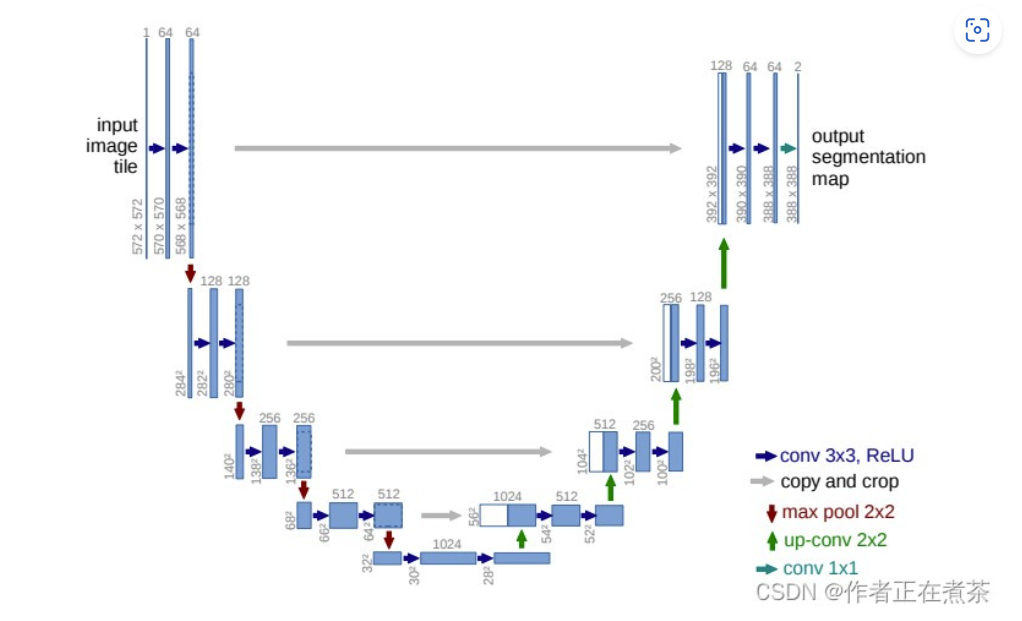
**Unet v1:** [[1505.04597] U-Net: Convolutional Networks for Biomedical Image Segmentation](https://arxiv.org/abs/1505.04597) 论文链接

**网络结构：**多层卷积层叠加,跳跃连接

代码里模型就是和下面这个图这么搭的，



**Yolov8+sam: 只训练了yolov8, sam无需训练**

**代码逻辑：**

**训练阶段：** 训练yolo的目标检测模型，让它可以给图片预测出比较准确的检测框，作为sam的prompt

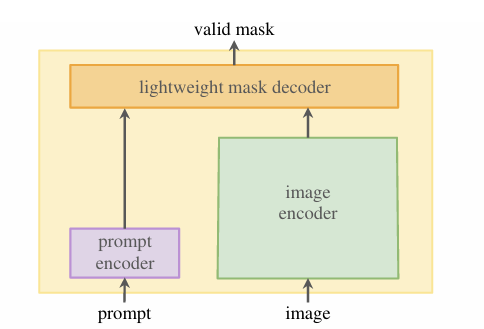
**预测阶段**： 图像输入 ➔ 训练好的目标检测模型Yolo ➔ 生成边界框（头、鳍、甲壳） 边界框 ➔ SAM 模型 ➔ 分割掩码（精确的头、鳍和甲壳分割） 输出分割结果

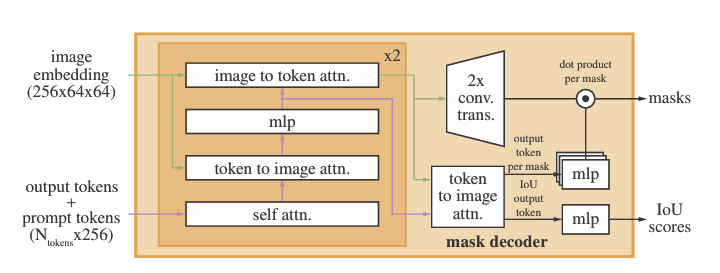
为什么选Sam?

[[2304.02643] Segment Anything](https://arxiv.org/abs/2304.02643) 论文链接  
  
1. 零样本分割（无需训练，给一个prompt(提示)，一个点、框、或者文本就能根据这些提示精准地生成分割结果，yolo预测的边界框正好能作为提示）

因为SAM 是一个经过大规模多样化数据集预训练的模型，具备很强的通用性，是分割领域的Foundation model，可以在不进行任何微调的情况下完成许多分割任务（即零样本分割）。这一特性使得 SAM 可以在各种场景和对象上直接应用，无需专门的数据和训练步骤

1. **Transformer 架构**





Sam是**Transformer 架构，**通过注意力机制，每一个特征点动态地关注图像的任意位置，使得每一层的感受野可以覆盖整个图像。

**双向交叉注意力**（bidirectional cross-attention）机制，这一机制帮助 SAM 在生成掩码时，将输入的图像信息与特定的提示信息（例如点击点、边界框等）进行有效结合。

· **图像特征与提示信息的交互**：将图像的视觉特征与用户提供的提示信息（例如点击、边界框）进行信息融合，使模型可以根据提示信息的上下文调整对图像的理解。

· **掩码生成的双向交互**：既从图像特征中提取信息，又在生成掩码时将图像和提示信息进行反复的相互作用，使得掩码生成过程更加精准。

在 SAM 的 Mask Decoder 中，双向交叉注意力分为两个主要方向：

1. 从提示信息到图像特征的注意力：

提示信息（Prompt Embedding）作为查询向量（query），用于在图像特征中进行注意力操作。

这使得提示信息可以从图像特征中“看到”需要关注的区域，并获取图像中的细节。

1. 从图像特征到提示信息的注意力：

图像特征作为查询向量，而提示信息则作为键（key）和值（value）。

这允许图像特征结合提示信息的上下文，使图像特征中的每个部分能够接收到提示信息的引导，从而在掩码生成时对提示信息做出响应。