

中图分类号: TP391

单位代号: 10280

密 级: 公开

学 号: 21721570

# 上海大学



## 硕士学位论文

SHANGHAI UNIVERSITY  
MASTER'S DISSERTATION

题 目	深度学习在动态骨显像和 PET/CT 中的骨科相关感染诊断 研究
-----	--

作 者 聂良兵

学科专业 计算机应用技术

导 师 李成范

完成日期 二〇二四年三月



姓    名：聂良兵

学号：21721570

论文题目：深度学习在动态骨显像和 PET/CT 中的骨科相关感染诊断研究

# 上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主    席：

委    员：

导    师：

答辩日期：        年    月    日



姓 名：聂良兵

学号：21721570

论文题目：深度学习在动态骨显像和 PET/CT 中的骨科相关感染诊断研究

## 上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

日期： 年 月 日

## 上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

学位论文作者签名：

导师签名：

日期： 年 月 日

日期： 年 月 日



## 上海大学工学硕士学位论文

# 深度学习在动态骨显像和 PET/CT 中 的骨科相关感染诊断研究

作 者: 聂良兵

导 师: 李成范

学科专业: 计算机应用技术

计算机工程与科学学院

上海大学

2024 年 3 月



A Dissertation Submitted to Shanghai University for the  
Degree of Master in Engineering

# Deep Learning in the Diagnosis of Orthopedic Related Infections Using Dynamic Bone Scintigraphy and PET/CT

Candidate: Liangbing Nie

Supervisor: Chengfan Li

Major: Computer Applications Technology

Schoool of Computer Engineering and Science  
Shanghai University  
March, 2024



## 摘要

骨科相关感染（假体关节感染和骨折相关感染）会导致慢性疼痛和其他造成患者不适的症状，对人类的身心健康危害的同时还需要高昂的医疗费用。因此，对患者的及时评估与准确诊断是治疗、预防并发症和良好预后的关键。核医学影像是骨科相关感染中常见的诊断手段之一。动态骨显像用于诊断假体关节感染，PET/CT 则用于诊断骨折相关感染。医生在动态骨显像中观察每张图像中的组织摄取形态和不同图像之间时序上的生理代谢差异这些影像学特征作为依据来进行评估与诊断。在 PET/CT 中则通过观察软组织的损伤程度、骨骼的断裂程度、以及高摄取区域呈现的形式等多方面的影像学特征，进行综合分析并作出最终的诊断决策。

随着人工智能的发展，越来越多计算机视觉的深度学习方法以其优异的性能、较高的准确率和较短的推理时间等特点被逐步应用于医学影像之中。然而，对于骨科相关感染，目前还没有基于深度学习的全自动诊断方法的相关研究。采用深度学习来实现该方法时，需要去研究解决一系列问题。例如数据集的缺乏、动态骨显像存在生理高摄取区域的干扰、具有时序性的动态骨显像包含三种相、PET/CT 中病灶形态和位置多样且只占据整个影像中一小部分以及 PET 和 CT 均为三维影像等等。因此，鉴于核医学影像及相关疾病的特征，可以将骨科相关感染的自动诊断方法视为医学影像的计算机辅助诊断领域中的检测与分类任务。针对上述问题，本文的主要研究内容总结如下：

(1) 数据集的构建与预处理方法的设计：对于骨科相关感染数据集的缺乏，本文作者与上海市第六人民医院核医学科合作，共同构建了两个关键数据集。其中，动态骨显像数据集纳入了从 2016 年 1 月至 2021 年 6 月之间 455 名患者的影像数据和相应的完整诊断报告，而 PET/CT 影像数据集纳入了 2016 年 11 月到 2021 年 12 月之间 281 名患者的相关数据。在医学专家的指导下，对数据集进行了详细的标注工作，以确保数据的准确性和可靠性。其次，针对动态骨显像和 PET/CT 影像，分别设计了有效数据预处理方法，用于优化图像质量、标准化数据格式，以及降低无关区域和生理高摄取区域对模型性能的不利影响。

(2) 基于  $^{99m}\text{Tc}$ -MDP 动态骨显像的假体关节感染辅助诊断框架：针对动态骨显

像中假体关节感染诊断任务，构建了一种辅助诊断框架，由数据预处理和分类模型 DBS-eNet 组成。DBS-eNet 结合了三维卷积和 ConvLSTM，可以更好地利用动态骨显像中不同相以及相中的时序性图像序列。其中，三维卷积在时序上提取不同相下每一张图像的生理摄取形态特征，而 ConvLSTM 捕捉和丰富单一图像生理形态摄取特征的同时还捕获图像序列中生理代谢的差异变化特征。实验结果表明，对于假体膝关节感染和假体髋关节感染的诊断任务，该框架在五折交叉验证中优于其他卷积神经网络，准确率分别达到了 86.48% 和 86.33%；在独立验证中与核医学专家们比较，准确率在膝关节上提升了 13.55%，髋关节上保持一致。

(3) 基于  $^{18}\text{F}$ -FDG PET/CT 三维影像的下肢骨折相关感染检测与诊断框架：针对 PET/CT 影像中下肢骨折相关感染诊断任务，提出了由病灶检测网络和病灶诊断分类网络构成的两阶段框架 3DFRINet。病灶检测网络通过双分支设计和注意力模块可以结合两个模态的多尺度细节特征，从而准确地检测出病灶区域。病灶诊断分类网络通过最大强度投影将立体影像转换为平面图像，获取有效特征的同时对候选区域进一步挖掘病灶的代谢形态特征，实现对骨折相关感染的准确分类。实验结果表明，该框架在下肢 PET/CT 影像中病灶检测性能优于其他 YOLO 模型， $\text{AP}_{25}$  达到了 0.9234。在骨折相关感染病灶的诊断分类中，相比较于初级核医学医生准确率提升了 11.27%，比高级核医学医生则提升了 4.23%。

(4) 骨科相关感染的自动诊断与可视化方法：为了完善假体关节感染和骨折相关感染的自动化辅助诊断流程，本文采用了 PyQt6 框架，设计了一个骨科相关感染的自动诊断与可视化方法，用以提供全面而便捷的可视化浏览和分析功能。该方法整合了(2) 和 (3) 的框架，支持假体关节感染的辅助诊断和骨折相关感染的辅助检测诊断。此外，该方法支持二维、三维和 PET/CT 融合可视化，以满足不同医学影像模态的需求。该方法在可视化方面提供多种影像分析功能，包括坐标与影像值的展示、切换视图、旋转和缩放等等。

**关键词：**医学影像，骨科相关感染，图像分类，目标检测，可视化

## ABSTRACT

Orthopedic related infection (prosthetic joint infection and fracture related infection) causes chronic pain and other symptoms that cause discomfort to the patient, posing a risk to human physical and mental health along with high medical costs. Therefore, timely evaluation and accurate diagnosis of patients are essential for treatment, prevention of complications and good prognosis. Nuclear medicine imaging is one of the common diagnostic tools in orthopedic related infections. Dynamic bone scintigraphy is used to diagnose prosthetic joint infection, while PET/CT is used to diagnose fracture related infection. In dynamic bone scintigraphy, the physician observes the tissue uptake pattern in each image and the physiological and metabolic differences in the time series between images as the basis for evaluation and diagnosis. In PET/CT, various imaging features such as the extent of soft tissue damage, the degree of fracture of the bone, and the form of the uptake area are observed and analyzed to make a final diagnostic decision.

With the development of artificial intelligence, more and more deep learning methods for computer vision have been gradually applied to medical imaging with their excellent performance, high accuracy and short inference time. However, for orthopedic related infection, there is no research related to a fully automated diagnostic method based on deep learning. When deep learning is used to implement this method, a series of problems need to be researched and solved. For example, there is a lack of datasets, dynamic bone scintigraphy has interference from physiological high uptake regions, dynamic bone scintigraphy with temporal sequencing contains three phases, lesion morphology and location in PET/CT are diverse and occupy only a small portion of the whole image, and both PET and CT are three-dimensional images, and so on. Therefore, considering the characteristics of nuclear medicine images and related diseases, automatic diagnostic methods for orthopedic related infection can be regarded as a detection and classification task in the field of computer aided diagnosis of medical images. To address the above issues, the main research of this paper is summarized as follows:

(1) Construction of datasets and design of the preprocessing method: For the lack of orthopedic related infection datasets, the author collaborated with the Nuclear Medicine Department of Shanghai Sixth People's Hospital to construct two key datasets. Among them, the dynamic bone scintigraphy dataset incorporated the imaging and corresponding complete diagnostic reports of 455 patients between January 2016 and June 2021, while the PET/CT imaging dataset incorporated the relevant data of 281 patients between November 2016 and December 2021. Detailed labelling work was carried out on the datasets under the guidance of medical experts to ensure the accuracy and reliability of the data. Next, effective data preprocessing methods were designed for dynamic bone scintigraphy and PET/CT images, respectively, for optimizing the image quality, standardizing the data format, and reducing the detrimental effects of irrelevant and physiologically high uptake regions on the model performance.

(2) An assisted diagnostic framework for prosthetic joint infection based on  $^{99m}\text{Tc}$ -MDP dynamic bone scintigraphy: For the diagnostic task of prosthetic joint infection in dynamic bone scintigraphy, an assisted diagnostic framework is constructed, consisting of data pre-processing and classification model, DBS-eNet. DBS-eNet combines 3D convolution and ConvLSTM, which can better utilize temporal image sequences in different phases as well as in-phase from dynamic bone scintigraphy. Among them, 3D convolution extracts the physiological uptake morphological features of each image under different phases on the temporal sequence, while ConvLSTM captures and enriches the physiological morphological uptake features of a single image while also capturing the features of differential physiological metabolism changes in the image sequence. The experimental results show that for the diagnostic tasks of prosthetic knee infection and prosthetic hip infection, the framework outperforms other convolutional neural networks in five-fold cross-validation, with accuracies of 86.48% and 86.33%, respectively, and in independent validation comparing with nuclear medicine experts, the accuracies are improved by 13.55% on the knee, and remain consistent on the hip.

(3) A framework for the detection and diagnosis of lower extremity fracture related infection based on  $^{18}\text{F}$ -FDG PET/CT three-dimensional imaging: For the diagnosis task of lower extremity fracture related infection in PET/CT images, a two-stage framework

3DFRINet consisting of a lesion detection network and a lesion diagnostic classification network is proposed. The lesion detection network can accurately detect lesion regions by combining multi-scale detailed features of both modalities through the dual-branching design and the attention module. The lesion diagnostic classification network converts the cubic image into the planar image through maximum intensity projection, acquires effective features while further mining the metabolic-morphological features of the lesion for the candidate region, and achieves accurate classification of fracture related infection. The experimental results show that the framework outperforms other YOLO models in lesion detection in lower limb PET/CT images, with an AP<sub>25</sub> of 0.9234. In the diagnostic classification of fracture related infection lesions, the accuracy of the framework is improved by 11.27% in comparison with junior nuclear medicine physicians, and 4.23% compared with senior nuclear medicine physicians.

(4) The automated diagnostic and visualization method for orthopedic related infection: To improve the automated assisted diagnosis process for prosthetic joint infection and fracture related infection, this paper adopts the PyQt6 framework to design an automated diagnosis and visualization method for orthopedic related infection, which is used to provide comprehensive and convenient visualization browsing and analysis functions. The method integrates the frameworks of (2) and (3) to support assisted diagnosis of prosthetic joint infection and assisted detection and diagnosis of fracture related infection. In addition, the method supports 2D, 3D and PET/CT fusion visualization to meet the needs of different medical imaging modalities. The method provides a variety of image analysis functions in terms of visualization, including the display of coordinates and image values, switching views, rotation and scaling, etc.

**Keywords:** Medical imaging; Orthopedic related infection; Image classification; Object detection; Visualization



# 目 录

摘要 .....	I
ABSTRACT .....	III
插图索引 .....	XI
表格索引 .....	XIII
第一章 绪论 .....	1
1.1 课题来源 .....	1
1.2 研究背景与意义 .....	1
1.3 研究内容与创新点 .....	5
1.4 本文的整体结构 .....	7
第二章 国内外研究现状 .....	9
2.1 图像分类 .....	9
2.1.1 图像分类算法 .....	10
2.1.2 医学影像分类算法 .....	17
2.2 图像检测 .....	19
2.2.1 图像检测算法 .....	19
2.2.2 医学影像检测算法 .....	24
2.3 本章小结 .....	26
第三章 基于 $^{99m}\text{Tc}$ -MDP 动态骨显像的假体关节感染辅助诊断框架 .....	27
3.1 问题分析 .....	27
3.2 数据预处理 .....	30
3.2.1 数据清洗与标注 .....	31
3.2.2 预处理算法 .....	32
3.2.3 数据增强 .....	35
3.3 DBS-eNet .....	35
3.3.1 模型结构 .....	35

3.3.2	ConvLSTM .....	36
3.3.3	损失函数 .....	38
3.3.4	可视化模块 .....	39
3.4	实验 .....	39
3.4.1	实验环境 .....	39
3.4.2	数据集 .....	40
3.4.3	评估指标 .....	40
3.4.4	实验结果 .....	41
3.4.5	可视化分析 .....	45
3.5	本章小结 .....	46
<b>第四章</b>	<b>基于<sup>18</sup>F-FDG PET/CT 三维影像的下肢骨折相关感染检测与诊断框架</b> ..	<b>47</b>
4.1	问题分析 .....	47
4.2	数据预处理 .....	50
4.2.1	数据清洗与标注 .....	51
4.2.2	PET/CT 的转换与配准 .....	52
4.2.3	人体区域保留 .....	54
4.2.4	数据增强 .....	56
4.3	框架设计 .....	57
4.3.1	病灶检测网络 322-YOLO .....	59
4.3.2	病灶诊断分类网络 .....	62
4.4	实验 .....	63
4.4.1	实验环境 .....	63
4.4.2	数据集 .....	63
4.4.3	评价指标 .....	64
4.4.4	实验结果 .....	65
4.4.5	可视化分析 .....	68
4.5	本章小结 .....	68
<b>第五章</b>	<b>骨科相关感染的自动诊断与可视化方法</b> .....	<b>71</b>
5.1	问题分析 .....	71

5.2 总体设计 .....	72
5.2.1 界面设计 .....	73
5.3 自动诊断 .....	76
5.3.1 辅助诊断模块 .....	76
5.4 可视化 .....	77
5.4.1 文件解析模块 .....	77
5.4.2 可视化模块 .....	78
5.4.3 影像分析模块 .....	81
5.5 案例分析 .....	82
5.6 本章小结 .....	85
<b>第六章 总结与展望 .....</b>	<b>87</b>
6.1 总结 .....	87
6.2 展望 .....	88
<b>参考文献 .....</b>	<b>89</b>
<b>攻读硕士学位期间取得的研究成果 .....</b>	<b>101</b>
<b>致 谢 .....</b>	<b>103</b>



## 插图索引

图 1.1	$^{99m}\text{Tc}$ -MDP 动态骨显像示意图 .....	3
图 1.2	$^{18}\text{F}$ -FDG PET/CT 示意图 .....	4
图 1.3	本文的研究内容 .....	6
图 1.4	论文的整体结构 .....	8
图 2.1	卷积神经网络示意图 .....	11
图 2.2	不同卷积神经网络的模块示意图 .....	12
图 2.3	Vision Transformer 示意图 .....	14
图 2.4	循环神经网络示意图 .....	15
图 2.5	四个角度最大强度投影图像作为输入进行预测的示意图 .....	18
图 2.6	U-Net 整体结构示意图 .....	20
图 2.7	R-CNN 整体结构示意图 .....	22
图 2.8	YOLO 整体结构示意图 .....	23
图 2.9	融合 PET 和 CT 的目标分割方法 .....	25
图 3.1	髋关节置换手术后的动态骨显像 .....	28
图 3.2	整体研究流程 .....	29
图 3.3	动态骨显像的预处理流程 .....	30
图 3.4	动态骨显像的数据标注 .....	31
图 3.5	三种预处理算法的结果示意图 .....	32
图 3.6	基于深度学习的多特征神经网络模型 DBS-eNet .....	35
图 3.7	ROC 曲线、校准曲线与决策曲线分析 .....	44

图 3.8 DBS-eNet 的 Grad-CAM 热力图 .....	45
图 3.9 DBS-eNet 的 t-SNE 散点图 .....	45
图 4.1 两例患者的 PET/CT .....	48
图 4.2 整体研究流程 .....	50
图 4.3 PET/CT 的数据标注 .....	51
图 4.4 PET 和 CT 在物理空间中所占据的区域 .....	53
图 4.5 人体区域保留算法效果图 .....	54
图 4.6 PET/CT 的数据增强方法 .....	56
图 4.7 两阶段的检测与诊断框架 .....	58
图 4.8 病灶诊断分类网络 .....	62
图 4.9 ROC 曲线、校准曲线和决策曲线分析 .....	68
图 4.10 322-YOLO 的检测结果 .....	69
图 5.1 总体设计 .....	72
图 5.2 系统界面 .....	74
图 5.3 可视化界面 .....	75
图 5.4 辅助诊断模块的运行流程 .....	76
图 5.5 三种颜色映射 .....	79
图 5.6 案例可视化 .....	83
图 5.7 自动诊断方法 .....	84

## 表格索引

表 3.1	实验配置 .....	39
表 3.2	动态骨显像的分布情况 .....	40
表 3.3	髋部动态骨显像上不同预处理算法之间的比较 .....	42
表 3.4	DBS-eNet 在五折交叉验证和独立验证上的性能表现 .....	43
表 3.5	DBS-eNet 与分类网络在五折交叉验证中的比较 .....	43
表 3.6	DBS-eNet 在独立测试集上与专业医生的比较 .....	44
表 4.1	下肢骨折相关感染的 PET/CT 影像数据集特征 .....	64
表 4.2	病灶检测网络与其他 YOLO 的检测性能比较 .....	65
表 4.3	病灶检测网络 322-YOLO 的消融实验 .....	66
表 4.4	有无额外分类器的病灶检测网络的比较 .....	67
表 4.5	3DFRINet 与核医学医生的比较 .....	67



# 第一章 绪论

## 1.1 课题来源

本课题来源于上海市公共卫生重点学科建设项目“基于新技术、新方法的放射卫生检测适宜技术研究”（编号：GWVI-11.1-40）。本文的研究受到上海市第六人民医院核医学科的专家技术支持和数据支持并获得上海市第六人民医院伦理委员会的批准（编号：SH6H-2022-KY-113(K)）。

## 1.2 研究背景与意义

骨骼是一种坚硬的器官<sup>[1]</sup>，构成了大多数脊椎动物骨骼的一部分。骨骼保护身体的其他器官，产生红细胞和白细胞，储存矿物质，为身体提供结构和支持，并使活动成为可能。骨骼有各种形状和大小，有复杂的内部和外部结构<sup>[2]</sup>。它们重量轻，但坚固耐用，具有多种功能。然而，骨骼系统也不免于面临各种疾病和损伤的威胁，其中感染性疾病是其中之一。

骨科相关感染是骨科手术和损伤后常见的并发症，涉及骨骼、关节、软组织等。主要包括骨髓炎、关节炎、滑膜炎、以及骨折相关感染和假体关节感染等。其中，骨折相关感染和假体关节感染是骨科手术中可怕且严重的并发症<sup>[3]</sup>。假体关节感染发生在人工关节（如人工髋关节、人工膝关节）的植入区域，对于新植入假体关节的患者来说有 2%-2.4% 的发病率，对于经历过翻修手术的患者可能达到 20%<sup>[4]</sup>。作为关节置换术后潜在的致命并发症之一<sup>[5-6]</sup>，假体关节感染可引起慢性疼痛、关节功能降低和其他导致患者身体不适的症状<sup>[7]</sup>。假体关节感染常用的诊断方式有实验室检测、血清学、关节液抽取、微生物培养，但这些方式都没有完美的敏感性和特异性<sup>[8-13]</sup>。假体关节感染的治疗方式包括手术清除感染组织、更换假体以及抗生素治疗。

另一方面，骨折相关感染曾称为创伤后骨髓炎，是骨损伤或创伤性手术后最严重的肌肉骨骼并发症之一<sup>[14]</sup>，可能涉及手术创口感染、软组织受损引起的感染等，导致患病率增加、医疗费用昂贵和严重的诊断挑战<sup>[15]</sup>。在骨折手术后，骨折相关感染的发病率为 1%-19%<sup>[16-17]</sup>。随着病情的发展，骨髓炎可伴有软组织感染，脓毒性关节炎伴或不伴皮肤窦道等一系列临床表现<sup>[18]</sup>，可能需要抗生素治疗、手术清创以及骨

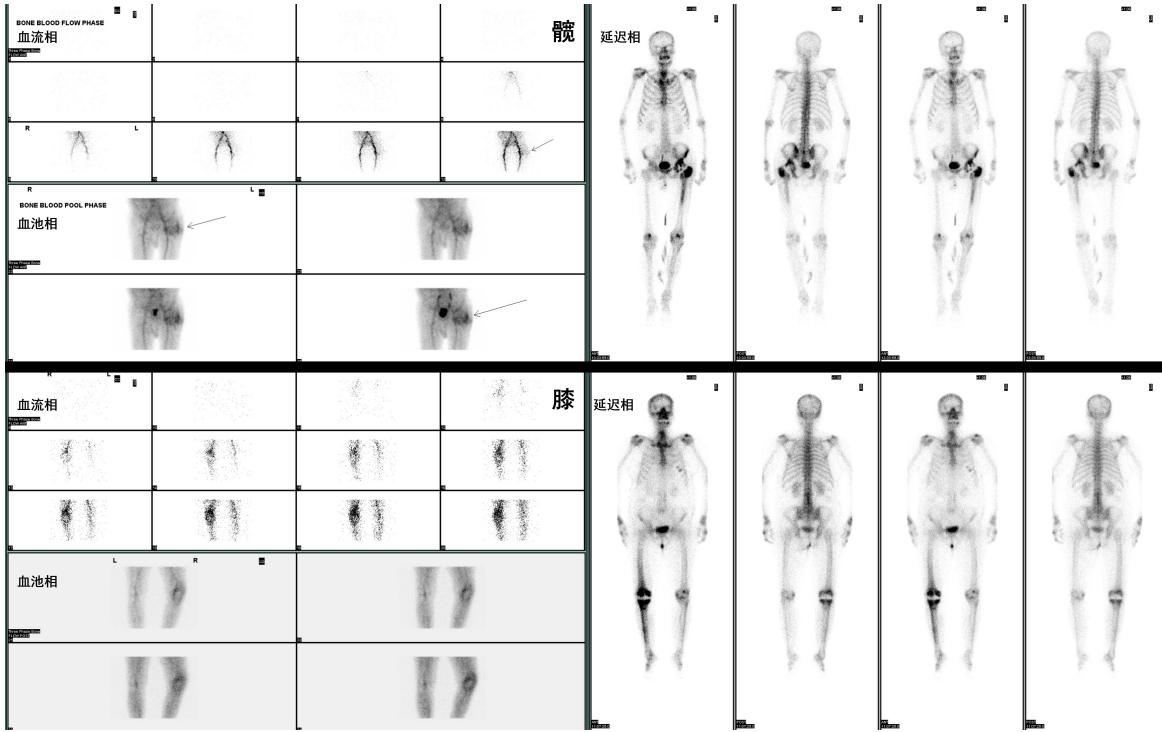
折修复等综合治疗手段。

虽然假体关节感染和骨折相关感染在发病原因和治疗方式上存在差异，但它们共同强调了在骨骼系统健康维护中的重要性。在美国初次全髋和全膝置换手术以及翻修手术的数量从 2012 年的 52,134 增加到 2020 年的 2,224,587 例。其中，全髋置换手术占总手术的 52.1%，全膝置换手术占 34.6%<sup>[19]</sup>。随着关节置换手术的数量逐年增加，美国每年大约有 11,000 个患者确诊假体关节感染<sup>[20]</sup>。仅美国在 2017 年的假体关节感染治疗的估计成本就超过了 10 亿美元<sup>[21]</sup>，而在美国骨折相关感染每年治疗的成本估计约为假体关节感染的四倍<sup>[22]</sup>。早期诊断、科学治疗和术后的有效康复是保障患者骨骼健康的关键，对于降低感染风险、减轻患者痛苦，降低医疗成本，以及维护骨骼功能都至关重要。

作为一种无侵入式的诊断方式，医学成像技术，例如超声、X 射线、计算机断层成像（Computed Tomography, CT）、正电子发射断层显像（Positron Emission Tomography, PET）、单光子发射计算机断层成像术（Single-Photon Emission Computed Tomography, SPECT）、核共振成像（Magnetic Resonance Imaging, MRI）和骨三相（Three-Phase bone Scan, TBS）等等，可以提供人体内部结构和功能信息，被用于疾病的早期检测、诊断和治疗。其中，核医学成像技术通过注射放射性标记的物质，如放射性同位素，来追踪和测量骨髓、关节和周围软组织的代谢活动，从而提供更全面的信息。

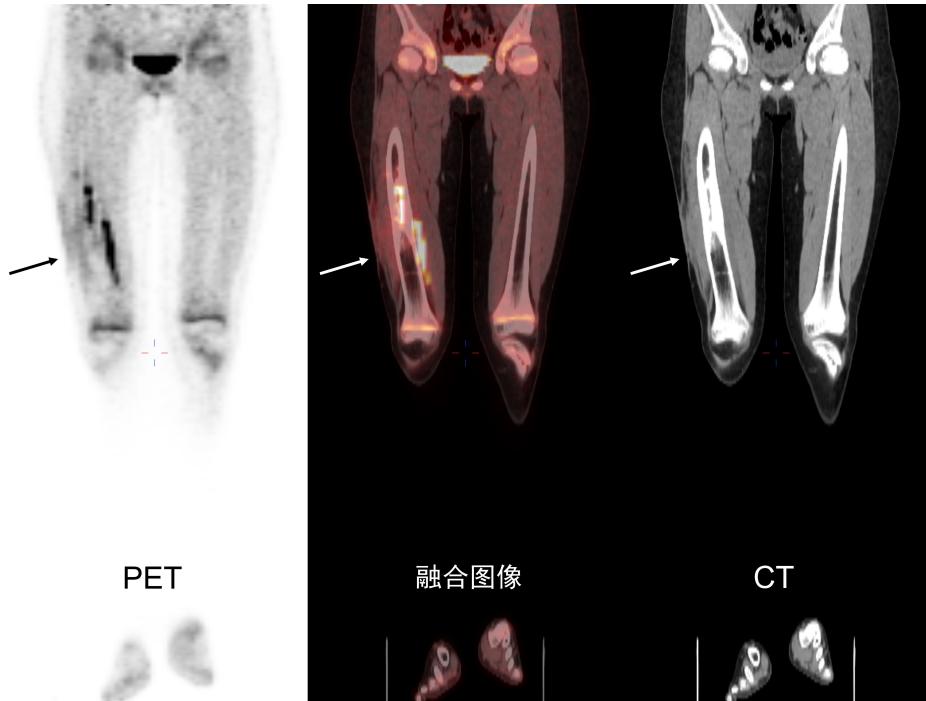
对于假体关节感染，核医学成像可以显示人工关节周围是否存在异常的代谢活动，揭示潜在的感染区域。如图1.1，锝亚甲基二膦酸盐 (<sup>99m</sup>Tc-Methylene Diphosphonate, <sup>99m</sup>Tc-MDP) 动态骨显像，也称之为骨三相，是核医学成像中第一个用于诊断假体关节感染的广泛、可靠且简单的影像模态<sup>[5]</sup>。Verberne 等人<sup>[23]</sup>的研究表明骨显像诊断全髋置换手术后的假体关节感染具有 80% 的灵敏度和 69% 的特异度。Kumar 等人<sup>[24]</sup>发现动态骨显像在评估脓毒性的（或疼痛的）髋关节假体的敏感性、特异性和准确性分别为 81%、86% 和 82%。Zhang 等人<sup>[25]</sup>通过在血流相中设定阈值的方法发现动态骨显像对膝关节假体感染的诊断灵敏度为 94.3%，特异度为 56.0%。根据以上文献和临床实践，动态骨显像诊断假体关节感染通常具有较高的敏感性，但特异性较低，尤其是经历过全膝置换手术的患者。

对于骨折相关感染，核医学成像则能够帮助定位骨折处是否存在感染迹象，有助于及早发现并采取适当的治疗措施。氟代脱氧葡萄糖 (<sup>18</sup>Fluorodeoxyglucose, <sup>18</sup>F-FDG) PET/CT 作为核医学的成像方式之一，常用于骨折相关感染的诊断<sup>[26]</sup>，如图1.2。

图 1.1  $^{99m}\text{Tc}$ -MDP 动态骨显像示意图Figure 1.1 Schematic of  $^{99m}\text{Tc}$ -MDP dynamic bone scintigraphy

它结合了解剖和生理信息，具有更高的空间分辨率和更好的量化可能性<sup>[27]</sup>。Lemans 等人<sup>[28]</sup>通过在  $^{18}\text{F}$ -FDG PET/CT 上量化分析，对骨折相关感染的准确率为 83%，阴性预测值为 0.91。Wanter 等人<sup>[29]</sup>研究发现， $^{18}\text{F}$ -FDG PET/CT 在慢性骨髓炎和骨科植入物相关感染中，阴性预测值为 0.89、准确率为 82%。总体而言， $^{18}\text{F}$ -FDG PET/CT 在诊断骨折相关感染方面显示出令人满意的准确性。

近几年来，人工智能不断地发展，其在医学影像方面的应用也逐渐变成了关注的焦点。越来越多的深度学习方法被应用于医学影像领域的各种任务之中，如影像降噪，疾病分类，器官、组织和病灶的分割与检测、患者的预后效果预测等等<sup>[30-31]</sup>。以卷积神经网络（Convolutional Neural Network, CNN）、循环神经网络（Recurrent Neural Network, RNN）和基于注意力机制的 Transformer 为代表的深度学习方法因其优异的性能和较短的推理时间，经常被用来辅助核医学医生更迅速、准确地诊断病灶区域，减轻工作量。Liu 等人<sup>[32]</sup>（2022）提出了一种创新性多尺度 CNN 用于 PET 中阿尔茨海默病的诊断。Takahashi 等人<sup>[33]</sup>（2022）提出了一个在 PET/CT 中对乳腺癌进行诊断的深度学习模型。然而，目前并没有用于假体关节感染和骨折相关感染这类骨科相关感染的自动化辅助诊断方法。为了有效地构建这种自动化辅助诊断方法，

图 1.2  $^{18}\text{F}$ -FDG PET/CT 示意图Figure 1.2 Schematic of  $^{18}\text{F}$ -FDG PET/CT

有几个问题必须前瞻性地加以考虑：

(1) 数据集缺失。目前在网络上缺乏公开的针对假体关节感染的动态骨显像数据集和针对骨折相关感染的 PET/CT 数据集。此外，医学影像数据集通常规模较小，来源相对单一，模型训练具有挑战性。

(2) 髋关节的动态骨显像中存在生理上的高摄取区域，如膀胱和肾脏，这可能会使神经网络在进行分类时难以学习真正有用的信息并专注于病灶区域。其次，动态骨显像中包含血流相、血池相和延迟相（图1.1），且血流相和血池相中的图像序列具有时序性。因此，需要深入考虑如何充分利用动态骨显像的这些特点。最后，神经网络是一个缺乏解释的黑盒，这使得医生很难理解其预测的依据。

(3) 尽管 CT 可以记录骨骼以及附近软组织的变化，但对骨折相关感染不太敏感。而 PET 中可以显示人体不同区域的代谢活性，且较高的标准摄取值反映了患有骨折相关感染的风险增加，但需要与炎症和愈合区分开来。其次，骨折相关感染的病灶区域仅占 PET/CT 整体影像的一小部分区域，且病灶的形状和位置多样，摄取的形式和程度也有所区别。因此，需要考虑到这些特征复杂性和多样性。最后，PET 和 CT 都是三维的影像数据，这使得可以更全面地了解患者的解剖结构和生理活动。然而，这也给数据处理和模型设计带来了挑战。

(4) 骨科相关感染的自动化辅助诊断方法应以简洁直观、易于操作的形式呈现，尤其对缺乏计算机和人工智能专业知识的用户尤为重要。首先，该方法需要整合骨科相关感染的自动诊断流程，具备智能分析的能力，以协助医生更快速地做出准确的诊断决策。其次，该方法需要提供图形化界面，以方便用户操作。最后，医学影像的可视化是必不可少的功能之一，并且提供一些影像分析功能，满足大多数医学影像模态的需求，以便用户可直观地阅览影像。

由此，本文提出了四个主要的研究目标：

(1) 构建数据集，收集真实患者的假体关节感染或骨折相关感染的影像数据和诊断报告，并进行数据清洗、标注和预处理工作。

(2) 设计一个在动态骨显像上基于深度学习的假体关节感染辅助诊断框架，自动得出诊断结果，并通过类激活图等方法给医生提供决策依据和可信度。

(3) 设计一个在三维 PET/CT 上基于深度学习的骨折相关感染检测与诊断框架，从 PET/CT 中自动定位病灶区域并判断是否存在感染。

(4) 设计一个骨科相关感染的自动诊断与可视化方法，不仅提供智能化功能辅助非专业人员进行假体关节感染和骨折相关感染的自动诊断功能，还能够满足用户简单直观地阅览与分析医学影像。

### 1.3 研究内容与创新点

根据 1.2 小节所述，本文围绕动态骨显像假体关节感染诊断分类任务、PET/CT 骨折相关感染检测诊断任务以及骨科相关感染的自动诊断与可视化方法展开研究，主要研究内容和创新点如图1.3所示，包含以下四个部分：

(1) 数据集的构建与预处理方法的设计：鉴于假体关节感染的动态骨显像数据集和骨折相关感染的 PET/CT 数据集的匮乏，本文作者与上海市第六人民医院核医学科合作，共同构建了两个重要的数据集。动态骨显像数据集纳入了在 2016 年 1 月至 2021 年 6 月期间 455 名患者的影像数据和相应的诊断报告，而 PET/CT 影像数据集则包含了 2016 年 11 月到 2021 年 12 月期间的 281 名患者的影像数据和诊断报告。在医学专家的指导下，通过仔细阅览诊断报告和使用标注工具 ITK-SNAP 对数据集进行了详尽的标注工作，以确保数据的精准性和可靠性。针对动态骨显像和 PET/CT 影像，本文分别构建了高效的数据预处理方法。其中，动态骨显像采用了三种预处

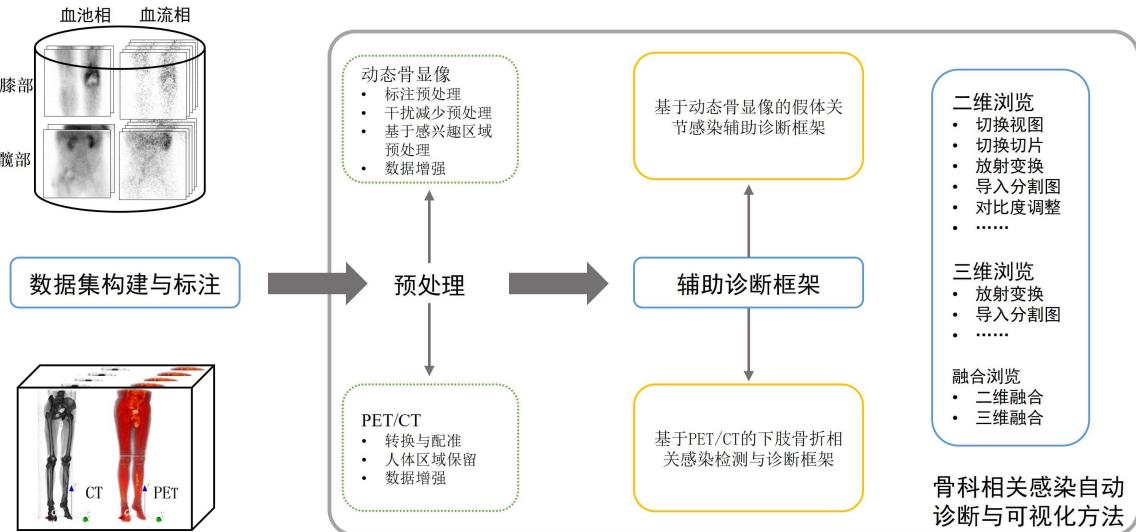


图 1.3 本文的研究内容

Figure 1.3 The research content of this paper

理算法和数据增强手段，而 PET/CT 影像则包括转换与配准、人体区域保留以及数据增强。这些预处理方法优化图像质量、标准化数据格式，并减缓无关区域和生理高摄取区域对模型性能的不良影响。

(2) 基于  $^{99m}\text{Tc}$ -MDP 动态骨显像的假体关节感染辅助诊断框架：为解决动态骨显像中假体关节感染诊断任务，构建了一种辅助诊断分类框架，由数据预处理和分类模型 DBS-eNet 组成。DBS-eNet 结合了三维卷积和 ConvLSTM，可以更有效地利用动态骨显像中不同相和相中的时序性图像序列。三维卷积用于在时序上提取不同相下每一张图像的生理摄取形态特征，ConvLSTM 则捕捉和丰富单一图像的生理形态摄取特征，同时捕获图像序列中生理代谢的差异变化特征。实验结果表明，在五折交叉验证中，该框架在假体膝关节感染和假体髋关节感染的诊断任务上优于其他 CNN，分类准确率分别达到了 86.48% 和 86.33%；在独立验证中，该框架在假体膝关节感染上分类准确率达到了 87.74%，相比核医学专家提升了 13.55%；在假体髋关节感染上分类准确率达到了 86.36%，与核医学专家们一致。统计学分析表明，该框架的诊断性能在假体膝关节感染上优于核医学专家，在假体髋关节上与核医学专家相当。

(3) 基于  $^{18}\text{F}$ -FDG PET/CT 三维影像的下肢骨折相关感染检测与诊断框架：为解决 PET/CT 影像中下肢骨折相关感染诊断任务，本文提出了一种全自动两阶段的检测与诊断框架 3DFRINet。其中，3DFRINet 由病灶检测网络和病灶诊断分类网络构成。

通过双分支设计和注意力模块，病灶检测网络能够捕获和融合不同模态的多尺度细节特征，从而精准地检测出病灶区域。病灶诊断分类网络通过最大强度投影将立体影像转换为平面图像，在获取有效特征的同时，对候选区域进行进一步挖掘，以捕捉骨折相关感染的代谢形态特征，从而实现准确的分类。实验结果表明，3DFRINet 在下肢 PET/CT 影像中病灶检测性能优于其他 YOLO 模型，其 AP<sub>25</sub> 达到了 0.9234。在骨折相关感染病灶诊断分类任务中，该方法上达到了 91.55% 的分类准确率和 0.9250 的 F1，相比较于与初级核医学医生分别提升了 11.27% 和 0.1045，比高级核医学医生分别提升了 4.32% 和 0.0419。统计学分析表明，3DFRINet 具有与初级核医学医生更好或相同的诊断能力，与高级核医学医生相当的诊断性能。

(4) 骨科相关感染的自动诊断与可视化方法：为了完善整个假体关节感染和骨折相关感染的自动化辅助诊断流程，本文采用 PyQt6 框架，设计了一个骨科相关感染的自动诊断与可视化方法。该方法旨在为医生提供一款功能强大的辅助工具，具备全面而便捷的影像阅览和智能分析功能。在智能分析方面，该方法整合了(2)和(3)中的框架，以支持假体关节感染的诊断分类和骨折相关感染的自动检测和分类。在影像阅览方面，该方法支持二维、三维和融合可视化，以满足不同医学影像模态的需求。二维可视化兼容大多数医学影像模态，提供展示视图、坐标和所在坐标影像值的信息，切换视图、几何变换等功能；三维可视化支持 PET 和 CT，提供了旋转、缩放等功能；融合可视化支持 PET 和 CT 融合阅览，提供了二维和三维可视化功能。

## 1.4 本文的整体结构

本文基于作者攻读硕士学位期间的研究方向和参与的课题，分为六个章节，如图1.4所示。各章节的内容安排如下：

第一章详细介绍了假体关节感染和骨折相关感染。接着阐述了深度学习方法与医学影像结合在骨科相关感染自动诊断方面的研究背景和研究意义，并介绍了研究的内容、目标和创新点。

第二章介绍了与本文研究内容相关的近几年国内外相关工作，总结了目前医学影像中分类和检测任务的研究进展，最后阐述了本文中假体关节感染和骨折相关感染诊断任务的特点，分析了任务的特殊性和现有的可行方法。

第三章从问题定义出发，详细介绍了基于<sup>99m</sup>Tc-MDP 动态骨显像的假体关节感

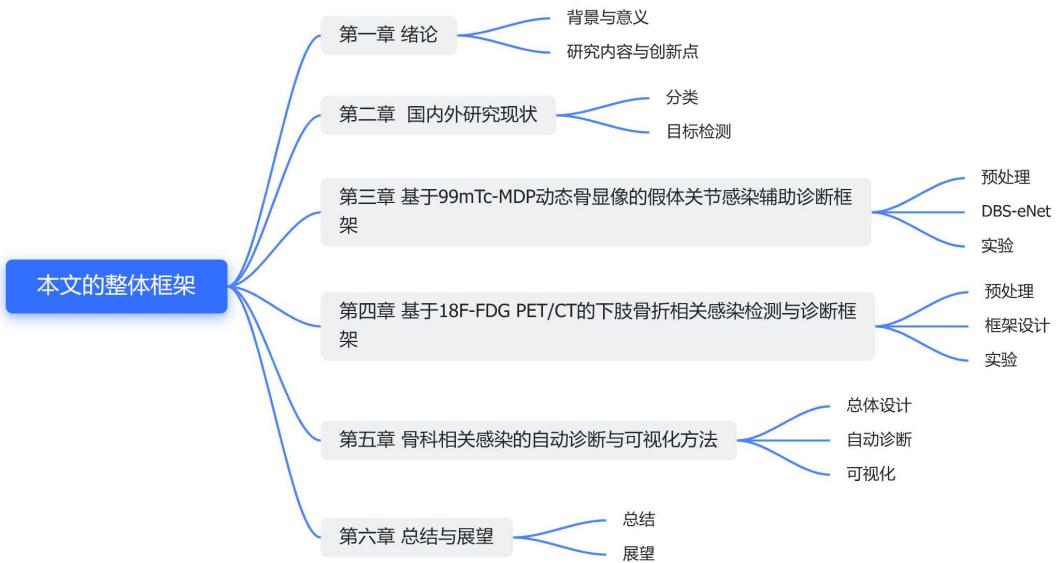


图 1.4 论文的整体结构

Figure 1.4 The overall structure of the paper

染辅助诊断框架。内容包括数据预处理方法、深度学习方法中模型的整体结构、损失函数以及用于展示模型可解释性的可视化方法。接着，通过一系列实验评估了该诊断框架的有效性和较好的诊断性能。

第四章同样从问题定义入手，详细介绍了基于<sup>18</sup>F-FDG PET/CT三维影像的下肢骨折相关感染检测与诊断框架。内容包括数据预处理方法和框架设计。随后，通过一系列实验评估了该框架在病灶检测和诊断方面的良好表现。

第五章详细地介绍了骨科相关感染的自动诊断与可视化方法，拥有智能影像分析与影像阅览功能。

第六章对本文的研究工作进行了总结，并提出了未来可以进一步研究的方向。

## 第二章 国内外研究现状

本章旨在综述国内外关于本文研究内容的相关工作。首先，本章将介绍当前计算机视觉领域中与分类和检测任务相关的神经网络模型。特别地将重点介绍并分析分类任务中的 CNN 系列、Transformer 系列和 RNN 系列模型和在检测任务中的 U-Net 系列目标分割模型、R-CNN 和 YOLO 系列目标检测模型。其次，本章将详细介绍医学影像领域中分类任务与检测任务的最新研究进展，并对本文所涉及的假体关节感染和骨折相关感染诊断任务的特殊性和现有方法进行深入分析与总结。

### 2.1 图像分类

分类任务是指输入数据经过模型计算后分为不同的类别或者标签的任务，这是深度学习中最常见和最基本的问题之一。其中，模型需要从训练数据集中学习数据特征与标签之间的关系，用以对未见过的数据正确地从已知的多个标签集合中选择最合适的一个或几个标签。在计算机视觉领域中，对图像、视频或其他视觉数据的特征提取一直以来都是研究的核心问题。赋予计算机类似于人类一样的智慧视觉能力是多年来许多研究人员共同追求的目标和努力方向。随着相关技术的不断发展，尤其是图像识别技术的不断进步，在计算机视觉领域之中的分类任务在医疗、航海、交通等多个现实场景中有着诸多良好的应用前景，包括人脸识别、航空遥感和医学影像等等。

通常情况下，分类任务的主要流程包括特征提取和模型设计两个关键步骤。合适的特征提取和模型设计可以相辅相成，以实现较优异的性能表现。在早期，图像识别技术<sup>[34]</sup>主要采用了基于先验知识设计的手工特征。原理在于对于不同的应用场景，需要由相应领域的专家提供专业知识去设计所需要的特征。然而，这种针对某一的领域特性和相应的数据类型进行的手工编码，难以去处理海量的数据，效率及其低下<sup>[35]</sup>。此外，人工设计的特征的普遍性不强，图像中可能存在噪声，难以涵盖所有情形，很可能导致分类的结果并不理想。而在图像领域中，常见的形状、纹理、颜色等浅层视觉特征和局部二值模式、方向梯度直方图等局部不变特征，虽然具有一定的普适性，但应用到现实场景具体图像时，针对性不足，尤其是面对一些存在

复杂场景的情形，更加难以去准确使用手工特征去描述目标图像<sup>[36]</sup>。

在模型方面，传统的机器学习模型实现简单，决策依据直观可理解，但对于复杂图像如图像噪声干扰较多，不同类别之间差异小等问题，其性能会大大下降。此外，传统机器学习依赖于特征提取，难以实现全自动化，性能极大受限于手工特征的有效性。随着计算机硬件的快速发展，特别是图形处理器（Graphics Processing Unit, GPU），以及大规模训练数据集（ImageNet<sup>[37]</sup>、CelebA<sup>[38]</sup>等）的出现，深度学习方法在计算机视觉领域中其优越性在不断凸显。相比较于传统的图像识别技术，深度学习方法不再需要人工设计的手工特征，而是通过神经网络模型自主地在训练数据集中学习到潜在有效的特征，这很好地将特征提取融合到模型设计之中，解决了人工设计特征的难题。由此，深度学习方法开始逐步替代传统的机器学习方法成为了主流，得到了飞跃式的进展。

本文围绕动态骨显像和 PET/CT 进行相关研究。由于动态骨显像和 PET/CT 的特点，导致难以直接应用深度学习方法进行影像特征提取。例如，动态骨显像与 PET 可以展示人体组织生理代谢的强弱，而 CT 可以展示人体组织的解剖结构信息。影像中的无关区域与干扰区域会影响最终分类的精度，且单独的图像特征并不充足。动态骨显像还具有时序上的特征；而 PET/CT 还具有更全面的立体特征。

### 2.1.1 图像分类算法

常见的机器学习模型有 K-近邻算法，决策树、支持向量机等等。K-近邻算法（K-Nearest Neighbor, KNN）<sup>[39]</sup>基于一个简单的假设：相似的样本具有相似的标签。该算法通过计算样本之间的距离来代表样本之间的相似性，在训练数据集中根据每个未知样本最相似的 K 个邻居来对该未知样本进行分类。该算法简单直观易于理解，但计算和内存的开销大且对不平衡数据集不友好。决策树<sup>[40]</sup>的基本思想是通过一系列的对某一特征值的条件筛选来对数据进行划分，直至划分到叶子结点得到最终的预测结果。该算法易于理解和解释，具有可视性，但容易过拟合和达到局部最优解。支持向量机（Support Vector Machine, SVM）<sup>[41]</sup>的基本思想是寻找一个可以将不同类别样本隔开的超平面，并且最大化样本与超平面之间的最小距离。该算法具有较强的泛化能力和对噪声的鲁棒性，但仅适用于二分类问题，且计算和内存的开销大。

如今，深度学习方法逐渐成为主流，逐步取代了传统机器学习方法的地位。CNN

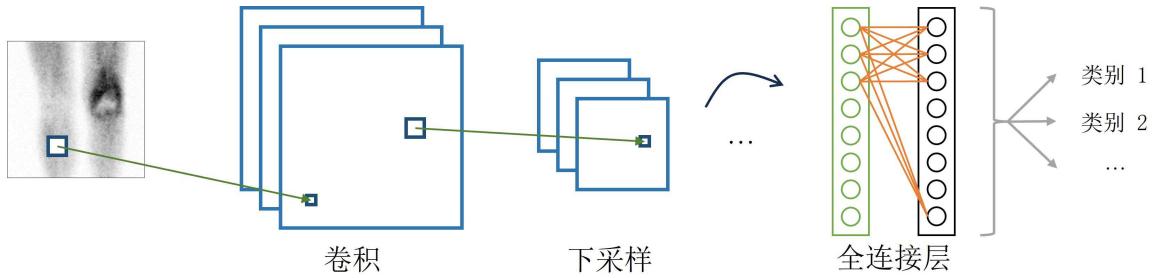


图 2.1 卷积神经网络示意图

Figure 2.1 Schematic of the convolutional neural network

是深度学习方法中具有代表性的方法之一，通常由卷积层，下采样层，全连接层组成，如图2.1所示。卷积层是核心组成部分，最大的特点在于参数共享机制，即通过相同的参数提取输入数据中不同位置的同一特征。参数共享机制可以大大减少参数的数量，提高模型的泛化能力。卷积层中若干个卷积核都具有可学习的参数，而该参数通过梯度反向传播算法进行训练调整。对于输入到卷积层的二维矩阵数据  $I$  而言，输出的像素点  $O_{i,j}$  可由公式2.1表示：

$$O_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{i+m, j+n} \times K_{m,n} \quad (2.1)$$

其中， $i, j$  表示输出矩阵的位置， $M, N$  是二维卷积的尺寸大小， $K_{m,n}$  表示卷积核中的权重值。其中，输出图像  $O$  每个位置的值，由卷积核  $K$  会沿着二维矩阵数据  $I$  进行窗口滑动依次计算得出。

接下来将介绍与分析一些经典的 CNN 模型：LeNet<sup>[42]</sup>（1998）是历史上第一个成功应用于手写数字识别任务的 CNN，并且为后来的更深层次的 CNN 的发展奠定了基础。LeNet 的整体结构非常简单，一共有七层，包括交替的卷积和池化层，以及全连接层和输出层。它的成功证明了 CNN 在图像识别任务上的有效性，也使得它成为深度学习领域中里程碑式的模型之一。

AlexNet<sup>[43]</sup>（2012）可以视为第一个现代深度 CNN 模型，与 LeNet 有着相似的网络结构，但网络层数更深。它的主要贡献在于：(1) 引入了 ReLU 激活函数去加速训练过程并减轻了梯度消失问题。(2) 为了减少过拟合，引入了 Dropout<sup>[44]</sup>正则化技术和数据增强方法用以提升模型的泛化能力和鲁棒性并扩充了数据集的大小。AlexNet 的出现为后续的 CNN 的设计和训练提供了重要的启发和借鉴。

VGG<sup>[45]</sup>（2014）的网络结构简单而有效。与 AlexNet 相比较，它的卷积核更小，

都是 $3 \times 3$ 的大小。这不仅可以减少模型参数量，还增加了模型的非线性能力。此外，VGG的网络结构可以更深，通常为16或19层，这使得网络可以捕获更好的复杂特征。VGG在图像分类任务上取得了很大的成功，其清晰的网络结构和易于理解的设计使得它成为了深度学习入门的经典模型之一。

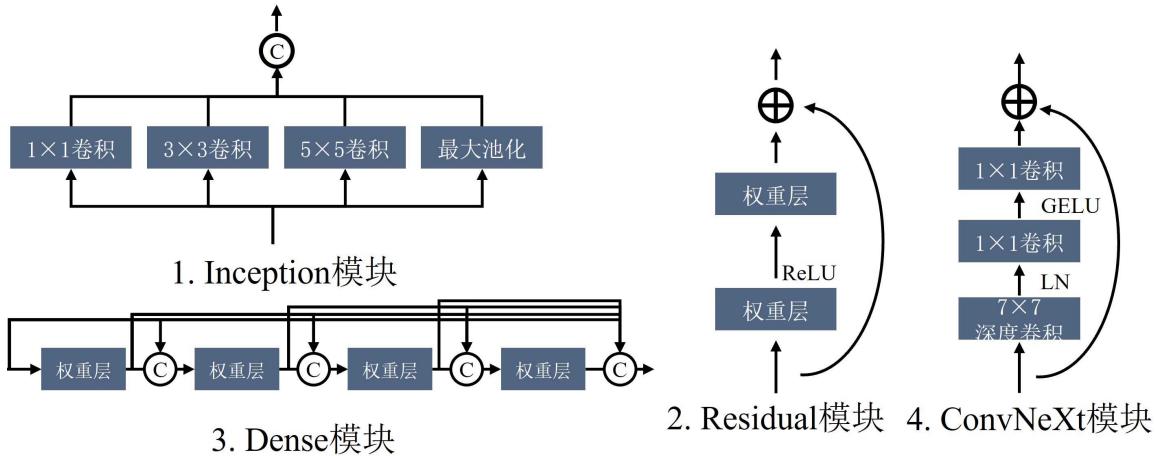


图 2.2 不同卷积神经网络的模块示意图

Figure 2.2 Schematic of the modules of different convolutional neural networks

GoogleNet<sup>[46]</sup>（2014），也称为Inception网络，初次引入了基础模块（Inception模块，图2.2）的概念，用以搭建一个稀疏且高性能的神经网络模型。Inception模块采用高度并行的设计方式，在多个不同尺度上同时执行卷积和池化操作，以捕获不同层次的特征。它将不同大小的卷积核和池化核组合在一起，让神经网络模型自主学习最合适特征提取方式。尽管GoogleNet的网络结构为22层，但它的参数量是AlexNet的 $1/15$ ，同时是VGG的 $1/3$ 。因此，GoogleNet是在计算资源有限的情况下一个较好选择。此外，GoogleNet使用全局平均池化层替代了全连接层来整合特征，将每个特征图转换为一个单独的特征值，在大大减少参数量的同时降低了过拟合的风险。

ResNet<sup>[47]</sup>（2015）的核心创新点在于引入了Residual模块（图2.2），用以解决神经网络层数加深时出现的问题，如梯度爆炸、梯度消失以及神经网络模型能力退化。其中，模型能力退化并不是由过拟合造成的，而是由于模型层数的增加导致训练误差的增加。从LeNet到GoogleNet模型发展的历程来看，神经网络模型的网络结构越深，意味着可以提取到越丰富的特征，并且，在更深的网络结构下提取的特征图含有更充足的语义信息。但是实验表明随着神经网络模型不断加深，模型的性能越差。

Residual 模块通过引入跳跃连接的方式，即将浅层的输入直接与深层的输出相加，保持梯度的流动，提升信息的传播效率，来解决梯度和训练困难的问题。ResNet 可以构建非常深的网络，甚至可以超越 1000 层。相比于传统的网络模型，ResNet 更容易优化和训练。ResNet 也是深度学习领域中里程碑式的成果之一。

DenseNet<sup>[48]</sup>（2017）在 ResNet 的基础上更进一步，充分利用了跳跃连接，提出了具有密集连接的 Dense 模块，如图2.2。Dense 模块为了确保信息的最大流动，将所有层相互连接，其中每个卷积层的输入是前面所有层的特征图拼接而成，输出会保留下来并传给后面所有的层进行拼接。这种密集连接的设计使得网络可以更充分地利用前面提取的特征信息，提升网络模型的表达能力，也有助于解决梯度消失和信息稀疏的问题。DenseNet 通过密集连接的设计，实现了全局特征的重用，从而提高了网络对复杂特征的捕获能力。其创新的网络结构和有效的特征融合方式丰富了深度学习模型的设计空间，提供了新的研究思路与方法。

ConvNeXt<sup>[49]</sup>（2022）从 ResNet 模型出发，在 Swin Transformer<sup>[50]</sup>的设计启发中汲取精华，在网络结构和训练策略上进行了多项改进，得到了性能超越 Swin Transformer 的 CNN。ConvNeXt 的基础模块如图2.2所示。由于 Swin Transformer 采用较大的自注意力窗口，自注意力中中间隐藏层的特征图通道数更大并且深度可分离卷积与注意力机制的计算具有相似性，ConvNeXt 模块采用了倒置瓶颈结构，即中间大、两头小，并使用了大卷积核（ $7 \times 7$ ）的深度可分离卷积。此外，在细节优化上，将激活函数 ReLu 改为了 GELU，归一化函数 Batch Normalization<sup>[51]</sup>改成了 Layer Normalization。ConvNeXt 表明了纯粹的 CNN 经过精心设计，依然可以与以 Transformer 为代表的新模型展开竞争。

相比较于 CNN，最初为自然语言处理任务设计的 Transformer 具有一些独特优势。即其自注意力机制在应用于图像时能够全面覆盖整个图像区域并聚焦于高度相关的特定区域。Dosovitskiy 等人<sup>[52]</sup>于 2020 年的开创性研究引入了 Vision Transformer（ViT），这是将 Transformer 模型成功应用于图像领域的首例，如图2.3所示。ViT 通过将图片分割成若干固定大小的小块（patches），并为每个小块赋予位置编码后构成序列输入到 Transformer 网络中。该模型通过捕捉图像中全局和局部的信息，来实现对图像特征的学习与分类。

Swin Transformer<sup>[50]</sup>（2021）是一种新型的 Transformer，专为视觉任务设计。其创新之处在于引入了一个称为”shifted window”的机制，它允许跨越多尺度的特征交

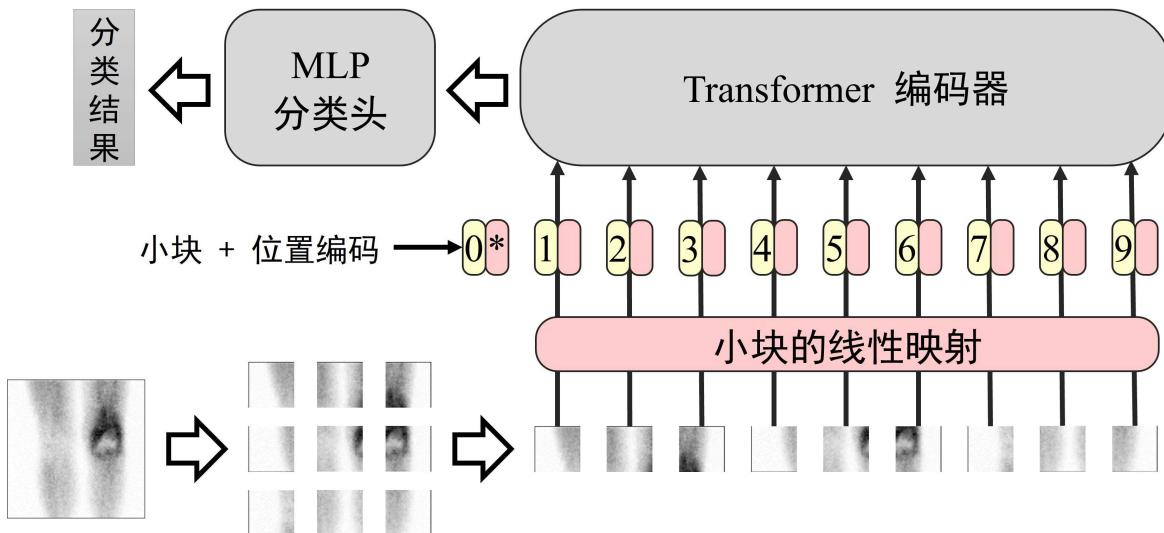


图 2.3 Vision Transformer 示意图

Figure 2.3 Schematic of Vision Transformer

互。与传统 Transformer 相比，Swin Transformer 通过此机制显著地减少了计算复杂性，使得在处理图像等高分辨率数据时更为高效。其工作原理是将图像分成多个小块，然后在这些小块中局部地应用自注意力机制，之后通过滑动（shift）这些小块的方式实现了全局信息的整合。

Transformer in Transformer<sup>[53]</sup> (TNT, 2021) 是另一种创新 Transformer 架构，它在传统的自注意力机制中嵌入了一个自注意力机制。这个设计的关键是它内部具有两级 Transformer 结构：第一级抓取像素级的特征，而第二级聚焦于区域特征，从而对整个图像进行特征表示。这种内外 Transformer 的组合使得这个架构能够更有效地捕获图像内部的复杂结构。

综上所述，CNN 和新兴的 Transformer 都是计算机视觉领域中的强大工具，它们在特征提取和模式识别方面展现出杰出的优势。对于 CNN 而言，其卷积层具有固定且较小的感受野，这使其能够有效地识别局部特征，如边缘、形状和纹理等，但可以通过多层堆叠以捕捉全局上下文信息。而 Transformer 通过自注意力机制可以捕捉到全局上下文信息，从而实现对整张图像的理解。然而，Transformer 模型的训练需要依赖大量数据，而数据集小是医学影像领域普遍面临的问题。因此，CNN 成为了更值得探索的方法。对于动态骨显像和 PET/CT 诊断分类任务而言，CNN 可以有效地学习到病灶区域中局部关键性特征，例如生理摄取的形态和解剖结构的细节。然而，对于具有时序性特点的动态骨显像，由于 CNN 并非专门为处理序列数据而设计，其

在这方面的能力是受限的。

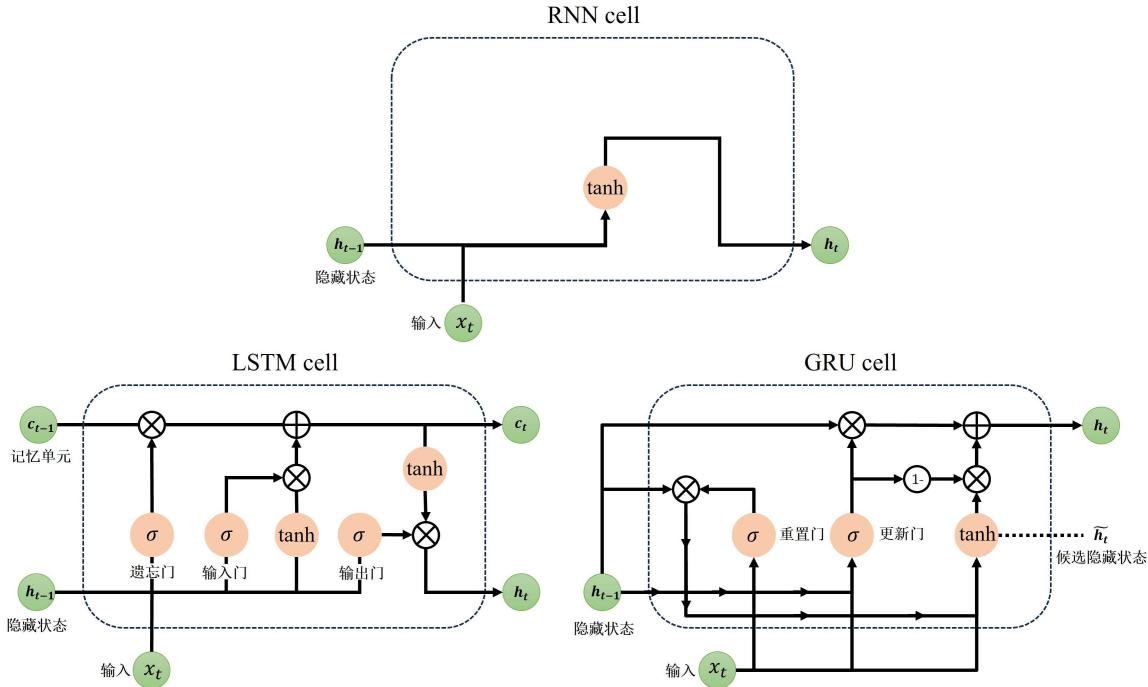


图 2.4 循环神经网络示意图

Figure 2.4 Schematic of the recurrent neural network

对比 CNN, RNN 作为深度学习方法之一, 是一种专门用于处理序列数据的神经网络模型, 非常适用于时间序列分析任务之中。RNN 的主要原理在于它通过不断地更新隐藏状态 (hidden state) 去保存序列中每个时间点上一定的历史信息, 并用于当前时间点的决策或下一时间点的预测, 如图 2.4。RNN 的结构可以是序列到序列的形式, 但当 RNN 应用于时间序列分析中的分类任务时, 其结构可以变成序列到单一输出的形式。对于输入的序列数据  $X = \{x_1, \dots, x_t, \dots, x_T\}$  而言, 在每个时间点  $t$  上, RNN 通过以下公式更新隐藏状态  $h$ :

$$\begin{aligned} h_t &= \tanh(W_h h_{t-1} + W_i x_t) \\ y^t &= \sigma_s(W_h h_{t-1}) \end{aligned} \quad (2.2)$$

其中,  $\tanh$  表示双曲正切函数,  $W_h$  和  $W_i$  分别表示上一隐藏状态到当前隐藏状态的权重和当前输入到当前隐藏状态的权重,  $\sigma_s$  表示 Softmax 函数, 用于计算所属不同类别的概率分布,  $y^T$  表示 RNN 中的单一输出。

RNN 在理论上可以捕捉长期依赖。但是在实际中, 随着时间的增长, 梯度消失

或梯度爆炸的问题会出现，并且 RNN 学习和保留过去的历史信息的能力是有限的。为了解决 RNN 中存在的缺陷，研究人员提出了它的变体，长短时记忆网络（Long Short Term Memory, LSTM）<sup>[54]</sup> 和门控循环单元（Gated Recurrent Unit, GRU）<sup>[55]</sup>。

LSTM<sup>[54]</sup>（2010）相比较于 RNN 拥有一个更为复杂的结构，如图2.4所示。它在 RNN 的基础上添加了三个门控记忆单元（遗忘门、输入门和输出门）以及一个独立的记忆单元  $c_t$  来更新和输出隐藏状态。在长序列数据的处理中，LSTM 展现出其强大的能力，通过动态地调节记忆单元，以实现时序信息的有效保留。在训练过程中，LSTM 通过精确的门控机制，执行三项关键操作：首先，它利用遗忘门从单元状态中丢弃非核心信息；其次，它通过输入门引导新的信息加入单元状态；最终，它通过输出门精心挑选必要信息来构建输出。这一过程赋予了其在长时间跨度上维持和更新历史信息的能力，并捕捉到数据的深层依赖关系。由此机制，LSTM 有效地克服了传统 RNN 面临的梯度消失与梯度爆炸问题。

GRU<sup>[55]</sup>（2014）相比较于 LSTM，简化内部繁琐的门控机制，如图2.4所示。它在 LSTM 的基础上将隐含状态与记忆单元合并为一个统一的隐藏状态，并通过两种门控单元（更新门与重置门）来控制信息流动。更新门的作用是调节前序隐藏状态对当前隐藏状态的影响程度，允许模型在必要时保存跨越较长时间区间的信息。而重置门则决定在当前隐藏状态的更新过程中，前序隐藏状态的参与程度，这样使得模型可以有选择性地忘记或保留过往信息。GRU 通过更加灵活和精简的门控策略来维持长期的状态信息，并同时捕获短期的上下文特征，这一策略在许多实验中被验证能够有效预防梯度消失的问题。

Yiugit 等人<sup>[56]</sup>于 2021 年提出了两种简单且有效的 GRU 变体：GRU1 和 GRU2。通过一种基于经验的方法，这些学者试图确定一组能够提高 GRU 准确度或降低损失的系数  $\epsilon, \alpha$ 。其中， $\epsilon$  确定门控循环单元的作用力度， $\alpha$  确定门控当前单元门的作用力度。相比较于 GRU，GRU1 增加了当前单元的作用，而 GRU2 减少了循环单元的作用。值得说明的是，对标准 GRU 算法进行这样微小且简便的调整能够显著提升其性能。

综上所述，RNN 及其变体为处理序列数据提供了富有弹性的学习架构。RNN 以其简单的循环连接结构，去捕获时序信息，但仍存在梯度消失或爆炸的问题。LSTM 和 GRU 则引入门控机制，有效地保存长期依赖关系，优化了信息的存储与过滤，降低了梯度消失和爆炸的风险。对于动态骨显像中的假体关节感染诊断任务而言，不

仅需要关注每张图像中的病灶摄取形态特征，还需要抓取跟时序相关的生理代谢差异变化特征，RNN 和 CNN 的有效结合将适用于该任务的模型设计，同样也是图像序列分类任务的重点研究方向之一。

### 2.1.2 医学影像分类算法

计算机辅助诊断 (computer aided diagnosis, CAD) 是通过计算机对病患的医学影像，如 PET、CT、MRI 等，进行分析并辅助医生完成相应疾病的诊断流程。计算机依靠逐渐流行的深度学习技术，可以自动地从目标数据中学习获得更深层次的特征，并且排除人为因素的影响，实现全自动化的辅助诊断。基于医学影像的分类诊断是计算机辅助诊断的重要研究方向之一。对于不同的医学影像而言，由于其本身成像技术不同，从而拥有不同的特性和成像效果，这也是科研人员不可忽视的研究基础之一。根据医学影像的特点，有效设计合适的图像处理算法和模型结构有助于提升分类诊断任务的质量。

Cheng 等人<sup>[57]</sup> (2017) 提出了一个级联的三维 CNN 用于脑部 PET 影像的分类诊断。该方法首先划分整个 PET 脑影像为数个三维的局部子区域。接着，利用一个三维卷积网络来提取每个子区域中更深层次的鉴别特征。最后，应用一个高级别的三维卷积网络来融合各个子区域特征，实现对阿尔茨海默症患者与健康对照组的有效区分。Liu 等人<sup>[58]</sup> (2018) 在阿尔茨海默病的诊断领域，采用了创新性的方法，将 CNN 与 RNN 相结合。把三维 PET 影像分解为二维切片序列，通过二维卷积网络学习切片内部的特征和 RNN 掌握切片间的序列特征，来展现对阿尔茨海默病的高效分类性能。Khagi 等人<sup>[59]</sup> (2020) 提出了一种用于脑部 MRI 和 PET 影像进行分类诊断的三维卷积架构 divNet。它可以有效区分阿尔茨海默症、轻度认知障碍和正常，用于协助临床医生进行初步快速检测来确定患者的病情。Lee 等研究者<sup>[60]</sup> (2021) 对 PET 影像进行了深入探究，评估了 Inception3D、ResNet3D 以及 VGG3D 三种深度学习架构在阿尔茨海默病诊断中的性能表现。结果显示，这些模型在诊断任务中平均准确率高达 95%。在进一步的外部验证中，VGG3D 具有最佳的泛化能力。Liu 等人<sup>[32]</sup> (2022) 提出一种创新性多尺度 CNN 来用于脑部 MRI 影像中的阿尔茨海默病的诊断。该网络通过引入通道注意机制，改善通道间的相互依赖关系，自适应地调整通道维度上的特征，来增强它的特征表达能力。大量实验证明，通过将 MRI 划分为白质 (white matter) 和灰质 (gray matter) 来进行训练，该网络以更少的参数和更低的计算

复杂度获得了最优异的性能。Takahashi 等人<sup>[33]</sup>（2022）探索了一种创新的图像预处理方式，即采用了多角度的最大强度投影（Maximum Intensity Projection, MIP），如图2.5所示。他们将 MIP 预处理后的三维 PET 影像输入至深度学习模型中进行训练，使其能够识别乳腺癌的标志性特征。在随后的性能评估中，该模型展现出了超越没有经验的放射科医生的诊断能力，与三名专业放射科医生相当的诊断能力。

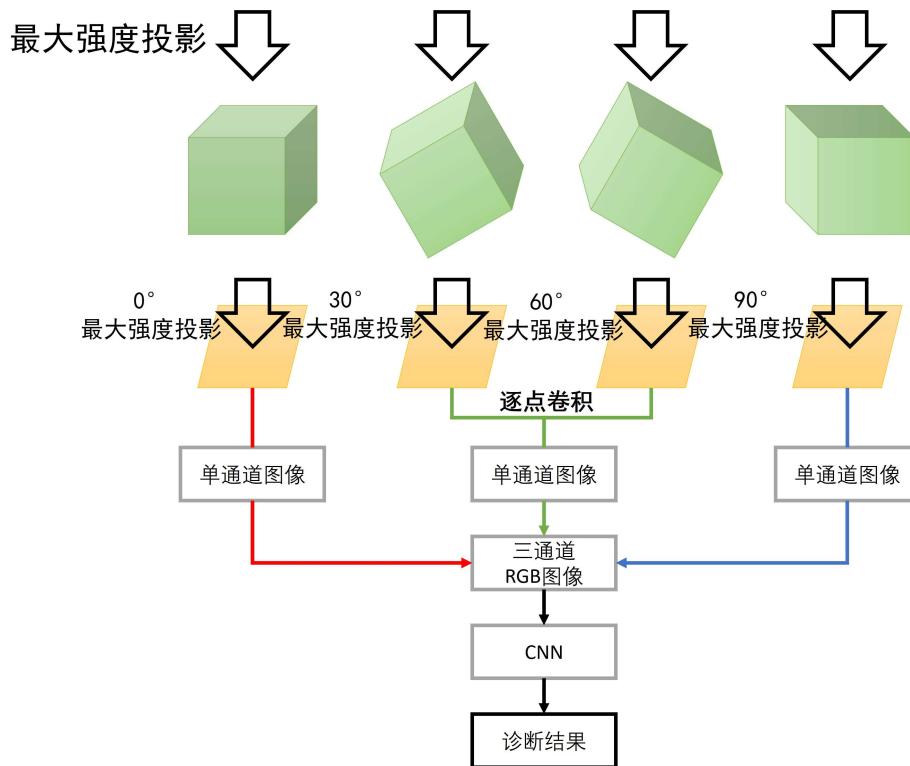


图 2.5 四个角度最大强度投影图像作为输入进行预测的示意图

Figure 2.5 Schematic of four angular maximum intensity projection images as input for prediction

综上所述，对于动态骨显像中假体关节感染任务，以上研究存在一些可借鉴之处。例如，将时序性图像的固有三维性质，直接采用三维 CNN 进行特征提取似乎是一种直观的策略。然而，传统的 CNN 在捕捉长期时序依赖方面存在局限性。由此，结合使用 CNN 和 RNN，以便在时间序列的内部和之间提取综合特征，对于该特定医学任务来说可能更加精准和适宜。同时，也应该重视任务的特定难点，比如无关区域干扰和生理状态下高摄取区域可能对分类精度造成负面影响。鉴于此，研究并设计针对动态骨显像的数据预处理方法，去有效减少负面影响，进一步优化深度学习模型的性能，是提高诊断精准度的关键。

在 PET/CT 影像中的骨折相关感染任务中，考虑到 PET/CT 为三维影像数据，直

接采用三维 CNN 是一个不错的选项，以便充分利用体积数据的空间信息。同时，考虑到计算复杂度和数据维度的挑战，采用 MIP 对数据进行降维处理也是一个有效的替代方案。此外，针对 PET/CT 影像的特定情况，如多病灶的存在和相对较小的病灶区域在整体影像中难以突显导致传统的深度学习模型可能无法有效集中于关键的病灶区域的问题，需要特别关注。因此，开展针对 PET/CT 影像的专业预处理方法研究和引入检测算法定位病灶区域，并对其进行截取或突出处理，可以在不损失重要信息的前提下减少背景噪声的干扰，进而提高深度学习模型在骨折相关感染检测任务中的准确性和效率。

## 2.2 图像检测

检测任务通常指输入数据中一个或多个目标物体的识别与空间定位过程。图像检测，在计算机视觉领域内，是解决“识别图像中存在的目标对象以及这些目标对象的准确位置”的关键性问题。该任务适用于众多领域，例如安防监控、自动驾驶、医学影像分析等等。

在医学影像分析中，检测任务特别强调了对疾病标志物的精准定位与识别，如肿瘤、异常组织等，从而为临床诊断提供关键的影像学依据。随着深度学习技术的发展，检测任务通常是要求模型通过边界框来准确发现并定位目标对象。在深度学习与医学影像分析相结合的背景下，检测任务有时会通过像素级分类，即图像分割，来增强目标的定位精度。无论是哪种定位方式，这都需要检测模型具有理解和区分图像中各种复杂特征的能力，以便实现对不同的目标对象的准确识别和区分。

在骨折相关感染的诊断流程中，医生通常集中审视病灶区域在生理代谢方面展现的摄取形态，因为与感染相关的特征通常紧密聚焦于这些区域。对此，探讨检测任务中的神经网络架构显得尤为重要。本节将概述有关于此方面的研究现状，并分析这些架构应用于 PET/CT 影像上进行骨折相关感染病灶区域检测的可行性。

### 2.2.1 图像检测算法

在执行检测任务的深度学习模型分类上，一般可划分为目标分割模型与目标检测模型两大类。目标分割模型，以 U-Net<sup>[61]</sup>为代表，旨在实现对图像中每个像素所属类别的识别，从而获取精确的目标形状与位置。这类模型在医学影像中尤为重要，因为它允许细致划分病灶区域，为后续的量化分析提供了基础。相对地，目标检测模

型，如区域卷积神经网络（Region with CNN features, R-CNN）<sup>[62]</sup>和 YOLO（You Only Look Once）<sup>[63]</sup>，往往通过预测边界框来定位目标对象，提供了一种相对较快且效率更高的检测方法，尽管其定位的精细程度相对较低。无论是哪种类型的模型，其设计和优化都在不断地促进医学影像分析与计算机视觉领域的研究进展，继而改善临床诊断与治疗的质量。

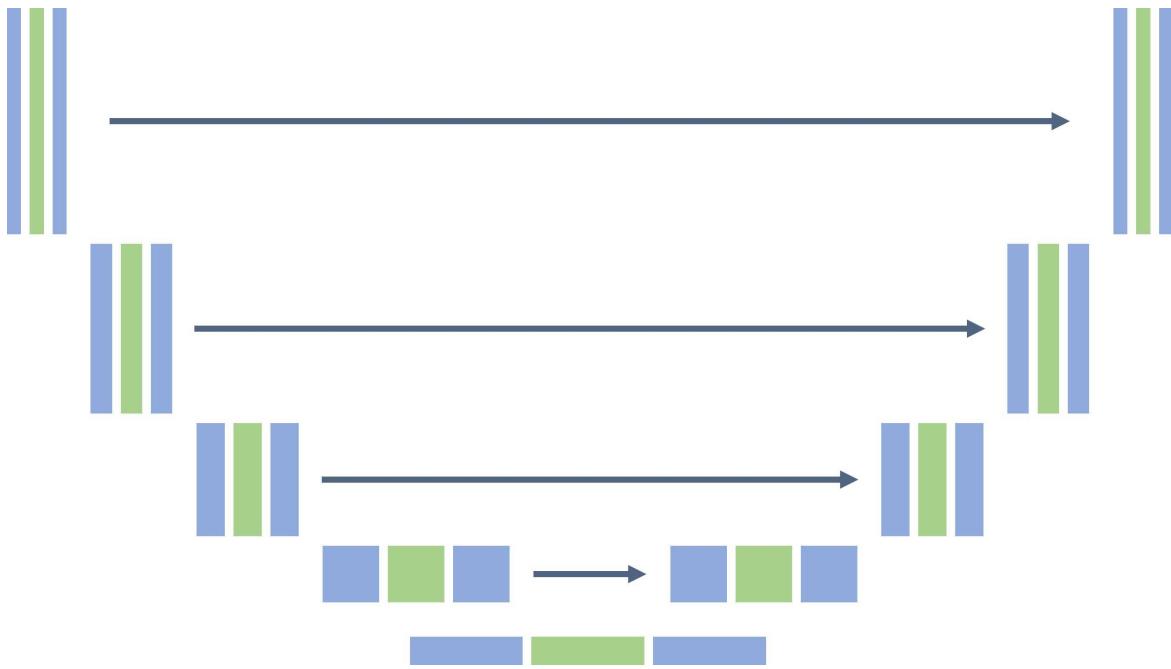


图 2.6 U-Net 整体结构示意图

Figure 2.6 Schematic of the overall structure of U-Net

在目标分割模型中，U-Net<sup>[61]</sup>（2015）的提出开辟了医学影像分割任务的新纪元。它是一种专为医学图像分割设计的深度学习模型结构，摒弃了之前模型对大量数据依赖的限制，在较少标注样本的情况下仍能实现高效能的模型训练。在结构上，如图2.6所示，整体呈现为一种“U”字形的对称结构，包括下采样（收缩路径）和上采样（扩展路径）两个部分，能够捕获上下文中多尺度的语义信息。下采样阶段逐步提取特征并降低数据维度大小，而上采样阶段逐步恢复数据维度大小并进行特征解码。同时，通过跳跃连接将下采样阶段的语义信息与上采样阶段的位置信息相结合，以增强特征并实现精确的定位。

U-Net++<sup>[64]</sup>（2018）是 U-Net<sup>[61]</sup>的嵌套版本，强化了特征学习能力，增强了模型对复杂医学影像的分割准确度。主要的创新点包括跳跃连接的重新设计和深度监督。在跳跃连接方面，U-Net++ 添加了密集的跳跃路径，允许模型结合来自不同阶段的特

征，从而提升了捕捉细节、特征融合与学习的能力。在深度监督方面，U-Net++为了充分利用多尺度的语义信息，在每一层的嵌套路径上，都可以输出一个分割图结果。通过所有的分割图结果一起联合优化，有效地强化了模型对不同尺度特征的学习。

Attention U-Net<sup>[65]</sup> (2018) 是一种集成了注意力机制的改进型 U-Net<sup>[61]</sup>。这种机制可以使得模型更加关注于影像中的关键部分，从而提高特征的区分力和医学影像分割的准确性。其核心的创新点在于跳跃连接中加入了注意力门控模块，以便过滤和引导传统的跳跃连接特征，使得网络能够集中学习更有价值的信息。

U-Net 3+<sup>[66]</sup> (2020) 是在 U-Net 的基础上发展出来的一种医学影像分割模型。该模型通过引入全尺度跳跃连接和深度监督机制，增强了对多尺度特征的融合，并提高了模型的学习能力。全尺度跳跃连接保证了各个解码阶段中全尺度特征的充分利用，而深度监督机制在各个解码阶段直接引入了额外的监督，确保了模型在多层次结构上的有效特征学习和提取。

TransU-Net<sup>[67]</sup> (2021) 是一种结合了卷积强大的局部特征提取能力和 Transformer 的全局自注意力机制的新型网络架构。它先使用卷积提取空间特征，再利用 Transformer 编码器以捕获长距离上下文关系，最后通过解码器来输出精确的分割结果。这种混合架构使得 TransU-Net 在分割精度上超越了传统的 CNN 网络。

综上，U-Net 及其变体具有强大的特征学习能力，在医学影像处理中得到了广泛的应用。每个变体都是为了解决特定问题而设计，比如改进特征的捕获能力、更精确地分割图像中的特定部分等。随着深度学习技术的进步，将会有更多的 U-Net 变体被提出来解决更复杂的医学影像分割任务。在 PET/CT 中，用于识别骨折相关感染病灶的任务面临的挑战在于感染区域的形态分布呈现多样性，包括但不限于点状、簇状或弥散性分布。这种分布的复杂多样性导致利用像素级定位进行精确标注和检测变得较为困难。因此，本研究的重点在于目标检测模型。

在目标检测模型中，深度学习方法主要分为两大类：二阶段检测器和一阶段检测器。

二阶段检测器的工作流程分为两个阶段：1. 生成目标的候选区域；2. 提取这些区域的特征进行分类和边界框的精调。在具有里程碑式的二阶段检测器中，R-CNN<sup>[62]</sup> (2014) 是第一个开创性的二阶段检测器，整体结构如图2.7所示。在 R-CNN 模型中，首先由选择性搜索（Selective Search）算法依据图像特征如颜色、纹理、形状等，负责挑选出潜在含有目标对象的候选区域。随后，对这些区域进行标准化预处理，并

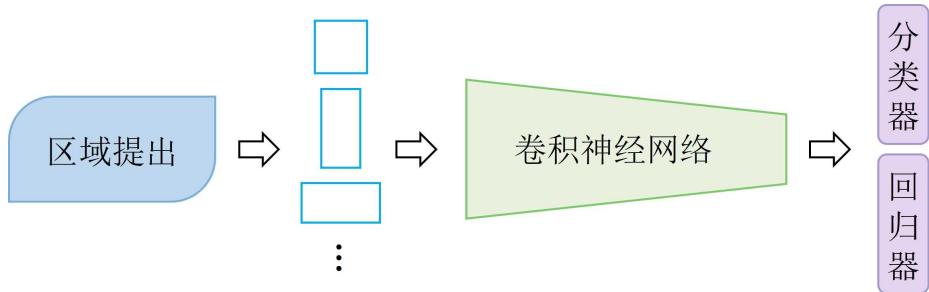


图 2.7 R-CNN 整体结构示意图

Figure 2.7 Schematic of the overall structure of R-CNN

通过预训练的 CNN 提取特征。最终，由分类器对特征区域进行分类，并通过边界框回归器精确调整目标对象的位置，以提升目标检测的准确性。虽然 R-CNN 在其推出时对目标检测领域产生了巨大的影响，但它也存在一些缺点，例如速度慢、训练需要对不同部分分阶段进行，存储需求高。为了克服这些缺陷，后续基于 R-CNN 的改进模型，如 Fast R-CNN<sup>[68]</sup> 和 Faster R-CNN<sup>[69]</sup>，被提出。

Fast R-CNN<sup>[68]</sup>（2015）对整个目标检测流程进行了端到端的优化，使得模型可以同时执行目标分类任务和目标定位任务，从而提高了训练速度、检测效率和检测精度。该方法的核心创新点有两个：一是共享了 CNN 提取的整个图像的特征，采用了感兴趣区域池化技术，在该特征图上对多个大小不同候选区域映射为固定大小的特征。二是将目标分类和边界框回归进行结合，并提出了多任务损失函数用来同时优化分类和回归任务。

Faster R-CNN<sup>[69]</sup>（2015）在 Fast R-CNN 的基础上创新地引入了区域提议网络（Region Proposal Network, RPN），它能够替代掉选择性搜索算法，并且学习如何快速且高质量地生成候选区域。RPN 与 Fast R-CNN 的结构无缝地整合在一起，进一步提升了目标检测的效率与准确性。这不仅实现了整个过程端到端的训练和较之前模型的显著加速，还成为目标检测领域的又一里程碑。

一阶段检测器，相比较于二阶段检测器，跳过了候选区域的生成阶段，直接对图像中的目标进行位置和类别的预测。它们通常使用全卷积网络来分析图像，使得在推理过程中能够获得更快的速度。

YOLO<sup>[63]</sup>（2016）是一阶段检测器中非常流行且影响深远的一种端到端的目标检测器。其核心思想是把目标检测任务转换为一个回归问题来解决。YOLO 采用一个单一的神经网络结构去全图预测，如图2.8，其结构可大致分为三个部分：用于特

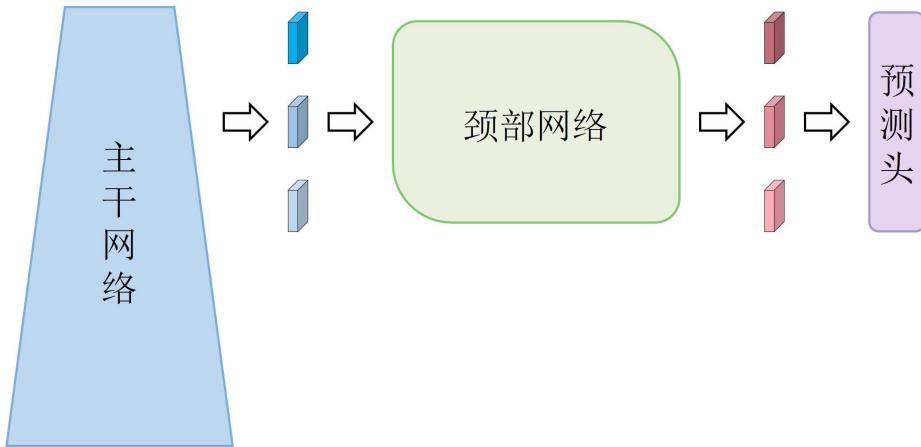


图 2.8 YOLO 整体结构示意图

Figure 2.8 Schematic of the overall structure of YOLO

特征提取的主干网络，用于特征融合的颈部网络和用于预测的头部网络。YOLO 将输入图像划分为多个网格，每个网络预测多个目标的边界框和类别以及一个置信度。置信度反映了边界框包含目标的可能性。由于 YOLO 的设计简单直观和检测速度高效，它在现实世界中应用广泛，并且不断地在各个版本中得到改进，并确保在速度和准确度间取得更好的平衡。

YOLOv2<sup>[70]</sup>（2017）是 YOLO<sup>[63]</sup>的改进模型。它在保持高速检测的特点上，通过引入批标准化、使用更高分辨率输入、增加锚框来改善边界框预测、更换主干网络以及利用多尺度训练来提高对不同尺寸目标的检测能力，从而大幅提升了检测精确度，尤其在小目标的检测上。

YOLOv3<sup>[71]</sup>（2018）进一步提高了检测性能。它引入了多尺度预测来加强对小目标的识别能力，还通过加深主干网络来增加网络的特征学习能力和保持了合理的运算效率。它还采用了逻辑回归进行类别的预测，并优化了描框的机制以更好地适应不同形状的目标。整体上，YOLOv3 在速度与准确度之间获得了更优的平衡。

YOLOv4<sup>[72]</sup>（2020）是目标检测领域中又一重要进展，其在 YOLOv3<sup>[71]</sup>的基础上，综合了多项新技术来进一步提升检测准确率并且保证高速处理。该版本采用了跨阶段部分（Cross Stage Partial）<sup>[73]</sup>结构应用于主干网络中，引入了 Mish<sup>[74]</sup>激活函数，使用了路径聚合网络（PANet）<sup>[75]</sup>和空间金字塔池化（SPP）<sup>[76]</sup>，增强了网络的特征提取和特征融合能力，同时包含了多种检测技术，如自适应锚框计算等等。

YOLOv5<sup>[77]</sup>（2020）是 YOLO 系列的又一迭代，而其自身就迭代了多个版本。它采用了模块化的设计，可以选择不同大小的模型，更易于使用。YOLOv5 包含了自动

学习的锚框尺寸、多尺度训练、权重剪枝优化等技术，使模型既可以在资源有限的设备上运行，也能在需要高精确度的场合中保持较高性能。

YOLOX<sup>[78]</sup>（2021）是一个新兴的目标检测架构，它在 YOLOv3 的基础上做了优化和改进，目标是在性能和灵活性之间取得更好的平衡。YOLOX 的主要特点是它放弃了原有的锚框机制，采用了无锚框的设计，这使得网络在训练时对各种大小的目标有更好的适应能力。除此之外，YOLOX 采用了解耦的头部设计，在分类和回归两个不同的任务上分开进行预测。

综上所述，一阶段目标检测器得益于算法和计算能力的提升，在目标检测任务中表现出色。它们在处理速度上显著优于二阶段检测器，满足实时处理的需求。速度的提升并未以牺牲准确率为代价，经过精心的网络设计和训练，一阶段检测器在不断地缩小与二阶段检测器的精度差距，甚至超越了它们。因此，在 PET/CT 的病灶检测任务中，面对感染区域的复杂多样性，选择一阶段检测器提供了一个既快速又准确的检测策略。

## 2.2.2 医学影像检测算法

在医学影像领域中，基于深度学习的检测算法已经逐步应用于现代的诊断过程之中。随着医学影像技术的发展，如 MRI、CT 和 PET 等成像技术不断进步，产生了大量的高维度和高质量的医学影像数据。这些影像含有丰富的生物医学信息，对于疾病的早期发现、诊断和治疗监测提供了重要依据。为高效、准确地解读这些复杂的数据，基于深度学习的医学影像检测算法被广泛研究和部署。

Tang 等人<sup>[79]</sup>（2018）的研究提出了一种创新的架构，专门用于 CT 中的肺结节检测。该研究构建了一个受 U-Net 启发并结合 3D Faster R-CNN 的检测模型用于提出肺结节的候选框，还设计了一个三维卷积神经网络去降低假阳性。Li 等人<sup>[80]</sup>（2019）在神经网络设计中巧妙地融合了人类的专业知识，提出了一种具有位置感知注意力机制的多视图特征金字塔网络。该网络专门针对常见的病变检测任务，在 CT 的三种不同视角下，有效地利用三维医学影像信息。Al-masni 等人<sup>[81]</sup>（2020）提出了一种全自动、两阶段的集成深度学习框架，旨在高效地检测脑微出血。第一阶段中，YOLO 被采用来识别潜在的脑微出血候选区域；随后，在第二阶段，引入了一个三维卷积网络来减少检测结果中的假阳性。Mei 等人<sup>[82]</sup>（2021）在 YOLOv4 的基础上，结合各种现有的技术，如深度过参数化卷积层、卷积注意力模块和焦点损失函数，并进

行了冗余的通道修剪，提出了一种更高效且实用的肺结节检测器 YOLO-lung。Lee 等人<sup>[83]</sup>（2021）提出了一个创新的头颈部肿瘤分割框架。此框架采用了 U 形结构，设计了一种具有双分支和交叉注意力模块的编码器以及一个解码器。此外，模型集成的应用进一步强化了框架的分割性能。Wang 等人<sup>[84]</sup>（2022）提出了一种创新的多尺度自适应注意力特征融合网络，专注于 PET/CT 中的胰腺病变检测。该网络的核心创新体现在基于多模态的双特征提取网络和特征融合网络，以获得丰富的多尺度特征的同时让语义信息更加集中，提高了检测的精确性。Huang 等人<sup>[85]</sup>（2022）利用深度学习技术，提出了一种融合了 PET 高敏感度和 CT 精细解剖信息的创新性目标分割方法，如图2.9所示。此方法在编码阶段采用双模态的输入策略，并在解码阶段对不同模态信息进行特征融合，最大化地挖掘 PET 和 CT 各自的优势。这一方法可以有效地提高肿瘤分割性能。

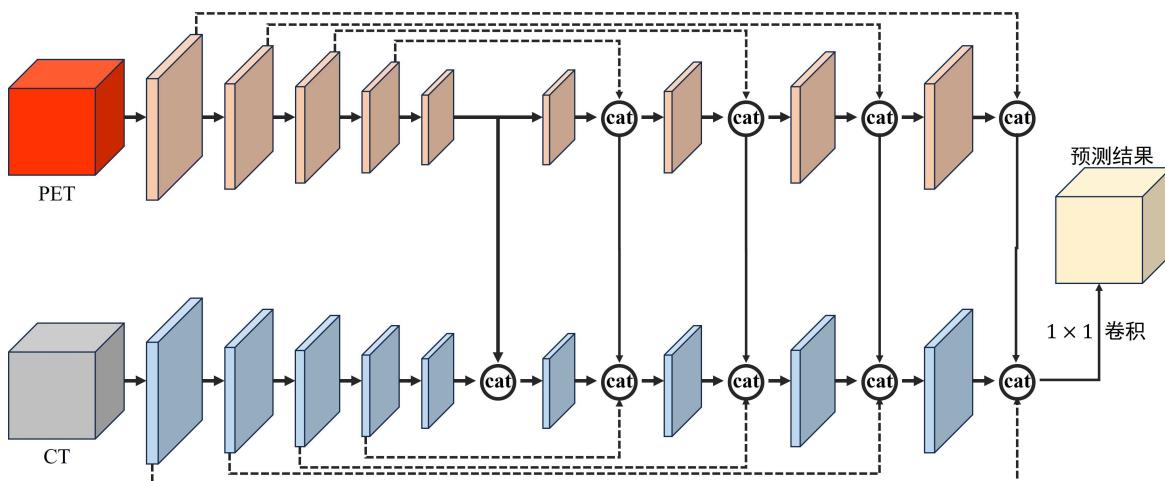


图 2.9 融合 PET 和 CT 的目标分割方法

Figure 2.9 Object segmentation method by fusing PET and CT

通过上述研究，可以明确认识到在 PET/CT 中定位骨折相关感染的任务需要充分利用 PET 与 CT 两种不同模态提供的信息。整合和发挥两者独特优势至关重要。因此，探讨一个双分支网络架构是一个非常值得考虑的方向。同时，还需面对 PET/CT 中高摄取区域所带来的干扰问题，并采用创新性的实用方法予以解决。借鉴以往研究中有效的技术方案，并结合本研究任务的具体特性，提高定位准确性是本研究努力的核心目标。

## 2.3 本章小结

本章首先回顾了分类任务的核心流程及其发展变化。继而，着重解析了 CNN、ViT 和 RNN 的基础理论和关键特性，并针对本研究所涉及的分类任务，探讨了这些模型的适用性。同时，本章亦涵盖了近期医学影像领域分类研究的进展，并详细审视了这些进展对于本项研究所具有的借鉴意义。此外，特别强调了深度学习领域中检测算法的研究，梳理与总结了当下普遍应用的网络架构的发展脉络，在医学影像检测领域的最新研究动态，并从中归纳出适于本研究检测任务的方法，为后面的研究提供了坚实的基础。

### 第三章 基于<sup>99m</sup>Tc-MDP 动态骨显像的假体关节感染辅助诊断框架

上一章介绍了用于分类和检测任务中的人工智能方法，以及它们在医学影像领域中的应用。基于上一章的分析，本章主要研究用于辅助诊断假体关节感染的诊断框架。该框架包括预处理算法和基于时序性影像的分类模型 DBS-eNet（Dynamic Bone Scintigraphy effective neural Network, 动态骨显像有效神经网络）<sup>[86]</sup>。首先，预处理算法从动态骨显像中获取有效信息。然后，由 DBS-eNet 进一步地提取动态骨显像时序上的生理代谢差异性变化特征和其中每张图像中的组织摄取形态特征来实现疾病的自动诊断。在实验中，该框架和目前主流的分类网络、专业的核医学专家进行了对比，结果表明该方法可以有效且准确地辅助诊断假体关节感染，同时具有潜在的临床应用价值。

#### 3.1 问题分析

在面对动态骨显像中的假体关节感染的诊断任务时，必须综合考虑医生对假体关节感染的诊断流程、动态骨显像的性质和辅助诊断框架的有效设计。因此，本文需要思考以下几个问题：

(1) 关节置换手术后动态骨显像的公开数据集的缺乏。本文作者与上海市第六人民医院核医学科医生收集了许多患者影像数据并进行了清洗，构建了一个由经历过髋关节置换手术或者膝关节置换手术的患者的真实动态骨显像数据集。然而，该数据集存在着数据规模较小、数据不平衡的问题，增加了人工智能诊断框架的设计要求与训练难度。

(2) 髋关节置换手术后的动态骨显像存在着生理高摄取区域。在人体的髋部附近存在着膀胱和肾脏这类自然高摄取区域，同时假体关节疼痛区域也会出现高摄取的情况，如图3.1所示。这两者具有相似的图像特征，同时出现会导致神经网络在训练过程中产生混淆，无法学习到真实有用的信息，难以关注于病灶区域。

(3) 动态骨显像本身的影像特点。动态骨显像包括三种不同条件下采集的影像，分别是在注射后立刻动态扫描（2秒/帧，20帧）用于获取灌注信息的血池相、灌注

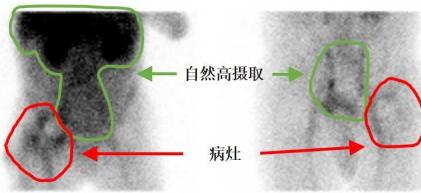


图 3.1 髋关节置换手术后的动态骨显像

Figure 3.1 Dynamic bone scintigraphy after hip replacement

后动态扫描（1分钟/帧，5帧）可帮助诊断炎症状况或血液供应问题的血流相<sup>[87]</sup>和在延迟代谢阶段中（示踪剂注射后4小时，1帧）扫描的延迟相。因此，主要存在的问题是需要确定要使用哪些相。

(4) 动态骨显像具有时间连续性。动态骨显像中，每张图像拥有组织摄取形态特征，一系列图像之间在时序上存在生理代谢差异性变化特征。因此，如何利用动态骨显像的时序性特征也是一个需要研究的主要问题之一。

(5) 神经网络是一个缺乏解释的黑盒。DBS-eNet 是一个端到端的神经网络，可直接通过数据输入到模型之中获得最终结果。这导致医生很难以理解或者分析模型所提取和学习到的相关知识。这不仅缺乏决策过程中的透明性，还没有对模型的决策结果的解释，无法给医生带来可信度以及临床诊断的依据。

在本章研究中，设计了一个基于<sup>99m</sup>TC-MDP 动态骨显像的框架，用于辅助诊断假体关节感染。该框架针对上述问题一一进行优化改进：

(1) 通过使用数据增强方法（镜像、旋转）和损失函数 Focal Loss<sup>[88]</sup>上的改进（结合 Label smoothing<sup>[89]</sup>），来解决数据集规模较小和数据不平衡的问题。

(2) 提出了三种数据预处理算法，分别是标准预处理算法、干扰减少预处理算法和基于感兴趣区域预处理算法，来获取图像序列中有效的信息。其中，标准预处理算法可以用于经历过膝关节置换术后患者的动态骨显像，同时这三种预处理算法都可以用于经历过髋关节置换术后患者的动态骨显像。

(3) 动态骨显像由血流相、血池相和延迟相组成。由于非感染与感染在延迟相中均可以呈现阳性，而血流相和血池相对于假体关节感染的诊断更具有敏感性。当血池相和血流相呈现阳性表明感染的可能性较大，当血池相和血流相呈现阴性的时候可以排除假体关节感染。由此，在专家的建议下，本章的研究主要关注于动态骨显像的血流相和血池相。

(4) 针对动态骨显像的时序性图像序列特性, 提出了一个基于深度学习和多特征的有效分类网络, DBS-eNet。该网络采用了三维深度卷积和卷积长短期记忆 (Convolution Long Short Term Memory, ConvLSTM)<sup>[90]</sup>作为基础构建块, 能够提取预处理后的动态骨显像的影像学特征和时序特征, 并且获得最终的预测结果。

(5) 为了验证所提出框架的决策结果的有效性和可信度, 本章引入了用于可视化模型对某一类输入图像的兴趣区域的 Grad-CAM (Gradient-weighted Class Activation Mapping)<sup>[91]</sup> 和用于可视化分类模型提取的特征的可区分性的 t-SNE (t-distributed Stochastic Neighbor Embedding)<sup>[92]</sup>。

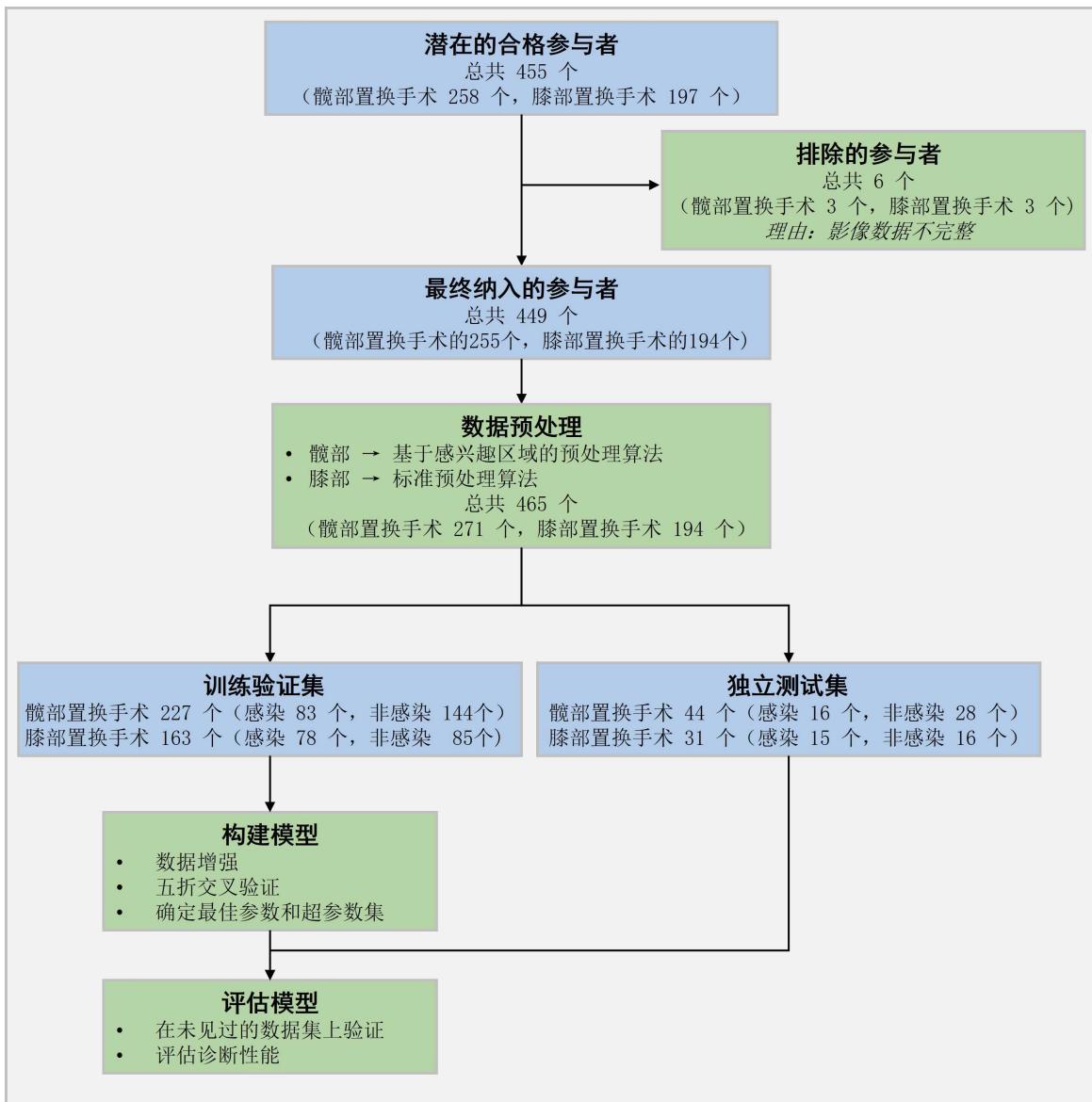


图 3.2 整体研究流程

Figure 3.2 Overall research process

本章研究的整体流程如图3.2所示。首先，通过最初的数据收集与筛选，最终得到了包含227个髋关节置换手术后的患者数据（感染组与非感染组：83 vs. 144）和163个膝关节置换手术后的患者数据（感染组与非感染组：78 vs. 85）的训练验证集，以及由44个髋关节置换手术后的患者数据（感染组与非感染组：16 vs. 28）和31个膝关节置换手术后的患者数据（感染组与非感染组：15 vs. 16）组成的独立测试集。其次，设计和改良模型结构。通过在训练验证集中五折交叉验证的结果，适应性地调整模型结构、超参数和训练策略，来获取最终的模型和最佳的超参数。最后，在独立测试集上采用分类任务中的评估指标进一步全面地评估模型的综合性能。

## 3.2 数据预处理

为了有效地解决动态骨显像数据规模较小、数据集不平衡以及存在自然高摄取区域的问题，本节设计了一个数据预处理流程，如图3.3所示。通过该流程将动态骨显像转换成有效信息。具体流程如下：

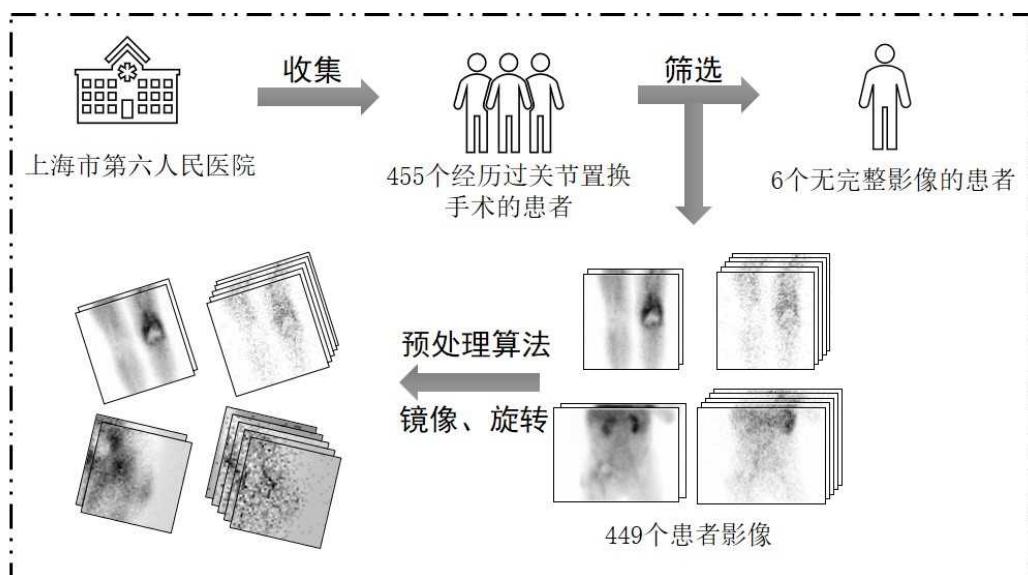


图 3.3 动态骨显像的预处理流程

Figure 3.3 Preprocessing flow of dynamic bone imaging

(1) 数据清洗与标注：对收集的经历过关节置换手术的患者数据进行清洗，并且进行类别、干扰区域、感兴趣区域的标注；

(2) 预处理算法：需要将获取的原生数据转换成可有效区分的数据，去除掉自然高摄取区域的干扰；

(3) 数据增强：通过镜像与旋转生成新的数据样本，来增加数据的多样性，以缓解数据集过小的问题。

### 3.2.1 数据清洗与标注

本章所采集的动态骨显像数据都是以 DICOM (Digital Imaging and Communications in Medicine) 格式存储。DICOM 被广泛用于存储、交换和传输医学图像。作为现代放射成像的核心，DICOM 已被医院和医疗软件行业广泛采用，支持各种不同医疗设备获取的影像模态，如超声、CT 和 MRI 等<sup>[93]</sup>。

数据预处理的第一步是对收集到的患者数据进行清洗，如排除掉影像数据不完整、缺乏完整诊断结果的病例。此外，为了保护病人的个人隐私信息，对 DICOM 格式文件进行脱敏。因此，DICOM 格式文件仅仅保留与图像数据相关的信息。在经过上述操作后，本文作者在专业的核医学医生的指导下使用 ITK-SNAP 软件 (v3.8.0, <http://www.itksnap.org>) 对所有的经历过关节置换手术的患者的动态骨显像进行标注，最终标注结果由专业医生进行了核查。标注根据患者治疗过程中获取的诊断报告为依据，为每一个患者的动态骨显像确定病灶的类别信息和位置信息，还为经历过髋关节置换手术的患者的动态骨显像标注了干扰区域和感兴趣区域，如图3.4所示。

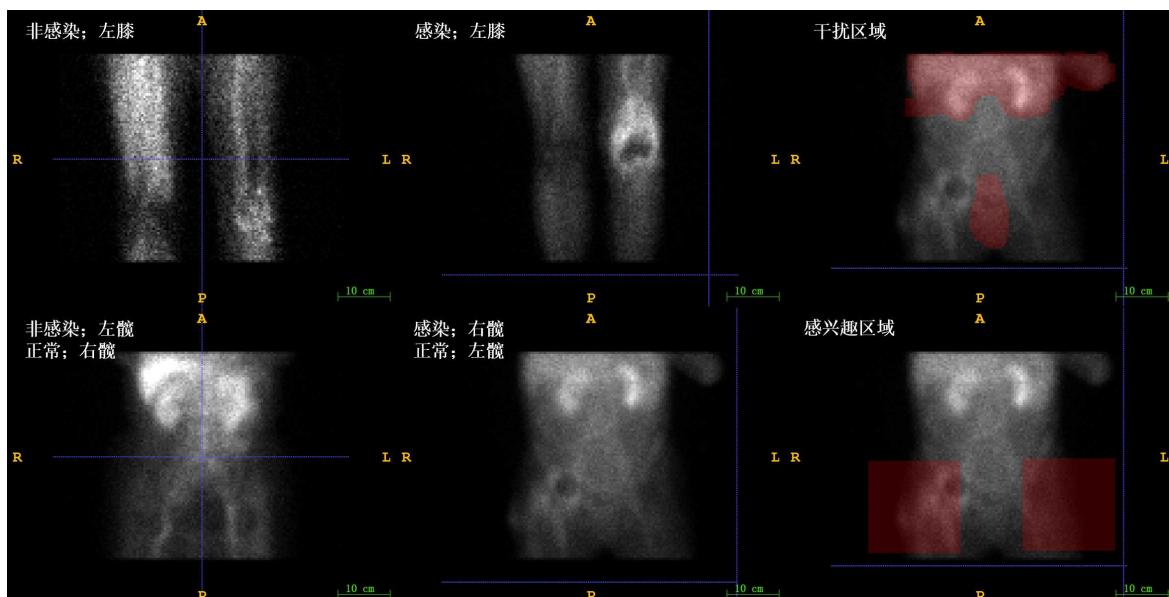


图 3.4 动态骨显像的数据标注

Figure 3.4 Data annotation for dynamic bone imaging

在数据标注完成之后，将对数据进行进一步的整理工作。首先，去除掉数据集中

存在的重复的影像数据。其次，将所有的影像数据和标注数据转换为 NPZ 格式。其中，NPZ 格式文件是 NumPy<sup>[94]</sup>中的一种以二进制的形式存储的格式文件，可用于以未压缩形式的存储多个 NumPy 数组对象。通过采用 NPZ 格式，将影像数据和对应的标注数据存放在一个文件之中。

### 3.2.2 预处理算法

由于原生的影像数据在不同个体之间的数值波动范围差异大以及存在自然高摄取区域，会极大地影响分类性能。因此，本节提出了三种预处理算法，将动态骨显像的影像数据转换为有效信息。预处理后得到的影像数据通过可视化后如图3.5所示。

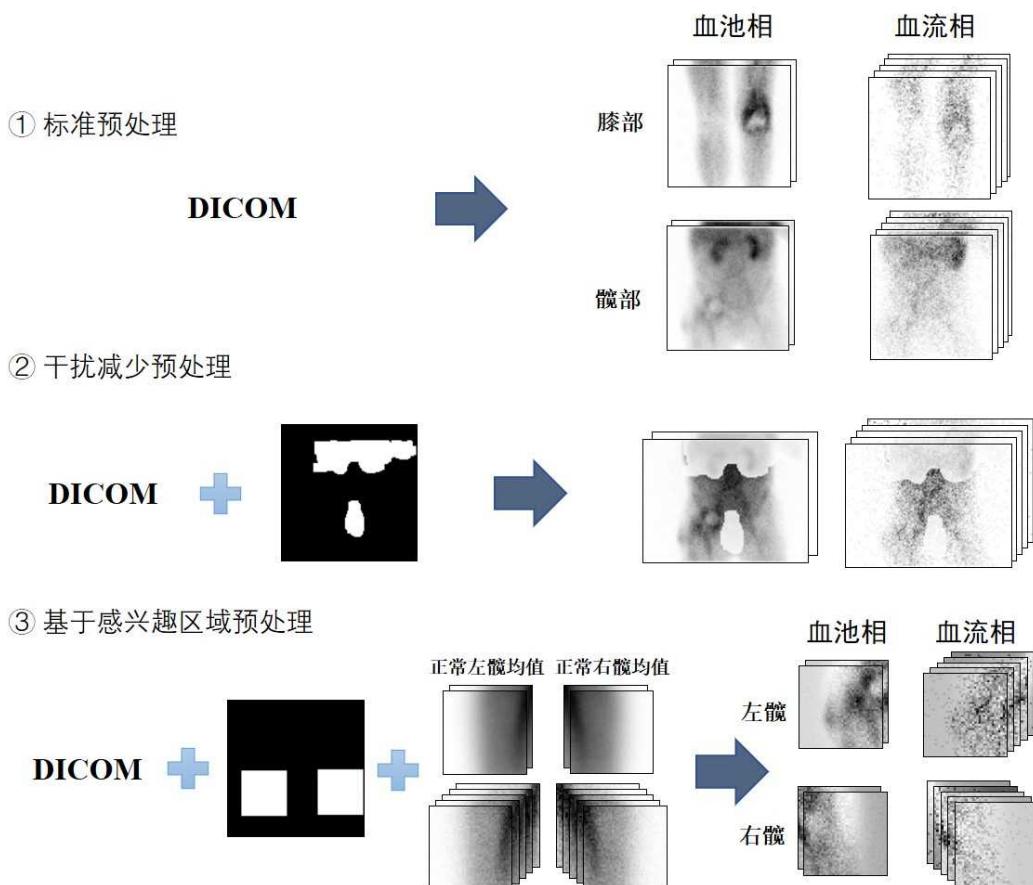


图 3.5 三种预处理算法的结果示意图

Figure 3.5 Schematic of the results of the three preprocessing algorithms

#### (1) 标准预处理

标准预处理的关键步骤就是执行最小最大归一化，通过设定图像数据的下限和上限来调整数据的范围。其基本原理在于病灶区域的上限往往显著高于一般的正常

区域，而一般的正常区域的数值范围相对较为固定，因此通过调整图像数据的上限，能够更好地突显出潜在的病灶区域，提高模型对于病灶特征的感知能力。标准预处理算法过程如下。

读取  $^{99m}\text{Tc}$ -MDP 动态骨显像中的血流相  $F$  和血池相  $P$ 。其中，血流相  $F$  一共有 20 张  $128 \times 128$  的二维图像，而血池相  $P$  一共有 5 张  $128 \times 128$  的二维图像。随后计算出血流相和血池相中的最大值，如公式3.1：

$$\begin{aligned} F_{max} &= \max(\max(F_1), \max(F_2), \dots, \max(F_{20})) \\ P_{max} &= \max(\max(P_1), \max(P_2), \dots, \max(P_5)) \end{aligned} \quad (3.1)$$

最后，分别对血流相和血池相数据进行截断和归一化处理，如公式3.2所示：

$$\begin{aligned} F_i^{m,n} &= \begin{cases} 1, & \text{if } F_i^{m,n} \geq \alpha F_{max} \\ 0, & \text{if } F_i^{m,n} \leq 0 \\ F_i^{m,n}/\alpha F_{max}, & \text{otherwise} \end{cases} \\ P_j^{m,n} &= \begin{cases} 1, & \text{if } P_j^{m,n} \geq \alpha P_{max} \\ 0, & \text{if } P_j^{m,n} \leq 0 \\ P_j^{m,n}/\alpha P_{max}, & \text{otherwise} \end{cases} \end{aligned} \quad (3.2)$$

其中， $1 \leq i \leq 20, 1 \leq j \leq 5, 1 \leq m, n \leq 128$ ， $i, j$  分别表示血流相和血池相中相应序号的图像， $m, n$  分别表示二维图像中像素的行坐标和列坐标， $\alpha \in [0, 1]$  是自定义的超参数，用于调整归一化区间的最大值，默认值为 0.5。

## (2) 干扰减少预处理

干扰减少预处理在标准预处理的基础上，考虑到了自然高摄取区域这类干扰区域，如肾脏、膀胱，与高摄取的病灶区域具有一定的相似性，减少了神经网络对干扰区域与病灶区域的混淆，用以提升分类模型的性能。对于动态骨显像中的多张二维图像数据，每个值表示在相应的人体位置上吸收药物的大致放射强度。该值越大，表示此处药物吸收得多，代表着此处的生理代谢更加活跃。因此，干扰减少预处理的基本原理是降低自然高摄取区域的放射强度值来减少它的干扰能力。干扰减少预处理算法过程如下。

与标准预处理类似，首先读取动态骨显像中的血流相  $F$  和血池相  $P$  并计算出血

流相和血池相中的最大值。然后，使用公式3.3通过干扰区域  $A_{IR}$  来降低自然高摄取区域的放射强度：

$$F_i^{m,n} = \begin{cases} \alpha_{IR} F_i^{m,n}, & \text{if } (m, n) \in A_{IR} \\ F_i^{m,n}, & \text{otherwise} \end{cases} \quad P_j^{m,n} = \begin{cases} \alpha_{IR} P_j^{m,n}, & \text{if } (m, n) \in A_{IR} \\ P_j^{m,n}, & \text{otherwise} \end{cases} \quad (3.3)$$

其中， $\alpha_{IR} \in [0, 1]$  表示对干扰区域的衰竭程度，默认值为 0.4。最后通过标准预处理的公式3.2对血流相和血池相数据进行处理。

### (3) 基于感兴趣区域预处理

基于感兴趣区域预处理来源于核医学医生在诊断过程中通过分析正常髋部与关节置换手术后的髋部之间的差异做出决策。由此，通过基于感兴趣区域处理后将得到差异特征，不同于标准预处理和干扰减少预处理得到的强度特征。基于感兴趣区域预处理的算法过程如下。

首先，根据标注大小为  $40 \times 40$  的感兴趣区域，从动态骨显像中将该区域裁剪出来。其次，根据类别信息  $c$  和位置信息  $p$ ，计算正常的左髋和右髋的平均值，如公式3.4所示：

$$\begin{aligned} FNL_i^{m,n} &= \text{mean}\left(\sum_{c_i=\text{正常}, p_i=\text{左}} F_i^{n,m}\right) & PNL_i^{m,n} &= \text{mean}\left(\sum_{c_i=\text{正常}, p_i=\text{左}} P_i^{n,m}\right) \\ FNR_i^{m,n} &= \text{mean}\left(\sum_{c_i=\text{正常}, p_i=\text{右}} F_i^{n,m}\right) & PNR_i^{m,n} &= \text{mean}\left(\sum_{c_i=\text{正常}, p_i=\text{右}} P_i^{n,m}\right) \end{aligned} \quad (3.4)$$

其中， $i, j, m, n$  与公式3.2一致， $FNL, PNL$  分别表示正常左髋血流相和血池相的平均值， $FNR, PNR$  分别表示正常右髋血流相和血池相的平均值。最后，对于其他的经历过关节置换手术的兴趣区域的血流相和血池相  $FROI, PROI$ ，将根据公式3.5进行计算，获得与正常髋部之间的差异特征：

$$\begin{aligned} FROI_i^{m,n} &= \begin{cases} FROI_i^{m,n} - FNL_i^{m,n}, & \text{if } p = \text{左} \\ FROI_i^{m,n} - FNR_i^{m,n}, & \text{if } p = \text{右} \end{cases} \\ PROI_i^{m,n} &= \begin{cases} PROI_i^{m,n} - PNL_i^{m,n}, & \text{if } p = \text{左} \\ PROI_i^{m,n} - PNR_i^{m,n}, & \text{if } p = \text{右} \end{cases} \end{aligned} \quad (3.5)$$

其中， $p$  表示经历过关节置换手术的髋部的位置信息。

### 3.2.3 数据增强

由于数据集规模较小，为了补充数据集，本节采用了镜像和旋转的方式来生成新的训练样本。镜像将对整个动态骨显像中血流相和血池相的图像进行相同的水平翻转，同时旋转也是对整个动态骨显像中血流相和血池相的图像从 $-20^{\circ}$ 到 $20^{\circ}$ 中随机选取一个角度进行旋转。数据增强采用在线策略，即在训练过程中随机地应用在预处理后得到的强度特征或者差异特征。在本章研究中，所有的数据增强方法都将以50%的概率进行采用。

## 3.3 DBS-eNet

动态骨显像拥有两个具有时序的血流相和血池相，它们展示了一个患者在注射药剂后不同时间点下的状态。由于CNN和ConvLSTM在提取特征能力上拥有各自的适用领域，本章提出了结合了CNN和ConvLSTM的DBS-eNet。该模型设计了卷积对输入的特征进行下采样和图像特征提取，有助于减少模型的计算量和参数量，并且采用了ConvLSTM继续丰富图像特征并提取时序特征。卷积和ConvLSTM的组合可以增强对动态骨显像的特征提取能力，用以提供图像以及时序上的混合特征信息。

### 3.3.1 模型结构

根据动态骨显像是具有时序性的单通道二维图像序列，针对性地构建了基于深度学习和多特征的有效分类网络DBS-eNet的三个部分，模型结构如图3.6所示。

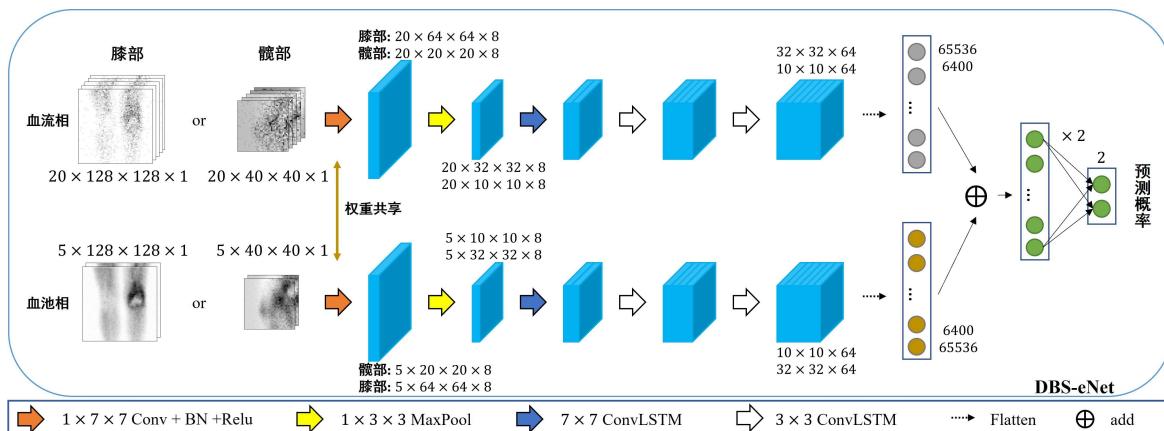


图 3.6 基于深度学习的多特征神经网络模型 DBS-eNet

Figure 3.6 DBS-eNet: A multi-feature neural network based on deep learning

第一部分是一个共享权重且用于提取生理摄取形态特征的  $1 \times 7 \times 7$  的三维卷积层、批归一化层 (Batch Normalization)<sup>[51]</sup>、校正线性单元 (Rectified Linear Unit, ReLU) 和用于下采样的最大池化层 (Max Pooling) 组成。其中， $1 \times 7 \times 7$  三维卷积层是时序卷积，在时序维度上权重的尺寸为 1，等价于对每一张图像进行特征提取。又由于血流相和血池相是一个患者在不同时间点下的状态，即在血流相和血池相上病灶区域处于同一位置。因此，时序卷积设定为共享权重的形式，去提取血流相和血池相上每一张图像中同一病灶的特征，同时减少参数量。批归一化层用在卷积层之后可以避免数据分布发生变化，用以保留网络各层在训练过程中的学习成果。ReLU 激活函数可以有效解决梯度消失的问题并加快模型的收敛速度。最大池化层则可以减少特征的维度的同时提取出特征中更好更强的语义信息。第一部分这样的设计，可以在减少参数量和计算量的同时保持特征的主要语义信息。

第二部分是一个由一层  $7 \times 7$  的 ConvLSTM 和两层  $3 \times 3$  的 ConvLSTM 组成。这一设计旨在图像处理的过程中更加全面地捕捉和丰富生理形态摄取特征并且捕获序列之中生理代谢差异变化特征。首先， $7 \times 7$  的 ConvLSTM 通过更大的感受野去丰富影像中病灶的摄取形态特征。接着，两层  $3 \times 3$  的 ConvLSTM 以较小的卷积核去捕捉到病灶摄取形态中的各种细节和纹理特征。最后，时序感知的 ConvLSTM 在更好地理解单一图像内容的基础上，还能够在图像序列中建模和捕捉时间上的关联性，从而更全面地分析病灶区域中药剂的摄取形式和程度变化。因此，第二部分在图像特征和时序特征上取得了良好的平衡，为整个网络的性能提供了强大的表达能力。

第三部分是特征融合层和两层全连接层，融合来自于血流相和血池相中提取的图像特征和时序特征并计算诊断结果的概率。首先，特征融合层通过逐元素地融合来更全面准确地反映图像序列的复杂特征。接着，第一层全连接层引入 GELUs (Gaussian error linear units) 和 Dropout<sup>[44]</sup>。这不仅有助于引入非线性元素，增强网络对特征的表达能力，同时还一定程度上减轻了过拟合的风险。最后，第二层全连接层的输出经过 Dropout 层的处理，进一步提升了网络的泛化能力。第三部分的设计有助于确保网络在处理两个相的时序图像特征时保持稳健性，提供可靠而精准的结果。

### 3.3.2 ConvLSTM

ConvLSTM 是 DBS-eNet 网络中的核心模块。由于动态骨显像是一个图像序列，重要的不仅是图像特征，还需要捕获整体图像序列之间的时序特征。因此，与普通

的卷积神经网络相比，卷积与 LSTM 相结合的 ConvLSTM 能够捕获到动态骨显像中的时空特征，更加有效地处理复杂的时序信息，从而使网络在分析动态骨显像时表现得更为出色。

在常见的序列建模模型中，LSTM<sup>[54]</sup>作为一种特殊的 RNN 结构，专门设计来解决传统 RNN 中存在的梯度消失或梯度爆炸的问题，且已经被证明在建模远程依赖关系方面是稳定而强大的。通过记忆单元和门控机制来控制信息的流动，可以将梯度留在记忆单元之中，防止梯度太快消失，以更有效地捕捉长期依赖关系。LSTM 的关键公式如3.6所示：

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \text{Tanh}(W_c x_t + U_c h_{t-1} + b_c) \\ h_t &= o_t \odot \text{Tanh}(c_t) \end{aligned} \quad (3.6)$$

其中， $i_t, f_t, o_t, h_t, c_t$  分别表示  $t$  时刻下的输入们、遗忘门、输出门、隐藏状态和记忆单元， $\odot$  表示逐元素相乘。多个 LSTMs 可以堆叠起来形成更加复杂的结构，用以解决更多更复杂的序列建模任务。

尽管 LSTM 可以解决序列建模任务，但是无法处理更加复杂的序列问题。为了解决这个问题，ConvLSTM 在 LSTM 的基础上进行了扩展，将输入到状态和状态到状态的转换过程修改为卷积计算，关键公式如3.7所示：

$$\begin{aligned} i_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \odot \mathcal{C}_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \odot \mathcal{C}_{t-1} + b_f) \\ \mathcal{C}_t &= f_t \odot \mathcal{C}_{t-1} + i_t \odot \text{Tanh}(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \odot \mathcal{C}_t + b_o) \\ \mathcal{H}_t &= o_t \odot \text{Tanh}(\mathcal{C}_t) \end{aligned} \quad (3.7)$$

其中， $*$  表示为卷积计算， $\odot$  表示逐元素相乘，输入  $\mathcal{X}_t$ 、记忆单元的输出  $\mathcal{H}_t$  和状态信息  $\mathcal{C}_t$  都是二维数据，即空间维度。为了更好地理解输入和状态信息，可以将它们视为一个空间网格。ConvLSTM 通过输入和状态信息的局部邻域来确定网格中某个位置的未来状态。这可以很简单地通过卷积操作来实现状态到状态和输入到状态的

转变。为了确保状态信息在计算过程中空间大小不变且与输入相同，在进行卷积计算之前需要进行填充（padding）。

### 3.3.3 损失函数

损失函数用于假体关节感染任务中的作用是评估 DBS-eNet 结果的准确程度。由于任务本身收集的髋部数据中存在类别不平衡的情况，不同类别的样本数量差异较大，这可能导致模型学习过程中存在一定的偏差。由于感染与非感染的样本分布不均匀，模型可能更容易偏向于在数量中占主导地位的类别，而对于少数类别的样本学习效果相对较弱。为了解决这一问题，本章引入并设计了带有标签平滑（Label smoothing）<sup>[89]</sup>的  $\alpha$  平衡版的 Focal Loss<sup>[88]</sup>用于髋部动态骨显像的差异特征，解决类别不平衡与类别之间的相似性，如公式3.8所示：

$$\begin{aligned} \tilde{y}_{n,c} &= \text{Softmax}(x_{n,c}) = \frac{\exp(x_{n,c})}{\sum_{i=1}^C \exp(x_{n,i})} \\ y_{n,c} &= \begin{cases} 1 - \epsilon & \text{if } c = \mathcal{C} \\ \frac{\epsilon}{C-1} & \text{if } c \neq \mathcal{C} \end{cases} \\ \mathcal{L}_{fl}(x, y) &= -\frac{1}{N} \sum_{n=1}^N \alpha_{n,c} \cdot (1 - \sum_{c=1}^C \tilde{y}_{n,c} y_{n,c})^\gamma \cdot \log \sum_{c=1}^C \tilde{y}_{n,c} y_{n,c} \end{aligned} \quad (3.8)$$

其中， $x_{n,c}$  是模型的输出结果，表示样本  $n$  对于类别  $c$  的预测值； $\tilde{y}_{n,c}$  是  $x_{n,c}$  经过 Softmax 后的预测结果； $C$  表示类别数量； $y_{n,c}$  是标签平滑后的真实标签值； $\epsilon$  是平滑参数； $\mathcal{C}$  表示正确的类别； $\alpha_{n,c}$  表示样本  $n$  对于所属类别  $\mathcal{C}$  的权重因子； $\gamma \geq 0$  表示可调聚焦参数； $N$  表示批大小。

而对于类别更为平衡的膝部动态骨显像的强度特征，则采用了如公式3.9所示的加权交叉熵损失函数：

$$\begin{aligned} y_{n,c} &= \begin{cases} 1 & \text{if } c = \mathcal{C} \\ 0 & \text{if } c \neq \mathcal{C} \end{cases} \\ \mathcal{L}_{ce} &= -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C w_c \log \tilde{y}_{n,c} y_{n,c} \end{aligned} \quad (3.9)$$

其中， $\tilde{y}_{n,c}, y_{n,c}, C, \mathcal{C}, N$  同上式3.8一致； $w_c$  表示对于类别  $c$  的权重因子。

### 3.3.4 可视化模块

为了了解 DBS-eNet 的决策基础，采用了 Grad-CAM 和 t-SNE。Grad-CAM 是一种用于解释深度学习模型预测的可视化技术，特别适用于图像分类任务。其基本原理是在模型通过前向传播预测后计算反向传播的梯度，随后通过对梯度进行全局平均池化得到相应特征图的权重生成类激活图。本章将 Grad-CAM 应用在框架中的 DBS-eNet 最后一个 ConvLSTM 层上来计算类激活图。该类激活图可以显示出动态骨显像中不同区域的重要性。t-SNE 是一种降维和可视化高维数据的技术。其核心思想是通过在低维空间中保留原始数据点之间的相似性关系，将高维数据映射到一个更易于理解和可视化的低维空间。本章通过采用保持不同样本之间相对距离的 t-SNE 方法，将 DBS-eNet 中提取的最后的特征向量（1024 维）降至 2 维，并将其可视化在二维散点图上来显示出该特征向量的有效性和可区分性。其中，每个点代表在特征空间中的一个感染者或者非感染者的样本。

## 3.4 实验

### 3.4.1 实验环境

本章中的所有实验均在 PAI 上海大学机器学习平台上进行。该平台是由上海大学计算机学院设计和构建的多加速器异构集群。本章实验的软硬件配置如表3.1所示。PAI 平台是上海大学首台采用自主开发的基于容器的 PAI 管理系统的异构计算平台，

表 3.1 实验配置

Table 3.1 Experimental setup

环境	配置
操作系统	Linux Ubuntu 20.04.2
CPU	Intel(R) Xeon(R) Gold 5115 CPU @ 2.40GHz
GPU	NVIDIA Tesla V100, 32GB
内存	128GB
编程语言	Python 3.8.17
深度学习框架	Pytorch 1.8.0

专注于支持机器学习和深度学习的科研和教学活动。自 2018 年 6 月以来，该平台一直提供稳定的计算服务。在硬件方面，该平台包括 2 个登录/管理节点、1 套 KNL 计

算节点、2个FPGA节点、10个双GPU节点、1台配备四个GPU的DGX-station、4个I/O节点，以及一套200T光纤存储阵列，同时配备有1套100G IB高速网络、1套100G OPA高速网络和1套千兆管理网络。这一多样化的硬件配置为各种计算需求提供了灵活的支持。平台的管理软件由上海大学计算机学院自主开发完成，学生们负责相应的软件维护和升级工作。PAI平台不仅面向全校师生提供了丰富的教学资源，还为各类计算任务提供了强大的计算服务。

### 3.4.2 数据集

本文作者与上海市第六人民医院核医学科进行合作构建了假体关节感染动态骨显像数据集。该数据集包含髋关节置换手术和膝关节置换手术后患者的动态骨显像，分类详情见表3.2所示，一共纳入了从2016年1月至2021年6月之间455例具有最终诊断结果的患者，其中258名患者经历了髋关节置换手术，197名患者经历了膝关节置换手术。在查看了所有患者的影像后，其中6例患者由于影像不全被排除。最终纳入并分析了449例患者。所有患者在经历了膝关节置换手术或髋关节置换手术后均出现疼痛或关节功能障碍，且都进行了翻修手术。假体关节感染的最终诊断结果主要基于2021年的EBJIS诊断指南的定义<sup>[95]</sup>。在本章的数据集中，一共有192例感染患者和257例非感染患者。

表 3.2 动态骨显像的分布情况

Table 3.2 Distribution of dynamic bone scintigraphy

标签	髋部	膝部
感染	99	93
非感染	156	101

### 3.4.3 评估指标

假体关节感染诊断任务属于分类问题，本章选用了分类任务中常用的评估指标，如准确率(Accuracy, Acc)、特异率(Specificity, Spec)、敏感率(Sensitivity, Sen)、F1值(F1 Score)，阴性预测值(Negative Predictive Value, NPV)、阳性预测值(Positive Predictive Value, PPV)和接收者操作特征曲线下面积(Area Under the receiver operating characteristic Curve, AUC)。其中，准确率表示对假体关节感染预测准确的样本在总

样本的占比，但在样本不均衡的情况下，不能够很好地评估模型性能的优劣。特异率表示对实际不患有假体关节感染的患者诊断为未感染的概率。敏感率表示对实际上患有假体关节感染的患者诊断为感染的概率。阳性预测值表示为被诊断为感染的患者实际上患有假体关节感染的概率。阴性预测值表示为被诊断为未感染的患者实际上未患有假体关节感染的概率。F1 值是阳性预测值和敏感率的综合评价指标。AUC 是敏感性和特异性的综合评价指标。它们的计算公式3.10所示：

$$\begin{aligned}
 Acc &= \frac{TP + TN}{Total} & F1 &= \frac{2 \times PPV \times Sen}{PPV + Sen} \\
 Spec &= \frac{TN}{TN + FP} & Sen &= \frac{TP}{TP + FN} \\
 NPV &= \frac{TN}{TN + FP} & PPV &= \frac{TP}{TP + FP} \\
 AUC &= \frac{\sum_{i \in P} rank_i - \frac{M(1+M)}{2}}{M \times N}
 \end{aligned} \tag{3.10}$$

其中， $TP, TN, FP, FN$  分别是真阳性、真阴性、假阳性、假阴性样本的数量； $Total$  表示样本的整体数量； $P$  表示阳性的类别； $M, N$  分别是阳性样本和阴性样本的数量； $rank_i$  是第  $i$  个样本的序号，所有的预测样本根据阳性预测值从小到大进行排序。

此外，实验中采用了 Bootstrap<sup>[96]</sup>来计算这些评估指标的置信区间，并采用麦克内马尔检验（McNemar test）<sup>[97]</sup>进行统计学分析，来验证 DBS-eNet 与专业的核医学医生之间的差异性。更进一步地采用了校准曲线（Calibration Curve）和决策曲线分析（Decision Curve Analysis, DCA）<sup>[98]</sup>来更加全面地评估本章框架的潜在应用价值和有效性。校准曲线可以描绘模型输出的概率与实际观察到的事件发生概率之间的关系，在本章研究中用于评估 DBS-eNet 预测假体关节感染的概率与实际假体关节感染的概率之间的一致性。决策曲线分析是一种用于评估模型在不同概率阈值下其决策效果的方法。它通常应用于医学领域，特别是在评估临床预测模型的效果。决策曲线分析有助于确定模型在不同概率阈值下相对于其他决策策略的净收益，并为医学决策提供了实用的信息。

### 3.4.4 实验结果

实验中，对于构建的动态骨显像数据集先进行上述中的数据预处理后，可以从经历了髋关节置换手术的患者的动态骨显像中获得差异特征（ $25 \times 40 \times 40$ ）以及从经历了膝关节置换手术的患者的动态骨显像中获得强度特征（ $25 \times 128 \times 128$ ），用于

模型训练。对于模型的训练配置，采用了随机梯度下降（Stochastic Gradient Descent, SGD）优化器，学习率设置为 0.01，动量（momentum）为 0.8，权重衰减（weight decay）为 0.0005。批大小（batch size）设置为 4，Dropout 的比例设置为 0.2，以及整个训练周期（epoch）为 80。在训练周期的一半和四分之三，学习率将乘以 0.1。

为了评估不同预处理算法在髋部动态骨显像的效果，进行了三种预处理的对比实验，如表3.3所示。基于感兴趣区域预处理的效果最佳，优于干扰减少预处理与标准预处理。这是由于基于感兴趣区域预处理，相比较于标准预处理去掉了自然高摄取区域以外，还使用差异特征替代强度特征，从而避免了在数据集较小的情况下，自然高摄取区域给神经网络带来的偏差干扰。干扰减少预处理则是由于简单但不自然地减弱自然高摄取区域强度的方式，造成了动态骨显像上存在断崖式的强度变化，从而导致了模型的综合性能下降。

表 3.3 髋部动态骨显像上不同预处理算法之间的比较

Table 3.3 Comparison between different preprocessing methods on hip dynamic bone scintigraphy

准确率 (%)	五折交叉验证	独立验证
标准预处理	76.23	73.21
干扰减少预处理	69.32	68.83
基于感兴趣区域预处理	86.33	86.36

DBS-eNet 的综合性能实验结果如表3.4所示。在五折交叉验证中，对假体膝关节感染的准确率和 AUC 值分别为 86.48% (80.98%, 91.41%) 和 0.889 (0.825, 0.933)，对假体髋关节感染的准确率分别为 86.33% (81.50%, 90.75%) 和 0.866 (0.809, 0.911)。在独立验证中，对假体膝关节感染的准确率和 AUC 值分别达到 87.74% (82.58%, 92.90%) 和 0.957 (0.905, 0.973)，对假体髋关节感染的准确率分别达到 86.36% (81.81%, 90.91%) 和 0.906 (0.859, 0.946)。这表明了 DBS-eNet 具有较高的综合性能。

为了更好地评估 DBS-eNet，实验中采用了五折交叉验证的方式对比了几个主流的卷积神经网络（VGG<sup>[45]</sup>，ResNet<sup>[47]</sup>，DenseNet<sup>[48]</sup>和 ConvNeXt<sup>[49]</sup>）。为了将卷积神经网络应用于预处理后的动态骨显像数据，首先将预处理得到的血池相和血流相数据重采样到  $64 \times 64 \times 64$ 。其次，二维卷积神经网络的卷积操作和下采样操作都修改为相应的三维操作。最后，卷积神经网络在卷积部分采用了共享权重，然后提取的特征进行融合并放入全连接层中得到预测结果。

表 3.4 DBS-eNet 在五折交叉验证和独立验证上的性能表现

Table 3.4 the Performance of DBS-eNet on five-fold cross validation and independent validation

	Acc(%)	Spec(%)	Sen(%)	F1	PPV(%)	NPV(%)	AUC
<b>假体膝关节感染</b>							
五折交叉验证	86.48 (80.98, 91.41)	83.53 (75.00, 91.25)	89.67 (82.61, 95.89)	0.8631 (0.8050, 0.9167)	84.29 (75.31, 90.81)	90.83 (82.56, 95.83)	0.889 (0.825, 0.933)
独立验证	87.74 (82.58, 92.90)	86.25 (78.31, 93.51)	89.33 (82.19, 95.65)	0.8772 (0.8182, 0.9290)	86.60 (77.52, 93.42)	89.66 (81.92, 95.78)	0.957 (0.905, 0.973)
<b>假体髋关节感染</b>							
五折交叉验证	86.33 (81.50, 90.75)	92.34 (87.97, 96.38)	76.03 (67.06, 85.14)	0.8026 (0.7341, 0.8639)	86.27 (76.71, 92.86)	87.12 (81.58, 92.31)	0.866 (0.809, 0.911)
独立验证	86.36 (81.81, 90.91)	87.86 (82.14, 93.10)	83.75 (75.31, 91.03)	0.8178 (0.7464, 0.8784)	80.25 (70.93, 88.11)	90.50 (85.50, 94.89)	0.906 (0.859, 0.946)

评估指标的值 (95% 置信区间)

如表3.5所示，可以看出 DBS-eNet 的性能优于其他的卷积神经网络，在膝关节置换手术和髋关节置换手术的五折交叉验证集上都具有最好的准确率、灵敏性、F1 值和阴性预测值，以及良好的特异性和阳性预测值。这表明卷积和 ConvLSTM 的结合相比较于三维卷积更适合于具有时序性的图像序列，能够更有效地捕获动态骨显像中的时空特征信息，有效地提高了模型的分类性能。

表 3.5 DBS-eNet 与分类网络在五折交叉验证中的比较

Table 3.5 Comparison between DBS-eNet and classification networks in five-fold cross validation

	假体膝关节感染						假体髋关节感染							
	Acc	Spec	Sen	F1	PPV	NPV	AUC	Acc	Spec	Sen	F1	PPV	NPV	AUC
VGG11	84.05%	80.00%	88.67%	0.8407	82.76%	90.77%	0.884	79.23%	89.58%	60.74%	0.6112	-	81.95%	0.691
VGG19	81.00%	82.35%	79.92%	0.7764	84.20%	86.13%	0.875	81.93%	95.81%	57.79%	0.6988	<b>89.40%</b>	79.84%	0.851
ResNet18	83.37%	<b>94.12%</b>	71.42%	0.7976	<b>93.04%</b>	79.24%	0.847	72.00%	75.17%	66.84%	0.5923	-	-	0.715
ResNet50	64.36%	52.94%	77.08%	0.6763	64.07%	73.16%	0.601	57.99%	80.00%	20.00%	0.1067	-	-	0.351
DenseNet121	79.72%	81.18%	77.67%	0.7803	83.69%	82.68%	0.844	79.62%	<b>95.17%</b>	52.50%	0.5975	-	78.76%	0.791
ConvNeXt-T	52.16%	100.00%	0.00%	0.0000	-	52.16%	0.493	60.83%	78.62%	29.41%	0.2265	-	-	0.680
ConvNeXt-B	52.16%	100.00%	0.00%	0.0000	-	52.16%	0.432	54.94%	57.24%	50.59%	0.3346	-	-	0.544
DBS-eNet	<b>86.48%</b>	83.53%	<b>89.67%</b>	<b>0.8631</b>	84.29%	<b>90.83%</b>	<b>0.889</b>	<b>86.33%</b>	92.34%	<b>76.03%</b>	<b>0.8026</b>	86.27%	<b>87.12%</b>	<b>0.866</b>

此外，本章在独立测试集上将 DBS-eNet 与 3 位具有 10 年以上诊断经验的核医学专家进行比较。为了排除其他因素（例如医生可能接触到某些患者病例），只保留独立测试集中的图像并随机打乱，三位核医学专家独立通过各自的临床经验得出诊断结果。比较结果如表3.6所示，DBS-eNet 在独立测试集上假体膝关节感染的准确率为 87.74%，高于三位核医学医生的 70.97%、74.19%、74.19%。相比之下，DBS-eNet

表 3.6 DBS-eNet 在独立测试集上与专业医生的比较

Table 3.6 Comparison of DBS-eNet with specialist physicians on the independent test set

	假体膝关节感染						假体髋关节感染					
	Acc	Spec	Sen	F1	PPV	NPV	Acc	Spec	Sen	F1	PPV	NPV
医生 1	70.97%	43.75%	100.00%	0.7692	62.50%	100.00%	86.36%	89.29%	81.25%	0.8125	81.25%	89.29%
医生 2	74.19%	56.25%	93.33%	0.7778	66.67%	90.00%	86.36%	85.71%	87.50%	0.8235	77.78%	92.31%
医生 3	74.19%	50.00%	100.00%	0.7895	65.22%	100.00%	86.36%	85.71%	87.50%	0.8235	77.78%	92.31%
DBS-eNet	<b>87.74%</b>	<b>86.25%</b>	89.33%	<b>0.8772</b>	<b>86.60%</b>	89.66%	<b>86.36%</b>	87.86%	83.75%	0.8178	80.25%	90.50%

的分类性能与三位核医学医生在独立测试集上假体髋关节感染的诊断能力相当，具有相同的准确率和其他相当的评估指标。三位医师诊断结果与 DBS-eNet 对经历了膝关节置换手术的患者的预测结果进行比较， $p$  值分别为 0.004、0.008、0.012，均小于 0.05，表明该差异有统计学意义。对于经历了髋关节置换手术的患者，三位医师诊断结果与 DBS-eNet 之间比较的  $p$  值均为 1，表明该差异无统计学意义。

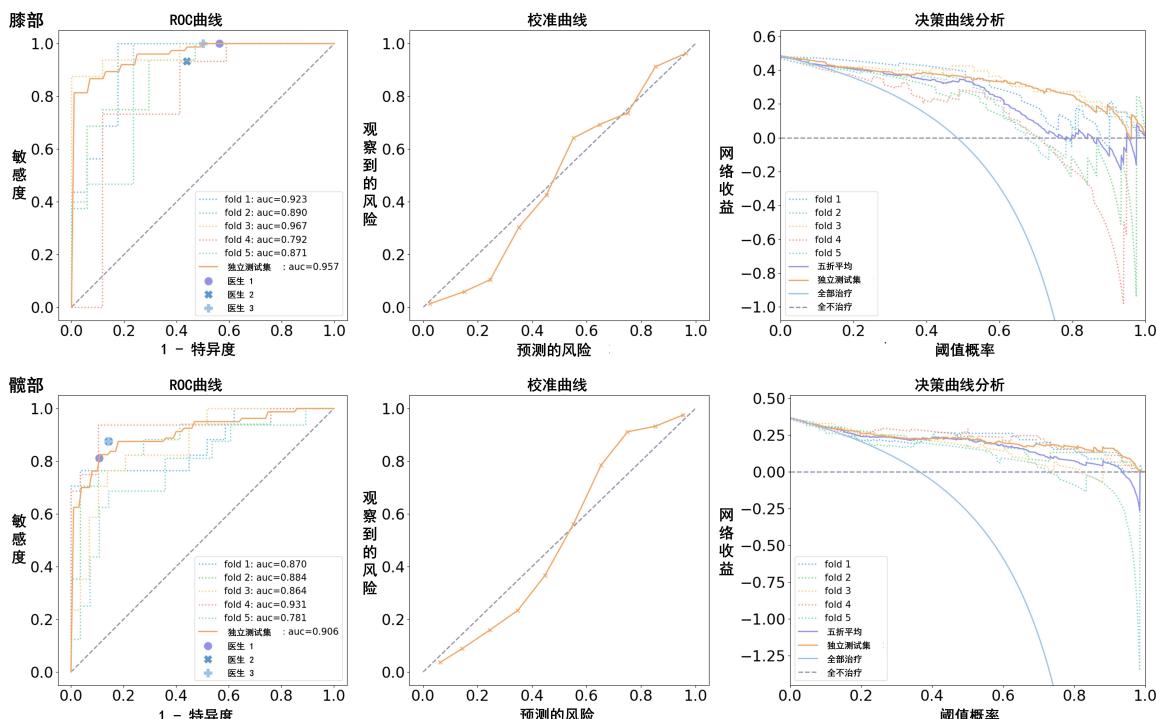


图 3.7 ROC 曲线、校准曲线与决策曲线分析

Figure 3.7 ROC curve, calibration curve and decision curve analysis

图3.7分别展示了本章框架在膝关节置换手术和髋关节置换手术动态骨显像数据集的 ROC 曲线、校准曲线与决策曲线分析。ROC 曲线展示了本章框架与三名专家相当的分类性能。其次，本章提出的框架的校准曲线较为贴近 45 度对角线，这反映了在膝关节置换手术和髋关节置换手术的数据集上假体关节感染的预测概

率与假体关节感染的真实概率具有良好的一致性。此外，本章提出的框架在独立测试集上的决策曲线分析表明，假体关节感染的预测在所有阈值概率下拥有一个较高的净收益，而在五折交叉验证中，存在一到两折在高概率阈值下为负的网络净收益的情况。因此，在大多数情况下，在临床实践中是有价值且有益的。

### 3.4.5 可视化分析

动态骨显像与 Grad-CAM 获得的热力图融合后的图像如图3.8所示。动态骨显像

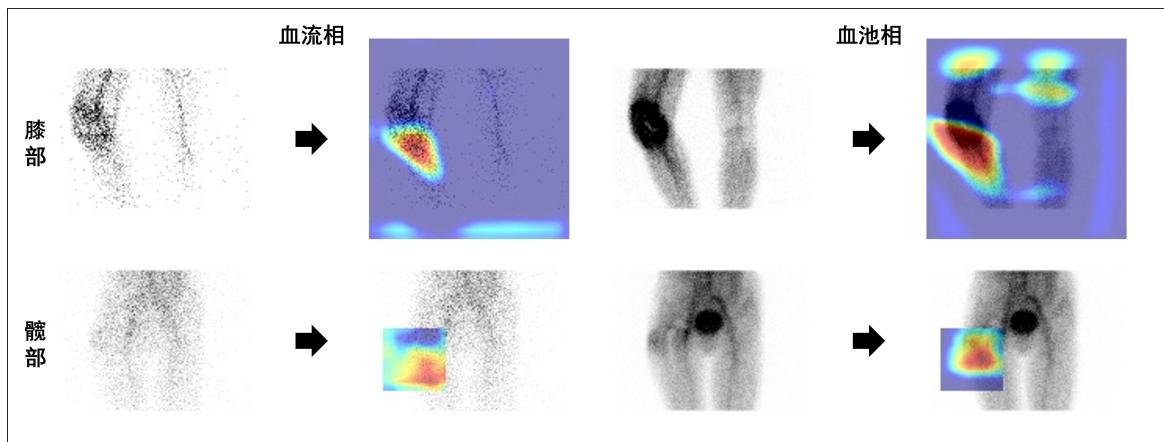


图 3.8 DBS-eNet 的 Grad-CAM 热力图

Figure 3.8 Grad-CAM heatmap of DBS-eNet

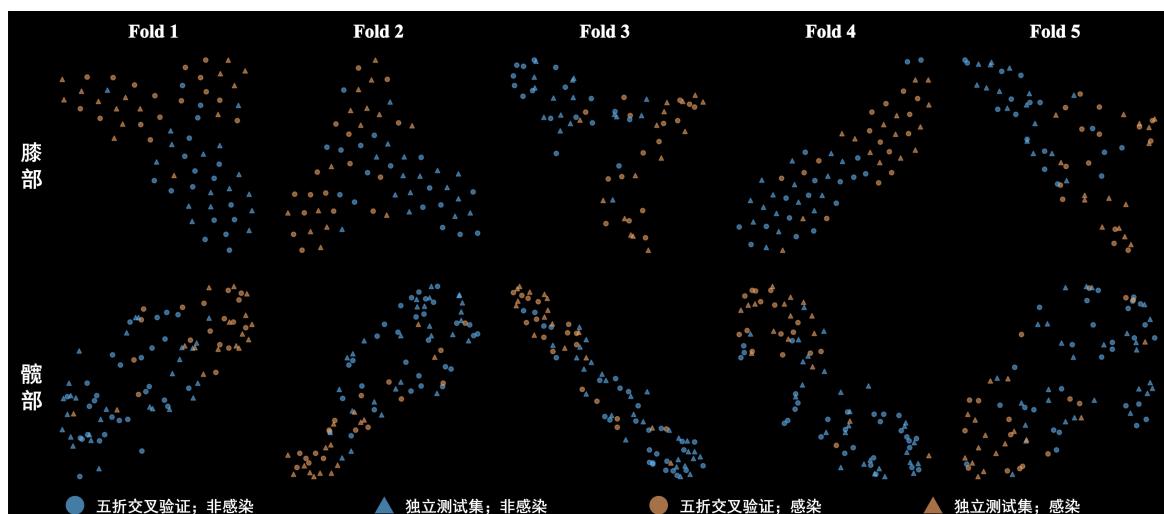


图 3.9 DBS-eNet 的 t-SNE 散点图

Figure 3.9 Scatter plot of t-SNE for DBS-eNet

中血流相和血池相的病灶区域都是重要性最大的区域，这表明 DBS-eNet 在预测这些

影像的最终结果时可以正确聚焦病灶区域，也验证了本章提出的方法符合核医学医生在诊断决策中的判断。

在图3.9中，t-SNE 得到的感染样本和未感染样本的特征点在膝关节置换手术和髋部关节置换手术数据集中都大致地被划分为两簇，体现了 DBS-eNet 提取的特征本身具有较好的可区分性，表明了该框架的良好性能。通过 Grad-CAM 和 t-SNE 方法，可以增加本章提出的框架的可解释性。此外，在对经历了关节置换手术的患者进行诊断时，如果医生之间的判定结果一致性较低，通过可视化框架的决策过程以及总结模型学习到的特定模式和影像学特征，可以为医生提供额外的意见作为参考，不仅有助于提高疾病诊断的准确率和效率，同时为后续的治疗和预后提供了有力的信息支持。

### 3.5 本章小结

本章主要提出了一个基于  $^{99m}\text{Tc}$ -MDP 动态骨显像的假体关节感染辅助诊断框架。该框架包含了获取强度特征的标准预处理、获取差异特征的基于感兴趣区域预处理和用来提取生理摄取形态特征和时序性差异变化特征的 DBS-eNet。实验结果表明，在综合性能表现上，该框架在假体膝关节感染上达到了 86.48% (80.98%, 91.41%) 的准确率和 0.889 (0.825, 0.933) 的 AUC，在假体髋关节感染上达到了 86.33% (81.50%, 90.75%) 的准确率和 0.866 (0.809, 0.911) 的 AUC，优于其他卷积神经网络。在与核医学专家的比较中，该框架的性能在诊断假体膝关节感染上优于专家，在诊断假体髋关节感染上与专家相当。这揭示了人工智能方法在诊断假体关节感染的可行性和诊断框架的有效性。在可视化结果中，该框架在动态骨显像上关注区域与专家相似，并且提出的特征具有明显的可区分性。这表明它可以有效辅助医生进行诊断决策，提升诊断质量。

## 第四章 基于<sup>18</sup>F-FDG PET/CT 三维影像的下肢骨折相关感染检测与诊断框架

上一章介绍了应用于动态骨显像中进行假体关节感染诊断的辅助诊断框架，本章将根据 PET/CT 影像中骨折相关感染的诊断流程，研究患者下肢潜在病灶的检测与诊断，并提出了全自动两阶段的下肢骨折相关感染检测与诊断框架。第一阶段通过双分支三维定位网络检测病灶区域，第二阶段通过卷积神经网络进一步地进行病灶分类，预测是否存在感染。检测部分通过 PET 主分支和 CT 辅分支的网络结构设计有效地提取两种不同模态的特征并进行逐阶段地充分融合生理代谢信息和解剖组织信息。诊断分类部分对检测的病灶进行最大强度投影转换为二维影像，在降低数据和模型规模的同时保留有效特征，通过分类器计算得到预测结果。该框架可以较为准确地检测出病灶并拥有较高的诊断分类准确率。

### 4.1 问题分析

在应对 PET/CT 三维影像中下肢骨折相关感染的诊断任务时，需综合考虑医生对于骨折相关感染的诊断流程、PET/CT 影像的特性，以及人工智能在该诊断任务中的应用可能面临的问题，可以概括如下：

(1) 缺乏关于下肢骨折相关感染的公开数据集。目前网络上暂无公开的关于骨折相关感染的数据集。由此，本文作者与上海市第六人民医院核医学科医生合作收集了来源于真实患者下肢的 PET/CT 影像，并构建了与真实分布一致的下肢骨折相关感染的 PET/CT 影像多模态数据集。然而，数据集存在与第三章一样的问题（规模小和不平衡）。这种数据集的特性对于模型的训练提出了额外的挑战。

(2) CT 中骨折相关感染病灶的形状与分布区域多样。在骨折相关感染的临床诊断流程中，首要任务是明确病灶的具体位置。尽管骨折相关感染的病灶位置与骨骼强度相关，但由于人体骨骼广泛分布于全身，与一些组织和器官相比，并不像它们存在于一个相对固定的位置或者范围内。其次，病变的程度也有所差异，可能会涉及骨骼的不同部位以及软组织的破裂和损伤，也可能因个体差异而不同，导致病灶形状多变，如图4.1所示。因此，对于骨折相关感染的病灶位置，是否能够准确地检

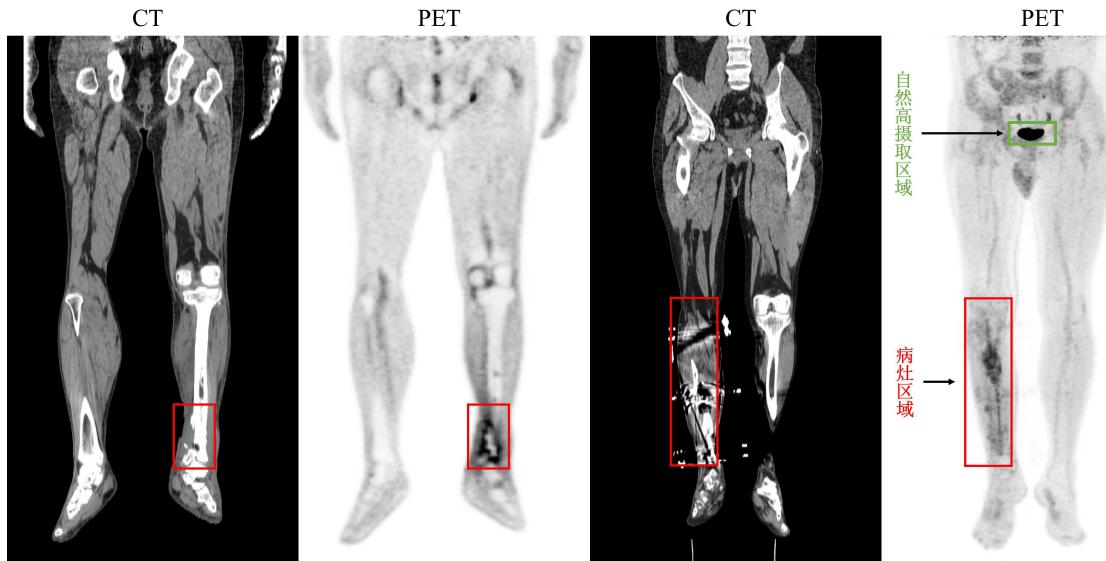


图 4.1 两例患者的 PET/CT

Figure 4.1 PET/CT of two patients

测出至关重要。

(3) PET 中高摄取区域呈现的形式和程度各不相同。从核医学影像的角度分析, PET 影像可以通过人体的组织细胞对于示踪剂氟代脱氧葡萄糖的摄取程度来展示人体的生理代谢信息。生理代谢强的区域一般都有着对于示踪剂的摄取程度高的特点, 但呈现的形式与程度也各不相同。此外, 人体中存在的自然高摄取区域(膀胱、肾脏等), 以及炎症和愈合过程都会导致周围的软组织的摄取能力增加。骨折相关感染的病灶区域也因自身病变程度的不同而展示出各种不同的摄取形式和程度, 如图4.1所示。因此, 如何避免其他高摄取区域造成的混淆是一个本章研究的主要问题。

(4) PET 和 CT 影像之间不同的特征。PET 和 CT 是两种不同的医学影像模态, 各自可以表现不同的内容信息。PET 影像可以展现出人体的生理代谢信息, 而 CT 影像可以展现出人体的解剖组织信息。同时, PET 和 CT 均为三维影像。因此, 充分利用 PET 和 CT 影像的两种不同信息以及它们本身的三维性质同样是一个值得研究的问题。

(5) 病灶检测中正负样本不平衡。在本章的下肢骨折相关感染数据集中, 一个病人一般仅有一个或二个病灶。对于病灶检测任务而言, 所需要检测的目标数量少, 会导致影像中的正负样本数量差异巨大, 这是一个严重的正负样本不平衡问题。此外, 目前的检测方法中, 有基于先验框的方法, 也有不基于先验框的方法。哪种方法更加适用于本章的研究也是值得去考虑的问题。

根据上述的几个问题，本章提出了两阶段的下肢骨折相关感染的检测诊断框架，该框架主要从以下几个方面进行针对性地研究：

(1) 针对数据集的问题，本章设计了一套针对于 PET/CT 影像的数据预处理方法 (PET/CT 配准、去除机床等无关区域)，以及其中的适应性修改的数据增强方法 (翻转、Mixup<sup>[99]</sup> 和 RICAP<sup>[100]</sup>) 能够极大地增加数据集的多样性，缓解数据集较小的问题。经过这样整个预处理流程的 PET/CT 影像可以去除掉较大的无关区域，有效地提升模型的性能和预测结果。

(2) 针对骨折相关感染病灶难以准确检出的问题，设计了一个并行双分支的病灶定位三维网络模型。通过主干网络的并行双分支设计，加强对 PET 和 CT 影像特征的提取和融合能力，有效地学习病灶在生理代谢中摄取能力较高、呈点状或弥漫性的特点，以及在解剖组织中骨骼部位存在缺失、断裂或金属植入物的特点，将丰富的语义信息提供给后续的特征融合子网和预测子网。

(3) 针对高摄取区域之间易产生混淆的问题，本章添加自然高摄取区域的标注，并且通过 CT 影像的补充和注意力机制来排除其他情况带来的干扰。由于自然高摄取区域 (膀胱) 的位置相对固定的特点，相比于病灶区域，更容易让定位网络自主学习到，进而能够有效地减少自然高摄取区域的干扰。此外，通过 CT 影像的补充以及注意力机制，更好地排除掉与骨骼不相关的高摄取区域，进而促进定位性能的提升。

(4) PET 和 CT 影像之间的特征相互补充和相互促进驱使本章引入并设计并行双分支病灶定位网络。其次，考虑到 PET/CT 影像的三维性质，定位网络中的二维计算方式都设计为相应的三维计算方式。这样的设计可以更好地适用于三维影像的结构，以更全面地捕捉影像的立体信息，更准确地学习病灶在深度方向上的分布和形态，从而提高网络对于骨折相关感染任务中的立体病灶的处理效果。

(5) 对于正负样本不平衡的问题，本章研究通过多方面的改进策略进行优化。首先，引入了新的检测目标 (高摄取区域)；其次，在训练过程中，采用了 SimOTA 标签匹配策略，以实现对正样本数量的动态匹配；最后，在病灶定位网络的特征融合网络和预测网络中进行了修剪设计，以缓解正负样本极端不平衡的问题。此外，先验框可以通过人工设计或者在数据集中采用聚类方法生成，人工设计难度较大，而采用聚类方法，在本章研究中存在数据集较小的问题，可能会导致生成的先验框存在一定的偏差。因此，采用无先验框的目标检测方法。

整体流程如图4.2所示。首先，通过最初的数据收集、筛选与标注，最终得到了

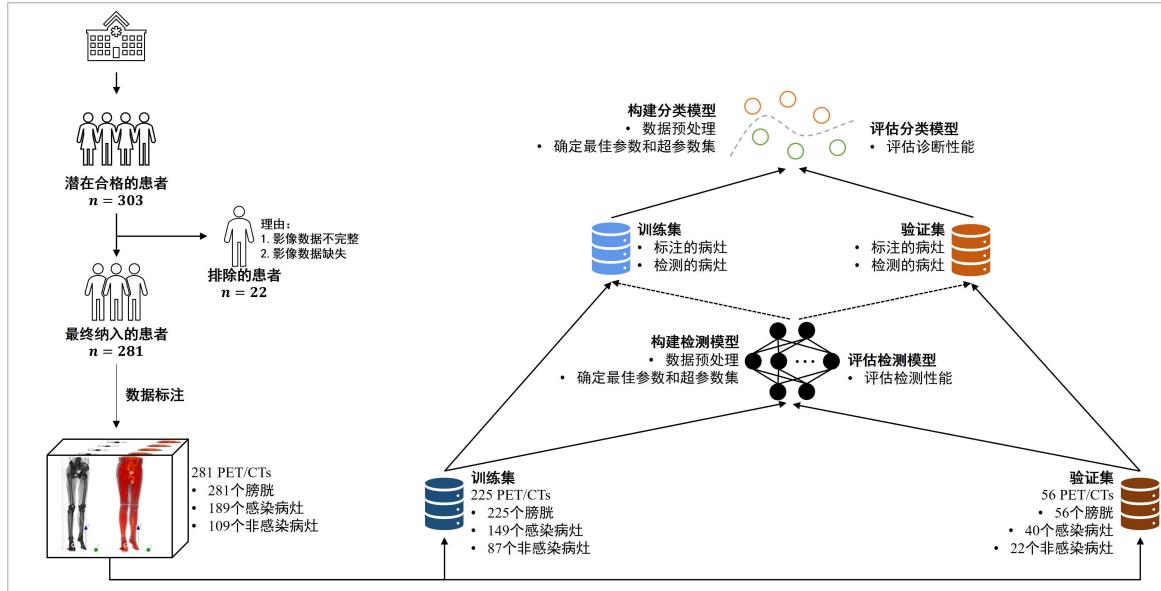


图 4.2 整体研究流程

Figure 4.2 Overall research process

包含了 281 名患者的 PET/CT 影像数据。其中，有 281 个膀胱标注、180 个感染病灶标注和 109 个非感染病灶标注。其次，按照 4: 1 的患者个体数量比例进行划分，得到了包含了 225 个患者 PET/CT 影像数据的训练集（膀胱、感染病灶与非感染病灶：225 vs. 149 vs. 87）和包含了 56 个患者 PET/CT 影像数据的验证集（膀胱、感染病灶与非感染病灶：56 vs. 40 vs. 22）。随后，设计和构建第一阶段的检测模型。通过使用训练集的训练结果，适应性地调整模型结构、超参数和训练策略，并利用验证集来进一步评估该模型的检测性能，获得最终的检测模型和最佳超参数。最后，设计和构建第二阶段的分类模型。利用第一阶段中检测模型在训练集和验证集中的检测结果添加到第二阶段的训练集和验证集中来扩充数据集的规模。同样地使用训练集的训练结果来适应性地调整模型、超参数和训练策略，并利用验证集来进一步评估该模型的分类性能，获得最终的分类模型和最佳超参数。最终二阶段的框架由最佳的检测模型和最佳的分类模型组成。

## 4.2 数据预处理

针对 PET/CT 影像中骨折相关感染的检测诊断任务，本章提出了一套完整的 PET/CT 影像的预处理算法，其具体流程如下：

(1) 数据清洗与标注：清洗收集的患者 PET/CT 影像数据并进行三维目标框及其

类别标注；

(2) PET/CT 的转换与配准：首先，将 CT 影像中的像素值转换为用于测量组织对 X 射线的衰减程度的亨氏单位（Hounsfield Unit, HU），并且将 PET 影像中的像素值转换为用于半量化分析生理代谢活动的标准摄取值（Standardized Uptake Value, SUV）。然后，由于 PET 与 CT 影像的采样分辨率不一致，因此通过配准将 PET 和 CT 影像在物理空间中的原点、体素间距、尺寸大小变成一致。

(3) 人体区域保留：考虑到 PET/CT 影像中非人体区域所占空间较大，并且机床属于纯粹的无关干扰区域，因此可彻底剔除该部分以提高影像质量和解剖学解析度。

(4) 数据增强：通过翻转、适应性修改的 Mixup 和 RICAP (Random Image Cropping And Patching) 增加 PET/CT 影像中的检测目标数量，并扩充数据集的规模。

#### 4.2.1 数据清洗与标注

本章所采集的 PET/CT 同样是以 DICOM 格式存储。首先，对收集到的 PET/CT 进行清洗，例如排除掉不完整或存在缺失的影像数据。此外，为了保护病人的个人隐私信息，对 DICOM 格式文件进行脱敏，去除掉 DICOM 格式文件中有关于病人、医院、医生、设备制造商等等所有隐私数据。在经过上述操作后，本文作者在专业的核医学医生的指导下使用 ITK-SNAP 对所有患者的 PET/CT 进行标注。最终标注结果由专业医生进行了核查。标注根据患者诊断报告为依据，在每一个患者的 PET/CT 影像上采用不同的颜色标注膀胱、感染病灶和非感染病灶，如图4.3所示。

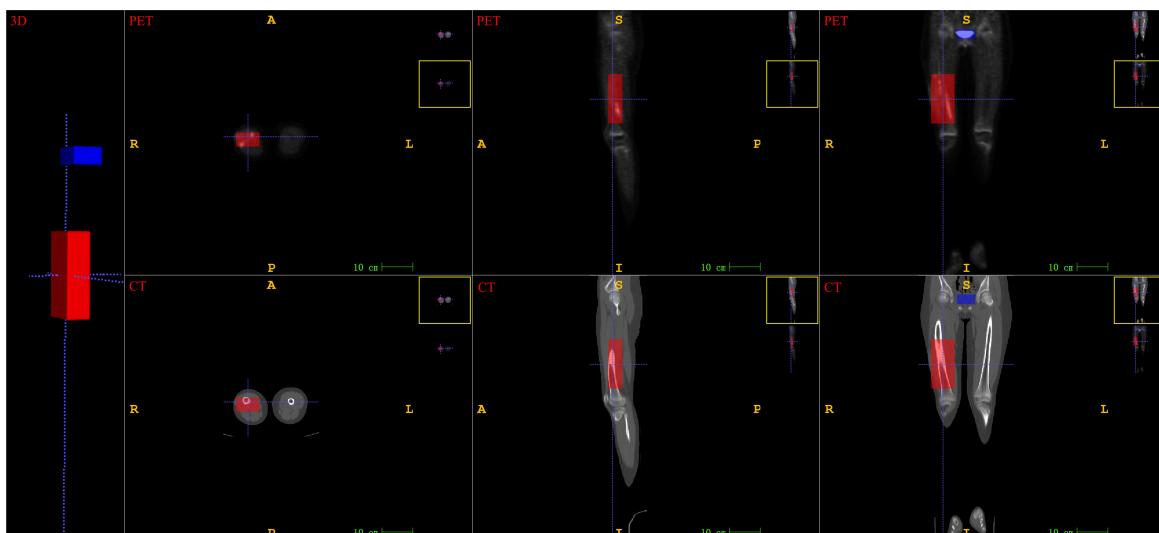


图 4.3 PET/CT 的数据标注

Figure 4.3 Data annotation for PET/CT

#### 4.2.2 PET/CT 的转换与配准

HU 是由 CT 扫描仪测量的人体组织相对于水的相对线性密度值。它提供了一种标定和对比度一致的方法，使得不同的 CT 扫描仪和不同扫描参数下的图像能够具有相似的密度值。此外，不同类型的组织对 X 射线具有不同的吸收能力，它们在 CT 中会展示出不同的亮度。将这些亮度转换为 HU 可以更加直观地表示组织的相对密度值，有助于区分不同的组织，也更方便于医生与研究人员进行更精确的定量分析。由此，本章将以 DICOM 格式存储的 CT 影像的像素值转换为 HU 值。

读取 CT 的 DICOM 文件，提取该文件中所需要的元数据：标签为 (7FE0,0010) 且关键字为 Pixel Data 的像素值  $P_{ct}$ 、标签为 (0028,1052) 且关键字为 Rescale Intercept 的截距  $b$  和标签为 (0028,1053) 且关键字为 Rescale Slope 的斜率  $m$ 。最后，通过公式4.1转换为 HU 值：

$$HU = m \times P_{ct} + b \quad (4.1)$$

SUV 是一种用于衡量 PET 图像中肿瘤摄取的相对标准单位，通常用于定量分析和治疗反馈。SUV 可以减少多种因素带来的影响，如患者体重、注射剂量等等，使得不同患者、不同设备和不同扫描时期的 PET 影像结果具有可比性。此外，SUV 值可以用于评估肿瘤或异常组织的活性、生物学特性以及治疗效果。由此，本文将以 DICOM 格式存储的 PET 影像的像素值转换为 SUV 值。

读取 PET 的 DICOM 文件。首先提取标签为 (0010,1030) 且关键字为 Patient Weight 的体重  $w$ ，使用公式4.2计算病人的体重指标  $\mathcal{BW}$ ：

$$\mathcal{BW} = 1000 \times w \quad (4.2)$$

其次，提取标签为 (0008,0021) 且关键字为 Series Date 的扫描日期  $s_d$ 、标签为 (0008,0031) 且关键字为 Series Time 的扫描时间  $s_t$  和标签为 (0018,1072) 且关键字为 Radiopharmaceutical Start Time 的测量时间  $m_t$ ，通过公式4.3来计算扫描到测量之间经过的时间  $\mathcal{T}$ ：

$$\mathcal{T} = seconds(s_d s_t - s_d m_t) \quad (4.3)$$

其中， $seconds(\cdot, \cdot)$  表示计算两个时间戳之间的秒数。然后，提取标签为 (0018,1074) 且关键字为 Radionuclide Total Dose 的注射剂量  $d$  和标签为 (0018,1075) 且关键字为

Radionuclide Half Life 的半衰期  $f$ , 使用公式4.4来计算药物活性  $\mathcal{AC}$ :

$$\mathcal{AC} = d \times 2^{-\frac{T}{f}} \quad (4.4)$$

最后, 提取标签为(7FE0,0010)且关键字为 Pixel Data 的像素值  $P_{pt}$ 、标签为(0028,1052)且关键字为 Rescale Intercept 的截距  $b$  和标签为(0028,1053)且关键字为 Rescale Slope 的斜率  $m$ , 通过公式4.5转换为 SUV 值:

$$SUV = (m \times P_{pt} + b) \times \mathcal{BW} \div \mathcal{AC} \quad (4.5)$$

PET 和 CT 是核医学影像中常见的两种三维模态影像。它们由一组像素或体素组成, 每个像素或体素之间存在间距并且占据了空间中一部分物理区域, 这与其他图像之间存在着很大的差异。如图4.4所示, PET 和 CT 影像所占据的物理空间区域由影像的原点、间距、尺寸和方向余弦矩阵所确定。其中, 原点是在每一维上的索引均为 0 的体素在世界坐标系上的位置; 间距是每个维度上体素之间的间距; 尺寸是每一个维度上体素的数量; 方向余弦矩阵是每个轴的方向。

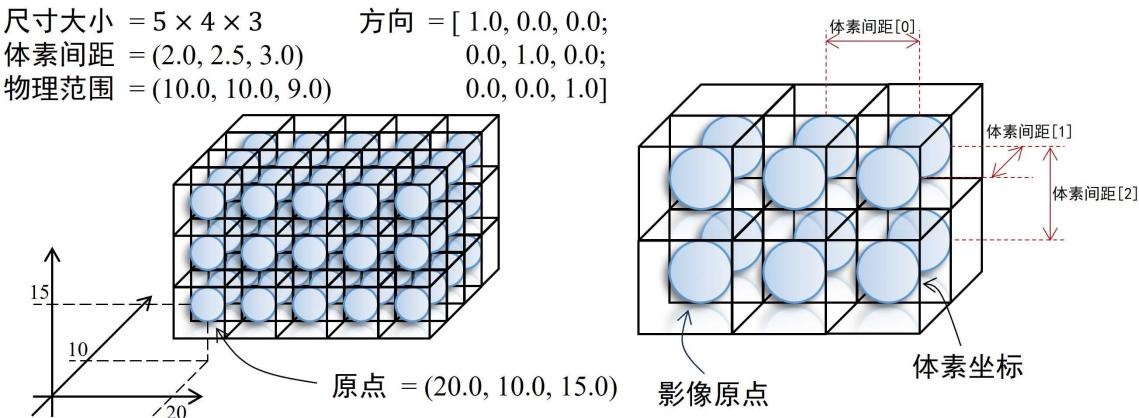


图 4.4 PET 和 CT 在物理空间中所占据的区域

Figure 4.4 The area occupied by PET and CT in the physical space

在本章收集的数据集中, PET 和 CT 的原点、体素间距与尺寸大小都存在差异。因此, 为了充分利用与融合 PET 和 CT 两种模态影像数据, 采用了 SimpleITK 工具库<sup>[101]</sup>进行配准。在经过前面处理计算得到的 HU 和 SUV 替换掉原来的像素值的基础上, 使用 SimpleITK 中的函数 Resample 按照 CT 的原点、体素间距与尺寸大小对 PET 进行重采样。

### 4.2.3 人体区域保留

在对 PET 和 CT 进行转换和配准之后，由于 PET 和 CT 图像中含有过多与任务无关的区域以及 CT 中存在的扫描设备机床容易影响模型训练和检测性能。本章提出了人体区域保留算法，保留前后效果如图4.5所示。人体区域保留算法流程如下：



图 4.5 人体区域保留算法效果图

Figure 4.5 Result of the body retention method

(1) 读取 CT 影像中的 HU 三维数组  $Arr_{HU}$  和 PET 影像中的 SUV 三维数组  $Arr_{SUV}$ 。通过在 CT 和 PET 上分别进行图像二值化计算出初步的分割掩码图  $maskB_{HU}$  和  $maskB_{SUV}$ ，如公式4.6所示：

$$maskB_{HU}^{i,j,k} = \begin{cases} 1, & \text{if } Arr_{HU}^{i,j,k} > -200 \\ 0, & \text{otherwise} \end{cases} \quad (4.6)$$

$$maskB_{SUV}^{i,j,k} = \begin{cases} 1, & \text{if } Arr_{SUV}^{i,j,k} > 0.001 \\ 0, & \text{otherwise} \end{cases}$$

其中， $i, j, k$  表示三个维度上的索引。

(2) 使用一个核半径为 2 且形态为球形的图像滤波  $F_2$  对二值化得到的分割掩码图进行形态学闭操作，用以去除小孔，平滑边界，使得分割掩码图更加完整，如公式4.7所示：

$$\begin{aligned} maskB_{HU} &= (maskB_{HU} \oplus F_2) \ominus F_2 \\ maskB_{SUV} &= (maskB_{SUV} \oplus F_2) \ominus F_2 \end{aligned} \quad (4.7)$$

其中， $\oplus$  表示形态学上的膨胀操作， $\ominus$  表示形态学上的腐蚀操作。

(3) 计算分割掩码图  $maskB_{HU}$  和  $maskB_{SUV}$  的最大连通量。首先，扫描分割掩码图，对其中的每个连通区域标记一个唯一的标签  $L$ 。然后，统计每一个标记区域的体素数量  $S$ 。最后，选择具有最大体素数量的标记区域，即最大连通量，如公式4.8所示：

$$L_{max} = \arg \max_{i=1}^N S_i \quad (4.8)$$

其中， $N$  表示标签的总数， $S_i$  表示标签为  $i$  的区域的体素数量。再通过公式4.9计算仅保留最大连通量的分割掩码图并计算出交集：

$$\begin{aligned} mask_{HU}^{i,j,k} &= \begin{cases} 1, & \text{if } maskB_{HU,max}^{i,j,k} = L_{HU,max} \\ 0, & \text{otherwise} \end{cases} \\ mask_{SUV}^{i,j,k} &= \begin{cases} 1, & \text{if } maskB_{SUV,max}^{i,j,k} = L_{SUV,max} \\ 0, & \text{otherwise} \end{cases} \\ mask &= mask_{HU} \cap mask_{SUV} \end{aligned} \quad (4.9)$$

其中， $maskB_{HU,max}$  和  $maskB_{SUV,max}$  为计算最大连通量时对每个连通区域标记了唯一标签的三维数组。 $L_{HU,max}$  和  $L_{SUV,max}$  分别是  $maskB_{HU,max}$  和  $maskB_{SUV,max}$  的最大连通量的标签。

(4) 使用一个核半径为 20 且形态为球形的图像滤波  $F_{20}$  对最后的分割掩码图  $mask$  进行形态学闭操作，去除伪影，使得分割掩码图更加完整，得到最终的人体区域掩码图  $mask_{body}$ ，如公式4.10所示：

$$mask_{body} = (mask \oplus F_{20}) \ominus F_{20} \quad (4.10)$$

(5) 根据  $mask_{body}$  计算出可以包围人体区域的最小立方体，对 CT、PET 和  $mask_{body}$  进行裁剪。最后，根据  $mask_{body}$  将属于非人体区域中的 CT 的 HU 值设为-1000，PET 的 SUV 值设为 0，如公式4.11所示：

$$\begin{aligned} Arr_{HU}^{i,j,k} &= -1000 && \text{where } mask_{body}^{i,j,k} = 0 \\ Arr_{SUV}^{i,j,k} &= 0 && \text{where } mask_{body}^{i,j,k} = 0 \end{aligned} \quad (4.11)$$

#### 4.2.4 数据增强

由于骨折相关感染的 PET/CT 数据集规模较小、检测目标数量少，导致在一定程度上增加了模型的训练难度。对于这种小规模和少样本的情况可能影响到模型对于骨折相关感染的全面理解和准确检测的问题，本章研究需要采用有效数据增强方法以提高模型的泛化性能，并使其更好地适应不同类型的病灶区域。因此，提出了三种数据增强的方法，如图4.6所示，分别是镜像翻转、适应性修改的 Mixup 和 RICAP。

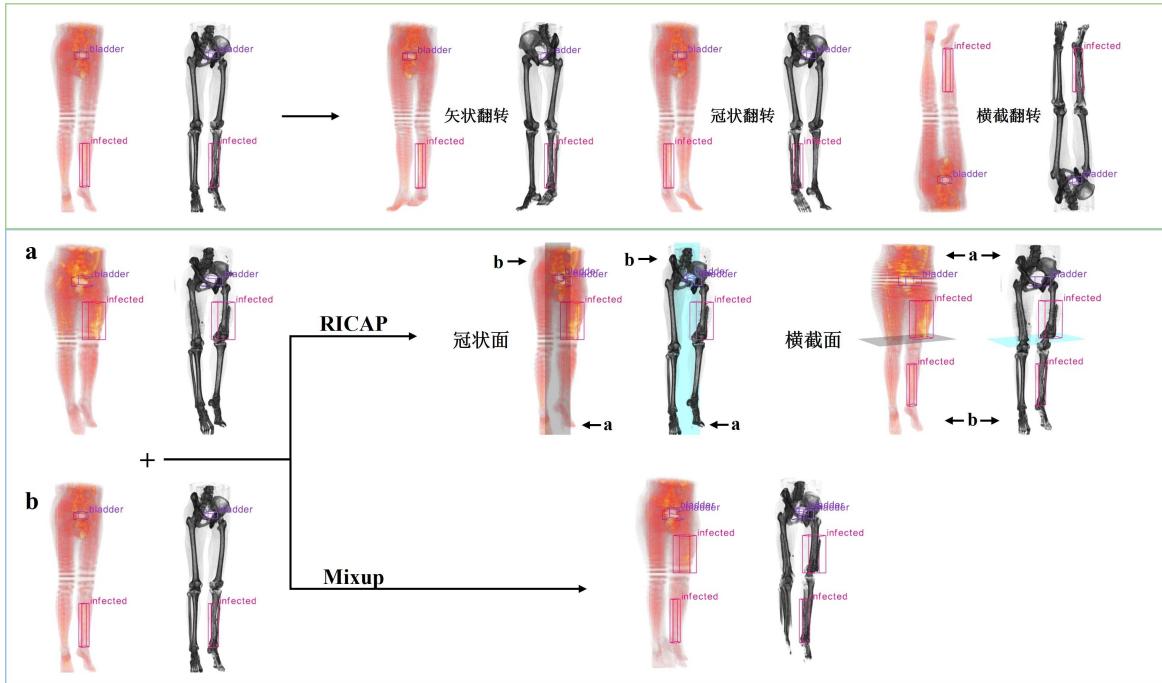


图 4.6 PET/CT 的数据增强方法

Figure 4.6 Data Augmentations for PET/CT

翻转属于单样本的数据增强方法，对单个样本随机地在矢状轴、管状轴和横断轴的其中一个方向上翻转。而 Mixup 和 RICAP 则是属于多样本的数据增强方法，通过各自的方式使用两个样本去合成一个新的样本。

为了将原用于分类任务中的 Mixup 应用于目标检测任务，Mixup 数据增强方法进行了适应性地修改。对于两个输入样本  $image_1$  和  $image_2$  以及它们的标签  $box_1$  和  $box_2$ ，输入样本进行等比例的混合生成新的输入样本  $image_{new}$ ，而标签采用并集的方式合成新的标签  $box_{new}$ ，如公式4.12所示：

$$\begin{aligned} image_{new} &= 0.5 \times image_1 + 0.5 \times image_2 \\ box_{new} &= box_1 \cup box_2 \end{aligned} \quad (4.12)$$

在本章研究中，同样将适用于分类任务中的 RICAP 进行了适应性地修改，以更加适合于骨折相关感染病灶检测任务。首先，随机地在矢状轴和横断轴方向上二选一。随后，在选中的方向上随机地选择需要裁剪的区域并对输入样本和标签进行裁剪。最后，根据裁剪区域的相对位置，拼接裁剪后的两个输入样本及其标签来生成新的输入样本和新的标签。裁剪区域的计算方式如下：

假设输入样本的大小为  $(W, H, D)$ 。如果在矢状轴上进行裁剪与拼接，则对于两个输入样本的裁剪区域分别为  $(1, 1, 1) \sim (\omega, H, D)$  和  $(\omega + 1, 1, 1) \sim (W, H, D)$ 。如果在横断轴上进行裁剪，则对于两个输入样本的裁剪区域分别为  $(1, 1, 1) \sim (W, H, \delta)$  和  $(1, 1, \delta + 1) \sim (W, H, D)$ 。其中， $\omega$  和  $\delta$  的计算方式如4.13所示：

$$\begin{aligned} \omega = round(\omega) \quad \omega &\sim U(0.4 \times W, 0.6 \times W) \\ \delta = round(\delta) \quad \delta &\sim U(0.4 \times D, 0.6 \times D) \end{aligned} \quad (4.13)$$

其中， $round(\cdot)$  表示取整函数， $X \sim U(a, b)$  表示  $X$  服从区间  $(a, b)$  上的均匀分布。

### 4.3 框架设计

PET/CT 影像是两种模态组合的三维影像数据，它们展示了一个患者在注射药剂后的解剖组织信息和生理代谢信息。针对 PET 与 CT 影像之间存在互补互助的特性。本章在骨折相关感染的诊断方法中设计了一个两阶段的病灶检测诊断分类网络框架 3DFRINet，整体架构如图4.7所示。整个框架分为病灶检测网络与病灶诊断分类网络，采用先检测后分类的两阶段设计，可以排除掉与病灶无关的大部分区域，使得病灶诊断分类网络更加关注于病灶区域的生理代谢信息。病灶检测网络 322-YOLO (3 dimensions 2 paths 2 heads YOLO) 通过双分支设计和注意力模块，进一步优化对骨折相关感染病灶的定位。按照 YOLO 的结构设计也提高网络对于小尺度和多尺度

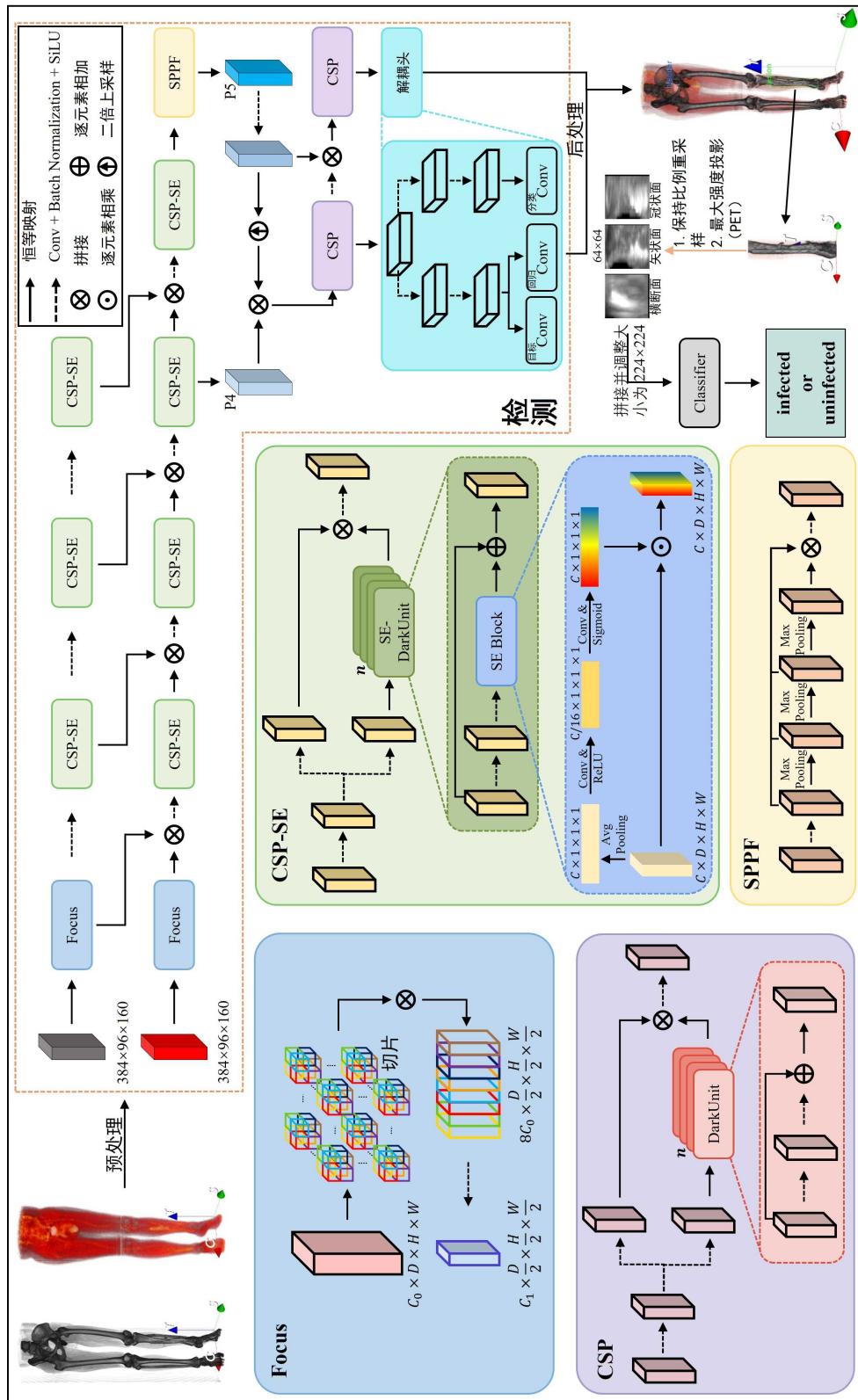


图 4.7 两阶段的检测与诊断框架

Figure 4.7 Two-stage detection and diagnosis framework

病灶的检测精度，同时有效地减少了错检率。病灶诊断分类网络对潜在病灶进一步挖掘病灶的代谢形态特征，实现对骨折相关感染的准确分类。

#### 4.3.1 病灶检测网络 322-YOLO

经过本章引入的预处理流程后，将 PET 和 CT 影像数据重采样至  $384 \times 96 \times 160$ ，并将 CT 的 HU 以 300 为窗位，1500 为窗宽进行归一化，PET 的 SUV 小于 0 的统一为 0。随后，被输入到 322-YOLO 病灶检测网络，对骨折相关感染病灶区域进行检测和识别。322-YOLO 在 YOLOX<sup>[78]</sup>的基础上改进而来，由用于特征提取的双分支主干网络，整合两个不同层次下的特征图的颈部网络以及用于预测目标的种类、大小和位置的头部网络组成。

双分支主干网络可以分为 PET 主分支和 CT 辅助分支，当 PET 主分支持续提取生理代谢信息特征的同时，来自于 CT 辅助分支的解剖信息特征也被逐阶段地融入生理代谢信息特征之中。PET 主分支主要由 Focus、CSP-SE 和 SPPF 模块构成，而 CT 辅助分支主要由 Focus 和 CSP-SE 模块构成。Focus 模块从三个维度上以一个像素间隔对输入数据进行切片，随后拼接切片得到的 8 个子块，然后通过卷积进行特征提取与下采样。它将 PET 中的生理信息和 CT 中的解剖信息从空间维度堆叠到通道维度，在不牺牲特征信息的前提下，提高卷积操作对生理和解剖信息的感知范围，为网络提供了更广泛的上下文信息，有助于准确定位和识别潜在病灶。CSP-SE 模块 (the Cross Stage Partial-Squeeze and Excitation dark unit) 在 dark unit 的基础上结合了跨阶段部分结构<sup>[73]</sup>和通道注意力模块<sup>[102]</sup>。跨阶段部分结构允许特征在不同层之间进行更有效的信息传递和融合，有助于捕捉不同尺度和抽象级别的特征。同时，通道注意力模块的引入使得 CSP-SE 模块能够根据特定通道的重要性调整特征的权重，进一步提升网络对骨折相关感染病灶的关注度。通过充分结合了多层次的特征表达和通道级的信息调控，为网络在处理骨折相关感染特征上提供了更为丰富和精准的上下文信息。SPPF 模块是由 SPP<sup>[76]</sup>改进而来，具有更少的参数，更快的计算速度。它可以整合不同尺度的特征，以更好地适应骨折相关感染病灶的多样性（例如病灶区域大小、生理代谢或解剖结构上的差异），减轻对单一尺度特征的过度依赖，全面地捕捉病灶区域的上下文信息。

用于训练模型的输入样本为  $SUV, HU \in R^{C \times D \times H \times W}$ ，其中  $C, D, H, W$  分别是输入样本的通道数、深度、高度和宽度。由于正负样本存在极度不平衡的问题，因此病

灶检测网络不再使用最大的同时也是正负样本最不平衡的特征图  $P_3$ 。则由双分支主干网络提取的两个尺度下的特征图  $P_4$  和  $P_5$  的计算方式如公式4.14所示：

$$\begin{aligned} P_1^{HU} &= \text{Focus}(HU) & P_1^{SUV} &= \text{Focus}(SUV) \\ P_i^{HU} &= \text{CSP-SE}(\text{CBS}(P_{i-1}^{HU})) \\ P_j^{SUV} &= \text{CSP-SE}(\text{CBS}(P_{j-1}^{HU} \oplus P_{j-1}^{SUV})) \\ P_4 &= P_4^{SUV} & P_5 &= \text{SPPF}(P_5^{SUV}) \end{aligned} \quad (4.14)$$

其中， $P_i^{HU}(i = 2, 3, 4)$  表示为第  $i$  阶段中 CT 辅助分支提取的特征图， $P_j^{SUV}(j = 2, 3, 4, 5)$  表示为第  $j$  阶段中 PET 主分支融合并提取的特征图，CBS 表示一个由卷积、批归一化和 SiLU 构成的基础模块， $\oplus$  表示拼接操作。

颈部网络的结构由自顶向下和自底向上两条路径组成，以融合不同层次下的特征图，来提高对病灶的准确定位和识别。自顶向下的路径通过将丰富的语义信息（如病灶和高摄取区域的类别信息）传播到浅层特征图，为网络提供更高层次的语义理解。与此同时，自底向上的路径则将病灶和高摄取区域的位置、大小信息传递到深层特征图，使得不同层次的特征在融合后既包含丰富的语义信息，又包含位置和大小信息。这种双向信息传递的机制有助于更全面地捕捉病灶的特征，从而提高检测准确性。与此同时，自底向上的路径聚合了所有特征层的路径，从而缩短了底层与顶层特征之间的距离。这样的结构有助于更快地传播底层特征到上层，提高了网络对于病灶不同层次特征的有效利用，增强了整体网络的表达能力。通过自顶向下的路径计算得到  $P_4^{neck}$  和自底向上的路径计算得到  $P_5^{neck}$ ，如公式4.15所示：

$$\begin{aligned} P_4^{neck} &= \text{CSP}(\text{Upsample}(\text{CBS}(P_5)) \oplus P_4) \\ P_5^{neck} &= \text{CSP}(\text{CBS}(P_5) \oplus \text{CBS}(P_4^{neck})) \end{aligned} \quad (4.15)$$

其中，Upsample 表示 2 倍上采样。

头部网络采用解耦的模式，通过在颈部网络融合的特征中独立地进行目标的类别、位置和大小的预测，有效减少了不同预测任务之间的互相干扰。这种解耦设计使得模型可以更准确地预测定位病灶和高摄取区域的类别，同时在空间上准确定位

病灶和高摄取区域的位置和大小，从而提高了整体检测性能，如公式4.16所示：

$$\begin{aligned} P_c &= \text{CBS}(\text{CBS}(P_n^{neck})) & P_r &= \text{CBS}(\text{CBS}(P_n^{neck})) \\ \text{cls}_n &= \text{Conv}_1(P_c) & \text{reg}_n &= \text{Conv}_2(P_r) \\ && \text{obj}_n &= \text{Conv}_3(P_r) \end{aligned} \quad (4.16)$$

### (1) 无先验框

322-YOLO 采用了无先验框的方式。在解耦头中，独立预测的目标类别设置为 2 类，分别是病灶和膀胱；目标大小从 4 个值改成了 6 个值，即预测框在三维网格中位于左上后方的三个偏移量以及预测框的宽、高和深。在检测网络最终预测的不同层次下的特征图中，根据目标的中心位置以及预定义的范围去分配相应的正样本。

### (2) 分配策略

322-YOLO 作为一个无先验框的目标检测方法，对每一个目标不仅仅只分配一个位于中心位置的正样本，同时还关注其他高质量的预测。特别地，322-YOLO 将目标的中心位置为中心点，以 2.5 为半径的正方体区域设置为正样本，同时也将目标框中的区域设置为正样本。通过这种正样本分配方式去关注些高质量的预测有可能带来有益的梯度，这从一定程度上来说也可以缓解训练过程中正负样本的极端不平衡。

在病灶检测任务中，标签分配策略也是很重要的一部分。322-YOLO 采用了 SimOTA<sup>[78]</sup>，以免去手工选定参数的方式，让真实框自动地匹配多个最合适的预测框。SimOTA 可以视为一个动态 top-k 的过程：首先，计算每个真实框与预测框之间的成对成本  $c$ ，通过使用预测框  $p_n$  与真实框  $g_m$  损失来表示，如公式4.17所示：

$$c_{mn} = L_{mn}^{cls} + \lambda L_{mn}^{reg} \quad (4.17)$$

其中， $\lambda$  表示平衡系数， $L_{mn}^{cls}$  和  $L_{mn}^{reg}$  是真实框  $g_m$  与预测框  $p_n$  之间的分类损失和回归损失。对于每个真实框，在满足前面所述的正样本分配策略的基础上，进一步地选择前 k 个最小成对成本的预测框。其中，成对成本越低表示真实框与预测框的匹配度越高。最终，这些满足条件的预测框将设置为正样本。

### (3) 损失函数

在 SimOTA 标签分配策略对每个目标都分配相应的正负样本之后，322-YOLO 采用了二元交叉熵损失函数（Binary Cross Entropy Loss）用于计算区分正负样本的损失  $\mathcal{L}_{obj}$  和目标类别的损失  $\mathcal{L}_{cls}$ 。而对于目标尺度大小的损失  $\mathcal{L}_{reg}$ ，322-YOLO 则

使用了交并比损失函数 (IoU Loss)。对于预测框  $p = (p^{obj}, p^{cls}, p^{reg})$  和真实框  $g = (g^{obj}, g^{cls}, g^{reg})$ , 三种损失的计算方式如公式4.18所示:

$$\begin{aligned}\mathcal{L}_{obj} &= - \sum_{i=0}^I (g_{bi}^{obj} \log p_{bi}^{obj} + (1 - g_{bi}^{obj}) \log(1 - p_{bi}^{obj})) \\ \mathcal{L}_{cls} &= - \sum_{k=0}^K (g_{bk}^{cls} \log p_{bk}^{cls} + (1 - g_{bk}^{cls}) \log(1 - p_{bk}^{cls})) \\ \mathcal{L}_{reg} &= \sum_{k=0}^K (1 - (\frac{g_{bk}^{reg} \cap p_{bk}^{reg}}{g_{bk}^{reg} \cup p_{bk}^{reg}})^2)\end{aligned}\quad (4.18)$$

其中,  $I$  表示所有预测框的数量,  $K$  表示 SimOTA 标签分配策略对每个真实框匹配到的相应的正样本的数量。最终的损失函数如公式4.19所示:

$$\mathcal{L} = \mathcal{L}_{obj} + \mathcal{L}_{cls} + 5 \times \mathcal{L}_{reg} \quad (4.19)$$

#### 4.3.2 病灶诊断分类网络

首先, 设置置信度阈值为 0.1, 最大输出的检测框数量为 100, 让目标检测网络 322-YOLO 输出所有符合条件的预测框。然后, 设置交并比阈值为 0.01, 去除掉相同类别下重叠程度高于交并比阈值的预测框。最后, 筛选得到类别为病灶的预测框。

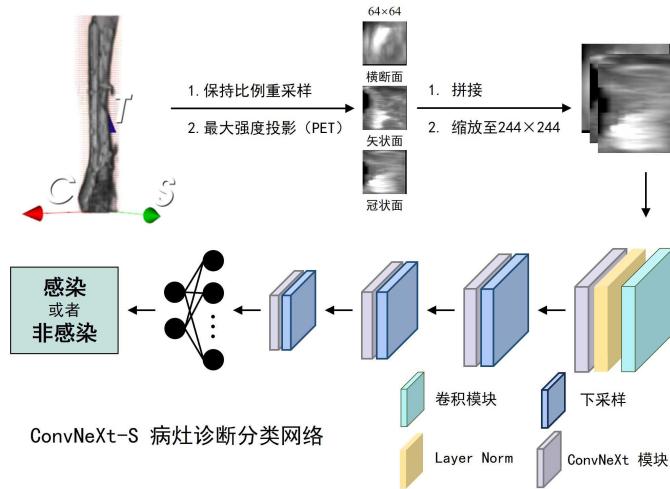


图 4.8 痘痘诊断分类网络

Figure 4.8 Lesion diagnosis classification network

根据得到的病灶预测框, 在预处理后的 PET 影像上进行裁剪, 得到病灶区域的 PET 影像。然后对病灶区域的 PET 影像进行等比例的缩放并在四周进行填充, 使得

最终获得的 PET 影像为  $64 \times 64 \times 64$ ，并将 PET 的 SUV 小于 0 的设置为 0。其中，PET 中填充区域的值为 0。在矢状轴、冠状轴和横断轴上进行最大强度投影得到的三张单通道图像拼接成一张三通道图像，并重采样至  $244 \times 244$  的大小后，输入到病灶诊断分类网络 ConvNeXt-S 之中，对病灶区域进行诊断分类，如图4.8所示。

将训练模型的图像输入到 ConvNeXt-S 诊断分类网络中，经过 ConvNeXt-S 的卷积层、Layer norm 层、ConvNeXt 模块、下采样层、平均池化层和全连接层得到最终的诊断分类结果。其中，在训练过程中，使用了镜像（水平或翻转）和 Mixup 来弥补数据集不足的问题，损失函数采用的是同第三章一样的 Focal Loss 改进版，如公式3.8所示。

## 4.4 实验

### 4.4.1 实验环境

本章所有实验均使用 PAI 上海大学机器学习平台上进行，具体详细的配置信息与第三章一致。实验配置主要包括两个 NVIDIA Tesla V100 (32GB) GPU、Linux Ubuntu 20.04.2 操作系统、Pytorch 1.8.0 深度学习框架、Python 3.8.17 编程语言、SimpleITK 2.0.2 医学影像处理库、opencv-python 4.5.4.60 图像处理库以及 numpy 1.24.3 数据处理库。

### 4.4.2 数据集

本章采用的下肢骨折相关感染的 PET/CT 影像数据集由本文作者与上海市第六人民医院核医学科合作构建。纳入数据集的标准：存在既往的下肢骨折手术治疗；存在术后症状，如疼痛、活动受限或红肿；进行过  $^{18}\text{F}$ -FDG PET/CT 扫描； $^{18}\text{F}$ -FDG PET/CT 扫描时间与既往手术之间的时间间隔超过 3 个月； $^{18}\text{F}$ -FDG PET/CT 扫描时间与手术后的时间间隔少于 1 个月。排除标准：存在鼻窦道、脓等典型感染的临床表现；术后未行常规细菌培养以及病理分析； $^{18}\text{F}$ -FDG PET/CT 扫描前一周内使用抗生素；不完整的临床数据；现有资料不能确定最终诊断。在数据集方面，一共纳入了从 2016 年 11 月到 2021 年 12 月之间 303 名潜在合格的患者。在阅览所有患者的 PET/CT 影像后，22 名患者因为影像缺失或不完整被排除掉。最终，具有 298 个病灶区域的 281 名患者被纳入本章的研究中，详细信息如表4.1所示。本章研究中所有患

者骨折相关感染的最终诊断是基于手术中在深部组织中提取并微生物培养或组织病理学分析以及至少 6 个月的临床随访。

表 4.1 下肢骨折相关感染的 PET/CT 影像数据集特征

Table 4.1 Characteristics of PET/CT datasets for lower limb fracture-related infections

特征	数量
骨折类别, 数量 (%)	
股骨	91 (30.5)
胫腓骨	169 (56.7)
足	36 (12.1)
下肢	2 (0.7)
诊断结果, 数量 (%)	
感染	190 (63.8)
非感染	108 (36.2)

#### 4.4.3 评价指标

骨折相关感染诊断任务包含分类和目标检测两个子任务。分类任务采用的评估指标与第三章一致。Acc 评估下肢骨折相关感染的诊断准确率；Spec 评估实际不患骨折相关感染的患者诊断未感染的概率；Sen 评估实际患有骨折相关感染的患者诊断为感染的概率。PPV 评估被诊断为骨折相关感染的患者实际患有的概率；NPV 评估被诊断未患骨折相关感染的患者实际未感染的概率。F1 是 PPV 和 Sen 的综合评价指标，AUC 是敏感性和特异性在不同阈值设定下生成的 ROC 曲线下的面积。

在目标检测任务中, 本章选用了常用的评估指标, 如平均精度 (Average Precision, AP)、精确率 (Precision)、召回率 (Recall) 和 F1 值 (F1 score), 它们的计算方式如公式4.20所示:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 F1score &= \frac{2 \times Precision \times Recall}{Precision + Recall} \\
 AP &= \sum_{r_i} (Recall_{r_i} - Recall_{r_{i-1}}) \times Precision_{r_i}
 \end{aligned} \tag{4.20}$$

其中,  $TP$  是被正确检测为正类别的样本数量,  $FP$  是被错误检测为正类别的样本数

量,  $FN$  是被错误地标记为负类别的样本数量。 $Recall_{r_i}$ ,  $Precision_{r_i}$  表示根据模型输出和真实标签, 在置信度为  $r_i$  下的精确率和召回率。平均精度衡量了在不同置信度阈值下目标检测网络的精确率和召回率之间的平衡, 是一个综合性指标; 精确率表示预测为正类别的样本中实际为正类别的比例; 召回率表示实际为正类别的样本中被正确预测为正类别的比例。

#### 4.4.4 实验结果

本章提出的病灶检测网络 322-YOLO 跟其他 YOLO 系列模型的病灶检测性能进行了比较, 如表4.2所示。其中, 为了 YOLO 系列模型适用于三维的 PET/CT 影像数据, 将 YOLO 系列模型的预处理、模型结构、后处理部分中都从相应的二维操作修改为了三维操作。在综合检测性能上, 322-YOLO 不仅拥有较小的计算量和参数量, 还表现出了最佳的病灶定位检测性能。此外, 322-YOLO 的计算量和参数量还不到在  $AP_{25:75}$  上排名第二的 YOLOX-M 的一半。在其他关于病灶检测性能的细分评估指标  $AP_{25}$ 、 $Recall_{25}$ 、 $F1_{25}$  和  $Recall_{50}$  中, 322-YOLO 都是最佳的。这表明 322-YOLO 在病灶检测、计算量和参数量上的综合表现能力最佳。值得说明的是, 322-YOLO 在验证集中检测到了 75 个病灶区域, 其中有 63 个与专业医生标记的病灶区域相匹配, 有 9 个是未标记的骨折相关感染病灶区域, 只有 3 个是完全误检。

表 4.2 病灶检测网络与其他 YOLO 的检测性能比较

Table 4.2 Comparison of detection performance between lesion detection network and other YOLO

模型	$AP_{25:75}$	$AP_{25}$	$Precision_{25}$	$Recall_{25}$	$F1_{25}$	$AP_{50}$	$Precision_{50}$	$Recall_{50}$	$F1_{50}$	flops	params
YOLOv3	0.1941	0.5656	0.6842	0.6290	0.6566	0.1152	0.1579	0.1452	0.1515	368.79G	172.95M
YOLOv4	0.2553	0.7377	0.8000	0.7742	0.7871	0.1405	0.2667	0.2581	0.2624	278.54G	170.69M
YOLOv5-S	0.2019	0.6876	0.7377	0.7258	0.7318	0.0540	0.1803	0.1774	0.1789	35.65G	15.84M
YOLOv5-M	0.2475	0.7265	0.7460	0.7581	0.7520	0.1644	0.3016	0.3065	0.3040	99.17G	50.12M
YOLOv5-L	0.2113	0.7109	0.6957	0.7742	0.7349	0.0963	0.1884	0.2097	0.1990	216.93G	114.87M
YOLOX-S	0.3286	0.8085	0.8154	0.8548	0.8351	0.3032	<b>0.4154</b>	0.4355	<b>0.4254</b>	59.15G	21.24M
YOLOX-M	0.3673	0.8049	<b>0.8281</b>	0.8548	0.8415	<b>0.3073</b>	0.4063	0.4194	0.4128	152.03G	62.29M
YOLOX-L	0.3069	0.7892	0.8000	0.8387	0.8194	0.1738	0.3077	0.3226	0.3151	310.91G	136.53M
322-YOLO	<b>0.3834</b>	<b>0.9234</b>	0.8000	<b>0.9677</b>	<b>0.8839</b>	0.3041	0.3600	<b>0.4355</b>	0.3977	63.71G	26.04M

\*<sub>25:75</sub> 表示在  $\text{IoU} \in [0.25, 0.75]$  以 0.05 为步长时 \* 的平均值; \*<sub>25</sub> 表示在  $\text{IoU}=0.25$  时 \*; \*<sub>50</sub> 表示在  $\text{IoU}=0.50$  时 \*;

为了验证病灶检测网络中多个改进方法的有效性, 如模型结构、数据增强, 本章设计了消融实验。实验步骤如下:

- (1) 采用 YOLOX-S 作为基准模型在本章数据集中进行训练与评估;

- (2) 在(1)的基础上，将主干网络最后一个阶段特征提取模块的组合 SPP+CSP 修改为 CSP+SPPF，即 SPP→SPPF\*；
- (3) 在(2)的基础上，仅采用主干网络中提取的 P4 和 P5 特征图，颈部网络和头部网络结构也进行相应的修改，即 only use P4/P5；
- (4) 在(3)的基础上，在主干网络部分添加通道注意力模块，即 SE block；
- (5) 在(4)的基础上，修改单分支主干网络为双分支的结构，即 two path backbone；
- (6) 在(5)的基础上，在训练过程中随机使用适应性修改的 Mixup，即 Mixup；
- (7) 在(6)的基础上，在训练过程中再随机使用适应性修改的 RICAP，即 RICAP。

在表4.3中可以看出，在综合检测性能指标 AP<sub>25:75</sub> 中，除了 only use P4/P5 以外，所有在模型结构上的改进方法对病灶检测网络的检测性能都有所提升。对于 only use P4/P5 而言，计算量和参数量的减少，使得病灶检测性能的下降是可以接受的。数据增强方法 Mixup 在 AP<sub>25:75</sub> 上进一步提升了模型检测病灶的能力，而 RICAP 在 AP<sub>25</sub> 和 AP<sub>50</sub> 上进一步达到了最优的结果。通过病灶检测网络的消融实验可以得出，病灶检测性能 AP<sub>25:75</sub> 提升了 16.7%，AP<sub>25</sub> 达到了 0.9234，AP<sub>50</sub> 达到了 0.3041，这表明了本章研究中针对病灶检测任务提出的所有改进的有效性。

表 4.3 病灶检测网络 322-YOLO 的消融实验

Table 4.3 Ablation experiments on the lesion detection network 322-YOLO

	病灶				膀胱		flops	params
	AP <sub>25:75</sub>	AP <sub>25</sub>	AP <sub>50</sub>	AP <sub>25:75</sub>	AP <sub>25</sub>	AP <sub>50</sub>		
YOLOX-S	0.3286	0.8085	0.3032	0.7676	1.0000	0.8418	59.15G	21.24M
+ SPP→SPPF*	0.3390	0.7829	0.2812	0.7543	1.0000	0.8870	59.15G	21.24M
+ only use P4/P5	0.3248	0.8506	0.1650	0.7714	1.0000	0.8487	35.08G	18.22M
+ SE block	0.3610	0.8943	0.1965	0.7523	1.0000	0.8355	35.09G	18.24M
+ two path backbone	0.3895	0.9105	0.2746	0.7442	0.9643	0.8857	63.71G	26.04M
+ Mixup	<b>0.4084</b>	0.9010	0.2795	0.7341	1.0000	0.8615	63.71G	26.04M
+ RICAP	0.3834	<b>0.9234</b>	<b>0.3041</b>	0.7456	<b>1.0000</b>	<b>0.9279</b>	63.71G	26.04M

为了验证单阶段算法与双阶段算法的效果差异，进行了对比实验，如表4.4所示。如果仅使用 322-YOLO 自身的类别预测来进一步细分病灶类别，病灶的检测效果较差。而采用两阶段的病灶检测诊断分类网络框架 3DFRINet (322-YOLO + Classifier)，即通过添加二阶段的病灶诊断分类网络对已定位的病灶进行进一步地细分，可以大大的提高病灶检测性能。与单阶段的 322-YOLO 的病灶检测性能相比，3DFRINet 在

表 4.4 有无额外分类器的病灶检测网络的比较

Table 4.4 Comparison of lesion detection networks with and without additional classifiers

	感染			非感染		
	AP <sub>25</sub>	AP <sub>50</sub>	AP <sub>25:75</sub>	AP <sub>25</sub>	AP <sub>50</sub>	AP <sub>25:75</sub>
322-YOLO	0.6395	0.3108	0.3129	0.2193	0.0022	0.0433
322-YOLO + Classifier	0.8016	0.3346	0.3790	0.6330	0.1253	0.1808

感染病灶中的 AP<sub>25</sub>、AP<sub>50</sub> 和 AP<sub>25:75</sub> 分别提升了 25.35%、7.66% 和 21.12%，在非感染病灶中的 AP<sub>25</sub>、AP<sub>50</sub> 和 AP<sub>25:75</sub> 分别提升了 188.65%、5595.45% 和 317.55%。这种现象产生的原因在于，感染病灶与非感染病灶之间存在一定的相似性。加之，定位网络主要分析整体影像的特征，而不是特定病灶的特征，从而引起识别的不准确。同时，类别不平衡也使得定位网络在病灶的预测上倾向于感染类别。由此可知，这个实验的结果证明了两阶段的检测分类框架比单阶段算法的性能更好。

表 4.5 3DFRINet 与核医学医生的比较

Table 4.5 Comparison of 3DFRINet with nuclear medicine physicians

	Acc(%)	Spec(%)	Sen(%)	NPV(%)	PPV(%)	F1	AUC	p
3DFRINet	91.55 (83.10,97.18)	90.32 (77.78,100.00)	92.50 (83.72,100.00)	90.32 (78.57,100.00)	92.50 (82.93,100.00)	0.9250 (0.8533,0.9756)	0.9331 (0.8547,0.9870)	-
初级医生 1	74.65	93.55	60.00	64.44	92.31	0.7273	-	<b>0.0018</b>
初级医生 2	80.28	80.65	80.00	75.76	84.21	0.8205	-	0.0768
初级医生 3	74.65	77.42	72.50	68.57	80.56	0.7632	-	<b>0.0075</b>
高级医生 1	87.32	90.32	85.00	82.35	91.89	0.8831	-	0.4531
高级医生 2	85.92	90.32	82.50	80.00	91.67	0.8684	-	0.2188
高级医生 3	87.32	93.55	82.50	80.56	94.29	0.8800	-	0.3750

为了进行更加全面的对比评估实验，本章将提出的两阶段的病灶检测诊断分类网络框架 3DFRINet 与 6 名核医学影像科的临床医生进行比较，如表 4.5 所示。其中，有 3 名初级核医学医生和 3 名高级核医学医生。从对比结果可以得出，3DFRINet 拥有 91.55% 的准确率、92.50% 的敏感率、90.32% 的阴性预测值和 0.9250 的 F1 上，均优于 3 名初级核医学医生和 3 名高级核医学医生。此外，其 90.32% 特异率和 92.50% 的阳性预测值都位列前三。这表明了 3DFRINet 相较于 3 名初级核医学医生和 3 名高级核医学医生具有优秀的骨折相关感染诊断分类的性能表现。通过统计学分析，3DFRINet 与两名初级核医学医生在准确率上的 p 值小于 0.05，这说明 3DFRINet 与两名初级核医学医生在准确率上存在显著差异，而与 1 名初级核医学医生和 3 名高级核医学医生在准确率无显著差异。由此可知，3DFRINet 具有与初级核医学医生更

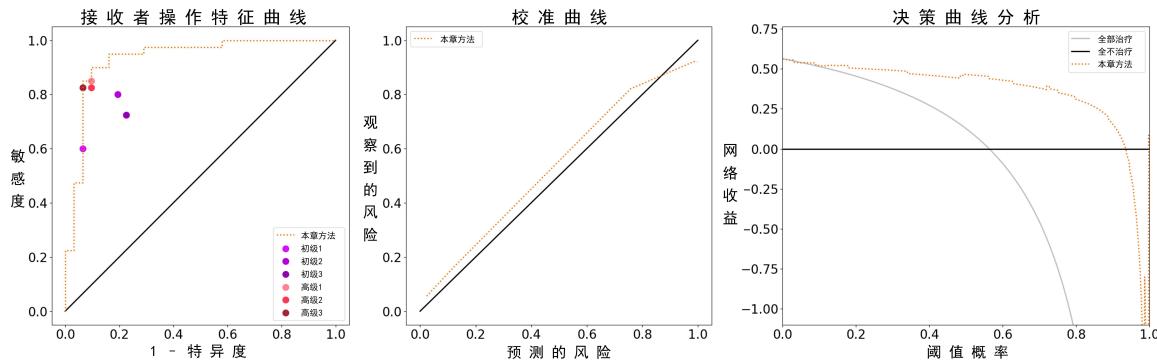


图 4.9 ROC 曲线、校准曲线和决策曲线分析

Figure 4.9 ROC curve, calibration curve and decision curve analysis

好或相同的诊断能力，与高级核医学医生相当的诊断准确率。

如图4.9所示，从 ROC 曲线上可以得出 3DFRINet 的诊断分类性能与三名高级核医学医生相当。3DFRINet 的校准曲线较接近与 45 度对角线，这反映了其在三维 PET/CT 影像上骨折相关感染的预测概率与骨折相关感染的真实概率具有良好的一致性。决策曲线分析表明，3DFRINet 在阈值概率较高的情况下存在负面收益，但大多数情况下是存在收益的且有价值的。

#### 4.4.5 可视化分析

图4.10展示了 322-YOLO 检测结果的可视化图，包括 PET 的三维点云、CT 的三维重建和 PET 的最大强度投影图。其中，青色和蓝色的边界框表示真实标签，而紫色和绿色的边界框表示检测结果。322-YOLO 展现了出色的检测能力，不仅能够准确定位已标注的病灶区域，而且能够有效检测出未标注的病灶区域（第 1 行第 1 列，第 3 行第 1 列）。值得注意的是，在病灶区域过大的情况下，322-YOLO 还能通过使用两个边界框（第 1 行第 2 列）精准地定位病灶中更值得关注的子区域。此外，无论 PET 图像中是否存在高摄取区域，或者 CT 图像中是否存在骨侵蚀、骨折或内固定物等情况，322-YOLO 模型都能够进行综合考虑并做出准确的检测与定位（第 2 行和第 3 行）。这表明了 322-YOLO 对于复杂场景和不同病变情况的强大适应性和鲁棒性。

### 4.5 本章小结

本章主要提出了一个基于  $^{18}\text{F}$ -FDG PET/CT 影像的两阶段的下肢骨折相关感染病灶检测诊断分类网络框架 3DFRINet，由预处理、322-YOLO 病灶检测网络和

ConvNeXt-S 病灶诊断分类网络组成。预处理通过去除无关区域并保留人体区域的 PET SUV 和 CT HU 输入到 3DFRINet 中。在 322-YOLO 病灶检测网络中，通过双分支主干网络结构设计与通道注意力模块有效地提取并融合 PET 的生理代谢信息和 CT 的解剖结构信息，从而充分利用不同模态的特点准确地检测病灶。通过最大强度投影将检测到的病灶从三维转换为二维，在降低数据和模型规模的同时获取有效特征，ConvNeXt-S 病灶诊断分类网络达到了优异的分类性能。在下肢骨折相关感染的诊断任务中，该框架在检测方面达到了 0.9234 的 AP<sub>25</sub> 和 0.3041 的 AP<sub>50</sub>，在诊断分类方面达到了 91.55% (83.10%, 97.18%) 的准确率、0.9250 (0.8533, 0.9756) 的 F1 和 0.9331 (0.8547, 0.9870) 的 AUC，这表明该框架具有优越的病灶检测性能和良好的诊断分类性能。在与核医学医生的比较中，该框架相当或优于初级核医学医生，且与

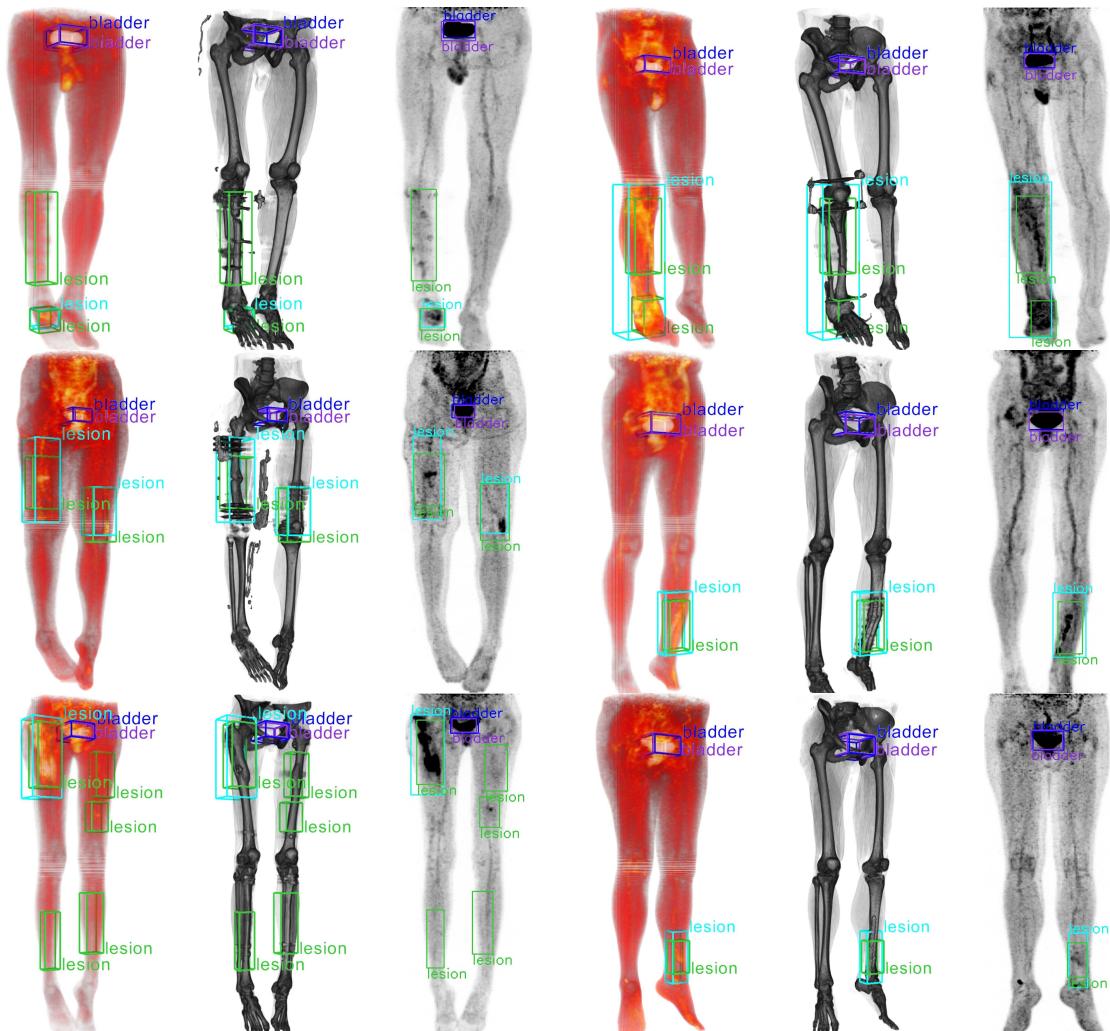


图 4.10 322-YOLO 的检测结果

Figure 4.10 Detection results of 322-YOLO

高级核医学医生相当。这进一步表现了深度学习方法在诊断骨折相关感染的可行性与该检测诊断框架的有效性。

## 第五章 骨科相关感染的自动诊断与可视化方法

第三章和第四章分别详细地介绍了应用于动态骨显像和 PET/CT 中以人工智能为基础的假体关节感染辅助诊断框架和骨折相关感染辅助检测诊断框架。由于这两个框架无法提供便利且直观的辅助诊断以及可视化，本章继续在这一研究方向上深入，结合这两种人工智能方法并采用最新的 PyQt6 框架，去设计一种新颖的骨科相关感染的自动诊断与可视化方法。该方法旨在构建一个骨科相关感染的自动辅助诊断全流程，提供全面的医学影像可视化、阅览和分析功能，以促进临床诊断中假体关节感染与骨折相关感染的诊断决策和提高准确性。

### 5.1 问题分析

在设计假体关节感染和骨折相关感染的自动诊断与可视化方法时，必须深入考虑核医学专家在诊断这类骨科相关感染疾病的具体流程和对医学影像的可视化需求。这一方法需要融合多种先进的图像处理技术，同时也要求透彻地理解并仿真医生的诊断流程。这要求此方法不仅要精确处理和展示医学影像，还具备医学的决策智能，以提供直观、准确的诊断辅助。因此，本章主要包含以下两个方面的研究：

(1) 模拟核医学医生的诊断流程。在假体关节感染和骨折相关感染诊断任务中，核医学专家遵循的实际临床诊断流程具体包括：首先，在动态骨显像或 PET/CT 中，精确定位到潜在的病灶区域。随后，核医学专家依靠专业的工具对所定位的病灶区域进行深入的分析和评估。最后，核医学专家将依托其丰富的临床经验和细致的视觉评估作出最终的诊断决策。

由此，在本章中，所研究的自动诊断方法需要集成第三章和第四章设计的图像处理技术。这些图像处理技术能够从复杂的动态骨显像和 PET/CT 影像中提取到关键的生物标志物，通过端到端地方式直接在医学影像中进行辅助诊断。特别是第三章采用了 CNN 和 ConvLSTM，用于深入地学习动态骨显像中的图像特征和序列变化，这对于髋关节和膝关节假体感染的准确诊断至关重要。同时，第四章利用基于深度学习的检测和分类算法处理 PET/CT 影像，以获得更加全面立体的解剖结构和生理代谢信息，进而有效地锁定病灶区域，并对骨折相关感染进行精准诊断。

(2) 不同医学影像模态的可视化。在核医学专家的临床诊断流程中, 不同医学影像的可视化是最基础的需求之一。可视化需要能够处理、解析和展示不同来源的影像数据, 如动态骨显像和 PET/CT (PET SUV 和 CT HU), 保证医生在做出诊断决策时能够全面考量所有相关影像信息。

由此, 在本章中, 所研究的可视化方法需要使得医学影像可根据专业医生的具体需求进行个性化调整, 包括调整对比度、缩放、以及使用不同的着色方案以适应不同的医学影像模态和增强特定组织或病变的可视性等。同时, 也需要特别注重和优化医生的视觉体验, 提供直观的用户界面来展示各种各样的影像数据。通过提供分层视图、切换视图、融合影像和三维重建功能等, 帮助医生从不同角度分析单一模态和多模态影像, 从而得到更全面的诊断信息。

## 5.2 总体设计

本章设计了一个基于深度学习和 PyQt6 框架的骨科相关感染自动诊断和可视化方法, 旨在提供精确地诊断辅助。如图 5.1 所示的方法流程涵盖了四个关键模块和相应的工作流程。整个方法流程由四个主要步骤构成:

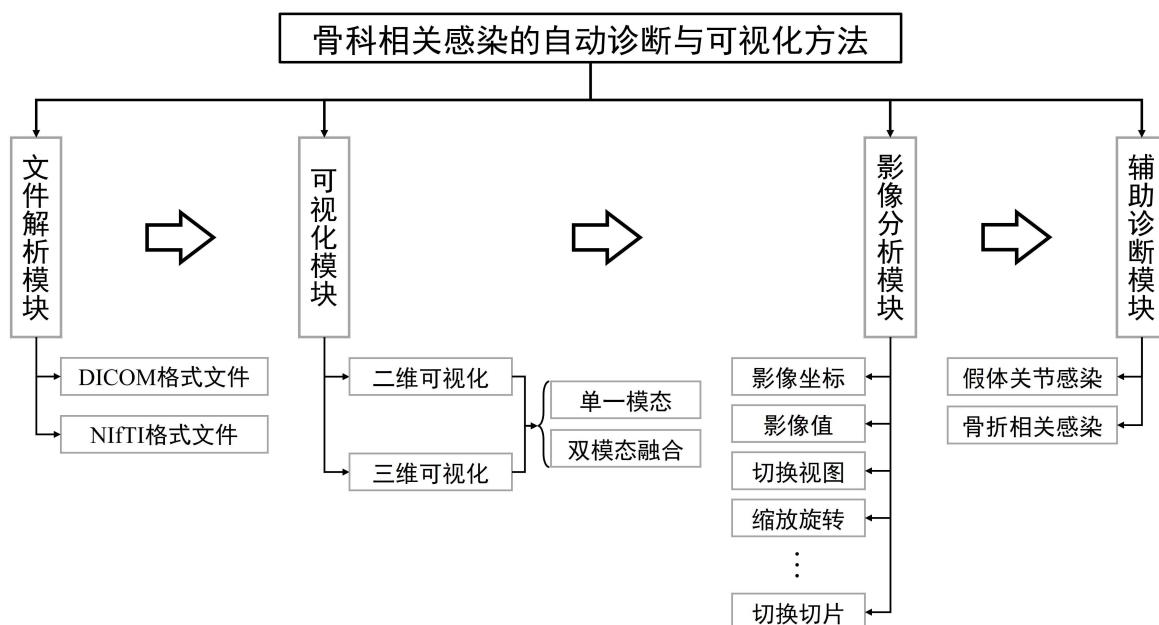


图 5.1 总体设计

Figure 5.1 Overall design

1. 文件解析: 被用户选定的 DICOM 或 NIfTI 文件经过读取和解析, 其中根据各

自的文件规范执行相应的解码流程，旨在准确抽取所需的医学信息数据。

2. 可视化：经过解析得到的数据在预处理后转化为视觉表现形式，以直观展示医学影像。该预处理包括标准化、配准和融合，以提供丰富且细致的视觉信息。
3. 影像分析：在实现影像数据的视觉呈现后，影像分析步骤提供了一系列工具，用于在多个视角和尺度下观察医学影像，包括视图切换、仿射变换等等操作。
4. 辅助诊断：在用户通过可视化和影像分析获得初步印象后，辅助诊断步骤通过第三章和第四章中描述的深度学习框架，为假体关节感染和骨折相关感染提供辅助诊断能力。

从文件解析开始，DICOM 和 NIfTI 格式的医学影像数据由文件解析模块进行读取与解析。DICOM 数据由于其独有的详细性和医疗信息的丰富性，需要处理其复杂的元数据以提取诊断所需的重要数据；而 NIfTI 格式则常用于功能成像和三维影像处理，因此其解析旨在获取空间结构信息以便后续分析。

在完成解析之后，可视化模块将根据用户选择的医学影像数据和可视化功能进行相应的处理。该模块支持单模态和双模态影像的二维和三维可视化。在可视化过程中，医学影像数据均被默认设置为三维结构。对于原本是二维的医学影像数据，将自动扩展出一个大小为一的维度，以便与既定的三维数据结构保持一致性。因此，本章可视化模块的数据处理和分析侧重于三维及以下维度的医学影像。

影像分析模块根据用户的需求，对医学影像执行特定的操作。此模块功能广泛，包括展示影像坐标与在特定坐标上用于临床评估的影像值、切换不同视图，例如矢状面、冠状面和横断面、以及执行仿射变换操作等。每项功能都旨在为医生提供详细的医学影像信息，从而更准确地定位和诊断潜在的病灶区域。

最终，在完成医学影像的可视化以及初步分析后，辅助诊断模块融合了假体关节感染辅助诊断框架与骨折相关感染检测诊断框架，可以给用户提供医学辅助诊断建议。该建议包括在动态骨显像中对患者的膝部或髋部的假体进行诊断，以及利用 PET/CT 影像检测与识别患者下肢中潜在的感染区域。辅助诊断模块基于医学影像特征辅助医生评估病灶的性质以及严重程度，为医学诊断提供更高的准确性与效率。

### 5.2.1 界面设计

图5.2呈现的是所设计的骨科相关感染的自动诊断与可视化方法的界面。界面主要分为侧边导航栏和中心的可视化显示区域。在侧边导航栏中，展示了 DICOM 和

NIfTI 文件由文件解析模块处理所得的层次化结构信息。该信息采用了可折叠的列表按钮的形式，详细地列出了患者的姓名拼音与影像扫描的具体时间，并在其内部展示了归属于当前患者的影像数据，例如系列描述与影像尺寸信息。侧边栏顶部设置了多个功能按钮，包括打开文件、删除已解析的影像数据、二维影像展示、三维影像展示、二维影像融合以及三维影像融合。其中，功能按钮将作用于侧边导航栏中被勾选的医学影像数据。



图 5.2 系统界面

Figure 5.2 System interface

中央的可视化显示区域将依据操作者在侧边导航栏中执行的操作，以标签页的形式对医学影像数据进行可视化显示。图5.3展示了四种不同可视化功能下每个标签页顶部配备的工具栏，其中融合了影像分析模块与辅助诊断模块的众多功能。

在影像的二维可视化中，工具栏前端不仅显示了当前视图信息，同时展示了位于可视化区域中心蓝色十字准星所指示的像素位置坐标及其相关的影像值，这些信息用于评估和分析至关重要。值得说明的是，坐标可以进行修改，可视化区域也会随之进行更新。工具栏包含了多种功能按钮：视图按钮用于变换不同的视角；操作按钮允许用户复原图像、执行旋转及进行镜像翻转；鼠标左键图标按钮则使用户能够通过鼠标左键拖拽影像或移动蓝色虚线十字准星；切换按钮则可以让鼠标滑轮的

功能在切换切片和缩放之间进行转换；对比度按钮则用以调整可视化影像的对比度。此外，还包含了导入分割图像或者边界框的标签按钮，调节标签透明度的滑块，以及用于给出辅助诊断建议的 AI 按钮。

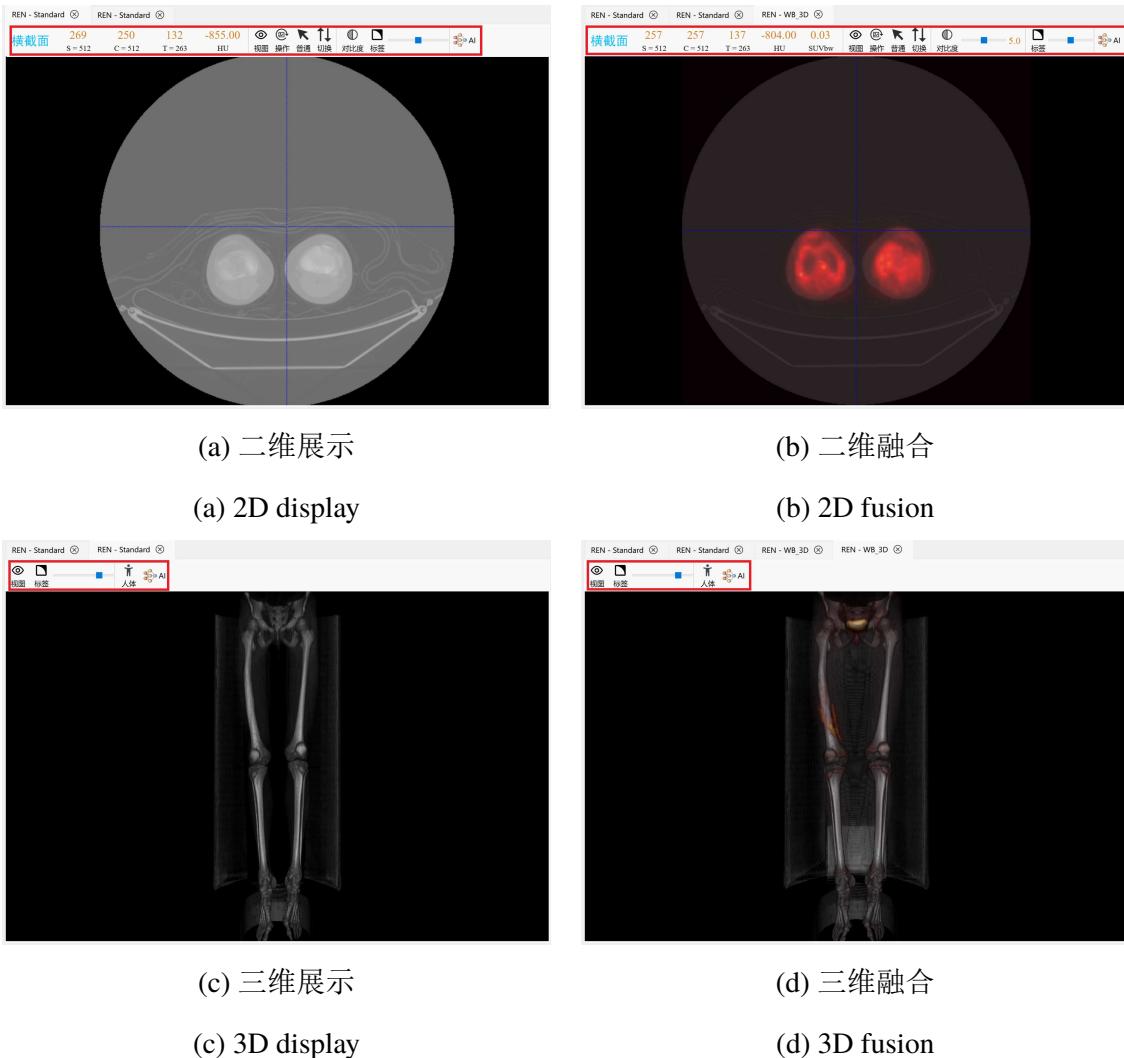


图 5.3 可视化界面

Figure 5.3 Visual interface

在影像的三维可视化工具栏中，视图功能按钮使用户能够选择展示整体影像数据或仅展示根据标签图导入的边界框所分割的部分；标签功能按钮方便用户导入标注病灶区域的边界框；透明度调节滑块通过精细控制，使得边界框标签透明度得以调整；而人体功能按钮则在 PET/CT 影像融合展示中，提供了移除机床视图的选项；同样地，AI 按钮针对三维影像提供辅助诊断功能。此外，用户可通过鼠标左键操控，使影像围绕其几何中心在三维空间内进行旋转，而鼠标滚轮则赋予了以影像中心作

为锚点进行放大或缩小的功能。

## 5.3 自动诊断

辅助诊断模块，作为本章研究中自动诊断方法的核心组成部分，集成了基于<sup>99m</sup>Tc-MDP 动态骨显像的假体关节感染辅助诊断框和基于<sup>18</sup>F-FDG PET/CT 三维影像的下肢骨折相关感染检测与诊断框架。在用户界面上，该集成的自动诊断方法通过一个简洁的 AI 按钮呈现于标签页的顶部工具栏中，从而在可视化显示区域中直观地提供使用接入点。

### 5.3.1 辅助诊断模块

通过简单地点击 AI 按钮，辅助诊断模块便可迅速触发内嵌的深度学习算法，进而自动对当前可视化的医学影像进行分析。特别地，辅助诊断模块会从医学影像的元数据中自动提取关键信息，以确保算法的适用性。算法开始对影像数据进行深入分析时，将对病灶区域进行精准的辅助诊断，并以可视化的方式展现出来。

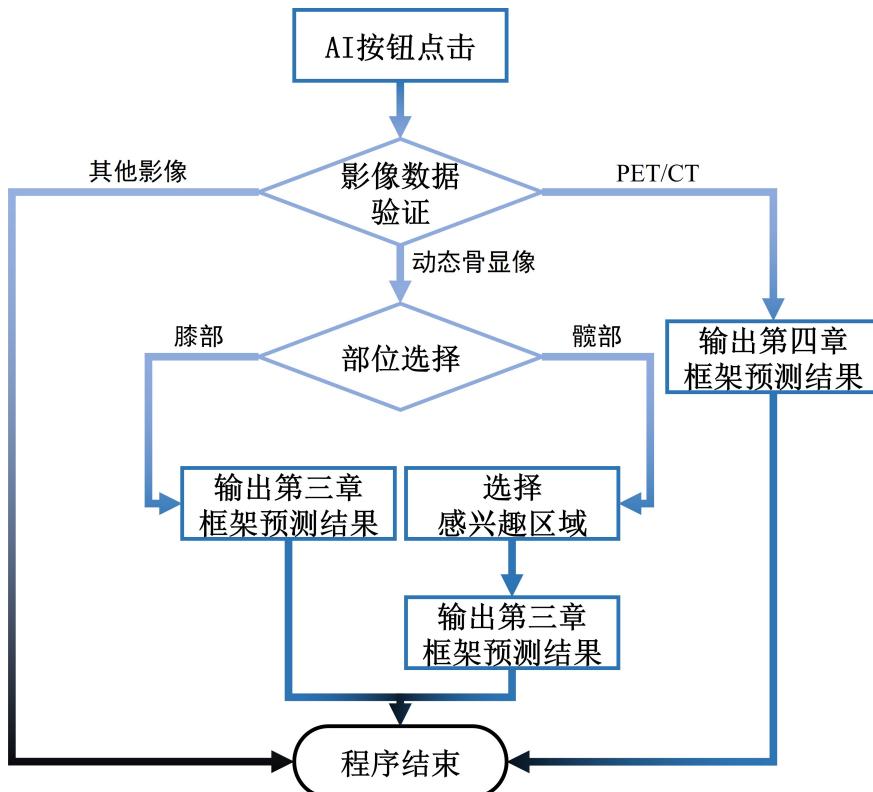


图 5.4 辅助诊断模块的运行流程

Figure 5.4 The processing flow of the auxiliary diagnosis module

图5.4展示了辅助诊断模块的整体运行流程。在初始阶段，对医学影像的元数据进行解析，从而识别所可视化的影像的模态。当遇到不兼容的影像模态时，辅助诊断模块会通过信息框提示用户，流程于此终止。

若影像模态确认为动态骨显像时，则会引导用户在膝部和髋部之间进行选择。选定膝部后，辅助诊断模块会立即采用第三章的框架，执行病灶的辅助诊断，同时通过计时弹窗提示程序正在运行中以及运行时间。髋部选择路径稍有不同，流程会进入一个选择感兴趣区域的环节。用户通过鼠标左键按下生成一个 $40 \times 40$ 的红色边框，并在可视化区域中移动到理想的区域。随后，释放鼠标左键，辅助诊断模块将以鼠标位置为中心的 $40 \times 40$ 的区域设置为感兴趣区域并提供给第三章的框架用于进一步分析和诊断。

若影像模态确认为 PET/CT 时，辅助诊断模块会自主调用第四章的检测诊断框架，开始识别和评估潜在的病灶区域，同时通过计时弹窗提示用户。在算法运行结束后，辅助诊断结果不仅以弹窗的形式对用户进行提示，同时在 PET/CT 中绘制出特定的边界框标注，从而实现了一个高效、直观且用户交互性强的辅助诊断工作流。

## 5.4 可视化

在本章研究中，可视化方法发挥着至关重要的作用，其由文件解析、可视化以及影像分析三大关键功能模块构成。文件解析模块构成了可视化方法的基石，它的主要职责在于解析影像文件数据，并将其重塑为三维医学影像，为之后的展示与分析奠定重要基础。可视化模块则构成了可视化方法的核心，该模块负责将解析后的医学影像数据转化并映射至自然图像领域中常规应用的色彩空间，其目的是提高影像的可读性的同时，使得医学影像的关键信息在视觉上更加明显。最后，影像分析模块代表了可视化方法的进阶扩展，它操控着对视觉呈现的医学影像的进一步图像处理，如调整对比度和缩放等。此模块便于用户进行细节分析，包括识别和解读影像上的关键特征，从而全面地挖掘影像所蕴含的深层次信息。

### 5.4.1 文件解析模块

在初始阶段，文件解析模块会将医学影像中原始的像素值转换为更适合用于评估和分析的影像值，用以确保了影像信息的临床相关性，还为后续的量化分析提供了定量的指标。例如，CT 影像中原始像素值被转换成 HU，PET 影像的原始像素值

则被转换为 SUV。

对于 DICOM 文件，解析过程如下：

- 1) 读取需解析的 DICOM 文件所在路径下的所有 DICOM 文件。
- 2) 根据元数据中的 Study Instance UID 将这些文件划分到不同的研究组中。
- 3) 根据元数据中的 Series Instance UID 进一步在研究组中划分为不同的系列组。
- 4) 根据每个文件的影像大小，在系列组中再进一步划分为不同的尺寸组。
- 5) 根据元数据中的 Slice Location 或 Instance Number 对尺寸组中的文件进行排序。
- 6) 依次将每个尺寸组的文件组成一个完整的医学影像，并转换像素值为影像值。
- 7) 数据以研究组、系列组、尺寸组和医学影像为顺序的层次化结构进行展示。

在 NIfTI (Neuroimaging Informatics Technology Initiative) 文件的解析过程中，由于这类文件并不包含像 DICOM 格式那样丰富的元数据，因此其解析过程相对简单。解析一开始是读入 NIfTI 文件，并为之分配随机生成的 Study Instance UID 与 Series Instance UID。接着读取该文件中的影像数据并获取影像大小。最终解析后的数据将按照一种与 DICOM 相同的层次化结构进行呈现。

#### 5.4.2 可视化模块

在文件解析结束之后，可视化模块将对不同模态的医学影像数据采取差异化的可视化处理策略，具体可分为二维和三维两个维度，各自又包含单一模态展示和双模态融合展示两种处理方法。

对于二维展示，为了将医学影像值有效地映射到自然图像领域中的 RGB 色彩空间，可视化模块采用了最大最小值归一化方法，确保医学影像值可以在保持其诊断信息的同时，以一种标准化且用户直观理解的视觉形式进行展现。其计算方式如下：

$$v = \text{floor}\left(\frac{v - \min(v)}{\max(v) - \min(v)} \times 255\right) \quad (5.1)$$

其中， $\text{floor}(\cdot)$  表示向下取整函数， $\max(\cdot)$  表示计算最大值的函数， $\min(\cdot)$  表示计算最小值的函数。

颜色映射是用于将数据值映射到颜色的一种方法，可以使数据分布、趋势和特征以更加直观的方式得以呈现。图5.5展示了三种颜色映射：`binary` 是一种简单的黑白二色的颜色映射，其中较小的值对应黑色而较大的值对应白色；`gray` 颜色映射通过灰阶色调展示图像或数据的亮度等级；`hot` 是一种常用于表示温度或热度的颜色映

射，整体上从黑色渐变到红色、黄色最终达到白色。这些颜色映射各具特色，能有效地增强数据的视觉效果和解读性。



图 5.5 三种颜色映射

Figure 5.5 Three color maps

接下来，不同模态的影像值将采用不同的颜色映射，以增强影像中各区域之间的对比度和其可视性效果，如算法5.1所示。

---

#### 算法 5.1：医学影像着色处理

---

输入：医学影像  $\mathcal{I}$

输出：颜色映射  $cm$

```

1 if  $\mathcal{I} = RGB$  then  $cm \leftarrow$  None;           // 无需进行着色处理
2 else if  $\mathcal{I} \in \{CT, MRI\}$  then  $cm \leftarrow$  gray;    // gray 体现不同解剖结构的层次
3 else if  $\mathcal{I} \in \{PET, NM\}$  then
4   if 双模态融合可视化 then  $cm \leftarrow$  hot;    // hot 在融合时强调高摄取区域
5   else  $cm \leftarrow$  binary;                      // 单模态可视化
6 else  $cm \leftarrow$  gray;

```

---

在完成医学影像的颜色映射之后，重要的一步是考虑可视化区域的尺寸  $(w, h)$ 、三维医学影像在各个方向上的尺寸  $(x, y, z)$  以及体素间距 ( $\text{Spacing} = (s_x, s_y, s_z)$ )。这一步骤是为了保证医学影像能够完整地填充整个可视化区域，以便用户能够得到一个清晰且准确的视图。因此，在进行三维医学影像切片的可视化处理时，需要计算相应的缩放比例，以便影像切片能够适配可视化区域。相关的缩放比例计算公式如下：

$$\begin{aligned}
 S_x^h &= s_y \times \min\left(\frac{h}{z \times s_z}, \frac{w}{y \times s_y}\right) & S_x^w &= s_z \times \min\left(\frac{h}{z \times s_z}, \frac{w}{y \times s_y}\right) \\
 S_y^h &= s_x \times \min\left(\frac{h}{z \times s_z}, \frac{w}{x \times s_x}\right) & S_y^w &= s_z \times \min\left(\frac{h}{z \times s_z}, \frac{w}{x \times s_x}\right) \\
 S_z^h &= s_x \times \min\left(\frac{h}{x \times s_x}, \frac{w}{y \times s_y}\right) & S_z^w &= s_y \times \min\left(\frac{h}{x \times s_x}, \frac{w}{y \times s_y}\right)
 \end{aligned} \tag{5.2}$$

其中,  $\mathcal{S}^h, \mathcal{S}^w$  分别代表三维影像中, 垂直于  $\cdot$  轴的切片在可视化过程中, 所需应用在高度与宽度方向上的缩放比例。

最终, 三维医学影像的对应切片将由可视化模块根据预设的坐标及视图参数来进行展现。当进行双模态影像的融合展示时, 首先是在两种不同模态的影像之间采用第四章中介绍的方法进行配准。完成配准后, 依据设定的坐标和视图参数, 两种模态的影像切片将被融合展示。图像融合可以采用以下加权平均的计算公式:

$$\text{Image} = \alpha \times \text{Image}_1 + \beta \times \text{Image}_2 \quad (5.3)$$

其中,  $\text{Image}_1, \text{Image}_2$  分别代表了待融合的两幅图像, 而融合系数  $\alpha, \beta$ , 分别取值为 0.3 和 0.7, 用以控制融合过程中各图像的权重。

对于三维展示, 本章提出的方法采用了 Visualization Toolkit (VTK) 库来实现医学影像的三维可视化处理。首先, 将医学影像数据包含的三维矩阵和体素间距以及自定义的影像原点导入到 VTK 中, 并转换成适合 VTK 处理的图像格式。为了确保生成的三维影像中心位于坐标原点  $(0, 0, 0)$ , 将医学影像的原点设定为:

$$\begin{aligned} o_x &= -0.5 \times \text{size}_x \times \text{spacing}_x \\ o_y &= -0.5 \times \text{size}_y \times \text{spacing}_y \\ o_z &= -0.5 \times \text{size}_z \times \text{spacing}_z \end{aligned} \quad (5.4)$$

其中,  $o_x, o_y, o_z$  分别代表医学影像在三维空间中的原点坐标,  $\text{size}_x, \text{size}_y, \text{size}_z$  分别代表医学影像在各轴方向上的大小,  $\text{spacing}_x, \text{spacing}_y, \text{spacing}_z$  则为各轴方向上的体素间距。

随后, 可视化模块使用了基于光线投射的体渲染技术, 该技术通过射出光线, 对体数据进行采样与体素插值, 从而生成包含了透明度、颜色以及光照效果的三维影像, 创造出具有立体感的体渲染图, 如图5.3(c)所示。

考虑到该体渲染技术的特性, 可视化模块设定了将医学影像的影像值转换为相应颜色以及透明度的映射表, 并设置了必要的光照参数, 包括环境光照、漫反射光照和镜面反射光照的强度。在三维模型构建完成后, 通过调整观察相机的位置和方向, 可以生成完整详尽的三维视图, 以供医学专家进行深入分析和解读。具体而言, 为展示完整的三维医学影像, 相机的视角设定为  $60^\circ$ , 焦点设置在  $(0, 0, 0)$ , 朝上方

向设置为  $(0, 0, 1)$ , 位置定位在  $(0, P_y, 0)$ 。其中  $P_y$  的计算公式如下:

$$P_y = -0.5 \times \max(size_x \times spacing_x, size_z \times spacing_z) \div \tanh(\pi/6) \quad (5.5)$$

在双模态融合展示方面, 可视化模块依据以上方法, 分别对两种不同模态的医学影像进行独立的三维渲染并直接进行展示, 如图5.3(d)所示。

### 5.4.3 影像分析模块

在完成可视化模块的处理后, 影像分析模块会给二维和三维展示的医学影像提供各种功能。如图5.3(a)和5.3(b)所示, 在二维影像分析时, 包含了全部功能的工具栏位于标签页顶部。包含视图、坐标和对应影像值的信息区域与可视化区域内的影像切片及蓝色十字标记相互联动, 任何一处发生变动, 其他位置也会随之更新。主要的映射关系包括:

1. 鼠标位置  $\rightarrow$  场景位置  $\rightarrow$  图像坐标
2. 图像坐标  $\rightarrow$  场景位置

坐标信息可以直接修改, 也可以通过鼠标滑轮。鼠标左键可移动蓝色十字, 视图也可以进行切换。操作按钮提供了实施旋转、镜像翻转以及图像重置的功能。普通按钮允许用户修改鼠标左键的操作, 以实施影像的平移变换, 而切换按钮则改变鼠标滚轮的作用, 从而实现影像的缩放变换。所有这些操作均通过应用仿射变换矩阵来完成。二维仿射变换矩阵如下所示:

$$\begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (5.6)$$

其中,  $a$  和  $d$  表示缩放因子,  $b$  和  $c$  用于旋转变换之中,  $t_x$  和  $t_y$  表示平移量。通过调整这些参数, 可以实现不同的仿射变换, 如下所示:

旋转	缩放	镜像	
$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(5.7)

通过对对比度按钮可以调节视图中医学影像的对比度，这是通过在弹出的窗口中输入窗位  $w_c$  和窗宽  $w_w$ ，或者最大值  $ma$  和最小值  $mi$  来实现的。具体而言，对医学影像值  $v$  的计算公式如下所示：

$$\begin{aligned} ma &= w_c + 0.5 \times w_w \\ mi &= w_c - 0.5 \times w_w \\ v &= \frac{\min(\max(mi, v), ma) - mi}{ma - mi} \end{aligned} \quad (5.8)$$

标签按钮可用来导入 NIfTI 格式的标签数据，以渲染分割图或显示边界框；标签滑块则可调整标签图的透明度。

在三维影像分析时，所有功能均汇聚于标签页顶部的工具栏中，如图5.3(c)和5.3(d)所示。视图按钮允许用户选择性地渲染特定的影像数据，以此进行独立的视图展示。标签按钮用于导入 NIfTI 或 Json 格式的标签数据，这使得用户能够渲染出三维的立体边界框和关联分类标签。通过标签滑块，用户可细致地调整标签图层的透明度，以便更好地视察底层的医学影像。人体按钮采用了第四章中介绍的人体保留算法，专门针对 PET/CT 影像数据进行优化处理，去除掉机床等无关区域。此外，鼠标左键的操作，实现了观察摄像头在以原点为中心的虚拟球面的移动，来达成三维医学影像的旋转变换。同时，通过鼠标滑轮调节，可以改变三维影像与观察摄像头之间的距离，从而实现三维影像的缩放效果。

## 5.5 案例分析

在本节中，将示范自动诊断与可视化方法对上海市第六人民医院核医学科 2016 年的三个病例的应用实践：

- (1) 患者 A 在膝关节置换手术后两月接受了膝部的动态骨显像扫描，医生通过影像分析初步判定膝假体处可能感染，并在评估病症、临床表现以及影像与血液检验结果后，确诊为膝关节假体感染；
- (2) 患者 B 在髋关节置换术后两年进行了髋部动态骨显像扫描。经影像学分析后，医生怀疑髋假体发生感染，最终依据全面诊断确诊为假体髋关节感染；
- (3) 患者 C 在右股骨骨折治疗后接受 PET/CT 扫描。经过影像学判断，疑似存在感染性病灶。借助微生物培养、病理分析及临床综合判断，确认患有骨折相关感染。

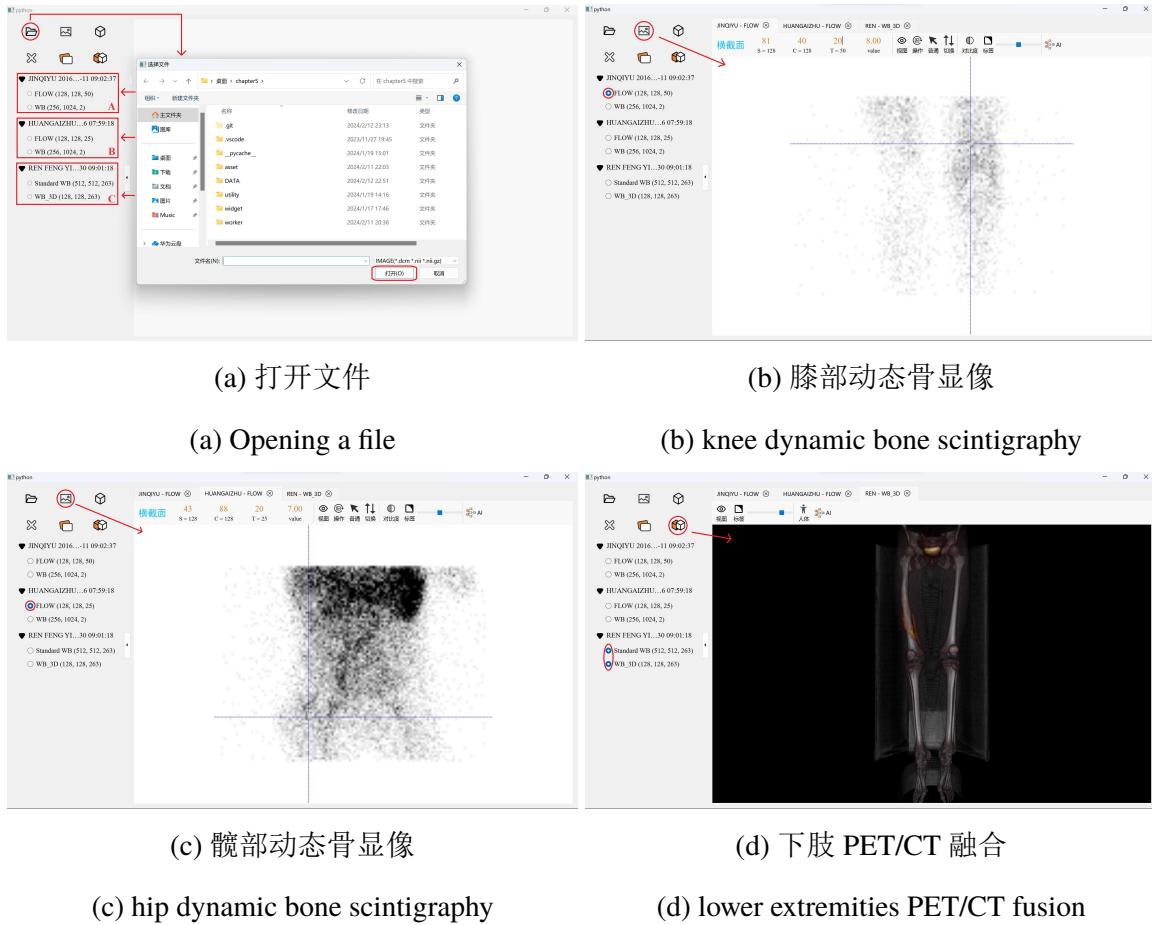


图 5.6 案例可视化

Figure 5.6 Case visualization

如图5.6(a)所示，通过在侧边导航栏顶部的“打开文件”功能按钮，可以在启动的文件选择窗口中依次打开患者 A、患者 B 和患者 C 的 DICOM 医学影像文件。影像文件被打开并成功解析后，结果在侧边导航栏中清晰地展示了出来。

如图5.6(b)所示，通过选中患者 A 的动态骨显像并点击“二维影像展示”按钮，从而呈现了患者 A 膝部的二维可视化影像。图5.6(c)则展现了类似操作过程，勾选患者 B 的动态骨显像并点击相同的功能按钮，实现了髋部影像的二维可视化。随后，使用可视化区域顶部的影像分析工具。通过调整影像的对比度，以获得更清晰细致的可视化效果。在此基础上，灵活地切换不同的切片，搜索并定位到最合适的影像切片。最后，移动蓝色十字光标至特定区域，以查看相应的影像值，为进一步的分析和评估提供关键信息。

通过可视化区域顶部工具栏最右侧的 AI 功能按钮，可以激活辅助诊断模块的自动诊断功能，并获得基于深度学习方法生成的辅助诊断建议。图5.7(a)所展示的是患

者 A 的案例，辅助诊断结果建议患者存在假体关节感染的情况。同时，图5.7(b)中以透明红色矩形标识了患者 B 髋部中的感兴趣区域，并表明了患者 B 存在假体关节感染的诊断建议。

图5.6(d)展现了患者 C 的 PET 和 CT 影像数据，通过选中并点击“三维影像融合”按钮后，得到的下肢立体融合影像的可视化图。此后，利用影像分析工具可以对该立体影像进行缩放和旋转操作，以便对特定感兴趣的区域进行详细的观察和分析。点击 AI 辅助诊断按钮后，如图5.7(c)所呈现的，患者 C 被智能辅助诊断为存在一个感染区域，并且该区域在患者 C 的右股骨处被一个青色立体框所标记。最后借助于视图按钮，还能够从不同角度观察并分析感染区域的影像数据，如图5.7(d)所示。



图 5.7 自动诊断方法

Figure 5.7 Automatic diagnosis method

通过对三个具体病例的详细展示和分析，提供的辅助诊断建议与经验丰富的医生所作出的诊断高度一致。这表明本章所介绍的自动诊断与可视化方法，不仅赋予了医生一个实用的影像分析工具，还可以通过整合先进的深度学习技术提出极具参考价值的辅助诊断建议。

## 5.6 本章小结

本章提出了一种用于骨科相关感染的自动诊断和可视化方法。此方法结合了前两章提出的假体关节感染辅助诊断框架与骨折相关感染检测与诊断框架，以便进行对动态骨扫描和 PET/CT 影像的自动化处理和分析。同时，本方法还包含了从文件读取解析到医学影像的各种可视化，以及涵盖多项分析工具的影像分析，实现对医学影像中潜在病灶的定位与分析。本章最后通过三个真实病例的分析，验证了提出的辅助诊断建议与患者真实情况之间基本一致。这证明了本章方法，不仅为用户提供了一个全面而高效的可视化交互和分析工具，而且有助于促进医生的临床决策和提高准确性。



## 第六章 总结与展望

### 6.1 总结

本文聚焦于在核医学影像中进行骨科相关感染疾病的辅助诊断，具体涉及假体关节感染和骨折相关感染。动态骨显像具有时间序列特性且包含三种不同阶段的相，PET 和 CT 可以提供更全面的三维影像信息，以及动态骨显像和 PET/CT 中均存在自然高摄取区域的干扰等等。针对这些问题和特点，本文提出了基于动态骨显像的假体关节感染辅助分类诊断框架和基于下肢 PET/CT 影像中骨折相关感染的检测与诊断框架。本文的主要工作总结如下：

(1) 面对骨科相关感染领域医学影像数据集的不足，本文作者与上海市第六人民医院核医学科进行了合作，共同构建了两个核心数据集。动态骨显像数据集涵盖了 2016 年 1 月至 2021 年 6 月期间 455 位患者的影像及其匹配的诊断报告。PET/CT 数据集则收集了 2016 年 11 月至 2021 年 12 月期间 281 位患者的医学影像和诊断报告。在医学专家的指导下，对这两个数据集进行了精确的数据标注，确保其高度的准确性和可信度。此外，本文还特别为动态骨显像和 PET/CT 影像设计了专门的数据预处理方法，旨在统一数据格式，提升影像质量，并减少无关区域和自然高摄取区域带来的不利影响。

(2) 为提升动态骨显像在假体关节感染诊断中的应用效果，本文提出了一个包含了预处理和融合三维卷积与 ConvLSTM 的辅助诊断模型（DBS-eNet）的辅助诊断框架。DBS-eNet 充分挖掘了动态骨显像多个相及其内部的图像序列信息。通过三维卷积技术，模型能够在时间维度上有效提取每幅图像的生理摄取形态特征，而 ConvLSTM 则在强化图像的形态特征的同时，能够捕捉图像序列中生理代谢活动的动态变化。通过实验证明，DBS-eNet 在处理假体膝关节感染与假体髋关节感染诊断任务时表现出色，准确率在五折交叉验证测试中分别达到了 86.48% 和 86.33%，并在与核医学专家的独立验证比较中，膝关节诊断准确率提高了 13.55%，髋关节诊断准确率与专家持平。

(3) 面向下肢骨折相关感染的 PET/CT 影像诊断任务，本文提出了一个创新的两阶段框架——3DFRINet。该框架包含了病灶检测网络与病灶诊断分类网络。病灶检

测网络通过双分支网络的架构设计和注意力机制，高效融合不同模态下的多尺度特征，实现病灶的精确定位。紧随其后的病灶诊断分类网络，运用最大强度投影技术将三维影像信息投影至二维，不仅保留了关键特征，也深入分析了感染病灶的代谢形态，以达到精准分类的目标。实验结果显示，3DFRINet 在下肢 PET/CT 影像病灶检测任务中表现卓越，超越了其他 YOLO 模型，其 AP<sub>25</sub> 指标高达 0.9234；而在对骨折相关感染病灶进行诊断分类时，相比于资历较浅的核医学医师，准确率提高了 11.27%，与资深核医学医师相比也提升了 4.23%。

(4) 本文最后设计了一套针对骨科相关感染的自动诊断与可视化方法。该方法融合了(2)和(3)的模型，不仅支持假体关节感染的诊断分类工作，还能自动检测并分类骨折相关感染的病灶。此外，为了适应各种医学影像模态，该方法支持二维、三维以及 PET/CT 融合影像的可视化分析。在功能性设计上，该方法为二维影像提供了诸如医学影像值及其坐标呈现、视图切换等功能；而在三维影像方面，则允许用户进行旋转、缩放等操作，以促进对复杂医学影像数据的理解和分析。

综上所述，本文实现了假体关节感染和骨折相关感染的完整诊断流程。实验对比、可视化方法以及案例分析都证实了本文研究的可信度。本文可以为医疗诊断提供了客观的建议，助力医学影像智能诊断领域向规范化及标准化迈进。

## 6.2 展望

本文成果虽已取得一定的成就，但在未来的研究中仍存在以下几个发展方向：

(1) 尽管本文所构建的数据集对于实验而言足够，但它的样本规模较小、类别分布不均及数据来源单一。未来可以收集和补充更多不同医院的患者数据，来提升模型的泛化能力，并减少数据集本身带来的偏误。

(2) 在假体髋关节感染的诊断任务中，仍然依赖人工标注的兴趣区域。在未来的研究中，开发一套完全自动化的诊断系统将是有价值的探索方向。

(3) 理论上，动态骨显像的延迟相对假体关节感染的诊断贡献不大，未来可以考虑进行验证。SPECT 也被用于诊断假体关节感染，结合深度学习和 SPECT 进行辅助诊断，可能具有研究价值。PET/CT 中，衰减校正与否对 PET 在深度学习中诊断效果的影响值得深入研究，同时伪影消除技术去除 CT 中的金属伪影是否能提高深度学习模型的性能一样值得研究。

## 参考文献

- [1] MORGAN E F, GERSTENFELD L C. The bone organ system: form and function [M]//Marcus and Feldman's Osteoporosis. Elsevier, 2021: 15-35.
- [2] DE BUFFRÉNIL V, DE RICQLÈS A J, ZYLBERBERG L, et al. Vertebrate skeletal histology and paleohistology[M]. crc Press, 2021.
- [3] SIDIROPOULOS K, CHRISTOFILOS S I, TSIKOPoulos K, et al. Viral infections in orthopedics: A systematic review and classification proposal[J]. World Journal of Orthopedics, 2022, 13(11): 1015.
- [4] SCONFRENZA L M, SIGNORE A, CASSAR-PULLICINO V, et al. Diagnosis of peripheral bone and prosthetic joint infections: overview on the consensus documents by the eanm, ejbis, and esr (with escmid endorsement)[J]. European radiology, 2019, 29: 6425-6438.
- [5] NICCOLI G, MERCURIO D, CORTESE F. Bone scan in painful knee arthroplasty: obsolete or actual examination?[J]. Acta Bio Medica: Atenei Parmensis, 2017, 88 (Suppl 2): 68.
- [6] USUELLI F G, INDINO C, MACCARIO C, et al. Infections in primary total ankle replacement: anterior approach versus lateral transfibular approach[J]. Foot and Ankle Surgery, 2019, 25(1): 19-23.
- [7] ALEKSYNIENE R, IYER V, BERTELSEN H C, et al. The role of nuclear medicine imaging with 18f-fdg pet/ct, combined 111in-wbc/99mtc-nanocoll, and 99mtc-hdp spect/ct in the evaluation of patients with chronic problems after tka or tha in a prospective study[J]. Diagnostics, 2022, 12(3): 681.
- [8] MEERMANS G, HADDAD F S. Is there a role for tissue biopsy in the diagnosis of periprosthetic infection?[J]. Clinical Orthopaedics and Related Research®, 2010, 468: 1410-1417.

- [9] PARVIZI J, JACOVIDES C, ADELI B, et al. Mark b. coventry award: synovial c-reactive protein: a prospective evaluation of a molecular marker for periprosthetic knee joint infection[J]. Clinical Orthopaedics and Related Research®, 2012, 470(1): 54-60.
- [10] FINK B, GEBHARD A, FUERST M, et al. High diagnostic value of synovial biopsy in periprosthetic joint infection of the hip[J]. Clinical Orthopaedics and Related Research®, 2013, 471: 956-964.
- [11] PAUL H Y, CROSS M B, MORIC M, et al. The 2013 frank stinchfield award: diagnosis of infection in the early postoperative period after total hip arthroplasty[J]. Clinical Orthopaedics and Related Research®, 2014, 472(2): 424-429.
- [12] FINK B, STEURER M, HOFÄCKER S, et al. Preoperative pcr analysis of synovial fluid has limited value for the diagnosis of periprosthetic joint infections of total knee arthroplasties[J]. Archives of Orthopaedic and Trauma Surgery, 2018, 138(6): 871-878.
- [13] CLAASSEN L, RADTKE K, ETTINGER M, et al. Preoperative diagnostic for periprosthetic joint infection prior to total knee revision arthroplasty[J]. Orthopedic Reviews, 2014, 6(3).
- [14] GOVAERT G A, KUEHL R, ATKINS B L, et al. Diagnosing fracture-related infection: current concepts and recommendations[J]. Journal of orthopaedic trauma, 2020, 34(1): 8.
- [15] METSEMAKERS W J, SMEETS B, NIJS S, et al. Infection after fracture fixation of the tibia: analysis of healthcare utilization and related costs[J]. Injury, 2017, 48(6): 1204-1210.
- [16] KTISTAKIS I, GIANNOUDI M, GIANNOUDIS P V. Infection rates after open tibial fractures: are they decreasing?[J]. Injury, 2014, 45(7): 1025-1027.
- [17] MALHOTRA A K, GOLDBERG S, GRAHAM J, et al. Open extremity fractures: impact of delay in operative debridement and irrigation[J]. Journal of Trauma and Acute Care Surgery, 2014, 76(5): 1201-1207.

- [18] BHOIL A, CAW H, VINJAMURI S. Role of 18f-flurodeoxyglucose in orthopaedic implant-related infection: Review of literature and experience[J]. Nuclear Medicine Communications, 2019, 40(9): 875-887.
- [19] American joint replacement registry (ajrr): 2021 annual report[R]. Rosemont, IL: American Academy of Orthopaedic Surgeons (AAOS), 2021.
- [20] KAMATH A F, ONG K L, LAU E, et al. Quantifying the burden of revision total joint arthroplasty for periprosthetic infection[J]. The Journal of arthroplasty, 2015, 30(9): 1492-1497.
- [21] PREMKUMAR A, KOLIN D A, FARLEY K X, et al. Projected economic burden of periprosthetic joint infection of the hip and knee in the united states[J]. The Journal of arthroplasty, 2021, 36(5): 1484-1489.
- [22] THAKORE R V, GREENBERG S E, SHI H, et al. Surgical site infection in orthopedic trauma: a case-control study evaluating risk factors and cost[J]. Journal of clinical orthopaedics and trauma, 2015, 6(4): 220-226.
- [23] VERBERNE S, RAIJMAKERS P, TEMMERMAN O. The accuracy of imaging techniques in the assessment of periprosthetic hip infection: a systematic review and meta-analysis[J]. JBJS, 2016, 98(19): 1638-1645.
- [24] KUMAR R, KUMAR R, KUMAR V, et al. Comparative analysis of dual-phase 18f-fluoride pet/ct and three phase bone scintigraphy in the evaluation of septic (or painful) hip prostheses: A prospective study[J]. Journal of Orthopaedic Science, 2016, 21(2): 205-210.
- [25] ZHANG F, SHEN C, YU J, et al. The temporal impact of prosthesis implantation and semi-quantitative criteria on the diagnostic efficacy of triple-phase bone scanning for periprosthetic joint infection[J]. Orthopaedic Surgery, 2022, 14(7): 1438-1446.
- [26] ZHANG Q, DONG J, SHEN Y, et al. Comparative diagnostic accuracy of respective nuclear imaging for suspected fracture-related infection: a systematic review and bayesian network meta-analysis[J]. Archives of Orthopaedic and Trauma Surgery, 2021, 141: 1115-1130.

- [27] GHOLAMREZANEZHAD A, BASQUES K, BATOULI A, et al. Clinical nononcologic applications of pet/ct and pet/mri in musculoskeletal, orthopedic, and rheumatologic imaging[J]. AJR. American journal of roentgenology, 2018, 210(6): W245.
- [28] LEMANS J V, HOBBELINK M G, IJPMA F F, et al. The diagnostic accuracy of <sup>18</sup>f-fdg pet/ct in diagnosing fracture-related infections[J]. European Journal of Nuclear Medicine and Molecular Imaging, 2019, 46: 999-1008.
- [29] WENTER V, MÜLLER J P, ALBERT N L, et al. The diagnostic value of [18 f] fdg pet for the detection of chronic osteomyelitis and implant-associated infection[J]. European journal of nuclear medicine and molecular imaging, 2016, 43: 749-761.
- [30] 刘超, 邱金旭, 刘世成. 人工智能在影像诊断中的应用及研究进展[J]. 中国实验诊断学, 2023, 27(873-875).
- [31] 李小雷. 人工智能在医学影像图像处理中的研究进展[J/OL]. 中国医学计算机成像杂志, 2023, 29(454-457). DOI: [10.19627/j.cnki.cn31-1700/th.2023.04.001](https://doi.org/10.19627/j.cnki.cn31-1700/th.2023.04.001).
- [32] LIU Z, LU H, PAN X, et al. Diagnosis of alzheimer' s disease via an attention-based multi-scale convolutional neural network[J]. Knowledge-Based Systems, 2022, 238: 107942.
- [33] TAKAHASHI K, FUJIOKA T, OYAMA J, et al. Deep learning using multiple degrees of maximum-intensity projection for pet/ct image classification in breast cancer[J]. Tomography, 2022, 8(1): 131-141.
- [34] FU K S, et al. Pattern recognition and image processing[J]. IEEE transactions on computers, 1976, 100(12): 1336-1346.
- [35] 季长清, 高志勇, 秦静, 等. 基于卷积神经网络的图像分类算法综述[J]. 计算机应用, 2022, 42(4): 1044.
- [36] 杨真真, 匡楠, 范露, 等. 基于卷积神经网络的图像分类算法综述.[J]. Journal of Signal Processing, 2018, 34(12).
- [37] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

- [38] LIU Z, LUO P, WANG X, et al. Large-scale celebfaces attributes (celeba) dataset[J]. Retrieved August, 2018, 15(2018): 11.
- [39] GUO G, WANG H, BELL D, et al. Knn model-based approach in classification[C]// On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings. Springer, 2003: 986-996.
- [40] SUTHAHARAN S, SUTHAHARAN S. Decision tree learning[J]. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, 2016: 237-269.
- [41] SUTHAHARAN S, SUTHAHARAN S. Support vector machine[J]. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 2016: 207-235.
- [42] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [43] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [44] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1): 1929-1958.
- [45] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. CoRR, 2014, abs/1409.1556. <https://api.semanticscholar.org/CorpusID:14124313>.
- [46] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [47] 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. 2016: 770-778.

- [48] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [49] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-11986.
- [50] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [51] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448-456.
- [52] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J/OL]. CoRR, 2020, abs/2010.11929. <http://arxiv.org/abs/2010.11929>.
- [53] HAN K, XIAO A, WU E, et al. Transformer in transformer[J]. Advances in Neural Information Processing Systems, 2021, 34: 15908-15919.
- [54] MEMORY L S T. Long short-term memory[J]. Neural computation, 2010, 9(8): 1735-1780.
- [55] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J/OL]. CoRR, 2014, abs/1412.3555. <http://arxiv.org/abs/1412.3555>.
- [56] YİĞİT G, AMASYALI M F. Simple but effective gru variants[C]//2021 international conference on INnovations in intelligent SysTems and applications (INISTA). IEEE, 2021: 1-6.
- [57] CHENG D, LIU M. Classification of alzheimer' s disease by cascaded convolutional neural networks using pet images[C]//Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8. Springer, 2017: 106-113.

- [58] LIU M, CHENG D, YAN W, et al. Classification of alzheimer' s disease by combination of convolutional and recurrent neural networks using fdg-pet images[J]. *Frontiers in neuroinformatics*, 2018, 12: 35.
- [59] KHAGI B, KWON G R. 3d cnn design for the classification of alzheimer' s disease using brain mri and pet[J/OL]. *IEEE Access*, 2020, 8: 217830-217847. DOI: [10.1109/ACCESS.2020.3040486](https://doi.org/10.1109/ACCESS.2020.3040486).
- [60] LEE S Y, KANG H, JEONG J H, et al. Performance evaluation in [18f] florbetaben brain pet images classification using 3d convolutional neural network[J]. *Plos one*, 2021, 16(10): e0258214.
- [61] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015: 234-241.
- [62] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [63] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [64] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSH N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, 2018: 3-11.
- [65] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[J/OL]. *CoRR*, 2018, abs/1804.03999. <http://arxiv.org/abs/1804.03999>.

- [66] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2020: 1055-1059.
- [67] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J/OL]. CoRR, 2021, abs/2102.04306. <http://arxiv.org/abs/2102.04306>.
- [68] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [69] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
- [70] REDMON J, FARHADI A. Yolo9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [71] REDMON J, FARHADI A. Yolov3: An incremental improvement[J/OL]. CoRR, 2018, abs/1804.02767. <http://arxiv.org/abs/1804.02767>.
- [72] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J/OL]. CoRR, 2020, abs/2004.10934. <http://arxiv.org/abs/2004.10934>.
- [73] WANG C Y, LIAO H Y M, WU Y H, et al. Cspnet: A new backbone that can enhance learning capability of cnn[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.
- [74] MISRA D. Mish: A self regularized non-monotonic activation function[J/OL]. CoRR, 2019, abs/1908.08681. <http://arxiv.org/abs/1908.08681>.
- [75] 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. 2018: 8759-8768.
- [76] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

- [77] JOCHER G, NISHIMURA K, MINERVA T, et al. Yolov5 <https://github.com/ultralytics/yolov5>[J]. Accessed March, 2020, 7: 2021.
- [78] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J/OL]. CoRR, 2021, abs/2107.08430. <http://arxiv.org/abs/2107.08430>.
- [79] TANG H, KIM D R, XIE X. Automated pulmonary nodule detection using 3d deep convolutional neural networks[C]//2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018: 523-526.
- [80] LI Z, ZHANG S, ZHANG J, et al. Mvp-net: multi-view fpn with position-aware attention for deep universal lesion detection[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer, 2019: 13-21.
- [81] AL-MASNI M A, KIM W R, KIM E Y, et al. A two cascaded network integrating regional-based yolo and 3d-cnn for cerebral microbleeds detection[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 1055-1058.
- [82] MEI S, JIANG H, MA L. Yolo-lung: a practical detector based on imporved yolov4 for pulmonary nodule detection[C]//2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2021: 1-6.
- [83] LEE J, KANG J, SHIN E Y, et al. Dual-path connected cnn for tumor segmentation of combined pet-ct images and application to survival risk prediction[M]//3D Head and Neck Tumor Segmentation in PET/CT Challenge. Springer, 2021: 248-256.
- [84] WANG H, WU Z, WANG F, et al. Maff: Multi-scale and self-adaptive attention feature fusion network for pancreatic lesion detection in pet/ct images[C]//Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering. 2022: 1412-1419.
- [85] HUANG Z, ZOU S, WANG G, et al. Isa-net: Improved spatial attention network for pet-ct tumor segmentation[J]. Computer Methods and Programs in Biomedicine, 2022, 226: 107129.

- [86] NIE L, SUN Z, SHAN F, et al. An artificial intelligence framework for the diagnosis of prosthetic joint infection based on 99mtc-mdp dynamic bone scintigraphy[J]. European Radiology, 2023: 1-10.
- [87] SCHAUWECKER D S. The scintigraphic diagnosis of osteomyelitis.[J]. AJR. American journal of roentgenology, 1992, 158(1): 9-18.
- [88] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [89] ZHANG C B, JIANG P T, HOU Q, et al. Delving deep into label smoothing[J]. IEEE Transactions on Image Processing, 2021, 30: 5984-5996.
- [90] SHI X, CHEN Z, WANG H, et al. Convolutional lstm network: A machine learning approach for precipitation nowcasting[J]. Advances in neural information processing systems, 2015, 28.
- [91] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.
- [92] VAN DER MAATEN L, HINTON G. Visualizing data using t-sne.[J]. Journal of machine learning research, 2008, 9(11).
- [93] MILDENBERGER P, EICHELBERG M, MARTIN E. Introduction to the dicom standard[J]. European radiology, 2002, 12: 920-927.
- [94] HARRIS C R, MILLMAN K J, VAN DER WALT S J, et al. Array programming with numpy[J]. Nature, 2020, 585(7825): 357-362.
- [95] MCNALLY M, SOUSA R, WOUTHUYZEN-BAKKER M, et al. The ebjis definition of periprosthetic joint infection: a practical guide for clinicians[J]. The bone & joint journal, 2021, 103(1): 18-25.
- [96] HESTERBERG T. Bootstrap[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2011, 3(6): 497-526.

- [97] LACHENBRUCH P A. Mcnemar test[J]. Wiley StatsRef: Statistics Reference Online, 2014.
- [98] VICKERS A J, ELKIN E B. Decision curve analysis: a novel method for evaluating prediction models[J]. Medical Decision Making, 2006, 26(6): 565-574.
- [99] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[C]//International Conference on Learning Representations. 2018.
- [100] TAKAHASHI R, MATSUBARA T, UEHARA K. Ricap: Random image cropping and patching data augmentation for deep cnns[C]//Asian conference on machine learning. PMLR, 2018: 786-798.
- [101] YANIV Z, LOWEKAMP B C, JOHNSON H J, et al. Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research[J]. Journal of digital imaging, 2018, 31(3): 290-303.
- [102] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.



## 攻读硕士学位期间取得的研究成果

### 一、论文

[1] Liangbing Nie, Zhenkui Sun, Fengling Shan, Chengfan Li, Xuehai Ding and Chentian Shen, An artificial intelligence framework for the diagnosis of prosthetic joint infection based on  $^{99m}\text{Tc}$ -MDP dynamic bone scintigraphy, European Radiology, 2023, 33(10): 6794-803, DOI: 10.1007/s00330-023-09687-w. (SCIE 2 区 Top, IF=5.9)

[2] Chengfan Li, Liangbin Nie, Zhenkui Sun, Xuehai Ding, Quanyong Luo, and Chentian Shen, 3DFRINet: A Framework for the Detection and Diagnosis of Fracture Related Infection in Low Extremities Based on  $^{18}\text{F}$ -FDG PET/CT 3D Images, Computerized Medical Imaging and Graphics. (SCIE 2 区, IF=5.7, 二审中)

[3] Nianting Ju, Liangbing Nie, Yang Wang, Liying Hou, Chengfan Li, Xuehai Ding, Quanyong Luo and Chentian Shen, Predicting  $^{18}\text{F}$ -FDG SUVs of metastatic pulmonary nodes from CT images in patients with differentiated thyroid cancer by using a convolutional neural network, Frontiers in Endocrinology, 2023 14, DOI: 10.3389/fendo.2023.1127741. (SCIE 2 区, IF=5.2, 共同一作)

### 二、专利

[1] 人工智能模型训练方法及装置 - CN202310324627.1 - 发明已公开  
[2] 人工智能模型训练方法及装置、计算机可读存储介质 - CN202310183902.2 -  
发明已公开

### 三、项目

1. 上海市公共卫生重点学科建设项目“基于新技术、新方法的放射卫生检测适宜技术研究”(GWVI-11.1-40), 参与。
2. 上海市军民融合项目“基于机器视觉的智能射击辅助训练系统”(HCXBCY-2023-050), 参与。
3. 上海市第六人民医院项目“分化型甲状腺癌  $^{131}\text{I}$  治疗数据库软件”(22H00352) 和“基于深度学习的肾脏图像 ROI 分割与慢性肾病分类方法研究”(23H00752), 参与。



## 致 谢

犹如一轮盛大的乐章即将落幕，我的硕士生涯快要步入历史的长河。遥望那些滚滚流转的岁月，它们如同一部错落有致、峥嵘曲折的史诗，让我感触良多。我于那五彩斑斓的回忆中漫步，时空似乎变得让人眷恋：春日阳光灿烂，仿佛将一切美好尽收眼底。夏日微风徐来，轻轻抚过肌肤，带走了炎炎烈日的烈热。秋日的落叶纷落，像是岁月的诗行，寂静中述怀过往。冬日的夕阳，又像是初下的暖阳，意味着岁末和新的开始。

在这个冒险与学习的旅程中，我经历了人生的极致起落。情感的峰谷，随着日子的推移，不停地跳动着人生的节奏，如诗如画。科研的曲折，像是星辰引导的道路，让我经历了苦辣酸甜，收获了成长的痛苦与喜悦。须知疼痛之后，才能学会真正的坚韧和坚定。我在平淡日子中，学会面对质朴的自我，包容并热爱我所在的世界。

作为一名硕士研究生，我不仅在知识的海洋中畅游，同样在人生的沙漠中找到了绿洲。那些举世无双的伙伴和朋友，如同金子一般珍贵，在我生命中留下了彩虹般的痕迹。而最重要的是，我找到了那个愿意与我共度余生，波澜岁月，风雨共渡的伴侣。对所有这些，我怀感激之情，无以言表，只愿感恩生活的每一个奇迹。

梦想照耀未来，星辰指引道路。人生大道上，我将带着满载的希望与热忱，继续前行不息，永不言弃。

最后，我愿对我尊敬的李成范老师和丁雪海老师表达最深的感恩。感谢丁老师在学业上的精心指导，使我在知识的森林中找到了道路。感谢老师们在日常生活中的照料和关心，使我在陌生的城市中找到了家的温暖。您们的一切，如同灯塔，指引我照亮前程。我将带着感激，踏上新的征程。

