



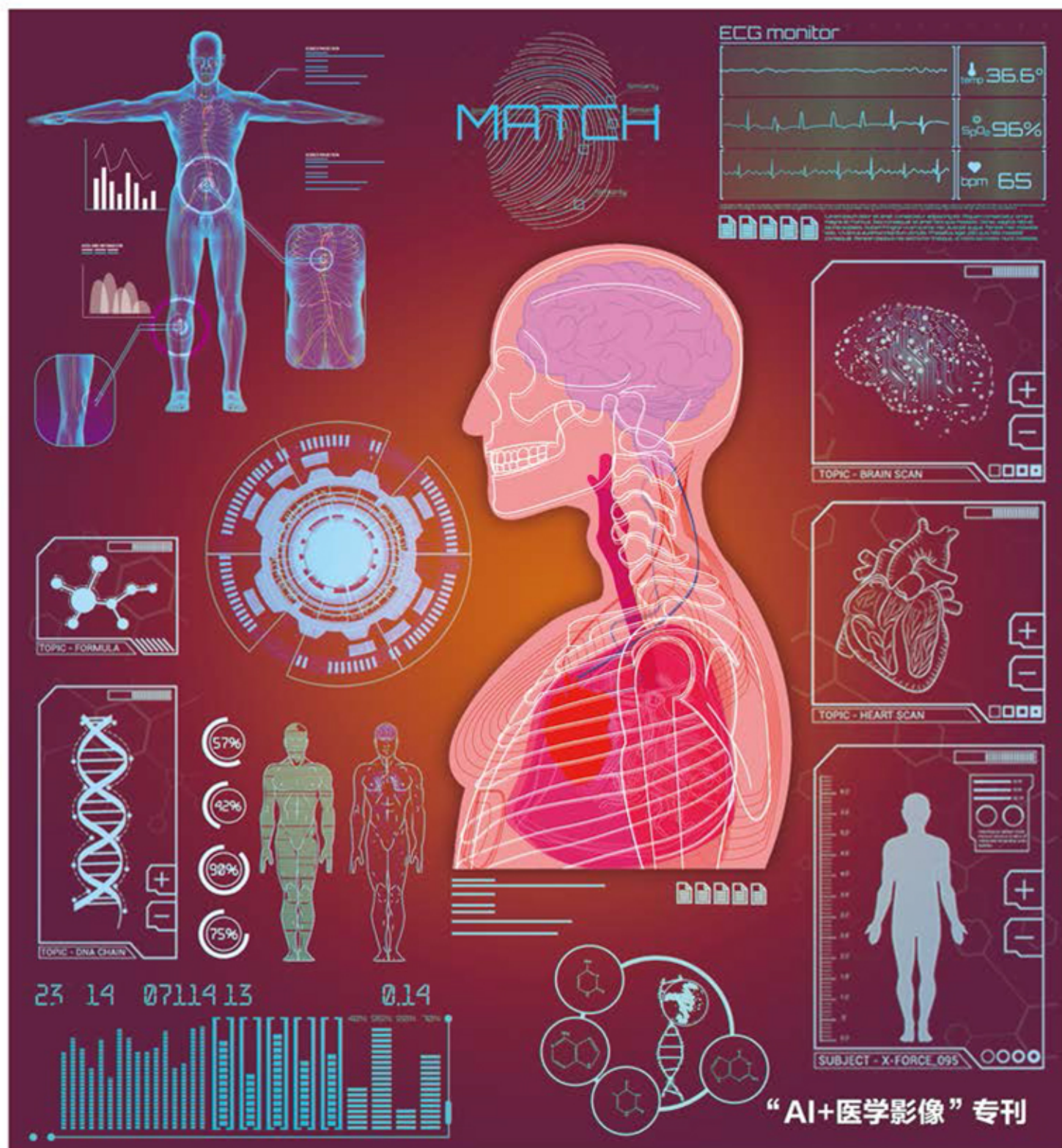
JOURNAL OF IMAGE AND GRAPHICS

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

2020
10
VOL.25

ISSN1006-8961
CN11-3758/TB





第25卷第10期（总第294期）

2020年10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 顾行发

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@radi.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 Gu Xingfa

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@radi.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

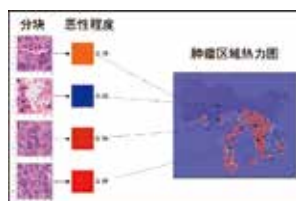
ISSN 1006-8961

CODEN ZTTXFZ

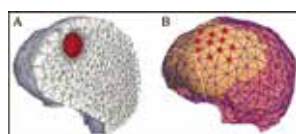
国外发行代号 M1406

国内邮发代号 82-831

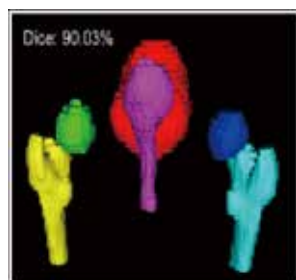
国内定价 60.00元



深度学习在组织病理学中的应用综述(第1982页)



机器学习在术中光学成像技术中的应用研究(第1994页)



2D级联CNN模型的放疗危及器官自动分割(第2110页)

序言	安虹, 张云泉 I
编者按	II

综述

深度学习在医学影像中的应用综述

施俊, 汪琳琳, 王珊珊, 陈艳霞, 王乾, 魏冬铭, 梁淑君, 彭佳林, 易佳锦, 刘盛锋, 倪东, 王明亮, 张道强, 沈定刚	1953
---	------

深度学习在组织病理学中的应用综述

金旭, 文可, 吕国锋, 石军, 迟孟贤, 武铮, 安虹	1982
------------------------------------	------

机器学习在术中光学成像技术中的应用研究

张崇, 王坤, 田捷	1994
------------------	------

前沿进展

医学3D计算机视觉: 研究进展和挑战

杨健程, 倪冰冰	2002
----------------	------

点云算法在医学领域的研究进展

李美佳, 于泽宽, 刘晓, 颜荣耀, 于媛媛, 王大明, 陈涓, 陆军, 祁鹏, 王俊杰, 刘杰	2013
--	------

肝脏肿瘤CT图像深度学习分割方法综述

马金林, 邓媛媛, 马自萍	2024
---------------------	------

磁共振影像深度学习在精神分裂症诊断中的应用综述

宋佩伦, 王雅萍, 耿秀娟, 张航, 宋学勤	2047
------------------------------	------

新型冠状病毒肺炎(COVID-19)医学影像AI诊断研究进展

孟琢, 李铭辉	2058
---------------	------

等强度婴儿脑MR图像分割的深度学习方法综述

张航, 王雅萍, 耿秀娟, 付鹏飞	2068
-------------------------	------

医学影像疾病诊断的残差神经网络优化算法研究进展

周涛, 霍兵强, 陆惠玲, 师宏斌	2079
-------------------------	------

IVIM及纹理分析在术前预测宫颈癌类型和淋巴结转移研究进展

李翠平, 董江宁	2093
----------------	------

4D时空纵向分析在生物医学领域中的应用现状与趋势

徐春园, 曾晓天, 宋泽雨, 唐晓英	2100
--------------------------	------

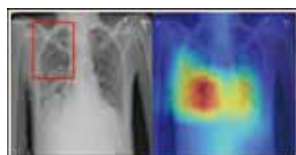
计算机断层扫描图像

2D级联CNN模型的放疗危及器官自动分割

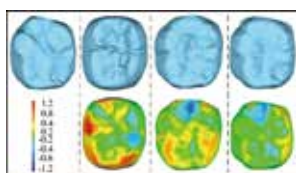
石军, 赵敏帆, 薛旭东, 郝晓宇, 金旭, 安虹, 张红雁	2110
--------------------------------------	------

融合双注意力机制3D U-Net的肺肿瘤分割

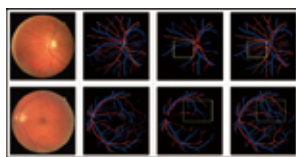
郝晓宇, 熊俊峰, 薛旭东, 石军, 文可, 韩文廷, 李尧扬, 赵俊, 傅小龙	2119
--	------



密集挤压激励网络的多标签胸部X光片疾病分类
(第2238页)



高分辨率深度生成网络的缺失牙体形态设计
(第2249页)



语义融合眼底图像动静脉分类方法(第2259页)

深度学习结合影像组学的肝脏肿瘤CT分割

刘云鹏, 刘光品, 王仁芳, 金冉, 孙德超, 邱虹, 董晨, 李瑾, 洪国斌 2128

超级计算支撑的新冠肺炎CT影像综合分析辅助系统应用

康波, 郭佳, 王帅, 徐波, 孟祥飞 2142

鼻咽癌原发肿瘤放疗靶区的自动分割

薛旭东, 郝晓宇, 石军, 丁轶, 魏伟, 安虹 2151

磁共振图像

组卷积轻量级脑肿瘤分割网络

赵奕名, 李锦, 关欣 2159

深度迭代融合的脑部磁共振图像颅骨去除网络

姚发展, 李智, 王丽会, 程欣宇, 张健 2171

融合密集残差块和GAN变体的医学图像非刚性配准

张桂梅, 胡强, 龚磊 2182

超声图像

结合分段频域和局部注意力的超声甲状腺分割

胡屹杉, 秦品乐, 曾建潮, 柴锐, 王丽芳 2195

面向乳腺超声图像分割的混合监督双通道反馈U-Net

贡荣麟, 施俊, 王骏 2206

乳腺超声双模态数据的协同约束网络

杨子奇, 龚勋, 朱丹, 郭颖 2218

乳腺超声肿瘤动态噪声指数及分割算法

邹海鹏, 龚勋, 胡毕杰, 罗俊 2229

研究应用

密集挤压激励网络的多标签胸部X光片疾病分类

张智睿, 李锦, 关欣 2238

高分辨率深度生成网络的缺失牙体形态设计

郭闯, 戴宁, 田素坤, 孙玉春, 俞青, 刘浩, 程筱胜 2249

语义融合眼底图像动静脉分类方法

高颖琪, 郭松, 李宁, 王恺, 康宏, 李涛 2259

AS-PANet:改进路径增强网络的重叠染色体实例分割

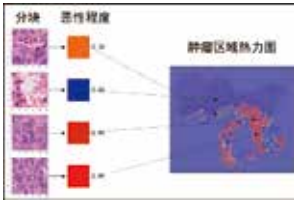
林成创, 赵淦森, 尹爱华, 丁笔超, 郭莉, 陈汉彪 2271

TSCNN:面向可穿戴心电信号监测与分析的卷积神经网络

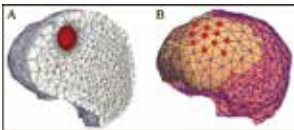
孟琨, 葛康, 宋阳, 杨东溟 2281

CONTENTS

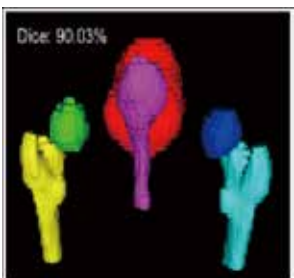
JOURNAL OF IMAGE AND GRAPHICS



A Survey on Applications of Deep Learning in Histopathology(P1982)



Review: the application of machine learning in intraoperative optical imaging technologies(P1994)



Automatic segmentation of organs-at-risk in radiotherapy using 2D cascade-CNN model(P2110)

Review

Applications of deep learning in medical imaging: a survey

Shi Jun, Wang Linlin, Wang Shanshan, Chen Yanxia, Wang Qian, Wei Dongming, Liang Shujun, Peng Jialin, Yi Jiajin, Liu Shengfeng, Ni Dong, Wang Mingliang, Zhang Daoqiang, Shen Dinggang 1953

Survey on the applications of deep learning to Histopathology

Jin Xu, Wen Ke, Lyu Guofeng, Shi Jun, Chi Mengxian, Wu Zheng, An Hong 1982

Review: the application of machine learning in intraoperative optical imaging technologies

Zhang Chong, Wang Kun, Tian Jie 1994

Frontier

Advances and challenges in medical 3D computer vision

Yang Jiancheng, Ni Bingbing 2002

Progress of point cloud algorithm in medical field

Li Meijia, Yu Zekuan, Liu Xiao, Yan Rongyao, Yu Yuanyuan, Wang Daming, Chen Juan, Lu Jun, Qi Peng, Wang Junjie, Liu Jie 2013

Review of deep learning segmentation methods for CT images of liver tumors

Ma Jinlin, Deng Yuanyuan, Ma Ziping 2024

Survey of deep learning in MRI-based diagnosis of schizophrenia

Song Peilun, Wang Yaping, Geng Xiujuan, Zhang Hang, Song Xueqin 2047

Progress of artificial intelligence diagnosis and Prognosis technology for COVID-19 medical imaging

Meng Lu, Li Ronghui 2058

Review of deep learning methods for isointense infant brain MR image segmentation

Zhang Hang, Wang Yaping, Geng Xiujuan, Fu Pengfei 2068

Progress of residual neural network optimization algorithm for medical imaging disease diagnosis

Zhou Tao, Huo Bingqiang, Lu Huiling, Shi Hongbin 2079

Advances in preoperative identification of subtypes and lymph node metastasis of cervical cancer by IVIM and texture analysis

Li Cuiping, Dong Jiangning 2093

Research status and trend of 4D spatiotemporal longitudinal analysis in biomedical field

Xu Chunyuan, Zeng Xiaotian, Song Zeyu, Tang Xiaoying 2100

Computerd Tomography Image

Automatic segmentation of organs-at-risk in radiotherapy using 2D cascade-CNN model

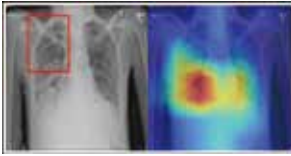
Shi Jun, Zhao Minfan, Xue Xudong, Hao Xiaoyu, Jin Xu, An Hong, Zhang Hongyan 2110

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

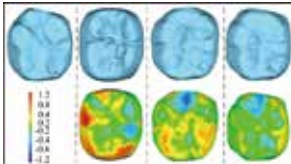
Hao Xiaoyu, Xiong Junfeng, Xue Xudong, Shi Jun, Wen Ke, Han Wenting, Li Xiaoyang, Zhao Jun, Fu Xiaolong 2119

Accurate segmentation method of liver tumor CT based on the combination of deep learning and radiomics

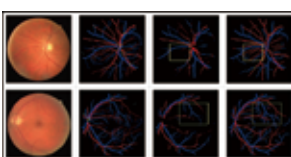
Liu Yunpeng, Liu Guangpin, Wang Renfang, Jin Ran, Sun Dechao, Qiu Hong, Dong Chen, Li Jin, Hong Guobin 2128



Multilabel chest X-ray disease classification based on a dense squeeze-and-excitation network (P2238)



Morphological design of missing tooth driven by high-resolution deep generation network(P2249)



Arteriovenous classification method in fundus images based on semantic fusion(P2259)

Supercomputing-supported COVID-19 CT image comprehensive analysis assistant system

Kang Bo, Guo Jia, Wang Shuai, Xu Bo, Meng Xiangfei 2142

Auto-segmentation of high-risk primary tumor gross target volume for the radiotherapy of nasopharyngeal carcinoma

Xue Xudong, Hao Xiaoyu, Shi Jun, Ding Yi, Wei Wei, An Hong 2151

Magnetic Resonance Image

Lightweight brain tumor segmentation algorithm based on a group convolutional neural network

Zhao Yiming, Li Qiang, Guan Xin 2159

Deep iterative fusion network on skull removal of brain magnetic resonance images

Yao Fazhan, Li Zhi, Wang Lihui, Cheng Xinyu, Zhang Jian 2171

Non-rigid medical image registration based on residual-in-residual dense block and GAN

Zhang Guimei, Hu Qiang, Gong Lei 2182

Ultrasound Image

Ultrasound thyroid segmentation based on segmented frequency domain and local attention

Hu Yishan, Qin Pinle, Zeng Jianchao, Chai Rui, Wang Lifang 2195

Hybrid supervised dual-channel feedback U-Net for segmentation of breast ultrasound images

Gong Ronglin, Shi Jun, Wang Jun 2206

Cooperative suppression network for bimodal data in breast cancer classification

Yang Ziqi, Gong Xun, Zhu Dan, Guo Ying 2218

Dynamic noise index and segmentation algorithm for breast ultrasound tumor

Zou Haipeng, Gong Xun, Hu Bijie, Luo Jun 2229

Research and Application

Multilabel chest X-ray disease classification based on a dense squeeze-and-excitation network

Zhang Zhirui, Li Qiang, Guan Xin 2238

Morphological design of missing tooth driven by high-resolution deep generation network

Guo Chuang, Dai Ning, Tian Sukun, Sun Yuchun, Yu Qing, Liu Hao, Cheng Xiaosheng 2249

Arteriovenous classification method in fundus images based on semantic fusion

Gao Yingqi, Guo Song, Li Ning, Wang Kai, Kang Hong, Li Tao 2259

AS-PANet: a chromosome instance segmentation method based on improved path aggregation network architecture

Lin Chengchuang, Zhao Gansen, Yin Aihua, Ding Bichao, Guo Li, Chen Hanbiao 2271

TSCNN: a convolutional neural network for the monitoring and analysis of the electrical signals of the heart from wearable devices

Meng Lu, Ge Kang, Song Yang, Yang Dongming 2281

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2020)10-2218-11

论文引用格式: Yang Z Q, Gong X, Zhu D and Guo Y. 2020. Cooperative suppression network for bimodal data in breast cancer classification. Journal of Image and Graphics, 25(10): 2218-2228 (杨子奇, 龚勋, 朱丹, 郭颖. 2020. 乳腺超声双模态数据的协同约束网络. 中国图象图形学报, 25(10): 2218-2228) [DOI:10.11834/jig.200246]

乳腺超声双模态数据的协同约束网络

杨子奇¹, 龚勋¹, 朱丹¹, 郭颖²

1. 西南交通大学信息科学与技术学院, 成都 610031; 2. 华北理工大学附属医院, 唐山 063000

摘要: **目的** 通过深度学习对乳腺癌早期的正确诊断能大幅提高患者生存率。现阶段大部分研究者仅采用 B 型超声图像作为实验数据, 但是 B 型超声自身的局限性导致分类效果难以提升。针对该问题, 提出了一种综合利用 B 型超声和超声造影视频来提高分类精度的网络模型。**方法** 针对 B 型超声图像及造影视频双模态数据的特性设计了一个双分支模型架构。针对传统提取视频特征中仅使用单标签的不足, 制定了病理多标签预训练。并设计了一种新的双线性协同机制, 能更好地融合 B 型超声和超声造影的特征, 提取其中的病理信息并抑制无关噪声。**结果** 为了验证提出方法的有效性, 本文设计了 3 个实验, 前两个实验分别对 B 型超声和超声造影进行预训练。在造影分支, 使用根据医学领域设计的病理多标签进行预训练。最后, 采用前两个实验的预训练模型进行第 3 个实验, 相比单独使用 B 型超声图像精度提升 6.5%, 比单用超声造影视频精度提高 7.9%。同时, 在使用双模态数据里, 本文方法取得了最高精度, 相比排名第 2 的成绩提高了 2.7%。**结论** 本文提出的协同约束网络, 能对不同模态的数据进行不同处理, 以提取出其中的病理特征。一方面, 多模态数据确实能从不同角度展示同一个病灶区, 为分类模型提供更多的病理特征, 进而提高模型的分类精度。另一方面, 合适的融合方式也至关重要, 能最大程度地利用特征并抑制噪声。

关键词: 超声造影; 双分支网络; 双线性融合; 协同注意力; 乳腺癌分类

Cooperative suppression network for bimodal data in breast cancer classification

Yang Ziqi¹, Gong Xun¹, Zhu Dan¹, Guo Ying²

1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China;

2. North China University of Science and Technology Affiliated Hospital, Tangshan 063000, China

Abstract: **Objective** Computer-aided breast cancer diagnosis is a fundamental problem in the field of medical imaging. Correct diagnosis of breast cancer through deep learning can immensely improve the patients' survival rate. At present, most researchers only use B-mode ultrasound images as experimental data, but the limitation of B-mode ultrasound data makes it difficult to achieve a high classification accuracy. With the development of medical images, contrast-enhanced ultrasound (CEUS) video can provide accurate pathological information by observing the dynamic enhancement of the lesion area in temporal sequence. In view of the above ultrasound image problems, this paper proposes a network model that can comprehensively utilize B-mode ultrasound video data and CEUS video data to improve the classification accuracy. **Method** First, a dual-branch model architecture is designed on the basis of the characteristics of two-stream structure and dual-mo-

收稿日期: 2020-05-31; 修回日期: 2020-07-07; 预印本日期: 2020-07-14

基金项目: 国家自然科学基金项目(61876158); 四川省重点研发项目(2019YFS0432)

Supported by: National Natural Science Foundation of China (61876158); Sichuan Science and Technology Program (2019YFS0432)

dal data. One branch uses a frame of B-mode ultrasound video data and Resnet34 network model to extract pathological features. The other branch uses ultrasound contrast data and R (2 + 1) network model to extract temporal sequence information. Second, pathological multilabel pretraining is designed in this branch using 10 pathological information in CEUS video data because of the shortcoming of traditional video feature extraction. After the two-branch network, the characteristics of B-mode ultrasound data and CEUS video data are obtained. We perform bilinear fusion on the obtained features to better integrate the features of B-mode ultrasound and CEUS. To extract pathological information and suppress irrelevant noise, the extracted and fused features from the two-branch network are processed using the attention mechanism to obtain the attention weight of the corresponding feature, and the corresponding weight is applied to the original feature. Weighted ultrasound and contrast features are obtained. Finally, the features obtained through the attention mechanism are bilinearly fused to obtain the final features. **Result** This article designed three experiments, where the first two experiments are pretraining on B-mode ultrasound and CEUS to verify the effectiveness of the proposed method and select the network with the strongest feature extraction ability for ultrasound data. In the B-mode ultrasound data pretraining experiment, the classic VGG (visual geometry group) 16-BN (batch normalization), VGG19-BN, ResNet13, ResNet34, and ResNet50 networks were selected as the backbone network of the ultrasound branch for training to select the network with the strongest extraction ability for ultrasound images. The final classification results of each network are 74.2%, 75.6%, 80.5%, 81.0%, and 92.1%. Considering that the accuracy of the Resnet50 network in the test set is only 79.3%, which is relatively different from the accuracy of the training set and resulting in serious overfitting, the Resnet34 network is used as the backbone network of B-mode ultrasound data. In the pretraining experiment of the CEUS branch, the current mainstream P3D, R3D, CM3, and R (2 + 1) D convolutional networks are used as the backbone network of the CEUS branch for training. The final classification results of each network are 75.2%, 74.6%, 74.1%, and 78.4%, and the R (2 + 1) D network with better results in the experiment is selected as the backbone network of the CEUS branch. Pretraining using pathological multilabels is designed in accordance with the medical field. The accuracy of the experiment combining the two data is improved by 6.5% compared with the use of B-mode ultrasound images alone and improved by 7.9% compared with the single-use CEUS video. At the same time, the proposed method achieves the highest accuracy in the use of bimodal data, which increases by 2.7% compared with the highest score. **Conclusion** The proposed cooperative suppression network can process different modal data differently to extract the pathological features. On the one hand, multimodal data can certainly display the same lesion area from different angles, providing many pathological features for the classification model, thereby improving its classification accuracy. On the other hand, a proper fusion method is crucial because it can maximize the use of features and suppress noise.

Key words: contrast-enhanced ultrasound (CEUS); double-branch network; bilinear fusion; collaborative attention; breast cancer classification

0 引言

乳腺癌是女性最常见的疾病,也是造成女性病亡的主要原因之一。据统计结果显示,乳腺癌位居中国癌症病死率的前5名(Chen等,2015)。美国癌症协会发布的统计结果显示,乳腺癌的发病率呈逐年上升和年轻化的趋势(Wild,2014)。在早期及时发现乳腺癌可以提高患者的生存率,因此早期的准确诊断起着至关重要的作用。其中B型超声诊断技术因其具备安全性、低成本和高通用性等特点广泛用于早期乳腺癌的检测。但是,其诊断准确性取决于超声医师的经验,不同级别的医师之间的诊断

差异可能会超过30%(Hooley等,2013)。

对于当前影像医师人力缺失以及诊断严重依赖操作者经验的问题,计算机辅助诊断被视为其解决的重要办法。目前该领域的研究主要有手工提取特征的传统方法和自动学习特征的深度学习方法。与传统方法相比,深度学习在图像识别中有着极强的自动提取特征能力,可以解决对手工特征提取的依赖,获得了许多研究人员的关注。通过一种完全计算机化的系统(基于ANN(artificial neural network)),将超声图像与乳腺结构信息相结合来识别和区分良性和恶性乳腺癌病例(Mohammed等,2018)。基于现有CNN(convolutional neural network)网络对乳腺超声图像进行乳腺病变检测和分类(Cao

等,2019),在分类方法中,为了分析预训练对分类结果的影响,比较了4种情况(随机初始化的ROI(region of interest)、迁移学习的ROI、随机初始化的全尺寸图像和迁移学习的全尺寸图像)下基于不同CNN网络分类方法的性能。基于一种使用多尺度内核和跳跃连接的深度卷积神经网络来诊断乳腺超声图像(Qi等,2019)。该网络由两部分组成,用于识别恶性肿瘤并以级联方式识别结节,从而提高分类精度。此外,结合传统方法与基于GAN(generative adversarial network)的方法进行数据增强后(Al-Dhabyani等,2019),使用卷积神经网络方法和迁移学习方法进行实验,一定程度上能缓解当前医学研究小样本的困境。乳腺癌分类在深度学习下取得了不错的进展,但是目前的研究大都是针对B型超声图像直接迁移自然图像的分类方法。然而自然数据与医学影像差异较大,导致网络泛化能力不强,同时小概率情况下B型超声会产生伪像及伪影,不能展示病灶区血流信息等局限性(Cantisani等,2014),导致基于B型超声的深度学习方法难以进一步提升精度。随着医

学影像技术的发展,超声造影(contrast-enhanced ultrasound, CEUS)视频可以通过观察病变区域在时间序列上的动态增强而提供更精确的病理信息,逐渐成为比传统B型超声更有效的临床诊断技术(Wubulihasimu等,2018)。本文尝试同时利用B型超声和超声造影两种模态数据以提高乳腺癌的分类精度。

图1是双模态数据的一个样本,图1(a)为B型超声视频的图例;图1(b)为CEUS视频的图例,从左到右每幅图像为对应视频的一帧。根据超声成像的特点,本文使用亮度值来量化帧间的差异。对不同模态下视频每一个像素的亮度值进行了3维可视化。从图中可以发现B型超声视频的亮度值在时间维度上仅略有波动。而超声造影视频周围正常组织中的亮度值只是细微的波动,但中心病变组织中的亮度值波动很大。因此,B型超声数据帧间差异小,包含了大量的空间信息,而超声造影数据帧间差异大,具有时间特征。B型超声和超声造影数据是病灶区不同角度的展示,如何合理地从这两种数据中提取对应的特征是提高分类精度的关键。

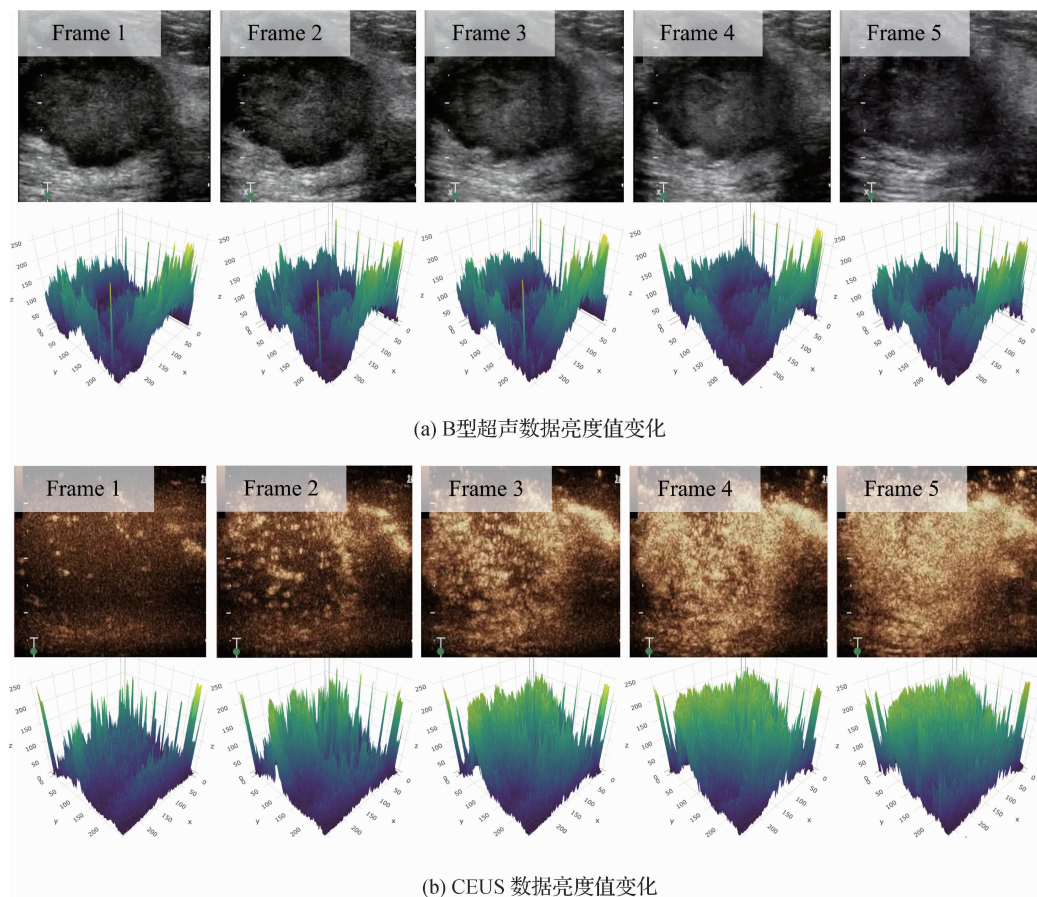


图1 双模态数据图和亮度值图

Fig. 1 Comparison chart of dual-mode data and brightness value

((a) brightness value of B-ultrasound video; (b) brightness value of CEUS video)

目前,在基于两种不同数据的视频分类方法中,双流法是常见的一种。双流法(Simonyan 和 Zisserman, 2014)中的一个流学习单个视频帧的空间上下文,另一个流从堆叠的视频光流中建模运动特征。之后,改进双流网络的方法不断提出。Feichtenhofer 等人(2016)使用不同的融合策略,分析了两个分支网络的性能差异。时间段网络(Wang 等, 2016)通过将一个长视频分成 n 个段,然后分别放入两个流中,最后整合 n 个段的特征进行预测。为了使网络学习时域中不同帧之间的相关性,从而使网络易于识别主要动作,Zhou 等人(2018)提出了时间关系网络(temporal relation network, TRN)。双流法既能保留不同数据类型的个性,又能提取其中的共性。因此广泛应用于适合双模态数据的分类。基于此,本文在双流法的基础上设计了双分支网络。

超声造影作为一种新模态的视频数据,如何理解其中的信息至关重要。目前,3D 卷积是视频理解的一种主流方法。Carreira 和 Zisserman (2017) 受到 Inception-V1 (Szegedy 等, 2015) 的启发,提出了 I3D (inflated 3D convnet) 网络,该网络在每个初始模块中使用不同大小的 3D 卷积内核,并且使用 $1 \times 1 \times 1$ 卷积核进行降维。为了减少参数量大小,Qiu 等人(2017)提出了一个伪 3D 残差网络(P3D ResNet),该网络使用大小为 $1 \times 3 \times 3$ 的 2D 空间卷积和 $3 \times 1 \times 1$ 的 1D 时间卷积,而不是大小为 $3 \times 3 \times 3$ 的 3D 卷积,最终在减少参数量的同时取得更好的效果。Tran 等人(2018)通过 3D CNN 重新审视了时间推理在动作识别中的作用,基于此设计了一个新的时空卷积块 $R(2+1)D$ 网络,该网络将 3D 卷积滤波器分解为独立的空间和时间成分以显著提高准确性。研究者们通过 Action-VLAD (vector of aggregate locally descriptor) (Girdhar 等, 2017) 代替传统的平均池化和最大池化,该方法可以在整个视频中汇总有关场景外观和人员动作,不必将每个帧都唯一地分配给单个帧行动。由于超声造影视频更多是亮度值的变化而非自然物体的运动,因此传统的 3D 卷积并不适合该数据,而 $2D+1D$ 的方式更能从时间维度体现亮度值的变化,因此,本文采用了 $R(2+1)D$ 提取超声造影中的时间特征。

双模态数据分别具有空间和时间的特性,合适的融合方式能充分利用不同数据包含的信息。当前特征融合方式多种多样,Two-Stream 网络(Simonyan

和 Zisserman, 2014) 采用 add 融合将两个特征连接,达到增加通道数的效果。TRN 网络(Zhou 等, 2018) 模型使用 concat 融合将两个特征向量组合成复向量,从而增加每一维信息量的效果。在细粒度图像分类中,为了学到两个不同网络分支中的位置信息和外观信息,使用了双线性融合方法融合不同的分支(Lin 等, 2015)。此外,注意力机制可以使神经网络关注于对分类结果更有帮助的信息,并抑制其中的噪声信息。在之后的研究中,Wang 等人(2017)首次成功将深度卷积网络与人类视觉注意力机制进行有效结合,通过堆叠多个注意力模块构建注意力残差网络。SENet (squeeze-and-excitation network) (Hu 等, 2020) 显式地建模特征通道之间的相互依赖关系。通过对文本和问句生成注意力权值,提出了协同注意力(Lu 等, 2016)。基于 B 型超声和超声造影数据的特殊性,结合当前领域的方法,本文设计了一种双线性协同融合机制,尽可能地提取双模态数据中各自有用的信息以提升乳腺癌的分类精度。

1 方 法

当前,联合超声数据和造影数据已成为乳腺癌诊断的通用方法。造影数据和超声数据作为表现乳腺病灶区的不同形式,能更全面地描述病灶区的特征,对乳腺癌诊断具有极大的帮助。为了提取其中的特征,本文提出了协同约束网络。网络架构如图 2 所示,从图 2 可以看到,网络被设计为双分支架构分别处理超声和造影数据。同时,为了更好地融合造影和超声分支,本文提出了双线性协同融合机制。

1.1 超声数据和造影数据

合理地利用数据必须先理解数据的本质,在临床中,医师通常利用 B 型超声数据和超声造影数据来诊断乳腺癌。对于 B 型超声数据,医师主要观察其中的纹理与形态,例如其中的回声密度、边缘是否规则、有无钙化,能从中了解乳腺肿块的部位、大小、囊实性以及有无转移等。而超声造影主要提供的是病灶区动态变化,例如病灶区的增强幅度、增强时相、增强顺序等。这与从图 1 获得的结果吻合,B 型超声图像包含丰富的形状和纹理等外观特征,超声造影视频在时间维度包含大量的病理时序变化。

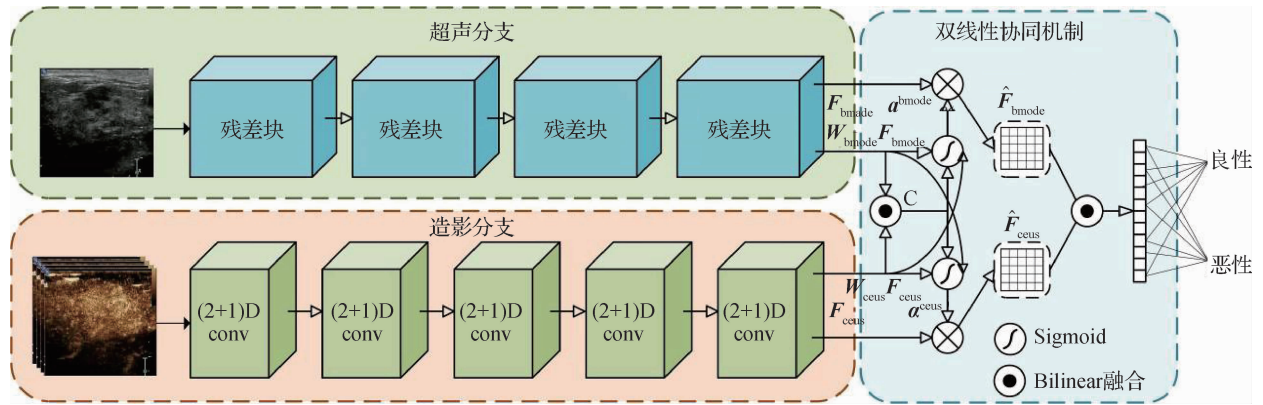


图2 协同约束网络架构

Fig. 2 The cooperative suppression network architecture

因此,B型超声视频的单帧图像足以表示原视频。而造影视频能提供病灶区的动态变化,因此网络需要输入整个造影视频以提取其中的时序信息。本文选取B型超声视频亮度值最大的一帧作为超声分支的输入数据,定义为 S 。而超声造影数据,出于减小计算复杂度和数据冗余的需要,需要从原视频中提取尽量少的帧并使这些帧包含尽量多的时间信息表示原始的造影视频。参考视频识别领域的文献,本文从原视频中提取16帧作为输入数据,标记为 V ,计算为

$$V = \left\{ f^j \mid f_{bri}^j = \frac{[\max(f_{bri}) - \min(f_{bri})]}{16} \times i + f_{bri}^{\min} \right\} \quad (1)$$

式中, f_{bri}^j 表示造影视频中第 j 帧的亮度值,首先计算出所有帧中亮度最大值 $\max(f_{bri})$ 和最小值 $\min(f_{bri})$,然后将亮度差16等分,最后根据亮度差获得16帧,其中 $0 \leq i < 16$ 。最终得到网络的输入数据 (V, S) 。

1.2 超声分支

图像分类研究一直飞速发展,在超声领域也是如此。根据B型超声图像分类领域的发展,本文选择了ResNet-34网络作为超声分支的主干网络。在深度学习中,网络的深度对网络的特征提取能力有极大影响,更深的网络能从图像中获得更高层的特征信息。然而,在实践中,大部分过深的网络并未取得良好的结果,因为网络过深会导致梯度难以传播到深的网络层,进而导致分类精度下降。为了解决该问题,He等人(2016)提出了ResNet,通过在网络层之间引入残差连接,能在不降低梯度的情况下将

网络的梯度传递到更深的层,从而使网络层数尽可能多,进而从输入的图片中提取更多的深层特征信息,提高网络的特征提取能力。

在超声图像处理领域,数据一直是稀缺的。因此为了避免过拟合,网络必须精心设计。在超声分支设计中,每层引入Batch Normalization和Dropout,同时ResNet-34的所有通道被减半以减少整个网络的参数量。该分支的损失函数为

$$L_{bmode} = \frac{1}{N} \sum_i -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (2)$$

式中, p_i 表示第 i 个样本预测为正的的概率。 y_i 是该样本的正确标签, N 表示总样本的数量。最终求得超声分支的损失值 L_{bmode} 。

1.3 造影分支

相比图像,造影视频中包含更多的病理信息,同时也包含更多的噪声信息,在提取病理信息的同时必须抑制噪声。在视频分类任务中,常见的做法是一个视频对应一个类别,通过网络预测这个类别。然而视频由多帧组成,包含大量的信息,一个独立的标签不足以表述一个视频的特征,因此神经网络根据这个单独的标签学习到的特征是很片面的。在临床医学中,医师根据造影视频诊断时,主要是观察病灶区的增强幅度、增强时相、增强顺序等一系列特征来判断当前病灶区的良恶性。因此,与简单的良恶性标签相比,用上述特征组成的多标签更适合表示造影视频,并能使网络聚焦对良恶性权重更高的特征。

因此,结合临床医学的实践,本文采用了新的标签去描述一个造影视频,如表1所示。本文根据医

师的建议采用了10个标签描述视频,每个标签采用onehot编码,即用二元特征表示该类别,相比直接用良恶性做标签,该标签能更准确地描述视频中包含的造影信息并抑制视频中的噪声。如图3所示,基于临床医学设计的标签,能使造影分支提取出的特征具有一定的中层语义,而这些中层语义,恰恰是进行乳腺癌诊断的重要病理特征。

表1 造影视频病理的多标签结构

Table 1 Multi-label structure of pathology of CEUS

病理名称	种类	onehot 编码
增强强度	高	1 0 0
	中	0 1 0
	低	0 0 1
增强时相	快进	1 0 0
	同进	0 1 0
	慢进	0 0 1
增强后病灶范围	扩大	1 0 0 0
	不变	0 1 0 0
	缩小	0 0 1 0
	难以分辨	0 0 0 1
增强顺序	向心性	1 0
	非向心性	0 1
增强均匀性	均匀	1 0
	不均匀	0 1
增强完整性	完整	1 0
	不完整	0 1
蟹足症	无	1 0
	有	0 1
滋养血管	无	1 0
	有	0 1
增强后边界	清楚	1 0 0
	不清楚	0 1 0
	难以分辨	0 0 1
增强后形态	规则	1 0 0
	不规则	0 1 0
	难以分辨	0 0 1

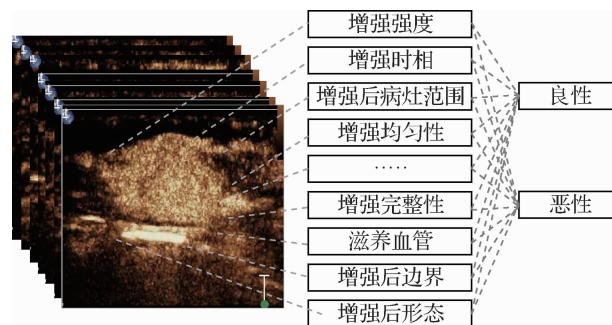


图3 造影信息提取图示

Fig. 3 Information of CEUS extraction diagram

造影分支的主干网络为 $R(2+1)D$, 是一个精心设计的3D网络,通过2D卷积和1D卷积的相互配合,在大幅度减少网络参数和计算量的同时保证了高效的时序特征提取能力。该网络的损失函数为

$$L_{\text{ceus}} = \frac{1}{N} \sum_i - \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (3)$$

式中, p_{ic} 表示 i 样本在类别 c 中的概率, y_{ic} 是该样本的标签值。计算出每个样本的所有类别值,然后求均值得到最终的 L_{ceus} 值, M 是样本的总类别数, N 为样本的总数量(该 loss 仅在超声分支预训练的时候使用)。

1.4 双线性协同融合机制

如前所述,超声数据和造影数据是病灶区不同视角的呈现,分别从空间和时间维度表达了病理特征。在采用超声分支提取超声视频中的外观特征,采用造影分支提取造影视视频的时序特征后,更重要的是如何采用合适的融合方法。

通常的融合方法都是按特征的对应维度一一融合,因此适合每一部分对应关联的两个特征进行融合。但是本文提取出的超声特征的整个维度和造影特征的每一个维度都关联,造影的时间特征是超声空间特征上的连续变化。其次,超声视频上的外观特征是组成造影时空特征的基础,时间特征又是空间特征的动态表达。因此空间特征和时间特征是密切相关、相互影响的。在融合这两种特征时必须考虑到它们之间的形态差异与本质区别。

据此,本文提出了双线性协同融合机制。具体结构如图3所示。经过双分支网络后,分别得到超声的特征 F_{bmode} 和造影的特征 F_{ceus} 。首先对得到的特征进行双线性融合,得到新的特征 C ,双线性特征融合计算方式为矩阵外积,通过外积能将超声空间

特征的整个特征与造影时序特征的每一维进行融合,充分考虑空间对每一维时间的影响,得到更准确的融合特征

$$C = F_{ceus}^T \times F_{bmode} \quad (4)$$

为了进一步提取高层的特征,抑制噪声,本文对提取的特征使用注意力机制进一步处理。该注意力机制为

$$a^{ceus} = S(W_{bmode} F_{bmode} + (W_{ceus} F_{ceus}) C) \quad (5)$$

$$a^{bmode} = S(W_{ceus} F_{ceus} + (W_{bmode} F_{bmode}) C) \quad (6)$$

式中, W_{bmode} , W_{ceus} 分别为卷积核, $S(\cdot)$ 为 sigmoid 跃迁函数。在该机制中,首先对 F_{ceus} 进行卷积操作,然后与前一步融合的特征 C 进行计算,再加上对 F_{bmode} 卷积后的特征。最后采用 sigmoid 函数将特征值的范围限制到 $0 \sim 1$ 。 a^{ceus} 和 a^{bmode} 即最后得到的对应特征的矩阵权重。然后将对应权重应用到原特征,得到加权后的 $\hat{F}_{ceus}$ 和 $\hat{F}_{bmode}$

$$\hat{F}_{ceus} = \sum_{n=1}^N a_n^{ceus} F_{ceus}^T, \hat{F}_{bmode} = \sum_{n=1}^N a_n^{bmode} F_{bmode}^T \quad (7)$$

从式(2)可以清晰地看出,注意力机制建立在双线性融合的特征 C 上。在后续的计算中,使用卷积操作同时纳入了 $\hat{F}_{ceus}$ 和 $\hat{F}_{bmode}$ 特征的信息。最终得到的注意力权重合理地融合了超声特征和造影特征。最后,对经过注意力机制得到的特征进行双线性融合,得到最终的特征

$$C = \hat{F}_{ceus}^T \times \hat{F}_{bmode} \quad (8)$$

2 实验及分析

2.1 实验数据

本文使用的双模态数据一共 1 105 个样本,其中有 587 个为良性数据,518 个为恶性数据,每个样本文件中均包含 B 型超声视频、超声造影视频和对应的病理信息。以上所有数据均来自于四川省某医院超声科,其标签(良性和恶性)均由专业医师标注,保证所有的样本都是可靠的。

在训练阶段,为了契合网络的输入,需要对数据进行预处理操作。由于 B 型超声图像的特殊性,常规的数据增强方法(例如旋转、移位和颜色抖动)会产生形变,因此不适用于此数据集。因此,水平翻转和比例不变缩放方法适用于医学数据扩增。数据扩增后,被缩放到 256×256 像素大小以输入网络进行

训练。

2.2 实验结果分析

为了进一步提高网络的特征提取能力和验证本文方法的有效性,本文针对提出的网络特别设计了 3 个实验。第 1 个实验单独使用超声数据训练超声分支;第 2 个实验使用针对造影数据设计的多标签信息对造影分支进行训练;第 3 个实验使用前两个实验获得的各分支的预训练模型对双分支网络参数进行初始化以获得更好的结果。实验的硬件平台为: Intel(R) Xeon(R) CPU E3-1231 v3 @ 3.40 GHz (NVIDIA GeForce GTX 1080Ti) 32 GB; 软件环境为: Ubuntu 18.04, Pytorch 1.1.0。

2.2.1 超声实验

超声分支主要是从超声数据中提取超声的形状、纹理等外观特征。本文首先对超声分支进行单独训练,以提高超声分支对超声图像的特征提取能力。传统的预训练模型大都是在 ImageNet 数据集上训练的,然而鉴于自然图像与超声图像巨大的差异,ImageNet 并不适合超声领域的预训练数据,因此本文单独收集了超声图像,包括良性 1 633 幅,恶性 1 078 幅,对超声分支进行训练。本文选取了经典的 VGG-BN (visual geometry group batch normalization) 网络与 ResNet 网络作为超声分支的主干网络进行训练,以选取针对超声图像提取能力最强的网络。

该实验使用 Adam 优化器,初始学习率为 0.01,每经过 5 个批次学习率下降 0.001。采用分类精度和 F1 值作为衡量指标。实验结果如表 2 所示。

表 2 超声分支基于不同主干网络测试结果

Table 2 Ultrasonic branches test results based on different backbone networks

模型	/%			
	训练集精度	训练集 F1	测试集精度	测试集 F1
VGG16-BN	74.2	72.2	73.5	72.7
VGG19-BN	75.6	74.8	73.2	73.9
ResNet18	80.5	76.5	79.0	74.7
ResNet34	81.0	80.0	80.3	75.5
ResNet50	92.1	89.5	79.3	70.8

注:加粗字体表示每列的最优结果。

从表 2 中可以看到,ResNet 系列的模型精度总体来说都是高于 VGG 的。基于 ResNet 的残差机

制,使得能设计出更深的网络,从而获得更高的特征提取能力。表2中的ResNet也是网络越深精度越高,但是网络过深的另一个问题是学习参数量过大导致过拟合。ResNet50虽然在训练集上取得了92.1%的精度,但是和测试集的79.3%相差了12%,已经产生了严重的过拟合。因此根据实际情况,本文选择了ResNet34作为超声分支的主干网络,在获得最高的特征提取能力的同时避免了过拟合,另一方面较少的参数量能进一步提高网络的运行速度。

2.2.2 造影实验

造影视频是对乳腺癌病灶区的动态描述,其中包含着大量与B型超声视频不同的病理特征,合理利用这些特征能提高乳腺癌诊断的准确率。鉴于视频难以建模和计算量巨大的问题,从视频中提取特征信息一直是一项困难的任务。为了提高造影分支的有效性,本文通过病理多标签的方式对视频进行更详细的特征描述,每个视频由表1的10个标签组成,每个标签的onehot编码串联成一个向量为当前视频的描述标签进行多标签分类。然后通过网络来训练得到该标签的信息。在实验中,优化器为SGD (stochastic gradient descent),初始学习率为0.1,每5个批次下降0.01。最终实验结果如表3所示。

表3 基于不同主干网络的造影分支测试结果

Table 3 Contrast branches test results based on different backbone network

模型	/%			
	训练集精度	训练集 F1	测试集精度	测试集 F1
P3D	75.2	74.2	74.8	74.3
R3D	74.6	74.5	74.3	72.5
CM3	74.1	73.1	73.6	73.3
R(2+1)D	78.4	77.8	78.1	77.6

注:加粗字体表示每列的最优结果。

从表3中可以看到,R(2+1)D方法获得了最好的结果,分类精度和F1值分别为78.4%和77.8%,相比其他方法最高提升4.3%和4.7%。根据多分类标签训练出来的网络能从造影视频中提取相应的病理信息,而这些病理信息正是诊断乳腺癌良恶性的关键。同时采用了多标签信息后,对造影视频有了更详细的描述,能在一定程度上防止过拟合。

2.2.3 协同约束网络实验

针对前两阶段训练出来的预训练模型,本文迁移到最终的神经网络进行微调。在该实验中,输入数据为超声视频和造影图像,标签为最终的良恶性分类。模型训练的输入尺寸为256×256像素,优化器为Adam,学习率为0.001,每5个批次学习率自动衰减1%。为全面评估模型性能,本文使用精度、召回率、精确率和F1值对模型进行定量分析。在当前超声领域,针对超声数据与造影数据的方法并不多见,因此本文通过与视频理解领域使用广泛的方法进行对比来验证本文提出方法的有效性。实验结果见表4。

表4 本文算法与其他方法对比结果

Table 4 Comparison between the proposed method and other methods

模型	/%			
	精度	召回率	精确率	F1
Two-stream	80.5	79.8	81.4	80.6
MSN	83.4	81.6	84.5	83.0
NGMN	84.8	84.9	84.3	84.6
TRN	83.1	83.2	82.2	82.7
Action-VLAD	80.7	81.9	80.8	81.3
P3D	84.6	84.3	83.8	84.1
本文算法(未预训练)	85.3	85.2	84.2	84.7
本文算法(预训练)	87.5	87.7	86.8	87.2

注:加粗字体表示每列的最优结果。

在4个不同指标中,本文方法综合而言均取得了最好结果。本文方法分类精度达到87.5%,相较其他方法上升了3%左右。

从图4可以发现,在训练集与测试集的实验结果中,本文算法均获得稳定提升。通过其他方法可以发现,基于超声图像的特殊性,大部分方法的召回率和精确率有较大落差,而本文方法在这两个指标上均取得了不错的成绩,最终在F1值上也获得了87.2%的优异成绩。与单独的两分支相比,双分支网络的分类精度是远高于两个分支的,证实了:结合B型超声信息与超声造影信息能提供多角度的病理信息,进而提高分类精度。同时,实验对比了本文方法在未预训练和预训练情况下的结果,可以发现经过预训练后取得了更优异的成绩,证明了超声造影分支通过多标签预训练确实提高了模型的精度。更详细

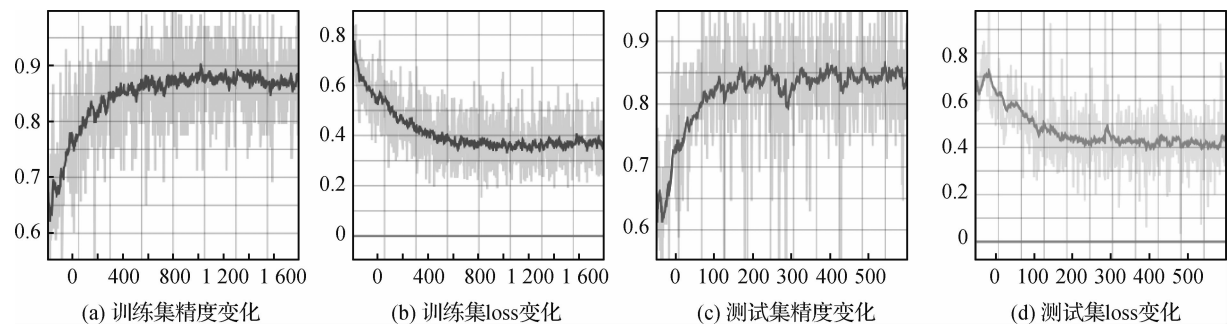


图4 训练期间本文算法精确度和损失值变化情况

Fig.4 Changes of proposed method's accuracy and loss during training

((a) accuracy values of training set; (b) loss value of training set; (c) accuracy values of test set; (d) loss value of test set)

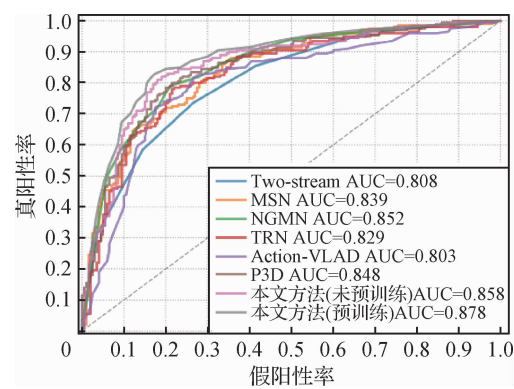


图5 本文算法与其他方法对比

Fig.5 Comparison of proposed method with other methods

的可视化结果由图5给出,图中展示了不同方法精度,可以发现,本文方法也取得了最优结果。

为了进一步验证本文算法的有效性,本文设计了消融实验。结果如表5所示,从表中可以发现,单独使用超声和造影数据取得的分类精度均低于超声和造影融合实验,说明使用双模态数据进行分类相比单模态的数据能获得更多病理信息,为乳腺诊断提供更多的特征信息。同时,本文针对双线性协同机制进行了消融实验,通过对比双线性协同融合机制和常用的拼接融合,可得本文提出的融合机制能将分类精度提高2.7%,能更好地融合超声数据和造影数据中的病理信息,以此提高分类精度。

表6为本文算法与常见的融合算法进行对比实验结果,其中Sum为求和融合,Max为最大值融合,Conv为卷积融合,Biillin为对双线性融合。从表中可以发现,采用Sum和Max方法并未得到很好的效果,这是因为这两种方法直接对超声特征和造影特征进行像素级别融合,而没有考虑两者特征的差异性。而卷积融合具有学习能力,因此有一定的适

应性。双线性融合是将不同特征的每一个维度进行笛卡儿集的融合,而实验中的超声空间特征和造影时间特征也是每一维交互联系的,因此恰好适合本文模型。在此基础之上,为了进一步抑制双模态数据中的造影,本文引入注意力机制,设计了双线性协同融合机制,通过协同注意力,可以让超声分支和造影分支的病理信息相互促进,噪音信息相互抑制,进而使网络获得进一步的提升。

表5 本文算法消融实验结果

Table 5 Results of ablation experiment

模型	/%			
	精度	召回率	精确率	F1
仅超声	80.3	74.3	76.8	75.5
仅造影	78.1	77.1	78.3	77.6
超声+造影 (常用拼接融合)	84.8	84.9	84.3	84.6
超声+造影 (双线性协同融合)	87.5	87.7	86.8	87.2

注:加粗字体表示每列的最优结果。

表6 不同融合方式对分类精度的影响

Table 6 Influence of different fusion methods on classification accuracy

融合方法	/%			
	精度	召回率	精确率	F1
Sum	82.4	81.3	83.4	82.3
Max	81.6	81.4	82.2	81.8
Conv	83.4	83.6	84.3	83.9
Biillin	85.0	84.5	84.8	84.6
本文	87.5	87.7	86.8	87.2

注:加粗字体表示每列的最优结果。

3 结 论

本文的主要研究内容是采取双模态数据进行乳腺癌智能诊断,即联合 B 型超声图像和超声造影视频作为神经网络的输入数据来进行乳腺癌诊断。为此,本文设计了双分支网络架构,能对不同模态的数据进行不同处理,以提取其中的病理特征。同时,为了融合这两种不同特征,本文设计了双线性协同融合机制,根据空间和时间特征的特性采取合适的融合方式。

最终实验结果中本文方法相比其他方法的分类精度提高了 2.7%,取得了最好的成绩。相比其他方法,利用双模态数据能从不同角度展示同一个病灶区,为分类模型提供更多的病理特征,进而提高模型的分类精度。但是,基于双模态设置的双分支网络加之造影分支的 3D 网络导致网络的总参数量过大,实时效率不高。

乳腺癌的病理表现是多方面的,仅凭单独 B 型超声图像进行诊断具有局限性,使用多模态的数据势必是更有效的方向。研究利用多模态超声图像,分析不同模态各自包含的信息,如何充分利用它们的特征,并适当地融合病理信息,是我们下一步的工作。

参考文献 (References)

- Al-Dhabyani W, Gomaa M, Khaled H and Fahmy A. 2019. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *International Journal of Advanced Computer Science and Applications*, 10(5): 618-627 [DOI: 10.14569/IJACSA.2019.0100579]
- Cantisani V, Grazhdani H, Fioravanti C, Rosignuolo M, Calliada F, Messineo D, Bernieri M G, Redler A, Catalano C and Ambrosio F D. 2014. Liver metastases: contrast-enhanced ultrasound compared with computed tomography and magnetic resonance. *World Journal of Gastroenterology*, 20(29): 9998-10007 [DOI: 10.3748/wjg.v20.i29.9998]
- Cao Z T, Duan L X, Yang G W, Yue T and Chen Q. 2019. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Med Imaging*, 19(1): 51 [DOI: 10.1186/s12880-019-0349-x]
- Carreira J and Zisserman A. 2017. Quo Vadis, action recognition? A new model and the kinetics dataset//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 6299-6308 [DOI: 10.1109/CVPR.2017.502]
- Chen W Q, Zheng R S, Baade P D, Zhang S W, Zeng H M, Bray F, Jemal A, Yu X Q and He J. 2016. Cancer statistics in China, 2015. *CA: Cancer Journal for Clinicians*, 66(2): 115-132 [DOI: 10.3322/caac.21338]
- Feichtenhofer C, Pinz A and Zisserman A. 2016. Convolutional two-stream network fusion for video action recognition//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 1933-1941 [DOI: 10.1109/CVPR.2016.213]
- Girdhar R, Ramanan D, Gupta A, Sivic J and Russell B. 2017. Action-VLAD: learning spatio-temporal aggregation for action classification//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 971-980 [DOI: 10.1109/CVPR.2017.337]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hooley R J, Scoutt L M and Philpotts L E. 2013. Breast ultrasonography: state of the art. *Radiology*, 268(3): 642-659 [DOI: 10.1148/radiol.13121606]
- 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]
- Lin T Y, RoyChowdhury A and Maji S. 2015. Bilinear CNN models for fine-grained visual recognition//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: Chile: IEEE: 1449-1557 [DOI: 10.1109/iccv.2015.170]
- Lu J S, Yang J W, Batra D and Parikh D. 2016. Hierarchical question-image co-attention for visual question answering//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red Hook: ACM: 289-297
- Mohammed M A, Al-Khateeb B, Rashid A N, Ibrahim D A, Ghani M K A and Mostafa S A. 2018. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers and Electrical Engineering*, 70: 871-882 [DOI: 10.1016/j.compeleceng.2018.01.033]
- Qi X F, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q and Yi Z. 2019. Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical Image Analysis*, 52: 185-198 [DOI: 10.1016/j.media.2018.12.006]
- Qiu Z F, Yao T and Mei T. 2017. Learning spatio-temporal representation with pseudo-3D residual networks//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: Italy: IEEE: 5533-5541 [DOI: 10.1109/ICCV.2017.590]
- Simonyan K and Zisserman A. 2014. Two-stream convolutional networks

- for action recognition in videos//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge; USA: ACM; 568-576
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston; IEEE; 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Tran D, Wang H, Torresani L, Ray J, LeCun Y and Paluri M. 2018. A closer look at spatiotemporal convolutions for action recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE; 6450-6459 [DOI: 10.1109/CVPR.2018.00675]
- Wang F, Jiang M Q, Qian C, Yang S, Li C, Zhang H G, Wang X G and Tang X O. 2017. Residual attention network for image classification//Proceedings of 2017 Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE; 6450-6458 [DOI: 10.1109/CVPR.2017.683]
- Wang L M, Xiong Y J, Wang Z, Qiao Y, Lin D H, Tang X O and Van Gool L. 2016. Temporal segment networks: towards good practices for deep action recognition//Proceedings of the 14th European Conference on Computer Vision. Amsterdam; Springer; 20-36 [DOI: 10.1007/978-3-319-46484-8_2]
- Wild C P. 2014. International agency for research on cancer//Wexler P, ed. Encyclopedia of Toxicology. Amsterdam; Academic Press; 1067-1069 [DOI: 10.1016/B978-0-12-386454-3.00402-4]
- Wubulhasimu M, Maimaitusun M, Xu X L, Liu X D and Luo B M. 2018. The added value of contrast-enhanced ultrasound to conventional ultrasound in differentiating benign and malignant solid breast lesions: a systematic review and meta-analysis. Clinical Radiology, 73(11): 936-943 [DOI: 10.1016/j.crad.2018.06.004]
- Zhou B L, Andonian A, Oliva A and Torralba A. 2018. Temporal relational reasoning in videos//Proceedings of the 15th European Conference on Computer Vision. Munich; Springer; 803-818 [DOI: 10.1007/978-3-030-01246-5_49]

作者简介



杨子奇, 1995年生, 男, 硕士研究生, 主要研究方向为医学图像处理。

E-mail: xiyaozhe@my.swjtu.edu.cn



龚勋, 通信作者, 男, 副教授, 主要研究方向为计算机视觉及模式识别、医学图像处理等。

E-mail: xgong@home.swjtu.edu.cn

朱丹, 女, 硕士研究生, 主要研究方向为医学图像处理。

E-mail: 1337156613@qq.com

郭颖, 女, 硕士研究生, 住院医师, 主要研究方向为心脏病、甲状腺疾病和乳腺疾病的图像诊断。

E-mail: 826310603@qq.com