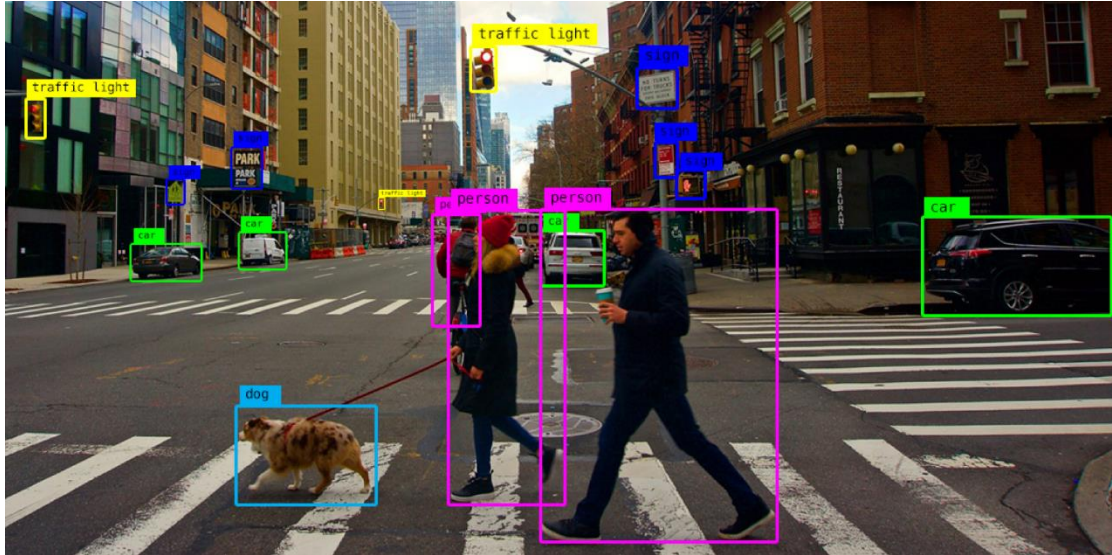


目标检测中的 YOLO 网络长什么样？

机器视觉中有一类称为“目标检测”的任务，目的是从一张图片中检测出行人、汽车、建筑物等目标，并把它们框出来，打上类别标签。



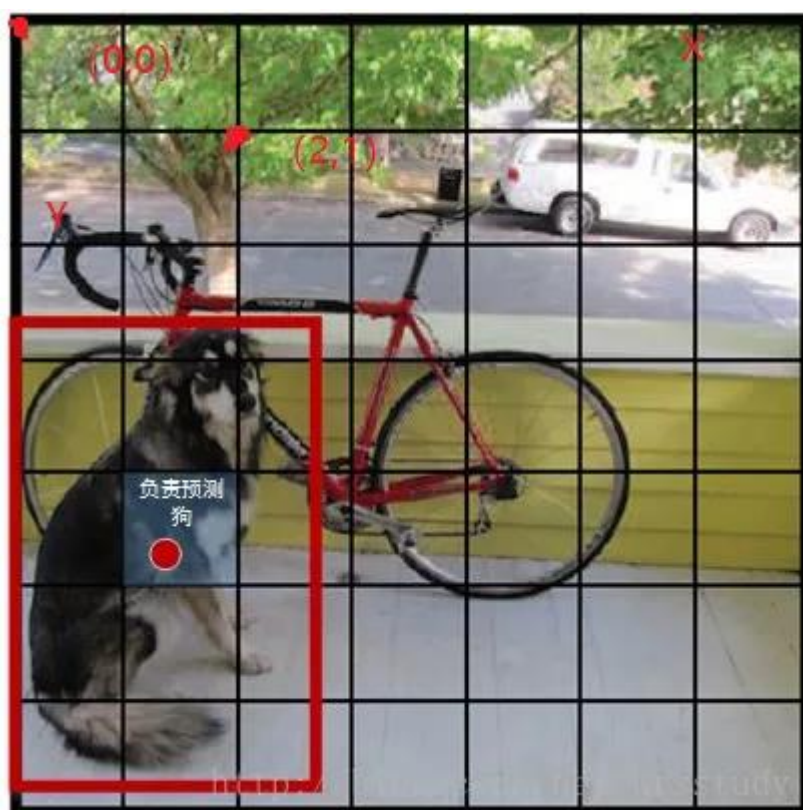
在实际应用场景中，一幅照片中可能出现多个目标，每个目标的类型、位置、大小都有差异，如何把这些目标全部检测出来呢？

一种方式是用不同大小的窗口在图片上滑动，并利用一个深度神经网络来判断框住的部分是不是一个完整的目标。找出所有目标位置后，再利用另一个深度神经网络来判断每个位置的目标是什么。这种基于滑动窗口的方法可以得到较好的效果，但速度较慢。



为了提高检测速度，研究者提出一种称为 YOLO 的实时目标检测方法。这一方法将图片分成很多小格子，神经网络可以同时预测每个小格子所代表的目标，因此不再需要大量的滑动窗口，从而实现更快速的检测。

那么，如何训练一个神经网络实现这一预测能力呢？首先，如果某个目标的中心点落在这个小格子内，则该目标就由这个小格子所代表。例如，图中红色点是狗的中心位置，这一点所处的小格子就负责代表这条狗。有了这一定义，我们就可以对每个小格子所代表的目标进行标记，包括目标的类别、位置、大小等。利用这些标注信息，就可以训练神经网络了。



在预测时，神经网络对每个小格子都预测若干可能的目标区域，最后选那些最可靠的预测作为检测结果。因为对所有小格子是同时预测的，所以可将图片中的所有目标一次性检测出来。

