

多尺度融合编码与自注意力的肺部CT分割算法

彭佑菊^{1*}, 徐杨^{1,2#}, 蒋诗怡¹, 熊举举¹

¹贵州大学大数据与信息工程学院, 贵州 贵阳

²贵阳铝镁设计研究院有限公司, 贵州 贵阳

收稿日期: 2023年7月19日; 录用日期: 2023年9月11日; 发布日期: 2023年9月18日

摘要

计算机断层扫描(CT)在当前是一种辅助检测肺炎的有效手段, 但病理表征的复杂性给医生诊断时带来不便, 难以准确地对图像进行分割。为进一步辅助医生根据病理表征诊断病情, 本文基于U-net提出了一种多尺度融合编码网络, 并结合自注意力机制, 力图在辅助医生判断的角度提供可行性方案。为了获取不同尺度的语义信息, 首先在编码器部分设计了一种多尺度融合编码器模块, 提取不同尺度的特征, 充分感知语义信息; 同时在编码器和解码器之间的跳连部分引入了改进的自注意力机制, 使得网络更好地关注不同语义特征的相关性; 最后, 采用融合Dice损失函数, Focal损失函数, 交叉熵损失函数构建的多级损失函数, 更好地约束训练。通过训练公开的数据集, 得到分割结果表明Dice相似系数、精确率(Precision)、召回率(Recall)分别为75.37%、77.03%、71.87%, 优于其他的模型。我们验证了改进的网络能够在图像分割过程中提升网络性能的可能性。

关键词

肺炎, 图像分割, 多尺度融合编码器, 自注意力机制, 深度学习

Lung CT Image Segmentation Algorithm with Multi-Scale Fusion Encoder and Self-Attention

Youju Peng^{1*}, Yang Xu^{1,2#}, Shiyi Jiang¹, Juju Xiong¹

¹College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

²Guiyang Aluminum-Magnesium Design and Research Institute Co. LTD., Guiyang Guizhou

Received: Jul. 19th, 2023; accepted: Sep. 11th, 2023; published: Sep. 18th, 2023

*第一作者。

#通讯作者。

文章引用: 彭佑菊, 徐杨, 蒋诗怡, 熊举举. 多尺度融合编码与自注意力的肺部 CT 分割算法[J]. 建模与仿真, 2023, 12(5): 4616-4630. DOI: 10.12677/mos.2023.125421

Abstract

Computed tomography (CT) is currently an effective means to assist in the detection of new coronary pneumonia, but the complexity of pathological representations brings inconvenience to doctors when diagnosing lesions, and it is difficult to accurately segment the images. In order to further assist doctors in diagnosis based on pathological representations, we propose a multi-scale fusion coding network based on U-net, combined with the self-attention mechanism, trying to provide a feasible solution from the perspective of assisting the doctor's judgment. In order to obtain semantic information of different scales, firstly, a multi-scale fusion encoder module is designed in the encoder part to extract features of different scales and fully perceive the semantic information; at the same time, a skip connection between the encoder and decoder is introduced. The improved self-attention mechanism makes the network pay more attention to the correlation of different semantic features; finally, the multi-level loss function of the fusion Dice loss function, Focal loss function, and cross-entropy loss function is used to better constrain training. Through training the public data set, the results show that the Dice similarity coefficient, precision rate and recall rate are 75.37%, 77.03%, and 71.87%, respectively, which are better than other existing network models. We have verified the feasibility of the improved network being able to improve network performance during image segmentation.

Keywords

Pneumonia, Image Segmentation, Multi-Scale Fusion Encoder, Self-Attention Mechanism, Deep Learning

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前为止, 全球累计肺炎确诊病例突破 4 亿例, 累计死亡病例超 600 万[1]。目前主要的肺炎筛查方法是逆转录聚合酶链反应(RT-PCR) [2], 但最近研究表明, RT-PCR 敏感性不够高, 导致 RT-PCR 筛查肺炎的阳性率并不十分理想, 有时并不能满足临床的需要。在这种情况下, 可以采用断层扫描(computed tomography, CT)和磁共振成像(magnetic resonance imaging, MRI) [3]检测来进一步诊断患者是否患病, 一般情况下, 医生可以通过患者的 CT 影像图来判断病人是否感染肺炎[4]。最近研究表明, 胸部 CT 对于分析肺炎的敏感度高达 0.97 [5], 非常有利于肺炎的筛查, 但是 CT 扫描费时低效, 还需要专业医生进行辅助判断, 需要耗费巨大的人力财力。

随着语义分割技术的快速发展, 不同的神经网络模型在医疗影像领域中也得到了广泛应用, 卷积神经网络的(Convolutional Neural Network, CNN)成功应用于图像分割领域[6], 为图像特征提取提供了全新的解决方法。2015 年 Long 等[7]提出了全卷积网络(Fully Convolutional Network, FCN), 使用卷积层替换全连接层, 使图像达到像素级分类, 从而实现语义级别的分割效果。Ronneberger 等[8]提出了 U-Net, 采用了编码器 - 解码器的结构, 编码器通过卷积与池化操作对输入图像进行下采样编码, 解码器通过上采样操作将编码特征映射成掩模图, 得到语义分割结果。Zhou 等人[9]受 U-Net 启发提出了 U-Net++模型, 通过增加编码器与解码器之间的细粒度信息重新设计了跳连模块。与 FCN 相比 U-Net 作为骨干网络分割

效果提升明显,但存在着特征提取不充分、不同层级语义信息利用率低等现象。文献[10]中的 Attention-Unet 模型,就是将注意力机制引入到解码器部分,从而将注意力集中在需要提取的区域;文献[11]中提出的 R2U-Net,将 RNN 网络和 ResNet 网络的结构整合到了编码器-解码器结构里。文献[12]中提出 Inf-Net 从肺部 CT 图片中自动分割感染区域,利用并行部分解码器(Parallel Partial Decoder, PPD)用于聚合高级特征(结合上下文信息)并生成全局图,作为后续步骤的初始指导区域。为了进一步挖掘边界线索,利用一组隐式重复的反向注意力(Reverse Attention, RA)模块和显式边缘注意引导来建立区域和边界的关系。文献[13]在 U-Net 的基础上,在编码器模块增加了残差网络,使用多尺度的图像输入,同时在编解码之间增加了空洞空间金字塔池化(ASPP),用来提取上文的多尺度信息,最后将提取到的多尺度信息与编码部分级联起来,结合注意力机制来提高图像分割的精度。文献[14]提出的是一种改进 U-Net 网络的肺结节分割方法。该方法替换上采样操作,最小化像素点与标签的损失,并融合了通道注意力空间注意力机制,增加全局相关性,能够准确地分割出肺结节区域,具有良好的分割性能。文献[15]提出一种基于低秩张量自注意力重构的语义分割网络 LRSAR-Net,其中低秩张量自注意力重构模块用来获取长范围的信息,低秩张量自注意力模块先生成多个低秩张量,构建低秩自注意力特征图,然后将多个低秩张量注意力特征图重构成高秩注意力特征图,通过自注意力模块计算相似度矩阵来获取长范围的语义信息。文献[16]提出了 MiniSeg,通过较少的参数量构建图像分割网络模型。但上述方法依然存在着一些问题,如准确率不高,分割结果边界不明显,没有充分整合网络高低层特征信息等特点。

近年来,肺部分割工作发展迅速,但由于肺部 CT 以及病灶的特殊性,如标注的 CT 图像较少、肺部 CT 图像组织结构复杂、病灶部位分布不均、轮廓不明显、CT 切片中感染病灶的大小和位置多变等,会造成梯度爆炸、特征利用率低等问题,给模型的训练增加难度,难以提取有效特征,分割精确性低,大部分肺炎公开数据都集中在诊断上,只有极少部分数据集提供了分割标签[17]。如果仅仅使用原始 U-Net 对其训练,存在梯度消失、特征利用率低等问题,最终导致模型的分割准确率难以提高。为了解决上述问题,本文在 U-net 基础上提出了一种多尺度融合编码网络(Multi-Scale Fusion Encoder Net, MSFENet),该网络在编解码器的基础上设计了多尺度融合编码模块,充分利用不同尺度的信息;同时为了进一步解决梯度消失以及上下文特征利用不全的问题,在跳连部分添加改进的自注意力模块;最后为了加速网络收敛使用了多级损失函数约束训练。本文的贡献如下:

- 1) 本文在编码器模块提出了多尺度融合编码模块,融合了编码部分的高低层信息,该编码器能保留网络模型中低级轮廓信息和高级语义信息,有助于充分感知深入的特征,为后续的信息提取提供帮助。
- 2) 本文在编码器和解码器的跳连部分添加了改进的自注意力模块,减少了对外部信息的依赖,捕捉特征的内部相关性,并解决梯度爆炸即全局上下文信息提取不全的问题。
- 3) 最后,本文将 Dice 损失函数、Focal 损失函数、交叉熵损失函数进行融合构建多级损失函数,从而解决正负样本分布不均以及肺部 CT 的相关问题,更好地约束网络训练。

2. 本文方法

2.1. 网络结构

本文的网络整体框架如图 1 所示,网络主要包括 3 个部分:编码器、解码器和自注意力模块。其中在编码器部分,本文采用残差模块作为基础网络模块,在网络高层部分采用多尺度融合编码融合不同尺度的语义特征信息,并将融合后的映射特征图输入给后续解码器部分,同时为了防止下采样过程中特征信息的丢失,编码器和解码器的跳连之间引入了改进的自注意力机制,去关注不同尺度的特征信息,更好的细化特征映射质量,融合不同阶段的语义信息。

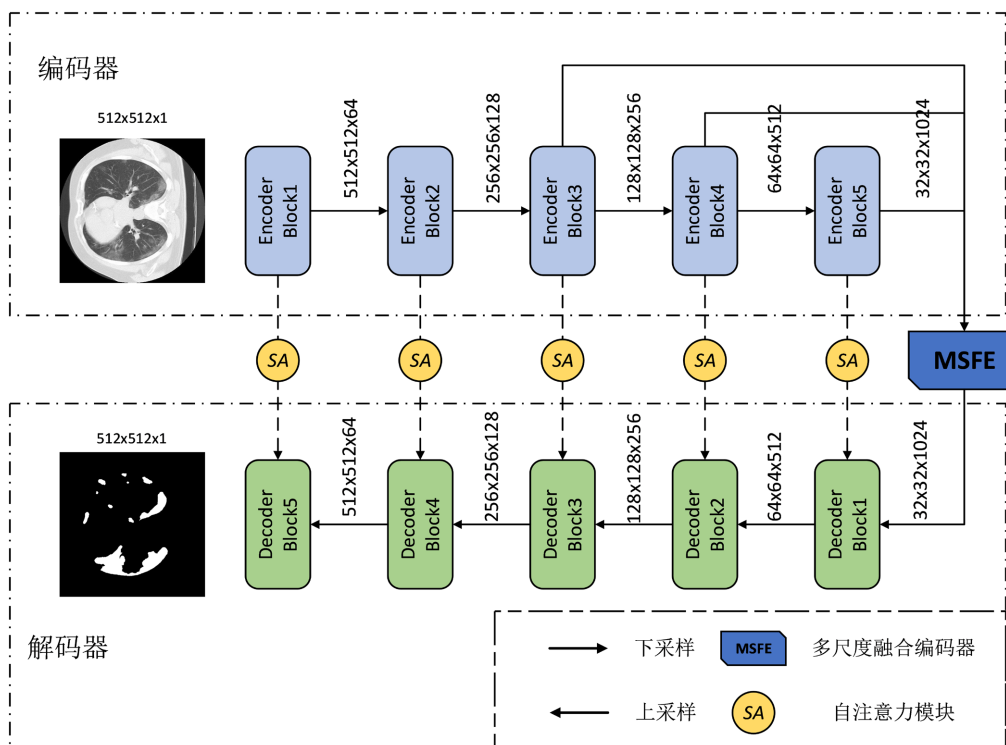


Figure 1. Network structure of model
图 1. 网络模型结构

2.2. 编码器解码器模块

编码器 - 解码器结构是当前语义分割领域的流行框架，其中最常见 U-Net 网络也是许多新网络的设计基础[18] [19]，它能够实现端到端的像素级分割。编码器作为整个网络的特征提取部分，将原始输入图像编码成低分辨率的特征图，解码器则负责将特征图还原成原始分辨率的像素分类预测图。

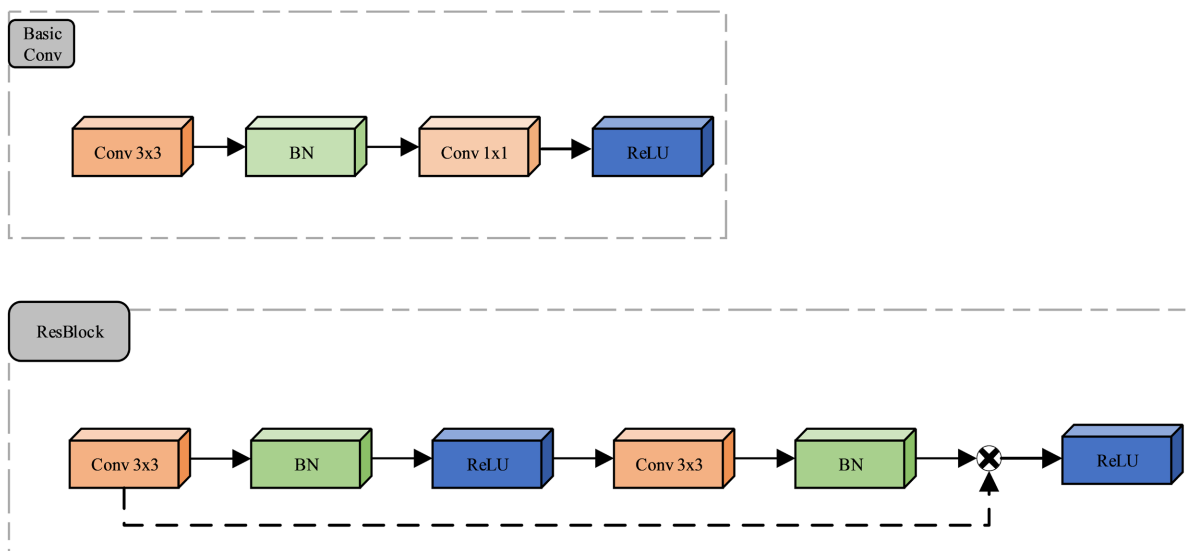


Figure 2. Residual module structure
图 2. 残差模块结构

本文采用残差模块[20]作为基础网络模块。早期深度学习普遍认为网络深度越深,模型的表现就越好,因为如果神经网络越深就能够提取到更高级的语义信息。但实验发现网络越深越难训练,并且浅层网络和深层网络的着重点也有所不同。而残差网络的提出使得特征映射对输出的变化更加敏感,能够保留编码部分中不同层级丢失掉的信息。本文提出的 MSFENet 网络是通过重新设计传统的 U-Net 所得的,该网络结构包括下采样路径和上采样路径,下采样路径能够增大接受域同时也能节约计算成本,而上采样路径就是在恢复下采样路径中图片缩小的分辨率。编码器的残差模块结构如图 2 所示,它包含一个基础的卷积运算(Basic Conv)和一个残差模块(ResBlock),其中基础的卷积运算包括卷积层、批量归一化(Batch Normalization)和线性整流单元。解码器部分由五个解码器模块组成,其中模块的结构与编码器模块类似,将编码器模块的下采样部分换成上采样,通道数随着分辨率上升而减少,通过上采样还原分辨率。

2.3. 多尺度融合编码模块

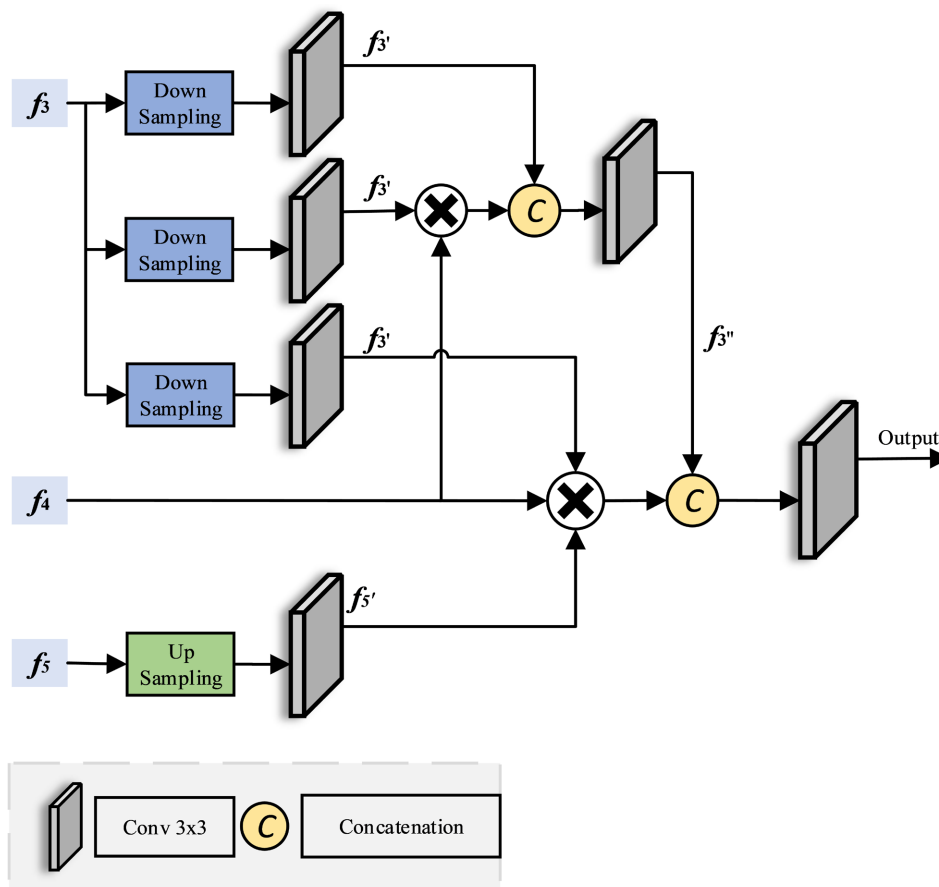


Figure 3. Multi-scale fusion encoder
图 3. 多尺度融合编码器

低层网络的感受野较小,小目标细节信息和边缘信息较多,但缺少丰富的语义信息;高层网络的感受野较大,有丰富的语义信息,但包含的小目标信息较少。因医学影像中待分割的目标区域尺度变换大、形状不规则、边界模糊,以及目标区域与周围组织强度不均匀等因素[21],只学习单一尺度的图像特征可能会导致一些重要的上下文信息被忽略,从而导致网络的分割效果不佳。而 CT 图像中不仅包含着许多细节特征,而且也富含高级语义信息,所以为了更好地融合网络中低层的边缘信息和高层语义信息,本

文设计了多尺度融合编码模块，将低层的特征添加至相邻高层信息中进行融合，同时也将高层的特征添加到相邻的低层组合成新的特征，以此来聚合不同尺度的信息，对于输入的肺部 CT 图像，首先在提取编码器模块第 3、4、5 层($f_i, i=3,4,5$)的高层信息后通过多尺度融合编码融合信息之后输出。如图 3 所示，网络的第 3 层的编码器通过一个下采样再通过一个 3×3 的卷积后输出得到 f'_3 ，再通过卷积和拼接操作将 f'_3 和 f_4 的特征融合在一起得到 f_3'' ，然后将第 5 层的编码器输出信息进行上采样后通过卷积得到 f'_5 ，之后与第 4 层编码器和 f_3'' 的输出信息进行融合，最后通过拼接和卷积操作得到输出特征图 Output，公式如式(1)~(4)。

$$f'_3 = \text{Conv}(\text{Down}(f_3)) \quad (1)$$

$$f'_5 = \text{Conv}(\text{Down}(f_5)) \quad (2)$$

$$f_3'' = \text{Conv}\left(\text{Concat}\left(f_3', (f'_3 \otimes f_4)\right)\right) \quad (3)$$

$$\text{Output} = \text{Conv}\left(\text{Concat}\left(f_3'', (f_4 \otimes f'_5)\right)\right) \quad (4)$$

通过多尺度融合编码器，可以从多个感受野中学习丰富的上下文信息，而且可以根据不同尺度的特征分配不同的学习权重，从而更好的融合不同的语义与边缘轮廓信息。

2.4. 改进的自注意力机制

注意力机制最早在自然语言处理和机器翻译对齐文本中提出并使用，后面逐渐也应用到计算机视觉领域，证实了在视觉和卷积神经网络中引入注意力机制可以显著地提升网络的性能。本文选择在网络中引入了改进的自注意力机制(self attention) [22]，自注意力机制是 transformer [23]中的重要组成部分，自注意力机制的引入是为了捕捉输入序列各项之间的依赖关系，尤其是捕捉那些在序列模型中很容易丢失掉的长范围依赖关系，这种依赖关系在计算机视觉中非常重要。改进的注意力机制如图 4 所示，它表示从集合 x 到集合 y 的映射，一般模型如式(5)所示，

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \quad (5)$$

其中， x 为模型的输入， y 表示为一个与输入具有相同尺寸的输出图像， i 和 j 均为两个集合的元素下标， $f(x_i, x_j)$ 表示作用在输入序列第 i, j 两个位置上的标量函数，即权重系数； $g(x_j)$ 是作用在输入中第 j 个位置上的某种变换函数； $C(x)$ 为归一化函数。

在此基础上引入残差块，如式(6)所示：

$$Z_i = W_z y_i + p(x_i) \quad (6)$$

如上式所示，得到输出 y_i 后，首先对其进行线性变换 $W_z y_i$ ，然后再与 x_i 进行合成，再加上 $p(x_i)$ 表示池化操作残差连接，之后得到输出特征 Z_i 。

设输入的特征图维度为 $H \times W \times d_v$ ，其中 H, W 和 d_v 分别表示每张特征图的高度、宽度以及通道数(通道数也是特征维度)，整个处理过程如下：

- 1) 对输入序列分别进行三个 1×1 的卷积操作(输出通道为 $\frac{d_v}{2}$)，得到三个尺寸为 $H \times W \times \frac{d_v}{2}$ 的特征图；
- 2) 对三个特征图进行“reshape”操作得到三个特征图的矩阵表示，将这三个矩阵分别记作 Q, K 和 V ，其中 Q, K, V 表示 $W_\theta x, W_\phi x$ 和 $W_g x$ ；

- 3) 利用 softmax (QTK)计算归一化的注意力权重系数;
- 4) 将注意力权重系数作用到 V 上, 即计算 softmax (QTK) V, 得到一个 $H \times W \times \frac{d_v}{2}$ 的矩阵;
- 5) 将第 4)步得到的矩阵进行形状变换, 得到一个 $H \times W \times \frac{d_v}{2}$ 维张量, 将该张量记为 y;
- 6) 对张量 y 进行 $1 \times 1 \times 1$ 的卷积操作(输出特征图通道数为 d_v), 得到一个 $H \times W \times d_v$ 维张量, 该张量即为 $W_2 y$, 将输出特征图的维度恢复至原始维度, 然后再将其与池化处理操作后的 x 相加得到输出 Z。

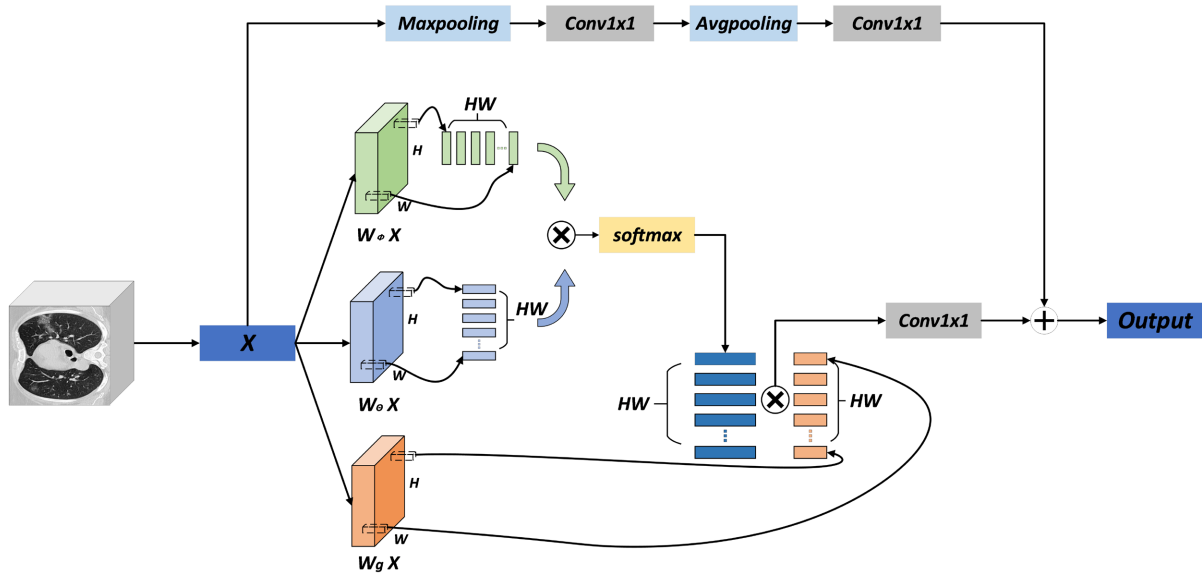


Figure 4. Improved self-attention mechanism
图 4. 改进的自注意力机制

这样一来引入的改进的自注意力机制不仅可以保证输入与输出的特征图具有相同的尺寸, 而且它更好的关注某一个像素与其他位置像素的关系, 从而更好的把握全局特征。

2.5. 多级损失函数

在语义分割过程中, 常用的损失函数是交叉熵损失函数和 Dice 损失函数。但是由于肺部 CT 和病灶的特殊性, 本文采用了一种 Dice 损失函数、Focal 损失函数以及交叉熵损失函数相结合的多级损失函数。Dice 损失函数关注的点在于标签部分和预测部分的重合情况, 在样本标签不均的情况下也表现得很好, Dice 损失函数定义如式(7)所示。Focal 损失函数是基于二分类交叉熵(BCE)改进而来的, 该损失函数降低了大量简单负样本在训练中所占的权重, 可以解决训练中正负样本比例严重失衡的问题, Focal 损失函数定义如式(8)所示。交叉熵损失函数可以表征真实样本标签和预测概率之间的差值, 定义如式(9)。

$$L_{Dice} = 1 - \frac{2 \times (\text{pred} \cap \text{true})}{\text{pred} \cup \text{true}} \quad (7)$$

$$L_{FL} = -\alpha_i (1 - p_i)^{\gamma} \log(p_i) \quad (8)$$

$$L_{CE} = -\alpha_i \log(p_i) \quad (9)$$

其中, pred 表示预测值的集合, true 表示真实值的集合, α_i 是类别权重参数, $\alpha_i \in [0, 1]$, $(1 - p_i)^{\gamma}$ 是样

本难度权重因子, γ 是聚焦参数, p 表示预测值与真实值相等的概率, 当预测值与真实值相等时, p_i 与 p 值相等, 否则 $p_i = 1 - p$ 。

综合上述函数的优缺点, 本文采用三种损失函数结合的多级损失函数来解决本文梯度饱和、样本不均衡以及肺炎 CT 的相关问题。即

$$L = L_{Dice} + L_{FL} + L_C \quad (10)$$

以此来更好的约束网络训练。

3. 实验结果及分析

3.1. 数据集介绍

本文的数据集其中一个是由意大利医疗和介入放射协会收集[24]的肺炎-CT-Seg 数据集, 该数据集包含超过 40 名肺炎患者的 100 张 CT 图像, 另一个数据集是由 Ma 等创建[25]的肺炎-CT-Seg-Benchmark 数据集, 其中包含 3520 张 20 例标记的肺炎患者肺部三维 CT 影像, 切片大小为 512×512 像素。如图 5 所示, 分别展示了 a、b、c 三组数据集里边的 CT 图像和标签, 其中: 第一行表示肺部 CT 切片图像, 第二行表示医生手工标注相应的 CT 图像的金标准。在本文的实验中, 首先将两个数据集的 3620 张 CT 图像混合在一起, 从混合数据集中随机挑选 50 张 CT 图像进行测试, 其余的用作训练验证。

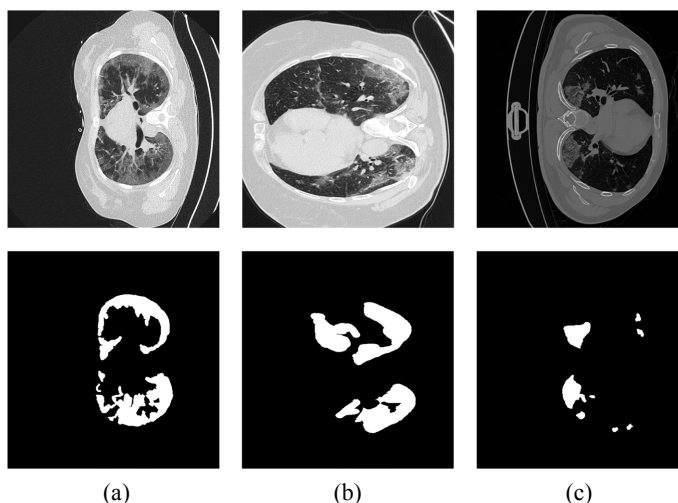


Figure 5. CT image and gold standard
图 5. CT 图像与金标准

3.2. 数据预处理

数据预处理包括图像增强和图像裁剪变换。首先采用一种非常经典的直方图均衡化算法——CLAHE 算法(Contrast Limited Adaptive Histogram Equalization)来进行数据增强, CLAHE 算法的主要作用是增强图像的对比度同时能够抑制噪声; 然后对图像进行了裁剪, 裁掉图像中多余的黑边部分; 最后对图像做了仿射变换, 对图像整体放缩百分之二十, 旋转图像变换角度正负 15 度。通过以上操作, 有利于后续的特征提取和网络训练。如图 6 所示, 图 6 表示原图与通过 CLAHE 算法进行图像增强的图像对比、经过裁剪后的图像对比以及预处理前后数据集的 CT 值分布直方图对比(其中蓝色部分代表原始图像的 CT 值分布, 橙色代表经过数据增强后的 CT 值分布), 图 7 表示图像增强和裁剪后再进行仿射变换的 CT 图像与标注金标准。

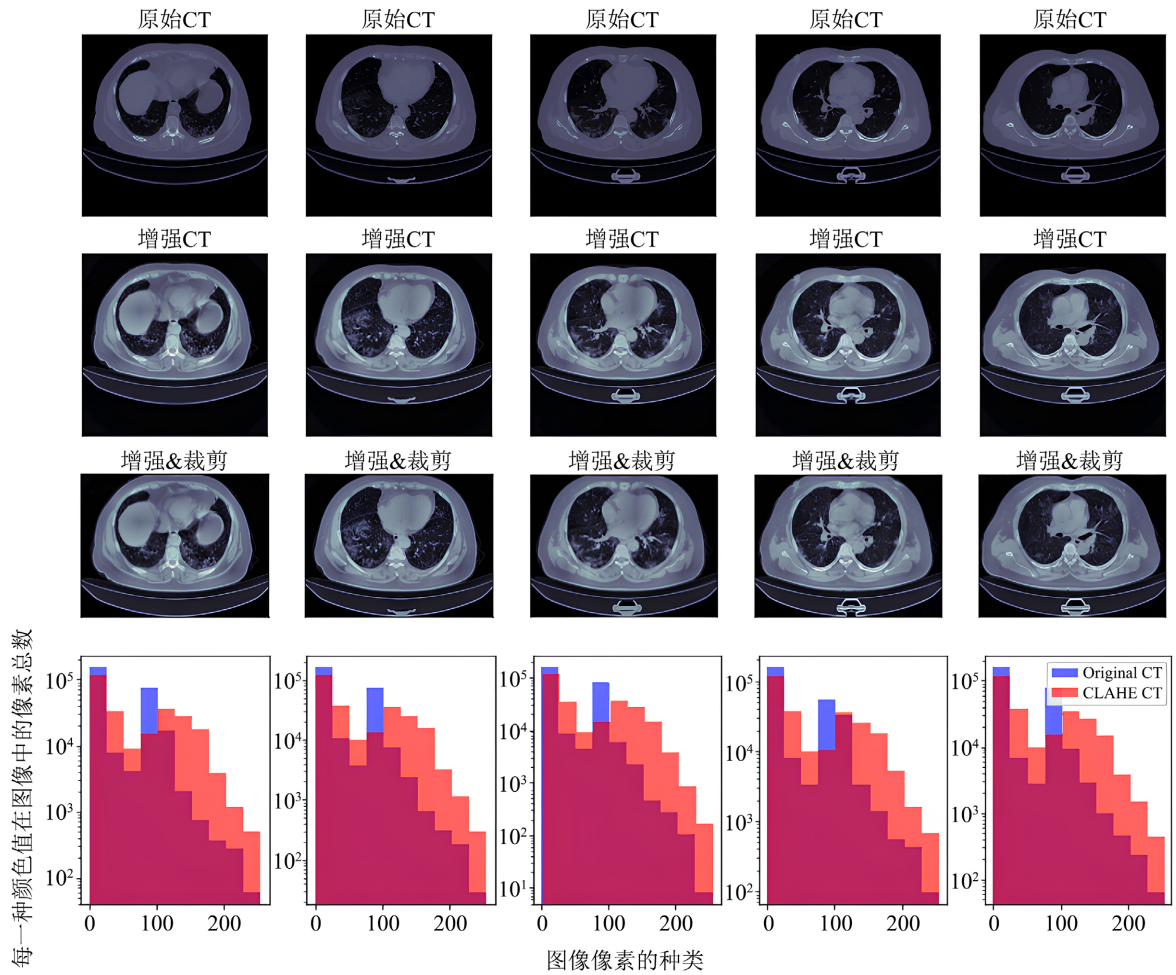


Figure 6. Image enhancement, cropping and distribution histogram

图 6. 图像增强、裁剪与分布直方图

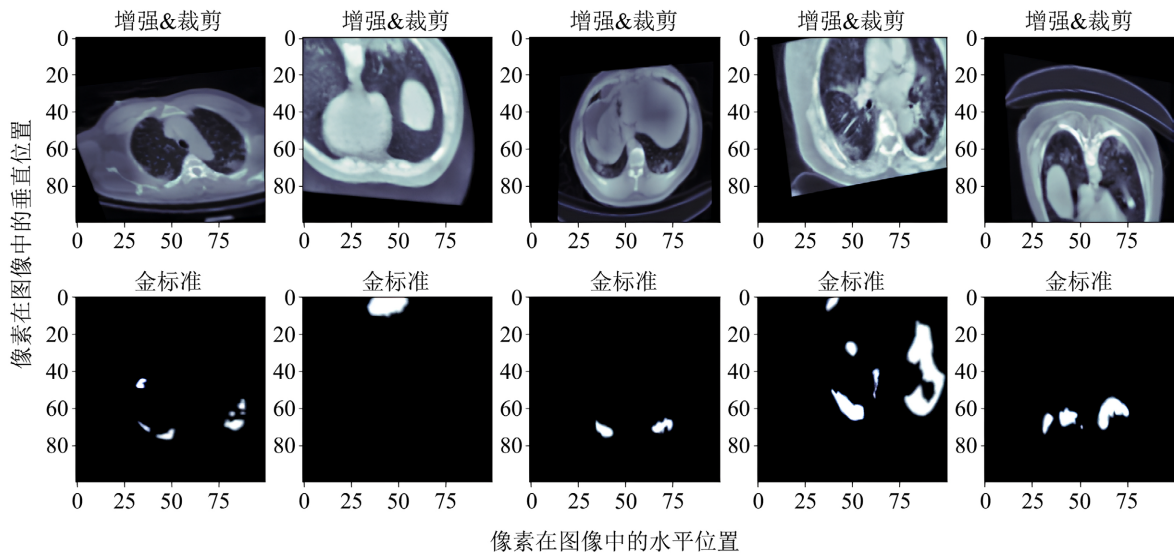


Figure 7. CT image after affine transformation and corresponding gold standard

图 7. 仿射变换后的 CT 图像及对应的金标准

3.3. 实验参数设置

本文实验的环境选择 CPU 为 Intel i7-12700, 内存为 64G, 使用 NVIDIA GeForce RTX 3090 GPU。梯度下降法训练模型, 同时选用 PyTorch 的深度学习框架, 学习率设置为 $5e-4$, 优化器选择 Range 优化器来更新参数。如图 8, 图中展示了训练过程中训练集和验证集 epoch 与相应的 loss 之间的关系, 可以看出, 随着网络迭代次数的增加, 训练集和验证集的损失函数逐渐降低, 当样本的训练次数到达 100 之后, 验证集的损失函数趋近于稳定, 所以本文将网络的训练的 epoch 设置为 100, 同时将网络的批次大小设置为 4, 方便网络训练。

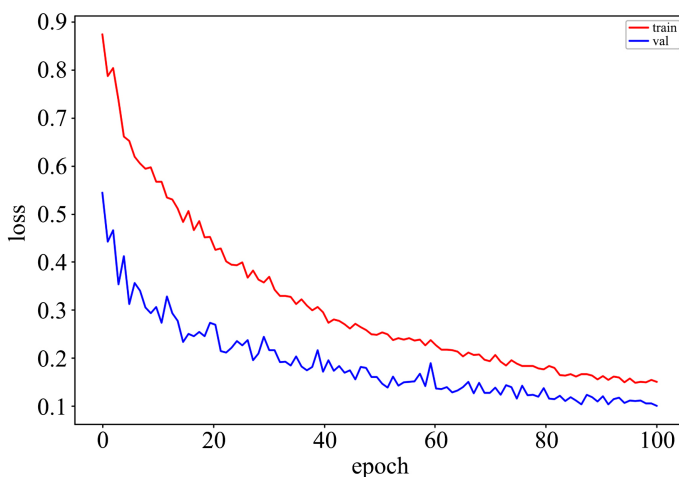


Figure 8. Training set and validation set loss function
图 8. 训练集与验证集损失函数

3.4. 评价指标

本文采用常见的 3 个医学影像分割评价指标对模型的性能进行评估, 这些评价指标分别是 Dice 相似性系数(dice similarity coefficient, DSC)、精确率(Precision)和召回率(Recall)。计算公式分别如下:

$$DICE = \frac{2TP}{2TP + FP + FN} \quad (11)$$

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

其中, TP 表示真正例(true positive), 即算法预测的结果与人工分割的结果一样的像素个数; FP 表示假正例(false positive), 即在算法分割区域内, 在人工分割区域以外的像素个数; FN 表示假反例(false negative), 在图像分割中表示在人工分割区域内, 在算法分割区域外的像素个数。

3.5. 实验结果

3.5.1. 消融实验

为了证实本文改进的模块可以有效提升模型的分割性能, 本文做了 3 组对比实验来说明本文增加的多尺度融合编码模块(MSFE)和自注意力模块(self-attention)可以在不同的程度上提升对 CT 图像的分割能力, 如表 1 所示。

Table 1. Ablation experiments**表 1.** 消融实验

Model	DSC/%	Precision/%	Recall/%
原始网络	70.55	72.01	67.39
原始网络 + MSFE	73.87	76.03	68.97
原始网络 + attention	71.42	74.09	69.85
MSFENet (本文方法)	75.37	77.03	71.87

表中的原始网络是指不额外添加其他模块的网络结构, 由表可知, 添加了多尺度融合解码器模块 (MSFE) 之后, 网络的性能得到了提升, Dice 相似性系数 (dice similarity coefficient, DSC) 从原始网络的 70.55% 增加至 73.87%, 精确率 (Precision) 从原始网络的 72.01% 增加至 76.03%, 召回率 (Recall) 从 67.39% 增加至 68.97%。同时, 在原始网络结构上增加了自注意力机制后, Dice 相似性系数、精确率和召回率分别由 70.55% 增加至 71.42%、72.01% 增加至 74.09%、67.39% 增加至 69.85%。同时增加多尺度融合解码器和自注意力机制之后, 网络的 Dice 相似性系数、精确率和召回率在原始网络的基础上依次提升了 4.82%、5.02%、2.48%, 说明本文设计的相关模块不仅可以整合不同尺度的特征信息, 还能够更好地关注特征的内部相关性, 从而提升网络的性能。

3.5.2. 对比实验

为了更好的证明本文改进的模块可以有效的提升分割性能, 我们将本文提出的方法与传统的一些图像分割模型进行比较, 如 FCN、SegNet [26]、U-Net、Inf-Net、CAREs-Unet [27], 比较结果如表 2。

Table 2. Comparison of different segmentation algorithms**表 2.** 不同分割网络对比

Model	DSC/%	Precision/%	Recall/%
FCN [7]	63.51	65.33	60.87
SegNet [26]	65.58	67.81	62.04
U-Net [8]	68.14	70.05	64.99
Inf-Net [12]	72.35	73.66	68.86
CAREs-Unet [27]	73.54	75.43	70.02
MSFENet (ours)	75.37	77.03	71.87

由表中数据可知, 本文的实验方法的 Dice 相似系数 (DSC)、精确率 (Precision)、召回率 (Recall) 为 75.37%、77.03%、71.87%, 由表可知本文改进的网络模型在图像分割的过程中性能优于上述的其他分割网络。得益于本文增加的多尺度融合编码器和自注意力机制, 与传统网络相比, 本文的网络模型性能提升幅度较大; 与 Inf-Net 相比, 本文的网络在 Dice 相似性参数、精确率、召回率上分别提升了 3.02%、3.37%、3.01%; 与 CAREs-Unet 相比, 本文改进的网络在 Dice 相似性参数、精确率、召回率上分别提升了 1.83%、1.6%、1.85%; 因为本文在基础网络上增加了多尺度融合解码器模块和引入了自注意力机制, 与其他网络相比,

网络的性能在不同程度上都得到了提升，验证了本文改进的网络能够在图像分割过程中提升网络性能的可行性。

如图 9 所示，本文展示了 6 个不同的肺部 CT 图像分割结果的可视化，并且与当前先进的两种肺炎分割方法结果进行比较。

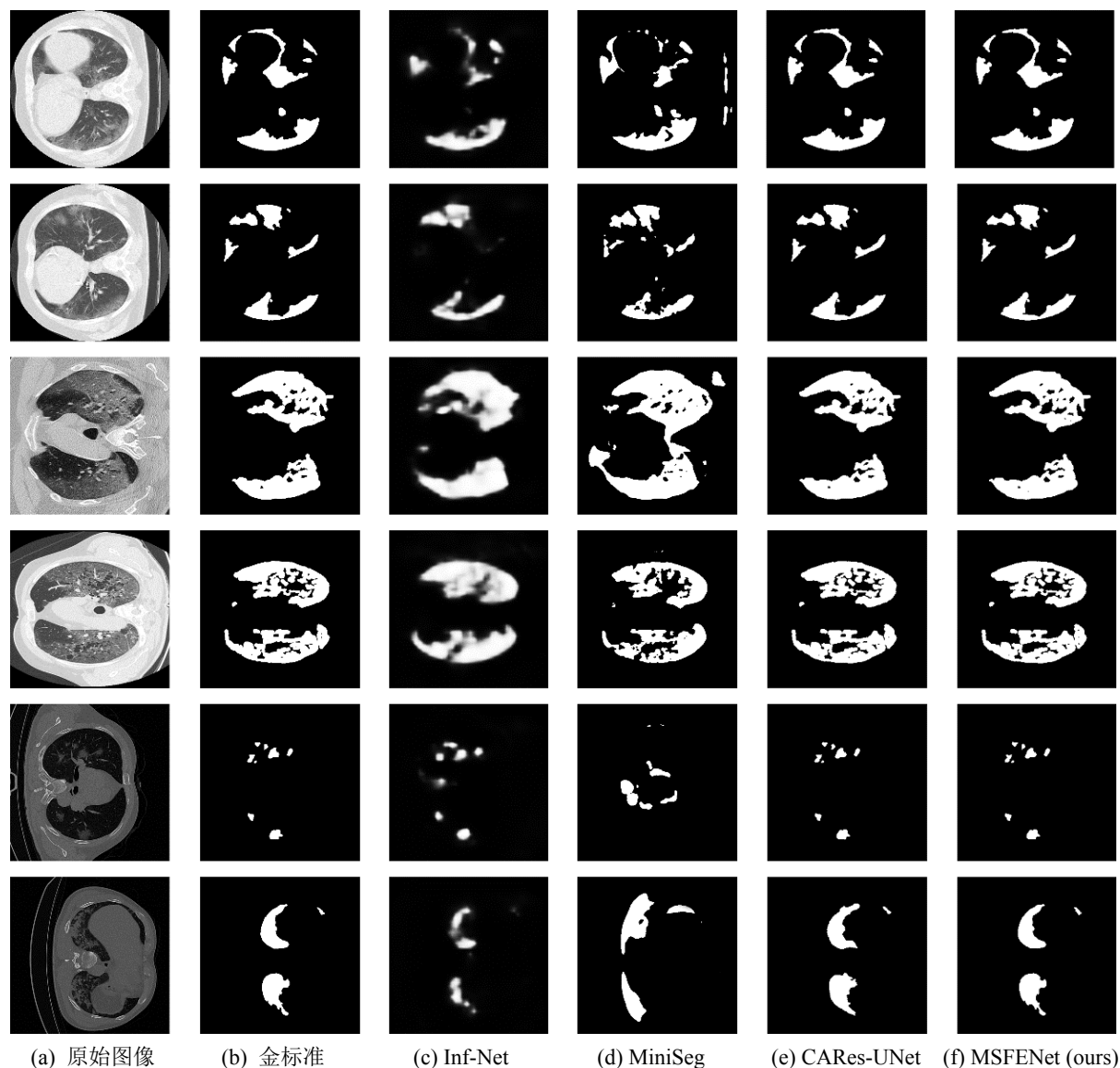


Figure 9. Segmentation results of different algorithms

图 9. 不同算法的分割结果

其中，图中的图 9(a)表示的是作为输入的肺部 CT 图像，图 9(b)是数据集给出的金标准标注，图 9(c)是 Inf-Net 模型输出的分割结果，图 9(d)是通过 MiniSeg 模型训练后的分割结果，图 9(e)是 CARes-UNet 模型输出的分割结果，图 9(f)是使用本文改进的模型对数据集的分割结果。得益于本文改进的多尺度融合解码器和引入的自注意力机制，与 Inf-Net 模型的训练结果相比，本文改进的网络训练结果图像更清晰，对于图像的边缘处理得更好；MiniSeg 网络虽然轮廓相对清晰，但错分区域较多，相比之下本文的模型的

分割结果更接近于金标准；与 CAREs-Net 网络的训练结果相比，得益于设计的多尺度融合编码器模块，本文改进的模型在 CT 图像的分割上对于图像中的微小细节处理得更好。从图 9 中可以看到，本文改进的网络对于图像中的细节处理得更好，精确率也更高，边缘轮廓更加突出，接近于金标准图像。说明了本文改进的方法在肺炎病灶分割上确实更精确。不足之处是本文模型与其他分割模型一样，尽管对病灶分割的性能有一定的提升，但都具有在模糊边界上的分割效果不佳的缺陷。

3.5.3. 不同损失函数的对比实验

在网络训练中，不同的 Loss 函数会对网络的分割性能产生不同的影响，为了选择合适的 Loss 函数，对不同 Loss 函数对网络分割的性能进行了比较，如表 3。表中列出了不同损失函数下的网络分割效果，因为不同的 Loss 函数具有各自的缺陷，例如：1) 交叉熵损失函数的缺点是当医学图像中的类别不均衡时，训练就会被像素较多的类主导，无法学习到较小的类的特征；2) Focal 损失函数具有需要手动调节参数的缺点；3) Dice 损失函数在极端情况下(网络预测接近 0 或 1 时)，对应点的梯度值会极小，Dice Loss 会存在梯度饱和的情况。因此在模型训练中选择 Dice 损失函数、Focal 损失函数以及交叉熵损失函数融合的多级损失函数，有利于解决正负样本不均衡、难易样本不平衡、梯度饱和等问题，如表中结果所示，选择融合的多级损失函数用于网络训练，可以使得网络的性能得到明显提升。

Table 3. Loss function comparison
表 3. 损失函数对比

Loss	DSC/%	Precision/%	Recall/%
Dice	72.41	73.96	68.85
CE + Focal	72.05	73.24	69.02
Dice + Focal	74.82	75.87	69.01
Dice + Focal + CE (ours)	75.37	77.03	71.87

4. 结束语

本文提出了一种多尺度融合编码网络(MSFENet)，通过融合不同尺度的特征，更好地保留网络中的低层轮廓信息和高层语义信息，同时在编码器和解码器的跳连部分引入了自注意力机制，使模型能够更好地捕捉数据或者特征的内部相关性，最后融合 Dice 损失函数，Focal 损失函数，交叉熵损失函数作为多级损失函数，更好地约束网络训练。实验结果表明，本文改进的图片分割方法在 Dice 相似系数、精确率(Precision)、召回率(Recall)上达到 75.37%、77.03%、71.87%，与其他网络相比，具有更好的分割效果。本文设计的网络模型能够提升对肺炎肺部 CT 图像的分割效果，为辅助肺炎的检测与治疗提供了一种新的方法和思路，在未来的研究中，为了解决可移动设备部署的问题，我们将对网络进行轻量化处理。

基金项目

贵州省科技计划项目(黔科合支撑[2023]一般 326)。

参考文献

- [1] Thompson, R. (2020) Pandemic Potential of 2019-nCoV. *The Lancet Infectious Diseases*, **20**, 280. [https://doi.org/10.1016/S1473-3099\(20\)30068-2](https://doi.org/10.1016/S1473-3099(20)30068-2)

- [2] 唐江平, 周晓飞, 贺鑫, 等. 基于深度学习的新型冠状病毒肺炎诊断研究综述[J]. 计算机工程, 2021, 47(5): 1-15.
- [3] Zhang, Z., Romero, A., Muckley, M.J., *et al.* (2019) Reducing Uncertainty in Undersampled MRI Reconstruction with Active Acquisition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 2049-2058. <https://doi.org/10.1109/CVPR.2019.00215>
- [4] Rubin, G.D., Ryerson, C.J., Haramati, L.B., *et al.* (2020) The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Radiology*, **296**, 172-180. <https://doi.org/10.1148/radiol.2020201365>
- [5] Ai, T., Yang, Z., Hou, H., *et al.* (2020) Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, **296**, E32-E40. <https://doi.org/10.1148/radiol.2020200642>
- [6] Girshick, R., Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [7] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [8] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M. and Frangi, A.F., Eds., *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [9] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., *et al.* (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al.*, Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [10] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas.
- [11] Alom, M.Z., Hasan, M., Yakopcic, C., *et al.* (2018) Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation.
- [12] Fan, D.P., Zhou, T., Ji, G.P., *et al.* (2020) Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. *IEEE Transactions on Medical Imaging*, **39**, 2626-2637. <https://doi.org/10.1109/TMI.2020.2996645>
- [13] 钱宝鑫, 肖志勇, 宋威. 改进的卷积神经网络在肺部图像上的分割应用[J]. 计算机科学与探索, 2020, 14(8): 1358-1367.
- [14] 王馨, 强彦, 杨晓棠, 等. 基于双注意力 3D-UNet 的肺结节分割网络模型[J]. 计算机工程, 2021, 47(2): 307-313.
- [15] 张桃红, 郭徐徐, 张颖. LRSAR-Net 语义分割模型用于肺炎 CT 图片辅助诊断[J]. 电子与信息学报, 2022, 44(1): 48-58.
- [16] Qiu, Y., Liu, Y., Li, S., *et al.* (2020) Miniseg: An Extremely Minimum Network for Efficient COVID-19 Segmentation.
- [17] 郭宇, 李瑞冰, 刘莎, 等. 基于机器学习的肺癌图像辅助诊断应用研究[J]. 中国医学装备, 2021, 18(3): 124-128.
- [18] 杨莉, 万旺根. 基于预训练-微调策略的肺炎预测模型[J]. 计算机工程, 2022, 48(3): 17-22.
- [19] Raj, A.N.J., Zhu, H., Khan, A., *et al.* (2021) ADID-UNET—A Segmentation Model for COVID-19 Infection from Lung CT Scans. *PeerJ Computer Science*, **7**, e349. <https://doi.org/10.7717/peerj-cs.349>
- [20] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [21] 王雪, 李占山, 吕颖达. 基于多尺度感知和语义适配的医学图像分割算法[J]. 吉林大学学报(工学版), 2022, 52(3): 640-647.
- [22] Wang, X., Girshick, R., Gupta, A., *et al.* (2018) Non-Local Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. <https://doi.org/10.1109/CVPR.2018.00813>
- [23] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [24] 刘锐, 丁辉, 尚媛园, 等. 肺炎医学影像数据集及研究进展[J]. 计算机工程与应用, 2021, 57(22): 15-27.
- [25] Ma, J., Wang, Y., An, X., *et al.* (2021) Toward Data-Efficient Learning: A Benchmark for COVID-19 CT Lung and Infection Segmentation. *Medical Physics*, **48**, 1197-1210. <https://doi.org/10.1002/mp.14676>
- [26] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture

for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>

- [27] Xu, X., Wen, Y., Zhao, L., *et al.* (2021) CAREs-UNet: Content-Aware Residual UNet for Lesion Segmentation of COVID-19 from Chest CT Images. *Medical Physics*, **48**, 7127-7140. <https://doi.org/10.1002/mp.15231>