

基于Mamba模型的U型医学图像分割网络

勾金伟, 李代民, 刘 锋*

重庆理工大学数学科学学院, 重庆

收稿日期: 2025年9月19日; 录用日期: 2025年10月22日; 发布日期: 2025年10月30日

摘 要

在医学图像分割网络结构中, 卷积神经网络(CNN)无疑是最强大的网络, 它凭借强大的局部特征提取能力, 取得了很多突破性进展, 成为该领域的主流方法之一。然而, CNN固有的局部感知机制使其在捕捉图像中长距离依赖信息时存在明显的局限性, 这在一定程度上制约了其分割性能。Transformer模型凭借自注意力机制, 在处理长距离依赖关系上展现出显著优势, 能够有效挖掘图像全局范围内的特征关系。但自注意力机制的计算复杂度随输入序列长度的平方而增加, 导致Transformer模型在处理高分辨率医学影像时, 往往面临计算成本过高、内存占用量大的问题。为解决该问题, 本文提出一种新的分割网络——M-UNet。该网络的创新在于将Mamba模型特有的VSS块结构与空洞空间卷积金字塔池化模块进行结合。利用VSS块的线性复杂度特性, 在实现高效全局信息建模的同时, 大幅降低计算内存。其次, 借助ASPP模块多尺度空洞卷积的优势, 精准捕捉医学图像中不同尺度的关键特征, 弥补了CNN在多尺度特征提取上的不足。为验证M-UNet的性能, 本文在Synapse数据集和ACDC数据集上进行了实验。实验结果表明, 相较于传统基于CNN的分割方法, M-UNet在分割精度和计算速度两方面均有显著提升。

关键词

医学图像分割, 长距离建模, 卷积神经网络, Mamba

U-Shaped Medical Image Segmentation Network Based on the Mamba Model

Jinwei Gou, Daimin Li, Feng Liu*

School of Mathematical Sciences, Chongqing University of Technology, Chongqing

Received: September 19, 2025; accepted: October 22, 2025; published: October 30, 2025

Abstract

In the architecture of medical image segmentation networks, Convolutional Neural Networks (CNNs)

*通讯作者。

文章引用: 勾金伟, 李代民, 刘锋. 基于 Mamba 模型的 U 型医学图像分割网络[J]. 计算机科学与应用, 2025, 15(10): 341-350. DOI: 10.12677/csa.2025.1510273

are undoubtedly the most powerful networks, achieving significant breakthroughs due to their strong local feature extraction capabilities, making them one of the mainstream methods in this field. However, the inherent local perception mechanism of CNNs imposes clear limitations when capturing long-distance dependency information in images, which somewhat restricts their segmentation performance. Conversely, Transformer models, with their self-attention mechanism, exhibit significant advantages in handling long-distance dependencies, enabling effective exploration of the feature relationships across the entire image. However, the computational complexity of the self-attention mechanism increases quadratically with the length of the input sequence, leading to issues of excessively high computational costs and large memory usage when dealing with high-resolution medical images. To address this issue, this paper proposes a new segmentation network—M-UNet. The innovation of this network lies in the combination of the unique VSS block structure of the Mamba model with the dilated spatial convolution pyramid pooling module. By leveraging the linear complexity characteristic of the VSS block, it efficiently models global information while significantly reducing computational memory. Additionally, utilizing the multi-scale dilated convolution advantages of the ASPP module allows for precise capture of key features at different scales within medical images, compensating for CNNs' shortcomings in multi-scale feature extraction. To validate the performance of M-UNet, experiments were conducted on the Synapse dataset and the ACDC dataset. The experimental results indicate that compared to traditional CNN-based segmentation methods, M-UNet shows significant improvements in both segmentation accuracy and computational speed.

Keywords

Medical Image Segmentation, Long-Distance Modeling, CNN, Mamba

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

1.1. 基于传统的医学图像分割方法

传统图像分割技术的核心依托于多学科理论, 这些方法长期作为图像分割领域的核心方法, 为后续相关方法的演进奠定了重要基础。从技术分类来看, 传统图像分割方法主要可划分为三大类。第一种是基于阈值的分割方法, 该方法核心思路是通过设定合理的灰度阈值, 将图像中像素点的灰度值与阈值进行比较, 从而将图像划分为前景与背景等不同区域。Sandhya 等[1]提出在图像的多级分割中使用 HSO (Harmony Search Optimization)概念, 该算法使用 Otsu 和 Kapur 提出的目标函数指导候选解, 从图像直方图内部适当的搜索空间中得到候选解, HSO 算子不断演化候选解直到找到最佳阈值。Weglinski 等人[2]提出了用中值滤波减少噪声再用区域生长法来进行图像的分割。Kaur 等[3]提出一种混合多级阈值算法, 将直觉模糊集和 tsallis 熵相结合, 用于从模糊边界和对比度差的图像中自动描述肿瘤组织区域; 第二种是基于水平集的分割方法, 水平集方法是由 Osher 和 Sethian [4]提出的, 该方法借助水平集函数将图像中动态变化的曲线或曲面嵌入到高维空间中, 进而实现对图像中目标边界的精准捕捉与分割。Singh 等人[5]针对图像中脑肿瘤分割的需求, 提出了一种自动稀疏约束水平集方法。通过这一设计, 有效提升了图像分割的精确度。Zabir [6]等人根据水平集方法中初始轮廓选择的问题, 提出了一种基于迭代距离正则化水平集演化的分割方案, 并且还在 BRATS2012 公开数据集的测试中展现出来显著的优势。Hachemi 等人[7]深受启发, 他们将水平集方法与聚类算法进行融合, 引入模糊 C 均值聚类 FCM 的核心概念, 优化了传

统水平集分割流程。第三种是基于区域的分割方法, Toennies [8]运用医学图像全自动分割的需求, 提出了一种自适应区域生长算法。该算法的核心点在于, 能够从感兴趣区域的自身特征中自主学习并生成区域, 无需人工预设参数。后来, Rajeswari [9]等人还在区域生长方法中初始种子点选择这一关键环节, 提出了一种基于改进区域生长的全图像分割方法。在传统图像分割算法体系中, 普遍存在一个共性问题, 那就是对人工先验知识的高度依赖, 具有很强的主观性, 受这些主观经验的影响, 传统分割方法在处理医学图像时, 往往难以精准捕捉病变区域的真实边界与完整的形态, 难以提高分割精度。

1.2. 基于深度学习的医学图像分割方法

因为传统图像分割算法依赖主观经验, 且传统方法在复杂图像中分割精度不高, Ronneberger [10]等人就对 FCN 的架构进行了创新, 提出了一种名为 UNet 的新型网络。Wang 等人[11]使用 3DU-Net 对 MRI 图像中的脑瘤进行分割, 并在 UNet 使用非线性 Leaky ReLU 函数作为激活函数提升了模型的精度。Long [12]等人做出了创新, 提出了一种端到端的图像分割模型 FCN, FCN 将 CNN 中的全连接层转换成卷积层, 通过上采样操作将 CNN 特征恢复成输入图像大小, 以真实图像作为判断依据, 让网络直接做像素级别的识别和分割。Yang [13]等人提出一种密集全卷积网络。与 U-Net 相比, 放弃了上采样路径。Luo 等人[14]用 3D-CNN 的方法提取不同模态图像空间特征, 这种方法可适应大小不同的病变区域。在端到端图像分割技术的演进中, Zhao 等人[15]提出了 CDDFuse 模型, 解决了多模态数据的精准分割问题。在本文中, 我们提出了 Mamba 与 U-Net 分割网络的结合, 利用 Mamba 模型特有的 VSS 模块与空洞空间卷积金字塔池化模块进行融合, 进一步解决了内存占用大和边界处分割精度不高的问题。与 VM-UNet 相比, 都减少了计算的线性复杂度, 在处理高分辨率图像时速度快, 内存效率高, 且都具有强大的全局建模能力。然后也有一些不同之处, 虽然 VM-UNet 的感受野较传统 CNN 有提升, 但还是有限, 在下采样时会降低特征图的分辨率, 而加入的金字塔池化模块不仅不会降低分辨率, 还会扩大感受野。且在参数上也小于以往的分割网络。这是因为金字塔模块里的空洞卷积可以在不损失分辨率的情况下来扩大感受野, 它将不同空洞卷积组合起来, 能看到不同尺度的特征信息。

2. 实验方法

2.1. SSM 原理

状态空间模型在深度学习领域, 它被用来处理序列数据, 捕捉数据中的长期依赖关系。SSM 的核心思想是将系统的当前状态进行抽象表示, 通过隐式状态来捕捉序列中的关联信息。状态空间模型主要由状态方程和观测方程构成, 用于构建时刻 t 输入 $x(t) \in R$ 和输出 $y(t) \in R$ 的关系。其中, 隐藏状态 $h(t) \in R^n$ 的维度为 n 。一般形式为:

$$h'(t) = Ah(t) + Bx(t) \quad (1)$$

其中, $h'(t)$ 是当前状态 $h(t)$ 的导数, $x(t)$ 是系统在 t 时刻的输入, $A \in R^{n \times n}$ 表示状态矩阵, $B \in R^{n \times 1}$ 表示线性投影参数。为了满足深度学习的需求, 很多时候需要对连续的状态空间模型进行离散化处理, 将连续时间的状态方程和观测方程转换为离散时间的形式, 以便在计算机上进行数值计算。首先, 引入一个时间尺度参数 Δ , 并使用固定的离散化规则将 A 和 B 转化为 $\tilde{A}$ 和 $\tilde{B}$, 一般形式如下:

$$\bar{A} = e^{A\Delta} \quad (2)$$

$$\tilde{B} = (\Delta A)^{-1} (e^{A\Delta} - I) \Delta B \quad (3)$$

所以, 就可以将式(1)变化为:

$$h'(t) = \bar{A}h(t) + \bar{B}x(t) \quad (4)$$

最后通过全局卷积的方式对式(4)进行计算:

$$h_{k+1} = A_d h_k + B_d x_k \quad (5)$$

其中, k 表示离散时间步。

2.2. ASPP

在图像分割任务中, 感受野的大小直接决定了模型对图像全局信息和局部细节的捕捉能力。然而, 传统卷积通过堆叠层数或下采样扩大感受野, 但这会导致分辨率下降和细节丢失, 且感受野固定, 难以同时适应不同尺度的目标。为了解决本论文的问题, 在经过理论论证后, 我们选择了 ASPP 模块, 它可以并行多个具有不同空洞率的空洞卷积层, 能够在不降低分辨率的情况下, 有效地扩大感受野。同时我们还进行了实验验证, 实验结果表明, 该模块的引入有效减少了分割精度。特别是在小目标分割上非常明显。ASPP 是图像分割中用于捕捉多尺度上下文信息的核心模块, 其核心是通过不同膨胀率的空洞卷积并行提取特征, 模拟金字塔池化的多尺度感受野效果。ASPP 的主要实现方式分为四步。首先通过在卷积核之间引入空洞, 可以扩大感受野, 而不增加参数数量和计算复杂度, 其次是多尺度膨胀率, ASPP 模块并行地使用不同膨胀率的空洞卷积(r1、r6、r12、r18, 空洞率越大, 感受野越大)。然后池化, ASPP 模块还包含一个全局平均池化操作, 目的是将特征图的空间维度压缩为 1×1 , 从而获取全局信息。最后特征融合, 最后将不同膨胀率的空洞卷积输出和全局平均池化进行融合, 得到最终的特征图。它的模块构造结构图如图 1。

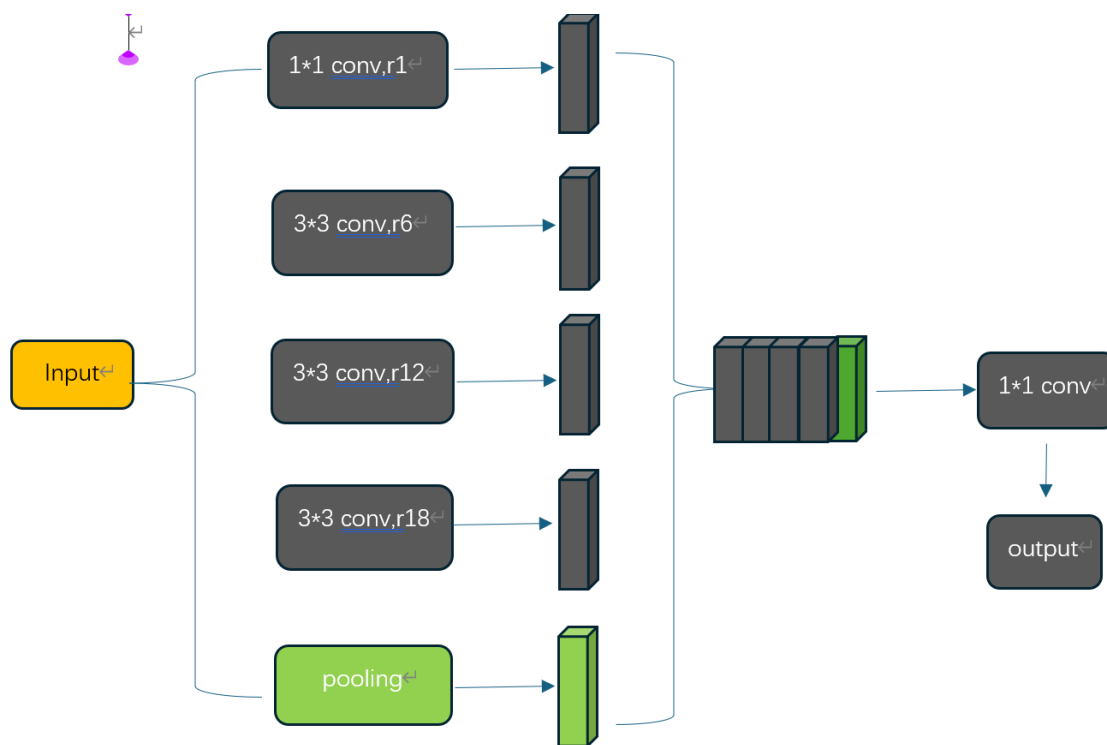


Figure 1. ASPP structure diagram

图 1. ASPP 结构图

2.3. VSS

在图像分割中, VSS (Visual State Space)模块是 Vision Mamba UNet (VM-UNet)的核心模块, 专门用

于捕捉大范围的上下文信息，有助于提高模型对全局信息的建模能力，同时保持线性计算复杂度。VSS 模块的输入首先进行归一化处理，之后被分成两个分支。第一个分支通过一个线性层后接一个激活函数。第二个分支则依次通过一个线性层、一个深度可分离卷积和一个激活函数处理后，送入 2D-Selective-Scan (SS2D) 模块进行进一步的特征提取。然后，使用 Layer Normalization 对特征进行归一化。最后，通过逐元素乘法操作，将两个分支的输出合并。再使用一个线性层混合特征，并通过与残差连接的组合形成 VSS 块的输出。它的框架图如图 2 所示。

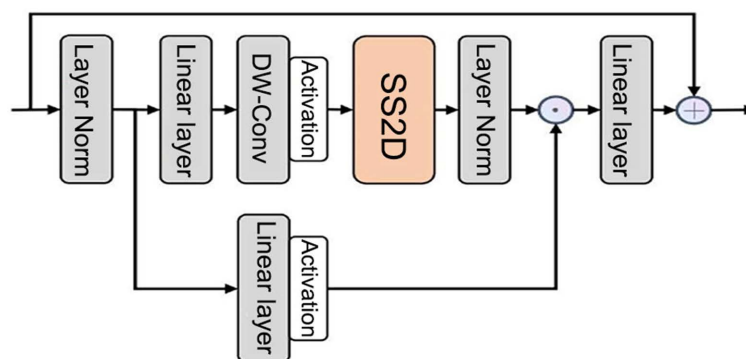


Figure 2. VSS framework diagram
图 2. VSS 框架图

2.4. 总体框架

M-UNet 的医学图像分割基本流程为输入数据预处理、编码器编码，提取特征、解码器恢复细节、输出分割图，各模块协同配合实现高效医学图像分割，首先，输入数据预处理：Patch Embedding 奠定多尺度基础网络首先通过 Patch Embedding 模块对输入图像进行初始化处理，将原始图像划分为尺寸为 4×4 的不重叠图像补丁，并将每个补丁嵌入至 96 维的特征空间中。然后再逐层下采样，强化全局与局部信息捕捉，预处理后的特征图像进入编码器，通过“VSS (Visual State Space) 块 + Patch Merging 层”的组合结构，实现多层次特征提取与逐步下采样。

其次，将处理好的图片输入编码器，经过编码器的多轮下采样处理。这一过程不仅实现了多尺度特征的有效提取，更通过维度与分辨率的动态调整，显著增强了网络对图像局部细节与全局关联信息的捕捉能力。在编码器与解码器之间的跳跃连接层，引入 ASPP 模块，该模块通过搭载不同扩张率的空洞卷积操作，可同时捕捉多尺度上下文信息。解码器采用与编码器对称的结构设计，以“VSS 块 + Patch Expanding 层”为核心，结合跳跃连接实现特征的高效重建与空间细节恢复。经过多轮上采样处理，最终使输出特征图恢复至与输入图像一致的原始分辨率。同时，为避免特征在多次下采样与上采样过程中丢失关键空间细节，编码器与解码器之间引入跳跃连接层，该层将编码器各层输出的高分辨率特征，直接传递至解码器的对应层级。

最后，经过解码器的特征重建与细节恢复后，网络最终输出与输入图像尺寸一致的分割结果，就是最终的预测图。M-unet 的总体框架图如图 3 所示。

3. 实验与结果

3.1. 数据集

3.1.1. Synapse 多器官影像分割数据集

该数据集是公开数据集，我们在 MICCAI 2015 多图谱腹部标记挑战赛中使用了 30 张腹部 CT 扫描，

总共有 3779 张轴向增强腹部临床 CT 图像。每个 CT 体由 85 到 198 个 512×512 像素的切片组成。本文选用了 18 个病例(2212 切片)作为训练集, 12 个病例(1567 切片)作为验证集, 对 8 个腹部脏器(主动脉、胆囊、脾脏、左肾、右肾、肝脏、胰腺、胃)进行随机分割, 最后, 我们报告了平均 DSC 和平均 Hausdorff 距离(HD)。

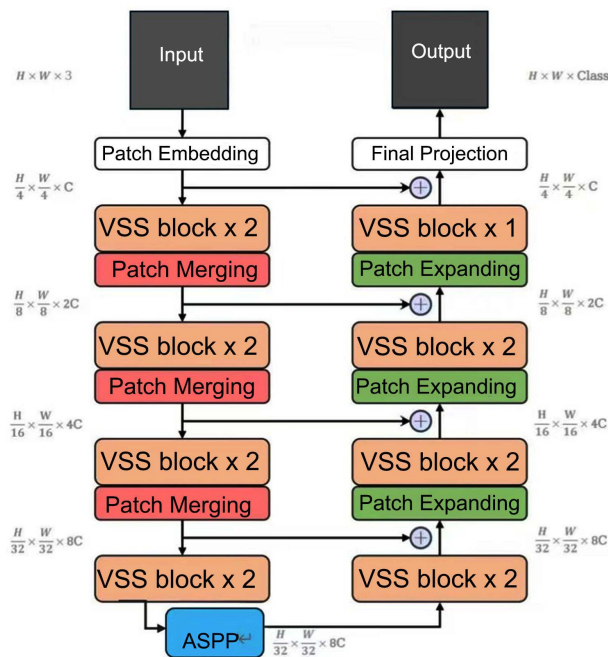


Figure 3. Overall framework diagram

图 3. 总体框架图

3.1.2. ACDC (Automatic Cardiac Diagnosis Challenge)数据集

该数据集是一个用于心脏分析和自动诊断的挑战数据集, 他包含了 100 名不同患者的 MRI 扫描图像, 该数据集对心脏关键结构进行了精准标注, 具体涵盖左心室、右心室及心肌三大核心区域。从数据维度来看, 每位患者的 MRI 个体数据由 28 至 40 层连续切片构成, 切片厚度范围为 5 毫米至 10 毫米。在数据划分上, 将数据集按 7:1:2 的比例划分为三个子集。

3.2. 评价指标

在医学图像分割中, Dice 系数、HD95 (95% Hausdorff 距离)、精确度(Precision)、召回率(Recall)、IoU (交并比)是评估模型性能的核心指标。Dice 系数衡量两个集合(预测掩模 P 和真实掩模 T)的重叠程度, 取值范围为[0, 1], 值越大表示分割效果越好。

$$\text{Dice} = \frac{2|TP|}{|T| + |P|} \quad (6)$$

Hausdorff 距离衡量两个集合之间的最大不匹配距离, 而 HD95 是其 95%分位数版本, 忽略 5%的极端离群点, 避免噪声或异常值对结果的影响, 而 HD95 是对其的优化, 用于评估模型预测的目标边界与真实边界的偏差程度。HD95 越小, 表示预测与真实掩模的边界越接近。

$$\text{HD}(P, T) = \max \left\{ \sup_{p \in P} \inf_{t \in T} d(p, t), \sup_{t \in T} \inf_{p \in P} d(t, p) \right\} \quad (7)$$

$$\text{HD95}(P, T) = 95\text{th percentile of all distances in HD calculation} \quad (8)$$

精确度又称准确率，核心衡量模型预测为目标的像素中，真正是目标像素的比例——即评估预测结果的纯度。真阳性(TP)，假阳性(FP)，真阴性(TN)，假阳性(FP)。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

召回率衡量真实正样本中被模型正确预测的比例，反映模型捕捉目标的能力。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

IoU 是衡量预测掩模与真实掩模的重叠区域占两者并集的比例，取值范围为[0, 1]，值越大表示分割效果越好。

$$\text{IoU} = \frac{|\text{TP}|}{|\text{PUG}|} \quad (11)$$

4. 实验细节

对于所有的实验，为了提高模型的泛化能力，我们应用了简单的数据增强。根据经验，我们在网络的编码器中加载了预训练模型的权重，使用 SGD 优化器进行训练，学习率为 0.01，动量为 0.9，权衰减为 0.0001。ACDC 数据集的批大小为 24，Synapse 数据集的默认训练迭代次数分别为 20 k 和 14 k。所有实验均在 python3.13、pytorch2.6、cuda12.6 和 ubuntu22.4 的环境，并使用单个 Nvidia RTX4060Ti GPU 进行。

4.1. 实验结果对比

4.1.1. Synapse 多器官分割数据集上的实验结果

表 1 呈现了 M-UNet 模型在 Synapse 多器官分割数据集上的实验结果。从核心指标来看，该模型在平均 dice 相似系数(DSC)及各单个器官的 DSC 指标上均实现了当前最优表现，其平均 DSC 达到 80.79%，相较于同实验中的其他对比方法(如传统 CNN 模型、基于 Transformer 的分割模型等)，该数值呈现出显著的性能提升，充分验证了 M-UNet 在多器官分割任务中的综合优势。由分割结果可见，M-UNet 在不同尺寸、不同解剖复杂度的器官分割中均展现出出色的适应性：其中，对左肾、肝脏、胰腺与胃的分割 DSC 分别达到 86.01%、94.76%、64.19%和 82.38%。尤其是针对肝脏这类体积较大、边界相对清晰的器官，模型实现了 94.76%的高 DSC 分数。这一结果充分证明，M-UNet 在多器官分割场景中，不仅对较大尺寸器官的分割具备高鲁棒性，同时在复杂解剖结构的精准分割上也展现出优异性能。

Table 1. Experimental results on the Synapse dataset

表 1. Synapse 数据集上的实验结果

Methods	DSC↑	HD↓	Aorta	Callbladder	Kidney (L)	Kidney (R)	Liver	Pancreas	Spleen	Stomach
V-Net [16]	68.81	/	75.3	51.87	77.1	80.75	87.84	40.05	80.56	56.98
DARR [17]	69.77	/	74.7	53.77	72.31	73.24	94.08	54.18	89.9	45.96
U-Net [18]	76.85	39.70	89.1	69.72	77.77	68.6	93.43	53.93	86.67	75.58
R50 U-UNet [19]	74.68	36.87	87.7	63.66	80.6	78.19	93.74	56.9	85.87	75.58
R50 Att-UNet [19]	75.57	36.97	55.9	63.91	79.2	72.71	93.56	49.37	87.19	74.95
R50 ViT [19]	71.29	32.87	73.7	55.13	75.8	72.2	91.51	45.99	81.99	73.95
M-Unet	80.79	26.31	87.5	69.1	86.01	78.64	94.76	64.19	86.74	82.38

4.1.2. ACDC 数据集上的实验结果

表 2 呈现了 M-UNet 模型在自动心脏诊断挑战(ACDC)数据集上的实验结果。从核心评估指标来看,在 dice 相似系数(DSC)、召回率(Recall)与交并比(IoU)三大关键指标上,均全面超越其他对比方法,具体数值分别达到 88.01%、89.01%和 80.71%。与当前最优模型 Swin-UNet 相比,这三项指标的提升幅度分别为 2.23 个百分点、2.18 个百分点和 2.17 个百分点,提升效果显著。最主要是其特征提取模块能够更精准地捕捉心脏结构的关键语义信息,且该模型在全局上下文建模层面,该模型有效整合了不同尺度的特征关联,为整体分割精度的提升奠定了基础。不过,实验结果也暴露出模型的一处待优化方向:其 95%豪斯多夫距离(HD95)指标相对较高。这一现象表明,尽管 M-UNet 的整体分割区域与真实标注的重叠度有异,但在处理心脏细微解剖边缘处时,仍存在局部精度不足的问题。由于当图像中出现边界模糊或标注微小偏差时,模型对局部边界的预测误差会被放大,最终导致 HD95 数值升高。

综合来看,尽管 M-UNet 在边缘分割精度上仍有进一步优化的空间,但凭借其在 DSC、Recall、IoU 等核心指标上的出色表现,已充分证明其在分割任务中的强大优势。

Table 2. Experimental results on the ACDC dataset
表 2. ACDC 数据集上的实验结果

	Average					Right ventricle		Myocardium		Left ventricle	
Methons	DSC↑	HD95↓	Precision↑	Recall↑	IoU↑	DSC↑	HD95↓	DSC↑	HD95↓	DSC↑	HD95↓
V-Net [16]	80.87	3.06	89.21	81.38	72.76	81.43	3.66	75.79	3.06	85.38	2.47
U-Net [19]	84.23	2.73	89.71	82.72	75.51	85.72	1.75	77.41	3.72	89.55	2.73
TransUNet [20]	86.56	2.28	84.63	83.42	76.77	87.15	2.97	83.34	1.89	89.2	1.97
Swin-UNet [21]	85.78	2.47	86.56	86.83	78.54	88.54	2.08	81.82	2.03	86.97	3.31
M-UNet	88.01	2.86	88.53	89.01	80.71	88.46	2.81	83.96	2.43	91.62	3.35

5. 消融实验

5.1. 分辨率的影响

关于输入分辨率的影响。该网络的默认输入分辨率为 224×224 。在这里,我们还跑了高分辨率 512×512 上训练的结果,如表 3 所示。当使用 512×512 作为输入时,我们保持相同的批次大小,对于该网络,将分辨率从 224×224 更改为 512×512 ,导致平均 DSC 提高了 6.89%,但是,这会导致更大的计算成本。因此,考虑到计算成本,本文中的所有实验比较均以 224×224 的分辨率进行。

Table 3. The impact of input resolution
表 3. 输入分辨率的影响

Resolution	Average DSC	Aorta	Callbladder	Kidney (L)	Kidney (R)	Liver	Pancreas	Spleen	Stomach
224	80.79	87.46	69.1	86.01	78.64	94.76	64.19	86.74	82.38
512	87.68	90.76	72.33	86.04	83.41	95.43	74.39	88.37	84.36

5.2. ASPP

为验证各模块的有效性,本研究在 ACDC 数据集上设计实验。从参数量维度来看,纯 VMamba 模型展现出最优的轻量化优势,参数量为所有测试模型中最小,但受限于特征提取能力,其 DSC 值明显低于

其他集成模块的模型。当为 VMamba 引入 ASPP 模块后, 模型参数量虽有小幅增加, 但 DSC 值实现显著跃升, 且整体参数量仍低于除纯 VMamba 外的其他模型, 有效缓解了参数太大的问题。在未引入成 ASPP 模块时, 基础 Mamba 模型在 ACDC 数据集上的平均 DSC 为 86.31%。而引入 ASPP 模块后, 模型在数据集上的分割性能得到突破性提升, 平均 DSC 分别提升至 87.79%。这一结果充分证明, ASPP 模块通过搭载不同扩张率的卷积操作, 能够高效捕捉多尺度上下文信息。为进一步验证 VMamba 与 ASPP 组合的优越性, 在保留 ASPP 模块的前提下, 将核心的 VMamba 分别替换为 SwinUNet 这一主流分割模型。实验结果表明, SwinUNet 与 ASPP 的组合虽能在两个数据集上实现 DSC 提升, 但伴随参数量的明显增加, 且最终性能仍逊色于 VMamba 与 ASPP 的组合。如表 4 所示。

Table 4. Ablation experiment of ASPP

表 4. ASPP 的消融实验

Dataset	Methods	Params	DSC↑
ACDC	Swin-UNet	27.16	85.78
	Vmamba	19.12	86.31
	Swin-UNet + ASPP	33.84	86.93
	M-UNet	25.81	87.79

6. 实验总结

在本文中, 我们介绍 M-UNet 模型, 通过创新性融合 VSS 模块与 ASPP 模块, 成功突破了传统 CNN 在捕捉长距离依赖信息时的固有局限, 同时实现了计算成本的有效降低。实验结果表明, 该模型在 ACDC 及 Synapse 多器官分割数据集上均表现优异, 展现出卓越的特征建模能力与全局信息捕获性能。然而, 本研究仍存在一些缺点, 第一, 模型当前性能对预训练权重的初始化存在较强依赖, 后续可探索针对医学图像分割任务特性设计的端到端预训练方法, 进一步释放模型性能潜力; 第二, 边缘区域的分割精度仍有提升空间, 尤其在 ACDC 数据集中, 模型的 HD95 指标相对较高, 表明对心脏细微解剖边缘的预测精度不足, 未来需重点优化边缘特征提取机制; 如可以加入边界感知损失函数, 注意力机制来优化这个问题。第三, 研究目前仅聚焦于单模态医学图像分割任务, 后续计划扩展至多模态场景, 并探索模型在 3D 医学图像分割中的泛化能力。

综上所述, 模型的性能优势源于两大核心模块的协同作用, 首先, VSS 块凭借线性计算复杂度特性, 实现了高效的全局上下文建模, 与 Swin-UNet 的模型相比, M-UNet 在显著降低计算开销的同时, 仍保持了优异的分割精度, 解决了传统 Transformer 模型“高精度伴随高计算成本”的痛点。其次, ASPP 模块通过多尺度空洞卷积结构, 强化了模型对医学图像中不同尺度特征的处理能力, 能够精准捕捉病灶、器官边界等关键语义信息, 在心脏结构分割、多器官协同分割等复杂医学任务中, 显著提升了分割结果的准确性。

参考文献

- [1] Sandhya, G., Babu Kande, G. and Satya Savithri, T. (2017) An Efficient Approach for the Detection of White Matter, Gray Matter, and Cerebrospinal Fluid from MR Images of the Brain Using an Advanced Multilevel Thresholding. 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 27-28 February 2017, 422-426. <https://doi.org/10.1109/aeicb.2017.7972345>
- [2] Weglinski, T. and Fabijanska, A. (2011) Brain Tumor Segmentation from MRI Data Sets Using Region Growing Approach. Proceedings of the 7th International Conference on Perspective Technologies and Methods in MEMS Design, Polyana, 11-14 May 2011, 185-188.

- [3] Kaur, T., Saini, B.S. and Gupta, S. (2018) A Novel Fully Automatic Multilevel Thresholding Technique Based on Optimized Intuitionistic Fuzzy Sets and Tsallis Entropy for MR Brain Tumor Image Segmentation. *Australasian Physical and Engineering Sciences in Medicine*, **41**, 41-58.
- [4] Osher, S. and Sethian, J.A. (1988) Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics*, **79**, 12-49. [https://doi.org/10.1016/0021-9991\(88\)90002-2](https://doi.org/10.1016/0021-9991(88)90002-2)
- [5] Singh, N. and Choudhary, N. (2017) Automatic Localization and Level Set Based Energy Minimization for MRI Brain Tumor. 2017 *International Conference on Computer, Communications and Electronics (Comptelix)*, Jaipur, 1-2 July 2017, 130-134. <https://doi.org/10.1109/comptelix.2017.8003951>
- [6] Zabir, I., Paul, S., Rayhan, M.A., Sarker, T., Fattah, S.A. and Shahnaz, C. (2015). Automatic Brain Tumor Detection and Segmentation from Multi-Modal MRI Images Based on Region Growing and Level Set Evolution. 2015 *IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, Dhaka, 19-20 December 2015, 503-506. <https://doi.org/10.1109/wiecon-ece.2015.7443979>
- [7] Beddad, B. and Hachemi, K. (2016) Brain Tumor Detection by Using a Modified FCM and Level Set Algorithms. 2016 *4th International Conference on Control Engineering & Information Technology (CEIT)*, Hammamet, 16-18 December 2016, 1-5. <https://doi.org/10.1109/ceit.2016.7929114>
- [8] Pohle, R. and Toennies, K.D. (2001) Segmentation of Medical Images Using Adaptive Region Growing. *SPIE Proceedings*, **4322**, 1337-1346. <https://doi.org/10.1117/12.431013>
- [9] Sukanya, A., Rajeswari, R. and Subramaniam Murugan, K. (2020) Region Based Coronary Artery Segmentation Using Modified Frangi's Vesselness Measure. *International Journal of Imaging Systems and Technology*, **30**, 716-730. <https://doi.org/10.1002/ima.22412>
- [10] 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
- [11] Wang, L., Xie, C. and Zeng, N. (2019) RP-Net: A 3D Convolutional Neural Network for Brain Segmentation from Magnetic Resonance Imaging. *IEEE Access*, **7**, 39670-39679. <https://doi.org/10.1109/access.2019.2906890>
- [12] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/cvpr.2015.7298965>
- [13] Yang, B. and Zhang, W. (2019) FD-FCN: 3D Fully Dense and Fully Convolutional Network for Semantic Segmentation of Brain Anatomy. arXiv: 1907.09194.
- [14] Luo, M., Huang, J. and Yang, F. (2014) Multimodal 3D Convolutional Neural Networks Features for Brain Tumor Segmentation. *Science Technology and Engineering*, **14**, 78-83.
- [15] Zhao, Z., Bai, H., Zhang, J., Zhang, Y., Xu, S., Lin, Z., et al. (2023) CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 5906-5915. <https://doi.org/10.1109/cvpr52729.2023.00572>
- [16] Milletari, F., Navab, N. and Ahmadi, S. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3dv.2016.79>
- [17] Fu, S., Lu, Y., Wang, Y., Zhou, Y., Shen, W., Fishman, E., et al. (2020) Domain Adaptive Relational Reasoning for 3D Multi-Organ Segmentation. In: Martel, A.L., et al., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*, Springer, 656-666. https://doi.org/10.1007/978-3-030-59710-8_64
- [18] 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
- [19] Ibtehaz, N. and Rahman, M.S. (2020) MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks*, **121**, 74-87. <https://doi.org/10.1016/j.neunet.2019.08.025>
- [20] Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., et al. (2024) TransUNet: Rethinking the U-Net Architecture Design for Medical Image Segmentation through the Lens of Transformers. *Medical Image Analysis*, **97**, Article ID: 103280. <https://doi.org/10.1016/j.media.2024.103280>
- [21] 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