Modeling and Simulation 建模与仿真, 2025, 14(4), 799-810
Published Online April 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.144331

Hans汉斯

基于VMamba-CNN混合的结直肠癌切片图像分割

王劭羽[1]，陈庆奎[1]，黄　陈[2]

[1]上海理工大学光电信息与计算机工程学院，上海
[2]上海交通大学医学院附属第一人民医院胃肠外科，上海

收稿日期：2025年3月23日；录用日期：2025年4月16日；发布日期：2025年4月24日

摘　要

该研究提出一种基于VMamba和卷积神经网络(CNN)混合架构的结直肠癌(CRC)病理切片图像分割方法VMDC-Unet，旨在解决传统方法在处理肿瘤异质性、复杂背景及模糊边界时的不足。该方法通过融合VMamba模型的长距离依赖处理能力和CNN的局部特征提取优势，引入改进的ConvNext模块以增强细粒度特征提取，并设计局部自注意力机制优化跳跃连接的特征融合效率。实验结果表明，在SJTU_GSFPH和Glas数据集上，VMDC-Unet的分割精度与泛化能力均优于其他基准模型，消融实验进一步验证了各模块的有效性。该工作为医学图像分割提供了多模型协同的新思路，其结合全局依赖建模与局部特征强化的策略，为CRC精准诊疗提供了可靠的技术支持。

关键词

医学图像分割，卷积神经网络，结直肠癌，VMamba

Colorectal Cancer Slice Image Segmentation Based on VMamba-CNN Hybrid

Shaoyu Wang[1], Qingkui Chen[1], Chen Huang[2]

[1]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai
[2]Department of Gastrointestinal Surgery, The First People's Hospital Affiliated to School of Medicine, Shanghai Jiaotong University, Shanghai

Received: Mar. 23rd, 2025; accepted: Apr. 16th, 2025; published: Apr. 24th, 2025

Abstract

This study proposes a VMDC-Unet method based on a hybrid architecture of VMamba and

文章引用: 王劭羽, 陈庆奎, 黄陈. 基于 VMamba-CNN 混合的结直肠癌切片图像分割[J]. 建模与仿真, 2025, 14(4): 799-810. DOI: 10.12677/mos.2025.144331

convolutional neural network (CNN) for colorectal cancer (CRC) pathological image segmentation, aiming to address the limitations of traditional methods in handling tumor heterogeneity, complex backgrounds, and blurred boundaries. The method integrates VMamba's long-range dependency modeling capability with CNN's local feature extraction strength. It introduces an enhanced ConvNext module to improve fine-grained feature representation and designs a local self-attention mechanism to optimize feature fusion efficiency in skip connections. Experimental results demonstrate that VMDC-Unet outperformed baseline models in segmentation accuracy and generalization capability on both SJTU_GSFPH and Glas datasets. Ablation studies further verified the effectiveness of each component. The work provides a novel multi-model collaboration strategy for medical image segmentation, where the combination of global dependency modeling and local feature enhancement offers reliable technical support for precise CRC diagnosis and treatment.

Keywords

Medical Image Segmentation, Convolutional Neural Network, Colorectal Cancer, Vmamba

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

结直肠癌(Colorectal Cancer, CRC)在全球范围内的发病率和死亡率居高不下，新增病例和死亡人数在所有恶性肿瘤中位列前三[1]。在 CRC 的临床诊断中，病理切片分析是确诊和治疗计划制定的重要环节。通过病理切片图像的精确分割，医生能够更准确地识别肿瘤的边界、大小、形态等特征，从而为患者提供更为精确的治疗策略[2]。

传统的人工分割过程耗时且依赖医生的经验和主观判断，导致分割结果的一致性和可重复性受限[3]。随着医疗影像数据量的增加，医生的分析负担进一步加重，影响了诊断的准确性和临床效率[4]。随着医学图像技术的快速发展，自动化的医学图像分割方法在提高诊断效率和准确性方面显示出巨大的潜力[5]。自动化分割方法可以减少医生的主观判断误差，提高病理诊断的一致性和可重复性，对改善结直肠癌患者的预后和提高治疗效果具有重要的临床意义。

尽管传统的医学图像分割技术，如基于阈值的分割[6]、边缘检测[7]、区域生长[8]等方法，在某些情况下能够取得一定的分割效果，但它们通常依赖于图像的特定特征，在处理具有复杂背景和模糊边界的图像时，分割精度和鲁棒性往往不足[9]。随着深度学习技术的兴起，基于卷积神经网络(CNN)和 Transformer 模型的方法已经在医学图像分割领域取得了显著的进展，这些方法通过学习图像的深层特征，提高了分割的准确性[10]。然而，现有方法在处理长范围依赖和捕捉局部细节方面仍存在不足，特别是在面对肿瘤组织的微小变化和复杂的病理切片背景时，这些方法往往难以达到理想的分割效果[11]。

卷积神经网络(CNN)由于其直接从原始图像数据中学习层次化特征的能力，在医学图像分割领域取得了显著的进展，其中 U-Net [12]是一个具有里程碑意义的模型，它采用编码器–解码器结构和跳跃连接来实现精确的定位和分割。U-Net 的多个变体，如 U-Net++ [13]、Attention U-Net [14]、Res-UNet [15]和 U-Net3+ [16]，通过引入更复杂的跳跃连接、注意力机制和残差连接网络，进一步提高了分割的准确性。尽管这些基于 CNN 的方法在图像分割领域取得了显著成就，但它们仍面临着局部感受野限制，难以捕捉图像中的长距离依赖和复杂上下文信息。

Transformer 架构最初用于自然语言处理任务，但最近已被成功应用于医学图像分割，其自注意力机

制有效地模拟了图像中的长距离依赖和上下文信息。例如，TransUNet [17]结合了 CNN 和 Transformer 的优势，利用 CNN 提取局部特征，同时使用 Transformer 编码器捕捉全局上下文。Swin-UNet [18]基于 Swin Transformer，通过层次结构和滑动窗口平衡局部和全局特征提取。UCTransNet [19]重新思考了跳跃连接在 U-Net 结构中的作用，引入 Channel Transformer (CTrans)模块来替换普通的跳跃连接。nnFormer [20]是专为医学图像分割设计的 Transformer 架构，利用自注意力机制建模长距离依赖。此外，ConvNext [21]是一种新一代 CNN 架构，旨在提高图像分割任务中的效率和性能。尽管基于 Transformer 的模型在分割性能上显示出显著的改进，但它们的计算复杂性较高，需要大量的计算资源和更长的训练时间。

状态空间序列模型(SSMs) [22]在长序列数据处理中表现出色，Mamba 模型[23]通过选择机制进一步改进了 SSMs，允许模型以依赖输入的方式选择相关信息。VMamba [24]在视觉任务中展现了显著优势，不仅保留了全局特征提取能力，还显著降低了 Transformer 的计算复杂度，凸显了其在长距离依赖建模中的潜力。

本文提出了一种基于 VMamba 与 CNN 融合的结直肠癌图像分割方法，旨在克服 CNN 和 Transformer 的局限性。通过整合 VMamba 的长距离依赖建模能力和 CNN 的局部特征提取优势，我们的模型在分割精度和计算效率上取得了平衡。此外，改进的 ConvNext 模块通过细节增强卷积进一步提升了细粒度特征提取能力，而专为跳跃连接设计的局部自注意力机制则优化了特征融合过程，在降低计算复杂度的同时保持了高分割精度。

为了解决这些挑战，我们提出了一种结合 VMamba 和 CNN 的混合架构 VMDC-Unet。VMamba 模型以其通过状态空间表示处理长距离依赖的效率而闻名，它补充了 CNN 的局部特征提取能力。这种混合方法旨在利用两种模型的优势，实现卓越的分割性能。

我们的贡献有三个方面：

1) 混合 VMamba 和 CNN 结构：在结直肠癌图像分割领域引入混合 VMamba-CNN 结构的研究。通过结合这两种强大的模型，我们旨在提高分割的准确性和效率，展示在医学图像分割任务中的强烈竞争力。

2) 改进的 ConvNext 模块：我们改进了 ConvNext 模块，通过结合细节增强卷积，允许上采样模块更好地提取和恢复图像中的细粒度特征，从而提升分割的精度。

3) 用于跳跃连接的自注意力机制：我们提出了一种专为跳跃连接设计的局部自注意力机制。这种机制在减少计算复杂性的同时，改善了编码器和解码器特征的融合。我们的实验表明，这种方法不仅保持了高分割准确性，还提高了模型的效率。

2. 方法

我们提出的 VMDC-Unet 如图 1 所示，采用 U 型架构，主要模块包括编码器、跳跃连接和解码器。

编码器将输入图像分割为 4 × 4 不重叠块，并通过线性嵌入层投影到特征维度 C (默认 96)，转换后的块输入四个编码器模块中来生成特征提取，每个编码器模块包括两个 VMamba 层和一个下采样层。具体来说，Vmamaba 负责进行提取图像全局特征，下采样层负责减少输入特征的高度和通道数。每个解码器也由两个 DeConvNext 层和上采样层组成，从而将提取的上下文特征与编码器的多尺度特征进行融合。其中，DeConvNext 层负责提取图像局部特征，上采样层负责恢复特征的高度和宽度，经过四个解码器之后，使用最终投影层来恢复特征的大小，来匹配分割目标。

2.1. VSS 块

VSS 块是 VMamba 架构中的核心模块，用于捕获图像上下文信息并提取特征，其结构如图 2 所示。

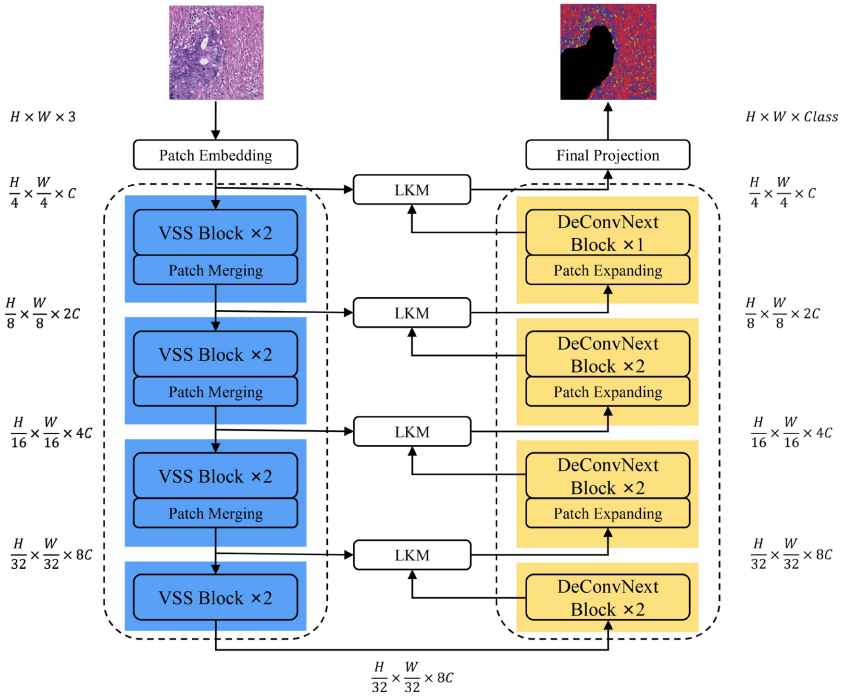

Figure 1. VMDC-Unet overall structure diagram
图 1. VMDC-Unet 整体结构图

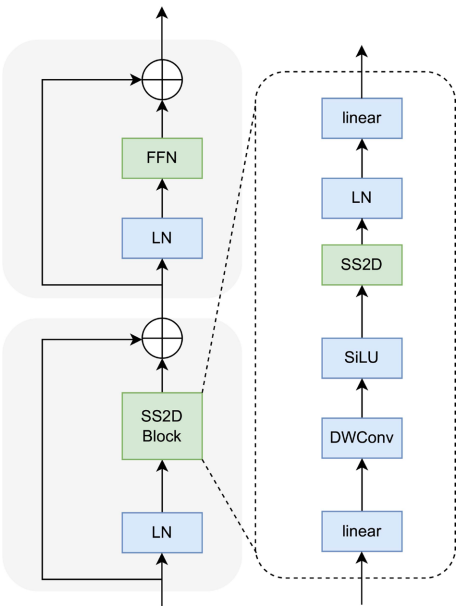

Figure 2. VSS block structure diagram
图 2. VSS 块结构图

VSS 块由一个网络分支和两个残差连接组成，输入数据首先通过一个线性层进行特征变换。随后通过深度可分离卷积层，这种卷积层首先对每个输入通道进行空间卷积，然后所有通道共享一个 1 × 1 的卷积层来混合特征，这有助于减少计算量。并通过一个 SiLU 非线性激活函数来引入非线性特性，增强模型的表达能力。随后，由 VSS 块的核心 SS2D 执行选择性扫描操作，通过交叉扫描、选择扫描、交叉合并来提

取图像的全局信息。最后，将特征通过线性层进行混合，并通过残差连接与输入相加，形成 VSS 块的输出。

SS2D 模块的设计目的是将一维选择性扫描的概念扩展到二维视觉领域，主要包括交叉扫描、S6 块、交叉合并。输入图像首先被沿四个不同的方向(从左上到右下、从右下到左上、从右上到左下、从左下到右上)展开成序列。这一步骤将图像分割成多个序列，使得每个序列都包含了图像在特定方向上的信息。

每个序列通过独立的 S6 块进行处理，S6 块受 Mamba 模型中的选择性扫描机制启发，通过调整状态空间模型(SSM)的参数，根据输入动态选择性地保留相关信息，同时过滤掉不相关的信息。

经过 S6 块处理后的序列从四个方向重新合并，以恢复输出图像到与输入相同的尺寸。这一步骤确保了图像特征的完整性，并为后续的处理步骤提供了一个统一的特征表示。

2.2. DeConvNext 块

DeConvNext 块在 ConvNextV2 模块的基础上进行了创新，引入了差分增强卷积(DeConv)，以显著提升局部特征提取能力。其核心思想在于利用多个并行卷积分支分别捕捉输入图像中的不同细节信息，并通过差分运算突出局部变化。具体来说，DeConv 模块包括原始卷积(VC)、中心差异卷积(CDC)、角度差异卷积(ADC)、水平差异卷积(HDC)和垂直差异卷积(VDC)五个分支，每个分支侧重于提取图像中不同方向和尺度的高频信息，如边缘和轮廓。通过计算这些分支输出之间的差异，模块能够显式地编码先验信息，从而更敏感地捕捉局部细节。

为了将这些并行分支高效地整合到一个统一的卷积操作中，DeConv 模块采用了重参数化技术。该技术在训练阶段保持多个并行卷积分支以充分利用它们各自的特征提取优势，而在推理阶段，通过对各分支相应位置的权重进行逐点相加，实现将多个卷积层“融合”为一个标准卷积层。这不仅简化了模型结构，降低了计算成本，同时避免了额外的参数开销，从而在保持高分割精度的同时提升了模型的运行效率。图 3 直观展示了这一过程：首先利用 DeConv 对输入数据进行预处理，通过重参数化后获得精细化的局部特征，为后续深层特征分析奠定坚实基础。随后，输入数据通过深度可分离卷积层，该层采用了 7 × 7 的大卷积核，代替了传统的 3 × 3 卷积核。这种设计选择使得网络能够在捕获更广阔范围的上下文信息的同时，减少参数数量和计算量，从而在保持高效性能的同时，降低模型的复杂性。

进一步地，特征图经过 1 × 1 的逐点卷积进行通道数的调整，这不仅有助于网络学习到更有效的特征表示，而且保持了计算的高效率。紧接着，LayerNorm 提供了一种稳定的归一化机制，为深层网络训练提供了帮助。再次通过逐点卷积和 GELU 非线性激活函数，网络能够进一步学习复杂的特征映射，这对于提高分割精度至关重要。

最终，全局响应归一化层(GRN)的应用，为网络带来了全局归一化的能力，通过规范化整个特征图的响应，增强了特征之间的对比度，使得网络能够更清晰地区分不同的特征，从而进一步提升了模型的分割性能和泛化能力。

2.3. 跳跃连接

在卷积神经网络(CNN)的设计中，跳跃连接是一种常见的技术，用于缓解深层网络中的梯度消失问题，并增强特征的传递。传统上，跳跃连接通过两种方式实现：一种是 concatenation，即将不同层的通道直接拼接起来；另一种是 addition，即将特征图进行简单的相加。然而，concatenation 可能会增加特征的冗余性，导致网络需要额外的努力来识别并忽略不重要的特征。而 addition 的方式虽然简单，但可能不足以有效地整合复杂的特征信息。

为了解决这些问题，本文提出了一种基于局部自注意力机制的改进方法，如图 4 所示。自注意力机制能够捕捉序列内部的长距离依赖关系，通过计算序列中每个元素对其他所有元素的注意力权重，实现

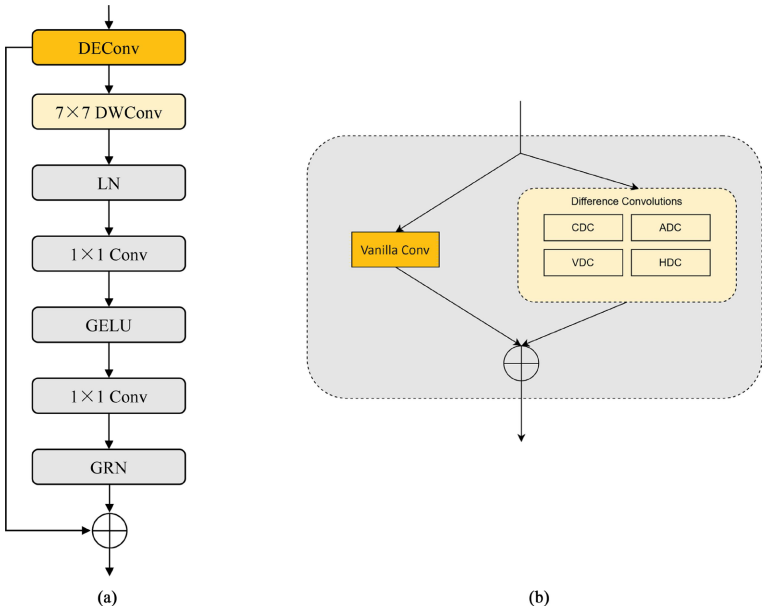

Figure 3. (a) DeConvNext block diagram; (b) DeConv structure diagram
图 3. (a) DeConvNext 块结构图；(b) DeConv 结构图

上下文特征的有效融合。具体来说，模型为序列中的每个元素生成查询(Q)、键(K)和值(V)，这些是通过不同的线性变换得到的。然后，模型计算查询与所有键的兼容性，通常采用点积操作来实现，并使用 softmax 函数对得到的注意力矩阵进行归一化，以获得每个元素对其他元素的注意力权重。最后，利用这些权重对值进行加权求和，得到加权的平均表示，这个表示将作为当前元素的更新。

在本文的方法中，我们采用了一种创新的策略：利用上采样层的局部特征来生成查询和键，而将下采样层的全局特征作为值。通过这种方式生成注意力矩阵，并且根据该注意力矩阵来提取下采样层传过来的全局特征。我们不仅能够保留局部特征的细节，还能够更好的融合全局上下文信息，提高模型对复杂特征的处理能力。

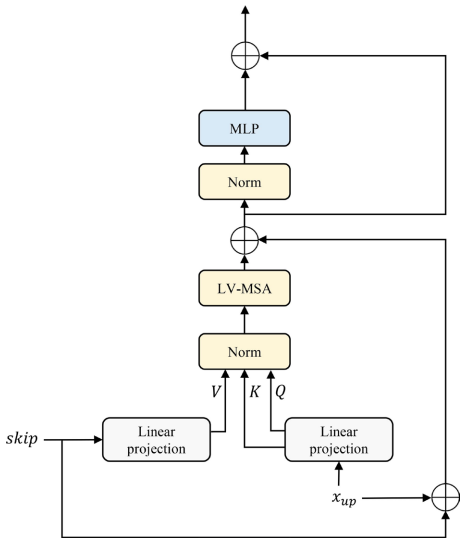

Figure 4. Skip connection structure diagram
图 4. 跳跃连接结构图

3. 实验

3.1. 数据集

本文采用了两个不同的数据集，分别为上海交通大学附属第一人民医院胃肠外科提供的数据集(简称SJTU_GSFPH)以及一个公开的结直肠癌分割挑战数据集(简称 Glas)。

SJTU_GSFPH 数据集包含了 2014 年 1 月至 2018 年 12 月期间经术后病理证实的II、III期 CRC 患者的临床资料、术后病理和随访结果等。其中左半结肠和直肠癌共 546 例，黏液腺癌 103 例，腺癌 717 例。该数据集共有 996 张 HE 切片图像，图像尺寸为 1276 × 689，标注图包括四个类别，其中，黑色表示肿瘤实质、蓝色表示间质细胞、红色表示间质中的胶原等(非细胞)成分、黄色表示间质中的坏死。每张图像被对半切分后。该数据集数量被扩充至 1992 张，随后被分为 1400 个训练图像和 592 个测试图像。

Glas 数据集是一个专门针对结直肠癌组织切片图像的医学图像分割挑战数据集，它包含了 165 张来自 16 张不同患者 H&E 染色的全幻灯片图像(WSIs)。这些图像均采用 MIRAXMIDI 幻灯片扫描仪以 20 倍物镜放大和 0.465 微米像素分辨率进行扫描，以获取高清晰度的图像数据。数据集中的每张图像尺寸均为 775 × 522 像素，并提供了详细的实例分割标注，精确地标识出每个腺体的边界和流明区域。为了支持模型的训练与评估，这 165 张图像被划分为 85 张训练图像，80 张测试图像。

3.2. 实现细节

输入图像被缩放至 256 × 256 分辨率，训练模型 batch 大小为 8，采用 AdamW 优化器，初始学习率为 1e−3。采用余弦退火学习率调节算法(Cosine Annealing Learning Rate)作为调度器，最大迭代次数为 50 次，最小学习率为 1e−5。训练次数设置为 300 次。实验使用 pytorch 框架在单个 NVIDIA RTX4070 GPU 上进行，实验采用 DICE、IOU 和 Hausdorff 距离(HD95)作为我们的模型评估指标。

3.3. 实验结果及性能分析

3.3.1. SJTU_GSFPH 数据集

为了全面评估所提出的 VMDC-Unet 模型在结直肠癌切片图像分割任务中的有效性，本文将其与多种基于 CNN 和 Transformer 的先进模型进行了比较，包括 U-Net 及其变体(如 R34-U-Net、UNet++、Attention-Unet)以及基于 Transformer 的模型(如 TransUnet, Swin-Unet、UCTransNet、nnFormer)，同时以 VMUnet 作为基准模型。

实验结果显示，我们提出的 CNN-VMamba 混合方法(即 VMDC-Unet)在两个关键的分割性能指标——平均交并比(MIOU)和 Dice 系数上均取得了最佳性能。具体来说，VMDC-Unet 的 MIOU 达到了 79.4%，Dice 系数达到了 88.51%，显著优于其他比较模型。这一结果表明，通过结合 CNN 和 VMamba 网络的优势，VMDC-Unet 能够更准确地识别和分割结直肠组织切片中的腺体结构。

从表 1 的对比结果可以看出，经典的 U-Net 模型表现最弱，引入残差网络的 R34-U-Net 和多尺度连接的 UNet++在性能上有所改进，但提升有限。Attention-Unet 通过注意力机制显著增强了目标区域的分割能力。基于 Transformer 的模型(如 TransUnet、Swin-Unet 和 UCTransNet)进一步提升了分割性能，其中 TransUnet 的 MIOU 达到 78.14%，Dice 系数为 87.52%，展现了全局信息建模能力的重要性。相比之下，VMUnet 使用 VMamba 模块，性能进一步提升至 MIOU 为 78.42%，Dice 系数为 87.95%，HD95 降至 12.87。最终，提出的 VMDC-Unet 模型通过对网络结构的优化，实现了最佳分割性能，其 MIOU 和 Dice 系数分别提升了 1%和 1.6%，HD95 降低至 10.95。这一结果验证了 VMDC-Unet 在复杂腺体分割任务中的有效性和优越性。

图 5 展示了不同模型在 SJTU_GSFPH 数据集上的可视化分割结果。通过直观比较，我们可以观察到

VMDC-Unet 在细节处理和边界识别上的优势。在复杂的组织结构和腺体边界模糊的情况下，VMDC-Unet 能够生成更加平滑和准确的分割边缘，减少了对肿瘤和间质的分割误差。

Table 1. Comparison results of different models on the SJTU_GSFPH dataset
表 1. 不同模型在 SJTU_GSFPH 数据集上的对比结果

Method	Avg MIOU ↑	Avg DICE ↑	Avg HD95 ↓
U-Net	61.65	70.21	24.73
R34-U-Net	63.30	72.83	23.58
UNet++	64.12	72.95	23.07
Attention-Unet	68.61	75.36	20.51
TransUnet	78.14	87.52	15.26
Swin-Unet	77.82	86.73	16.71
UCTransNet	78.03	87.23	15.82
nnFormer	77.65	86.29	14.77
VMUnet	78.42	87.95	12.87
VMDC-Unet	**79.40**	**89.51**	**10.95**

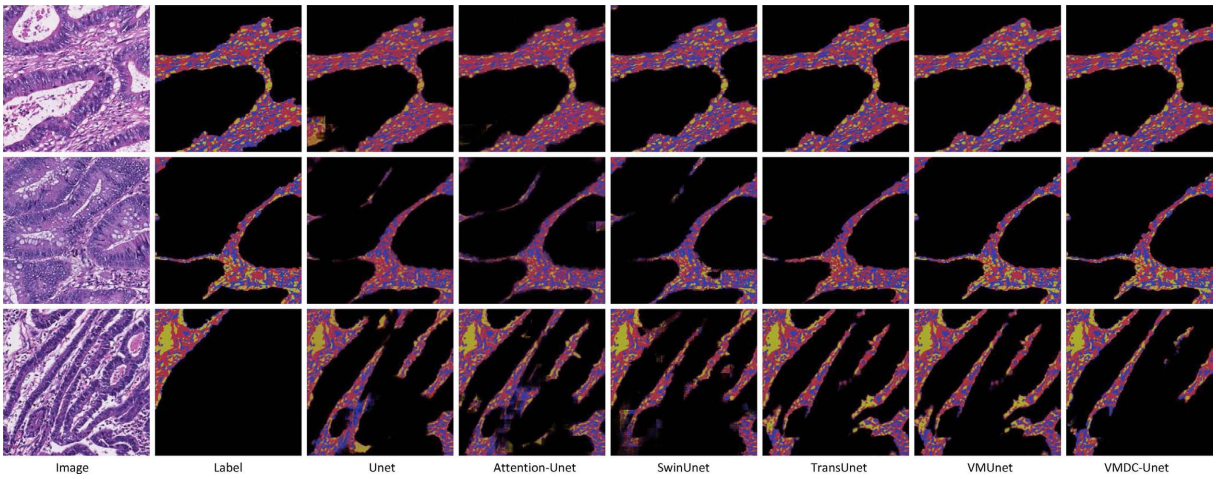

Figure 5. Visualization of segmentation effects of different models on the SJTU_GSFPH dataset
图 5. 不同模型在 SJTU_GSFPH 数据集上的分割效果可视化

3.3.2. Glas 数据集

为了进一步验证所提出 VMDC-Unet 模型的泛化能力，我们在另一个公开的结直肠癌组织切片数据集——Glas 数据集上进行了评估。实验结果如表 2 所示。与 SJTU_GSFPH 数据集的表现一致，VMDC-Unet 在 Glas 数据集上依然展现了最优性能，其平均交并比(MIOU)和 Dice 系数分别达到了 83.21%和 91.63%，同时 HD95 指标进一步降低至 6.74，充分体现了其在不同数据集上的优越性和出色的泛化能力。

从表 2 的对比结果可以看出，U-Net 的基础性能较弱，而引入残差模块的 R34-U-Net 和多尺度设计的 UNet++提升有限。Attention-Unet 和基于 Transformer 的模型(如 TransUnet 和 Swin-Unet)在捕获全局和局部特征方面表现优异，其中 Swin-Unet 达到 MIOU 82.13%和 Dice 89.62%。VMUnet 通过全局采用 VMamba 模块进一步提升，MIOU 和 Dice 分别为 82.87%和 91.09%。最终，VMDC-Unet 获得了最佳性

能，并验证了其卓越的泛化能力。

图 6 展示了不同模型在 Glas 数据集上的可视化分割结果。可视化结果与 SJTU_GSFPH 数据集相仿，VMDC-Unet 生成了更好的分割结果，比其他基准模型的结果相比更接近真实值。可以很容易地看出，我们提出的方法不仅突出了正确的显著区域，消除了混淆的假阳性病灶，而且还产生了连贯的边界。这些观察表明，VMDC-Unet 能够在保留细节形状信息的同时进行更精细的分割。

值得注意的是，在两个不同的数据集上，Swin-Unet 和 Transunet 的表现不同。在 SJTU_GSFPH 数据集上，由于该数据集腺体结构复杂且背景干扰较多，全局注意力机制能够更好地捕捉全局语义上下文关系，因此 TransUnet 表现更优。而 Glas 数据集中腺体的边界更加规则、纹理较为清晰，因此 Swin-Unet 的局部特性提取能力得以充分发挥。而结合了特征提取和全局语义捕捉的 VMDC-Unet 则可以在两个数据集都有不错的效果。

综上所述，无论是在 SJTU_GSFPH 数据集还是 Glas 数据集上，VMDC-Unet 都展现出了卓越的分割性能和泛化能力。这些结果表明，VMDC-Unet 不仅适用于特定的数据集，而且能够很好地处理来自不同患者和不同染色条件下的图像。未来的工作将集中在进一步优化模型结构，探索新的训练策略，并在更多的数据集上进行评估，以实现更高的分割精度和更好的临床应用前景。

Table 2. Comparison results of different models on the Glas dataset
表 2. 不同模型在 Glas 数据集上的对比结果

Method	Avg MIOU ↑	Avg DICE ↑	Avg HD95 ↓
U-Net	74.71	85.46	15.30
R34-U-Net	76.22	87.15	10.27
UNet++	77.03	87.56	12.72
Attention-Unet	80.63	88.80	9.10
TransUnet	80.4	88.43	8.27
Swin-Unet	82.13	89.62	9.00
UCTransNet	82.25	90.18	7.52
VMUnet	82.87	91.09	7.52
VMDC-Unet	**83.21**	**91.63**	**6.74**

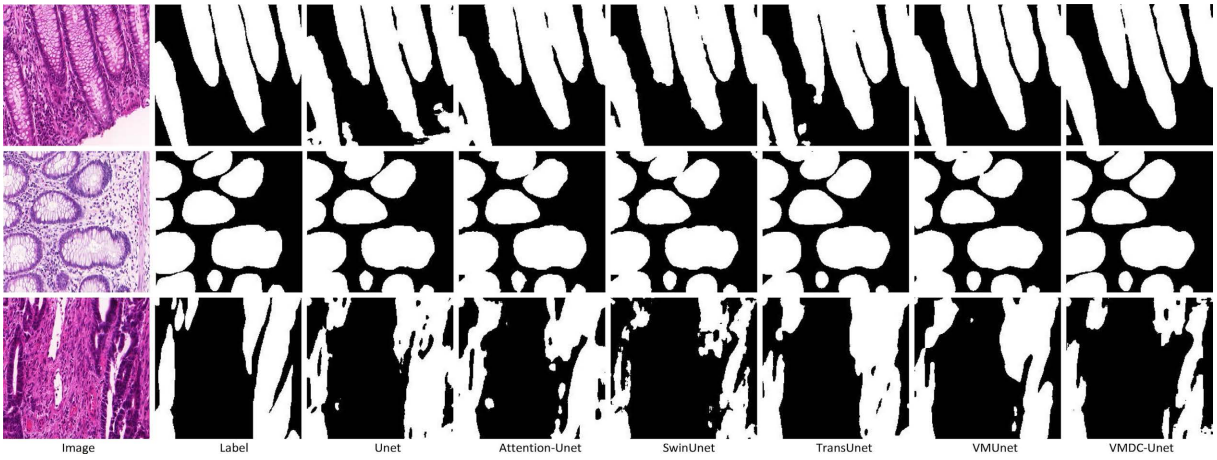

Figure 6. Visualization of segmentation effects of different models on the Glas dataset
图 6. 不同模型在 Glas 数据集上的分割效果可视化

3.4. 消融实验

为了全面评估所提出的 VMDC-Unet 模型中各个组件的贡献，本文进行了一系列的消融实验。这些实验特别关注了对 ConvNext 模块的改进——DeConvNext 模块，以及在跳跃连接中引入的自注意力机制(Self-Att)的有效性。通过这些实验，我们旨在验证这些技术是否真正增强了模型的分割性能，并探索它们在处理复杂生物医学图像时的潜力。

实验结果如表 3 所示，从中观察可得，将原始的 ConvNext 模块替换为 DeConvNext 模块，无论是否引入自注意力机制，模型的性能都有轻微提升，特别是在 SJTU_GSFPH 数据集上。DeConvNext 模块在特征恢复阶段的表现更为突出，表明其在特征重建方面的效果可能更加显著。

进一步分析发现，在所有配置中，引入自注意力机制的模型均比简单的相加操作表现得更好，尤其在图像细节和上下文信息的处理上展现出了明显的优势。特别是，当 DeConvNext 模块和自注意力机制结合使用时，在 MIOU 和 Dice 系数上均达到了最佳性能，分别为 79.4%和 88.51% (SJTU_GSFPH 数据集)，以及 83.21%和 91.63% (Glas 数据集)。这一结果进一步验证了 DeConvNext 模块和自注意力机制在提升分割精度方面的有效性与互补性。

表 4 展示了不同模型的参数量和浮点运算(Flops)，为模型的计算复杂度和资源效率提供了直观对比。可以看出，VMDC-Unet 模型拥有最高的参数量(197.47 M)，但其 FLOPS 却是最低的(6.66 亿次浮点运算)，这得益于 VMamba 模块，它拥有与 Transformer 类似的全局特征提取能力的同时大幅降低了计算复杂度。通过将 VMamba 与 CNN 结合，VMDC-Unet 在参数量与计算效率之间达成了最优折中，表现出了不错的性能和泛化能力。

Table 3. Ablation experiment results
表 3. 消融实验结果

Method				SJTU_GSFPH		Glas	
BaseLine	ConvNext	DeConv	Self-Att	MIOU	DICE	MIOU	DICE
√				78.41	87.90	82.87	91.09
√	√			79.02	88.25	82.66	90.72
√	√	√		79.21	88.43	83.10	91.35
√	√		√	79.23	88.47	83.02	91.14
√	√	√	√	79.40	88.51	83.21	91.63

Table 4. Total model parameters and total number of floating point operations in one forward propagation
表 4. 模型总参数量和一次向前传播的浮点运算总数

Method	Params (M)	Flops
U-Net	25.81	29.02
TransUnet	105.32	32.25
Swin-Unet	27.18	7.74
VMDC-Unet	**197.47**	**6.66**

4. 结论

本文提出的 VMDC-Unet 模型在结直肠癌病理切片图像分割任务中展现了卓越的性能。通过结合

VMamba 和 CNN 的优势，该模型不仅提高了分割的准确性，还增强了对长距离依赖的处理能力。通过引入改进的 ConvNext 模块和自注意力机制，VMDC-Unet 在细节恢复和特征融合方面表现出色，显著提升了分割精度。

实验结果表明，在 SJTU_GSFPH 和 Glas 两个数据集上，VMDC-Unet 均取得了最佳的 MIOU 和 DICE 系数，同时 HD95 指标显著降低，充分验证了模型在不同数据集上的优异泛化能力和高效性。此外，通过消融实验评估了模型中各关键组件的作用，验证了 DeConvNext 模块和自注意力机制在提升局部特征提取能力和增强模型感知细节及上下文信息方面的重要性。实验结果表明，这些创新模块显著改善了分割精度，并且通过重参数化技术的应用，有效降低了计算开销，提高了整体模型的运行效率。

综上所述，VMDC-Unet 在分割精度、模型泛化能力和计算效率之间实现了良好的平衡，为结直肠癌病理切片图像分割提供了一种创新而有效的解决方案。未来工作将集中在进一步优化网络结构、探索更高效的训练策略以及在更多临床数据集上的验证，力求为结直肠癌的精确诊疗提供更可靠的技术支持。

致　谢

本研究受到了国家自然科学基金项目(61572325)；上海市重点科技项目(19DZ1208903)的资助，本文作者团队对上述项目的支持表示感谢。

基金项目

国家自然科学基金项目(61572325)；上海市重点科技项目(19DZ1208903)。

参考文献

[1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **71**, 209-249. https://doi.org/10.3322/caac.21660

[2] Rawla, P., Sunkara, T. and Barsouk, A. (2019) Epidemiology of Colorectal Cancer: Incidence, Mortality, Survival, and Risk Factors. *Gastroenterology Review*, **14**, 89-103. https://doi.org/10.5114/pg.2018.81072

[3] Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A. and Sethi, A. (2017) A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Transactions on Medical Imaging*, **36**, 1550-1560. https://doi.org/10.1109/tmi.2017.2677499

[4] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., *et al.* (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[5] Thakur, N., Yoon, H. and Chong, Y. (2020) Current Trends of Artificial Intelligence for Colorectal Cancer Pathology Image Analysis: A Systematic Review. *Cancers*, **12**, Article 1884. https://doi.org/10.3390/cancers12071884

[6] Otsu, N. (1975) A Threshold Selection Method from Gray-Level Histograms. *Automatica*, **11**, 23-27.

[7] Canny, J. (1986) A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679-698. https://doi.org/10.1109/tpami.1986.4767851

[8] Adams, R. and Bischof, L. (1994) Seeded Region Growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 641-647. https://doi.org/10.1109/34.295913

[9] 殷晓航, 王永才, 李德英. 基于 U-Net 结构改进的医学影像分割技术综述[J]. 软件学报, 2021, 32(2): 519-550.

[10] 张玮智, 于谦, 苏金善, 等. 从 U-Net 到 Transformer: 深度模型在医学图像分割中的应用综述[J]. 计算机应用, 2024, 44(z1): 204-222.

[11] 陈哲, 童基均, 潘哲毅. 基于 Attention-ResUNet 的肝脏肿瘤分割算法[J]. 计算机时代, 2023(10): 100-104.

[12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[13] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al.*, Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[14] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999.

[15] Xiao, X., Lian, S., Luo, Z. and Li, S. (2018) Weighted Res-UNet for High-Quality Retina Vessel Segmentation. 2018 9*th International Conference on Information Technology in Medicine and Education* (*ITME*), Hangzhou, 19-21 October 2018, 327-331. https://doi.org/10.1109/itme.2018.00080

[16] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., *et al.* (2020) UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *ICASSP* 2020—2020 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Barcelona, 4-8 May 2020, 1055-1059. https://doi.org/10.1109/icassp40776.2020.9053405

[17] Chen, J., Lu, Y., Yu, Q., *et al.* (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv: 2102.04306.

[18] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al.* (2023) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T. and Nishino, K., Eds., *Computer Vision—ECCV* 2022 *Workshops*, Springer, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9

[19] Wang, H., Cao, P., Wang, J. and Zaiane, O.R. (2022) Uctransnet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 2441-2449. https://doi.org/10.1609/aaai.v36i3.20144

[20] Zhou, H.Y., Guo, J., Zhang, Y., *et al.* (2021) nnFormer: Interleaved Transformer for Volumetric Segmentation. arXiv: 2109.03201.

[21] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T. and Xie, S. (2022) A Convnet for the 2020s. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 11966-11976. https://doi.org/10.1109/cvpr52688.2022.01167

[22] Gu, A., Goel, K. and Ré, C. (2021) Efficiently Modeling Long Sequences with Structured State Spaces. arXiv: 2111.00396.

[23] Gu, A. and Dao, T. (2023) Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv: 2312.00752.

[24] Liu, Y., Tian, Y., Zhao, Y., *et al.* (2024) VMamba: Visual State Space Model. arXiv: 2401.10166.