

RM Transunet: 基于小样本数据的肺癌脑转移瘤分割

杨玉婷*, 祝汉灿#

绍兴文理学院数理信息学院, 浙江 绍兴

收稿日期: 2024年3月26日; 录用日期: 2024年4月30日; 发布日期: 2024年4月30日

摘要

脑肿瘤的语义分割是一项基本的医学图像分析任务, 可以协助临床医生诊断患者, 并持续关注病灶部分的变化情况。得益于深度学习的发展, 医学图像自动分割取得了很大的进步。然而, 现有的深度学习分割模型依赖于庞大的训练数据支撑, 在临床上, 医学数据通常数据量较小。为了改善这些问题, 本文提出了一种新的深度医学图像分割框架, 称为Residual Multise Transunet (RM Transunet), 以提高小样本医学图像的语义分割质量。本文提出的RM Transunet遵循了Transunet的设计, 融合了CNN和Transformer, 并引入了Residual Block和MulSE, 有效地在不同尺度的特征之间建立全局依赖关系, 以便充分利用这些获得的多尺度特征表示。针对医学图像分割的实验证明了RM Transunet的有效性, 并表明我们的方法明显优于当下的方法。本研究的贡献不仅在于提供了一种新的思路来解决医学图像领域小样本数据的挑战, 同时也为临床决策和治疗优化提供了有益的参考。

关键词

深度学习, 脑肿瘤分割, 肺癌脑转移瘤, 小样本数据

RM Transunet: Segmentation of Lung Cancer Brain Metastases Based on Small Sample Data

Yuting Yang*, Hancan Zhu#

School of Mathematical Information, Shaoxing University of Arts and Sciences, Shaoxing Zhejiang

Received: Mar. 26th, 2024; accepted: Apr. 30th, 2024; published: Apr. 30th, 2024

*第一作者。

#通讯作者。

文章引用: 杨玉婷, 祝汉灿. RM Transunet: 基于小样本数据的肺癌脑转移瘤分割[J]. 计算机科学与应用, 2024, 14(4): 358-367. DOI: 10.12677/csa.2024.144105

Abstract

Semantic segmentation of brain tumors is a basic medical image analysis task that can assist clinicians in diagnosing patients and continuously paying attention to changes in the lesions. Thanks to the development of deep learning, great progress has been made in automatic segmentation of medical images. However, existing deep learning segmentation models rely on the support of huge training data. In clinical practice, medical data usually have a small amount of data. To improve these problems, this paper proposes a new deep medical image segmentation framework called residual multise transunet (RM Transunet) to improve the semantic segmentation quality of small sample medical images. The RM Transunet proposed in this article follows the design of Transunet, integrates CNN and Transformer, and introduces Residual Block and MuilSE to effectively establish global dependencies between features at different scales in order to make full use of these obtained multi-scale feature representations. Experiments on medical image segmentation demonstrate the effectiveness of RM Transunet and show that our method significantly outperforms current methods. The contribution of this study is not only to provide a new idea to solve the challenge of small sample data in the field of medical images, but also to provide a useful reference for clinical decision-making and treatment optimization.

Keywords

Deep Learning, Brain Tumor Segmentation, Lung Cancer Brain Metastases, Small Sample Data

Copyright © 2024 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

脑肿瘤分为原发性和继发性肿瘤。原发性脑肿瘤起源于脑细胞, 而继发性肿瘤是指从其他器官转移到大脑[1]。肺癌是中国及世界发病率和死亡率均为第一位的恶性肿瘤, 且发病率逐年上升。肺癌脑转移作为一类瘤继发性肿瘤, 是肺癌患者最常见的并发症之一, 且脑转移瘤是肺癌常见的死亡原因之一[2]。当患者出现脑转移瘤, 患者预后情况通常非常不理想, 生活质量大幅度下降。

随着医学成像设备的快速发展和普及, 成像技术在临床中得到了广泛应用。医学图像在诊断和治疗过程中起着至关重要的作用, 而医学图像处理和分析是一个复杂而关键的步骤。医学图像分割能够从特定组织图像中提取关键信息, 实现医学图像信息可视化。分割后的图像被提供给医生用于组织体积的定量分析、诊断、病理改变组织的定位、解剖结构的描绘、治疗计划等不同任务。因此, 准确且高效的分割脑肿瘤区域在临床上具有重要意义。基于现代科技的发展, 深度学习算法在计算机视觉领域取得了显著成就。深度学习算法所具备的自动、高效等优点, 已被证实在医学图像的分割领域同样具有较高的准确性。通过算法对医学图像中具有特殊意义的部分进行特征提取, 获取相关信息并进行自动化地精准分割, 可以为临床诊断和病理研究提供可靠的依据, 帮助医生做出更准确的诊断。

尽管深度学习分割方法在许多领域都取得了显著进展, 但在医学图像处理领域还存在一定挑战。首先, 现有的深度学习模型通常需要大量的标注数据进行训练, 而在医学图像等领域, 获取大规模标注数据是一项耗时且费力的任务当数据集非常小时, 深度学习算法的性能会受到限制, 传统的机器学习算法和深度学习算法通常难以发挥出其应有的效果。其次, 医学图像病理信息复杂且无规律可循, 对于不同

的病理数据特征各异, 现有的视觉领域的深度学习模型通常在训练数据上表现出色, 但泛化能力有限, 预测效果可能表现不佳。此外, 深度学习模型通常需要大量的计算资源进行训练, 这对于许多实际应用来说是一个挑战, 尤其是在一些复杂任务和深层网络结构中。因此, 如何在小样本数据情况下准确分割出肺癌脑转移瘤区域, 成为了医学图像分割领域中的一个难题。

针对上述问题, 本文提出了一种针对小样本数据的肺癌脑转移瘤分割算法, 结合 CNN 和 transformer 的优点, 将 Residual Block 引入编码器, 在下采样期间构建远程依赖关系和全局上下文连接, 在跳过链接部分利用多尺度的自我注意机制有效地在不同尺度的特征之间建立全局依赖关系, 在上采样过程中充分挖掘数据信息, 最后将融合的特征逐渐恢复到与像素级预测的输入图像相同的分辨率。得益于这些改进, 我们高效利用了小样本数据。我们评估了 RM Transunet 在小样本数据上的分割性能, 并对比了当下热门的分割网络结构, 实验结果证明了本文所提出的方法相较于传统的深度学习方法, 在准确性方面有显著的提高。

2. 相关工作

图像分割是指将图像中感兴趣的结构或区域进行像素级别的标记, 以实现精确的定位和量化分析[3]。神经网络发展初期的医学图像分割领域主要由卷积神经网络(CNN)主导。2015年, Ronneberger [4]等学者提出了 U-net 模型, 模型采用对称的编码器-解码器架构, 并在之间运用了跳跃连接, 有效捕获局部和全局特征, 从而提高分割精度。U-net 在多个医学图像分割任务中广泛应用, 被认为是医学图像分割领域的经典模型之一。Attention U-net [5]是 U-net 的一种变体, 它引入了注意力机制, 通过对不同区域的特征进行加权, 使得网络可以更加关注与局部区域, 帮助网络更加专注地学习和关注图像中的重要特征, 且能够更好地适应不同场景和不同尺度下的图像分割任务, 从而提高图像分割的准确性和效率。在[6] [7] [8] [9]中研究者们使用带有注意力机制的 U-net 结构进行肿瘤区域的分割和病理预测等任务, 并取得了有效突破。通常情况下, 适量的加深网络层次有利于更多的信息发掘, 但随着层次的加深, 梯度消失和梯度爆炸, 网络退化问题等问题也相继出现。Resnet [10]网络通过残差学习的思想解决了深度神经网络的退化问题, 促进了神经网络的进一步发展。在[11]中, 通过引用残差结构和注意力机制, 有效提高了分割性能。尽管利用 CNN 的卷积操作提取图像局部特征对于识别图像和分割等任务非常有效, 但是 CNN 在进行卷积操作时, 没有考虑输入序列中不同位置之间的关系, 而只是从局部区域中提取特征, 因此 CNN 的位置信息处理能力相对较弱, 在某些任务中存在一定的局限性, 如处理多尺度信息和边缘细节。此外, CNN 对于位置的变换比较敏感, 这可能导致模型在对数据进行平移、旋转等预处理操作时鲁棒性较差[12]。

为了解决这个问题, Google Brain 的研究人员提出了 Transformer [13]结构, 一种基于自注意力机制的神经网络架构。Transformer 通过自注意力机制来捕捉输入序列中各个位置之间的依赖关系, 从而更有效地建模长距离依赖关系, 并对图像的全局特征进行提取。2020年 Google 提出视觉 Transformer [14]结构, 用来处理计算机视觉任务的模型。视觉 Transformer 将输入的图像划分成一组固定大小的图像块, 并将这些图像块表示为序列。取得了与传统 CNN 相当的性能, 并具有更好的可扩展性、模型可解释性和泛化能力。基于此, 图像分割领域涌现出大量优秀的模型结构[15] [16] [17] [18], 在 Transformer 基础上进行了创新和改进, 以更好地适应不同的分割任务和数据特征。然而, Transformer 模型的自注意力机制需要大量的计算资源和内存, 对于大规模的图像数据, 可能需要巨大的模型和计算能力, 使得其在实际应用中可能受到限制。

最近的研究表明, 将 Transformer 模块结合到 U 形全卷积结构中, 可以充分发挥两种结构各自的优势。例如, 可以先使用 CNN 进行特征提取, 然后利用 Transformer 进行全局关系建模或上下文理解, 或者将两者结合起来以满足特定任务的需求, 同时学习图像的局部和全局特征[19]。Chen [20]等人将

Transformer 模块引入到 U-Net 结构中, 设计了用于医学图像分割的 Transunet 网络结构。在[21]中将 Transformer 模块作为编码器和解码器之间的残差长连接部分, 用于学习不同尺度特征图的全局信息, 在[22]中将 Transformer 模块作为最底层的特征提取模块, 获取全局信息的同时尽可能地优化参数量。除了将卷积和 Transformer 进行基础的融合, 在[23] [24]中, 研究者们将融合后的结构中加入注意力机制, 进一步优化模型能力。

注意力机制作为一种常用的深度学习技术, 其主要目的是通过在模型中引入可学习的注意力权重, 使模型能够动态地关注输入数据中的重要部分, 从而提高模型的性能。根据其在模型中的应用方式和设计特点, 注意力机制可以分为多种不同类型。其中, Transformer 中的自注意力机制[13]是一种常用于序列数据处理的技术, 它允许模型根据序列中不同位置的信息动态地调整注意力权重; 注意力池化[25]将注意力机制引入到池化层中, 使得网络能够有选择性地关注和提取最相关的特征信息, 从而提高了模型的性能; Woo 等人[26]在其研究中提出了空间注意力和通道注意力两种常用于图像处理任务的注意力机制。前者旨在通过学习图像中不同区域的重要性权重来定位目标, 而后者则专注于学习不同通道之间的关系, 以提取更具有区分性的特征表示。这些不同类型的注意力机制为模型提供了处理复杂数据的灵活和有效能力。Hu 等人[27]提出的 SE 注意力机制通过动态调整通道的重要性权重, 有效地增强模型对于重要特征的关注度, 并降低对于无关特征的干扰, 进而提高模型的性能表现。在图像处理领域, 通过学习和应用注意力机制, 模型能够更有效地感知输入数据中的关键信息, 更好地理解图像的局部和全局特征, 从而在图像分割等任务中更准确地定位和识别目标。

3. 数据和方法

3.1. 数据集

肺癌脑转移瘤数据: 我们利用浙江省人民医院收集的 100 例肺癌脑转移瘤的病例对所提出的方法进行验证。在专业医生的协助下, 我们对收集到的数据进行了肿瘤区域的分割, 并将其作为后续实验的金标准。

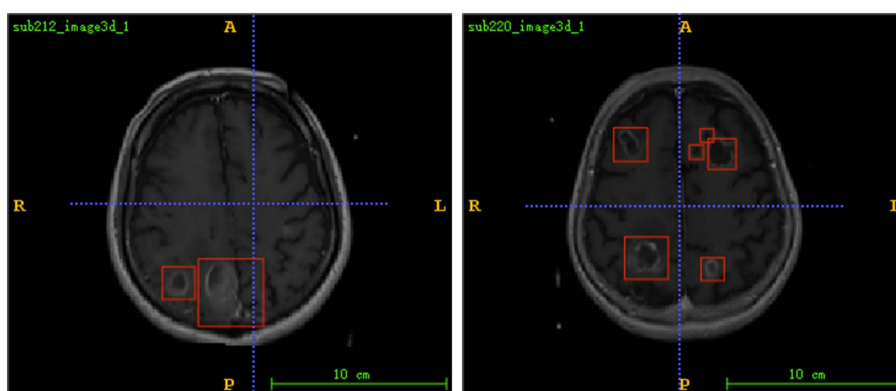


Figure 1. View of brain metastases from lung cancer

图 1. 肺癌脑转移瘤视图

由于 MRI 图像中还包含非大脑区域, 如硬脑膜、颅骨、脑膜和头皮等。这些非大脑区域的存在通常会降低模型在分类任务中的性能。因此, 我们首先对所有数据进行了颅骨剥离, 并进行了归一化处理、偏置场校正和图像配准等预处理步骤。通过这些预处理步骤, 我们获得了更规范、更清晰的图像数据。处理后的肺癌脑转移瘤和腺癌脑转移瘤的可视化如图 1 所示。

3.2. 模型介绍

RM Transunet 的网络结构图如图 2 所示, 主要包括编码器, 解码器和加入了多尺度注意力机制的跳跃连接三部分。

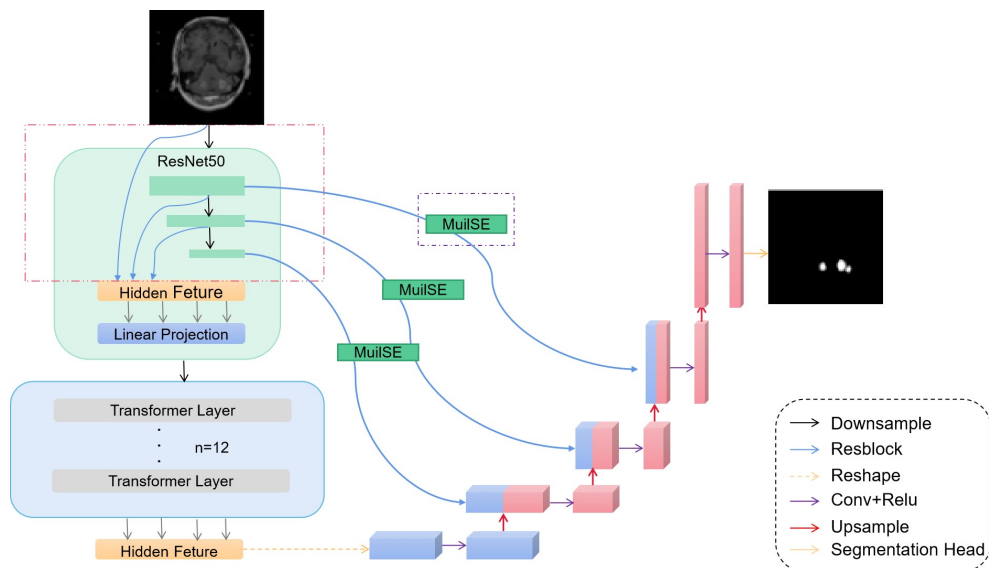


Figure 2. RM Transunet network structure diagram
图 2. RM Transunet 网络结构图

编码器: 编码器的主要任务是对输入图像进行特征提取和压缩, 并且保留重要的特征信息。编码器部分主要经历 CNN 和 Transformer 两个阶段。对于输入图像, 首先经过 3 个次下采样, 通过一系列卷积层和池化层逐渐改变图像的空间尺寸和特征通道数。每次下采样过程将分辨率减半, 并在每个中间阶段将通道维度加倍。对于给定的大小为 $H \times W \times C$ 的输入, 输出特征被重塑为 $H/2^i \times W/2^i \times D / 2^i$ ($i \in \{1, 2, 3\}$) 的大小。在该过程中, 每次进行下采样后的结果都要经过一个 Resblock, 并将 Resblock 的输出与最后一次下采样的结果相加, 经过线性映射后输入第二阶段, Resblock 部分的具体操作如图 3 所示。在第二阶段, 我们使用了 12 层 Transformer 结构对第一阶段提取的特征进行进一步提取, 以便细化特征并建立全局联系。在编码器部分我们进行了这些层的组合, 使网络能够充分提取图像中的局部和全局特征。

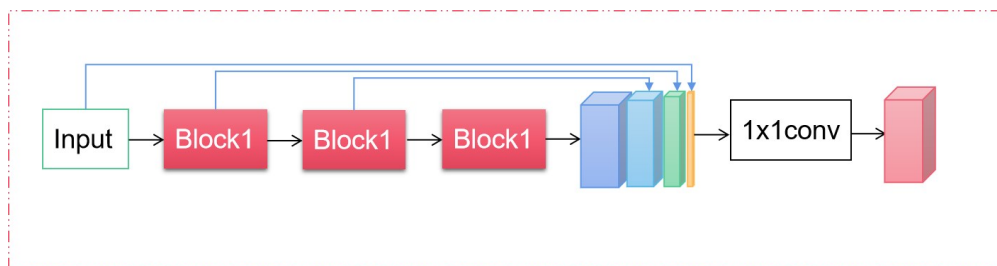


Figure 3. Resblock part specific operation diagram
图 3. Resblock 部分具体操作示意图

解码器: 解码器的任务是将编码器所提取的低维表示还原为像素级别的分割结果, 从而完成肿瘤分割任务。网络整体结构遵循 U 型结构, 编码器第一阶段提取的特征表示通过跳过连接的方式传递给解码

器。在编码器的第二阶段，底层特征直接作为底层解码器的输入，经过上采样后与跳过连接的输出逐层相加。随后，一系列卷积层和上采样层逐步将特征图的分辨率增加，使其恢复到原始输入图像的尺寸。解码器的最后一层采用一个 $1 \times 1 \times 1$ 的卷积层作为分割头，在不改变输出结果大小的同时，将每个像素点分配到不同的类别中，实现图像分割。通过这种设计在解码器中建立长期依赖关系和全局上下文交互，以获得更好的解码性能。

多尺度的注意力机制：在跳过链接部分，我们加入了一个多尺度的注意

力机制 Mult Attention SE (MASE)，具体操作过程如图 4 所示。在 MASE 中，输入特征首先经过多尺度的卷积操作，再通过 SEblock 以产生加权的特征表示。最后，将不同尺度的 SEblock 输出特征相加。该机制在不同尺度上对输入特征进行多层次的关注，从而捕捉到更加丰富和多样化的特征信息。这种方法有效地利用输入特征中的信息，提升对信息的关注度和特征提取能力，从而有助于在解码阶段实现更好的性能表现。

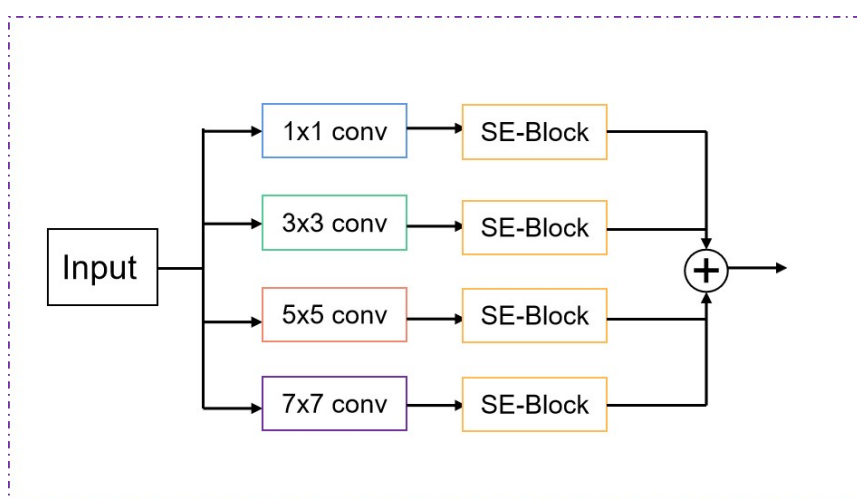


Figure 4. Mult Attention SE multi-scale attention mechanism

图 4. Mult Attention SE 多尺度注意力机制

3.3. 损失函数

针对本文的分割任务，我们选用了 Dice 损失函数。Dice 系数通常被用作一种集合相似度度量函数，用于衡量两个样本之间的相似程度。其取值范围在 0 到 1 之间[28]。数值越接近 1，表示两个样本之间的相似度越高，即分割结果越准确。Dice 系数的计算公式如下所示：

$$\text{Dice Coefficient} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

其中 $|X \cap Y|$ 表示集合 X 和 Y 的交集， $|X|$ 和 $|Y|$ 分别表示 X 、 Y 中元素的个数，对于分割任务而言， $|X|$ 和 $|Y|$ 分别表示分割的 Ground True 和 Predict mask。

Dice Loss 的目标是最小化预测分割结果与真实分割结果之间的差异，以使 Dice 系数尽可能地接近 1，从而提高分割模型的性能，进而提高分割模型的性能。Dice 损失和 Dice 系数的关系是： $\text{Dice Loss} = 1 - \text{Dice Coefficient}$ ，由此得到 Dice Loss 的公式为：

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

3.4. 实现细节

针对我们的实验数据, 考虑到数据量较少可能导致结果的偶然性, 我们将所有病人数据转换成 2D 切片, 并从中筛选出所有包含病变区域的数据进行实验。具体地, 我们将筛选之后的 602 个数据按照 7:1:2 的比例随机分为训练集, 验证集, 测试集。即 421 个病例切片作为训练集, 用于模型的训练过程, 60 个病例切片作为验证集, 用于调整模型的超参数和监测训练的进展, 121 个病例切片作为测试集, 以检验实验结果。我们的测试数据集在病变大小、位置和卒中类型等方面具有高度的多样性, 旨在验证模型的泛化能力和鲁棒性。

我们使用了两张 NVIDIA RTX 3090 GPU 进行模型训练, 批量大小设置为 8, 学习率为 $1e-4$, 权值衰减为 $1e-6$ 。经过多次实验对比, 发现在训练 50 次后, 模型已经达到了较好的效果, 并且性能相对稳定, 每次训练过程大约需要 1.7 小时。

为了更好地进行对比研究, 我们进行了消融实验。在消融实验中, 我们使用了不包含 Resblock 和 MuilSE 的原始模型结构进行训练。在这些实验中, 我们保持了与之前一致的参数设置, 即批量大小为 8, 学习率为 $1e-4$, 权值衰减为 $1e-6$, 训练次数为 50 次, 并使用相同的数据划分。

通过实验的评估, 我们能够全面了解模型的性能和稳定性。这种实验设计使得我们能够对肺癌脑转移瘤的分割任务进行可靠且细致的研究。

4. 实验结果

4.1. 消融研究及实验结果

我们首先进行消融研究, 以评估加入 Resblock 和 MuilSE 的重要性。为了使评估结果更准确可靠, 我们选取了多个评价指标, 共同评估模型在分割任务中的不同方面的性能表现, 包括对目标的准确识别、像素级的预测准确度等。实验结果如表 1 中所示。消融研究的结果显示, 当 Resblock 被移除时, 模型失去了残差连接和深层特征复用的能力, 导致了性能下降。同样地, 当 MuilSE 模块被移除时, 模型无法有效地利用多尺度的特征信息和注意力机制, 进而影响了分割性能。因此, 这两种设计模块的加入对于模型的整体性能起到了关键作用。这证实了这两种设计对于模型性能的重要性。

Table 1. Ablation study experimental results

表 1. 消融研究实验结果

Resblock	MuilSE	DICE	IOU	ACC	REC	PRE
√	√	0.9065	0.8474	0.9989	0.9199	0.9159
√	×	0.9041	0.8408	0.9989	0.9176	0.9085
×	√	0.9034	0.8442	0.9989	0.9122	0.9170
×	×	0.9007	0.8382	0.9987	0.9043	0.9169

4.2. 与现有分割方法的比较

针对我们的模型特点和结构, 我们对比了经典的全卷积网络结构 U-net 和引入了注意力机制的 Attention U-net 两种经典的分割模型。在对比实验中, 均采用 U-Net 和 Attention U-Net 验证集上的最优模型进行测试, 实验结果如表 2。其中, 在准确率和精确率方面, U-net 与我们的模型相当, 这表明 U-net 在正确分类像素方面具有很高的准确性, 然而, U-net 的召回率和精确率低于我们的模型, 而 Attention U-net 的各项指标结果均不佳, 我们推测是因为样本数据量一定程度上影响了其性能。这些都证实了我们的模

型在小样本分割任务中能够产生高度准确的结果, 并且对目标的识别和定位都具有很好的表现。

Table 2. Comparative study experimental results
表 2. 对比研究实验结果

Model	DICE	IOU	ACC	REC	PRE
Our Model	0.9065	0.8474	0.9989	0.9199	0.9159
U-net	0.8820	0.8057	0.9984	0.8511	0.9409
Attention U-net	0.6466	0.5275	0.9955	0.7504	0.6930

在图 5 中, 我们展示了在实验数据集上我们的模型与其他基线模型相比较的定性结果, 白色区域表示标签 1, 即肿瘤标签部分。上到下分别是三个不同病例数据的图像, 从左到右分别是精标准 GT, 本文提出的模型, U-net 结构和 Attention U-net 的分割结果。我们可以观察到基线模型不能很好地预测分割边界, 尤其是不规则的病变区域边缘, 且会遗漏较小的病变区域。相比之下, 我们提出的模型可以较为准确地预测分割边界, 即使对于非常小的病变也可以精准地刻画。这是由于我们的模型有效地使用 CNN 和 Transformer 的优点来捕获细粒度的像素级细节, 并通过残差和注意力机制建长距离联系。

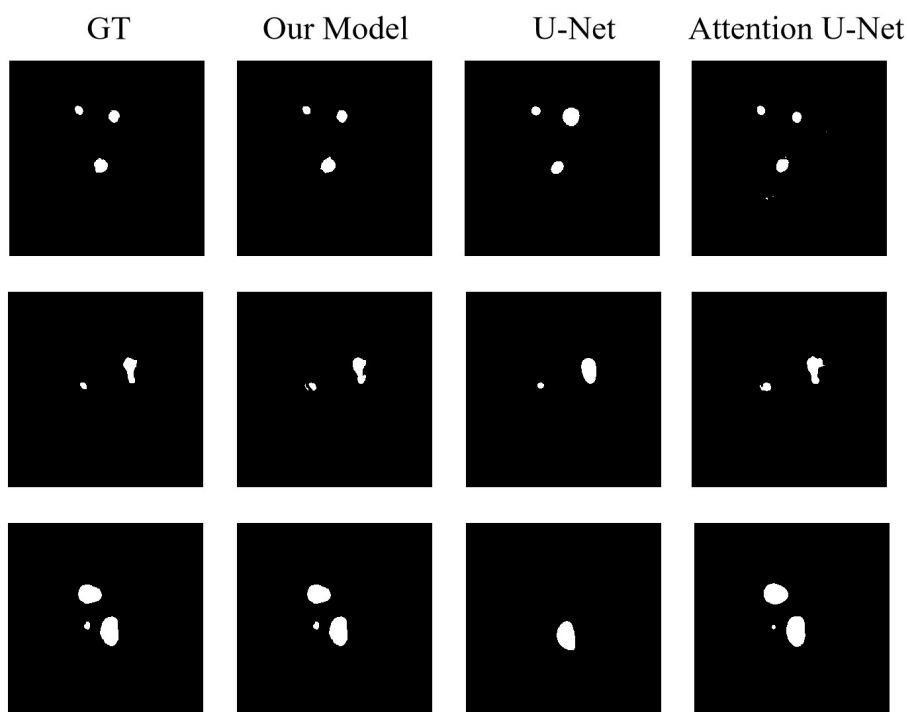


Figure 5. Visualization of segmentation results
图 5. 分割结果可视化

5. 讨论与结论

在本文中, 我们介绍了 RM Transunet, 这是一种利用残差结构和多尺度注意力机制对小样本脑肿瘤进行语义分割的新架构。我们提出的模型遵循 Transunet 的网络设计, 并在编码器部分引入了残差连接, 以辅助特征提取。在编码器与解码器的中间引入了加入不同尺度的注意力机制的跳跃连接, 以获取更全面的特征, 充分挖掘小样本数据的信息。我们对该方法进行了充分验证, 证明了其有效性。

在未来的研究中, 我们计划收集更多不同类型的样本用于模型训练, 以增强模型的鲁棒性和泛化能力。通过引入更多样本, 可以更好地适应不同情况下的变化, 从而提高模型的性能和适用性。我们也将持续改进算法性, 如改进注意力机制、优化损失函数或网络结构等, 以进一步提高分割精度和效率。

参考文献

- [1] 石远凯, 孙燕, 于金明, 等. 中国肺癌脑转移诊治专家共识(2017年版)[J]. 中国肺癌杂志, 2017, 20(1): 1-13.
- [2] Davidson, M.R., Larsen, J.E., Yang, I.A., *et al.* (2010) MicroRNA-218 Is Deleted and Downregulated in Lung Squamous Cell Carcinoma. *PLOS ONE*, **5**, e12560. <https://doi.org/10.1371/journal.pone.0012560>
- [3] Song, Y., Ren, S., Lu, Y., *et al.* (2022) Deep Learning-Based Automatic Segmentation of Images in Cardiac Radiography: A Promising Challenge. *Computer Methods and Programs in Biomedicine*, **220**, Article ID: 106821. <https://doi.org/10.1016/j.cmpb.2022.106821>
- [4] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, Munich, 5-9 October 2015, 234-241.
- [5] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas.
- [6] Trebing, K., Stańczyk, T. and Mehrkanoon, S. (2021) SmaAt-UNet: Precipitation Nowcasting Using a Small Attention-UNet Architecture. *Pattern Recognition Letters*, **145**, 178-186. <https://doi.org/10.1016/j.patrec.2021.01.036>
- [7] Islam, M., Vibashan, V., Jose, V., *et al.* (2020) Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. *Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019*, Shenzhen, 17 October 2019, 262-272. https://doi.org/10.1007/978-3-030-46640-4_25
- [8] Noori, M., Bahri, A. and Mohammadi, K. (2019) Attention-Guided Version of 2D UNet for Automatic Brain Tumor Segmentation. *Proceedings of the 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, 24-25 October 2019, 269-275. <https://doi.org/10.1109/ICCKE48569.2019.8964956>
- [9] Guo, C., Szemenyei, M., Yi, Y., *et al.* (2021) Sa-Unet: Spatial Attention U-Net for Retinal Vessel Segmentation. *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, 10-15 January 2021, 1236-1242. <https://doi.org/10.1109/ICPR48806.2021.9413346>
- [10] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *The 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>
- [11] Maji, D., Sigedgar, P., Singh, M., *et al.* (2022) Attention Res-UNet with Guided Decoder for Semantic Segmentation of Brain Tumors. *Biomedical Signal Processing and Control*, **71**, Article ID: 103077. <https://doi.org/10.1016/j.bspc.2021.103077>
- [12] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [13] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [15] Liu, Z., Lin, Y., Cao, Y., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *The Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [16] Wang, W., Chen, C., Ding, M., *et al.* (2021) TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference*, Strasbourg, 27 September-1 October 2021, 109-119. https://doi.org/10.1007/978-3-030-87193-2_11
- [17] Hatamizadeh, A., Tang, Y., Nath, V., *et al.* (2022) UNETR: Transformers for 3D Medical Image Segmentation. *The Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 3-8 January 2022, 1748-1758. <https://doi.org/10.1109/WACV51458.2022.00181>
- [18] Kim, N., Kim, D., Lan, C., *et al.* (2022) ReSTR: Convolution-Free Referring Image Segmentation Using Transformers. *The Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 18124-18133. <https://doi.org/10.1109/CVPR52688.2022.01761>
- [19] He, K., Gan, C., Li, Z., *et al.* (2023) Transformers in Medical Image Analysis. *Intelligent Medicine*, **3**, 59-78. <https://doi.org/10.1016/j.imed.2022.07.002>

-
- [20] Chen, J., Lu, Y., Yu, Q., *et al.* (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.
- [21] Wang, H., Cao, P., Wang, J., *et al.* (2022) UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *The Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 2441-2449. <https://doi.org/10.1609/aaai.v36i3.20144>
- [22] Li, Y., Wang, Z., Yin, L., *et al.* (2021) X-Net: A Dual Encoding-Decoding Method in Medical Image Segmentation. *The Visual Computer*, **39**, 2223-2233.
- [23] Fang, J., Yang, C., Shi, Y., *et al.* (2022) External Attention Based TransUNet and Label Expansion Strategy for Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*, **23**, 19054-19063. <https://doi.org/10.1109/TITS.2022.3154407>
- [24] Yang, Y. and Mehrkanoon, S. (2022) AA-TransUNet: Attention Augmented TransUNet for Nowcasting Tasks. *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, 18-23 July 2022, 1-8. <https://doi.org/10.1109/IJCNN55064.2022.9892376>
- [25] Lin, M., Chen, Q. and Yan, S. (2013) Network in Network.
- [26] Woo, S., Park, J., Lee, J.-Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. *The Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [27] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [28] Li, X., Sun, X., Meng, Y., *et al.* (2020) Dice Loss for Data-Imbalanced NLP Tasks. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 2020, 465-476. <https://doi.org/10.18653/v1/2020.acl-main.45>