

深度学习框架下的医学图像分割算法研究

张婷玮 龙艳彬 邓凯文

辽宁科技大学, 中国·辽宁 鞍山 114051

摘要: 针对医学图像的高维性、复杂性和高精度要求, 论文综述了深度学习框架下的医学图像分割算法。深度学习模型能够自适应地从大量数据中学习并提取多层次特征, 为医学图像分割提供了高精度、高鲁棒性和可扩展性强的解决方案。论文重点讨论了 U-Net、Transformer、Mamba 以及分割一切模型 (SAM) 等先进模型在医学图像分割中的应用, 并从多个维度进行了综合对比分析。此外, 论文还总结了当前医学图像分割研究面临的挑战, 并对未来研究方向进行了展望。

关键词: 医学图像分割; U-Net; Transformer; Mamba; 分割一切模型 (SAM)

Research on Medical Image Segmentation Algorithm under Deep Learning Framework

Tingwei Zhang Yanbin Long Kaiwen Deng

University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China

Abstract: In response to the high dimensionality, complexity, and precision requirements of medical images, this paper reviews medical image segmentation algorithms under the framework of deep learning. Deep learning models can adaptively learn and extract multi-level features from large amounts of data, providing a high-precision, robust, and scalable solution for medical image segmentation. The paper focuses on the application of advanced models such as U-Net, Transformer, Mamba, and Segmentation Everything Model (SAM) in medical image segmentation, and conducts comprehensive comparative analysis from multiple dimensions. In addition, the paper also summarizes the challenges faced by current medical image segmentation research and provides prospects for future research directions.

Keywords: medical image segmentation; U-Net; Transformer; Mamba; Split All Models (SAM)

0 前言

在医学研究与临床实践中, 医学图像分割扮演着至关重要的角色。医学图像, 如 CT、MRI 和超声图像等, 为医生提供了患者体内结构和病理变化的直观展示。通过对这些图像进行精确分割, 医生能够更准确地识别器官、组织以及病变区域, 从而为疾病诊断、治疗策略制定以及临床研究提供关键信息^[1]。

然而, 医学图像分割面临着诸多挑战。首先, 医学图像通常具有高维性和复杂性, 包含大量的细节信息和噪声, 这增加了分割的难度。其次, 不同患者之间的解剖结构和病理变化存在差异, 使得分割算法需要具备高度的自适应性和鲁棒性。最后, 医学图像分割还需要满足高精度要求, 以确保诊断的准确性和可靠性^[2]。

1 基于 U-Net 的医学图像分割研究

1.1 U-Net 模型的基本架构和原理

U-Net 是一种经典的卷积神经网络 (CNN) 架构, 专为生物医学图像分割而设计。其名称来源于其网络结构的“U”形布局, 该布局由编码器 (下采样路径) 和解码器 (上采样路径) 组成, 通过跳跃连接 (skip connections) 将编码器的特征图直接传递给解码器, 从而保留更多的图像细节^[3-4]。

1.2 U-Net 在医学图像分割中的应用案例

U-Net 在医学图像分割领域具有广泛的应用, 包括但不限于细胞分割、器官分割、病变区域检测等。例如, 在细胞显微图像分割中, U-Net 能够准确地识别并分割出细胞轮廓; 在肺部 CT 图像分割中, U-Net 可以精确地分割出肺部区域, 为肺部疾病的诊断提供重要信息; 在脑部 MRI 图像分割中, U-Net 能够识别并分割出不同的脑组织, 如灰质、白质和脑脊液等^[5]。

1.3 U-Net 的改进版本和变体模型

为了进一步提高 U-Net 的性能, 研究人员提出了多种改进版本和变体模型。例如, 3D U-Net 将二维卷积替换为三维卷积, 以处理三维医学图像; ResU-Net 引入了残差连接 (residual connections), 以增强网络的特征提取能力; Attention U-Net 则通过引入注意力机制, 使网络能够更关注图像中的重要区域。此外, 还有诸如 Dense U-Net、UNet++ 等变体模型, 通过增加网络密度、改进跳跃连接等方式, 进一步提升了 U-Net 的分割性能^[6-7]。

在讨论 U-Net 的改进版本和变体模型的软件实现程序时, 我们需要明确每个模型的核心改进点, 并基于这些点来设计相应的网络架构。以下是对几种主要 U-Net 变体模型的简要软件实现框架, 使用 Python 和深度学习框架 (如

TensorFlow 或 PyTorch) 作为示例。

1.3.1 3D U-Net

3D U-Net 的核心在于将二维卷积替换为三维卷积, 以处理三维医学图像。以下是一个基于 PyTorch 的简化 3D U-Net 实现框架:

```
```python
import torch
import torch.nn as nn

class ConvBlock3D(nn.Module):
 def __init__(self, in_channels, out_channels):
 super(ConvBlock3D, self).__init__()
 self.conv1 = nn.Conv3d(in_channels, out_channels,
kernel_size=3, padding=1)
 self.bn1 = nn.BatchNorm3d(out_channels)
 self.relu = nn.ReLU(inplace=True)
 self.conv2 = nn.Conv3d(out_channels, out_channels,
kernel_size=3, padding=1)
 self.bn2 = nn.BatchNorm3d(out_channels)
 def forward(self, x):
 x = self.conv1(x)
 x = self.bn1(x)
 x = self.relu(x)
 x = self.conv2(x)
 x = self.bn2(x)
 x = self.relu(x)
 return x

class DownSampling3D(nn.Module):
 def __init__(self, in_channels, out_channels):
 super(DownSampling3D, self).__init__()
 self.maxpool = nn.MaxPool3d(kernel_size=2)
 self.convblock = ConvBlock3D(in_channels, out_
channels)
 def forward(self, x):
 x = self.maxpool(x)
 x = self.convblock(x)
 return x

class UpSampling3D(nn.Module):
 def __init__(self, in_channels, out_channels,
bilinear=True):
 super(UpSampling3D, self).__init__()
 self.up = nn.ConvTranspose3d(in_channels, out_
channels // 2, kernel_size=2, stride=2)
 self.convblock = ConvBlock3D(out_channels, out_
channels)
 def forward(self, x1, x2):
 x1 = self.up(x1)
```

Assume x1 and x2 have the same spatial dimensions

```
x = torch.cat([x2, x1], dim=1)
```

```
x = self.convblock(x)
```

```
return x
```

Define the full 3D U-Net architecture here...

(For brevity, only key components are shown)

```
...
```

### 1.3.2 ResU-Net

ResU-Net 引入了残差连接, 以增强网络的特征提取能力。以下是一个基于 PyTorch 的简化 ResU-Net 实现框架:

```
```python
class ResBlock(nn.Module):
    def __init__(self, in_channels, out_channels):
        super(ResBlock, self).__init__()
        self.conv1 = nn.Conv2d(in_channels, out_channels,
kernel_size=3, padding=1)
        self.bn1 = nn.BatchNorm2d(out_channels)
        self.relu = nn.ReLU(inplace=True)
        self.conv2 = nn.Conv2d(out_channels, out_channels,
kernel_size=3, padding=1)
        self.bn2 = nn.BatchNorm2d(out_channels)
        if in_channels != out_channels:
            self.shortcut = nn.Sequential(
                nn.Conv2d(in_channels, out_channels, kernel_
size=1),
                nn.BatchNorm2d(out_channels)
            )
        else:
            self.shortcut = nn.Sequential()
    def forward(self, x):
        out = self.conv1(x)
        out = self.bn1(out)
        out = self.relu(out)
        out = self.conv2(out)
        out = self.bn2(out)
        out += self.shortcut(x)
        out = self.relu(out)
        return out
```

Define the full ResU-Net architecture here...

(For brevity, only the ResBlock is shown)

1.4 U-Net 在医学图像分割中的性能评估

U-Net 在医学图像分割中的性能评估通常基于多个方面, 包括分割精度、计算效率、鲁棒性等。常用的评估指标包括 Dice 系数、IoU (交并比)、精确度、召回率等。Dice 系数和 IoU 是衡量分割结果与真实标签之间重叠程度的常用指标, 值越高表示分割性能越好。精确度和召回率则分别

反映了分割结果的准确性和完整性。此外, 计算效率也是评估 U-Net 性能的重要指标之一, 特别是在处理大规模医学图像数据时^[8-10]。

2 基于 Transformer 的医学图像分割研究

2.1 Transformer 模型的基本原理和优势

Transformer 模型最初是为自然语言处理 (NLP) 任务而设计的, 其核心思想是通过自注意力机制 (self-attention mechanism) 来捕捉序列数据中的长距离依赖关系。Transformer 由编码器和解码器两部分组成, 其中编码器负责将输入序列转换为一系列隐藏状态, 而解码器则利用这些隐藏状态生成输出序列。

在 Transformer 中, 自注意力机制通过计算输入序列中不同位置之间的相关性得分, 来动态地调整每个位置的表示。这种机制使得 Transformer 能够捕捉到输入数据中的全局上下文信息, 从而在处理长序列数据时表现出色。

2.2 Transformer 在医学图像分割中的应用与挑战

尽管 Transformer 最初是为 NLP 任务而设计的, 但近年来其在计算机视觉领域的应用也取得了显著进展。在医学图像分割中, Transformer 模型被用于捕捉图像中的全局上下文信息, 以提高分割性能。

然而, 将 Transformer 应用于医学图像分割也面临一些挑战。第一, 医学图像通常具有高分辨率和复杂的解剖结构, 这使得 Transformer 在处理这些图像时需要消耗大量的计算资源和内存。第二, 医学图像分割任务通常需要精细的边界定位, 而 Transformer 的自注意力机制可能会在处理局部细节时表现不佳。因此, 如何在保持 Transformer 全局上下文捕捉能力的同时, 提高其局部细节处理能力, 是当前研究的重要方向。

2.3 Transformer 的改进版本和变体模型在医学图像分割中的表现

为了克服 Transformer 在医学图像分割中的挑战, 研究人员提出了多种改进版本和变体模型。例如, ViT (Vision Transformer) 将 Transformer 应用于图像分类任务, 通过将图像分割成一系列小块 (patches) 并作为序列输入到 Transformer 中, 实现了对图像的全局上下文捕捉。在此基础上, 研究人员进一步提出了适用于医学图像分割的改进版本, 如 Swin Transformer 和 TransUNet 等。

Swin Transformer 通过引入局部自注意力机制和移位窗口 (shifted windows) 策略, 有效地降低了计算复杂度, 并提高了对局部细节的捕捉能力。TransUNet 则将 Transformer 与 U-Net 相结合, 利用 U-Net 的跳跃连接来保留图像的细节信息, 并通过 Transformer 捕捉全局上下文信息。这些改进版本和变体模型在医学图像分割任务中取得了显著的性能提升。

2.4 Transformer 与 U-Net 等模型的对比分析

与 U-Net 等传统 CNN 模型相比, Transformer 在医学

图像分割中具有一些独特的优势。第一, Transformer 能够捕捉到图像中的全局上下文信息, 这对于理解复杂的解剖结构和病变区域非常重要。第二, Transformer 的并行处理能力使得其在处理大规模医学图像数据时更加高效。然而, Transformer 也面临一些挑战, 如计算资源消耗大、局部细节处理能力不足等。

3 基于 Mamba 的医学图像分割研究

3.1 Mamba 模型的基本特点和优势

Mamba 是一种新型的深度学习架构, 它基于选择性状态空间模型 (Selective State Space Models, SSMs) 来改进传统的状态空间模型。Mamba 模型的基本特点和优势主要体现在以下几个方面:

①选择性机制: Mamba 引入了一种新颖的选择性机制, 使得模型可以根据输入动态调整其行为。这种机制能够有效地过滤掉无关信息, 并加强与任务相关的信息, 从而提高模型的性能和效率。

②线性时间复杂度: Mamba 模型的设计具有线性时间复杂度, 这意味着在处理长序列数据时, 其计算复杂度随着输入序列长度的增加而线性增长。这一特性使得 Mamba 在处理大型复杂数据集时具有显著优势, 尤其是在医学图像分割等需要处理大量空间和时间关系的任务中。

③硬件感知算法: Mamba 模型采用了一种硬件感知算法, 该算法充分利用 GPU 的内存层次结构来提高计算速度和降低内存需求。这种算法结合了递归神经网络 (RNN) 的递归计算效率和卷积神经网络 (CNN) 的并行处理优势, 使得 Mamba 在处理长序列数据时表现出更高的计算效率和性能。

④多模态数据融合: Mamba 模型在融合多模态数据方面表现出色, 能够处理来自不同来源和格式的数据, 从而提高诊断准确性和患者的治疗效果。

3.2 Mamba 在医学图像分割中的应用案例

在医学图像分割领域, Mamba 模型已经展现出了巨大的潜力和优势。以下是一些具体的应用案例:

①器官分割: Mamba 模型被广泛应用于医学图像的器官分割任务中。通过利用 Mamba 的选择性机制和线性时间复杂度优势, 模型能够高效地处理复杂的医学图像数据, 并准确地分割出目标器官。

②皮肤病变分割: 在皮肤病变的分割任务中, Mamba 模型也取得了显著的效果。模型能够准确地识别出皮肤病变区域, 为医生提供重要的诊断依据。

③ HC-Mamba 模型: 研究者提出了 HC-Mamba 模型, 这是一种基于 Mamba 架构的混合卷积技术视觉模型。该模型引入了扩张卷积技术和深度可分离卷积, 以捕获更广泛的上下文信息并减少计算成本。实验结果表明, HC-Mamba 在医学图像分割任务中展现出了竞争性的表现。

3.3 Mamba 模型的改进和优化方法

为了进一步提高 Mamba 模型的性能和效率, 研究者们提出了多种改进和优化方法:

①模型架构优化: 通过简化传统 SSM 架构, 合并 H3 和多层感知器 (MLP) 块, 形成了一个均匀堆叠的结构。这一优化不仅简化了模型的结构, 还提高了模型的灵活性和效率。

②损失函数设计: 在医学图像分割任务中, 损失函数的设计对模型的性能有着重要影响。研究者们提出了多种综合损失函数, 如结合 mIoU 损失、Dice 损失和边界损失等, 以全面优化分割模型的性能。

③数据增强和预处理: 为了提高模型的泛化能力和鲁棒性, 研究者们采用了数据增强和预处理技术。通过对原始医学图像进行旋转、缩放、翻转等操作, 生成更多的训练样本, 从而增强模型的泛化能力。

4 SAM 大模型的医学图像分割研究

4.1 SAM 模型的基本架构和原理

SAM (Segment Anything Model) 是一个创新的图像分割模型, 其核心架构和原理如下:

基本架构:

SAM 模型主要由三大模块构成: 图像编码器 (Image Encoder)、提示编码器 (Prompt Encoder) 和掩码解码器 (Mask Decoder)。

①图像编码器: 该模块负责将输入的图像转换为向量表示。它通常采用预训练的 Vision Transformer (ViT) 作为核心, 以适应高分辨率的输入。图像被切割成多个小块, 每个小块的颜色值被转化为一个向量, 作为 Transformer 的输入序列。

②提示编码器: 这个模块将用户提供的分割提示 (如点、框、文本等) 转换为向量表示。这些提示向量与图像向量拼接在一起, 共同作为 Transformer 的输入。

③掩码解码器: 该模块将图像和提示的向量表示融合, 并解码出最终的分割掩码。这个掩码表示了每个像素的标签, 从而实现了图像的分割。

原理:

SAM 模型基于 Transformer 的自注意力机制, 计算图像中每个像素块与其他像素块以及用户提示之间的相关性。这种全局的依赖关系捕捉能力使得 SAM 能够准确地理解图像的结构和语义信息, 以及用户的意图。通过训练, SAM 模型学会了如何根据输入的图像和提示生成准确的分割掩码。

4.2 SAM 在医学图像分割中的应用潜力

SAM 模型在医学图像分割领域具有巨大的应用潜力, 主要体现在以下几个方面:

①零样本分割能力: SAM 能够通过用户提供的简单提示 (如点、框) 对图像进行分割, 而无需预先训练或学习特

定领域的知识。这使得 SAM 在医学图像分割中, 尤其是在数据稀缺的情况下, 具有显著的优势。

②高效性和灵活性: SAM 模型的训练过程基于大规模的自然图像数据集, 这使得它能够快速适应新的医学图像数据。同时, 其强大的泛化能力和灵活性使得 SAM 能够处理多种不同的医学图像分割任务。

③可解释性: SAM 模型的训练过程使用可视化技术, 如特征图分析和梯度映射, 这有助于理解模型的决策过程, 从而提高模型的可解释性。这对于医学图像分割来说至关重要, 因为医生需要了解模型的分割依据, 以便做出更准确的诊断。

4.3 SAM 模型的训练和优化策略

为了训练和优化 SAM 模型, 研究者们采取了多种策略:

①大规模数据集训练: SAM 模型使用了一个包含数亿个掩码和数百万张图像的巨大数据集进行训练。这确保了模型能够学习到丰富的图像特征和分割能力。

②优化算法和技巧: 在训练过程中, 研究者们采用了多种优化算法和技巧, 如混合精度训练、模型并行和数据并行等, 以加速训练过程并提高模型的性能。

③正则化和数据增强: 为了防止过拟合, 研究者们采用了正则化方法, 并在训练过程中使用了数据增强技术, 如旋转、缩放和翻转等, 以增加模型的泛化能力。

4.4 SAM 与其他先进模型的对比分析

与其他先进的医学图像分割模型相比, SAM 模型具有以下优势:

①更强的泛化能力: 由于 SAM 模型使用了大规模的数据集进行训练, 并采用了 Transformer 的自注意力机制, 因此它具有更强的泛化能力, 能够处理多种不同的医学图像分割任务。

②更高的准确性: 在多个医学图像分割任务上, SAM 模型的表现优于其他先进的模型。这得益于其强大的图像特征提取能力和准确的分割掩码生成能力。

③更好的可解释性: 与其他深度学习模型相比, SAM 模型的可解释性更强。研究者们使用了可视化技术来展示模型的决策过程, 这使得医生能够更直观地理解模型的分割依据。

然而, SAM 模型也存在一些局限性, 如处理三维医学图像的能力有限、对特定医学领域的适应性需要进一步提高等。未来, 研究者们将继续优化 SAM 模型, 以更好地满足医学图像分割的需求。

5 结语

论文深入探讨了医学图像分割领域的最新进展, 特别是聚焦于 SAM 大模型和 Mamba 模型等先进技术的应用。通过详细分析这些模型的基本架构、原理、应用潜力、训练优化策略以及与其他模型的对比分析, 论文揭示了它们在医

学图像分割中的巨大潜力和实际价值。

在 SAM 大模型的研究中,我们强调了其基于 Transformer 的自注意力机制在捕捉全局依赖关系方面的优势,以及其在零样本分割能力、高效性和灵活性、可解释性等方面的独特之处。这些特性使得 SAM 模型在医学图像分割中,尤其是在处理复杂结构和精细细节时,表现出色。

对于 Mamba 模型,我们探讨了其基于选择性状态空间模型的设计,以及这种设计如何带来线性时间复杂度、硬件感知算法和多模态数据融合等优势。这些特性使得 Mamba 模型在处理大型复杂数据集时更加高效,同时在融合多模态数据方面表现出色,为医学图像分割提供了更多的可能性。

通过论文的研究,我们不仅深入理解了这些先进模型的工作原理和优势,还通过对比分析揭示了它们在医学图像分割中的实际表现。这些研究成果为未来的医学图像分割研究提供了重要的参考和启示。

未来的医学图像分割研究将更加注重模型的泛化能力、计算效率、可解释性和鲁棒性等方面的提升。随着技术的不断进步和应用场景的不断拓展,我们有理由相信,未来的医学图像分割技术将为医学研究和临床实践带来更多的创新和突破。

参考文献:

- [1] HSU C Y, TSENG C C, LEE S L, et al. Image classification using convolutional neural networks with different convolution operations[C/OL].//2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan),2020:1-2.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2024,39(4):640-651.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[J]. Springer International Publishing,2015.
- [4] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2023,40(4):834-848.
- [5] 许伟鑫.基于深度卷积网络的肺结节分割与良恶性分类算法研究[D].重庆:重庆大学,2021.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//GUYON I, LUXBURG U, BENGIO S, et al. Advances in Neural Information Processing Systems,2022:45-48.
- [7] CHEN J, LU Y, YU Q, et al. TransUNet: transformers make strong encoders for medical image segmentation[J]. Cornell University - arXiv, Cornell University - arXiv,2021:123-128.
- [8] ZHANG Y, LIU H, 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.
- [9] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[J]. arXiv preprint arXiv,2024.
- [10] ZHU L, LIAO B, ZHANG Q, et al. Vision Mamba: efficient visual representation learning with bidirectional state space model[J]. arXiv preprint arXiv,2024:34-37.