

融合多维注意力机制 CNN 皮肤肿瘤 图像分割提取

高正君¹, 张佩炯², 司小强¹

(1. 甘肃省人民医院整形美容外科, 兰州 730030;

2. 兰州资源环境职业技术大学 电力工程学院, 兰州 730022)

摘要: 针对卷积神经网络 (CNN) 在医学图像分割时, 受皮肤病损图像多样性、分割目标位置、形状及尺度变化等因素影响, 提出了一种基于传统卷积神经网络综合注意力模块图像分割算法; 首先利用 U-Net 主干网络的优势, 其目的让图像特征提取更完善; 其次, 由空间、通道、尺度构成的综合注意力机制对目标病灶区域进行检测识别, 利用通道级联把来自编码器中低级图像特征和解码器中高级图像特征注意力结合起来进行权值自适应融合, 提升了网络对样本病灶区的关注度和辨识力, 突出强调最相关的特征通道和多尺度间最显著的特征图; 通过对 ISIC2018 数据集及医院整形外科提供患者不同类型的皮肤肿瘤图像进行分割测试, 并将注意力模块随机组合形成的不同算法进行指标评价对比, 所提出算法的平均分割精度可达 92.89%; 实验结果表明, 所提出算法是有效可行的, 在多维度下分割处理带复杂背景的皮肤病灶图像时有更高的鲁棒性。

关键词: 皮肤肿瘤; 图像分割; 注意力模块; 池化; 卷积神经网络

Combined Multidimensional Attention Mechanism Convolutional Neural Network in Skin Tumor Image Segmentation

GAO Zhengjun¹, ZHANG Peijiong², SI Xiaoqiang¹

(1. Department of Plastic Surgery, Gansu Provincial Hospital, Lanzhou 730030, China;

2. School of Electrical Engineering, Lanzhou Resource & Environment Voc-Tech University, Lanzhou 730022, China)

Abstract: In medical image segmentation, convolutional neural network (CNN) is affected by the diversities of skin lesions images, the location, shape and scale changes of segmentation targets, and other factors. A multi-dimensional attention module based on the space, channel and scale is proposed to optimize the CNN image segmentation algorithm. Firstly, the advantage of U-NET backbone network is used to make the image feature extraction more perfect. Secondly, the comprehensive attention mechanism is composed of multidimensional space, channels and scale, which detects and identifies the target lesion area, by using the lower-level channel cascade from the encoder image features and the decoder senior image, the adaptive weighting fusion of attention is unified, and enhances the awareness of the network on the sample lesions and discrimination, and highlights the most relevant characteristics of the channel, and emphasizes the most salient feature maps between multiple scales. The segmentation test was carried out on the ISIC2018 data set, the different types images of skin tumors are provided by hospital plastic surgery department, and the index evaluation and comparison of different algorithms formed by random combination of attention modules show that, the average segmentation accuracy of the proposed algorithm reaches 92.89%. Experimental results show that the proposed algorithm is effective and feasible, which has higher robustness in the segmentation of cutaneous lesions under the complex background.

Keywords: skin tumor; image segmentation; attention module; pooling; convolutional neural network

0 引言

据全球调查数据统计, 皮肤肿瘤及皮肤癌的发病率以 3%~8% 的比例增长, 且死亡率也在逐年上升, 皮肤肿瘤及皮肤癌疾病正严重危害着人类的健康^[1]。皮肤肿瘤最有

效的解决办法就是能早期精准诊断后进行病灶的根除, 随着计算机人工智能技术的不断发展, 医学图像的精确分割对于病理临床诊断评估、治疗方案的制定及疾病状况的监测有着举足轻重的作用^[2]。通常, 医学图像的成像类型多种多样, 皮肤镜图像可通过无创性显微成像技术来获取,

收稿日期: 2022-02-06; 修回日期: 2022-03-07。

基金项目: 国家自然科学基金(61862039, 61462059); 甘肃省人民医院内科研基金重点学科项目(20GSSY1-3)。

作者简介: 高正君(1984-), 女, 甘肃兰州人, 硕士, 主治医师, 主要从事皮肤肿瘤及瘢痕机制方向的研究。

司小强(1975-), 男, 甘肃静宁人, 博士, 主任医师, 主要从事皮肤肿瘤及创面愈合等方向的研究。

通讯作者: 张佩炯(1984-), 男, 甘肃武山人, 硕士, 工程师, 主要从事人工智能优化算法方向的研究。

引用格式: 高正君, 张佩炯, 司小强. 融合多维注意力机制 CNN 皮肤肿瘤图像分割提取[J]. 计算机测量与控制, 2022, 30(8): 161-168.

图像可以清晰的显示皮肤肿瘤的病灶特征,但由于病例的多样性,导致病灶间的差别微乎其微,医生通过肉眼观察方式分析判断病灶类别会变得十分困难。倘若能借助计算机辅助诊断,这样一来可以大大缓解医生的就诊压力,方便于偏远山区的患者进行远程会诊,简化了患者和医生之间的就诊繁琐程序。因此,医学图像的处理对医生更快捷准确的诊断患者的病情有着重要意义。

医学图像的分割是对图像中健康皮肤组织上所携带的病灶区域进行提取,简单地讲就是将一幅医学数字图像分割成不同的区域,按照灰度、颜色、纹理等特征,把同一图像区域中的特征相似性和异同性分割出来,为医生诊断提供可靠信息,图像分割结果的好坏直接影响计算机对图像视觉的理解^[3]。医学图像智能化分割技术通常有:阈值分割技术、区域分割技术、边缘检测分割技术。2013年,徐舒畅等人选用独立色彩浓度阈值实现了皮肤镜图像的病灶分割,阈值分割技术对皮肤镜图像中颜色分布很依赖,由于皮肤镜中有气泡、毛发、伪影等干扰性介质,对分割结果影响还是比较大的^[4]。2015年,Sumithra等人使用区域增长算法实现了皮肤镜病灶图像的分割,在区域增长算法中,需要人为设置种子点,这样的话易受主观因素的干扰,智能化程度较低^[5]。2018年,Vesal等人通过改进U-Net网络,应用于完整尺寸皮肤镜图像分割,将分割结果与人工分割结果相对比,Jac相似系数提升到76%。虽然计算速度较快,但在皮损区的边界处常会充斥着毛发,图像分割的精确度影响较大^[6]。

现如今,智能优化算法中的深度卷积神经网络技术已广泛应用于图像处理,并且在医学图像分割领域取得了很大的进展。与其它智能算法方法相比,卷积神经网络(CNN)可以从大量数据集中自动学习最有用的特征^[7]。然而,现有的CNN大多面临以下问题:首先,在卷积层设计时,特别是对于小目标而言,不同的空间位置使用相同的权值,这可能会导致缺乏空间感,从而降低了处理柔性形状和位置结构的性能^[8]。其次,CNN通常使用大量的特征通道,而这些通道可能是冗余的,许多网络如U-Net使用具有不同语义信息的低级和高级特性的连接,分割任务时,突出相关通道而抑制不相关通道有利于分割任务^[9]。第三,CNN通常提取多尺度特征来处理不同尺度的目标,但还是缺乏对特定图像最适合的分割尺度意识,由于嵌套的非线性结构,CNN决策的局限性变大,限制了它们在临床决策中的应用。

2021年,陈法法等人以U-net网络为主干网络,融合了双注意力机制对锈蚀图像进行分割,虽然在分割图像时准确率达到95%,但由于缺少尺度注意力模块,图像的提取时间较长,敏感度较低^[10]。基于传统的卷积神经网络在处理图像时存在的局限性,我们提出了一种综合注意模块网络CA-Net图像分割算法,为了实现这些注意力模块,考虑到空间位置、特征通道和尺度指标的相互配合,同时受U-Net网络架构的启迪,设计了双通道多尺度空间注意力

模块、残差网络通道注意力模块和自适应选择特征的尺度注意力模块,设计的编码模块以多尺度输入残差网络为基础层,利用综合注意力模块,分割提取自适应相关的空间区域、特征通道和尺度。本文提出的综合注意模块网络CA-Net图像分割算法,输入的是皮肤肿瘤图片,输出的是一个二值化图像,模型通过识别图像的病灶区和皮肤健康区来检测皮肤肿瘤的病灶边缘并提取病灶区的特征,然后与数据库中的真实标签图进行比对,作出相关评估来验证本文所提算法的性能。

1 相关理论原理

1.1 卷积神经网络原理

卷积神经网络从类型上来说属于深度神经网络,多层感知器的设计,实施较少的预处理后,可实现特征分析处理^[11]。其网络架构一般由输入层、卷积层、池化层、全连接层及输出层顺序连接而成^[12]。

卷积层的作用将上一层的输出进行卷积处理后提取卷积特征,在卷积层中,输入数据要进行卷积处理计算,通过卷积核的稀疏连接和权值共享形式,最终将卷积后的特征图作为输出^[13]。卷积核与输入数据中的感受野进行一次卷积操作得到输入数据的局部特征,通过平移的方式与输入数据中的多个感受野重复进行多次卷积得到多个局部特征,最终组合成全局特征作为一种输出特征^[14]。图1为2D矩阵的卷积过程。

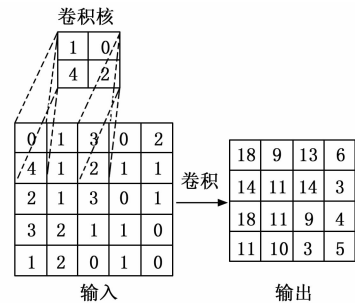


图1 2D矩阵卷积过程

池化层的作用是对上一层的输出进行降维和抽象处理,提取局部特征^[15]。常见的池化方法有最大池化和平均池化,其池化过程如图2所示。池化操作将卷积的特征激活转换为维数更低的池化特征,进而对噪声具有很好的鲁棒性^[16]。在相邻区域里,由同一个卷积核生成合并特征,该特征有助于降低神经网络中激活函数对像素和神经网络架构的敏感性^[17]。在参数设定上,池化也需要指定合并区域的大小和步长。

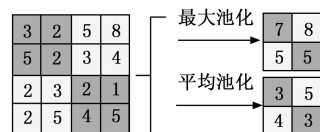


图2 最大池化和平均池化过程

1.2 激活函数

在神经网络架构中, 各层之间连接都需要激活函数进行非线性变化处理, 从而提高了神经网络的模拟学习能力。卷积神经网络中最常用的激活函数为 Sigmoid 函数和线性修正单元 ReLU 函数^[18]。

Sigmoid 函数的定义:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

ReLU 函数的定义:

$$f(x) = \max\{0, 1\} = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

Sigmoid 函数值的变化范围为 $[0, 1]$, 该函数的特点是在中间小范围内有梯度增量, 倘若网络输入的初始化权值处于映射值两端, 会出现梯度增量为 0, 此时卷积神经网络在学习训练时收敛速度变慢。而 ReLU 函数恰好弥补了 Sigmoid 函数的局限性, 在 $x > 0$ 时, 梯度增量为 1, 缓解了梯度消失的难题, 且卷积神经网络在训练时收敛速度加快^[19]。

2 综合注意力算法模块设计

本文提出的综合注意力网络模型 CA-Net 如图 3 所示, U-Net 主干网是由编码器和解码器组成的点到点可训练的网络^[20]。该编码器作为一个特征提取器, 在多个尺度上连续获取高维特征, 解码器利用这些编码特征恢复分割图像目标。在 CA-Net 网络中有 4 个空间注意力模块, 分别为 SA1—SA4; 4 个通道注意力模块, 分别为 CA1—CA4; 1 个尺度注意力模块 LA。通过增加卷积块, 同时实现对特征图的空间位置、通道和尺度的综合注意力引导, 空间注意力用于增强特征上感兴趣的区域, 同时滤除潜在的无关背景部分。提出的多尺度空间注意力模块是一种低分辨率(如 SA1) 非局部模块和其它分辨率(如 SA2~SA4) 双路径组合, 我们也可称它为空间关节注意模块, 该模块可加强像素间的关系, 目的能使该网络更好地聚焦于分割目标。

2.1 联合空间注意力模块

在神经网络中, 联合 4 个空间注意力模块来学习 4 种不同分辨率的图像。首先, 对于低分辨率图像(如 SA1) 的空间注意力, 使用非局部模块来捕捉整体图像有较好认知的所有像素间的特征。SA1 模块结构如图 4 所示。其中 x 代表大小为 $256 \times H \times W$ 输入特征映射, 256 是输入通道号, H 、 W 分别表示高度和宽度。采用输出通道号为 64 三个并行的卷积层对 x 进行降维, 分别得到 3 个压缩特征映射 x' 、 x'' 和 x''' , 且它们具有相同的大小 $64 \times H \times W$ 。这 3 个特征映射可以重构成大小为 $64 \times HW$ 的 2D 矩阵。

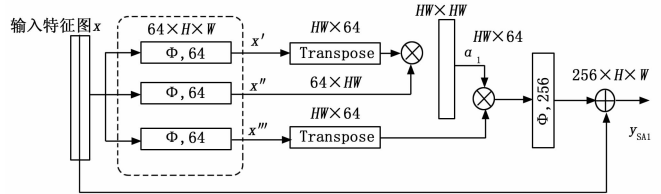


图 4 空间注意力模块结构 SA1

空间注意力系数可通过式 (3) 定义:

$$\alpha_1 = \sigma(x'^T \cdot x'') \quad (3)$$

其中: x'^T 表示 x' 的转置矩阵, σ 代表 sigmoid 激活函数, $\alpha_1 \in (0, 1)^{HW \times HW}$, α_1 将每个像素的特征表示为所有像素特征的加权和, 以保证图像像素间的相互作用。

修正后的降维特征图可用式 (4) 来表示:

$$\hat{x} = \alpha_1 \cdot x'''^T \quad (4)$$

其中: \hat{x} 表示降维重构成大小为 $64 \times H \times W$, 最终残差连接有利于促进在训练期间的信息传递。SA1 的输出由式 (5) 得到:

$$y_{SA1} = \Phi^{256}(\hat{x}) + x \quad (5)$$

非局部模块应用到高分辨率的特征映射从此在局限性, 需将注意力闸门 (AG) 扩展到 SA2~SA4 的空间注意力模块中自学习注意力系数^[21]。由于单个 AG 会引发噪声空间注意力映射, 因此提出了一种双路径空间注意力, 该空间

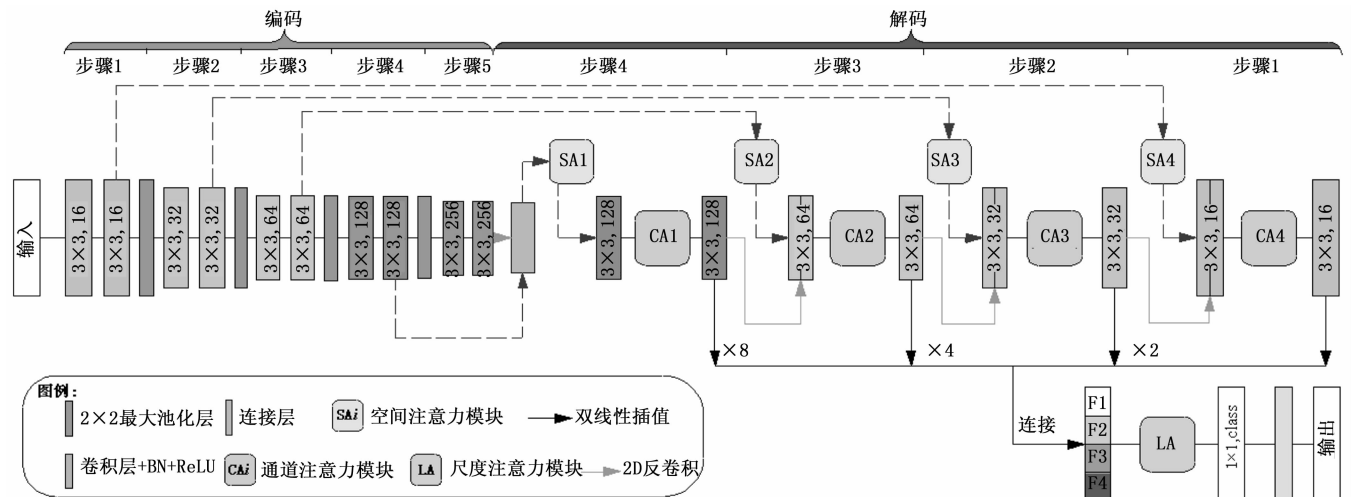


图 3 综合注意力模块流程图

注意力是利用两个 AG 并行增强对兴趣区域的注意, 并降低注意力映射中的噪声, 从而可提高图像分割的鲁棒性。单路径空间注意力 SA 如图 5 所示, 双路径空间注意力结构 SA2—SA4 如图 6 所示。

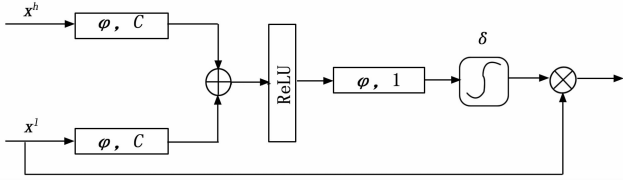


图 5 单路径空间注意力模块结构 SA

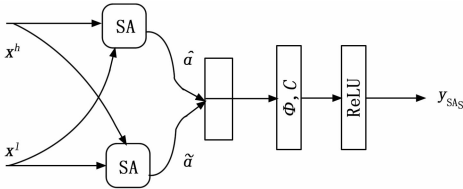


图 6 双路径注意力模块结构 SA2—SA4

在单路径空间注意力模块中, x^l 表示在编码器中尺度大小为 s 的低级特征图, x^h 表示从解码器底部向上采样的高级特征图, 尺度大小为 $s+1$, x^l 和 x^h , 由 ReLU 函数获得的特征映射紧接着被输入到另一个 1×1 卷积中, 该卷积输出一个通道号, 为了得到像素的注意力系数, 由激活函数追踪通道号, 注意力系数 $\alpha \in [0, 1]^{H \times W}$ 。在双路径 AG 中, 两种路径的空间注意力映射分别由 $\hat{\alpha}$ 和 $\tilde{\alpha}$ 来表示, 双路径 AG 的输出 SA_s ($s=2, 3, 4$) 可由式 (6) 定义:

$$y_{SA_s} = \text{ReLU}[\Phi^C((x^l \cdot \hat{\alpha}) \Theta(x^l \cdot \tilde{\alpha}))] \quad (6)$$

其中: Θ 代表通道级联, Φ^C 通道号为 C 的 1×1 卷积, 空间注意力模块中 s 分别取 2, 3, 4 时, C 分别为 64, 32, 16。

2.2 通道注意力模块

利用通道级联把来自编码器中低级的图像特征和解码器中高级的图像特征空间注意力结合起来。编码器中的特征通道主要包含低级图像特征信息, 而解码器的特征通道则包含更多的语义信息^[22], 为了更好地利用特征通道最有效的信息, 通道注意力模块能自动突显相关的特征通道信息, 同时抑制掉无关的特征通道信息^[23]。通道注意力模块结构如图 7 所示。

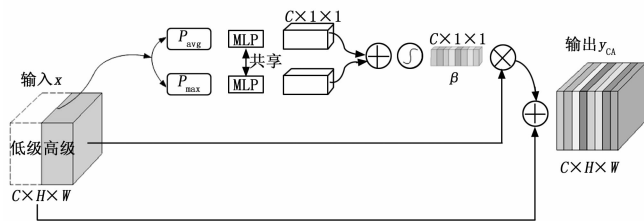


图 7 通道注意力模块结构

通常特征通道是利用平均池化来激发的, 在通道注意

力模块中, 采用最大池化来保持更多的信息。 x 代表通道号为 C 的级联输入特征映射, 平均池化 P_{avg} 和最大池化 P_{max} 两者配合来获取每个通道中的所有信息, 其中 $P_{\text{avg}}(x) \in R^{C \times 1 \times 1}$, $P_{\text{max}}(x) \in R^{C \times 1 \times 1}$ 。多层感知 (MLP) 的介入主要用来获取通道注意力系数 $\beta, \beta \in [0, 1]^{C \times 1 \times 1}$, MLP 由两个全连接层构成, 处理后的结果送入 sigmoid 激活函数可得到 β , 通道注意力模块输出 y_{CA} 可由式 (7) 定义:

$$y_{CA} = x \cdot \beta + x \quad (7)$$

2.3 尺度注意力模块

在不同尺度下, U-Net 主干网络可以获取特征映射。为了更好地处理不同尺度的目标图像, 需要将已预测的图像特征结合起来。然而对于既定的目标图像, 各种尺度的特征映射可能与目标图像的相关性差异较大, 为了能自动确定每个像素的尺度权值, 使网络能自适应于给定的输入图像的尺度, 提出了尺度注意力模块, 该模块能自动学习每个图像的尺度权重, 以捕捉不同尺度下的图像特征。尺度注意力模块结构如图 8 所示。

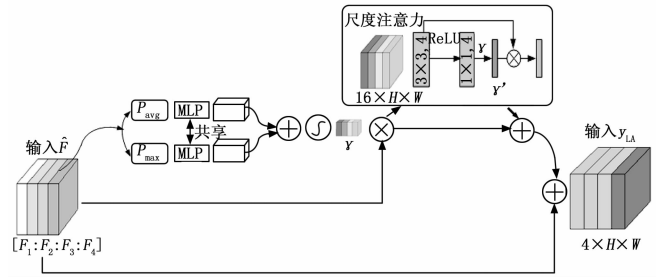


图 8 尺度注意力模块结构

从图 8 可以看出, 它的输入是由解码器中获得不同尺度级联的插值特征映射组成, 为了降低计算机的运行负荷, 特征映射压缩成 4 个通道, 这 4 个通道由 1×1 卷积构成, 不同尺度图像的压缩结果级联成混合的特征映射 \hat{F} 。与通道注意力模块类似, 把平均池化、最大池化以及多层感知 (MLP) 三者结合起来得到尺度注意力系数 $\gamma, \gamma \in [0, 1]^{4 \times 1 \times 1}$ 。为了给每个图像像素配置多尺度注意力权值, 采用了空间注意力模块 LA^* , LA^* 由 3×3 和 1×1 卷积层组成, γ^* 代表像素尺度注意力系数, $\gamma^* \in [0, 1]^{1 \times H \times W}$, 第 1 层卷积使用 ReLU 激活函数, 处理得到 4 个输出通道, 第 2 层卷积使用 Sigmoid 激活函数, 处理得到 4 个输出通道。尺度系数不断将权值配置到图像目标域, 无论图像目标尺度有多大, 尺度注意模块能够使卷积神经网络与最适宜的尺度相匹配, 不相关的区域值变小, 图像目标域变大, 分割提取精度增强。尺度注意力模块输出 y_{LA} 可由式 (8) 定义:

$$y_{LA} = \hat{F} \cdot \gamma \cdot \gamma^* + \hat{F} \cdot \gamma + \hat{F} \quad (8)$$

3 实验验证及结果分析

3.1 实验数据采集及参数设置

为了验证本文提出综合注意力模块在进行卷积神经网络深度学习后, 能够准确有效地分割病灶图像, 本文使用了 ISIC2018 数据集及医院病理科提供的病理切片, 该数据

集包含 2 594 张图像和它们的真值。所采用的实验软件平台为 PyTorch, 硬件环境为 NVIDIA Geforce GTX 1080 Ti GPU。我们将数据集随机分为 1 816、260 和 518 分别进行训练、验证和测试。皮肤肿瘤图像分割数据集的原始大小是从 720×540 到 6 708×4 439 范围内, 将每张图像的大小调整为 256×342, 并通过均值和标准差进行归一化。在训练过程中为了让图像数据增强, 把图像随机的按水平、垂直及在 $(-\pi/6, \pi/6)$ 任意角度内旋转, 图像裁剪大小为 224×300。神经网络采用自适应矩估计进行训练, 初始学习率为 10^{-4} , 权值衰减为 10^{-8} , 批量大小为 16, 迭代次数为 300。学习速率每 256 次衰减 0.5, 特征通道号设为 16, 每次下采样后增加一倍。

3.2 评价指标

对皮肤镜图像进行分割提取, 需生成二值掩码, 这样才能对病灶区和正常区进行区分, 评估本文算法性能的优越性需通过评价指标来决策。

常见的评价指标包括敏感性 (Se)、相似系数 (Dice)、平均对称表面距离 (Assd)、准确性 (Acc) 及并交比 (Jac), 使用骰子损失函数对每个网络进行训练, 它们的定义如下:

$$Se = \frac{E_{TP}}{E_{TP} + E_{FN}} \quad (9)$$

$$Dice = \frac{2|R_a \cap R_b|}{|R_a| + |R_b|} \quad (10)$$

$$Assd = \frac{1}{|S_a| + |S_b|} \times \left(\sum_{a \in S_a} d(a, S_b) + \sum_{b \in S_b} d(b, S_a) \right) \quad (11)$$

$$Acc = \frac{E_{TP} + E_{TN}}{E_{TP} + E_{TN} + E_{FP} + E_{FN}} \quad (12)$$

$$Jac = \frac{E_{TP}}{E_{FP} + E_{FN} - E_{TP}} \quad (13)$$

式 (9)、式 (12)、式 (13) 中, E_{TP} 、 E_{TN} 、 E_{FP} 、 E_{FN} 分别表示真阳性、真阴性、假阳性、假阴性区域像素点数量。式 (10) 中 R_a 和 R_b 分别表示按照算法分割的图像区域集合和真值集合, $Dice \in (0, 1)$ 。式 (11) 中 S_a 和 S_b 分别表示智能化分割边界点集合和真值集合, $d(v, S_a) = \min_{w \in S_a} (\|v - w\|)$ 表示从点 v 到 S_a 集合中所有点的最小欧氏距离。

骰子损失函数 L_{Dice} 的定义为:

$$L_{Dice} = 1 - \frac{2 \sum_{i,j} (t_{ij} p_{ij})}{2 \sum_{i,j} (t_{ij} p_{ij}) + \sum_{i,j} (t_{ij}^2) + \sum_{i,j} (p_{ij}^2)} \quad (14)$$

其中: t_{ij} 和 p_{ij} 分别表示像素 (i, j) 的真值和预测值。通过 $\sum_{i,j} (t_{ij} p_{ij})$ 即真值和预测值的内积之和来计算并交比 Jac, 进而提高了皮损域的综合关注度。

3.3 本文算法的实验验证

3.3.1 不同因素下空间注意力的指标影响

我们首先在不使用通道注意力和尺度注意力模块的情况下研究了空间注意力模块的有效性, 并从 3 种状况下比

较了多层空间注意力在皮肤病损图像分割中的视觉特征, 该视觉比较如图 9 所示。1) 在空间注意力模块 SA1-4 中使用了标准单路径 AG, 该状况的视觉图用 s-AG 来表示。2) 在空间注意力模块 SA1-4 中使用了双路径 AG, 该状况的视觉图用 t-AG 来表示。3) 仅在空间注意力 SA1 中使用非局部模块, 该状况的视觉图用 n-Local 来表示。本文提出的在 SA1 中使用的非局部模块和 SA2-4 双路径 AG 联合空间注意力方法分割视觉图用 Js-A 表示。对于 U-Net 网络, 跳过连接是通过编码器和解码器串接中的相应特性来实现的, 不同方法之间的量化评估如表 1 所示。

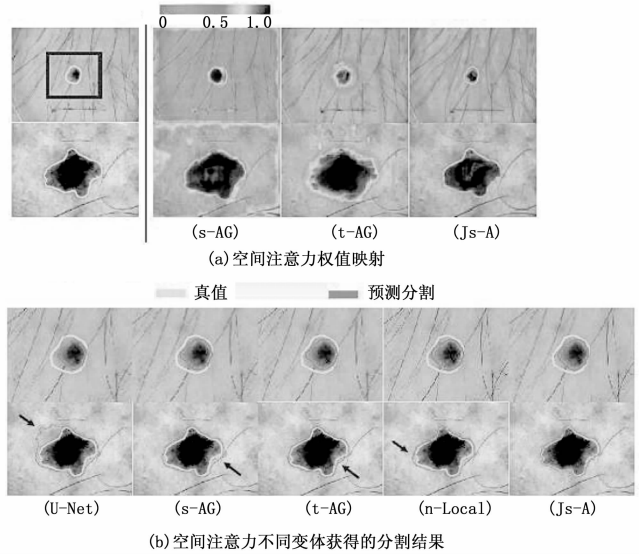


图 9 不同空间注意力结构在皮肤病损分割中的视觉比较

表 1 不同空间注意力结构的皮肤病损图像分割定量评价

| 方法 | Para/M | Dice/% | Assd/pix |
|---------------------|------------|-------------------|------------------|
| Baseline(U-net) | 1.9 | 87.77±3.51 | 1.23±1.07 |
| 单路径空间注意力(s-AG) | 2.1 | 88.46±3.37 | 1.18±1.24 |
| 双路径空间注意力(t-AG) | 2.4 | 89.18±3.29 | 0.90±0.50 |
| 非局部模块空间注意力(n-Local) | 1.9 | 90.15±3.21 | 0.65±0.72 |
| 联合空间注意力(Js-A) | 2.0 | 90.83±3.31 | 0.81±1.06 |

从图 9 (a) 可以看出, 单路径 AG 几乎关注每一个图像像素, 这意味着它的分散性较大。虽然双路径 AG 优于单途径 AG, 但自适应程度不够强。相比之下, 本文提出的空间联合注意力方法比上述方法更关注分割目标。图 9 (b) 给出的分割结果可以看出, 在神经网络中引入空间注意块大大提高了分割精度, 此外, 本文提出的联合空间注意力 Js-A 方法在两个分割案例下都比其他空间注意方法得到了更好的结果。尤其在第二个病例中, 病变形状复杂, 边界模糊, 提出的联合空间注意力 Js-A 保持了较好的结果。

从表 1 可以看出, 所有使用空间注意力图像分割都比 Baseline 的分割精度更高。同时, 我们观察到双通道空间注意力闸门 AG 比单通道的更有效, 联合空间注意力闸门 AG

优于其他空间注意力闸门，相似系数 Dice 从 88.46% 提高到 90.83%。

3.3.2 不同因素下通道注意力的指标影响

在通道注意力方法比较中，仅引入了通道注意力模块来验证所提出方法的有效性。首先研究了通道注意力模块在网络中插入位置的影响：1) 仅在编码器中，2) 仅在解码器中，3) 同时插入编码器和解码器中。这 3 种情况分别表示为 C-A (Enc)，C-A (Dec) 和 C-A (Enc& Dec)。与此同时还比较了在使用和不使用最大池化时对通道注意力模块的影响，该影响视觉比较如图 10 所示。不同方法之间的量化评估如表 2 所示。

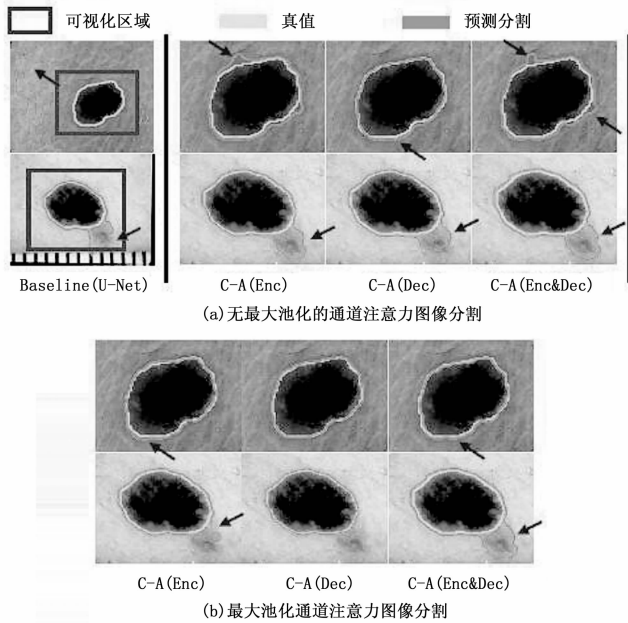


图 10 不同通道注意力方法在皮肤病损分割中的视觉比较

表 2 不同通道注意力的皮肤病损图像分割定量评价

| 方法 | P_{max} | Para/M | Dice/% | Assd/pix |
|------------------|-----------|------------|-------------------|------------------|
| Baseline(U-net) | — | 1.9 | 87.77±3.51 | 1.23±1.07 |
| C-A (Enc) | × | 2.7 | 91.06±3.17 | 0.73±0.56 |
| C-A (Enc) | ✓ | 2.7 | 91.36±3.09 | 0.74±0.55 |
| C-A (Dec) | × | 2.7 | 91.56±3.17 | 0.64±0.45 |
| C-A (Dec) | ✓ | 2.7 | 91.68±2.97 | 0.65±0.54 |
| C-A (Enc& Dec) | × | 3.4 | 90.85±3.42 | 0.92±1.40 |
| C-A (Enc& Dec) | ✓ | 3.4 | 91.63±3.50 | 0.58±0.41 |

从图 10 可以看出，当皮肤镜图像纹理复杂时，Baseline (U-Net) 的图像分割性能较差，而通道注意力模块的介入提高了目标分割的准确性。显然，本文提出的通道注意力模块 C-A (Dec) 的图像分割精度高于其它模块。

从表 2 可以看出，通过在不同的因素下进行定量比较分析，通道注意力模块的介入确实提高了分割性能，而且带有最大池化信息的通道注意力模块比只使用平均池化的通道注意力模块性能更好。此外从评价数据发现，插入解

码器中的通道注意力模块要比插入编码器中或同时插入编码器和解码器中的性能更好，C-A (Dec) 的 Dice 为 91.68%。

3.3.3 不同因素下尺度注意力的指标影响

我们只引入了尺度注意力模块来验证所提出的尺度注意力方法的有效性。设 L-A (1- K) 表示来自不同尺度级联特征映射的尺度注意力，为了研究特征映射尺度大小对分割的影响，我们分别取 $K=2, 3, 4, 5$ 进行比较。不同尺度注意力方法对皮损分割的量化评估如表 3 所示，不同尺度下的特征融合图像分割中的视觉比较如图 11 所示，皮肤镜图像上尺度注意力的可视化如图 12 所示。

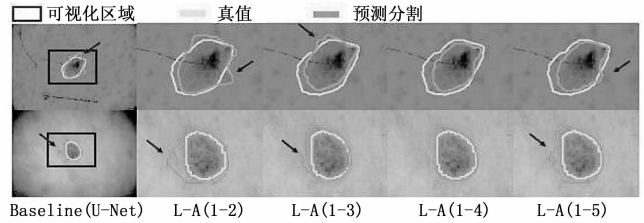


图 11 不同尺度下的特征融合图像分割中的视觉比较

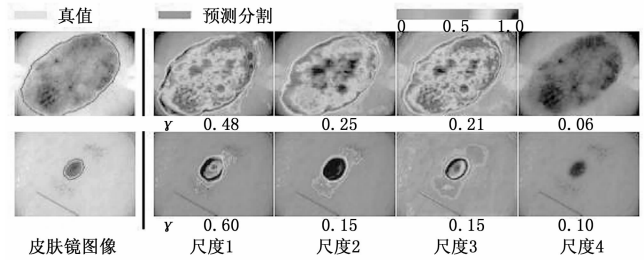


图 12 皮肤镜图像上尺度注意力的可视化

从图 12 可以看出，暖色代表较高尺度注意力分割区域，其中每幅图下面的数字表示尺度注意力系数 γ 。图中有大病灶和小病灶之分，可以观察到，在尺度 2 和 3 下，跟小病灶相比，大病灶具有较高的 γ ，而在尺度 1 下，跟大病灶相比，小病灶具有较高的 γ ，像素尺度注意力映射表明，在第 1 排图中尺度 2 下和第 2 排图中尺度 1 下尺度注意力最集中，这说明神经网络在学习训练时，自动倾向聚焦在相应的尺度上，对不同大小的病造进行分割。

表 3 不同尺度注意力方法对皮肤病损分割的量化评估

| 方法 | Para/M | Dice/% | Assd/pix |
|------------------|------------|-------------------|------------------|
| Baseline(U-net) | 1.9 | 87.77±3.51 | 1.23±1.07 |
| L-A (1-2) | 2.0 | 91.21±3.33 | 1.00±1.36 |
| L-A (1-3) | 2.0 | 91.53±2.52 | 0.70±0.61 |
| L-A (1-4) | 2.0 | 91.58±2.48 | 0.66±0.47 |
| L-A (1-5) | 2.0 | 89.67±2.40 | 0.82±0.50 |

从表 3 可以看出，在多尺度下融合特征优于 Baseline (U-Net)。当我们将尺度从 1 到 4 图像特征融合时，Dice 和 Assd 分别得到了最佳值 91.58% 和 0.66。然而，当我们将这 5 个尺度的图像特征融合在一起时，分割的准确率会降

低。这表明在低分辨率的特征映射下, 不适合预测像素的细节特征。因此在后续的实验 A 中。我们只融合了尺度 1 到 4 的特征。

3.4 皮肤肿瘤图像分割提取

甘肃省人民医院整形外科、皮肤科、提供了自 2012 年至 2019 年经皮肤镜检查后做了病理切片的皮肤肿瘤图像, 该图像宽高比各不相同。由于原始图像中存在噪声, 且在图像中除了病灶区域外, 还有大量的健康皮肤背景区域, 因此输入图像在网络训练前首先要预处理, 即需进行图像增强以便使病灶区和健康区有很大的辨识度, 对于标签图, 需要对图像进行二值化; 之后将训练样本送入模型训练, 得到优化分割模型, 再将测试样本送到已训练好的模型中测试, 获取分割结果, 为了更好地验证本文提出的融合综合注意力卷积神经网络对病灶图像的分割, 分别对提供的不同病灶样本进行了病灶区分割提取, 并与 U-Net 算法和所提出的注意力模块任意组合构成的算法机制进行对比, 分割结果如图 13 所示, 涉及的病灶类型有黑色素瘤 (A)、混合痣 (B)、基底细胞癌 (C)、角化棘皮瘤 (D)、鳞状细胞癌 (E)、皮内痣 (F)、腺样囊性癌 (G) 及脂溢性角化 (H) 等。

为了研究不同注意力组合的效果, 将本文提出的 CA-Net 与空间、通道和尺度 3 种基本注意力的任意 6 种组合状态进行了比较。局部注意力方法与综合注意力方法在皮肤损伤分割中的量化评估如表 4 所示, 其中, SA 表示多尺度联合空间注意力, CA 表示仅在解码器架构中使用的通道注意力, LA 表示尺度注意力。

与 Baseline (U-net) 相比, SA、CA 和 LA 图像分割性能方面有了很大改善。与单一的注意力方法相比, 两种注意力方法任意组合在一起效果更好。为了评估本文提出的综合注意力 CA-Net 在皮肤肿瘤图像分割的有效性和精确性, 在不同注意力组合下计算了其相似性 Dice、Assd、敏感性 Se 以及准确性 Acc 的值, 从表 4 可以看出, 在综合注意力下, 评估值都优于其他方法, 其对应值分别为 92.08%、0.58、93.77% 及 92.89%。

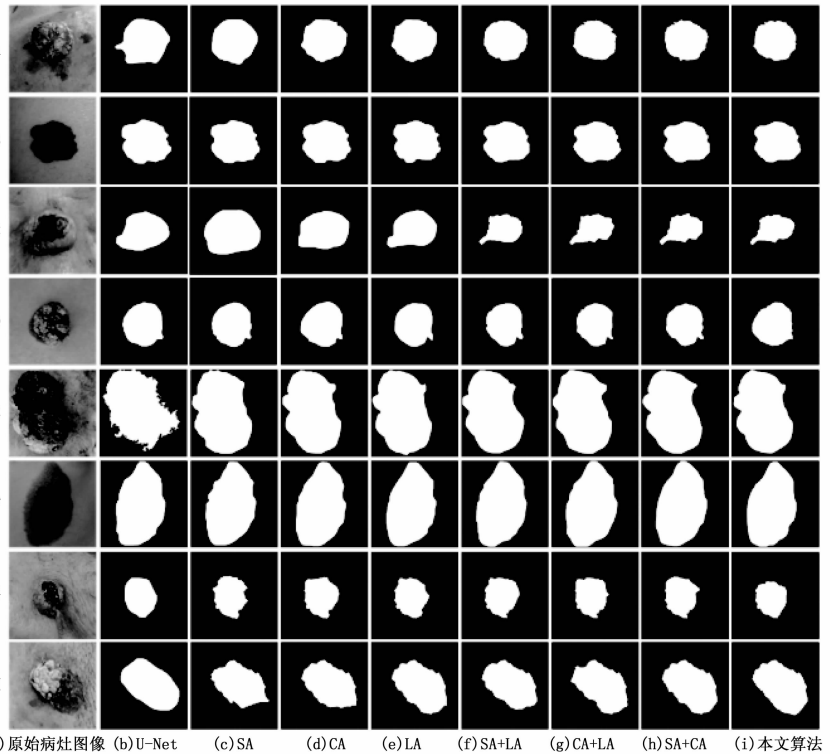


图 13 不同算法的分割结果图

从分割结果可以看出, 由于原始病灶图像的多样性和图像本身辨识度的不同, 利用不同算法在图像分割时病灶区提取的信息浓度差异较大。纵观这 8 类样本, 采用 U-Net 时, 较易分割提取到携带与病灶无关的信息成分, 部分图像辨识度低时, 对分割效果影响尤为不利。另外分别让 SA、CA、LA 注意力模块介入到卷积神经网络时, 分割模型在一定程度上削弱了网络退化, 提取到样本病灶边界更加圆滑, 从而大幅度减少了图像信息损失, 但由于空间、通道、尺度注意力模块的单一化, 在图像语义信息捕捉上适应性较差。基于此因, 构建了对 SA、CA、LA 任意两个注意力模块随机组合混搭, 我们发现, SA+CA 注意力模块组合在病灶图像分割区的信息浓度更为丰富, 在一定范围内提高了不同维度图像的特征适应性, 但由于权重在深度学习时分配的差异性, 对感兴趣的目标分割区敏感性偏低, 分割的准确性不够。而采用了本文提出的综合注意力模块算法后, 综合注意力机制更加关注目标域的信息捕捉,

表 4 局部注意力方法与综合注意力方法在皮肤病损分割中的量化评估

| 方法 | Para/M | Dice/% | Assd/pix | Se/% | Acc/% |
|-----------------|--------|------------|-----------|------------|------------|
| Baseline(U-net) | 1.9 | 87.77±3.51 | 1.23±1.07 | 85.36±3.81 | 86.54±2.01 |
| SA | 2.0 | 90.83±3.31 | 0.81±1.06 | 89.87±3.17 | 88.63±2.47 |
| LA | 2.0 | 91.58±2.48 | 0.66±0.47 | 90.17±2.34 | 91.17±1.38 |
| CA | 2.7 | 91.68±2.97 | 0.65±0.54 | 91.83±2.51 | 91.37±2.88 |
| SA+LA | 2.1 | 91.62±3.13 | 0.70±0.48 | 91.76±3.51 | 91.14±3.69 |
| CA+LA | 2.7 | 91.75±2.87 | 0.67±0.48 | 91.84±2.89 | 91.55±2.34 |
| SA+CA | 2.8 | 91.87±3.00 | 0.73±0.69 | 92.93±3.71 | 91.88±2.26 |
| 综合注意力 | 2.8 | 92.08±2.67 | 0.58±0.39 | 93.77±2.23 | 92.89±2.23 |

既有效的增强了模型分割的效果,又有效避免了特征向量上冗余的信息,平均分割的精确度可达 92.89%,从而为皮肤肿瘤智能化初步诊断提供可靠依据,在不同的医学图像分割任务中得以推广。

4 结束语

通过融合空间注意力模块、通道注意力模块及尺度注意力模块的综合注意力方法来改进传统的卷积神经网络,从而对医学图像进行分割,由于病灶等分割目标的空间位置、形状和尺度变化较大,利用该方法使神经网络在深度学习目标的空间位置和大小上,对病灶图像目标位置能实现精确分割,最大限度的关注相关通道和尺度是提高分割性能的有效途径。得出了如下结论:

1) 使用本文提出的综合注意力方法改进优化传统的卷积神经网络,可使用较少的参数获得更准确的病灶分割。

2) 在预测分割目标时,CA-Net 就如何预测每个空间位置、特征映射通道和尺度方面能很好地把控。更重要的是,我们在单向传递中获得这些注意系数时不需要额外的计算,从而很大程度地缩减了深度学习的时间。

3) 由于编码器充当特征提取器,因此在 CA-Net 网络模块设计中,大多数注意力模块都在解码器中,若在编码器中加入注意力模块,可能会导致一些潜在的图像被抑制过滤掉,在解码器中设计注意力模块能突出病灶目标图像特征的相关性。

4) 提出的综合注意力卷积神经网络 (CA-Net) 能够自适应地关注空间位置、特征通道和目标尺度。在现有的空间注意力和通道注意力方法的启发下,对网络进行了进一步的改进,提高了网络深度学习能力,更专注于感兴趣的目标区域,提高分割精度。在下一阶段的任务中,准备尝试将 CA-Net 应用于其他医学成像模式乃至 3D 医学图像的分割。

参考文献:

- [1] JAFARI M, et al. FU-Net: Multi-class image segmentation using feedback weighted U-net [C] //International Conference on Image and Graphics, Springer, 2019: 529-537.
- [2] QIN Y, KAMNITSAS K, ANCHA S, et al. Autofocus layer for semantic segmentation [J]. Proc. MICCAI, 2018, 603-611.
- [3] SCHLEMPER J, OKTAY O, SCHAAP M, et al. Attention gated networks: Learning to leverage salient regions in medical images [J]. IEEE Trans. Med. Image, 2019, 53: 197-207.
- [4] 徐舒畅. 采用独立色素浓度分布分割皮损图像 [J]. 中国图象图形学报, 2013, 18 (11): 1452-1456.
- [5] SUMITHRA R, SUHIL M, GURU D S. Segmentation and classification of skin lesions for disease diagnosis [J]. Procedia Computer Science, 2015, 45: 76-85.
- [6] VESAL S, RAVIKUMAR N, MAIER A. Skin Net: a deep learning framework for skin lesion segmentation [J]. arXiv preprint ar Xiv: 1806. 09522, 2018.
- [7] CHEN R, CHEN H, REN J, et al. Explaining neural networks semantically and quantitatively [J]. Proc. ICCV, 2019:

9187-9196.

- [8] YUAN Y, CHAO M, LO Y C. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance [J]. IEEE Transactions on Medical Imaging, 2017, 36 (9): 1876-1886.
- [9] 陈法法, 成孟腾, 杨蕴鹏, 等. 融合双注意力机制和 U-Net 网络的锈蚀图像分割 [J]. 西安交通大学学报, 2021, 55 (12): 119-128.
- [10] 李飞, 王超, 浦东, 等. 基于卷积注意力的输电线路防震锤检测识别 [J]. 计算机测量与控制, 2021, 29 (1): 1-10.
- [11] SARKER M M K, RASHWAN H A, AKRAM F, et al. SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks [C] //International Conference on Proc. MICCAI, 2018: 21-29.
- [12] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks [J]. Nature, 2017, 542 (7639): 115.
- [13] REBECCA S L, KIMBERLY M D, AHMEDIN J. CA: a cancer journal for clinicians [J]. Cancer Statistics, 2019, 69 (1): 7-34.
- [14] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification [C] //International Conference on Proc. CVPR, 2017: 3156-3164.
- [15] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deep Lab: semantic image segmentation with deep convolutional nets [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017 (99): 1.
- [16] CHEN Y, DAI X, LIU M, et al. Dynamic convolution: attention over convolution kernels [C] //International Conference on Proc. CVPR, 2020: 11030-11039.
- [17] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation [J]. IEEE Trans. Pattern Anal., 2016, 565-571.
- [18] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. Trans. Pattern Anal., 2017 Mach. Intell., 2017, 39 (12): 2481-2495.
- [19] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C] //International Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.
- [20] BI L, KIM J, AHN E, et al. Dermoscopic image segmentation via multistage fully convolutional networks [J]. IEEE Trans. Bio-med. Eng., 2017, 64 (9): 2065-2074.
- [21] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [M]. 2015: 234-241.
- [22] SCHWARZ S T, et al. In vivo assessment of brainstem depigmentation in Parkinson disease: potential as a severity marker for multicenter studies [J]. IEEE Trans., 2016, 283 (3): 789-798.
- [23] TORRENTS-BARRENA J, PIELLA G, MASOLLER N, et al. Segmentation and classification in MRI and US fetal imaging: recent trends and future prospects [J]. IEEE Med. Image Anal., 2019, 51: 61-88.