问题：

1. 自然图像和医学图像之间存在领域差距，例如计算机断层扫描（CT）和磁共振成像（MRI）等医学成像模态之间存在较大的区别
2. 应用数据类型是3D时，2D的Transformer模型不能利用平面的上下文信息，会有限制
3. 3D带注释的数据集是非常耗时的

创新：

1. 引入自监督预训练学习框架
2. 3D Swin UNETR架构

自监督的预训练：

1. 使用Swin UNETR编码器进行预训练
2. 使用多目标损失函数
3. 使用三个不同的代理任务，旨在更好的帮助模型理解医学图像

为了执行任务将额外的三个投影头（projection heads）附加到了编码器上

1. 预训练之后的正式训练--使用完整的Swin UNETR模型
2. 预训练数据增强：对三维图像数据进行随机剪裁以获得子体积，然后对每个子体积进行两次随机旋转和cutout（遮挡部分区域）数据增强。

网络模型：

先进行自监督的预训练-模型在没有手动标签的情况下，从CT图像中学到了有用的特征表示

自监督的预训练：

通过没有标签的数据-5,050张CT图像 来对模型进行预训练，通过代理任务来预训练模型

预训练任务：图像修补（image inpainting）、3D旋转预测和对比学习

正式训练：

数据集：MSD、BTCV

题目：

时间：

网络模型提出的原因：

1. 应用数据类型是3D时，2D的Transformer模型不能利用平面的上下文信息

（2）3D带注释的数据集非常耗时，难获取

网络模型基础：

创新点：

（1）3D Swin UNETR分割网络架构

1. 自监督的预训练架构

1.使用Swin UNETR编码器部分作为自监督预训练网络架构，在预训练期间添加额外的重建头，输出三种的自监督预训练任务的结果

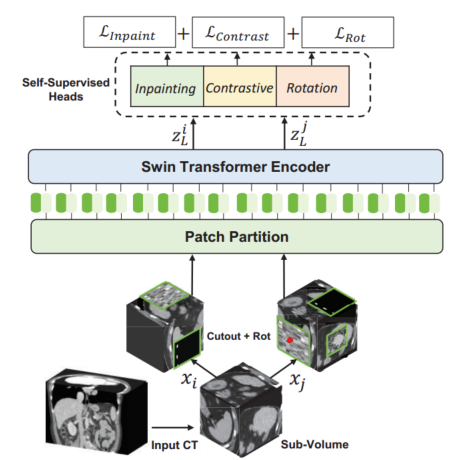
1. 将输入的CT图像随机裁剪成子体积，对每个子体积进行两次随机旋转和cutout（遮挡部分区域）数据增强
2. 使用三个任务进行自监督学习：图像修补（image inpainting）、3D旋转预测和对比编码

用于自监督学习的对比编码方法，它通过最大化相同类别样本之间的信息关联度，同时最小化不同类别样本之间的信息关联度，以改进视觉表示学习

这部分解释了在给定一批增强的子体积（一种数据块）的情况下，对比编码通过最大化正对（来自相同子体积的增强样本）之间的互信息，同时最小化负对（来自不同子体积的视图）之间的互信息，从而实现更好的表示学习。

对比编码额外的重建头：线性层

1. 对于三个任务各自设置不同的损失函数，将三个损失函数集结合在一起形成混合损失函数



实验：

1. 预训练：

预训练数据集：有5个公共CT数据集，包括5050个3D样本

胸部、腹部和头部/颈部的3D体积的相应数量分别为2018、1520和1223

1. 正式训练

两个公开数据集：

BTCV：30个腹部3D CT样本，进行13个器官的分类

MSD：包括来自不同器官和图像模态的10个分割任务

效果：

掩模分割："掩模分类"是一种图像分割任务，它涉及将图像中的每个像素分配给特定的类别或标签。每个像素都被分配一个类别标签，用以表示它属于哪个对象或区域。这些类别标签通常以掩模（mask）的形式表示，其中每个掩模代表一个不同的类别。

掩模分类通常与语义分割和实例分割密切相关：

语义分割：在语义分割中，图像中的每个像素被分配到一个语义类别，例如道路、汽车、树木等。这意味着所有具有相同语义的像素都属于相同的类别，但不区分不同的实例。例如，所有的汽车像素都属于同一类别，但不区分不同的汽车。

实例分割：在实例分割中，图像中的每个像素被分配到一个唯一的实例标识符，这意味着不仅要区分不同的对象或物体，还要区分同一类别中的不同实例。例如，如果图像中有多辆汽车，每辆汽车都会有一个唯一的标识符，以区分它们。

掩模分类结合了这两种任务的特点，它不仅为图像中的每个像素分配一个类别标签，还为每个类别创建一个掩模，以表示该类别的像素分布。这使得在分割任务中可以更好地了解不同类别的位置和形状。

总之，掩模分类是一种图像分割任务，它旨在将每个像素分配给不同的类别，并为每个类别生成掩模以表示像素分布。这有助于更精细地理解图像中的对象和类别分布。