

全局与局部注意力机制的胃部肿瘤分割算法

徐凯成¹, 方志军¹, 蔡清萍², 卫子然², 高永彬¹, 姜晓燕¹

1. 上海工程技术大学电子电气工程学院, 上海 201600; 2. 上海长征医院, 上海 200003

【摘要】通过CT实现术前胃部肿瘤诊断是一种潜在高效的技术方法,而准确的肿瘤影像分割是实现该方法的关键。为了能够精确地提取到肿瘤区域,提出一种基于注意力机制的2D分割网络GLat-Net对上腹部CT影像中的胃部肿瘤区域进行分割,通过增加对肿瘤周围区域的关注,从全局和局部两个角度提取有效的上下文信息;同时在解码模块中引入权重模块突出具有代表性的特征。通过实验结果证明,相比较于其他前沿分割方法,该算法在胃部肿瘤分割上有更高的准确度。

【关键词】胃部肿瘤;上腹部CT;深度学习;医学影像分割, GLat-Net

【中图分类号】TP391;R318

【文献标志码】A

【文章编号】1005-202X(2021)04-0446-06

Gastric tumor segmentation algorithm based on global-local attention mechanism

XU Kaicheng¹, FANG Zhijun¹, CAI Qingping², WEI Ziran², GAO Yongbin¹, JIANG Xiaoyan¹

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201600, China; 2. Shanghai Changzheng Hospital, Shanghai 200003, China

Abstract: Using CT to realize preoperative gastric tumor diagnosis is a potentially efficient technical method, of which accurate tumor image segmentation is the prerequisite for preoperative diagnosis. In order to achieve a better performance on the extraction of tumor, a 2D segmentation network (GLat-Net) based on attention mechanism is proposed to complete the segmentation of the gastric tumor on upper abdominal CT images, which focuses on the area around the tumor and extracts effective contextual information from global and local perspectives. Furthermore, the weight module is introduced to decoder module for highlighting the representative features. The experiment results show that the proposed approach outperforms current state-of-the-art methods on gastric tumor segmentation, having a higher segmentation accuracy.

Keywords: gastric tumor; upper abdominal CT; deep learning; medical image segmentation; GLat-Net

前言

胃癌作为全球范围内致死率最高的癌症之一^[1]。胃癌的术前分期对治疗策略的制定起着至关重要的作用^[2]。上消化道内镜检查(胃镜)是胃癌检测的金标准,但由于其本身是一种侵入性检查,不仅对操作医师的技术要求较高,且患者需要承受一定的痛苦^[3]。通过CT诊断胃癌是一种高效无痛的方法,也成为未来医学工程重点发展的胃癌诊断技术手段^[4]。然而由于专家资源分布不均衡,在CT上实现精准的

术前分期难度很高,事实上即使是由多位专家共同完成诊断,其结果仍然会存在很强的主观性。基于深度学习的方法由于其出色的自学习能力成为医学图像分析的主要选择之一,在减少专家重复性工作的同时提供了更加客观的参考。

分割作为计算机视觉的基本任务之一,被广泛地应用于各种领域。在医学影像中,通过对病灶及器官组织区域的提取可以获得许多有价值的信息,如尺寸、位置等,这些信息对最后诊断提供了很大的帮助。医学影像分割在诸多医学影像应用中起着至关重要的作用,其基本原理是通过灰度、颜色、纹理、亮度和对比度来识别病变或器官区域所对应的像素或体素^[5-6]。与传统自然图像相比,医学影像分割的难点主要有数据量少,图像质量低(对比度、模糊、噪音、伪影和失真),不同样本间器官组织差别较大等^[7-8]。

【收稿日期】2020-10-24

【基金项目】上海市科委重点项目(18411952800)

【作者简介】徐凯成,硕士研究生,主要研究方向:图像处理, E-mail: xukch@foxmail.com

【通信作者】方志军,博士,硕士生导师,教授,主要研究方向:图像处理、视频编码以及模式识别, E-mail: zjfang@sues.edu.cn

近几年,由于U-Net网络在医学影像分割任务中的出色表现^[9],被许多学者选择为基础框架,并以此展开进一步研究。对于医学影像来说,复杂的背景信息会导致提取的特征中包含多余噪声,而现有的一些分割算法并没有针对编码过程中提取的粗糙特征进行精细筛选^[10-11],这增加了解码模块还原对象的难度。因此,使用合适的方法来获取上下文信息并实现特征的筛选十分关键。Fu等^[12]尝试获取全局依赖来丰富局部特征,这类方法太过于依赖全局信息而忽略了局部信息的自身价值;Gu等^[13]采用了不同感受野的尺寸来获取更多的上下文信息,该方法虽然提升了感受野,但是这些信息仍然是局部的,即信息来源为一部分周围像素。

针对胃部肿瘤较难精确分割的问题,本文提出了一种基于注意力机制的2D分割网络GLat-Net来实现上腹部CT影像胃部肿瘤区域的分割。并引入一种注意力机制模块,分别从全局和局部两个角度来突出有价值信息和抑制无用信息;同时采用权值模块决定解

码过程中通道上的权值,通过对通道上的特征进行筛选来提升分割的效果。相较于已有的方法,其主要优势在于同时采用了全局与局部视角,以共同监督互相约束的方式来对粗糙特征进行精细筛选,从而有效的提高了胃部肿瘤分割的准确率。本文所提出的方法应用到上海长征医院临床上腹部CT影像数据分析中,结果表明GLat-Net可以精确地提取胃部肿瘤,相较于其他前沿分割方法有着更好的表现。

1 分割方法

本文提出运用GLat-Net网络结构作为分割网络,主要由编码、GL-Attention模块和解码3部分构成,如图1所示。本文采用ResNet-34作为模型的backbone来对输入进行特征提取,为了获取更丰富的上下文信息,使用全局-局部注意力机制,结合全局和局部两个角度来提取更丰富的特征。在编码过程中,本文采用了加权模块来学习通道之间的相关性并给通道赋予权值,进一步提升分割表现。

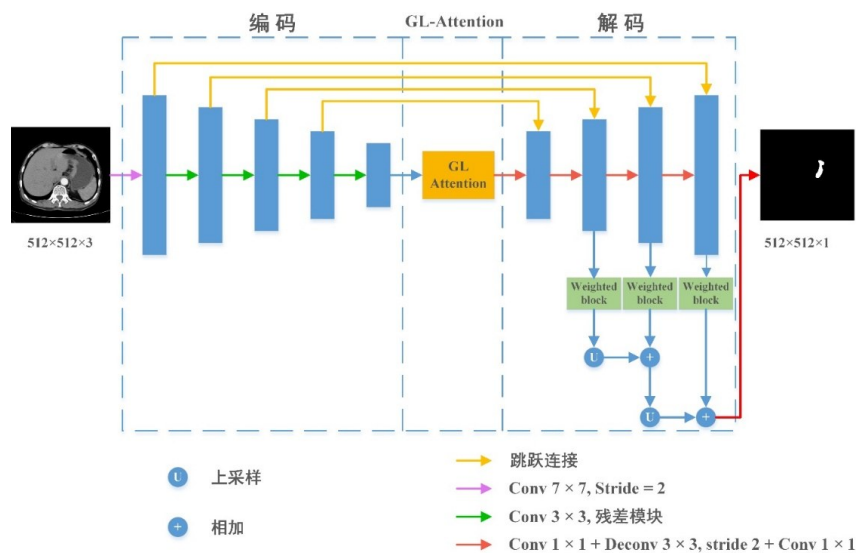


图1 GLat-Net结构示意图
Fig.1 GLat-Net architecture

1.1 编码

由于数据量少的限制,为了降低欠拟合和过拟合的风险,本文选择了ResNet-34作为网络的主体。与传统的U-Net结构(每个编码模块包含两个卷积层和一个池化层)相比,ResNet-34的跳跃连接在抑制梯度消失现象上具有更好的表现,并且在训练中提升了网络的收敛能力^[14]。

1.2 全局-局部注意力机制

由于受到容积效应的影响,肿瘤在CT上呈现的密度会与实际密度存在偏差。胃部肿瘤的分割主要面临以下几个困难:肿瘤形状的多样性,不同分期下

肿瘤尺寸的差异性以及组织纹理的模糊性。如图2所示,红色标注区域为肿瘤位置。因此,为了能更好地提取出肿瘤,需要增加对肿瘤周围区域的关注。Li等^[15]提出,全局视角虽然能提供有价值的上下文信息,但是图像中并不是所有对象都对物体定位起着积极作用,而物体周边的局部信息能对特殊区域内的推理提供线索。因此,对于医学影像这类具有复杂背景的数据而言,合理利用局部信息是十分必要的。基于此,本文提出一种全局-局部注意力机制,从全局和局部两个角度突出有效信息,抑制干扰信息,提升肿瘤提取的准确性。

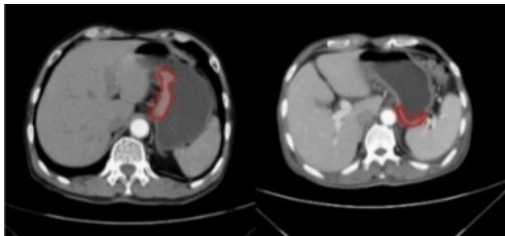


图2 不同数据样本间的多样性
Fig.2 Diversity of different data samples

在图3所示的GL-Attention结构中,分别设置了全局和局部双路径来实现特征的共同筛选。本文按水平与垂直两个方向来获取有价值的全局上下文信息,首先将输入张量 x 的行元素输入到双向长短期记忆网络(biLSTM)中,得到与 x 尺寸一致的张量 x' ,随后将 x' 的列元素也以同样方式输入到biLSTM中并得到张量 x'' ,通过两个双向的biLSTM使得感受野能够覆盖整个输入图,该张量中每个元素代表全局视角

下的原输入 x 对应位置的权值。类似地,为了获取局部上下文信息,在局部路径上采用了金字塔池化层^[16],池化核尺寸分别设置为:2×2、3×3和5×5,多尺寸的池化核提供了不同大小的感受野,这使得网络能够适应目标物体尺寸上的差异。采用双线性插值来统一经金字塔池化后得到特征图像尺寸,将不同池化层的输出张量拼接后,通过1×1卷积得到与输入张量 x 通道数一致的局部特征图像,该张量中每个元素分别代表局部视角下的原输入 x 对应位置的权值。将全局特征图像和局部特征图像进行逐元素相乘经过softmax函数后得到最终的注意力图 a :

$$a = \sigma \left(R_v \left(R_h(x) \right) \times P(x) \right) \tag{1}$$

其中, $R(x)$ 和 $P(x)$ 分别代表两次的biLSTM和金字塔池化, σ 为softmax函数。效果如图4所示。将注意力图 a 与输入特征 x 相乘后便得到筛选后的特征,最后,将输入张量 x 与筛选后的特征进行相加。

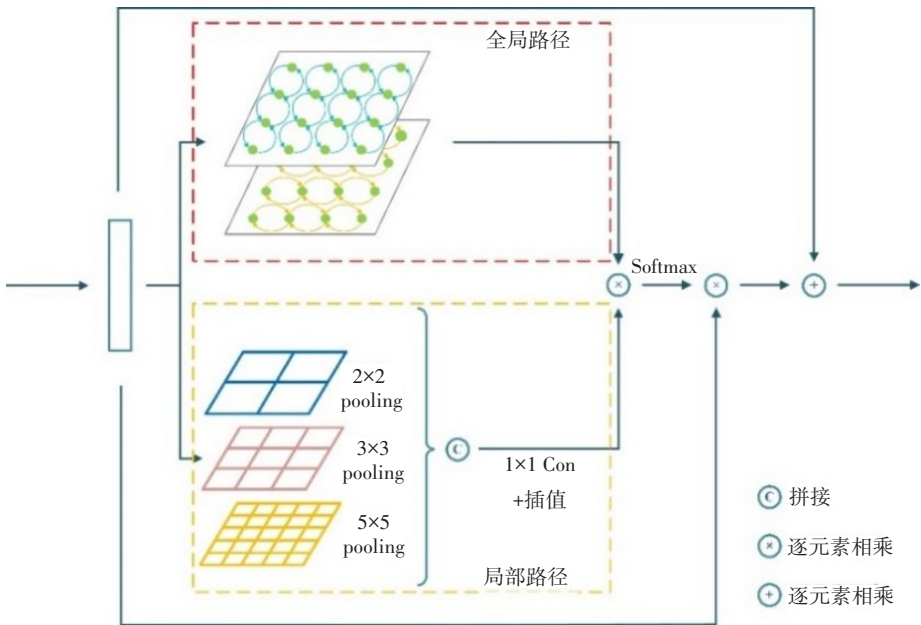


图3 GL-Attention结构图
Fig.3 GL-Attention architecture

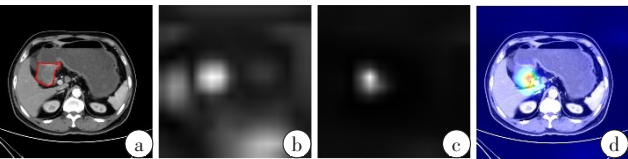


图4 GL-Attention效果的可视化
Fig.4 Visualization of GL-Attention

图 a~d 分别为原图及其标注、高级特征、注意力图以及注意力热图

1.3 解码权重模块

在解码过程中,网络会根据在编码中得到的高级特征进行图像还原,值得注意的是,并不是所有高级特

征都有助于提升分割精度。为了能学习它们之间的相关性以及不同通道间的贡献,本文采用了一种权重模块来获取各个通道的权值,实现通道上的特征筛选^[17]。

如图5所示,首先通过全局平均池化层对输入张量 $t \in R^{W \times H \times 1}$ 做逐通道的统计,统计结果 $u \in R^{1 \times 1 \times 1}$ 中的每个元素 $u_m(m=1, 2, \dots, 1)$ 可以用来描述对应通道上的整个图像。根据统计结果 u ,采用两个1×1的卷积与ReLU和Sigmoid函数来获取通道权重 $A \in R^{1 \times 1 \times 1}$,通道权重 A 中每个元素 $a^m(m=1, 2, \dots, 1)$ 代表着通道 m 的权值。值得注意的是,为了减少计算消耗,第一个1×1卷积层的输出张量通道数被压缩为 $1/r$, 1 为原通道数,

r 为压缩率,第二个 1×1 卷积层将压缩后的通道还原为1。将生成的权重 A 与原输入张量 t 对应相乘得到筛选后的结果 t' 。具体计算过程如下所示:

$$u = \frac{1}{H \times W} \sum_{i=1}^{H \times W} (v \otimes t) \quad (2)$$

$$att = S(w_2 \otimes (R(w_1 \otimes u))) \quad (3)$$

$$t' = t * att \quad (4)$$

其中, \otimes 为卷积运算, v, w_1 和 w_2 分别为对应特征的权

重矩阵, R 和 S 分别代表ReLU函数和Sigmoid函数,通过实验证明,当解码权重块应用于 d_2, d_3 和 d_4 的侧输出时效果最为明显,其中, $d_n(n=1, 2, 3, 4)$ 分别代表第 n 个解码模块。由于不同解码权重模块的输出尺寸不同,为了使输出张量间能够进行累加运算,采用了插值策略(用 U 表示)来统一 t' 的尺寸,计算过程如下所示:

$$Q = t'_3 + U(t'_2 + U(t'_1)) \quad (5)$$

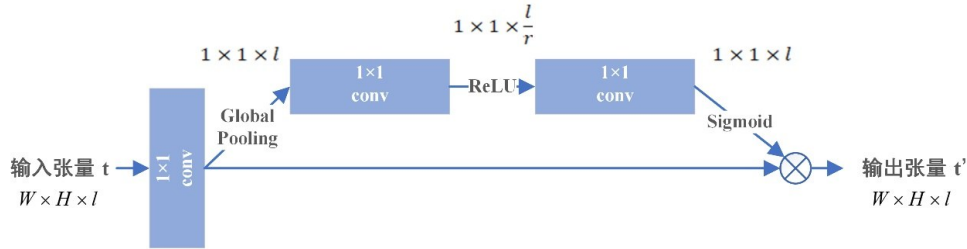


图5 解码权重模块说明

Fig.5 Illustration of weight module in decoder module

1.4 loss 函数

由于肿瘤的尺寸在不同分期阶段所呈现出的尺寸差异较大。当背景的像素个数远远大于肿瘤的像素个数时,正负样本之间的差距会严重失衡。标准交叉熵损失函数并不适用于这种情况。为了解决此类问题,采用了对于类别不均衡数据中表现更好的 dice coefficient loss 函数^[18-19],定义如下:

$$L_{\text{dice}} = 1 - \frac{\sum_c 2 w_c \sum_i p(c, i) g(c, i)}{\sum_c \sum_i p^2(c, i) + \sum_c \sum_i g^2(c, i)} \quad (6)$$

其中, N 和 C 分别代表像素和类别的总数, $p(c, i) \in [0, 1]$ 是预测为 c 类的概率。相似地, $g(c, i) \in \{0, 1\}$ 是第 c 类的标签, $\sum_c w_c = 1, \sum_c w_c = 1$ 代表不同类的权重。

整体loss函数定义如下:

$$L = L_{\text{dice}} + L_{\text{reg}} \quad (7)$$

$$L_{\text{reg}} = \lambda \|w\|_2^2 \quad (8)$$

其中, L_{reg} 代表L2正则化函数来防止网络在训练过程中出现过拟合, λ 是正则化系数, w 是权重系数向量。

2 实验与分析

2.1 实验数据

本实验为了讨论GLat-Net应用于胃部肿瘤分割在临床诊断上的有效性,采用了上海长征医院临床中440组包含不同分期的上腹部CT数据,图像尺寸为 512×512 ,其中,390组将作为训练集,50组作为测试集。

2.2 实验设置

本次实验使用Pytorch库来搭建网络模型,操作

系统为Ubuntu 16.04,显卡为11G的NVIDIA GeForce GTX 1080 Ti。

在训练中,总迭代次数total_iter为150,初始学习率base_lr设置为0.0002,采用了poly学习率衰减策略来调整学习率,更新公式为:

$$\text{lr} = \text{base_lr} \times \left(1 - \frac{\text{iter}}{\text{total_iter}}\right)^{\text{power}} \quad (9)$$

其中,power设置为0.9,实验采用Adam算法对网络进行优化。

2.3 实验结果分析

为了验证本文所提出的GLat-Net的有效性,通过计算交并比(IoU)和准确率(Acc),将其与当前前沿的分割算法U-Net^[10]、Dense U-Net^[20]、CE-Net^[13]和Attention U-Net^[21]进行对比,对比参数IoU和Acc,见公式(10)和(11):

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{FN} + \text{TP}} \quad (10)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \quad (11)$$

其中,TP, FN, FP, TN分别代表真阳性,假阴性,假阳性以及真阴性像素的数量。经计算,获得胃部肿瘤分割效果对比如表1所示。

从量化结果可以看到,对于U-Net、Dense U-Net、CE-Net和Attention U-Net,CE-Net在肿瘤分割任务上表现更为出色,其在IoU和Acc两个指标上分别达到了79.80%和99.47%,而本文提出的GLat-Net达到了80.74%和99.53%。相较于传统的U-Net,CE-Net在IoU上提升了0.41%,而GLat-Net提升了1.35%。由此可以得出GLat-Net输出的分割结果与标注有更

表 1 胃部肿瘤分割的量化结果(%)

Tab.1 Quantitative results of gastric tumor segmentation (%)

方法	IoU	Acc
U-Net	79.39	99.43
Dense U-net	79.44	99.44
Attention U-Net	78.95	99.43
CE-Net	79.80	99.47
GLat-Net	80.74	99.53

高的重合率,在原图中更多的像素被准确地分类。ROC 曲线对比如图 6 所示。

分割的可视化结果如图 7 所示,胃部肿瘤的位置被用不同的颜色标出,图 7a 为 CT 原图。作为参考标准,图 7b 中,原图中肿瘤的正确位置被用红色标出。图 7c~e 分别为 U-Net、Attention U-Net 以及 CE-Net 的分割可视化结果,分别用绿色、蓝色、青色标记,本实验分割结果图 7f 由黄色标出。相较于对比的分割算法,GLat-Net 在肿瘤的提取方面取得了更好的效果。通过对比可视化结果中红色方框标出的区域,发现

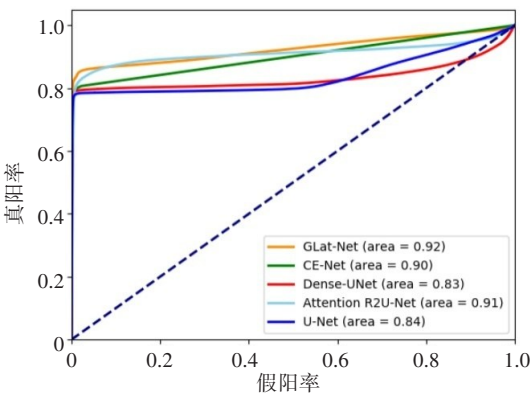


图 6 ROC 曲线对比

Fig.6 Comparison of ROC curve

在 U-Net、Attention U-Net 和 CE-Net 的结果中存在着间断的现象,且在轮廓上存在较明显的误差,但是 GLat-Net 生成的结果中并不存在这类问题,肿瘤区域的像素在 GL-Attention 的全局局部共同约束下增强了肿瘤区域像素间的相关性,有效地抑制了间断现象的发生,且在轮廓上也有更好的表现。

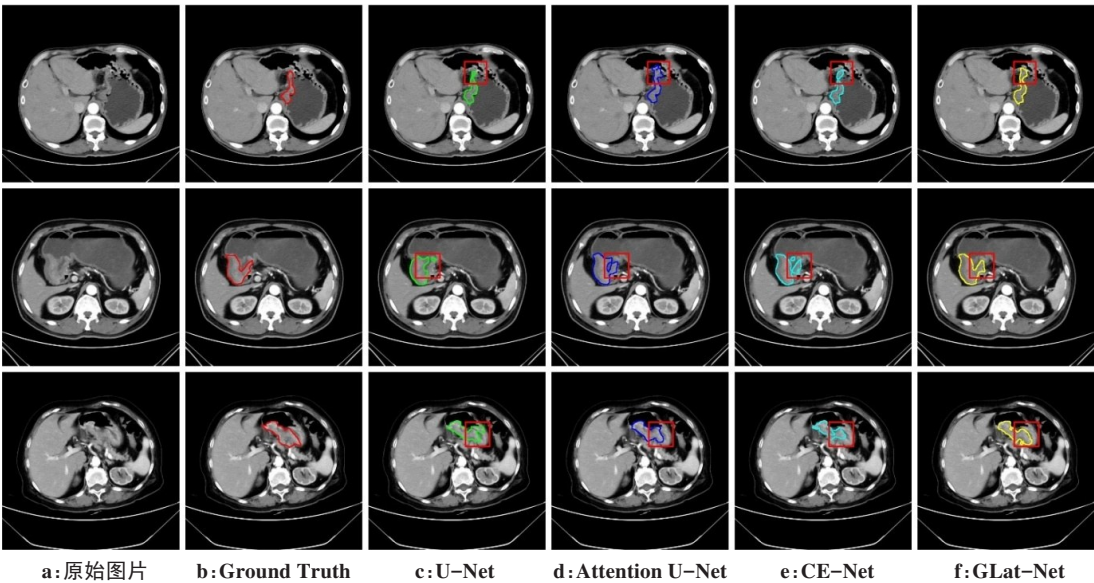


图 7 胃部肿瘤分割结果

Fig.7 Results of gastric tumor segmentation

1~3 行分别代表不同患者的胃部肿瘤 CT 影像

2.4 消融实验

训练曲线图如图 8 所指示,相较于 U-Net, ResNet-34 的曲线收敛地更快。

在表 2 中,给出了权重模块在不同解码模块上的表现,根据实验结果可以证明,当解码权重模块被应用在第 2、3、4 层解码层时效果最佳。为了评估网络中各个部分的表现,每个模块都分别单独进行了实验,来验证本文所提出的 GLat-Net 的有效性。如表 3

所示,基准模型是以 ResNet-34 为主体,由编码和解码两部分组成的 U 型结构网络,单独进行实验后,得到 IoU 和 Acc 结果分别为 79.37% 和 99.45%。添加了本文提出的 GL-Attention 模块后, IoU 和 Acc 的量化结果分别为 80.44% 和 99.52%,解码权重模块单独作用时, IoU 和 Acc 中的量化结果分别为 80.59% 和 99.52%。综上可见,本文所提出的方法有效地提升了胃肿瘤分割的准确度。

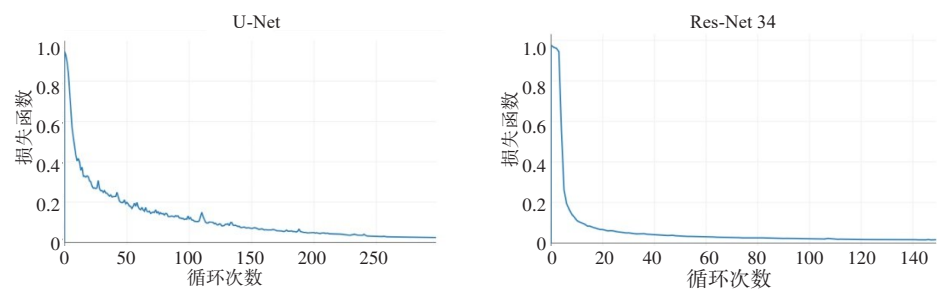


图8 U-Net和Res-Net 34训练曲线图

Fig.8 Training curves of U-Net and Res-Net 34

表2 解码权重模块的表现对比(%)

Tab.2 Performance comparison of decoder module with or without weight module (%)

权重模块	IoU	Acc
不使用权重模块	79.37	99.45
d1+d2	79.51	99.50
d2+d3	79.95	99.50
d3+d4	80.17	99.51
d1+d2+d3	79.49	99.49
d2+d3+d4	80.59	99.52
d1+d2+d3+d4	79.86	99.50

表3 消融实验量化结果(%)

Tab.3 Quantitative results of ablation experiment (%)

方法	IoU	Acc
Base Network	79.37	99.45
Base Network+GL-Attention	80.44	99.52
Base Network+DecoderWeighted block	80.59	99.52
Base Network+GL-Attention+DecoderWeighted block	80.74	99.53

3 总结

通过CT实现胃部肿瘤医学影像的分割和快速诊断成为术前的重要依据。本文提出了一种基于注意力机制的分割模型GLat-Net的胃部肿瘤区域提取方法。该方法结合全局和局部两个视角来提取上下文信息,并根据这些信息来突出有效特征和抑制无用特征,降低了CT影像上肿瘤不确定性特征的干扰;同时采用了解码权重模块来学习通道间的相关性,通过对通道赋予权重提升了分割效果。实验证明该方法能够在胃部肿瘤分割任务上得到准确的分割结果,与前沿分割网络CE-Net相比,该方法在交并比上提升了0.94%。

【参考文献】

[1] TORRE LA, BRAY F, SIEGEL R L, et al. Global cancer statistics, 2012 [J]. CA Cancer J Clin, 2015, 65(2): 87-108.

[2] SENDLER A, DITTLER H J, FEUSSNER H, et al. Preoperative staging of gastric cancer as precondition for multimodal treatment[J]. World J Surg, 1995, 19(4): 501-508.

[3] 张瑞, 雷林, 彭绩, 等. 胃癌筛查的研究进展[J]. 中国肿瘤, 2020, 29(6): 411-418.

ZHANG R, LEI L, PENG J, et al. Research progress of gastric cancer screening[J]. China Cancer, 2020, 29(6): 411-418.

[4] 薛萍, 王峰. MRI与CT对胃癌术前分期诊断的作用研究[J]. 影像研究与医学应用, 2020, 4(20): 40-41.

XUE P, WANG F. Study on the role of MRI and CT in preoperative diagnosis of gastric cancer [J]. Journal of Imaging Research and Medical Applications, 2020, 4(20): 40-41.

[5] MAHMOUD A A, EL-RABAIE E S, TAHA T E, et al. Medical image segmentation techniques, a literature review, and some novel trends [J]. MJEEER, 2018, 27(2): 23-58.

[6] AGRAVAT R R, RAVAL M S. Deep learning for automated brain tumor segmentation in MRI images [M]//Soft Computing Based Medical Image Analysis. Academic Press, 2018: 183-201.

[7] SEIBERT J A. X-ray imaging physics for nuclear medicine technologists. Part 1: basic principles of X-ray production[J]. J Nucl Med Technol, 2004, 32(3): 139-147.

[8] HESAMIAN M H, JIA W, HE X, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. J Digit Imaging, 2019, 32(4): 582-596.

[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. Springer, Cham, 2015: 234-241.

[10] FU H Z, CHENG J, XU Y W, et al. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation[J]. IEEE Trans Med Imaging, 2018, 37(7): 1597-1605.

[11] ALOM M Z, HASAN M, YAKOPCIC C, et al. Recurrent residual convolutional neural network based on U-Net (r2u-net) for medical image segmentation[J]. arXiv preprint arXiv: 1802.06955, 2018.

[12] FU J, LIU J, TIAN H, et al. Dual attention Network for scene segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 3146-3154.

[13] GU Z W, CHENG J, FU H Z, et al. CE-Net: context encoder network for 2D medical image segmentation[J]. IEEE Trans Med Imaging, 2019, 38(10): 2281-2292.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. The IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[15] LI J N, WEI Y C, LIANG X D, et al. Attentive contexts for object detection[J]. IEEE Trans Multimedia, 2016, 19(5): 944-954.

[16] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing net-work[C]. The IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 2881-2890.

[17] ZHANG Z, FU H, DAI H, et al. ET-Net: a generic edge-attention guidance network for medical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 442-450.

[18] CRUM W R, CAMARA O, HILL D L. Generalized overlap measures for evaluation and validation in medical image analysis[J]. IEEE Trans Med Imaging, 2006, 25(11): 1451-1461.

[19] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmen-tation [C]. 2016 4th International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.

[20] GUAN S, KHAN A A, SIKDAR, S et al. Fully dense U-Net for 2D sparse photoacoustic tomography artifact removal[J]. IEEE J Biomed Health Inform, 2020, 24(2): 568-576.

[21] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: learning where to look for the pancreas[J]. arXiv preprint arXiv: 1804.03999, 2018.

(编辑:薛泽玲)