## "超人类机器视觉"是不存在的

我想再向大众解释一下,人工智能(AI)领域的"历史性突破"——"超人类机器视觉"是怎么来的。之前的文章解释过,但可能信息被埋在比较长的内容里,很多人没看到,以至于仍然蒙在鼓里。对于这种关键问题,简洁易懂是非常重要的,所以我决定在极短的篇幅之内把它解释清楚。

简言之,AI 领域的所谓"机器视觉",一个重要问题就是让机器回答"图片上是什么东西?"这个问题。比如图片上是一辆汽车,如果你说是"汽车",就算对了。所谓"Top-5"标准,就是每张图片给 5 次机会,你说出 5 种东西的名字,只要其中一个对了就算对。比如图片上是一辆汽车,你说"猫,吉他,风扇,汽车,橙子",这也算你对了。已经察觉到问题了吗?我们继续……

用大量图片做这个测试,统计"识别率",这就是很多机器视觉专家做的事情。有人用这个方式做了一个"人机对比"实验,发现机器的识别率超过了人,这就是"超人类视觉"(super-human level vision)的由来。"超人类视觉"被认为是 AI 领域的历史性突破,也就是这些年 AI 如此火热的原因。

但很多人没看出来,"Top-5"是非常不科学,不合理的。因为如果是人见过的东西,他只需要一次就能对,毫不含糊,另外 4 次机会完全没必要,而机器经常一次猜不对,需要另外 4 次机会才能蒙混过关。人精确地知道这个东西是什么,而机器只是在猜测"它可能是 A,也可能是 B,C,D,E",一共 5 种可能。所以对于同一个图片,虽然人和机器按照"Top-5"都算"对了",他们的准确程度其实大不一样。这就像设计一个考试,每道选择题本来只有一个正确答案,但却给了 5 次机会做对。这样就没法分清优等生和差等生了,甚至差等生有时候表现比优等生还好,因为不确切知道答案,瞎蒙都能做对。

我想一般人都理解这里的问题,然而"Top-5"标准却是 AI 领域所谓"超人类视觉"的来源。其实就算用如此不公平的标准,机器的识别率也没超出很多,几乎可以作为"噪音"忽略。当年参加测试的只有一个人,这个人不是从别处请来的"独立实验者",而是参与此项目的一个学生。这个人的名字叫 Andrej Karpathy,他后来成为了 Tesla 公司的 AI Director。根据如此偏颇,甚至可能是作弊的方式,他们宣称"机器视觉超越了……人类"。

你可能以为"Top-5"虽然不科学,不合理,也许也没什么太大的害处。你可要小心了,它其实可以致命,而且正在威胁着很多人的生命。人的生存环境里,往往是没有 5 次机会来判断一个东西是什么的,实际上经常只有一次机会,不能有任何含糊。比如,在马路上把"卡车"识别为"白板",是可以致命的。然而这种识别错误,就是 Tesla 的Autopilot 多次导致致命车祸的原因。详情可以参考我这篇 2016 年的文章。

这个测试其它的问题还有很多。比如,测试用的图片都是光照良好情况下的清晰图片,没有自然环境的各种复杂性,比如暗光,夜景,遮挡,阴影,反光,镜面,折射,模糊等。另外,只识别出物体"叫什么名字",并不等于知道了它的 3D 形状和边界,并不等于可以拿起,操作,或者避开物体,并不等于可以依靠这个技术来做"自动驾驶"。