

人脸识别的偏见背后 是算法盲点还是叵测人心

皮尤研究中心的一项最新研究发现，Facebook 的新闻图片中，男性出现的频率是女性的两倍，且大部分的图片是关于男性的。

考虑到当下有 43% 的美国成年公民主要通过 Facebook 获取新闻资讯，皮尤研究中心使用机器视觉测试了 2018 年 4 月至 6 月期间 17 个全国性新闻媒体在 Facebook 上所发布新闻图片的性别比例情况。测试算法最终识别出了 53067 人，其中女性占 33%，男性占 67%，差距悬殊。但在现实生活中，美国人口的性别比例大致是均衡的。

那么，是谁在“扭曲”两性？

全媒派综合皮尤研究中心以及 MIT 媒体实验室科学家 Joy Buolamwini 的研究发现，走近人脸识别下两性的比例失衡现象，一起探究：为何有时候，算法眼中的你处在可男可女的模糊地带？进一步讲，性别之外还有哪些偏见？为了应对这种状况，我们可以做些什么？

人脸识别下的性别失衡

皮尤的这份报告指出，在 Facebook 上不同类型的新闻报道中，女性在图片中的“在场”情况始终低于男性。在与经济有关的帖子中，只有 9% 的图片为纯女性内容，与此形成鲜明对比的是，纯男性图像占到了 69%。女性在娱乐新闻图片中拥有更多展示机会，但总体上仍低于男性。

你也许会为女性身影的稀少而感到疑惑，这在一定程度上和更大的社会现实相关。比如，在关于职业足球队的新闻报道中，识别出的图像大多都是男性；在针对美国参众两院（女性占比为 25%）的报道中，识别出的女性面孔当然要比男性少得多。

抛开这些颗粒度更小的细节不谈，这份研究仍然揭示了一些值得警觉的现状：在 Facebook 新闻图像中，男性比女性更为突显；在两人或更多人的群像中，男性往往比女性多。同时，男性会占据更大的视觉空间。

研究人员还测量了图像中女性面部与男性面部的大小情况（目前的技术只能捕捉人脸的大小情况，忽略了头发、珠宝和头饰等因素的影响）。结果显示，男性面孔平均占到的图像面积更大，这种差异导致图像中男性平均面部尺寸比女性大出了 10%。在 Facebook 的图像中，这表现为男性人物能给读者带来更大的视觉冲击。

具体来看，在涉及经济相关的帖子中，女性面孔的平均尺寸比男性小 19%，但在娱乐相关的内容中，女性面孔的平均尺寸比男性大 7%。

像面部识别这样的机器视觉工具，正被越来越广泛地应用在执法、广告及其他领域，对性别的识别，是其基本功能之一。

在现实生活中，识别你周围人的性别再简单不过，但是对于计算机而言，它的工作需要经历怎样的步骤？

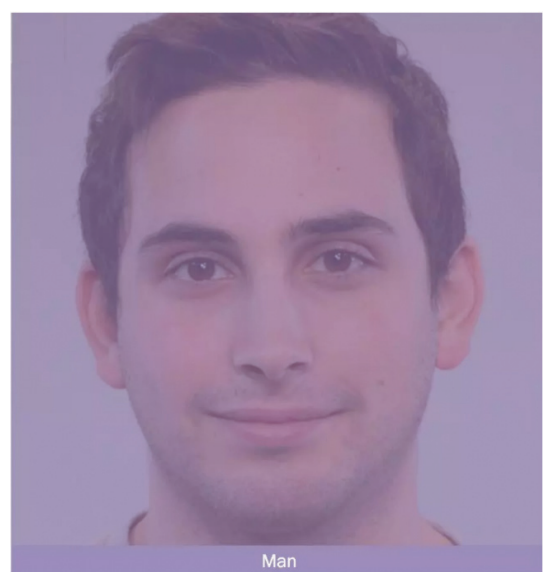
计算机如何“看出”你的性别？

“在给算法‘喂入’成千上万个图像案例后，作为一个‘成熟的算法’，面部识别系统自己就能学会如何辨别男性和女性。”这种回答虽然可以解释上文的疑问，但对于“黑箱”外的我们，可能并不容易理解这一学习过程。

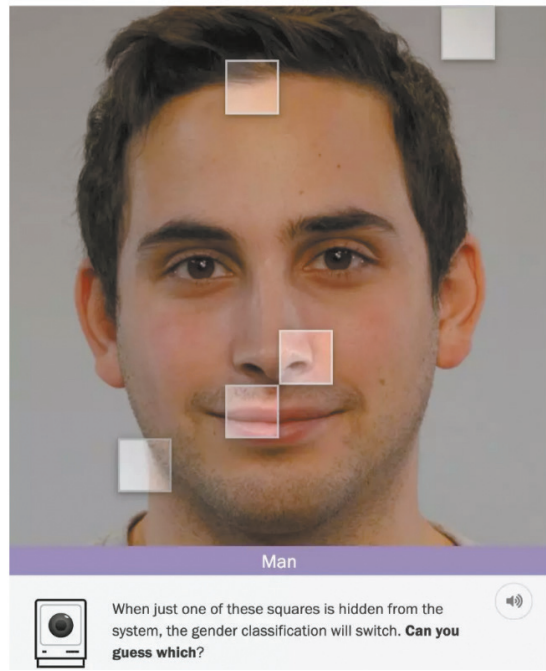
为了更好地了解这一过程中的规则，皮尤研究中心进行了一个有趣的实验，他们将自己中心工作人员的图像上传到机器视觉系统，并对图像内容进行部分遮挡，希望从中寻得规律，找到是哪些面部区域会让算法做出或改变决策。

在这个“人机博弈”的交互挑战中，你不妨也大胆猜测下，哪些部分影响了系统的判断？

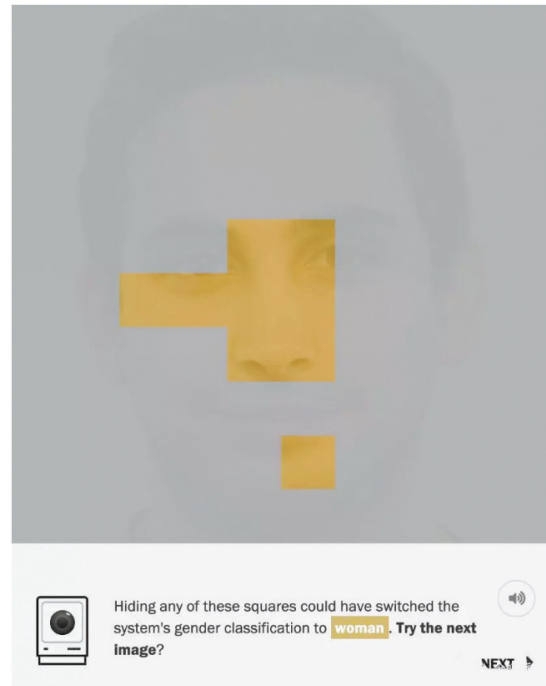
首先，输入一张清晰的图片到机器视觉系统，此时，不管是算法还是你，都可以清楚地判断出照片中人物的性别。



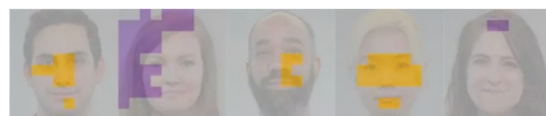
接下来，照片中出现了若干方框，提示信息告诉你，“选中某一方框意味着，在图片中遮挡隐藏该部分内容，你的选择有可能影响性别判断。”



最后，当你完成选择后，图片将呈现出能影响性别分类改变的所有区域。



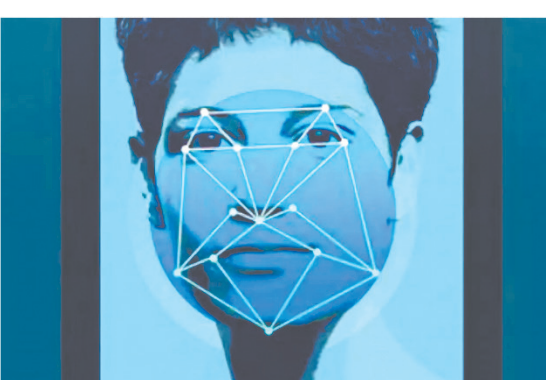
下面的这组图片，是互动实验中的部分结果图。当你选中了画面中紫色或者黄色的区域时，都会带来识别系统的决策改变。性别多元化的当下，在现实生活中，性别的识别虽也不易，但皮尤通过这个实验，更清楚地展示出，在算法系统中，让机器坚决肯定地说出被测者的性别，实在太难了。



仔细看这张图，你还能发现什么？——有时，导致模型识别发生改变的人脸部分，或许和我们预期的有很大的出入。比如，第四张图中，遮住人们的脸部，会导致系统识别发生改变，但更多的时候，令算法产生相反判断的“干扰区域”，其实是脸部边缘、发根、嘴角等区域。

从这些实验案例中，你也许还发现了，没有一个统一、稳定的规律能够对这一现象进行解释。有时，遮住某个被测者的脸部中间会导致性别识别发生变化，但以同样方式遮住另一个，并不一定会得到相同的结果。

机器学习的确可以极大地提高我们处理数据的效率，但与传统的计算机程序不同，机器学习遵循一系列严格的步骤，它们的决策方式在很大程度上隐而不显，并且高度依赖于用来训练自身的数据。这些特点可能导致，机器学习工具产生更难以被人理解和提前预测到的系统性偏差。

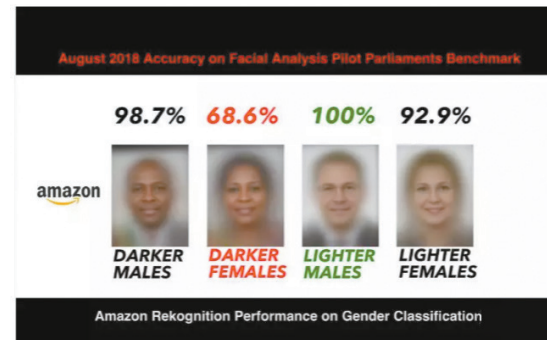


从这个角度看，皮尤研究中心用一个简化的实验，展示了用于训练算法的数据是如何将隐藏的偏差、意外的错误引入到了系统结果中。研究人员表示，随着算法正在人类社会中发挥越来越重要的决策影响力，了解它们的局限、偏差具有重要意义。

“偏见”带来了什么

最近，包括图灵奖获得者 Yoshua Bengio 在内的 26 位 AI 领域顶尖研究者，在一篇公开博文中要求亚马逊立即停止向警方出售其人工智能服务 Amazon Rekognition。亚马逊云计算部门前首席科学家 Anima Anandkumar 等人也加入了这一联合呼吁。

此前，多伦多大学的研究人员 Deborah Raji 和麻省理工学院媒体实验室的研究人员 Joy Buolamwini 撰写了研究报告，指出亚马逊的 Rekognition 在检测图像中肤色较深的女性性别时，要比判断肤色较浅男性性别的错误率高得多。该研究成果也得到了学者们的支持，但亚马逊曾对两人撰写的这篇报告及研究方法提出过异议。



亚马逊面部识别系统对不同肤色、性别的测试准确率

Joy Buolamwini 主导了一个名为 Gender Shades 的 AI 研究项目，在研究了各个领先科技公司的面部识别系统后发现，所有系统在识别男性脸上表现更佳，所有系统在识别浅色脸上准确率更高。深色皮肤女性的平均识别错误率高达 35%，深色皮肤男性的为 12%，浅色皮肤女性为 7%，浅色皮肤男性的错误率不超过 1%。

面部识别系统的“偏见”可能带来什么？



Google 将这位用户的朋友识别为“大猩猩”

“不管其正确性如何，面部识别技术都可能被滥用。” Joy 说道。准确或者不准确地使用面部识别技术对他人的身份、面孔、性别进行分析，都可能侵犯到他人的自由。比如，不准确的识别可能会使得无辜者蒙冤，受到执法人员的无理审查，这并不是假想的情况。英国非盈利组织“老大哥观察”（Big Brother Watch UK）曾发布一份报告，强调伦敦警察厅使用的面部识别技术有超过 90% 的性别识别错误率。去年夏天，英国媒体报道了这样一则新闻，一名年轻黑人男性因为面部识别技术失误，被误认为嫌犯而在众目睽睽之下遭到警方搜身检查。

一份被泄露的报告也显示，IBM 为执法机构提供了技术，能够根据头发颜色、肤色和面部特征等搜索视频中的人物。这一消息，引发了人们对警方将利用该技术关注特定种族的担忧。

著名非裔记者、平权运动家 Ida B. Wells 被识别为男性。

为了减少搜索人脸所需的时间，执法部门正在大量使用性别分类。如果需匹配人脸的性别是已知的，通过简单的二分法，就可以大量减少需要处理的潜在匹配数。性别分类正广泛应用到警务活动中。当这些有偏差的识别系统被广泛应用到社会生活中，就可能造成更糟糕的后果。

Joy Buolamwini 在 TED 上发表题为 How I'm

fighting bias in algorithms 的演讲

在 TED 演讲中，Joy 和大家分享了一个小故事：

在同样的光线条件下，面部识别系统只能检测到浅肤色的参与者；只有戴上白色面具，才能检测到深肤色的参与者。“在人工智能工具确定人脸的身份或者辨别表情信息前，最基本的前提是，检测出人脸。但是，面部识别系统在检测黑皮肤个体上，屡次失败。我只能安慰自己，算法不是种族主义者，是自己的脸太黑了。” Joy 说道。

偏差来自哪里

如果对比开发者自己声明的准确率和研究者们得出的研究结论，会发现一个有趣的事情：公司发布的数据和独立第三方的外部准确率总是有所出入。那么，是什么导致了这一差异？

Joy 提醒我们关注基准数据集的偏差。“当我们讨论面部分析技术的准确性时，是通过一系列图像或者视频测试来完成的。这些图像数据构成了一个基准，但并不是所有的基准都是平等的。”

亚马逊相关负责人表示，该公司使用了超过 100 万张面孔数据作为基准来检测产品的准确率。但是，不要被这个看起来很大的样本迷惑了。“因为我们不知道基准数据的详细人口统计学数据。没有这些信息，我们就无法判断，是否在基准选择上，就可能埋下了种族、性别或者肤色等偏见的可能。”

Facebook 曾宣布，在名为 Labeled Faces in the Wild（简称 LFW，为世界上人脸识别最知名的数据集之一）的数据集测试中，自己面部识别系统的准确率高达 97%。但当研究人员查看这个所谓的黄金标准数据集时，却发现这个数据集集中有近 77% 的男性，同时超过 80% 是白人。

为了在数据层就尽可能最大限度地剔除偏差，Joy 提出，应当构建更具包容性的基准数据集。为了平衡基准数据，她列出了世界上妇女在议会所占比例最高的十个国家，其中卢旺达以超过 60% 的女性比例领先世界。考虑到北欧国家和少数非洲国家具有典型代表性，Joy 选定了 3 个非洲国家和 3 个北欧国家，通过选择来自这些国家年轻、深肤色的个体数据来平衡数据集的皮肤等类型。

正是基于这个更平衡的数据集，他们对亚马逊、Kairos、IBM、Face++ 等公司的面部识别系统进行了重新评估。在 2018 年 8 月的研究中，他们发现，亚马逊和 Kairos 在白人男性识别工作上表现优秀，但亚马逊对有色人种的女性面孔识别的准确率很低，仅为 68.6%。

亚马逊的面部识别系统为奥普拉·温弗里的这张图片打上了男性标签，并给出了数据置信度。

Joy 表示，现实世界中的脸部识别要比实验检测更为复杂和困难，他们建立的基准数据集也并非完全经受得住考验，“但这就好比跑步比赛，在基准测试中的出色表现，起码能保证你不会刚起步就摔倒。”

即便在同样的基准下，面部识别系统的准确度数字可能也会发生变化。人工智能并不完美。在这种情况下，通过提供置信度给用户更具体的判断信息是一个有用的做法。

面部识别技术已经被广泛地应用在大规模监视、人工智能武器化和更多的执法环境中。但是，这项强大的技术是在没有得到充分监督的情况下，快速发展着。

为了减少对面部识别技术的滥用，算法正义联盟（Algorithmic Justice League）和隐私与技术中心（Center on Privacy & Technology）发起了“安全面孔承诺”（Safe Face Pledge）活动。在目前，包括亚马逊在内的很多科技公司尚未加入这一承诺。“根据我们的研究，贸然向执法部门或者政府机构出售面部识别系统将是不负责任的。”作为算法正义联盟创始人之一的 Joy 希望，在未来，更多的机构能加入到“安全面孔承诺”，能够负责任地、符合道义地为面部分析技术的发展付出行动。

毕竟，算法偏见的背后，其实是我们人类自己的偏见。📌



（本版由腾讯传媒 全媒派授权转载）