



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

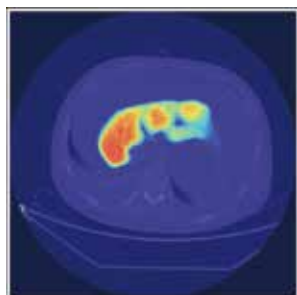
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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

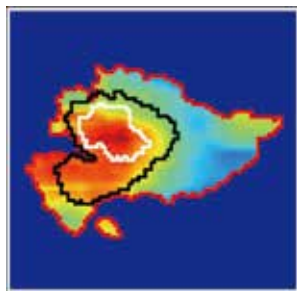
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



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

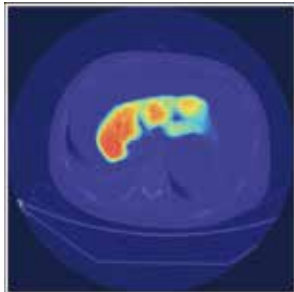
吕鹏飞, 王莹, 王思齐, 于晓升, 吴成东 2293

CONTENTS

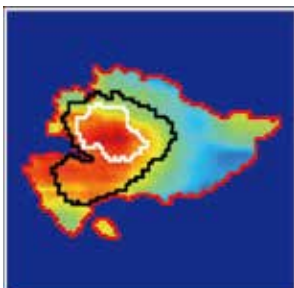
JOURNAL OF IMAGE AND GRAPHICS



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, Hei 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: TP391.4 文献标识码: A 文章编号: 1006-8961(2021)09-2181-12

论文引用格式: Fang C and Bai Z Y. 2021. Multi-scale bone lesion detection based on attention mechanism and deformable convolution. Journal of Image and Graphics, 26(09): 2181-2192 (方成, 柏正尧. 2021. 融合注意力机制与可变形卷积的多尺度骨病变检测. 中国图象图形学报, 26(09): 2181-2192) [DOI: 10.11834/jig.200476]

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*

云南大学信息学院, 昆明 650500

摘要: **目的** 在计算机断层扫描(computed tomography, CT)影像中对骨组织部位进行自动分析和检测, 对于骨科疾病的早期诊断具有重要意义, 然而基于人工分析诊断的方法存在效率较低、诊断的准确性和客观一致性无法保证等问题。为此, 本文研究构建一个骨组织病变检测的级联神经网络模型, 以期为骨科医生的诊断提供支持。**方法** 在影像预处理阶段使用改进的增强方法对CT影像进行对比度增强并获取影像中的人体有效部位; 根据骨骼组织CT值(Hounsfield unit, HU)的分布范围进行阈值分割, 得到大致的骨组织区域; 以级联目标检测模型为研究基线, 结合注意力机制与可变形卷积增加特征图的全局上下文的相关性, 以适应形态多变的骨病灶; 通过特征融合模块促进不同尺度特征信息之间的融合, 并在多个尺度特征图上分别进行骨组织病变训练和预测。**结果** 在DeepLesion数据集上进行实验, 实验结果表明, 本文网络对骨病变检测的召回率(recall)、准确率(precision)、F1分数、平均精度(average precision, AP)分别为0.85、0.613、0.712以及0.816; 较对照组中性能最优的通用CT病变检测网络对骨病变检测的召回率提升0.15。**结论** 本文提出的网络模型对CT骨组织病变具有较好的检测效果, 能够对骨组织病变判别诊断提供辅助支持, 提高诊断效率, 降低漏诊风险。**关键词:** 骨病变检测; 多尺度目标检测; 注意力机制; 医学影像处理; 级联神经网络

Multi-scale bone lesion detection based on attention mechanism and deformable convolution

Fang Cheng, Bai Zhengyao*

School of Information Science and Engineering, Yunnan University, Kunming 650500, China

Abstract: **Objective** Since frequent orthopedic diseases cause serious harm to human body, automatic analysis and detection of bone tissue position in computed tomography (CT) has crucial clinical significance for early diagnosis of orthopedic diseases. The method based on manual analysis and diagnosis of bone tissue in CT image has problems such as low efficiency. The accuracy and objective consistency of diagnosis cannot be achieved. Therefore, A cascaded neural network model for bone tissue lesion detection has been demonstrated to aid decision support for orthopedic surgeons' diagnosis. **Method** The proposed bone lesion detection algorithm has mainly consist of four steps. At first, convert original data into CT value data in terms of the conversion formula and relative files that have been illustrated by the National Institutes of Health Clinical Center (NIHCC), USA, in the preprocessing stage. The segmentation has been conducted via the mean value of CT value data as the threshold in order to filter out most of the non-human body parts in the image. The segmentation cannot be filtered out entirely due to the high CT value of the bed plate material of CT equipment. Rectangle kernels (RK)-based

收稿日期: 2020-08-20; 修回日期: 2020-10-22; 预印本日期: 2020-10-29

* 通信作者: 柏正尧 baizhy@ynu.edu.cn

opening operation in morphological operations have benefited to filter out the CT bed plate from CT image. According to the characteristics of bone tissue in CT image, a contrast enhancement method based on Gamma transform is proposed to enhance the contrast of CT images. Next, the approximate bone tissue area in enhanced CT images have been calculated via thresholding based on the distribution range of the Hounsfield unit (HU) of the bone tissue in the CT image. The cascaded object detection model has been set up as our baseline. The attention mechanism and deformable convolution have been integrated to increase the global context relevance of the feature map based on the bone lesions with variable shapes. At last, the feature fusion module has been used to strengthen the fusion of feature information at various scales. A multi-scale feature map for the training and prediction of bone tissue lesions has been sorted out. **Result** Four designated groups of comparative experiments in the context of the network structure have compared with modeling capability. The detected model has been mainly examined based on average precision (AP). 1) ResNet50, ResNet101 and ResNeXt101 have been used as feature extraction networks to complete training and testing based on naïve Cascade R-CNN (region-convolutional neural network) model to calculate the baseline. The analyzed results have shown that ResNeXt101 has the best optimization based on the AP up to 0.543 to get the baseline. 2) Feature pyramid networks (FPN), path aggregation feature pyramid networks (PAFPN), neural architecture search-feature pyramid networks (NAS-FPN) and naive structure have been adapted to complete model training and testing based on the calculated Cascade R-CNN. The feature fusion module based on the best value has been sorted out. The best PAFPN based on the AP increased to 0.721 has been leaked out the feature fusion module. 3) Two groups of comparative experiments have been illustrated. Firstly, batch normalization (BN) and group normalization (GN) modules have been used in the head of R-CNN for entire training and testing. The results have shown that the performance of GN is better than BN with 0.723 AP. Attention mechanism block and deformable convolution block have been embedded in model to verify their effectiveness in the next step. The verified results have shown the effectiveness of attention mechanism and deformable convolution module. The AP have been improved to 0.816. 4) The trained model and other object detection network models have been calculated to compare the testing value of each model. The research experiments results have been achieved based on DeepLesion dataset. The results have shown as below: 1) the recall is 0.85; 2) the precision is 0.613; 3) the F1-score is 0.712; 4) the AP is 0.816. The performance has been significantly improved in comparison of the existing universal CT lesion detection models based on the recall rates of 0.574 and 0.70. **Conclusion** The main methods such as HU value threshold segmentation and morphological operations have been used to filter out most of the non-bone tissue area in the CT image at the image preprocessing stage. The bone tissue area has been highlighted coupled with the enhancement of the image contrast further. The training model has been accelerated to reduce the interference of noise. The second group of experimental results have improved the fusion of low-level feature information, high-level feature information and enhances the location information of high-level features. The semantic information of low-level features based on the multi-scale feature pyramid fusion module has been embedded in the network structure. The detection performance has been significantly improved at the end. The third group of experimental results have been concluded based on the enhancing adaptation of attention mechanism and the weight of irrelevant information deduction. The operation of deformable convolution module has realized the network adaptation further based on multi-shapes and sizes convolution kernels the fourth group of experiment results have achieved via the comparison experiment between our model and other object detection models. The metrics including recall, precision, F1-score and AP have been mainly evaluated in these models. The experimental results have demonstrated that the model analysis has a good detection effect on the CT bone tissue lesions in the context of upgrading diagnostic efficiency, missed diagnosis deduction and quick diagnosis and treatment. The differential diagnosis of bone tissue lesions has been aided effectively. The real-time detection capability can be strengthened via the deduction of model parameters quantity and the time of training and judging.

Key words: bone lesion detection; multi-scale object detection; attention mechanism; medical image process; Cascade R-CNN

0 引言

骨病是一种广泛存在的疾病,包括脊柱炎、颈椎病、骨质增生以及骨质疏松等。据统计,我国每年骨科创伤病例高达2 000多万(仲剑平,2009),具有危害大、发病突然且前期不易发觉等特点(汪家旺等,2004),临床治疗中需要较为有效的确诊手段来支撑医生的治疗方案。

随着医疗放射技术的快速发展以及人们对自身健康关注度的持续升高,我国各级医院每天都会产生大量的影像资料。然而,当前对这些影像的分析诊断基本停留在传统的手工分析阶段,导致放射科医生的工作量激增。同时,每天面对大量的重复性工作,可能因为疲劳出现疏忽,造成误诊和漏诊,况且医生对影像的分析诊断是一个相对主观的过程,与自身业务水平、经验以及当时的状态有一定关联。因此,采用深度学习技术对骨科计算机断层扫描进行自动分析和诊断,辅助骨科医生提升诊断效率、降低漏诊率具有重要的现实意义。

目前,不少研究团队利用深度卷积神经网络(deep convolutional neural network, DCNN)对计算机断层扫描(computed tomography, CT)数据集 DeepLesion 中的8种类型病变进行了深入研究。Jiang 等人(2020)提出 ElixirNet,采用改进的区域生成网络(region proposal network, RPN)策略平衡正负样本,自动建立推荐区域之间的特征关系,提高了分类和回归效率。Zhang 等人(2020)提出了3D-AG(3D aggregated)是以 Faster R-CNN(region convolutional neural networks, R-CNN)(Ren 等,2017)为研究基线的3D多类型病变检测模型,在RPN中引入聚合分类器分支,通过特征聚合和局部放大以增强网络分类性能。Yan 等人(2019)提出的 LesaNet 通过放射报告与标签之间的关系学习对病变图像注释。这些研究表明将深度学习技术应用到骨病变检测具有可行性。然而,由于 DeepLesion 数据集中的骨病变数据过少及骨科疾病CT影像存在形态多样和特征不明显等情况,导致上述病变检测模型对骨病变的检测性能较低。

DeepLesion 数据集的骨科病变主要集中在脊柱骨、肋骨和骨盆等部位,常见病变类型包括骨质疏松、骨质增生、骶骨病变以及脊柱骨病变等。骨质疏

松的CT表现主要为骨密度减低、骨皮质变薄分层等。骨质增生硬化的CT表现主要为骨质密度增高、骨皮质增厚等。骶骨病包括骶骨浆细胞瘤、软骨肉瘤和尤文肉瘤等(赵浩锋等,2019)。骶骨浆细胞瘤表现为溶骨性骨质破坏、偏心性、呈软组织肿块;骶骨软骨肉瘤表现为骶前巨大软组织肿块、密度较高、边界不清;骶骨尤文肉瘤表现为溶骨性骨质破坏,软组织肿块突出于骶前,边界不清,偏心性。脊柱骨病包括脊柱裂、蝴蝶椎、血管瘤、骨嗜酸性肉芽肿、脊索瘤和骨髓瘤。脊柱裂的CT表现为脊椎后部中线处纵行或斜行裂隙、骨皮质光滑完整;蝴蝶椎的CT表现为椎体内纵行裂隙、边缘骨皮质样硬化。血管瘤分为良性血管瘤和侵袭性血管瘤。良性血管瘤的CT表现为椎体中央低密度、边界清晰、内见多发斑点状高密度;侵袭性血管瘤的CT表现为椎体内蜂窝状改变、骨皮质膨胀。骨嗜酸性肉芽肿的CT表现为地图样骨质破坏,严重者呈铜钱样扁平。脊索瘤的CT表现为溶骨性骨质破坏、骨皮质膨胀等。骨髓瘤分为弥漫性骨质疏松型、单骨性、多骨性和硬化型。弥漫性骨质疏松型的CT表现为有粟粒状破坏,可伴有压缩性骨折;单骨性的CT表现为膨胀性骨质破坏、椎体呈现小脑回样改变;多骨性的CT表现为穿凿样骨质破坏、骨皮质变薄;硬化型的CT表现为骨质破坏区内可见散在骨质硬化、边缘硬化(董越等,2017)。

随着医疗影像数据的累积以及计算能力的快速提升,使得深度学习技术在医疗影像领域得到了应用、提升了诊断效率,并在病灶分类、分割以及检测等领域取得了一系列成果。本文使用深度学习方法对骨病患者的CT影像进行分析研究,构建一个用于骨病变检测模型,辅助拟定治疗方案。

1 相关工作

1.1 基于深度学习的目标检测算法

基于深度学习的目标检测算法主要分为基于候选区域检测方法和单次目标检测方法。基于候选框的检测算法有 Faster R-CNN 和 Cascade R-CNN(Cai 和 Vasconcelos,2021)等,主要思路是在特征图上通过两个阶段检测网络逐步实现目标区域的分类和位置回归,达到检测模型的优化。第1阶段由RPN在

特征图上生成一系列候选区域,第2阶段通过区域卷积神经网络对候选区域进行更准确的分类和位置回归。单次目标检测算法主要有YOLO(you only look once)(Redmon等,2016)和RetinaNet(Lin等,2020)等,主要思路是直接在特征图上进行分类和位置回归操作,获得检测结果。

由于网络结构方面的差异导致两种算法的性能有所不同。通常,基于候选框的算法在检测精度上占优,单次检测算法在检测速度上占优。本文研究的是医学图像中骨病灶的检测,对检测精度的需求高于检测速度,因此选择基于候选框的检测算法为本文模型的研究基准。

Faster R-CNN 系列检测算法由 RPN 和单个 R-CNN 组成,考虑到模型训练过程中正负样本平衡性的问题,通常将交并比(intersection over union, IoU)设置为0.5,即预测区域与真实标记区域的交集与并集的比值大于0.5时判定区域含有骨病灶。IoU的表达式为

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

式中, A 表示预测区域, B 表示真实标记区域。

然而,交并比阈值设置为0.5时样本区域内含有较多背景信息,此时将其作为正样本训练网络可能会使网络学到过多的非目标特征信息(即背景信息),导致误检率升高;但直接使用更高的交并比阈值可能会造成网络无法获取足够数量的骨病灶样本,导致正负样本数量不均衡,漏检率升高。Cascade R-CNN 通过级联三阶段的 R-CNN 逐步提高 IoU 阈值,避免了直接提高 IoU 带来的风险。前一个 R-CNN 的回归区域作为后一个 R-CNN 的训练输入条件,如图1所示。

相较于 Faster R-CNN 系列检测网络,Cascade R-CNN 通过级联3个 R-CNN 进一步提高了模型的检测性能,因此本文选择 Cascade R-CNN 为模型基础研究框架。

1.2 残差网络

两种典型残差网络基本组成单元如图2所示,图2(a)为ResNet(He等,2016)的瓶颈单元,即BottleNeck。图2(b)为ResNeXt(Xie等,2017)模块,包含32条相同的卷积路径,每条路径包含的参数量较少。

通常地,提升模型性能的方法是加深或加宽网

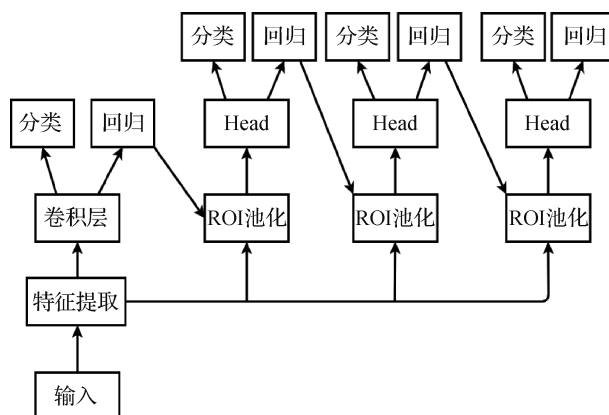


图1 Cascade R-CNN 结构

Fig. 1 Structure of Cascade R-CNN

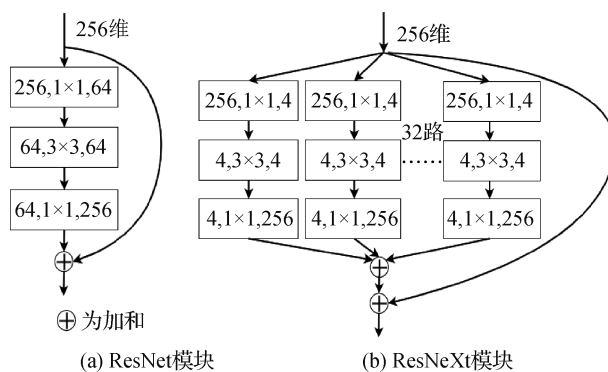


图2 ResNet与ResNeXt模块结构

Fig. 2 Unit structure of ResNet and ResNeXt((a) block of ResNet; (b) block of ResNeXt)

络结构。然而随着参数增加,网络设计难度和计算开销会快速增长。ResNeXt网络采用平行堆叠相同结构的模块(即分组卷积)代替ResNet网络的三层卷积模块,在不明显增加参数量级的情况下提升了模型性能。表1展示了ResNet101和ResNeXt101在输入图像为 768×768 像素时的主要结构和输出特征尺寸。本文在不明显增加计算量的情况下,以ResNeXt101作为特征提取网络(backbone),并将conv2—conv5阶段输出的特征图作为多尺度特征基准。

1.3 可变形卷积

卷积神经网络的特征提取过程主要依赖卷积核实现,传统卷积核的几何结构固定,每次执行卷积操作仅提取固定的局部特征信息。Dai等人(2017)提出可变形卷积网络(deformable convolution networks, DCN),使得卷积核能够产生自由形变以适应几何形变的物体。

表1 ResNet101 与 ResNeXt101 网络结构
Table 1 Structure of ResNet101 and ResNeXt101

阶段	输出/像素	ResNet101(He 等, 2016)	ResNeXt101(Xie 等, 2017)
conv1	384 × 384	7 × 7, 64, stride = 2 3 × 3, maxpool, stride = 2	7 × 7, 64, stride = 2 3 × 3, maxpool, stride = 2
conv2	192 × 192	$\begin{pmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{pmatrix} \times 3$	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3, 128, c = 32 \\ 1 \times 1, 256 \end{pmatrix} \times 3$
conv3	96 × 96	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{pmatrix} \times 4$	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3, 256, c = 32 \\ 1 \times 1, 256 \end{pmatrix} \times 4$
conv4	48 × 48	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1\,024 \end{pmatrix} \times 23$	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512, c = 32 \\ 1 \times 1, 1\,024 \end{pmatrix} \times 23$
conv5	24 × 24	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2\,048 \end{pmatrix} \times 3$	$\begin{pmatrix} 1 \times 1, 1\,024 \\ 3 \times 3, 1\,024, c = 32 \\ 1 \times 1, 2\,048 \end{pmatrix} \times 3$

对几何形状变化较大和不规则的骨病灶直接使用固定尺寸的卷积核较难适应,因此引入 DCN 以适应不同几何形状的目标病灶。

1.4 注意力机制

注意力机制类似于视觉注意力机制,扫描整幅图像找到重点区域以获取更多有关特征信息,忽略无关细节。注意力机制有助于利用有限资源,从大量信息中快速筛选有价值的信息。Cao 等人(2019)提出 GCNet(global context networks),不仅集成了 NLB(non-local block)(Wang 等,2018)的全局能力,且具备 SENet(squeeze-and-excitation networks)(Hu 等,2020)自动调整通道依赖关系和轻量化的特点。

本文在特征提取网络中引入 GC(global context)模块促进网络从特征繁杂的 CT 图像中快速提取骨病变相关特征信息、抑制无关信息,如图 3 所示。

2 数据集与预处理

2.1 实验数据集

实验数据来自美国国立卫生研究院临床中心(National Institutes of Health Clinical Center, NIH-CC)发布的 DeepLesion 数据集。数据集中共有 32 120 幅使用实体肿瘤疗效评价标准(response evaluation criteria in solid tumour, RECIST)(Eisenhauer 等,2008)进行病灶标记的 CT 图像(Yan 等,

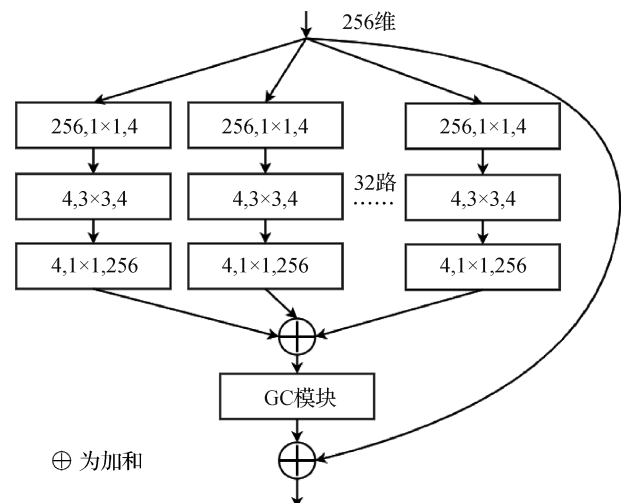


图3 ResNeXt 引入 GC 模块

Fig. 3 Structure of ResNeXt with global context block

2017),已知 30% 数据给定了病变类型,其中 247 幅图像标定为含有骨病变;另外 70% 数据未说明具体病变类型。

为了更好地训练本文网络模型,需要对数据集进行预处理。首先,剔除成像质量太差的 2 幅图像,如图 4(a)(b)所示;其次,挑选更多的骨病变 CT 图像;由于 CT 图像中骨骼区域特征明显,易于区分,据此,本文从 70% 未指明病变类型的数据中挑选出 155 幅病灶标记框位于骨骼区域的图像作为本文实验数据,如图 4(c)(d)所示,绿框即为真实病灶部

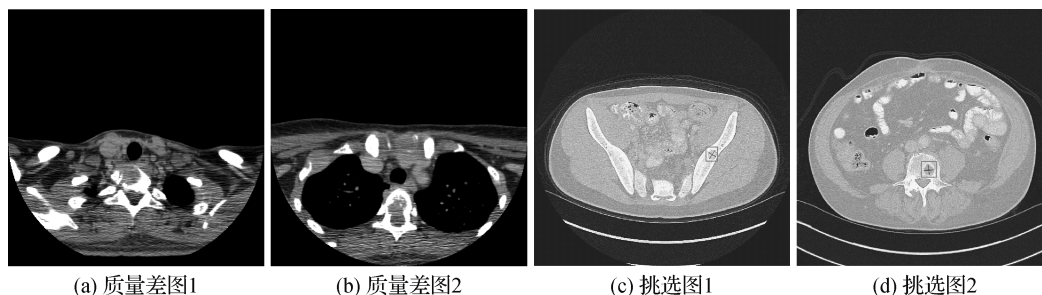


图4 数据筛选示例

Fig.4 Selected samples ((a) poor quality image 1;(b) poor quality image 2;(c) selected image 1;(d) selected image 2)

位标记。因此,本文使用的原始实验数据共由400幅CT图像组成。

2.2 图像转换

数据集中存储的CT图像均为16位单通道PNG(portable network graphics)图像,根据NIHCC提供的标签文件,将16位CT图像转换为CT值数据,再转换为8位图像,转换公式为

$$HU = P_{16} - 32\ 768 \quad (2)$$

$$P_8 = \frac{HU - Lo}{Hi - Lo} \times 255 \quad (3)$$

式中, HU 、 P_{16} 和 P_8 分别表示CT值数据 HU 、原始16位图像 P_{16} 以及转换得到的8位图像 P_8 的像素值, Lo 和 Hi 分别表示标签文件中DICOM(digital imaging and

communications in medicine)窗的最小值和最大值。

原始CT数据直接经过式(2)和式(3)转换得到的8位图像亮度偏高、对比度偏低,直接用于网络训练不利于对病灶区域特征信息的提取。

2.3 图像增强与大致区域提取

图5展示了预处理与骨骼大致区域提取过程。

1)获取有效人体区域。思路是使用图像处理算法滤除CT图像中非人体区域。首先将原图转换为CT值数据,并将其归一化至 $[-1\ 000\ HU, +1\ 000\ HU]$;其次以CT值数据的均值为阈值进行分割,过滤掉部分非必要区域;然后使用中值滤波过滤白噪点;最后使用开运算操作,除去图像中的两条高亮度CT床背板弧线,如图5(c)(d)所示。

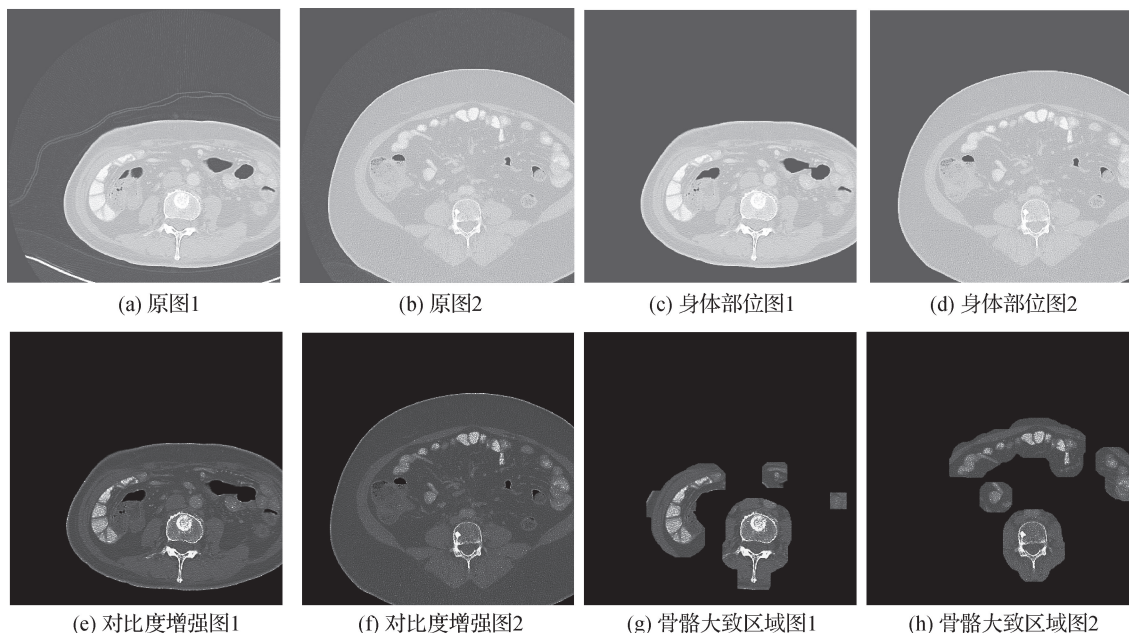


图5 预处理与骨骼区域提取

Fig.5 Preprocessing and bone tissue extraction((a) original image 1;(b) original image 2;(c) body part image 1;(d) body part image 2;(e) contrast enhancement 1;(f) contrast enhancement 2;(g) approximate bone region of image 1;(h) approximate bone region of image 2)

2) 对比度增强。为进一步提升骨骼部位的对比度, 本文提出一种基于伽马变换的改进方法进行对比度增强, 其表达式为

$$y = c \times (1.0 - (1.0 - x)^{\frac{1}{\gamma}}) \quad (4)$$

式中, c 为常数, x 为归一化输入值, γ 为变换因子, y 为结果值, 结果如图 5(e)(f) 所示。

3) 获取大致骨骼区域。根据骨骼 CT 值大致范围(汪家旺等, 2004), 使用阈值分割获取骨骼的大致区域, 结果如图 5(g)(h) 所示, 并以此训练网络。

预处理改善了图像的对比度, 使病灶特征更加明显, 降低了提取有效特征的难度; 同时, 获取骨骼大致区域降低了非有效区域对模型的干扰。

2.4 数据扩增

训练深度学习网络模型, 通常需要使用大量带标签的数据; 本文使用的原始实验数据仅 400 幅 CT 图像, 不足以很好地完成模型训练; 因此需要对数据进行扩增。常见的数据扩增方法有随机翻转、裁剪、平移、缩放、旋转、添加噪声和仿射变换等。先将数据集分别以 85% 和 15% 的比例随机划分为训练集和测试集, 再分别进行 10 倍扩增, 得到训练集 3 400 幅图像、测试集 600 幅图像。

3 网络模型

由于实验数据存在病灶对比度低和尺寸过小等情况, 因此模型在训练中既要克服目标特征不明显的情况, 又要具备对小目标病灶的检测能力。基于实验数据的特征和检测性能需求, 本文以 Cascade R-CNN 为基础框架进行改进。首先选择性能更佳的特征提取网络, 并引入注意力机制和可变形卷积进一步提升特征提取能力; 其次通过特征融合模块增强病灶特征; 然后采用多尺度检测方式提升对不同尺度病灶的适应能力; 最后在 R-CNN 网络的 Head 结构增加卷积操作增强子网络的分类和位置回归性能。

3.1 特征融合

针对浅层特征语义信息较少、几何信息较多, 而深层特征语义信息较多、几何信息粗略的特点, 特征金字塔网络(feature pyramid network, FPN)(Lin 等, 2017)通过自顶向下和横向连接促进不同尺度特征信息融合, 达到多尺度特征融合的效果。PAFPN(path aggregation feature pyramid networks)(Liu 等,

2018)通过增加自底向上的路径, 缩短了从低层特征到高层特征的路径, 降低了特征信息的流动难度。NAS-FPN(neural architecture search-feature pyramid networks)(Ghiasi 等, 2019)通过强化学习自动搜索得到特征融合模型结构, 具有自由跨范围融合特性。通过 PAFPN 模块促进低层特征信息与高层特征信息相互融合, 增强病灶部位特征。如图 6 所示, P_2 — P_6 分别表示不同尺度特征, D_2 — D_6 对应的输出结果, “+”表示特征相加。

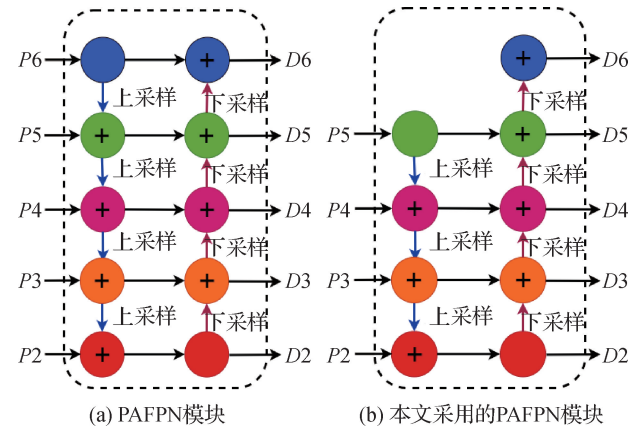


图6 特征融合模块结构图

Fig. 6 Structure of feature fusion block

((a) PAFPN block; (b) our PAFPN block)

3.2 网络结构

本文网络结构如图 7 所示, 主要由特征提取、特征融合、RPN 与 3 个级联 R-CNN 构成。特征提取网络获取多尺度特征信息, 以 ResNeXt101 为特征提取网络, 并引入 DCN 和 GC 模块; 采用迁移学习策略, 冻结阶段 conv1 的预训练网络权重, 对其他层的权重进行微调训练, 输出 conv2—conv5 共 4 个阶段特征, 供后续多尺度特征融合与检测。特征融合部分使用 PAFPN 模块对多尺度特征信息进行融合, 将 4 种不同尺度特征图进行自顶向下、横向连接以及自底向上的特征融合, 并输出 5 种尺度特征。RPN 和级联 R-CNN 网络逐步提升模型的检测精度, 在多尺度特征图上分别生成预测区域, 通过预设的 IoU 阈值进行正负样本分类和位置回归, 并将位置回归结果作为骨病灶的大概率区域传送至下一个 R-CNN 继续训练。

3.3 级联 R-CNN 网络结构

为适应不同尺寸病灶, 分别对 5 种尺度特征图上的目标感兴趣区域(region of interest, ROI)进行 ROI align 操作。R-CNN 结构如图 8 所示, 图中多尺

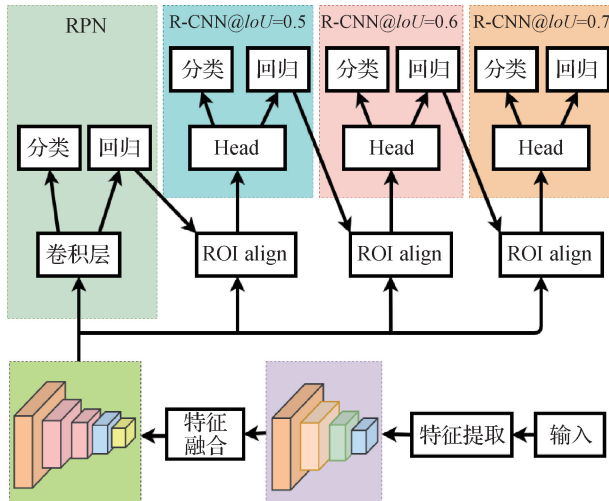


图7 本文模型结构

Fig. 7 Structure of our model

度特征为特征融合模块输出的5个尺度特征图,特征图上的ROI区域表示上一层网络对目标病灶区域的回归结果。首先通过ROI align将不同尺度特征图上满足条件的候选区域变换到相同尺寸;其次进行卷积操作;最后通过两次全连接操作(fully connected, FC)再对结果进行分类和回归。

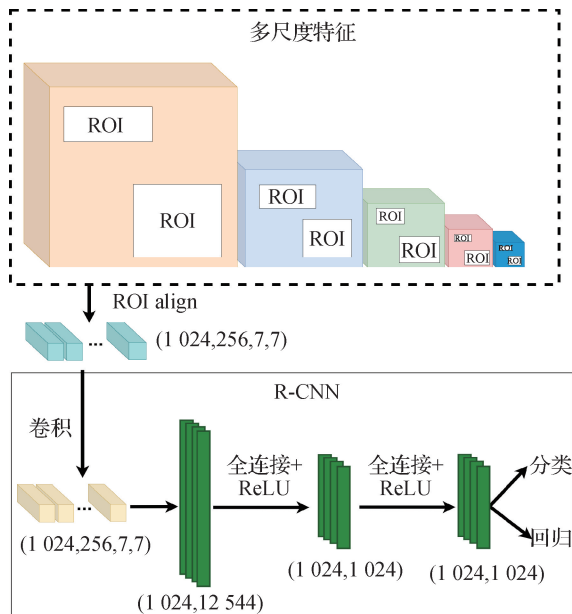


图8 R-CNN结构

Fig. 8 Structure of our R-CNN

4 实验与结果分析

4.1 实验设置

实验使用的深度学习框架为 Pytorch 1.5.0、

mmdetection 2.3.0 (Chen 等, 2020) 和 python3.6, 操作系统为 Ubuntu 18.04 LTS。在 Intel i7-6700K 处理器、16 GB 内存、11 GB 显存 NVIDIA GeForce RTX2080Ti 显卡的设备上完成训练, 由 OpenCV 实现预处理, 数据扩增由 imgaug 库完成。采用 SGD (stochastic gradient descent) 优化器, 初始学习率设置为 0.01, 前 500 次迭代执行学习率线性预热方式, 增长率为 0.001; batch size 为 4、epoch 为 30, 输入图像尺寸设置为 768 × 768 像素。为克服训练样本不足可能带来的训练困难、过拟合以及泛化能力差等问题, 本文实验(包括对照组实验)均使用 MS COCO (Microsoft common objects in context) 预训练权重进行迁移学习训练。

4.2 实验设计

为对比验证不同模块、结构以及训练方法对模型检测性能的影响, 设置 4 组实验。

第 1 组: 为选取性能较优、训练难度适宜的特征提取网络, 对比 3 种典型残差网络并以平均精度 (average precision, AP) 为性能衡量指标进行选取。

第 2 组: 为进一步实现特征信息融合, 提升模型检测性能, 对比使用典型特征融合模块的检测性能指标并选择合适的特征融合模块。

第 3 组: 为验证归一化、注意力机制以及可变形卷积对模型的有效性, 首先在 R-CNN 中使用 BN (batch normalization) 和 GN (group normalization), 对比性能指标; 其次引入注意力机制 GC 和可变形卷积 DCN, 验证其有效性。

第 4 组: 为验证本文模型的有效性, 在相同条件下对 Faster R-CNN (Ren 等, 2017)、Cascade R-CNN (Cai 和 Vasconcelos, 2021)、Libra R-CNN (Pang 等, 2019)、FCOS (fully convolutional one-stage object detection) (Tian 等, 2019)、RetinaNet (Lin 等, 2020) 和本文模型进行训练和测试, 对比各项指标。

4.3 模型评价指标

本文以召回率 (recall)、准确率 (precision)、F1 分数以及平均精度 (AP) (取 $IoU = 0.5$) 作为模型评价指标, 具体为

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

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$F1 = 2 \times \frac{R \times P}{R + P} \quad (7)$$

式中, R 、 P 和 $F1$ 分别表示召回率、准确率和 $F1$ 分数。 TP 表示骨病灶被正确预测, FN 表示将骨病灶错误预测为背景、 FP 表示将背景预测为骨病灶。

AP 用于衡量模型对某一类目标的平均精度; 以积分的方式计算平滑后的 P-R (precision-recall) 曲线 (Padilla 等, 2020) 下方的面积作为最终的 AP 值, 即

$$AP = \int_0^1 P(R) d(R) \quad (8)$$

式中, P 和 R 分别表示准确率和召回率。

4.4 实验结果分析

第 1 组实验的目的为选定适用于本文数据的特征提取网络, 即在 Cascade R-CNN 网络模型中分别以 ResNet50 (He 等, 2016)、ResNet101 (He 等, 2016) 以及 ResNeXt101 (Xie 等, 2017) 为特征提取网络, 以 AP 为评价指标, 进行对比实验, 实验结果如表 2 所示。可以看出, ResNeXt101 对本文数据的特征提取性能优于其他两类, 主要由于 ResNeXt 中平行堆叠的分组卷积拓扑结构代替 ResNet 的残差卷积结构增强了网络的特征提取能力。

表 2 不同特征提取网络的实验结果

Table 2 Experimental results of different backbone

特征提取网络	AP
Cascade R-CNN + ResNet50	0.420
Cascade R-CNN + ResNet101	0.471
Cascade R-CNN + ResNeXt101	0.543

注: 加粗字体为最优结果。

第 2 组实验以前一组实验的结果为基础, 即以 ResNeXt101 为特征提取网络, 通过实验对比验证不含特征融合模块以及分别引入 FPN、PAFPN 和 NAS-FPN 3 种特征融合模块时对本文网络的性能指标的影响。结果如表 3 所示, 可以看出引入 FPN 后模型的检测性能有较大提升, 表明特征融合模块促进了不同尺度特征信息之间的相互作用。对比引入 FPN、PAFPN 以及 NAS-FPN 的实验结果可知, PAFPN 对本文网络的特征信息融合效果优于其他两类, 表明经过两路特征融合之后骨病灶特征更加凸显。

第 3 组实验分为两部分, 第 1 部分验证在 R-CNN 中使用 BN (Ioffe 和 Szegedy, 2015) 与 GN (Wu 和 He, 2018) 对模型检测性能的影响; 第 2 部分验证注意力模块 GC 与可变形卷积 DCN 对模型

检测性能的有效性, 实验结果如表 4 所示。可以看出, 通过 GN 进行归一化使本文模型检测性能略有提升 (第 1、2 行); 而在网络中引入 DCN 与 GC 模块后, 模型检测性能提升效果显著 (第 3、4 行)。

表 3 特征融合模块的实验结果

Table 3 Experimental results of feature fusion modules

特征融合网络	AP
Cascade R-CNN + ResNeXt101	0.543
Cascade R-CNN + ResNeXt101 + FPN	0.655
Cascade R-CNN + ResNeXt101 + PAFPN	0.721
Cascade R-CNN + ResNeXt101 + NAS-FPN	0.671

注: 加粗字体为最优结果。

表 4 不同模块平均精度对比实验结果

Table 4 Comparison of modules' AP

BN	GN	GC	DCN	AP
✓				0.721
	✓			0.723
	✓	✓		0.764
	✓	✓	✓	0.816

注: 加粗字体为最优结果。

第 4 组实验是在相同条件和训练策略下进行, 对比本文模型与 Faster R-CNN (Ren 等, 2017)、RetinaNet (Lin 等, 2020)、FCOS (Tian 等, 2019)、Libra R-CNN (Pang 等, 2019)、Cascade R-CNN (Cai 和 Vasconcelos, 2021) 的实验结果如表 5 所示, 其中, 3D-AG (3D aggregated) 和 ElixirNet 的结果分别来自 Zhang 等人 (2020) 和 Jiang 等人 (2020) 的方法。由于 RetinaNet、FCOS 和 Libra R-CNN 含有 FPN 且均在多尺度特征上进行检测, 因此本组实验模型均引入 FPN 进行横向对比。

本组实验结果分析: 1) 多类型病灶检测网络 3D-AG 和 ElixirNet 对骨组织病变的检测召回率指标低于其他模型的主要原因是一方面二者使用的原始数据中骨病灶 CT 数量过少, 另一方面是为了平衡其他类型病变的检测性能, 丢失了部分特征较弱和尺寸过小的骨病灶的检测能力。2) 召回率、准确率、 $F1$ 分数和 AP 指标表明本文模型对骨病灶的检测性能优于其他模型, 一方面是由于 GC 与 DCN 模块提升了网络模型对形态多变的骨病灶部位特征信息提取能力, 另一方面通过 PAFPN 增强了病灶特征

信息并与搭配的级联 R-CNN 网络逐步提升了模型的检测性能。3) 对比表 5 中第 3、6、7 行的结果, 尽管使用相同的条件及训练策略, 召回率却未获得较大提升, 说明 Faster R-CNN、Cascade R-CNN 和 Libra R-CNN 均存在漏检较多的情况。然而准确率高于第 4 行和第 5 行的单次检测网络, 说明基于候选框的检测算法经过多次分类和回归之后获取了较为准确的位置特征信息。4) 表 5 中 RetinaNet 取得较高的召回率, 可知其检测到较多的真实病灶; 但其准确率偏低, 说明检测结果中含有较多的假阳性病灶。

表 5 不同模型实验结果

Table 5 Experimental results of different models

对比模型	<i>R</i>	<i>P</i>	<i>F1</i>	<i>AP</i>
ElixirNet	0.574	—	—	—
3D-AG	0.700	—	—	—
Faster R-CNN	0.674	0.512	0.581	0.601
RetinaNet	0.828	0.349	0.491	0.712
FCOS	0.751	0.410	0.530	0.610
Cascade R-CNN	0.732	0.529	0.614	0.655
Libra R-CNN	0.817	0.481	0.605	0.707
本文	0.850	0.613	0.712	0.816

注: 加粗字体为各列最优结果, “—”表示无数据。

各对照模型的实验结果 P-R 曲线如图 9 所示, 本文模型 P-R 曲线包围的面积大于其他模型, 结合表 5 中的 F1 分数进行分析, 结果表明本文网络模型的检测性能优于其他对比模型。

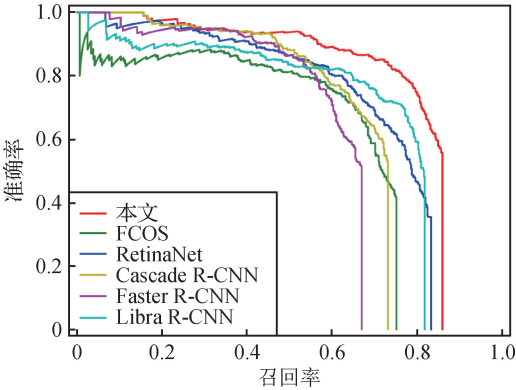


图 9 对照模型 P-R 曲线($IoU=0.5$)

Fig. 9 Precision-recall curve of each model($IoU=0.5$)

4.5 测试效果图

本文和对照组模型的检测结果(任取 3 幅)如

图 10 所示, 绿色框为真实病灶区域, 红色框为模型检测到的病灶部位。从第 1 幅图像的检测结果可知, 仅本文模型和 FCOS 检测到病灶, 本文模型检测结果与标签的重叠和得分均高于后者, 本文模型与 FCOS、Libra R-CNN、RetinaNet 的检测结果中均出现了假阳性, 说明影像中存在与骨病灶特征相似的正常特征导致误检。从第 2 幅图像的检测结果来看, 所有模型均检测到骨病灶, 本文检测结果最优。也可以发现基于候选区域的网络对小目标病灶的检测得分强于单次检测网络。从第 3 幅图像的检测结果来看, 仅 Cascade R-CNN 模型未检测到骨病灶, 单从检测得分上看本文结果次于 Libra R-CNN, 但 Libra R-CNN 的检测结果中出现了假阳性, 其误检率较高。综合检测结果, 表明本文网络模型对骨病灶的检测效果优于其他网络模型。

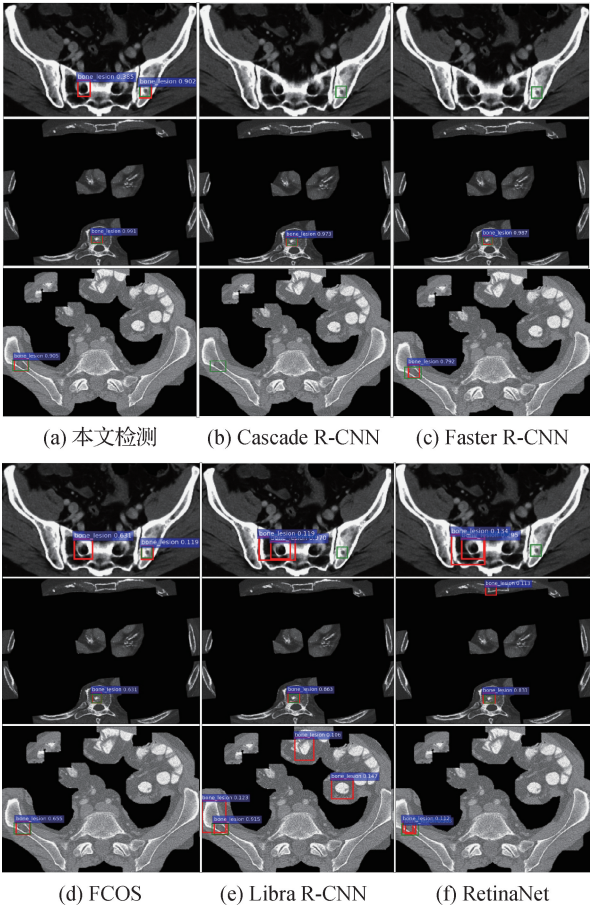


图 10 不同部位骨病灶的检测结果

Fig. 10 Detection results of different bone lesions

((a) ours; (b) Cascade R-CNN; (c) Faster R-CNN; (d) FCOS; (e) Libra R-CNN; (f) RetinaNet)

5 结 论

本文提出一种融合注意力机制和可变形卷积的多尺度级联检测网络模型。首先,通过全局上下文建模注意力机制使得网络自动聚焦目标病灶部位的特征信息,以优化浅层特征的几何细节信息和深层特征的全局信息,增强网络对骨病灶的敏感度。采用可变形卷积方式,使得网络具备覆盖尺寸差异和形态多变的骨病灶的检测能力。其次,根据浅层特征丰富的细节信息和深层特征丰富的语义信息,使用特征融合网络模块促进浅深层特征信息之间的相互融合,进一步凸显骨病灶部位的特征信息。最后,在 RPN 和级联 R-CNN 中分别对每个尺度的特征信息进行分类和回归,以提升网络对不同尺寸骨病灶的检测能力。

通过一系列对比实验验证了本文方法的有效性,召回率、准确率、F1 分数和平均准确率分别为 0.85、0.613、0.712 和 0.816,优于 Zhang 等人(2020)和 Jiang 等人(2020)的检测方法。本文方法的不足之处是对特征不明显的小尺寸病灶和大病灶内含有小病灶特征的骨病变会发生漏检和误检。

下一步的工作包括:1) 建立相邻 CT 扫描断面之间的上下文关系,使得模型能够学习到更多骨病灶邻近的特征信息,降低特征不明显和尺寸过小的骨病灶的漏检;2) 使用弱监督学习方式从原始数据集中自动获取到更多的骨病变类型 CT 图像数据,进一步增强模型的泛化性能。

参考文献 (References)

Cai Z W and Vasconcelos N. 2021. Cascade R-CNN; high quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5): 1483-1498 [DOI: 10.1109/TPAMI.2019.2956516]

Cao Y, Xu J R, Lin S, Wei F Y and Hu H. 2019. GCNet; non-local networks meet squeeze-excitation networks and beyond [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1904.11492.pdf>

Chen K, Wang J Q, Pang J M, Cao Y H, Xiong Y, Li X X, Sun S Y, Feng W S, Liu Z W, Xu J R, Zhang Z, Cheng D Z, Zhu, C C, Cheng, T H, Zhao Q J, Li B Y, Lu X, Zhu R, Wu Y, Dai J F, Wang J D, Shi J P, Ouyang W L, Loy C and Lin D H. 2020. MM-detection [CP/OL]. [2020-08-08]. <https://github.com/open-mmlab/mmdetection>

Dai J F, Qi H Z, Xiong Y W, Yi L, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 3 [DOI: 10.1109/ICCV.2017.89]

Dong Y, Cui J F and Li X H. 2017. CT Diagnosis of Musculoskeletal System. Beijing: Science Press (董越, 崔久法, 李小龙. 2017. 骨关节肌肉系统 CT 诊断. 北京: 科学出版社)

Eisenhauer E, Therasse P, Bogaerts J, Schwartz L, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubenstein L, Shankar L, Kaplan R, Lacombe D and Verweij J. 2008. 32 INVITED new response evaluation criteria in solid tumors: revised RECIST guideline version 1.1. *European Journal of Cancer Supplements*, 6(12): #13 [DOI: 10.1016/S1359-6349(08)71964-5]

Ghiasi G, Lin T Y, Pang R M, Le Q V and Brain G. 2019. NAS-FPN: learning scalable feature pyramid architecture for object detection [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1904.07392.pdf>

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, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hu J, Shen L, Albanie S, Sun G and Wu E H. 2020. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023 [DOI: 10.1109/TPAMI.2019.2913372]

Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift//Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France: JMLR: 448-456

Jiang C H, Wang S J, Xu H, Liang X D and Xiao N. 2020. ElixirNet: relation-aware network architecture adaptation for medical lesion detection [EB/OL]. [2020-03-03]. <https://arxiv.org/pdf/2003.08770.pdf>

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: 2117-2125 [DOI: 10.1109/CVPR.2017.106]

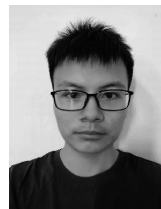
Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2020. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2): 318-327 [DOI: 10.1109/TPAMI.2018.2858826]

Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]

Padilla R, Netto S L and da Silva E A B. 2020. A survey on performance metrics for object-detection algorithms//Proceedings of 2020

- International Conference on Systems, Signals and Image Processing. Niteroi, Brazil; IEEE; 2157-870 [DOI: 10.1109/IWSSIP48289.2020.9145130]
- Pang J M, Chen K, Shi J P, Feng H J, Ouyang W L and Lin D H. 2019. Libra R-CNN: towards balanced learning for object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE [DOI: 10.1109/CVPR.2019.00091]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 779-788 [DOI: 10.1109/CVPR.2016.91]
- 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]
- Tian Z, Shen C H, Chen H and He T. 2019. FCOS: fully convolutional one-stage object detection [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1904.01355.pdf>
- Wang J W, Wang D H, Zhang L L, Yu T F and Weng X J. 2004. Relationship between bone CT value and bone's minim element. Chinese Journal of Medical Imaging Technology, 20(9): 1328-1330 [汪家旺, 王德杭, 张廉良, 俞同福, 翁学军. 2004. 骨组织 CT 值与骨结构成分间的关系研究. 中国医学影像技术, 20(9): 1328-1330] [DOI: 10.3321/j.issn:1003-3289.2004.09.006]
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local Neural Networks [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1711.07971.pdf>
- Wu Y X and He K M. 2018. Group normalization [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1803.08494.pdf>
- Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1611.05431.pdf>
- Yan K, Peng Y F, Sandfort V, Bagheri M, Lu Z Y and Summers R M. 2019. Holistic and comprehensive annotation of clinically significant findings on diverse CT images: learning from radiology reports and label ontology [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1904.04661.pdf>
- Yan K, Wang X S, Lu L and Summers R M. 2017. DeepLesion: automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/1710.01766.pdf>
- Zhang N, Cao Y, Liu B Y and Luo Y. 2020. 3D aggregated faster R-CNN for general lesion detection [EB/OL]. [2020-07-20]. <https://arxiv.org/pdf/2001.11071.pdf>
- Zhao H F, Gao P R, Li H L and Qu J R. 2019. CT and MRI features of the malignant tumors of the sacrum. Journal of Practical Medical Imaging, 20(6): 607-609. [赵浩锋, 高朋瑞, 黎海亮, 曲金荣. 2019. 骶骨恶性肿瘤的 CT 及磁共振成像表现. 实用医学影像杂志, 20(6): 607-609] [DOI: 10.16106/j.cnki.cn14-1281/r.2019.06.023]
- Zhong J P. 2009. Chinese Yearbook of Surgery. Shanghai: Second Military Medical University Press (仲剑平. 2009. 中国外科年鉴. 上海: 第二军医大学出版社)

作者简介



方成, 1989 年生, 男, 硕士研究生, 主要研究方向为基于深度学习的医疗图像处理。

E-mail: 281011824@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