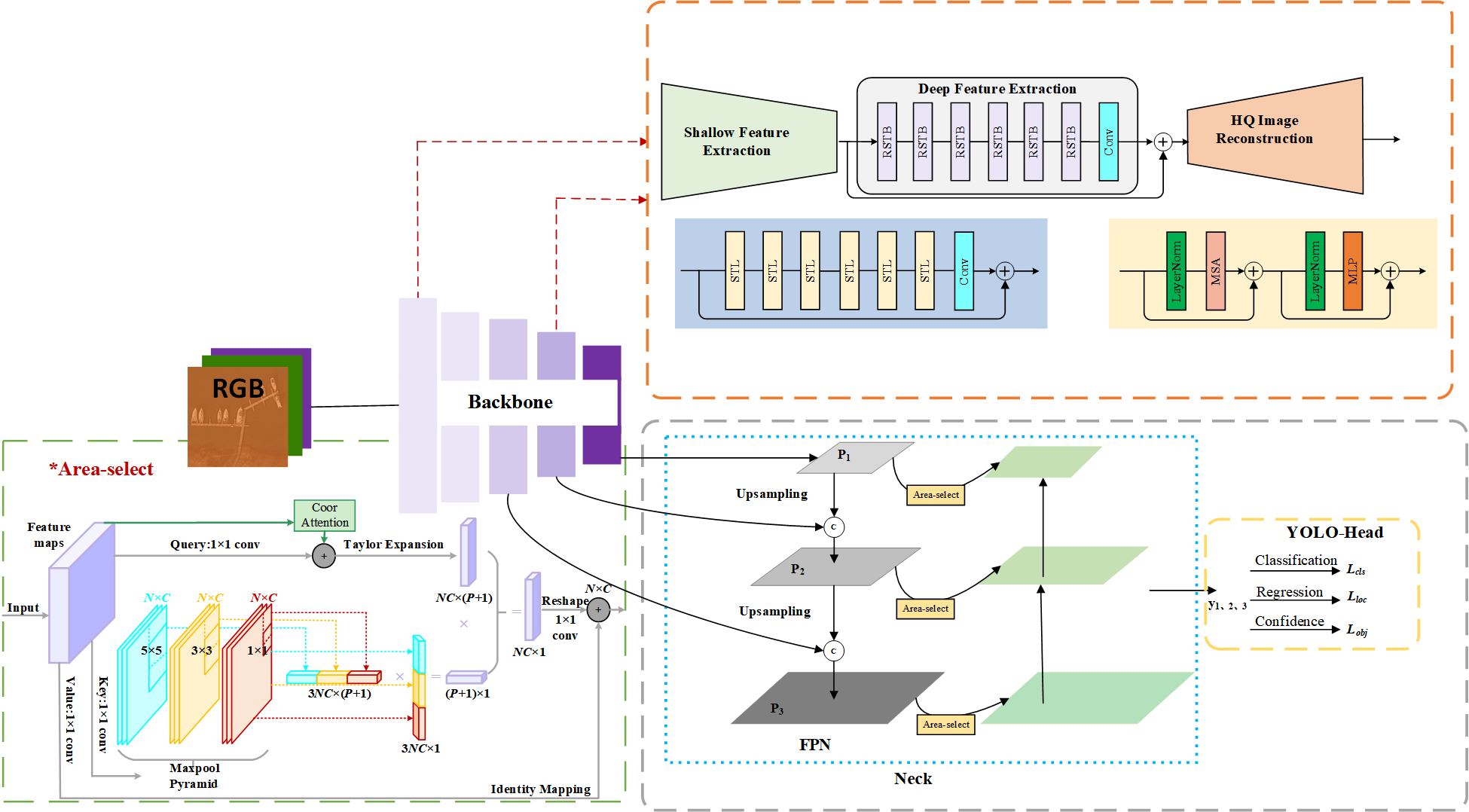
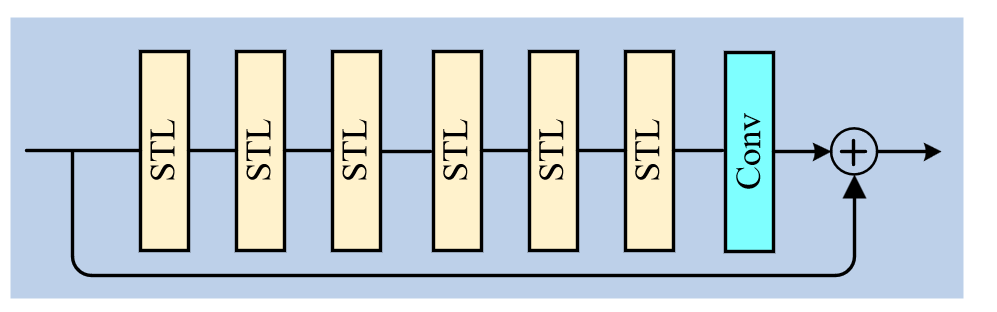
整体框架



创新点1(ISRB,Image super resolution branching)

使用了一个图像超分结构，增强了小目标的检测。在图像超分部分的，络结构主要分为 3 个部分，分别是浅层[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/Wenyuanbo/article/details/_blank)模块，深层特征提取模块和高质量图像重建模块。其中前后两个模块都是基于 CNN 的，中间模块则主要使用 SwInTransformer。

深层特征提取模块由若干个残差 SwInTransformer 块 (RSTB) 和卷积块构成，具体结构如下图。



(1) 首先将来自浅层特征提取模块的特征图分割成多个不重叠的 patch embeddings；

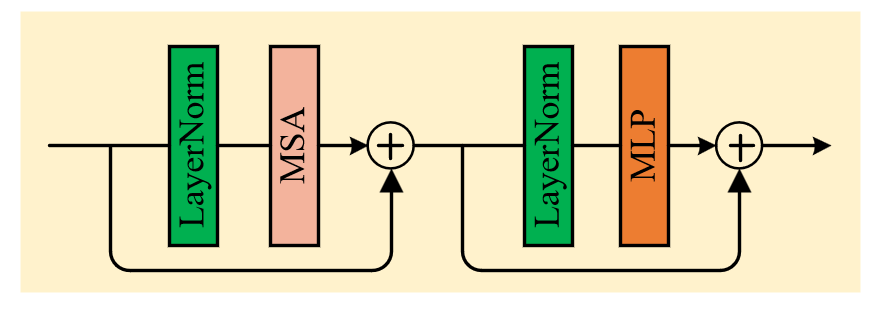
(2) 再通过多个串联的残差 SWin Transformer 块 (RSTB);

(3) 将多个不重叠的 patch embeddings 重新组合成与输入特征图分辨率一样；

(4) 最后通过一个卷积层 (1 层或3 层卷积) 输出；

(5) 在每个 RSTB 中都引入残差连接。

残差 SwInTransformer 块 (RSTB) 中的 STL 就是 SwIn Transformer Layer 的意思，具体结构如下图。



(1) 首先通过一个归一化层 LayerNorm；

(2) 再通过多头自注意力 (Multi-head Self Attention) 模块；

(3) 在多头自注意力结尾引入残差；

(4) 再通过一个归一化层 LayerNorm；

(5) 最后通过一个多层感知机 MLP；

(6) 结尾同样引入残差。

（概括来说，我们是这么操作的，首先数据集输入的图片，我们先缩放到512\*512，然后，我们在计算loss那里输入我们没有缩放的图像，将这个损失函数加到原本的损失函数里面，所以我们的损失函数会比原来的yolo多一项。）

第一项是我们的总损失

第二项是我们的传统yolo损失

第三项是我们超分结构的损失

结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***ISRB*** | **TAB** | **TAB** |  |
| *A* |  |  |  | *71.2* |
| *E* | **√** |  |  | *76.5* |