

湖州师范学院

2023 届毕业设计(论文)

课 题 名 称: 基于 YOLOv5 的肺结节目标检测
技术研究

课 题 名 称 (英文): Research on Lung Nodules
Target Detection Technology Based on YOLOv5

学 生 姓 名: 张慧清 学 号: 2019082324

专 业 名 称: 计算机科学与技术

指 导 教 师: 郝秀兰 职 称: 副教授

所 在 学 院: 信息工程学院

完 成 日 期: 2023 年 3 月 22 日

教务处制表

基于 YOLOv5的肺结节目标检测技术研究

摘要：传统的医生看病人的肺部 CT 图像主要依靠肉眼的方式，随着深度学习的发展，可以将深度学习的技术应用到 CT 图像的识别中来。由于深度学习对肺结节的检测效果好、速度快，从而被广泛应用于医疗系统。在本实验中采取 YOLOv5作为基础模型展开对肺结节检测的研究，选择 Pytorch 深度学习包构建神经网络，采用公开肺结节数据集(LIDC-IDRI)作为训练数据，作者对其进行了标注，标注工作由4名经验丰富的医生独立完成，以确保数据集的准确性。数据集中包含了来自1010个患者的1018套全肺 CT 扫描图像。对训练集和验证集采取9：1的分配策略进行训练。查看随着训练加深时：损失的变化情况、正确分辨肺结节的能力，以及得到模型的平均检测精度。实验表明，在训练200步时，检测的精度达到了0.7202。mAP 在取0.5的置信度下为0.6678，值得注意的是，召回率达到了0.6270。该实验数据证明，YOLOv5对肺结节有较好的检测能力。

YOLOv5模型相对于之前的 YOLO 系列算法，创新性的提出了正样本分配策略来加强模型的查全能力。依旧采用单阶段检测结构使得网络保持较高的检测速度，对网络结构和预测头也做了一定优化。所以本实验采取 YOLOv5来实现对肺结节的检测。

关键词：肺结节, CT 图像, 识别检测, YOLOv5, 深度学习

Research on lung nodules target detection technology based on YOLOv5

Abstract: Traditional doctors mainly rely on the naked eye to see patients' lung CT images. With the development of deep learning, deep learning technology can be applied to the recognition of CT images. Deep learning is widely used in medical system because of its good effect and fast speed in the detection of pulmonary nodules. In this experiment, YOLOv5 was adopted as the basic model to carry out the research on the detection of pulmonary nodules, Pytorch deep learning package was selected to build the neural network, and open lung nodules data set (LDC -IDRI) was used as the training data. The author annotated the data, which was independently completed by four experienced doctors to ensure the accuracy of the data set. The data set included 1018 whole-lung CT scans from 1,010 patients. A 9:1 allocation strategy was adopted for training the training set and verification set. Look at the changes in losses, the ability to correctly distinguish lung nodules, and the average detection accuracy of the obtained model as the training progresses. The experiment shows that the accuracy of detection reaches 0.7202 when training 200 steps. mAP is 0.6678 with confidence of 0.5, and it is worth noting that the recall rate reaches 0.6270. The experimental data proved that YOLOv5 had a good ability to detect pulmonary nodules.

Compared with previous YOLO series algorithms, YOLOv5 model innovatively proposes a positive sample allocation strategy to enhance the model's full search ability. The single-phase detection structure is still used to maintain a high detection speed, and the network structure and prediction head are optimized to some extent. Therefore, YOLOv5 was adopted in this study to realize the detection of pulmonary nodules.

Keywords: Lung nodules, CT images, Recognition and detection, YOLOv5, deep learning

目 录

第一章 绪 论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	1
1.3 课题研究的主要内容	2
第二章 相关理论基础	4
2.1 肺结节检测基础	4
2.2 YOLO 算法介绍	5
2.2.1 神经网络概述	5
2.2.2 神经元模型	5
2.2.3 卷积神经网络	6
2.2.4 激活函数	6
第三章 构建基于 YOLOv5 的肺结节检测模型	9
3.1 目标检测网络	9
3.2 YOLOv5 网络结构	9
3.3 YOLOv5 模型检测流程	11
3.4 YOLOv5 模型优化策略	11
3.4.1 IoU 与非极大抑制策略	11
3.4.2 特征提取网络与预测头	12
3.4.3 正样本分配策略	13
第四章 仿真实验	14
4.1 数据获取与构建	14
4.1.1 数据集选取	14
4.1.2 数据集标注信息	14
4.1.3 数据集预处理	15
4.2 YOLOv5 模型构建	15
4.3 实验结果分析	16
4.3.1 评价指标	16
4.3.2 实验结果	17
4.3.3 肺结节检测结果展示	19
第五章 总结与展望	21
参考文献	22
致谢	23

第一章 绪论

1.1 研究背景和意义

肺癌是全球致死率最高的疾病，肺癌病人死亡率近几年排在恶性肿瘤第一位，无论是在繁华的城市还是偏远的乡村，成为威胁着人类生命健康的巨大毒瘤。到目前为止，医学界对晚期癌症的治疗手段很难达到一个稳定且有效的水平，那么对于作为肺癌早期形态的肺结节的诊断和治疗就显得尤为重要，“早发现早治疗”的基本思想在一定程度上可以提高肺癌病人的治愈率。

电子计算机断层扫描仪器（CT）对于肺癌的筛查起着主要作用，由于其操作较简单、效率较高、扫描出来的图像清晰，可以清楚准确的反映出肺结节的位置和大小。有着其他检查途径不可代替的诊断价值。医生在拿到扫描影像以后，根据以往经验对其大小和形状以及所处的位置进行全面的诊断，但是许多因素的存在会导致医生的误诊，比如在观察大量 CT 图像以后，医生可能会出现眼部疲劳导致误诊，或者思绪开小差对诊断结果造成的主观性误差，那么此时应用计算机辅助系统来诊断肺结节就显得尤为重要，一方面减轻了医生的工作量，另外也提升了对肺结节的检测精度。

随着深度学习的发展，计算机技术在医疗领域也得到了很大的普及，医学影像的进展速度也随之加快。计算机扫描得到的医学图像包含数十上百张图片，这些图片又包含着目标信息不同的角度，形成了一个切片，计算机对切片的推理过程较为复杂，原因是切片本身的坐标具有不确定性，而病变更是五花八门，会增加推理的复杂程度，机器学习算法的渗透可以减少这些阻碍，本研究采用 YOLO v5 模型，对肺结节公开数据集进行训练，训练得到的模型要对肺结节 CT 图像有较强的检测能力。

1.2 国内外研究现状

上世纪中后叶，对于医学图像的处理主要运用数学建模来完成医疗任务的分析工作，以较低级的像素处理复杂的图片信息。非机器学习类的方法主要依靠支持向量机（SVM）原理，分割往往达不到深度学习算法的分割效果，基准是检测算法的准确率。但由于肺结节的异质性和复杂性，会增加分割的难度以及分割的准确性，造成这一难题的原因是结节周围的视觉特征相似性，另外并排结节的存在会加大分割难度，并排结节在强度这一特征上与肺壁的相似度很高，毛玻璃结节具有不透明的性质，它在周围的对比度较低，那么简单的阈值分割和形态检测方法，就不能有效分割这些结节，OTSU 算法恰巧可以处理此类结节，OTSU 基于一维灰度直方图来实现自动分割，该算法具有计算简单方便，无参数无监督的优点，但 OTSU 不适用于 GGO 结节。潘子妍等提出了一种新的基于 XGBoost 的多个特征融合方法^[1]，并对其进行了实验验证。首先采用门限划分方法对各结点进行候补，然后采用不变矩方法、快速稳健的特征方法、灰度浓缩矩阵方法对各结点进行了多个候补结点的局部及全局特征的提取。最后，利用 XGBoost-决策树的方法，对有误的结节进行剔除。陶善礼提出了一种基于互熵法的局部 Loss 丢失函数^[2]，解决了正反两个样本之间的不平衡问题，从而改变了模型梯度的更新方向。李琬玮在前期工作中，提出了一种新的 CT 影像预处理算法^[3]，该算法在不同的假阳性情况下，对不同类型的 CT 影像进行了检测，其灵敏度可提高 1%-3%。单基甸在肺 CT 图像上对肺结节大小的分布特征进行了研究^[4]，因为肺结节在平面上可以被矩形框标出一个类似于圆形的、小的靶子。于是，Faster-R-CNN 网络成为了一个主要的网络，它可以在网络中调整所产生的锚框的比例，从而得到一个更适合于它的锚框，并且，它还可以在卷积层之后加入一个反卷积层，将一些浅层的特征还原出来，然后再将特征图放大，从而提高该网络对于肺结节的检测能力。刘洋使用基于像素伽马变换增强进行图像质量提升^[5]，使用最大类间方

差法进行前背景分离,结合形态学操作完成肺实质修补,用于细化颗粒、平滑边缘和填充血管,精细化分割肺部实质,实现显著性区域增强。闫思桃在 YOLOv3 的基础上进行改进^[6],并基于改进后的模型搭建基于深度学习的肺结节筛查系统,并加入了病人信息管理模块。霍恩泽试在 FasterR-CNN 中,选择适用于多种肿瘤的参照框架,使其与当地肿瘤的特征相匹配,并利用 Focal 分类器的损失函数对非对称的背景和肺部肿瘤进行均衡^[7]。徐时怀在前期研究基础上,提出了一种基于 CascadeR-CNN 的新型多级串联检测器^[8],并对其进行了验证。采用 CascadeR-CNN 算法,通过逐步提高图像中的边缘盒的质量,从而提高图像中的病变位置。张仕瑞提出了“分离的交叉棱锥”网络^[9],在该网络中将产生的对象的测量框架可信度与位置坐标进行了部分的分离,并将其上下交错接收到不同深度的特征图。席孝倩在前期工作中,针对不同类型的肺结节,研究了一种新的肺结节自动监控方法^[10],并在此基础上,对其进行了详细的分析,并对其进行了详细的分析。潘成成提出一种用于候选肺结节检测的 3D DPN-U Faster R-CNN 模型^[11]。该模型是在 3DDPN-Unet 为特征提取网络的基础上改进的,而 3D DPN-Unet 网络是由 3D 双路径网络(DualPath Network, DPN)为构建块在 Unet 结构的基础上设计的。

对于小直径的结节,它与周围噪声较为相似,在降噪的处理中,很容易漏掉这些小的结节,因此需要深度学习算法对其进行检测。随着计算机和有关医学图像处理技术的发展,医学诊断技术的发展迅猛向前,许多研究成果都是通过提取图像的数据特征来训练模型,进而缩短训练模型的时间,大大减少了任务量。

Dong 等人提出了一个基于密集连接的三维肺结节分割网络模型^[12],该网络模型由三个模块组成,包括一个编码器(E)、一个解码器(D)和一个检测器(A)。其中 E 模块提取了肺结节的形状异质性特征,D 模块提取了肺结节的局部和全局尺度特征,A 模块则提取了肺区域内和肺间的空间异质性特征。B 模块根据上述特征从 D 模块中学习感兴趣区域分类。其利用 3D ResNet 网络提取肺结节形状异质性特征,并将三维图像转换为多视角分割图像,然后使用 FasterR-CNN 检测算法确定感兴趣区域是否属于结节。

近年来,深度学习在肺结节检测方面确实有了较好的检测能力。本实验采用检测能力较成熟的 YOLOv5 模型来检测肺结节,测试实验模型的检测能力,为基于深度学习的肺结节研究提供数据支持。

1.3 课题研究的主要内容

本实验采用 YOLOv5 模型,采用 Pytorch 科学计算包构建深度学习网络。对肺结节公开数据(LIDC-IDRI)集进行目标检测。首先研究数据集的结构与标注信息,再对标注信息进行格式转换,输入到网络模型中,调整模型的 batch-size、epoch、学习率等相关参数,经过 GPU 加速的环境训练,得到肺结节检测的平均检测精度值(mAP)、精度与置信度和召回率的函数图像,以及分类准确性的判断矩阵等相关指标,分析实验结果。研究方案如图 1-1 所示。

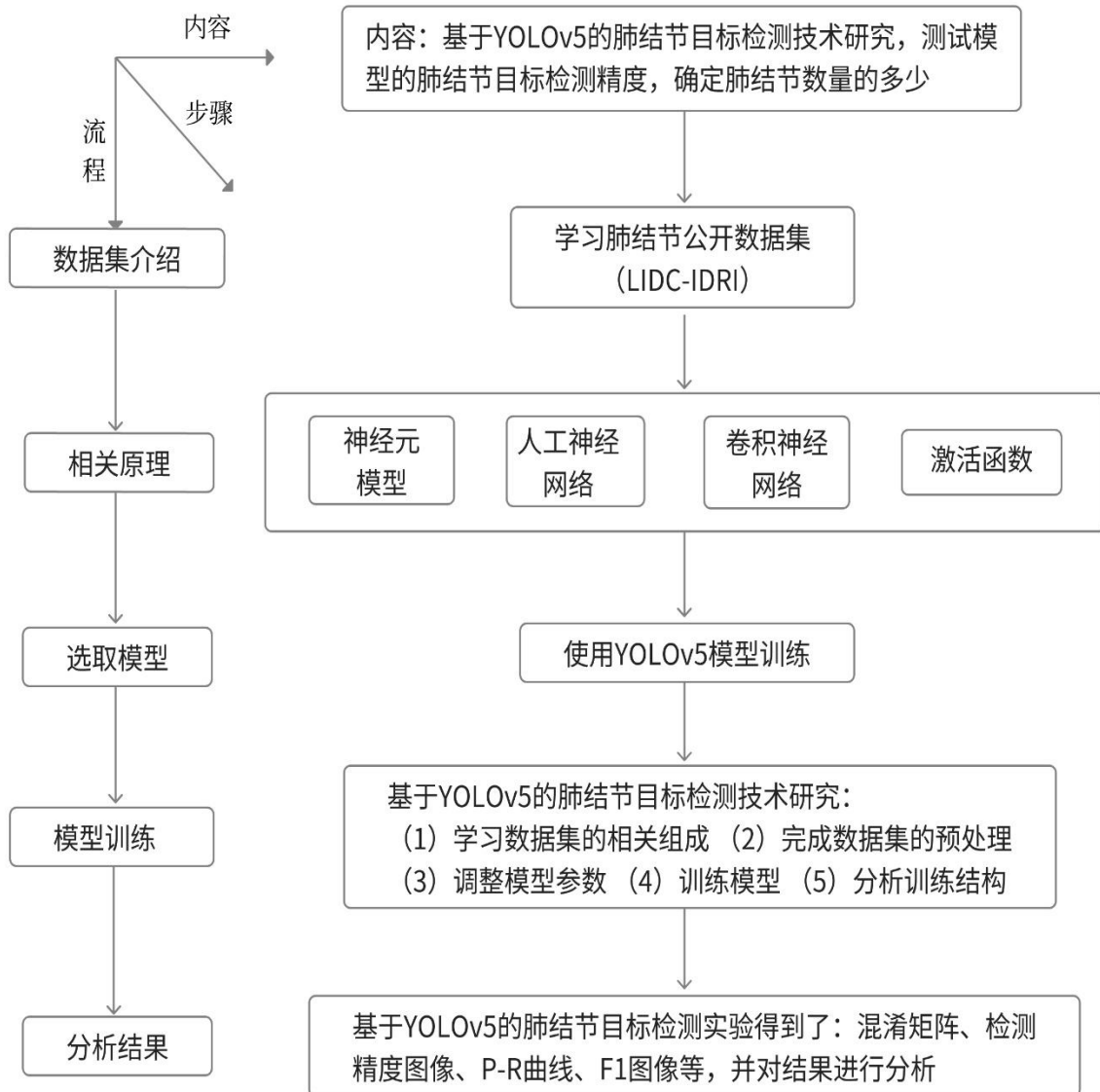


图 1-1 研究方案

第二章 相关理论基础

2.1 肺结节检测基础

肺结节是检测肺癌的开端，因为大多数病人可以自行缓解甚至自愈，病情稳定或者是无症状的患者并不需要治疗，所以判断肺结节是良性还是恶性的指标是肺结节的大小，一般肺结节直径较小的为良性，直径较大的判定为恶性。良性结节对人体没有影响，不具有致癌性，不会累及其他器官；恶性结节不易控制且生长速度过快，容易扩散从而累及其他器官。

最常见的结节类型是 SNS，而亚实性结节中的纯磨玻璃结节是早期一些专家判肺癌的重要依据，纯磨玻璃结节在影像学检查上是肺内模糊的结节影，其密度要稍微高于周围的肺间质，通俗来说就是一团比附近组织密度稍大的浑浊，结节内部的支气管以及血管的轮廓仍可见，但是因为结节的大小与肺癌息息相关，所以准确测量结节的直径就显得尤为重要。一些学者就对结节的直径与诱发肺癌的关系进行了研究^[13]，最终根据 ELCAP 数据库确定了在肺结节直径小于 5mm 时，无明显的症状和体征，为良性结节，发展成为恶性结节的概率为 1%；结节直径在 6mm~10mm 之间时恶化的概率为 24%；结节直径在 11mm~30mm 之间时恶化的概率为 33%；结节直径在 30mm 以上者，发展成为恶性结节的概率为 80%。因此，为了提高检查效率，本文将良性结节定义为肺结节直径在 30mm 以下，将恶性结节定义为结节直径在 30mm 或以上。

目前肺结节的诊断鉴别主要依靠 X 射线检查：肺结节病的首要发现是异常的胸部 X 射线表现，通过断层扫描对病人的肺部进行扫描，来获得病人的肺部影像，然后据此判断是否有结节，以及结节的类型。如图 2-1 所示，为某肺结节病人的 CT 图像。

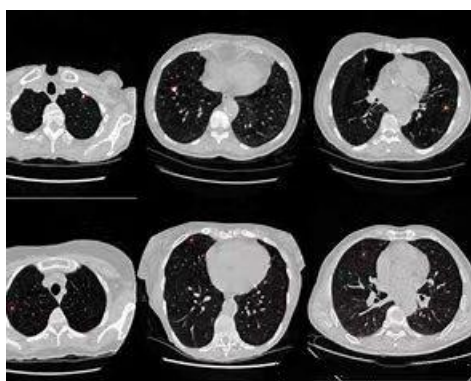


图 2-1 肺结节病人 CT 图像

早期的肺结节检测过程很简单，首先是利用 X 射线获得患者的肺部 CT 图像，然后会有专业的医生观察影像中是否存在结节，如存在，判断结节类型。但是，这完全依赖于医生的专业知识水平以及经验，无法保证诊断结果的准确率，而且每名患者的肺部 CT 有高达几十上百张肺部不同角度的影像，医生要全部浏览过后才能给出判定结果。而实际生活中，影像科的医生一天往往会接收许多患者的 CT 影像，如此大的阅片量给医生带来了极大的负担。而长时间浏览相似的物品时，势必会出现视觉疲劳。因此，难免会出现误诊漏诊现象。为了降低误诊率，引进了 CAD 系统进行辅助。目前，通过计算机进行断层扫描，然后利用 CAD 系统寻找具有特定的密度（肺结节）或者特定的物理特征，在肺结节检测领域中较为活跃。

2.2 YOLO 算法介绍

2.2.1 神经网络概述

深度学习指能够像人的大脑一样学习大量数据并进行工作的智能技术。它是机器学习的一个子集，由神经网络层驱动，按照人类大脑对工作的处理方式松散建模而成，接着通过训练大量数据配置神经网络的神经，那么训练后的模型可以处理新的数据，此过程不需要人为参与。

20 世纪 40 年代初，麦克洛克一皮茨模型是世界上第一个人工神经元模型，到 20 世纪 50 年代末，弗兰克·罗森布莱特提出的感知机模型使人类迈入了探索人工智能的时代，但是模型中遗留的异或问题却困扰了其近 20 年。一直到八十年代中叶，杰弗里·辛顿发明了两级或更高层次的人工神经网络，这一问题得到了很好的解决，并在辨识与预报方面取得了很大的进展，在当时引起了很大的轰动。20 世纪初深度神经网络被辛顿提出，将 ANN 的研究越发往深度探索。现如今，人工智能与我们的生活息息相关，随处可见机器学习的影子，被广泛应用于医疗卫生、网络安全、社交媒体、数字助手等方面，并取得了显著成就。

目标检测是利用计算机来识别目标，从而提高准确率，解放双眼，传统的目标检测大体可以分为三个步骤，分别是选择候选区域、提取相关特征向量以及划分目标区域。但是也存在很大的弊端，比如选择候选区域时不能准确地识别出结节所在区域，大多为手动标注且噪声太大，导致运算速度慢，假阳性率比较高；在提取相关特征向量时需要人工设定特征描述符，一些具有代表性的信息捕捉不到。而应用了深度学习的目标检测算法则很好的解决了以上问题。

2.2.2 神经元模型

20 世纪 40 年代，McCulloch and Pitts 将神经元抽象为如图所示的简单模型，这就是沿用至今的“M-P 神经元模型”。神经元接收到来自 n 个其他神经元传递过来的输入信号，这些信号通过带权重的连接进行传递，神经元接收到的总输入值将与神经元的阈值进行比较，再经过“激活函数”处理以产生神经元的输出。将很多这种神经元按某种层级联系在一起，就形成了一个神经网络，如图 2-2 所示。

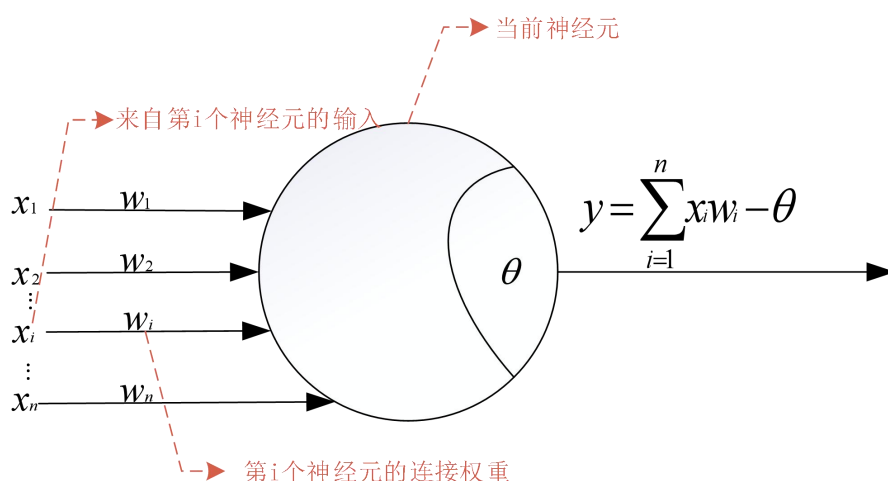


图 2-2 神经元结构图

2.2.3 卷积神经网络

典型的卷积神经网络^[14]由三部分构成：卷积层、池化层和全连接层。卷积层负责提供图像中的局部特征；池化层用来大幅度降低参数量级；全连接层类似传统神经网络的部分用来输出想要的结果。生物神经网络使大脑能够以复杂的方式处理大量信息。在生物神经网络中神经元之间彼此相连，当有刺激传来时，神经元会发生“兴奋”反应，即向传导方向连接的神经元释放乙酰胆碱，由于不同电性的离子的流动，此时神经元的内部电位发生变化，当电位的差值增加到一定数值之后，此神经元被激活，开始向连接的神经元释放化学物质。人工神经网络就是模拟生物神经系统所做出的交互反应，被广泛应用于互联网络。神经网络包括输入层、输出层、隐含层。输入层接收特征向量 x ；输出层产出最终的预测 h ；隐含层介于输入层与输出层之间。人工神经网络具有很强的学习能力。在获取一个训练集之后，通过学习观察各部分的特点来提取特征，各个特征之间用不同的网络结点连接，通过训练网络权重改变连接的强度，直到最终得出正确的结果。

2.2.4 激活函数

激活函数的作用是去线性化，最终决定了要发射给下一个神经元的内容，神经网络结点的计算就是加权求和，在加上偏置项，假设模型的输出 y 和输入 x_i 满足以下关系，数学函数可以表达为：

$$y = \sum_i \omega_i x_i + b \quad (\text{公式 2-1})$$

其中 $\omega_i, b \in R$

所设想的激活函数是如图 2-3 所示的 Step Function 函数，把输入值经过映射得到输出值“0”或“1”，显而易见，“1”表示神经元兴奋，“0”表示神经元受到抑制。但是阶跃函数在它的不连续、不光滑的性质方面差强人意，所以，常用的激活函数是 Sigmoid 函数，如图 2-3 所示。

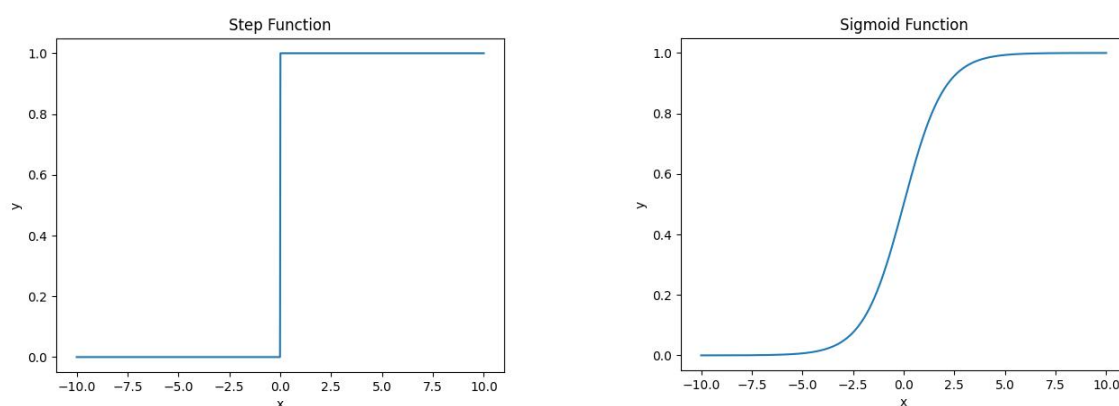


图 2-3 激活函数

典型的 Sigmoid 函数将在较大范围内来回变化的输入值缩小挤压到(0, 1)输出值范围内，因此它有时也被成为“挤压函数”。但是在使用神经网络训练时，这些激活函数经常会碰到梯度消失的问题，影响输出结果且 Sigmoid 函数执行指数运算，计算机运算速度较低，输出结果不是以 0 为中心的，这会降低权重更新的效率。因此研究出了能克服梯度消失等问题的无界非多项式激活函数--ReLU 函数，数学表达式为：

(公式 2-2)

$$\text{ReLU}(x) = \max(0, x)$$

函数图像如图 2-4 所示：

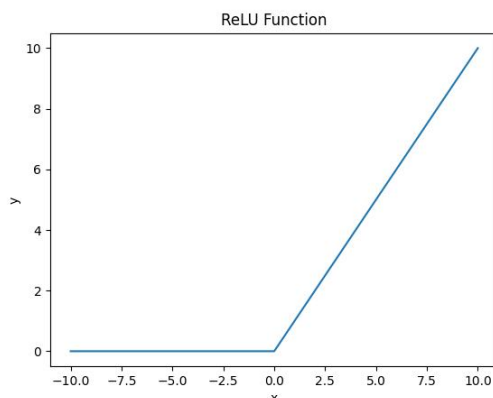


图 2-4 ReLU 函数

从函数图像可以看出，在进行训练的时候，对一些区域的数据比较敏感，而一些区域的数据却不敏感。在反向传播过程中，如果输入负数，输出的结果总是 0。为了保持负半轴的信息，又在 ReLU 函数的基础上做出了改进，从而生成了很多的衍生函数，其中具有代表性的是 Leaky ReLU 函数。

其数学表达式为：

$$\text{Leaky ReLU}(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (\text{公式 2-3})$$

其中 α 的值通常为 0.01。Leaky ReLU 函数的范围是负无穷到正无穷，该函数输出对负值输入有很小的坡度，能减少静默神经元的出现，有助于扩大 ReLU 函数的范围。函数图像如图 2-5 所示。

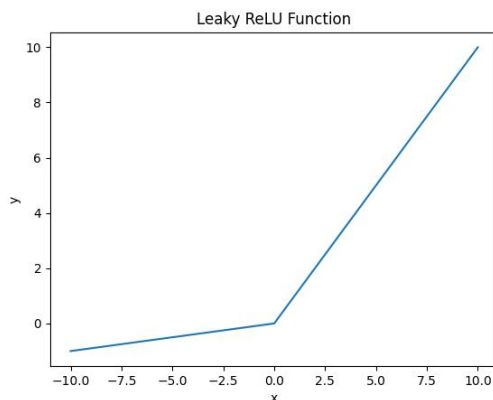


图 2-5 Leaky ReLU 函数图

ELU 函数的引入提供了一种新的思路。相比于 ReLU，ELU 为负值，因此该激励函数的均值趋近于 0。平均激活率趋近 0，学习速度会加快。其数学表达式为：

$$\text{ELU}(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \quad (\text{公式 2-4})$$

与 ReLU 函数相比, ELU 函数很好的缓解了 ReLU 函数存在的问题, 且以 0 为输出结果的中心。接下来开始处理偏置偏移的影响, ELU 函数对于偏置偏移的影处理, 主要迫使正常梯度更加接近单位自然梯度, 正常梯度与单位自然梯度的接近会使均值向 0 加速学习, ELU 函数在较小的输入下是有一定的输出的, 这部分输出会饱和至负值, 具有一定的抗干扰能力, 从而减少向前传播的突发变异和干扰信息。其函数图像如图 2-6 所示。

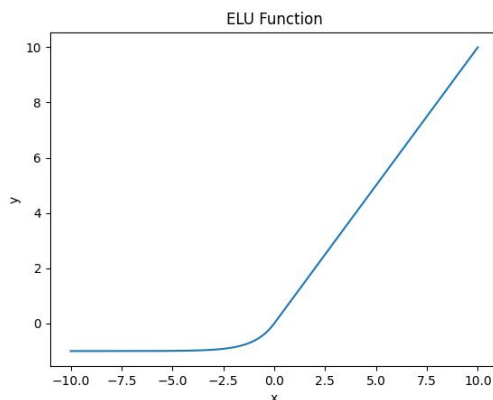


图 2-6 ELU 函数图

但是由于 ELU 函数在负输入时涉及到非线性, 因此计算速度会有所降低。

Softmax 激活函数在数学领域被称为归一化指数函数, 用以处理分类问题, 在多元分类问题中, 超过两个类标签则需要添加类成员关系。Softmax 函数可以将长度为 K 的任意实数向量压缩在 $(0, 1)$ 范围内, 且这些值的累加和为 1。其数学表达式为:

$$\text{Soft max}(x) = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad (\text{公式 2-5})$$

Softmax 函数在零点不可微, 一般作为神经网络的最后一层, 接收上一层网络的输入, Softmax 函数在负输入的梯度为 0, 这意味着权重不会在反方向传播期间产生更新, 因此会产生永不激活的死亡神经元。其函数图像如图 2-7 所示:

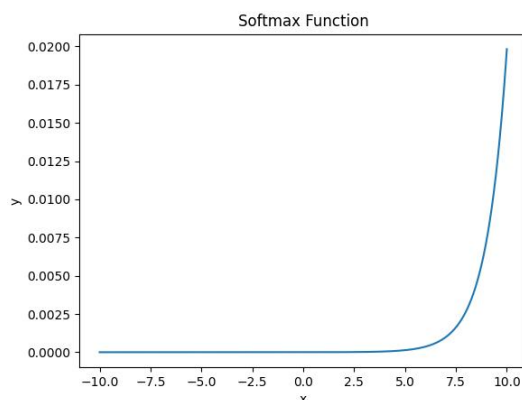


图 2-7 Softmax 函数图

单阶段目标检测网络是首先由 Eigen 等 9 人提出的，以 DCNN 为分类器构建的，该网络的原理是越过建议生成阶段，直接作用于全部图片，对其进行特征提取并分类，且检测速度比 R-CNN 系列的网络更快。因此越来越多的人研究 one-stage。

1.输入端

608*608

2.Backbone

Focus CBL CSP1_1 CBL CSP1_3 CBL CSP1_3 CBL SPP CSP2_1 CBL 上采样

3.Neck

Concat CSP2_1 CBL 上采样 Concat CSP2_1 CBL Concat CSP2_1 CBL Concat CSP2_1 CBL

4.Prediction

76*76*255

38*38*255

19*19*255

CBL = Conv BN LeakyRelu

Res unit = CBL CBL add

CSP1_X = CBL Res unit Conv X个残差组件 Concat BN LeakyRelu CBL

CSP2_X = CBL CBL Conv 2*X个 Concat BN LeakyRelu CBL

Focus = slice slice slice slice Concat CBL

SPP = CBL Maxpool Maxpool Maxpool Maxpool Concat CBL

结构图中基本组件：

CBL: 由 Conv+Bn+Leaky_relu 激活函数三者组成。

Res unit: 借鉴 Resnet 网络中的残差结构，让网络可以构建的更深。

CSPX: 借鉴 CSPNet 网络结构, 由卷积层和 X 个 Res unit 模块和 Concat 组成。

SPP: 采用 $1*1, 5*5, 9*9, 13*13$ 的最大池化的方式, 进行多尺度融合。

Concat: 张量拼接, 会扩充两个张量的维度, 例如 $26*26*256$ 和 $26*26*512$ 两个张量拼接, 结果是 $26*26*768$ 。

add: 张量相加, 张量直接相加, 不会扩充维度, 例如 $104*104*128$ 和 $104*104*128$ 两个张量相加, 结果还是 $104*104*128$ 。YOLO5 的性能测试图如下图 3-2 所示:

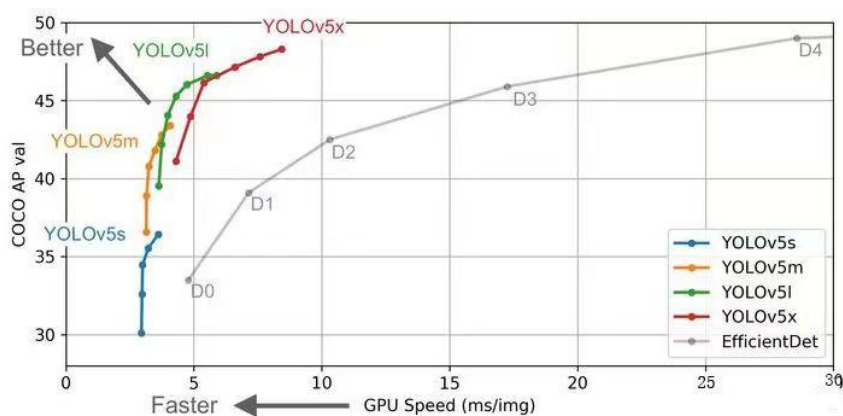


图 3-2 YOLOv5 性能测试图

YOLOv5s 的网络最小、速度最快、AP 精度也最低, 其它三种网络在 YOLOv5 的基础上, 随着网络的深度和宽度的增加, AP 精度也在不断的提升, 但对速度的要求也在不断提高。其特征提取过程如图 3-3 所示。

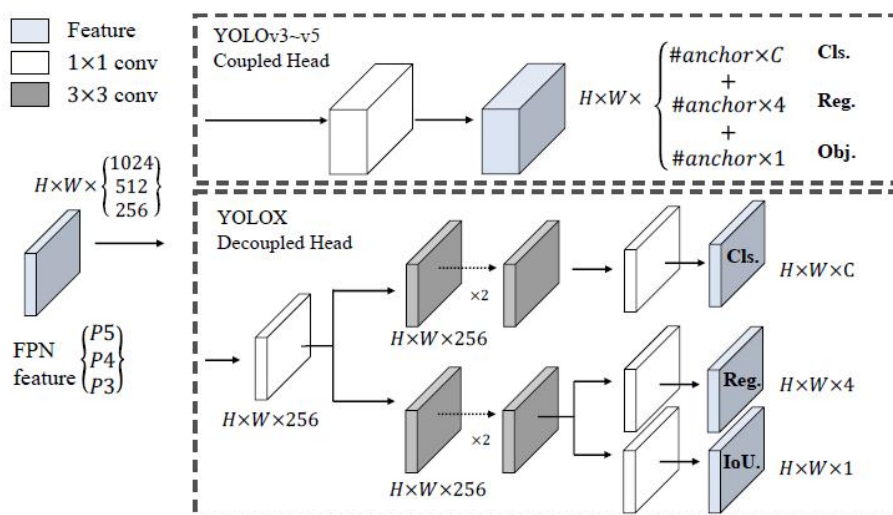


图 3-3 特征提取过程

YOLOv5 模型在对小目标检测方面有较为成熟的检测能力, 肺结节目标较小, 因此本实验采用 YOLOv5 作为基础模型实现对肺结节的检测。YOLO 算法由特征提取网络(Backbone)、强化特征层(Neck)、预测头(Head)组成。在特征提取网络中, 使用多个卷积层和池化层的交叉组合, 提高检测图像的通道, 并且降低输入图像的大小, 提取出来最精简的高纬度特征, 由上采样扩充数据集的大小, 通过下采样精简图像信息, 降低运算的复杂度。经过提取特征层和强化提取, 完成了对不同大小图像的特征提取。由预测头预测物体的种类和位置。

3.3 YOLOv5 模型检测流程

YOLOv5 作为一种单阶段检测算法，对图像特征提取和检测一次完成^[15]，因此检测速度较快，该模型将输入的图像进行分割，将图像分割成大小相同的网格，每个网格只负责中心位置落在该网格的物体。每个网格周围会产生不同大小的边界框，模型会根据边界框和实际标注框的交并比(IoU)来选择最优边界框预测网格中物体的类别位置等相关参数。最后采用非极大抑制(NMS)的做法，选取一定的阈值(confidence)大小，挑选出最优边界框作为本次预测的输出。具体检测流程如图 3-4 所示

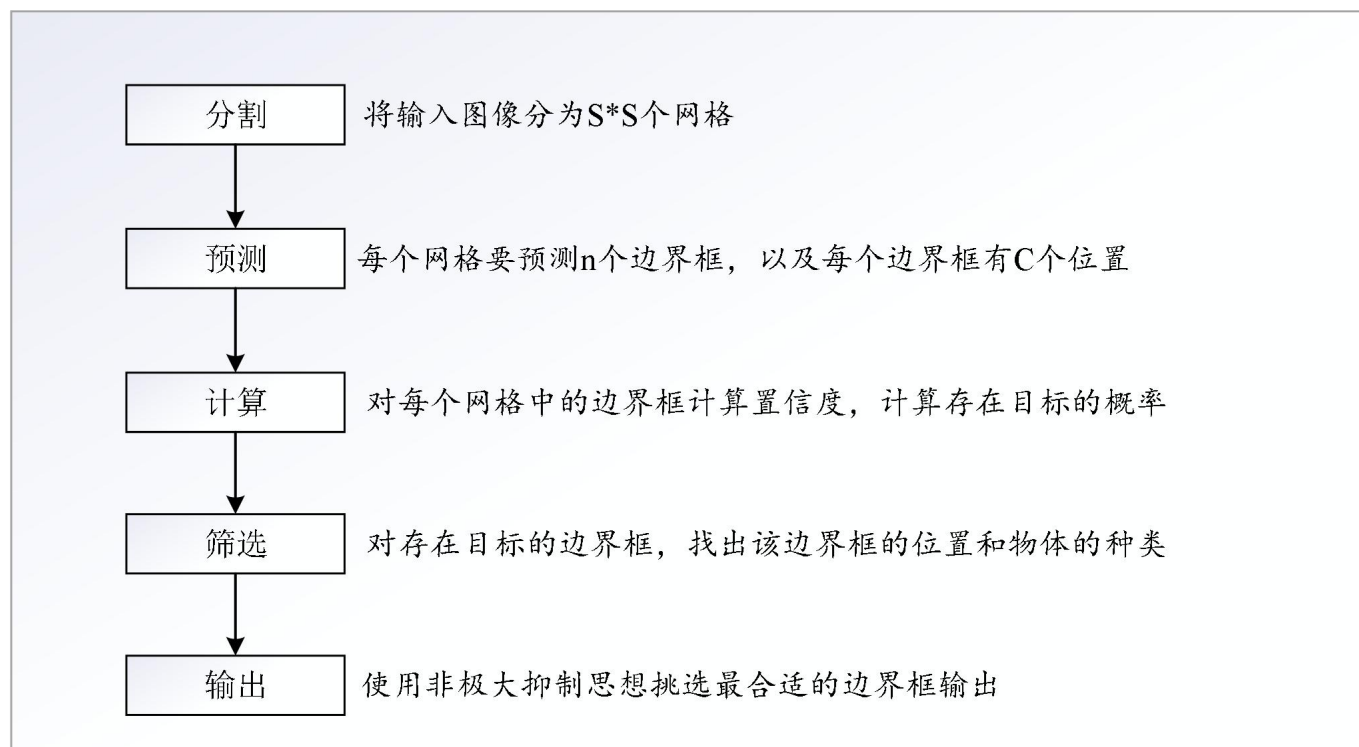


图 3-4 YOLO 检测流程

3.4 YOLOv5 模型优化策略

3.4.1 IoU 与非极大抑制策略

在网格周围有 B 个边界框，例如 YOLOv1 每个网格周围就有 2 个边界框。实际标签和候选边界框的交集比并集的值(IoU)可以很好的解决这个问题。计算公式如下：

$$IoU = \frac{A \cup B}{A \cap B} \quad (\text{公式 3-1})$$

使用 IoU 可以挑选出于实际标签框的候选框来进行预测，如图 3-5 所示，使用非极大抑制操作即采用这种思想。在 YOLOv1 中一般采用挑选最大的 IoU 值最大的候选框，这种方法会让算法快速定位到目标检测框，然而这种方法带来的问题是一个网格中只能有一个预测框，模型规定将图像划分为 7*7 的网格，这样最多可以检测 49 个目标，YOLOv5 避免了上述问题，采用更密集的网格划分，同时给每个网格匹配更多的候选框。这样就大大提高了模型对小目标的检测能力。

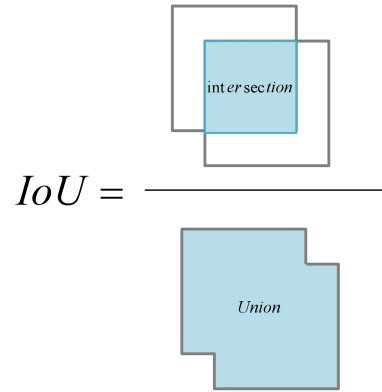


图 3-5 IoU 示意图

非极大抑制(NMS)操作是将满足置信度预测的候选框筛选出来,作为最终的输出结果。在网格周围的大量候选框中,选出当前最优结果。这里出现的一个缺点就是有可能该物体有多个匹配框,而算法模型仅仅暴力的选取了当前最优,而忽略了整体最优。YOLOv5 算法中采用正样本分配策略解决了以上问题。

3.4.2 特征提取网络与预测头

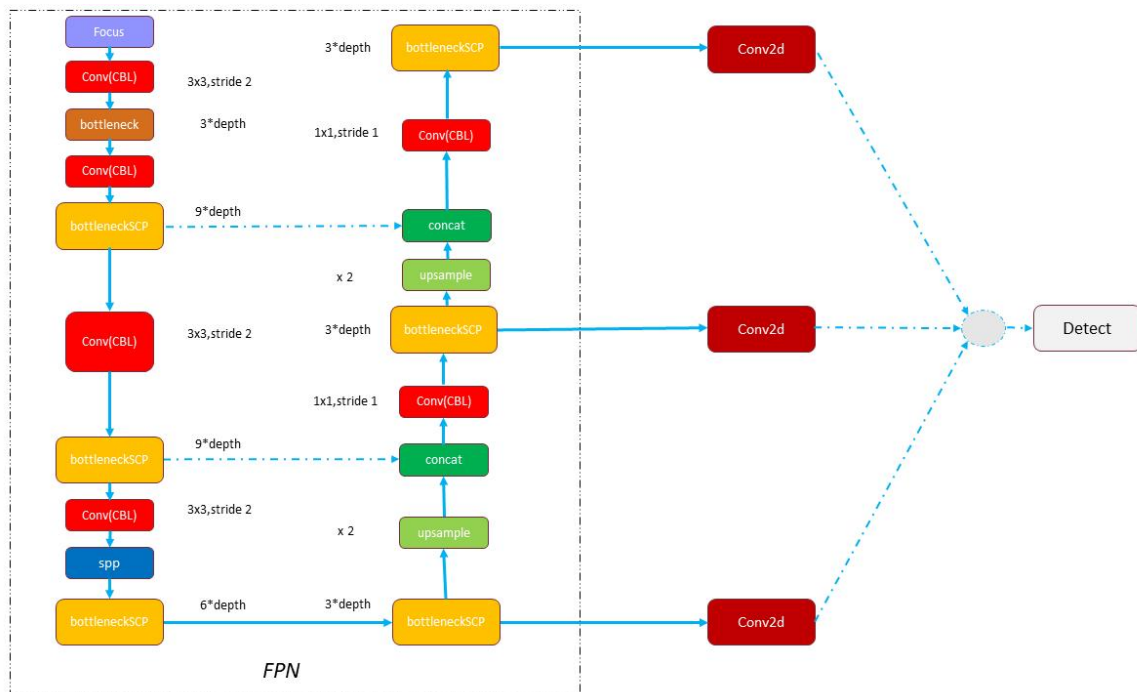


图 3-6 特征提取网络

YOLOv5 的特征提取操作采用了 New CSP-Darknet53 网络,如图 3-6 所示,在第一阶段放弃了池化操作,也就是没有进行下采样和全连接操作,避免了在前期特征提取中丢失太多的特征细节,将 stride 的值加倍后,得到的特征图减半。在 C3 操作中,把输出的结果卷积后与原来的输出拼接,经过多次的卷积和 C3 提取特征信息。在前特征提取完毕后使用金字塔池化(SPPF),将不同大小的特征图融合,得到图像各个尺寸上的特征信息。在特征提取的第二阶段,采用向上采样和第一阶段的输出特征信息拼接,再反复多次进行卷积和 C3 操作,得到不同尺寸、不同维度的特征图。

YOLOv5 使用 SPPF 构建 neck 模块, SPPF 可以对不同大小的特征信息加工, 使用多尺度的池化操作, 将它们压缩成大小相同的特征信息, 将这些特征信息堆叠输出。堆叠后的特征信息从下往上, 特征图的大小依次递减, 所以又称为金字塔池化。Neck 模块整合了不同层次的特征信息, 提高模型的提取能力。此外, YOLOv5 引入了 FPN(Feature Pyramid Network)模块, 增强的特征图的信息展现表现能力。在金字塔的不同层级上动态的融合特征图, 模型同时检测不同尺度的目标, 使模型检测速度和检测精度有很大的提高。

3.4.3 正样本分配策略

YOLO 的初代版本中, 由于网格数量较少, 并且采用非极大抑制策略使网络检测小目标的能力有限。YOLOv5 采用正样本分配策略, 可以将该网格的中心进行一定的偏移, 原来的模型只有一个网格预测该物体, 将次中心偏移后可以有更多的网格来预测该物体, 值得注意的是, 我们只能设定 2 个方向偏移, 不然会造成下方的网格和上方网格重复的情况, 没有提升精度还增加了运算复杂度。这种偏移的策略使更多的网格可以参与筛选, 可以有效的提升模型的召回率, 如图 3-7 所示。

经过上述操作, 非极大抑制算法需要筛选的网格增多, YOLOv5 抛弃了只用 IoU 作为筛选指标这种策略, 采取了多指标加权得分作为初筛, 再采用 `ground_truth` 来检测与预测框的匹配程度, 挑选出更优秀的边界框。值得一提的是, YOLOv4 模型中已经加入了对候选框的缩放, 设置合适的比例来匹配标注框。初筛的指标包括 `anchor` 的长宽比例、IoU 的大小、计算的类别是否正确等。给以上指标设置不同的权重, 计算出每个候选框的得分信息。

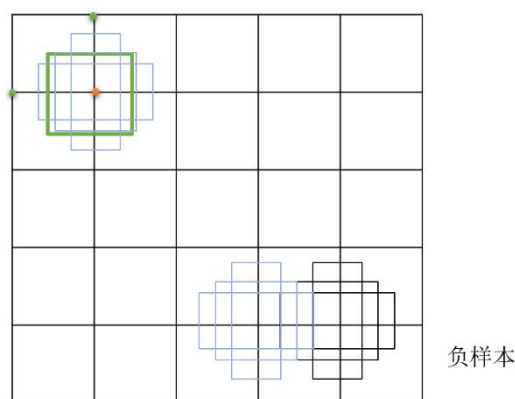


图 3-7 正样本分配策略

第四章 仿真实验

4.1 数据获取与构建

4.1.1 数据集选取

本研究利用国际公开 LIDC-IDRI 肺结节数据库，它包含了大量的 CT、X 光片、X 线等影像资料，并配以相应的病灶标记，是目前全球最大的、面向公众的肺结节图像数据库。这个图像数据库包含了来自 1,010 个患者的 1018 套全肺 CT 扫描图像（扫描层厚 0.6~5 mm, 512×512 像素），数据总容量为 124 GB。该数据库为每例 1 个文件夹，包括 100-300 幅全套 CT 影像，并配有一个记录文档，每一幅影像均有 4 名资深胸片医生做二级标记。在第一阶段，每一名医生都可以独立地进行诊断，并对病患位置进行标注。在这个过程中，会标注三类：1) ≥ 3 mm 的结节，2) < 3 mm 的结节，3) ≥ 3 mm 的非结节。（官网描述：“nodule ≥ 3 mm”，“nodule < 3 mm”，“non-nodule ≥ 3 mm”，详见 Summary）。接下来的第二个环节，就是每位医生都会将另外三个医生的标记重新检查一遍，并确定最终的诊断结果。这种两步式的注释能够尽量完全地注释所有的结果，同时又不会引起 forced consensus。数据集内容如表 4-1 所示。

表 4-1 数据集

collection statistics	updated 3/21/2012
数据大小	124G
图像类型	CT(computed tomography), 243,958 张 DX(digital radiography) cR(computed radiography)
图片数	244,527
患者数	1010
系列数(Number of Series)	1,018 CT 290 CR/DX
研究数(Number of studies)	1038

到现在为止，已经检测到了一千零十二个案例，每一个案例都有相应的结构：

- 1) LIDC-IDRI-XXXX / Study Instance UID / Series Instance UID / *.dcm, *.xml;
- 2) XXXX : 0000 至 1012;
- 3) Study Instance UID : 对应每一个案例的检查实例编号;
- 4) Series Instance UID : 对应于不同的检查序号;
- 5) *.dcm, *.xml : 分别对应于每一张 dcm 切片和 xml 标注文件。

4.1.2 数据集标注信息

模块处理的数据为 $\text{slicer} * \text{rows} * \text{cols}$ 尺寸的三维矩阵 D。比如：对于病例 LIDC-IDRI-0001，即为 13 512*512 的矩阵，总共有 133 张切片，每一张的尺寸是 512*512，按照一定的次序存入二进制文件中，

每个像素大小为 2 字节(对应 short 类型)，数据集标注如图 4-1 所示。

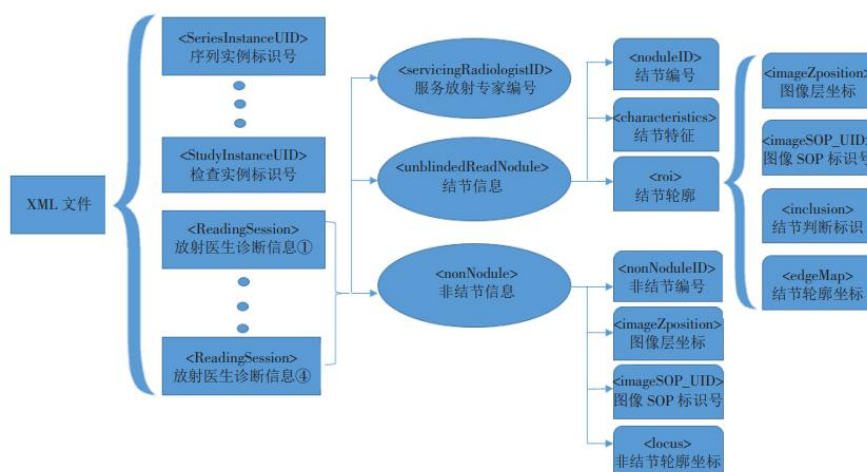


图 4-1 XML 标注信息说明

4.1.3 数据集预处理

在进行目标检测训练之前，需要对数据集进行预处理。首先是图片的格式问题，数据集中图片是 DICOM 格式，不能直接用于模型的训练，要将其转换为 jpg 格式。其次 LIDC-IDRI 数据集虽然是专门用于肺结节检测研究，但是数据集中仍然有大量数据是属于正常数据，即不含肺结节的数据，带有标注的数据占比很小，因此需要对数据集进行预处理。首先将数据集分为标注肺结节的和未标注肺结节的（即正常无结节），保留标注肺结节的数据集，将未标注数据集的数据随机抽取 2% 进行保留，其他数据舍弃。经训练，未标注肺结节的数据集仍占比较重，采用数据增强方式对数据进行扩展。为了加快模型收敛，并对数据集进行标准化处理。

4.2 YOLOv5 模型构建

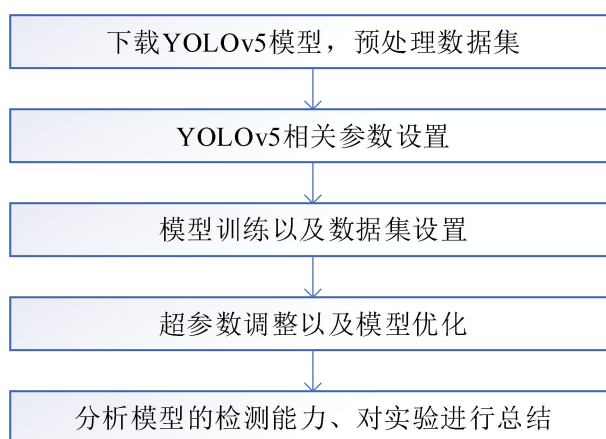


图 4-3 实验步骤

在数据集预处理结束后，要开始使用 YOLOv5 模型实现对肺结节的检测，实验检测流程如图 4-3 所示。从 github 网站上下载 YOLOv5 的模型以及代码，设置好环境的运行环境，然后对网络进行训练

和测试，具体实验步骤如下：

- 1) 在 mask 上标记所有医生标注的肺结节，将带有标记的 mask 切片都选出来，同时选择相同数量的没有标注的 mask 切片。
- 2) 使用 find contours 方法为每个带有 mask 的切片绘制边界框，处理图像，并突出图像中感兴趣的区域，并用 YOLOv5 的格式保存 512*512 图像和对应的标签
- 3) 对 YOLOv5 模型相关参数进行设置，例如：channel 的缩放系数(depth_multiple)、CSP 模块的层缩放系数(width_multiple)、标签种类数量(nc)、标签名称等，配置内容如表 4-2 所示。

表 4-2 YOLOv5 训练参数设置

参数名称	具体规格
depth_multiple	0.33
width_multiple	0.50
nc	1
names	nodule
learning_rate	0.001
最大 epoch	160

- 4) 将处理好的图像和标签送到模型中训练，模型使用梯度下降法(SGD)策略提升检测精度，在训练过程中使用预训练模型 YOLOv5.pt 来加快网络模型的收敛速度。
- 5) 对模型进行超参数优化，即使用遗传算法对超参数已经交叉变异已获得更好的结果，而不需要传统的网格搜索。这种优化思想使用在模型的 anchor 检测头的变异优化当中。
- 6) 对验证集中的图像传入当前的 best.pt 模型中得到验证集预测结果
- 7) 对上述预测结果传入损失函数计算损失，并通过反向传播计算损失的大小，使用优化器减少损失，然后开始下一轮训练
- 8) 网络训练完毕得到成熟的检测模型，将测试图像输入到模型，得到肺结节预测结果
- 9) 分析检测结果，计算平均检测精度(mAP)、召回率(recall)等。

4.3 实验结果分析

4.3.1 评价指标

评价指标分为精度评价指标和速度评价指标。

- 1) 精度评价指标：map（平均准确度均值，精度评价），准确率（Accuracy），混淆矩阵（Confusion Matrix），精确率（Precision），召回率（Recall），平均正确率（AP）等；
- 2) 速度评价指标：FPS（指每秒处理的图片数量或者处理每张图片所需时间）
- 3) TP（True Positive）指正确检测到目标；TN（True Negative）指正确检测到非目标；FP（False Positive）指误检测；FN（False Negative）指漏检测。

其他评价指标计算公式如下：

$$Precision = \frac{TP}{TP + FP} \quad (\text{公式 4-1})$$

表示预测结果为肺结节且预测结果正确的样本数与所有预测正确样本数的比值。

$$Recall = \frac{TP}{TP + FN} \quad (\text{公式 4-2})$$

表示预测结果为肺结节且预测结果正确的样本数与标注样本数的比值。

经过实验训练结果如表 4-3 所示。

表 4-3 训练结果

参数名称	具体数值
检测精度	0.72026
mAP@.5	0.66771
mAP@.5:.95	0.40883
训练集总损失	0.034104
训练集分类损失	0
验证集边界框损失	0.020337
Recall	0.62704

4.3.2 实验结果

- 1) 训练得到的混淆矩阵如图 4-4 所示，对正样本和负样本的分辨能力较优秀，background FP 为 0，也就是几乎全部分辨出了正确的背景。肺结节的检测正确率达到了 90%，只有 10%的肺结节被错误的检测成了背景，实验模型的检测能力较好，对肺结节和背景有较强的分辨能力。

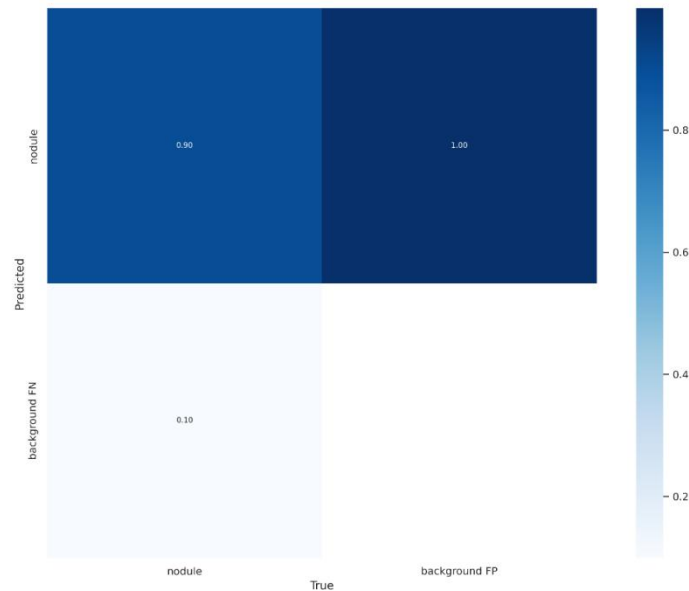
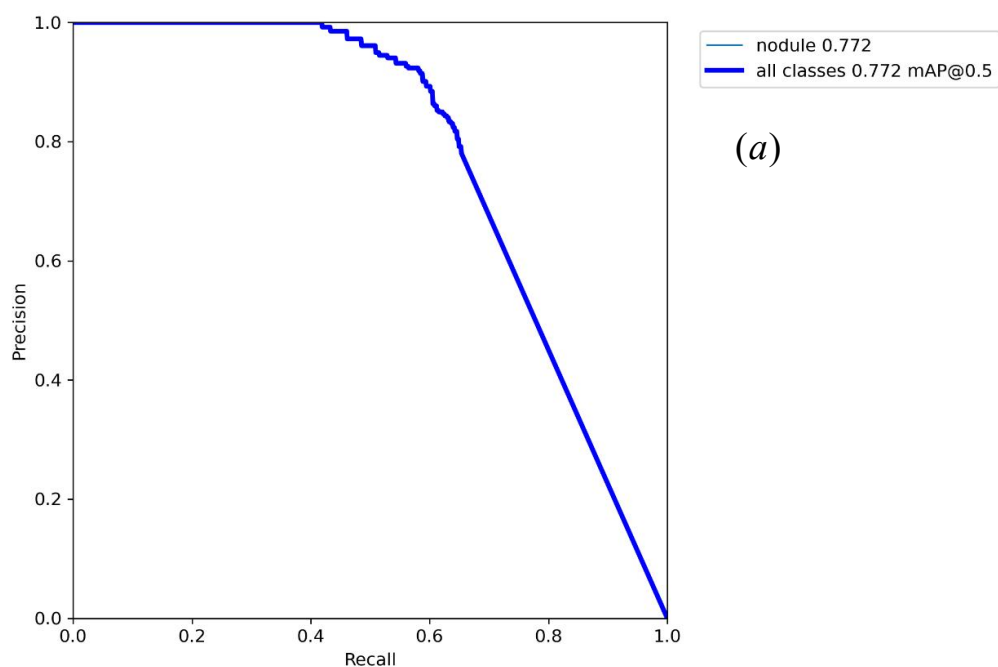


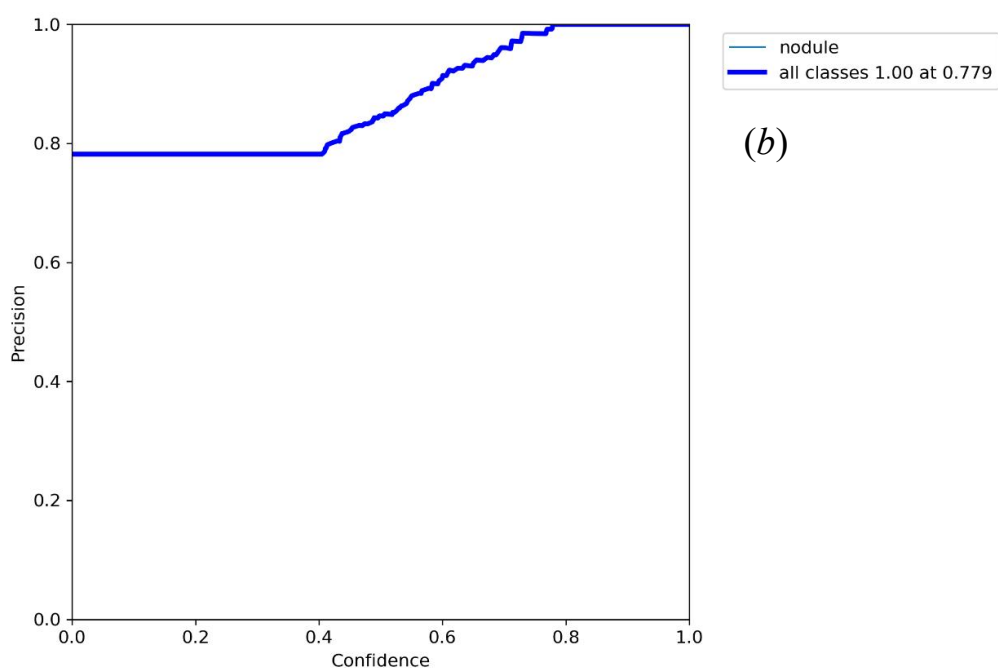
图 4-4 混淆矩阵

- 2) 检测精度图像如图 4-5 所示，两张图像分别反映了在不同的召回率和置信度的情况下，检测精度的变化曲线。从图 4-5(a)可以看出当召回率为 0.4 时检测精度趋近于 1，当召回率较高时，那检测的

准确度会大大下降。从图 4-5 (b)中可以看出当置信度设置 0.8 时，检测精度趋向 1，实验表明，在取合适的置信度的情况下，该模型的检测精度较高。



(a)



(b)

图 4-5 检测精度图像

- 3) 在不同的置信度的情况下，召回率会发生较大的变化，如图 4-6 所示，置信度设置的较高时，会很大程度上影响模型的检索能力，很多正样本就找不出来了；如果置信度设置的较低，召回率是会有所提升，但是检测的精度又受到了影响。因此，选择合适的置信度才可以让该模型的检测能力最大程度的发挥出来。

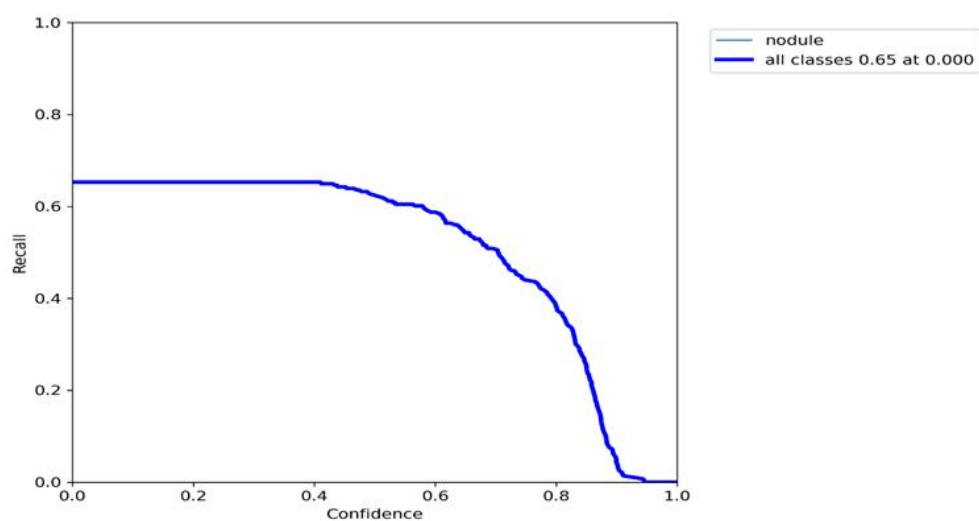


图 4-6 召回率曲线图

4.3.3 肺结节检测结果展示

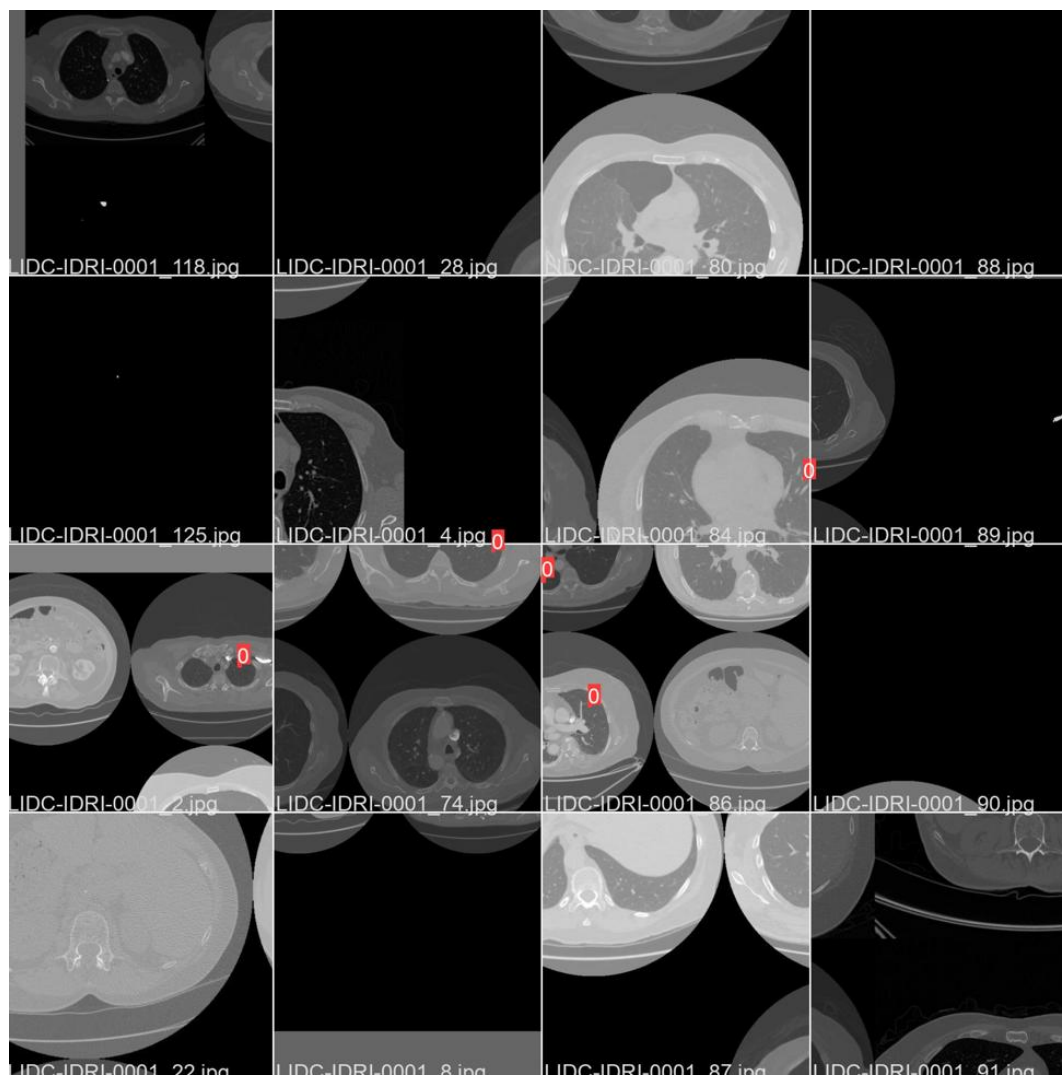


图 4-7 检测结果

选择训练最成熟的模型 `best.pt` 在测试集上进行预测，得到的预测结果如图 4-7 所示，有肺结节的地方在图中标记了出来。得益于公开数据集的数据量很大，训练样本充足，所以得到的检测结果较好，图片的肺结节都标注了出来。由于 GPU 资源不充足，本实验训练步数仅有 160 步，该模型的检测能力还可以进一步的提高，本实验中不再继续深究。

第五章 总结与展望

肺部 CT 图像需要耗费大量时间和精力，因其形态各异，使得研究自动肺结节检测系统变成一个具有挑战性的工作。在深度学习算法中，YOLO 系列算法不断改进，本文是基于 YOLOv5 的肺结节检测算法，为了提高检测精度，在其中加入了时空注意力机制，大大提高了检测性能。现对本文进行深刻的总结：

- (1) 查阅并学习肺结节检测基础，然后学习深度基础知识，掌握卷积神经网络以及目标检测网络等相关算法，注意力机制等算法基础，为本文肺结节检测网络的构建打下坚实基础。
- (2) 了解了当下流行的各种目标检测网络，最终选择检测速度较快的 YOLOv5 模型为本实验的基础架构，并依此建立肺结节检测网络。
- (3) 通过网络了解肺部 CT 相关知识，确定以 LIDC-IDRI 数据集为本文的数据源，同时进行数据筛选、数据预处理，使之便于提出的算法相匹配。然后经过多次实验最终建立起肺结节检测网络，并通过调整主干网络的训练参数筛选出最佳性能的网络，最终网络在减低对设备要求的条件下增加了检测精确度与查全率。

随着深度学习研究领域的热度高涨，医学图像的识别也获得突破性进展，本文提出基于 YOLOv5 的肺结节检测算法，大大提高了检测的精度，缓解了医学影像医生人力资源短缺的诸多问题，具有很高的学术价值，进而提高了应用价值。但是，该工作没有做到尽善尽美，总体还有很大的改良空间：

- (1) 本文构建了较为复杂的深度学习模型，网络训练耗费了大量时间，调参的时间代价较高，因此，该模型仍有较大的改进空间。
- (2) 本文的鲁棒性有待提高，在对结节的处理上，是以 3mm 为分界限，因为进行更细致的划分来提高检测准确率是进一步要考虑的问题。
- (3) 在整体流程方面，本文算法集中于帮助医学影像医师进行肺结节的筛查工作，这只是检查过程的一部分，后续还可以分析结节的良恶性，进一步了解患者病情。

参考文献

- [1] 潘子妍,邢素霞,逢键梁,申楠,王瑜,刘子骄,鞠子涵.基于多特征融合与 XGBoost 的肺结节检测[J].中国医学物理学杂志,2021,38(11):1371-1376.
- [2] 陶善礼. 基于 YOLOv3 网络的 CT 图像肺结节检测方法的研究与实现[D].重庆邮电大学,2022.
- [3] 李琬祎. 基于神经网络的 CT 图像肺结节检测研究[D].电子科技大学,2022.
- [4] 单基甸. 基于 CT 图像的肺结节检测算法研究[D].哈尔滨工业大学,2021.
- [5] 刘洋. 基于深度学习的肺结节检测系统研究与实现[D].重庆大学,2021.
- [6] 闫思桃. 基于深度学习的肺结节筛查系统的研究与实现[D].北京邮电大学,2021.
- [7] 霍恩泽. 基于深度学习的磨玻璃肺结节自动检测算法研究[D].东南大学,2021.
- [8] 徐时怀. 基于级联卷积神经网络的医学影像肺部小结节检测算法研究[D].中国计量大学,2021.
- [9] 张仕瑞. 肺结节检测及分类的深度学习方法研究[D].天津大学,2020.
- [10] 席孝倩,刘威.基于目标检测算法的肺结节辅助诊断系统[J].计算机与现代化,2020(11):1-7.
- [11] 潘成成. 基于区域卷积神经网络的医学图像识别研究[D].贵州大学,2020.
- [12] DONG Zhe,YAO Hong Yan,YU Shi Cheng,HUANG Fei,LIU Jian Jun,ZHAO Yan Lin,WANG Qi Qi.Changes in Notified Incidence of Pulmonary Tuberculosis in China, 2005 – 2020[J].Biomedical and Environmental Sciences,2023,36(02):117-126.
- [13] Hu Yaowen et al. Pulmonary nodules recognition based on parallel cross-convolution[J]. Multimedia Tools and Applications, 2022, 81(20) : 29137-29158.
- [14] Balci Mehmet Ali et al. A Series-Based Deep Learning Approach to Lung Nodule Image Classification[J]. Cancers, 2023, 15(3) : 843-843.
- [15] Sindhu Ramachandran S. et al. Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans[C]. Houston, Texas, United States, 2018.

致谢

转眼四年的大学生活即将划上句号，回首过往，耳边回荡着老师们的谆谆教导，在此论文完成之际，谨向四年来给予我帮助和关心的良师益友致以最真挚的谢意！

首先，我要特别感谢我的导师郝秀兰老师，郝老师在学术研究中沉稳扎实，刻苦钻研，心无旁骛，为我们树立了榜样。从课题目标的制定、所遇问题的解决和最终课题的核心搭建，郝老师步步悉心指导。其次感谢我的家人朋友们在背后的默默付出，让我勇于直面困难，怀有更加坚定的信心解决疑难，直至顺利毕业。

最后感谢评审和答辩组的老师们对我毕业论文的指导与评价，感谢你们的辛苦付出，祝桃李满天下。