## 数据获取，预处理，模型框架，实验结果与分析（消融实验）

#### 数据获取

##### 直接通过系统导入高质量的钢材表面图片。

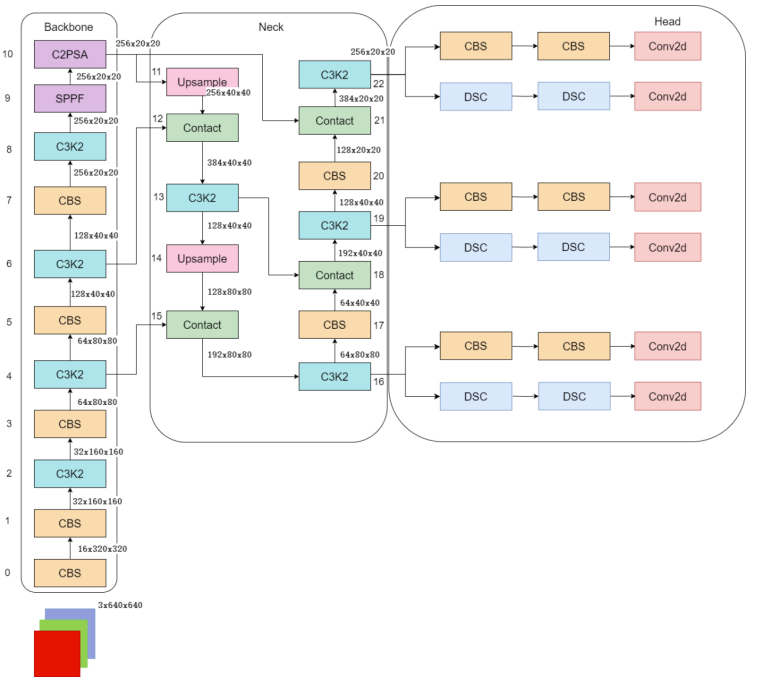
##### 通过摄像头现场拍摄图片并实时检验。

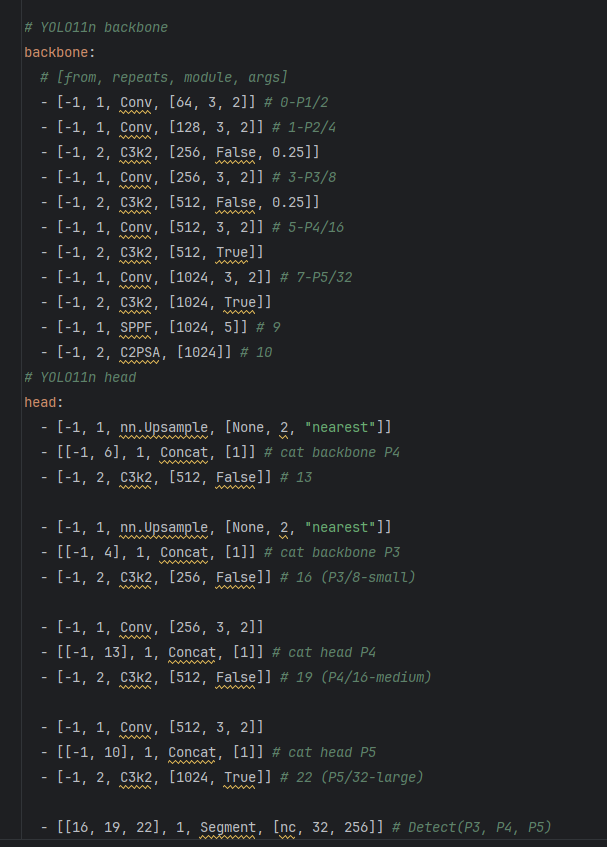
#### 数据预处理

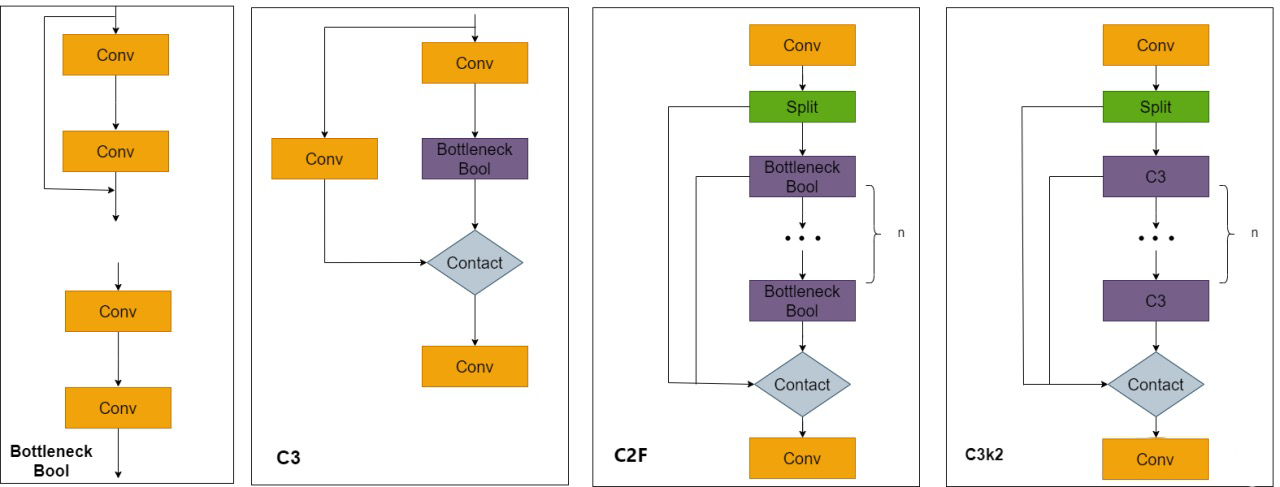
对于谢韦尔数据集我们进行了文本和图像上处理，首先读取csv文本数据，对数据标注进行了解读，并根据需要对标注数据进行了格式转化，然后，根据标注信息对一万多张图像进行了可视化，通过分类统计出无缺陷钢板数量有5902张，将无缺陷图像去除；有缺陷的6666张，在设计训练集和验证集时，缺陷图像1: 897, 3: 5150, 4: 801, 2: 247，由于不同缺陷之间数量极不平衡，对不同类型的缺陷采用了不同程度的数据增强(旋转，切割，合并，噪音等)。同时也度东北大学的NEU-DET数据集进行了标签合并。将两个数据集合并。

#### 模型框架

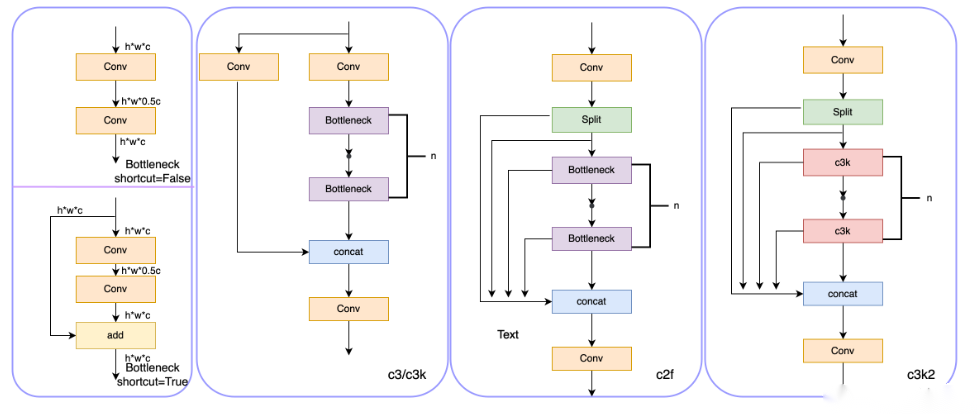
##### YOLOV-11-seg模型网络架构



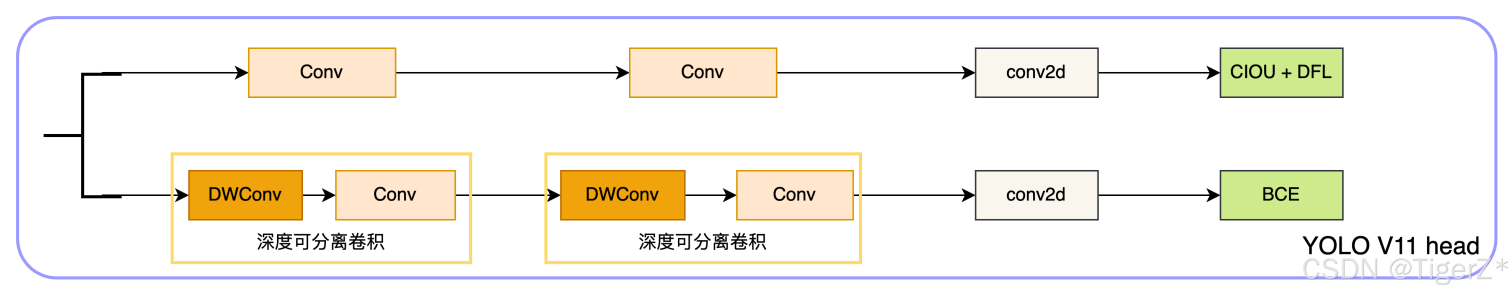




主干网络（backbone）



颈部网络（neck）



头部网络（head）

| 层索引 | 模块类型 | 参数说明 | 功能解析 | 输出特征图尺度变化 |
| --- | --- | --- | --- | --- |
| 0 | Conv | [64, 3, 2] | 64通道，3x3卷积，步长2 | 输入尺寸的1/2 |
| 1 | Conv | [128, 3, 2] | 128通道，3x3卷积，步长2 | 输入尺寸的1/4 |
| 2 | C3k2 | [256,False,0.25] | 含2个C3模块，256通道 | 保持尺寸（1/4） |
| 3 | Conv | [256, 3, 2] | 256通道，步长2 | 输入尺寸的1/8 |
| 4 | C3k2 | [512,False,0.25] | 含2个C3模块，512通道 | 保持尺寸（1/8） |
| 5 | Conv | [512, 3, 2] | 512通道，步长2 | 输入尺寸的1/16 |
| 6 | C3k2 | [512, True] | 含2个C3模块，启用Shortcut | 保持尺寸（1/16） |
| 7 | Conv | [1024, 3, 2] | 1024通道，步长2 | 输入尺寸的1/32 |
| 8 | C3k2 | [1024, True] | 含2个C3模块，启用Shortcut | 保持尺寸（1/32） |
| 9 | SPPF | [1024, 5] | 空间金字塔池化，池化核5 | 保持尺寸（1/32） |
| 10 | C2PSA | [1024] | 融合PSA注意力机制的CSP模块 | 保持尺寸（1/32） |

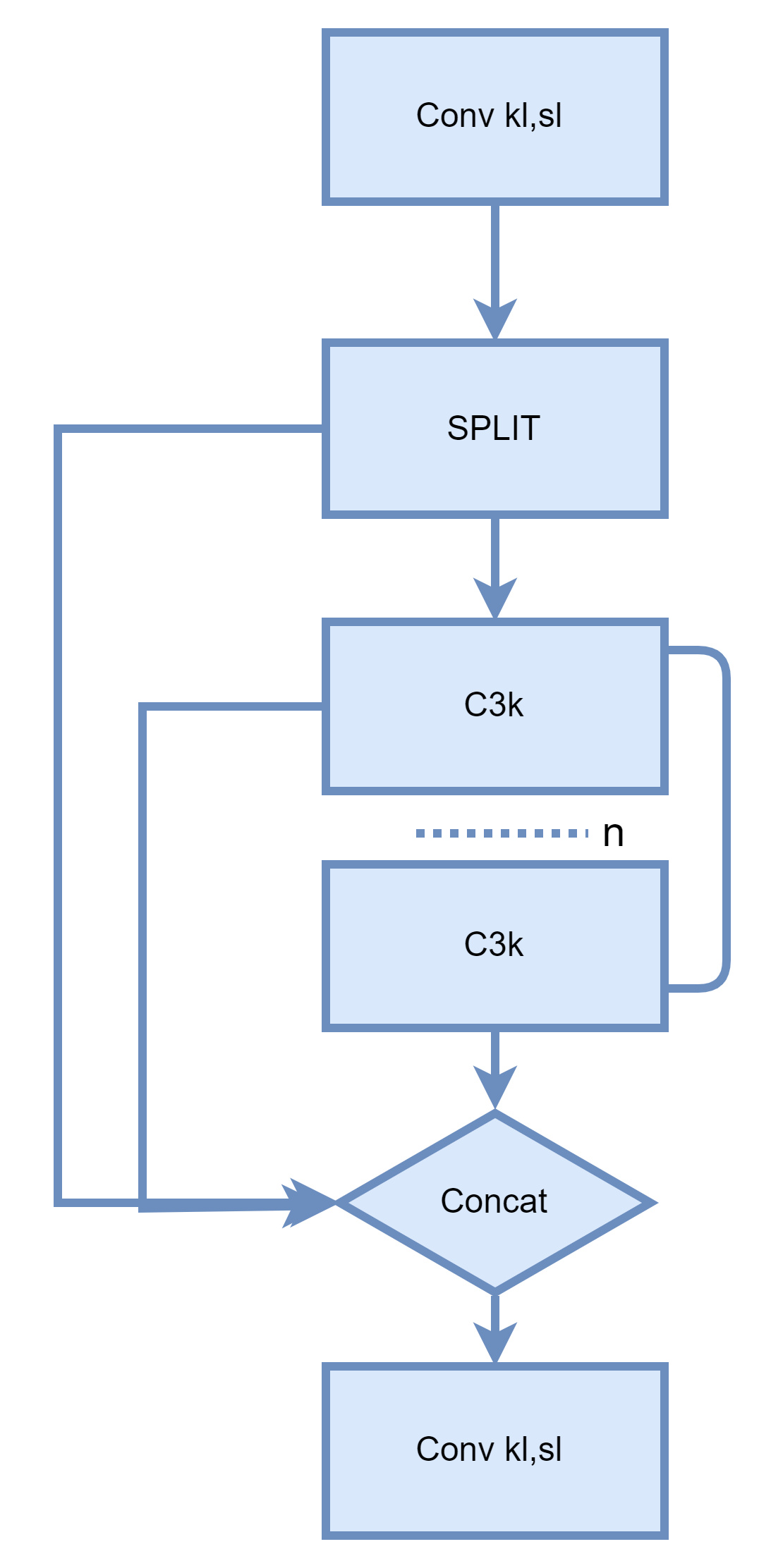
其中yolov11使用了C3K2模块（提取不同层次的特征），有以下两个部分构成:

ConvModule：基础卷积模块，包含卷积层、归一化层和激活函数，用于特征提取。

Bottleneck堆叠(选择是否进行残差链接)：多个Bottleneck模块串联，每个Bottleneck可选择是否集成C3k模块(允许自定义卷积核大小，适应不同尺寸的图像特征)。

无C3k模块时：Bottleneck结构与YOLOv8的C2f(加速处理)模块(堆叠多个卷积层实现跨通道特征融合)类似，包含两个卷积层和残差连接。

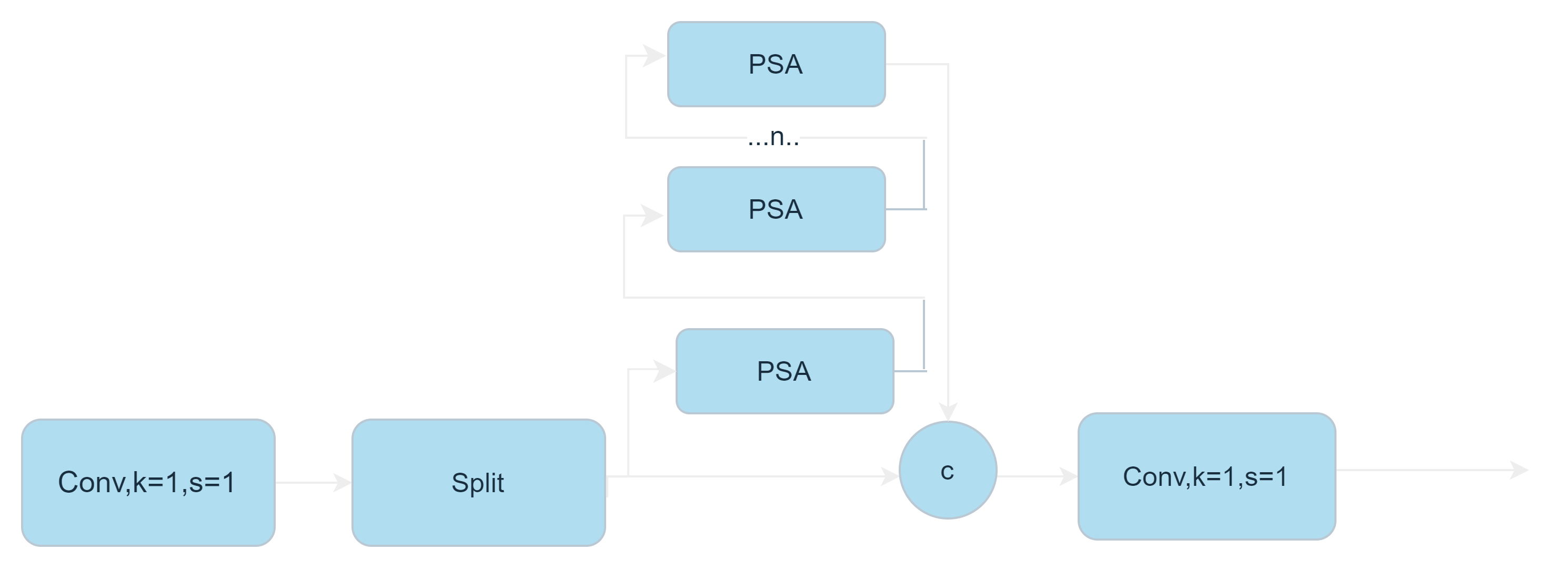
集成C3k模块时（YOLO11改进点）：Bottleneck内部新增3个卷积层和子Bottleneck模块，增强特征融合能力



yolov11也使用了C2PSA ，该模块是 YOLO11 中用于增强[特征提取](https://so.csdn.net/so/search?q=%E7%89%B9%E5%BE%81%E6%8F%90%E5%8F%96&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/m0_49699378/article/details/_blank)的一个高级模块，结合了 CSP (Cross Stage Partial) 结构和 PSA (Pyramid Squeeze Attention) 注意力机制，从而提升了多尺度特征提取能力。

CSP结构是将原输入分成两个分支，分别进行卷积操作使得通道数减半，然后一个分支进行Bottleneck \* N操作，然后concat两个分支，使得BottlenneckCSP的输入与输出是一样的大小。可以学习到更多特征

其中PSA使用不同大小的卷积核（使用3×3、5×5、7×7等卷积核并行提取特征，生成多尺度特征图），每个卷积核能够在不同的感受野上提取特征，从而捕捉多尺度信息。卷积核越大，感受野越大，适合检测大物体；卷积核越小，适合捕捉小物体的细节，同时结合 Squeeze-and-Excitation (SE) 模块(动态调整每个特征通道的权重来提升模型的表示能力，从而使模型能够更加关注重要的特征并忽略不相关的信息)对特征通道进行加权，从而增强网络对不同尺度目标的注意力聚焦。



YOLOv11的Neck（特征融合图）是目标检测模型中连接Backbone（主干网络）和Head（检测头）的核心模块，其核心功能是通过多尺度特征融合与信息交互优化，提升模型对目标特征的表达能力。接收来自Backbone的不同层级特征图（例如P3/P4/P5），通过\*\*上采样（Upsample）和拼接（Concat）\*\*操作，将低分辨率语义信息与高分辨率细节信息融合。

Head部分 进行特征解码与预测生成

输入：接收来自Backbone（主干网络）和Neck（特征融合层）的多尺度特征图。

输出：生成以下三类预测结果：

边界框（Bounding Box）：目标的位置和大小（通常表示为中心点坐标、宽高）。

类别概率（Class Probability）：目标属于每个类别的置信度。

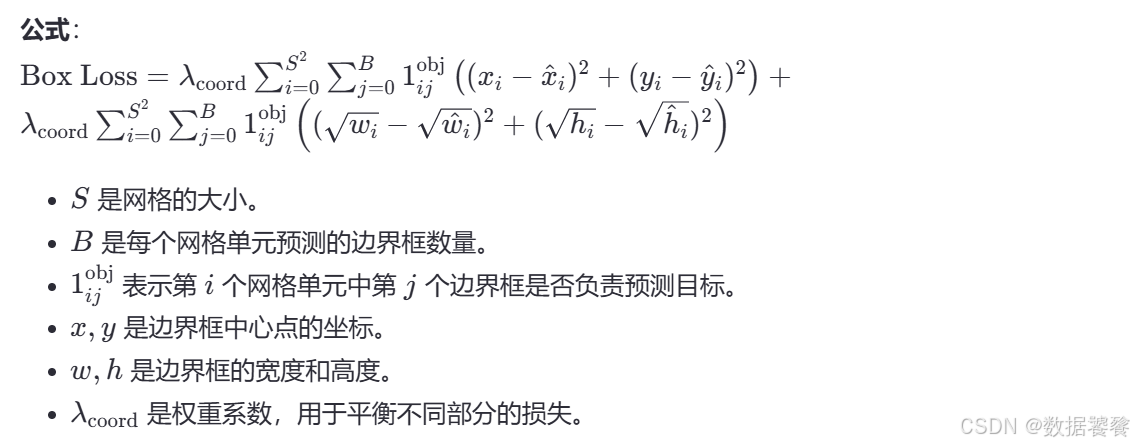
置信度（Confidence Score）：预测框内存在目标的概率。

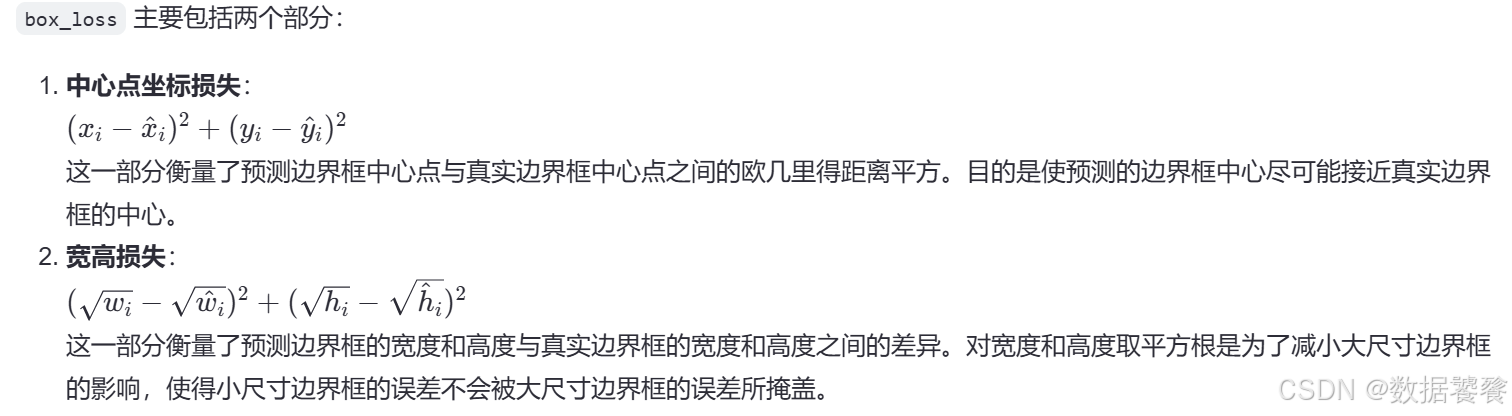
Head部分进行损失函数对接

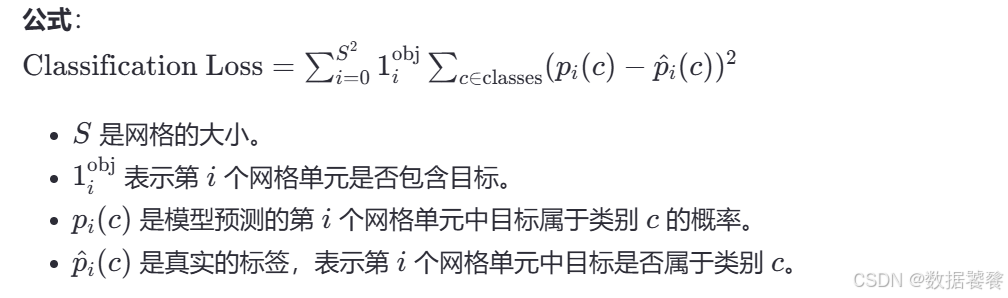
定位损失：box\_ loss（考虑重叠区域、中心点距离、长宽比）。

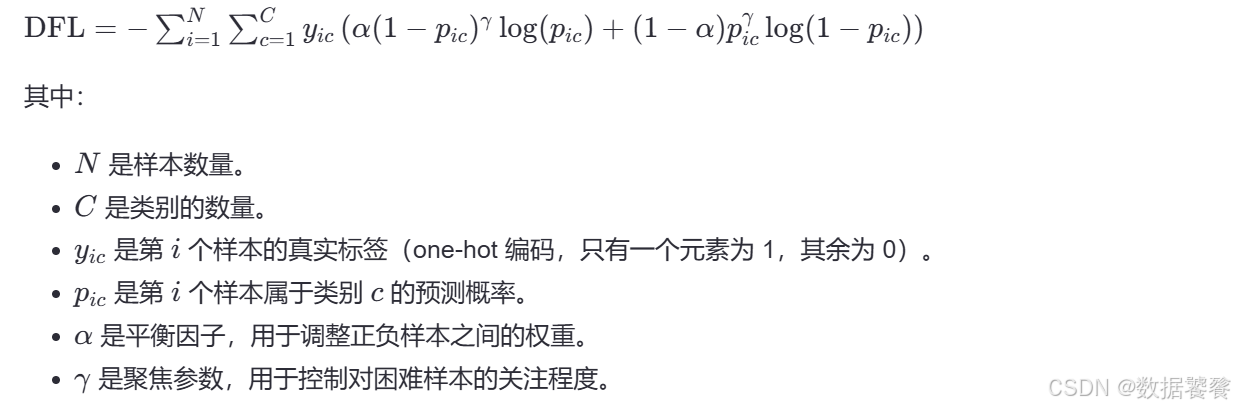
分类损失：cls\_loss（解决类别不平衡问题）。

置信度损失：dfl\_loss二元交叉熵（区分前景与背景）。











YOLOv11 的总损失函数由 ​定位损失（PWIoU + DFL-Pro）​、分类损失（MetaVFL）​、置信度损失（TALv4）​ 及 ​自监督对比损失 构成：

#### 消融实验

以yolov11-seg模型为基础模型，进行了两方面改进，引入多尺度卷积注意力模块MSCAM以及添加了自适应阈值焦点损失（ATFL）。设计了yov11-seg，yov11-seg+MSCAM，yov11-seg+ATFL，yov11-seg，+MSCAM+AFTL四个实验组。

自适应阈值焦点损失（ATFL）是一种动态调整损失权重的损失函数，通过降低易分类样本的影响，增强对难分类样本的关注，从而提升模型性能，特别是在类别不平衡的情况下。

多尺度卷积注意力模块MSCAM可以在增强特征图的同时降到底计算成本，提升模型性能。