convnext噪声太多，严重过拟合，在训练集上通过噪声来分类（配合特征图可视化和训练日志）

尝试使用bilateral filter先对图像去噪再送入convnext。（图片）

选择原因：去噪+保边。但效果还是不好。猜想原因可能是滤波的同时损害了部分特征。无法做到既完整地保留特征又有效地去噪。

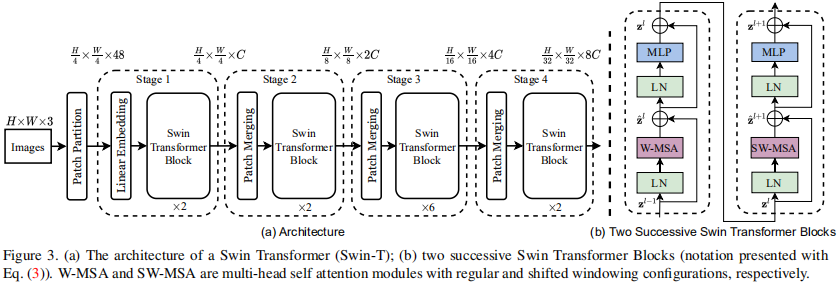
（选讲）

Transformer的自注意力机制能够有效去噪，VIT是对图像全局使用自注意力机制的网络，但是使用全局自注意机制带来的问题是参数量过大，但是该任务下样本量太小，VIT的训练效果不好，因此选用局部注意力

因此使用了swin网络，swin是使用局部注意力机制的网络，它具有强大的去噪和提取特征能力的同时参数量也很小，适合我们的任务。和SOTA的Convformer和metaformer效果差不多

算法描述：去学一下swin

W-MSA：



传统划分图像（ViT）4\*4，why？因为图像数据量太大了，分组减少参数量，但是对于我们的任务，参数量还是太大了。

但swin运用了CNN的思想，划分成7\*7的块，在进行池化操作（三次），增加感受野，

每一个block是Patch merging和swin transformer block的堆叠

Patch merging可以理解为下采样（结合PPT里面的图，P162前两步操作把空间维度放到通道维度）

MLP：1\*1卷积

P168，左边的W-MSA（不重叠窗口），对每一个窗口进行transformer，多重注意力，

右边的（重叠窗口P173），可以实现窗口之间信息交互

数据预处理