何晓晨, 丁德锐. 基于 Transformer 特征交互的 U 型肺部病灶图像分割网络[J]. 智能计算机与应用,2025,15(10):74-81. DOI:10.20169/j.issn.2095-2163.251011

基于 Transformer 特征交互的 U 型肺部病灶图像分割网络

何晓晨, 丁德锐
(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘 要: 在肺部病灶图像中,病灶区域形状多变、结构信息捕捉困难且对噪声敏感等,这导致分割网络无法学习到肺部病灶固有的形状特征并提取到精确的结构信息,从而造成分割结果的不清晰。针对上述问题,本文提出了一种基于 Transformer 特征交互的 U 型肺部病灶图像分割网络。该网络以 Swin-Transformer 层提取浅层输入的长距依赖特征,有效定位形状和大小差异较大的结构目标,并在一定程度上抑制噪声的影响;设计了特征交互模块建模编-解码层的全局上下文信息,实现了不同层间的交互,丰富了语义特征,有效还原了结构细节;通过引入亚像素卷积来代替传统上采样,有效地处理了低分辨率分割图,提高了图像的细节和清晰度。实验结果表明,文中所提算法具有良好的分割效果,可以生成清晰、准确的分割图像。

关键词: 肺部病灶图像; U 型分割网络; 结构信息; Transformer 层; 特征交互; 全局上下文; 亚像素卷积
中图分类号: TP391.7 **文献标志码**: A **文章编号**: 2095-2163(2025)10-0074-08

A U-shaped-network based on Transformer and feature interaction for lung lesion image segmentation

HE Xiaochen, DING Derui
(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Lung lesion images are usually sensitive to noise, variable shape of the lesion area, and difficult to capture structural information. As a result, the segmentation network cannot learn the inherent shape features of the lung lesion and extract accurate structural information, resulting in unclear segmentation results. To solve these problems, this paper proposes a U-shaped lung lesion image segmentation network based on Transformer feature interaction. The Swin-Transformer layer is firstly used to extract the long-distance dependent features of shallow inputs, and hence effectively locate the target structures with large shapes and size differences, and suppress the influence of noises to a certain extent. The feature interaction module is designed to model the global contextual information of the codec layer, realize the interaction of different layers, enrich the semantic features, and effectively restore the structural details. Subpixel convolution is introduced to replace the traditional sampling method, which can effectively process the segmentation image with low resolution and improve the detail and clarity of the image. The experimental results show that the proposed algorithm has a good segmentation effect and can generate clear and accurate segmentation images.

Key words: lung lesion image; U-shaped segmentation network; structural information; Transformer layer; feature interaction; global context; subpixel convolution

0 引 言

在临床上,计算机断层扫描(CT)图像常用于对患者的肺部病灶进行诊断和评估。通过准确分割出肺部病灶区域,可以获取病灶的位置、形状、大小等关键信息,从而为医生做出后续诊断和制定诊疗方案提供重要的辅助[1]。肺部病灶的准确分割对于肺部疾病的早期发现、定量评估和治疗监测具有重要意义。然而,肺部病变的多样性以及图像采集过程中噪声的影响,则对准确分割肺部病灶图像造成了不小的挑战[2]。

在医学图像分割领域,以 UNet[3] 及其变体为代表的编解码网络是最常用的方法。然而由于卷积操作的窗口大小有限,就只能对局部区域进行感知和

基金项目: 国家自然科学基金(61973219)。
作者简介: 何晓晨(1999—),女,硕士研究生,主要研究方向:深度学习,图像处理。Email:hexiaochen0810@163.com; 丁德锐(1981—),男,博士,教授,博士生导师,主要研究方向:深度学习,图像处理,智能算法研究。
收稿日期: 2024-01-31

特征提取；当图像中的目标物体在尺度上发生变化时，将可能无法准确地进行分割。针对这些问题，基于注意力机制和特征金字塔的方法近年来备受关注。注意力机制可以帮助模型在提取特征时，根据上下文信息动态调整不同位置的重要性权重，有助于模型更好地理解特征的语义结构。特征金字塔结构可以通过在不同尺度上对图像进行特征表示，就能有效捕捉不同尺度物体的特征[4]。以上方法虽然提高了分割精度，但是传统的注意力机制，例如 SE[5]、ECA[6]、CBAM[7]通常是基于局部邻域的关联性来计算注意力权重，限制了模型对全局上下文信息的建模能力；特征金字塔结构，例如 PSP[8]、AHSP[9]、CE[10]在不同尺度上提取特征会导致特征冗余的问题。

针对肺部病灶图像的特点、传统注意力机制的局限性以及特征金字塔结构的冗余性，本文提出了一种新的肺部病灶分割网络，主要的工作和创新点如下所示：

(1)在 U 型网络结构的基础上，提出一种基于 Transformer 特征交互的肺部病灶分割网络，有效地关注全局上下文信息，捕捉目标的结构特征。

(2)在编码器的低层添加 Swin-Transformr 层，实现长距远程依赖特征的提取，抑制噪声，有效定位复杂多变的肺部病灶。

(3)不同于一般的跳跃连接，在编解码不同层之间设计特征交互模块，实现编码层和解码层不同尺度信息的特征交互，并且对注意力模块的输出特征信息进行修正，调整注意力的错误偏差，改善图像的结构细节。

(4)在解码器的上采样部分，不同于传统的双线性上采样以及反卷积上采样，通过引入亚像素卷积，实现对缩小后特征图的放大处理。

1　本文方法

本文致力于提高病变阴影区域模糊不规则、结构不清晰的肺部病灶图像的处理能力。因此，提出的模型不仅需要学习到全局的形状分布特征，还要能实现对于该类图像结构细节的恢复。为此，本文设计了一种基于 Transformer 特征交互的 U 型肺部病灶图像分割网络(A Transformer and Feature interaction-based U-shaped network for lung lesion image segmentation, TF-Unet)，设计结构如图 1 所示。

因为 Unext[11]网络提出的 Tok-MLP 块可以有效地标记和投射卷积特征，且该网络参数数量少、计算复杂度低，所以本文也采用了 Unext 型的网络构架。具体地，输入的肺部病灶 CT 图像在编码器的前 2 个阶段分别添加 Swin-Transformer[12]层，用于弥补卷积窗口局限性导致的全局形状分布信息的丢失，并在一定程度上降低噪声带来的不良影响；在 Tok-MLP 块中，分别跨越宽度和高度移动卷积特征，创建跨越宽度和高度的随机窗口，并在窗口内计算注意力，进而向编码器高层阶段得到的全局特征中引入更多的局域特征，丰富语义信息；然后，构建特征交互模块对编解码不同尺度层的信息进行关注与修正，从而恢复更多的病灶结构细节；最后，在解码器的每个上采样阶段采用亚像素卷积[13]操作，从而对图像进行有效放大，实现病灶区域的准确分割。

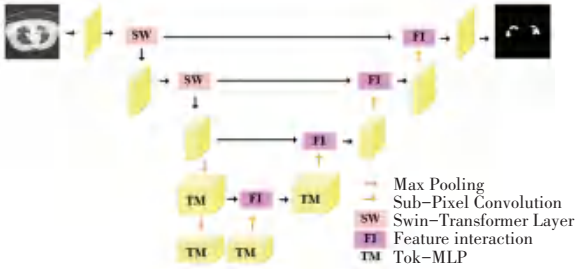

图 1　基于 Transformer 特征交互的 U 型肺部病灶图像分割网络

Fig. 1　U-shaped lung lesion image segmentation network based on Transformer and feature interaction

1.1　Swin-Transformer 层

由于 U 型网络的卷积操作对输入图像进行特征提取时，卷积核的大小是固定的，导致低层阶段编码的信息较为局部。当出现病变阴影大于卷积窗口覆盖面积时，就会难以捕捉到病灶区域的形状分布特点。此外，医疗设备采集的 CT 图像不可避免地会伴随一定的噪声，从而影响网络的分割精度。

相较于卷积操作只是学习并更新卷积核窗口内的参数，Swin-Transformer 则是采用全局自注意力机制方法，对整张图像上每个像素点之间计算注意力权重，因此该方法能够有效地学习整张图像上病灶特征的内在关联，捕捉到整个病灶区域的形状分布特征。为此，本文在低层编码阶段添加了 SWin-Transformer 层，即移位窗口 Transformer 层具体设计结构如图 2 所示。当 Transfomer 由语言任务应用到视觉任务时，图像中像素的高分辨率导致了计算复杂度剧增，移位窗口的方法则将自注意力的计算限制在不重叠的局部窗口，同时允许跨窗口连接，从而提高了 Transformer 在视觉任务的应用效率。首先，Swin-Transformer 层中第一个部分使用一个规则的

窗口划分策略(W-MSA),从左上角像素开始,将 8×8 的特征图均匀划分为 2×2 个、大小为 4×4 的窗口;然后,第二个部分采用来自前一层移位的窗口配置(SW-MSA),窗口从左上角分别向右侧和下方偏移 2 个像素,W-MSA 和 SW-MSA 的窗口划分如图 3 所示。移位窗口配置方法引入了前一层相邻非重叠窗口之间的联系。

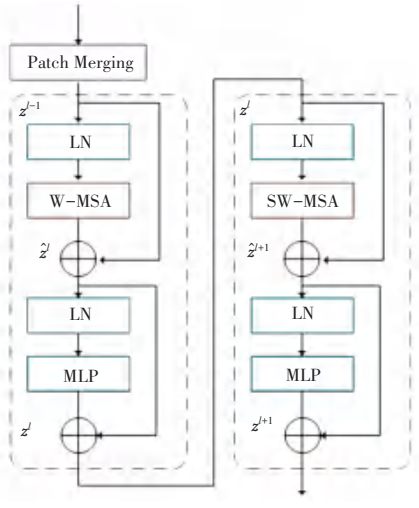

图 2 Swin-Transformer 层

Fig. 2 Swin-Transformer layer

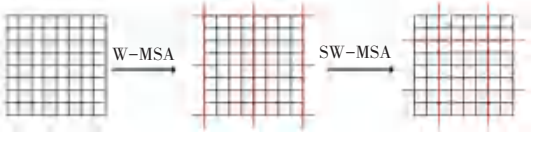

图 3 窗口划分方法

Fig. 3 Window partitioning method

使用移位窗口划分的方法,Swin-Transformer 层的计算规则如下:

$$\begin{cases} \hat{z}^{l} = \mathrm{W-MSA}(\mathrm{LN}(z^{l-1})) + z^{l-1} \\ z^{l} = \mathrm{MLP}(\mathrm{LN}(\hat{z}^{l})) + \hat{z}^{l} \\ \hat{z}^{l+1} = \mathrm{SW-MSA}(\mathrm{LN}(z^{l})) + z^{l} \\ z^{l+1} = \mathrm{MLP}(\mathrm{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \end{cases} \tag{1}$$

其中,W-MSA 表示常规窗口分区配置的多头自注意力;SW-MSA 表示移位窗口分区配置的多头自注意力;$\hat{z}^{l}$ 表示(S)W-MSA 操作的输出特征;z^{l} 表示 MLP 操作的输出特征。

1.2 特征交互模块

肺部病灶形状分布不规则,致使其结构信息的捕捉较为困难。传统的 U 型网络在处理图像分割任务时,只是在编码的阶段中添加注意力机制来突出病灶信息的权重,却忽略了解码阶段在逐步恢复图像细节过程中更加需要关注病灶区域的细节特征并进行细微调整,因此只能得到粗略的病灶分割图。类似于人类的视觉特性,当关注一个实体时,首先会从全局角度定位潜在目标物体,然后不断调整视线以关注该实体的不同部位;而特征交互模块(FI)的基本思想是先利用轴向注意力机制来对全局信息进行关注,准确定位病灶结构信息,获得全局注意力权重图,然后通过对注意力图的修正,不断细化病灶的边缘细节。

传统的特征金字塔结构通常采取多路分支,同时使用不同大小卷积核来进行不同尺度特征的提取。不仅计算复杂度高,而且还忽略了对提取的特征做出修正;而特征交互模块是将已有的编解码不同层之间的不同尺度的信息进行拼接融合,在不增加网络复杂度的前提下对提取的多尺度信息进行学习并做出调整,提高了对病灶细节信息预测的准确度,实现了对于肺部图像的高效处理。

在医学图像领域,大部分模型都采用 U 型架构,使用在编码层与对应的相同分辨率的解码层的输出特征之间进行相加或者拼接操作的跳跃连接的方式,来实现编解码相同层之间信息的融合。鉴于此类方式过于单一,传统特征金字塔结构的不足以及人类视觉的特性,本文在解码阶段设计了特征交互模块,如图 4 所示。

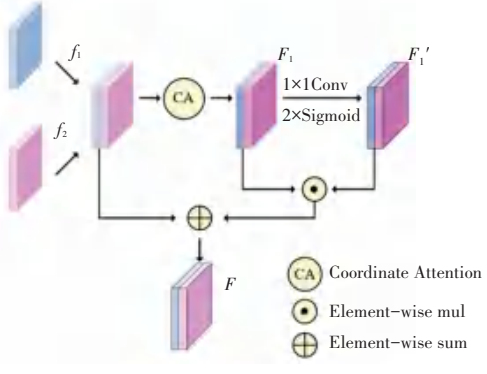

图 4 特征交互模块

Fig. 4 Feature interaction module

首先,将编解码不同层之间的不同尺度的信息进行拼接融合;然后,使用轴向注意力[14](CA)对特征信息分配权重;其次,通过 1×1 卷积和 2 个 Sigmoid 操作对注意力图进行调整修正;最后,再与注意力之前特征相加得到结构细化后的分割图。具体计算规则如下:

$$\begin{cases} F_1 = \mathrm{CA}(f_1 + f_2) \\ F_1' = 2 \times \mathrm{Sigmoid}(1 \times 1\ \mathrm{Conv} F_1) \\ F = (f_1 + f_2) + F \cdot F_1' \end{cases} \tag{2}$$

其中，f_1 和 f_2 分别表示编码层和其对应解码上一层的输出特征；CA 表示轴向注意操作；F_1 表示经过轴向注意机制得到的注意权重图；$F_1^{'}$ 表示经过权重调整后得到的修正图；F 表示特征交互模块的输出特征图。

特征交互模块采用轴向注意力，因其不仅仅能捕获跨通道的信息，还能捕获方向感知和位置感知的信息。此外，轴向注意力的计算复杂度较低，有助于模型在不引入过多的参数的条件下更加精准地定位和识别感兴趣的病灶结构细节。这里的 Sigmoid 操作取 2 倍，其目的是将注意力的可调整幅度限制到 0~2 之间，从而使病灶信息的权重得到进一步加强，非病灶信息的权重得到更好的抑制，因此能够突显病灶信息与背景信息之间的对比，避免模糊预测。后续得到的轴向注意力权重修正图不仅能够更好地突出某些病灶细节的注意力，同时也能有效抑制干扰信息。本模块的设计丰富了跳跃连接的形式，弥补了特征金字塔的不足以及较为完善地模拟了人类视觉的特性，在应用中取得了良好的表现。

1.3　亚像素卷积

解码阶段对于恢复肺部病灶结构细节至关重要，常见的上采样方法有双线性插值上采样和反卷积上采样。其中，双线性插值上采样是基于单线性插值，其方法实现简单，无需训练，但是会丢失信息；反卷积上采样，顾名思义，上采样的过程是卷积操作，就是通过先在待上采样的特征图周围补 0 后再卷积，提高输出的分辨率。该方法需要训练，相对于前者能更好地还原特征信息，但是补 0 会引入无效信息，甚至对梯度优化带来不利影响。为了克服上述不足，本文采用亚像素卷积（具体结构如图 5 所示），在不失真的情况下增加图像的分辨率，并提高图像的质量和视觉效果。

在相机成像过程中，图像数据经过离散化处理，每个像素代表成像面上的一个区域的颜色。由于感光元件的限制，相邻像素之间存在间隔，即像素之间有一定的物理距离。这些物理距离内的像素被称为亚像素。从宏观角度看，这些亚像素被视为相邻像素的一部分；但从微观角度看，这些亚像素实际上是独立存在的，并且可能存在微小的差异。因此，亚像素可以看作是对图像细节的更精细的表示。根据相邻像素之间插值情况的不同 ，可以调整亚像素的精度（见图 5），例如将低分辨率 128×128 的特征图放大 3 倍，也就是每个像素从横向和纵向上细分为 3 个像素点，变成 512×512 的尺寸大小。这样就要先进行隐藏层卷积操作，产生 3×3、9 张相同大小的特征图，然后把这 9 张特征图拼接成一张放大 3 倍的大图。具体操作是对 9 张特征图分别取第一个像素点，拼成 3×3 的像素块作为高分辨率图的第一个像素块，每个像素点都执行同样的操作，即低分辨率图一个像素点对应高分辨率图一个 3×3 的像素块，最终实现特征图的有效放大。

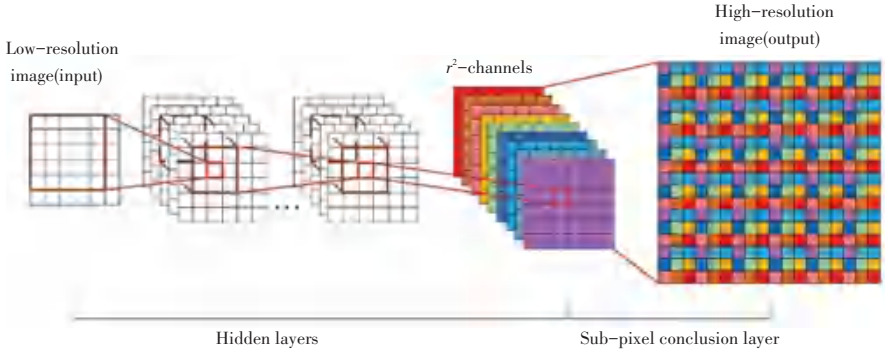

图 5　亚像素卷积

Fig. 5　Subpixel convolution

1.4　损失函数

面向肺部病灶图像进行语义分割的本质为像素点进行逐个分类，故而采用如下交叉熵作为基本损失函数：

$$L_{\mathrm{BCE}} = -\sum_{i=1}^{n} y_i \log y_i^{'} \tag{3}$$

其中，L_{BCE} 表示基本交叉熵损失；y_i 表示标签值；$y_i^{'}$ 表示预测值。

对于只分割前景和背景的二值分割场景，当前景像素的数量远远小于背景像素的数量时，损失函数中的 $y_i = 0$ 的成分就会占据主导，使得模型严重偏向背景。此外，交叉熵损失只是像素的整体概率

分布,对于图像分割任务,却并未考虑目标的整体结构。基于此,本文进一步添加了 IoU 损失 L_{IoU}:

$$L_{\mathrm{IoU}} = 1 - \frac{A \cap B}{A \cup B} \tag{4}$$

其中, A 表示计算的特征图, B 表示标签图。

整体的损失函数 L_{total} 由 IoU 损失函数(L_{IoU})与交叉熵损失函数(L_{BCE})联合组成,实现了整体结构和小目标分割的权衡:

$$L_{\mathrm{total}} = L_{\mathrm{IoU}} + \lambda L_{\mathrm{BCE}} \tag{5}$$

2 实验与结果分析

在本节中,首先对实验配置进行介绍,如数据集、实验设置、评价指标等;接着,将本文设计的 U 型网络与 7 个用于肺部病灶图像分割的深度学习网络进行性能对比;最后,详细分析数值实验结果与可视化实验结果。通过消融实验验证 Swin-Transformer 层、特征交互模块和亚像素卷积的有效性。

2.1 数据集

本文实验主要采用 2 个公开且具备典型性的权威肺部病灶图像集,即 COVID-19 CT scan 数据集和 MS COVID-19 数据集。具体信息如下。

(1)COVID-19 CT scan 数据集[15]:该数据集由 20 个注释的 COVID-19 胸部 CT 序列组成。每一个 CT 序列内的病灶图像都是由专业放射科医生进行验证标注。另外,每个序列内图像的分辨率为 512×512,平均切片个数为 176。

(2)MS COVID-19 数据集[16]:该数据集由意大利医学和介入放射学会整理发布,共收集了 40 多名肺部感染病人的 100 张轴向 CT 图像。

2.2 实验设置

实验训练与测试主要在配有 NVIDIA GTX 3090Ti 显卡的 Windows10 操作系统上进行,且本文设计的 TF-Unet 主要利用 PyTorch 深度学习框架进行搭建。

(1)基本设置:本文采用标准裁剪和随机翻转方式进行数据增强,使用 5-折交叉验证来进行网络训练与测试。具体地,将数据集平均分成 5 份,每次随机选取 4 份进行网络训练,1 份进行测试。训练及测试过程重复 5 次,直至测试完所有数据为止。

网络训练时,采用 Poly 学习率策略,即首先将初始学习率设置为 1×10^{-3},随后使用 Adam 随机梯度下降算法,以 1×10^{-4} 的权重衰减率进行网络训练,且数据集在网络中训练的往返次数为 80,批量数为 8。另外,便于公正客观地比较 TF-Unet 与其余网络的分割效果,本文对于其余网络进行同样的设置。

(2)模块设置:对于设计的 TF-Unet 内编解码各阶段输出特征图尺寸大小分别为{256×256,128×128,64×64,32×32,16×16},输出特征图通道数分别设定为{32,64,128,160,256}。

2.3 评价指标

为了评估 TF-Unet 对于肺部病灶图像的病灶区域与非病灶区域的分割效果。本文采用医学影像分析中常用的 5 种评价指标来衡量构建网络的表现性能,包括:Dice、mIoU、SEN、以及 SPC。其中,Dice 为 Dice 相似系数,主要用来衡量整体的像素估计类别与真实类别之间的相似性;mloU 为平均交并比,用于计算病灶区域像素正确归类与真实病灶区域像素的比率;SEN 为灵敏度,表示网络对于病灶区域像素的正确判断归类程度;SPC 为特异性,表示网络对于非病灶区域像素的正确判断归类程度。对于这 4 个性能指标而言,其数值越大,表示网络分割肺部病灶图像的效果越好。4 个评价指标表达式为:

$$\mathrm{Dice} = \frac{2 \times \mathrm{TP}}{\mathrm{FN} + 2 \times \mathrm{TP} + \mathrm{FP}} \tag{6}$$

$$\mathrm{mIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{FP} + \mathrm{TP}} \tag{7}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8}$$

$$\mathrm{SPC} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{9}$$

其中,TP 表示病灶区域像素归类正确率;TN 表示非病灶区域像素归类正确率;FP 表示病灶区域像素归类错误率;FN 表示非病灶区域像素归类错误率; k 表示图像分割区域类别数。

为了量化网络对于图像内分割区域边界的敏感程度,本文还采用 HD(Hausdorff Distance)距离来度量图像分割前后 2 组像素点集之间的最近距离,其表达式为:

$$h(A,B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a,b) \} \} \tag{10}$$

其中, a 和 b 分别表示像素集合 A 和 B 内的点; $d(a,b)$ 表示像素点 a 和 b 之间的欧氏距离。对于 HD 而言,其数值越小,代表着网络的分割效果越好。

2.4 对比验证

为了验证所构建 U 型网络的优越性,本文进行了大量的对比实验。参与对比实验的分割网络共 7 个,分别为 UNet[3]、UNet++[17]、PraNet[18]、

MedT[19]、MiniSeg[10]、MT-UNet[20]和 UNext[11]。

2.4.1　数值结果对比

采用 2.2 节规定的实验设置，在 2 个数据集上进行网络性能对比实验，最终得到了相应的结果。需要说明的是，受篇幅限制，在此仅以在 COVID-19 CT scan 数据集上的结果为例，详细分析比较各个网络的表现性能。

表 1 记录了 TF-Unet 与 7 个其他优秀模型在 COVID-19 CT scan 数据集上取得的 5 个性能指标结果。从表 1 可以看出，基于纯卷积的模型性能要比融合注意力机制的模型性能差一些，原因是纯卷积的方法缺乏对非线性特征的提取能力以及无法建模全局上下文信息。本文所设计的 Transformer 特征交互的 U 型网络在 5 个评价指标上皆取得了最优的结果。Dice 系数达到了 0.791 6，平均 IoU 达到了 0.853 2，SEN 达到了 0.869 5，SPC 达到了 0.997 7，HD 距离达到了 0.536 4。

表 1　TF-Unet 与非轻量级网络在 COVID-19 CT scan 上的指标结果

Table 1　Experimental results of the TF-Unet and the non-lightweight networks on COVID-19 CT scan networks　%

模型	Dice	mIoU	SEN	SPC	HD
UNet[3]	64.96	75.78	82.27	98.82	131.84
UNet++[17]	68.52	78.50	71.77	99.14	75.63
PraNet[18]	76.51	82.70	76.25	99.12	74.85
MedT[19]	73.67	77.04	82.15	99.32	63.38
MiniSeg[10]	77.35	82.55	84.06	98.46	70.62
MT-UNet[20]	70.25	73.73	67.09	99.27	76.62
UNext[11]	78.01	84.78	84.59	99.57	56.42
TF-UNet	**79.16**	**85.32**	**86.95**	**99.77**	**53.64**

因此，本文所构建的 TF-UNet 面向肺部病灶图像能够取得较好的分割结果。相似的结果与结论在 MS COVID-19 数据集上同样可以得到。

2.4.2　可视化结果对比

为了清晰直观地体现本文模型在肺部病灶图像的分割效果，图 6 展示了 TF-Unet 与现有最新图像分割网络的可视化对比实验结果，可以清晰地发现 TF-Unet 对于肺部病灶图像的分割效果明显优于其它网络。特别是第 2 列、第 5 列内，对于小目标对象以及不连续的目标对象，本文模型更加敏感且分割精度更高，这主要得益于 Swin-Transformer 层对于全局信息的把握、特征交互模块对于多尺度信息的捕捉以及亚像素卷积对于图像分辨率的有效还原。

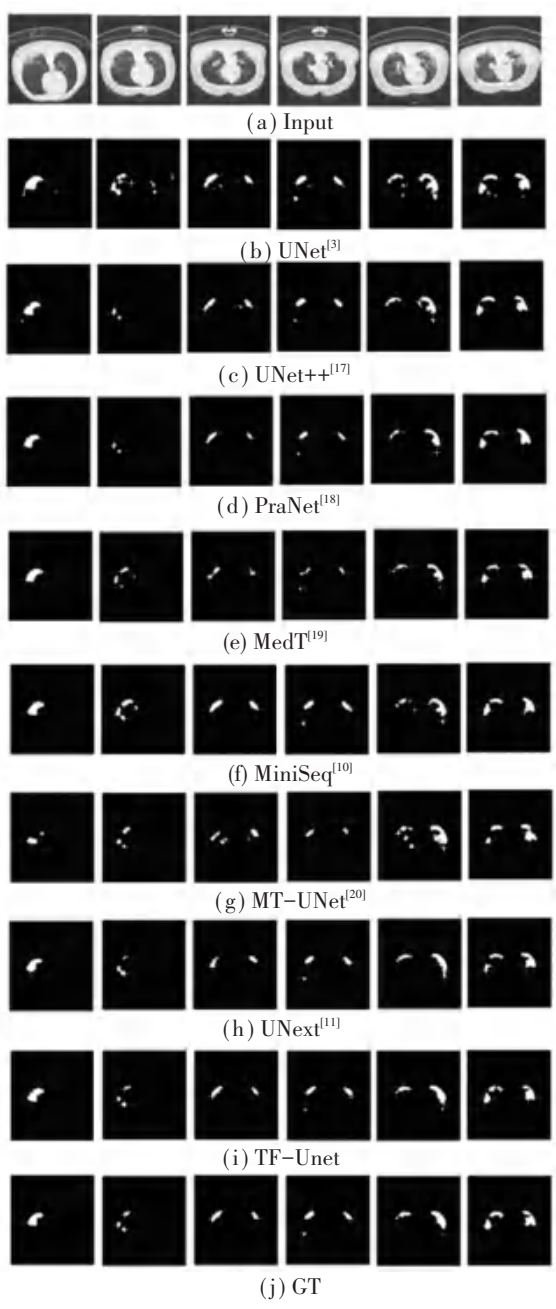

图 6　不同分割网络输出的可视化对比分割图

Fig. 6　Visualization segmentation output comparison of different segmentation networks

2.5　消融实验

本文在肺部病灶区域分割模型中引入了 Swin-Transformer 层、特征交互模块和亚像素卷积，以改善分割性能。为了验证这些模块的有效性，本文使用 UNext[11]作为基线模型，并进行了消融实验。具体实验设置包括以下几种模型：基线模型；基线模型+Swin-Transformer 层；基线模型+Swin-Transformer 层+特征交互 FI 模块；线模型+Swin-Transformer 层+特征交互 FI 模块+亚像素 sub-pixel 卷积、即本文提出的 TF-Unet 网络模型。通过消融实验结果，可以评

估这些模块对于肺部病灶区域分割性能的影响。参与消融实验的所有 U 型网络见表 2。

表 2 参与消融实验的所有 U 型网络

Table 2 All U - shaped networks involved in the ablation experiments

模型	SW	FI	Sub-pixel
基线模型			
SW	√		
SW+FI	√	√	
TF-Unet	√	√	√

随后,以在 MS COVID-19 数据集上取得的实验结果为例,详细分析设计的各个模块的有效性。表 3 记录了参与消融实验的网络在 MS COVID-19 数据集上取得的 5 个性能指标值。对比基线模型和加入 SW 层的网络取得的 5 个指标值可以看出,前 4 个指标值随着 Swin-Transformer 层的加入而有所提升,最后一个指标值随着其加入而降低,这说明 Swin-Transformer 层能够在低层编码阶段有效地捕捉全局信息,并且抑制图像内噪声产生的不良影响。通过对比加 SW 和加 SW+FI 取得的指标值,则可以看出指标 Dice、mIoU、SEN、SPC 分别提升了 0.44%、0.66%、3.76%、0.45%,指标 HD 降低了 2.58%。这证实了特征交互模块能够融合多尺度不同特征,恢复图像的结构细节。比较加 SW+FI 和 TF-Unet 取得的最终指标值可以发现,指标 Dice、mIoU、SEN、SPC 分别提升了 0.08%、0.32%、0.22%、0.20%,指标 HD 降低了 0.83%,这表明亚像素卷积能够有效放大图像。从 TF-Unet 取得的指标结果可以看出,所有指标值都达到了最优。

表 3 参与消融实验的分割网络在 MS COVID-19 上的实验结果对比

Table 3 The experimental results comparison of all segmentation networks involved in the ablation experiments on MS COVID-19 %

模型	Dice	mIoU	SEN	SPC	HD
基线模型	75.17	81.42	80.39	96.98	76.84
SW	75.61	81.79	81.38	97.37	74.89
SW+FI	76.05	82.45	85.14	97.82	72.31
TF-Unet	**76.13**	**82.77**	**85.36**	**98.02**	**71.48**

为了直观地了解各模块的作用,本文选取了部分分割结果进行对比,具体如图 7 所示。从第 3 列的前 2 张分割图可以看出,Swin-Transformer 层的全局注意力可以捕捉到遗漏的病灶信息;对比第 2 列的第 2、3 张分割图可以发现,特征交互模块可以更好地恢复图像的结构细节,特别是对于微小非病灶区域的判别能力很强;对比第 1 列的第 3、4 张分割图可以看到,亚像素卷积可以对分割的缩略图进行有效的放大;将各种模型与标签图对比,可以发现在纹理结构较为复杂的 CT 图像中,TF-Unet 得到的分割图最为精准完整。

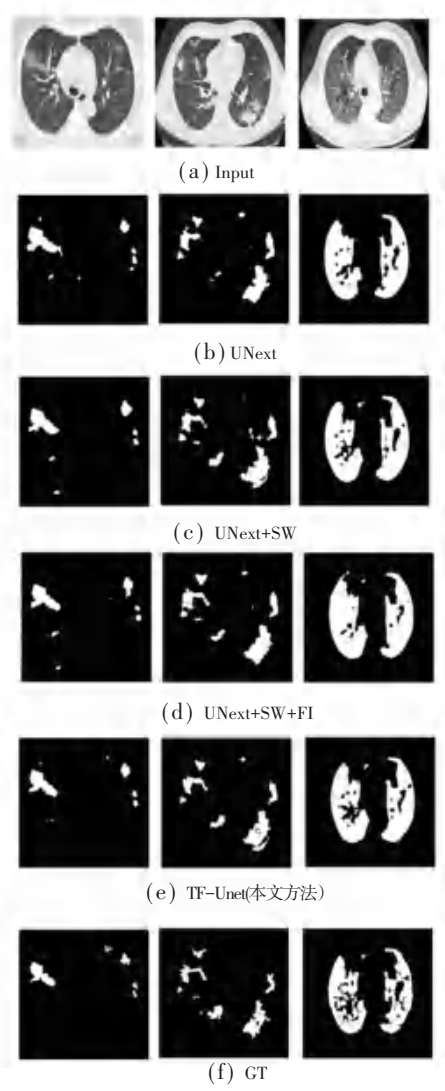

图 7 消融实验输出的可视化对比分割图

Fig. 7 Visualization segmentation output comparison of ablation experiment

3 结束语

针对肺部病灶图像采用 U 型网络自动分割的问题,本文提出了一种基于 Transformer 特征交互的 U 型网络 TF-Unet。首先,在低层的编码环节添加了 Swin-Transformer 层,有助于网络在提取低层信息时,强化对全局特征信息的关注,全面定位复杂多变的肺部病变阴影;改造跳跃连接的固有设计,搭建特征交互模块融合编解码不同层之间的多尺度信

息，修正注意力图的操作能够更加清晰地恢复病灶结构细节；将亚像素卷积应用到上采样的过程中，实现分割结果图的高效还原。此外，本模型构建的联合损失函数，对于整体结构和小目标的分割起到加强的效果。实验结果表明：本文提出的 U 型网络分割方法的性能明显优于其他医学图像分割模型。在以后的工作中，将对网络进行半监督或无监督学习的训练方式[21]，减少对于大量标注的医学数据的依赖，有效利用无标注图像来提高分割精度。

参考文献

[1] 袁甜，程红阳，陈云虹，等. 基于 U-Net 网络的肺部 CT 图像分割算法[J]. 自动化与仪器仪表，2017(6)：59-61.

[2] 崔文成，王可丽，邵虹. 基于稠密块和注意力机制的肺部病理图像异常细胞分割[J]. 智能科学与技术学报，2023，5(4)：525-534.

[3] RONNEBERGER O，FISCHER P，BROX T. U-Net：Convolutional networks for biomedical image segmentation [C]// Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI 2015). Cham：Springer，2015：234-241.

[4] 樊圣澜，柏正尧，陆倩杰，等. 基于 Transformer 网络的 COVID-19 肺部 CT 图像分割[J]. 中国图象图形学报，2023，28(10)：3203-3213.

[5] HU Jie，SHEN Li，ALBANIE S，et al. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2018：7132-7141.

[6] WANG Qilong，WU Banggu，ZHU Pengfei，et al. ECA-Net：Efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2020：11534-11542.

[7] WANG W，TAN X，ZHANG P，et al. A CBAM based multiscale transformer fusion approach for remote sensing image change detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2022，15：6817-6825.

[8] YAN L，LIU D，XIANG Q，et al. PSP net-based automatic segmentation network model for prostate magnetic resonanceimaging [J]. Computer Methods and Programs in Biomedicine，2021，207：106211.

[9] QIU Yu，LIU Yun，LI Shijie，et al. MiniSEG：An extremely minimum network for efficient COVID-19 segmentation [J]. Proceedings of the AAAI Conference on Artificial Intelligence，2021，35(6)：4846-4854.

[10] MEI Haiyang，JI Gepeng，WEI Ziqi，et al. Camouflaged object segmentation with distraction mining [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2021：8772-8781.

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

[12] LIU Ze，LIN Yutong，CAO Yue，et al. Swin transformer：Hierarchical vision transformer using shifted windows [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway，NJ：IEEE，2021：10012-10022.

[13] SHAO G，SUN Q，GAO Y，et al. Sub-pixel convolutional neural network for image super-resolution reconstruction [J]. Electronics，2023，12(17)：3572.

[14] WEN Ge，LI Shaobao，LIU Fucai，et al. YOLOv5s-CA：A modified YOLOv5s network with coordinate attention for underwater target detection[J]. Sensors，2023，23(7)：3367.

[15] JUN M，CHENG G. COVID-19 CT lung and infection segmentation，Dataset[EB/OL]. (2020-04-20). https://www.kaggle.com/andrewmvd/covid 19- ct-scans.

[16] JENSSEN H B. Covid-19 radiology-data collection and preparation for artificial intelligence[EB/OL]. (2020-04-10). https://sirm.org/category/ sen- zacategoria/covid-19/.

[17] JIA Yutong，LIU Lei，ZHANG Chenyang. Moon impact crater detection using nested attention mechanism based UNet++[J]. IEEE Access，2021，9：44107-44116.

[18] LIU Limei，LIU Meng，MENG Kexin，et al. Camouflaged locust segmentation based on PraNet[J]. Computers and Electronics in Agriculture，2022，198：107061.

[19] VALANRASU J M J，OZA P，HACIHALILOGLU I，et al. Medical transformer：Gated axial-attention for medical image segmentation [C]//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention(MICCAI 2021). Cham：Springer，2021：36-46.

[20] WANG Hongyi，XIE Shi′ao，LIN Lanfen，et al. Mixed transformer U-Net for medical image segmentation [C]// Proceedings of 2022 IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP). Piscataway，NJ：IEEE，2022：2390-2394.

[21] 郭敏，张熙涵，李阳. 融合注意力的教师互一致性半监督医学图像分割[J]. 计算机工程，2024，50(9)：313-323.