# YOLOP

**YOLOP: You Only Look Once for Panoptic Driving Perception**

论文地址：

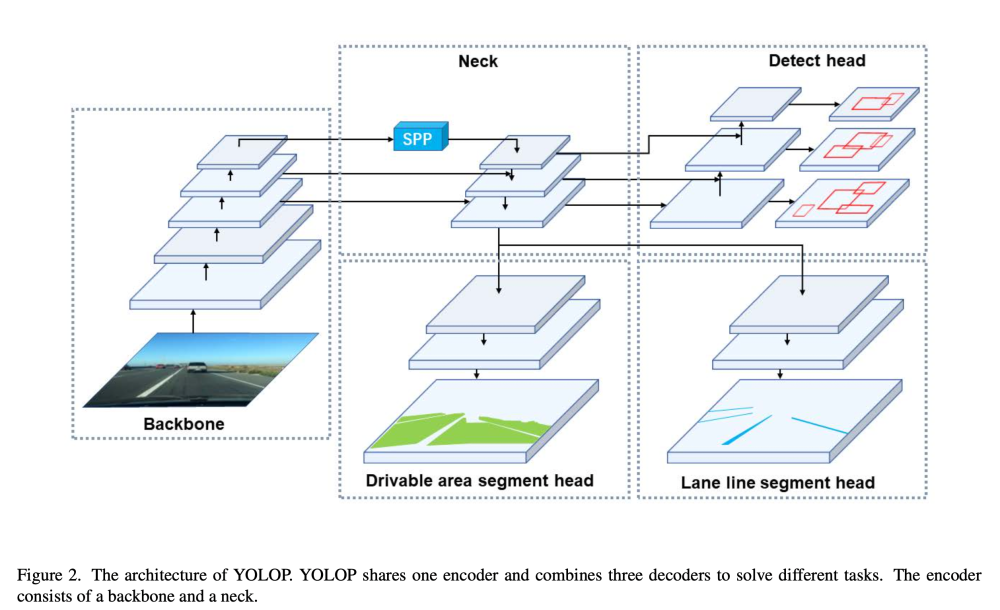
<https://arxiv.org/pdf/2108.11250.pdf>

代码地址：

<https://github.com/hustvl/YOLOP>

### 网络结构

YOLOP包含一个共享编码器去执行多任务，neck之后接了2D检测头（**Detect Head**），可行驶区域分割头（**Driable Area Segment Head**），车道线分割头（**Lane Line Segment Head**）。



#### Encoder

该网络中的编码器由骨干网络与Neck网络构成：

**Backbone** ：骨干网络用于提取输入图像的特征，考虑到YOLOv4在2D目标检测中的优异性能，选择CSPDarkNet作为骨干网络。

**Neck** ：Neck用于对骨干网络提取的特征进行融合，主要采用了SPP与FPN构建Neck模块。SPP用于生成融合不同尺度的特征；FPN则在不同语义层面融合特征，使得生成特征包含多尺度、多语义级信息。

#### Decoders

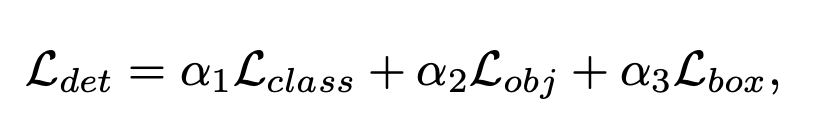
YOLOP包含三个用于三个任务的解码器：

**Detect Head** 类似YOLOv4，采用了基于Anchor的多尺度检测机制。首先，采用PAN进行更优特征融合，然后采用融合后特征进行检测：多尺度特征的每个grid被赋予三个先验anchor(包含不同纵横比)，检测头将预测位置偏移、高宽、类别概率以及预测置信度。

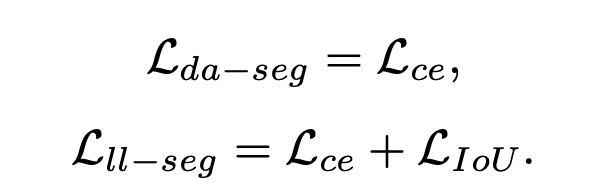
**Driable Area Segment Head & Lane Line Segment Head** 驾驶区域分割头与车道线分割头采用了相同结构。将FPN的输出特征(分辨率为[w/8, H/8,256])送入到分割分支。设计的分割分支非常简单，通过三次上采样处理输出特征尺寸为*(W, H, 2)*，代表每个像素是驾驶区域/车道线还是背景的概率。由于Neck中已包含SPP模块，并未像PSPNet添加额外的SPP模块。此外，采用了最近邻上采样层以降低计算量。因此，分割解码器不仅具有高精度输出，同时推理速度非常快。

#### Loss Function

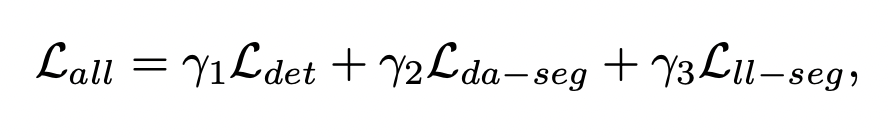
检测头loss：



可行驶区域和车道线检测loss：

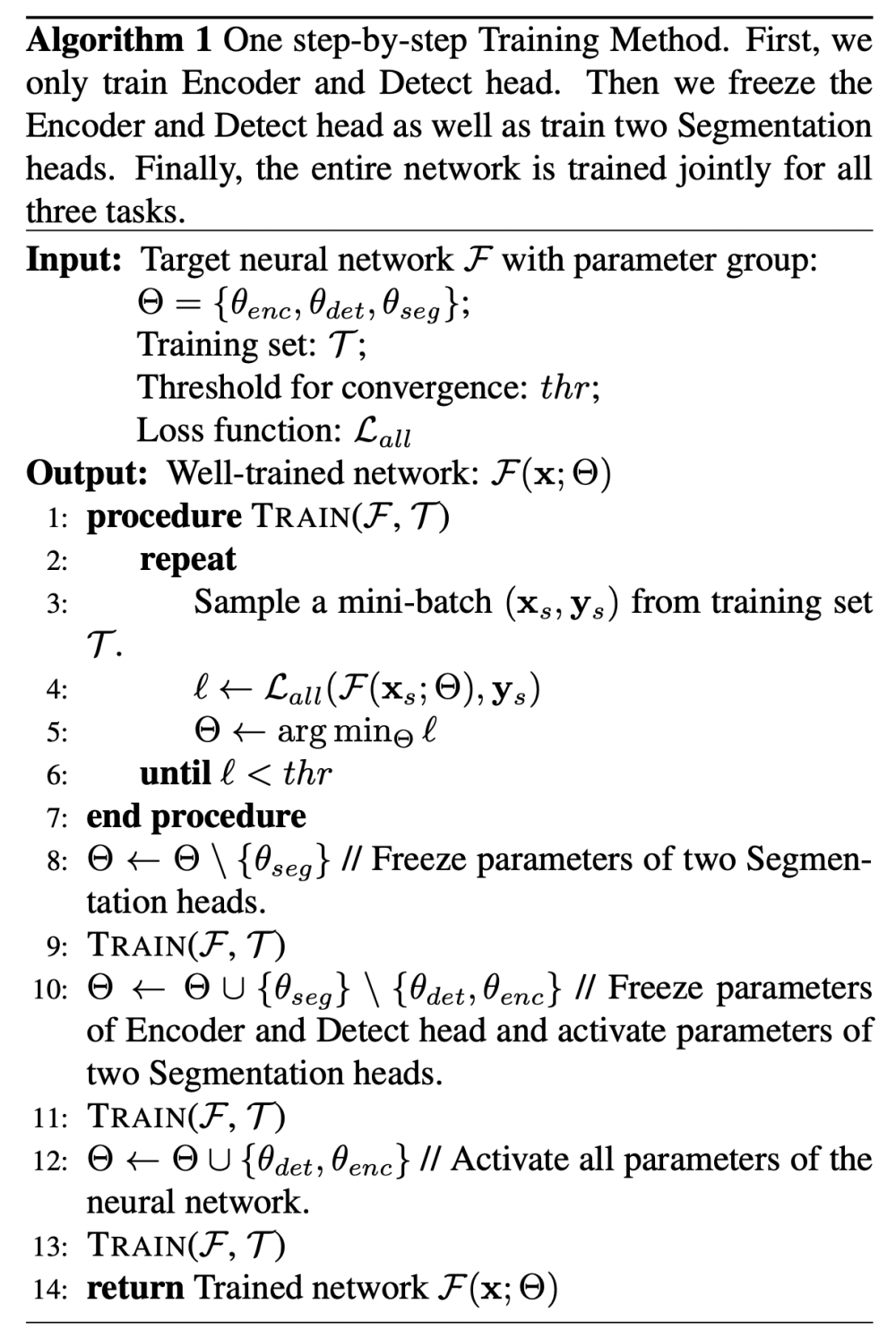


总体loss：



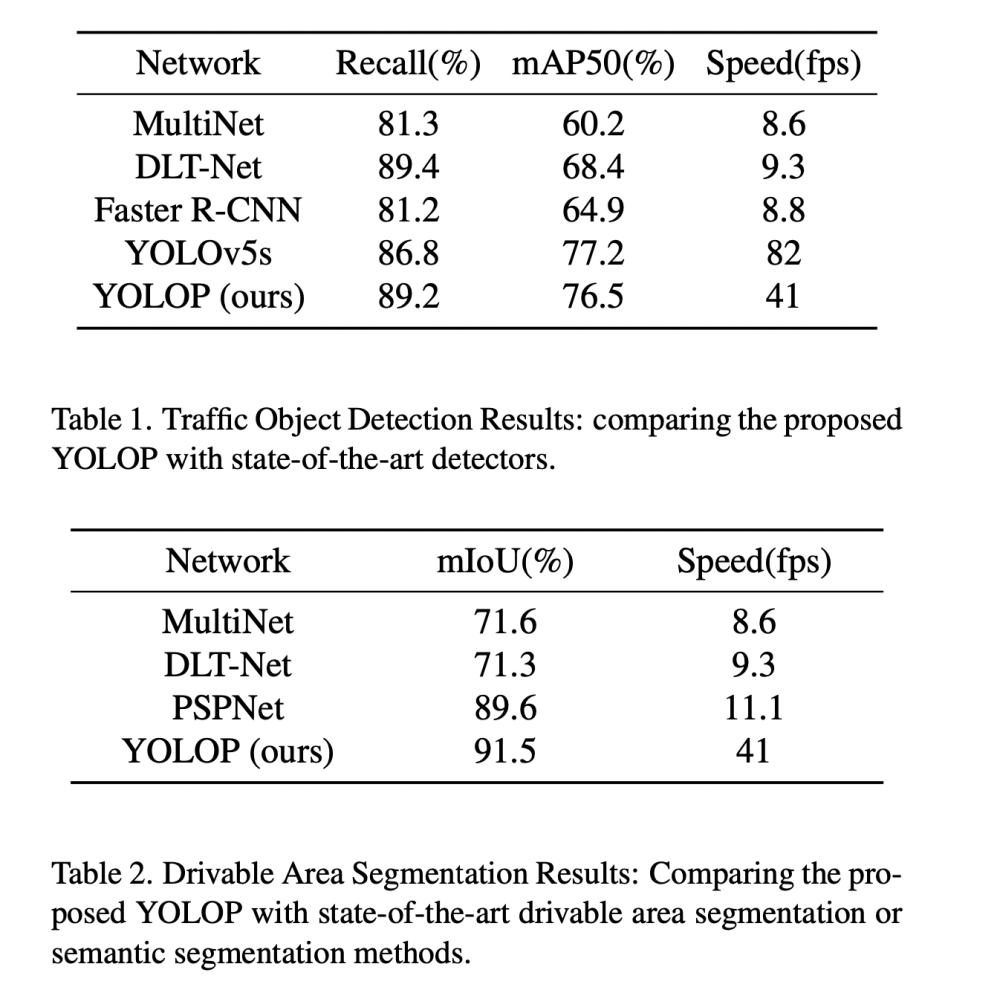
#### Training Paradigm

其尝试了不同范式训练上述模型。最简单的是端到端训练，然后三个任务可以进行联合学习。也尝试了交替优化算法：即每一步聚焦于一个或多个相关任务，而忽视不相关任务。下图给出了本文所用到的step-by-step训练方案。



### 结果

最后YOLOP在BDD100K上取得了较好结果，且计算速度明显更快。



### 效果

