

融合多尺度特征和双分支并行的肺结节图像分割网络^①



王超学, 王 磊

(西安建筑科技大学 信息与控制工程学院, 西安 710055)

通信作者: 王 磊, E-mail: wanglei1998531@163.com

摘 要: 肺结节图像的准确分割对于肺癌的早期诊断具有重要意义, 针对肺结节图像尺度多样、边缘模糊导致特征提取不充分和细节信息丢失问题, 本文提出一种融合多尺度特征和双分支并行的肺结节图像分割网络 RAVR-UNet. 首先, 针对 U-Net 网络在编码阶段无法充分提取肺结节特征, 采用双分支并行特征聚合网络提取肺结节图像中的特征信息, 减少特征编码时的信息损失. 其次, 通过引入 Agent_ViT 模块, 在保持线性计算的基础上, 增强全局信息建模能力. 然后, 为恢复下采样期间丢失的肺结节空间信息, 在解码阶段加入多尺度特征融合模块. 最后, 设计混合损失函数以缓解肺结节图像分割任务中正负样本不平衡问题. 在 LIDC-IDRI 公开数据集上的实验结果表明, 所提网络的相似系数、交并比分别达到 93.15%、87.63%, 优于主流肺结节分割算法且分割结果更接近真实值.

关键词: 肺结节图像分割; 多尺度特征融合; U-Net; ViT; 代理注意力

引用格式: 王超学, 王磊. 融合多尺度特征和双分支并行的肺结节图像分割网络. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9832.html>

Pulmonary Nodule Image Segmentation Network Integrating Multi-scale Features and Dual-branch Parallelism

WANG Chao-Xue, WANG Lei

(School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Accurate image segmentation of pulmonary nodules is of great significance for the early diagnosis of lung cancer. To solve the problem of insufficient feature extraction and detail loss caused by multiple scales and blurred edges of pulmonary nodules image, this study proposes a pulmonary nodule image segmentation network named RAVR-UNet, which incorporates multi-scale features and double-branch parallel. Firstly, given the inability of the U-Net network to fully extract pulmonary nodule features in the coding stage, a double-branch parallel feature aggregation network is used to extract the feature information from pulmonary nodule images to reduce the information loss during feature coding. Secondly, the Agent_ViT module is introduced to enhance the capability of global information modeling while maintaining linear computation. Then, to recover the lost pulmonary nodule spatial information during subsampling, a multi-scale feature fusion module is added in the decoding stage. Finally, a mixed loss function is designed to alleviate the imbalance between positive and negative samples in the pulmonary nodule image segmentation task. Experimental results on the LIDC-IDRI public dataset show that the similarity coefficient and intersection over union (*IoU*) of the proposed network reach 93.15% and 87.63%, respectively, which is better than the mainstream pulmonary nodal segmentation algorithms, and the segmentation results are closer to the real values.

Key words: pulmonary nodule image segmentation; multi-scale feature fusion; U-Net; ViT; agent attention

^① 基金项目: 国家自然科学基金面上项目 (62072363, 32471597)

收稿时间: 2024-10-05; 修改时间: 2024-10-23; 采用时间: 2024-11-19; csa 在线出版时间: 2025-03-04

肺癌是当今世界上最常见的癌症之一,早期肺癌多表现为肺结节^[1].在肺部CT图像中,准确识别肺结节对于肺癌的早期诊断具有重要意义^[2].

近年来,基于深度学习的肺结节识别已经成为当前研究热点.其中,以卷积神经网络为基础的U-Net^[3]及其多种改进变体^[4-10]在肺结节图像分割任务中得到广泛应用.然而,卷积运算在局部感受野内进行特征提取的特性限制了上述网络在浅层捕获全局信息的能力.Transformer^[11-14]等网络的提出可以有效缓解这一问题,但却容易忽略或者破坏某些局部细节.

为充分发挥卷积神经网络与Transformer模型各自的优势和特点,研究者们相继提出多种混合模型^[15-21]应用于肺结节图像分割领域,显著提高模型分割性能.尽管现有研究已经在肺结节分割领域取得一定进展,但仍存在以下问题:首先,现有的分割网络大多采用单分支网络结构,无法有效捕捉不同尺度的特征信息.其次,连续的采样操作造成部分细节信息丢失和感受野受限问题,从而影响整体分割效果.

针对上述问题,在前人研究基础上,本文提出一种

融合多尺度特征和双分支并行的肺结节图像分割网络,简称为RAVR-UNet,主要工作如下.

(1) 设计双分支并行特征聚合网络,提取图像多级别特征信息并采用注意力门模块融合两个分支特征.

(2) 设计Agent_ViT模块并将其引入到编解码器之间,使得网络可以在保持线性计算的同时享受更大甚至全局感受野的优势.

(3) 提出的多尺度特征融合模块,采用不同膨胀率的卷积操作来捕获多尺度特征信息,在上采样过程中保留更多的细节和上下文信息,减少信息丢失.

(4) 网络训练过程中采用交叉熵损失函数和Dice相似系数加权结合的混合损失函数,缓解图像中前景和背景不平衡问题.

1 RAVR-UNet 网络

本文提出的RAVR-UNet网络结构如图1所示.该网络整体框架基于双分支编解码器-单解码器架构,主要由双分支并行聚合网络、Agent_ViT和解码器3个部分组成,接下来将重点介绍网络的改进部分.

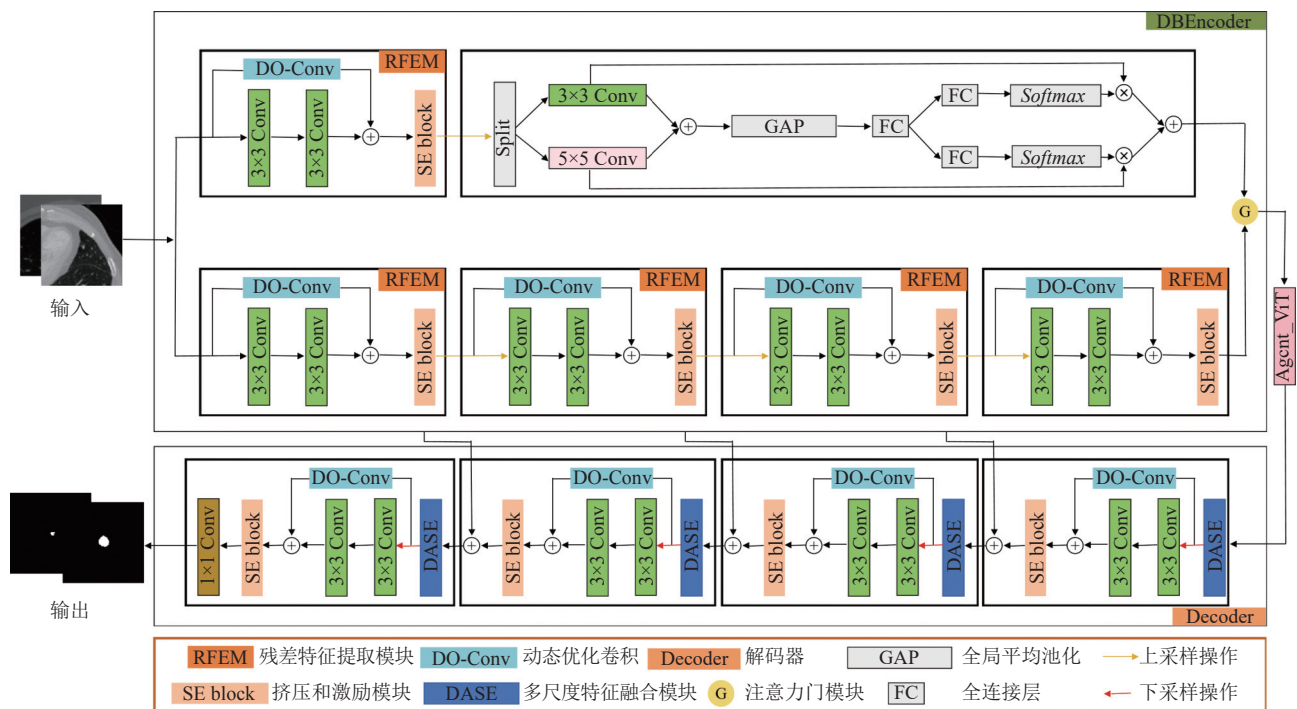


图1 模型结构

1.1 双分支并行聚合网络

1.1.1 残差特征提取模块

本文采用残差特征提取模块(residual feature extr-

action module, RFEM)代替原始卷积层作为编码器、解码器的基础架构.该模块基于原始U-Net编码器的残差结构,融合了深度过参数化卷积(depthwise over-

parameterized convolutional layer, DO-Conv)^[22]、压缩和激活模块 (squeeze-and-excitation, SE)^[23]。其中, 深度过参数化卷积通过引入冗余参数和更复杂的权重来增强卷积操作的特征提取能力。SE 注意力模块通过全局平均池化操作来排除空间依赖性, 增强模型对关键特征的代表能力, 使得模型能够自动学习不同通道特征的重要性。SE 模块的实现分为两个阶段: 压缩 (squeeze) 和激活 (excitation)。其架构如图 2 所示。

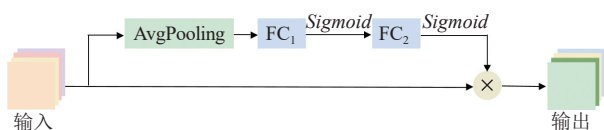


图 2 Squeeze-and-excitation 模块

1.1.2 双分支编码结构

与传统的 U-Net 框架不同, 在编码器部分, 特征提取网络采用双分支编码器机制 (double branch encoder,

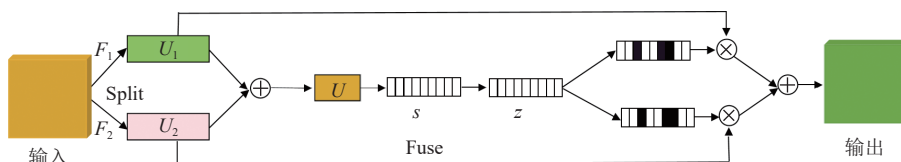


图 3 SKConv 模块

1.1.3 注意力门模块

在双分支编码结构中, 两个分支从原始图像中分别提取全局和局部语义信息, 简单地将这两种复杂度差异较大的特征信息进行组合可能会导致部分信息丢失和模型性能下降。对此本文提出注意力门模块, 模型结构如图 4 所示。该模块通过动态选择机制调整特征融合的方式, 实现不同层次特征信息的交互, 增强网络对肺结节区域的感知能力。

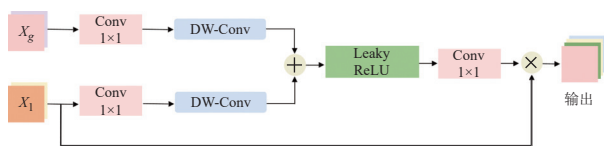


图 4 注意力门模块

1.2 Agent_ViT 模块

Vision Transformer 编码器中多头自注意力 (multi-head self attention, MHSA) 需要计算多个注意力头的输出, 每个头都要进行注意力权重的计算和线性变换, 存在计算复杂度高的问题。对此本文提出改进的 ViT 模块 Agent_ViT, 如图 5 所示。使用 agent attention^[25]代替

DBEncoder)。通过这种方法, 输入特征映射遍历两个不同的编码器分支, 不仅可以增强模型的特征学习能力, 而且能够在特征提取过程中保留更多的细节信息。

在等效尺度下进行下采样操作后, 两个分支收敛于初始编码器块以产生相同的特征。随后, 类似于最初 U-Net 架构, 第 1 个分支对高分辨率特征逐步进行下采样操作, 这种分层输出和输入方法可以获取更精准的特征映射, 捕获局部特征。第 2 个分支对该特征进行 4 倍下采样操作, 得到的结果细节性较差, 关注的是整体结构。随后, 该分支进行选择性核卷积运算^[24] (selective kernel convolution, SKConv)。通过引入选择性内核机制, 在不增加网络复杂度的情况下有效地扩大感受野, 提高网络对不同尺度特征的捕捉能力, 允许网络在每个位置上选择性聚焦, 从而在局部细节和全局特征之间找到最佳权衡点。SKConv 主要包括 3 个步骤: Split、Fuse 和 Select, 其架构如图 3 所示。

ViT 中的多头自注意力机制, 使得网络可以在保持线性计算量的同时享受更大甚至全局感受野的优势。

Agent attention 由两个 Softmax 注意力组成, 即 agent 组合和 agent 广播。具体而言, A 首先被视为 query 并在 A 、 K 、 V 之间执行注意力计算, 即 agent 聚合。然后, A 作为 key 和 value 执行第 2 个注意力计算, 使用 query 矩阵 Q , 将全局信息从 agent 特征广播到每个 query 标记并获得最终输出, 如式 (1)、式 (2) 所示。通过这种方式, 避免了 Q 和 K 之间成对相似度的计算, 同时通过 agent token 保留每个 query、key 对之间的信息交换。

$$Q^A = \text{Attn}^S(Q, A, \text{Attn}^S(A, K, V)) \quad (1)$$

$$\begin{aligned} O^A &= \sigma(QA^T)\sigma(AK^T)V = \varphi_q(Q)\varphi_k(K)^T V \\ &= \text{Attn}^{\varphi_q/\varphi_k}(Q, K, V) \end{aligned} \quad (2)$$

其中, Q 、 K 、 $V \in \mathbb{R}^{n \times C}$ 表示 query、key 和 value 矩阵, $\sigma(\cdot)$ 为 Softmax 函数, $A \in \mathbb{R}^{n \times C}$ 为 agent tokens。通过这种方式避免了 Q 和 K 之间成对相似度的计算, 同时通过 agent token 保留每个 query、key 对之间的信息交换。

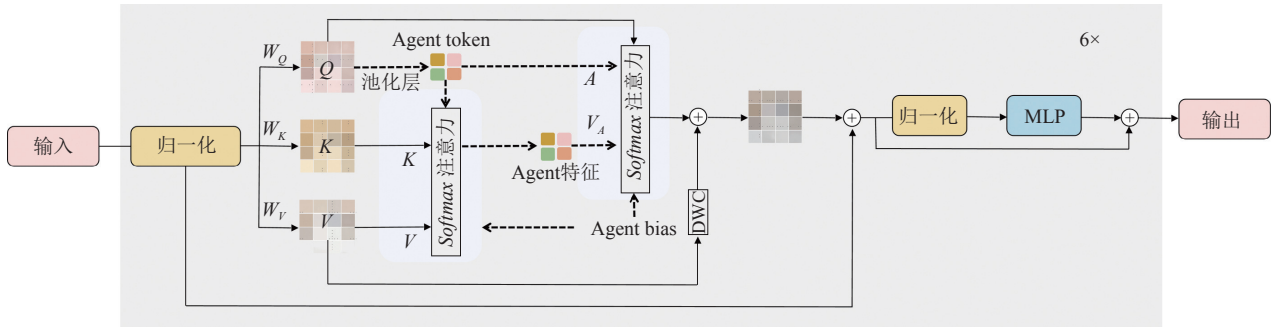


图5 Agent_ViT层

1.3 多尺度特征融合模块

在下采样过程中执行最大池化操作会丢失一些重要的空间信息, 导致分割准确率降低. 为使模型能够更好地结合上下特征层信息, 本文提出了基于ASPP^[26]的多尺度特征融合模块DASE, 模型架构如图6所示. DASE模块是一个多分支的模块, 在每个尺度上都有不同的感受野, 使得网络可以在各个尺度上充分探索

目标特征的局部和全局空间信息. 同时, 在DASE最后加入ScSE模块, 使得模型关注特征通道之间的关系来增强有意义的特征而抑制无用特征. 在转置卷积前加入多尺度特征融合模块, 不仅可以在上采样过程中保留更多的细节和上下文信息, 减少信息丢失, 而且提高小尺度物体的分割精度, 有助于处理边界模糊和尺度多样的肺结节分割问题.

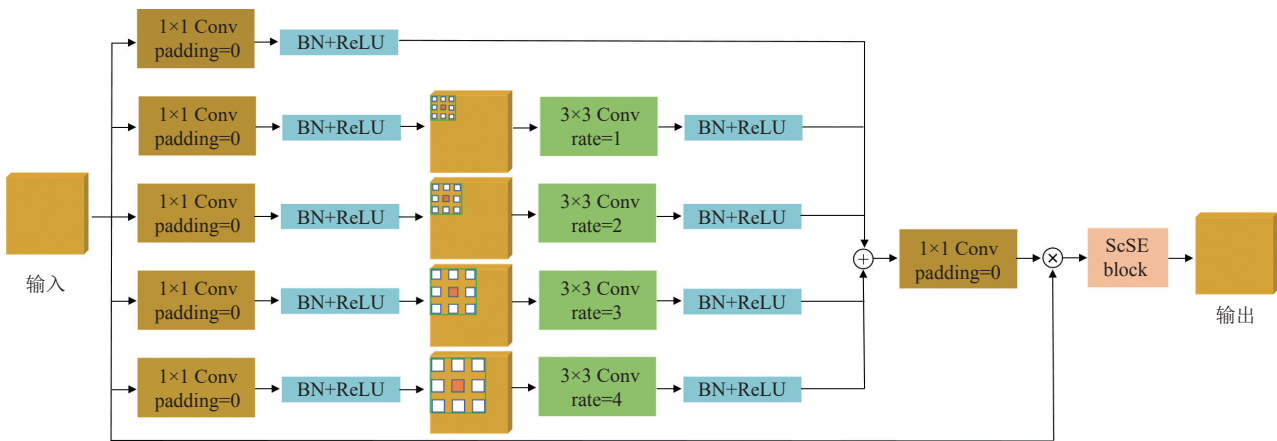


图6 多尺度特征融合模块DASE

ScSE模块是由空间压缩通道激励模块和通道压缩空间激励模块组成, 该模型架构如图7所示.

问题, 本文提出一种新的混合损失函数, 将Dice损失函数和BCE损失函数相结合, 如式(3)–式(5)所示:

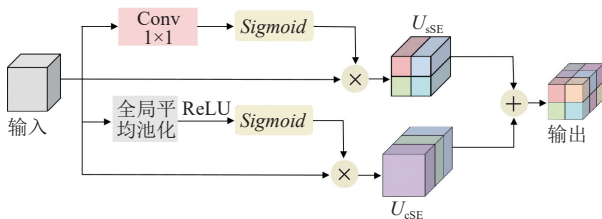


图7 ScSE模块

1.4 混合损失函数

在肺结节图像分割中, 肺结节在肺部图像中的比例远小于其他区域, 导致正负样本极度不平衡. 针对该

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

$$L_{BCE} = -\frac{1}{N} \sum_{n=1}^N w_n [y_n \log x_n + (1 - y_n) \log(1 - x_n)] \quad (4)$$

$$L = \alpha \times L_{BCE} + (1 - \alpha) \times L_{Dice} \quad (5)$$

其中, N 表示样本中所有像素点数量, x_i 代表第 i 个像素点被预测为正样本的概率值, y_i 代表第 i 个像素点的实际类别, 其中 $y_i=1$ 表示正样本, $y_i=0$ 表示负样本, w_i 为超参数, $|X \cap Y|$ 代表 X 和 Y 的交集, $|X|+|Y|$ 代表 X

和 Y 的元素个数, α 为权重因子, 用于控制 $Dice$ loss 和 BCE loss 权重系数的超参数.

2 实验与结果分析

2.1 数据集

本文采用的实验数据集来源于公开的肺部影像数据库 LIDC-IDRI^[27], 该数据集中每张图像像素为 512×512 , 如图 8(a) 所示. 由于原始 CT 图像中肺结节占比较小, 会产生类不平衡问题, 从而影响网络的训练过程. 因此, 为了避免其他肺部组织对实验结果的影响, 以原始图像中肺结节为中心裁剪图像. 最终得到像素为 128×128 的图像, 构成实验所用的数据集, 如图 8(b) 所示. 该数据集共有 15 096 张图像, 其中 9 057 张 (60%) 用于训练, 3 019 张 (20%) 用于验证, 3 020 张 (20%) 用于测试.

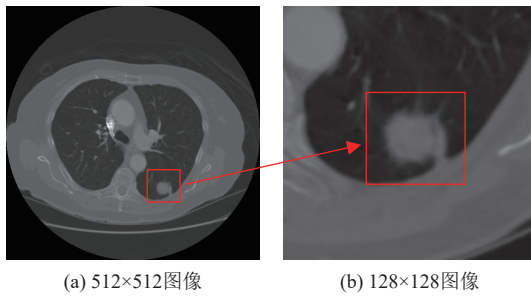


图 8 数据处理前后对比

2.2 实验设置

本文采用 PyTorch 1.11.0 作为核心框架, 并结合 Python 3.7 搭建网络. 在硬件配置上, 采用 NVIDIA GeForce RTX3090 GPU, 并在 Windows 10 操作系统上运行. 在模型训练过程中采用 Adam 优化算法, epoch 设置为 100, batch size 为 16, 初始学习率为 $1E-5$.

2.3 评价指标

本文采用 $Dice$ 相似系数、交并比 IoU 、精确率、召回率以及 $F1-Score$ 衡量所改进模型效果. 具体公式如式 (6)–式 (10) 所示:

$$Dice = \frac{2 \times TP}{(FP + TP) + (FN + TP)} \quad (6)$$

$$IoU = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

其中, TP 表示正确分类为正例的样本数, FP 表示错误分类为正例的样本数, TN 表示正确分类为负例的样本数, FN 表示错误分类为负例的样本数.

2.4 损失函数对比实验

为验证所提出的混合损失函数的有效性, 本文分别使用 BCE loss、 $Dice$ loss 和提出的混合损失函数来训练模型, 实验对比结果如表 1 所示. 从表中可以发现当网络使用混合损失函数的模型效果要高于单一损失函数的结果.

表 1 损失函数对比实验 (%)

| 损失函数 | IoU | $Dice$ | $Precision$ | $Recall$ |
|---------------|-------|--------|-------------|----------|
| BCE loss | 86.12 | 91.48 | 92.22 | 91.97 |
| $Dice$ loss | 86.38 | 91.78 | 92.44 | 91.39 |
| Proposed loss | 87.63 | 93.15 | 93.96 | 93.87 |

为了评估不同权重因子对模型分割性能的影响, 本文采用不同权重因子 α 训练模型以寻求最优的混合损失函数效果. 将其取值为 $[0, 1]$, 以 0.2 为间隔开展一系列消融实验, 可视化如图 9 所示. 对比实验结果中的各项评价指标确定 α 的最优值为 0.5.

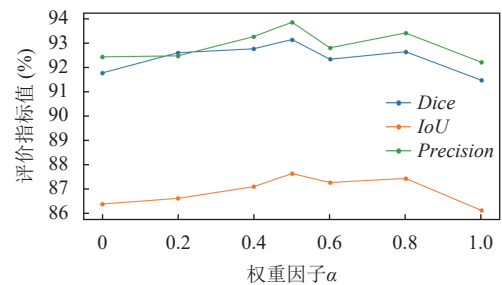


图 9 超参数 α 消融实验

2.5 消融实验

为验证 RAVR-UNet 网络中不同模块的有效性, 本文在相同的实验配置下对 LIDC-IDRI 数据集进行消融研究. 以 U-Net 作为基线网络, 验证添加 RFEM、Agent_ViT、DASE、DBEncoder 模块后的肺结节分割效果. 消融实验所得结果如表 2 所示, 与基线网络 U-Net 相比, 本文模型对于肺结节的分割效果明显更佳.

由表 2 可知, 在 U-Net 网络中加入 RFEM 模块后, 在测试集结果中, $Dice$ 和 IoU 的值分别提高了 0.58% 和 0.42%. 在此基础上, 分别添加 Agent_ViT、DASE 和

DBEncoder 模块,网络性能分别提高 0.78% 和 1.22%、0.60% 和 0.84%、0.84% 和 1.55%。此外,将 RFEM、DASE 和 Agent_ViT 这 3 个模块同时整合到 U-Net 中,网络性能相对于单一模块整合有显著提升。最后将以上 4 个模块都集成到 U-Net 中, *Dice* 和 *IoU* 值与 U-Net 网络相比分别提高了 2.85% 和 3.06%, 本文提出的 RAVR-UNet 在 LIDC-IDRI 数据集上表现最好,可以实现最佳分割效果。

为了更直观地展现该模型在肺结节分割任务中的性能表现,本文绘制了训练过程中损失函数变化曲线,如图 10 所示。由图中可以看出训练损失在 epoch 小于

20 时下降较快,并且迭代次数达到 55 轮以后,损失函数逐渐趋于稳定,最后收敛在 1% 左右。同时绘制了 *Dice* 和 *IoU* 曲线图,从曲线中可以更直观地看出本文所提出的模型分割性能达到最优。

表 2 消融实验性能对比 (%)

| RFEM | Agent_ViT | DASE | DBEncoder | <i>IoU</i> | <i>Dice</i> |
|------|-----------|------|-----------|------------|-------------|
| √ | — | — | — | 84.99 | 90.88 |
| √ | √ | — | — | 86.21 | 91.66 |
| √ | — | √ | — | 85.83 | 91.48 |
| √ | — | — | √ | 86.54 | 91.72 |
| √ | — | √ | √ | 87.23 | 92.37 |
| √ | √ | — | √ | 87.57 | 92.79 |
| √ | √ | √ | √ | 87.63 | 93.15 |

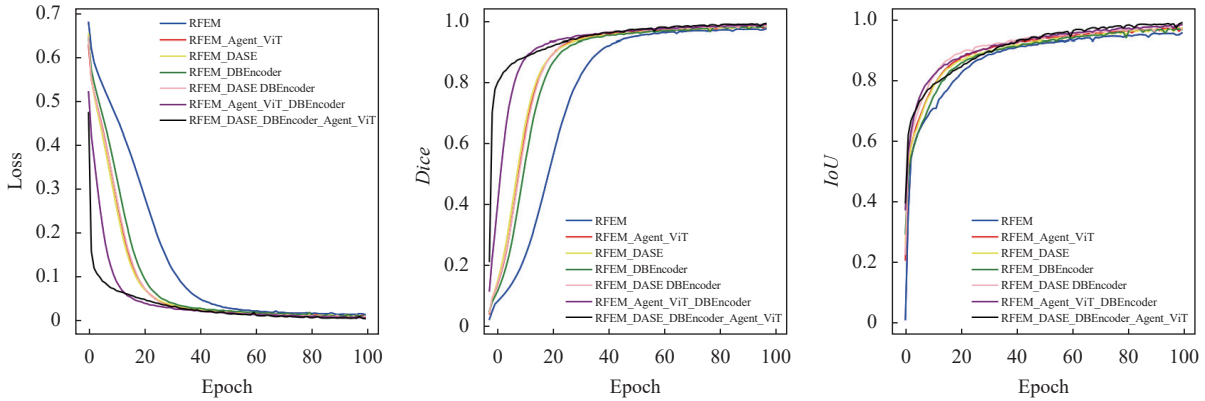


图 10 消融实验曲线对比

2.6 对比实验

为了验证 RAVR-UNet 在肺结节分割任务上的有效性,本文将其与现有的开源医学图像分割模型 Swin-UNet^[16]、TransUNet^[15]、Attention U-Net^[5]、UNet++^[4]、ResUNet++^[6]、U-Net^[3]、SMR-UNet^[28]、HTC-Net^[18]、DA-TransUNet^[17]在 LIDC-IDRI 数据集上进行对比实验和分析。在实验中,将相同的训练策略应用于 RAVR-UNet 和上述开源网络,以确保对比实验的公平性。如表 3 所示,本文所提出的方法在肺结节分割任务中获得了最优的分割效果。

图 11 展示了肺结节分割的部分示例,为清晰地观察肺结节的分割情况,本文从 9 个对比模型中随机选取两个进行展示。可以看出 U-Net、SMR-UNet 和 RAVR-UNet 都能够分割出肺结节区域,但存在一定欠分割和过分割现象。本文提出的 RAVR-UNet 网络可以有效区分肺结节和背景区域,其分割结果更接近真实值。对比其他医学图像分割模型,RAVR-UNet 能够更好地解决肺结节分割问题,表现出较高精确性和稳健性。

表 3 对比实验性能对比 (%)

| 方法 | <i>IoU</i> | <i>Dice</i> | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
|-----------------|--------------|--------------|------------------|---------------|-----------------|
| U-Net | 84.57 | 90.30 | 92.61 | 90.60 | 90.45 |
| Attention U-Net | 83.23 | 89.31 | 91.17 | 90.07 | 89.45 |
| ResUNet | 83.97 | 89.92 | 91.51 | 91.01 | 90.12 |
| ResUNet++ | 83.84 | 89.65 | 91.18 | 90.67 | 89.75 |
| TransUNet | 80.35 | 86.89 | 89.88 | 87.64 | 87.04 |
| Swin-UNet | 72.16 | 81.05 | 81.81 | 85.48 | 81.59 |
| HTC-Net | 82.05 | 88.52 | 92.75 | 86.77 | 88.55 |
| DA-TransUNet | 80.37 | 87.11 | 89.21 | 88.98 | 87.45 |
| SMR-UNet | 86.88 | 91.87 | 93.12 | 92.78 | 91.96 |
| RAVR-UNet | 87.63 | 93.15 | 93.96 | 93.87 | 92.33 |

此外,为了更直观地展示本文所提出的网络优于其他开源模型,绘制了不同模型在训练过程中训练集和验证集的损失函数及 *Dice*、*IoU* 曲线图,如图 12、图 13 所示。随着 Epoch 增加,所有模型的损失函数值呈现出下降趋势, Swin-UNet 下降速度较慢, RAVR-UNet 下降较快且较早趋于平稳。所有模型的 *Dice* 值和 *IoU* 值随着训练次数的增加而增加,并逐渐趋于平稳。从整体表现来看, RAVR-UNet 稳定性高于其他网络,综合表现最佳。

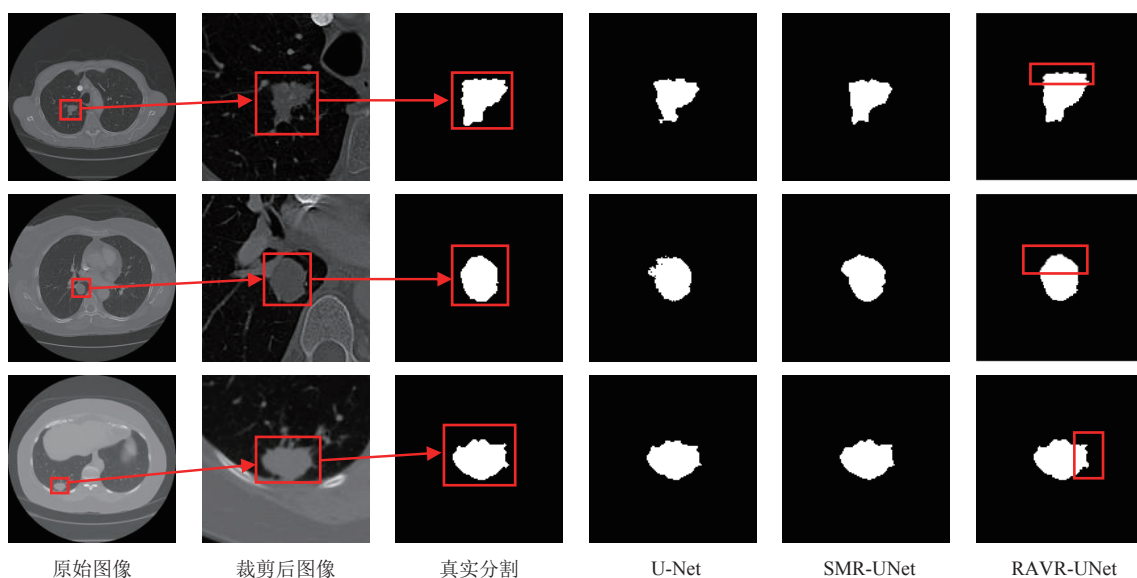


图 11 部分对比实验可视化效果

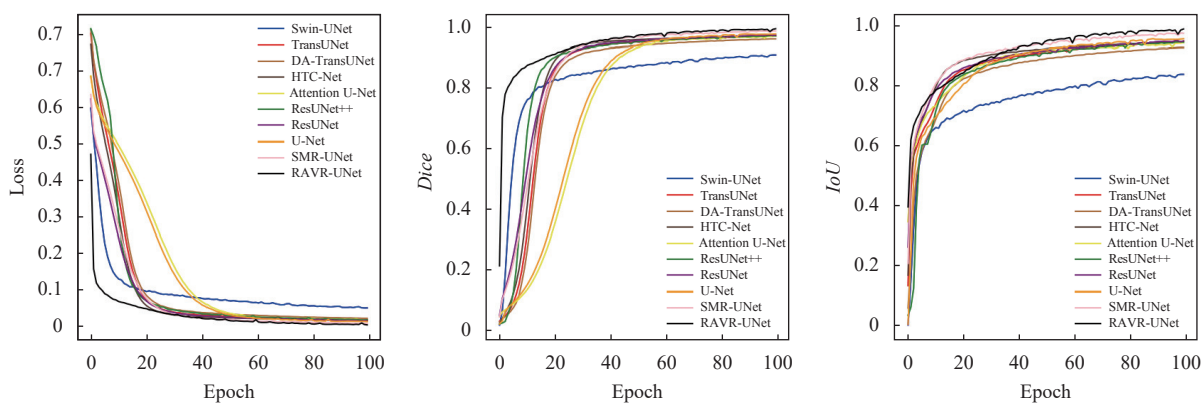


图 12 训练集对比实验曲线

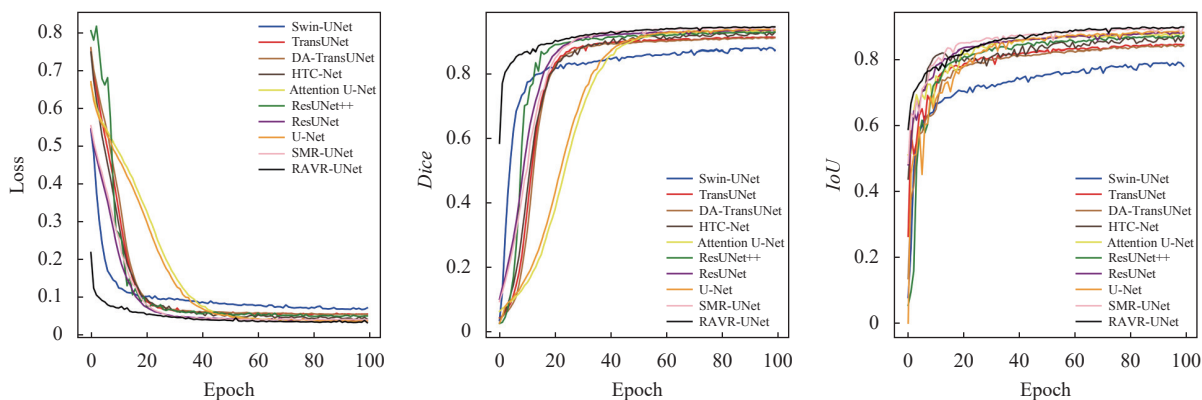


图 13 验证集对比实验曲线

3 结束语

本文针对目前肺结节图像分割研究中存在的问题, 基于 U-Net 提出一种改进的肺结节图像分割算法

RAVR-UNet, 并对该算法改进的各个结构进行说明. 实验结果表明, RAVR-UNet 在 LIDC-IDRI 公共数据集上表现出良好的分割性能. 然而, 由于实验硬件条件限制,

本研究基于 2D 切片数据训练模型。相较于三维数据,其忽略了切片间的关系,导致上下文信息丢失。因此,未来将重点研究三维肺结节图像分割,充分利用切片间的信息以实现更精确的分割效果。

参考文献

- 1 Liu CX, Pang MY. Lung CT image segmentation via dilated U-Net model and multi-scale gray correlation-based approach. *Circuits, Systems, and Signal Processing*, 2024, 43(3): 1697–1714. [doi: [10.1007/s00034-023-02532-x](https://doi.org/10.1007/s00034-023-02532-x)]
- 2 尹周一, 王梦圆, 游伟程, 等. 2022 美国癌症统计报告解读及中美癌症流行情况对比. *肿瘤综合治疗电子杂志*, 2022, 8(2): 54–63.
- 3 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 8th International Conference on Medical Image Computing and Computer-assisted Intervention–MICCAI 2015*. Munich: Springer, 2015. 234–241.
- 4 Zhou ZM, Rahman Siddiquee MM, Tajbakhsh N, *et al.* UNet++: A nested U-Net architecture for medical image segmentation. *Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada: Springer, 2018. 3–11.
- 5 Oktay O, Schlemper J, Folgoc LL, *et al.* Attention U-Net: Learning where to look for the pancreas. *arXiv:1804.03999*, 2018.
- 6 Jha D, Smedsrud PH, Riegler MA, *et al.* ResUNet++: An advanced architecture for medical image segmentation. *Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM)*. San Diego: IEEE, 2019. 225–2255.
- 7 Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of the 4th International Conference on 3D Vision (3DV)*. Stanford: IEEE, 2016. 565–571.
- 8 Huang HM, Lin LF, Tong RF, *et al.* UNet 3+: A full-scale connected UNet for medical image segmentation. *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona: IEEE, 2020. 1055–1059.
- 9 Selvadass S, Bruntha PM, Sagayam KM, *et al.* SATUNet: Series atrous convolution enhanced U-Net for lung nodule segmentation. *International Journal of Imaging Systems and Technology*, 2024, 34(1): e22964. [doi: [10.1002/ima.22964](https://doi.org/10.1002/ima.22964)]
- 10 Keetha NV, Samson ABP, Annavarapu CSR. U-Det: A modified U-Net architecture with bidirectional feature network for lung nodule segmentation. *arXiv:2003.09293*, 2020.
- 11 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 12 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*, 2020.
- 13 Carion N, Massa F, Synnaeve G, *et al.* End-to-end object detection with Transformers. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 213–229.
- 14 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 9992–10002.
- 15 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*, 2021.
- 16 Cao H, Wang YY, Chen J, *et al.* Swin-UNet: UNet-like pure Transformer for medical image segmentation. *Proceedings of the 2022 European Conference on Computer Vision*. Tel Aviv: Springer, 2022. 205–218.
- 17 Sun GQ, Pan YZ, Kong WK, *et al.* DA-TransUNet: Integrating spatial and channel dual attention with Transformer U-Net for medical image segmentation. *Frontiers in Bioengineering and Biotechnology*, 2024, 12: 1398237. [doi: [10.3389/fbioe.2024.1398237](https://doi.org/10.3389/fbioe.2024.1398237)]
- 18 Tang H, Chen YB, Wang T, *et al.* HTC-Net: A hybrid CNN-Transformer framework for medical image segmentation. *Biomedical Signal Processing and Control*, 2024, 88: 105605. [doi: [10.1016/j.bspc.2023.105605](https://doi.org/10.1016/j.bspc.2023.105605)]
- 19 Manzari ON, Kaleybar JM, Saadat H, *et al.* BEFUNet: A hybrid CNN-Transformer architecture for precise medical image segmentation. *arXiv:2402.08793*, 2024.
- 20 Guo XY, Lin X, Yang X, *et al.* UCTNet: Uncertainty-guided CNN-Transformer hybrid networks for medical image segmentation. *Pattern Recognition*, 2024, 152: 110491. [doi: [10.1016/j.patcog.2024.110491](https://doi.org/10.1016/j.patcog.2024.110491)]
- 21 Zhang HK, Lian J, Yi ZT, *et al.* HAU-Net: Hybrid CNN-Transformer for breast ultrasound image segmentation. *Biomedical Signal Processing and Control*, 2024, 87: 105427. [doi: [10.1016/j.bspc.2023.105427](https://doi.org/10.1016/j.bspc.2023.105427)]
- 22 Cao JM, Li YY, Sun MC, *et al.* DO-Conv: Depthwise over-

- parameterized convolutional layer. *IEEE Transactions on Image Processing*, 2022, 31: 3726–3736. [doi: [10.1109/TIP.2022.3175432](https://doi.org/10.1109/TIP.2022.3175432)]
- 23 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 24 Li X, Wang WH, Hu XL, *et al.* Selective kernel networks. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 510–519.
- 25 Han DC, Ye TZ, Han YZ, *et al.* Agent attention: On the integration of softmax and linear attention. *Proceedings of the 18th European Conference on Computer Vision*. Milan: Springer, 2023. 124–140.
- 26 Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. 2018. 833–851.
- 27 Armato III SG, McLennan G, Bidaut L, *et al.* The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 2011, 38(2): 915–931. [doi: [10.1118/1.3528204](https://doi.org/10.1118/1.3528204)]
- 28 Hou JC, Yan C, Li R, *et al.* Lung nodule segmentation algorithm with SMR-UNet. *IEEE Access*, 2023, 11: 34319–34331. [doi: [10.1109/ACCESS.2023.3264789](https://doi.org/10.1109/ACCESS.2023.3264789)]

(校对责编: 王欣欣)