

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2020)10-2119-09

论文引用格式: Hao X Y, Xiong J F, Xue X D, Shi J, Wen K, Han W T, Li X Y, Zhao J and Fu X L. 2020. 3D U-Net with dual attention mechanism for lung tumor segmentation. Journal of Image and Graphics 25(10): 2119-2127(郝晓宇, 熊俊峰, 薛旭东, 石军, 文可, 韩文廷, 李骁扬, 赵俊, 傅小龙. 2020. 融合双注意力机制 3D U-Net 的肺肿瘤分割. 中国图象图形学报 25(10): 2119-2127 [DOI: 10.11834/jig.200282])

## 融合双注意力机制 3D U-Net 的肺肿瘤分割

郝晓宇<sup>1</sup>, 熊俊峰<sup>4,5</sup>, 薛旭东<sup>2</sup>, 石军<sup>1</sup>, 文可<sup>1</sup>, 韩文廷<sup>1</sup>, 李骁扬<sup>2</sup>, 赵俊<sup>4</sup>, 傅小龙<sup>3</sup>

1. 中国科学技术大学计算机科学与技术学院, 合肥 230026; 2. 中国科学技术大学附属第一医院 肿瘤放疗科, 合肥 230001; 3. 上海交通大学附属胸科医院放射肿瘤科, 上海 200030;
4. 上海交通大学生物医学工程学院, 上海 200240; 5. 腾讯医疗健康, 上海 200000

**摘要:** 目的 精确的肺肿瘤分割对肺癌诊断、手术规划以及放疗具有重要意义。计算机断层扫描(computed tomography, CT)是肺癌诊疗中最重要的辅助手段,但阅片是一项依靠医生主观经验、劳动密集型的工作,容易造成诊断结果的不稳定。实现快速、稳定和准确的肺肿瘤自动分割方法是当前研究的热点。随着深度学习的发展,使用卷积神经网络进行肺肿瘤的自动分割成为了主流。本文针对 3D U-Net 准确度不足,容易出现假阳性的问题,设计并实现了 3 维卷积神经网络 DAU-Net(dual attention U-Net)。方法 首先对数据进行预处理,调整 CT 图像切片内的像素间距,设置窗宽、窗位,并通过裁剪去除 CT 图像中的冗余信息。DAU-Net 以 3D U-Net 为基础结构,将每两个相邻的卷积层替换为残差结构,并在收缩路径和扩张路径中间加入并联在一起的位置注意力模块和通道注意力模块。预测时,采用连通域分析对网络输出的二值图像进行后处理,通过判断每个像素与周围 26 个像素的连通关系获取所有的连通域,并清除最大连通域外的其他区域,进一步提升分割精度。结果 实验数据来自上海胸科医院,总共 1 010 例肺癌患者,每例数据只包含一个病灶,专业的放射科医师提供了金标准,实验采用十折交叉验证。结果表明,本文提出的肺肿瘤分割算法与 3D U-Net 相比,Dice 系数和哈斯多夫距离分别提升了 2.5% 和 9.7%,假阳性率减少了 13.6%。结论 本文算法能够有效提升肺肿瘤的分割精度,有助于实现肺癌的快速、稳定和准确分割。  
**关键词:** U-Net; 计算机断层扫描(CT); 肺部肿瘤; 分割; 注意力机制

## 3D U-Net with dual attention mechanism for lung tumor segmentation

Hao Xiaoyu<sup>1</sup>, Xiong Junfeng<sup>4,5</sup>, Xue Xudong<sup>2</sup>, Shi Jun<sup>1</sup>, Wen Ke<sup>1</sup>,  
Han Wenting<sup>1</sup>, Li Xiaoyang<sup>2</sup>, Zhao Jun<sup>4</sup>, Fu Xiaolong<sup>3</sup>

1. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China;
2. Department of Radiation Oncology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China;
3. Department of Radiation Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai 200030, China;
4. School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
5. Tencent HealthCare, Co. Ltd., Shanghai 200000, China

**Abstract:** **Objective** Precise lung tumor segmentation is a necessary step in computer-aided diagnosis, surgical planning, and radiotherapy of lung cancer. Computed tomography (CT) images are important auxiliary tools in clinical medicine. The diagnosis of lung cancer tumors is labor intensive that requires professional radiologists to carefully examine hundreds of CT

收稿日期: 2020-06-10; 修回日期: 2020-07-07; 预印本日期: 2020-07-14

基金项目: 国家重点研发计划项目(2016YFB1000403); 中央高校基本科研业务费专项资金资助

**Supported by:** National Key Research and Development Program of China(2016YFB1000403); Fundamental Research Funds for the Central Universities

slices for finding and confirming the location of tumor lesions, and final reports need to be verified by other experienced radiologists. This process consumes time and effort. Doctors commonly make different diagnoses at the same time, and the same doctor may make different decisions at different times because of the difference in their subjective experience. To solve the above problems, increasing scientific researchers have devoted to the field of medical imaging by continuously promoting the combination of artificial intelligence and medical imaging, and the automatic segmentation of lung tumors has been widely investigated. To address the problems that 3D U-Net is insufficiently accurate and is prone to produce false positive pixels, this paper proposes a new network named dual attention U-Net (DAU-Net) that incorporates dual attention mechanisms and residual modules. A post processing method based on connected component analysis is used to remove the false positive regions outside the region of interest.

**Method** In accordance with the characteristics of lung CT images, we proposed a pipeline to preprocess CT images, which was divided into three steps. Standardizing pixel pitch was the first step that needs to be performed because different pixel spacings will affect the speed and quality of network convergence in the training process. The thickness of all 2D slices is 5 mm, and the range of in-plane resolution varies from 0.607 mm to 0.976 mm. Thus, linear interpolation was applied to each CT slice to obtain 1 mm in-plane resolution. The interpolated CT images still exist in 3D form. The window width and window level were then set to 1 600 and -200, respectively, that is, the pixel values in the CT image greater than 600 were set to 600 and those less than -1 000 were set to -1 000. The intensity values of images were truncated to the range of  $[-1\ 000, 600]$  and linearly normalized to  $[0, 1]$  to enhance the regions of interest when using CT images, which is helpful for the automatic segmentation of lesions. This step will make the size of each CT image less than  $N \times 512 \times 512$ , where  $N$  is the number of slices. After padding to  $N \times 512 \times 512$ , the CT images and their corresponding annotations were cropped to a constant size of  $N \times 320 \times 260$  from a fixed coordinate  $(0, 90, 130)$  of the very beginning slice, and interpolation was used to scale the size of the images to  $64 \times 320 \times 260$ . The main architecture of the network adopts the 3D form of the U-Net by replacing every two adjacent convolutional layers with a residual structure and adding two attention mechanisms to the middle of the contraction path and the expansion path to obtain DAU-Net. The network can alleviate degradation, gradient disappearance, and gradient explosion caused by the increase in the depth of the neural network by adding the residual structures. Similar to U-Net, encoder-decoder networks can merge high-resolution feature maps with position information and low-resolution feature maps with contextual information through skip connections to capture targets of different scales. However, they cannot take advantage of the positional relationship of different objects in global images and association between different categories. To retain the advantages of encoder-decoder structures and overcome the above problems, a position attention module and a channel attention module connected in parallel are combined with 3D U-Net. The position attention module can encode context information from a wide range into local features and the channel attention module can find the dependency relationship between different channels, thereby strengthening the interdependent features. The network can perform end-to-end training and it was trained by optimizing soft dice loss in this work. After inference, connected component analysis is used to remove the false positive regions that are wrongly segmented by only keeping the largest connected component and discarding other parts. Considering that this paper uses a 3D CNN (convolutional neural network), a 26-neighborhood connected component analysis method is used to determine the connection relationship between a central pixel and its 26 adjacent pixels. The output of the network has two channels, and softmax is used to make the output between zero and one. In binarization, only the channel index with a high probability is selected to obtain the final binary result where the connected component analysis method is applied. This postprocessing method effectively improves the segmentation accuracy and decreases the false positive rate (FPR). The premise of using this method is that the dataset we use contains only one lesion per case.

**Result** We retrospectively collected data from patients in Shanghai Chest Hospital from 2013 to 2017. The study was approved by Shanghai Chest Hospital, Shanghai Jiao Tong University. Ethical approval (ID: KS 1716) was obtained for use of the CT images. Experienced radiologists provided the gold standard of each case. In the experiment, we compared the standard 3D U-Net and the reproduced 3D attention U-Net. The experiment used 10-fold cross-validation for all networks, and we adopted the widely used Dice, Hausdorff distance (HD), FPR, and true positive rate to evaluate the predicted outputs. The results show that the proposed DAU-Net has powerful performance in the lung tumor segmentation task, and the postprocessing method can effectively reduce the interference of false positive regions on the segmentation results. Compared with 3D

U-Net, Dice and HD are improved by 2.5% and 9.7%, respectively, and FPR is reduced by 13.6%. **Conclusion** The proposed lung tumor segmentation algorithm can effectively improve the accuracy of tumor segmentation and help to achieve rapid, stable, and accurate segmentation of lung cancer.

**Key words:** U-Net; computed tomography (CT); lung tumor; segmentation; attention mechanism

## 0 引言

肺癌对我国居民的身体健康造成了极大威胁,发病率和死亡率呈现逐年上升的趋势(陈万青等, 2010)。目前,计算机断层扫描(computed tomography, CT)是肺癌诊断和治疗的重要辅助工具。绝大多数肺癌患者直到晚期才被确诊,错失了最佳的治疗时机,特别是在医疗资源相对匮乏的地区,因此对早期癌症病灶的检测和诊断显得尤为关键。在传统临床中,对肺癌肿瘤的诊断需要专业的放射医师对每一张 CT 切片进行认真细致的检查,寻找并确认病灶位置,同时需要其他有经验的医师进行核验,这一过程费时费力。此外,由于不同医生主观经验的差别,容易产生同时不同诊,甚至出现同一个医生在不同时间的诊断结果也不相同的情况。为了解决上述问题,快速、稳定和精准的肺肿瘤分割算法已经成为研究热点。

传统的肺部肿瘤分割方法主要可以分为两类,第 1 类为基于区域的方法,第 2 类为基于边缘的方法。基于区域的方法可分为阈值法(Wei 等, 2009)、区域生长法(Parveen 和 Kavitha, 2013)和分水岭算法(Kanitkar 等, 2015);基于边缘的方法可细分为微分算子法和活动轮廓法,活动轮廓法又包括 Snake 模型(Mirderikvand 等, 2016)和水平集算法(Farag 等, 2013)。2009 年,Wei 等人提出了一种优化后的阈值方法,融合了 OSTU(大津法)和遗传算法,用于分割肺部的感兴趣区域。Parveen 和 Kavitha 于 2013 年使用区域生长法实现了肺部病灶区域的分割,但是该方法的结果会包括胸腔轮廓、动脉等其他组织。Kanitkar 等人(2015)采用标记控制的分水岭算法,通过标记前景和背景,解决了分水岭算法的过分割问题,实现了肺癌病灶的自动检测,但该方法同样无法获取精确的肿瘤轮廓。Mirderikvand 等人(2016)使用图割算法(graph cut)和 Snakes 算法实现了对 CT 图像中肺结节的精确分割。Farag 等人(2013)提出了一种通用的肺结节形

状模型,并使用水平集算法完成对肺结节的自动分割,但是需要首先建立肺结节的先验形状模型,使得算法过程十分复杂。

随着计算机视觉技术和深度学习的发展,同时得益于不断提高的计算能力和持续增长的可用数据量,深度学习技术在医疗影像分析领域不断取得重大突破。在众多深度学习算法中,卷积神经网络(convolutional neural network, CNN)成为医疗影像分析领域的主流,与传统方法相比, CNN 能够更有效解决医疗影像分析中病灶检测、分割和分类等任务。1995 年, CNN 已经被应用于医学影像分析领域(Lo 等, 1995),但之后一段时间发展缓慢。直到 2012 年,具有划时代意义的 AlexNet(Krizhevsky 等, 2012)被提出,卷积神经网络的大规模应用才真正开始。全卷积神经网络(fully convolutional network, FCN)能够完成像素级别的分类,奠定了图像语义分割的基础。U-Net 被提出后(Ronneberger 等, 2015),逐渐成为了目前医学影像分割领域最常用的卷积神经网络结构。后来更多类似的网络出现,如 Zhou 等人(2019)提出的 U-Net++, Li 等人(2018a)提出的 H-DenseUNet(hybrid densely connected U-Net)等。这些网络的诞生推动了卷积神经网络在医疗影像领域的发展,高精度的肺肿瘤自动分割成为可能。

除了不同卷积神经网络结构本身的发展,一些特殊的模块,例如注意力(attention)模块被设计出来,用于提升语义分割的性能。Oktay 等人(2018)提出了集成注意力门(attention gate, AG)的 attention U-Net,注意力门很容易集成到标准的卷积神经网络结构中,抑制模型与任务无关的部分,同时加强学习与任务有关的特征。Fu 等人(2019)提出了具有通道注意力和位置注意力机制的双注意力机制网络(dual attention network, DANet),通过两种注意力机制增强不同通道、不同位置间具有依赖的特征,进而提升模型的整体精度。双注意力机制为本文的研究工作提供了思路。

本文采用 3 维卷积神经网络,以适应肺部肿瘤

的 3 维空间特性,实现肺肿瘤端到端的分割。在 3D U-Net 的基础上,将网络结构中每两个相邻的卷积层替换为残差模块,同时在网络结构中添加了位置注意力机制和通道注意力机制,最终得到 3 维双注意力机制 U-Net (dual attention U-Net, DAU-Net)。在预测时,本文采用基于连通域分析的后处理方法,清除假阳性区域,进一步提升肺肿瘤分割的准确度。

1 本文方法

1.1 网络结构概述

本文提出了一种以 3D U-Net 为基础的新型卷积神经网络,网络结构如图 1 所示。该模型采用典型的编码器-解码器结构,具有两条对称的路径,左边的路径称为收缩路径(编码器),右边的路径称为扩张路径(解码器)。收缩路径用于下采样,提取深层的语义特征,由于收缩路径中包含了池化层,所以特征图的尺寸不断缩小。扩张路径用于上采样,与收缩路径有所不同,池化层被替换为了上采样层,以逐渐恢复图像的分辨率,进而达到端到端分割的目的。收缩路径与扩张路径的对应层之间通过跳转链接相连。跳转链接将收缩路径中具有位置信息的高分辨率特征图传递到扩张路径,与具有上下文语义信息的低分辨率特征图融合,能够捕获不同尺度的

目标,从而实现像素级别精细分割的目的。

理论上,卷积神经网络越深性能越好,但是由于神经网络反向传播过程中梯度连乘的原因,网络加深会产生梯度消失或梯度爆炸,使得网络训练困难。针对以上问题,He 等人(2016)提出了残差结构,有助于缓解神经网络深度增加带来的网络退化、梯度消失和梯度爆炸等问题。为了增强卷积神经网络的性能,本文将 3D U-Net 中每两个相邻的卷积层替换为残差模块(如图 1 中所示的残差模块)。

具有编码器-解码器结构的卷积神经网络在医学图像领域的语义分割任务中具有突出表现,但该类网络由于感受野范围的限制,无法在局部特征中编码范围更广的上下文信息,也不能利用不同通道间的依赖关系。为了解决这个问题,同时保留编码器-解码器结构优点,本文在 3D U-Net 的编码器和解码器中间添加了并联在一起的位置注意力模块和通道注意力模块,二者的输出求和后输入解码器。使得原本应用于 2 维卷积神经网络的两种注意力模块在本文中被拓展为 3 维结构。两个模块分别在空间和通道维度获取全局范围内的特征依赖,并对特征图中任何两个具有关联的特征进行加强,从而提升网络的表示能力。网络结构如图 1 所示。图中的数字代表了残差模块中两个卷积层具有的卷积核数。

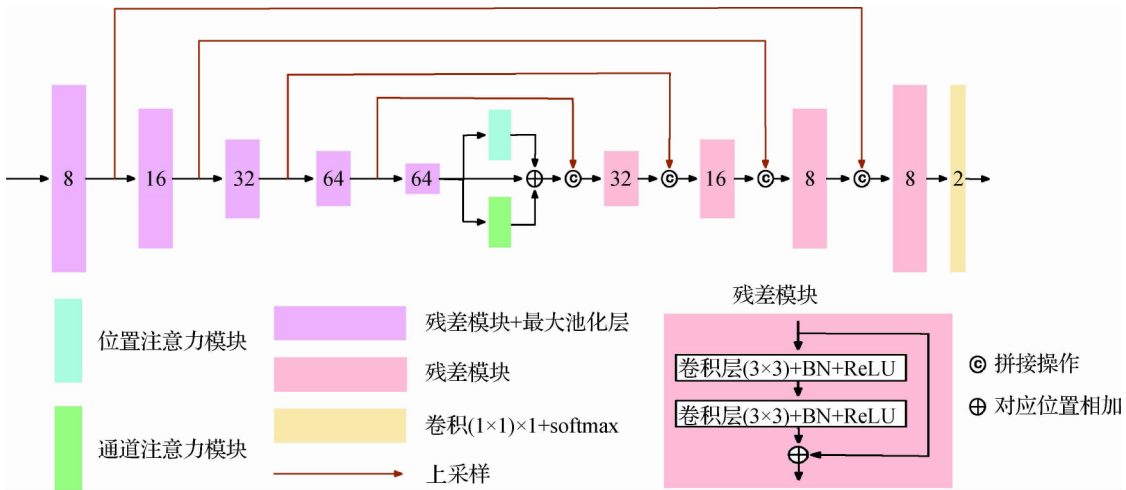


图 1 DAU-Net 网络结构图

Fig. 1 Structure diagram of DAU-Net

1.2 位置注意力模块

位置注意力模块能够在局部特征中编码更长范围内的上下文信息,从而提升局部特征的表达能

其结构如图 2 所示。位置注意力模块的工作流程如下:

- 1) 特征图A 分别通过3个卷积层,得到3个新



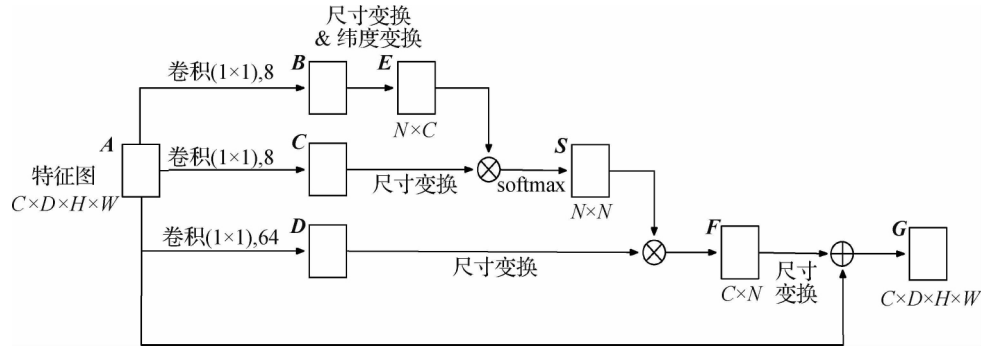


图2 位置注意力模块结构图

Fig.2 Structure diagram of position attention mechanism

的特征图  $B$ 、 $C$  和  $D$ 。

2) 对  $B$  进行尺寸变换 (reshape) 和维度变换 (transpose) 得到  $E$ , 尺寸从  $C \times D \times H \times W$  变为  $N \times C$ , 其中  $N$  的大小为  $D \times H \times W$ 。

3)  $E$  与  $C$  相乘, 再通过 softmax 函数得到空间监督图  $S$ , 这时  $S$  的尺寸为  $N \times N$ 。

4) 将  $D$  的维度 reshape 为  $C \times N$ , 之后与  $S$  相乘得到  $F$ , 再乘以系数  $\alpha$ , 之后将维度 reshape 为  $C \times D \times H \times W$ 。其中  $\alpha$  初始化为 0, 并通过训练过程进行学习。

5) 最后  $F$  与  $A$  相加得到  $G$ ,  $G$  即为位置注意力模块的输出。

### 1.3 通道注意力模块

通道注意力模块用于寻找不同通道之间的依赖关系, 并对依赖的特征进行加强, 其结构如图 3 所示。通道注意力模块的工作流程如下:

1) 分别对特征图  $A$  进行 reshape、reshape、reshape 和 transpose, 得到  $B$ 、 $C$  和  $D$ 。

2) 将  $D$  进行 reshape 和 transpose, 得到大小为

$N \times C$  的特征图, 与  $C$  相乘再通过 softmax 得到  $X$ ,  $X$  的大小为  $C \times C$ 。

3)  $B$  与  $X$  相乘得到  $E$ 。

4)  $E$  与系数  $\beta$  相乘后, reshape 为  $C \times D \times H \times W$ , 其中  $\beta$  初始化为 0, 并通过训练过程学习。

5) 最后  $E$  与  $A$  相加得到  $F$ ,  $F$  为通道注意力模块的输出。

### 1.4 评估指标与损失函数

1) 采用医学影像分割领域中使用最广泛 Dice 相似系数进行评估, 计算为

$$D = \frac{2|P \cap G|}{|P| + |G|} \quad (1)$$

式中,  $P$  和  $G$  分别为预测结果和真实标注, 而损失函数  $loss$  计算为

$$L = 1 - D \quad (2)$$

2) 采用哈斯多夫距离 (Hausdorff distance, HD) 进行评估, 定义为

$$H(A, B) = \max \left\{ \max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b) \right\} \quad (3)$$

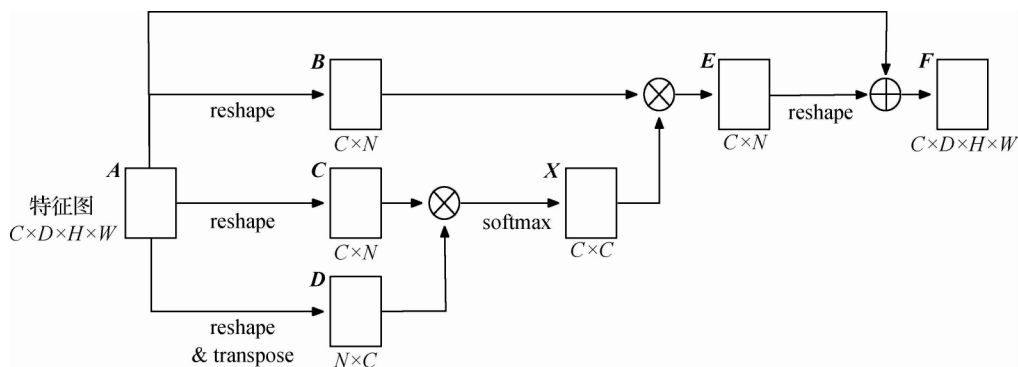


图3 通道注意力模块结构图

Fig.3 Structure diagram of channel attention mechanism

式中,  $a$ ,  $b$  分别是  $A$  和  $B$  上的像素点,  $d$  为欧氏距离。

3) 采用假阳性率( false positive rate, FPR) 和真阳性率( true positive rate, TPR) 在像素级别对分割结果进行评估, 分别定义为

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

式中,  $FP$  表示将背景错误预测为肿瘤的部分,  $TN$  表示正确预测为背景的部分,  $FN$  表示被错误预测为背景的肿瘤部分。

## 2 实验结果与分析

### 2.1 数据集

本文回顾性研究了 2013—2017 年在上海胸科医院就诊的 1 010 例肺癌患者数据, 该数据集首先应用于表皮生长因子受体( epidermal growth factor receptor, EGFR) 基因突变的检测任务( Li 等, 2018b)。该数据集由上海交通大学附属胸科医院认证, 同时 CT 图像的使用通过了道德认证( ethical approval, ID: KS1716)。每例数据都包含金标准, 由经验丰富的影像科医师手动勾画。每例 CT 图像, 由 90 ~ 130 幅尺寸为  $512 \times 512$  像素、厚度为 5 mm 的 2 维切片构成。切片内部像素的间距为 0.607 ~ 0.976 mm 之间, 每例数据均只含有一个肿瘤病灶。

### 2.2 预处理与后处理

1) 不同病例的 CT 切片厚度相同, 但是切片内部的像素间距不同, 所以首先对不同病例的像素间距进行调整, 采用线性插值的方式, 将不同病例每张切片的像素间距调整为 1 mm, 调整之后的 CT 图像仍以 3 维的形式存在。此时, 切片的尺寸会小于  $512 \times 512$  像素, 具体尺寸取决于切片的原始像素间距。其次设置 CT 图像的窗宽和窗位, 起到清除干扰、增强感兴趣区域的作用。本文将 CT 值大于 600 HU( Hounsfield unit) 设置为 600 HU, CT 值小于 -1 000 HU 设置为 -1 000 HU, 然后将像素值归一化至  $[0, 1]$ 。由于 CT 影像中每张切片的边界与人体组织区域之间具有大范围的黑色区域, 其中可能存在的成像仪器金属骨架或尾影会对模型的收敛效果产生影响, 而且会占用大量显存。所以, 本文采用

以下步骤对 CT 图像进行裁剪, 去除黑色区域( 示意图如图 4) :

(1) 将 3 维 CT 图像作为若干张 2 维切片的组合;

(2) 切片内像素间距调整为 1 mm 后, 尺寸会小于  $512 \times 512$  像素, 将每张切片的尺寸零填充为  $512 \times 512$  像素;

(3) 以左上角坐标为( 90, 130) 的像素作为起始点, 将切片的尺寸裁剪为  $320 \times 260$  像素。

(4) 将属于同一个病例的切片按原始顺序进行组合, 得到尺寸为  $N \times 320 \times 260$  的 CT 图像, 其中  $N$  为切片数量;

(5) 通过差值的方式将输入图片的尺寸缩放为  $64 \times 320 \times 260$ 。

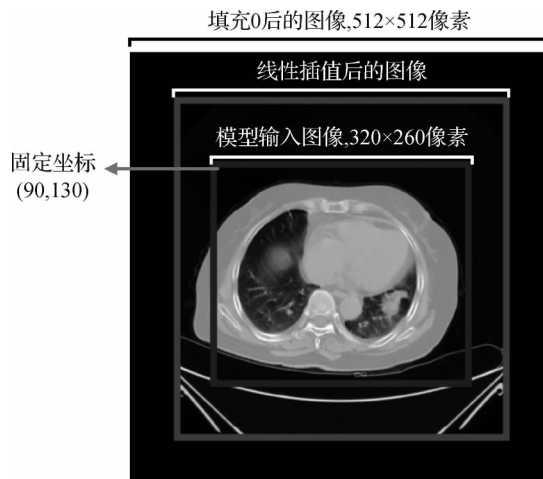


图4 数据预处理示意图

Fig. 4 Schematic diagram of data preprocessing

2) 得到卷积神经网络的输出后, 采用基于连通域分析的后处理方法清除假阳性区域。连通域( connected component) 是指图像中具有相同像素值且位置相邻的前景像素点组成的图像区域, 而连通域分析是指将各个连通域找出并进行标记。在本文涉及的肺肿瘤分割场景下, 前景为神经网络输出二值化后的阳性像素点。在连通域标记时, 由于 CT 图像本身是 3 维的特点, 采用基于二十六邻域的分析方法, 判断某像素与周围 26 个点的值是否相同, 相同则认为两个像素属于同一连通域。根据对数据集的分析, 每例数据中只包含一个感兴趣区域( 肺肿瘤病灶), 所以在得到网络输出的二值图像后, 采用连通域分析法判断每个像素与周围 26 个像素的连通关系, 获得所有的连通域并计算体积, 选择体积

最大的连通域作为最终的肿瘤区域。

### 2.3 训练细节以及参数设置

首先将所有数据随机划分为 10 份, 采用十折交叉验证。依次选择其中的一折作为测试集, 剩余数据按照 8 : 2 的比例划分为训练集和验证集。每折数据在训练时, 数据量分别为: 训练集 646 例、验证集 162 例以及测试集 202 例。

网络结构采用 Pytorch 1.2.0 实现, 使用英伟达 P40 显卡进行训练。训练时使用 Adam 作为优化器, 权重衰减设置为 0.000 01, 批尺寸设置为 4。采用余弦退火( Loshchilov 和 Hutter 2017) 学习率变化方案, 初始学习率设置为 0.01, 最小学习率设置为 0.000 01, 学习率变化迭代周期设置为 4。总共训练 150 次迭代, 保存在验证集上 Dice 系数最高的 3 组权重。在测试时, 计算 3 组权重的算数平均值作为模型的最终结果。

### 2.4 实验结果分析

表 1 列举了不同模型在肺肿瘤数据集上的实验结果。实验中使用的模型均采用 3 维卷积神经网络, 分割结果为十折交叉验证的平均值。在实验中, 本文除了对比 3D U-Net 之外, 还复现了 3D 版本的 Attention U-Net( Oktay 等 2018)。如表 1 所示, 本文提出的肺肿瘤分割算法在 Dice 系数、哈斯多夫距离(HD)、真阳性率(TPR)和假阳性率(FPR)等评估方法上都取得了最优的结果。对比 3D U-Net, 本文方法的 Dice 和 HD 分别提升了 2.5% 和 9.7%, FPR 减少了 13.6%。在不使用后处理的情况下, 3D U-Net 的 Dice 分割结果为 0.751, 3D Attention U-Net 的分割结果与 3D U-Net 相比有所提高, Dice 系数为 0.754, 但本文提出的网络 DAU-Net 的结果能够达到 0.770。

表 1 分割结果对比

Table 1 Comparison of segmentation results

网络结构	HD	Dice	TPR	FPR/10 <sup>-3</sup>
3D U-Net	2.410	0.751	0.771	0.191
3D U-Net + 后处理	2.432	0.759	0.764	0.162
3D Attention U-Net	2.381	0.754	0.779	0.186
3D Attention U-Net + 后处理	2.405	0.764	0.773	0.154
DAU-Net	<b>2.127</b>	0.769	<b>0.797</b>	0.170
DAU-Net + 后处理	2.151	<b>0.774</b>	0.789	<b>0.153</b>

注: 加粗字体为最优值, 网络均为 3 维结构。

在使用基于连通域分析的后处理方法后, 不同网络的 Dice 都有所提升, 3D U-Net 的分割结果从 0.751 提升至 0.759, 3D Attention U-Net 的结果从 0.754 提升至 0.764, DAU-Net 从 0.770 提升至 0.774。同时, 3D U-Net 的 FPR 从 0.191 降低至 0.162, 3D Attention U-Net 的 FPR 从 0.186 降低至 0.154, 而本文提出的 DAU-Net 具有最低的 FPR 值 (0.170), 并通过后处理方法降低至 0.153。在使用 HD 对网络性能进行评估时, 3D Attention U-Net 的结果与 3D U-Net 相似, 分别为 2.381 和 2.410, 而 DAU-Net 的结果为 2.127, 应用后处理方法后 HD 降低至 2.151, 但仍有 9.7% 的提升。

图 5 分别为本文算法、3D Attention U-Net 和 3D U-Net 算法的分割结果图。

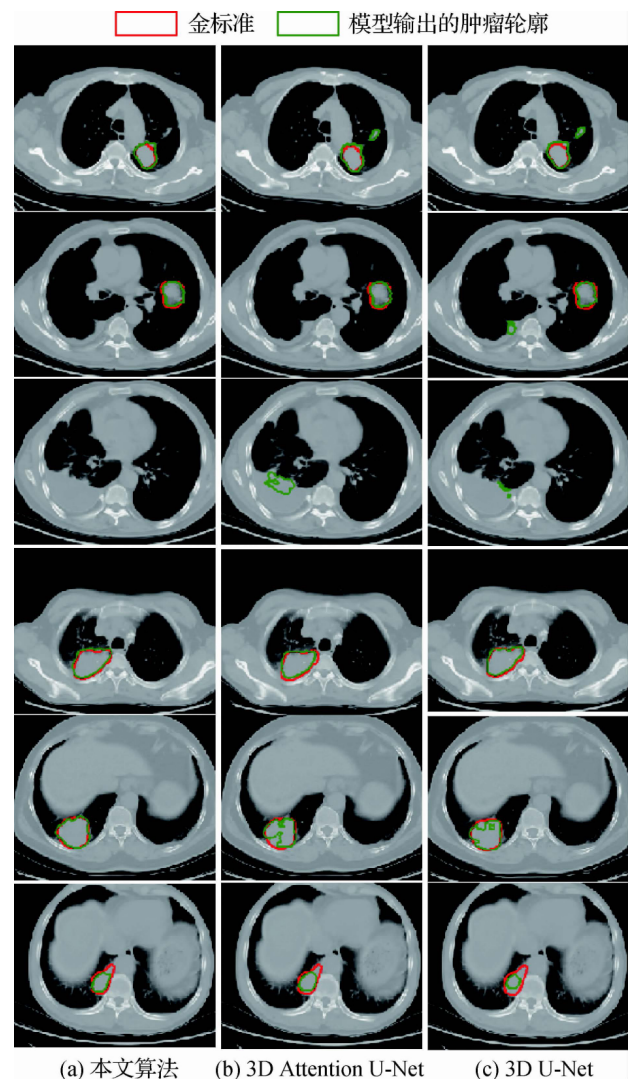


图 5 预测结果对比图

Fig. 5 Comparison of prediction results

((a) ours; (b) 3D Attention U-Net; (c) 3D U-Net)



根据图 5 第 1、2 行所示情况,预测结果与金标准重合度高,不同模型在分割肿瘤方面都具有较高准确度,但 3D Attention U-Net 和 3D U-Net 在同一张切片上的分割得到了两个感兴趣区域,而实际的肿瘤区域只有一个,产生了假阳性,错误地将正常组织作为病灶进行分割,本文算法通过基于连通域分析的后处理,清除了假阳性区域,去除肿瘤区域之外被错误分割的正常组织。此外,由于肿瘤区域分布在连续的 CT 切片上,当使用神经网络进行分割时,假阳性区域还可能会出现在肿瘤区域之外的不同切片上,如图 5 第 3 行中 3D Attention U-Net 和 3D U-Net 的分割结果,原本没有病灶的切片被错误分割,但本文算法减少了这一情况的出现。由图 5 第 4—6 行可以看到,3D Attention U-Net 的分割结果与金标准相差较大。3D U-Net 的分割结果同样较差,而且分割得到的感兴趣区域内含有空洞。与二者相比,DAU-Net 具有最准确的分割结果,与金标准的轮廓有更多的重合。如图 5 第 6 行所示,当肿瘤区域与正常组织的边界不够清晰时,卷积神经网络的效果不够理想,与金标准相比相差较多,但本文模型仍具有更好的性能。

### 3 结 论

本文提出一种融合双注意力机制与残差结构的 3 维卷积神经网络 DAU-Net。该模型具有编码器—解码器结构,能够对使用不同层级的特征进行融合,并通过在收缩路径和扩张路径中间添加位置注意力和通道注意力模块,搜索同一通道内不同距离之间、以及不同通道之间深层语义特征的依赖关系,并对具有依赖关系的特征进行增强,从而提升网络的分割性能。此外,本文采用连通域分析作为后处理方法,清除肿瘤之外被错误分割的假阳性区域,进一步提升分割准确度。实验结果表明,与 3D U-Net 和 3D Attention U-Net 相比,本文模型能够更精确地完成 CT 中肺部肿瘤的分割任务。但本文提出的后处理方法具有一定的局限性,只适用单发的肺肿瘤分割,后续工作将聚焦在多发的肺肿瘤病例,结合实例分割的思想,实现更具普遍性的深度学习算法。

### 参考文献(References)

Chen W Q, Zhang S W and Zou X N. 2010. Estimation and projection of

lung cancer incidence and mortality in China. *Chinese Journal of Lung Cancer*, 13(5): 488-493 (陈万青, 张思维, 邹小农. 2010. 中国肺癌发病死亡的估计和流行趋势研究. *中国肺癌杂志*, 13(5): 488-493) [DOI: 10.3779/j.issn.1009-3419.2010.05.20]

Farag A A, El Munim H E A, Graham J H and Farag A A. 2013. A novel approach for lung nodules segmentation in chest CT using level sets. *IEEE Transactions on Image Processing*, 22(12): 5202-5213 [DOI: 10.1109/TIP.2013.2282899]

Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE: 3146-3154 [DOI: 10.1109/CVPR.2019.00326]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Kanitkar S S, Thombare N D and Lokhande S S. 2015. Detection of lung cancer using marker-controlled watershed transform//*Proceedings of 2015 International Conference on Pervasive Computing*. Pune: IEEE: 1-6 [DOI: 10.1109/PERVASIVE.2015.7087031]

Krizhevsky A, Sutskever I and Hinton G E. 2012. Imagenet classification with deep convolutional neural networks//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe [s. n.]: 1097-1105

Li X M, Chen H, Qi X J, Dou Q, Fu C W and Heng P A. 2018a. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Transactions on Medical Imaging*, 37(12): 2663-2674 [DOI: 10.1109/TMI.2018.2845918]

Li X Y, Xiong J F, Jia T Y, Shen T L, Hou R P, Zhao J and Fu X L. 2018b. Detection of epithelial growth factor receptor (EGFR) mutations on CT images of patients with lung adenocarcinoma using radiomics and/or multi-level residual convolutionary neural networks. *Journal of Thoracic Disease*, 10(12): 6624-6635 [DOI: 10.21037/jtd.2018.11.03]

Lo S C B, Chan H P, Lin J S, Li H, Freedman M T and Mun S K. 1995. Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, 8(7-8): 1201-1214 [10.1016/0893-6080(95)00061-5]

Loshchilov I and Hutter F. 2017. SGDR: stochastic gradient descent with warm restarts [EB/OL]. [2020-06-10]. <https://arxiv.org/pdf/1608.03983.pdf>

Mirdrikvand N, Naderan M and Jamshidnezhad A. 2016. Accurate automatic localisation of lung nodules using graph cut and snakes algorithms//*Proceedings of the 6th International Conference on Computer and Knowledge Engineering*. Mashhad: IEEE: 194-199 [DOI: 10.1109/ICCKE.2016.7802139]

Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich P M, Misawa K,



Mori K, McDonagh S, Hammerla N Y, Kainz B, Glocker B and Rueckert D. 2018. Attention U-Net: learning where to look for the pancreas [EB/OL]. [2020-06-10]. <https://arxiv.org/pdf/1804.03999.pdf>

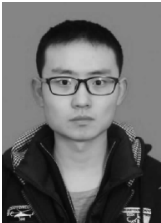
Parveen S S and Kavitha C. 2013. Detection of lung cancer nodules using automatic region growing method//Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies. Tiruchengode: IEEE: 1-6 [DOI: 10.1109/ICCC-NT.2013.6726669]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4\_28]

Wei Y, Chang C X, Jia T and Xu X H. 2009. Segmentation of regions of interest in lung CT images based on 2-D OTSU optimized by genetic algorithm//Proceedings of 2009 Chinese Control and Decision Conference. Guilin: IEEE: 5185-5189 [DOI: 10.1109/CCDC.2009.5195024]

Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2019. UNet + +: redesigning skip connections to exploit multiscale features in image segmentation [EB/OL]. [2020-06-10]. <https://arxiv.org/pdf/1912.05074.pdf>

## 作者简介



郝晓宇, 1996年生, 男, 硕士研究生, 主要研究方向为人工智能与医疗大数据、并行计算。

E-mail: hxy2018@mail.ustc.edu.cn



韩文廷, 通信作者, 男, 副教授, 主要研究方向为并行算法与软件、深度学习算法与编程框架、大数据并行存储与处理。

E-mail: han@ustc.edu.cn

熊俊峰, 男, 博士, 工程师, 主要研究方向为人工智能与医疗大数据。E-mail: jfxiong@sjtu.edu.cn

薛旭东, 男, 博士, 工程师, 主要研究方向为肿瘤放射物理与人工智能在医学中的应用。E-mail: xuexudong511@163.com

石军, 男, 硕士研究生, 主要研究方向为人工智能与医疗大数据。E-mail: shijun18@mail.ustc.edu.cn

文可, 男, 博士研究生, 主要研究方向为人工智能与医疗大数据。E-mail: wenke51@mail.ustc.edu.cn

李骁扬, 男, 博士, 工程师, 主要研究方向为肿瘤放射治疗。E-mail: dxxiaoyangli@163.com

赵俊, 男, 教授, 主要研究方向为医学成像及医学图像处理。E-mail: junzhao@sjtu.edu.cn

傅小龙, 男, 放疗科主任, 主任医师, 主要研究方向为放疗的新技术、时间剂量分割、化疗、分子靶向药物和手术等综合性治疗, 以及功能性影像应用和个体化治疗。

E-mail: 13651635013@126.com