

基于深度学习结合解剖学注意力机制的肺结节良恶性分类

刘雲, 王一达, 张成秀, 杨光, 王成龙

华东师范大学物理与电子科学学院上海市磁共振重点实验室, 上海 200062

【摘要】肺结节作为肺癌的初期表现,及时的发现和准确的良恶性诊断对于疾病的治疗具有重要的意义。为了提高肺部CT图像中肺结节良恶性的诊断率,提出一种基于3D ResNet的卷积神经网络,并通过加入解剖学注意力模块有效地提高了肺结节良恶性的分类精度。此外,该方法通过自动分割以获取注意力机制所需的感兴趣区域,实现整个流程的全自动化。解剖学注意力的添加能更好地捕捉图像中的局部纹理信息,进一步提取对于肺结节良恶性诊断有用的特征。本文方法在LIDC-IDRI数据集上进行验证。实验结果表明与传统的3D ResNet及其他现有的方法相比,本文方法在分类精度上有显著的提高,在独立测试集上的最终分类的AUC达到0.973,准确率为0.940。由此可见,本文方法能在辅助医生对肺结节的诊断中起到重要作用。

【关键词】肺结节;注意力机制;CT图像;深度学习

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2022)11-1441-07

Classification of benign and malignant pulmonary nodules by deep learning with anatomy-based attention mechanism

LIU Yun, WANG Yida, ZHANG Chengxiu, YANG Guang, WANG Chenglong

Shanghai Key Laboratory of Magnetic Resonance, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China

Abstract: Pulmonary nodule is the initial sign of lung cancer, and the timely detection and accurate diagnosis of malignant and benign nodules have great significance for the treatment of diseases. In order to improve the diagnostic accuracy of benign and malignant pulmonary nodules in pulmonary CT images, a novel convolution neural network based on 3D ResNet is proposed, with anatomy-based attention mechanism for greatly improving the classification accuracy of pulmonary nodules. In addition, the region of interest required for the attention mechanism is obtained by automatic segmentation, thereby achieving the full automation of the whole process. The addition of anatomy-based attention mechanism can better capture local texture information in CT images and further extract useful features for diagnosing benign and malignant pulmonary nodules. The proposed method is verified in LIDC-IDRI data set. The results show that the proposed method greatly improves classification accuracy as compared with other existed methods and traditional 3D ResNet, achieving an area under the receiver operating characteristic curve (AUC) of 0.973 and an accuracy of 94.0% in the independent test set. The proposed method has the potential to assist doctors in the diagnosis of pulmonary nodules.

Keywords: pulmonary nodule; attention mechanism; CT image; deep learning

前言

作为全球发病率第二、死亡率第一的恶性肿瘤,肺癌已经严重影响人类的健康^[1]。肺癌的早期表现

就是肺结节,及时地对肺结节的良恶性做出诊断可以使患者及早接受治疗。目前临床上,医生主要通过观察肺部CT图像来进行肺结节诊断,CT图像具有分辨率高和无创等优点。美国肺部筛查试验的报告表明,使用CT图像进行诊断使得肺癌死亡率降低20%^[2];但早期肺癌从影像上不易诊断,仍然存在较高的误诊率。

近年来,机器学习及深度学习方法的进步极大地提升了计算机辅助诊断方法的能力。目前,已有不少用于诊断肺结节良恶性分类的传统机器学习方法^[3-5]和深度学习方法^[6-8]。2015年,Madero Orozco

【收稿日期】2022-07-13

【基金项目】中国博士后科学基金(2021M691038);上海市浦江人才计划(2020PJD016)

【作者简介】刘雲,硕士研究生,研究方向:医学图像处理,人工智能,E-mail: yiliumri@gmail.com

【通信作者】王成龙,博士,助理研究员,研究方向:医学图像处理、人工智能,E-mail: clwang@phy.ecnu.edu.cn

等^[9]对肺结节图像进行小波变换,再对小波变换后的图像提取组学特征,并输入支持向量机,实现对肺结节良恶性的分类。2017年,Alilous等^[10]通过对肺部CT图像提取影像组学特征获取结节内部纹理变化,并将组学特征输入支持向量机进行良恶性分类。2018年,Causey等^[11]比较了深度学习以及传统影像组学方法对于肺结节良恶性的预测效果。2019年,Xie等^[12]提出一种协同深度学习模型(MV-KBC),在LIDC-IDRI数据集上的分类AUC达到95.70%。

尽管目前已在肺结节良恶性的分类上有了一定的研究成果,但是由于肺结节相对于肺部CT图像来说,体积较小、位置不固定且诊断的假阳性较高,所以如何选择合适的网络架构来进一步提高分类的准确率,仍面临巨大的挑战。在卷积神经网络中添加注意力机制能有效地减少网络对于无关信息的关注,提高网络的性能^[13]。Sun等^[14]提出基于解剖学的注意力机制,利用脑部组织解剖学的先验知识在脑部MRI图像上的感兴趣区域分割任务上取得了较好的效果。鉴于此,本研究提出一种基于解剖学注意力机制的分类网络,在LIDC-IDRI数据集上取得了较好的分类效果,并通过自动分割肺结节获得注意力机制所需要的感兴趣区域,减轻医生勾画感兴趣区域的负担。

1 材料和方法

1.1 实验数据

LIDC-IDRI数据集是由美国国家癌症研究会创建的肺部CT图像公开数据库^[15]。LIDC-IDRI包含从7个学术中心和8个医学影像公司收集的1018个实例,每一个实例都包含至少一个CT扫描图像和标注文件(XML格式文件)。标注文件包含被最多4名放射科医生标注的感兴趣区域以及9个肺结节征象如钙化、分叶征、毛刺征等。1018个实例包含7371个至少被一名放射科医生标记为“结节”的病灶,其中2669个病灶被至少一名放射科医生标记为“结节直径大于3 mm”。

本研究去除了小于3 mm的结节以及仅被一个医生标记的数据^[16]。放射科医生用1到5表示肺结节的恶性程度,数值越大表明结节恶性的可能性越大。通过对多名医生标记的恶性程度求平均值得到结节的最终恶性程度^[17]。平均恶性程度大于3的标记为恶性即正样本,小于3的标记为良性即负样本,等于3的被认为是人类不易区分的、金标准不明确的数据,所以剔除等于3的标记^[16-18]。在对数据进行刷选后,剩余624个实例,共1226个结节,按照7:3的比例将数据随机拆分成训练集和独立测试集。入组标准详见图1。

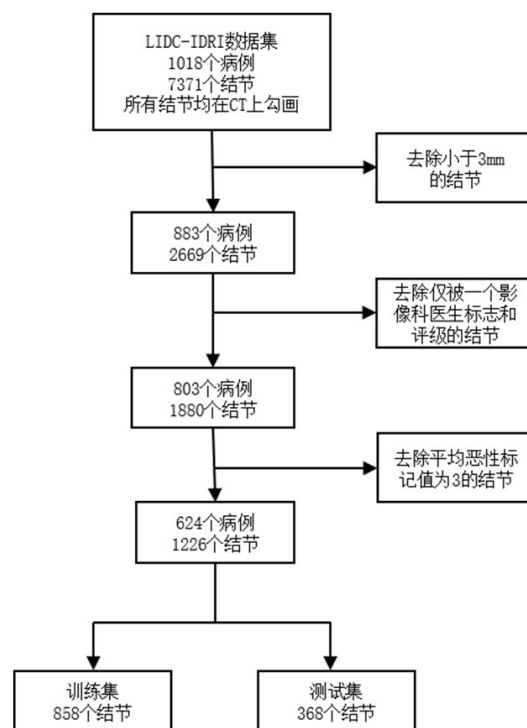


图1 入组标准

Figure 1 Inclusion criteria

1.2 数据预处理

本实验共使用624个实例。为了避免噪声对图像对比度的影响,参照nnU-Net^[19]里的预处理方式将图像像素灰度值小于0.5%和大于99.5%的部分裁切掉。计算公式如式(1)所示。

$$x'_i = \begin{cases} P_{0.5} & (x_i \leq P_{0.5}) \\ x_i & (P_{0.5} < x_i < P_{99.5}) \\ P_{99.5} & (P_{99.5} \leq x_i) \end{cases} \quad (1)$$

其中, $P_{0.5}$ 、 $P_{99.5}$ 代表图像X的0.5%及99.5%灰度值。 x'_i 为像素 x_i 处理后所得像素值。

然后将图像归一化到0到1之间。计算公式如式(2)所示。

$$x''_i = \frac{x'_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

其中, x_{\min} 和 x_{\max} 分别代表图像灰度最大值及最小值。

由于原始图像分辨率不统一,所以通过双线性插值和最近邻分别对原始CT图像和感兴趣区域进行插值,固定分辨率为1.0 mm × 1.0 mm × 1.5 mm。为了保证进入网络的图像大小一致,将每个结节的外接边框向外延伸n个像素(n=35)以确保之后的随机裁切能得到足够大的图像,裁切方法如图2所示,结节区域统计分析见图3。

1.3 硬件及软件平台

实验所用计算机硬件配置为操作系统Ubuntu18.04,显卡型号为GeForce RTX 2080 Ti,选择pytorch1.7作为程序框架,Python版本3.7。

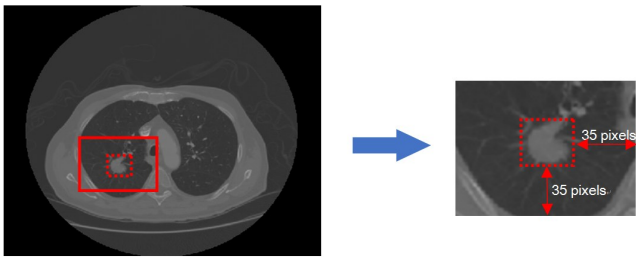


图 2 肺结节局部 ROI 裁切示意图
Figure 2 Local ROI patches of lung nodules

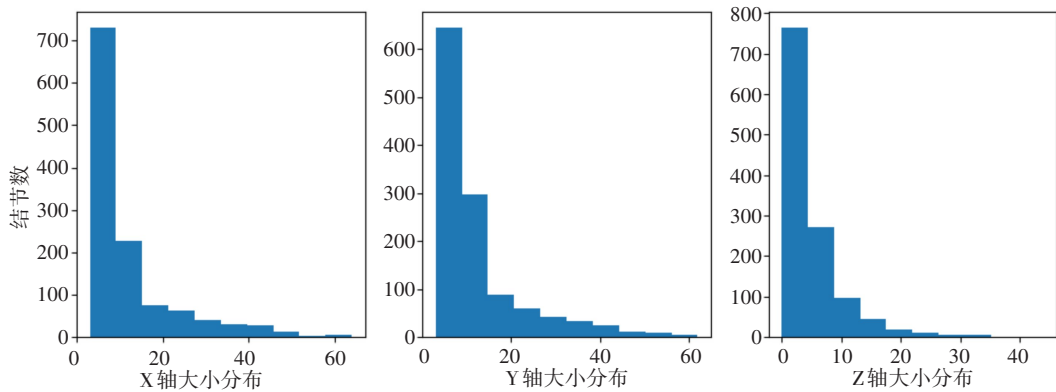


图 3 结节区域的统计分析
Figure 3 Statistical analysis of nodule area

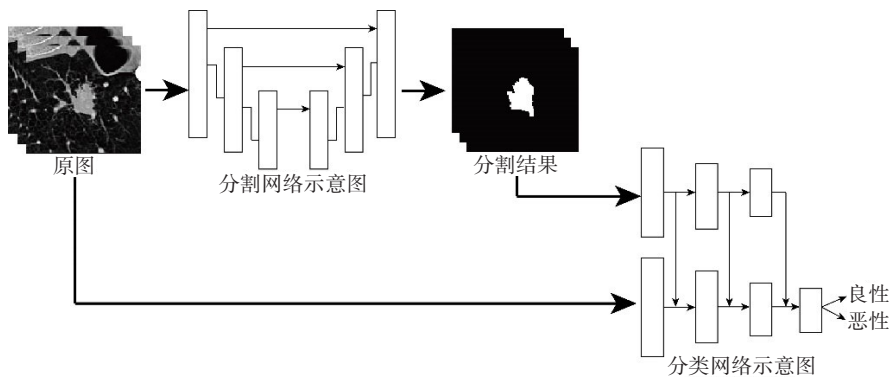


图 4 实验流程图
Figure 4 Experimental workflow

分割实验的训练过程通过随机裁切来增加训练数据量,预测阶段则使用滑动窗口的方式进行预测,滑动窗口的重叠部分为图像训练大小的 50%。实验对于重叠部分的多次预测结果还添加了高斯重要性加权,以避免拼接伪影,减少靠近边界位置的影响,最后通过平均以及 Softmax 的非线性变换获得最后的分割概率。分类实验的训练过程采用随机裁剪,测试阶段裁剪的方式是以前景为中心裁切。除了通过裁切的方式增加数据量,训练过程中对训练数据采取的数据扩增方法还包括随机旋转 90°、更改图像对比度、仿射变换和添加高斯噪声。

1.5 肺结节分割网络结构

本实验使用 Isensee 等^[19]提出的 nnU-Net。对图

1.4 实验流程

实验流程如图 4 所示。具体步骤如下:(1)将裁切好的图像及感兴趣区域随机裁切获得 80×80×56 的大小;(2)将裁剪好的图像输入分割网络以自动获得结节的感兴趣区域;(3)将分割好的感兴趣区域和原始图像随机以结节为中心裁剪到 64×64×32 的大小,之后一起输入分类网络,获得最终良恶性诊断的概率。

像按照结节区域裁切获得一个较小的矩形区域再输入 nnU-Net 进行训练。图 5 为分割网络结构图,网络一共经过 4 次下采样,在每一次卷积层后面加入实例标准化和 LeakyReLU 非线性变换。每一次下采样的特征图都通过跨层连接的形式加到上采样的特征图上,以避免下采样中信息的丢失。

1.6 基于解剖学注意力机制的分类网络结构

1.6.1 解剖学注意力机制 如图 6 所示,解剖学注意力机制的输入包括两张特征图:从原始图像卷积得到的特征图以及从感兴趣区域卷积得到的特征图。两张特征图相加的结果分别经过两个卷积和 Sigmoid 非线性变换,再分别与原来的特征图逐像素相乘,分别获得对应的加权特征矩阵,两个特征矩阵逐像素

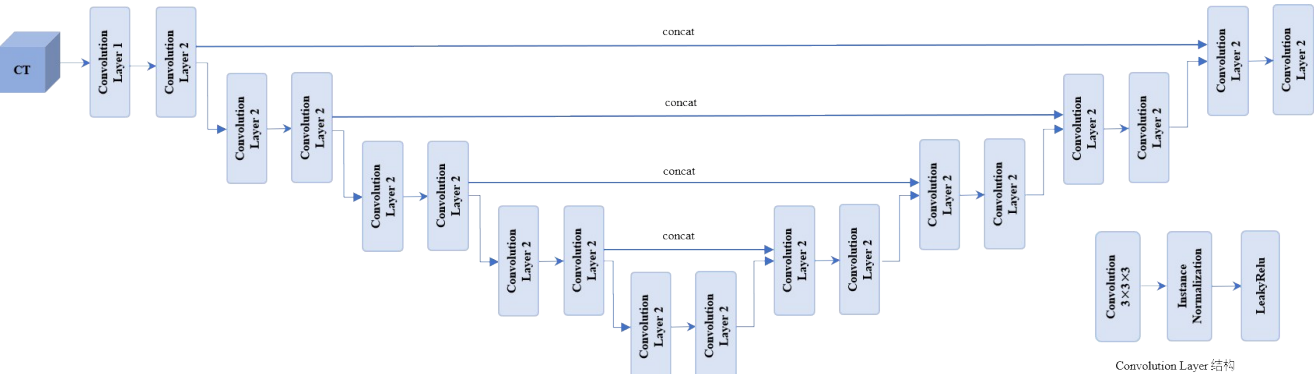


图5 肺结节分割网络结构

Figure 5 Pulmonary nodule segmentation network structure

相加获得加权特征图。

通过在卷积神经网络里添加注意力机制可以调整特征图的权重分配,让网络学习最应该学习的地方,忽略输入中的背景干扰以及冗余信息。基于注意力机制,网络也可以更好地捕捉到图像中的局部纹理信息,对肺结节的良恶性做出更准确的分类^[14]。

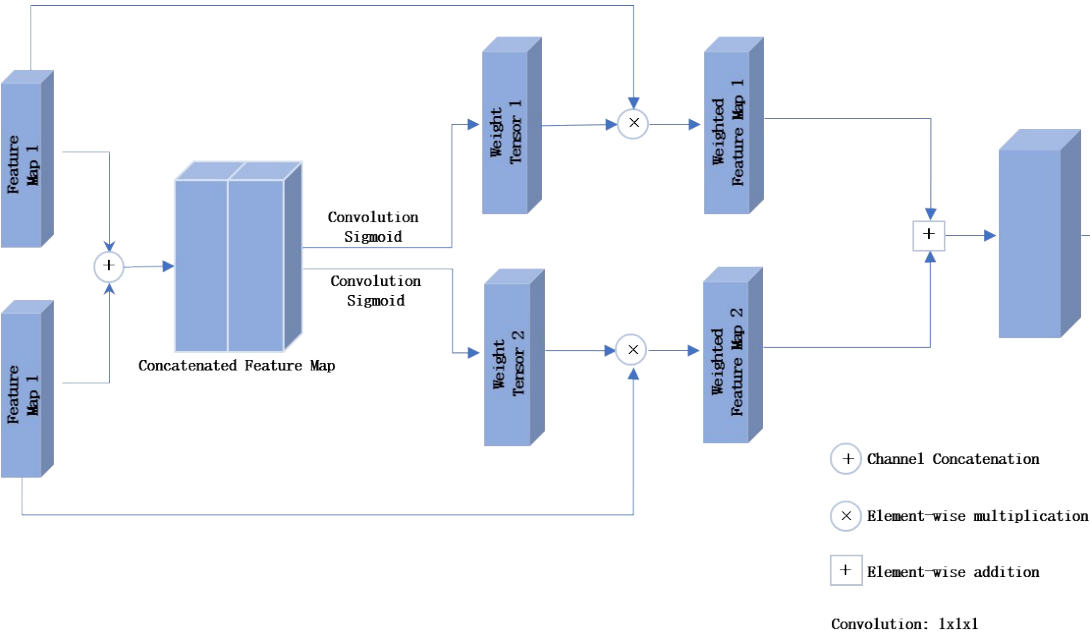


图6 解剖学注意力机制结构

Figure 6 Architecture of anatomy-based attention mechanism

1.6.2 分类网络结构 本实验搭建的分类网络是以3D 残差网络 ResNet^[20]为基础,残差网络在收敛速度和分类效果上都具有自己的优势。如图 7 所示,将 CT 图像和感兴趣区域分别输入网络,对于 CT 图像经过一个卷积块 (Convolution Block) 和 4 个残差块 (Residual Block),最后再经过平均池化和全连接层获得最后的良恶性分类概率。感兴趣区域图像也同样经过 4 层卷积块,经过相同层级卷积块的 CT 特征图与感兴趣区域特征图同时输入解剖学注意力模块。残差连接的添加能很好地避免信息丢失。注意力机制将 CT 和感兴趣区域的特征图分别经过 1×1×1

的卷积获得对应的权重矩阵,再与原来的特征图逐点相乘,获得相应的加权特征图。两个加权特征图逐点相加所获得加权求和特征图便作为下一次残差块的输入。经过最后一个残差块之后的特征图通过平均池化来减少空间信息,最后经过全连接层输出良恶性的分类概率。网络一共经过 5 次下采样,通过融合图像及感兴趣区域的特征图能更好地提取 CT 图像局部的纹理特征。

网络经过的卷积块包括卷积、ReLU 非线性变换、批标准化以及最大池化。其中,最大池化是用来对感兴趣区域进行 2 倍下采样。网络通过添加注意

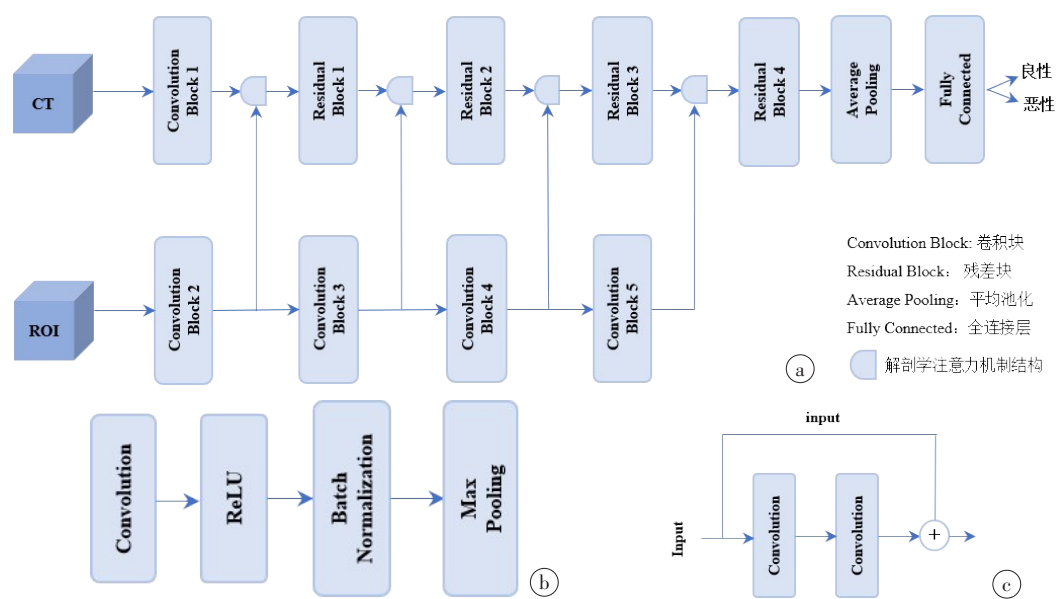


图 7 分类网络结构图
Figure 7 Classification network structures
a: 肺结节分类网络结构图; b: 卷积块结构; c: 残差块结构

力机制模块来自自动学习特征图的最优权重,以实现更好的分类效果。

1.7 训练过程

分割网络训练使用的损失函数为 Dice 和交叉熵损失函数:

$$L_{total} = L_{Dice} + wL_{CE} \quad (3)$$

其中, L_{total} 为总的损失函数; L_{Dice} 为 Dice 损失函数; L_{CE} 为交叉熵损失函数; w 为权重。

本研究设置 w 为 1, 优化器为 SGD, 初始学习率为 0.01, 动量为 0.99, 动量衰减为 3×10^{-5} , 批量为 14, 最大 epochs 数为 200, 分割网络采用 5 折交叉验证进行训练。

分类网络训练使用的能量函数为二值交叉熵函数, 优化器为 Adam, 初始学习率 0.000 1, 批量为 10, 最大 epochs 数为 300。学习率策略使用的是自适应调整学习率, 当验证集的损失维持 10 代不下降, 就将学习率降为原来的 1/10。训练中使用了提前停止, 即当验证集的损失在 50 个 epochs 中变化小于 0.000 1 时, 提前终止网络的训练。

为了避免过拟合, 对训练数据进行在线扩增, 扩增的方式包括随机旋转 90°、随机调节对比度、随机仿射变换和随机添加高斯噪声。

1.8 评价指标

分割网络的输出为像素级别的分类概率, 将输出的分割概率经过 Softmax 以获得最后的分割结果, 本研究利用 Dice 系数作为分割的评价指标, Dice 值的范围为 0~1。当 Dice 值越接近于 1 时, 分割结果越好; 越接近于 0 时, 分割结果越差。

$$Dice = \frac{2TP}{FP + 2TP + FN} \quad (4)$$

其中, TP 表示正样本预测为正样本即真阳性, FP 为负样本预测为正样本即假阳性, TN 为负样本预测为负样本即真阴性, FN 为正样本预测为负样本即假阴性。

分类网络的输出为良恶性的概率, 本研究利用操作者特性 (Receiver Operating Characteristic, ROC) 曲线及其曲线下面积 (Area Under Curve, AUC) 来评价分类网络的性能^[21]。ROC 曲线可以直接地反映二分类问题的分类效果, 当曲线越倾向于左上, AUC 的值越接近于 1, 说明分类器的分类越准确。此外, 本研究使用的评价指标还包括精确率、召回率、准确率、假阳率和 F1 分数, 如表 1 所示。这些评价指标均通过混淆矩阵来比较分类网络的分类结果和影像科医生的标注。

表 1 分类评价指标公式
Table 1 Formulas of classification metrics

评价指标	计算公式
精确率	$\frac{TP}{TP + FP}$
召回率	$\frac{TP}{TP + FN}$
准确率	$\frac{TP + TN}{TP + FN + FP + TN}$
假阳率	$\frac{FP}{FP + TN}$
F1 分数	$\frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$

2 结果

首先将图像输入分割网络,再将分割网络的输出作为感兴趣区域和原始图像一起输入基于 ResNet18 的分类网络。为了验证本文方法,在 LIDC-IDRI 数据集上进行训练和测试。将网络的主干网络替换成 ResNet34 和 ResNet50,分别测试分类的性能;通过消融实验验证解剖学注意力机制的有效性;除此之外,还对本文方法与最近发表的两篇基于 LIDC-IDRI 数据集的肺结节分类方法进行比较。

肺结节的分割任务采用 5 折交叉验证进行测试,采用集成测试(Ensemble Testing)法对测试集进行测试。验证集 5 折分割结果显示 Dice 精度分别为:

0.831、0.821、0.826、0.829、0.827;测试集分割结果为 0.819,分割精度 Dice 达到 0.8 以上,基本可以满足定位和注意力机制所需要的感兴趣区域。本研究还计算了多个医生之间手动分割的 Dice,在 368 个测试集数据上,医生的平均分割 Dice 只能达到 0.73。自动分割的精度高于医生的手动分割,且速度也更快,可以很好地为下游的分类任务提供输入。肺结节良恶性的分类结果如表 2 所示。可以看出本研究所提出的添加了解剖学注意力机制的基于 ResNet18 的分类网络结果相较于最近提出的一些分类方法以及传统的 ResNet 网络,AUC 都有显著的提高,精确率、准确率和 F1 分数都是最高的,召回率比较高,由此可见,本文方法对肺结节良恶性的诊断有一定的作用。

表 2 基于 LIDC-IDRI 数据集的不同网络分类效果的比较
Table 2 Comparison of classification performances of different networks in LIDC-IDRI data set

方法	AUC	精确率	召回率	准确率	假阳率	F1 分数
ResNet18	0.959	0.625	0.928	0.880	0.130	0.747
ResNet34	0.962	0.709	0.871	0.907	0.084	0.782
ResNet50	0.948	0.756	0.842	0.918	0.064	0.797
ResNet50+注意力	0.973	0.699	0.929	0.910	0.094	0.798
ResNet34+注意力	0.972	0.788	0.900	0.935	0.057	0.840
HSCNN ^[16]	0.856	-	0.705	0.842	0.111	-
HESAM ^[18]	0.957	0.720	0.843	0.910	0.077	0.776
本文方法	0.973	0.833	0.857	0.940	0.040	0.845

3 讨论

本研究首先将肺部 CT 图像输入分割网络,并将分割网络的输出作为感兴趣区域和 CT 图像一起输入到加了解剖学注意力机制的 3D 残差网络 ResNet 模型中,对肺结节良恶性进行分类,取得了相比于现有方法更高的分类精度。分割网络实现了感兴趣区域的自动获取,而注意力机制的添加很好地帮助网络学习到良恶性分类需要关注的地方,及时更新特征图的权重也能更好地学习到图像中的纹理变化,对提升肺结节分类精度起到了重要的作用。本研究通过自动分割所获得的结节的区域相比于医生之间的手动分割精度有一个显著的提高,将分割结果作为后续分类的注意力机制更具有鲁棒性更客观。相比现有的一些分类方法,本文方法的精度也是非常高且实现了整个流程的全自动化,极大地解决了不同医生之间勾画不一致的问题。

HSCNN 网络^[16]在训练过程中利用多头输出实现良恶性和征象的同时输出,通过征象使网络关注区分良恶性时更应关注的位置,而本研究直接通过

加入结节的感兴趣区域,让网络关注到结节区域,获得了更高的分类精度且不需要医生对征象的标注以及减少了由于不同医生之间标注不一致所带来的误差。但是肺结节分布于肺部的各个区域,目前还未实现结节的检测,后续工作也可以加入结节的检测以实现真正的全自动化。

4 结论

本研究将图像输入分割网络以获得分类网络注意力机制所需要的感兴趣区域,再将分割网络的输出与原始图像输入分类网络,实现肺结节良恶性分类的自动流程。同时,通过在分类网络中引入解剖学注意力机制,在 LIDC-IDRI 数据集上获得了优于传统 ResNet 以及近期工作的结果。

本研究主要创新点有:(1)提出一个从肺结节分割到分类的全自动化流程。将自动分割得到的肺结节区域作为网络的注意力机制,将卷积神经网络的识别重点集中于局部肺结节上。在预测过程中,全过程无需人工介入,从而实现端到端的全自动诊断

流程。(2)本文方法肺结节自动分割的精度高于医生的手动分割结果,且分割速度更快,分割结果一致性更好,可以更好地为下游的分类任务提供注意力信息。(3)首次提出将解剖学注意力机制加入分类网络中,使网络更好地关注到肺结节的局部纹理信息,从而提高网络的分类精度。注意力机制即插即用,具有很好的灵活性,可方便地扩展于其他的主干网络。分类网络的整体结构较为简单,带来了更好的分类鲁棒性。

未来的工作也将着眼于实现对结节的自动检测以实现整个流程真正的全自动化,同时也会考虑到网络的可解释性,让医生能够相信网络的输出。

【参考文献】

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA J Clin, 2021, 71(3): 209-249.
- [2] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening[J]. New Eng J Med, 2011, 365(5): 395-409.
- [3] Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer [J]. Trans Lung Cancer Res, 2017, 6(1): 86.
- [4] Beig N, Khorrami M, Alilou M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas[J]. Radiology, 2019, 290(3): 783-792.
- [5] Chen CH, Chang CK, Tu CY, et al. Radiomic features analysis in computed tomography images of lung nodule classification[J]. PLoS One, 2018, 13(2): e0192002.
- [6] Ma J, Song Y, Tian X, et al. Survey on deep learning for pulmonary medical imaging[J]. Front Med, 2020, 14(4): 450-469.
- [7] Nasrullah N, Sang J, Alam MS, et al. Automated lung nodule detection and classification using deep learning combined with multiple strategies[J]. Sensors (Basel, Switzerland), 2019, 19(17): 3722.
- [8] Tong C, Liang B, Su Q, et al. Pulmonary nodule classification based on heterogeneous features learning[J]. IEEE J Sel Area Comm, 2021, 39(2): 574-581.
- [9] Madero Orozco H, Vergara Villegas OO, Cruz Sanchez VG, et al. Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine[J]. Biomed Eng Online, 2015, 14: 9.
- [10] Alilou M, Orooji M, Madabhushi A. Intra-perinodular textural transition (Ipris): a 3D descriptor for nodule diagnosis on lung CT [C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Lecture Notes in Computer Science. Cham: Springer, 2017: 647-655.
- [11] Causey JL, Zhang J, Ma S, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans[J]. Sci Rep, 2018, 8(1): 9286.
- [12] Xie Y, Xia Y, Zhang J, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT [J]. IEEE Trans Med Imaging, 2019, 38(4): 991-1004.
- [13] Mnih V, Heess N, Graves A. Recurrent models of visual attention[M]// Gerd N. Advances in neural information processing systems. San Francisco: Morgan Kaufmann Publishers, 2014: 2204-2212.
- [14] Sun L, Shao W, Zhang D, et al. Anatomical attention guided deep networks for ROI segmentation of brain MR images[J]. IEEE Trans Med Imaging, 2020, 39(6): 2000-2012.
- [15] Armato SG, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans[J]. Med Phys, 2011, 38(2): 915-931.
- [16] Shen S, Han SX, Aberle DR, et al. An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification[J]. Expert Syst Appl, 2019, 128: 84-95.
- [17] Shen W, Zhou M, Yang F, et al. Multi-scale convolutional neural networks for lung nodule classification[J]. Inf Process Med Imaging, 2015, 24: 588-599.
- [18] Lei Y, Tian Y, Shan H, et al. Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping[J]. Med Image Anal, 2020, 60: 101628.
- [19] Isensee F, Jaeger PF, Kohl SA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nat Methods, 2021, 18(2): 203-211.
- [20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [21] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms[J]. Pattern Recogn, 1997, 30(7): 1145-1159.

(编辑:谭斯允)