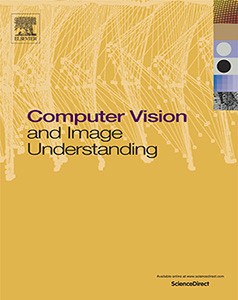
[计算机视觉与图像理解223 (2022) 103525](https://doi.org/10.1016/j.cviu.2022.103525)



目录列表可从以下网址获得[科学指导](http://www.elsevier.com/locate/cviu)

计算机视觉和图像理解

期刊主页:[www.elsevier.com/locate/cviu](http://www.elsevier.com/locate/cviu)

[](http://crossmark.crossref.org/dialog/?doi=10.1016/j.cviu.2022.103525&domain=pdf)深度学习用于深度伪造的创建和检测:综述

陈氏阮[a](#_bookmark0),[∗](#_bookmark7)，郭越雄阮[b](#_bookmark1)阮晋勇[a](#_bookmark0)阮德清[a](#_bookmark0)天慧[c](#_bookmark2)赛义德·纳哈万迪[d](#_bookmark3)清潭阮[e](#_bookmark4)Quoc-Viet Pham[f](#_bookmark5),

阮广明[g](#_bookmark6)

澳大利亚维多利亚州迪肯大学信息技术学院

b .澳大利亚昆士兰州格里菲斯大学信息和通信技术学院

c .大韩民国庆北Kumoh国家技术研究所信通技术融合研究中心

澳大利亚维多利亚州迪肯大学智能系统研究与创新研究所

越南胡志明市胡志明市技术大学信息技术学院

f .大韩民国釜山釜山国立大学第四次工业革命领袖教育韩国东南中心

g .拉米·UMR·CNRS，8201，法国瓦朗谢讷，法兰西高等理工大学

一个真实的信息

由Nikos Paragios传达

*关键词:*

Deepfakes

人脸操作人工智能深度学习自动编码器

开始

法医调查

A B S T R A C T

深度学习已被成功应用于解决从大数据分析到计算机视觉和人类水平控制的各种复杂问题。然而，深度学习的进步也被用来创建可能对隐私、民主和国家安全造成威胁的软件。最近出现的深度学习应用之一是deepfake。Deepfake算法可以创建虚假的图像和视频，人类无法将它们与真实的图像和视频区分开来。因此，能够自动检测和评估数字视觉媒体完整性的技术的提出是必不可少的。本文介绍了一个算法的调查，用来创造深度假，更重要的是，方法提出检测深度假在迄今为止的文献。我们对与deepfake技术相关的挑战、研究趋势和方向进行了广泛的讨论。通过回顾deepfakes的背景和最先进的deepfake检测方法，本研究提供了deepfake技术的全面概述，并有助于开发新的和更强大的方法来处理日益具有挑战性的deepfakes。

# 介绍

在狭义的定义中，deepfakes(源于“深度学习”和“fake”)是通过一些技术创建的，这些技术可以将目标人的面部图像叠加到源人的视频上，以制作目标人做或说源人做的事情的视频。这就构成了deepfakes的一个类别，即面子互换。从更广泛的定义来看，deepfakes是人工智能合成的内容，也可以分为其他两类，即假唱和木偶大师。假唱指的是经过修改使嘴部动作与录音一致的视频。木偶大师deepfakes包括一个目标人物(木偶)的视频，该目标人物跟随坐在摄像机前的另一个人(大师)的面部表情、眼睛和头部运动而被动画化([Agarwal et al.](#_bookmark21),[2019](#_bookmark21)).

虽然一些deepfake可以通过传统的视觉效果或计算机图形方法来创建，但是最近用于deep fake创建的常见底层机制是深度学习模型，例如自动编码器和生成对抗网络(GANs ),其已经广泛应用于计算机视觉领域([Vincent et al.](#_bookmark180),[2008](#_bookmark180);[Kingma and Welling](#_bookmark106),[2013](#_bookmark106);[Goodfellow et al.](#_bookmark75),[2014](#_bookmark75);[Makhzani et al.](#_bookmark129),

∫通讯作者。

*电子邮件地址:*[thanh.nguyen@deakin.edu.au](mailto:thanh.nguyen@deakin.edu.au)(阮晋勇)。

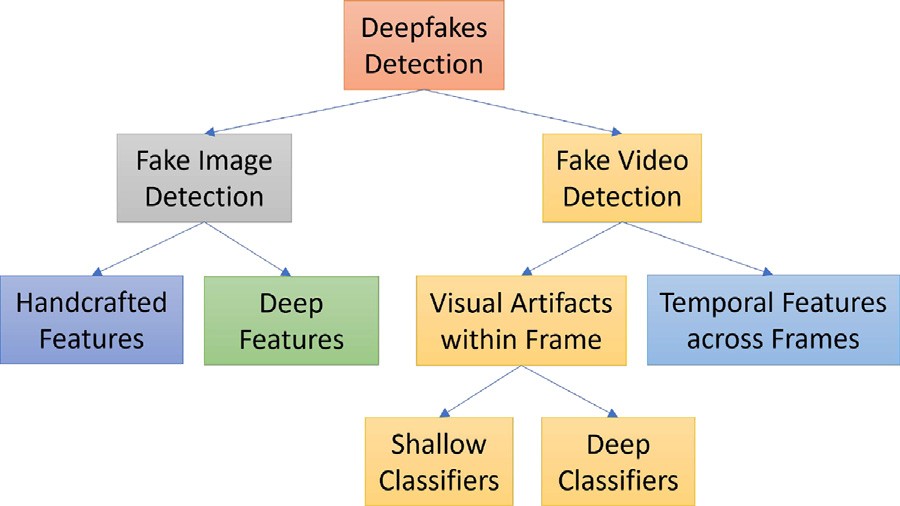
<https://doi.org/10.1016/j.cviu.2022.103525>

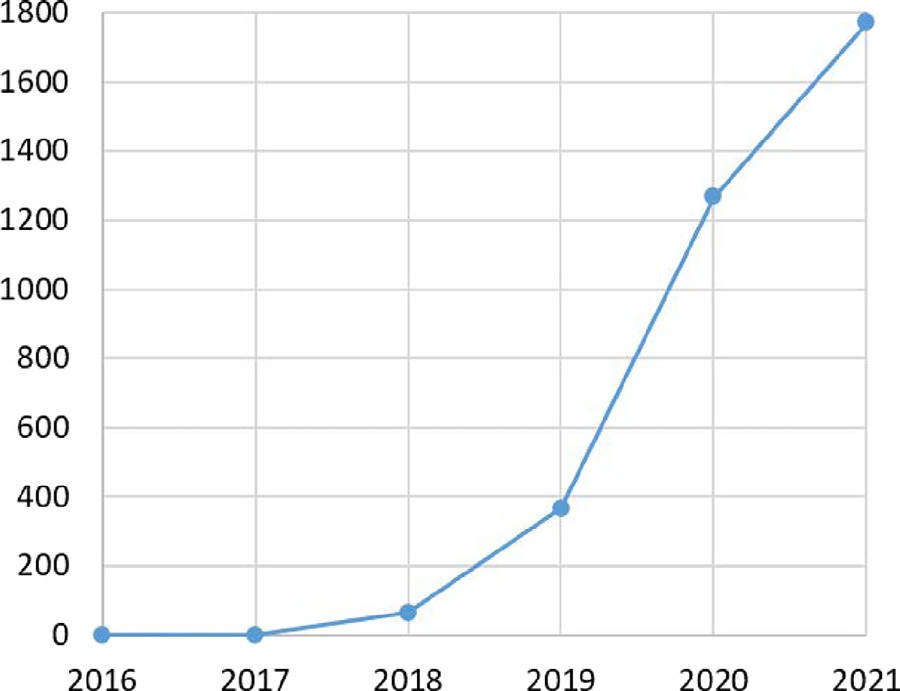
[2015](#_bookmark129);[Tewari et al.](#_bookmark168),[2018](#_bookmark168);[Lin et al.](#_bookmark123),[2021](#_bookmark123);[Liu et al.](#_bookmark125),[2021](#_bookmark125)).这些模型用于检查一个人的面部表情和动作，并合成另一个人做出类似表情和动作的面部图像([Lyu](#_bookmark127),[2018](#_bookmark127)).Deepfake方法通常需要大量的图像和视频数据来训练模型，以创建照片级的图像和视频。由于名人和政治家等公众人物可能在网上有大量视频和图像，他们是deepfakes的最初目标。Deepfakes被用来在色情图片和视频中将名人或政治家的脸换成身体。第一个deepfake视频出现在2017年，其中一个名人的脸被换成了一个色情演员的脸。当deepfake方法可以被用来制作世界领导人的视频，并以伪造为目的进行虚假演讲时，这对世界安全构成了威胁([Bloomberg](#_bookmark28),[2018](#_bookmark28);[Chesney and Citron](#_bookmark37),[2019](#_bookmark37);[Hwang](#_bookmark94),[2020](#_bookmark94)).因此，Deepfakes可以被滥用来造成国家之间的政治或宗教紧张局势，愚弄公众并影响选举活动的结果，或者通过制造假新闻来制造金融市场的混乱([Zhou and Zafarani](#_bookmark195),[2020](#_bookmark195);[Kaliyar et al.](#_bookmark100),[2021](#_bookmark100);[Guo et al.](#_bookmark82),[2020](#_bookmark82)).它甚至可以用来生成假的地球卫星图像

2022年2月8日收到；2022年7月18日收到修订版；2022年7月27日接受

2022年7月31日上线

Elsevier Inc .保留所有权利。



图一。2016-2021年deepfakes相关论文数量，从[https://app.dimensions.ai](https://app.dimensions.ai/)在2021年底，搜索关键字“deepfake”适用于学术论文的全文。

包含并不真正存在的对象以迷惑军事分析家，例如，创建一座横跨河流的假桥，尽管实际上并没有这样的桥。这可能会误导在战斗中被引导过桥的部队([Tucker](#_bookmark176),[2019](#_bookmark176);[Fish](#_bookmark69),[2019](#_bookmark69)).

由于创建逼真的数字人的民主化具有积极的意义，因此deepfakes也有积极的用途，例如它们在视觉效果、数字化身、snapchat过滤器中的应用，为失去声音的人创建声音，或者在不重拍的情况下更新电影片段([Marr](#_bookmark133),[2019](#_bookmark133)).Deepfakes可以在摄影、视频游戏、虚拟现实、电影制作和娱乐中产生创造性或生产性影响，例如，外国电影的逼真视频配音、通过历史人物的复活进行教育、购物时虚拟试穿衣服等等([Mirsky and Lee](#_bookmark138),[2021](#_bookmark138);[Verdoliva](#_bookmark178),[2020](#_bookmark178)).然而，恶意使用deepfakes的数量在很大程度上超过了正面使用。先进的深度神经网络的发展和大量数据的可用性使得伪造的图像和视频对于人类甚至复杂的计算机算法来说几乎无法区分。如今，创建这些经过处理的图像和视频的过程也简单得多，因为只需要一张目标个人的身份照片或一段简短的视频。越来越少的努力需要制作一个令人震惊的令人信服的镜头。最近的进展甚至可以创建一个只有静止图像的深度假([Zakharov et al.](#_bookmark189),[2019](#_bookmark189)).因此，Deepfakes不仅会影响公众人物，也会影响普通人。例如，一个声音deepfake被用来诈骗一位首席执行官243，000美元([Damiani](#_bookmark50),[2019](#_bookmark50)).最近发布的一款名为DeepNude的软件显示了更多令人不安的威胁，因为它可以将一个人转变为未经同意的色情([Samuel](#_bookmark160),[2019](#_bookmark160)).同样，中国应用程序Zao最近也火了起来，因为不太熟练的用户可以将自己的脸贴到电影明星的身上，并将自己插入知名电影和电视片段([The Guardian](#_bookmark169),[2019](#_bookmark169)).这些伪造形式对侵犯隐私和身份造成了巨大威胁，并影响到人类生活的许多方面。

因此，在数字领域寻找真相变得越来越重要。当处理deepfakes时，这甚至更具挑战性，因为它们主要用于恶意目的，并且现在几乎任何人都可以使用现有的deepfake工具来创建deepfakes。迄今为止，已经提出了许多方法来检测deepfakes([Lyu](#_bookmark128),[2020](#_bookmark128);[Guarnera et al.](#_bookmark79),[2020c](#_bookmark79);[Jafar et al.](#_bookmark97),[2020](#_bookmark97);[Trinh et al.](#_bookmark175),[2021](#_bookmark175);[Younus and Hasan](#_bookmark188),[2020](#_bookmark188)).它们中的大多数都是基于深度学习的，因此，深度学习方法的恶意和积极使用之间的战斗已经出现。为了应对面部交换技术或深度伪造的威胁，美国国防高级研究计划局(DARPA)启动了一项媒体取证研究计划(名为媒体取证或媒体取证)以加速开发

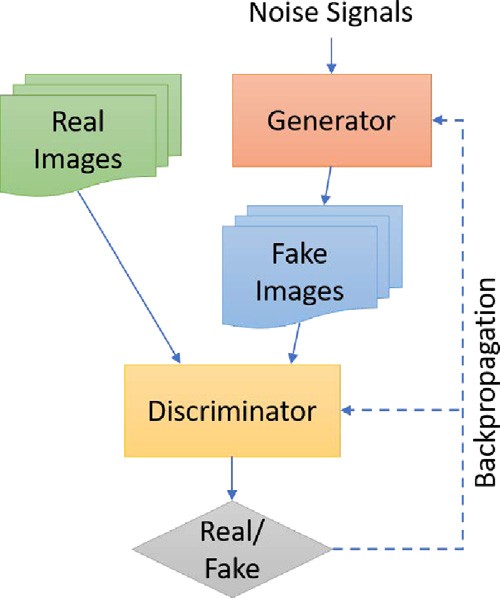
图二。与deepfake检测方法相关的评论论文类别，其中我们将论文分为两大类，即伪图像检测和人脸视频检测。

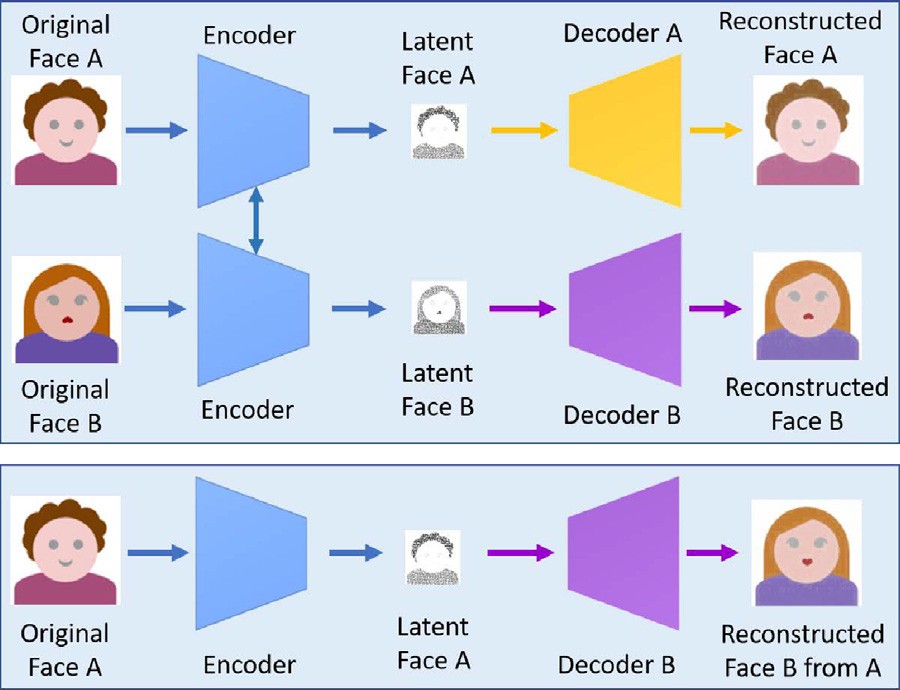
伪数字视觉媒体检测方法([Turek](#_bookmark177),[2019](#_bookmark177)).最近，脸书公司与微软公司和人工智能联盟合作发起了Deepfake检测挑战，以促进更多的研究和开发，检测和防止deepfakes被用来误导观众([Schroepfer](#_bookmark162),[2019](#_bookmark162)).数据来自[https://app.dimensions.ai](https://app.dimensions.ai/)2021年底显示，近年来deepfake论文数量大幅增加([Fig. 1](#_bookmark8)).虽然获得的深度造假论文数量可能低于实际数量，但这一课题的研究趋势明显增强。

已有关于创建和检测deepfakes的调查论文，请参见[Tolosana et al.](#_bookmark174)([2020](#_bookmark174)),[Verdoliva](#_bookmark178)([2020](#_bookmark178))和[Mirsky and Lee](#_bookmark138) ([2021](#_bookmark138)).举个例子，[Mirsky and Lee](#_bookmark138) ([2021](#_bookmark138))侧重于再现方法(即，改变目标的表情、嘴型、姿势、凝视或身体)，以及替换方法(即，通过交换或转移方法替换目标的面部)。[Verdoliva](#_bookmark178)([2020](#_bookmark178))将检测方法分为常规方法(例如，不使用任何外部数据进行训练的盲方法、基于一类传感器和基于模型的方法、以及具有手工制作特征的监督方法)和基于深度学习的方法(例如，CNN模型)。[Tolosana et al.](#_bookmark174)([2020](#_bookmark174))根据创建deepfakes的方式对创建和检测方法进行了分类，包括全脸合成、身份交换、属性操作和表情交换。另一方面，我们用不同的视角和分类法进行调查。我们根据数据类型(即图像或视频)对deepfake检测方法进行分类，如中所示[Fig.2](#_bookmark9).对于伪图像检测方法，我们关注所使用的特征，即它们是手工特征还是深层特征。对于伪视频检测方法，基于该方法是使用跨帧的时间特征还是视频帧内的视觉伪像，来识别两个主要的子类别。我们还广泛讨论了深度伪造检测和多媒体取证问题的挑战、研究趋势和方向。

# Deepfake创作

Deepfakes由于篡改视频的质量以及其应用程序对从专业到新手的各种计算机技能的广泛用户的易用性而变得流行。这些应用大多是基于深度学习技术开发的。深度学习以其表示复杂和高维数据的能力而闻名。具有这种能力的深度网络的一个变体是深度自动编码器，它已经被广泛使用[applied](#_bookmark149) 对于维数减少和图像压缩([Punnap-purath and Brown](#_bookmark149),[2019](#_bookmark149);[Cheng et al.](#_bookmark34),[2019](#_bookmark34);[Chorowski et al.](#_bookmark44),[2019](#_bookmark44)).deepfake创建的第一次尝试是FakeApp，由Reddit用户使用自动编码器-解码器配对结构([Faceswap](#_bookmark62),[2022](#_bookmark62);[FakeApp](#_bookmark64),[2022](#_bookmark64)).在该方法中，自动编码器提取人脸图像的潜在特征，解码器用于重建人脸



图3。使用两个编码器-解码器对的deepfake创建模型。两个网络使用相同的编码器，但不同的解码器用于训练过程(上图)。用普通编码器对A面的图像进行编码，并用解码器B进行解码，以创建一个深度赝品(下图)。重建的图像(在底部)是具有面部a的嘴形状的面部B。面部B最初具有颠倒的心脏的嘴，而重建的面部B具有传统心脏的嘴。

图像。为了在源图像和目标图像之间交换面部，需要两个编码器-解码器对，其中每对用于在图像集上训练，并且编码器的参数在两个网络对之间共享。换句话说，两对具有相同的编码器网络。这种策略使得普通编码器能够找到并学习两组人脸图像之间的相似性，这相对来说并不具有挑战性，因为人脸通常具有相似的特征，例如眼睛、鼻子、嘴的位置。[Fig. 3](#_bookmark10) 显示了一个deepfake创建过程，其中人脸A的特征集与解码器B相连接，以从原始人脸A重建人脸B。这种方法应用于多个作品中，如DeepFaceLab([DeepFaceLab](#_bookmark52),[2022b](#_bookmark52))，DFaker([DFaker](#_bookmark54),[2022](#_bookmark54))，DeepFake\_tf(基于tensorflow的deepfakes)([DeepFake\_tf](#_bookmark53),[2022](#_bookmark53)).通过添加在VGGFace([Keras-VGGFace](#_bookmark104),[2022](#_bookmark104))到编码器-解码器架构，基于生成对抗网络的deepfakes的改进版本([Goodfellow et al.](#_bookmark75),[2014](#_bookmark75))，即faceswap-GAN，是在[Faceswap-GAN](#_bookmark63)([2022](#_bookmark63)).VGGFace感知损失被添加到

图4。GAN架构由一个发生器和一个鉴别器组成，每个都可以通过神经网络来实现。整个系统可以用反向传播来训练，这允许两个网络提高它们的能力。

经过充分的训练后，两个网络都提高了它们的能力，即，发生器𝐺能够产生与真实图像非常相似的图像，而鉴别器𝐷能够很好地将伪图像与真实图像区分开。

[Table1](#_bookmark12)总结了流行的deepfake工具及其典型特征。其中，一个突出的方法是人脸合成[based](#_bookmark102)上的GAN模型，即StyleGAN，是在[Karraset al.](#_bookmark102)([2019](#_bookmark102)).StyleGAN的动机是风格转移([Huang et al.](#_bookmark91),[2017](#_bookmark91))具有能够创建逼真人脸图像的特殊生成器网络架构。在传统的GAN模型中，例如GAN的渐进生长(PGGAN)([Karras et al.](#_bookmark101),[2017](#_bookmark101))，信号噪声(潜在代码)被馈送到代表发生器的前馈网络的输入层。在StyleGAN中，有两个网络被构建并链接在一起，一个是𝑓绘图网络，另一个是𝑔.合成网络潜在代码𝑧 ∈ 𝑍首先被转换成𝑤 ∈ 𝑊(其中

*𝑊是一个中间潜在空间)通过一个非线性函数𝑓*

*𝑍 → 𝑊，其特征在于由几个完全连接的层组成的神经网络(即映射网络)。使用仿射变换，中间表示𝑤专用于样式𝑦 = (𝑦𝑠，𝑦𝑏)，该样式将被提供给自适应实例规范化(AdaIN)操作，指定为:*

使眼球运动更加真实，与输入更加一致

𝑦adain(𝑥)=𝑦

*𝑥𝑖 − 𝜇(𝑥𝑖) + 𝑦*

(2)

面孔，有助于消除分割遮罩中的伪像，引导

*𝑖 𝑠,𝑖*

*𝜎(𝑥𝑖)*

*𝑏,𝑖*

到更高质量的输出视频。该模型有助于创建64 × 64、128 × 128和256 × 256分辨率的输出。另外，来自FaceNet实现的多任务卷积神经网络(CNN)([FaceNet](#_bookmark61),[2022](#_bookmark61))用于进行人脸检测[more](#_bookmark48)稳定和面部对齐更可靠。周期根([Cycle-GAN](#_bookmark48),[2022](#_bookmark48))用于该模型中的生成网络实现。

传统的GAN模型包括两个神经网络:生成器和鉴别器，如[Fig.4](#_bookmark11).给定具有𝑝𝑑𝑎𝑡𝑎分布的真实图像𝑥的数据集，生成器的目标

*𝐺将产生类似于真实图像𝑥的图像𝐺(𝑧，其中𝑧是具有𝑝𝑧.分布的噪声信号鉴别器的目的*

*𝐺要正确分类𝐺生成的图像和𝑥.的真实图像训练鉴别器𝐷以提高其分类能力，即最大化𝐷(𝑥，这表示𝑥是真实图像而不是𝐺.生成的伪造图像的概率另一方面，𝐺被训练以最小化其输出被𝐷分类为合成图像的概率，即最小化1 𝐷(𝐺(𝑧)).这是两个玩家𝐷和𝐺之间的极小极大博弈，可以用下面的价值函数来描述(*[Goodfellow et al.](#_bookmark75),[2014](#_bookmark75)):

最小最大𝑉 (𝐷，𝐺) = E𝑥∼𝑝 (𝑥)[log 𝐷(𝑥)]

其中每个特征图𝑥𝑖被单独归一化。StyleGAN生成器架构允许通过不同的比例修改样式来控制图像合成。此外，该方法使用两个潜在代码来生成给定比例的图像，而不是在训练期间使用一个随机潜在代码。更具体地，两个潜在代码𝑧1和𝑧2被分别馈送到映射网络来创建

*𝑤1和𝑤2通过在交叉点之前应用𝑤1和在交叉点之后应用𝑤2来控制样式。*[Fig. 5](#_bookmark13) 演示通过在三种不同比例下混合两种潜在代码创建的图像示例，其中每个样式子集控制图像的独立的有意义的高级属性。换句话说，StyleGAN的生成器架构能够学习高级属性的分离(例如，当在人脸上训练时的姿势和身份)，并且能够对人脸合成进行直观的、特定比例的控制。

# 深度伪造检测

Deepfake检测通常被认为是一个二元分类问题，其中分类器用于在真实视频和篡改视频之间进行分类。这种方法需要一个大的真假视频数据库来训练分类模型。假冒的数量

*𝐺 𝐷*

*𝑑𝑎𝑡𝑎*

+e𝑧∼𝑝𝑧(𝑧)[log(1-𝐷(𝐺(𝑧)))](1)

视频越来越可用，但在为验证各种检测方法设定基准方面仍然有限。解决

## 表1

著名deepfake工具总结。 工具链接主要功能

面部交换<https://github.com/deepfakes/faceswap> -使用两对编码器-解码器。

-共享编码器的参数。

面部交换-甘<https://github.com/shaoanlu/faceswap-GAN> 对抗损失和感知损失(VGGface)被添加到

自动编码器架构。

少镜头人脸平移<https://github.com/shaoanlu/fewshot-face-translation-GAN> -使用预先训练的人脸识别模型来提取潜在的

氮化镓加工的嵌入技术。

* 合并从FUNIT模块获得的语义先验([Liuet al.](#_bookmark124),[2019](#_bookmark124))和黑桃([Park et al.](#_bookmark146),[2019](#_bookmark146)).

DeepFaceLab<https://github.com/iperov/DeepFaceLab> -从Faceswap方法扩展到新型号，例如H64，

H128、LIAEF128、SAE([DeepFaceLab](#_bookmark51),[2022a](#_bookmark51)).

* 支持多种人脸提取模式，如S3FD、MTCNN、dlib或manual([DeepFaceLab](#_bookmark51),[2022a](#_bookmark51)).

制造商<https://github.com/dfaker/df> - DSSIM损失函数([DSSIM](#_bookmark59),[2022](#_bookmark59))用于重建人脸。

* 基于Keras库实现。

DeepFake\_tf<https://github.com/StromWine/DeepFake_tf> 类似于DFaker，但基于tensorflow实现。AvatarMe<https://github.com/lattas/AvatarMe> -从任意“野外”图像中重建3D人脸。

* 可以通过6K分辨率的3D人脸从单个低分辨率图像中重建真实的4K([Lattas et al.](#_bookmark113),[2020](#_bookmark113)).

牵线木偶<https://hyperconnect.github.io/MarioNETte> -保留目标的少数镜头脸部重现框架

身份。

* 身份自适应不需要额外的微调阶段([Ha et al.](#_bookmark84),[2020](#_bookmark84)).

DiscoFaceGAN<https://github.com/microsoft/DiscoFaceGAN> -生成具有独立潜伏的虚拟人的面部图像

身份、表情、姿势和照明的变量。

* 将3D先验嵌入对抗学习([Deng et al.](#_bookmark55),[2020](#_bookmark55)).

StyleRig<https://gvv.mpi-inf.mpg.de/projects/StyleRig> -使用钻机般的控件创建面部肖像图像

通过3D可变形人脸模型进行预训练和固定造型。

* 无需手动注释的自我监督([Tewari et al.](#_bookmark167),[2020](#_bookmark167)).

变脸者<https://lingzhili.com/FaceShifterPage> -通过开发和集成高保真度的面部交换

目标属性。

* 可以应用于任何新的人脸对，而不需要特定对象的训练([Li et al.](#_bookmark114),[2019a](#_bookmark114)).

FSGAN<https://github.com/YuvalNirkin/fsgan> -面部交换和重现模型，可应用于

成对的面孔，而不需要对这些面孔进行训练。

* 调整姿势和表情变化([Nirkin et al.](#_bookmark144),[2019](#_bookmark144)).

StyleGAN<https://github.com/NVlabs/stylegan> -基于风格提出了一种新的GANs生成器架构

转让文献。

* 新的体系结构导致高级属性的自动、无监督的分离，并且能够对图像的合成进行直观的、特定比例的控制([Karras et al.](#_bookmark102),[2019](#_bookmark102)).

Face2Face<https://justusthies.github.io/posts/face2face/> -单目目标视频序列的实时面部再现，

例如Youtube视频。

* 由源演员制作目标视频的面部表情动画，并在

照片级时尚([Thies et al.](#_bookmark172),[2016](#_bookmark172)).

神经纹理<https://github.com/SSRSGJYD/NeuralTexture> -作为场景捕捉过程的一部分而学习的特征地图

并作为地图存储在3D网格代理的顶部。

* 可以在静态和动态环境中以实时速率一致地重新呈现或操纵现有视频内容([Thies](#_bookmark171)

[et al.](#_bookmark171),[2019](#_bookmark171)).

可转换瓶颈网络

“照我做的做”运动转移

<https://github.com/kyleolsz/TB-Networks> -一种对图像内容进行精细3D操作的方法。

* 使用可转换瓶颈框架在CNN模型中应用空间转换([Olszewski et al.](#_bookmark145),[2019](#_bookmark145)).

<https://github.com/carolineec/EverybodyDanceNow> -自动将运动从源转移到目标

通过学习视频到视频的翻译。

* 可以创建具有多个主题的动作同步舞蹈视频([Chan et al.](#_bookmark33),[2019](#_bookmark33)).

神经发声木偶<https://justusthies.github.io/posts/neural-voice-puppetry> -一种音频驱动的面部视频合成方法。

-从音频序列中合成一个正在说话的头部的视频

另一个人使用三维人脸图像。([Thies et al.](#_bookmark170),[2020](#_bookmark170)).

这个问题，[Korshunov and Marcel](#_bookmark110)([2019](#_bookmark110))使用开源代码Faceswap-GAN([Faceswap-GAN](#_bookmark63),[2022](#_bookmark63)).来自公开可用的VidTIMIT数据库的视频([VidTIMIT database](#_bookmark179),[2022](#_bookmark179))用于生成低质量和高质量的deepfake视频，这些视频可以有效地模仿面部表情、嘴部动作和眨眼。这些视频随后被用来测试各种deepfake检测方法。测试结果表明，流行的基于VGG的人脸识别系统([Parkhi et al.](#_bookmark147),[2015](#_bookmark147))和Facenet([FaceNet](#_bookmark61),[2022](#_bookmark61);[Schroff et al.](#_bookmark163),[2015](#_bookmark163))不能有效地检测deepfakes。

诸如唇同步方法的其他方法([Chung et al.](#_bookmark45),[2017](#_bookmark45);[Suwajanakorn et al.](#_bookmark166),[2017](#_bookmark166);[Korshunov and Marcel](#_bookmark109),[2018b](#_bookmark109))和图像[quality](#_bookmark72) 支持向量机(SVM)度量标准([Galbally and Mar-cel](#_bookmark72),[2014](#_bookmark72))在应用于从这个新产生的数据集中检测deepfake视频时产生非常高的错误率。这引起了人们对未来发展更强大的方法的迫切需要的关注，这些方法可以从真品中检测出深度赝品。

本节介绍了深度假冒检测方法的概况，其中我们将它们分为两大类:假冒图像检测方法和假冒视频检测方法([Fig.2](#_bookmark9)).后者是有区别的

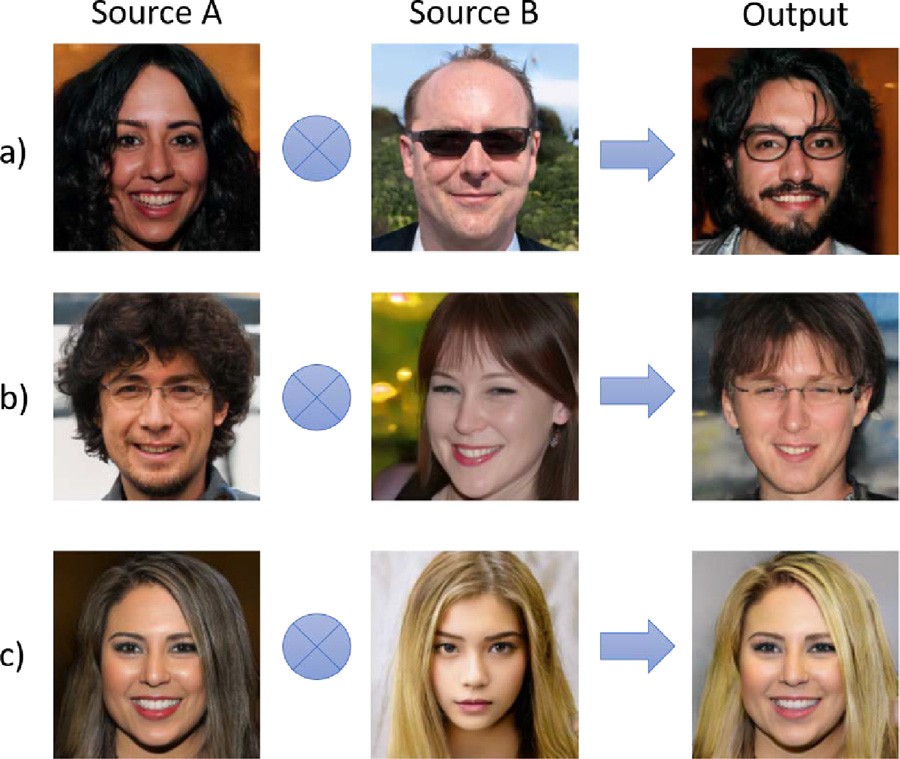


图5。使用StyleGAN混合样式的示例:输出图像是通过从源B复制指定的样式子集并从源获取其余部分来生成的

形容词（adjective的缩写）(A)从源B复制粗略样式将生成具有来自源B的高级特征以及来自源A的所有颜色和更精细面部特征的图像；(B)如果从B复制中等分辨率的风格，则输出图像将具有来自B的较小比例的面部特征，并且保留来自A的姿势、一般面部形状和眼镜；(c)如果从源B复制精细样式，则生成的图像将具有源B的色彩方案和微观结构([Karras et al.](#_bookmark102),[2019](#_bookmark102)).

分成两个小组:基于单个视频帧的方法中的视觉伪像和基于帧的方法中的时间特征。虽然大多数基于时间特征的方法使用深度学习递归分类模型，但是使用视频帧内的视觉伪像的方法可以由深度或浅层分类器来实现。

* 1. *伪图像检测*

Deepfakes越来越不利于隐私、社会安全和民主([Chesney and Citron](#_bookmark35),[2018a](#_bookmark35)).一旦这种威胁被引入，检测deepfakes的方法已经被提出。早期的尝试是基于从赝品图像合成过程中的赝品和不一致性获得的手工特征。最近的方法，例如，[de Lima et al.](#_bookmark121)([2020](#_bookmark121))和[Amerini and Caldelli](#_bookmark23)([2020](#_bookmark23))，已经普遍应用深度学习来自动提取显著的和有区别的特征，以检测deepfakes。

* + 1. *基于手工特征的方法*

尽管GAN的开发正在进行，并且频繁引入了许多新的GAN扩展，但是大多数关于GAN生成的图像的检测的工作没有考虑检测模型的泛化能力。[Xuan et al.](#_bookmark183) ([2019](#_bookmark183))使用图像预处理步骤，例如高斯模糊和高斯噪声，以去除GAN图像的低水平高频线索。这增加了真实图像和伪造图像之间的像素级统计相似性，并且允许取证分类器学习更多内在的和有意义的特征，这比先前的图像取证方法([Yang et al.](#_bookmark186),[2016](#_bookmark186);[Bayar and Stamm](#_bookmark27),[2016](#_bookmark27))或图像隐写分析网络([Qian et al.](#_bookmark150),[2015](#_bookmark150)).[Zhang et al.](#_bookmark191)([2017](#_bookmark191))使用单词袋方法来提取一组紧凑特征，并将其馈送到各种分类器，例如SVM([Wang et al.](#_bookmark181),[2017](#_bookmark181))，随机森林(RF)([Bai](#_bookmark25),[2017](#_bookmark25))和多层感知器(MLP)([Zheng et al.](#_bookmark193),[2016](#_bookmark193))用于区分交换的人脸图像和真实的人脸图像。在深度学习生成的图像中，由GAN模型合成的图像可能是最难检测的，因为它们基于GAN学习复杂输入数据的分布并生成具有相似输入分布的新输出的能力而具有真实感和高质量。

另一方面，[Agarwal and Varshney](#_bookmark22)([2019](#_bookmark22))将基于GAN的deepfake检测作为假设测试问题，其中使用认证的信息论研究引入了统计框架([Maurer](#_bookmark137),[2000](#_bookmark137)).定义了合法图像和由特定GAN生成的图像的分布之间的最小距离，即oracle误差。分析结果表明，当GAN不太精确时，该距离增加，并且在这种情况下，更容易检测到deepfakes。在高分辨率图像输入的情况下，需要极其精确的GAN来产生这种方法难以检测的伪图像。

* + 1. *基于深层特征的方法*

人脸交换在视频合成、肖像变形，特别是身份保护方面有许多引人注目的应用，因为它可以用一组库存图像中的人脸代替照片中的人脸。然而，它也是网络攻击者用来渗透识别或认证系统以获得非法访问的技术之一。深度学习(如CNN和GAN)的使用使得交换的人脸图像对取证模型来说更具挑战性，因为它可以保留照片的姿势、面部表情和光照([Korshunova et al.](#_bookmark111),[2017](#_bookmark111)).

[Hsu et al.](#_bookmark90)([2020](#_bookmark90))介绍了一种用于检测deepfake图像的两阶段深度学习方法。第一阶段是基于普通伪特征网络(CFFN)的特征提取器[Chopra et al.](#_bookmark43)([2005](#_bookmark43))被使用。CFFN包含几个密集单元，每个单元包括不同数量的密集块([Huang et al.](#_bookmark91),[2017](#_bookmark91))以提高输入图像的代表能力。基于成对信息的使用，通过CFFN学习过程提取伪图像和真实图像之间的区别特征，**成对信息是每对两个输入图像的标签。如果两幅图像属于同一类型，即假-假或真-真，则成对标签为1。相反，如果它们属于不同的类型，即假-真，则成对标签为0**。然后，基于CFFN的区别特征被馈送到神经网络分类器，以区分欺骗性图像和真实图像。该方法对假人脸和假普通图像检测都有效。一方面，人脸数据集是从CelebA([Liu et al.](#_bookmark126),[2015](#_bookmark126))，包含10，177个身份和202，599个各种姿势和背景杂波的对齐人脸图像。五种GAN变体用于生成大小为64 × 64的假图像，包括深度卷积GAN (DCGAN)([Radford et al.](#_bookmark151),[2015](#_bookmark151))，瓦瑟斯坦·甘(WGAN)([Arjovsky et al.](#_bookmark26),[2017](#_bookmark26))，带梯度惩罚的WGAN

(WGAN-GP)([Gulrajani et al.](#_bookmark81),[2017](#_bookmark81))，最小二乘GAN([Mao et al.](#_bookmark131),

[2017](#_bookmark131))，以及PGGAN([Karras et al.](#_bookmark101),[2017](#_bookmark101)).为了验证所提出的方法，总共获得了385，198幅训练图像和10，000幅真实和虚假测试图像。另一方面，从ILSVRC12提取一般数据集([Russakovsky et al.](#_bookmark157),[2015](#_bookmark157)).高保真自然图像的大规模GAN训练模型[synthesis](#_bookmark190)(比根) ([Brock et al.](#_bookmark29),[2018](#_bookmark29))，自我关注甘([Zhanget al.](#_bookmark190),[2019](#_bookmark190))和光谱归一化GAN([Miyato et al.](#_bookmark140),[2018](#_bookmark140))用于生成大小为128 × 128的假图像。训练集包括600，000幅假图像和真实图像，而测试集包括10，000幅两种类型的图像。实验结果显示了所提出的方法相对于其竞争方法(例如在中介绍的那些方法)的优越性能[Farid](#_bookmark65) ([2009](#_bookmark65)),[Mo et al.](#_bookmark141) ([2018](#_bookmark141)),[Marra et al.](#_bookmark134)([2018](#_bookmark134))和[Hsu et al.](#_bookmark89)([2018](#_bookmark89)).

同样的，[Guo et al.](#_bookmark83)([2021](#_bookmark83))提出了一个CNN模型，即SCnet，来检测deepfake图像，这些图像是由基于Glow的面部伪造工具([Kingma and Dhariwal](#_bookmark105),[2018](#_bookmark105)).由发光模型合成的假图像([Kingma and Dhariwal](#_bookmark105),[2018](#_bookmark105))把面部表情恶意篡改了。这些图像是超现实的，具有完美的视觉质量，但它们仍然有微妙或明显的操纵痕迹，这是由SCnet利用的。由于分层特征提取块，SCnet能够自动学习图像数据的高级取证特征，该分层特征提取块由四个卷积层堆叠而成。每一层学习一个新的

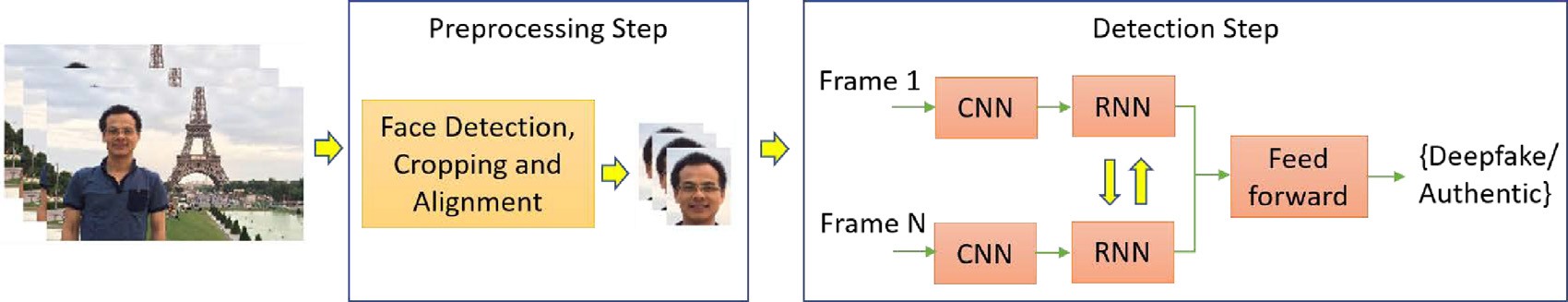
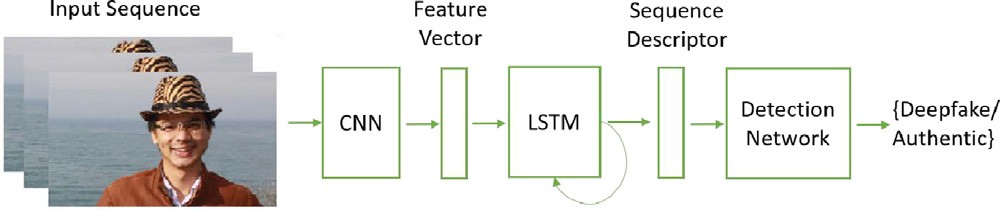


图6。一种用于人脸操纵检测的两步过程，其中预处理步骤旨在检测、裁剪和对齐一系列帧上的人脸，第二步通过组合卷积神经网络(CNN)和递归神经网络(RNN)来区分被操纵的和真实的人脸图像([Sabir et al.](#_bookmark158),[2019](#_bookmark158)).

前一层的一组特征映射，每个卷积运算定义如下:

*𝑖*

*𝑓 (𝑛) = ∑ 𝑓 (𝑛−1) ∗ 𝜔(𝑛) + 𝑏(𝑛)*

*𝑗*

*𝑖*

*𝑖𝑗*

*𝑗*

*𝑖=1*

(3)

其中𝑓 (𝑛)是𝑛th图层的𝑗th要素地图，𝜔(𝑛)的权重是

图7。一种利用卷积神经网络和

*𝑗*

𝑖 𝑗

*𝑖𝑗 th层，以及𝑏(𝑛)*

长短期记忆(LSTM)提取给定视频序列的时间特征，

𝑛 𝑗中第个卷积核的第个信道

是𝑛th层中𝑗th卷积核的偏差项。使用包含321，378的数据集来评估所提出的方法[face](#_bookmark105)通过应用发光模型创建的图像([Kingmaand Dhariwal](#_bookmark105),[2018](#_bookmark105))到CelebA人脸图像数据集([Liu et al.](#_bookmark126),[2015](#_bookmark126)).评价结果表明，SCnet模型具有较高的准确性[and](#_bookmark18)比中提出的Meso-4模型更好的概括[Afcharet al.](#_bookmark18)([2018](#_bookmark18)).

最近，[Zhao et al.](#_bookmark192)([2021](#_bookmark192))提出了一种利用局部源特征的自洽性进行深度伪造检测的方法，所述局部源特征是图像的与内容无关的空间局部信息。这些特征可以来自成像管道、编码方法或图像合成方法。**假设修改后的图像在不同位置具有不同的源特征，而原始图像在不同位置具有相同的源特征。这些源特征以下采样特征图的形式表示，由CNN模型使用称为成对自洽学习的特殊表示学习方法提取。**这种学习方法旨在惩罚引用来自同一图像的位置的具有低余弦相似性得分的特征向量对。同时，它也惩罚来自不同图像的具有高相似性得分的对。学习的特征图然后被馈送到用于深度伪造检测的分类方法。在七个流行的数据集上对该方法进行了评估，包括[FaceForensics](#_bookmark60)++ ([Rossler et al.](#_bookmark156),[2019](#_bookmark156))，DeepfakeDetection([Dufour andGully](#_bookmark60),[2019](#_bookmark60))，Celeb-DF-v1 & Celeb-DF-v2([Li et al.](#_bookmark119),[2020b](#_bookmark119))，深度假货检测挑战赛(DFDC)([Dolhansky et al.](#_bookmark56),[2020](#_bookmark56))，DFDC预览版([Dolhansky et al.](#_bookmark57),[2019](#_bookmark57))，以及DeeperForensics-1.0([Jiang et al.](#_bookmark99),[2020](#_bookmark99)).实验结果表明，该方法优于目前最先进的方法。然而，当处理由直接输出整个图像的方法生成的伪图像时，它可能具有局限性，该整个图像的源特征在每个图像内的所有位置上都是一致的。

* 1. *虚假视频检测*

大多数图像检测方法不能用于视频，因为视频压缩后帧数据的严重退化([Afchar et al.](#_bookmark18),[2018](#_bookmark18)).此外，视频具有在各组帧之间变化的时间特性，因此它们对于被设计成仅检测静态伪图像的方法来说是具有挑战性的。本小节重点介绍深度伪造视频检测方法，并将其分为两个小组:采用时间特征的方法和探索帧内视觉伪像的方法。

* + 1. *跨视频帧的时间特征*

基于在deepfakes的合成过程中没有有效地加强时间一致性的观察，[Sabir et al.](#_bookmark158)([2019](#_bookmark158))

它们通过序列描述符来表示。由全连接层组成的检测网络被用于将序列描述符作为输入，并计算帧序列属于真实或深度伪造类([Güera and Delp](#_bookmark80),[2018](#_bookmark80)).

利用视频流的时空特征来检测deepfakes。视频操作是在逐帧的基础上进行的，因此由面部操作产生的低级伪影被认为进一步表现为跨帧的不一致的时间伪影。基于卷积网络的集成，提出了递归卷积模型[DenseNet](#_bookmark41)([Huang et al.](#_bookmark91),[2017](#_bookmark91))和门控重现单位细胞([Choet al.](#_bookmark41),[2014](#_bookmark41))来利用跨帧的时间差异(参见[Fig. 6](#_bookmark14)).所提出的方法在包括1000个视频([Rossler et al.](#_bookmark156),[2019](#_bookmark156))，并显示出有希望的结果。

同样的，[Güera and Delp](#_bookmark80)([2018](#_bookmark80))强调deepfake视频包含帧内不一致和帧间时间不一致。然后，他们提出了时间感知管道方法，使用CNN和长短期记忆(LSTM)来检测deepfake视频。CNN用于提取帧级特征，然后将这些特征送入LSTM以创建时间序列描述符。完全连接的网络最终用于根据序列描述符将篡改的视频与真实视频进行分类，如所示[Fig. 7](#_bookmark15).使用600个视频的数据集获得了大于97%的准确率，其中包括从多个视频托管网站收集的300个deepfake视频和从[Laptev et al.](#_bookmark112)([2008](#_bookmark112)).

另一方面，2000年提出了使用一种生理信号，即眨眼，来检测深度假货[Li et al.](#_bookmark116)([2018](#_bookmark116))根据观察，deepfakes中的一个人眨眼的频率比未经过篡改的视频中的人少得多。一个健康的成年人通常会眨眼2到10秒，每次眨眼需要

然而，Deepfake算法经常使用在线可用的人脸图像进行训练，这些人脸图像通常显示睁开眼睛的人，即，在互联网上发布的很少图像显示闭着眼睛的人。因此，如果没有人眨眼的图像，deepfake算法就没有能力生成可以正常眨眼的假脸。换句话说，deepfakes中的眨眼率远低于正常视频。辨别真假视频，[Li et al.](#_bookmark116)([2018](#_bookmark116))裁剪视频中的眼睛区域，并将它们分布到长期递归卷积网络(LRCN)([Donahue et al.](#_bookmark58),[2015](#_bookmark58))进行动态预测。LRCN由基于CNN的特征提取器、基于长短期记忆(LSTM)的序列学习和基于全连接层的状态预测组成，以预测眼睛睁开和闭合的概率

状态。眨眼表现出强烈的时间依赖性，因此LSTM的实现有助于有效地捕捉这些时间模式。

最近，[Caldelli et al.](#_bookmark30)([2021](#_bookmark30))提出了使用光流来测量沿着帧序列的时间轴的信息，用于视频深度伪造检测。光流是在视频的两个时间上不同的帧上计算的矢量场，可以描述场景中物体的运动。预期光流场在合成创建的帧和自然生成的帧之间是不同的([Amerini et al.](#_bookmark24),[2019](#_bookmark24)).与原始视频相比，插入deepfake视频中的嘴唇、眼睛或整张脸的不自然运动将引入独特的运动模式。基于这种假设，由光流场组成的特征被输入到CNN模型中，用于区分deepfakes和原始视频。更具体地说，ResNet50架构([He et al.](#_bookmark85),[2016](#_bookmark85))被实现为用于实验的CNN模型。使用FaceForensics++数据集获得的结果([Rossler et al.](#_bookmark156),[2019](#_bookmark156))表明，这种方法在分类精度方面与最先进的方法不相上下。还实验了这种特征与基于帧的特征的组合，这导致了改进的deepfake检测性能。这证明了光流场在捕获视频帧时间轴上的不一致性用于深度伪造检测中的有用性。

* + 1. *视频帧中的视觉伪像*

正如在前面的小节中可以注意到的，使用跨视频帧的时间模式的方法主要基于深度逆流网络模型来检测深度伪造视频。这一小节研究通常将视频分解成帧的另一种方法，并探索单帧内的视觉伪像以获得鉴别特征。然后，这些特征被分配到深度或浅度分类器中，以区分真假视频。因此，我们在这一小节中根据分类器的类型对方法进行分组，即深度或浅度。

*深度分类器。Deepfake视频通常以有限的分辨率创建，这需要仿射人脸扭曲方法(即缩放、旋转和剪切)来匹配原始视频的配置。由于扭曲的面部区域和周围环境之间的分辨率不一致，这个过程留下了可以被CNN模型检测到的伪像，例如VGG16(*[Simonyan and Zisserman](#_bookmark164),[2014](#_bookmark164))、ResNet50、ResNet101和ResNet152([He et al.](#_bookmark85),[2016](#_bookmark85)).在中提出了一种深度学习方法，该方法基于在deepfake生成算法的面部扭曲步骤期间观察到的伪像来检测deep fake[Li and Lyu](#_bookmark118)([2018](#_bookmark118)).所提出的方法在两个deepfake数据集上进行评估，即UADFV和DeepfakeTIMIT。UADFV数据集([Yang et al.](#_bookmark185),[2019](#_bookmark185))包含49个真视频和49个假视频[with](#_bookmark109)总共32，752帧。DeepfakeTIMIT数据集([Korshunovand Marcel](#_bookmark109),[2018b](#_bookmark109))包括一组64 × 64大小的低质量视频和另一组128 × 128的高质量视频，每个质量组总共有10，537个原始图像和从320个视频中提取的34，023个伪造图像。将所提出的方法的性能与其他流行的方法进行比较，例如两种deepfake检测MesoNet方法，即Meso-4和MesoInception-4([Afchar et al.](#_bookmark18),[2018](#_bookmark18))，头部姿势([Yang et al.](#_bookmark185),[2019](#_bookmark185))，以及人脸篡改检测方法双流NN([Zhou et al.](#_bookmark194),[2017](#_bookmark194)).该方法的优点是在训练检测模型之前不需要生成deepfake视频作为负样本。相反，在将高斯模糊应用于随机选取的缩放图像并扭曲回原始图像之前，通过提取原始图像的面部区域并将其对齐到多个尺度来动态生成负样本。与需要预先生成深度伪像的其他方法相比，这减少了大量的时间和计算资源。[Nguyen et al.](#_bookmark143)([2019](#_bookmark143))提出了使用胶囊网络来检测被操纵的图像和视频。最初引入胶囊网络是为了解决当应用于逆图形任务时CNN的局限性，逆图形任务的目的是找到物理的

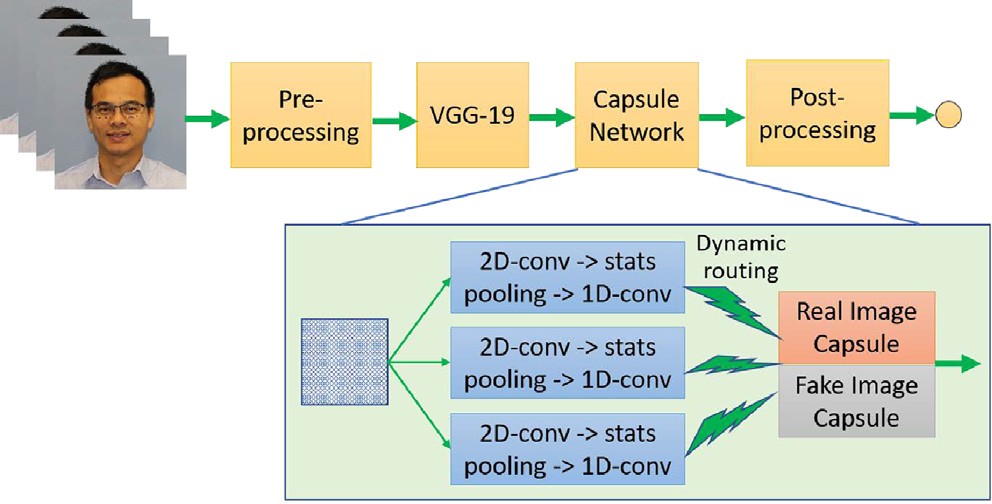


图8。胶囊网络采用从VGG-19网络获得的特征([Simonyanand Zisserman](#_bookmark164),[2014](#_bookmark164))来区分假图像或视频与真图像或视频(上)。预处理步骤检测人脸区域并将其缩放到128 × 128的大小，然后使用VGG-19提取胶囊网络的潜在特征，该胶囊网络包括三个主胶囊和两个输出胶囊，一个用于真实图像，一个用于虚假图像(下图)。统计池构成了处理伪造检测的胶囊网络的重要部分([Nguyen et al.](#_bookmark143),[2019](#_bookmark143)).

用来产生世界图像的过程([Hinton et al.](#_bookmark88),[2011](#_bookmark88)).基于动态路由算法的胶囊网络的最新发展([Sabour et al.](#_bookmark159),[2017](#_bookmark159))展示了其描述对象部分之间的层次姿态关系的能力。这种开发被用作检测伪造图像和视频的管道中的一个组件，如中所示[Fig. 8](#_bookmark16).采用动态路由算法，通过多次迭代将三个胶囊的输出路由到输出胶囊，以区分伪图像和真实图像。该方法通过覆盖大范围伪造图像和视频攻击的四个数据集进行评估。它们包括著名的Idiap研究所重放攻击数据集([Chingovska et al.](#_bookmark38),[2012](#_bookmark38))，deepfake面部交换数据集由[Afchar et al.](#_bookmark18)([2018](#_bookmark18))，面部重现FaceForensics数据集([Rössler et al.](#_bookmark155),[2018](#_bookmark155))，由Face2Face方法产生([Thies et al.](#_bookmark172),[2016](#_bookmark172))，以及由生成的完全计算机生成的图像数据集[Rahmouni et al.](#_bookmark152)([2017](#_bookmark152)).在所有这些数据集上，与竞争方法相比，所提出的方法产生了最佳的性能。这显示了胶囊网络在构建通用检测系统中的潜力，该系统可以有效地用于各种伪造图像和视频攻击。

*浅层分类器。深度伪造检测方法主要依赖于伪造和真实图像或视频之间的人为事实或内在特征的不一致性。*[Yang et al.](#_bookmark185) ([2019](#_bookmark185))提出了一种通过观察包括头部方向和位置的3D头部姿态之间的差异的检测方法，该3D头部姿态是基于中央面部区域的68个面部标志来估计的。检查3D头部姿态是因为在深度假脸生成管道中有一个缺点。将提取的特征送入SVM分类器，获得检测结果。在两个数据集上的实验表明，与同类方法相比，该方法具有更好的性能。第一个数据集，即UADFV，由49个深度伪视频和它们各自的真实视频([Yang et al.](#_bookmark185),[2019](#_bookmark185)).第二个数据集包括241个真实图像和252个深度伪图像，这是DARPA MediFor GAN图像/视频挑战中使用的数据子集([Guan et al.](#_bookmark76),[2019](#_bookmark76)).类似地，在中研究了一种基于眼睛、牙齿和面部轮廓的视觉特征来利用深度伪造和面部操纵的伪影的方法[Matern et al.](#_bookmark136)([2019](#_bookmark136)).视觉伪像源于缺乏全局一致性、对入射照明的错误或不精确的估计，或者对底层几何形状的不精确估计。对于deepfakes检测，利用眼睛和牙齿区域中丢失的反射和丢失的细节，以及基于面部标志从面部区域提取的纹理特征。因此，使用眼睛特征向量、牙齿特征向量和从全脸裁剪提取的特征。提取特征后，使用两种分类器，包括逻辑回归和小神经网络，对真实视频中的假货进行分类。对视频进行的实验

从YouTube下载的数据集显示，就接收器工作特性曲线下的面积而言，最佳结果为0.851。然而，所提出的方法有一个缺点，即要求图像满足某些先决条件，例如睁开眼睛或看到牙齿。

光响应非均匀性(PRNU)分析的使用是在1999年提出的[Koopman et al.](#_bookmark107) ([2018](#_bookmark107))从真品中检测出deepfakes。PRNU是传感器图案噪声的组成部分，其归因于硅晶片的制造缺陷以及由于硅晶片的物理特性的变化而导致的像素对光的不一致的灵敏度。PRNU分析被广泛使用[in](#_bookmark122) 图像取证([Rosenfeld and Sencar](#_bookmark154),[2009](#_bookmark154);[Li and Li](#_bookmark117),[2011](#_bookmark117);[Lin andLi](#_bookmark122),[2016](#_bookmark122);[Scherhag et al.](#_bookmark161),[2019](#_bookmark161);[Phan et al.](#_bookmark148),[2018](#_bookmark148))并提倡用在[Koopman et al.](#_bookmark107) ([2018](#_bookmark107))因为交换的面部应该改变视频帧的面部区域中的局部PRNU模式。视频被转换成帧，这些帧被裁剪到被质疑的面部区域。然后，被裁剪的帧被顺序分成八组，其中为每组计算平均PRNU模式。计算标准化的交叉相关分数，用于比较这些组中的PRNU模式。创建了一个测试数据集，包括10个真实视频和16个经过处理的视频，其中假视频是通过DeepFaceLab工具([DeepFaceLab](#_bookmark52),[2022b](#_bookmark52)).该分析显示了在深度赝品和真品之间的平均归一化互相关分数方面的显著统计差异。因此，这一分析表明

尽管数据集更大，但PRNU在深度伪造检测方面有潜力

需要进行测试。

当带着怀疑的眼光看一个视频或图像时，用户通常想要搜索它的来源。然而，这种工具目前还不可行。[Hasan and Salah](#_bookmark86)([2019](#_bookmark86))提出了使用区块链和智能合约来帮助用户检测deepfake视频，其基础是假设视频只有在其来源可追踪时才是真实的。每个视频都与一个链接到其父视频的智能合约相关联，并且每个父视频都有一个到其分层结构中的子视频的链接。通过这个链，即使视频已经被复制多次，用户也可以可信地追溯到与原始视频相关联的原始智能合同。智能契约的一个重要属性是星际文件系统的唯一散列，该文件系统用于以分散的和内容可寻址的方式存储视频及其元数据([IPFS](#_bookmark95),[2022](#_bookmark95)).智能合约的主要特性和功能针对几种常见的安全挑战进行了测试，例如分布式拒绝服务、重放和中间人攻击，以确保解决方案满足安全要求。这种方法是通用的，并且可以扩展到其他类型的数字内容，例如图像、音频和手稿。

# 讨论和未来的研究方向

在深度学习的支持下，deepfakes可以比以往任何时候都更容易地创建。由于社交媒体平台的发展，这些虚假内容的传播也更快([Zubiaga et al.](#_bookmark197),[2018](#_bookmark197)).有时候，假货并不需要传播给大量的观众来造成有害的影响。出于恶意目的创建deepfakes的人只需要将它们发送给目标受众，作为他们破坏策略的一部分，而无需使用社交媒体。例如，这种方法可以被试图影响政治家等重要人物决策的情报机构所利用，[leading](#_bookmark36)对国家和国际安全的威胁([Chesney andCitron](#_bookmark36),[2018b](#_bookmark36)).为了解决deepfake的报警问题，研究界已经致力于开发deepfake检测算法，并且已经报告了大量的结果。本文回顾了最先进的方法，并在中总结了典型的方法[Table 2](#_bookmark17).值得注意的是，那些使用先进的机器学习来创建deepfakes的人与那些努力检测deepfakes的人之间的战斗正在加剧。

Deepfakes的质量一直在提高，性能

检测方法需要相应改进。灵感

人工智能破坏的东西也可以被人工智能修复([Floridi](#_bookmark70),[2018](#_bookmark70)).检测方法仍处于其早期阶段，并且已经提出和评估了各种方法，但是使用片段数据集。一种提高检测方法性能的方法是创建deepfakes的不断增长的更新基准数据集，以验证检测方法的持续发展。这将促进检测模型的训练过程，尤其是那些基于深度学习的检测模型，其需要大的训练集([Dolhansky et al.](#_bookmark56),[2020](#_bookmark56)).

提高deepfake检测方法的性能非常重要，尤其是在交叉伪造和跨数据集的情况下。大多数检测模型是在相同伪造和数据集内实验中设计和评估的，这不能确保它们的泛化能力。一些先前的研究已经解决了这个问题，例如，在[Wang et al.](#_bookmark182)([2020](#_bookmark182)),[Caldelli et al.](#_bookmark30)([2021](#_bookmark30)),[Zhao et al.](#_bookmark192)([2021](#_bookmark192)),[Cozzolino et al.](#_bookmark47)([2018](#_bookmark47))和[Marra et al.](#_bookmark135)([2019](#_bookmark135))，但是在这个方向上还需要做更多的工作。一个接受过特定伪造训练的模型需要能够对抗另一个未知的伪造，因为潜在的deepfake类型在现实世界中通常是未知的。同样，当前的检测方法主要集中在deepfake生成管道的缺点上，即找到竞争对手的弱点来攻击它们。这种信息和知识在敌对环境中并不总是可用的，在敌对环境中，攻击者通常试图不暴露这种深度伪造的创建技术。最近关于恶意干扰攻击来欺骗基于DNN的检测器的工作使得深度伪造检测任务更加困难([Gandhi and Jain](#_bookmark73),[2020](#_bookmark73);[Hussain et al.](#_bookmark93),[2021](#_bookmark93);[Carlini and Farid](#_bookmark32),[2020](#_bookmark32);[Yang et al.](#_bookmark184),[2021](#_bookmark184);[Yeh et al.](#_bookmark187),[2020](#_bookmark187)).这些是检测方法开发的真正挑战，未来的研究需要专注于引入更鲁棒、可扩展和可推广的方法。

另一个研究方向是将检测方法整合到

社交媒体等分销平台，以提高其应对deepfakes广泛影响的有效性。使用有效检测方法的筛选或过滤机制可以是简单的[mented](#_bookmark36)在这些平台上，为了减轻deepfakes检测([Chesneyand Citron](#_bookmark36),[2018b](#_bookmark36)).法律可以要求拥有这些平台的科技公司迅速移除deepfakes，以减少其影响。此外，水印工具还可以集成到人们用来制作数字内容的设备中，以创建不可变的元数据来存储原创细节，如时间和位置[multimedia](#_bookmark36)内容及其未经篡改的证明([Chesneyand Citron](#_bookmark36),[2018b](#_bookmark36)).这种集成很难实现，但解决方案可能是使用颠覆性的区块链技术。区块链已在许多领域得到有效应用，但迄今为止很少有研究解决基于该技术的deepfake检测问题。因为它可以创建一系列唯一的不可更改的元数据块，所以它是数字起源解决方案的一个很好的工具。区块链技术对这个问题的集成已经证明了某些结果([Hasan and Salah](#_bookmark86),[2019](#_bookmark86))但这个研究方向还远未成熟。

使用检测方法来识别假货是至关重要的，但是-

了解发布deepfakes的人的真实意图更加重要。这需要用户根据发现deepfake的社会背景做出判断，例如，谁发布了它以及他们对它说了什么([Read](#_bookmark153),[2019](#_bookmark153)).这是至关重要的，因为deepfakes越来越逼真，人们高度期待检测软件将落后于deepfake创建技术。因此，研究deepfakes的社会背景以帮助用户做出这样的判断是值得的。

视频和照片已被广泛用作警方调查和司法案件的证据。它们可以由具有计算机或执法背景以及在收集、检查和分析数字信息方面的经验的数字媒体取证专家在法庭上作为证据介绍。机器学习和人工智能技术的发展可能已经被用来修改这些数字内容，因此专家的意见可能不足以鉴定这些证据，因为即使是专家也是如此

## 表2

著名deepfake检测方法综述。

方法分类器/技术

处理所用数据集的关键功能

眨眼([Li et al.](#_bookmark116),[2018](#_bookmark116))

帧内和时间不一致([Güeraand Delp](#_bookmark80),[2018](#_bookmark80))

LRCN——用LRCN来学习眨眼的时间模式。

-根据观察，deepfakes的闪烁频率比正常情况小得多。

CNN和LSTM CNN被用于提取帧级特征，

这些序列被分发给LSTM以构建用于分类序列描述符。

视频由49个采访和演示视频组成，以及它们对应生成的deepfakes。

视频从多个网站获得的600个视频的集合。

使用面部扭曲伪像([Li and Lyu](#_bookmark118),[2018](#_bookmark118))

VGG16([Simonyan andZisserman](#_bookmark164),[2014](#_bookmark164)),

ResNet模型([Heet al.](#_bookmark85),[2016](#_bookmark85))

基于扭曲的面部区域和周围环境之间的分辨率不一致，使用CNN模型来发现伪像。

视频- UADFV([Yang et al.](#_bookmark185),[2019](#_bookmark185))，包含49个真视频和49个假视频共32752帧。

-deepfakelimit([Korshunov and Marcel](#_bookmark109),[2018b](#_bookmark109))

MesoNet([Afchar et al.](#_bookmark18),[2018](#_bookmark18))

CNN -引入了两个深度网络，即Meso-4和MesoInception-4，以在介观分析级别检查deepfake视频。

-在deepfake和FaceForensics数据集上获得的准确率分别为98%和95%。

视频两个数据集:由在线视频构成的deepfake数据集和由Face2Face方法创建的FaceForensics数据集([Thies et al.](#_bookmark172),[2016](#_bookmark172)).

眼睛、教师和面部纹理([Matern et al.](#_bookmark136),[2019](#_bookmark136))

逻辑回归和神经网络

* 利用面部纹理差异，以及deepfakes的眼睛和牙齿区域缺少的反射和细节。
* 逻辑回归和神经网络用于分类。

视频从YouTube下载的视频数据集。

具有RCN的时空特征([Sabir et al.](#_bookmark158),[2019](#_bookmark158))

使用集成了卷积网络DenseNet([Huang et al.](#_bookmark91),[2017](#_bookmark91))和门控重现单位细胞([Cho et al.](#_bookmark41),[2014](#_bookmark41))

Videos FaceForensics++数据集，包括1000个视频([Rossler et al.](#_bookmark156),[2019](#_bookmark156)).

LSTM的时空特征([Chintha et al.](#_bookmark39),[2020](#_bookmark39))

卷积双向递归LSTM网络

* XceptionNet CNN用于面部特征提取，而音频嵌入通过堆叠多个卷积模块获得。
* 使用了两个损失函数，即交叉熵和kull back-lei bler散度。

视频面部取证++([Rossler et al.](#_bookmark156),[2019](#_bookmark156))和Celeb-DF (5639个deepfake视频) ([Li et al.](#_bookmark119),[2020b](#_bookmark119))数据集和ASVSpoof 2019逻辑访问音频数据集([Todisco et al.](#_bookmark173),[2019](#_bookmark173)).

PRNU的分析([Koopman et al.](#_bookmark107),[2018](#_bookmark107))

音素-视位不匹配([Agarwalet al.](#_bookmark20),[2020b](#_bookmark20))

PRNU -数码相机光敏传感器因其工厂缺陷而产生的噪声模式分析。

* + 探索真实和深度伪造视频之间的PRNU模式的差异，因为面部交换被认为会改变当地的PRNU模式。

CNN——利用口型(即视位)与口语音素之间的不匹配。

* + 专注于与M、B和P音素相关的声音，因为它们需要完全的嘴部闭合，而deepfakes经常错误地合成它。

由作者创建的视频，包括10个真实视频和16个使用DeepFaceLab([DeepFaceLab](#_bookmark52),[2022b](#_bookmark52)).

来自Instagram和YouTube的四个野生假唱视频([www.instagram.com/bill\_posters\_uk](http://www.instagram.com/bill_posters_uk)和)以及其他是使用合成技术创建的，

即音频到视频(A2V)([Suwajanakorn et al.](#_bookmark166),[2017](#_bookmark166))和文本转视频(T2V)([Fried et al.](#_bookmark71),[2019](#_bookmark71)).

使用基于归因的可信度(ABC)指标([Fernandes](#_bookmark66)

[et al.](#_bookmark66),[2020](#_bookmark66))

使用外观和行为([Agarwal](#_bookmark19)

[et al.](#_bookmark19),[2020a](#_bookmark19))

ResNet50型号([Heet al.](#_bookmark85),[2016](#_bookmark85)),

经过VGGFace2的预培训([Cao et al.](#_bookmark31),[2018](#_bookmark31))

基于面部和行为特征的规则。

* ABC指标([Jha et al.](#_bookmark98),[2019](#_bookmark98))用于在不访问训练数据的情况下检测deepfake视频。
* 原始视频的ABC值大于0.94，而deepfakes的ABC值较低。

使用ResNet-101([He et al.](#_bookmark85),[2016](#_bookmark85))而静态面部生物测定是使用VGG([Parkhi et al.](#_bookmark147),[2015](#_bookmark147)).

视频VidTIMIT和从COHFACE获得的另外两个原始数据集(<https://www.idiap.ch/dataset/cohface>)和来自YouTube。COHFACE的数据集([Fernandes et al.](#_bookmark67),[2019](#_bookmark67))和YouTube被商业网站用来生成两个deepfake数据集[https://deepfakesweb.com](https://deepfakesweb.com/)而另一个deepfake数据集是DeepfakeTIMIT([Korshunov and Marcel](#_bookmark108),[2018a](#_bookmark108)).

视频世界领导人数据集([Agarwal et al.](#_bookmark21),[2019](#_bookmark21))，FaceForensics++([Rossler et al.](#_bookmark156),[2019](#_bookmark156))，Google/Jigsaw deepfake检测数据集([Dufour and Gully](#_bookmark60),[2019](#_bookmark60))，DFDC([Dolhanskyet al.](#_bookmark57),[2019](#_bookmark57))和Celeb-DF([Li et al.](#_bookmark119),[2020b](#_bookmark119)).

FakeCatcher([Ciftciet al.](#_bookmark46),[2020](#_bookmark46))

CNN提取人像视频中的生物信号，并使用它们作为真实性的隐式描述符，因为它们在deepfakes中没有在空间和时间上得到很好的保存。

视频UADFV([Yang et al.](#_bookmark185),[2019](#_bookmark185))，FaceForensics([Rössler et al.](#_bookmark155),[2018](#_bookmark155))，FaceForensics++([Rossleret al.](#_bookmark156),[2019](#_bookmark156))，名人DF([Li et al.](#_bookmark119),[2020b](#_bookmark119))，以及142个视频的新数据集，独立于生成模型、分辨率、压缩、内容和上下文。

情感音频-视觉情感线索([Mittal](#_bookmark139)

暹罗网络([Chopra et al.](#_bookmark43),[2005](#_bookmark43))

提取人脸和语音的模态和情感嵌入向量用于深度伪造检测。

视频深度限制([Korshunov and Marcel](#_bookmark108),[2018a](#_bookmark108))和DFDC([Dolhansky et al.](#_bookmark57),[2019](#_bookmark57)).

[et al.](#_bookmark139),[2020](#_bookmark139))

(下一页继续)

无法辨别被操纵的内容。这一点需要在当今的法庭上加以考虑，因为大量的数字操作方法([Maras and Alexandrou](#_bookmark132),[2019](#_bookmark132)).

因此，数字媒体取证结果必须被证明是有效和可靠的，然后才能在法庭上使用。这需要仔细记录取证过程的每一步以及如何得出结果。机器学习和人工智能算法可以用来

表2(续)。

方法分类器/技术

处理所用数据集的关键功能

头部姿势([Yanget al.](#_bookmark185),[2019](#_bookmark185))

胶囊-法医([Nguyen et al.](#_bookmark143),[2019](#_bookmark143))

使用面部区域的68个标志提取SVM特征。

* + 使用SVM使用提取的特征进行分类。

胶囊网络-由VGG-19网络提取的潜在特征

([Simonyan and Zisserman](#_bookmark164),[2014](#_bookmark164))被送入胶囊网络进行分类。

* + 动态路由算法([Sabour et al.](#_bookmark159),[2017](#_bookmark159))用于通过多次迭代将三个卷积胶囊的输出路由到两个输出胶囊，一个用于伪图像，另一个用于真实图像。

视频/图像

视频/图像

* UADFV由49个深度假视频和各自的真视频组成。
* 来自DARPA MediFor GAN图像/视频挑战赛的241幅真实图像和252幅深度伪图像。

四个数据集:Idiap研究所重放攻击([Chingovska et al.](#_bookmark38),[2012](#_bookmark38))，deepfake换脸由[Afchar et al.](#_bookmark18)([2018](#_bookmark18))，面部重现FaceForensics([Rössler et al.](#_bookmark155),[2018](#_bookmark155))，以及使用([Rahmouni et al.](#_bookmark152),[2017](#_bookmark152)).

预处理结合深度网络([Xuan et al.](#_bookmark183),[2019](#_bookmark183))

分析卷积轨迹([Guarnera et al.](#_bookmark77),[2020a](#_bookmark77))

DCGAN、WGAN-GP和PGGAN。

KNN、SVM和线性判别分析(LDA)

* 增强深度学习模型的泛化能力以检测GAN生成的图像。
* 去除伪图像的低级特征。
* 迫使深度网络更加关注真假图像的像素级相似性，以提高泛化能力。

利用期望最大化算法提取与GAN基图像深度伪造生成器卷积生成过程相关的局部特征。

图像-真实数据集:CelebA-HQ([Karras et al.](#_bookmark101),[2017](#_bookmark101))，包括高质量的人脸图像

1024 × 1024分辨率。

* + 假数据集:由DCGAN生成([Radfordet al.](#_bookmark151),[2015](#_bookmark151))，WGAN-GP([Gulrajani et al.](#_bookmark81),[2017](#_bookmark81))和PGGAN([Karras et al.](#_bookmark101),[2017](#_bookmark101)).

图片来自CelebA和

相应的deepfakes由五个不同的GANs创建(组级deep

白化和着色变换[Cho et al.](#_bookmark40),[2019](#_bookmark40)斯塔根[Choi et al.](#_bookmark42),[2018](#_bookmark42),

阿特根([He et al.](#_bookmark87),[2019](#_bookmark87))，StyleGAN([Karras](#_bookmark102)

[et al.](#_bookmark102),[2019](#_bookmark102))，StyleGAN2([Karras et al.](#_bookmark103),[2020](#_bookmark103))).

词汇袋和浅层量词([Zhang et al.](#_bookmark191),[2017](#_bookmark191))

SVM，RF，MLP使用单词包提取鉴别特征

方法，并将这些特征输入SVM、射频和MLP进行二元分类:无辜与捏造。

众所周知的LFW人脸数据库([Huanget al.](#_bookmark92),[2007](#_bookmark92))，包含13，223张图像，分辨率为250 × 250。

成对学习([Hsuet al.](#_bookmark90),[2020](#_bookmark90))

防御深伪中的对抗性扰动([Gandhi andJain](#_bookmark73),[2020](#_bookmark73))

美国有线电视新闻网连接到CFFN

VGG([Parkhi et al.](#_bookmark147),[2015](#_bookmark147))和ResNet([Heet al.](#_bookmark85),[2016](#_bookmark85))

两阶段过程:使用基于暹罗网络架构的CFFN的特征提取([Chopraet al.](#_bookmark43),[2005](#_bookmark43))并使用CNN进行分类。

* 引入敌对的扰动来增强deepfakes和愚弄deepfake探测器。
* 使用Lipschitz正则化和深度图像先验技术提高deepfake检测器的准确性。

图片-脸部图片:来自CelebA的真实图片([Liuet al.](#_bookmark126),[2015](#_bookmark126))，以及DCGAN生成的伪([Radford et al.](#_bookmark151),[2015](#_bookmark151))，WGAN

([Arjovsky et al.](#_bookmark26),[2017](#_bookmark26))，WGAN-GP([Gulrajaniet al.](#_bookmark81),[2017](#_bookmark81))，最小二乘GAN([Mao et al.](#_bookmark131),[2017](#_bookmark131))，以及PGGAN([Karras et al.](#_bookmark101),[2017](#_bookmark101)).

-一般图片:来自ILSVRC12的真实图片([Russakovsky et al.](#_bookmark157),[2015](#_bookmark157))，以及BIGGAN生成的假的([Brock et al.](#_bookmark29),[2018](#_bookmark29))，自我关注甘([Zhang et al.](#_bookmark190),[2019](#_bookmark190))和光谱归一化GAN([Miyato et al.](#_bookmark140),[2018](#_bookmark140)).

图片5000来自CelebA的真实图片([Liu et al.](#_bookmark126),[2015](#_bookmark126))和5000个通过“少数镜头面部平移GAN”方法创建的伪图像([Few-Shot Face Translation GAN](#_bookmark68),[2022](#_bookmark68)).

面部x光检查([Li et al.](#_bookmark115),[2020a](#_bookmark115))

CNN -尝试定位目标和原始人脸之间的混合边界，而不是捕捉特定操作的合成伪影。

-可以在没有假图像的情况下进行训练。

图像人脸取证++([Rossler et al.](#_bookmark156),[2019](#_bookmark156))，DeepfakeDetection (DFD)([Dufour and Gully](#_bookmark60),[2019](#_bookmark60))，DFDC([Dolhansky et al.](#_bookmark57),[2019](#_bookmark57))和Celeb-DF([Li et al.](#_bookmark119),[2020b](#_bookmark119)).

使用CNN生成的图像的常见伪像([Wang et al.](#_bookmark182),[2020](#_bookmark182))

ResNet-50([He et al.](#_bookmark85),[2016](#_bookmark85))用ImageNet预先训练([Russakovskyet al.](#_bookmark157),[2015](#_bookmark157))

使用由高性能无条件GAN模型，即PGGAN([Karras et al.](#_bookmark101),[2017](#_bookmark101))并评估分类器推广到其他CNN合成图像的效果。

图像CNN生成的图像的新数据集，即ForenSynths，由来自StyleGAN([Karras et al.](#_bookmark102),[2019](#_bookmark102))，超分辨率方法([Dai et al.](#_bookmark49),[2019](#_bookmark49))和FaceForensics++([Rossleret al.](#_bookmark156),[2019](#_bookmark156)).

在基于GAN的图像上使用卷积轨迹([Guarnera](#_bookmark78)

[et al.](#_bookmark78),[2020b](#_bookmark78))

KNN、SVM和LDA训练期望最大化算法

([Moon](#_bookmark142),[1996](#_bookmark142))通过表示图像生成期间由GANs留下的卷积轨迹的指纹来检测和提取区别特征。

图像由十个GAN模型生成的图像数据集，包括CycleGAN([Zhu et al.](#_bookmark196),[2017](#_bookmark196))，StarGAN([Choi et al.](#_bookmark42),[2018](#_bookmark42))，AttGAN([He et al.](#_bookmark87),[2019](#_bookmark87))，GDWCT([Cho et al.](#_bookmark40),[2019](#_bookmark40))，StyleGAN

([Karras et al.](#_bookmark102),[2019](#_bookmark102))，StyleGAN2([Karras et al.](#_bookmark103),[2020](#_bookmark103))，PGGAN([Karras et al.](#_bookmark101),[2017](#_bookmark101))，FaceForensics++([Rossler et al.](#_bookmark156),[2019](#_bookmark156))，IMLE([Li et al.](#_bookmark120),[2019b](#_bookmark120))，还有黑桃([Park et al.](#_bookmark146),[2019](#_bookmark146)).

使用CNN提取的深度特征([Guoet al.](#_bookmark83),[2021](#_bookmark83))

一个新的CNN模型，即SCnet

基于CNN的SCnet能够自动学习图像数据的高级取证特征，这要归功于分层特征提取块，它是通过堆叠四个卷积层形成的。

图像通过应用发光模型([Kingma andDhariwal](#_bookmark105),[2018](#_bookmark105))到CelebA人脸图像数据集([Liu et al.](#_bookmark126),[2015](#_bookmark126)).

支持确定数字媒体的真实性，并具备[obtained](#_bookmark96) 准确可靠的结果，例如，[Su et al.](#_bookmark165) ([2017](#_bookmark165))和[Iulianiet al.](#_bookmark96)([2018](#_bookmark96))，但这些算法大多无法解释。这

为人工智能在取证问题中的应用创造了巨大的障碍，因为不仅取证专家通常不具备计算机算法方面的专业知识，而且计算机专业人员也不能

正确解释结果，因为这些算法大多是黑盒模型([Malolan et al.](#_bookmark130),[2020](#_bookmark130)).这一点更为关键，因为具有最准确结果的最新模型是基于由许多神经网络参数组成的深度学习方法。研究人员最近试图创造白盒和可解释的检测方法。一个例子是由提出的方法[Giudice et al.](#_bookmark74) ([2021](#_bookmark74))中，他们使用离散余弦变换统计来检测所谓的特定GAN频率，以区分真实图像和深度赝品。通过分析特定的频率统计，该方法可以用于数学地解释多媒体内容是否是深度伪造的以及为什么是。在这个领域必须进行更多的研究，因此计算机视觉中可解释的人工智能是一个研究方向，需要在数字媒体取证中促进和利用人工智能和机器学习的进步和优势。

# 结论

Deepfakes已经开始侵蚀人们对媒体内容的信任，因为看到它们不再等同于相信它们。它们可能对目标造成困扰和负面影响，加剧虚假信息和仇恨言论，甚至可能刺激政治紧张局势，煽动公众，引发暴力或战争。这在当今尤为重要，因为制造深度假货的技术越来越容易获得，社交媒体平台可以快速传播这些假货内容。这项调查提供了deepfake创建和检测方法的及时概述，并对该领域的挑战、潜在趋势和未来方向进行了广泛的讨论。因此，这项研究对于人工智能研究社区开发解决deepfakes的有效方法将是有价值的。

# 著作权贡献声明

陈氏阮:概念化，方法论，调查，写作。郭越雄阮:概念化，写作-审查和编辑。阮晋勇:方法论，写作-初稿。阮德清:可视化，写作-原始草案。Thien Huynh-The:验证、写作-审核和编辑。Saeid Na- havandi:验证、写作-审核和编辑。Thanh Tam Nguyen:可视化，写作-评论和编辑。Quoc-Viet Pham:验证，写作-审查和编辑。阮孟明:调查，写作

–原始草案。

# 竞争利益声明

作者声明，他们没有已知的竞争财务利益或个人关系，这可能会影响本文报道的工作。

# 参考

[Afchar,大流士，诺齐克，文森特，山木，俊一，越前伊佐夫，2018。MesoNet:一个紧凑的面部视频伪造检测网络。In: 2018 IEEE国际信息取证和安全研讨会，WIFS。IEEE，第1–7页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb1)

[Agarwal,Shruti，Farid，Hany，El-Gaaly，Tarek，Lim，Ser-Nam，2020a。探测深层-从外表和行为上造假视频。IEEE国际研讨会信息取证和安全，WIFS。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb2)

Agarwal，Shruti，Farid，Hany，Fried，Ohad，Agrawala，Maneesh，2020b。从音位-视位不匹配检测深度伪造视频。IEEE/CVF计算机视觉和模式识别研讨会会议录。第660-661页。[Agarwal,Shruti、Farid、Hany、顾、、何、明明、长野、Koki、李、郝、](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb4)

[2019.保护世界领导人免受深度造假。在:计算机视觉和模式认可研讨会，第1卷，第38-45页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb4)

Agarwal，Sakshi，Varshney，Lav R，2019。深度伪造检测的极限:一个稳健的估计观点。arXiv预印本[arXiv:1905.03493](http://arxiv.org/abs/1905.03493).

阿梅里尼，艾琳，卡尔德利，罗伯托，2020。通过基于LSTM的分类器利用预测误差不一致性来检测deepfake视频。2020年ACM信息隐藏和多媒体安全研讨会会议录。第97-102页。

阿梅里尼，艾琳，加尔特里，莱昂纳多，卡尔德利，罗伯托，德尔宾博，阿尔贝托，2019。基于CNN的光流视频深度防伪检测。IEEE/CVF国际计算机视觉研讨会会议录。第1205-1207页。

[Arjovsky,马丁，钦塔拉，苏史密斯，博图，莱昂，2017。瓦瑟斯坦生成敌对网络。国际机器学习会议。PMLR，第214-223页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb8)

[Bai,双，2017。在深度卷积神经网络上生长随机森林场景分类。专家系统。应用程序71，279–287。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb9)

贝亚尔，贝尔哈森，斯塔姆，马修c，2016。使用新卷积层的通用图像操作检测的深度学习方法。在:会议录

第四届ACM信息隐藏和多媒体安全研讨会。第5-10页。[彭博，](https://fortune.com/2018/09/11/deep-fakes-obama-video/)2018.伪造视频是如何变得容易的，为什么这么可怕。[https:](https://fortune.com/2018/09/11/deep-fakes-obama-video/)

[//fortune.com/2018/09/11/deep-fakes-obama-video/](https://fortune.com/2018/09/11/deep-fakes-obama-video/).

布洛克，安德鲁，多纳休，杰夫，西蒙扬，凯伦，2018。用于高保真自然图像合成的大规模GAN训练。arXiv预印本[arXiv:1809.11096](http://arxiv.org/abs/1809.11096).

[Caldelli,罗伯托，加尔特里，莱昂纳多，阿梅里尼，艾琳，德尔宾博，阿尔贝托，2021。视觉的基于流的细胞神经网络检测遗忘深度篡改。模式识别。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb13)

[列特。146, 31–37.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb13)

[Cao,琼，沈，李，谢，，帕克希，奥姆卡尔M，齐塞曼，安德鲁，2018。VGGFace2:一个用于跨姿势和年龄识别人脸的数据集。在:第13届IEEE2018年FG自动人脸和手势识别国际会议。IEEE，第67-74页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb14)

卡利尼，尼古拉斯，法里德，汉尼，2020。用白盒和黑盒攻击躲避deep假图像检测器。载于:IEEE/CVF计算机视觉会议论文集

和模式识别研讨会。第658-659页。

Chan，Caroline，Ginosar，Shiry，周，Tinghui，Efros，Alexei A，2019。大家现在跳舞。IEEE/CVF计算机视觉国际会议论文集。第5933-5942页。

[Cheng,、孙、、竹内、正、加藤、二郎，2019。活力](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb17)

[基于压缩的卷积自动编码器图像压缩。IEEE Trans。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb17)[多媒体。22 (4), 860–873.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb17)

[Chesney,罗伯特，香橼，丹妮尔·济慈，2018a。深度造假:一个迫在眉睫的挑战隐私、民主和国家安全。Democr。国家安全局。107.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb18)

2018年b d . k . Citron，Chesney。兴奋剂上的虚假信息:深度造假的威胁。

<https://www.cfr.org/report/deep-fake-disinformation-steroids>.

[Chesney,罗伯特，香橼，丹妮尔，2019。Deepfakes和新的假情报战争；后真相地缘政治时代的到来。外国Aff。98, 147.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb20)

[钦科夫斯卡，伊万娜，安茹斯，安德烈，马塞尔，塞巴斯蒂安，2012。论的有效性人脸反欺骗中的局部二值模式。载于:国际会议录](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb21)

[生物统计学特别兴趣小组会议。IEEE，第1–7页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb21)

[Chintha,阿卡什语、泰语、鲍语、索拉瓦迪语、萨尼亚特·贾维德语、巴特语、卡尔塔维亚语、希克森语、安-德雷阿，赖特，马修，普图查，雷蒙德，2020。递归卷积结构用于音频欺骗和视频deepfake检测。IEEEJ。Sel。顶端。签名。过程。14(5), 1024–1037.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb22)

Cho，Wonwoong，Choi，Sungha，Park，David Keetae，Shin，Inkyu，Choo，Jaegul，2019。通过逐组深度白化和着色变换进行图像到图像的转换。《IEEE/CVF计算机视觉和模式会议论文集》

认可。第10639-10647页。

Cho、Kyunghyun、Van merrinboer、Bart、Gulcehre、卡格拉尔、Bahdanau、Dzmitry、Bougares、Fethi、Schwenk、Holger、Bengio、Yoshua，2014年。使用用于统计机器翻译的RNN编码器-解码器学习短语表示。arXiv预印本[arXiv:1406.1078](http://arxiv.org/abs/1406.1078).

Choi，Yunjey，Choi，Minje，Kim，Munyoung，Ha，Jung-Woo，Kim，Sunghun，Choo，Jaegul，2018。StarGAN:用于多领域图像到图像翻译的统一生成对抗网络。美国电气与电子工程师协会会议录

计算机视觉和模式识别。第8789-8797页。

[Chopra,Sumit，Hadsell，Raia，LeCun，Yann，2005年。学习相似性度量dis刑事上，应用于人脸验证。在:IEEE计算机学会计算机视觉和模式识别会议，CVPR 05，第1卷。IEEE，第539-546页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb26)

[乔洛夫斯基，Jan，Weiss，Ron J，Bengio，Samy，van den Oord，a Ron，2019。无人监督的](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb27)

[使用WaveNet自动编码器的语音表示学习。IEEE/美国计算机学会会刊。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb27)[有声语言。过程。27 (12), 2041–2053.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb27)

[Chung,Joon Son，Senior，Andrew，Vinyals，Oriol，Zisserman，Andrew，2017。唇读野外的句子。In: 2017年IEEE计算机视觉与模式会议认可，CVPR。IEEE，第3444–3453页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb28)

[Ciftci,Umur Aybars，，Ilke，Yin，Lijun，2020。FakeCatcher:检测合成](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb29)

[使用生物信号的肖像视频。IEEE Trans。肛门模式。马赫。知识..](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb29)

科佐利诺、达维德、提斯、胡斯图斯、罗斯勒、安德烈亚斯、里斯、克里斯蒂安、涅纳、马蒂亚斯、维尔多里瓦、路易莎、2018。ForensicTransfer:用于伪造检测的弱监督域适应。arXiv预印本[arXiv:1812.02510](http://arxiv.org/abs/1812.02510).

2022年的CycleGAN。<https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>.

戴，陶，蔡，张剑瑞，张永兵，夏，舒涛，张，雷，2019。单幅图像超分辨率的二阶注意网络。《IEEE/CVF计算机视觉和模式识别会议论文集》。第11065-11074页。

新泽西州Damiani，2019。一个声音deepfake被用来诈骗一个CEO 243，000美元。[https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-](https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/)

[used-to-scam-a-ceo-out-of-243000/](https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/).

[DeepFaceLab，](https://mrdeepfakes.com/forums/thread-deepfacelab-explained-and-usage-tutorial)2022a。DeepFaceLab:解释和使用教程。<https://mrdeepfakes.com/forums/thread-deepfacelab-explained-and-usage-tutorial>.

DeepFaceLab，2022b。<https://github.com/iperov/DeepFaceLab>.

[DeepFake\_tf，](https://github.com/StromWine/DeepFake_tf)2022.DeepFake\_tf:基于tensorflow的DeepFake。<https://github.com/StromWine/DeepFake_tf>.

邓，于，杨，娇龙，陈，董，文，方，童，辛，2020。基于3D模拟对比学习的非纠缠可控人脸图像生成。《IEEE/CVF计算机视觉和模式识别会议论文集》。第5154-5163页。

DFaker，2022。<https://github.com/dfaker/df>.

多尔汉斯基，布赖恩，比顿，乔安娜，普弗劳姆，本，陆，季阔，，拉斯，王，孟林，坎顿费勒，克里斯蒂安，2020。deepfake检测挑战数据集。arXiv预印本[arXiv:2006.07397](http://arxiv.org/abs/2006.07397).

多尔汉斯基，布莱恩，霍维斯，拉斯，普弗洛姆，本，巴拉姆，妮可，费雷尔，克里斯蒂安坎顿，2019。deepfake检测挑战赛(DFDC)预览数据集。arXiv预印本

[arXiv:1910.08854](http://arxiv.org/abs/1910.08854).

多纳休，杰弗里，安妮·亨德里克斯，丽莎，瓜达拉马，塞尔吉奥，罗尔巴克，马库斯，韦努戈帕兰，苏巴希尼，萨恩科，凯特，达雷尔，特雷弗，2015。用于视觉识别和描述的长期递归卷积网络。IEEE计算机视觉和模式识别会议论文集。第2625-2634页。

[DSSIM,](https://github.com/keras-team/keras-contrib/blob/master/keras_contrib/losses/dssim.py)2022.[https://github.com/keras-team/keras-contrib/blob/master/keras\_contrib/](https://github.com/keras-team/keras-contrib/blob/master/keras_contrib/losses/dssim.py)

[losses/dssim.py](https://github.com/keras-team/keras-contrib/blob/master/keras_contrib/losses/dssim.py).

杜福尔，尼克，古利，安德鲁，2019。为deepfake检测研究贡献数据。<https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html>.

面网，2022。<https://github.com/davidsandberg/facenet>.

[Faceswap,](https://github.com/deepfakes/faceswap)2022.面向所有人的Deepfakes软件。<https://github.com/deepfakes/faceswap>.

Faceswap-GAN，2022。<https://github.com/shaoanlu/faceswap-GAN>.

FakeApp，2022。FakeApp 2.2.0。<https://www.malavida.com/en/soft/fakeapp/>.

[法里德，哈尼，2009。图像伪造检测。IEEE信号处理。玛格。26 (2), 16–25.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb48)费尔南德斯、史蒂文、拉吉、桑尼、埃维兹、里卡德、潘努、焦德·辛格、贾哈、苏米特·Ku-马尔、奥尔蒂斯、艾迪、温蒂拉、尤斯蒂娜、索尔特、玛格丽特、2020。使用基于属性的置信度度量检测深度伪造视频。IEEE/CVF计算机视觉和模式识别研讨会会议录。第308-309页。

费尔南德斯、史蒂文、拉吉、桑尼、奥尔蒂斯、艾迪、温蒂拉、尤斯蒂娜、萨尔特、玛格丽特、乌罗斯维奇、戈尔达纳、杰哈、苏米特、2019。使用神经ODE预测深度假视频的心率变化。IEEE/CVF国际会议论文集

计算机视觉研讨会。第1721-1729页。

[Few-Shot](https://github.com/shaoanlu/fewshot-face-translation-GAN)脸翻译甘，2022。<https://github.com/shaoanlu/fewshot-face-translation-GAN>.

鱼，t，2019。深度造假:人工智能操纵的媒体将被“武器化”来欺骗军方。[https://www.express.co.uk/news/science/1109783/deep-fakes-ai-](https://www.express.co.uk/news/science/1109783/deep-fakes-ai-artificial-intelligence-photos-video-weaponised-china)

[artificial-intelligence-photos-video-weaponised-china](https://www.express.co.uk/news/science/1109783/deep-fakes-ai-artificial-intelligence-photos-video-weaponised-china).

[Floridi,卢西亚诺，2018。人工智能、deepfakes和ectypes的未来。菲洛斯。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb53)

[技术。31 (3), 317–321.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb53)

[Fried,奥哈德、特瓦里、阿尤什、佐尔霍夫、迈克尔、芬克斯坦、亚当、谢克特曼、伊莱、高曼，丹B，热那亚，凯尔，金，泽宇，西奥博尔特，克里斯蒂安，阿格拉瓦拉，马涅什，2019.基于文本的发言者视频编辑。ACM Trans图表。38 (4), 1–14.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb54)

[Galbally,哈维尔，马塞尔，塞巴斯蒂安，2014。基于普通图像的人脸反欺骗质量评估。第22届国际模式识别会议。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb55)

[IEEE，第1173–1178页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb55)

[Gandhi,Apurva，Jain，Shomik，2020。对抗性干扰愚弄深度假探测器。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb56)[IEEE国际神经网络联合会议。IEEE，第1–8页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb56)

朱迪切，奥立弗，瓜内拉，卢卡，巴蒂亚托，塞巴斯蒂安，2021。对抗深度假货

检测GAN DCT异常。arXiv预印本[arXiv:2101.09781](http://arxiv.org/abs/2101.09781).

[好伙计，伊恩、普盖-阿巴迪、让、米尔扎、迈赫迪、徐、冰、沃德-法利、大卫、奥泽尔、谢尔吉尔、库维尔、亚伦、本吉奥、约舒厄，2014年。生成对抗网络。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb58)[神经Inf。过程。系统。27, 2672–2680.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb58)

[Guan,迎海，科萨克，马克，罗伯逊，埃里克，李，尤扬，耶茨，艾米N，德尔-加多、安德鲁、周、丹尼尔、凯哈、蒂莫西、史密斯、杰夫、菲库斯、乔纳森、2019.MFC数据集:媒体取证挑战的大规模基准数据集](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb59) [评价。IEEE计算机视觉冬季应用研讨会。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb59)

[IEEE，第63–72页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb59)

瓜内拉，卢卡，朱迪切，奥利佛，巴蒂亚托，塞巴斯蒂安，2020a。通过分析卷积轨迹进行深度伪造检测。IEEE/CVF计算机视觉和模式识别研讨会会议录。第666-667页。

[Guarnera,卢卡，朱迪切，奥利弗，巴蒂亚托，塞巴斯蒂安，2020b。对抗深度伪造](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb61)

[曝光图像上的卷积轨迹。IEEE访问8，165085–165098。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb61)

[Guarnera,卢卡，朱迪切，奥利佛，纳斯塔西，克里斯蒂娜，巴蒂亚托，塞巴斯蒂安，2020c。表示“在… 之前”: pre-adolescent ｜ precaution ｜ precede深度伪造图像的初步取证分析。年:AEIT国际年刊会议，AEIT。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb62)

[Güera,大卫，德尔普，爱德华j，2018。基于递归神经网络的深度伪造视频检测网络。载于:第15届IEEE高级视频和信号国际会议](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb63)

[基于监视，AVSS。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb63)

古尔拉贾尼，伊桑，艾哈迈德，法鲁克，阿尔乔夫斯基，马丁，杜穆林，文森特，库维尔，亚伦，2017。改进了瓦瑟斯坦·甘斯的训练。arXiv预印本[arXiv:1704.00028](http://arxiv.org/abs/1704.00028).

[Guo,斌，丁，亚三，姚，林娜，梁，，于，，2020。虚假的未来社交媒体上的信息探测:新观点和趋势。ACM计算机。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb65)

[Surv。53 (4), 1–36.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb65)

[Guo,支青，胡，，夏，明，杨，，2021。基于辉光的盲检测面部伪造。多媒体工具应用程序80 (5)，7687–7710。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb66)

[Ha,宋珠，克斯纳，马丁，金，贝木苏，徐，锡春，金，东英，2020。牵线木偶:保留看不见的目标的身份的少数镜头脸部重现。在:AAAI人工智能会议论文集，第34卷，(07)，页10893–10900.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb67)

[Hasan,哈亚河，萨拉赫，哈立德，2019。打击deepfake视频使用区块链和智能合同。IEEE接入7，41596–41606。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb68)

何，，张，向宇，任，，孙，简，2016。用于图像识别的深度残差学习。IEEE计算机视觉和模式识别会议论文集。第770-778页。

[He,、左、、阚、美娜、山、、陈、西林，2019。AttGAN:只改变你想要的面部属性编辑。IEEE Trans。图像处理。28 (11), 5464–5478.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb70)

[Hinton,Geoffrey E .，Krizhevsky，Alex，Wang，Sida D .，2011年。转变自动-编码器。国际人工神经网络会议。施普林格，pp44–51.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb71)

[Hsu,志忠，李，佳彦，庄，一休，2018。学习识别假脸野外的图像。在:2018年国际计算机研讨会，消费者和控制，是3C。IEEE，第388–391页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb72)

[Hsu,志忠，庄，一休，李佳彦，2020。深度伪图像检测基于成对学习。应用科学。10 (1), 370.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb73)

黄，高，刘，庄，范德马腾，劳伦斯，温伯格，基连Q，2017。密集连接的卷积网络。IEEE计算机视觉和模式识别会议论文集。第4700-4708页。

[Huang,加里b，拉梅什，马努，伯格，塔玛拉，博学米勒，埃里克，2007年。有标签的野外人脸:用于研究无约束条件下人脸识别的数据库环境。技术报告07–49，马萨诸塞大学，阿默斯特。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb75)

侯赛因、谢兹恩、尼卡拉、帕尔思、杰雷、马尔哈尔、库尚法尔、法里纳兹、麦考利、朱里安，2021。对抗的deepfakes:评估deepfake检测器对对抗例子的脆弱性。IEEE/CVF计算机视觉应用冬季会议论文集。第3348-3357页。

[Hwang,T., 2020.Deepfakes:基于威胁的评估。技术报告，中心乔治城大学安全和新兴技术。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb77)

IPFS，2022年。IPFS为分布式网络提供动力。<https://ipfs.io/>.

[Iuliani,马西莫、舒拉尼、达萨拉、丰塔尼、马尔科、梅乌奇、萨维里奥、皮瓦、亚历山德罗、2018.一个用于MP4类文件无监督分析的视频取证框架集装箱。IEEE Trans。Inf。法医安全。14 (3), 635–645.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb79)

[Jafar,穆萨·塔伊塞尔、阿巴布内、穆罕默德、祖比、穆罕默德、哈桑、阿马尔、2020.deepfake视频的取证与分析。在:第十一届国际信息与通信系统会议。IEEE，第053–058页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb80)

[Jha,苏斯米特、拉杰、桑尼、费尔南德斯、史蒂文、贾哈、苏米特K、贾哈、萨默什、贾莱安、布莱恩、维尔马，Gunjan，Swami，Ananthram，2019。基于归因的置信度度量深度神经网络。神经Inf。过程。系统。32, 11826–11837.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb81)

姜，李明，李，任，吴，韦恩，钱，陈，阿来，陈变，2020。DeeperForensics-1.0:用于真实世界人脸伪造检测的大规模数据集。《IEEE/CVF计算机视觉和模式识别会议论文集》。第2889-2898页。

[Kaliyar,Rohit Kumar，Goswami，Anurag，Narang，Pratik，2021。DeepFakE:改进基于张量分解的深度神经网络假新闻检测。J.超级切割。77 (2), 1015–1037.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb83)

卡拉斯，泰罗，艾拉，提莫，莱内，萨穆利，莱蒂宁，亚科，2017。为了提高质量、稳定性和多样性而逐步种植甘蔗。arXiv预印本[arXiv:1710年。10196](http://arxiv.org/abs/1710.10196).

Karras，Tero，Laine，Samuli，Aila，Timo，2019。一种基于风格的生成对抗网络生成器体系结构。《IEEE/CVF计算机视觉和模式识别会议论文集》。第4401–4410页。

Karras，Tero，Laine，Samuli，Aittala，Miika，Hellsten，Janne，Lehtinen，Jaakko，Aila，Timo，2020。stylegan图像质量的分析与改进。载于:IEEE/CVF计算机视觉和模式识别会议论文集。第8110-8119页。

Keras-VGGFace，2022。Keras-VGGFace:用Keras框架实现VGGFace。<https://github.com/rcmalli/keras-vggface>.

Kingma，Diederik P .，Dhariwal，Prafulla，2018年。辉光:具有可逆1 × 1卷积的生成流。收录于:第32届国际神经信息处理系统会议录。第10236–10245页。

金玛，迪德里克p，韦林，马克斯，2013年。自动编码变分贝叶斯。arXiv预印本[arXiv:1312.6114](http://arxiv.org/abs/1312.6114).

库普曼，玛丽莎，罗德里格斯，安德里亚·马卡鲁拉，杰拉德茨，芝诺，2018。deepfake视频操作的检测。第20届爱尔兰机器视觉和图像处理会议。第133-136页。

科尔舒诺夫，帕维尔，马塞尔，塞巴斯蒂安，2018a。Deepfakes:人脸识别的新威胁？评估和检测。arXiv预印本[arXiv:1812.08685](http://arxiv.org/abs/1812.08685).

[科尔舒诺夫，帕维尔，马塞尔，塞巴斯蒂安，2018年b。tam-中说话人不一致性检测pered视频。在:第26届欧洲信号处理会议。IEEE，第2375-2379页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb92)

[科尔舒诺夫，帕维尔，马塞尔，塞巴斯蒂安，2019。漏洞评估和检测deepfake视频。2019年ICB国际生物识别大会。IEEE，pp。1–6.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb93)

科尔舒诺娃，伊琳娜，史，，丹布雷，乔尼，泰斯，卢卡斯，2017。基于卷积神经网络的快速人脸交换。IEEE计算机视觉国际会议论文集。第3677-3685页。

[Laptev, 伊万，马斯扎克，马钦，施密德，科迪莉亚，罗森菲德，本杰明，2008。学问电影中真实的人类行为。在:IEEE计算机视觉和模式识别。IEEE，第1–8页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb95)

拉塔斯、亚历山卓斯、莫斯乔格卢、斯蒂利亚诺斯、格切尔、巴里斯、普卢皮斯、斯蒂利亚诺斯、特里安塔夫-卢、瓦西里奥斯、戈什、阿比吉特、扎费里欧、斯特凡诺斯、2020。AvatarMe:现实可渲染的3D面部重建“在野外”。《IEEE/CVF计算机视觉和模式识别会议论文集》。第760-769页。

李，凌志，包，建民，杨，郝，陈，董，文，方，2019a。面向高保真和遮挡感知的人脸交换。arXiv预印本[arXiv:1912.13457](http://arxiv.org/abs/1912.13457).

李，凌志，鲍，建民，张，婷，杨，郝，陈，董，文，方，郭，吴建民，2020a。人脸x光更一般的人脸伪造检测。《IEEE/CVF计算机视觉和模式识别会议论文集》。第5001–5010页。

[Li,月尊，常，明清，吕，四维，2018。在ictu oculi:揭露人工智能创造通过检测眨眼来伪造视频。2018年IEEE国际研讨会信息取证和安全，WIFS。IEEE，第1–7页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb99)

[Li,张村，李，岳，2011。颜色分离光响应非均匀性形成数字图像取证。IEEE Trans。电路系统。视频技术。22 (2), 260–271.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb100)

李，岳尊，吕，四维，2018。通过检测面部扭曲伪影来曝光deepfake视频。arXiv预印本[arXiv:1811.00656](http://arxiv.org/abs/1811.00656).

李，岳尊，杨，辛，孙，濮，齐，洪刚，吕，四维，2020b。Celeb-DF:用于deepfake取证的大规模挑战数据集。《IEEE/CVF计算机视觉和模式识别会议论文集》。第3207-3216页。

李、柯、张、天昊、马利克、吉滕德拉、2019b。通过条件IMLE从语义布局合成多样的图像。IEEE/CVF计算机视觉国际会议论文集。第4220-4229页。

德利马，奥斯卡，富兰克林，肖恩，巴苏，什雷什塔，卡尔沃斯基，布莱克，乔治，安内特，2020。基于时空卷积网络的深度伪造检测。arXiv预印本[arXiv:2006.14749](http://arxiv.org/abs/2006.14749).

[Lin,徐峰，李，张村，2016。基于摄像机的大规模图像聚类指纹。IEEE Trans。Inf。法医安全。12 (4), 793–808.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb105)

[Lin, 贾成，李，杨，杨，关慈，2021。FPGAN:人脸识别方法社交机器人的生成对抗网络。神经网络。133, 132–147.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb106)

刘，明宇，黄，荀，Mallya，Arun，Karras，Tero，Aila，Timo，Lehtinen，Jaakko，Kautz，2019年1月。少镜头无监督图像到图像翻译。IEEE/CVF计算机视觉国际会议论文集。第10551-10560页。

[Liu,明宇，黄，荀，余，佳惠，王，丁纯，Mallya，Arun，2021。图像和视频合成的生成对抗网络:算法和应用程序。继续。IEEE 109 (5)，839–862。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb108)

刘，，罗，平，王，肖刚，唐，肖鸥，2015。野外深度学习人脸属性。IEEE计算机视觉国际会议论文集。第3730-3738页。

[Lyu,](http://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072)四维，2018。眨眼间就能检测出“deepfake”视频。<http://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072>.

[Lyu,四维，2020。Deepfake检测:当前的挑战和下一步。在:IEEE国际多媒体及展览研讨会。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb111)

马克扎尼，阿里雷扎，施伦斯，黄邦贤，贾伊特利，纳夫迪，古德菲勒，伊恩，弗雷，布伦丹，2015。敌对的自动编码器。arXiv预印本[arXiv:1511.05644](http://arxiv.org/abs/1511.05644).

[Malolan,Badhrinarayan，Parekh，Ankit，Kazi，Faruk，2020。可解释的深度伪造使用目视判读方法进行检测。第三届国际会议信息和计算机技术。IEEE，第289–293页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb113)

[Mao,徐东、李、青、谢、浩然、刘、、王、甄、斯莫-莱伊，斯蒂芬，2017。最小二乘生成对抗网络。在:诉讼中IEEE计算机视觉国际会议。第2794-2802页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb114)

[Maras,玛丽-海伦，亚历山德鲁，亚历克斯，2019。确定视频视频的真实性人工智能时代的信任和深度假视频的觉醒。里面的](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb115)

[J.证据证明23 (3)，255–262。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb115)

马尔，b，2019。人工智能deepfakes的最佳(也是最可怕的)例子。<https://www.forbes.com/sites/bernardmarr/2019/07/22/the-best-and-scariest-examples-of-ai-enabled-deepfakes/>.

[Marra,弗朗西斯科、格拉尼耶洛、迭戈、科佐利诺、达维德、维尔多里瓦、路易莎，2018。社交网络中GAN生成的虚假图像检测。在:2018年IEEEMIPR多媒体信息处理和检索会议。IEEE，pp。384–389.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb117)

[Marra,弗朗西斯科、萨尔托里、克里斯蒂亚诺、博阿托、朱利亚、维尔多里瓦、路易莎，2019。增加的gan生成图像的检测和分类学习。在:2019年IEEE信息取证与安全国际研讨会，WIFS。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb118)

[Matern,法尔科，里斯，克里斯蒂安，斯塔明格，马克，2019。利用视觉假象揭露deepfakes和面部操纵。计算机的IEEE冬季应用视觉工作坊。IEEE，第83–92页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb119)

[Maurer,尤里·m .，2000年。认证理论和假设检验。IEEE Trans。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb120)

[告知。理论46 (4)，1350-1356。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb120)

[Mirsky,伊斯罗尔，李，柯文，2021。deepfakes的创建和检测:综述。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb121)

[ACM计算机。Surv。54 (1), 1–41.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb121)

米塔尔，特里莎，巴塔查里亚，乌塔兰，钱德拉，罗汉，贝拉，阿尼基特，马诺查，迪内什，2020。情感不会说谎:一种使用视听情感线索的深度伪造检测方法。3，arXiv预印本[arXiv:2003.06711](http://arxiv.org/abs/2003.06711).

宫藤，泰克，片冈，俊树，小山，正德，吉田，雄一，2018。生成对抗网络的谱归一化。arXiv预印本[arXiv:1802年。05957](http://arxiv.org/abs/1802.05957).

莫，，陈，柏林，罗，魏琦，2018。基于卷积神经网络的假脸识别。第六届ACM信息隐藏和多媒体安全研讨会论文集。第43-47页。

[Moon,托德k，1996年。期望值最大化算法。IEEE信号处理。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb125)

[玛格。13 (6), 47–60.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb125)

[Nguyen,Huy H .，Yamagishi，Junichi，越前，Isao，2019。胶囊取证:使用cap-苏乐网络检测伪造的图像和视频。在:IEEE国际会议声学、语音和信号处理。IEEE，第2307–2311页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb126)

尼尔金，尤瓦尔，凯勒，约西，哈斯纳，塔尔，2019。FSGAN:主题不可知的脸交换和重演。IEEE/CVF计算机视觉国际会议论文集。第7184-7193页。

Olszewski，Kyle，Tulyakov，Sergey，Woodford，Oliver，Li，Hao，Luo，，2019。可转换的瓶颈网络。IEEE/CVF计算机视觉国际会议论文集。第7648-7657页。

Park，Taesung，Liu，Ming-Yu，Wang，Ting-Chun，Zhu，Jun-Yan，2019。空间自适应归一化的语义图像合成。《IEEE/CVF计算机视觉和模式识别会议论文集》。第2337-2346页。

帕克希，奥姆卡尔m，韦达尔迪，安德烈亚，齐塞尔曼，安德鲁，2015。深度人脸识别。《英国机器视觉会议论文集》，BMVC。第41.1至41.12页。

[Phan,Quoc-Tin，Boato，Giulia，De Natale，Francesco G.B .，2018。精确且可扩展基于相机指纹稀疏表示的图像聚类。IEEE Trans。Inf。法医安全。14 (7), 1902–1916.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb131)

[彭那普拉思，Abhijith，Brown，Michael S .，2019。学习原始图像重建-感知深度图像压缩器。IEEE Trans。肛门模式。马赫。智能。42 (4),1013–1019.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb132)

[Qian,银隆，董，景，王，魏，谭，铁牛，2015。steganal的深度学习通过卷积神经网络进行分析。在:媒体水印、安全性和法医学，第9409卷，第94090页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb133)

拉德福德，亚历克，梅斯，卢克，钦塔拉，苏密特，2015。深度卷积生成对抗网络的无监督表示学习。arXiv预印本[arXiv:1511.06434](http://arxiv.org/abs/1511.06434).

[Rahmouni,尼古拉斯，诺齐克，文森特，山木，俊一，越前伊佐，2017。区分-用卷积神经网络从自然图像中引导计算机图形。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb135)[载于:IEEE信息取证和安全研讨会，WIFS。IEEE，第1–6页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb135)

[Read,](http://nymag.com/intelligencer/2019/06/how-do-you-spot-a-deepfake-it-might-not-matter.html)米（meter的缩写））, 2019.你能认出假的吗？有关系吗？<http://nymag.com/intelligencer/2019/06/how-do-you-spot-a-deepfake-it-might-not-matter.html>.

[罗森菲尔德，库尔特，森卡尔，胡斯列夫塔哈，2009年。基于PRNU的鲁棒性研究相机识别。载于:媒体取证和安全，第7254卷，第72540页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb137)

罗斯勒、安德烈亚斯、科佐利诺、达维德、维尔多里瓦、路易莎、里斯、克里斯蒂安、提斯、胡斯图斯、涅纳、马蒂亚斯、2018。FaceForensics:用于人脸伪造检测的大规模视频数据集。arXiv预印本[arXiv:1803.09179](http://arxiv.org/abs/1803.09179).

罗斯勒、安德烈亚斯、科佐利诺、达维德、维尔多里瓦、路易莎、里斯、克里斯蒂安、提斯、胡斯图斯、涅纳、马蒂亚斯，2019。学习检测被操纵的面部图像。IEEE/CVF计算机视觉国际会议论文集。第1-11页。

[鲁萨科夫斯基，奥尔加、邓、贾、苏、郝、克劳斯、乔纳森、萨特西、桑吉夫、马、肖恩、黄、智恒、卡帕西、安德烈、科斯拉、阿迪蒂亚、伯恩斯坦、迈克尔等，2015。ImageNet大规模视觉识别挑战。里面的J. Comput。Vis。115 (3),211–252.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb140)

[Sabir,埃克拉姆、程、加辛、贾斯瓦尔、阿尤什、阿卜德勒马吉德、瓦勒、马西、亚科布、纳塔拉詹，Prem，2019。人脸操作的递归卷积策略视频中的检测。载于:IEEE计算机视觉会议论文集和模式识别研讨会，第3卷，(1)，第80-87页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb141)

萨布尔，萨拉，弗罗斯特，尼古拉斯，辛顿，杰弗里e，2017。cap- sules之间的动态路由。《第31届神经信息处理系统国际会议论文集》。第3859-3869页。

塞缪尔s，2019。一个家伙做了一个deepfake应用程序，把女人的照片变成裸体。不太顺利。<https://www.vox.com/2019/6/27/18761639/ai-deepfake-deepnude-app-nude-women-porn>.

[Scherhag,乌尔里希，德比亚西，卢卡，拉斯盖布，克里斯蒂安，布施，克里斯托夫，乌赫尔，安德烈亚斯，2019.基于PRNU分析的人脸变形攻击检测。IEEE Trans。生物群落。贝哈夫。身份科学。1 (4), 302–317.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb144)

[施若普费尔，](https://ai.facebook.com/blog/deepfake-detection-challenge)米（meter的缩写））, 2019.创建数据集和对deepfakes的挑战。<https://ai.facebook.com/blog/deepfake-detection-challenge>.

施罗夫，弗洛里安，卡列尼琴科，德米特里，菲尔宾，詹姆斯，2015。FaceNet:用于人脸识别和聚类的统一em- bedding。IEEE计算机视觉和模式识别会议论文集。第815-823页。

Simonyan，Karen，Zisserman，Andrew，2014。用于大规模图像识别的非常深的卷积网络。arXiv预印本[arXiv:1409.1556](http://arxiv.org/abs/1409.1556).

[Su,、李、崔华、赖、岳聪、杨、2017。一种快速的伪造检测基于指数傅里叶矩的视频区域复制算法。电气电子工程师学会反式。多媒体。20 (4), 825–840.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb148)

[Suwajanakorn，苏帕索恩，塞茨，史蒂文M，凯梅尔马赫-什利泽曼，爱尔兰共和军，2017。表示“联合的”从音频中学习对口型。ACM Trans图表。36 (4),1–13.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb149)

特瓦里、阿尤什、埃尔加里布、穆罕默德、巴拉杰、高拉夫、伯纳德、弗洛里安、塞德尔、汉斯-彼得、佩雷斯、帕特里克、佐尔霍费尔、迈克尔、特奥巴尔特、克里斯蒂安、2020。StyleRig:操纵StyleGAN对肖像图像进行3D控制。《IEEE/CVF计算机视觉和模式识别会议论文集》。第6142-6151页。

[Tewari,阿尤什、佐尔霍弗、迈克尔、伯纳德、弗洛里安、加里多、帕布罗、金、铉佑、佩雷斯，帕特里克，西奥多，克里斯蒂安，2018。高保真单目人脸重建基于无监督的基于模型的人脸自动编码器。IEEE Trans。模式肛门。马赫。智能。42 (2), 357–370.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb151)

《卫报》，2019年。中国deepfake应用程序Zao在病毒式传播后引发隐私争议。<https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zao-triggers-privacy-fears-viral>.

[Thies,胡斯图斯、埃尔加里布、穆罕默德、特瓦里、阿尤什、特奥巴尔特、克里斯蒂安、马蒂亚斯，2020年。神经语音木偶:音频驱动的面部重演。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb153)[年:欧洲计算机视觉会议。施普林格，第716-731页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb153)

[Thies,胡斯图斯，佐勒费尔，迈克尔，内尔，马蒂亚斯，2019。延迟神经渲染:使用神经纹理的图像合成。ACM Trans图表。38 (4), 1–12.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb154)

提斯，胡斯图斯，佐尔霍夫，迈克尔，斯塔明格，马克，西奥博尔特，克里斯蒂安，涅纳，马蒂亚斯，2016。Face2Face:实时人脸捕捉和重现RGB视频。IEEE计算机视觉和模式识别会议论文集。第2387-2395页。

托迪斯科、马西米利亚诺、王、辛、维斯曼、维尔、萨希杜拉、Md、德尔加多、赫克-托尔、诺茨奇、安德烈亚斯、山木、俊一、埃文斯、尼古拉斯、金努宁、、李、孔艾克、2019。ASVspoof 2019:欺骗和伪造音频检测的未来前景。arXiv预印本[arXiv:1904.05441](http://arxiv.org/abs/1904.05441).

[Tolosana,鲁本，维拉-罗德里格斯，鲁本，菲尔雷斯，朱利安，莫拉莱斯，艾塔米，奥尔特加-加西亚，哈维尔，2020。深度伪装与超越:一项关于面部操纵和假检测。Inf。聚变64，131–148。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb157)

Trinh，Loc，Tsang，Michael，Rambhatla，Sirisha，Liu，Yan，2021。通过动态原型进行可解释且值得信赖的deepfake检测。IEEE/CVF计算机视觉应用冬季会议论文集。第1973-1983页。塔克，帕特里克，2019。最新的人工智能武器:“深度伪造”地球照片。[https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-](https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/)

[faking-whole-world-and-china-ahead/155944/](https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/).

[Turek,](https://www.darpa.mil/program/media-forensics)米（meter的缩写））, 2019.媒体取证(MediFor)。<https://www.darpa.mil/program/media-forensics>.

[弗多里瓦，路易莎，2020。媒体取证和深度造假:概述。IEEE J. Sel。顶端。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb161)

[签名。过程。14 (5), 910–932.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb161)

VidTIMIT数据库，2022。<http://conradsanderson.id.au/vidtimit/>.

文森特，帕斯卡尔，拉罗歇尔，雨果，本吉奥，约舒厄，曼扎戈尔，皮埃尔-安托万，2008。用去噪自动编码器提取和合成鲁棒特征。《第25届机器学习国际会议论文集》。第1096-1103页。

[Wang,辛，托姆，尼古拉，科德，马蒂厄，2017。凝视潜在支持向量机用于通过弱监督区域选择改进的图像分类。模式认识一下。72, 59–71.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb164)

王，李胜玉，王，奥利弗，张，理查德，欧文斯，安德鲁，埃夫罗斯，阿列克谢，2020。CNN生成的图像非常容易辨认...暂时如此。《IEEE/CVF计算机视觉和模式识别会议论文集》。第8695-8704页。

[Xuan,新生，彭，博，王，魏，董，景，2019。关于…的推广甘图像取证。在:中国生物识别会议。斯普林格，](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb166) [第134-141页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb166)

[Yang,朝飞，丁，利亚，陈，，李，海，2021。防御gan基通过感知变形的敌对面孔进行深度伪装攻击。在:IEEE国际神经网络联合会议。IEEE，第1–8页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb167)

[Yang,辛，李，岳尊，吕，四维，2019。使用不一致的揭露深层假货头部姿势。载于:IEEE国际声学、语音和信号会议处理，ICASSP。IEEE，第8261–8265页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb168)

[Yang,、倪、荣荣、赵、姚，2016。基于重捕获图像取证拉普拉斯卷积神经网络。国际数字技术研讨会水印。施普林格，第119-128页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb169)

叶，金元，陈，西文，蔡尚伦，王，盛德，2020。用对抗性攻击扰乱基于图像平移的深度伪造算法。IEEE/CVF计算机视觉应用冬季会议论文集。第53-62页。

[Younus, 穆罕默德·阿克兰，哈桑，塔哈·穆罕默德，2020。有效快速深度造假基于哈尔小波变换的检测方法。国际会议计算机科学与软件工程。IEEE，第186-190页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb171)

扎哈罗夫，叶戈尔，Shysheya，Aliaksandra，布尔科夫，叶戈尔，伦皮茨基，维克托，2019。现实神经说话头模型的少镜头对抗学习。IEEE/CVF计算机视觉国际会议论文集。第9459-9468页。

[Zhang,韩，古德菲勒，伊恩，，迪米特里斯，奥登纳，奥古斯都，2019。自我关注生成性对抗网络。国际机器学习会议。PMLR，第7354-7363页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb173)

[Zhang,应，郑，，事情，Vrizlynn L.L .，2017。自动人脸交换及其应用检测。第二届国际信号与图像处理会议，ICSIP。IEEE，第15–19页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb174)

赵，天辰，徐，向，徐，，丁，惠，熊，，夏，魏，2021。用于deepfake检测的学习自洽性。IEEE/CVF计算机视觉国际会议论文集。第15023–15033页。

[Zheng,李雷、杜夫纳、斯特凡、伊德里西、哈立德、加西亚、克里斯托夫、巴斯科特、阿提拉、2016.用于降维和人脸的连体多层感知器身份证明。多媒体工具应用程序75 (9)，5055–5073。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb176)

[Zhou,彭，韩，辛童，莫拉里乌，弗拉德，戴维斯，拉里，2017。双流神经网络篡改人脸检测网络。在:IEEE计算机视觉和模式识别研讨会。IEEE，第1831-1839页。](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb177)

[Zhou,信义，扎法拉尼，礼萨，2020。假新闻调查:基本理论，检测方法和机会。ACM计算机。Surv。53 (5), 1–40.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb178)

朱，俊彦，朴泰星，伊索拉，菲利普，埃夫罗斯，阿列克谢A，2017。使用循环一致对抗网络的不成对图像到图像翻译。IEEE计算机视觉国际会议论文集。第2223-2232页。

[Zubiaga,阿凯茨，阿克，阿赫梅特，邦切娃，卡琳娜，利亚卡塔，玛利亚，普罗克特，罗布，2018。社交媒体中谣言的发现和解决:一项调查。ACM计算机。Surv。51 (2), 1–36.](http://refhub.elsevier.com/S1077-3142(22)00111-4/sb180)