

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-13

论文引用格式: Wang Zhengyu, Song Jie, Wang Cailing. Pathology image nuclear segmentation with dual conditional diffusion model[J/OL]. Journal of Image and Graphics, XXXX: 1-13. DOI: 10.11834/jig.260234. (王征宇, 宋杰, 王彩玲. 双重条件扩散模型的病理图像细胞核分割[J/OL]. 中国图象图形学报, XXXX: 1-13. DOI: 10.11834/jig.260234. ) [DOI: 10.11834/jig.260234]

## 双重条件扩散模型的病理图像细胞核分割

王征宇, 宋杰, 王彩玲

南京邮电大学自动化学院, 南京 210023

**摘要:** 目的 数字病理图像细胞核精准分割是病理量化分析的关键问题。数据驱动的深度分割方法需较大规模训练数据集, 虽能较好处理语义推理, 但受逐像素分类影响, 对重叠、成团细胞核的分割能力有限。深度扩散生成模型通过迭代去噪可隐式建模细胞核空间几何约束, 提供了一种有别于传统模型的新型学习机制。本文提出一种融合判别式分割和生成式扩散的双重条件扩散模型, 用于病理图像细胞核分割。方法 首先, 基于去噪扩散概率模型构建双重条件扩散网络, 将分割任务定义为条件掩码生成问题, 通过迭代去噪预测精细化结果; 其次, 借助注意力机制设计空间与语义互补先验, 将其嵌入扩散时间步中, 引导模型聚焦细胞核个体区域; 此外, 引入随机潜在嵌入, 显著减少训练和采样时间步数。结果 在乳腺、结肠、肾脏、肺、前列腺及胃等多组病理图像上的实验表明, 本文模型在平均交并比等指标上均优于现有数据驱动和扩散生成深度分割方法, 分别超出次优模型 1.03 个百分点和最新模型 2.17 个百分点。消融实验验证了各模块的有效性。通过耦合判别式分割与生成式学习, 本文模型超出基线扩散模型 3.68 个百分点。结论 本文提出的双重条件扩散模型通过融合判别式分割与生成式扩散形成互补机制, 在复杂重叠细胞核结构的分割任务上取得较好效果, 为高精度细胞核分割提供了新思路。

**关键词:** 数字病理; 细胞核分割; 去噪扩散概率模型; Transformer; 潜在嵌入

## Pathology image nuclear segmentation with dual conditional diffusion model

Wang Zhengyu, Song Jie, Wang Cailing

College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

**Abstract: Objective** Accurate nuclear segmentation in digital pathology images is a fundamental yet challenging task in quantitative pathological analysis, playing a critical role in cancer diagnosis, grading, prognosis assessment, and treatment evaluation. Manual annotation of nuclei is labor-intensive, time-consuming, and suffers from poor reproducibility, necessitating automated segmentation methods. Data-driven deep learning approaches, particularly those based on convolutional neural networks and Transformer architectures, have achieved considerable success in this domain. These methods typically require large-scale annotated training datasets and can effectively handle semantic reasoning of nuclei. However, due to their inherent reliance on pixel-wise classification, they often struggle to accurately segment individual nuclei in challenging scenarios such as overlapping, clustered, or tightly packed structures, where spatial geometric information is crucial. For example, in histopathology images of breast or prostate tissues, nuclei frequently form dense clusters with ambiguous boundaries, making individual instance separation particularly difficult. In contrast, deep diffusion generative models offer a fundamentally different learning paradigm. Through an iterative forward noising and reverse denoising pro-

收稿日期: 2026-04-22; 修回日期: 2026-05-26

\* 通信作者: 宋杰 j.song@njupt.edu.cn

基金项目: 国家自然科学基金项目(项目编号: 62001247); 江苏省前沿技术研发计划(项目编号: BF2024070)

Supported by: National Natural Science Foundation of China(62001247); Frontier Technologies R&D Program of Jiangsu(BF2024070)

©中国图象图形学报版权所有

cess, these models can implicitly capture the underlying spatial and geometric constraints of nuclear structures without requiring explicit shape priors. This capability makes them particularly promising for medical image segmentation tasks where structural integrity and boundary delineation are paramount. Nevertheless, directly applying diffusion models to nucleus segmentation faces two major challenges: the difficulty of incorporating discriminative priors to guide the generation process, and the high computational overhead associated with numerous denoising steps. To address these issues, this paper proposes a novel dual-condition diffusion model that synergistically integrates discriminative segmentation and generative diffusion for high-precision pathological image nucleus segmentation. **Method** The proposed framework consists of three collaboratively designed components. First, based on the denoising diffusion probabilistic model, the framework constructs a dual-condition diffusion network that reformulates the nucleus segmentation task as a conditional mask generation problem. Unlike standard diffusion models that predict noise at each timestep, this network directly predicts refined segmentation masks, enabling more efficient and stable training. Specifically, the network takes the initial coarse mask produced by the discriminative segmenter as input and iteratively refines it through a series of denoising steps. Second, to address the limitation that diffusion models lack explicit spatial awareness, the model leverages attention mechanisms to design spatial and semantic complementary priors. These priors capture both local texture information and global contextual relationships of nuclear structures. The spatial prior focuses on edge and boundary information, while the semantic prior encodes category-level consistency across the entire nucleus region. They are embedded into each time step of the diffusion learning process, serving as conditional guidance that directs the diffusion process to focus on individual nucleus regions rather than being distracted by background clutter or ambiguous boundaries. This design ensures that the generative refinement remains grounded in meaningful anatomical constraints throughout all denoising steps. Third, recognizing the computational inefficiency of traditional diffusion models, the model introduces a random latent embedding strategy. Specifically, Gaussian-sampled latent embeddings are injected into each layer of the diffusion refiner, enabling the model to learn denoising mappings with larger step sizes. This strategy significantly reduces both the number of training timesteps and the number of sampling timesteps required during inference, achieving a favorable trade-off between segmentation accuracy and computational efficiency without compromising generation quality. The three components are trained jointly in an end-to-end manner, allowing the discriminative segmenter to provide high-quality spatial initializations and the diffusion refiner to leverage complementary priors for structural optimization. **Result** Extensive experiments are conducted on publicly available multi-organ and multi-disease histopathology datasets, including breast, colon, kidney, lung, prostate, and stomach tissues, all with pixel-level nucleus annotations. The proposed model is compared against a comprehensive set of state-of-the-art methods, including approaches based on convolutional neural networks, Transformer-based architectures, and recent diffusion-based segmentation models. Quantitative evaluation employs two widely accepted metrics: the Dice similarity coefficient and the mean intersection over union. Experimental results consistently demonstrate that the proposed model outperforms all existing data-driven and diffusion-based deep segmentation methods across multiple organs, outperforming the suboptimal model by 1.03 percentage points and the latest model by 2.17 percentage points in terms of mean Intersection over Union. Consistent improvements are observed on all tested datasets, confirming the robustness and generalizability of the approach across different tissue types. Ablation studies are conducted to systematically validate the contribution of each core module. Removing the discriminative segmenter, the diffusion refiner, or the complementary prior embedding each leads to noticeable performance degradation, confirming that all three components play essential and non-redundant roles. Through coupling discriminative segmentation with generative learning, integrating discriminative initialization, generative refinement, and spatial-semantic guidance, the model achieves a substantial improvement of 3.68 percentage points in mean Intersection over Union on multi-organ image data compared to the baseline. These gains are both statistically significant and practically meaningful for clinical applications. Qualitative results further illustrate that the model produces segmentation masks with sharper boundaries, fewer isolated false positives, and better separation of touching and overlapping nuclei compared to competing methods. **Conclusion** This paper addresses the challenging problem of segmenting complex overlapping and clustered nucleus structures in digital pathology images by proposing a dual-condition diffusion model that integrates discriminative segmentation and generative diffusion in a complementary manner. The framework introduces three key innovations: a dual-condition diffusion network formulating segmentation as conditional mask

generation, spatial and semantic complementary priors embedded via attention mechanisms to guide the diffusion process, and a random latent embedding strategy that reduces computational overhead. Extensive experimental results demonstrate that the proposed method consistently outperforms existing state-of-the-art approaches. Ablation studies confirm the necessity and synergy of the three core components. The model achieves a favorable balance between segmentation accuracy and computational efficiency, making it suitable for potential clinical deployment. While the current work focuses on nucleus segmentation, the proposed paradigm may also be applicable to other medical image segmentation tasks requiring precise structural delineation under challenging conditions. Despite the promising results, the proposed method has one limitation. The iterative denoising process incurs higher computational cost compared to single-stage discriminative methods. Future work will focus on model compression techniques, such as knowledge distillation and pruning, as well as more efficient sampling strategies to reduce inference time while maintaining segmentation accuracy.

**Key words:** Digital Pathology; Nucleus Segmentation; Denoising Diffusion Probabilistic Model; Transformer; Latent Embedding

论文引用格式:[DOI: 10. 11834/jig. 260234]

## 0 引言

数字病理图像细胞核精准分割在癌症分级(Qiu 等, 2021)、有丝分裂检测(Cireşan 等, 2013)和病理标本感兴趣分子量化(Zhao 等, 2025)等方面扮演着重要角色。手动注释细胞核既费力又耗时,且可追溯性差。由于细胞核在形状、重叠程度和染色质方面存在差异,精确提取细胞核完整结构具有挑战性。深度学习可分层抽取原始图像的抽象特征,为解决病理图像细胞核精准分割提供了有效途径(Shi 等, 2025)。

现有基于 U-Net 架构的判别式分割网络(Zhou 等, 2021),主要包括卷积神经网络(convolutional neural network, CNN)的变种(Song 等, 2023; Zhang 等, 2024)和 Transformer 嵌入(Valanarasu 等, 2021; He 等, 2023),并采用交叉熵损失或 Dice 损失来学习输入图像到分割掩码的超分辨映射。尽管取得了巨大成功,但这些方法只专注在像素特征空间中学习细胞间的决策边界,并未捕捉细胞底层的空间几何结构信息(Kazerouni 等, 2023),因此无法学习到内在的结构特征,导致对模糊边界和重叠遮挡的处理能力受限。

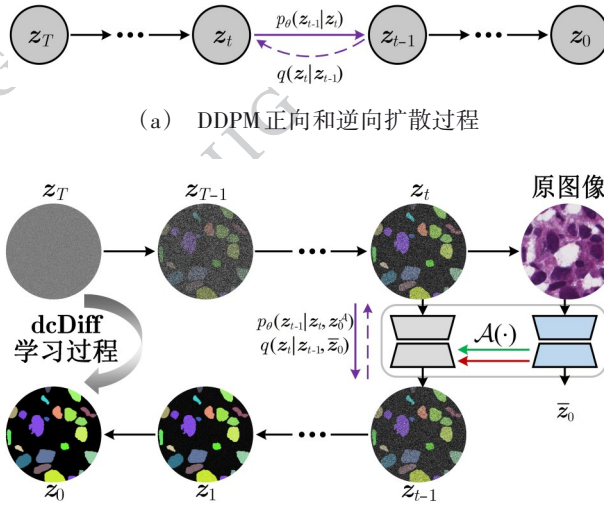
作为主流生成式学习方法,去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)具备捕捉复杂数据分布的能力,利用迭代去噪机制,实现从随机采样的高斯噪声到目标图像或潜在分布的高质量映射(Croitoru 等, 2023)。与判别式分割模型相比,生成式扩散分割模型通过学习分割掩码的分布

或条件分布,生成分割结果,且由于采样的多样性,能够衡量生成的分割结果的不确定性。最近,相关研究表明,扩散模型在语义分割和实例分割等任务中,性能较之数据驱动的深度分割方法具有很强的竞争力(Kazerouni 等, 2023),并且由于分割掩码本质上是由细胞语义决定的,因此,将判别特征融入到扩散学习过程中,可提升其分割精度和泛化能力。如图 1 所示。

基于此,本文探索如何高效融合判别式分割和生成式学习的互补优势,提出了一种新颖的双重条件扩散模型(dual conditional diffusion model, dcDiff)。本文模型将细胞核分割任务定义为 DDPM 条件掩码生成问题,利用其迭代去噪过程生成掩码预测并减少误差。此外,针对扩散生成模型逐像素概率建模的局限性,该模型通过嵌入自注意力和交叉注意力机制融合多尺度信息,将空间和语义互补先验集成到每个扩散时间步,以增强其细胞核完整结构建模能力,并通过在模型中引入随机潜在嵌入,减少训练和采样过程的扩散时间步数。得益于其计算可行性,所提出的 dcDiff 为细胞核分割的扩散生成提供了新的范式。通过乳腺、胃、前列腺等多组病理图像的分割实验表明:该模型在多个指标上均优于目前数据驱动和扩散生成深度模型,并在相同的模型迭代次数下能够取得最佳性能。

本文的主要贡献如下:

- 1) 提出了双重条件扩散模型 dcDiff。该模型融合了判别式分割和生成式学习的互补优势,形成了一种替代传统深度分割模型的新型学习机制;
- 2) 在 dcDiff 中嵌入双重 Transformer 耦合编码模块,用于学习空间和语义互补先验,并结合扩散生成



(b) 本文提出 dcDiff 模型, 将细胞核分割定义为 DDPM 条件掩码生成问题

图1 dcDiff 分割模型示例

Fig. 1 Illustration of the dcDiff diffusion process for segmentation ((a) Forward and reverse diffusion processes of DDPM; (b) dcDiff formulates the nucleus segmentation task as a DDPM conditional mask generation problem)

过程, 增强了模型对细胞核完整结构的建模能力;

3) 在 dcDiff 的扩散层中引入了随机潜在嵌入机制。该机制通过向各层注入随机采样向量, 使模型能够采用大步长去噪策略, 从而减少训练和采样过程中的时间步数。

## 1 相关工作

### 1.1 基于数据驱动的深度神经网络细胞核分割

数据驱动深度分割方法主要包括卷积编解码 (Song 等, 2023; Zhang 等, 2024; Ronneberger 等, 2015) 和 Transformer 嵌入 (Valanarasu 等, 2021; He 等, 2023; Cao 等, 2022; Zeng 等, 2023)。前者通过下采样编码器提取多尺度语义特征, 并利用上采样解码器逐级恢复空间细节, 最终生成像素级分割掩码; 而后者利用自注意力机制对图像块序列进行全局上下文建模, 通过捕获像素间的长程依赖关系来增强分割特征的代表能力。受此启发, 文献 (Valanarasu 等, 2021) 提出了医学 Transformer 模型 (medical Transformer, MedT), 该模型通过引入门控轴向注意力机制扩展了基于注意力机制的分割架构 (Wang 等, 2020), 并为后续设计特定于医学分割领域的 Swin-UNet (Cao 等, 2022) 结构提供可能。文献 (He 等, 2023) 给出了一个 Swin Transformer 驱动的分割

框架 TransNuSeg, 进一步验证了 Transformer 在组织病理细胞核分割中的适用性和潜力。文献 (Zhang 等, 2023) 给出了一个集成了 Swin Transformer 和卷积神经网络的分组注意力架构, 有效融合了图像的全局信息与局部信息, 提高了医学图像分割结果的精确度。

### 1.2 基于深度扩散模型的语义分割

随着扩散生成模型的兴起, 研究者提出了一系列深度扩散分割模型。一般而言, 对于基于 DDPM 的分割方法, 扩散生成通过两个参数化的马尔可夫链, 在  $T$  个时间步内迭代地向输入图像  $z_0$  添加高斯噪声  $z_T \sim \mathcal{N}(0, I)$ , 然后学习逆向扩散过程恢复数据分布。前向过程  $q(\cdot)$  由联合分布给出:

$$q(z_{1:T}|z_0) = \prod_{t=1}^T q(z_t|z_{t-1}) \quad (1)$$

式中,  $q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t}z_{t-1}, \beta_t I)$ 。

$\beta_t \in (0, 1)$  表示方差调度 (Ho 等, 2020)。经过连续迭代, 任意时间步  $t$  图像  $z_t$  可通过以下公式获得:

$$q(z_t|z_0) = \mathcal{N}(z_t; \sqrt{\bar{\alpha}_t}z_0, (1-\bar{\alpha}_t)I) \quad (2)$$

式中,  $\alpha_t = 1 - \beta_t$ ,  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 。逆过程  $p_\theta(\cdot)$  的目标是学习从噪声  $z_T$  恢复目标  $z_0$ 。为此, 需要构建一个参数化的神经网络来逼近联合分布:

$$p_\theta(z_{0:T}) = p(z_T) \prod_{t=1}^T p_\theta(z_{t-1}|z_t) \quad (3)$$

式中,  $p_\theta(z_{t-1}|z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t), \Sigma_\theta(z_t, t))$ ,  $\mu_\theta(z_t, t)$ ,  $\Sigma_\theta(z_t, t)$  和  $\theta$  分别表示去噪网络的均值、方差和参数。条件的引入增强了扩散的可控性和可解释性, 例如生成内容和结构完整性 (Rombach 等, 2022)。SegDiff (Amit 等, 2021) 和 cDAL (Hejrati 等, 2024) 使用了 U-Net 架构, 输入图像通过图像编码器作为条件传递。MedSegDiff (Wu 等, 2024) 引入了动态条件编码, 为每个采样步骤建立状态自适应条件。而 GenSelfDiff-HIS (Purma 等, 2025) 采用了一种不同的策略: 在组织病理图像分割的自监督学习框架中, 将以 U-Net 为基础的图像空间生成式扩散过程作为前置任务, 即先利用大量无标注图像训练扩散模型学习图像生成能力, 再将预训练后的 U-Net 迁移到分割任务中进行微调。值得注意的是, 去噪扩散隐式模型 (Song 等, 2020), 通过非马尔可夫推理过程显著加速了采样, 从而减少扩散模型的计算开销。与现有研究不同, 本文提出了一种双重条件扩散网络, 旨在学习一种受空间与语义先验引导的细胞核形态分布模式, 该模式通过嵌入细胞核的空间和语义先验, 能够有效建模细胞核的完整结构, 并结合随机潜在嵌入减少训练和采样过程中的时间步数。

## 2 本文方法

本章首先阐述 dcDiff 的设计动机, 然后在此基

础上对 dcDiff 各个组成部分进行详细描述。

### 2.1 研究动机

为探究不同深度分割模型的机理以及各自优缺点, 进而引出 dcDiff 的设计思路, 本文选取 Kumar 等采集并标注的 TCGA 全切片图像 (whole slide images, WSI) 集 (Kumar 等, 2017) 作为算法比较, 并提出了三个关键发现, 如表 1 所示: 1) 由于计算效率, 数据驱动的深度分割网络也许是可行的, 但其分割精度易受背景复杂性的影响; 2) 直接应用扩散模型, 需要成百上千步的神经网络估计, 导致模型和计算复杂度高; 3) 条件增强的扩散模型有助于实现精度和效率的折中。

因此, 本文提出 dcDiff 模型, 采用 U 型网络 (Ronneberger 等, 2015) 作为判别式分割器  $\varphi(\cdot)$  和扩散优化器  $z_\theta(\cdot)$ , 其中扩散优化器使用分割掩码输出作为输入进行迭代优化, 并以 Transformer 耦合编码嵌入  $\mathcal{A}(\cdot)$  为条件, 引导扩散学习, 建模细胞核完整结构, 如图 2 所示。此外, 在条件扩散优化器实现中, 采用 (Benny 等, 2022) 建模思想学习  $z_0$  预测代替噪声预测, 并引入随机潜在嵌入, 用以提升训练和采样效率。

### 2.2 判别式分割器

假设  $(x, z_0)$  表示训练样本集, 其中  $x \in \mathbb{R}^{H \times W}$  为空间分辨率为  $H \times W$  的病理图像,  $z_0 \in \{0, 1\}^{H \times W}$  是图

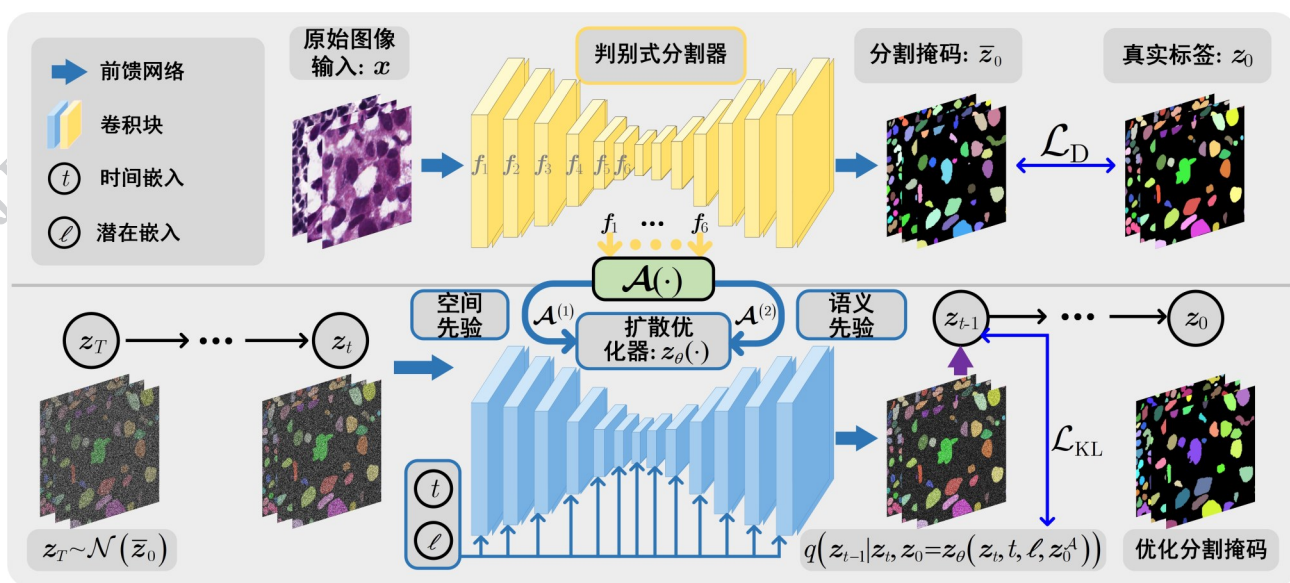


图 2 本文提出的 dcDiff 分割框架

Fig 2 Overview of the proposed dcDiff segmentation framework

像对应的人工标注。预训练的判别式分割器通过提取多尺度特征表示  $\{f^m\}_{m=1}^M$  ( $M=6$ ), 通道数为  $\{64, 128, 128, 256, 256, 256\}$ , 预测初始的粗糙分割概率图为  $\bar{z}_0 \in \mathbb{R}^{h \times w}$ , 其中每个像素值介于 0 到 1 之间, 表示该像素属于细胞核的概率。

### 2.3 基于随机潜在嵌入的扩散优化器

扩散优化器以辅助 Transformer 嵌入  $\mathcal{A}(\{f^m\})$

为条件, 对初始分割掩码  $\bar{z}_0$  进行迭代优化。具体地,

扩散优化器前向扩散过程使用  $\bar{z}_0$  作为输入添加高斯分布噪声, 而逆向学习过程考虑 Transformer 嵌入条件  $\mathcal{A}(\{f^m\})$ , 对采样进行校正。其过程可形式化描述为:

$$z_T \Leftrightarrow \dots \Leftrightarrow z_t \xrightarrow[\mathcal{Q}(z_t|z_{t-1}, \bar{z}_0)]{p_\theta(z_t|z_{t-1}, z_0^A)} z_{t-1} \Leftrightarrow \dots \Leftrightarrow z_0 \quad (4)$$

式中,  $z_1, \dots, z_T$  是与标签  $z_0$  尺寸一致的潜在变量。在前向扩散过程中, 扩散优化器根据预设的噪声调度  $\beta_1, \dots, \beta_T$  逐步向数据中添加高斯噪声。其前向过程  $q(z_{1:T} | z_0, \bar{z}_0)$  定义为:

$$q(z_{1:T} | z_0, \bar{z}_0) = \prod_{t=1}^T q(z_t | z_{t-1}, \bar{z}_0) \quad (5)$$

$$q(z_t | z_{t-1}, \bar{z}_0) = \mathcal{N}(z_t; \sqrt{1 - \beta_t} z_{t-1}, \beta_t \bar{z}_0) \quad (6)$$

式中  $\alpha_t = 1 - \beta_t$ ,  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ , 则在任意时间步  $t$  采样  $z_t$  可通过以下方式得到:

$$q(z_t | z_0, \bar{z}_0) = \mathcal{N}(z_t; \sqrt{\bar{\alpha}_t} z_0, (1 - \bar{\alpha}_t) \bar{z}_0) \quad (7)$$

式中, 该分布的均值参数可定义为随着时间步  $t$  增加的初始分割  $\bar{z}_0$  与真实标签  $z_0$  间的一种加权插值。需要注意的是, 与标准 DDPM 不同, 本文扩散优化器的前向过程并非向真实分割掩码  $z_0$  中添加噪声, 而是以判别式分割器的初始预测  $\bar{z}_0$  作为起点, 通过一个带噪声调度的马尔可夫链, 逐步向其中混入真实标签  $z_0$  的信息, 并同时添加高斯噪声。这一设计的物理意义在于: 在时间步  $t$  较小时,  $z_t$  保留了  $\bar{z}_0$  的大部分结构信息, 引导扩散过程从较优的初始状态开始; 随着  $t$  增大,  $z_t$  逐渐趋向于纯噪声与真实标签的混合, 迫使模型学习从粗糙预测到精细结构的映射。

扩散优化器的逆向过程  $p_\theta(z_{0:T} | \mathcal{A}(\{f^m\}))$  可视为一个马尔可夫链, 初始状态为  $\mathcal{A}(\{f^m\})$  与  $z_0$  的逐元素乘积, 即  $z_0^A = z_0 \odot \mathcal{A}(\{f^m\})$ , 式中  $\odot$  表示哈达玛

乘积。Transformer 嵌入旨在扩散过程中关注细胞核的空间和语义信息, 引导扩散生成更接近于真实标签的分割掩码。具体建模如下:

$$p_\theta(z_{0:T} | z_0^A) = p(z_T | z_0^A) \prod_{t=1}^T p_\theta(z_{t-1} | z_t, z_0^A) \quad (8)$$

$$p_\theta(z_{t-1} | z_t, z_0^A) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t, z_0^A), \Sigma_\theta(z_t, t, z_0^A)) \quad (9)$$

其输出表达为  $z_0 = z_\theta(z_t, t, z_0^A)$ , 最后通过后验分布  $q(z_{t-1} | z_t, z_0 = z_\theta(z_t, t, z_0^A))$  完成对  $z_0$  的采样。

在标准的去噪扩散概率模型中, 逆向过程通常假设去噪分布为单峰高斯分布, 这要求去噪步长足够小以保证生成质量, 从而导致训练与推理过程需要数百甚至上千个时间步。然而, 当去噪步长增大时, 真实去噪分布往往呈现多模态特性, 难以用简单高斯分布建模。

为了解决这一问题, 本文在扩散优化器中引入了随机潜在嵌入策略。具体而言, 就是在扩散优化器的每一层中注入一个服从于标准高斯分布的随机潜在嵌入  $\sim \mathcal{N}(\cdot; 0, I)$ , 使得模型对标签的预测不再是仅依赖于当前有噪标签  $z_t$  和时间嵌入  $t$  的确定性映射, 而是表示为  $z_0 = z_\theta(z_t, t, \cdot, z_0^A)$ 。此时, 模型的:

$$p_\theta(z_{t-1} | z_t) = \int p(\cdot) \cdot q(z_{t-1} | z_t, z_0 = z_\theta(z_t, t, \cdot, z_0^A)) d \quad (10)$$

式中,  $p(\cdot)$  为随机潜在先验分布, 通常取为标准高斯分布。该过程可以通过计算后验  $q(z_{t-1} | z_t, z_0 = z_\theta(z_t, t, \cdot, z_0^A))$  估计  $z_{t-1}$  的分布。

该建模通过引入随机潜在嵌入策略的方式使得逆向去噪分布  $p_\theta(z_{t-1} | z_t)$  不再局限于单峰高斯分布的假设, 而是能够表达复杂的多模态分布。正是这种多模态表达能力, 使得扩散优化器在学习过程中能够采用较大的扩散步长和较大的方差  $\beta_t$  扰动  $T$  个与协同增强, 本文设计了一个新颖的 Transformer 耦合编码嵌入模块  $\mathcal{A}(\cdot)$ 。该模块包含两条学习路径, 即像素路径(标记为“(1)”)和语义路径(标记为“(2)”), 二者通过交叉注意力机制实现信息交互。具体地, 对于给定第  $L$  个模块的输入, 令  $f_i^{(1)}$  和  $f_i^{(2)}$  分别表示像素特征与语义特征。在像素路径中, 使用初始语义特征  $f_i^{(2)}$  作为键(Key)和值(Value)与经过多头自注意力机制更新后的像素查询(Query)特征  $\hat{f}_i^{(1)}$  进行交叉注意力计算, 从而增强

像素特征的语义判别能力得到像素路径输出

$f_{i+1}^{(1)}$ 。计算过程如下:

$$\bar{f}_i^{(1)} = LN(f_i^{(1)}), \bar{f}_i^{(2)} = LN(f_i^{(2)}) \quad (11)$$

$$f'_i{}^{(1)} = MSA(\bar{f}_i^{(1)}, \bar{f}_i^{(1)}, \bar{f}_i^{(1)}) + f_i^{(1)} \quad (12)$$

$$\hat{f}_i^{(1)} = MSA(LN(f'_i{}^{(1)}), \bar{f}_i^{(2)}, \bar{f}_i^{(2)}) + f'_i{}^{(1)} \quad (13)$$

$$f_{i+1}^{(1)} = MLP(LN(\hat{f}_i^{(1)})) + \hat{f}_i^{(1)} \quad (14)$$

式中  $MSA$  表示多头自注意力 (multi-head self-attention) 机制,  $LN$  为层归一化 (layer normalization),  $MLP$  为多层感知器 (multilayer perceptron)。由于像素路径输出  $f_{i+1}^{(1)}$  已经学习到精细的局部细节信息, 因而通过进一步与语义特征交互可提升其语义判别能力。另一方面, 语义路径使用更新后的像素特征  $f_{i+1}^{(1)}$  作为键 (key) 和值 (value) 用于提供额外先验, 与初始语义特征  $f_i^{(2)}$  进行交叉注意力计算, 以达到全局完整性修正的目的。计算过程如下:

$$\bar{f}_i^{(2)} = LN(f_i^{(2)}) \quad (15)$$

$$f'_i{}^{(2)} = MSA(\bar{f}_i^{(2)}, f_{i+1}^{(1)}, f_{i+1}^{(1)}) + f_i^{(2)} \quad (16)$$

$$f_{i+1}^{(2)} = MLP(LN(f'_i{}^{(2)})) + f'_i{}^{(2)} \quad (17)$$

迭代有监督地训练上述过程直到收敛, 使得局部细节与全局语义在网络的每一层相互作用。

值得注意的是,  $f_{i+1}^{(1)}$  和  $f_{i+1}^{(2)}$  将作为增强后的查询特征输入下一模块。由于梯度可以通过两条路径反向传播, 这种耦合设计使得各模块能够同时补偿局部细节和全局语义的信息损失。最终, 通过条件嵌入为扩散学习提供了丰富的空间先验与语义先验, 前提下, 大幅减少训练与采样所需的时间步数。从而

有效提升了分割的精度与鲁棒性。

## 2.5 损失函数设计

本文模型采用端到端的联合训练策略, 总损失函数由判别式分割器的损失和扩散优化器的损失加权组成。

判别式分割器采用交叉熵损失  $\mathcal{L}_{CE}$  与 Dice 损失  $\mathcal{L}_{DC}$  的加权组合。交叉熵损失  $\mathcal{L}_{CE}$  对每个像素独立分类误差提供密集的监督信号; Dice 损失  $\mathcal{L}_{DC}$  直接优化分割区域的重叠程度。两者结合可以兼顾像素级分类精度与区域级分割质量。计算公式如下:

$$\mathcal{L}_D = \mathcal{L}_{CE} + \lambda_{DC} \mathcal{L}_{DC} \quad (18)$$

式中,  $\lambda_{DC}$  用于平衡两类损失对模型优化的影响。

扩散优化器的训练基于 DDPM 负对数似然的变分上界。对于给定样本对  $(x, z_0)$  和第  $t$  步的潜在变量  $z_t$ , 模型采用 KL 散度对扩散优化器进行优化, 用于约束逆向分布逼近真实去噪分布, 并为大步长训练提供稳定监督。计算公式如下:

$$\mathcal{L}_{KL} = KL[q(z_t | z_{t-1}, z_0, \bar{z}_0) \| p_\theta(z_t | z_{t-1}, z_0^*)] \quad (19)$$

式中,  $KL[\cdot]$  表示 KL 散度计算。

## 3 实验结果及分析

### 3.1 数据集

癌症基因组图谱 (the cancer genome atlas, TCGA) 是被广泛用于细胞核分割的病理数据集, 因为它覆盖了多家医院、多位病人、不同器官以及疾病状态信息。为了全面评估所提方法在分割任务上的性能, 本文在多疾病状态和多器官数据集 TCGA 数据集和 TCGA WSI 数据集 (Kumar 等, 2017; Kumar 等, 2019) 上进行了实验。这两个数据集包含来自多个人体器官 (如乳腺、肝脏、前列腺) 的组织病理学图像, 所有这些图像都提供了像素级的细胞核分割标注, 其中, 每张图像大小为  $1000 \times 1000$ , 为确保评估的严谨性, 数据按 8:1:1 的比例划分为训练集、验证集和测试集。同时实验应用了旋转和翻转等几何增强技术, 以增强模型对组织形态变化的鲁棒性。

### 3.2 实现细节

为了便于算法的验证和比较, 本文涉及的所有网络均在统一的框架下训练和测试。实验采用单张 NVIDIA A100 GPU (40 GB 内存) 的服务器和 PyTorch 1.10.0 平台。网络训练周期设置为 250, 批

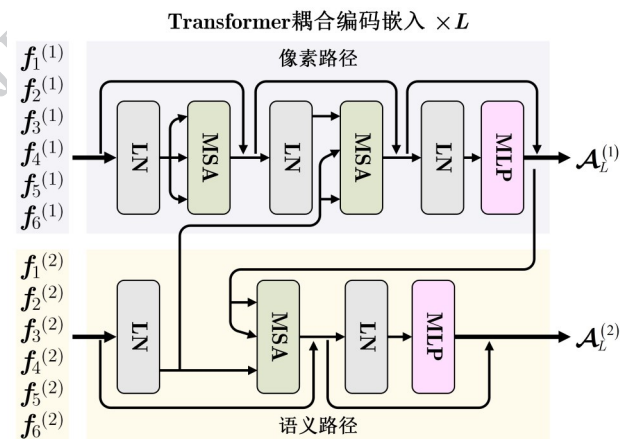


图3 Transformer耦合编码嵌入模块

Fig. 3 The Transformer coupled encoding module

大小设置为2,采用Adam优化器进行优化,初始学习率为 $1 \times 10^{-4}$ ,并结合余弦退火学习率调度器进行训练。调度器在100个训练迭代周期内将学习率从其初始值平滑衰减至最小值 $1 \times 10^{-5}$ 。此外,在实

验中,dcDiff的扩散步数设置为 $T = 4$ ,并采用标准的线性噪声调度和 $\lambda_{Dice} = 1.0$ 以平衡交叉熵损失和Dice损失。判别式分割器在联合训练阶段进行微调,以在保留其特征提取能力的同时更好地适配优化任务。

### 3.3 分割效果比较

本文将dcDiff与现有数据驱动和扩散生成深度分割方法进行基准比较,包括基于CNN的Snake R-CNN (Song 等, 2023)、基于Transformer的MedT (Valanarasu 等, 2021)和TransNuSeg (He 等, 2023),以及四种扩散概率去噪模型的扩展:SegDiff (Amit 等, 2021)、MedSegDiff (Wu 等, 2024)、GenSelfDiff-HIS (Purma 等, 2025)和cDAL (Hejrati

等, 2024)。为确保公平比较,所有方法均在相同的超参数配置和硬件设置下进行训练和评估,以消除外部因素对性能比较的影响。

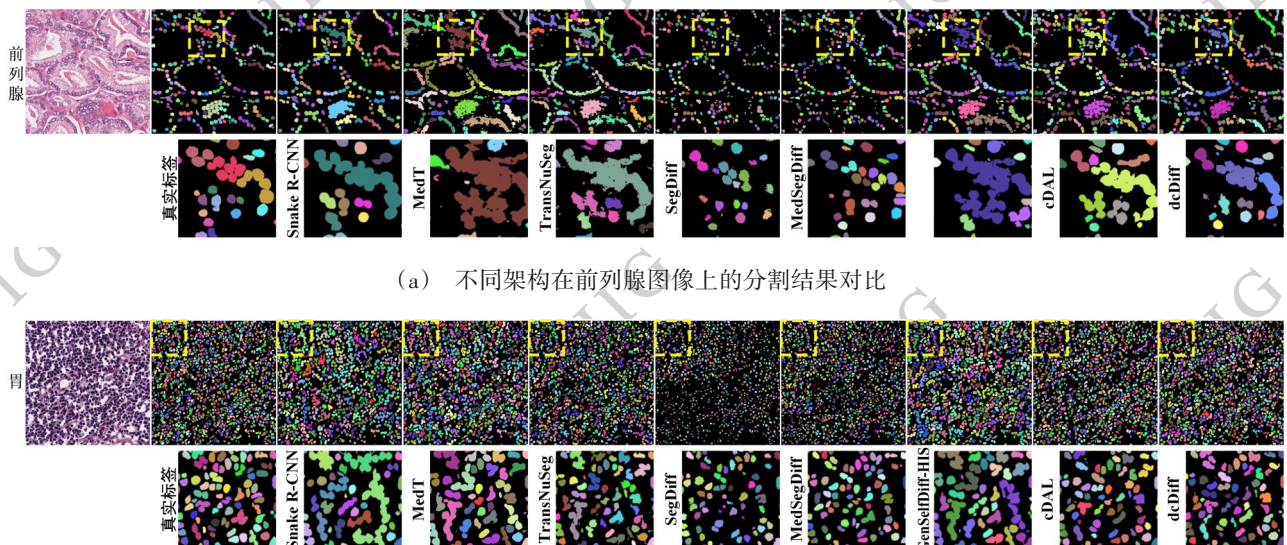
#### 3.3.1 定性分析

如图4和图5所示,dcDiff在多个器官和不同

果,其中放大视图突出了dcDiff的优势,能够有效处理复杂背景、染质稀疏和成团细胞的分割问题。突出其在病理图像多尺度、多样性细胞核分割中的泛化能力。图5展示了所提出方法应用于相同器官疾病状态下的细胞核分割方面优于其他分割方法。图4展示了该方法应用于不同器官图像的可视化结果,其中的放大视图表明dcDiff通过耦合判别式学习与生成式学习精准建模细胞核完整结构。

#### 3.3.2 定量分析

表1展示了各模型在参数量、计算量与推理速度上的对比。与传统的去噪扩散概率模型相比,SegDiff与MedSegDiff由于采用较大的扩散步数,参数量(parameters, Params)均超过了130M,每秒帧数(frames per second, FPS)不足0.02,难以应用于临床实际。dcDiff通过随机潜在嵌入减少了扩散步数,参数量降至40.3M, FPS提升至2.2,同时mIoU为67.4%,仍优于二者。这说明dcDiff的大步长去噪策略能够在保持精度的前提下显著降低计算开销。与轻量级模型Snake R-CNN和自监督扩散模型GenSelfDiff-HIS相比,虽然二者的推理速度



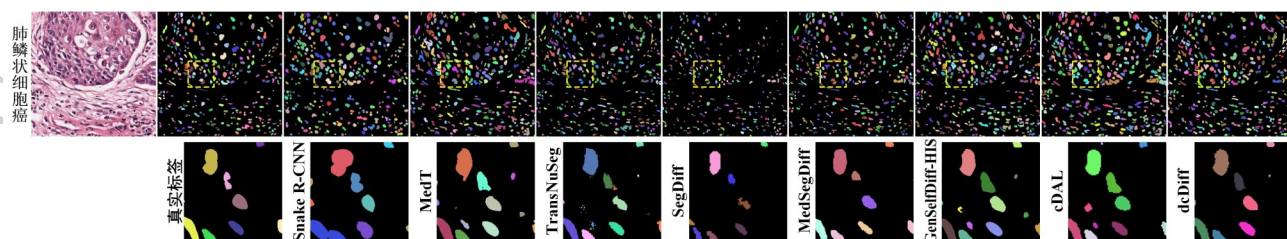
(a) 不同架构在前列腺图像上的分割结果对比

(b) 不同架构在胃部图像上的分割结果对比

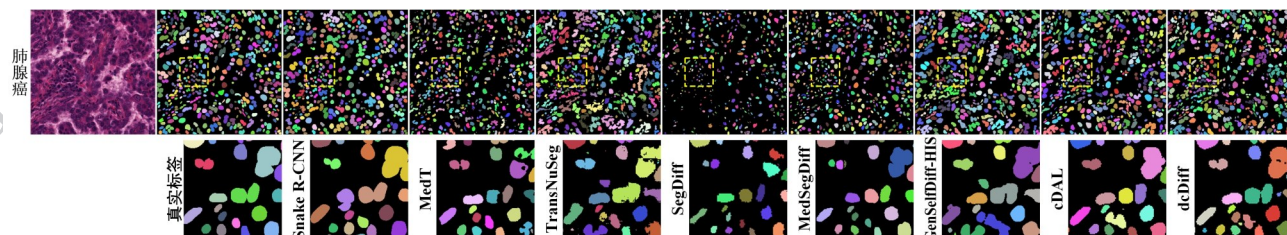
((a)Comparison of segmentation results of different architectures on prostate image; (b)Comparison of segmentation results of different architectures on stomach image)

图4 不同架构在两个不同器官上的分割结果对比

Fig. 4 Comparison of segmentation results of different architectures on two distinct organs



(a) 不同架构在肺鳞状细胞癌图像上的分割结果对比



(b) 不同架构在肺腺癌图像上的分割结果对比

((a)Comparison of segmentation results of different architectures on lung squamous cell carcinoma image ;(b)Comparison of segmentation results of different architectures on lung adenocarcinoma image )

图5 不同方法在同一器官不同疾病状态下的分割结果对比

Fig. 5 Comparison of segmentation results among different methods on the same organ across different disease states

较快,但是其mIoU分别仅65.4%和65.3%。前者受限于单步判别式推理,后者仅在预训练阶段使用扩散,分割时仍为判别式推理。dcDiff将扩散优化器嵌入推理阶段,以适度增加计算量为代价,mIoU提升2.0%,结构分割能力更强。此外,相较于采用单尺度条件嵌入策略的cDAL,dcDiff以相近的

参数量将mIoU提升了2.3%,在分割精度上取得优势,这表明dcDiff在精度与效率之间取得了更优的平衡。因此,dcDiff在医学图像分割任务中较好地实现了性能和效率之间的折中。与现有的深度扩散分割模型相比,dcDiff能够显著缓解计算效率瓶颈。

表1 dcDiff与其他模型的分割性能比较

Table 1 Performance comparison of dcDiff with other segmentation models

模型	参数量 (Params)	浮点运算次数 (FLOPs)	每秒帧数 (FPS)	平均交并比 (mIoU)
Snake R-CNN	1.4M	0.4G	30.3	65.4%
MedT	1.4M	10.6G	1.4	66.2%
TransNuSeg	17.8M	165.9G	30.8	65.2%
SegDiff	132.9M	823315.5G	0.01	65.0%
cDAL	39.8M	324.6G	4.4	65.1%
MedSegDiff	153.8M	289572.5G	0.02	66.2%
GenSelfDiff-HIS	31.3M	84.5G	23.1	65.3%
<b>dcDiff</b>	40.3M	297.1G	2.2	<b>67.4%</b>

注:黑色加粗字体表示最优分割结果。

表2和表3总结了比较方法在多疾病状态和多器官数据集上的定量评估结果。相比于现有深度分

割模型,dcDiff在Dice相似系数(Dice similarity coefficient, Dice)和平均交并比(mean intersection over

union, mIoU)指标上均取得了优越的分割性能。

如表2所示, dcDiff在乳腺、结肠、肾脏、肺、前列腺和胃的分割指标上取得了最优性能,而在膀胱和肝脏分割中, dcDiff也展现出了具有竞争力的结果。相较于其他去噪扩散概率模型, dcDiff以判别式分割器提供的初始分割掩码作为起点,在重叠细胞核场景下更易保持实例独立性。同时,其他去噪扩散概率模型的条件嵌入为单尺度特征,缺乏结构感知。dcDiff的Transformer耦合模块同时提供像素路径与语义路径先验,使得去噪过程能感知细胞核的完整形态。因此dcDiff在肺、胃等器官病理细胞核分割任务中,相较于现有的最优去噪扩散方法, Dice系数分别实现0.86%和0.68%的性能增益。这有力验证了本文方法在不同器官的病理图像细胞核分割任

务中具备的独特优势,证明了双条件引导的有效性。

如表3所示, dcDiff在所有测试场景下均取得了最佳结果:当器官类型相同但疾病状态与训练数据不同(SD)时、当器官类型不同但疾病状态相同(DS)时和当器官类型和疾病状态都不同(DD)时。相比其他的去噪扩散概率模型, dcDiff在不同

分割场景下均取得了最优的性能,其中,得益于dcDiff的判别分割掩码与生成式扩散学习的互补机制, dcDiff在器官类型和疾病状态与训练数据都不同的分割场景下Dice系数仅下降0.40%,远优于其他的去噪扩散概率模型。这有力验证了双重条件增强了本文提出模型的鲁棒性。因此,本文提出的方法展示了整体更优的综合性能。

表2 不同方法在多器官图像上的定量比较结果,分别展示了八个不同器官测试图像的评估指标

Table 2 Quantitative comparison results of different methods on multi-organ images, presenting metrics for test images of eight different organs respectively

模型	器官																全部图像	
	膀胱		乳腺		结肠		肾脏		肝脏		肺		前列腺		胃			
	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)
Snake R-CNN	81.31	70.43	<u>78.35</u>	<u>62.29</u>	76.02	62.29	79.77	66.24	77.02	62.52	79.04	67.01	76.04	64.53	79.76	65.88	78.26	65.35
MedT	81.37	70.10	77.06	63.17	75.51	60.67	78.08	64.23	76.63	62.05	79.89	67.23	74.64	63.82	79.43	71.01	77.65	64.25
TransNuSeg	<b>84.46</b>	<b>73.25</b>	78.24	64.34	77.87	63.76	78.77	65.04	<b>78.43</b>	62.65	79.03	65.41	78.26	64.32	<u>86.57</u>	<u>76.32</u>	79.25	66.33
SegDiff	77.61	63.86	76.87	64.28	71.87	60.74	76.34	63.84	70.64	57.38	73.52	62.31	76.81	62.66	81.88	70.70	77.47	63.00
MedSegDiff	81.88	70.41	77.92	64.70	75.63	61.02	<u>80.04</u>	<u>66.61</u>	72.39	58.24	77.62	63.69	78.05	64.25	83.76	72.08	78.51	66.23
GenSelfDiff-HIS	82.28	70.13	77.81	63.73	78.09	64.58	79.26	65.73	77.07	<b>63.15</b>	<u>80.16</u>	<u>67.41</u>	<u>79.82</u>	<u>66.50</u>	85.87	75.25	77.96	65.25
cDAL	82.06	70.06	78.09	64.49	<u>78.36</u>	<u>64.78</u>	79.59	66.55	75.74	61.18	78.86	65.08	79.20	64.64	85.98	75.97	<u>79.45</u>	<u>66.39</u>
<b>dcDiff</b>	<u>82.98</u>	<u>71.11</u>	<b>78.61</b>	<b>64.84</b>	<b>79.11</b>	<b>65.44</b>	<b>80.21</b>	<b>67.03</b>	<u>77.22</u>	<u>62.98</u>	<b>81.02</b>	<b>68.15</b>	<b>80.37</b>	<b>67.24</b>	<b>86.67</b>	<b>76.48</b>	<b>80.45</b>	<b>67.42</b>

注:黑色加粗字体表示最优结果,有下划线“\_”的数字表示次优结果。

### 3.4 消融实验

dcDiff包含三个关键组件:1)判别式分割器 $\phi$ ; 2)扩散优化器 $z_0$ ; 3)Transformer耦合编码嵌入 $A$ 。为定量评估各组件对dcDiff分割性能的贡献,具体而言,消融实验分别移除或保留这三个关键模块,以观察各组件的独立作用与协同效应。消融实验的结果如表4所示,其中“ $\phi$ ”表示使用了该模块,“ $\bar{\phi}$ ”表示未使用该模块。由表4可得出以下结论:1)单独移除扩散优化器或判别式分割器均会导致分割性能明显下降,这是因为移除前者使模型失

去生成式精化能力;移除后者则使模型缺乏可靠的空间先验,初始掩码质量不足;2)移除Transformer耦合编码嵌入模块也会导致分割性能变差,是因为扩散过程无法获得充分的语义引导,空间与语义信息的缺失导致分割结果边界模糊、结构完整性不足;3)

当三个模块同时使用时,模型在乳腺、肺、胃等多个器官数据集上均取得最优性能,验证了三个模块的有效性。实验结果表明,判别式分割器、扩散优化器与Transformer耦合编码嵌入模块各自对提升分割性能具有显著贡献,各模块的协同作用共同保障

表3 不同架构在多疾病图像上的定量比较结果,分别展示了三种疾病状态下测试图像的评估指标

Table 3 Quantitative comparison results of different architectures on multi-disease images, respectively showing the evaluation metrics of test images under three disease states

模型	SD		DS		DD	
	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)
Snake R-CNN	75.75	59.69	78.10	64.02	71.84	56.81
MedT	79.08	68.08	80.48	67.83	79.93	66.80
TransNuSeg	79.90	66.81	79.03	65.41	69.87	55.31
SegDiff	70.91	55.03	76.87	63.44	77.64	66.75
MedSegDiff	71.11	55.44	77.62	63.69	79.18	67.45
GenSelfDiff-HIS	79.02	67.81	80.96	68.01	65.03	53.45
cDAL	77.11	68.87	80.85	67.97	74.27	59.94
<b>dcDiff</b>	<b>80.12</b>	<b>68.22</b>	<b>81.02</b>	<b>68.15</b>	<b>80.62</b>	<b>67.80</b>

注:黑色加粗字体表示最优结果。

表4 在乳腺、肺和胃三个不同器官数据集上,对dcDiff中提出的判别式分割器模块 $\varphi$ 、扩散优化器模块 $z_\theta$ 、Transformer耦合编码嵌入模块 $\mathcal{A}$ 进行的消融实验结果

Table 4 Ablation study results of the discriminative segmenter module  $\varphi$ , diffusion refiner module  $z_\theta$ , and Transformer-coupled encoding embedding module  $\mathcal{A}$  proposed in dcDiff on three different organ datasets: breast, lung, and stomach

$\varphi$	$z_\theta$	$\mathcal{A}$	乳腺		肺		胃	
			Dice (%)	mIoU (%)	Dice (%)	mIoU (%)	Dice (%)	mIoU (%)
			72.58	57.33	79.35	65.58	84.46	73.73
			76.77	64.20	74.53	63.01	82.88	71.22
			77.84	63.86	79.95	66.04	85.64	75.63
			<b>78.61</b>	<b>64.84</b>	<b>81.02</b>	<b>68.15</b>	<b>86.67</b>	<b>76.48</b>

注:黑色加粗字体表示最优结果,“ ”表示使用了该模块,“ ”表示未使用该模块。

了dcDiff的最优性能。

## 4 结论

本文针对数字病理图像细胞核分割中重叠、成团结构难以精准分割的问题,提出了一种融合

判别式分割与生成式扩散的双重条件扩散模型dcDiff。主要结果与不足如下:

1)构建了基于DDPM的双重条件扩散模型,由判别式分割器与扩散优化器互补级联。其中,判别式分割器基于U型网络架构提取多尺度卷积特征,生成粗糙分割掩码,为扩散过程提供空间先验;扩散优化器使用初始掩码作为输入,通过迭代去噪过程逐步优化分割结果,并将细胞核分割任务定义为条件掩码生成问题。所提出模型通过扩散生成能够有效建模复杂细胞核完整结构。

2)提出了Transformer耦合编码嵌入模块。该模块包含像素与语义两条学习路径,并利用交叉注意力实现深度融合,通过将空间与语义互补先验嵌入扩散学习中,以引导扩散过程聚焦细胞核个体区域,从而有效解决了传统方法中空间与语义信息丢失问题。

3)提出了随机潜在嵌入策略。该策略将服从标准高斯分布的随机潜在嵌入注入扩散优化器的每一层,使模型能够学习大步长下的去噪映射,在保持分割精度的同时大幅减少所需时间步数,显著提升了模型的训练与推理效率。

4)本文方法虽在细胞核分割任务上取得显著成效,但仍存在一定局限性。多组器官和不同疾病状态数据集上的实验结果表明:dcDiff在Dice系数和平均交并比等指标上均优于现有数据驱动和扩散生成深度分割模型。然而,dcDiff的参数数量和计算复杂度相较于轻量级CNN模型仍偏高,这主要是因为去噪扩散概率模型中迭代去噪架构虽能提升分割精度,但其多步生成特性决定了需要较大规模的网络容量来建模复杂的数据分布。

综上,本文提出的双重条件扩散模型通过融合判别式分割与生成式扩散的互补优势,在复杂重叠细胞核结构的分割任务上取得了较好效果,为高精度细胞核分割提供了新的解决方案。未来工作将从模型结构压缩及高效采样策略等方向进一步探索,以推动该方法在临床场景中的实际应用。

## 参考文献 (References)

- Amit T, Shaharbany T, Nachmani E and Wolf L. Segdiff: Image segmentation with diffusion probabilistic models [EB/OL]. [2021-12-01].

- <https://arxiv.org/pdf/2112.00390.pdf>
- Benny Y and Wolf L. 2022. Dynamic dual-output diffusion model//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans IEEE: 11482-11491. [DOI: 10.1109/CVPR52688.2022.01119]
- Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. 2022. Swin-unet: Unet-like pure transformer for medical image segmentation//European conference on computer vision. Cham: Springer Nature Switzerland: 205-218.[DOI: 10.1007/978-3-031-25066-8\_9]
- Ciṙsan D C, Giusti A, Gambardella L M and Schmidhuber J. 2013. Mitosis detection in breast cancer histology images with deep neural networks//International conference on medical image computing and computer-assisted intervention. Berlin, Heidelberg: Springer Berlin Heidelberg: 411-418. [DOI: 10.1007/978-3-642-40763-5\_51]
- Croitoru F A, Hondru V, Ionescu R T and Shah M. 2023. Diffusion models in vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 45 (9) : 10850-10869. [DOI: 10.1109/TPAMI.2023.3261988]
- He Z, Unberath M, Ke J and Shen Y. 2023. Transnuseg: A lightweight multi-task transformer for nuclei segmentation//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, : 206-215.[DOI: 10.1007/978-3-031-43901-8\_20]
- Hejrati B, Banerjee S, Glide-Hurst C and Dong M. 2024. Conditional diffusion model with spatial attention and latent embedding for medical image segmentation//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland: 202-212. [DOI: 10.1007/978-3-031-72114-4\_20]
- Ho J, Jain A and Abbeel P. Denoising diffusion probabilistic models. 2020. Advances in neural information processing systems, 33: 6840-6851.[DOI: 10.48550/arXiv.2006.11239]
- Kazerouni A, Aghdam E K, Heidari M, Azad R, Fayyaz M, Hacihaliloglu I, et al. 2023. Diffusion models in medical imaging: A comprehensive survey. Medical image analysis, 88: 102846. [DOI: 10.1016/j.media.2023.102846]
- Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A and Sethi A. 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE transactions on medical imaging, 36(7): 1550-1560.[DOI: 10.1109/TMI.2017.2677499]
- Kumar N, Verma R, Anand D, Zhou Y, Onder O P, Tsougenis E, et al. 2019. A multi-organ nucleus segmentation challenge. IEEE transactions on medical imaging, 39 (5) : 1380-1391. [DOI: 10.1109/TMI.2019.2947628]
- Purma V, Srinath S, Srirangarajan S, Kakkar A and Prathosh A P. 2025. GenselfDiff-HIS: generative self-supervision using diffusion for histopathological image segmentation. IEEE Transactions on Medical Imaging, 44 (2) : 618-631. [DOI: 10.1109/TMI. 2024.3453492]
- Qiu Z, Pan Y, Wei J, Wu D, Xia Y and Shen D. 2021. Predicting symptoms from multiphase MRI via multi-instance attention learning for hepatocellular carcinoma grading//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing: 439-448.[DOI: 10.1007/978-3-030-87240-3\_42]
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 10684-10695. [DOI: 10.1109/CVPR52688.2022.01042]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international Publishing: 234-241.[DOI: 10.1007/978-3-319-24574-4\_28]
- Shi J, Wang T T, Zhu Z Q, Zhao M F, Wang B X and An H. 2025. Deep learning-based medical image segmentation methods. Journal of Image and Graphics, 30(6): 2161-2186 (石军, 王天同, 朱子琦, 赵敏帆, 王炳勋, 安虹. 2025. 基于深度学习的医学图像分割方法综述. 中国图象图形学报, 30(6): 2161-2186) [DOI: 10.11834/jig.240467]
- Song J, Cai Z, Song Y, Lian Z and Xiao L. 2023. Learnable Snake R-CNN for Instance-Level Biomedical Image Segmentation//2023 IEEE International Conference on Image Processing (ICIP). IEEE: 2920-2924.[DOI: 10.1109/ICIP49359.2023.10222722]
- Song J, Meng C and Ermon S. 2020. Denoising diffusion implicit models [EB/OL].[2020-10-06]. <https://arxiv.org/pdf/2010.02502.pdf>
- Valanarasu J M J, Oza P, Hacihaliloglu I and Patel V M. 2021. Medical transformer: Gated axial-attention for medical image segmentation//International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing: 36-46.[DOI: 10.1007/978-3-030-87193-2\_4]
- Wang H, Zhu Y, Green B, Adam H, Yuille A and Chen L C. 2020. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation//European conference on computer vision. Cham: Springer International Publishing: 108-126. [DOI: 10.1007/978-3-030-58548-8\_7]
- Wu J, Fu R, Fang H, Zhang Y, Yang Y, Xiong H, et al. 2024. Medsegdiff: Medical image segmentation with diffusion probabilistic model//Medical imaging with deep learning. PMLR: 1623-1639. [DOI: 10.48550/arXiv.2211.00611]
- Zeng H, Shan X, Feng Y and Wen Y. 2023. Msaanet: Multi-scale axial attention network for medical image segmentation//2023 IEEE International Conference on Multimedia and Expo (ICME). IEEE: 2291-2296.[DOI: 10.1109/ICME55011.2023.00391]
- Zhang W, Lian J and Dong B. 2024. Multi-Scale Position-Aware Cell

Nucleus Mask Attention for Tumor Budding Detection [C]//2024 IEEE International Conference on Multimedia and Expo (ICME). IEEE: 1-6.[DOI: 10.1109/ICME57554.2024.10688170]

Zhang X F, Zhang S, Zhang D H and Liu R. 2023. Group attention based medical image segmentation model. Journal of Image and Graphics, 28(10): 3231-3242 (张学峰, 张胜, 张冬晖, 刘瑞. 2023. 引入分组注意力的医学图像分割模型. 中国图象图形学报, 28(10): 3231-3242)[DOI:10.11834/jig.220748]

Zhao Y, Liu H, Yan J, Tian Y and Han X. 2025. Hypergraph Self-Supervised Learning for Survival Prediction on Whole Slide Images//2025 IEEE International Conference on Multimedia and Expo (ICME). IEEE: 1-6. [DOI: 10.1109/ICME59968.2025.11209897]

Zhou T, Dong Y L, Huo B Q, Liu S and Ma Z J. 2021. U-Net and its

applications in medical image segmentation: a review. Journal of Image and Graphics, 26(9): 2058-2077 (周涛, 董雅丽, 霍兵强, 刘珊, 马宗军. 2021. U-Net网络医学图像分割应用综述. 中国图象图形学报, 26(9): 2058-2077) [DOI: 10.11834/jig.200704]

### 作者简介

王征宇,男,硕士研究生,主要研究方向为生物医学图像处理、深度学习。E-mail: 1048685115@qq.com

宋杰,通信作者,男,副教授,主要研究方向为深度学习、生物医学图像计算。E-mail: j.song0041@gmail.com

王彩玲,女,副教授,主要研究方向为图像处理、模式识别与机器学习、计算机视觉。E-mail: wangcl@njupt.edu.cn