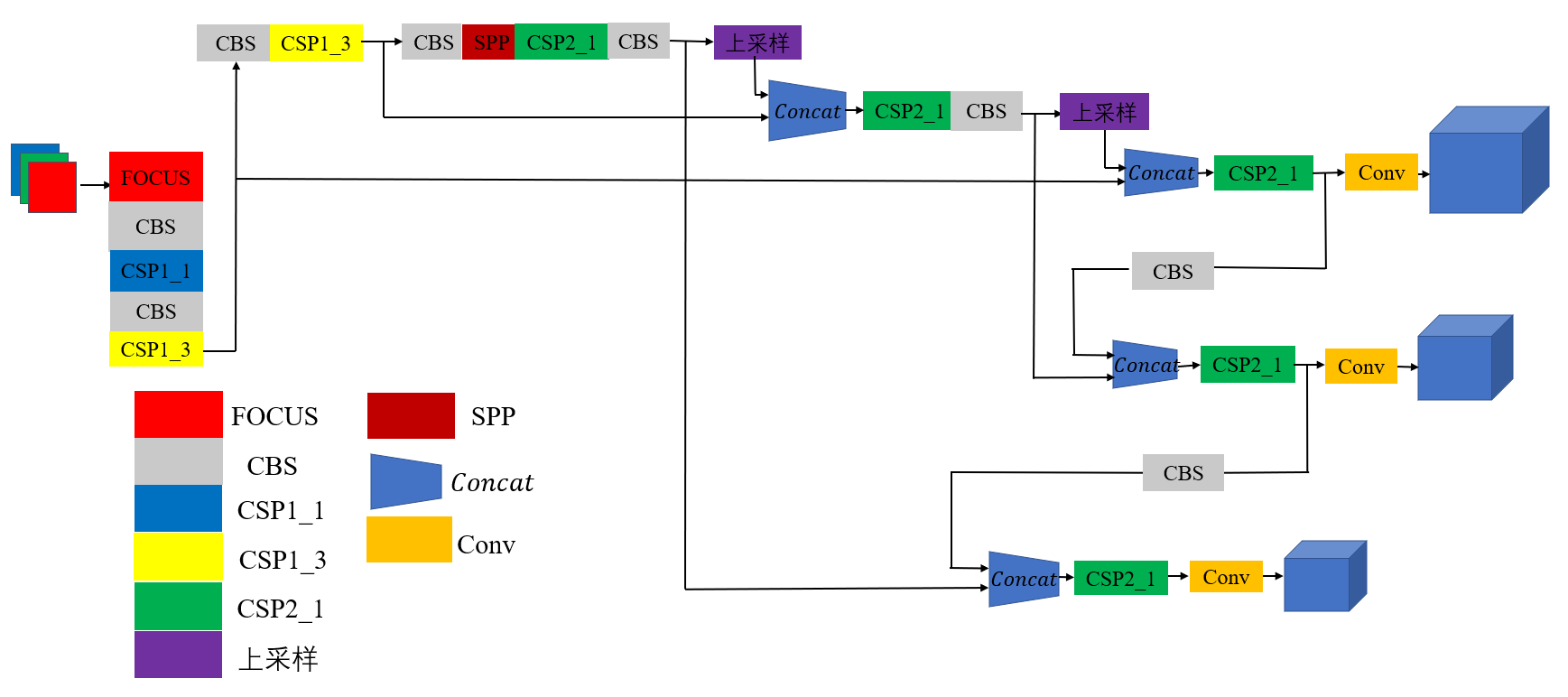
YOLOv5网络结构学习

YOLOv5网络结构图



上图中Conv为卷积结构，金字塔池化（Spatial Pyramid Pooling）结构用于将特征图转化为固定大小的特征向量，CBS模块则为卷积加BN（Batch Normalization）加激活函数的套件。其中BN的作用为在深度学习过程中保证每一层神经网络的输入的分布相同，用于解决随训练网络深度加深导致收敛速度变慢的问题。CSP结构为残差结构，为增加层与层之间的反向传播梯度值，避免网络加深而造成的梯度消失。Concat为特征融合操作。

YOLOv5网络主要有四部分组成，分别为输入端、Backbone（主干网络）、Neck（颈部网络）、Prediction（预测端）。

输入端包含Mosaic数据增强、自适应锚框计算、自适应图片缩放三部分：

(1) 其中YOLOv5的输入端采取的数据增强方式与YOLOv4相同，都使用了Mosaic方式。Mosaic数据增强以CutMix数据增强方式为基础，使用四张图片进行随机缩放、裁剪、排布方式对图片进行拼接。增加了小型识别目标，提升了在对小目标识别时的效果。

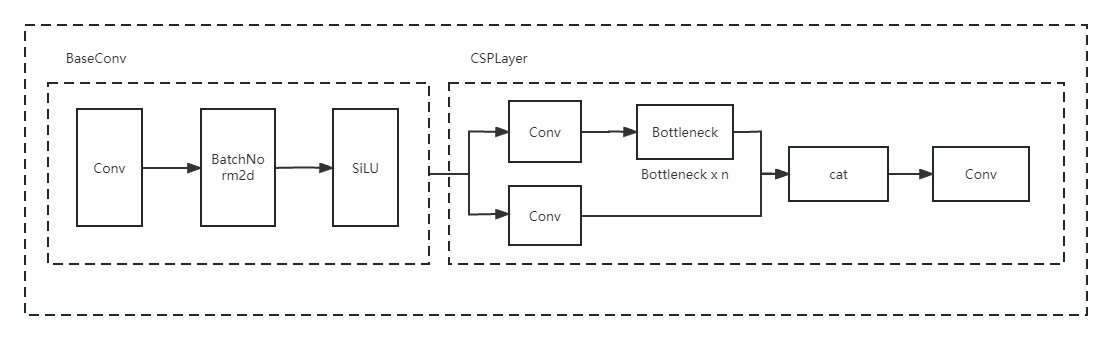
(2) 在使用YOLO算法进行网络训练前，需设定锚框长宽。对不同数据集，YOLO都会设定一个不同的锚框长宽。例如，YOLOv5设定的初始锚框值为：[116,90,156,198,373,326]、[30,61,62,45,59,119]、[10,13,16,30,33,23]，与YOLOv3、YOLOv4不同，YOLOv5将计算锚框初始值的步骤加入到代码中，不需要单独计算，而是在每次训练时会计算对于不同训练集的锚框值，将其与真实值做比对，以此来不断迭代更新网络参数。

(3) 在通常情况下，使用目标检测算法的图片大小、长宽比都不相同，因此在将图片送入网络之前，需要将图片进行处理，调整为同意尺寸再送入检测网络。在进行缩放调整时，检测的图片通常会产生黑边，若黑边在图片中所占比例较大则会产生信息冗余从而使推理速度下降，影响性能。在YOLOv5中对图像缩放时会自适应的为原始图片添加少量黑边，从而减少缩放后图片中黑边所占比例，以此减少推理过程中的计算量，增强性能。

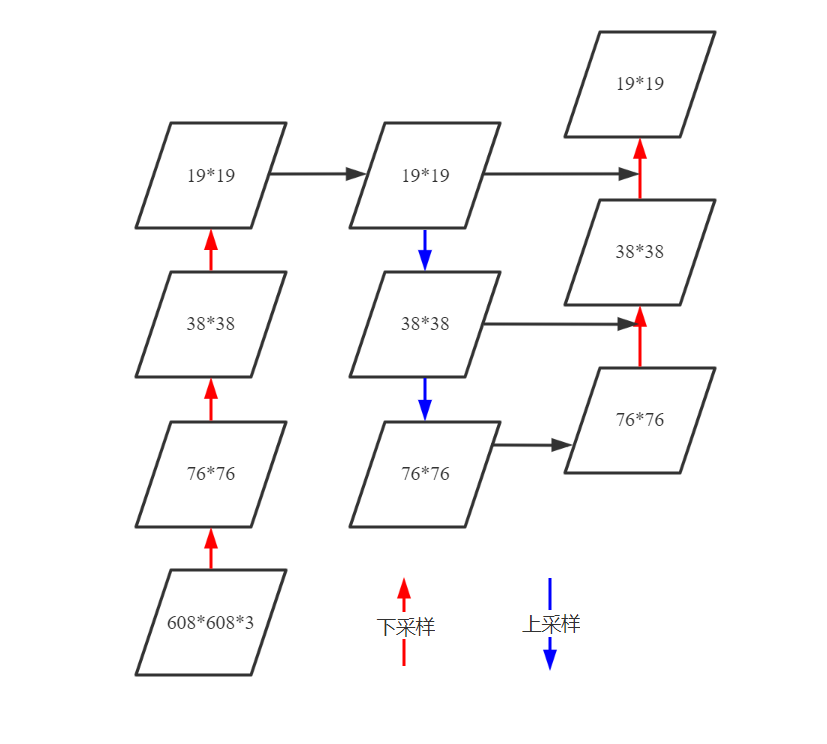
Backbone主干网络主要包含Focus和CSP两个结构。在YOLOv5中相较之前版本多出了Focus结构，用于切片操作。以YOLOv5s为例：将输入的原始图像由608\*608\*3切片为304\*304\*12的特征图，再通过32个卷积核计算变为304\*304\*32的特征图。与YOLOv4相比，v5版本增加了一个CSP结构分别用于主干网络和颈部网络。提取出大量特征信息用于解决梯度重复问题。

Neck颈部网络主要使用FPN+PAN结构，使用上采样操作来得到特征图，将自下而上的特征金字塔运用在PAN中。

CSP结构：

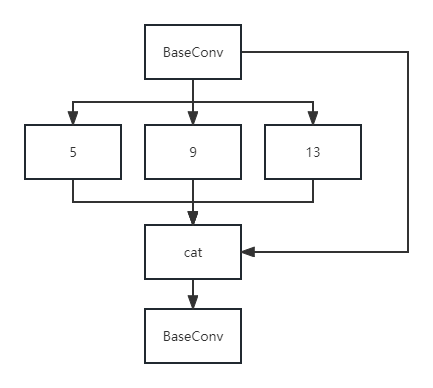


FPN结构：



YOLOv5中对输入图采用3\*3的卷积核且步长为2，进行卷积处理得到各不同层数的特征图，在FPN中将上层所包含的较强的语义信息通过上采样自顶向下传达融合，在金字塔结构中则是将每一级的特征映射作为输入使用3\*3的卷积核处理，且其输出通过横向连接加入到FPN路径中的同级特征图中。融合位置信息及特征信息。

SPPBottleneck：



Dark5部分添加了BottleNecck模块，在BaseConv与CSPLayer中添加了SPPBottleneck部分，这部分通过使用3个不同池化核大小的最大池化进行特征提取再合并，以提高网络的感受野。