

基于VAN_ResUNet的肝脏CT图像分割

李 阳

(中国石油大学(北京)理学院, 北京 102200)

摘要: 在医学影像领域, 肝脏CT图像分割对肝病的早期诊断和治疗规划不可或缺, 但肝脏的复杂形态与周围组织的低对比度使得肝脏分割尤为困难。为此, 提出一种基于VAN_ResUNet的肝脏CT图像分割方法。首先, 在编码层嵌入视觉注意力网络, 加强模型对肝脏细微结构的捕捉能力; 其次, 引入大核注意力机制, 通过分解卷积捕获长距离关系, 加强特征之间的通道交互性; 最后, 利用改进的大核注意力特征融合模块跨层连接特征来提取多尺度信息, 并拼接两条支路关注全局和局部特征, 以显著提升分割准确性。在公开数据集上的实验表明, 所提模型的平均交并比、平均精确率、Dice系数分别为97.02%, 96.10%, 95.12%, 优于现有算法。

关键词: CT图像; 肝脏分割; VAN; LKA; FLKA; ResUNet

DOI: 10.11907/rjdk.241074

中图分类号: TP391.14

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)008-0188-08



Liver CT Image Segmentation Based on VAN_ResUNet

LI Yang

(College of Science, China University of Petroleum(Beijing), Beijing 102200, China)

Abstract: In the field of medical imaging, liver CT image segmentation is indispensable for early diagnosis and treatment planning of liver disease. However, the complex morphology of the liver and the low contrast of surrounding tissues make liver segmentation particularly difficult. Therefore, a liver CT image segmentation method based on VAN_ResUNet is proposed. Firstly, embedding a visual attention network in the encoding layer enhances the model's ability to capture subtle liver structures; Secondly, introducing a large kernel attention mechanism to capture long-distance relationships through decomposed convolutions and enhance channel interactivity between features; Finally, the improved large kernel attention feature fusion module is used to connect features across layers to extract multi-scale information, and two branches are concatenated to focus on global and local features, significantly improving segmentation accuracy. Experiments on public datasets have shown that the proposed model has an average intersection to union ratio, average accuracy, and Dice coefficient of 97.02%, 96.10%, and 95.12%, respectively, which are superior to existing algorithms.

Key Words: CT image; liver segmentation; VAN; LKA; FLKA; ResUNet

0 引言

肝癌是一种发展迅速且及其严重的恶性疾病, 发病率世界排名第5, 死亡率世界排名第3, 素有“癌中之王”称号。肝癌治疗困难的原因主要在于发现较晚, 因此早期诊断便显得尤为重要。肝脏分割是肝脏疾病诊断和评估的基础, 准确提取肝脏区域可使医生更清晰地观察肝脏形态、结构和组织密度, 有助于发现和诊断肝脏疾病。

随着医学影像技术不断进步, 计算机断层扫描(Computed Tomography, CT)在临床肝脏疾病的诊断和治疗规划

中起到越来越关键的作用, 但肝脏的解剖结构包括血管、胆管、肝组织等组织, 因此, 准确、自动地分割肝脏区域挑战巨大。传统肝脏分割方法受肝脏形态学变异、血管与肝实质之间的复杂交互、噪声等因素影响, 存在精度不高和依赖于专业知识的局限性, 因此构建一种能适应复杂解剖结构的肝脏CT自动分割方法具有重要意义。

早期医学图像分割方法主要依赖传统的计算机视觉技术, 包括阈值分割、区域生长、边缘检测等^[1-3]。这些方法在一些简单场景下表现良好, 但医学图像灰度不均匀、噪声大、结构复杂, 因此准确率较低。随着计算机视觉技术的发展, 基于特征的方法逐渐兴起, 通常利用图像中的

收稿日期: 2024-01-24

扫描二维码阅读全文:

作者简介: 李阳(1997-), 女, 硕士, DOBI医疗科技有限公司算法工程师, 研究方向为医疗肿瘤识别。本文通讯作者: 李阳。



纹理、形状、强度等特征进行分割,例如支持向量机、K-means、模糊C-means等^[4-7]。然而,这些方法通常需要手工设计特征,在面对不同图像数据时需要调整参数,缺乏通用性。

深度学习技术的崛起推动了医学图像分割领域的快速发展,特别是卷积神经网络(Convolutional Neural Network, CNN)能使模型从数据中学习特征表示,避免了手工设计特征的繁琐过程^[8]。例如,U-Net、SegNet、DeepLab等深度学习架构被广泛应用于肝脏CT图像分割任务,性能得到了显著提升^[9-11]。同时,注意力网络在其他任务的优秀表现让很多学者将注意力机制应用于肝脏CT图像分割领域,以增强模型对感兴趣区域的程度来提升分割性能。这类方法通常引入注意力模块或其他机制使模型自动学习图像中不同区域的重要性,从而更好地捕捉目标边界和细节,例如DANet、SE-Net等^[12,13]。

虽然,上述网络在一定程度提升了图像的分割性能,但仍存在一些不足。U-Net并未充分融合不同尺度特征,采用编码器—解码器结构先逐渐降低空间分辨率再逐步升高,阻碍了网络中部分信息的传递,尤其是远距离特征。SegNet上采样部分仅使用池化索,导致恢复的图像细节不够精细。DeepLab模型因空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)结构十分复杂,导致模型计算成本较高。其他基于自注意力机制的网络分割模型破坏了图像的二维结构。基于普通卷积的注意力网络分割模型缺乏通道适应性。

为了解决上述问题,本文提出一种基于VAN_ResUNet肝脏分割模型,旨在探索一种基于深度学习的肝脏CT图像分割方法提升分割准确性。通过学习大规模肝脏CT数据来改进模型,以克服其他深度学习方法中存在的缺点,使其能适应肝脏CT图像复杂的解剖结构,避免模型受到不同患者及医学图像噪声的影响。

1 相关工作

1.1 基于深度学习的图像分割

近年来,深度学习在医学图像分割领域取得了显著进展,LeNet、AlexNet等经典网络在图像分类领域的成功应用引起了广泛关注^[14,15]。研究者们开始探索将CNN应用于医学图像分割任务,例如U-Net设计了一种称为“跳跃连接”的结构,在不同层次捕捉图像的局部、全局信息,为后续模型的发展提供了方向,例如V-Net、Attention U-Net等^[16,17]。

为了更好地捕捉图像的多尺度信息和上下文关系,研究者们提出考虑空间层次结构的网络,引入空洞卷积、注意力机制等技术使深度学习模型更全面地理解医学图像的结构和特征^[18]。例如,谷歌团队提出的DeepLab v2采用了空洞卷积来扩大感受野,提升了上下文感知能力,并引

入多尺度信息融合模块,通过池化操作在不同尺度下提取特征^[19]。Zhao等^[20]提出的PSPNet引入了金字塔池化模块将特征图划分为不同区域,允许网络在不同分辨率下聚合信息。Zhou等^[21]基于U-Net提出改进的U-Net++,通过设计更深的U形网络结构,引入密集的特征融合模块传递底层与顶层特征之间的信息,还嵌套的将编码器的每个层与解码器的多个层连接起来,以提升不同层级的特征表达能力。Yang等^[22]提出的DenseASPP结合了DenseNet与ASPP模块以提取更大感受野,使DenseASPP在处理复杂纹理图像及小物体分割时表现更优秀^[23,24]。Zhang等^[25]提出的ResUNet结合了ResNet的残差连接,减少了深度神经网络中的梯度消失问题,意味着网络可学习更复杂的特征。其中,U型网络结构包含编码器与解码器,通过跳跃连接低级和高级特征,将有助于模型进行信息传递和处理不同尺寸的输入图像,无需预先调整图像大小,增加了模型灵活性。

1.2 注意力机制

在传统图像分割任务中,常规方法主要依赖卷积神经网络等进行特征提取和像素分类。然而,这些方法有时难以捕捉全局与局部信息之间的复杂关系,特别在处理大尺寸图像或含有细小的结构图像时更困难,因此很多学者将注意力机制引入图像分割领域。例如,Vaswani等^[26]提出的自注意力机制在自然语言处理领域中得到了广泛应用,Transformer模型引入的自注意力机制允许模型在处理每个像素时,动态调整图像其他位置的关注程度,对输入序列中的不同位置分配不同权重,从而更好地捕捉长距离依赖关系。Hu等^[13]在SE-Net中引入通道注意力机制,使模型能更集中地关注图像中的不同区域,有助于提升模型对空间关键区域的感知能力。Woo等^[27]提出的CBAM结合了空间与通道注意力机制,允许模型在通道维度选择性地放大对目标有用的特征,以提升图像分割任务性能。谷歌团队提出的DeepLabv3使用空洞卷积与多尺度注意力机制,允许模型在不同尺度上感知图像特征,以提升模型对不同尺度结构的分割性能^[28]。Wang等^[29]提出的非局部神经网络NLNN,通过建模图像中任意两个位置之间的像素关系,以捕捉全局上下文,提升分割性能。

2 本文方法

本文提出一种基于VAN_ResUNet的肝脏分割模型,具体流程如图1所示。首先,对三维CT图像数据进行图像预处理以提高图像质量;其次,将预处理后的图像输入VAN_ResUNet网络中得到预测结果;最后,对损失函数采用深度监督衰减策略,在一定程度上降低数据不平衡造成的影响,以得到最终的预测分割结果^[30]。

2.1 图像预处理

图像预处理在肝脏图像分割中具有重要作用,可使图

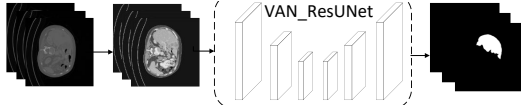


Fig. 1 Process of liver segmentation model based on VAN_ResUNet

图1 基于VAN_ResUNet的肝脏分割模型流程

像在不同条件下具备一致性。为了增强图像对比度并抑制噪声,首先,设定图像灰度阈值为 ± 200 ,将阈值外的灰度值全部截断,大于图像灰度阈值的像素灰度值设置为200,小于图像灰度阈值的像素灰度值设置为-200;其次,为了实现减少计算复杂性加速模型收敛目的,对三维图像进行降采样,具体为对 x, y 轴降采样0.5倍;再次,为了去除不同切片之间的尺度差异,对切片维度 z 轴进行归一化;最后,为了包含更多肝脏区域周围的上下文信息,寻找肝脏区域开始和结束的切片并设定扩张参数向外扩张,在原始图像中保留截取的切片区域以得到预处理后的CT图像。

2.2 网络结构

2.2.1 总体结构

VAN_ResUNet网络结构基于ResUNet分割网络,结合了残差连接和U-Net,有助于更好地传递梯度以减少训练中的梯度消失或爆炸问题。具体为,通过跳跃连接结合深层与浅层特征,从不同角度提取特征,网络模型结构如图2所示。由此可见,网络主要由编码器和解码器组成,编码器首先通过堆叠卷积层池化层,以逐步减小特征图尺寸来提取图像的全局和局部特征,其次将编码器的不同层次特征与上采样后的特征进行融合。

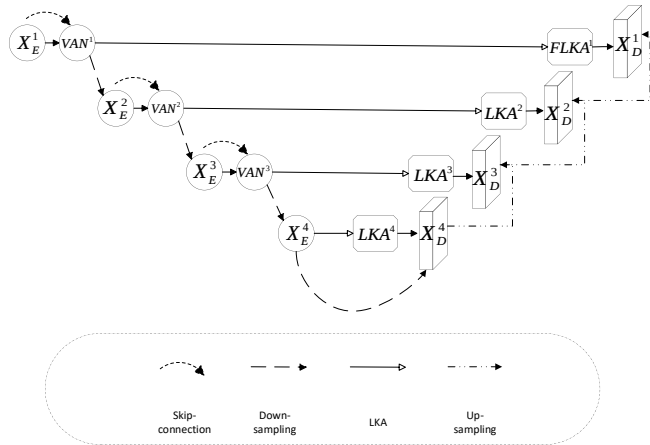


Fig. 2 VAN_ResUNet network structure

图2 VAN_ResUNet网络结构

本文模型在ResUNet基础上加入了视觉注意力网络(Visual Attention Network, VAN)与大核注意力(Large Kernel Attention, LKA),解决了ResUNet无法直接提取与肝脏分割相关特征的问题^[31]。由于编码器在处理长序列时存在困难,本文将视觉注意力网络VAN嵌入在编码器之间以解决长距离依赖性,在编码器提取特征时有助于提升模型对输入数据的表征能力。同时,在编码器与解码器之间加入大核注意力机制LKA及基于注意力特征融合改进的

FLKA模块。其中,LKA基于深度卷积与深度空洞卷积,不但使模型避免了因网络加深,感受野越来越小造成的空间分辨率下降导致的肝脏边界信息丢失问题,还通过逐点卷积加深了网络通道的交互性。基于注意力特征融合改进的FLKA模块采用全局与局部注意力两条支路,通过动态融合跨层特征以更好地捕捉细粒度特征与图像整体信息,并采用 1×1 卷积和自适应池化应对不同输入以增加网络灵活性。

2.2.2 LKA结构

注意力机制能使模型更关注图像的特定区域,既有助于更好地捕捉关键特征,还能灵活适应不同的输入关系,有助于减少模型关注图像中的冗余信息,从而提升模型的计算效率和泛化能力。目前,大多数注意力机制基于卷积堆叠,但卷积操作采用静态权重,无法针对不同区域特征进行动态调整,缺乏适应性。例如,对不同区域进行相同处理会影响肝脏CT图像分割精确度,且肝脏CT图像对比度低、图像信息模糊,因此难以区分肝脏内部与边缘区域、病变与正常组织。此外,应用最广的自注意力机制将二维图像视为一维序列,破坏了图像中的二维结构,并且自注意力机制只考虑了空间维度的适应性而忽略了通道维度的适应性。

为此,本文采用一种新的线性注意力机制大核注意力LKA,吸收了卷积与自注意力机制的优势,包括局部信息提取、长距离依赖性和适应性,内部采用空洞卷积捕获对比度较低的肝脏区域上下文信息。同时,该机制既加大了对对比度较低区域的权重,还考虑了通道维度适应性,可有效关联分布在不同通道的低对比度区域,以提升模型识别能力。LKA注意力的生成可看作为一个自适应选择过程,根据输入图像选择特征并自动忽略低对比度图像伴随的噪声,关键步骤是生成注意力权重,即不同部分的重要性权重,原因为准确得到注意力权重就需要了解不同特征之间的关系,因此本文使用大核卷积来构建相关性并生成注意力以捕获不同特征之间的关系,但大核卷积的计算参数较多、成本很高。

为了解决上述问题,本文将大核卷积分解为深度卷积、深度空洞卷积和逐点卷积,如图3所示。具体为,将一个 $K \times K$ 卷积分解为一个 $K/d \times K/d$ 深度空洞卷积 Conv_{dwd} 、 $(2d-1) \times (2d-1)$ 深度卷积 Conv_{dw} 和 1×1 卷积 $\text{Conv}_{1 \times 1}$,以少量的计算成本和参数来捕获长距离关系。

LKA模块中的计算流程为:

$$\text{att} = \text{Conv}_{1 \times 1}(\text{Conv}_{\text{dwd}}(\text{Conv}_{\text{dw}}(F))) \quad (1)$$

$$\text{output} = \text{att} \otimes F \quad (2)$$

式中: F 为输入特征; att 表示每个特征的重要性。

首先,将特征 F 输入深度卷积,使用分组卷积以保持特征图的大小并减少模型参数量;其次,采用深度空洞卷积提取更广泛范围的特征,在保持分辨率和参数效率的同时避免丢失边缘区域的信息;再次,采用逐点卷积进行特征

通道整合得到注意力权重 att ;最后,将输入特征 F 与注意力权重重点乘得到 LKA 的输出。

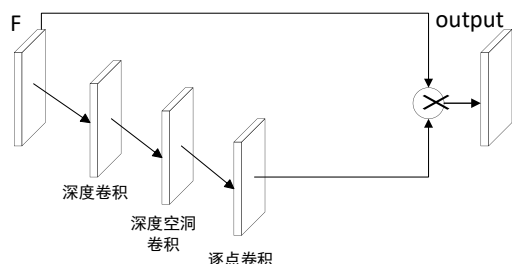


Fig. 3 LKA structure

图3 LKA 结构

LKA 相较于一般的注意力机制而言,无需额外的归一化函数,原因为注意力机制的关键是基于输入特征自适应地调整输出而非归一化注意力图。本文模型融合了 LKA 结合卷积网络与自注意机制的优点,LKA 不仅实现了空间维度的适应性,还实现了通道维度的适应性。

2.2.3 VAN 结构

基于 LKA 的注意力机制在分割肝脏 CT 图像的过程中关注了肝脏不同位置的影响,但未利用不同尺度的肝脏 CT 图像。由于肝脏 CT 低对比度图像常伴随着边界模糊问题,仅利用单一尺度的图像会导致边界分割不精准。

为此,本文在编码器结构内部嵌入视觉注意力网络 VAN,在 LKA 的基础上增强网络在不同层级上的特征提取能力,可帮助模型捕捉到更细微的边缘和结构特征,并在不同尺度捕获边界信息并整合,以准确定位肝脏与周围组织的边界。此外,这种模块嵌入方式使网络更容易调整,且基于大核注意力 LKA 设计的视觉注意力网络 VAN 内部使用了残差连接,将输入输出相加以改善梯度回传、加深网络,提升了模型训练稳定性。

VAN 结构对输入值进行降采样,使用步幅数控制降采样率。首先,在降采样后将一个阶段中所有层保持相同的空间分辨率和通道数;其次,依次堆叠使用批归一化、 1×1 卷积、GELU 激活、LKA 和前馈网络来提取特征。考虑到编码器深层特征与浅层特征的适应性,分别设计了 VAN-V1、VAN-V2、VAN-V3、VAN-V4 等 4 种注意力网络结构,如图 4 所示。由此可见,VAN-V1 结构简单,能灵活嵌入到不同的编码层中,包括 1×1 卷积 pwconv 与一种基于高斯误差的激活函数,该函数相较于 RELU 并未将小于 0 的值全部置 0,随后将激活后的值输入大核注意力 LKA 中的 1×1 卷积 pwconv 中进行通道交互。VAN-V2 在 VAN-V1 基础上增加了多层感知机结构,有利于学习非线性关系的特征,并在每个残差连接后立即使用 LKA 单元以更早地在网络中提升全局感知能力。VAN-V3 对 VAN-V2 的先后顺序进行了调整,输入特征先进入多层感知机提取特征,再输入 LKA 注意力中,在数据到达 LKA 单元和降维卷积层之前,通过增加额外的卷积与激活层使网络深入地分析和处理特征。VAN-V4 在 VAN-V2 的基础上在多层感知机

内部加入大核注意力 LKA,可分别在两个阶段提取关键特征,避免在初始特征阶段的特征整合过程中影响最终的融合权重。

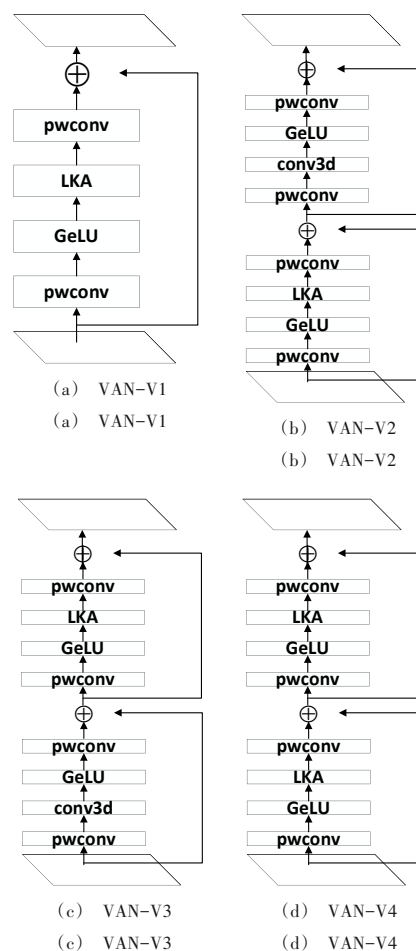


Fig. 4 4 different VAN structures

图4 4种不同的VAN结构

2.2.4 FLKA 结构

大多数特征融合方法通过构建复杂的路径来组合不同层的特征,融合方式通过简单的加法或拼接等线性组合实现。然而,肝脏 CT 图像器官结构多样、不同个体解剖结构的差异使肝脏 CT 图像形态复杂,上述特征融合可能会造成信息冗余,即过度强调不同肝脏 CT 图像的重复信息,忽略了图像之间的差异,模型对不同形态结构的肝脏 CT 图像的适应性较差。例如,LKA 仅关注同一层中的软特征选择,并未跳过连接进行跨层融合,且只有一种形式的注意力权重,适应性较差。

为此,本文借鉴注意力特征融合 (Attention Feature Fusion, AFF) 思想改进 LKA 模块,提出一种大核注意特征融合模块 FLKA,通过上下文规模感知的方式动态、自适应地融合接收到的特征^[32]。具体为,将 FLKA 模块融合全局与局部支路的特征作为输出,局部支路在点卷积中加入大核卷积 LKA 有助于提取肝脏的局部细节特征,例如肝脏器官边界、血管等异质性结构,从而更好地捕捉不同肝脏的

差异。

本文对特征进行长连接映射,改进后的FLKA模块结构如图5(a)所示。首先,给定两个特征图 X 、 Y ,分别代表低层特征映射、网络深层的高级语义特征映射,大核注意特征融合FLKA如式(3)所示;其次,图5(a)中虚线表示 $1 - M(X \oplus Y)$,特征融合的权值 $M(X \oplus Y)$ 和 $1 - M(X \oplus Y)$ 取值范围在0~1之间,这使得网络能在 X 、 Y 之间进行加权平均。

$$Z = M(X \oplus Y) + (1 - M(X \oplus Y)) \otimes Y \quad (3)$$

式中: Z 为融合特征; $\oplus$ 表示初始求和形式的特征集成; M 为多尺度大核注意模块MS-LKA。

多尺度大核注意模块MS-LKA结构如图5(b)所示,MS-LKA核心在于通过改变空间池化大小,在多个尺度上实现通道注意力操作。同时,选择点卷积pwconv利用每个

空间位置的通道交互来聚合上下文,分别设计全局特征提取 $G(x)$ 与局部特征提取 $L(x)$ 两条支路。局部特征提取可表示为:

$$L(x) = pwconv(l(g(pwconv(x)))) \quad (4)$$

式中: $pwconv$ 为3维 1×1 卷积; g 为激活函数GELU; l 为大核注意力LKA。

全局特征提取可表示为:

$$G(x) = pwconv(l(g(pwconv(adp(x))))) \quad (5)$$

式中: adp 为全局平均池化。

通过MS-LKA得到的细化特征 X_0 为:

$$X_0 = X \otimes M(X) = X \otimes \sigma(L(X) \oplus G(X)) \quad (6)$$

式中: $M(X)$ 为MS-LKA产生的注意权重; $\oplus$ 表示加法; $\otimes$ 表示元素级乘法; σ 为激活函数PReLU。

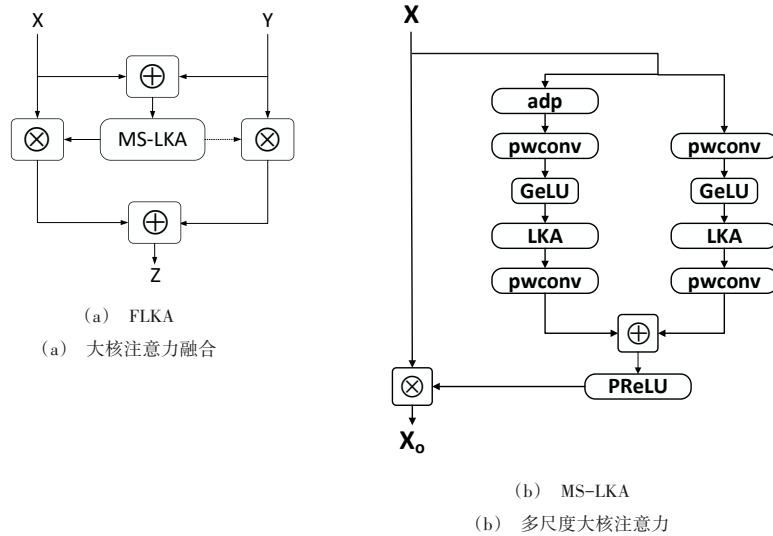


Fig. 5 FLKA structure

图5 FLKA结构

综上所述,本文所提模块克服了传统特征融合方法的局限性,创新地实现了动态跨层整合特征,解决了传统方法中存在的提取信息重复的问题,提升了模型对融合区域的关注程度,特别是处理多变的异质肝脏CT图像。

2.3 损失函数

本文采用的损失函数为Tversky Loss,主要用于处理不平衡类别问题,特别适用于正例像素数量远小于负例像素数量的肝脏CT图像分割领域^[33]。

$$LOSS = \frac{TP}{TP + \alpha FP + \beta FN} \quad (7)$$

式中:TP为真正例;FP为假正例;FN为假负例; α 、 β 为平衡系数,用于调整假正例和假负例的权重。

本文设置 $\alpha=0.3$ 、 $\beta=0.7$,首先通过Tversky Loss计算解码器不同层输出,其次对不同层输出进行加权,最后对权重系数采用深度监督系数衰减策略,在训练不同阶段动态调整监督信号的重要性。具体为,训练初期较大的系数可使网络更关注于学习底层特征;训练后期较小的系数可使

网络更专注于学习高级的抽象特征。

为此,本文使用深度监督系数衰减策略,使其在网络各层次都能起到作用,有助于减少梯度消失问题,使网络更快收敛。具体地,将初始深度监督系数选为0.4,当训练次数大于30时系数以0.8的倍数进行衰减。

3 实验结果与分析

3.1 数据集

本文使用LiTS2017公开数据集进行性能评估,该数据集包含131例患者的腹部CT扫描3D图像,数据存储为.nii格式,每个CT扫描图像都有放射科医生标注的肝脏和肝脏肿瘤位置的分割标签,图像原始分辨率为 512×512 。本文经过预处理后将图像大小处理为 256×256 ,训练、验证、测试集的比例分别为7:2:1。为了防止模型发生过拟合,还对原始图像进行图像翻转、平移等数据增强操作。

3.2 评价指标

为了评估模型效果,分别采用 Dice 系数、平均交并比 (Mean Intersection to Union Ratio, MIoU)、平均精确率 (Mean Accuracy Map, Map) 对模型进行量化。其中, Dice 系数用来衡量两个集合的相似度,通常用于评估图像的分割准确性,即衡量算法预测的分割结果与实际分割结果的重叠程度; MIoU 衡量算法预测的分割结果与实际分割结果的交并比的平均值,即 IoU 的均值; Map 通常用于评估目标检测与图像分割等任务的性能,在图像分割中 Map 既考虑了不同类别的分割精度,还计算了他们的平均值。

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (8)$$

$$MIoU = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (9)$$

$$AUC_i = \sum_{i=1}^N (R_n - R_{n-1}) \times P_n \quad (10)$$

$$Map = \frac{1}{C} \sum_{i=1}^C AUC_i \quad (11)$$

式中: A 为预测标签; B 为真实标签; N 为类别总数; TP_i 表示第 i 个类别的真正例,即预测为该类别且真实标签也为该类别的像素数; FP_i 表示第 i 个类别的假正例,即预测为该类别但真实标签为其他类别的像素数; FN_i 表示第 i 个类别的假负例,即未能正确预测为该类别的像素数; N 为召回率的阈值数; R_n 、 P_n 分别为在第 n 个召回率阈值下的召回率和查准率; C 为类别总数。

3.3 实验结果

为了验证 VAN_ResUNet 分割肝脏 CT 图像的性能,在 同一数据集上与不同模型及不同模块进行比较。首先,针对 4 种 VAN 结构分别进行消融实验来验证 VAN-V1 的效果;其次,为了验证 VAN_ResUNet 网络每个注意力模块的必要性,分别去除不同层的注意力来评估模型评估;再次,为了评估 VAN-V1 卷积核大小对网络性能及计算成本的影响,比较 VAN-V1 结构在 3 种卷积核大小下的实验结果;最后,将本文方法与 U-Net、U-Net++、SegNet、KiU-Net、ResUNet 模型进行比较^[34]。

3.3.1 不同 VAN 结构比较

本文针对 4 种 VAN 结构分别进行消融实验,实验结果如表 1 所示。由此可知, VAN-V1 结构的 Dice 系数为 95.12%, 平均交并比为 97.02%, 平均精确率为 96.10%, 3 种评价指标均优于其余 3 种 VAN 结构。其中, VAN-V1 相较于 VAN-2、VAN-3 结构的 Dice 系数分别提升 3.38%、4.24%; VAN-V4 在 4 种 VAN 结构中的表现最差,性能甚至低于未改进的基础网络,原因为编码器嵌入 VAN-V4 会导致网络过深,造成最底层编码器输出的特征图太小,从而损失部分深层信息。实验表明,最合适嵌入在编码器的 VAN 结构为 VAN-V1。

3.3.2 LKA 嵌入不同编码层对网络模型的影响

为了验证 VAN_ResUNet 结构连接在编码器与解码器

Table 1 Influence of 4 VAN structures on model performance

表 1 4 种 VAN 结构对模型性能的影响

VAN 不同结构	Dice	MIoU	Map
VAN_V2	0.917 4	0.939 5	0.927 4
VAN_V3	0.908 8	0.935 8	0.919 5
VAN_V4	0.889 0	0.919 6	0.897 2
VAN_ResUNet	0.951 2	0.970 2	0.961 0

之间每个注意力机制的有效性,分别去除不同层的注意力进行消融实验,结果如表 2 所示。由此可知,去除编码器 1 与解码器 1 之间的 FLKA1 模块后, Dice 系数下降 1.24%; 去除编码器 2 与解码器 2 之间的 LKA2 模块后, Dice 系数下降 2.13%; 去除编码器 3 与解码器 3 之间的 LKA3 模块后, Dice 系数下降 2.29%; 去除编码器 4 与解码器 4 之间的 LKA4 模块后, Dice 系数下降 2.55%。实验表明,去除每个模块都会导致网络性能下降,因此 VAN_ResUNet 的每个注意力模块对模型均具有积极作用。

Table 2 Impact of different layers of attention modules on model performance

表 2 不同层注意力模块对模型性能的影响

注意力层	Dice	MIoU	Map
不融入 FLKA1	0.938 8	0.958 2	0.947 3
不融入 LKA2	0.929 9	0.949 7	0.939 3
不融入 LKA3	0.928 3	0.957 6	0.938 0
不融入 LKA4	0.925 7	0.952 3	0.935 1
VAN_ResUNet	0.951 2	0.970 2	0.961 0

3.3.3 不同卷积核大小的 LKA 比较

卷积核大小是深度学习中一个重要的超参数,直接影响着模型的感受野和特征提取能力。小卷积核能在保持计算效率的同时增加感受野,例如多个 3×3 卷积核的组合相当于一个大卷积核。此外,多个小卷积核的非线性层叠加可增加网络表示能力;大卷积核具有更大的感受野,既更容易捕捉全局特征和更多上下文信息,还能降低网络深度,减轻梯度消失问题、简化模型训练过程。

为此,本文对 LKA 结构采用不同的卷积核大小进行评估,分别计算了在不同卷积核大小下模型的 Dice 系数及单层 LKA 模块的参数成本,结果如表 3 所示。由此可知,当卷积核大小为 7 时 Dice 指标最优,但考虑到不同编码层的特征图大小不一,本文在 VAN_ResUNet 中采用的 LKA 结构的卷积核大小也有所不同。多次实验发现, LKA3 与 LKA4 处于网络深层,卷积核为 7×7 更适合; LKA2 与 FLKA1 输入的特征图较大,卷积核为 14×14 更适合。

Table 3 Influence of different LKA convolutional kernel sizes on models

表 3 LKA 不同卷积核大小对模型的影响

方法	K	膨胀率	参数/M	GFLOPs	Dice
LKA	5	2	0.20	0.34	0.925 0
LKA	7	3	0.50	2.60	0.937 5
LKA	14	2	1.71	3.40	0.926 3

3.3.4 不同模型比较

为了验证 VAN_ResUNet 网络的性能, 将其与 5 种现有网络进行比较, 并令各模型的编码层数量、解码层数量及其他训练参数与本文模型保持一致, 结果如表 4 所示。由此可知, VAN_ResUNet 在 3 个指标上均优于其他网络, 相较于基础网络 ResUNet 的 Dice 系数、平均交并比、平均精确率分别提升 3.27%、1.32%、3.75%, 相较于 KiU-Net 模型在 3 个指标上分别提升 3.75%、0.82%、3.04%, 相较于 Attention U-Net 模型在 3 个指标上分别提升 4.92%、2.18%、5.35%, 相较于 SegNet 在 3 个指标上分别提升 7.67%、3.65%、7.79%, 相较于 U-Net 在 3 个指标上分别提升 6.23%、2.88%、6.28%。

为了更好地呈现各方法的实际分割效果, 取一实例进行可视化展示, 如图 6 所示。由此可见, U-Net、SegNet 方

Table 4 Performance comparison of various algorithms

表 4 各算法的性能比较

模型	Dice	MIoU	Map
U-Net ^[9]	0.888 9	0.941 4	0.898 2
SegNet ^[10]	0.874 5	0.933 7	0.883 1
Attention U-Net ^[17]	0.902 0	0.948 4	0.907 5
KiU-Net ^[34]	0.913 7	0.962 0	0.930 6
ResUNet ^[25]	0.918 5	0.957 0	0.923 5
VAN_ResUNet	0.951 2	0.970 2	0.961 0

法边界丢失现象严重, 原因可能为模型并未融合深层与浅层的关键特征。本文所提网络使用跨层注意力特征融合方法避免了边界丢失, 还使分割结果更接近真实图像, 既未在肝脏 CT 图像边界少分割, 也没有在其余部分多分割。

总而言之, 本文算法在实际数据集上的分割精度更高, 解决边界问题更出众, 可视化结果进一步证明了本文算法在特征融合中引入大核注意力机制的有效性。

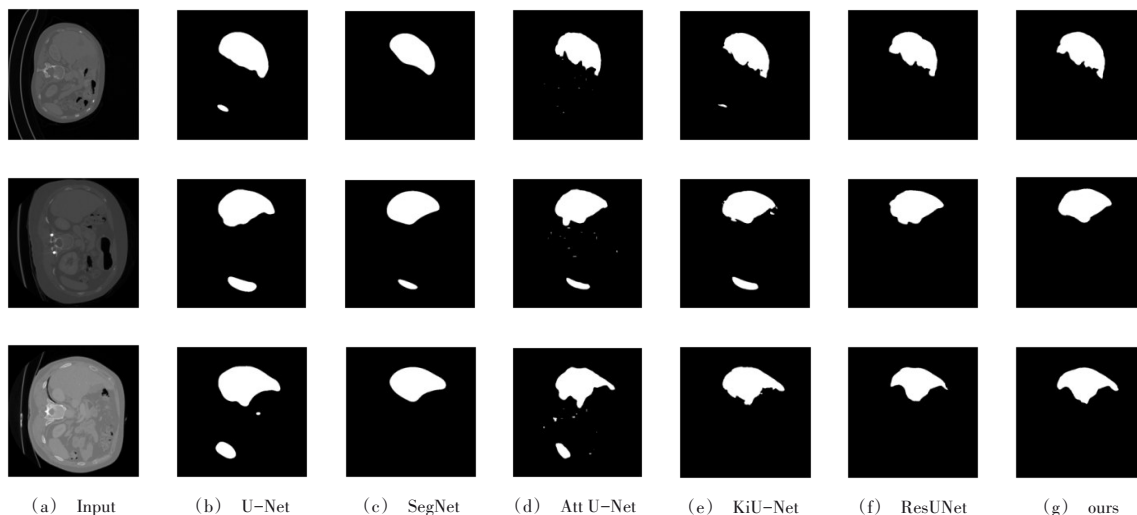


Fig. 6 Comparison of visual segmentation results

图 6 可视化分割结果比较

4 结语

本文提出一种高效的肝脏 CT 分割网络, 创新性地将视觉注意力网络 VAN 嵌入编码器, 使其在提取特征时更关注目标区域。首先, 为了使模型更适应肝脏 CT 数据, 提出 4 种 VAN 结构进行比较分析; 其次, 在编码器与解码器之间加入大核注意力机制 LKA, 在保证特征图大小的同时加强了特征之间的通道交互性; 再次, 基于注意力特征融合 AFF 的思想, 改进了最外层 LKA 注意力模块; 最后, 提出 FLKA 合并局部特征提取和全局特征提取两条支路的输出, 使网络能跨层融合不同的语义特征。

实验表明, 基于 VAN_ResUNet 肝脏 CT 图像分割网络在 Dice 系数、平均交并比、平均精确率指标上优于比较网络。然而, 本文方法仍存在不足之处: 只针对最外层编码器、解码器加入了改进的 FLKA, 原因在于 FLKA 结构很深, 如果每层都堆叠 FLKA 会丢失特征融合的部分信息。

未来, 将考虑进行一步对模型结构进行改良, 在无需过深的结构下仍能实现跨层特征交互。

参考文献:

- [1] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [2] ADAMS R, BISHOF L. Seeded region growing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641-647.
- [3] CANNY J. A Computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.
- [4] HARALICK R M, SHAPIRO L G. Image segmentation techniques [C]// 2017 Asia Modelling Symposium, 2017: 103-107.
- [5] CORINNA C, VLADIMIR V. Machine learning [M]. Berlin: Springer, 1995.
- [6] MACQUEEN J. Some methods for classification and analysis of multivariate observations [C]// Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967: 282-297.
- [7] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms

- [M]. Berlin: Springer, 1981.
- [8] LECUN Y, BOSER B, DENKER J, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Computation*, 1989, 1(4): 541–551.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]// *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015: 234–241.
- [10] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(12): 2481–2495.
- [11] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848.
- [12] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]// *2019 IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 3141–3149.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [14] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2): 84–90.
- [16] MILLETARI F, NAVAB N, AHMADI S A S. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]// *2016 Fourth International Conference on 3D Vision*, 2016: 565–571.
- [17] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: learning where to look for the pancreas[DB/OL]. <https://arxiv.org/abs/1804.03999>.
- [18] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[DB/OL]. <https://arxiv.org/abs/1511.07122>.
- [19] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848.
- [20] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]// *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6230–6239.
- [21] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: a nested U-Net architecture for medical image segmentation[DB/OL]. <https://arxiv.org/abs/1807.10165>.
- [22] YANG M, YU K, ZHANG C, et al. DenseASPP for semantic segmentation in street scenes[C]// *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 3684–3692.
- [23] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]// *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261–2269.
- [24] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[DB/OL]. <https://arxiv.org/abs/1706.05587>.
- [25] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-Net[DB/OL]. <https://arxiv.org/abs/1711.10684>.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL]. <https://arxiv.org/abs/1706.03762>.
- [27] WOO S, PARK J, LEE J, et al. CBAM: convolutional block attention module[DB/OL]. <https://arxiv.org/abs/1807.06521>.
- [28] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848.
- [29] WANG X, GIRSHIC R, GUPTA A, et al. Non-local neural networks[C]// *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7794–7803.
- [30] LEE C Y, XIE S, GALLAGHER P, et al. Deeply-supervised nets[DB/OL]. <https://arxiv.org/abs/1409.5185>.
- [31] GUO M, LU C, LIU Z, et al. Visual attention network[J]. *Computational Visual Media*, 2023, 9(4): 733–752.
- [32] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]// *2021 IEEE Winter Conference on Applications of Computer Vision*, 2021: 3559–3568.
- [33] SALEHI S S M, ERDOGMUS D, GHOLIPOUR A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[DB/OL]. <https://arxiv.org/abs/1706.05721>.
- [34] JOSE J M, SINDAGI V, HACIHALILOGLU I, et al. KiU-Net: towards accurate segmentation of biomedical images using over-complete representations[DB/OL]. <https://arxiv.org/abs/2006.04878>.
- [35] HUANG Y L, HU K, GUO J Q, et al. Liver CT image segmentation based on multi-scale feature fusion and dual attention mechanism[J]. *Computer Science*, 2022, 49(S2): 549–557.
- 黄扬林, 胡凯, 郭建强, 等. 基于多尺度特征融合和双重注意力机制的肝脏CT图像分割[J]. *计算机科学*, 2022, 49(S2): 549–557.

(责任编辑:刘嘉文)