



目标检测

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数**, **统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

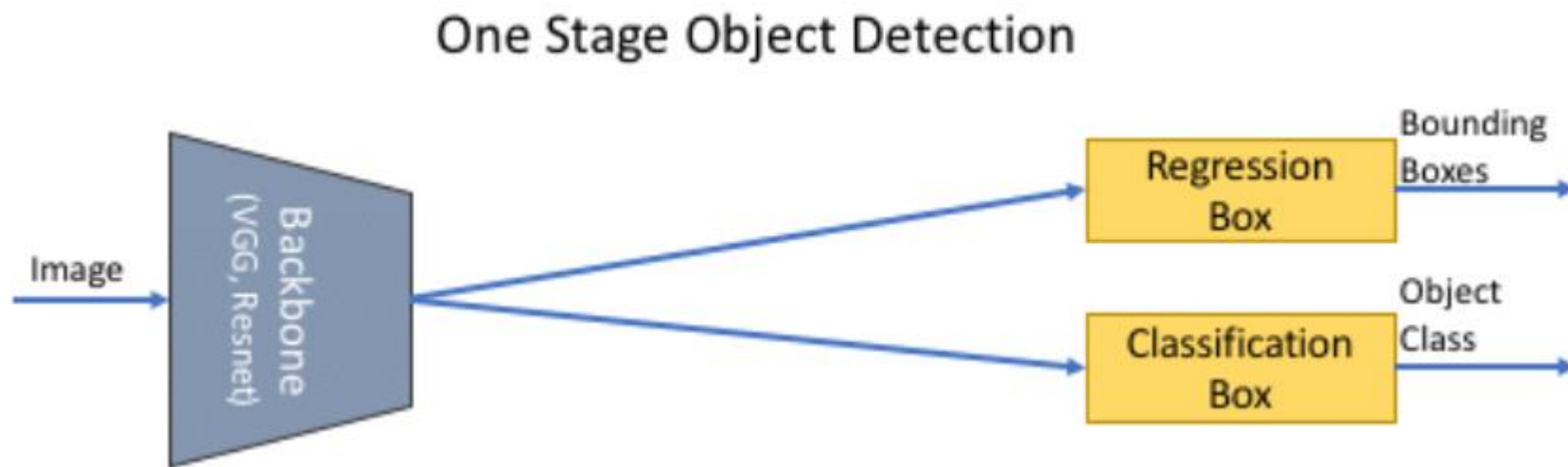
https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

* 版权声明:

- 仅限用于个人学习
- 禁止用于任何商业用途

单阶段目标检测

单阶段目标检测是计算机视觉领域中用于检测图像或视频中物体位置的一种技术。与传统的两阶段目标检测方法（如Faster R-CNN）不同，单阶段目标检测方法（如YOLO和SSD）将目标检测任务简化为一个单一的神经网络模型，直接从输入图像中预测目标的类别和位置。



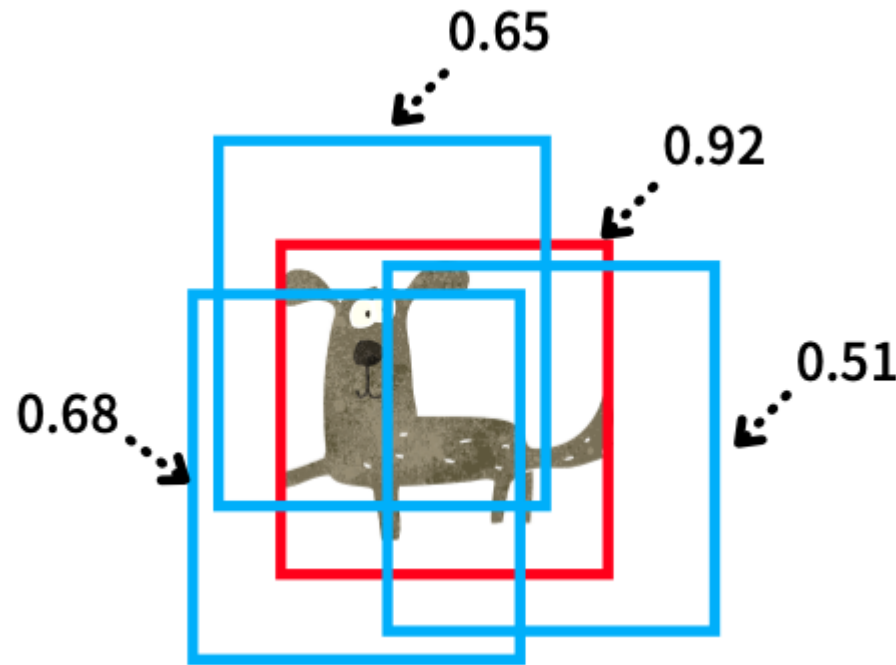
NMS (Non-Maximum Suppression)

NMS (Non-Maximum Suppression) 是一种常用的目标检测算法中的后处理步骤，用于消除重叠的边界框并选择最佳的边界框。在目标检测任务中，检测器通常会在图像中检测到多个候选目标，这些目标可能会有很大的重叠，因此需要一个方法来筛选出最合适的目标框。

NMS 的基本思想是，对于每个类别，首先根据检测器输出的置信度对所有边界框进行排序，然后从置信度最高的边界框开始，逐个地将与该边界框重叠度（通常使用IoU，即交并比）高于设定阈值的其他边界框去除，直到所有边界框都被处理过一次。

NMS 的主要步骤包括：

1. 对所有边界框按照置信度进行排序。
2. 选择置信度最高的边界框，并将其添加到最终输出的结果中。
3. 计算该边界框与其他所有边界框的重叠度。
4. 去除重叠度高于设定阈值的边界框。
5. 重复步骤2-4，直到所有边界框都被处理。



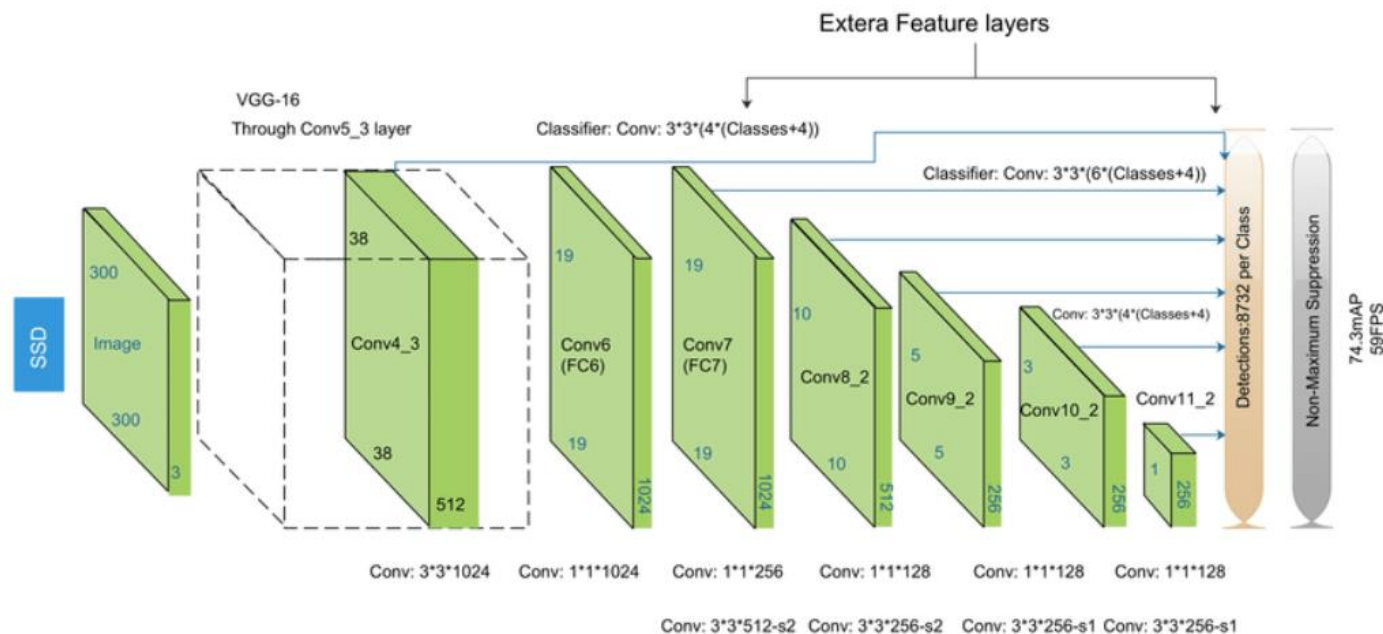
单阶段目标检测 - SSD

SSD (Single Shot MultiBox Detector) 是一种经典的单阶段目标检测算法，由Wei Liu等人于2016年提出。SSD结合了高效的目标检测方法和密集的目标预测，具有较高的检测精度和较快的检测速度。

如图所示的SSD结构，其是基于VGG-16构建，在模型的浅层SSD可以检测到较小的目标，而较深的层检测出较大的目标。

SSD 的一些关键特点：

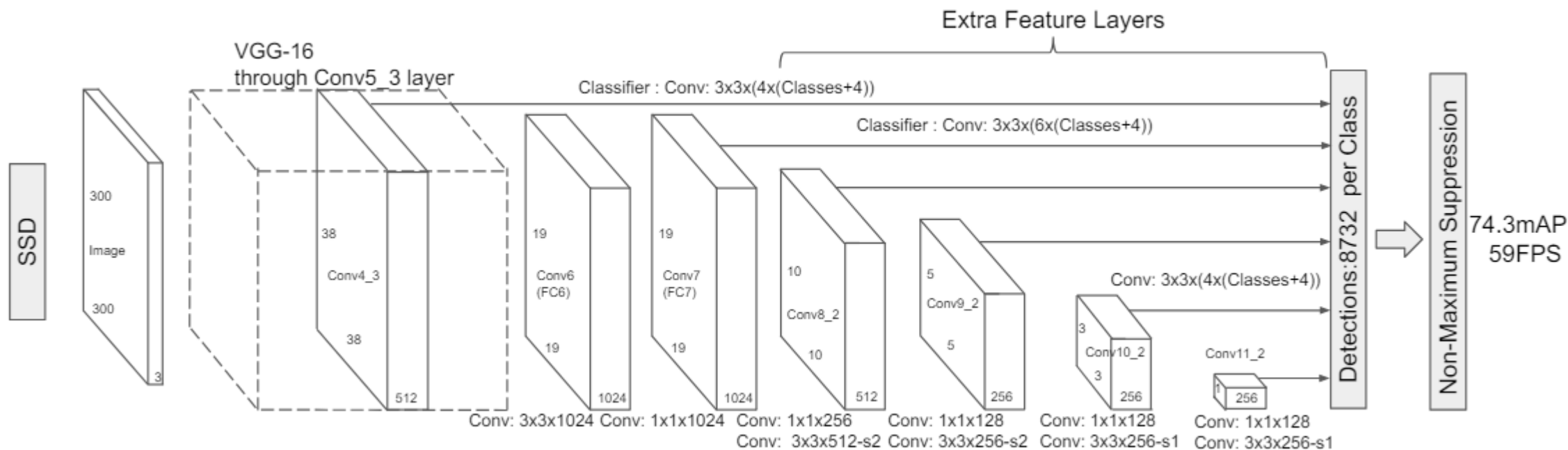
- **多尺度特征图**：SSD通过在不同层级的特征图上进行目标检测，可以有效地检测不同尺度和长宽比的目标。
- **先验框 (Prior Boxes)**：SSD引入了先验框的概念，即预定义的一系列框，用于在不同位置和尺度上进行目标检测。
- **多任务学习**：SSD采用了多任务学习的方式，同时优化目标类别的预测和目标位置的回归。
- **损失函数**：SSD使用了一个综合的损失函数，包括目标分类损失和目标定位损失。
- **实时性**：相比于一些传统的两阶段目标检测方法，SSD具有更快的检测速度。



单阶段目标检测 - SSD

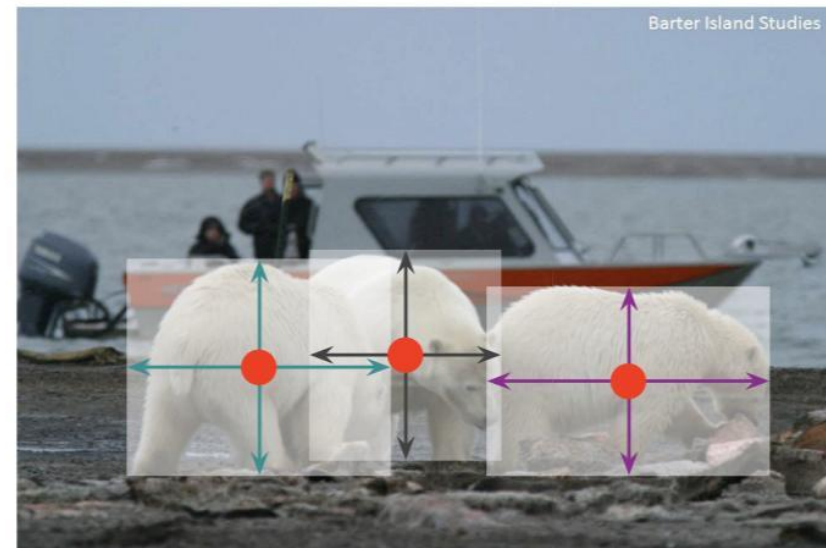
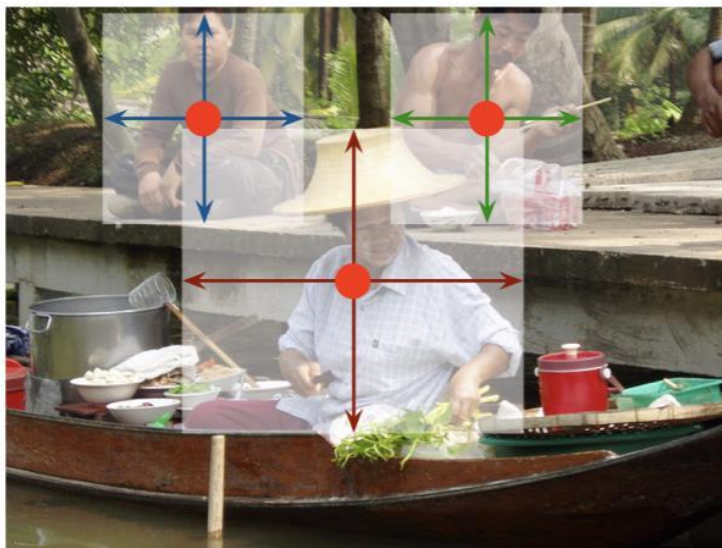
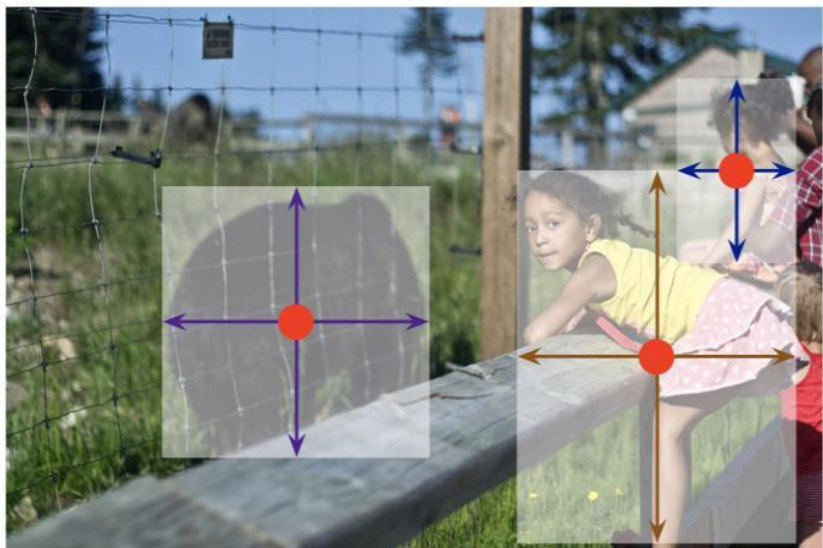
网络结构主要分为以下几个部分：

- **VGG16 Base Layer:** SSD网络以VGG16作为基础的特征提取层Base Layer，选取其中的Conv4_3作为第一个特征层用于目标检测
- **Extra Feature Layer:** 额外的添加了几个特征层用于目标检测，将VGG16中的FC7改成了卷积层Conv7，同时增加了Conv8、Conv9、Conv10、Conv11几个特征层。特征金字塔，在多个尺度上进行目标检测以提高检测精度。
- **Detection Layer:** 类别预测和框预测(框的表示由 (cx, cy, w, h) 来表示)。
- **NMS:** 用于消除重叠的边界框并选择最佳的边界框。



单阶段目标检测 - CenterNet (anchor-free)

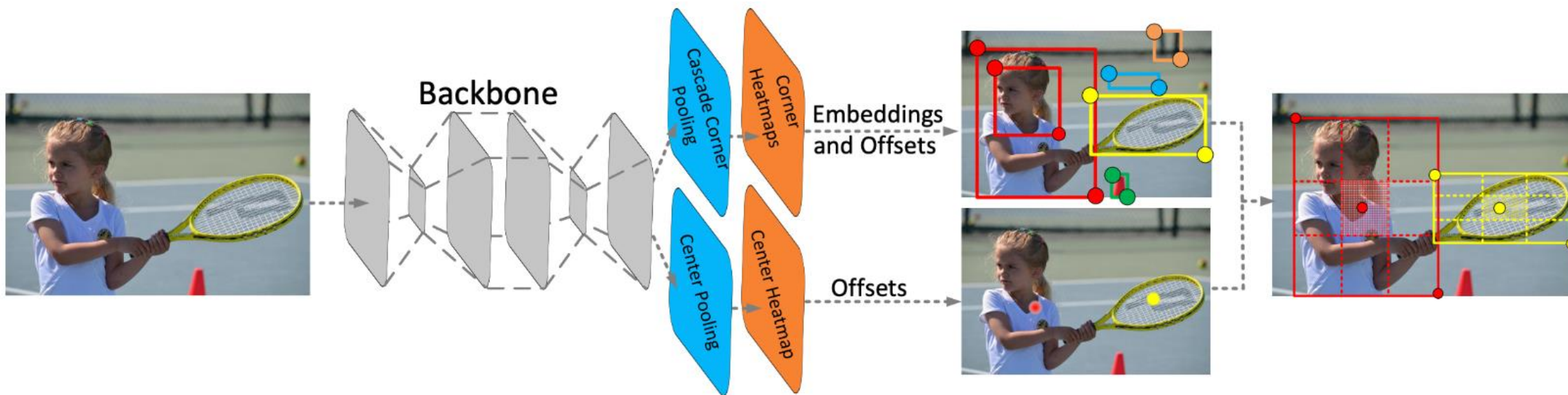
CenterNet核心思想是通过检测目标的中心点来实现目标检测，而不是直接预测边界框。这种方法使得CenterNet在目标检测任务中取得了很好的性能，同时具有较高的速度和精度。



单阶段目标检测 - CenterNet

网络结构主要分为以下几个部分：

- **骨干网络 (Backbone Network)**：通常采用一些流行的卷积神经网络（如ResNet、Hourglass等）作为基础网络，用于提取输入图像的特征。
- **中心点预测头 (Center-Point Head)**：这部分网络用于预测图像中每个目标的中心点。通常采用卷积神经网络结构，输出一个热图，其中每个点表示可能是目标中心点的概率。
- **边界框预测头 (Bounding Box Head)**：在检测到目标的中心点之后，通过预测目标的宽度、高度等信息来生成目标的边界框。





Thank

You