智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

融合CNN和ViT的乳腺超声图像肿瘤分割方法

彭雨彤,梁凤梅

引用本文:
彭雨彤,梁凤梅. 融合CNN和ViT的乳腺超声图像肿瘤分割方法[J]. 智能系统学报, 2024, 19(3): 556-564.
PENG Yutong, LIANG Fengmei. Tumor segmentation method for breast ultrasound images incorporating CNN and ViT[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(3): 556-564.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202304046

DOI: 10.11992/tis.202304046
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20230831.1117.002

融合 CNN 和 ViT 的乳腺超声图像肿瘤分割方法

彭雨彤, 梁凤梅
(太原理工大学 电子信息与光学工程学院, 山西 晋中 030600)

摘　要: 针对乳腺超声图像肿瘤区域形状大小差异大导致分割困难, 卷积神经网络(convolutional neural networks, CNN)建模长距离依赖性和空间相关性方面存在局限性, 视觉 Transformer(vision Transformer, ViT)要求数据量巨大等问题, 提出一种融合 CNN 和 ViT 的分割方法。使用改进的 Swin Transformer 模块和基于可形变卷积的 CNN 编码器模块分别提取全局特征和局部细节特征, 设计使用交叉注意力机制融合这两种尺度的特征表示, 训练过程采取二元交叉熵损失混合边界损失函数, 有效提高分割精度。在两个公共数据集上的实验结果表明, 与现有经典算法相比所提方法的分割结果有显著提升, dice 系数提升 3.841 2%, 验证所提方法的有效性和可行性。

关键词: 卷积神经网络; 乳腺超声图像分割; Swin Transformer; 交叉注意力机制; 混合损失函数; 可形变卷积; 多头跳跃注意力; 深度学习
中图分类号: TP391　　**文献标志码**: A　　**文章编号**: 1673−4785(2024)03−0556−09

中文引用格式: 彭雨彤, 梁凤梅. 融合 CNN 和 ViT 的乳腺超声图像肿瘤分割方法 [J]. 智能系统学报, 2024, 19(3): 556–564.
英文引用格式: PENG Yutong, LIANG Fengmei. Tumor segmentation method for breast ultrasound images incorporating CNN and ViT[J]. CAAI transactions on intelligent systems, 2024, 19(3): 556–564.

Tumor segmentation method for breast ultrasound images incorporating CNN and ViT

PENG Yutong, LIANG Fengmei
(College of Electronic Information and Optical Engineering, Taiyuan University of Technology, Jinzhong 030600, China)

Abstract: A segmentation method that fuses CNN and ViT is proposed to address the problems of large differences in shape and size of tumor regions of breast ultrasound images that lead to difficulty in segmentation, limitations in long-range dependency and spatial correlation in convolutional neural network (CNN) modeling, and the huge amount of data required by vision Transformer (ViT). Global and local detail features were extracted using a modified Swin Transformer module and a CNN encoder module based on deformable convolution, respectively. The design uses a cross-attention mechanism to fuse the feature representations of the two scales, and the training process adopts a binary cross-entropy loss combined with a boundary loss function. This approach effectively improves the segmentation accuracy. Experimental results on two public datasets show that the segmentation findings of the proposed method have been significantly improved compared with those of the existing classical algorithms, with a 3.841 2% improvement in the dice coefficient. This outcome verifies the effectiveness and feasibility of the proposed method.

Keywords: convolutional neural network; breast ultrasound image segmentation; Swin Transformer; crossover attention mechanism; hybrid-loss function; deformable convolution; multihead skip attention; deep learning

根据我国国家癌症中心 2022 年 2 月发布的最新一期全国癌症统计数据显示[1], 乳腺癌是我国女性发病率第一且死亡率第四的癌症, 分别占比 29.05% 和 6.39%。在我国, 乳腺癌的发病率和死亡率呈逐年上升趋势[2]。乳腺超声成像由于其无创、无放射性、低成本的特点[3] 被广泛用于临床检测乳腺肿瘤, 也成为资源匮乏国家和地区进行大规模乳腺癌筛查和诊断的最合适方法。而即使是有经验的放射科医师也难以准确和快速地标记乳腺超声图像的病变区域。随着人工智能和深度学习技术的发展, 乳腺超声图像的计算机辅助诊断(computer aided diagnosis, CAD)技术也应运而生, 而乳腺超声图像 CAD 系统中的一个重要内

收稿日期: 2023−04−24.　**网络出版日期**: 2023−08−31.
基金项目: 山西省重点研发计划项目 (202102030201012).
通信作者: 梁凤梅. E-mail: fm_liang@163.com.

容就是乳腺肿瘤区域的精确分割。

近年来兴起的深度学习分割方法将图像的分割转换为逐像素的分类问题[4], 可以获得更好的分割结果[5]。卷积神经网络(convolutional neural networks, CNN)被广泛应用于医学图像分割的各个方面[6]。U-Net[7] 是比较流行的基于 CNN 的体系结构。Almajalid 等[8] 首次使用 U-Net 从超声图像中分割乳腺病变, 与以往的基于图的和基于模糊 c 均值聚类的分割方法对比有更好的分割结果。与一般图像相比, 由于乳腺超声图像肿瘤区域纹理、形状、大小差异大[9], 而且乳腺超声图像中有许多与乳腺肿瘤病变距离较远但外观相似的正常像素[10], 直接采用 U-Net 进行分割得不到理想的结果。为了提高分割精度, Zhuang 等[11] 提出了残差扩张注意力门 U-Net(residual dilated attention gate UNet, RDAU-Net), 它使用扩张的残差块和注意力门分别替换 U-Net 中的基本块和原始跳跃连接, 提高了肿瘤分割的整体灵敏度和准确性。虽然基于 CNN 的网络结构在乳腺超声图像分割任务中取得了显著成功, 但由于其有限的感受野和固有的归纳偏置[12], 在图像中建立全局上下文和长距离依赖存在限制, 导致医学影像任务的分割效果也受限。

最近, 在自然语言处理领域中表现出色的 Transformer 的激励下, 视觉 Transformer(vision transformer, ViT)[13] 被提出。ViT 可以利用多头注意力机制, 有效地构建长距离依赖关系并捕获全局上下文。最典型的是 Valanarasu 等[14] 提出的医用 Transformer(medical Transformer, MedT)中的门控轴向注意层, 用于构建多头注意力模块。和 CNN 相比, ViT 的缺点是必须在大数据集上进行预训练。而且 ViT 很难处理高分辨率图像, 因为其自注意力的计算复杂度和输入图像大小成平方增长关系。此外, 当 ViT 用于图像处理领域时, 二维图像被切片并作为一维序列送入模型[13], 仅关注全局上下文, 低分辨率特征不能通过直接上采样到全分辨率来有效恢复详细的定位信息, 从而导致分割结果粗糙。

为了提高基于 ViT 的方法对小数据集的适应性, 一些研究人员已经进行了相关尝试。比如提出混合 CNN 和 ViT 的方法对图像进行分割, 结合 CNN 的局部特征表示能力和 ViT 的全局特征表示能力。例如 Chen 等[15] 提出的 TransUnet, 通过跳跃连接来自 CNN 和 ViT 块的局部和全局特征图, 构成类似 U-Net 的模型结构。Zhang 等[16] 提出了 TransFuse, 采取并行分支结构和 BiFusion 模块结合 CNN 和 ViT 的特征信息。但这些方法也存在着不能有效融合信息并保持特征一致性[17]、对于乳腺超声图像肿瘤区域的细节信息学习不够、分割精度不高等问题。

因此, 为了进一步提高乳腺超声图像肿瘤区域的分割效果, 本文提出一种新的乳腺超声图像肿瘤区域分割模型, 采用多尺度的 Swin Transformer 混合 CNN 结构, 引入交叉注意力机制融合粗粒度和细粒度的特征表示。在公共数据集 BUSI[18] 和 Dataset B[19] 上的实验结果验证了本文提出算法的有效性。

1　本文模型结构

为了解决乳腺超声图像分割的难点, 本文提出了融合 CNN 和 Swin Transformer 架构的乳腺肿瘤分割模型, 结构如图 1 所示。首先对乳腺超声图像进行预处理, 再送入 CNN 学习局部特征表示, 接着使用改进的 Swin Transformer 块学习全局上下文表示, 创造性地提出了三层级残差连接架构, 将 CNN 和 Swin Transformer 信息交互, 之后将学习到的高级特征和低级特征使用交叉注意力机制融合, 最后由分割头生成分割掩膜图。

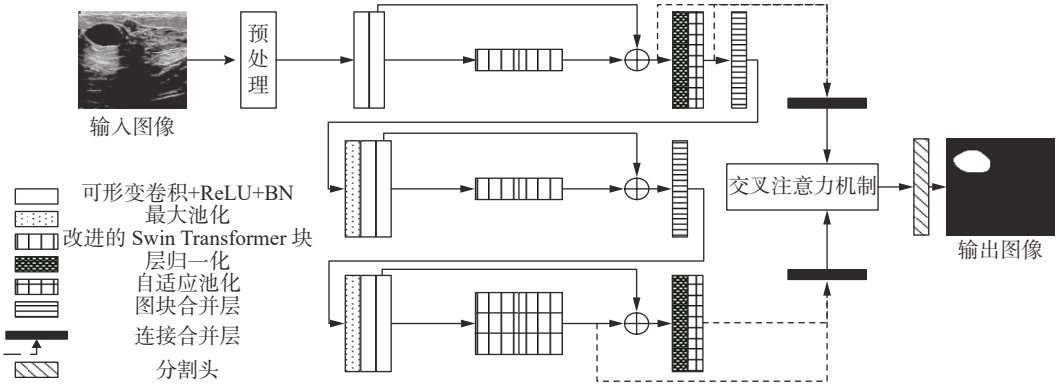

图 1　模型结构

Fig. 1　Overall structure of model

1.1　基于可形变卷积的 CNN 结构

本文采用 U-Net 网络的编码器层代表 CNN 的典型结构，即输入层、卷积池化层、全连接层、输出层。所有卷积操作后都会伴随 ReLU 激活和批归一化（batch normalization, BN），来保障每层网络的输入都是相同的分布。肿瘤区域形状不规则，和背景区域的亮度差异不大，区分边界和背景区域非常困难。而可形变卷积[20]具有较强的空间适应性，在普通卷积的基础上引入了一组偏移量和权重值共同学习，大幅度提高网络在学习细节信息上的表现，因此引入了可形变卷积代替普通卷积作为 CNN 特征提取结构，可形变卷积的原理如图 2 所示。本文所设计的 CNN 特征提取结构具有三层卷积操作，前两次卷积操作后由最大池化将 CNN 和 Swin Transformer 块残差连接的特征图进行下采样。

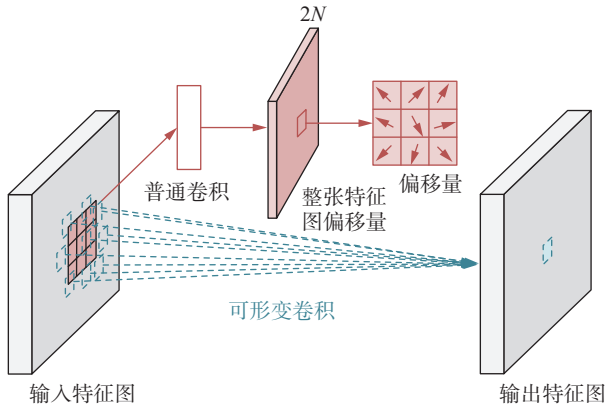

图 2　可形变卷积示意

Fig. 2　Schematic diagram of deformable convolution

1.2　改进的 Swin Transformer 结构

和原始的 ViT 始终对图片全局做自注意力计算不同，Swin Transformer 是对窗口做自注意力计算[21]，大大减少了计算量，把计算量从平方量级变成线性量级，因此引入经典的 Swin Transformer 结构作为典型 ViT 结构。Swin Transformer 采用四阶段金字塔结构，结构如图 3 所示。

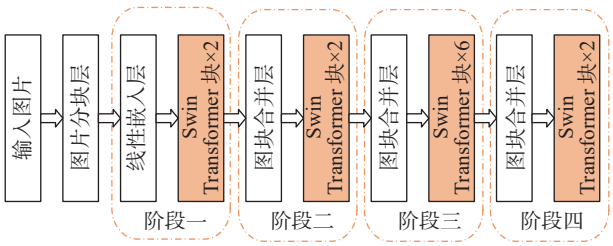

图 3　Swin Transformer 示意

Fig. 3　Diagram of Swin Transformer

Swin Transformer 包括两个连续的 Swin Transformer 块。用基于窗口的多头自注意力（window multi-head self attention, W-MSA）和基于移位窗口的多头自注意力（shifted window multi-head self attention, SW-MSA）来代替原始的 ViT 中的多头自注意力（multi-head self attention, MSA）。在 W-MSA 中，自注意力将应用于大小为 $M\times M$ 的局部窗口。然而，由于没有跨窗口的连接，它的建模能力有限。为了缓解这种情况，引入了 SW-MSA，其利用与 W-MSA 模块的输入相比移位的窗口配置，这是为了确保具有跨窗口连接。该过程用如下公式表示：

$$A(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}\left(\frac{\boldsymbol{Q}(\boldsymbol{K})^{\mathrm{T}}}{\sqrt{d}}+B\right)\boldsymbol{V} \tag{1}$$

其中：$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为每个像素点的查询（query）、键（key）、值（value）进行注意力 A 的计算，B 是 Swin Transformer 中的相对位置偏差。

为了更好地利用来自不同阶段的特征映射，将 Swin Transformer 中的注意力机制从 MSA 改进为多头跳跃注意力（multi-head skip self attention, MSKA）。MSA 只是计算与自身的相似度，很难有效地激活一个注意头中的单个类别。而利用 MSKA 计算语义相同的两个不同特征映射的相似度，更好地利用了注意力机制，能够激活一个注意头中的单个类别。使用前一阶段的输出 M 作为键、值，使用特征映射 F 作为查询。基于窗口的多头跳跃注意力（window multi-head skip self attention, W-MSKA）设置为

$$A(\boldsymbol{Q}_{\mathrm{F}},\boldsymbol{K}_{\mathrm{M}},\boldsymbol{V}_{\mathrm{M}})=\mathrm{softmax}\left(\frac{\boldsymbol{Q}_{\mathrm{F}}(\boldsymbol{K}_{\mathrm{M}})^{\mathrm{T}}}{\sqrt{d}}+B\right)\boldsymbol{V}_{\mathrm{M}} \tag{2}$$

其中：$\boldsymbol{Q}_{\mathrm{F}}$ 是跳跃注意力的查询，是F的线性变换；$\boldsymbol{K}_{\mathrm{M}}$ 和 $\boldsymbol{V}_{\mathrm{M}}$ 是键和值，都是M的线性变换。改进后的 Swin Transformer 块的结构如图 4 所示。

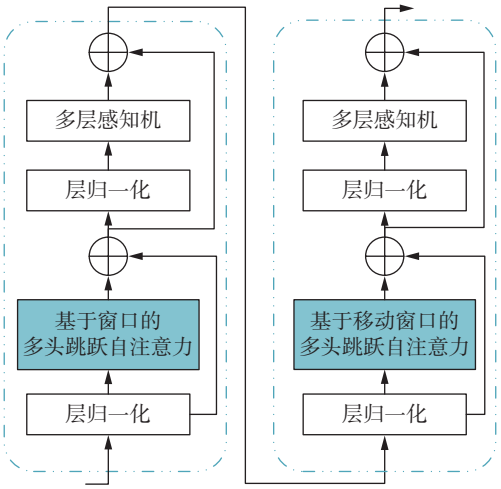

图 4　改进的 Swin Transformer 块示意

Fig. 4　Diagram of improved Swin Transformer block

1.3　基于交叉注意力机制的特征融合模块

为了有效融合改进 Swin Transformer 编码器

传递的嵌入信息, 引入交叉注意力机制[22] 融合每个级别的特征。具体地说, 在融合之前, 两个级别的类标记被交换, 这意味着一个级别的类标记与另一个级别的标记连接。然后, 每个新的嵌入被单独地馈送通过用于融合的模块, 并且最终被反向投影到其自身的级别。这种与其他级别标记的交互使类标记能够与它们的跨级别标记共享丰富的信息。

根据这种思想提出的特征融合模块如图 5 所示, 用P^l和P^s分别表示来自 CNN 和来自 Swin Transformer 块的不同分辨率的特征图。由 Transformer 编码器输出的最高分辨率视觉标记和高分辨率关键点标记, 和来自 CNN 的低分辨率标记重新组合生成新的视觉标记和关键点标记, 作为多尺度交叉注意力模块的输入。为了避免关键点标记的冗余和多次融合关键点标记, 模块采用了移动关键点标记策略。该策略的具体做法: 在输入到该模块前, 高、低分辨率视觉标记分别拼接固定的高、低分辨率关键点标记, 将包含关键点和视觉特征的高分辨率标记作为键和值, 低分辨率标记作为查询, 然后进行交叉注意力计算。当进行下一次交互时, 该关键点标记和其他的高、低分辨率视觉标记重新拼接。

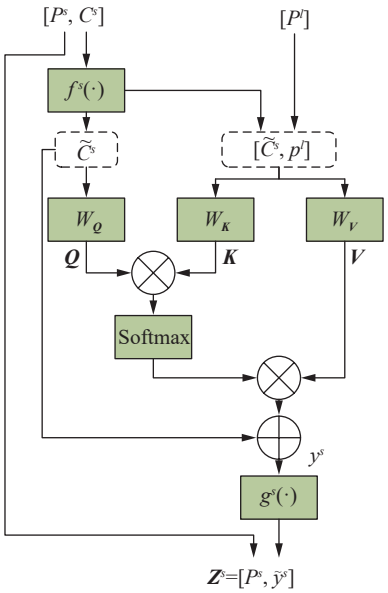

图 5　交叉注意力机制特征融合示意

Fig. 5　Diagram of cross attention for feature fusion

1.4　混合损失函数

根据在预测标签和真实标签之间每个像素存在的偏差, 损失函数用来在训练期间矫正网络参数。医学成像中常用的分割损失函数是交叉熵和 dice 得分。由于乳腺超声图像中肿瘤的像素所占比例很小, 肿瘤分割存在严重的数据不平衡问题。因此, 该模型将主要学习具有大量像素的非肿瘤样本的特征。二元交叉熵损失函数被广泛用于解决上述问题。交叉熵测量图像中每个像素的预测概率和校正概率的对数值。对于图像二值分割的问题, 二元交叉熵在像素级分类上表现良好。二元交叉熵损失函数的数学表达式为

$$L_{\mathrm{BCE}}=-\frac{1}{n}\sum_{i=1}^{n}[y_i\log p(y_i)+(1-y_i)\log(1-p(y_i))] \tag{3}$$

式中y_i是分割图中每个像素值的预测, 取值为 0 或 1。在医学图像中, 边界是不同组织之间区别较为明显的地方, 在人工分割时是区分不同目标的重要参考。然而在卷积神经网络中, 边界作为不同类别相交位置, 与网络分割区域的特征如形态、强度相似性等不同, 属于高频信息, 如果不加以一定约束, 特征在不断卷积的过程中不断被抽象, 容易丢失这些细节信息导致网络分割错误。

为了提升网络分割精度, 有必要对网络施加额外的边界约束以提高网络对边界信息的利用提升网络的分割性能。同时, 依靠设置交叉熵损失权重来缓解数据分布不平衡, 通过对目标施加额外的监督损失, 边界损失[23] 也能更进一步解决分割精度不高的问题。边界损失能在不增加网络复杂性的同时充分发掘区域和边界互补信息辅助区域分割。二元边界损失函数的数学表达式为

$$P=\frac{1}{|B_p|}\sum_{x\in B_p}[\![d(x,B_g)<\theta]\!] \tag{4}$$

$$R=\frac{1}{|B_g|}\sum_{x\in B_g}[\![d(x,B_p)<\theta]\!] \tag{5}$$

$$L_{\mathrm{Boundary}}=1-\frac{2P\cdot R}{P+R} \tag{6}$$

式中: B_g和B_p分别为真实标签和分割预测结果的边界, $d(\cdot)$是以像素为单位测量的欧几里得距离, θ是预定义的阈值。采用混合损失函数的方法会获得更好的分割结果, 因此本文在实验训练过程中使用混合交叉熵损失和边界损失计算预测的分割结果损失, 损失函数的定义为

$$L=\alpha L_{\mathrm{BCE}}+(1-\alpha)L_{\mathrm{Boundary}} \tag{7}$$

其中α为超参数, 需要进行实验确定适合值。

2　实验仿真及结果分析

实验所使用的计算机系统环境为 Ubuntu20.04, 采用 Python 编程语言, Pycharm 作为集成开发环境(IDE), Pytorch1.11 作为深度学习框架, 在 10 GB 显存的 RTX3080 GPU 上进行训练。输入图像大小为 224 像素×224 像素, 训练使用 AdamW 优化器, 初始学习率为 0.000 1, 权重衰减为 0.000 05,

数据批处理大小设置为 4，最大训练 80 000 轮次。

2.1　数据集和数据预处理

使用公开的乳腺超声图像数据集 BUSI[18] 和 Dataset B[19] 进行对比实验，来验证本文提出算法的可行性和有效性。每个数据集都包含原始乳腺超声图像和由专业医师标注好的肿瘤区域。其中，BUSI 数据集包含 133 张正常超声图像、437 张良性肿瘤图像和 210 张恶性肿瘤图像。Dataset B 数据集包含 110 张良性肿瘤图像和 53 张恶性肿瘤图像。为了更直观地评估本文方法在恶性肿瘤和良性肿瘤上的分割性能，实验只在 BUSI 和 Dataset B 中的良性肿瘤和恶性肿瘤图像上进行训练和测试，不使用正常超声图像进行训练。

考虑到数据集过小的问题，本文使用数据增广方法来扩充数据，包括平移、随机水平翻转、随机裁剪等。将 BUSI 扩张为 3 235 张图像，将 Dataset B 扩张为 815 张图像，按照 4∶1 随机划分为训练集、测试集。实验使用双线性插值技术将数据集中的所有图像大小统一为 224 像素×224 像素，并进行 3×3 中值滤波器处理。同时，使用最近邻插值技术将所有样本对应的真实标签（Ground Truth）大小也统一为 224×224。

2.2　评价指标

在每个实验中，为了定量分析分割效果，本文采用了常用的 5 个评价指标I，包括 dice 系数、Jc 指数、95%Hausdorff 距离（Hd95）、准确度和召回率。

dice 系数可以计算分割结果和真实标签之间的相似度，取值范围是 [0,1]，取值越大证明分割结果越好。计算公式为

$$I_{\mathrm{dice}}=\frac{2N_{\mathrm{TP}}}{2N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}} \tag{8}$$

其中：N_{TP}表示图像中正常且被模型预测正常的像素点数量，N_{FP}表示图像中肿瘤但被模型预测正常的像素点数量，N_{FN}表示图像中肿瘤且被模型预测肿瘤的像素点数量。

杰卡德指数（Jc 指数）也可以计算分割结果和真实标签的相似度，作为辅助评价指标使用，取值越大证明分割结果越好。

$$I_{\mathrm{Jc}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}} \tag{9}$$

Hausdorff 距离可以计算分割结果和真实标签边界的最大距离差异，主要用于衡量分割边界的准确性，取值越小证明分割结果越好，计算公式为

$$I_{\mathrm{Hd}}(A,B)=\max(h(A,B),h(B,A)) \tag{10}$$

$$h(A,B)=\max_{a\in A}\{\min_{b\in B}\|a-b\|\} \tag{11}$$

$$h(B,A)=\max_{b\in B}\{\min_{a\in A}\|b-a\|\} \tag{12}$$

其中：h是计算的 Hausdorff 距离，I_{Hd}是最终的 Hausdorff 距离结果，A是分割结果图，B是真实标签图。但在实际计算中往往不是取最大距离，而是取前 5% 距离，可以排除一些离群点距离对分割结果的影响。

准确率（precision）可以计算分割结果中真正的肿瘤区域的比例，取值越大证明分割结果越好。

$$I_{\mathrm{precision}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}} \tag{13}$$

召回率（recall）可以计算所有肿瘤区域被预测正确的比例，取值越大证明分割结果越好。

$$I_{\mathrm{recall}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}} \tag{14}$$

2.3　超参数设置

对本文提出的混合损失函数中的超参数α采用数据量较大的 BUSI 进行实验，对α取值 0.1~0.9，得出 0.4、0.5 是评价指标最大的两个值，最后取 0.45 进行实验验证，得出各评价指标的峰值。实验结果如图 6 所示。

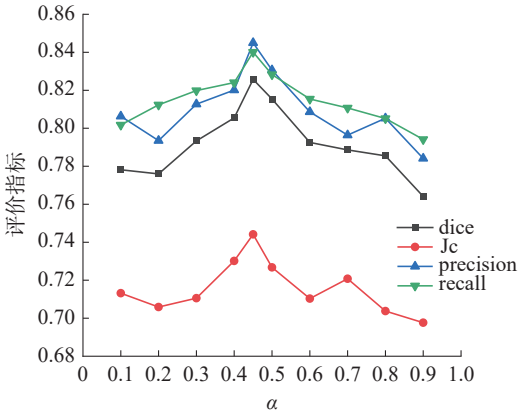

图 6　混合损失函数超参数设置结果

Fig. 6　Diagram of the results of the hyperparameter setting of the hybrid loss function

2.4　对比实验

为了验证所提出的算法的有效性，本文选择以下 6 个在医学图像分割方面具有代表性的网络结构与本文提出的算法进行对比。其中 U-Net 是经典 CNN 结构，Swin Transformer 是纯 ViT 结构，其余则都是混合 CNN-ViT 结构。

1) U-Net[4]：通过跳跃连接收缩路径和扩展路径，可以有效地融合低分辨率和高分辨率特征。

2) Swin Transformer[21]：利用 ViT 提取全局特征，通过基于窗口的多头注意力计算，可以实现跨窗口信息交互和线性计算复杂度。

3) TransUnet[15]：通过跳跃连接来自 CNN 和 ViT 的局部和全局特征图，构成类似 U-Net 的模

型结构。

4) TransDeeplab[24]: 利用带有移位窗口的 Swin Transformer 块扩展 deeplabv3+, 并对空洞空间金字塔池化模块建模。

5) SwinUnet[25]: 对照 Swin Transformer 块的图块拼接层设计了一个图块扩展层, 组成类似 U-Net 型的编码器–解码器结构。

6) TransNorm[26]: 将 ViT 整合到 U-Net 编码器和跳跃连接中, 利用两级注意力机制自适应地重校准跳跃连接路径。

在两个数据集 Dataset BUSI 和 Dataset B 中分别进行分割实验。由于考虑到不同硬件环境下的计算时间不同, 为了提高泛化性, 所以改用对比各个方法的参数量来进行间接说明计算时间。实验中采用的算法参数量如表 1 所示, 实验中测试集平均定量结果如表 2 和表 3 所示, 表中加粗字体为每列的最优值。

表 1　不同算法的参数量对比

Table 1　Comparison of parameter quantities of different algorithms

算法	参数量/10^6
U-Net	28.05
Swin Transformer	29.43
TransUnet	105.28
TransDeeplab	21.14
SwinUnet	27.17
TransNorm	117.63
本文算法	25.51

表 2　不同算法在 Dataset BUSI 上的分割结果

Table 2　Segmentation results of different algorithms on Dataset BUSI

算法	dice系数	Jc指数	Hd95	准确度	召回率
U-Net	0.787 320	0.703 703	56.255 785	0.815 321	0.807 724
Swin Transformer	0.781 523	0.702 417	63.498 802	0.795 067	0.813 038
TransUnet	0.797 532	0.714 024	53.049 520	0.828 517	0.815 834
TransDeeplab	0.803 847	0.717 159	55.842 163	0.813 009	0.838 311
SwinUnet	0.804 722	0.720 576	59.828 857	0.817 851	0.836 466
TransNorm	0.799 316	0.719 628	51.095 965	0.827 717	0.814 157
本文算法	**0.825 732**	**0.744 134**	**33.203 520**	**0.844 956**	**0.840 086**

表 3　不同算法在 Dataset B 上的分割结果

Table 3　Segmentation results of different algorithms on Dataset B

算法	dice系数	Jc指数	Hd95	准确度	召回率
U-Net	0.755 240	0.705 914	47.186 216	0.806 826	0.811 233
Swin Transformer	0.750 100	0.674 348	51.348 517	0.777 199	0.774 130
TransUnet	0.807 067	0.743 791	24.456 953	0.801 030	0.819 393
TransDeeplab	0.811 323	0.716 784	24.501 722	0.799 767	0.853 850
SwinUnet	0.806 981	0.700 275	30.370 696	0.783 342	0.818 952
TransNorm	0.781 180	0.700 323	24.035 719	0.799 767	0.853 850
本文算法	**0.825 857**	**0.740 428**	**23.125 318**	**0.827 233**	**0.859 421**

经过参数量对比, 可以看出本文所提出的算法参数量少于除了 TransDeeplab 的其他经典算法。由于本文主要解决的问题是乳腺超声图像的分割精度问题, 所以本文算法的运算量虽不是最优但在可接受的范围。

对比结果表明本文算法使用交叉注意力机制融合改进的 CNN 和 ViT, 对乳腺超声图像的分割结果都优于其他算法, 各个指标相较于其他算法都有明显提升, 和经典的 U-Net 算法相比 dice 系数提升了 3.841 2%, 这表明本文算法的有效性。在两个数据集上进行的实验结果可以证明本文提出算法的泛化性能, 相较于其他算法更具备实际应用价值。实验中部分测试集的分割结果如图 7 所示, 也可以看出采用混合边界损失对肿瘤边界的分割更准确。

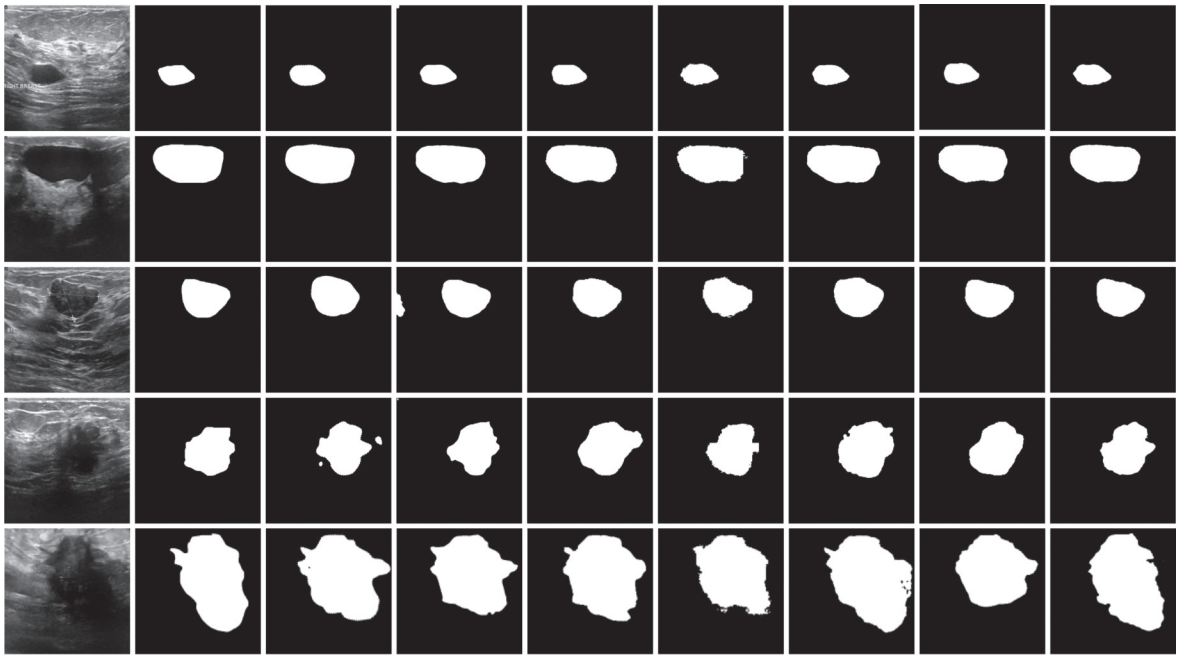

图 7　不同算法分割结果对比

Fig. 7　Comparison diagram of different algorithm segmentation results

2.5　消融实验

为了进一步说明本算法的有效性，使用数据量较大的 Dataset BUSI 进行 5 组消融实验，实验结果如表 4 所示。可以看出，单独使用 U-Net 的分割 dice 结果仅为 0.787 320，接续添加可形变卷积 (deformable convolution, DC)、混合损失函数 (hybrid loss, HL)、交叉注意力机制 (cross attention, CA)、多头跳跃注意力 (multi-head skip self attention, MSKA)4 个模块后分割结果各方面都有所提升，dice 系数、Jc 指数、准确度和召回率分别提升了 3.841 2%、4.043 1%、2.963 5% 和 3.236 2%，其中 DC 模块的提升效果最明显，表明了 DC、HL、CA、MSKA 等 4 个模块对乳腺超声图像肿瘤分割的可行性。

表 4　在 Dataset BUSI 上的消融实验分割结果

Table 4　Ablation experimental segmentation results on Dataset BUSI

算法	dice系数	Jc指数	Hd95	准确度	召回率
U-Net	0.787 320	0.703 703	56.255 785	0.815 321	0.807 724
U-Net+DC	0.803 680	0.722 510	37.643 451	0.829 810	0.816 059
U-Net+DC+HL	0.809 869	0.726 249	38.527 125	0.840 414	0.811 717
U-Net+DC+HL+CA	0.819 256	0.738 401	34.681 035	0.834 446	0.808 971
U-Net+DC+HL+CA+MSKA(本文算法)	**0.825 732**	**0.744 134**	**33.203 520**	**0.844 956**	**0.840 086**

3　结束语

本文提出一种融合 CNN 和 ViT 的乳腺超声图像肿瘤分割方法，该方法采用可形变卷积提升了 CNN 提取特征的能力。训练过程中采用二元交叉熵损失混合边界损失函数来优化提出的算法模型，对乳腺超声图像肿瘤区域的边界分割性能有所提升。对 Swin Transformer 的改进主要是将注意力机制从 MSA 改进为 MSKA，从计算和自身的注意力转变为计算两个相邻特征图的注意力，更好地利用了注意力机制。最后使用交叉注意力机制建立来自 CNN 的局部特征和来自 ViT 的全局特征的有效融合。通过对比实验和消融实验验证了所提出方法的有效性，并且在各个指标中都优于现有的经典算法，在智能医疗辅助诊断中具有发展前景。本文存在的不足在于对于恶性肿瘤的分割效果不如良性肿瘤，而且若没有 Ground Truth 情况下，只能依靠专业医师进行主观评价。以后的研究工作可以考虑对肿瘤进行分类后再分割，无监督分割，或者进行多任务处理等。

参考文献:

[1] ZHENG Rongshou, ZHANG Siwei, ZENG Hongmei, et al. Cancer incidence and mortality in China, 2016[J]. Journal of the national cancer center, 2022, 2(1): 1–9.

[2] 高艳多, 阎炯, 赵胜, 等. 1990—2019 年中国女性乳腺癌发病和死亡趋势的年龄−时期−队列模型分析 [J]. 中国预防医学杂志, 2022, 23(12): 909–916.
GAO Yanduo, YAN Jiong, ZHAO Sheng, et al. Trends in incidence and mortality of female breast cancer in China from 1990 to 2019 using age-period-cohort analysis model[J]. Chinese preventive medicine, 2022, 23(12): 909–916.

[3] XIAN Min, ZHANG Yingtao, CHENG H D, et al. Automatic breast ultrasound image segmentation: a survey [EB/OL]. (2017–04–04)[2023–04–24]. http://arxiv.org/abs/1704.01472.

[4] 苏丽, 孙雨鑫, 苑守正. 基于深度学习的实例分割研究综述 [J]. 智能系统学报, 2022, 17(1): 16–31.
SU Li, SUN Yuxin, YUAN Shouzheng. A survey of instance segmentation research based on deep learning[J]. CAAI transactions on intelligent systems, 2022, 17(1): 16–31.

[5] 施俊, 汪琳琳, 王珊珊, 等. 深度学习在医学影像中的应用综述 [J]. 中国图象图形学报, 2020, 25(10): 1953–1981.
SHI Jun, WANG Linlin, WANG Shanshan, et al. Applications of deep learning in medical imaging: a survey[J]. Journal of image and graphics, 2020, 25(10): 1953–1981.

[6] 张宇, 梁凤梅, 刘建霞. 融合类激活映射和视野注意力的皮肤病变分割 [J/OL]. 计算机工程与应用, 2022: 1–10. (2022–10–12) [2023–04–24]. https://kns.cnki.net/kcms/detail/11.2127.tp.20221011.1633.008.html.
ZHANG Yu, LIANG Fengmei, LIU Jianxia. Skin lesion segmentation based on classification activation mapping and visual field attention[J/OL]. Computer engineering and applications, 2022: 1–10. (2022–10–12) [2023–04–24]. https://kns.cnki.net/kcms/detail/11.2127.tp.20221011.1633.008.html.

[7] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234–241.

[8] ALMAJALID R, SHAN Juan, DU Yaodong, et al. Development of a deep-learning-based method for breast ultrasound image segmentation[C]//2018 17th IEEE International Conference on Machine Learning and Applications. Orlando: IEEE, 2018: 1103–1108.

[9] CHEN Gongping, DAI Yu, ZHANG Jianxun. RRCNet: refinement residual convolutional network for breast ultrasound images segmentation[J]. Engineering applications of artificial intelligence, 2023, 117: 105601.

[10] HE Qiqi, YANG Qiuju, XIE Minghao. HCTNet: a hybrid CNN-transformer network for breast ultrasound image segmentation[J]. Computers in biology and medicine, 2023, 155: 106629.

[11] ZHUANG Zhemin, LI Nan, JOSEPH RAJ A N, et al. An RDAU-NET model for lesion segmentation in breast ultrasound images[J]. PLoS One, 2019, 14(8): e0221535.

[12] D'ASCOLI S, TOUVRON H, LEAVITT M L, et al. ConViT: improving vision transformers with soft convolutional inductive biases[J]. Journal of statistical mechanics: theory and experiment, 2022, 2022(11): 114005.

[13] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020–10–22) [2023–04–24]. http://arxiv.org/abs/2010.11929.

[14] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: gated axial-attention for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 36–46.

[15] CHEN Jieneng, LU Yongyi, YU Qihang, et al. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021–02–08)[2023–04–24]. http://arxiv.org/abs/2102.04306.

[16] ZHANG Yundong, LIU Huiye, HU Qiang. TransFuse: fusing transformers and CNNs for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference. Strasbourg: ACM, 2021: 14–24.

[17] HEIDARI M, KAZEROUNI A, SOLTANY M, et al. HiFormer: hierarchical multi-scale representations using transformers for medical image segmentation[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2023: 6191–6201.

[18] AL-DHABYANI W, GOMAA M, KHALED H, et al. Dataset of breast ultrasound images[J]. Data in brief, 2020, 28: 104863.

[19] YAP M H, GOYAL M, OSMAN F, et al. Breast ultrasound region of interest detection and lesion localisation[J]. Artificial intelligence in medicine, 2020, 107: 101880.

[20] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 764–773.

[21] LIU Ze, LIN Yutong, CAO Yue, et al. Swin Transformer: hierarchical Vision Transformer using Shifted Windows[C]// 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 9992–10002.

[22] CHEN C F R, FAN Quanfu, PANDA R. CrossViT: cross-attention multi-scale vision transformer for image classification[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 347–356.

[23] BOKHOVKIN A, BURNAEV E. Boundary loss for remote sensing imagery semantic segmentation[C]//International Symposium on Neural Networks. Cham: Springer, 2019: 388–401.

[24] AZAD R, HEIDARI M, SHARIATNIA M, et al. Trans-DeepLab: convolution-free transformer-based DeepLab v3+ for medical image segmentation[C]//International Workshop on PRedictive Intelligence In MEdicine. Cham: Springer, 2022: 91–102.

[25] CAO Hu, WANG Yueyue, CHEN J, et al. Swin-unet: unet-like pure transformer for medical image segmentation[C]// Computer Vision – ECCV 2022 Workshops. Tel Aviv: ACM, 2022: 205–218.

[26] AZAD R, AL-ANTARY M T, HEIDARI M, et al. Trans-Norm: transformer provides a strong spatial normalization mechanism for a deep segmentation model[J]. IEEE access, 2022, 10: 108205–108215.

作者简介：

彭雨彤，硕士研究生，主要研究方向为医学图像处理。E-mail：pyt34567@163.com。

梁凤梅，副教授，博士，主要研究方向为图像处理与传输、智能信息处理。主持完成省自然科学基金 1 项、省科技成果推广项目 1 项、省技术创新项目 1 项。获得山西省科技进步二等奖 1 项（第一完成人）、山西省科技进步三等奖 2 项。发表学术论文 50 余篇。E-mail：fm_liang@163.com。

2024 IEEE“一带一路”人工智能可持续发展大会
2024 IEEE Belt and Road Congress on AI for Sustainable Development

2024 年是“一带一路”倡议“下一个金色十年”的崭新起点，也是高质量共建“一带一路”八项行动的重要节点。推动科技创新作为八项行动的关键内容，为通用人工智能时代下世界各国现代化发展带来无限可能，为一带一路可持续发展目标实现注入强劲动能。为此，中国人工智能学会联合 IEEE China Council 发起主办 2024 IEEE“一带一路”人工智能可持续发展大会，携手打造国际化交流平台，构建全球性共创模式，为一带一路高质量、可持续发展凝共识、集众智、聚合力、谋未来。

大会将于 2024 年 6 月 23—24 日在中国杭州举办，邀请国内外“一带一路”相关国家政府机构代表、科研院所及企业负责人、学术研究机构、社会组织各界嘉宾共同参加，围绕数据智能赋能带路可持续发展框架建立合作网络，推动“一带一路”通用人工智能可持续发展。详情可前往官网 https://www.caai.cn/index.php?s=/home/article/detail/id/3776.html。

联系方式

刘老师 010-82686687 aicon@caai.cn

邹老师 010-82686683 aicon@caai.cn