



伦敦大学学院  
计算神经科学教授、  
北京师范大学认知神经科学  
学习与学习国家重点室访问  
教授李兆平

# 人的眼睛为什么如此 擅长“抓重点”

专访英国伦敦大学学院计算神经科学家李兆平

我们能在过马路时及时发现一辆闯红灯汽车，这相当于视觉系统在1秒内就从几十本长篇小说中，找到了最重要的两三句话。

《环球科学》记者 廖红艳

半个世纪前，美国麻省理工学院的一位计算机系教授，给自己的学生布置了一个暑期项目，编写一个程序，让计算机能够识别出图片上的简单物体——将图片信息输入计算机，由计算机输出图片上物体的名称。结果出乎这位教授的意料，这个看似简单的项目，最终花了科学家几十年的时间。直到最近几年，由于计算速度的提高，以及大数据、深度学习算法的出现，计算机的识图能力才有了较大提升。

相比而言，人类大脑在图像识别方面的能力，则令人惊叹。我们可以在半秒内准确分辨出猫和狗；在瞬间认出隐藏在一群鸟中的河马；过马路时及时发现一辆正向我们冲来的汽车；或是在安静听课时，用眼睛瞬间“抓”住突然出现的蝴蝶。这些由大脑视觉系统下意识做出的简单行为，其实并不简单。正常情况下，人的眼睛每秒可以接收几十兆信息，信息量相当于几十张图片或是几十部长篇小说。但实际上，能被我们注意到的信息量上限，只相当于两三句话。也就是说，我们在过马路时及时发现一辆闯红灯汽车，这个动作，相当于视觉系统在1秒内就从几十本长篇小说中，找到了最重要的两三句话。

在完全无意识的状态下，视觉系统就将我们的注意力引导到了最特别、最重要的位置。亿万年前，我们的远古祖先正是凭借这个超级本领，才在危机四伏的环境中生存下来，并繁衍至今。但其中的生物机制是什么，科学家一直无法解释。直到1999年，当时任职于英国伦敦大学学院的李兆平才提出了一个全新的解释，她认为，在视觉中枢的VI区（视觉初级皮层，是第一个接受视网膜信息的大脑脑区），形成了一个视觉显著图，就像地图一样，视觉显著图会告诉我们哪个位置重要，吸引我们的注意力，并引导眼球自动往那个方向看。这个解释也被称为“VI显著图理论”（VI saliency hypothesis）。

新理论刚提出时，质疑声很多。因为传统观点认为，建立显著图的应该是与智力紧密相关的前额叶或类似的脑区。而早前的两项研究，让科学家相信，VI是低级脑区，完成不了这么高级的任务。一个是1953年美国神经生物学家史蒂夫·库夫勒（Steve Kuffler）发现，视网膜神经元只能对图像上很小一块区域（即该神经元对应的感受野）有反应，并且只有当区域内有自己“喜欢”的点时，这个神经元才会有所反应。另一个是10年后，库夫勒的学生戴维·胡贝尔（David Hubel）和托尔斯登·魏塞尔（Torsten Wiesel）发现，VI区单个神经元的感受野比视网膜的稍大，这里的神经元开始能提取图像上的“小线段”，并且不同神经元会偏爱不同属性的线段，比如某一颜色、某一朝向、是否运动（因这一发现，两人获得了1980年诺贝尔生理或医学奖）。

按照这个规律，科学家乐观地推测，随着视觉信息的深入，比VI高级的脑区（V2、V3、V4、V5等）的神经元能够认出的图像会越来越复杂，并最终在前额叶之类的脑区形成一个视觉显著图，引导注意力的方向。这个假设似乎也有道理，然而几十年过去了，科学家照着这个思路开展研究，进展却远不像当年在视网膜和VI区那么顺利。而李兆平的理论指出，VI区其实并不只是

在提取图像里的小线段，还能造出显著图来引导注意力。这个理论引起大家反思：注意力的介入，意味着部分视觉信息在大脑里被选择或丢失可能已经在VI区附近发生，传统思路没有考虑这些因素，这很可能是高级视觉皮层难以理解的原因。

“VI显著图理论”可以解释很多我们司空见惯的现象。比如，当你面对“万绿丛中一点红”的场景时，为什么我们的眼睛会在第一时间看向“一点红”。“VI显著图理论”认为，VI区有一个内部神经网络，网络中的神经元会相互作用——喜好相同的神经元相互抑制，即同性相斥。另外，注意力分配时遵循的是“竞价原则”。可以想像成在VI的一个输出区——上丘脑——有一个专门负责拍卖注意力的拍卖店，它不问其他，只看神经元的出价（发放脉冲的频率），哪个神经元的出价高，就会把注意力卖给谁，然后指挥眼睛往这个神经元对应的感受野看。

举个例子，当视觉输入的是一张满是绿棒仅有一根红棒的图片时，因为绿棒很多，喜欢绿棒的神经元会相互抑制；而喜欢红棒的神经元，因为红棒只有一根，不会被抑制，结果发放频率最高，最后赢得了注意。这也可以解释，为什么我们总是会被最特别的事物吸引。此外，VI区生成的视觉显著图，可以直接输入上丘脑——视网膜负责输入信息、VI区负责生成视觉显著图、上丘脑负责输出结果，一个完美的模型诞生了。

近年来，“VI显著图理论”引起了学术界的极大注意，获得了越来越多实验的支持。这些实验包括生理学、心理学和脑成像的观察，包括李兆平和北京同事的合作项目。2014年牛津大学出版社出版了她著的《理解视觉：理论、模型与数据》（Understanding vision: theory, models, and data），这也是计算视觉领域的第一本教科书。最近，现为伦敦大学学院计算神经科学教授、北京师范大学认知神经科学与学习国家重点实验室访问教授的李兆平接受“未来论坛”邀请，在中国科学技术馆做了一场名为《用人脑研究脑，是探索更是挑战》的“理解未来”讲座，向在场观众介绍了自己在探索视觉这个“心灵的窗户”过程中的一些感想。讲座结束后，《环球科学》对李兆平教授进行了专访。

**《环球科学》：**你最早是在加州理工学院学习物理，为什么会转到计算神经科学领域？

**李兆平：**上大学的时候，我喜欢物理，也喜欢思考与大脑有关的各种问题。比如男人和女人的大脑有什么不同，为什么有人会认为女人不适合学习物理？小孩子的大脑为什么可塑性那么强，可以学会那么多东西？不过，我真正进入神经相关领

域，是缘于听了当时加州理工学院物理系约翰·霍普菲尔德教授（John Hopfield，1982年提出了著名的Hopfield神经网络）的一场讲座，霍普菲尔德教授后来成了我的博士生导师。我的博士论文名是《嗅球模型及延伸》（A Model of the Olfactory Bulb and Beyond），当时的加州理工学院非常开明，虽然我的论文看起来与物理专业不太相符，但还是给我颁发了物理学博士学位。

**《环球科学》：**有人说，人工智能“深度学习”中的“深”指包含了很多层虚拟神经元，灵感就来自于人脑视觉皮层的层状结构。你认为这个说法准确吗？为什么？

**李兆平：**上世纪80年代我参与过人工神经网络的教学和研究，不过当时计算机太慢，只能实现浅度学习，比如实现二三层算法，而现在已经能做到上百层了。上世纪80年代，研究机器学习和计算神经的其实是同一批人，每年在这个领域会召开名为“神经信息处理系统”（Neural Information Progressing Systems）的研讨会。不过到了90年代，研究人员开始向两个方向分化，一个是偏工程方向，也就是机器学习；另一个是偏理科（生物）方向，就是计算神经科学。

至于深度学习是否借鉴了人脑视觉皮层层状结构这个问题，我觉得只能算是形式上的借鉴。现在人工智能中用到的虚拟神经元和神经网络，工作原理与我们大脑中的神经元和神经网络很不相同。大脑的大部分机理，科学家都还没有搞清楚，而且，即使机理搞清楚了，应用到计算机上，也还需要做大量的工作。比如，现在的计算机远不能像人脑一样，懂得引导视觉注意力方向。

**《环球科学》：**有人说，如果类比物理学的发展史，神经科学目前还处于牛顿力学之前的阶段，你怎么看这种观点？

**李兆平：**从牛顿力学开始，物理学家几乎可以在原理上去解释很多自然现象。相比之下，在神经科学领域，还没有出现像物理学中的牛顿力学那样的、可以照亮一大片领域的理论，所以我觉得这种说法有一定道理。

但也可能在攻克大脑的这场战役中，不存在一个大战场，只存在许多小战场，各个小战场虽然不是互通的，但10个小战场打胜后，也许能取得大战场的胜利。目前，我们已经在许多小块的领域做得非常漂亮，在许多小战场上取得了胜利，比如大脑奖赏机制、大脑定位系统，等等。我还是非常乐观，希望在有生之年，可以循着V1、V2、V3、V4深入地研究下去，搞清楚视觉系统。甚至更乐观一些，既然猴子大脑里有50%脑区是专管视觉，也许通过视觉系统这个“心灵的窗户”，我们可以获得突破，窥见整个大脑的奥秘。

**《环球科学》：**你觉得计算机最终能够模拟人脑吗？

**李兆平：**只是模拟不去理解是没有用的。试想一下，如果不能理解大脑运作的机制，拥有一个仿真的大脑又有什么用？反过来，如果能真正理解原理，就不需要拘泥于形式生硬地模仿，比如，鸟与飞机的关系。人类虽然从飞鸟获得灵感，但最终并没有模仿鸟的形态，制造出所谓的机器鸟，当科学家搞清楚了流体力学，他们就可以用钢筋铁骨制造出跟鸟完全不像

的飞机。

**《环球科学》：**你提出的“V1显著图理论”非常漂亮，但正如你所说，这是一个小战场的胜利，你的下一个战场是什么？

**李兆平：**V1区还有许多奥秘值得我们探索。最近我和北京师范大学认知神经科学与学习国家重点实验室的李武老师合作，首次发现在V1区，注意力的吸引使神经元的灵敏度得到提高（神经元因视觉输入而发放的电脉冲越强，灵敏度越高）。而过去认为，只有在高级脑区，神经元才会发生因为获得注意力，灵敏度提高的情况。最近还研究了V2是否也参与了建立视觉显著图。虽然目前在V2方面还没有太大进展，但我比较乐观，希望能用我的一生，破解视觉的奥秘，甚至大脑的奥秘。

**《环球科学》：**1999年，“V1显著图理论”刚提出时，受到了许多质疑，一些科学家甚至认为是天方夜谭，但渐渐地科学界开始接受这个理论，特别是最近几年，越来越受重视，为什么学界的态度会发生转变？

**李兆平：**一个理论从提出到被大多数人接受，要经历一段时间。伟大的麦克斯韦的电磁学理论也花了很多年才被接受。当然，我的理论不像他的理论那么伟大，所以等更多年才被接受也很可能，我们也应该继续对它进行实验检验。在2008年，我用行为实验验证了“V1显著图理论”的一个意想不到的预测。这可能是促使学界态度转变的一个比较重要的因素。“V1显著图理论”虽然可以解释，为什么与众不同的事物会吸引我们的注意力，但所有理论都应该能用来推测未知现象，否则都只能算是“马后炮”。

所以，我设计了一个行为实验来验证自己的理论的一个预测。首先，请自愿者通过立体眼镜，左右眼分别看两幅不同的图——左眼看到的图上，几乎所有的棒都向左倾，只有一根棒向右倾，右眼看到的是一张空白图，要求自愿者尽快找到唯一的右倾棒。然后，通过眼动跟踪，记录自愿者看向右倾棒的时间。接下来，研究人员移动左图上的一个左倾棒，到右图的相应位置，得到两张新图。再请自愿者戴上立体眼镜，左右眼分别看两张新图，找出朝向不同的棒。结果发现，尽管两种情况下，视觉认知最终看到的图像是一样的（都是左眼输入图和右眼输入图的叠加），但相比第一种情况，第二情况下，自愿者需要花费更多时间，才能找到那根右倾棒。通过眼动跟踪，我们发现，即使明确要求自愿者尽快找到不同朝向的棒，自愿者还是会不由自主地看向跟任务无关的、只有右眼可见的那根唯一的棒。

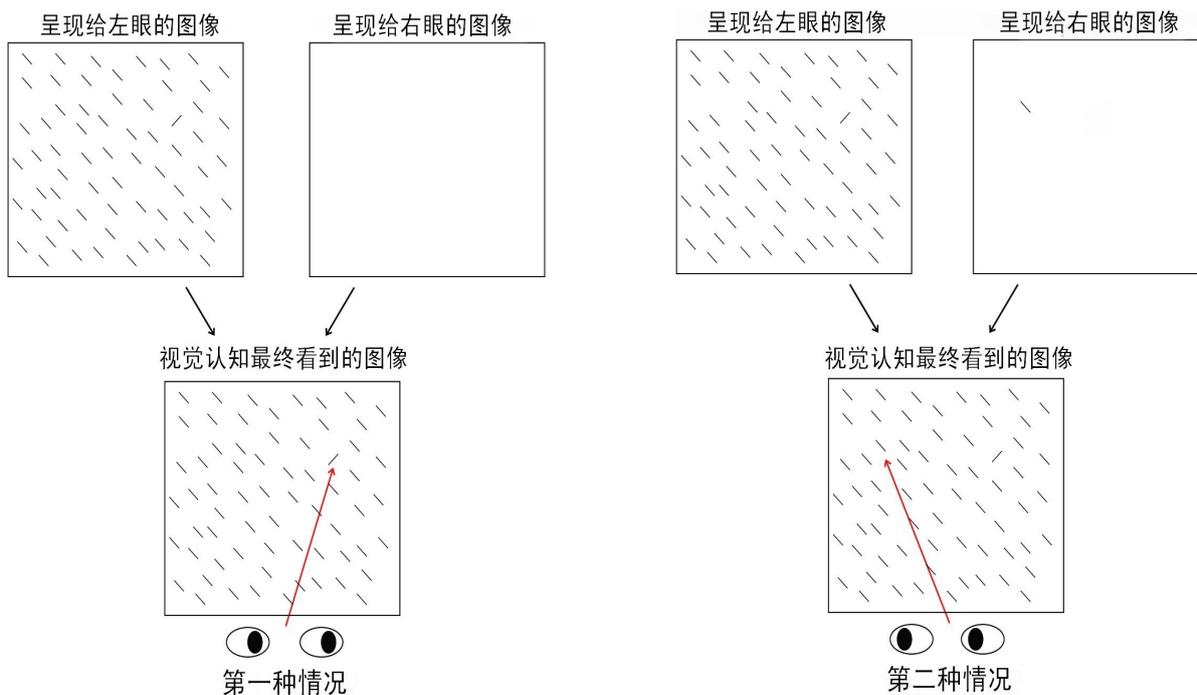
按照我们的理论，在第一种情况下，右倾棒是唯一的显著点，但第二种情况，会出现两个视觉显著点——一个是唯一的右倾棒，另一个是唯一在右眼画面中的那根棒。这也解释了，为什么自愿者的眼动轨迹经常会先看唯一的右倾棒，后看唯一的右倾棒。而且，在所有视觉皮层中，只有V1区的神经元存在左右眼输入偏好，有些细胞喜欢左眼输入的信号，有些细胞喜欢右眼输入的信号。这说明生成视觉显著点的脑区，就在V1区。这个实验还告诉我们，其实“看”和“见”是分开。

**《环球科学》：**“V1显著图理论”的提出对神经科学领域意味着

## 左右眼实验

研究人员设计了一个行为实验来验证自己的理论的一个预测。首先，请自愿者通过立体眼镜，左右眼分别看两幅不同的图——左眼看到的图上，几乎所有的棒都向左倾，只有一根棒向右倾，右眼看到的是一张空白图，要求自愿者尽快找到唯一的右倾棒。然后，通过眼动跟踪，记录自愿者看向右倾棒的时间。接下来，研究人员移动左图上的一个左倾棒，到右图的相应位置，得到两张新图。再请自愿者戴上立体眼镜，左右眼分别看两张新图，找出朝向不同的棒。

结果发现，尽管两种情况下，视觉认知最终看到的图像是一样的（都是左眼输入图和右眼输入图的叠加），但相比第一种情况，第二种情况下，自愿者需要花费更多时间，才能找到那根右倾棒。通过眼动跟踪，我们发现，即使明确要求自愿者尽快找到不同朝向的棒，自愿者还是会不由自主地看向跟任务无关的、只有右眼可见的那根唯一的棒。



什么？它对普通大众的生活会产生影响吗？

**李兆平：**长期以来，科学家对V2区的研究进展不如预期的那么好，可能就在于对V1区了解还不足。我们现在知道，视觉信息被V1区作用而受到了过滤，有些信息被收录，有些信息被删除。如果信息在进入V2区前被大量筛除，我们还按照传统的方法，去测量V2区的感受野，经常碰壁也就不足为奇了。接下来，“V1显著图理论”也许可以指导对V2及以上视觉皮层的研究。

视觉注意力分配，可以用到很多生活场合，比如，在广告和电影画面中，如何引导注意力。注意力分两种，一种自上而下，叫有意识注意；另一种自下而上，叫无意识注意，我们的理论主要可以解释无意识注意。在传统电影中，导演一般利用摄影机视角、灯光、选择性聚焦乃至音效等诸多手段，来引导观众的注意力。但如果电影工作者借鉴我们的理论，通过设计一些唯一的特征——这些唯一的特征在VR（虚拟）环境下还有更多的设计途径，引导观众的无意识注意，或许可以不露痕迹地让观众的眼睛向某一方看。

2016年7月10日发表于《环球科学》  
（科学美国人中文版）  
微信公众号“科研圈”



环球科学 SCIENTIFIC AMERICAN