

基于SegFormer的超声影像图像分割

杨靖怡, 李芳, 康晓东, 王笑天, 刘汉卿, 韩俊玲

引用本文

杨靖怡, 李芳, 康晓东, 王笑天, 刘汉卿, 韩俊玲. [基于SegFormer的超声影像图像分割](#) [J]. 计算机科学, 2023, 50(6A): 220400273-6.

YANG Jingyi, LI Fang, KANG Xiaodong, WANG Xiaotian, LIU Hanqing, HAN Junling. [Ultrasonic Image Segmentation Based on SegFormer](#) [J]. Computer Science, 2023, 50(6A): 220400273-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[CT影像阶段化目标检测方法研究](#)

Study on Phased Target Detection in CT Image

计算机科学, 2023, 50(6A): 220200063-10. <https://doi.org/10.11896/jsjcx.220200063>

[基于CT图像语义的COVID-19实例分割与分类网络](#)

COVID-19 Instance Segmentation and Classification Network Based on CT Image Semantics

计算机科学, 2023, 50(6A): 220600142-9. <https://doi.org/10.11896/jsjcx.220600142>

[基于球簇聚类的超像素分割迭代算法](#)

Superpixel Segmentation Iterative Algorithm Based on Ball-k-means Clustering

计算机科学, 2023, 50(6A): 220600114-7. <https://doi.org/10.11896/jsjcx.220600114>

[基于主动学习和U-Net++分割的芯片封装空洞率的研究](#)

Study on BGA Packaging Void Rate Detection Based on Active Learning and U-Net++ Segmentation

计算机科学, 2023, 50(6A): 220200092-6. <https://doi.org/10.11896/jsjcx.220200092>

[综合应用Faster R-CNN和U-net的心脏MRI图像分割](#)

Cardiac MRI Image Segmentation Based on Faster R-CNN and U-net

计算机科学, 2023, 50(6A): 220600047-9. <https://doi.org/10.11896/jsjcx.220600047>

基于 SegFormer 的超声影像图像分割

杨靖怡¹ 李芳^{1,2} 康晓东¹ 王笑天¹ 刘汉卿¹ 韩俊玲¹

¹ 天津医科大学医学影像学院 天津 300202

² 北京市化工职业病防治院 北京 100093

(yangjy9912@163.com)

摘要 超声影像分割既是医学影像图像处理的重要环节,也是临床诊断的常用技术手段。文中提出将 SegFormer 网络模型用于实现医学超声影像图像的精准分割。一方面,将超声标签图转化为单通道形式,并对其进行二值化处理,以完成对数据集图像的预处理;另一方面,采用迁移学习的方式载入预训练模型,用于微调已经训练好的模型参数,并选用带有动量的随机梯度下降优化器来加速收敛速度及减小震荡。与 FCN,UNet 和 DeepLabV3 的对比实验结果表明,该模型在乳腺结节超声影像数据集上的各项评估指标均为最优,mIoU,Acc,DSC 和 Kappa 分别为 81.32%,96.22%,88.91% 和 77.85%。实验结果还表明,该模型在不同超声影像数据集中表现出了良好的鲁棒性。

关键词: SegFormer; 图像分割; 超声影像; Transformer

中图分类号 TP391

Ultrasonic Image Segmentation Based on SegFormer

YANG Jingyi¹, LI Fang^{1,2}, KANG Xiaodong¹, WANG Xiaotian¹, LIU Hanqing¹ and HAN Junling¹

¹ School of Medical Image, Tianjin Medical University, Tianjin 300202, China

² Beijing Chemical Occupational Disease Control Hospital, Beijing 100093, China

Abstract Ultrasonic image segmentation is not only an important part of medical image processing, but also a common technical means of clinical diagnosis. In this paper, the SegFormer network model is proposed to realize the accurate segmentation of medical ultrasound images. On the one hand, the ultrasonic label image is transformed into a single channel and processed by binarization to complete the preprocessing of the data set image; on the other hand, the pre-training model is loaded into the pre-training model to fine-tune the trained model parameters, and a random gradient descent optimizer with momentum is selected to accelerate the convergence speed and reduce the oscillation. Experimental results show that, compared with FCN, UNet and DeepLabV3, all the evaluation indexes of the proposed model are the best in the breast nodule ultrasound image data set, and the evaluation indexes of mIoU, Acc, DSC and Kappa is 81.32%, 96.22%, 88.91% and 77.85% respectively. The experimental results also show that the model is robust in different ultrasonic image data sets.

Keywords SegFormer, Image segmentation, Ultrasonic image, Transformer

1 引言

超声影像具有无损伤、灵敏度高、成像迅速、成本低廉等特点,目前已被广泛应用于临床检测中。然而,超声影像常存在噪声斑点、信号衰减等问题的干扰,往往会损害图像质量,导致图像对比度降低。同时,在临床上对于超声影像的评估与诊断往往高度依赖于临床诊断人员的主观判断,具有个体化差异等问题。应用计算机辅助检查技术完成超声影像图像分割,可将超声影像中具有特殊含义的部分分割出来,并提取相关特征,为临床医生在诊疗的过程中提供可靠的诊断依据,辅助临床医生做出更准确、更客观的诊断。因此,将更精准、更有效的图像分割技术应用于超声影像中尤为重要^[1]。

目前,超声影像图像分割方式大致可分为两种,分别是基于传统的超声影像分割方法与基于深度学习的超声影像分割

方法。对于传统的图像分割方法,从整体上可分为基于边缘检测、基于区域及将两者相结合 3 种类型。然而,超声影像具有噪声大、区域模糊、边界信息弱、感兴趣特征区域难以定位等问题,导致以上分割方法无法精准定位分割区域^[2]。近年来,深度学习已经成为最主要的机器学习工具,并且被广泛应用于各研究领域,尤其是图像分析处理与计算机视觉^[3]。基于深度学习的超声影像的图像分割也表现出了巨大的潜力,它可直接处理超声影像,并可自动学习从原始的超声影像中获取的抽象特征。

基于卷积神经网络(Convolution Neural Network, CNN)是目前图像分割领域应用得最广的一种深度学习方法。Girshick 等^[4]将 R-CNN 应用于淋巴结的超声影像分割中,由于此方法需要预先提取特征信息,因此计算量巨大,效率较低。自 2015 年,Long 等^[5]提出了端到端、像素到像素的全卷积

基金项目:京津冀协同创新项目(17YEXTZC00020)

This work was supported by the Beijing-Tianjin-Hebei Collaborative Innovation Project(17YEXTZC00020).

通信作者:康晓东(kxd2004@126.com)

神经网络(Fully Convolutional Networks, FCN),可输入随机大小的图像,实现语义级别的图像分割任务。Yap 等^[6]用乳腺肿瘤超声影像数据集训练 FCN,由于恶性乳腺肿瘤尺寸多变并且训练图像数量较少,导致分割效果欠缺。Ronneberger 等^[7]提出了基于 UNet 网络模型的图像分割并将其应用于生物医学图像分割,UNet 网络拥有对称的网络结构和跳跃连接,适用于医学图像分割,解决了样本不足的问题,但对于阴影较多、边界模糊的超声影像来说,其分割效果并不理想。之后,不少学者将 UNet 网络进行改进,将其更好地应用于超声影像中,Zhuang 等^[8]在 UNet 的解码器中引入空洞卷积以增加感受野,但其在多尺度特征提取方面尚有不足。为了解决这个问题,Ibtehaz 等^[9]设计了 Multires block 加入 UNet 网络,用于提取多尺度信息,并且同时可使网络轻量化。He 等^[10]设计了深度注意力 UNet 网络,实现了在超声影像中对甲状腺轮廓实行自动分割的效果。Wang 等^[11]在 UNet 中融入了空间注意力、通道注意力、多尺度特征模块的 IU-Net 模型,解决了超声影像下甲状腺尺寸多变的问题。Carneiro 等^[12]将 DNN 应用于超声心动图中左心室的分割,其能从原始超声影像中学习高级特征。Anas 等^[13]基于残差网(ResNet)进行深度框架的优化,使用指数加权图提高局部预测。Milleari 等^[14]设计霍夫-CNN,用于多个深部脑区域的监测与分割。

基于上述内容,本文提出将 SegFormer 网络模型应用于对医学超声影像的图像分割,为验证此网络模型在医学图像分割领域,尤其是超声影像分割领域中的优异表现及鲁棒性。一方面,将超声标签图转化为单通道形式,并将其二值化处理后再转换为像素值为 0 和 1 的连续标签,以完成对数据集图像的预处理;另一方面,采用迁移学习的方式载入预训练模型,微调已经训练好的模型参数,并选用带有动量的随机梯度下降优化器来加速收敛速度及减小震荡。本文共设计了两个实验:首先,搭建优化的 SegFormer 网络模型,并将其与其他 3 种常见分割网络模型进行对比研究,包括基于全卷积神经网络的 FCN 模型和 Unet 模型以及基于残差网络的 DeepLabV3^[15]模型,验证了 SegFormer 网络模型结构的可靠性及准确性。同时,为验证此模型适用于各种超声影像分割并为临床分割工作提供了参考意义及临床实用价值,将多个医学超声影像数据集,包括甲状腺结节超声影像数据集、颈部臂丛神经影像结节超声数据集和血管瘤超声影像数据集这 3 个数据集分别输入此网络进行训练,并对分割训练结果进行研究评估。实验结果表明,对于乳腺结节超声数据集,本文方法在平均交并比(mean Intersectin over Union, mIoU)、准确率(Accuracy, Acc)、Dice 相似度系数(Dice Similarity Coefficient, DSC)及 Kappa 系数 4 个指标上均高于其他 3 种网络模型。并且,对于不同的超声影像数据集,本文方法表现出了较好的鲁棒性,可为临床超声影像分割工作提供参考意义及使用价值。

2 SegFormer 框架的特点

多层次的 Transformer 是 SegFormer 网络模型中最主要的组成部分,是 SegFormer 在简单结构下仍能保持较高分割精度的关键。

2.1 Transformer

目前,Transformer 在自然语言处理、图像处理等领域

发挥着重要的作用。多头自我注意力(self-attention)机制是其关键特征,Transformer 的结构图如图 1 所示。自我注意力机制可理解为是图形化的归纳偏置,它将序列中的所有标记与基于相关性的池化操作连接起来。高效自我注意力模型是长序列模型应用时的关键。由于需要被处理的文字、图像、视频等常都由大量像素或标记组成,因此处理长序列的效率对于 Transformer 的广泛使用至关重要。多头自我注意力机制进一步完善了自我注意力层,通过使用多头自我注意力机制,可使序列中的每个元素学习从序列中的其他标记收集信息,扩展了此模型专注于不同位置的能力,并且其给出了注意力层的多个表示子空间(Representation Subspaces),从而提高了注意力层的性能。Vision Transformer(ViT)^[16]是首先将 Transformer 引入语义分割的一批算法,ViT 可以将每个图像视为一系列标记,然后将它们馈送到多个 TransFormer 层进行分类。DeiT^[17]进一步探索了一种数据使用高效化的训练策略和 Vit 的蒸馏方法。PVT^[18]首次在 TransFormer 中引入金字塔结构的作品,展示了在密集预测任务中单一 TransFormer 主干卷积神经网络(Convolutional Neural Networks, CNN)的潜力。之后的 TransFormer 加强了特征的局部连续性,并去除了固定大小的位置嵌入,以提高 TransFormer 在密集预测任务中的性能^[19]。

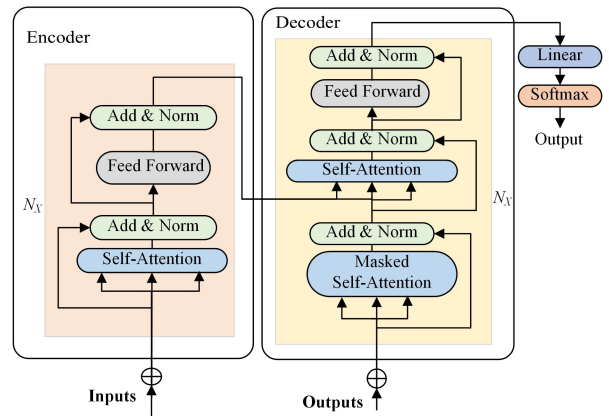


图 1 Transformer 的结构

Fig. 1 Architecture of Transformer

2.2 SegFormer

SegFormer 将层次化 Transformer 编码器与轻量级全多层感知(All-MLP)解码器相结合,其被应用于多个经典语义分割数据集中,与其他模型相比,在精度与速度方面均有优异的表现,是一种结构简单、高效且鲁棒性极佳的网络^[20]。SegFormer 模型的结构如图 2 所示。SegFormer 由分级 Transformer 编码器和轻量级多层感知机解码器两部分组成。

2.2.1 分级 Transformer 编码器

SegFormer 采用了 Mix Transformer(MiT)编码器,在给定的输入图像的情况下,生成类似 CNN 的多级特征(Multi-level Features),可以同时提供高分辨率的粗特征(Coarse Features)和低分辨率的细特征(Fine-grained Features),使得像素分类更准确,边缘分割更精细。

SegFormer 的编码器由 3 个模块组成,分别是高效自注意力模块、图像块重叠融合模块和 Mix-FFN 结构。

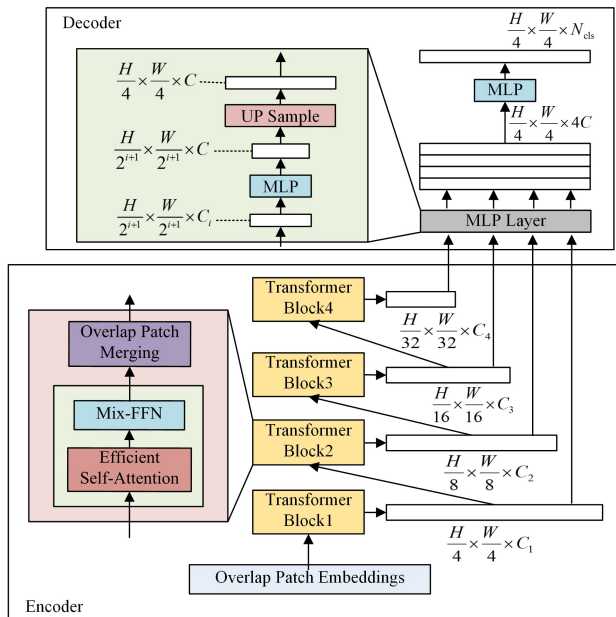


图 2 SegFormer 的结构

Fig. 2 Architecture of SegFormer

(1) 高效自注意力模块。在最初的多头自注意力(multi-head self-attention)过程中,每个 Q, K, V 都具有相同的维度 $N \times C$, 其中 $N = H \times W$ 是序列的长度。多头自注意力的计算式如式(1)所示:

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{\text{head}}}}\right)V \quad (1)$$

此过程的复杂程度高,不宜应用于分辨率较大的图像,因此 SegFormer 对此进行了改进,将 Q, K, V 的维度设置为 $N \times C, N/R \times C, N/R \times C$ 。其过程如式(2)~式(3)所示:

$$\hat{K} = \text{Reshape}\left(\frac{N}{R}, C \cdot R\right)(K) \quad (2)$$

$$K = \text{Linear}(C \cdot R, C)(\hat{K}) \quad (3)$$

(2) 像素块(Patch)重叠融合模块。最初的像素块融合过程将非重叠的像素块结合,导致像素块之前无法保留局部连续性,SegFormer 设置了超参数 S , 将切割后的带有重叠部分的像素块按照一定步长合并,从而保证了像素块周围的连续性。

(3) Mix-FFN 结构。传统的 TransFormer 使用固定的位置编码(Positional Encoding, PE)来引入位置信息,当测试分辨率和训练分辨率不同时,需要对嵌入的位置代码进行插值或做固定分辨率的滑动窗口测试,导致精度下降或计算效率降低等问题。SegFormer 中引入了 Mix-FFN, 避免训练过程中 PE 带来的影响,且避免以零填充导致缺失位置泄露。Mix-FFN 通过在前馈网络(FFN)中直接使用 3×3 Conv 来传递位置信息。Mix-FFN 的计算式如式(4)所示:

$$x_{\text{out}} = \text{MLP}(\text{GELU}(\text{Conv}_{3 \times 3}(\text{MLP}(x_{\text{in}})))) + x_{\text{in}} \quad (4)$$

其中, x_{in} 是来自自注意模块的特征。在 Mix-FFN 中,每一个 FFN 中都融合了一个 3×3 的卷积和一个 MLP。由于运用了深度卷积,因此在此模型训练过程中计算效率得到大大提升。

2.2.2 轻量级多层感知机解码器

SegFormer 集成了一个仅由 MLP 层组成的轻量级解码器,避免了手动操作或应用大量计算组件。构成此简易解码器的关键就是其分层 TransFormer 编码器。相比传统的

CNN 编码器,分层 TransFormer 编码器具有更大的有效接收场(Effective Receptive Field, ERF),可用于融合多级特征,生成最终的语义分割掩码(Mask)。轻量级多层感知机解码器由 4 个主要部分组成,如式(5)所示。首先,编码器的多级功能 F_i 会将不同层的特征通过一个线性层来统一通道维度。其次,将特征上采样至 $1/4$ 分辨率,并拼接至一起,再采用一个线性层为级联特征 F 进行融合。最后,另一个线性层使用融合的特征预测具有 $H/4 \times W/4 \times N_{\text{cls}}$ 分辨率的分割掩模。在解码器中只包含 4 个部分,共应用 6 个线性层,避免了其他复杂操作,使其计算量与参数量非常小,因此保证了此算法高效的特性。

$$\begin{aligned} \hat{F}_i &= \text{Linear}(C_i, C)(F_i), \forall i \\ \hat{F}_i &= \text{Upsample}\left(\frac{W}{4} \times \frac{W}{4}\right)(\hat{F}_i), \forall i \end{aligned} \quad (5)$$

$$F = \text{Linear}(4C, C)(\text{Concat}(\hat{F}_i)), \forall i$$

$$M = \text{Linear}(C, N_{\text{cls}})(F)$$

其中, M 表示预测的掩码, $\text{Linear}(C_i, C)$ 表示分别以 C_i 和 C 作为输入和输出向量维度的线性层。

3 本文方法

本文尝试利用 SegFormer 实现超声影像图像分割。本文方法包括数据集整理、数据仿真和损失函数优化 3 个步骤。

首先,本文对超声数据集中的图像进行预处理操作,将标签图转化为单通道形式,再将标签图以背景部分及待分割特征区域部分划分为两类,并将其进行二值化处理后转换成像素值为 0 和 1 的连续标签。完成对数据集图像的预处理后,将此数据集以 8:2 的比例划分为训练集与验证集。

其次,通过两个相关实验设计验证基于 SegFormer 超声影像图像分割的有效性。在实验 1 中,将 SegFormer 网络模型与基于全卷积网络的 FCN 模型和 UNet 模型以及基于残差网络的 DeepLabV3 模型同时应用于乳腺结节超声影像这一特定数据集中,通过实验结果进行对比分析,验证 SegFormer 网络模型在医学超声影像分割处理中的有效性。在实验 2 中,进一步将 SegFormer 网络模型应用于多个医学超声影像数据集,利用评估指标对实验结果进行定量与定性的分析,以展现此网络模型良好的鲁棒性,证明将此网络模型应用于临床超声影像分割工作中的参考意义及临床使用价值。

最后,采用迁移学习的方式载入预训练模型,微调已经训练好的模型参数以获得更好的分割结果。同时,在网络模型中采取数据增强策略,以此扩大训练样本数量,数据增强内容包括对原图像进行随机填充剪裁、随机水平翻转、随机垂直翻转及图像归一化操作,使图像的数量相当于增加至原来的 4 倍。在此模型中,选用带有动量的随机梯度下降(Stochastic Gradient Descent with Momentum, SGDM)优化器来加速收敛速度及减小震荡,并对单次输入网络的样本数量与学习率进行优化调整,在减少迭代次数的同时增加损失函数的收敛程度与梯度下降方向的准确性。

在实验设计中,首先分别对 3 个数据集中的图像进行格式转换,将其转换为 PNG 无损压缩格式,标签图用与实验 1 相同的方式进行预处理,其次利用 3 个数据集中的训练集图像训练已构建好的 SegFormer 网络模型,并用验证集得出评估网络训练的结果。

4 实验设计与数据集仿真

为验证此模型的优越性与鲁棒性,本文共设计了两个实验。实验1是将优化的 SegFormer 网络模型与 FCN 网络模型、UNet 网络模型和 DeepLabV3 网络模型进行对比研究,应用同一乳腺结节超声数据集训练以上4种网络模型,对训练结果进行验证并做出定量与定性分析,以验证本文方法的有效性 & 可靠性。实验2是将 SegFormer 网络模型应用于不同数据集中进行分割训练。

4.1 数据集与实验环境

本文共使用4个公开数据集进行实验。4个公开数据集分别是:乳腺结节超声影像数据集、甲状腺结节超声影像数据集、颈部臂丛神经超声影像数据集和血管瘤超声影像数据集。在数据集中,每张图像均有各自对应的标签图。其中,甲状腺超声影像数据集来自百度 Ai Studio 平台,其余3者均来自 Kaggle 网络平台。

乳腺结节超声数据集收集于2018年,共由780张图像组成,来自于600名年龄在25~75岁之间的女性患者。图像尺寸为500×500像素,图像为PNG。数据集被分为3类,分别是良性结节、恶性结节及正常样本。本文只使用该数据集中的良性结节与恶性结节数据共645张图像做超声影像分割研究。甲状腺结节超声影像数据集上传于2020年,图像为BMP,选取其中250张图像进行超声影像分割研究。颈部臂丛神经超声影像数据集来源于2016年 Kaggle 特征预测大赛(Featured Prediction Competition),数据集中的图像数据均由接受过专业培训的人员标注,本文选取数据集中含有颈部臂丛神经影像的图像进行研究,清洗掉其中不含神经的颈部超声影像,共得到2323张图像,图像为580×420像素,图像格式为TIF。血管瘤超声影像数据集上传于2021年,共214张图像,图像格式为PNG。以上数据集均按8:2的比例划分为训练集和验证集。

本实验以 PaddlePaddle 框架为实验环境,操作系统为 Ubuntu16.04, GPU 的版本为 Tesla V100,内存与显卡均为32GB。采用SGDM优化器,epoch的训练轮数为50,初始学习率设置为0.01,每次输入网络模型的样本数量为4。

4.2 实验评价指标

基于对医学超声影像的分割,本文中评估模型质量主要通过4个指标进行判断,它们分别是 mIoU, Acc, DSC 和 Kappa 系数。

当分别对每个类别的数据集单独进行推理计算时,计算出预测区域和实际区域交集与预测区域和实际区域的并集,取交集除以并集的结果为交并比。如果将所有类别得到的交并比的结果取平均,即得到 mIoU, mIoU 是语义分割中一项重要的评价指标。Acc 指图像中类别预测正确的像素占图像中总像素和的比例,准确率越高,模型的质量就越好。mIoU 和 Acc 的计算式如下:

$$mIoU = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \quad (6)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

其中, TP 代表实际为阳性样本且被预测为阳性样本的特征的数量, TN 代表实际为阴性样本且被预测为阴性样本的特征的数量, FP 代表预测为阳性样本但实际是阴性样本

的特征的数量, FN 代表预测为阴性样本但实际是阳性样本的特征的数量。

对于分割过程中的评价标准主要采用 Dice 相似系数,它是一种集合相似度度量函数,通常用于计算两个样本的相似度,其取值范围为[0,1], Dice 系数越接近1,代表模型质量越高。其计算式如下:

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \quad (8)$$

其中, X 代表真实值,即图像中需要被分割的区域, Y 代表预测值,即图像经过深度学习网络被预测的分割区域。

Kappa 系数用于检验一致性的指标,可以用于衡量分类的效果。Kappa 系数的计算主要基于混淆矩阵,取值范围为-1到1之间,通常大于0。Kappa 系数越高,代表模型质量越好。其具体表达式如下:

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (9)$$

其中, P_o 为分类器的准确率, P_e 为随机分类器的准确率。

4.3 实验过程及结果

实验1将 SegFormer 与常见的医学图像分割网络模型 FCN, UNet, DeepLabV3 进行对比研究,并基于医学图像分割评价标准得出对比实验结果,如表1所列。

表1 不同网络对比实验的结果

Table 1 Comparative results of different algorithms

(单位:%)

Algorithm	mIoU	Accuracy	DSC	Kappa
FCN	77.40	95.22	86.16	72.32
UNet	71.11	93.88	81.08	62.26
DeepLabV3	79.21	96.12	87.42	74.85
SegFormer	81.23	96.22	88.91	77.85

注:表中加粗数据为最优结果

将 SegFormer 网络模型应用于乳腺结节超声影像,使 mIoU 达到 81.23%, Dice 系数可达到 88.91%, 基于像素分类的准确率可达 96.22%, 衡量分类效果的 Kappa 系数可达 77.85%。SegFormer 网络模型在本文所涉及的4项评估指标上均优于其他网络,表现优异。

训练过程中验证集上的准确率与 mIoU 上升曲线的对比结果如图3所示, DeepLabV3、FCN 网络模型在验证集上的 mIoU 均在10轮后趋近于饱和, SegFormer 网络模型在验证集上的 mIoU 一直处于上升的趋势,说明其具有很强大的特征学习能力。UNet 网络模型在验证集上的 mIoU 虽然也处于上升趋势,但其最优 mIoU 值落后于 SegFormer 网络模型最优 mIoU 值近20%。

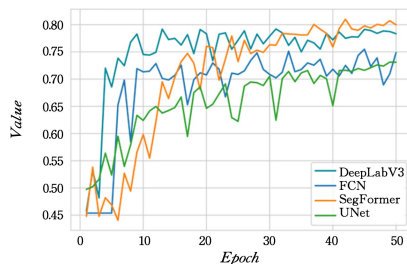


图3 不同网络的 mIoU 比较

Fig. 3 mIoU comparison of different algorithms

图4为各模型实例分割的效果图,选取验证集中乳腺结节影像共6张,其中良性结节影像与恶性结节影像各3张。

在前 3 张超声影像中,由于良性乳腺结节的区域较小并且边缘规整、轮廓清晰,而且结节部位与背景区域之间有明显像素值区别,4 个网络均能较好、完整地分割出乳腺结节部位,各网络分割效果无明显区别。对于后 3 张实例分割图中的恶性结节,各网络分割效果差异明显,由于恶性乳腺结节体积大、边缘不整齐、结节内部灰度像素值不均匀等原因,导致恶性结节的分割效果不如良性结节。

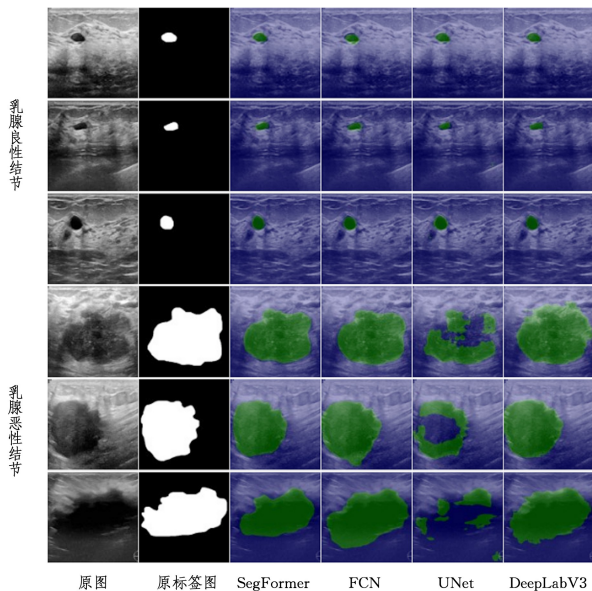


图 4 不同网络下乳腺结节超声实例的分割图

Fig. 4 Ultrasonic case segmentation images of breast nodules under different algorithms

以标签图为基准,UNet 网络模型不能良好分割出完整的恶性结节区域,但只能对恶性结节区域进行散在分割,分割效果较差。FCN 网络模型与 DeepLabV3 网络模型虽然能完整地分割出乳腺恶性结节区域,但就分割结果而言,恶性结节周围均出现较多的假阳性区域,无法进行有效分割。SegFormer 网络模型的分割结果形态更逼近于标签图,虽在部分细节部位仍不能精准分割结节区域,但与其他 3 种网络模型相比,其分割效果最优。

在实验 2 中,为进一步证明本文提出的 SegFormer 网络模型确实有效并验证其鲁棒性与将其应用于超声影像中的临床价值,本文将 SegFormer 网络模型应用于其他 3 个种类的超声影像数据集的图像分割,包括甲状腺结节超声影像数据集、血管瘤超声影像数据集、颈部臂丛神经超声影像数据集以及乳腺结节超声影像数据集。实验 2 将 SegFormer 网络模型应用中的各数据集结合实验 1 中乳腺结节超声影像数据集进行实验,所得评价指标结果如表 2 所列,甲状腺、臂丛神经、血管瘤的实例分割结果如图 5 所示。

表 2 不同数据集的分割结果

Table 2 Segmentation results of different data sets

Dataset type	mIoU	Accuracy	DSC
甲状腺结节	90.50	97.04	94.91
血管瘤	83.90	94.89	81.83
颈部臂丛神经	65.15	97.51	74.07
乳腺结节	81.23	96.22	88.91

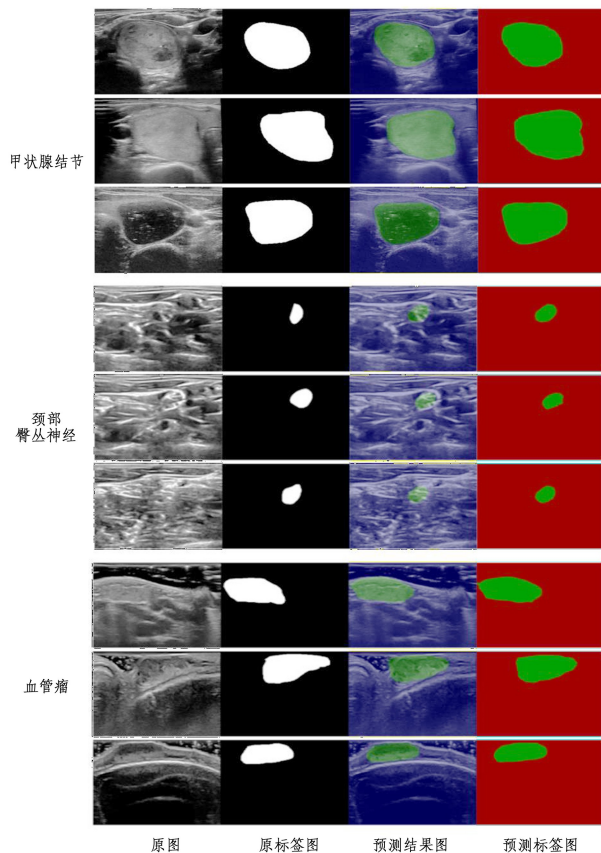


图 5 不同数据集的超声实例分割图

Fig. 5 Ultrasonic case segmentation images of different data sets

从表 2 可以看出,SegFormer 网络模型在甲状腺结节超声数据集中表现最优,mIoU、准确率、Dice 系数分别可达 90.50%,97.04%及 94.91%。同时,此网络在血管瘤与乳腺结节超声影像数据集中均表现良好。对于颈部臂丛神经来说,由于其需要被分割出的区域均较小,且边界与周围区域交界不清,背景信息复杂、噪声大,使得颈部臂丛神经验证集的评价结果稍差,这说明此网络应将提升分割精度作为今后改进此网络的主要目标。

结束语 针对超声影像分割,本文提出了基于 SegFormer 的超声影像图像分割,一方面,将超声标签图转化为单通道形式,并对其进行二值化处理为像素值为 0 和 1 的连续标签,以完成对数据集图像的预处理;另一方面,采用迁移学习的方式载入预训练模型微调已经训练好的模型参数,并选用带有动量的随机梯度下降优化器来加速收敛速度及减小震荡。与 FCN,UNet 和 DeepLabV3 的对比实验的结果表明,该模型在评估指标上最优。实验结果同时表明,该模型具有简单、高效、鲁棒性强等优点。

超声影像成像主要依靠临床医生的实际操作,随着临床医生使用超声探头时的角度与力度的改变,超声的成像结果也随之改变。超声影像往往是连续的动态影像,由于 SegFormer 轻便的网络结构,可迅速、实时地分割出超声影像中需被分割的特征区域,及时为临床医生提供客观的、可靠的临床诊断依据。

尽管 SegFormer 可应用于各种超声影像,并均取得了较好的分割效果,但对边缘的精准分割及对较小分割目标的分割精度仍有待加强。这也是我们今后将进一步研究的方向。

参考文献

- [1] ZHOU T, DONG Y L, HUO B Q, et al. U-Net and its applications in medical image segmentation: a review[J]. *Journal of Image and Graphics*, 2021, 26(9): 2058-2077.
- [2] KANG X D. *Medical Image Processing*[M]. People's Medical Publishing House, 2009: 200.
- [3] CHEN C, QI F. Review on Development of Convolutional Neural Network and Its Application in Computer Vision[J]. *Computer Science*, 2019, 46(3): 63-73.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 580-587.
- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 3431-3440.
- [6] YAP M H, GOYAL M, OSMAN F M, et al. Breast ultrasound lesions recognition: end-to-end deep learning approaches [J]. *Journal of Medical Imaging*, 2018, 6(1): 011007.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]// *Proceedings of the 2015 International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015: 234-241.
- [8] ZHUANG Z, LI N, JOSEPH RAJ A N, et al. An RDAU-NET model for lesion segmentation in breast ultrasound images [J]. *PloS One*, 2019, 14(8): e0221535.
- [9] IBTEHAZ N, RAHMAN M S J N N. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. *Neural Networks*, 2020, 121: 74-87.
- [10] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn [C]// *Proceedings of the IEEE International Conference on Computer Vision*. 2017: 2961-2969.
- [11] WANG B, LI M X, LIU X. Ultrasound Image Segmentation Method of Thyroid Nodules Based on the Improved U-Net Network [J]. *Journal of Electronics & Information Technology*, 2022, 44(2): 514-522.
- [12] CARNEIRO G, NASCIMENTO J, FREITAS A. Robust left ventricle segmentation from ultrasound data using deep neural networks and efficient search methods [C]// *Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging*. From Nano to Macro: IEEE, 2010: 1085-1088.
- [13] ANAS E M A, NOURANIAN S, MAHDAVI S S, et al. Clinical target-volume delineation in prostate brachytherapy using residual neural networks [C]// *Proceedings of the 2017 International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017: 365-373.
- [14] MILLETARI F, AHMADI S A, KROLL C, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound [J]. *Computer Vision and Image Understanding*, 2017, 164: 92-102.
- [15] CHEN L C, PAPANDEOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. *arXiv: 1706. 05587*, 2017.
- [16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. *arXiv: 2010. 11929*, 2020.
- [17] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]// *Proceedings of the 2020 European Conference on Computer Vision*. Springer, 2020: 213-229.
- [18] WANG W, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 568-578.
- [19] TAY Y, DEGHANI M, BAHRI D, et al. Efficient transformers: A survey [J]. *arXiv: 2009. 06732*, 2020.
- [20] XIE E, WANG W, YU Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. *Advances in Neural Information Processing Systems*, 2021, 34.



YANG Jingyi, born in 1999, undergraduates. Her main research interests include medical imaging diagnosis and so on.



KANG Xiaodong, born in 1964, Ph.D., professor. His main research interests include medical image processing and medical information system integration.