

文章编号 1004-924X(2026)04-0652-19

基于扩散模型的细胞图像与掩膜联合生成数据扩充方法

徐凯歌¹, 孙晓明^{1*}, 陈泳吉¹, 陈言¹, 龙兵¹, 韩永琦¹, 苏伯超², 魏昌宁²

(1. 哈尔滨理工大学 黑龙江省激光光谱技术及应用重点实验室, 黑龙江 哈尔滨 150006;

2. 深圳职业技术大学 未来技术学院, 广东 深圳 518055)

摘要:为了解决细胞图像分割任务中标注数据不足及传统增强方法局限性的问题,本研究首先提出了一种基于扩散模型的细胞图像与掩膜联合生成的数据扩充方法。该方法通过构建由噪声预测 U-Net 与噪声采样器组成的联合生成框架,将细胞图像和掩膜按通道堆叠为四通道联合图像,实现单阶段联合生成高质量合成数据,避免了多阶段生成引入的误差。模型融合时间嵌入与多头自注意力机制,以增强对细胞结构特征与掩膜对应关系的建模能力,并采用均方误差损失函数与余弦退火学习率调度优化训练过程。在 CryoNuSeg 和 ISBI2012 基准数据集上进行的全面实验结果表明,所提出的方法显著提高了下游分割模型的性能。在 DPM++ 2M Karras 采样设置下, CryoNuSeg 的 IoU 与 Dice 分别达到 62.50% 与 75.78%, 优于翻转、旋转等传统增强手段。结果验证了联合生成增强在扩展数据多样性与提升分割精度方面的优越性,为细胞图像分割在标注稀缺场景下提供了高效可行的数据扩充方案。

关键词:细胞图像分割; 扩散模型; 数据扩充; UNet; 联合生成

中图分类号: TP394.1 **文献标识码:** A

doi: 10.37188/OPE.20263404.0652

CSTR: 32169.14.OPE.20263404.0652

A diffusion model based data augmentation method for joint generation of cell images and masks

XU Kaige¹, SUN Xiaoming^{1*}, CHEN Yongji¹, CHEN Yan¹, LONG Bing¹,

HAN Yongqi¹, SU Bochao², WEI Changning²

(1. Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China;

2. Shenzhen Polytechnic University, Shenzhen Polytechnic University, School of future technology, Shenzhen 518055, China)

* Corresponding author, E-mail: sunxiaoming@hrbust.edu.cn

Abstract: To address the issues of insufficient annotated data samples and limitations of traditional augmentation methods in cell image segmentation tasks, this study first proposed a data augmentation ap-

收稿日期: 2025-11-12; **修订日期:** 2025-12-12.

基金项目: 广东省教育厅普通高等教育项目 (No. 2023kcxt077); 深圳市高层次人才科研创业基金 (No. 6023330002k); 广东省综合类高校创新团 6024210083K 队项目 (No. 2020kcxt047); 深圳职业技术大学高校启动基金 (No. 6022312031k); 深圳职业技术大学基金项目 (No. 6023271042K, No. 6024330002K, No. 6024210083K, No. 6025310049K); 黑龙江省外国专家项目 (No. 110000223020238000)

proach based on diffusion models for the joint generation of paired cell images and masks. The method constructed an end-to-end joint generation framework consisting of a noise-prediction U-Net and a dedicated noise sampler. It stacked raw cell images and their corresponding masks into four-channel joint images by channel, enabling single-stage joint generation of high-quality synthetic data and effectively avoiding cumulative errors introduced by multi-stage generation pipelines. The model integrated temporal positional embedding and multi-head self-attention mechanism to enhance the model's ability to model the fine-grained corresponding relationships between cell structural features and semantic masks. It also adopted the mean squared error loss function and cosine annealing learning rate scheduler to optimize the training dynamics and stabilize convergence. Comprehensive experimental results conducted on the CryoNuSeg and ISBI2012 benchmark datasets demonstrate that the proposed method significantly improves the performance of downstream segmentation models. Under the DPM++ 2M Karras sampling setting, the IoU and Dice scores on the CryoNuSeg dataset reach 62.50% and 75.78% respectively, which outperform conventional augmentation techniques such as flipping, rotation, and scaling. The results fully verify the superiority of joint generation augmentation in expanding data diversity and improving segmentation accuracy, providing an efficient, feasible, and generalizable data augmentation solution for cell image segmentation in scenarios with scarce annotations.

Key words: cell image segmentation; diffusion models; data augmentation; UNet; joint generation

1 引言

传统数据增强^[1-2]是指通过对原始医学图像实施一系列简单的几何变换和图像处理操作,生成更多训练样本以优化模型性能。这种方法无需依赖复杂的模型或算法,常用于提升深度学习模型在医学图像分割任务中的泛化能力。常见的传统数据增强技术包括旋转、平移、缩放、裁剪、翻转以及亮度调整等^[3]。这些变换能够显著增加训练数据的多样性,从而有效降低模型过拟合的风险。例如,Zeiser等人^[4]利用水平镜像、缩放和调整大小等数据增强手段,训练U-Net模型对乳腺X光照片中的肿块进行分割,实验结果显示,该方法显著提升了模型的性能。类似地,Manjunath等人^[5]通过数据增强技术训练改进后的ResUNet架构,用于CT图像中的肝脏及其肿瘤分割。此外,调整图像的颜色和光照条件^[6]也是传统数据增强的重要手段。例如,通过添加噪声和模糊处理,可以模拟真实医学图像中常见的噪声干扰和模糊现象,从而增强模型对实际数据的适应能力。Nalepa等人^[7]通过回顾脑肿瘤分割中的数据增强技术,指出噪声添加和高斯模糊等方法在处理有限或不平衡的医疗数据集时尤为重要,并显著提升了模型

在脑肿瘤分割任务中的鲁棒性和精度。此外,集成成像和消串扰技术^[8]广泛应用于3D显示领域,能减少相邻图像元之间的串扰,提升图像清晰度和精度,这些技术对医学图像处理也提供了有益的启示。

尽管传统数据增强方法在医学图像分割中表现优异,但其局限性也明显。这些方法依赖人工设计的变换规则,难以充分模拟医学图像中的复杂变化。因此,面对更具挑战性的分割任务时,传统增强方法的适用性可能受限。

使用混合算法增强数据是指将多种数据增强方法结合起来,通过融合不同的传统和现代算法,提高医学图像分割模型的泛化能力和鲁棒性。与单一的数据增强方法相比,混合算法可以更加灵活地生成多样化的训练样本,从而增强模型对不同类型图像特征的适应性。常见的混合算法包括结合几何变换与噪声处理、深度学习生成模型与传统图像处理技术的融合等。例如,He等人^[9]提出了一种使用生成对抗网络(Generative Adversarial Networks, GAN)恢复肝脏CT图像特征的方法,通过从低质量图像生成高质量图像,视为一种数据增强形式。该方法显著提高了肝脏区域分割的精度,避免了过拟合问题,并在训练3D卷积神经网络(Con-

volutional Neural Network, CNN) 时展现出优势。类似地, Abdelhafiz 等人^[10]开发了一种用于乳腺癌质量分割的卷积神经网络模型, 其训练可能利用了传统数据增强技术如旋转、裁剪和翻转, 从而提升了分割任务的准确度和鲁棒性。研究显示, 这种方法能够有效处理图像中的噪声和不规则性。此外, Verma 等人^[11]提出了一种脑 MRI 图像分割的混合方法, 将核直觉模糊熵 C 均值与直觉粒子群优化算法相结合。该方法利用 IPSO 的全局搜索能力优化 KIFECM 的聚类过程, 从而有效缓解传统聚类算法易陷入局部最优的问题。通过融合全局优化策略与模糊聚类机制, 该方法在真实与模拟脑数据上的分割性能均得到显著提升。另外, Isensee 等人^[12]提出了基于 UNet 的医学图像分割自适应框架 nnU-Net, 包含自动数据增强和其他优化策略。该框架成功应用于肺 CT 图像分割, 通过混合数据增强技术显著提升了模型的准确性和泛化能力, 尤其在处理多种组织类型的图像时表现出色。

使用混合算法进行数据增强已获得一定的成功, 但它也面临一些挑战。首先, 混合方法可能引入相矛盾的信息分布。此外, 不同算法之间的融合可能会带来额外的复杂性, 特别是在选择合适的融合策略时。

生成对抗网络 (GAN)^[13-14] 是一种深度学习方法, 通过生成器和判别器的对抗训练, 能够生成与真实数据相似的虚拟样本。在医学图像分割中, GAN 被广泛用于数据增强, 特别是在数据样本不足的情况下。通过生成逼真的图像, GAN 能够在不增加标注负担的前提下, 提供更多的训练样本, 从而提高分割模型的鲁棒性和泛化能力。利用 GAN 增强数据的方法, 通常通过生成与原始图像风格一致的新图像, 来提高模型对不同形态、不同组织结构的分割能力。例如, Wei X 等人^[15]使用了一种增强的 GAN 和基于掩码区域的 CNN 架构, 用于 CT 图像的自动肝脏分割, 其中包括使用生成的合成数据进行数据增强。这种方法使模型能够学习多种肝脏形态, 从而提高了肝脏分割的精度。同样, Singh V 等人^[16]提出了一种使用生成对抗网络改进乳腺癌数字乳腺 X 线照片分割的数据增强

方法, 展示了 GAN-based 增强在提高分割精度方面的有效性。Mondal 等人^[17]提出了一种基于生成对抗学习 (GAN) 的少样本三维多模态医学图像分割方法, 可视为通过生成模型进行数据增强的一种形式。这种方法通过生成逼真的合成样本, 适用于数据有限的场景, 能够有效提升模型的分割性能。Chen C 等人^[18]提出了一种无监督的双向跨模态适应方法, 使用 GAN, 可应用于心脏 MRI 图像分割。该技术增强了模型处理不同成像模态的能力, 从而改善了分割表现。

基于 GAN 的增强方法在医学图像分割中取得了显著成效, 但仍然存在一些问题。首先, GAN 模型的训练过程可能较为复杂, 需要大量的计算资源。其次, 生成的伪图像质量可能受限于模型的训练效果, 低质量的生成图像可能会影响分割结果。并且 GAN 网络是由一个原输入生成一个输出, 受到原样本的限制, 难以生成大量广泛的数据, 同时 GAN 模型的训练复杂性和图像质量的控制仍然是当前研究的主要挑战。

使用生成对抗网络 (GANs) 和变分自编码器 (Variational Autoencoders, VAEs) 进行数据增强是医学图像分析中的关键策略, 特别适用于处理有限或不平衡的数据集。这些生成模型能够合成与原始数据相似的图像, 从而增加训练数据的多样性和数量。在脑部 MRI 肿瘤分割领域, Li 等人^[19]提出了一种多模态数据增强框架 Tumor-GAN, 利用 GANs 生成合成脑肿瘤图像, 增强训练数据集并提高分割模型的准确性, 从而显著改善脑肿瘤分割模型的性能。对于肺部疾病检测, 使用胸部 CT 和 X 光图像, Al-Sheikh 等人^[20]开发了一种多类深度学习架构, 通过数据增强技术处理肺部图像的变异性, 提升了不同类型肺部疾病的分类准确性。在糖尿病视网膜病变图像分类领域, Chokuwa 等人^[21]利用 VAEs 进行跨域泛化, 帮助模型适应不同来源的视网膜图像变化, 增强了其在现实场景中的鲁棒性和表现。这些研究强调了数据增强在医学图像分析中的重要性, 展示了生成模型如何有效应对数据有限和域变异性挑战。

基于掩膜条件控制的扩散模型增强: 传统扩散模型通常用于生成单一 RGB 图像或灰度

图像。然而,在医学图像分割中,需要同时生成图像和对应的掩膜,许多研究者尝试使用扩散模型进行数据增强,但大多采用两阶段方法:首先利用无条件扩散模型生成掩膜,然后以该掩膜作为条件控制生成图像,或反之。这种两阶段生成方式因掩膜作为中间阶段,可能导致形态畸变,并因引入额外模型而影响生成图像与掩膜的匹配程度。Shao 等人^[22]提出了 Diffuse-Expand 方法,通过训练一个既能分割又能分类医学图像和掩膜的模型作为分类器,利用分类器的输出控制扩散模型进行两阶段生成:首先生成掩膜,然后以掩膜为条件生成图像。在共享参数的扩散模型中,他们生成图像-掩膜对,扩展了 2D 医学图像数据集,从而提升了如 CO-VID-19 和 CGMH Pelvis 数据集上分割模型的性能。Du 等人^[23]训练了一个无条件扩散模型生成掩膜,然后利用其 ArSDM (Adaptive Refinement Semantic Diffusion Model) 以掩膜为条件生成结肠镜图像。该方法能合成高质量的结肠息肉图像,支持下游任务如息肉分割和检测,训练效果可与真实数据相当。Macháček 等人^[24]提出了一种基于掩膜条件控制的潜扩散模型,用于生成胃肠道息肉图像。该方法以生成的分割掩膜为条件,生成高保真的合成息肉图像及其对应的真实掩膜。通过逐步引入噪声并最终生成清晰图像,扩散模型能够有效地模拟和增强图像数据的多样性,这对于医疗领域尤其重要,因为标注数据通常非常有限且获取困难。生成的合成数据不仅帮助提升模型的训练效果,还能减少对大量手动标注的依赖。例如,在医学图像分割中,扩散模型生成的合成数据集可以达到与使用大量手动标注数据集相当的准确性^[25-26]。

扩散模型^[27-28]的一个显著优势是其能够生成多样化的分割结果,并提供像素级的不确定性评估。这种特性允许生成多个可能的分割掩码,从而提高分割的鲁棒性和准确性^[29-30]。这些研究表明,扩散模型在生成合成医学图像和掩膜以进行数据增强方面是有效的。然而,它们均采用两阶段生成过程,可能在图像与掩膜的关系捕捉上存在局限。本文的工作旨在解决这一问题,提出单阶段扩散模型,联合生成图像-掩膜对,确保更好

的对齐并减少形态畸变风险。

为了解决细胞图像分割任务中标注数据不足以及传统数据增强方法的局限性,本文提出了一种基于扩散模型的细胞图像与掩膜联合生成数据扩充方法。本文的主要贡献如下:

(1)提出了一种将细胞图像与掩膜联合生成的扩充方法,通过构建噪声预测 U-Net 与噪声采样器的联合生成框架,利用四通道联合图像的方式,避免了多阶段生成引入的误差,从而实现了一步生成高质量合成数据。

(2)在该模型中,融合了时间嵌入和多头自注意力机制,增强了对细胞特征及掩膜关系的建模能力。通过这种设计,模型能够更准确地捕捉细胞图像与其掩膜之间的复杂关系,提高了生成数据的质量。

(3)在训练过程中,采用均方误差 (Mean Squared Error, MSE) 损失函数与余弦退火学习率调度来优化模型训练,使得模型在数据扩充过程中获得更好的训练稳定性与性能。

(4)通过在 CryoNuSeg 和 ISBI2012 数据集上的实验验证,联合生成增强方法在扩展数据多样性和提升分割精度方面具有明显优势,为医学图像分割任务提供了一种高效且创新的数据扩充解决方案。

2 方 法

本文提出的细胞图像及其掩膜联合生成扩充方法,其核心在于设计一个基于扩散模型的联合生成框架。该框架通过学习原始数据集的分布,生成与原始数据一致的细胞图像及其对应掩膜,从而有效扩充训练数据集,提升细胞分割模型的性能。图 1 为细胞图像与掩膜联合生成增强方法的流程。

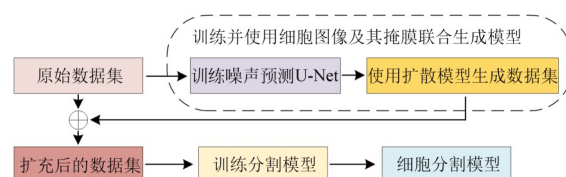


图1 细胞图像与掩膜联合生成增强方法流程

Fig. 1 Cell image and mask joint generation enhancement method flow

如图 1 所示,该框架主要由两部分构成:噪声预测 U-Net 和噪声采样器。噪声预测 U-Net 负责学习原始数据集的特征分布,而噪声采样器则从噪声中逐步恢复出高质量的细胞图像及其掩膜。具体而言,训练过程分为两个阶段:首先,利用原始数据集的训练集对噪声预测 U-Net 进行训练,使其能够准确预测并去除图像中的噪声;其次,通过训练好的噪声预测 U-Net 生成扩充数据集,并将其与原始数据集结合作为增强数据集,用于后续细胞分割模型的训练。

2.1 细胞图像与掩膜联合生成增强架构

细胞图像与掩膜联合生成增强架构通过通道联合生成 RGBA 四通道联合图像,并利用预测 U-Net、时间嵌入架构和多头注意力机制,从噪声中一步生成高质量细胞图像及其掩膜。

该架构首先将掩膜取值范围调整至与图像一致(0~255),掩膜最小值和最大值分别对应 0 和 255。若细胞图像为灰度图像,则将其复制为 R, G, B 三通道,掩膜作为 A 通道(Alpha 透明度通道),通过通道维度的 Concat 操作形成具有 RGBA 四通道的联合图像。训练过程包括正向和反向两个阶段:正向过程从原始联合图像逐步添加噪声,直至变为完全噪声,通常基于马尔科夫链,每一步依赖前一步结果;反向过程从完全噪声图像出发,通过预测 U-Net 逐步去噪,恢复至原始联合图像,时间嵌入架构辅助分辨去噪步骤,多头注意力机制增强特征提取能力。如图 2 所示为细胞图像与掩膜联合生成增强架构,该过程体现了图像与掩膜的内在联系。训练完成后,模型从随机噪声生成 RGBA 四通道联合图像,随后解除联合操作,A 通道直接形成具有高二值化特性的新掩膜,无需额外二值化处理,RGB 通道构成新细胞图像,若为灰度图像则需灰度化处理。新生成的细胞图像与掩膜天然具备分割对应关系,避免了传统方法中先生成图像再生成掩膜或反之的多阶段流程。相较于依赖多个模型的生成方式,该方法通过单个模型一步到位完成生成,减少了额外模型引入的损失。其核心在于学习细胞图像与掩膜的内生逻辑,使生成结果高效且能成对呈现图像与掩膜的分割。

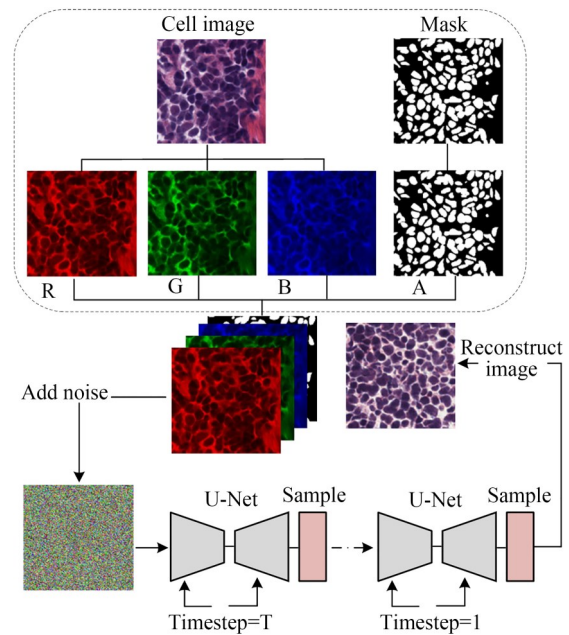


图 2 细胞图像与掩膜联合生成增强架构

Fig. 2 Joint image-mask generation enhancement architecture

2.2 噪声预测 U-Net 架构

在噪声预测 U-Net 框架内采用了一个 4 层的编码器-解码器结构,如图 3 所示。与通常的图像扩散模型生成不同的是,在本研究中自编码器编码解码的对象是细胞图像及其细胞分割掩膜按通道堆叠得到的四通道图像,以此让噪声预测 U-Net 在前向推理时充分利用细胞图像及其分割掩膜的特征相关性,学习其间的高阶特征,也能让噪声预测 U-Net 学习到细胞分割图像和掩膜间

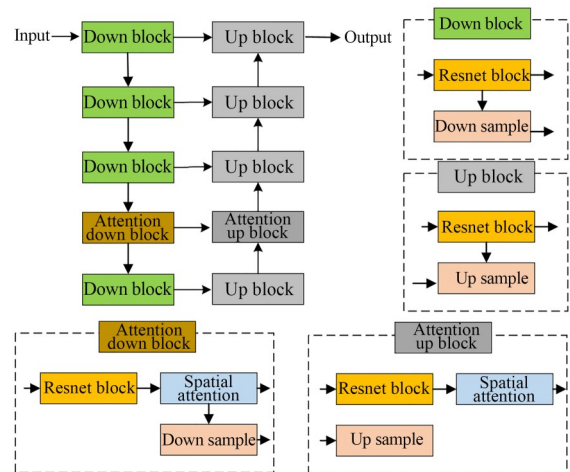


图 3 噪声预测 U-Net 架构

Fig. 3 Noise estimation U-Net

的位置、结构、特征相关性,以便让扩散模型生成的细胞图像及其掩膜是对应的,以及掩膜为细胞图像的分割结果,这是让细胞图像及其细胞掩膜联合生成的关键所在。

该结构编码模块和解码模块包含 resnet 模块、空间注意力模块、上下采样模块,以此更好地提取细胞与其掩膜的分割对应相关信息,增强编码和解码阶段语义特征的提取和融合。每层下采样会将特征图尺寸缩小为原本的 1/2,然后快速连接至解码部分,然后每层上采样会将特征图变为原本的 2 倍。图 3 为噪声预测 U-Net 的具体架构。

2.3 时间嵌入架构

噪声预测 U-Net 通过对比预测的噪声与实际添加的噪声进行训练。在该 U-Net 的残差块中,加入了时间嵌入(Timestep Embeddings),以向网络传递当前图像添加了多少步噪声。时间嵌入采用位置嵌入(Position Embedding)方式,并使用余弦嵌入方法进行编码。

时间嵌入架构在 U-Net 的残差块中起到了增强作用,从而优化和改进噪声预测过程。具体来说,时间嵌入在噪声预测 U-Net 中扮演着关键角色,尤其是在帮助网络识别图像中添加了多少步噪声的情况下。时间嵌入通过位置嵌入实现,采用余弦嵌入方式。这种位置嵌入方法能够将每个时间步的信息转化为高维向量,作为网络的输入,并与其他特征共同用于训练。

时间嵌入与残差块的结合显著提升了网络对图像噪声的预测能力,使得模型能够根据不同的噪声级别做出不同的响应。在本模型中,时间嵌入机制的核心作用是将每个时间步(timesteps)信息转化为高维向量(即时间嵌入),以便帮助模型在每个时间步骤中更有效地进行去噪和图像生成。图 4 展示了时间嵌入在网络中的作用,并与残差块结构相结合。

(1)时间步提取与时间嵌入:在训练过程中,模型首先随机选择一个时间步。每个时间步代表了图像当前的噪声级别。时间步信息会被转换为时间嵌入(Timestep Embedding),它通过一个 Dense Layer(密集层)进行处理,通常采用正弦或余弦编码将时间步信息嵌入到一个高维空间中。这些嵌入向量将作为输入传递给网络的

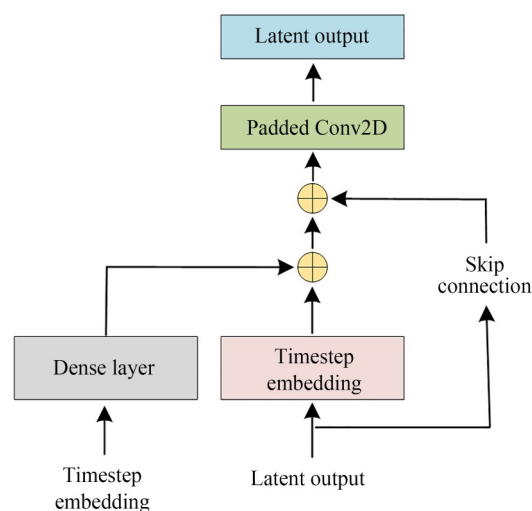


图4 时间步嵌入架构

Fig. 4 Timestep embedding module

其他部分。

(2)时间嵌入的作用:时间嵌入会与模型的主干(如卷积层或残差块)进行连接,提供关于当前噪声水平的信息,帮助模型判断如何去噪。在图 4 中,Timestep Embedding 与 Dense Layer 相连接,通过残差连接(Skip Connection)将其与上层的潜在输出(Latent Output)结合,确保信息流动和有效的特征提取。

(3)在残差块中的应用:残差块通过 Skip Connection 将时间嵌入的输出与其他层的输出结合,形成一个“捷径路径”,从而使模型在学习时更加稳定,避免了梯度消失问题。Timestep Embedding 和 Latent Output 通过这类跳跃连接(Skip Connection)共同影响每个时间步的去噪过程,确保在每个时间步上模型能结合噪声信息和图像的潜在特征进行有效的去噪。

(4)模型输出:经过时间嵌入处理后,模型的输出是通过 paddedConv2D(带 padding 的 2D 卷积层)处理的潜在输出,进一步融合时间步信息来生成去噪图像。

在本架构中,时间嵌入通过与其他层的连接和跳跃连接(Skip Connection)共同作用于残差块,帮助模型根据不同的时间步(噪声水平)进行去噪处理。图 4 中的结构图清晰地展示了时间嵌入如何通过 Dense Layer 生成嵌入向量,并与其他网络层(如卷积层和潜在输出)结合,形成有效的去噪路径。

2.4 多头自注意力模块

多头自注意力机制(Multi-Head Self-Attention)是自注意力机制的一种扩展,旨在通过多个注意力头并行地学习输入数据的不同表示,从而增强模型的表达能力。本文使用的多头注意力机制用在了网络的下采样和上采样部分主要目的是:在自然语言处理(Natural Language Processing, NLP)和计算机视觉(Computer Vision, CV)等领域,Transformer架构广泛采用多头自注意力机制,以实现对不同部分信息的捕获和组合^[31]。下面详细描述多头自注意力模块的结构和计算过程。

在多头自注意力机制中,首先需要对输入序列进行线性变换,将输入序列映射到多个子空间。给定输入序列 X ,每个注意力头都有一组独立的权重矩阵,分别用于生成查询(Query)、键(Key)和值(Value)向量。

计算单个头的注意力时,对于每个注意力头 i ,首先计算查询向量 Q_i 和键向量 K_i 之间的相似性。通常采用点积操作来度量查询和键之间的相似度,然后进行缩放处理。具体计算如式(1):

$$A_i = \text{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right), \quad (1)$$

其中: A_i 是第 i 个注意力头的注意力权重矩阵, d_k 是键向量的维度。Softmax函数用于将点积结果转换为概率分布,表示各个位置的重要性。使用这些注意力权重对值向量 V_i 进行加权求和,得到每个头的输出:

$$\text{Attention}_i = A_i V_i. \quad (2)$$

多头自注意力机制的关键特点是并行计算多个注意力头,每个头在不同的子空间中独立学习并生成其对应的输出。假设有 h 个注意力头,最终的输出是所有注意力头的输出的拼接。具体来说,将所有 h 个头的输出拼接起来后,应用一个线性变换矩阵得到最终的多头自注意力输出。

2.5 损失函数设计

基于均方误差(Mean Squared Error, MSE)损失的计算可以分为两种情况:一种是预测噪声(ϵ),它是典型的噪声预测模型。在扩散模型中,通常希望模型能够准确地预测在加噪声过程中引入的噪声部分;另一种是预测图像样本“sample”,直接对图像进行预测。预测噪声“ep-

silon”的情况下,模型的输出是预测的噪声,而目标是实际的噪声,如公式(3);预测+样本“sample”的情况下计算的是样本预测的均方误差损失,如公式(4):

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|\hat{\epsilon}_i - \epsilon_i\|_2^2, \quad (3)$$

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i - x_i\|_2^2, \quad (4)$$

其中: $\hat{\epsilon}_i$ 是模型预测的噪声, ϵ_i 是实际噪声, N 是批次大小; $\hat{x}_i$ 是模型预测的样本图像, x_i 为实际的干净图像。公式(5)采用信噪比(Signal-to-Noise Ratio, SNR)加权。加权方式基于目标图像的加权比例:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \text{snr_weights}_i \|\hat{\epsilon}_i - \epsilon_i\|_2^2, \quad (5)$$

$$\text{snr_weights}_i = \frac{\alpha_i}{1 - \alpha_i}, \quad (6)$$

其中, snr_weights_i 是通过加权系数 α_i 计算得到的信噪比(SNR)权重。

2.6 生成增强训练方法

训练增强模型时采样的扩散模型采样方法(或者称为调度器)为DDPM(Denoising Diffusion Probabilistic Models, DDPM)^[32],但DDPM所需的采样步数过长,需要的采样步数为1 000步,消耗时间过长,所以采取所需采样步数较少但生成的图像效果较好的两种采样方法DDIM^[33]和DPM++ 2M Karras^[34],这两种方法仅需要50步采样就能采样生成高质量的图像,其他的大部分采样方法存在生成图像质量不高或者生成崩坏问题。DDIM是一种简单但效果不错的采样方法。DPM++ 2M Karras属于DPM-Solver++采样方法的一种,优势是效果好且多样。

训练分割模型将使用原始训练集和扩充训练集进行增强训练,扩充数据集与原数据集合起来即为增强数据集(Augmented Dataset),因此增强分割数据集采用如图5所示格式。细胞图像及其掩膜生成增强模型使用原始数据集的训练集进行训练,而不使用验证集和测试集的任何数据,避免间接获取验证和测试数据特征对之后的验证和测试造成影响,同时训练集、验证集、测试集的图像是先划分好再裁剪,避免图像来源交叉对验证和测试造成影响。

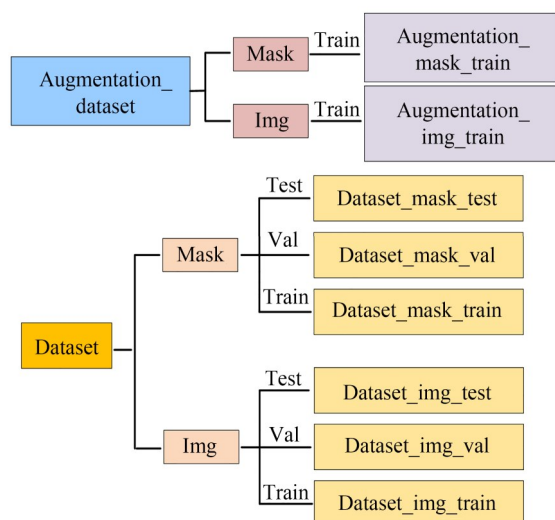


图5 增强数据集配置

Fig. 5 Augmented segmentation dataset configuration

训练细胞图像及其掩膜生成增强模型完成后,设定采样方法生成扩充数据集(Augmentation dataset),生成的图像和掩膜均一一成对,并且扩充数据集仅有训练集,用于增强数据集的训练,而不影响验证集和测试集的正常验证和测试。增强数据集流程如图6所示。

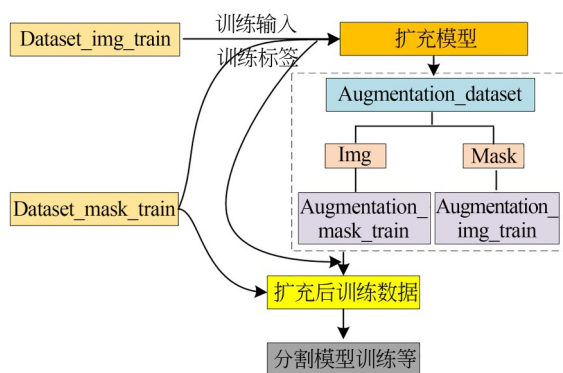


图6 扩充数据集流程

Fig. 6 Augmentation segmentation dataset flow

作为基准的U-Net包含5层对称的编码器-解码器,初始通道数配置为{16, 32, 64, 128, 256},解码阶段采用双线性插值进行2倍上采样,通过跳跃连接将对应编码器层的高分辨率特征与深层语义特征进行通道拼接,后接相同的 3×3 卷积模块进行特征融合,使用零填充保持特征图空间分辨率。该基础U-Net的参数数量约为1.2 M(与具体输入尺寸相关),在保证分割性能的同时

具备较高的计算效率,适合在本实验中作为数据扩充消融和对比的基准模型。

3 实验过程及结果分析

3.1 数据集及图像预处理

本文使用CryoNuSeg^[35]数据集和ISBI2012数据集^[36]来验证所提方法的有效性和先进性。

3.1.1 CryoNuSeg数据集

CryoNuSeg数据集由维也纳医科大学团队于2021年发布,旨在解决术中病理快速制备引起的细胞核分割问题,如冰晶伪影、染色质低对比度及核边界断裂等。该数据集包含30张冰冻切片图像(分辨率 512×512 pixels),覆盖甲状腺、乳腺、胃等10种器官,共约30 000个细胞核实例。数据集提供三种独立标注版本,包括重复标注和第三方校验,并附有二值掩码、距离图等辅助数据用于模型训练。

使用其中的mask binary标注作为训练集标签,且不强制为重叠细胞边缘添加缝隙。本文把30张图像按6:2:2比例划分为训练集、验证集和测试集,随后将图像裁剪为分辨率 256×256 pixels的图像块。裁剪后,得到72张训练集图像、24张验证集图像和24张测试集图像。

3.1.2 ISBI2012数据集

ISBI2012数据集由2012年IEEE生物医学成像国际研讨会发布,旨在为神经元膜(属于神经元细胞)分割任务提供基准测试。数据集基于果蝇一龄幼虫腹侧神经索的透射电镜切片图像,包含30张训练图像(分辨率 512×512 pixels,空间分辨率 $4 \times 4 \times 50$ nm/ pixel)及专家标注的二值标签,以及30张测试图像。该数据集的挑战性在于低信噪比、复杂的神经元膜拓扑和亚像素级的边界分离需求。所有标注由领域专家手动完成并经过一致性校验,成为早期U-Net模型优化的关键基准。

数据集为灰度图像,包含30张训练集和30张测试集图像,由于没有验证集,本文把按4:1比例从训练集中划分部分图像作为验证集。裁剪后,得到96张训练集图像、24张验证集图像和120张测试集图像。

3.2 训练策略与参数设置

本文提出的细胞图像及其掩膜生成增强模

型是在 Linux 系统上使用 Python 编程语言编码的,并以 Pytorch 为深度学习框架进行开发。实验过程中,本文使用 Visual Studio Code 作为主要的编程环境。训练设备配备 24 GB 的内存,CPU 为 Xeon(R) Platinum 8362,主频为 2.80 GHz, GPU 为 RTX 3090。使用 Adam 优化器,动量为 0.9。训练策略为余弦退火学习,以平滑训练过程。在余弦退火策略中,学习率以余弦函数的形式减小,这能够确保学习率减小更平滑,防止由于学习率下降得太快而使得模型无法正常收敛。初始学习率为 0.000 1。批量大小设置为 16。本文每个分割模型训练 100 个 epochs。

3.3 评价指标

本文以 U-Net 作为基准对比了基于扩散模型的细胞图像与掩膜联合生成扩充方法与传统扩充方法的性能。

在医学图像上训练的扩散模型,无法使用一般的图像质量评估方法进行评估,如 FID (Fréchet Inception Distance)^[37] 分数或 Lpips (Learned Perceptual Image Patch Similarity)^[38],均是使用在自然图像上训练得到的预训练模型的特征层输出评估原始图像和生成的图像的相似度,但通常所使用的自然图像如 ImageNet 等

绝大部分数据集不包括医学图像数据,尤其是细胞图像数据,在这些自然图像数据集训练得到的预训练模型无法评估细胞图像。既然难以评估细胞图像与原图像的指标,因此直接使用模型扩充增强产生的效果来对增强方法进行评估。

实验评价指标包括 IoU, Dice 系数、参数数量以及计算复杂性,本文由于是用于衡量增强方法带来的精度提升,不涉及参数数量以及计算复杂性对比,因此选用 IoU, Dice 来对精度进行衡量。其中 TP 代表实际为正类且识别为正类,FP 代表实际负类且为识别正类, FN 代表实际正类但为识别负类, TN 代表实际负类且识别负类。

则 IoU 和 Dice 计算公式如方程(7)和(8)所示:

$$IoU = \frac{TP}{TP + FN + FP}, \quad (7)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP}. \quad (8)$$

3.4 对比实验结果及分析

为了证明本文所提出的细胞图像和掩膜联合生成对分割模型训练的增强效果,与传统数据增强方法如翻转、旋转等在基准模型上进行训练对比,评估不同方法的指标,如表 1 所示。

表 1 对比实验的分析结果

Tab. 1 Comparative experiment analysis results

(%)

Augmentation	CryoNuSeg		ISBI2012	
	IoU	Dice	IoU	Dice
Org	56.24	70.50	78.56	87.83
Org+ flip	57.97	72.51	82.90	90.46
Org+ rotate	55.72	71.01	79.85	88.31
Org+ flip+rotate	58.26	72.62	81.56	89.42
Diffusion Model-based data augmentation ^[39]	—	—	83.10	89.97
GAN-based data augmentation ^[40]	60.80	75.60	79.58	88.47
Org+ Augmentation_x2(DPM++)	62.50	75.78	85.25	91.85

实验结果表明,不同的数据增强策略对细胞图像分割性能存在显著差异。在控制扩充数据量不超过原始数据集两倍的约束下,传统几何变换方法中,水平/垂直翻转(Flip)增强效果显著优于旋转(Rotate)。以 ISBI2012 数据集为例,Flip 增强使 IoU 和 Dice 分别提升 4.34% 和 2.63% (78.56%→82.90%, 87.83%→90.46%),而 Ro-

tate 仅带来 1.29% 和 0.48% 的微弱提升。这种现象可能源于旋转操作引入的细胞结构形变与真实生物形态存在偏差,而翻转操作更符合细胞的空间对称特性。

特别的,多种增强方式的混合使用并未产生协同效应。在 ISBI2012 数据集上,Flip+Rotate 组合的 IoU (81.56%) 反而低于单一 Flip 增强效

果,Dice系数亦下降1.04%。这暗示传统增强方法的组合可能引发过拟合或语义失真,尤其在细胞边界模糊区域。然而在CryoNuSeg数据集上,混合增强使IoU和Dice分别提升2.02%和2.12%(56.24%→58.26%,70.50%→72.62%),说明数据分布差异对增强策略敏感性存在影响。

本文采用基于扩散模型的数据增强方法^[39],在ISBI2012数据集上实现了显著的性能提升。具体而言,IoU提高了4.54%(从78.56%提升至83.10%),Dice系数也提升了2.14%(从87.83%提升至89.97%)。然而,由于CryoNuSeg数据集的模型输入为灰度图像,而该数据集为RGB图像,二者在输入格式上的不匹配导致无法直接进行实验。另一方面,基于生成对抗网络(GAN)的增强方法^[40]。通过生成器与判别器的对抗训练生成新的图像样本,能够有效捕捉图像中的复杂细节,并提升模型对细胞结构的处理能力。在CryoNuSeg数据集上的实验结果表明,GAN增强方法使IoU提高了4.56%(从56.24%提升至60.80%),Dice系数提高了5.10%(从70.50%提升至75.60%),进一步验证了其在数据增强中的有效性。

为了直观观察到本文提出细胞图像和掩膜联合生成增强方法Org+Augmentation_x2(DPM++)与其他传统增强方法对基准模型分割效果提升的对比,其中Augmentation_x2是一个扩充后的数据集,其大小是原始训练集的两倍。图7和图8展示了Org+Augmentation_x2(DPM++)与其他增强方法训练得到的基准模型在测试集上的分割图对比,其中图7是CryoNuSeg作为原始数据集的对比,图8是ISBI作为原始数据集的对比。

第一行和第二行分别展示了输入的图像(Input)和作为正确分割结果的掩膜(Group Truth, GT),第三行到第六行为其他增强方法训练得到的基准模型的输出结果,最后一行是本文设计的增强方法训练得到的基准模型输出结果。

从对比图分析,在ISBI2012数据集上作为原始数据集进行对比发现,Org,Org+Rotate,Org+Flip+Rotate增强方法训练得到的基准模型推理的分割结果有不少存在大片将大块应为图像细胞的区域分割成背景的情况,使得分割结果出现

大片分割错误的黑色块,在CryoNuSeg数据集上也会出现更多把图像细胞分割成背景的情况,但Flip增强与本文提出的联合生成增强方法则不会明显出现该问题,且本文提出的联合生成增强方法效果更好。从对比图可以直观地看出在CryoNuSeg数据集和ISBI2012数据集上,使用细胞图像及其掩膜联合生成增强方法增强后相较于增强方法,与Ground Truth更为接近,可知其数据集扩充增强效果优于其他增强方法。

从对比数据分析,本文提出的细胞图像及其掩膜联合生成增强方法(Augmentation_x2,DPM++ 2M Karras采样方法二倍增强)展现出明显优势。在CryoNuSeg数据集上,其IoU和Dice分别达到62.50%和75.78%,相较于传统最佳组合(Flip+Rotate)提升4.24%和3.16%。

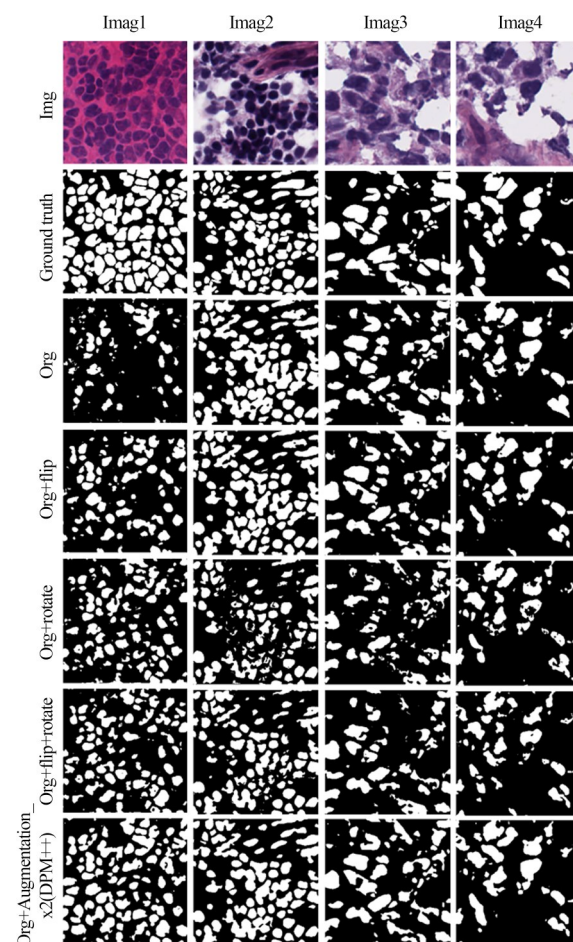


图7 联合扩充增强方法和其他增强方法在CryoNuSeg数据集的对比图

Fig. 7 Comparison of the joint augmentation and other augmentation on CryoNuSeg dataset

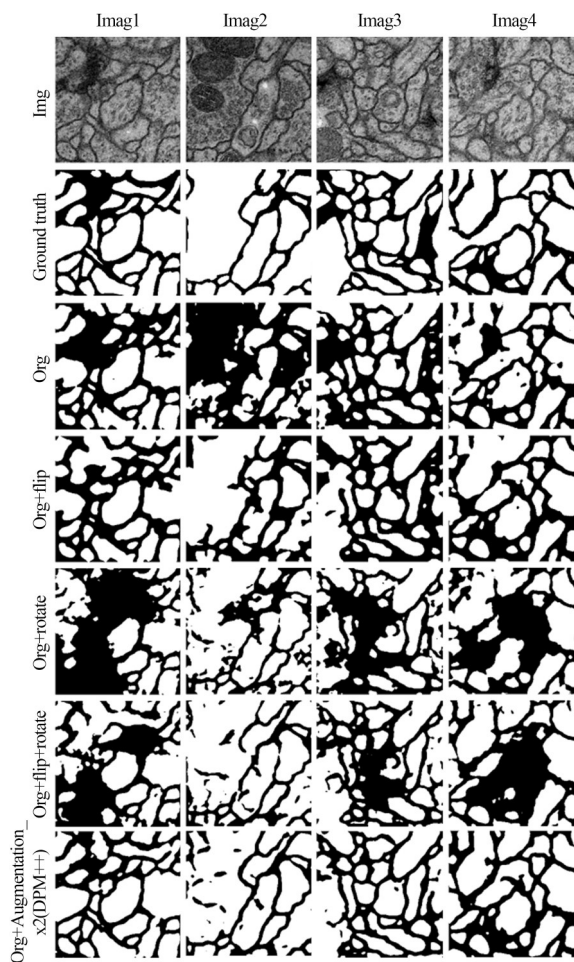


图8 联合扩充增强方法和其他增强方法在ISBI2012数据集的对比图

Fig. 8 Comparison of the joint augmentation and other augmentation on ISBI2012 dataset

在ISBI2012数据集上亦实现85.25%的IoU和91.85%的Dice,验证了该方法在保持细胞结构一致性的同时扩展形态学特征的优越性。实验数据证明,针对细胞图像特性设计的增强策略相比通用几何变换,能更有效地提升分割网络对不同形态细胞以及多样性场景的表征能力。即便通过生成增强的方法得到的是现实不存在的细胞图像,也可能学习到了不存在的特征,但模型学习允许冗余,在学习多样化的特征方面对模型的提升更大。

为了生成高质量的联合数据,本文将细胞图像与对应的掩膜图像按通道堆叠,生成了四通道联合图像,并使用该数据集进行训练。图9展示了四通道数据和联合生成的高质量合成数据(彩图见期刊电子版)。每一行的第一列为原

始细胞图像,接下来的三列分别展示了原始图像在RGB三个颜色通道,红色R、绿色G和蓝色B的分离效果。第五列A为原始数据的Mask,第六列Joint Generation展示了通过本文方法联合生成的图像,将RGB通道和掩膜信息结合,生成了包含更多细节和增强特征的合成图像,最后一列Generation Mask为联合生成图像的Mask。这些图示清晰地展示了RGB通道与掩膜通道的分离,以及联合生成的图像的细节效果。

为了验证本文提出联合生成的数据扩充方法的有效性,对于扩充前后的数据集本文使用了UNet, UNet++^[41], FCN^[42], Deeplabv3_Resnet50^[43]和Deeplabv3_Resnet101^[43]这些模型作为对比的分割模型,详细结果如表2所示。该表展示了不同模型在原始数据集CryoNuSeg和ISBI2012与扩充后的数据集CryoNuSeg_Augmented_x2和ISBI2012_Augmented_x2上的性能表现。通过数据扩充,所有模型的性能得到了显著提升。在CryoNuSeg和ISBI2012数据集及其扩充版本上,所有模型的分割精度均有显著提升,特别是在IoU和Dice系数方面。具体而言,UNet模型在CryoNuSeg数据集上的IoU由56.24%提高至62.50%,Dice由70.50%提升至75.78%,分别提升了6.26%和5.28%。对于UNet++,其IoU从68.40%提升至76.28%,Dice从80.85%提升至86.41%,IoU提升了7.88%,Dice提升了5.56%。FCN模型表现尤为突出, IoU从66.71%提升至79.58%,Dice从79.41%提升至88.46%,其中IoU提升了12.87%,Dice提升了9.05%。Deeplabv3_Resnet50和Deeplabv3_Resnet101在CryoNuSeg数据集上的IoU和Dice分别也有不同程度的提升,前者IoU从64.65%提升至73.31%,Dice从77.85%提升至84.41%;后者IoU从68.24%提升至73.13%,Dice从80.85%提升至84.31%。尽管提升幅度相对较小,但依然显示出数据扩充的有效性。在ISBI2012数据集上,UNet模型的IoU从78.56%提升至85.25%,Dice从87.83%提升至91.85%,分别提升了6.69%和4.02%。UNet++的IoU从85.34%提升至88.66%,Dice从90.71%提升至92.81%,虽然提升幅度相对较

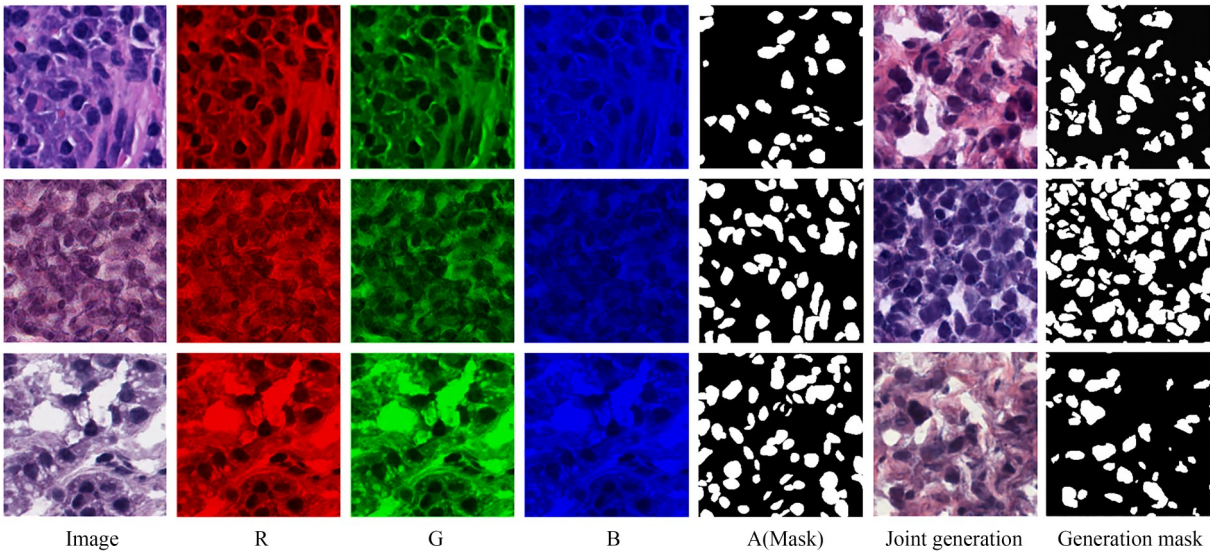


图 9 四通道数据与联合生成的高质量合成数据

Fig. 9 Four-channel data and jointly generated high-quality synthetic data

表 2 不同分割模型在扩充前后的数据集的分割结果

Tab. 2 Segmentation results of different models on the original and augmented datasets (%)

Model	CryoNuSeg		ISBI2012		CryoNuSeg_ Augmented_x2		ISBI2012_ Augmented_x2	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
UNet	56.24	70.50	78.56	87.83	62.50	75.78	85.25	91.85
UNet++ ^[41]	68.40	80.85	85.34	90.71	76.28	86.41	86.63	92.81
FCN ^[42]	66.71	79.41	85.07	91.86	79.58	88.46	86.05	92.48
Deeplabv3_Resnet50 ^[43]	64.65	77.85	83.66	91.06	73.31	84.41	86.40	92.64
Deeplabv3_Resnet101 ^[43]	68.24	80.85	82.86	90.19	73.13	84.31	84.91	91.77

小,但依然具有较好的提升效果。对于 FCN, IoU 由 85.07% 提升至 88.05%,Dice 由 91.86% 提升至 92.48%,表现出平稳的提升。Deeplabv3_Resnet50 和 Deeplabv3_Resnet101 分别在 IoU 和 Dice 方面取得了 2%~4% 的提升,前者 IoU 从 83.66% 提升至 86.40%,Dice 从 91.06% 提升至 92.64%;后者 IoU 从 82.86% 提升至 86.40%,Dice 从 90.19% 提升至 91.77%。

总体而言,数据扩充方法在所有模型上均显著提高了分割精度,特别是在 IoU 和 Dice 指标上,证明了扩充数据集对于提高模型性能的有效性,尤其在 CryoNuSeg 数据集上,性能提升幅度较大。

3.5 消融实验结果及分析

3.5.1 不同采样方法的消融实验

本文基于细胞及其掩膜联合扩散架构构建

增强数据集对分割模型进行训练。为了证明该增强方法可以提升分割模型的分割精确度能力,并且验证原训练集数据的作用,本节对增强方法进行消融实验。首先使用细胞图像及其掩膜生成增强模型生成扩充数据集,扩充数据集的数据量与原训练集成比例,生成不同比例的扩充数据集,与原数据集作为增强数据集。评估不同比例的扩充数据集,以及不使用原始训练集的情况下对分割模型训练精度造成的影响。不同采样方式扩充的图像以及原数据集图像对比如图 10 所示。

每个消融实验和对比实验都采用 CryoNuSeg 数据集和 ISBI2012 数据集进行训练和测试,训练得到 CryoNuSeg 的细胞及其掩膜联合增强模型,和 ISBI2012 的细胞及其掩膜联合增强模型,这两个模型分别生成两种数据集的扩充数据

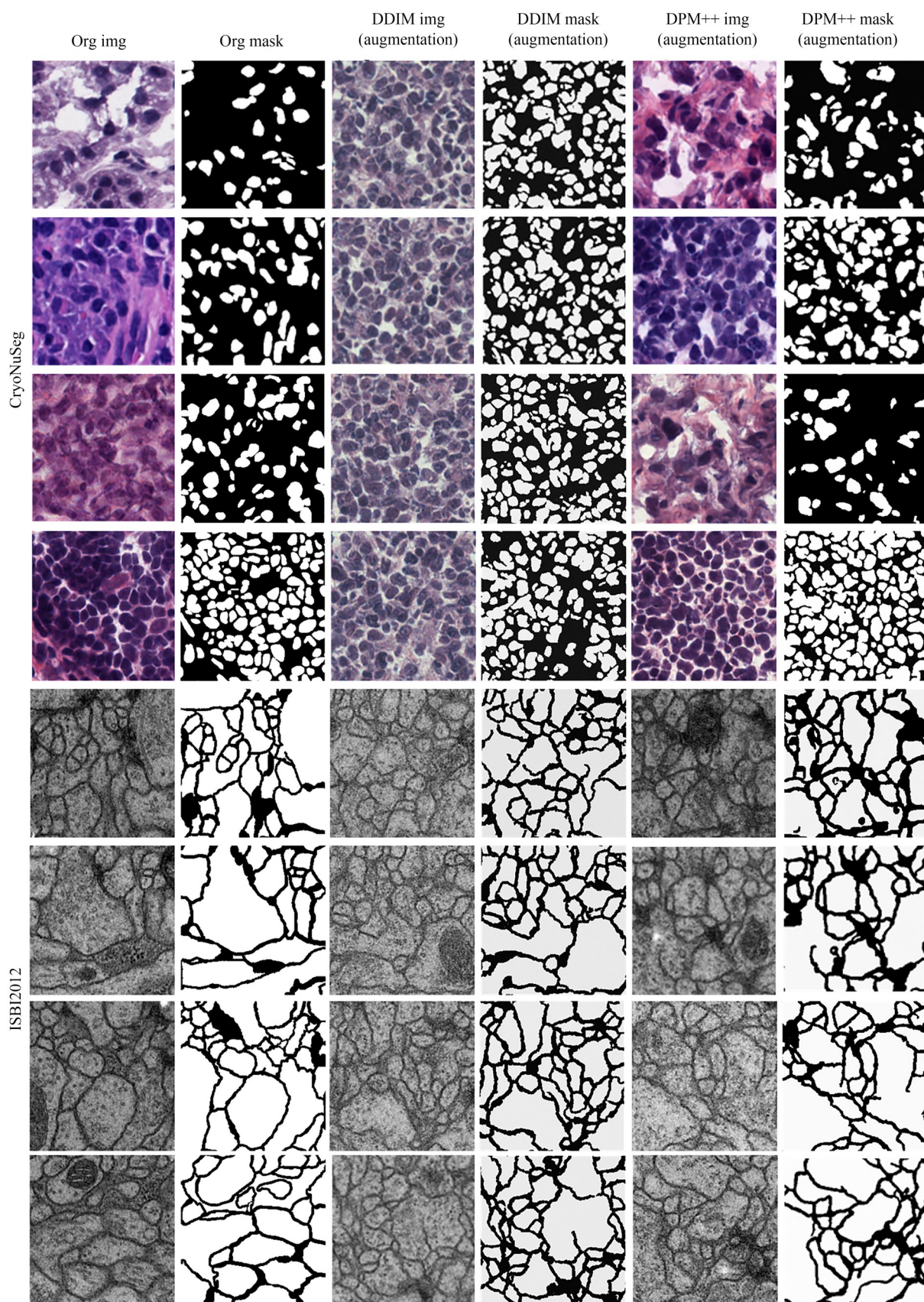


图 10 不同采样方法生成的细胞图像及掩膜对比

Fig. 10 Comparison of cell images and masks generated by different sampling methods

集独立测试,在整个消融实验中每个数据集的细胞及其掩膜联合增强模型仅训练一次。并且训练完成的细胞及其掩膜联合增强模型均使用 DDIM 和 DPM++ 2M Karras 两种采样方法分别进行采样生成的细胞图像及掩膜在分割模型上训练测试。

可以发现 DDIM 采样方法虽然提供了更多不同的细胞形态,生成的细胞及掩膜在色彩与密度方面过于同质化,这个问题在 ISBI2012 上,由于该数据集为灰度图像没有颜色的影响,采样的图像正常,所以 DDIM 采样方法相较于 RGB 细胞图像数据集来说在灰度图像数据集上表现更好。而 DPM++ 2M Karras 采样方法则展现出更好的多样性,生成的细胞图像及掩膜涵盖了原始训练集中所可能的各种颜色分布和密度分布,其在 RGB 细胞图像数据集的表现和灰度图像数据集的表现都不错。

为了验证细胞及其掩膜联合数据扩充方法对于分割模型训练精度提升的有效性,设计消融实验在基线 U-Net 上进行测试。

对于按不同比例进行扩充并与原始数据集混合成的增强数据集进行增强训练的实验,首先

生成 1 倍于原始训练集的扩充数据集称为 Augmentation_x1,然后按照 Augmentation_x1 的扩充次序继续扩充至 2 倍于原始训练集的扩充数据集称为 Augmentation_x2 以此类推直到 Augmentation_x4 为止,与原始训练集 Org 混合则成为增强训练集。对两种采样方法都进行这一过程。

表 3 与表 4 的消融实验揭示了扩充数据规模、采样方法与模型性能间的复杂关系。实验表明:

(1)原始数据与增强数据的协同作用:当单独使用 DDIM 采样的 Augmentation_x1 数据时, CryoNuSeg 数据集 IoU 下降 7.78% (56.24%→48.46%), Dice 下降 7.65% (70.50%→62.85%),证明脱离原始数据的纯生成式增强会导致特征分布偏移。而将 Augmentation_x1 与 Org 数据混合后,DDIM 与 DPM 方法 IoU 分别提升 2.55% 和 4.57%,说明原始数据对维持生成数据的生物学合理性具有关键锚定作用。

(2)采样方法对增强效果的敏感性: DPM++ 2M Karras(以下简称 DPM++)相比 DDIM 展现出更强的鲁棒性。在 Augmentation_x1 独立增强时,DPM++ 使 CryoNuSeg 的

表 3 消融实验的分析结果

Tab. 3 Analysis results of ablation experiments

(%)

Augmentation	CryoNuSeg		ISBI2012	
	IoU	Dice	IoU	Dice
Org	56.24	70.50	78.56	87.83
Augmentation_x1(DDIM)	48.46	62.85	78.26	87.65
Augmentation_x1(DPM++ 2M Karras)	57.33	71.01	82.74	90.41
Org+ Augmentation_x1(DDIM)	58.79	72.95	83.72	90.98
Org+ Augmentation_x1(DPM++ 2M Karras)	60.81	74.66	83.54	90.85

表 4 不同扩充尺度的消融实验

Tab. 4 Ablation study on different augmentationment scales

(%)

Augmentation	CryoNuSeg				ISBI2012			
	IoU		Dice		IoU		Dice	
	DDIM	DPM++	DDIM	DPM++	DDIM	DPM++	DDIM	DPM++
Org	56.24		70.50		78.56		87.83	
Org+ Augmentation_x1	58.79	60.81	72.95	74.66	83.72	83.54	90.98	90.85
Org+ Augmentation_x2	59.13	62.50	73.19	75.78	84.61	85.25	91.51	91.85
Org+ Augmentation_x3	57.91	63.70	72.01	76.64	84.55	85.71	91.48	92.14
Org+ Augmentation_x4	64.55	64.16	77.08	76.84	84.76	86.40	91.61	92.55

IoU 达到 57.33%, 接近原始数据水平 (56.24%), 而 DDIM 则显著退化至 48.46%。这种差异可能源于采样自 DPM++ 的增强图像具有更广泛的多样性和更好的细胞结构。当增强规模扩大至 Augmentation_x4 时, DPM++ 在 ISBI2012 数据集上的 IoU 提升至 86.40% (+7.84% vs Org), 显著优于 DDIM 的 84.76% (+6.20% vs Org), 验证了 DPM++ 在高阶增强任务中的优势, 而仅在灰度图像的细胞任务中被 DDIM 采样方法微小超越 (被超越不足 0.2%)。当然花费更多时间生成扩充数据集以大量的数据扩充直到达到数据扩充效果的边界也是一种以时间换取训练数据量的可行策略。

如图 11 所示, 随着增强倍数增加, 模型精度呈现先升后稳的趋势。以 dpm 在 CryoNuSeg 的 Dice 系数为例, Augmentation_x2 时达到 75.78% (+5.28% vs Org), Augmentation_x3 时增至 76.64% (+0.86% 增幅), Augmentation_x4 时则回落至 76.84%, 增幅趋近于零。这表明 2 倍增强已能覆盖主要特征空间, 而更高倍数的增强可能引入冗余或噪声样本。值得注意的是, Augmentation_x4 在 ISBI2012 上仍保持 Dice 系数 92.55% 的峰值, 说明数据需求规模与数据集固有复杂度正相关。

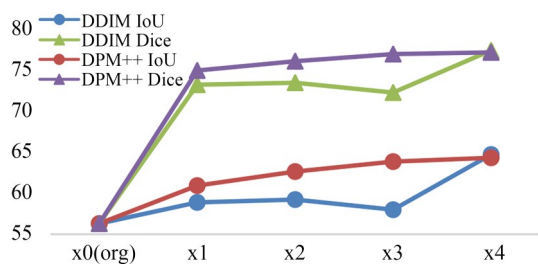


图 11 不同增强倍数的指标折线图

Fig. 11 Line charts of metrics across different augmentation factors

基于实验, 本文推荐采用 Org + Augmentation_x2 (DPM++) 作为均衡精度与计算成本的优选方案, 其在 CryoNuSeg 与 ISBI2012 上的平均 IoU 提升分别为 6.26% 和 6.69%, 而扩充和训练时间约为 Augmentation_x4 的一半。对于高分辨率细胞图像分析场景, 若计算资源充足, 可优先扩展 DPM++ 至 Augmentation_x4 以获取额

外 1.2%~1.7% 的性能增益。

3.5.2 网络模型中模块的消融实验

所提出的改进主要集中在时间嵌入和多头自注意力机制两个模块, 本文通过设计消融实验来分析这两个模块对模型性能的影响。如表 5 所示, 分别探讨了去除时间嵌入和去除多头自注意力机制时模型的表现, 以评估它们各自对最终结果的贡献。

表 5 模块消融实验

Model	CryoNuSeg_Augmented_x2		ISBI2012_Augmented_x2	
	Augmented_x2		Augmented_x2	
	IoU	Dice	IoU	Dice
Baseline	56.53	68.98	80.81	88.01
Baseline+时间嵌入	61.69	73.87	85.09	90.89
Baseline+多头注意力机制	59.76	72.27	82.83	89.22
Our	62.50	75.78	85.25	91.85

实验结果表明, 模型架构和策略的选择对图像分割任务的性能具有显著影响。在 CryoNuSeg_Augmented_x2 数据集上, Baseline 模型的 IoU 和 Dice 分别为 56.53% 和 68.98%, 在 ISBI2012_Augmented_x2 数据集上, IoU 和 Dice 分别为 80.81% 和 88.01%。引入时间嵌入后, CryoNuSeg_Augmented_x2 数据集的 IoU 和 Dice 分别提升至 61.69% 和 73.87%, 而在 ISBI2012_Augmented_x2 数据集上, IoU 和 Dice 分别达到 85.09% 和 90.89%, 显示了时间嵌入在提升模型性能方面的有效性。时间嵌入通过引入时间步信息, 使模型能够理解图像在不同时间步的变化, 尤其是在扩散过程中噪声的引入和去噪阶段。这种设计使得模型能够更好地捕捉图像的动态演变, 提升了对细节和噪声的处理能力, 从而显著改善了分割精度。

与之相比, 采用多头自注意力机制后, 模型在 CryoNuSeg_Augmented_x2 数据集上的 IoU 和 Dice 分别为 59.76% 和 72.27%, 在 ISBI2012_Augmented_x2 数据集上的 IoU 和 Dice 分别为 82.83% 和 89.22%, 尽管相较于 Baseline 取得了明显提升, 但整体效果仍略逊色于引入时间嵌入的模型。多头自注意力机制通过多个

注意力头并行地关注图像的不同特征区域,从而捕捉更为丰富的全局上下文信息。这使得模型能够更好地理解图像中的长距离依赖关系,增强了对图像全局特征的捕捉能力,进而提升了分割精度。

最终,本文提出同时使用这两个模块的方法在两个数据集上的分割表现均优于其他方法。在 CryoNuSeg_Augmented_x2 数据集上, IoU 和 Dice 分别达到了 62.50% 和 75.78%, 而在 ISBI2012_Augmented_x2 数据集上, IoU 和 Dice 分别为 85.25% 和 91.85%, 显著优于 Baseline 和其他增强策略。这一结果表明, 结合时间嵌入和多头自注意力机制的方法, 在提升模型性能方面发挥了重要作用, 能够更好地处理图像的动态变化和全局上下文信息, 从而在图像分割任务中取得了更好的效果。

3.5.3 损失函数选取

通过对比均方误差(MSE)和 Hybrid 这两种类型的损失函数, 如公式(3)和公式(9)所示。并在表6展示了这些损失函数在图像分割任务中的表现。

$$L_{\text{hybrid}} = L_{\text{MSE}} + \lambda_{\text{L1}} \cdot L_{\text{L1}}, \quad (9)$$

其中: $\hat{y}_i$ 表示模型预测的值, y_i 表示真实值, N 表示样本数量, $L_{\text{L1}} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$ 表示 L1 损失, 计算的是绝对误差的平均值, λ_{L1} 表示一个权重系数, 控制 MSE 损失和 L1 损失的相对重要性。

表6 不同损失函数的对比实验

Tab. 6 Comparison of different loss functions (%)

损失函数	CryoNuSeg_		ISBI2012_	
	Augmented_x2		Augmented_x2	
	IoU	Dice	IoU	Dice
MSE	59.18	71.15	84.40	90.30
Hybrid	53.67	65.54	81.67	87.70
Our	62.50	75.78	85.25	91.85

实验结果表明, 不同损失函数对图像分割任务的性能具有显著影响。在 CryoNuSeg_Augmented_x2 数据集上, 使用 MSE 损失时, 模型的 IoU 和 Dice 分别为 59.18% 和 71.15%, 表现较差; 而 Hybrid 损失的效果有所提升, IoU 和 Dice 分别为 53.67% 和 65.54%。这可能是因为

MSE 损失对噪声和异常值敏感, 无法有效应对复杂的局部特征或数据噪声, 从而导致性能较差; 而 Hybrid 损失尽管结合了 MSE 和 L1 损失, 但仍未能充分克服数据集中的噪声问题和复杂特征。相比之下, 采用自定义损失函数后, 模型性能显著提升, IoU 和 Dice 分别为 62.50% 和 75.78%, 表明该损失函数更好地适应了数据集的特性, 尤其在处理噪声和局部特征时表现出色。

在 ISBI2012_Augmented_x2 数据集上, 自定义损失函数同样取得了最优结果, IoU 和 Dice 分别为 85.25% 和 91.85%, 明显优于 MSE 和 Hybrid 损失。这表明, 在 ISBI2012 数据集上, 自定义损失函数能够更好地处理数据的多样性和特征, 从而提升分割性能。总体而言, 本文使用的损失函数能显著提高模型在不同数据集上的分割性能, 尤其在应对数据集特性和噪声方面展现出更好的适应性。

4 结 论

为解决细胞图像分割任务中标注数据不足及传统增强方法局限性的问题, 本研究提出了一种基于扩散模型的细胞图像与掩膜联合生成数据扩充方法。该方法通过构建噪声预测 U-Net 与噪声采样器组成的联合生成框架, 将细胞图像和掩膜按通道堆叠为四通道联合图像, 实现一步生成高质量合成数据, 避免了多阶段生成引入的误差。模型融合时间嵌入与多头自注意力机制, 增强了对细胞特征及掩膜关系的建模能力, 并采用均方误差损失函数与余弦退火学习率调度优化训练过程。实验结果表明, 在 CryoNuSeg 和 ISBI2012 数据集上的性能显著提升, 特别是在 DPM++ 2M Karras 采样下, IoU 和 Dice 系数分别达到 62.50% 和 75.78% (CryoNuSeg), 优于传统翻转、旋转等增强手段, 验证了联合生成增强方法在扩展数据多样性与提升分割精度方面的优势。

综上所述, 本研究不仅在解决细胞图像分割任务中的数据不足问题方面取得了突破, 还提出了一种高效的数据扩充方法, 并验证了其在实际应用中的效果。未来可进一步探索扩散模型在多模态数据生成中的应用, 以及在实时临床场景

中的优化部署,推动该技术在医学图像分割中的广泛应用。

作者贡献声明:

徐凯歌:方法的提出,论文构思和撰写;

孙晓明:论文审核与编辑写作;

陈泳吉:实验方法的实现及数据整理;

陈言:实验数据分析;

龙兵:实验数据验证;

韩永琦:实验数据验证;

苏伯超:论文审核与编辑写作;

魏昌宁:论文审核与编辑写作。

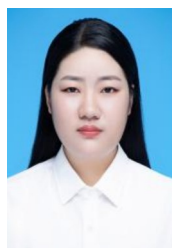
参考文献:

- [1] 马国祥,严传波,张志豪,等. 基于数据增强的CT图像病灶检测方法[J]. 计算机系统应用, 2021, 30(10): 187-194.
MA G X, YAN C B, ZHANG Z H, *et al.* Lesion detection method for CT images based on data augmentation[J]. *Computer Systems & Applications*, 2021, 30(10): 187-194. (in Chinese)
- [2] 康瀚,苏志金. 图像数据增强技术原理与发展综述[J]. 信息技术, 2024(9): 176-185.
KANG L, SU Z J. Principle and prospect of image data enhancement technology[J]. *Information Technology*, 2024(9): 176-185. (in Chinese)
- [3] 李帆雅,张泽辉,陈博洋,等. 面向复杂工业场景的人体姿态估计性能增强方法[J]. 仪器仪表学报, 2025, 46(8): 255-265.
LI F Y, ZHANG Z H, CHEN B Y, *et al.* Human pose estimation performance enhancement method for complex industrial scenes[J]. *Chinese Journal of Scientific Instrument*, 2025, 46(8): 255-265. (in Chinese)
- [4] ZEISER F A, COSTA C A D, ZONTA T, *et al.* Segmentation of masses on mammograms using data augmentation and deep learning[J]. *Journal of Digital Imaging*, 2020, 33(4): 858-868.
- [5] MANJUNATH R V, KWADIKI K. Automatic liver and tumour segmentation from CT images using Deep learning algorithm[J]. *Results in Control and Optimization*, 2022, 6: 100087.
- [6] 邹鸿博,章彪,王子川,等. 基于光照模型的细胞内镜图像不均匀光照校正算法[J]. 中国光学(中英文), 2024, 17(1): 160-166.
ZOU H B, ZHANG B, WANG Z C, *et al.* Non-uniform illumination correction algorithm for cytoscopy images based on illumination model[J]. *Chinese Optics*, 2024, 17(1): 160-166. (in Chinese)
- [7] NALEPA J, MARCINKIEWICZ M, KAWULOK M. Data augmentation for brain-tumor segmentation: a review[J]. *Frontiers in Computational Neuroscience*, 2019, 13: 83.
- [8] 邓慧,吕国皎,杨梅,等. 基于掩膜板阵列的消串扰集成成像3D显示方法[J]. 液晶与显示, 2022, 37(5): 592.
DENG H, LÜ G J, YANG M, *et al.* Crosstalk-free integral imaging 3D display method based on a mask array[J]. *Chinese Journal of Liquid Crystals and Displays*, 2022, 37(5): 592. (in Chinese)
- [9] HE R N, XU S Q, LIU Y S, *et al.* Three-dimensional liver image segmentation using generative adversarial networks based on feature restoration[J]. *Frontiers in Medicine*, 2022, 8: 794969.
- [10] ABDELHAFIZ D, BI J B, AMMAR R, *et al.* Convolutional neural network for automated mass segmentation in mammography[J]. *BMC Bioinformatics*, 2020, 21(1): 192.
- [11] VERMA D, VERMA H, TIWARI P K. A hybrid approach for MRI brain image segmentation using KIFECM-IPSO algorithm[J]. *Expert Systems with Applications*, 2025, 268: 126239.
- [12] ISENSEE F, JAEGER P F, KOHL S A A, *et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. *Nature Methods*, 2021, 18(2): 203-211.
- [13] 田卓展,陈纯毅,胡小娟,等. 融合生成对抗网络的大气无线光信道密钥提取[J]. 光学精密工程, 2025, 33(3): 486-496.
TIAN Z Z, CHEN C Y, HU X J, *et al.* Secret key extraction from atmospheric wireless optical channels by combining with generative adversarial network[J]. *Opt. Precision Eng.*, 2025, 33(3): 486-496. (in Chinese)
- [14] 安义帅,付钰,俞艺涵,等. 基于数据增强与特征挖掘的异常流量检测方法[J]. 通信学报, 2025, 46(8): 16-30.
AN Y S, FU Y, YU Y H, *et al.* Anomaly traffic detection method based on data augmentation and

- feature mining [J]. *Journal on Communications*, 2025, 46(8): 16-30. (in Chinese)
- [15] WEI X Q, CHEN X W, LAI C, *et al.* Automatic liver segmentation in CT images with enhanced GAN and mask region-based CNN architectures [J]. *BioMed Research International*, 2021, 2021: 9956983.
- [16] SINGH V K, RASHWAN H A, ROMANI S, *et al.* Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network [J]. *Expert Systems with Applications*, 2020, 139: 112855.
- [17] MONDAL A K, DOLZ J, DESROSIERS C. Few-shot 3D Multi-Modal Medical Image Segmentation Using Generative Adversarial Learning [EB/OL]. 2018; *arXiv*: 1810.12241. <https://arxiv.org/abs/1810.12241>
- [18] CHEN C, DOU Q, CHEN H, *et al.* Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, 39(7): 2494-2505.
- [19] LI Q Y, YU Z B, WANG Y B, *et al.* TumorGAN: a multi-modal data augmentation framework for brain tumor segmentation [J]. *Sensors*, 2020, 20(15): 4203.
- [20] AL-SHEIKH M H, DANDAN OAL, AL-SHAMAYLEH A S, *et al.* Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images [J]. *Scientific Reports*, 2023, 13: 19373.
- [21] CHOKUWA S, KHAN M H. Generalizing across domains indiabetic retinopathy via variational auto-encoders [C]. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2023 Workshops*. Cham: Springer, 2023: 265-274.
- [22] SHAO S T, YUAN X H, HUANG Z, *et al.* DiffuseExpand: Expanding Dataset for 2D Medical Image Segmentation Using Diffusion Models [EB/OL]. 2023; *arXiv*: 2304.13416. <https://arxiv.org/abs/2304.13416>
- [23] DU Y H, JIANG Y C, TAN S Y, *et al.* ArSDM: colonoscopy images synthesis with adaptive refinement semantic diffusion models [C]. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2023*. Cham: Springer, 2023: 339-349.
- [24] MACHÁČEK R, MOZAFFARI L, SEPASDAR Z, *et al.* Mask-conditioned latent diffusion for generating gastrointestinal polyp images [C]. *4th Workshop on Intelligent Cross-Data Analysis and Retrieval. Thessaloniki Greece*. ACM, 2023: 1-9.
- [25] ESCHWEILER D, YILMAZ R, BAUMANN M, *et al.* Denoising diffusion probabilistic models for generation of realistic fully-annotated microscopy image datasets [J]. *PLoS Computational Biology*, 2024, 20(2): e1011890.
- [26] SARAGIH D G, HIBI A, TYRRELL P N. Using diffusion models to generate synthetic labeled data for medical image segmentation [J]. *International Journal of Computer Assisted Radiology and Surgery*, 2024, 19(8): 1615-1625.
- [27] 闫光辉, 吴佰靖, 马龙. LightDiffu-DCE: 基于光照强度扩散的低光照图像增强 [J]. *光学精密工程*, 2025, 33(7): 1114-1129.
- YAN G H, WU B J, MA L. LightDiffu DCE: low light image enhancement based on light intensity diffusion [J]. *Opt. Precision Eng.*, 2025, 33(7): 1114-1129. (in Chinese)
- [28] 肖珀麟, 刘进锋. 扩散模型与傅里叶滤波器结合的图像去雾方法 [J]. *液晶与显示*, 2024, 39(11): 1532-1543.
- XIAO P L, LIU J F. Image dehazing method combining diffusion model and Fourier filter [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(11): 1532-1543. (in Chinese)
- [29] WOLLEB J, SANDKÜHLER R, BIEDER F, *et al.* Diffusion Models for Implicit Image Segmentation Ensembles [EB/OL]. 2021; *arXiv*: 2112.03145. <https://arxiv.org/abs/2112.03145>
- [30] RAHMAN A, VALANARASU J M J, HACIHALILOGLU I, *et al.* Ambiguous medical image segmentation using diffusion models [C]. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 11536-11546.
- [31] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [J]. *Advances in neural information processing systems*, 2017, 30.
- [32] ZHANG Z Y, YAO L H, WANG B, *et al.* DiffBoost: enhancing medical image segmentation via text-guided diffusion model [J]. *IEEE Transac-*

- tions on Medical Imaging, 2025, 44 (9) : 3670-3682.
- [33] SONG J M, MENG C L, ERMON S. Denoising Diffusion Implicit Models [EB/OL]. 2020: *arXiv*: 2010.02502. <https://arxiv.org/abs/2010.02502>
- [34] SONG Y, SOHL-DICKSTEIN J, KINGMA D P, *et al.* Score-based Generative Modeling through Stochastic Differential Equations [EB/OL]. 2020: *arXiv*: 2011.13456. <https://arxiv.org/abs/2011.13456>
- [35] MAHBOD A, SCHAEFER G, BANCHER B, *et al.* CryoNuSeg: a dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images [J]. *Computers in Biology and Medicine*, 2021, 132: 104349.
- [36] ISBI Challenge. ISBI Challenge Dataset [DB/OL]. (2012) [2025-02-10]. http://brainiac2.mit.edu/isbi_challenge/.
- [37] HEUSEL M, RAMSAUER H, UNTERTHINER T, *et al.* GANs trained by a two time-scale update rule converge to a local nash equilibrium [C]. *Neural Information Processing Systems.*, 2017
- [38] ZHANG R, ISOLA P, EFROS A A, *et al.* The unreasonable effectiveness of deep features as a perceptual metric [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 586-595.
- [39] PARK T, LIU M Y, WANG T C, *et al.* Semantic image synthesis with spatially-adaptive normalization [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 2332-2341.
- [40] MIROWSKI P, FABIJANŠKA A. Diffusion model-based synthesis of brain images for data augmentation [J]. *Biomedical Signal Processing and Control*, 2026, 113: 108940.
- [41] ZHOU Z W, RAHMAN SIDDIQUEE M M, TAJBAKHS N, *et al.* UNet++: a nested U-Net architecture for medical image segmentation [C]. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer, 2018: 3-11.
- [42] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 7-12, 2015, Boston, MA, USA. IEEE, 2015: 3431-3440.
- [43] CHEN L C, PAPANDREOU G, SCHROFF F, *et al.* Rethinking Atrous Convolution for Semantic Image Segmentation [EB/OL]. 2017: *arXiv*: 1706.05587. <https://arxiv.org/abs/1706.05587>

作者简介:



徐凯歌(1994—),女,河南信阳人,博士研究生,2017年于郑州大学获得学士学位,2021年于哈尔滨理工大学获得硕士学位,现就读于哈尔滨理工大学,主要从事机器视觉及图像处理方面的研究。E-mail: xukaige0112@163.com

通讯作者:



孙晓明(1983—),女,黑龙江海林人,博士,教授,博士生导师,2007年、2012年于哈尔滨工业大学分别获得硕士、博士学位,现为哈尔滨理工大学教授,主要从事机器视觉、图像处理和信号特征提取等方面的研究。E-mail: sunxiaoming@hrbust.edu.cn