

# 多视图与注意力机制结合的肺结节检测算法

刘宇博, 刘国柱, 史操, 许灿辉

(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

**摘要:** 针对传统的计算机辅助诊断系统对低剂量肺部 CT 图像结节检出率低、假阳性高等问题, 提出一种类 U-Net 网络和基于注意力机制的两阶段肺结节检测模型。为了提高肺结节的检测速度和检出率, 首先构建了一个三维网络用于候选结节的检测, 充分利用结节的三维信息提高候选结节的检出率的同时, 优化了检出速度; 然后采用多视图输入方式以保证对结节空间特征的获取, 将结节在三维空间下的矢状面、冠状面、水平面等 9 个角度下的切片一起输入网络, 利用 ViT 模型做特征提取器并结合特征金字塔网络实现对结节的分类, 将所有切片结果融合以实现假阳性结节的筛查。最终在 LUNA16 数据集上的实验结果表明, 所提出的模型准确率达到 94.7%, 提高了准确率的同时降低了误诊率和漏诊率。

**关键词:** 深度学习; 肺结节检测; 候选结节; 假阳性筛查; 敏感度

**DOI:** 10.15938/j.jhust.2022.06.014

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1007-2683(2022)06-0115-09

## Pulmonary Nodules Detection Algorithm Combining Multi-view and Attention Mechanism

LIU Yu-bo, LIU Guo-zhu, SHI Cao, XU Can-hui

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

**Abstract:** To solve the problems of low detection rate and high false positive of nodules in low-dose lung CT images by traditional computer-aided diagnosis system, a two-stage pulmonary nodules detection model based on U-Net network and attention mechanism was proposed. In order to improve the detection speed and detection rate of pulmonary nodules, a 3D network was constructed to detect the candidate nodules firstly. It optimized the detection speed while the three-dimensional information of nodules was fully utilized to improve the detection rate of the candidate nodules. Then, the multi-view input method was used to ensure that the spatial features of nodules was obtained. The sections from 9 angles in three-dimensional space, including sagittal plane, coronal plane and horizontal plane, were input into the network together. The ViT network was used as a feature extractor and combined with the feature pyramid network to achieve the classification of nodules, and we fused all section results to achieve the screening of false positive nodules. The final experimental results on LUNA16 data set show that the accuracy of the proposed model reaches 94.7%, which improves the accuracy and reduces the rate of misdiagnosis and missed diagnosis.

**Keywords:** deep learning; pulmonary nodules detection; candidate nodes; false-positive reduce; sensitivity

收稿日期: 2021-09-01

基金项目: 国家自然科学基金(61806107, 61973180, 62002190).

作者简介: 刘宇博(1997—), 男, 硕士研究生;

史操(1981—), 男, 博士, 讲师.

通信作者: 刘国柱(1965—), 男, 硕士, 教授, E-mail: LGZ\_0228@163.com.

## 0 引言

随着空气质量的恶化,肺癌已成为全球致死率最高的疾病之一<sup>[1]</sup>,肺癌的早期形态表现为肺结节,尽管只有少数的肺结节是恶性病变,但结节的特征往往能表示结节的性质,因此肺结节的检测工作就变得至关重要。随着计算机断层扫描等技术的快速发展,低剂量的CT扫描成为一种十分有效地检测方法,但也大大增加了放射科医生的工作量。

设计先进的计算机辅助诊断系统能够辅助医生进行肺结节的筛查和精确定位<sup>[2-3]</sup>,它们不受情绪和疲劳的影响,可以快速的以相同的质量扫描影像中的任何地方。传统的肺结节检测方法主要是根据设定阈值来实现结节的分割<sup>[4]</sup>,这种方式很难将血管组织跟结节进行区分,同时也受CT图像质量的影响。后来,研究人员使用机器学习的方法进行肺结节检测,手动从候选结节中提取图像的底层特征,然后将这些特征输入到传统分类器中<sup>[5-6]</sup>,这样做结节的形状、尺寸及纹理等特征均需要人工提取,自动化程度较低,且训练量十分庞大,不易完成。随着深度学习的发展,研究人员将卷积神经网络引入到图像和视频的处理中取得了不错的效果。XIE等<sup>[7]</sup>最先使用Faster R-CNN算法来检测肺结节,该算法在LUNA16数据集上取得了86.4%的平均敏感度,但没有对针对小结节的检测进行调整,同时对结节空间特征提取有限。由于GPU限制,无法将二维检测直接推广到三维检测,因此DING等<sup>[8]</sup>利用二维区域建议网络从单个二维图像中提取建议框,然后将他们组合成三维建议框,实现了对结节的三维检测。刘迪等<sup>[9]</sup>提出了首先使用3D Faster R-CNN模型进行候选结节的检测,对候选结节进行 $32 \times 32 \times 32$ 大小的截取,再通过三维卷积神经网络进行假阳性的判断,这样做可以考虑结节的三维性质并取得不错的效果,但存在检测较慢且效率较低等问题。DOU等<sup>[10]</sup>针对不同的结节大小提出了一种多尺度输入的三维卷积神经网络,以此来适应肺结节自身尺度的变化,但该方法使用多个三维平行支路大大增加了计算量和模型复杂度。HONG等<sup>[11]</sup>将注意力机制与三维卷积神经网络相结合,并通过大量实验证明,注意力机制在肺结节检测中有良好表现。尽管目前很多网络取得了不错的效果,但目前的现有算法仍存在漏诊率较高、误诊率较高、检测效率较低等问题。

针对这些问题,本文分阶段的构建了两个网络

来实现对肺结节的检测。第一步建立了一个三维区域建议网络来直接预测结节的位置,既保证了检测速度的同时,充分考虑结节的三维特征,最大化的找出疑似结节的区域来防止漏诊。第二步建立基于注意力机制的多视图分类网络,将特征金字塔(feature pyramid networks, FPN)<sup>[12]</sup>与ViT网络<sup>[13]</sup>结合进行特征提取,将多视图结果融合,实现对疑似肺结节的精准分类来防止误诊。

## 1 肺结节检测模型

对于检测任务,分为one-stage和two-stage两种模式。在one-stage模式下,大体有两种处理方法,第一种是将边界框和概率进行同时预测,比如文[14],第二种是只对默认框的概率进行预测,而不是生成建议框,比如文[15]。对于two-stage模式,第一步通过检测网络得到候选结节;第二步在得到的候选结节中进一步剔除假阳性结节。一般来说单步检测速度比较快,分两步检测的准确率会更好。经实验论证,针对肺结节检测任务,两步检测网络整体性能明显优于单步检测。

因此,本文采用two-stage的检测模式,设计了两个网络来实现检测和去除假阳性结节的工作,首先参考了3D U-Net<sup>[16]</sup>网络的网络结构,构建三维区域建议网络,实现第一步的候选结节检测工作,然后提出了基于注意力机制的多视图分类网络实现第二步的去除假阳性结节的工作。

整个方法的实现流程如图1所示。首先对输入的CT序列进行预处理,根据给出的掩码图来实现肺实质部分的分割,然后对肺实质部分进行检测,输出所有疑似结节的区域,根据坐标将疑似结节裁剪后送入分类网络进行再次分类筛选,判断是否为真正结节。

### 1.1 候选结节检测

候选结节检测要尽最大可能的保证结节的召回率,本文方法受3D U-Net网络的启发,编码器与解码器之间采用跳跃连接的方式,在解码器输出部分,使用三维区域建议网络来实现结果的输出,这样做是参考单步检测系统将预测的建议框直接作为检测结果。

#### 1.1.1 候选结节检测网络结构

目标检测通常采用基于图像的训练方法,将整张图片作为网络输入,但由于GPU内存的限制,对于3D网络来说将整张CT图像输入到网络中是有困难的。为了解决这一问题,本模型将从肺扫描的

图片中截取  $96 \times 96 \times 96$  的 3D 块来作为网络的输入。

本网络结构保留了编码器和解码器中间的跳跃连接结构,它可以使网络最大程度的获取多尺度信息。因为结节的大小不一,该结构有助于得到结节在不同尺度下的特征信息。网络的输出层使用三维区域建议网络,该结构可以使网络直接生成建议并

返回疑似结节的坐标值,网络结构如图 2 所示。网络主干分为前馈和反馈两个路径,前馈网络先是经过了三层三维卷积,每一层为 24 个大小为  $3 \times 3 \times 3$  的卷积核,然后是经过 4 个残差块(Res blocks),每个残差块结束时经过一个三维最大池化层,池化层的大小为  $2 \times 2 \times 2$  且步长为 2。反馈网络先是经过一个反卷积层,反卷积层的步长为 2 且卷积核大小为  $3 \times 3 \times 3$ ,然后再经过一个反卷积层和一个组合单元,组合单元负责将一个前馈网络的 blob(前馈网络生成的特征组合)和一个反馈网络的 blob(反馈网络生成的特征组合)相连并将结果送到一个残差块中,在第二个组合单元中加入位置信息(Location 块)将组合结果送入残差块后得到特征图的尺寸为  $24 \times 24 \times 24 \times 131$ ,然后经过两层大小为  $1 \times 1 \times 1$  的卷积,通道数分别为 64 和 15,然后对输出层的输出维度进行调整,最终得到的输出结果尺寸为  $24 \times 24 \times 24 \times 3 \times 5$ 。

结合 RPN<sup>[17]</sup> 网络的工作方法,我们在特征图的每个位置上都设置三个不同尺度的 Anchor,大小分别为 10、30、60 mm,所以我们将网络的 4 维输出调整为  $24 \times 24 \times 24 \times 3 \times 5$ ,其中锚盒的数量为  $24 \times 24 \times 24 \times 3 \times 5$  个回归值为  $(o, d_x, d_y, d_z, d_r)$  分别表示预测概率、结节的坐标与直径。

原网络是通过上采样的过程将输出调整为原输入图像的大小,而在本方法中,模型上采样到  $24 \times 24 \times 24$  大小后不再上采样,而是通过特征图对通道数和输出结果进行调整,调整后每个点的 Anchor 位置如图 3 所示。这样,既可以保留目标的多尺度信息,又可以将检测结果直接输出,大大提高了检测效率。同时 3 个不同大小的 Anchor 可以更好地适应结节的大小,使得网络对小目标的检测更加敏感。

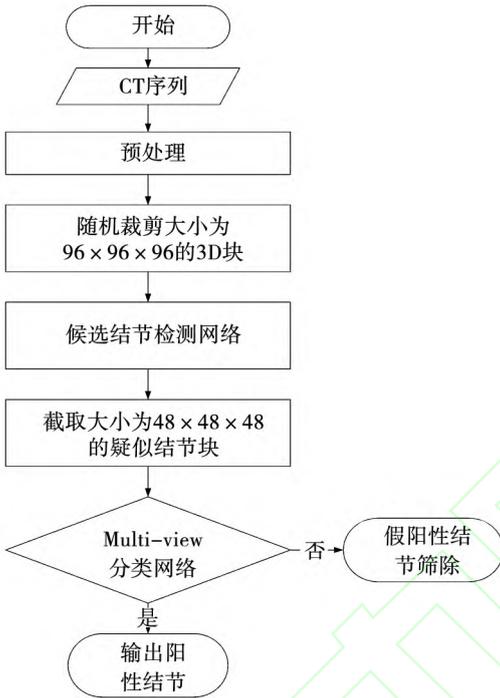


图 1 本文方法算法流程图  
Fig. 1 Algorithm flow chart

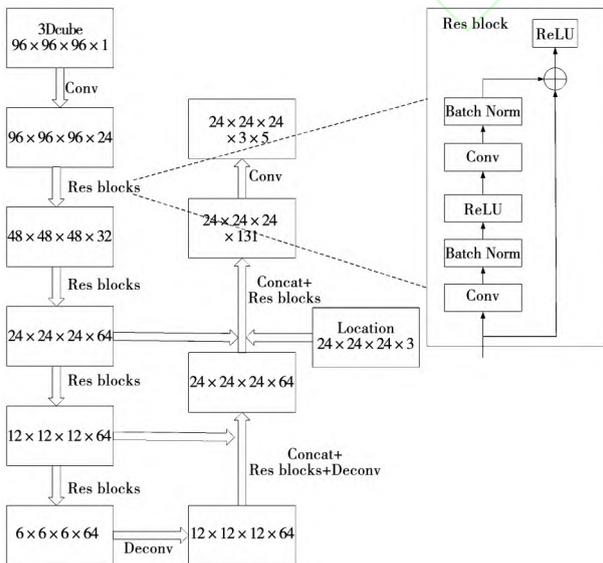


图 2 候选结节检测网络结构图  
Fig. 2 Network structure of candidate nodule detection

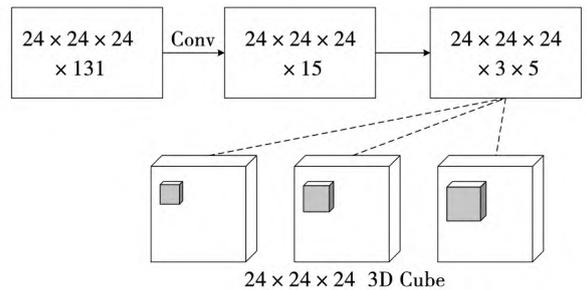


图 3 调整后锚点 Anchors 示意图  
Fig. 3 Adjusted Anchors schematic

### 1.1.2 候选结节检测网络损失函数

损失函数是由预测锚盒的分类部分损失函数和回归部分损失函数组成,使用 IOU 来确定每个锚盒

的标签,当锚盒与靶结节的 IOU 大于 0.5 时预测为阳性样本,IOU 小于 0.02 时预测为阴性样本。锚盒的预测概率用  $\hat{p}$  表示,锚盒的真实标签用  $p$  表示,分类的损失函数计算公式如下:

$$L_{cls} = p \log(\hat{p}) + (1 - p) \log(1 - \hat{p}) \quad (1)$$

式中  $p \in \{0, 1\}$ 。

对于回归部分的损失,用  $(x_g, y_g, z_g, r_g)$  表示目标结节的真实位置和直径,用  $(x_a, y_a, z_a, r_a)$  表示锚点的位置和直径,用  $(x, y, z, r)$  表示预测结节的位置和直径。回归的损失函数计算公式如下:

$$L_{reg} = \sum_k S(G_k, P_k) \quad (2)$$

$$G_k = \left( \frac{x_g - x_a}{r_a}, \frac{y_g - y_a}{r_a}, \frac{z_g - z_a}{r_a}, \log\left(\frac{r_g}{r_a}\right) \right) \quad (3)$$

$$P_k = \left( \frac{x - x_a}{r_a}, \frac{y - y_a}{r_a}, \frac{z - z_a}{r_a}, \log\left(\frac{r}{r_a}\right) \right) \quad (4)$$

其中  $S(\cdot)$  为光滑  $L1$  范数函数。式(3)表示真实边界框的回归值,式(4)表示预测边界框的回归值; $k$  为批量锚点的索引。

模型总的损失函数为:

$$L = L_{cls} + p^* L_{reg} \quad (5)$$

当目标结节为阳性样本时  $p^* = 1$ ,反之为阴性样本时  $p^* = 0$ 。这样设计损失函数是由于既要保证预测框更加接近于真实框,又需要对框内目标进行分类判断,故损失函数由两部分组成。在训练时负样本的数量会远远多于正样本的数量,然而模型的回归损失只与正样本有关,因此该损失函数可以在尽量保证正负样本平衡的同时,更好地进行回归损失计算,并在特殊情况下简化了损失函数。

### 1.2 假阳性结节筛除

在第一阶段对 CT 图像进行检测后得到大量的疑似结节,其中含有很多假阳性结节,第二阶段的任务是进行假阳性结节的筛除工作。由于疑似结节的数量很多,导致使用三维网络进行逐一甄别会造成计算需求量大、计算时间长、硬件需求量高等问题,从而降低了模型的整体效率。针对此问题本文采用多视图网络的处理方法,通过从结节不同的角度进行切割,来实现网络的多视图输入,最大程度的保留结节的三维空间信息。为了获取结节不同尺度下的特征信息,将 ViT 特征提取器与特征金字塔网络进行结合。最终将所有视图下得到的结果相融合实现对目标结节的分类,既解决了检测速度慢的问题,又保证了模型准确率。

#### 1.2.1 Multi-view 分类网络

结节具有一定的空间特征,仅从单一视角观察

很容易忽略其空间特征,因此本网络从多个视角下观察结节特征,构建多输入网络并使用 ViT 模型和特征金字塔网络进行特征提取,最后将所有视图下得到的结果相融合实现,构造了一个多视图分类网络。

基于注意力机制的网络在图像领域取得了很好的效果。传统的卷积网络存在着感受野固定且局限的问题,注意力机制可以在不影响结果的情况下取代卷积网络中卷积层的作用,从而学习到对图像的全局理解,更好的获取图片的关键信息,捕捉关键特征,增加了模型的可解释性,可以有效地从模型中检查注意力的分布。

在 ViT 网络的基础上,为了更好地获取结节特征,本文受特征金字塔网络的启发,对不同尺度的特征图和原图像进行特征提取。对于每一个输入图像而言,每一层不同尺度的特征图都有不同的语义信息,这样做可以最大限度的保存图片的语义信息。同时底层的特征图中具有更多更清晰的位置信息以及物体的纹理图案,而高层的特征图中有更高的语义信息,使用特征金字塔网络能够在增加较少计算量的前提下,将二者进行融合。多视图分类网络的模型结构如图 4 所示。

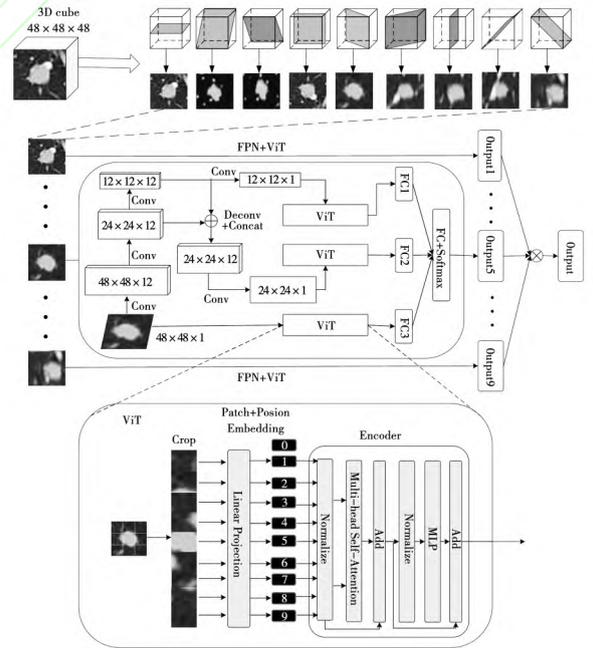


图 4 Multi-view 分类网络

Fig. 4 Multi-view classification of network

在第一阶段中,我们通过检测网络得到了疑似结节的坐标并将其从预处理后的图像中裁剪出一个以结节坐标为中心的  $48 \times 48 \times 48$  的 3D 图像块。从图像块的 9 个角度进行切割,包括结节的冠状面、矢

状面、水平面及其他 6 个对称平面,其中,上述 6 个对称平面是从对角线出发对图像块进行切割得到的,具体操作如图 4 最上方所示,并将得到的图片固定成相同大小。对每张图片我们进行相同操作,首先对图片进行一层卷积操作,卷积核大小为  $3 \times 3$  且卷积核个数为 12,然后再经过两层卷积,每层卷积都使用 12 个大小为  $3 \times 3$  的卷积核,每层卷积结束后都经过一个最大池化层,池化层的大小为  $2 \times 2$  且步长为 2。将后两层的特征图通过反卷积调整后相连,并通过卷积操作分别将通道数调整为 1,最后得到了两个通道数为 1 且尺度不同的特征图和一张原图像。

将得到的 3 个图像分别送入到 ViT 网络中,ViT 网络的具体操作如图 4 中 ViT 模块图所示。首先对每一张图进行裁剪处理,用  $(H, W, C)$  依次表示输入图片的高、宽和通道数,然后将图片分成大小为  $P \times P$  的 patches,那么图像切割后 patches 的数量  $N = \frac{H}{P} \times \frac{W}{P}$ 。在图 4 的示意图中每张图像大小为  $(48, 48, 1)$ ,每个 patch 的大小为  $16 \times 16$ ,图中白色线表示切割方式。然后将每个 patch 通过 Linear 层拉平成 1 维向量并对每个 patch 进行位置编码。将编码后的向量送入到 Encoder 层中,Encoder 层除了包括两个归一化层和两个叠加层外还有一个多头自注意力层<sup>[18]</sup>(multi-head self-Attention, MSA)和一个多层感知器。

在 Encoder 层内,注意力机制的主要工作原理是通过计算每个输入 patch 之间的相似性得分来找出关键特征,从而得到整个输入中相对重要的部分。每一个 patch 都通过线性变换得到  $Query(Q)$ 、 $Key(K)$ 、 $Value(V)$  3 个向量,其中  $Key$  和  $Value$  是成对存在的,通过计算  $Q$  和  $K$  之间的相似度作为重要性权重,对得到的权重通过 Softmax 函数进行归一化,然后将结果与  $V$  相乘得到加权后的  $V$ ,从而达到对重要 patch 的 value 值进行保留,对不重要的 patch 的 value 值进行削弱的效果。自注意力机制(self-attention)则是仅关注自身,即注意力的相关性计算仅在输入特征内部元素之间,不需要其他额外信息,此时  $Q = XW_Q$ 、 $K = XW_K$ 、 $V = XW_V$ , $X$  为输入的 patch 所拉平的 1 维向量。多头自注意力机制则是将  $Q$ 、 $K$ 、 $V$  映射到不同的子空间中,每个空间进行自注意力的计算然后将每个空间的结果连到一起,这样可以更好地捕捉每个输入之间的联系以及更好地获取关键特征,同时可以有效地防止模型的过拟合。多头自

注意力机制公式为:

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

$$head_i = Attention(XW_i^Q, XW_i^K, XW_i^V) \quad (7)$$

$$MSA(X) = \text{Concat}(head_1, \dots, head_n)W^O \quad (8)$$

其中: $\sqrt{d_k}$ 为缩放因子来避免点积带来的方差影响; $d_k$ 的值由 patch 的大小和  $h$  决定, $d_k = \frac{P \times P \times C}{h}$ ; $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$ 、 $W^O$  为可学习训练的矩阵; $X$  为输入的 patch 所拉平的 1 维向量; $h$  为子空间数量。我们可以通过递归计算所有层的权重矩阵得到经过 MSA 层后的注意力图像,以此来查看模型注意力的分布,如图 5 所示。在注意力集中的部分,结节位置有明显的亮度变化和轮廓突出。最后图片在经过一个多层感知器后得到一个全连接层输出,然后将 3 个图像得到的全连接层输出进行拼接得到该视角下的全连接层输出,通过 Softmax 函数得到预测结果。对不同视角下的 9 张图进行相同操作后,将得到的结果进行融合,得到结节最终的分类结果。

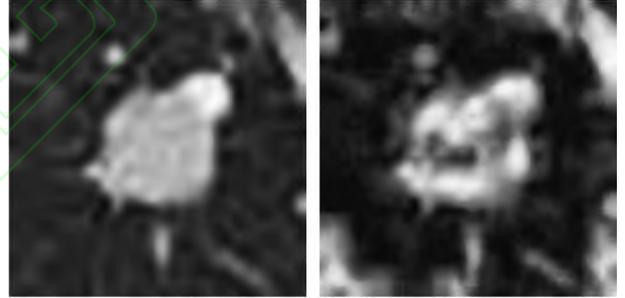


图 5 左为输入原图像,右为经过 MSA 后的注意力图像  
Fig. 5 The left is the original input image, and the right is the attention image after MSA

## 2 实验

### 2.1 数据及预处理

本文使用的数据集为 LUNA16 数据集,该数据集来源于 Lung Image Database Consortium (LIDC-IDRI)<sup>[19]</sup> 公开数据库,数据集中包含 888 例由四名放射科医生标注的患者的 1 186 个结节标记,结节的直径范围为  $3 \sim 30$  mm 之间。每一套 CT 由 .mhd 文件和 .raw 文件共同给出,两个文件分别给出了 CT 的一些基本信息和具体数据,数据集还给出了所有 CT 中医生所标记结节的具体位置信息,以及每套 CT 标准的肺实质分割掩码图。实验时,对 LUNA16 数据集进行十折交叉验证。

第1步所有的原始数据转换成亨斯菲尔德单位(HU),这是描述辐射密度的标准定量尺度,每个组织都有其特定的HU范围;第2步由于该数据集提供了标准分割的掩码图,因此求取掩码图的边缘得到含有肺实质的区域,然后对掩码图进行像素间隔为 $1 \times 1 \times 1$ 的重采样;第3步,对掩码图进行膨胀操作,并对原始数据应用操作后的掩码图,并将掩码外的数据值统一置为170(普通组织的亮度);第4步将原始数据裁剪在 $-1200 \sim 600$ 之间,然后将数据归一化至 $0 \sim 255$ ,将原始数据同样进行像素间隔为 $1 \times 1 \times 1$ 的重采样;第5步再截取区域内的数据即可得到预处理后的图片,预处理后的图片如图6所示。

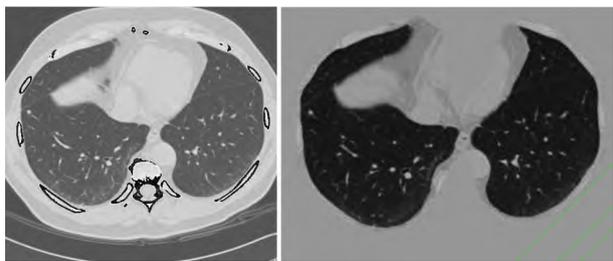


图6 左为CT序列原图像,右为预处理后图像

Fig. 6 The original CT sequence image is on the left, and the pre-processed image is on the right

## 2.2 实验评价指标

对于肺结节检测的评价标准如表1所示。

表1 肺结节检测评价标准

Tab. 1 Criteria for detection and evaluation of pulmonary nodules

标准结果			
诊断结果	阳性(病例)	阴性(对照)	合计
阳性	真阳性(TP)	假阳性(FP)	$TP + FP$
阴性	假阴性(FN)	真阴性(TN)	$FN + TN$
合计	$TP + FN$	$FP + TN$	$TP + FP + FN + TN$

其中:TP表示本应为结节且模型预测结果也为结节的个数;FP表示本不是结节但预测结果为结节的个数;FN表示本应为结节预测但结果不是结节的个数;TN表示本不是结节且预测结果也不是结节的个数。本文采用敏感度(TPR)、特异性(TNR)、准确率(ACC)、受试者操作特征自由响应曲线(free-response receiver operating characteristic, FROC)和竞争性指标(competition performance metric, CPM)作为实验性能的评价指标,其中敏感度、特异性、准确率计算公式如下:

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$TNR = \frac{TN}{TN + FP} \quad (10)$$

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (11)$$

## 2.3 实验结果和分析

### 2.3.1 实验环境

本实验的实验环境为Ubuntu16.04操作系统,硬件为1块NVIDIA RTX3090显卡,编程语言为Python3.7,使用的深度学习框架为PyTorch。

### 2.3.2 消融实验结果分析

本文使用k折交叉验证的方式对数据集进行训练和测试。由于在k折交叉验证中,不同k的取值会对实验结果产生影响,因此本文首先通过实验对k的取值进行确定。实验结果见表2。

表2 不同k值下模型表现

Tab. 2 Model performance at different k-values

k	TPR/%	TNR/%	ACC/%
5	89.8	92.3	91.0
7	92.6	94.9	93.7
10	94.6	96.8	95.7
15	93.9	97.1	95.5

由于进行交叉验证需要大量的时间,本文选择使用5、7、10、15折交叉验证进行实验。由表2可知使用10折和15折时取得不错效果,但折数越多训练数据也变多,训练时间也会相应变长。因此综合考虑,本文采用10折交叉验证的方式进行实验,并分别在10折数据上进行TPR、TNR、ACC的计算,最终结果取十次的平均值,结果如表3所示。可见,模型在各项指标下的表现优异且平稳。

表3 10折交叉验证下模型表现

Tab. 3 10 fold cross-validation under the model performance

Fold	TPR/%	TNR/%	ACC/%
0	93.8	96.3	95.1
1	94.6	95.9	95.3
2	94.6	96.8	95.7
3	93.9	97.5	95.7
4	93.4	97.0	95.2
5	93.1	95.9	94.5
6	95.3	97.0	96.2
7	96.0	96.8	96.4
8	95.2	96.4	95.8
9	93.6	97.1	95.4

为了检验第二阶段假阳性剔除模型中每个部分的具体效果,本文进行了五组消融实验,实验结果见表 4。其中第 1 组为使用 ResNet50 作为骨干网络进行特征提取并进行分类得到的结果,第 2 组为只使用 ViT 网络作为特征提取器来实现分类的结果,前两组主要用于展示注意力机制对特征提取部分的提升效果,第 3 组为使用 FPN + ViT 网络相融合进行特征提取实现分类的结果,对比可得,如果仅使用 ViT 网络仍会有一些数量的漏诊、假阳结节的出现,加入 FPN 网络后,可以获得不同尺度下特征图的语义信息,最大限度的保存图片特征信息,敏感度和特异性均有提高,第 4 组为使用 Multi-view + ViT 网络得到的结果,由于结节具有一定的空间特征,仅从单一视角观察很容易忽略其空间特征,所以采用多视图输入与第一组相比,结果有一定提高,第 5 组将三部分进行融合后得到 Multi-view + FPN + ViT 网络的分类结果,网络最终得到了 94.6% 的敏感度和 96.8% 的特异性,并将最终的准确率提高到了 95.7%。

表 4 消融实验结果  
Tab. 4 Ablation results

方法	TPR/%	TNR/%	ACC/%
ResNet50	77.8	81.3	79.5
ViT	79.6	83.9	81.7
FPN + ViT	85.6	88.8	87.1
Multi-view + ViT	87.6	91.1	89.2
Multi-view + FPN + ViT	94.6	96.8	95.7

### 2.3.3 对比实验结果分析

为了检验模型在假阳性结节剔除阶段的提升效果,本文采用敏感度和准确率两个评价指标与之前类似方法进行对比。为了展示模型在检测效率和运行效率方面的表现,本文采用每秒传输帧数(frames per second, FPS)和浮点运算次数(floating point operations, FLOPs)来对模型进行评价,其中, FPS 表示每秒钟检测图像的帧数以此来表示模型的检测效率, FLOPs 用于衡量模型的复杂度以此来表示模型的运行效率,对比结果如表 5 和表 6 所示。

文[20]采用的是多视图网络的输入方式,使用二维卷积神经网络进行特征提取;文[21]设计了一个三维卷积神经网络用于检测和分类;文[22]设计了一个深度三维 DPN 网络来进行结节分类,由表 5 可知,相比于二维检测方法,本文方法可以很大程度提高敏感度和准确率,相比于纯三维检测方法,本文

方法极大地降低了计算成本,无论是检测效率还是运行效率都明显优于三维网络,在保证敏感度和准确率的同时保证了模型的检测效率和运行效率。

表 5 不同分类器的分类结果比较

Tab. 5 Comparison of classification results of different classifiers

方法	2D/3D	TPR/%	ACC/%	FPS	FLOPs/亿
文[20]方法	2D	85.4	84.9	45.13	6.88
文[21]方法	3D	85.96	85.62	20.08	28.52
文[22]方法	3D	95.80	90.44	16.14	44.63
本文方法	3D+2D	94.6	95.7	40.72	8.21

为了进一步检验模型的整体性能,本文采用 CPM 评分和 FROC 曲线将其在相同数据集上与当前比较新的其他模型进行对比,如图 7、表 6 所示。

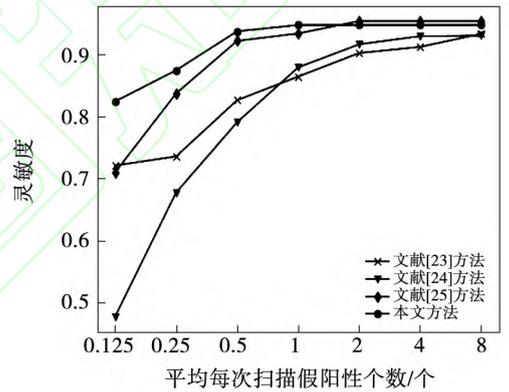


图 7 对比模型的 FROC 曲线对比

Fig. 7 Compare the FROC curve of the mod

由表 6 可知,在假阳性率分别为 0.125、0.25、0.5、1、2、4、8 时,本文所提出的方法无论是敏感度还是 CPM 评分都取得了不错的效果,文[23]采用的单阶段检测方法,虽检测速度快但检测效果与两阶段检测方法相比还是有很大的差距;文[24]是采用多尺度策略进行三维深度卷积神经网络来实现肺结节的检测,该方法无论在敏感度还是 CPM 评分上都逊色于本文提出的方法;文[25]为 S4ND 模型,该模型使用单个网络的单个前馈进行检测,构建了较为深层的三维卷积神经网络,虽然在假阳性率在 2 到 8 时的敏感度要比本文方法略高,但三维网络需要更多的计算成本和检测时间,而本方法使用二维网络对结节进行快速且准确的分类大大提高了模型效率,同时当假阳性率足够低时本文方法的表现效果更好。由表 7 可知,由于本文使用十折交叉验证,故测试过程中参与检测的样本个数为 89 个,在模型运行效率对比上,本文方法相对于两个三维方法在速度

上有着很大优势,与二维方法相比,随速度稍有不及,但由表6可知准确率有显著提高,因此综合来看,本文所提出的两阶段检测方法,在检测效率和运行效

率上有良好的表现,并取得了94.63%的敏感度和0.9182的CPM评分。

表6 不同方法的实验结果对比及 CPM 评分

Tab. 6 Comparison of experimental results and CPM score of different methods

方法	假阳性率							CPM
	0.125	0.25	0.5	1	2	4	8	
文[23]方法	0.720 4	0.735 0	0.825 5	0.863 4	0.901 3	0.911 2	0.932 1	0.841 0
文[24]方法	0.477 3	0.676 9	0.790 0	0.878 2	0.915 3	0.928 4	0.929 2	0.795 0
文[25]方法	0.709 3	0.836 2	0.920 8	0.932 7	0.952 7	0.952 7	0.952 7	0.896 7
本文方法	0.823 4	0.873 5	0.936 2	0.946 3	0.946 3	0.946 3	0.946 3	0.918 2

表7 对比实验运行效率

Tab. 7 Comparison of experimental operation efficiency of different methods

方法	FPS	FLOPs/ 亿	检测样 本总数	检测总 时长/min	平均时长 (秒/样本)
文[23]方法	55.28	12.38	89	7.92	5.45
文[24]方法	25.23	28.78	89	17.6	12
文[25]方法	10.30	43.34	89	44	30
本文方法	40.72	8.21	89	11	7.5

### 3 结 论

针对低剂量 CT 图像的肺结节检测任务,本文提出了一种两阶段的检测方法,第一步通过候选结节检测网络对 CT 图像进行检测,得到疑似结节;第二步通过基于注意力机制的多视图多尺度分类网络,对疑似结节进行筛选确认。使用 ViT 网络做特征提取器来更好的获取图片的关键信息,捕捉关键特征,增加了模型的可解释性;利用 FPN 网络能够最大限度的保存图片的高层语义信息和底层空间信息;选择多视图输入的方式可以最大程度的获取结节的三维空间特征。本文方法相比于其他二维检测方法有着更高的敏感度和准确率,相比于三维方法在增加检测效率和降低计算成本的同时保证了检测结果的准确率,降低了假阳率和漏诊率,对于医生的辅助诊断有一定的帮助和实际应用价值。

### 参 考 文 献:

- [1] TORRE Lindsey A, SIEGEL Rebecca L, JEMAL Ahmedin. Lung Cancer Statistics[J]. *Advances in Experimental Medicine & Biology* 2016, 893: 1.
- [2] 刘露, 杨培亮, 孙巍巍, 等. 深度置信网络对孤立性肺结节良恶

性的分类[J]. *哈尔滨理工大学学报* 2018, 23(3): 13.

LIU Lu, YANG Peiliang, SUN Weiwei, et al. DBN Classifier for Classification of Benign and Malignant Solitary Pulmonary Nodule [J]. *Journal of Harbin University of Science and Technology*, 2018, 23(3): 13.

- [3] 王卫兵, 王卓, 徐倩, 等. 基于三维卷积神经网络的肺结节分类[J]. *哈尔滨理工大学学报*, 2021, 26(4): 7.

WANG Weibing, WANG Zhuo, XU Qian, et al. Lung Nodule Classification Based on 3D Convolutional Neural Network [J]. *Journal of Harbin University of Science and Technology*, 2021, 26(4): 7.

- [4] 石羽, 曹晓光, 张弘, 等. 肺结节自动检测算法研究[J]. *仪器仪表学报* 2006, 27(3): 2265.

SHI Yu, CAO Xiaoguang, ZHANG Hong, et al. Study of Automatic Detection Algorithm for Lung Nodules [J]. *Chinese Journal of Scientific Instrument* 2006, 27(3): 2265.

- [5] 陈树越, 晁亚, 邹凌. 基于几何特征的孤立性肺结节检测[J]. *生物医学工程学杂志* 2016, 33(4): 680.

CHEN Shuyue, CHAO Ya, ZOU Ling. Detection of Solitary Pulmonary Nodules Based on Geometric Features [J]. *Journal of Biomedical Engineering* 2016, 33(4): 680.

- [6] 陈侃, 李彬, 田联房. 基于局部阈值和聚类中心迭代的肺结节检测算法[J]. *计算机科学* 2012(2): 302.

CHEN Kan, LI Bin, TIAN Lianfang. Pulmonary Nodules Detection Algorithm Based on Local Threshold and Iterative of Clustering Center [J]. *Computer Science*, 2012(2): 302.

- [7] XIE Hongtao, YANG Dongbao, SUN Nannan, et al. Automated Pulmonary Nodule Detection in CT Images Using Deep Convolutional Neural Networks [J]. *Pattern Recognition* 2019, 85: 109.

- [8] DING Jia, LI Aoxue, HU Zhiqiang, et al. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks [C]// *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham 2017: 559.

- [9] 刘迪, 王艳娇, 徐慧. 基于深度学习的医学图像肺结节检测[J]. *微电子学与计算机* 2019, 36(5): 5.

LIU Di, WANG Yanjiao, XU Hui. Pulmonary Nodule Detection of Medical Image Based on Deep Learning [J]. *Microelectronics &*

- Compute 2019 ,36(5):5.
- [10] DOU Qi , CHEN Hao , YU Lequan , et al. Multi-level Contextual 3D CNNs for False Positive Reduction in Pulmonary Nodule Detection [J]. IEEE Transactions on Biomedical Engineering , 2016: 1558.
- [11] 洪敏杰 , 吴刚 , 刘星辰 , 等. 基于注意力机制的肺结节检测算法[J]. 计算机工程与设计 2021 ,42(1):83.  
HONG Minjie , WU Gang , LIU Xingchen , et al. Detection Algorithm of Lung Nodule Based on Attention Mechanism[J]. Computer Engineering and Design 2021 ,42(1):83.
- [12] LIN Tsunyi , DOLLAR Piotr , GIRSHICK Ross , et al. Feature Pyramid Networks for Object Detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2017: 2117.
- [13] DOSOVITSKIY Alexey , BEYER Lucas , KOLESNIKOV Alexander , et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [J]. arXiv preprint arXiv: 2010.11929 2020.
- [14] REDMON Joseph , FARHADI Ali. YOLO9000: Better , Faster , Stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2017: 7263.
- [15] LIU Wei , ANGUELOV Dragomir , ERHAN Dumitru , et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision. Springer , Cham 2016: 21.
- [16] CICEK Ozgun , ABDULKADIR Ahmed , LIENKAMP Soeren S. , et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Springer , Cham , 2016: 424.
- [17] REN Shaoqing , HE Kaiming , GRISHICK Ross , et al. Faster R-cnn: Towards Real-time Object Detection with Region Proposal Networks [J]. Advances in Neural Information Processing Systems 2015 ,28: 91.
- [18] TAN Hongchen , LIU Xiuping , TIAN Shengjing , et al. MHSA-Net: Multi-Head Self-Attention Network for Occluded Person Re-Identification [J]. arXiv preprint arXiv:2008.04015 2020.
- [19] ARMATO Samuel G. , ROBERTS Rachael Y. , McNitt-Gray Michael F. , 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]. Academic Radiology 2007 ,14(12):1455.
- [20] SETIO Arnaud , CIOMPI Francesco , LITJENS Geert , et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks [J]. IEEE Transactions on Medical Imaging 2016 ,35(5):1160.
- [21] LIAO Fangzhuo , LIANG Ming , LI Zhe , et al. Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network [J]. IEEE Transactions on Neural Networks & Learning Systems , 2019 ,30(11): 3484.
- [22] ZHU Wentao , LIU Chaochun , FAN Wei , et al. DeepLung: Deep 3d Dual Path Nets for Automated Pulmonary Nodule Detection and Classification [C]//2018 IEEE Winter Conference on Applications of Computer Vision ( WACV) . IEEE , 2018: 673.
- [23] 刘若愚 , 刘立波. 基于改进全卷积网络模型的肺结节检测 [J]. 激光与光电子学进展 2020 ,57(16):10.  
LIU Ruoyu , LIU Libo. Detection of Pulmonary Nodules Based on Improved Full Convolution Network Model [J]. Laser & Optoelectronics Progress 2020 ,57(16):10.
- [24] GU Yu , LU Xiaoqi , YANG Lidong , et al. Automatic Lung Nodule Detection Using a 3D Deep Convolutional Neural Network Combined with a Multi-scale Prediction Strategy in Chest CTs [J]. Computers in Biology and Medicine , 2018 ,103:220.
- [25] KHOSRAVANNaji , BAGCI Ulas. S4ND: Single-shot Single-scale Lung Nodule Detection [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer , Cham 2018: 794.

( 编辑: 温泽宇)