GAN[18](#_bookmark27).他们提出的方法实现了作为最佳假货检测性能的最终96.3%的AUC

对新的上下文和操作技术具有鲁棒性。

眨眼[[106]](#_bookmark147)也被用来检测假视频。在…里[[92],](#_bookmark133)作者提出了一种叫做DeepVision的算法来分析眨眼模式的变化。他们的方法是基于快超面的融合[[107]](#_bookmark148) 和眼睛长宽比(耳朵)[[108]](#_bookmark149)检测面部并获得

18<https://github.com/shaoanlu/faceswap-GAN>

表四

身份交换:不同最新检测方法的比较。每个公共数据库获得的最佳结果用粗体标注。斜体的结果表示它们发表于[[26](#_bookmark69)]，但原著中没有。FF++ = FACEFORENSICS++，AUC =曲线下面积，ACC。=准确度，EER =等错误率，TCR =真实分类率。

**学习方法分类器最佳性能数据库**

SVM PCA+LDA

科尔舒诺夫和马塞尔(2018)

[[1]](#_bookmark51)

视听功能主成分分析+RNN

**能效比= 3.3%**

**能效比= 8.9%**

**deepfakelimit(LQ)**

**DeepfakeTIMIT(总部)**

Matern等人(2019年)

逻辑回归

AUC = 85.1%自有

*AUC = 70.2% UADFV*

[[89]](#_bookmark130) 视觉特征

MLP

*AUC = 77.0%*

*AUC = 77.3%*

*deepfakelimit(LQ)*

*DeepfakeTIMIT(总部)*

*AUC = 78.0% FF++ / DFD*

*AUC = 66.2% DFDC预览版*

*AUC = 55.1% Celeb-DF*

AUC = 89.0% UADFV

杨等(2019)

|  |  |  |  |
| --- | --- | --- | --- |
| [[90]](#_bookmark131) |  | *AUC = 53.2%* | *DeepfakeTIMIT(总部)* |
|  |  | *AUC = 47.3%* | *FF++ / DFD* |
|  |  | *AUC = 55.9%* | *DFDC预览* |
|  |  | *AUC = 54.6%* | *名人DF* |
| Agarwal和Farid (2019)头部姿势和面部特征 | SVM | AUC = 96.3% | 所有者(FaceSwap，总部) |
| Jung等人(2020年)眨眼 | 距离 | Acc。= 87.5% | 自己的 |
|  |  | AUC = 97.7% | UADFV |
| 李等(2019)人脸扭曲特征CNN | | **AUC = 99.9%**  **AUC = 99.7%** | **deepfakelimit(LQ)**  **DeepfakeTIMIT(总部)** |
|  | | *AUC = 93.0%* | *FF++ / DFD* |
|  | | ***AUC = 75.5%*** | *DFDC预览* |
|  | | *AUC = 64.6%* | *名人DF* |
|  | | Acc。= 98.4% | 自己的 |
|  | | *AUC = 84.3%* | *UADFV* |

头部姿势特征SVM

*AUC = 55.1%*

*deepfakelimit(LQ)*

[[91]](#_bookmark132)

[[92]](#_bookmark133)

[[26]](#_bookmark69),[[93]](#_bookmark134)

Afchar等人(2018年)[[94]](#_bookmark135)

介观特征CNN

*AUC = 87.8%*

*AUC = 68.4%*

Acc。' 90.0%

Acc。' 94.0%

Acc。' 98.0%

Acc。' 83.0%

Acc。' 93.0%

Acc。' 96.0%

*DeepfakeTIMIT(LQ)DeepfakeTIMIT(HQ) FF++ (DeepFake，LQ) FF++ (DeepFake，HQ)ff++(DeepFake，RAW) FF++ (FaceSwap，LQ) FF++ (FaceSwap，HQ) FF++ (FaceSwap，RAW)*

隐写分析特征

*AUC = 75.3% DFDC预览版*

*AUC = 54.8% Celeb-DF*

*AUC = 85.1% UADFV*

周等(2018)

[[95]](#_bookmark136)

+

深度学习功能

美国有线电视新闻网

SVM

*AUC = 83.5%*

*AUC = 73.5%*

*deepfakelimit(LQ)*

*DeepfakeTIMIT(总部)*

*AUC = 70.1% FF++ / DFD*

*AUC = 61.4% DFDC预览版*

*AUC = 53.8% Celeb-DF*

罗斯勒等人(2019年)[[12]](#_bookmark59)

介观特征隐写分析特征深度学习特征

美国有线电视新闻网

Acc。' 94.0%

**Acc。' 98.0%**

**Acc。' 100.0%**

Acc。' 93.0%

**Acc。' 97.0%**

**Acc。' 99.0%**

ff++(LQ deep fake)

**FF++ (DeepFake，HQ) FF++ (DeepFake，RAW) FF++ (FaceSwap，LQ) FF++ (FaceSwap，HQ) FF++ (FaceSwap，RAW)**

*AUC = 65.8% UADFV*

Nguyen等人(2019年)

|  |  |  |
| --- | --- | --- |
| [[96]](#_bookmark137) | *AUC = 55.3%* | *DeepfakeTIMIT(总部)* |
|  | *AUC = 76.3%* | *FF++ / DFD* |
|  | 能效比= 15.1% | FF++ (FaceSwap，HQ) |
|  | *AUC = 53.6%* | *DFDC预览* |
|  | *AUC = 54.3%* | *名人DF* |
|  | *AUC = 61.3%* | *UADFV* |
| Nguyen等人(2019)深度学习的特点是胶囊网络 | *AUC = 78.4%*  *AUC = 74.4%* | *deepfakelimit(LQ)*  *DeepfakeTIMIT(总部)* |
|  | *AUC = 96.6%* | *FF++ / DFD* |
|  | *AUC = 53.3%* | *DFDC预览* |
|  | *AUC = 57.5%* | *名人DF* |

深度学习特色AE +多任务学习

*AUC = 62.2%*

*deepfakelimit(LQ)*

[[97]](#_bookmark138)

Dang等人(2019年)[[17]](#_bookmark63)

多尔汉斯基等人(2019年)[[83]](#_bookmark124)

王与丹切娃(2020)[[98]](#_bookmark139)

深度学习特征CNN +注意机制AUC = 99.4%深度学习特征CNN精度= 93.0%

深度学习特征3DCNN TCR = 95.13%

能效比= 3.1%

**TCR = 92.25%**

**召回率= 8.4%**

DFFD

**DFDC预览**

**ff++(LQ deep fake)ff++(LQ face swap)**

顾时代与Delp (2018)[[99]](#_bookmark140)

**AUC = 96.3%**

图像+时间特征CNN + RNN Acc。= 97.1%自有

萨比尔等人(2019年)[[98]](#_bookmark139)

图像+时间特征CNN + RNN AUC = 96.9%

**ff++(LQ deep fake)ff++(LQ face swap)**

Tolosana等人(2020年)[[100]](#_bookmark141)

面部区域特征CNN

**AUC = 100.0% UADFV**

**AUC = 99.4% FF++ (FaceSwap，HQ) AUC = 91.0% DFDC预览版**

**AUC = 83.6% Celeb-DF**

眼睛的长宽比。最后，提取基于眨眼次数和周期的特征来判断视频的真假。这种方法在专有数据库上实现了最终87.5%的准确率。

另一个有趣的研究方向是基于人脸处理管道中包含的伪像的检测。在…里[[93],](#_bookmark134)李和吕认为，一些深度伪造算法只能创建有限分辨率的图像，需要进一步扭曲以匹配源视频中的原始人脸。这种转换在最终的DeepFake视频中留下了独特的人工痕迹。因此，作者提出了一种基于CNN的检测系统，以便从检测到的面部区域和周围区域检测这种伪影的存在。从头开始训练了四个不同的CNN模型:VGG16[[109],](#_bookmark150)ResNet50、ResNet101和ResNet152[[110].](#_bookmark151)他们提出的检测方法使用UADFV和DeepfakeTIMIT数据库进行了测试，优于这些数据库的最先进水平。

李等人后来在1995年提出[[26]](#_bookmark69) 中提出的工作的改进版本[[93].](#_bookmark134) 在这种情况下，作者包括了一个新的空间金字塔池模块，以更好地处理分辨率的变化[[111].](#_bookmark152) 这种检测方法使用不同的数据库进行评估，在其中一些数据库中实现了最先进的结果。

文献中还提出了基于介观和隐写分析特征的方法。Afchar等人在[[94]](#_bookmark135) 由几层组成的两个不同的网络，以便集中于图像的介观属性:I)由4个卷积层和随后的全连接层(Meso-4)组成的CNN网络，以及ii)由中介绍的初始模块的变体组成的Meso-4的修改[[112],](#_bookmark153) 名为《盗梦空间4》。他们提出的方法最初是使用一个私人数据库对DeepFakes进行测试的，实现了98.4%的假货检测准确率，达到了最佳性能。该预训练的检测模型在中针对未知数据库进行了测试[[26],](#_bookmark69) 在某些情况下证明是一种健壮的方法，例如使用FaceForensics++时。

周等人提出了一种用于人脸操纵检测的双流网络。特别地，作者考虑了两个流的融合:I)基于CNN GoogLeNet的人脸分类流[[112]](#_bookmark153) 以检测人脸图像是否是伪造的，以及ii)路径三元组流，其使用具有三元组丢失的图像补片的隐写分析特征和用于分类的SVM来训练。最初的系统被训练来检测表情互换操作。尽管如此，李等人在1999年评价[[26]](#_bookmark69) 预训练模型(使用SwapMe应用程序训练)检测身份互换操作的泛化能力，是针对最近的Celeb-DF数据库的最稳健的方法之一[[26].](#_bookmark69)

Ro ssler等人使用FaceForensics++数据库对不同的假货检测方法进行了详尽的分析[[12].](#_bookmark59)评估了五种不同的检测系统:I)通过手工隐写分析特征训练的基于CNN的系统[[113],](#_bookmark154)ii)基于CNN的系统，其卷积层是专门设计来抑制图像的高级内容的[[114],](#_bookmark155) *iii)基于CNN的系统，具有计算四个统计值(平均值、方差、最大值和最小值)的全局池层*[[115],](#_bookmark156)㈣

CNN MesoInception-4检测系统描述于[[94],](#_bookmark135)最后v)基于CNN的系统XceptionNet[[116]](#_bookmark157) 使用ImageNet数据库进行预训练[[117]](#_bookmark158) 并且针对面部操纵检测任务被重新训练。总的来说，基于XceptionNet架构的检测系统在DeepFakes和FaceSwap这两种类型的操作方法中都提供了最好的结果。此外，考虑不同的视频质量水平来评估检测系统，以便模拟许多社交网络的视频处理。在这个真实场景中，当降低视频质量时，所有检测系统的精度都下降，这表明在真实场景中这项任务是多么具有挑战性。

最近在计算机视觉中考虑的深度学习方法已经被应用于进一步改进身份交换操作的检测。在…里[[96],](#_bookmark137) Nguyen等人提出了一个CNN系统，该系统使用多任务学习来同时检测虚假视频并定位被操纵的区域。他们考虑了一种基于自动编码器的检测系统。具体来说，他们建议使用Y形解码器，以便在分类、分割和重建任务之间共享有价值的信息，通过减少损失来提高整体性能。他们提出的方法用FaceForensics++数据库的FaceSwap操作方法进行了评估[[87],](#_bookmark128) 达到15.07% EER的最佳性能，远远超过其他检测方法。此外，这个模型似乎不能很好地推广到其他数据库，结果低于80%的AUC。

后来，同样的作者在[[97]](#_bookmark138)一种基于新型胶囊网络的防伪系统。这种方法比传统的CNN使用更少的参数，具有相似的性能[[118]–[120].](#_bookmark160)所提出的检测系统最初使用准确率高于90%的FaceForensics++数据库进行评估。相同的预训练检测模型在中针对未见过的数据库进行了测试[[26],](#_bookmark69)表现出较差的概括结果，正如在大多数伪造检测系统中发生的那样。

注意力机制也被应用于进一步改进检测系统的训练过程。Dang等人在[[17]](#_bookmark63) 对不同面部操作的全面分析。他们提出了一种基于CNN和注意力机制的检测系统，来处理和改进分类器模型的特征图。他们提出的注意力地图可以很容易地实现，并通过包含单个卷积层、其相关的损失函数和掩蔽随后的高维特征，插入到现有的主干网络中。他们提出的检测方法用DFFD数据库进行了测试(基于以前的FaceForensics++数据库和来自互联网的视频集的组合)。特别是，对于身份互换检测，他们提出的方法实现了99.43%的AUC和3.1%的EER。尽管由于考虑了不同的实验方案，很难在研究之间提供公平的比较，但很明显，他们的检测方法提供了最先进的结果。

在…里[[83],](#_bookmark124) 除了DFDC数据库的描述之外，作者提供了使用三个简单检测系统的基线结果:I)由6个卷积层和1个全连接层组成的小型CNN模型，用于检测低电平

图像操作，ii)仅使用面部图像训练的XceptionNet模型，以及iii)使用完整图像训练的XceptionNet模型。基于XceptionNet的检测系统，仅考虑人脸图像(非全图)，提供了93.0%的精度和8.4%的召回率的最佳结果。

研究了基于3DCNN的深度学习方法[[121]](#_bookmark161) 以便考虑空间和运动信息。特别地，作者提出了基于I3D的伪检测器[[122]](#_bookmark162) 和3D ResNet[[123]](#_bookmark163) 方法，在FaceForensics++的低质量视频上取得了令人满意的结果。

文献中还研究了不仅基于图像级特征，而且基于视频帧的时间级特征的检测系统。顾时代和德尔普提出于[[99]](#_bookmark140) 自动检测虚假视频的时间感知管道。他们考虑了CNN和RNNs的组合。对于CNN，作者使用了InceptionV3[[124]](#_bookmark164) 使用ImageNet数据库进行预训练[[117].](#_bookmark158)对于RNN系统，他们考虑了一个由2048个内存块组成的隐藏层的LSTM模型。最后，包含了两个完全连接的层，提供了帧序列是真是假的概率。他们的方法使用专有数据库进行评估，最终准确率为97.1%。

在这方面，Sabir等人提出了一种基于使用流中存在的时间信息来检测虚假视频的方法[[98].](#_bookmark139) 这种模型背后的直觉是利用跨帧的时间差异。因此，他们认为递归卷积网络类似于[[99],](#_bookmark140) 在本研究中接受端到端培训，而不是使用预先培训的模型。他们提出的检测方法通过FaceForen- sics++数据库进行了测试，对于DeepFake和FaceSwap方法，分别实现了96.9%和96.3%的AUC结果。在分析中只考虑了低质量的视频。

最后，研究了每个面部区域检测虚假视频的辨别能力[[100].](#_bookmark141) 作者考虑了一个基于XceptionNet的假货检测系统。在实验框架中考虑了来自第一代和第二代的数据库，结论是与第一代相比，在第二代的最新DeepFake视频数据库中实现了较差的假货检测结果，对于DFDC预览和Celeb-DF数据库的结果分别为91.0%和83.6% AUC。重要的是要强调，与[[26],](#_bookmark69)为每个数据库专门训练了一个单独的假货检测系统。

总之，尽管文献中提出了许多不同的方法，但它们对看不见的数据库都显示出较差的概括结果，如表中所示[IV.](#_bookmark28)此外，我们还强调了大多数方法在第二代DeepFake数据库上实现的较差检测结果，其结果低于60% AUC。

1. 属性操作
2. *操作技术和公共数据库*

这种面部处理包括在图像中修改面部的一些属性，例如头发或皮肤的颜色、性别、年龄、添加眼镜等。尽管基于GAN的框架在一般图像翻译和操作方面取得了成功[[43],](#_bookmark85)[[71],](#_bookmark112)[[125]–[129],](#_bookmark167)和

特别是对于面部属性操作[[43],](#_bookmark85)[[44],](#_bookmark86)[[68],](#_bookmark109)[[130]–[134],](#_bookmark170) 据我们所知，这一领域的研究很少有公开的数据库。主要原因是大多数GAN方法的代码都是公开的，因此研究人员可以很容易地生成他们自己的假数据库。因此，本节旨在强调该领域最新的GAN方法，从旧的到最近的，并提供相应代码的链接。

在…里[[130],](#_bookmark168)作者介绍了可逆条件遗传算法(IcGAN)[19](#_bookmark29) 用于复杂图像编辑，作为与条件GAN (cGAN)一起使用的编码器的联合[[135].](#_bookmark171)这种方法在属性操作方面提供了准确的结果。但是，它严重地改变了人的面部身份。

Lample等人在1999年提出[[133]](#_bookmark169) 一种编码器-解码器架构，其被训练成通过直接在潜在空间中解开图像的突出信息和属性值来重建图像[20](#_bookmark30).然而，正如IcGAN方法所发生的那样，生成的图像可能缺少一些细节或出现意外的失真。

一种叫做StarGAN的增强型方法[21](#_bookmark31)被提议于[[43].](#_bookmark85)在StarGAN方法之前，许多研究已经显示了一般两个领域的图像到图像翻译的有希望的结果。然而，很少有研究专注于处理两个以上的领域。在这种情况下，直接的方法是为每对图像域独立地建立不同的模型。StarGAN提出了一种新的方法，只使用一个模型就可以为多个领域执行图像到图像的翻译。作者通过属性分类损失和循环一致性损失训练了一个条件属性转移网络。与以前的方法相比，取得了良好的视觉效果。然而，它有时包括来自输入人脸图像的不期望的修改，例如肤色。

几乎在同一时间，何等人提出[[68]](#_bookmark109)阿特甘[22](#_bookmark32)这是一种新颖的方法，它从潜在表示中去除了严格的属性无关约束，只将属性分类约束应用于生成的图像，以保证属性的正确变化。AttGAN在逼真的属性操作上提供了最先进的结果，同时保留了其他面部细节。

文献中提出的最新方法之一是STGAN[23](#_bookmark33) [[44].](#_bookmark86) 一般来说，属性操作可以通过结合编码器-解码器或g an来实现。然而，正如刘等人所评论的。[[44],](#_bookmark86)编码器-解码器中的瓶颈层通常提供模糊和低质量的操作结果。为了改善这一点，作者提出并结合了选择性传输单元与编码器-解码器，以同时改善属性操作能力和图像质量。因此，STGAN最近在属性操作方面超越了最先进的技术。

尽管大多数属性操作方法的代码是公开可用的，但是缺乏

19<https://github.com/Guim3/IcGAN>

20<https://github.com/facebookresearch/FaderNetworks>

21<https://github.com/yunjey/stargan/blob/master/README.md>22<https://github.com/LynnHo/AttGAN-Tensorflow>

23<https://github.com/csmliu/STGAN>

表五

**属性操作:不同最新检测方法的比较。每个公共数据库获得的最佳结果用粗体标注。AUC =曲线下面积，ACC。=精度，EER =等误差率。**

**学习方法分类器最佳性能数据库(代)**

王等(2019)

[[56]](#_bookmark97)

纳塔拉杰等人(2019年)[[58]](#_bookmark99)

局部骨质量

GAN管道以SVM Acc为特色。= 84.7%自有

隐写分析功能CNN Acc。= 99.4%自有

(StarGAN/CycleGAN)

(接口根/样式根)

Bharati等人(2016年)[[136]](#_bookmark172)

贾恩等人(2019年)[[137]](#_bookmark173)

塔里克等人(2018年)

深度学习功能(面部贴片)

深度学习功能(面部贴片)

总体Acc。=总体符合率为96.2%。= 87.1%

总体Acc。= 99.6%的整体Acc。= 99.7%

CNN + SVM

AUC = 99.9%

Own(名人修图，ND-IIITD修图)

自己的

(ND-IIITD修图，StarGAN)

自己的

[[138]](#_bookmark174) 深度学习功能CNN

AUC = 74.9%

(ProGAN，

Adobe Photoshop)

**EER = 1.0%**

Dang等人(2019年)[[17]](#_bookmark63)

深度学习特征CNN +注意机制AUC = 99.9%

**DFFD(Facebook/StarGAN)**

王等(2019)[[139]](#_bookmark175)

马拉等人(2019年)[[60]](#_bookmark101)

张等(2019)[[140]](#_bookmark176)

Rathgeb等人(2020年)[[141]](#_bookmark177)

深度学习功能DRN AP = 99.8%拥有深度学习功能CNN +增量学习Acc。= 99.3%自有

频谱域特性GAN鉴频器Acc。= 100%自有PRNU功能评分级融合EER = 13.7%自有

(5个公共应用程序)

(辉光/StarGAN)

(Adobe Photoshop)

(StarGAN/CycleGAN)

在比较不同的操纵检测方法时，公共数据库和实验方案的结果至关重要。否则，不可能在研究之间进行公平的比较。到目前为止，就我们所知，DFFD数据库[[17]](#_bookmark63)似乎是唯一一个考虑这种面部处理的公共数据库。该数据库包括分别通过FaceApp和StarGAN方法生成的18，416和79，960幅假图像。

1. *操纵检测*

属性操作最初是在人脸识别领域进行研究的，目的是了解生物特征识别系统对整形手术、化妆品、化妆或遮挡等物理因素的鲁棒性如何[[142]–[147].](#_bookmark179)然而，最近像FaceApp这样的移动应用程序的成功激发了研究界对数字人脸属性操作的检测。桌子[V](#_bookmark34)提供了该领域最相关方法的比较。我们为每项研究提供了与研究方法、分类器、最佳性能和数据库相关的信息。

一些作者建议分析内部GAN管道，以检测真实图像和操纵图像之间的不同伪像。类似于整个人脸合成操作，王等人在[[56]](#_bookmark97)由于逐层神经元激活模式可以捕获对面部操纵检测系统很重要的更细微的特征，因此监控神经元行为也可以作为检测假脸的资产。他们提出的方法名为FakeSpoter，从深度人脸识别系统(VGG人脸)中提取真实和虚假人脸的神经元覆盖行为作为特征[[63],](#_bookmark104) OpenFace[[64],](#_bookmark105) 和FaceNet[[65]),](#_bookmark106) 然后为最终分类训练一个SVM。作者测试了

他们提出的方法使用了来自CelebA-HQ的真实面孔[[48]](#_bookmark90)和FFHQ[[41]](#_bookmark83) 通过界面生成的数据库和合成人脸[[66]](#_bookmark107) 还有StyleGAN[[41],](#_bookmark83) 为了获得最佳性能，使用FaceNet模型实现了最终84.7%的操纵检测准确度。

受隐写分析启发的伪造检测系统也已经被研究。如第节所述。[III-B](#_bookmark15)对于整个人脸合成，Nataraj等人在[[58]](#_bookmark99)基于像素共生矩阵和CNN相结合的检测系统。他们使用StarGAN方法基于属性操作创建了一个新的假数据集[[43]](#_bookmark85) 通过CelebA数据库进行培训[[52],](#_bookmark93)达到最佳结果的最终99.4%的准确度。

许多研究也专注于纯粹的深度学习方法，要么向网络提供面部补丁，要么提供完整的面部。在…里[[136],](#_bookmark172)Bharati等人提出了一种基于受限玻尔兹曼机器(RBM)的深度学习方法，以检测人脸图像的数字修饰。检测系统的输入由人脸图像组成，以学习区分特征来将每个图像分类为原始图像或修饰图像。关于数据库，作者从原始ND-IIITD数据库生成了两个假数据库(集合B[[148])](#_bookmark180) 和一组从网上下载的名人面部图像。假图像是使用专业软件PortraitPro Studio Max生成的[24](#_bookmark36)考虑的因素包括皮肤纹理、眼睛形状、鼻子、嘴唇和整个面部、微笑的突出度、嘴唇形状和眼睛颜色。他们提出的方法在名人和ND-IIITD修饰数据库中分别实现了96.2%和87.1%的操纵检测总准确率。

24<https://www.anthropics.com/portraitpro/>

中提出了一种基于非重叠面补片的类似方法[[137].](#_bookmark173)Jain等人提出了一种由6个卷积层和2个全连接层组成的CNN特征提取器。此外，剩余连接被认为是受ResNet架构的启发[[110].](#_bookmark151)最后，使用SVM进行最终分类。关于实验框架，在[[136]](#_bookmark172)被认为是。此外，作者考虑了通过StarGAN方法创建的假图像[[43],](#_bookmark85) 使用CelebA数据库进行培训[[52].](#_bookmark93) 总的来说，在两种操纵方法中都获得了良好的检测结果，实现了几乎100%的操纵检测准确性。

基于完整人脸的深度学习方法在文献中得到进一步研究，总体上取得了非常好的结果。塔里克等人在2005年评估了[[138]](#_bookmark174) 使用不同的CNN架构，如VGG16[[63],](#_bookmark104)VGG19[[63],](#_bookmark104)雷斯内特[[110],](#_bookmark151)或XceptionNet[[116],](#_bookmark157)除其他外。对于真实的人脸图像，CelebA数据库[[52]](#_bookmark93)被用过。关于伪图像，考虑了两种不同的方法:I)基于GAN的机器方法，特别是程序方法[[48],](#_bookmark90) 以及ii)基于Adobe Photo- shop CS6的手动方法，包括诸如化妆、眼镜、太阳镜、头发和帽子之类的操作。对于实验评估，考虑了不同大小的图像(从32×32到256×256像素)。对于机器创建的场景，获得了最终99.99%的AUC，而对于人类创建的场景，对于最佳CNN模型，该值降低到最终74.9%的AUC。因此，在机器和人类创建的假图像之间观察到操纵检测性能的高度退化。

*×*

*×*

注意力机制也被应用于进一步改进检测系统的训练过程。如前所述，Dang等人在[[17]](#_bookmark63)一个能够检测不同类型假货的系统。他们使用注意机制来处理和改进CNN模型的特征图。关于属性操作，考虑了两种不同的方法:I)通过公共FaceApp软件创建的假图像，具有多达28种不同的可用过滤器，考虑诸如头发、年龄、眼镜、胡须和肤色等方面；以及ii)通过StarGAN方法创建的假图像[[43],](#_bookmark85)多达40种不同的过滤器。他们提出的方法使用他们的新型数据库DFFD进行了测试，取得了接近1.0% EER(和99.9% AUC)的非常好的结果。

王等人进行了年[[139]](#_bookmark175)一项有趣的研究使用了Adobe Photoshop的公开商业软件(人脸识别液化工具[[149])](#_bookmark181)为了合成新面孔，还需要一位专业艺术家来处理50张真实的照片。作者开始通过亚马逊机械土耳其公司(AMT)进行人体研究，向参与者展示真实和虚假的图像，并要求他们将每张图像分类。取得的结果表明这项任务对人类来说是多么具有挑战性，最终准确率为53.5%，接近概率(50%)。在人类研究之后，作者提出了两种不同的自动模型:I)基于扩张残差网络(DRN)的全局分类模型，以预测人脸是否具有

是否被扭曲，以及ii)基于光流场的局部扭曲预测器，以便识别操纵发生的位置，并反转它们。中提出的PWC-Net方法[[150]](#_bookmark182)被认为是计算从原始到操纵的流量，反之亦然。自动和手动人脸合成操作的性能分别达到99.8%和97.4%。

工作[[60]](#_bookmark101) 由Marra等人完成，也在第二节中描述。[III-B](#_bookmark15)当新的gan出现在网络中时，能够正确地执行辨别，并且基于XceptionNet模型，其提出的操纵检测方法实现了99.3%的准确度。

基于从频谱域提取的特征而不是原始图像像素的检测系统由张等人在[[140].](#_bookmark176) 给定一幅图像作为输入，他们对每个RGB通道应用2D DFT，得到每个通道一幅频率图像。关于分类器，他们提出了AutoGAN，这是一种GAN模拟器，可以在任何图像中合成GAN伪影，而无需访问任何预先训练的GAN模型。使用看不见的GAN模型测试了他们提出的方法的推广能力。特别是斯塔根[[43]](#_bookmark85) 高根呢[[126]](#_bookmark166)在评估中被考虑。对于StarGAN方法，使用频域(100%)获得了良好的检测结果。然而，对于GauGAN方法，观察到系统性能的高度退化，50%的精度。作者声称，这是由于GauGAN的发生器与CycleGAN(用于训练)的发生器完全不同。

最后，Rathgeb等人在[[141]](#_bookmark177) 基于光响应非均匀性(PRNU)的检测系统。具体地说，融合了从跨图像单元的PRNU模式提取的空间和光谱特征的分析中获得的分数。他们提出的方法在使用5种不同的移动应用程序创建的私有数据库上进行评估，在操纵检测方面实现了平均13.7%的EER。

总结这一部分，我们可以看到，大多数属性操作检测系统的核心都是基于深度学习技术的，总体上提供了接近100%准确性的非常好的结果，如表中所示[V.](#_bookmark34) 这主要是由于伪图像中存在GAN指纹信息而产生的。然而，如在整个面部合成操作中所指出的，最近的研究已经在文献中提出从假图像中去除这种GAN指纹，同时保持非常逼真的外观[[16],](#_bookmark62)[[78],](#_bookmark119) 这甚至对最先进的操纵检测器来说也是一个挑战。

1. 表情互换
2. *操作技术和公共数据库*

这种操作也称为面部重现，包括修改人的面部表情。我们重点介绍最受欢迎的技术Face2Face和NeuralTextures，它们用另一个人(也在一个视频中)的面部表情替换视频中一个人的面部表情。据我们所知，这方面唯一可用的数据库是FaceForensics++数据库[[12],](#_bookmark59) 人脸取证的扩展[[87].](#_bookmark128)

最初，FaceForensics数据库侧重于Face2Face方法[[45].](#_bookmark87)这是一种计算机图形方法，它将源视频的表达转移到目标视频，同时保持目标人的身份。这是通过手动选择关键帧实现的。具体而言，每个视频的第一帧用于获得临时面部身份(即，3D模型)，并在剩余帧中跟踪表情。然后，通过将每帧的源表情参数(即76个Blendshape系数)转移到目标视频来生成假视频。后来，这些作者在FaceForen- sics++中提出了一种基于神经文本的新的学习方法[[46].](#_bookmark88)这是一种使用原始视频数据来学习目标人的神经纹理的渲染方法，包括渲染网络。特别是，作者在其实现中考虑了Pix2Pix中使用的基于补丁的GAN损耗[[126].](#_bookmark166)只修改了嘴对应的面部表情。需要注意的是，所有数据都可以在FaceForensics++ GitHub上获得[25](#_bookmark38).总共有1000个从Youtube上提取的真实视频。关于被操纵的视频，有2000个假视频可用(每个被考虑的假接近有1000个视频)。此外，重要的是强调考虑不同的视频质量等级，特别是:i) RAW(原始质量)，ii) HQ(等于23的恒定速率量化参数)，以及iii) LQ(等于40的恒定速率量化参数)。这方面模拟了社交网络中通常应用的视频处理技术。

除了在视频级别的表情交换操作中考虑的Face2Face和NeuralTexture技术之外，最近还提出了不同的方法来改变图像和视频中的面部表情。一个非常流行的方法出现在[[151].](#_bookmark183)Averbuch-Elor等人提出了一种使用不同主题的视频自动动画化静止肖像的技术，将视频主题的表现力转移到目标肖像。与需要来自输入和目标人脸的视频的Face2Face和NeuralTexture方法不同，在[[151]](#_bookmark183) 只需要目标的图像。在这方面，最近提出了一种新的方法[[152],](#_bookmark184)在一次和几次学习中都提供了非常好的结果。

最后，我们还重点介绍了图像级的其他流行方法。比如FaceApp等移动应用[26](#_bookmark39)允许轻松地改变微笑的程度，从开心到生气。这些方法基于当前的GAN架构。例如，Choi等人在[[43]](#_bookmark85)StarGAN将输入图像改变为不同表情水平的潜力，如生气、高兴、中性、悲伤、惊讶和恐惧。其他最近的改进伪图像的图像质量和参数的控制编辑的方法是InterFaceGAN[[66],](#_bookmark107)乌甘[[153],](#_bookmark185)斯特甘[[44],](#_bookmark86) 还有阿特根[[68].](#_bookmark109)

25<https://github.com/ondyari/FaceForensics>

26<https://apps.apple.com/gb/app/faceapp-ai-face-editor/id1180884341>

1. *操纵检测*

本节旨在使用FaceForensics++数据库提供视频级别的表情交换检测器的概述，因为就我们所知，这是该领域研究的唯一公开可用的数据库。可以使用第节中描述的相同方法检测图像级别(非视频)的操作。[III-B](#_bookmark15) 和[V-B.](#_bookmark35)

桌子[VI](#_bookmark40)提供了表达式交换检测领域中最相关的方法的比较。对于每项研究，我们都包括与方法、分类器、最佳性能和数据库相关的信息。我们用粗体突出显示了唯一的公共数据库FaceForensics++所取得的最佳结果。值得注意的是，在某些情况下，会考虑不同的评估指标(如AUC和EER)，这使得很难在研究之间进行公平的比较。

下面的一些方法已经在第节中讨论过了。[IV-B](#_bookmark25)为了身份互换。在这里，我们总结了他们在检测表达交换操作方面取得的成果。初步研究集中在虚假视频中存在的视觉特征，如眼睛的颜色、缺少反射等。在…里[[89]](#_bookmark130) Matern等人使用FaceForensics++测试了所提出的方法，只有Face2Face操作技术，实现了最佳性能的最终86.6% AUC。文献中也研究了基于介观和隐写分析特征的方法。在…里[[94],](#_bookmark135) 使用来自FaceForensics++数据库的Face2Face假视频对所提出的方法进行了测试[[12],](#_bookmark59) 总体上取得了良好的效果，尤其是对于原始质量的视频。同样的方法后来在1996年进行了测试[[12]](#_bookmark59) 反对虚假视频，

与Face2Face相比，获得了较低精度的结果。

最近的深度学习方法也被应用，效果很好。在…里[[12],](#_bookmark59)基于XceptionNet的检测系统在Face2Face和NeuralTextures操作中提供了最好的结果，接近100%的原始质量。此外，考虑不同的视频质量水平来评估检测系统，以便模拟许多社交网络的视频处理。在这个真实场景中，所有检测系统的准确性都随着视频质量而降低，正如在身份交换操作中发生的那样。

在…里[[96],](#_bookmark137) 使用FaceForensics++数据库对所提出的基于多任务学习的方法进行了评估。对于Face2Face方法，在HQ视频上实现了7.1%的EER，而对于NeuralTexture方法，在操纵检测中EER增加了一点，最终达到7.8%。

注意机制最近在[[17]](#_bookmark63)进一步改进培训过程。使用DFFD数据库测试了所提出的检测方法，该数据库对于表达式交换操作仅基于来自FaceForensics++数据库的数据。所提出的方法实现了AUC = 99.4%和EER = 3.4%。

研究了基于3DCNN的深度学习方法[[121]](#_bookmark161)以便考虑空间和运动信息。类似于身份交换操作，作者提出了基于I3D的伪造检测器[[122]](#_bookmark162)和3D ResNet[[123]](#_bookmark163) 方法，在FaceForensics++数据库的低质量视频上取得了有希望的结果。

表六

表情互换:不同最新检测方法的比较。每个公共数据库获得的最佳结果用粗体标注。FF++ = FACEFORENSICS++，AUC =曲线下面积，ACC。=准确率，EER =等差错率，TCR =真实分类率。

**学习方法分类器最佳性能数据库(代)**

Matern等人(2019年)

[[89]](#_bookmark130)

视觉特征逻辑回归，MLP AUC = 86.6% FF++ (Face2Face，RAW)

Acc。= 83.2% FF++ (Face2Face，LQ)

Acc。= 93.4% FF++ (Face2Face，HQ)

Afchar等人(2018年)[[94]](#_bookmark135)

罗斯勒等人(2019年)

介观特征CNN

介观特征

Acc。= 96.8% FF++ (Face2Face，RAW) Acc。' 75% FF++(神经纹理，LQ) Acc . '85% FF++(神经纹理，HQ) Acc。' 95% FF++(神经纹理，原始)Acc '91% FF++ (Face2Face，LQ)

**Acc。98% FF++ (Face2Face，HQ)**

[[12]](#_bookmark59)

隐写分析具有深度学习功能

美国有线电视新闻网

**Acc。100% FF++ (Face2Face，RAW) Acc。81% FF++(神经纹理，LQ) Acc。' 93% FF++(神经纹理，HQ) Acc '99% FF++(神经纹理，原始)**

Nguyen等人(2019年)[[96]](#_bookmark137)

EER = 7.1% FF++ (Face2Face，HQ)

EER = 7.8% FF++(神经结构，总部)

Dang等人(2020年)[[17]](#_bookmark63)

深度学习功能自动编码器

王与丹切娃(2020)[[98]](#_bookmark139)

深度学习特征CNN +注意机制AUC = 99.4%

深度学习特征3DCNN TCR = 90.27%

**能效比= 3.4%**

TCR = 80.5%

**FF++ (Face2Face，-) FF++ (Face2Face，LQ)**

**FF++(神经纹理，LQ)**

萨比尔等人(2019年)[[98]](#_bookmark139)

阿梅里尼等人(2019年)[[154]](#_bookmark186)

图像+时间特征CNN + RNN Acc。= 94.3 FF++ (Face2Face，LQ)

图像+时间特征CNN +光流Acc。= 81.6% FF++ (Face2Face，-)

另一条有趣的线索是基于对图像和时间信息的分析。在…里[[98],](#_bookmark139) 使用FaceForensics++数据库测试了所提出的基于递归卷积网络的方法，对于Face2Face技术实现了94.3%的AUC结果。在分析中只考虑了低质量的视频。最后，在[[154],](#_bookmark186)Amerini等人提出采用光流场来利用可能的帧间差异，使用PWC-Net方法[[150].](#_bookmark182)光流是在连续帧中计算的矢量场，用于提取场景中的表观运动。这种方法的使用是有动机的，因为由于嘴唇、眼睛等的异常运动，假视频应该具有不自然的光流。使用VGG16和ResNet50网络获得了初步结果，获得了Acc。= 81.6%，以实现操纵检测的最佳性能。

最后，如前所述，这里报道的大多数用于表达交换检测的方法也已经用于身份交换检测。[IV-B.](#_bookmark25) 一般来说，伪造检测器似乎可以学习相似的特征来区分真实和伪造的内容，在两种类型的操作中都取得了良好的结果。我们强调了新技术的潜力，如在训练过程中更好地指导网络的注意机制，如所示[[17],](#_bookmark63) 检测身份互换和表达互换操作的AUC结果达到99.4%。

1. 其他面部操作方向

前面描述的四类人脸处理技术是最近几年最受关注的技术，但它们并不能完全代表所有的技术

可能的面部操作。本节讨论了人脸处理中其他一些具有挑战性和危险性的方法:人脸变形、人脸去识别和基于音频或文本的人脸合成(即音频到视频和文本到视频)。

1. *面部变形*

人脸变形是一种人脸操作，可用于创建模拟两个或更多个人的生物特征信息的人工生物特征人脸样本[[155],](#_bookmark187)[[156].](#_bookmark188)这意味着新的变形人脸图像将成功地通过这两个或更多个人的人脸样本的验证，从而对人脸识别系统造成严重威胁[[157],](#_bookmark189) [[158].](#_bookmark190) 从这个意义上说，与本次调查中涵盖的四种主要类型相比，面部变形是一种不同类型的面部操作。此外，值得注意的是，人脸变形主要集中在图像级创建假样本，而不是视频，如身份交换操作。

最近在人脸变形领域有大量的研究。Scherhag等人已经发表了该领域的非常完整的综述。[[156]](#_bookmark188) 在2019年包括变形技术和变形攻击检测器。尽管有大量的出版物，这一领域的研究仍处于起步阶段，有许多公开的问题和挑战。重要的是要强调缺乏公开可用的数据库和基准，这使得很难在研究之间进行公平的比较。为了克服这一点，Raja等人最近提出了一个有趣的变形攻击检测框架[[159],](#_bookmark191)包括公开可用的数据库、评估平台和基准-

标记[27](#_bookmark42).该数据库包括由150名受试者的1，800张照片组成的变形图像和真实图像。变形图像使用6种不同的算法生成，呈现了各种各样的可能方法。

关于人脸变形检测器，在文献中已经基于不同的特征提出了不同的方法，例如:由于混合操作而减少人脸细节[[160],](#_bookmark192) 传感器模式噪声的傅立叶频谱[[161],](#_bookmark193) 面部标志之间的差异[[162],](#_bookmark194) [[163],](#_bookmark195) 和纯深度学习功能[[164],](#_bookmark196)[[165].](#_bookmark197)此外，为了恢复共犯的面部图像，已经研究了基于面部变形的方法[[166],](#_bookmark198) [[167].](#_bookmark199)

1. *人脸去识别*

人脸去识别(de-ID)的主要目的是去除人脸图像或视频上的身份信息，以保护个人隐私[[168].](#_bookmark200)这可以通过几种方式实现。最简单的方法可以是通过模糊或像素化来模糊面部(例如，在谷歌地图街景中)。更复杂的方法试图提供具有不同身份的人脸图像，但保持所有其他因素(姿势、表情、照明等)。)不变。所以face de-ID的概念很笼统。实现面部去识别的一个可能的选择是通过面部身份交换。

该领域的早期工作是基于将人脸去ID应用于静止图像。在…里[[169]](#_bookmark201)Gross等人提出了一个de-ID的多因素框架，它结合了线性、双线性和二次模型。他们表明，他们的方法能够保护隐私，同时保留表情变化人脸数据库的数据效用。最近，基于生成性深度神经网络(特别是GAN)的图像合成方法的发展已经激发了新的人脸识别方法，例如[[170]–[175],](#_bookmark203)它使用合成的人脸来代替原来的人脸。还有，在[[176],](#_bookmark204)作者建议使用半对抗性网络(SAN)来混淆任意基于人脸的性别分类器。

最近，在[[177]](#_bookmark205)Gafni等人在2019年提出了一种方法，即使在无约束的视频中，该方法也可以提供令人信服的性能。他们的方法是基于一个对抗的自动编码器和一个经过训练的人脸分类器。这样他们可以获得丰富的潜在空间，嵌入身份和表达信息。还有，在[[178]](#_bookmark206) 提出了一种基于深度迁移模型的人脸识别方法。该方法将非身份相关的人脸属性视为原始人脸的风格，并使用训练好的人脸属性转移模型来提取和映射它们到不同的人脸，在图像和视频中都取得了非常有前景的结果。

该领域的一些其他相关研究通过消除不需要的或受保护的信息，如身份、性别或面部表情，直接在面部表征或深度面部模型上工作[[179]–[181].](#_bookmark208)一旦受保护的信息被解开，就可以基于新的表示生成人脸图像或视频，其中受保护的信息已经被消除、减少或模糊。

27<https://biolab.csr.unibo.it/fvcongoing>

1. *音频到视频和文本到视频*

面部表情交换的一个相关主题是从音频或文本合成视频。这些类型的视频面部操作也被称为唇-sinc深度假动作[[182].](#_bookmark209)流行的例子可以在网上看到[28](#_bookmark43)[29](#_bookmark44).

关于从音频合成假视频(音频到视频)，Suwajanakorn等人在[[125]](#_bookmark165) 一种合成高质量视频的方法，该视频是一个人(在这种情况下是奥巴马)以准确的唇形同步说话。为此，他们使用了这个人之前几个小时的视频和一段新的录音作为输入。在他们的方法中，他们采用了一种递归神经网络(基于LSTMs)来学习从原始音频特征到口型的映射。然后，基于每一帧的嘴部形状，他们合成了高质量的嘴部纹理，并将其与3D姿势匹配进行合成，以创建新的视频来匹配输入的音频轨道，从而产生逼真的效果。

在…里[[183],](#_bookmark210) Song等人提出了一种基于新的条件递归生成网络的方法，该网络在递归单元中结合了图像和音频特征以获得时间依赖性，并且还结合了一对时空鉴别器以获得更好的图像/视频质量。因此，他们的方法可以将嘴唇和嘴巴以及表情和头部姿势变化作为一个整体进行建模，从而获得更加真实的结果。源代码在GitHub中是公开的[30](#_bookmark45).还有，在[[184]](#_bookmark211)Song等人提出了一种动态方法，该方法不采用如[[125].](#_bookmark165)在他们的方法中，他们能够通过从输入视频执行3D人脸模型重建以及递归网络将源音频转换为表情参数来生成非常真实的假视频。最后，他们引入了一种新的视频渲染网络和一种动态编程方法来构建一个时间连贯和照片般逼真的视频。视频结果显示在互联网上[31](#_bookmark46).

另一个有趣的方法在[[185].](#_bookmark212) 周等人提出了一个新的框架，称为“去纠缠视听系统”,它使用去纠缠的视听表示生成高质量的人脸说话视频。音频和视频语音信息都可以用作输入指导。GitHub中提供了源代码[32](#_bookmark47).关于从文本合成假视频(文本到视频)，Fried等人在[[186]](#_bookmark213) 一种方法，该方法将人说话的视频和想要说出的文本作为输入，并合成一个新的视频，其中人的嘴与新单词同步。特别地，他们的方法自动地用音素、视位、3D人脸姿态和几何形状、反射率、表情和每帧的场景照明来注释输入的正在说话的头部视频。最后，循环视频生成网络创建与编辑过的抄本相匹配的照片级真实视频。赝品的例子

用这种方法生成的视频是公开可用的[33](#_bookmark48).

28<https://www.youtube.com/watch?v=VWMEDacz3L4>29<https://www.bbc.com/news/technology-48607673>

30[https://github.com/susanqq/Talking脸一代](https://github.com/susanqq/Talking_Face_Generation)

31<https://wywu.github.io/projects/EBT/EBT.html>

32<https://github.com/Hangz-nju-cuhk/Talking-Face-Generation-DAVS>

33<https://www.ohadf.com/projects/text-based-editing/>

据我们所知，目前还没有与音频和文本到视频伪造检测内容相关的公开可用的数据库和基准。关于这个主题的研究通常是通过使用公开可用的实现(如本节中描述的实现)合成内部数据来进行的。

最近的研究分析了检测音频和文本转视频的虚假内容有多容易。在…里[[182],](#_bookmark209)Agarwal等人提出了一种伪检测方法，该方法利用了口型(视位)和口语音素之间存在的不一致性。他们把注意力集中在一些特定的视位上，在这些视位上，嘴必须完全闭上，并观察到这种情况在许多被操纵的视频中并没有发生。他们提出的方法取得了良好的效果，特别是当视频长度增加时。

1. 结束语

受数字人脸识别(尤其是深度伪造)不断取得成功的推动，这项调查提供了该领域的全面全景，包括最新的详细信息:

*I)面部操作的类型，ii)面部操作技术，iii)用于研究的公共数据库，以及iv)每个面部操作组的检测基准，包括通过最具代表性的操作检测方法获得的关键结果。*

一般来说，大多数当前的人脸操纵似乎容易在受控场景下被检测到，即，当在与它们被训练的条件相同的条件下评估假检测器时。这一事实已经在本次调查中包含的大多数基准测试中得到证明，在操纵检测中实现了非常低的错误率。然而，这种情况可能不太现实，因为伪造的图像和视频通常在社交网络上共享，受到压缩级别、大小调整、噪声等高度变化的影响。此外，面部操作技术也在不断改进。这些因素激发了对伪检测器在未知条件下的泛化能力的进一步研究。这方面在不同的著作中已有初步研究[[16],](#_bookmark62) [[59]–[61].](#_bookmark103) 未来的研究可能会与最新的出版物保持一致[[187],](#_bookmark214)

[[188]](#_bookmark215) 因为他们不需要假视频进行训练，提供了更好的对看不见的攻击的概括能力。

在特征或分数级别上的融合技术可以提供对不同场景的更好的伪检测器适应性[[189]–[191].](#_bookmark217)事实上，不同的伪检测方法已经基于不同信息源的组合，例如周等人在[[95]](#_bookmark136)一种基于隐写分析和纯深度学习特征相结合的检测系统，而Rathgeb等人在[[141]](#_bookmark177)空间和光谱特征的组合。另外两种有趣的融合方法最近在[[192],](#_bookmark218)[[193],](#_bookmark219)结合RGB、深度和红外信息来检测物理面部攻击。此外，已经提出了面部加权方法，以便使用多帧来检测虚假视频[[194].](#_bookmark220) 最后，融合其他信息来源，如上传到社交网络时伴随视频的文本、按键或音频，对于改进检测器非常有价值[[195]–[198].](#_bookmark222)

除了仅基于图像/视频信息的传统伪检测器之外，应当研究新的方案，以便提供更鲁棒的工具。这方面的一个例子是Tursman等人在[[199].](#_bookmark223) 作者提出在捕捉时通过社交验证来检测虚假内容:真实性的仲裁者是一组同步捕捉发言者的摄像机，集体达成共识，然后实时签署他们的视频为“真实”。像这样的方法可以进一步保护媒体内容免受攻击。

接下来，我们将重点介绍每个面部处理组需要改进的关键方面和未来趋势:

* 人脸合成:目前的操作通常基于GAN架构，如StyleGAN，提供非常逼真的图像。然而，大多数检测机可以轻松区分真假图像，准确率接近100%。这是由于假图像以特定的GAN指纹为特征而产生的。但是，如果我们能够消除那些GAN指纹或添加一些噪声模式，同时保持非常逼真的合成图像，会怎么样呢？最近的方法集中在这一研究路线上，即使对于最好的操纵检测系统来说，这也是一个挑战[[16],](#_bookmark62) [[78],](#_bookmark119) [[200].](#_bookmark224)
* 身份交换:尽管文献中提出了许多不同的方法，但是很难确定哪一个是最好的。这是由许多不同的因素造成的。首先，大多数方法都是针对特定的数据库和压缩级别进行训练的，总体来说效果非常好。然而，它们都对未知条件表现出较差的概括结果。此外，不同的度量(即，Acc。、AUC、EER等。)和实验方案通常被认为无助于实现研究之间的公平比较。应该进一步考虑所有这些方面，以便在该领域取得进展。

此外，我们希望强调在第二代最新DeepFake数据库(如DFDC和Celeb-DF)中实现的检测结果[[26],](#_bookmark69) [[83].](#_bookmark124) 而假检测器在第一代数据库如UADFV和FaceForensics++中已经达到接近100%的AUC结果[[12],](#_bookmark59)[[82],](#_bookmark123)在最新的数据库上，它们都有很高的性能下降，特别是对于Celeb-DF数据库，在大多数情况下AUC结果低于60%。因此，需要付出更多的努力来进一步改善当前的假货检测系统，例如，通过大规模的挑战和基准，如最近的DFDC[34](#_bookmark50).

* 属性操作:与人脸合成(GAN指纹移除)相同的方面也适用于此，因为大多数操作都是基于GAN架构的。此外，值得注意的是用于研究的公共数据库很少(只有DFFD数据库是公开的[[17]),](#_bookmark63)以及缺乏标准的实验方案来对研究进行公平的比较。

34<https://deepfakedetectionchallenge.ai/>

* 表达式交换:与随着改进的深度伪造数据库的发布而迅速发展的身份交换相反，据我们所知，表达式交换中唯一的公共数据库是FaceForensics++。该数据库的特征在于易于检测的视觉伪影，因此在几种伪检测方法中实现了接近100%的AUC结果。我们鼓励研究人员基于最新技术生成并公开更真实的数据库[[125],](#_bookmark165) [[151],](#_bookmark183) [[184].](#_bookmark211)

所有这些方面，加上改进的GAN方法的发展和最近的DeepFake检测挑战(DFDC ),将培育新一代逼真的伪造图像/视频[[70]](#_bookmark111)以及更先进的面部操作检测技术。

感谢

这项工作得到了以下项目的支持:PRIMA(h 2020-MSCA-ITN-2019-860315)、侵入-ETN(h 2020-MSCA-ITN-2019-860813)、BIBECA (MINECO/FEDER RTI2018-

101248-B-I00)、Bio-Guard(2017年BBVA科学调查设备基金会)和埃森哲。Ruben Tolosana得到了马德里社区教育、青年和体育理事会和社会欧洲基金会的支持。

参考

1. 页（page的缩写）Korshunov和S. Marcel，“深度伪装:面对认知的新威胁？评估和检测”，arXiv预印本arXiv:1812.08685，2018。
2. D.Citron，“深度造假如何破坏真理并威胁民主”，2019年。【在线】。可用:[https://www.ted.com](https://www.ted.com/)
3. R.Cellan-Jones，“Deepfake视频在9个月内翻了一番”，2019年。【在线】。可用:<https://www.bbc.com/news/technology-49961089>
4. BBC Bitesize，“Deepfakes:它们是什么，为什么我要做一个？”2019.【在线】。可用:<https://www.bbc.co.uk/bitesize/articles/zfkwcqt>
5. 形容词（adjective的缩写）Swaminathan，M. Wu和K.J.R. Liu，“通过固有指纹进行数字图像取证”，IEEE信息取证和安全汇刊，第3卷，第1期，第101–117页，2008年。
6. H.Farid，“图像伪造检测”，IEEE信号处理杂志，第26卷，第2期，第16–25页，2009年。
7. 米（meter的缩写））Stamm和K. Liu，“使用统计固有指纹的图像操作的取证检测”，IEEE信息取证和安全汇刊，第5卷，第3期，第492–506页，2010年。
8. 形容词（adjective的缩写）Rocha、W. Scheirer、T. Boult和S. Goldenstein，“看不见的视觉:数字图像和视频取证的当前趋势和挑战”，ACM计算调查，第43卷，第4期，第1–42页，2011年。
9. 南Milani、M. Fontani、P. Bestagini、M. Barni、A. Piva、M. Tagliasacchi和S. Tubaro，“视频取证概述”，《亚太信号与信息处理学报》，第1卷，第1–18页，2012年。
10. 形容词（adjective的缩写）Piva，“图像取证概述”，ISRN信号处理，第2013卷，第1–22页，2013年。
11. 页（page的缩写）Korus，“数字图像完整性-保护和验证技术调查”，数字信号处理，第71卷，第1–26页，2017年。
12. 形容词（adjective的缩写）罗斯勒、d .科佐利诺、l .弗多里瓦、c .里斯、j .提斯和

米（meter的缩写））niener，“FaceForensics++学习检测被操纵的面部图像”，在Proc。2019年IEEE/CVF计算机视觉国际会议。

1. J.Galbally、S. Marcel和J. Fierrez，“生物识别反欺骗方法:人脸识别调查”，IEEE Access，第2卷，第1530-1552页，2014年。
2. 形容词（adjective的缩写）Hadid、N. Evans、S. Marcel和J. Fierrez，“欺骗攻击下的生物识别系统:评估方法和经验教训”，IEEE信号处理杂志，第32卷，第5期，第20–30页，2015年。
3. 南Marcel，M. Nixon，J. Fierrez，N. Evans，《生物识别反欺骗手册》(第二版)，2019年。
4. J.内维斯、r .托洛萨纳、r .维拉-罗德里格斯、v .洛佩斯、h .普罗恩科a

J.Fierrez，“GANprintR:改进的假货和对人脸操纵检测技术的评估”，IEEE信号处理精选期刊，2020年。

1. H.Dang，F. Liu，J. Stehouwer，X. Liu和A. Jain，“关于数字人脸操作的检测”，在Proc。2020年IEEE/CVF计算机视觉和模式识别会议。
2. 碳（carbon的缩写）Canton，L. Davis，E. Delp，P. Flynn，S. McCloskey，L. Leal-Taixe，

页（page的缩写）Natsev和C. Bregler，“计算机视觉和模式识别在媒体取证中的应用”，IEEE/CVF计算机视觉和模式识别会议，2019年。【在线】。可用:<https://sites.google.com/view/mediaforensics2019>

1. B.Biggio，P. Korshunov，T. Mensink，G. Patrini，D. Rao和A. Sadhu，“合成现实:检测视听假货的深度学习”，2019年国际机器学习会议。【在线】。

可用:<https://sites.google.com/view/audiovisualfakes-icml2019/>

1. 长度Verdoliva和P. Bestagini，“多媒体取证”，载于ACM多媒体，2019年。【在线】。可用:<https://acmmm.org/tutorials/#tut3>
2. K.Raja，N. Damer，C. Chen，A. Dantcheva，A. Czajka，H. Han和R. Ramachandra，“生物识别中的深度伪造和表示攻击研讨会”，IEEE计算机视觉应用冬季会议，2020年。【在线】。可用:<https://sites.google.com/view/wacv2020-deeppab/>
3. 米（meter的缩写））Barni、S. Battiato、G. Boato、H. Farid和N. Memon，“野外多媒体取证”，IEEE模式识别国际会议，2020年。【在线】。可用:<https://iplab.dmi.unict.it/mmforwild/>
4. 碳（carbon的缩写）布雷格勒、m .科韦尔和m .斯莱尼，“视频重写:用音频驱动视觉语言”，《计算机图形学》，第31卷，第2期，第353-361页，1997年。
5. D.P .金玛和m .韦林，“自动编码变分贝叶斯”，载于

*继续。2013年学习表征国际会议。*

1. I .古德菲勒、j .普热-阿巴迪、m .米尔扎、b .徐、d .沃德-法利、

南奥泽尔，a .库维尔和y .本吉奥，“生成性对抗性网络”，在Proc。神经信息处理系统进展，2014。

1. Y.李，x .杨，p .孙，h .齐，和S. Lyu，“Celeb-DF:一个大规模的具有挑战性的数据集，用于DeepFake取证”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别会议。
2. I. Yerushalmy和H. Hel-Or，“基于镜头和传感器像差的数字图像伪造检测”，《国际计算机视觉杂志》，第92卷第1期，第71–91页，2011年。
3. A.C. Popescu和H. Farid，“揭露彩色滤光片阵列插值图像中的数字伪造”，IEEE信号处理汇刊，第53卷，第10期，第3948–3959页，2005年。
4. H.曹和A.C. Kot，“数字图像取证中去马赛克规则的精确检测”，IEEE信息取证与安全汇刊，第4卷第4期，第899–910页，2009年。
5. Z.林，何俊杰，唐晓霞，唐春华，“基于DCT系数分析的JPEG图像篡改快速自动检测”，模式识别，第42卷，第11期，第2492-2501页，2009年。
6. 陈耀林和徐志泰，“通过压缩伪像的周期性分析检测JPEG图像的再压缩，用于篡改检测”，IEEE信息取证和安全学报，第6卷，第2期，第396–406页，2011年。
7. I. Amerini、L. Ballan、R. Caldelli、A. Bimbo和G. Serra，“基于SIFT的复制-移动攻击检测和转换恢复取证方法”，IEEE信息取证和安全汇刊，第6卷，第3期，第1099-1110页，2011年。
8. D.科佐利诺、g·波吉和l·弗多里瓦，“Splicebuster:一种新的盲图像拼接检测器”，正在进行中。IEEE信息取证与安全国际研讨会，2015，第1–6页。
9. 形容词（adjective的缩写）Gironi，M. Fontani，T. Bianchi，A. Piva和M. Barni，“检测帧删除和插入的视频取证技术”，正在进行中。IEEE声学、语音和信号处理国际会议，2014年，第6226–6230页。
10. Y.吴，江，孙，王，“基于速度场一致性揭露视频帧间伪造”，正在进行中。IEEE声学、语音和信号处理国际会议，2014年，第2674–2678页。
11. B.c .霍斯勒和M. C .斯塔姆，“检测视频速度操纵”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
12. H.Allcott和M. Gentzkow，“2016年选举中的社交媒体和假新闻”，《经济展望杂志》，第31卷第2期，第211-36页，2017年。
13. D.拉泽尔，鲍姆，本克勒，贝林斯基，格林希尔，

氟（fluorine的缩写）Menczer，M. J. Metzger，B. Nyhan，G. Pennycook，D. Rothschild等人，“假新闻的科学”，《科学》，第359卷，第6380期，第1094–1096页，2018年。

1. I. M. J. Kietzmann，L.W. Lee和T. Kietzmann，“深度假动作:不给糖就捣蛋？”《商业视野》,第63卷，第2期，第135-146页，2020年。
2. 长度Verdoliva，“媒体取证和深度伪造:概述”，arXiv预印本arXiv:2001.06564，2020。
3. T.Karras，S. Laine和T. Aila，“一个基于风格的生成式对抗网络的生成器架构”，在Proc。2019年IEEE/CVF计算机视觉和模式识别会议。
4. E.Gonzalez-Sosa，J. Fierrez，R. Vera-Rodriguez和F. Alonso- Fernandez，“用于野外识别的面部软生物识别:最近的工作、注释和COTS评估”，IEEE信息取证和安全汇刊，第13卷，第8期，第2001–2014页，2018年。
5. Y.Choi，M. Choi，M. Kim，J. Ha，S. Kim和J. Choo，“StarGAN:用于多域图像到图像翻译的统一生成对抗网络”，正在进行中。2018年IEEE/CVF计算机视觉和模式识别会议。
6. 米（meter的缩写））刘，丁，夏，刘，丁，左，文，“一个统一的选择性传输网络的任意图像属性编辑”，正在进行。2019年IEEE/CVF计算机视觉和模式识别会议。
7. J.提斯、m .佐尔霍夫、m .斯塔明格、c .塞奥巴尔特和m .尼尔，“Face2face:实时人脸捕捉和RGB视频的再现”，正在进行中。2016年IEEE/CVF计算机视觉和模式识别会议。
8. J.提斯、m .佐尔霍·费尔和m .内尔，“延迟神经渲染:使用神经纹理的图像合成”，《美国计算机学会图形学报》，第38卷，第66期，第1-12页，2019年。
9. AI生成的10万张脸，2018。【在线】。可用:[https:](https://generated.photos/)

[//generated.photos/](https://generated.photos/)

1. T.Karras、T. Aila、S. Laine和J. Lehtinen，“为提高质量、稳定性和变化性而逐步种植甘蔗”，正在进行中。2018年国际学习表征会议。
2. 氟（fluorine的缩写）马拉、d·格拉尼耶洛、l·弗多里瓦和g·波吉，“甘人会留下人工指纹吗？”进行中。IEEE多媒体信息处理与检索会议，2019，第506–511页。
3. 米（meter的缩写））Albright和S. McCloskey，“通过反演的震源发生器属性？”进行中。2019年IEEE/CVF计算机视觉和模式识别研讨会。
4. 形容词（adjective的缩写）Jain，P. Majumdar，R. Singh和M. Vatsa，“通过DAD-HCNN检测GANs和修饰基于数字的变更”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
5. Z.刘，罗，王，唐，“野外深度学习人脸属性”。2015年IEEE/CVF计算机视觉国际会议。
6. D.易，雷，廖，李，“从零开始学习人脸表征”，arXiv预印本arXiv:1411.7923，2014。
7. 曹，沈，谢，帕克希，塞斯曼，“人脸识别数据集”，正在进行中。2018年自动人脸和手势识别国际会议。
8. 南McCloskey和M. Albright，“使用颜色线索检测GAN生成的图像”，arXiv预印本arXiv:1812.08247，2018年。
9. R.王，l .马，F. Juefei-Xu，X. Xie，J. Wang和Y. Liu，“FakeSpot- ter:识别人工智能合成假脸的简单基线”，arXiv预印本arXiv:1909.06122，2019年。
10. 长度葛奈拉、o .朱迪切和s .巴蒂亚托，“通过分析卷积轨迹进行深度伪造检测”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
11. 长度Nataraj，T. Mohammed，B. Manjunath，S. Chandrasekaran，A. Flen- ner，J. Bappy和A. Roy-Chowdhury，“使用共生矩阵检测GAN生成的伪图像”，《电子成像》，第5期，第1–7页，2019年。
12. 名词（noun的缩写）余，l .戴维斯和m .弗里茨，“将假图像归因于甘:分析生成图像中的指纹”，正在进行中。2019年IEEE/CVF计算机视觉国际会议。
13. 氟（fluorine的缩写）Marra，C. Saltori，G. Boato和L. Verdoliva，“GAN生成图像的检测和分类的增量学习”，Proc .2019年IEEE信息取证与安全国际研讨会。
14. 名词（noun的缩写）Hulzebosch，Sarah Ibrahimi和Marcel Worring，“在真实世界场景中检测CNN生成的面部图像”。2020年IEEE/CVF计算机视觉和模式识别研讨会。
15. H.关，，罗伯逊，李，耶茨，德尔加多，

D.周，T. Kheyrkhah，J. Smith和J. Fiscus，“MFC数据集:用于媒体取证挑战评估的大规模基准数据集”，正在进行中。2019年IEEE计算机视觉冬季应用研讨会。

1. O.Parkhi，A. Vedaldi和A. Zisserman，“深度人脸识别”，载于

*继续。2015年英国机器视觉大会。*

1. B.Amos，B. Ludwiczuk和M. Satyanarayanan，“OpenFace:一个具有移动应用程序的通用人脸识别库”，CMU计算机科学学院，2016年。
2. 氟（fluorine的缩写）Schroff，D. Kalenichenko和J. Philbin，“FaceNet:人脸识别和聚类的统一嵌入”。2015年IEEE/CVF计算机视觉和模式识别会议。
3. Y.沈，顾俊杰，唐晓松，周炳良，“面向语义人脸编辑的汉字潜在空间解读”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别会议。
4. T.K. Moon，“期望值最大化算法”，IEEE信号处理杂志，第13卷，第6期，第47–60页，1996年。
5. Z.何，左，靳，山，陈，“只改变你想要的面部属性编辑”，IEEE图像处理学报，第28卷，第11期，第5464-5478页，2019年。
6. W.Cho，S. Choi，D. K. Park，I. Shin，和J. Choo，“通过群组方式深度增白和着色转换进行图像到图像的翻译”，正在进行中。2019年IEEE/CVF计算机视觉和模式识别会议。
7. T.Karras，S. Laine，M. Aittala，J. Hellsten，J. Lehtinen和T. Aila，“分析和改善StyleGAN的图像质量”，在Proc。2020年IEEE/CVF计算机视觉和模式识别会议。
8. J.朱，T. Park，P. Isola，A. Efros，“使用循环一致的对抗网络进行不成对的图像到图像的翻译”，正在进行中。2017年IEEE/CVF计算机视觉国际会议。
9. T.宫藤、t .片冈、m .小山和y .吉田，“生成性对抗网络的谱规范化”，正在进行中。2018年国际学习表征会议。
10. 米（meter的缩写））Bellemare，I. Danihelka，W. Dabney，S. Mohamed，B. Lak- shminarayanan，S. Hoyer和R. Munos，“作为有偏Wasserstein梯度解决方案的Cramer距离”，arXiv预印本arXiv:1705.10743，2017年。
11. 米（meter的缩写））宾考斯基、d萨瑟兰、m阿贝尔和a格雷顿，“去MMD甘斯的神秘化”，正在进行中。2018年国际学习表征会议。
12. 南瑞布菲，a .科列斯尼科夫，g .斯佩尔和c .兰伯特，“iCaRL:增量分类器和表征学习”，正在进行中。2017年IEEE/CVF计算机视觉和模式识别会议。
13. D.Kingma和P. Dhariwal，“Glow:具有可逆的1x1卷积的生成流”，正在进行中。神经信息处理系统进展，2018。
14. 形容词（adjective的缩写）Vaswani，N. Shazeer，N. Parmar，J. Uszkoreit，L. Jones，A. Gomez，

长度凯泽，和I. Polosukhin，“注意是你所需要的，”在Proc。神经信息处理系统进展，2017年，第5998–6008页。

1. D.Cozzolino，j .提斯，A. Ro ssler，m . niener和L. Ver- doliva，“SpoC:欺骗相机指纹”，arXiv预印本arXiv:1911.12069，2019年。
2. D.Cozzolino，j .提斯，A. Ro ssler，C. Riess，m . niener和L. Ver- doliva，“ForensicTransfer:弱监督域适应伪造检测”，arXiv预印本arXiv:1812.02510，2018年。
3. 米（meter的缩写））刘，欧文斯，埃夫罗斯，“打击假新闻:通过学习自我一致性的图像拼接检测”，在进步。2018年欧洲计算机视觉会议。
4. 页（page的缩写）周，x .韩，V. Morariu和L. Davis，“学习图像处理检测的丰富特征”，在Proc。2018年IEEE/CVF计算机视觉和模式识别会议。
5. Y.李，m .张，和S. Lyu，“在Ictu Oculi:揭露人工智能通过检测眨眼产生的假脸视频，”在Proc。2018年IEEE信息取证与安全国际研讨会。
6. B.多尔汉斯基、r .霍维斯、b .普弗劳姆、n .巴拉姆和c .费雷尔，“深度伪造检测挑战(DFDC)预览数据集”，arXiv预印本arXiv:1910.08854，2019年。
7. 碳（carbon的缩写）Sanderson和B. Lovell，“稳健和可扩展身份推理的多区域概率直方图”，在Proc。2009年国际生物识别会议。
8. K.张，张志军，李志军，乔，“多任务级联卷积网络的联合人脸检测与对齐”，IEEE信号处理通讯，第23卷第10期，第1499–1503页，2016年。
9. 谷歌人工智能，“为Deepfake检测研究贡献数据”，2019年。【在线】。可用:<https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html>
10. 形容词（adjective的缩写）罗斯勒、d .科佐利诺、l .弗多里瓦、c .里斯、j .提斯和

米（meter的缩写））niener，“人脸取证:人脸伪造检测的大规模视频数据集”，arXiv预印本arXiv:1803.09179，2018年。

1. 页（page的缩写）Pe rez，M. Gangnet和A. Blake，“泊松图像编辑”，《美国计算机学会图形学学报》，第22卷，第3期，第313-318页，2003年。
2. 氟（fluorine的缩写）Matern，C. Riess和M. Stamminger，“利用视觉人工制品揭露深度伪造和面部操纵”，正在进行中。2019年IEEE计算机视觉冬季应用研讨会。
3. X.杨，y .李，S. Lyu，“使用不一致的头部姿势揭露深度假动作”，在Proc。2019年声学、语音和信号处理国际会议。
4. 南Agarwal和H. Farid，“保护世界领导人免受深度欺诈”。2019年IEEE/CVF计算机视觉和模式识别研讨会。
5. T.Jung，S. Kim和K. Kim，“DeepVision:使用人眼眨眼模式检测Deepfakes”，IEEE Access，第8卷，第83 144–83 154页，2020年。
6. Y.李和S. Lyu，“通过检测面部扭曲伪影来揭露DeepFake视频”，在Proc。2019年IEEE/CVF计算机视觉和模式识别研讨会。
7. D.Afchar，V. Nozick，J. Yamagishi和I. Echizen，“MesoNet:一个紧凑的面部视频伪造检测网络”，正在进行中。2018年IEEE信息取证与安全国际研讨会。
8. 页（page的缩写）周，韩，V. Morariu，和L. Davis，“两个流的神经网络-篡改人脸检测的工作”，正在进行。2017年IEEE/CVF计算机视觉和模式识别研讨会。
9. H.Nguyen，F. Fang，J. Yamagishi，I. Echizen，“检测和分割操纵的面部图像和视频的多任务学习”，arXiv预印本arXiv:1906.06876，2019年。
10. H.H. Nguyen，J. Yamagishi和I. Echizen，“使用胶囊网络检测虚假图像和视频”，arXiv预印本arXiv:1910.12467，2019年。
11. E.Sabir，J. Cheng，A. Jaiswal，W. AbdAlmageed，I. Masi和

页（page的缩写）Natarajan，“视频中人脸操作检测的递归卷积策略”，正在进行中。2019年IEEE/CVF计算机视觉和模式识别研讨会。

1. D.Gu era和E. Delp，“使用递归神经网络的深度伪造视频检测”，正在进行中。2018年高级视频和基于信号的监控国际会议。
2. R.Tolosana，S. Romero-Tapiador，J. Fierrez和R. Vera-Rodriguez，“DeepFakes进化:面部区域和假货检测性能的分析”，arXiv预印本arXiv:2004.07532，2020。
3. 页（page的缩写）Korshunov和S. Marcel，“篡改视频中的说话人不一致性检测”，正在进行中。2018年欧洲信号处理会议。
4. J.Galbally、S. Marcel和J. Fierrez，“伪造生物识别检测的图像质量评估:虹膜、指纹和人脸识别的应用”，IEEE图像处理汇刊，第23卷，第2期，第710–724页，2014年。
5. I. Goodfellow、Y. Bengio和a .库维尔，深度学习，2016年。
6. D.King，“DLib-ML:机器学习工具包”，《机器学习研究杂志》，第10卷，第1755-1758页，2009年。
7. T.巴尔特鲁塞蒂斯、a .扎德、y .林和l .莫伦西，“开放脸2.0:面部行为分析工具包”，正在进行中。2018年自动人脸和手势识别国际会议。
8. R.Daza，A. Morales，J. Fierrez和R. Tolosana，“mEBAL:眨眼检测和注意力水平估计的多模态数据库”，arXiv预印本arXiv:2006.05327，2020。
9. R.Ranjan，V. M. Patel和R. Chellappa，“Hyperface:用于人脸检测、地标定位、姿态估计和性别识别的深度多任务学习框架”，IEEE模式分析和机器智能汇刊，第41卷第1期，第121-135页，2017年。
10. T.Soukupova和J. Cech，“使用面部标志的眨眼检测”，Proc .2016年计算机视觉冬季研讨会。
11. K.Simonyan和A. Zisserman，“用于大规模图像识别的极深度卷积网络”，arXiv预印本arXiv:1409.1556，2014年。
12. K.何，张，任，孙，“深度残差学习在图像识别中的应用”，中国教育科学出版社，2003年。IEEE/CVF计算机视觉和模式识别会议，2016，第770–778页。
13. ——《用于视觉识别的深度卷积网络中的空间金字塔池》，IEEE模式分析与机器智能汇刊，第37卷，第9期，第1904–1916页，2015年。
14. 碳（carbon的缩写）Szegedy，W. Liu，Y. Jia，P. Sermanet，S. Reed，D. Anguelov，

D.Erhan，V. Vanhoucke和A. Rabinovich，“用卷积深入研究”，正在进行中。2015年IEEE/CVF计算机视觉和模式识别会议。

1. D.科佐利诺、g .波吉和l .弗多里瓦，“重铸基于残差的局部描述符作为卷积神经网络:图像伪造检测的应用”，正在进行中。2017年ACM信息隐藏与多媒体安全研讨会。
2. B.Bayar和M. Stamm，“使用新卷积层的通用图像处理检测的深度学习方法”，Proc .2016年ACM信息隐藏与多媒体安全研讨会。
3. 名词（noun的缩写）Rahmouni，V. Nozick，J. Yamagishi和I. Echizen，“使用卷积神经网络区分计算机图形和自然图像”，在Proc。2017年IEEE信息取证与安全研讨会。
4. 氟（fluorine的缩写）Chollet，“异常:具有深度方向可分离进化的深度学习”，正在进行中。2017年IEEE/CVF计算机视觉和模式识别会议。
5. J.邓，董文伟，索彻，李，李，李，，“图像网络:一个大规模的层次图像数据库”，正在进行中。2009年IEEE/CVF计算机视觉和模式识别会议。
6. G.E. Hinton和A. Krizhevsky和S.D. Wang，“改造自动编码器”，人工神经网络国际会议，2011年，第44–51页。
7. 南Sabour，N. Frosst和G.E. Hinton，“Cap- sules之间的动态路由”，正在进行中。神经信息处理系统进展，2017年，第3856–3866页。
8. G.E. Hinton、S. Sabour和N. Frosst，“带EM路由的基质胶囊”，正在进行中。2018年国际学习代表研讨会。
9. Y.王和丹切娃:“一个视频胜过1000个谎言。比较用于检测Deepfakes的3DCNN方法。2020年IEEE自动人脸和手势识别国际会议。
10. J.Carreira和A. Zisserman，“Quo Vadis，动作识别？一个新的模型和动力学数据集”，正在进行中。2017年IEEE/CVF计算机视觉和模式识别会议。
11. K.Hara，H. Kataoka，Y. Satoh，“时空3D CNNs能追溯3D CNNs和Imagenet的历史吗？”进行中。2018年IEEE/CVF计算机视觉和模式识别会议。
12. 碳（carbon的缩写）Szegedy，V. Vanhoucke，S. Ioffe，J. Shlens和Z. Wojna，“重新思考计算机视觉的初始架构”，正在进行中。2016年IEEE/CVF计算机视觉和模式识别会议。
13. 南Suwajanakorn、S. Seitz和I. Kemelmacher-Shlizerman，“合成奥巴马:从音频中学习口型同步”，《美国计算机学会图形学学报》，第36卷第4期，第1-13页，2017年。
14. 页（page的缩写）伊索拉，朱，周，埃夫罗斯，“图像到图像的翻译与有条件的对抗性网络”，正在进行。2017年IEEE/CVF计算机视觉和模式识别大会。
15. J.朱，r .张，D. Pathak，T. Darrell，A. Efros，O. Wang和

E.谢赫曼，“走向多模态图像到图像的翻译”，载于

*继续。神经信息处理系统进展，2017。*

1. T.Kim，M. Cha，H. Kim，J. Lee和J. Kim，"学习发现与生成性敌对网络的跨领域关系"，正在进行中。2017年机器学习国际会议。
2. D.Bau，J. Zhu，H. Strobelt，B. Zhou，J. Tenenbaum，W. Freeman和A. Torralba，“GAN解剖:可视化和理解生成性对抗性网络”，arXiv预印本arXiv:1811.10597，2018年。
3. G.佩拉瑙、J. V. D .韦杰尔、b .拉杜卡努和J. A .阿尔瓦雷斯，“用于图像编辑的可逆条件甘斯”，正在进行中。2016年神经信息处理系统研讨会进展。
4. 米（meter的缩写））李，左，张，“面部属性的深层身份意识转移”，arXiv预印本arXiv:1610.05586，2016。
5. W.沈和刘，“学习剩余图像的人脸属性操作”，在Proc。2017年IEEE/CVF计算机视觉和模式识别会议。
6. G.Lample，N. Zeghidour，N. Usunier，A. Bordes，L. Denoyer和

米（meter的缩写））Ranzato，“推子网络:通过滑动属性操纵图像”，正在进行中。神经信息处理系统进展，2017。

1. T.肖，洪军，马军，“优雅:与甘交换潜在编码以转移多种面部属性”，正在进行中。2018年欧洲计算机视觉会议。
2. 米（meter的缩写））Mirza和S. S. Osindero，“条件生成对抗性网络”，arXiv预印本arXiv:1411.1784，2014年。
3. 形容词（adjective的缩写）Bharati，R. Singh，M. Vatsa和k .鲍耶，“使用监督深度学习检测面部修饰”，《IEEE信息取证和安全学报》，第11卷，第9期，第1903-1913页，2016年。
4. 形容词（adjective的缩写）Jain，R. Singh和M. Vatsa，“关于检测GANs和基于合成变更的修饰”，在Proc。2018年生物识别理论、应用和系统国际会议。
5. 南Tariq，S. Lee，H. Kim，Y. Shin和S. Woo，“检测机器和人类在野外创造的假人脸图像”，Proc .多媒体隐私与安全国际研讨会，2018，第81–87页。
6. 南王，o .王，a .欧文斯，r .张，A. Efros，“通过编写Photoshop脚本检测Photoshop处理过的人脸”，正在进行中。2019年IEEE/CVF计算机视觉国际会议。
7. X.张，S. Karaman，S. Chang，“检测和模拟GAN伪图像中的伪影”，正在进行中。2019年IEEE信息取证与安全国际研讨会。
8. 碳（carbon的缩写）Rathgeb、A. Botaljov、F. Stockhardt、S. Isadskiy、L. Debiasi、A. Uhl和C. Busch，“基于PRNU的面部修饰检测”，IET生物识别，2020年。
9. J.Kim，J. Choi，J. Yi和M. Turk，“使用独立分量分析的有效表示对局部失真和部分遮挡鲁棒的人脸识别”，IEEE模式分析和机器智能汇刊，第12期，第1977-1981页，2005年。
10. 形容词（adjective的缩写）Dantcheva，C. Chen和A. Ross，“面部化妆品会影响人脸识别系统的匹配准确性吗？”进行中。国际生物识别会议:理论、应用和系统，2012年，第391–398页。
11. 名词（noun的缩写）美谛高丝、l .阿普维尔和j .杜杰雷，“基于纹理和形状分析的面部化妆检测技术”，正在进行中。2015年自动人脸和手势识别国际会议和研讨会。
12. 页（page的缩写）Majumdar，A. Agarwal，R. Singh和M. Vatsa，“通过部分篡改人脸来逃避人脸识别”，正在进行中。2019年IEEE/CVF计算机视觉和模式识别研讨会。
13. 碳（carbon的缩写）Rathgeb，A. Dantcheva，C. Busch，“人脸识别中面部美化的影响和检测:概述”，IEEE Access，第7卷，第152 667–152 678页，2019年。
14. 碳（carbon的缩写）Rathgeb、C. I. Satnoianu、N. E. Haryanto、K. Bernardo和

碳（carbon的缩写）Busch，“面部修饰的差分检测:多生物特征方法”，IEEE Access，第8卷，第106 373–106 385页，2020年。

1. 页（page的缩写）弗林，k .鲍耶和p .菲利普斯，“面部识别中时间依赖性的评估:一项初步研究”，正在进行中。2003年基于音频和视频的生物特征身份认证国际会议。
2. 调整和夸大五官。Adobe Photoshop，2016。【在线】。可用:<https://helpx.adobe.com/photoshop/how-to/face-aware-liquify.html>
3. D.孙，x .杨，M.Y .刘和J. Kautz，“普华永道网络:使用金字塔、弯曲和成本体积的光流的CNN”，在Proc。2018年IEEE/CVF计算机视觉和模式识别会议。
4. H.Averbuch-Elor，D. Cohen-Or，J. Kopf和M.F. Cohen，“将肖像带入生活”，《美国计算机学会图形学学报》，第36卷第6期，第196页，2017年。
5. E.扎哈罗夫，A. Shysheya，E. Burkov和V. Lempitsky，“现实的神经对话头部模型的少数镜头对抗学习”，在Proc。2019年IEEE/CVF计算机视觉国际会议。
6. D.朱，刘，江，高，吴，郭，“可追踪的多领域人脸翻译方法”，arXiv预印本arXiv:1907.11418。
7. I. Amerini，L. Galteri，R. Caldelli和A. Bimbo，“通过基于CNN的光流进行深度伪造视频检测”，正在进行中。2019年IEEE/CVF计算机视觉国际会议。
8. G.沃尔伯格，“图像变形:调查”，《视觉计算机》，第14卷，第8-9期，第360-372页，1998年。
9. 单位Scherhag，C. Rathgeb，J. Merkle，R. Breithaupt和C. Busch，“变形攻击下的人脸识别系统:调查”，IEEE Access，第7卷，第23 012–23 026页，2019年。
10. 米（meter的缩写））戈麦斯-巴列罗、c .拉斯盖伯、u .舍尔哈格和c .布施，“你的生物识别系统对变形攻击是否稳健？”进行中。2017年IEEE生物识别和取证国际研讨会。
11. 页（page的缩写）Korshunov和S. Marcel，“人脸识别对深度变形的脆弱性”，arXiv预印本arXiv:1910.01933，2019年。
12. K.Raja，M. Ferrara，A. Franco，L. Spreeuwers，I. Batskos，F. de Wit，

米（meter的缩写））戈麦斯-巴列罗、谢尔哈格、费希尔、文卡特什、辛格、李、贝热龙、伊萨德斯基、拉马钱德拉、拉斯盖卜、

D.Frings，U. Seidel，F. Knopjes，R. Veldhuis，D. Maltoni和

碳（carbon的缩写）Busch，“变形攻击检测-数据库、评估平台和基准测试”，arXiv预印本arXiv:2006.06458，2020。

1. 碳（carbon的缩写）恺自迩、马克鲁申、纽伯特、希尔德布兰特和

J.迪特曼，“模拟对照片身份文件的攻击和应用媒体取证检测面部变形，”在Proc。ACM信息隐藏与多媒体安全研讨会，2017，第21–32页。

1. 张立波，彭，龙，“利用传感器模式噪声的傅立叶谱进行人脸变形检测”，正在进行中。2018年多媒体与博览会国际会议。
2. 单位“使用面部标志检测变形的人脸图像”，正在进行中。2018年IEEE图像与信号处理国际会议。
3. 名词（noun的缩写）Damer，V. Bolle，Y. Wainakh，F. Boutros，P. Terho rst，A. Braun和A. Kuijper，“通过分析面部标志位移的有向距离来检测面部变形攻击”，德国模式识别会议，2018年。
4. 米（meter的缩写））Ferrara，A. Franco和D. Maltoni，“打印/扫描和异构图像源存在下的人脸变形检测”，arXiv预印本arXiv:1901.08811，2019年。
5. 单位Scherhag，C. Rathgeb，J. Merkle和C. Busch，“差分变形攻击检测的深度人脸表示”，arXiv预印本arXiv:2001.01202，2020。
6. 米（meter的缩写））Ferrara，A. Franco和D. Maltoni，“面对demorphing”，IEEE信息取证和安全汇刊，第13卷，第4期，第1008-1017页，2017年。
7. 氟（fluorine的缩写）彭，张立波，龙，闵，“FD-GAN:人脸变形生成对抗网络恢复同伙人脸图像”，IEEE Access，第7卷，第75页122–75页131，2019。
8. R.Gross，L. Sweeney，F. De la Torre和S. Baker，“基于模型的人脸识别”，正在进行中。2006年IEEE/CVF计算机视觉和模式识别研讨会。
9. R.Gross、L. Sweeney、J. Cohn、F. De la Torre和S. Baker，“人脸去识别”，载于《保护视频监控中的隐私》。斯普林格，2009年，第129-146页。
10. B.Meden，R. C. Mall，S. Fabijan，H. K. Ekenel，v . struc和P. Peer，“利用生成性深度神经网络进行人脸识别”，《IET信号处理》，第11卷，第9期，第1046-1054页，2017年。
11. K.Brkic，I. Sikiric，T. Hrkac和Z. Kalafatic，“我认识那个人:图像中人的生成性全身和面部去识别”，正在进行中。2017年IEEE/CVF计算机视觉和模式识别研讨会。
12. Q. Sun，L. Ma，S.O. Joon，L.V. Gool，B. Schiele和M. Fritz，“通过头部修复进行自然有效的模糊处理”，正在进行中。2018年IEEE/CVF计算机视觉和模式识别会议。
13. B.Meden，v . emersicˇ，v . struc，和P. Peer，“k-Same-Net:k-anonymous with Generative Deep Neural Networks for Face de identification”，《熵》，第20卷，第1期，第60页，2018年。
14. 南郭，冯，李，山和董，“将多样性集成到基于神经网络的人脸识别中”，正在进行中。中国控大会，2018。
15. 潘耀林，黄明杰，丁国栋，吴建林，张建生，“K-Same- Siamese-GAN:基于生成对抗网络的K-Same算法在人脸图像识别中的应用”，正在进行中。2019年高级视频和基于信号的监控国际会议。
16. 动词 （verb的缩写）Mirjalili，S. Raschka和A. Ross，“FlowSAN:增强隐私的半对抗性网络，以混淆任意基于人脸的性别分类器”，IEEE Access，第7卷，第99 735-99 745页，2019年。
17. O.Gafni，L. Wolf和Y. Taigman，“视频中的实时人脸去识别”，arXiv预印本arXiv:1911.08348，2019年。
18. Y.李、吕世安，“不丢面子的去身份化”，正在进行中。2019年ACM信息隐藏和多媒体安全研讨会。
19. 米（meter的缩写））Alvi，A. Zisserman和C. Nella˚ ker，“睁一只眼闭一只眼:从深度神经网络嵌入中明确去除偏差和变异”，在Proc。2018年欧洲计算机视觉会议。
20. 形容词（adjective的缩写）Morales，J. Fierrez和R. Vera-Rodriguez，“SensitiveNets:学习应用于人脸识别的不可知论表示”，arXiv预印本arXiv:1902.00334，2019年。
21. 南龚，刘，贾殷，“去偏置人脸识别”

*arXiv预印本arXiv:1911.08080，2019。*

1. 南Agarwal，H. Farid，O. Fried和M. Agrawala，“从音位-视位不匹配中检测深度伪造视频”，在Proc。2020年IEEE/CVF计算机视觉和模式识别研讨会。
2. Y.宋，朱军，李，王，齐，“基于条件循环对抗网络的人脸生成研究”，正在进行中。2019人工智能国际联合会议。
3. 长度宋，吴文伟，钱，何，罗正荣，“每个人都在说话:让我说你想说的话”，arXiv预印本arXiv:2001.05201，2020。
4. H.周，刘玉英，刘志军，罗平，王晓翔，"用非对抗性视听表征生成谈话面孔"，正在进行中。2019年AAAI人工智能大会。
5. O.放大图片作者:Jeffrey a .

D.B. Goldman、K. Genova、Z. Jin、C. Theobalt和M. Agrawala，“基于文本的有声视频编辑”，《美国计算机学会图形学学报》，第38卷第4期，第1-14页。

1. H.Khalid和S. S. Woo，“OC-fake detect:使用一类变分自动编码器对Deepfakes进行分类”。2020年IEEE/CVF计算机视觉和模式识别研讨会。
2. 南费尔南德斯、拉杰、埃维兹、潘努、贾、奥尔蒂斯、
   1. Vintila和M. Salter，“使用基于属性的置信度度量检测深度伪造视频”，在Proc。2020年IEEE/CVF计算机视觉和模式识别研讨会。
3. J.Fierrez，A. Morales，R. Vera-Rodriguez和D. Camacho，“生物统计学中的多重分类器。第1部分:基础与综述，《信息融合》，第44卷，第57–64页，2018。
4. ——“生物识别中的多分类器。第二部分:趋势与挑战，《信息融合》，第44卷，第103–112页，2018年。
5. R.S. M. Singh和A. Ross，“生物特征融合综合概述”，《信息融合》，第52卷，第187–205页，2019年。
6. 杨，朱，吴，叶，尤，朱，“多通道人脸识别系统中的通道选择方法”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
7. Z.于，秦，李，王，赵，雷，赵，"基于中心差分网络的多模态人脸反欺骗技术"，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
8. D.蒙塞拉特、郝、亚拉加达、贝雷迪、邵、

J.、E. Bartusiak、J. Yang、D. Gu era、F. Zhu和E. J. Delp，“自动人脸加权的深度假货检测”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。

1. T.Agrawal，R. Gupta和S. Narayanan，“通过融合分类和成对排序系统对虚假社交媒体使用的多模态检测”，在Proc。欧洲信号处理会议，2017，第1045–1049页。
2. K.Shu，A. Sliva，S. Wang，J. Tang和H. Liu，“社交媒体上的假新闻检测:数据挖掘视角”，ACM SIGKDD exploo-rations Newsletter，第19卷第1期，第22-36页，2017年。
3. K.Shu，D. Mahudeswaran，H. Liu，“FakeNewsTracker:假新闻收集、检测和可视化的工具”，计算和数学组织理论，第25卷，第1期，第60-71页，2019年。
4. 形容词（adjective的缩写）Morales，A. Acien，J. Fierrez，J. V. Monaco，R. Tolosana，R. Vera- Rodriguez和J. Ortega-Garcia，“在全球疫情中响应假新闻传播的击键生物测定学”，正在进行中。2020年IEEE计算机软件和应用会议研讨会。
5. E.Tursman，M. George，S. Kamara和J. Tompkin，“通过人脸几何一致性进行不可信的社交视频验证以打击深度伪造”，正在进行中。2020年IEEE/CVF计算机视觉和模式识别研讨会。
6. 名词（noun的缩写）Carlini和H. Farid，“用白盒和黑盒攻击逃避深度假图像检测器”，在Proc。2020年IEEE/CVF计算机视觉和模式识别研讨会。

Ruben Tolosana分别于2014年和2019年获得马德里自治大学的电信工程硕士学位和计算机与电信工程博士学位。2014年4月，他加入了马德里自治大学的生物识别和数据模式分析- BiDA实验室，目前在该实验室担任博士后研究员。从那以后，鲁本获得了几个奖项，如西班牙MECD的FPU研究奖学金(2015年)和欧洲生物技术奖

rics行业奖(2018)。他的研究兴趣主要集中在信号和图像处理、模式识别和机器学习，特别是在人脸处理、人机交互和生物识别领域。他是多种出版物的作者，并在许多不同的高影响力会议(例如，ICDAR、IJCB、ICB、BTAS、EUSIPCO等)上合作担任评论员。)和期刊(如IEEE TPAMI、TCYB、TIFS、TIP、ACM

CSUR等。).最后，他参与了几个国家和欧洲的项目，重点是在世界各地部署生物识别安全。

Ruben Vera-Rodriguez于2006年获得西班牙塞维利亚大学电信工程硕士学位，并于2010年获得英国斯旺西大学电气和电子工程博士学位。自2010年以来，他一直隶属于西班牙马德里自治大学生物识别小组，自2018年以来，他目前是该小组的副教授。他的研究兴趣包括信号和图像处理、模式识别和生物统计学，重点是签名、人脸、步态验证和

生物统计学的法医应用。他积极参与了几个国家和欧洲的生物识别项目。Ruben是2017年IEEE第51届国际卡纳汉安全与技术会议(ICCST)的项目主席；以及2018年第23届伊比利亚美洲模式识别大会(CIARP 2018)。

Julian Fierrez分别于2001年和2006年获得西班牙马德里理工大学电信工程硕士和博士学位。自2002年以来，他一直在马德里理工大学生物识别小组工作。自2004年以来，他一直在马德里自治大学工作，目前担任副教授。2007年至2009年，他以玛丽·居里奖学金在美国密歇根州立大学做访问研究员。他的研究兴趣包括信号和

图像处理、模式识别和生物统计学，重点是多生物统计学、生物统计学评估、系统安全、法医学和生物统计学的移动应用。他积极参与了多个欧盟生物识别项目(如TABULA RASA和BEAT ),并因其研究产生了显著影响。他获得了多项荣誉，包括2006年EAB欧洲生物识别产业奖、2012年EURASIP最佳博士奖、授予马德里社区40岁以下科学技术领域最佳研究人员的Miguel Catalan奖，以及2017年IAPR青年生物识别研究者奖。他是《IEEE信息取证与安全汇刊》和《IEEE图像处理汇刊》的副主编。

艾塔米·莫拉莱斯于2006年获得大加那利岛拉斯帕尔马斯大学电信工程硕士学位，并于2011年获得大加那利岛拉斯帕尔马斯大学博士学位。自2017年以来，他是马德里自治大学的副教授。他曾在密歇根州立大学生物识别研究实验室、香港理工大学生物识别研究中心、博洛尼亚大学生物识别系统实验室和Schepens进行研究

眼科研究所。他在国际期刊和会议上发表了70多篇科学文章。他与其他大学和私人实体合作，参与了国家和欧盟项目，如UAM大学、UPM大学、欧洲旅游管理学院、因陀罗大学、Union Fenosa大学、Soluziona大学或埃森哲大学。他的研究兴趣集中在模式识别、计算机视觉、机器学习和生物信号处理。他曾获得ULPGC、La Caja de Canarias、SPEGC和COIT颁发的奖项。

Javier Ortega-Garcia分别于1989年和1996年获得西班牙马德里理工大学电气工程硕士学位和电气工程博士学位(优等生)。他目前是西班牙马德里自治大学信号处理讲座的正教授，开设生物识别和数字信号处理课程。他是生物识别和数据模式分析小组BiDA-Lab的创始人和主管。他撰写了300多篇国际论文，

包括书籍章节、参考期刊和会议论文。他的研究兴趣集中在用于安全、电子健康和用户档案应用的生物特征模式识别(在线签名验证、说话人识别、人机交互)。他主持了Odyssey-04、说话人识别研讨会、ICB-2013、第六届IAPR国际生物识别会议和ICCST2017、第51届IEEE国际卡纳汉安全技术会议。