Computer Science and Application 计算机科学与应用, 2024, 14(12), 97-110
Published Online December 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.1412244

Hans汉斯

TGU-Net：基于纹理与语义感知增强的3D肝脏及其肿瘤分割

朱　峰，吴　俊，张　航

扬州大学信息工程学院，江苏　扬州

收稿日期：2024年11月2日；录用日期：2024年12月2日；发布日期：2024年12月11日

摘　要

肝肿瘤的早期诊断对提高患者生存率至关重要，而精准的肝肿瘤分割在诊疗过程中具有关键作用。然而，传统的分割方法依赖于医生的手动操作，既耗时耗力，也容易受到医生主观经验的影响。近年来，卷积神经网络和Transformer等技术在肝肿瘤分割上取得了一定进展，但仍面临特征提取不足和收敛速度慢等挑战。具体而言，现有方法通常过于关注肿瘤整体形状、位置等全局信息，而忽视了肿瘤边缘模糊、内部结构复杂等局部细节，这些细节对提高分割精度至关重要。同时，尽管Transformer在捕捉长距离依赖和全局上下文方面具有优势，但未能有效结合肝肿瘤的结构特征，影响了模型的分割效果和效率。为解决这些问题，本文基于3D-UNet提出改进的TGU-Net。首先在跳跃连接中加入了纹理增强模块(Texture Enhancement Module)，通过多分支、多尺度3D卷积核选择机制，更好地提取局部特征并捕捉边缘的细微梯度变化，从而提高模型对边缘细节的敏感度和分割精度。其次，在3D-UNet的瓶颈层引入了3D Cross-Shaped Transformer模块(Cross-Shaped Transformer)，结合3D Transformer的建模能力与Cross-Shaped自注意力机制，使模型更精准地聚焦于肿瘤区域的语义信息，提高对肿瘤复杂形态的理解能力。为进一步提高模型的训练效率，本文在该模块之前加入3D深度可分离卷积的先验层(Local Encoding Module)，通过分离空间和通道的卷积操作，提升了特征提取的效率并加快训练速度。在LiTS2017数据集上的实验验证表明，TGU-Net的IOU和Dice指标分别提升了3.89和2.57个百分点，相较于多种SOTA算法表现优异，证明了其在肝肿瘤分割任务中的优越性。

关键词

肝肿瘤分割，3D-UNet，Transformer，深度学习，医学图像分割

TGU-Net: 3D Liver and Tumor Segmentation Based on Texture and Semantic Perception Enhancement

Feng Zhu, Jun Wu, Hang Zhang

College of Information Engineering, Yangzhou University, Yangzhou Jiangsu

文章引用: 朱峰, 吴俊, 张航. TGU-Net: 基于纹理与语义感知增强的 3D 肝脏及其肿瘤分割[J]. 计算机科学与应用, 2024, 14(12): 97-110. DOI: 10.12677/csa.2024.1412244

Received: Nov. 2nd, 2024; accepted: Dec. 2nd, 2024; published: Dec. 11th, 2024

Abstract

Early diagnosis of liver tumors is critical for improving patient survival rates, and precise liver tumor segmentation plays a key role in treatment planning. However, traditional segmentation methods rely on manual operations by clinicians, which are time-consuming, labor-intensive, and often influenced by subjective experience. Recently, technologies like convolutional neural networks (CNNs) and Transformers have achieved progress in liver tumor segmentation, yet challenges remain in feature extraction and model convergence speed. Specifically, existing methods often overemphasize global features, such as the overall shape, location, and size of the tumor, while overlooking local details, including blurred tumor edges and complex internal structures, which are essential for improving segmentation accuracy. Although Transformers excel at capturing long-range dependencies and global context, they have yet to effectively incorporate the structural characteristics of liver tumors, impacting segmentation performance and model efficiency. To address these issues, this paper proposes an enhanced TGU-Net model based on the 3D-UNet architecture. First, a Texture Enhancement Module is introduced into the skip connections, employing a multi-branch, multi-scale 3D convolutional kernel selection mechanism. This module better captures local features and fine gradient changes around tumor edges, thereby enhancing the model's sensitivity to edge details and improving segmentation accuracy. Next, a 3D Cross-Shaped Transformer module is incorporated in the bottleneck layer of 3D-UNet. By combining the 3D Transformer's modeling capability with Cross-Shaped self-attention, the model achieves more precise focus on the semantic information of tumor regions, enhancing its ability to understand complex tumor morphologies. To further improve training efficiency, a Local Encoding Module using 3D depthwise separable convolutions is added before this module, separating spatial and channel convolutions to accelerate training and improve feature extraction efficiency. Experimental validation on the LiTS2017 dataset demonstrates that TGU-Net improves IOU and Dice scores by 3.89 and 2.57 percentage points, respectively, outperforming multiple state-of-the-art algorithms and underscoring its superiority in liver tumor segmentation tasks.

Keywords

Liver Tumor Segmentation, 3D-UNet, Transformer, Deep Learning, Medical Image Segmentation

1. 引言

肝肿瘤是泌尿系统常见的恶性肿瘤之一，其中透明细胞癌最常见，发病率在欧美国家高于亚洲，城市高于农村。肿瘤可分为良性和恶性，良性包括血管平滑肌脂肪瘤等，恶性包括透明细胞癌、肝细胞癌等[1]。随着疾病进展，常出现无痛血尿、腰痛、腰部肿块等症状，提示肿瘤已侵入肝盂。早期发现和治疗是提高患者生存率的关键，而肝肿瘤分割作为医学图像处理的重要任务，其精度直接影响诊断与治疗。传统手动分割耗时且受医生主观因素影响，已无法满足临床需求，自动分割技术的开发成为迫切需求。精确分割可帮助医生更准确判断肿瘤特征，制定更合适的治疗方案，减少不必要的手术操作，提高手术质量与患者生存质量。肝肿瘤分割的研究也能推动医学图像处理技术的进步。然而，肝肿瘤的图像边缘

模糊和形状不规则给分割带来挑战，现有网络难以捕捉这些特征，因此需要改进算法以提高分割精度和效率[2]。

近年来，深度学习技术在医学图像分割领域取得了显著进展，极大地提升了多种医学任务的分割精度。其中，U-Net [3]是 Ronneberger 等人在 2015 年提出的一种经典模型。其“U”形结构包括一个对称的编码器和解码器，通过下采样提取特征，并利用上采样恢复空间信息。同时，U-Net 通过跳跃连接将编码器中的特征与解码器中的特征相融合，从而实现对图像的精准分割。然而，U-Net 的网络结构相对简单，导致其特征提取能力较弱，从而在某些分割任务中表现不佳。为了解决这一问题，U-Net++ [4]作为 U-Net 的改进版本，通过引入更密集的跳跃连接和嵌套结构，增强了不同尺度特征之间的交互，显著提升了医学图像分割的精度，特别适用于细微结构的分割任务。然而，U-Net++的复杂结构导致了计算量和参数量的显著增加，进而降低了训练效率。此外，过多的跳跃连接可能引入冗余信息，影响模型的性能。ResUnet [5]和 ResUnet++ [6]结合了 U-Net 和残差网络的优势，通过引入残差模块来增强特征提取能力，并有效解决了深层网络中的梯度消失问题，使其在处理复杂结构和模糊边界的医学图像分割任务中表现更加出色。在 3D 医学图像分割中，3D-UNet [7]将 U-Net 扩展至支持三维医学体素数据的网络结构，采用 3D 卷积和 3D 池化操作来处理三维体积数据，从而在所有三个维度上捕捉上下文信息。这种设计在分割器官、肿瘤等具有立体结构的任务中表现优异。nnU-Net [8]是一种自适应的医学图像分割框架，支持 2D 与 3D 医学图像分割，其核心思想是在不设计新网络结构的前提下，通过自动调整现有的 U-Net 架构来适应不同的任务。nnU-Net 能够根据数据集的特点自动调整网络的超参数(如网络深度、卷积核大小、学习率等)，为每个具体的分割任务生成最优的 U-Net 配置。其强大的通用性和无需人为设计模型的特性，使 nnU-Net 在多种医学图像分割任务中表现出色。尽管如此，卷积网络在长距离依赖的感知能力上较弱，这在某些语义复杂的分割任务中可能限制其高精度表现，例如肝肿瘤的分割任务。

最近，基于 Transformer 的医学图像分割网络相较于传统的卷积神经网络获得了越来越多的关注。例如，TransUnet [9]和 TransFuse [10]是结合了 Transformer 和 U-Net 的医学图像分割模型，前者将 Transformer 的全局信息提取能力与 U-Net 的局部特征捕获相结合，从而能够更好地捕捉复杂图像中的全局和细节信息，提升了分割精度，特别适用于具有复杂结构和多尺度特征的医学图像。同时，SwinUnet [11]作为一种基于 Swin Transformer [12]的医学图像分割模型，结合了 U-Net 的结构和 Swin Transformer 的局部－全局特征提取能力，通过滑动窗口注意力机制显著增强了 Transformer 的全局感受野。这些方法均基于 Transformer 网络，专注于实现 2D 医学图像分割。而在 3D 任务中，Transformer 也引起了关注。例如，SwinUnet r [13]将 SwinUnet 扩展至 3D 医学图像分割，通过将 3D Patch 划分为 Token，实现了对体素中的长距离依赖感知。此外，UNETR [14]是另一种结合 U-Net 和 Transformer 的医学图像分割模型，在传统 U-Net 的编码器－解码器结构中引入了 Transformer，增强了特征提取能力和长距离依赖建模。UNETR 使用卷积操作提取局部特征，并通过 Transformer 捕捉全局上下文，从而在分割任务中更好地理解图像信息。其跳跃连接机制也促进了特征融合，提高了分割精度，特别适合复杂结构的医学图像，如肿瘤和器官的分割。然而，与卷积神经网络相比，Transformer 模块通常缺乏足够的先验知识，这导致在计算注意力时需要更长的时间，可能陷入局部最优解，进而增加训练时间。

卷积神经网络与 Transformer 已广泛应用于肝肿瘤分割算法中。Shuanhu 等人提出了 TD-Net [15]，用于 CT 图像中肝肿瘤的自动分割。TD-Net 结合了 Transformer 和方向信息，采用共享编码器提取多层次特征，并设计了两个解码分支，分别生成初始分割图和方向信息。通过四个跳跃连接保留空间信息，并在第四个跳跃连接中引入 Transformer 模块以提取全局上下文。然而，由于肝肿瘤图像中包含大量局部信息，TD-Net 在保留空间信息时未能很好地平衡全局信息与局部信息的提取。Yang 等人[16]提出了一种结合层次化 Swin Transformer 和 CNN 的双编码路径结构，将其嵌入编码器和解码器中，通过融合长距离依

赖和多尺度上下文连接，捕捉不同语义尺度的粗略特征。然而，Swin Transformer 仍然面临收敛速度较慢的问题。Ruiyang 等人提出的 DHT-Net [17]，通过动态层次变换网络(DHTrans)和边缘聚合块(EAB)，用于肝肿瘤的自动分割，辅助放射科医生进行临床诊断。DHTrans 通过动态自适应卷积自动感知肿瘤位置，并利用不同感受野大小的分层操作学习肿瘤的多样特征，增强语义表示。同时，它通过聚合全局和局部的纹理信息，充分捕捉肿瘤区域的复杂形态特征。EAB 则用于提取网络浅层的细粒度边缘特征，增强肝脏和肿瘤区域的边界清晰度。然而，DHT-Net 仅依靠自适应卷积提供的先验信息，仍需要在训练过程中进一步优化肝脏和肿瘤区域边界的生成。综上所述，近年来这些方法虽然在肝肿瘤分割任务中取得了进展，但仍存在两个主要问题：其一，特征提取过程中往往偏重语义信息，而忽视了肝肿瘤数据中细微的局部信息；其二，Transformer 的应用未能根据肝肿瘤的特性进行有效的先验设置，导致收敛速度较慢。

本文旨在解决现有肝肿瘤分割方法在特征提取和 Transformer 适应性方面的不足，提出了一种基于 3D-UNet 的改进网络结构 TGU-Net，以增强对肝肿瘤局部信息的捕捉能力和提高分割效果。为了增强网络对肝肿瘤的局部信息提取能力，也为了使得 Transformer 在提取肝脏肿瘤的过程中拥有合适的先验信息，本文的贡献点如下：

1. 增强网络对肝肿瘤的纹理信息提取能力：本文提出的 TGU-Net 在特征提取过程中，针对肝肿瘤图像中的细微局部信息进行了有效增强。通过在前几个跳跃连接中引入基于纹理增强的模块(TEM)，采用多分支、多尺度的 3D 卷积核选择注意力机制，TGU-Net 能够更精准地捕捉肿瘤周围的细微梯度变化，从而提高分割精度。这一改进解决了现有方法中过于关注全局语义信息而忽视局部信息的问题。

2. 针对肝脏肿瘤设计更为合适的 Transformer：为了提升 Transformer 在肝肿瘤分割中的表现，本文在 3D Unet 的瓶颈层中引入了基于十字交叉窗口注意力机制的 3D Transformer 模块(CST)。该模块结合了 3D Transformer 的建模能力与精准聚焦特性，能够更好地捕捉肿瘤复杂形态的语义信息。这一设计有效改善了 Transformer 未能根据肝肿瘤不规则的形状进行有效设计的局限性。为进一步加快模型收敛速度，本文在 3D Transformer 模块前加入了 3D 深度可分离卷积先验模块(LEM)。

3. 精度提升：通过改进，TGU-Net 在 LiTS2017 数据集上的评估结果显示，IOU 指标和 Dice 指标分别提升了 3.89 和 2.57 个百分点，超越了多种最新的 SOTA 算法。

2. 方法

2.1. 方法架构

图 1 展示了本文提出的改进版 TGU-Net 网络结构，输入的图像尺寸为 $I \in \mathbb{R}^{H\times W\times D}$。$H$，$W$，$D$ 分别代表 3D 输入图像的长度，宽度与深度，相较于原始 3D-UNet，该改进网络在前三个阶段的跳跃连接中加入了一种纹理增强模块(TEM, Texture Enhancement Module)。在网络的 BottleNeck 中，我们还设计了语义感知增强模块(CST, Cross-Shaped Transformer)，旨在提升肿瘤语义信息的提取效果。我们还设计了一种深度可分离先验模块(LEM, Local Encode Module)与 CST 相结合，加快 Transformer 的收敛速度。通过在 3D-UNet 的 BottleNeck 上采用深度可分离卷积与十字交叉窗口注意力机制的 3DTransformer 的混合设计，从而增强模型的整体语义理解能力。TGU-Net 的设计综合考虑了语义信息与纹理信息，在提取特征的过程中对纹理信息和语义信息进行增强。网络最终输入的分割图像尺寸为 $O \in \mathbb{R}^{H\times W\times D}$。

2.2. 纹理增强模块

本文设计了一种 3D 多分支选择性核的纹理增强模块(TEM)，其构建基于对 SKNet [18]心理念的深度结合，并将这些关键组件改进为适合处理三维数据的形式。通过将 SKNet 的核心模块重组为 3D 结构，TEM 模块能够更高效地捕获并强化输入数据中的纹理特征。

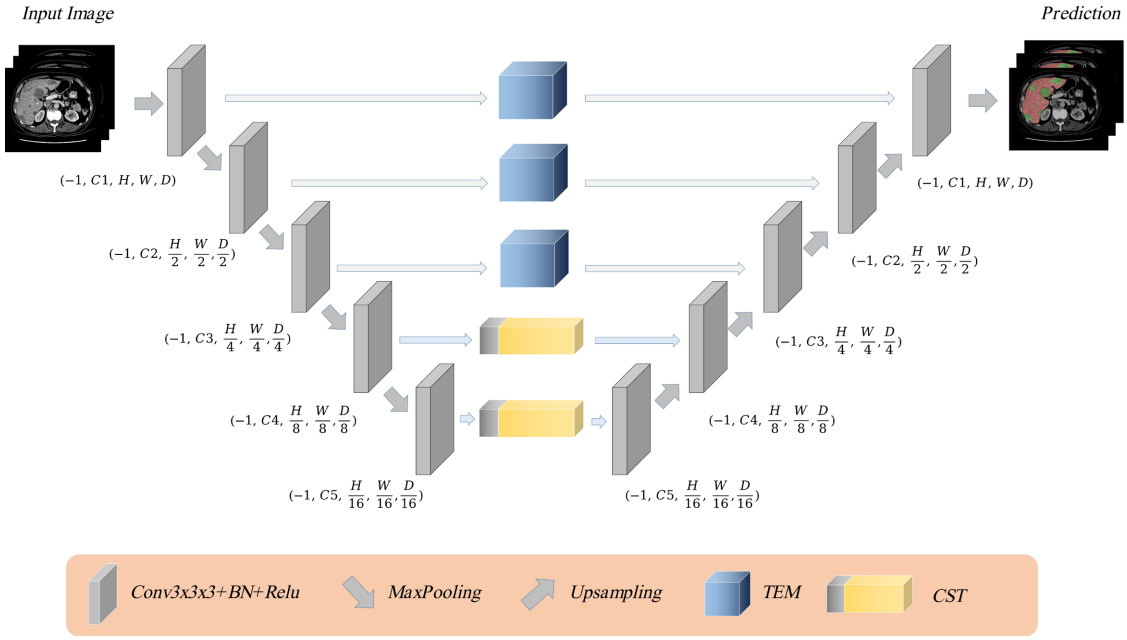

Figure 1. TGU-Net network architecture, where TEM represents the Texture Enhancement Module, and CST stands for the Cross-Shaped Transformer
图 1. TGU-Net 网络结构图，其中 TEM 为纹理增强模块，CST 为语义感知增强模块

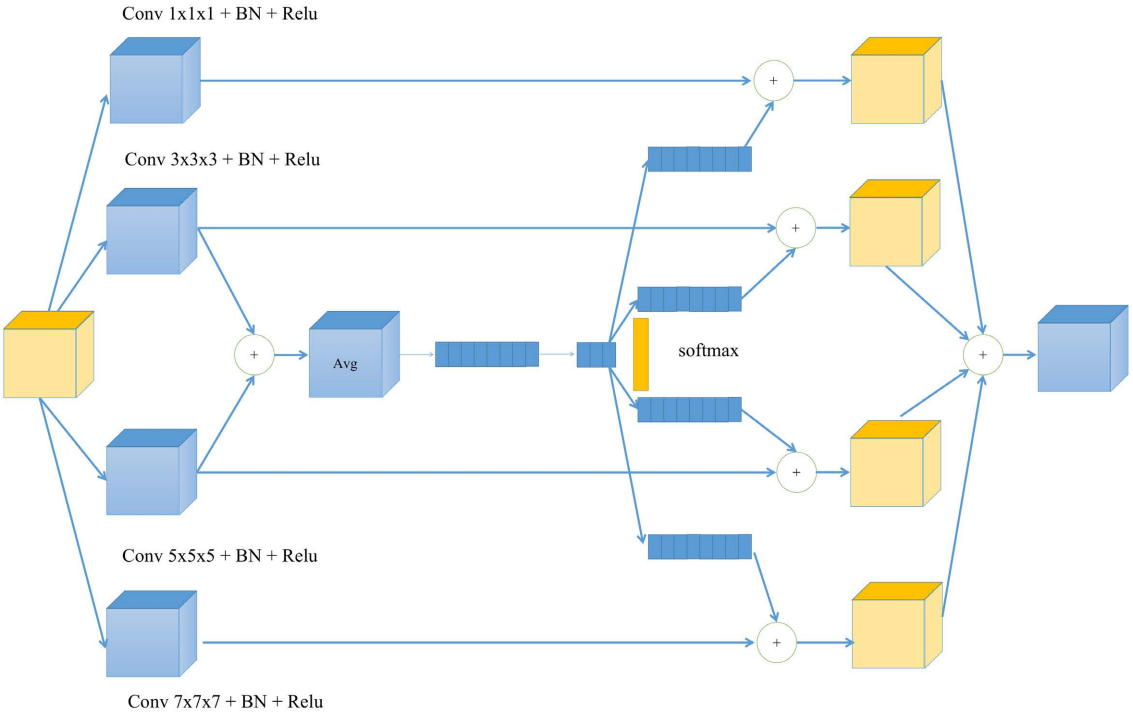

Figure 2. TEM network structure, where blue represents the network modules, and yellow indicates the feature maps
图 2. TEM 网络结构，其中，蓝色为网络模块，黄色为特征图

如图 2 所示，对于输入的特征图，其尺寸为 $f \in \mathbb{R}^{H\times W\times D\times C}$，其中 H、W、D 分别表示高度、宽度和深度，C 代表通道数或特征维度。该特征图随后被输入到四个并行且独立的分支中，每个分支使用不同大

小的卷积核来提取多尺度特征信息。具体来说，第一个分支采用的是 3 × 3 × 3 卷积核，第二个分支使用 5 × 5 × 5 卷积核，第三个分支采用 7 × 7 × 7 卷积核，最后，第四个分支选择了 9 × 9 × 9 卷积核，以确保从不同尺度充分捕捉输入特征中的关键信息。

$$U = U_{3\times3\times3} + U_{5\times5\times5} + U_{7\times7\times7} + U_{9\times9\times9} \tag{1}$$

如公式 1 所示，不同大小的卷积核可以捕捉特征图中不同层次和范围的细节，从而增强模型的特征表示能力。除了 3D 卷积外，TEM 在特征提取过程中引入了 3D 批量归一化(3D Batch Normalization, 3DBN)和 ReLU 激活函数。在四个分支的卷积操作完成后，每个分支将生成一组特征图。然后，这些特征图相加后，经过 3D 全局均值池化层，该池化层通过对空间维度(H、W、D)进行全局平均值计算，将特征图压缩为仅包含通道信息的特征向量。随后，该特征向量将通过两个连续的 1 × 1 × 1 卷积层处理计算注意力分数，首先通过降维降低计算复杂度，接着再升维，使其恢复到与初始通道数 C 相同的维度。

$$\mathrm{d} = \max\left(\frac{C}{r}, L\right) \tag{2}$$

如公式 2 所示，为两个 1 × 1 × 1 卷积层相连，分别用于降维和升维操作，其中 r 代表降维因子，L 是常量。这一降维－升维过程帮助模型学习更高级的特征表达。完成这些操作后，四个分支通过 Softmax 函数计算相应的注意力权重，输出的权重矩阵形状为 $1\times1\times1\times C$，表示每个通道的权重分布。接着，这些注意力分数与分支特征图逐元素相乘，实现特征加权处理。通过这种加权，模型可以突出重要特征并抑制不重要的特征，增强判别能力。最终，将四个加权后的特征图进行相加融合，得到最终输出的特征图。该融合特征图汇集了多尺度的关键信息。TEM 模块的设计主要是为了显著增强网络在捕捉纹理信息方面的能力，提升整体性能。这一设计基于对现有模型在肾脏肿瘤图像处理中表现不足的深入分析。传统注意力机制在处理复杂的细微纹理信息时，常常难以精确捕捉这些特征，尤其当纹理细节十分复杂时更为困难。为了解决这一问题，TEM 模块融合了 SKNet 和 SENet 的策略。SKNet 通过动态选择不同尺寸的卷积核，满足了纹理特征提取的需求，确保模型能够准确捕捉目标区域的感受野。与此同时，SENet 通过关注通道间的关系，增强了模型捕捉全局纹理信息的能力，在 3D-UNet 中引入 TEM 模块后，模型在纹理信息提取方面表现出了显著的改进，可以有效提升整体性能，旨在实现更加全面、精确的肾脏肿瘤分割任务。

2.3. 语义感知增强模块

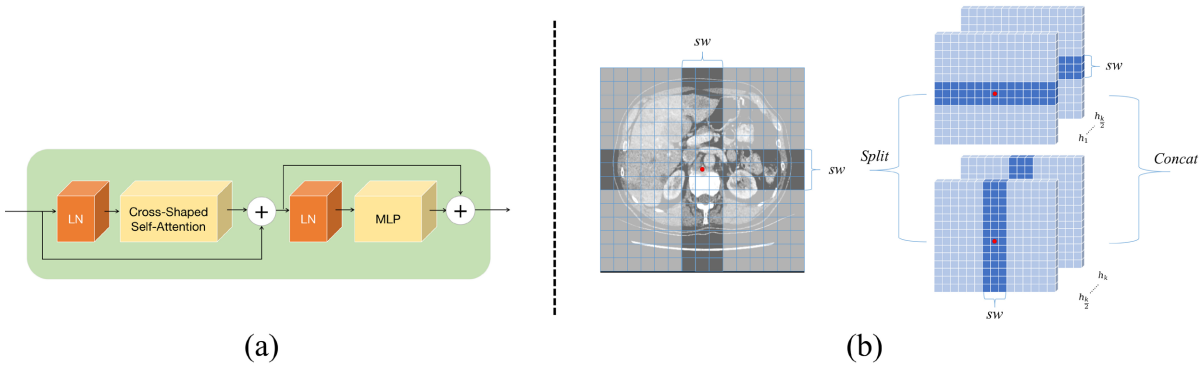

Figure 3. Cross-Shaped Transformer network structure. (a) shows the components of the Cross-Shaped Transformer, and (b) illustrates the cross-shaped region in the Cross-Shaped Self-Attention

图 3. Cross-Shaped Transformer 网络结构：(a)展示了基于十字交叉窗口注意力机制 Transformer 的组成部分，(b)展示了十字交叉窗口注意力机制

肾肿瘤区域往往是不规则的，而现有的 Transformer [19]在进行语义感知的过程中，往往只使用 Patch 作为 Token，而 Patch 局限了 Transformer 获得更广的长距离依赖，基于此，我们根据肾肿瘤的特点，设计了一种基于十字交叉窗口注意力机制 Transformer。

为了使网络结构有效捕捉长距离依赖，本文在 BottleNeck 中设计了 Cross-Shaped Transformer (CST) 作为核心组件，如图 3(a)所示。CST 包括层归一化、十字交叉窗口注意力机制(CSA)和多层感知机(MLP)。该结构中的每个层归一化模块均配备了残差连接，以有效减轻训练过程中的梯度消失问题。多层感知机由两层构成，采用 GELU 激活函数，以增强模型的非线性表达能力。CST 的计算过程可如下所示：

$$\hat{X}_l = \mathrm{CSA}\left(\mathrm{LN}\left(X_{l-1}\right)\right) + X_{l-1} \tag{3}$$

$$X_l = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{X}_l\right)\right) + \hat{X}_l \tag{4}$$

在此公式(1)和(2)中，X_{l-1} 代表前一层的输出结果，而 X_l 则表示当前层的输出。在该结构中，特征图被传递到长距离依赖感知模块。在 CST 的设计中，构建了一个独特的交叉窗口，允许在水平和垂直条带上并行进行自注意力计算，从而实现十字交叉窗口注意力机制(CSA)，其结构如图 3(b)所示。通过这种设计，模型能够有效捕捉图像中远距离像素之间的关系，从而增强整体特征提取的能力。

在 CSA 中，输入特征首先通过线性投影映射到 K 个头中。随后，每个头在水平或垂直条带上进行局部自注意力的计算。在进行水平条带的自注意力计算时，特征 X 被均匀划分为多个相同宽度的、不重叠的水平条带 $\left[X^1, X^2, \cdots, X^M\right]$，每个条带中包含 $sw \times W$ 个 token。这里的 sw 表示条带的宽度，可以根据需要进行调整，以在学习能力与计算复杂性之间实现平衡。形式上，假设第 k 个头的查询、键和值的投影维度分别为 W_k^Q， W_k^K， W_k^V。最终，第 k 个头在水平条带上执行自注意力计算后的输出可表述为：

$$X = \left[X^1, X^2, \cdots, X^M\right] \tag{5}$$

$$Y_k^i = \mathrm{Att}\left(X^1 W_k^Q, X^2 W_k^K, X^M W_k^V\right) \tag{6}$$

$$\mathrm{HAtt}_k\left(X\right) = \left[Y_k^1, Y_k^2, \cdots, Y_k^M\right] \tag{7}$$

在该模型中 $X^i \in \mathbb{R}^{(sw \times W) \times C}$， $M \in H/sw$， $i = 1, \cdots, M$， $W_k^Q \in \mathbb{R}^{C \times d_k}$， $W_k^K \in \mathbb{R}^{C \times d_k}$， $W_k^V \in \mathbb{R}^{C \times d_k}$ 分别代表第 k 个头所使用的查询、键和值的投影矩阵，而 C/K 则用于表示特征维度的划分方式。针对垂直条带的自注意力计算，采用类似的推导方法来获得结果，第 k 个头的输出用 $\mathrm{VAtt}_k\left(X\right)$ 来表示。考虑到头颈部医学图像的特性，假定其中不存在任何方向性偏差。本研究将所有的 K 个头分为两组，每组包含 $K/2$ 个头，通常情况下 K 是一个偶数。在这两组中，第一组负责执行水平条带的自注意力计算，而第二组则专注于垂直条带的自注意力计算。最终，两组的输出结果将被拼接在一起，以形成完整的特征表示。

$$\mathrm{CSA}\left(X\right) = \mathrm{Concat}\left(\mathrm{head}_1, \cdots, \mathrm{head}_k\right) W^O \tag{8}$$

$$\mathrm{head}_k = \begin{cases} \mathrm{HAtt}_k\left(X\right), & k = 1, \cdots, \dfrac{K}{2} \\ \mathrm{VAtt}_k\left(X\right), & k = \dfrac{K}{2}, \cdots, K \end{cases} \tag{9}$$

在公式(6)与(7)中，CSA 代表十字交叉窗口注意力机制，W^O 被用作将自注意力的输出转换为目标输出维度(通常设置为 C)的标准投影矩阵。正如前面所提到的，设计自注意力机制时一个关键思路是将多头注意力划分为若干组，并针对不同组应用各自的自注意力操作。通过这种分组方法，Transformer 模块中每个 token 的注意力范围得以有效扩展。这种策略与传统自注意力机制形成鲜明对比，后者通常对所有多头使

用统一的自注意力计算方式。基于十字交叉窗口作为 token，更符合肾肿瘤区域的不规则性，可以更好的适配肾肿瘤区域，提升分割精度。

2.4. 深度可分离先验层

由于 Transformer 在捕捉长距离依赖关系时通常会面临较慢的计算速度和高内存占用问题，因此本文还设计了基于深度可分离卷积层[20]的 LEM (Local Encoding Module)，为 CST (Cross-Shaped Transformer) 提供偏置归纳。这一设计的目的是通过引入深度可分离卷积的特性，有效提高特征提取的效率，从而在保持模型表达能力的同时显著减少计算复杂度和参数数量。这种方式不仅能加快模型，还能提升在处理复杂特征时的灵活性和响应能力，从而进一步增强 CST 在长距离依赖捕捉中的表现，如图 4 所示：

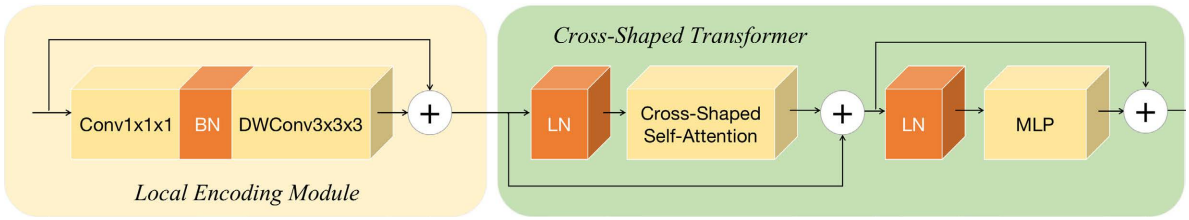

Figure 4. Local Encoding Module and Cross-Shaped Transformer network structure
图 4. Local Encoding Module 和 Cross-Shaped Transformer 的网络结构图

对于输入特征图，在进行先验感知的 LEM 之后，有：

$$\mathrm{LEM}(X)=\mathrm{DConv}_{3\times3\times3}\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1\times1}(X)\right)\right)+X \tag{10}$$

在公式(10)中，LEM 模块由 1 × 1 × 1 卷积、批量归一化和 3 × 3 × 3 深度可分离卷积组成。经过 LEM 模块处理的特征会被送入 CST。

2.5. 损失函数

在本文中，我们采用了交叉熵损失函数与 Dice 损失函数两种损失函数。交叉熵损失函数(Cross-Entropy Loss)是语义分割任务中常用的一种损失函数，它衡量的是模型输出的预测概率分布与真实标签分布之间的差异。在语义分割中，交叉熵损失函数用于逐像素地计算预测类别与真实类别之间的差异。具体来说，对于每个像素，网络会预测它属于各个类别的概率，而交叉熵损失会根据该预测概率与真实类别之间的差异来更新网络的权重。交叉熵损失函数的公式如下：

$$L_{\mathrm{CE}}=-\sum_{i=1}^{C}g_i\cdot\log(p_i) \tag{11}$$

C 表示类别的总数，g_i 是第 i 类的真实标签，采用 one-hot 编码，即属于该类时为 1，否则为 0，p_i 是模型预测属于第 i 类的概率。

$$L_{\mathrm{Dice}}=1-\frac{2\sum_{i=1}^{N}p_ig_i}{\sum_{i=1}^{N}p_i^2+\sum_{i=1}^{N}g_i^2} \tag{12}$$

p_i 是模型预测的第 i 个像素的值，通常是一个介于 0 和 1 之间的概率。g_i 是第 i 个像素的真实标签，通常是二进制值。0 或 1。N 是像素的总数。

$$L=L_{\mathrm{CE}}+L_{\mathrm{Dice}} \tag{13}$$

最终，本文使用的损失为交叉熵损失与 Dice 损失的和。

3. 实验

3.1. 数据集

LiTS (The Liver Tumor Segmentation Benchmark)数据集[21]专注于肝脏及其肿瘤的自动化分割，广泛应用于医学图像分割研究。该数据集由来自 7 个不同医学中心的 CT 图像组成，包含 131 例训练数据和 70 例测试数据，涵盖了各种肝脏和肿瘤的形态学特征。这一数据集被用于 ISBI 2017、MICCAI 2017、MICCAI 2018 等多个国际竞赛，并作为 MSD (Medical Segmentation Decathlon)的 Task03，推动了相关分割算法的进展，成为评估肝脏及肝脏肿瘤分割算法的标准基准。

3.2. 数据预处理

为了使网络能够更有效地提取图像中的关键信息，并为后续实验奠定基础，数据预处理过程经过了以下关键：

1) 重采样：我们对原始图像及其标签进行了空间重采样，以确保它们符合预设的体素间距。具体操作包括：在 X 轴和 Y 轴方向上，使用双三次插值法(B-spline)将图像重采样至 0.767578125 毫米的间距；而在 Z 轴方向，基于最近邻插值法将图像调整至 1.0 毫米的间距。标签图像的重采样也采用最近邻插值法，以保证其离散性质不受破坏，避免误差引入。

2) 标签二值化处理：由于插值过程可能对标签数据产生不必要的值变化，因此在重采样后，需对标签进行二值化处理。具体方法是：将所有体素值大于 0.1 的标记为 1，表示前景(如肿瘤或器官)，而小于或等于 0.1 的体素则标记为 0，表示背景。此步骤确保了在插值过程中产生的噪声被有效去除，增强了标签数据的准确性与纯净性。

3) 图像归一化：为了消除由于成像设备和参数差异带来的灰度范围不一致问题，所有图像数据进行了归一化处理。通过将图像灰度值映射到 0 到 1 的范围内，保证了图像间的一致性，减少了不同来源数据的偏差。此外，这种归一化还能改善模型的收敛性，提升训练过程的稳定性和效率。

4) Patch 裁剪：原始的 3D 图像大小为(512, 512, D)，若直接输入会导致显存占用较大，所以使用随机 Patch 裁剪的策略，裁剪出(64, 64, 64)大小的区域传入网络。

这些预处理步骤显著提升了数据的质量，确保模型在一致、优化的数据集上进行训练，从而为模型的精度和性能提供了坚实的基础。

3.3. 实验设置

本文实验的训练配置如下。硬件方面，使用了 4 张 NVIDIA GeForce RTX 3090 显卡，每张显卡配备 24 GB 显存，总共 96 GB 的显存支持大规模深度学习模型的训练。处理器采用了 Intel Xeon Platinum 8280L CPU，主频为 2.60 GHz，提供强大的计算能力。内存为 32 GB，能够满足训练时数据处理和加载的需求。操作系统为 Linux，具有良好的稳定性和兼容性。训练框架采用了 Pytorch，这是一个灵活且高效的深度学习框架。包管理工具使用 Anaconda，用于管理依赖环境。训练过程中的模型性能和指标可通过 Tensorboard 进行可视化和监控。在超参数方面，我们使用了基于 Iteration 的方式进行迭代，一共训练 5000 个 iteration，可以达到收敛状态，我们使用了 Adam 优化器，学习率设置为 1e−3，采用 Poly 学习率调整策略，在 5000 一个 Iterations 中线性调整至 1e−5，Batch Size 设置为 8。其中损失收敛图如图 5 所示。

3.4. 实验指标

在对 3D 医学图像分割任务进行评估时，采用了一系列综合性指标以量化分割结果与真实标签之间的相似度。这些指标包括交并比(IoU)、Dice 系数、精确率(Precision)、召回率(Recall)以及 Kappa 系数，

这五个指标[22]相互补充，共同构建了一个全面的评估框架。具体而言，IoU 和 Dice 系数用于定量评估分割区域与真实区域的重叠程度，Precision 和 Recall 则分别反映了模型的预测准确性与完整性，而 Kappa 系数则衡量了模型预测结果与真实标签之间的一致性。这种多维度的评价方法能够更准确地反映分割算法在实际应用中的性能表现。

$$\text{IoU} = \frac{\text{Overlap}}{\text{Union}} \tag{14}$$

$$\text{Dice} = \frac{2*TP}{FP + 2*TP + FN}. \tag{15}$$

$$\text{Precision} = \frac{TP}{FP + TP} \tag{16}$$

$$\text{Recall} = \frac{TP}{FN + TP} \tag{17}$$

$$\text{Kappa} = \frac{P_0 - P_c}{1 - P_c} \tag{18}$$

在(14)~(18)中，相关的指标包括真阳性(*TP*)、假阳性(*FP*)和假阴性(*FN*)。具体而言，真阳性指的是预测区域与实际标注区域重叠的像素点的数量；假阳性则是指那些被错误标记为正样本的像素，即预测区域中存在但实际上不在真实标注区域内的像素；而假阴性则是那些在真实标注区域中存在但未被预测区域所覆盖的像素点数量。在公式(18)中，观察到的准确率 P_0 是一个重要的统计量，它反映了模型预测结果与实际标签之间一致的比例。另一方面，期望准确率 P_c 则表示在随机情况下，模型预测一致的概率。这两个指标为评估模型的性能提供了基础，有助于量化预测结果的可靠性与有效性。

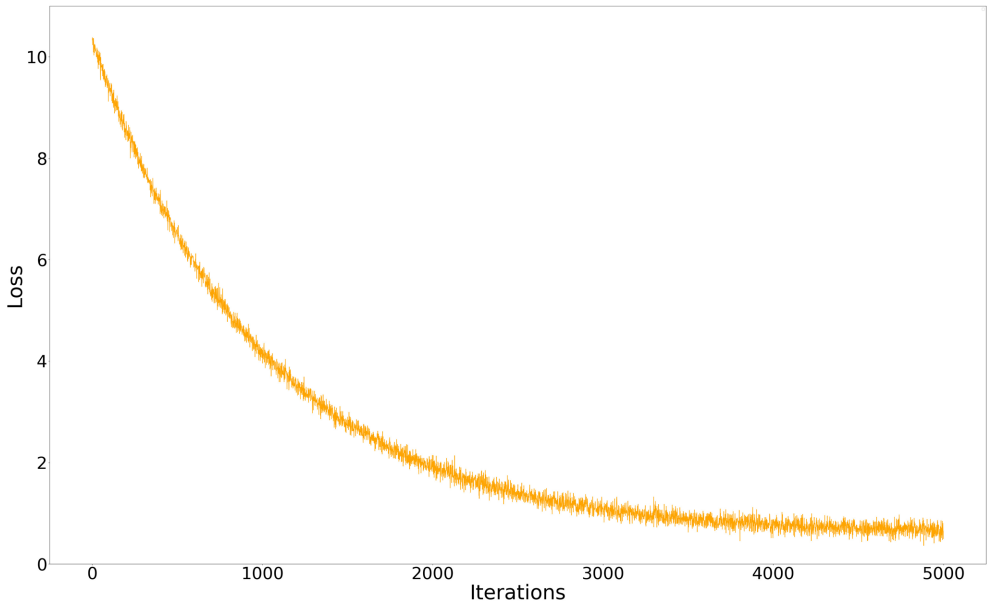

Figure 5. The curve of loss values changes with iterations
图 5. 损失值随着 iteration 的变化曲线

3.5. 对比实验

在本节的对比实验中，针对肾脏肿瘤分割任务，对六种不同算法进行了系统性的性能评估，这些算

法包括 3DU-Net [7]、Unetr [14]、nnformer [23]、SwinUnetr [13]、3DResUnet [24]和 Medformer [25]。其中，Medformer 在肾脏肿瘤分割领域被认为是近年来取得最先进精度的模型。为全面分析各算法的效果，采用了五个关键性能指标进行评估：IoU、Dice 系数、精确率、召回率和 Kappa 系数。这些指标提供了一个客观的框架，用于比较模型之间的性能差异。实验结果表明，所提出的算法在所有评估指标上均显示出显著的优势，尤其在分割精度和结果一致性方面的表现尤为突出。

Table 1. Quantitative analysis results of different methods
表 1. 不同方法的定量分析结果

方法	IoU	Dice	Precision	Recall	Kappa
3DU-Net [7]	74.18	80.85	83.85	83.05	79.04
Unetr [14]	71.05	76.55	81.20	78.30	73.90
nnformer [23]	75.13	81.27	84.72	82.75	78.72
SwinUnetr [13]	76.30	82.21	86.82	83.63	80.18
3DResUnet [24]	76.95	82.52	86.40	83.46	80.60
Medformer [25]	77.59	82.81	88.79	84.08	81.21
TGU-Net	78.29	83.42	89.34	84.31	81.78

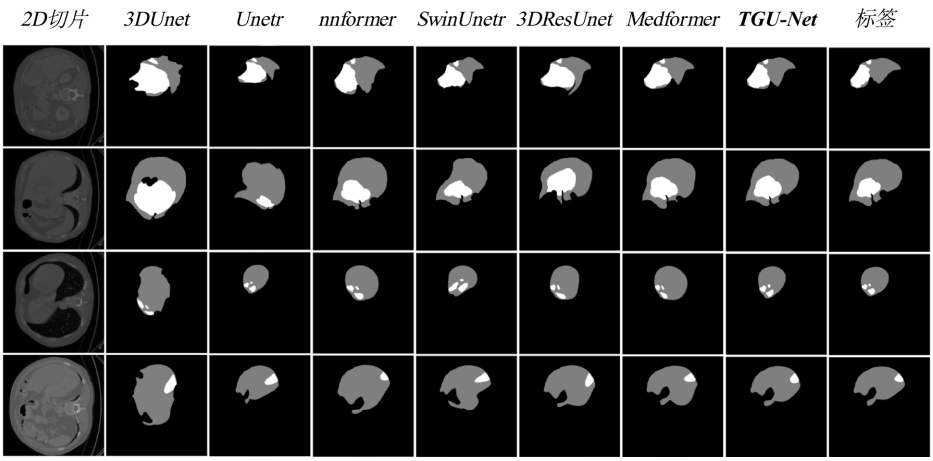

Figure 6. Visualization of segmentation results for different methods
图 6. 不同方法的分割结果

根据表 1 的数据显示，本文提出的算法在交并比(IoU)方面取得了 78.29 的高分，显著高于其他比较模型，这一结果表明该算法在分割区域的重叠精度上具有更为卓越的能力。该性能不仅体现了算法在处理复杂肾脏肿瘤形态时的鲁棒性，还反映出其对边界细节的准确捕捉能力。在 Dice 系数评估中，本文算法的得分为 83.42，再次领先于所有对比模型。作为评估分割准确性的重要指标，这一结果显示出本文算法在匹配分割区域的能力上明显优于其他方法，能够有效应对肾脏肿瘤形状和大小的变化。Precision 和 Recall 两个指标进一步证明了本文算法在平衡假阳性与假阴性方面的优势。具体而言，本文算法的 Precision 达到了 89.34，表明其在分割过程中能够有效减少误判，提高对目标区域的准确识别。同时，Recall 值为 84.31，显示出该算法在检测肾脏肿瘤方面的高效性，显著降低了漏检的风险。在 Kappa 系数评估中，本文算法以 81.78 的得分继续领先。Kappa 系数用于衡量分类结果的一致性，而高 Kappa 值表明该算法在面对多种病例和复杂形态时，能够保持较高的分类一致性和可靠性。综上所述，本文提出的算法在

肾脏肿瘤分割任务中，各项指标均优于传统的 3DU-Net、Unetr、nnformer、SwinUnetr、3DResUnet 以及 Medformer 等方法，展现了其在复杂医学图像处理中的明显优势。这些实验结果进一步验证了所提出的改进策略在实际应用中的有效性，为未来的肾脏肿瘤图像分割技术的研究提供了重要参考。

如图 6 所示，可视化结果清晰地表明，本文提出的算法在肾脏肿瘤分割任务中表现优于其他六种算法。通过对比可以发现，本文方法生成的分割结果与真实标签(GT)具有更高的相似度。在肿瘤边缘的处理上，所提算法展现了更高的精细度，而在肿瘤内部区域的完整性方面也体现出更优的表现。

相比之下，其他算法在处理肿瘤边缘时常出现不连续或模糊的情况，且在捕捉肿瘤内部复杂形态时存在一定的误差。相对而言，本文算法不仅使分割边界更为清晰，而且能够准确保持肿瘤区域的形状和大小，从而显著提高了与 GT 的重叠度。这些结果充分展示了所提算法在肾脏肿瘤分割中的更强能力和稳定性。

3.6. 消融实验

在本研究中，针对模型性能的提升，我们引入了三个改进模块：纹理信息增强模块(TEM)、形状特征增强模块(CST)以及深度可分离先验模块(LEM)。为全面验证每个模块对模型性能的具体贡献，我们进行了系统的消融实验。实验中，我们将 TEM、CST 和 LEM 逐一集成到 3DU-Net 模型中，观察各自的性能提升效果。消融有助于理解各模块在整体框架中的作用，还能揭示它们如何相互作用以优化分割结果。如表 2 所示。

Table 2. Ablation experimental results
表 2. 消融实验结果

方法	IoU	Dice	Precision	Recall	Kappa
3DU-Net	74.18	80.85	83.85	83.05	79.04
3DU-Net + LEM	75.86	81.46	84.87	83.86	79.52
3DU-Net + TEM	76.97	82.50	88.86	84.47	80.59
3DU-Net + CST	76.20	81.63	86.04	84.24	79.75
TGU-Net	78.29	83.42	89.34	84.31	81.78

引入纹理增强模块(TEM)后，模型性能表现出显著的提升。具体而言，交并比(IoU)从 75.86 提升至 76.97，而 Dice 系数则由 81.46 增加至 82.50，Precision 则实现了从 84.87 到 88.86 的显著跃升。这些变化表明，TEM 模块在提升模型对细节纹理信息的敏感性方面起到了关键作用，从而在分割精度和一致性上实现了显著改善。同时，语义增强模块 CTS 的引入同样对模型性能产生了正面的影响，尽管提升幅度相较于 TEM 模块稍显逊色。通过使用 Mix-CS 模块，IoU 小幅上升至 76.20，Precision 也从 84.87 提高到 86.04。这一结果反映了 CST 模块在增强模型对全局语义信息感知能力方面的作用，尤其在提升分割精度方面展现出良好的效果。LEM 对于精度提升是微小的。但 3 个改进都加到 3DU-Net 中后，IoU 指标达到了 78.29。说明 3 个改进点可以相互兼容。

4. 结论

在本文中，我们提出了一种改进的 TGU-Net 网络结构，以应对肾脏肿瘤分割中面临的局部信息提取和全局语义适应性不足的问题。TGU-Net 的核心创新在于引入了纹理增强模块(TEM)、语义感知增强模块(CST)以及深度可分离卷积的 Local Encoding Module (LEM)。TEM 模块通过 3D 多分支选择性核的设计，强化了模型对图像中微小纹理特征的捕捉能力，确保对肿瘤边缘模糊和内部异质性结构的精准提取。

CST 模块则利用十字交叉窗口自注意力机制提升了模型对长距离依赖关系和全局语义信息的捕捉能力，使网络能够更精准地聚焦肿瘤区域复杂的形态和细节。为进一步提高模型效率，LEM 模块通过空间和通道分离的 3D 卷积操作降低了计算复杂度，加快了模型的收敛速度。实验结果显示，TGU-Net 在精度和效率上显著优于传统的 3D-UNet 结构，在肾脏肿瘤分割任务中表现出色，展示了其在医学图像分割领域的广阔应用前景。

参考文献

[1] 丛文铭. 肝脏肿瘤临床病理学研究的回顾与展望[J]. 第二军医大学学报, 2002, 23(5): 468-470.

[2] 杨柳. 临床 CT 图像中肝脏肿瘤分割研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2013.

[3] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[4] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[5] Diakogiannis, F.I., Waldner, F., Caccetta, P. and Wu, C. (2020) Resunet-A: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **162**, 94-114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

[6] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., Lange, T.D., Halvorsen, P., *et al*. (2019) Resunet++: An Advanced Architecture for Medical Image Segmentation. 2019 *IEEE International Symposium on Multimedia*, San Diego, 9-11 December 2019, 225-2255. https://doi.org/10.1109/ism46123.2019.00049

[7] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O. (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: *Lecture Notes in Computer Science*, Springer, 424-432. https://doi.org/10.1007/978-3-319-46723-8_49

[8] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H. (2020) NNU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nature Methods*, **18**, 203-211. https://doi.org/10.1038/s41592-020-01008-z

[9] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.

[10] Zhang, Y., Liu, H. and Hu, Q. (2021) Transfuse: Fusing Transformers and CNNs for Medical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 14-24. https://doi.org/10.1007/978-3-030-87193-2_2

[11] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al*. (2023) Swin-U-Net: U-Net-Like Pure Transformer for Medical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9

[12] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al*. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 9992-10002. https://doi.org/10.1109/iccv48922.2021.00986

[13] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R. and Xu, D. (2022) Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: *Lecture Notes in Computer Science*, Springer, 272-284. https://doi.org/10.1007/978-3-031-08999-2_22

[14] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., *et al*. (2022) UNETR: Transformers for 3D Medical Image Segmentation. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 3-8 January 2022, 1748-1758. https://doi.org/10.1109/wacv51458.2022.00181

[15] Di, S., Zhao, Y., Liao, M., Zhang, F. and Li, X. (2023) TD-Net: A Hybrid End-to-End Network for Automatic Liver Tumor Segmentation from CT Images. *IEEE Journal of Biomedical and Health Informatics*, **27**, 1163-1172. https://doi.org/10.1109/jbhi.2022.3181974

[16] Yang, Z. and Li, S. (2023) Dual-Path Network for Liver and Tumor Segmentation in CT Images Using Swin Transformer Encoding Approach. *Current Medical Imaging*, **19**, 1114-1123. https://doi.org/10.2174/1573405619666221014114953

[17] Li, R., Xu, L., Xie, K., Song, J., Ma, X., Chang, L., *et al*. (2023) DHT-Net: Dynamic Hierarchical Transformer Network for Liver and Tumor Segmentation. *IEEE Journal of Biomedical and Health Informatics*, **27**, 3443-3454. https://doi.org/10.1109/jbhi.2023.3268218

[18] Li, X., Wang, W., Hu, X. and Yang, J. (2019) Selective Kernel Networks. 2019 *IEEE/CVF Conference on Computer*

Vision and Pattern Recognition, Long Beach, 15-20 June 2019, 510-519. https://doi.org/10.1109/cvpr.2019.00060

[19] Dosovitskiy, A. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.

[20] Kaiser, L., Gomez, A.N. and Chollet, F. (2017) Depth Wise Separable Convolutions for Neural Machine Translation.

[21] Bilic, P., Christ, P., Li, H.B., *et al*. (2023) The Liver Tumor Segmentation Benchmark (Lits). *Medical Image Analysis*, **84**, Article 102680.

[22] Taha, A.A. and Hanbury, A. (2015) Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Medical Imaging*, **15**, 1-28. https://doi.org/10.1186/s12880-015-0068-x

[23] Zhou, H.Y., Guo, J., Zhang, Y., *et al*. (2021) NN Former: Interleaved Transformer for Volumetric Segmentation. arXiv: 2109.03201.

[24] Zhang, C., Ai, D., Feng, C., Fan, J., Song, H. and Yang, J. (2020) Dial/Hybrid Cascade 3dresunet for Liver and Tumor Segmentation. *Proceedings of the* 2020 4*th International Conference on Digital Signal Processing*, New York, 19-21 June 2020, 92-96. https://doi.org/10.1145/3408127.3408201

[25] Chowdary, G.J. and Yin, Z. (2024) Med-Former: A Transformer Based Architecture for Medical Image Classification. In: *Lecture Notes in Computer Science*, Springer, 448-457. https://doi.org/10.1007/978-3-031-72120-5_42