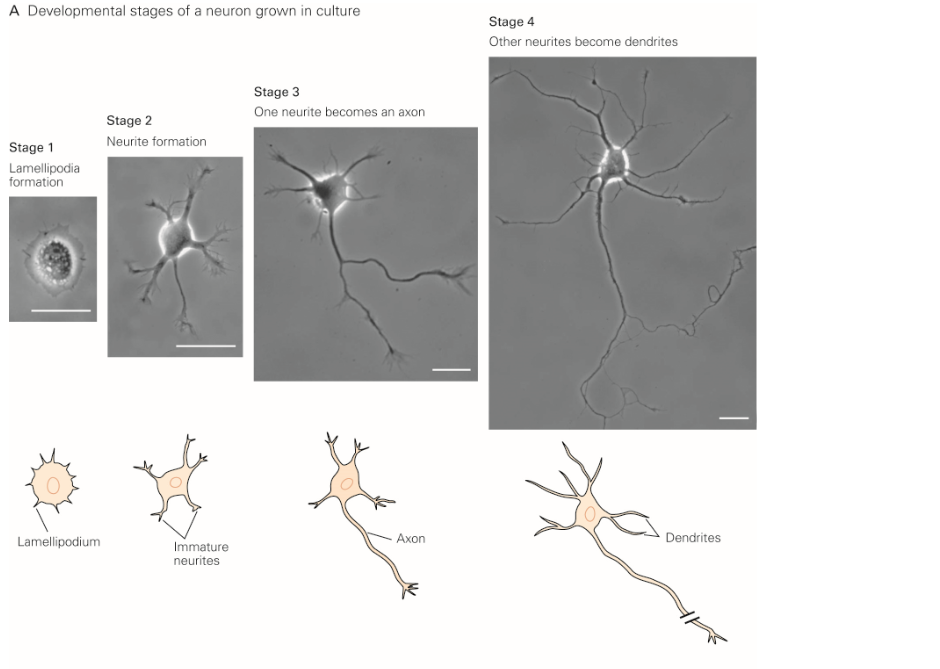
智力的本质和稀缺的专注力

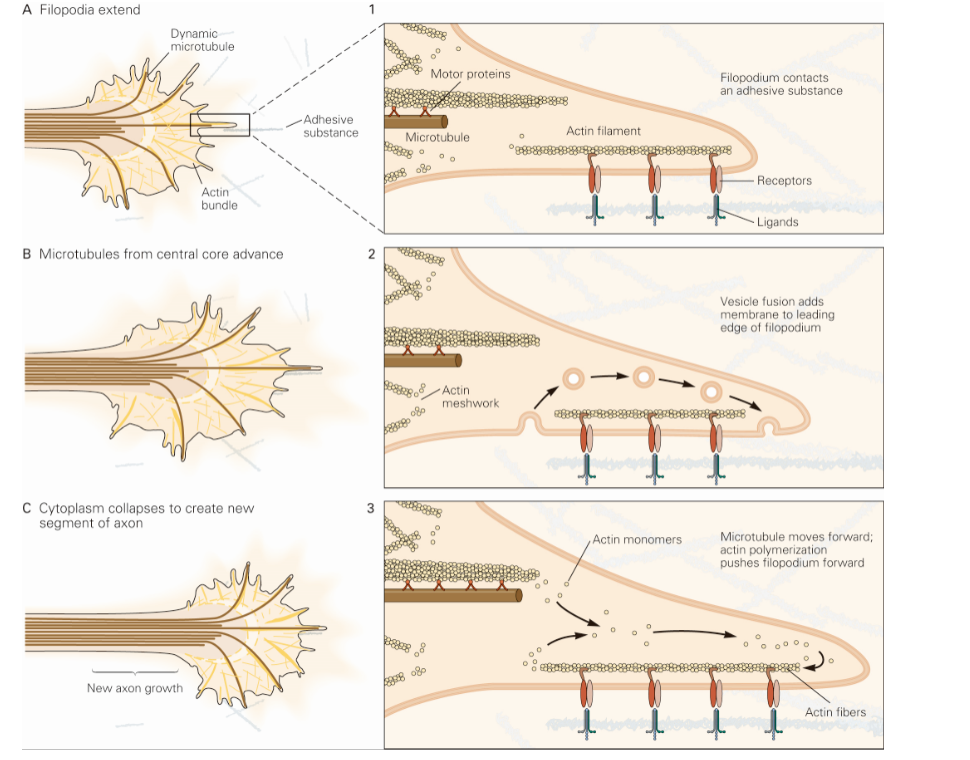
上次还是有同事说我写的东西有点太复杂了，我尽量不用专用词汇，来看看我们是否能构造一个简单的能让机器理解人类语言的模型。要让机器理解人类的语言，那应该先搞清楚人类是怎么认知，思考和处理语言的。为了描述简单，我们只涉及视觉，听觉，认知，工作记忆，持久记忆和能感知感觉的镜像神经系统。如果说要理解这些内容是需要初中的物理知识，高中的一些生物知识和本科计算机里面的数字逻辑等知识，并且您也不打算学这些，我建议您就不需要往下看了。

为了构造一个稍微健壮的模型，构造它的基本元素就需要尽量简单并且可以被重复实验来验证，里面会没有复杂的数学公式也不会出现让人琢磨不透的文字，当然也有可能是您想多了。

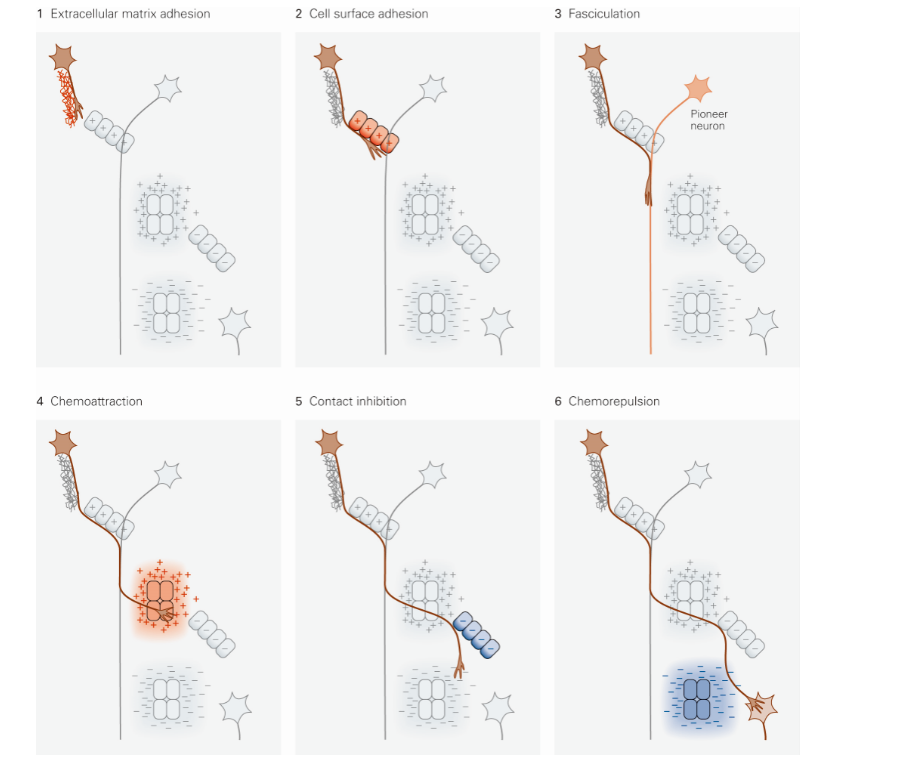
先讲讲神经元生长吧，因为理解了这个点，后面就会比较容易理解其他的点。大多数人都能想像豆芽是怎么生长的吧，如果不知道，其他植物也行。我们的神经科学家也在神经细胞培养组织里面为我们准备了这样几组神经细胞生长的照片。这个好像跟想像中的豆芽生长不太像啊。对，细心的同学应该想到了植物是靠生长激素生长的，所以它会一直朝产生生长激素的组织的方向生长。神经元生长的时候会先向多个方向长出还不成熟的触角。这个跟植物不太一样，神经元是单个细胞，细胞骨架蛋白维持了这些触角的生长，如果某个触角上的机动蛋白微丝失去了平衡，细胞骨架就会安排这个触角成为一个能传递信号的触手，我们叫它轴突，其他触角则会变成接收信号的触手，我们叫它树突。



当神经元的一个触角选定为轴突并准备向目标对象生长的时候，轴突的顶端会生成一个特殊的结构，我们叫它生长锥。生长锥上有多个伪足，当伪足遇到了刺激信号，比如说遇到了粘合物质，生长锥就会延着粘合物质生长。为了不引入太多专业名词，可以通过我们的神经科学家为我们精心绘制的图片来了解这个过程。



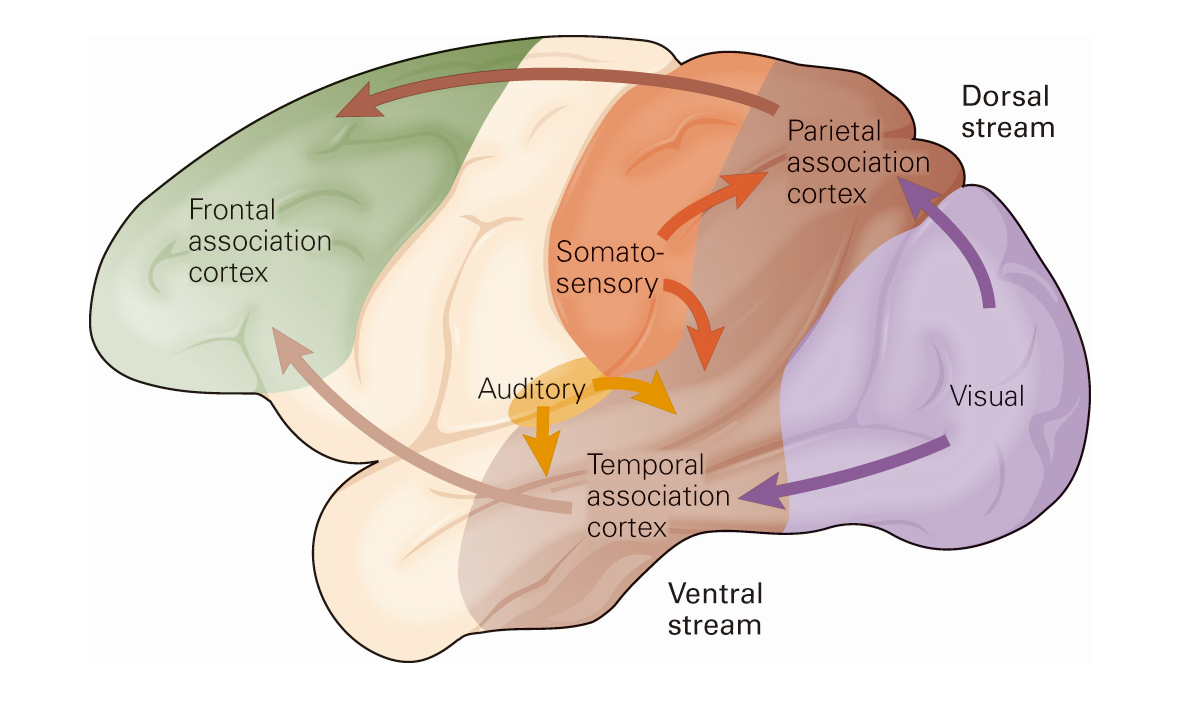
看到这里是不是已经头晕了，这跟人类的语言，认知和思考没有半点关系吧。对，我讲的这些东西的目的不是想把大家侃晕，相反，我倒是想大家可以稍微保持一点点注意力能继续看完后面的介绍。因为后面马上就会讲到一个很神奇的故事，一百年前的神经科学家把它比作是史诗级的爱情故事。我觉得那位科学家是有点太激动了，我们还是用一颗平常心来看待这个事情吧。其实我前文也提到过神经元存在的目的是什么，它需要把信号传递给下一个目标对象。但很显然它不能凭空把信号传递给目标对象，那怎么办呢？在茫茫的生物组织里面，哪一个会是神经元的目标对象？这个时候“神迹”出现了，它会引导神经轴突向它的目标对象生长直至轻触目标对象。这看起来很神奇，但其实有用到轴突生长锥的特性和生物组织的蛋白质分布。在胚胎的时候，上皮细胞会分化成神经元细胞并通过蛋白质引导到目标对象上，我们称它为先锋神经元。一旦完成了这个从无到有的质变连接后，后面的其他神经元就会顺着先锋神经元的轴突生长到目标对象上。在生物生长的过程中这些轴突粘连在一起形成神经束并伴随生物生长一起变化。我们的神经科学家为我们整理了下面六种引导方式：1.细胞外基质的生长化学分子引导轴突生长。2.其他神经元细胞表面引导轴突生长。3.先锋细胞轴突引导轴突生长。4.细胞分泌的可溶解化学分子信号引导轴突生长。5.调节对象细胞表面表达排斥信号使得轴突转向。6. 细胞分泌的可溶解化学分子信号排斥轴突生长并使得其转向。



了解到这里，我也想要把你们引导到我的目标对象上了。不过在这之前，我还是想让大家适当休息一下，聊一些轻松一点的话题。比如说这个从0到1的过程，我不知道我小时候是怎么就学会说话了，我妈说我学会说话要比正常的小伙伴要晚。好吧，虽然知道这个过程挺漫长，不过起码还是学会了说话。再到后来就是按照这个社会原本的生活轨迹成长，和发小爬树抓知了，上小学上初中，之后考试上高中上大学然后工作。这似乎已经到达了我们的人生目标。但不知道什么原因，就我个人而言，似乎有一种 “神迹” 引导到我去寻找那么一两个 “爱好” 。这些爱好我曾经想放弃过，比如说人是怎么思考的，因为那时候没有太多理论或者实验能让我搞清楚这些事情。现在神经科学家已经做出了很多的探索，结合现在的心理学和计算机学科，我打算尝试构造一个比较简单的人类思考模型，为了验证这个模型的有效性就是让机器也能理解人类语言。另一种辅助性的验证方法就是大家能从这个模型中更加清楚自己是怎么思考的，并能从里面获得益处。

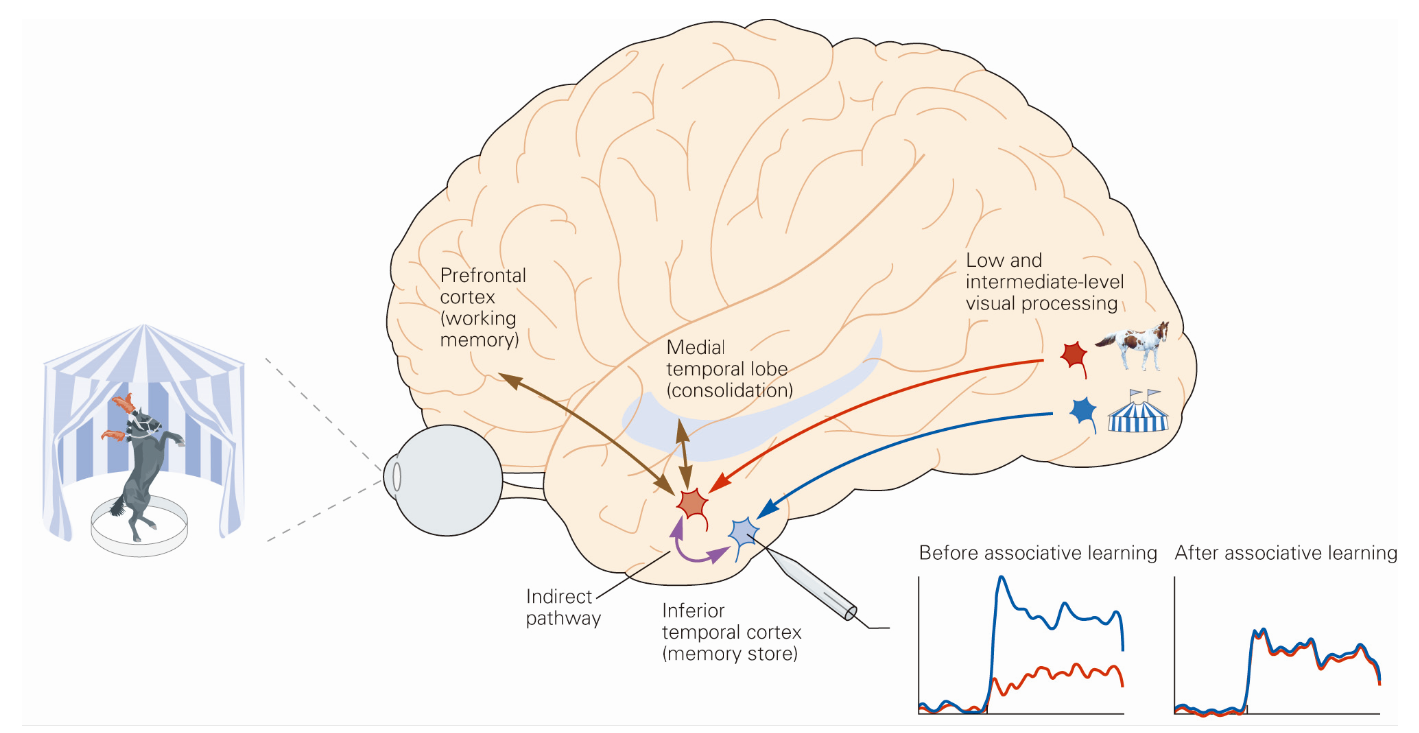
大家是不是有过这样的感觉，当我们非常愉悦跟别人交流的时候，对方说香蕉，我们是不是会在脑子里面回想起这样一个画面，里面有一个金黄色的香蕉。但当对方跟我们说pisang的时候，我估计大部分人都会愣住几秒然后也没认出来。这是怎么发生的呢？从神经科学的角度看，它们涉及到一个认知的过程，因为我们学中文的时候有人告诉过我们香蕉是什么意思，但是突然听到pisang这个词的时候，没有人告诉我这代表着什么，所以我们也就没有办法认出它来。但如果我跟你说pisang是马来语香蕉的意思，你是不是会呈现一个香蕉的画面并且学会了这个单词。

从心理学角度来讲，听到香蕉后会在大脑里面呈现一个画面其实是一个催眠的过程。这在大脑里面是怎么发生的呢？一般人都有视觉，听觉，体感，嗅觉，味觉五种输入型感觉，而视觉处理在大脑皮层中占据了1/3的区域，其次是体感和听觉。下面是神经科学家为我们制作的处理视觉，听觉在大脑皮层分布的区域以及它们的联合区域。为了不引入太多专有名词并只关注在认知思考过程，我们只要了解以下区域就好了，位于后脑勺的视觉处理区域，位于耳朵后上方的听觉处理区域，位于听觉区域和视觉区域间的大脑下方的认知区域，位于前额的工作记忆区域和位于听觉区域上方工作记忆区域后下方的镜像感知与处理区域。



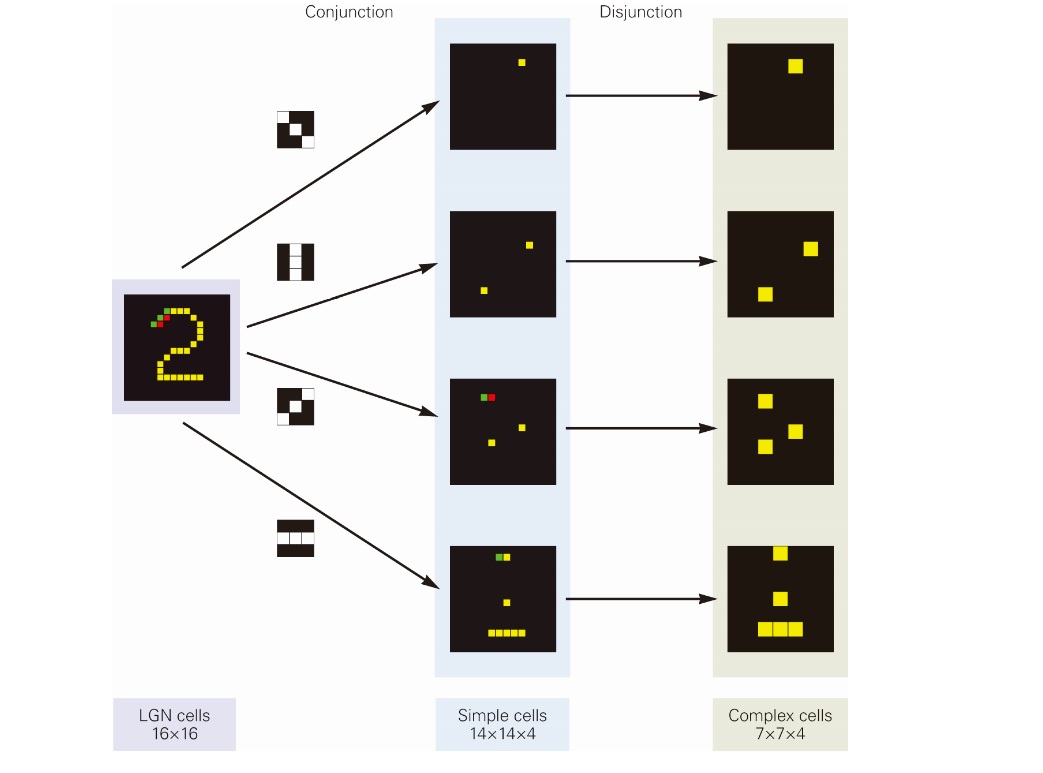
当我们学会香蕉这个名词后，其实它是一个实物，我们通过视觉可以激活它视觉上的特征并将这些特征在原来的认知基础上集合成一个新的神经元群，这一小撮神经元的连接就代表香蕉。那有同学会问了，那么多香蕉，为什么一小撮神经元的连接就能表示它？在刚开始的时候看到香蕉会结合已有的认知神经元连接激活连接成一个新的连接集合，后来我们的大脑里弱化了细节但保留了香蕉的主要特征，当听到香蕉这个词后关联到了对应的认知区域，认知区域会反向把香蕉特征投射到视觉区域和工作记忆区域，所以我们就看到了香蕉的画面。但听到香蕉的时候并不是什么时候都会呈现香蕉的画面，比如说在上英语课的时候，听到香蕉就会回想banana这个单词的声音而忽略了画面的呈现，这个就跟每个人提取认知信息的偏好顺序和当时的神经元活动状态有关。

除了认识新事物会产生新的神经元连接，两个认知对象之间也可以产生新的连接。我们的神经生理学家在大脑认知区域找到了马和帐篷的认知神经元。在开始的时候马的信号不能激活帐篷的认知神经元，但当看到马戏团帐篷里面马的表演后，再输入马的信号能激活帐篷的认知神经元。也就是说我们习得了这种认知关联。从下图可以看到工作记忆也可以激活认知神经元，这也说明除了外感觉输入信号可以关联认知，思考的时候工作记忆也可以关联认知。那其实一个有吸引力的观点是只要认知区域神经细胞被激活，它就有可能和其他细胞按照神经元生长的方式产生新连接，比如关联认知对象的意义，功能和用途等。神经纤维平均的生长速度为每天0.5到5毫米，我们的头发正常生长速度为每天0.27到0.4毫米，如果我们内心比较平静，又比较专注，那其实获取神经元生长的资源会很多，加上认知区域神经纤维已经有一定规模，很大的可能性是它们彼此之间的距离非常短，这种连接的可塑性就会比较强。上面实验提到的只是视觉对象的连接，可以猜想这简单的模型也可以获得抽象的认知，比如在小的时候，如果已经有香蕉的概念，然后再认知了大香蕉是指比别的香蕉要大的香蕉，大苹果是指比别的苹果更大的苹果，慢慢就保留了大的含义弱化或抑制了具体的视觉场景。

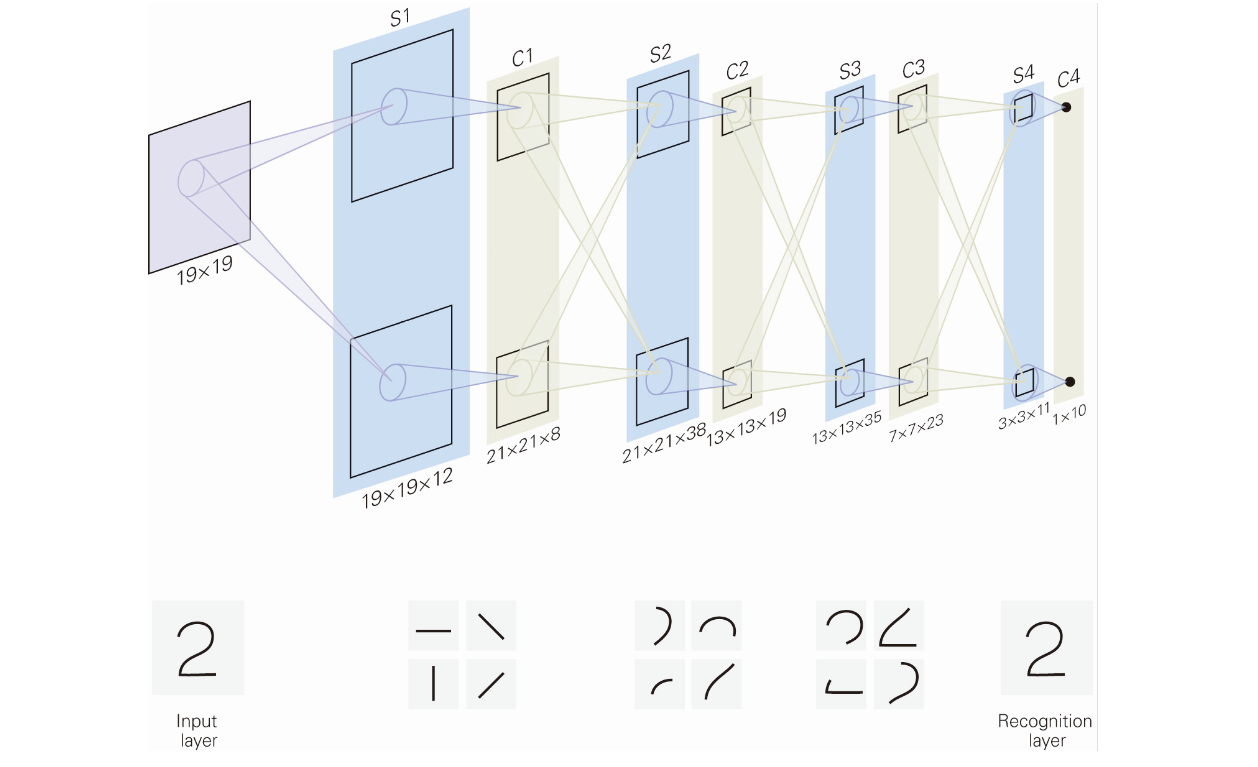


如果要让机器也能理解人类语言除了要和人有一致的认知外，还有一个问题是人是怎么思考的，如何和人交流。这个也是神经科学还比较谨慎的地方，但是有实验证明猴子有一套镜像神经系统，它会在猴子做出动作之前激活，也就是在计划阶段，猴子可以在这个阶段选择做还是不做。听起来和人类的思考模式是一样的。更吸引人的是猴子在观察其他猴子或者人类在计划做这些动作的时候，比如先看一下杯子，也会激活猴子的镜像神经系统，等动作完成后猴子的镜像神经系统也停止了放电，这说明猴子通过镜像神经系统可以判断其他个体的意图。也就是说镜像系统可以基于经验让我们理解对方的动作或者意图。这个实验也给出了一个方案可以根据已有的经验去理解对方的意图从而完成交流。现代的心理学刚好可以补上这些心理活动的经验，假设这个机器人是一位智者，他拥有我们的认知能力并且和我们的认知一致，同时也在认知里面混有人类心理活动的经验，那跟他交流的时候，他会知道我们的意图从而帮助我们。那构建这样一套认知系统是否可行呢？现代基于神经网络的人工智能已经可以识别文字，图像和声音，可以下围棋打星际甚至开汽车，但它们都不会像人一样思考，那差距在哪里呢？我们对比一下利用神经网络的视觉处理，里面会涉及到一些专业名词，不想看的同学可以跳过。

先看一个简单的计算机认知机的实现，它模拟了大脑皮层的中级视觉处理，里面神经元是按层级结构连接，前面的一层会连到后面的一层，每一层有多个神经元执行类似的操作。下图是识别两个2的示意图，黄色是两个2共有的部分，红色是其中一个2的部分，绿色是另一个2的部分，在从输入层到第一层的简单神经元上，先对它们进行位置上的方向识别，也就是“\“，”|“，”/“，“—”四个方向上的识别，如果每一小块都满足条件的话，就激活下一层的对应的神经元，所以我们会看到这两个2在位置处理的时候“\”和“|”方向上没有什么区别，在“/”和“—“方向上黄色神经元被两个2激活，红色神经元被一个2激活，绿色被另一个2激活。然后第一层的两个相邻的简单神经元连接到第一层的一个复杂神经元上，如果其中任何一个神经元激活，那个复杂神经元就会激活。这样看似有微小差异的输入在最终结果反映是一样的。



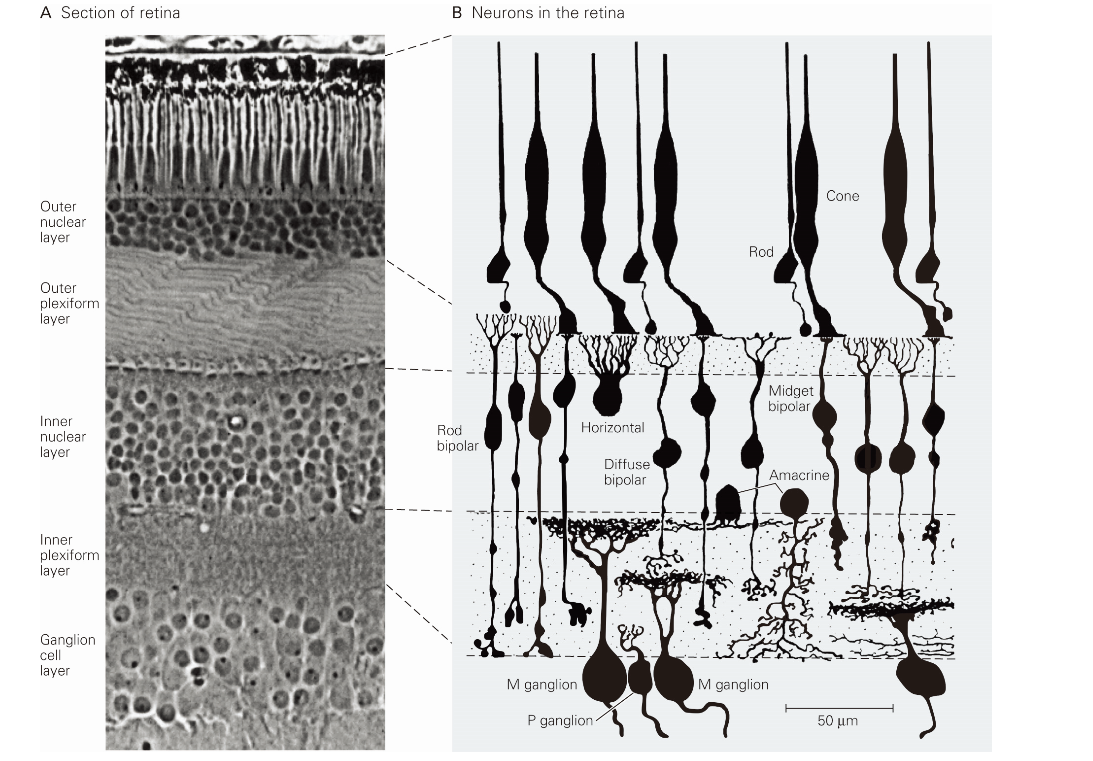
我们再扩展一下加到4层神经网络，每一层都有简单神经元和复杂神经元，利用神经网络的优势是能很快计算出结果。在理论上说它每一层是串行的，每一层里面每一个神经元运算是并行的，它的设计也是来源于大脑的视觉中级处理，这样它可以在有限的时间得出结果。比如说下图是一个处理手写数字的认知模型，在第一层处理方位信息并在复杂神经元上保持它们相对的空间稳定性，第二层处理这些这些方位信息连接处的特征，在第三层处理更复杂的特征，比如下面输入层的2的处理。最终在第4层上有10个复杂神经元代表数字0到9.



虽然这个模型借鉴了大脑视觉处理的原理，但是错误率比较高，后来科学家在神经元连接上加上了权重，并且对神经网络进行训练从而得到这些权重信息，这个过程我们叫它规则学习。规则学习有时候又分为两种，一种叫监督学习，另一种叫无监督学习。监督学习有一个额外的“老师“来评估整个网络的表现并且发出奖励或者错误信号来修改权重信息，其中最流行的学习方法就是反向传播，当输出得到一个错误信号后，反向传播从输出层按照神经元连接相反的方向去调节连接权重直到输入层。无监督学习是靠自学的方式去调整权重，比如说调整一个权重保持其他权重不变。在某些特殊领域，反向传播学习对工程师比较有用，比如用这个方法开发出识别手写数字的商业应用至今还在用，不过还不太清楚生物神经网络是不是存在这种学习机制，接下来我们来了解一下正真的处理视觉的神经网络吧。

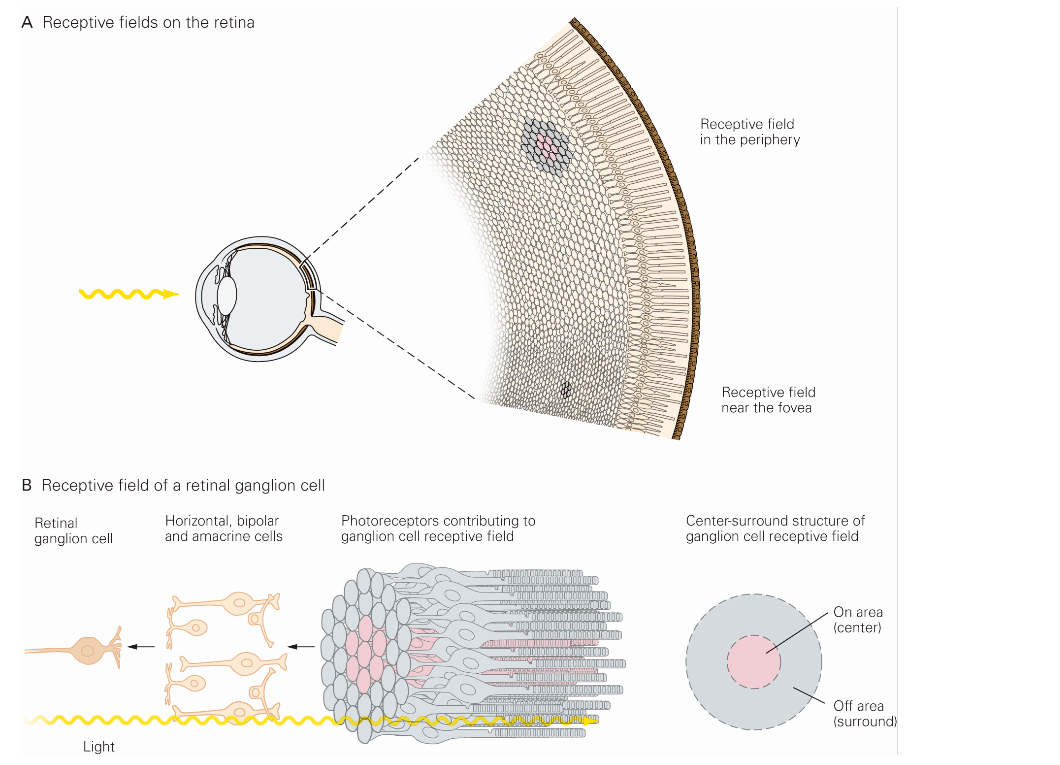
人类视觉系统分低中高三级视觉处理，低级处理在视网膜进行，中级处理在大脑视觉区域进行，高级处理指视觉对认知影响和处理引发的协同动作。

低级视觉处理的主要器官是视网膜，它有三层五类神经细胞构成，我们的神经科学家们用光学显微镜拍出了下面的视网膜分层照片，最内层由光感受器长成锥形的视锥细胞和光感受器长成杆状的视状细胞构成，中间层由双极细胞和无长突神经细胞构成，外层由神经节细胞构成。



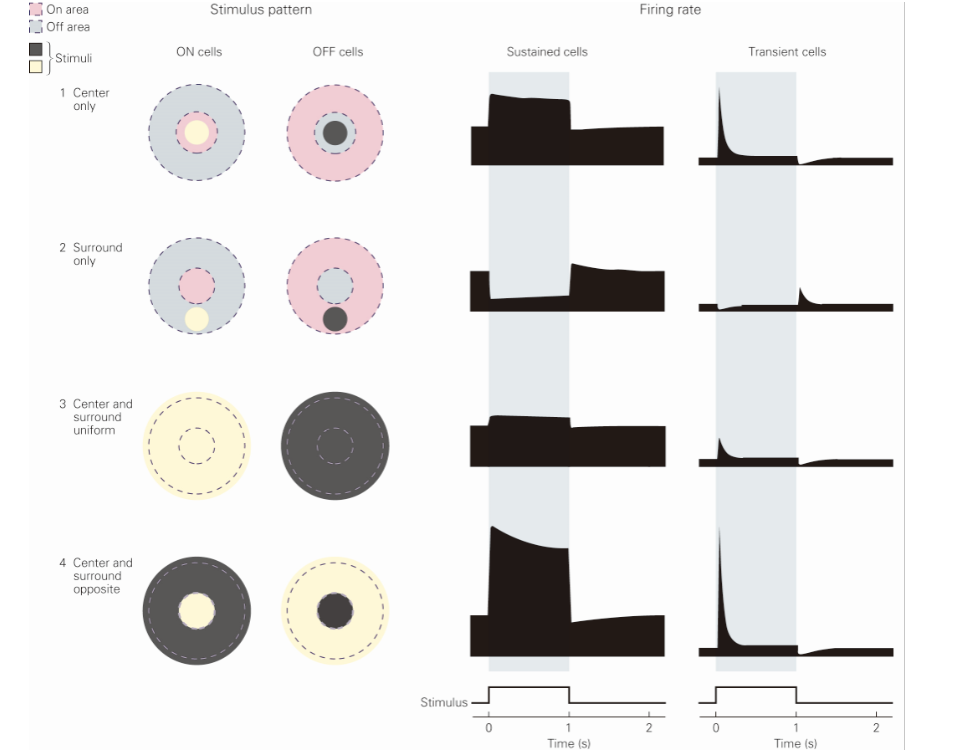
里面涉及到比较多的专业细节，为了让大家比较容易懂，我们关注在几个相关的视网膜特性上。视网膜最内层是接收光子并转换为神经信号的区域，人的视网膜里面大概包含1亿个能感受明暗的视杆细胞和6百万个能感受光颜色的视锥细胞。一般人有红绿蓝三种视锥细胞，其中90%的视锥细胞集中在没有中层和外层细胞遮挡的大约1平方毫米的小凹处，其他视锥细胞会和视杆细胞分布在视网膜其他区域。虽然视杆细胞外面还有中外两层细胞，但光还是能透过它们到达视杆细胞光感受器上，在这些接收区域，神经科学家发现了两类由视杆细胞和其他细胞构成的结构：中心细胞接收光线细胞群放电的结构和中心细胞光线移除细胞群放电的结构。

这好像又出现了专用词汇，还好我们的神经科学家帮我们绘制了这样的第一种结构的图片。它们的特点是最内层有几个光感受细胞处于中心，其他光感受细胞围绕着中心细胞，通过中层的若干双极细胞和无长突神经细胞连接到外层的一个神经节细胞。这种结构的功能是当光点只在中心区域或者围绕区域会激活神经节细胞，光在或者不在整个中心和围绕区域都不会激活神经节细胞。有些小伙伴可能比较难理解了，如果学过大学里数字逻辑课里面的异或门就比较好理解这些了，它的原理是有两个输入和一个输出，只有当一个输入有信号的时候才有输出信号，如果两个输入都有信号或者都没信号不会有信号输出。如果小伙伴还没理解也没关系，后面我们会详细说到。



我们的神经科学家用实验测试了刚才提到的两种结构。下图第一列圆圈表示我们提到的第一类中心接收光线细胞群放电结构的示意图，第二列圆圈表示提到的第二类中心光线移除细胞群放电结构的示意图，第三列是第一类结构对应条件下神经节细胞兴奋的时间，第四列是第二类结构对应条件下神经节细胞的兴奋时间。刚开始的时候两组结构都接收到灰色的光。在第一行我们点亮第一类结构的中心区域，神经节细胞持续兴奋了一段时间，我们移除第二类结构中心区域的灰光，神经节也兴奋了一段较短时间。在第二行里面，我们点亮第一类结构的围绕区域，神经节细胞反应被抑制了，移除第二类结构围绕区域的部分灰光，神经节细胞反应也被抑制。在第三行里面，我们给第一类结构的中心和围绕部分都打上亮光，神经节细胞有反应但是幅度很小，我们移除第二类结构中心和围绕区域的灰光，神经节的反应幅度也很小。但在第四行中，我们给第一类结构中心打上亮光，围绕区域移除灰光，神经节反应强烈，我们再给第二类结构中心移除灰光，围绕区域打上亮光，神经节反应也很强烈。所以当视网膜的外围接收区域和靠近小凹的接收区域填充这些结构后，第一类结构会对边界和轮廓特别敏感，第二类结构对移动物体会比较敏感。

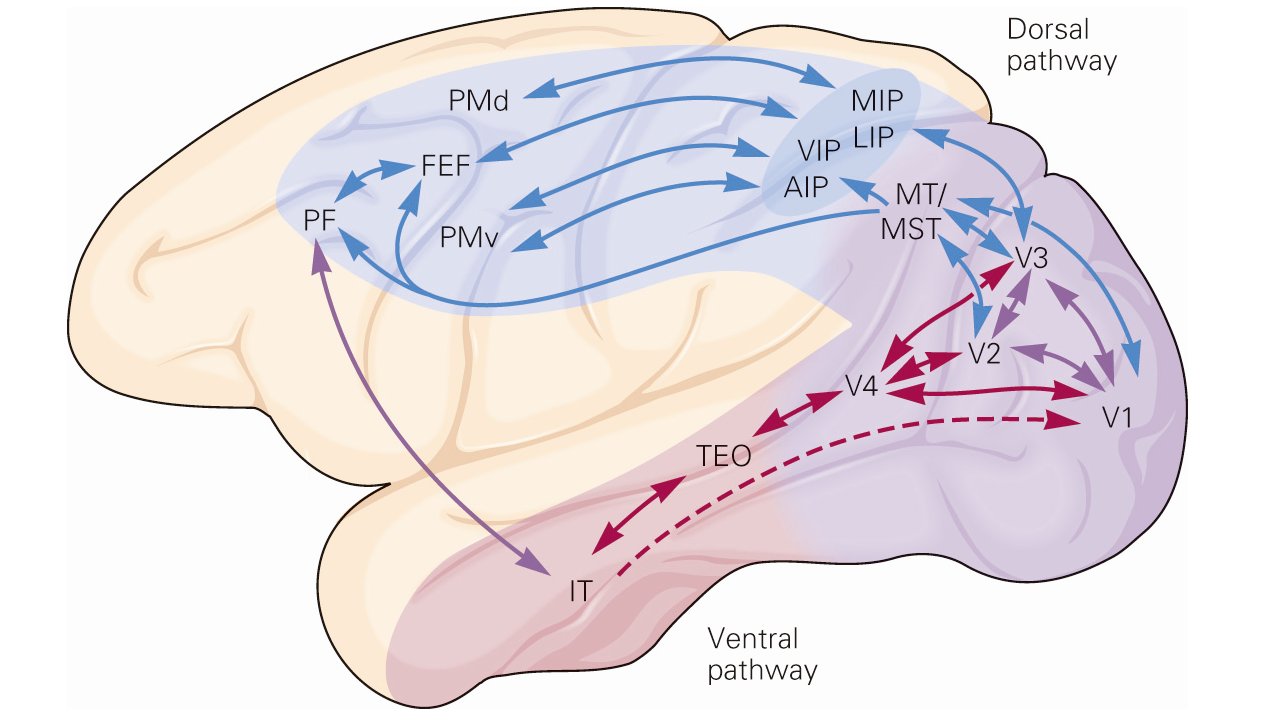
在这里我感觉找到了某些证据来证明我们古人留给我们的一些忠告，比如古时候遇到老虎或者其他猛禽的时候千万不要乱动，找一个下风口并且周围环境复杂的地方呆着别动，也行还可能不被发现，可一旦跑的话，它本能就会扑上来。开玩笑了，其实当我们用余光扫视周围环境的时候确实会对物体的轮廓和运动比较敏感，在后面提到的视觉处理与行动中会起引导作用。还有些有趣的事情。当我们遇到危险的时候有时候会本能呆着不动，难道是因为这个原因我们的祖先才在原始社会存活了下来吗？另外动物是我们的朋友，当了解了它们后，会发现能很好和它们相处。



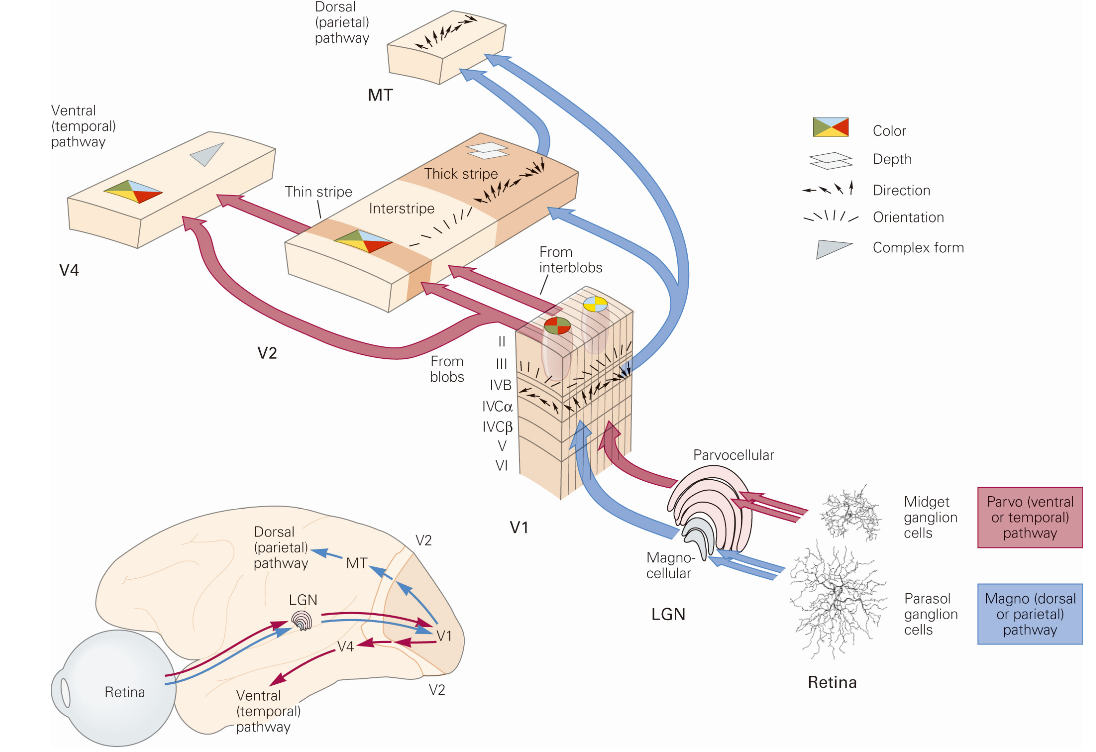
这里说一段题外话，如果不是计算机从业人员建议直接跳过这一段吧。记得几年前有一个同事在交大读研究生，他的算法课大作业两个星期没完成，让我帮忙看一下，我周六的时候花了一个上午帮他写了两个算法。题目是求一个城市的轮廓，第一个算法中规中矩，在第二个算法的时候我投机取巧用了一个桶排序，从理论上把这个题目的时间复杂度降到O(n)。我觉得这个求轮廓的算法时间复杂度已经不能再降了。但后来发现人类大脑处理这些问题都是O(1)。像上面提到的用第一类结构求轮廓那就是O (1) 的时间复杂度。只要看到这个物体立马就能提取它的轮廓。

接下来又是毁我们三观的时刻，如果感觉吃力的话，可以适当休息一下。

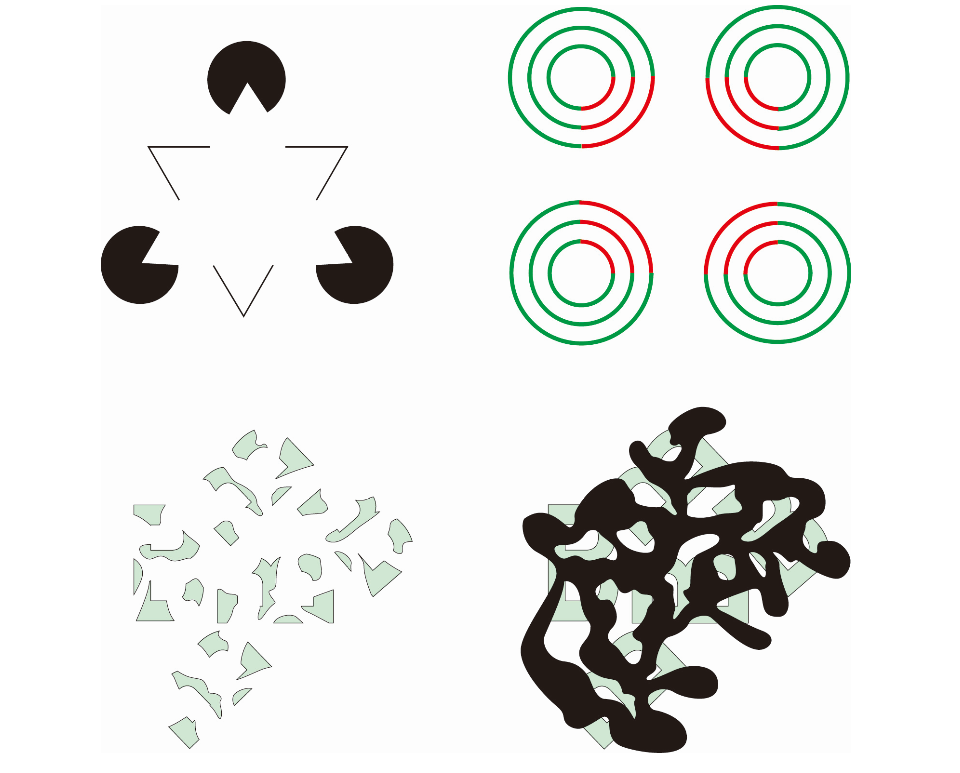
希望看到这里的同学都能够有比较好的状态，刚才我们提到的两类结构它们连接的是不同的神经节细胞，一个是感知边界和轮廓，另一个是感知移动物体。我们神经科学家发现这样类似的神经节细胞大概有20种，它们有的感知空间粗细，有的对光谱进行过滤，最终有100万的这些神经节细胞经过丘脑后接入大脑皮层。这个时候我们到达了视觉中级处理层，上面提到用神经网络做成的应用来识别手写数字，它的模型来源就是视觉中级处理层，比如说视觉区域分四个区域：第一视觉区域V1，第二视觉区域V2，第三视觉区域V3和第四视觉区域V4，第一视觉区域会呈现很多带方向的短线碎片，然后提供这些信息来处理物体轮廓。



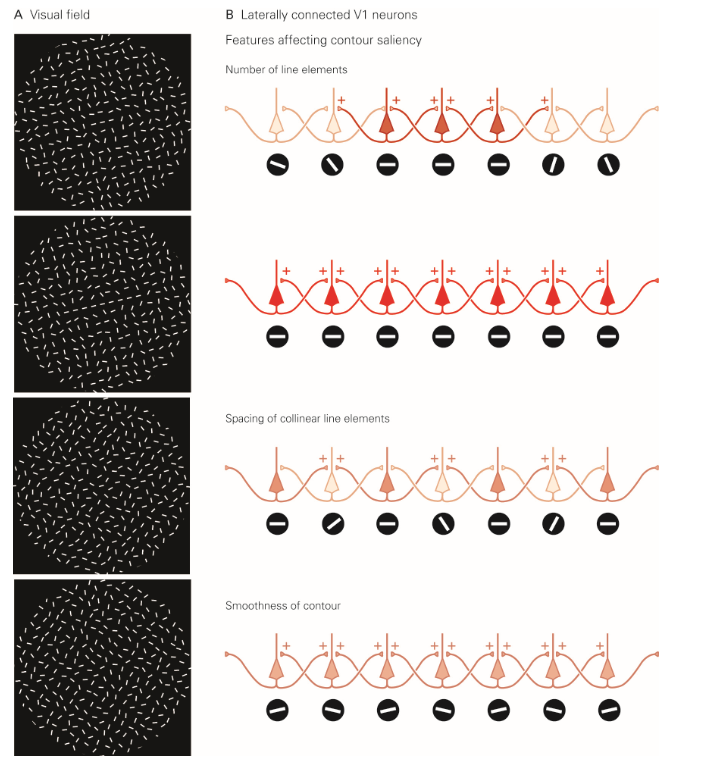
如果我们细看的话，第一视觉区域会并行处理这些信息，比如把大脑皮层的第四层第一节的有关短线段的颜色，形状方向，移动方向和深度信息传递给第二视觉区域的不同位置。



因为我们是通过局部的线段方向和对比来构造物体的表面和轮廓，这导致我们能构造不在视觉区域的轮廓和表面。比如说下左上图可以看到一个白色的三角形，右上图可以看到粉红色的正方形。在左下图，我们可以看到不规则的且没有关联的碎片，但当散乱的碎片之间覆盖一些黑色区域后，我们在右下图可以比较容易看到这些碎片代表字母B。



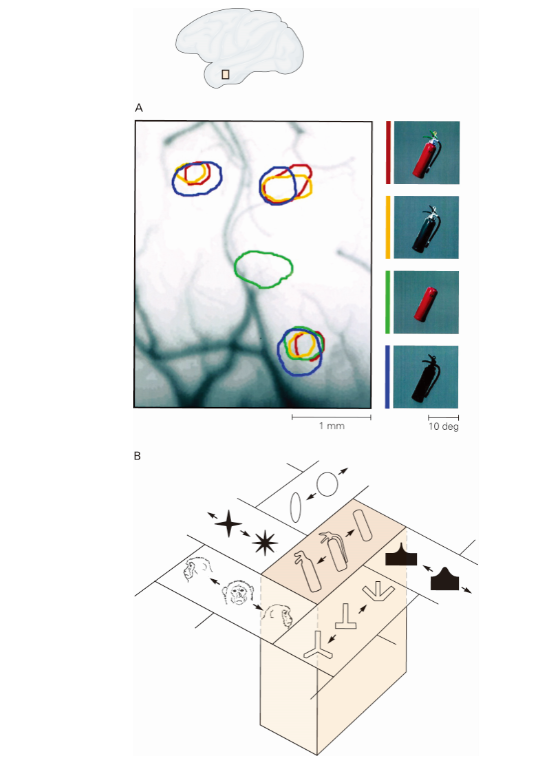
感觉好神奇，现在的人工智能估计很难识别它们出来吧。在视觉处理的过程中大脑皮层帮我们做了很多事情，比如在复杂的背景下识别轮廓。像下面这张图的第一行，我们在第一视觉区域神经元阵列里面会激活三个神经元，而且这些神经元还会把兴奋信号传递给旁边的神经元，所以可以看到五个输出信号。像第二行就明显感觉到轮廓了，第三行中间水平的线段距离较远，而且邻近区域穿插了其他方向的线段，也能感觉出来，就是反映没有那么强烈。第四行中间没有水平的线段，但是它们接近水平方向，而且彼此间距离较近没有太多其他方向反应强烈的线段，它们之间的角度也比较平滑，我们能有一定反应。



我们在认知事物的时候，除了能快速感知它们的轮廓外，还有一些加速我们认知过程的方法，比如说我们在看到事物一部分的时候或者看不太清楚的时候，会快速脑补其他的部位，然后如果有需要会通过已有认知事物的特征去检查这个事物。比如说我们检查视力时候看的视力表，这里又要提一下毁我们三观的事实。在视网膜上，我们眼睛唯一能看清楚的物体的地方是在一个大约1平方毫米的小凹处，我们注视物体的时候物体反射光线正对小凹的位置，有经验的同学能通过这个特征观察到别的同学在看什么。小凹处是没有视锥细胞的，所以在晚上我们的小凹是不工作，看到的景象也只有灰黑色，这个时候我们看不清物体只能脑补一下是什么。还有一个错觉是在红光下我们看到香蕉是黄色，其实也是我们脑补的，因为我们看到物体的颜色是因为它们反射了对应的光线，而红色是单色光，没办法让香蕉反射黄光，但我们的意识里面它是黄色的。另一个有趣的实验是在视网膜上有一个盲点，在视网膜把神经节纤维传入大脑的时候，有一块区域是没有光接收器的，所以在这个位置也不会接收到镜像。那我们为什么没有察觉呢？在我们双眼看物体的时候，是没有视觉区域刚好掉入两个眼睛的盲点。如果你像验证它可以做一个盲点实验，在一张A4的纸上相隔一个手掌的水平位置画两个稍大一点的圆或十字图形，一定要醒目。然后用右手遮住你的左眼，用右眼水平盯着A4纸上左边的图案，然后慢慢靠近或者远离A4纸，在大概60公分的地方你会发现右眼余光看不到A4纸右边的图案，记得有同事问我盲人看东西是什么感觉，就跟这个一样，不是黑色，就是单纯看不见。如果右边的图案换成竖着的直线，在同一个位置右边原来看不到的地方会脑补出直线。在我们认知的时候，如果有人朝我们走来，他们在我们视网膜呈现的图像越来越大，但是我们激活的认知区域还是同一个地方，所以并不会因为远近的关系我们就认不出他们了。

## 

在认知物体的时候其实已经涉及到了高级视觉处理了，我们对世界的视觉经验都是基于对象中心的，它由很多的视觉特征组成，而且我们通常会把它关联到特别的经历，其他已经记住的对象和其他感官上。比如说看到灭火器我们就想到了消防演习。如果我们看到了不同的灭火器会是什么样？我们的神经科学家做了一个这样的实验，当我们看到第一个红色灭火器的时候，在下图画红圈的地方会被激活，看到黑色灭火器的时候会激活画黄圈区域的神经元，看到红色柱状物的时候会激活绿圈区域神经元，看到黑色灭火器会激活蓝圈区域神经元。这说明我们存储的认知事物是按一定阵列结构存储对象的。



不管灭火器是什么颜色，细节上有什么不同，当我们识别出它来后都会激活一样的认知神经元，这使得我们在处理信息的时候可以忽略细节从而加速我们的活动。我们在一个复杂环境中通常会快速移动我们的身体，头或者用余光使得我们很快找到感兴趣的对象，这个过程中我们会忽略扫视过程的其他物体并且不会引起认知神经元激活。这里面还涉及到一个主动关注和被动关注，当你看这篇文章的时候就是主动关注，当这个时候突然有一束闪光，你的注意力可能就被动从这篇文章上离开了。