摘 要

精准的医学影像辅助诊断技术通过对病灶区域的准确定位与特征建模，为临床医生提供直观、可靠的诊断依据，在提升影像判读效率、降低漏诊与误诊风险以及实现肺癌的早期发现与精准干预等方面具有重要意义。然而，囊腔型肺癌在 CT 影像中通常呈现出结构复杂、边界模糊及尺度变化显著等特点，给自动化分析方法带来了较大挑战。现有基于深度学习的医学影像分析模型在病灶区域建模能力与整体判别性能之间仍存在一定权衡，同时对关键区域信息的利用仍有待进一步加强。

针对上述问题，本文围绕囊腔型肺癌 CT 影像辅助诊断任务，研究并构建了基于 ROI 与注意力机制的医学影像分析方法，重点提升模型对关键病灶区域的感知能力与判别性能。本文的主要研究工作与结论如下：

（1）提出了一种基于 ROI 与注意力机制的囊腔型肺癌 CT 影像辅助诊断模型 LungROINet。该模型通过引入 ROI 区域建模策略，引导网络重点关注潜在病灶区域，并结合多尺度注意力机制对局部与全局特征进行增强建模，从而提升对囊腔型肺癌关键影像特征的表达能力。实验结果表明，LungROIaNet 在囊腔型肺癌辅助诊断任务中取得了较好的性能，在检测与良恶性判别精度方面相较于多种主流方法具有一定优势，验证了所提出方法在复杂医学影像场景下的有效性。

（2）围绕医学影像中标注成本高、数据获取受限等问题，进一步研究了弱监督条件下的医学影像分割方法，分析并引入了基于注意力机制的弱监督分割模型 MambaWS-SegNet。该模型通过弱标注信息构建有效监督信号，并结合注意力机制与区域建模策略，实现对病灶区域的精细分割。相关实验结果表明，该方法在多种医学影像分割数据集上均表现出良好的分割性能，体现了其在弱监督医学影像分析任务中的应用潜力。

（3）在上述研究基础上，本文对 ROI 建模与注意力机制在不同医学影像任务中的应用效果进行了综合分析，探讨了其在复杂病灶结构建模、小目标识别以及弱监督学习场景中的优势与局限性。研究结果表明，合理引入 ROI 与注意力机制有助于增强模型对关键区域的判别能力，为后续医学影像辅助诊断方法的研究提供了有益参考。

关键词：肺癌；ROI；注意力机制；医学影像

Abstract

Accurate medical image-based auxiliary diagnosis, through precise localization and modeling of lesion regions, provides clinicians with clear and intuitive diagnostic information, playing a critical role in improving image interpretation efficiency, reducing the risk of missed and misdiagnosed cases, and enabling early detection and intervention for lung cancer. However, cystic-cavity lung cancer in CT images is characterized by complex internal structures, blurred lesion boundaries, and significant scale variations, which pose substantial challenges to existing deep learning–based medical image analysis methods. In particular, current models often suffer from limited capability in effectively modeling region-specific features while maintaining robust diagnostic performance.To address these challenges, this thesis investigates medical image analysis methods based on region-of-interest (ROI) modeling and attention mechanisms for cystic-cavity lung cancer auxiliary diagnosis, and proposes two representative frameworks targeting different supervision settings. The main contributions and conclusions are summarized as follows:

(1) LungROINet: A ROI-Based Auxiliary Diagnosis Network for Cystic-Cavity Lung Cancer CT Images.To enhance lesion-focused feature representation, LungROINet introduces ROI-guided region modeling to emphasize potential lesion areas, combined with multi-scale attention mechanisms to jointly capture local details and global contextual information. Experimental results demonstrate that LungROINet achieves competitive performance in cystic-cavity lung cancer auxiliary diagnosis tasks, including lesion detection and benign-malignant classification, validating its effectiveness in handling complex lesion structures in CT images.

(2) MambaWS-SegNet: A Weakly Supervised Medical Image Segmentation Network Based on Mamba and Attention Mechanisms.Considering the high annotation cost and limited availability of pixel-level labels in medical imaging, this thesis further explores weakly supervised medical image segmentation. MambaWS-SegNet leverages weak annotations to construct effective supervision signals and integrates attention-based region modeling with Mamba-based feature learning to achieve fine-grained lesion segmentation. Experimental results on multiple medical image segmentation datasets indicate that the proposed framework exhibits favorable segmentation performance and strong robustness under weak supervision.

(3) Comprehensive Analysis of ROI Modeling and Attention Mechanisms in Medical Image Analysis.Based on the above studies, this work further analyzes the role of ROI modeling and attention mechanisms across different medical imaging tasks and supervision settings. The results demonstrate that appropriate integration of ROI-based region modeling and attention mechanisms can effectively enhance the discriminative representation of lesion regions, providing practical insights and methodological references for future research on medical image-based auxiliary diagnosis.

**Keywords:** Lung Cancer; ROI; Attention Mechanism; Medical Image

目 录

[摘 要 I](#_Toc220685397)

[Abstract II](#_Toc220685398)

[图和附表清单（非必要项） 5](#_Toc220685399)

[第1章 引 言 1](#_Toc220685400)

[1.1 研究背景与意义 1](#_Toc220685401)

[1.2 国内外研究现状 3](#_Toc220685402)

[1.2.1 基于深度学习的医学图像目标检测方法 3](#_Toc220685403)

[1.2.2 基于弱监督的医学图像分割方法 4](#_Toc220685404)

[1.2.3 现有研究中存在的问题 4](#_Toc220685405)

[1.3 研究内容与创新点 5](#_Toc220685406)

[1.3.1 研究内容 5](#_Toc220685407)

[1.3.2 创新点 6](#_Toc220685408)

[1.4 本文的组织结构 6](#_Toc220685409)

[第2章 相关理论和技术 8](#_Toc220685410)

[2.1 医学图像语义分析相关定义 8](#_Toc220685411)

[2.2 基于DETR的医学图像目标检测相关理论 9](#_Toc220685412)

[2.2.1 目标检测中的特征提取与多尺度表示方法 10](#_Toc220685413)

[2.2.2 注意力机制的引入 13](#_Toc220685414)

[2.2.3 实时端到端检测模型RT-DETR的基本思想 19](#_Toc220685415)

[2.3 基于 ROI 粗定位引导的医学图像半监督语义分割相关技术 23](#_Toc220685416)

[2.3.1 基于 U-Net 的医学图像语义分割模型基础 23](#_Toc220685417)

[2.3.2 医学图像半监督语义分割问题形式 25](#_Toc220685418)

[2.3.3 ROI 粗定位引导的关键区域建模 25](#_Toc220685419)

[2.3.4 教师—学生框架与 ROI 约束的半监督学习 27](#_Toc220685420)

[2.4 频域相关理论 28](#_Toc220685421)

[2.4.1 傅里叶变换与频谱表征 29](#_Toc220685422)

[2.4.2 DCT 的能量压缩性质与特征紧凑表示 29](#_Toc220685423)

[2.4.3 DWT 的多尺度频带分解与空间局部性 30](#_Toc220685424)

[2.5 本章小结 30](#_Toc220685425)

[第3章 基于 DETR 的医学图像目标检测模型 31](#_Toc220685426)

[3.1 囊腔型肺癌目标检测模型LungROINet 32](#_Toc220685427)

[3.1.1 部分卷积与多尺度注意力融合模块PMA 33](#_Toc220685428)

[3.1.2 基于多实例学习的Patch级分类与图像级聚合 35](#_Toc220685429)

[3.1.3 ROI 区域获取与高斯加权软池化 36](#_Toc220685430)

[参考文献 38](#_Toc220685431)

图和附表清单（非必要项）

（章标题“插图清单”“附表清单”或“插图和附表清单”用中文三号黑体，居中，单倍行距，段 前24磅，段后18磅。内容部分中文用小四号宋体，英文和数字用小四号Times New Roman字体，两端对齐，首行缩进2字符，行距20磅。）

（插图和附表清单为非必要项。论文中图表较多时，可分别列出“插图清单”和“附表清单”，“插图清单”在前，“附表清单”在后。插图较多而附表较少、或者插图较少而附表较多、或者二者均较少时，可将插图和附表合编为“插图和附表清单”。清单中应有正文对应的编号、题目和页码。插图和附表清单另起一页，置于目录之后。）

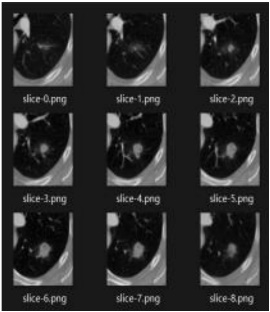
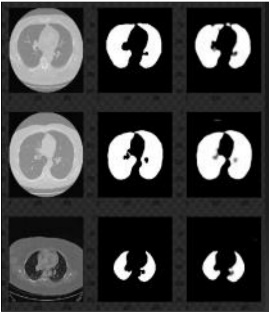
（示例）

1. 引 言

## 研究背景与意义

肺癌是全球范围内发病率和死亡率均居前列的恶性肿瘤之一，其早期发现与精准诊断对提高患者生存率具有重要的临床意义 [1]。随着医学影像技术的不断发展，计算机断层扫描（Computed Tomography, CT）因其成像速度快、空间分辨率高等优势，已被广泛应用于肺部疾病的筛查与诊断过程 [2]。在临床实践中，放射科医生通常需要结合 CT 影像中病灶的形态、位置及内部结构等信息，对肺部病变进行综合判断。然而，受限于影像数据规模庞大、病灶类型复杂以及医生经验差异等因素，人工阅片过程往往耗时耗力，且存在一定的主观性和不确定性 [3]。因此，借助人工智能技术对医学影像进行自动分析，为临床诊断提供可靠的辅助支持，已成为当前医学影像领域的重要研究方向。

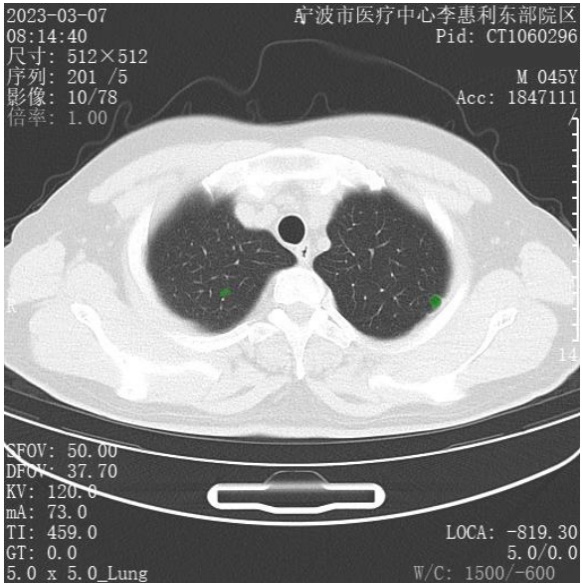
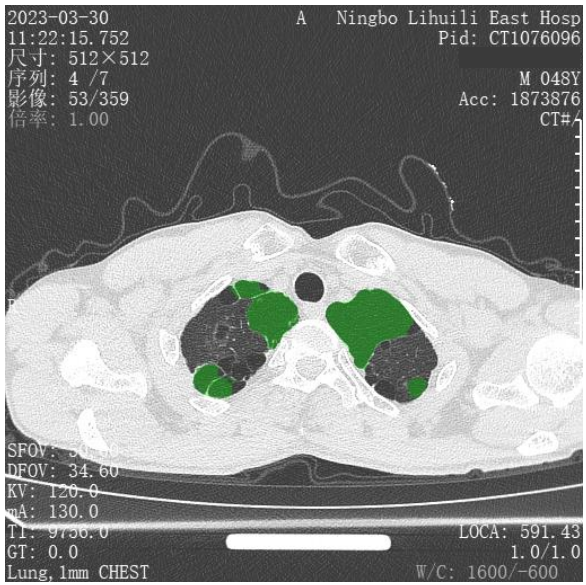
近年来，深度学习方法在医学影像分析任务中取得了显著进展，特别是在病灶检测、分割及分类等方面展现出优于传统方法的性能 [4]。医学图像语义分析作为医学影像理解的关键技术之一，旨在通过对影像中不同语义区域进行建模，为后续诊断、治疗规划及疗效评估提供依据。根据建模粒度和任务目标的不同，医学图像语义分析通常可分为像素级分割与区域级检测两类。像素级分割侧重于对病灶边界和内部结构进行精细刻画，适用于对空间精度要求较高的应用场景；区域级检测则通过定位关键病灶区域，在保留整体语义信息的同时降低计算复杂度，更适合小目标或多病灶场景。图 1.1 给出了医学图像分割与检测任务的典型示例，其中图 1.1(a) 和图 1.1(b) 展示了像素级分割在器官影像中的应用效果，能够对病灶区域进行精细勾画；图 1.1(c) 则展示了区域级检测与原图联合建模的示例，保留了完整影像上下文信息，更有利于小目标病灶的定位与分析。

（a）肺结节原图 （b）肺结节掩码 (c)肺部原图和掩码

图1.1　医学图像分割示例

然而，在实际临床应用中，肺部病灶在 CT 影像中往往呈现出形态多样、尺度差异显著以及边界模糊等特点，这对自动化分析方法提出了更高要求。尤其是囊腔型肺癌（Cystic-Cavity Lung Cancer），作为一种具有代表性的小目标病灶类型，其在 CT 影像中通常表现为壁厚不均的低密度空腔结构，病灶区域面积较小，内部组织形态复杂，且易与正常肺组织或其他良性病变产生混淆 [5]。如图 1.2 所示，囊腔型肺癌在不同患者和不同层面 CT 图像中的表现差异明显，部分病灶仅在局部切片中呈现出轻微结构异常，给病灶的准确定位和语义建模带来了较大挑战。当病灶尺寸较小时，影像噪声和背景组织对模型判断的干扰进一步加剧，导致诊断不确定性显著增加，对模型在局部结构建模和上下文语义理解方面提出了更高要求。

（a）小目标CT图 （b）大目标CT图

图1.2　分割效果经典实例

针对上述问题，区域兴趣（Region of Interest, ROI）建模思想为医学影像分析提供了一种有效的解决途径。通过引导模型重点关注潜在病灶区域，ROI 方法能够在降低背景干扰的同时增强对关键区域特征的表达能力，从而提升小目标病灶的检测与判别效果 [6]。与此同时，注意力机制在深度学习模型中的引入，使网络能够根据不同空间位置和特征通道的重要性自适应分配计算资源，在复杂结构建模和多尺度特征融合方面表现出良好优势。将 ROI 建模与注意力机制相结合，有助于进一步提升模型对囊腔型肺癌等复杂病灶的感知能力与判别性能。

此外，医学影像分析任务还普遍面临标注成本高、数据获取受限等现实问题。高质量的像素级标注通常需要专业影像医生耗费大量时间进行精细勾画，难以在大规模数据集中推广应用。因此，在有限标注条件下探索具备良好性能的弱监督医学影像分析方法，对于提升模型的实用性与可推广性具有重要意义。弱监督方法通过利用图像级标签、区域标注或不完全标注信息构建有效监督信号，为医学影像分割与辅助诊断提供了一种兼顾性能与成本的解决方案。

基于上述研究背景，本文以囊腔型肺癌 CT 影像辅助诊断为研究对象，围绕 ROI 建模与注意力机制在医学影像分析中的应用展开研究。一方面，针对囊腔型肺癌小目标、结构复杂的特点，设计基于 ROI 与注意力机制的辅助诊断模型，以提升对关键病灶区域的识别与判别能力；另一方面，结合弱监督学习思想，探索在有限标注条件下实现有效医学影像分割的方法。相关研究有助于推动医学影像辅助诊断技术在复杂临床场景中的应用，为提高肺癌诊断效率与准确性提供技术支持与理论参考。

## 国内外研究现状

近年来，随着医学影像数据的快速增长以及临床对精准诊断需求的不断提升，医学图像语义分析已成为计算机视觉与医学影像分析领域的重要研究方向之一[7]。医学图像语义分析的核心目标是对影像中的解剖结构、病灶区域或功能异常进行自动或半自动识别与建模，从而为疾病诊断、治疗规划及预后评估提供可靠依据[8]。在肺部 CT 影像分析场景中，由于病灶类型多样、尺度差异显著且存在边界模糊等问题，如何实现对关键病灶区域的准确定位与有效表征，仍是医学影像辅助诊断研究中的重要问题。

早期医学影像分析方法多依赖传统图像处理技术，如阈值分割、区域生长、活动轮廓模型以及基于形态学运算的边缘检测等[9]。此类方法通常依赖人工设定参数或先验规则，在特定条件下能够取得一定效果，但在复杂成像模态和多样化病理变化场景中，往往存在鲁棒性不足、泛化能力有限的问题[10]。随着深度学习技术的发展，数据驱动方法逐渐成为医学图像分析的主流研究方向，相关研究主要围绕医学图像目标检测与语义分割等任务展开。

### 基于深度学习的医学图像目标检测方法

医学图像目标检测旨在对影像中的异常区域进行自动定位，是医学影像辅助诊断中的重要环节。随着卷积神经网络的发展，大量基于深度学习的目标检测方法被引入医学影像分析领域，并在肺结节检测、病灶筛查等任务中取得了良好效果[11]。早期方法多采用两阶段检测框架，通过候选区域生成与区域分类相结合，实现对病灶区域的定位与判别。

近年来，单阶段目标检测方法凭借结构简洁、推理速度快等优势，逐渐在医学影像检测任务中得到应用。相关研究通过引入多尺度特征融合、特征金字塔结构以及上下文信息建模策略，提升模型对不同尺度病灶的检测能力[12]。然而，在肺部 CT 影像中，尤其是针对囊腔型肺癌等小目标病灶，目标检测仍面临一定挑战。一方面，小目标病灶在整幅影像中占比极低，容易被背景组织淹没；另一方面，病灶边界模糊且形态多变，增加了检测模型对关键区域准确定位的难度[13]。

### 基于弱监督的医学图像分割方法

医学图像分割旨在对病灶区域进行像素级精细刻画，是辅助诊断和定量分析的重要基础。然而，高质量像素级标注通常需要专业医生进行逐像素勾画，标注成本高、耗时长，限制了分割模型在大规模医学影像数据中的应用[15]。为缓解这一问题，弱监督医学图像分割方法逐渐受到研究者关注。

弱监督分割方法通常利用图像级标签、点标注、涂鸦标注或粗粒度区域标注等弱监督信息，构建近似监督信号以指导模型训练[16]。相关研究通过区域约束、伪标签生成以及一致性正则化等策略，在降低标注成本的同时尽可能保持分割精度。近年来，结合注意力机制的弱监督分割方法通过增强模型对关键区域的响应能力，在复杂病灶分割任务中展现出较好的性能[17]。

在肺部 CT 影像中，囊腔型肺癌等病灶通常呈现出结构复杂、边界模糊及内部异质性强等特点，使得弱监督分割任务更具挑战性。如何在有限标注条件下有效利用上下文信息与区域约束，实现对病灶区域的稳定分割，仍是当前弱监督医学图像分割研究的重要问题[18]。

### 现有研究中存在的问题

深度学习方法在医学图像语义分析与辅助诊断领域中取得了显著进展，在提高诊断效率、减轻医生工作负担等方面展现出良好的应用前景。然而，在复杂肺部 CT 影像场景下，尤其是针对囊腔型肺癌等具有小目标、结构复杂和边界模糊特征的病灶，现有研究方法仍存在一定不足，主要体现在以下几个方面：

（1）囊腔型肺癌在 CT 影像中通常表现为尺度较小、形态不规则的低密度结构，其区域在整幅影像中占比极低。现有深度学习模型在多次下采样过程中容易丢失关键局部信息，导致小目标病灶在复杂背景中难以被准确定位或完整建模。同时，为提升分割或检测精度而引入更深或更复杂的网络结构，往往会显著增加模型参数量和计算开销，影响模型在临床实际环境中的应用效率。

（2）多数医学图像分析方法采用全局特征学习策略，缺乏对潜在病灶区域的显式建模与有效约束，导致模型在训练过程中容易受到大量无关背景信息的干扰。尤其在肺部 CT 影像中，正常肺组织结构复杂、密度变化显著，若缺乏有效的区域引导机制，模型难以在特征学习过程中稳定聚焦于与诊断相关的关键区域，从而影响检测与分割结果的可靠性和一致性。

（3）受限于医学影像像素级标注成本高、获取难度大的现实条件，弱监督医学图像分割方法逐渐受到关注。然而，在囊腔型肺癌等边界不清晰、内部结构复杂的病灶场景中，弱监督信号往往难以准确反映真实病灶分布，伪标签噪声较大，易导致模型训练过程不稳定。此外，现有弱监督方法在充分利用上下文信息和区域一致性约束方面仍存在不足，其分割结果在不同样本和不同成像条件下的泛化能力有待进一步提升。

综上所述，如何在保证计算效率的前提下提升小目标病灶的检测与分割精度，如何增强模型对关键病灶区域的感知能力，以及如何在弱监督条件下提高分割结果的稳定性与鲁棒性，仍是当前医学图像辅助诊断研究中亟需解决的问题。

## 研究内容与创新点

### 研究内容

为解决当前医学影像辅助诊断中针对囊腔型肺癌 CT 影像所面临的小目标病灶难以准确定位、关键区域建模能力不足以及标注成本高等问题，本文围绕基于深度学习的医学图像目标检测与弱监督分割方法开展研究，主要研究内容如下：

针对问题（1），考虑到囊腔型肺癌在 CT 影像中具有多区域分布、小目标比例高以及结构复杂等特点，本文研究面向小目标病灶的医学图像目标检测方法。通过引入区域兴趣（ROI）建模思想，结合多尺度特征融合与注意力机制，引导模型重点关注潜在病灶区域，在增强关键区域特征表达能力的同时抑制无关背景干扰，从而提升模型对囊腔型肺癌病灶的定位与判别能力，构建面向囊腔型肺癌 CT 影像的辅助诊断模型。

针对问题（2），为进一步提升模型对关键病灶区域的感知能力，本文在目标检测框架中引入注意力机制与区域约束策略，从空间和特征层面对病灶区域进行强化建模。通过在特征提取与融合过程中对不同区域和尺度信息进行自适应加权，使模型能够更加稳定地聚焦于与诊断相关的局部区域，从而提高在复杂背景条件下的检测鲁棒性与结果一致性。

针对问题（3），考虑到医学影像像素级标注成本高、获取难度大的现实问题，本文进一步研究基于弱监督学习的医学图像分割方法。在仅依赖有限弱监督信息的条件下，探索利用区域约束与特征建模策略实现对囊腔型肺癌病灶区域的有效分割。通过引入适用于弱监督场景的特征学习与约束机制，提高模型在复杂病灶分割任务中的稳定性与泛化能力，为后续定量分析与辅助诊断提供支持。

### 创新点

本文围绕囊腔型肺癌 CT 影像辅助诊断任务，在医学图像目标检测与弱监督分割两个方面开展研究，主要创新点如下：

（1）提出了一种面向囊腔型肺癌 CT 影像的 ROI 引导医学图像目标检测网络 LungROINet。该方法在目标检测框架中引入 ROI 建模与注意力机制，通过多尺度特征融合增强模型对小目标病灶的感知能力，在复杂背景条件下实现对囊腔型肺癌病灶区域的稳定定位与判别，为囊腔型肺癌的辅助诊断提供了一种有效的建模思路。

（2）针对医学影像标注成本高的问题，提出了一种基于弱监督学习的医学图像分割网络 MambaWS-SegNet。该方法在弱监督条件下结合区域约束与特征建模策略，实现对囊腔型肺癌病灶区域的有效分割，在减少对像素级精细标注依赖的同时，提升了分割结果的稳定性与鲁棒性。

（3）从医学影像实际应用需求出发，系统地探索了 ROI 建模、注意力机制与弱监督学习在囊腔型肺癌 CT 影像分析中的协同作用，为复杂小目标病灶的检测与分割提供了统一的建模思路与实践参考，对医学影像辅助诊断方法的设计具有一定的借鉴意义。

## 本文的组织结构

本文围绕囊腔型肺癌 CT 影像的辅助诊断问题，结合医学图像目标检测与弱监督分割方法，对相关模型与技术展开系统研究。全文共分为五章，各章节的组织结构安排如下：

第一章为引言。首先介绍医学图像辅助诊断的研究背景与研究意义，分析囊腔型肺癌在 CT 影像中的成像特点及其在自动分析中的难点；随后综述国内外在医学图像目标检测与弱监督分割领域的研究现状，总结现有研究中存在的问题；在此基础上明确本文的研究内容与主要创新点，最后对全文的组织结构进行说明。

第二章为相关理论与关键技术。本章介绍医学图像分析中常用的深度学习基础理论与方法，包括卷积神经网络的基本结构、多尺度特征建模思想以及目标检测与图像分割的基本原理，为后续模型设计与方法研究提供理论基础。

第三章为基于 ROI 的囊腔型肺癌医学图像目标检测方法研究。针对囊腔型肺癌小目标比例高、结构复杂且易受背景干扰的问题，提出面向囊腔型肺癌 CT 影像的辅助诊断模型 LungROINet。本章重点介绍模型的整体框架与关键模块设计，包括区域兴趣建模、多尺度特征融合及注意力机制等，并通过对比实验与消融实验验证所提出方法在病灶定位与判别任务中的有效性。

第四章为基于弱监督学习的囊腔型肺癌医学图像分割方法研究。针对医学影像像素级标注成本高的问题，提出弱监督分割模型 MambaWS-SegNet，在有限监督信息条件下实现对囊腔型肺癌病灶区域的有效分割。本章详细介绍模型结构设计与弱监督策略，并通过实验分析模型在复杂病灶分割任务中的性能表现与稳定性。第五章为总结与展望。本章对全文的研究工作进行总结，概括本文提出方法的主要成果与不足，并结合医学影像辅助诊断领域的发展趋势，对未来的研究方向进行展望。

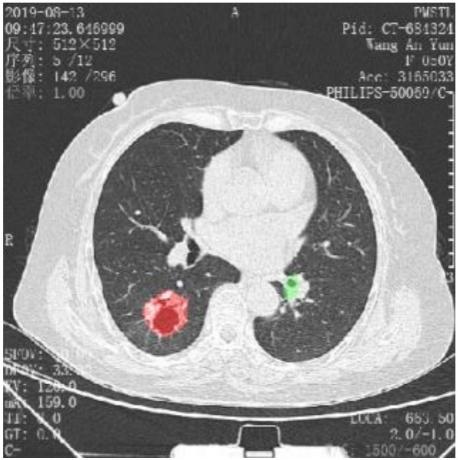
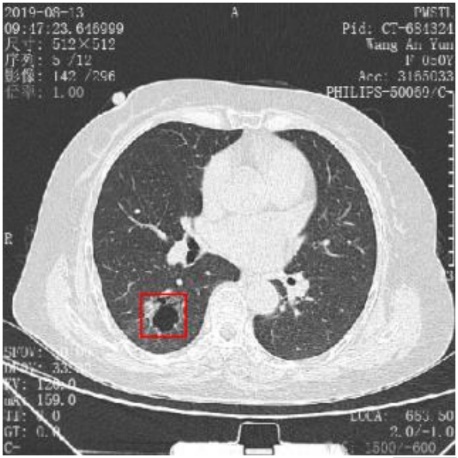
1. 相关理论和技术

本文围绕囊腔型肺癌 CT 影像的辅助诊断任务，重点研究医学图像目标检测与语义分割方法，并在此基础上探讨弱监督条件下的医学图像分析策略。为后续模型设计与方法研究提供理论支撑，本章将介绍医学图像语义分析的相关定义、基于深度学习的医学图像目标检测与语义分割基本理论，以及弱监督医学图像分割的相关概念。

## 医学图像语义分析相关定义

语义分割（Semantic Segmentation）旨在对医学图像中的每一个像素赋予语义类别标签，用于刻画不同组织结构或病灶区域在空间上的精确分布。该任务关注像素级别的语义一致性，输出与原图尺寸一致的分割掩膜。在肺部 CT 影像中，语义分割常用于将图像划分为背景、正常肺组织及病灶区域等，如图 2.1(a) 所示。

目标检测（Object Detection）则以实例级目标为研究对象，通过预测目标的边界框位置及其类别标签，实现对病灶位置和类别的联合建模。与语义分割相比，目标检测不涉及像素级的精细勾画，输出结果相对粗粒度，但在病灶快速定位、筛查与候选区域生成等任务中具有较高的应用价值，如图 2.1(b) 所示。

（a）语义分割 （b）目标检测

图2.1 语义分割和目标检测

从数学角度看，可将医学图像表示为定义在二维空间上的函数：

其中 表示图像中的像素位置， 表示该位置处的像素值（灰度值或 RGB 值）， 为图像通道数（对于 RGB 图像，）， 表示图像的定义域。

在语义分割任务中，其目标是学习一个从像素空间到语义标签空间的映射函数：

其中 表示预定义的语义类别集合，共包含 类。对于任意像素位置 ，模型预测其语义标签为 。该映射函数通常由深度神经网络进行近似学习，推理阶段输出与输入图像尺寸一致的语义分割结果。

对于目标检测任务，其目标是学习从图像到目标集合的映射：

其中 表示预测目标数量， 表示第 个目标的边界框（可用中心点与宽高或左上—右下坐标表示）， 为目标类别标签， 为对应的置信度分数。通过该形式化定义，目标检测能够为后续的病灶分析提供候选区域与定位信息，从而辅助临床筛查与诊断流程。

## 基于DETR的医学图像目标检测相关理论

医学图像目标检测在临床辅助诊断中主要用于对病灶区域进行快速定位与初步判别，常作为后续精细分割、定量测量与随访评估的前置步骤。与自然场景图像相比，医学影像尤其是 CT 切片通常存在组织结构复杂、灰度分布不均以及目标尺度偏小等特点，使得检测模型不仅需要具备较强的局部细节刻画能力，还需能够有效整合跨区域的上下文信息，以降低漏检与误检风险。近年来，随着 Transformer在计算机视觉领域的广泛应用，基于自注意力机制的端到端目标检测范式逐渐成为研究热点。其中，DETR将目标检测统一建模为集合预测问题，通过全局建模方式为复杂场景下的目标定位提供了新的研究思路；在此基础上，面向实际应用需求的实时 DETR系列模型进一步推动了检测框架在效率与可用性方面的发展。在医学图像目标检测任务中，现有方法通常面临以下挑战：

（1）小目标占比高，尺度差异明显。病灶区域往往呈现小尺度、弱纹理或边界模糊等特征，不同患者及不同病程阶段的病灶形态差异显著，导致检测模型在多尺度表征过程中易出现漏检现象。

（2）背景干扰强，目标与非目标相似度高。CT 影像中正常组织结构复杂，部分区域在灰度或形态上与病灶区域相近，模型容易将背景结构误判为异常目标，从而产生误检。

（3）上下文依赖显著，局部感受野受限。医学影像中的目标判别往往依赖于邻近结构及整体形态线索，仅依靠局部感受野提取的特征难以稳定地区分相似组织结构。

（4）精度与效率之间存在权衡。临床应用场景通常期望在较短时间内完成筛查与提示，而高精度检测模型往往伴随较高的计算开销，使其在实际部署中面临一定限制。

针对上述问题，研究者提出了一系列优化策略，以提升医学图像目标检测的性能与适用性，主要体现在以下几个方面：

（1）特征提取与多尺度表示。通过优化特征提取网络并引入多尺度特征表示方式，使模型能够同时利用高层语义信息与低层细节特征，从而增强对小目标及尺度变化显著病灶区域的感知能力。

（2）注意力机制的引入。利用注意力机制对特征图中不同空间位置或通道维度进行自适应加权，引导模型重点关注与病灶相关的关键区域信息，有效抑制复杂背景和无关结构的干扰，提升模型在复杂医学影像场景下的判别能力与鲁棒性。

（3）端到端目标检测范式。将目标检测建模为统一的端到端学习过程，直接从输入影像映射到目标预测结果，减少对人工先验规则与复杂中间处理步骤的依赖，有助于简化检测流程并提高模型整体的一致性与泛化能力。

因此，本文将围绕目标检测中的特征提取与多尺度表示方法、注意力机制在检测任务中的应用以及端到端目标检测范式三个方面，对基于 DETR 的医学图像目标检测相关理论进行系统介绍，为后续章节中结合具体模型的设计与改进奠定理论基础。

### 目标检测中的特征提取与多尺度表示方法

特征提取是目标检测任务中的前置关键环节，其目标是将原始医学影像转换为具有判别性的特征表示，为后续目标定位与类别判别提供信息基础。在医学 CT 影像中，病灶目标往往呈现尺度小、对比度低、形态不规则以及边界模糊等特点，且同时伴随复杂的正常解剖结构与成像噪声干扰，导致检测模型容易出现漏检与误检。特别是在囊腔型肺癌相关病灶中，囊壁、壁结节等关键结构通常占据像素比例较小，对特征表达的细粒度能力提出更高要求。因此，若特征提取阶段不能有效保留关键结构信息并形成足够的语义表达，后续检测模块即便采用更复杂的预测策略，也难以获得稳定可靠的检测结果。

在基于DETR的端到端检测范式中，特征提取模块不仅承担“提取局部纹理”的角色，还直接决定Transformer编码器输入表示的质量。DETR将目标检测建模为集合预测问题，并通过全局注意力建立跨区域依赖关系[19]，因此其输入特征既需要包含局部细节（如边缘、纹理与微小结构），也需要具备一定的全局语义表达能力以支撑后续全局建模。通常，DETR系列框架会先由骨干网络提取特征，再将特征组织为适合注意力建模的输入表示（如图2.2所示），该过程对特征的分辨率、语义层级与信息完整性均较为敏感。进一步地，面向实际部署需求的实时DETR系列模型在继承端到端思想的同时，更强调特征提取阶段的效率与可用性[20]。

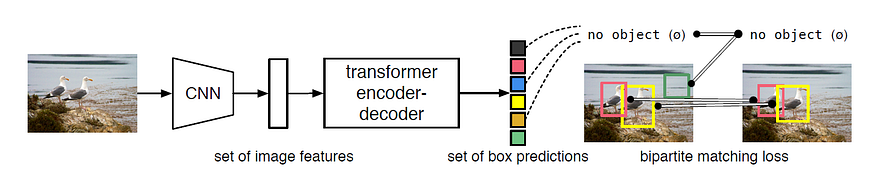


图2.2　DETR的目标检测特征提取与输入表示示意图

因此，面向 DETR/RT-DETR体系的特征提取设计通常更加关注两类能力：其一是骨干网络对医学影像关键结构的表征能力，其二是多尺度特征表示对小目标与尺度变化的适配能力。基于这一认识，现有研究一般从“轻量化特征提取”和“多尺度特征融合”两条主线对特征提取阶段进行改进，并逐步形成较为成熟的技术路线。

**（1）轻量级特征提取网络**

骨干网络（Backbone）负责将输入图像映射为高维特征表示，是检测模型性能与效率的主要决定因素之一。以 ResNet 为代表的经典卷积骨干通过残差连接有效缓解深层网络训练难题，使网络在保持较强表征能力的同时提升可训练性（如图2.3所示）[21]。对于医学影像检测而言，较强的骨干表征能力有助于提取病灶的细粒度纹理与边缘特征，从而提高对微小目标和低对比目标的敏感性。然而，临床辅助诊断场景通常还要求模型具备较好的部署友好性，因此近年来许多工作进一步强调在保证特征表达的前提下降低计算与参数开销，使模型能够在资源受限环境下稳定运行。轻量化思路一般体现在：优化网络结构设计以减少冗余计算、采用更高效的特征提取单元或通道配置，从而在精度与效率之间取得更平衡的折中。需要注意的是，在 DETR/RT-DETR 等端到端范式中，骨干输出特征将直接影响后续全局注意力建模的效果，因此轻量化设计往往需要兼顾语义信息的完整性与特征分辨率的合理性，避免因过度压缩导致小目标信息丢失。

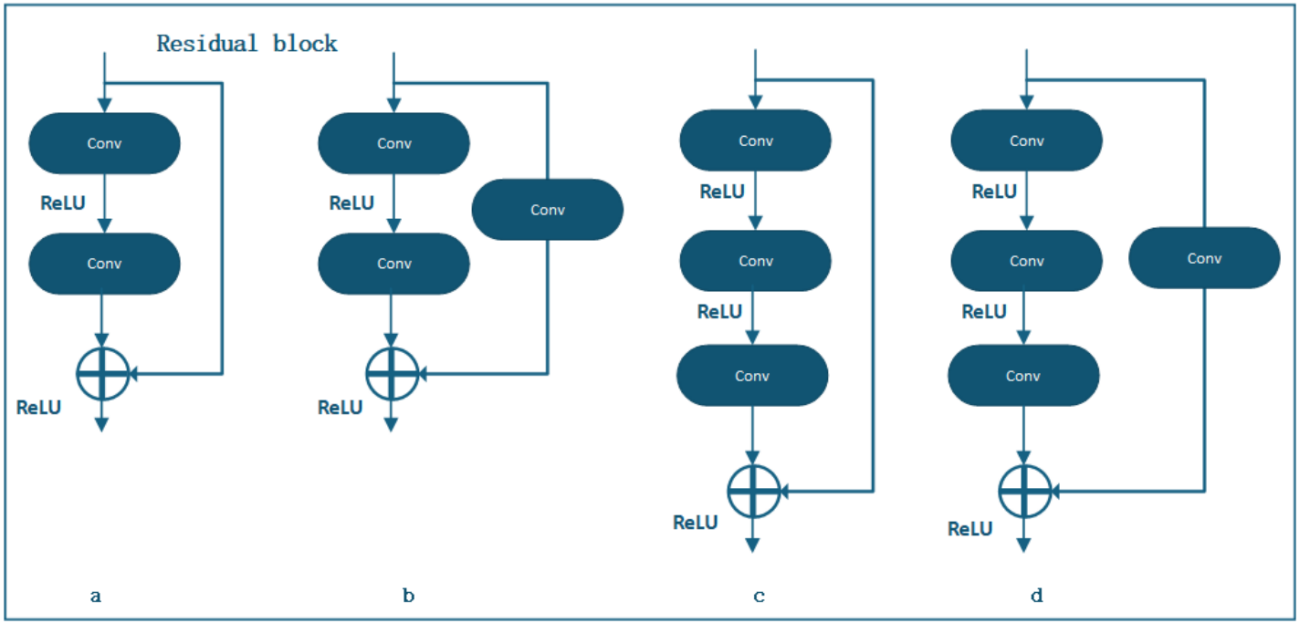


图2.3　ResNet的四种残差结构

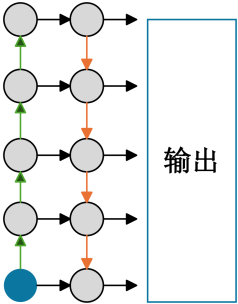
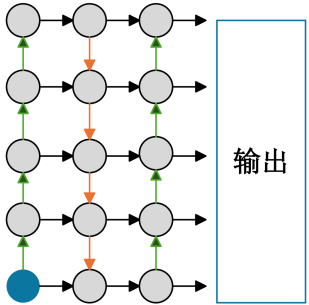
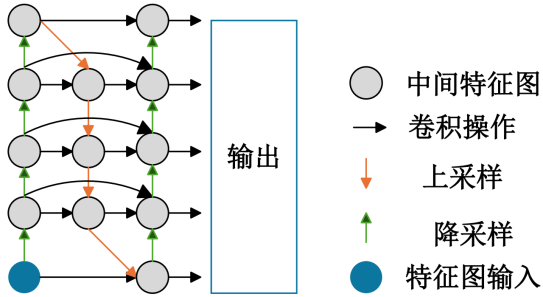
**（2）多尺度特征表示与融合**

医学 CT 影像中的病灶目标通常呈现尺度小、形态差异大、边界不清等特点，单一尺度的特征表示往往难以同时兼顾细粒度结构与高层语义信息。一般而言，浅层特征图分辨率较高，能够保留较丰富的边缘与纹理细节，但语义表达能力相对有限；深层特征图语义信息更强，能够提供更稳定的目标判别线索，但空间分辨率较低，容易造成小目标位置信息的丢失。因此，通过多尺度特征表示与融合实现不同层级特征的互补，是提升医学图像目标检测性能的关键手段之一。如图2.4所示，特征金字塔网络（FPN）、路径聚合网络（PANet）以及双向特征金字塔网络（BiFPN）是目标检测领域中具有代表性的三类多尺度特征融合结构。

FPN 由 Lin 等提出[22]，旨在缓解目标检测中的尺度变化问题。其核心思想是构建自顶向下的语义传递路径，并通过横向连接将高层语义特征融合到低层高分辨率特征中，从而在保持空间细节的同时增强特征的语义表达能力。如图2.4(a)所示，FPN 通过逐层上采样并与对应尺度的浅层特征进行融合，使模型在不同尺度上均具备较强的检测能力。对于医学 CT 中尺度较小的病灶目标，该结构能够在一定程度上缓解深层特征分辨率不足导致的小目标漏检问题。

在 FPN 的基础上，Liu 等提出 PANet[24]，进一步引入自底向上的路径增强机制，以强化低层细节信息向高层语义特征的传递过程。如图2.4(b)所示，PANet 在特征金字塔中增加自底向上的聚合路径，使不同尺度特征之间的信息流动更加充分，从而提升整体特征表达的完整性。由于医学影像中目标与部分背景结构在外观上具有一定相似性，PANet 通过增强细节与语义的双向融合，有助于提高模型对关键结构的辨识能力并降低误检风险。

BiFPN 由 Tan 等在 EfficientDet 框架中提出[23]，其在 PANet 的双向融合思想基础上进一步强调高效的跨尺度特征融合方式。如图2.4(c)所示，BiFPN 通过引入可学习的融合权重对来自不同尺度的特征进行加权组合，使模型能够根据任务需求自适应调整各尺度特征的贡献比例，从而提升融合效果；同时，BiFPN 对冗余连接进行简化，在保证特征融合能力的前提下有效降低了计算复杂度，更适合对推理效率有要求的实际应用场景。

（a）FPN （b）PANet (c)BiFBN

图2.4　卷积特征融合方法

从更一般的角度来看，多尺度特征融合可抽象为不同尺度特征在空间域上的加权组合过程。设来自不同层级的多尺度特征分别为 与 ，常见融合方式可表示为加权组合后通过卷积映射得到融合特征：

(2-1)

其中 与 为融合权重，用于平衡不同尺度特征在融合结果中的占比。

进一步地，当引入频域分支对特征进行增强时，多尺度融合过程可扩展为空间域与频域特征的协同建模，其表达形式可写为：

(2-2)

其中 表示空间域特征分支， 表示频域特征分支。通过空间域与频域特征的联合融合，可以进一步增强模型对纹理变化与结构差异的刻画能力，为后续医学图像目标检测任务提供更加充分的多尺度特征表示。

### 注意力机制的引入

医学图像目标检测的核心目标是在复杂组织背景中实现病灶区域的可靠定位与初步判别。与自然场景相比，CT 等医学影像存在灰度对比度有限、结构纹理相似度高、目标尺度偏小且形态差异显著等特点，使得检测模型不仅需要具备对局部细节的敏感性，还需要整合跨区域的上下文线索以降低漏检与误检风险。传统卷积神经网络（CNN）主要通过局部卷积核堆叠逐层扩大感受野，但其信息交互在空间上具有“逐层传递”的特点，当目标判别高度依赖长距离结构关系（例如病灶与邻近血管、胸膜或支气管走向的相对关系）时，仅依赖局部感受野容易出现语义歧义。为此，近年来研究者逐渐引入注意力机制（Attention），通过显式建模特征之间的相关性实现全局信息聚合，从而增强模型在复杂医学影像背景下的特征表达能力与目标判别鲁棒性。

注意力机制的直观含义可概括为“在给定输入的情况下，自适应强调重要信息并抑制无关信息”。从数学角度，注意力机制可被抽象为一个从输入特征到加权输出特征的映射过程，如图2.5所示，其一般形式可写为：

其中 表示输入特征（可为序列或二维特征图展开后的序列）， 表示可学习参数， 表示注意力算子。该抽象形式强调：注意力并非固定规则，而是依据输入内容动态变化的“自适应选择机制”。

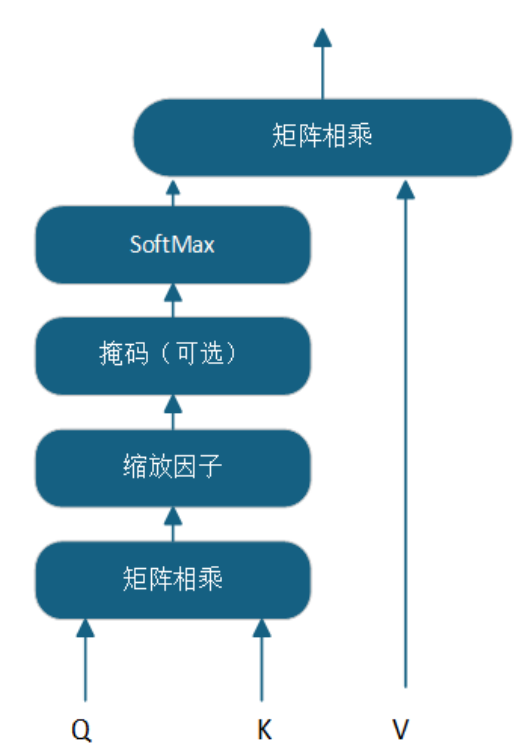


图2.5　缩放点积注意力

**（1）自注意力：突破局部感受野的全局依赖建模**

在视觉任务中，自注意力（Self-Attention）通过在同一输入内部建立任意位置之间的依赖关系，使特征能够直接与远距离位置进行信息交互。如图2.6所示，给定输入特征图 ，通常先将其展平为长度为 的序列：

(2-4)

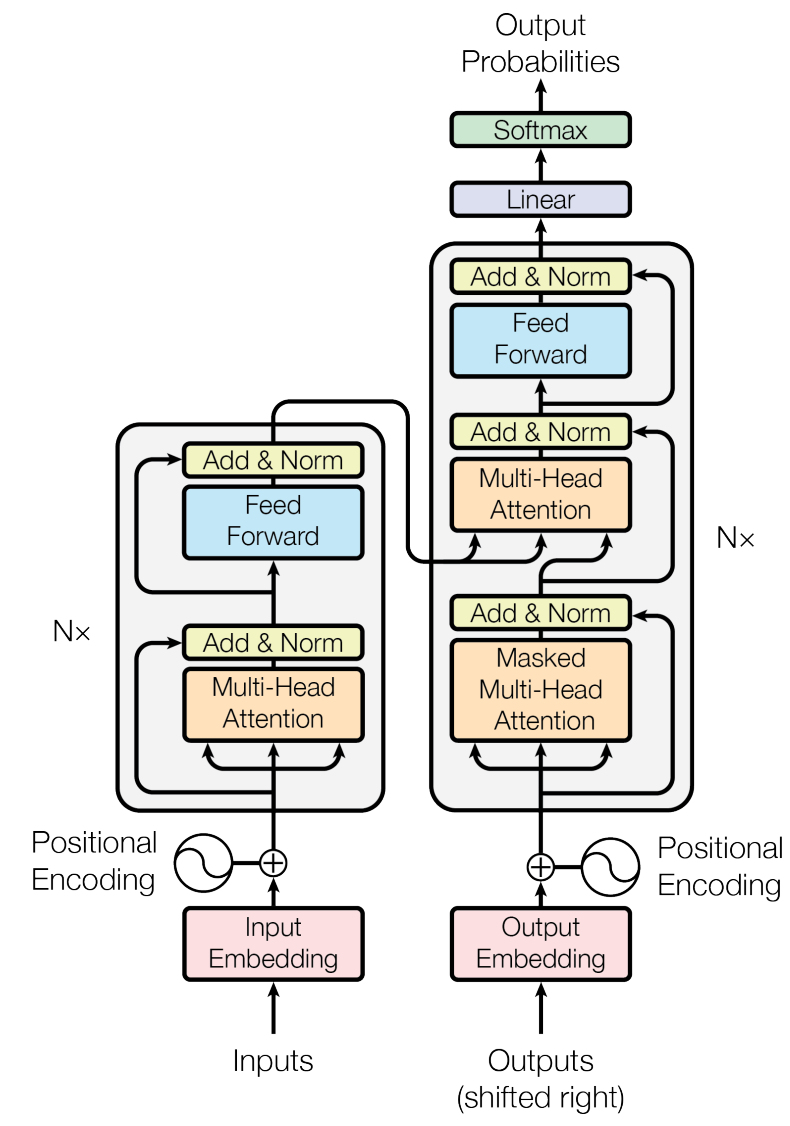


图2.6　Transformer 总体结构

随后通过线性投影构造查询（Query）、键（Key）和值（Value）三类向量：

(2-5)

其中 为可学习矩阵， 为投影后的特征维度。三者的意义可理解为： 表示“我在找什么”， 表示“我有什么线索”， 表示“我携带的信息内容”。自注意力的核心计算采用缩放点积注意力（Scaled Dot-Product Attention），定义为：

(2-6)

其中 给出了所有位置之间的相似度矩阵，分母 用于缩放内积以缓解维度增大导致的数值不稳定；Softmax 将相似度归一化为注意力权重，从而实现对 的加权求和。对第 个位置而言，其输出特征可写为：

(2-7)

上述公式体现了注意力机制的关键性质：输出 并非仅由局部邻域决定，而是由整幅图像范围内的全局特征加权融合得到，这使模型能够在医学影像中利用更丰富的上下文线索来区分外观相似的组织结构。

**（2）位置编码：将空间结构显式注入注意力计算**

注意力机制本身对输入序列的排列是“置换等价”的，为使模型感知空间位置信息，通常引入位置编码（Positional Encoding）。如图2.6所示，可将位置编码 与输入特征相加得到带位置信息的序列表示：

(2-8)

对于二维图像特征，位置编码可分解为水平与垂直两个方向的编码并组合。常见的正弦位置编码形式为：

(2-9)

其中 表示位置索引， 表示维度索引。通过引入 ，注意力权重不仅由外观特征决定，也会隐式关联空间布局信息，从而更符合医学影像中“结构相对位置”对判别的重要性。

**（3）多头注意力：多子空间并行建模与表征增强**

单头注意力往往在一个投影子空间内建模相关性，可能不足以同时捕捉局部纹理、形态边界与全局结构关系。为此，多头注意力（Multi-Head Attention）通过并行多个注意力头，在不同子空间学习互补关系。如图2.7所示，多头注意力定义为：

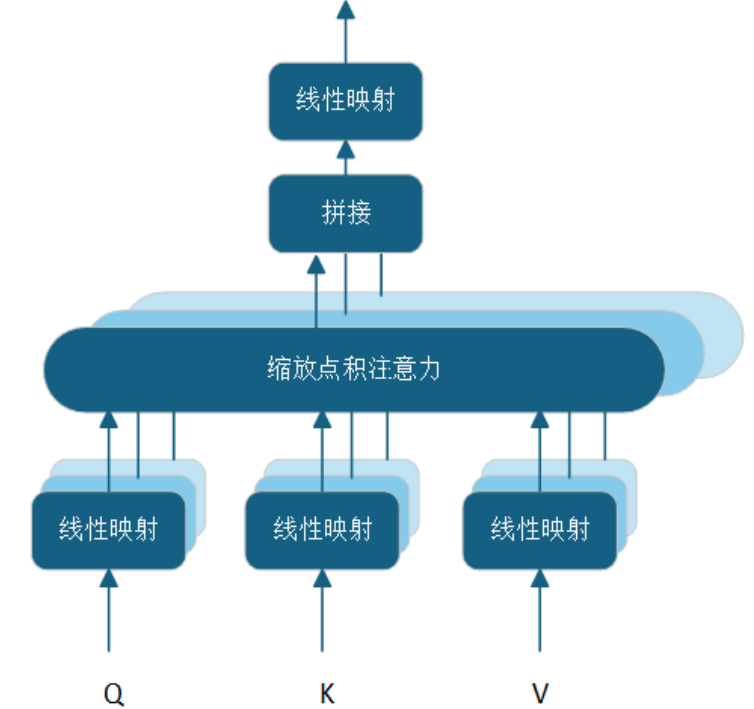


图2.7　多头注意力结构

(2-10)

其中第 个注意力头为：

(2-11)

为头数， 为输出映射矩阵。直观上，多头注意力能够让模型“同时用多种视角看同一幅影像”，例如一个头偏向聚焦边缘结构，另一个头偏向聚焦整体形态轮廓，从而提升对小目标和模糊边界目标的稳定性。

**（4）Transformer 基本块：残差、归一化与前馈网络的组合**

在 DETR 及其衍生方法中，注意力通常嵌入 Transformer 编码器—解码器结构中。为稳定训练并增强非线性表达能力，Transformer 块通常包含残差连接（Residual）、层归一化（LayerNorm）与前馈网络（FFN）。如图2.6所示，一个典型子层输出可写为：

(2-12)

(2-13)

其中 为层归一化算子，FFN 通常为两层全连接：

(2-14)

为非线性激活函数（如 ReLU/GELU）。该结构通过“注意力（全局交互）+ FFN（逐位置非线性变换）”的组合，实现对复杂医学影像特征的深层建模。

**（5）从注意力到目标检测：集合预测与查询机制**

基于注意力机制的 DETR 框架将目标检测视为集合预测问题，通过解码器中的查询向量（Object Queries）与编码器输出特征进行交互，实现目标的直接定位与分类［26］。如图2.8所示，设编码器输出为 ，查询向量集合为 （ 为查询数量），解码器通过交叉注意力实现查询与图像特征的匹配，其形式为：

(2-15)

其中 为更新后的查询特征。随后，每个查询输出一个分类结果与边界框参数。边界框常用中心点与宽高参数化：

(2-16)

在医学图像中，由于单幅图像往往目标数量有限，查询机制能够减少密集候选产生带来的背景干扰，使模型更倾向于输出“少而准”的病灶提示结果。



图2.8　DETR总体结构

**（6）匹配思想与训练稳定性：二分图匹配的集合一致性约束**

为保证预测集合与真实目标集合的一致性，DETR 在训练阶段引入集合匹配思想。令真实目标集合为 ，预测集合为 ，通过匈牙利匹配获得最优对应关系［26］，其可抽象为最小化匹配代价：

(2-17)

其中 为匹配排列， 为代价函数（可由分类一致性与框回归误差构成）。在本节中不展开具体损失细节，仅**强调**：该匹配机制使查询预测具备“集合级一致性”的训练约束，从而提升端到端检测的稳定性与可解释性。

**（7）面向医学图像检测的三点作用总结**

结合上述机制，注意力机制在医学图像目标检测中的引入可概括为三方面价值（并为后续章节铺垫）：

1）全局结构关系建模：自注意力通过式(2-6)–(2-7)建立跨区域依赖，使模型能够利用整体结构线索辅助局部判别，缓解医学影像中目标与背景外观相似导致的误检问题。如图2.6所示，这种全局交互对于复杂组织结构的区分具有重要意义。

2）查询驱动的目标表达：通过式(2-15) 的交叉注意力，查询向量与编码器特征完成显式交互，使目标定位由“密集候选筛选”转向“集合预测匹配”。如图2.8所示，该范式更契合医学影像中目标数量较少但需高置信提示的应用特征。

3）可部署导向的效率探索：在保持注意力建模优势的基础上，后续实时 DETR 系列模型进一步探索高效注意力计算与工程化简化，使端到端检测思想更贴近实际部署需求［27］。如图2.8所示，实时化思路强调在计算开销可控的前提下维持对关键区域的关注能力，为后续章节引入 ROI 约束或弱监督线索提供了可衔接的建模基础。

### 实时端到端检测模型RT-DETR的基本思想

端到端 Transformer 检测方法将目标检测从“密集候选框生成 + NMS 后处理”的传统流程，转化为集合预测（set prediction）问题：模型直接输出固定数量的预测集合，并通过匹配机制完成监督，从而减少后处理依赖并提升推理链路的一致性。然而，在实时检测场景中，端到端范式的主要瓶颈并不只在解码器，而往往出现在多尺度特征交互导致的计算冗余，以及查询（query）质量不稳定带来的收敛与精度损失。为此，RT-DETR[28]在保持端到端优势的同时，围绕“如何用更少的代价获取更有效的特征交互”与“如何让更可靠的查询进入解码器”两个关键点展开设计，其整体流程如图2.9所示。

在医学 CT 等影像的目标检测任务中，病灶目标通常呈现“小尺度、弱对比、形态差异大”的特点，同时背景结构复杂且与病灶相似度高，这使得模型必须同时具备细粒度定位能力与跨区域上下文建模能力。综合来看，实时端到端检测在该类任务中常遇到以下挑战：其一，小目标密集与尺度差异显著导致多尺度表征不足易漏检；其二，背景干扰强、相似结构多导致误检风险提升；其三，上下文依赖明显，仅局部特征难以稳定判别；其四，精度与速度难以兼得，部署侧对延迟敏感进一步放大这一矛盾。针对上述问题，RT-DETR 主要通过“高效多尺度编码”与“高质量查询选择”来减少无效计算并提升有效交互，其编码器演进思路如图2.10所示，跨尺度融合模块示例如图2.11所示，不确定性引导的查询选择可视化如图2.12所示。

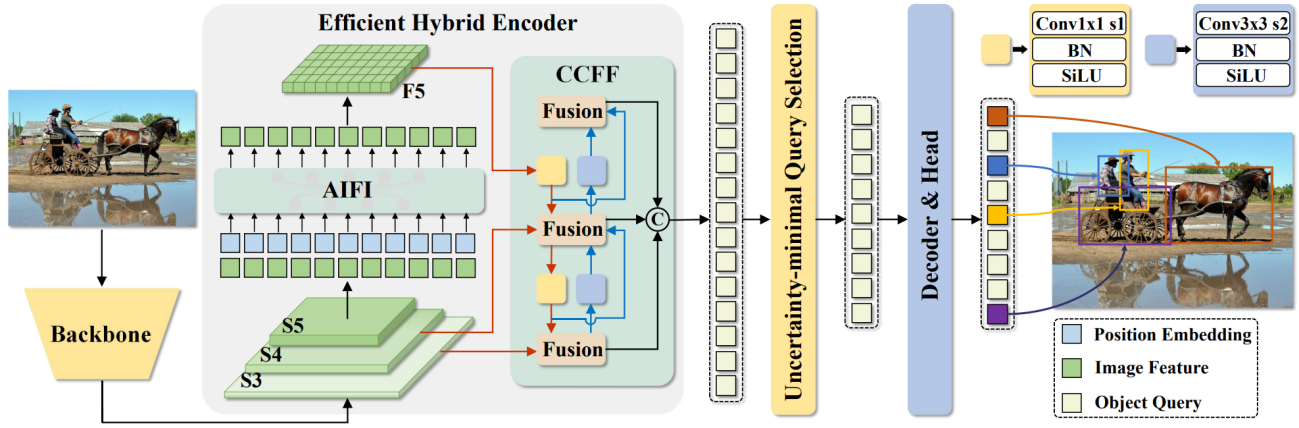


图2.9　RT-DETR 端到端实时检测总体框架

**（1）多尺度特征编码与计算效率优化**

设输入图像为 ，骨干网络提取多尺度特征：

(2-14)

其中 为 backbone， 表示尺度层级。

将二维特征展平为序列：

(2-15)

多尺度拼接得到总序列：

(2-16)

在全局自注意力中，注意力矩阵为：

(2-17)

输出为：

(2-18)

当 时，计算代价随 增长，可写为：

(2-19)

对于高分辨率医学图像，多尺度拼接会显著增大 ，从而使编码端成为实时瓶颈之一。RT-DETR 的关键思想是：如图2.10所示，将多尺度信息利用拆分为“尺度内交互”和“跨尺度融合”，让更贵的注意力交互优先集中在语义更强的高层特征上，同时通过较轻量的融合模块完成尺度间的信息整合。形式化地，记尺度内交互为 ，跨尺度融合为 ，则：

(2-20)

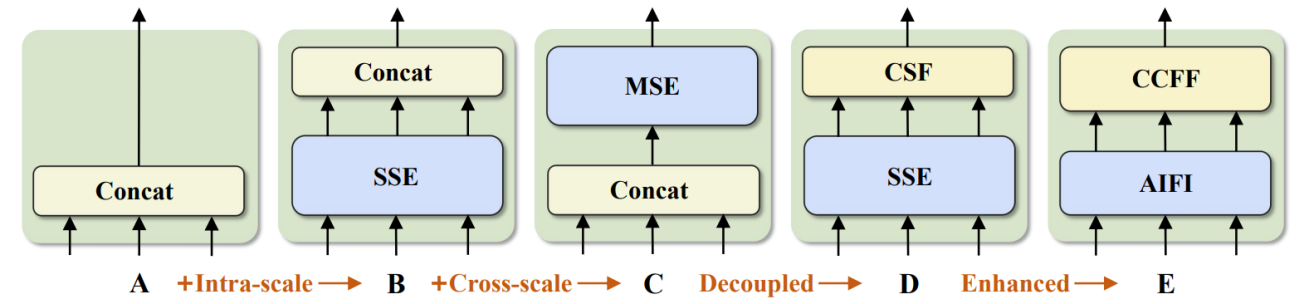


图2.10　多尺度编码器结构演进示意

其中 为编码器输出的融合特征，用于后续查询交互。跨尺度融合模块可抽象为对齐与融合操作（图2.11给出示意）：

(2-21)

(2-22)

其中 可以理解为卷积式或门控式融合算子，用于将不同尺度的互补信息整合到统一空间，降低“全都做注意力交互”的冗余。

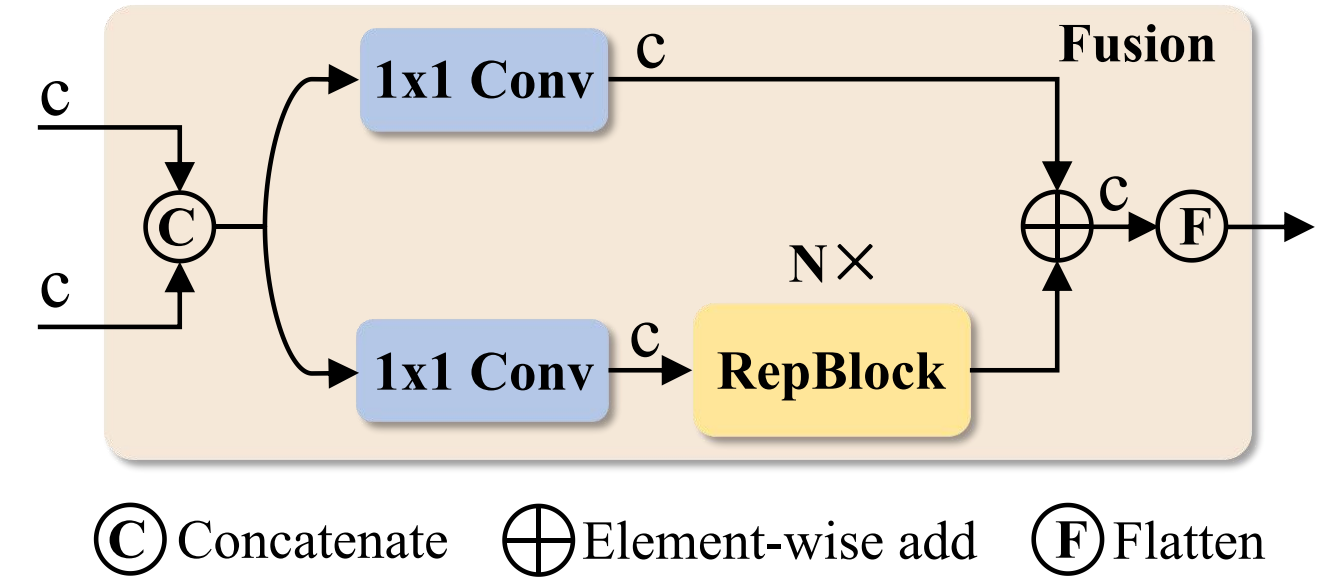


图2.11　跨尺度融合模块示意

**（2）查询选择机制与解码器输入质量提升**

端到端检测最终输出固定数量 个预测：

(2-23)

其中 为类别预测， 为边界框预测。

训练阶段通常需要建立 GT 与预测的一对一匹配：

(2-24)

这种匹配机制减少了重复预测，但也意味着：如果初始 query 不够“像目标”，decoder 需要更多层去修正，既影响收敛也影响实时性。RT-DETR 的核心改进之一是“uncertainty-minimal query selection”：如图2.12所示，它不只依据分类分数挑选 query，而是引入一种“分类—定位一致性”的不确定性度量，倾向于选择“分类和定位都更可靠”的特征进入 decoder。可将不确定性写为：

(2-25)

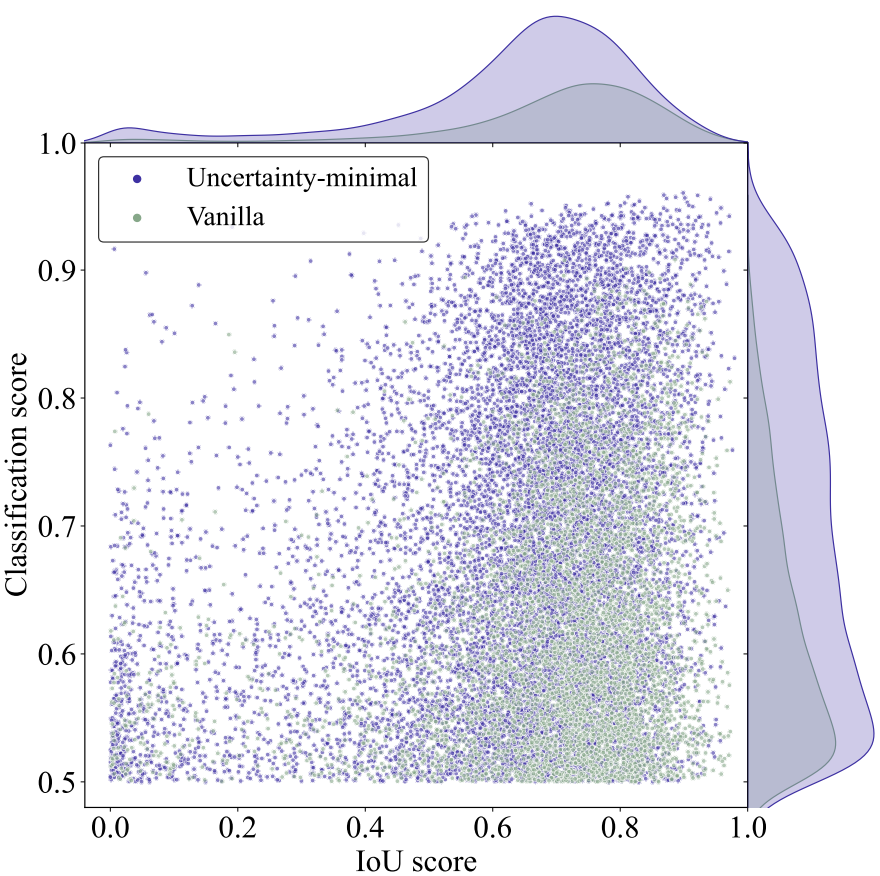


图2.12　不确定性最小查询选择的可视化结果

其中 表示与定位质量相关的得分（如 IoU 质量估计或回归置信）， 表示与分类相关的得分/分布， 越大表示“分类认为像目标但定位不稳”（或相反），从而不适合作为高质量 query。在此基础上，查询选择可写为：

(2-26)

其中 为权重， 为初始 query 集合。直观解释是：在同样分类分数下，更偏好不确定性更小的候选特征，从而减少 decoder 的无效修正。同时，训练目标可抽象为：

(2-27)

其中 约束不确定性， 为其权重。该写法用于表达“把 query 质量显式纳入优化”的思想层逻辑，不涉及具体实现细节。

**（3）速度可调：decoder 深度当成部署旋钮**

为适配不同部署场景，RT-DETR 强调端到端检测也需要“可调节性”：在不改变整体训练范式的情况下，可通过调整 decoder 迭代步数实现速度—精度折中。设 decoder 的迭代更新为：

(2-28)

其中 为初始 query， 为编码器输出特征。每步预测为：

(2-29)

推理阶段可取前 步的输出作为结果

(2-30)

当 较小，延迟更低；当 增大，通常可获得更充分的特征交互与更稳定的框回归效果。该机制使端到端检测在实时部署时具备更好的工程可控性，与“无需 NMS 的统一推理链路”一起构成 RT-DETR 的实用性核心。

## 基于 ROI 粗定位引导的医学图像半监督语义分割相关技术

### 基于 U-Net 的医学图像语义分割模型基础

医学图像语义分割的目标是对图像中每一个像素赋予明确的语义标签，从而实现对病灶区域的精确定位与形态刻画。在现有方法中，基于编码器–解码器结构的 U-Net[38] 模型由于其结构简洁、特征利用充分，在医学图像分割领域得到了广泛应用，尤其适用于样本规模有限且对定位精度要求较高的医学场景。

U-Net 采用对称的编码器–解码器架构，如图 2.13 所示。编码器部分通过多层卷积与下采样操作逐步提取高层语义特征，以增强模型对复杂结构的判别能力；解码器部分则通过上采样逐级恢复空间分辨率，实现对目标区域的精细重建。与传统全卷积网络不同，U-Net 在编码器与解码器之间引入了跳跃连接（skip connection），将浅层的空间细节特征直接传递至解码端，有效缓解了下采样过程中空间信息丢失的问题。设输入医学图像为：

(2-31)

其中 、 分别表示图像的空间尺寸， 表示通道数。U-Net 的编码过程可视为一系列特征映射函数的级联：

(2-32)

其中 表示第 层编码模块， 为对应尺度下的特征表示。解码阶段则通过上采样操作逐步恢复分辨率，并与对应编码层的特征进行融合：

(2-33)

其中 表示解码模块， 表示特征拼接操作。最终，网络输出与输入图像同尺度的像素级预测结果。

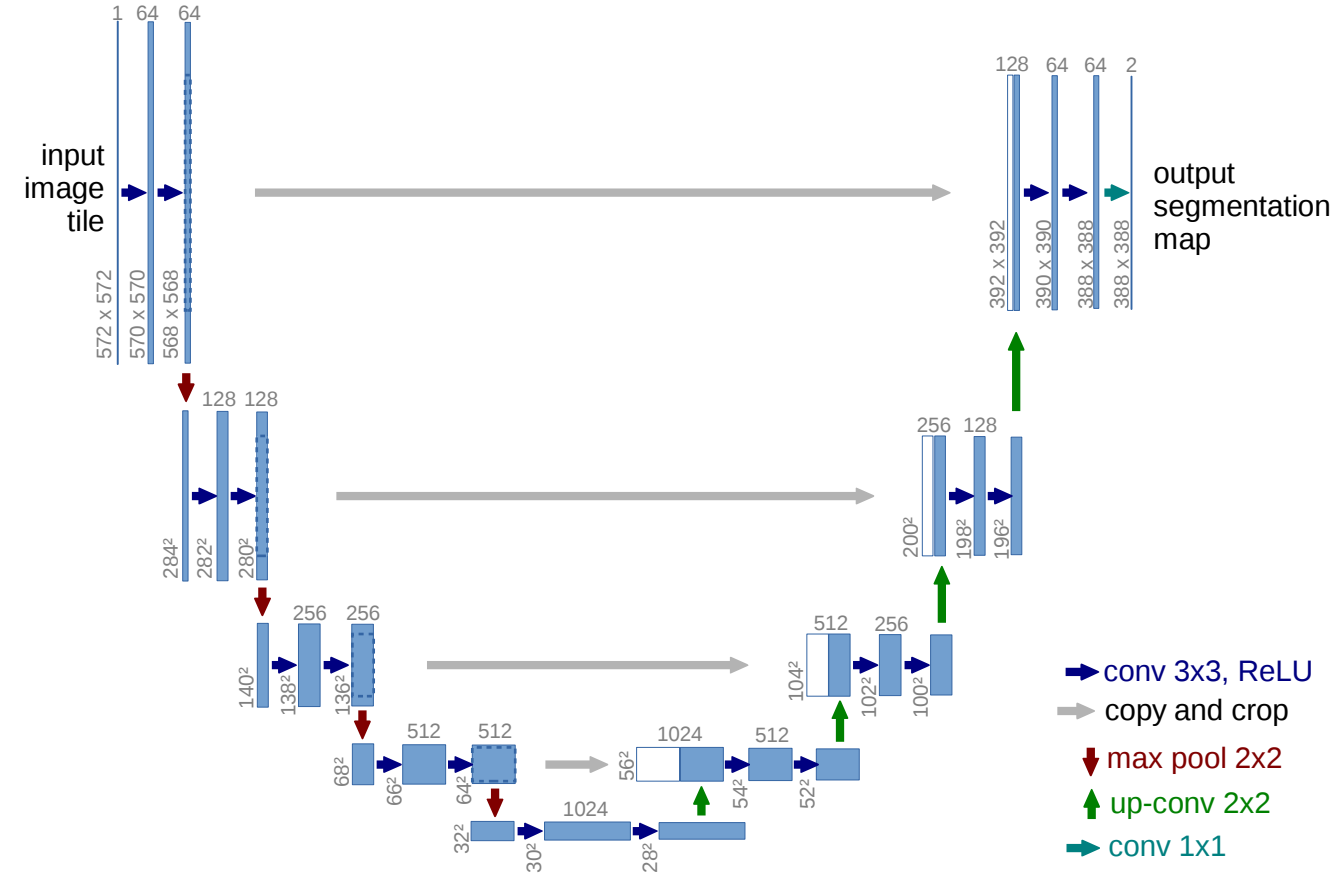


图2.13　U-Net模型示意图

得益于跳跃连接机制，U-Net 能够在保持高层语义表达能力的同时，充分利用低层空间细节信息，使其在病灶边界刻画与小目标分割方面表现稳定。因此，U-Net 及其变体已成为医学图像语义分割领域的重要基础模型，并常被用作弱监督或半监督分割方法中的核心分割骨干网络。

在 U-Net 框架被广泛应用的基础上，后续研究逐渐关注模型在不同医学图像任务中的泛化性与工程可复现性问题。为减少人工调参对分割性能的影响，Isensee [39]等提出了 nnU-Net（no-new-U-Net）框架，其核心思想并非设计全新的网络结构，而是基于 U-Net 体系对数据预处理、网络配置、训练策略和后处理流程进行系统化自适应调整。nnU-Net 根据具体任务的数据分辨率、体素间距和样本规模，自动选择二维或三维 U-Net 结构，并动态配置网络深度、卷积核尺寸及下采样策略，从而在多种医学图像分割基准任务中取得稳定且具有竞争力的性能。

### 医学图像半监督语义分割问题形式

医学图像语义分割旨在学习从影像空间到像素级语义标签空间的映射，以实现对病灶区域的精确描绘。然而，在 CT 等医学影像中获取像素级标注需要专业医生逐层勾画，标注成本高、耗时长，导致可用的精细标注样本规模通常受限。为缓解这一矛盾，半监督语义分割（Semi-supervised Segmentation）通过同时利用少量标注数据与大量无标注数据，在降低人工标注依赖的同时尽可能保持分割精度与泛化能力[32-33]。

设训练集由两部分组成：少量有标注样本 与大量无标注样本 。其中 表示输入 CT 图像（或 patch）， 表示像素级标签图。分割网络记为 ，输出为像素级类别概率图 。在经典的监督学习情形下，分割训练可通过交叉熵、Dice 等损失进行优化[30]；而在半监督设置中，需要为无标注样本构造可学习信号（如伪标注、一致性约束等），并与监督信号共同驱动训练[32-33]。

### ROI 粗定位引导的关键区域建模

在临床辅助诊断流程中，“先定位、后精分”是一种更贴近实际的范式：先通过粗粒度定位（例如检测框或候选区域）锁定可疑病灶，再在该区域内进行精细分割。对于 CT 影像而言，这种 ROI（Region of Interest）粗定位具有两点直接收益：其一，降低背景组织对模型学习的干扰，使网络更集中地建模病灶边界与纹理细节；其二，在小目标占比高的场景下，ROI 有助于提升有效分辨率与目标显著性，从而减少漏检漏分带来的误差积累。基于 ROI 的两阶段思想在目标检测与语义分割中已被广泛采用，ROI 引导的整体流程如图2.13所示。例如以候选区域（ROI）为核心的两阶段框架能够将计算资源聚焦于可能存在目标的位置，并通过 ROI 特征对齐进一步提升定位与分割效果[29,31]。因此，将 ROI 粗定位引入半监督语义分割，可作为一种“弱化背景 + 强化目标”的先验约束，为后续伪标注学习提供更可靠的空间范围。

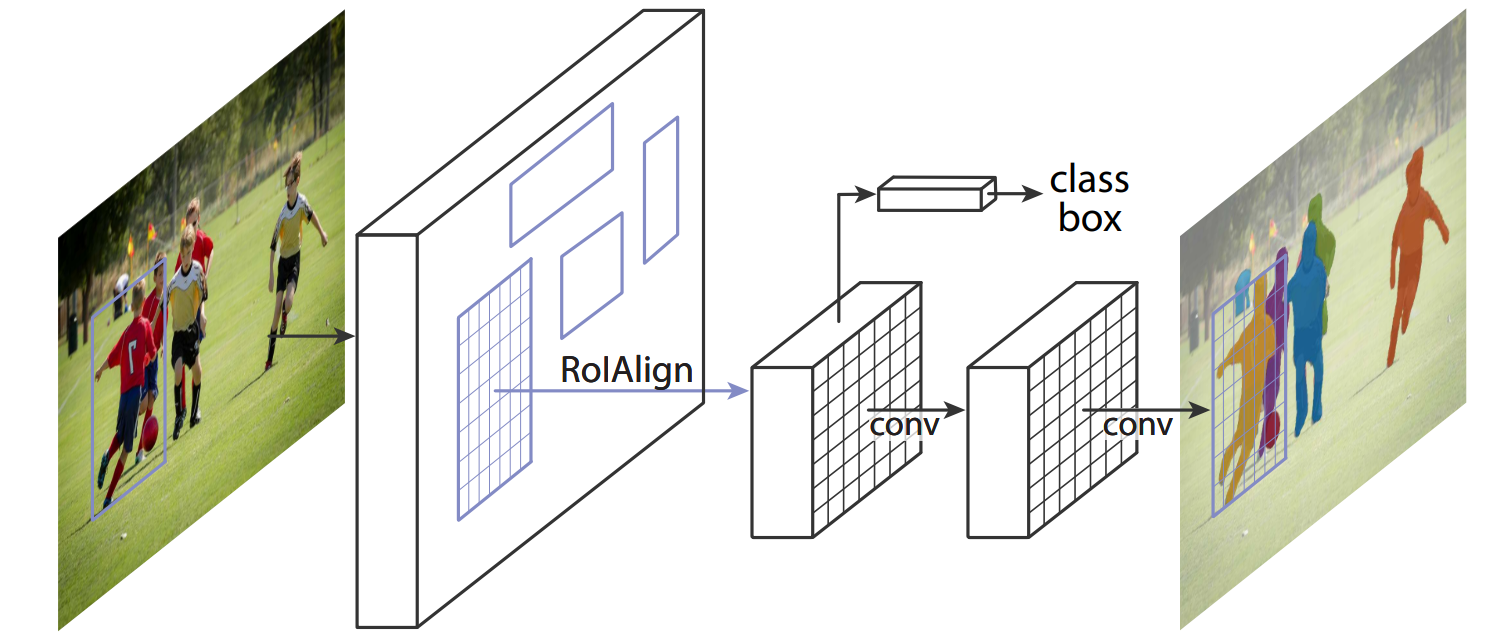


图2.14　RoI 引导的候选区域特征对齐与分割流程示意图

**（1）医学图像分割任务中 ROI 的常见形式**

医学场景下 ROI 粗定位通常不追求“像素级精确”，而强调“高召回的候选覆盖”，常见形式包括：

1）检测框 ROI：通过检测器输出 bounding box，得到病灶候选区域。

2）候选掩码 ROI：通过粗分割/阈值/弱定位得到二值候选掩码；

3）裁剪与放大 ROI：基于候选区域对原图裁剪并重采样，提高目标占比。

这些 ROI 形式本质上都是在空间上施加约束：让分割模型的学习与优化更多发生在“可能是病灶”的区域内，而不是在整幅图像上被大量背景像素主导。

**（2）ROI 引导对半监督学习的意义**

半监督分割的关键困难之一在于：无标注数据的学习信号往往来自模型自身（或教师模型）的预测，这些预测在背景复杂、目标模糊时容易产生噪声。ROI 约束提供了一种直接的降噪方式：只在 ROI 内生成/筛选伪标注，并在 ROI 内施加一致性正则，从而把“最不确定、最像背景”的区域尽量排除在自训练闭环之外。换言之，ROI 的作用不是替代半监督策略，而是为半监督策略提供一个更稳定的空间支撑，使伪标注学习从“全图盲学”转为“在候选区域内精学”。

为便于形式化描述，引入 ROI 掩码 ，其中 表示像素 位于 ROI 内。对有标注样本，监督损失可在 ROI 内计算为：

(2-34)

其中 为像素集合，。式(2-30)体现了 ROI 作为空间约束的基本方式：对 ROI 外像素不强行学习像素级判别，从而减少背景主导的问题。

### 教师—学生框架与 ROI 约束的半监督学习

在半监督语义分割中，教师—学生（Teacher–Student）框架是一类应用极为广泛的策略：教师网络为无标注样本提供伪标注或软监督信号，学生网络在监督损失与无监督损失的共同驱动下学习，并通过一致性约束降低预测波动[32-33]。在 ROI 引导的设置中，教师—学生框架可以自然地与 ROI 约束融合：教师先在 ROI 内生成更可信的伪标注，学生只在 ROI 内学习伪标注与一致性，从而提升自训练闭环的稳定性。

**（1）伪标注生成：从教师预测到 ROI 内标签**

记教师网络为 ，对无标注样本输出概率图 。最常用的硬伪标注形式为：

(2-35)

式(2-35)将软概率离散化为像素标签，用于把无标注样本转化为“可训练样本”。在 ROI 引导下，伪标注的使用通常进一步结合置信度筛选：只让教师在 ROI 内且高置信的像素进入无监督损失。

**（2）ROI + 置信度筛选：降低噪声伪标注的影响**

设教师在像素 的最大类别概率为 ，阈值为 。则无监督伪标注损失可写为：

(2-36)

式(2-32)同时引入两种“过滤”：其一是置信度过滤 ，其二是 ROI 过滤 。二者共同作用可显著降低伪监督噪声对训练的干扰，使无标注数据的学习更加“稳”。

**（3）一致性约束：在 ROI 内保持预测稳定**

一致性正则（Consistency Regularization）强调：同一无标注样本在不同扰动/增强下的预测应保持一致[33]。令 表示两种数据增强，对应学生预测为 ，。一致性损失可写为：

(2-37)

式(2-34)强调一致性只在 ROI 内约束：即把“稳定性学习”集中在关键区域，使模型在病灶边界与内部结构上的预测更鲁棒。该思想可用教师-学生框架示意图进行说明（如图2-14所示），突出“教师提供目标、学生拟合目标、两视角一致”的训练闭环[32]。综合监督项与无监督项，最终优化目标为：

(2-38)

其中 为权重系数，用于平衡监督学习与半监督学习的贡献。至此，ROI 约束、伪标注学习与一致性正则形成了一个较为通用的半监督分割学习范式：ROI 降低空间噪声 → 伪标注提供学习信号 → 一致性抑制预测抖动，为后续章节在具体模型中落地这些思想提供理论铺垫。

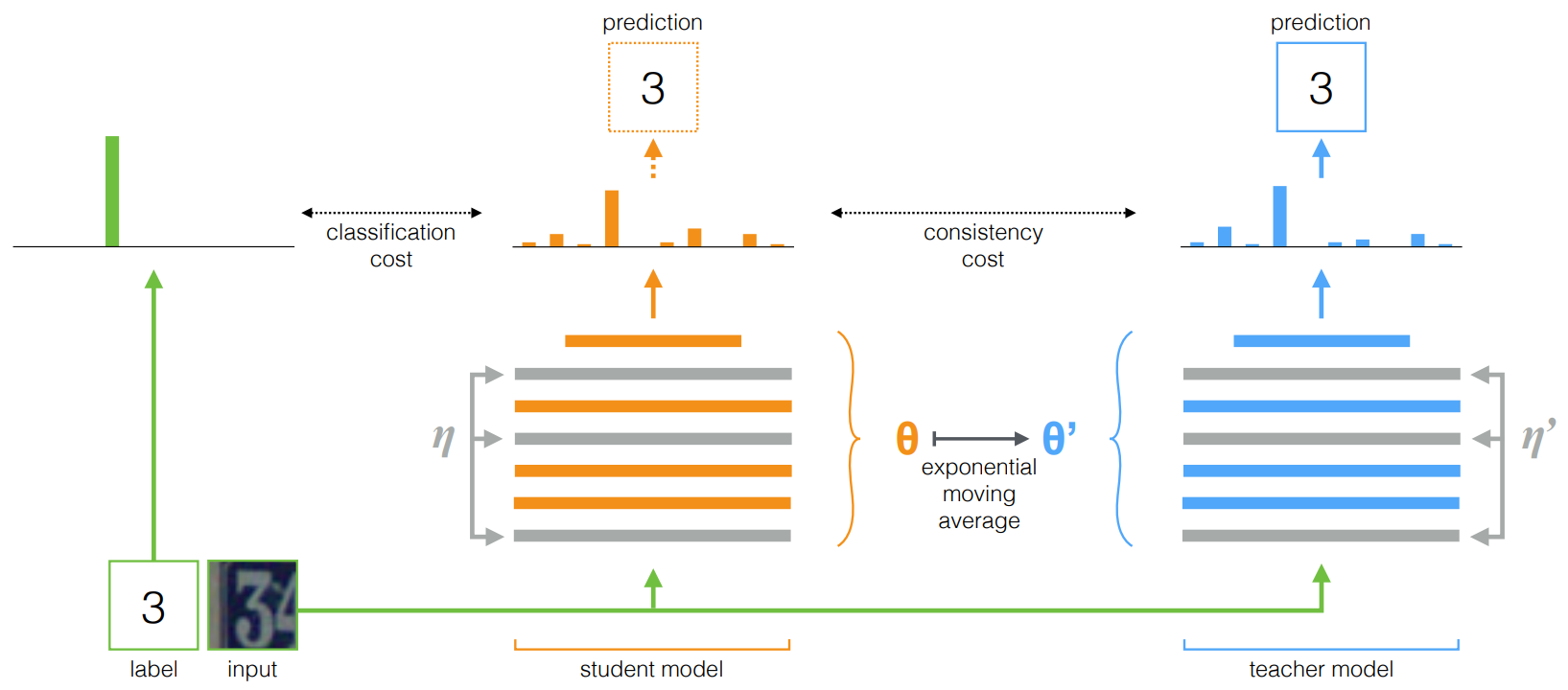


图2.15　教师-学生一致性学习的半监督训练框架示意图

## 频域相关理论

对医学图像特征在频域进行分析与建模具有天然优势：一方面，空间域与频域之间的变换在理论上是可逆的（不引入信息“凭空消失”的额外假设），另一方面，频域表示往往能把“结构信息”和“细节/噪声信息”以不同频带的形式分离出来，使得网络更容易对不同层次的模式进行建模与抑噪。在医学图像（尤其是 CT）中，病灶边界、纹理粗糙度、空洞/结节的局部对比等，常对应于中高频成分；而器官轮廓、整体密度梯度等更偏向低频成分。因此，将特征在频域中进行分解、压缩或重加权，常被用作补充空间域卷积/注意力的上下文建模手段，从而提升对小目标、弱边界和复杂背景的鲁棒性。

常见的频域变换包括傅里叶变换（Fourier Transform, FT/DFT）、离散余弦变换（Discrete Cosine Transform, DCT）与离散小波变换（Discrete Wavelet Transform, DWT）等。下面分别给出其核心形式与在图像特征建模中的典型含义。

### 傅里叶变换与频谱表征

傅里叶变换的核心思想是：将信号表示为一组不同频率正弦/余弦基函数的线性叠加。对连续时间信号 ，其傅里叶变换可写为式(2-36)：

(2-39)

其中， 为频域表示， 表示频率， 为虚数单位。式(2-35)说明：频域的每个频点刻画了信号在该频率基上的“投影强度”。在图像任务中，网络输入/特征是离散采样形式，更常用离散傅里叶变换（DFT）。对大小为 的二维离散信号（图像或特征图），二维 DFT 可写为式(2-37)：

(2-40)

其中， 为离散频率索引， 为频谱系数。通常低频（靠近原点）对应全局结构与缓慢变化区域，高频对应边缘、纹理与突变细节。为了在深度模型中更直观地刻画频带能量分布，常对频谱幅值进行统计，例如式(2-38)所示的频谱能量（或功率谱密度的离散形式）：

(2-41)

式(2-37)可用于衡量特征图整体频率响应强度；进一步也可按频带划分（低/中/高频）分别求和，用于构建“频带注意/频带约束”等机制（即对不同频段赋予不同权重），以抑制噪声频段并增强病灶相关细节频段[34]。

### DCT 的能量压缩性质与特征紧凑表示

与 DFT 不同，DCT 使用实数余弦基，具有更强的能量压缩特性：在自然图像及大量医学图像中，能量往往集中在低频 DCT 系数上，因此 DCT 常用于压缩与去冗余表示（例如 JPEG 的经典做法）。对一维长度为 的序列 ，DCT-II 的形式可写为式(2-39)：

(2-42)

其中 为归一化系数，常取

(2-43)

（归一化的选择会影响系数尺度，但不改变“能量更集中于低频”的经验规律）。在图像/特征图中，常对行列分别做一维 DCT 得到二维 DCT；对深度网络而言，DCT 的价值主要体现在紧凑表征和频带可控，即保留少量低频系数即可重构主体结构，从而为轻量化的全局建模提供低成本通道，通过选择性保留/抑制部分频段，实现对平滑背景与高频噪声的分离，帮助网络在弱边界与低对比区域更稳定地学习判别线索[35]。

### DWT 的多尺度频带分解与空间局部性

DFT/DCT 的基函数在全局范围内展开，空间局部性较弱；而 DWT 通过尺度函数与小波函数实现“空间—频率”联合建模，兼具多尺度与局部性优势，更适合刻画医学图像中“局部突变 + 多尺度结构”并存的特点。对二维信号 ，单层二维 DWT 通常把信息分解为四个子带：低频近似子带 以及三个高频细节子带 、、（分别对应水平/垂直/对角方向的细节响应）。其一层分解可概括为式(2-41)：

(2-44)

从意义上讲：保留主要结构（如器官形态、整体密度分布），有助于稳定的上下文理解；捕获边缘与纹理细节（如结节边界、细小空洞壁、局部噪声纹理），可用于细粒度判别与边界增强。进一步做多层 DWT 时，递归分解 子带即可形成金字塔式多尺度频带表示，这与分割/检测网络常见的“多尺度特征金字塔”在思想上是一致的：用低分辨率承载全局语义，用高分辨率承载细节与边界，同时通过跨尺度融合获得兼顾定位与识别的表示[36–37]。

## 本章小结

本章围绕医学图像语义分析任务，对相关理论与关键技术进行了系统梳理。首先给出了医学图像语义分析与目标检测的基本定义，介绍了基于 DETR 的医学图像目标检测相关理论，为病灶区域的粗定位提供理论基础。随后，从 U-Net 体系出发，概述了医学图像语义分割的基本问题与多尺度特征建模思想，并进一步总结了 ROI 粗定位引导下的半监督语义分割框架及其典型形式。最后，介绍了频域分析在医学图像特征建模中的基本理论，为后续结合空间域与频域信息的模型设计提供了理论支撑。

1. 基于 DETR 的医学图像目标检测模型

近年来，深度学习在医学图像智能分析领域取得了显著进展，尤其在肺部 CT 结节/病灶自动检测方面，为早期筛查与临床辅助诊断提供了技术支撑。然而，相较自然场景目标检测，医学 CT 往往存在组织结构复杂、噪声与伪影干扰、对比度不均衡等问题；同时病灶区域常呈现小尺度、边界模糊、形态不规则等特征，导致模型在定位精度、误检控制与鲁棒性之间更难兼顾。对于囊腔型肺癌而言，其多表现为薄壁空腔结构，且与部分良性病变外观相似，容易造成漏检与误检；而检测框的轻微偏差又会直接影响后续 ROI 粗定位的可靠性，从而限制后续精细分析的效果。

为解决上述问题，本文提出面向医学 CT 的实时端到端检测模型 LungROINet，以“为后续分割生成高质量 ROI 粗定位”为目标，在保持推理效率的前提下提升小目标可感知性、降低复杂背景误检，并增强区域级特征的稳定性。具体而言，本文围绕以下三个方面进行改进：

（1）小病灶特征易弱化。囊腔型病灶通常占比小、纹理弱，常规特征提取在下采样与融合过程中易丢失关键信息，导致漏检。为此，本文在特征提取阶段引入轻量的多尺度增强思路，同时保留局部细节与跨尺度语义信息，从而提升对小目标与薄壁结构的表达能力。

（2）背景干扰强导致误检。CT 中正常组织结构复杂，部分区域与病灶外观相近，模型易被背景牵引产生误检或定位不稳。为此，本文采用区域级学习与联合建模的思路，使模型在定位与判别过程中更强调潜在病灶区域、抑制无关背景响应，从而提升复杂背景下的鲁棒性。

（3）ROI 表征对框误差敏感。囊腔型病灶边界模糊，框回归易波动，轻微偏移即可引入过多背景或遗漏关键上下文，影响后续 ROI 表征。为此，本文引入软聚合与权重化汇聚的区域特征构建方式，使聚合过程更关注核心区域并保留必要上下文，从而降低对边界框误差的敏感性。

基于上述设计，LungROINet 构建了一个面向囊腔型肺癌的实时检测与 ROI 粗定位框架，为后续基于 ROI 的精细分析提供更可靠的区域先验。本文结构安排如下：第 3.1 节介绍模型框架与关键设计；第 3.2 节给出实验设置、数据集与评价指标；第 3.3 节展示对比实验结果及分析；第 3.4 节通过消融实验验证各设计的有效性；第 3.5 节对本章工作进行总结。

## 囊腔型肺癌目标检测模型LungROINet

为应对囊腔性肺癌在 CT 图像中“小目标占比高、边界模糊、背景干扰强”等检测难点，本文在 RT-DETR 的端到端检测框架上构建了 LungROINet，并从以下三个方面进行改进：

（1）多尺度小目标特征增强（PMA 模块）。针对囊腔性病灶尺度小、纹理弱导致的特征表达不足问题，LungROINet 引入 PMA 模块，在特征提取阶段融合多尺度注意力并结合部分卷积的高效建模方式，以更低的计算代价突出小病灶的关键响应，从而提升后续候选目标定位的稳定性与召回能力。

（2）定位-判别联合建模（MIL 双分支结构）。针对复杂肺部背景下“定位不稳”与“良恶性判别易受干扰”的问题，LungROINet 采用 MIL-based dual-branch 的检测头设计，将候选区域的定位学习与良恶性判别过程进行协同建模，通过区域级证据聚合增强对有效病灶线索的选择能力，从而降低背景结构带来的误检风险。

（3）边界误差鲁棒的区域特征聚合（GWSPR）。针对囊腔病灶边界模糊、候选框标注与预测存在偏差时容易造成区域特征不稳定的问题，LungROINet 引入 Gaussian-Weighted Soft ROI Pooling（GWSPR），以连续可微的高斯加权方式对 ROI 内特征进行软聚合，在保留上下文信息的同时降低对候选框精确边界的敏感性，为后续 ROI 粗定位引导的精细分割提供更可靠的区域表示。该网络结构整体设计如图 3.1 所示，接下来本小节将围绕上述三方面内容进行详细介绍。

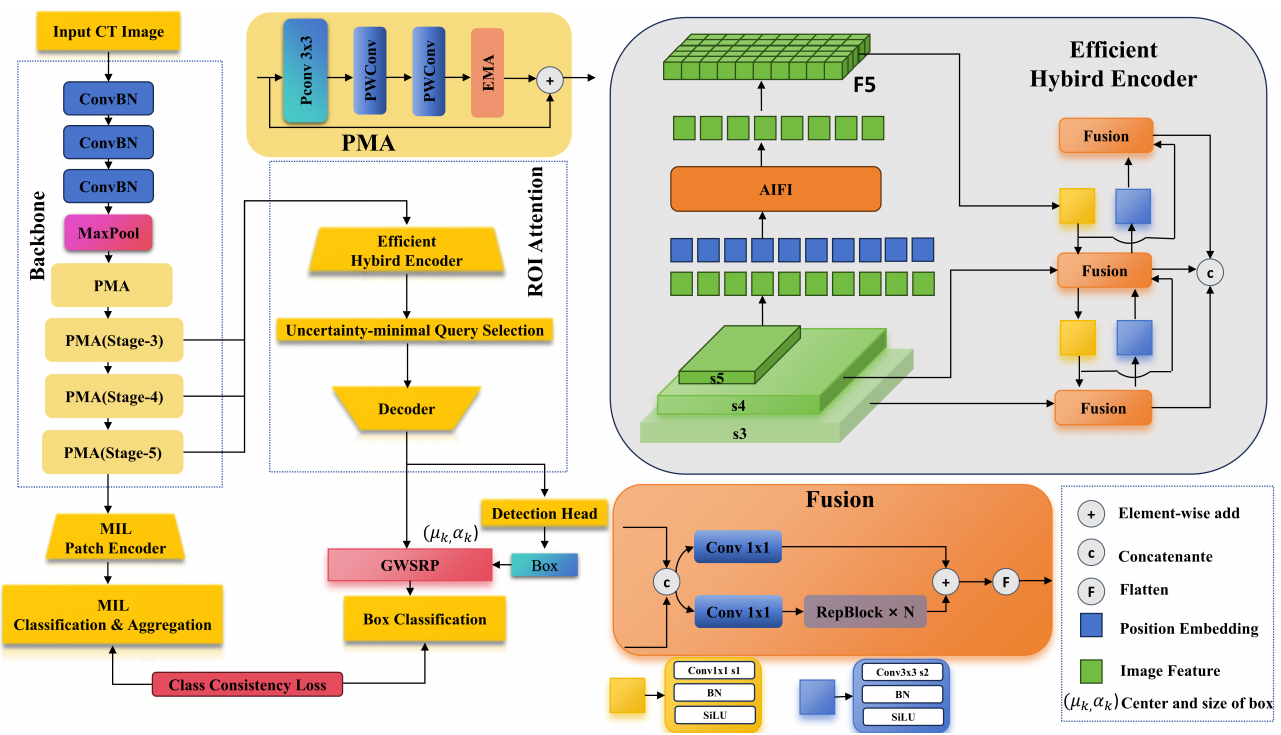


图3.1　LungROINet网络结构示意图

### 部分卷积与多尺度注意力融合模块PMA

在肺部 CT 图像的结节检测任务中，尤其是针对囊腔性结节等小尺度、低对比度、形态复杂的病灶区域，传统卷积神经网络在特征提取阶段往往面临两方面不足：一方面，标准卷积在全通道范围内进行计算，参数量和计算开销较大，不利于实时检测场景；另一方面，局部卷积操作的感受野有限，难以同时兼顾局部结构细节与全局上下文信息，从而影响小目标的定位精度与鲁棒性。

针对上述问题，LungROINet 在特征提取阶段引入了部分卷积与多尺度注意力相结合的特征增强模块（PMA），以在保持网络轻量化的前提下提升对小结节区域的判别能力。该模块通过部分卷积、通道交互与多尺度注意力建模三者的协同设计，实现对局部结构信息与全局语义信息的高效融合，其整体结构如图 3.2 所示。

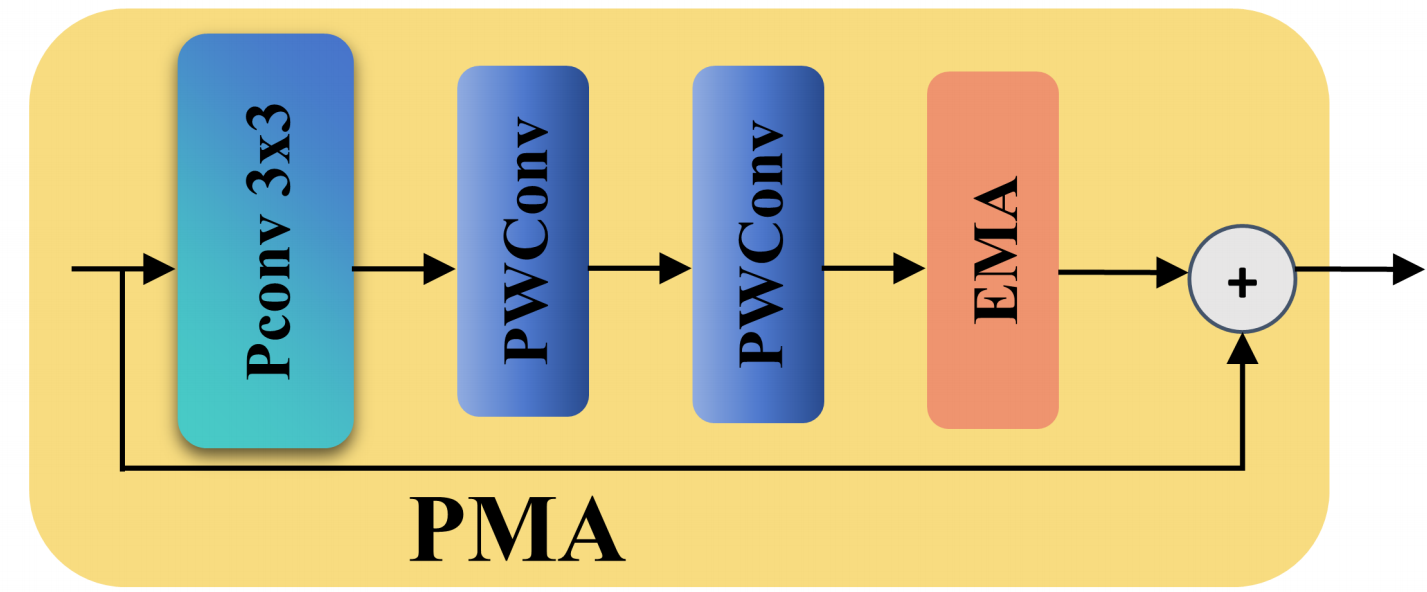


图3.2　PMA模块示意图

首先，针对标准卷积在通道维度上的冗余计算问题，PMA 模块采用部分卷积（Partial Convolution）策略[40]，仅对输入特征图中的部分通道执行空间卷积运算，其余通道通过恒等映射直接传递，从而在显著降低计算复杂度的同时保留关键信息。设输入特征图为则部分卷积操作可表示为：

(3-1)

其中 和 分别表示卷积核尺寸与步长。通过仅对 个通道进行卷积计算（），有效减少了特征提取阶段的计算负担。对应的计算复杂度（以 FLOPs 计）可近似表示为：

(3-2)

相较于全通道卷积，该方式在保持特征表达能力的同时显著提升了推理效率。随后，为弥补部分卷积在通道交互能力上的不足，PMA 模块引入逐点卷积（Pointwise Convolution）对压缩后的特征通道进行融合与重组，其计算形式为：

(3-3)

该操作在不增加空间计算复杂度的前提下增强了不同通道之间的信息交互能力，有助于提升特征判别性。

在此基础上，为进一步建模跨尺度上下文依赖关系，PMA 模块引入高效多尺度注意力机制（Efficient Multi-scale Attention, EMA）[41] ，如图 3.2 所示。该机制通过在不同空间尺度上对特征进行重排与聚合，使网络能够同时关注局部细节结构与长程依赖信息。设融合后的特征表示为 ，则多尺度注意力增强过程可概括为：

(3-4)

其中 表示多尺度注意力映射函数，用于对不同尺度特征进行加权建模。最后，PMA 模块通过残差连接将增强后的特征与原始输入特征进行融合，以稳定网络训练过程并保留原始语义信息，其输出形式为：

(3-5)

通过上述设计，PMA 模块在保证计算效率的同时，实现了对小尺度结节区域更为敏感的特征建模，为后续的区域定位与分类任务提供了更加鲁棒的特征基础。实验结果表明，该模块在提升检测精度的同时有效保持了模型的实时推理能力，为临床应用场景下的快速筛查提供了支撑。

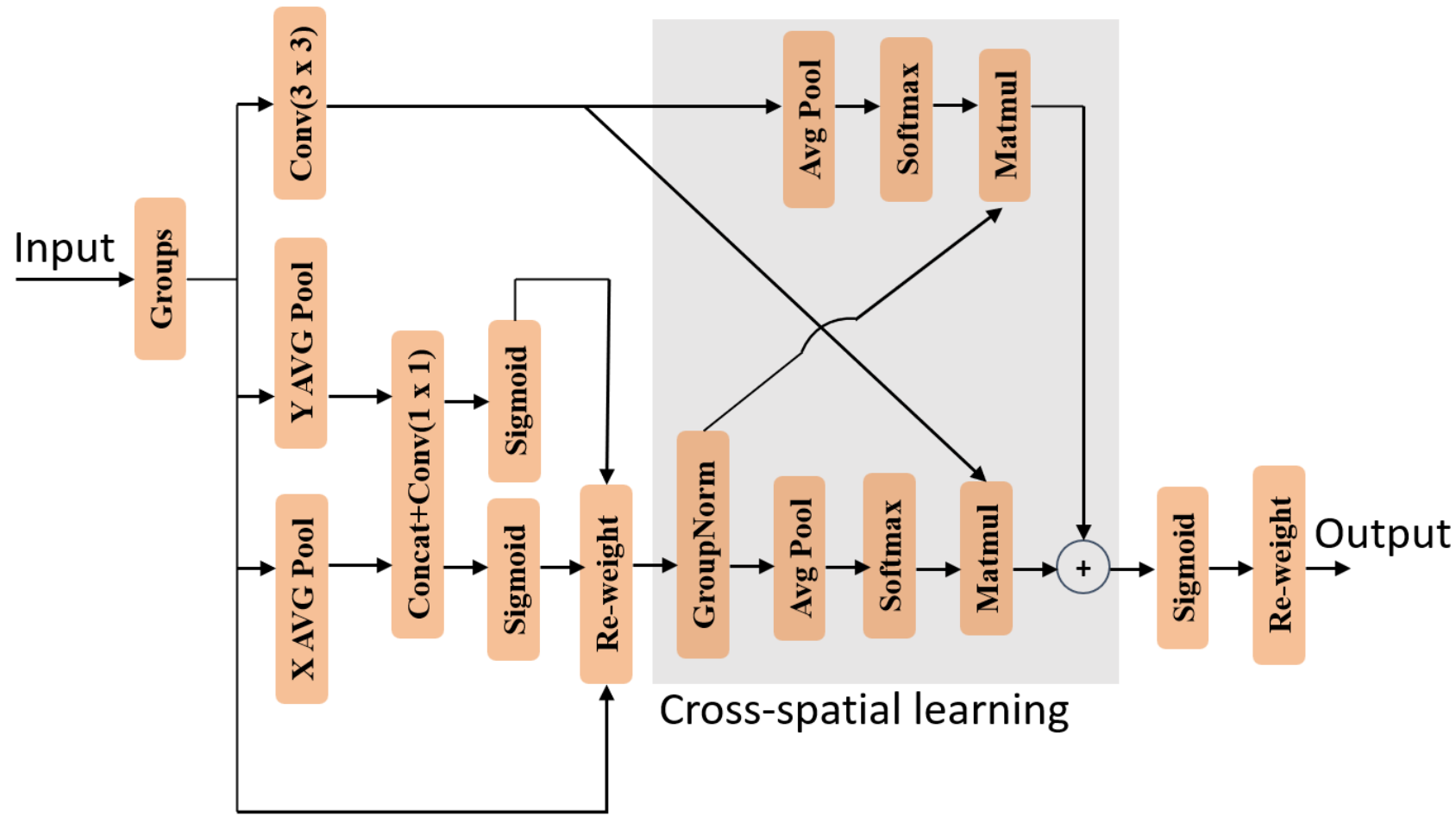


图3.3　EMA模块结构

### 基于多实例学习的Patch级分类与图像级聚合

囊腔性肺癌在 CT 图像中通常呈现病灶区域占比小、空间分布稀疏且位置不确定等特点，仅依赖全局特征进行判别容易被大量背景信息稀释关键线索。为此，本文引入多实例学习（Multiple Instance Learning, MIL）思想[42]，并采用基于 MIL 的双分支结构对局部区域进行建模：一条分支负责对局部实例进行判别与聚合，另一条分支用于保持检测框架下的定位信息表达，从而实现“定位—判别”协同学习，其整体流程如图 3.4 所示。

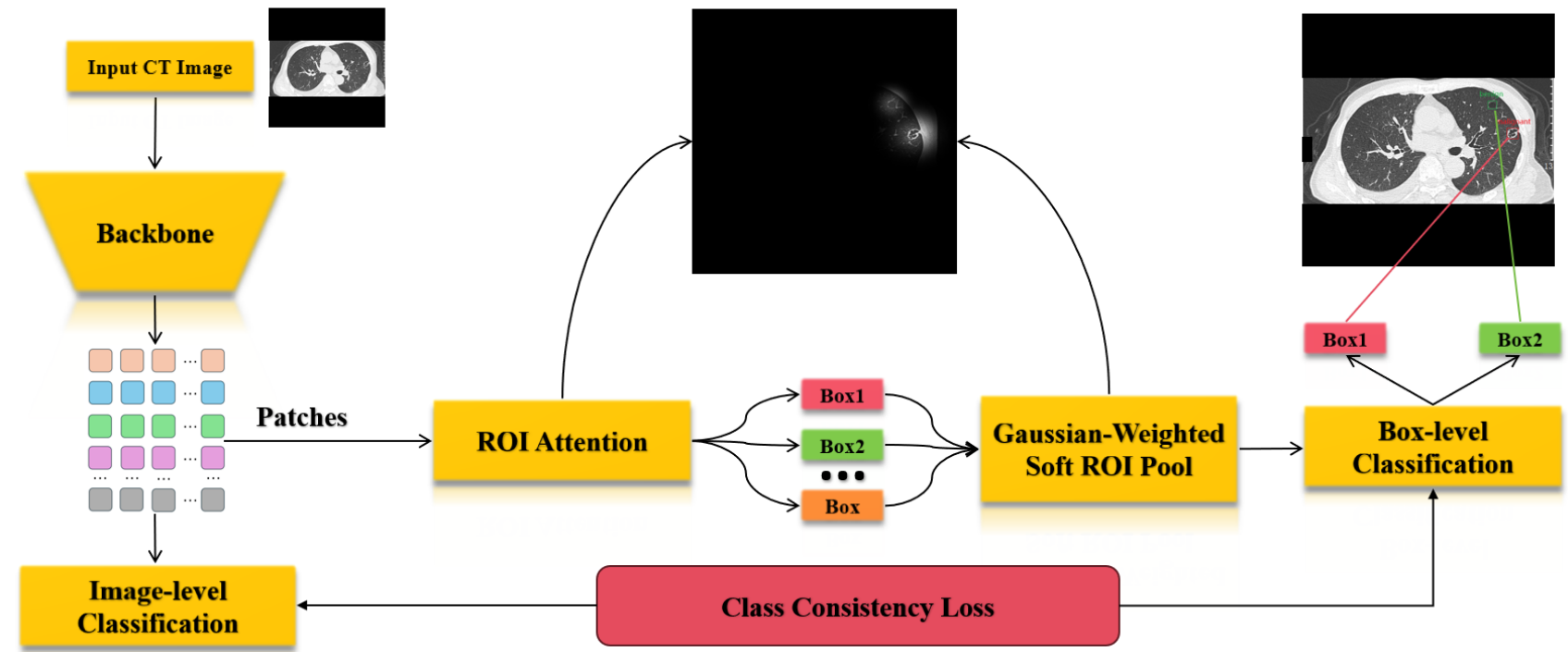


图3.4　MIL双分支结构图

具体而言，设输入图像经骨干网络与特征增强模块后得到二维特征图 。为引入多实例建模，本文将 在空间维度上按固定网格进行划分，得到一组局部子区域，每个子区域对应一个局部特征向量，视为一个“实例”（即一个 patch）。进一步地，为与检测分支的多候选建模形式保持一致，本文将这些实例组织为 ：其中 表示第 个候选集合（或候选组）， 表示该集合中的第 个局部实例。这样，一幅图像即可被表示为多个实例集合构成的“bag”，从而满足 MIL 的建模前提。

**（1）实例级分类建模与“无发现”类别**

在 MIL 框架下，每个局部实例 通过轻量级 MLP 输出各类别的 logit 预测 。考虑到医学图像中大量实例并不包含病灶信息，若强制其归入某一病灶类别会显著增加误检风险，本文额外引入“无发现”类别 来显式刻画背景实例。相应地，实例级概率定义为：

(3-6)

(3-7)

其中 为 Sigmoid 函数， 为目标类别集合。上述形式本质上将“是否存在有效发现”作为门控条件，使类别预测仅在非 的概率质量上进行分配，从而降低背景噪声对类别判别的干扰。

**（2）基于 LSE 的图像级聚合**

在得到所有实例的类别概率后，本文采用 Log-Sum-Exp（LSE）池化[43]将实例级预测聚合为图像级输出。对任意类别 ，图像级预测定义为：

(3-8)

LSE 聚合在形式上兼顾最大池化与平均池化的特性：既能强调少量高置信度实例对整体判别的贡献，又能避免完全忽略低响应实例所包含的上下文信息，因而更适合病灶稀疏分布的医学影像多实例场景。

通过上述“特征图划分实例—实例级判别—图像级聚合”的 MIL 流程，模型能够从局部区域中自动挖掘与病灶相关的判别线索，并形成稳定的图像级预测结果。

### ROI 区域获取与高斯加权软池化

在本文框架中，每个预测框都会被视为一个潜在包含诊断意义结构的感兴趣区域（ROI）。需要强调的是，这些 ROI 不是外部固定候选框，而是由检测分支在推理过程中动态生成，因此能够随输入自适应地聚焦于更可能包含病灶信息的区域。为了对这些 ROI 的区域表征进行统一的提取、细化与利用，本文在检测输出的基础上引入以下两步关键设计：ROI 注意力特征提取与高斯加权软区域池化，如图3.5所示。

**（1）ROI 注意力特征提取与候选框预测**

在完成多尺度增强后，来自不同尺度的特征（如 ）被送入 RT-DETR 的编码器与解码器进行集合预测。RT-DETR 解码器输出的一组查询中，第 个查询对应一个潜在 ROI。本文直接利用这组查询 ，将其作为“区域选择”的条件信息：对每个查询 ，ROI Attention 模块以 为输入，并在不改变原有解码流程的前提下得到该查询对应的初步 ROI 特征 。



图3.5　恶性结节软高斯ROI池化示意图

在几何信息方面，对每个查询 ，直接利用 RT-DETR 的边界框回归分支输出，用于预测该 ROI 对应的框中心与尺度参数。该预测为后续的软池化提供了“中心位置”与“空间扩散范围”的几何先验，使得区域聚合既可保持语义一致性，也能在不同输入尺度下具有可比性。

**（2）高斯加权软 ROI 池化（GWSRP）**

传统的硬裁剪/硬池化通常只保留预测框内部特征，并将框外上下文完全丢弃；当预测框存在轻微偏移或边界模糊时，容易引入边界伪影、空间细节损失，并对框误差高度敏感。为缓解该问题，本文采用高斯加权软 ROI 池化：将每个 ROI 看作以预测框中心为均值的二维高斯感受野，中心区域权重大、边缘区域权重小，从而在“聚焦 ROI”与“保留必要上下文”之间取得折中。

具体而言，对第 个 ROI，在空间位置 处定义软注意力权重场为：

(3-9)

其中， 为空间尺寸， 与 表示位置 相对 ROI 中心的归一化距离：

(3-10)

表示 ROI 中心位置， 控制高斯核在两个方向的扩散范围（由检测分支预测的尺度相关参数给出）。在得到权重场 后，对 ROI 特征进行逐元素加权，得到最终用于分类/判别的 ROI 表征：  
 (3-11)

其中 为逐元素相乘。与硬池化相比，该机制能够在候选框存在轻微误差时仍保持区域响应的平滑性与鲁棒性，同时避免边界处信息突变；并且由于权重以中心为主导，模型对小尺度结节的关注更集中（图3.5展示了该特性）。

### 损失函数

本文的优化目标建立在 RT-DETR 的原始训练目标之上，记为 。该项由查询（query）层面的分类损失、Smooth 边界框回归损失以及用于框对齐的 IoU 损失构成，使得模型保持较强的检测能力。为了进一步利用多实例学习（MIL）框架在图像级判别中的优势，本文在此基础上额外引入两项约束：1）图像级二元交叉熵损失 ，用于监督 patch 分支聚合后的图像级预测，从而提升良恶性等图像级分类的稳定性；2）跨分支一致性损失 ，用于约束 patch 分支的预测与 ROI 分支“映射到 patch 空间”的预测保持一致，强化两分支的信息协同。

**（1）总体损失**

(3-12)

其中， 与 为平衡系数（你论文原文设为 1.0），分别控制图像级监督项与跨分支一致性项对总损失的贡献。

**（2）ROI分支到Patch分支 的映射概率**

跨分支一致性需要将 ROI 分支的预测“对齐”到 patch 网格空间。设 ROI 分支对第 个 ROI 的类别预测为 ，并且该 ROI 对 patch 网格位置 的注意力权重为 。则 ROI到patch 的映射概率定义为逐元素加权：

(3-13)

这里的含义是：ROI 分支给出的类别置信度 在空间上由注意力权重 进行“分配/投影”，从而得到对应 patch 位置的类别概率形式。

**（3）跨分支一致性损失**

patch 分支在每个 patch 位置 上对类别 的预测概率记为 。为了使 patch 分支与 ROI 分支投影后的结果一致，本文采用 KL 散度衡量二者差异，并对所有 patch 位置与类别求和/平均：

(3-14)

其中 为 Kullback–Leibler 散度， 为类别集合， 分别为 patch 网格的高与宽（即特征图被划分后的行列数）。该项约束的直观意义是：如果某些 patch 在 patch 分支中倾向于某类，那么 ROI 分支在对应空间位置投影出来的概率也应与其一致，从而减少分支间预测冲突、提升整体判别鲁棒性。

## 实验细节

### 实验环境设置

为保证实验结果的可复现性与公平性，本文所有实验均在统一的软硬件环境下完成。实验平台采用 Ubuntu 20.04.6 操作系统，深度学习框架为 PyTorch 2.7.0，并基于 CUDA 加速进行模型训练与推理。硬件配置方面，实验在配备 NVIDIA RTX 4090 显卡（显存 24 GB）、64 GB 内存以及 Intel® Core™ i9-12900K 处理器的工作站上完成，该配置能够满足实时目标检测模型在大规模 CT 图像数据上的训练与推理需求。

在模型训练阶段，优化器选用 AdamW 优化算法，其在保持较快收敛速度的同时能够有效抑制过拟合问题[44]。初始学习率设置为 ，权重衰减系数同样设为 ，并采用梯度裁剪策略，将梯度范数限制为 1.0，以增强训练过程的数值稳定性。批量大小设置为 8，模型共训练200个 epoch，以充分学习肺部 CT 图像中的判别性特征。

在数据预处理方面，所有输入图像统一缩放至 分辨率，并基于训练集的均值与标准差进行归一化处理，以减小不同扫描条件带来的分布差异。同时，为提升模型的泛化能力，训练过程中引入了基于 Albumentations 库的数据增强策略[45]，包括颜色扰动以及一定概率的高斯模糊操作（概率为 30%），以模拟实际临床场景中成像条件的多样性。

上述实验环境与参数配置在所有对比方法中保持一致，从而确保不同模型之间的性能对比具有公平性与可信度。

### 数据集介绍

为全面评估所提出模型在囊腔性肺癌检测与分类任务中的有效性与泛化能力，本文选取了两个具有代表性的胸部 CT 数据集进行实验验证，分别为公开数据集 LUNA16 以及由临床专家构建的囊腔性肺癌专用数据集。两个数据集在病灶形态、标注方式及临床复杂度方面具有明显差异，能够从不同层面检验模型在真实应用场景中的适用性。

**（1）LUNA16 公共肺结节数据集**

LUNA16 数据集来源于 LIDC-IDRI 项目，是当前肺结节检测领域广泛采用的标准评测数据集之一[46]。该数据集包含 888 例胸部 CT 扫描，共计 1,186 个由多名放射科专家一致确认的肺结节标注。所有 CT 图像均为轴向切片，结节尺寸分布具有较大跨度，涵盖了从微小结节到中等尺寸结节的多种情况。

在实验设置中，本文遵循主流划分策略，将 LUNA16 数据集按 8:1:1 的比例划分为训练集、验证集和测试集，以保证实验结果的可比性与公平性。该数据集主要用于评估模型在常规肺结节检测场景下的基础性能，为后续在更复杂病灶类型上的验证提供参考基线。

**（2）囊腔型肺癌 CT 数据集**

除公共数据集外，本文进一步引入了一个更具临床挑战性的囊腔性肺癌 CT 数据集。该数据集由中国浙江省宁波市医疗中心李惠利医院，地理坐标为北纬 29.51°，东经 121.33°。团队整理并标注，所有病例均经病理或长期随访结果确认，具有较高的医学可靠性。数据集共包含 342 张轴向 CT 图像，按病灶性质分为恶性和良性两类，其中重点关注囊腔性肺癌这一具有薄壁结构、小尺度、低对比度等特点的病灶类型。

由于囊腔性肺癌在 CT 图像中常仅占据极少像素，且与正常组织或良性病变在外观上高度相似，该数据集在目标定位与分类层面均具有较高难度。实验中，该数据集同样按 80% 用于训练、10% 用于验证、10% 用于测试的比例进行划分，以模拟真实临床筛查场景下模型的应用条件。

为减少不同成像设备与扫描参数带来的影响，所有 CT 图像在输入模型前均进行了统一的预处理操作，包括插值缩放至固定分辨率以及强度归一化处理，以保证模型训练与推理过程的稳定性。

综合来看，LUNA16 数据集提供了标准化、规模化的评测基础，而囊腔性肺癌数据集则进一步考察模型在复杂、临床真实场景下的检测与分类能力，两者的结合能够更全面地验证所提方法的有效性与鲁棒性。

表3.1　数据集构成

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据集名称 | 类别 | 图像数量 | 病例数量 | 数据类型 |
| LUNA16 | 肺结节 | 1186 | 888 | CT |
| 囊腔型肺癌数据集 | 恶性肺结节 | 70 | 70 | CT |
| 良性肺结节 | 272 | 106 | CT |

参考文献

1. Prabhu S, Prasad K, Robles-Kelly A, et al. AI-based carcinoma detection and classification using histopathological images: A systematic review[J]. Computers in Biology and Medicine, 2022, 142: 105209.
2. Opoka L, Szturmowicz M, Osińska K, et al. CT imaging features of thin-walled cavitary squamous cell lung cancer[J]. Advances in Respiratory Medicine, 2019, 87(2): 114–117.
3. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis[J]. Annual Review of Biomedical Engineering, 2017, 19(1): 221–248.
4. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998–6008..
5. Chen J, et al. TransUNet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
6. Dosovitskiy A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
7. 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]. Medical Image Analysis, 2017, 42: 1–13.
8. Lin T Y, et al. Microsoft COCO: Common objects in context[C]//Computer Vision–ECCV 2014. Cham: Springer, 2014.
9. Redmon J, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
10. Zhao Y, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024.
11. Chen J, et al. Run, don’t walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023.
12. Woodring J H, Fried A M, Chuang V P. Solitary cavities of the lung: diagnostic implications of cavity wall thickness[J]. American Journal of Roentgenology, 1980, 135(6): 1269–1271.
13. Womack N A, Graham E A. Epithelial metaplasia in congenital cystic disease of the lung: Its possible relation to carcinoma of the bronchus[J]. The American Journal of Pathology, 1941, 17(5): 645.
14. Tian Y, Ye Q, Doermann D. YOLOv12: Attention-centric real-time object detectors[J]. arXiv preprint arXiv:2502.12524, 2025.
15. Tang M, Djelouah A, Perazzi F, Boykov Y, Schroers C. Normalized cut loss for weakly-supervised CNN segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 1818–1827.
16. Wang A, Xu M, Zhang Y, Islam M, Ren H. S2ME: Spatial-spectral mutual teaching and ensemble learning for scribble-supervised polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention (MICCAI). 2023: 35–45.
17. Wei J, Hu Y, Cui S, Zhou S K, Li Z. WeakPolyp: You only look bounding box for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention (MICCAI). 2023: 757–766.
18. ZMilletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). 2016: 565–571.
19. Carion N, Massa F, Synnaeve G, et al. End-to-End Object Detection with Transformers[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2020: 213-229.
20. Lv W, Xu S, Wang Y, et al. RT-DETR: Real-Time Detection Transformer[EB/OL]. arXiv:2304.08069, 2023.
21. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.
22. Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2117-2125.
23. Tan M, Pang R, Le Q V. EfficientDet: Scalable and Efficient Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10781-10790.
24. Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 8759-8768.
25. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]//Advances in Neural Information Processing Systems (NeurIPS). 2017: 5998-6008.
26. Carion N, Massa F, Synnaeve G, et al. End-to-End Object Detection with Transformers[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2020: 213-229.
27. Zhu X, Su W, Lu L, et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection[C]//Proceedings of the International Conference on Learning Representations (ICLR). 2021..
28. Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen. DETRs Beat YOLOs on Real-time Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024.
29. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
30. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015: 234-241.
31. He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2017: 2961-2969.
32. Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//*Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
33. Sohn K, Berthelot D, Li C L, et al. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence[C]//*Advances in Neural Information Processing Systems (NeurIPS)*. 2020.
34. Oppenheim A V, Schafer R W. *Discrete-Time Signal Processing*[M]. 3rd ed. Upper Saddle River: Pearson, 2010.
35. Ahmed N, Natarajan T, Rao K R. Discrete cosine transform[J]. *IEEE Transactions on Computers*, 1974, C-23(1): 90-93.
36. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way[M]. 3rd ed. Amsterdam: Elsevier/Academic Press, 2009.
37. Daubechies I. *Ten Lectures on Wavelets*[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1992.
38. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015.
39. Isensee F, Petersen J, Klein A, et al. nnU-Net: A Self-configuring Method for Deep Learning-based Biomedical Image Segmentation[J]. Nature Methods. 2021.
40. Liu G, Reda F A, Shih K J, et al. Image Inpainting for Irregular Holes Using Partial Convolutions[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 85–100.
41. Li Y, Wang Y, Zhang B, et al. Efficient Attention Mechanisms for Visual Recognition: A Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
42. Ali A, Saenko K. Confidence-rated multiple instance boosting for object detection[C]//Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014: 1417-1424.
43. Pinheiro P O, Collobert R. From Image-level to Pixel-level Labeling with Convolutional Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 1713–1721
44. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization[C]//Proceedings of the International Conference on Learning Representations (ICLR).2019.
45. Buslaev A, Iglovikov V I, Khvedchenya E, et al. Albumentations: Fast and flexible image augmentations[J]. Information, 2020, 11(2): 125.
46. 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]. Medical Image Analysis, 2017, 42: 1–13.
47. 刘裕国, 杨柳, 张洋, 等. 雾锺来袭,如何突围[N/OL]. 人民日报, 2013-01-12[2013-11-06]. http://papcr.pco.ecom.cn/rmrb/htm1/2013-01/12/nw.D110000 renmrb \_ 20130112 \_2-04.htm.

（示例12：档案）

（著录格式：［序号］ 主要责任者. 题名: 档案编号[A]. (更新或修改日期)[引用日期]. 出版地: 出版社, 出版年: 起止页码.）

1. 北京市人民政府办公厅. 关于转发北京市企业投资项目核准暂行实施办法的通知: 京政办发[2005]37号[A/OL]. (2005-07-12) [2011-07-12]. http://china.findlaw.cn/fagui/p\_1/ 39934.html.

（示例13：标准）

（著录格式：［序号］ 主要责任者. 题名: 标准名称: 标准号[S]. 出版地: 出版社, 出版年: 起止页码.）

1. 全国信息与文献标准化技术委员会. 文献著录: 第4部分 非书资料: GB/T 3792.4—2009[S]. 北京: 中国标准出版社, 2010: 3.
2. 国家环境保护局科技标准司. 土壤环境质量标准: GB 15616—1995[S/OL]. 北京: 中国标准出版社, 1996: 2-3[2013-10-14]. http://www.mee.gov.cn/ywgz/ fgbz/bz/bzwb/trhj/199603/ W020070313485587994018.p