

# 基于 Mamba 的高分辨率数字病理 图像分类研究



## 重庆大学硕士学位论文 (专业学位)

学生姓名：朱 翔

指导教师：黄 昊 教师

专业学位类别：电子信息

(软件工程)

研究方向：计算机视觉

答辩委员会主席：熊庆宇 教授

授位时间：2025 年 6 月

# **Research on High-Resolution Digital Pathological Image Classification Based on Mamba**



A thesis submitted to Chongqing University  
in partial fulfillment of the requirement  
for the professional degree of Master of  
electronic information

by

**Xiang Zhu**

(Software Engineering)

**Supervisor: Prof. Sheng Huang**

**June , 2025**

## 摘要

医学影像分析作为疾病诊断的“金标准”，是临床医疗的重要支撑。其中，数字病理图像分类作为医学影像分析的核心领域，不仅是构建临床辅助诊断技术体系的关键，更是实现精准医疗的重要技术路径。在数字病理图像分类任务中，由于病理图像的高分辨率特性以及微小病灶区域等原因，通常采用多实例学习框架，但大量实例会导致计算成本高、实例关系建模不充分等问题。当前研究表明 Mamba 架构具有极强的长序列建模能力且仅有线性计算复杂度，适合解决上述病理图像分类难题，但其中仍然存在空间建模不充分、与病理图像多实例学习的序列不变性存在冲突等问题。为攻克这一挑战，实现 Mamba 架构在病理图像领域的高效适配，本研究从多扫描 Mamba 和扫描不变性 Mamba 两个关键视角展开深入探索。

(1) 针对 Mamba 一维序列建模所造成图像空间结构捕获能力不足的问题，本研究提出基于空间信息增强的多扫描 Mamba 多实例学习框架 ( $M^2S\text{-MIL}$ )。该框架创新性地设计了多个扫描模式分支作为 Mamba 的输入，提出了区域螺旋扫描和网格扫描等新型扫描模式，显著提升 Mamba 对局部连续性、旋转不变性及同一序列不同方向空间关联信息的捕获能力。同时，通过设计基于分块局部注意力的 2D 空间上下文感知模块，进一步强化序列对 2D 空间上下文的感知能力，极大提高 Mamba 对空间关系的捕捉，有力推动 Mamba 模型与病理图像分类任务的适配进程。大量实验表明，综合多种扫描模式分支并引入空间信息，能够切实解决 Mamba 一维实例序列空间建模不充分的问题，显著提高模型性能，同时在数字病理图像分类任务中一定程度缓解了使用 Mamba 模型出现的建模倾向差异难题。

(2) 针对 Mamba 作为序列相关模型，与病理图像多实例学习的序列不变性存在冲突的问题，本研究提出基于对比学习策略的扫描不变 Mamba 多实例学习 ( $SMC\text{-MIL}$ ) 框架。通过设计双分支实例序列对比学习框架，来约束同一 Mamba 从不同扫描序列获得的包表示及分类结果保持一致，以引导 Mamba 学习与序列顺序无关的特征。同时引入基于实例评估的序列对比增强方法，来增加对比学习难度，迫使 Mamba 在输入序列差异极大的情况下，学习到对病理图像形成一致表示和决策的能力，成功赋予 Mamba 良好的序列不变建模和实例识别能力。实验结果显示， $SMC\text{-MIL}$  在继承了 Mamba 优异计算效率和建模能力的同时，有效降低模型对输入序列顺序的敏感性，专注于实例无序关系建模，挖掘出更具显著性判别特征的实例，更好地适配病理图像分类任务。

**关键词：**数字病理图像分类；Mamba 架构；多实例学习；弱监督学习

## Abstract

As the “gold standard” of disease diagnosis, medical image analysis is an important support for clinical treatment. Among them, digital pathological image classification, as the core field of medical image analysis, is not only the key to the construction of clinical assisted diagnosis technology system, but also an important technical path to achieve precision medicine. In digital pathological image classification tasks, challenges arise from insufficient instance relationship modeling and high computational resource consumption due to the large number of instances. Recent studies demonstrate that the Mamba architecture, characterized by strong long-sequence modeling capabilities and linear computational complexity, is suitable for addressing these issues. However, limitations remain, including inadequate spatial modeling and conflicts with the sequence invariance requirements of multi-instance learning for pathological images. To address these challenges and enable efficient adaptation of the Mamba architecture in pathological image analysis, this study explores two key perspectives: multi-scan Mamba and scan-invariant Mamba.

(1) To overcome the spatial modeling limitations of Mamba’s one-dimensional sequence processing, we propose the Multi-scan Mamba with Spatial Enhancement ( $M^2S-MIL$ ) framework. This framework introduces innovative scanning modes (e.g., regional spiral scanning and grid scanning) as inputs to Mamba, enhancing its ability to capture local continuity, rotation invariance, and spatial correlations across different directions. Additionally, a 2D spatial context sensing module based on block-local attention is designed to reinforce 2D spatial relationship perception. This approach significantly improves Mamba’s spatial modeling capabilities, effectively bridging the gap between Mamba’s architecture and pathological image classification tasks. Extensive experiments validate that integrating multi-scan branches and spatial information alleviates Mamba’s one-dimensional spatial limitations, enhances model performance, and mitigates discrepancies in modeling trends for digital pathological image classification.

(2) To resolve conflicts between Mamba’s sequence-dependent processing and the sequence invariance requirements of multi-instance learning, we develop the Scan-invariant Mamba via Contrast Learning (SMC-MIL) framework. This framework employs a two-branch instance sequence contrast learning mechanism to constrain consistent bag representations and classification results across different scan sequences, guiding Mamba to learn order-agnostic features. A sequence contrast enhancement method based on instance

evaluation is introduced to increase learning difficulty, forcing Mamba to develop consistent representations and decisions even for highly variant input sequences. Experimental results show that SMC-MIL maintains Mamba's computational efficiency while reducing its sensitivity to input sequence order. By focusing on instance disorder relationship modeling and discriminative feature mining, SMC-MIL better aligns with the requirements of pathological image classification tasks. This study systematically addresses the limitations of Mamba in pathological image analysis through architectural innovations, providing a robust foundation for future advancements in medical image-assisted diagnosis.

**Key words:** Whole Slide Image Classification; Mamba Architecture; Multiple Instance Learning; Weakly Supervised Learning

# 目 录

|  |     |
|--|-----|
| <b>摘要</b> .....                          | I   |
| <b>Abstract</b> .....                    | II  |
| <b>图目录</b> .....                         | VI  |
| <b>表目录</b> .....                         | VII |
| <b>1 绪 论</b> .....                       | 1   |
| 1.1 研究背景及意义.....                         | 1   |
| 1.2 国内外研究现状与挑战 .....                     | 3   |
| 1.2.1 数字病理图像分类研究现状 .....                 | 3   |
| 1.2.2 视觉 Mamba 研究现状.....                 | 4   |
| 1.3 本文研究内容与创新点 .....                     | 6   |
| 1.4 本文组织结构 .....                         | 8   |
| <b>2 相关研究技术与理论</b> .....                 | 9   |
| 2.1 数字病理图像多实例学习.....                     | 9   |
| 2.2 Mamba 架构 .....                       | 10  |
| 2.3 对比学习 .....                           | 11  |
| 2.3.1 基于负样本的对比学习 .....                   | 12  |
| 2.3.2 无负样本的对比学习.....                     | 12  |
| 2.4 数据集及评价指标 .....                       | 12  |
| 2.4.1 数据集.....                           | 13  |
| 2.4.2 评价指标 .....                         | 15  |
| 2.5 病理图像的预处理过程 .....                     | 15  |
| 2.6 本章小结 .....                           | 16  |
| <b>3 基于多扫描 Mamba 的高分辨率病理图像分类研究</b> ..... | 17  |
| 3.1 研究动机 .....                           | 17  |
| 3.2 问题描述和符号定义.....                       | 18  |
| 3.3 基于空间信息增强的多扫描 Mamba 模型.....           | 19  |
| 3.3.1 方法概述 .....                         | 19  |
| 3.3.2 预处理与特征提取 .....                     | 20  |
| 3.3.3 多扫描 Mamba 模块 .....                 | 20  |
| 3.3.4 2D 空间上下文感知模块 .....                 | 22  |

|   |           |
|---|-----------|
| 3.3.5 基于 Mamba 的实例聚合 .....                | 22        |
| 3.3.6 模型优化和推理.....                        | 23        |
| <b>3.4 实验设置及结果分析.....</b>                 | <b>23</b> |
| 3.4.1 实验设置 .....                          | 23        |
| 3.4.2 基准数据集实验结果.....                      | 25        |
| 3.4.3 消融实验 .....                          | 29        |
| 3.4.4 可视化分析.....                          | 33        |
| <b>3.5 本章小结 .....</b>                     | <b>34</b> |
| <b>4 基于扫描不变性 Mamba 的高分辨率病理图像分类研究.....</b> | <b>36</b> |
| <b>4.1 研究动机 .....</b>                     | <b>36</b> |
| <b>4.2 基于序列差异化对比的 Mamba 模型.....</b>       | <b>37</b> |
| 4.2.1 方法概述 .....                          | 37        |
| 4.2.2 预处理与特征提取 .....                      | 37        |
| 4.2.3 基于 Mamba 的双分支对比架构.....              | 38        |
| 4.2.4 差异化序列生成（简单与困难分支） .....              | 39        |
| 4.2.5 基于 Mamba 的实例聚合 .....                | 40        |
| 4.2.6 基于一致性的模型优化 .....                    | 41        |
| 4.2.7 模型推理 .....                          | 41        |
| <b>4.3 实验设置及结果分析.....</b>                 | <b>41</b> |
| 4.3.1 实验设置 .....                          | 42        |
| 4.3.2 基准数据集实验结果.....                      | 43        |
| 4.3.3 消融实验 .....                          | 48        |
| 4.3.4 可视化分析.....                          | 55        |
| <b>4.4 本章小结 .....</b>                     | <b>55</b> |
| <b>5 总结与未来展望 .....</b>                    | <b>58</b> |
| <b>5.1 总结 .....</b>                       | <b>58</b> |
| <b>5.2 未来展望 .....</b>                     | <b>59</b> |
| <b>参考文献 .....</b>                         | <b>61</b> |
| <b>附 录 .....</b>                          | <b>66</b> |
| A. 作者在攻读硕士学位期间的论文和专利目录.....               | 66        |
| B. 作者在攻读硕士学位期间参与的科研项目 .....               | 66        |
| C. 学位论文数据集.....                           | 67        |
| <b>致 谢 .....</b>                          | <b>68</b> |

## 图目录

|   |    |
|---|----|
| 图 1.1 Camelyon16 数据集中肿瘤区域相对于正常区域的比例示例图 .....                            | 1  |
| 图 1.2 两种多实例学习范式示意图 .....  | 2  |
| 图 1.3 Mamba 模型与 MIL 任务差异图 .....   | 5  |
| 图 1.4 本文的主要研究内容图 .....  | 7  |
| 图 2.1 Mamba 架构示意图 .....   | 11 |
| 图 2.2 对比学习结构示意图 .....   | 12 |
| 图 2.3 数字病理图像的预处理过程示意图 .....   | 16 |
| 图 3.1 常规扫描、网格扫描和区域螺旋扫描示意图 .....   | 17 |
| 图 3.2 基于多扫描 Mamba 的多实例学习架构示意图 .....                                     | 19 |
| 图 3.3 不同 Mamba 模型的性能比对柱状图 .....   | 30 |
| 图 3.4 不同区域数量的模型性能折线图 .....  | 31 |
| 图 3.5 原始 Mamba(baseline) 和 M <sup>2</sup> S-MILL 聚合特征 t-SNE 可视化结果 ..... | 33 |
| 图 3.6 多个 Mamba 变体和 M <sup>2</sup> S-MIL 关注的图块可视化结果 .....                | 34 |
| 图 4.1 SMC-MIL 模型总览图 .....   | 38 |
| 图 4.2 SMC-MIL 模型推理流程示意图 .....   | 42 |
| 图 4.3 不同 Mamba 变体的扫描不变性评估的双样本 t 检验 .....                                | 50 |
| 图 4.4 不同 Mamba 模型性能比对柱状图 .....  | 51 |
| 图 4.5 不同序列增强率 $\alpha\%$ 下的模型性能折线图 .....                                | 52 |
| 图 4.6 不同序列掩码率 $\beta\%$ 下的模型性能折线图 .....                                 | 53 |
| 图 4.7 与 SOTA 方法性能比较的 Bootstrap 检验 .....                                 | 53 |
| 图 4.8 原始 Mamba(baseline) 和 SMC-MIL 关注的图块可视化结果 .....                     | 56 |
| 图 4.9 原始 Mamba(baseline) 和 SMC-MIL 聚合特征 t-SNE 可视化结果 .....               | 57 |

## 表目录

|  |    |
|--|----|
| 表 2.1 癌症诊断、分型病理图像数据集详情 .....   | 13 |
| 表 2.2 生存预测病理图像数据集详情 .....  | 13 |
| 表 3.1 基于空间信息增强的多扫描 Mamba 模型的实验环境及配置 .....                              | 24 |
| 表 3.2 M <sup>2</sup> S-MIL 在 CAMELYON 数据集的癌症诊断性能比较 (%) .....           | 25 |
| 表 3.3 M <sup>2</sup> S-MIL 在 TCGA-NSCLC 数据集的亚型分类性能比较 (%) .....         | 26 |
| 表 3.4 M <sup>2</sup> S-MIL 在 TCGA-BRCA 数据集的亚型分类性能比较 (%) .....          | 26 |
| 表 3.5 M <sup>2</sup> S-MIL 在 BRACS 数据集官方划分下的粗粒度分类性能对比 (%) .....        | 27 |
| 表 3.6 以 UNI 为离线特征提取器的癌症诊断和分型性能 .....                                   | 28 |
| 表 3.7 M <sup>2</sup> S-MIL 在以 R50 为特征提取器的三个主要数据集上的生存预测结果 (%) .....     | 28 |
| 表 3.8 M <sup>2</sup> S-MIL 在以 UNI 为特征提取器的三个主要数据集上的生存预测结果 (%) .....     | 29 |
| 表 3.9 在三个主要数据集上，不同组件对性能的影响 .....                                       | 30 |
| 表 3.10 不同大小窗口对性能影响 .....   | 31 |
| 表 3.11 M <sup>2</sup> S-MIL 以 PLIP 为离线特征提取器的亚型分型性能 .....               | 32 |
| 表 3.12 M <sup>2</sup> S-MIL 以 PLIP 为离线特征提取器在 TCGA-NSCLC 上的亚型分型性能 ..... | 32 |
| 表 3.13 M <sup>2</sup> S-MIL 以 PLIP 为离线特征提取器在 TCGA-BRCA 上的亚型分型性能 .....  | 33 |
| 表 4.1 基于扫描不变性 Mamba 的高分辨率病理图像分类研究的实验环境及配置 .....                        | 42 |
| 表 4.2 SMC-MIL 在 CAMELYON 数据集的癌症诊断性能比较 (%) .....                        | 43 |
| 表 4.3 SMC-MIL 在 TCGA-NSCLC 数据集的亚型分类性能比较 (%) .....                      | 44 |
| 表 4.4 SMC-MIL 在 TCGA-BRCA 数据集的亚型分类性能比较 (%) .....                       | 44 |
| 表 4.5 SMC-MIL 在 BRACS 数据集官方划分下的粗粒度分类性能对比 (%) .....                     | 45 |
| 表 4.6 SMC-MIL 在 BRACS 数据集官方划分下的细粒度分类性能对比 (%) .....                     | 45 |
| 表 4.7 SMC-MIL 在 BRACS 数据集五折交叉验证下的细粒度分类性能对比 (%) .....                   | 46 |
| 表 4.8 SMC-MIL 在 R50 提取特征的三个主要数据集上的生存预测结果 (%) .....                     | 47 |
| 表 4.9 SMC-MIL 在 UNI 提取特征的三个主要数据集上的生存预测结果 (%) .....                     | 48 |
| 表 4.10 在三个主要数据集上，不同组件对性能的影响 .....                                      | 49 |
| 表 4.11 不同 Mamba 变体的扫描不变性评估对比 .....                                     | 50 |
| 表 4.12 不同的分支对比策略对模型性能的影响 .....   | 51 |
| 表 4.13 讨论不同扫描模式的组合对模型性能的影响 .....                                       | 52 |
| 表 4.14 不同推理方式的性能对比 .....   | 54 |
| 表 4.15 SMC-MIL 以 PLIP 为离线特征提取器的癌症诊断性能 .....                            | 54 |
| 表 4.16 SMC-MIL 以 PLIP 为离线特征提取器在 TCGA-NSCLC 上的亚型分型性能 .....              | 54 |

表 4.17 SMC-MIL 以 PLIP 为离线特征提取器在 TCGA-BRCA 上的亚型分型性能 ..... 55

# 1 绪 论

本章内容共分为四节，第一节介绍本文的研究背景及意义；第二节总结医学病理图像领域及视觉 Mamba 的国内外研究现状，并对其面临的挑战进行分析；第三节介绍本文的研究内容与创新点；第四节对本文组织结构进行概括。

## 1.1 研究背景及意义

随着全球人口老龄化进程的不断加速，以及疾病发生率呈现显著增长趋势，临床医学领域对精准高效的疾病诊断技术需求日益凸显。依据《“健康中国 2030”规划纲要》<sup>[1]</sup>，我国人口老龄化进程加快，慢性疾病负担日益沉重，这无疑对医疗诊断的准确性与效率提出了更高要求。病理诊断作为疾病的“金标准”，在临床决策中占据关键地位<sup>[2-4]</sup>。而传统依靠病理医生在显微镜下人工阅片的方式存在明显短板。如在病理诊断实践中，细胞与组织形态学的显著异质性特征以及肿瘤表型的微小异质性差异，共同构成了该领域的核心挑战。这种复杂性要求病理医师不仅需要具备深厚的专业知识储备，更依赖于长期临床实践经验的支撑。然而，病理医师培养周期长且成本高昂的现状，与临床需求之间的矛盾日益凸显。同时，视觉观察的主观性叠加切片染色过程中产生的颜色偏差与形态差异，可能导致诊断结果的不一致性，甚至引发误判。此外，高分辨率病理图像通常包含亿级像素信息，而病灶区域仅占图像面积的极小比例（如图 1.1）病理医师需在有限时间内完成大量图像的精读分析，显著降低了人工诊断的效率与准确性。

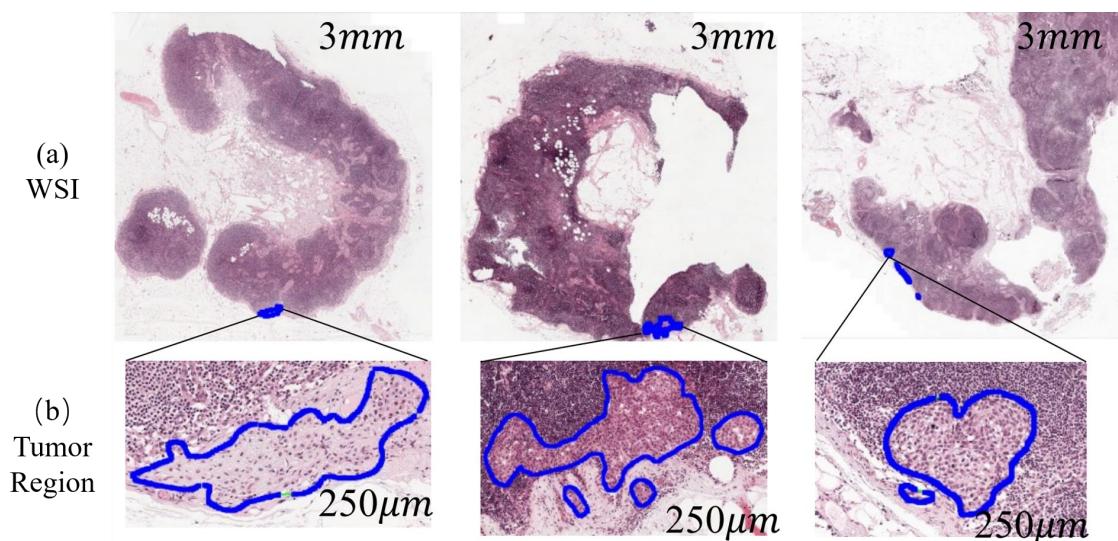


图 1.1 Camelyon16 数据集中肿瘤区域相对于正常区域的比例示例图。

Fig. 1.1 Example of the proportion of the tumor region to the normal region in the Camelyon16 dataset.

与此同时，随着人工智能在各领域的蓬勃发展，AI 赋能医疗给病理智能诊断提供了可能。国务院印发的《新一代人工智能发展规划》<sup>[5]</sup>着重指出，要加速人工智能在医疗领域的应用，全力推动医疗影像辅助判读等技术的进步。利用高分辨率数字病理扫描仪通过高效精准的数字化流程，可以将组织切片转换为亿级像素的全幅切片图像（Whole Slide Image，下文简称病理图像或 WSI）<sup>[6]</sup>，而这些图像便于存储、传输与共享，打破了时空限制。加之计算机视觉技术的突飞猛进，图像处理、人工智能及机器学习算法持续创新，让病理图像自动化分析从设想走向现实。通过对海量数字病理图像的深度学习，计算机模型有能力捕捉到人类肉眼难以察觉的病变特征，以提高病理诊断的效率与准确性。

然而，针对如此高分辨的 WSI 图像而言，传统深度学习算法仍有不小挑战：数十亿像素的图片如直接输入到传统的分类器中，将导致难以承担的计算成本；而若采用采样出的低分辨率的缩略图作为输入，则会不可避免的丢失大量关键信息，造成性能上的损失。同时，由于病理图像的标注需要大量专家知识，因此细粒度的图像标注过程既耗时又耗力。当前，在仅提供低精度标注的场景下，如何充分挖掘未标注数据的潜力以优化模型效能，已成为数字病理图像分析领域的核心研究方向之一。

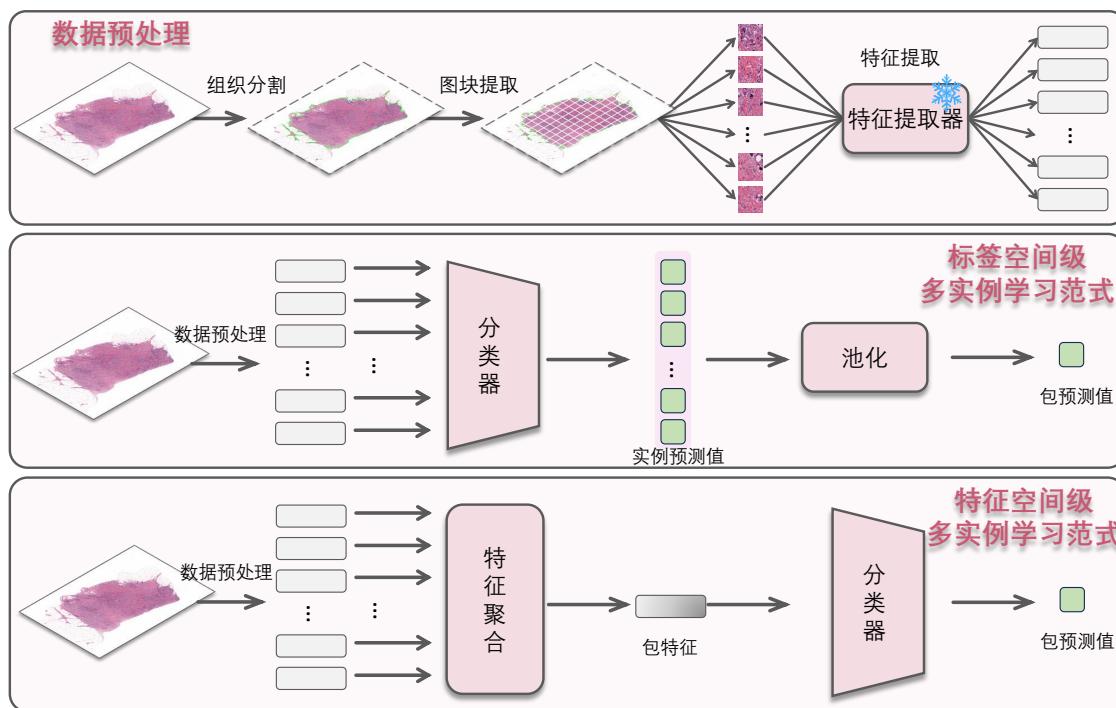


图 1.2 两种多实例学习范式示意图。

Fig. 1.2 The illustration of two paradigms for multiple instance learning.

多实例学习 (Multiple Instance Learning, MIL)<sup>[7,8]</sup> 作为一种特殊的弱监督学习

框架是业内解决这一类问题的常见技术手段。具体而言，每张病理图像都被视为一个“包”（Bag），并会被进一步分割成大量图块（Patch），这些图块被定义为“包”中的“实例”（Instance）。其目标是在分类标签（如“恶性”或“良性”）作用于整个包的情况下，通过分析实例与包标签之间的关联性，挖掘出关键实例特征，并基于这些信息完成对整张图像的分类。然而，在病理图像的实际应用场景中，多实例学习框架虽然得到了广泛应用<sup>[9-13]</sup>，但仍然面临着诸多现实问题。由于尺寸巨大，切分后的实例数目众多，导致在使用具有较强建模能力的传统 ViT<sup>[14]</sup>进行实例处理时所需的计算资源消耗巨大，只能采取减少实例数目或更换为线性注意力机制的折中方案。

而近期，一些基于状态空间模型（State Space Model, SSM）<sup>[15]</sup>的架构为解决上述问题，提供了新的思路。Mamba<sup>[16,17]</sup>是最近提出的一种具有代表性的基于 SSM 架构的模型，与 ViT 相比具有更好的长序列建模能力，而且仅需线性的计算资源。自从其引入到视觉领域后，在分类<sup>[18-21]</sup>、检测<sup>[22-25]</sup>、分割<sup>[26-29]</sup>等任务中已经取得了显著的表现，展现出了宏大的应用前景。然而作为原本被提出解决序列问题的 Mamba 总是倾向于建模出一种前序 Token 与后序 Token 的因果关系，这与病理图像多实例学习的序列不变特性并不相符<sup>[30]</sup>。并且，其以一维序列作为输入，常常无法完全捕获图像中的二维空间信息。鉴于此，针对高分辨率数字病理图像分类任务，通过构建基于 Mamba 的模型框架弥合序列相关的 Mamba 模型与序列无关的 WSI 分类任务之间的性能差异，深度挖掘病理图像的空间特征信息，进而优化模型的分类效能，推动数字病理图像计算机辅助诊断准确率的显著提升，具有重要的临床应用价值与学术研究意义。

## 1.2 国内外研究现状与挑战

### 1.2.1 数字病理图像分类研究现状

鉴于数字病理图像分辨率极高，同时细粒度像素级标注较为匮乏，当前研究工作普遍借助多实例学习框架攻克数字病理图像的分类难题。多实例学习，英文表述为 Multiple Instance Learning，简称为 MIL，部分文献也将其译为多示例学习。多实例学习的经典范式最早由 Dietterich TG 等学者于 1997 年在药物活性预测研究中首次提出并成功应用<sup>[7]</sup>。作为弱监督学习领域的典型框架，该技术已受到学术界的广泛关注。其训练数据通常由带有类别标签的包集合构成，每个包内包含多个未标注实例，构成独特的弱监督学习模式。正包中至少含有一个正实例，负包中只包含负实例。在多实例学习的训练过程中，只需要提供粗粒度的包级标签（如病人是否患癌），而在实际场景中粗粒度的包级标签往往是易获取的，从而使得多实例学习非常契合多种实际场景，很多问题可以自然的表述为多实例问题。以药物活性预测问题为例<sup>[7]</sup>，在该问题里，主要的任务目标在于对某一种新分子是否适宜用于制药进行预测。实际情况是，每一个分子都存在着多种可能的低能形状。所获得

的标签信息仅仅是明确了哪些分子是适合制药的，然而对于究竟是分子的哪一种具体低能形状在其是否适合制药这一判定中起到了关键决定性作用，却并不清楚。鉴于这样的情况，在处理该问题时，便将每一个分子当作是一个“包”，而分子的每一种低能形状则对应成为这个“包”中的一个“实例”，如此一来，就形成了多实例学习中所定义的实例与包的关系模式。由于其独有的特性，多实例学习在各个领域都得到了广泛的发展和应用，如目标检测<sup>[31-34]</sup>，语义分割<sup>[35]</sup>，图像和视频分类<sup>[36-41]</sup>，文本分类<sup>[42]</sup>，医学图像分类<sup>[9,11,12,43-45]</sup>等。

当前，多实例学习方法主要分为标签空间方法与特征空间方法。标签空间方法基于实例级标签推断聚合策略直接预测包标签，其实例间关系建模能力有限，且在大规模数据集中面临计算效率瓶颈<sup>[46]</sup>。鉴于此，当前研究重点逐渐转向特征空间方法，这类方法通过整合包内所有实例的特征信息，构建包的嵌入特征表示，进而实现包级标签的预测。例如，Ilse 等人<sup>[43]</sup>提出的 AB-MIL，在多实例学习框架中引入基于注意力的聚合层，通过神经网络以参数化方式量化每个实例对于包级特征嵌入的贡献。Li 等学者<sup>[9]</sup>在研究中提出并构建了 DSMIL 方法，通过融合自监督对比学习与金字塔特征融合机制高效提取实例级特征，并创新性地引入非局部注意力模块聚合多尺度实例信息，从而显著提高 WSI 全切片图像分类任务的准确率。DTFD-MIL<sup>[12]</sup>引入双层多实例学习框架，将图像分割成多个伪包，通过聚合伪包的预测结果，推断整个病理图像的标签。Zhu 等人<sup>[47]</sup>提出 DGR-MIL 模型，通过引入可学习的全局向量，精准识别最具代表性的实例，并采用交叉注意力机制计算实例间的重要性，有效挖掘病理图像中的关键特征。Shao 等人<sup>[11]</sup>将 Transformer 架构引入多实例学习框架中，利用空间位置编码保留空间拓扑结构，并通过注意力机制实现实例间关系的多粒度建模。

Transformer 中自注意力机制计算成本高昂，限制了其在大规模图像数据处理任务中的推广与应用。因此，Tang 等人<sup>[13]</sup>通过将病理图像划分为多个区域，显著减少 Token 数量，并引入线性自注意力机制，降低计算成本的同时有效建模实例间的关系。然而，这种轻量型的自注意力机制会在一定程度上削弱模型对复杂实例关系的建模能力。最近提出的 Mamba 模型<sup>[16]</sup>因其高效的计算效率和强大的实例关系建模能力成为了热门模型，但作为与输入顺序相关的序列模型，Mamba 对实例扫描模式非常敏感，这与多实例学习对序列不变性的要求相冲突，影响了其在病理图像分类任务中的稳定性，限制了其在病理图像分类中的潜力。

### 1.2.2 视觉 Mamba 研究现状

Mamba 最初是在自然语言处理领域提出的，作为一种更有效地处理长序列的架构<sup>[16]</sup>。与 Transformer<sup>[48]</sup>等传统模型不同，Mamba 采用了一种选择性状态空间模型（SSM<sup>[49]</sup>），该模型基于内容动态地过滤和处理信息，使模型能够选择性地记住或忽略部分输入。Mamba 在处理速度和可扩展性方面提供了显著的改进，特别是对于较长的序列。Zhu 等人首先将 Mamba 引入视觉领域研究人们的视野<sup>[18]</sup>，

其在具有强大的建模能力的同时仅需线性计算资源的特点，让其迅速在分类、检测、分割等诸多领域取得令人惊奇的效果。

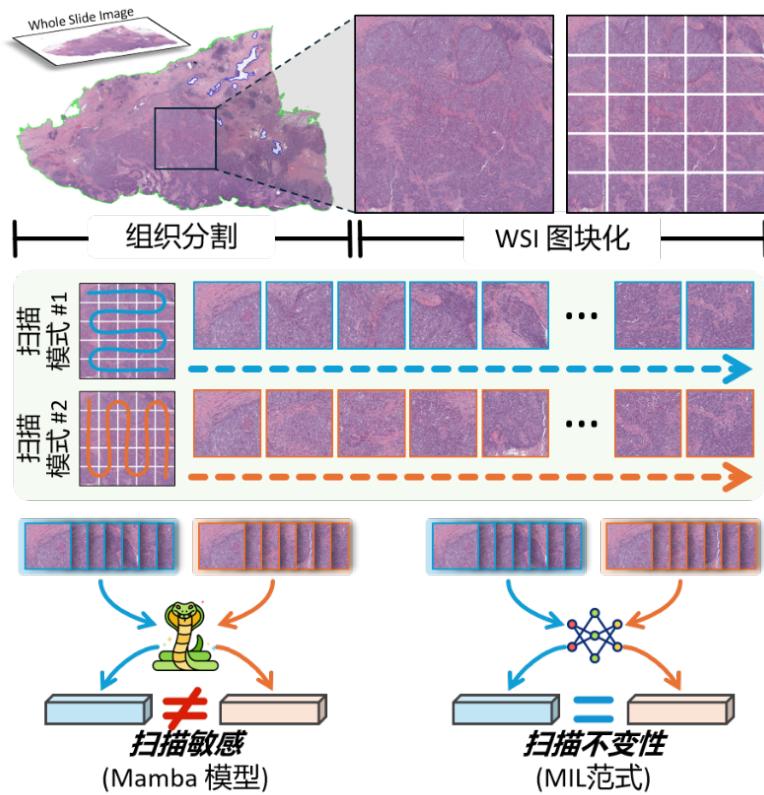


图 1.3 Mamba 模型与 MIL 任务差异示意图。

Fig. 1.3 Discrepancy between Mamba and MIL for computational pathology tasks.

然而，Mamba 作为一种自回归模型，在视觉任务与多实例学习任务中固有地表现出差异，如图1.3所示：Mamba 假设之前和之后的斑块之间存在因果关系<sup>[50]</sup>，而在图像中，不同斑块之间不存在这种因果关系，甚至在多实例学习范式中更是将各个实例视为“平等”实例，并无顺序关联。因此，许多研究通过聚合多种扫描模式来减轻 Mamba 对序列顺序的敏感性。例如，Vim<sup>[18]</sup> 通过组合来自两个分支（向前和向后）的结果来减轻序列顺序的影响。而 Vmamba<sup>[50]</sup> 采用了四种扫描模式：左上到右下、右下到左上、右上到左下、左下到右上。这些模式综合多个信息序列以提高性能。LocalMamba<sup>[30]</sup> 引入了一种新的局部扫描策略，将图像划分为不同的窗口，综合横向、纵向、局部等多种扫描模式，在保持全局视角的同时有效地捕获局部依赖，大大提高捕获有效图像表示方面的性能。

对于病理图像分类而言，MamMIL<sup>[51]</sup> 结合了 Bi-SSM 和 2D-CAB 模块，整合了来自原始、反向和局部序列的扫描模式信息，从而增强了 WSI 分类。此外，MambaMIL<sup>[52]</sup> 将序列重塑成一个矩形，并应用序列重排序操作来垂直扫描新序列，提高了其模拟长序列的能力。然而，这些方法过度关注输出序列信息的整合，未能

从根本上突破序列输入顺序对 Mamba 建模能力的约束，仅仅是一种妥协的策略，其依旧高度依赖扫描模式和集成形式，使得模型性能不够稳定性。

### 1.3 本文研究内容与创新点

本文研究了 Mamba 架构在病理图像分类中的应用。虽然 Mamba 架构在处理病理图像分类任务时，可以解决由于实例数目过多导致的问题，例如计算资源消耗巨大、实例间关系建模不充分等，但其仍然存在其无法捕获二维空间关系、以及与病理图像多实例学习的序列不变性存在冲突等问题，与病理图像分类任务并不完全适配。如何缓解甚至消除二者其中的差异依旧是将高效且性能卓越的 Mamba 架构适配到病理图像领域一个重要的且有挑战性的问题。因此，本文研究了 Mamba 架构在病理图像分类中的应用，分别从多扫描模式融合 Mamba 和扫描不变性 Mamba 的两个角度开展以下研究：

#### (1) 基于多扫描 Mamba 的高分辨率病理图像分类研究

针对 Mamba 一维序列建模所造成图像空间结构捕获能力不足的问题，本文从 Mamba 整合多种扫描模式的角度出发，提出基于空间信息增强实现的多扫描 Mamba 的多实例学习框架（**M**ulti-scan **M**amba with **S**patial information **e**nhan**c**ement，简称 **M**<sup>2</sup>**S-MIL**）。由于常用扫描模式未关注到图像的局部连续性、旋转不变性等特性，该框架采用了多个扫描模式分支作为 Mamba 的输入，通过设计了区域螺旋扫描和网格扫描等新的扫描模式，提高 Mamba 对局部连续性、旋转不变性以及同一序列不同方向的空间关联信息的捕获能力。同时为进一步提升序列的 2D 空间上下文的感知能力，设计了基于分块局部注意力的 2D 空间上下文感知模块，利用区域内的信息整合更进一步提高 Mamba 对空间能力的捕捉，使 Mamba 模型更适配于病理图像分类任务。在 3 个 WSI 分析子任务和 7 个数据集上的大量实验表明，综合多种扫描模式分支，引入空间信息，能够切实解决 Mamba 从一维实例序列中空间建模不充分的问题，显著提高模型性能，同时在数字病理图像分类任务中一定程度缓解了使用 Mamba 模型出现的建模倾向差异难题。

#### (2) 基于扫描不变性 Mamba 的高分辨率病理图像分类研究

本文针对 Mamba 作为序列相关模型，与病理图像多实例学习的序列不变性存在冲突的问题，从构建扫描不变性 Mamba 角度出发，提出了基于对比学习策略实现的扫描不变 Mamba 的多实例学习（**S**can-invariant **M**amba with **C**ontrastive **L**earning，简称 **SMC-MIL**）框架。为了指导 Mamba 学习到不同实例扫描序列中一致的包特征，本文设计了双分支实例序列对比学习框架，通过约束同一 Mamba 从不同扫描序列获得的包表示及分类结果保持一致，来引导 Mamba 学习与序列顺序无关的特征。同时引入基于实例评估的序列对比增强方法，来增加对比学习难度，迫使 Mamba 在输入序列差异极大的情况下，学习到对病理图像形成一致表示和决策的能力，成功赋予 Mamba 良好的序列不变建模和实例识别能力。实验结果显示，

SMC-MIL 在继承了 Mamba 优异计算效率和建模能力的同时，有效降低模型对输入序列顺序的敏感性，专注于实例无序关系建模，挖掘出更具显著性判别特征的实例，更好地适配病理图像分类任务。

本文提出的 M<sup>2</sup>S-MIL 模型与 SMC-MIL 模型分别从多扫描模式 Mamba 和扫描不变性 Mamba 的两个角度出发，对数字病理图像分类任务中 Mamba 的应用进行深入的剖析和研究。M<sup>2</sup>S-MIL 模型通过综合多种扫描模式的特征信息，并引入空间上下文感知模块，进一步提升 Mamba 对 2D 空间信息的捕获能力，缓解原本一维 Mamba 模型对 2D 病理图像空间建模能力不足的问题；SMC-MIL 模型则进一步突破扫描顺序对 Mamba 建模能力的约束，弥合了原本 Mamba 的序列敏感性与病理图像多实例学习中序列不变性假设的冲突，使 Mamba 架构更加适配病理图像分类任务，并提高了模型挖掘更具判别性区域的能力，从而提升模型性能。

图 1.4 展示了本文的主要研究内容和贡献。

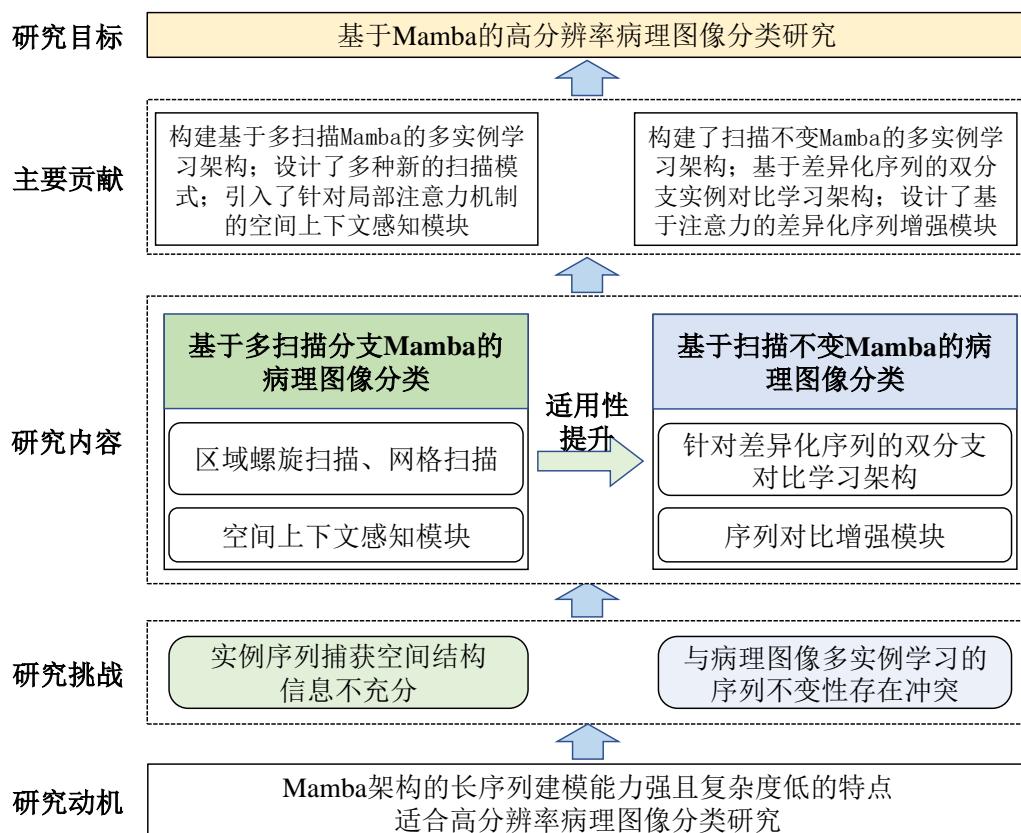


图 1.4 本文的主要研究内容。

Fig. 1.4 The main research content of this paper.

## 1.4 本文的组织结构

本文由 5 个章节构成，从多个维度深入探讨基于 Mamba 的高分辨率数字病理图像分类研究。下面为各章节的具体介绍：

**第一章：绪论。**本章首先阐述高分辨率数字病理图像分类任务的研究背景与科学意义，在系统梳理国内外前沿研究进展的基础上，重点剖析基于视觉 Mamba 架构的数字病理图像分类技术现状及存在的挑战。最后对本文的研究内容和组织结构进行概述。

**第二章：相关研究技术与理论。**首先系统阐述本课题涉及的基础理论框架，涵盖病理图像多实例学习范式、Mamba 模型架构原理及对比学习机制等核心内容。在此基础上，详细解析所用实验数据集、性能评估指标体系及数字病理图像预处理技术路线。

**第三章：基于多扫描融合 Mamba 的高分辨率医学病理图像分类研究。**首先，通过深入剖析现有 Mamba 病理图像分类算法在空间结构建模能力方面的局限性，提出了基于多扫描 Mamba 的高分辨率病理图像分类算法 ( $M^2S\text{-MIL}$ )，最后通过在 3 个 WSI 任务和 7 个数据集上的大量实验证明了  $M^2S\text{-MIL}$  模型的有效性。

**第四章：基于扫描不变性 Mamba 的高分辨率医学病理图像分类研究。**首先对基于多扫描序列融合 Mamba 的病理图像分类算法的不足进行分析介绍，提出了基于扫描不变性 Mamba 的病理图像分类算法 (SMC-MIL)，然后介绍了提出的 SMC-MIL 模型的模型框架以及其中的双分支实例序列对比学习框架和序列对比增强模块，并通过系统的实验证明 SMC-MIL 模型充分赋予了 Mamba 扫描不变属性。

**第五章：总结与未来展望。**总结并分析了本文提出的基于 Mamba 的高分辨率病理图像分类算法研究的成果及不足，并对未来的研究方向与内容进行了展望。

## 2 相关研究技术与理论

本章内容共分为六节，第一节详细介绍病理图像分类任务以及主流的多实例学习架构；第二节介绍 Mamba 架构；第三节介绍第四章方法应用到的对比学习方法及相关理论；第四节对本文所使用的数据集和评价指标进行介绍；第五节对病理图像的预处理过程进行详细介绍，第六节对本章内容进行总结。

### 2.1 数字病理图像多实例学习

数字病理图像分类属于前沿且极具潜力的交叉研究领域，它将病理学、图像分析和计算机科学等多学科知识进行深度整合。此领域借助先进算法、机器学习模型和人工智能技术，以实现对全切片图像（Whole Slide Image，简称 WSI 或病理图像）的全面、精确分析解读，这也是它的核心目标。病理图像分类任务通常可分为癌症诊断、亚型分类、生存预测等任务，分别从这些任务角度中辅助病理学家诊断疾病。数字病理领域，超高分辨率全切片图像（WSI）要完成像素级标注，需投入大量时间与人力。这一难题致使基于像素级标签开展的传统深度学习病理图像分类，难以施展。在此背景下，凭借独特优势，基于多实例学习范式的数字病理图像分类，逐渐成为该领域的常用方法。

多实例学习（Multiple Instance Learning, MIL）<sup>[7]</sup>是弱监督学习的重要分支，其核心思想是处理数据标注在“包”（Bag）级别而非单个实例级别的场景。具体而言，每个数据包包含多个实例，但仅包级别标签已知。经典假设为：若一个包中至少存在一个正实例（Positive Instance），则该包被标记为正类；若所有实例均为负类，则包标记为负类。此范式由 Dietterich 等人在 1997 年药物分子活性预测研究中首次提出，并逐渐拓展至计算机视觉、自然语言处理及生物医学等领域。

具体来说，对于在病理图像分类任务而言，一张 WSI 图像  $X$  将作为一个包，将之切分为多个小图块，将每个图块 Patch 作为一个实例  $x_i$ ，进而得到实例集合  $X = \{x_i\}_{i=1}^N$ 。需注意，不同包所包含的实例数目  $N$  并不固定。在任何分类任务里，包  $X$  都附带已知的标签信息  $Y \in \{0, 1\}$ ，而实例标签  $y_i$  的具体取值  $\{0, 1\}$  处于未知状态。以二分类任务为例，包的预测标签  $Y$  满足公式：

$$Y = \begin{cases} 0 & \text{if } \sum_i^N y_i = 0, \\ 1 & \text{otherwise.} \end{cases} \quad (2.1)$$

而多实例学习模型  $\mathcal{M}(\cdot)$  的目标是利用所有实例预测包的标签  $\hat{Y} \leftarrow \mathcal{M}(X)$ 。

早期的方法倾向于预测出每个实例的标签后，再以此推断出包标签。这种方法称之为标签空间级的方法。这类方法非常直观，充分利用了多实例学习的定义，但其实例间关系建模能力有限，且在大规模数据集中面临计算效率瓶颈<sup>[46]</sup>。鉴于

此，当前研究重点逐渐转向特征空间方法，这类方法通过整合包内所有实例的特征信息，构建包的嵌入特征表示，进而实现包级标签的预测。大量的研究证明基于特征空间的方法表现出比标签空间级方法更强的性能。

自 Ilse 等人<sup>[43]</sup>提出 AB-MIL 后，基于注意力的特征空间方法已成为业内主流。该方法通过构建基于上下文感知的注意力机制，动态计算每个实例在包特征嵌入空间中的注意力权重，并据此生成最终的包级特征表示。例如，DTFD-MIL<sup>[12]</sup>引入双层多实例学习框架，将图像分割成多个伪包，同样使用注意力机制聚合伪包的预测结果，推断整个病理图像的标签。Zhu 等人<sup>[47]</sup>提出 DGR-MIL 模型，通过引入可学习的全局向量，精准识别最具代表性的实例，并采用交叉注意力机制计算实例间的重要性，有效挖掘病理图像中的关键特征。Shao 等人<sup>[11]</sup>将 Transformer 架构引入多实例学习框架中，利用空间位置编码保留空间拓扑结构，并通过注意力机制实现实例间关系的多粒度建模。然而 Transformer 中自注意力机制计算成本高昂，诸多工作只能采取缩减 token 数量或采用线性的注意力机制等折中方式将这种复杂模型应用于病理图像分类任务中<sup>[13]</sup>。

## 2.2 Mamba 架构

如图2.1，Mamba 架构是由 Albert Gu 和 Tri Dao 等人于 2023 年提出的新型序列建模框架<sup>[16]</sup>，其核心创新在于将状态空间模型（State Space Model, SSM）与选择性机制（Selective Mechanism）相结合，突破了传统 Transformer 模型在长序列建模中的局限性。该架构通过建立线性时间复杂度与序列长度的关系，显著提升了模型处理长序列依赖的能力，在自然语言处理、基因组学分析、计算机视觉等领域展现出显著优势。

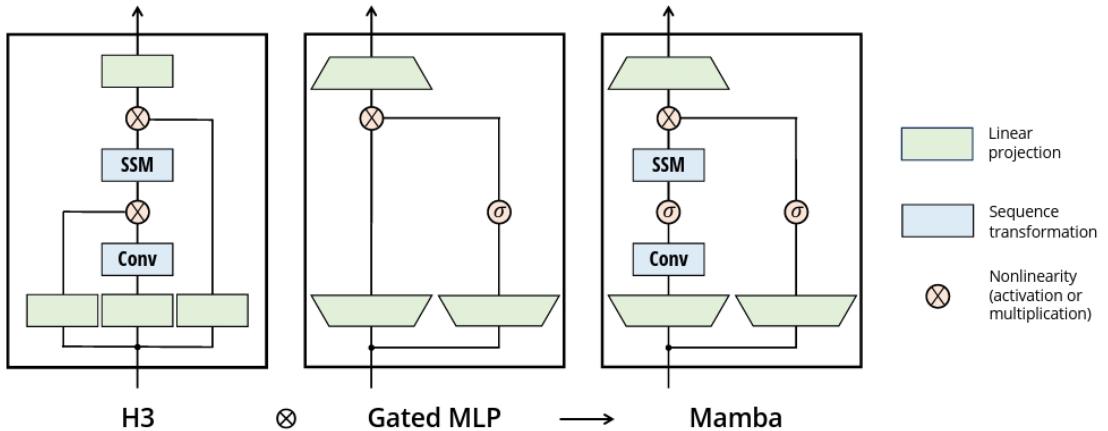
从数学基础来看，Mamba 架构以连续时间状态空间模型为理论框架。其核心微分方程可表述为：

$$\begin{aligned} h'(t) &= \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) &= \mathbf{C}h(t) + \mathbf{D}x(t), \end{aligned} \tag{2.2}$$

其中  $\mathbf{A}$  为状态转移矩阵， $\mathbf{B}$ 、 $\mathbf{C}$  分别为输入/输出投影矩阵， $\mathbf{D}$  为跳跃连接项。通过零阶保持法（Zero-order Hold）的离散化处理，如公式 2.3 所示：

$$\begin{aligned} \bar{\mathbf{A}} &= \exp(\Delta \mathbf{A}), \\ \bar{\mathbf{B}} &= (\Delta \mathbf{A})^{-1}(\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}. \end{aligned} \tag{2.3}$$

模型将连续系统转化为适用于离散序列的递归计算形式，这一过程在保持系统动态特性的同时实现了计算可行性。与传统递归神经网络不同，Mamba 引入了输入依赖的参数动态化机制，通过函数  $\mathbf{B}, \mathbf{C}, \Delta = f_{\theta}(x_t)$  实现参数矩阵的实时生成，使模型能够根据当前输入特征自适应调整状态转移模式。

图 2.1 Mamba 架构示意图。<sup>[16]</sup>Fig. 2.1 Illustration of the Mamba architecture.<sup>[16]</sup>

在架构设计层面，Mamba 通过三个关键创新实现突破：首先，提出选择性扫描算法（Selective Scan Algorithm），通过时间步长参数  $\Delta$  的动态调节，实现高频信号密集采样与低频信号稀疏处理的智能平衡；其次，开发硬件感知并行化技术，利用 GPU 内存层级特性优化扫描操作的计算模式，在保持线性时间复杂度的同时实现并行加速；最后，采用模块化堆叠结构，通过交替堆叠 SSM 块与门控前馈网络构建深度模型，配合残差连接和层归一化技术确保梯度有效传播。实验数据显示，在 PG19 长文本建模任务中，该架构的困惑度相比 Transformer 降低 23%，训练速度提升 5.2 倍<sup>[16]</sup>。

相较于传统 Transformer 的  $O(N^2)$  复杂度，Mamba 在序列长度为 N 时实现  $O(N)$  的时间与空间复杂度，这一特性使其在万 Token 级别的高分辨数字病理图像等场景展现独特优势。但其本质作为序列敏感型模型，对输入实例的排列顺序具有高度依赖性。这一特性与多实例学习任务中固有的包级序列不变性要求存在根本性冲突，同时也和视觉任务存在偏差，导致模型在病理图像分类等需弱监督定位的关键场景中易受实例顺序扰动影响，制约模型性能的稳定性与可重复性，并影响其在视觉任务更广泛的应用。

### 2.3 对比学习

对比学习作为一种自监督学习范式，通过构建正负样本对的差异性建模实现数据表征学习，已经逐渐成为计算机视觉领域的研究热点之一。其核心在于最大化语义相似样本的嵌入相似度，同时最小化不相关样本的关联性。该方法突破了传统监督学习对人工标注的依赖，其表达式如下所示：

$$\mathcal{L} = -\log \frac{\exp(\cos(f(x), f(x^+))/\tau)}{\exp(\cos(f(x), f(x^+)) + \sum_{j=1}^N \exp(\cos(f(x), f(x^-)))}. \quad (2.4)$$

图2.2展示了一些具有代表性的对比学习结构。根据对比学习是否有负样本可以将对比学习分为两类：基于负样本的对比学习、无负样本的对比学习，以下将分别介绍。

### 2.3.1 基于负样本的对比学习

基于负样本的对比学习方法通过显式地构造负样本对来实现特征判别，代表模型包括 SimCLR<sup>[53]</sup> 和 MoCo<sup>[54]</sup>。其核心在于构建动态负样本队列，利用 InfoNCE 损失函数进行噪声对比估计。以 MoCo v2 为例，通过动量编码器（动量系数 0.999）维护包含 65,536 个负样本的特征字典，在 ImageNet 线性评估协议下达到 71.1% 的 Top-1 准确率<sup>[55]</sup>。但该方法面临假负例（False Negative）问题，当负样本中真实正例比例超过 5% 时，模型性能下降 23%<sup>[56]</sup>

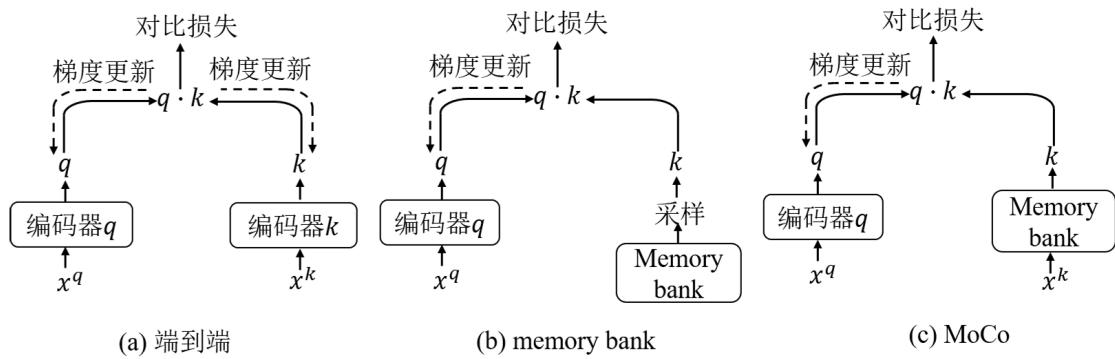


图 2.2 对比学习结构示意图。

Fig. 2.2 Contrastive Learning structure diagram.

### 2.3.2 无负样本的对比学习

为解决负样本偏差问题，BYOL<sup>[57]</sup> 等框架摒弃负样本依赖，采用非对称网络架构进行自蒸馏学习。通过在线编码器（Online Encoder）与目标编码器（Target Encoder）的交互更新，结合预测头（MLP Projector）实现特征一致性约束。实验表明，在 ImageNet 数据集上，BYOL 使用 ResNet-50 骨干网络可达到 74.3% 的线性分类准确率，且对批处理大小敏感性降低 80%。但该方法需严格防止模型坍塌（Model Collapse），通常需保留 BatchNorm 层作为隐式正则化器。

## 2.4 数据集及评价指标

病理图像分类任务通常可分为癌症诊断、亚型分类、生存预测等任务。从计算机视觉视角来看，癌症诊断本质上是图像二分类任务；癌症亚型分类，属于图像多分类范畴；而癌症生存预测，对应图像回归预测任务。接下来这一节，将围绕这三项子任务，对相关数据集、模型评估指标等展开介绍。

表 2.1 癌症诊断、分型病理图像数据集详情。# 代表具体数目

Table 2.1 Cancer diagnosis, classification WSI dataset details. # denotes the number of it.

| 数据集        | # 病例 | #WSI | # 平均图块 | # 最大图块 | # 最小图块 |
|------------|------|------|--------|--------|--------|
| CAMELYON   | 370  | 899  | 8108   | 44402  | 140    |
| TCGA-NSCLC | 956  | 1053 | 10302  | 43637  | 85     |
| TCGA-BRCA  | 1062 | 1131 | 8746   | 36618  | 283    |
| BRACS      | 189  | 547  | 8429   | 22020  | 106    |

表 2.2 生存预测病理图像数据集详情。# 代表具体数目

Table 2.2 Survival prediction Pathology image dataset details. # denotes the number of it.

| 数据集       | # 病例 | # WSI | # 平均图块 | # 最大图块 | # 最小图块 |
|-----------|------|-------|--------|--------|--------|
| TCGA-LUAD | 478  | 541   | 10629  | 45783  | 85     |
| TCGA-LUSC | 478  | 512   | 10395  | 33594  | 158    |
| TCGA-BLCA | 376  | 457   | 14364  | 34174  | 414    |

## 2.4.1 数据集

### ①癌症诊断数据集

**CAMELYON-16**<sup>[58]</sup> 与 **CAMELYON-17**<sup>[59]</sup> 作为当前开放获取规模最大的乳腺癌淋巴结转移病理诊断基准数据集, 均提供病灶转移状态的二元分类标注(转移/非转移)。其中, CAMELYON-16 数据集的训练集包含由荷兰拉德堡德大学医学中心及乌德勒支大学医学中心联合提供的 270 张全切片数字病理图像(Whole Slide Image, WSI), 并配备独立划分的 129 张官方测试样本。CAMELYON-17 数据集则进一步整合荷兰五所医疗机构的病理资源, 共收录 1000 张经过专业标注的 WSI 影像。考虑到 CAMELYON-17 官方测试集 500 张样本的标注信息尚未开放, 本研究仅采用其已公开的 500 张训练样本(对应 100 个临床病例及其影像学诊断结论)。尽管不少研究, 如文献<sup>[9,11,13,60]</sup>, 采用 CAMELYON-16<sup>[58]</sup> 数据集, 对不同模型在癌症诊断任务中的性能展开评估。然而, 该数据集仅包含 400 张病理图像, 图像数量的局限性, 削弱了评估结果的有效性。鉴于此, 本文参照 CLAM<sup>[10]</sup> 的设置, 将 CAMELYON-16 与 CAMELYON-17 数据集整合, 最终形成包含 370 个病例、899 张有效样本(阴性样本 591 例, 阳性样本 308 例)的融合数据集(见表2.1)。经预处理后, 所有病理图像在 20 倍光学放大倍率下完成分块处理, 单张 WSI 平均生成 8108 个有效图像块。

### ②亚型分类数据集

本文采用了 TCGA 肺癌数据集、浸润性乳腺癌数据集、和乳腺癌亚型数据集作为亚型分类的评估数据集。其相关数目详情如表2.1所示。

TCGA 肺癌数据集 (**TCGA-NSCLC**) 是由美国国家癌症研究所 (National Cancer Institute, NCI) 开发的癌症数据共享系统 (The Cancer Genome Atlas, 简称 TCGA)。TCGA 包括肺鳞状细胞癌 (TCGA-LUSC) 和肺腺癌 (TCGA-LUAD) 两类数据, TCGA-LUSC 的 478 个病例中有 512 个 WSI, TCGA-LUAD 478 个病例中有 541 个 WSI, 共 1053 个 WSI。该数据集仅给出了图像级别的标签。和 CAMELYON 数据集相比, 它的阳性样本中肿瘤区域明显大很多。TCGA 数据集采用了跟 CAMELYON 数据集一样的数据预处理方法。经过处理后, 每一幅放大 20 倍的图像大概能产生 10300 个图像块。TCGA 数据集仅仅提供了图像类别的标注信息, 并没有提供细致的像素标注内容。

浸润性乳腺癌数据集 (**Breast Invasive Carcinoma, BRCA**) 是肿瘤基因组图谱大型国际研究项目 (TCGA) 的一个子项目, 致力于了解和研究乳腺癌 (BRCA) 的基因组特征。BRCA 项目系统地收集了来自不同乳腺癌患者队列的肿瘤样本, 包括原发肿瘤和转移瘤, 以及正常对照组织样本 (如非癌乳腺组织)。其中浸润性导管癌 (IDC) 779 张, 浸润性小叶癌 (ILC) 198 张。

**BRACS**<sup>[61]</sup> 是一个同时用于粗粒度和细粒度分型的乳腺癌亚型数据集。其建立在帕斯卡尔基金会 (IRCCS Fondazione Pascale)、美国国家研究委员会 (National Research Council CNR) 的高性能计算与网络研究所 (ICAR) 和 IBM 苏黎世研究中心之间的一项协议之上, 该协议旨在“通过组织学图像的自动分析来识别乳腺癌病理学中的非典型肿瘤的方法和工具的开发”。在粗粒度的 BRACS 分型任务中, 每个 WSI 可分为良性 (Benign)、非典型 (Atypical) 和恶性肿瘤 (Malignant)。在细粒度的 BRACS 分型任务中, 将每一种 WSI 分为正常 (Normal)、病理良性 (Pathological Benign, PB)、普通型导管增生 (Usual Ductal Hyperplasia, UDH)、不典型导管增生 (Atypical Ductal Hyperplasia, ADH)、扁平上皮不典型 (Flat Epithelial Atypia, FEA)、导管原位癌 (Ductal Carcinoma in Situ, DCIS) 和浸润性癌 (Invasive Carcinoma, IC)。其中 Normal、PB、UDH 被分为良性病灶; FEA、ADH 被归类为非典型 (Atypical); DCIS 和 IC 被归为恶性肿瘤 (Malignant)。BRCAS 数据集被官方划分为 395 个 WSIs 的训练集、65 个 WSIs 的验证集和 87 个 WSIs 的测试集以用于评估。

### ③生存预测数据集

癌症预后生存分析的核心, 是量化评估患者群体时序生存概率的分布特征, 具体涵盖术后存活期预测、复发风险分层及生存状态演化建模等临床关键维度。本研究选取 TCGA 数据库中肺腺癌 (**TCGA-LUAD**)、肺鳞癌 (**TCGA-LUSC**) 及膀胱尿路上皮癌 (**TCGA-BLCA**) 三个项目的临床随访数据构建愈后预测模型。需要特别说明的是, 与上文所述诊断及分型任务不同, 生存预测模型以独立病例为最小分析单元: TCGA-LUAD 与 TCGA-LUSC 项目沿用分型研究中的原始影像数据, 但重新构建了包含生存时间、复发状态等临床终点的标注体系; TCGA-BLCA 数据

集则专门收录 376 例膀胱尿路上皮癌（Bladder Urothelial Carcinom）患者的完整临床随访记录。三个数据集的相关数目详情见表2.2。在实验设计层面，通过整合多中心临床数据构建生存分析框架，重点考察模型对患者个体化生存曲线的预测效能。

#### 2.4.2 评价指标

在癌症诊断与分型分类模型的性能评估中，本研究仿照相关文献<sup>[13,60]</sup>，采用标准化的评价指标体系。基于分类任务的特征差异，分别选用准确率、ROC 曲线下面积（AUC）及 F1 分数作为核心评估参数：针对具有二分类特性的肿瘤诊断问题，采用二元 F1 分数（F1-binary）进行量化分析；而对于涉及多类别判别的癌症亚型分类任务，则选用宏观 F1 分数（F1-macro）作为评价基准。AUC 对类别分布不均衡具有较强鲁棒性，且对分类决策阈值变化具有较低敏感性，能够有效量化模型的分类效能。鉴于上述优势，本研究将 AUC 作为核心评价指标，并在消融实验环节着重分析模型的 AUC 表现。

在生存预测任务中，本研究依据文献<sup>[62-66]</sup> 的设置，选取一致性指数（Concordance Index, C-index）<sup>[67]</sup> 作为生存分析模型的核心评估指标。该指标在继承 AUC 评估特性的基础上，针对生存数据特有的删失（censored data）问题进行了算法扩展，可系统评估模型的判别效能——即通过个体风险评分对生存时间进行可靠排序的能力。其数学表达式如2.5所示：

$$\text{C-index} = \frac{\sum_{i \neq j} \delta_j \cdot I(T_i < T_j) \cdot I(\hat{S}_i > \hat{S}_j)}{\sum_{i \neq j} \delta_j \cdot I(T_i < T_j)}, \quad (2.5)$$

其中  $T_j$  表示样本  $j$  的实际生存时间， $\hat{S}_j$  为模型预测的生存概率， $\delta_j \in \{0, 1\}$  为事件状态指示器（ $\delta = 1$  表示发生终点事件）。指示函数  $I(\cdot)$  在条件成立时取 1，否则为 0。该计算架构通过约束条件  $\delta_j = 1$ ，确保仅纳入实际发生事件的样本对进行统计评估，从而提升结果的可信度。指标取值范围具有明确临床解释性：当 C-index = 1 时表征模型具备完全预测能力，而 C-index = 0.5 则等效于随机猜测。相较于传统 AUC 指标，C-index 的优势在于能够有效处理生存分析中普遍存在的删失数据问题，其评估机制更符合临床研究数据的实际分布特征。

### 2.5 病理图像的预处理过程

本研究针对数字病理图像高分辨率特性（1-10 亿像素）带来的计算挑战，采用多阶段预处理框架实现高效特征提取（图 2.3）。

**分割组织。**首先基于 Openslide 库读取 WSI 数据，通过多尺度分割策略提取有效组织区域。具体操作是先在 32 倍降采样层级下，将图像转换到 HSV 色彩空间，使用中值滤波对图像边缘进行平滑处理，之后依据饱和度通道阈值生成初始

二值掩码；随后实施形态学闭运算消除微米级孔洞，并通过面积阈值过滤非组织区域<sup>[10]</sup>。为提升算法鲁棒性，系统自动生成包含分割参数及处理日志的元数据文件，支持人工复核与参数微调。

**划分图块并提取特征。**在 20 倍光学放大倍率下，采用滑动窗口法从有效组织区域裁剪  $256 \times 256$  像素图像块，存储为标准化 hdf5 格式并记录空间坐标信息。该尺度选择平衡了细胞形态学特征保留与计算效率，单个 WSI 可产生  $10^3\text{-}10^4$  量级图像块（视组织面积而定）。在特征提取环节，本研究采用了两种特征提取器：其一，采用在 ImageNet 数据集上进行预训练的 ResNet-50 模型<sup>[68]</sup>，该模型截断至第三残差块，随后应用自适应均值池化操作，把图像块编码成维度为 1024 的向量；其二，在“Mass-100K”上预训练的通用的、用于病理学的自监督视觉编码器 UNI<sup>[69]</sup>，同样输出 1024 维向量。将提取的特征作为在线深度学习模型监督学习的输入，能大幅缩短训练时间、降低计算成本。基于此构建的特征矩阵，让单个 GPU 在数小时内，就能完成万级全切片图像（WSI）的模型训练。

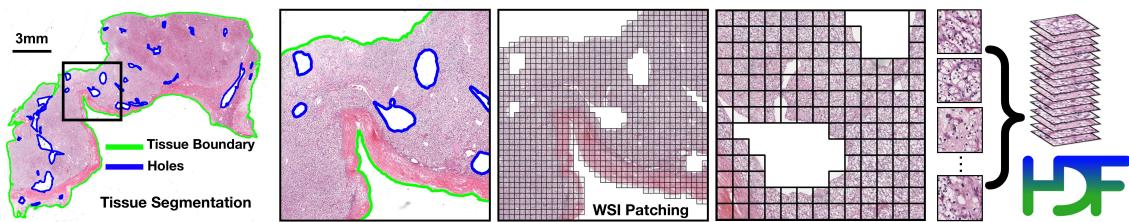


图 2.3 数字病理图像的预处理过程示意图。<sup>[10]</sup>

Fig. 2.3 Illustration of the WSI preprocessing process.<sup>[10]</sup>

## 2.6 本章小结

本章首先详细介绍病理图像分类常用的多实例学习范式的分类及发展过程，其次介绍了 Mamba 架构的特点、优势及迁移至视觉领域所面临的挑战。然后对后续研究工作所涉及到的对比学习等相关技术进行了详细介绍，根据是否使用负样本将其分为基于负样本对比学习和无负样本的对比学习进行了详细阐述。接着，介绍了本文方法所涉及到的数字病理图像分类子任务、数据集和评价指标等，最后介绍了病理图像的预处理阶段。

### 3 基于多扫描 Mamba 的高分辨率病理图像分类研究

本章研究了基于多扫描 Mamba 的高分辨率病理图像分类研究，通过整合多种扫描模式分支的特征信息，引入病理图像的空间信息，提高 Mamba 建模 2D 空间关系的能力，进而提高高分辨率病理图像分类性能。本章内容共分为四节，第一节介绍研究的动机；第二节介绍问题的描述与符号定义；第三节介绍本章提出的基于空间信息增强的多扫描 Mamba 模型；第四节给出实验设置和结果分析；第五节对本章进行总结与讨论。

#### 3.1 研究动机

病理图像的多实例学习范式中，由于病理图像的分辨率过高，往往面临实例数量庞大、信息繁杂等问题，因此如何高效地提取实例关系并突出关键实例特征一直是多实例学习的核心。近年来，Mamba 模型凭借其高效的计算速度和强大的数据建模能力，逐渐被引入病理图像多实例学习领域，并在特征聚合方面取得了瞩目的效果。然而，Mamba 模型本质是一类序列学习方法，其输入是单个序列，而单个序列无法体现出视觉的空间关系，因而单个序列扫描模式的 Mamba 模型的性能会受其空间关系建模的限制。

为了缓解上述所提到的问题，业内主流的解决方案是将多个扫描模式序列的信息进行整合。然而，如图3.1左图所示，大多数方法虽然通过类似横向和竖向的扫描模式缓解了 Mamba 序列敏感特性和视觉任务的差异，但是依旧只是在一维序列中进行，并没有充分利用视觉任务的特性如局部连续性、旋转不变性等，更没有关注到空间信息的捕获。为此，本章从引入局部连续性、旋转不变性的视觉特性出发，提供了两种新的扫描模式，综合多个扫描模式分支以建模出更全面特征。此外，本章还通过进一步综合局部空间信息，来提高模型对视觉任务的适配。

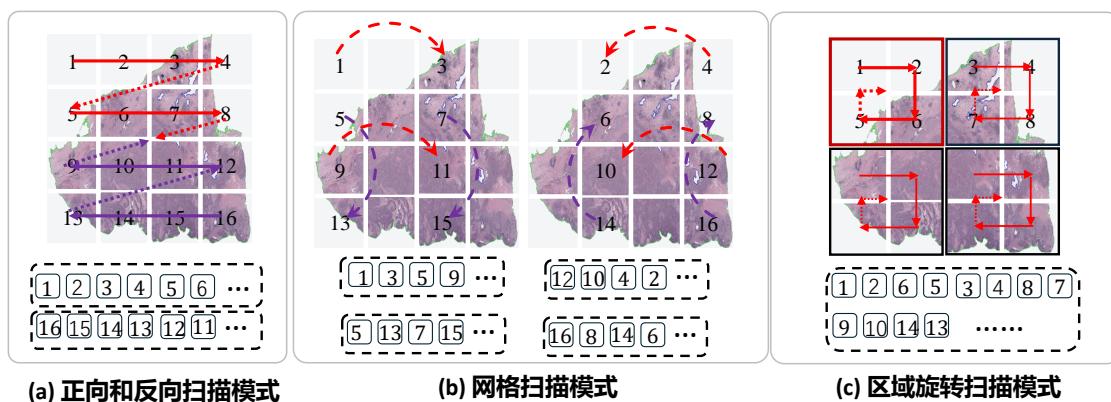


图 3.1 常规扫描、网格扫描和区域螺旋扫描示意图。

Fig. 3.1 Schematic diagram of conventional scan, grid scan and region spiral scan.

### 3.2 问题描述和符号定义

本研究采用多实例学习范式构建高分辨率病理图像分类模型，其数据集由 WSI 全切片图像与对应标签构成的样本对  $(X, Y)$  组成，

$$D = \{(X_1, Y_1), \dots, (X_j, Y_j), \dots, (X_n, Y_n)\}, \quad (3.1)$$

其中， $n$  为数据集样本数量。每一个 WSI 切片  $X_j$  都被视为一个包，将其 Patch 视为该包中的实例，可以表示为  $X_j = \{x_i^j\}_{i=1}^N$ 。实例数量  $N$  因包而异。包的类别由其中实例类别综合可得。如果一个包包含至少一个阳性实例，则认为它是阳性的；否则，它被归类为负包：

$$Y(X) = \begin{cases} 0 & \text{if } \sum_{x \in X} y(x) = 0, \\ 1 & \text{otherwise.} \end{cases} \quad (3.2)$$

在多实例学习范式中，样本包级标签  $Y(X)$  作为监督信号已知，而实例级标签  $y(x)$  则保持隐式状态。该学习机制的核心目标在于通过聚合实例层特征信息，即用所有实例  $\hat{Y} \leftarrow \mathcal{M}(X)$  预测包标签，最终实现病理图像的准确分类。根据最近流行的方法，MIL 预测过程可以分为三个步骤：实例特征提取、实例特征聚合和包分类。具体来说，这个过程可以定义如下：

$$\hat{Y} \leftarrow \mathcal{M}(X) := \mathcal{C}(\mathcal{A}(\mathcal{F}(X))), \quad (3.3)$$

其中  $\mathcal{F}(\cdot)$ 、 $\mathcal{A}(\cdot)$  和  $\mathcal{C}(\cdot)$  分别是上述步骤的映射函数。

由于同一包中存在大量实例，因此 WSI 分类问题可以看作是一个长序列建模问题。基于现代状态空间模型 (SSM)，前人设计了结构化状态空间序列模型 (S4) 和 Mamba 等几种适合长序列建模的方法。经典的 S4 模型可以写成线性递归式：

$$\begin{aligned} h_t &= \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t, \\ y_t &= \mathbf{C}h_t. \end{aligned} \quad (3.4)$$

其中  $\bar{\mathbf{A}}$  表示离散化后的状态矩阵， $\mathbf{B}$  和  $\mathbf{C}$  表示投影参数， $\bar{\mathbf{B}}$  为离散化后的  $\mathbf{B}$ 。另外，模型可以使用卷积模式计算输出：

$$\begin{aligned} \bar{\mathbf{K}} &= (\mathbf{C}\bar{\mathbf{B}}, \mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}, \dots, \mathbf{C}\bar{\mathbf{A}}^{N-1}\bar{\mathbf{B}}), \\ Y &= X * \bar{\mathbf{K}}, \end{aligned} \quad (3.5)$$

其中  $K \in R^L$  表示一个结构化卷积核， $N$  表示输入序列  $X$  的长度。

Mamba 进一步将选择机制集成到 S4 模型中，通过高效的硬件感知并行算法，

使参数动态依赖于输入。这允许 Mamba 根据当前上下文有选择地沿着序列传播或忘记信息，从而实现高效的长序列建模。因此，结合 Mamba 的多实例学习范式可以被定义：

$$\hat{Y} \leftarrow \mathcal{M}(X) := \mathcal{C}(\mathcal{A}(\Psi(\mathcal{F}(X)))), \quad (3.6)$$

其中  $\Psi(\cdot)$  表示为 Mamba 模型。

### 3.3 基于空间信息增强的多扫描 Mamba 模型

#### 3.3.1 方法概述

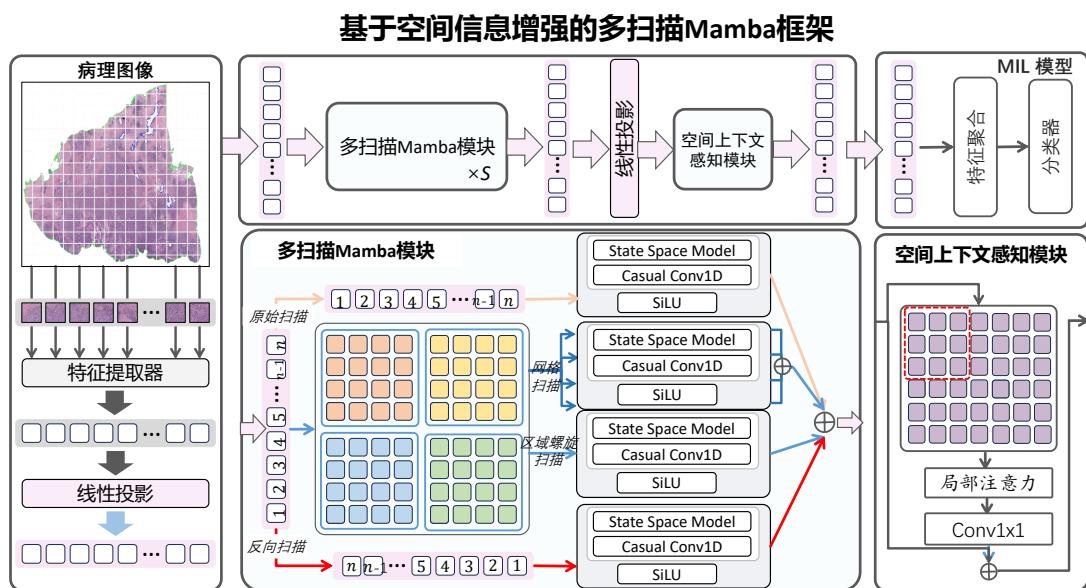


图 3.2 基于多扫描 Mamba 的多实例学习架构示意图。

Fig. 3.2 Illustration of the multi-instance learning architecture with multi-scan Mamba

为了解决 3.1 小节中的问题，如图 3.2 所示，本章提出了一种基于空间信息增强的多扫描 Mamba 模型，该模型既发挥了 Mamba 长序列建模能力强的特点，同时也缓解其和视觉 2D 图像任务间的差异。具体而言，本研究通过添加区域螺旋扫描和网格扫描等多种扫描模式分支，引入局部连续性、旋转不变性等的 2D 图像特性。此外，为了进一步整合 2D 空间关系信息，本章引入一个 2D 空间上下文感知模块。总的来说，本研究的贡献如下：

- 本研究提出了一种基于多扫描 Mamba 的多实例学习架构，命名为  $M^2S\text{-MIL}$ 。 $M^2S\text{-MIL}$  通过添加区域螺旋扫描和网格扫描等扫描模式，引入 2D 图像的局部依赖和旋转不变性等，使之更适合病理图像分类任务。
- 针对视觉领域的 2D 属性，设计了新的扫描模式：区域螺旋扫描模式和网格扫描模式，引入了更多局部依赖、旋转不变性等 2D 空间信息，缓解顺序敏感的 Mamba 与病理图像分类任务间的差异。

- 为了进一步整合提升 2D 空间关系信息，本章设计了一个 2D 空间上下文感知模块，采用分块局部注意力，进一步融合局部相邻的实例信息，进一步提高模型性能。

### 3.3.2 预处理与特征提取

给定一个 WSI 图片，参照 CLAM<sup>[10]</sup> 的设置，本章将之裁剪成一系列  $256 \times 256$  的非重叠图块，放大倍数为 20 倍，同时丢弃背景区域（饱和度  $< 15$ ）。本章采用 ResNet-50<sup>[70]</sup> 或 UNI<sup>[69]</sup> 作为底层特征提取器  $\mathcal{F}(\cdot)$ ，对每个实例提取特征得到 1024 维的原始特征。为了减少特征冗余和计算消耗，使用多层感知器（MLP）层将这些特征的维度从 1024 减少到 512。数字病理图像的特征提取过程为：

$$z_i = \mathcal{F}_\varphi(x_i) \in \mathbb{R}^d. \quad (3.7)$$

在该模型架构中， $\mathcal{F}_\varphi(\cdot)$  被定义为骨干特征提取网络，其参数  $\varphi$  表示特征学习模块的可训练参数集，通常其不与后续网络一起端到端训练， $d = 512$ 。最后，一个 WSI 由一个连续长序列表示。本章将之定义为原始扫描序列  $Z_{origin} = \{z_i\}_{i=1}^N$ ，其中  $z_i$  是  $X$  的第  $i$  个实例的提取特征， $N$  是实例的个数。

### 3.3.3 多扫描 Mamba 模块

为了使 Mamba 能充分学习到实例间的相互依赖，并缓解其本身因果建模倾向所带来的负面影响，本章方法提出使用多种扫描模式，以消除单一扫描序列所建模出的偏执，捕获实例间更全面的特征信息。如图 3.2 所示，在多扫描 Mamba 模块中可分为四个扫描分支：原始扫描序列分支、反向扫描序列分支、区域螺旋扫描分支和网格扫描分支。首先，原始扫描模式序列被输入 Mamba 模型以提供常规特征：

$$Z'_{origin} = \Psi_{origin}(Z_{origin}), \quad (3.8)$$

其中， $\Psi_{origin}(\cdot)$  同 3.2 小节所示为 Mamba 模型。反向扫描序列以初步缓解 Mamba 不过分学习到正向序列的特有顺序偏执：

$$Z'_{flip} = \Psi_{flip}(Z_{flip}) = \Psi_{flip}(\{z_i\}_{i=N}^1). \quad (3.9)$$

为进一步引入空间信息，本章精心设计了两种新的扫描模式：区域螺旋扫描和网格扫描模式。其中，区域螺旋扫描被设计以引入局部连续性和旋转不变性等 2D 图像属性，网格扫描模式用于在一个序列中捕获多个方向上的信息。

具体而言，输入的实例特征序列首先被重塑为 2D 特征图，初始的二维特征图形式为  $H \in \mathbb{R}^{N \times D}$ ，之后转变为  $H \in \mathbb{R}^{[\sqrt{N}] \times [\sqrt{N}] \times D}$ 。重塑后便可如图 3.1 所示的

网格扫描方式。网格子序列 1 可表示为：

$$Z_{grid1} = \left\{ z_{ij} \mid i = 2k + 1, j = 2m + 1, k, m \in \mathbb{N}_0, k, m \leq \left\lfloor \frac{\lceil \sqrt{N} \rceil}{2} \right\rfloor \right\}. \quad (3.10)$$

同理，网格序列集合 2、3、4 可分别表示为：

$$\begin{aligned} Z_{grid2} &= \left\{ z_{ij} \mid i = 2k, j = 2m + 1, k, m \in \mathbb{N}_0, k, m \leq \left\lfloor \frac{\lceil \sqrt{N} \rceil}{2} \right\rfloor \right\}, \\ Z_{grid3} &= \left\{ z_{ij} \mid i = 2k + 1, j = 2m, k, m \in \mathbb{N}_0, k, m \leq \left\lfloor \frac{\lceil \sqrt{N} \rceil}{2} \right\rfloor \right\}, \\ Z_{grid4} &= \left\{ z_{ij} \mid i = 2k, j = 2m, k, m \in \mathbb{N}_0, k, m \leq \left\lfloor \frac{\lceil \sqrt{N} \rceil}{2} \right\rfloor \right\}. \end{aligned} \quad (3.11)$$

网格扫描模式所得的四个子序列(都是原始特征集合的一部分)被分别输入 Mamba 后，将所输出子序列拼接在一起并还原回原本的位置。

$$Z'_{grid} = Concat(\Psi_{grid}^1(Z_{grid1}), \Psi_{grid}^2(Z_{grid2}), \Psi_{grid}^3(Z_{grid3}), \Psi_{grid}^4(Z_{grid4})). \quad (3.12)$$

网格扫描模式在同一序列中综合多方向信息，以缓解单一扫描模式序列所带来的顺序偏向。

同时，如图 3.1 所示，特征图  $H$  被均匀划分为  $L \times L$  个非重叠式区域，以便于进行区域螺旋扫描。每个区域由  $M \times M$  个实例构成，并且满足关系： $L \times M = \lceil \sqrt{N} \rceil$ 。此外，为确保区域数量为  $(L, L)$  时能整除特征图大小，本章将对特征图进行零填充操作。因此特征图  $H$  被划分为  $\{H^l\}_{l=1}^{L^2}$ ,  $H \in \mathbb{R}^{M \times M \times D}$ 。本章在划分后的区域内分别进行区域螺旋扫描，得到序列  $Z_{spiral}$  ( $Z_{spiral}$  由螺旋扫描唯一确定)，并将螺旋扫描序列输入到 Mamba 进行建模，得到序列：

$$Z'_{spiral} = \Psi_{spiral}(Z_{spiral}). \quad (3.13)$$

该扫描模式相较于 LocalMamba<sup>[30]</sup> 不仅引入了局部连续性，同时引入了 2D 病理图像的旋转不变性，进一步加强了空间信息，提高 Mamba 模型对视觉特性的建模能力。

最后，将四种扫描模式经由 Mamba 建模的实例序列分别还原至原始位置后，相互融合，得出多扫描分支 Mamba 模块的输出序列：

$$Z_{multi\_scan} = Merge(Z'_{origin}, Z'_{flip}, Z'_{grid}, Z'_{spiral}), \quad (3.14)$$

其中  $Merge(\cdot)$  表示将四个分支序列对应融合成一个序列。本章将  $Z_{multi\_scan}$  作为多扫描 Mamba 模块的输出。除非特殊说明，本章中的多扫描 Mamba 模块层数  $S$  设为 1。综合了多种扫描模式的 Mamba，不仅保留了 Mamba 本身极强的建模能力，并且还引入了二维空间信息、视觉特性等，更加适配病理图像分类任务。

### 3.3.4 2D 空间上下文感知模块

为了进一步提升模型捕获空间信息的能力，本章将综合多种扫描信息的序列特征输入到基于分块局部注意力的 2D 空间上下文感知模块中。如图 3.2 所示，本章将输出的 1D 特征序列再次重塑为 2D 特征图 ( $H' \in \mathbb{R}^{\lceil \sqrt{N} \rceil \times \lceil \sqrt{N} \rceil \times D}$ )，然后将特征划分为固定大小 ( $M' \times M'$ ) 的局部窗口（不同于上一小节的区域），每窗口内计算局部注意力。并通过  $1 \times 1$  卷积还原原本的分辨率，避免信息损失。记某一个窗口内的输入特征为  $H'_{win}$ ，则窗口内特征图可表示为：

$$\begin{aligned} Q_{win} &= \text{Conv}_{1 \times 1}(H'_{win}), \quad K_{win} = \text{Conv}_{1 \times 1}(H'_{win}), \quad V_{win} = \text{Conv}_{1 \times 1}(H'_{win}), \\ H''_{win} &= \text{Conv}_{1 \times 1}\left(\sigma\left(\frac{Q_{win}K_{win}^T}{\sqrt{d_k}}\right)V_{win}\right), \end{aligned} \quad (3.15)$$

其中  $\sigma$  是 softmax 函数， $Q_{win}$ ,  $K_{win}$  和  $V_{win}$  分别是局部注意力机制的窗口内查询特征、窗口内键特征、窗口内值特征。最后，全局的特征图  $H'''$  是通过分块局部注意力结果  $H''$  和多扫描分支 Mamba 模块的结果  $H'$  残差连接而得。

$$H''' = H'' + H', \quad (3.16)$$

因此，本章可以获得一个包含 2D 空间信息特征图  $H''' \in \mathbb{R}^{\lceil \sqrt{N} \rceil \times \lceil \sqrt{N} \rceil \times D}$ ，并将之转换为 1D 实例序列  $Z_{output} = \{z_{output}^1, z_{output}^2, \dots, z_{output}^N\}$ 。

### 3.3.5 基于 Mamba 的实例聚合

在特征融合阶段，本小节采用基于注意力机制的特征融合策略。具体而言，构建包含双全连接层的上下文感知注意力模块，通过可训练参数动态计算每个实例特征的注意力权重。该权重能够有效量化各实例特征对包级分类决策的贡献程度，进而通过加权平均操作将实例级特征整合为最终的包特征表示：

$$\mathbf{b} = \mathcal{A}(Z_{output}) = \sum_{i=1}^N a_i z_{output}^i. \quad (3.17)$$

在该模块中， $a_i$  被定义为第  $i$  个实例特征对应的可训练注意力权重参数。该参数通过反向传播机制动态优化，用于量化实例特征对包级分类决策的贡献程度，并满

足归一化约束条件  $\sum_{i=1}^N a_i = 1$ :

$$a_i = \frac{\exp\{u^T \tan(vz_{output}^i)^T\}}{\sum_{j=1}^N \exp\{u^T \tan(vz_{output}^j)^T\}}, \quad (3.18)$$

其中， $u$  和  $v$  分别表示两个全连接层的网络参数。

### 3.3.6 模型优化和推理

预测的包标签可以通过将包的表示输入分类器  $\mathcal{C}(\cdot)$  得到:

$$\hat{Y} \leftarrow \mathcal{C}(\mathbf{b}) := \mathcal{C}(\mathcal{A}(Z_{output})). \quad (3.19)$$

本章内容仅使用常规的分类损失。包级分类预测的损失值通过交叉熵损失函数计算，其数学定义如式 3.20 所示：

$$\mathcal{L}_{cls} = Y \log \hat{Y} + (1 - Y) \log(1 - \hat{Y}), \quad (3.20)$$

其中， $Y$  表示包的真实标签， $\hat{Y}$  则表示包预测标签。综合上述内容，该模型的优化过程能够用如下公式来呈现，并且通过反向传播算法来求解得出整个模型的最优参数。

$$\{\hat{\Psi}, \hat{\mathcal{R}}, \hat{u}, \hat{v}, \hat{\mathcal{C}}\} \leftarrow \arg \min_{\Psi, \mathcal{R}, u, v, \mathcal{C}} \mathcal{L} = \mathcal{L}_{cls}, \quad (3.21)$$

其中  $\Psi$  是本章所有 Mamba 架构的参数， $\mathcal{R}$  表示本章空间上下文感知模块的参数， $u, v$  分别表示基于注意力机制的聚合模块中两个全连接层的网络参数， $\mathcal{C}$  为分类器的网络参数。

在推理阶段，本章采用和训练阶段保持一致的流程。同样采用四种扫描模式获取四个扫描序列分支分别输入 Mamba 架构，而后经过 2D 空间上下文感知模块进行空间结构信息增强，最后采用基于注意力机制的聚合模块聚合包特征，并通过分类器进行推断分类。

## 3.4 实验设置及结果分析

在本节中，首先介绍了本章方法的实验设置，包括实验所用环境、数据集、评估指标和实施细节，然后分析了基于多扫描 Mamba 的高分辨率病理图像分类实验结果，接下来对模型的各个模块以及超参数进行了消融实验和分析，最后对模型所关注的重点区域进行了可视化分析。

### 3.4.1 实验设置

#### (1) 实验环境

本章节的实验工作基于实验室高性能计算平台完成，该平台搭载 AMD Ryzen 5700X 20 核中央处理器与 3 块 NVIDIA RTX 3090 图形处理器。实验框架基于 Python 编程语言构建，采用 OpenSlide 库实现病理图像的高效读取，基于 Matplotlib 库完成数据可视化。详细的实验环境参数配置见表 3.1：

表 3.1 基于空间信息增强的多扫描 Mamba 模型的实验环境及配置

Table 3.1 Experimental environment and configuration of M<sup>2</sup>SMIL

| 设备     | 配置               | 版本          | 数量 |
|--------|------------------|-------------|----|
| 操作系统   | Ubuntu           | 20.04.4 LTS | 1  |
| CPU    | AMD Ryzen 5700X  | -           | 1  |
| GPU    | GeForce RTX 3090 | -           | 3  |
| IDE    | VsCode           | 1.86.2      | 1  |
| 编程语言   | Python           | 3.10.11     | -  |
| 深度学习框架 | Pytorch          | 2.0.1       | -  |

## (2) 数据集

本章通过三个 WSI 分析子任务来评估 M<sup>2</sup>S-MIL 模型：癌症诊断、亚型分类和生存预测。对于癌症诊断任务，尽管已有大量研究<sup>[9,11,60]</sup> 基于 CAMELYON-16<sup>[58]</sup> 数据集对不同模型在癌症诊断任务中的性能进行了评估，但该数据集仅包含 400 张病理图像，数量上的局限性对模型评估的全面性与准确性造成了显著影响。因此，为克服这一问题，本章参照文献<sup>[10,13]</sup> 的做法，将 CAMELYON-16<sup>[58]</sup> 数据集与 CAMELYON-17<sup>[59]</sup> 数据集进行合并，构建了一个包含 899 张病理图像的综合性 CAMELYON 数据集。这一策略在一定程度上缓解了数据量匮乏导致的评估困境，为模型性能评估提供了更为丰富和可靠的数据基础。对于分型任务，本章使用 TCGA-NSCLC、TCGA-BRCA 和 BRACS<sup>[61]</sup> 数据集。为了评估生存预测的准确性，本章使用 TCGA-LUAD、TCGA-LUSC、TCGA-BLCA 来评估生存预测任务的性能。

## (3) 评估指标

对于诊断和分型，本章利用 Accuracy、AUC 和 F1-score 来评估模型的性能，其中 AUC 是本章主要的评价指标，在消融实验的讨论中，本章仅报告 AUC。对于生存预测，本章报告了所有数据集的 C-index。默认情况下，除 BRACS 数据集外，本章采用 5 折交叉验证进行实验。而对于 BRACS 数据集，本章也遵循了文献<sup>[69]</sup> 的设置，遵循官方数据划分标准，将数据集划分为训练集（395 例）、验证集（65 例）和测试集（87 例），该划分方案有效保证了数据分布一定的合理性与模型评估的客观性，而后进行粗粒度的三分类实验。并且进行了三次独立的重复实验，以尽量减少官方划分的随机影响。

## (4) 实施细节

按照之前的工作<sup>[10,12,60]</sup>, 每个 WSI 图片被裁剪成一系列  $256 \times 256$  的不重叠的贴片, 放大倍数为 20 倍, 同时丢弃背景区域(饱和度  $< 15$ )。本章使用在 ImageNet<sup>[71]</sup> 中预训练的 ResNet-50 模型<sup>[68]</sup> 和在  $1 \times 10^6$  WSIs 的  $1 \times 10^8$  patches 的大型内部组织学数据集上预训练的最新基础模型 UNI<sup>[69]</sup> 作为骨干网络, 从每个 patch 中提取初始特征向量, 其维度为 1024。初始特征向量随后通过全连接层降为 512 维表示。在处理所有数据集时, 将学习率设定为  $2 \times 10^{-4}$ , 权重衰减系数设为  $10^{-5}$ 。学习率的调整采用余弦退火策略, 该策略能够使学习率随着训练轮次的增加而动态、平滑地变化, 有助于模型更有效地收敛。所有模型均运用早停策略开展了 200 个轮次的训练。早停策略可在验证集性能不再提升时及时停止训练, 防止模型过拟合, 从而保证模型在测试集上具备良好的泛化能力。早停策略的耐心值除 CAMELYON 数据集外都设置为 20, 而 CAMELYON 设置为 30。本章未引入梯度裁剪或梯度累积等优化技巧, 以确保实验结果的纯粹性与方法对比的公平性。所有模型均基于标准训练流程完成参数优化, 未采用任何额外的正则化技术或训练技巧。批量大小设置为 1。

### 3.4.2 基准数据集实验结果

#### (1) 癌症诊断和分型

表 3.2 M<sup>2</sup>S-MIL 在 CAMELYON 数据集的癌症诊断性能比较 (%). 最优结果用粗体表示, 次优结果使用下划线表示。

Table 3.2 Comparison of cancer diagnostic performance of M<sup>2</sup>S-MIL on the CAMELYON dataset (%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | Accuracy                           | AUC                                | F1-score                           |
|--------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| AB-MIL <sup>[43]</sup>               | $88.08 \pm 1.87$                   | $91.18 \pm 1.60$                   | $83.59 \pm 2.75$                   |
| CLAM <sup>[10]</sup>                 | $87.98 \pm 2.17$                   | $91.70 \pm 2.15$                   | $83.38 \pm 3.09$                   |
| DSMIL <sup>[9]</sup>                 | $88.42 \pm 1.77$                   | $91.50 \pm 1.23$                   | $83.82 \pm 2.95$                   |
| TransMIL <sup>[11]</sup>             | $88.75 \pm 1.04$                   | $91.82 \pm 1.82$                   | $84.47 \pm 1.57$                   |
| DTFD-MIL <sup>[12]</sup>             | $88.07 \pm 1.13$                   | $91.61 \pm 1.73$                   | $83.24 \pm 1.67$                   |
| MHIM-MIL <sup>[60]</sup>             | <u><math>89.42 \pm 1.16</math></u> | <u><math>92.55 \pm 1.11</math></u> | $84.63 \pm 2.13$                   |
| IBMIL <sup>[72]</sup>                | $88.98 \pm 2.08$                   | $91.75 \pm 2.09$                   | $84.61 \pm 3.07$                   |
| SRMamba <sup>[52]</sup>              | $88.75 \pm 1.47$                   | $91.82 \pm 0.93$                   | $84.27 \pm 2.08$                   |
| WiKG <sup>[73]</sup>                 | $88.19 \pm 2.30$                   | $91.42 \pm 1.44$                   | $82.93 \pm 3.41$                   |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | $89.20 \pm 1.48$                   | $92.29 \pm 1.32$                   | <u><math>85.20 \pm 1.86</math></u> |
| <b>M<sup>2</sup>S-MIL(Ours)</b>      | <b><math>89.76 \pm 1.33</math></b> | <b><math>93.02 \pm 1.20</math></b> | <b><math>85.48 \pm 2.43</math></b> |

表 3.2 给出了各种 MIL 方法在 CAMELYON 上的癌症诊断性能。这些结果表明, 本章的方法在所有基准测试的所有指标下都实现了最佳性能。具体来说, 在 CAMELYON 数据集上, 本章的方法在 Accuracy、AUC 和 F1-score 方面分别比次

优的方法提高了 0.34%、0.47% 和 0.28%。同样值得注意的是，与同样基于 Mamba 架构的 SRMamba 相比，所提出的 M<sup>2</sup>S-MIL 的性能得到了显着提高。M<sup>2</sup>S-MIL 在 Accuracy、AUC 和 F1-score 方面分别比 SRMamba 高 0.78%，1.2% 和 1.21%。这验证了本章的假设，即本章设计的架构成功引入了空间信息，使模型更好捕获 2D 空间信息与视觉特性。

表 3.3 M<sup>2</sup>S-MIL 在 TCGA-NSCLC 数据集的亚型分类性能比较 (%). 最优结果用粗体表示，次优结果使用下划线表示。

Table 3.3 Comparison of sub-typing performance of M<sup>2</sup>S-MIL on the TCGA-NSCLC dataset (%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB-MIL <sup>[43]</sup>               | 90.42±1.76        | 95.23±1.46        | 89.93±1.80        |
| CLAM <sup>[10]</sup>                 | 87.98±2.17        | 91.70±2.15        | 83.38±3.09        |
| DSMIL <sup>[9]</sup>                 | 90.33±1.97        | 95.68±1.01        | 89.91±2.19        |
| TransMIL <sup>[11]</sup>             | 90.13±1.07        | 94.92±1.02        | 89.52±1.07        |
| DTFD-MIL <sup>[12]</sup>             | 90.71±1.66        | 95.64±1.44        | 90.29±1.83        |
| MHIM-MIL <sup>[60]</sup>             | <u>91.18±1.79</u> | 95.84±1.41        | <u>90.91±1.97</u> |
| IBMIL <sup>[72]</sup>                | 89.76±2.29        | 95.49±1.18        | 89.43±2.21        |
| SRMamba <sup>[52]</sup>              | 90.43±3.26        | 95.07±1.96        | 90.18±3.21        |
| WiKG <sup>[73]</sup>                 | 89.67±2.49        | 94.96±1.66        | 89.53±2.43        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 90.81±2.41        | <u>96.20±1.16</u> | 90.49±2.41        |
| <b>M<sup>2</sup>S-MIL(Ours)</b>      | <b>91.48±2.29</b> | <b>96.65±1.08</b> | <b>91.19±2.40</b> |

表 3.3 给出了各种 MIL 方法在 TCGA-NSCLC 上的癌症分型性能。这些结果同样如图 3.2，本章的方法在所有基准测试的所有指标下都实现了最佳性能。具体来说，在 TCGA-NSCLC 数据集上，本章的方法在 Accuracy、AUC 和 F1-score 方面分别比 SOTA 的方法提高了 0.3%、0.45% 和 0.28%。与同样基于 Mamba 架构的 SRMamba 相比，所提出的 M<sup>2</sup>S-MIL 的性能提升分别为 1.05%、1.58% 和 1.01%。这验证了本章所提出模型的有效性。

表 3.4 M<sup>2</sup>S-MIL 在 TCGA-BRCA 数据集的亚型分类性能比较 (%). 最优结果用粗体表示，次优结果使用下划线表示。

Table 3.4 Comparison of sub-typing performance of M<sup>2</sup>S-MIL on the TCGA-BRCA dataset (%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                     | Accuracy   | AUC        | F1-score   |
|------------------------|------------|------------|------------|
| AB-MIL <sup>[43]</sup> | 86.80±2.56 | 91.17±1.69 | 72.42±3.48 |
| CLAM <sup>[10]</sup>   | 86.15±3.98 | 91.38±1.67 | 72.29±4.21 |

表 3.4 (续)  
Table 3.4 (continued)

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| DSMIL <sup>[9]</sup>                 | 86.58±3.86        | 91.80±2.09        | 72.63±4.49        |
| TransMIL <sup>[11]</sup>             | 86.22±3.30        | 91.61±1.81        | 72.20±4.28        |
| DTFD-MIL <sup>[12]</sup>             | 86.04±3.05        | 91.34±1.95        | 70.76±4.20        |
| MHIM-MIL <sup>[60]</sup>             | 86.80±2.56        | 92.24±1.63        | 74.77±7.72        |
| IBMIL <sup>[72]</sup>                | 85.54±3.20        | 91.41±2.20        | 71.14±3.60        |
| SRMamba <sup>[52]</sup>              | 88.14±3.05        | <u>93.31±1.40</u> | <u>75.61±4.00</u> |
| WiKG <sup>[73]</sup>                 | <u>88.53±2.28</u> | 92.14±2.06        | 75.11±3.98        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 88.25±2.06        | 92.58±1.75        | 75.15±2.91        |
| <b>M<sup>2</sup>S-MIL(our)</b>       | <b>89.26±2.37</b> | <b>93.50±1.70</b> | <b>76.62±3.47</b> |

表 3.4 报告了 BRCA 数据集上的亚型分类性能比较。观察结果与表格 3.2, 表 3.3 所示一致。本章的模型在所有指标上都达到了最先进的性能。与次优方法相比, 该方法在 BRCA 数据集上的 Accuracy、AUC 和 F1-score 分别提高了 0.73%、0.19% 和 1.01%。在与非 Mamba 模型相比, AUC 增加了 0.92%, 这体现了 Mamba 性能的强大建模能力。

表 3.5 M<sup>2</sup>S-MIL 在 BRACS 数据集官方划分下的粗粒度分类 (三分类) 性能对比 (%). 最优结果用粗体表示, 次优结果使用下划线表示。

Table 3.5 Performance comparison of M<sup>2</sup>S-MIL coarse-grained classification (3-class) under the official classification of BRACS dataset (%)). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB-MIL <sup>[43]</sup>               | 58.05±1.44        | 80.68±0.35        | 55.83±1.97        |
| CLAM <sup>[10]</sup>                 | 57.85±0.94        | 83.54±0.69        | 58.41±1.08        |
| DSMIL <sup>[9]</sup>                 | 55.21±0.85        | 82.54±0.57        | 51.67±0.75        |
| TransMIL <sup>[11]</sup>             | 54.17±2.55        | 79.32±0.31        | 50.68±2.40        |
| DTFD-MIL <sup>[12]</sup>             | 58.22±6.26        | <u>85.96±2.13</u> | <u>59.70±2.86</u> |
| MHIM-MIL <sup>[60]</sup>             | 57.78±1.28        | 81.03±0.13        | 54.74±1.56        |
| IBMIL <sup>[72]</sup>                | 58.30±3.58        | 80.07±2.42        | 58.27±4.88        |
| SRMamba <sup>[52]</sup>              | <u>58.38±5.21</u> | 83.12±3.90        | 57.87±6.94        |
| WiKG <sup>[73]</sup>                 | 51.74±2.46        | 76.71±1.80        | 48.71±2.31        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 56.02±2.36        | 82.11±0.54        | 55.58±3.65        |
| <b>M<sup>2</sup>S-MIL(Ours)</b>      | <b>59.38±1.13</b> | <b>86.21±0.46</b> | <b>60.13±1.79</b> |

表 3.5 展示了 BRACS 数据集上依照官方划分的粗粒度亚型分类性能。结果表明, 与基线相比, 本章的方法仍然具有优势。例如, 本章的方法在这任务上分别将

次优的方法中的 Accuracy 和 AUC 提高了 1.0% 和 0.25%，对比 SRMamba，AUC 更是提高了 2.09%。这些结果都清楚地表明，本章的方法的显著性能改进并不仅仅归功于 Mamba 本身强大的建模能力，并且再次证实了本章的方法充分引入了空间关系和视觉特性，帮助模型更好的捕获和建模实例间的关系，大大提升模型性能。

表 3.6 以 UNI 为离线特征提取器的癌症诊断和分型性能。最优结果用粗体表示。

Table 3.6 Cancer diagnosis and Sub-typing performance with UNI as the offline feature extractors.

The best results are shown in bold.

| 方法  | CAMELYON                  | NSCLC             | BRCA              |
|-----|---------------------------|-------------------|-------------------|
| UNI | ABMIL <sup>[43]</sup>     | 99.36±0.40        | 98.24±0.59        |
|     | Mamba <sup>[16]</sup>     | 99.44±0.47        | 98.45±0.53        |
|     | <b>M<sup>2</sup>S-MIL</b> | <b>99.51±0.22</b> | <b>98.55±0.56</b> |

由于使用 UNI 特征提取器进行诊断和分型任务的基线方法表现出足够强的性能，使得进一步的比较没有意义（见表3.6），因此本文仅关注 UNI 特征在生存预测的影响。

## (2) 生存预测

表 3.7 M<sup>2</sup>S-MIL 在以 R50 为特征提取器的三个主要数据集上的生存预测结果 (%)。

最优结果用粗体表示，次优结果使用下划线表示。

Table 3.7 Survival Prediction results on three main datasets extracted by R50(%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | BLCA              | LUAD              | LUSC              |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB - MIL <sup>[43]</sup>             | 57.97±5.43        | 60.56±5.59        | 59.11±2.47        |
| CLAM <sup>[10]</sup>                 | 58.73±3.24        | 60.45±5.94        | 60.71±2.91        |
| DSMIL <sup>[9]</sup>                 | 58.29±3.53        | 59.78±5.53        | 60.21±3.90        |
| TransMIL <sup>[11]</sup>             | 57.00±3.60        | 63.06±2.97        | 57.46±4.69        |
| DTFD-MIL <sup>[12]</sup>             | 59.85±4.18        | 60.65±5.62        | 59.03±2.54        |
| MHIM-MIL <sup>[60]</sup>             | 59.70±4.06        | 63.26±1.00        | <u>60.91±4.26</u> |
| IBMIL <sup>[72]</sup>                | 59.09±5.42        | 60.69±5.71        | 58.50±2.13        |
| SRMamba <sup>[52]</sup>              | 60.94±2.19        | 61.47±4.29        | 58.35±3.95        |
| WiKG <sup>[73]</sup>                 | 58.04±3.30        | 62.77±4.43        | 58.10±4.95        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | <u>60.98±2.73</u> | <u>63.28±3.75</u> | 60.64±3.83        |
| <b>M<sup>2</sup>S - MIL</b>          | <b>61.96±1.28</b> | <b>63.88±6.92</b> | <b>61.22±4.30</b> |

表 3.8 M<sup>2</sup>S-MIL 在以 UNI 为特征提取器的三个主要数据集上的生存预测结果 (%)。

最优结果用粗体表示, 次优结果使用下划线表示。

Table 3.8 Survival Prediction results on three main datasets extracted by UNI(%)(%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                     | BLCA              | LUAD              | LUSC              |
|--|-------------------|-------------------|-------------------|
| AB - MIL <sup>[43]</sup>               | 59.42±5.96        | 63.37±3.65        | 59.25±4.74        |
| CLAM <sup>[10]</sup>                   | 59.68±4.72        | <u>65.64±3.69</u> | 59.69±2.88        |
| DSMIL <sup>[9]</sup>                   | 60.83±4.08        | 64.41±2.30        | 60.58±4.25        |
| TransMIL <sup>[11]</sup>               | 59.14±5.34        | 61.97±5.17        | 58.78±3.54        |
| DTFD - MIL <sup>[12]</sup>             | 59.64±6.28        | 65.19±1.56        | 60.27±5.39        |
| MHIM - MIL <sup>[60]</sup>             | 60.78±3.71        | 64.88±0.71        | <u>61.43±3.77</u> |
| IBMIL <sup>[72]</sup>                  | 59.33±4.02        | 64.56±4.24        | 61.12±5.10        |
| SRMamba <sup>[52]</sup>                | <u>61.46±3.00</u> | 64.82±5.05        | 59.30±5.69        |
| WiKG <sup>[73]</sup>                   | 59.53±3.12        | 63.93±3.87        | 60.32±3.25        |
| R <sup>2</sup> T - MIL <sup>[13]</sup> | 59.42±3.15        | 64.69±2.67        | 61.06±3.85        |
| <b>M<sup>2</sup>S - MIL</b>            | <b>61.74±4.80</b> | <b>66.17±3.77</b> | <b>62.05±4.30</b> |

表 3.7 和表 3.8 分别给出了三个生存预测数据集以 R50 提取特征和以 UNI 提取特征的实验结果。本章提出的 M<sup>2</sup>S-MIL 模型表现出很强的性能, 在 BLCA 数据集上的 C-index 得分为 61.96%, 在 LUAD 上的得分为 63.88%, 在 LUSC 上的得分为 61.22%。其明显优于比较的方法, 在每个数据集上分别比次优方法提高了 0.98%, 0.68% 和 0.31%。此外, 即使在利用基础模型提取的高质量特征时, M<sup>2</sup>S-MIL 也会产生较好的效果, 同样达到 SOTA 的性能效果。特别是, 与 SRMamba 相比, 它在 LUAD 和 LUSC 数据集上的性能分别提高了 1.35% 和 2.65%。这些结果强调了本章提出的策略的一致性和可靠性, 强调了其在生存预测结果方面的有效性。

### 3.4.3 消融实验

#### (1) 不同模块的有效性分析

为了研究 M<sup>2</sup>S-MIL 模型中每个模块对模型的影响, 本文在 CAMELYON、TCGA-NSCLC 和 TCGA-BRCA 三个数据集上进行了消融实验, 此部分实验分为六个模型, 分别为基准模型 (Origin Mamba), 只添加网格扫描分支的基准模型 (+Grid), 只添加区域螺旋扫描的基准模型 (+RSS.), 应用多扫描分支的基准模型 (+Multi-scan.), 只应用 2D 空间上下文感知模块的基准模型 (+2D-SCA.) 以及最终模型 (+Multi-scan.+2D-SCA.)。

在三个数据集上的模块消融实验如表 3.9 所示。OriginMamba 意味着简单地将 Mamba 整合到 ABMIL 中。Grid 表示添加了网格扫描模式; RSS 表示只添加了区域螺旋扫描模式分支; +Multi-scan. 表示应用了多扫描模式分支; +2D-SCA. 表示应用 2D 空间上下文模块; +MS.+2D-SCA. 表示同时运用上述两个模块, 即最终模型。

表 3.9 在三个主要数据集上，不同组件对性能的影响。最优结果用粗体表示。

Table 3.9 Module ablation experiments of M<sup>2</sup>S-MIL on CAMELYON, NSCLC and BRCA. The best results are shown in bold.

| 方法                    | CAMELYON          | NSCLC             | BRCA              |
|-----------------------|-------------------|-------------------|-------------------|
| Origin Mamba          | 91.57±1.37        | 95.03±1.89        | 92.56±2.11        |
| +Grid.                | 92.11±1.53        | 95.64±1.97        | 92.71±1.90        |
| +RSS.                 | 92.30±1.12        | 95.66±1.45        | 92.89±1.02        |
| +Multi - scan.        | 92.72±1.82        | 96.25±1.44        | 93.01±1.82        |
| +2D - SCA.            | 91.78±1.21        | 95.43±1.26        | 92.89±1.65        |
| <b>+MS.+2D - SCA.</b> | <b>93.02±1.20</b> | <b>96.65±1.08</b> | <b>93.50±1.70</b> |

本章的基线指的是将原始的 Mamba 架构合并到 ABMIL 中。首先，可以看到分别加入网格扫描模式和区域螺旋扫描模式的分支时，都有一定的性能提升，而将多扫描模式分支输入集成到基线模型中时，该模型在 CAMELYON、NSCLC 和 BRCA 数据集上的 AUC 分别提高了 1.15%、1.22% 和 0.45%，表明提出综合多种扫描模式的 Mamba 能够更好的建模实例间的 2D 空间关系，从而更适配图像任务。实验结果表明，与 OriginMamba 相比，2D 空间上下文感知模块的引入分别增强了 Mamba 对序列内实例间的空间信息捕获能力。将这两个模块引入 OriginMamba 后，完整的 M<sup>2</sup>S-MIL 取得了最好的性能（在 CAMELYON 上 AUC 为 93.02%，在 TCGA 上 AUC 为 96.65%，在 BRCA 上 AUC 为 93.50%）。

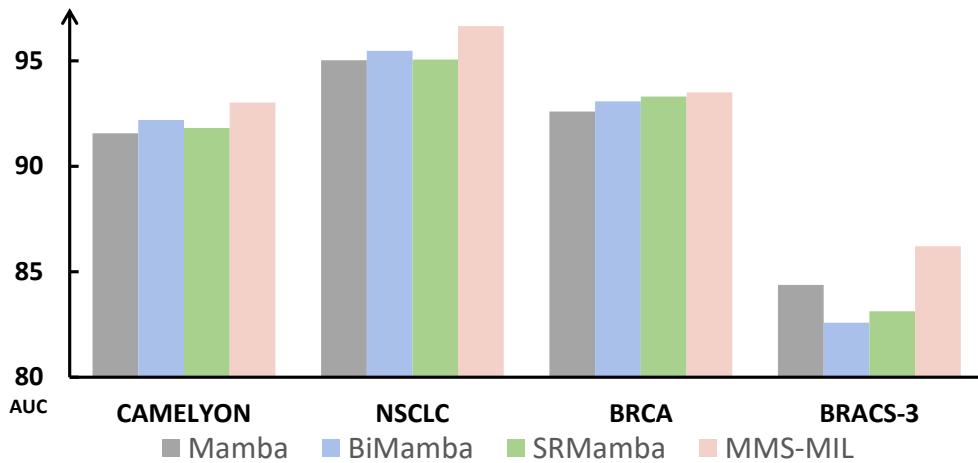


图 3.3 不同 Mamba 模型的性能比对柱状图。

Fig. 3.3 Bar chart of performance comparison of different Mamba models.

## (2) 不同 Mamba 模型性能比对

为了进一步验证本章的方法的优越性来源于本章的架构而不是 Mamba 本身，

本章对三种变体进行了比较实验：图3.3中的 Mamba Block、BiMamba 和 SRMamba。结果表明， $M^2S\text{-MIL}$  在所有四个数据集上都取得了最佳性能。本章的方法不仅在性能上超过了原始的单分支 Mamba，而且在与没有引入局部连续性和旋转不变性等视觉图像的聚合方法（例如，BiMamba 和 SRMamba）的比较中表现出许多优势。实验结果表明，本章的引入 2D 特性的 Mamba 比其他多扫描综合的 Mamba 算法具有更强的实例建模能力，更强的空间信息捕获能力，能够提取更多的判别性特征。

### （3）讨论区域数量对模型性能的影响

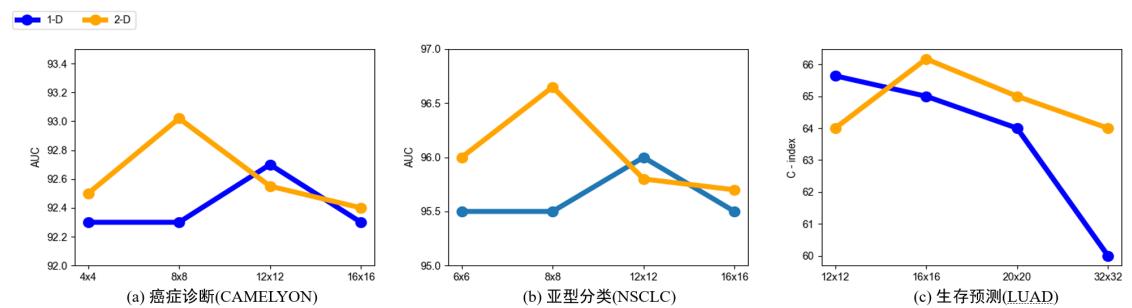


图 3.4 不同区域数量的模型性能折线图。

Fig. 3.4 Breakpoint plots of model performance for different number of regions.

本小节系统地研究了不同区域划分策略在本章的方法中的影响。图 3.4展示了不同区域划分在病理分类任务上的性能效果，其中 1-D 代表并不将之重塑为 2D 特征图而只进行一维序列的分区。从与 1D 的对比实验可以看出，二维区域划分方式优于一维区域划分方式，因为它保留了更多的原始图像的空间信息。并且本章认为，本章的性能提升并不仅仅依赖于对序列简单的划分区域，侧面印证了本章方法，通过引入更多空间信息从而提高了模型性能。此外，另一个现象是，使用过小或过大的区域会使性能下降。这是因为小区域大幅降低了 Mamba 建模二维关系的能力（近似正常原始序列输入），而过大的区域将原本相距较远不具备连续性的特征关联在一起导致性能损耗。因此，采用适中的区域划分是多扫描 Mamba 输入的最佳选择。

### （4）窗口大小对 2D 空间上下文感知模块的影响分析

表 3.10 不同大小窗口对性能影响。最优结果用粗体表示。

Table 3.10 Effect of different window sizes on performance. The best results are shown in bold.

| 方法           | CAMELYON                           | NSCLC                              | BRCA                               | BRACS3                             |
|--------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| W/o          | $92.72 \pm 1.82$                   | $96.25 \pm 1.44$                   | $93.01 \pm 1.82$                   | $85.24 \pm 0.32$                   |
| $3 \times 3$ | $92.78 \pm 1.52$                   | $96.31 \pm 1.33$                   | $93.37 \pm 1.93$                   | $85.71 \pm 1.80$                   |
| $5 \times 5$ | <b><math>93.02 \pm 1.20</math></b> | <b><math>96.65 \pm 1.08</math></b> | <b><math>93.50 \pm 1.70</math></b> | <b><math>86.21 \pm 0.46</math></b> |

本小节探究了不同窗口大小的 2D 空间上下文感知模块对 M<sup>2</sup>S-MIL 的性能影响。可以看到越大窗口范围可以整合更多的空间信息，所带来的性能提升越大。但是由于计算窗口注意力所需要的计算资源随窗口大小增大而增大，所以平衡计算资源与性能之间选择适中的窗口大小是 M<sup>2</sup>S-MIL 的最佳选择。

### (5) 使用 PLIP 作为特征提取器的性能结果

表 3.11 M<sup>2</sup>S-MIL 以 PLIP 为离线特征提取器的癌症诊断性能。最优结果用粗体表示，次优结果用下划线表示。

Table 3.11 Cancer diagnosis performance of M<sup>2</sup>S-MIL with PLIP as the offline feature extractors. The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法   | Accuracy                    | AUC               | F1 - score        |
|------|-----------------------------|-------------------|-------------------|
| PLIP | AB - MIL                    | 90.09±1.63        | <u>94.55±1.91</u> |
|      | OriginMamba                 | 90.87±1.52        | 94.44±2.19        |
|      | BiMamba                     | 90.54±1.89        | 94.20±1.99        |
|      | SRMamba                     | <u>91.87±0.85</u> | 94.20±2.16        |
|      | <b>M<sup>2</sup>S - MIL</b> | <b>92.21±0.60</b> | <b>95.24±1.85</b> |

表 3.12 M<sup>2</sup>S-MIL 以 PLIP 为离线特征提取器在 TCGA-NSCLC 上的亚型分型性能。最优结果用粗体表示，次优结果用下划线表示。

Table 3.12 Cancer sub-typing performance of M<sup>2</sup>S-MIL on the TCGA-NSCLC dataset with PLIP as the offline feature extractors. The optimal experimental results are marked in bold, and the sub-optimal experimental results are underlined.

| 方法   | Accuracy                    | AUC               | F1 - score        |
|------|-----------------------------|-------------------|-------------------|
| PLIP | AB - MIL                    | 89.94±1.84        | 94.45±1.64        |
|      | OriginMamba                 | <u>90.89±1.60</u> | 95.57±1.60        |
|      | BiMamba                     | 90.70±1.95        | 95.42±1.68        |
|      | SRMamba                     | 89.85±2.07        | <u>95.72±1.51</u> |
|      | <b>M<sup>2</sup>S - MIL</b> | <b>91.56±1.76</b> | <b>96.22±1.12</b> |

为了进一步探索本章方法的有效性，本小节还进行了将同样采用自监督预训练的 PLIP<sup>[74]</sup> 作为特征提取器的相关实验，以验证本章方法的普遍性。如表3.11、3.12和表3.13所示，可以看出实验结果与 R50 特征一致。本章的方法优于其他基于 Mamba 的方法。在 CAMELYON 数据集上，它在 Accuracy、AUC 和 F1-score 方面分别比次优的方法提高了 0.34%、0.69% 和 1.09%。这些数据在 NSCLC 数据集中分别为 0.57%，0.4% 和 0.4%。最后，BRCA 数据集的改进也值得注意，Accuracy

表 3.13  $M^2S\text{-MIL}$  以 PLIP 为离线特征提取器在 TCGA-BRCA 上的亚型分型性能。最优结果用粗体表示，次优结果用下划线表示。

Table 3.13 Cancer sub-typing performance of  $M^2S\text{-MIL}$  on the TCGA-BRCA dataset with PLIP as the offline feature extractors. The optimal experimental results are marked in bold, and the sub-optimal experimental results are underlined.

| 方法   | Accuracy                            | AUC                                | F1-score                           |
|------|-------------------------------------|------------------------------------|------------------------------------|
| PLIP | AB-MIL                              | $83.81 \pm 3.72$                   | $90.18 \pm 2.99$                   |
|      | OriginMamba                         | $86.75 \pm 5.76$                   | $92.90 \pm 2.44$                   |
|      | BiMamba                             | <u><math>87.73 \pm 3.06</math></u> | $92.64 \pm 2.28$                   |
|      | SRMamba                             | $86.40 \pm 3.31$                   | $92.17 \pm 2.67$                   |
|      | <b><math>M^2S\text{-MIL}</math></b> | <b><math>88.82 \pm 3.22</math></b> | <b><math>93.40 \pm 1.24</math></b> |

提高了 1.09%，AUC 提高了 0.5%，F1-score 提高了 0.65%。这进一步表明，即使使用非传统的 ResNet-50 特征提取器，本章方法的性能改进也不仅仅是依赖于简单的多个 Mamba 分支合并，而是正确捕获了空间信息与视觉特征。

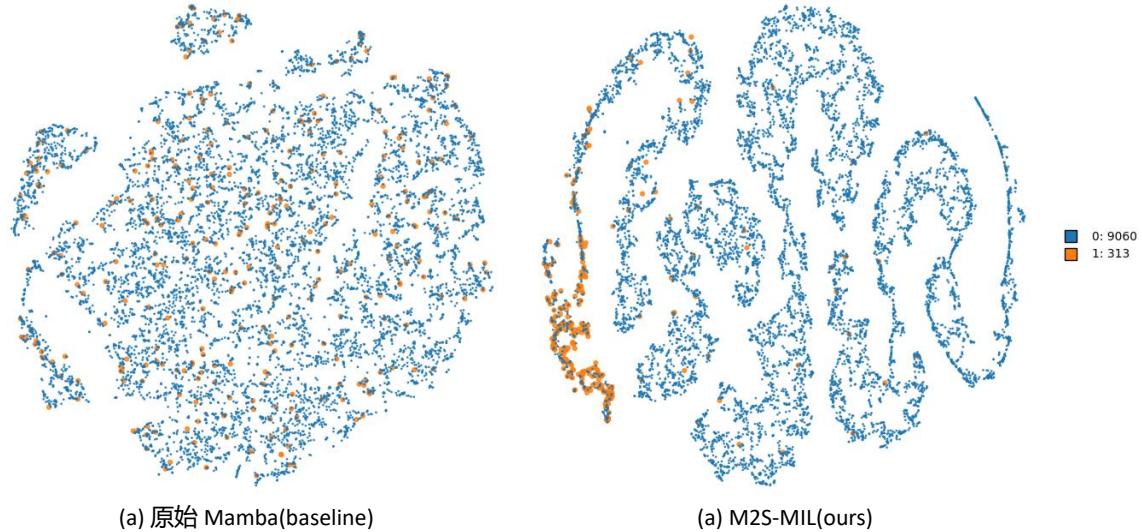


图 3.5 原始 Mamba(baseline) 和  $M^2S\text{-MIL}$  聚合特征 t-SNE 可视化结果。

Fig. 3.5 The t-SNE visualization of aggregated features from Mamba(baseline) and SMC-MIL.

#### 3.4.4 可视化分析

为了直观地验证本章方法的可解释性，本章用 t-SNE 可视化了同一个 WSI 切片的所有实例经过 Mamba 和  $M^2S\text{-MIL}$  所建模出后的特征，结果展示在图3.5。可以看到  $M^2S\text{-MIL}$  更具有条理，而原始 Mamba 的结果更为混乱。说明  $M^2S\text{-MIL}$  使 Mamba 的实例间的关系建模能力更为出色。

并且本章将多种 Mamba 变体与本章的方法在 Camelyon-16 数据集上产生的高

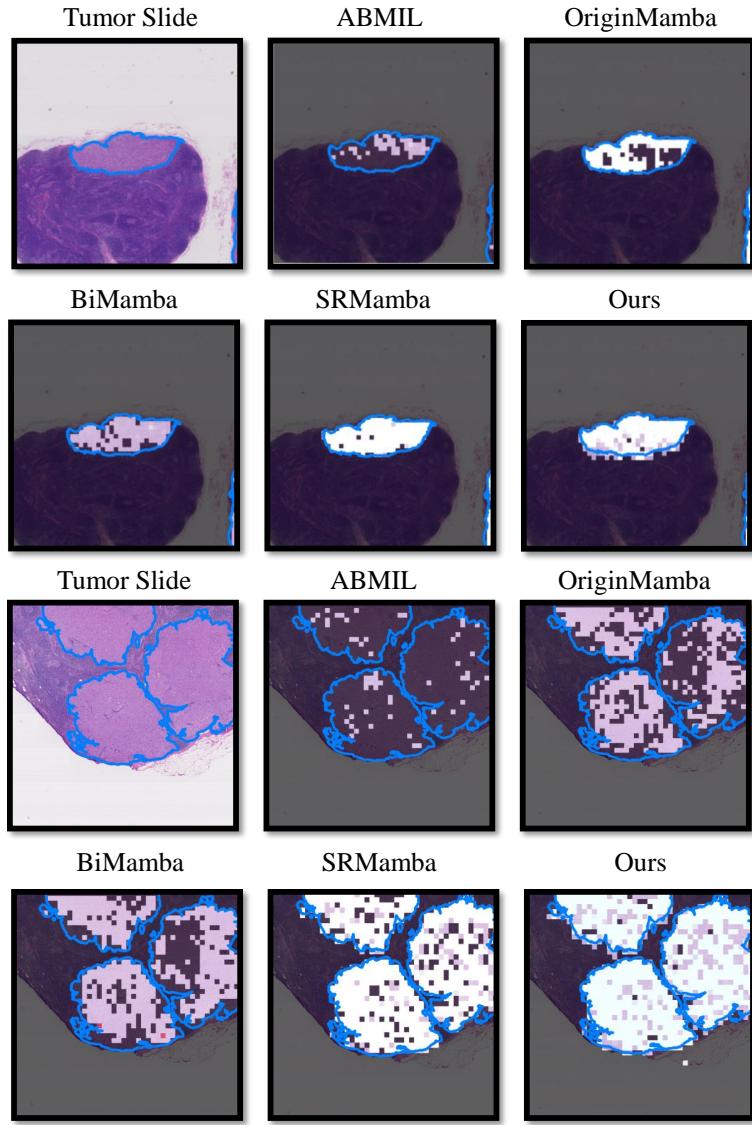
图 3.6 多个 Mamba 变体和 M<sup>2</sup>S-MIL 关注的图块可视化结果

Fig. 3.6 Visualization results of patches focused by other Mamba variants(baseline) and SMC-MIL.

注意力分数的实例可视化，如图3.6所示。蓝色的线条勾勒出肿瘤区域。越亮的斑块表明注意力得分越高。可视化结果清楚地表明，与原始 Mamba 相比，M<sup>2</sup>S-MIL 能够更准确地选择显著斑块。

### 3.5 本章小结

本章针对序列关系建模模型 Mamba 在处理数字病理图像分类任务时，对视觉任务特性和空间关系捕捉不充分的问题，提出了一种简单有效的解决思路，并提出一种新的多实例学习方法：基于空间信息增强的多扫描 Mamba 的数字病理图像分类（Multi-scan Mamba with Spatial information enhancement，简称 M<sup>2</sup>S-MIL）。该方法通过综合多种扫描模式的信息，针对病理图像空间信息，设计了区域螺旋扫

描和网格扫描模式，增强 Mamba 对病理图像空间关系的建模。同时利用 2D 空间上下文感知模块，进一步对整合后的序列进行空间关系增强，提升模型性能。在 3 个 WSI 子任务以及 7 个数据集上的大量对比及消融实验充分证明了本章提出的基于空间信息增强的多扫描 Mamba 的多实例学习架构综合了多种扫描模式分支，引入空间信息，解决了数字病理图像分类任务中使用 Mamba 模型存在的图像空间结构捕获能力不足的问题。

## 4 基于扫描不变性 Mamba 的高分辨率病理图像分类研究

上一章研究了基于多扫描 Mamba 的高分辨率病理图像分类算法，在保留 Mamba 高效的建模能力同时，通过引入 2D 图像空间结构信息，增强 Mamba 对实例间空间结构关系的建模能力，并且通过集成多扫描分支的方式缓解了 Mamba 对顺序的依赖，提高了模型性能。然而，该算法仅关注于输出序列信息的整合，未能从根本上突破序列输入顺序对 Mamba 建模能力的约束，仅仅是一种妥协的策略，依旧高度依赖扫描模式与集成方式。因此，在上一章的基础上，本章主要研究基于扫描不变性 Mamba 的高分辨率病理图像分类算法，通过构建成对的差异化实例序列，并在此基础上开展对比学习，促进 Mamba 模型忽略序列顺序，挖掘并聚合关键实例特征。本章内容共分为四节，第一节介绍本章的研究动机；第二节介绍本章提出的基于扫描不变性 Mamba 的高分辨率图像分类算法；第三节给出实验设置和结果分析；第四节对本章进行小结。

### 4.1 研究动机

在经典的 MIL 范式中，使用预训练模型将实例转化成离线特征，然后聚合成包级别的特征表示，以进行后续分析。在这个范式中，由于病理图像的高分辨率特性，导致其划分后的实例数目过多，因此 WSI 分类可以被视为一个长序列建模问题，通过将包中的大量实例视为标记来处理，能够建模实例与整个包之间的空间上下文相关性，以捕捉判别性信息。一些基于状态空间模型（State Space Model, SSM）的方法被设计来有效对长序列进行建模。Mamba 是最近提出的一种具有代表性的基于 SSM 架构的模型，与 ViT 相比具有更好的长序列建模性能，但需要的计算资源更少<sup>[18]</sup>。然而，在 Mamba 模型和 MIL 任务之间存在一个根本的区别：如图 1.3 所示，Mamba 作为扫描序列敏感模型，其性能高度依赖实例的输入顺序，这与多实例学习的序列不变性要求存在本质冲突，导致模型在病理图像分类任务中易受随机序列干扰而产生性能波动。

为了调和扫描敏感的 Mamba 和扫描不变的 MIL 之间的差异，以前的工作包括上一章的方法都主要是将不同扫描模式的图像 Patch 序列输入到 Mamba 中，然后将多个输出序列聚合。例如，MamMIL<sup>[51]</sup> 结合了 Bi-SSM 和 2D-CAB 模块以整合来自原始序列、反向序列和局部序列的扫描模式信息。而 MambaMIL<sup>[52]</sup> 将序列重塑为矩形，并应用序列重新排序操作来垂直扫描新序列。然而，用于生成实例序列的扫描模式数量有限，限制了模型对扫描模式不变特征的学习能力。因此，这并没有从根本上减轻扫描模式对 Mamba 性能的影响，仅仅是一种妥协的策略，其依旧高度依赖扫描模式和集成形式。

## 4.2 基于序列差异化对比的 Mamba 模型

### 4.2.1 方法概述

虽然 Mamba 在实例特征学习和对大包（长序列）的处理效率上具有突出的优势，但它是扫描敏感的，这与 MIL 的假设不一致。在 MIL 中，Mamba 的输出不应该依赖于实例输入顺序，也不应该对实例数量敏感。换句话说，MIL 对于实例输入是扫描不变的。在这种情况下，本章精心设计了一个基于对比学习的扫描不变 Mamba 模型，旨在更好地将先进的 Mamba 模型适应于 MIL 任务。

如图 4.1 所示，本章的模型由两个 Mamba 分支组成。一个是“Easy”的分支，而另一个是“Hard”的分支。在“Easy”分支中，对高度关注的实例通过 Mixup 生成越来越多的突出实例，从而降低包分类的难度。而在“Hard”分支中，将容易分类的实例进行掩码，以提高包分类的难度。显然，这两个分支在扫描顺序、实例数目和包分类难度上存在较大差异。因此，本章方法采用对比学习强制 Mamba 对这两个不同序列输出相同的结果，从而赋予 Mamba 这样一个扫描不变的属性。此外，极高的分支差异也使得 Mamba 通过对比学习进一步增强了泛化能力。

本研究的主要贡献可以概括为以下几点：

- 本研究提出了一个基于 Mamba 的多实例学习框架，命名为 SMC-MIL，用于计算病理学。SMC-MIL 通过实例序列对比学习来引导 Mamba 学习扫描不变的包表示，这更适合基于 MIL 的 WSI 分类范式。
- 设计了一个双分支实例序列对比学习框架，旨在实现两个极端不同的输入实例序列之间的一致性，消除序列排列对最终包决策的影响。该框架为使用 SSM 处理 MIL 任务提供了一个简单而有效的解决方案。
- 提出了一种序列对比度增强方法，该方法基于实例分类困难程度的评估，而有意地打乱、增强和掩盖实例。该方法不仅增加了两个输入对比序列在顺序和长度上的差异，而且扩大了识别判别区域的难度差异。显然，这种方法增加了对比学习的难度，从而迫使 Mamba 在不考虑其图块序列排列的情况下学习 WSI 的判别特征。

### 4.2.2 预处理与特征提取

遵从文献<sup>[10]</sup>的设置，每张病理图像被裁剪成一系列  $256 \times 256$  的非重叠补丁，放大倍数为 20 倍，同时丢弃背景区域（饱和度  $< 15$ ）。本章采用 ResNet-50<sup>[70]</sup> 或 UNI<sup>[69]</sup> 作为底层特征提取器  $\mathcal{F}(\cdot)$ 。因此，本章定义原始扫描序列为：

$$Z = \{z_i\}_{i=1}^N, \quad (4.1)$$

其中  $z_i$  是  $X$  的第  $i$  个实例的提取特征， $N$  是实例的个数。另一个扫描序列定义为  $Z' = \text{Resort}(Z)$ （即  $Z'$  通过对第一个扫描序列的顺序进行随机重排获得， $\text{Resort}(\cdot)$  表示随机重排操作）。

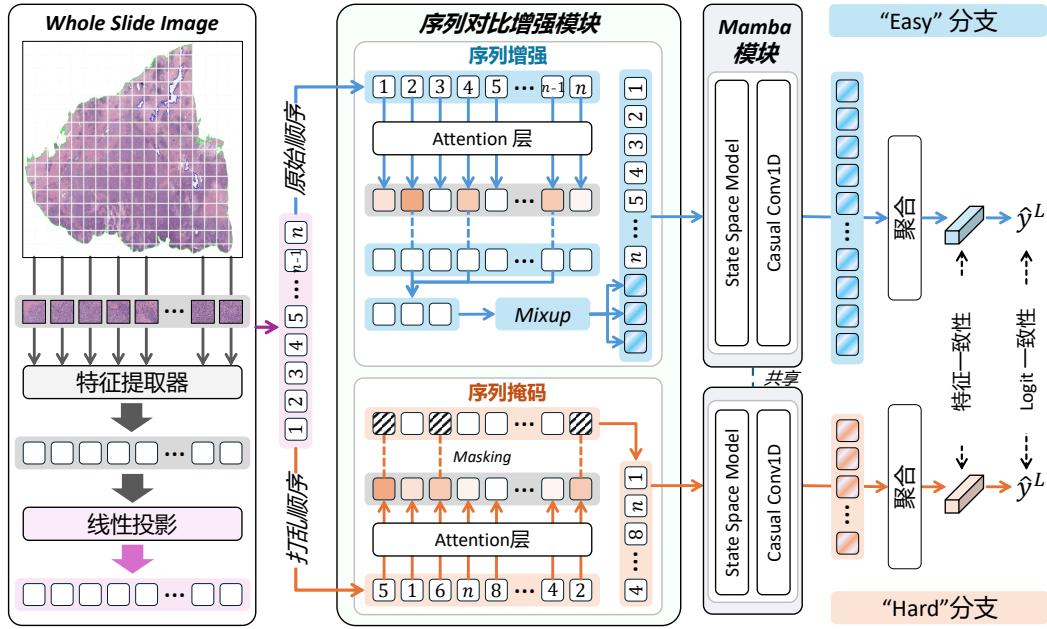


图 4.1 SMC-MIL 模型总览图。

Fig. 4.1 Overview of proposed SMC-MIL.

#### 4.2.3 基于 Mamba 的双分支对比架构

通过进行序列对比学习，引导 Mamba 学习扫描不变的包特征。在该方法中，本章设计了一个双分支实例序列对比学习框架，旨在实现两个极端不同的输入实例序列之间的一致性，并消除序列排列对最终包决策的影响。

假设差异化序列对比增强模块为  $S(\cdot)$ ，则通过差异化序列生成模块可得到两个极端不同的序列：

$$\begin{aligned} Z_1 &= S(Z), \\ Z_2 &= S(Z'). \end{aligned} \tag{4.2}$$

给定一个 Mamba 模块  $\Psi(\cdot)$  与一个注意力聚合模块  $\phi(\cdot)$ ，分别对实例序列进行实例关键建模和特征聚合，为了方便后续表示，该聚合过程记作  $\rho(\cdot)$ ，统称为 Mamba 聚合器。根据上述步骤，则差异化的两个序列分支的包特征分别为：

$$\begin{aligned} \mathbf{b}_1 &= \rho(Z_1), \\ \mathbf{b}_2 &= \rho(Z_2). \end{aligned} \tag{4.3}$$

这两个特征都是同一个病理图像的有效表示，因此应该反映一致的视觉内容，本章称之为特征一致性。

给定一个分类器  $\mathcal{C}(\cdot)$ , 分别对这些特征进行分类, 获取其分类标签:

$$\begin{aligned}\hat{Y}_1 &= \mathcal{C}(\mathbf{b}_1), \\ \hat{Y}_l &= \mathcal{C}(\mathbf{b}_2).\end{aligned}\tag{4.4}$$

同样的, 这两个标签应该反映一致分类结果, 本章称之为标签一致性。

通过特征一致性和标签一致性的约束, 反向优化引导 Mamba 架构学习到与扫描无关的包特征, 从根本上突破 Mamba 架构对输入顺序的依赖, 使 Mamba 架构更适配高分辨率病理图像分类任务。

#### 4.2.4 差异化序列生成 (简单与困难分支)

基于扫描序列的顺序、长度以及发现判别区域的难易程度的三个角度, 本章提出了一种序列对比度增强方法进行差异化序列生成, 即构造“Easy”分支和“Hard”分支。根据所在分支的不同, 上小节提到的序列对比增强模块  $S(\cdot)$ , 可分为两个子模块: 序列增强模块  $S_A(\cdot)$  和序列掩码模块  $S_M(\cdot)$ 。下面分别详细介绍这两个部分。

##### 序列增强 (“Easy” 分支)

如图 4.1 顶部所示, 本章设计了一个基于注意力的序列增强模块来构建“Easy”分支。其核心思想是使用注意力机制识别高关注实例, 然后通过 Mixup 对这些已识别的实例执行实例级增强。此过程生成新实例来补充原始序列。本章将每个实例的注意力得分定义为:

$$A = [a_1, \dots, a_i, \dots, a_N] = \phi(Z),\tag{4.5}$$

其中  $\phi(\cdot)$  为注意力模块。然后将其进行排序:

$$I = [i_1, i_2, \dots, i_N] = Sort(A),\tag{4.6}$$

其中  $i_1$  为注意力得分最高实例的索引,  $i_N$  为得分最低实例的索引。本章将实例的  $\alpha\%$  作为进行序列增强的实例。其对应的索引是  $\lfloor \alpha\% \times N \rfloor$ , 其对应的实例注意力分数是  $I_h = [i_t]_{t=1}^{\lfloor \alpha\% \times N \rfloor}$ 。则增强序列可表示为:

$$Z_{easy} \leftarrow S_A(Z) := concat\{\varphi(Z_{I_h}), Z\},\tag{4.7}$$

其中  $\varphi(\cdot)$  是所选实例上的 Mixup 操作,  $S_A(\cdot)$  是序列增强模块,  $concat$  则代表将进行 Mixup 后的实例同样拼接到原本实例序列上。注意, 这里的注意力机制与后续聚合模块中的注意力模型共享参数。因此, 在“Easy”分支中输入 Mamba 的序列长度为  $\lfloor \alpha\% \times N \rfloor + N$ 。

##### 序列掩码 (“Hard” 分支)

如图 4.1 底部所示，本章采用基于注意力的掩蔽机制来掩盖“Hard”分支中容易区分的实例，迫使 Mamba 增强其鉴别可判别性区域的能力。对于分类模型，本章认为具有最高和最低关注值的实例是最容易区分的。因此，本章试图掩盖这些实例，以迫使模型学习更精确的分类边界。具体来说，本章生成一个  $n$  维掩码向量  $M$  如下所示：

$$M = [m_1, \dots, m_i, \dots, m_N], \quad (4.8)$$

其中  $m_i \in \{0, 1\}$ 。如果  $m_i = 1$ ，则第  $i$  个实例被保留； $m_i = 0$  时，第  $i$  个实例被掩码。本章将高关注值和低关注值的掩码率分别设置为  $\beta_h\%$  和  $\beta_l\%$ 。那么，高值对应的序列下标为  $I_{\beta_h} = [i_t]_{t=1}^{\lfloor \beta_h \% \times N \rfloor}$ ，低值对应的序列下标为  $I_{\beta_l} = [i_t]_{t=N-\lfloor \beta_l \% \times N \rfloor}^N$ 。因此，

$$m_i = \begin{cases} 0 & i \in I_{\beta_h} \cup I_{\beta_l}, \\ 1 & \text{other.} \end{cases} \quad (4.9)$$

最后，掩码后的序列可以表示为：

$$Z_{hard} \leftarrow S_M(Z') := M \cdot Z', \quad (4.10)$$

其中  $S_M(\cdot)$  是序列掩码模块。为了便于讨论，本文假设  $\beta_h = \beta_l$ 。

“Easy” 分支保留了原本的序列顺序，同时增加了序列长度，并且通过 Mixup 操作添加了多样性，降低了 Mamba 捕获可判别性区域的难度；“Hard” 分支打乱了原本的序列顺序，同时缩减了序列长度，并且通过掩盖掉易于分辨的实例，提高了 Mamba 捕获可判别性区域的难度。上下两个分支形成了鲜明的对比。如此差异化的序列分支，加大了对比学习任务的难度，迫使 Mamba 学习到更清晰的分类边界，以提高模型鲁棒性。

#### 4.2.5 基于 Mamba 的实例聚合

本小节介绍基于 Mamba 的实例聚合，即上文所提到的 Mamba 聚合器  $\rho(\cdot)$ 。如上所述，“Easy” 和 “Hard” 分支的输入实例序列分别为  $Z_{easy}$  和  $Z_{hard}$ 。 $\Psi(\cdot)$  是上文提到的一个 Mamba 块。这两个分支的包表示可以通过执行基于注意力机制的聚合来获得，

$$\begin{aligned} \mathbf{b}_1 &= A_1 \cdot \Psi(Z_{easy}) = \phi(\Psi(S_A(Z))) \cdot \Psi(S_A(Z)), \\ \mathbf{b}_2 &= A_2 \cdot \Psi(Z_{hard}) = \phi(\Psi(S_M(Z'))) \cdot \Psi(S_M(Z')), \end{aligned} \quad (4.11)$$

其中  $A_1$  和  $A_2$  分别是  $\Psi(Z_{easy})$  和  $\Psi(Z_{hard})$  的注意力分数。类似地，它们的预测包标签可以通过将包的表示输入分类器  $\mathcal{C}(\cdot)$  得到，

$$\begin{aligned} \hat{Y}_1 &\leftarrow \mathcal{C}(\mathbf{b}_1) := \mathcal{C}(A_1 \cdot \Psi(Z_{easy})), \\ \hat{Y}_2 &\leftarrow \mathcal{C}(\mathbf{b}_2) := \mathcal{C}(A_2 \cdot \Psi(Z_{hard})). \end{aligned} \quad (4.12)$$

#### 4.2.6 基于一致性的模型优化

由于这两个分支的扫描序列都来自同一个包，因此它们的包表示和分类应该是一致的。它们的包表示的一致性损失为：

$$\mathcal{L}_{b_{con}} = -\frac{\text{Softmax}(\mathbf{b}_1) \log(\mathbf{b}_2) + \text{Softmax}(\mathbf{b}_2) \log(\mathbf{b}_1)}{2}. \quad (4.13)$$

而它们的包分类的一致性损失为：

$$\mathcal{L}_{l_{con}} = -\frac{\text{Softmax}(\hat{Y}_1) \log(\hat{Y}_2) + \text{Softmax}(\hat{Y}_2) \log(\hat{Y}_1)}{2}. \quad (4.14)$$

除了一致性损失之外，本章还向“Easy”分支引入了分类损失，以确保预测标签与其真实值之间的一致性，

$$\mathcal{L}_{cls} = Y \log \hat{Y}_1 + (1 - Y) \log(1 - \hat{Y}_1), \quad (4.15)$$

其中  $Y$  是真实值。综上所述，该模型可通过处理以下优化问题得到：

$$\{\hat{\Psi}, \hat{\phi}, \hat{\mathcal{C}}\} \leftarrow \arg \min_{\hat{\Psi}, \hat{\phi}, \hat{\mathcal{C}}} \mathcal{L} = \mathcal{L}_{cls} + \gamma(\mathcal{L}_{b_{con}} + \mathcal{L}_{l_{con}}), \quad (4.16)$$

其中  $\hat{\Psi}, \hat{\phi}, \hat{\mathcal{C}}$  是本章网络的参数， $\gamma$  是平衡分类损失和一致性损失的参数。

#### 4.2.7 模型推理

在模型训练阶段，本章采用双分支对比学习架构，通过约束其中特征一致性和标签一致性，反向梯度优化 Mamba 架构。而在模型推理环节，本章仅借助单分支 Mamba 模型来执行图像的标签预测工作。详细来讲，本章分别对以下两种推理模式进行了尝试：

**采用序列增强模块进行推理：**给定一张图像  $X$  经过图像预处理操作后得到实例特征序列，送入到序列增强模块获得新的序列。继续送入 Mamba 模块后建模各实例间信息，建模出更好的特征。最终，通过聚合模块输出包特征，将其输入分类网络以生成最终预测结果。具体推理流程详见图 4.2(a)。

**不采用序列增强模块进行推理：**本章还尝试了不采用序列增强模块直接进行输入至 Mamba 模块进行推理。推理流程详情见图 4.2(b)所示。

在不特别说明的情况下，本文采用序列增强模块进行推理。

### 4.3 实验设置及结果分析

在本节中，首先介绍了本章方法的实验设置，包括实验使用数据集、评估指标和实验细节，然后分析了基于扫描不变性 Mamba 的高分辨率病理图像分类实验结

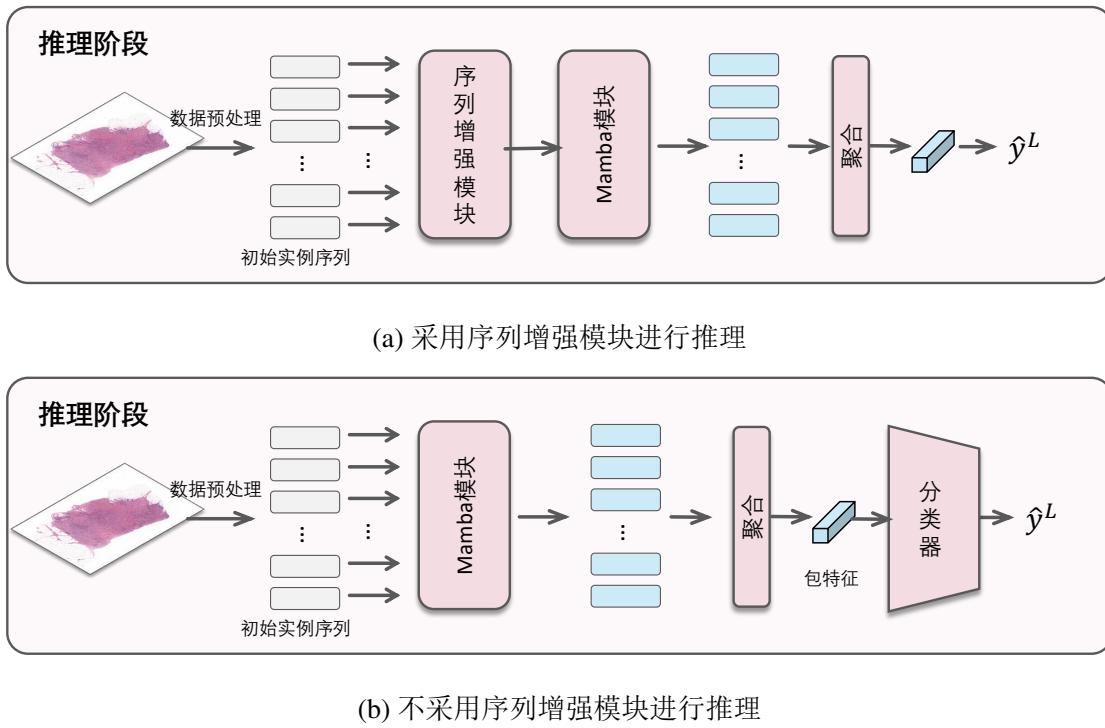


图 4.2 SMC-MIL 模型推理流程示意图。

Fig. 4.2 Schematic diagram of SMC-MIL model inference process.

果，接下来对模型的各个模块以及超参数进行了消融实验和分析，最后对模型所关注的重点区域进行了可视化分析。

### 4.3.1 实验设置

#### (1) 实验环境

本章实验依托于和上一章实验同样的设置环境和平台。详细的实验环境参数配置见表 4.1：

表 4.1 基于扫描不变性 Mamba 的高分辨率病理图像分类的实验环境及配置

Table 4.1 Experimental environment and configuration of SMC-MIL

| 设备     | 配置               | 版本          | 数量 |
|--------|------------------|-------------|----|
| 操作系统   | Ubuntu           | 20.04.4 LTS | 1  |
| CPU    | AMD Ryzen 5700X  | -           | 1  |
| GPU    | GeForce RTX 3090 | -           | 3  |
| IDE    | VsCode           | 1.86.2      | 1  |
| 编程语言   | Python           | 3.10.11     | -  |
| 深度学习框架 | Pytorch          | 2.0.1       | -  |

#### (2) 数据集

本章同样通过三个 WSI 分析子任务来评估 SMC-MIL 模型：癌症诊断、亚型分类和生存预测。对于癌症诊断任务，本章采用和上一章同样的设置，采用 CAMELYON 数据集进行评估。对于分型任务，本章使用 TCGA-NSCLC、TCGA-BRCA 和 BRACS<sup>[61]</sup> 数据集。同时为了评估生存预测的准确性，本章使用 TCGA-LUAD, TCGA-LUSC, TCGA-BLCA 来评估生存预测任务的性能。

### (3) 评估指标

对于诊断和分型，本章利用 Accuracy、AUC 和 F1-score 来评估模型的性能，其中 AUC 是本章主要的评价指标，在消融实验的讨论中，本章仅报告 AUC。对于生存预测，本章报告了所有数据集的 C-index。默认情况下，本章采用 5 折交叉验证，这与性能比较的基线方法一致。具体来说，对于 BRACS 数据集，本章也遵循了来自官方的划分方案，在官方分区中将数据集分为 395:65:87 的训练集、验证集和测试集。本章进行了三次独立的重复实验，以尽量减少官方划分的随机影响。此外，为了在实验中保持与其他数据集的一致性，本章还对 BRACS 的 7 类分类任务进行了 5 次交叉验证实验，标记为 BRACS-7<sup>\*</sup>，并报告在表4.7中。

### (4) 实施细节与对比方法

在实施细节上，本章采用与上一章实验同样的实验细节设置。在对比方法上，本文采用了 10 种广泛使用的 MIL 方法以及上一章研究的方法与本章方法进行比较。其中包括：AB-MIL<sup>[43]</sup>, CLAM<sup>[10]</sup>, DSMIL<sup>[9]</sup>, TransMIL<sup>[11]</sup>, DTFD-MIL<sup>[12]</sup>, IBMIL<sup>[72]</sup>, MHIM-MIL<sup>[60]</sup>, SRMamba<sup>[52]</sup>, WIKG<sup>[73]</sup>, R<sup>2</sup>T-MIL<sup>[13]</sup> 和 M<sup>2</sup>S-MIL。其中 SRMamba 与上一章所提出的方法 M<sup>2</sup>S-MIL 是基于 Mamba 架构的方法，也是本章主要的基准方法。本文在相同的设置下复现了这些方法的结果以进行比较。

#### 4.3.2 基准数据集实验结果

##### (1) 癌症诊断和分型

表 4.2 SMC-MIL 在 CAMELYON 数据集的癌症诊断性能比较 (%). 最优结果用粗体表示，次优结果使用下划线表示。

Table 4.2 Comparison of cancer diagnostic performance of SMC-MIL on the CAMELYON dataset (%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                       | Accuracy   | AUC        | F1-score   |
|--------------------------|------------|------------|------------|
| AB-MIL <sup>[43]</sup>   | 88.08±1.87 | 91.18±1.60 | 83.59±2.75 |
| CLAM <sup>[10]</sup>     | 87.98±2.17 | 91.70±2.15 | 83.38±3.09 |
| DSMIL <sup>[9]</sup>     | 88.42±1.77 | 91.50±1.23 | 83.82±2.95 |
| TransMIL <sup>[11]</sup> | 88.75±1.04 | 91.82±1.82 | 84.47±1.57 |
| DTFD-MIL <sup>[12]</sup> | 88.07±1.13 | 91.61±1.73 | 83.24±1.67 |
| MHIM-MIL <sup>[60]</sup> | 89.42±1.16 | 92.55±1.11 | 84.63±2.13 |
| IBMIL <sup>[72]</sup>    | 88.98±2.08 | 91.75±2.09 | 84.61±3.07 |
| SRMamba <sup>[52]</sup>  | 88.75±1.47 | 91.82±0.93 | 84.27±2.08 |

表 4.2 (续)  
Table 4.2 (continued)

| 方法                        | Accuracy          | AUC               | F1-score          |
|---------------------------|-------------------|-------------------|-------------------|
| WiKG [73]                 | 88.19±2.30        | 91.42±1.44        | 82.93±3.41        |
| R <sup>2</sup> T-MIL [13] | 89.20±1.48        | 92.29±1.32        | 85.20±1.86        |
| 第三章方法                     | <u>89.76±1.33</u> | <u>93.02±1.20</u> | <u>85.48±2.43</u> |
| <b>SMC-MIL(Ours)</b>      | <b>89.97±1.78</b> | <b>93.81±1.65</b> | <b>85.76±1.77</b> |

表 4.3 SMC-MIL 在 TCGA-NSCLC 数据集的亚型分类性能比较 (%). 最优结果用粗体表示, 次优结果使用下划线表示。

Table 4.3 Comparison of sub-typing performance of SMC-MIL on the TCGA-NSCLC dataset (%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                        | Accuracy          | AUC               | F1-score          |
|---------------------------|-------------------|-------------------|-------------------|
| AB-MIL [43]               | 90.42±1.76        | 95.23±1.46        | 89.93±1.80        |
| CLAM [10]                 | 87.98±2.17        | 91.70±2.15        | 83.38±3.09        |
| DSMIL [9]                 | 90.33±1.97        | 95.68±1.01        | 89.91±2.19        |
| TransMIL [11]             | 90.13±1.07        | 94.92±1.02        | 89.52±1.07        |
| DTFD-MIL [12]             | 90.71±1.66        | 95.64±1.44        | 90.29±1.83        |
| MHIM-MIL [60]             | 91.18±1.79        | 95.84±1.41        | 90.91±1.97        |
| IBMIL [72]                | 89.76±2.29        | 95.49±1.18        | 89.43±2.21        |
| SRMamba [52]              | 90.43±3.26        | 95.07±1.96        | 90.18±3.21        |
| WiKG [73]                 | 89.67±2.49        | 94.96±1.66        | 89.53±2.43        |
| R <sup>2</sup> T-MIL [13] | 90.81±2.41        | 96.20±1.16        | 90.49±2.41        |
| 第三章方法                     | <u>91.48±2.29</u> | <u>96.65±1.08</u> | <u>91.19±2.29</u> |
| <b>SMC-MIL(Ours)</b>      | <b>91.85±2.37</b> | <b>97.10±1.46</b> | <b>91.61±2.40</b> |

表 4.4 SMC-MIL 在 TCGA-BRCA 数据集的亚型分类性能比较 (%). 最优结果用粗体表示, 次优结果使用下划线表示。

Table 4.4 Comparison of sub-typing performance of SMC-MIL on the TCGA-BRCA dataset (%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法            | Accuracy   | AUC        | F1-score   |
|---------------|------------|------------|------------|
| AB-MIL [43]   | 86.80±2.56 | 91.17±1.69 | 72.42±3.48 |
| CLAM [10]     | 86.15±3.98 | 91.38±1.67 | 72.29±4.21 |
| DSMIL [9]     | 86.58±3.86 | 91.80±2.09 | 72.63±4.49 |
| TransMIL [11] | 86.22±3.30 | 91.61±1.81 | 72.20±4.28 |
| DTFD-MIL [12] | 86.04±3.05 | 91.34±1.95 | 70.76±4.20 |
| MHIM-MIL [60] | 86.80±2.56 | 92.24±1.63 | 74.77±7.72 |

表 4.4 (续)  
Table 4.4 (continued)

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| IBMIL <sup>[72]</sup>                | 85.54±3.20        | 91.41±2.20        | 71.14±3.60        |
| SRMamba <sup>[52]</sup>              | 88.14±3.05        | 93.31±1.40        | 75.61±4.00        |
| WiKG <sup>[73]</sup>                 | 88.53±2.28        | 92.14±2.06        | 75.11±3.98        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 88.25±2.06        | 92.58±1.75        | 75.15±2.91        |
| 第三章方法                                | <u>89.26±2.37</u> | <u>93.50±1.70</u> | <u>76.62±3.47</u> |
| <b>SMC-MIL(our)</b>                  | <b>90.16±1.83</b> | <b>93.98±1.02</b> | <b>78.11±3.30</b> |

表 4.5 SMC-MIL 在 BRACS 数据集官方划分下的粗粒度分类（三分类）性能对比 (%). 最优结果用粗体表示，次优结果使用下划线表示。

Table 4.5 Performance comparison of SMC-MIL coarse-grained classification (3-class) under the official classification of BRACS dataset (%)). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB-MIL <sup>[43]</sup>               | 58.05±1.44        | 80.68±0.35        | 55.83±1.97        |
| CLAM <sup>[10]</sup>                 | 57.85±0.94        | 83.54±0.69        | 58.41±1.08        |
| DSMIL <sup>[9]</sup>                 | 55.21±0.85        | 82.54±0.57        | 51.67±0.75        |
| TransMIL <sup>[11]</sup>             | 54.17±2.55        | 79.32±0.31        | 50.68±2.40        |
| DTFD-MIL <sup>[12]</sup>             | 58.22±6.26        | 85.96±2.13        | 59.70±2.86        |
| MHIM-MIL <sup>[60]</sup>             | 57.78±1.28        | 81.03±0.13        | 54.74±1.56        |
| IBMIL <sup>[72]</sup>                | 58.30±3.58        | 80.07±2.42        | 58.27±4.88        |
| SRMamba <sup>[52]</sup>              | 58.38±5.21        | 83.12±3.90        | 57.87±6.94        |
| WiKG <sup>[73]</sup>                 | 51.74±2.46        | 76.71±1.80        | 48.71±2.31        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 56.02±2.36        | 82.11±0.54        | 55.58±3.65        |
| 第三章方法                                | <u>59.38±1.13</u> | <u>86.21±0.46</u> | <u>60.13±1.79</u> |
| <b>SMC-MIL(Ours)</b>                 | <b>59.78±1.18</b> | <b>87.74±0.36</b> | <b>61.13±2.08</b> |

表 4.6 SMC-MIL 在 BRACS 数据集官方划分下的细粒度分类（七分类）性能对比 (%). 最优结果用粗体表示，次优结果使用下划线表示。

Table 4.6 Performance comparison of SMC-MIL's fine-grained classification (7-class) under the official classification of BRACS dataset (%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                     | Accuracy          | AUC        | F1-score   |
|------------------------|-------------------|------------|------------|
| AB-MIL <sup>[43]</sup> | 30.44±2.67        | 73.92±0.29 | 28.76±2.12 |
| CLAM <sup>[10]</sup>   | <u>37.60±1.52</u> | 76.83±1.29 | 35.68±2.84 |

表 4.6 (续)  
Table 4.6 (continued)

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| DSMIL <sup>[9]</sup>                 | 27.68±5.10        | 76.66±0.83        | 27.56±4.34        |
| TransMIL <sup>[11]</sup>             | 26.83±5.48        | 70.34±2.57        | 25.28±3.15        |
| DTFD-MIL <sup>[12]</sup>             | 35.02±2.76        | <u>79.17±0.80</u> | <u>35.92±6.29</u> |
| MHIM-MIL <sup>[60]</sup>             | 34.39±1.55        | 74.20±0.46        | 33.09±1.08        |
| IBMIL <sup>[72]</sup>                | 32.74±2.36        | 74.80±0.29        | 30.58±2.94        |
| SRMamba <sup>[52]</sup>              | 33.99±4.27        | 74.92±2.00        | 33.94±3.26        |
| WiKG <sup>[73]</sup>                 | 33.13±2.43        | 74.14±1.23        | 32.93±4.39        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 26.64±1.02        | 74.02±1.40        | 26.67±1.59        |
| 第三章方法                                | 34.16±2.04        | 77.37±1.24        | 34.37±2.73        |
| <b>SMC-MIL(Ours)</b>                 | <b>38.29±0.87</b> | <b>79.57±0.42</b> | <b>37.93±1.33</b> |

表 4.7 SMC-MIL 在 BRACS 数据集五折交叉验证下的细粒度分类性能对比 (%). 最优结果用粗体表示, 次优结果使用下划线表示。

Table 4.7 Performance comparison of SMC-MIL for fine-grained classification (7-class) under five-fold cross validation on BRACS dataset(%). The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | Accuracy          | AUC               | F1-score          |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB-MIL <sup>[43]</sup>               | 51.18±2.38        | 82.09±2.76        | 46.58±2.81        |
| CLAM <sup>[10]</sup>                 | 51.74±6.49        | 82.88±2.15        | 49.26±7.70        |
| DSMIL <sup>[9]</sup>                 | 48.81±5.89        | 83.56±1.45        | 45.40±6.66        |
| TransMIL <sup>[11]</sup>             | 45.34±7.70        | 80.61±2.64        | 42.90±8.73        |
| DTFD-MIL <sup>[12]</sup>             | 51.92±2.16        | 82.81±2.00        | 38.76±3.80        |
| MHIM-MIL <sup>[60]</sup>             | 45.31±4.93        | 81.54±2.80        | 41.41±4.51        |
| IBMIL <sup>[72]</sup>                | 51.19±3.23        | 81.69±2.63        | 47.45±2.84        |
| SRMamba <sup>[52]</sup>              | <u>55.04±2.86</u> | 83.89±2.95        | <u>52.00±3.76</u> |
| WiKG <sup>[73]</sup>                 | 52.28±5.47        | 82.82±3.04        | 49.44±6.40        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 52.84±3.81        | 83.96±1.73        | 49.11±7.11        |
| 第三章方法                                | 54.88±1.26        | <u>84.20±0.92</u> | 51.89±4.38        |
| <b>SMC-MIL(our)</b>                  | <b>56.85±2.52</b> | <b>85.11±2.14</b> | <b>55.68±2.56</b> |

表4.2呈现了各类 MIL 方法在 CAMELYON 数据集上的癌症诊断性能。在该数据集上, 本章所采用的方法表现卓越。从关键指标来看, Accuracy 相较于次优方法提升了 0.21%, AUC 提高了 0.79%, F1-score 增加了 0.28%。与除本文上一章所提方法外的 SOTA 方法相比, 本章方法在 AUC 上有 1.27% 的提升。和同样基于 Mamba 架构的 SRMamba 对比, 所提出的 SMC-MIL 在 AUC 性能方面优势明显, 高出 1.27%。这些数据充分表明, 在 CAMELYON 数据集的所有基准测试指标下,

本章方法实现了最佳性能，有力地验证了消除 Mamba 对扫描模式的依赖能够更好地适配 MIL 任务这一假设。

表4.3展示了各类 MIL 方法于 TCGA-NSCLC 数据集上的癌症分型性能。在 TCGA-NSCLC 数据集中,本章方法同样表现出色。Accuracy 较次优方法提升 0.37%, AUC 提升 0.45%, F1 - score 提升 0.42%。相较于除上一章外的 SOTA 方法, 本章方法在 AUC 上提升了 0.90%。相较于基于 Mamba 架构的 SRMamba, SMC - MIL 在 AUC 性能上高出 1.58%。由此可见, 在 TCGA-NSCLC 数据集的各项基准测试指标中, 本章方法同样展现出最优性能, 进一步验证了消除 Mamba 对扫描模式的依赖有助于其更好地适应 MIL 任务这一观点。

表4.4报告了 BRCA 数据集上的亚型分类性能比较。观察结果与表格 4.2 和 4.3 所示一致。本章的模型在所有指标上都达到了最先进的性能。与次优方法相比, 该方法在 BRCA 数据集上的 Accuracy、AUC 和 F1-score 分别提高了 0.9%、0.48% 和 1.49%。在与 SRMamba 模型相比, Accuracy、AUC 和 F1-score 分别增加了 2.02%, 0.67%, 2.5%。

表4.5, 表4.6和表4.7展示了 BRACS 数据集上亚型分类性能。其中, 表 4.5 和 表 4.6 报告了在 BRACS 数据集上使用官方划分的 3 类和 7 类分类的性能结果。结果表明, 与基线相比, 本章的方法仍然具有优势。例如, 本章的方法在这两个任务上分别将 AUC 中的 SRMamba 提高了 4.62% 和 4.65%。同时, 为了与前文保持设置一致性, 表 4.7 展示了在 BRACS 五折交叉验证下细粒度分类的结果。与次优方法相比, 本章方法在该设置下的 Accuracy、AUC 和 F1-score 分别提高了 1.81%、1.51% 和 3.68%。这些结果都清楚地表明, 本章方法的显著性能提升不仅仅归功于 Mamba 本身强大的建模能力, 还证实了该方法的扫描不变性能能够更好帮助 Mamba 解决 MIL (多实例学习) 问题。

## (2) 生存预测

表 4.8 SMC-MIL 在 R50 提取特征的三个主要数据集上的生存预测结果 (%)。

最优结果用粗体表示, 次优结果使用下划线表示。

Table 4.8 Survival Prediction results on three main datasets extracted by R50(%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                       | BLCA                               | LUAD             | LUSC             |
|--------------------------|------------------------------------|------------------|------------------|
| AB-MIL <sup>[43]</sup>   | <b><math>57.97 \pm 5.43</math></b> | $60.56 \pm 5.59$ | $59.11 \pm 2.47$ |
| CLAM <sup>[10]</sup>     | $58.73 \pm 3.24$                   | $60.45 \pm 5.94$ | $60.71 \pm 2.91$ |
| DSMIL <sup>[9]</sup>     | $58.29 \pm 3.53$                   | $59.78 \pm 5.53$ | $60.21 \pm 3.90$ |
| TransMIL <sup>[11]</sup> | $57.00 \pm 3.60$                   | $63.06 \pm 2.97$ | $57.46 \pm 4.69$ |
| DTFD-MIL <sup>[12]</sup> | $59.85 \pm 4.18$                   | $60.65 \pm 5.62$ | $59.03 \pm 2.54$ |
| MHIM-MIL <sup>[60]</sup> | $59.70 \pm 4.06$                   | $63.26 \pm 1.00$ | $60.91 \pm 4.26$ |
| IBMIL <sup>[72]</sup>    | $59.09 \pm 5.42$                   | $60.69 \pm 5.71$ | $58.50 \pm 2.13$ |
| SRMamba <sup>[52]</sup>  | $60.94 \pm 2.19$                   | $61.47 \pm 4.29$ | $58.35 \pm 3.95$ |

表 4.8 (续)  
Table 4.8 (continued)

| 方法                                   | BLCA              | LUAD              | LUSC              |
|--------------------------------------|-------------------|-------------------|-------------------|
| WiKG <sup>[73]</sup>                 | 58.04±3.30        | 62.77±4.43        | 58.10±4.95        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 60.98±2.73        | 63.28±3.75        | 60.64±3.83        |
| 第三章方法                                | <u>61.96±1.28</u> | <u>63.88±6.92</u> | <u>61.22±4.30</u> |
| <b>SMC-MIL</b>                       | <b>62.49±3.11</b> | <b>65.05±4.05</b> | <b>61.64±5.02</b> |

表 4.9 SMC-MIL 在 UNI 提取特征的三个主要数据集上的生存预测结果 (%)。

最优结果用粗体表示，次优结果使用下划线表示。

Table 4.9 Survival Prediction results on three main datasets extracted by UNI(%).

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法                                   | BLCA              | LUAD              | LUSC              |
|--------------------------------------|-------------------|-------------------|-------------------|
| AB-MIL <sup>[43]</sup>               | 59.42±5.96        | 63.37±3.65        | 59.25±4.74        |
| CLAM <sup>[10]</sup>                 | 59.68±4.72        | 65.64±3.69        | 59.69±2.88        |
| DSMIL <sup>[9]</sup>                 | 60.83±4.08        | 64.41±2.30        | 60.58±4.25        |
| TransMIL <sup>[11]</sup>             | 59.14±5.34        | 61.97±5.17        | 58.78±3.54        |
| DTFD-MIL <sup>[12]</sup>             | 59.64±6.28        | 65.19±1.56        | 60.27±5.39        |
| MHIM-MIL <sup>[60]</sup>             | 60.78±3.71        | 64.88±0.71        | 61.43±3.77        |
| IBMIL <sup>[72]</sup>                | 59.33±4.02        | 64.56±4.24        | 61.12±5.10        |
| SRMamba <sup>[52]</sup>              | 61.46±3.00        | 64.82±5.05        | 59.30±5.69        |
| WiKG <sup>[73]</sup>                 | 59.53±3.12        | 63.93±3.87        | 60.32±3.25        |
| R <sup>2</sup> T-MIL <sup>[13]</sup> | 59.42±3.15        | 64.69±2.67        | 61.06±3.85        |
| 第三章方法                                | <u>61.74±4.80</u> | <u>66.17±3.77</u> | <u>62.05±4.30</u> |
| <b>SMC-MIL</b>                       | <b>62.02±5.22</b> | <b>66.79±3.47</b> | <b>63.47±5.09</b> |

表 4.8 和表 4.9 分别给出了使用 R50 和 UNI 作为特征提取器在三个生存预测数据集上的实验结果。本章提出的 SMC-MIL 模型表现出很强的性能，在 BLCA 数据集上的 C-index 得分为 61.96%，在 LUAD 上的得分为 65.05%，在 LUSC 上的得分为 61.64%。其明显优于比较的方法，在每个数据集上分别比除上一章外的次优方法提高了 0.98%，1.00% 和 0.73%。此外，即使在利用基础模型提取的高质量特征时，SMC-MIL 也会产生不错的性能。特别地，与 SRMamba 相比，它在 BLCA、LUAD 和 LUSC 数据集上的性能分别提高了 0.56%、1.87% 和 4.17%。这些结果强调了本章提出的策略的一致性和可靠性，强调了其在预测生存结果方面的有效性。

### 4.3.3 消融实验

#### (1) 不同模块的有效性分析

为了研究 SMC-MIL 模型中每个模块对模型的影响，本文在 CAMELYON、TCGA-NSCLC 和 TCGA-BRCA 三个数据集上进行了消融实验，此部分实验分为五个模型，

分别为基准模型（Origin Mamba），应用双分支一致性约束的基准模型（+Con.），添加序列对比增强模块的基准模型（+Aug.），（+Con.+Aug.+Samp.）表示使用采样策略扩展“hard”分支以匹配“Easy”分支的长度，以及最终模型（+Con.+Aug.）。

表 4.10 在三个主要数据集上，不同组件对性能的影响。最优结果用粗体表示，次优结果用下划线表示。

Table 4.10 Module ablation experiments of SMC-MIL on CAMELYON, NSCLC and BRCA. The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法               | CAMELYON          | NSCLC             | BRCA              |
|------------------|-------------------|-------------------|-------------------|
| Origin Mamba     | 91.57±1.37        | 95.03±1.89        | 92.56±2.11        |
| +Con.            | 93.07±1.10        | 96.21±1.37        | 93.44±2.21        |
| +Aug.            | 91.71±1.36        | 95.29±1.21        | 93.17±1.56        |
| +Con.+Aug.+Samp. | <u>93.13±1.90</u> | <u>96.41±1.13</u> | <u>93.47±1.17</u> |
| +Con.+Aug.       | <b>93.81±1.65</b> | <b>96.65±1.46</b> | <b>93.98±1.02</b> |

在三个数据集上的模块消融实验如表4.10所示。其报告了 SMC-MIL 中不同模块在三个数据集上的影响。OriginMamba 意味着简单地将 Mamba 整合到 ABMIL 中，Aug. 表示应用了序列对比学习，Samp. 表示使用采样策略扩展“hard”分支以匹配“Easy”分支的长度，Con. 表示应用一致性约束。本章的基线指的是将原始的 Mamba 架构合并到 ABMIL 中。首先，将对比学习和一致性损失框架集成到基线模型中。该策略在 CAMELYON、NSCLC 和 BRCA 数据集上的 AUC 分别提高了 1.50%、1.18% 和 0.88%，表明提出的一致性约束有效地指导 Mamba 从不同的扫描序列中学习到更详细的、与序列顺序无关的特征。实验结果表明，与 OriginMamba 相比，序列增强和掩码模块的引入分别增强了 Mamba 对序列内显著区域的区分能力。将这两个模块引入 OriginMamba 后，完整的 SMC-MIL 取得了最好的性能（在 CAMELYON 上 AUC 为 93.81%，在 TCGA 上 AUC 为 96.65%，在 BRCA 上 AUC 为 93.98%）。

## （2）扫描不变性的验证分析

为了验证本章的方法使 Mamba 能够构建更多的扫描不变性特征，本章将三个随机选择的扫描序列输入到训练的 Mamba 变体中进行推理，并分别计算其 AUC 的平均值和标准差。实验结果见表4.11。与基线相比，本章的方法在 CAMELYON 和 NSCLC 数据集上的标准差降低了 2 倍，在 BRACS-7 数据集上的标准差降低了 3 倍。值得注意的是，本章的方法不仅在所有数据集中实现了最高的平均 AUC，而且还显示出最小的标准差，并且与基准实验（表4.2、4.3、4.4和表4.7）中的正常输入的实验结果相差无几。这强调了本章的方法在指导 Mamba 模型学习扫描序列的更多不变特征方面的有效性，优于其他 Mamba 变体方法。

表 4.11 不同 Mamba 变体的扫描不变性评估对比。最优结果用粗体表示。

Table 4.11 Scan-invariant Evaluation of different Mamba variants. The optimal experimental results are marked in bold.

| 变体             | CAMELYON          | NSCLC             | BRCA              | BRACS-7*          |
|----------------|-------------------|-------------------|-------------------|-------------------|
| Mamba          | 89.29±0.51        | 94.34±1.07        | 91.96±1.20        | 78.08±0.79        |
| BiMamba        | 86.91±0.61        | 94.78±1.11        | 92.32±1.21        | 77.37±0.69        |
| SRMamba        | 89.99±0.46        | 94.73±0.62        | 91.99±1.78        | 77.27±0.31        |
| <b>SMC-MIL</b> | <b>93.09±0.25</b> | <b>96.12±0.50</b> | <b>93.00±0.95</b> | <b>84.30±0.23</b> |

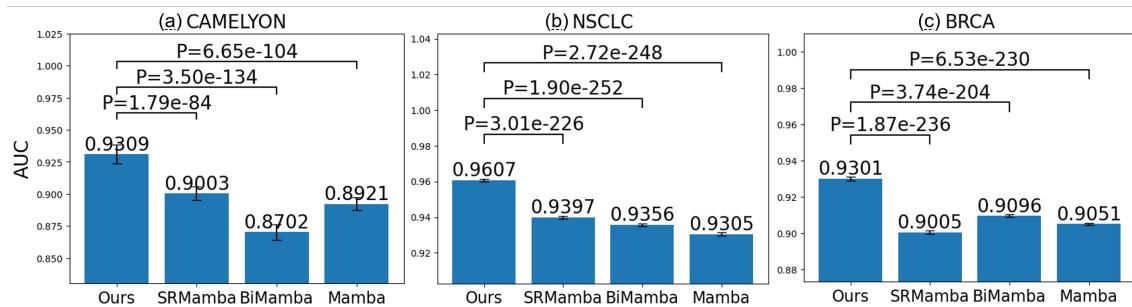


图 4.3 不同 Mamba 变体的扫描不变性评估的双样本 t 检验

Fig. 4.3 Two-sample t-test for scanning invariance evaluation of different Mamba variants.

为了进一步验证本章方法的扫描不变性，本章重复了验证过程，将随机化的数量增加到 100，并进行了双样本 t 检验，如图 4.3 所示。结果显示，本章的方法不仅在平均值方面优于其他 Mamba 方法，而且 p 值远低于 0.05。这表明，在统计意义上，其他 Mamba 方法与本章方法具有显著的差异。

### (3) 不同 Mamba 模型性能比对

为了进一步验证本章的方法的优越性来源于本章的架构而不是 Mamba 本身，本章对三种变体进行了比较实验：图 4.4 中的 Mamba Block、BiMamba 和 SRMamba。结果表明，SMC-MIL 在所有四个数据集上的所有六个任务中都取得了最佳性能。本章的方法不仅在性能上超过了原始的单分支 Mamba，而且在多序列（或扫描）聚合方法（例如，BiMamba 和 SRMamba，以及上一章提出的方法）中表现出许多优势。实验结果表明，扫描不变 Mamba 算法比多扫描集成 Mamba 算法具有更强的实例建模能力，能够提取更多的判别性特征。

### (4) 不同的分支对比策略对模型性能的影响

表 4.12 列出了五种不同分支对比策略下 SMC-MIL 的性能。结果表明，“Easy vs. Hard” 策略产生最佳性能。表现如此优异的主要原因是“Easy” 分支帮助 Mamba 产生可靠的包表示和分类，而“hard” 分支迫使模型不仅要学习扫描无关的特征，还要从被屏蔽序列中探索更多的判别实例，以实现两个分支之间的一致性。此外，

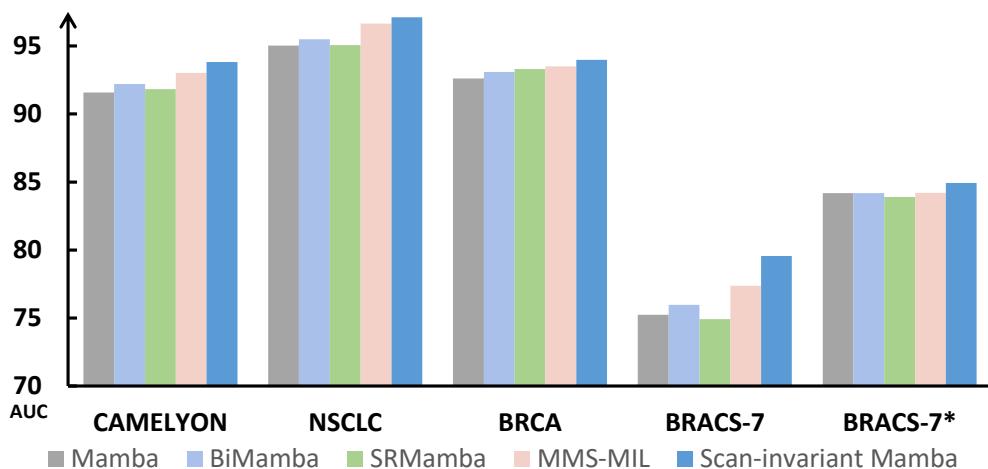


图 4.4 不同 Mamba 模型性能比对柱状图。

Fig. 4.4 Bar chart of performance comparison of different Mamba models.

表 4.12 不同的分支对比策略对模型性能的影响。最优结果使用粗体表示，次优结果使用下划线表示。

Table 4.12 Effects of different branch contrast strategies on model performance. The best results are shown in bold.

| 方法                   | CAMELYON          | NSCLC             | BRACS-3           | BRACS-7           |
|----------------------|-------------------|-------------------|-------------------|-------------------|
| Vanilla vs. Hard     | <u>93.13±1.90</u> | 96.30±1.23        | 84.89±0.73        | <u>76.52±0.32</u> |
| Easy vs. Vanilla     | 92.64±1.33        | <u>96.55±1.21</u> | 82.66±4.39        | 77.44±0.95        |
| Hard vs. Hard        | 90.39±2.55        | 95.79±1.48        | 82.57±1.83        | 77.22±2.65        |
| Easy vs. Easy        | 91.66±2.10        | 96.35±1.23        | <u>85.63±0.84</u> | 74.48±1.61        |
| <b>Easy vs. Hard</b> | <b>93.81±1.65</b> | <b>96.65±1.46</b> | <b>87.74±0.36</b> | <b>79.57±0.42</b> |

两个分支的差异也有利于性能提升。“Hard vs. Hard” 策略表现最差。本章将其性能较差归因于在生成包表示和分类过程中缺乏来自“Easy” 分支的可靠指导。

### (5) 讨论不同扫描模式的组合对模型性能的影响

比较不同的扫描模式可能会产生不同的结果。因此，本章还研究了使用不同形式的对比对 SMC-MIL 性能的影响。如表4.13所示，在 CAMELYON 数据集上，通过比较原始扫描模式和随机扫描模式可以获得最佳性能，而在其他数据集上，原始扫描模式和反向扫描模式可以获得最佳效果。因此，在 CAMELYON 实验中，本章使用原始扫描模式和随机扫描模式来生成不同顺序的序列，对于剩余的数据集，本章使用原始扫描模式和反向扫描模式。本小节同样对比了上一章提出的网格扫描和区域螺旋扫描模式，同样取得不错的性能。这侧面说明第三章提出的扫描模式补足了原始序列没有的空间信息。

表 4.13 讨论不同扫描模式的组合对模型性能的影响。最优结果用粗体表示。

Table 4.13 Effect of combinations of different scan patterns on model performance. The best results are shown in bold.

| 方法                  | CAMELYON          | NSCLC             | BRCA              | BRACS-7*          |
|---------------------|-------------------|-------------------|-------------------|-------------------|
| Origin vs. Backward | 92.64±1.36        | <b>96.65±1.46</b> | <b>93.98±1.02</b> | <b>85.11±2.14</b> |
| Origin vs. Random   | <b>93.81±1.65</b> | 96.40±1.09        | 93.20±1.50        | 84.75±2.79        |
| Origin vs. Grid.    | 92.70±2.80        | 96.32±1.78        | 93.62±1.65        | 84.59±1.33        |
| Origin vs. RSS.     | 92.26±2.17        | 95.78±1.28        | 93.73±1.98        | 84.48±2.00        |
| Random vs. Random   | 92.10±2.20        | 95.79±1.08        | 92.87±1.02        | 83.45±1.74        |

#### (6) 讨论序列增强率的影响 $\alpha\%$

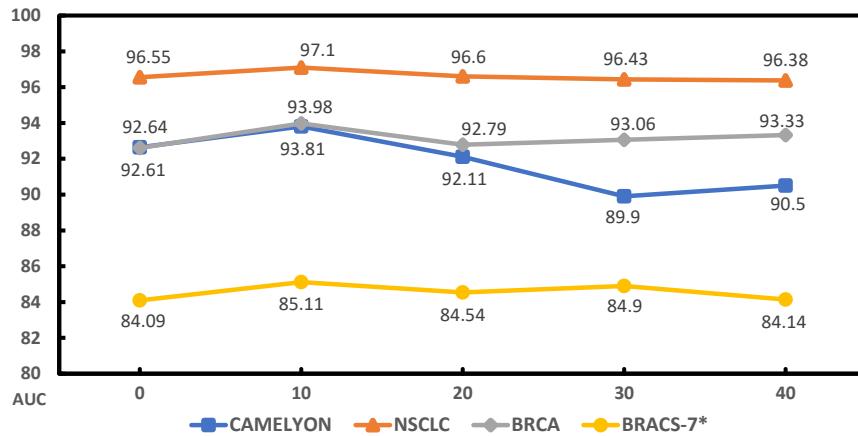
图 4.5 不同序列增强率  $\alpha\%$  下的模型性能折线图。Fig. 4.5 Line plot of model performance for different sequence enhancement rates  $\alpha\%$ .

图4.5讨论了超参数序列增强率  $\alpha\%$  的影响。为了更好地突出实验结果，本章将 BRACS-7\* 结果绘制在副轴上。如图所示，除 CAMELYON 外，大多数数据集在序列增强率为 0.1 时达到最佳性能，CAMELYON 在  $\alpha\% = 40\%$  时达到峰值。为了保持一致性，本章将所有数据集的序列扩增率  $\alpha\%$  默认设置为 10%。

#### (7) 讨论序列掩码率的影响 $\beta\%$

图 4.6讨论了超参数序列掩码率  $\beta\%$  的影响。本小节与前一小节一致，同样在次级轴上绘制了 BRACS-7\* 结果。所有数据集在 10% 的序列掩码率下达到最佳性能。因此，除非另有说明，否则本章默认设置  $\beta\% = 10\%$ 。

#### (8) Bootstrap 检验

本小节对本章方法和 SOTA 方法进行了有放回的随机抽样 Bootstrap 检验，共采样 100 次，得到的结果如图 4.7所示。可以看出，本章的方法显著高于 SOTA，因

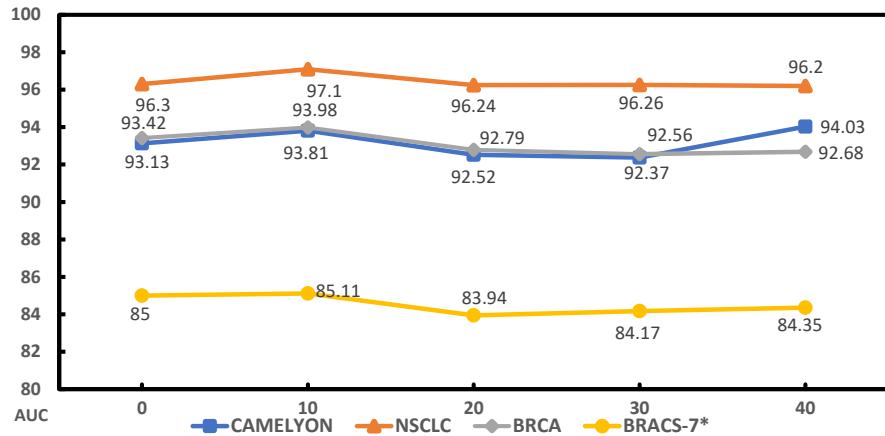
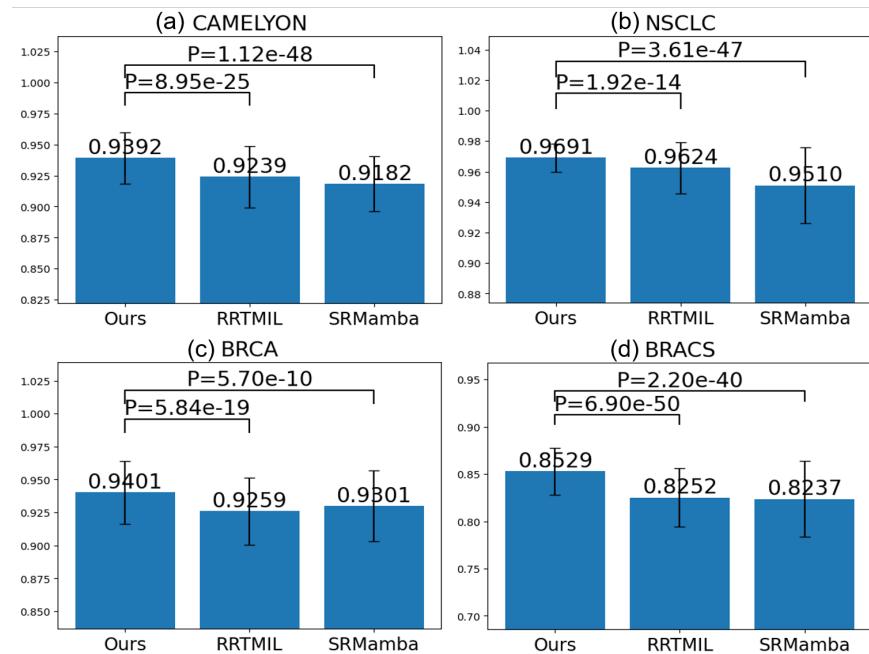
图 4.6 不同序列掩码率  $\beta\%$  下的模型性能折线图。Fig. 4.6 Line plot of model performance for different sequence masking rates  $\beta\%$ .

图 4.7 与 SOTA 方法性能比较的 Bootstrap 检验。

Fig. 4.7 Bootstrap test for performance comparison with SOTA method.

为  $p$  值远小于 0.05。

### (9) 两种推理方式的性能对比

本小节讨论了使用不同推理方式对性能结果的影响，结果报告在表 4.14 中。可以看出两者并没有明显的差异，这更证明本章所提出模型的泛化性能。默认设置下，本文采用有序列增强模块的推理模式进行推理。

### (10) 使用 PLIP 作为特征提取器的性能结果

为了进一步探索本章方法的有效性，本小节还进行了将同样采用自监督预训练

表 4.14 不同的推理方式的性能对比。最优结果用粗体表示。

Table 4.14 Performance comparison of different inference modes. The best results are shown in bold.

| 方法         | CAMELYON          | NSCLC             | BRACS-3           | BRACS-7           |
|------------|-------------------|-------------------|-------------------|-------------------|
| w/ 序列增强模块  | <b>93.81±1.65</b> | 96.65±1.46        | <b>87.74±0.36</b> | <b>79.57±0.42</b> |
| w/o 序列增强模块 | 93.70±1.42        | <b>96.67±1.31</b> | 87.65±4.17        | 79.43±0.95        |

表 4.15 SMC-MIL 以 PLIP 为离线特征提取器的癌症诊断性能。最优结果用粗体表示，次优结果用下划线表示。

Table 4.15 Cancer diagnosis performance of SMC-MIL with PLIP as the offline feature extractors.

The optimal experimental results are marked in bold, and the sub-optimal are underlined.

| 方法   | Accuracy       | AUC               | F1-score          |
|------|----------------|-------------------|-------------------|
| PLIP | AB-MIL         | 90.09±1.63        | <u>94.55±1.91</u> |
|      | OriginMamba    | 90.87±1.52        | <u>94.44±2.19</u> |
|      | BiMamba        | 90.54±1.89        | <u>94.20±1.99</u> |
|      | SRMamba        | <u>91.87±0.85</u> | <u>94.20±2.16</u> |
|      | 第三章方法          | <u>92.21±0.60</u> | <u>95.24±1.85</u> |
|      | <b>SMC-MIL</b> | <b>92.50±0.87</b> | <b>95.79±1.31</b> |

表 4.16 SMC-MIL 以 PLIP 为离线特征提取器在 TCGA-NSCLC 上的的亚型分型性能。最优结果用粗体表示，次优结果用下划线表示。

Table 4.16 Cancer sub-typing performance of SMC-MIL on the TCGA-NSCLC dataset with PLIP as the feature extractors. The optimal experimental results are marked in bold, and the sub-optimal experimental results are underlined.

| 方法   | Accuracy       | AUC               | F1-score          |
|------|----------------|-------------------|-------------------|
| PLIP | AB-MIL         | 89.94±1.84        | 94.45±1.64        |
|      | OriginMamba    | 90.89±1.60        | <u>95.57±1.60</u> |
|      | BiMamba        | 90.70±1.95        | <u>95.42±1.68</u> |
|      | SRMamba        | 89.85±2.07        | <u>95.72±1.51</u> |
|      | 第三章方法          | <u>91.56±1.76</u> | <u>96.22±1.12</u> |
|      | <b>SMC-MIL</b> | <b>92.30±0.60</b> | <b>96.78±1.68</b> |

的 PLIP<sup>[74]</sup> 作为特征提取器的相关实验，以验证本章方法的普遍适用性。如表 4.15、4.16 和 4.17 所示，可以看出实验结果与 R50 特征一致。本章的方法优于其它基于 Mamba 的方法。在 CAMELYON 数据集上，它在精度、AUC 和 F1-score 方面分别比次优的方法提高了 0.29%、0.55% 和 0.53%。这些指标在 NSCLC 数据集中分别为 0.74%，0.36% 和 1.15%。最后，BRCA 数据集的改进也值得注意，Acc 提高了 0.93%，AUC 提高了 1.31%，F1-score 提高了 0.37%。这进一步表明，即使使用非传统的 ResNet-50 特征提取器，本章方法的性能改进也不仅仅是依赖于简单的多个

表 4.17 SMC-MIL 以 PLIP 为离线特征提取器在 TCGA-BRCA 上的亚型分型性能。最优结果用粗体表示，次优结果用下划线表示。

Table 4.17 Cancer sub-typing performance of SMC-MIL on the TCGA-BRCA dataset with PLIP as the feature extractors. The optimal experimental results are marked in bold, and the sub-optimal experimental results are underlined.

|      | 方法             | Accuracy          | AUC               | F1-score          |
|------|----------------|-------------------|-------------------|-------------------|
| PLIP | AB-MIL         | 83.81±3.72        | 90.18±2.99        | 67.94±4.22        |
|      | OriginMamba    | 86.75±5.76        | 92.90±2.44        | 74.35±7.23        |
|      | BiMamba        | 87.73±3.06        | 92.64±2.28        | 74.07±4.75        |
|      | SRMamba        | 86.40±3.31        | 92.17±2.67        | 72.39±4.62        |
|      | 第三章方法          | <u>88.82±3.22</u> | <u>93.40±1.24</u> | <u>75.00±2.50</u> |
|      | <b>SMC-MIL</b> | <b>89.75±0.54</b> | <b>94.71±1.49</b> | <b>75.37±0.98</b> |

Mamba 分支合并，而是正确捕获了空间信息与视觉特征。

### (11) Mamba 效率优势的验证分析

为了说明

#### 4.3.4 可视化分析

为了直观地验证本章方法的可解释性，本章将 OriginMamba 和本章的方法在 Camelyon-16 数据集上产生的高注意力分数的实例可视化，如图4.8所示。蓝色的线条勾勒出肿瘤区域。越亮的斑块表明注意力得分越高。可视化结果清楚地表明，与原始 Mamba 相比，SMC-MIL 能够更准确地选择显著斑块。

并且为了更加充分验证本章方法的扫描不变性，本章用 t-SNE 可视化了同一个 WSI 切片经过 N 个不同扫描模式所建模出的 N 个包特征，结果展示在图4.9（展示图中 N 为 1000）。可以看到 SMC-MIL 建模出的包特征明显更为集中，而原始 Mamba 的建模更为分散。实验充分说明本章所提出的基于序列差异化对比的 Mamba 模型具有扫描不变性。

## 4.4 本章小结

本章针对原始 Mamba 倾向性建模出因果关系与病理图像多实例学习的序列不变性存在冲突的问题，以引导 Mamba 建模与顺序无关的包特征为切入点，提出了一种基于 Mamba 和对比学习的 MIL 框架，以解决扫描敏感 Mamba 和扫描不变 MIL 之间的差距。该框架通过对不同扫描分支的结果应用一致性约束，指导 Mamba 从不同序列中学习包的扫描不变特征。此外，序列对比度增强分别用于增强和掩码序列，迫使 Mamba 识别更多的鉴别斑块。在 3 个 WSI 子任务以及 12 个基准上的大量对比及消融实验表明，该框架优于最新的其他方法，并且本章的方法解决了多扫描信息整合的策略存在无法穷举所有扫描情况，高度依赖扫描模式

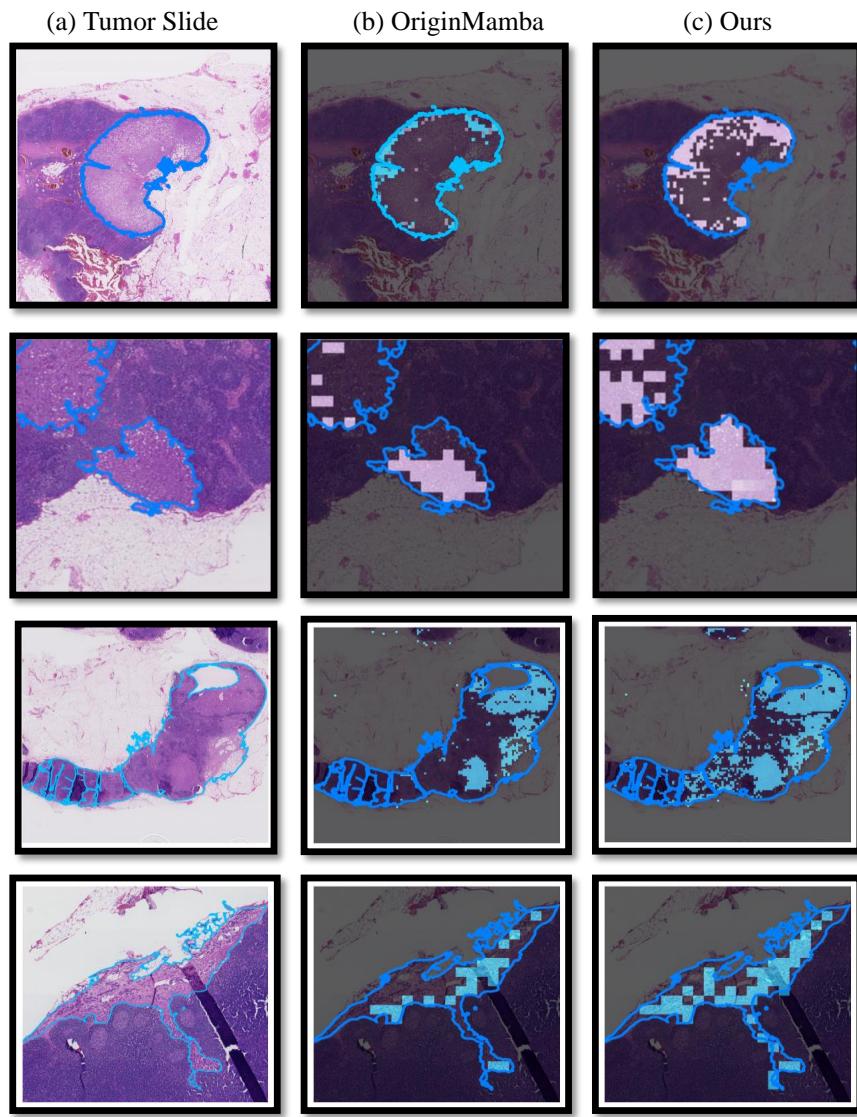


图 4.8 原始 Mamba(baseline) 和 SMC-MIL 制作的图块可视化结果

Fig. 4.8 Visualization results of patches focused by original Mamba(baseline) and SMC-MIL.

和集成方式的问题，有效地减轻了 Mamba 在 MIL 任务中的局限性，使其更适合这些任务，并为 Mamba 在视觉任务中的进一步应用铺平了道路。

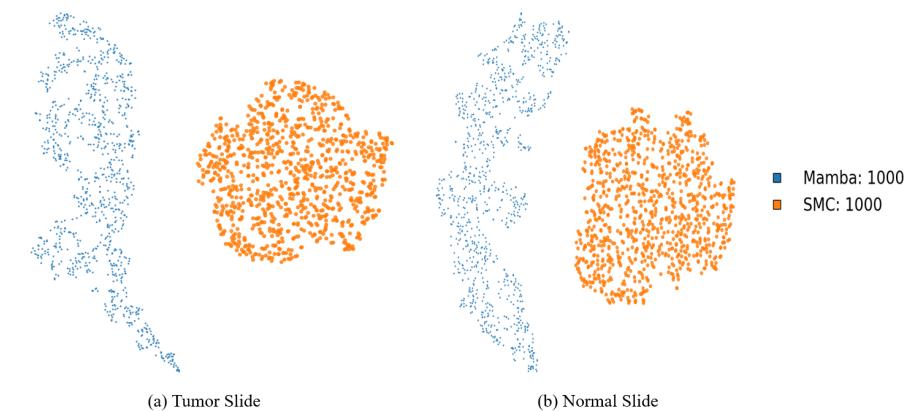


图 4.9 原始 Mamba(baseline) 和 SMC-MIL 聚合特征 t-SNE 可视化结果

Fig. 4.9 The t-SNE visualization of aggregated features from Mamba(baseline) and SMC-MIL.

## 5 总结与未来展望

本章内容共分为两节，第一节对本文研究内容与方法进行总结；第二节会阐述本文提出方法存在的局限之处，同时对后续的研究方向展开展望。。

### 5.1 总结

本研究聚焦于 Mamba 架构在病理图像分类领域的应用，旨在解决该架构在面对病理图像分类任务时所暴露出的一系列问题。病理图像分类任务由于实例数目众多，常导致计算资源消耗巨大，同时实例间关系建模也往往不够充分。Mamba 架构虽在一定程度上有效应对了这些挑战，但其仍存在无法充分捕获二维空间关系，以及与病理图像多实例学习的序列不变性存在冲突的问题。

为解决上述问题，实现 Mamba 架构在病理图像领域的高效适配，本研究从多扫描模式融合 Mamba 和扫描不变性 Mamba 两个独特角度展开深入研究：

(1) 在基于多扫描 Mamba 的高分辨率病理图像分类研究方面，鉴于 Mamba 一维序列建模所造成图像空间结构捕获能力不足的问题，研究提出了基于空间信息增强实现的多扫描 Mamba 的多实例学习框架 ( $M^2S$  - MIL)。该框架创新性地设计了多个扫描模式分支作为 Mamba 的输入，提出了区域螺旋扫描和网格扫描等新型扫描模式，显著提升了 Mamba 对局部连续性、旋转不变性以及同一序列不同方向空间关联信息的捕获能力。同时，通过设计基于分块局部注意力的 2D 空间上下文感知模块，进一步强化序列对 2D 空间上下文的感知能力，通过区域内信息整合，极大地提高了 Mamba 对空间能力的捕捉，有力地推动了 Mamba 模型与病理图像分类任务的适配进程。在 3 个 WSI 分析子任务和 7 个数据集上进行的大量实验充分证明，综合多种扫描模式分支并引入空间信息，能够切实解决 Mamba 从一维实例序列中空间建模不充分的问题，显著提高模型性能，同时起到一定缓解数字病理图像分类任务中使用 Mamba 模型时出现的建模倾向差异的作用。

(2) 在基于扫描不变性 Mamba 的高分辨率病理图像分类研究方面，针对 Mamba 作为序列相关模型，与病理图像多实例学习的序列不变性存在冲突的问题，研究提出了基于对比学习策略的扫描不变 Mamba 的多实例学习 (SMC-MIL) 框架。为引导 Mamba 学习到不同实例扫描序列中一致的包特征，精心设计了双分支实例序列对比学习框架，强制在两个极端不同的扫描序列之间，通过相同的 Mamba 获得的包表示及分类 logits 值保持一致性。同时，还设计了一种基于实例评估的序列对比增强方法，不仅增加了对比学习的难度，还迫使 Mamba 在输入序列差异极大的情况下，学习对 WSI 图像形成一致的表示和决策，成功赋予了 Mamba 良好的序列不变建模和判别实例识别能力。同样在 3 个 WSI 分析子任务和 12 个基准上开展的广泛系统实验结果显示，SMC-MIL 能够在让 Mamba 继承其优异计算效率和建模

能力的同时，有效降低模型对输入序列顺序的敏感性，专注于实例无序关系建模，挖掘出更具显著性判别的特征，从而更好地适配病理图像分类任务。

总体而言，本研究提出的 M<sup>2</sup>S-MIL 模型和 SMC-MIL 模型，分别从多扫描模式 Mamba 和扫描不变性 Mamba 两个维度，对数字病理图像分类任务中 Mamba 的应用进行了全面且深入的剖析与研究。M<sup>2</sup>S-MIL 模型通过综合多种扫描模式的特征信息，并引入空间上下文感知模块，显著提升了 Mamba 对 2D 空间信息的捕获能力，有效缓解了原本 Mamba 一维序列建模所造成图像空间结构捕获能力不足的问题；SMC-MIL 模型则进一步突破了扫描顺序对 Mamba 建模能力的束缚，成功弥合了原本 Mamba 的序列敏感性与病理图像多实例学习中序列不变性假设之间的冲突，使 Mamba 架构更加契合病理图像任务，同时提高了模型挖掘更具判别性区域的能力，进而提升了模型性能。

## 5.2 未来展望

本文分析了在 Mamba 架构适配病理图像分类任务方面的挑战，以数据的视觉特征和任务特点为切入点，从多扫描信息融合和构造扫描不变性 Mamba 两个角度入手，提出了基于空间信息增强的多扫描 Mamba 模型和基于差异化序列对比的扫描不变 Mamba 模型来解决高分辨率病理图像分类问题，取得了一定成果，但仍存在诸多可拓展与深入研究的方向。

在扫描优化层面，未来可进一步探索更高效的扫描模式设计。现有的区域螺旋扫描和网格扫描虽已取得良好效果，但随着病理图像复杂性的不断增加，可能需要研发更具针对性、能够更精准捕捉图像特征的扫描模式。例如，可以结合病理图像中不同组织类型、病变特征的分布特点，设计自适应的扫描模式，使 Mamba 能够更快速、准确地获取关键信息。

对于对比学习策略，也有很大的提升空间。当前的双分支实例序列对比学习框架和基于实例评估的序列对比增强方法已发挥重要作用，但可以尝试引入更复杂的对比学习算法，如结合生成对抗网络（GAN）的思想，让 Mamba 在与生成器的对抗过程中，学习到更具鲁棒性和判别性的特征表示。同时，在对比学习的过程中，如何更好地平衡模型的训练效率和性能提升，也是需要深入研究的问题。

从应用拓展角度来看，本文的研究依旧聚焦于多实例学习范式中将离线的特征进行特征聚合。未来可以迁移至特征提取，或者尝试打破范式，探索在线特征提取与实例聚合的病理图像分类方法。

此外，与临床实践的结合也是未来研究的重要方向。目前的研究更多停留在实验室环境下的模型性能验证，而将这些模型真正应用到临床诊断中，还需要解决一系列实际问题，如模型的可解释性、与现有临床诊断流程的兼容性等。通过与临床医生合作，深入了解临床需求，能够使模型的改进更贴合实际应用，为病理诊断提供更有力的支持。

在数据层面，随着医疗数据的不断积累，如何更好地利用大规模病理图像数据训练 Mamba 模型也是未来研究的重点。一方面，可以探索更有效的数据增强方法，在不改变图像真实病理特征的前提下，增加数据的多样性，提高模型的泛化能力；另一方面，如何从海量数据中筛选出对模型训练最有价值的数据，避免冗余数据对模型性能的影响，也是需要解决的问题。

最后，随着人工智能技术的不断发展，跨领域技术融合将为病理图像分类带来新的机遇。例如，将 Mamba 架构与知识图谱技术相结合，利用知识图谱中丰富的医学知识，引导 Mamba 更好地理解病理图像中的特征和关系，从而提高分类的准确性和可靠性。总之，未来在 Mamba 架构应用于病理图像分类领域的研究中，仍有广阔的探索空间，有望通过不断的技术创新和应用拓展，为病理诊断领域带来更显著的变革。

## 参考文献

- [1] 中共中央、国务院. “健康中国 2030” 规划纲要[R]. 北京, 2016.
- [2] Pinckaers H, Van Ginneken B, Litjens G. Streaming convolutional neural networks for end-to-end learning with multi-megapixel images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(3): 1581-1590.
- [3] Lu M Y, Chen T Y, Williamson D F, et al. Ai-based pathology predicts origins for cancers of unknown primary[J]. Nature, 2021, 594(7861): 106-110.
- [4] 侯乃侨, 翟炜, 郑军华. 显微知著: 人工智能在肾癌数字病理中的应用现状及挑战[J]. 中国癌症防治杂志, 2024, 16(2): 143-151.
- [5] 国务院. 新一代人工智能发展规划[R]. 北京, 2017.
- [6] 唐仲平, 崔权哲, 杨李波, 等. 全切片图像扫描技术在临床病理诊断工作中的应用[J]. 河北医科大学学报, 2018, 39(10): 5.
- [7] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial intelligence, 1997, 89(1-2): 31-71.
- [8] Herrera F, Ventura S, Bello R, et al. Multiple instance learning[M]. Springer, 2016.
- [9] Li B, Li Y, Eliceiri K W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14318-14328.
- [10] Lu M Y, Williamson D F, Chen T Y, et al. Data-efficient and weakly supervised computational pathology on whole-slide images[J]. Nature Biomedical Engineering, 2021, 5(6): 555-570.
- [11] Shao Z, Bian H, Chen Y, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification[C]//Proceedings of the Advances in Neural Information Processing Systems. 2021: 2136-2147.
- [12] Zhang H, Meng Y, Zhao Y, et al. Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18802-18812.
- [13] Tang W, Zhou F, Huang S, et al. Feature re-embedding: Towards foundation model-level performance in computational pathology[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11343-11352.
- [14] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[A]. 2020.
- [15] Hamilton J D. State-space models[J]. Handbook of econometrics, 1994, 4(1): 3039-3080.
- [16] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[A]. 2023.
- [17] Dao T, Gu A. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality[C]//Proceedings of the International Conference on Machine Learning. 2024.

- [18] Zhu L, Liao B, Zhang Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model[C]//Proceedings of the International Conference on Machine Learning. 2024: 62429-62442.
- [19] Yue Y, Li Z. Medmamba: Vision mamba for medical image classification[A]. 2024.
- [20] 白玉, 吴昊琦, 张丽丽, 等. SpiralMamba: 一种用于高光谱图像分类的轻量级 Mamba 网络 [J/OL]. 电讯技术, 2025, 15(4): 1-10. DOI: 10.20079/j.issn.1001-893x.241031001.
- [21] Yao J, Hong D, Li C, et al. Spectralmamba: Efficient mamba for hyperspectral image classification[A]. 2024.
- [22] 石洋宇, 谢承杰, 郑棣文, 等. 基于 Mamba-CNN 的多尺度异常行为检测方法[J/OL]. 北京航空航天大学学报, 2025: 1-12. DOI: 10.13700/j.bh.1001-5965.2024.0416.
- [23] 胡晓伟, 闫奕昕, 王大为, 等. 基于 YOLOM 算法的路面病害轻量化检测方法[J/OL]. 中国公路学报, 2024, 37(12): 381-391. DOI: 10.19721/j.cnki.1001-7372.2024.12.016.
- [24] Chen T, Ye Z, Tan Z, et al. Mim-istd: Mamba-in-mamba for efficient infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 6(1): 361-387.
- [25] Zhang H, Chen K, Liu C, et al. Cdmamba: Incorporating local clues into mamba for remote sensing image binary change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 5(4): 361-387.
- [26] 彭圆圆, 黎浩洋, 李文, 等. LTDA-Mamba: 基于 CNN-Mamba 混合网络的视网膜血管分割算法[J]. 光学学报, 2024, 4(2): 325-338.
- [27] 崔丽群, 郝思雅, 栾五洋. 基于 Mamba 的轻量级三维点云实例分割算法[J]. 计算机工程与应用: 1-10.
- [28] Li G, Huang Q, Wang W, et al. Selective and multi-scale fusion mamba for medical image segmentation[J]. Expert Systems with Applications, 2025, 261(2): 125518.
- [29] Wu R, Liu Y, Liang P, et al. H-vmunet: High-order vision mamba unet for medical image segmentation[J]. Neurocomputing, 2025: 129447.
- [30] Huang T, Pei X, You S, et al. Localmamba: Visual state space model with windowed selective scan[A]. 2024.
- [31] Hoffman J, Pathak D, Darrell T, et al. Detector discovery in the wild: Joint multiple instance and representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2015: 2883-2891.
- [32] Wu J, Yu Y, Huang C, et al. Deep multiple instance learning for image classification and auto-annotation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2015: 3460-3469.
- [33] 陈钰. 基于 ELDA 集成的多实例目标检测模型[D]. 上海交通大学, 2016.
- [34] Tang P, Wang X, Wang A, et al. Weakly supervised region proposal network and object detection [C]//Proceedings of the European Conference on Computer Vision. 2018: 352-368.
- [35] Xu G, Song Z, Sun Z, et al. Camel: A weakly supervised learning framework for histopathology image segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer

- Vision. 2019: 10682-10691.
- [36] 唐大伟, 鲁东明, 许端清, 等. 壁画图像分类中的分组多实例学习方法[J]. 中国图象图形学报, 2014, 19(5): 708-715.
- [37] Huang G, Huang S, Huangfu L, et al. Weakly supervised patch label inference network with image pyramid for pavement diseases recognition in the wild[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2021: 7978-7982.
- [38] Wang X, Wang B, Bai X, et al. Max-margin multiple-instance dictionary learning[C]//Proceedings of the International Conference on Machine Learning. 2013: 846-854.
- [39] Chen Y, Bi J, Wang J Z. Miles: Multiple-instance learning via embedded instance selection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 1931-1947.
- [40] Rahmani R, Goldman S A. Missl: Multiple-instance semi-supervised learning[C]//Proceedings of the International Conference on Machine Learning. 2006: 705-712.
- [41] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning [C]//Proceedings of the Advances in Neural Information Processing Systems. 2002: 458 -469.
- [42] Zhou Z H, Sun Y Y, Li Y F. Multi-instance learning by treating instances as non-iid samples[C]//Proceedings of the International Conference on Machine Learning. 2009: 1249-1256.
- [43] Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning[C]//Proceedings of the International Conference on Machine Learning. 2018: 2127-2136.
- [44] Kanavati F, Toyokawa G, Momosaki S, et al. Weakly-supervised learning for lung carcinoma classification using deep learning[J]. Scientific Reports, 2020, 10(1): 9297.
- [45] Chikontwe P, Kim M, Nam S J, et al. Multiple instance learning with center embeddings for histopathology classification[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2020: 519-528.
- [46] Campanella G, Hanna M G, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images[J]. Nature Medicine, 2019, 25(8): 1301-1309.
- [47] Zhu W, Chen X, Qiu P, et al. Dgr-mil: Exploring diverse global representation in multiple instance learning for whole slide image classification[C]//Proceedings of the European Conference on Computer Vision. 2024: 333-351.
- [48] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing Systems. 2017: 18923 -18935.
- [49] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.
- [50] Liu Y, Tian Y, Zhao Y, et al. Vmamba: Visual state space model[A]. 2024.
- [51] Fang Z, Wang Y, Wang Z, et al. Mammil: Multiple instance learning for whole slide images with state space models[A]. 2024.
- [52] Yang S, Wang Y, Chen H. Mambamil: Enhancing long sequence modeling with sequence re-ordering in computational pathology[C]//Proceedings of the International Conference on Medical

- Image Computing and Computer-Assisted Intervention. 2024: 296-306.
- [53] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the International Conference on Machine Learning. 2020: 1597-1607.
- [54] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9729-9738.
- [55] Chen X, Fan H, Girshick R, et al. Improved baselines with momentum contrastive learning[A]. 2020.
- [56] Chen T S, Hung W C, Tseng H Y, et al. Incremental false negative detection for contrastive learning[C]//Proceedings of the International Conference on Machine Learning. 2021: 21271-21284.
- [57] Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[C]//Proceedings of the Advances in Neural Information Processing Systems. 2020: 21271-21284.
- [58] Bejnordi B E, Veta M, Van Diest P J, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer[J]. JAMA, 2017, 318(22): 2199-2210.
- [59] Bandi P, Geessink O, Manson Q, et al. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge[J]. IEEE Transactions on Medical Imaging, 2018, 38(2): 550-560.
- [60] Tang W, Huang S, Zhang X, et al. Multiple instance learning framework with masked hard instance mining for whole slide image classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4078-4087.
- [61] Brancati N, Anniciello A M, Pati P, et al. Bracs: A dataset for breast carcinoma subtyping in h&e histology images[J]. Database, 2022, 2022: baac093.
- [62] Yao J, Zhu X, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks[J]. Medical Image Analysis, 2020, 65: 101789.
- [63] Chen R J, Lu M Y, Weng W H, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4015-4025.
- [64] Zhou F, Chen H. Cross-modal translation and alignment for survival analysis[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 21485-21494.
- [65] Wulczyn E, Steiner D F, Xu Z, et al. Deep learning-based survival prediction for multiple cancer types using histopathology images[J]. PLOS ONE, 2020, 15(6): e0233678.
- [66] Zhu X, Yao J, Zhu F, et al. Wsisa: Making survival prediction from whole slide histopathological images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

- Recognition. 2017: 7234-7242.
- [67] Harrell Jr F E, Lee K L, Mark D B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors[J]. Statistics in Medicine, 1996, 15(4): 361-387.
- [68] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [69] Chen R J, Ding T, Lu M Y, et al. Towards a general-purpose foundation model for computational pathology[J]. Nature Medicine, 2024, 30(3): 850-862.
- [70] Royer-Carfagni G. Can a moment-curvature relationship describe the flexion of softening beams? [J/OL]. European Journal of Mechanics - A/Solids, 2001, 20(2): 253-276. DOI: [https://doi.org/10.1016/S0997-7538\(00\)01128-1](https://doi.org/10.1016/S0997-7538(00)01128-1).
- [71] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4): 699-711.
- [72] Lin T, Yu Z, Hu H, et al. Interventional bag multi-instance learning on whole-slide pathological images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 19830-19839.
- [73] Li J, Chen Y, Chu H, et al. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11323-11332.
- [74] Huang Z, Bianchi F, Yuksekgonul M, et al. A visual–language foundation model for pathology image analysis using medical twitter[J]. Nature Medicine, 2023, 29(9): 2307-2316.

## 附录

### A. 作者在攻读硕士学位期间的论文和专利目录

- [1] Tang W, Zhou F, 导师 Huang S , 作者 **Zhu X** , et al. Feature re-embedding: Towards foundation model-level performance in computational pathology[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2024: 11343-11352.
- [2] 作者 **Zhu X** , 导师 Huang S , Tang W, et al. Adaptive Learning of Instance Representatives in Dual Spaces for Medical Image Classification[J]. Neural Computing and Applications (中科院 SCI 二区, 返修中)
- [3] 作者 **Zhu X** , 导师 Huang S , Liu B, et al. Scan-invariant Mamba with Contrastive Learning in Computational Pathology[J]. IEEE Transactions on Medical Imaging (SCI 一区, Submitted)
- [4] 导师 黄晟 , 作者 **朱翔** , 刘倍言, 董佳俊, 吴治霖, 邱可真, 洪明坚, 葛永新, 徐玲, 杨梦宁. 一种基于提示分割任务的混合监督文本检测方法: CN118781606A[P]. 2024-10-15
- [5] 导师 黄晟 , 方珩, 唐文浩, 作者 **朱翔** , 石华展, 杨俣豪, 官子焱, 洪明坚, 葛永新, 徐玲. 一种基于空间上下文感知的全视野数字切片图像分类方法: CN118570534A[P]. 2024-08-30
- [6] 导师 黄晟 , 张小先, 唐文浩, 作者 **朱翔** , 徐玲, 葛永新, 杨梦宁, 张小洪. 一种基于掩码的数字病理图像分类方法: CN116486159A[P]. 2023-07-25

### B. 作者在攻读硕士学位期间参与的科研项目

- [1] 国家自然科学基金面上项目, 少样本学习特征生成与鲁棒性关键技术研究
- [2] 重庆市自然科学基金面上项目, 文本描述协同的双向生成式少样本学习研究

### C. 学位论文数据集

| 关键词   |            | 密级     | 中图分类号         |  |  |
|---|------------|--------|---------------|--|--|
| 数字病理图像分类; Mamba 架构;<br>多实例学习; 弱监督学习;  |            | 公开     | TP            |  |  |
| 学位授予单位名称  | 学位授予单位代码   | 学位类别   | 学位级别          |  |  |
| 重庆大学  | 10611      | 专业硕士   | 硕士            |  |  |
| 论文题名  |            | 并列题名   | 论文语种          |  |  |
| 基于 Mamba 的高分辨率医学<br>病理图像分类研究  |            | /      | 汉语            |  |  |
| 作者姓名  | 朱翔         | 学号     | 202224131038t |  |  |
| 培养单位名称  |            | 培养单位代码 |               |  |  |
| 重庆大学  |            | 10611  |               |  |  |
| 学科专业  | 研究方向       | 学制     | 学位授予年         |  |  |
| 软件工程  | 计算机视觉      | 3 年    | 2025 年        |  |  |
| 论文提交日期  | 2025 年 6 月 | 论文总页数  | 68            |  |  |
| 导师姓名  | 黄晟         | 职称     | 教授            |  |  |
| 答辩委员会主席   |            |        |               |  |  |
| 电子版论文提交格式   |            |        |               |  |  |
| 文本 <input checked="" type="checkbox"/> 图像 <input type="checkbox"/> 视频 <input type="checkbox"/> 音频 <input type="checkbox"/> 多媒体 <input type="checkbox"/> 其他 <input type="checkbox"/> |            |        |               |  |  |

## 致 谢

行文至此，我的毕业论文即将画上句号，回顾这段求学历程，心中满是感慨。转眼间，从本科到硕士，已在重庆度过了七年，我的求学之路也告一段落。临近道别之际，借此机会给每一个支持和帮助过我的人们说一声感谢。

首先，感谢我的导师黄晟教授。在整个研究生阶段，导师给予了我无微不至的指导和关怀。是他严谨的工作态度与悉心的教导，让我不断进步，最终完成了这篇毕业论文。在我遇到困难和挑战时，导师总是耐心倾听、指导并鼓励我，帮助我面对困难。在此，我向导师黄晟教授表示最诚挚的感谢！

其次，我想感谢师兄师姐师弟师妹和同门们。三年来，我们一起吃饭、一起学习、一起进步。一个人是做不好科研，与人交流合作，才能明白自己的缺陷与局限。特别地，我想感谢张译师兄，如果不是张译师兄，我不会有认识到黄老师及课题组其他家人有机会，是他的温柔让我感受到兄长的关怀与照顾，进而产生想加入该课题组的想法。我还想特别感谢周锋涛师兄、唐文浩师兄和石华展师弟。与两位师兄的交流探讨常常让我受益匪浅，他们学术上的见解让我获益良多；与师弟的交流也让清晰自身的缺陷和问题。与三位的思想碰撞是我感觉三年科研生活中思想的高光。此外，还得特别感谢我的室友兼同门兼好兄弟严杰轩。他在本文的撰写过程中提供了很多建议与帮助，没有他的指导，我的文字表达将更加一塌糊涂。

另外，感谢我的家人。无论是在生活中还是在学业上，他们始终是我坚强的后盾，一直默默支持着我，尤其是我的父亲，他用他的身躯帮我抵挡了太多社会的压力与淤泥，使我能够在象牙塔般的校园中安心学习。感谢我的女朋友，与我分享快乐，遇到挫折时给予我鼓励，帮助我渡过难关，不断前行。

感谢足球，让我学会了团结与坚持，让我遇到了很多生动的人，收获了珍贵的感情。在重大虎溪操场奔跑的这几年，我把开心、低落、满足、遗憾、高兴、幽怨等情绪全都留在了这里。我会时刻记住那些瞬间，带着它们奔跑在未来更广阔的绿茵场上。

感谢自己，虽然没有追上寄予厚望的自己但也按时长大。心态更加成熟，眼界更加开阔。感谢自己做出的每一个选择、每一份努力，造就了现在的自己。

最后，感谢百忙之中参与评阅和答辩的各位专家、教授。

朱翔

二〇二五年五月 于重庆