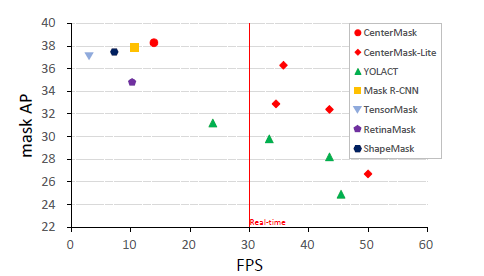
**CenterMask : Real-Time Anchor-Free Instance Segmentation**

**1. Backgroud**

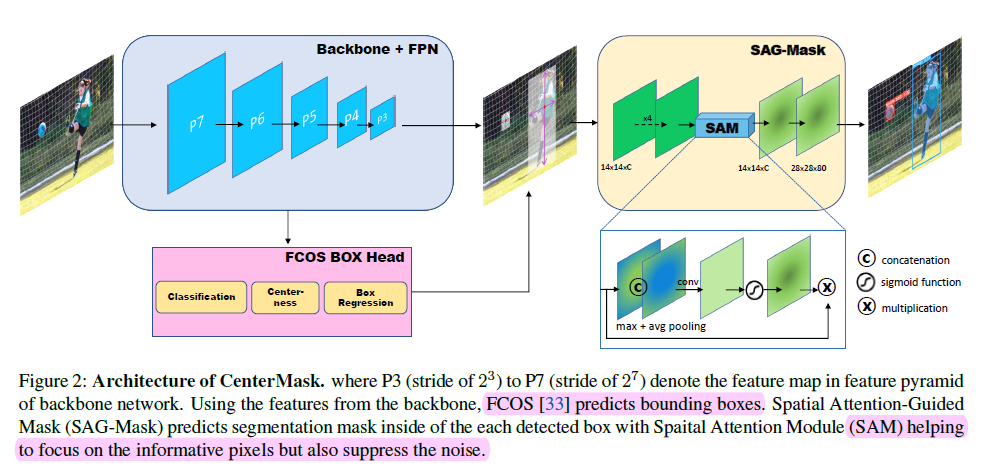
·Mask R-CNN（Faster-RCNN）与YOLACT（RetinaNet）最大的缺点就是检测器严重依赖于预先设置的锚框，该预定义锚对超参数（例如输入大小，纵横比，比例等）和不同的数据集非常敏感。此外，为了提高召回率，设置了很多锚框，从而导致正/负样本的不平衡以及较高的计算/内存成本。

·为了解决锚框的这些缺点，目前更倾向于通过使用角/中心点从锚框向无锚的方向转变，提出CenterMask（基于FCOS）这与基于锚框的检测器相比，效率更高，性能更好。



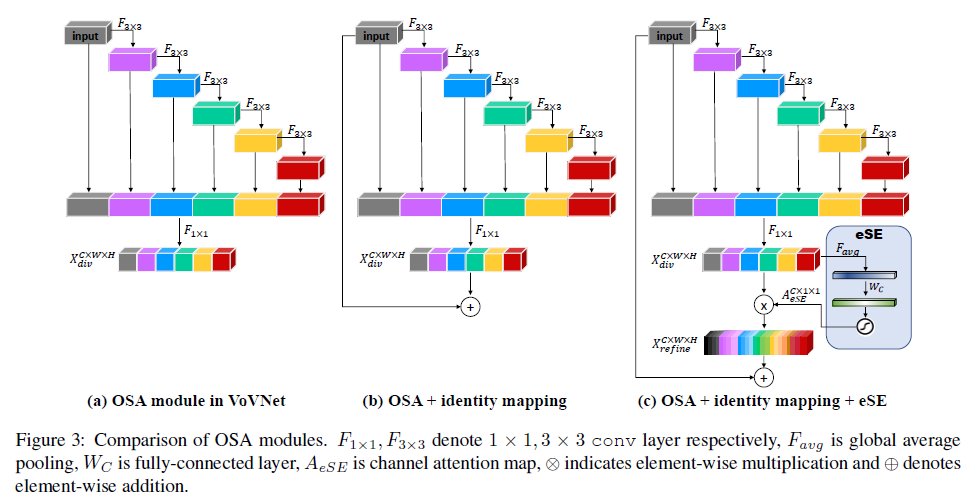
**2. Architecture**

CenterMask结构组成：➀BAckbone+FPN；➁FCOS BOX Head；➂SAG-Mask

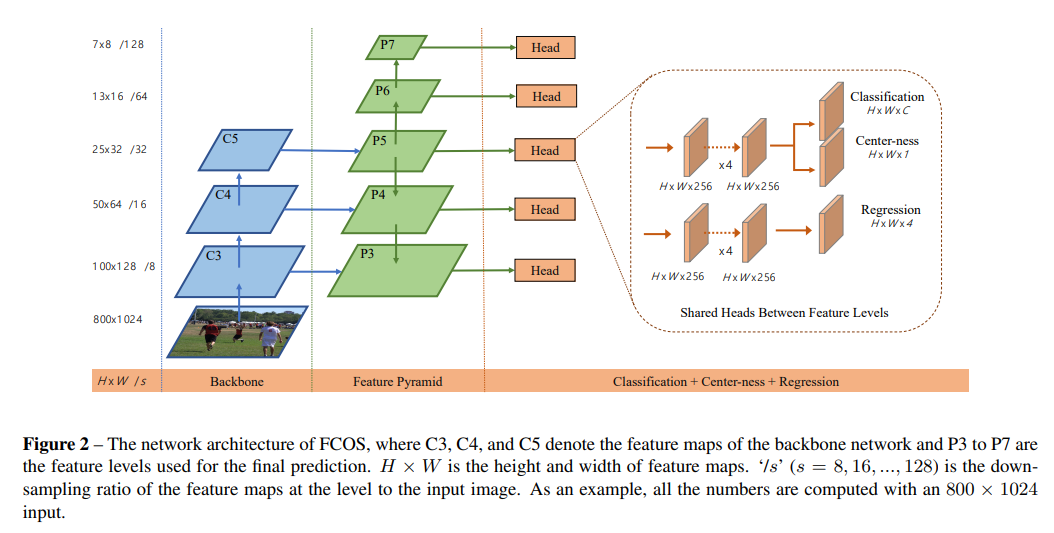


**BackBone**

VoVNetV2：基于VoVNet有两点改进,一是引入残差结构，二是重新设计了SE模块，将其中的2个FC替换为1个FC，得到了新的SE模块eSE。



**FCOS**



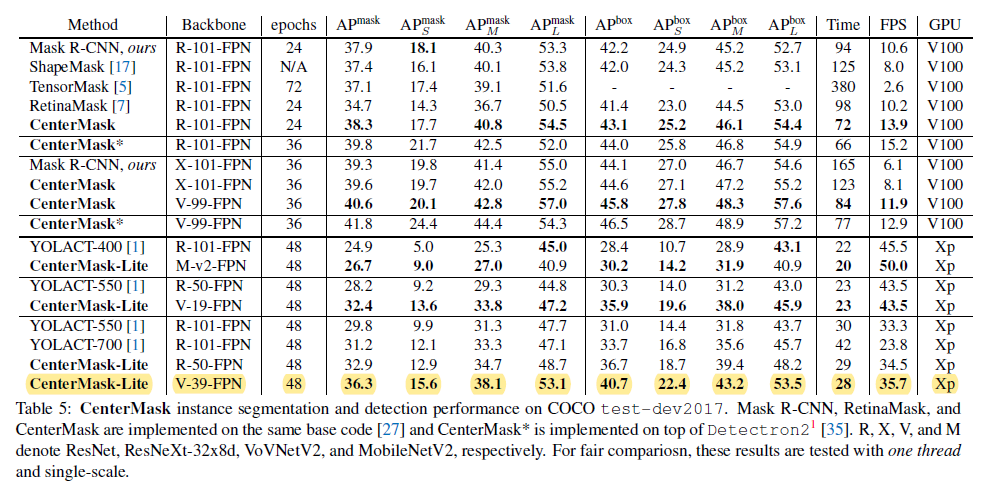
**SAG-Mask**

如上图所示，进行Max与Avg以及Conv等操作

**Mutil-loss**

L = Lcls+Lcenter+Lbox+Lmask，分别为分类损失、中心损失、回归损失、Mask损失

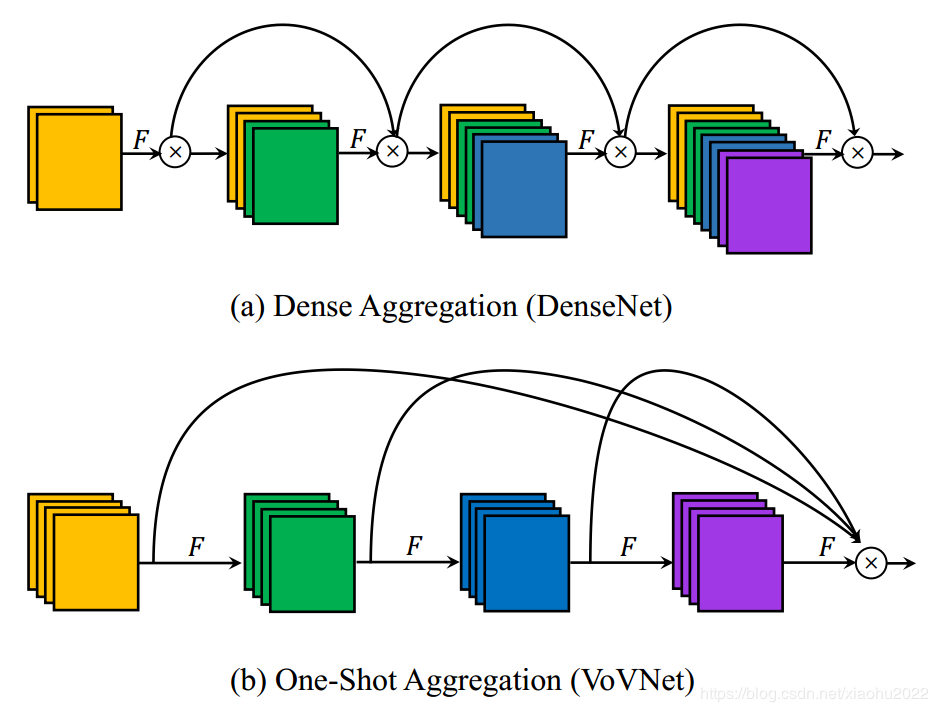
1. **Result**

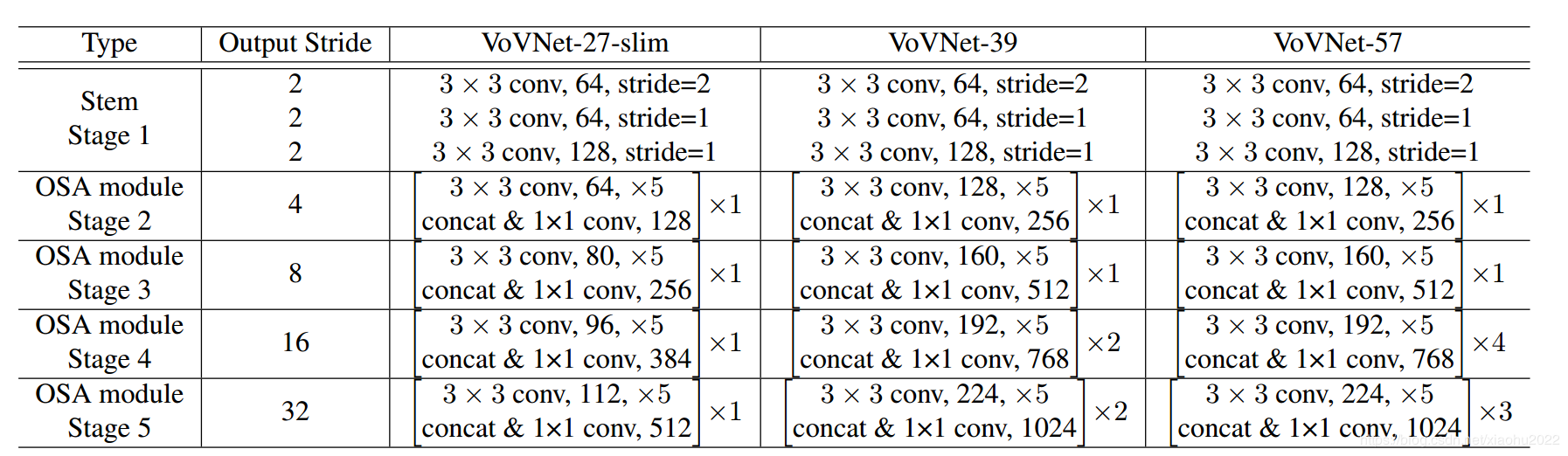


**VoVNet**

高效网络设计要素:（1）计算量（2）参数量（3）内存成本访问，当输入和输出的channel数相同时MAC才取下界（4）GPU计算效率，相比FLOPs，我们更应该关注的指标是FlOPs per Second，即用总的FLOPs除以总的GPU推理时间，这个指标越高说明GPU利用越高效。

VoVNet首先是一个由3个3x3卷积层构成的stem block，然后4个阶段的OSA模块，每个stage的最后会采用一个stride为2的3x3 max pooling层进行降采样，模型最终的output stride是32。与其他网络类似，每次降采样后都会提升特征的channel数





**FCOS**

点（x,y）落在ground truth之内，并且类别与真实标签一致，才被视作正样本。除了分类以外，坐标回归是一个4D实向量（l,t,r,b）,代表中心坐标到边界框四条边的距离。

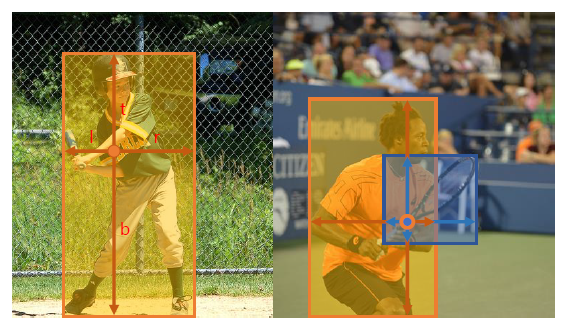
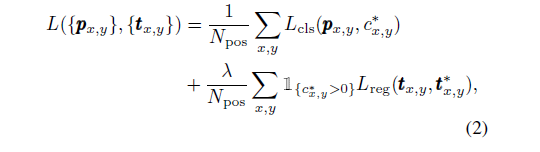
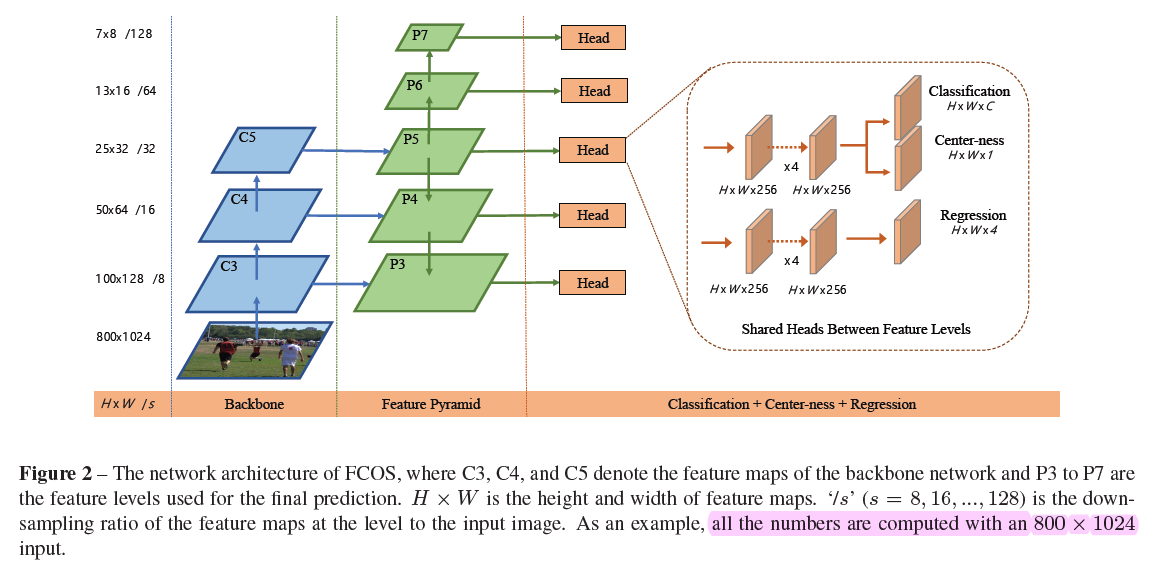


Figure 1 – As shown in the left image, FCOS works by predicting a 4D vector (l; t; r; b) encoding the location of a bounding box at each foreground pixel (supervised by ground-truth bounding box information during training). The right plot shows that when a location residing in multiple bounding boxes, it can be ambiguous in terms of which bounding box this location should regress.

**Loss Function**





对于基于锚的检测器，可以通过降低正锚框所需的IOU分数在一定程度上补偿由于步幅较大而导致的较低召回率。