# TransUnet：Transformer为医学图像分割提供强大的编码器

Jieneng Chen[1)]，Yongyi Lu[1)]，Qihang Yu[1]，Xiangde Luo[2]，Ehsan Adeli[3]，Yan Wang[4]，Le Lu[5]，Alan L.Yuille[1]和Yuyin Zhou[3]

[1]约翰霍普金斯大学
[2]电子科技大学
[3]斯坦福大学
[4]华东师范大学
[5]PAII公司

抽象的。医学图像分割是开发医疗系统，特别是疾病诊断和治疗计划的必要前提。在各种医学图像分割任务中，Ushaped架构（也称为U-Net）已经成为事实上的标准，并取得了巨大的成功。然而，由于卷积操作的固有局部性，U-Net通常在显式建模长程相关性方面表现出局限性。为序列到序列预测而设计的变压器已经作为具有固有全局自注意机制的替代架构出现，但由于低层次细节不足，可能导致有限的定位能力。在本文中，我们提出了TransUnet，它具有Transformers和U-Net的优点，是医学图像分割的一种强有力的选择。一方面，转换器对来自卷积神经网络（CNN）特征图的标记化图像块进行编码，作为用于提取全局上下文的输入序列。另一方面，解码器对编码特征进行上采样，然后将其与高分辨率CNN特征图相结合，以实现精确定位。
我们认为，变换器可以作为医学图像分割任务的强编码器，与U-Net相结合，通过恢复局部空间信息来增强更精细的细节。Transunet在不同的医学应用中，包括多器官分割和心脏分割，取得了优于各种竞争方法的性能。代码和模型可在https://github.com/Beckschen/ TransUNet.

## 1 介绍

卷积神经网络（CNNs），特别是全卷积网络（FCNs）[8]在医学图像分割中占主导地位。在不同的变体中，U-Net[12]，它由一个对称的编码器-解码器网络组成，通过跳跃连接来增强细节保留，已成为事实上的选择。基于这种方法，已经在广泛的医学应用中取得了巨大的成功，例如心脏分割

磁共振（Mr）[16计算机断层扫描（CT）器官分割[7,17,19]和息肉分割[20]来自结肠镜检查视频。

尽管基于CNN的方法具有特殊的表示能力，但由于卷积运算的固有局部性，它们在建模显式长程关系时通常表现出局限性。因此，这些架构通常产生较弱的性能，特别是对于在纹理、形状和尺寸方面表现出较大的患者间差异的目标结构。为了克服这一限制，现有研究提出建立基于CNN特征的自我注意机制。13、15]. 另一方面，为序列到序列预测而设计的变换器已经作为替代架构出现，其完全采用分配卷积算子，并且完全依赖于注意机制[14]. 与先前的基于CNN的方法不同，Transformer不仅在建模全局上下文方面很强大，而且在大规模预训练下对下游任务表现出优越的可转移性。在机器翻译和自然语言处理（NLP）领域取得了广泛的成功。3,14]. 最近，对于各种图像识别任务，尝试也已经达到甚至超过了最先进的性能[4,18].

在本文中，我们提出了第一个研究，探索变压器在医学图像分割方面的潜力。然而，有趣的是，我们发现简单的使用（即，使用转换器对标记化的图像块进行编码，然后直接将隐藏的特征表示上采样为全分辨率的密集输出）不能产生令人满意的结果。

这是由于变换器将输入视为1D序列，并且只关注在所有阶段对全局上下文进行建模，因此导致缺乏详细定位信息的低分辨率特征。并且该信息不能通过直接上采样到全分辨率来有效地恢复，因此导致粗糙的分割结果。另一方面，CNN架构（例如，U-Net[12]）提供了一种用于提取低级视觉线索的途径，其可以很好地弥补这种精细的空间细节。

为此，我们提出了第一个医学图像分割框架TransUnet，它从序列到序列预测的角度建立了自注意机制。为了补偿Transformer带来的特征分辨率损失，Transunet采用了混合CNN-Transformer架构，以利用来自CNN特征的详细高分辨率空间信息和Transformer编码的全局上下文。受U形架构设计的启发，由变压器编码的自我注意特征随后被上采样，以与从编码路径中跳过的不同高分辨率CNN特征相结合，从而实现精确定位。我们表明，这样的设计允许我们的框架保留变压器的优势，也有利于医学图像分割。实验结果表明，与之前的基于CNN的自我关注方法相比，我们的基于变压器的架构提供了一种更好的方式来利用自我关注。 此外，我们观察到，更密集地合并低级特征通常会导致更好的分割精度。大量的实验证明，

我们的方法在各种医学图像分割任务上优于其他竞争方法。

## 2 相关作品

将细胞神经网络与自我注意机制相结合。各种研究试图通过基于特征图对所有像素的全局交互进行建模，将自注意机制整合到细胞神经网络中。例如，Wang等人设计了一个非局部算子，它可以插入到多个中间卷积层中。15]. 基于编码器-解码器U型架构，Schlemper等人[13]提出了集成到跳跃连接中的附加注意门模块。与这些方法不同的是，我们在我们的方法中使用了变压器来嵌入全局自注意。

变形金刚。变压器最早是由[14]用于机器翻译，并在许多自然语言处理任务中建立了最新技术。为了使变压器也适用于计算机视觉任务，已经进行了一些修改。例如，Parmar等人[11]仅在每个查询像素的局部邻域中而不是全局地应用自关注。Child等人[1]提出了稀疏变换器，其采用全局自关注的可扩展近似。最近，视觉变压器（Vit）[4]通过直接将具有全局自关注的变压器应用于全尺寸图像，实现了ImageNet分类的最新技术。据我们所知，所提出的TransUnet是第一个基于Transformer的医学图像分割框架，它建立在非常成功的Vit的基础上。

## 3 方法

给定图像X $R^{H\times W\times C}$，其空间分辨率为H W，通道数为C。我们的目标是预测相应的大小为H w的逐像素标签图。最常见的方式是直接训练CNN（例如，UNET）以首先将图像编码成高级特征表示，然后将其解码回全空间分辨率。与现有方法不同，我们的方法通过使用变压器将自注意机制引入到编码器设计中。我们将首先在章节中介绍如何直接应用转换器对分解图像块的特征表示进行编码3.1. 然后，将在第节中阐述TransUnet的总体框架3.2.

### 3.1 变压器作为编码器

**图像序列化以下**[4]，我们首先通过重新执行标记化将输入X整形为一系列平坦的2D面片 $\{X^i \in R^{P\cdot C} \mid I= 1, .., N\}$，其中每个块的大小为pxp，并且N=硬件是图像的数量补丁（即，输入序列长度）。

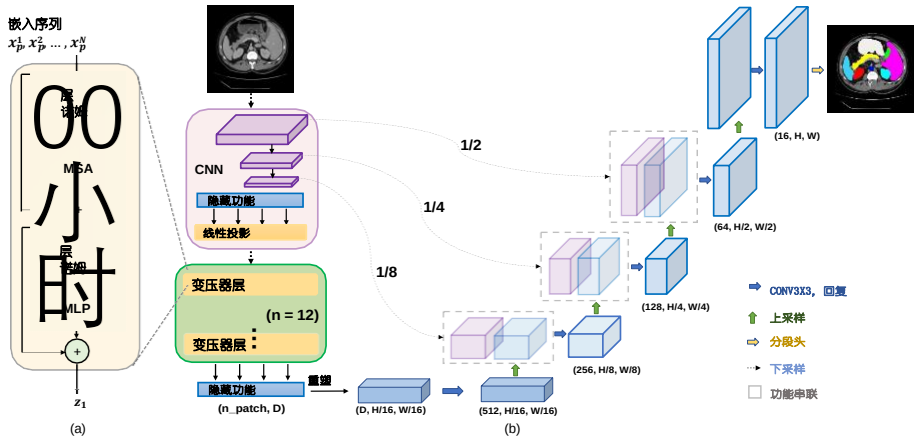


图1：框架概述。（a）变压器层示意图；（B）拟议的Transunet的结构。

面片嵌入我们使用可训练线性投影将矢量化面片$X_p$映射到潜在的d维嵌入空间。为了对面片空间信息进行编码，我们学习特定的位置嵌入，这些位置嵌入被添加到面片嵌入以保留位置信息，如下所示：

$$Z_0=[X^1E; X^2E; \bullet \bullet \bullet ;X^Ne]+e_{位置}, \quad (1)$$

其中$E \in R^{(P^2 \cdot C)\times D}$是面片嵌入投影，并且$E_{位置} \in R^{Vx D}$表示位置嵌入。

变换器编码器由L层多头自注意（MSA）和多层感知器（MLP）块组成（等式（2)(3)）.因此，第层的输出可以写为：

$$Z_{\ell}=MSA(ln(Z_{\ell-1}))+Z_{\ell-1}, \quad (2)$$

$$Z_{\ell}=MLP(ln(Z^{'}_{\ell}))+Z^{'}_{\ell}, \quad (3)$$

其中，ln,()表示层归一化算子，$Z_L$是编码图像表示。变压器层的结构如图所示1(a).

### 3.2 TransUNet

出于分割的目的，一种直观的解决方案是简单地上采样EN-编码特征表示$Z_LP^2$到用于预测密集输出的全分辨率。这里，为了恢复空间顺序，编码的特征的大小首先应从硬件重塑Ture $P^2$至$\frac{H}{P}\times\frac{W}{P}$。我们使用1×1卷积

为了将重构特征的通道大小减少到类别数，然后将特征图直接双线性上采样到全分辨率HW，以预测最终的分割结果。在后面第节的比较中4.3因此，在解码器设计中，我们将这种朴素的上采样基线表示为“无”。

如上所述，尽管将变压器与朴素上采样相结合已经产生了合理的性能，但这种策略并不是最优的。

分段中变压器的年龄，因为通常远小于
因此，原始图像分辨率HW不可避免地导致
低级细节（例如，器官的形状和边界）。因此，为了补偿这种信息损失，TransUnet采用混合CNN-Transformer架构作为编码器以及级联上采样器，以实现精确定位。拟议的TRANSUNET的概况如图所示1.

**CNN-Transformer混合作为编码器。而不是使用纯变压器作为编码器（第3.1），**TransUnet采用CNN-Transformer混合模型，其中首先使用CNN作为特征提取器来生成输入的特征图。块嵌入被应用于从CNN特征图而不是从原始图像中提取的11块。

我们选择这种设计是因为1）它允许我们在解码路径中利用中间高分辨率CNN特征图；以及2）我们发现混合CNN-变换器编码器比简单地使用纯变换器作为编码器执行得更好。

级联上采样器我们介绍了一种级联上采样器（CUP），它由多个上采样步骤组成，用于解码隐藏的特征以输出最终的分割掩码。在将隐藏特征序列整形为的形状后，我们通过级联来实例化CUP
用于达到从到H×W的全分辨率的多个上采样块，
其中每个块由2上采样算子、3×3卷积层
和ReLU层。

我们可以看到，CUP与混合编码器一起形成了一个U形架构，该架构通过跳跃连接实现了不同分辨率级别的特征聚合。CUP的详细架构以及中间跳跃连接可以在图中找到1(b).

## 4 实验和讨论

### 4.1 数据集和评估

Synapse多器官分割数据集[1].我们在MICCAI 2015多图谱腹部标记挑战中使用了30个腹部CT扫描，总共有3779个轴向对比增强腹部临床CT图像。

每个CT体积由85198个512512像素的切片组成，体素空间分辨率为([0.54 0.54][0.98 0.98][2.5 5.0])。以下[5]，我们报道了平均DSC和平均Hausdorff距离（HD）

[1] https://www.synapse.org/#!Synapse:syn3193805/wiki/217789

对8个腹部脏器（主动脉、胆囊、脾脏、左肾、右肾、肝脏、胰腺、脾脏、胃）随机分割18个训练病例（2212个轴位切片）和12个病例进行验证。

**自动心脏诊断挑战**[2]. ACDC Challenge收集从MRI扫描仪获取的不同患者的检查。屏气下采集Cine Mr图像，一系列短轴层面覆盖心脏，从左心室底部到心尖部，层厚5~8mm。短轴面内空间分辨率从0.83到1.75mm$^2$/像素。

每个患者扫描都用左心室（LV）、右心室（RV）和心肌（MYO）的真实情况进行人工注释。我们报告了平均DSC，随机分为70个训练病例（1930个轴向切片），10个病例用于验证，20个病例用于测试。

表1：Synapse多器官CT数据集的比较（平均DICE评分%和平均Hausdorff距离（mm），以及每个器官的DICE分数%）。

| 框架 | | 平均值 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 编码器 | 解码器 | DSC ↑ | HD ↓ | 主动脉 | 胆囊 | 肾脏（L） | 肾脏（R） | 肝 | 胰腺 | 脾 | 胃 |
| V-Net[9] | | 68.81 | - | 75.34 | 51.87 | 77.10 | 80.75 | 87.84 | 40.05 | 80.56 | 56.98 |
| 达尔[5] | | 69.77 | - | 74.74 | 53.77 | 72.31 | 73.24 | 94.08 | 54.18 | 89.90 | 45.96 |
| R50 | U-网[12] | 74.68 | 36.87 | 84.18 | 62.84 | 79.19 | 71.29 | 93.35 | 48.23 | 84.41 | 73.92 |
| R50 | 阿图努内[13] | 75.57 | 36.97 | 55.92 | 63.91 | 79.20 | 72.71 | 93.56 | 49.37 | 87.19 | 74.95 |
| 维生素[4] | 没有人 | 61.50 | 39.61 | 44.38 | 39.59 | 67.46 | 62.94 | 89.21 | 43.14 | 75.45 | 69.78 |
| 维生素[4] | 杯子 | 67.86 | 36.11 | 70.19 | 45.10 | 74.70 | 67.40 | 91.32 | 42.00 | 81.75 | 70.44 |
| R50-维生素[4] | 杯子 | 71.29 | 32.87 | 73.73 | 55.13 | 75.80 | 72.20 | 91.51 | 45.99 | 81.99 | 73.95 |
| **TransUNet** | | **77.48** | **31.69** | 87.23 | 63.13 | 81.87 | 77.02 | 94.08 | 55.86 | 85.08 | 75.62 |

### 4.2 实施细节

对于所有实验，我们应用简单的数据增强，例如随机旋转和翻转。

对于纯粹的基于变压器的编码器，我们简单地采用Vit[4]具有12个变压器层。对于混合编码器设计，我们结合了ResNet-50[6]和Vit，表示为“R50-Vit”，通过本文。所有变压器主干（即Vit）和ResNet-50（表示为“R-50”）都在ImageNet上进行了预训练[2]. 除非另有说明，否则输入分辨率和面片大小p被设置为224、224和16。因此，我们需要在CUP中连续级联四个2上采样块，以达到全分辨率。并且用SGD优化器以0.01的学习率、0.9的动量和1e-4的权重衰减来训练模型。对于ACDC数据集和Synapse数据集，默认批量大小为24，默认训练迭代次数分别为20K和14K。所有实验均使用单个NVIDIA RTX2080Ti GPU进行。

以下[17,19]，以逐层的方式推断所有的3D体积，并且将预测的2D切片堆叠在一起以重建用于评估的3D预测。

2 https://www.creatis.insa-lyon.fr/Challenge/acdc/

### 4.3 与现有技术的比较

我们在Synapse多器官分割数据集上进行了主要实验，将我们的TransUnet与之前的四种技术进行了比较：1）V-Net[9];2）达尔[5];3) U-Net [12]和4）Attnunet[13].

为了证明我们的CUP解码器的有效性，我们使用Vit[4]作为编码器，并比较分别使用朴素上采样（“无”）和CUP作为解码器的结果；为了证明我们的混合编码器设计的有效性，我们使用CUP作为解码器，并分别比较使用Vit和R50-Vit作为编码器的结果。为了与Vithybrid基线（R50-Vit-Cup）和我们的TransUnet进行公平的比较，我们还替换了U-Net的原始编码器[12]和Attnunet[10]使用ImageNet预训练的ResNet-50。根据DSC和平均豪斯多夫距离（单位：mm）的结果列于表中1.

首先，我们可以看到，与Vit-None相比，Vit-Cup在平均DSC和Hausdorff距离方面分别提高了6.36%和3.50 mm。这一改进表明，我们的CUP设计提供了比直接上采样更好的解码策略。类似地，与Vit-Cup相比，R50-Vit-Cup在DSC和Hausdorff距离方面也有3.43%和3.24mm的额外改进，这证明了我们的混合编码器的有效性。在R50-Vit-CUP的基础上，我们的TransUnet也配备了跳接，在基于变压器的模型的不同变体中实现了最佳效果。

其次，桌子1 还示出了所提出的TransUnet相对于现有技术具有显著的改进，例如，考虑到平均DSC，性能增益范围从1.91%到8.67%。特别地，直接将变压器应用于多器官分割产生了合理的结果（Vit-Cup的67.86%DSC），但不能与U-Net或Attnunet的性能相匹配。这是由于变换器可以很好地捕获有利于分类任务的高层语义，但缺乏用于分割医学图像精细形状的低层线索。另一方面，将变压器与CNN组合，即R50-VITCUP，优于V-NET和DARR，但仍产生比纯基于CNN的R50-U-NET和R50-ATTNUNET差的结果。最后，当通过跳跃连接与U-Net结构结合时，所提出的Transunet建立了新的技术发展水平，其性能分别比R50-Vit-Cup和之前最好的R50-Attnunet高出6.19%和1.91%。 显示了TransUnet学习高层语义特征和低层细节的强大能力，这在医学图像分割中是至关重要的。对于平均Hausdorff距离，也可以看到类似的趋势，这进一步证明了我们的TransUnet相对于这些基于CNN的方法的优势。

### 4.4 分析研究

为了全面评估所提出的TransUnet框架并验证其在不同设置下的性能，进行了各种消融研究。

包括：1）跳接次数；2）输入分辨率；3）序列长度和斑块大小；4）模型尺度。

跳过连接数。如上所述，集成U-netlike跳跃连接有助于通过恢复低级空间信息来增强更精细的分割细节。该消融的目的是测试在TransUnet中添加不同数量的跳跃连接的影响。通过将跳跃连接的数量改变为0（R50-Vit-Cup）/1/3，图中总结了所有8个测试器官的平均DSC的分割性能2. 请注意，在“1-skip”设置中，我们仅在1/4分辨率比例下添加跳跃连接。我们可以看到，添加更多的跳跃连接通常会导致更好的分割性能。最佳平均DSC和HD是通过将跳跃连接插入到除输出层以外的CUP的所有三个中间上采样步骤来实现的，即在1/2、1/4和1/8分辨率范围内（如图所示1). 因此，我们的TransUnet采用了这种配置。还值得一提的是，较小器官（即主动脉、胆囊、肾脏、胰腺）的性能增益比较大器官（即肝脏、脾脏、胃）的性能增益更明显。这些结果加强了我们最初的直觉，即将类似U-Net的跳跃连接集成到变压器设计中，以便能够学习精确的低层次细节。

作为一个有趣的研究，我们在跳跃连接中应用加性变压器，类似于[13]，并且发现这种新型的跳跃连接甚至可以进一步提高分割性能。由于GPU内存的限制，我们在1/8分辨率比例跳跃连接中使用了一个光变压器，同时保持其他两个跳跃连接不变。结果，这种简单的改变导致1. 4%DSC的性能提升。

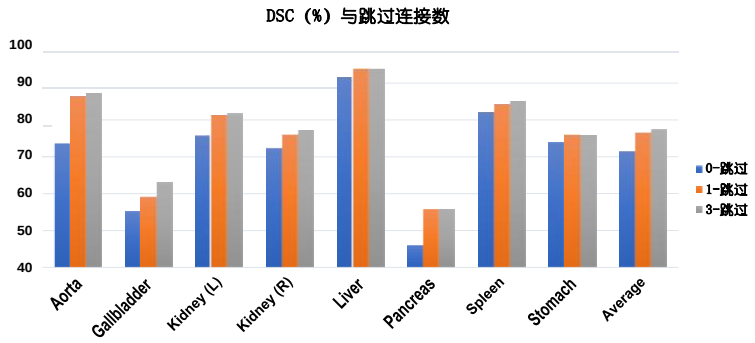


图2：关于TransUnet中跳跃连接数量的消融研究。

输入分辨率的影响。TransUnet的默认输入分辨率为224224。这里，我们还提供了在高分辨率512 512上训练TransUnet的结果，如表中所示2. 当使用512 512作为输入时，我们保持相同的面片大小（即，16），这导致近似的

5× 变压器的较大序列长度。如同[4]表明，增加有效序列长度显示出稳健的改进。对于TransUnet，将分辨率标度从224224改变到512512导致平均DSC提高6.88%，代价是更大的计算成本。因此，考虑到计算成本，本文中的所有实验比较都是在默认分辨率为224224的情况下进行的，以证明TransUnet的有效性。

表2：输入分辨率影响的消融研究。

| 分辨率 | 平均DSC | 主动脉 | 胆囊 | 肾脏（L） | 肾脏（R） | 肝脏 | 胰腺 | 脾脏 | 胃 |
|---|---|---|---|---|---|---|---|---|---|
| 224 | **77.48** | 87.23 | 63.13 | 81.87 | 77.02 | 94.08 | 55.86 | 85.08 | 75.62 |
| 512 | **84.36** | 90.68 | 71.99 | 86.04 | 83.71 | 95.54 | 73.96 | 88.80 | 84.20 |

**关于斑块大小/序列长度的影响。**

我们还研究了补丁大小对TransUnet的影响。结果总结在表中3. 可以观察到，通常用较小的块尺寸获得较高的分割性能。注意，变换器的序列长度与补丁大小的平方成反比（例如，补丁大小16对应于序列长度196，而补丁大小32具有较短的序列长度49），因此，减小补丁大小（或增加有效序列长度）显示出稳健的改进，因为变换器对于较长的输入序列编码每个元素之间更复杂的相关性。遵循Vit中的设置[4]，我们使用1616作为本文中的默认补丁大小。

表3：关于补片大小和序列长度的消融研究。

| 补丁大小 | 序列长度 | 平均DSC | 主动脉 | 胆囊 | 肾脏（L） | 肾脏（R） | 肝脏 | 胰腺 | 脾脏 | 胃 |
|---|---|---|---|---|---|---|---|---|---|---|
| 32 | 49 | 76.99 | 86.66 | 63.06 | 81.61 | 79.18 | 94.21 | 51.66 | 85.38 | 74.17 |
| 16 | 196 | 77.48 | 87.23 | 63.13 | 81.87 | 77.02 | 94.08 | 55.86 | 85.08 | 75.62 |
| 8 | 784 | **77.83** | 86.92 | 58.31 | 81.51 | 76.40 | 93.81 | 58.09 | 87.92 | 79.68 |

模型缩放。最后但并非最不重要的是，我们提供了对不同模型大小的Transunet的消融研究。特别地，我们研究了两种不同的TransUnet配置，“基本”和“大型”模型。对于“基础”模型，隐藏尺寸D、层数、MLP尺寸和头数分别设置为12、768、3072和12，而对于“大型”模型，这些超参数设置为24、1024、4096和16。从表中4 我们得出结论，更大的模型导致更好的性能。考虑到计算成本，我们对所有的实验都采用“基础”模型。

表4：模型规模的消融研究。

| 模型比例 | 平均DSC | 主动脉 | 胆囊 | 肾脏（L） | 肾脏（R） | 肝脏 | 胰腺 | 脾脏 | 胃 |
|---|---|---|---|---|---|---|---|---|---|
| 基于 | 77.48 | 87.23 | 63.13 | 81.87 | 77.02 | 94.08 | 55.86 | 85.08 | 75.62 |
| 大的 | 78.52 | 87.42 | 63.92 | 82.17 | 80.19 | 94.47 | 57.64 | 87.42 | 74.90 |

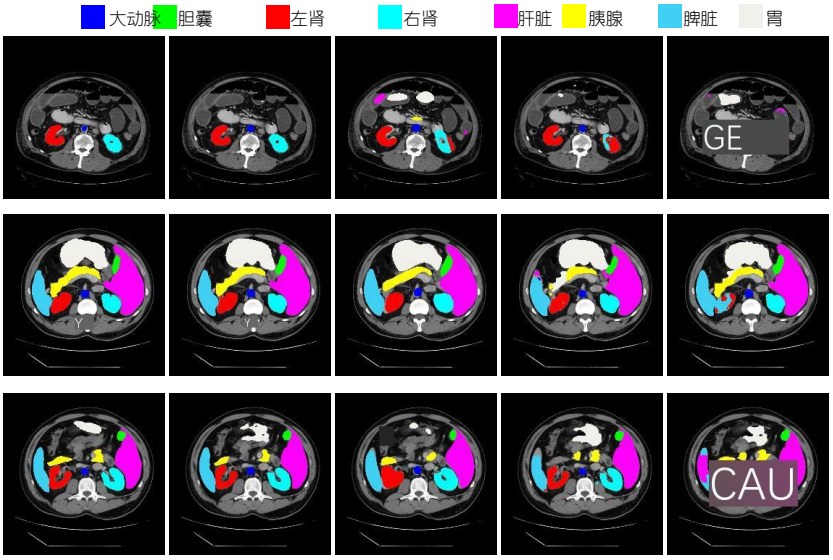


图3：通过可视化定性比较不同方法。从左至右：（a）Ground Truth，（B）TransUnet，（C）R50-Vit-Cup，（d）R50-AttNunet，（e）R50U-Net。我们的方法预测较少的假阳性，并保持更精细的信息。

表5：DSC中ACDC数据集的比较（%）。

| 框架 | 平均值 | RV | Myo | LV |
|---|---|---|---|---|
| R50-U-Net | 87.55 | 87.10 | 80.63 | 94.92 |
| R50-AttnUNet | 86.75 | 87.58 | 79.20 | 93.47 |
| 维特杯 | 81.45 | 81.46 | 70.71 | 92.18 |
| R50-ViT-CUP | 87.57 | 86.07 | 81.88 | 94.75 |
| TransUNet | 89.71 | 88.86 | 84.53 | 95.73 |

### 4.5 可视化

我们提供了Synapse数据集的定性比较结果，如图所示3. 可以看出：1）纯粹基于CNN的方法U-Net和Attnunet更可能对器官进行过度分割或分割不足（例如，在第二行中，脾脏被Attnunet过度分割而被Unet分割不足），这表明基于Transformer的模型，例如我们的TransUnet或R50-Vit-Cup具有更强的编码全局上下文和区分语义的能力。2）第一行中的结果显示，与其他方法相比，我们的TransUnet预测的假阳性更少，这表明TransUnet在抑制这些噪声预测方面比其他方法更有优势。
3）为了在基于变压器的模型中进行比较，我们可以观察到，关于边界和形状，R50-Vit-Cup的预测往往比Transunet的预测更粗糙（例如，第二个模型中的胰腺预测

行）。此外，在第三行中，TransUnet正确地预测了左肾和右肾，而R50-Vit-Cup错误地填充了左肾的内孔。这些观察结果表明，TransUnet能够进行更精细的分割并保留详细的形状信息。这是因为TransUnet同时享有高级全局上下文信息和低级细节的好处，而R50-Vit-Cup仅依赖于高级语义特征。这再次验证了我们最初的直觉，即将类似U-Net的跳跃连接集成到变压器设计中，以实现精确定位。

### 4.6 泛化到其他数据集

为了展示我们的TransUnet的泛化能力，我们进一步评估了其他成像模式，即针对自动心脏分割的Mr数据集ACDC。我们观察到TransUNET相对于纯CNN方法（R50-UNET和R50-ATTNUNET）和其他基于Transformer的基线（Vit-CUP和R50-Vit-CUP）的一致改进，这与Synapse CT数据集的先前结果相似。

## 5结论

变压器被认为是具有强大的内在自我关注机制的架构。在这篇论文中，我们首次研究了变压器在一般医学图像分割中的应用。为了充分利用Transformer的强大功能，提出了TransUnet，它不仅通过将图像特征视为序列来编码强全局上下文，而且通过U型混合架构设计很好地利用了低级CNN特征。作为主流的基于FCN的医学图像分割方法的替代框架，Transunet的性能优于各种竞争方法，包括基于CNN的自注意方法。

致谢。这项工作得到了Lustgarten基金会胰腺癌研究的支持。

## 参考文献

1. Child, R., Gray, S., Radford, A., Sutskever, I.: 用稀疏变换器生成长序列。arXiv预印本arXiv: 1904.10509 (2019)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: 大规模分层图像数据库。见: 2009年IEEE计算机视觉与模式识别会议。PP。248 - 255.IEEE (2009)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: 用于语言理解的深度双向转换器的预训练。arXiv预印本arXiv: 1810.04805 (2018)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S.等人: 图像价值16x16字: 大规模图像识别的变压器。见: ICLR (2021)

5. Fu, S., Lu, Y., Wang, Y., Zhou, Y.、Shen, W.、Fishman, E.、Yuille, A.: 3D多器官分割的领域自适应关系推理。见: 医学图像计算和计算机辅助干预国际会议。PP。656 – 666.斯普林格 (2020)
6. He, K., Zhang, X., Ren, S., Sun, J.: 图像识别的深度残差学习。见: IEEE计算机视觉和模式识别会议论文集。PP。770 – 778 (2016)
7. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-DenseUNET: 用于从CT体积分割肝脏和肿瘤的混合密集连接UNET。IEEE医学成像汇刊37 (12), 2663 – 2674 (2018)
8. Long, J., Shelhamer, E., Darrell, T.: 用于语义分割的全卷积网络。见: IEEE计算机视觉和模式识别会议论文集。PP。3431 – 3440 (2015)
9. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: 用于体积医学图像分割的全卷积神经网络。见: 3DV (2016)
10. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B.等人: 注意U-Net: 学习在哪里寻找胰腺。MIDL (2018)
11. Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., Tran, D.: 图像转换器。见: 机器学习国际会议。PP。4055 – 4064.PMLR (2018)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: 生物医学图像分割的卷积网络。见: 医学图像计算和计算机辅助干预国际会议。PP。234 – 241.斯普林格 (2015)
13. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: 注意力门控网络: 学习利用医学图像中的显著区域。医学图像分析53, 197 – 207 (2019)
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N.、Kaiser, L.、Polosukhin, I.: 你需要的只是关注。见: 神经信息处理系统的进展。PP。5998 – 6008 (2017)
15. Wang, X., Girshick, R., Gupta, A., He, K.: 非局部神经网络。见: IEEE计算机视觉和模式识别会议论文集。PP。7794 – 7803 (2018)
16. Yu, L., Cheng, J.Z., Dou, Q., Yang, X., Chen, H., Qin, J., Heng, P.A.: 使用密集连接的体积ConvNets的自动3D心血管Mr分割。见: 医学图像计算和计算机辅助干预国际会议。PP。287 – 295.斯普林格 (2017)
17. Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E.K., Yuille, A.L.: 循环显著性转换网络: 整合小器官分割的多阶段视觉线索。见: IEEE计算机视觉和模式识别会议论文集。PP。8280 – 8289 (2018)
18. Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H.等: 用变形金刚从序列到序列的角度重新思考语义分割。arXiv预印本arXiv: 2012.15840 (2020)
19. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L.: 腹部CT扫描中胰腺分割的固定点模型。见: 医学图像计算和计算机辅助干预国际会议。PP。693 – 701.斯普林格 (2017)
20. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNET++: 用于医学图像分割的嵌套U-Net架构。见: 医学IM中的深度学习

由于长度过长，标题被取消 13

临床决策支持的年龄分析和多模态学习，3 - 11. 斯普林格（2018）