计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(5): 1290−1304, 2026

基于拓扑先验和双分支视觉 Mamba 的皮肤病变分割

耿蕾蕾[1,2] 许振熙[1,2] 纪则轩[3] 武思寒[1,2] 范琳伟[1,2] 刘 慧[1,2]

[1]（山东财经大学计算机与人工智能学院 济南 250014）

[2]（山东省数字经济轻量智算与可视化重点实验室（山东财经大学） 济南 250014）

[3]（南京理工大学计算机科学与工程学院 南京 210094）

（leileigeng_njust@163.com）

Skin Lesion Segmentation Based on Topological Prior and Dual-Branch Vision Mamba

Geng Leilei[1,2], Xu Zhenxi[1,2], Ji Zexuan[3], Wu Sihan[1,2], Fan Linwei[1,2], and Liu Hui[1,2]

[1]（*School of Computer and Artificial Intelligence*, *Shandong University of Finance and Economics*, *Jinan* 250014）

[2]（*Shandong Key Laboratory of Lightweight Intelligent Computing and Visualization for Digital Economy* (*Shandong University of Finance and Economics*), *Jinan* 250014）

[3]（*School of Computer Science and Engineering*, *Nanjing University of Science and Technology*, *Nanjing* 210094）

Abstract Segmenting skin lesions from dermatoscopic images is crucial for the quantitative analysis and early diagnosis of skin cancer. However, automatic segmentation remains challenging due to blurred lesion boundaries, low contrast between lesions and surrounding skin, and the presence of artifacts, all of which complicate the segmentation process. Although the visual state space model based on Mamba architecture demonstrates notable advantages over Transformer models in terms of linear computational complexity and long-range dependency modeling, it still struggles to preserve the topological integrity of segmentation results. To address this issue, a topological prior guided dual-branch vision Mamba network (TDBVM) is proposed. The proposed architecture adopts a dual-encoder design to extract topological priors and deep semantic features independently. In particular, the topology branch incorporates a multi-color space topological component extraction module to generate topological prior maps. These priors are fused with visual features through a multi-scale fusion mechanism, effectively guiding feature learning, enhancing the model's ability to capture complex lesion boundaries and morphological variations, and suppressing artifacts. Experimental results on ISIC2018, ISIC2017, ISIC2016 and PH^2 datasets demonstrate that the proposed method significantly outperforms existing state-of-the-art approaches in terms of both segmentation accuracy and topological structure preservation, showing better segmentation results and robust generalization capability.

Key words skin lesion segmentation; visual state space model; topological data analysis; persistent homology; dual-branch network

收稿日期: 2025-06-26; **修回日期**: 2025-12-19

基金项目: 国家自然科学基金项目 (62572245, 62576193); 山东省自然科学基金项目 (ZR2021MF039, ZR2025MS985); 山东省泰山学者特聘专家基金项目 (tstp20221137); 山东省泰山学者基金项目 (tsqn202507240); 山东省重点研发计划项目 (2025CXPT100); 济南市人才发展专项资金项目 (202333037)

This work was supported by the National Natural Science Foundation of China (62572245, 62576193), the Natural Science Foundation of Shandong Province (ZR2021MF039, ZR2025MS985), the Special Funds of Taishan Scholars Project of Shandong Province (tstp20221137), the Taishan Scholars Project of Shandong Province (tsqn202507240), the Key Research and Development Program of Shandong Province (2025CXPT100), and the Jinan Talent Development Project - Science and Technology (202333037).

通信作者: 刘慧（liuh_lh@sdufe.edu.cn）

摘　要　皮肤病变分割对于皮肤癌的定量分析与早期诊断具有重要意义。然而，由于病灶边界模糊、对比度低及伪影干扰等因素，自动分割任务仍面临较大挑战。尽管 Mamba 架构的视觉状态空间模型凭借线性复杂度和对长距离依赖的建模能力，在效率与性能上相较 Transformer 展现出明显优势，但在保持分割结果拓扑结构完整性方面仍存在不足。为缓解上述问题，提出拓扑先验引导的双分支视觉 Mamba 网络（TDBVM）。该网络采用双编码器架构，分别提取拓扑结构先验与深度语义特征。其中拓扑分支引入多颜色空间拓扑组件提取模块以生成拓扑信息图。通过多尺度融合机制，拓扑先验有效引导视觉特征学习，增强模型对复杂边界与病灶形态的感知能力，并有效抑制伪影干扰。在 ISIC2018, ISIC2017, ISIC2016, PH^2 数据集上的实验结果表明，所提方法在分割精度与拓扑结构保持方面均显著优于现有先进方法，表现出较好的分割性能与良好的泛化能力。

关键词　皮肤病变分割；视觉状态空间模型；拓扑数据分析；持续同调；双分支网络

中图法分类号　TP391

DOI: 10.7544/issn1000-1239.202550492　　**CSTR:** 32373.14.issn1000-1239.202550492

皮肤病变是一类全球高发且类型复杂的皮肤疾病，包括黑色素瘤、基底细胞癌、脂溢性角化病等多种类型。近年来，随着发病率的不断上升，皮肤病变的早期筛查与诊断受到越来越多的关注。精确的病变区域分割是计算机辅助诊断的重要环节，不仅能够提高病变识别的准确性，还能为病灶面积量化、形态学特征分析和疾病分期提供量化依据，从而为临床医生的诊断与治疗决策提供有力支持[1]。因此，自动化分割方法已成为该领域的研究热点。然而，由于各种复杂因素的影响，皮肤病变分割仍面临诸多挑战：1)病灶边界模糊，且与背景对比度极低；2)病变区域在形态和颜色上具有高度差异性；3)存在毛发、标尺、人工标注等伪影干扰。

针对医学图像分割任务，尤其是皮肤病变分析，已有大量研究开展。早期方法主要包括经典图像处理技术和传统机器学习算法[2-5]。近年来，随着深度卷积神经网络（CNN）的发展，UNet[6] 架构凭借其能够有效保留空间细节信息的 U 形全卷积网络设计，展现出强大的分割能力，迅速成为医学图像分割的主流框架。随后，研究者们提出多种创新变体，例如：UNet++[7] 通过密集跳跃连接优化特征传递效率；ResUNet[8] 结合残差结构改善梯度传播；UNet3+[9] 通过全尺度跳跃连接实现更丰富的上下文融合；RCAR-UNet[10] 在 UNet 的跳跃连接中引入全局最大池化与全局平均池化操作融合图像的全局上下文信息。尽管 UNet 及其变体取得了显著进展，但 CNN 框架受限于局部感受野，难以充分建模全局上下文信息，限制了其分割性能的进一步提升。为弥补这一不足，研究者借鉴 Transformer[11] 在自然语言处理中的成功经验，将其引入医学图像分割任务。Transformer 的自注意力机制擅长建模长距离依赖关系，提升了模型对复杂结构和全局语义的理解能力。近年来，TransUNet[12], Swin-UNet[13], UNETR[14], MedT[15], SegFormer[16] 等基于 Transformer 的架构相继被提出，并在多个医学图像分割任务上性能优越，但效率受限于二次方复杂度。为寻求全局建模能力与计算高效性的更优结合，状态空间模型（state space model, SSM）及其变体 Mamba[17] 等新架构应运而生。Mamba 及其核心的选择性状态空间 S6 机制以线性复杂度高效建模长距离依赖，对皮肤病变分割任务具有重要意义。其中线性复杂度可显著提升高分辨率皮肤镜图像的处理效率，而长距离建模有助于更准确地捕捉病灶的整体形态和边界连贯性。针对视觉任务优化的视觉状态空间模型（visual state space model, VSS），如 Vision Mamba[18]，进一步增强了 2 维空间特征的捕获能力，这些先进的 SSM 架构被迅速引入医学图像分割领域[19-21]。然而，尽管现有深度学习方法在皮肤病分割性能上有所提升，但在精确分割模糊边界、毛发和其他伪影干扰时仍面临挑战。例如，当病灶与背景的对比度较低时，UNet[6] 难以准确界定病灶区边界，而引入注意力机制的 Attention-UNet[22] 在分割区域上方出现噪声，如图 1 示例 1 所示。进一步讲，当病灶区域内部颜色变化较大时，UNet[6] 和 Attention-UNet[22] 的分割结果内部出现空洞，如图 1 示例 2 所示。在更为复杂的场景下，特别是当边界模糊且存在毛发干扰时，UNet[6], Attention-UNet[22] 和 VM-UNet[19] 均难以获得准确的分割边界，如图 1 示例 3 所示。此外，后 2 种方法在 3 个样例上所产生的分割边界平滑度均不理想。

现有深度学习方法，包括采用先进 SSM 架构的

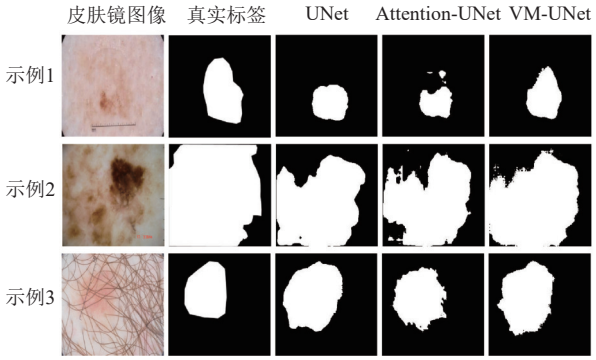


Fig. 1 An example of skin lesion segmentation

图 1 皮肤病变分割示例

模型，在处理复杂边界和保持拓扑连通性方面仍表现出明显的不足，凸显了它们在捕捉图像全局结构信息方面的局限性，例如示例 2 所示的 3 个分割结果中均存在空洞。为了克服这一挑战，拓扑数据分析(topological data analysis, TDA)为建模图像的全局结构与内在形状特征提供了强有力的数学工具[23-25]。TDA 能够捕捉数据的内在“形状”与连接关系，持续同调(persistent homology, PH)作为一个核心工具，能够在多尺度下稳定地量化拓扑特征，例如第 0 维 Betti 数量化连通分量。由持续同调生成的持久性图(persistence diagram, PD)对噪声和变形展现出显著的鲁棒性，为区分病灶与背景、保持分割结构完整性提供了独特的全局视角。鉴于持续同调的潜在优势，研究者已探索将其融入深度分割模型，其应用策略主要分为 2 类：1)设计拓扑感知的损失函数，以 Betti 数或持久性图距离作为正则化项，直接优化分割结果的拓扑一致性[26-27]，但此方式对特征学习过程的影响较为间接；2)将拓扑特征作为先验或辅助输入来补充全局结构信息，已在癌症评分预测[23]、图像分类[24]、医学图像去噪[25]等任务中展现出良好效果，并被验证能够提升分割性能[28-30]。然而，现有的融合策略往往停留在较为表层的整合，缺乏将拓扑信息作为显式结构先验进行深度嵌入，并有效引导例如 Vision Mamba[18] 这类先进高效模型特征学习的机制。

本文认为，图像具有全局、抽象和结构化的拓扑特征，这类特征与深度网络所学习到的局部化、语义化和细节导向的视觉特征之间存在显著差异。若在单一编码器中直接混合处理，可能会导致特征学习的相互干扰，难以充分发挥各自的独特优势。因此，设计一种能够独立处理并随后有效融合这些信息的网络架构至关重要。基于此考虑，本文提出了一种拓扑先验引导的双分支视觉 Mamba 网络(topological prior guided dual-branch vision Mamba network, TDBVM)，用于应对皮肤病变分割中的挑战。本文的主要工作贡献有 3 点：

1)提出了一种拓扑先验引导的双分支视觉 Mamba 网络(TDBVM)，该架构并行提取全局拓扑信息与局部视觉特征，并通过多尺度融合机制整合两者的优势，提升了分割性能。

2)提出了一个多颜色空间协同拓扑组件提取模块，该模块利用持续同调从多颜色空间信息中提取鲁棒的拓扑先验，引导网络理解低对比度或复杂形态病变的结构一致性，增强对模糊或复杂边界的感知能力。

3)在多个公开的皮肤病变分割数据集上评估了 TDBVM 的分割性能。结果表明，与现有先进方法相比，TDBVM 在 Dice 相似系数、交并比、准确率等多个分割指标上表现出明显优势，提升了皮肤病变的分割精度。

1 相关工作

1.1 基于视觉状态空间模型的医学图像分割

Mamba[17] 等状态空间模型凭借核心的选择性状态空间 S6 机制实现了线性复杂度下的高效长距离依赖建模。在此基础上，针对视觉任务优化的 VSS 迅速兴起。Zhu 等人[18] 使用双向 SSM 和位置嵌入建模视觉任务中的空间信息。Liu 等人[31] 通过设计 2 维选择性扫描等策略，将 VSS 拓展至图像分析领域。VSS 架构凭借其在全局上下文理解和计算效率方面的显著潜力，迅速在医学图像分割领域得到应用。近年来，越来越多的研究尝试将 VSS 模块与经典的 UNet 架构相结合，提出了多种混合模型。例如，Ruan 等人[19] 将 VSS 集成到 UNet 的编码-解码路径中；Chen 等人[20] 引入多尺度 VSS 表示以增强不同分辨率下的特征表达；Zhang 等人[21] 对 VSS 整体集成结构进行了优化，进一步提升了多层次特征融合效果；丁建睿等人[32] 将 VSS 模块与邻域注意力模块相结合，以强化局部与全局特征提取。这些 VSS 融合模型在多器官分割与病灶分割等任务中，通常能以较低的计算开销实现较高的分割精度。

然而，尽管这些基于 VSS 的模型在标准评估指标上表现优异，它们主要侧重于深度网络对纹理、语义信息等复杂视觉特征的高效建模。这些模型缺乏显式约束分割结果拓扑结构正确性的机制，导致其在保持目标区域的连通性、避免无关孔洞或内部断裂等方面存在不足。尤其在处理形态复杂、边界模

糊或受强伪影干扰的皮肤镜图像时，单纯依赖视觉特征提取的局限性更为突出。因此，引入拓扑数据分析作为补充性结构先验，并探索其与 VSS 模型的协同增强机制，为突破上述瓶颈提供了理论动因与技术基础。

1.2 拓扑数据分析在医学图像分析中的应用

为弥补深度学习模型对全局结构信息建模的不足，医学图像分割领域引入了 TDA。TDA 的核心工具是持续同调方法，该方法通过递增阈值 p 对图像进行子水平集滤波，在多尺度下提取图像的拓扑不变量，如 Betti 数。其中第 0 维 Betti 数 β_0 描述独立连通分量的数量，第 1 维 Betti 数 β_1 表示环路，如孔洞、空腔等。拓扑结构的演变过程被编码为持久性图，即一个 2 维散点图，图中每个点 $\sigma_i=(t_i^{\mathrm{b}},t_i^{\mathrm{d}})$ 表示一个拓扑特征自出现至消失的生命周期。据此，拓扑先验被定义为基于持久性图导出的数学约束，用于约束分割结果的解剖学结构的完整性，例如目标区域的连通性、无非病理学孔洞等。近年来，持续同调与深度学习的结合受到广泛关注，在多类医学图像任务中取得了良好效果。例如 Peng 等人[33] 对由持续同调提取的拓扑特征与深度视觉特征进行融合。Qaiser 等人[34] 在肿瘤分割中将拓扑特征与卷积神经网络的末端输出联合建模。Leventhal 等人[24] 则将拓扑结构作为先验知识嵌入图分类框架。然而，现有方法普遍采用向量化的拓扑表示，仅在浅层特征或解码器末端进行单阶段先验信息嵌入，拓扑约束无法有效贯穿于特征学习全过程。这种浅层融合策略使得拓扑引导作用呈现间接性与片段化，难以在保持全局连通性的同时精确刻画局部边界细节，反映了拓扑结构约束与视觉语义表征的融合仍面临挑战。

2 本文方法

2.1 视觉状态空间模块

本文采用 VSS 模块作为核心特征提取单元之一，用于高效捕获图像的 2 维空间特征并有效建模长距离依赖关系，其结构如图 2 所示。

VSS 模块的具体流程如下：输入特征首先进行层归一化处理，随后分成 2 个并行分支。第 1 个分支由 1 个线性层和 1 个激活函数组成。在第 2 个分支中，归一化后的特征经线性变换层处理后，通过深度可分离卷积提取局部空间特征，随后经激活函数处理，送入 SS2D 模块进行进一步的特征提取。其中 SS2D 模块的关键流程包括 3 步：首先，将输入的 2 维特征图 $\boldsymbol{X}_t\in\mathbb{R}^{H\times W\times C}$ 通过交叉扫描机制沿多个方向展开为 1 维序列 $\boldsymbol{x}_t\in\mathbb{R}^{C}$。随后，这些序列输入至选择性状态空间单元 S6，并根据如下状态更新式(1)进行建模和特征输出。

$$\boldsymbol{h}_t=\bar{\boldsymbol{A}}\boldsymbol{h}_{t-1}+\bar{\boldsymbol{B}}\boldsymbol{x}_t,\quad \boldsymbol{y}_t=\boldsymbol{C}\boldsymbol{h}_t, \tag{1}$$

其中 $\bar{\boldsymbol{A}}$ 和 $\bar{\boldsymbol{B}}$ 分别为离散状态转移矩阵与输入矩阵，均依赖于输入序列 $\boldsymbol{x}_t$，由时间步长参数 Δ 和模型原始连续参数计算得出，能够根据输入动态调节信息流，体现了模型的“选择性”特性。最后，VSS 模块通过交叉合并机制将所有方向的输出序列 $\boldsymbol{y}_t$ 重构为 2 维特征图 $\boldsymbol{Y}\in\mathbb{R}^{H\times W\times C}$。重构后的特征 $\boldsymbol{Y}$ 再次进行层归一化处理，并与第 1 个分支的特征进行逐元素相乘操

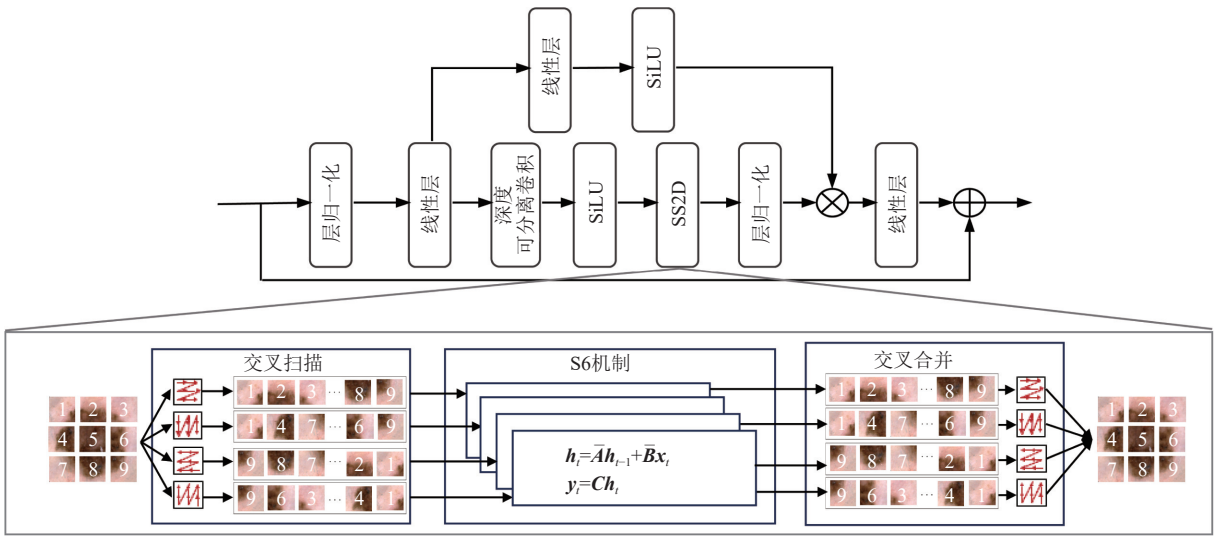


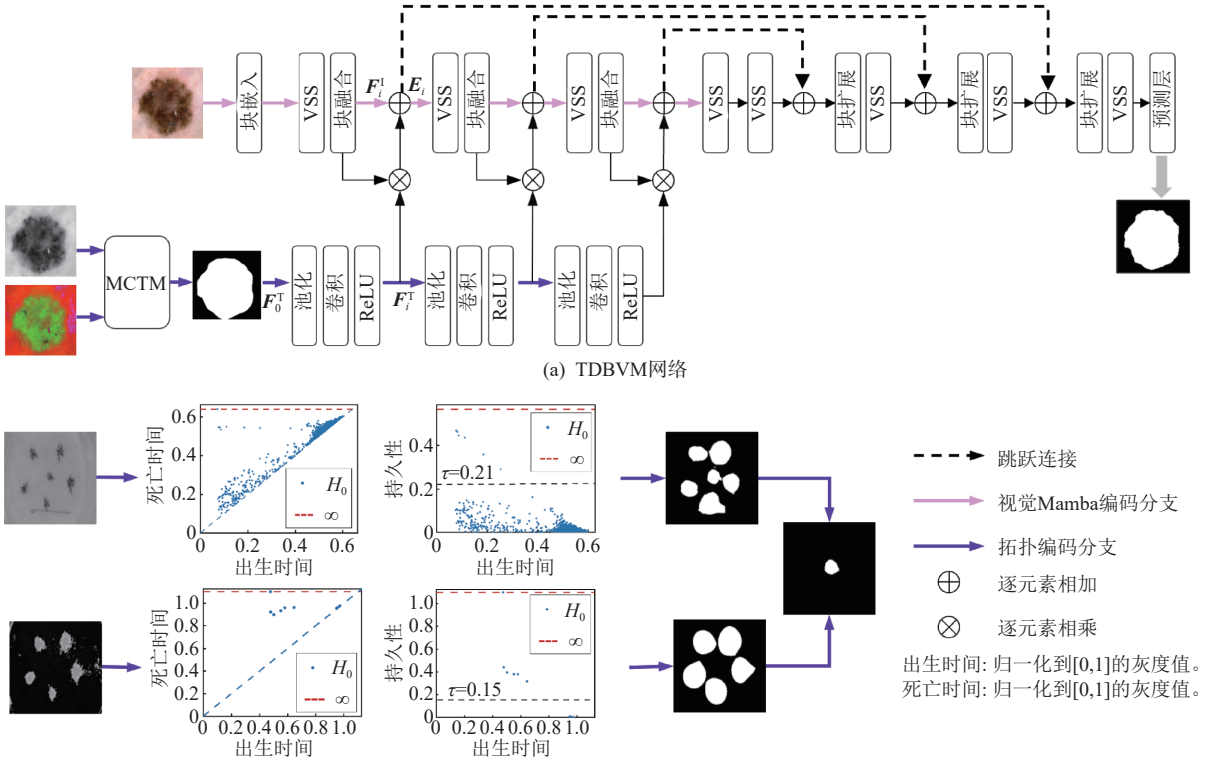
Fig. 2 Visual state space model

图 2 视觉状态空间模型

作。融合后的特征经过 1 个线性变换层与原始输入进行残差连接,最终形成 VSS 模块的输出。

2.2 TDBVM 网络

近年来,以 VSS 为核心的 Mamba 架构在多个视觉任务中的性能与计算效率之间实现了良好的平衡[35-36]。然而,基于 VSS 的分割模型主要侧重于视觉特征的高效建模,在保持分割结果的拓扑结构完整性以及精确描绘复杂空间结构方面,尤其是在处理边界模糊的皮肤病变时,仍存在一定局限性。为此,本文引入拓扑数据分析方法以提取拓扑结构先验,并提出 TDBVM 用于提升皮肤病变分割的准确性与结构一致性。TDBVM 采用 U 型架构,由拓扑先验编码分支与图像特征编码分支 2 个协同工作的编码器,以及 1 个共享解码器构成。其整体结构如图 3 所示。



(a) TDBVM网络

(b) MCTM模块

Fig. 3 Topological prior guided dual-branch vision Mamba network

图 3 拓扑先验引导的双分支视觉 Mamba 网络

TDBVM 网络输入为皮肤镜图像 $\boldsymbol{I}\in\mathbb{R}^{H\times W\times 3}$。在拓扑编码分支中,首先通过本文提出的多颜色空间协同的拓扑组件提取模块(multi-color space collaborative topological component extraction module, MCTM)从输入图像生成初始拓扑先验图 $\boldsymbol{F}_0^{\mathrm{T}}$。随后,该分支通过 4 层卷积 Conv 和池化 Pool 操作进行逐层下采样,以提取多尺度拓扑特征图序列 $\{\boldsymbol{F}_i^{\mathrm{T}}|1\leqslant i\leqslant 4\}$,更新公式为

$$\boldsymbol{F}_{i+1}^{\mathrm{T}}=ReLU\left(Conv\left(Pool\left(\boldsymbol{F}_i^{\mathrm{T}}\right)\right)\right)。\tag{2}$$

在图像特征编码分支中,本文构建了基于视觉 Mamba 的编码器提取图像的深度语义特征,其核心组件为 VSS。该编码器包含 4 个阶段,输出多尺度语义特征图序列 $\{\boldsymbol{F}_i^{\mathrm{I}}|1\leqslant i\leqslant 4\}$。为弥补语义编码过程中潜在的结构信息损失,并利用拓扑特征的全局视角进行引导,本文设计了拓扑引导的融合机制:在每个阶段,将图像语义特征图 $\boldsymbol{F}_i^{\mathrm{I}}$ 与相应拓扑特征图 $\boldsymbol{F}_i^{\mathrm{T}}$ 进行融合,生成拓扑引导的融合特征图 $\{\boldsymbol{E}_i|1\leqslant i\leqslant 4\}$,定义为:

$$\boldsymbol{E}_i=\boldsymbol{F}_i^{\mathrm{T}}\odot\boldsymbol{F}_i^{\mathrm{I}}+\boldsymbol{F}_i^{\mathrm{I}},\tag{3}$$

其中⊙表示逐像素相乘的融合操作。融合特征 $\boldsymbol{E}_i$ 随后将通过 VSS 模块进行特征增强与整合,进而得到下一层的语义特征 $\boldsymbol{F}_{i+1}^{\mathrm{I}}=VSS(\boldsymbol{E}_i)$。通过此融合机制,拓扑结构先验信息可有效引导视觉特征学习,增强模型对边界结构与复杂病灶形态的感知能力。最终,融合特征序列 $\{\boldsymbol{E}_i\}$ 将通过跳跃连接传递至解码器的相应层级,用于生成精细的分割结果。

2.3 多颜色空间协同的拓扑组件提取模块

为了提取能够反映皮肤病变核心的拓扑结构和抑制常见伪影干扰，并可作为显式先验引导后续分割的特征图，本文设计了 MCTM。该模块以持续同调理论为基础，对图像结构进行拓扑分析，并引入多颜色空间信息以提升拓扑先验的鲁棒性与准确性。主要步骤如下：

首先，对输入图像 $\boldsymbol{I}$ 的灰度版本 $\boldsymbol{I}_{\mathrm{G}}$ 应用子水平集滤波，构建立方复形序列。在此基础上，通过持续同调分析提取拓扑特征，并生成持久性图，图 4(c)(d)分别展示了皮肤镜图像在不同过滤阈值 p 下的二值化形态及其对应的持久性图。持久性图揭示了噪声点与病灶区域所对应的拓扑组件在分布特征上的差异。持久性图中的每个点 $\sigma_i=(t_i^{\mathrm{b}},t_i^{\mathrm{d}})$ 表示一个拓扑组件的生命周期 $l_i=t_i^{\mathrm{d}}-t_i^{\mathrm{b}}$，其中 t_i^{b} 和 t_i^{d} 分别为其出生时间和死亡时间。本文方法主要关注 Betti 数 β_0，即图像中的连通区域。通过筛选持续时间较长的稳定结构如病灶主体区域，同时抑制短持续时间的噪声如毛发伪影来获得更可靠的拓扑先验。随后，设定动态阈值 τ，对满足 $l_i\geqslant\tau$ 的拓扑组件进行筛选，从而构成初始显著拓扑结构集合 $\{\cup_{i=1}^{N}TopC_i^{\mathrm{I}}\}$：

$$TopC_m^{\mathrm{I}}=\{\sigma_i\,|\,l_i\geqslant\tau\},\tau=\frac{l_1+l_2}{2}=\frac{\left|t_1^{\mathrm{d}}-t_1^{\mathrm{b}}\right|+\left|t_2^{\mathrm{d}}-t_2^{\mathrm{b}}\right|}{2},\quad(4)$$

其中生命周期相差最长的 2 个拓扑组件为 σ_1 和 σ_2。随后，为进一步去除伪影并提升先验的特异性，引入颜色空间信息进行协同优化。计算输入图像色调通道 $\boldsymbol{I}_{\mathrm{H}}$ 的持续同调，同样提取生命周期超过阈值 τ 的拓扑组件集 $\{\cup_{j=1}^{N}TopC_j^{\mathrm{H}}\}$。基于大量实验和统计经验，色调通道中的显著组件多数对应非病灶干扰，如气泡、人工标记等伪影。因此，本文通过像素集求差操作，将其从初始灰度通道的拓扑组件集合中剔除，得到优化的拓扑组件集合

$$\left\{\cup_{i=1}^{N}TopC_i'\right\}=\left\{\cup_{i=1}^{N}TopC_i^{\mathrm{I}}\right\}\setminus\left\{\cup_{j=1}^{N}TopC_j^{\mathrm{H}}\right\},\quad(5)$$

$$TopC_i'=\begin{cases}\varnothing,\ \text{if}\ \dfrac{\left|TopC_i^{\mathrm{I}}\cap TopC_j^{\mathrm{H}}\right|}{\left|TopC_i^{\mathrm{I}}\right|}\geqslant t,\\ TopC_i^{\mathrm{I}},\ \text{其他情况。}\end{cases}\quad(6)$$

在实际应用中，阈值 t 一般设为 0.6。最后，将经过颜色空间优化后的拓扑组件 $\{\cup_{i=1}^{N}TopComp_i\}$ 集合生成最终的先验拓扑信息图 $\boldsymbol{F}_0^{\mathrm{T}}$。图 3(b)展示了经多源信息验证的稳定拓扑结构，通常对应潜在的病灶区域，可为 TDBVM 网络中双分支特征融合提供更具判别性的全局结构引导。

2.4 损失函数

本文使用融合交叉熵损失 L_{Bce} 与 Dice 损失 L_{Dice} 构成的混合损失函数 L_{Total}，

$$L_{\mathrm{Total}}=\lambda L_{\mathrm{Bce}}+(1-\lambda)L_{\mathrm{Dice}},\quad(7)$$

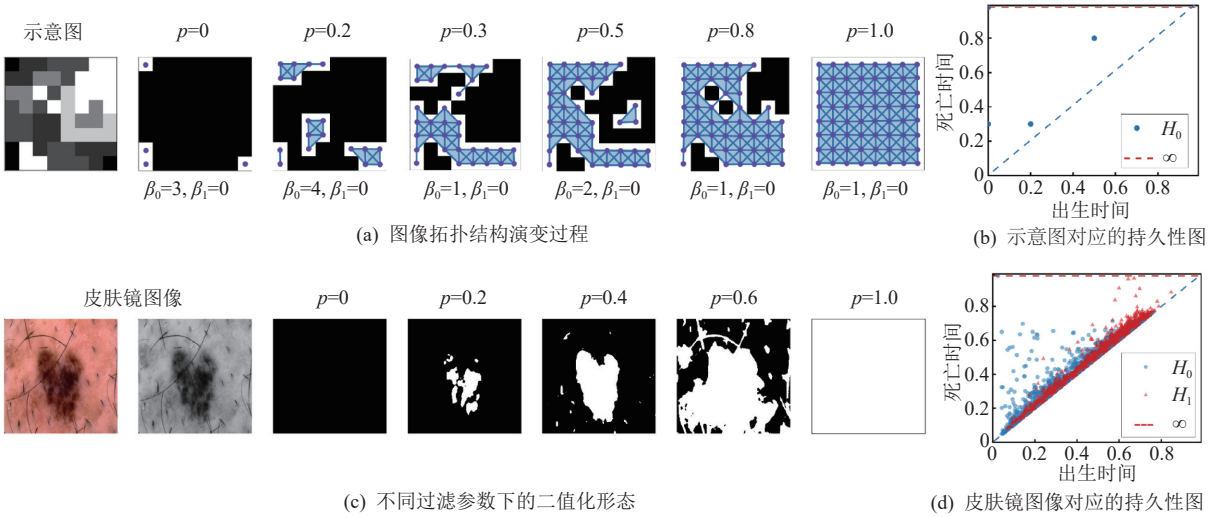
其中 λ 为超参数，$\lambda\in[0,1]$，用于平衡融合交叉熵损失 L_{Bce} 和 L_{Dice}。L_{Bce} 定义为：

$$L_{\mathrm{Bce}}=-\frac{1}{N}\sum_{i=1}^{N}\left[x_i\lg(y_i)+(1-x_i)\lg(1-y_i)\right],\quad(8)$$

而 L_{Dice} 定义为：

$$L_{\mathrm{Dice}}=1-\frac{2|\boldsymbol{X}\cap\boldsymbol{Y}|}{|\boldsymbol{X}|+|\boldsymbol{Y}|},\quad(9)$$

其中 $N=H\times W$ 为图像的总像素数，x_i 和 y_i 分别为第 i 个像素的真实标签和模型预测值；$|\boldsymbol{X}|$ 和 $|\boldsymbol{Y}|$ 分别为真实标签总数和预测结果中前景像素的总数。



(a) 图像拓扑结构演变过程　(b) 示意图对应的持久性图

(c) 不同过滤参数下的二值化形态　(d) 皮肤镜图像对应的持久性图

Fig. 4 Sequence of cubical complexes and persistence diagram

图 4 立方复形序列和持久性图

3 实验及结果分析

3.1 数据集

本文方法基于4个公开皮肤镜图像数据集开展实验,包括国际皮肤影像协同组织(International Skin Imaging Collaboration, ISIC)发布的ISIC2016[37]、ISIC-2017[38]和ISIC2018[39]数据集,以及葡萄牙生物医学图像分析实验室(Laboratory for Biomedical Image Analysis, LBIA)发布的PH^2数据集[40]。为确保评估的客观性,本文严格按照各数据集官方划分进行训练和测试,未进行跨年份的数据混合,各数据集的划分详见表1。为增强模型的泛化能力并有效缓解过拟合,训练过程中采用了随机水平翻转和随机旋转等数据增强策略。模型训练选用AdamW优化器,初始学习率设为5×10^{-4},批处理大小(batch size)为16,训练总轮次(epochs)为100。此外,网络模型VMamba-S[31]的编码器和解码器权重均基于ImageNet-1K预训练参数进行初始化,以提高特征提取效果并加快模型收敛速度。

Table 1 Division of Dataset

表1 数据集的划分

序号	名称	训练集样本数量	测试集样本数量
1	ISIC2018	1 886	808
2	ISIC2017	2 150	600
3	ISIC2016	900	379
4	PH^2	-	200

注:"-"表示未用于训练。

3.2 评价指标

为了综合评估方法的分割性能,本文采用交并比(intersection over union, IoU)、Dice相似系数(Dice similarity coefficient, DSC)、准确率(accuracy, Acc)、敏感性(sensitivity, Se)、豪斯多夫距离(Hausdorff distance, HD)、平均对称表面距离(average symmetric surface distance, ASSD)这6个指标对分割结果进行定量分析。IoU和DSC用于量化分割结果与真实标签的面积重叠程度,评估分割结果的准确性;Acc反映了所有像素被正确分类的整体比例,衡量总体分割精度;Se则量化了模型正确识别前景像素的能力,评估病灶检出能力。此外,HD和ASSD用于评估分割边界与真实边界之间的距离,评估两者轮廓的一致性,其中HD对异常值更敏感,ASSD反映了平均的边界误差。

3.3 实验结果

为了充分评估本文方法的性能,选取9种性能领先的模型进行对比实验,涵盖3类代表性架构:第1类为基于CNN的分割模型,包括经典的UNet[6]及其优化变体UNet++[7]和Attention-UNet[22]。第2类为融合注意力机制的模型,TransFuse[41]采用并行CNN与双分支架构Transformer实现多尺度特征融合;UTNetV2[42]在U型结构中引入Transformer模块增强上下文建模能力;MALUNet[43]结合4种注意力机制构建多级特征交互;DCSAU-Net[44]通过通道分割和软注意力策略提升特征表达;EGE-UNet[45]设计了基于Hadamard积的注意力机制,并引入多尺度融合方案。第3类为基于视觉状态空间建模的结构,采用VSS模块构建的VM-UNet[19]属于此类。

表2展示了不同方法在ISIC2018数据集上的IoU, DSC, Acc, Se, HD, ASSD。实验结果表明,本文方法Se位列第2,其余5项指标均取得最优,表明本文方法在边界精度和区域一致性方面具有明显优势。相较于双分支架构的TransFuse[41]和先进的VM-UNet[19],本文方法在多数指标上表现更优,仅在Se中存在微小差距。值得注意的是,本文方法在HD和ASSD上的表现最佳,验证了其在病灶边界精确定位中的有效性。

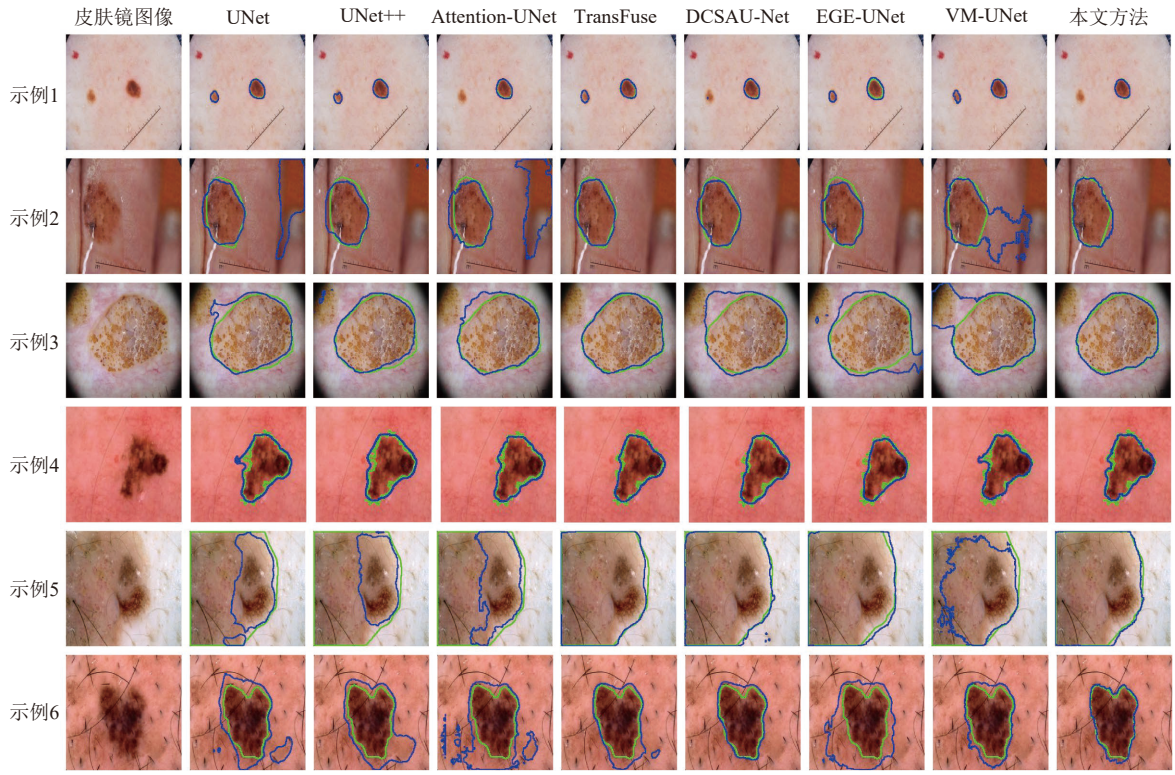
Table 2 Average Segmentation Performance Comparison of Different Methods on ISIC2018 Dataset

表2 不同方法在数据集ISIC2018上的平均分割性能比较

方法	IoU	DSC	Acc	Se	HD	ASSD
UNet	0.778 5	0.875 4	0.940 8	0.854 7	26.71	5.85
UNet++	0.784 1	0.879 0	0.942 6	0.855 7	27.78	5.73
Attention-UNet	0.786 9	0.880 7	0.942 9	0.865 0	26.54	5.81
UTNetV2	0.789 5	0.882 5	0.943 4	0.876 1	25.68	5.56
TransFuse	0.806 1	0.891 5	0.945 2	**0.912 1**	24.00	5.02
MALUNet	0.802 3	0.890 1	0.947 1	0.897 2	23.27	5.13
DCSAU-Net	0.797 7	0.887 5	0.945 9	0.876 5	22.12	5.80
EGE-UNet	0.800 6	0.889 3	0.945 4	0.901 4	24.17	5.72
VM-UNet	0.809 5	0.894 7	0.948 6	0.897 7	22.86	5.02
TDBVM	**0.820 0**	**0.901 1**	**0.951 2**	0.911 8	**21.98**	**4.93**

注:黑体数值表示最优值。

图5展示了不同方法在ISIC2018数据集部分样本上的皮肤镜图像分割可视化结果。可视化结果表明,多数方法在处理病灶与周围非病变区域颜色相似的样本时存在误分割现象。例如,图5示例1中,病灶左侧的非病变区域与病灶颜色相似,导致多数



注：绿线代表真实标签，蓝线代表分割结果。

Fig. 5 Segmentation results comparison of different models on ISIC2018 dataset

图 5 不同模型在 ISIC2018 数据集上的分割结果对比

方法将其误判为病灶区域；图 5 示例 2 中，UNet[6] 和 Attention-UNet[22] 将右侧背景噪声误识为病灶区域，而 VM-UNet[19] 则未能准确预测右侧边界。此外，分割边界模糊的样例时，CNN 框架的模型、DCSAU-Net[44]、EGE-UNet[45]、VM-UNet[19] 均表现不佳，如图 5 示例 3 所示。如图 5 示例 4，在低对比度复杂边界场景下，多数模型未能准确分割病灶左侧突起区域，而本文方法通过拓扑先验约束实现了更精确的边界定位。图 5 示例 5 和示例 6 分别反映了病灶内部颜色差异显著、毛发干扰等复杂情形，本文方法在这些干扰因素下仍能保持分割结果的结构连贯性与拓扑一致性，验证了其在复杂场景的鲁棒性。

在 ISIC2017 数据集上，本文方法在所有 6 个评价指标中均取得最优性能，具体结果见表 3。图 6 展示 3 个代表性示例的分割结果。图 6 示例 1 中毛发遮挡和手术标记干扰导致其他模型误分割病变区域，而本文方法通过拓扑约束有效抑制了干扰。图 6 示例 2 的病灶内部颜色差异显著，导致基于 CNN 的模型误将病灶分割为 2 个连通区域且边界缺失，而本文方法保持了结构完整性。图 6 示例 3 所示的极低对比度场景中，CNN 类模型和 TransFuse[41] 出现背景误分割，DCSAU-Net[44] 和 VM-UNet[19] 仅识别出部分区域，而 EGE-UNet[45] 和本文方法能分割出完整边界。上述结果表明，基于拓扑先验引导与全局拓扑结构建模的本文方法，在应对边界模糊、毛发干扰、人工标记、颜色干扰及低对比度等挑战时具有显著优势。

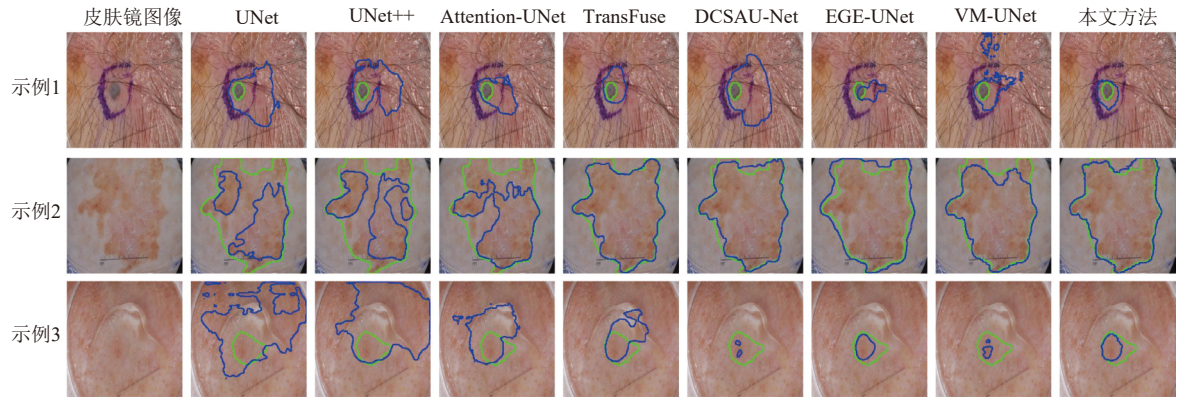
Table 3 Average Segmentation Performance Comparison of Different Methods on ISIC2017 Dataset

表 3 不同方法在数据集 ISIC2017 上的平均分割性能比较

方法	IoU	DSC	Acc	Se	HD	ASSD
UNet	0.761 5	0.864 6	0.956 1	0.836 9	27.98	6.21
UNet++	0.766 4	0.867 8	0.956 2	0.856 9	24.98	5.80
Attention-UNet	0.772 1	0.871 4	0.958 6	0.837 5	25.66	5.73
UTNetV2	0.771 9	0.871 2	0.956 2	0.847 3	25.13	5.32
TransFuse	0.793 7	0.871 9	0.954 3	0.910 2	24.00	5.05
MALUNet	0.786 2	0.880 6	0.961 1	0.846 5	24.61	5.14
DCSAU-Net	0.787 4	0.881 1	0.960 6	0.872 0	21.56	5.13
EGE-UNet	0.796 0	0.885 3	0.961 5	0.863 0	21.53	5.42
VM-UNet	0.798 4	0.887 9	0.962 4	0.888 5	21.60	5.02
TDBVM	**0.810 8**	**0.895 5**	**0.965 5**	**0.911 3**	**21.28**	**4.97**

注：黑体数值表示最优值。

ISIC2016 数据集的定量评估结果如表 4 所示，本文方法在全部 6 项指标上均取得最优性能。图 7 进



注：绿线代表真实标签，蓝线代表分割结果。

Fig. 6 Segmentation results comparison of different models on ISIC2017 dataset

图 6 不同模型在 ISIC2017 数据集上的分割结果对比

Table 4 Average Segmentation Performance Comparison of Different Methods on ISIC2016 Dataset

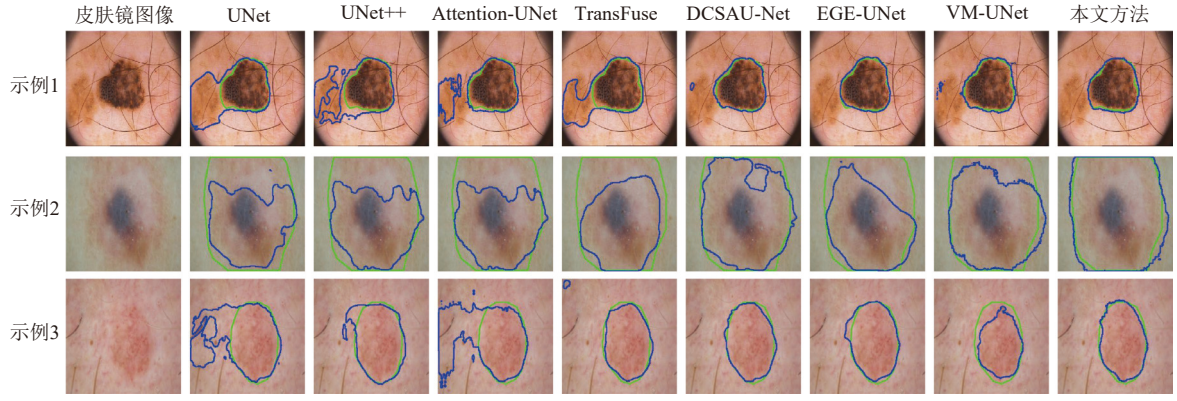
表 4 不同方法在数据集 ISIC2016 上的平均分割性能比较

方法	IoU	DSC	Acc	Se	HD	ASSD
UNet	0.828 9	0.906 4	0.948 9	0.887 4	25.38	5.87
UNet++	0.833 7	0.909 3	0.949 1	0.905 8	24.96	5.71
Attention-UNet	0.840 6	0.913 4	0.951 6	0.906 5	23.37	5.50
UTNetV2	0.847 9	0.917 7	0.952 1	0.917 3	22.96	5.24
TransFuse	0.853 3	0.915 2	0.950 9	0.925 6	22.27	4.95
MALUNet	0.859 4	0.923 6	0.958 6	0.921 8	20.61	4.73
DCSAU-Net	0.849 7	0.918 7	0.954 0	0.923 5	19.91	5.83
EGE-UNet	0..848 3	0.917 9	0.953 1	0.929 8	19.34	5.62
VM-UNet	0.869 3	0.930 1	0.960 8	0.923 9	19.21	4.52
TDBVM	**0.877 1**	**0.934 5**	**0.962 7**	**0.943 7**	**19.02**	**4.41**

注：黑体数值表示最优值。

一步展示了该数据集上的部分分割可视化结果。由于图 7 示例 1 左侧正常皮肤与病灶区颜色相近，大多数方法将该区域的全部或部分误分为病灶，而本文方法能够精准区分病灶与周围正常组织。图 7 示例 2 的病灶区域内部存在明显颜色差异，其余方法均未能识别上方病灶区域，而仍能实现病灶区域的完整分割。图 7 示例 3 存在明显毛发遮挡，且病灶与背景对比度较低，CNN 框架的模型和 EGE-UNet[45] 倾向于将左侧正常皮肤分割为病灶区域，VM-UNet[19] 在病灶上方存在分割遗漏，而本文方法依然能够准确划分完整的病灶轮廓。

本文方法在 PH^2 数据集上进行了独立测试，鉴于该数据集样本量有限，采用在 ISIC2017 数据集上预训练的模型进行测试，以验证模型的泛化能力。实验结果表明，本文方法在 IoU，DSC，Acc，HD，ASSD 共 5 个指标上均达到最优性能，在 Se 指标排名第 2，如表 5 所示。PH^2 数据集上的测试结果表明，模型在训练过程中未出现过拟合现象。值得注意的是，在对分割边界敏感的 IoU 指标上，本文方法相比次优模型提升了 1 个百分点。此外，在 HD 和 ASSD 指标上也优于现有方法，基于拓扑先验引导的双分支结



注：绿线代表真实标签，蓝线代表分割结果。

Fig. 7 Segmentation results comparison of different models on ISIC2016 dataset

图 7 不同模型在 ISIC2016 数据集上的分割结果对比

Table 5 Average Segmentation Performance Comparison of Different Methods on PH^2 Dataset

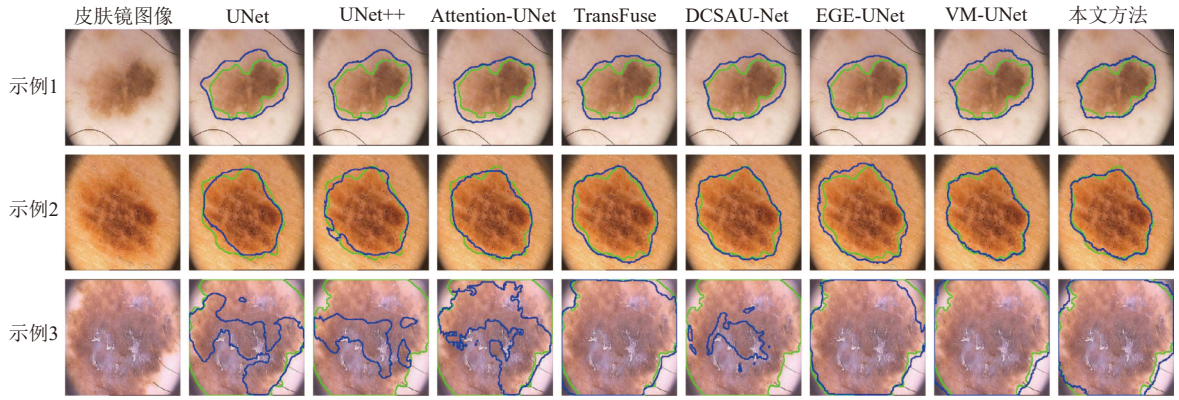
表 5 不同方法在数据集 PH^2 上的平均分割性能比较

方法	IoU	DSC	Acc	Se	HD	ASSD
UNet	0.801 8	0.883 8	0.936 2	0.882 0	24.79	7.87
UNet++	0.802 7	0.894 4	0.936 3	0.883 8	24.53	7.91
Attention-UNet	0.804 7	0.895 6	0.936 5	0.892 7	23.31	7.66
UTNetV2	0.826 9	0.900 2	0.942 5	0.910 7	21.63	6.53
TransFuse	0.834 5	0.904 7	0.937 9	**0.963 6**	21.57	6.27
MALUNet	0.842 5	0.910 6	0.941 3	0.938 7	21.21	6.73
DCSAU-Net	0.815 5	0.898 4	0.933 9	0.906 6	21.71	6.94
EGE-UNet	0.832 0	0.908 3	0.939 3	0.931 4	20.80	6.38
VM-UNet	0.853 4	0.920 9	0.947 8	0.943 2	19.08	5.71
TDBVM	**0.867 1**	**0.928 8**	**0.953 3**	0.945 5	**18.22**	**4.99**

注：黑体数值表示最优值。

构有助于提升边界分割的精度。图 8 的示例 1 和示例 2 中的病灶边界呈现复杂锯齿状结构。对比方法倾向于连接锯齿边界的顶点，从而边界过度光滑。而本文方法通过拓扑先验约束，保留了病灶的真实边界形态，图 8 示例 1 中表现尤为突出。图 8 示例 3 病灶内部特征差异显著，基于 CNN 的方法受限于局部感受野，难以实现病变区域的完整分割，TransFuse[41]，EGE-UNet[45]，VM-UNet[19] 也出现了分割错误。相比之下，本文方法有效捕捉了低对比度区域的全局结构特征。

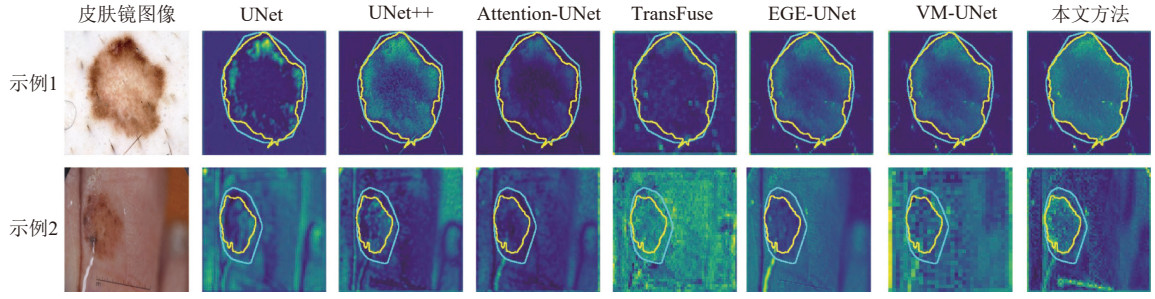
上述结果表明，本文方法在多个公开数据集及复杂场景下均展现出优越的分割能力，尤其在提升分割区域完整性与边界准确性方面具有显著优势，验证了其拓扑先验与视觉建模协同机制的有效性与泛化性。为进一步探究本文方法性能优势的来源，图 9 展示了不同方法在编码器不同层级所提取的特征图，以分析其结构感知能力与抗干扰性能。如图 9 示例 1，在浅层特征中，本文方法在拓扑先验的引导下，能够更清晰地表达病灶区域的结构边界，对毛发等伪影干扰展现出更强的鲁棒性，有效提升了前景与背景的区分能力。如图 9 示例 2，在深层特征中，本文方法能有效抑制病灶周围区域中被误判为病灶的噪声响应，表现出更优的判别能力，如对于病灶右侧区域。上述不同层级特征的可视化对比结果进一步验证了所提拓扑先验引导机制与双分支融合策略在增强模



注：绿线代表真实标签，蓝线代表分割结果。

Fig. 8 Segmentation results comparison of different methods on PH^2 dataset

图 8 不同方法在 PH^2 数据集上的分割结果对比



注：青线代表真实标签，黄线代表 MCTM 提取的拓扑组件。

Fig. 9 Visualization results of feature maps extracted by the encoder

图 9 编码器中抽取的特征图可视化结果

型结构感知能力与对干扰信息抑制方面的有效性。

3.4 统计显著性检验

为了验证本文方法在基线模型上的性能提升。本文对4个数据集的实验结果进行了统计显著性检验。本文采用配对样本 t 检验，通过分析 t 统计量和 p 值对性能差异进行了量化评估。t 统计量用于衡量2种模型间平均性能差异的大小，其值越大表明差异越显著；p 值用于衡量观察到的差异出现的概率，通常认为 $p<0.05$ 具有统计学意义。表6中的结果显示，本文方法在4个数据集上均优于所有列出的对比方法。具体而言，在4个数据集上，与 DCSAU-Net[44]，EGE-UNet[45] 和 VM-UNet[19] 的比较中，p 值亦明显小于0.05，表明性能提升在统计学上具有显著性；而与其余模型比较，p 值均远小于0.05，进一步证实了性能提升的高度显著性。综上所述，TDBVM 在4个数据集上的分割性能提升均具有统计学显著性。

Table 6 Paired *t*-Test Between TDBVM and Other Models

表6 TDBVM 与其他模型之间的配对样本 *t* 检验

方法	数据集	t 检验	p 值	显著性差异
UNet	ISIC2018	4.617 0	4.53E−06	是
	ISIC2017	7.171 9	2.02E−12	是
	ISIC2016	4.913 8	1.33E−06	是
	PH^2	6.330 2	1.59E−09	是
UNet++	ISIC2018	3.929 6	9.24E−05	是
	ISIC2017	6.113 9	1.68E−09	是
	ISIC2016	6.028 0	3.95E−07	是
	PH^2	6.498 3	6.38E−10	是
Attention-UNet	ISIC2018	4.261 4	2.27E−05	是
	ISIC2017	4.666 1	3.73E−06	是
	ISIC2016	4.945 3	1.15E−06	是
	PH^2	8.013 7	9.25E−14	是
UTNetV2	ISIC2018	3.871 6	1.21E−04	是
	ISIC2017	3.956 2	8.48E−05	是
	ISIC2016	3.610 7	3.61E−04	是
	PH^2	4.792 1	3.22E−06	是
TransFuse	ISIC2018	3.218 2	1.34E−03	是
	ISIC2017	4.251 3	2.44E−05	是
	ISIC2016	3.723 0	2.27E−04	是
	PH^2	6.027 0	7.97E−09	是
MALUNet	ISIC2018	4.150 9	3.68E−05	是
	ISIC2017	4.863 4	1.54E−06	是
	ISIC2016	4.972 1	1.18E−06	是
	PH^2	6.823 0	1.05E−10	是
DCSAU-Net	ISIC2018	2.521 8	1.19E−02	是
	ISIC2017	2.355 9	1.88E−02	是
	ISIC2016	2.237 5	2.56E−02	是
	PH^2	3.939 5	1.13E−04	是
EGE-UNet	ISIC2018	2.710 1	6.87E−03	是
	ISIC2017	2.361 2	1.83E−02	是
	ISIC2016	2.351 8	2.01E−02	是
	PH^2	5.108 1	7.59E−07	是
VM-UNet	ISIC2018	2.021 9	4.36E−02	是
	ISIC2017	2.150 3	3.19E−02	是
	ISIC2016	2.112 5	3.51E−02	是
	PH^2	3.947 1	1.09E−04	是

3.5 消融实验

3.5.1 权重参数 λ 的设置

本文对混合损失函数中权重系数 λ 的设置进行了敏感性分析，讨论其对模型分割性能的影响。保持网络结构与其余训练参数，调整交叉熵损失项与 Dice 损失项的权重比值。所采用的损失函数形式如式(8)所示，λ 为控制2项损失相对权重的超参数，取值范围为 [0,1]。本文方法以 IoU 和 DSC 作为性能评估指标。随着 λ 值的增大，IoU 与 DSC 值先升后降，当 λ=0.5 时，性能达到最优，如图10所示。结果表明，当损失权重设置均衡时，本文方法可兼顾交叉熵损失在像素级分类精度上的优势与 Dice 损失在区域优化方面的能力，从而获得更优的综合分割性能。

3.5.2 TDBVM 结构消融分析

为了验证本文提出的颜色空间协同拓扑组件提取模块 MCTM 和拓扑引导的双分支编码器的有效性及其协同作用，实验以基于视觉 Mamba 的 U 形架构为基线模型，设置了3种不同的网络模型开展消融实验，分别为不使用 MCTM 模块和双分支编码器、不使用双分支编码器、不使用 MCTM 模块，结构设置如图11所示。实验在3个公开数据集上进行，并将结果在表7、表8、表9中展示。

为评估 MCTM 模块效果，仅使用了单编码器，即在输入端将原始皮肤镜图像与 MCTM 提取的拓扑先验图 $\boldsymbol{F}_0^{\mathrm{T}}$ 通过通道拼接，并作为 Mamba 编码器的输入，结构如图11(b)所示。MCTM 能有效提取低对比度病灶的拓扑结构并抑制伪影干扰，如图12所示。因此，相较于基线模型，图11(b)模型在3个数据集上的各项指标均有所提升。在 ISIC2018 数据集上，图11(b)所示算法在 IoU 和 Se 指标上分别优化了0.61个百分点和1.38个百分点；在 ISIC2017 上，IoU 和 DSC 分别提升了0.53个百分点和0.32个百分点；在 ISIC2016 上，IoU 和 Se 分别提升了0.35个百分点和0.3个百分点。结果表明，即使仅在输入阶段引入 MCTM 生成的拓扑先验信息，也能增强模型的结构

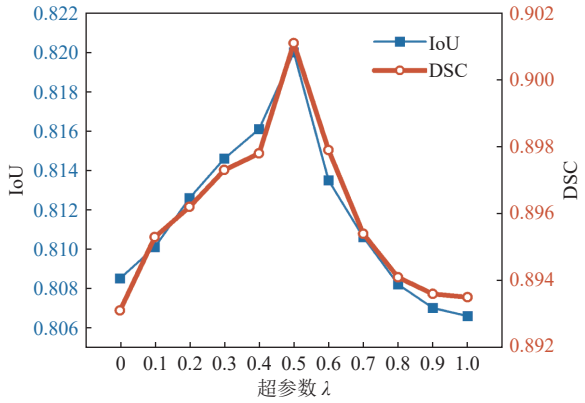


Fig. 10 Setting of weight parameter λ

图 10 权重参数 λ 的设置

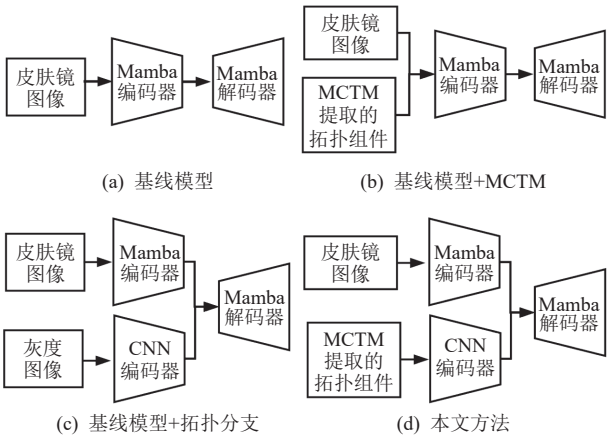


Fig. 11 Module structure of ablation study

图 11 消融实验模块结构

感知能力和分割精度。

为独立评估双分支结构的作用，图 11(c)模型在基线结构基础上引入一条并行的灰度图像 CNN 编码路径，但该路径不包含 MCTM 模块，并与 Mamba 分支的特征进行融合，结构如图 11(c)所示。实验结果

Table 7 Ablation Study Results on ISIC2018 Dataset

表 7 ISIC2018 数据集上的消融实验结果

方法	IoU	DSC	Acc	Se
基线模型	0.809 5	0.894 7	0.948 6	0.897 7
基线模型+MCTM	0.815 6	0.897 3	0.949 3	0.911 5
基线模型+拓扑分支	0.816 0	0.898 7	0.949 9	0.911 6
本文	**0.820 0**	**0.901 1**	**0.951 2**	**0.911 8**

注：黑体数值表示最优值。

Table 8 Ablation Study Results on ISIC2017 Dataset

表 8 ISIC2017 数据集上的消融实验结果

方法	IoU	DSC	Acc	Se
基线模型	0.798 4	0.887 9	0.962 4	0.888 5
基线模型+MCTM	0.803 7	0.891 1	0.963 8	0.891 5
基线模型+拓扑分支	0.804 3	0.891 6	**0.964 2**	0.880 5
本文	**0.806 5**	**0.892 9**	0.963 2	**0.913 8**

注：黑体数值表示最优值。

Table 9 Ablation Study Results on ISIC2016 Dataset

表 9 ISIC2016 数据集上的消融实验结果

方法	IoU	DSC	Acc	Se
基线模型	0.869 3	0.930 1	0.960 8	0.923 9
基线模型+MCTM	0.872 8	0.932 1	0.961 9	0.926 9
基线模型+拓扑分支	0.874 1	0.932 8	0.962 1	0.932 8
本文	**0.877 1**	**0.934 5**	**0.962 7**	**0.943 7**

注：黑体数值表示最优值。

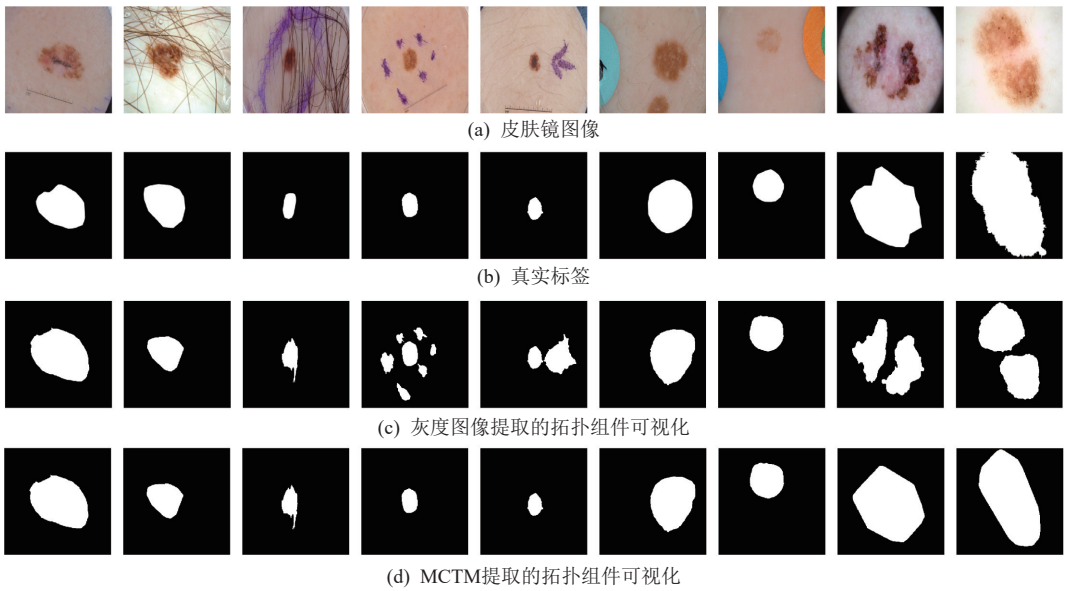


Fig. 12 Optimization of MCTM in topological components extraction from gray images

图 12 MCTM 在灰度图像拓扑组件提取中的优化

表明,该结构在部分数据集上带来性能提升,但整体提升效果具有不确定性。相较于基线模型,图 11(c)所示模型在 ISIC2018 上,IoU 提升了 0.65 个百分点,但在 ISIC2017 上,Se 却下降了 0.80 个百分点。这说明,在缺乏拓扑先验引导的情况下,双分支结构本身对模型性能的提升是有限的,在部分场景或评价指标上引发性能退化。

本文提出的完整模型 TDBVM 融合了 MCTM 模块与拓扑引导的双分支架构,结构如图 11(d)所示。实验结果表明,相比仅引入 MCTM 的模型(图 11(b))和仅采用基础双分支结构的模型(图 11(c)),TDBVM 的分割性能均有显著提升。由此验证了 MCTM 提供的拓扑先验与双分支架构实现的多尺度特征融合之间存在的协同效应,表明 TDBVM 在提升模型结构感知能力与分割鲁棒性方面具有优势和有效性。

4 结 论

本文针对皮肤病变分割中普遍存在的边界模糊、拓扑结构复杂且难以保持以及伪影干扰显著等挑战,提出了一种拓扑先验引导的双分支视觉 Mamba 网络(TDBVM)。该网络采用协同双编码器架构,其中拓扑分支利用多颜色空间协同的拓扑组件提取模块来提取鲁棒的全局拓扑先验;图像特征分支借助高效的视觉 Mamba 捕捉深度视觉语义信息。通过多层级特征融合机制,拓扑先验能够有效引导视觉特征学习过程,增强了模型对复杂边界和病灶形态结构的感知能力,并有效抑制了噪声干扰。在 ISIC2018, ISIC2017, ISIC2016, PH^2 这 4 个公开数据集上的实验结果表明,本文方法在分割精度和病灶区域拓扑结构的保持方面均优于多种现有先进方法,有力验证了拓扑先验与高效视觉状态空间模型深度协同的有效性,为复杂医学图像分割任务,特别是皮肤病变分割,提供了一种具有前景的新解决方案。

作者贡献声明:耿蕾蕾负责模型设计与分析、论文撰写和修改;许振熙负责框架设计与实现以及实验整理与分析;纪则轩负责理论指导;武思寒负责实验可视化与文字校对;范琳伟负责实验指导;刘慧提出指导意见并修改论文。

参 考 文 献

[1] Mirikharaji Z, Abhishek K, Bissoto A, et al. A survey on deep learning for skin lesion segmentation[J]. Medical Image Analysis, 2023, 88: 102863

[2] Liu Fuchang, Chen Qiang, Sun Quansen, et al. Texture classification and shape statistics variational approach for segmentation of left ventricle Tagged MR images[J]. Journal of Software, 2009, 20(1): 30−40 (in Chinese)
(刘复昌, 陈强, 孙权森, 等. 结合纹理与形状的 Tagged MR 图像左心室分割算法[J]. 软件学报, 2009, 20(1): 30−40)

[3] Xie Lizhi, Zhou Mingquan, Tian Yun, et al. Cerebrovascular segmentation based on region growing and localadaptive C-V model[J]. Journal of Software, 2013, 24(8): 1927−1936 (in Chinese)
(解立志, 周明全, 田沄, 等. 基于区域增长与局部自适应 C-V 模型的脑血管分割[J]. 软件学报, 2013, 24(8): 1927−1936)

[4] Che Na, Che Xiangjiu, Gao Zhanheng, et al. Secondary segmentation algorithm for magnetic resonance brain image based on local entropy minimization[J]. Journal of Computer Research and Development, 2010, 47(7): 1294−1303 (in Chinese)
(车娜, 车翔玖, 高占恒, 等. 基于局部熵最小化的核磁共振脑图像二次分割算法[J]. 计算机研究与发展, 2010, 47(7): 1294−1303)

[5] Chen Yunjie, Zhang Jianwei, Wei Zhihui, et al. Brain MRI segmentation using the active contours based on Gaussian mixture models[J]. Journal of Computer Research and Development, 2007, 44(9): 1595−1603 (in Chinese)
(陈允杰, 张建伟, 韦志辉, 等. 基于高斯混合模型的活动轮廓模型脑 MRI 分割[J]. 计算机研究与发展, 2007, 44(9): 1595−1603)

[6] Ronneberger O, Fischer P, Brox T. UNet: Convolutional networks for biomedical image segmentation [C] //Proc of the 18th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2015: 234−241

[7] Zhou Zongwei, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation [C] //Proc of the 4th Int Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Berlin: Springer, 2018: 3−11

[8] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C] //Proc of the 29th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770−778

[9] Huang Huimin, Lin Lanfen, Tong Ruofeng, et al. UNet3+: A full-scale connected UNet for medical image segmentation [C] //Proc of the 45th IEEE Int Conf on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2020: 1055−1059

[10] Sun Ying, Ding Weiping, Huang Jiashuang, et al. RCAR-UNet: Retinal vessels segmentation network based on rough channel attention mechanism[J]. Journal of Computer Research and Development, 2023, 60(4): 947−961 (in Chinese)
(孙颖, 丁卫平, 黄嘉爽, 等. RCAR-UNet: 基于粗糙通道注意力机制的视网膜血管分割网络[J]. 计算机研究与发展, 2023, 60(4): 947−961)

[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] //Proc of the 31st Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6000−6010

[12] Chen Jieneng, Lu Yongyi, Yu Qihang, et al. Trans-UNet: Transformers make strong encoders for medical image segmentation [J]. arXiv

preprint, arXiv: 2102.04306, 2021

[13] Cao Hu, Wang Yueyue, Chen Joy, et al. Swin-UNet: UNet-like pure Transformer for medical image segmentation [C] //Proc of the 17th European Conf on Computer Vision. Berlin: Springer, 2022: 205−218

[14] Hatamizadeh A, Tang Yucheng, Nath V, et al. UNETR: Transformers for 3D medical image segmentation [C] //Proc of the 17th IEEE Winter Conf on Applications of Computer Vision. Piscataway, NJ: IEEE, 2022: 1748−1758

[15] Valanarasu J M J, Oza P, Hacihaliloglu I, et al. Medical Transformer: Gated axial-attention for medical image segmentation [C] //Proc of the 24th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2021: 36−46

[16] Xie Enze, Wang Wenhai, Yu Zhiding, et al. SegFormer: Simple and efficient design for semantic segmentation with Transformers [C] //Proc of the 35th Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2021: 12077−12090

[17] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces [J]. arXiv preprint, arXiv: 2312.00752, 2023

[18] Zhu Lianghui, Liao Bencheng, Zhang Qian, et al. Vision Mamba: Efficient visual representation learning with bidirectional state space model [J]. arXiv preprint, arXiv: 2401.09417, 2024

[19] Ruan Jiacheng, Li Jincheng, Xiang Suncheng. VM-UNet: Vision Mamba UNet for medical image segmentation [J]. arXiv preprint, arXiv: 2402.02491, 2024

[20] Chen Chaowei, Yu Li, Min Shiquan, et al. MSVM-UNet: Multi-scale vision Mamba UNet for medical image segmentation [C] //Proc of the 18th IEEE Int Conf on Bioinformatics and Biomedicine. Piscataway, NJ: IEEE, 2024: 3111−3114

[21] Zhang Mingya, Yu Yue, Gu Limei, et al. VM-UNET-V2: Rethinking vision Mamba UNet for medical image segmentation [C] //Proc of the 20th Int Symp on Bioinformatics Research and Applications. Berlin: Springer, 2024: 335−346

[22] Oktay O, Schlemper J, Folgoc L L, et al. Attention-UNet: Learning where to look for the pancreas [J]. arXiv preprint, arXiv: 1804.03999, 2018

[23] Rammal A, Assaf R, Goupil A, et al. Machine learning techniques on homological persistence features for prostate cancer diagnosis[J]. BMC Bioinformatics, 2022, 23(1): 476

[24] Leventhal S, Gyulassy A, Heimann M, et al. Exploring classification of topological priors with machine learning for feature extraction[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 30(7): 3959−3972

[25] Malek A A, Alias M A, Razak F A, et al. Persistent homology-based machine learning method for filtering and classifying mammographic microcalcification images in early cancer detection[J]. Cancers, 2023, 15(9): 2606

[26] Hu Xiaoling, Li Fuxin, Samaras D, et al. Topology-preserving deep image segmentation [C] //Proc of the 33rd Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2019: 5657−5668

[27] He Hongliang, Wang Jun, Wei Pengxu, et al. TopoSeg: Topology-aware nuclear instance segmentation [C] //Proc of the 19th IEEE/CVF Int Conf on Computer Vision. Piscataway. NJ: IEEE, 2023: 21307−21316

[28] Haft-Javaherian M, Villiger M, Schaffer C B, et al. A topological encoding convolutional neural network for segmentation of 3D multiphoton images of brain vasculature using persistent homology [C] //Proc of the 30th IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops. Piscataway. NJ: IEEE, 2020: 990−991

[29] Gupta S, Hu Xiaoling, Kaan J, et al. Learning topological interactions for multi-class medical image segmentation [C] //Proc of the 17th European Conf on Computer Vision. Cham: Berlin: Springer, 2022: 701−718

[30] Wyburd M K, Dinsdale N K, Namburete A I L, et al. TEDS-Net: Enforcing diffeomorphisms in spatial transformers to guarantee topology preservation in segmentations [C] //Proc of the 24th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2021: 250−260

[31] Liu Yue, Tian Yunjie, Zhao Yuzhong, et al. VMamba: Visual state space model [C] //Proc of the 38th Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2024: 103031−103063

[32] Ding Jianrui, Zhang Ting, Liu Jiadong, et al. A medical video segmentation algorithm integrating neighborhood attention and state space model[J]. Journal of Electronics & Information Technology, 2025, 47(5): 1582−1595 (in Chinese)
(丁建睿, 张听, 刘家栋, 等. 融合邻域注意力和状态空间模型的医学视频分割算法[J]. 电子与信息学报, 2025, 47(5): 1582−1595)

[33] Peng Yaopeng, Wang Hongxiao, Sonka M, et al. PHG-Net: Persistent homology guided medical image classification [C] //Proc of the 19th IEEE Winter Conf on Applications of Computer Vision. Piscataway, NJ: IEEE, 2024: 7583−7592

[34] Qaiser T, Tsang Y-W, Taniyama D, et al. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features [J]. Medical Image Analysis, 2019, 55: 1−14

[35] Xing Zhaohu, Tian Ye, Yang Yijun, et al. SegMamba: Long-range sequential modeling Mamba for 3D medical image segmentation [C] //Proc of the 27th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2024: 578−588

[36] Liu Jiarun, Yang Hao, Zhou Hongyu, et al. Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining [C] //Proc of the 27th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2024: 615−625

[37] Gutman D, Codella N C F, Celebi E, et al. Skin lesion analysis toward melanoma detection: A challenge at the 13th Int Symp on Biomedical Imaging [J]. arXiv preprint, arXiv: 1605.01397, 2016

[38] Codella N C F, Gutman D, Celebi M E, et al. Skin lesion analysis toward melanoma detection: A challenge at the 14th Int Symp on Biomedical Imaging [J]. arXiv preprint, arXiv: 1710.05006, 2017

[39] Codella N, Rotemberg V, Tschandl P, et al. Skin lesion analysis toward melanoma detection: A challenge at the 15th Int Symp on Biomedical Imaging [J]. arXiv preprint, arXiv: 1902.03368, 2018

[40] Mendonça T, Ferreira P M, Marques J S, et al. PH^2—A dermoscopic image database for research and benchmarking [C] //Proc of the 35th Annual Int Conf of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE, 2013: 5437−5440

[41] Zhang Yundong, Liu Huiye, Hu Qiang. TransFuse: Fusing Trans-

formers and CNNs for medical image segmentation［C］//Proc of the 24th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2021: 14−24

［42］ Gao Yunhe, Zhou Mu, Liu Di, et al. A data-scalable Transformer for medical image segmentation: Architecture, model efficiency, and benchmark［J］. arXiv preprint, arXiv: 2203.00131, 2022

［43］ Ruan Jiacheng, Xiang Suncheng, Xie Mingye, et al. MALUNet: A multi-attention and light-weight UNet for skin lesion segmentation［C］//Proc of the 16th IEEE Int Conf on Bioinformatics and Biomedicine. Piscataway, NJ: IEEE, 2022: 1150−1156

［44］ Xu Qing, Ma Zhicheng, He Na, et al. DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation［J］. Computers in Biology and Medicine, 2023, 154: 106626

［45］ Ruan Jiacheng, Xie Mingye, Gao Jingsheng, et al. EGE-UNet: An efficient group enhanced UNet for skin lesion segmentation［C］//Proc of the 26th Int Conf on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2023: 481−490



Geng Leilei, born in 1984. PhD, associate professor, master supervisor. Member of CCF. Her main research interests include image processing, computer vision, and high-dimensional data analysis.

耿蕾蕾, 1984 年生。博士，副教授，硕士生导师。CCF 会员。主要研究方向为图像处理、计算机视觉、高维数据分析。



Xu Zhenxi, born in 2001. Master candidate. Her main research interests include image processing and computer vision.

许振熙, 2001 年生。硕士研究生。主要研究方向为图像处理、计算机视觉。



Ji Zexuan, born in 1984. PhD, associate professor, PhD supervisor. Member of CCF. His main research interests include pattern recognition, computer vision, image processing, and medical imaging.

纪则轩, 1984 年生。博士，副教授，博士生导师。CCF 会员。主要研究方向为模式识别、计算机视觉、图像处理、医学影像学。



Wu Sihan, born in 2001. Master candidate. Her main research interests include image processing and computer vision.

武思寒, 2001 年生。硕士研究生。主要研究方向为图像处理、计算机视觉。



Fan Linwei, born in 1991. PhD, associate professor, master supervisor. Member of CCF. Her main research interests include image processing and computer vision.

范琳伟, 1991 年生。博士，副教授，硕士生导师。CCF 会员。主要研究方向为图像处理、计算机视觉。



Liu Hui, born in 1978. PhD, professor, PhD supervisor. Member of CCF. Her main research interests include machine learning, data mining, and image processing.

刘　慧, 1978 年生。博士，教授，博士生导师。CCF 会员。主要研究方向为机器学习、数据挖掘、图像处理。