眼见未必是真！如何识别“AI换脸”？

在人工智能时代，我们眼睛见到的未必是真的。“AI换脸”技术的学术名词叫做深度伪造技术（Deepfake），也就是深度学习（Deep learning）+伪造（fake），是机器深度学习技术和人脸假冒的结合体。大家可能对“AI换脸”技术有种熟悉又茫然的感觉，但不论是某些短视频下方“该视频疑似使用智能合成技术，请谨慎识别”的显著标识，还是大家有时使用的发型换脸软件，都表明“AI换脸”技术已经融入到大众生活中。

俗话说得好，有人的地方就有江湖。看似有趣好玩的“AI换脸”技术，实则背后暗藏着众多黑灰产业链——换脸变成美女的诈骗犯罪团伙，传播虚构照片、视频的卖家，以及一些换他人脸、谋我之利的个人行为，还有对公众人物的鬼畜恶搞和造谣等，这些都让我们看到了这项神奇技术的另一面。

正确地认识是合理利用的前提。“AI换脸”技术是如何定义的，它是如何实现换脸的，又能应用在哪些领域呢？

在深圳中国国际高新技术成果交易会上，参观者在展台体验“AI换脸”。（资料图片） 新华社记者毛思倩 摄

“AI换脸”是如何实现的？

“AI换脸”最开始走入大众视野是在2007年，外网Reddit社区上一位名为deepfakes的用户利用该技术将成人影像女主角换脸为一些好莱坞知名女星并发布，最终由于影响过于恶劣而被封号。

但该用户后来直接公开了该技术的源代码，由此衍生出了各类换脸小程序和App等。我国的换脸潮则是从2019年B站杨幂版黄蓉的突然出圈以及“ZAO”软件的爆火开始的，直到现在，换脸视频在B站仍是一种流量密码。

根据研究，当下的数字图像篡改技术大致可分为合成、润饰、计算机生成、变种、增强和绘画六类。

“AI换脸”则是在深度学习算法的加持下，计算机篡改人脸图像的plus版。它在人脸识别技术及图形学伪造技术的基础上，利用一些深度学习算法，更高效、逼真地将图片或视频中源人脸替换到目标人脸上去。

从专业层次上讲，换脸技术的深度学习算法主要包括生成对抗网络（Generative Adversarial Network，GAN），卷积神经网络（Convolutional Neural Network，CNN），循环神经网络（Recurrent Neural Network，RNN）和编码器-解码器（Encoder-Decoder）四种。

其中，CNN和RNN是分别具有不同功能的前馈神经网络和递归神经网络；GAN是一种通过生成模型和判别模型互相博弈的方法来学习数据分布的生成式网络；而编码器-解码器则分别是将图像编制为计算机数据信号以及信号还原为图像的工具。

另外，替换过程中应用到的技术，除了大家比较熟悉的人脸识别技术，还包括图形学伪造技术，它能利用3D模型追踪人脸关键点并将源人脸仿射变换到目标人脸上。

一般来说，“AI换脸”技术被定义为基于人脸图像识别的深度伪造技术，而深度人脸伪造技术包括身份替换、面部重演、属性编辑、人脸生成等。大众所熟知的换脸技术只是其中身份替换的一种方式，仅仅是深度伪造技术中小小一隅罢了。

识别“AI假脸”的简单方法。

“AI换脸”主要应用场景

目前，我国的“AI换脸”技术主要应用于影视娱乐领域，例如影视剧里的角色换脸、短视频的换脸制作、换妆换脸软件等。

除此之外，“AI换脸”在纪录片、地图实景、监控等涉及隐私保护的方面也发挥了不小的作用。

与此同时，随着深度伪造技术的全面发展，它也在游戏、虚拟现实VR等领域为用户提供了更真实的游戏体验感。同时，深度伪造技术将与过世亲人视频通话或在VR中再相见变成了可能，让我们看到了未来这类技术的另一种美好。

如何轻松识别“AI假脸”？

“AI换脸”逐渐成为一种新型的诈骗手段，伪造的内容难以被肉眼或耳朵分辨出来。如何精准识别“AI换脸”成为一个亟待解决的问题。北京邮电大学网络空间安全学院教授周琳娜接受记者采访时表示，“AI换脸”检测技术有多种手段，对于公众而言，可以依据“AI假脸”的bug进行简单识别。

首先是注意观察“AI假脸”的纹理特征。比如，伪造后的视频人物的眼睛或牙齿轮廓细节容易不一致；两只眼睛瞳孔的颜色不一样或瞳孔中心反射的细节不一样；或是很多伪造视频由于视频分辨率低于原始视频分辨率，伪造出来视频的牙齿边缘过于整齐等。

其次，注意观察“AI假脸”的生物特征。“AI假脸”有可能不符合正常人的生理特征，比如，一个健康成年人一般间隔2—10秒眨一次眼，每次眨眼用时0.1—0.4秒，而在伪造视频中，人的眨眼频率可能不符合上述规律。

再次，注意观察“AI假脸”的嘴部固有特征。比如，由于人嘴部的运动是最为频繁且快速的，因此AI软件无法真实准确地渲染其连续动作。基于此，可以作为判断视频真假的因素之一。

此外，伪造后的视频会造成一定的视频抖动，导致视频出现帧间不一致情况。

在周琳娜看来，“AI换脸”的负面影响出现在电信诈骗中，会造成整个社会公信力降低的问题。“AI换脸”技术是一个新兴事物，它能给我们的社会带来很多便捷，但是也会影响到我们的生活。因此必须结合社会需求、技术能力，完善和制定好相关法律法规，做好“AI换脸”技术的应用和治理。

技术加持用“魔法”打败“魔法”

9月8日，2024年国家网络安全宣传周在南沙正式拉开帷幕，网络安全博览会也同步举行。在中国移动展位的现场，一个“AI换脸”视频被实时鉴伪技术精准识别为伪造，整个过程仅需5秒左右。深度合成换脸实时鉴伪体验区吸引了许多观众的围观，纷纷上手体验深度鉴伪技术。“我们先是模拟诈骗分子‘AI换脸’的过程，再去反向学习技术，进而升级迭代反制系统。”中国移动信息安全与管理运行中心副经理王晓晴介绍，他们在实名入网业务上开展了试点应用，基于鉴伪能力可实时检测实名验证过程中的AI人脸，主动防范“AI换脸”线上开卡、补卡等风险发生，鉴伪的精确率达90%以上。

“一旦辨别为‘AI人脸’就出会弹窗警告，提醒相关用户存在的风险。”王晓晴表示，深度鉴伪能力还将在5G新通话等业务场景中试点应用，对新通话内容进行伪造检测、内容有害检测等，及时发现涉诈有害风险，有效防范涉诈等违法违规事件。

记者在中国移动展位看到，一位观众体验名为羲和·防诈卫士的全场景防诈服务，运用这一服务远程临时停机用户家人的手机，防止诈骗电话继续轰炸。王晓晴介绍，该项服务通过AI自动检测到用户家人接听疑似诈骗电话时，将自动向用户及家人发送必达通知，实现识别预警。若家人处于通话中并未看到提醒消息，用户可向家人发送语音预警，直接在当前来电中插播语音提醒。同时用户可以远程临时停机家人的手机，防止诈骗电话继续轰炸。王晓晴表示，这一服务提供25种骗局套路文本的识别，目前准确率达到99%。

综合广州日报、北京科技报报道