# **Yolo11-seg**

**YOLO11文档：**[**https://docs.ultralytics.com/models/yolo11/**](https://docs.ultralytics.com/models/yolo11/)

**YOLO11 论文：https://arxiv.org/abs/2410.17725**

数据处理：

1. 根据任务要求按照‘metadata\_splits.csv’文件中’split\_open’列对数据集进行划分，分成三类数据集Train，Test，Val。
2. 将COCO格式数据集转换成Yolo所需要的格式。具体步骤如下
   1. 读取train.json文件（也就是annotation）
   2. 处理annotation中用RLE编码的segmentation字段
   3. 生成与图片一一对应的mask（掩码）文件 以YOLO所需要的txt格式保存
   4. 至此YOLO数据集处理完毕

图片包含 表格

描述已自动生成

**Train：使用YOLOv11-seg 进行训练**

**Val(验证)：**

**展示IOU结果和混淆矩阵**

图表, 条形图

描述已自动生成

**Predict:**

**展示的是实际mask和预测mask的对比**

**Deeplabv3+resnet101**

Deeplabv3论文：https://arxiv.org/pdf/1802.02611

可以提一提迁移学习这个概念，因为这个模型就是在torch库里（from torchvision.models.segmentation import deeplabv3\_resnet101, DeepLabV3\_ResNet101\_Weights） 别人预训练好的分割模型，在这基础上进行修改。

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

表格

描述已自动生成

**模型架构逻辑可以了解一下：DeepLabv3 Network Architecture**

**DeepLabv3 is specifically designed for semantic segmentation, focusing on assigning a class label to every pixel in an image. Its architecture can be divided into several key components:DeepLabv3 专为语义分割而设计，专注于为图像中的每个像素分配一个类标签。它的架构可以分为几个关键组件：**

**1. Backbone Network (Feature Extractor):**

* **DeepLabv3 typically uses deep CNNs like ResNet as the backbone.DeepLabv3 通常使用 ResNet 或深度 CNN 作为主干。**
* **These networks are responsible for extracting hierarchical features from the input image.这些网络负责从输入图像中提取分层特征。**
* **DeepLabv3 incorporates dilated/atrous convolutions in the later layers of the backbone to:DeepLabv3 在主干的后几层中整合了扩张/闭孔卷积，以：**
  + **Expand the receptive field without reducing the spatial resolution.在不降低空间分辨率的情况下扩大感受野。**
  + **Preserve finer details necessary for segmentation.**

**2. Atrous Spatial Pyramid Pooling (ASPP):2. Atrous 空间金字塔池化 （ASPP）：**

**The ASPP module is a hallmark of DeepLabv3, designed to capture multi-scale contextual information.ASPP 模块是 DeepLabv3 的标志，旨在捕获多尺度上下文信息。**

* **Parallel Atrous Convolutions:**
  + **Uses multiple convolutions with different dilation rates to capture features at varying scales.使用具有不同膨胀率的多个卷积来捕获不同比例的特征。**
  + **Ensures robust detection of objects with different sizes.确保可靠地检测不同大小的物体。**
* **Image-Level Context:图像级上下文：**
  + **Adds a global average pooling branch to encode image-wide context information.添加全局平均池化分支以对图像范围的上下文信息进行编码。**
* **1x1 Convolutions:**
  + **Reduces dimensionality and combines multi-scale features.降低维度并组合多尺度特征。**

**ASPP outputs a feature map that aggregates information from different scales, providing a rich context for segmentation.ASPP 输出一个特征图，该图聚合了来自不同尺度的信息，为分割提供了丰富的上下文。**

**3. Decoder (in DeepLabv3+):**

**While DeepLabv3 does not have a dedicated decoder, DeepLabv3+ extends it by introducing a decoder module to refine segmentation outputs:虽然 DeepLabv3 没有专用的解码器，但 DeepLabv3+ 通过引入解码器模块来优化分割输出，从而对其进行了扩展：**

* **Combines low-level features from earlier layers of the backbone with high-level features from ASPP.将主干网早期层的低级功能与 ASPP 的高级功能相结合。**
* **Uses upsampling (bilinear interpolation) to recover spatial resolution and improve boundary precision.使用上采样 （双线性插值） 来恢复空间分辨率并提高边界精度。**

**4. Final Classifier:**

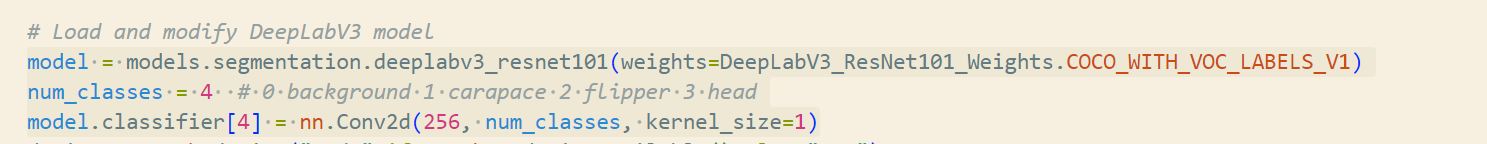
* **The output of the ASPP (or decoder in DeepLabv3+) is passed through a 1x1 convolutional layer to produce logits.ASPP（或 DeepLabv3+ 中的解码器）的输出通过 1x1 卷积层生成 logit。**
* **The logits represent class probabilities for each pixel in the image.logit 表示图像中每个像素的类概率。**

**架构逻辑结束**

**我的Deeplabv3模型重点逻辑：**

**数据处理：与YOLO不同 生成的是与原图名字一一对应的.png格式的mask图片，**

**加载预训练模型:** **DeepLabV3\_ResNet101\_Weights.COCO\_WITH\_VOC\_LABELS\_V1**

**自定义输出类别为4类 （包括背景）** 

**重点就是下面这些：**

**Criterion：改进的类别权重处理（Class Weights）**

**Optimizer：使用 AdamW 优化器**

**Scheduler：修改了学习率调度器**

* **CosineAnnealingWarmRestarts:**
  + **T\_0: 初始周期长度（10 个 epoch）。**
  + **T\_mult: 每次周期倍增因子（2），使学习率周期逐渐加长。**
  + **eta\_min: 最小学习率（1e-6），保证学习率不会完全消失。**

**优势:**

* **模拟退火策略，学习率呈余弦波下降：**
  + **在每个周期内学习率逐渐降低，接近最优解时探索范围更小。**
  + **周期结束时重启学习率，有助于跳出局部最优。**
* **避免了固定学习率或单调递减学习率可能导致的训练效率低或陷入局部最优问题。**

文本

描述已自动生成

**剩下的就是evaluate和predict 与YOLO大差不差。**