

动态时序引导的病理图像分类扩散模型

赵赫阳¹, 陈如豪¹, 罗一恺¹, 张 鹏¹, 王文剑^{2*}

(1. 山西大学计算机与信息技术学院, 山西太原 030006;

2. 山西大学计算智能与中文信息处理教育部重点实验室, 山西太原 030006)

摘要: 基于深度学习的病理图像分类是实现高效、可靠病理辅助诊断的关键技术基础,对病理图像智能分析、病变区域精准识别与组织类型自动判别具有重要研究意义。当前主流病理图像分类方法多依赖单一时间点的静态特征提取,难以有效表征病理状态从正常、发育异常到恶性转化的连续演进过程,无法充分建模组织层面复杂的层级结构与空间分布特征,导致对相似病理区域、细微病变结构的识别能力存在明显局限,模型判别精度与泛化性能难以进一步提升。针对上述问题,本文提出一种动态时序引导的病理图像分类扩散模型(Dynamic Temporal-guided Diffusion for Pathology, DT-DPath),构建统一的病理图像分类框架。该模型以扩散模型为基础,将前向加噪与反向去噪过程模拟为组织形态学的退化与重构过程,通过聚合扩散过程中不同噪声水平下的时序动态特征,构建能够刻画病理演化过程的判别性表征,精准捕捉细胞层面的细微结构差异,并有效建模组织内部的复杂层级与空间分布关系。在特征提取阶段,本文设计时序动态通道注意力机制(Temporal Dynamic Channel Attention, TDCA),将通道注意力扩展至时序维度,自适应聚焦于关键时序阶段的特征响应,增强模型对时序演化信息的表征能力。为优化训练效率,本文采用基于时序贡献的非均匀重要性采样策略,依据各时间步的分类显著性与信息密度动态构建采样分布,优先保留高判别力的中间去噪阶段,在保持分类性能的同时显著提升模型训练速度与稳定性。实验在结直肠癌、乳腺癌及淋巴结转移等多种病理图像公开数据集上开展,与CNN(Convolutional Neural Networks)、Transformer、扩散模型分类器以及病理基础模型等当前先进方法对比,结果表明DT-DPath在分类准确率、精确率、召回率及 F_1 分数等指标上全面领先,性能提升了0.65~3.59个百分点。消融实验验证了时序动态通道注意力与非均匀重要性采样的有效性,可视化分析显示模型可精准聚焦病理判别关键区域,特征聚类效果显著优于传统方法。本文方法实现了从静态特征建模到动态时序演化表征的范式转变,为病理图像智能分类提供了新的技术路径,也为扩散模型在病理图像分析领域的应用提供了可解释的研究思路。

关键词: 扩散模型;时序特征引导;时序通道注意力;非均匀重要性采样;图像分类;病理图像

基金项目: 国家自然科学基金(No. 62506220, No. 62476157, No. U21A20513);山西省基础研究计划资助项目(No.202403021222024);山西省专利转化专项计划项目(No.20250016);山西省高等学校科技创新项目(No.2024L002)

中图分类号: R361.3; TP391.41

文献标识码: A

文章编号: 0372-2112(2026)04-1736-14

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20251079

Dynamic Temporal-Guided Diffusion Model for Pathological Image Classification

ZHAO Heyang¹, CHEN Ruhao¹, LUO Yikai¹, ZHANG Peng¹, WANG Wenjian^{2*}

(1. School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China;

2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China)

Abstract: Deep learning-based pathological image classification is a key technical foundation for efficient and reliable computer-aided pathological analysis, which is of great significance for intelligent analysis of pathological images, accurate identification of lesion regions and automatic discrimination of tissue types. Current mainstream pathological image classification methods mostly rely on static feature extraction at a single time point, which makes it difficult to effectively represent the continuous evolution process of pathological states from normal, developmental abnormalities to malignant transformation, and cannot fully model the complex hierarchical structure and spatial distribution features at the tissue level. As a result, the recognition ability for similar pathological regions and subtle lesion structures is obviously limited, and the discrimination accuracy and generalization performance of the model are difficult to further improve. To address the above issues, this paper proposes a dynamic temporal-guided diffusion model for pathology (DT-DPath), which constructs a unified classification framework for pathological images. Based on the diffusion model, this framework simulates the forward

noise addition and reverse denoising processes as the degradation and reconstruction of tissue morphology. By aggregating temporal dynamic features at different noise levels during the diffusion process, it constructs discriminative representations that depict the pathological evolution process, accurately captures subtle structural differences at the cellular level, and effectively models the complex hierarchical and spatial distribution relationships within tissues. In the feature extraction stage, this paper designs a temporal dynamic channel attention (TDCA) mechanism, which extends channel attention to the temporal dimension and adaptively focuses on feature responses at key temporal stages to enhance the model's ability to represent temporal evolution information. To optimize training efficiency, a non-uniform importance sampling strategy based on temporal contribution is adopted, which dynamically constructs the sampling distribution according to the classification saliency and information density of each time step, and preferentially retains intermediate denoising stages with high discriminative power. This strategy significantly improves training speed and stability while maintaining classification performance. Experiments are conducted on a variety of public pathological image datasets including colorectal cancer, breast cancer and lymph node metastasis. Compared with state-of-the-art methods such as CNN, Transformer, diffusion model classifiers and pathological foundation models, the results show that DT-DPath achieves comprehensive superiority in accuracy, precision, recall and F_1 -score, with performance improvements ranging from 0.65~3.59 percentage points. Ablation experiments verify the effectiveness of temporal dynamic channel attention and non-uniform importance sampling, and visualization analysis shows that the model can accurately focus on key regions for pathological discrimination, with significantly better feature clustering effects than traditional methods. The proposed method realizes the paradigm shift from static feature modeling to dynamic temporal evolution representation, provides a new technical path for intelligent classification of pathological images, and also offers an interpretable research idea for the application of diffusion models in the field of pathological image analysis.

Keywords: diffusion model; temporal feature guidance; temporal channel attention; non-uniform importance sampling; image classification; pathological image

Foundation Item(s): National Natural Science Foundation of China (No.62506220, No.62476157, No.U21A20513); Fundamental Research Program of Shanxi Province (No.202403021222024); Special Project for Patent Transformation in Shanxi Province (No.20250016); Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (No.2024L002)

0 引言

病理学图像作为肿瘤诊断的金标准,能够提供丰富的组织学和细胞学信息,有助于准确识别癌变区域、评估肿瘤分级与分期,从而为临床制定精准治疗方案提供重要依据^[1-4]。然而,病理切片通常呈现出高度复杂的形态学特征,如细胞核的多形性与异型性,腺体结构从轻度紊乱到完全破坏的谱系变化,以及伴随的间质纤维化反应与炎性细胞浸润等。这些复杂的形态表现使得病理学图像的自动分析面临巨大挑战。传统的人工诊断方法存在效率低、依赖专家经验、主观性强及可重复性差等局限,易导致诊断结果的不一致性^[5-6]。在此背景下,融合先进图像处理技术与机器学习算法,实现对病理特征的自动识别、定量分析和准确分类,已成为提升病理辅助诊断的重要途径。

近年来,深度学习在图像识别与分类领域的显著进展,为病理学图像分析开辟了新途径。卷积神经网络(Convolutional Neural Networks, CNN)凭借其局部连接与特征提取,能够有效提取多尺度视觉特征,在多项医学图像分析任务中表现出色^[7]。例如, Peng 等人^[8]通过迁移学习结合 ResNet50 模型实现了结直肠

癌病理图像的分类; Khayatian 等人^[9]提出了一种基于 EfficientNetV2B0 与 CatBoost 的混合框架,在多个病理数据集上取得了良好的分类效果; Hamid 等人^[10]提出的 Alter-AttUNet 通过空间注意力门机制,有效缓解了稀疏标注和类别不平衡问题,提升了病理图像分类性能; Murtaza 等人^[11]设计的 BMICNet 架构采用两阶段训练策略,进一步提升了乳腺癌病理图像的分类性能。然而, CNN 在建模长程依赖关系方面存在局限,难以充分捕获病理图像中复杂的全局上下文信息。为弥补这一不足,研究者开始引入 Transformer 架构,通过图像分块与自注意力机制实现病理图像像素级长距离依赖建模^[12-15]。Kolla 等人^[16]提出改进的 Swin-Transformer V2 架构,显著增强了多尺度特征融合能力; Liu 等人^[17]设计的层次图金字塔变换器(Hierarchical Graph Pyramid Transformer, HGPT)利用几何先验提升了分类精度; Ding 等人^[18]通过稀疏交互与权重共享机制提升了 Transformer 的分类性能和计算效率。总体而言,基于 Transformer 的方法通过自注意力机制捕获全局静态特征,进一步推动了病理图像分类性能的提升。

尽管 CNN 与 Transformer 在病理图像分类中表现出强大的特征提取能力,但二者均存在一个根本性局

限,即依赖于单一时间点的静态特征表示,未能有效建模病理形态演变中的时序动态信息。这种对时序动态性(Temporal Dynamics, TD)的忽视,本质上偏离了病理发展的生理学基础——疾病的发生与发展是一个连续、渐进的演化过程。以结直肠癌为例,从正常结肠黏膜经腺瘤样息肉、低级别上皮内瘤变至高级别上皮内瘤变,最终演变为浸润性腺癌,这一病程伴随着细胞核异型性逐渐加重、腺体结构从轻度紊乱到完全破坏的谱系变化,以及间质纤维化与炎性浸润的动态响应^[1,5]。类似地,在前列腺癌中,从良性前列腺增生经不同级别的前列腺上皮内瘤变至浸润性腺癌,细胞与组织形态呈现出明确的时序过渡特征。这种从正常架构经发育异常到恶性转化的动态演化过程,蕴含着丰富的生物学先验知识。然而,现有方法普遍未能有效利用这一信息。传统模型仅基于单一时间点的静态图像进行特征建模,难以捕捉随病程推进的形态学演化轨迹,导致在识别形态相似亚型及检测早期病变方面表现受限。

基于上述分析,本文提出一种动态时序引导的全新病理学分类扩散框架(Dynamic Temporal-guided Diffusion for Pathology, DT-DPath)。该框架通过显式建模病理特征的时序演化过程,实现对组织形态连续变化的有效捕捉。DT-DPath以扩散模型为基本架构,在前向过程中模拟病理图像的渐进退化,并在反向去噪过程中重构其形态恢复路径,从而构建从高噪声退化状态到清晰原始图像的完整演化轨迹。在反向去噪的中间阶段,模型提取并聚合不同噪声水平下的时序动态特征,形成从宏观器官结构到微观细胞形态重构的判别性演化序列。为增强对关键演化阶段的感知能力,本文设计了时序动态通道注意力机制(Temporal Dynamic Channel Attention, TDCA),该机制能够自适应地聚焦于对病理诊断最具判别力的中间去噪状态。同时,本文引入一种基于时序贡献的非均匀重要性采样策略,优先保留对分类任务最具信息量的动态中间状态,在保持模型表征能力的同时显著提升训练效率。在NCT-CRC-HE-100K、BreakHis和PCam等多个病理图像数据集上的实验表明,DT-DPath在分类任务中均优于CNN、Transformer、扩散分类模型及当前病理领域最先进的基础模型,验证了所提出方法的有效性与先进性。

本文的主要贡献如下:

(1)提出了一种动态时序引导的病理图像分类扩散模型(DT-DPath)。该模型将扩散过程转化为病理形态的退化-重构模拟,通过聚合多时间步动态特征构建判别性演化轨迹,实现了从静态快照到动态轨迹的表征范式转变,有效克服了传统方法难以捕捉病理

状态连续演化过程的根本局限。

(2)设计了一种时序动态通道注意力机制,将通道注意力扩展至时序维度,使其能够随去噪进程自适应地调整权重,增强对病理演化关键中间阶段的特征响应,提升特征的判别能力与鲁棒性,从而弥补了传统静态注意力机制忽略时序判别信息的固有缺陷。

(3)提出了一种基于时序贡献的非均匀重要性采样策略,通过动态评估各时间步的分类显著性与信息密度构建采样分布,优先保留最具判别力的中间去噪阶段,在维持分类性能的同时显著降低训练时间与计算开销,有效克服了均匀采样在低信息密度阶段的资源浪费问题。

1 动态时序引导的病理图像分类扩散模型

1.1 模型整体架构

本文提出的DT-DPath采用端到端训练的条件扩散框架。模型参数首先通过自然图像预训练权重进行初始化,以增强 saliency map 生成的稳定性,并结合L2正则化进一步提升特征提取的鲁棒性。随后,模型在病理数据集上进行fine-tune,以适配特定任务的特征分布。其核心在于通过多时间步动态特征融合,构建能够捕捉病理形态连续演化的判别性时序表征。

与静态空间特征建模的传统方法不同,DT-DPath以扩散模型为主体框架,更加注重对动态时间特征的捕捉。传统生成式扩散模型仅通过加噪-去噪过程学习图像生成,而DT-DPath则充分利用其多时间步特性,系统性地提取并整合不同时间步的动态特征向量,构建具有时序演化特性的判别性表示。该机制能够从更深层次揭示病理图像的复杂演化规律,不仅提供更丰富的多尺度语义信息,还能实现更精确的特征表示,从而显著提升分类性能。整体架构如图1所示。

具体工作流程如下:模型首先对原始图像 \mathbf{x}_0 逐步添加噪声,生成一系列退化图像 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$,最终得到接近高斯分布的噪声图像 \mathbf{x}_T 。在反向扩散过程中,模型通过U形结构逐步去噪,同时引入时序动态通道注意力机制(TDCA),自适应地增强对病理演化关键阶段敏感性的特征通道,从而引导模型聚焦于细胞核形态、腺体结构等关键区域。为提升训练效率与模型稳定性,本文在扩散过程中引入重要性采样策略(图1中虚线区域)。该策略根据前向扩散中的噪声强度分布,优先选取对病理动态演化建模贡献显著的时间步进行采样,对部分贡献较低的时间步,则提前终止采样并返回空向量。这一设计使模型能够更加专注于关键演化阶段,有效提高计算资源的利用效率。最终,模型将筛选得到的动态特征向量进行时序

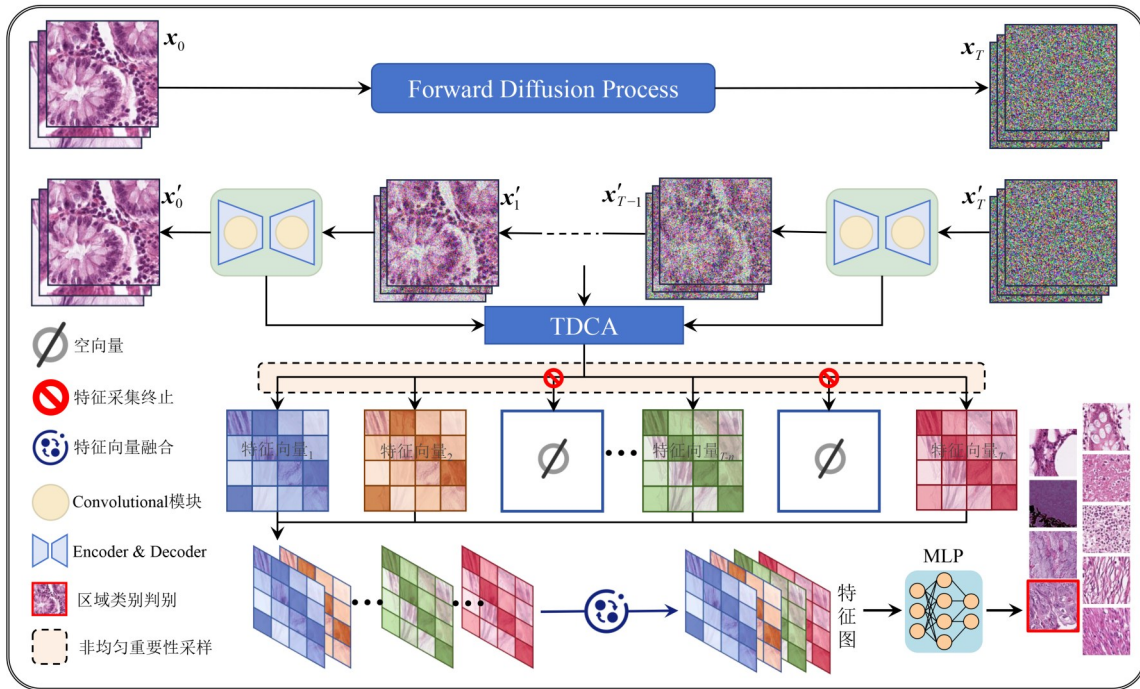


图1 DT-DPath 模型架构

Figure 1 DT-DPath model architecture

融合,形成能够完整捕捉病理演化过程的特征序列,作为分类判断依据。

1.2 动态时序引导的特征提取

DT-DPath的核心在于将扩散模型的加噪-去噪过程转化为病理形态的退化-重构模拟,从而显式构建从高噪声退化状态到清晰原始图像的完整时序演化轨迹。这种动态时序引导的特征提取范式,突破了传统静态单帧特征建模的局限,通过系统性聚合反向扩散过程中不同噪声水平下的中间表示,形成能够刻画病理状态连续演进的判别性序列,精准捕捉细胞层面细微结构差异与组织层面复杂层级关系的动态变化规律。

在前向加噪过程中,模型逐步向原始图像 x_0 添加高斯噪声,生成一系列退化图像 x_1, x_2, \dots, x_T , 直至接近纯噪声分布 x_T 。这一过程模拟了病理形态从清晰完整到逐步破坏的退化路径,可形式化为

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

其中, x_t 表示时间步 t 的数据, β_t 是控制噪声强度的调度假参数, \mathbf{I} 为单位矩阵。

在反向去噪过程中,模型从噪声图像 x_T 出发,逐步恢复原始内容,重建细胞与组织结构在时序上的形态演化轨迹。这一逆向链定义为

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

其中, μ_θ 和 Σ_θ 是通过神经网络学习均值与方差。

DT-DPath 的特征提取发生在反向扩散的中间阶

段,对于选定的时间步 i ,利用 U-Net 从含噪图像 x_i 与时间嵌入 t 中提取动态特征向量:

$$f_i = \text{U-Net}_\phi(x_i, t; \text{TDCA})$$

其中, ϕ 为网络参数。这些特征 $f_i (i = 1, 2, \dots, k)$ 捕捉了从宏观器官结构崩解到微观细胞形态逐步重构的瞬时状态,蕴含丰富的时序演化信息。将多个时间步的动态特征进行时序连接与融合,构建完整特征序列:

$$\mathbf{F} = [f_1; f_2; \dots; f_k]$$

最终,该序列输入多层感知机实现分类预测。通过这一动态时序引导机制,模型不仅保留了扩散过程的多尺度语义信息,还显式表征了病理演化从正常到异常再到恶性转化的连续轨迹(如图2所示),显著提升了对相似病理区域的区分能力与对早期病变的敏感性。

1.3 时序动态通道注意力机制

DT-DPath 采用多时间步动态特征融合策略实现分类任务。传统注意力机制通常仅在单一特征图上分配权重,属于静态建模方式,而病理检测中的关键线索往往蕴含在形态演化的中间阶段。为此,本文提出一种时序动态通道注意力机制。该机制将传统静态的注意力权重扩展为随时间步变化的动态函数,使其能够在反向去噪的每一阶段,自适应地增强对病理演化最具判别性的中间状态特征的响应。

具体而言,在反向去噪的每个时间步 t , TDCA 通过全局平均池化(Global Average Pooling, GAP)与双层

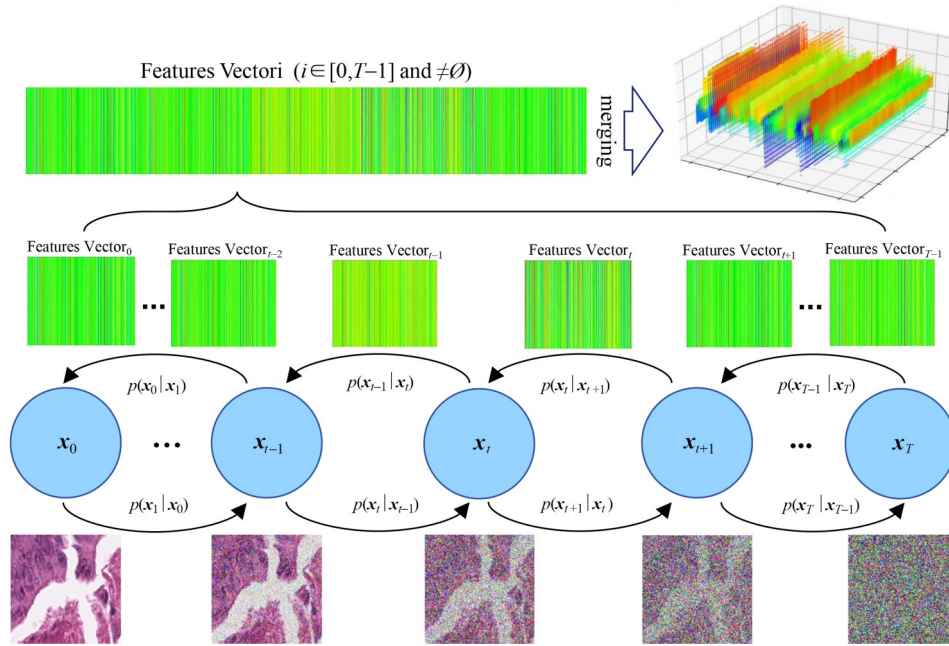


图2 扩散模型时序动态特征生成流程

Figure 2 Flowchart of temporal dynamic feature generation in diffusion model

全连接网络,生成随时间变化的通道级权重,用于增强反映病理形态渐变的关键通道响应。假设不同时间步获取的动态特征响应可表示为 $f_i \in \mathbb{R}^{C \times H \times W}$,其中 C 为通道数, H 和 W 分别为特征向量的高和宽。通过全局平均池化操作压缩空间维度,得到每个通道在当前时间步 t 的全局响应,其计算方式如下:

$$S_{c,t} = \text{GAP}(f)_{c,t} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_{c,i,j,t}$$

其中, $f_{c,i,j,t}$ 表示特征向量 f_i 在通道 c 、位置 (i, j) 的值, GAP 表示该通道在整个特征向量上的平均响应。由此得到 $S_{c,t}$ 表示时间步 t 中的第 c 个通道的全局平均响应,它反映了该通道在时间步 t 的重要程度,用于评估各通道在当前演化阶段中的重要性程度。基于这一响应,通过全连接层进一步生成通道注意力权重的过程可表示为

$$z_{c,t} = \text{ReLU}(\text{FC}(S_{c,t})) \beta_{c,t} = \sigma(\text{FC}(z_{c,t}))$$

其中, σ 是Sigmoid函数, FC 代表全连接层, $\beta_{c,t}$ 代表时间步 t 中第 c 个通道所对应的注意力权重。通过这种方式,TDCA能够为每个时间步中的特征通道动态分配差异化权重,从而有效增强与分类任务最相关的特征响应。

最终,将通道注意力权重与原始动态特征响应 $f_{c,t}$ 逐通道相乘,得到经时序加权的优化特征:

$$f'_{c,t} = f_{c,t} \circ \beta_{c,t}$$

通过该时序加权机制,TDCA能够在反向去噪过程中实现动态注意力分配,在早期阶段侧重于噪声抑

制,中期聚焦于形态结构增强,晚期则致力于边界细节精化。这一过程显著提升了模型对病理演化轨迹的建模与表征能力。

1.4 非均匀重要性采样

DT-DPath需从上千个时间步中筛选对分类最具贡献的少数阶段。若均匀采样,所有时间步等概率参与训练,将浪费大量计算在低信息密度的纯噪声或饱和清晰阶段。为此,本文提出非均匀重要性采样策略,依据各时间步 t 在分类梯度显著性与病理信息密度两方面的综合贡献,动态构建采样概率分布。具体而言,在反向扩散训练过程中,该策略通过以下数学表达式为每个时间步 t 分配相应的采样权重:

$$w_t = \sigma \cdot \text{Salienc}(f_t) + (1 - \sigma) \cdot \text{Densin}(f_t)$$

其中,显著性项 $\text{Salienc}(f_t)$ 通过分类损失对 f_t 的梯度范数衡量其对最终预测的贡献;信息密度项 $\text{Densin}(f_t)$ 通过特征图熵评估形态复杂性。二者共同识别出病理形态渐变的关键中间态,例如腺体结构从完整到半破坏的过渡阶段。 $\text{Densin}(f_t)$ 表示动态响应 f_t 的时序信息密度,定义为特征图的空间信息熵:

$$\text{Densin}(f_t) = - \sum_{i=1}^H \sum_{j=1}^W p(f_{t,i,j}) \log p(f_{t,i,j})$$

其中, $p(f_{t,i,j})$ 为特征图 f_t 在位置 (i, j) 的归一化激活值; H 和 W 分别为特征图的高和宽; $\sigma \in [0, 1]$ 为平衡参数,通过验证集优化以平衡显著性与信息密度的贡献。

基于重要性权重 w_t ,可以构造非均匀采样概率分布:

$$P(t) = \frac{w_t}{\sum_{\tau=1}^T w_\tau}$$

其中, T 为扩散过程中的总时间步数。通过该概率分布, 模型更倾向于选择那些包含更丰富病理信息的时间步, 例如反映细胞核轮廓重构或腺体边界恢复等关键演化阶段的时间步。

1.5 联合损失函数

DT-DPath 的训练过程通过优化以下联合损失函数实现:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{denoise}} + (1 - \alpha) \mathcal{L}_{\text{class}}$$

其中, $\mathcal{L}_{\text{denoise}}$ 为扩散模型的去噪损失, $\mathcal{L}_{\text{class}}$ 为分类交叉熵损失, α 为加权系数, 通过验证集上的分类性能动态调整, 以平衡去噪重建与分类判别的收敛速度。训练中通过非均匀重要性采样选择关键时间步, 确保计算资源集中于对病理演化最具贡献的中间阶段。

去噪任务保证中间表示忠实于病理形态的恢复路径, 分类任务引导模型聚焦演化过程中的高判别性阶段。二者协同优化, 使模型在学习高质量时序表征的同时, 直接服务于最终分类目标, 显著提升训练效率与模型稳定性。

2 实验和结果

2.1 实验设置

2.1.1 数据集

本研究选取结直肠癌 (NCT-CRC-HE-100K)、乳腺癌 (BreakHis) 以及淋巴结转移 (PCam) 3 个基准数据集进行验证。NCT-CRC-HE-100K 代表了从正常组织、癌前病变再到浸润性癌的早-中期连续谱系; BreakHis 聚焦于多种从良性向恶性转化的中间过渡亚型; 而 PCam 则体现了癌细胞晚期远处转移的微小灶特征。三者涵盖了从初期异常→中间过渡→晚期扩散的完整病理演进时序过程, 为验证本文所提出的动态时序引导框架在捕捉病理演化全过程中的有效性提供了数据实验基础。具体实验数据如表 1 所示, 图 3 展示了 3 个数据集中各类组织类型的病理图像示例。

表 1 实验数据集信息

Table 1 Experimental dataset information

数据集	组织类型	亚型	分类目标	训练样本	测试样本
NCT-CRC-HE-100K	结直肠癌	9	组织类型	70 000	15 000
BreakHis	乳腺癌	8	癌症类型	5 537	1 186
PCam	淋巴结	2	转移状态	262 144	32 768

数据集 1: NCT-CRC-HE-100K^[19] 是专用于结直肠癌组织病理学图像分类的基准数据集, 共包含 100 000 张 H&E 染色的人类结直肠癌及正常组织病理学图像,

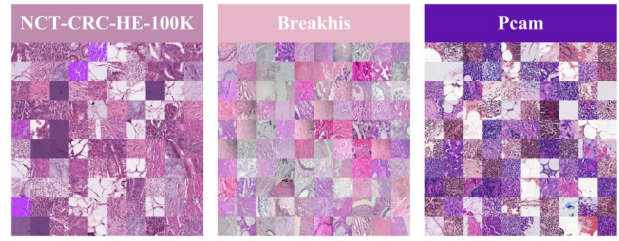


图 3 数据集示例

Figure 3 Dataset examples

涵盖 9 种组织类型, 分别为脂肪组织 (Adipose tissue, ADI)、背景 (Background, BACK)、组织碎片 (Debris, DEB)、淋巴细胞 (Lymphocytes, LYM)、黏液 (Mucus, MUC)、平滑肌 (Smooth muscle, MUS)、正常结肠黏膜 (Normal colon mucosa, NORM)、癌相关基质 (Cancer-associated stroma, STR) 以及结直肠癌腺癌上皮 (Colorectal adenocarcinoma epithelium, TUM)。所有图像均采用 Macenko 方法进行颜色标准化处理^[20]。

数据集 2: BreakHis^[21] 是一个专门用于乳腺癌组织病理学图像分类的基准数据集, 共收集 7 909 张图像, 涵盖 2 480 张良性和 5 429 张恶性肿瘤样本, 并进一步细分为 4 种良性与 4 种恶性, 共计 8 种组织学亚型。

数据集 3: PCam^[22] 是一个专用于淋巴结转移检测的病理图像数据集, 包含 327 680 张图像, 阴性与阳性样本约各占 50%, 用于判断淋巴结切片中是否存在转移癌细胞的二分类任务。

2.1.2 多分类评价指标

本研究采用准确率、精确率、召回率及 F_1 分数这一组指标, 以系统评估模型的综合性能与稳健性。

准确率 (Accuracy, Acc) 用于衡量模型正确分类的样本数占总样本数的比例。在多分类任务中, 其计算公式可表示为

$$\text{Accuracy} = \frac{\sum_{i=1}^K \text{TP}_i}{N}$$

其中, K 类别总数, TP_i 表示第 i 类的真正例数, N 为整体样本数。Acc 反映了模型总体上的正确预测比例。

精确率 (Precision, Pre) 是通过计算每个类别的精确率, 然后对这些精确率求平均得到的。在多分类任务中, 总体精确率的计算公式为

$$\text{Precision} = \frac{1}{K} \sum_{i=1}^K \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}$$

其中, FP_i 表示第 i 类的假正例数。Pre 反映了模型在所有类别上预测的平均准确性。

召回率 (Recall, Rec) 是通过计算每个类别的召回率, 然后对这些召回率求平均得到的。在多分类任务中, 总体召回率的计算公式为

$$\text{Recall} = \frac{1}{K} \sum_{i=1}^K \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}$$

其中, FN_i 表示第 i 类的假负例数。Rec 反映了模型在所有类别上对实际样本的平均覆盖能力。

F_1 分数 (F_1 -Score) 是通过计算每个类别的 F_1 分数, 然后对这些 F_1 分数求平均得到的。在多分类任务中, 总体 F_1 分数的计算公式为

$$F_1 = \frac{1}{K} \sum_{i=1}^K 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

2.1.3 实验参数设置

本文所有实验均在 PyTorch 2.1 实现, 并在单个 Ubuntu 20.04 服务器上完成。硬件配置包括一颗 AMD EPYC 776 364 核处理器、503 GiB 内存、1 020 GB 硬盘, 以及一张显存为 80 GB 的 NVIDIA RTX A800 显卡, 训练过程通过 CUDA 11.8 进行加速。

实验所使用的图像数据在输入前均经过统一预处理。首先进行标准化, 将像素强度值归一化至 $[0, 1]$ 区间, 并调整至相同尺寸; 随后应用一系列数据增强策略, 包括随机旋转 (角度范围: $\pm 15^\circ$)、水平翻转和随机缩放 (缩放范围: 0.9~1.1)。缩放区间的设定基于病理图像的结构特性, 旨在避免组织与细胞形态因过度形变而失真, 同时提升模型对尺度变化的鲁棒性。

本实验共进行 100 轮训练, 并引入早停机制以防止过拟合。模型优化器选用 Adam, 初始学习率设置为 1×10^{-4} , 并采用阶梯式衰减策略, 每经 10 轮训练学习率减半。此外, 网络结构中加入了丢弃率为 0.4 的 Dropout 层, 以进一步增强模型正则化效果。所有实验结果均基于测试集进行评估, 该集合的标签信息在训练过程中完全不可见, 从而确保评估结果能够有效反映模型在真实场景中的泛化能力。

2.2 实验结果与分析

2.2.1 对比模型概述

为了验证本文方法的有效性, 我们将其与近年来几类主流的病理图像分类模型进行性能比较。所选取的对比模型涵盖典型 CNN、Transformer、扩散模型分类器以及病理基础模型, 具体说明如下:

(1) VGG16^[23]。该模型基于经典的 VGG (Visual Geometry Group) 架构, 通过多层卷积和池化结构有效提取图像的低级和高级特征。本研究采用预训练权重进行参数初始化, 并借助迁移学习方法, 使模型在组织病理图像的分类任务中展现出较高的准确性。

(2) DenseNet201^[24]。该模型采用密集连接机制将每一层与前面所有层直接相连, 实现特征值的重用与梯度的高效传播。在本文中, DenseNet201 能够以更少的参数量提取丰富的多尺度特征, 缓解梯度消失问题, 并显著提高对小病灶和复杂组织结构的识别能力。

(3) Xception-TL^[25]。该模型在 Inception 架构基础上引入深度可分离卷积, 通过解耦通道与空间相关性, 在保持高模型表达能力的同时显著降低了计算复杂度。在本文中, Xception-TL 通过迁移学习能够高效捕捉细胞核形态、腺体结构等细微差异, 并提升癌变区域定位的精确度。

(4) ResNet-50 (Transfer Learning, TL)^[8]。该模型通过引入残差连接有效解决深度网络的退化问题。在本文中, ResNet-50 (TL) 凭借其强大的深度特征提取能力, 通过迁移学习有效区分高分化与低分化肿瘤区域, 并增强对图像噪声及染色差异的鲁棒性。

(5) ViT-B/16^[26]。该模型是原始 Vision Transformer 的基本变体, 通过将图像分块并应用自注意力机制处理序列, 实现高效的全局建模。在病理图像分析中, ViT-B/16 能够捕捉长程依赖关系, 适用于大规模组织结构识别, 但需通过预训练提升在小数据集上的性能。

(6) DeiT-Base^[27]。该模型是数据高效的图像 Transformer, 通过知识蒸馏和注意力机制优化训练过程, 实现更少的计算资源消耗。在组织病理学图像分类中, DeiT-Base 能够快速适应特定任务, 提升对细胞形态和组织纹理的表征能力, 尤其在资源受限场景下表现出色。

(7) Swin-Transformer V2^[28]。作为一种改进的 Transformer 架构, 该模型采用分层窗口注意力机制, 提高了计算效率和特征提取能力。在医学影像分析中, Swin-Transformer V2 通过处理多尺度输入, 进一步提升了对病理图像的建模精度, 适用于肿瘤检测和分割任务。

(8) HGPT^[17]。该模型是一种层次图金字塔 Transformer, 通过有效地利用组织分布的几何表示, 进一步提升了模型对病理图像的建模能力。HGPT 能够整合多层次图结构信息, 实现对复杂组织形态的精确分类。

(9) Diffusion classifiers^[29]。该模型利用扩散模型作为零样本分类器, 通过密度估计实现图像分类, 而无需额外训练。在病理图像任务中, Diffusion classifiers 能够从生成式预训练中提取判别性特征, 支持零样本场景下的肿瘤识别。

(10) ImmunoDiff^[30]。该模型是一种基于扩散的免疫表型预测框架, 通过模拟成像变化预测。在组织病理学分析中, ImmunoDiff 能够生成后处理图像, 并整合解剖学知识提升预测准确性。

(11) UNI^[31]。该模型是一个通用的病理图像基础模型, 通过在大规模病理数据集上进行自监督预训练, 结合对比学习和掩码图像建模, 能够提取丰富的视觉特征, 适用于多种病理图像分类任务。其在细胞

形态和组织结构建模方面表现出较强的泛化能力。

(12)CONCH^[32]。该模型是一个专为病理图像设计的多模态基础模型,通过联合图像和文本的预训练,结合自监督和监督学习,能够有效捕捉病理图像中的语义信息和上下文特征。其在跨数据集任务中展现出较好的鲁棒性。

2.2.2 分类性能对比

为全面评估 DT-DPath 在病理图像分类中的性能,本文选取了 NCT-CRC-HE-100K、BreakHis 和 PCam 这 3 个涵盖不同组织类型与分类难度的公开数据集,与 CNN、Transformer、扩散模型以及基础模型进行了系统对比,实验结果如表 2 所示。

表 2 在 NCT-CRC-HE-100K、BreakHis、PCam 数据集上的分类性能对比

单位:%

Table 2 Comparison of classification performance on the NCT-CRC-HE-100K, BreakHis, and PCam datasets

unit:%

Model(Year)	NCT-CRC-HE-100K				BreakHis				PCam			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
VGG16 ^[23]	95.74	94.26	94.98	94.61	85.49	86.21	85.96	86.14	96.50	96.40	96.40	96.40
DenseNet201 ^[24]	97.50	97.45	97.40	97.42	87.20	87.10	87.05	87.08	98.90	98.90	98.80	98.90
Xception-TL ^[25]	97.80	97.75	97.70	97.72	93.32	92.98	92.36	92.44	98.20	98.15	98.10	98.12
ResNet-50 ^[8]	98.70	94.33	92.33	93.32	91.50	91.20	91.00	91.10	98.60	98.55	98.50	98.52
ViT-B/16 ^[26]	96.80	96.72	96.68	96.70	92.10	91.95	91.88	91.91	98.40	98.36	98.32	98.34
DeiT-Base ^[27]	97.30	97.25	97.20	97.22	92.80	92.65	92.58	92.61	98.65	98.60	98.56	98.58
Swin-Trans V2 ^[28]	97.85	97.80	97.76	97.78	93.40	93.25	93.18	93.21	98.90	98.86	98.82	98.84
HGPT ^[17]	98.20	98.16	98.12	98.14	93.90	93.75	93.68	93.71	99.05	99.01	98.97	98.99
Diff-classifiers ^[29]	98.53	98.49	98.47	98.51	94.37	94.33	94.31	94.35	96.20	96.16	96.14	96.18
ImmunoDiff ^[30]	98.47	98.93	98.91	98.95	94.81	94.77	94.75	94.79	97.10	97.06	97.04	97.08
UNI ^[31]	98.80	98.78	98.75	98.76	93.00	92.95	93.00	92.97	99.00	99.00	98.95	98.97
CONCH ^[32]	96.50	96.48	96.45	96.46	92.80	92.75	92.80	92.77	98.80	98.80	98.75	98.77
DT-DPath (our)	99.33	99.31	99.28	99.29	94.85	94.67	94.85	94.70	99.45	99.32	99.31	99.32

注:加粗数据表示最优值。

在 NCT-CRC-HE-100K 结直肠癌组织病理图像的九分类任务中,DT-DPath 取得了 99.33% 的准确率,相较于各类型模型中的最优基线均有提升:比 ResNet-50(98.70%) 提高 0.63 个百分点,较传统 Transformer 类最佳模型 HGPT(98.20%) 提升 1.13 个百分点,优于近期扩散模型 ImmunoDiff(98.97%) 0.36 个百分点,较基础模型 UNI(96.80%) 提升幅度达 2.53 个百分点。在其他关键指标上,DT-DPath 亦全面高于其他对比模型:精确率(Pre)达 99.31%,较 ImmunoDiff(98.93%) 提升 0.38 个百分点;召回率(Rec)为 99.28%,较 ImmunoDiff(98.91%) 提高 0.37 个百分点; F_1 分数达到 99.29%,较 ImmunoDiff(98.95%) 提升 0.34 个百分点。参考上述对比数据和数据集特性,在 NCT-CRC-HE-100K 九分类任务中 DT-DPath 准确率达 99.33%,较最佳基线提升 0.36~2.53 个百分点,有效克服了九类组织形态高度相似的难点,通过动态时序特征融合精准捕捉病理演化差异。

在 BreakHis 八亚型乳腺癌病理图像分类任务中,DT-DPath 实现了 94.85% 的准确率,显著优于其他类型的最佳模型:较 CNN 类最优模型 Xception-TL 提升 1.53 个百分点,较 Transformer 类最佳模型 HGPT 提高 0.95 个百分点,较基础模型 UNI(93.00%) 与 CONCH

(92.80%) 分别领先 1.85 个百分点与 2.05 个百分点。值得一提的是,尽管扩散模型在本任务中表现优异,Diffusion classifiers 和 ImmunoDiff 分别达到 94.37% 与 94.81% 的准确率,DT-DPath 仍实现对二者的超越,并在 F_1 分数上以 94.70% 优于 ImmunoDiff 的 94.79%。参考上述对比数据和数据集特性,在 BreakHis 八亚型任务中 DT-DPath 准确率 94.85%,领先扩散模型 0.04~1.53 个百分点,TDCA 与重要性采样聚焦腺体崩解关键阶段,实现细粒度亚型突破。

在 PCam 淋巴结转移二分类任务中,DT-DPath 取得了 99.45% 的准确率,在所有对比方法中表现最佳。具体而言,其准确率较 CNN 类最佳模型 DenseNet201(98.90%) 提升 0.55 个百分点,较传统 Transformer 类最佳模型 HGPT(99.05%) 提升 0.40 个百分点,较基础模型 UNI(99.00%) 提升 0.45 个百分点,并显著优于扩散模型 ImmunoDiff(97.10%) 2.35 个百分点。在其他关键指标上,DT-DPath 亦全面领先:精确率(Pre)达 99.32%,较 HGPT(99.01%) 提升 0.31 个百分点;召回率(Rec)为 99.31%,较 HGPT(98.97%) 提高 0.34 个百分点; F_1 分数达到 99.32%,较 HGPT(98.99%) 提升 0.33 个百分点。

参考上述对比数据和数据集特性,在 PCam 二分

类任务中 DT-DPath 准确率 99.45%，提升 0.40~2.35 个百分点，通过增强转移灶动态成形中间态，显著降低误判并提升微小病灶定位精度。

综上所述，DT-DPath 在 3 个数据集的各项评价指标上均全面优于所有对比模型。实验结果显示，其平均准确率较最佳 CNN 模型提升 0.90 个百分点，较最佳传统 Transformer 模型提升 0.83 个百分点，较最佳扩散模型提升 0.95 个百分点，较最佳基础模型提升 1.61 个百分点。性能优势的背后反映了建模思路的根本差异，传统模型受限于对病理图像某一瞬态的静态表征，而 DT-DPath 通过对特征动态演化过程的显式建模实现了有效突破。尽管扩散模型引入了过程信息，但其去噪过程缺乏针对病理语义的引导机制。相比之

下，DT-DPath 通过时序依赖注意力与重要性采样策略，能够精准聚焦于病理诊断中最具判别力的动态中间状态，构建从微观细胞到宏观组织的多层次时序一致性表示。这一设计突破了传统静态特征建模的固有局限，在复杂多变的病理场景中展现出更强大的泛化能力。

2.2.3 消融实验

为验证 DT-DPath 中各模块的有效性，本文在 NCT-CRC-HE-100K、BreakHis 和 PCam 这 3 个数据集上进行了系统的消融实验。实验中设置使用默认均匀采样时间步进行特征聚合的扩散模型作为 baseline，通过逐步引入不同组件，定量评估了各模块对模型性能贡献，具体结果如表 3 所示。

表 3 在 NCT-CRC-HE-100K、BreakHis、PCam 数据集上的消融实验

单位：%

Table 3 Ablation experiments on the NCT-CRC-HE-100K, BreakHis, and PCam datasets

unit：%

模型			NCT-CRC-HE-100K				BreakHis				PCam			
Baseline	重要性采样	TDCA	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
√			97.42	97.28	97.27	97.28	91.23	91.45	91.51	91.48	96.56	96.75	96.72	96.74
√	√		97.93	97.94	97.89	97.91	91.66	91.67	91.99	91.63	97.12	97.33	97.32	97.33
√		√	99.23	99.21	99.20	99.20	93.55	93.21	93.25	93.23	99.13	99.19	99.20	99.20
√	√	√	99.33	99.31	99.28	99.29	94.85	94.67	94.85	94.70	99.45	99.32	99.31	99.32

注：加粗数据表示最优值。

在 NCT-CRC-HE-100K 数据集上，基线模型的分类准确率为 97.42%，精确率为 97.28%，召回率为 97.27%， F_1 分数为 97.28%。通过将均匀采样替换为重要性采样，准确率提升至 97.93% (+0.51 个百分点)，精确率、召回率和 F_1 分数分别增至 97.94%、97.89% 和 97.91%，表明重要性采样能有效选择关键时间步，增强模型对病灶区域等重要信息的关注。在基线模型中加入时序通道注意力机制后，准确率显著提高至 99.23% (+1.81 个百分点)，精确率、召回率和 F_1 分数分别达到 99.21%、99.20% 和 99.20%，显示该机制能充分挖掘不同通道和时间步的特征，进一步提升分类性能。最终，结合重要性采样和时序通道注意力机制的 DT-DPath 模型在 NCT-CRC-HE-100K 上取得最佳表现，准确率为 99.33%，精确率为 99.31%，召回率为 99.28%， F_1 分数为 99.29%。在 BreakHis 数据集上，基线模型的准确率为 91.23%，而引入重要性采样后会增至 91.66% (+0.43 个百分点)，加入时序通道注意力机制后达 93.55% (+2.32 个百分点)，组合两者后最高为 93.85%，精确率、召回率和 F_1 分数同样稳步提升，分别从 91.45%、91.51%、91.48% 增至 93.67%、93.85%、93.70%。在 PCam 数据集上，基线模型准确率为 96.56%，引入重要性采样后增至 97.12% (+0.56 个百分点)，加入时序通道注意力机制后达 99.13% (+2.75 个百分点)，两者组合后最高为 99.45%，精确

率、召回率和 F_1 分数同样稳步提升，分别从 96.75%、96.72%、96.74% 增至 99.32%、99.31%、99.32%。3 个数据集的实验结果一致表明，重要性采样和时序通道注意力机制显著提升了模型性能，且两者具有互补性，充分验证了 DT-DPath 模型设计的合理性和必要性。

为了进一步探究扩散过程中不同去噪阶段特征对分类性能贡献，以及多时序特征聚合的有效性，我们以基线模型提取的原始清晰图像特征 ($t=0$) 为基础，在特征聚合过程中分别仅叠加反向扩散的早期 ($t=600\sim 800$ ，接近正常组织恢复阶段)、中期 ($t=300\sim 500$ ，异常发育过渡阶段) 和晚期 ($t=1\sim 200$ ，恶性特征显现阶段) 有限时间步内的特征，观察各阶段特征对分类性能的增益效果。实验结果如表 4 所示。可以看出，叠加早期特征 (对应正常组织形态恢复) 的增益效果相对有限，甚至在某些数据集上可能导致轻微性能波动，这与早期阶段噪声较高、形态信息较为接近正常组织、判别性较弱的特性相符；叠加中期特征 (对应从正常到异常的发育过渡) 带来了一定正向提升，但幅度相对温和；而叠加晚期特征 (对应恶性转化阶段) 带来的性能提升最为显著，在 3 个数据集上均表现出最强的增益。这表明扩散模型在后期去噪阶段生成的表征，对病理演化中最为关键的恶性形态变化 (如腺体结构的明显破坏、细胞核异型性的显著加重等) 具有最强的捕捉能力。

表 4 原始特征叠加不同去噪阶段单一时间步特征的性能对比

单位:%

Table 4 Performance comparison of original features superimposed with single-timestep features from different denoising stages

unit:%

特征采样范围	NCT-CRC-HE-100K				BreakHis				PCam			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
仅 $t=0$ 原始特征	97.31	97.25	97.24	97.24	91.18	91.19	91.23	91.22	96.59	96.72	96.82	96.81
$t=0+$ 早期特征($t \approx 700$)	96.86	96.85	97.84	97.84	90.88	90.89	90.82	90.83	96.05	96.02	96.02	96.01
$t=0+$ 中期特征($t \approx 400$)	97.22	97.18	97.15	97.16	91.02	91.11	91.08	91.09	97.31	97.34	97.30	97.31
$t=0+$ 晚期特征($t \approx 100$)	99.19	99.19	99.10	99.11	93.95	94.08	94.03	94.07	99.09	99.12	99.09	99.10
$t=0+$ 全部时序特征	99.33	99.31	99.28	99.29	94.85	94.67	94.85	94.70	99.45	99.32	99.31	99.32

注:加粗数据表示最优值。

同时为了探究多时序特征聚合的有效性,我们考察了按重要性采样权重由高到低选取的不同数量关键时间步进行聚合的效果。实验结果如表 5 所示。结果显示,随着聚合时间步数量增加,模型性能在 3 个数据集上均呈现先快速上升后趋于稳定的趋势:聚合 10 个高贡献时间步时提升显著;15~25 个时间步

后边际收益明显减小;达到 30 个时间步时,性能已近乎饱和。这表明通过非均匀重要性采样策略,模型仅需从总时间步中浓缩约 3% 的关键阶段即可充分捕获病理演化轨迹的主要判别信息,显著提升了训练与推理效率,同时验证了时序特征聚合策略的高效性。

表 5 聚合不同数量关键时间步特征的性能对比

单位:%

Table 5 Performance comparison of aggregating features from varying numbers of key timesteps

unit:%

聚合时间步数	NCT-CRC-HE-100K				BreakHis				PCam			
	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1	Acc	Pre	Rec	F_1
1	97.31	97.25	97.24	97.24	91.18	91.19	91.23	91.22	96.59	96.72	96.82	96.81
10	98.42	98.28	98.27	98.28	93.23	93.45	93.51	93.48	97.56	97.75	97.72	97.74
15	98.86	98.84	98.87	98.85	93.92	93.91	93.87	93.90	98.81	98.83	98.81	98.82
25	99.12	99.15	99.12	99.14	94.38	94.38	94.43	94.40	99.08	99.10	99.08	99.10
30	99.33	99.31	99.28	99.29	94.85	94.67	94.85	94.70	99.45	99.32	99.31	99.32

注:加粗数据表示最优值。

2.3 可视化分析

2.3.1 注意力热图可视化分析

为展示时序动态通道注意力机制(TDCA)在 DT-DPath 模型中捕捉病理演化关键中间态的能力,图 4 对比了不同模型生成的注意力热图。该热图通过对 TDCA 模块输出的通道注意力权重进行空间平均与归一化处理得到,用以反映模型对图像不同区域的关注程度。图中颜色越亮表示注意力权重越高,颜色越暗则表示权重越低。

通过与其他比较模型的注意力热图进行对比,本文所提出的 DT-DPath 模型展现出了显著的优势。例如在 NCT-CRC-HE-100 $\times 10^3$ 数据集上,其他模型的注意力热图中亮度最高的区域多集中在左侧红框位置,该区域虽然特征较为显眼,但实际上是一个黏液污染区域。而 DT-DPath 模型在关注到这一污染区域的同时,能够将更多的注意力集中在右侧橙框所在的黏液丝上,这一区域才是真正具有类别判别意义的关键特征所在。这表明 DT-DPath 模型能够更准确地识别和聚焦于对分类有重要意义的组织结构和细胞形态,而不被表面的、不具代表性的特征所干扰。在图 4 第 4 列中,可

以看到 DT-DPath 模型对上侧区域的结直肠癌上皮细胞的轮廓区分更为细致。模型能够捕捉到上皮细胞的细胞核、细胞质以及细胞边界等关键的形态学特征,并将注意力集中在这些区域,从而更准确地进行分类。

2.3.2 t -SNE 可视化分析

为了进一步验证 DT-DPath 模型在特征提取和分类任务中的性能,本文通过 t -SNE (t -distributed Stochastic Neighbor Embedding) 对 DT-DPath 在不同训练阶段提取的分类结果进行了 t -SNE 降维,并绘制了相应的可视化图以观察数据的分布情况。

DT-DPath 模型在 NCT-CRC-HE-100K 数据集上的 t -SNE 可视化结果如图 5 所示,分别对应训练 5 轮、15 轮、25 轮、35 轮、45 轮和 55 轮后的分类结果特征分布。从图 5 可以看出,随着训练的进行,特征点的分布逐渐从分散变为聚类。在训练初期(5 轮,图 5(a))所示,分类结果特征点分布较为分散,不同类别的特征之间存在一定的重叠,这表明模型在训练初期尚未完全学习到数据的内在结构,特征表示能力较弱。随着训练的进行(15~35 轮,图 5(b)和(c))所示,特征点逐渐形成较为明显的聚类,不同类别的特征开始分

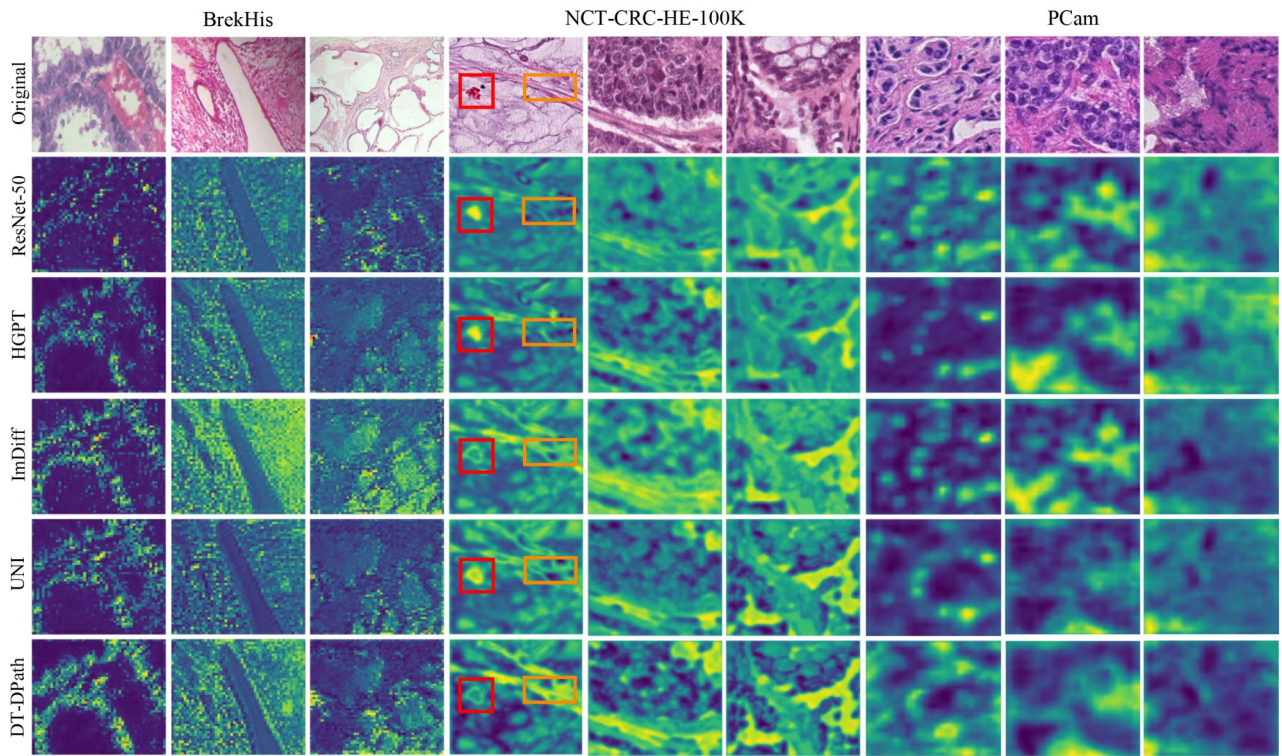


图4 注意力热图可视化

Figure 4 Attention heatmap visualization

离,这表明 DT-DPath 逐渐学会了捕捉关键的形态学特征,并能够有效区分不同病理区域。在训练后期(45 轮至 55 轮,图 5(e)和(f)),特征分布进一步优化,

不同类别的特征点形成了清晰的聚类,几乎没有明显的重叠,这表明模型在训练后期已经具备了较强的特征提取和分类能力。

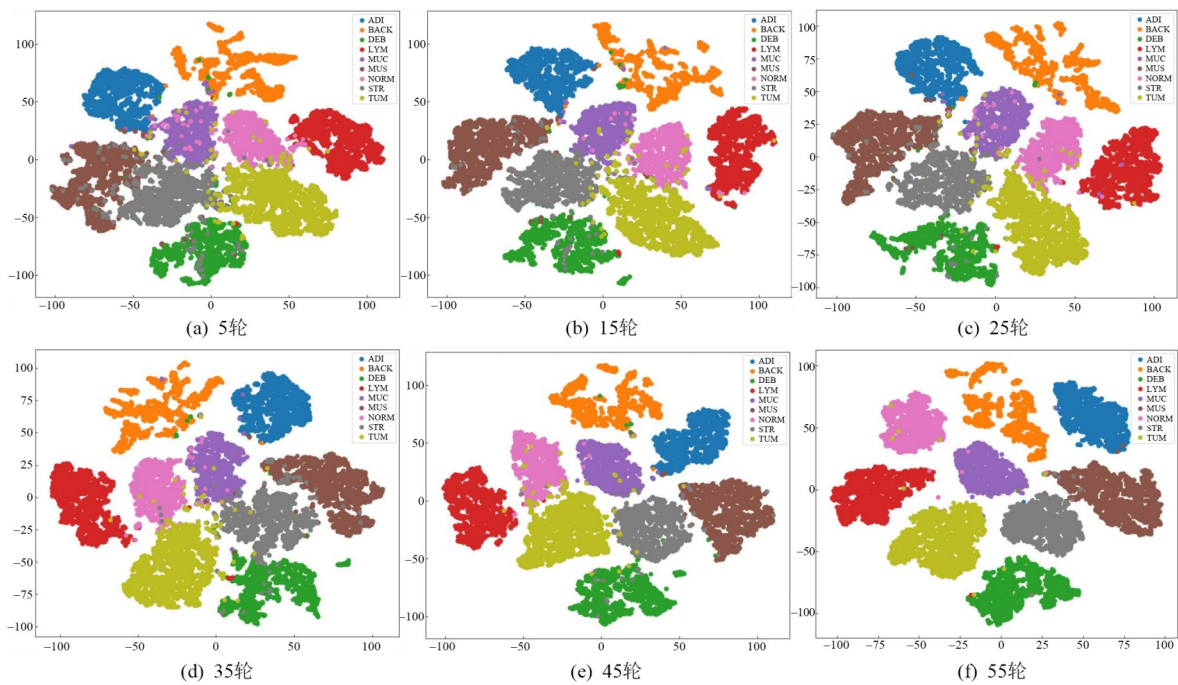


图5 t-SNE 可视化

Figure 5 t-SNE visualization

2.3.3 时序阶段与病理形态注意力演化可视化分析

为了进一步探究所提出 DT-DPath 模型中扩散去噪过程是否能够隐式捕捉病理组织从正常到恶性转化的连续形态学演化规律,我们选取 NCT-CRC-HE-100K 数据集中的一张典型结直肠腺癌图像作为代表性病例,分析其在反向扩散过程中的关键时间步表现及其与真实病理结构变化的对应关系。图 6 展示了该

图像在扩散去噪过程中的 4 个典型阶段(早期 $t \approx 700$ 、中期 $t \approx 400$ 、中后期 $t \approx 250$ 、晚期 $t \approx 100$)的去噪中间图像(上行)及其对应的注意力热图(下行),右侧为所有关键时间步动态特征聚合后的综合热图。随着去噪进程的推进,图像从高噪声的模糊状态逐步恢复清晰的组织细节,同时注意力分布呈现出明显的演化规律。

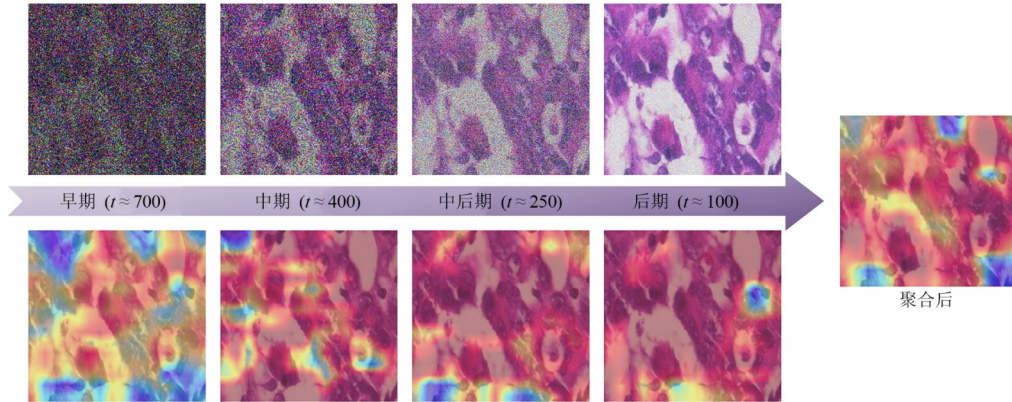


图 6 时期时间步的特征表现与病理形态注意力演化可视化

Figure 6 Visualization of feature manifestations and pathological morphological attention evolution across time steps

从早期阶段开始,注意力较为分散,主要捕捉图像的整体轮廓和大尺度组织分布,类似于病理演化中晚期恶性浸润后组织严重破坏的残余宏观特征。进入中期和中后期,注意力逐步集中到腺体结构的渐进重构、细胞核异型性和不规则腺腔边界,这些区域正是从高级别上皮内瘤变向浸润性腺癌过渡的关键形态学变化点,具有最高的判别信息密度。晚期阶段注意力进一步精炼到细胞核深染、核浆比增大及浸润前沿等细微恶性细节,接近原始图像的真实病理表征。最终聚合后的综合热图清晰地突出了对癌变最具诊断价值的区域,表明 DT-DPath 通过多时间步动态特征融合,能够有效模拟并捕捉病理组织从宏观破坏到微观恶性特征连续演化的过程,从而为分类决策提供更具生物学意义的时序判别信息。

3 结论

本文提出了一种动态时序引导的病理图像分类扩散模型 DT-DPath,该模型以扩散模型为核心,通过模拟组织形态学的退化-重构过程,将去噪中间的多时间步表示转化为病理状态从正常到恶性连续演化的动态表征,从而突破了传统方法对单一时间点静态特征的依赖。DT-DPath 的核心创新在于将扩散去噪过程重新诠释为病理形态演化的生物学模拟,并通过聚合多时间步动态特征构建判别性演化轨迹。同时,引入时序动态通道注意力机制,实现对关键演化阶段的自适应特征增强,并结合非均匀重要性采样策略,

优化训练效率并突出最具判别力的中间态。这些贡献使扩散模型从生成范式扩展到病理分类的判别框架,并在多项基准任务中显著优于现有最先进方法。实验结果表明,DT-DPath 在多个公开病理图像数据集上的分类性能全面领先,可视化与定量分析进一步证实了对病理连续谱系敏感中间态的有效捕捉。本文尝试为扩散模型在医学成像领域的可解释性应用提供一种新的思路,具有理论价值。未来可进一步探索真实纵向病理序列数据下的时序建模,以进一步验证和扩展本方法的生物学意义。

参考文献

- [1] Biller L H, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: A review[J]. JAMA, 2021, 325(7): 669-685.
- [2] Sawicki T, Ruzkowska M, Danielewicz A, et al. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis[J]. Cancers, 2021, 13(9): 2025.
- [3] Center M M, Jemal A, Smith R A, et al. Worldwide variations in colorectal cancer[J]. CA: A Cancer Journal for Clinicians, 2009, 59(6): 366-378.
- [4] Boyle P, Leon M E. Epidemiology of colorectal cancer[J]. British Medical Bulletin, 2002, 64(1): 1-25.
- [5] Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, et al. Colorectal carcinoma: A general overview and future per-

- spectives in colorectal cancer[J]. *International Journal of Molecular Sciences*, 2017, 18(1): 197.
- [6] Dekker E, Tanis P J, Vleugels J L, et al. Colorectal cancer[J]. *The Lancet*, 2019, 394(10207): 1467-1480.
- [7] Sarvamangala D R, Kulkarni R V. Convolutional neural networks in medical image understanding: A survey[J]. *Evolutionary Intelligence*, 2022, 15(1): 1-22.
- [8] Peng Chuncheng, Lee B R. Enhancing colorectal cancer histological image classification using transfer learning and ResNet50 CNN model[C]//*Proceedings of 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*. Piscataway: IEEE, 2023: 36-40.
- [9] Khayatian D, Maleki A, Nasiri H, et al. Histopathology image analysis for gastric cancer detection: A hybrid deep learning and catboost approach[J]. *Multimedia Tools and Applications*, 2025, 84(19): 21777-21803.
- [10] Ben Hamida A, Devanne M, Weber J, et al. Weakly supervised learning using attention gates for colon cancer histopathological image segmentation[J]. *Artificial Intelligence in Medicine*, 2022, 133: 102407.
- [11] Murtaza G, Shuib L, Mujtaba G, et al. Breast cancer multi-classification through deep neural network and hierarchical classification approach[J]. *Multimedia Tools and Applications*, 2020, 79(21): 15481-15511.
- [12] He Kelei, Gan Chen, Li Zhuoyuan, et al. Transformers in medical image analysis[J]. *Intelligent Medicine*, 2023, 3(1): 59-78.
- [13] Shamshad F, Khan S, Zamir S W, et al. Transformers in medical imaging: A survey[J]. *Medical Image Analysis*, 2023, 88: 102802.
- [14] 李广丽, 叶艺源, 吴光庭, 等. 联合多视角Transformer编码与在线融合互学习的乳腺癌病理图像分类模型[J]. *电子学报*, 2024, 52(7): 2369-2381.
Li Guangli, Ye Yiyuan, Wu Guangting, et al. Breast cancer pathological image classification model via combining multi-view transformer coding and online fusion mutual learning[J]. *Acta Electronica Sinica*, 2024, 52(7): 2369-2381. (in Chinese)
- [15] 宋艳涛, 路云里. SwinT-Unet: 基于双通道自注意力机制的超声图像分割方法[J]. *电子学报*, 2024, 52(11): 3835-3846.
Song Yantao, Lu Yunli. SwinT-Unet: Ultrasound image segmentation based on two-channel self-attention mechanism[J]. *Acta Electronica Sinica*, 2024, 52(11): 3835-3846. (in Chinese)
- [16] Kolla B, Venugopal P. An integrated approach for magnification independent breast cancer classification[J]. *Biomedical Signal Processing and Control*, 2024, 88: 105594.
- [17] Liu M X, Liu Y Z, Xu P B, et al. Exploiting geometric features via hierarchical graph pyramid transformer for cancer diagnosis using histopathological images[J]. *IEEE Transactions on Medical Imaging*, 2024, 43(8): 2888-2900.
- [18] Ding Meidan, Qu Aiping, Zhong Haiqin, et al. A transformer-based network for pathology image classification[C]//*Proceedings of 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Piscataway: IEEE, 2021: 2028-2034.
- [19] Kather J N, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study[J]. *PLoS Medicine*, 2019, 16(1): e1002730.
- [20] Spanhol F A, Oliveira L S, Petitjean C, et al. A dataset for breast cancer histopathological image classification[J]. *IEEE Transactions on Biomedical Engineering*, 2016, 63(7): 1455-1462.
- [21] Veeling B S, Linmans J, Winkens J, et al. Rotation equivariant CNNs for digital pathology[C]//*Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention*. Heidelberg: Springer, 2018: 210-218.
- [22] Albashish D, Al-Sayyed R, Abdullah A, et al. Deep CNN model based on VGG16 for breast cancer classification[C]//*Proceedings of 2021 International Conference on Information Technology (ICIT)*. Piscataway: IEEE, 2021: 805-810.
- [23] Zhong Ziliang, Zheng Muhang, Huafeng Mai, et al. Cancer image classification based on DenseNet model[J]. *Journal of Physics: Conference Series*, 2020, 1651(1): 012143.
- [24] Gupta M, Verma N, Sharma N, et al. Deep transfer learning hybrid techniques for precision in breast cancer tumor histopathology classification[J]. *Health Information Science and Systems*, 2025, 13(1): 20.
- [25] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//*Proceedings of the 9th International Conference on Learning Representations*. [S.l.]: ICLR, 2021: 1-21.
- [26] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[C]//*Proceedings of the 38th International Conference on Machine Learning*. [S.l.]: PMLR, 2021: 10347-10357.
- [27] Li Meng. Transformer-based self-supervised learning and

distillation for medical image classification: Improving colorectal cancer detection on NCT-CRC-HE-100K with Swin-T V2[C]//Proceedings of 2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE). Piscataway: IEEE, 2024: 644-648.

- [28] Liu Ze, Hu Han, Lin Yutong, et al. Swin transformer V2: Scaling up capacity and resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 12009-12019.
- [29] Li A C, Prabhudesai M, Duggal S, et al. Your diffusion model is secretly a zero-shot classifier[C]//Proceedings of

2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 2206-2217.

- [30] Bhattacharya M, Huang J, Sher A F, et al. Immunodiff: A diffusion model for immunotherapy response prediction in lung cancer[PP/OL]. V1.arVix (2025-05-29)[2026-01-21]. <https://arxiv.org/abs/2505.23675>.
- [31] Chen R J, Ding Tong, Lu M Y, et al. Towards a general-purpose foundation model for computational pathology[J]. Nature Medicine, 2024, 30(3): 850-862.
- [32] Lu M Y, Chen Bowen, Williamson D F K, et al. A visual-language foundation model for computational pathology[J]. Nature Medicine, 2024, 30(3): 863-874.

作者简介



赵赫阳 男,2004年1月生于山西省太原市。山西大学计算机与信息技术学院硕士研究生。主要研究方向为医学图像处理。
E-mail: zhaohayang1@sxu.edu.cn



张鹏 男,1992年3月出生于山西省朔州市。山西大学计算机学院副研究员、硕士生导师。主要研究方向为计算机视觉、医学影像。
E-mail: pengzhang@sxu.edu.cn



陈如豪 男,2005年3月生于山东省聊城市。现就读于山西大学计算机与信息技术学院。主要研究方向为医学图像处理。
E-mail: 202301006105@email.sxu.edu.cn



王文剑 女,1968年10月出生于山西省太原市。现为山西大学计算智能与中文信息处理教育部重点实验室副主任、教授、博士生导师。主要研究方向为机器学习、数据挖掘。
E-mail: wjwang@sxu.edu.cn



罗一恺 男,2005年5月生于浙江省台州市。现就读于山西大学计算机与信息技术学院。主要研究方向为人工智能与深度学习。
E-mail: luoyikai@sxu.edu.cn