



# MSAG-TransNet:肺部 CT 图像中新型冠状病毒感染区域的分割模型

祝鹏烜,黄体仁,李 旭

(浙江理工大学理学院,杭州 310018)

**摘 要:** 针对现有新型冠状病毒感染区域的分割方法存在形态特征提取不充分、感染区域检测不完整以及背景混淆等问题,提出了一种肺部 CT 图像中新型冠状病毒感染区域的分割新模型:MSAG-TransNet 模型。该模型在 U 型网络的基础上增加了多尺度特征抽取模块、Transformer 语义增强模块和多重注意力门模块等 3 个新模块。首先设计了多尺度特征抽取模块来增强骨干网络的特征提取能力,通过多分支结构的深度可分离卷积,充分提取感染区域的形态特征;其次,设计了 Transformer 语义增强模块来捕获图像全局位置信息,整合局部形态特征;最后,设计了多重注意力门模块,将提取的特征与对应上采样过程的门信号拆分成不同分区,然后利用注意力门抑制各分区的无效特征,得到最终分割结果。该模型在两个公开的新型冠状病毒感染 CT 数据集上进行实验,实验结果显示:分割图像的 Dice 系数分别为 82.03%和 76.67%,精确率为 77.27%和 72.34%,交并比为 69.53%和 62.16%;与其他主流模型相比,该模型能够提取更丰富的形态特征,检测到更完整的感染区域,并且得到更精准的分割结果。该模型可以更精确的定位和量化新型冠状病毒感染区域,为临床诊疗提供可靠参考。

**关键词:** 新型冠状病毒感染;CT 图像;图像分割;卷积神经网络;U 型网络;Transformer

**中图分类号:** TP391.4

**文献标志码:** A

**文章编号:** 1673-3851 (2023) 11-0734-11

**引文格式:**祝鹏烜,黄体仁,李旭. MSAG-TransNet:肺部 CT 图像中新型冠状病毒感染区域的分割模型[J]. 浙江理工大学学报(自然科学),2023,49(6):734-744.

**Reference Format:** ZHU Pengxuan, HUANG Tiren, LI Xu. MSAG-TransNet: Segmentation model of COVID-19 infected areas in lung CT images[J]. Journal of Zhejiang Sci-Tech University,2023,49(6):734-744.

## MSAG-TransNet: Segmentation model of COVID-19 infected areas in lung CT images

ZHU Pengxuan, HUANG Tiren, LI Xu

(School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** To solve the problems of insufficient morphological feature extraction, incomplete detection of infected areas and background confusion in the existing segmentation methods for infected areas of COVID-19, a new segmentation model MSAG-TransNet for infected areas of COVID-19 in lung CT images was proposed. This model adds three new modules to the U-shaped network: a multi-scale feature extraction module, a Transformer semantic enhancement module and a multi-attention gate module. Firstly, the multi-scale feature extraction module was designed to enhance the feature extraction ability of the backbone network, and a deep separable convolution of a multi-branch structure was used to fully extract the morphological features of the infected areas. Secondly, the global position information of the image was captured through the designed Transformer semantic enhancement module, and local morphological features were integrated. Finally, the multi-attention gate module was designed to split the extracted features and the gate signal corresponding to the upsampling process into different partitions,

收稿日期:2023-03-03 网络出版日期:2023-06-07

基金项目:浙江省自然科学基金项目(LQ21F030019)

作者简介:祝鹏烜(1993—),男,浙江江山人,硕士研究生,主要从事医学图像分割方面的研究。

通信作者:李 旭,E-mail:lixu0103@163.com

and then attention gates were used to suppress invalid features in each partition to obtain the final segmentation result. The model was tested on two publicly available COVID-19 CT datasets. Experimental results show that the Dice indexes of the segmented image are 82.03% and 76.67%, the accuracy rates are 77.27% and 72.34%, and the intersection and merger ratios are 69.53% and 62.16%, respectively. Compared with other mainstream models, this model can extract richer morphological features, detect more complete infected areas, and obtain more accurate segmentation results. Therefore, this model can more accurately locate and quantify the infected areas of COVID-19, and provide reliable guidance for clinical diagnosis and treatment.

**Key words:** COVID-19; CT image; image segmentation; convolutional neural network; U-shaped network; Transformer

## 0 引 言

新型冠状病毒具有快速的传播能力和超强的感染性,自 2019 年底开始在全球各地不断蔓延,严重冲击国内外的医疗系统,对人类健康造成了极大威胁。据世卫组织统计,全球累计新型冠状病毒感染确诊病例超 7.5 亿、死亡病例超 680 万<sup>[1]</sup>。在疫情高峰期,医疗人员严重短缺,医疗资源挤兑现象时有发生。CT 检查因能观察到磨玻璃影等典型的新型冠状病毒感染表现,成为新型冠状病毒感染的重要影像学检查手段。但是,观察大量肺部 CT 图像并作出诊断是一项繁重的工作,且会因经验不足而导致误诊和漏诊<sup>[2]</sup>。临床上,对感染区域的分割一般由初级医师手动勾画,再由多名经验丰富的放射科医师审核。250 张 CT 图像切片的分割需要大约 400 min<sup>[3]</sup>。利用现有的计算机智能辅助诊断系统可以快速地从肺部 CT 图像中自动分割出新型冠状病毒感染区域,不仅可以有效评估患者感染的严重程度及病变情况,提高诊断的准确率和效率<sup>[4-5]</sup>,而且能够避免医患交叉感染的风险。将性能更好的端到端语义分割模型移植到这些计算机智能辅助诊断系统中并进行算法升级<sup>[6]</sup>,可以提升感染区域分割精度,从而更好地分析肺部状况,制定治疗方案等。因此,设计一种分割结果更准确的模型,有助于医生诊疗新型冠状病毒感染,对疾病诊断、疫情控制和患者治疗等都有重要意义。

传统的图像分割方法有阈值分割、边缘检测等,但是这些方法难以准确分割 CT 图像中灰度差异小或者边界模糊的目标,整体分割效果并不理想。随着深度神经网络技术的蓬勃发展,医学图像分割算法的性能得到了有效提升<sup>[7]</sup>。Ronneberger 等<sup>[8]</sup>提出了基于全卷积神经网络设计的 U-Net 模型,该模型最初应用于细胞分割任务。由于该模型采用了 U 型结构,并使用跳跃连接将对应分辨率的特征信息

引入上采样过程中,可以有效利用高级语义信息和低级特征,为后续分割处理提供不同尺度的信息。受此启发,各种 U-Net 变种模型被广泛应用于医学图像分割任务中,并取得了优异的分割性能<sup>[9]</sup>。Zhou 等<sup>[10]</sup>提出了 UNet++ 模型,通过嵌套不同深度的 U-Net,对跳跃连接进行了改进,并对网络增加了深监督,在肺结节数据集上提升了分割性能。Oktay 等<sup>[11]</sup>提出了 Attention U-Net 模型,设计了注意力门模块,使得模型可以自动学习聚焦目标结构,抑制图像中无关区域的信息。Isensee 等<sup>[12]</sup>提出了自适应深度学习框架 nnU-Net 模型,该模型充分利用了不同数据集的特性,可以自主将相应的基本架构转移到不同数据集和分割任务中,该模型在多项比赛和医学图像分割任务中表现优异。Hatamizadeh 等<sup>[13]</sup>提出了 UNETR 模型,直接将 Transformer 编码器<sup>[14]</sup>提取的特征跳跃连接到解码器中,该模型在多器官数据集上表现出较高的分割性能。Zhang 等<sup>[15]</sup>提出了 TransFuse 模型,该模型可以将 Transformer 编码器分支和卷积编码器分支中的多尺度特征进行有效融合,在息肉分割等任务上取得了较精确的结果。在肺部 CT 图像分割任务中,Chen 等<sup>[16]</sup>利用大量肺部 CT 图像数据训练 UNet++ 模型,使模型的分割准确率达到与放射科医生相当的水平。谢娟英等<sup>[17]</sup>提出了 XR-MSF-Unet 模型,引入了融合多尺度特征的注意力模块,强化了模型的细节分割性能。顾国浩等<sup>[18]</sup>在 U-Net 基础上引入了循环残差模块和自注意力机制,加强了模型对特征信息的抓取,从而使模型在应对新型冠状病毒感染区域与血管相融时具有更高的分割精度。宋瑶等<sup>[19]</sup>采用预训练网络 EfficientNet-B0 进行特征提取,采用 DUpsampling 代替传统上采样,使设计的模型获取更多细节信息,在新型冠状病毒感染公开数据集上提高了分割性能。余后强等<sup>[20]</sup>提出了 PCA-Unet 模型,用 PCA

对肺部 CT 图像中的新型冠状病毒感染区域进行特征预提取,去除了图像中的噪声等因素干扰,使模型能够更准确地分割目标区域。

上述医学图像分割模型在新型冠状病毒感染区域的分割中存在以下问题:首先,由于肺部 CT 图像中新型冠状病毒感染区域形态各异、大小不一且纹理复杂,现有模型存在形态特征提取不充分的问题。其次,新型冠状病毒感染区域通常分散于肺部 CT 图像的不同位置,由于卷积运算感受野较小,长距离空间相关信息的学习能力有限,难以准确定位感染区域。Transformer 编码器可以通过自注意力机制很好地进行全局和长距离语义信息的交互,捕获全局语义信息,但其对局部信息的特征抽取能力较弱。因此仅使用卷积或仅使用 Transformer 编码器不能有效整合全局位置信息和局部形态特征,导致感染区域检测结果不完整。最后,新型冠状病毒感染区域通常表现为磨玻璃影、实变影及斑片状模糊阴影,感染区域可弥漫至整个肺区,且易与肺部气管、血管等混淆<sup>[21-22]</sup>。如果无法有效抑制无关背景的特征信息,那么肺部 CT 图像中的正常区域易被错误检测为感染区域,导致背景混淆,从而降低分割性能。

针对分割任务中感染区域形态特征没有得到充分挖掘、感染区域检测的不完整以及背景混淆等问题,本文提出了一种肺部 CT 图像中新型冠状病毒感染区域的分割模型。本文将该模型命名为 MSAG-TransNet,由每个模块英文名称的字母组合而成。该模型首先设计多尺度特征抽取模块(Multi-scale feature extraction module, MS)来增强骨干网络的特征提取能力,通过多分支结构,利用多尺度深度可分离卷积提取不同形态感染区域的空间信息,使用空间注意力机制实现多尺度信息交互,充分提取感染区域的形态特征。其次,设计了 Transformer 语义增强模块(Transformer semantic enhancement module, Trans),将 CT 图像及特定形态特征拆分成序列特征,并输入 Transformer 编码器捕获全局位置信息,整合局部形态特征,充分利用卷积和 Transformer 的优点。最后,设计了多重注意力门模块(Multi-attention gate module, MAG),将提取的特征与对应上采样过程的门信号拆分成不同分区,利用注意力门抑制各分区的无效特征,从而突出感染区域的特征信息,得到最终分割结果。由于该模型融合了多尺度的形态特征和全局信息,并能抑制无效特征,因此有望提高新型冠状病毒感染区域检测的完整性和背景区分度,进而提升

模型的分割性能。将该模型移植到现有的计算机智能辅助诊断系统中,有望在不改变硬件的情况下提高分割结果的精度,更好地辅助医生诊疗患者。

## 1 本文模型

本文提出的 MSAG-TransNet 模型是一个端到端的语义分割模型。该模型采用编解码器体系,模型结构示意图如图 1 所示。该模型的骨干网络采用 U-Net 的架构方式,保留在医学图像分割任务上表现优异的 U 型结构和跳跃连接。整体网络共分 5 层,每层编码器由 1 个卷积块和 MS 模块组成,每个卷积块由 2 个相连的  $3 \times 3$  卷积和线性整流函数(Rectified linear unit, Relu)组成,其中: $C$  表示卷积块输出通道数。肺部 CT 图像经过 5 个 MS 模块增强的编码器进行 4 次下采样,抽取到尽可能充分的多尺度形态特征。在模型第 1 层和第 5 层设置 Trans 模块,其中第 1 层的输入由肺部 CT 图像和第 1 层编码器抽取的特征堆叠而成。第 1 层的输出分为两部分:一部分与第 5 层编码器抽取的特征进行堆叠,利用 Transformer 编码器进一步学习特征之间的远程依赖关系;另一部分与模型最终的上采样结果相连接。这一设计可以加强网络最低层和最顶层的特征交互,整合全局和局部信息。下采样特征经过 4 次上采样恢复原有分辨率,这样会使得图像信息丢失。为减少有用信息丢失,并使模型有效抑制肺部 CT 图像中的背景信息,在每层跳跃连接中加入本文设计的 MAG 模块。

### 1.1 多尺度特征抽取模块:MS 模块

由于患者感染程度不同,新型冠状病毒感染区域形状多变,大小不一。即使同一感染者,在病程不同时期,其肺部感染区域也在不断变化<sup>[23]</sup>。在感染区域的分割任务中,需要利用各种形态的信息。但是单一尺寸的卷积核对特征抽取的能力不足,在分割时无法有效提取目标不同形态的特征,因此本文设计了 MS 模块,以获得感染区域更丰富的形态特征。

该模块通过多分支结构来实现多尺度特征的抽取,并通过空间注意力机制<sup>[24]</sup>提取 CT 图像中感染区域不同尺度的重要空间信息,最后对不同尺度卷积核抽取的特征进行加权,得到最终的特征图,实现多尺度信息的交互。MS 模块流程如图 2 所示,其中: $O_i$  表示该模块的输出特征图; $F_i$  表示该模块的输入特征图;“DW,  $1 \times 7$ ”表示卷积核大小为  $1 \times 7$  的深度可分离卷积操作,其他类似;“ $1 \times k$ ”“ $k \times 1$ ”“ $2 \times 2$ ”表示对应卷积核大小的常规卷积操作。



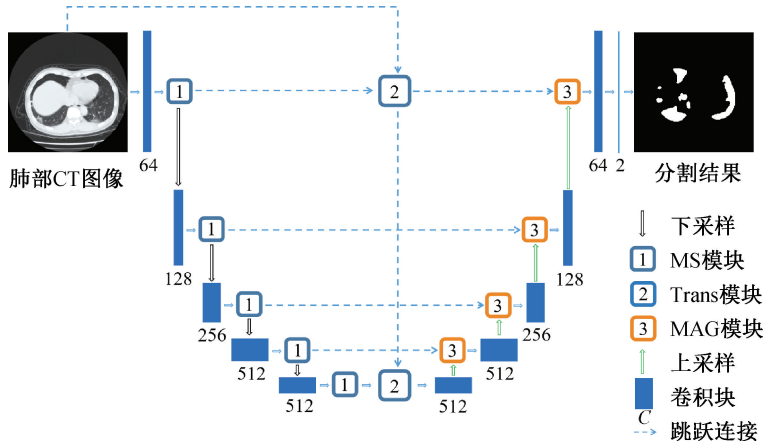


图 1 MSAG-TransNet 模型结构示意图

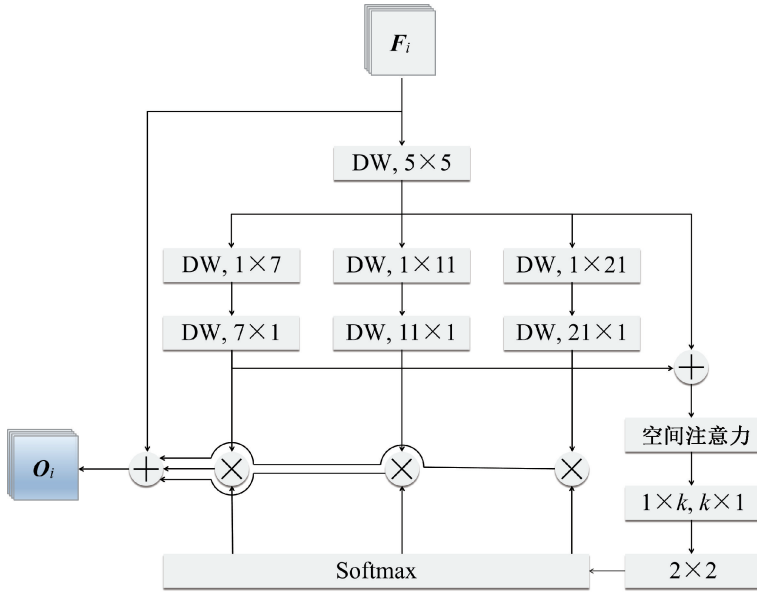


图 2 MS 模块流程图

首先将骨干网络第  $i$  层卷积块提取的特征图  $F_i$  作为输入传入该模块,进行一次深度可分离卷积运算以进一步抽取特征,然后通过一个多分支深度可分离卷积抽取多尺度特征,每个分支的输出特征  $S_j$  的计算公式为:

$$S_j = f_j(f_{DW}(F_i)) \quad (1)$$

其中:  $f_{DW}()$  表示深度可分离卷积操作,卷积核大小为  $5 \times 5$ ;  $f_j()$  表示第  $j$  个分支的计算函数。在每个分支中串联使用两个带状深度可分离卷积来近似不同尺度的大核深度可分离卷积。参考 Guo 等<sup>[25]</sup>提出的多分支模型的卷积核大小设置方法,本文设置 3 个分支,每个分支的卷积核大小分别为 7、11 和 21。使用深度可分离卷积和串联使用  $1 \times k, k \times 1$  的带状卷积来代替一个  $k \times k$  大小的卷积核,以减少模型的参数量,提高程序运行效率。带状卷积核作为传统方形网格卷积核的补充,抽取更多不同形

态的特征。这里的分支数和卷积核大小的设置均可以根据任务不同进行调整组合。

为了有效整合各个分支的信息,本文对不同分支提取的特征  $S_j$  与  $f_{DW}(F_i)$  进行逐元素(Element-wise)相加,随后用卷积建模不同通道间的关系得到中间特征  $S$ ,  $S$  的计算公式为:

$$S = f_{\text{Conl}}(f_{DW}(F_i) + \sum_{j=1}^3 S_j) \quad (2)$$

其中:  $f_{\text{Conl}}()$  表示卷积操作,卷积核大小为  $1 \times 1$ 。然后对  $S$  应用 Woo 等<sup>[24]</sup>提出的空间注意力机制进一步提取出重要的空间信息,得到空间注意力特征  $A$ ,  $A$  的计算公式为:

$$A = f_{\text{Con7}}(f_{\text{Cat}}(f_{\text{Ave}}(S), f_{\text{Max}}(S))) \quad (3)$$

其中:  $f_{\text{Cat}}()$ 、 $f_{\text{Ave}}()$ 、 $f_{\text{Max}}()$  分别表示通道堆叠操作、通道平均池化和通道最大池化;  $f_{\text{Con7}}()$  表示卷积核大小取  $7 \times 7$  的卷积操作。

为了自适应地选择不同分支的权重,参考 Li 等<sup>[26]</sup>提出的“引导特征”(Attention guided feature)方法,进一步对得到的双通道空间注意力特征  $\mathbf{A}$  进行步长为 2 的卷积操作,生成一个单通道空间分辨率减半的引导特征  $\mathbf{s}$ ,降低空间分辨率和通道数以提高效率。随后引导特征  $\mathbf{s}$  经转置卷积生成 3 通道的注意力特征,用于自适应地选择不同空间尺度的信息。这里对注意力特征的通道应用指数函数归一化运算,得到第  $j$  个分支的自适应权重  $w_j$ ,计算过程可用式(4)–(5)表示:

$$\mathbf{s} = f_{\text{ReLU}}(f_{\text{BN}}(f_{\text{Convl}}(f_{\text{Conlk}}(\mathbf{A})))) \quad (4)$$

$$w_j = \frac{\exp(f_{\text{T}}(\mathbf{s})_i)}{\exp(f_{\text{T}}(\mathbf{s})_1) + \exp(f_{\text{T}}(\mathbf{s})_2) + \exp(f_{\text{T}}(\mathbf{s})_3)} \quad (5)$$

其中:  $f_{\text{ReLU}}()$ 、 $f_{\text{BN}}()$  分别表示线性整流激活函数和批量归一化函数;  $f_{\text{Convl}}()$ 、 $f_{\text{Conlk}}()$  分别表示卷积核大小  $k \times 1$ 、 $1 \times k$  的卷积操作;  $f_{\text{T}}()$  表示卷积核大小为 2 的转置卷积操作。

该模块的最终输出特征图  $\mathbf{O}_i$  为输入特征  $\mathbf{F}_i$  与各个不同分支特征的加权求和的残差连接,可用式(6)表示为:

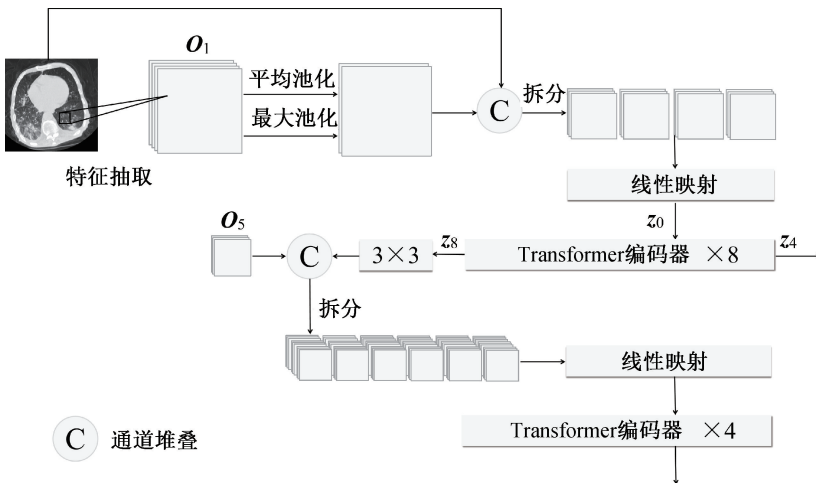
$$\mathbf{O}_i = \mathbf{F}_i + \sum_{j=1}^3 \mathbf{S}_j \times w_j \quad (6)$$

## 1.2 Transformer 语义增强模块: Trans 模块

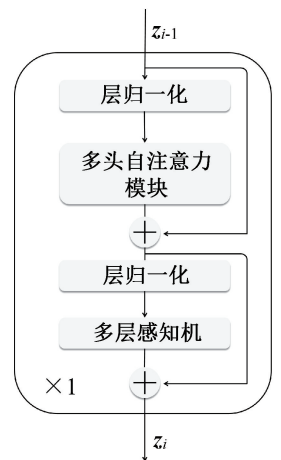
由于肺部 CT 图像中的感染区域分散,因此较难得到位置准确、结构完整的新冠病毒感染区域分割图。本文针对感染区域检测不完整的问题设计了 Trans 模块。该模块通过改进 Dosovitskiy

等<sup>[27]</sup>提出的 ViT 模型,可以有效整合全局位置信息和局部形态特征。ViT 模型将原本应用于自然语言处理领域的 Transformer 推广到了计算机视觉领域,首先将图像拆分变形并重组为序列特征,然后利用线性映射改变序列特征维度,最后把序列特征作为输入,利用 Transformer 编码器来捕获丰富的全局语义信息。但简单地把输入图像分割成几个较大的图像块很难抓取到图像的低层特征,如边和角等局部形态特征。因此将特定卷积提取的形态特征与 Transformer 捕获的全局信息融合,能充分利用卷积和 Transformer 的优点。

本文设计的 Trans 模块流程如图 3(a)所示。首先将模型第 1 层编码器抽取的形态特征  $\mathbf{O}_1$  进行最大池化和平均池化操作,得到最大池化特征和平均池化特征,并将这 2 个特征与原图堆叠成一个 3 通道的输入图像。其次将维度为  $H \times W \times 3$  的输入图像拆分成维度为  $P \times P \times 3$  的小图像块,拆分后的小图像块个数为  $N = H \times W / P^2$ 。所有小图像块经过变形重组得到维度为  $N \times L$  的序列特征,并通过线性映射,将原来的  $N \times L$  维序列特征变成  $N \times D$  维序列特征  $\mathbf{z}_0$ 。然后将  $\mathbf{z}_0$  输入 8 个堆叠的 Transformer 编码器中,其中第 4 个 Transformer 编码器的输出特征  $\mathbf{z}_4$ ,通过 MAG 模块跳跃连接最终的上采样结果。第 8 个 Transformer 编码器的输出特征  $\mathbf{z}_8$ ,进行 1 个  $3 \times 3$  卷积操作后,与最终的下采样特征  $\mathbf{O}_5$  堆叠,并进行拆分和线性映射。最后将该特征传入 4 个堆叠的 Transformer 编码器中,进一步融合全局信息。



(a) Trans 模块流程图



(b) Transformer 编码器流程图

图 3 Trans 模块及 Transformer 编码器流程图

Transformer 编码器流程如图 3(b)所示,其中:  $z_{i-1}$ 、 $z_i$  分别表示每个编码器的输入和输出。

Transformer 编码器由一个多头注意力模块和多层感知机组成,其间采用残差连接,并在输入前进行层

归一化(Layer normalization, LN)。输出  $z_i$  可由式(7)–(9)计算得到:

$$z = f_{LN}(z_{i-1}) \quad (7)$$

$$z_{mid} = f_{MSA}(z) + z_{i-1} \quad (8)$$

$$z_i = f_{MLP}(f_{LN}(z_{mid})) + z_{mid} \quad (9)$$

其中:  $z$ 、 $z_{mid}$  分别表示层归一化序列特征和多头注意力提取的中间特征;  $f_{MSA}()$ 、 $f_{MLP}()$ 、 $f_{LN}()$  分别表示多头注意力、多层感知机和层归一化的函数。

多头自注意力是自注意力机制的一种推广,一个多头自注意力模块包含多个平行的自注意力头,第  $h$  个头的自注意力分数  $A_h$  通过式(10)–(11)计算:

$$\begin{cases} Q_h = zW_{h,1}, \\ K_h = zW_{h,2}, \\ V_h = zW_{h,3} \end{cases} \quad (10)$$

$$A_h = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{D_h}}\right) V_h \quad (11)$$

其中:  $Q_h$ 、 $K_h$ 、 $V_h$  分别表示第  $h$  个自注意力头的询问、键和值;  $W_{h,1}$ 、 $W_{h,2}$  和  $W_{h,3}$  分别是  $Q_h$ 、 $K_h$  和  $V_h$  的参数矩阵;  $D_h$  是一个比例因子,可以防止内积过大。多头注意力的输出可用式(12)表示:

$$f_{MSA}(z) = [A_1; A_2; \dots; A_n] W_{MSA} \quad (12)$$

其中:  $W_{MSA}$  是参数矩阵。Transformer 编码器中的多层感知机包含 2 层线性变换和非线性激活操作,由式(13)计算:

$$f_{MLP}(f_{LN}(z_{mid})) = \sigma(f_{LN}(z_{mid})W_1 + b_1)W_2 + b_2 \quad (13)$$

其中:  $W_1$ 、 $W_2$  分别是第 1、2 层的权重;  $b_1$  和  $b_2$  分别是第 1、2 层的偏置;  $\sigma()$  代表非线性激活函数,在本文中使用高斯误差线性单元激活函数(Gaussian error linear unit, GELU)。

### 1.3 多重注意力门模块:MAG 模块

编码器下采样过程会极大压缩图像的信息,特征图的分辨率会随着下采样过程降低。在解码器部分,通过转置卷积或者上采样操作将这些低分辨率特征图恢复到原始分辨率大小时,需要填补很多空白内容。如果这个过程缺乏足够多的辅助信息,会导致大量有用特征与无关背景特征混杂。采用跳跃连接可以把相应分辨率的特征引入上采样或转置卷积过程,为最后的图像分割处理提供更多的信息。但是将编码器提取的特征图通过跳跃连接与解码器上采样的特征进行简单通道堆叠,会混入干扰分割的无效背景特征。因此,为了进一步突出对分割任务有用的图像信息,抑制无效背景特征,本文采用注意力门机制对特征进行有效融合。

本文对注意力门进行改进,设计了 MAG 模块。将特征图拆分成多个较小的特征块,每个小特征块学习不同的注意力门参数,自适应地利用门信号对不同区块中的像素点进行特征融合,进一步提高模型抑制无效背景特征的能力,从而精细化分割结果。MAG 模块的流程如图 4 所示。其中:  $o_{ij}$  为编码器第  $i$  层抽取的特征  $O_i$  拆分后的小特征块;  $G_i$  为解码器部分第  $i+1$  层特征上采样后的特征,在该模块中作为门信号;  $g_{ij}$  为  $G_i$  拆分后的小特征块;  $U_i$  为该模块第  $i$  层的最终输出。

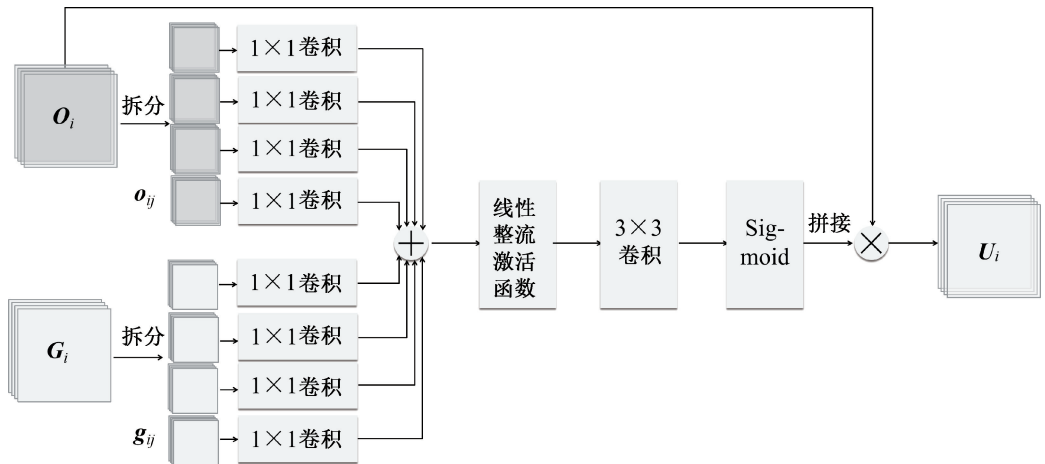


图 4 MAG 模块流程图

首先把  $O_i$  拆分成  $M$  个小特征块  $o_{ij}$ ,  $O_i$  的维度为  $C \times H_o \times W_o$ ,  $o_{ij}$  的维度为  $C \times h_o \times w_o$ 。拆分保持特征通道数不变,拆分后的特征块个数  $M =$

$(H_o/h_o) \times (W_o/w_o)$ 。其中:  $H_o$  和  $W_o$  是  $O_i$  的高和宽;  $h_o$  和  $w_o$  是  $o_{ij}$  的高和宽。门信号  $G_i$  按同样的方式拆分成  $M$  块,每块  $g_{ij}$  与  $o_{ij}$  相对应,由  $g_{ij}$

作为  $\mathbf{o}_{ij}$  的门信号进行后续的操作。拆分后的特征块分别进行不同的  $1 \times 1$  卷积操作,并与线性整流激活函数和 Sigmoid 结合,生成各自的权重图  $\mathbf{y}_{ij}, \mathbf{y}_{ij}$  由式(14)计算:

$$\mathbf{y}_{ij} = \text{Sigmoid}(f_{\text{Con3}}(f_{\text{ReLU}}(f_{\text{Con1}}(\mathbf{o}_{ij}) + f_{\text{Con1}}(\mathbf{g}_{ij})))) \quad (14)$$

其中:  $f_{\text{Con3}}()$  表示卷积核大小为  $3 \times 3$  的卷积操作。将不同的权重图  $\mathbf{y}_{ij}$  按位置拼接,得到与原始特征图大小相同的权重图  $\mathbf{Y}$ ,MAG 模块的最终输出  $\mathbf{U}_i$  由式(15)计算:

$$\mathbf{U}_i = \mathbf{O}_i \times \mathbf{Y} \quad (15)$$

#### 1.4 损失函数

本文模型运用交叉熵(Cross entropy)作为逐像素(pixel-wise)监督的损失函数  $L_{\text{ce}}, L_{\text{ce}}$  可用式(16)计算:

$$L_{\text{ce}} = -[Y_{\text{seg}} \cdot \log P_{\text{seg}} + (1 - Y_{\text{seg}}) \cdot \log(1 - P_{\text{seg}})] \quad (16)$$

其中:  $Y_{\text{seg}}$  和  $P_{\text{seg}}$  分别是实际的标签值和预测为正类的概率值。

## 2 实验结果与分析

### 2.1 数据集及评价指标

#### 2.1.1 实验数据集

本文将 MSAG-TransNet 模型与 7 种主流模型在两个公开的新型冠状病毒感染 CT 数据集上进行对比实验。数据集 1 来源于新型冠状病毒感染 CT 肺部感染分割数据集<sup>[3]</sup>,该数据集由 20 例新型冠状病毒感染患者的 CT 图像以及专家手动勾画的标签图像组成。文件以 NifTi 格式提供,使用 nibabel 包读取。CT 图像的分辨率为  $512 \times 512$ ,原始数据的 CT 图像共有 3520 个切片。标签图像分为 2 类,分别是背景和感染区域。将原始数据中 CT 图像和标签图像转换成 Png 格式,剔除全阴的样本后剩余 1844 个切片,将其按病例划分为训练集和测试集,分别包含 15、5 个病例的切片数据,其中 1482 个切片用于训练,362 个切片用于测试。

数据集 2 来源于网络公开数据<sup>[28]</sup>,该数据集由 199 例新型冠状病毒感染患者的 CT 图像以及专家手动勾画的标签图像组成。CT 图像的分辨率为  $512 \times 512$ ,原始数据的 CT 图像中共有 13705 个切片。将原始数据中 CT 图像和标签图像转换成 Png 格式,剔除全阴的样本后剩余 4981 个切片,将其按病例划分为训练集及测试集,分别包含 160、39 个病例的切片数据,其中 3985 个切片用于训练,996 个

切片用于测试。

训练中对图像进行随机缩放和长宽的扭曲、翻转图像、改变饱和度和明度等数据增强操作。

#### 2.1.2 评价指标

本文采用 Dice 系数、精确率、交并比和 Hausdorff 距离这 4 个广泛使用的指标对实验结果进行量化评价,其计算公式为:

$$D_{\text{sc}}/\% = \frac{2T_p}{T_N + 2T_p + F_p} \times 100 \quad (17)$$

$$P_r/\% = \frac{T_p}{T_p + F_N} \times 100 \quad (18)$$

$$I_{\text{ou}}/\% = \frac{T_p}{F_p + T_p + F_N} \times 100 \quad (19)$$

$$H_D = \max(\max_{p_B \in P_B} \{\min_{q_B \in G_B} \|p_B - q_B\|\}, \max_{q_B \in G_B} \{\min_{p_B \in P_B} \|q_B - p_B\|\}) \quad (20)$$

其中:  $D_{\text{sc}}, P_r, I_{\text{ou}}$  分别表示 Dice 系数、精确率和交并比;  $H_D$  表示 Hausdorff 距离, mm;  $T_p$  表示预测结果是正类,实际是正类的样例;  $F_p$  表示预测结果是正类,实际是负类的样例;  $T_N$  表示预测结果是负类,实际是负类的样例;  $F_N$  表示预测结果是负类,实际是正类的样例;  $P_B, G_B$  分别表示预测结果和实际标签的边界点集;  $p_B, q_B$  分别表示  $P_B$  和  $G_B$  上的像素点的坐标。精确率越高,预测正确的感染区域占有感染区域的比例越高。Dice 系数和交并比指标越高,模型整体分割性能越好。Hausdorff 距离越小,预测结果的边界距离和实际标签的边界越近。

### 2.2 实验配置及参数设置细节

本文模型通过 Pytorch 工具包在一块 NVIDIA GeForce RTX 3060 GPU 上进行训练。为了提高运行效率,将训练图片尺寸变为  $256 \times 256$ , batchsize 设置为 2。本文使用 SGD 优化器进行网络优化,初始学习率设置为  $1 \times 10^{-2}$ ,动量设置为 0.99,权重衰减设置为  $5 \times 10^{-5}$ 。学习率依据“polyLR”<sup>[12]</sup>更新,总共经过 500 轮的迭代训练。其他对比模型除 nnU-Net 外也按上述同样的方法进行训练,nnU-Net 使用其默认的训练策略。本文模型在两个数据集上的训练曲线如图 5 所示,显示了损失函数值随训练轮数增加的变化情况。

### 2.3 实验结果分析

#### 2.3.1 消融实验结果分析

为了证实模型中增加的不同模块的有效性,本文进行如下实验:在原始骨干网络的基础上,先后加



入 MS 模块、Trans 模块和 MAG 模块进行对比。实验结果如表 1 所示。在骨干网络上增加 MS 模块后,Dice 系数、精确率和交并比分别提升了 3.24%、5.01%和 4.37%,各项指标增幅明显,这表明模型提取了更为丰富的形态特征。继续在模型中加入 Trans 模块后,精确率提升 11.25%,这表明模型检测到了更完整的感染区域。由于模型将图像内更多的点预测为感染区域,精确率虽然大幅提升,但模型

同时把更多的背景误分为目标,导致 Dice 系数和交并比仅提升了 0.72%和 1.01%。最后引入的 MAG 模块可以有效抑制无效背景特征,虽然精确率有所下降,但是 Dice 系数和交并比相较于加入了前两个模块的模型进一步提升了 1.23%和 1.75%,模型整体表现更佳。这表明了本文所提出的改进设计可以有效提升分割效果。

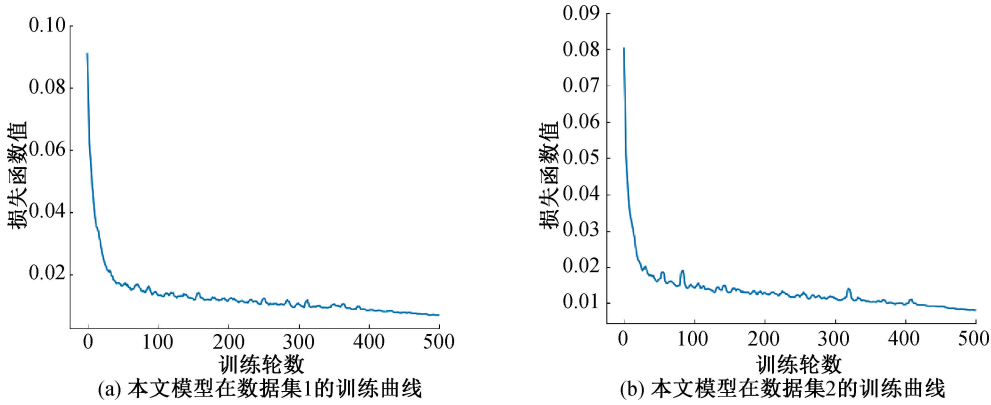


图 5 本文模型在两个数据集的训练曲线

表 1 消融实验评价指标					
MS 模块	Trans 模块	MAG 模块	$D_{SC}/\%$	$P_r/\%$	$I_{oU}/\%$
×	×	×	76.84	69.82	62.40
✓	×	×	80.08	74.83	66.77
✓	✓	×	80.80	86.08	67.78
✓	✓	✓	82.03	77.27	69.53

注:“×”表示该实验不包含此模块,“✓”表示该实验包含此模块。

图 6 展示了消融实验的分割结果。从图 6 中可以看出,增加 MS 模块后得到的分割结果具有更准确的形态细节。继续增加 Trans 模块后,模型将图像中更多的点预测为感染区域,但是检测到更多感染区域的同时也将更多的背景错误分类。最后增加 MAG 模块,误分的无关背景得到了有效抑制。该结果进一步表明本文设计的各个模块有效。

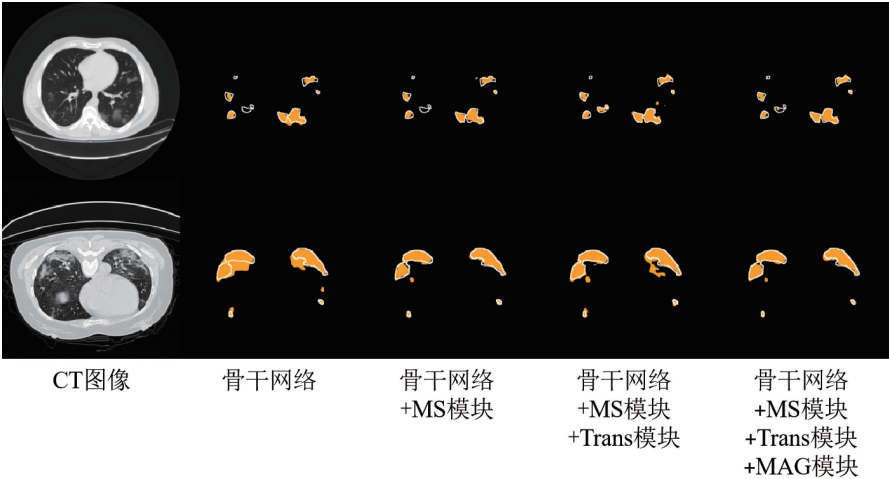


图 6 消融实验的分割结果

注:色块为模型的分割结果,白色线条为实际标签轮廓线。下同。

2.3.2 分割结果的对比分析

为了进一步验证所提出模型的有效性,本文将其在上述公开数据集上与 7 种主流模型进行了比较,这些对比模型包括 U-Net<sup>[8]</sup>、Attention

U-net<sup>[11]</sup>、UNETR<sup>[13]</sup>、nnU-Net<sup>[12]</sup>、UNet++<sup>[10]</sup>、TransFuse<sup>[15]</sup>及 BSNet<sup>[6]</sup>。表 2 和表 3 分别显示了以上不同模型在两个数据集上获得的 Dice 系数、精确率、交并比及 Hausdorff 距离。



表 2 数据集 1 中不同模型的评价指标

模型	$D_{sc}/\%$	$P_r/\%$	$I_{oU}/\%$	$H_D/mm$
U-Net	76.84	69.82	62.40	39.87
Attention U-net	78.19	71.69	64.19	36.72
UNETR	60.79	52.62	43.66	58.93
nnU-Net	76.73	67.94	62.24	38.90
TransFuse	64.13	57.59	47.20	57.89
UNet++	80.11	74.52	66.82	38.13
BSNet	80.65	83.52	67.57	37.83
本文模型	82.03	77.27	69.53	35.77

表 3 数据集 2 中不同模型的评价指标

模型	$D_{sc}/\%$	$P_r/\%$	$I_{oU}/\%$	$H_D/mm$
U-Net	71.83	62.42	56.05	51.80
Attention U-net	74.82	67.80	59.76	42.51
UNETR	68.64	64.24	52.25	63.19
nnUNet	71.77	62.73	55.97	49.77
TransFuse	70.44	68.33	54.37	49.74
UNet++	75.04	68.79	60.05	47.86
BSNet	74.07	76.90	58.82	43.94
本文模型	76.67	72.34	62.16	38.66

实验结果表明,相较于 U-Net,本文模型在两个数据集上的 Dice 系数分别提升了 5.19% 和 4.84%,精确率提升了 7.45%和 9.92%,交并比提升了 7.13%和 6.11%,Hausdorff 距离降低了 4.1 mm 和 13.14 mm。相较于第二优的模型,Dice 系数分别提升了 1.92%和 1.63%,交并比提升了 2.71%和 2.11%,Hausdorff 距离降低了 0.95 mm 和 3.85 mm。这表明本文模型在新型冠状病毒感染区域分割的结果更精准。

图 7 显示了本文模型与其他模型的分割结果。结合图 7、表 2 和表 3 的评价指标结果可知:

a)U-Net 模型采用的 U 型结构和跳跃连接加强了编解码器之间的信息传递,但其使用单一尺度的卷积进行特征提取,无法充分获取图像信息,导致实验结果的分割精确率不高。

b)利用 Attention U-net 模型得到的评价指标结果明显优于 U-Net,表明注意力门机制的引入可以提升模型分割肺部 CT 图像中新型冠状病毒感染区域的性能。

c)利用 UNETR 模型得到的一些分割结果在形态上与实际标签差距大,该模型使用 Transformer 编码器完全替代卷积编码器,虽然提升了模型学习远程空间依赖性的能力,但是会损失很多局部细节信息。

d)nnU-Net 模型使用的网络结构与 U-Net 一

致,在实验中采用其默认的自适应训练方式,利用数据集特性自动设置超参数。由于其没有在本质上解决感染分割任务的问题,该模型得到的评价指标结果相较于 U-Net 没有提升。

e)TransFuse 模型在编码器中以并行方式融合了卷积分支和 Transformer 分支,整合了全局语义信息和局部特征信息,相较于完全使用 Transformer 作编码器的 UNETR 提升了分割性能。由于 Transformer 的归纳偏置相比于卷积更少,需要大量数据才能得到更好的实验结果。在增加了数据量的第二个数据集上,该模型的精确率高于 U-Net 等卷积神经网络模型,但模型整体分割效果仍不理想。

f)利用 UNet++模型进行实验得到的评价指标优于上述模型,该模型嵌套不同深度的 U-Net 来提取不同层次特征,并在网络中加入深监督,有效整合不同层次的特征。但该模型检测的感染区域仍然有所缺失,形态细节仍有不足。

g)BSNet 模型在精确率上优于其他模型,该模型设计了双分支语义增强模块和镜像对称边界引导模块,对不同高级特征之间的互补关系进行建模,从而能够生成更完整的分割结果。但其误分了更多的背景区域,这使得该模型在其他指标上的表现不如本文提出的模型。

h)本文提出的 MS 模块可以提取 CT 图像中感染区域不同尺度的重要空间信息,充分融合多分支信息增强特征提取能力,分割的感染区域形态更接近实际标签。Trans 模块对全局语义信息的提取更加准确,且能有效整合局部特征,可以检测出更完整的感染区域。MAG 模块通过对特征图的拆分精细分配有用特征信息与无效背景特征的权重,有效抑制肺部无关背景信息。与上述主流方法相比,本文所提出的模型能更准确、更完整地定位和分割肺部 CT 图像中新型冠状病毒感染区域。

各个模型的每轮平均训练时间及每张切片的平均测试时间如表 4 所示。每轮训练时间基于数据集 1 的实验计算。本文模型在数据集 1 上的每轮训练时长为 175.85 s,每张切片的测试时间为 0.18 s。虽然训练时间较长,但测试时间与其他模型相差无几,将训练完成的模型移植到计算机智能辅助诊断系统中可以在几乎不影响效率的情况下得到更好的分割结果。

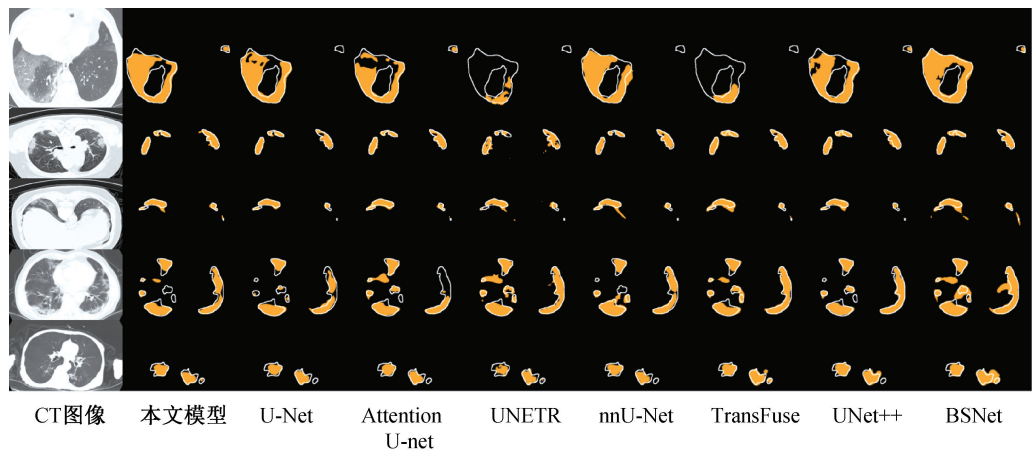


图 7 不同分割模型的分割结果

表 4 各个模型的平均训练时间及平均测试时间

模型	平均训练时间/s	平均测试时间/s
U-Net	72.30	0.11
Attention U-net	72.95	0.12
UNETR	104.35	0.14
nnUNet	155.65	0.22
TransFuse	152.40	0.18
UNet++	143.25	0.17
BSNet	125.20	0.15
本文模型	175.85	0.18

3 结 语

为了提升肺部 CT 图像中新型冠状病毒感染区域的分割性能,本文提出了 MSAG-TransNet 模型。该模型在 U 型网络的基础上增加了 MS 模块、Trans 模块和 MAG 模块。消融实验结果表明,所加的 3 个模块都能提升模型的评价指标。设计的 MS 模块有效整合了不同形态感染区域的空间信息,可以有效增强模型的特征抽取能力;Trans 模块有效整合了全局位置信息和局部形态特征,可以使模型检测到更完整的新型冠状病毒感染区域;MAG 模块可以使模型有效抑制无关背景特征。与 7 种当前主流的分割模型相比,本文模型的各项算法评价指标更优。因此,本文提出的模型能够有效提升肺部 CT 图像中新型冠状病毒感染区域的分割性能,更好地辅助医生诊疗新型冠状病毒感染患者。

参考文献:

[1] World Health Organization. WHO coronavirus(COVID-19) dashboard[EB/OL]. (2023-01-30)[2023-03-03]. <https://covid19.who.int/>.

[2] 左斌,李菲菲. 基于注意力机制和 Inf-Net 的新冠肺炎

图像分割方法[J]. 电子科技, 2023, 36(2): 22-28.

[3] Ma J, Wang Y X, An X L, et al. Toward data-efficient learning; a benchmark for COVID-19 CT lung and infection segmentation[J]. Medical Physics, 2021, 48 (3): 1197-1210.

[4] Zhao W T, Jiang W, Qiu X G. Deep learning for COVID-19 detection based on CT images[J]. Scientific Reports, 2021, 11:14353.

[5] Wu J T, Leung K, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study[J]. The Lancet, 2020, 395 (10225): 689-697.

[6] Cong R M, Zhang Y M, Yang N, et al. Boundary guided semantic learning for real-time COVID-19 lung infection segmentation system[J]. IEEE Transactions on Consumer Electronics, 2022, 68(4): 376-386.

[7] 彭璟,罗浩宇,赵淦森,等. 深度学习下的医学影像分割算法综述[J]. 计算机工程与应用,2021, 57(3): 44-57.

[8] 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.

[9] 王卓英,童基均,蒋路茸,等. 基于 U-Dense-net 网络的 DSA 图像冠状动脉血管分割[J]. 浙江理工大学学报(自然科学版), 2021, 45(3): 390-399.

[10] Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation [C] // International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3-11.

[11] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning where to look for the pancreas[EB/

OL]. (2018-5-20) [2023-03-03]. <https://arxiv.org/abs/1804.03999>.

[12] Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. *Nature Methods*, 2021, 18(2): 203-211.

[13] Hatamizadeh A, Tang Y C, Nath V, et al. UNETR: Transformers for 3D medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA. IEEE, 2022: 1748-1758.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-12-06) [2023-03-03]. <https://arxiv.org/abs/1706.03762>.

[15] Zhang Y D, Liu H Y, Hu Q. TransFuse: Fusing transformers and CNNs for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 14-24.

[16] Chen J, Wu L L, Zhang J, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography[J]. *Scientific Reports*, 2020, 10(1): 19196.

[17] 谢娟英, 张凯云. XR-MSF-Unet: 新冠肺炎肺部 CT 图像自动分割模型[J]. *计算机科学与探索*, 2022, 16(8): 1850-1864.

[18] 顾国浩, 龙英文, 吉明明. U-Net 改进及其在新冠肺炎图像分割的应用[J]. *中国医学物理学杂志*, 2022, 39(8): 1041-1048.

[19] 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.

[20] 余后强, 徐怿璠, 徐静蕾, 等. 基于 PCA 的 Unet 网络用于新冠肺炎 CT 图像分割[J]. *现代信息科技*, 2022, 6(20): 94-97.

[21] 姬广海, 黄满华, 张庆, 等. 新型冠状病毒肺炎 CT 表现及动态变化[J]. *中国医学影像技术*, 2020, 36(2): 242-247.

[22] 许玉环, 吕晓艳, 张见增, 等. 新型冠状病毒肺炎不同临床分型的 CT 特征[J]. *中国医学影像学杂志*, 2020, 28(12): 887-890.

[23] 丁晨宇, 秦立新, 余辉山, 等. 新型冠状病毒肺炎患者的 CT 表现及动态变化特点[J]. *临床放射学杂志*, 2021, 40(1): 55-59.

[24] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[25] Guo M H, Lu C Z, Hou Q, et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation[EB/OL]. (2022-09-18) [2023-03-03]. <https://arxiv.org/abs/2209.08575>.

[26] Li X, Wang W H, Hu X L, et al. Selective kernel networks [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2020: 510-519.

[27] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2023-03-03]. <https://arxiv.org/abs/2010.11929>.

[28] Roth H, Xu Z Y, Tor-Díez C, et al. Rapid artificial intelligence solutions in a pandemic: The COVID-19-20 lung CT lesion segmentation challenge[J]. *Medical Image Analysis*, 2022, 82: 102605.

(责任编辑: 康 锋)