



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  
11  
VOL.26

ISSN1006-8961  
CN11-3758/TB





第26卷第11期（总第307期）  
2021年11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

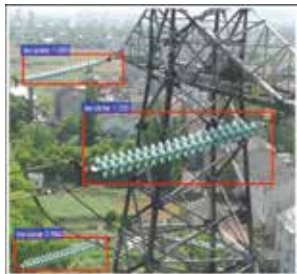
**Superintended by** Chinese Academy of Sciences  
**Sponsored by** Aerospace Information Research Institute, CAS  
China Society of Image and Graphics  
Institute of Applied Physics and Computational Mathematics

**Editor-in-Chief** Wu Yirong  
**Editor, Publisher** Editorial and Publishing Board of Journal of Image and Graphics  
**Address** No. 19, North 4<sup>th</sup> 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

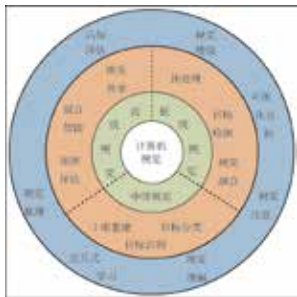
**Distributed by** Beijing Bureau for Distribution of Newspapers and Journals  
**Domestic** All Local Post Offices in China  
**Overseas** China International Book Trading Corporation  
(P.O.Box 399, Beijing 100048, P.R.China)  
**Printed by** Beijing Kexin Printing Co., Ltd.

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



面向改进尺度缩放网络的绝缘子识别(第2561页)



混合增强视觉认知架构及其关键技术进展(第2619页)



激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

编者按 ..... I

## 电力视觉前沿技术

输电线路部件视觉缺陷检测综述

赵振兵, 蒋志钢, 李延旭, 威银城, 翟永杰, 赵文清, 张珂 ..... 2545

面向改进尺度缩放网络的绝缘子识别

赵文清, 张海明, 徐敏夫 ..... 2561

最优知识传递宽残差网络输电线路螺栓缺陷图像分类

威银城, 金超熊, 赵振兵, 丁洁涛, 吕斌 ..... 2571

可变形NTS-Net的螺栓属性多标签分类

张珂, 何颖宣, 赵凯, 冯晓晗, 赵振兵, 马占宇 ..... 2582

嵌入双注意力机制的Faster R-CNN航拍输电线路螺栓缺陷检测

威银城, 武学良, 赵振兵, 史博强, 聂礼强 ..... 2594

轻量化航拍图像电力线语义分割

许刚, 李果 ..... 2605

## 综述

混合增强视觉认知架构及其关键技术进展

王培元, 关欣 ..... 2619

深度学习人脸特征点自动定位综述

徐亚丽, 赵俊莉, 吕智涵, 张志梅, 李劲华, 潘振宽 ..... 2630

## 图像处理和编码

多尺度特征复用混合注意力网络的图像重建

卢正浩, 刘丛 ..... 2645

## 图像分析和识别

自监督深度离散哈希图像检索

万方, 强浩鹏, 雷光波 ..... 2659

L-UNet: 轻量化云遮挡道路提取网络

许苗, 李元祥, 钟娟娟, 左宗成, 熊伟 ..... 2670

基于多尺度特征多对抗网络的雾天图像识别

陈硕, 钟汇才, 李勇周, 王师峰, 杨建刚 ..... 2680

## 图像理解和计算机视觉

结合动态图卷积和空间注意力的点云分类与分割

宋巍, 蔡万源, 何盛琪, 李文俊 ..... 2691

激光点云的稀疏体素金字塔邻域构建与分类

陶帅兵, 梁冲, 蒋腾平, 杨玉娇, 王永君 ..... 2703

## 计算机图形学

单幅植物叶片图像的3D重建

任非儿, 刘通, 杨龙 ..... 2713

## 医学图像处理

深度学习多部位病灶检测与分割

黎生丹, 柏正尧 ..... 2723

## 遥感图像处理

对抗学习遥感图像场景识别

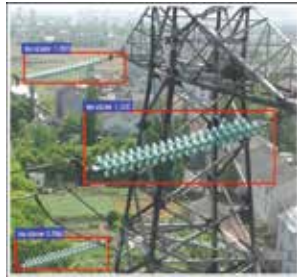
李彤, 张钧萍 ..... 2732

融合特征的卫星视频车辆单目标跟踪

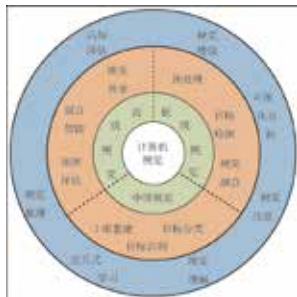
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 ..... 2741

# CONTENTS

## JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)



Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

### Frontier Technology of Power Computer Vision

Overview of visual defect detection of transmission line components

Zhao Zhenbing, Jiang Zhigang, Li Yanxu, Qi Yincheng, Zhai Yongjie, Zhao Wenqing, Zhang Ke ..... 2545

Insulator recognition based on an improved scale-transferrable network

Zhao Wenqing, Zhang Haiming, Xu Minfu ..... 2561

Image classification method of transmission line bolt defects using the optimal knowledge transfer wide residual network

Qi Yincheng, Jin Chaoxiong, Zhao Zhenbing, Ding Jietao, Lyu Bin ..... 2571

Multi-label classification method of bolt attributes based on deformable NTS-Net

Zhang Ke, He Yingxuan, Zhao Kai, Feng Xiaohan, Zhao Zhenbing, Ma Zhanyu ..... 2582

Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism

Qi Yincheng, Wu Xueliang, Zhao Zhenbing, Shi Boqiang, Nie Liqiang ..... 2594

Research on lightweight neural network of aerial powerline image segmentation

Xu Gang, Li Guo ..... 2605

### Review

Hybrid enhanced visual cognition framework and its key technologies

Wang Peiyuan, Guan Xin ..... 2619

Automatic facial feature points location based on deep learning: a review

Xu Yali, Zhao Junli, Lyu Zhihan, Zhang Zhimei, Li Jinhua, Pan Zhenkuan ..... 2630

### Image Processing and Coding

Multiscale feature reuse mixed attention network for image reconstruction

Lu Zhenghao, Liu Cong ..... 2645

### Image Analysis and Recognition

Self-supervised deep discrete hashing for image retrieval

Wan Fang, Qiang Haopeng, Lei Guangbo ..... 2659

L-UNet: lightweight network for road extraction in cloud occlusion scene

Xu Miao, Li Yuanxiang, Zhong Juanjuan, Zuo Zongcheng, Xiong Wei ..... 2670

Haze image recognition based on multi-scale feature and multi-adversarial networks

Chen Shuo, Zhong Huicai, Li Yongzhou, Wang Shizheng, Yang Jiangang ..... 2680

### Image Understanding and Computer Vision

Dynamic graph convolution with spatial attention for point cloud classification and segmentation

Song Wei, Cai Wanyuan, He Shengqi, Li Wenjun ..... 2691

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud

Tao Shuaibing, Liang Chong, Jiang Tengping, Yang Yujiao, Wang Yongjun ..... 2703

### Computer Graphics

3D reconstruction of a single plant leaf image

Ren Feier, Liu Tong, Yang Long ..... 2713

### Medical Image Processing

Multiorgan lesion detection and segmentation based on deep learning

Li Shengdan, Bai Zhengyao ..... 2723

### Remote Sensing Image Processing

Remote sensing image scene recognition based on adversarial learning

Li Tong, Zhang Junping ..... 2732

Integrating multiple features for tracking vehicles in satellite videos

Han Mingfei, Li Shengyang, Wan Xue, Xuan Shiyu, Zhao Zifei, Tan Hong, Zhang Wanfeng ..... 2741

中图法分类号: TP399 文献标识码: A 文章编号: 1006-8961(2021)11-2723-09

论文引用格式: Li S D and Bai Z Y. 2021. Multiorgan lesion detection and segmentation based on deep learning. Journal of Image and Graphics, 26(11): 2723-2731 (黎生丹, 柏正尧. 2021. 深度学习多部位病灶检测与分割. 中国图象图形学报, 26(11): 2723-2731) [DOI:10.11834/jig.200353]

## 深度学习多部位病灶检测与分割

黎生丹, 柏正尧\*

云南大学信息学院, 昆明 650091

**摘要:** 目的 多部位病灶具有大小各异和类型多样的特点, 对其准确检测和分割具有一定的难度。为此, 本文设计了一种 2.5D 深度卷积神经网络模型, 实现对多种病灶类型的计算机断层扫描 (computed tomography, CT) 图像的病灶检测与分割。方法 利用密集卷积网络和双向特征金字塔网络组成的骨干网络提取图像中的多尺度和多维度信息, 输入为带有标注的中央切片和提供空间信息的相邻切片共同组合而成的 CT 切片组。将融合空间信息的特征图送入区域建议网络并生成候选区域样本, 再由多阈值级联网络组成的 Cascade R-CNN (region convolutional neural networks) 筛选高质量样本送入检测与分割分支进行训练。结果 本文模型在 DeepLesion 数据集上进行验证。结果表明, 在测试集上的平均检测精度为 83.15%, 分割预测结果与真实标签的端点平均距离误差为 1.27 mm, 直径平均误差为 1.69 mm, 分割性能优于 MULAN (multitask universal lesion analysis network for joint lesion detection, tagging and segmentation) 和 Auto RECIST (response evaluation criteria in solid tumors), 且推断每幅图像平均时间花费仅 91.7 ms。结论 对于多种部位的 CT 图像, 本文模型取得良好的检测与分割性能, 并且预测时间花费较少, 适用病变类别与 DeepLesion 数据集类似的 CT 图像实现病灶检测与分割。本文模型在一定程度上能满足医疗人员利用计算机分析多部位 CT 图像的需求。

**关键词:** 深度学习; 计算机断层扫描 (CT) 图像; 病灶检测; 病灶分割; DeepLesion

## Multiorgan lesion detection and segmentation based on deep learning

Li Shengdan, Bai Zhengyao\*

School of Information Science and Engineering, Yunnan University, Kunming 650091, China

**Abstract:** **Objective** Most of the computed tomography (CT) image analysis networks based on deep learning are designed for a single lesion type, such that they are incapable of detecting multiple types of lesions. The general CT image analysis network focusing on accurate, timely diagnosis and treatment of patients is urgently needed. The public medical image set is quite difficult to build because doctors or researchers must process the existing CT images more efficiently and diagnose diseases more accurately. To improve the performance of CT image analysis networks, several scholars have constructed 3D convolutional neural networks (3D CNN) to extract substantial spatial features which have better performances than those of 2D CNN. However, the high computational complexity in 3D CNN restricts the depth of the designed networks, resulting in performance bottlenecks. Recently, the CT image dataset with multiple lesion types, DeepLesion, has contributed to the universal network construction for lesion detection and segmentation task on CT images. Different lesion scales and types cause a large burden on lesion detection and segmentation. To address the problems and improve the performance of CT image analysis networks, we propose a model based on deep convolutional networks to accomplish the tasks of multi-organ

收稿日期: 2020-07-17; 修回日期: 2020-11-14; 预印本日期: 2020-11-21

\* 通信作者: 柏正尧 baizhy@ynu.edu.cn

lesion detection and segmentation on CT images, which will help doctors diagnose the disease quickly and accurately.

**Method** The proposed model mainly consists of two parts. 1) Backbone networks. To extract multi-dimension, multi-scale features, we integrate bidirectional feature pyramid networks and densely connected convolutional networks into the backbone network. The model's inputs are the combination of CT key slice and the neighboring slices, where the former provides ground truth information, and the latter provide the 3D context. Combining the backbone network with feature fusion method enables the 2D network to extract spatial information from adjacent slices. Thus, the network can use features of the adjacent slices and key slice, and network performance can be improved by utilizing the 3D context from the CT slices. Moreover, we try to simplify and fine tune the network structure such that our model has a better performance as well as low computational complexity than the original architecture. 2) Detection and segmentation branches. To produce high-quality, typical proposals, we place the features fused with 3D context into the region of proposal network. The cascaded R-CNN (region convolutional neural network) with gradually increasing threshold resamples the generated proposals, and the high-quality proposals are fed into the detection and segmentation branches. We set the anchor ratios to 1 : 2, 1 : 1, and 2 : 1, and the sizes in region of proposal networks to 16, 24, 32, 48, and 96 for the different scales of lesions. We take different cascaded stages with different value of intersection over union such as 0.5, 0.6, and 0.7 to find the suitable cascaded stages. The original region of interest (ROI) pool method is substituted with ROI align for better performances. **Result** We validate the network's performance on the dataset DeepLesion containing 32 120 CT images with different types of lesions. We split the dataset into three subsets, namely, training set, testing set, and validating set, with proportions of 70%, 15%, and 15%, respectively. We employ the stochastic gradient descent method to train the proposed model with an initial learning rate of 0.001. The rate will drop to 1/10 of the original value in the fourth and sixth epoch (eight epochs in total for training). Four groups of comparative experiments are conducted to explore the effects of different networks on detection and segmentation performance. Multiple network structures such as feature pyramid networks (FPN), bidirectional feature pyramid networks (BiFPN), feature fusion, and different number of cascade stages and segmentation branch are considered in our experiments. Experimental results show that BiFPN can function well in the detection task compared with FPN. Moreover, detection performance is greatly improved by using the feature fusion method. As the number of cascaded stages increases, detection accuracy drops slightly, while the performance of segmentation improves greatly. In addition, the networks without a segmentation branch can detect lesions more accurately than those with a segmentation branch. Hence, we recognize a negative relationship between detection and segmentation tasks. We can select different structures for distinct requirements on detection or segmentation accuracy to achieve satisfying results. If doctors or researchers want to diagnose lesions more accurately, the baseline network without a segmentation branch can meet the requirements. For more accurate segmentation results, baseline network with three cascaded stages network can achieve the goal. We present the results from the three-stage cascaded networks. The results show that the average detection accuracy of our model on the DeepLesion test set is 83.15%, while the average distance error between the segmentation prediction result and the real weak label of response evaluation criteria in solid tumors (RECIST)'s endpoint is 1.27 mm, and the average radius error is 1.69 mm. Our network's performance in segmentation is superior to the multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation and auto RECIST. The inference time per image in our network is 91.7 ms. **Conclusion** The proposed model achieves good detection and segmentation performance on CT images, and takes less time to predict. It is suitable for accomplishing lesion detection and segmentation in CT images with similar lesion types in the DeepLesion dataset. Our model trained on DeepLesion can help doctors diagnose lesions on multiple organs using a computer.

**Key words:** deep learning; computed tomography (CT) images; lesion detection; lesion segmentation; DeepLesion

## 0 引言

计算机通过对大量医学影像数据的监督学习,

能够提供参考的病变位置,辅助医生缩减人工诊断时间,迅速为患者提供治疗方案。然而公开的医学影像数据集较少且获取难度较大,这就要求更高效地处理现有数据,提高诊断精度。现阶段设计的网

络通常仅适用于诊断某一种类型的病灶。由于缺乏通用网络,不便于对患者进行全面检查,无形之中增加烦琐步骤,耽误病情及时诊断。因此,建立通用病灶诊断网络和提高病灶诊断精度成为研究热点。为尽可能获取计算机断层扫描(computed tomography, CT)图像中的空间信息,增强网络特征提取能力,提出了3维卷积神经网络(three dimension convolutional neural networks, 3D CNN),以连续的CT切片组作为3D网络的输入,对比2D网络,空间信息的加入使网络得到更充分的训练,性能得到提高。谢未央等人(2019)以3D ResNeXt(Xie等,2017)为骨干网络并结合区域建议网络(region of proposals networks, RPN)(Ren等,2017),在公开的LUNA16(lung nodule analysis)数据集(Setio等,2017)上实现对肺结节的精准检测。Taha等人(2018)提出一种端对端的网络Kid-Net,以3D CT切片作为输入,实现对肾脏血管和动、静脉的分割。Roth等人(2018)提出一种3D FCN(fully convolutional networks)模型,实现腹部CT图像的语义分割,该模型的Dice平均分数达到90%。通过3D CNN提取和利用空间信息,网络性能在一定程度上得到提高。然而,3D CNN带来的大量可学习参数和计算量阻碍了构建更加深层和复杂的网络,使得性能受到制约。而2D网络以单一的CT切片作为输入,由于缺少对空间信息的利用,网络训练不够充分,同样面临性能受到制约的情况。随着具有多种部位和病变的大型CT图像数据集DeepLesion

(Yan等,2018b)的公开,为设计通用病灶诊断网络提供了便利。针对该数据集,Yan等人(2019)提出一种通用病灶分析网络(multitask universal lesion analysis network for joint lesion detection, tagging and segmentation, MULAN),采用2.5D网络并结合特征融合办法,利用2D卷积提取和利用CT切片组的空间信息,对比3D CNN,不仅网络计算量下降,而且使用更深层网络使得检测精度得到较大提高。

为进一步提高通用病灶诊断网络的精度,本文设计一种2.5D的深度卷积神经网络模型,实现对多种病灶类型的CT图像的病灶检测与分割。该模型经过基于迁移学习的特征提取网络后生成候选区域样本,再经过检测分支与分割分支预测检测框和分割区域。与MULAN相比,本文模型引入更优秀的特征提取网络获取检测和分割所需的特征,而级联网络筛选高阈值的候选样本对分割分支训练,因而具有更出色的分割性能。

## 1 相关技术

### 1.1 Mask R-CNN

Mask R-CNN(region convolutional neural networks)(He等,2017)是计算机视觉领域通用的实例分割框架,可以实现对目标的检测与分割任务,如图1所示。为使特征图中的感兴趣区域(regions of interest, ROI)能更精确地采样,引入ROI Align操作。

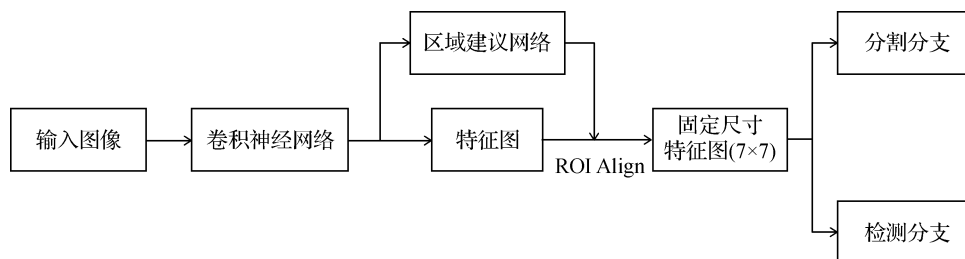


图1 Mask R-CNN 框架(He等,2017)

Fig. 1 The Mask R-CNN framework(He et al., 2017)

### 1.2 密集连接网络

为解决深度卷积神经网络出现的梯度消失问题,Huang等人(2017)提出密集卷积网络(densely connected convolutional networks, DenseNet),如图2所示。

### 1.3 BiFPN

传统的特征金字塔网络(feature pyramid networks, FPN)(Lin等,2017)将不同分辨率的特征图

直接相加,实际上每个特征图对网络输出的贡献不同。针对这个问题,Tan等人(2020)提出双向特征金字塔网络(bi-directional feature pyramid networks, BiFPN),如图3所示(图中, $P_3 \sim P_7$ 表示从低阶到高阶的5个具有不同分辨率的输入特征)。

### 1.4 Cascade R-CNN

传统目标检测网络仅使用单一阈值计算候选样

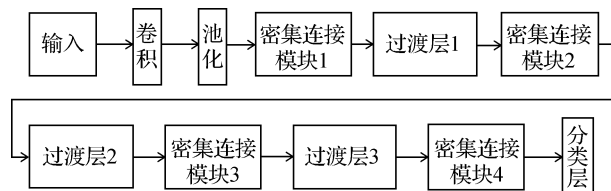


图2 密集连接网络结构(Huang等,2017)

Fig. 2 The DenseNet architecture(Huang et al., 2017)

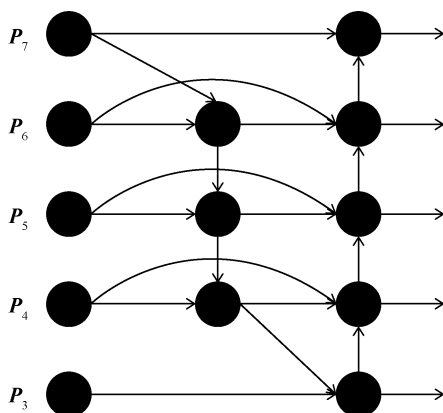


图3 BiFPN结构(Tan等,2020)

Fig. 3 BiFPN architecture(Tan et al., 2020)

本的交并比(intersection over union, IOU),并且依据设定的阈值划分正负样本。为解决各候选样本具有不同 IOU 的问题, Cai 和 Vasconcelos (2018) 提出 Cascade R-CNN。图4为三阶段级联网络,其中,级联网络由原先的 ROI pool 替换为 ROI Align。 $B$  代表回归框,  $C$  代表类别概率,  $H$  代表检测分支, Align 代表 ROI Align,  $F$  代表特征图,  $B_0$  代表 RPN 输出的候选样本。

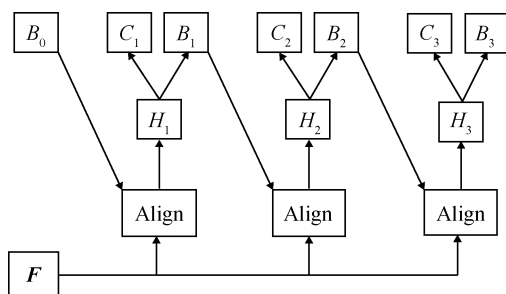


图4 三阶段级联网络(Cai和Vasconcelos等,2018)

Fig. 4 Three stage cascaded network  
(Cai and Vasconcelos, 2018)

### 1.5 迁移学习

在深度学习领域常常在某一较大型数据集上完成预训练的网络,然后经过微调该模型来学习与其

相关的其他数据集。这种迁移学习的操作可以较大地减少训练时间和加速网络收敛,是目前构建深度学习网络必不可少的一部分。

## 2 数据集

### 2.1 数据集简介

实验采用 DeepLesion 数据集,由美国国立卫生研究院临床中心团队依据历史医学数据构建,数据集集中的 CT 图像来自 4 427 名患者,共 10 594 个病例,有 32 120 个病变标注,包含多种病灶类型且每幅图像存在 1~3 个病灶,是目前最大的公开 CT 图像数据集(Yan等,2018b),分为训练集、验证集与测试集,比例分别为 70%、15% 和 15%。

### 2.2 数据集的预处理

为突出病灶特征,对 CT 图像进行增强。1) 采用 3D 增强。位于相邻切片上的病灶大概率在中央切片的相同位置,可依据病灶大小读取病变更明显的切片作为中央切片。将中央切片的序号加上或减去一个随机值,上限为病灶短半径的一半乘以像素的实际距离,再除去切片间距并取整,之后读取新序号的切片实现 3D 增强。2) 读取中央切片后,根据数据集提供的切片间隔值,利用线性插值与相邻切片生成 8 张新相邻切片,减去 32 768,得到图像的真实 HU(Hounsfield unit)值。3) 由于不同部位的 CT 图像 HU 值不同,先采用对比增强突出病灶特征,再利用归一化操作避免不同 HU 值对网络训练的影响。4) 裁剪图像黑边以及随机调整尺寸,用于减少不必要的计算量以及数据增广。

## 3 模型设计

### 3.1 基于迁移学习的特征提取网络

本文模型基于 Mask R-CNN,采用 DenseNet-121 结合 BiFPN 设计特征提取网络。实验发现,去除官方版本 DenseNet 的最后一层 dense block 4 与第 3 层的 transition layer 后,不仅能降低计算量,而且性能有所提升。因此,本文网络对 DenseNet-121 进行结构简化。特征提取网络如图 5 所示,其中 ts1、ts2 和 ds3 是 DenseNet-121 中第 1、2 层的 transition layer 以及 dense block 3 的简称。为方便说明,图 5 中的切片图像未经裁剪。

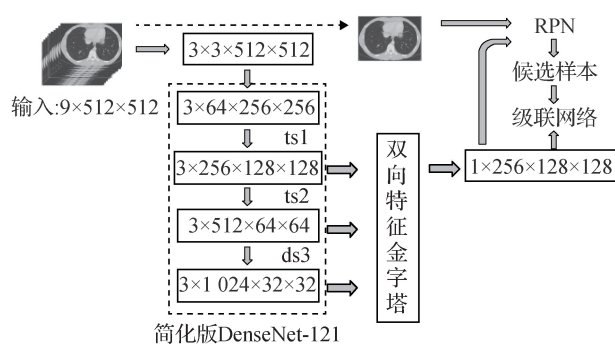


图5 特征提取网络

Fig. 5 Feature extraction network

由于裁剪 CT 图像的黑边后,特征图在 BiFPN 中会出现维度不匹配情况,为避免此类情况发生,本文对 BiFPN 网络进行简化,去除了官方版本的  $P_6$  和  $P_7$ ,并且相应调整了结构,如图 6 所示。图中 BiFPN 的输入和输出的卷积层通道数为 256,输出具有最高分辨率的特征图。由于各个病灶之间的大小差异很大,范围从 0.21 ~ 342.5 mm,因此选用最高分辨率的特征图对网络进行训练。

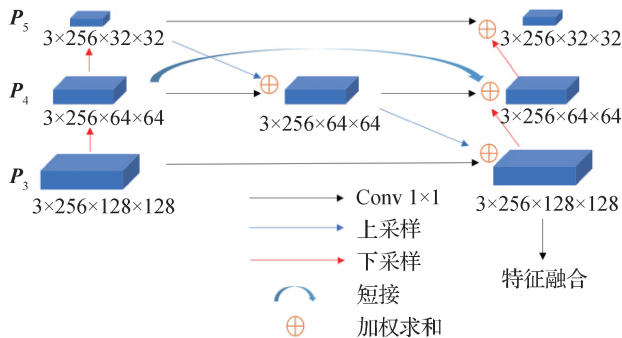


图6 简化版 BiFPN

Fig. 6 Simplified BiFPN architecture

特征融合如图 7 所示,特征图在第 0 维联结,再经过一层 2D 卷积后实现相邻切片与中央切片的特征融合,以达到提取空间信息的目的。图 7 中,  $C$  代表卷积核的通道数,  $H$  和  $W$  分别代表特征图的长和宽。

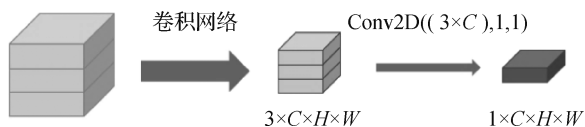


图7 特征融合

Fig. 7 Feature fusion

### 3.2 病灶检测与分割

本文设计的网络结构中的级联网络分为单阶

段、级联两阶段与级联三阶段,图 4 为级联三阶段的结构。各阶段的候选样本特征图经过 ROI Align 和各自的检测分支后输出回归框和类别概率,再将回归框解码后送入下一阶段,如此循环往复。分割分支不采用级联结构,输入是由设计的级联网络最后阶段生成的候选正样本,之后输出预测的分割结果,检测与分割分支如图 8 所示。图 8 中,FC 代表全连接层,Conv 代表 2D 卷积。

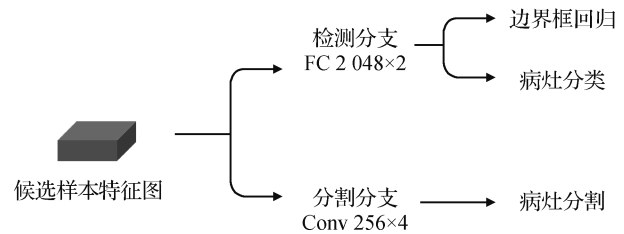


图8 检测与分割分支

Fig. 8 Detection and segmentation branch

## 4 实验结果与分析

### 4.1 实验设置

本文网络在 Pytorch 框架下实现,模型在 RTX 2060 Super 显卡上训练,使用随机梯度下降算法训练 8 轮,每批送入的样本数量为 2,基础学习率为 0.001,并在第 4 和第 6 轮分别下降至原先的 1/10。RPN 的锚框尺寸设置为 [16, 24, 32, 48, 96],比例设置为 [1:2, 1:1, 2:1]。级联网络各阶段参数设置如下:阈值为 [0.5, 0.6, 0.7],边框解码权重分别为 [1, 1, 0.5, 0.5], [2, 2, 1, 1] 和 [3, 3, 1.5, 1.5]。检测分支的类别损失权重为 1,边框回归损失权重为 10,分割分支损失权重为 1。

为对比不同网络结构对性能的影响,设置了 4 组实验。基准网络为基于迁移学习的特征提取网络和单阶段检测与分割网络,即简化版 DenseNet-121 与 BiFPN 为特征提取网络,加入特征融合方法,以及由单一阈值构成的检测与分割分支的模型。

第 1 组实验包括:1) 采用 BiFPN 的基准网络,不加入特征融合;2) 采用 FPN 的基准网络,不加入特征融合。第 2 组实验包括:1) 采用 BiFPN 的基准网络;2) 采用 BiFPN 的基准网络,但不加入特征融合;3) 采用 FPN 的基准网络;4) 采用 FPN 的基准网络,但不加入特征融合。第 3 组实验包括:1) 基准

网络;2)基准网络级联两阶段;3)基准网络级联三阶段。第4组实验包括:对第3组的3个对比试验采取剔除分割任务操作并进行性能对比。

## 4.2 评价指标

### 4.2.1 检测分支

采用与MULAN相同的评价指标,检测类别分为病灶和背景两类。灵敏度(sensitivity)计算为

$$S = \frac{TP}{TP + FN} \quad (1)$$

式中, $TP$ 代表真阳性(true positive),指实际是病灶且预测为病灶, $FN$ 代表假阴性(false negatives),指实际是病灶而预测为背景。由于一幅CT切片上可能存在多处病变,因此使用无限制受试者工作特征(free-response receiver operating characteristic, FROC)评价检测性能。为与其他方法进行对比,将每幅图像负样本数为[0.5,1,2,4]的平均灵敏度作为对比基准。

### 4.2.2 分割分支

DeepLesion数据集没有提供像素级别的分割标签,采用实体瘤疗效评价标准(response evaluation criteria in solid tumors, RECIST)(Eisenhauer等,2009)构建分割的软标签,真实标注框、RECIST以及检测与分割结果如图9所示。加粗的绿色框和绿色的长短轴分别表示真实标注框和RECIST,红色长短轴及其构建的轮廓是网络预测的分割结果,红色细框表示预测框,上方的0.976表示预测为病灶的概率。分割指标采用与MULAN相同的两项指标:

1)从真实标签的端点到预测结果的平均距离误差;

2)预测结果与真实标签直径长度的平均误差。

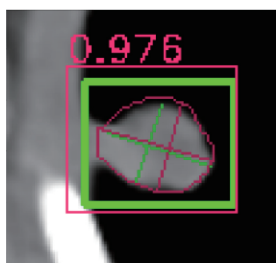


图9 真实标注框、RECIST以及检测与分割结果

Fig.9 Ground truth, RECIST, detection and segmentation results

## 4.3 实验结果分析

由于根据官方标准划分数据集,因此本文实验结果可与采用同样数据集的其他方法进行对比。

在DeepLesion测试集上的检测性能对比与消融实验见表1,分割性能对比与消融实验如表2所示。在表1中,ULDor(universal lesion detector)(Tang等,2019)、3DCE(3D context enhanced region-based CNN)(Yan等,2018a)和RetinaNet(Zlocha等,2019)实验仅包括病灶检测,Auto-RECIST(Tang等,2018)仅包括病灶分割。MULAN\*为MULAN不将额外标签数据用于训练优化层的实验性能。表3为网络在推断阶段对每幅图像的推理时间,值越小表示预测给定图像的速度越快。MULAN\*的推断时间取自公开的代码日志,其他实验在同一设备上,由于实验设备不同,该对比仅供参考。

1)第1组实验。从表1可知,BiFPN病灶检测性能明显优于FPN,说明引入权重的BiFPN能够更好地平衡不同特征图的影响,从而提高检测性能。

2)第2组实验。从表1可以看出,加入特征融合后,网络性能都得到提升,而BiFPN结合特征融合的网络性能优于其他网络,说明特征融合通过提取CT相邻切片中的空间信息,能够较大幅度地提高网络特征提取能力,进而提高检测精确度。

3)第3组实验。从表1可知,随着级联阶段的增多,检测性能有所下降,说明设定值为0.5的阈值能较好地符合候选样本的IOU值,而过大的阈值容易造成过拟合而导致精度下降。从表2可知,随着级联阶段的增加,分割精度有较大提升,并且级联三阶段网络分割误差最小,本文的实验结果即为级联三阶段网络的预测结果图。从表3可知,级联三阶段网络每幅图像推断时间最短,说明符合要求高阈值的候选样本数量较少,因而处理速度得到加快。

4)第4组实验。从表1可知,去除分割分支后的网络检测性能都有所提高。结合特征融合的BiFPN网络在去除分割任务后的检测精度达到最优,为84.71%,说明分割任务与检测任务存在耦合关系。

结合表1和表2可知,由于分割分支仅对正样本进行病灶分割,级联阶段越多,生成的候选区域越接近真实标签,分割性能提升越明显。然而级联阶段越多,过拟合对检测性能影响越大,导致性能下降。因此可以针对不同精确度需求,灵活分配网络级联的阶段和任务。对病灶初检,选择单阶段无分割任务的网络,提供更精确的病灶检测结果;对病灶需要观察是否变化,选择级联三阶段的网络,提供较为精确的分割结果用于诊断参考。

表 1 不同方法在 DeepLeison 测试集上的检测性能 (灵敏度) 对比与消融实验研究  
Table 1 Comparison of detection accuracy (sensitivity) and ablation studies on the test set of DeepLesion among different methods

| 方法                        | 每幅图像负样本数     |              |              |              | 均值           |
|---------------------------|--------------|--------------|--------------|--------------|--------------|
|                           | 0.5          | 1            | 2            | 4            |              |
| ULDor( Tang 等,2019)       | 52.86        | 64.8         | 74.84        | 84.38        | 69.22        |
| 3DCE( Yan 等,2018a)        | 62.48        | 73.37        | 80.7         | 85.65        | 75.55        |
| RetinaNet( Zlocha 等,2019) | 72.15        | 80.07        | 86.4         | 90.77        | 82.35        |
| MULAN * ( Yan 等,2019)     | -            | -            | -            | -            | 84.24        |
| FPN 无特征融合                 | 65.88        | 75.67        | 82.25        | 87.46        | 77.81        |
| BiFPN 无特征融合               | 67.00        | 76.2         | 82.94        | 88.09        | 78.56        |
| FPN                       | 74.96        | 82.45        | 87.13        | 91.21        | 83.94        |
| BiFPN( 基准网络)              | 75.47        | 82.76        | 87.85        | 91.00        | 84.27        |
| BiFPN( 基准网络) + 无分割        | <b>75.61</b> | <b>82.89</b> | <b>88.25</b> | <b>92.09</b> | <b>84.71</b> |
| BiFPN + 级联两阶段             | 72.42        | 81.9         | 87.36        | 91.41        | 83.27        |
| BiFPN + 级联两阶段 + 无分割       | 73.37        | 82.61        | 87.89        | 91.67        | 83.89        |
| BiFPN + 级联三阶段             | 73.09        | 81.33        | 87.03        | 91.16        | 83.15        |
| BiFPN + 级联三阶段 + 无分割       | 73.85        | 81.54        | 87.37        | 91.28        | 83.51        |

注:加粗字体表示各列最优结果,“-”表示该项指标不详,无分割表示剔除分割任务。

表 2 不同方法在 DeepLeison 测试集上的分割性能 (分割误差) 对比与消融实验研究  
Table 2 Comparison of segmentation accuracy (error) and ablation studies on the test set of DeepLesion among different methods

| 方法                        | /mm         |             |
|---------------------------|-------------|-------------|
|                           | 端点平均距离      | 直径平均误差      |
| Auto RECIST( Tang 等,2018) | -           | 1.71        |
| MULAN * ( Yan 等,2019)     | 1.43        | 1.97        |
| BiFPN( 基准网络)              | 1.49        | 2.07        |
| BiFPN( 基准网络) + 级联两阶段      | 1.38        | 1.88        |
| BiFPN( 基准网络) + 级联三阶段      | <b>1.27</b> | <b>1.69</b> |

注:加粗字体表示各列最优结果,“-”表示该项指标不详。

4.4 多部位病灶检测与分割预测结果

图 10 是 BiFPN 级联三阶段网络在 DeepLesion 测试集的部分预测结果,绿色的加粗框和长短轴分别表示真实标注框和 RECIST,其他颜色表示网络预测结果,上方标注的数字表示预测的概率。由图 10

表 3 不同方法每幅图像推断时间对比  
Table 3 Comparison of inference time per image among different methods

| 方法                    | 每幅图像推断时间 /ms |
|-----------------------|--------------|
|                       |              |
| MULAN * ( Yan 等,2019) | 102.2        |
| BiFPN( 基准网络)          | 95.7         |
| BiFPN( 基准网络) + 级联两阶段  | 93.8         |
| BiFPN( 基准网络) + 级联三阶段  | <b>91.7</b>  |

注:加粗字体表示最优结果。

可知,对于数据集中大小各异的病灶,本文网络的检测与分割表现良好,预测结果有可能更接近真实病灶,也可能预测出负样本或者真实缺失标签。如图 10(f),从放大图可以看出,黄色预测结果的覆盖区域要比真实标签更加贴近骨盆病灶,其概率为 0.949。而图 10(b)中,左侧的紫色预测结果标记一处阴影区域,其概率为 0.912,有可能是未标记的真实病变区域或者负样本。

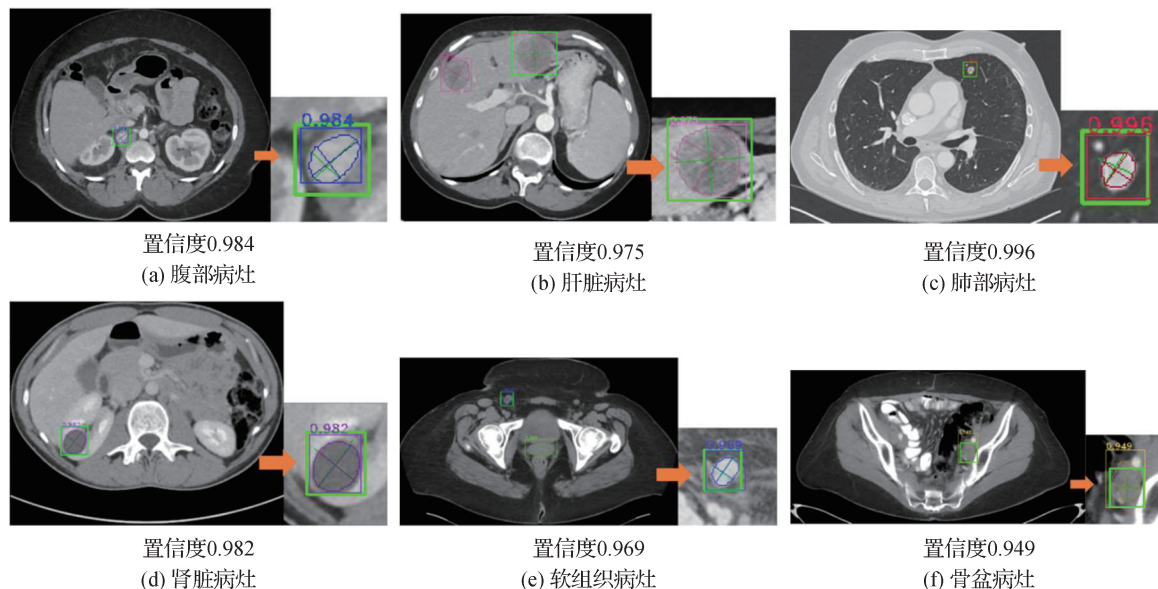


图 10 实验结果

Fig. 10 The experiment results

((a) abdomen lesion; (b) liver lesion; (c) lung lesion; (d) kidney lesion; (e) soft tissue lesion; (f) pelvis lesion)

## 5 结 论

本文提出一种基于深度卷积网络的多部位 CT 图像病灶检测与分割方法。该方法以 DenseNet 与 BiFPN 为骨干网,并且结合特征融合办法,使网络提取和利用切片组中的空间信息。同时,后续的级联网络筛选高质量的候选样本送入检测与分割分支训练,分割性能比 MULAN 方法得到进一步提升。实验表明,在包含多部位和多病变 CT 图像的 DeepLesion 数据集上,本文方法能够较准确且快速实现病灶检测与分割。

目前,制约检测性能提升的因素主要是标签数量不足,原因如下:1) DeepLesion 数据集中的切片标注较为耗时,而专家更关注明显或重要的病变区域,没有标注所有切片的病变区域,标签数据总体上偏少;2) 某些类别的病灶比较罕见,导致训练出现正负样本比例不均衡问题,网络检测该类型的病变效果较差。在今后的工作中,将研究如何利用已有的真实标签数据,挖掘存在于相邻切片的真实标签并补充到原有标签集中,从而增加网络训练的数据量,提高检测准确度。

## 参考文献 (References)

- Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]
- Eisenhauer E A, Therasse P, Bogaerts J, Schwartz L H, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D and Verweij J. 2009. New response evaluation criteria in solid tumours; revised RECIST guideline (version 1.1). European Journal of Cancer, 45 (2): 228-247 [DOI: 10.1016/j.ejca.2008.10.026]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- 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: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Roth H R, Shen C, Oda H, Oda M, Hayashi Y, Misawa K and Mori K. 2018. Deep learning and its application to medical image segmentation. Medical Imaging Technology, 36 (2): 63-71 [DOI: 10.11409/mit.36.63]

- Setio A A A, Traverso A, De Bel T, Berens M S N, Van Den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci M E, Geurts B, Van Der Gugten R, Heng P A, Jansen B, De Kaste M M J, Kotov V, Lin J Y H, Manders J T M C, S  nora-Mengana A, Garc  a-Naranjo J C, Papavasileiou E, Prokop M, Saletta M, Schaefer-Prokop C M, Scholten E T, Scholten L, Snoeren M M, Torres E L, Vandemeulebroucke J, Walasek N, Zuidhof G C A, Van Ginneken B and Jacobs C. 2017. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical Image Analysis*, 42: 1-13 [DOI: 10.1016/j.media.2017.06.015]
- Taha A, Lo P, Li J N and Zhao T. 2018. Kid-net: convolution networks for kidney vessels segmentation from CT-volumes//Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention. Granada, Spain: Springer; 463-471 [DOI: 10.1007/978-3-030-00937-3\_53]
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE; 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]
- Tang Y, Harrison A P, Bagheri M, Xiao J and Summers R M. 2018. Semi-automatic RECIST labeling on CT scans with cascaded convolutional neural networks//Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer; 405-413 [DOI: 10.1007/978-3-030-00937-3\_47]
- Tang Y B, Yan K, Tang Y X, Liu J M, Xiao J and Summers R M. 2019. Uldor: A universal lesion detector for CT scans with pseudo masks and hard negative example mining//The 16th IEEE International Symposium on Biomedical Imaging (ISBI 2019). Venice, Italy: IEEE; 833-836 [DOI: 10.1109/ISBI.2019.8759478]
- Xie S N, Girshick R, Doll  r P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE; 5987-5995 [DOI: 10.1109/CVPR.2017.634]
- Xie W Y, Chen Y B, Wang J Y, Li Q and Chen Q. 2019. Detection of pulmonary nodules in CT images based on convolutional neural networks. *Computer Engineering and Design*, 40 (12): 3575-3581 (谢未央, 陈彦博, 王季勇, 李强, 陈群. 2019. 基于卷积神经网络的CT图像肺结节检测. *计算机工程与设计*, 40 (12): 3575-3581) [DOI: 10.16208/j.issn1000-7024.2019.12.035]
- Yan K, Bagheri M and Summers R M. 2018a. 3D context enhanced region-based convolutional neural network for end-to-end lesion detection//Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer; 511-519 [DOI: 10.1007/978-3-030-00928-1\_58]
- Yan K, Tang Y, Peng Y, Sandfort V, Bagheri M, Lu Z Y and Summers R M. 2019. MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation//Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen, China: Springer; 194-202 [DOI: 10.1007/978-3-030-32226-7\_22]
- Yan K, Wang X S, Lu L and Summers R M. 2018b. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, 5 (3): #036501 [DOI: 10.1117/1.JMI.5.3.036501]
- Zlocha M, Dou Q and Glocker B. 2019. Improving retinaNet for CT lesion detection with dense masks from weak RECIST labels//Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen, China: Springer; 402-410 [DOI: 10.1007/978-3-030-32226-7\_45]

## 作者简介



黎生丹, 1995年生, 男, 硕士研究生, 主要研究方向为基于深度学习的医学影像处理。

E-mail: 646286091@qq.com



柏正尧, 通信作者, 男, 教授, 主要研究方向为信号处理、图像处理、模式识别。

E-mail: baizhy@ynu.edu.cn