

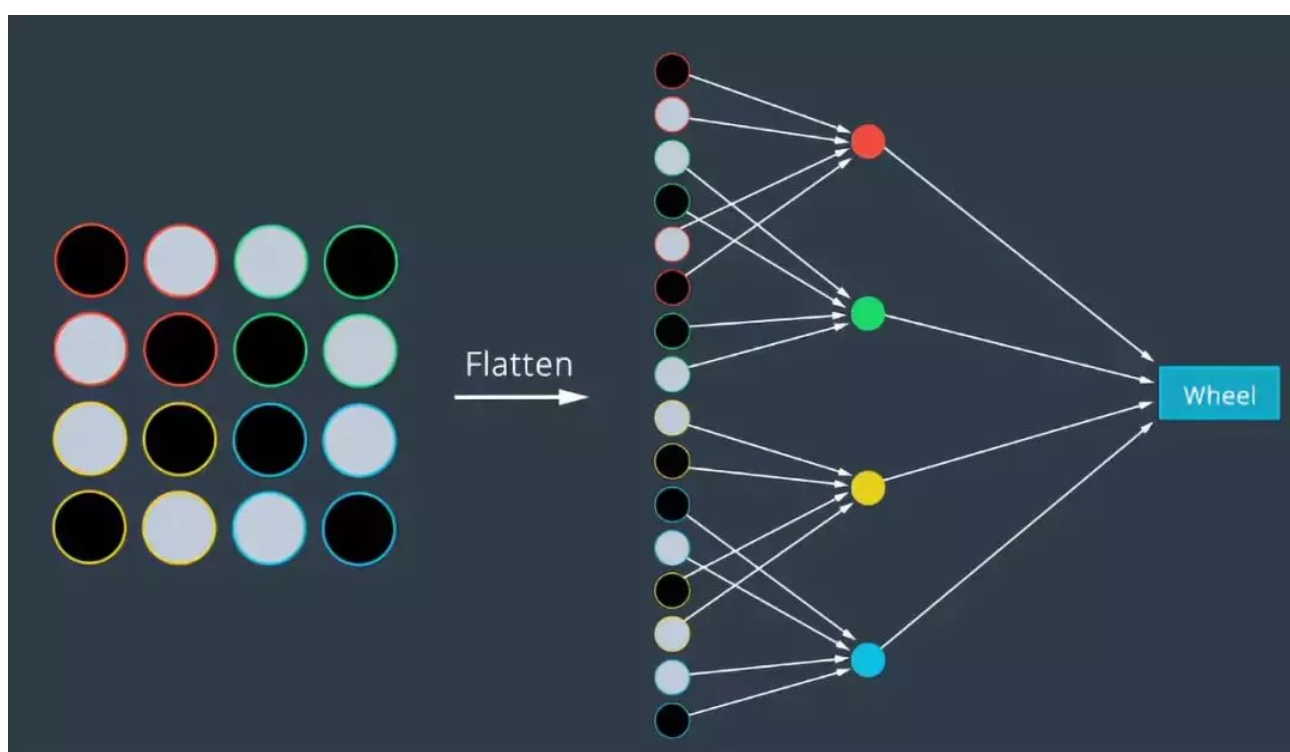
Apollo自动驾驶入门课程第⑤讲 — 感知（下）



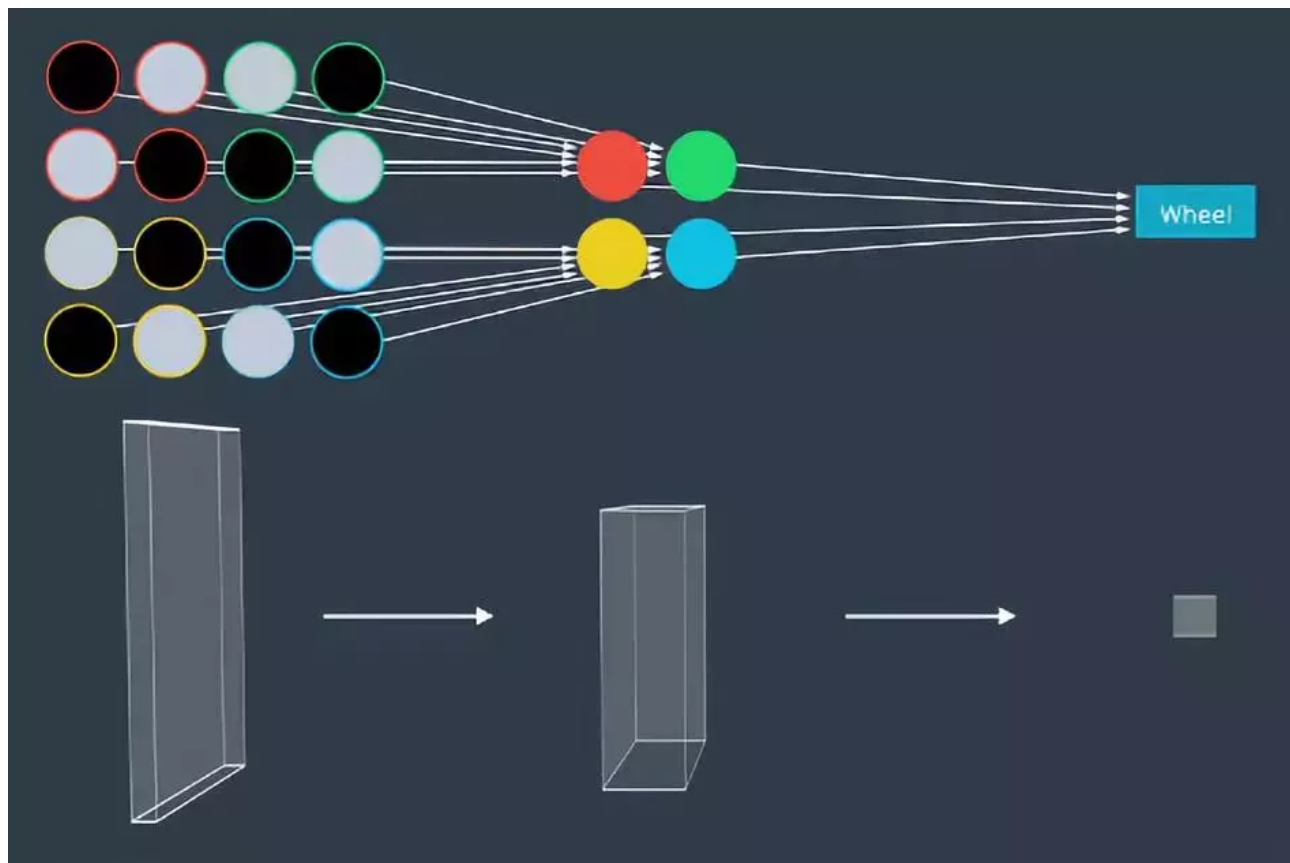
第五课，感知（下）

1 | 卷积神经网络

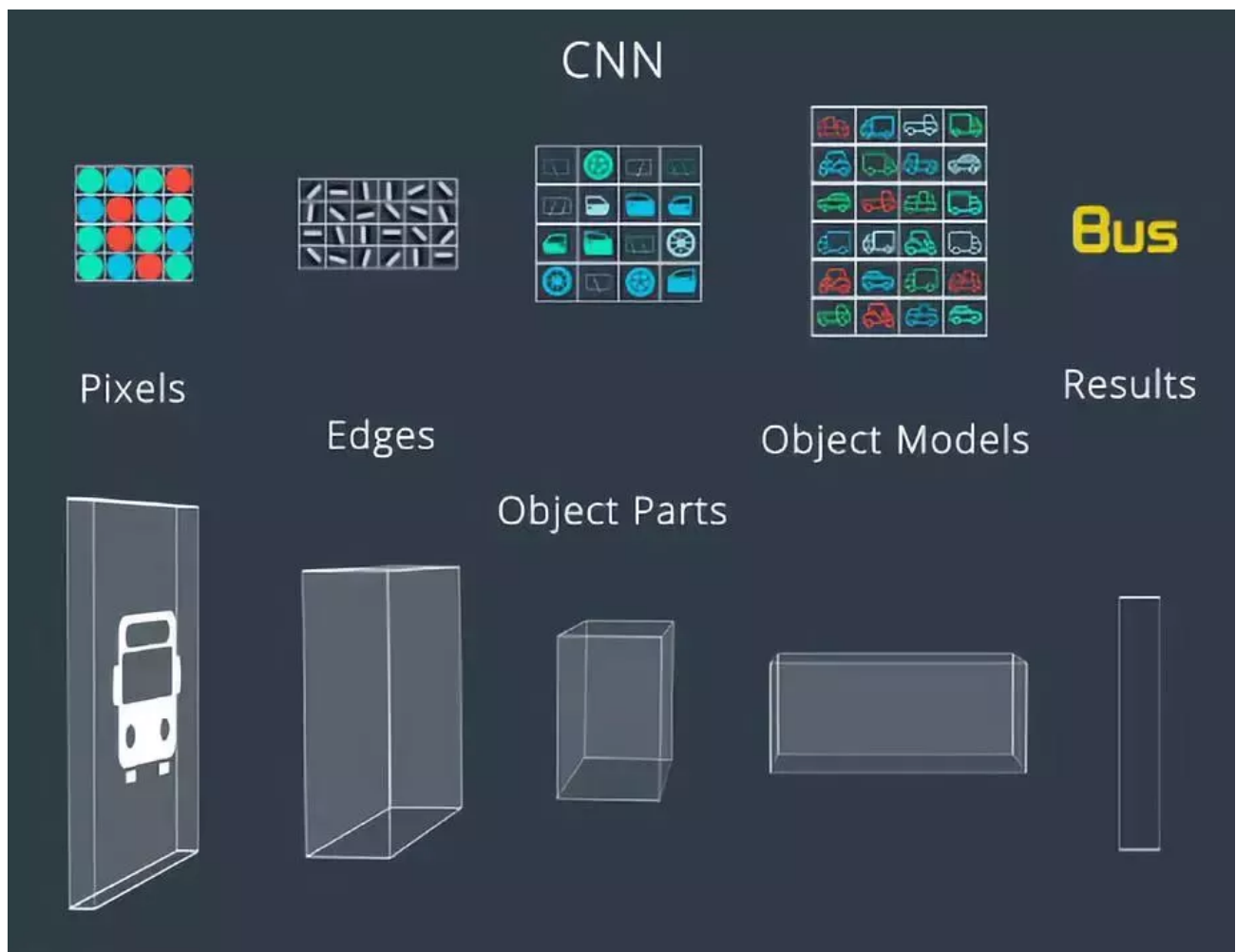
卷积神经网络CNN是一种人工神经网络，它对感知问题特别有效，CNN接受多维输入包括定义大多数传感器数据的二维和三维形状。如果使用标准神经网络对图像进行分类，则需要通过一种方法将图像连接到网络的第一层，这属于一维。标准做法是通过将图像矩阵重塑为一个矢量，并在一个大行中连接所有列，将图像“展开”为一维像素阵列。



然而这种方法打破了图像中所嵌入的空间信息，如果图像中有车轮，则车轮中的所有像素将散布在整个像素阵列中。如果将其散布在一个维度上，神经网络很难从图像中提取车轮。CNN通过维持输入像素之间的空间关系来解决这个问题，CNN通过将过滤器连续滑过图像来收集信息，每次收集信息时，只对整个图像的一小部分区域进行分析，称为**卷积**。



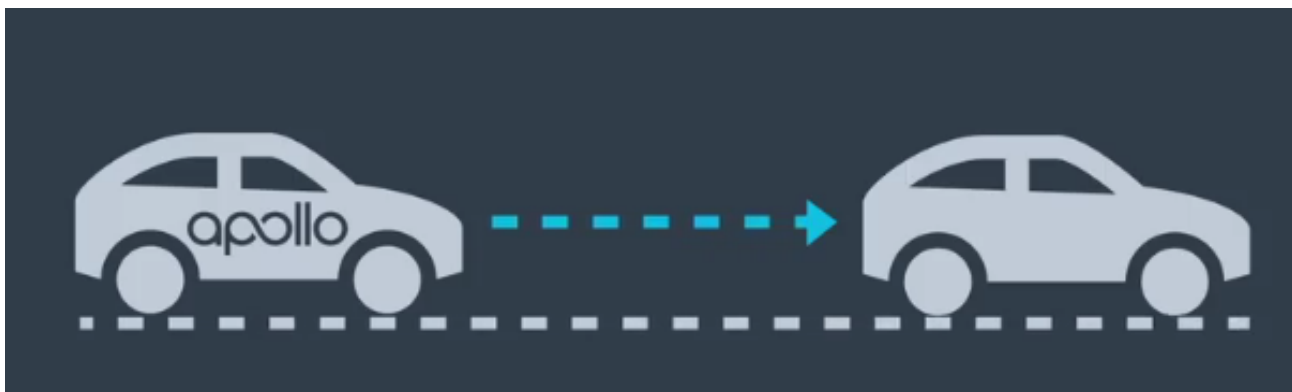
当在整个输入图像上对一个过滤器进行卷积时，将该信息与下一个卷积层相关联。例如CNN识别第一个卷积层中的基本边缘和颜色信息，然后通过在第一层上卷积新过滤器，CNN使用边缘和颜色信息来归纳更复杂的结构如车轮、车门、挡风玻璃；而另一个卷积可使用车轮、车门、挡风玻璃识别整个车辆；最后神经网络可使用这一高阶信息对车辆进行分类。



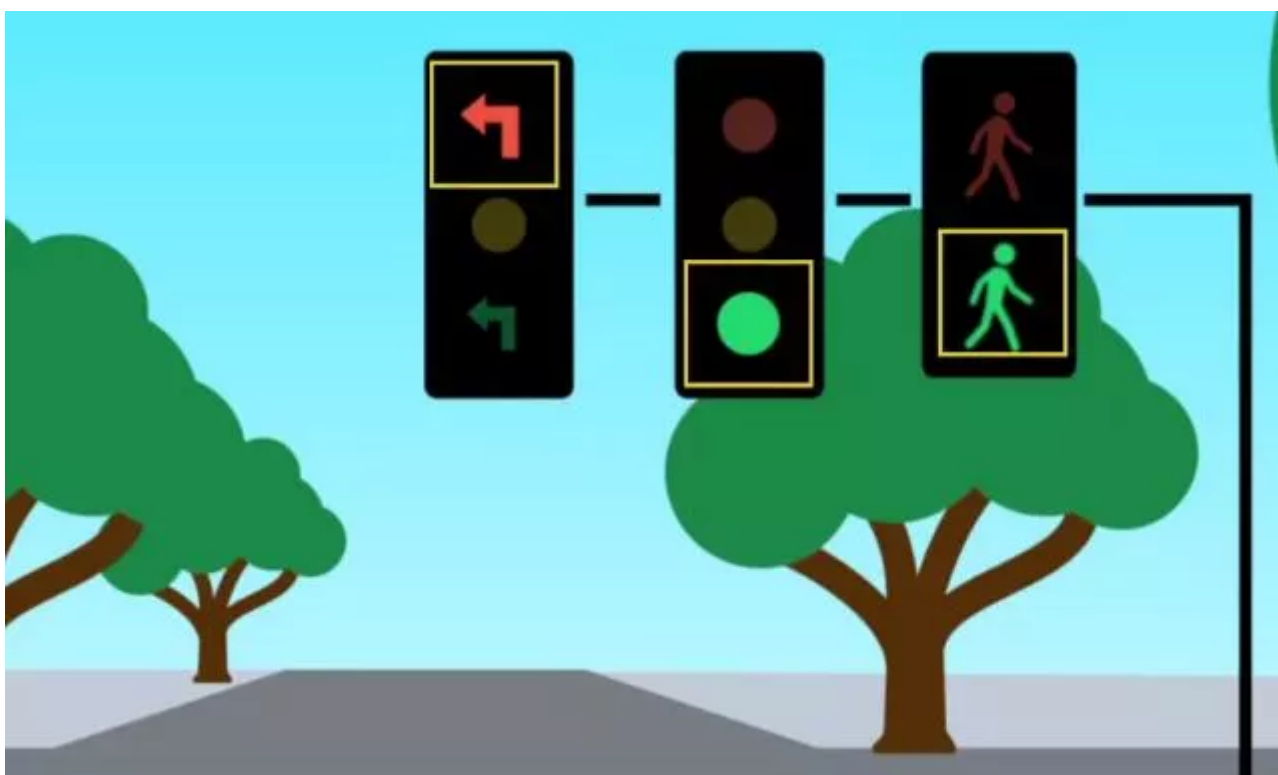
CNN根据其任务查找真正需要的特征，任务可能是图像检测、分类、分割、其他类型的目标。

2 | 检测与分类

感知任务中，首先是**障碍物检测和分类**，在驾驶过程中会遇到许多障碍物，静态障碍物包括墙壁、树木、杆子、建筑物；动态障碍物包括行人、自行车、汽车。计算机首先需要知道障碍物的位置，然后进行分类。在路中行驶的无人驾驶车可能会探测到许多不同的物体，汽车根据所感知的物体类型来确定路径和速度。如果感知到前方有一辆自行车，汽车可能会决定减速和变道，以便安全驶过自行车。但是如果感知到前方有另一辆车，并预测到前方车辆也将以接近限速的速度行驶，可能会保持其速度和车道。



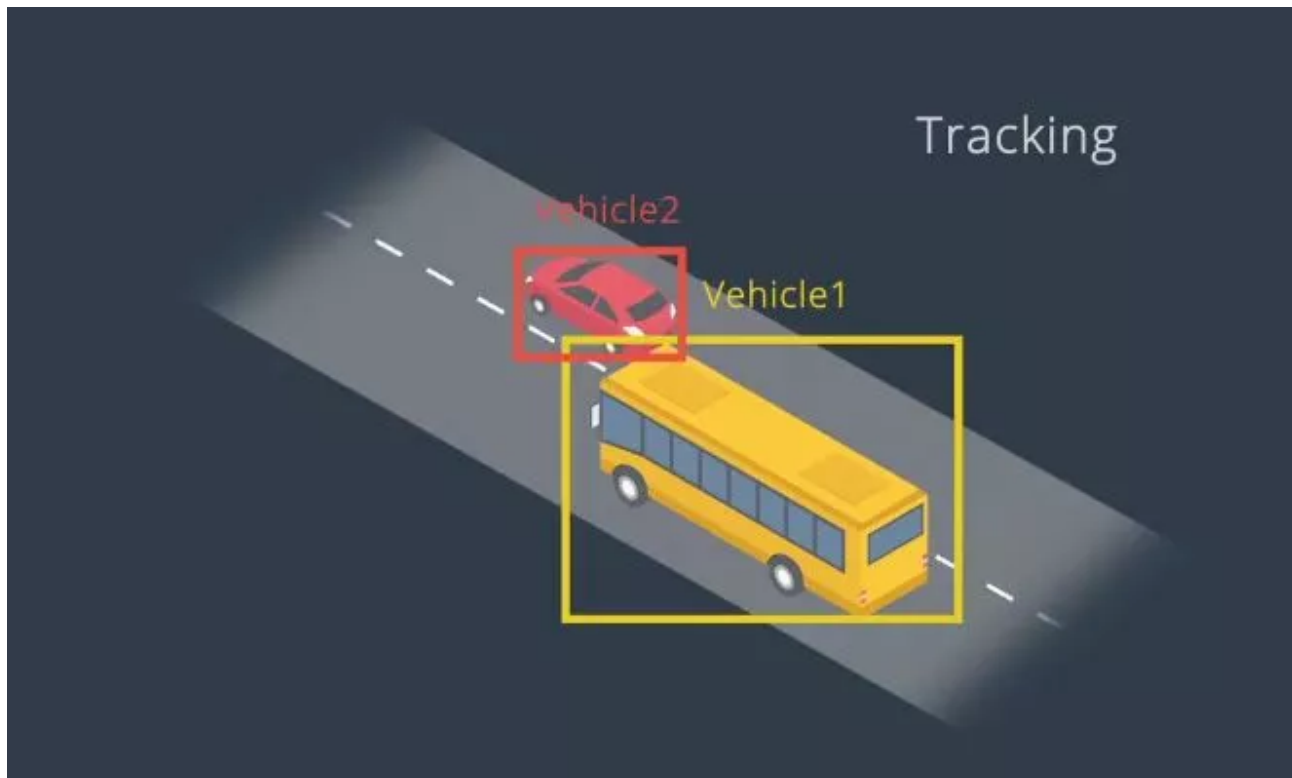
将另一个示例为交通信号灯检测分类，首先将使用计算机视觉对图像中的交通信号灯进行定位，然后根据灯光显示颜色对交通信号灯进行分类。



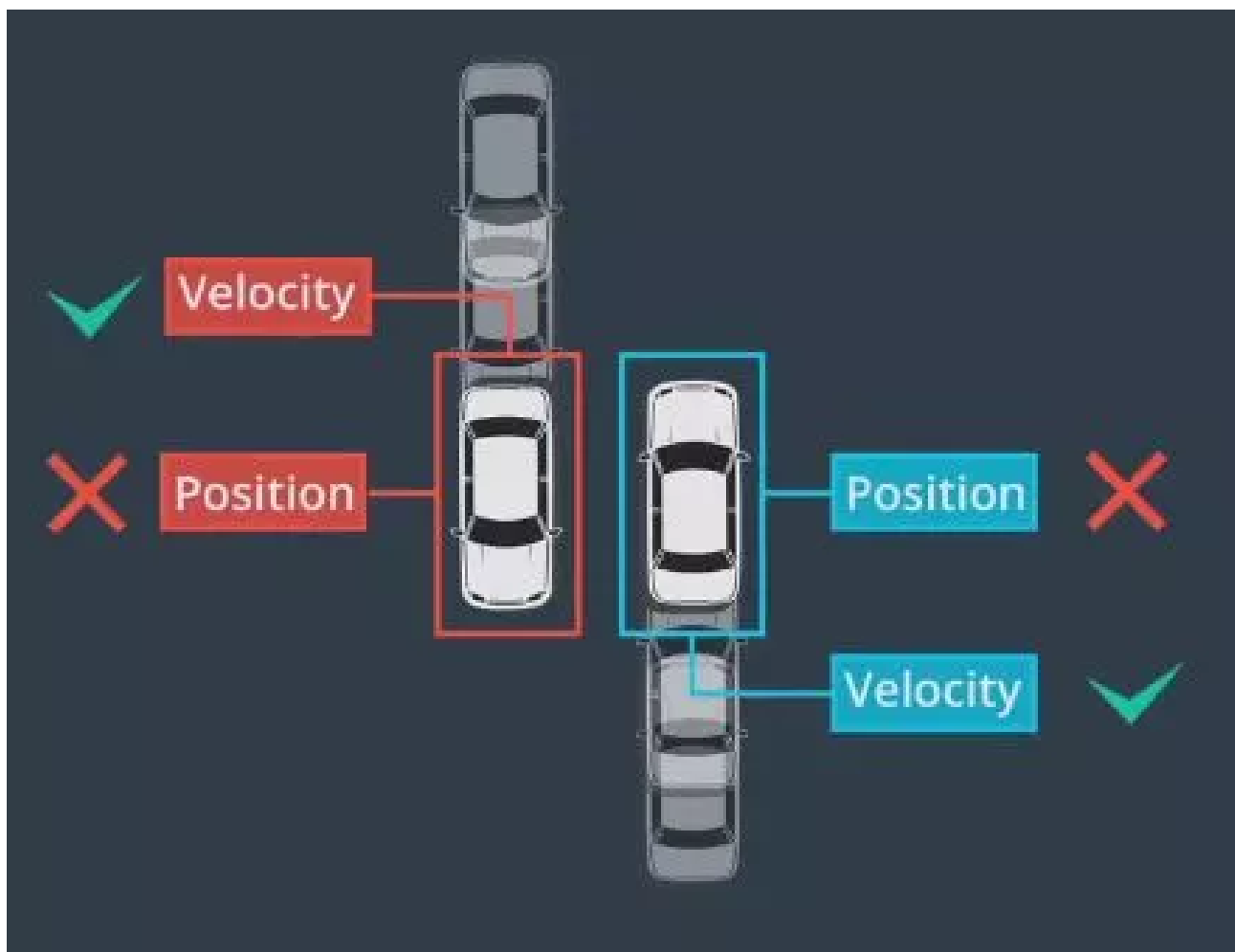
无人驾驶中，使用什么算法来对障碍物进行检测和分类？可以先使用**检测CNN**来查找图像中的对象的位置，在对图像中的对象进行定位后，可以将图像发送给另一个CNN进行分类。也可以使用单一CNN体系结构对对象进行检测和分类，一种通常的做法为在单个网络体系结构的末端附加几个不同的“头”，一个头可能执行检测，另一个则可能执行分类。经典体系结构为 R-CNN、Fast R-CNN、Faster R-CNN、YOLO、SSD等。

检测完对象后，需要追踪它们追踪的意义是什么？如果我们对每个帧中的每个对象进行检测并用边界框对每个对象进行标识，那么跨帧追踪对象会带来哪些好处？

首先追踪在检测失败时是至关重要的，如果你在运行检测算法时，对象被其他对象遮挡一部分，则检测算法可能会失败，但追踪可以解决遮挡问题。另一个原因在于追踪可以保留身份，障碍物检测的输出为包含对象的边界框，但是对象没有与任何身份关联，单独使用对象检测时，计算机不知道一个帧中的哪些对象与下一帧中的哪些对象相对应，该任务对人类来说很简单，但对汽车来说很困难。



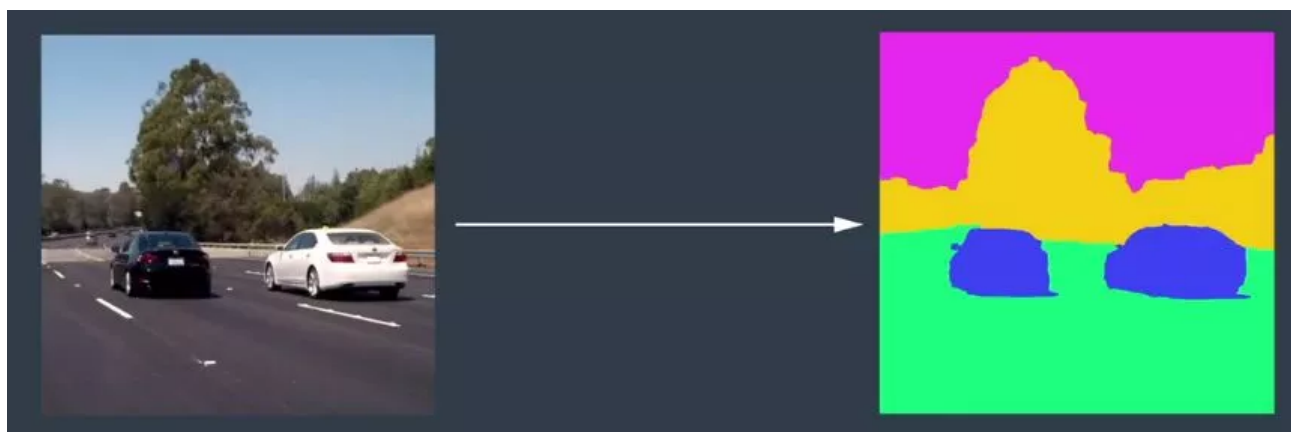
追踪的第一步为确认身份，通过查找特征相似度最高的对象，将在之前的帧中检测到的所有对象与在当前的帧中检测到的对象进行匹配。对象具有各种特征，可能基于颜色、基于形状，计算机视觉算法可以计算出复杂的图像特征，如局部二值模式和方向梯度直方图。当然也需要考虑连续视频帧中，两个障碍物之间的位置和速度，由于两帧之间的对象位置和速度没有太大变化，该信息也可以帮助快速找到匹配的对象。



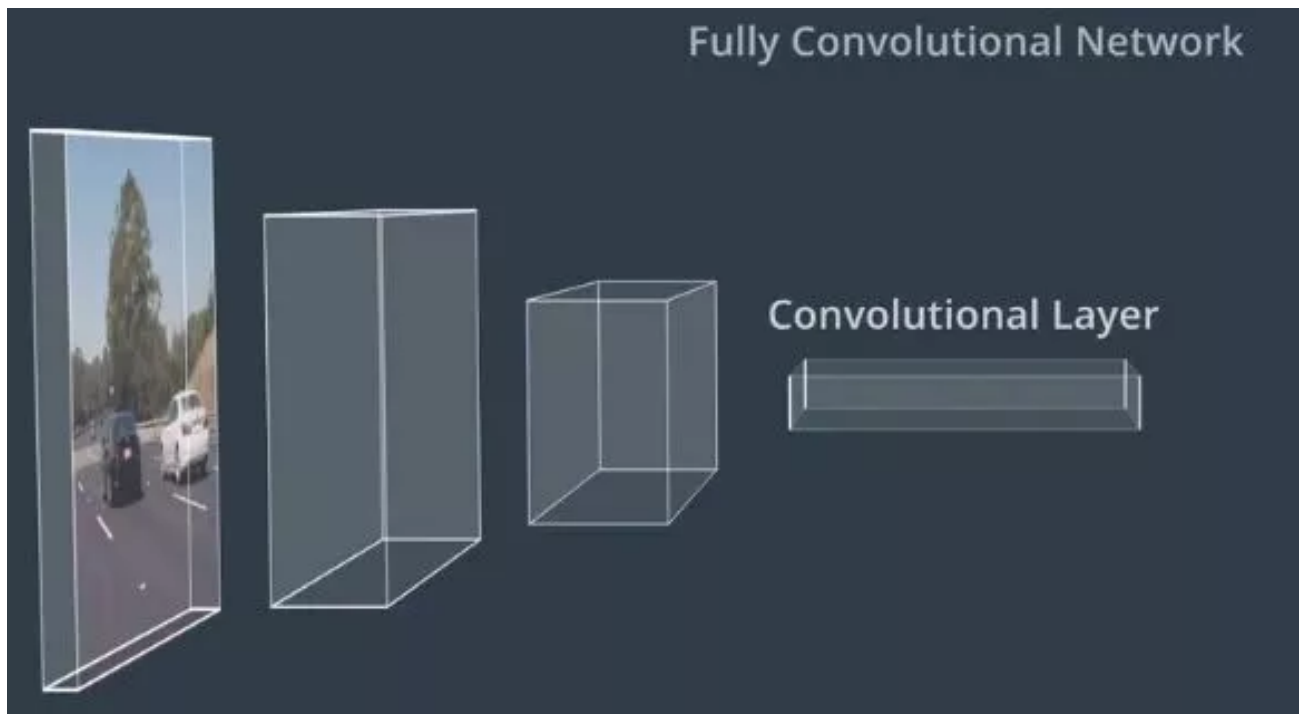
在确定身份后，可以使用对象的位置并结合预测算法以估计在下一个时间步的速度和位置，该预测可帮助识别下一帧中的相应对象。

4 | 分割

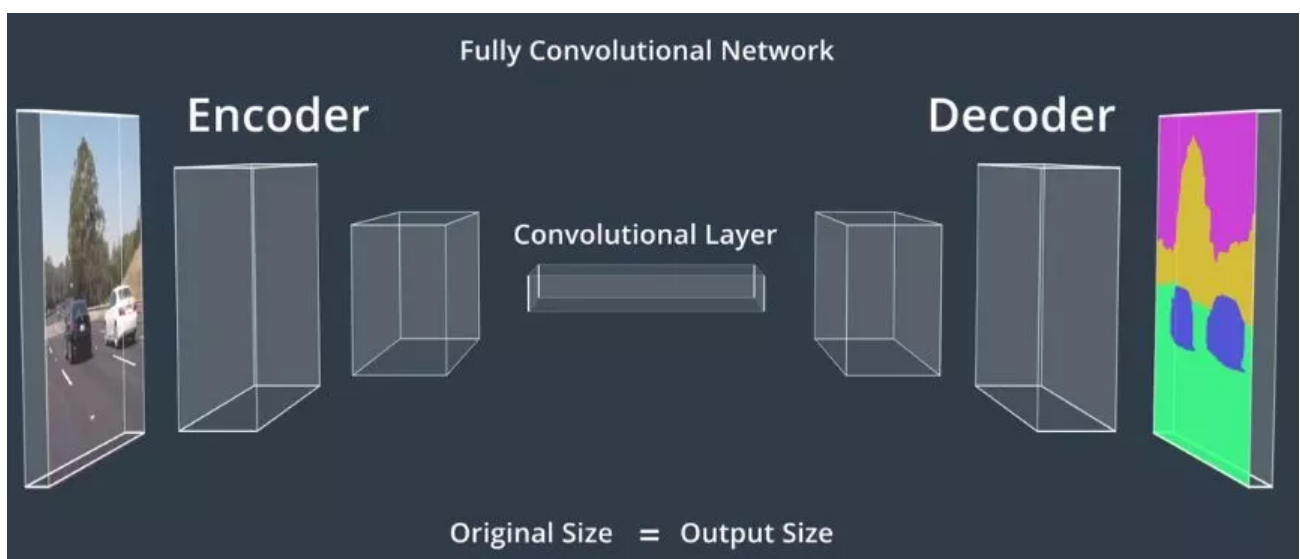
语义分割涉及对图像的每个像素进行分类，用于尽可能详细地了解环境并确定车辆可驾驶区域。



语义分割依赖于一种特殊类型的CNN，被称为**全卷积网络FCN**。FCN用卷积层来替代传统CNN体系结构末端的平坦层，现在网络中的每一层都是卷积层，因此名称为“**全卷积网络**”。



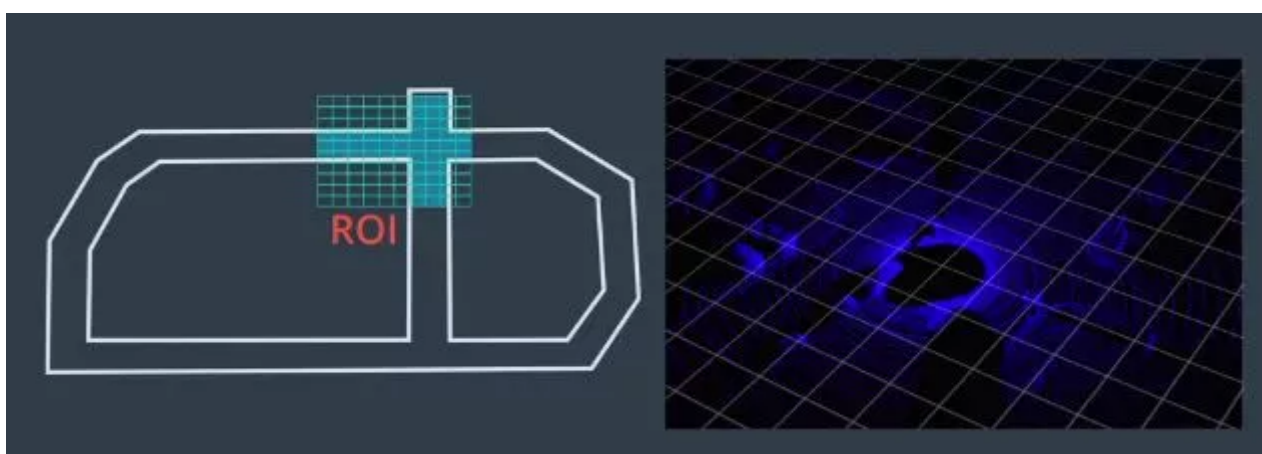
FCN提供了可在原始输入图像之上叠加的逐像素输出，必须考虑的一个复杂因素是大小。在典型的CNN中经过多次卷积之后所产生的输出比原始输入图像小得多。然而为了分割像素，输出尺寸必须与原始图像的尺寸相匹配。为了完成，可以对中间输出进行上采样处理，直到最终输出的大小与原始输出图像的大小相匹配。网络的前半部分通常被称为**编码器**，因为这部分网络对输入图像的特征进行了提取和编码。网络的后半部分通常被称为**解码器**，因为它对这些特征进行了解码并将其应用于输出。



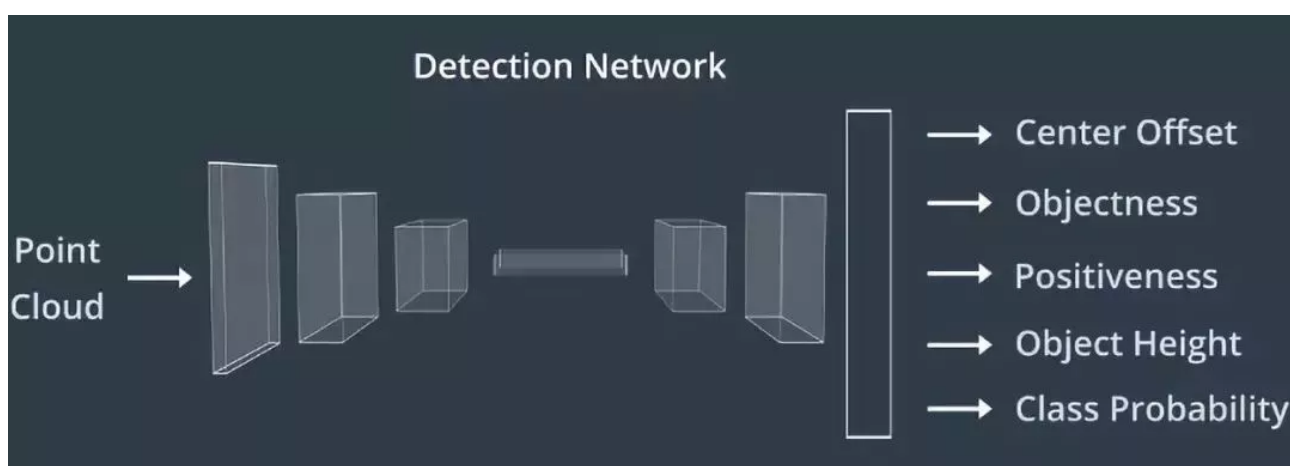
激光雷达数据提供了用于构建世界视觉表征的足够空间信息，计算机视觉技术不仅可以使摄像头图像进行对象分类，还可以使用点云和其他类型的空间相关数据进行对象分类。

5 | Apollo 感知

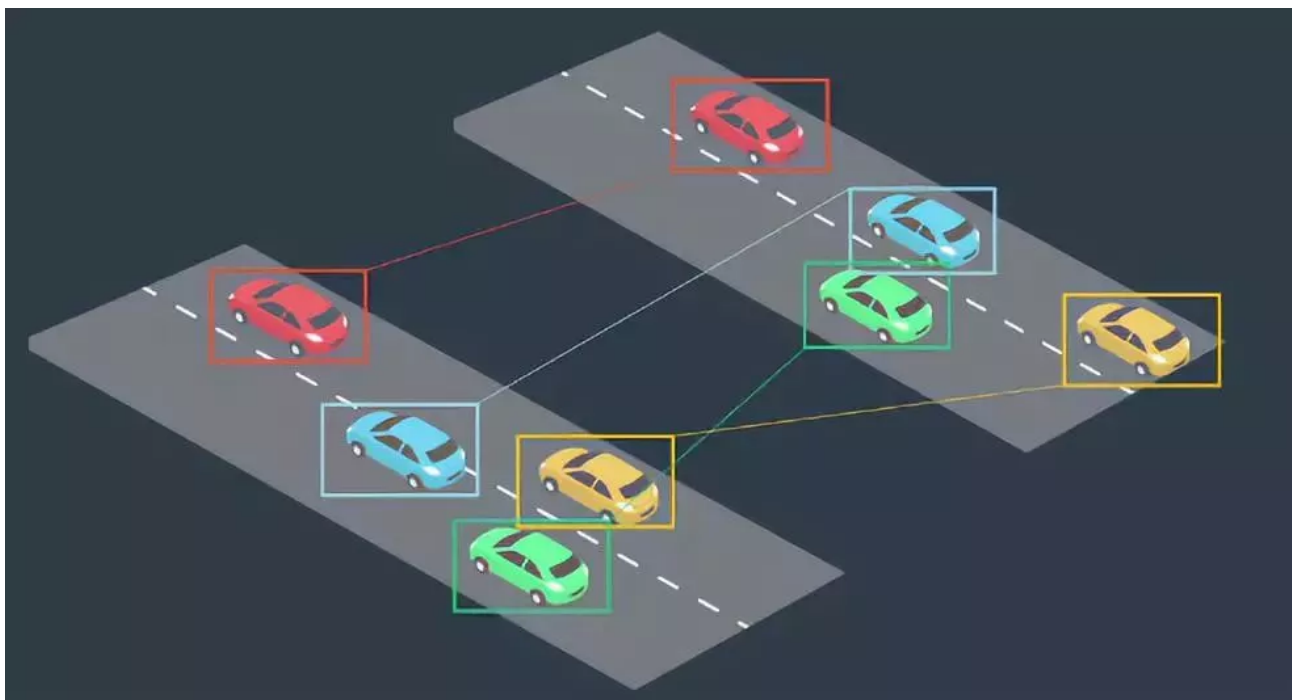
Apollo 开放式软件栈可感知障碍物、交通信号灯、车道。对于三维对象检测，Apollo 在高精度地图上使用感兴趣区域ROI来重点关注相关对象。Apollo将ROI过滤器应用于点云和图像数据，以缩小搜索范围并加快感知。



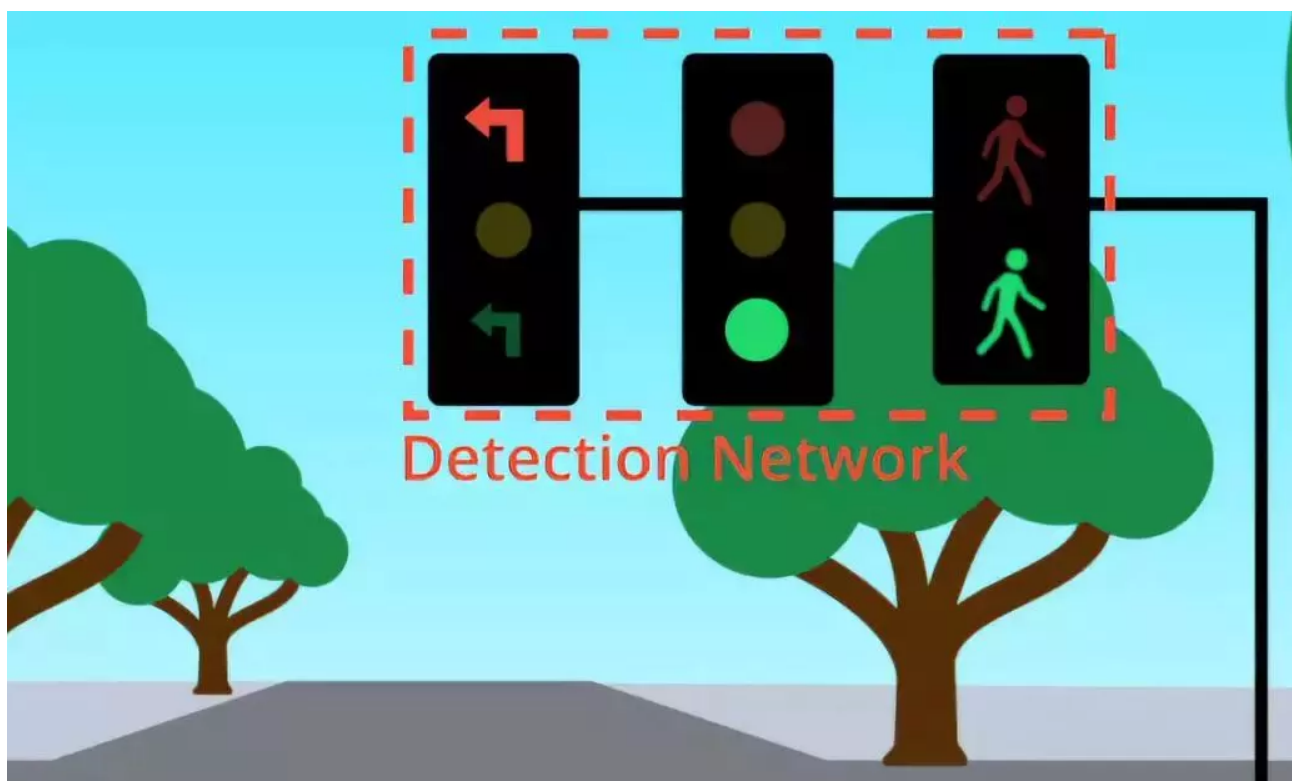
然后通过检测网络馈送已过滤的点云，输出用于构建围绕对象的三维边界框。



最后使用被称为检测跟踪关联的算法来跨时间步识别单个对象，该算法先保留在每个时间步要跟踪的对象列表，然后在下一个时间步中找到每个对象的最佳匹配。

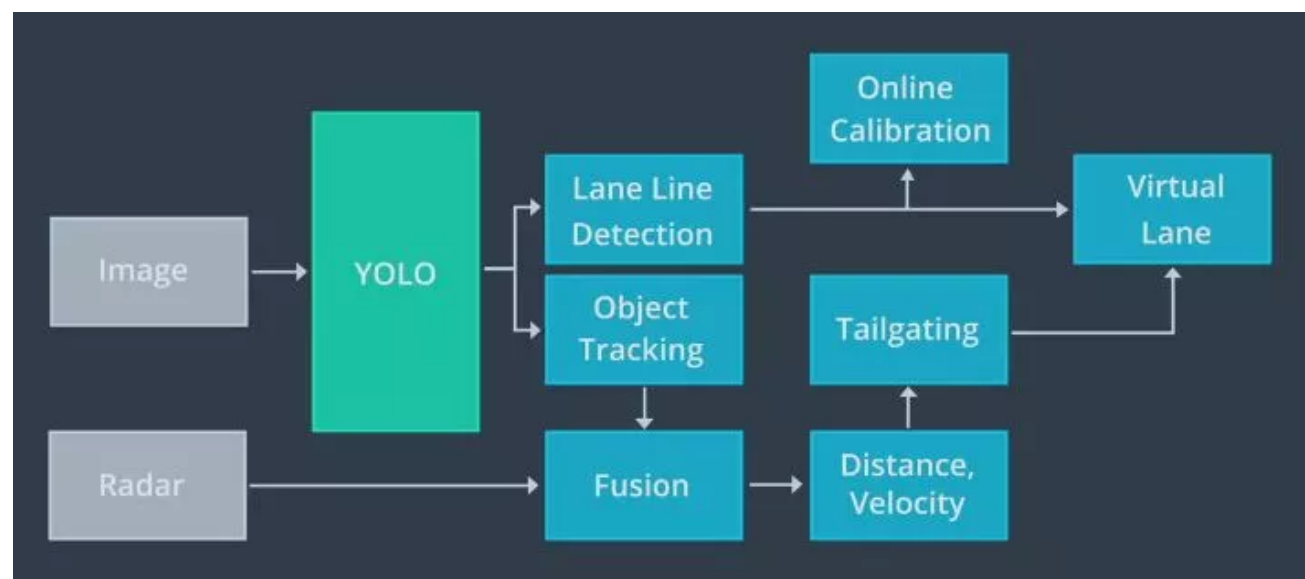


对于交通信号灯的分类，Apollo先使用高精度地图来确定前方是否存在交通信号灯。如果前方有交通信号灯，则高精度地图会返回灯的位置。后续根据位置，摄像头搜索范围，在摄像头捕获到交通信号灯图像后。Apollo使用检测网络对图像中的灯进行定位，然后从较大的图像中提取交通信号灯。将裁剪的交通灯图像提供给分类网络以确定灯颜色，如果有许多灯则系统需要选择哪些灯与其车道相关。



Apollo使用 **YOLO 网络**来检测车道线、动态物体其中包括车辆、卡车、骑自行车的人、行人。在经过YOLO网络检测后，在线检测模块会并入来自其他传感器的数据对车道线预测进行调整，

车道线最终被并入名为“虚拟车道”的单一数据结构中。同样也通过其他传感器的数据对YOLO网络所检测到的动态对象进行调整，以获得每个对象的类型、位置、速度、前进方向。虚拟通道和动态对象均被传递到规划与控制模块。



6 | 传感器数据比较

感知通常依赖于摄像头、激光雷达、雷达，如图显示了这三种传感器的优缺点。摄像头非常适用于分类，在Apollo中摄像头主要用于交通信号灯分类、车道检测。激光雷达的优势在于障碍物检测，即使在夜间仍能准确地检测障碍物。雷达在探测范围和应对恶劣天气方面占优势，通过融合这三种传感器的数据可实现最佳聚合性能，被称为“传感器融合”。

	Camera	LiDAR	Radar	Camera+Radar+LiDAR
Object Detection	●	●	●	●
Object Classification	●	●	●	●
Range of Visibility	●	●	●	●
Lane Tracking	●	●	●	●
Functionality in Bad Weather	●	●	●	●
Functionality in Poor Lighting	●	●	●	●

● Good

● Mixed

● Poor

雷达已经在汽车上使用很多年，在各种系统中都需要雷达，如自适应巡航控制、盲点警告、碰撞警告和碰撞预防系统等。尽管雷达技术已经成熟，它仍在不断进步，作用不断提升。其他传感器测量速度的方法是计算两次读数之间的差距，而雷达则通过多普勒效应来直接测量速度。多普勒效应根据对象在远离还是接近你，测量出雷达的频率变化。就像消防车警报器一样，当车辆正在远离你和驶向你时，听起来声是不一样的。多普勒效应对传感器融合至关重要。因为它可以把速度作为独立的测量参数，从而提升了融合算法的收敛速度。雷达还可以生成环境的雷达地图，进而实现定位。因为雷达波在坚硬表面会回弹。因此，它可以直接测量对象距离，无需在视线范围内也可以。雷达可以看到其他车辆底部。并发现可能会被阻挡的建筑物和对象。在车上的所有传感器中，雷达是不容易受雨雾影响的。而且视野宽阔，可达 150 度，距离可达 200 多米。与激光雷达和摄像头相比，雷达分辨率较低，尤其是在垂直方向，分辨率非常有限。分辨率低意味着来自静态物体的反射可能产生问题。例如，街道上检修孔盖或汽水罐，可能产生很高的雷达反射率，但他们并不大。我们将其称为雷达杂波。因此，当前的车载雷达通常会忽视静态物体。

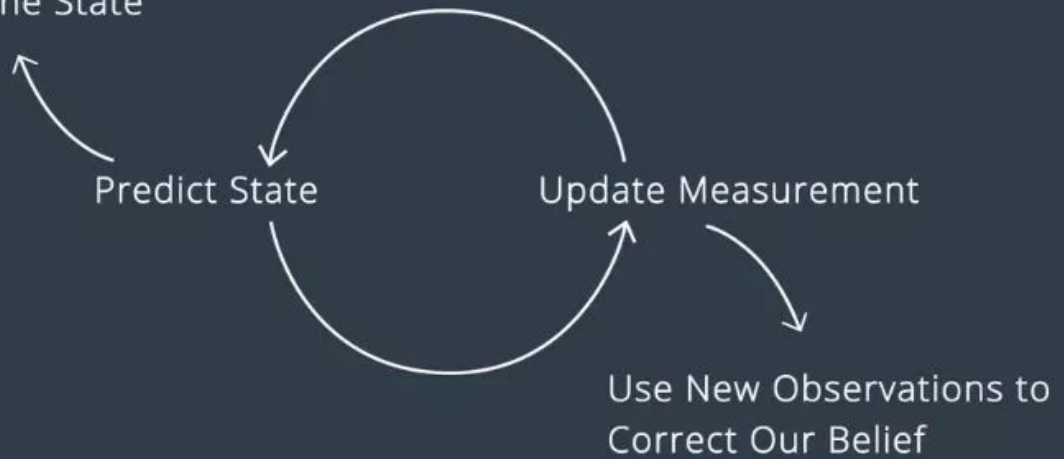
激光雷达是激光探测与测量的简称，而雷达则是无线电探测与测量的简称。雷达使用无线电波，而激光雷达则使用红激光束来确定传感器和附近对象的距离。目前的激光雷达大多使用 900 纳米光波光度的光源。但部分激光雷达使用的光波长度更长，在雨雾中性能更好。当前的激光雷达使用旋转座架发射激光，扫描周边环境。激光是脉冲式的，脉冲被对象反射，然后返回一个点云，来代表这些物体。激光雷达的空间分辨率远远高于雷达。因为激光束越聚焦，垂直方向的扫描层数量就越多，因此每层的激光雷达的密度也越高。目前，激光雷达还不能直接测量对象的速度，必须使用两次或多次扫描之间的位置差来确定。激光雷达受天气和传感器清洁程度影响也很大，因此需要保持清洁。它们块头也比其他传感器更大，因此也很难安装，除非你只想在车顶安装一个大的激光扫描器。

7 | 感知融合策略

Apollo使用激光雷达和雷达来检测障碍物，用于融合输出的主要算法为**卡尔曼滤波**。卡尔曼滤波有两个步骤：第一步为预测状态，第二步是更新测量结果。设想正在跟踪一名行人，这里的状态表示行人的位置和速度，从已经掌握的行人状态开始，使用这些信息来执行卡尔曼滤波的第一步，即预测行人在将来的状态；下一步为误差结果更新，使用新的传感器来更新所认为的行人状态，卡尔曼滤波算法是预测和更新步骤的无限循环。

Two-Step Estimation Problem

Use Information We Have
to Predict the State



实际上有两种测量结果更新步骤：同步和异步。同步融合同时更新来自不同传感器的测量结果，而异步融合则逐个更新所收到的传感器测量结果。传感器融合可提高感知性能，因为各传感器相辅相成，融合也可以减少跟踪误差。