

王炳亮, 孙丰杰, 原震. 基于综合注意力机制的卷积神经网络医学图像分割研究[J]. 智能计算机与应用, 2025, 15(5): 211-215. DOI: 10.20169/j.issn.2095-2163.250530

基于综合注意力机制的卷积神经网络医学图像分割研究

王炳亮, 孙丰杰, 原震

(浪潮通信信息系统有限公司 优化分析产品研发部, 济南 250000)

摘要: 深度卷积神经网络被广泛应用于医学图像分割,并展现出良好的性能。本文提出了一种基于综合注意力机制的卷积神经网络来进行医学图像分割,从而提高分割性能。通过与一些流行的深度学习分割方法相比较,实验结果表明本文所提出的方法在分类性能上具有一定的竞争力。在 ISIC2018 数据集上的 *Dice* 和 *IoU* 得分分别为 89.72%, 82.93%, 在 CVC-Clinic-DB 数据集上的 *Dice* 和 *IoU* 得分分别为 89.86%, 82.05%。

关键词: 空间注意力; 通道注意力; 卷积神经网络; 医学图像分割

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2025)05-0211-05

Research on medical image segmentation based on convolutional neural network with integrated attention mechanism

WANG Bingliang, SUN Fengjie, YUAN Zhen

(Optimization and Analysis Product R&D Department, Inspur Communication Information System Co., Ltd., Jinan 250000, China)

Abstract: Deep convolutional neural networks are widely used for medical image segmentation and achieve good performance. This paper proposes a convolutional neural network based on the integrated attention mechanism for medical image segmentation tasks to improve the segmentation performance. By comparing with some popular deep learning segmentation methods, the experimental results show that the method proposed in this paper is competitive in classification performance. The *Dice* and *IoU* scores on the ISIC2018 dataset are 89.72% and 82.93%, and on the CVC-ClinicDB dataset are 89.86% and 82.05%, respectively.

Key words: spatial attention; channel attention; convolutional neural network; medical image segmentation

0 引言

开发自动、准确和鲁棒的医学图像分割方法一直是医学图像的主要问题之一,因其对促进定量病理评估、治疗计划开展和监测疾病进展非常重要。从医学扫描影像中分割出器官或病变位置有助于临床医生做出准确的诊断,减轻医生的工作负担,提高医生的工作效率。然而,一些病变结构的尺度和形状变化较大,很难构建先验模型,如皮肤病变图像^[1]。此外,一些结构在大图背景下可能有较大的位置和方向变化,如磁共振成像中的胎盘^[2-3]。为了获得良好的分割性能,就有必要设计一种分割方法自动识别目标的位置和大小。

目前,深度卷积神经网络(Convolutional Neural

Network, CNN)得到了快速发展,在许多医学图像分割任务中取得了最先进的性能。与传统的分割方法相比, CNN 具有更高的特征表示能力,能够从大量的医学图像数据集中学习到肿瘤特征。然而,当现有的大多数 CNN 方法利用卷积层来提取医学图像特征时,在图像不同的区域中使用共享权值,这将会导致在处理复杂医学图像分割任务时,对肿瘤的不同形状和位置辨识过程中缺乏空间感知能力,从而降低模型的分割性能。

为了应对以上挑战,注意力机制引起了越来越多的研究者的关注。注意力机制可以模仿人类的行为,在特征映射中关注与肿瘤最相关的信息,同时抑制不相关的部分。U-Net 模型通过跳过层将编码器和解码器连接起来,减少了特征提取的误差,在医

作者简介: 孙丰杰(1993—),男,中级工程师,主要研究方向:大数据,网络优化;原震(1985—),男,中级工程师,主要研究方向:大数据,数据治理,数据要素流通。

通信作者: 王炳亮(1985—),男,硕士,中级工程师,主要研究方向:深度学习,数据挖掘,大数据,数据安全。Email: wangbl@inspur.com。

收稿日期: 2023-10-28

哈尔滨工业大学主办 ◆ 科技创新与应用

学图像分割^[4]中取得了可观成效。Oktay 等学者^[5]提出了一种基于注意力的 U-Net 模型,通过抑制输入图像中的不相关区域,同时突出特定局部区域的显著特征,从而提高模型的分割性能。He 等学者^[6]提出了一种集成到 U-Net 模型中的残差学习框架,用于提取更深层次网络的特征。Cao 等学者^[7]提出了一种基于变压器的 U 形编码器结构 Swin-UNet。该结构通过跳变连接将提取的上下文特征与编码器的多尺度特征融合,以补偿下采样造成的空间损失。

对于人工智能系统来说,当应用于医疗诊断时,准确性是非常重要的。在研究工作中,本文利用通道注意力、空间注意力和多尺度注意力来增强网络对肿瘤的关注,以此来提高模型的分割性能。

1 综合注意力卷积神经网络分割模型

1.1 综合注意力卷积神经网络模型架构设计

本文提出了一个综合注意力卷积神经网络模型

用于医学图像分割,主要包括编码器、解码器、跳跃连接、空间注意模型(CA),通道注意模型(SA)和尺度特征注意模型(SFA)。编码器包括 5 个卷积层,通过下采样来提取医学图像中的肿瘤低级语义信息。编码器还包括 5 个反卷积层,用于进一步从编码器提取到的低级信息中提取高级语义信息,并将图像逐步恢复到原图大小。跳跃连接是为了防止在下采样过程中信息的丢失,进而避免由此带来的低级语义信息特征的丢失。空间注意模型可以对医学图像中的肿瘤位置给予重要关注,使得网络能够学习到更多的肿瘤特征。在卷积过程中,并不是所有的通道都可以提取到重要的肿瘤特征信息,本文使用了通道注意模型,以此对不同通道设置不同的关注程度。在面对肿瘤位置和大小多样性的情况下,本文设计了尺度注意模块提取跨尺度的肿瘤特征。综合注意力卷积神经网络模型结构如图 1 所示。

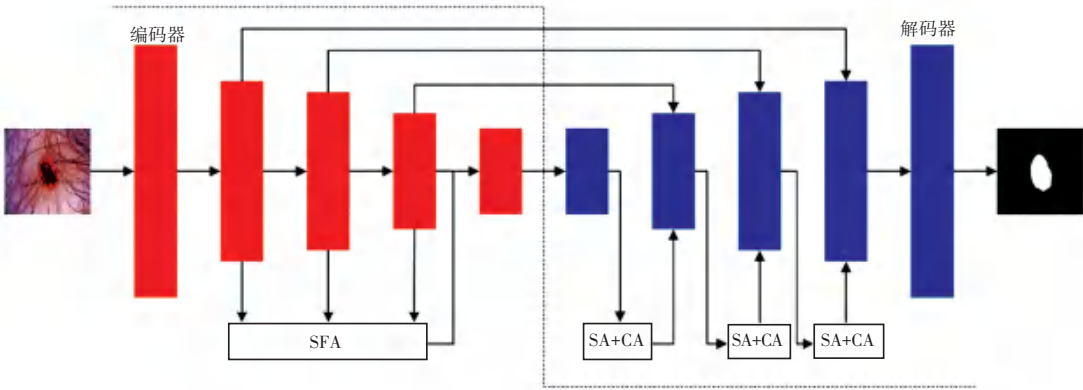


图 1 综合注意力卷积神经网络模型结构

Fig. 1 Integrated attention convolutional neural network model structure

由图 1 可知,本文将医学图像送入到网络中进行特征提取以此来进行肿瘤位置的分割。在编码器中,为了提取到不同尺度的肿瘤特征信息,使得提取到的肿瘤特征更加细节,本文设计了尺度特征注意模型(SFA),模型结构如图 2 所示。对此可以定义为:

$$B_1 = Conv_{3 \times 3}(Conv_{3 \times 3}(Conv_{1 \times 1}(X))) \quad (1)$$
$$B_2 = Conv_{3 \times 3}(Conv_{1 \times 1}(X)) \quad (2)$$
$$B_3 = Conv_{1 \times 1}(X) \quad (3)$$
$$Output_{SFA} = B_1 + B_2 + B_3 \quad (4)$$

其中, $Conv_{3 \times 3}$ 表示卷积核大小为 3 的卷积层; $Conv_{1 \times 1}$ 表示卷积核大小为 1 的卷积层; $Output_{SFA}$ 表示尺度特征注意模型的输出。

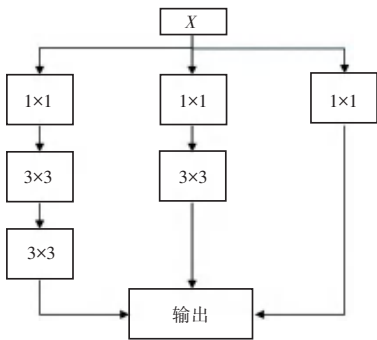


图 2 尺度特征注意模型

Fig. 2 Scale feature attention model

空间注意模型可以对肿瘤区域之间的特征进行增强,以在全局视野中突出肿瘤区域。模型细节如

图3所示。通过特征图之间相乘可以得到一个矩阵,通过 Sigmoid 函数得到权重系数 α ,然后再与输入图像 x 相乘,从而使网络对每个特征图赋予不同的重要程度,其计算公式定义如下:

$$\alpha = \text{Sigmoid}(x \cdot x^T) \quad (5)$$

$$\text{output}_{SA} = x + x \cdot \alpha \quad (6)$$

其中, x 表示输入的特征图; x^T 表示特征图的转置; output_{SA} 表示空间注意模型的输出。

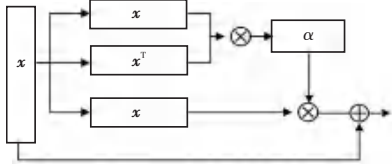


图3 空间注意模型

Fig. 3 Spatial attention model

来自编码器的特征通道大多包含低级信息,而来自解码器的对应通道包含更多的语义信息。因此,对于分割任务可能具有不同的重要性。为了更好地利用最有用的特征通道,研究引入了通道关注来自自动突出相关的特征通道,同时抑制不相关的通道。通道注意模块设计如图4所示。研究给出的数学计算公式如下:

$$\beta = \text{Sigmoid}(\text{MLP}(P_{\max}(X), P_{\text{avg}}(X))) \quad (7)$$

$$\text{output}_{CA} = X + X \cdot \beta \quad (8)$$

其中, X 表示特征图; $P_{\max}$ 表示最大池化; P_{avg} 表示平均池化; MLP 表示多层感知机; output_{CA} 表示通道注意模型的输出。

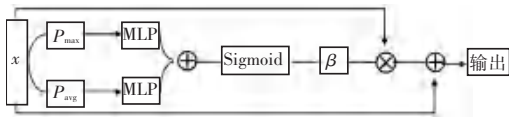


图4 通道注意模型

Fig. 4 Channel attention model

1.2 损失函数

在医学图像分割过程中,为了提高医学图像的分割性能,本文将交叉熵损失函数与 Dice 损失函数相结合作为本文的损失函数。交叉熵损失描述预测值与基础真值之间的距离,而 Dice 损失度量预测值与基础真值之间的一致性程度,数学定义具体如下:

$$\text{Loss}_D = 1 - \frac{2 \sum_{i=1}^n p_i y_i}{\sum_{i=1}^n (p_i + y_i) + \lambda} \quad (9)$$

$$\text{Loss}_B = - \frac{1}{n} \sum_{i=1}^n [y_i \cdot \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (10)$$

$$\text{Loss}_H = \text{Loss}_B + \text{Loss}_D \quad (11)$$

其中, Loss_B 表示交叉熵损失, Loss_D 表示 Dice 损失。

2 实验结果分析

2.1 实验参数设置

所有的方法都是在 Pytorch 框架下实现的。本文采用 Adam 优化算法^[8-11]在 NVIDIA 4090 上进行训练,显存为 24 G,初始学习率设置为 0.000 1, batch_size 设置为 16, $L2$ 正则化系数设置为 $1e-5$,迭代次数为 500 次。

为了评估模型的分割性能,使用以下指标进行评估: Dice 、 IoU 、精确度 (Precision) 和灵敏度 (Sensitivity),可由如下公式计算求出:

$$\text{Dice} = \frac{2 |GT \cap P|}{|GT| + |P|} \quad (12)$$

$$\text{IoU} = \frac{|GT \cap P|}{|GT \cup P|} \quad (13)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (15)$$

其中, GT 表示真实的肿瘤区域; P 表示网络预测的肿瘤区域; TP, FP, FN 分别表示真阳性、假阳性和真阴性。

2.2 数据集

本文实验使用了 2 个公开可用的公共医学图像分割数据集: ISIC2018 和 CVC-ClinicDB。ISIC 2018 数据集^[12]由国际皮肤成像协作组 (ISIC) 发布,是一个大规模的皮肤镜图像数据集,包含 2 594 张图像和标签,并且由专业的皮肤科医生审查和策划过。CVC-ClinicDB 数据集^[13]是 MICCAI 2015 结肠镜视频自动息肉检测子挑战赛训练阶段的官方数据集,该数据由从结肠镜检查视频中提取的 612 张静态图像和标签组成。

本文对 2 种不同的数据集进行预处理,将图像裁剪为 256×256 。并通过均值和标准差进行归一化作为网络的输入,再对图像进行亮度增强、随机水平和垂直翻转来进行图像增强。

2.3 实验结果分析

在本文中,为了评估提出模型的有效性,将分割

结果与 U-Net^[4]、Att-UNet^[6]、ResUNet^[6]、Swin-UNet^[7] 和 UNeXt^[14] 进行了比较。

本文使用了 2 个公开的医学图像公开数据集对所提出的模型进行验证,所有的方法都采用了相同的训练方式。分割结果的定量分析见表 1 和表 2。

根据表 1 和表 2 的结果可以看出,本文提出的方法在几个指标上优于以上方法,这表明了提出方法的有效性。与基线模型 U-Net 相比,本文提出的方法在 ISIC2018 数据集和 CVC-ClinicDB 数据集上都取得了较高的分割性能。与 U-Net 相比,在 ISIC2018 数据集上, *Dice*、*IoU*、*Precision* 和 *Sensitivity* 分别提高了 1.80%、2.41%、3.03% 和 2.46%。在 CVC-ClinicDB 数据集上, *Dice*、*IoU*、*Precision* 和 *Sensitivity* 分别提高了 2.35%、1.80%、1.27% 和 3.38%。这种良好的性能得益于 U-Net 的稳定性和所提模块的有效性,综上分析表明本文的方法分割得更好。

表 1 ISIC 2018 数据集分割结果与其他模型的定量比较
Table 1 Quantitative comparison of ISIC 2018 dataset segmentation results with other models %

模型	<i>Dice</i>	<i>IoU</i>	<i>Precision</i>	<i>Sensitivity</i>
U-Net	87.92	80.52	87.49	88.35
Att-UNet	88.01	80.82	87.64	89.34
ResUNet	88.42	81.25	88.57	89.28
Swin-UNet	88.73	81.96	89.24	89.87
UNeXt	86.19	78.36	88.35	89.65
本文模型	89.72	82.93	90.52	90.81

表 2 CVC-ClinicDB 数据集分割结果与其他模型的定量比较
Table 2 Quantitative comparison of CVC - ClinicDB dataset segmentation results with other models %

模型	<i>Dice</i>	<i>IoU</i>	<i>Precision</i>	<i>Sensitivity</i>
U-Net	87.61	80.25	89.46	87.65
Att-UNet	88.32	81.13	89.23	88.50
ResUNet	88.12	80.82	88.32	87.43
Swin-UNet	88.64	81.03	88.97	89.72
UNeXt	88.83	81.74	89.21	90.21
本文模型	89.86	82.05	90.73	91.03

图 5 和图 6 显示了 ISIC2018 数据集和 CVC-ClinicDB 数据集的分割可视化结果。第 1 行表示原始图像,第 2 行表示原始图像的标签,第 3~8 行分别表示 U-Net, Att-UNet, ResUNet, Swin-UNet 和 UNeXt。在视觉上来看,本文提出的方法在有效识别肿瘤位置的同时取得了良好的分割性能。

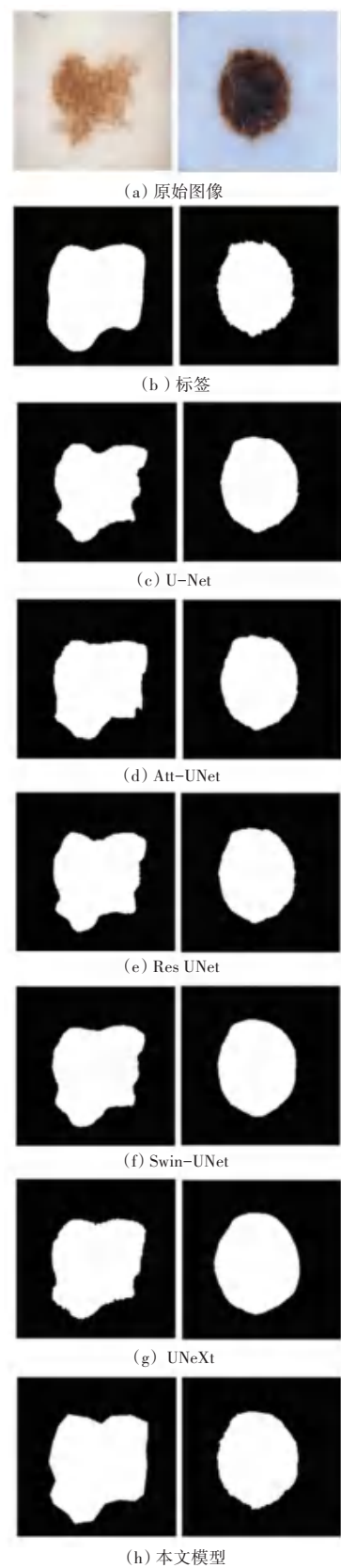


图 5 ISIC2018 数据集的分割视图比较
Fig. 5 Comparison of segmented views of ISIC2018 dataset

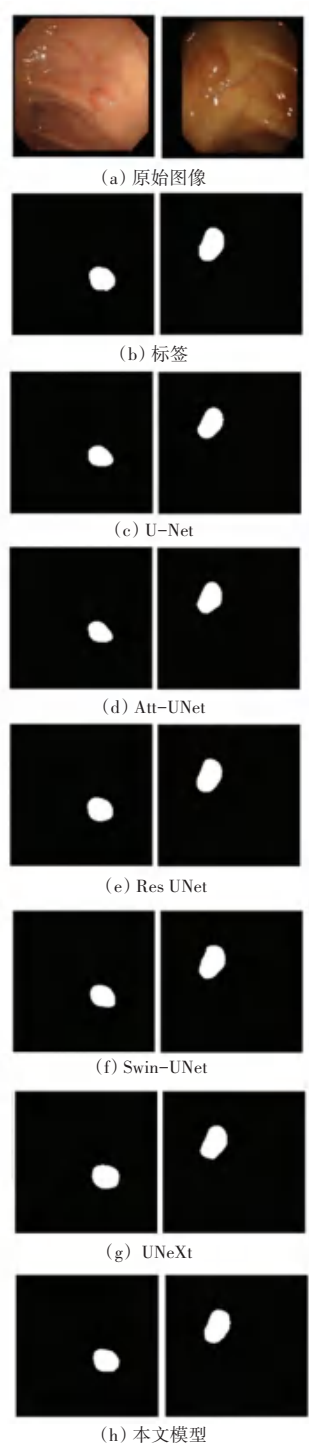


图 6 CVC-ClinicDB 数据集的分割视图比较

Fig. 6 Comparison of segmented views of CVC-ClinicDB dataset

3 结束语

本文提出了一种基于综合注意力机制的卷积神经网络模型用于医学图像分割。实验结果证明,该模型在 ISIC2018 数据集和 CVC-ClinicDB 数据集上取得了较好的分割性能,为医学图像分割的发展提供了一种解决方法。

参考文献

[1] CODELLA N, ROTEMBERG V, TSCHANDL P, et al. Codella et al. Skin lesion analysis toward melanoma detection. 2018: A challenge hosted by the international skin imaging collaboration (ISIC) [J]. arXiv preprint arXiv,1902.03368,2019.

[2] WANG Guotai, ZULUAGA M A, LI Wenqi, et al. DeepIGeoS: A deep interactive geodesic framework for medical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019, 41(7): 1559-1572.

[3] SALEHIS S M, HASHEMI S R, VELASCO-ANNIS C, et al. Real-time automatic fetal brain extraction in fetal MRI by deep learning [C]// 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Piscataway, NJ: IEEE, 2018: 720-724.

[4] 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.

[5] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention unet: Learning where to look for the pancreas[J]. arXiv preprint arXiv,1804.03999,2018.

[6] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for imagerecognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2016: 770-778.

[7] CAO Hu, WANG Yueyue, CHEN J, et al. Swin-UNet: UNet-like pure transformer for medical image segmentation[J]. arXiv preprint arXiv,2105.05537,2021.

[8] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv,1412.6980,2014.

[9] REN Yupeng, HUANG Jisheng, HONG Zhiyou, et al. Image based concrete crack detection in tunnels using deep fully convolutional networks[J]. Construction and Building Materials, 2020, 234:117367.

[10]BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(12):2481-2495.

[11]WANG Junfeng, LIU Fan, YANG Wenjie, et al. Pavement crack detection using attention UNet with multiple sources [C]// Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision. Cham:Springer,2020:664-672.

[12]LEI Bi, KIM J, AHN E, et al. Dermoscopic image segmentation via multistage fully convolutional networks[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(9):2065-2074.

[13]BERNAL J, SNCHEZ F J, FERNNDEZ-ESPARRACH G, et al. WMDOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics,2015, 43:99-111.

[14]VALANARASU J M J, PATEL V M. Unext: Mlp-based rapid medical image segmentation network [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2022: 23-33.