



图片：Public Domain

## 真实场景下的视觉搜索



Spencer, 心理学

我们每天从睁开眼睛开始，每时每刻都在做的一件事就是找东西，更书面化一点，视觉搜索。比如在杂乱无章的书桌上找钥匙，在拥挤的食堂找同伴的面孔，在街道上找一个特定的店铺，这时候你会意识到你在做这件事，因为费劲。但大多时候，视觉搜索这件事发生地悄无声息，你根本不会觉察到，比如你想打开一个桌面上的文件，你得先找到这个文件的位置。虽然这时候的搜索发生地非常自动化，但我们的大脑其实在这个过程中做了不少工作，就像我在一篇文章里读到的很喜欢的一句话，“The results presented here are in many ways a case study of one of the central lessons of cognitive science: that how effortless and natural a process feels is a poor guide to how much work is supporting that process beneath the surface.”（大意：这个结果从多方面例证了认知科学的启示：一个感觉起来自然不费劲的加工过程背后其实有大量的支持工作）。

在视觉搜索这个任务以及背后的认知过程，心理学家 / 脑科学家的研究非常多，也开始地很早。大多数的视觉搜索研究都采用比较简单的范式（图1），比如在多个色块中搜索一个不同颜色的色块，在多个有一定朝向的线段中搜索一个不同朝向的线段，或者采用更复杂一点的刺激，搜索字母或者图形等。这些采用简单刺激的研究的好处在于剥离掉很多无关因素的影响，但相应地，和我们实际生活中的搜索任务相差较多。所以现在很多的研究开始关注真实场景下的视觉搜索（**visual search in a real-world scene**），这些研究结果也可以为机器视觉提供支持：如何在机器上重建人类的视觉搜索能力。

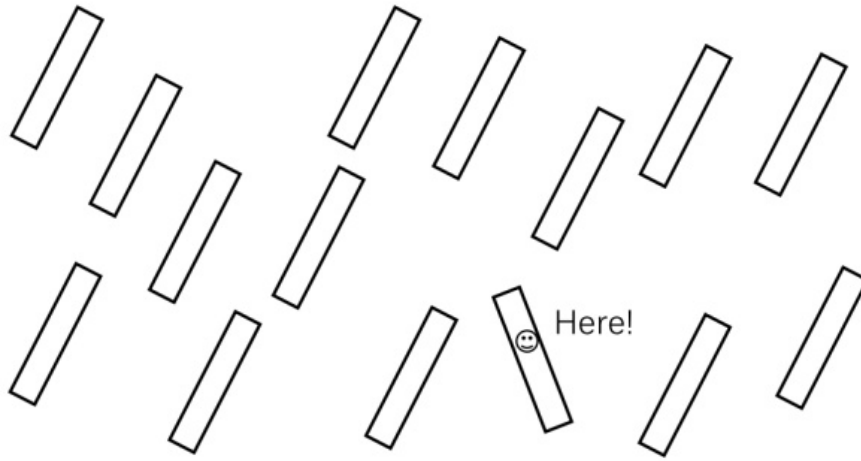


图 1 快来找我呀！

那么，真实场景下的视觉搜索研究和简单的视觉搜索关注的问题有什么不一样呢？

### 第一：情境线索(contextual cues)

在简单的视觉搜索任务中，目标（你要找的东西）出现的位置通常是随机的，可能出现在场景中的任意位置。但是在真实场景下的视觉搜索任务，目标的可能位置是受到限制的。比如你要下面这张图（图 2）上找直升机，那么你肯定会在天上（图片的上方）去找，而不会从沙漠（图片的下方）去找，这时候“天空”就是情境线索，限制了目标可能出现的位置。相反地，如果要去搜索吉普车，“沙漠”就成了情境线索。

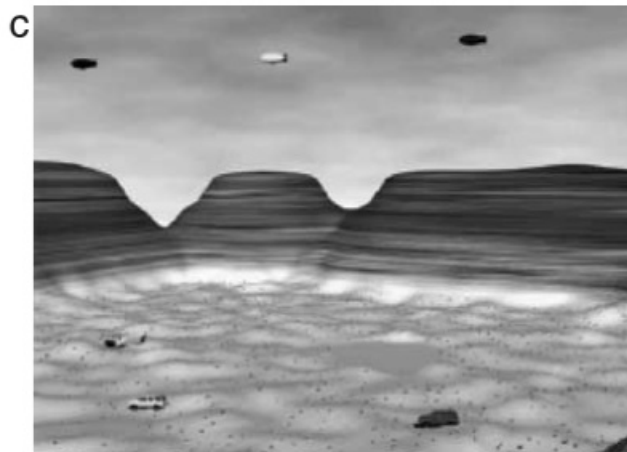


图 2 找直升机 / 吉普车

### 第二：物物共存关系(object-object occurrence)

物物共存关系指的是有些物体在真实生活中总是成对出现的，但你在搜索

物体 a 的时候可以依赖于物体 b 的位置。如图 3 所描述的，如果当物体（烟囱）没有出现在预期的位置（房顶），而是出现在树上的时候，视觉搜索成绩就会下降。从眼动的数据（图中的绿点）中，我们也可以看出被试从去预期的位置（房顶）搜索烟囱。



图 3 咦 烟囱去哪儿了? (Eckstein, 2011)

这两个因素在我们真实的搜索中起到关键性的作用，因此在构建模型过程中纳入这些简单的视觉搜索范式没有覆盖的因素也是至关重要的。Torralba 等人构建了一个基于情境线索的模型：情境引导模型（contextual guidance model, 图 4）。

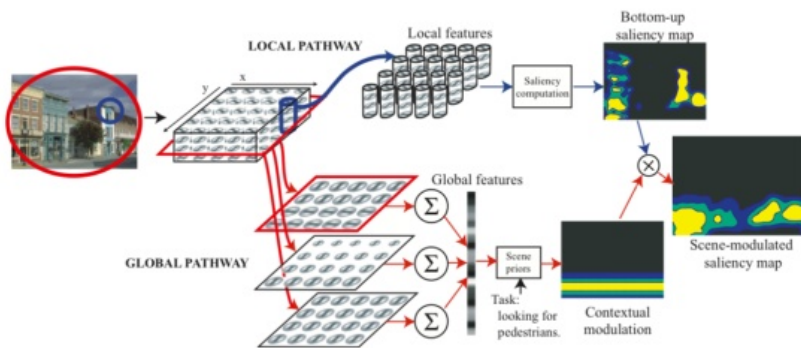


图 4 情境引导模型

这个模型基于两条通路：局部特征通路和全局特征通路。

局部特征通路指的是独立地计算各个空间位置的特征，构建一个突显地图。这个突显地图的产生依赖于颜色、纹理等基本的特征。从心理学的概念来讲，这是自下而上的一个加工过程，我们的注意力总是不自觉地被突显程度高的刺激吸引。多个脑成像的研究表明，这个自下而上的突显地图反映在从初级视觉皮层到顶叶区域 (Li, 2002; Goldberg, Bisley, Powell, & Gottlieb, 2006; Gottlieb, Kusunoki, & Goldberg, 1998)。而全局特征通路计算的整个情境因素，从而提供目标可能出现的位置的信息。研究者希望通过

这个模型来预测人在真实场景下的视觉搜索行为。研究者发现，这个情境引导模型在预测人的早期的注视行为上是较为有效的，而且跟仅仅基于突显地图构建的模型比较来看也是更好的（图 5）。

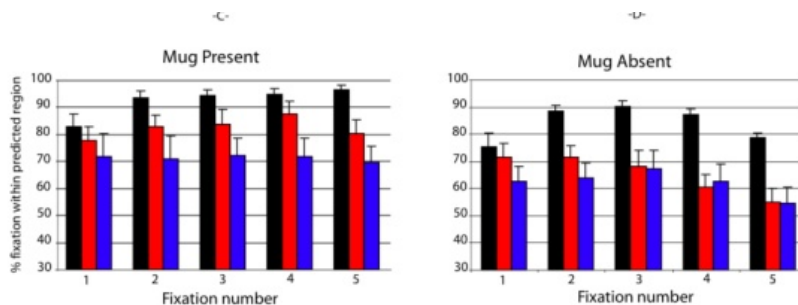


图 5 黑：真实观察者数据；红：情境引导模型的预测数据；蓝：基于突显地图的模型的预测数据

完。这篇文章没有总结句。

这篇文章的内容主要基于 *Journal of vision* 上的一篇综述，很有意思的是这篇综述的补充材料里还有对三位现实生活中的视觉搜索专家（一名渔夫、一名放射科医生、一名卫星图像分析员）的采访，可以说非常接地气了。

距离我上次写专栏文章已经快五个月了（发现知乎的插入图片功能好用了很多！），一方面是因为忙于搬砖，一方面也是感到生动有趣地写科普是一件很难的事情。令人高兴的是，我的博士期间的第一篇文章昨天正式 online 了，所以今天也算心安理得地不务正业一下。与大家共勉～

引用

Eckstein, M. P. (2011). Visual search: a retrospective...*Journal of Vision*, 11(5), 14-14.

Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766–786.

等。

[查看知乎原文](#)

客官，这篇文章有意思吗？

好玩！[下载 App 接着看](#) (๑•ㅅ•) ✧

[再逛逛吧](#) ' \_ >`

他们是当之无愧的 90 年代英伦摇滚巅峰代表



[下载「知乎日报」客户端查看更多](#)