

Apollo自动驾驶入门课程第④讲 — 感知（上）



本周我们将介绍感知，了解车辆如何利用感知元件感知周围环境，了解不同的感知任务，例如分类、检测和分割，并学习对感知而言至关重要的卷积神经网络。

在开车时，我们用眼睛来判断速度、车道位置、转弯位置。在无人驾驶车中，情况类似，只不过需要使用静态摄像头和其他传感器来感知环境，使用大量计算机视觉技术。

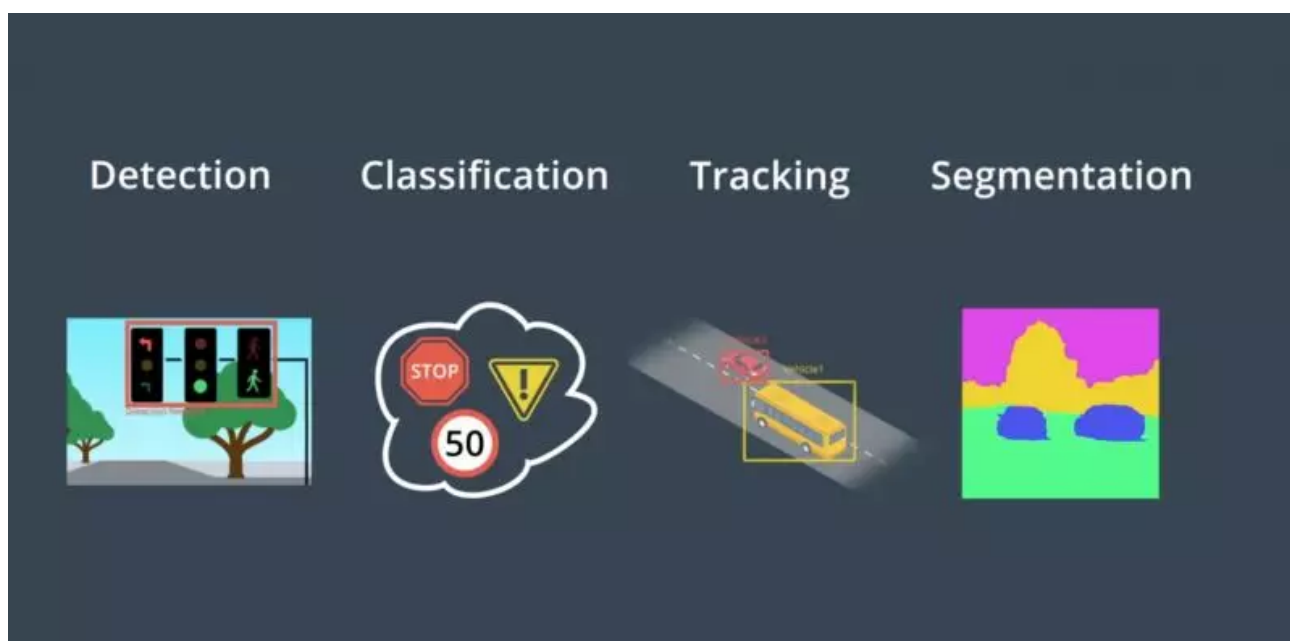
第四课，感知

1 | 感知的概述

我们人类天生就配备多种传感器，眼睛可以看到周围的环境，耳朵可以用来听，鼻子可以用来嗅，也有触觉传感器，甚至还有内部传感器，可以测量肌肉的偏转。通过这些传感器，我们可以感知到我们周围的环境。我们的大脑每分每秒都在进行数据处理，大脑的绝大部分都是用于感知。

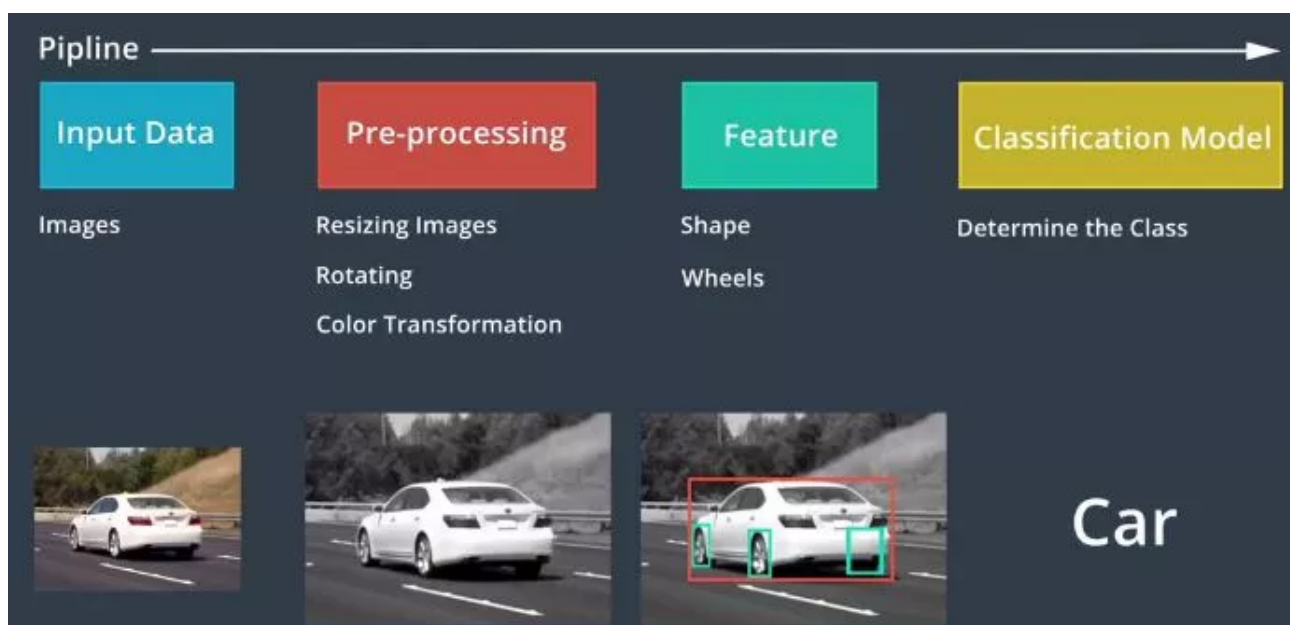
现在，无人驾驶车辆也在做这些事情，只不过他们用的不是眼睛而是摄像头。但是他们也有雷达和激光雷达，它们可以帮忙测量原始距离，可以得到与周围环境物体的距离。对于每个无人驾驶汽车，它的核心竞争力之一是利用海量的传感器数据，来模仿人脑理解这个世界。谈论传感器时也会涉及到神经网络、深度学习、人工智能。

作为人类，我们可以自动识别图像中的物体，甚至可以推断这些物体之间的关系。但是对于计算机而言图像只是红、绿、蓝色值的集合。无人驾驶车有四个感知世界的核心任务：**检测**——指找出物体在环境中的位置；**分类**——指明对象是什么；**跟踪**——指随时间的推移观察移动物体；**语义分割**——将图像中的每个像素与语义类别进行匹配如道路、汽车、天空。



将分类作为研究计算机视觉一般数据流程的例子。图像分类器是一种将图像作为输入，并输出标识该图像的标签的算法，例如交通标志分类器查看停车标志并识别它是停车标志、让路标志、限速标志、其他标志。分类其甚至可以识别行为，比如一个人是在走路还是在跑步。

分类器有很多种，但它们都包含一系列类似的步骤。首先计算机接收类似摄像头等成像设备的输入。然后通过预处理发送每个图像，预处理对每个图像进行了标准化处理，常见的预处理包括调整图像大小、旋转图像、将图像从一个色彩空间转换为另一个色彩空间，比如从全彩到灰度，处理可帮助我们的模型更快地处理和学习图像。接下来，提取特征，特征有助于计算机理解图像，例如将汽车与自行车区分开来的一些特征，汽车通常具有更大的形状并且有四个轮子而不是两个，形状和车轮将是汽车的显著特征。最后这些特征被输入到分类模型中。此步骤使用特征来选择图像类别，例如分类器可以确定图像是否包含汽车、自行车、行人、不包含这样的对象。

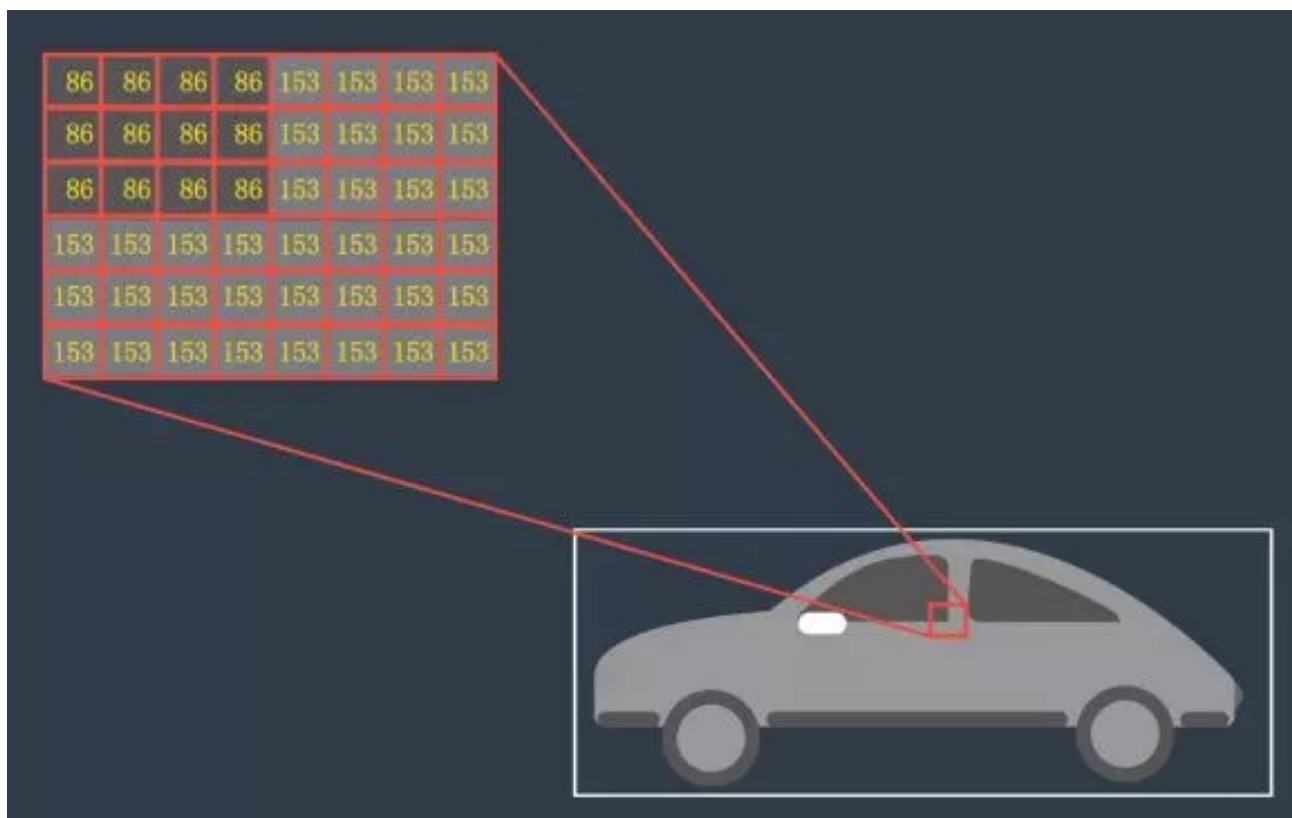


为了完成这些视觉任务，需要建立模型，模型是帮助计算机了解图像内容的工具。

3 | 摄像头图像

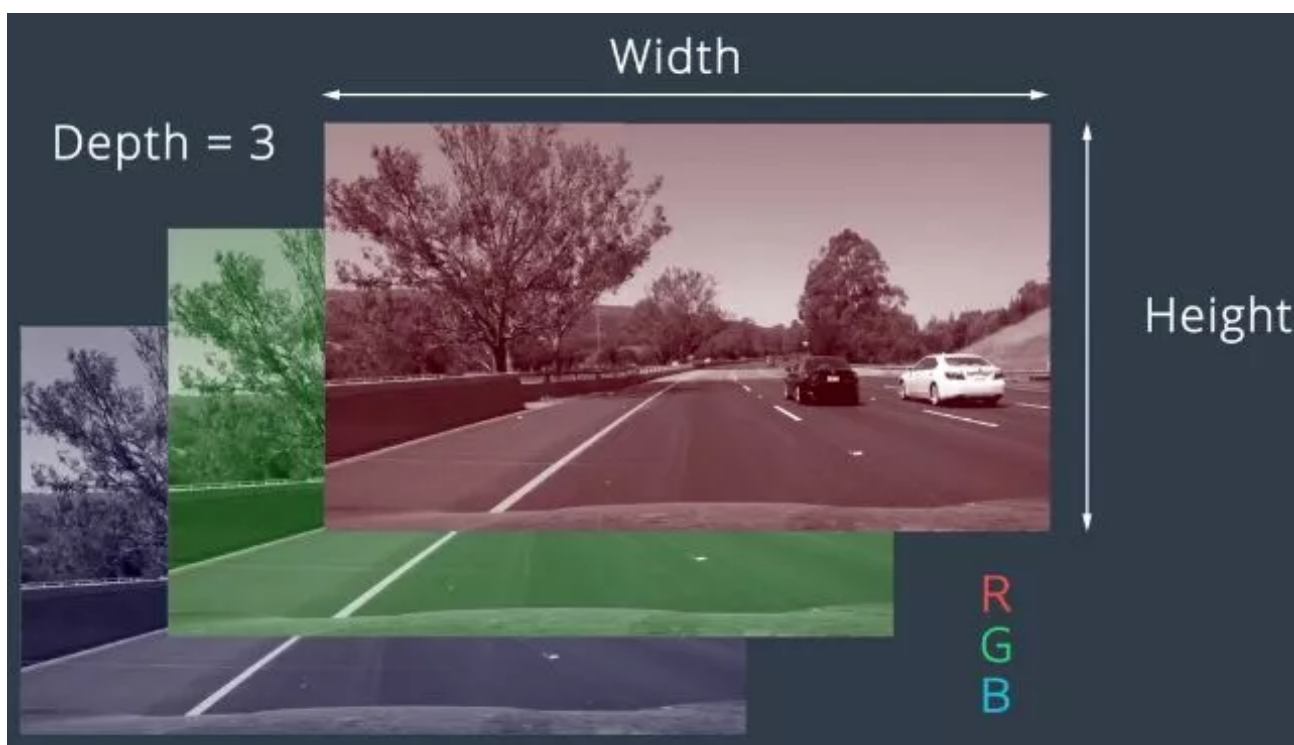
不论计算机在执行什么识别任务，通常在开始时将摄像头图像作为输入。

摄像头图像是最常见的计算机视觉数据，以这张汽车照片为例，让我们看看计算机如何认为这实际上是一辆汽车的图像。从计算机的角度来看，图像只是一个二维网格被称为矩阵，矩阵中的每个单元格都包含一个值，数字图像全部由像素组成，其中包含非常小的颜色或强度单位，我们可以对其中的数字做出非常多的处理。通常这些数字网格是许多图像处理技术的基础，多数颜色和形状转换都只是通过对图像进行数学运算以及逐一像素进行更改来完成。



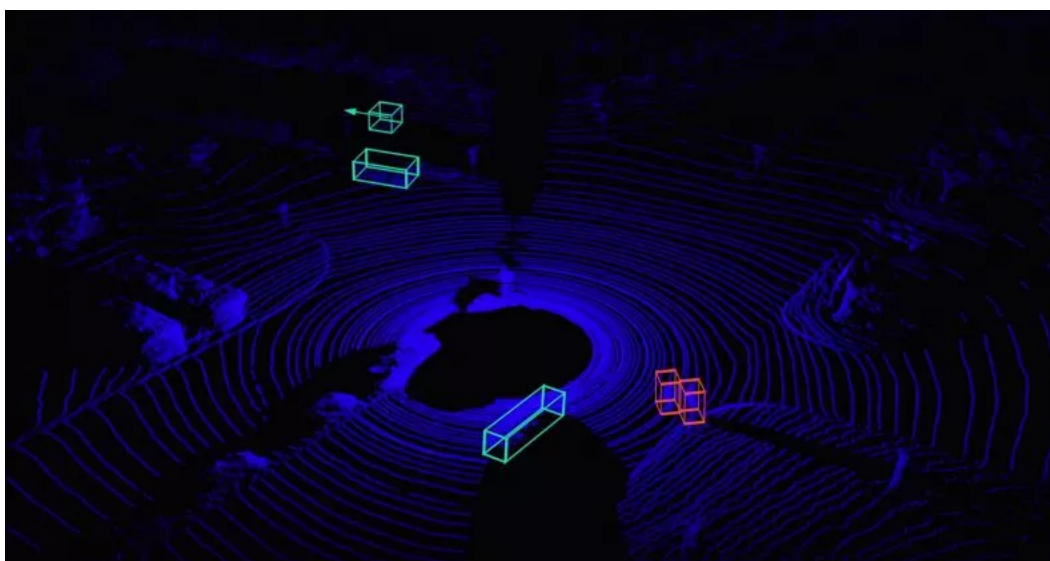
以上是我们将一个图像分解为二维灰度像素值网络，彩色照片是相似的，但是更复杂一些。

彩色图像被构建为值的三维立方体，每个立方体都有高度、宽度和深度，深度为颜色通道数量。大多数彩色图像以三种颜色组合表示红色、绿色、蓝色，称为RGB图像。对于RGB图像来说，深度值是3，因此可用立方体来表示。



4 | LiDAR图像

感知扩展到传感器，而不仅仅是摄像头。激光雷达传感器创建环境的点云表征，提供了难以通过摄像头图像获得的信息如距离和高度。激光雷达传感器使用光线尤其是激光来测量与环境中反射该光线的物体之间的距离，激光雷达发射激光脉冲并测量物体，将每个激光脉冲反射回传感器所花费的时间。反射需要的时间越长，物体离传感器越远，激光雷达正是通过这种方式来构建世界的视觉表征。



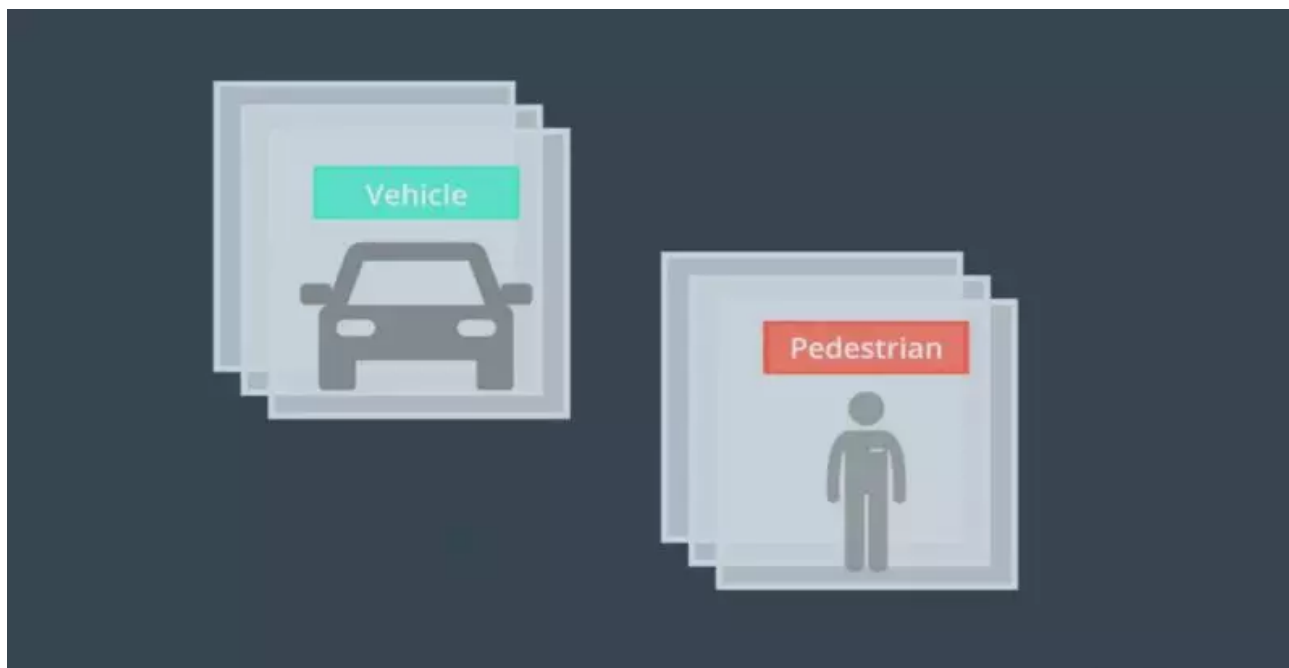
激光雷达通过发射光脉冲来检测汽车周围的环境，蓝色点表示反射激光脉冲的物体，中间的黑色区域是无人驾驶车本身占据的空间。由于激光雷达测量激光反射束，它收集的数据形成一团云或“点云”，点云中的每个点代表反射回传感器的激光束，可以告诉我们关于物体的许多信息例如其形状和表面纹理。这些数据提供了足够的对象检测、跟踪、分类信息。正如我们所看，在点云上执行的检测和分类结果为红点为行人，绿点表示其他汽车。

激光雷达数据提供了用于构建世界视觉表征的足够空间信息，计算机视觉技术不仅可以使用摄像头图像进行对象分类，还可以使用点云和其他类型的空间相关数据进行对象分类。

5 | 机器学习

机器学习是使用特殊算法来训练计算机从数据中学习的计算机科学领域。通常，这种学习结果存放在一种被称为“模型”的数据结构中，有很多种模型，事实上“模型”只是一种可用于理解和预测世界的数据结构。机械学习诞生于20世纪60年代，但随着计算机的改进，在过去的20年中才真正的越来越受欢迎。

机器学习涉及使用数据和相关的真值标记来进行模型训练，例如可能会显示车辆和行人的计算机图像以及告诉计算机哪个是哪个的标签。我们让计算机学习如何最好地区分两类图像，这类机器学习也称为监督式学习，因为模型利用了人类创造的真值标记。



可以假想一个类似的学习过程，但这次使用的是没有真值标记的车辆与行人图像，让计算机自行决定哪些图像相似、哪些图像不同，这被称为无监督学习。不提供真值标记，而是通过分析输入的数据，计算机凭借自行学习找到区别。

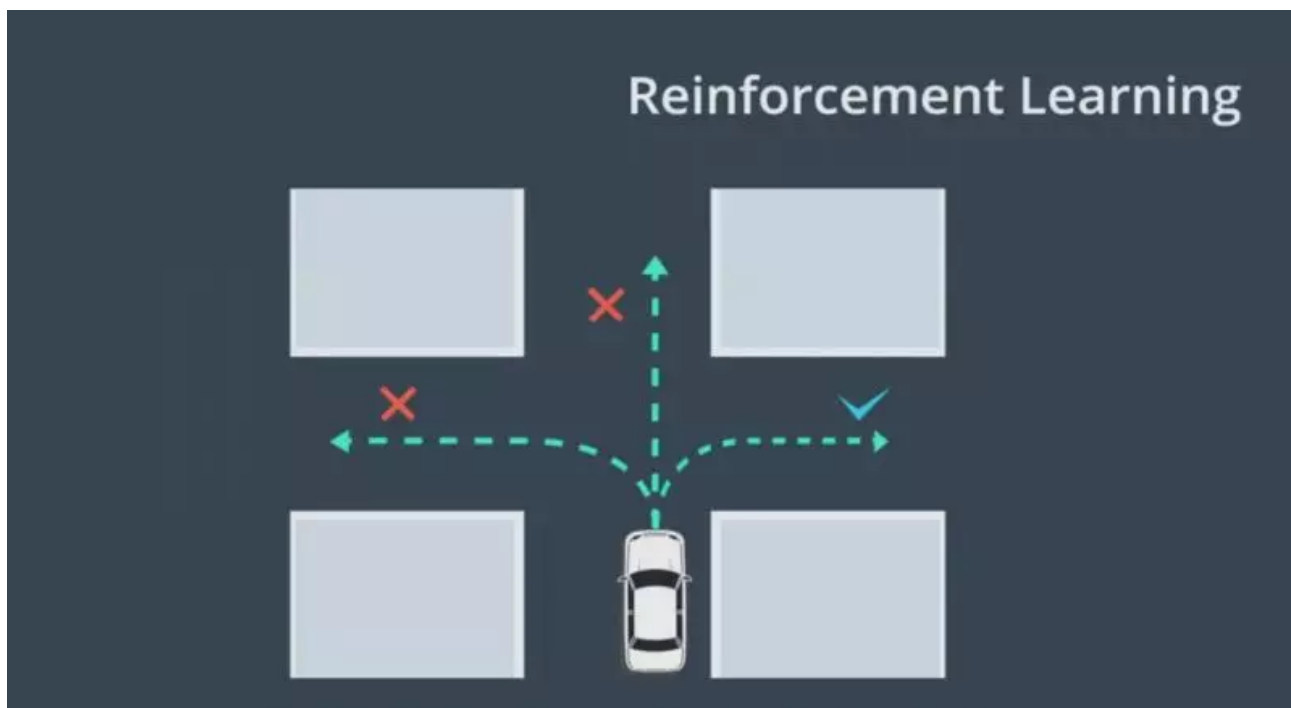
Unsupervised Learning



半监督式学习是将监督学习和无监督学习的特点结合在一起，该方法使用少量的标记数据和大量的未标记数据来训练模型。

强化学习是另一种机器学习，强化学习涉及允许模型通过尝试许多不同的方法来解决这个问题，然后衡量哪种方法最为成功，计算机将尝试许多不同的解决方案，最终使其方法与环境相适应。

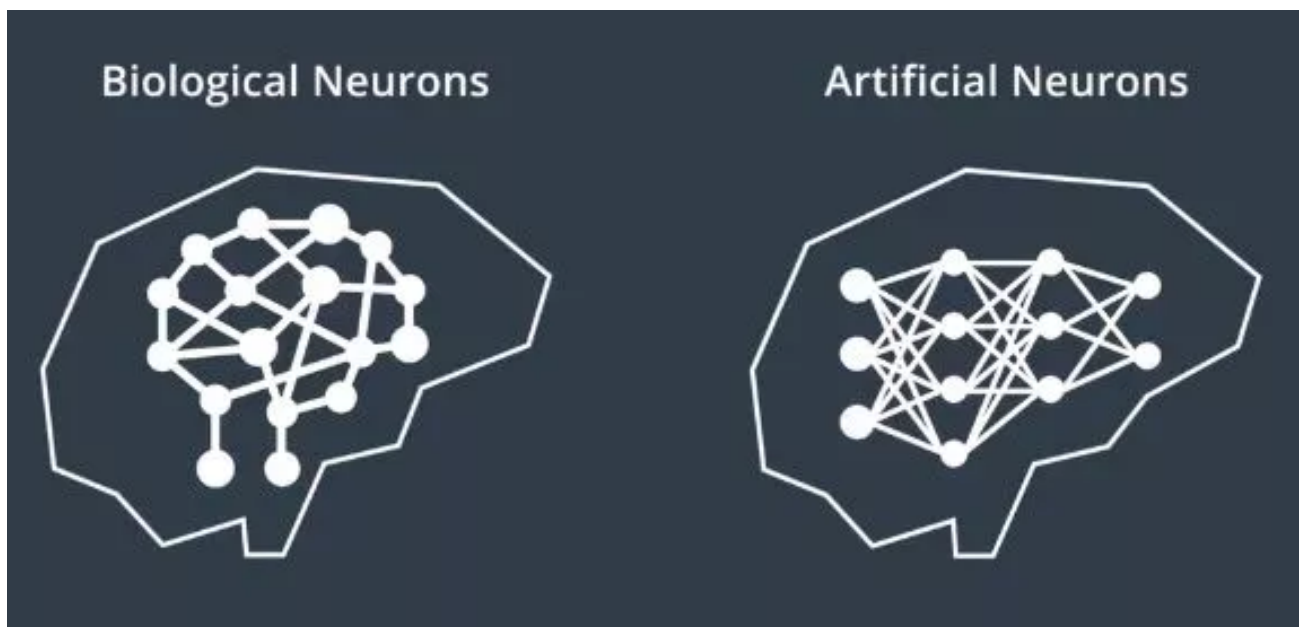
例如在模拟器中，强化学习智能体可训练汽车进行右转，智能体将在初始位置发动车辆，然后以多种不同的方向和速度进行实验性驾驶，如果车辆实际完成了右转，智能体会提高奖励即得分。



起初车辆可能无法找到执行转弯的方法，然而就像人类那样，车辆最终会从一些成功的右转经验中学习，最后学会如何完成任务。

6 | 神经网络

人工神经网络用于无人驾驶车，受到构成人类神经系统的生物神经元启发，生物神经元通过相互连接构成了神经网络或神经网络，通过类似的方式将人工神经元层连接起来以创建用于机器学习的人工神经网络。

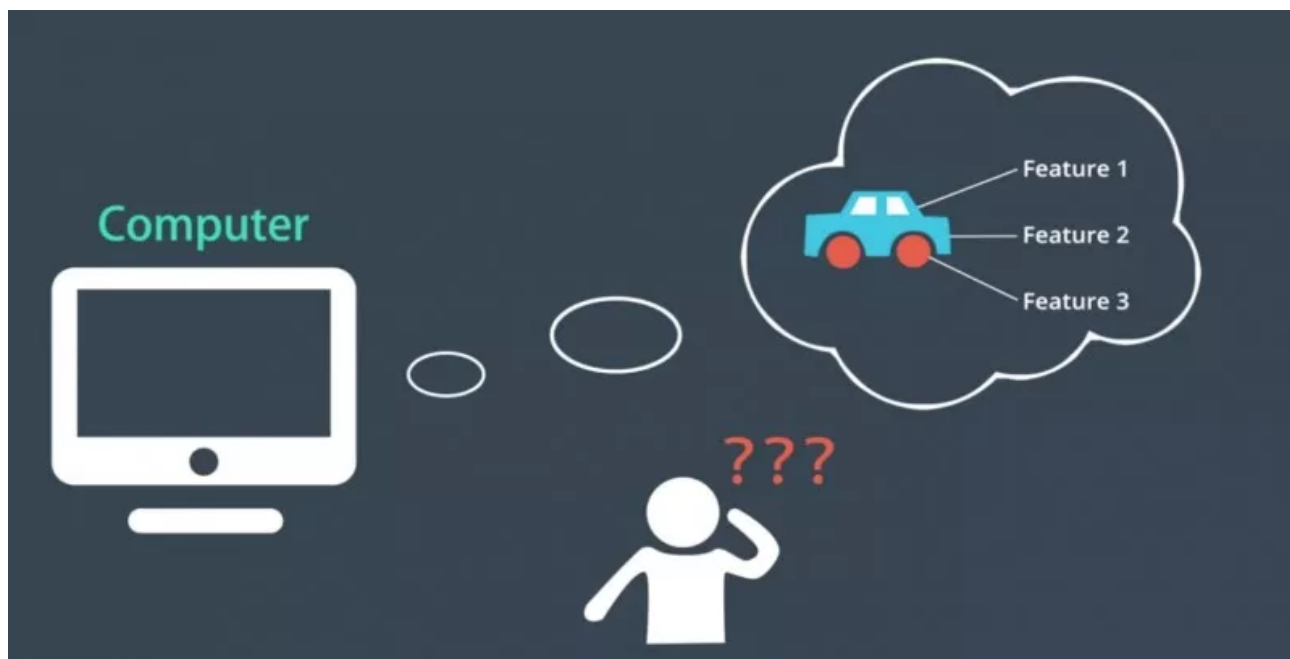


人工神经网络是通过数据来学习复杂模式的工具，神经网络由大量的神经元组成，人工神经元负责传递和处理信息，也可以对这些神经元进行训练。可以将这些图像识别为车辆，无论它们是黑是白、或大或小，你甚至可能不知道自己如何知道它们是车辆，也许是某些特征触发了你的反应，如车轮、车灯、车窗。人工神经网络具有类似的运作方式。人工神经网络通过密集训练，计算机可以辨别汽车、行人、交通信号灯、电线杆。它们学习了用于执行任务的模型，只是我们可能很难直观地理解该数学模型。

当看到该图像时，你的大脑如何工作？你的大脑可能会将图像分为几部分然后识别特征如车轮、车窗、颜色，然后大脑将使用这些特征对图像进行检测和分类。



例如在确定图像是否为车辆时，大脑可能不会认为颜色是关键特征。因为汽车有多种颜色，所以大脑会将更多权重放在其他特征上并降低颜色的重要性。



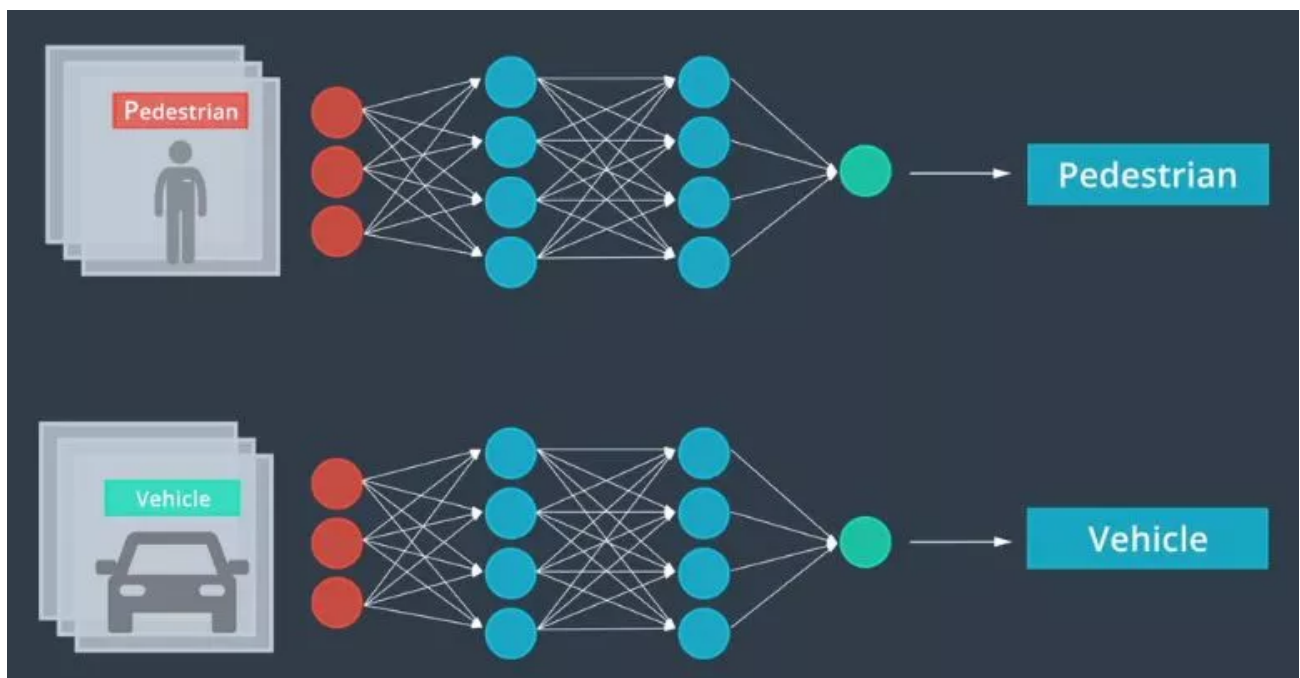
同样，神经网络也会从图像中提取许多特征，但这些特征可能是我们人类无法描述或甚至无法理解的特征。但我们最终并不需要理解，计算机将调整这些特征的权重，以完成神经网络的最终任务。

7 | 反向波算法

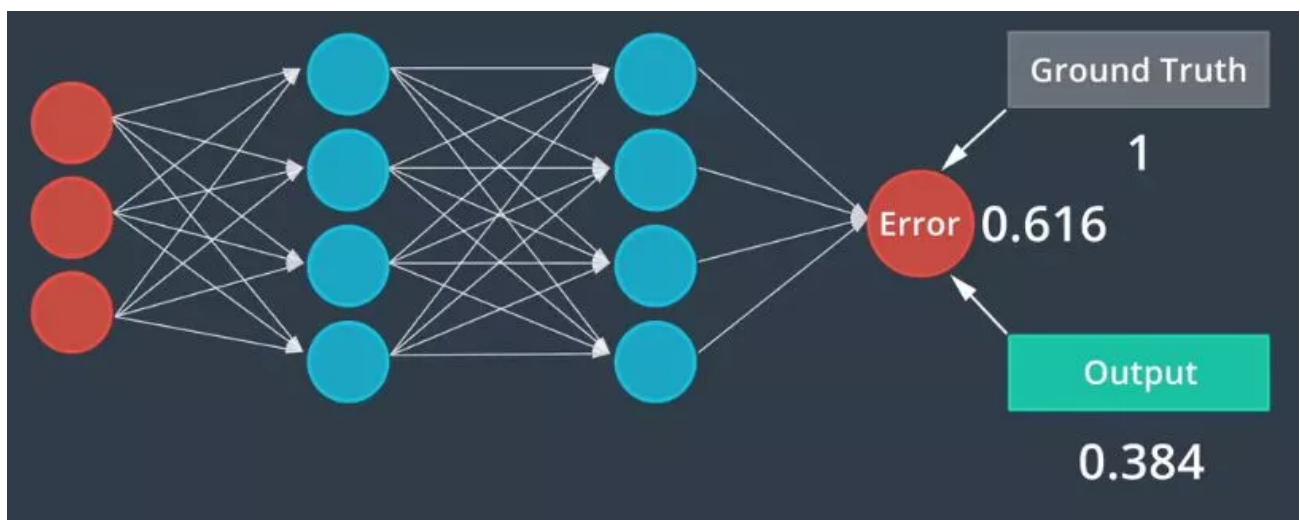
学习有时称为训练，由三步循环组成——**前馈、误差测定、反向传播**。

首先随机分配初始权重即人工神经元的值，通过神经网络来馈送每个图像产生输出值，这被称为前馈。

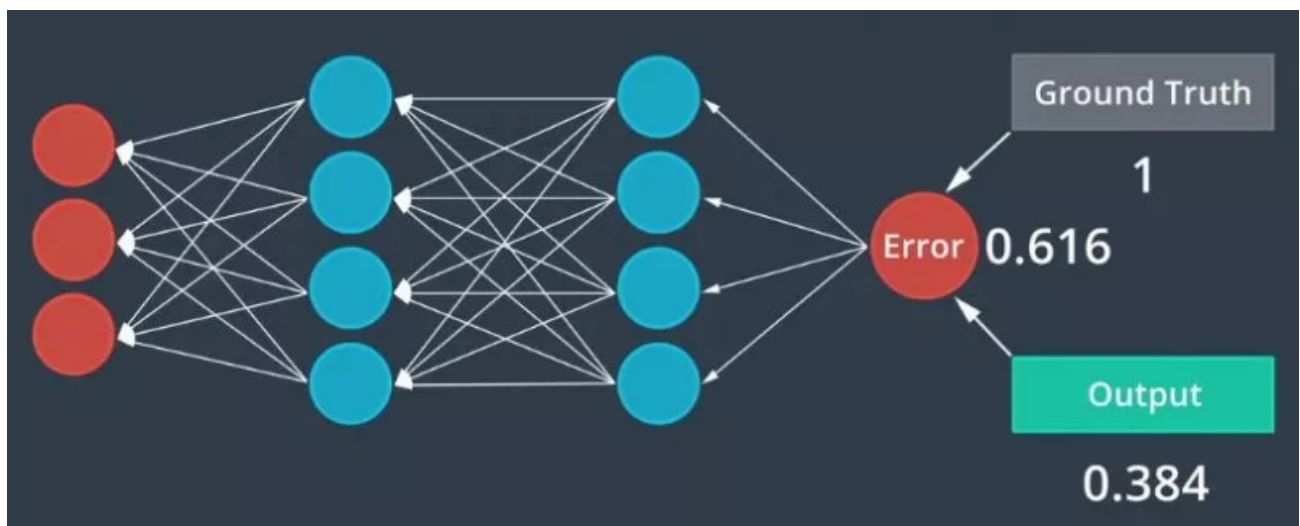
继续开车可以发现，有些点右边只有一棵树，也可以排除。



下一步为误差测定，误差是真值标记与与前馈过程所产生输出之间的偏差。



下一步是反向传播，通过神经网络反向发送误差，此过程类似前馈过程，只是以相反方向进行。



每个人工神经元都对其值进行微调，这是基于通过神经网络后向传播的误差，可生成更准确的网络。一个训练周期：包括前馈、误差测定、反向传播还远远不够。为了训练网络，通常需要数千个这样的周期，最终结果应该是模型能够根据新数据做出准确预测。