



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象 图形学报

2021

09

VOL.26

ISSN1006-8961
CN11-3758/TB



“深度学习+智慧医疗”专刊



第26卷第9期（总第305期）
2021年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

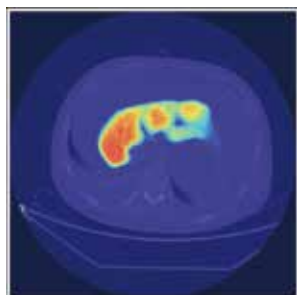
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

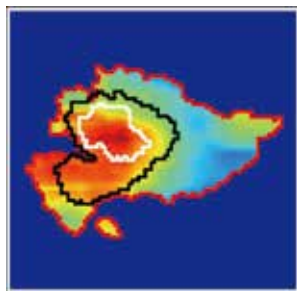
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



肺部影像解剖结构分割数据集及应用(第2111页)



LFSCA-UNet: 基于空间与通道注意力机制的肝纤维化区域分割网络(第2121页)



语义拉普拉斯金字塔多中心乳腺肿瘤分割网络(第2193页)

编者按 I

综述

深度医学图像配准研究进展: 迈向无监督学习

马露凡, 罗凤, 严江鹏, 徐哲, 罗捷, 李秀 2037

U-Net网络医学图像分割应用综述

周涛, 董雅丽, 霍兵强, 刘珊, 马宗军 2058

医学图像深度学习技术: 从卷积到图卷积的发展

唐朝生, 胡超超, 孙君顶, 司马海峰 2078

多尺度变换像素级医学图像融合: 研究进展、应用和挑战

周涛, 刘珊, 董雅丽, 霍兵强, 马宗军 2094

数据集论文

肺部影像解剖结构分割数据集及应用

覃文军, 李小硕, 周庆华, 刘盼, 杨金柱 2111

计算机断层扫描图像

LFSCA-UNet: 基于空间与通道注意力机制的肝纤维化区域分割网络

陈弘扬, 高敬阳, 赵地, 吴忌, 陈金军, 全显跃, 李欣明, 薛峰, 周沐瑶, 柏冰冰 2121

融合上下文和多尺度特征的胸部多器官分割

吉淑滢, 肖志勇 2135

注意力机制下密集空洞卷积的肺部图像分割

郭宁, 柏正尧 2146

融合注意力机制和特征金字塔网络的CT图像肺结节检测

张福玲, 张少敏, 支力佳, 周涛 2156

改进Faster R-CNN模型的CT图磨玻璃密度影目标检测

杨淑莹, 邓东升, 郑清春 2171

融合注意力机制与可变形卷积的多尺度骨病变检测

方成, 柏正尧 2181

磁共振图像

语义拉普拉斯金字塔多中心乳腺肿瘤分割网络

王黎, 曹颖, 郭顺超, 唐雷, 鄧子翔, 王荣品, 王丽会 2193

引入注意力机制和多视角融合的脑肿瘤MR图像U-Net分割模型

罗恺锴, 王婷, 叶芳芳 2208

集成注意力增强和双重相似性引导的多模态脑部图像配准

田梨梨, 程欣宇, 唐堃, 张健, 王丽会 2219

LRUNet: 轻量级脑肿瘤快速语义分割网络

何康辉, 肖志勇 2233

自适应多模态特征融合胶质瘤分级网络

王黎, 曹颖, 田梨梨, 陈祈剑, 郭顺超, 张健, 王丽会 2243

研究应用

面向运动想象脑电图识别的镜卷积神经网络

罗靖, 王耀杰, 刘光明, 王晓帆, 鲁晓峰, 黑新宏 2257

先验权重共享码本下内窥镜图像大肠病变分类

朱霆威, 李胜, 何熊熊 2270

融合空洞卷积与注意力的胃癌组织切片分割

陈颖锲, 李晗, 周雪婷, 万程 2281

视觉显著性的眼底图像视盘检测

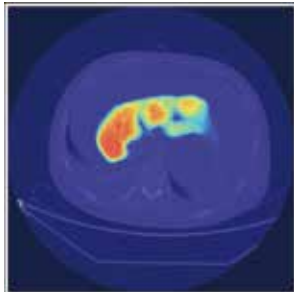
吕鹏飞, 王莹, 王思齐, 于晓升, 吴成东 2293

CONTENTS

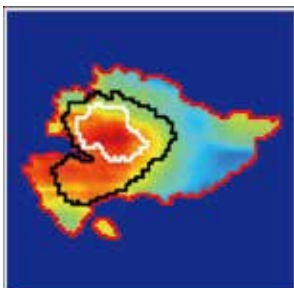
JOURNAL OF IMAGE AND GRAPHICS



Pulmonary image anatomical structure segmentation dataset and applications(P2111)



LFSCA-UNet: liver fibrosis region segmentation network based on spatial and channel attention mechanisms(P2121)



Semantic Laplacian pyramids network for multicenter breast tumor segmentation(P2193)

Review

- Deep-learning based medical image registration pathway: towards unsupervised learning
Ma Lufan, Luo Feng, Yan Jiangpeng, Xu Zhe, Luo Jie, Li Xiu 2037
- U-Net and its applications in medical image segmentation:a review
Zhou Tao, Dong Yali, Huo Bingqiang, Liu Shan, Ma Zongjun 2058
- Deep learning-based medical images analysis evolved from convolution to graph convolution
Tang Chaosheng, Hu Chaochao, Sun Junding, Sima Haifeng 2078
- Research on pixel-level image fusion based on multi-scale transformation: progress application and challenges
Zhou Tao, Liu Shan, Dong Yali, Huo Bingqiang, Ma Zongjun 2094

Dataset

- Pulmonary image anatomical structure segmentation dataset and applications
Tan Wenjun, Li Xiaoshuo, Zhou Qinghua, Liu Pan, Yang Jinzhu 2111

Computerd Tomography Image

- LFSCA-UNet: liver fibrosis region segmentation network based on spatial and channel attention mechanisms
Chen Hongyang, Gao Jingyang, Zhao Di, Wu Ji, Chen Jinjun, Quan Xianxue, Li Xinming, Xue Feng, Zhou Muyao, Bai Bingbing 2121
- Integrated context and multi-scale features in thoracic organs segmentation
Ji Shuying, Xiao Zhiyong 2135
- The integration of attention mechanism and dense atrous convolution for lung image segmentation
Guo Ning, Bai Zhengyao 2146
- Detection of pulmonary nodules in CT images by combining an attention mechanism and a feature pyramid network
Zhang Fuling, Zhang Shaomin, Zhi Lijia, Zhou Tao 2156
- Ground-glass opacity target detection in CT scans based on improved Faster R-CNN model
Yang Shuying, Deng Dongsheng, Zheng Qingchun 2171
- Multi-scale bone lesion detection based on attention mechanism and deformable convolution
Fang Cheng, Bai Zhengyao 2181

Magnetic Resonance Image

- Semantic Laplacian pyramids network for multicenter breast tumor segmentation
WANG Li, CAO Ying, Guo Shunchao, Tang Lei, Kuai Zixiang, Wang Rongpin, Wang Lihui 2193
- U-Net segmentation model of brain tumor MR image based on attention mechanism and multi-view fusion
Luo Kaikai, Wang Ting, Ye Fangfang 2208
- Multimodal brain image registration with integrated attention augmentation and dual similarity guidance
Tian Lili, Cheng Xinyu, Tang Kun, Zhang Jian, Wang Lihui 2219
- LRUNet: a lightweight rapid semantic segmentation network for brain tumors
He Kanghui, Xiao Zhiyong 2233
- Adaptive multi-modality fusion network for glioma grading
Wang Li, Cao Ying, Tian Lili, Chen Qijian, Guo Shunchao, Zhang Jian, Wang Lihui 2243

Research and Application

- Mirror convolutional neural network for motor imagery electroencephalogram recognition
Luo Jing, Wang Yaojie, Liu Guangming, Wang Xiaofan, Lu Xiaofeng, He Xinhong 2257
- Colorectal lesion classification of endoscopic images based on priori weight shared codebook application
Zhu Tingwei, Li Sheng, He Xiongxiang 2270
- The fusing of dilated convolution and attention for segmentation of gastric cancer tissue sections
Chen Yingsi, Li Han, Zhou Xueting, Wan Cheng 2281
- Optic disc detection based on visual saliency in fundus image
Lyu Pengfei, Wang Ying, Wang Siqi, Yu Xiaosheng, Wu Chengdong 2293

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)09-2135-11

论文引用格式: Ji S Y and Xiao Z Y. 2021. Integrated context and multi-scale features in thoracic organs segmentation. Journal of Image and Graphics, 26(09): 2135-2145 (吉淑滢, 肖志勇. 2021. 融合上下文和多尺度特征的胸部多器官分割. 中国图象图形学报, 26(09): 2135-2145) [DOI: 10.11834/jig.200558]

融合上下文和多尺度特征的胸部多器官分割

吉淑滢, 肖志勇*

江南大学人工智能与计算机学院, 无锡 214122

摘要: **目的** 肿瘤周围高危器官的准确分割是图像引导放射治疗中的关键步骤,也是对抗肺癌和食道癌,规划有效治疗策略的重要组成部分。为了解决不同患者之间器官形状和位置的复杂变化情况以及计算机断层扫描(computed tomography, CT)图像中相邻器官之间软组织对比度低等问题,本文提出了一种深度学习算法对胸部CT图像中的高危器官进行细分。**方法** 以U-Net神经网络结构为基础,将冠状面下的3个连续切片序列即2.5D(2.5 dimension)数据作为网络输入来获取切片联系,同时利用高效全局上下文实现不降维的跨通道交互、捕获单视图下切片序列间的长距离依赖关系、加强通道联系和融合空间全局上下文信息。在编码部分使用金字塔卷积和密集连接的集成提取多尺度信息,扩大卷积层的感受野,并将解码器与编码器每层进行连接来充分利用多尺度特征,增强特征图的辨识度。考虑到CT图像中多器官形状不规则且紧密相连问题,加入深度监督来学习不同层的特征表示,从而精准定位器官和细化器官边界。**结果** 在ISBI(International Symposium on Biomedical Imaging)2019 SegTHOR(segmentation of thoracic organs at risk in CT images)挑战赛中,对40个胸部多器官训练样本进行分割,以Dice系数和HD(Hausdorff distance)距离作为主要评判标准,该方法在测试样本中食道、心脏、气管和主动脉的Dice系数分别达到0.855 1、0.945 7、0.923 0和0.938 3,HD距离分别为0.302 3、0.180 5、0.212 2和0.191 8。**结论** 融合全局上下文和多尺度特征的算法在胸部多器官分割效果上更具竞争力,有助于临床医师实现高效的诊断与治疗。**关键词:** 多器官分割;伪三维;高效全局上下文;金字塔卷积;多尺度特征

Integrated context and multi-scale features in thoracic organs segmentation

Ji Shuying, Xiao Zhiyong*

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

Abstract: **Objective** Automatic segmentation of organs at risk (OAR) in computed tomography (CT) has been an essential part of implementing effective treatment strategies to resist lung and esophageal cancers. Accurate segmentation of organs' tumors can aid to interpretate inherent position and morphological changes for patients via facilitating adaptive and computer assisted radiotherapy. Manual delineation of OAR cannot be customized in the future. Scholars have conducted segmentation manually for heart-based backward esophagus spinal and cord-based upper trachea based on intensity levels and anatomical knowledge. Complicated variations in the shape and position of organs and low soft tissue contrast between neighboring organs in CT images have caused emerging errors. The CT-based images for lifting manual segmentation skill for thoracic organs have been caused time-consuming. Nonlinear-based modeling of deep convolutional neural networks (DCNNs) has been presented tremendous capability in medical image segmentation. Multi organ segmentation deep learning

收稿日期:2020-09-15;修回日期:2021-01-21;预印本日期:2021-01-28

* 通信作者: 肖志勇 zhiyong.xiao@jiangnan.edu.cn

基金项目:江苏省优秀青年基金项目(BK20190079)

Supported by: Natural Science Foundation of Jiangsu Province for Excellent Young Scholars(BK20190079)

skill has been applied in abdominal CT images. The small size and irregular shape for automatic segmentation of the esophagus have not been outreached in comparison with even larger size organs. Two skills have related to 3D medical image segmentation have been implemented via the independent separation of each slice and instant 3D convolution to aggregate information between slices and segment all slices of the CT image in. The single slice segmentation skill cannot be used in the multi-layer dependencies overall. Higher computational cost for slices 3D segmentation has been operated via all layers aggregation. A 2.5D deep learning framework has been illustrated to identify the organs location robustly and refine the boundaries of each organ accurately. **Method** This network segmentation of 2.5D slice sequences under the coronal plane composed of three adjacent slices as input can learn the most distinctive of a single slice deeply. The features have been presented in the connection between slices. The image intensity values of all scans were truncated to the range of $[-384, 384]$ HU to omit the irrelevant information in one step. An emerging attention module called efficient global context has been demonstrated based on the completed U-Net neural network structure. The integration for effective channel attention and global context module have been achieved. The global context information has been demonstrated via calculating the response at a location as the weighted sum of the features of all locations in the input feature map. A model has been built up to identify the correlation for channels. The effective feature map can obtain useful information. The useless information can be deducted. The single view long distance dependency between slice sequences can be captured. Attention has been divided into three modules on the aspect of context modeling module, feature conversion module and fusion module. Unlike the traditional global context module, feature conversion module has not required dimensionality to realize the information interaction between channels. The channel attention can be obtained via one dimensional convolution effectively. The capability of pyramid convolution has been used in the encoding layer part. Extracted multi-scale information and expanded receptive field for the convolution layer can be used via dense connection. The pyramid convolution has adapted convolution kernels on different scales and depths. The increased convolution kernels can be used in parallel to process the input and capture different levels of information. Feature transformation has been processed uniformly and individually in multiple parallel branches. The output of each branch has been integrated into the final output. Multi-scale feature extraction based on adjusting the size of the convolution kernel has been achieved than the receptive field resolution down sampling upgration. Multi-layer dense connection has realized feature multiplexing and ensures maximum information transmission. The integration of pyramid convolution and dense connection has obtained a wider range of information and good quality integrated multi-scale images. The backward gradient flow can be smoother than before. An accurate multi-organs segmentation have required local and global information fusion, decoder with each layer of encoders connecting network and the low level details of different levels of feature maps with high level semantics in order to make full use of multi-scale features and enhance the recognition of feature maps. The irregular and closely connected shape of multi-organs in CT images can be calculated at the end. Deep supervision has been added to learn the feature representations of different layers based on the sophisticated feature map aggregation. The boundaries of organs and excessive segmentation deduction in non-organ images and network training can be enhanced effectively. More accurate segmentation results can be produced finally. **Result** In the public dataset of the segmentation of thoracic organs at risk in CT images (SegTHOR) 2019 challenge, the research has been performed CT scans operation on four thoracic organs (i.e., esophagus, heart, trachea and aorta), take Dice similarity coefficient (DSC) and Hausdorff distance (HD) as main criteria, the Dice coefficients of the esophagus, heart, trachea and aorta in the test samples reached 0.855 1, 0.945 7, 0.923 0 and 0.938 3 separately. The HD distances have achieved 0.302 3, 0.180 5, 0.212 2 and 0.191 8 respectively. **Conclusion** Low level detailed feature maps can capture rich spatial information to highlight the boundaries of organs. High level semantic features have reflected position information and located organs. Multi scale features and global context integration have been the key step to accurate segmentation. The highest average DSC value and HD obtained for heart and Aorta have achieved its high contrast, regular shape, and larger size compared to the other organs. The esophagus had the lowest average DSC and HD values due to its irregularity and low contrast to identify within CT volumes more difficult. The research has achieved a DSC score of 85.5% for the esophagus on test dataset. Experimental results have shown that the proposed method has beneficial for segmenting high risk organs to strengthen radiation therapy planning.

Key words: multi-organ segmentation; pseudo three dimension; efficient global context; pyramid convolution; multi-scale features

0 引言

放射线治疗是治疗胸部癌症(肺癌或食道癌等)的有效手段。放射治疗的成功取决于射线照射肿瘤及其周围高危器官的剂量,为了避免过量照射周围正常组织造成放射性损伤,准确分割计算机断层扫描(computed tomography, CT)图像中的高危器官是关键任务。传统方法是专家手动勾勒器官,费时且烦琐,因此精确分割多器官的自动化方法成为医学图像研究中的热点。

CT图像的自动分割有助于诊断CT图像中有风险的胸腔器官,但存在三方面的挑战:1)医学图像是3维的,需要处理的数据量和计算量都很大;2)一些器官紧密相邻,CT图像中的轮廓对比度低;3)不同患者的某些器官的形状和位置差异很大。早期的多器官分割是基于区域生长的分割算法(Wu等,2008;Zhou和Bai,2007),这是一种半自动分割方法,依赖种子点的正确位置获得了强大性能。随着卷积神经网络在图像处理和分割中的优异表现,逐步用于医学图像分割,经典方法是Ronneberger等人(2015)提出的U-Net网络,该网络上采样层和下采样层之间的跳跃连接可以使下采样层提取的特征直接传递到上采样层,从而得到精准的像素定位,提升分割性能,但是U-Net网络不能获取切片之间的相关信息,容易出现欠分割。近年来,陆续提出了许多编码器—解码器体系结构的相似模型,例如U-Net++(Zhou等,2018)和U-Net 3+(Huang等,2020)等。U-Net++通过引入嵌套和密集的跳跃连接,将深层与浅层特征进行融合,减小编码器与解码器之间的语义差距,但是不能从全面的范围探索足够的特征。为了获取足够的空间信息,Milletari等人(2016)将2D U-Net改进为3D V-Net,直接对体素进行处理来充分挖掘上下文信息。Li等人(2018)提出了一种用于肝和肿瘤的H-DenseUNet分割,提取切片内和切片间特征,并通过混合特征融合层共同优化。Dou等人(2016)提出了3D深度监督网络(deeply-supervised net, 3D-DSN)来解决肝脏分割问题。3D-DSN涉及向隐藏层注入额外的监督,以抵消逐渐消失的不利影响。

3D卷积神经网络(3D convolutional neural network, 3D-CNN)模型能够充分利用空间相关特征,但

训练耗时且参数过多,计算资源需求大。2D-CNN模型较为轻量且收敛,但不能充分考虑切片间的相关性。为解决这一问题,本文选取2.5D(dimension)的数据,由3张相邻的切片作为输入组成,不仅可以学习单张切片的最具区别性的特征,也能考虑到切片之间的联系,获取空间上下文信息。同时基于Cao等人(2019)提出的全局上下文模块进行改进,得到一个新的注意力模块,通过不降维来对特征图的空间和通道添加注意力,增加重要特征的权重,同时减弱不重要特征的影响。在解码器层,每层添加了金字塔卷积,捕获低分辨率和高分辨率特征图的多尺度信息,增加感受野,减少特征损失。不同于U-Net同尺度的特征图融合,网络中每一个解码器层都融合了来自编码器中的小尺度和同尺度的特征图,将多尺度下的细粒度语义和粗粒度语义结合,增强切片间相关性并规范化切片间预测,以使其更加连贯。本文方法在ISBI(International Symposium on Biomedical Imaging)2019胸腔高危器官分割Seg-THOR(segmentation of thoracic organs at risk in CT images)挑战赛中表现优异(Trullo等,2019)。

1 本文方法

1.1 全局上下文模型回顾

在卷积神经网络中,传统的卷积层只能在局部领域内建立像素关系,其长距离依赖关系需要通过深度叠加卷积层来获取,但是不断地加深网络会导致计算量大,优化困难。全局上下文网络(global context network, GCNet)采用压缩激励模块建立通道尺寸之间的相互依赖关系,引入空间注意力机制来建立远程依赖关系,并从全局角度有效地对全局上下文建模,更好地捕捉多器官特征,其网络结构如图1所示。

给定长、宽和通道数分别为 W 、 H 和 C 的特征图 $F \in \mathbf{R}^{C \times H \times W}$,首先经过上下文建模模块,通过简化版自注意力机制,逐元素相乘将所有位置的特征结合在一起,形成全局上下文特征 $X \in \mathbf{R}^{C \times 1 \times 1}$;再通过特征转化模块捕获各通道之间的相互依存关系,加强有用信息,压缩无用信息。中间层用 1×1 卷积降低维数,得到特征图 $Y \in \mathbf{R}^{C/r \times 1 \times 1}$, r 表示压缩比例,取 $r = 16$;最后经过融合模块逐元素相加,将全局上下文特征合并到原先的特征图中,得到与输入有相同维度的特征图 $F_c \in \mathbf{R}^{C \times H \times W}$,具体为

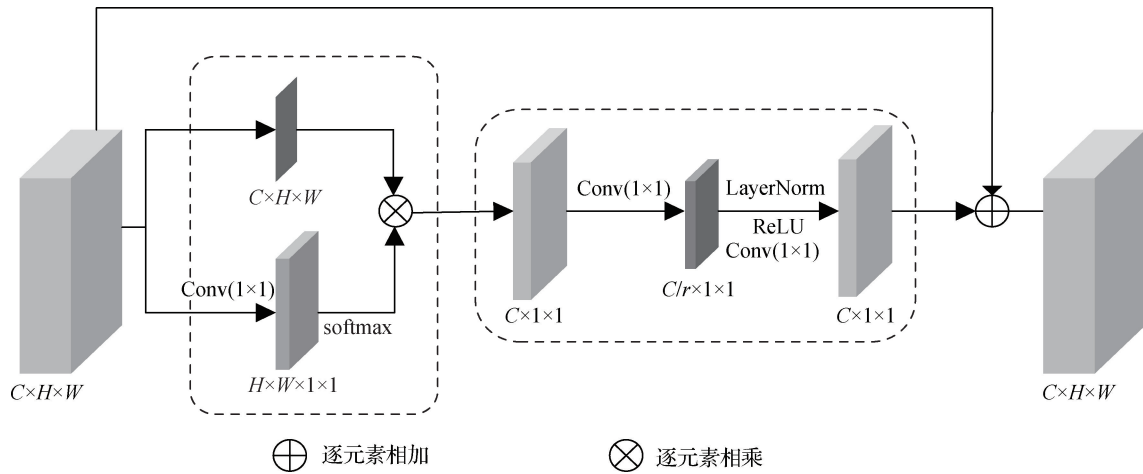


图1 全局上下文网络结构

Fig. 1 GCNet structure diagram

$$F_c = F + W_{v2} \sigma \left(W_{v1} \sum_{j=1}^{N_p} \frac{\exp(W_k x_j)}{\sum_{m=1}^{N_p} \exp(W_k x_m)} x_j \right) \quad (1)$$

式中, $N_p = H \times W$, σ 表示层标准化(layer normalization, LN)和 ReLU 激活函数, x_j 代表特征图中查询点的位置, 得到全局注意力池化的总权重, W_k 、 W_{v1} 和 W_{v2} 是经过 1×1 卷积后的特征权重向量。

1.2 高效全局上下文模型

GCNet 是 Non-local (Wang 等, 2018) 和 SENet (Hu 等, 2018) 的结合, 同时考虑了通道注意和空间注意, 本文对其通道注意进行改进, 空间注意仍遵循 GCNet。GCNet 中通道注意是通过降维来捕捉非线性的跨通道交互, 而 Wang 等人 (2020) 提出的高效通道注意力网络 (efficient channel attention network, ECANet) 表明避免降维和适当的跨通道交互对于学习高性能和高效率的通道注意力是重要的。

本文引入 ECANet 来改进 GCNet, 得到一个新的注意力网络为高效全局上下文网络 (efficient global context network, EGCNet), EGCNet 保留了 GCNet 捕获长距离依赖的优秀特性, 同时引入了 ECANet 的不降维跨通道交互学习的特点, 更好地获取全局和局部信息, 其跨通道信息交互模块如图 2 所示。

EGCNet 转化模块中不进行降维, 直接通过大小为 k 的快速 1 维卷积捕获局部跨通道交互信息, k 代表局部跨信道交互的覆盖率, 即通道附近有多少个邻居参与这个信道的注意力预测。通过高效信道注意力的通道权重的特征向量公式为

$$w = \sigma(C^{1D}_k(y)) \quad (2)$$

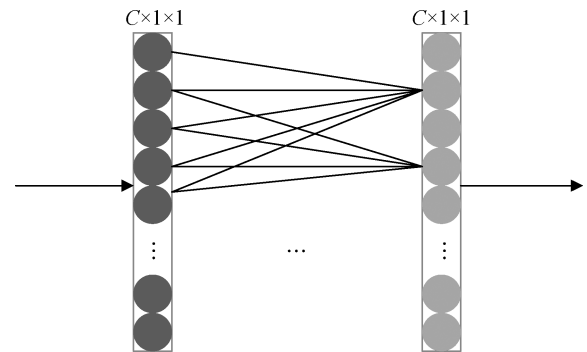


图2 跨通道信息交互模块

Fig. 2 Cross-channel information exchange module

式中, C^{1D} 代表 1 维卷积, y 为输入的特征, σ 是 sigmoid 函数, k 为涉及的参数信息, 对 k 进行手动调参, 实验表明, k 取 3 时达到最优效果。

1.3 金字塔卷积模型

理论上 CNN 网络架构具有非常大的感受野, 但实际上小得多, 且通过卷积或池化操作增加感受野都会造成信息损失, 无法捕获足够的高级语义, 导致特征图辨识度较低, 影响模型性能 (Zhou 等, 2015)。本文使用 Duta 等人 (2020) 提出的金字塔卷积 (pyramidal convolution, PyConv) 模型, 包含不同大小和深度的卷积核, 除增加感受野外, 可以并行使用增加的卷积核处理输入, 捕获不同级别的信息, 且保持与标准卷积相似的计算参数。标准卷积与金字塔卷积的对比如图 3 所示。

图 3(a) 是标准卷积示意图, 包含单一类型的核, 卷积核的空间分辨率为 K^2 , 深度等于输入特征图的通道数 C_i , 执行 FM_0 个相同空间分辨率和深度

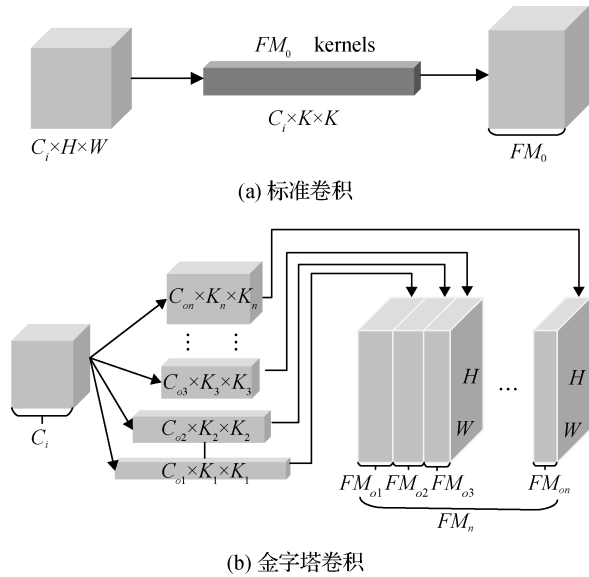


图3 不同卷积对比图

Fig. 3 Comparison of different convolutions

((a) standard convolution; (b) pyramidal convolution)

的卷积核得到深度为 FM_0 的输出特征。卷积特征变换中,每个空间位置的视野主要由预定义的内核大小控制,缺少大的感受野。图3(b)是PyConv示意图,包含一个由 n 层不同类型的卷积核构成的金字塔。在PyConv的每个级别,卷积核包含不同的空间分辨率,从金字塔的底部到顶部逐层增加内核大小为 $K_0, K_1, K_2, \dots, K_n$,随着空间大小的增加,卷积核的深度从底部到顶部逐层降低为 $C_{o1}, C_{o2}, C_{o3}, \dots, C_{on}$,输入特征 FM_i 经过不同的卷积核得到输出特征的深度 $FM_{o1}, FM_{o2}, FM_{o3}, \dots, FM_{on}$,拼接成最终的输出特征的深度 FM_n 。为了在PyConv的每个级别上使用不同深度的卷积核,应用了分组卷积,将输入特征图分为不同的组,特征变换在多个并行分支中均匀、单独地执行处理,每个分支的输出合并为最终输出,增强了卷积核之间的连通性。

这种双重定向的金字塔内核类型,在增加卷积核大小的同时,减少卷积核的深度,提供了非常多样的集成网络在学习过程中可以探索的内核类型,不同类型的内核带来了更多的补充信息,有助于提高网络的识别性能。

1.4 深度监督

本文分割胸部中的食管、心脏、主动脉和气管等器官,这些器官紧密相连,软组织对比度低,在网络训练时容易出现分类错误。为解决该问题,在解码层的每一层添加深度监督,即添加额外的损失进行

反向传播。辅助损失注入中间层能正规化网络,减轻梯度消失,降低中底层有效特征的提取和训练难度,使分类错误最小化,提升分割性能。

深度监督流程如图4所示,可以看出,特征图 M 经过4步编码/解码得到特征图 M_0, M_1, M_2, M_3 ,得到的每个特征图通过上采样从低分辨率恢复到高分辨率,使用 softmax 函数得到分割概率图 $Z \in \mathbb{R}^{5 \times H \times W}$,数字5为通道数,表示背景、食管、心脏、主动脉和气管的分割结果。每个概率图都与真实标签(ground truth, GT)计算损失,分别得到4个 loss 值,最后按系数求和得到总的 loss 值。

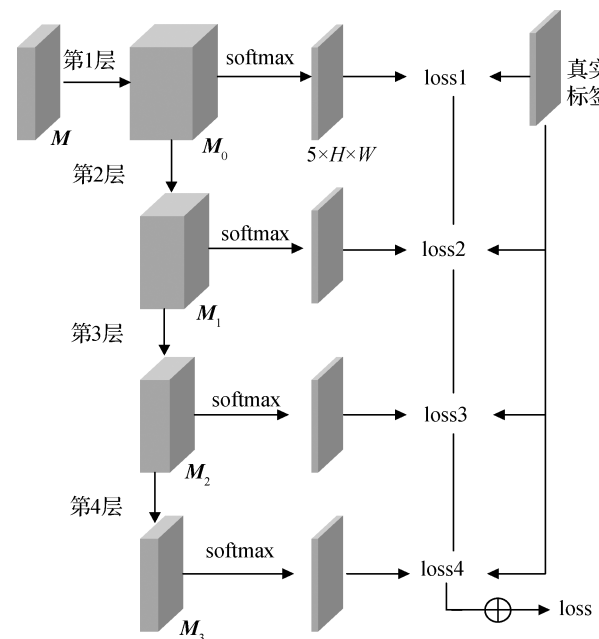


图4 深度监督流程图

Fig. 4 Flow chart of depth supervision

1.5 网络模型

为了准确进行多器官分割,本文在U-Net的基础上设计了一个新的多尺度融合型类U结构,主要包括编码部分、多尺度连接和解码部分。网络结构如图5所示。编码部分包含4个下采样块和1个输入块,输入块由两个 3×3 卷积核组成,在每个卷积层之后添加批标准化(batch-normalization, BN)(Ioffe和Szegedy, 2015)和ReLU激活函数。为了获取多尺度信息,在下采样块中,每个卷积块联合使用PyConv和密集连接提取多尺度特征,增加感受野,PyConv添加在两个卷积层中间,同样在每个卷积层之后添加BN和ReLU激活函数,缓解梯度爆炸和加速网络收敛。同时,在每个卷积块中加入密集连接

模块(Huang等,2017),对采集的特征重复使用,减少特征损失。下采样采用 2×2 的最大池化,步长为2,下采样块的结构如图6所示。由于编码层在下采样过程会损失空间信息和位置信息,为了从全面范围探索足够信息,本文在解码器和对应的编码器及其以上位置都添加跳跃连接,获取编码器的片间联系和解码器的片内联系,多尺度捕获细粒度的细节信息和粗粒度的语义信息。在跳跃连接层中添加EGC(efficient global context),该模块计算每个像素点特定的全文信息,自动获取每个特征通道的重要程度,提升有用特征并抑制无用特征,对整个输入特

征图进行上下文建模,感受野可以覆盖整个特征图。同时在跳跃层添加残差连接,确保向后传递梯度流更加平滑,在一定程度上缓解了网络因层数过深导致的退化问题。解码部分包含4个上采样卷积块和1个输出块,上采样操作使用 2×2 的反卷积和两个 3×3 的卷积,每个卷积层后都有BN和ReLU操作,最终得到与输入图像相同分辨率的特征图。最后输出块使用 1×1 的卷积核与softmax分类器获得多器官的分割结果。在每个上采样块中加入深度监督,使网络的前中期可以对分割器官的类别进行有效判定,增强识别能力。

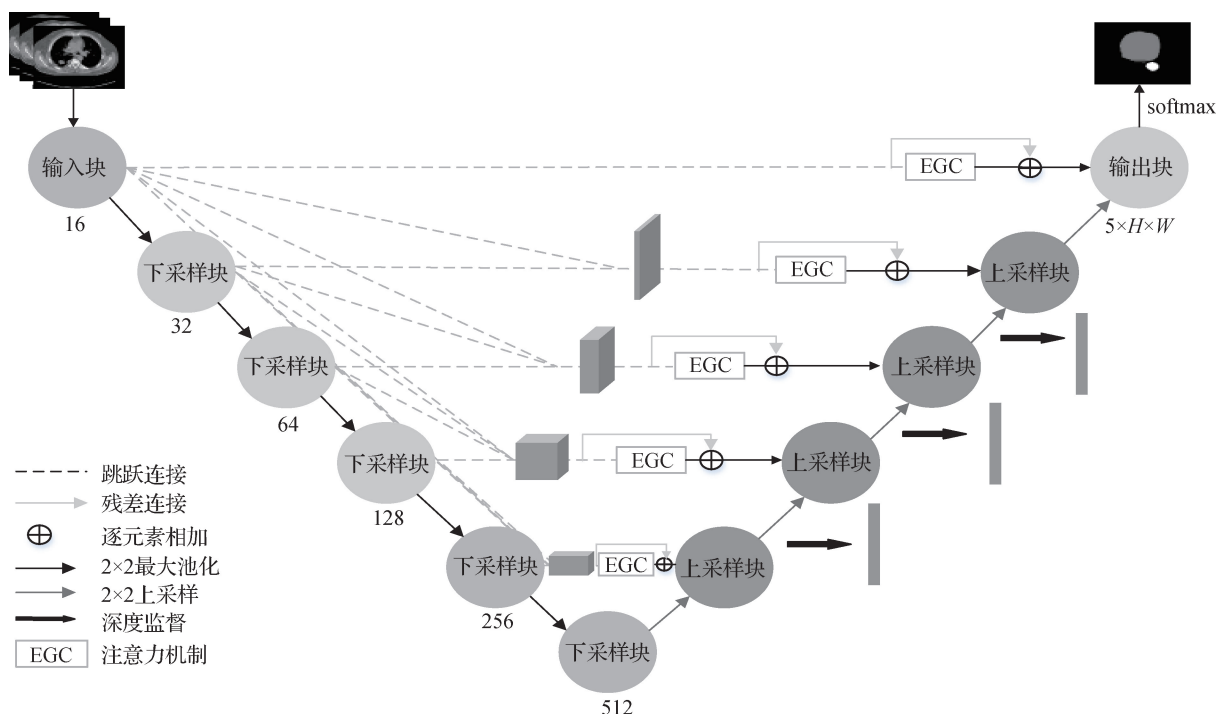


图5 网络结构图

Fig. 5 The network structure diagram

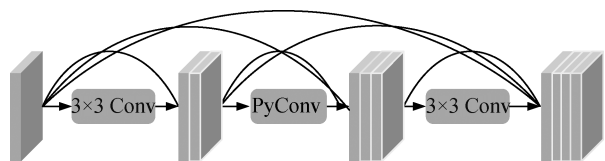


图6 下采样块结构

Fig. 6 Down block structure

1.6 损失函数

本文多器官分割中,食管体积较小,占据图像的小部分。网络会过多地考虑背景体素,而忽略前景体素,易出现类别不平衡现象。Dice损失是基于区域的损失函数,仅关注前景,不考虑整幅图像中的背

景体素,比交叉熵损失(cross entropy loss, CE Loss)有明显优势。本文分割背景、食管、心脏、气管和主动脉,对每个器官分别求其Dice损失并取平均,其定义为

$$DL = \frac{1}{C} \sum_{c=1}^C \left(1 - \frac{2 \sum_i p_c(i) g_c(i)}{\sum_i p_c^2(i) + \sum_i g_c^2(i)} \right) \quad (3)$$

式中, C 代表器官的类别数,取 $C = 5$, N 代表每幅图像的体素点个数, $p_c(i)$ 代表每个体素中器官类别为 c 的预测概率值, $g_c(i)$ 代表每个体素中器官类

别为 c 的真实值 (GT)。

网络训练过程采用深度监督进行模型优化,将解码部分每层得到的预测结果与 GT 比较,计算得到各层损失,最终损失是各层损失按一定比例相加的结果。每层损失均按 DL 进行计算,最终的损失定义为

$$L(Y, \hat{Y}) = \alpha \sum_{b=1}^3 DL_b + DL_4 \quad (4)$$

式中, $\sum_{b=1}^3 DL_b$ 是网络前 3 层的 Dice 损失总和, DL_4 是网络最终层的 Dice 损失值, α 是取前 3 层损失值的比例权重,取值为 0.33。

2 实验结果与分析

2.1 实验数据集

实验使用 ISBI SegTHOR 挑战赛数据集, (<https://competitions.codalab.org/competitions/21145>), 该数据集专注于食管、心脏、主动脉和气管等高危器官,共 60 幅 CT 图像,其中 40 幅用于训练,20 幅用于测试。用于训练的每幅图像包含 4 个器官的真实标签,即手动分割结果。每个 CT 扫描的面内尺寸为 512×512 像素,切片数量在 150 ~ 284 之间。实验使用 4 折交叉验证方法,网络模型对 20 例测试数据的测试结果在 SegTHOR 线上提交。

实验训练数据集较少,网络训练易出现过拟合问题,本文对所有训练数据进行水平和垂直翻转,按 60% 比例进行缩放实现数据扩充。为了减少计算量和内存空间,保留感兴趣区域 (region of interest, ROI) 作为网络输入,每幅图像沿 x 轴和 y 轴中心裁剪为 400×400 像素, z 轴保留相同数量的切片。考虑到器官的唯一性和连贯性,对预测结果进行后处理操作,使用最大连接算法,删除器官周围较小区域,填充器官内部不平滑区域。

2.2 参数设置

实验硬件环境为 NVIDIA RTX 2080 Ti GPU, Intel Core i7 处理器,基于 Python 的 Pytorch 库实现。网络采用 kaiming_normal (He 等, 2015) 进行权重初始化,随机梯度下降法 (stochastic gradient descent, SGD) 作为网络反向传播优化器,参数学习率 (learning rate, lr) 为 0.01, 动量 (momentum) 为 0.9, 权重衰减 (weight-decay) 为 0.000 01, 训练的批大小

(batch size) 取 4。考虑到训练数据较少,为了更好地验证网络模型的性能,使用 4 折交叉验证方法取平均值作为最后的实验结果。

2.3 评价指标

使用 Dice 相关性系数 (Dice similarity coefficient, DSC) 和豪斯多夫距离 (Hausdorff distance, HD) 评估细分的准确性。

DSC 指标用来衡量手动分割与自动细分之间的重叠度,两者重合度越高,分割效果越好,是评估细分准确性广泛使用的指标,其定义为

$$DSC(T, P) = \frac{2|P \cap T|}{|P| + |T|} \quad (5)$$

HD 指标是描述自动分割组中的点到手动分割组中对应最接近的点的最大距离,单位为 mm,其定义为

$$HD(T, P) = \max_{t \in T} (\max_{p \in P} (\sqrt{T^2 - P^2})) \quad (6)$$

式中, T 表示手动分割的真实结果, P 表示算法预测的分割结果。

2.4 实验结果

2.4.1 样本处理对实验结果的影响

由于 SegTHOR 数据集中大多数 CT 图像的对比度较低,因此进行预处理提高精度。依据医学图像理论,将所有 CT 图像的 Hu 值截断到 $[-384, 384]$ 范围内,并对图像的均值和方差归一化,增加对比度,忽略不相关信息。预处理前后器官分割的 DSC 和 HD 指标的结果对比如表 1 所示。可以看出,未处理的各项指标的精度均较差,因为未处理的图像对比度低,背景噪声较多,从而产生错误的分割结果。处理后,不同患者的图像对比度一致,网络得到较好的学习判别能力,预测的结果精度得到很大提升。

表 1 预处理与未处理的分割结果比较
Table 1 Comparison of segmentation results
between preprocessed and unprocessed

器官	未处理		预处理	
	DSC	HD/mm	DSC	HD/mm
食管	0.704 9	0.534 1	0.855 1	0.302 3
心脏	0.900 3	1.003 1	0.945 7	0.180 5
气管	0.842 5	0.458 9	0.923 0	0.212 2
主动脉	0.897 2	0.559 3	0.938 3	0.191 8

注:加粗字体表示各项最优结果。

为验证模型的分割性能,选取前 100 个 epoch 验证损失,如图 7 所示。可以看出,未处理的图像损失波动较大,训练不稳定,损失值大于处理过的图像,预测结果较差。处理过的图像在经历约 20 个 epoch 训练后,损失函数值趋于平稳,变化不明显,训练稳定。

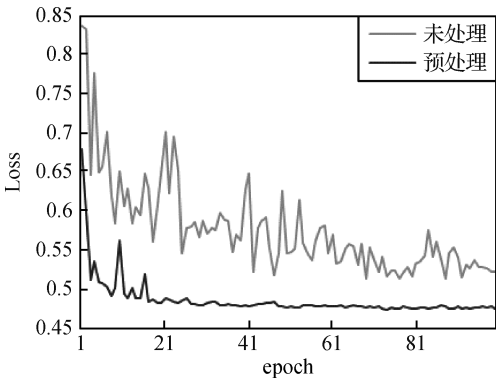


图 7 验证损失对比图

Fig. 7 Comparison of validation loss

2.4.2 高效全局上下文模型对实验结果的影响

在高效全局上下文模型中,卷积核大小为 k 的快速 1 维卷积代表局部跨信道交互的覆盖率, k 值的选取一定程度上影响了模型性能,分别对 k 取 3、5、7、9 进行测试,结果如表 2 所示。可以看出,随着领域参数 k 值的变化,平均 DSC 和 HD 在 $k = 3$ 时达到最优的结果,其他值虽然相似性系数变化不大,但是预测结果与真实结果的边缘差距较高。出现这一现象的原因在于训练的模型层次较深时,卷积核增大,使计算量增加,计算性能也会降低。

表 2 领域参数选取结果比较

Table 2 Comparison of field parameter results

领域参数 k	平均 DSC	平均 HD/mm
3	0.915 8	0.221 7
5	0.906 0	0.280 8
7	0.902 8	0.364 1
9	0.907 2	0.378 5

注:加粗字体表示各列最优结果。

同时为了验证 ECANet 对 GCNet 模块改进的有效性,将本文网络中的 EGC 模块替换为 GC,其他部分保持不变,进行对比实验,结果如表 3 所示。可以看出,加入 EGC 模块的分割精度整体高于 GC 模

块,其中 HD 指标更为明显,可见不降维的跨通道交互能有效结合空间和通道的注意力,充分获取局部特征信息和全局上下文信息,器官分割边界得到精准细化。

表 3 GC 和 EGC 的结果比较

Table 3 Comparison of results between GC and EGC

器官	GC		EGC	
	DSC	HD/mm	DSC	HD/mm
食管	0.843 0	0.413 3	0.855 1	0.302 3
心脏	0.944 6	0.184 9	0.945 7	0.180 5
气管	0.905 7	0.300 3	0.923 0	0.212 2
主动脉	0.919 1	0.552 7	0.938 3	0.191 8

注:加粗字体表示各项最优结果。

2.4.3 网络模型对实验结果的影响

为了验证基于 U-Net 的改进策略对分割性能的影响,将多个改进策略进行对比,实验结果如表 4 所示。可以看出,U-Net 对处理后的图像进行分割,心脏和主动脉因为具有较高的对比度、规则的形状和较大的尺寸,获得了较优的 DSC 和 HD;食管由于不规则性和低对比度,在 CT 图像中难以识别,DSC 和 HD 值表现更美。器官的边缘信息受到损失,分割不连续且器官周围出现较小的分割器官,出现过分割和欠分割现象。在 U-Net 基础上加上深度监督 (deeply-supervised, DS),通过获取多个尺度下的损失,引导网络对有效特征提取,气管获得了较高的评估结果。在此基础上再分别加入 EGC 模块、金字塔卷积和一个解码与所有编码之间的跳跃连接构成的多尺度融合 (multi-scale fusion, MF) 模块。MF 模块将高分辨率的局部信息和低分辨率的全局信息充分融合,从而探索足够信息。EGC 模块通过不降维的跨信道交互策略提高通道注意力学习。两种模型的分割结果明显优于未添加的模型,EGC 模块的效果更加明显,证实了 EGC 模型的有效性。

图 8 是样本在不同模型下得到的分割结果,选取了验证集中图像对比度和器官结构均不同的 4 位患者进行对比。绿色、红色、蓝色和黄色分别代表心脏、食道、气管和主动脉区域,为了可视化分割结果,选择矢状面作为对比面。本文模型的实验结果分割平滑,没有出现较明显的欠分割和过分割现象,多个器官的分割结果最接近于真实标签。

表 4 各种网络结构的分割结果比较

Table 4 Comparison of segmentation results of various network structures

模型	DSC				HD/mm			
	食管	心脏	气管	主动脉	食管	心脏	气管	主动脉
U-Net	0.833 1	0.925 0	0.903 7	0.908 0	0.569 0	0.288 2	0.743 8	0.666 7
U-Net + DS	0.821 5	0.916 6	0.910 1	0.898 6	0.673 2	0.419 9	0.258 6	1.296 5
U-Net + MF + DS	0.837 6	0.937 2	0.909 9	0.920 6	0.416 6	0.269 7	0.313 5	0.307 4
U-Net + EGC + DS	0.853 4	0.941 7	0.918 6	0.924 7	0.394 3	0.205 2	0.202 8	0.272 9
U-Net + EGC + DS + MF	0.855 1	0.945 7	0.923 0	0.938 3	0.302 3	0.180 5	0.212 2	0.191 8

注:加粗字体表示各列最优结果。

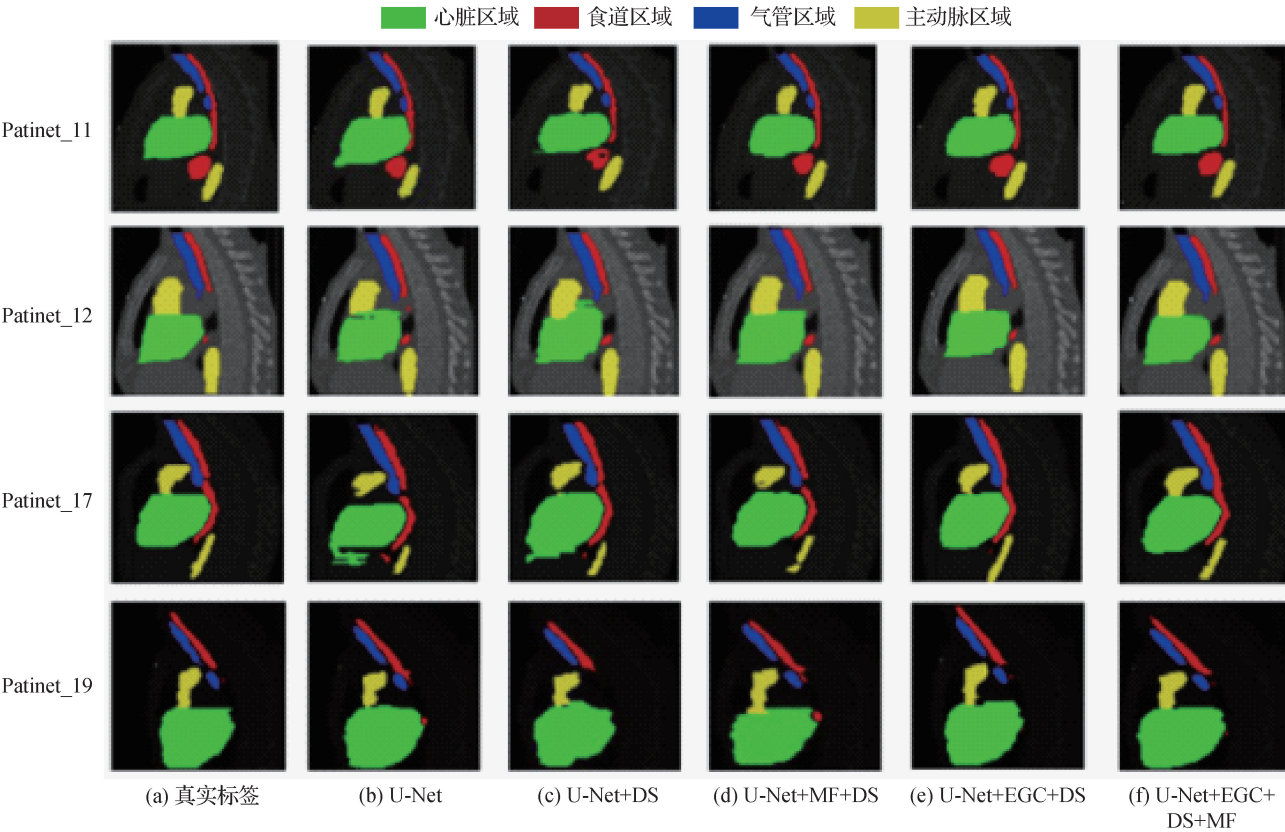


图 8 不同模型的分割结果

Fig. 8 Segmentation results on different models

((a) ground truth; (b) U-Net; (c) U-Net + DS; (d) U-Net + MF + DS; (e) U-Net + EGC + DS; (f) U-Net + EGC + DS + MF)

2.4.4 与其他算法的比较

为进一步验证本文方法的性能,在 SegTHOR 数据集上与其他分割方法进行对比,结果如表 5 所示。Kim 等人(2019)采用级联的 2D 网络,对多视图进行集成,分割心脏等大器官精度较高,但对于食管等小器官精度较低,3 维数据 z 轴包含大量空间信息,多视图集成需裁剪切片大小,可能丢失切片之间的

联系,导致分割效果差。Zhang 等人(2019)和 Chen 等人(2019)都采用先粗分割再细分割的方法,对 4 个器官先进行定位再基于位置进行精确分割,其中定位是关键,如果定位精度较低,则分割结果会很差。本文方法充分利用了相邻切片并学习特定切片的信息来帮助识别不明显器官。实验结果表明,本文方法取得了较好的细分结果。

表 5 不同分割算法的比较

Table 5 Comparison of segmentation results on different models

方法	DSC				HD/mm			
	食管	心脏	气管	主动脉	食管	心脏	气管	主动脉
Kim 等人(2019)	0.751 8	0.932 8	0.888 5	0.891 9	0.926 7	0.218 4	0.888 5	1.130 0
Zhang 等人(2019)	0.773 2	0.938 4	0.893 9	0.923 2	1.677 4	0.208 9	0.274 1	0.308 1
Chen 等人(2019)	0.816 6	0.932 9	0.891 0	0.923 2	0.491 4	0.241 7	0.274 6	0.308 1
本文	0.855 1	0.945 7	0.923 0	0.938 3	0.302 3	0.180 5	0.212 2	0.191 8

注:加粗字体表示各列最优结果。

3 结 论

针对医学上胸部多器官形状位置复杂和目标区域分割精度低的问题,从广泛使用的 U-Net 中抽象设计出类 U 的结构来自动分割食管、心脏、主动脉和气管等胸部器官。采用高效通道注意模块来改进全局上下文模块,从空间和通道上,建立了有效的特征图远程依赖关系。为了扩大感受野,捕获更多图像的细节特征,在编码层将金字塔卷积和密集连接进行集成。此外,通过编码器和所有解码器之间完整的跳跃连接以及深入的监督功能,将不同级别的特征图的底层细节和高层语义融合在一起。实验结果表明,提出的方法具有优越性,得到了准确高效的细分结果,精度值相对于其他对比方法,具有较大提升。

食管细小且狭长,周围边界模糊,分割效果受到局限,今后将考虑使用边缘检测方法来改善食管等小物体的分割性能,提高胸腔器官的分割结果。同时,由于训练样本数量有限,在未来的工作中,将使用更多的数据来评估本文方法的泛化能力。

参考文献 (References)

- Cao Y, Xu J R, Lin S, Wei F Y and Hu H. 2019. GCNet: non-local networks meet squeeze-excitation networks and beyond//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE: #00246 [DOI: 10.1109/ICCVW.2019.00246]
- Chen P, Xu C H, Li X Y, Ma Y Y and Sun F L. 2019. Two-stage network for OAR segmentation [EB/OL]. [2020-8-31]. http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_4.pdf
- Dou Q, Chen H, Jin Y M, Yu L Q, Qin J and Heng P A. 2016. 3D deeply supervised network for automatic liver segmentation from CT volumes//Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer: 149-157 [DOI: 10.1007/978-3-319-46723-8_18]
- Duta I C, Liu L, Zhu F and Shao L. 2020. Pyramidal convolution: rethinking convolutional neural networks for visual recognition [EB/OL]. [2020-08-31]. <https://arxiv.org/pdf/2006.11538.pdf>
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1026-1034 [DOI: 10.1109/ICCV.2015.123]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 4700-4708 [DOI: 10.1109/CVPR.2017.243]
- Huang H M, Lin L F, Tong R F, Hu H J, Zhang Q W, Iwamoto Y, Han X H, Chen Y W and Wu J. 2020. UNet 3+: a full-scale connected UNet for medical image segmentation//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona, Spain: IEEE: 1055-1059 [DOI: 10.1109/ICASSP40776.2020.9053405]
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. [2020-08-31]. <https://arxiv.org/pdf/1502.03167.pdf>
- Kim S, Jang Y, Han K, Shim H and Chang H J. 2019. A cascaded two-step approach for segmentation of thoracic organs [EB/OL]. [2020-08-31]. http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_3.pdf
- Li X M, Chen H, Qi X J, Dou Q, Fu C W and Heng P A. 2018. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Transactions on Medical Im-

- ging, 37(12): 2663-2674 [DOI: 10.1109/TMI.2018.2845918]
- Milletari F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//Proceedings of the 4th International Conference on 3D Vision. Stanford, USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- 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, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Trullo R, Petitjean C, Dubray B and Ruan S. 2019. Multiorgan segmentation using distance-aware adversarial networks. Journal of Medical Imaging, 6(1): #014001 [DOI: 10.1117/1.JMI.6.1.014001]
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11534-11542 [DOI: 10.1109/CVPR42600.2020.01155]
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Wu J, Ye F, Ma J L, Sun X P, Xu J and Cui Z M. 2008. The segmentation and visualization of human organs based on adaptive region growing method//Proceedings of the 8th IEEE International Conference on Computer and Information Technology Workshops. Sydney, Australia: IEEE: 439-443 [DOI: 10.1109/CIT.2008.Workshops.24]
- Zhang L, Wang L S, Huang Y J and Chen H. 2019. Segmentation of thoracic organs at risk in CT images combining coarse and fine network [EB/OL]. [2020-08-31]. http://ceur-ws.org/Vol-2349/SegTHOR2019_paper_5.pdf
- Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2015. Object detectors emerge in deep scene CNNs [EB/OL]. [2020-08-31]. <https://arxiv.org/pdf/1412.6856.pdf>
- Zhou Y X and Bai J. 2007. Multiple abdominal organ segmentation: an atlas-based fuzzy connectedness approach. IEEE Transactions on Information Technology in Biomedicine, 11(3): 348-352 [DOI: 10.1109/TITB.2007.892695]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. Unet++: a nested u-net architecture for medical image segmentation//Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介



吉淑滢, 1996年生, 女, 硕士研究生, 主要研究方向为医学图像分割、目标检测、计算机视觉。

E-mail: 3354876639@qq.com



肖志勇, 通信作者, 男, 副教授, 主要研究方向为人工智能、机器视觉、图像/视频处理。

E-mail: zhiyong.xiao@jiangnan.edu.cn