

### **FDiff-Fusion:基于模糊逻辑驱动的医学图像扩散融合网络分割模型**

耿胜, 丁卫平, 鞠恒荣, 黄嘉爽, 姜舒, 王海鹏

引用本文

耿胜, 丁卫平, 鞠恒荣, 黄嘉爽, 姜舒, 王海鹏. [FDiff-Fusion:基于模糊逻辑驱动的医学图像扩散融合网络分割模型](#)[J]. 计算机科学, 2025, 52(6): 274-285.

GENG Sheng, DING Weiping, JU Hengrong, HUANG Jiashuang, JIANG Shu, WANG Haipeng. [FDiff-Fusion:Medical Image Diffusion Fusion Network Segmentation Model Driven Based onFuzzy Logic](#) [J]. Computer Science, 2025, 52(6): 274-285.

---

### **相似文献推荐 (请使用火狐或 IE 浏览器查看文章)**

**Similar articles recommended (Please use Firefox or IE to view the article)**

#### [基于高斯偏置自注意力和交叉注意力的医学图像分割模型](#)

Gaussian-bias Self-attention and Cross-attention Based Module for Medical Image Segmentation  
计算机科学, 2024, 51(11A): 240300071-9. <https://doi.org/10.11896/jsjcx.240300071>

#### [集成全尺度融合和循环注意力的医学图像分割网络](#)

Medical Image Segmentation Network Integrating Full-scale Feature Fusion and RNN with Attention  
计算机科学, 2024, 51(5): 100-107. <https://doi.org/10.11896/jsjcx.230400114>

#### [复杂网络社团发现综述](#)

Survey of Community Discovery in Complex Networks

计算机科学, 2023, 50(11A): 230100130-11. <https://doi.org/10.11896/jsjcx.230100130>

#### [面向轻量化医学图像分割网络的神经结构搜索](#)

Neural Architecture Search for Light-weight Medical Image Segmentation Network  
计算机科学, 2022, 49(10): 183-190. <https://doi.org/10.11896/jsjcx.210800052>

#### [融合时序监督和注意力机制的脉络膜新生血管分割](#)

Choroidal Neovascularization Segmentation Combining Temporal Supervision and Attention  
Mechanism

计算机科学, 2021, 48(8): 118-124. <https://doi.org/10.11896/jsjcx.200600150>

# FDiff-Fusion: 基于模糊逻辑驱动的医学图像扩散融合网络分割模型

耿 胜 丁卫平 鞠恒荣 黄嘉爽 姜 舒 王海鹏

南通大学人工智能与计算机学院 江苏 南通 226019

(gszs369@163.com)

**摘要** 医学图像分割在临床诊疗和病理分析中具有重要的应用价值。近年来,去噪扩散模型在图像分割建模方面取得了显著成功,其能够更好地捕获图像中的复杂结构和细节信息。然而,利用去噪扩散模型进行医学图像分割的方法大多忽略了分割目标的边界不确定性和区域模糊因素,从而造成了最终分割结果的不稳定性和不准确性。为了解决这一问题,提出了一种基于模糊逻辑驱动的医学图像扩散融合网络分割模型(FDiff-Fusion)。该模型通过将去噪扩散模型集成到经典 U-Net 网络中,有效地从输入医学图像中提取丰富的语义信息。由于医学图像的分割目标边界不确定性和区域模糊化现象普遍存在,因此在 U-Net 网络的跳跃路径上设计了一种模糊学习模块。该模块为输入的编码特征设置多个模糊隶属度函数,以描述特征点之间的相似程度,并对模糊隶属度函数应用模糊规则处理,从而增强了模型对不确定边界和模糊区域的建模能力。此外,为了提高模型分割结果的准确性和鲁棒性,在测试阶段引入了基于迭代注意力特征融合的方法。该方法将局部上下文信息添加到注意力模块中的全局上下文信息中,以融合每个去噪时间步的预测结果。实验结果显示,与现有的先进分割网络相比,FDiff-Fusion 在 BRATS 2020 脑肿瘤数据集上获得的平均 Dice 分数和 HD95 距离分别为 84.16% 和 2.473mm,在 BTCV 腹部多器官数据集上获得的平均 Dice 分数和 HD95 距离分别为 83.82% 和 7.98mm,表现出良好的分割性能。

**关键词:** 去噪扩散模型; U-Net 网络; 医学图像分割; 模糊学习; 迭代注意力特征融合

中图分类号 TP391

## FDiff-Fusion: Medical Image Diffusion Fusion Network Segmentation Model Driven Based on Fuzzy Logic

GENG Sheng, DING Weiping, JU Hengrong, HUANG Jiashuang, JIANG Shu and WANG Haipeng

School of Artificial Intelligence and Computer Science, Nantong University, Nantong, Jiangsu 226019, China

**Abstract** Medical image segmentation has important application value in clinical diagnosis, treatment and pathological analysis. In recent years, denoising diffusion models have achieved remarkable success in image segmentation modeling, which can better capture complex structure and detail information in images. However, most of the methods using the denoising diffusion model for medical image segmentation ignore the boundary uncertainty and region ambiguity of the segmentation target, resulting in the instability and inaccuracy of the final segmentation results. In order to solve this problem, a medical image diffusion fusion network segmentation model driven based on fuzzy logic (FDiff-Fusion) is proposed. By integrating the denoising diffusion model into the classical U-Net network, this model can effectively extract rich semantic information from inputting medical images. Since the boundary uncertainty and region blurring of medical image segmentation are common, a fuzzy learning module is designed on the jump path of U-Net network. The module sets several fuzzy membership functions for the input encoded features to describe the similarity degree between the feature points, and applies fuzzy rules to the fuzzy membership functions, thus enhancing the modeling ability of the model to the uncertain boundary and fuzzy region. In addition, in order to improve the accuracy and robustness of the model segmentation results, a method based on iterative attention feature fusion is introduced in the test phase, which adds lo-

到稿日期:2024-06-03 返修日期:2024-08-30

基金项目:国家重点研发计划(2024YFE0202700);国家自然科学基金(62006128, 62102199, 62471259, 62406153);江苏省自然科学基金(BK20231337);江苏省双创博士计划;江苏省高等学校自然科学研究面上项目(23KJB520031, 24KJB520032);南通市科技局基础科学研究项目(JC2021122);中国博士后科学基金(2022M711716);江苏省实践创新计划资助项目(SJ CX24\_2017)

This work was supported by the National Key R&D Program of China (2024YFE0202700), National Natural Science Foundation of China (62006128, 62102199, 62471259, 62406153), Natural Science Foundation of Jiangsu Province (BK20231337), Double-Creation Doctoral Program of Jiangsu Province, General Program of the Natural Science Foundation of Jiangsu Province Higher Education Institutions (23KJB520031, 24KJB520032), Basic Science Research Program of Nantong Science and Technology Bureau (JC2021122), China Postdoctoral Science Foundation (2022M711716) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJ CX24\_2017).

通信作者:丁卫平(dw p9988@163.com)

cal context information to the global context information in the attention module to fuse the prediction results of each denoising time step. Experimental results show that compared with existing advanced segmentation networks, the average Dice score and the average HD95 distance obtained by FDiff-Fusion on BRATS 2020 brain tumor dataset are 84.16% and 2.473mm, respectively. The mean Dice score and the mean HD95 distance obtained on BTCV abdominal multi-organ dataset are 83.41% and 7.98mm, respectively, showing good segmentation performance.

**Keywords** Denoising diffusion model, U-Net network, Medical image segmentation, Fuzzy learning, Iterative attention feature fusion

## 1 引言

脑肿瘤是一种在脑组织内形成的异常生长,可能是良性或恶性的,对患者的生活和整体健康造成严重影响。据统计,全球每年约有数十万人被诊断患有脑肿瘤,其中恶性肿瘤的发病率呈现上升趋势<sup>[1]</sup>。目前,脑肿瘤的早期诊断和治疗成为临床关注的重点之一。然而,其在脑组织内的复杂位置和结构,以及不同类型肿瘤的异质性,使得其诊断和治疗变得极具挑战性。此外,腹部多器官的分割是一项重要的医学图像处理任务,旨在从腹部影像中准确提取多个器官的轮廓和边界,如肝脏、肾脏、胰腺等,以帮助医生进行诊断、手术规划和治疗监测<sup>[2]</sup>。传统用于脑肿瘤和腹部多器官分割的方法往往需要医生具备丰富的经验和专业知识,且分割过程复杂、耗时,容易受到主观因素和误差的影响。因此,自动化的医学图像分割技术对于提高诊断准确性和临床效率具有重要意义,有助于医生及时发现病变区域,提高患者的治疗效果和生存率。

医学图像分割是一种利用计算机视觉和图像处理技术将医学影像中感兴趣的结构或区域准确地标记或分割出来的过程<sup>[3]</sup>。它在医学影像分析中起到了至关重要的作用,有助于医生进行诊断、监测疾病和制定治疗方案。通过自动识别和量化影像中的病变区域,医生能够更快速、更准确地制定治疗方案,从而提高患者的治疗效果。然而,由于医学图像具有复杂性和多样性,例如噪声、分割边缘不确定性和区域模糊化等,目前对医学图像的分割仍面临着许多挑战和困难<sup>[4]</sup>。

随着深度学习的不断进步,卷积神经网络(Convolutional Neural Network, CNN)以其强大的特征表示能力在医学图像分割领域发挥着关键作用<sup>[5-6]</sup>。近年来,学者们通过引入 Transformer 结构,进一步提升了医学图像分割的性能<sup>[7]</sup>。研究显示, CNN 和 Transformer 能够有效学习医学图像中更具代表性的特征,因此 CNN 和 Transformer 在各种分割任务中取得了显著的性能提升<sup>[8-9]</sup>。其中, U-Net 由 Ronneberger 等<sup>[10]</sup>于 2015 年提出,是一种深度学习模型,主要用于图像分割任务。其独特的 U 形结构,使其特别适用于医学图像分割等领域。U-Net 包含对称的编码器(收缩路径)和解码器(扩展路径),通过跳跃连接将编码器和解码器的不同层级连接在一起。这种结构设计有助于捕获图像中的局部和全局信息,从而提高分割的准确性。SegResNet 是一种基于残差连接的改进型 U-Net 架构,它利用变分自编码器添加重建分支,提高了模型的特征提取能力<sup>[11]</sup>。该架构旨在克服神经网络训练中的梯度消失问题,并加速模型的收敛过程<sup>[12]</sup>。SegResNet 在医学图像分割等任务中表现出色,兼具深度学习模型的高表达能力和 U-Net 结构的优势。通过引入残差连接,

SegResNet 能够更好地捕获图像中的特征信息,提高分割的准确性和鲁棒性。TransBTS 是一种基于编码器-解码器结构的分割网络<sup>[13]</sup>,首先利用 3D CNN 提取体积空间特征,并对输入的 3D 图像进行下采样,以产生紧凑的体积特征图,从而有效捕获局部 3D 上下文信息;然后,每个卷积被重塑为一个令牌并输入到 Transformer 进行全局特征的建模;最后, 3D CNN 解码器采用 Transformer 的特征嵌入并执行渐进上采样来预测全分辨率分割图。UNETR 利用 ViT 作为编码器直接对全局特征进行建模,并使用具有跳跃连接的基于 CNN 的解码器输出分割结果<sup>[14-15]</sup>。Swin UNETR 利用 Swin-Transformer 作为编码器来提取多尺度特征,并采用基于 CNN 的解码器来生成结果,实现最先进的医学图像分割结果<sup>[16-17]</sup>。Chen 等<sup>[18]</sup>提出的 TransUNet 兼具 Transformers 和 U-Net 的优点。Transformer 对来自 CNN 所提取的特征图的标记化图像块进行编码,并将其作为提取全局上下文信息的输入序列。随后,解码器对编码特征进行上采样操作,再将其与高分辨率 CNN 特征图相结合以实现精确定位。

在医学图像处理领域,去噪一直是一个基础且至关重要的问题<sup>[19]</sup>。传统的去噪方法通常依赖于滤波器或变换域处理,比如小波变换、均值滤波、高斯滤波等,但这些方法往往难以有效应对复杂的噪声并同时保持图像细节<sup>[20]</sup>。近年来,随着深度学习技术的快速发展,去噪扩散模型逐渐受到了广泛的关注。它的基本原理是利用深度神经网络模型,通常是 U-Net 网络模型,通过训练将含有噪声的输入图像映射到相应的干净图像空间<sup>[21]</sup>。相较于传统的医学图像分割方法,去噪扩散模型具备更强的非线性建模能力和更优越的泛化性能,能够更有效地捕获图像中的复杂结构和细节信息<sup>[22]</sup>。在 Wu 等<sup>[23]</sup>提出的 MedSegDiff 中,为了增强 DPM 中医学图像分割的逐步区域注意力,他们提出了动态条件编码,为每个采样步骤建立了状态自适应条件;并且,他们进一步提出特征频率解析器,以消除该过程中高频噪声分量的负面影响。在 MedSegDiff 的基础上, Wu 等提出了一种新颖的基于 Transformer 的条件 U-Net 框架,以及一种新的频谱空间 Transformer 来建模噪声和语义特征之间的相互作用。这种架构的改进催生了一种新的基于扩散的医学图像分割方法,称为 MedSegDiff-V2<sup>[24]</sup>,它显著提高了 MedSegDiff 的分割性能。Xing 等<sup>[25]</sup>提出了 Diff-UNet,并将其用于医学体积分割。他们将扩散模型集成到标准 U 形架构中,有效地从输入体积中提取语义信息,从而为医学体积分割提供出色的像素级表示。为了增强扩散模型预测结果的鲁棒性,他们还在推理过程中引入了基于步进不确定性的融合模块,以合并每个步骤的扩散模型的输出。Wolleb 等<sup>[26]</sup>利用扩散模型来解决二维医学图像分割

的问题,并在测试阶段通过对每个扩散步骤的输出结果进行求和来进行融合,以提升分割结果的鲁棒性。

尽管去噪扩散模型在图像分割领域取得了巨大的成功,但对医学图像的分割仍然存在一些挑战和待解决的问题。例如,医学图像具有较大的个体差异性,常受到多种形式的噪声干扰,其中包括电子设备的信号干扰、成像过程中的运动伪影,以及基于物理特性的噪声。同时,医学图像中的分割目标的分割边界不确定性和区域模糊化现象普遍存在。因此,上述模型难以有效地处理不确定的分割边界和模糊的区域。

为了克服这些问题,本文提出了一种基于模糊逻辑驱动的医学图像扩散融合网络分割模型——FDiff-Fusion。该分割方法通过引入模糊学习模块(Fuzzy Learning Module, FLM)来增强 U-Net 网络对医学图像的边界不确定和区域模糊的建模能力,并且设计了一个基于迭代注意力机制的融合模块(Iterative Attention Fusion, IAF)对去噪过程的多个时间步的预测结果进行融合,以便在测试阶段得到更准确的分割结果。本文方法旨在增强医学图像分割的准确性和鲁棒性,以及更有效地处理分割边界不确定和区域模糊的情况,这对于改善医学图像分割技术并提高临床诊断的准确性具有重要意义。

本文的主要贡献如下:

1)为了克服医学图像的分割边界不确定和区域模糊等问题,引入了一个模糊学习模块,通过设置多个模糊隶属度函数和使用模糊规则处理来增强模型对分割边界不确定和区域模糊的建模能力;

2)为了提高模型预测结果的准确性和鲁棒性,提出一个基于迭代注意力机制的融合模块,将每个时间步分割图的局部上下文信息添加到注意力模块中的全局上下文信息中来对最终的分割结果进行优化;

3)为了全面评估 FDiff-Fusion 模型的分割性能,使用多个先进的深度学习模型在 BRATS 2020 脑肿瘤数据集和 BTCV 腹部多器官数据集上进行了对比实验,实验结果证明了该模型的有效性。

本文第 2 章介绍了相关基础知识,主要介绍了 U-Net 网络模型、模糊学习理论和去噪扩散模型;第 3 章介绍了 FDiff-Fusion 的模型架构;第 4 章介绍了实验结果并对实验结果进行分析;最后进行总结与展望。

## 2 基础知识

### 2.1 U-Net 网络模型

U-Net 网络的核心思想是将编码器和解码器相结合,以实现端到端的图像分割。U-Net 特有的 U 形结构,使得网络能够同时捕获图像的全局和局部信息,从而有效地克服传统方法的不足。同时,U-Net 网络还采用了跳跃连接,使得网络能够在不同层级之间传递信息,进一步提升了分割的准确性和鲁棒性。下面对 U-Net 网络结构的组成部分进行介绍。

编码器(收缩路径):编码器负责将输入图像进行多次下采样,以提取图像的高级特征。通常采用卷积层和池化层构建编码器,通过逐步减小特征图的空间分辨率,同时增加特征

图的通道数,来捕获更丰富的特征信息。这一过程有助于将输入图像中的细节信息逐步提炼为更高级别的抽象特征,为后续的图像分割任务提供了基础。

解码器(扩展路径):解码器负责将编码器提取的特征图进行上采样操作,并与对应的编码器层特征图连接,以恢复原始图像的空间分辨率,并生成分割结果。解码器包含反卷积层和跳跃连接,跳跃连接有助于网络更充分地利用不同层级的特征信息,以提升分割的精度。

跳跃连接:U-Net 网络中的跳跃连接是其关键特征之一。通过跳跃连接,编码器中的特征图可以直接与解码器中相应层级的特征图相连接。这种连接机制使得网络能够更充分地利用不同层级的特征信息,防止信息丢失,从而提高了分割的准确性和鲁棒性。

### 2.2 模糊学习理论

模糊学习是一种基于模糊集理论的机器学习方法,旨在处理数据中的模糊性或不确定性<sup>[27]</sup>。模糊集理论用于描述事物之间的模糊关系,并对不确定性进行建模。以下是模糊学习的一些关键概念和方法。

模糊集理论:模糊集理论引入模糊概念,描述事物之间的不确定性和模糊性。在该理论中,一个元素可以部分属于一个集合,而不是完全属于或不属于,这主要通过隶属度函数表示<sup>[28-29]</sup>。

模糊推理:通过引入模糊集合和模糊逻辑运算来处理模糊性问题<sup>[30]</sup>。模糊集合是一个元素的隶属度(Membership Degree)在 0 到 1 之间的函数,表示该元素属于该集合的程度。模糊逻辑运算包括模糊与、模糊或、模糊非等运算,用于处理模糊命题之间的关系。其基本思想是根据模糊规则进行推断,其中模糊规则由条件部分和结论部分组成。条件部分是一组模糊命题,结论部分是对应的模糊结论。模糊推理引入了模糊推理引擎,通过模糊规则和输入的模糊命题进行推理,从而得到模糊的结论<sup>[31]</sup>。

边缘检测:模糊学习也可以用于医学图像中的边缘检测任务。传统的医学图像分割算法往往难以有效地处理模糊的边界或复杂的背景情况,而模糊学习方法可以更好地识别模糊边缘,并提高边缘检测的准确性<sup>[32-33]</sup>。

模糊隶属度函数:在一些基于传统的图像分割算法中,可以引入模糊隶属度函数来描述像素与不同组织或结构之间的隶属关系<sup>[34]</sup>。这种模糊隶属度函数可以通过模糊学习方法进行学习和优化,从而更准确地刻画图像中的模糊性和不确定性<sup>[35]</sup>。

### 2.3 去噪扩散模型

去噪扩散模型包括前向过程和反向过程。在前向过程(加噪过程)中,对图像逐步施加噪声,直至图像被破坏变成完全的高斯噪声图像;然后,在反向过程(去噪过程)中,使用深度神经网络来学习从高斯噪声图像还原为原始图像的过程<sup>[36-37]</sup>。具体来说,如图 1 所示,实线表示反向过程,虚线则表示前向过程。前向过程在原始图像 $x_0$ 上逐步添加高斯噪声,每一步得到的图像 $x_t$ 只与上一步得到的 $x_{t-1}$ 相关,直至第  $T$  步的图像 $x_T$ 变为纯高斯噪声图像。反向过程则是逐步去

除高斯噪声的过程,首先给定高斯噪声图像 $x_T$ ,通过逐步去噪操作,恢复到原始图像 $x_0$ 。

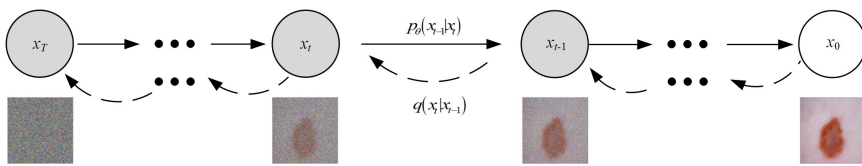


图1 去噪扩散模型的前向过程和反向过程

Fig. 1 Forward process and reverse process of denoising diffusion model

由于前向过程中图像 $x_t$ 只与上一步得到的 $x_{t-1}$ 相关,因此该过程可被视为马尔可夫过程,满足:

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (1)$$

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (2)$$

其中, $q(\cdot)$ 代表前向过程, $q(x_t | x_{t-1})$ 代表图像 $x_{t-1}$ 到图像 $x_t$ 的前向加噪过程, $\mathcal{N}(\cdot)$ 代表正态分布, $\mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$ 代表均值为 $\sqrt{1-\beta_t}x_{t-1}$ 、方差为 $\beta_t I$ 的正态分布。 $\beta_t \in (0, 1)$ 为高斯分布的方差超参序列,并且 $\beta_t$ 满足条件 $\beta_1 < \beta_2 < \dots < \beta_T$ 。

随后,利用重参数技巧,通过式(2)采样 $x_t$ ,即生成随机变量 $\epsilon_t \sim \mathcal{N}(0, 1)$ ,令 $\alpha_t = 1 - \beta_t$ ,以及 $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ ,从而可以得到:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t \quad (3)$$

在反向过程中,由于无法直接推断出 $q(x_{t-1} | x_t)$ ,因此使用深度学习模型 $p_\theta$ 来拟合分布 $q(x_{t-1} | x_t)$ ,模型参数为 $\theta$ :

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t) \quad (4)$$

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (5)$$

其中, $p_\theta(\cdot)$ 是用来拟合分布 $q(x_{t-1} | x_t)$ 的深度学习模型,通常是 U-Net 模型。 $p_\theta(x_{0:T})$ 表示从纯噪声图像 $x_T$ 逆向生成清晰图像 $x_0$ 的联合概率分布, $p(x_T)$ 表示纯噪声图像 $x_T$ 的概率,即 $p(x_T)$ 是标准正态分布。 $p_\theta(x_{t-1} | x_t)$ 表示在给定时间步 $t$ 的图像 $x_t$ 条件下,前一时间步 $x_{t-1}$ 的概率。 $\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$ 表示正态分布,均值为 $\mu_\theta(x_t, t)$ ,代表在时间步 $t$ 和给定图像 $x_t$ 的条件下,模型参数 $\theta$ 所预测的 $x_{t-1}$ 的均值;方差为 $\Sigma_\theta(x_t, t)$ ,代表在时间步 $t$ 和给定图像 $x_t$ 的条件下,模型参数 $\theta$ 所预测的 $x_{t-1}$ 的方差; $t$ 是加噪(或者去噪)的步数。虽然无法直接推导出 $q(x_{t-1} | x_t)$ ,但是在知道 $x_0$ 的情况下,可以通过贝叶斯公式得到 $q(x_{t-1} | x_t, x_0)$ 为:

$$q(x_{t-1} | x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I) \quad (6)$$

其中, $\mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$ 表示正态分布,均值是 $\tilde{\mu}(x_t, x_0)$ ,它依赖于当前时间步的图像 $x_t$ 和初始时间步图像 $x_0$ ;方差是 $\tilde{\beta}_t I$ , $\tilde{\beta}_t$ 表示时间步 $t$ 的噪声强度。在式(6)的推导过程中,通过贝叶斯公式可以将去噪扩散模型的反向过程转换为前向过程,且得到的概率密度函数和高斯概率密度函数的指数部分能够对应,即有:

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t \quad (7)$$

$$\tilde{\mu}(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 \quad (8)$$

其中, $\beta_t \in (0, 1)$ 为高斯分布的方差超参序列,并且满足条件

$\beta_1 < \beta_2 < \dots < \beta_T$ ,  $\alpha_t = 1 - \beta_t$ ,以及 $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ 。根据式(7)和式(8),可以通过计算得到 $q(x_{t-1} | x_t, x_0)$ 的分布。首先,根据式(3)可以得到:

$$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t) \quad (9)$$

然后将式(9)代入到式(8)中得到:

$$\tilde{\mu}_t = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t) \quad (10)$$

因为需要使用深度学习模型 $p_\theta$ 来拟合分布 $q(x_{t-1} | x_t)$ ,根据式(5),希望训练模型 $\mu_\theta(x_t, t)$ 以预估式(10)。由于 $x_t$ 在训练阶段会作为输入,因此它是已知的,可以转而让模型去预估噪声 $\epsilon_t$ ,即令:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) \quad (11)$$

其中, $\epsilon_\theta(x_t, t)$ 是通过神经网络(通常是 U-Net)从当前噪声图像 $x_t$ 和时间步 $t$ 预测的噪声成分。根据式(11),最终可得:

$$x_{t-1} = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right), \Sigma_\theta(x_t, t)\right) \quad (12)$$

式(12)描述了去噪扩散网络从时间步 $t$ 到时间步 $t-1$ 的条件概率分布,且是一个正态分布。其中,均值为 $\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t))$ ,表明 $x_{t-1}$ 的均值是根据当前时间步 $x_t$ 和一个由神经网络 $\epsilon_\theta$ 预测的噪声成分计算得到; $\Sigma_\theta(x_t, t)$ 为对应的方差,表示噪声在各个维度上的方差。

### 3 FDiff-Fusion

传统的医学图像分割方法通常是直接将医学图像作为输入,并通过训练模型来预测相应的分割标签图<sup>[38]</sup>。然而,与这种直接输入医学图像的方法不同,扩散模型采用了一种不同的策略。扩散模型的学习过程并非直接预测分割标签图,而是专注于学习去除高斯噪声的过程<sup>[39-40]</sup>。具体而言,在训练阶段,扩散模型接受医学图像以及带有高斯噪声的分割标签图作为输入,然后通过深度学习模型学习去除医学图像中的高斯噪声,生成清晰的分割结果<sup>[41]</sup>。这种方法的关键在于模型通过去噪的过程来提升分割结果的质量,而不是直接对医学图像进行分割预测<sup>[42]</sup>。然而,尽管去噪扩散模型在图像分割领域取得了不错的成果,但是在医学图像中对分割边界不确定性和区域模糊化的处理仍然存在挑战。为了克服上述问题,并提高分割结果的准确性和鲁棒性,本文提出了一种基

于模糊逻辑驱动的医学图像扩散融合网络分割模型——FDiff-Fusion,其结构如图2所示。

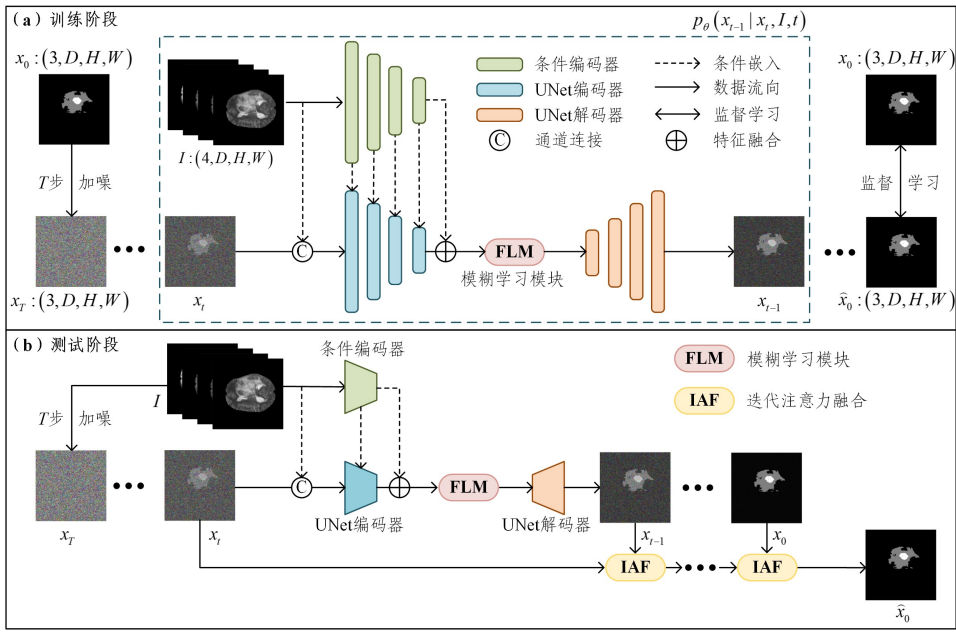


图2 FDiff-Fusion 模型结构

Fig. 2 FDiff-Fusion model structure

### 3.1 FDiff-Fusion 模型

FDiff-Fusion 模型由两个阶段组成,分别为前向过程和反向过程。在前向过程中,原始医学图像  $x_0 \in \mathbb{R}^{C \times D \times H \times W}$  通过逐步添加高斯噪声得到纯噪声图像  $x_T \in \mathbb{R}^{C \times D \times H \times W}$ 。具体而言,对于原始图像  $x_0$ ,使用  $T$  个时间步对马尔可夫过程执行变分推理,用来学习训练数据分布。此过程可以表示为:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t} x_{t-1}, (1 - \alpha_t) I) \quad (13)$$

其中,  $\mathcal{N}(x_t; \sqrt{\alpha_t} x_{t-1}, (1 - \alpha_t) I)$  表示  $x_t$  在给定的  $x_{t-1}$  时服从一个均值为  $\sqrt{\alpha_t} x_{t-1}$ 、方差为  $(1 - \alpha_t) I$  的正态分布,  $q(x_t | x_{t-1})$  代表图像  $x_{t-1}$  到图像  $x_t$  的前向加噪过程,  $t$  表示马尔可夫链中的第  $t$  个时间步,  $\alpha_t = 1 - \beta_t$  是控制第  $t$  个时间步添加高斯噪声强弱程度的噪声控制器,并且有  $\alpha_t \in (0, 1)$ ,  $\alpha_1 > \dots > \alpha_T$ 。

在反向过程中,训练 U-Net 通过学习去除添加的高斯噪声来使得纯噪声图片  $x_T$  恢复到原始图像  $x_0$ 。此过程可以表示为:

$$p_\theta(x_{0:T-1} | x_T) = \prod_{t=1}^T p_\theta(x_{t-1} | x_t) \quad (14)$$

其中,  $p_\theta(x_{0:T-1} | x_T)$  表示给定最终的纯噪声图像  $x_T$ ,生成前面所有图像  $x_0, x_1, \dots, x_{T-1}$  的联合概率分布,  $p_\theta(x_{t-1} | x_t)$  表示从时间步  $t$  的图像  $x_t$  生成前一步  $x_{t-1}$  的条件概率分布,  $\theta$  为模型参数。

反向过程主要通过去除添加的高斯噪声来得到清晰的分割结果。为了更好地学习去噪过程,采用了 U-Net 和一个用于对原始图像进行多尺度特征提取的 CNN 来学习去噪过程。

如图2所示,在测试阶段中,给定3D医学图像  $I \in \mathbb{R}^{C \times D \times H \times W}$  作为去噪扩散模型的每个时间步的条件嵌入。 $C, D, H, W$  分别为3D医学图像的通道数、深度、高度和宽度。首先,对3D医学图像  $I$  添加  $T$  步高斯噪声,从而获得纯噪声图像  $x_T$ 。随后,对纯噪声图像  $x_T$  进行逐步去噪操作,以第  $t$  个

时间步为例,将噪声图像  $x_t$  和原始医学图像  $I$  通过通道进行连接,并将连接后的结果输送到 U-Net 的编码器中,以获得多尺度特征  $f_t \in \mathbb{R}^{C \times D \times H \times W}$ 。同时,为了更好地引入原始3D医学图像作为条件嵌入,将3D医学图像  $I$  作为 CNN 的输入,以获得多尺度条件特征  $f_i \in \mathbb{R}^{C \times D \times H \times W}$ 。由于  $f_t$  和  $f_i$  包含相同数量和大小特征,因此将对应尺度的特征相加来获得融合特征。但是,因为3D医学图像中分割目标的边界不确定性和区域模糊化现象普遍存在,仅将 U-Net 的编码器所提取的多尺度特征  $f_t$  和 CNN 提取的多尺度特征  $f_i$  简单相加后作为每个时间步的编码特征,难以处理分割目标的不确定边界和模糊区域问题。为了解决这个问题,本文提出了一个模糊学习模块。一方面,原始3D医学图像包含准确的分割目标,但是很难确定其分割边缘;另一方面,当前时间步的分割图包含增强的分割目标,但是不够准确。因此,在 U-Net 的编码器和解码器的连接路径上使用了模糊学习模块来对编码特征进行模糊规则处理。接着,将经过 U-Net 编码器处理的特征通过跳跃连接输入到解码器中,得到第  $t$  个时间步的预测结果  $x_{t-1}$ 。最后,为了提高模型分割的准确性和鲁棒性,将每个时间步所获得的分割图通过迭代注意力进行融合,从而得到最终的分割结果  $\hat{x}_0 \in \mathbb{R}^{C \times D \times H \times W}$ 。

相对于测试阶段只将3D医学图像作为模型的输入,在训练阶段,本文将3D医学图像  $I \in \mathbb{R}^{C \times D \times H \times W}$  和对应的分割标签  $x_0 \in \mathbb{R}^{C \times D \times H \times W}$  同时输入到模型作为输入。其中,3D医学图像作为条件信息嵌入,对应的分割标签则用于对其进行  $T$  步加噪,从而得到高斯噪声图像  $x_T \in \mathbb{R}^{C \times D \times H \times W}$ 。经过  $T$  步去噪操作,可以得到模型预测的分割标签图  $\hat{x}_0 \in \mathbb{R}^{C \times D \times H \times W}$ ,随后将其与分割标签图  $x_0$  进行损失计算,从而更新模型的参数。FDiff-Fusion 是通过结合传统图像分割网络所使用的 MSE Loss, BCE Loss 和 Dice Loss 来进行训练的,所以模型的总损失  $\mathcal{L}_{total}$  为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{mse}}(\hat{x}_0, x_0) + \mathcal{L}_{\text{bce}}(\hat{x}_0, x_0) + \mathcal{L}_{\text{dice}}(\hat{x}_0, x_0) \quad (15)$$

### 3.2 模糊学习模块

为了解决 3D 医学图像中分割目标的边界不确定性和区域模糊化问题,本文提出了一个模糊学习模块。具体来说,对于 U-Net 所提取的特征,设计一个模糊残差连接作为模糊学习模块来进行处理。模糊学习模块包括模糊隶属度函数和模糊规则。如图 3 所示,对于每一个通道,  $M$  个模糊隶属度函数被应用到特征图中的每个特征点上。模糊隶属度函数在每个通道上保持不变,但在不同的通道上可以改变。每个模糊隶属度函数赋予每个通道的每个特征点一个模糊类标签,用来描述其与其他特征点之间的相似程度。在隶属函数层上,采用高斯函数作为隶属函数:

$$F'_{x,y,z,k,c} = e^{-\frac{(F_{x,y,z,c} - \mu_{k,c})^2}{\sigma_{k,c}^2}} \quad (16)$$

$$x=1, \dots, D; y=1, \dots, H; z=1, \dots, W$$

其中,  $F_{x,y,z,c}$  是通道  $C$  的特征图中的一个特征点,  $(x, y, z)$  代表特征点对应的位置;  $\mu_{k,c}$  和  $\sigma_{k,c}$  是第  $k$  个高斯隶属度函数的均值和标准差;  $F'_{x,y,z,k,c}$  表示通道  $C$  中位置为  $(x, y, z)$  的特征点的第  $k$  个输出模糊类标签。随后,将模糊规则“与(AND)”应用到每一个特征点的所有模糊隶属度函数上,通过式(17)得到输入特征图的最终模糊特征信息  $F_{x,y,z,c}$ :

$$F_{x,y,z,c} = \prod_{k=1}^M F'_{x,y,z,k,c} \quad (17)$$

式(17)根据每个隶属度函数的输出,可以得到每个特征点的模糊特征信息  $F_{x,y,z,c}$ ,进而就能得到 U-Net 编码器提取的特征图经过模糊学习模块处理后的模糊特征图  $\mathbf{F} \in \mathbb{R}^{C \times D \times H \times W}$ 。

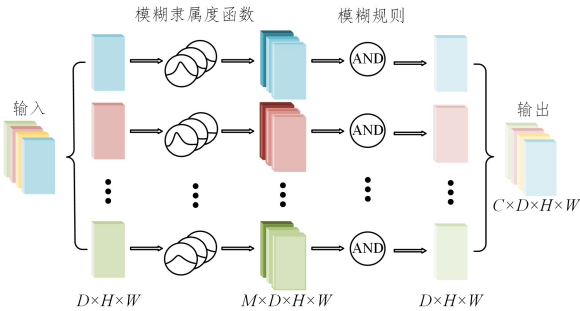


图3 模糊学习模块结构

Fig. 3 Fuzzy learning module structure

作为网络的一部分,模糊学习模块可以被视为由来自高斯隶属度函数的多个参数描述的模糊层。注意,参数  $\mu_{k,c}$  和  $\sigma_{k,c}$  在同一通道中保持不变,但是在不同通道之间可能有所不同。这是因为同一通道的特征点是由同一个卷积核提取的,会具有相似的特征;而不同通道的特征点采用不同的卷积核得到,具有不同的特征。此外,为了实现模糊学习模块,模糊隶属度函数的数量即  $M$ ,是唯一需要预设的超参数。

最后,如图 4 所示,模糊学习模块的输出为模糊度量  $\mathbf{F}$ ,其大小与经过 U-Net 处理过的特征图度量  $f_i$  相同,因此将它们通过残差连接进行特征融合,如式(18)所示:

$$\mathbf{F}_t = \text{BN}(\text{FLM}(f_i)) + \text{BN}(f_i) \quad (18)$$

其中,U-Net 的输出特征图和模糊度量通过批量归一化(Batch Normalization, BN)进行处理,以约束其动态范围。

然后,采用简单的加法运算来整合模糊逻辑信息。

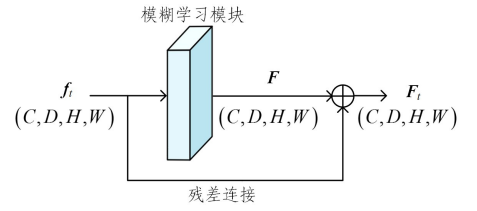


图4 模糊残差连接

Fig. 4 Fuzzy residual connection

### 3.3 迭代注意力融合

在传统的医学图像分割任务中,最后一轮的预测被作为模型的最终分割结果。然而,在 FDiff-Fusion 的反向过程中,每个时间步都会生成预测结果。因此,为了提高模型分割的准确性和鲁棒性,将反向过程中的每个时间步所获得的分割图进行融合来得到最终的分割结果。为此,本文设计了一个基于迭代注意力机制的融合模块来对模型的预测结果进行优化。传统的融合方式是采用简单的加法进行融合或者进行加权融合,与之不同,本文通过设计一个迭代注意力机制对去噪过程得到的分割图进行更深层次的融合。

#### 3.3.1 多尺度通道注意力融合

多尺度通道注意力模块(Multi-scale Channel Attention Module, MS-CAM)的关键思想是通过改变空间池的大小,在多个尺度上实现通道注意力。为了尽可能保持轻量级,将局部上下文添加到注意力模块中的全局上下文中,并且选择逐点卷积(Point-wise Conv, PWConv)作为局部通道上下文聚合器,它仅利用每个空间位置的逐点通道交互。如图 5 所示,给定一个具有  $C$  个通道的特征  $x \in \mathbb{R}^{C \times D \times H \times W}$ ,可以计算得到局部通道上下文  $L(x) \in \mathbb{R}^{C \times D \times H \times W}$ :

$$L(x) = \mathcal{B}(\text{PWConv}_2(\delta(\mathcal{B}(\text{PWConv}_1(x)))))) \quad (19)$$

其中,  $\delta$  表示修正线性单元(ReLU),  $\mathcal{B}$  表示批量归一化。  $\text{PWConv}_1$  和  $\text{PWConv}_2$  的内核大小分别为  $\frac{C}{r} \times C \times 1 \times 1$  和  $C \times \frac{C}{r} \times 1 \times 1$ 。

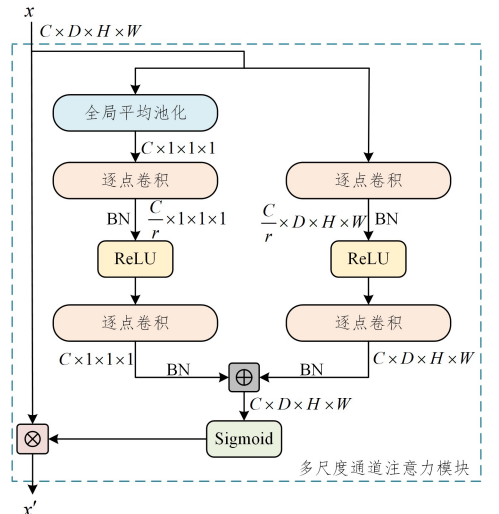


图5 多尺度通道注意力模块

Fig. 5 Multiscale channel attention module

值得注意的是,  $L(x)$  与输入特征具有相同的形状, 可以保留并突出低级特征中的微妙细节。同样地, 通过一个全局平均池化层(Global Average Pooling, GAP), 全局通道上下文  $G(x) \in \mathbb{R}^{C \times D \times H \times W}$  可表示为:

$$G(x) = \mathcal{B}(\text{PWConv}_2(\delta(\mathcal{B}(\text{PWConv}_1(g(x))))) \quad (20)$$

其中,  $g(x)$  是全局平均池化,  $\delta$  表示修正线性单元(ReLU),  $\mathcal{B}$  表示批量归一化,  $\text{PWConv}_1$  和  $\text{PWConv}_2$  表示逐点卷积。给定全局通道上下文  $G(x)$  和局部通道上下文  $L(x)$ , MS-CAM 的细化特征  $x' \in \mathbb{R}^{C \times D \times H \times W}$  可以由式(21)得到:

$$x' = x \otimes M(x) = x \otimes \sigma(L(x) \oplus G(x)) \quad (21)$$

其中,  $M(x) \in \mathbb{R}^{C \times D \times H \times W}$  表示 MS-CAM 生成的注意力权重,  $\sigma$  表示 sigmoid 激活函数,  $\oplus$  表示广播加法,  $\otimes$  表示逐元素乘法。

### 3.3.2 注意力融合

基于多尺度通道注意力模块, 注意力融合(Attention Fusion, AF)可以表示为:

$$Z = M(X \oplus Y) \otimes Y + (1 - M(X \oplus Y)) \otimes X \quad (22)$$

其中,  $Z \in \mathbb{R}^{C \times D \times H \times W}$  是融合后的特征,  $\oplus$  表示特征的逐元素求和操作。AF 如图 6(a) 所示, MS-CAM 模块的输出为特征融合权重  $M(X \oplus Y)$ , 虚线则表示  $1 - M(X \oplus Y)$ 。在这里, 融合权重的范围在 0 到 1 之间,  $1 - M(X \oplus Y)$  的范围也一样, 这使得网络能够在输入之间进行软选择或者加权平均。

### 3.3.3 迭代注意力融合

对于注意力模块的输入特征  $X$  和  $Y$ , 将其进行简单的逐元素求和操作可能会影响最终的融合权重。由于它仍然是特征融合问题, 一种直观的方法是由另一个注意力模块来融合输入特征。本文将这种两阶段方法称为迭代注意力融合(Iterative Attention Fusion, IAF), 如图 6(b) 所示。由此, 式(22)中的  $X \oplus Y$  可以被重新表述为:

$$X \oplus Y = M(X + Y) \otimes Y + (1 - M(X + Y)) \otimes X \quad (23)$$

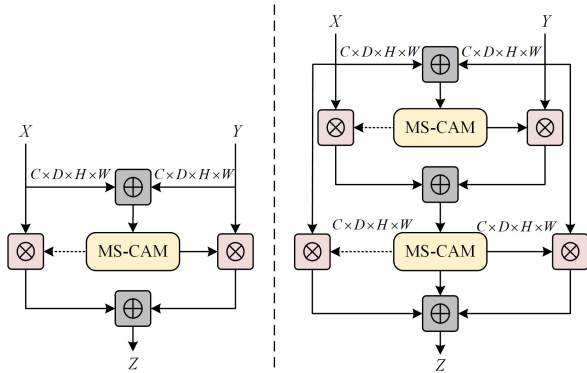


图 6 注意力融合和迭代注意力融合结构

Fig. 6 Attention fusion and iterative attention fusion structure

算法 1 展示了 FDiff-Fusion 模型的具体流程。

#### 算法 1 FDiff-Fusion 模型

输入: 医学图像  $I \in \mathbb{R}^{C \times D \times H \times W}$ , 掩码图  $x_0 \in \mathbb{R}^{C \times D \times H \times W}$

输出: 预测的分割标签图  $\hat{x}_0 \in \mathbb{R}^{C \times D \times H \times W}$

1. for epoch in max\_epoch do
2.  $t \sim \text{Uniform}(\{1, \dots, T\})$ ;
3. Sampling  $\beta_t$ ;

4.  $\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{i=1}^T \alpha_i$ ;
5.  $\epsilon_t \sim \mathcal{N}(0, 1)$ ;
6.  $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_t$ ; /\* 加噪过程 \*/
7. for  $t \leftarrow T; 1$  do /\* 去噪过程 \*/
8.  $f_t = \text{Conv}(I)$ ; /\* 使用 CNN 提取多尺度特征  $f_t \in \mathbb{R}^{C \times D \times H \times W}$  \*/
9.  $f_t = \text{UNetencoder}(x_t)$ ; /\* 将  $x_t$  作为 U-Net 的输入, 经过编码器的处理得到多尺度特征  $f_t \in \mathbb{R}^{C \times D \times H \times W}$  \*/
10.  $f_t = f_t + f_t$ ; /\* 将  $f_t$  和  $f_t$  对应尺度的特征相加来获得融合特征 \*/
11.  $F_t = \text{BN}(\text{FLM}(f_t)) + \text{BN}(f_t)$ ; /\* 对于 U-Net 编码器获得的特征通过模糊残差连接获得模糊特征信息 \*/
12.  $x_{t-1} = \text{UNetdecoder}(F_t)$ ; /\* 将模糊特征信息输入到 U-Net 解码器中, 得到时间步  $t$  的预测结果  $x_{t-1}$  \*/
13. end for
14. for  $t \leftarrow T-1; 1$  do /\* 迭代注意力融合 \*/
15.  $x_{t-1} = M(x_t \oplus x_{t-1}) \otimes x_{t-1} + (1 - M(x_t \oplus x_{t-1})) \otimes x_t$ ; /\* 逐时间步进行图像融合 \*/
16. return  $\hat{x}_0$ ;
17. end for

## 4 实验

### 4.1 数据集

本文使用了两个公开的分割数据集, 包括 BRATS 2020 脑肿瘤数据集和 BTCV 腹部多器官数据集。

BRATS(Brain Tumor Segmentation)2020 脑肿瘤数据集是一个用于脑肿瘤分割的公开数据集, 常用于医学图像处理<sup>[43]</sup>。该数据集包含来自不同医疗中心的 369 个对齐的多模态脑部 MRI 扫描图像, 每个模态都有  $155 \times 240 \times 240$  的体积。其中包括 4 个模态: T1 加权(T1)、T1 对比增强(T1Gd)、T2 加权(T2)和液体衰减反转恢复序列(FLAIR)<sup>[44]</sup>。每个模态的图像都经过配准和预处理, 以便在相同的空间位置进行比较。BRATS 2020 脑肿瘤数据集还提供了脑肿瘤的分割标签, 包括 4 种类型的肿瘤组织: 胶质母细胞瘤(Glioma)的坏死区域、囊肿、增强区域和正常组织。训练集、验证集和测试集的分割比例为 0.7, 0.1 和 0.2。

BTCV 数据集由 30 例 3D 腹部多器官图像组成, 每张 3D 图像有 13 个器官分割目标<sup>[45]</sup>。所有数据都重新采样到相同的空间。其中, 18 个 3D 腹部多器官图像用于模型的训练, 12 个图像用于测试。

### 4.2 实验参数设置

本实验使用基于 Windows11 系统的工作站, 实验平台为 PC(Intel(R) Core(TM) i9-10940X @ 3.30 GHz), 显卡为 NVIDIA GeForce RTX4090, 内存容量为 64 GB, 开发工具为 JetBrains PyCharm 2021.2.3 专业版, 使用 Python 语言实现实验中的相关算法。在实验初始阶段, 对所有实验数据样本引入随机裁减、空间填充、随机翻转、强度缩放和偏移来进行数据增强。在训练阶段, 损失函数结合了 Dice 损失、BCE 损失和 MSE 损失。对于 BRATS 2020 脑肿瘤数据集, 采用权重衰减为  $1 \times 10^{-3}$  的 AdamW 优化器, 最大迭代次数为 300, 训练批次大小为 2, 学习率为  $1 \times 10^{-4}$ , 并且使用余弦退火算法

更新学习率。每次迭代随机采样  $n$  个补丁(每个补丁的大小为  $96 \times 96 \times 96$ )进行训练。在测试中,将 DDIM 采样步数设置为 10,每个样本的大小为  $96 \times 96 \times 96$ 。滑动窗口重叠率为 0.5。对于 BTCV 数据集,采用权重衰减为  $1 \times 10^{-3}$  的 AdamW 优化器,最大迭代次数为 3 000,训练批次大小为 1,学习率为  $2 \times 10^{-4}$ ,并且使用余弦退火算法更新学习率。每次迭代随机采样  $n$  个补丁(每个补丁的大小为  $96 \times 96 \times 96$ )进行训练。在测试中,将 DDIM 采样步数设置为 10,每个样本的大小为  $96 \times 96 \times 96$ 。滑动窗口重叠率为 0.5。

### 4.3 评价指标

为了评价模型的分割性能,采用 Dice 相似系数(Dice Similarity Coefficient, DSC)和 HD95 距离(Hausdorff Distance at 95th Percentile)作为评价指标。

Dice 相似系数是一种常用的评价指标,用于衡量两个集合的相似性。其系数的取值范围为 0 到 1 之间,其中 1 表示完全重合,即模型预测的分割结果与真实分割结果完全一致;0 表示完全不重合,即两个分割结果没有任何重叠的像素。通常情况下,Dice 系数越接近 1,表示模型预测的分割结果与真实分割结果越相似,分割质量越高。相反,Dice 系数越接近 0,表示模型预测的分割结果与真实分割结果之间的差异越大,分割质量越低。DSC 计算公式如下:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (24)$$

其中,  $A$  和  $B$  分别表示地面真实标签和模型预测的结果;  $|A \cap B|$  表示模型预测的分割结果与真实分割结果的交集(即重叠的像素数);  $|A|$  和  $|B|$  分别表示地面真实标签和模型

预测结果的像素数。

Hausdorff 距离是一种衡量两个集合之间的相似度的度量,考虑了两个集合之间的最大距离。在图像分割任务中,一个集合代表真实的分割结果,另一个集合代表模型预测的分割结果。Hausdorff 距离衡量了两个分割结果之间的最大不一致程度,即一个分割结果中的点到另一个分割结果中的最远点的距离。

HD95 是 Hausdorff 距离的一个百分位数变体,它计算了在 95% 的情况下的距离。换句话说,HD95 表示真实分割边界到模型预测边界的距离中,95% 的距离值都小于或等于 HD95 的值。因此,HD95 越小,表示模型预测的分割结果与的匹配程度越好,分割质量越高。HD95 的计算式如下:

$$HD(A, B) = \max \left\{ \max_{g' \in A, p' \in B} \|g' - p'\|, \max_{p' \in A, g' \in B} \|p' - g'\| \right\} \quad (25)$$

其中,  $A$  和  $B$  分别表示地面真实标签和模型预测结果;  $\|g' - p'\|$  和  $\|p' - g'\|$  表示集合  $A$  和集合  $B$  之间的距离。HD95 使用地面真相和预测表面点集之间距离的第 95 个百分位数。因此,在计算 HD 时,极小一部分异常值的影响是最小的。

### 4.4 实验结果分析

为了验证 FDiff-Fusion 对于医学图像分割的有效性,在 BRATS 2020 脑肿瘤数据集和 BTCV 腹部多器官数据集上将 FDiff-Fusion 与 SOTA 分割模型进行了比较,定量结果如表 1 和表 2 所列。

表 1 BRATS 2020 脑肿瘤数据集的对比结果

Table 1 Comparison results of BRATS 2020 brain tumor dataset

方法	Dice 分数/%				HD95 距离/mm			
	全肿瘤	肿瘤核心	增强肿瘤	平均	全肿瘤	肿瘤核心	增强肿瘤	平均
UNETR	88.94	79.89	72.24	80.36	4.315	5.952	4.729	4.999
TransBTS	89.36	81.24	75.19	81.93	3.660	3.094	3.604	3.453
Diff-UNet	89.37	83.15	76.92	83.14	2.118	3.489	2.270	2.626
FDiff-Fusion	<b>90.51</b>	<b>84.37</b>	<b>77.61</b>	<b>84.16</b>	<b>2.115</b>	<b>3.086</b>	<b>2.220</b>	<b>2.473</b>

表 2 BTCV 腹部多器官数据集的对比结果

Table 2 Comparison results of BTCV abdominal multi-organ dataset

方法	平均									
	Dice 分数/%	HD95 距离/mm	脾脏/%	右肾/%	左肾/%	胆囊/%	肝脏/%	胃/%	主动脉/%	胰腺/%
VNet	68.81	27.45	80.56	80.75	77.10	51.87	87.84	56.98	75.34	40.05
TransUNet	78.03	30.72	85.62	77.98	82.64	64.54	93.94	77.02	87.21	55.34
Diff-UNet	82.17	10.23	86.61	83.19	83.79	<b>75.32</b>	94.65	75.02	86.62	<b>72.16</b>
FDiff-Fusion	<b>83.82</b>	<b>7.98</b>	<b>89.08</b>	<b>84.12</b>	<b>84.51</b>	74.01	<b>95.33</b>	<b>84.02</b>	<b>88.48</b>	70.98

在 BRATS 2020 脑肿瘤数据集上,将本文模型与当前广泛使用和认可的分割模型进行了比较,包括基于 CNN 方法的 UNETR、基于 Transformer 方法的 TransBTS 和基于扩散方法的 Diff-UNet。在 BTCV 腹部多器官数据集上,利用相同的训练和测试数据集,将本文模型与最先进的模型进行比较,包括 V-Net, TransUNet 和 Diff-UNet。本实验中的所有基线模型和 FDiff-Fusion 都是在相同数据集和相同计算机硬件下进行训练的,并且最好的训练模型由验证集选择。所有的定量和定性结果都是模型的直接输出,没有

经过任何后期处理过程。

表 1 列出了本文所提出的方法和 SOTA 方法在 BRATS 2020 脑肿瘤数据集的 3 个分区(全肿瘤、肿瘤核心、增强肿瘤)上的 Dice 分数和 HD95 距离及其对应的平均值。显然,本文方法不仅在 Dice 平均分数和 HD95 平均距离方面明显优于 SOTA 方法,而且在 3 个分区上的分数也比最先进的方法更高。3 个分区的平均 Dice 分数达到 84.16%,比第二名提高了 1.02%,比 UNETR 提高了 3.8%。同时,在 HD95 距离上,本文模型达到了 2.473 mm,排名

第一,明显优于最先进的方法。

表 2 中列出了 8 个腹部器官的 Dice 分数以及平均 Dice 分数和平均 HD95 距离,以比较 FDiff-Fusion 和最先进的分割方法。从表 2 看出,FDiff-Fusion 在 8 个器官上具有最佳的平均 Dice 分数和 HD95 距离。本文方法在 6 个器官上的 Dice 分数排名第一,平均 Dice 分数和 HD95 距离达到了 83.82%和 7.98mm,这表明本文方法在 BTCV 腹部多器官数据集上的分割性能比最先进的方法更具优越性。虽然 FDiff-Fusion 在胆囊和胰腺上的 Dice 分数排名第二,分别为 74.01%和 70.98%,但都只是略小于最佳分数(胆囊的 75.32%和胰腺的 72.16%)。

图 7 显示了在 BRATS 2020 脑肿瘤数据集上 FDiff-Fu-

sion 与几个基线模型的定性比较结果。其中,使用 T2 作为分割背景,GT 表示地面真实标签。在第一行中,与其他基线模型(由虚线方框标记)相比,本文提出的模型成功抑制了误报结果。在第二行中,与其他 SOTA 模型相比,FDiff-Fusion 生成了更精确的分割结果,噪声更少,分割边界更接近于地面真实标签,这得益于模糊学习模块的设计。在第三行中,本文模型实现了更准确和更完整的分割边界,这得益于本文提出的迭代注意力融合机制。实验表明,FDiff-Fusion 模型在分割边界和微小分割目标上实现了更精确的分割结果,而基线模型会丢失一些分割目标区域,或者在分割结果中包含其他非分割目标区域;并且对于目标区域边界的精确识别和分割,本文提出的模型也展现了明显的优越性。

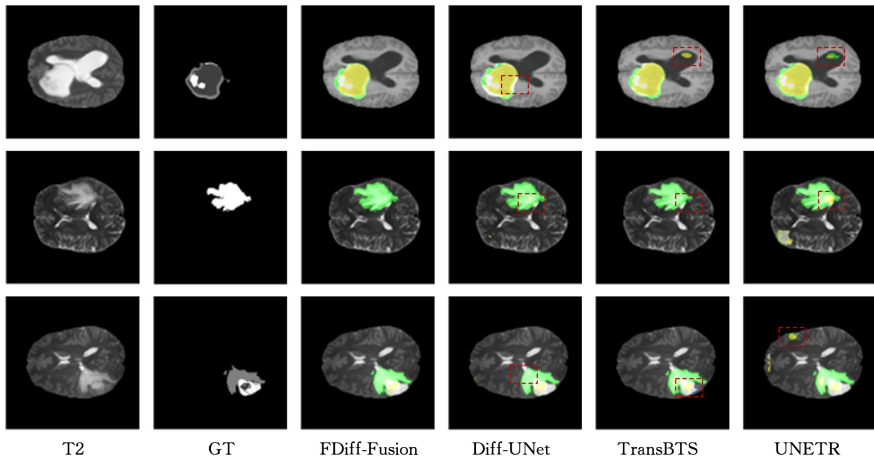


图 7 BRATS 2020 脑肿瘤数据集上 SOTA 方法定量比较的可视化

Fig. 7 Visualization of quantitative comparison of SOTA methods on BRATS 2020 brain tumor dataset

为了进一步展现本文模型对于医学图像的分割性能,在 BTCV 腹部多器官数据集上进行了分割实验,并选择了最终分割结果图的一个切片进行可视化。

从图 8 中可以清楚地看出,FDiff-Fusion 在 BTCV 腹部多器官数据集上表现良好,与基线模型相比,其分割结果

更加精准,与地面真值标签更加一致。这种优势来自于模糊学习模块和迭代注意力融合模块,在输入图像中,它能够捕获关键的形态和结构信息,并将这些信息融合到分割结果中,从而获得更精细的分割边界和更精确的器官区域。

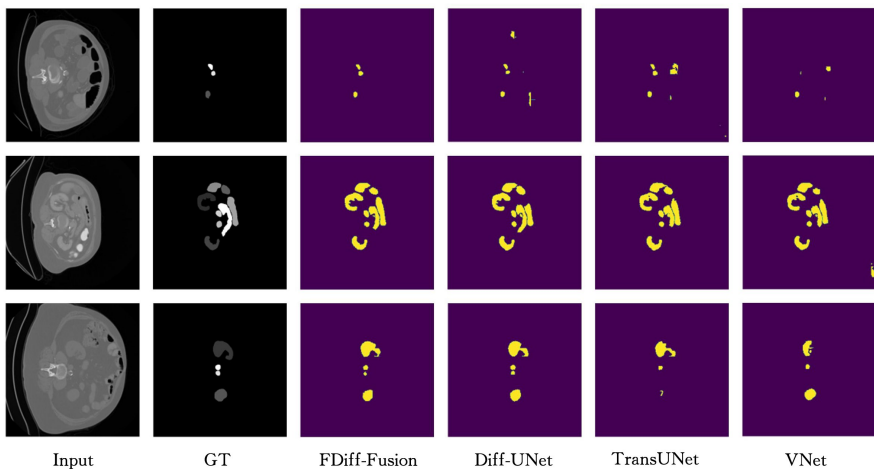
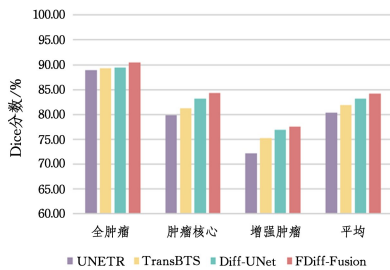


图 8 BTCV 腹部多器官数据集上 SOTA 方法定量比较的可视化

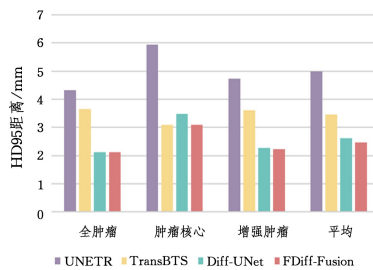
Fig. 8 Visualization of quantitative comparison of SOTA methods on BTCV abdominal multi-organ dataset

为了更加清晰地展示 FDiff-Fusion 在 3D 医学图像分割方面的优越性,本文采用柱形图对模型进行评估。如图 9 所示,其中图 9(a)和图 9(b)分别表示 BRATS 2020 脑肿瘤数据

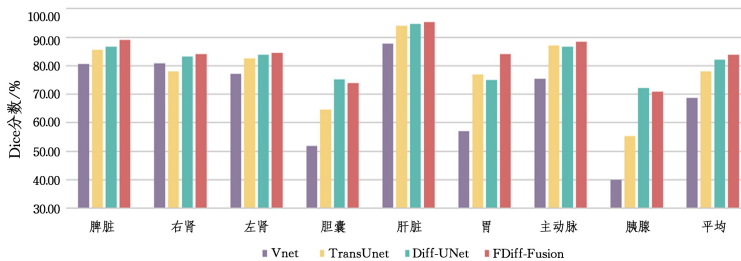
集的 Dice 分数和 HD95 距离结果柱形图,图 9(c)表示 BTCV 腹部多器官数据集的 Dice 分数结果柱形图。根据柱形图中柱子的高低,可以看出 FDiff-Fusion 模型具有更好的分割效果。



(a)BRATS 2020 脑肿瘤数据集 Dice 分数结果柱状图



(b)BRATS 2020 脑肿瘤数据集 HD95 距离结果柱状图



(c)BTCV 腹部多器官数据集 Dice 分数结果柱状图

图 9 不同模型在不同数据集上的分割结果柱形图对比

Fig. 9 Histogram comparison of segmentation results with different models on different datasets

### 4.5 消融实验

本节进行了全面的消融研究,以验证所提出的模糊学习模块和迭代注意力融合的有效性。实验结果如表 3 所列,其中 FLM 表示模糊学习模块,AF 表示注意力融合,IAF 表示迭代注意力融合。可以发现,basic+FLM 在全肿瘤(WT)、肿瘤核心(TC)和增强肿瘤上(ET)的平均 Dice 分数比 basic 的平均 Dice 分数高 0.65%,这说明本文设计的 FLM 能够有效地处理医学图像中的分割目标边界不确定和区域模糊化的现象,并有效提高模型的分割精度。此外,basic+FLM+AF 的 Dice 分数相比 basic+FLM 有明显提高,并且 basic+FLM+IAF 的 Dice 分数比 basic+FLM+AF 更高,说明本文引入的注意力融合机制对于提高模型的分割性能起到了明显的效果。basic+FLM+IAF 的平均 Dice 分数比 basic 高 1.24%,表明本文提出的模糊学习模块和迭代注意力机制对于医学图像分割具有更高的优越性。

表 3 BraTS 2020 数据集上不同模块的消融研究

Table 3 Ablation for different modules on BraTS 2020 dataset (%)

方法	WT	TC	ET	平均
basic	89.24	82.98	76.54	82.92
basic+FLM	90.12	83.64	76.98	83.57
basic+FLM+AF	90.42	83.88	77.46	83.92
basic+FLM+IAF	<b>90.51</b>	<b>84.37</b>	<b>77.61</b>	<b>84.16</b>

**结束语** 本文提出一种基于模糊逻辑驱动的医学图像扩散融合网络分割模型 FDiff-Fusion。为了解决医学图像普遍存在的分割边界不确定和区域模糊化现象,本文提出了两种新颖的技术来保证模型的性能,即模糊学习和迭代注意力融合。通过在 U-Net 模型中添加模糊学习理论,增强了 FDiff-Fusion 对模糊信息进行建模的能力。同时,对于去噪过程的每个时间步所预测的分割图,本文设计了一个基于迭代注意力特征融合的方法来融合每个时间步的预测结果,提高了模型分割的准确性和鲁棒性。FDiff-Fusion 在两个不同基准数

据集上的实验结果证明了本文模型相对于最先进模型的优越性,为医学图像分割领域提供了一种有效的解决方案。

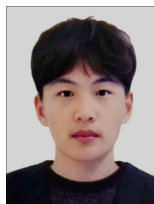
FDiff-Fusion 尽管在医学图像分割领域中取得了显著的成效,但仍存在许多方向可以进一步探索和完善。未来,将进一步优化 FDiff-Fusion 的模型结构和参数,以提高模型的泛化和迁移的能力,使其更好地适用于不同类型的医学图像。

### 参考文献

- [1] ULLAH Z, USMAN M, JEON M, et al. Cascade multiscale residual attention cnns with adaptive roi for automatic brain tumor segmentation[J]. Information Sciences, 2022, 608: 1541-1556.
- [2] TANG Y, YANG D, LI W, et al. Self-supervised pre-training of swin transformers for 3d medical image analysis[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 20730-20740.
- [3] HE Y, GE R, QI X, et al. Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(2): 2588-2601.
- [4] CHEN L, BENTLEY P, MORI K, et al. DRINet for medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2018, 37(11): 2453-2462.
- [5] LIU F, ZHANG Z, ZHOU R. Automatic modulation recognition based on CNN and GRU[J]. Tsinghua Science and Technology, 2021, 27(2): 422-431.
- [6] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: An end-to-end framework for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 15650-15664.
- [7] ZHANG Z M, GUO Y, MA C X, et al. GT-4S: Graph Transformer for Scene Sketch Semantic Segmentation[J]. Journal of Software, 2025, 36(3): 1375-1389.
- [8] ZHANG C, JIANG W, ZHANG Y, et al. Transformer and CNN

- hybrid deep neural network for semantic segmentation of very-high-resolution remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-20.
- [9] XIE Y, ZHANG J, SHEN C, et al. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation[C]// *Medical Image Computing and Computer Assisted Intervention — MICCAI 2021: 24th International Conference, Strasbourg, France, September 27-October 1, 2021, Proceedings, Part III 24*. Springer International Publishing, 2021: 171-180.
- [10] OLAF R, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015: 234-241.
- [11] ANDRIY M. 3D MRI brain tumor segmentation using autoencoder regularization[C]// *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries; 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*. Springer International Publishing, 2019: 311-320.
- [12] KINGMA D, WELLING M. Auto-encoding variational bayes [J]. *arXiv:1312.6114*, 2013.
- [13] WANG W, CHEN C, DING M, et al. Transbts: Multimodal brain tumor segmentation using transformer[C]// *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021: 109-119.
- [14] HATAMIZADEH A, TANG Y, NATH V, et al. Unetr: Transformers for 3d medical image segmentation[C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022: 574-584.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale[J]. *arXiv:2010.11929*, 2020.
- [16] HATAMIZADEH A, NATH V, TANG Y, et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images[C]// *International MICCAI Brainlesion Workshop*. Cham: Springer International Publishing, 2021: 272-284.
- [17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012-10022.
- [18] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation [J]. *arXiv: 2102.04306*, 2021.
- [19] HUANG D M, DAI L, WEI L F et al. A secure outsourcing fusion denoising scheme for multi-frame remote sensing images [J]. *Journal of Computer Research and Development*, 2017, 54(10): 2378-2389.
- [20] GOYAL B, DOGRA A, AGRAWAL S, et al. Image denoising review: From classical to state-of-the-art approaches[J]. *Information Fusion*, 2020, 55: 220-244.
- [21] CROITORU F, HONDRU V, IONESCU R, et al. Diffusion models in vision: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 10850-10869.
- [22] YANG L, ZHANG Z, SONG Y, et al. Diffusion models: A comprehensive survey of methods and applications[J]. *ACM Computing Surveys*, 2023, 56(4): 1-39.
- [23] WU J, FU R, FANG H, et al. Medsegdiff: Medical image segmentation with diffusion probabilistic model[C]// *Medical Imaging with Deep Learning*. PMLR, 2024: 1623-1639.
- [24] WU J, JI W, FU H, et al. Medsegdiff-v2: Diffusion based medical image segmentation with transformer [J]. *arXiv: 2301.11798*, 2023.
- [25] XING Z, WAN L, FU H, et al. Diff-unet: A diffusion embedded network for volumetric segmentation [J]. *arXiv: 2303.10326*, 2023.
- [26] WOLLEB J, ROBIN S, BIEDER F, et al. Diffusion models for implicit image segmentation ensembles[C]// *International Conference on Medical Imaging with Deep Learning*. PMLR, 2022: 1336-1348.
- [27] ZHOU T Y, DING W P, HUANG J S, et al. Fuzzy Logic Guided Deep Neural Network with Multi-granularity [J]. *Pattern Recognition and Artificial Intelligence*, 2023, 36(9): 778-792.
- [28] KUMAR D, AGRAWAL R K, KUMAR P. Bias-corrected intuitionistic fuzzy c-means with spatial neighborhood information approach for human brain MRI image segmentation[J]. *IEEE Transactions on Fuzzy Systems*, 2020, 30(3): 687-700.
- [29] 李季, 胡锦涛, 乔敏, 王艳. 一种针对脑部图像分割强度不均匀性的改进方法[J]. *重庆工商大学学报(自然科学版)*, 2023, 40(1): 34-39.
- [30] YANG L, WANG S, LIEW A. Fine-Grained Lip Image Segmentation using Fuzzy Logic and Graph Reasoning[J]. *IEEE Transactions on Fuzzy Systems*, 2024, 32(2): 349-359.
- [31] ZHOU M, SHANG C, LI G, et al. Transformation-based fuzzy rule interpolation with Mahalanobis distance measures supported by Choquet integral[J]. *IEEE Transactions on Fuzzy Systems*, 2022, 31(4): 1083-1097.
- [32] SONG S, JIA Z, YANG J, et al. Image segmentation based on fuzzy low-rank structural clustering[J]. *IEEE Transactions on Fuzzy Systems*, 2023, 31(7): 2153-2166.
- [33] LIU Z, TAN Y, HE Q, et al. SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(7): 4486-4497.
- [34] PHAM T. The Kolmogorov-Sinai entropy in the setting of fuzzy sets for image texture analysis and classification [J]. *Pattern Recognition*, 2016, 53: 229-237.
- [35] LI D, ZHANG H, LI T, et al. Hybrid missing value imputation algorithms using fuzzy c-means and vaguely quantified rough set [J]. *IEEE Transactions on Fuzzy Systems*, 2021, 30(5): 1396-1408.
- [36] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 6840-6851.
- [37] NICHOL A, DHARIWAL P. Improved denoising diffusion

- probabilistic models[C]// International Conference on Machine Learning. PMLR, 2021:8162-8171.
- [38] GAO Q, LI Z, ZHANG J, et al. CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization[J]. arXiv:2304.01814, 2023.
- [39] YUE J, FANG L, XIA S, et al. Dif-fusion: Towards high color fidelity in infrared and visible image fusion with diffusion models [J]. IEEE Transactions on Image Processing, 2023, 32: 5705-5720.
- [40] SONG J, MENG C, ERMON S. Denoising diffusion implicit models[J]. arXiv:2010.02502, 2020.
- [41] WANG Y, LIU H, FENG Y, et al. HeadDiff: Exploring Rotation Uncertainty with Diffusion Models for Head Pose Estimation [J]. IEEE Transactions on Image Processing, 2024, 33: 1868-1882.
- [42] GONG K, JOHNSON K, EL FAKHRI G, et al. PET image denoising based on denoising diffusion probabilistic model[J]. European Journal of Nuclear Medicine and Molecular Imaging, 2024, 51(2):358-368.
- [43] DING Y, YU X, YANG Y, et al. RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:3975-3984.
- [44] ULLAH Z, USMAN M, JEON M, et al. Cascade multiscale residual attention cnns with adaptive roi for automatic brain tumor segmentation[J]. Information sciences, 2022, 608:1541-1556.
- [45] FANG X, YAN P. Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction [J]. IEEE Transactions on Medical Imaging, 2020, 39(11):3619-3629.



**GENG Sheng**, born in 2001, postgraduate. His main research interests include granular computing, fuzzy theory and deep learning.



**DING Weiping**, born in 1979, Ph.D, professor, Ph.D supervisor. His main research interests include data mining, machine learning, granular computing, evolutionary computing, and big data analytics.

(责任编辑:柯颖)