基于改进UNet模型的眼球超声图像分割算法 研究

赵兵

上海理工大学机械工程学院,上海

收稿日期: 2024年10月8日; 录用日期: 2024年11月1日; 发布日期: 2024年11月8日

摘要

在医学图像分割领域,提高分割性能一直是一个具有挑战性的任务。超声图像具有边缘模糊、噪声污染 等缺点,为了解决眼球超声图像分割结果不理想这一难题,本文提出了一种基于UNet的改进分割算法。 首先,本文采用了残差网络(ResNet)结合UNet,有效地解决了模型退化的问题,进一步提高了模型的精 度和泛化能力;其次,在主干特征提取部分引入高效多尺度注意力(EMA)机制,以增强分割模型的特征 表示能力;最后,通过RAVIR数据集进行泛化性实验,证明了所提出模型的泛化能力。实验结果显示, 改进的UNet算法在超声眼球图像数据集上获得的MIoU和Dice的值分别达到了82.5%和82.3%,相比 UNet模型分别提升了1.1%和1.4%,具有更好的医学图像分割效果。

关键词

图像分割,残差网络,UNet,注意力机制

Research on Segmentation Algorithm of Eye Ultrasound Image Based on Improved UNet Model

Bing Zhao

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 8th, 2024; accepted: Nov. 1st, 2024; published: Nov. 8th, 2024

Abstract

In the field of medical image segmentation, improving segmentation performance has been a challenging task. Ultrasound images have disadvantages such as blurred edges and noise pollution; in order to solve the complex problem of unsatisfactory segmentation results of eye ultrasound images, this paper proposes an improved segmentation algorithm based on UNet. Firstly, the residual network (ResNet) combined with UNet is used in this paper to effectively solve the problem of model degradation and further improve the accuracy and generalization ability of the model; secondly, the efficient multi-scale attention (EMA) mechanism is introduced in the central feature extraction part to enhance the feature representation ability of the segmentation model; finally, generalizability experiments are carried out with the RAVIR dataset, which proves the generalization ability of the proposed model's generalization ability. The experimental results show that the improved UNet algorithm achieves 82.5% and 82.3% values of MIoU and Dice on the ultrasound eye image dataset, which are 1.1% and 1.4% higher than the UNet model, and it has better medical image segmentation results.

Keywords

Image Segmentation, Residual Network, UNet, Attention Mechanism

Copyright © 2024 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u>

CC O Open Access

1. 引言

医学图像分割,如 CT 或 MRI 等,是医学图像分析的重要基础。随着智能设备的普及使得居民用眼 过度,视网膜血管形态的变化已被证实与多种临床症状有关,包括眼科疾病甚至全身性疾病。其中老年 性黄斑变性、青光眼和糖尿病视网膜病变是主要的致盲疾病。然而,眼球区域分割过于依赖临床经验丰 富的医生,要对视网膜血管进行客观和定量的分析依赖于精确的血管分割,这既耗时又耗力。综上原因, 越来越多的学者引入深度学习,将 CNN 应用到眼球超声图像的语义分割中,如何将医学图像快速、准确 地从背景中分割出重要特征成为一个重要的研究话题。

近年来,基于深度学习的语义分割算法被广泛使用。Serdar 等人[1]使用 UNet 模型对结肠组织学图像 进行了细胞核分割研究。通过执行多类语义分割,结肠组织学图像通过每个像素都被归类为 6 种类型。 最后精确度达到了 95.23%,交并比(IoU)的值为 48.57%,取得了满意的分割效果。Yin 等人[2]在 UNet 模 型中引入跳跃连接(Skip-Connections)和图像滤波模块,以恢复下采样造成的空间损失和引导图像还原结 构信息。实现了端到端的训练和快速推理,在不影响参数的情况下精确度仍然有 0.7%的提升。Yang 等人 [3]提出了一种结合边缘感知和变换器结构的编码器——解码器网络,称为 EPT-Net。有效解决了卷积运 算的固有定位性在模拟远距离依赖性方面存在局限性。Ling 等人[4]提出了一种单级多任务注意力网络 (MTANet)。首先在分割任务中设计了一个反向加法注意力模块用于融合全局图中的区域和边界搜索;其 次在分类任务中使用了一个注意力瓶颈模块用于图像特征和临床特征的融合。这可以有效地对图像中的 对象进行分类,同时为每个医疗对象生成高质量的分割图像,最终的 MIoU 的值达到了 71.07%。

针对以上研究都获得了不错的精确度,根据眼球超声图像的轮廓、灰度、颜色等特征,面对的仅是 一个二分类的问题,需要从眼底图像中提取精细的血管。本文提出一种基于改进 UNet 的分割模型,用于 获得更为精准的视网膜下血管图像。首先,我们在每个下采样和上采样中引入残差链接(ResNet),通过一 个核为1的卷积操作实现维度匹配,以改善模型的训练效率和综合性能;其次,在编码器上引入一种高 效多尺度注意力机制(EMA) [5]以增强模型捕捉特征的能力;最后,多项实验结果表明改进后的模型满足 医学图像分割任务的需求。

赵兵

2. UNet 网络模型

UNet 网络是 2015 年在 MICCAI 提出的语义分割算法[6],结构如图 1 所示。它基于 FCN 的思想[7], 采用编码器 - 解码器结构。输入图像经过卷积和最大池化,生成五个特征层,之后进行特征融合得到一个有效特征层,最后对每个特征点进行分类,有效解决了生物医学图像问题。



Figure 1. Network structure of UNet 图 1. UNet 网络结构

3. 网络残差模块



Figure 2. Network structure of ResNet 图 2. ResNet 网络结构

ResNet 在图像分类、目标检测、语义分割等多个领域都有广泛应用[8]。在语义分割任务中, ResNet 可以作为编码器提取图像特征,并将特征传递给解码器进行像素级别的语义分割。

ResNet 通过残差模块(Residual Module)和残差连接(Residual Connection)来构建网络,如图 2 所示,这使得它可以训练更深的网络而不会出现梯度消失的问题。在每个残差模块中增加一个跨层连接,让信息可以直接传递到后面的层次,从而保留原始特征,并避免特征逐层消失。其数学表达式为:

$$H(x) = F(x) + x \tag{1}$$

式中 x 表示捷径通道, F(x)表示求和前的特征映射通道, H(x)表示求和后的特征映射通道。使得网络的学习更加顺畅和稳定, 进一步提高了模型的精度和泛化能力。

4. EMA 注意力机制

EMA 是一种新颖的高效多尺度注意力模块,在保留每个通道信息的同时减少计算开销。EMA 模块 通过编码全局信息来重新校准每个并行分支中的通道权重,并通过跨维度交互来捕捉像素级别的关系, 从而提高模型处理特征和增强特征表示的能力。具体结构如图 3 所示。



Figure 3. EMA module network architecture 图 3. EMA 模块网络结构

在图 3 中, "G"表示输入通道被分成的组数。"X Avg Pool"和"Y Avg Pool"分别代表一维水平和垂直的全局池化操作。

为了使神经元巨大的局部感受野使神经元能够收集多尺度的空间信息。因此,EMA采用三个并行路 线来提取分组特征图的注意力权重描述。其中两条平行路线位于1×1分支,第三条路线位于3×3分支。 为了捕获所有通道的依赖关系并减轻计算开销,在1×1分支中采用两个一维全局平均池化操作分别沿两 个空间方向对通道进行编码,并且在3×3分支中仅堆叠单个3×3的卷积以捕获多尺度特征表示。此外, EMA采用跨不同空间维度方向以聚合空间信息的方法,有效地增强了特征的聚合。首先引入了两个张量, 一个是1×1分支的输出,另一个是3×3分支的输出。然后,利用二维全局平均池化对1×1支路的输 出进行全局空间信息编码,在通道特征联合激活机制前,将最小支路的输出直接转化为对应的维度形状 表示为*R*1^{1×C/G}×*R*3^{C/G×HW}。二维全局池化操作公式如下:

$$Z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} X_c(i, j)$$
⁽²⁾

为了提高模型的计算效率,在二维全局平均池化的输出处采用 Softmax 来拟合上述的线性变换。通 过将上述并行处理的输出与矩阵点积运算相乘得到了第一个空间注意力图。此外,同样利用二维全局平 均池化对 3 × 3 分支编码全局空间信息,1 × 1 分支在通道特征联合激活机制前直接转换为相应的维度形 状表示为 *R*₃^{1×C/G}×*R*₁^{C/G×HW}。在此基础上,导出了保留整个精确空间位置信息的第二空间注意力图。最 后,将每组内的输出特征图计算为生成的两个空间注意力权重值的集合,然后使用 Sigmoid 函数捕获像 素级的成对关系,输出特征映射以增强或减弱原始输入特征,从而得到最终输出。

5. 实验结果与分析

本次实验本文以 Windows11 为操作系统, Pytorch 为深度学习框架搭建实验平台,实验环境具体配置 见表 1。

Tabl	e 1. Experimental e	nvironment a	and config	guration
表 1.	实验环境与配置			

	配置
CPU	Intel(R) Core (TM) i7-12700H
GPU	NVIDIA Geforce RTX-3060 6G
深度学习框架	Pytorch1.8.1
编程语言	Python3.9
内存	32G

实验参数主要是 Adam 优化器,初始学习率为 0.01,100 个 epoch, batch 的大小为 2 张。

5.1. 评价指标

本文实验采用 Dice 系数(Dice coefficient)和 MIoU 的指标大小来衡量模型的分割效果。其中, Dice 系数是一种集合相似度度量函数,通常用于计算两个样本的相似度,取值范围在[0,1]:

$$Dice = \frac{2 |X \cap Y|}{|X| + |Y|} \tag{3}$$

式中: |*X*∩*Y*]是 *X*和 *Y*之间的交集; |*X*]和|*Y*]分表表示 *X*和 *Y*的元素的个数,其中,分子的系数为 2。 MIoU 是分别对每个真实标签和预测结果的交并比计算,然后再对所有类别的 IOU 求均值。MIoU 为:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FN + FP + TP}$$

$$\tag{4}$$

式中: k 为类别数; TP、FP 和 FN 分别表示正确识别的正样本数、识别错误的正样本数和识别错误的负 样本数。

5.2. 消融实验

本文实验采用来自荷兰的糖尿病视网膜病变的公开 FIVES 数据集[9], 其包含 20 个训练集和 20 个测 试集。为了验证残差网络模块、EMA 注意力机制模块的有效性, 与 ECA、ELA 注意力机制进行了一系 列的消融实验。实验结果如表 2 所示。

Table 2. Results of the ablation experiment 表 2. 消融实验结果

Model	ResNet	ECA	ELA	EMA	Dice (%)	MioU (%)
0					80.9	81.4
1	\checkmark				81.8	81.9
2		\checkmark			81.3	81.7
3			\checkmark		81.9	82.0
4				\checkmark	82.1	82.3
5	\checkmark			\checkmark	82.3	82.5

根据表 2 所示, Model 0 为未作出改进的 UNet 分割效果。通过引入上述不同的注意力机制可以看出, Dice 和 MIoU 的值均有提升,其中 Model 4 引入 EMA 注意力机制模块提升最大,分别为 1.2%和 0.9%。 最终本文将 Model 5 作为最终模型,在 Model 4 的基础上引入残差网络模块后分割性能提升最大,Dice 和 MIoU 的值分别提升了 1.4%和 1.1%。改进后的 UNet 模型和原始模型的分割效果如图 4 所示。



Figure 4. Comparison of segmentation effect of UNet model before and after improvement 图 4. 改进前后 UNet 模型分割效果对比

5.3. 对比实验

为了验证本文算法的优越性,本文将与图像分割领域主流的算法进行比较,即 PSPNet [10]、SegNet [11]、DeeplabV3 [12]等。采用 MIoU 作为评价指标,实验结果如表 3 所示。

Table 3. Comparison of experimental results 表 3. 对比实验结果			
Model	MioU (%)		
PSPNet	80.2		
SegNet	79.8		
DeeplabV3	80.9		
UNet	81.4		
Ours	82.5		

结果表明,UNet 算法相较于 PSPNet、SegNet、DeeplabV3 等模型,MIoU 的值最高,更适用于医学 图像分割的任务。同时,改进后的 UNet 模型所获得 MIoU 的值均高于其他目前主流的语义分割算法。本 文算法与上述算法相比,MIoU 的值分别提升了 2.3%、2.7%、1.6%和 1.1%。证实了改进算法的优越性和 综合性能。

5.4. 曲线训练分析

为了进一步测试本文改进算法对于眼球超声图像分割性能的提升效果,图 5 展示了训练损失的对比结果。



Figure 5. Comparison of training loss curves of UNet model before and after improvement 图 5. 改进前后 UNet 模型的训练损失曲线对比

从图 5 可以看出,在 25 个 epoch 之前,原始模型在训练集的损失函数有小阶段上升。相比之下,改进的模型训练误差在逐渐减小。在 25 个 epoch 之后,改进模型的曲线表现的更为稳定,加速了网络的收敛。这进一步地证实了 ResNet 和 EMA 模块的有效性。

5.5. 泛化性实验

为了验证本文提出的模型的泛化能力,将 UNet 改进前后的模型在 RAVIR 数据集上[13]进行对比实

验,其包含左眼和右眼的图像。该研究筛选出 200 张超声眼球图像并按照 8:2 的比例划分为训练集和测 试集,实验环境与表 1 相同,实验结果如表 4 所示。本文改进的 Unet 模型算法在 Dice 和 MIoU 的值分 别达到了 77.2%和 80.2%,相比于未改进的 UNet 模型分别提升了 1.9%和 0.8%。两种模型的分割结果如 图 6 所示,可以看出改进后的 UNet 在 RAVIR 数据集上有着更好的分割效果,证明了本文模型的泛化 能力。

Model	Dice (%)	MioU (%)
UNet	75.3	79.4
Ours	77.2	80.2

Table 4. Comparative experiments with the RAVIR dataset 表 4. RAVIR 数据集对比实验



Figure 6. Comparison of segmentation effect of UNet model on DAVIR dataset before and after improvement 图 6. 改进前后的 UNet 模型在 DAVIR 数据集上的分割效果对比

6. 结论

本文针对眼球超声图像边缘模糊导致不易分割的问题,在 UNet 模型的基础上进行了改进。首先在 UNet 的每个编码阶段之后引入一个 ResNet,以保留编码部分不同层丢失的信息。其次在每个卷积下采样 的过程中引入 EMA 注意力机制模块,以减少计算开销的同时保留每个通道的关键信息,提高模型的整 体分割性能。通过消融实验结果表明,本文改进的 UNet 模型 Dice 和 MIoU 的值达到了 82.3%和 82.5%, 这些值相比原始模型提升了 1.4%和 1.1%,验证了本文改进模型提出模块和引入 UNet 的有效性。通过对 比实验结果表明,本文改进的 UNet 模型的分割性能超过了 PSPNet、SegNet、DeeplabV3 等主流的语义 分割算法,评价指标 MIoU 的值提升了 1.1%至 2.7%,验证了本文模型的优越性。为了进一步验证改进算 法的分割能力,采用 RAVIR 数据集进行了泛化性实验,结果改进的 UNet 模型 Dice 和 MIoU 值达到了 77.2%和 80.2%,这表明本文模型卓越的泛化能力。综上实验结果,证实了本文算法在眼球区域血管分割 方面的功效和实用性。因此,本文改进的 UNet 模型能够有效地对眼球区域进行分割,满足医学图像的分 割要求。

参考文献

- Yildiz, S., Memis, A. and Varl, S. (2022) Nuclei Segmentation in Colon Histology Images by Using the Deep CNNs: A U-Net Based Multi-Class Segmentation Analysis. 2022 *Medical Technologies Congress*, Antalya, 31 October-2 November 2022, 1-4. <u>https://doi.org/10.1109/tiptekno56568.2022.9960188</u>
- [2] Yin, P., Yuan, R., Cheng, Y. and Wu, Q. (2020) Deep Guidance Network for Biomedical Image Segmentation. *IEEE Access*, 8, 116106-116116. <u>https://doi.org/10.1109/access.2020.3002835</u>
- [3] Yang, J., Jiao, L., Shang, R., Liu, X., Li, R. and Xu, L. (2023) EPT-Net: Edge Perception Transformer for 3D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **42**, 3229-3243.

赵兵

- [4] Ling, Y., Wang, Y., Dai, W., Yu, J., Liang, P. and Kong, D. (2024) MTANet: Multi-Task Attention Network for Automatic Medical Image Segmentation and Classification. *IEEE Transactions on Medical Imaging*, 43, 674-685. https://doi.org/10.1109/tmi.2023.3317088
- [5] Hu, X., Li, X., Huang, Z., Chen, Q. and Lin, S. (2024) Detecting Tea Tree Pests in Complex Backgrounds Using a Hybrid Architecture Guided by Transformers and Multi-Scale Attention Mechanism. *Journal of the Science of Food and Agriculture*, **104**, 3570-3584. <u>https://doi.org/10.1002/jsfa.13241</u>
- [6] Ronneberger, O., Philipp, F. and Thomas, B. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. <u>http://arxiv.org/abs/1505.04597</u>
- [7] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, 7-12 June 2015, 3431-3440. <u>https://doi.org/10.1109/cvpr.2015.7298965</u>
- [8] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 770-778. <u>https://doi.org/10.1109/cvpr.2016.90</u>
- [9] Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., et al. (2022) FIVES: A Fundus Image Dataset for Artificial Intelligence Based Vessel Segmentation. *Scientific Data*, 9, Article No. 475. https://doi.org/10.1038/s41597-022-01564-3
- [10] Li, Z. and Guo, Y. (2020) Semantic Segmentation of Landslide Images in Nyingchi Region Based on PSPNet Network. 2020 7th International Conference on Information Science and Control Engineering, Changsha, 18-20 December 2020, 1269-1273. <u>https://doi.org/10.1109/icisce50968.2020.00256</u>
- [11] Zheng, X., Zhang, S., Li, X., Li, G. and Li, X. (2021) Lightweight Bridge Crack Detection Method Based on SEGNet and Bottleneck Depth-Separable Convolution with Residuals. *IEEE Access*, 9, 161649-161668. https://doi.org/10.1109/access.2021.3133712
- [12] Cheng, L., Xiong, R., Wu, J., Yan, X., Yang, C., Zhang, Y., et al. (2024) Fast Segmentation Algorithm of USV Accessible Area Based on Attention Fast Deeplabv3. *IEEE Sensors Journal*, 24, 24168-24177. https://doi.org/10.1109/jsen.2024.3410403
- [13] Hatamizadeh, A., Hosseini, H., Patel, N., Choi, J., Pole, C.C., Hoeferlin, C.M., et al. (2022) RAVIR: A Dataset and Methodology for the Semantic Segmentation and Quantitative Analysis of Retinal Arteries and Veins in Infrared Reflectance Imaging. *IEEE Journal of Biomedical and Health Informatics*, 26, 3272-3283. <u>https://doi.org/10.1109/jbhi.2022.3163352</u>