

文章编号: 2096-1618(2023)01-0044-05

# 基于改进 U-Net 的新冠肺炎图像分割方法

杨梅, 罗建, 张晓倩, 金芊芊  
(西华师范大学电子信息工程学院, 四川 南充 637009)

**摘要:**近两年,新冠肺炎在全球暴发,给人类带来了极其严重的生命安全隐患,为尽可能地提高医生的诊断效率,研究新冠肺炎图像的病灶分割方法是极具价值的。利用 U-Net 网络模型作为基础网络,在第一、二层拼接前,引入改进的通道注意力,强化重要信息;在第三、四层拼接前,引入空间注意力,提取空间信息,对目标区域的细节信息进行补充;最后使用混合损失函数,加快网络收敛速度,避免样本不均衡。对比基础网络,改进后的网络模型既能够分割出较大的目标区域,也能够分割出较小的目标区域,更好地避免出现梯度消失问题,捕捉特征更加充分,有效提高分割性能和网络的可靠性。

**关键词:** U-Net; 高斯误差线性单元; 通道注意力; 空间注意力; 混合损失函数

**中图分类号:** TP391.41

**文献标志码:** A

**doi:** 10.16836/j.cnki.jcuit.2023.01.007

## 0 引言

2019 年底,部分地区暴发从未出现过的一种急性呼吸道疾病,世界卫生组织将此疾病命名为新型冠状病毒肺炎(COVID-19)<sup>[1]</sup>。新冠肺炎传染性十分强,传播速度极其迅速,很快就蔓延全球,引发了全世界的高度关注。截至 2022 年 3 月 23 日,世卫组织的数据显示,全球约有 4.7 亿人被确诊为新冠肺炎感染者,并且由于新冠肺炎暂时没有有效的救治措施,导致约 609 万人死亡。新冠肺炎不仅对人类的生命安全构成巨大威胁,还对全球经济产生严重的影响。

核酸检测是筛选 COVID-19 患者的黄金标准<sup>[2]</sup>,但是有研究表明,核酸检测的灵敏度低,约为 71%,需要多次检测才能准确诊断。与核酸检测相比,多项研究表明使用计算机断层扫描(CT)对 COVID-19 筛查更为敏感和有效。即使没有临床症状,胸部 CT 成像对新冠肺炎的检测也更为敏感。但是患者早期肺结节的特征在 CT 图像上并不能明显反映出来,需要经验丰富的医生精准识别并且标记出感染区域。因此,可以通过使用智能分割技术帮助医务人员进行诊断,减轻医务人员的工作强度,为救治患者争取有价值的时间<sup>[3]</sup>。

由于人工智能相关技术越来越先进,图像分割领域更加流行使用深度学习相关知识,并且获得的分割效果普遍比传统处理方法好。特别是后来出现的用于图像处理的卷积神经网络(CNN)<sup>[4]</sup>,为图像特征提取带来了另一种解决方案。2015 年,Long 等<sup>[5]</sup>创造性地

提出了一种在 CNN 基础上改进而来的全卷积神经网络(FCN),把卷积神经网络里面所有的全连接层转换成了卷积层,然后通过对图像在像素级上使用分类操作,避免了产生语义级图像分割的相关问题。Ronneberger 等<sup>[6]</sup>设计出具有编码-解码结构的 U-Net 网络模型,此模型在医学分割处理中表现突出,其中编码器经过卷积以及逐层下采样操作获得图像级别的特征。解码器部分将编码器传来的信号进行卷积以及逐层上采样操作,然后将其映射为对应的二值分割掩模,生成最终分割结果。在研究分割新冠肺炎 CT 图像时,很多学者采用 U-Net 作为基础网络进行改进创新。T Rajamani 等<sup>[7]</sup>设计出一种动态可变形的网络,该网络对 U-Net 进行了改进,在原始的 U-Net 中,加入了 criss-cross 注意力(CCA)模块,取得了比原始 U-Net 更好的分割效果。Zhao 等<sup>[8]</sup>在 U-Net 中采用空洞卷积,扩大了感受野,同时采用双重注意力机制,改进的模型分割性能有所提升。

但是,上述方法在对新冠肺炎 CT 图像进行分割时,容易出现梯度消失、特征未被充分利用等情况,致使最终分割性能比较低。本文使用 U-Net 作为基础网络,在编码部分和解码部分的每层拼接之前加入改进的通道注意力以及空间注意力。其中,本实验将通道注意力的激活函数改为高斯误差线性单元(GELU)激活函数,避免了梯度消失的问题,能更加迅速地更新梯度。同时,在深层部分使用空间注意力,更加突出特征图的空间信息,补充充足的细节信息。最后,使用由交叉熵损失函数以及 Dice 损失函数按比例构成的混合损失函数,既加快了网络的收敛速度,又保证了样本的均衡性。实验结果证明,本实验能更好地避免出现梯

度消失问题,捕捉特征更加充分,分割性能更好,且可靠性较高。

## 1 网络模型

### 1.1 网络模型介绍

本文 U-Net 作为基础网络,同时用改良的通道注意力以及空间注意力进行改进,提高了分割的准确性。U-Net 结构是一种标准的编码器-解码器结构。如图 1 所示,左侧可以视为编码器,右侧则视为解码器。编码器包含 4 个部分,其中每一部分都包括两个卷积层,并且每一个部分都通过 Maxpool 进行下采样。整个编码器模块从上到下一直在减小,逐步缩减特征图的分辨率,使其能更好地获取特征信息。解码器则是逐层恢复分割对象的细节信息以及空间维度,将目标区域准确地分割出来。解码器也包括 4 个部分,每一部分通过双线性插值法进行上采样以便增大分辨率,一直到和输入图像的分辨率几乎一样。同时 U-Net 还采用跳跃连接,即每次上采样后,将经过采样的特征图与编码器传来的具有一样分辨率的特征图进行拼接,以便于解码器能更加妥善地复原目标的相关信息。

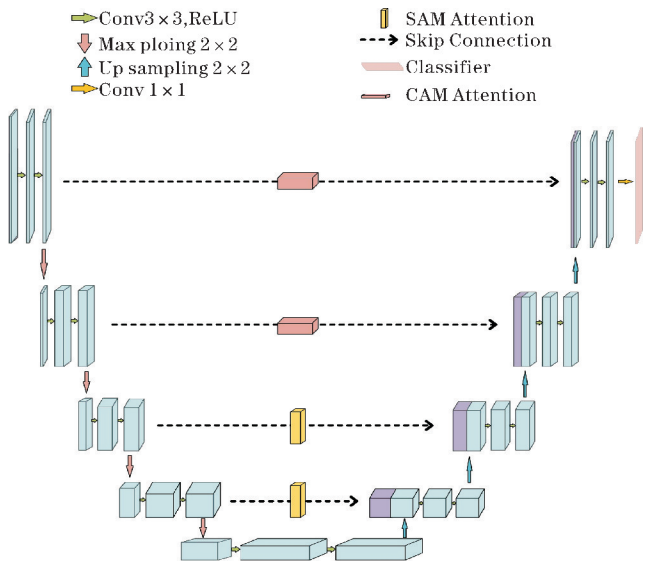


图1 网络结构

对每层编码部分传递的信息,使用注意力机制进行再次处理,以此减少无关区域的特征干扰,降低多余特征的数目。在第一、二层拼接前使用改进的通道注意力,强化特征图中的目标信息,同时抑制非目标信息,为后续特征图的操作提供更多的语义信息;在第三、四层拼接前使用空间注意力,补充感兴趣区域的空间信息,增强目标的细节特征。

### 1.2 激活函数

在神经网络中,能够通过激活函数向模型补充非线性元素,以便神经网络模型能更好地学习数据的特征。神经元的输出受随机正则化和激活函数的影响,但是正则化和激活函数是分开的两个步骤。高斯误差线性单元(GELU)<sup>[9]</sup>,则是把非线性因素和随机正则化结合在一起的激活函数。GELU 的权值取决于当前的输入 input 有多大的概率大于其余的 inputs,当输入的值较大时,权值较大概率赋为 1,保留 input,当输入较小时,权值较大概率为 0, input 被 dropout。

在 GELU 中,假设输入为  $x$ , mask 为  $m$ ,  $m$  服从伯努利分布:

$$\Phi(x) = P(X \leq x) \quad (1)$$

其中  $X$  服从标准正态分布,当输入  $x$  比较小,则输入更容易被 dropout 掉,反之,当输入  $x$  比较大,则输入更容易被保留,所以形成了随机依赖于输入的激活函数。

$$\text{GELU}(x) = xP(X \leq x) = x\Phi(x) \quad (2)$$

公式(2)可以近似替换为

$$\text{GELU}(x) = 0.5x(1 + \tanh(\sqrt{2/\pi}(x + 0.044715x^3))) \quad (3)$$

### 1.3 通道注意力

通道注意力是对各个通道的相关程度进行建模,就是利用深度学习得到各特征通道的重要性。然后,对各个通道给予相应的权重系数,以此加强网络所需要的特征,弱化无关特征。

通道注意力侧重于提取图片中的重要信息,将图片中的重要信息传递给深层的解码部分,以便在特征图中强化重要信息,提高分割的精度。

通道注意力机制<sup>[10]</sup>使大小为  $H \times W \times C$  特征图基于空间维度分别采用平均池化以及最大池化,压缩得到两个  $1 \times 1 \times C$  的特征图,接着将两个特征图分别传入共享的多层感知机,紧跟着把两个处理后的特征叠加,再通过激活函数获取权重系数,见图 2。最后,把权重系数和原来的特征进行相乘操作,就生成了强化后的新特征。其公式如下:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (4)$$

其中  $F$  为输入特征,  $\sigma$  表示 Sigmoid 激活函数。

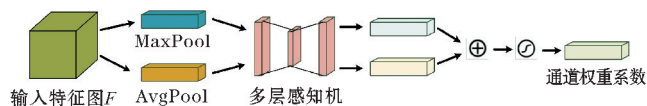


图2 通道注意力示意图

文中,将通道注意力中多层感知机的激活函数由 RELU 改为了 GELU,使通道注意力的非线性化提高,在进行反向传播的期间,能够有利于梯度的变动,有效避免梯度消失问题,提高分割的 Dice 评价指标。

#### 1.4 空间注意力

空间注意力是为了增强感兴趣区域的相关特征表示。实际上就是把原始图像中的空间位置信息经由空间转换模块转化到新的空间中,同时保存重要的信息。对任何一个位置产生一个权重掩模,然后再加权输出,故而可以加强目标区域,削弱无关背景区域。

空间注意力在通道维度中,采用平均池化操作以及最大池化操作来压缩通道。输入特征图为  $H \times W \times C$ , 将其在通道维度上,分别采用最大池化以及平均池化,然后获得  $H \times W \times 1$  的两个特征图,然后在通道维度对两个特征图采用拼接操作,得到  $H \times W \times 2$  的特征图,再经过卷积层将其还原为一个通道。采用  $7 \times 7$  的卷积核,并且保持一样的  $H$ 、 $W$ ,输出  $H \times W \times 1$  的特征图。然后,利用 Sigmoid 函数获得空间权重系数,并且和输入特征图进行乘法操作获得最终的特征图(图3)。

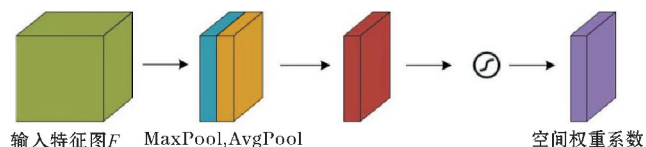


图3 空间注意力示意图

## 2 实验及结果分析

### 2.1 数据集

MosMedData 数据集收集了 1000 多个匿名病人的胸部 CT 数据<sup>[11]</sup>。收集的 CT 数据都包含特殊的标记,按照分类做标记,以便于反馈新冠肺炎患者胸部 CT 的病理异常情况。专家将其中 50 个 CT 扫描集做了分割标注,每张 CT 切片皆标明新冠肺炎独有的毛玻璃混浊以及硬化的区域,并且呈现肺组织异常。将 50 个 CT 扫描集及其 mask 都切为 2050 张  $512 \times 512$  的图片,再对数据图片使用归一化等预处理操作。

### 2.2 损失函数

本文采用混合损失函数,其由交叉熵损失函数以及 Dice 相似系数损失函数构成<sup>[12-13]</sup>,交叉熵损失函数的定义为

$$\text{Loss}_{\text{CE}} = - \sum_{i=1}^N \sum_{j=1}^L ((g_{ij} \lg p_{ij}) + (1 - g_{ij}) \lg (1 - p_{ij})) \quad (5)$$

式中, $N$  代表模型生成的分割预测图中所有的像素点, $L$  代表具有分割标签的掩膜图的所有像素点, $g_{ij}$  是预测图内第  $i$  个像素点以及对应的掩膜图内第  $j$  个像素点的实际类别, $p_{ij}$  为预测图内第  $i$  个像素点以及对应的掩膜图内第  $j$  个像素点的估计值。采用交叉熵损失函数能有效提高网络收敛速度,并且简化网络计算量。

Dice 损失函数的定义为

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2(\sum_{i=1}^N \sum_{j=1}^L g_{ij} p_{ij}) + \varepsilon}{\sum_{i=1}^N \sum_{j=1}^L g_{ij} + \sum_{i=1}^N \sum_{j=1}^L p_{ij} + \varepsilon} \quad (6)$$

式中的  $\varepsilon$  为光滑算子,数值非常小,其作用是为了防止出现分母为 0 的现象。Dice 损失函数能有效平衡样本不均的问题,使网络训练结果更加可靠。

混合损失函数定义为

$$\text{Loss} = \text{Loss}_{\text{CE}} + 0.2 \text{Loss}_{\text{Dice}} \quad (7)$$

本文的混合损失函数,既能加快网络收敛速度,又能保证样本的均衡性,提高了分割实验的效果。

### 2.3 评价指标

本实验使用 Dice 相似系数<sup>[14]</sup>以及 Kappa 系数<sup>[15]</sup>作为评估指标,其中 Dice 相似系数为最重要的评估指标。Dice 系数被用于评价预测值和真实值两者的相似程度,计算公式为

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (8)$$

其中  $A$  代表预测值, $B$  代表真实值。

Kappa 系数一般作为检测模型的一致性的方法,能够评价分类的性能。该系数根据混淆矩阵来进行运算,在  $-1 \sim 1$  取值,一般比 0 大。其公式如式(9)所示, $P_0$  为分类器的准确率, $P_e$  为随机分类器的准确率。Kappa 系数的值更接近 1,证明模型的可靠性更高。

$$\text{Kappa} = \frac{P_0 - P_e}{1 - P_e} \quad (9)$$

### 2.4 参数设置及训练

实验的硬件环境:CPU 为 Intel(R) Xeon(R) Gold 6271C CPU @ 2.60 GHz,4 个 CPU,内存为 100G,GPU 为 Tesla V100,显卡 32G。软件环境:Python 3.7,PyTorch 1.8.0,深度学习框架。将所用的数据集分成 70% 训练集、20% 验证集、10% 测试集。

在训练模型的过程中,学习率采用 PolynomialDecay 策略,初始学习率设置为 0.05,最终学习率为 0,迭代批量设置为 6,训练迭代次数为 40000。用 SGD 优化器进行优化,动量设置为 0.9,权重衰减系数设置为 0.00004。



本文模型在训练时,损失函数值的变化呈现如图 4 所示。可见,当迭代次数增大,网络的损失值慢慢减小。在迭代次数接近 30000 次左右,网络的损失值开始收敛,并且趋于稳定。

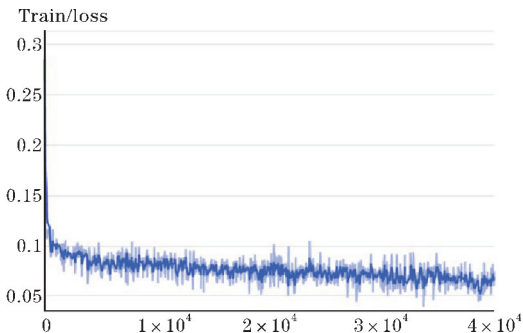


图 4 训练损失曲线

2.5 测试结果

为证明本文的分割性能有所提升,在保证数据预处理以及训练参数一样的情况下,将本文方法和 SegFormer\_B5<sup>[16]</sup>、U-Net 进行对比实验,相关结果如表 1 所示,其中,SegFormer\_B5 是一种基于 Transformer 的网络结构,是目前语义分割模型中,分割性能十分优秀的模型。观察表 1 的数据能够发现,SegFormer\_B5 的表现最差,可能是 SegFormer\_B5 更适合大量数据的训练,而本文中数据量有限。U-Net 的性能比 SegFormer\_B5

好,因为 U-Net 将深层信息与浅层信息拼接,能为后续处理特征图像,提供更加丰富的特征信息,但是来自编码部分的信息中有很多干扰信息,故而最终的分割性能比较差。本文的分割效果最好,在 U-Net 网络模型的编码部分以及解码部分的连接处加入注意力机制,有效地抑制了部分干扰信息,强化了有用的特征信息,便于后续进行分割,提升了分割性能。同时,比较各个模型的 Kappa 系数值,也可以发现本文模型的系数最高,可靠性最好。

表 1 各模型对比结果

模型	Dice	Kappa
SegFormer_B5	0.7520	0.5041
U-Net	0.7779	0.5558
本文	0.8046	0.6092

图 5 为各分割实验的分割结果,其中图 5(d)为真实标签。由图 5(a)可见,SegFormer\_B5 只能分割出较大的目标区域,并且区域分割不完整,边缘信息模糊,不能提取目标的细节信息。从图 5(b)可见,U-Net 也只能分割出较大的感兴趣区域,但是分割形状相对比较完整,细节信息比较丰富。而图 5(c)中,本文既能分割出较大的目标区域,也能分割出较小的目标,分割形状比 U-Net 模型完整,但是细节信息不够丰富,有待提升,总体而言,本文的分割性能好于 SegFormer\_B5、U-Net,并且更加可靠。

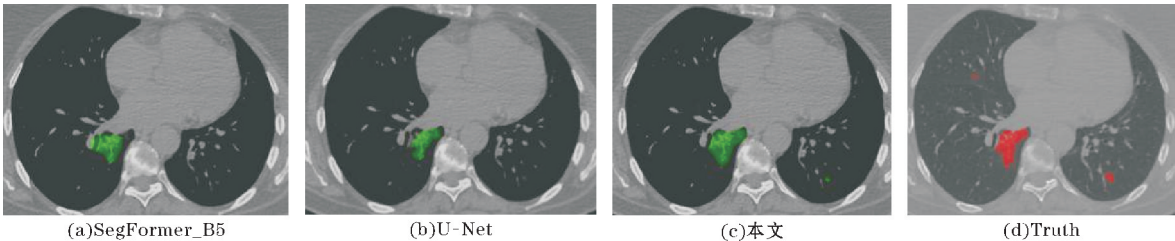


图 5 各分割实验结果对比图

3 结束语

为提高对新冠肺炎患者的诊治效率,使用人工智能技术帮助医生快速定位病灶区域是十分必要的。为改进传统网络结构对病灶区域分割不够完整,忽略较小目标区域等问题,设计了一种新的基于 U-Net 网络的改进方法,在 U-Net 的编码部分和解码部分前两层的连接中间引入改进的通道注意力机制,强化重要信息,抑制干扰信息,然后在后两层的连接中间引入空间注意力机制,提取空间信息,补充目标区域的细节信息。并且本文采用混合损失函数,有效保障了样本的

均衡性,提高了模型的收敛速度。通过对比实验可知,本实验改进的模型分割性能比传统的网络模型更好,可靠性也更高。但是本模型对目标区域的细节信息提取不够完整,导致边缘模糊,同时分割小目标的性能不够好,所以今后将继续研究如何提高细小目标区域的分割性能。

参考文献:

[1] Rt A. Pandemic potential of 2019-nCoV[J]. The Lancet Infectious Diseases,2020,20(3):280-280.  
[2] 李士雪,单莹. 新型冠状病毒肺炎研究进展述评

- [J]. 山东大学学报(医学版),2020,58(3):19-25.
- [3] 宋瑶,刘俊.改进 U-Net 的新冠肺炎图像分割方法[J]. 计算机工程与应用,2021,57(19):243-251.
- [4] Razavian A S, Azizpour H, Sullivan J, et al. CNN Features off-the-shelf: an Astounding Baseline for Recognition[C]. 2014 IEEE conference on computer vision and pattern recognition workshops. IEEE,2014:512-519.
- [5] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,39(4):640-651.
- [6] 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. Springer, Cham,2015:234-241.
- [7] Rajamani K T, Siebert H, Heinrich M P. Dynamic deformable attention network (DDANet) for COVID-19 lesions semantic segmentation[J]. Journal of Biomedical Informatics,2021:103816.
- [8] Zhao X, Zhang P, Song F, et al. D2a u-net: Automatic segmentation of covid-19 lesions from ct slices with dilated convolution and dual attention mechanism [J]. arXiv preprint arXiv,2021(2):05210.
- [9] 席佳妮.智能视觉物联网中基于车辆多属性检索的应用研究[D]. 呼和浩特:内蒙古大学,2021.
- [10] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV),2018:3-19.
- [11] Morozov S P, Andreychenko A E, Pavlov N A, et al. Mosmeddata: Chest ct scans with covid-19 related findings dataset[J]. arXiv preprint arXiv,2020(5):06465.
- [12] 付顺兵,王朝斌,罗建,等.基于改进 U-Net 模型的脑肿瘤 MR 图像分割[J]. 西华师范大学学报(自然科学版),2021,42(2):202-208.
- [13] 莫春梅,周金治,李雪,等.基于改进 U-Net 的肝脏分割方法[J]. 中国医学物理学杂志,2021,38(5):571-577.
- [14] 殷晓航,王永才,李德英.基于 U-Net 结构改进的医学影像分割技术综述[J]. 软件学报,2021,32(2):519-550.
- [15] 李婧.基于深度学习的极化 SAR 图像去噪与分类研究[D]. 阜新:辽宁工程技术大学,2021.
- [16] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems,2021,34:12077-12090.

## A Novel Coronary Pneumonia Image Segmentation Method based on Improved U-Net

YANG Mei, LUO Jian, ZHANG Xiaoqian, JIN Qianqian

( School of Electronic Information Engineering, Xihua Normal University, Nanchong 637009, China)

**Abstract:** In the past two years, the outbreak of new coronary pneumonia in the world has brought extremely serious security risks to human beings. In order to improve the efficiency of doctors in diagnosis as much as possible, it is of great value to study the lesion segmentation method of new coronary pneumonia images. This paper uses the U-Net network model as the basic network to process the pneumonia images, before the first and second layer splicing, the improved channel attention is introduced to strengthen important information; at the same time, before the third and fourth layers are spliced, the spatial attention is introduced to extract the spatial information, and the detailed information of the target area is supplemented. Finally, the mixed loss function is used to accelerate the convergence speed of the network and avoid the sample imbalance. Compared with the basic network, the improved network model can segment both large target areas and small target areas, and avoid the problem of gradient disappearance, and effectively improve the segmentation performance and the reliability of the network.

**Keywords:** U-Net; Gaussian error linear element; channel attention; spatial attention; mixed loss function