❖医学物理与工程学

Multi-scale feature weighted fusion and group gated attention UNet model for segmentation of endometrial cancer lesions in contrast-enhanced T1WI

LIU Chenxi[1], *LIU Liping*[2*], *REN Yulan*[3], *YU Changmin*[3]

(1. *College of Artificial intelligence, North China University of Science and Technology, Tangshan* 063210, *China*; 2. *School of Mechanical and Energy Engineering, Shanghai Electronic Information Vocational and Technical College, Shanghai* 201411, *China*; 3. *Department of Gynecologic Oncology, Fudan University Shanghai Cancer Center, Shanghai* 200032, *China*)

[**Abstract**] **Objective** To observe the value of multi-scale feature weighted fusion (MSWF) and group gated attention (GGA) UNet (MGA-UNet) model for segmentation of endometrial cancer (EC) lesions in contrast-enhanced T1WI (CE-T1WI). **Methods** CE-T1WI data of 327 EC patients were retrospectively enrolled and divided into training set (n=261) and validation set (n=66) at the ratio of 8 ∶ 2. Taken UNet as the backbone network, MSWF module and squeeze-and-excitation (SE) module were introduced into the encoder, and GGA mechanism was introduced into the skip connections, then a MGA-UNet model was constructed. The efficacy of the MGA-UNet model for segmenting EC lesions in CE-T1WI was observed and compared with other mainstream segmentation models. **Results** The accuracy, Recall, intersection over union and Dice similarity coefficient of MGA-UNet model for segmenting EC lesions in CE-T1WI was 96.84%, 95.73%, 92.37% and 88.96%, respectively, all superior to UNet, UNet++, Attention-UNet, SegNet, Swin-Unet and DeepLabV3+ models. **Conclusion** MGA-UNet model could be used to automately segment EC lesions in CE-T1WI.

[**Keywords**] endometrial neoplasms; magnetic resonance imaging; deep learning

DOI:10.13929/j.issn.1003-3289.2025.12.024

基于多尺度特征加权融合与分组门控注意力 UNet 模型用于分割对比增强 T1WI 中的子宫内膜癌病灶

刘辰熙[1],刘利平[2*],任玉兰[3],余长敏[3]

(1.华北理工大学人工智能学院,河北 唐山　063210;2.上海电子信息职业技术学院机械与能源工程学院,上海　201411;3.复旦大学附属肿瘤医院肿瘤妇科,上海　200032)

[摘　要] **目的** 观察基于多尺度特征加权融合(MSWF)与分组门控注意力(GGA)UNet(MGA-UNet)模型用于分割对比增强 T1WI(CE-T1WI)中的子宫内膜癌(EC)病灶的价值。**方法** 回顾性纳入 327 例 EC 患者的 CE-T1WI 数据,按8∶2比例将其分为训练集(n=261)与验证集(n=66)。以 UNet 为主干网络,将 MSWF 模块和通道注意力(SE)模块引入编码器,并将 GGA 机制引入跳跃连接处,构建 MGA-UNet 模型,观察其分割 CE-T1WI 中 EC 病灶的效能,并与其他主流分割模型进行对比。**结果** MGA-UNet 模型分割 CE-T1WI 中的 EC 病灶的准确率、召回率、交并比及戴斯相似系数分别为96.84%、95.73%、92.37%及 88.96%,均优于 UNet、UNet++、Attention-UNet、SegNet、Swin-Unet 及 DeepLabV3+模型。**结论** MGA-UNet 模型可自动分割 CE-T1WI 中的 EC 病灶。

[**第一作者**] 刘辰熙(2000—),男,河北保定人,在读硕士。研究方向:医学图像分割。E-mail: 15530828739@163.com

[**通信作者**] 刘利平,上海电子信息职业技术学院机械与能源工程学院,201411。E-mail: 11745430@qq.com

[**收稿日期**] 2025-05-08　　[**修回日期**] 2025-11-10

［关键词］ 子宫内膜肿瘤；磁共振成像；深度学习
［中图分类号］ R737.33；R445.2　［文献标识码］ A　［文章编号］ 1003-3289(2025)12-2050-06

子宫内膜癌(endometrial carcinoma，EC)是女性生殖系统常见恶性肿瘤，发病率和死亡率均呈逐年上升趋势[1-3]，通常采用以手术为核心的综合治疗方式，而为术前准确分期、识别图像中的病灶是制定手术策略和评估预后的关键。MRI 是 EC 术前评估的最佳影像学方法[4]，但具有主观依赖性。本研究以 UNet 架构为主干网络，通过残差架构将多尺度特征加权融合(multi-scale feature weighted fusion，MSWF)模块和通道注意力(squeeze-and-excitation，SE)模块引入编码器，并将分组门控注意力(group gated attention，GGA)机制引入跳跃连接处，构建基于 MSWF 与 GGA 的 UNet(MSWF and GGA UNet，MGA-UNet)模型，观察其用于分割对比增强 MR T1WI(contrast-enhanced T1WI，CE-T1WI)中的 EC 病灶的价值。

1　资料与方法

1.1　研究对象　回顾性收集 2018 年 12 月—2023 年 12 月复旦大学附属肿瘤医院 327 例经术后病理证实的 EC 患者，均为女性，年龄 26～78 岁、平均(53.9±9.5)岁；按 8∶2 比例将其随机分为训练集(n=261)与验证集(n=66)。排除标准：①图像质量差；②术前 MRI 未见明确病灶或无法准确勾画病灶 ROI；③缺少 T2WI 或 CE-T1WI；④于 MR 扫描前接受新辅助化学治疗。本研究经医院伦理审查委员会批准(1612167-18)，患者或家属均签署知情同意书。

1.2　仪器与方法　采用 Siemens Magnetom Skyra 3.0T MR 扫描仪及腹部线圈，嘱患者仰卧、头先进，平静呼吸下自髂嵴至耻骨联合下缘进行扫描。之后经静脉以 2～3 ml/s 流率注射钆喷酸葡胺(15 ml∶7.035 g)0.2 mmol/kg 体质量，延迟 120～180 s 采集轴位和冠状位脂肪抑制 CE-T1WI，TR 3.9 ms，TE 1.82 ms，矩阵 320×165，层厚 3.5 mm，FOV 440 mm×303 mm。

1.3　图像处理

1.3.1　标注　采用 3D Slicer 软件。由 1 名具有 3 年妇科影像学经验的主治医师 A 基于轴位 CE-T1WI，避开坏死、囊变、出血及子宫积液手动勾画病灶 ROI，利用三维重建技术重建病灶感兴趣体积(volume of interest，VOI)；之后由另 1 名具有 5 年妇科影像学经验的主治医师 B 再次勾画，且 1 个月后由医师 A 重复勾画，以观察者间及观察者内一致性良好的标注结果作为输入图像，并保存为 NIFTI 标准医学影像格式。

1.3.2　预处理　采用 PyCharm 2024.3.3 软件、Python3.8.8 语言和 SimpleITK 库，对 NIFTI 格式数据进行 2D 切片。首先将输入图像统一调整为 256×256 像素，以符合生理解剖特点的方式进行数据增强，包括小角度旋转(±10°)、调整亮度和对比度、添加椒盐噪声及高斯模糊，共获得 1 961 幅图像，包括训练集 1 537幅、验证集 424 幅。

1.4　MGA-UNet　MGA-UNet 网络主要由编码器-解码器结构和 GGA 模块组成。编码器采用双路径残差架构，主路通过级联 3×3 卷积层和 MSWF 模块实现多尺度特征动态融合、支路通过 1×1 卷积与 SE 模块增强注意力；采用残差连接融合上述 2 条路径特征，以有效捕捉并整合不同尺度下的病灶特征信息。解码器采用 2 个串联的 3×3 卷积提取特征，输出特征后经激活函数 Sigmoid 得到分割结果。同时将 GGA 模块引入跳跃连接，以提升模型对病灶区域的注意力。见图 1。

1.4.1　MSWF 模块　MSWF 模块由多尺度卷积分支、卷积块注意力模块(convolutional block attention module，CBAM)、权重生成网络及特征融合层组成(图 2)。输入特征图 $X \in R^{C\times H\times W}$ (C、H 及 W 分别代表输入特征图的通道数、高度及宽度)后，将通道维度均匀划分为 4 份，每份分别经过带不同大小(1×1、3×3、5×5、7×7)卷积核的卷积层处理，捕捉局部到全局的多尺度特征；每个分支输出后，引入 CBAM 模块，采用 SE 机制处理每个分支输出并生成每个通道权重，强调重要通道特征并抑制不相关通道；之后通过空间注意力机制进一步增强特征图空间表示能力，重点显示图像中病灶区域并抑制背景信息，进而获得特征图 Fi''。见公式(1)～(4)：

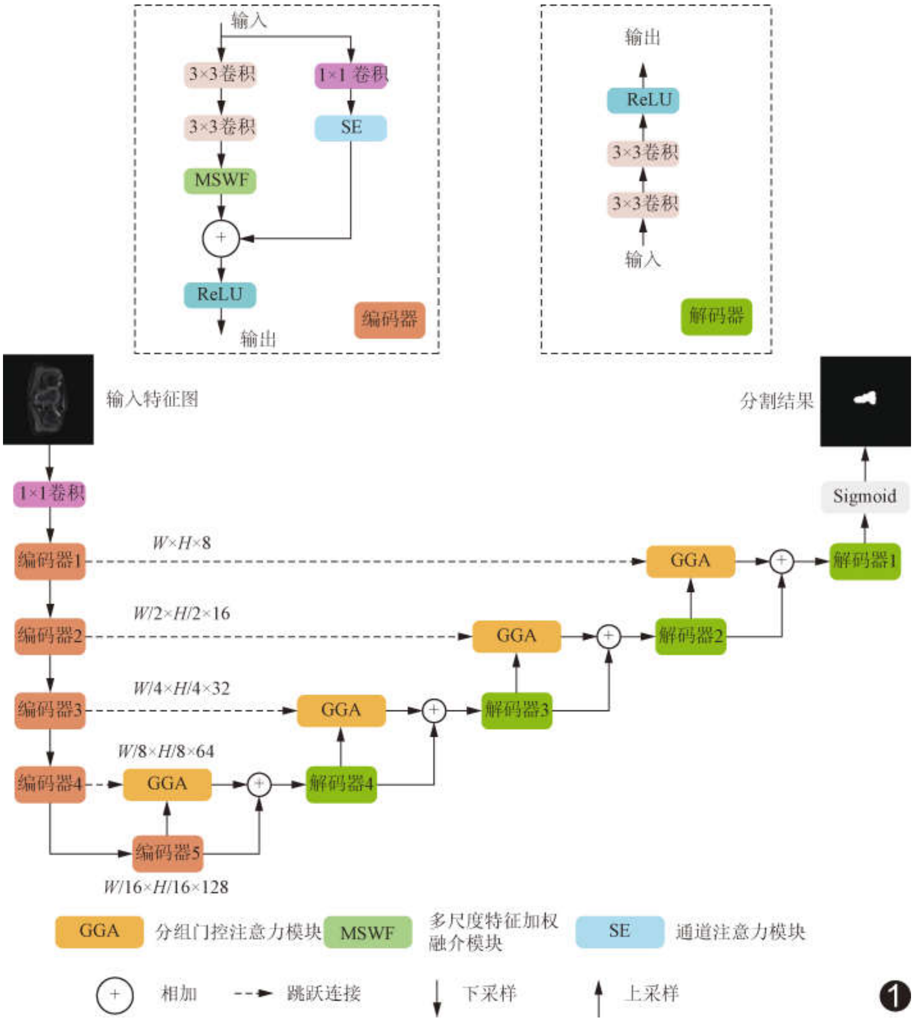
$$Xi = Split(X, i), i = 0,1,2,3 \tag{1}$$

$$F\,i = GroupConv(X\,i, K\,i, G\,i), i = 0,1,2,3 \tag{2}$$

$$Fi' = \sigma(Mc(Fi)) \odot Fi \tag{3}$$

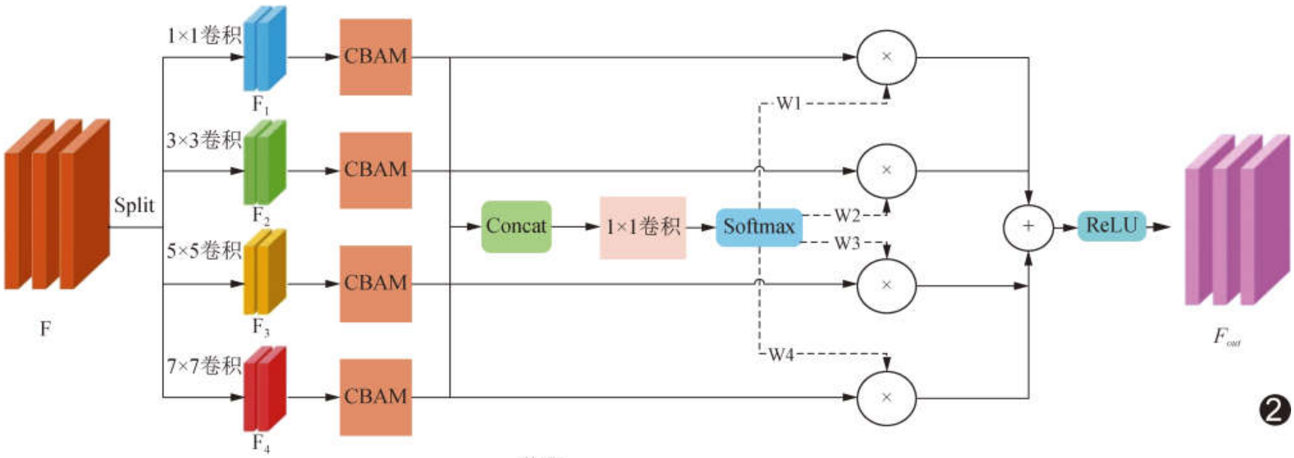
$$Fi'' = \sigma(Ms(Fi')) \odot Fi' \tag{4}$$

其中，Xi 为划分第 i 个通道后的子特征图，$Split$ 为沿通道维度的划分操作，X 为原始特征图的子集，Fi 为经分组卷积得到的第 i 个特征图，$GroupConv$ 表示分组卷积，Ki 为经 CBAM 增强后的第 i 个分支特征，Gi

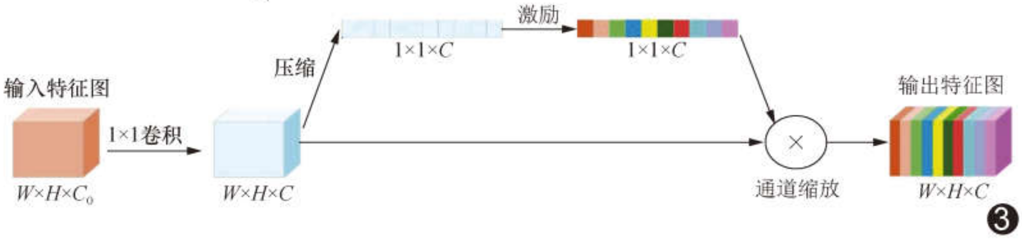


❶

图 1　MGA-UNet 网络结构图（W、H 分别代表输入特征图的宽度和高度）



❷



❸

图 2　MSWF 模块结构图　**图 3**　SE 模块结构图

为第 i 个分支的分组数，Fi' 表示经 SE 生成的特征图，$Mc(\cdot)$ 表示 SE 机制，Fi'' 表示经 CBAM 模块处理的特征图，σ 为 Sigmoid 激活函数，$Ms(\cdot)$ 表示空间注意力机制，$\odot$ 表示逐元素乘法。

将经 CBAM 模块处理后的各分支特征图拼接为高维特征图，并通过 1×1 卷积层减少通道数，将特征图映射到 4 个通道；经 Softmax 归一化生成 4 个分支对应权重，每个分支特征图与其对应权重相乘，将加权后的分支相加获得融合特征图；最后经线性整流函数（rectified linear unit，ReLU）激活函数层进行非线性变换，获得最终输出特征图 $Fout$。见公式(5)、(6)：

$$Fconv = Conv(Concat(F0'', F1'', F2'', F3''), 1 \times 1) \quad (5)$$

$$Fout = ReLU(\sum_{i=0}^{3} Softmax(Fconv)i \odot Fi'') \quad (6)$$

其中，$Fconv$ 是经卷积操作后的输出特征图，$Conv(1 \times 1)$ 表示 1×1 卷积，$Concat$ 表示拼

接操作，Fi'' 表示经 CBAM 模块处理的特征图，$Fout$ 为最终输出特征图，⊙ 表示逐元素乘法。

1.4.2　SE 模块[5]　SE 模块(图 3)是基于 SE 机制的特征增强结构。首先，通过 1×1 卷积将输入特征图($C_0 \times H \times W$)通道维度调整为 C，进行全局压缩与通道激励操作：全局压缩通过全局平均池化将输入特征图($C \times H \times W$)空间维度压缩为 1×1×C 通道描述符，捕获全局上下文信息；通道激励采用两层全连接层构建瓶颈结构，首层将通道维度压缩至 C/r，并经 ReLU 激活函数引入非线性表达能力，第 2 层可恢复至原通道维度，最终通过 Sigmoid 函数生成通道权重向量(范围 0～1)。

1.4.3　GGA 模块　GGA 模块由深度可分离卷积层、批量归一化层、Swish 激活函数层及 Sigmoid 函数层组成(图 4)。

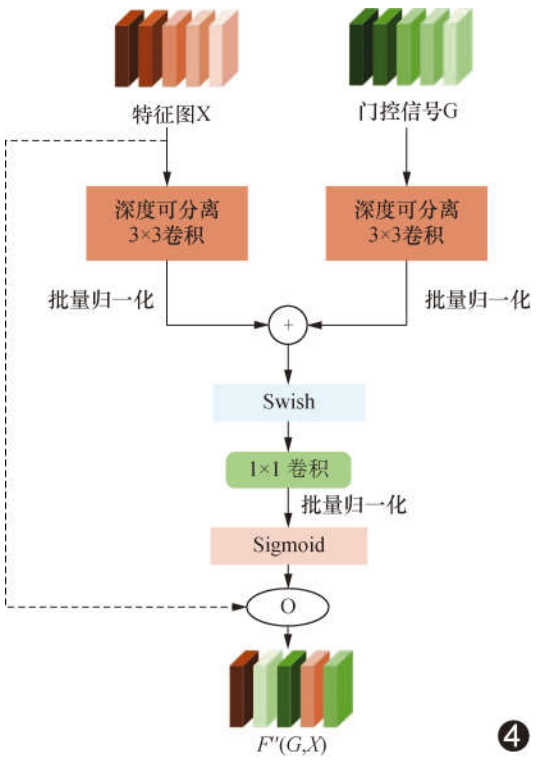


图 4　GGA 模块结构图

首先采用 2 个单独的 3×3 深度可分离卷积处理门控信号 G(来自跳跃连接的特征)和输入的上采样特征图 X，对卷积后特征进行批量归一化，并将批量归一化的门控信号与输入特征相加而实现特征间交互，以增强特征信息融合；之后经 Swish 激活函数引入非线性特性，生成特征图 $F'(G,X)$，通过 1×1 卷积和批量归一化进一步压缩特征信息，生成单通道特征图；最后利用 Sigmoid 函数将特征图转换为注意力系数(范围 0～1)，并与原始输入特征图 X 进行元素级别乘法运算，经注意力门控调节后获得特征图。见公式(7)、(8)：

$$F'(G,X) = \mathrm{Swish}(BN(DSC_g(G) + BN(DSC_x(X)))) \quad (7)$$

$$F''(G,X) = x \odot \sigma(BN(C(F'(G,X)))) \quad (8)$$

式中，$F'(G,X)$ 为经 Swish 函数处理的特征图，BN 为批量归一化，DSC_g 表示对来自跳跃连接的门控信号进行深度可分离卷积，G 表示来自跳跃连接的门控信号，DSC_x 表示对上采样特征图进行深度可分离卷积，X 表示上采样特征图，$F''(G,X)$ 表示经注意力门控调节后的特征图，C 为 1×1 卷积，σ 表示 Sigmoid 函数，⊙ 为元素级别的乘法运算。

1.4.4　损失函数　将分割 CE-T1WI 中的 EC 病灶的任务视为像素级二分类任务，分别将病灶和背景归为两类；以二元交叉熵(binary cross-entropy，BCE)为损失函数评估模型，见公式(9)：

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (9)$$

式中，$L(y,\hat{y})$ 表示损失函数，y_i 表示第 i 个样本的真实标签，$\hat{y}_i$ 表示模型对第 i 个样本的预测概率，N 表示样本总数。

1.5　运行环境　采用 PyTorch 框架，基于 PyCharm 平台开发的 Python 3.8.8，GeForce GTX 4070 GPU；输入图像分辨率配置为 256×256，训练过程中选择动量梯度下降优化器，采用动态学习率衰减策略；初始学习率设置为 0.001，训练 2 批次，训练迭代 100 轮。

1.6　评价指标　以准确率(Accuracy)、召回率(Recall)、交并比(intersection over union，IoU)和戴斯相似系数(Dice similarity coefficient，DSC)评估模型分割任务性能，并与 UNet[6]、UNet++[7]、Attention-UNet[8]、SegNet[9]、Swin-Unet[10] 和 DeepLabV3+[11] 分割结果进行对比。

1.7　消融实验　以原始 UNet 网络为基线模型，在此基础上加入 MSWF 模块、SE 模块及 GGA 模块，分析不同模块对模型性能的影响。

2　结果

采用 MGA-UNet 模型分割 CE-T1WI 中的 EC 病灶的 Accuracy、Recall、IoU 及 DSC 分别为 96.84%、95.73%、92.37% 及 88.96%，整体分割结果优于 UNet、UNet++、Attention-UNet、SegNet、Swin-Unet 和 DeepLabV3+，可解决以往区域边界不清晰和低对比度等问题。见图 5 和表 1。

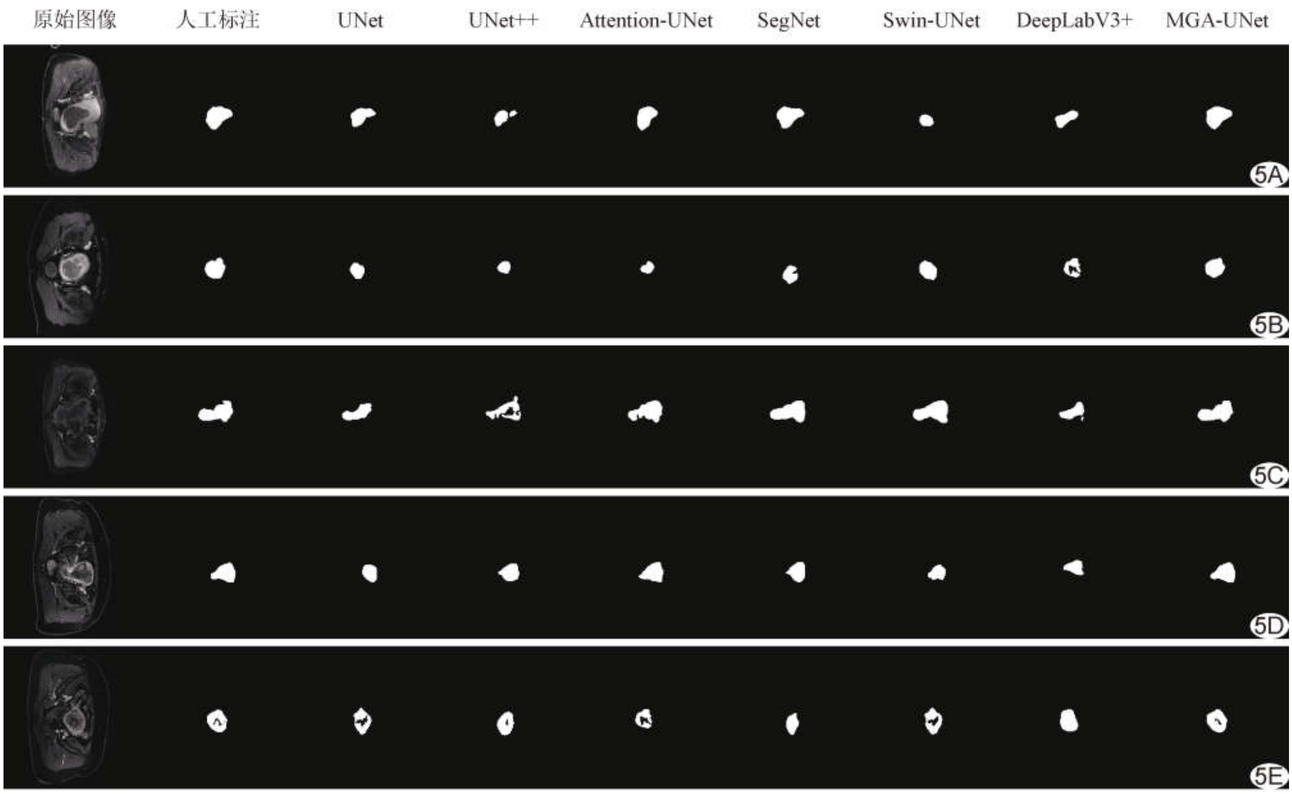


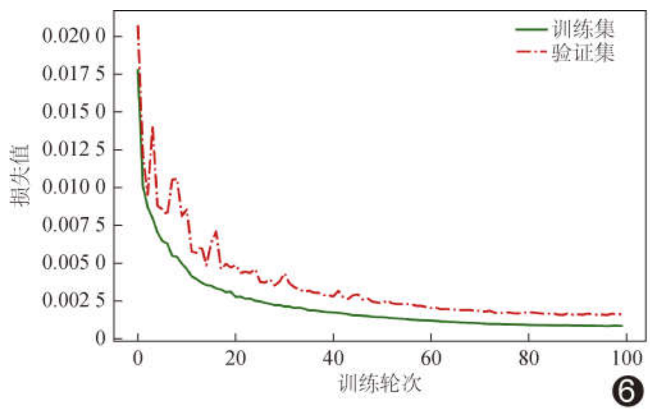
图 5　各模型分割 CE-T1WI 中的 EC 病灶示意图　A.患者女，59 岁，EC 分子分型为 p53abn，病理分期为Ⅲ期；B.患者女，43 岁，EC 分子分型为 MMRd，病理分期为Ⅱ期；C.患者女，62 岁，EC 分子分型为 NSMP，病理分期为Ⅲ期；D.患者女，57 岁，EC 分子分型为 p53abn，病理分期为Ⅱ期；E.患者女，46 岁，EC 分子分型为 NSMP，病理分期为Ⅱ期

表 1　各模型分割 CE-T1WI 中的 EC 病灶的性能

模型	Accuracy(%)	Recall(%)	IoU(%)	DSC
UNet	82.41	82.18	77.36	0.74
UNet＋＋	83.25	81.56	75.88	0.73
Attention-UNet	82.64	82.12	78.95	0.76
SegNet	84.22	83.62	77.64	0.76
Swin-Unet	83.94	84.11	79.23	0.77
DeepLabV3＋	85.91	86.02	79.54	0.80
MGA-UNet	96.84	95.73	92.37	0.89

MGA-UNet 模型的损失值随训练轮次增加而下降，且于训练 80 次后趋于平稳，见图 6；其 DSC 则随训练轮次增加而增加，最大值约 0.89，亦于训练 80 次后趋于平稳，见图 7。

消融实验显示，将 MSWF、SE 及 GGA 模块加入原始 UNet 网络后，其 Accuracy、Recall、IoU 和 DSC 均有所提高，各指标均以三者同时加入后（即 MGA-UNet 模型）最佳。见表 2。



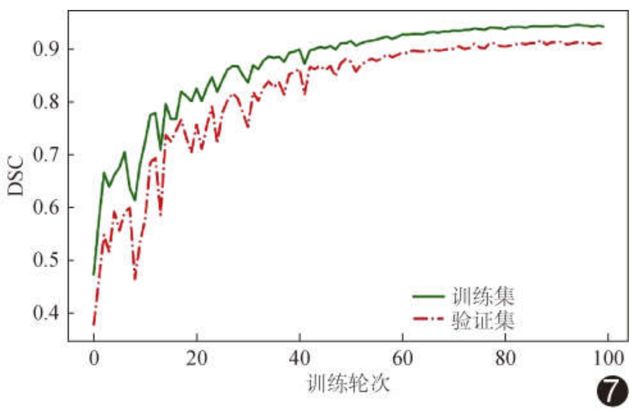


图 6　损失函数曲线图　　**图 7**　DSC 曲线图

表 2　MGA-UNet 模型消融实验

模型	Accuracy (%)	Recall (%)	IoU (%)	DSC
基线模型	82.41	82.18	77.36	0.74
基线模型+SE	84.21	83.72	79.25	0.80
基线模型+MSWF	85.37	87.84	82.51	0.83
基线模型+MSWF+SE	89.56	90.14	86.35	0.83
基线模型+GGA	85.14	86.53	83.43	0.82
基线模型+MSWF+SE+GGA(MGA-UNet)	96.84	95.73	92.37	0.89

3　讨论

高精度智能分割图像中的微小病灶是医学图像分析领域的研究热点,常用方法包括基于编码解码架构的 UNet 模型等[12]。目前利用人工智能精准分割 EC 病灶尚面临一定挑战,主要原因在于子宫内膜厚度个体差异较大,且 EC 呈复杂异质性生长模式,导致传统成像和分割方法难以充分识别并捕捉其特征,且子宫毗邻多个重要器官,一旦肿瘤浸润深肌层或向邻近组织扩散,可致病灶边界模糊不清,使基于 UNet 的端到端分割网络在特征空间建模时面临多尺度特征解耦与弱边界特征强化的双重挑战。

针对上述问题,本研究基于 U-Net 架构提出改进模型 MGA-UNet。MGA-UNet 编码器中,MSWF 模块采用 1×1~7×7 多尺度卷积核并行处理,使其能同时捕获微小病灶的精细纹理特征与周围解剖结构的上下文信息,解决单一尺度卷积对微小目标敏感性不足的问题;SE 模块则通过自适应通道权重分配,重点显示与微小病灶相关的关键特征通道,并有效抑制背景噪声的干扰;该模型在跳跃连接中引入 GGA 模块,通过空间注意力机制精确定位病灶位置,避免在多层特征融合过程中丢失微小目标关键信息;同时采用并行分组卷积提取多维度空间特征,并基于门控单元动态生成空间注意力权重图,有效解决病灶边界不清晰和低对比度问题,以增强分割精度。

本研究结果表明,采用 MGA-UNet 模型分割 CE-T1WI 中的 EC 病灶的 Accuracy、Recall、IoU 及 DSC 分别为 96.84%、95.73%、92.37%及 0.89,优于 UNet、UNet++、Attention-UNet、SegNet、Swin-Unet 和 DeepLabV3+等主流分割模型;消融实验结果进一步证实了 MSWF、SE 和 GGA 模块的价值,单独添加三者均可提升分割性能,联合添加更可达最佳效果。

综上所述,MGA-UNet 模型可用于自动分割 CE-T1WI 中的 EC 病灶。但本研究样本均来自单一中心,未通过跨中心外部验证评估模型泛化能力,有待后续进一步观察。

利益冲突:全体作者声明无利益冲突。

作者贡献:刘辰熙研究设计和实施、数据分析、撰写文章;刘利平统计分析、审阅文章;任玉兰指导、修改文章;余长敏图像分析和处理。

[参考文献]

[1] CROSBIE E J, KITSON S J, McALPINE J N, et al. Endometrial cancer[J]. Lancet, 2022,399(10333):1412-1428.

[2] MAHESHWARI E, NOUGARET S, STEIN E B, et al. Update on MRI in evaluation and treatment of endometrial cancer[J]. Radiographics, 2022,42(7):2112-2130.

[3] SHI R, ZHANG W, ZHANG J, et al. CircESRP1 enhances metastasis and epithelial-mesenchymal transition in endometrial cancer via the miR-874-3p/CPEB4 axis[J]. J Transl Med, 2022, 20(1):139.

[4] KELES D K, EVRIMLER S, MERD N, et al. Endometrial cancer: The role of MRI quantitative assessment in preoperative staging and risk stratification[J]. Acta Radiol, 2022, 63(8): 1126-1133.

[5] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020,42(8):2011-2023.

[6] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241.

[7] ZHOU Z, SIDDIQUEE M M R, TAJBAKHAH N, et al. UNet++: A nested U-Net architecture for medical image segmentation[J]. IEEE Trans Med Imaging, 2018,39:1856-1867.

[8] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[EB/OL].(2018-05-20)[2025-05-01]. https://arxiv.org/pdf/1804.03999.

[9] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017,39(12):2481-2495.

[10] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[EB/OL]. (2021-05-12)[2025-05-01]. https://arxiv.org/pdf/2105.05537.

[11] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision. Munich: Springer, 2018:833-851.

[12] 王翔,史操,袁正一.融合分区注意力 UNet 模型用于分割 MRI 中的膝关节软骨[J].中国医学影像技术,2024,40(5):764-768.