

# 集成自注意力机制的医学图像分割方法

赵凡, 张学典

(上海理工大学光电信息与计算机工程学院, 上海 200093)

**摘要:** 针对UNet架构在医学图像分割中捕捉局部特征及保留边缘细节的局限性, 提出了一种融合自注意力机制的改进型UNet算法。该算法基于传统编码-解码结构, 引入多尺度卷积(Multi-scale convolution, MSC)模块以实现多粒度特征提取, 同时集成卷积-自注意力(Convolution mixer attention, CMA)模块, 结合卷积层的局部特征建模和自注意力层的全局上下文建模。在BUSI和DDTI数据集分割任务中, 相比现有经典网络架构, 大量实验数据验证了本模型优异的分割能力。此外, 统计学数据分析、消融实验进一步验证了MSC和CMA模块的有效性。该研究为高精度医学图像分割提供了一种创新方法, 对于促进医学诊断的精确性和效率具有重要的理论与实践意义。

**关键词:** UNet; 医学图像分割; 卷积神经网络; 多尺度卷积; 注意力机制

中图分类号: TP391

文献标志码: A

## Medical Image Segmentation Method with Integrated Self-attention

ZHAO Fan, ZHANG Xuedian

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** Aiming at the limitations of the UNet architecture in capturing local features and preserving edge details in medical image segmentation, this paper presents an improved UNet algorithm integrating self-attention mechanism. The proposed algorithm is based on traditional encoder-decoder structure, incorporating a multi-scale convolution (MSC) block for multi-granularity feature extraction, and a convolution mixer attention (CMA) block, which combines the modeling of local features by convolutional layers with global contextual modeling by self-attention layers. In the segmentation task of BUSI and DDTI datasets, compared with the existing classical network architecture, a large number of experimental data verify the excellent segmentation ability of the model. Additionally, Statistical data analysis and ablation studies further confirm the effectiveness of the MSC and CMA modules. This research provides an innovative approach for high-precision medical image segmentation, holding significant theoretical and practical implications for enhancing the accuracy and efficiency of medical diagnoses.

**Key words:** UNet; medical image segmentation; convolutional neural network (CNN); multi-scale convolution (MSC); attention mechanism

## 引言

医学图像分割在精确分离医学影像中的组织、器官或病变区域方面发挥关键作用,对于医生定位和分析异常病变、促进早期检测与预防,以及治疗规划、监控和外科手术的精准定位信息提供了重要支持<sup>[1-3]</sup>。然而,医学图像的复杂性,尤其是像素的单一性和病变部位边界的模糊性,使得准确分割成为一项挑战。

深度学习技术的发展为医学图像分割带来了显著突破<sup>[4-5]</sup>。自2012年 AlexNet<sup>[6]</sup>展示了卷积神经网络在图像识别方面的优势以来,一系列改进及创新算法相继出现。2015年提出的全卷积网络(Fully convolutional network, FCN)<sup>[7]</sup>通过将卷积层替换成全连接层,实现了像素级预测。随后,UNet算法<sup>[8]</sup>基于FCN提出编码-解码结构,通过在跨层连接时使用拼接来实现更深层次的特征提取。此后,UNet的各种变体,如Nested UNet、Attention UNet和Residual UNet<sup>[9-11]</sup>,分别通过引入多层次子解码器、集成注意力机制和残差连接,不断提升医学图像分割效果。

尽管如此,U-Net模型主要依赖于卷积神经网络,这在捕捉全局上下文信息方面存在局限性。特别是在医学图像中,组织结构和病变之间常呈现广泛的空间关系,仅依靠局部信息可能难以准确分割复杂区域<sup>[12-15]</sup>。为此,近年来开始将自注意力机制(Transformer)<sup>[16]</sup>融入医学图像分割,例如MedT、TransUNet、Swin\_UNet和nnFormer等<sup>[17-20]</sup>,通过模型输入序列中元素之间的全局依赖关系建模,更好地理解图像内不同区域间的联系。而将Transformer融入UNet的网络多聚焦于局部特征学习,对小目标内部细节和外部轮廓边界的表达不足,不能完全满足医学实际应用中对分割精度的严格要求<sup>[21-22]</sup>。

鉴于此,本研究提出的集成自注意力分割网络(Multi-focus SegNet, MFSegNet)旨在优化UNet网络、提高分割精度和全局理解能力。本文模型在编码-解码结构中嵌入了卷积-自注意力(Convolution mixer attention, CMA)模块和多尺度卷积(Multi-scale convolution, MSC)模块,其中CMA模块结合卷积层和自注意力层,优化局部特征与全局上下文信息的融合;MSC模块在编码阶段通过不同尺寸卷积核构建语义特征金字塔<sup>[23]</sup>,实现多尺度特征提取,在解码阶段专注于提取局部细节和整体结构特征,从而提升分割结果的准确性。

## 1 集成自注意力分割网络模型 MFSegNet 的结构设计

### 1.1 改进的主干特征提取网络设计

为克服UNet模型在局部特征识别和边缘细节保持方面的限制,本研究提出了集成自注意力分割网络MFSegNet。本模型基于传统的编码-解码结构,同时融合了CMA模块和MSC模块,以增强医疗图像的语义分割性能。CMA模块在编码器和解码器之间进行信息的传递和融合,利用自注意力机制的全局上下文建模与卷积层的局部特征建模相结合,实现更精确的特征表达。MFSegNet网络架构如图1所示。

从图1可以看出,本模型的MSC模块被嵌入编码和解码网络的不同卷积层中,用于执行多尺度特征提取。在编码阶段,MSC模块通过结合不同尺度的卷积核( $1\times 1$ 、 $3\times 3$ 、 $7\times 7$ ),逐级增大通道数( $32\rightarrow 64\rightarrow 128\rightarrow 256\rightarrow 512$ ),实现多粒度的特征表达,构建起一个能够有效捕获丰富语义信息的特征金字塔结构。而在解码阶段,网络首先将上采样操作与编码特征融合,随后通过MSC模块将编码路径的语义特征融入细节恢复过程中,从而提升分割的精确度。这种对称的U形框架确保编码路径专注于提取图像的语义信息,而解码路径则专注于恢复图像的细节特征。改进模块及其细节将在1.2和1.3节中详述。

### 1.2 多尺度卷积 MSC 模块

针对U-Net网络中,边缘特征信息提取不足的缺点,本文提出MSC模块。如图2所示,通过不同大

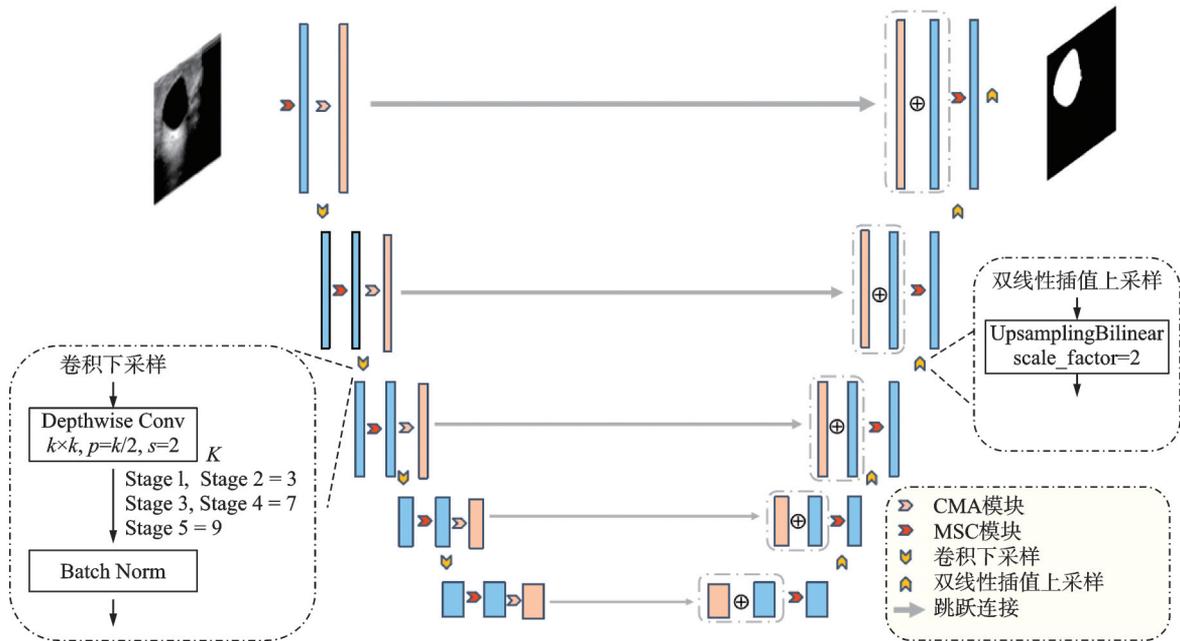


图1 MFSegNet 网络架构

Fig.1 Architecture of MFSegNet network

小的卷积核,识别不同对象的各种上下文信息,以增强编码器-解码器网络在图像分割任务中的特征表达能力,提升模型的分割性能与全局理解能力。

在编码路径中,MSC模块首先通过感受野为 $3 \times 3$ 的卷积核,快速捕获局部特征和边缘信息;然后引入Batch Norm(BN)层进行特征正则化,GELU层增强特征表达能力,接着 $1 \times 1$ 卷积进行维度缩减和非线性映射,实现跨通道交互和提升非线性能力,同时降维减小参数量;随后 $7 \times 7$ 的卷积获取更广泛的上下文信息,有助于全局结构的理解。大感受野用于识别图像中的较大目标特征,小感受野用于提取图像的细微特征,这一多尺度卷积结构在编码路径中构建了图像的语义特征金字塔结构,充分学习了图像的多样性与层次性。每个卷积后嵌入BN层和GELU层,以提高网络泛化能力。最后使用 $k \times k$ 卷积核、padding= $k/2$ 、步幅为2的卷积进行下采样。相较于传统的池化操作,它可以通过调整卷积核的尺寸和步幅来控制特征图的缩小程度,在保留边界信息的同时,更全面地捕获丰富的多尺度语义信息,从而进一步提升特征的表征能力。

在解码路径中,MSC模块表现出了较强的特征融合与上采样能力。首先通过双线性插值上采样,将分辨率恢复到编码前的大小,还原图像细节信息,以实现更连续平滑的特征图放大;随后,将激活前的特征图与激活后的特征图相加,实现残差链接;接下来同样是连续的 $3 \times 3$ 、 $1 \times 1$ 、 $7 \times 7$ 不同尺寸卷积,提取不同尺度的特征,恢复精细信息;同时每个卷积层后嵌入BN层和GELU继续增强解码特征的表达能力。在这一过程中,大感受野卷积强调精确定位,小感受野卷积则增强局部特征,促进多尺度特征的逐级融合。

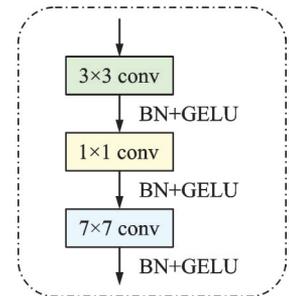


图2 MSC 模块

Fig.2 MSC block

### 1.3 卷积-自注意力 CMA 模块

本模型中提出 CMA 模块以增强医学图像分割网络的特征表达能力。该模块集成了卷积、残差连接以及自适应通道注意力计算,形成高效且可控的注意力特征增强单元,进而提高编码器的特征学习能力。该模块逐层递进处理特征信息,使模块具备了强大的特征提取和表示能力。

CMA 模块如图 3 所示,首先通过参数量较小、感受野较大的  $7 \times 7$  卷积,提取输入图像空间语义信息,后嵌入 BN 层和 GELU 融合多尺度特征并增加非线性,通过残差连接得到输出特征  $X'_1$ ;其次通过  $1 \times 1$  点卷积进行非线性映射和跨通道交互,融合来自不同卷积的特征,随后同样 BN 层和 GELU 与 block 1 的输出特征  $X'_1$  相加,实现残差连接得到  $X'_2$ ,进一步提升了特征的表达能力;接着通过  $1 \times 1$  卷积将通道数转换为输入特征图  $X_2$  的通道,其中 Sigmoid 激活函数学习每个特征通道的注意力权重,设置高注意力权重加强对目标和语义相关通道的特征表达,设置低注意力权重抑制不那么重要的通道弱化特征表达,以实现自适应的通道选择机制,得到的注意力系数和原始输入  $X$  对应的通道特征逐元素相乘,进行特征加权;最后与模块的输入特征  $X$  残差连接,得到最终输出特征  $Y$ 。本模块通过可学习的注意力权重,对特征进行自适应通道选择的操作,增强对重要特征通道的表达能力和对重要特征的识别能力,可表示为

$$X'_1 = \text{BN}(\alpha\{\text{Conv}_{7 \times 7}(X)\}) + X \tag{1}$$

$$X'_2 = \text{BN}(\alpha\{\text{Conv}_{1 \times 1}(X'_1)\}) + X'_1 \tag{2}$$

$$Y = \sigma\{\text{Conv}_{1 \times 1}(X'_2)\} \otimes X + X \tag{3}$$

式中: $X$  表示输入特征图; $X'_1$  表示 block 1 输出; $X'_2$  表示 block 2 输出; $Y$  表示输出的特征图; $\alpha$  表示 GELU 激活函数;BN 表示批量归一化 Batch Norm; $\sigma$  表示 Sigmoid 激活函数;“ $\otimes$ ”表示逐元素相乘;“+”表示逐元素相加。

## 2 实验及结果分析

在 BUSI 和 DDTI 两个公开数据集上进行分割实验,以验证本文优化网络的可行性和优越性。并选取经典医学图像分割网络 Nested UNet<sup>[9]</sup>、UNet、Att\_UNet、UNext<sup>[24]</sup>、MISSFormer<sup>[25]</sup>、UCTransNet<sup>[26]</sup>、Swin\_UNet、MTUNet<sup>[27]</sup> 以及本文提出的优化模型进行对比实验,在训练数据集进行训练,并在测试数据集上进行性能评估。

### 2.1 实验数据集与参数配置

#### 2.1.1 超声数据集

超声图像的低对比度、阴影丰富、背景噪声强烈和结构复杂等特点会导致病变区域边界模糊,因此给精确识别和分割带来困难。目前,深度学习技术在前列腺和颈动脉等超声图像的诊断中广泛应用,并取得了较好分割效果。然而,在乳腺和甲状腺超声图像分割中,由于乳腺的高脂肪组织内容和低对比度会造成图像边界模糊,以及甲状腺图像中不均匀的斑点噪声和复杂周围结构会带来多结节现象,因此对分割精度的要求极高,现有研究还存在一定局限性。基于此,本文分别使用乳腺癌分割、甲状腺癌分割两种不同的医学超声图像分割任务来评估本模型。乳腺癌分割任务使用乳腺超声图像(BUSI)

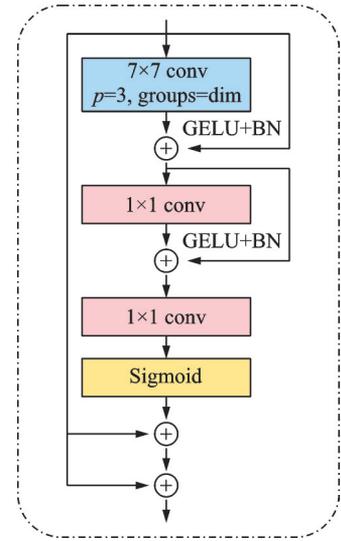


图3 CMA 模块  
Fig.3 CMA block

数据集<sup>[28]</sup>的公开数据,该数据集收集自600名女性患者的乳腺超声图像,包含78张图像,涵盖了133例正常病例、487例良性病例和210例恶性病例,每个病例都有相应的真实标签。本实验从数据集中选择了良性和恶性病例共647张图像。甲状腺癌分割任务使用甲状腺超声图像(DDTI)数据集<sup>[29]</sup>,该数据集由Pedraza等提供,包含来自单个设备的637张带像素级标签的超声甲状腺成像,其中包括甲状腺炎、囊性结节、腺瘤及甲状腺癌等病例图像。甲状腺结节与周围正常组织之间的低对比度和结节形状及边界的模糊性使得对分割模型的要求较高。

为评估该模型在BUSI和DDTI数据集上的性能,实验将所有图像调整为 $256 \times 256$ 的分辨率,两个数据集按8:2的比例随机分割为训练集和测试集,并采用5折交叉验证法,通过迭代结果计算平均值和标准差,以验证本模型在不同场景下的性能稳定性和可靠性。

### 2.1.2 实验参数配置

实验基于PyTorch框架,使用NVIDIA GeForce RTX 3090Ti GPU进行模型训练。在训练阶段,Epoch数设置为150,每个batch包含4张图像样本,采用SGD优化器,学习率为0.001,权重衰减系数为0.000 01,以提高模型的泛化性能。实验参数设置详见表1。

## 2.2 评价指标

为了定量评估优化模型的性能,选用召回率(Recall, Rec)、特异性(Specificity, Spec)、精确率(Precision, Pre)、 $F_1$ 值( $F_1$ -score)、准确率(Accuracy, Acc)和交并比(IoU)等评价指标来衡量分割性能。召回率指正确预测占实际为总的比例,召回率较高,表明模型对正例的识别能力较高;特异性指预测为总的所有负样本占实际为总的比例,较高的特异性表明模型对负例的识别能力较强;精确率指正确预测占所有预测为总的比例,精确率越高,表明模型分割效果越好; $F_1$ 值指模型精确率和召回率的综合加权, $F_1$ 值较高,表明精确率和召回率都较高;准确率指正确预测与所有预测的比例,准确率较高,表明模型预测正样本能力较强;交并比又称雅卡尔指数(Jaccard index),用来衡量两个样本的相似性,交并比较高,表明图像分割效果较好。各项指标分别定义为

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$F_1 = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \quad (7)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (8)$$

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

式中:TP表示预测结果为正类,真实为正类;FP表示预测结果为正类,真实为负类;FN表示预测结果为负类,真实为正类;TN表示预测结果为负类,真实为负类。

表1 实验参数设置

Table 1 Experimental parameter setting

参数	数值
Epoch	150
Batch大小	4
优化器	SGD
学习率	0.001
权重衰减系数	0.000 01
GPU	3 090Ti
框架	Pytorch

### 2.3 乳腺图像分割实验及分析

将本模型与其他几种图像分割算法在BUSI数据集上进行对比试验,以验证本文所提方法的有效性。表2列举了不同模型在乳腺分割上的分割结果。

表2 不同模型在BUSI数据集上的分割结果对比

Model	Rec	Spec	Pre	$F_1$	Acc	IoU	%
Nested UNet	82.85±2.87	98.13±1.07	77.84±3.87 <sup>*</sup>	76.30±3.38	96.03±2.07 <sup>*</sup>	67.27±3.05	
UNet	80.92±3.83	98.08±1.39	79.55±4.13	75.10±2.21	95.92±2.15	66.61±2.85	
Att_UNet	81.90±3.79	98.07±1.38	78.10±2.03	75.75±2.16	95.98±2.09	67.54±3.34	
UNext	74.77±4.03	98.34±1.09 <sup>*</sup>	78.03±3.07	70.22±3.31	95.57±2.14	61.12±3.07	
MISSFormer	76.36±4.21	96.39±1.04	68.06±3.95	68.95±2.23	94.25±2.62	58.39±3.22	
UCTransNet	71.45±4.17	97.62±1.47	72.66±3.09	62.98±3.09	94.73±2.20	52.94±3.46	
Swin_UNet	82.17±3.14	97.63±1.70	78.72±2.99	78.22±2.71 <sup>*</sup>	96.01±1.97	68.42±2.31 <sup>*</sup>	
MTUNet	82.54±2.91	97.59±1.73	75.96±2.71	76.77±3.07	95.73±2.91	67.33±3.98	
MFSegNet	<b>82.83±2.52</b>	<b>98.37±1.02</b>	<b>83.79±1.93</b>	<b>0.816±1.79</b>	<b>96.41±1.72</b>	<b>73.45±2.97</b>	

注:加粗表示最佳分割效果,“\*”表示使用 Student’s t-test。

由表2可知,在BUSI数据集分割中,本文模型召回率接近最优值,仅低于最优网络Nested UNet 0.21%,表明模型能够准确地捕获和分割图像中的目标区域,其较低的标准差进一步证明了模型的稳定性,这得益于自注意力机制对边缘信息的强化关注,更全面地覆盖目标区域,并更精确地识别图像中的目标结构;特异性高于次优网络Nested UNet且标准差低于次优网络,说明模型在负样本区分方面表现出色;精确率高于UNet 4.24%,并显示出最小的标准差,表明模型在预测为正样本时较高的准确性;准确率高出Nested UNet 3.8%,表明模型较强的正确预测能力,其CMA模块中融合的Sigmoid激活函数和注意力机制,通过专注于关键目标区域的学习,减少了不必要的预测误差,在提高预测准确性的同时,也增强了模型的整体准确率; $F_1$ 值高于Swin\_UNet 3.38%,说明模型能较好地平衡精确度和召回率;交并比高于Nested UNet 5.91%,远高于Nested UNet和所有其他模型,且标准差远低于其他模型,这表明分割结果与实际标签之间高度的一致性,同时展现了分割结果的稳定性和可靠性。这些结果进一步验证了自注意力机制和多尺度特征融合技术在边缘信息识别与处理方面的有效性,有助于生成更精确的分割结果,从而提升了整体分割性能。

可视化示例如图4所示,其中每行代表一个不同的超声图像样本,每列为不同网络的分割结果,以及原始输入图像和对应的真实分割(Ground truth, GT)结果。由图4可以看出,在处理边界清晰的组织图像时,各网络能够较为准确地分割乳腺结节。而在边界模糊的样本中,如MISSFormer、UCTransNet和MTUNet在边界丢失关键的边缘信息,并未能分割出乳腺结节的完整形状,Nested UNet、UCTransNet在某些样本上出现误分割。相比之下,本文模型在这些图像分割任务中表现较优异,其分割结果在与真实分割(GT)最为接近,无论是在整体性还是边缘细节的精确捕捉上,都显示了其优势,通过CMA模块中的自注意力机制更好地保留了关键的结构边缘,分割结果中几乎没有噪声的干扰,进一步验证了本模型在定位和分割关键区域方面的高精度和可靠性。

为进一步验证该方法优越性,本文采用 Student’s t-test 进行分析,利用 SPSS 软件比较本模型(实验组)与第2名方法(对照组)的性能差异,召回率(Rec)本模型排名第2不参与比较。对于特异性

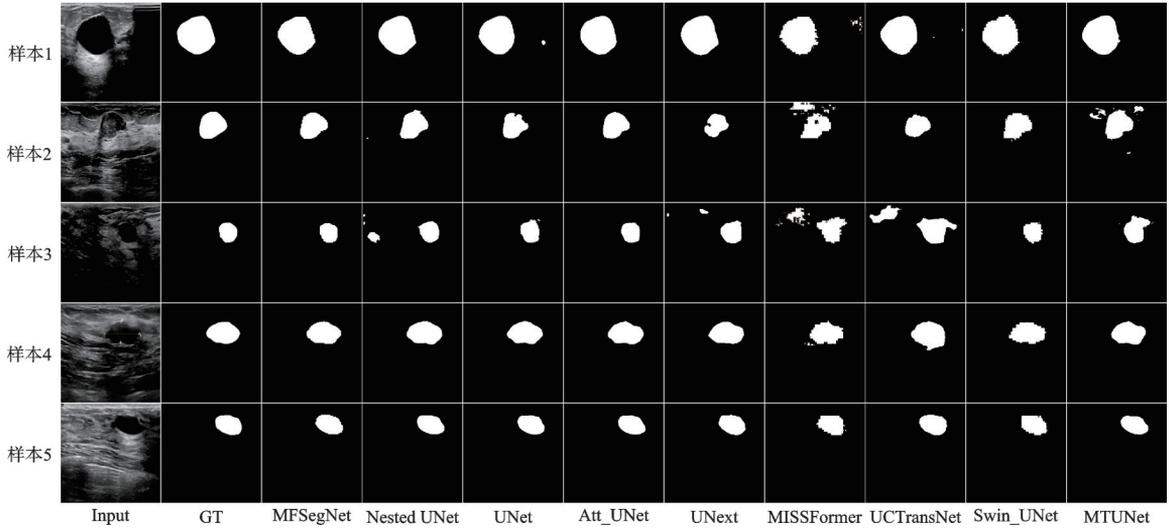


图4 各算法在BUSI数据集上的分割结果对比

Fig.4 Comparison of segmentation results of various methods on BUSI dataset

(Spec), 对照组模型为 UNext; 对于准确率 (Acc) 和精确率 (Pre), 对照组模型为 Nested UNet; 对于  $F_1$  值和交并比 (IoU), 对照组模型为 Swin\_UNet。计算  $p$  值如表 3 所示,  $p$  值均小于 0.05, 从统计学角度证明了本方法的优越性。

#### 2.4 甲状腺图像分割实验及分析

为进一步证明本文所提的分割模型网络模型确实有效并验证鲁棒性, 将其与其他几种图像分割算法在 DDTI 数据集上进行泛化试验, 以评估所提网络的分割性能。表 4、图 5 列举了不同模型在甲状腺分割上的分割结果。

表 4 为本模型在 DDTI 数据集上的定量分析结果。由表 4 可知, MFSegNet 的平均召回率最高、并具有相对较低的标准差, 这说明模型能够准确捕获和分割目标

表 3 Student's t-test 在 BUSI 数据集上的统计分析结果

Table 3 Statistical analysis results of student's t-test on BUSI dataset

评价指标	$p$
Spec	0.046 7
Pre	0.038 4
Acc	0.036 7
$F_1$	0.023 2
IoU	0.028 9

表 4 不同模型在 DDTI 数据集上的分割结果对比

Table 4 Comparison of segmentation results of different models on DDTI dataset

Model	Rec	Spec	Pre	$F_1$	Acc	IoU	%
Nested UNet	80.25±3.67	<b>98.35±2.01</b>	83.86±2.07	79.44±2.79	96.56±0.79*	69.16±2.91	
UNet	84.17±1.68*	97.85±2.17	80.34±2.71	79.58±2.34	96.38±0.91	69.18±2.89	
Att_UNet	82.97±2.59	97.93±2.61	82.29±3.09	79.72±2.49*	96.46±0.55	69.76±3.01*	
UNext	76.78±3.79	97.83±2.34	74.82±3.39	71.48±3.39	95.54±0.83	59.39±3.09	
MISSFormer	73.84±4.04	96.34±3.07	67.26±3.74	64.84±3.94	93.75±0.84	51.00±2.96	
UCTransNet	81.89±2.29	96.71±2.74	70.22±3.59	70.48±3.74	95.14±0.67	58.19±3.08	
Swin_UNet	82.28±2.91	97.95±2.71	80.29±2.77	78.78±3.04	96.46±0.92	68.21±2.92	
MTUNet	79.76±3.09	97.90±3.53	79.48±2.81	76.52±3.51	95.96±0.93	64.56±2.86	
MFSegNet	<b>84.27±2.06</b>	98.25±2.10	82.58±2.01	<b>81.23±2.27</b>	<b>96.74±0.41</b>	<b>71.36±2.57</b>	

注: 加粗表示最佳分割效果, "\*"表示使用 Student's t-test。

区域,  $F_1$  值高于次优模型 Att\_UNet 1.52%, 标准差低于其他网络, 说明能够专注于关键病变区域, 准确率比第 2 名模型 Nested UNet 高 0.18%, 交并比指标高于第 2 名模型 Att\_UNet 1.6%, 这是由于集成深度可分离卷积和自注意力机制增强了对医学图像全局病变结构和局部关键特征的表达能力, 使图像在边缘模糊、噪声水平高的病变区域也能够稳定、精确地识别出目标区域。实验数据证明, 本实验针对低对比度、阴影复杂及噪声高的超声图像时, 提升了分割精度, 同时展现了良好的泛化能力。

甲状腺实例分割结果如图 5 所示, 由图可知 Att\_UNet 和 MTUNet 在处理图像中的噪声方面存在不足, 错误地将噪声点识别为目标区域的部分, 导致分割结果存在明显误差。相反, MFSegNet 通过多尺度卷积对噪声有较强的抑制能力, 在图像噪声较多的情况下也能较好分割。特别在图 5 的最后一行甲状腺组织的边界与周围组织对比度不明显, 多数对比网络难以准确捕获完整的甲状腺组织, 导致信息丢失, 分割不完全。而 MFSegNet 通过引入注意力机制显著增强了对甲状腺区域特征的识别, 较好地分割出完整图像, 说明本模型能够更好地提高模型对甲状腺细微特征的捕捉能力, 即使是在边界模糊的情况下, 也能保证分割的完整性和精确性。

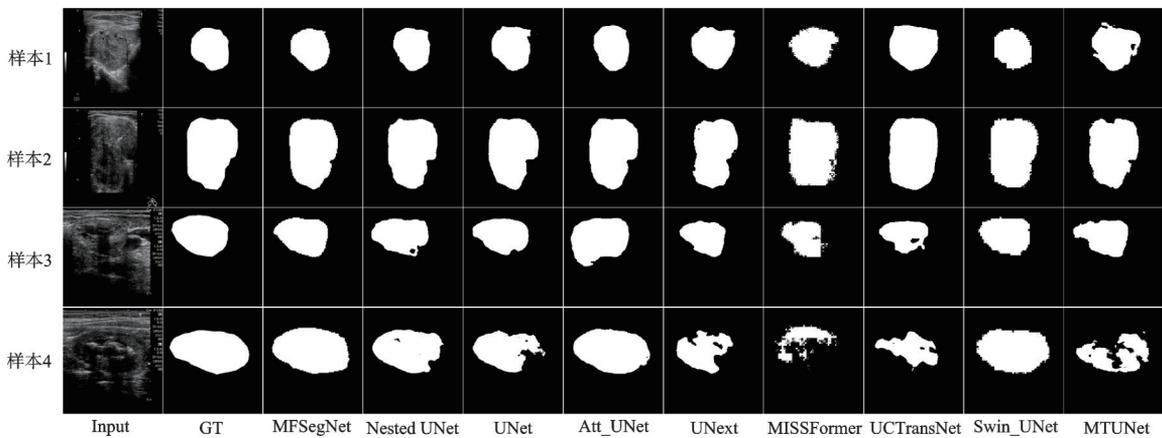


图 5 各算法在 DDTI 数据集上的分割效果对比

Fig.5 Comparison of segmentation results of various methods on DDTI dataset

为进一步验证实验结果, 同样采用 Student's t-test 进行统计分析, 对于特异性 (Spec) 和精确率 (Pre), 本模型排名第 2 不参与计算; 对于  $F_1$  值和交并比 (IoU), 对照模型为 Att\_UNet; 对于准确率 (Acc), 对照模型为 NestedUNet; 对于召回率 (Rec), 对照模型为 Att\_UNet。计算  $p$  值如表 5 所示,  $p$  值均小于 0.05, 通过统计学角度证明了本方法的优越性。

### 2.5 消融实验及有效性分析

本文在 BUSI 数据集上进行消融实验, 以评估 MFSegNet 嵌入模块的有效性, 其中实验涵盖了 3 个关键组件, 将 MSC 模块、CMA 模块去除作为 Baseline 模型, Baseline 的基础上加入 MSC 模块、CMA 模块逐步改进性能, 并采用  $F_1$  值、准确率及交并比对实验结果进行评价。通过在相同数据集上进行训练和测试, 模型在不同组件下的表现如表 6 所示, 引入 MSC 模块,  $F_1$  值和交并比分别提高了 1.27% 和 2.1%, 表明 MSC 模块通过多尺度卷积、下采样和上采样等操作, 增强了模型对不同尺度的特征的捕捉能力; 进一步添加 CMA 模型的指标得到较大提升,  $F_1$  值、准确率和交并比分别提高了

表 5 Student's t-test 在 DDTI 数据集上的统计分析结果

Table 5 Statistical analysis results of Student's t-test on DDTI dataset

评价指标	$p$
Rec	0.023 9
$F_1$	0.032 7
Acc	0.030 9
IoU	0.028 9

3.65%、0.26% 和 4.22%，表明 CMA 模块中引入的自注意力机制有助于模型更加准确关注重要的特征区域，增强模型的特征表达能力，显著改善了分割效果。

表 6 消融实验结果

Table 6 Ablation experimental results

组件	$F_1$	Acc	IoU
Baseline	77.96	96.15	69.23
Baseline +MSC Block	79.23	96.14	71.33
Baseline +MSC & CMA Block	<b>81.61</b>	<b>96.41</b>	<b>73.45</b>

注：加粗表示最佳分割效果。

### 3 结束语

本文在 UNet 网络的基础上，提出一种基于自注意力机制模块的卷积神经网络模型 MFSegNet。本分割模型嵌入了两个关键的模块：CMA 模块和 MSC 模块，并结合编码-解码结构模型，提升图像分割任务的性能。首先，CMA 模块引入混合器注意力机制，将卷积层对局部特征的建模和自注意力层对全局上下文的建模相融合。通过可分离卷积和乘法注意力操作，增强模型对全局依赖的学习能力，相比仅使用卷积层或自注意力层的网络更有效地提取不同尺度、层次特征，获得更好的分割效果；其次，MSC 模块引入多尺度卷积，相较于只使用固定尺寸卷积核的网络，多尺度卷积使模型能够更有针对性地捕获和利用关键特征，更好地提取病变的局部细节和整体结构特征；同时卷积下采样和双线性插值上采样操作，更全面地捕获多尺度语义信息，提升了分割性能；最后，在 BUSI 数据集上，特异性、精确率和准确率分别高于次优网络 0.24%、4.24% 和 3.8%。这些数据表明，本文方法在乳腺病变分割方面具有更好的性能。此外，该方法在甲状腺超声图像分割上的应用同样展现了其稳健性和泛化性，在 DDTI 数据集上的  $F_1$  值、准确率和交并比分别高于次优模型 1.52%、0.18% 和 1.6%，进一步验证了本模型在超声医学图像分割领域的广泛适用性和显著性能。

### 参考文献：

- [1] SHI F, WANG J, SHI J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19[J]. *IEEE Reviews in Biomedical Engineering*, 2020, 14: 4-15.
- [2] RAMESH K, KUMAR G K, SWAPNA K, et al. A review of medical image segmentation algorithms[J]. *EAI Endorsed Transactions on Pervasive Health and Technology*, 2021, 7(27): e6.
- [3] DUFFY S W, TABÁR L, YEN A M F, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women[J]. *Cancer*, 2020, 126(13): 2971-2979.
- [4] LOU A, GUAN S, LOEW M. DC-UNet: Rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation[C]//*Proceedings of Medical Imaging 2021: Image Processing*. [S.l.]: SPIE, 2021, 11596: 758-768.
- [5] CHOUDHARY A K, JANSCHKE A, GRUBESA T, et al. Grain size analysis in permanent magnets from Kerr microscopy images using machine learning techniques[J]. *Materials Characterization*, 2022, 186: 111790.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012. DOI: 10.1145/3065386.
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, [S.l.]: IEEE, 2015: 3431-3440.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//*Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*. Munich, Germany: Springer, 2015: 234-241.

- [9] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: A nested U-net architecture for medical image segmentation[EB/OL]. (2018-07-18). <https://doi.org/10.48550/arXiv.1807.10165>.
- [10] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[EB/OL]. (2018-04-11). <https://doi.org/10.48550/arXiv.1804.03999>.
- [11] ALOM M Z, YAKOPCIC C, HASAN M, et al. Recurrent residual U-Net for medical image segmentation[J]. *Journal of Medical Imaging*, 2019, 6(1): 014006.
- [12] 李清格, 杨小冈, 卢瑞涛, 等. 计算机视觉中的Transformer发展综述[J]. *小型微型计算机系统*, 2023, 44(4): 850-861.  
LI Qingge, YANG Xiaogang, LU Ruitao, et al. Transformer in computer vision: A survey[J]. *Journal of Chinese Computer Systems*, 2023, 44(4): 850-861.
- [13] 卢宏涛, 罗沐昆. 基于深度学习的计算机视觉研究新进展[J]. *数据采集与处理*, 2022, 37(2): 247-278.  
LU Hongtao, LUO Mukun. Survey on new progresses of deep learning based computer vision[J]. *Journal of Data Acquisition and Processing*, 2022, 37(2): 247-278.
- [14] 刘安安, 李天宝, 王晓雯, 等. 基于深度学习的三维模型检索算法综述[J]. *数据采集与处理*, 2021, 36(1): 1-21.  
LIU An'an, LI Tianbao, WANG Xiaowen, et al. Review of 3D model retrieval algorithms based on deep learning[J]. *Journal of Data Acquisition and Processing*, 2021, 36(1): 1-21.
- [15] 徐光宪, 冯春, 马飞. 基于UNet的医学图像分割综述[J]. *计算机科学与探索*, 2023, 17(8): 1776-1792.  
XU Guangxian, FENG Chun, MA Fei. Review of medical image segmentation based on UNet[J]. *Journal of Frontiers of Computer Science and Technology*, 2023, 17(8): 1776-1792.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-06-12). <https://doi.org/10.48550/arXiv.1706.03762>.
- [17] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: Gated axial-attention for medical image segmentation[C]//*Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021: 36-46.
- [18] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08). <https://doi.org/10.48550/arXiv.2102.04306>.
- [19] CAO H, WANG Y, CHEN J, et al. Swin-UNet: UNet-like pure transformer for medical image segmentation[C]//*Proceedings of European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 205-218.
- [20] ZHOU H Y, GUO J S, ZHANG Y H, et al. nnFormer: Interleaved transformer for volumetric segmentation[EB/OL]. (2022-02-04). <https://doi.org/10.48550/arXiv.2109.03201>.
- [21] AZAD R, AGHDAM E K, RAULAND A, et al. Medical image segmentation review: The success of U-Net[EB/OL]. (2022-11-27). <https://doi.org/10.48550/arXiv.2211.14830>.
- [22] 叶怀泽, 周子奇, 祁磊, 等. 基于特定领域解码的域泛化医学图像分割方法[J]. *数据采集与处理*, 2023, 38(2): 324-335.  
YE Huaize, ZHOU Ziqi, QI Lei, et al. Domain generalization via domain-specific decoding for medical image segmentation[J]. *Journal of Data Acquisition and Processing*, 2023, 38(2): 324-335.
- [23] 王子一, 徐苏平, 商琳. 非结构化数据的多粒度集成分类方法[J]. *南京航空航天大学学报*, 2020, 52(5): 723-728.  
WANG Ziyi, XU Suping, SHANG Lin. Multi-grained ensemble classification method for unstructured data[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2020, 52(5): 723-728.
- [24] VALANARASU J M J, PATEL V M. UNext: MLP-based rapid medical image segmentation network[C]//*Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2022: 23-33.
- [25] HUANG X, DENG Z, LI D, et al. Missformer: An effective transformer for 2D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2023, 42(5): 1484-1494.
- [26] WANG H, CAO P, WANG J, et al. UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI, 2022, 36(3): 2441-2449.
- [27] WANG H, XIE S, LIN L, et al. Mixed transformer U-Net for medical image segmentation[C]//*Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2022: 2390-2394.

- [28] AL-DHABYANI W, GOMAA M, KHALED H, et al. Dataset of breast ultrasound images[J]. *Data in Brief*, 2020, 28: 104863.
- [29] PEDRAZA L, VARGAS C, NARVÁEZ F, et al. An open access thyroid ultrasound image database[C]//Proceedings of the 10th International Symposium on Medical Information Processing and Analysis. [S.l.]: SPIE, 2015, 9287: 188-193.

**作者简介:**

赵凡(1998-),女,硕士研究生,研究方向:图像处理、医学影像分割,E-mail: zhaofan\_se@163.com。



张学典(1974-),通信作者,男,博士,教授,博士生导师,研究方向:计算机视觉、图像处理,E-mail: xd-zhang@usst.edu.cn。

(编辑:王静)