

# 基于改进Mamba的医学图像分割模型

高博艺, 丁学明, 胡鸿翔, 丁雪峰

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年2月24日; 录用日期: 2025年3月17日; 发布日期: 2025年3月26日

## 摘要

在医学图像分割任务中, 针对传统U型网络在膀胱肿瘤和视网膜眼底MRI图像分割中在处理复杂结构和细节上分割精度差的问题, 本研究提出了一种改进的U-Net网络模型——Akmamba-Net。该模型结合AKConv和Mamba-out模块, 有效提高了模型的特征提取能力。AKConv模块通过引入卷积操作与空间重采样机制, 增强了网络的适用性和灵活性, 尤其是在处理形状不规则的肿瘤边界时。Mamba-out模块则通过优化特征融合和增强细节信息, 进一步提升了模型的分割精度。实验结果表明, Akmamba-Net网络在视网膜眼底和膀胱肿瘤MRI图像分割任务中, Precision、Dice系数、IoU指标分别达到了97.2%、82.5%、71.9%和89.64%、89.98%、81.75%, 与U-Net和其他主流模型相比显著提高了分割的准确性, 能够有效地提高视网膜眼底、膀胱肿瘤的分割精度, 满足医学图像分割的需求。

## 关键词

U-Net, Mamba-Out, AKConv, 残差网络

# Medical Image Segmentation Model Based on Improved Mamba

Boyi Gao, Xueming Ding, Hongxiang Hu, Xuefeng Ding

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 24<sup>th</sup>, 2025; accepted: Mar. 17<sup>th</sup>, 2025; published: Mar. 26<sup>th</sup>, 2025

## Abstract

In medical image segmentation tasks, traditional U-Net networks often exhibit poor segmentation accuracy when handling complex structures and details in bladder tumor and retinal fundus MRI images. To address this issue, this study proposes an improved U-Net network model, Akmamba-Net. The model integrates the AKConv and Mamba-out modules, significantly enhancing the model's

feature extraction capability. The AKConv module improves the network's adaptability and flexibility by introducing convolution operations and spatial resampling mechanisms, particularly in handling irregular tumor boundaries. The Mamba-out module further enhances the segmentation accuracy by optimizing feature fusion and boosting detailed information. Experimental results show that Akmamba-Net achieves Precision, Dice coefficient, and IoU metrics of 97.2%, 82.5%, and 71.9% for retinal fundus segmentation, and 89.64%, 89.98%, and 81.75% for bladder tumor segmentation, respectively. Compared with U-Net and other mainstream models, Akmamba-Net significantly improves segmentation accuracy, effectively enhancing the segmentation precision of retinal fundus and bladder tumor regions, thus meeting the needs of medical image segmentation.

## Keywords

U-Net, Mamba-Out, AkConv, Residual Network

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

医学图像分割是医学图像处理中的一项重要任务，它直接影响到疾病的早期诊断、治疗计划的制定以及患者的治疗效果评估。随着计算机视觉和深度学习技术的快速发展，基于深度神经网络的医学图像分割方法在多个医学领域，尤其是肿瘤检测和器官分割中取得了显著的成果。医学影像技术在癌症诊断中发挥着不可或缺的作用，主要包括 X 射线检查、电子计算机断层成像(computed tomography, CT)、超声(ultrasound)、乳腺钼靶 X 线摄影检查、磁共振成像(magnetic resonance imaging, MRI)以及正电子发射断层成像(positron emission tomography, PET)等[1]。U-Net 网络是一种专门为医学图像分割任务设计的深度卷积神经网络，它首次由 Olaf Ronneberger 等人[2]在 2015 年提出，并迅速成为医学图像分割领域的标准模型。U-Net 网络在医学图像分析中的广泛应用，尤其是在细胞分割、肿瘤检测和器官分割等任务中，表现出了良好的分割性能。

曹心姿等人[3]提出了一种基于改进的 U-Net 网络的 Res-U-Net 模型，通过引入残差结构和数据标准化，有效提高了膀胱肿瘤 MRI 图像分割的准确性和鲁棒性。实验结果表明，改进后的模型在分割效果上优于原始 U-Net 模型，具有较高的研究价值。Li X 等人[4]提出了一种用于膀胱癌和膀胱壁 MRI 图像分割的新型深度学习模型 MH2AFormer (Multi-scale Hybrid Attention with Transformer)。该模型通过引入多尺度混合注意力机制和 Transformer 模块，有效提高了膀胱图像分割的准确性和效率，尤其在处理复杂肿瘤形态和模糊边界时表现出色。Wang Z 等人[5]提出了 Mmaba-Unet，结合了 U-Net 网络结构和 Mamba 模块的新型医学图像分割网络，通过高效的利用全局上下文建模和多尺度特征融合，显著提升了分割精度。S Deari 等人[6]提出的 BFMD SN U-Net 通过解决过拟合问题、增强特征学习和引入全局上下文信息，显著提高了视网膜血管分割的性能。胡鸿翔等人[7]提出了一种全分辨率卷积交互网络模型 FRCI-Net。该模型通过多分辨率卷积交互机制，在保持完整图像分辨率的同时，有效扩展特征图的水平和垂直方向，集成相邻阶段的多尺度特征图以补充上下文信息，强化特征表示，并通过改进的残差模块学习多分辨率表示，生成高精度预测结果。然而，尽管 U-Net 在医学图像分割中展现了优异的性能，但是在面对复杂的医学图像时，特别是处理结构不规则和纹理模糊的区域时，仍然存在一定的局限性。因此，如何进一步提高图像的分割精度和鲁棒性，成为了当前研究的重点之一。

为了克服现有方法带来的局限性，本文提出了一种基于改进 Mamba 的医学图像分割模型 Akmamba-Net。我们在 U-Net 网络模型的基础上，结合了 Mamba-out 模块，并引入了 AKConv 卷积模块，以增强模型对复杂形状和边界的捕捉能力。通过对 Mamba 模块的改进，优化了特征融合的策略，使得模型在处理复杂医学图像时能够更加精准地识别和分割视网膜眼底和膀胱肿瘤区域。此外，AKConv 模块进一步提升了模型对细节的处理能力，从而增强了分割精度。

本文的主要贡献可以概括为以下几点：

- 1) 提出了一种改进的 Mamba 模块 Mamba-out，增强了医学图像分割模型的特征融合能力，特别是在边缘区域的分割精度方面。
- 2) 设计了一个集成 AKConv 的 U-Net 网络，通过引入空间重采样和卷积模块，提高了模型对复杂边界和形态不规则的肿瘤区域的处理能力。
- 3) 在视网膜眼底和膀胱肿瘤 MRI 图像分割任务中，实验结果表明，改进后的模型在分割精度和鲁棒性上显著优于传统 U-Net 模型。

2. 网络模型

U-Net 的内部结构由编码器、跳跃连接层和解码器三个主要部分组成。编码器部分通过一系列卷积层和池化层逐步提取图像特征并降低空间分辨率；跳跃连接层连接着编码器和解码器，进一步处理图像的高级特征；解码器部分通过跳跃连接将编码器的低层特征与解码器的高层特征结合，从而精确恢复图像的细节，将特征图逐渐恢复到原始的输入图像尺寸，并逐步生成分割结果。最终，输出层通过卷积操作生成分割结果，输出与输入图像相同大小的图片。

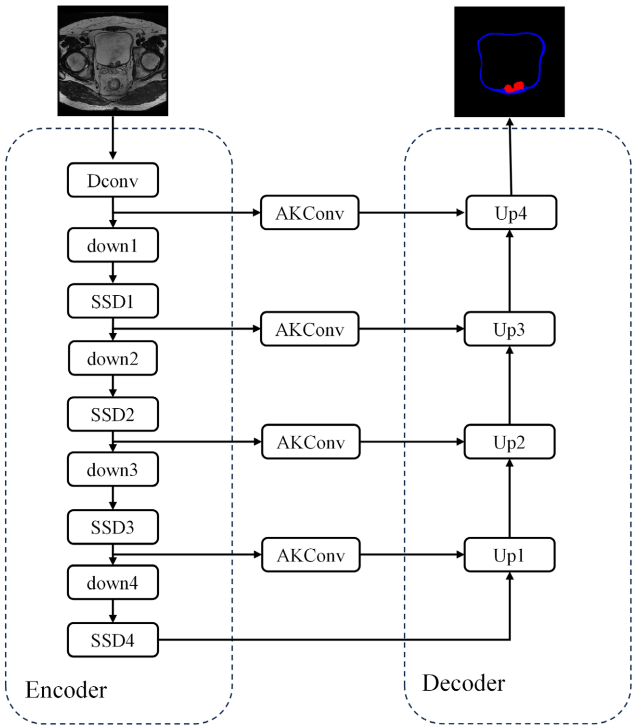


Figure 1. Akmamba-Net network architecture  
图 1. Akmamba-Net 网络结构

本文提出了一种改进的 U-Net 的分割网络 Akmamba-Net，在原本的 U-Net 模型架构中将普通卷积替

换为了 AKConv, 通过引入自适应的卷积核尺寸来更好地捕捉输入特征的空间信息。在标准卷积操作中, 卷积核的大小是固定的, 而 AKConv 通过引入可学习的参数, 使得卷积核的大小和形状能够动态调整, 从而更灵活地处理不同尺度和结构的图像特征。在跳跃连接层加入了 Mamba-out 模块, 能够有效处理图像中的复杂特征。Akmamba-Net 网络结构见图 1。

## 2.1. AKConv

AKConv 模块[8]是一种改进的卷积层, 提供了一种灵活的卷积机制, 通过引入动态偏移和双线性插值机制, 增强了卷积操作的适应性, 特别是在处理图像或特征图的空间变换时(例如目标的平移或旋转)。这种方法突破了传统卷积局限于固定的窗口和采样形状, 使得卷积操作可以更精准地适用于不同位置的目标和数据集。

首先, 输入特征图(x)经过  $p\_conv$  卷积层计算出一个偏移量(offset)。该偏移量反映了特征图中每个位置的位移量, 决定了卷积操作在采样时应当采用的新坐标系。接下来, 利用 offset 和预定义的常数生成新的空间位置(p)。这个新位置用于确定卷积操作在特征图中的采样位置, 确保卷积操作可以适应输入图像中的空间变化。

在生成新的采样位置后, AKConv 通过双线性插值计算每个采样点的权重。具体来说, 它通过计算四个邻近像素的距离, 得到每个像素的新位置的加权值, 这些权重会用于将四个邻近位置的特征值加权融合, 从而得到更精细的特征值。这些加权后的特征值构成了一个新的特征图( $x\_offset$ ), 该特征图即为输入图像经过偏移和插值后的重采样结果。

最后, 重采样后的特征图经过标准的卷积层(conv)、归一化层和 SiLU 激活函数进行进一步处理, 以提取更高层次的特征, 并生成最终的输出特征图(Output)。在整个过程中, AKConv 模块通过精细的空间调整和插值机制, 使得网络能够更好地处理图像中的平移、旋转等空间变换, 从而实现高效的特征提取。AKConv 模块网络结构见图 2。

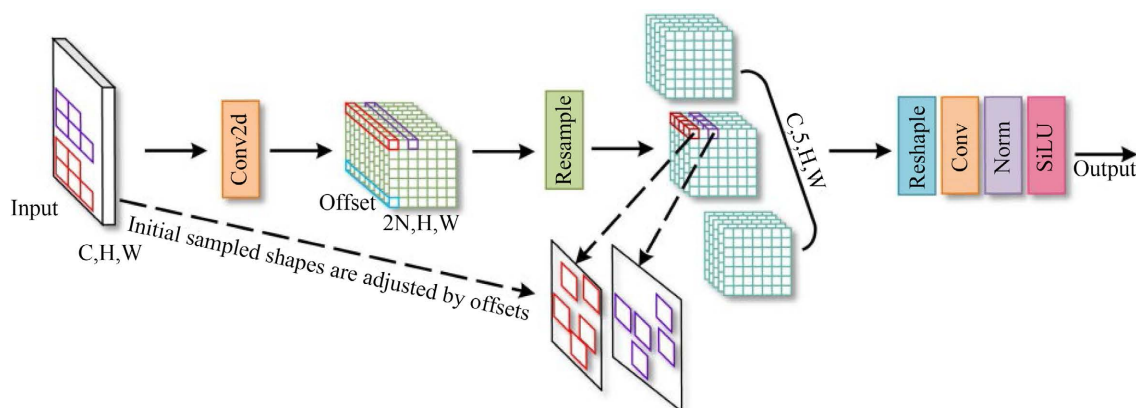


Figure 2. AKConv network architecture

图 2. AKConv 模块网络结构

总结来说, AKConv 通过三个关键步骤: 计算偏移量、进行双线性插值、重采样并卷积, 来实现更精细的特征提取。这种动态调整卷积采样位置的方式, 使得卷积神经网络能够在面对复杂的空间变换时, 仍然保持较强的适应性和表现力, 提高了特征提取的准确性。

## 2.2. Mamba-Out

本研究提出了 mamba-out 模块, 专门设计用于处理具有复杂时空结构的二维数据。该模块结合了卷

积神经网络(CNN)和特征投影技术,能够有效提取并融合输入数据中的空间和时间信息。模块的设计目标是在处理时序数据时保持计算效率,同时保留足够的特征表达能力。以下是 mamba-out 模块的详细架构与工作流程。

输入 Input 是一个形状为  $(B, C, H, W)$  的特征图,其中  $B$  是批次大小,  $C$  是通道数,  $H$  和  $W$  分别表示特征图的高度和宽度。在输入特征图进入网络之前,首先通过线性层进行初步的处理,得到新的特征图  $xz$ 。

接着,张量  $xz$  被拆分为两个部分:  $x$  和  $z$ ,其中  $x$  将用于后续的卷积操作,而  $z$  用于控制模型输出的缩放。此步骤的关键在于通过对  $z$  进行控制,实现模型输出的动态调整输出的大小。

处理后的  $x$  会传入一个深度可分离卷积层  $conv2d$ ,该层负责对数据进行空间特征的提取。卷积操作通过对每个输入通道应用独立的卷积核(组卷积),有效减少了计算量,并保留了重要的空间特征。卷积层的输出经过  $SiLU$  激活函数进行非线性转换,从而增强模型的表达能力。此阶段的操作通过以下公式(1)完。

$$x = SiLU(conv2d(x)) \quad (1)$$

网络的核心计算通过  $forward\_core$  函数完成,主要通过多次线性变换、卷积操作和状态更新来处理输入图片。在这一阶段,输入的张量  $x$  会被分解为多个维度,并通过时间序列处理模块进行处理。该模块通过迭代扫描,结合状态参数  $A_{logs}$  和时间参数  $D_s$ ,逐步更新状态信息,从而更好地学习时序特征。

通过这种方式,网络能够有效地捕捉时间依赖性和空间依赖性的信息,并在多个路径中融合这些信息。这些处理最终会生成四个张量  $y_1$ 、 $y_2$ 、 $y_3$ 、 $y_4$ ,分别代表不同状态下的输出。四个输出张量相加之后,得到最终的输出  $y$ ,如公式(2)所示。

$$y = y_1 + y_2 + y_3 + y_4 \quad (2)$$

之后  $y$  会经过  $LayerNorm$  层进行标准化,使其具有零均值和单位方差。之后,输出乘以由  $SiLU$  控制的缩放因子  $z$ ,动态调整输出的尺度。最终,处理后的输出通过一个线性层映射回原始的特征维度。

Mamba-out 模块为处理复杂的医学特征图片提供了一种高效且灵活的解决方案。通过深度可分离卷积、状态更新和动态缩放等技术,模型能够在计算效率和表达能力之间取得平衡。实验结果表明,该模块在医学图像分割任务中表现优异,具有广泛的应用前景。Mamba-out 模块网络结构见图 3。

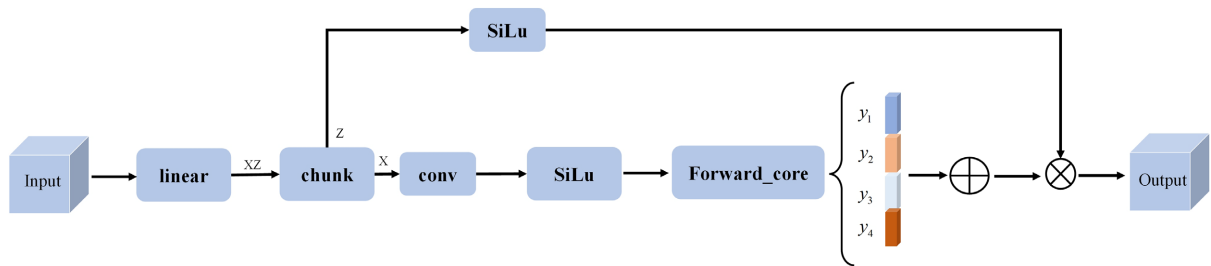


Figure 3. Mamba-out network architecture  
图 3. Mamba-out 模块网络结构

### 3. 实验结果和分析

#### 3.1. 数据集

本文采用 2 个数据集,分别是视网膜眼底图像公开数据集 DRIVE (Digital Retinal Images for Vessel Extraction)、ISICDM2019 临床数据分析挑战赛的基于磁共振成像的膀胱内外壁分割与肿瘤检测数据集,



膀胱数据采集自空军军医大学(第四军医大学),已经分别对肿瘤区域和膀胱壁区域进行了标记,标签使用 one-hot 编码进行处理。

### 3.2. 实验环境和参数配置

实验是在 Ubuntu22.04 操作系统上进行的,使用深度学习框架 Pytorch。GPU 配置为单卡的 NVIDIA GeForce GTX 4090, CUDA 的版本为 11.8。学习率为 0.0005,并设置了衰减权重,衰减率为 0.0001,每次训练批次(Batch)大小设置为 8,优化器选取了 Adam,并进行了 1000 个 Epoch 的训练周期。

### 3.3. 评价指标

本文采用精确率(Precision)、Dice 系数和 IoU 对模型的分割性能进行评价。

TP: 模型正确预测为正类的样本数。FP: 模型错误地预测为正类的样本数。精确率高意味着模型预测为正的样本中,正确的比例较高。

$$Precision = \frac{TP}{FP + TP} \quad (3)$$

Dice 系数是用于衡量两个样本集合相似度的指标,常用于语义分割任务中,计算的是预测结果和真实标签之间的重叠度。

$$Dice = \frac{2TP}{FP + 2TP + FN} \quad (4)$$

IoU 也被称为 Jaccard 指数,主要用于度量两个区域的重叠度,常用于目标检测和图像分割任务中。IoU 的定义为:

$$IoU = \frac{TP}{FP + TP + FN} \quad (5)$$

### 3.4. 对比试验

针对视网膜眼底图像公开数据集和膀胱肿瘤图像,为了验证提出的模型的性能,选取较为主流的分割模型 Unet++ [9]、Attention-unet [10]、Unet3+ [11]、Transunet [12],对比实验结果见表 1 和表 2。

**Table 1.** Comparison of experimental results  
**表 1.** 对比实验结果

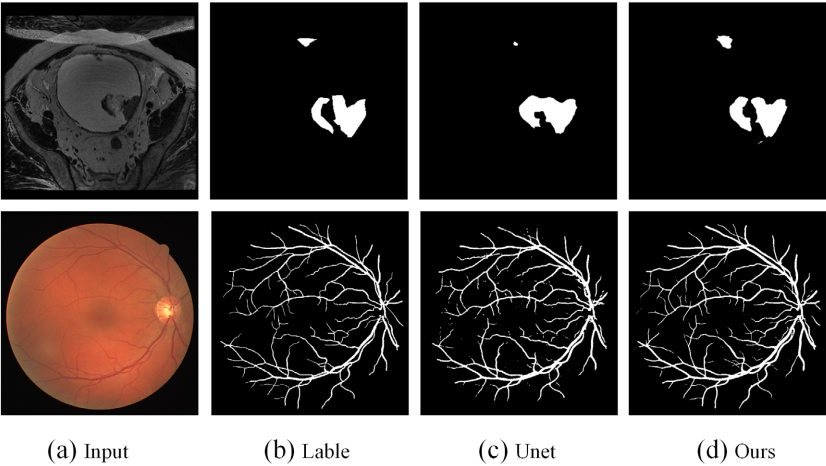
Model	Precision	Dice	IoU
Unet	95.61	82.22	71.23
Attention-unet	95.32	80.94	71.15
Unet++	95.64	81.92	70.95
Unet3+	97.03	80.32	71.04
Transunet	97.16	82.32	70.97
Ours	97.21	82.57	71.93

根据表 1 和表 2 的结果表明, Akmamba-Net 算法相较于 Unet、Attention-unet、Unet++、Unet3+、Transunet 等模型,在视网膜眼底 MRI 图像中 Precision 提升了 1.6%、1.89%、1.57%、0.18%、0.11%, Dice 系数提升了 0.35%、1.63%、0.65%、2.25%、0.25%, IoU 提升了 0.7%、0.78%、0.98%、0.89%、0.96%,

在膀胱肿瘤 MRI 图像中 Precision 提升了 8.52%、9.04%、2.5%、3.99%、3.32%，Dice 系数提升了 4.31%、0.75%、9.86%、2.33%、3.55%，IoU 提升了 6.72%、1.13%、13.83%、3.53%、6%，更适用于医学图像分割的任务，证实了改进算法的优越性和综合性能。AKConv 模块通过动态调整卷积核形状，有效捕捉了视网膜血管的细小分支和复杂拓扑结构，而 Mamba-out 模块通过跨层特征融合增强了血管边缘的连续性。膀胱肿瘤形态不规则且边界模糊，传统 U-Net 的固定卷积核难以适应此类特征。AKConv 的空间重采样机制以及 Mamba-out 动态缩放策略则抑制了背景噪声的干扰，在 Dice 系数上提升了 4.3%。视网膜眼底和膀胱肿瘤 MRI 图像相较于原始模型的分割效果见图 4。

**Table 2.** Comparison of experimental results  
**表 2.** 对比实验结果

Model	Precision	Dice	IoU
Unet	81.12	85.67	74.93
Attention-unet	80.60	89.23	80.52
Unet++	87.14	80.12	67.83
Unet3+	85.65	87.65	78.12
Transunet	86.32	86.43	75.65
Ours	89.64	89.98	81.65



**Figure 4.** Akmamba-Net model segmentation performance comparison  
**图 4.** Akmamba-Net 模型分割效果对比

3.5. 消融试验

为了验证本文模型性能的优越性，针对膀胱肿瘤图像，为了验证提出的模型的性能，将本文提出的 Akconv 模块、mamba-out 模块进行了一系列的消融实验，结果见表 3。

结果表明，针对膀胱肿瘤图像，Unet 通过引入 Akconv 和 mamba-out，引入 Akconv 后，Precision、Dice、IoU 提升了 6.22%、1.65%、2.31%，证明了 Akconv 能够提高特征提取的精度和鲁棒性，引入 mamba-out 后，Precision、Dice、IoU 提高了 7.6%、2.35%、4.08%，证明了 mamba-out 能够有效提取并融合输入数据中的空间和时间信息。

**Table 3.** Results of the ablation experiment  
**表 3.** 消融实验结果

Model	Precision	Dice	IoU
Unet	81.12	85.67	74.93
+Akconv	87.34	87.32	77.24
+mamba-out	88.72	88.02	79.01
+Akconv +mamba-out	89.03	88.97	80.21
Ours	89.64	89.98	81.65

#### 4. 结论

本文针对视网膜血管分割不佳和膀胱肿瘤图像分割边界模糊、误判等现象，提出了 Akmamba-Net 网络模型，创新性地提出了 Akconv 模块对输入的特征图进行空间调整，从而增强模型对空间信息的捕捉能力，为传统卷积操作引入了空间调整的能力，使得模型能够在处理空间复杂性时表现得更加灵活。同时引入 mamba-out 模块，使用了注意力机制来动态地加权输入特征，提高对关键区域的关注度，通过集成多尺度卷积和跨层信息传递，从而提升模型的分割精度。经过实验验证在视网膜眼底和膀胱肿瘤 MRI 图片上展现出了优越的分割性能，通过对比试验和消融实验证实了其在医学图像分割场景的应用潜力，满足医学图像分割要求。

#### 基金项目

国家自然科学基金资助项目(11502145)。

#### 参考文献

- [1] 诗慧陈, 维湘刘, 璟秦, 等. 基于深度学习和医学图像的癌症计算机辅助诊断研究进展[J]. 生物医学工程学杂志, 2017, 34(2): 314-319.
- [2] 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](https://doi.org/10.1007/978-3-319-24574-4_28)
- [3] 曹心姿, 梁秋源, 李瑞新, 等. 基于改进的 U-Net 网络在 MRI 膀胱肿瘤图像的分割[J]. 计算机与数字工程, 2021, 49(7): 1442-1447.
- [4] Li, X., Wang, J., Wei, H., Cong, J., Sun, H., Wang, P., et al. (2024) MH2AFormer: An Efficient Multiscale Hierarchical Hybrid Attention with a Transformer for Bladder Wall and Tumor Segmentation. *IEEE Journal of Biomedical and Health Informatics*, **28**, 4772-4784. <https://doi.org/10.1109/jbhi.2024.3397698>
- [5] Wang, Z., Zheng, J.Q., Zhang, Y., et al. (2024) Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation. arXiv: 2402.05079.
- [6] Deari, S., Oksuz, I. and Ulukaya, S. (2023) Block Attention and Switchable Normalization Based Deep Learning Framework for Segmentation of Retinal Vessels. *IEEE Access*, **11**, 38263-38274. <https://doi.org/10.1109/access.2023.3265729>
- [7] 胡鸿翔, 丁学明, 丁雪峰, 等. 基于全分辨率卷积交互网络的视网膜血管分割[J/OL]. 控制工程, 1-10. <https://kns.cnki.net/kcms2/article/abstract?v=cO3sNaablhqcqwEyCREI-3H4FJwn2ZHkyjyHEHX-Prtm9hCpIM1QFEopRixMXISJBerfqJL2nZfXQKC2n6Buw9QQ-TP-3-Vn3SxV3QISSY4rOEmSgRSYnf6LAaX-bci echdu43budi09ESr5qm edieyewlnksl4cyROQgkfcZNFjufgvl81EPR3j0BguDxA&uniplatform=NZKPT&language=CHS>, 2025-03-21.
- [8] Zhang, X., Song, Y., Song, T., et al. (2023) AKConv: Convolutional Kernel with Arbitrary Sampled Shapes and Arbitrary Number of Parameters. arXiv: 2311.11587.
- [9] 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](https://doi.org/10.1007/978-3-030-00889-5_1)
- [10] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999.
- [11] 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>
- [12] Chen, J., Lu, Y., Yn, Q., *et al.* (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv: 2102.04306.