新闻热线：010—58884102
E—mail：zdd@stdaily.com
科技日報
■责编 翟冬冬
2020年8月13日 星期四

AI被一件T恤蒙蔽“双眼” 见识不够是根本原因

视觉中国供图

本报记者 **谢开飞**
通 讯 员 **许晓凤 王忆希**

只要穿上一件印有特殊图案的T恤，就能骗过AI人体检测系统，从而达到“隐身”效果？

近日，这一场景真实上演。美国东北大学和麻省理工学院等研究机构，共同设计了基于对抗样本技术的T恤。据研究人员介绍，这是全球首个在非刚性物体（如T恤）上，进行的物理对抗性实验。AI人体检测摄像头无法准确地检测出穿着该T恤的行人，无论衣服发生怎样的褶皱或变形，都能达到“隐身”效果。

这件能让人在AI人体检测系统下“隐身”的T恤，其背后的原理是什么？这种缺陷会不会导致安全问题，要如何解决？科技日报记者就此采访了有关专家。

特殊图案便能骗过AI的“眼睛”

在本次实验中，一位穿着白T恤的男性和一位穿着黑T恤的女性从远处走来，在AI人体识别摄像头下，只能看到穿黑T恤女性的身影。

这是如何做到的？原来研究人员使用了一种被称为对抗攻击的方法来欺骗AI。仔细观察，白T恤上印有不同的色块，这些色块在人眼看来与普通图案无异，但对于机器来说，却会造成一定干扰。

中国科学院自动化研究所王金桥研究员解释说，科研人员对原T恤上的图案进行修改，通过技术手段生成具有较强干扰性的图案替换原有图案，改变了T恤原有的视觉外观，使得AI模型对数据标签的预测发生混淆和错误，从而到达攻击的目的。

“攻击者通过构造微不足道的扰动来干扰源数据，可以使得基于深度神经网络的人工智能算法输出攻击者想要的任何错误结果。而这类被干扰之后的输入样本被称之为对抗样本。”王金桥说。

对抗样本在实际中主要用来检验一些安全系数较高的系统，通过对抗的方式来提高AI模型的安全性，抵御可能面临的安全风险。比如刷脸支付，它必须具有一定的抗攻击能力，以便避免灾难性的后果，比如不能让攻击者简单地利用照片或者定向修改原输入就能破解用户支付系统。

有实验表明，对于一个正确分类的熊猫图像，在加入特定对抗样本的干扰之后，人眼看到的仍然是熊猫，但是AI图像识别模型却将其分类为长臂猿，且置信度高达99%。

不过，将对抗性图案印在衣服上这种欺骗AI的方式有一个缺陷，只要图案的角度和形状发生变化，就会轻易被识破。过去在设计对抗样本时，通常采用一些简单的变换，比如缩放、平移、旋转、亮度、对比度调整以及添加自适应的噪声等。

王金桥解释说，这些简单的变换，在产生静态目标的对抗样本时往往比较有效，但是针对行人这种非刚体的动态目标则容易失效。动态目标由于运动以及姿态变化，将导致这些简单变换发生较大的改变，从而使得对抗样本丧失原有的性质。

“相比过去设计的对抗样本，本次攻击的成功率更高。”福州大学数学与计算机科学学院、福建新媒体行业技术开发基地副主任柯逍博士指出，为应对人体移动造成的T恤形变，科研人员采用“薄板样条插值”的方法来建模行人可能发生的各种形变。同时，在训练阶段使用T恤上棋盘图案的格子来学习形变控制点位置变化关系，使得产生的对抗样本更加真实，对人体形变的贴合度更高。

AI视觉系统受到多方因素干扰

除了对抗攻击之外，在实际应用中的很多环境因素和人为因素，都可能导致AI人体检测出现失误。

如在自动驾驶场景下，由于天气条件恶劣（如大雪、大雾等）或者光线及路况复杂，导致前方人员成像模糊等，会极大影响前方目标检测性能。在监控场景下，可疑人员可能通过衣物、雨伞等的遮挡来干扰人工智能算法。

“排除本身紧急制动功能问题，具备行人检测功能的汽车也存在着无法及时、准确地检测出小目标人体等问题。”柯逍举例说，美国汽车协会曾对具备行人检测功能的多个品牌车辆做过一个测试，测试中用到的被撞目标包括成人假人与儿童假人。当车前出现儿童或汽车时速达到48千米时，仅一个品牌有一定概率检测出行人，其余3家品牌在两个场景下均未检测到行人。

为何在AI视觉识别技术下的目标检测模型如此脆弱？“在人类眼中，轻微的图像干扰并不会影响最终的判断，但对于AI模型来说却不是如此。”柯逍举例说，有相关实验表明，一个测试表现良好的图像检测与识别分类器，并没有像人类一样学习与理解目标图像真正底层的信息，而只是在训练样本上构建了一个表现良好的机器学习模型。

据了解，现有的AI视觉识别技术通常采用深度神经网络，本质上是一种特征深层映射，只是学习数据的统计特征或数据之间的关联关系，对数据量以及数据本身的丰富程度依赖较高，数据越多越丰富，则机器学习到的用于识别目标物的特征越具有判识度，也越能反映关联关系。

王金桥表示，但真实情况是，数据往往非常有限，使得神经网络学习到的模式也比较有限，难以让神经网络模型“见多识广”，导致其面对从未见过的数据时表现往往不尽如人意。另一方面，这种统计特征分布以及关联关系，一旦被攻击者获知或者破解，就有可能针对性地修改输入样本，从而改变模型的输出，达到攻击的目的。

AI视觉失灵易引发安全问题

穿上特殊T恤，达到所谓的“隐身”效果，其实就是混淆AI的视觉系统。AI目标检测技术的这种缺陷是否会导致安全问题的发生？

柯逍表示，美国汽车协会的汽车辅助驾驶案例中，行人被漏检或者未能及时被检测到，都可能导致交通事故的产生。此外，安防监控漏检危险人物与物品也可能导致安全隐患，不法分子可以利用对抗攻击来发现目标检测系统的漏洞，并进行攻击。

“安全问题的产生可能有模型本身缺陷问题，如泛化性能不足，训练数据单一，存在过拟合等现象。此时，应当尽可能地丰富训练数据，并在模型训练过程中加入防止过拟合的技术手段等来提升模型的实战能力。”王金桥认为，另一方面，实际系统中往往也需要考虑模型安全来增强结果可信度和模型的健壮性，加入攻击模型的预判，提高对抗样本的判别能力，从而降低安全风险。

当前，科研人员正不断提出精度更高、速度更快的AI目标检测模型，用于解决目标检测技术存在的漏检、误检、实时性与鲁棒性不强等问题。对于未来技术安全的构建，还需要做哪些努力？

王金桥认为，人工智能目前总体还处于起步阶段，现有的人工智能算法本质上还是学习简单的映射关系，并未真正地理解数据背后内容及潜在的因果关系。因此，其理论创新和产业应用还面临着诸多的技术难点，需要科研人员持续攻关，实现真正意义上的智能以降低应用的风险。

“其次，科研人员在进行技术研究以及新技术的应用过程中，应当尽可能地考虑各种安全问题，加入对抗样本防攻击模型，并做好相应的处理措施。”王金桥建议，从社会层面也应当建立和完善人工智能相关的法律法规，对技术的应用范围加以引导，对可能出现的安全问题作出相应的指导和规范，营造更加全面和成熟的科技创新环境。

靠“面子”进小区、取厕纸很方便，但未必安全

新华社记者 **杨有宗 何欣荣 龚雯**

点完餐看一下摄像头就能完成支付，住酒店刷脸后才能登记，上公厕用厕纸也得刷个脸才能取。随着人工智能的发展，人脸识别技术得到广泛应用，“脸”的应用场景被不断拓宽。

人脸识别技术看似“高大上”，但其存在的个人生物信息被过度采集和滥用的风险也不容忽视。相关专家表示，人脸识别技术不是万能的，收集、处理个人信息应当遵循合法、正当、必要的原则，基于身份验证等需要收集相关信息后也应尽到严格保管的责任和义务。中央网信办等部门近期也表示，针对面部特征等生物特征信息收集使用不规范等重点问题，App专项治理工作组将开展专题研究。

便利支付+身份认证 人脸识别应用场景不断拓展

在屏幕上点餐，选择刷脸支付，人脸比对后，输入手机号码后四位就完成付款。在上海一家商场的肯德基餐厅，记者观察发现，使用自助点餐机点餐的顾客中，选择“刷脸支付”的消费者占到两至三成。

除了大商场、大超市，部分便利店和街边小店里，刷脸支付设备也得到广泛应用。在上海陕西南路一家便利店，市民洪浩晨在购买一瓶饮料后，也通过刷脸方式完成付款。“从去年开始用刷脸支付就比较多，感觉比二维码方便。”洪浩晨说。

除了消费领域的便利支付，身份认证是人脸识别技术的另一大主战场。在全国机场和火车站的部分通道，乘客将身份证放在相应感应区，面部正视摄像头，每人只需几秒钟就可完成相关信息核验，快速进站。

一些公共服务机构还利用人脸识别技术来打击“黄牛”。复旦大学附属肿瘤医院一半以上的患者来自外省份甚至境外，该院去年就推出“人脸识别+身份绑定”系统，强化早高峰时段热门专家（特需）现场号源的管理。通过人脸识别系统绑定挂号人的身份，使得号贩子失去了现场“投机挂号”的操作空间。

不仅是医院，记者发现，各地政务类App中，刷脸登录、人脸验证已经广泛使用，如某地公积金App，用户就可通过人脸识别完成验证，线上支取公积金。上海市民胡志国说：“年纪大了，密码经常忘，尤其是登录一些不常用的App时都需要重置密码，刷脸就不存在这个问题。”

过度暴露隐私 人脸识别应用面临三大问题

不可否认，在人工智能成为新基建的背景下，人脸识别技术有其先进性和高效率。但任何先进技术的应用，都有其边界。在一些不必要的场景下，让渡自己的隐私，来换取一张通行证，必然会引起越来越多的反弹。

“现在上班刷脸打卡，工作时刷脸打开手机、笔记本电脑，午饭时刷脸支付，出差住酒店也得刷脸登记，甚至上公厕取厕纸都要刷脸，这张老脸是越刷越多，总感觉不对劲。”有网友如此感慨。

——强制刷脸遭质疑。记者梳理发现，关于公共场合使用人脸识别技术的争议和投诉正在增多。在合肥市“12345政府服务直通车”上，7月份有市民投诉：“繁华逸城”小区更换新物业后，办理门禁卡强制要求采集业主人脸信息。

对此，肥西县政府回复称，该项目的初衷是创建智慧平安小区，系统最终接入公安后台。“考虑老人和儿童人像采集不方便，可以办理门禁卡。”

上海一居民小区近期将小区门禁系统改为人脸识别系统。小区居民王女士说，改造前所有住户均需到物业采集人脸信息，“其实大家对采用人脸识别系统还是认可的，只是不知道个人信息是否会得到很好保护。物业为了让大家放心，出具了一份承诺书，承诺将严格保存收集的相关信息。”

——技术能力参差不齐。丰巢快递柜此前曾试点“刷脸取件”，其后被发现使用打印的取件人照片，也可以轻而易举地刷脸打开快递柜取件。丰巢回应称，“刷脸取件”功能仅为小范围试运营，并将测试版下线。

据悉，人脸识别技术可粗略分为基于2D人脸图像的技术和基于3D人脸图像的技术。通过照片即可完成人脸验证，大概率是采用了技术门槛较低的2D人脸图像认证。中国物流学会特约研究员杨达卿说，快递物流行业涉及消费者个人信息和财产，推广使用新技术时应慎之又慎。

——信息安全存隐忧。“密码泄露了，可以换一个，这脸部信息要是泄露了，可怎么换啊？”不少网友表示。

北京大学法学院教授薛军表示，人脸信息作为生物识别信息，一般来说伴随着人的一生，是不可更改的。这与手机号码之类的个人信息不一样，后者发生泄露，实在不行还可以换一个。但人脸信息发生泄露，不太可能去“换脸”。

不宜普遍适用 更不能在商业领域强制使用

人脸识别技术虽然有其优势，但并不能在生活各个领域普遍适用，更不能在部分商业领域强制使用。北京志霖律师事务所律师赵占领认为，收集使用个人信息，需要遵循三个原则，也就是合法、正当、必要原则，但目前缺少判断必要性的标准和依据。“目前收集人的脸部特征信息的商业机构，大多数是基于身份验证的需要。在企业收集这类信息后，能不能妥善保管并按照事先所告知的方式去使用相关信息，这也是大家最为担心的。”

薛军认为，利用人脸信息来快速、精确识别个人主体，对于个人行动轨迹的追踪非常高效，但因此对个人隐私权可能带来的侵害，也非常严重。

中国政法大学传播法研究中心副主任朱巍说，对收集个人生物信息的管理，核心在于对获取方的管理。“现实中遵循的原则应当是能不采集个人生物信息就不采集，而且宜通过立法，进一步明确具备采集资格的主体范围。消费者面对商家，也应当有控制、注销已被采集的生物信息的权利。”

情报所

保真且好看 自动上妆系统提升证件照颜值

科技日报讯（**黄龄亿** 记者**江耘** 通讯员**程振伟 孙垚棋**）近日，记者从杭州电子科技大学获悉，该校学生黄智坤建立了一套自动上妆系统，该系统能实现给人脸照片上妆且不失真，上妆程度、风格可控。经处理的照片既有美颜效果，又能达到证件照的要求。相关论文《基于现实世界的身份不变人脸自动上妆网络》近期收录于国际人工智能联合会议（IJCAI 2020）。

“现有的自动上妆方法中存在着两个问题，一是当背景复杂时，在生成带妆图像的同时图像的背景也会发生改变，导致图像失真。二是化妆程度不可控，导致上妆效果不理想。”黄智坤介绍，新设计的处理方法，能在不改变背景的条件下，准确、有效地完成人脸自动上妆，使生成的图像更加真实，更加贴近使用者的需求，又能用于身份证、护照等证件照，达到了实用和美观的平衡。

黄智坤还将这套保真自动上妆系统与现在普遍运用的社交工具上的美颜方法进行了比较。他表示，美颜照很难满足证件照的保真要求，但该自动上妆系统则可以。现有的美颜方法更多的是通过磨皮、改变轮廓等方法让图像中的人物更接近大众的审美，但改变轮廓的方法会使图像变得不真实。新的自动上妆系统只对人脸化妆区域进行处理，不会改变人物的轮廓，化妆前后人物图像的真实性会得到充分保证。

视觉中国供图

用“黑科技”赋能体育赛事 “智能亚运”解决方案揭晓

新华社讯（记者**夏亮 王梦**）近日，2022年杭州亚运会“智能亚运”重点项目解决方案揭晓，涵盖智能指挥、智能安防、智能生活等10个领域的30个解决方案将用“黑科技”为亚运赋能。

此次获奖的30个解决方案涵盖智能指挥、智能安防、智能安检、智能生活、智能表演、智能场馆、智能语言服务、智能出行、智能观赛、创新技术应用展示十大领域，广泛应用大数据、人工智能、虚拟现实、物联网、区块链等尖端技术，旨在通过科技手段提升赛事运行效率，为运动员、普通观众和游客提供优质服务，丰富赛事呈现和传播方式，创造交互式、沉浸式的观赛体验。

杭州亚组委副秘书长、杭州市副市长陈卫强表示，杭州亚组委将围绕杭州打造全国新型智慧城市建设“重要窗口”契机，发挥数字经济优势，调动社会各界力量，将创新科技融入亚运筹办各个环节，擦亮“智能亚运”金名片，为办成一届成功、精彩、圆满，令人难忘的体育文化盛会贡献智慧和力量。

“智能亚运”重点项目解决方案征集活动自2019年9月25日正式启动以来，吸引了众多高新企业、高等院校和科研机构为智能亚运献计献策，共征集到187份应征方案，经过亚组委和相关专家多轮评审，30份应征方案从中脱颖而出，将逐步应用于亚运筹办各领域。

据了解，下一步杭州亚组委将围绕“体育亚运”“城市亚运”“品牌亚运”主线，发挥杭州打造“全国数字经济第一城”的特色和优势，依托“城市大脑”建设，推进“智能亚运”的落地实施和推广应用，让“黑科技”在开闭幕式、火炬传递、赛事运行等场景高光亮相，让运动员和八方来宾切身感受到智慧城市的聪明，未来城市的魅力，开启“智能亚运”新时代。

合肥成立新一代人工智能产业发展联盟

新华社讯（记者**董雪**）记者从合肥市发改委获悉，为完善人工智能产业生态，合肥近日成立新一代人工智能产业发展联盟，该联盟由合肥综合性国家科学中心人工智能研究院牵头组建，拥有高等院校、科研院所、金融机构、企事业单位等各类成员单位322家。

据了解，推动人工智能产业政产学研用是该联盟的主要目标之一。该联盟将密切跟踪人工智能产业前沿，强化方向引领和指导，为重点企业、高等院校、科研院所等搭建专业支撑平台。同时，该联盟还将主动对接和借力各方资源，积极策划内容丰富、形式多样的对接活动，服务人工智能企业，助推人工智能产业加快做大做强做优。

牵头组建该联盟的合肥综合性国家科学中心人工智能研究院于今年4月揭牌运行，是由安徽省和中国科学院共同发起设立、参照国家实验室体制机制建设运行的新型研发机构。该研究院的目标包括构建面向产学研用创新环节的服务环境，形成大规模、开放式、可重构的人工智能应用试验环境和人才培养机制，培育一批国际领先的人工智能科技成果并产业化等。

据悉，合肥是全国重点布局人工智能产业的城市之一，近年来先后获批国家人工智能战略性新兴产业集群、新一代人工智能创新发展试验区，并拥有定位于人工智能领域的国家级产业基地“中国声谷”。截至今年上半年，“中国声谷”入园企业已有910户。

本报社址：北京市复兴路15号 邮政编码：100038 查询电话：58884031 总编室：58884048 58884050（传真） 策划运营中心：58884126 广告许可证：018号 本报激光照排 印刷：人民日报印刷厂 每月定价：33.00元 零售：每份2.00元