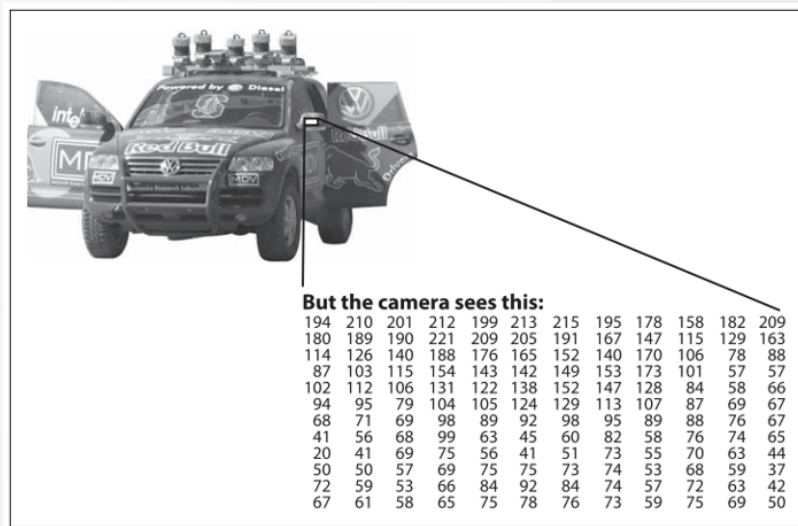


神经网络与深度学习

图像的特征



- 计算机图像是由一定数量的点阵像素构成的。如上所示，我们看到的是一辆车，但实际上计算机了解的是一种由各像素点的灰度值构成的矩阵，它并不能直接了解“这是一辆车”。
- 我们需要将“这是一辆车”这个事实用完全逻辑化的语言描述出来，让计算机建立一种函数，这个矩阵自变量 x 所相应的成果因变量 y 就是“车”。难度可想而知。



- 我们先看一种简朴的例子。一种数字 9，我们怎样描述它的形状？
- 目前有这么一种描述措施：这个数字的上半部分是一种圆，下半部分是靠右的一撇。用数学语言体现就是，上半部分能够以方程 $x_1^2 + y_1^2 = r_1^2$ ($r_1 > 0$) 拟合，下半部分能以 $x_2^2 + y_2^2 = r_2^2$ ($x_2 < 0, y_2 < 0, r_2 > 0$) 拟合，那么就能阐明这个图形极有可能是数字 9。

- **特征**：用来描述一种对象详细体现形式的逻辑语言。
- 如前页所述，“上半部分是圆，下半部分是靠右的一撇”就是数字 9 的**形状特征**。
- 特征是构成一种对象的必要但不充分条件，因为一种对象是由无数个特征构成的，在有限数量的特征里，我们永远只能**预估**该对象，而不能 100% 拟定该对象究竟是什么。

经典的图像特征之——Haar 特征



- 思索：我们一眼就能看出来这幅图像上有一栋房子。但任何一幅图像都是由一定数量的像素点构成的，我们是怎么从这些单纯的像素点里发觉了房子的呢？
- 这个问题还能够换一种问法：我们怎么懂得一幅图像中是有前景对象的，对于一张纯色画布，我们为何无法发觉任何对象？



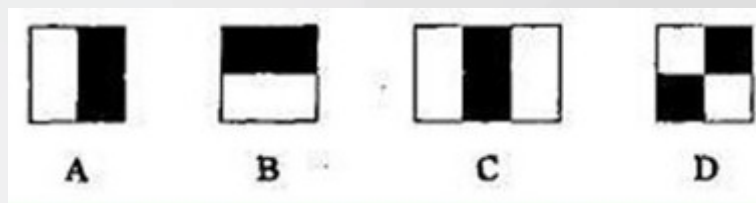
- 这涉及到一种现象：**颜色突变**。上图所圈出的区域中，都是颜色变化较大的区域。而我们就会自然地以为，这是物体的边沿，而由封闭边沿构成的区域就是物体。
- 不但是房子这个整体，房子里的门、窗户等元素我们都能发觉，而原因自然也是**颜色突变**。
- *思索：人能看见绝对透明（100%透明）的玻璃吗？*



○ 我们定义一种形状的时候，本质上就是在定义其产生**颜色突变**的像素点的大致相对位置。例如圆，在一种直角坐标系的图象上上，存在全部满足 $(x-a)^2+(y-b)^2=r^2$ 条件的坐标点的某个小邻域内有较大的**颜色突变**，那么这幅图像上就有一种圆的形状。左上方是一幅色盲测试图，不色盲的朋友都能看到左下角有一种蓝色的圆，而且是一种空心圆。

○ 思索：我们该怎样用颜色突变的方式定义一辆车？（先不考虑前身后身等复杂情况，假设全部的车都是像左下图这么固定角度）

- **Haar 特征**：对大量**同类的**图像样本进行学习，找出这些图像全部的产生了**颜色突变**的相对坐标位置，将这些**同类图片**作为**正样本**，同步给出**不属于该类的，远多于同类图片的负样本**，让机器对这些样本进行学习，并最终身成一种对象辨认模型，用来正确寻找一幅图像中你所需要寻找的元素。所以，Haar 特征广义上来说就是图像的形状特征。
- Haar 特征对于邻域的要求有三种：边沿特征、中心特征、对角线特征，其中边沿特征又分横向和纵向。经过计算白色区域像素亮度和与黑色区域像素亮度和之差来反应图像的颜色突变区域，即边沿区域。尤其地，对于 C 模型，需要将黑色区域乘以 2 作为其像素亮度和。



Haar 特征实例——人脸检测



- 我们用 OpenCV 中的一种训练得很好的人脸检测 Haar 模型，对一张照片进行人脸检测。成果如上。

Haar 特征的不足之处

- Haar 特征本质上是检测图像中的颜色突变的，所以对图像的形状较为敏感，同步对其他的特征（如颜色等）敏感度较低。而且 Haar 只合用于刚性运动物体的检测，而不合用于物体精细化的辨认。人脸检测和人脸辨认难度上完全不是一种数量级的。
- 实际上，任何一种单一图像特征都是不全方面的，是不能适应全部场景的。其实，图像本身只是由像素点构成的，图像的特征那是我们人为整顿出来的，严格地说它并不属于图像的一部分。那我们能不能从图像的像素点本身来寻找某些算法呢？

最基本的决策模型——感知器

- 生活中我们经常要对某些事情做出决策。但不论我们最终选择了什么，都一定是有原因的，而且原因往往不止一种。
- 目前我们来玩一种情景游戏。假如周末企业组织员工外出旅游，你是去还是不去呢？这是一种最简朴的感知器，最终止果只有两种，Yes 和 No。只是，成果虽然简朴，但是你会考虑诸多原因，最终决定成果。
- 假设你考虑下列三个原因：A. 天气怎样；B. 可否带伙伴一起去；C. 费用、伙食、旅馆等条件是否符合个人预期。

- 假设你是这么想的：和同事一块出去玩，是极难得的一件事，能带伙伴一起去那是锦上添花，不能也没关系；但天气是挺主要的，阴雨绵绵玩得不开心；旅馆嘛，这么多人一起，不会太好，但也不会太差，最主要的是享有。那么这时候，你的决策就和本地天气唯一有关，其他两个原因没有任何影响。
- 感知器的本质就是对各个输入量的加权和进行分析，做出 Yes or No 的决策的模型。

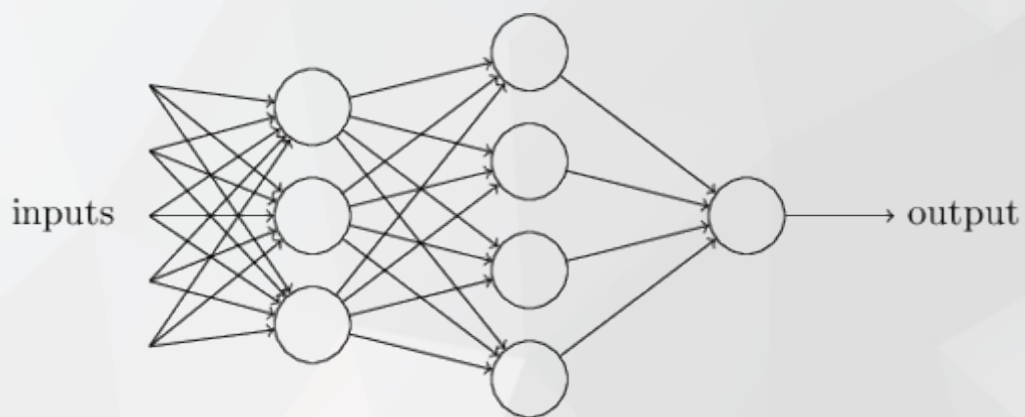


- 上述你考虑的三个原因能够作为感知器的三个输入量 x_1, x_2, x_3 ，并设主动的成果为 1，悲观的成果为 0。三者都有相应的权重 w_1, w_2, w_3 。现计算它们的加权和 $w_1x_1 + w_2x_2 + w_3x_3$ ，同步加上一种偏差值 b 。若其为正（不小于 0），则最终的回答是 Yes，不然（不不小于或等于 0）回答 No。
- 假如你的想法如前所述，那么你的感知器模型能够是这么的： $w_1 = 6, w_2 = 2, w_3 = 2, b = -5$ ，即判断 $6x_1 + 2x_2 + 3x_2 - 5$ 的正负。在全部的输入量均只能取 0 和 1 的前提下，我们能够很明显地看到，唯独当 $x_1 = 1$ 时，成果才为正，即决策为 Yes，只要 $x_1 = 0$ ，决策就必然为 No。
- 思索：若将 b 改为 -3 ，其他权重不变，那么决策会变成什么样？

较为复杂的感知器——多层感知器

- 《非诚勿扰》，一种非常流行的电视节目。男嘉宾上台后，在场的女嘉宾首先给男嘉宾打印象分，然后经过前后播放的多种不同的 AVR 以及多种提问对【自己是否乐意跟男嘉宾走】这个问题给出回答。
- 点评：
- 诸多事情的决策，并不是评估多种彼此独立的条件，往往某个条件是和之前的某些条件有关的。例如上面这个案例，第一印象好坏算作感知器的输入量 x_1 ，只有当它等于 1 时，背面的输入量才有意义。但也不是它等于 1 就一定表达最终的成果是 Yes，还需要分析其他的输入量才干得出最终止果。

- 那么，这种经过多层判断，得出最终成果的感知器，就叫多层感知器。如下图所示，是一个双层感知器，第二层的神经元的决策是跟第一层的决策相关的。



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/558126020065006130>