**计算机视觉工程实验三**

姓名：李智 学号：123106222804

## 实验目标

实现一种图像超分辨率方法在Set5数据集上的测试，得到超分辨率图像，测量其与原始真实图像之间的PSNR、SSIM指标值。对所选择方法的细节进行介绍，并试着讨论该方法可能存在的优缺点，以及可能的改进方向。

这里使用苏黎世联邦理工学院团队于2021年提出一个适用于图像修复的模型SwinIR，模型延续了Swin Transformer的设计思想，成功将Swin Transformer应用到了图像超分辨率的工作场景之中，用更少的参数获得更好的性能。

## 模型介绍

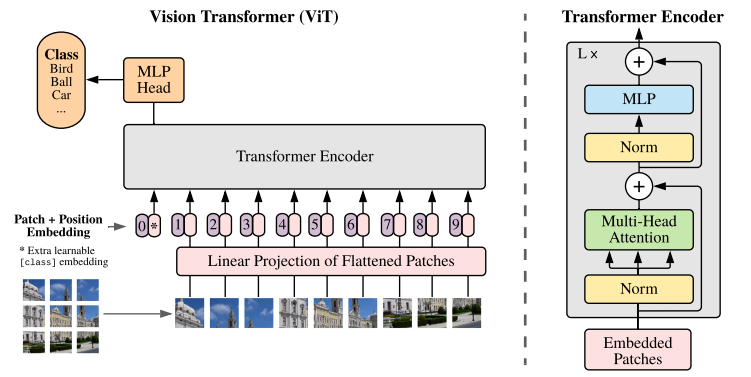
1. 算法背景

在图像超分辨率、图像去噪、压缩等图像修复任务中，卷积神经网络目前仍然是主流。很少有人尝试使用Transformers，但卷积神经网络普遍存在以下两个缺陷：

（1）图像和卷积核之间的交互是与内容无关的。

（2）在局部处理的原则下，卷积对于长距离依赖建模是无效的。

作为卷积的一个替代操作，Transformer设计了全连接的自注意力机制来捕捉全局信息，允许每一个位置直接与整个目标的其它位置进行交互，从而能有效的捕捉长距离的依赖关系。Transformer已经应用在多个CV的细分领域中。其中，最具代表性的是ViT（Vision Transformer），其模型如下：

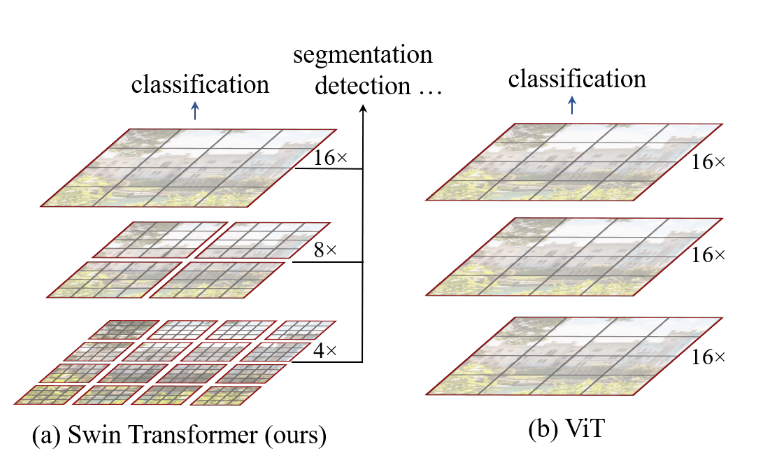


ViT将输入图片分为多个patch，再将每个patch投影为固定长度的向量送入Transformer，后续encoder的操作和原始Transformer中完全相同。由于Vision Transformer因为需要划分patch，因此具有以下两个缺点：

（1）边界像素不能利用patch之外的邻近像素进行图像恢复；

（2）恢复后的图像可能会在每个patch周围引入边界伪影，这个问题能够通过patch overlapping缓解，但会增加计算量。

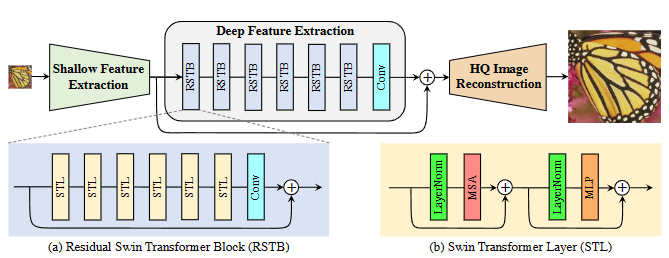
Swin Transformer结合了卷积和Transformer的优势，提出了一种包含滑窗操作，具有层级设计的Transformer结构，如图所示：



相较于ViT，Swin Transformer将，图片切分成若干个窗口，各个窗口内含有固定数量的patch。通过滑窗操作，模型能够学习到跨窗口的信息。同时通过下采样层，各个窗口内能够独立计算attention，优化了计算复杂度的问题，使得模型能够处理超分辨率的图片。SwinIR既为基于Swin Transformer提出的一种新的图像修复模型。

1. 原理分析：

SwinIR 的网络结构主要分为 3 个部分：浅层特征提取模块（shallow feature extraction）、深层特征提取模块（deep feature extraction）、高质量图像重建模块（highquality (HQ) image reconstruction modules）。其中前后两个模块都是基于 CNN 的，中间模块则主要使用 SwInTransformer。对所有的复原任务采用相同的特征提取模块，针对不同的任务采用不同的重建模块。



1. 浅层特征提取

浅层特征提取只使用一层3\*3卷积进行提取特征F0，如下：

1. 深层特征提取

深层特征提取模块由若干个残差 SwInTransformer 块 (RSTB) 和卷积块构成， 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，通过多个串联的残差 SWin Transformer 块 (RSTB)，再将多个不重叠的 patch embeddings 重新组合成与输入特征图分辨率一样，最后通过一个卷积层输出，并在每个 RSTB 中都引入残差连接，如下：

将提取到的浅层特征F0，使用深层特征提取模块HDF进一步提取特征。深层特征提取模块由K个residual Swin Transformer blocks(RSTB)和一个3×3卷积构成。

每个RSTB的输出F1,F2,FK，以及输出的深层特征FDK如式：

卷积层能够将卷积的归纳偏置引入基于Transformer的网络，为后续浅层、深层特征的融合奠定基础。

1. 图像重建模块

图像重建模块共提出了4种结构。

1) 经典超分 (卷积 + pixelshuffle 上采样 + 卷积)；

2) 轻量超分 (卷积 + pixelshuffle 上采样)；

3) 真实图像超分 (卷积 + 卷积插值上采样 + 卷积插值上采样 + 卷积)；

4) 像去噪和 JPEG 压缩去伪影 （卷积 + 引入残差）。

图像重建模块通过融合浅层特征F0和深层特征FDK来重建高质量图片IRHQ，式中HREC为重建模块。

浅层特征F0主要包含低频信息，而深层特征则专注于恢复丢失的高频信息。SwinIR采用一个长距离连接，将低频信息直接传输给重建模块，可以帮助深度特征提取模块专注于高频信息，稳定训练。在图像超分辨率任务中，通过sub-pixel convolution layer将特征上采样，实现重建。

1. 改进思路

针对SwinIR的改进方案有两个思路，其一是改进深层特征提取模块中的残差 SwInTransformer 块，这部分需要结合Transform以及Swin Transform的原理机制进行思考，较为复杂，另一部分是考虑改进图像的重建模块，这部分本质上其实就是卷积+上采样的组合，可以通过添加或替换不同的功能模块、或是变换组合顺序，实现不同种类的图像重建模块，这部分相对容易改进。结合上述思想，我提出了以下几个改进方案：

1. 添加图像全局特征信息的提取能力：

Swin Transform 采用基于窗口的局部注意力策略，即用划分局部窗口+滑动窗口的方式提取图像的特征，虽然这样做能够很大程度上平衡性能和计算开销，但却限制了使用大感受野来捕获全局信息并在早期层中建立长依赖关系。我们可以尝试将具有提取全局特征的卷积模块（如快速傅里叶卷积）替换掉RSTB模块、或是整个的深层特征提取模块中的初始卷积层（Conv），再者，可以通过引入一些残差模块，将全局特征核局部特征融合，以提高模型对图像整体信息的感知能力。

1. 增大窗口的尺寸，扩大窗口的感受野：

限制窗口大小虽然能节省计算成本，但只通过移位窗口逐步增加感受野，可能会牺牲自注意力机制的表征能力。因此，考虑增大窗口，并为此改进模型的自注意力机制，也许能一定程度上提高transformer使用全局信息的能力。

1. 在重建模块中引入残差连接：

在每个重建模块中添加残差连接，通过将原始输入与重建结果相加或连接，使模型能够学习到图像的细微差异和细节，帮助提高重建质量并减少信息损失。

## 结果分析

1. 测试环境：windows11 + python3.9 + pytroch

2. 运行分析：

（1）准备Set5数据：

Set5是一个常用的图像超分辨率重建领域的基准测试数据集，包含5张不同场景类型的图像，整体规模较小。常使用该数据集进行算法测试，并根据重建结果与原始高分辨率图像之间的差异来简单评估算法的准确性和视觉质量。将高分辨率的初始数据集放置在./testsets/Set5/HR目录下。

（2）使用Bicubic对Set5数据集进行下采样：

双三次插值是二维空间中最常用的插值方法，它能创造出比双线性插值更平滑的图像边缘。在这种方法中，函数f在点 (x,y) 的值可以通过矩形网格中最近的十六个采样点的加权平均得到，在这里需要使用两个多项式插值三次函数，每个方向使用一个。这里使用Bicubic下采样的方法将初始高分辨率图像缩放4倍，这种低分辨率到高分辨率的4倍缩放关系是超分辨率算法中常见的设置，将下采样的图片放置在./testsets/Set5/LR\_bicbic/X4目录下。

（3）使用预训练模型实现低分辨率图片的超分：

作者提供了在经典/轻量级/真实世界图像等场景的预训练模型，这里选用经典图像预训练模型（001\_classicalSR\_DIV2K\_s48w8\_SwinIR-M\_x4.pth），该模型使用DIV2K数据集进行训练和测试。DIV2K数据集是一个多样化的高分辨率图像超分辨率数据集，它包含了800个高分辨率图像，内容涵盖了自然风景、城市街景、人物肖像等多个领域，将该模型放置在./model\_zoo/swinir/目录下。

下载作者提供的测试代码和预训练模型，并将模型和数据集放置在相应的路径后，执行测试文件main\_test\_swinir.py。同时在命令行中设置如下参数：

task：任务类型，设置为classical\_sr

scale：放大的倍数，这个和预训练权重对应。设置对应权重x4

training\_patch\_size：用于训练 SwinIR 的patch大小，设置为48

model\_path：对应预训练模型的路径

folder\_lq：对应下采样后的低质量图片的文件夹路径

folder\_gt：对应原始高质量图片的文件夹路径

参数设置要求与预训练的模型任务完全一致，模型会读入folder\_lq路径的图像，并将超分后的图像与folder\_gt中的原始图像集做对比，计算评价指标，超分实现的图片放置在./results/swinir\_classical\_sr\_x4目录中。

（4）测量原始图像与真实图像的PSNR、SSIM指标值：

PSNR基于图像的均方误差（MSE）来度量原始图像与重建图像之间的差异；SSIM通过比较原始图像与重建图像的结构、亮度和对比度之间的相似性，给出一个0到1之间的相似性分数，图像测试成功后，得到的指标如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | baby | bird | butterfly | head | woman | avg(Set5) |
| PSNR(dB) | 32.47 | 33.41 | 28.21 | 30.04 | 29.84 | 30.80 |
| SSIM | 0.8728 | 0.9329 | 0.9211 | 0.7251 | 0.9120 | 0.8728 |

从PSNR的结果来看，Bird图像具有最高的峰值信噪比，而Butterfly图像具有最低的峰值信噪比。平均PSNR为30.80 dB，整体上图像的重建质量相对较高。从SSIM的结果来看，Bird图像具有最高的结构相似性指数，而Head图像具有最低的结构相似性指数。平均SSIM为0.8728，整体上图像的结构和内容相对保持较好的相似性。