

ω -net: 一种用于多种医学图像的二次特征提取方法

吴昊¹, 徐杨^{1,2*}, 曹斌^{1,2}¹贵州大学大数据与信息工程学院, 贵州 贵阳 550025;²贵阳铝镁设计研究院有限公司, 贵州 贵阳 550009

摘要 针对人为无法及时从医学图像发现病理特征, 最终导致病情恶化的问题, 提出一种二次特征提取方法(ω -net), 用来对肺、肝脏、细胞及脑质瘤进行分割。首先, 将全尺寸的 Unet 网络作为一次特征提取路径, 再将上采样路径的第三层作为起始层, 用来扩展二次特征提取路径, 以加强特征提取能力。其次, 为了建立长期的通道依赖关系以及增强特征的位置信息, 在不同的阶段引入两种新型注意力机制进行针对性优化。最后, 复现了 10 种经典网络, 在医学图像领域与基准网络 Unet 对比, 所提网络的常用指标平均交并比、敏感性、精确度和准确率最高分别提升了 0.0787、0.1287、0.1216、0.0201。经过在 4 种数据集上的指标数值和可视化结果比较, ω -net 的多项性能指标均优于其他网络, 证明了该网络的有效性与优越性。

关键词 医学图像; 语义分割; 二次特征提取; 注意力机制

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP213089

ω -net: A Secondary Feature Extraction Method for Multiple Medical Images

Wu Hao¹, Xu Yang^{1,2*}, Cao Bin^{1,2}¹College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, Guizhou, China;²Guiyang Aluminum Magnesium Design & Research Institute Co., Ltd., Guiyang 550009, Guizhou, China

Abstract The problem of finding pathological features artificially from medical images over time, which ultimately leads to deterioration, has increased significantly. Thus, it has become a crucial area of interest for researchers. This study introduced a secondary feature extraction method (ω -net) with the ability to segment lung, liver, nucleus, and brain tumors. First, we used the full-size Unet as the primary feature extraction path. Similarly, we used the third layer on the upsampling path as the starting layer to expand the secondary feature extraction path to enhance the feature extraction capability. Second, we introduced two new-attention mechanisms at various stages for targeted optimization to establish long-term channel dependence and enhance feature location information. Finally, the study reproduced 10 classic networks. With the application of Unet in the medical imaging field, the commonly used indicators, including mean intersection of union, sensitiveness, precision, and accuracy of the proposed network, increase by 0.0787, 0.1287, 0.1216, and 0.0201, respectively, compared with the benchmark network. The study evaluated the effectiveness and superiority of the introduced network and compared the index values and visualization results on four types of datasets. The results showed that the introduced network outperformed the other existing networks.

Key words medical image; semantic segmentation; secondary feature extraction; attention mechanism

1 引言

人为对临床相关的病理特征进行标注, 如细胞、肝脏、肺部等, 耗时耗力、精度低、速度慢。尤其在现实条件下, 每天获取的图像量呈指数级增长, 因此对处理速

度的要求十分苛刻。为了辅助医生实现快速准确的诊疗, 基于计算机辅助, 利用深度卷积架构进行自动分割已是当前时代的潮流。

自 2015 年 Ronneberger 等^[1]提出 Unet 用来分割细胞膜以来, 医学图像分割领域吸引了众多研究者的关

收稿日期: 2021-11-29; 修回日期: 2021-12-19; 录用日期: 2022-01-05; 网络首发日期: 2022-01-15

基金项目: 贵州省科技计划(黔科合支撑[2021]一般 176)

通信作者: *xuy@gzu.edu.cn

注, 由于其良好的结构和性能, 被频繁作为改进的基础网络应用于各种医学图像的分割。Xiao 等^[2]和 Guan 等^[3]分别受残差连接和密集连接的启发, 将 Unet 的每一个子模块分别替换为残差连接和密集连接的结构, 提出 ResUnet 和 DenseUnet, 分别用于视网膜血管的分割和图像伪影的去除。Ibtehaz 等^[4]受到启发, 将 MutiRes 模块与 Unet 结合, 同时他们认为低层次的特征和较高层次的特征存在语义差距, 不能直接通过跳跃路径拼接, 于是将跳跃路径替换为残差路径, 用于对皮肤病进行分割。Alom 等^[5]通过将残差连接和循环卷积结合, 替换 Unet 中的卷积块, 提出 R2Unet 用于对视网膜血管、皮肤病变及肺部进行分割。Zhou 等^[6]认为 Unet 最佳的深度在先验上是未知的, 并且跳跃连接只是对相同尺寸的特征层进行强制融合, 具有限制性。因此将跳跃连接替换为多个小型 Unet 的集合来抓取不同层次的特征, 通过叠加的方式对抓取的特征进行整合, 提出 Unet++ 用于对细胞膜等多种医学图像进行分割。Huang 等^[7]认为 Unet++ 尽管使用了密集的跳跃连接, 却没有充分利用好多尺度的信息, 便对上采样的每一层与同层及次层进行跳跃连接, 用于分割肝脏肿瘤。有学者认为仅靠网络结构本身无法完全解决问题, Oktay 等^[8]就在 Unet 中引入注意力机制, 在拼接之前使用了注意力模块, 重新调整了编码器的输出特征。Jha 等^[9]认为单个 Unet 的特征提取能力有限, 因此引入 VGG-19, 空洞空间金字塔池化 (ASPP) 构成第二个 Unet, 并对两个网络提取的特征进行融合, 提出 DoubleUnet 用于分割癌变皮肤和结肠。Valanarasu 等^[10]认为 Unet 的结构本身会导致分割时出现小目标丢失、对象边缘模糊等现象, 于是在每一次下采样时通过交叉残差融合模块 (CRFB) 与 Ki-Net 结合, 提出 KiUnet 用于分割 B 超图像。

基于以上分析, 现有的大量关于改进 Unet 的文献主要是从机制引入和结构本身两个方面进行改进的。但是模型过于复杂, 特征提取能力和泛化能力还有不足, 受此启发, 本文提出了一种二次特征提取网络 (ω -net)。同时, 针对二次特征提取阶段卷积过多的问题, 引入金字塔拆分注意力模块建立长距离通道依赖关系, 引入坐标注意力模块增强特征的位置信息, 并且二者能够为特征层调整不同的权重, 使模型可以有目的地学习。所提模型整体结构简单、易于理解、学习能力优越。

2 网络结构及算法原理

2.1 注意力机制

2.1.1 坐标注意力

由于压缩与激励 (SE) 机制^[11]只考虑了通道间信息的关系而忽视了位置信息的重要性, 无法捕获目标特征的空间信息。坐标注意力 (CA) 机制^[12]同时考虑了通道间的关系和位置信息, 不仅捕获了跨通道的肝

脏、细胞、肺等信息, 还包含方向与位置信息。在分成 X、Y 路径后, 首先利用两个一维全局池化将输入特征层沿垂直和水平方向分别聚合为两个对方向感知和对位置敏感的特征层; 然后将嵌入了特定方向信息的这两个特征层分别编码为两个注意力层, 每个注意力层都沿一个空间方向提取特征, 所计算的每个元素都能显示目标是否存在于相应的行和列内, 因此位置信息被保存在注意力层内; 最后通过乘法将两个方向的特征融合到整个输入特征层上, 以增强对目标的表示。这种编码过程可以更准确地定位目标的位置, 从而帮助整个模型更好地学习, 具体结构如图 1 所示。

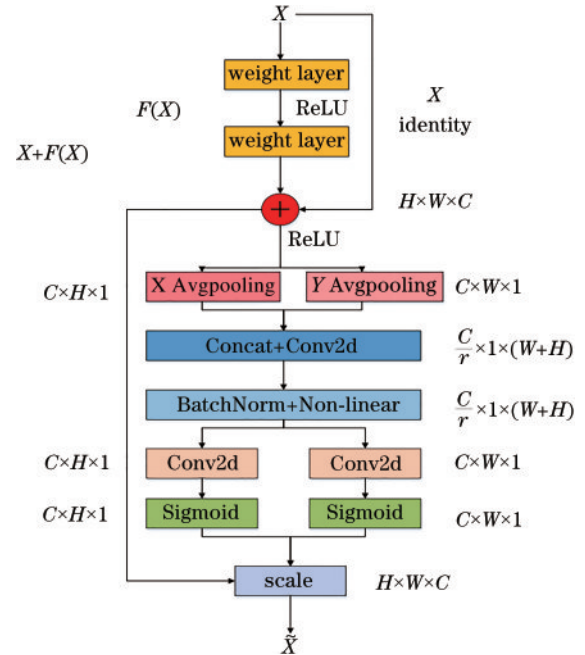


图 1 坐标注意力模块

Fig. 1 Coordinate attention module

2.1.2 金字塔拆分注意力

图 2 所示为金字塔拆分注意力 (PSA) 模块^[13], 其能够捕获更细微的多尺度空间信息, 建立更长距离的通道依赖关系, 学习更丰富的多尺度特征。首先, 利用 Split and Concat (SPC) 模块对通道进行切分, 针对每个通道特征层上的空间信息进行多尺度特征提取; 接着, 利用 SEWeight 模块获取不同尺度特征层的通道注意力向量; 然后, 利用 Softmax 对注意力向量进行重新标注, 得到新的多尺度通道交互之后的注意力权重; 最后, 对重新标注的权重和相应的特征层按元素进行点乘操作, 输出一个加权之后的特征层。

2.2 所提网络结构

2.2.1 总体介绍

由于过深的网络会出现梯度爆炸和消失现象, 模型无法学习, 而过浅的网络会让模型学习能力变差, 捕获不了目标像素, 于是借鉴 Unet 的 5 层结构, 设计出 ω -net, 如图 3 所示。首先由左边的网络对 4 种图像进行一次特征提取, 分为下采样阶段和上采样阶段。每次

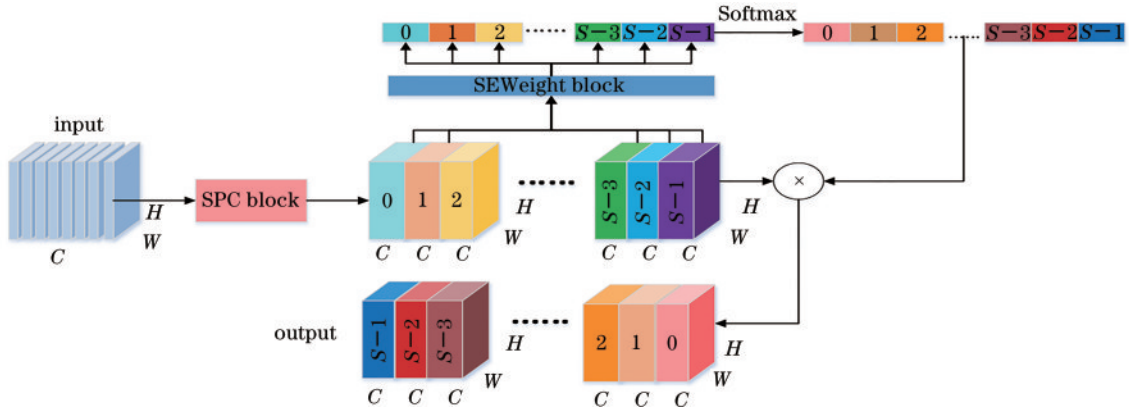


图 2 金字塔拆分注意力模块
Fig. 2 Pyramid split attention module

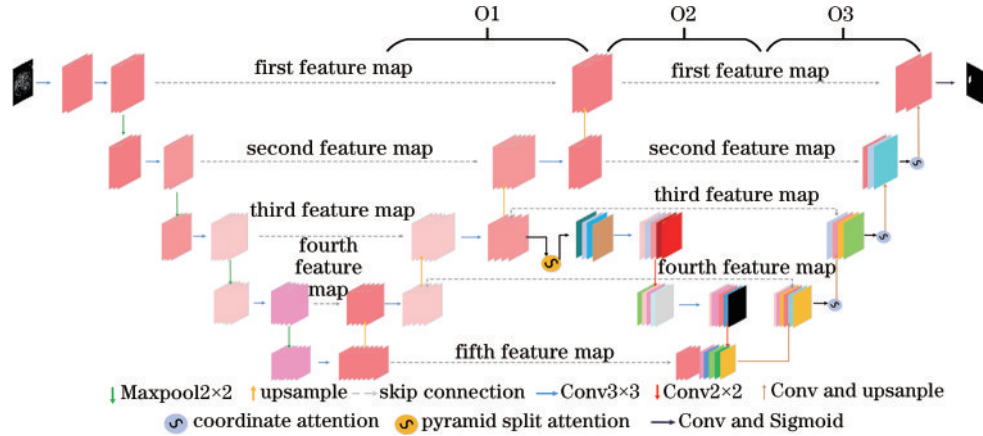


图 3 ω -net 结构
Fig. 3 ω -net structure

下采样通过 2 次 3×3 卷积和 1 次 2×2 最大池化；每次上采样会经过跳跃路径拼接同层特征，然后通过反卷积，获得信息分布较为粗略的特征层。接着将 O1 路径的第三层分为两条路径，O1 路径继续进行上采样和 3×3 反卷积，O2 路径将 O1 路径的第三层作为起始路径进行特征的二次提取，最后将两次提取的特征融合后输出。

由于一次特征提取的过程中采用了 2×2 的最大池化，会造成大量信息丧失，所以在二次特征提取的过程中用 2×2 的卷积进行下采样，将多余尺寸作为通道信息，并进行 3×3 的卷积，这样做能够保留更多的语义信息。同时为了弥补模型的缺陷，在 O2 路径的起始层引入 PSA，对提取的信息标注权重的同时建立长距离的通道依赖关系。

二次特征提取通过反复进行卷积对特征进行加强提取，引入注意力机制后会更有针对性地捕获特征，但是会造成信息损失的问题，所以利用 O1 路径各层所获得的特征层对 O3 路径进行层对层信息弥补，然后进行上采样。同样，特征层冗余问题也会引入，并且多次卷积后空间几何信息丧失严重，所以在每次拼接反卷积后引入新型注意力机制 CA。由于第一层已经拥有丰

富的位置信息，而第五层再经过池化会丢失信息，所以未引入 CA。

2.2.2 金字塔拆分注意力融入部分

对于输入特征层 X ，首先将其切分为 S 个部分 (X_0, X_1, \dots, X_{S-1})，然后对各部分提取尺度不同的特征，计算公式为

$$\begin{cases} K_i = 2 \times (i + 1) + 1 \\ G_i = 2^{\frac{K_i - 1}{2}}, F_i \in \mathbf{R}^{C' \times H \times W} \\ F_i = \text{Conv}(K_i \times K_i, G_i) X_i \\ i \in [0, S - 1] \end{cases}, \quad (1)$$

式中： K_i 代表第 i 个卷积核，分别取 (3, 5, 7, 9)； F_i 代表第 i 组特征层； G_i 是分组卷积的参数，分别取 (1, 4, 8, 16)。最后对所提取的多尺度特征通过 Concat 进行拼接，得到融合之后的特征层，公式为

$$F = \text{Concat}([F_0, F_1, \dots, F_{S-1}]), \quad (2)$$

式中： $F \in \mathbf{R}^{C \times H \times W}$ 。多尺度特征融合过程如图 4 所示。

在得到多尺度特征层之后，由 $Z_i = \text{SEWeight}(F_i)$ 计算出整个多尺度通道注意力权重向量 $Z = Z_0 \oplus Z_1 \oplus \dots \oplus Z_{S-1}$ ，利用 Softmax 来对通道注意力信

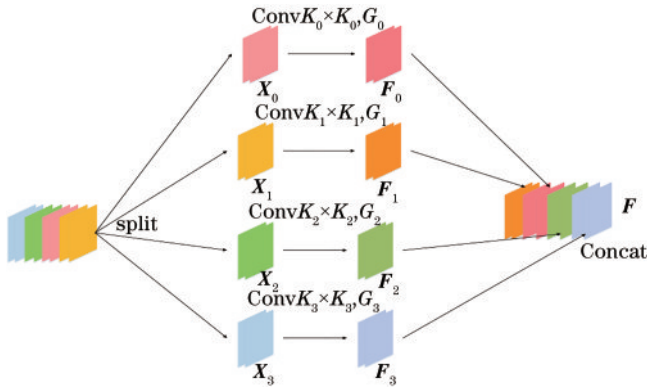


图 4 多尺度特征融合过程

Fig. 4 Multi-scale feature fusion process

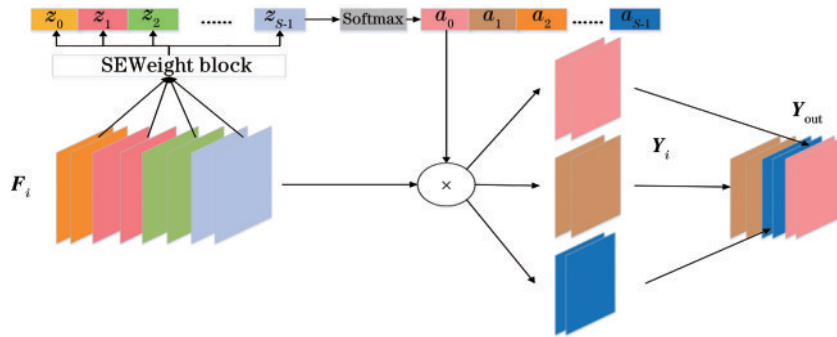


图 5 特征重标注过程

Fig. 5 Feature relabeling process

2.2.3 坐标注意力融入部分

将 O3 路径每一层定义为 x , 对输入 x , 使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核沿着水平坐标方向 X 和垂直坐标方向 Y 对每个通道进行编码。沿两个方向分别池化后进行拼接, 利用一个 1×1 卷积获得注意力特征层:

$$f = \delta \left\{ F_1 \left(\left[z^h, z^w \right] \right) \right\}, \quad (5)$$

式中: $f \in \mathbf{R}^{C/r \times (H+W)}$, r 是控制模块大小的超参数; F_1 是拼接操作。高度为 h 的第 c 个通道及宽度为 w 的第 c 个通道定义为

$$\begin{cases} z_c^w(w) = \frac{1}{H} \sum_{0 < j < H} x_c(j, w) \\ z_c^h(h) = \frac{1}{W} \sum_{0 < i < W} x_c(h, i) \end{cases}. \quad (6)$$

如图 6 所示, 经过激活函数获得非线性信息后, 将输出的结果重新按照水平和垂直方向分为两组特征层:

$$\begin{cases} g^h = \sigma \left[F_h(f^h) \right] \\ g^w = \sigma \left[F_w(f^w) \right] \end{cases} \quad (7)$$

然后对它们分别进行 1×1 卷积后, 利用 Sigmoid 函数获得注意力的权重信息。最终, 将输入的特征层

息进行权重重标注, 表达式为

$$a_i = \text{Softmax}(Z_i) = \frac{\exp Z_i}{\sum_{i=0}^{S-1} \exp Z_i}, \quad (3)$$

式中: $Z_i \in \mathbf{R}^{C \times 1 \times 1}$; $i \in [0, S-1]$ 且为正整数; SEWeight 来源于压缩与激励模块的权重处理部分。然后, 对对应尺度的特征层 F_i 与权重重标注后的注意力向量进行相乘, 计算公式为 $Y_i = F_i \otimes a_i$, $i \in [1, S-1]$ 。最后, 如式 (3) 所示对加权之后的特征层进行等尺度拼接, 输出一个多尺度信息更为丰富的特征层, 表达式为

$$Y_{\text{out}} = \text{Concat}([Y_0, Y_1, \dots, Y_{S-1}]). \quad (4)$$

特征重标注过程如图 5 所示。

信息与水平和垂直权重相乘, 得到坐标注意力特征层:

$$y_c(i, j) = x_c(i, j) \times g_c^w(i) \times g_c^h(j). \quad (8)$$

特征位置信息标注过程如图 6 所示。

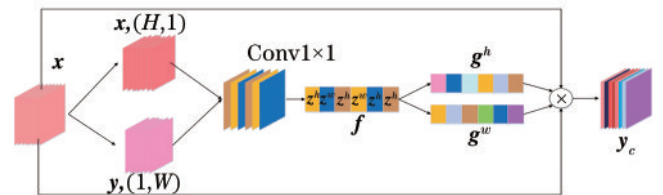


图 6 特征位置信息标注过程

Fig. 6 Feature location information labeling process

3 数据集与评价指标

3.1 数据集

为了验证所提算法的有效性 with 优越性, 在 Kagglelung、Liver、dsb2018 及 Brats18-19 四种数据集上对所提算法进行测试。其中 Kagglelung 数据集共 267 张, 每张图像大小为 512×512 ; Liver 数据集共 420 张, 每张图像大小为 512×512 ; dsb2018 数据集共 546 张, 每张图像大小为 256×256 ; Brats18-19 为 npy 文件, 共 22151 个文件, 图像大小为 512×512 。4 个数据集均配备有专家制作的标签, 总结如表 1 所示。

表 1 数据集的总结
Table 1 Dataset summary

Dataset	Number of images	Size	Segmentation part	Image (Label) format
Kagglelung	267	512×512	Lung	TIF(TIF)
Liver	420	512×512	Liver	PNG(PNG)
dsb2018	546	256×256	Cell	PNG(PNG)
Brats18-19	22151	512×512	Brain tumor	NPY(NPY)

3.2 评价指标

选用交叉熵损失^[14]作为损失函数,定义为

$$l = \frac{1}{N} \sum_j \left\{ -[y_j \cdot \log p_j + (1 - y_j) \cdot \log(1 - p_j)] \right\}, \quad (9)$$

式中: y_j 表示第 j 个样本的标签;目标像素为1,其余为0; p_j 表示第 j 个样本被预测为目标像素的概率。

为了能够直观地评估网络性能,选择6类评价指标对分割结果进行对比分析。准确率(A)是指正确分类的目标像素占整张图像像素的比例。敏感性(S)是指正确分类的非目标像素占真实非目标像素的比例。它们的计算公式分别为

$$A = \frac{N_{TN} + N_{TP}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (10)$$

$$S = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (11)$$

式中: N_{TP} 代表分类正确的目标像素量; N_{TN} 代表分类正确的背景像素量; N_{FP} 代表被错误分类为目标的背景像素量; N_{FN} 代表被错误分类为背景的目标像素量。精确度(P)是在预测结果上进行评价的,指预测为目标像素中真正的目标像素数量,计算公式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}. \quad (12)$$

Dice系数(D)^[15]是一种度量相似度的函数,通常用于计算两个样本的相似度,值的范围为 $[0, 1]$,计算公式为

$$D = \frac{2 \times N_{TP}}{(N_{TP} + N_{FP}) + (N_{TP} + N_{FN})}. \quad (13)$$

平均交并比(MIoU)是语义分割中常用的评估指标,通过计算标签与预测结果的差距交并比来衡量,定义为真正正例(交集)的数量除以真正正例、错误负例、错误正例的并集,计算公式为

$$R_{\text{MIoU}} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k (p_{ji} - p_{ii})}, \quad (14)$$

式中: p_{ij} 表示真实值为 i' ,被预测为 j' 的数量; $k+1$ 是类别个数,包含空类; p_{ii} 是真正的数量; p_{ji} 表示假负的数量。Dice系数对分割出的内部填充比较敏感,而Hausdorff距离(D_H)对分割出的边界比较敏感,度量两个点集间的最大不匹配程度,计算公式为

$$H(\mathbf{A}, \mathbf{B}) = \max[h(\mathbf{A}, \mathbf{B}), h(\mathbf{B}, \mathbf{A})], \quad (15)$$

$$D_H(\mathbf{A}, \mathbf{B}) = \max \left[\max_{\mathbf{a} \in \mathbf{A}} \left(\min_{\mathbf{b} \in \mathbf{B}} \|\mathbf{a} - \mathbf{b}\| \right), \max_{\mathbf{b} \in \mathbf{B}} \left(\min_{\mathbf{a} \in \mathbf{A}} \|\mathbf{b} - \mathbf{a}\| \right) \right], \quad (16)$$

式中: \mathbf{A} 为标签; \mathbf{B} 为分割结果。

4 实验结果和对比分析

4.1 实验配置

实验所用GPU为GEFORCE RTX3090,使用框架Pytorch1.7,在软件Pycharm上进行编译,训练阶段设置epoch为50,batch size为1,训练集与测试集的比例为8:2,初始学习率为 5×10^{-4} 。为了提高效率,设置early stop为6,选择Adam算法^[16]作为学习率调整策略,它利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率。它的优点主要在于经过偏置校正后,每一次迭代学习率都有个确定范围,使得参数比较平稳,定义为

$$\begin{cases} m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \end{cases} \quad (17)$$

$$\begin{cases} \widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \\ \widehat{v}_t = \frac{v_t}{1 - \beta_2^t} \end{cases}, \quad (18)$$

$$\theta_{t+1} = \theta_t - \widehat{m}_t \eta / (\sqrt{\widehat{v}_t} + \epsilon), \quad (19)$$

式中: t 表示次数,当 t 越来越大时, $1 - \beta_1^t$ 和 $1 - \beta_2^t$ 都趋近于1; $\eta / (\sqrt{\widehat{v}_t} + \epsilon)$ 为学习率; $\epsilon = 1 \times 10^{-8}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\eta = 0.001$; β_1 和 β_2 都趋近于1,控制指数衰减 ϵ ,防止分母为0; g_t 表示梯度; m_t 是梯度指数移动均值, \widehat{m}_t 是对 m_t 的纠正; v_t 是由梯度二阶矩求得的平方梯度, \widehat{v}_t 是对 v_t 的纠正。式(17)对梯度和梯度的平方进行滑动平均,每次的更新都和历史值相关。式(18)用来修正滑动平均。式(19)用来更新参数,来源于梯度下降公式。

4.2 实验结果分析及可视化

复现了多个经典网络,与所提方法进行实验对比,结果如表2~5所示。

4.2.1 脑质瘤分割

Brats18-19数据集包含2018年的数据集和2019年的数据集,本实验选用2018年的数据集作为训练集,2019年的数据集作为测试集。由表2各项指标可以看出,所提分割方法在多项指标上均为最优。其中,WT是整个肿瘤,包含ED浮肿区域、NET坏疽和ET增强肿瘤区域;TC是肿瘤的核心,包含ET增强肿瘤区域和NET坏疽。相较于基准网络Unet,所提网络在WT、TC、ET的Dice系数分别提高了0.1142、0.2160、0.1801,代表预测值与真实值更为相似;所提网络在WT、TC、ET的精确度分别提高了0.0210、0.1216、0.0771,代表像

表 2 脑质瘤分割实验结果对比

Table 2 Comparison of experimental results of brain tumor segmentation

Region	Parameter	Unet	ResUnet	DenseUnet	Unet++	ResUnet++	Fcn16s ^[17]	ω -net
WT	<i>D</i>	0.6754	0.7368	0.6713	0.6825	0.7609	0.7326	0.7896
TC		0.6330	0.6864	0.5508	0.7964	0.7513	0.7893	0.8490
ET		0.5145	0.6058	0.4951	0.6033	0.6369	0.6536	0.6946
WT	<i>P</i>	0.8104	0.8641	0.7784	0.9072	0.8723	0.9070	0.8314
TC		0.7779	0.6715	0.6627	0.9354	0.9155	0.8607	0.8995
ET		0.6465	0.8080	0.6332	0.7952	0.7238	0.8240	0.7236
WT	<i>S</i>	0.7469	0.7875	0.7667	0.7084	0.7962	0.7576	0.8326
TC		0.7552	0.9204	0.7620	0.8036	0.7626	0.8687	0.8839
ET		0.6563	0.6105	0.5679	0.6745	0.7406	0.6891	0.7591
WT	<i>D_H</i>	3.44	3.08	3.47	3.06	3.03	2.93	2.98
TC		2.06	2.47	3.05	1.77	2.07	1.93	1.67
ET		3.82	3.32	3.76	3.25	3.39	3.15	3.14
Number of parameters / 10 ⁶		118.42	124.42	155.12	139.74	142.41	512.25	133.81

素分类更为精确;所提网络在 WT、TC、ET 的敏感性分别提高了 0.0857、0.1287、0.1028;所提网络在 WT、TC、ET 的 Hausdorff 距离分别降低了 0.46、0.39、0.68,代表标签与预测结果像素间的差距减小。

由于 Brats18-19 数据集并不是图片格式,所以只

有标签图像用来对比,不同颜色代表肿瘤的不同区域分割情况,其中,绿色部分是水肿区域,黄色部分是增强肿瘤区域,红色部分是坏死。从图 7 不同网络的分割结果可以看出,所提网络能够获得更好的分割结果,对水肿、坏死区域分割结果更为精确。

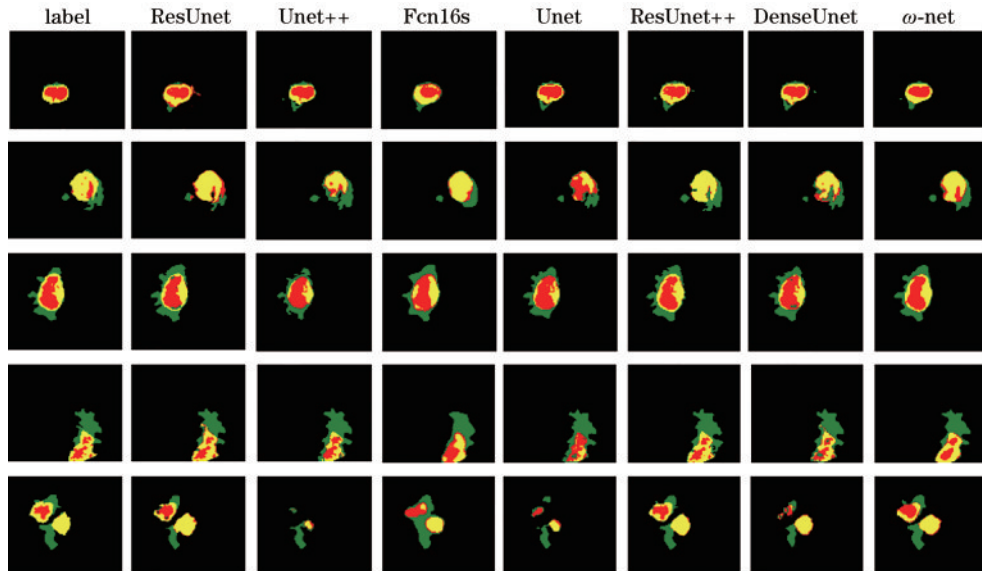


图 7 脑质瘤分割结果可视化

Fig. 7 Visualization of brain tumor segmentation results

4.2.2 细胞分割

通过对表 3 中的数据进行对比可知:所提方法均优于其他方法;与基准网络 Unet 比较, R_{MIoU} 提高了 0.0613, D_H 降低了 0.27, S 提高了 0.0907, A 提高了 0.0201, D 提高了 0.0596, 且参数量较小。

从图 8 不同网络的分割结果可以看出,所提网络可以正确分割出更多的细胞,且像素分类及定位更为准确。

4.2.3 肺部分割

对比表 4 中各项指标的数值可知:所提网络的多项指标均优于其他网络;对比基准网络 Unet, R_{MIoU} 提高了 0.0284, D_H 降低了 0.15, S 提高了 0.0101, P 提高了 0.0083, D 提高了 0.0043, 且参数量较小。

对比图 9 不同网络的分割结果可知, ω -net 能够更为精确地分类像素,相比其他网络,错分类、漏分类的情况较少,同时对气孔等细节的捕获能力较强。

表 3 细胞分割实验结果对比
Table 3 Comparison of cell segmentation experiment results

Network	R_{MIoU}	D_{H}	S	A	P	D	Number of parameters / 10^6
Unet	0.6931	5.23	0.7408	0.9373	0.9080	0.7805	118.42
CENet ^[18]	0.7116	5.20	0.7578	0.9453	0.8998	0.8041	139.73
Unet++	0.7488	5.09	0.8309	0.9553	0.8720	0.8355	139.74
AttUnet	0.7196	5.08	0.7726	0.9475	0.9020	0.8109	133.05
R2Unet	0.7221	6.15	0.6823	0.9248	0.8642	0.6105	149.12
R2AttUnet	0.7410	5.04	0.8177	0.9431	0.8683	0.8212	150.47
DenseUnet	0.6947	5.31	0.7344	0.9309	0.9129	0.7774	155.12
ω -net	0.7544	4.96	0.8315	0.9574	0.8789	0.8401	133.81

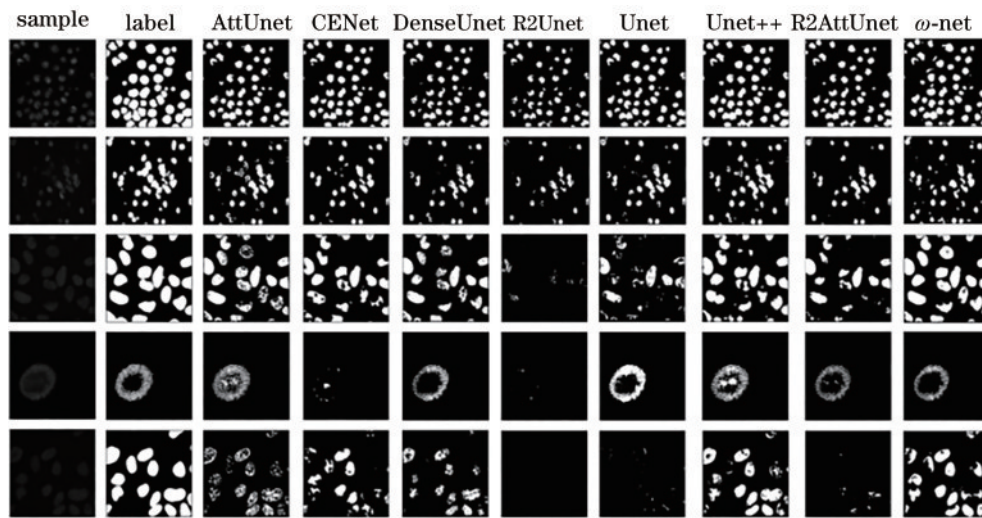


图 8 细胞分割结果可视化

Fig. 8 Visualization of cell nucleus segmentation results

表 4 肺部分割实验结果对比
Table 4 Comparison of lung segmentation experiment results

Network	R_{MIoU}	D_{H}	S	P	D	Number of parameters / 10^6
Unet	0.9507	6.07	0.9718	0.9685	0.9716	118.42
CENet	0.9538	6.03	0.9851	0.9680	0.9736	139.73
Unet++	0.9560	5.93	0.9850	0.9704	0.9748	139.74
AttUnet	0.9549	5.97	0.9809	0.9733	0.9742	133.05
R2Unet	0.9467	6.15	0.9821	0.9637	0.9698	149.12
ω -net	0.9791	5.92	0.9819	0.9768	0.9759	133.81

4.2.4 肝脏分割

对比表 5 中各项指标的数值可知:所提网络的多数指标均优于其他网络;相较于基准网络 Unet, R_{MIoU} 提高了 0.0787, D_{H} 降低了 0.41, S 提高了 0.0447, A 提高了 0.0036, P 提高了 0.0562, D 提高了 0.0457, 且参数量较小。

对比图 10 不同网络的分割结果可知, ω -net 可以更好地分割目标, 并且像素错分类、漏分类的现象较少, 对目标像素的定位也更加准确。

4.3 消融实验

为了展现所提 ω -net 的可靠性以及算法的不同改进部分对分割结果所形成的具体影响, 在 Liver 数据集上进行消融实验, 结果如表 6 所示。

由表 6 可知: 对比 Unet, 二次特征提取网络在选择第三层作为起始层后能够大幅度提升各指标值; 在第一层和第二层上指标下降, 是由于进行上采样后会损失大量的信息, 导致第二次提取特征时, 收获的起始特征数量过少; 而在第四层上, 由于上采样次数不够, 导致第一次未充分提取特征, 第二次特征提取变为第一

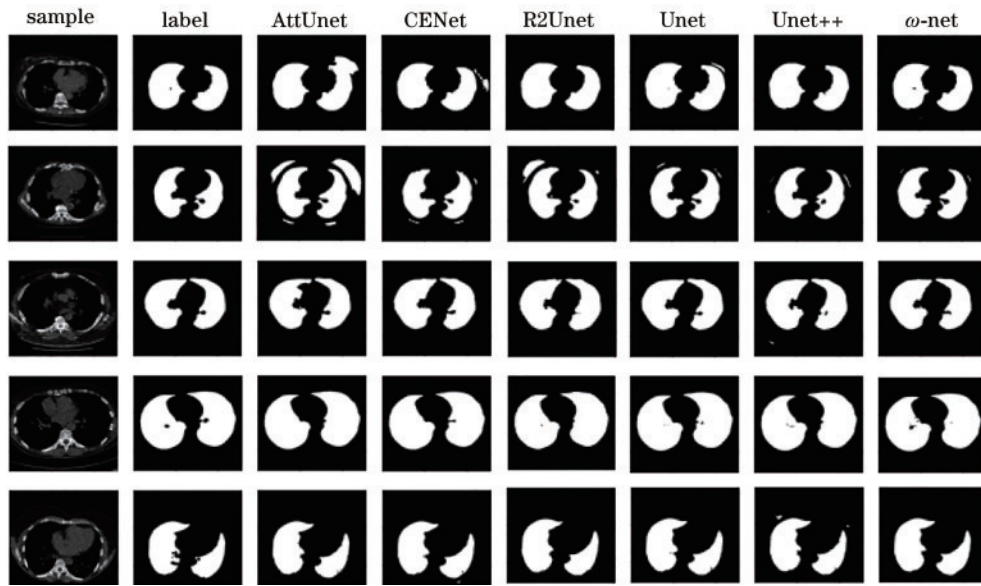


图 9 肺部分割结果可视化

Fig. 9 Visualization of lung segmentation results

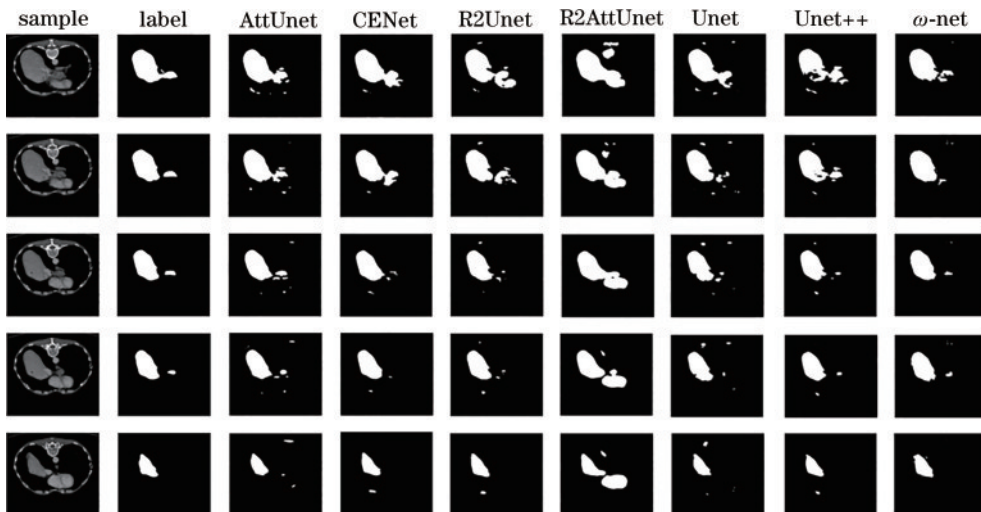


图 10 肝脏分割结果可视化

Fig. 10 Visualization of liver segmentation results

表 5 肝脏分割实验结果对比

Table 5 Comparison of liver segmentation experiment results

Network	R_{MIoU}	D_H	S	A	P	D	Number of parameters / 10^6
Unet	0.8185	6.62	0.8988	0.9865	0.8931	0.9001	118.42
CENet	0.8567	7.14	0.9498	0.9884	0.8990	0.9225	139.73
Unet++	0.8437	6.55	0.9771	0.9855	0.8530	0.9142	139.74
AttUnet	0.8409	7.87	0.9417	0.9883	0.8974	0.9132	133.05
R2Unet	0.7132	8.64	0.9565	0.9541	0.8872	0.8225	149.12
R2AttUnet	0.7558	7.46	0.7072	0.9296	0.9036	0.8454	150.47
ω -net	0.8972	6.21	0.9435	0.9901	0.9493	0.9458	133.81

次特征提取,从而失去应有的结果;由于卷积过多,为了建立长距离通道依赖关系,保证有效特征持续存在,引入 PSA,各项指标得以提升;同时为了增强特征的

几何信息,针对性引入 CA 机制,各项指标再次提升。实验结果表明,所提网络模型具有一定的有效性与合理性。

表 6 消融实验
Table 6 Ablation experiment

Model	R_{MIoU}	D_{H}	S	P	D
Unet	0.8185	6.62	0.8988	0.8931	0.9001
ω -net(first feature map)	0.8043	7.40	0.9001	0.9237	0.8909
ω -net(second feature map)	0.8069	7.94	0.9392	0.8515	0.8926
ω -net(third feature map)	0.8540	6.92	0.9404	0.9250	0.9249
ω -net(fourth feature map)	0.8394	7.32	0.9423	0.8768	0.9126
ω -net(third feature map) +PSA	0.8780	6.92	0.9429	0.9403	0.9235
ω -net(third feature map) +PSA+CA	0.8972	6.21	0.9435	0.9493	0.9458

5 结 论

受 Unet 启发,提出了一种二次特征提取网络 ω -net,用来对肝脏、肺部、脑质瘤及细胞进行分割,并通过引入两种注意力机制对网络进行针对性优化。 ω -net由CA、PSA、Unet及扩展路径组成,在二次特征提取阶段初期针对性地引入PSA,使得四种医学图像建立长距离的通道依赖关系,为后续阶段能够有针对性地捕获特征提供基础。在二次特征提取阶段中期:采用 2×2 卷积替换 2×2 池化,将多余的尺寸信息变为通道信息;对两次提取的特征进行融合,能够有效避免过多池化和卷积造成的信息丧失。在二次特征提取阶段末期,针对性地选用CA机制,每次信息融合后采用该机制赋予通道二次特征属性,避免过多通道带来的信息冗余。最终对比4种数据集下的实验结果,可以得出结论,所提网络模型能够很好地提升各项指标,有更好的泛化性。下一步工作主要是针对分割图像边缘模糊问题进行进一步的优化。

参 考 文 献

- [1] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [2] Xiao X, Lian S, Luo Z M, et al. Weighted res-UNet for high-quality retina vessel segmentation[C]//2018 9th International Conference on Information Technology in Medicine and Education (ITME), October 19-21, 2018, Hangzhou, China. New York: IEEE Press, 2018: 327-331.
- [3] Guan S, Khan A A, Sikdar S, et al. Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal [J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(2): 568-576.
- [4] Ibtehaz N, Rahman M S. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.
- [5] Alom M Z, Hasan M, Yakopcic C, et al. Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation[EB/OL]. (2018-02-20)[2021-05-08]. <https://arxiv.org/abs/1802.06955>.
- [6] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [7] Huang H M, Lin L F, Tong R F, et al. UNet 3: a full-scale connected UNet for medical image segmentation [C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing, May 4-8, 2020, Barcelona, Spain. New York: IEEE Press, 2020: 1055-1059.
- [8] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-net: learning where to look for the pancreas[EB/OL]. (2018-04-11)[2021-08-06]. <https://arxiv.org/abs/1804.03999>.
- [9] Jha D, Riegler M A, Johansen D, et al. DoubleU-net: a deep convolutional neural network for medical image segmentation[C]//2020 IEEE 33rd International Symposium on Computer-Based Medical Systems, July 28-30, 2020, Rochester, MN, USA. New York: IEEE Press, 2020: 558-564.
- [10] Valanarasu J M J, Sindagi V A, Hacihaliloglu I, et al. KiU-net: overcomplete convolutional architectures for biomedical image and volumetric segmentation[J]. IEEE Transactions on Medical Imaging, 2022, 41(4): 965-976.
- [11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [12] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [13] Zhang H, Zu K K, Lu J, et al. EPSANet: an efficient pyramid split attention block on convolutional neural network[EB/OL]. (2021-05-30) [2021-06-05]. <https://arxiv.org/abs/2105.14447>.
- [14] Zhang Z L, Sabuncu M R. Generalized cross entropy loss for training deep neural networks with noisy labels[EB/OL]. (2018-05-20) [2021-08-05]. <https://arxiv.org/abs/1805.07836>.
- [15] Bertels J, Eelbode T, Berman M, et al. Optimizing the dice score and Jaccard index for medical image segmentation: theory and practice[M]//Shen D G, Liu T M, Peters T

- M, et al. Medical image computing and computer assisted intervention-MICCAI 2019. Lecture notes in computer science. Cham: Springer, 2019, 11765: 92-100.
- [16] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2021-06-08]. <https://arxiv.org/abs/1412.6980>.
- [17] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [18] Gu Z W, Cheng J, Fu H Z, et al. CE-net: context encoder network for 2D medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.