

文章编号: 1671-7104(2023)02-0163-10

深度学习驱动的CT影像肺结节检测: 挑战、进展和展望

【作者】 谭双平¹, 张彤², 祖江林¹, 邓友锋¹, 吴下里¹, 魏杰¹, 严馨月²

1 武汉市江夏区第一人民医院, 武汉市, 430200

2 武汉大学 测绘遥感信息工程国家重点实验室, 武汉市, 430079

【摘要】 基于CT影像进行肺结节自动检测可以显著提高肺癌诊治效果。该文结合CT影像和肺结节的特点, 总结了基于各种深度学习模型的CT影像肺结节检测面临的挑战和近期研究进展, 重点分析了标志性成果和它们的优缺点。结合当前肺结节检测应用的现状, 对未来更好地应用和改进基于深度学习的CT影像肺结节检测技术提出了展望。

【关键词】 深度学习; CT影像; 肺结节检测

【中图分类号】 TP183

【文献标志码】 A

doi: 10.3969/j.issn. 1671-7104.2023.02.010

Deep Learning-driven Pulmonary Nodule Detection from CT Images: Challenges, Current Status and Future Directions

【Writers】 TAN Shuangping¹, ZHANG Tong², ZU Jianglin¹, DENG Youfeng¹, WU Xiali¹, WEI Jie¹, YAN Xinyue²

1 The First People's Hospital of Jiangxia District of Wuhan, Wuhan, 430200

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079

【Abstract】 Automatic detection of pulmonary nodule based on CT images can significantly improve the diagnosis and treatment of lung cancer. Based on the characteristics of CT image and pulmonary nodule, this study summarizes the challenges and recent progresses of CT image-based pulmonary nodule detection using various deep learning models. The study focuses on the review of major research developments by investigating their technical details, strengths and shortcomings. In light of the current application status of pulmonary nodule detection, a research agenda that aims to better apply and improve deep learning-driven pulmonary nodule detection technologies was given in this study.

【Key words】 deep learning, CT image, pulmonary nodule detection

0 引言

肺癌是当前我国以及全球范围内发病率较高的恶性肿瘤之一, 也是全球癌症死亡率最高的病种^[1]。若能在早期诊断出肺癌, 并有针对性地开展治疗, 将有效地提高治疗效果。肺结节是肺癌的早期表现形式之一, 人工判读胸部CT影像(thoracic computed tomography)检测肺结节(pulmonary nodule)是诊断早期肺癌的主要方式。

研究表明, 通过CT图像开展肺结节检测并治疗, 肺癌死亡率可以降低20%^[2]。但是人工诊断受到医师专业水平限制和图像差异的影响, 往往会出现漏检和错检的情况。同时依赖人工方式从大量

CT图像中检测复杂肺结节, 过程烦琐, 工作效率低下。随着图像解译技术的飞速进步, 计算机辅助检测(computer-aided detection)作为计算机辅助诊断(computer-aided diagnosis)的前提, 已经得到了广泛关注和应用。近年来, 机器学习特别是深度学习支持的图像解译方法已经越来越普遍地应用于CT影像的检测和诊断。许多研究者也针对CT影像和肺结节检测的特性和需求, 对原有目标检测模型做出了一些改进, 提出了一批新型自动检测算法, 显著改善了CT影像肺结节检测的精度和效率。基于CT影像的肺结节检测方法和技术取得了显著进步, 但是目前还缺少深入结合应用实践和问题本质的总结。

本研究主要关注CT影像的肺结节检测, 结合该应用问题的难题和挑战, 对近年来基于深度学习的CT影像肺结节检测方法进行归纳和讨论, 并深入分析当前研究进展和应用现状的脱节情况, 对未

收稿日期: 2022-07-31

基金项目: 武汉市卫生健康科研基金重点项目(WX20A11)

作者简介: 谭双平, E-mail: tanshuangpingzy@163.com

通信作者: 张彤, E-mail: zhangt@whu.edu.cn

来的研究方向进行梳理和总结。

1 CT图像的肺结节检测

1.1 CT图像肺结节的特点

早期肺癌初筛一般采用低剂量胸部计算机断层成像技术(low-dose computed tomography, LDCT),通过采用低辐射量的方式进行全肺断面扫描来发现肺结节。这种低辐射量扫描方式可以减少辐射危害,但是也会降低成像清晰度。目前有多种扫描方式,包括薄层扫描、高分辨率CT扫描、螺旋扫描、动态增强扫描、双能CT扫描等。对肺结节的进一步诊断往往会采用较小层厚的扫描方式,对CT图像进行冠状位、矢状位重建来辅助医生判断和确定结节特征。多断面扫描图像可以构成三维图像,从而更清晰地展示肺结节的形态以及与周围气管/血管的相对关系,有利于更精准地识别和定位肺结节以及确定肺结节类型,为后续治疗提供尽可能精准的先验信息。计算机辅助检测系统的算法设计也参考了医学上各个方位重构图像与三维重建辅助判断的方法。CT图像一般存储为DICOM(digital imaging and communications in medicine)格式,这是一种广泛采用的医学图像国际标准,包含了每位患者的个人信息、检查的元数据、分层和图像数据等。

肺结节通常认为是肺内密度较肺实质增高的潜在病灶,一般直径小于3 cm^[3]。肺结节的大小差异

很大,结节越大其恶性率越高,通常认为直径小于4~5 mm的肺结节是低危结节,大于8 mm则认为是高危结节^[4]。结节是否恶性要观察其生长状况。肺结节在内部组成上可划分为软组织、液体、脂肪和空气。在CT图像中,绝大部分肺结节的形态都呈不透明的圆球形肿块,仅有少数具有不规则形状。由于结节生长的方向异质性,部分肺结节带有棘状突起,甚至像多个结节融合而成,表现为分叶征。肺结节通常呈粗糙的外部轮廓,表现为毛刺征。结节-肺交界面虽然在CT成像上较为清晰,但是大部分较为毛糙。结节内部可分为空泡征、支气管充气征、空洞、液化坏死、钙化等多种类型。部分磨玻璃形态结节出现越来越频繁,其检测和诊断具有较大的不确定性^[5]。因为肺结节内外部形态特征的多样性,即使是有经验的放射医生往往也需要长时间判读,且不同放射医生的判读结果会有不同程度的差别。

肺结节在CT图像中表现形式多样,以LIDC-IDRI数据集为例,肺结节在钙化程度上分为6种情况^[6]。形状分类通过结节与线性、卵形、球形相似程度分为5种类型。边缘从明显到不明显分为5种程度。分叶征按照分叶浅深划分了5种分叶程度。有些肺结节边缘存在放射状线影即棘状突起呈毛刺征,根据突起的程度划分了5种毛刺征程度。纹理特征从磨玻璃征到实体纹理分为5种程度。图1展示了各种特征的肺结节CT图像示例。

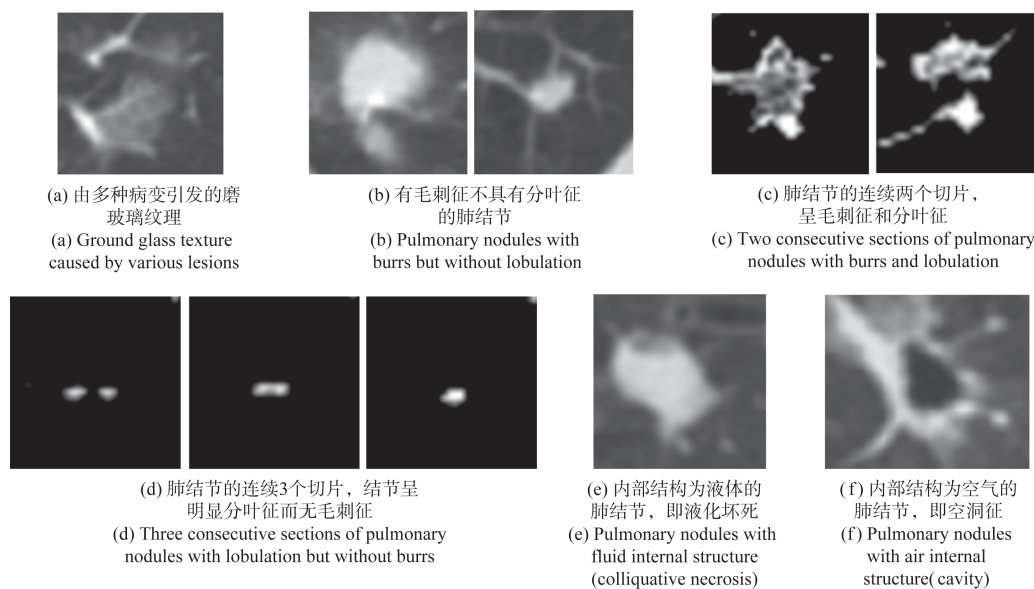


图1 肺结节CT图像示例

Fig 1 Example of CT images of pulmonary nodules

1.2 肺结节检测

与计算机视觉中的通用目标检测类似, CT影像的肺结节检测任务是在影像中识别出所有结节个体并尽可能精确地绘制出结节个体的范围。通常这个范围可用规则的矩形框给出, 也有一些研究者用圆球形更精确地表达肺结节的范围^[7]。通用目标检测可能涉及不止一种目标类型, 而肺结节检测只需要检测一种类别, 即是否为肺结节。在有充足的已标注训练样本条件下, 目标检测可用监督学习的方法训练得到检测器。

肺结节检测的完整技术流程通常包括预处理、候选肺结节检测和假阳性去除等3个步骤。其中预处理阶段主要是对原始CT图像进行降噪和增强等处理以及提取肺实质。肺实质的提取是将肺实质与其他肺部组织(如血管、气管等)分离, 从而去除干扰后期肺结节检测的信息。本研究主要关注后2个步骤, 即提取候选结节及改善假阳性过高的问题。提取候选结节需要保证尽量不遗漏可能的结节。传统候选结节检测依赖于底层特征和经典图像处理方法, 如利用局部形状和曲率特征、场景特征、基于距离阈值分割、三维几何特性和层次特征表征等^[8-12]。但是由于结节在尺寸、形状和类型上的异质性, 基于传统方法难以可靠地提取肺结节的统一特征表达。因此随着深度学习技术在计算机视觉领域的快速发展, CT图像的肺结节检测也开始大量应用基于深度学习的目标检测技术。候选结节检测过程中会尽可能地提取可能的结节, 不可避免地提取到大量假阳性结节, 这给临床诊断带来了不必要的困扰。因此如何高效而准确地去除假阳性结节也是肺结节检测中的关键步骤。传统方法主要是通过训练支持向量机或K近邻等分类器来识别假阳性结节^[8,11]。但肺结节的多样性使得这些传统方法难以准确地发现和去除假阳性结节, 因此近年来深度学习已成为假阳性去除任务的主流方法。肺结节检测通常采用矩形检测框的方式表达检测结果。除了目标检测常用的准确率(precision)、正确率(accuracy)和F1-score外, 肺结节检测的评价定义了特殊的评价指标。如FROC(free-response receiver operating characteristic)指标, 用其分别计算出每张扫描的假阳性率(average number of false positives per scan, avg. FPs)为0.125、0.25、0.5、1.0、2.0、4.0、8.0时对应的敏感度(sensitivity)。敏感度指正确检测出的肺结节占实际所有肺结节的比例,

敏感度高则漏诊率低。每张扫描的假阳性率表示误诊率, 在误诊率一定的情况下比较敏感度才有意义。与此同时定义了竞争性能指标(competition performance metric, CPM), 计算上述7个固定FPs对应的敏感度的平均值。FROC曲线是以敏感度为纵坐标, 以每张扫描的假阳性率为横坐标绘制的曲线图。CPM、FROC曲线和敏感度这3个判定指标是肺结节检测中最通用的指标, 往往同时用于网络性能评估^[17,13-14]。

随着肺结节检测的网络由二维逐渐向三维演化, 有研究者进一步提出了FROC_{IoU}指标^[15]。IoU(intersection over union)是指检测框与标注框的交集与并集的商, 相较于FROC指标, 若检测框与任何一个肺结节中心之间的距离在一定范围内则认为正确判定; 当IoU大于某个预先设定的值时, FROC_{IoU}指标才认为是正确判定。平均准确率(average precision, AP)是肺结节检测中常用的指标, 如AP₂₅表示在IoU ≥ 0.25 时为正确判定情况下的平均准确率; AP_s、AP_m、AP_l则分别对应小、中、大结节的平均准确率。也有学者通过计算各种大小结节的检出率来展示模型性能^[13]。

2 肺结节检测挑战

与自然图像中的目标检测不同, 由于CT图像自身以及肺结节的特点, 自动精确地提取肺结节还存在诸多挑战:

1) 肺结节的类型和种类多, 大小、形状各异。部分类型的结节检测较难, 如磨玻璃类结节常发于血管和肺实质之间, 灰度特异性较低。不同种类肺结节的数目之间存在很大差异, 如肺结节内部结构大多为软组织 and 液体, 内部为空气的结节数目较少, 尤其是肺结节内部为空气时与常见的结节特征存在较大差异, 使得常规的计算机视觉检测模型对该类型结节不敏感。同时结节个体大小也存在差异。因此通用的计算机视觉深度检测算法难以全面准确地提取所有的肺结节目标, 特别是假阳性结节的准确诊断是临床诊疗的重要难题^[3]。

2) 样本不平衡与成像差异。相对于计算机视觉的常规目标检测方法, 肺结节检测正样本数量较少。以LUNA16数据集为例, 888张CT图像中只包含1 186个正样本。同时由于CT图像设备由多家厂商供应, 肺结节在不同设备上的图像成像也可能存在一定差异。这一问题对训练兼容性强的候选结节检测器提出了挑战^[14,16]。

3) 肺结节作为小目标和稀疏目标给检测带来的困难。小目标的视觉特征信息显著少于常规尺寸目标,因而受到场景环境干扰大,其检测精度往往不能满足实际需求。由于肺结节通常较小,其带有的视觉特征不够显著,这给建立可靠的自动智能检测算法带来了困难^[13]。同时肺结节的出现频次不高,呈稀疏分布,这与自然图像一张包含多个密集目标的检测有较大区别,因此直接照搬计算机视觉的检测方法对于作为稀疏小目标的肺结节检测可能性能不佳。

4) 肺部CT图像数据量占用存储空间大。常用的LUNA16数据集存储空间超过100 G。由于内存等硬件的限制,深度神经网络并不能一次处理一组肺部CT图像,而是要将图片进行分割。现已有随机切割、以肺结节为中心切割以及沿着深度方向切割等方式。不合理的切割方式可能会造成信息丢失和割裂。如何更好地分割CT图像组以获取更好的结果仍有待研究。

5) 充分利用医学领域知识和临床经验。当前已有的肺结节深度学习检测方法基本上是将视觉领域的模型应用在CT图像上或做少量的改进。大部分成果停留在各种计算机视觉检测方法(如流程框架、基础学习网络、信息融合方法、样本不平衡应对策略和注意力机制等)的经验组合上,较少在建模时集成医学诊断知识和临床经验,少量研究停留在特征级的知识集成^[17]。在深度学习模型中集成高层次的领域和临床知识,不仅可以提高模型的可信度,也有利于改善模型的泛化能力。

6) 样本数据集的精确标注。无论是流行的基准数据集还是研究者采集的样本数据,都存在正样本的不确定性问题,即肺结节需要经验丰富的放射医生标注,标注难度大、成本高,不同放射医生的标注往往存在差异。即使是LUNA16数据集中得到4个专家共同标注的结节也只有777个^[18]。甚至有些结节的确定需要不同年份的CT图像判断其生长特性。仍有大量的样本无法得到确定标注,给训练可靠的检测器带来了困难。

7) 检测模型的可解释性。临床医学与普通的计算机视觉应用不同,诊断结果的容错性较低,不能完全依赖黑盒子的深度学习模型。现有的深度学习可解释模型较多关注较低层特征的重要性分析,缺乏关联显式医学知识的解释能力。如何增加深度学习检测模型的可解释性,让医生和患者能充分理解检测的原理,才能更好地评价模型的性能,做出

准确可靠的诊断。

3 肺结节的深度学习检测进展

目标检测是从图像或视频中对不同类型的物体进行识别和定位的任务,也是计算机视觉长期关注的热点研究方向之一。随着深度学习特别是卷积神经网络在图像解译中的突破性发展,目标检测开始广泛采用深度学习方法^[19]。

3.1 方法

受到计算机视觉领域目标检测研究的指引,当前基于CT图像的肺结节检测研究大量借鉴了视觉领域的研究成果,同时适当考虑CT图像中肺结节的成像特点,取得了一定的研究进展。下面首先关注整体框架和基础学习网络这两个肺结节检测问题,这也是所有检测任务面临的核心问题。然后对肺结节特征的提取和表征、样本不平衡等方面进行总结,并分析近年来深度学习驱动的肺结节自动检测的研究进展,这两方面则体现了研究人员各自的创新。每一部分先讨论该研究方向如何借鉴计算机视觉领域的相应方法,然后介绍肺结节检测研究中的具体案例,特别是改进的创新点。

1) 检测框架:当前基于深度学习的目标检测方法分为有锚框(anchor-based)和无锚框(anchor-free)两种方法。前者通过卷积神经网络提取的最终图像特征点为中心生成不同形状和尺寸的先验边框,然后进一步对这些边框进行回归精化,得到最终的检测框作为检测结果。有锚框方法的缺点在于锚框组合的尺寸和长宽比例需要手工设定,往往需要通过多次实验确定针对特定问题的合理锚框,确定锚框尺寸的计算量大。一旦目标更改还需重新设计锚框,如果目标尺寸变化较大,锚框设计将更加复杂。无锚框方法摒弃了先验边框,采用金字塔特征网络融合多尺度视觉特征获取目标的中心^[20]。然后依托该中心预测中心到4个边框的距离从而得到检测目标的范围^[21-22]。相对有锚框方法,无锚框方法由于不受锚框的约束,其训练难度更大,往往会带来较多的假阳性检测结果。选择适合的检测框架是进行自动目标检测的第一步也是最关键的一步。大部分深度学习驱动的肺结节检测方法采用两阶段策略:第一阶段通过常规的检测器发现尽可能多的候选肺结节区域;由于得到的候选区域含有大量假阳性结果,第二阶段需要过滤这些假阳性结果,并精化候选区域^[16]。两阶段策略的缺点是计算复杂度高,最终性能依赖第一阶段检测器的候选结果,对

于小型肺结节的检测性能不佳。

研究人员已经开始尝试建立整体端到端的肺结节单阶段检测框架^[14,23]。单阶段检测还可以扩展到更多的任务,如TANG等^[24]将结节检测、假阳性去除和结节分割等集成在一个端到端的三维卷积神经网络多任务学习框架中;HARSONO等^[25-26]选用inception 3D作为主干网络,建立了一种集成检测和分类的单阶段迁移学习方法,来应对样本稀缺的问题。随着无锚框方法的流行,通过一步检测方法可以直接提取肺结节的位置、半径和偏移量等空间信息来实现精确定位^[7]。

一些研究者针对CT图像和肺结节检测问题的特殊性,提出了特定的检测流程。KHOSRAVAN和BAGCI^[13]提出了在单一尺度上检测肺结节的单步检测流程方法,通过设计合理的卷积块和密集连接减少训练开销,利用合理的采样技术提高小结节的检测精度。PEZESHK等^[14]提出了多个三维卷积层和全连接层组成的两阶段检测流程,集成了非极大值抑制、候选图像块提取和假阳性去除等多个步骤,较好地解决了正样本缺少、参数训练困难和计算开销大的问题。LUO等^[7]提出了基于球形的肺结节表达方法和中心点匹配的无锚框检测方法,与传统基于外接矩阵框的表达方法和基于锚框的检测方法相比,改善了检测精度和效率。

2) 三维深度学习网络:在选定合适的检测框架后,就需要选取合理的深度学习网络作为肺结节检测的基础结构。由于肺结节检测的特殊性和挑战,直接应用计算机视觉的目标检测网络往往性能较差。许多研究者在已有的目标检测网络模型基础上进行了改进,如基于Faster R-CNN等^[27]。但是常规目标检测方法面向二维图像,直接将它们用于三维CT图像则无法充分利用三维空间信息。直接建立三维深度卷积神经网络提取三维CT图像序列中的肺结节也是普遍采用的方法,包括基于滑动窗口的三维卷积神经网络、支持不同输入图像尺寸的三维卷积神经网络、3D Faster R-CNN、三维区域候选网络(regional proposal network, RPN)和扩展到三维的ResNeXt网络等^[15,28-32]。由于性能的限制,整个肺部CT图像如果一次输入建立三维模型,则计算量过大。所以往往将图像采用固定维度切割成较小块或根据肺结节体积特征输入相邻几个CT图像,也有的学者将三维深度卷积神经网络用于假阳性去除步骤^[14,33]。

3) 肺结节特征的提取和表征:应对肺结节检

测挑战需要合理的归纳和表征肺结节的视觉特征。虽然深度学习方法提供了高层次抽象特征提取和表征的通用框架方法,但是还需要结合肺结节检测问题的挑战和特点,创新面向CT影像的新型特征集成表征方法。相比自然图像,CT影像带有丰富的多视图多尺度信息,因此建立需要多个网络或网络分支分别提取不同视图和不同尺度上的图像特征表达,然后融合学习用于识别肺结节的高层视觉或语义特征。

集成候选结节在不同视图上的二维图像块特征进行训练是恢复三维图像特征的一种有效策略^[34-35]。面向各种不同厚度以及一个纵轴切片的最大密度投影图像分别建立4个二维卷积神经网络并融合,可以较好地改善小结节检测精度和减少假阳性检测^[36]。XIE等^[16]为了整合结节的三维信息和学习结节的不同特征,分别对3种横断位切片训练了3种基于RPN的二维神经网络模型并融合获取候选检测。TAN等^[17]探索了特征级的两阶段融合方法,把各种纹理特征融入卷积神经网络,对预处理的候选图像块进行分类,从而识别肺结节。MEI等^[15]设计了一种非局部的图像切片组合模块来描述组内特征的关联情况,从而模拟医生对照不同切片来分析肺结节组成。ZHENG等^[36]基于人工诊断中利用最大密度投影图像的特点,面向不同厚层厚度的最大密度投影视图建立二维卷积神经网络,基于多视图网络融合来更好地利用CT影像的三维空间信息。

针对肺结节的形态大小差异,建立多尺度特征融合的网络是较为常见的思路。在特征提取和表达方面,金字塔多尺度模型得到了一定的应用,可以更好地支持不同尺寸肺结节的检测^[25,37]。KIM等^[38]建立了一种多尺度渐进特征集成卷积网络用于假阳性去除。孙华聪等^[39]提出了一种采用三维多尺度深度卷积神经网络的肺结节检测方法。GUO等^[40]提出了一种面向空间和通道信息的多尺度聚合网络。

在目标检测模型中引入注意力机制对提高模型性能有积极作用,注意力机制具有参数少、速度快、效果好等优点。在肺结节检测任务中,有研究者通过加入挤压和激活(squeeze-and-excitation)注意力模块,强化重要特征,提升了模型对通道的敏感性,获得了性能更高的模型^[7,23]。LUO等^[7]通过放缩图像与下采样后图像连接,为模型添加了空间注意力。基于多尺度特征的注意力网络能更好地利用CT影像中的细节信息,在候选结节检测和假阳性去除方面能起到显著的改善作用^[41-42]。

4) 改善正负样本不平衡: 在很多计算机视觉问题中, 正样本相对于负样本数量过少会导致检测模型精度下降。类似地, 由于标注费时费力, 带有精确标签的肺结节样本也比较缺乏。部分研究者借鉴了机器学习领域中常见的数据增广方法, 增加正样本数目^[43], 通过难例挖掘、负样本采样增强网络对易误判负样本的识别能力, 以及挖掘不同类型的难例训练多个具有针对性的模型融合检测结果等方法, 缓解正负样本不平衡问题, 以降低检测的假阳性率^[14,16]。对于特征提取和表达, 金字塔多尺度模型也得到了一定的应用, 可以更好地支持不同尺寸肺结节的检测^[25,37]。针对正样本不足以支持复杂深度学习建模训练的问题, HARSONO等^[25]在预训练扩展三维卷积网络(Inflated 3D ConvNet)基础上, 建立了一种单阶段的迁移学习方法, 显著改善了肺结节的检测和分类精度。GUO等^[40]则引入分布排序损失来改善样本不平衡问题。

基于以上总结和分析, 可以看出目前主流的肺结节检测研究的核心方法还是基于计算机视觉的核心目标检测技术。近年来, 越来越多地从两阶段检测转为单阶段检测, 从有锚框检测转向无锚框检测。同时大量使用RPN等流行的检测网络和三维卷积网络提取和过滤候选结节。部分研究还集成了数据增广、负样本采样、难例学习等改善正负样本不平衡问题。部分研究采用了挤压和激活网络等注意力机制, 提高了检测精度和可解释性。已有很多研究开始显式地顾及CT影像和肺结节检测问题的特点, 从改进检测流程、改进网络结构、提高信息融合质量和检测结果表达等方面进行了各种创新。很多研究者将各种视觉技术组合在一个集成的端到端的学习框架中, 如LI和FAN^[23,44]集成了自编码器、3D RPN网络, 融入了带有注意力机制的挤压和激励网络以提升模型对通道特征的相关描述能力。PENG等^[45]集成了Res2Net的多尺度特征提取和挤压激励单元来提升复杂背景下的肺结节检测精度与假阳性去除性能。HUANG等^[46]在一阶段检测框架中集成了特征融合和注意力机制。有些方法是结合肺结节检测任务要求, 对计算机视觉方法进行较为全面的改进, 如郭晓敏和黄新^[47]针对当前算法对小结节和近血管结节不敏感的问题, 在YOLOv3基础上, 分别对主干网络、激活函数、锚框优化和损失函数进行了改进, 提高了对常见结节类型的检测精度。以上这些方法和技术在面对肺结节样本差异大、不平衡等问题时, 可以灵活应用常规目标检测

方法, 改善检测器的特异性和稳定性, 提高了小结节检测的精度。

总体而言, 大多数研究还停留在直接应用或略微改进计算机视觉的各种检测方法。部分创新在利用CT影像特点方面有一定的进展, 但是大部分方法仍局限在特定的设计上, 缺乏通用的整体解决方案。在模仿人工诊断过程、网络建模设计、特征提取与集成、训练方法改进方面, 很少显式地利用医学专业知识。针对模型的不确定性和可解释性的研究还较少。

3.2 性能对比与讨论

为了便于算法研发和性能对比, 研究人员建立了多个肺结节CT影像基准数据集。因此可以基于这些数据集对经典和最新的检测算法进行对比和分析, 讨论各自的优缺点。表1列出了一些常见方法的性能对比结果。可以看出, 大部分研究基于LUNA16数据集进行性能测试和对比, 也有一些基于更早期的LIDC-IDRI数据集^[6,18]。总体而言, 近两年的研究已经取得了较高的检测精度, 特别是LUO等^[7]有针对性地提出了更合理的表达和基于无锚框检测的注意力机制, ZHANG等^[41]采用集成多尺度特征的注意力网络, GUO等^[40]提出了集成空间和通道信息的多尺度聚合网络, 显著提高了检测性能(FROC超过90%)。目前大部分检测方法都已经直接采用3D CNN或者RPN作为基本主干网络架构, 辅以各种集成、训练和损失函数的改进方法, 更好地利用了CT影像的多视图信息。针对三维卷积带来的较大计算压力, LIAO等^[31]提出了基于小块的训练方法, 有效地提高了训练效率。单阶段检测也逐渐成为主流, 还出现了将结节检测、假阳性去除和分割集成在一个端到端的3D CNN中, 进行多任务学习的方法, 更加有利于CT影像信息的紧凑提取和高效诊断^[24]。从表1方法可以看出, 基本框架、创新以及各种优化策略都存在很大的区别, 从检测结果看, 这些方法都发挥了一定的作用。三维建模的肺结节检测网络的功能指标普遍高于二维建模的网络, 多尺度特征的融合与注意力机制对性能的提升也发挥了很大的作用。我们可以认为各种创新方法都存在一定的互补性, 研究和应用人员可从中选取特定的基本框架、表达方法、特征先验、训练方法、损失函数、数据增广技术、网络集成方法、特征融合和注意力机制等, 合理组织和改进, 相信应该能进一步提高肺结节的检测性能。

表1 CT影像肺结节的深度学习方法对比

Tab.1 Comparison of deep learning-driven methods for pulmonary nodule detection based on CT images

基准数据集	文献	基本框架	核心创新	所取得的最优性能 (CPM)
LUNA16	KHOSRAVAN和BAGCI ^[13]	3D CNN	单一尺度上的单阶段集成检测	0.897
	XIE等 ^[16]	Faster R-CNN	融合基于3种横断位切片的检测网络	0.775
	PEZESHK等 ^[14]	3D CNN	基于数据增广训练的3D CNN多模型结果集成	0.832
	KIM等 ^[38]	3D CNN	多尺度渐进特征集成卷积网络	0.942
	LIAO等 ^[31]	3D RPN	分块训练的高效单阶段三维体检测	0.856
	LI和FAN ^[23]	3D RPN	基于自编码框架和挤压和激活注意力机制	0.862
	MEI等 ^[15]	3D PRN (ResNet50)	非局部的图像切片组合模块	0.874
	GUO等 ^[40]	U-形编码和解码器	融合了空间和通道信息的多尺度聚类网络; 分布排序损失应对不平衡	0.920
	ZHANG等 ^[42]	Faster R-CNN+U-Net	基于多尺度特征的注意力网络	0.927
LIDC-IDRI	TANG等 ^[24]	3D CNN	结节检测、假阳性去除和分割集成在3D CNN多任务学习框架	0.873
	LI和FAN ^[23]	3D RPN	基于自编码框架和挤压和激活注意力机制	0.773

4 讨论与总结

4.1 应用现状和分析

虽然目前肺结节检测技术已经取得了显著进步,在很多实际诊断的对比分析中,深度学习方法在某些场景下已经优于医生的人工判读,如亚实性结节的深度学习检测效果明显优于人工判读^[48]。但是自动而精确的检测仍是一个极具挑战性的目标。在多数临床实践中仍未普遍采用肺结节自动检测方法进行肺癌的早期诊断,基本还是依靠放射医生的经验人工判读。一部分原因是医生对基于机器学习的图像解译技术不够了解,更重要的是当前深度学习驱动的CT影像肺结节技术在检测精度、泛化能力、可解释性以及检测设备的集成方面还存在改进空间。首先在检测精度方面,虽然现在很多方法在基准数据集上取得了较好的检测精度,但是在实际CT影像中的表现不一定能满足实际需要,已有实证研究表明基于机器学习的自动检测算法假阳性较高^[49]。因此在实际判读诊断中,仍然需要通过医生的再次审核。专业医生对肺结节进行检测时,往往需要对比患者历次CT影像,进一步判断结节的生长情况从而对结节病变进行更准确的判断。患者

当时的健康状况和病史也会对肺结节的诊断产生影响,当前基于深度学习的检测方法不具有处理该种先验知识的能力,从而可能造成误诊。

此外,深度学习的黑盒子特性使得模型的结果难以通过医学知识进行解释和推理,自动检测工具难以取得医生和患者的信任。目前虽然有研究开始应用深度学习的可解释技术来对肺结节自动诊断结果进行评价和分析^[50],但是在深入和紧密集成领域知识和专家经验,构建更加透明、便于理解的深度学习模型方面还有很大的改进空间。最后,在使用的便利性方面,现有的CT设备自带图像存储和管理工具,自动检测算法并不能在实际运行中在线地集成到CT设备系统,需要另建专门的图像接入、训练和测试环境^[51],因此使用不够便捷,阻碍了肺结节自动检测算法的推广。

4.2 研究展望

根据前述的总结和分析,可以认为,随着深度学习和计算机视觉技术的发展,未来CT影像的肺结节自动智能检测以及辅助诊断将逐步走入临床应用。在实现这个目标之前,还需要开展许多研究工作。本小节归纳了未来的研究热点,首先从深度学习检测算法角度总结,然后结合临床应用讨论重点

研究方向。具体为:

1) 实例级分割。目前肺结节的检测结果基本上仍用规则矩形框表达。然而实际上肺结节的形态是极不规则的,因此实现实例级个体的精确边界分割和提取,是支持精确诊断的重要研究任务。

2) 半监督/自监督分割。由于目前高质量标注的肺结节数据集还比较缺乏,如何利用粗略标注或无标注临床影像数据,如何利用半监督学习或者目前流行的自监督学习技术,如何充分利用临床积累的大规模数据是未来研究的一大热点。

3) 不确定性建模。由于支持建立深度学习检测模型的训练样本标记存在相当的不确定性,影响了基于这些样本数据集训练的检测模型的泛化能力。当前绝大多数模型都没有考虑样本的不确定性,需要在样本选取、迭代训练、结果置信度表达等方面给予充分考虑。

4) 智能构建学习网络。结合前述的各种检测模型,可以发现网络结构的设计依赖于研发人员的经验,往往不是最优的模型架构。可以借鉴深度学习中智能网络架构的构建方法,选取合理的主干模型、网络拓扑结构和损失函数形式,科学采用和改进注意力机制,进一步提高检测模型的归纳泛化能力和预测精度。

肺结节精细检测的发展不仅需要影像学和计算机视觉学者的努力,更需要深入结合临床应用需求,通过临床辅助诊断应用的反馈逐步改进检测算法。因此面向临床应用的研究至关重要,包括:

1) 集成医学先验知识的检测方法。虽然已经有部分研究开始利用CT影像特点以及肺结节的医学知识,但是距离充分利用放射医生的专业领域知识还有相当的距离。随着先验知识驱动的机器学习研究的发展,如何在深度学习网络构建、样本选择、训练方法和结果验证方面融入领域先验知识是一个值得探讨的研究方向。特别是建立一个集成主动学习方法的自进化智能检测算法,将是较好地融合医学领域知识和医生经验的研究思路。

2) 建立基准数据集。相对计算机视觉领域的基准数据集,已有的肺结节检测基准数据集在图像规模、样本数量、代表性、样本的可靠性、细分类别的平衡等方面还有很大的差距。因此开展国际合作,建立更大、更全面、更精确的基准检测数据集

是一个支持未来研究发展的重要任务。这方面的研究从病例的选择,结节的标注,到评价指标的创新等都需要临床医生的深度参与。目前我国尚未建立基于中国病例特点的大规模基准数据集,这影响了研究成果的临床应用价值,因此建立有代表性的肺结节CT影像基准数据集将是一项重要的研究工作。

3) 模型的可解释性。深度学习模型的可解释性和可视化近年来取得了显著发展,有力支持了模型的推广和实际应用。但是在肺结节检测领域,可解释性的研究还很匮乏,阻碍了模型的实际推广。依托深度学习模型,结合可视化技术,对模型提取特征的有效性、先验知识的表达、检测结果的置信度等设计专门的解释方法,是吸引医生和患者采信肺结节自动检测结果的重要研究方向。

4) 临床诊断应用。肺结节自动检测算法必须服务于计算机辅助诊断的临床应用,因此将检测算法集成到现有的医学影像和病例管理系统中,充分考虑临床医生的使用习惯和诊断方式,设计便于使用的交互诊断界面和辅助诊断系统,将是肺结节深度学习检测方法走入实际应用的重要步骤。

5) 跨学科合作与交流。肺结节检测研究涉及计算机、人工智能、临床医学、生物学工程等多个学科。现在的研究范式基本还局限在计算机视觉的深度学习方法框架内,距离临床使用还有相当大的距离。临床医生对自动检测算法的信任度较低,研究人员与临床医生的交流也不够充分,因此建立持续的跨学科、跨机构交流和协作是实现肺结节智能检测进入临床诊断的必由之路。

4.3 总结

本研究从肺结节检测所采用的CT影像的特点出发,概括介绍了肺结节的特征,总结了计算机视觉中相关的目标检测方法,分析肺结节检测任务的特殊挑战。总结了近年来CT影像深度学习检测肺结节的有关进展,重点探讨了如何利用既有目标检测技术,并顾及了肺结节检测特点的各项创新改进。通过各方法在基准数据集上的性能对比,进一步分析了现有的肺结节检测方法的特色和优缺点。最后通过分析当前肺结节检测的应用现状,对未来如何更好地应用和改进深度学习CT影像肺结节检测技术提出了一些展望。我们认为随着机器学习和计算机视觉技术的飞速发展,基于医学影像的自动辅助诊断技术将迎来快速推广和

产业化契机。当前的瓶颈在于更好地在深度学习常规方法中结合医学专业和医学影像知识,提高检测精度和可解释性,从而更好地提升医疗服务水平。

参考文献

- [1] FERLAY J, SOERJOMATARAM I, DIKSHIT R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012[J]. *Int J Cancer*, 2015, 136(5): E359-E386.
- [2] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening[J]. *N Engl J Med*, 2011, 365(5): 395-409.
- [3] 周清华, 范亚光, 王颖, 等. 中国肺癌低剂量螺旋CT筛查指南(2018年版)[J]. *中国肺癌杂志*, 2018, 21(2): 67-75.
- [4] BAI C X, CHOI C M, CHU C M, et al. Evaluation of Pulmonary Nodules: clinical practice consensus guidelines for Asia[J]. *Chest*, 2016, 150(4): 877-893.
- [5] 刘宝东. 肺多发磨玻璃结节的诊治策略[J]. *中国肺癌杂志*, 2020, 23(8): 679-684.
- [6] ARMATO III S, 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.
- [7] LUO X D, SONG T, WANG G T, et al. SCPM-Net: an anchor-free 3D lung nodule detection network using sphere representation and center points matching[J]. *Med Image Anal*, 2022, 75: 102287.
- [8] MURPHY K, GINNEKEN B V, SCHILHAM A M R, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification[J]. *Med Image Anal*, 2009, 13(5): 757-770.
- [9] JACOBS C, RIKXOORT E M V, TWELLMANN T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images[J]. *Med Image Anal*, 2014, 18(2): 374-384.
- [10] HARDIE R C, ROGERS S K, WILSON T, et al. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs[J]. *Med Image Anal*, 2008, 12(3): 240-258.
- [11] NAQI S, SHARIF M, LALI I U. A 3D nodule candidate detection method supported by hybrid features to reduce false positives in lung nodule detection[J]. *Multimed Tools Appl*, 2019, 78: 26287-26311.
- [12] HAN H, LI L H, HAN F F, et al. Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme[J]. *IEEE J Biomed Health Inform*, 2015, 19(2): 648-659.
- [13] KHOSRAVAN N, BAGCI U. S4ND: single-shot single-scale lung nodule detection[C]// *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2018: 794-802.
- [14] PEZESHK A, HAMIDIAN S, PETRICK N, et al. 3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT[J]. *IEEE J Biomed Health Inform*, 2019, 23(5): 2080-2090.
- [15] MEI J, CHENG M M, XU G, et al. SANet: a slice-aware network for pulmonary nodule detection[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44(8): 4374-4487.
- [16] XIE H, YANG D B, SUN N N, et al. Automated pulmonary nodule detection in CT images using deep convolutional neural networks[J]. *Pattern Recognit*, 2019, 85: 109-119.
- [17] TAN J X, HUO Y M, LIANG Z R, et al. Expert knowledge-infused deep learning for automatic lung nodule detection[J]. *J XRay Sci Technol*, 2019, 27(1): 17-35.
- [18] SETIO A A A, TRAVERSO A, DE BEL T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge[J]. *Med Image Anal*, 2017, 42: 1-13.
- [19] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks[C]. *Advances in Neural Information Processing Systems*, 2012, NIPS 2012. Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012.
- [20] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. *IEEE Computer Society*, 2017.
- [21] KONG T, SUN F, LIU H, et al. Foveabox: beyond anchor-based object detector[J]. *IEEE Trans Image Processing*, 2020, 29: 7389-7398.
- [22] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2020.
- [23] LI Y M, FAN Y. DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection[J]. *Proc IEEE Int Symp Biomed Imaging*, 2020: 1866-1869.
- [24] TANG H, ZHANG C, XIE X. NoduleNet: decoupled false positive reduction for pulmonary nodule detection and segmentation[C]. *Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention, MICCAI 2019. Lecture Notes in Computer Science*, vol 11769. Springer, Cham.
- [25] HARSONO I W, LIAWATIMENA S, CENGGORO T W. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning[J]. *J King Saud Univ-Com Inform Sci*, 2020, 34(3): 567-577.

- [26] CARREIRA J, ZISSERMAN A. Quo Vadis, action recognition? a new model and the Kinetics dataset[C]. 17th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 21-26 July 2017, Honolulu, HI, USA.
- [27] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149.
- [28] GOLAN R, JACOB C, DENZINGER J. Lung nodule detection in CT images using deep convolutional neural networks[C]. International Joint Conference on Neural Networks. IEEE, 2016.
- [29] HUANG W K, XUE Y H, WU Y. A CAD system for pulmonary nodule prediction based on deep three-dimensional convolutional neural networks and ensemble learning[J]. PLoS One, 2019, 14(7): e0219369.
- [30] ZHU W T, LIU C C, FAN W, et al. Deep Lung: deep 3D dual path nets for automated pulmonary nodule detection and classification[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 12-15 March 2018, Lake Tahoe, NV, USA.
- [31] LIAO F Z, LIANG M, LI Z, et al. Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network[J]. IEEE Trans Neural Netw Learn Syst, 2019, 30(11): 3484-3495.
- [32] 谢未央, 陈彦博, 王季勇, 等. 基于卷积神经网络的CT图像肺结节检测[J]. 计算机工程与设计, 2019, 40(12): 3575-3581.
- [33] DING J, LI A X, HU Z Q, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks[C]// Proceedings of the 2017 International Conference on Medical Computing and Computer-Assisted Intervention. Berlin: Springer, 2017:559-567.
- [34] VAN GINNEKEN B, SETIO A, JACOBS C, et al. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans[C]// IEEE 12th International Symposium on Biomedical Imaging (ISBI), Brooklyn, NY, USA, 2015: 286-289.
- [35] SETIO A A A, CIOMPI F, LITJENS G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks[J]. IEEE Trans Med Imaging, 2016, 35(5): 1160-1169.
- [36] ZHENG S Y, GUO J P, CUI X N, et al. Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection[J]. IEEE Trans Med Imaging, 2020, 39(3):797-805.
- [37] LIU J, CAO L, AKIN O, et al. 3DFPN-HS2: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
- [38] KIM B C, YOON J S, CHOI J S, et al. Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection[J]. Neural Netw, 2019, 115: 1-10.
- [39] 孙华聪, 彭延军, 郭燕飞, 等. 3D多尺度深度卷积神经网络肺结节检测[J]. 中国图形图像学报, 26(7): 1716-1725.
- [40] GUO Z T, ZHAO L L, YUAN J L, et al. MSANet: multi scale aggregation network integrating spatial and channel information for lung nodule detection[J]. IEEE J Biomed Health Inform, 2022, 26(6): 2547-2558.
- [41] ZHANG C J, WANG L L, WU X, et al. 3D multi-branch encoder-decoder networks with attentional feature fusion for pulmonary nodule detection in CT scans[C]. International Joint Conference on Neural, 2021.
- [42] ZHANG H, PENG Y J, GUO Y F. Pulmonary nodules detection based on multi-scale attention networks[J]. Sci Rep, 2022, 12(1): 1466.
- [43] 凌铖. 基于生成对抗网络的医学图像数据增强技术研究[D]. 成都: 电子科技大学, 2020.
- [44] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.
- [45] PENG H X, SUN H C, GUO Y F. 3D multi-scale deep convolutional neural networks for pulmonary nodule detection[J]. PLoS One, 2021, 16(1): e0244406.
- [46] HUANG Y S, CHOU P R, CHEN H M, et al. One-stage pulmonary nodule detection using 3-D DCNN with feature fusion and attention mechanism in CT image[J]. Comput Methods Programs Biomed, 2022, 220: 106786.
- [47] 郭晓敏, 黄新. 改进YOLOv3算法在肺结节检测中的应用[J]. 激光杂志, 2022, 43(5): 207-213.
- [48] 黄文健, 国凤梅, 梁连英, 等. 基于深度学习的肺结节检测算法对亚实性结节的检出效果研究[J]. 影像研究与医学应用, 2021, 5(15): 27-29.
- [49] 李欣菱, 郭芳芳, 周振, 等. 基于深度学习的人工智能胸部CT肺结节检测效能评估[J]. 中国肺癌杂志, 2019(6): 336-340.
- [50] ZHENG S H, SHEN Z Q, PEI C H, et al. Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation[J]. Comput Methods Programs Biomed, 2021, 210: 106363.
- [51] HAN Y, QI H G, WANG L, et al. Pulmonary nodules detection assistant platform: An effective computer aided system for early pulmonary nodules detection in physical examination[J]. Comput Methods Programs Biomed, 2022, 217: 106680.