

# 基于改良编码与高斯过程的交互式医学图像分割<sup>①</sup>



张小瑞<sup>1,2</sup>, 莫云菲<sup>3</sup>, 孙伟<sup>4</sup>

<sup>1</sup>(南京信息工程大学 电子与信息工程学院, 南京 210044)

<sup>2</sup>(南京工业大学 计算机与信息工程学院, 南京 211816)

<sup>3</sup>(南京信息工程大学 计算机学院, 南京 210044)

<sup>4</sup>(南京信息工程大学 自动化学院, 南京 210044)

通信作者: 张小瑞, E-mail: [zxr365@126.com](mailto:zxr365@126.com)

**摘要:** 医学图像分割是众多医学临床应用的基础与关键组成. 近年来, 交互式分割方法凭借其在复杂临床任务中的高准确性和鲁棒性受到广泛关注. 然而, 现有基于深度学习的交互式分割方法在用户交互的利用上仍有不足, 特别是在交互编码设计和像素分类等方面. 针对上述问题, 本文提出了一种包含“近中心点”和“外边缘点”的混合交互设计, 以保障交互成本并对用户意图进行精准捕捉; 同时, 通过高斯衰减函数对现有测地线距离编码方法进行加权, 以降低图像噪声干扰, 提高交互编码的鲁棒性和准确性; 此外, 结合基于混合核函数的高斯过程分类方法, 在像素分类过程中对用户交互信息进行充分利用, 提升分割精度并赋予模型一定的可解释性. 实验结果表明, 本文所提方法在 MSD 数据集 4 个标志性子集的 5 项分割任务中均表现出较高的分割精度, 尤其在复杂任务 (如 Pancreas tumour 和 Colon 图像分割) 中, Dice 系数和 ASSD 值显著优于现有方法, 体现了其在精准分割和边界处理方面的优势.

**关键词:** 医学图像分割; 交互式图像分割; 深度学习; 高斯过程分类

引用格式: 张小瑞, 莫云菲, 孙伟. 基于改良编码与高斯过程的交互式医学图像分割. 计算机系统应用, 2025, 34(5): 138-150. <http://www.c-s-a.org.cn/1003-3254/9846.html>

## Interactive Medical Image Segmentation Based on Improved Encoding and Gaussian Process

ZHANG Xiao-Rui<sup>1,2</sup>, MO Yun-Fei<sup>3</sup>, SUN Wei<sup>4</sup>

<sup>1</sup>(School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

<sup>2</sup>(College of Computer and Information Engineering, Nanjing Tech University, Nanjing 211816, China)

<sup>3</sup>(School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China)

<sup>4</sup>(School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** Medical image segmentation serves as a fundamental and critical component in numerous clinical applications. Recent advancements in interactive segmentation methods have attracted significant attention due to their high accuracy and robustness in complex clinical tasks. However, current deep learning-based interactive segmentation methods exhibit limitations in leveraging user interactions, particularly in interactive encoding design and pixel classification. To address these limitations, this study proposes a hybrid interaction design incorporating “near-center points” and “outer-edge points”, which ensures low interaction costs while accurately capturing user intent. Additionally, the existing geodesic distance encoding method is enhanced by a Gaussian attenuation function to mitigate image noise interference and improve the robustness and accuracy of interaction encoding. Furthermore, a Gaussian process classification method based on a hybrid kernel function is integrated to fully exploit user interaction information during pixel classification, enhancing segmentation accuracy while endowing the model with interpretability. Extensive experiments on five

① 基金项目: 国家自然科学基金 (62272236, 62376128); 江苏省自然科学基金 (BK20201136, BK20191401)

收稿时间: 2024-10-29; 修改时间: 2024-11-19; 采用时间: 2024-12-04; csa 在线出版时间: 2025-03-31

CNKI 网络首发时间: 2025-04-01

segmentation tasks across four representative subsets of the medical segmentation decathlon (MSD) dataset demonstrate that the proposed method achieves consistently high segmentation accuracy. In particular, for complex tasks such as pancreas tumor and colon image segmentation, this method has significantly higher Dice coefficients and ASSD values than existing methods, showing its strengths in precise segmentation and boundary refinement.

**Key words:** medical image segmentation; interactive image segmentation; deep learning; Gaussian process classification

医学图像分割任务的主要目标在于从医学图像中准确分割出器官、病灶等用户感兴趣的结构,是医学三维重建与定量分析等任务中的关键步骤,更是病灶区域量化、治疗手段选择、放射治疗等重要技术的基础前提<sup>[1-3]</sup>.医学图像分割结果的准确性直接影响着医生对疾病状况的判断及后续治疗方案的选择,因此,如何在保证计算成本的基础上提升分割准确性一直是研究的焦点之一.

目前,借助大规模数据集和监督学习的支持,以FCN<sup>[4]</sup>、U-Net<sup>[5]</sup>等深度学习模型为代表的自动分割模型已经能够在一些简单的医学分割任务中实现与人类专家相媲美的效果<sup>[6-9]</sup>.然而,这一成就仅限于常规任务领域.在更为广泛、具体的应用场景下,受到患者隐私保护与样本标注成本限制,医学图像分割任务普遍面临着样本数据稀缺的挑战<sup>[10]</sup>.此外,在诸如肿瘤病灶区域图像分割等的复杂应用场景中,所涉及的显著病患个体差异及组织结构的复杂性等问题更进一步限制了自动图像分割方法的效果.相比之下,交互式图像分割方法借助用户针对性的指示交互信息对模型的分割过程进行指导,不仅降低了模型对于大规模数据集的依赖,也使得模型能够在复杂、多变的医学图像分割任务中保持较高的准确性和鲁棒性<sup>[11-15]</sup>.然而,如何准确捕捉用户意图并进行有效编码,同时结合用户交互信息和医学图像特性提取像素特征,从而实现精准分割,仍是该领域的关键研究挑战.

近年来,以卷积神经网络(CNN)为代表的深度学习技术在图像处理领域取得了显著进展,并被广泛应用于医学图像的交互式分割任务中.CNN凭借其强大的局部特征提取和多层次表示能力,能够对复杂的医学图像结构进行精确建模,使其在捕捉纹理、边缘以及空间语义信息方面表现出色.在交互式分割任务中,CNN方法通过将用户提供的交互信息(如点击、涂鸦等)编码为特定形式的引导特征,并与原始图像一同输入网络,以动态调整分割边界,展现出显著的灵活性与

高效性.这使得CNN在医学图像分割中具备自动化、可扩展性强和快速迭代的优点.然而,现有的交互编码方法在整合用户交互信息时,仍存在难以有效综合全局上下文信息的局限性.同时,这些方法容易受到噪声和灰度变化的干扰,特别是在医学图像边界模糊、对比度低的情况下,难以满足分割任务对精度和鲁棒性的高要求.此外,尽管基于CNN的自动图像分割方法在很多场景中表现出色,但许多现有的交互式图像分割方法在设计上依然主要是通过对自动分割框架的模块(如像素对比模块)进行改进,缺乏针对交互任务特点的专门优化.这种对自动分割框架的依赖限制了交互式方法在动态调整和灵活应对复杂分割场景的潜力.

针对上述问题,为了准确捕捉用户意图,充分利用用户交互信息,提升模型在应对复杂分割任务时的准确性和鲁棒性,本文提出了一种基于改良测地线距离编码和混合高斯过程分类的医学图像交互式分割方法,对交互编码方法、像素分类方法以及用户初次交互等模块进行了设计和改良.具体而言,本文的贡献主要包含以下几个部分.

(1) 利用高斯衰减函数对测地线距离进行加权,提出了一种高斯测地线距离编码方法,利用高斯衰减函数平滑过渡和快速衰减的特性,进一步降低图像噪声干扰,提高编码结果的鲁棒性和准确性;

(2) 结合基于混合核函数的高斯过程分类方法改进像素分类过程,利用用户指示交互建模像素特征与标签之间的映射关系,通过混合核函数捕捉医学图像中不同尺度的特征,在提高分割精度的同时,赋予模型一定的可解释性;

(3) 设计了一种“外边缘点”与“近中心点”结合的混合点击交互形式,结合分割目标的全局边界信息与局部结构信息,进一步丰富用户标注中蕴含的指示信息,提升分割的准确性、鲁棒性和效率.

本文在医学图像算法领域具有标志性的MSD数据集的4个标志性子集上进行了广泛实验,本文方

法在所涉及的5项分割任务中均表现出较高的分割精度,尤其在复杂分割任务中表现出色,分割效果显著优于现有方法。

## 1 相关工作

传统的交互式分割方法通常着眼于待分割图像的低级特征(灰度特征、颜色分布等)<sup>[16-18]</sup>,以此建立基于优化的图像模型。然而,医学图像普遍存在边界模糊、对比度低的问题,在处理医学图像分割任务时,其低级特征往往无法有效区分分割目标与背景,这就导致基于低级特征的传统交互式医学图像分割方法通常需要大量的用户交互和较长的用户时间才能获得较为可靠的结果。机器学习方法的引入一定程度上缓解了这一问题<sup>[19-21]</sup>。其通过主动选择候选区域向用户询问,从而获得大量信息性的用户反馈,以减少精确分割所需用户交互数量与用户时间。然而此类算法需要研究人员基于理解与经验自己设计、选择用于训练机器学习模型的特征,这不仅会耗费大量的时间与精力,也使得模型缺乏表达能力与适应性。

近年来,以卷积神经网络(CNN)为代表的深度学习技术在图像处理领域取得了显著的成效,这一框架也迅速被应用到交互式医学图像分割这一领域中<sup>[22-27]</sup>。如Wang等<sup>[10]</sup>提出的基于双CNN的DeepIGeoS框架,利用一个CNN生成初次自动分割结果,另一个CNN修正并细化用户交互信息。Koohbanani等<sup>[26]</sup>在卷积神经网络的基础上结合迁移学习来分割皮肤病变及其属性,利用多尺度信息融合生成预测结果。Sakinis等<sup>[27]</sup>在FCNN基础上提出了适用于2D医学图像的分割框架,支持用户通过新增或移动点击来细化分割结果。借助CNN强大的局部特征提取和特征表示能力,此类方法均取得了较为不错的成绩。整体而言,基于卷积神经网络的医学图像交互式分割方法通常遵循以下步骤:1)获取并编码用户交互信息;2)将待分割图像与编码的后用户交互输入到卷积神经网络,提取多尺度图像特征;3)逐像素分类,生成像素标签分类概率图;4)阈值化概率图,获取预测分割结果。在此过程中,用户的交互信息通常不会直接输入神经网络,而是首先经过点击编码,以进一步提取用户交互标注中包含的指示聚焦信息,获得相应的交互线索图后,再与待分割图像共同输入神经网络,从而指导图像及像素特征提取。

因此如何准确、有效编码用户交互信息至关重要,

专家学者提出了众多方案。例如,Li等<sup>[28]</sup>和Xu等<sup>[29]</sup>采用简单高效的欧几里得距离变换,将用户正负点击转换为两个欧氏距离映射。而Maninis等<sup>[23]</sup>和Wang等<sup>[30]</sup>则通过高斯函数将用户交互编码为高斯热图。这两种方法虽然简单直观,却没有综合考虑图像灰度变化与局部结构。在此基础上,Khan等<sup>[31]</sup>提出一种置信图(confidence map)编码方法,设计映射函数来计算目标图像中每个像素与用户标注极值点之间的关系,从而为每个像素赋予一个置信分数(confidence score)。虽然置信分数的映射关系可以依据不同的分割任务灵活调整,但其编码效果高度依赖于用户标注的准确性和映射关系函数的选择,因此在应对不同结构类型的分割任务时,需要重新设计合适的交互位置和映射关系函数。Wang等<sup>[10]</sup>则引入了测地线距离来对用户标注进行编码。这种方法不仅能体现像素间的几何距离关系,还综合了灰度变化、纹理细节等局部特征,但该编码方法易受噪声和灰度分辨率的影响,在处理复杂图像结构时,常常会出现路径不平滑或结果不自然的编码效果。

在依据用户交互和待分割图像通过神经网络获得多尺度图像特征之后,需要据此对待分割图像的每个像素的所属标签进行预测。目前常用的分类器有Sigmoid<sup>[5]</sup>、Softmax<sup>[4]</sup>、CRF(条件随机场)<sup>[32]</sup>等。其中Sigmoid分类器计算效率较高但仅适用于二元分类任务。Softmax分类器能够应对多类别分类任务,但在处理高维度、多类别问题时计算成本过高,因此通常结合交叉熵损失函数来衡量分类器输出的概率分布与真实类别分布之间的差异,优化Softmax分类器的性能,提高收敛速度和分类精度<sup>[33,34]</sup>。CRF作为后处理方法,利用像素间上下文关系优化分类结果,进一步优化分类器的输出结果,对边缘细节的分割效果有显著提升。然而,上述常规分类器均依赖逐像素对比分类,未能在像素分类过程中充分利用交互指示信息,这显然是一种浪费。为此,Zhou等<sup>[35]</sup>结合高斯过程分类模型与深度神经网络提出GPCIS模型,该模型在自然图像分割中效果显著,但在处理医学图像中复杂且形状不规则的分割目标时,难以满足任务需求。

此外,交互式分割方法依赖于用户的指示标注来指导后续分割过程,因此,用户交互的效率和质量直接影响模型的整体性能,为了实现在最小化用户交互成本的同时获取丰富指示信息这一目标,研究者们设计了多种交互形式,如框选<sup>[22]</sup>、涂鸦<sup>[24]</sup>和点击<sup>[11,23,30]</sup>等。显而易

见的是,框选、涂鸦等交互形式能够为模型提供更为丰富、具体的指示信息,但它们通常需要用户投入较多时间与精力以实现精准交互<sup>[23]</sup>。相比之下,点击交互因其灵活性和较低的用户成本而受到青睐。常规的点击交互设计通常会引导用户标注前景和背景两类点,但单一的前景、背景属性只能提供局部信息的输入,无法有效捕捉待分割目标的整体形状或边界信息。在此基础上, Xu等<sup>[29]</sup>提出以第1个点击信息生成注意力权重,这种方法有效引导后续分割过程,减少了后续点击所需的数量和精确度。但其过度依赖第1个点击的质量,使得模型的鲁棒性得不到保障,且对于形状复杂或不规则的目标物体,第1个点击生成的 attention 可能只涵盖一部分目标区域,导致模型对目标的整体理解不足。Maninis等<sup>[23]</sup>则尝试通过极值点的点击设计减少用户点击次数、丰富标注信息,但极值点交互对于用户交互的精

度要求过高,需要用户投入大量精力,反而增加了交互成本。针对这一问题, Luo等<sup>[14]</sup>提出了一种在待分割目标内部近边缘位置进行点击的交互设计,相较于极值点,这种交互设计有效减轻了用户负担,但集中在待分割目标内部的点击忽视了周围背景对目标的影响,在处理形状复杂目标时面临挑战。

## 2 方法

本文提出的基于改良测地线距离编码和混合高斯过程分类的医学图像交互式分割方法,旨在通过优化点击交互设计、交互编码方式与像素分类方法,在模型的各个部分实现对用户指示交互信息的准确提取与充分利用,从而提升在医学图像分割任务中分割精度和鲁棒性。整体结构如图1所示,可以分为3个关键模块:点击编码、特征提取和像素分类。

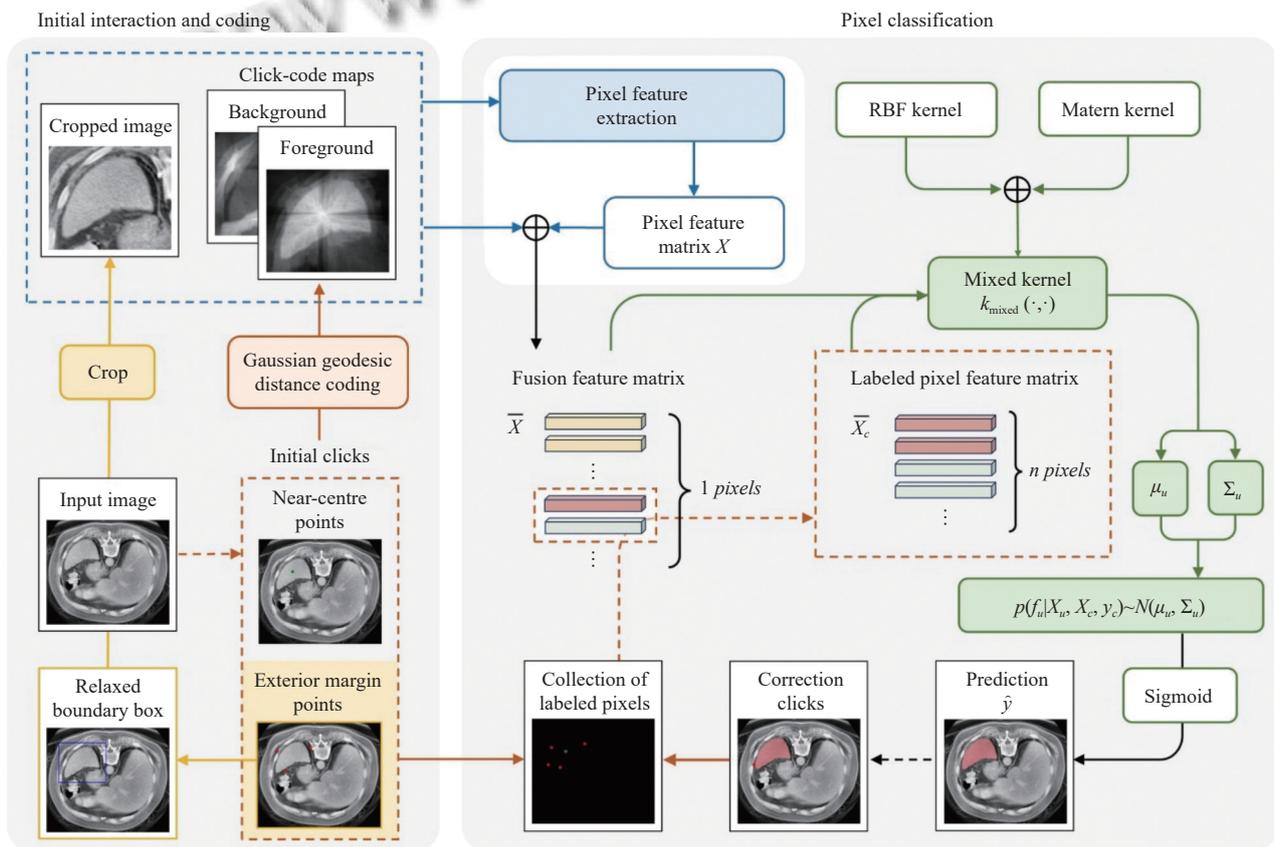


图1 整体结构图

在点击编码模块中,针对初次交互,本文提出了一种融合点击交互方法,引导用户在分割目标内部近中心区域与外侧边缘附近进行前景、背景的点击标注。通过获取到的外边缘点,定位分割目标所在区域,对待

分割图像进行宽松裁剪,以缩小后续处理范围。

在裁剪后的待分割图像的基础上,通过本文提出的高斯测地线距离编码方法对用户提供的初次交互进行编码。而后将编码图像与裁剪后的待分割图像进

行级联, 导入特征提取模块, 生成逐像素的像素特征矩阵。

在像素分类模块中, 本文结合高斯过程分类对现有医学图像交互式分割方法模块构成中的像素分类模块进行改良, 依据像素特征矩阵, 采用混合核函数对像素间关系进行捕捉, 利用用户标注信息建模像素特征与像素标签之间的映射关系, 对未标注像素进行分类, 获取预测的分割结果。用户可以对预测的分割结果进行修正, 直到获取满意结果。

### 2.1 融合点击交互

在交互式图像分割任务中, 用户初次交互是系统获取用户意图的关键步骤。其质量和精度决定了模型对目标区域的初步认知, 若无法从用户初次交互中获取充分、准确的指示信息, 系统可能会误判目标区域, 进而影响整个分割过程的效率和准确性。因此, 设计一个既能准确传递用户意图、又能降低用户操作负担的初次交互方式至关重要。虽然框选、涂鸦等交互形式能够提供更为丰富的信息, 却也需要用户付出更多的交互成本来实现精准标注。因此, 本文采用了点击交互作为基础交互形式, 以保证较低的用户交互成本, 减少交互误差。传统的“前景点+背景点”点击交互形式只能提供局部信息的输入, 无法有效捕捉待分割目标的整体形状或边界信息。而在复杂的医学图像中, 单一类型的交互信息难以完整表达目标对象的形状和边界信息。

因此, 本文对前景与背景的点击标注位置分别进一步的细化与规定, 设计了一种“近中心点+外边缘点”的融合交互形式。该交互形式引导用户提供两组关键点: 一组位于分割目标内部、靠近中心区域的前景点  $S_c = \{p_1, p_2, \dots, p_n\}$ , 另一组位于分割目标外侧、靠近边界的背景点  $S_e = \{q_1, q_2, \dots, q_m\}$ , 作为初始交互。该交互形式的优点在于, 规定点击位置增强了用户提供信息的针对性, 从中系统能够更准确地识别用户聚焦区域, 减少误判。外边缘点的标注还可以用于生成一个完整包含分割目标的宽松边界框  $Box = \{x_1, y_1, x_2, y_2\}$ 。基于该边界框, 本文对待分割图像  $I$  进行裁剪, 获得裁剪后的待分割图像  $I'$  以缩小后续处理范围, 从而显著减少后续计算的复杂度。提升计算效率的同时, 系统能够集中计算资源于相关区域, 从而提升分割的精度和性能。而在获取初次预测分割结果后的迭代修正阶段, 我们没有对用户交互位置进行具体规定, 用户仅需要直接在分割缺失以及溢出部分补充前景、背景点击标注。

### 2.2 高斯测地线距离编码

在交互式分割任务中, 用户提供的交互信息无法直接用于神经网络进行特征提取, 需要通过编码将这些交互指示信息转换为网络可识别的形式。此外, 准确有效的交互编码还能捕捉用户交互中所包含的聚焦信息, 为模型后续工作提供参考。

为综合考虑几何距离、灰度变化、纹理细节等上下文信息, 并减少医学图像噪声带来的干扰, 本文在测地线距离编码的基础上设计了一种改良的高斯测地线距离编码方法, 以准确引导网络识别图像中的目标特征, 提升编码在复杂分割任务中的准确性和鲁棒性。

具体而言, 本文采用高斯衰减函数对测地线距离进行加权, 以降低远距离路径对计算的影响。在医学图像中, 灰度值或纹理渐变常伴随着噪声或非均匀性, 尤其是在复杂组织或病变区域, 远距离像素间的相似性可能会掩盖真实边界信息。通过加权, 能够有效减小长路径带来的累积误差, 增强分割方法在高噪声或灰度不均匀区域中的鲁棒性, 确保模型对局部重要特征的敏感度更高, 避免远离交互点的噪声干扰影响整体计算。

在本文的设计中, 用户初次交互时提供的点击标注会依据位置类型被分为两组, 中心附近内部点(前景点)  $S_c$  可以帮助确定目标的核心区域, 而外边缘点(背景点)  $S_e$  则有助于准确描绘目标的边界, 避免模糊或噪声影响边界识别。基于此, 我们分别对  $S_c$  和  $S_e$  进行编码, 生成两张点击编码图像, 用于进一步提升模型对目标区域和边界的区分能力。假设  $i$  是裁剪后的待分割图像  $I'$  中的一个像素, 则裁剪后的待分割图像  $I'$  上, 从像素  $i$  到用户标注像素集合  $S$  (其中  $S \in \{S_e, S_c\}$ ) 的高斯测地线距离  $G(i, S, I')$  为:

$$G(i, S, I') = \min_{j \in S} \exp\left(-\frac{D_{\text{geo}}(i, j, I')}{2\sigma^2}\right) \quad (1)$$

$$D_{\text{geo}}(i, j, I') = \min_{p \in P_{i,j}} \int_0^1 \|\nabla I(p(s)) \cdot u(s)\| ds \quad (2)$$

其中,  $P_{i,j}$  是像素  $i$  和  $j$  之间所有可能路径的集合,  $p$  是路径集合中的一条可行路径, 用参数  $s \in [0, 1]$  参数化。 $u(s)$  是与路径  $p$  方向相切的单位向量, 定义为  $u(s) = p'(s) / \|p'(s)\|$ 。超参数  $\sigma$  控制高斯衰减的速度, 并在训练过程中通过优化自动调整。

### 2.3 像素特征提取

在获取裁剪后的待分割图像  $I'$  及其对应的点击编码图像后, 二者作为输入共同导入卷积神经网络。这一

过程旨在结合目标区域信息和用户交互信息,从而有效捕捉图像特征并提取像素特征. 为确保框架灵活性、可迁移性和可扩展性,本文的设计不依赖于固定的CNN结构,而是允许灵活替换并支持后续升级和发展.

针对本文聚焦的医学图像分割任务,我们采用在医学图像自动分割领域表现突出的 nnU-Net<sup>[33]</sup>作为特征提取模块,对裁剪后的待分割图像 $I'$ 的像素特征进行提取,以获得像素特征矩阵 $X$ . 随后,我们将该特征举证与裁剪后的待分割图像 $I'$ 以及来自点击编码模块的点击编码图像进行级联,而得到融合特征矩阵 $\bar{X}$ ,为后续像素分类提供依据. 为增强模型在处理医学图像时的稳定性和效率,本文用实例归一化层替换了批量归一化层,这种调整能够在更小批量的情况下仍保持较好的归一化效果. 同时,将特征通道数减少为原来的1/4,以在性能、存储开销和计算时间之间取得更好的平衡.

#### 2.4 基于混合核函数高斯过程分类的像素分类模块

在医学图像分割任务中,像素分类是整个流程的核心任务之一. 通过计算像素所属标签的概率,为每个像素分配前景或背景标签,从而精确提取用户感兴趣的结构. 然而,现有像素分类方法往往未能充分利用用户交互信息,导致模型准确性受到限制. 针对这一问题,本文采用一种基于高斯过程分类器 (Gaussian process classifier, GPC) 的像素分类方法,以用户标注的像素作为样本,通过核函数捕捉像素间的关系,从而对未标注的像素进行标签预测,实现在像素分类过程中对用户交互信息的进一步利用. 同时,针对医学图像分割任务复杂多样的特点,本文设计了一种结合多种特征的混合核函数,以提升模型对医学图像中模糊边界特征的捕捉能力,进一步提升像素分类的准确性与鲁棒性.

##### 2.4.1 基于高斯过程分类的像素分类模块

待分割图像上像素的特征向量 $x$ 与其对应的标签 $y$ 的映射关系可以表示为 $y = f(x)$ ,依据高斯过程假设,在不考虑噪声的情况下,映射函数 $f$ 服从一个0均值的高斯过程先验分布,即:

$$f \sim GP(0, K(x, x')) \quad (3)$$

其中, $K(\cdot, \cdot)$ 是核函数,用于衡量任意两个点 $x$ 和 $x'$ 之间的相似性. 以用户交互过程中标注像素的特征 $X_c$ 以及对应标签 $y_c$ 为训练样本,依据未标注像素的特征 $X_u$ ,推测未标注像素标签 $y_u$ 的任务可以通过如下公式表示:

$$p(y_u | X_u, X_c, y_c) = \int p(y_u | f_u) \cdot p(f_u | X_u, X_c, y_c) df_u \quad (4)$$

其中, $p(f_u | X_u, X_c, y_c)$ 是条件分布,表示在给定用户标注的像素信息的情况下,未标注像素对应的函数值 $f_u$ 的分布;而 $p(y_u | f_u)$ 则是一个基于 Sigmoid 函数的转换函数,用于将连续的函数值 $f_u$ 转换为离散的分类标签. 通过对未标注像素的特征进行推断,可以得出标签的概率分布.

接下来,对 $p(f_u | X_u, X_c, y_c)$ 进行计算:

$$p(f_u | X_u, X_c, y_c) = \int p(f_u | X_u, X_c, f_c) \cdot p(f_c | X_c, y_c) df_c \quad (5)$$

其中, $p(f_c | X_c, y_c)$ 为高斯后验分布:

$$p(f_c | X_c, y_c) = \frac{p(y_c | f_c) p(f_c | X_c)}{p(y_c | X_c)} \quad (6)$$

在高斯过程的框架下,标注像素和未标注像素对应函数值 $f_c$ 和 $f_u$ 的联合分布为:

$$\begin{bmatrix} f_c \\ f_u \end{bmatrix} \sim N\left(0, \begin{bmatrix} K_{cc} & K_{cu} \\ K_{uc} & K_{uu} \end{bmatrix}\right) \quad (7)$$

通过对高斯过程进行条件推断,可以得到未标注像素函数值的条件分布:

$$f_u | X_u, X_c, f_c \sim N(K_{uc} K_{cc}^{-1} f_c, K_{uu} - K_{uc} K_{cc}^{-1} K_{cu}) \quad (8)$$

然而,由于医学图像分割属于二元分类任务,像素标签 $y$ 是离散二元变量(0或1),因此似然函数 $p(y_c | f_c)$ 不能使用高斯分布来进行建模,这也就导致后验分布 $p(f_c | X_c, y_c)$ 难以计算. 为此,我们引入拉普拉斯近似法,通过对 $\log p(f_c | X_c, y_c)$ 进行二阶泰勒展开,得到后验分布的高斯近似. 近似后的高斯分布 $q(f_c | X_c, y_c)$ 为:

$$q(f_c | X_c, y_c) = N(f_c | \hat{f}_c, H^{-1}) \quad (9)$$

其中, $\hat{f}_c = \operatorname{argmax}_{f_c} \log p(f_c | X_c, y_c)$ ,是后验分布的最大后验估计值, $H = -\nabla^2 \log p(f_c | X_c, y_c) |_{f_c = \hat{f}_c}$ ,是 $\hat{f}_c$ 处的负 Hessian 矩阵. 通过高斯近似后,积分公式可以简化为:

$$p(f_u | X_u, X_c, y_c) \sim N(\mu_u, \Sigma_u) \quad (10)$$

其中,均值和协方差矩阵分别为:

$$\mu_u = K_{uc} K_{cc}^{-1} \hat{f}_c \quad (11)$$

$$\Sigma_u = K_{uu} - K_{uc} K_{cc}^{-1} K_{cu} + K_{uc} K_{cc}^{-1} H^{-1} K_{cc}^{-1} K_{cu} \quad (12)$$

最后,通过 Sigmoid 函数对预测结果进行处理,得到每个像素为前景(即 $y_u = 1$ )的概率分布,进而生成最终的预测分割结果. 为了保证全图像素的一致性,进一步避免用户的交互误差,我们对包含用户标注像素在内的所有像素都进行了像素标签预测,以获取最后的

预测分割结果。

#### 2.4.2 混合核函数设计

在实际应用中,核函数的选择对高斯过程分类的效果起着至关重要的作用.特别是在处理具有模糊边界和噪声干扰的医学图像时,单一核函数往往难以全面捕捉像素之间复杂且多样化的关系.这是因为单一核函数通常只能关注某一方面的特征,例如图像的空间结构或灰度信息,难以同时兼顾多维特征的差异性和互补性,尤其是在边界模糊、灰度不均匀或含有噪声的图像中表现尤为明显。

为解决这一问题,本文设计了一种结合多种特征的混合核函数,其由径向基核函数(RBF kernel)和马顿核函数(Matern kernel)组成.径向基核函数具有较强的局部敏感性,能够很好地捕捉图像中像素间的细微差异,尤其是在处理灰度变化显著的区域时表现优异.而马顿核函数则能够通过调整平滑性参数 $\nu$ ,在空间相关性和模型复杂度之间找到平衡,适合处理存在噪声或边界模糊的医学图像.具体公式定义如下:

$$k_{\text{mixed}}(\mathbf{x}_i, \mathbf{x}_j) = \alpha_1 \cdot \exp\left(-\frac{\|\mathbf{I}_i - \mathbf{I}_j\|^2}{2\tau^2}\right) + \alpha_2 \cdot \exp\left(1 + \frac{\sqrt{2\nu}d(\mathbf{x}_i, \mathbf{x}_j)}{\ell}\right) \quad (13)$$

其中,  $\mathbf{I}_i$ 和 $\mathbf{I}_j$ 分别表示像素 $i$ 和 $j$ 的像素值信息;  $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 分别表示从深度神经网络提取到的像素 $i$ 和 $j$ 的深度特征向量,  $d(\mathbf{x}_i, \mathbf{x}_j)$ 像素 $i$ 和 $j$ 在深度特征空间中的距离;  $\nu$ 是平滑性参数,针对医学图像常见的复杂性问题,如噪声、模糊边界和复杂结构等,本文采用了 Matern 5/2 核,即 $\nu = \frac{5}{2}$ ,以更灵活的建模像素间相似性;  $\alpha_1$ 和 $\alpha_2$ 是权重系数,  $\tau$ 和 $\ell$ 是用于调整核函数灵活性的超参数.通过结合这两种核函数,模型不仅能够更好地捕捉图像的局部结构信息,还能够面对噪声和模糊边界时具备更强的鲁棒性.具体而言, RBF 核函数有助于精确识别图像中的细节变化,而 Matern 核函数则通过其灵活性适应更广泛的像素间关系,提升分类器对不同复杂场景的适应性和泛化能力.这种混合核函数设计有效提高了模型在医学图像分割任务中的表现,特别是在边界模糊和噪声干扰较大的情况下,能够更准确地进行像素分类,提取目标区域。

#### 2.5 模型损失

模型训练所用复合损失函数 $L$ 结合了分割主任务

的损失和高斯过程的优化目标,以充分利用交互信息并优化分割性能,其具体形式为:

$$L = L_{\text{Main}} + \mu L_{\text{GP}} \quad (14)$$

其中,  $L_{\text{Main}}$ 为主任务损失,用于提升分割任务的区域覆盖和分类精度;  $L_{\text{GP}}$ 为高斯过程损失,用于优化高斯过程分类模块的性能; 损失权重参数 $\mu$ 用于平衡两部分损失的贡献值,本文中设置为0.001。

主任务损失 $L_{\text{Main}}$ 结合了 Dice 损失 $L_{\text{Dice}}$ 和交叉熵损失 $L_{\text{CE}}$ ,以分别优化分割区域的整体覆盖率和像素级分类精度,其定义为:

$$L_{\text{Main}} = \lambda_1 L_{\text{Dice}} + \lambda_2 L_{\text{CE}} \quad (15)$$

其中,  $\lambda_1$ 和 $\lambda_2$ 为权重系数,本文中均设置为0.5.高斯过程损失 $L_{\text{GP}}$ 于优化高斯过程分类模块的后验分布,确保混合核函数能够准确捕捉医学图像中的多尺度特征.其定义为:

$$L_{\text{GP}} = L_{\text{Pred}} + \omega L_{\text{Reg}} \quad (16)$$

其中 $L_{\text{Pred}}$ 为预测损失,采用负对数似然(NLL)来衡量高斯过程模型对训练数据的拟合程度;  $L_{\text{Reg}}$ 为正则化损失,用于限制高斯过程后验分布与先验分布的偏离程度,通过 KL 散度定义;  $\omega$ 为正则化权重,本文设置为0.01。

为了使模型能够从训练数据中自适应调整,本文框架所涉及的关键超参数是以端到端的方式从训练数据中自动学习的,包括用于高斯测地线距离编码的参数 $\sigma$ ,以及用于混合核函数高斯过程分类的 $\alpha_1$ 、 $\alpha_2$ 、 $\tau$ 和 $\ell$ ,以更好地捕捉医学图像的局部细节与全局特性,从而增强分割的准确性与鲁棒性。

### 3 实验

#### 3.1 数据集

本文在 medical segmentation decathlon (MSD) 数据集上验证了所提出的框架. MSD 数据集是一个涵盖了多种医学图像分割任务的大型标准化数据集,是医学图像分割算法研究领域的重要基准之一.该数据集由来自世界各地的研究机构和医疗中心提供,涵盖了10种不同的医学图像分割任务,包括脑部、心脏、肺部、肾脏等多个器官的分割.我们在其中4个具有代表性和挑战性的子集上进行了实验,如表1所示。

(1) 肺部子集(Task06\_Lung),包含63例非小细胞肺癌患者的术前薄层CT扫描.任务目标是分割肺内的

肿瘤. 本文从 63 例 CT 扫描图像中提取到包含分割目标的有效切片 467 片, 随机选取了其中 200 片作为训练集, 并从剩余切片中随机选择了 150 片作为测试集.

(2) 胰腺子集 (Task07\_Pancreas), 包含 281 例胰腺肿块切除患者的门静脉期 CT 扫描. 任务目标是分割胰腺实质和胰腺肿瘤. 本文从 281 例 CT 扫描图像中提取到包含分割目标的有效切片 2 193 片, 随机选取其中的 400 片作为训练集, 并从剩余的切片中随机选择了 200 片作为测试集.

(3) 脾脏子集 (Task09\_Spleen), 包含 61 例接受化疗治疗的肝脏转移瘤患者的门静脉期 CT 扫描. 任务目标是脾脏分割. 本文从 61 例 CT 扫描图像中提取到包含分割目标的有效切片 466 片, 随机选取其中 200 片作为训练集, 并从剩余的切片中随机选择 150 片作为测试集.

(4) 结肠子集 (Task10\_Colon), 包含 126 例接受原发结肠癌切除术患者的门静脉期 CT 扫描. 任务目标是分割结肠癌的原发肿瘤. 本文从 126 例 CT 扫描图像中提取到包含分割目标的有效切片 1 002 片, 随机选取其中的 400 片作为训练集, 并从剩余的切片中随机选择 200 片作为测试集.

表 1 所用 MSD 数据集子集概览

编号	名称	模态	分割目标	训练样本	测试样本
06	Lung	CT	肿瘤	200	150
07	Pancreas	CT	胰腺/肿瘤	400	200
09	Spleen	CT	脾脏	200	150
10	Colon	CT	肿瘤	400	200

### 3.2 评价指标

为了从客观数据角度验证本文算法的有效性, 对算法性能进行定量评估, 本文采用了戴斯系数 (*Dice*) 和平均表面距离 (*ASSD*) 作为衡量模型效果的统计指标. 戴斯系数 (*Dice coefficient*, *Dice*), 也称为戴斯相似性系数 (*Dice similarity coefficient*, *DSC*), 是用于衡量分割的模型预测结果和真实标签之间相似度的统计指标, 其公式如下:

$$Dice = \frac{2|\mathcal{R}_p \cap \mathcal{R}_g|}{|\mathcal{R}_p| + |\mathcal{R}_g|} \quad (17)$$

其中,  $\mathcal{R}_p$  和  $\mathcal{R}_g$  分别表示预测分割区域和真实分割区域.  $| \cdot |$  表示对应区域的像素数. 其结果通常乘以 100 表示为百分比形式, 即 *Dice* (%), 以直观体现预测结果与真实标签之间的相似度, 体现模型分割质量. 平均表面距离

(average symmetric surface distance, *ASSD*) 是用来衡量两个表面 (例如, 模型预测结果的表面和真实标签的表面) 之间平均距离的指标. 该指标反映了两个分割边界之间的对称性差异. 其计算公式如下:

$$ASSD = \frac{\sum_{i \in S_p} d(i, S_g) + \sum_{i \in S_g} d(i, S_p)}{|S_p| + |S_g|} \quad (18)$$

其中,  $S_p$  和  $S_g$  分别表示模型预测结果和真实标签的表面点集,  $d(i, S_g)$  表示点  $i$  到表面  $S_g$  之间的最短欧氏距离. 通常, *ASSD* 以像素 (pix) 为单位. *ASSD* 值越低, 表示分割结果越接近真实分割边界, 分割精度更高.

### 3.3 实验设置

本文在 PyTorch 框架<sup>[36]</sup>上进行实验, 所有实验均在一台配备 RTX 4090GPU (25.2 GB 显存)、16 核 AMD EPYC 9354 处理器、60.1 GB 内存及 751.6 GB 硬盘的 Ubuntu 20.04 系统服务器上进行. 使用的深度学习框架为 PyTorch 1.12.1, 并通过 CUDA 11.6 加速训练过程. 本文利用了 Monai 工具包来预处理医学图像及进行模型训练的优化. 实验所用的所有图像预先进行标准化处理, 将像素强度值归一化到 [0, 1] 范围内, 通过非局部均值滤波进行去噪处理, 统一调整至相同尺寸, 并使用数据增强技术, 如随机旋转、缩放和水平翻转, 以增强模型的鲁棒性.

实验共进行了 300 轮训练, 初始优化器为 Adam, 初始学习率设置为  $1E-4$ , 然后每 20 个 epoch 减半. 此外, 模型中引入了 Dropout 层, 丢弃率为 0.3, 以减少过拟合的风险. 所有实验结果均基于测试集的性能评估. 模型训练过程中, 仅在训练集上进行优化, 而文中所展示的分割性能实验均来自独立测试集的评估. 测试样本的标注信息在训练阶段完全不可见, 以模拟实际临床场景中的泛化能力.

为了验证模型训练的稳定性, 本文记录了损失值随迭代次数的变化情况. 图 2 展示了复合损失  $L$  的收敛曲线. 可以看到, 复合损失  $L$  随着训练轮数的增加逐渐降低, 并在约 150 轮后趋于稳定. 这表明, 本文提出的复合损失函数能够有效引导模型学习, 且优化过程收敛性良好.

### 3.4 参数优化

为了使模型能够根据训练数据自适应调整, 以更好地捕捉医学图像的局部细节与全局特性, 并提高分割的准确性与鲁棒性, 高斯测地线距离编码中的衰减

参数 $\sigma$ , 以及混合核函数高斯过程分类中的参数 $\alpha_1$ 、 $\alpha_2$ 、 $\tau$ 、 $\ell$ , 均通过端到端的方式从训练数据中自动学习. 训练开始前, 所有参数会被赋予合理的默认值作为初始状态. 而后, 参数值会在训练过程中通过反向传播不断更新, 最终收敛至最优值, 从而确保模型的最佳性能. 各参数初始默认值与最终收敛结果如表2所示.

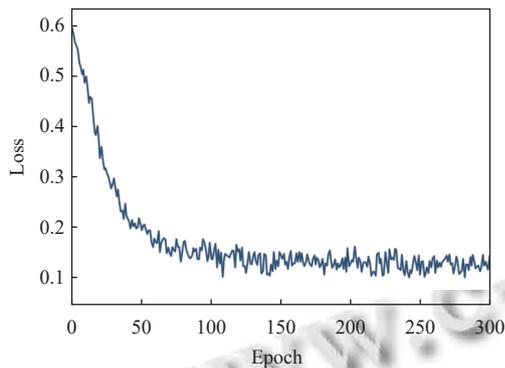


图2 复合损失 $L$ 的收敛曲线

表2 各参数初始值与优化结果

类别	$\sigma$	$\alpha_1$	$\alpha_2$	$\tau$	$\ell$
初始值	1.0	0.5	0.2	1.0	2.0
优化值	0.9	0.6	0.4	0.8	1.8

为了进一步验证自适应优化参数的优越性, 我们将采用自适应优化参数的模型与采用固定参数的模型在 Spleen 数据集与 Colon 数据集上进行直观的性能对比. 实验结果如表3所示. 结果表明, 与固定参数设置相比, 自适应优化后的参数能够更好地适应图像的复杂结构和噪声干扰, 从而提升分割精度.

表3 参数优化前后模型性能对比

数据集	参数设置	<i>Dice</i> (%)	<i>ASSD</i> (pix)
Spleen	固定参数	94.75±2.57	1.73±0.62
	优化参数	96.86±2.34	1.68±0.72
Colon	固定参数	78.94±2.83	4.24±1.95
	优化参数	82.27±3.27	3.12±1.07

### 3.5 模型对比

为了验证本文方法的有效性, 本文与现有的多种交互式分割方法以及自动分割方法在 MSD 数据集的 4 个子集上进行了对比.

#### 3.5.1 对比模型概述

对比模型选取 6 种交互式分割方法 (DEXTR<sup>[11]</sup>、InterCNN<sup>[12]</sup>、DeepIGeoS<sup>[10]</sup>、BS-IRIS<sup>[13]</sup>、MIDeepSeg<sup>[14]</sup>、TIS<sup>[15]</sup>) 以及 1 种自动分割方法 (nnU-Net<sup>[37]</sup>).

对以上模型的说明如下.

(1) DEXTR<sup>[11]</sup>: deep extreme cut (DEXTR) 将用户提供的矩形框选编码为欧几里得距离图作为软约束输入, 结合卷积编码-解码网络进行端到端训练, 并通过条件随机场解决重叠问题.

(2) InterCNN<sup>[12]</sup>: 在自动分割方法基础上训练卷积神经网络, 不限于二分类任务, 能通过模拟真实用户的迭代交互提高编辑性能.

(3) DeepIGeoS<sup>[10]</sup>: 通过一个单独的 CNN 结构来改进已获取的自动分割结果, 利用测地线距离变换将用户输入与 CNN 融合, 并通过保留分辨率的网络和条件随机场 (CRF) 进一步细化分割.

(4) BS-IRIS<sup>[13]</sup>: 将交互式图像分割建模为马尔可夫决策过程 (MDP), 利用强化学习 (RL) 迭代优化, 采用多代理强化学习共享体素级策略, 同时引入边界感知奖励机制, 从而在用户交互提示的基础上高效生成高质量分割掩码.

(5) MIDeepSeg<sup>[14]</sup>: 通过使用指数化测地线距离编码用户提供的内部边缘点, 结合少量用户点击进行信息融合, 实现对新见和未见过对象的高效交互式分割.

(6) TIS<sup>[15]</sup>: 基于 Transformer 架构, 将用户点击的像素与具有相似特征的像素进行分组来细化自动分割结果, 支持同时对多个类别的分割掩码进行细化编辑.

(7) nnU-Net<sup>[37]</sup>: 一种基于深度学习的自动化分割方法, 能够根据不同任务自动配置预处理、网络架构、训练和后处理步骤.

#### 3.5.2 整体性能对比

我们在 MSD 数据集中 4 个子集的 5 项分割任务上对本文方法以及 7 种对比方法的整体性能进行了比较, 实验结果如表4和图3所示. 表4量化对比了 MSD 上 4 个子集的 5 项分割任务上各模型分割性能, 加粗字体表示最优分割精度. 图3直观展示了各方法的最终分割效果. 可以看出, 在 MSD 上 4 个子集的 5 项分割任务上, 本文方法在 *Dice* 系数和 *ASSD* 值方面均表现优异, 尤其是在 Lung 子集和 Pancreas 子集的 Pancreas tumour 分割任务中, 本文方法的 *Dice* 系数分别达到了 87.33% 和 83.31%, 显著优于其他对比方法. 此外, 本文方法在 *ASSD* 值上也保持较低水平, 体现了其在分割精度与边界误差之间的良好平衡. 例如, 在 Pancreas 子集的 Pancreas parenchyma 分割任务和 Colon 子集的分割任务中, 本文方法的 *ASSD* 值分别为

1.96 和 3.12, 均优于大多数对比方法. 相比于较为常规的分割任务 (如 Spleen), 本文方法在复杂分割任务 (如 Pancreas tumour 和 Colon) 中的表现尤为出色, 相较于其他对比方法, *Dice* 系数和 *ASSD* 值均有显著提升. 这

表明我们提出的基于高斯测地线距离编码和混合高斯过程分类的交互式分割方法, 能够有效捕捉用户交互提示信息, 进而实现精准分割, 尤其在应对复杂分割任务时展现出明显的优势.

表 4 所用数据集上各模型分割性能

Method	Lung		Pancreas				Spleen		Colon	
	<i>Dice</i> (%)	<i>ASSD</i> (pix)	Pancreas parenchyma		Pancreas tumour		<i>Dice</i> (%)	<i>ASSD</i> (pix)	<i>Dice</i> (%)	<i>ASSD</i> (pix)
			<i>Dice</i> (%)	<i>ASSD</i> (pix)	<i>Dice</i> (%)	<i>ASSD</i> (pix)				
nnU-Net	64.09	4.86	82.16	3.19	49.34	8.42	77.38	3.37	44.84	8.94
DEXTR	82.06±1.96	2.84±1.24	83.43±3.14	2.26±1.31	74.31±2.43	3.52±1.19	94.18±3.25	2.67±1.36	69.27±2.54	4.97±2.16
InterCNN	80.07±2.65	2.98±1.19	82.31±3.28	2.31±1.12	74.17±2.91	3.92±1.45	95.18±4.70	1.23±0.39	69.58±2.97	4.77±1.79
DeepIGeoS	81.74±1.72	3.16±1.41	82.77±1.51	2.15±0.48	75.36±2.60	4.04±2.36	96.39±2.22	1.58±1.24	70.61±2.46	4.53±2.49
BS-IRIS	81.67±2.14	3.31±1.68	85.16±1.34	2.64±1.04	76.49±2.48	2.97±1.56	95.13±3.34	1.94±0.48	71.27±1.82	4.38±1.91
MIDeepSeg	82.31±3.58	2.92±1.57	84.69±4.03	2.04±0.87	70.34±4.36	3.61±2.07	<b>96.93±1.43</b>	<b>1.18±0.44</b>	71.89±3.09	4.17±2.14
TIS	85.07±1.55	2.49±0.44	87.72±1.28	<b>1.92±0.59</b>	77.91±2.07	2.94±1.64	95.11±2.75	1.57±0.75	76.03±1.68	3.92±1.58
Ours	<b>87.33±2.38</b>	<b>2.31±0.73</b>	<b>89.84±2.04</b>	1.96±0.67	<b>83.31±2.49</b>	<b>2.87±1.36</b>	96.86±2.34	1.68±0.72	<b>82.27±3.27</b>	<b>3.12±1.07</b>

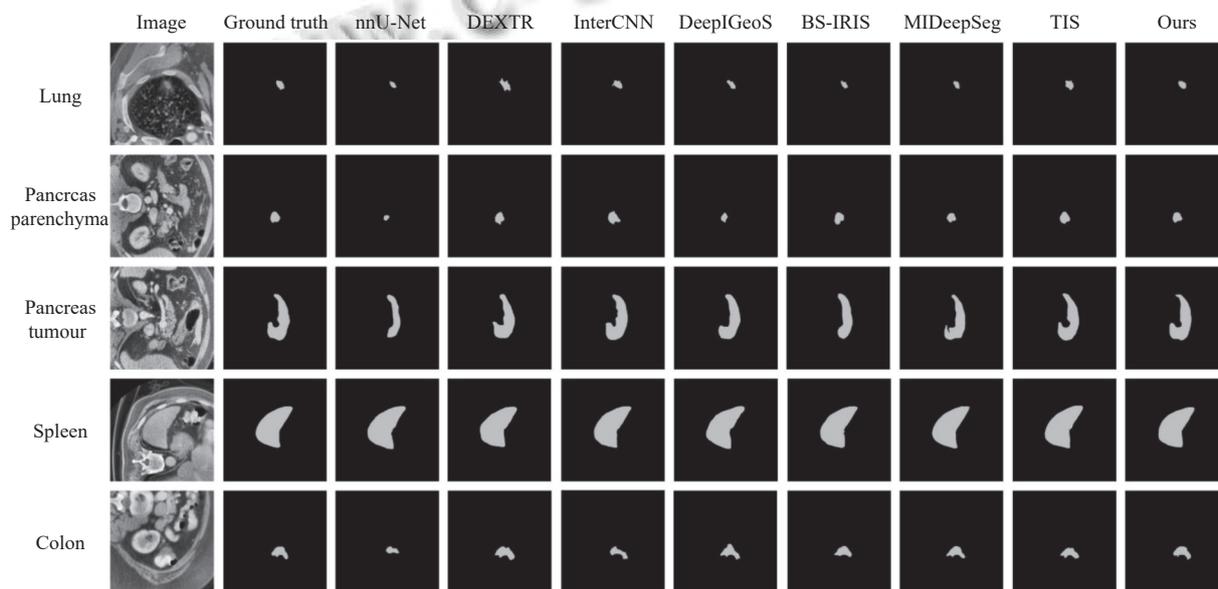


图 3 所用数据集上各模型分割性能的可视化

### 3.5.3 迭代性能对比

在交互式分割任务中, 随着用户点击数量的增加, 分割结果通常会逐步得到改进. 为了验证本文模型的迭代优化效果, 我们在 Pancreas 子集的 Pancreas parenchyma 和 Pancreas tumour 两项分割任务上进行了同一交互序列下不同分割方法的迭代性能对比. 实验结果如图 4 所示, 随着点击标注数量的增加, 本文提出的方法在 *Dice* 系数方面显示出了稳定的性能提升, 明显优于其他对比方法. 这一趋势表明, 我们的方法能够有效利用用户的交互信息, 在每轮点击中逐步细化分割边界, 从而在多次交互后实现更高

的分割精度. 这种持续的性能提升反映了方法在处理复杂医学图像时的优势, 特别是在细节捕捉和边界定义方面.

### 3.5.4 编码性能对比

同时, 为了探究本文在交互编码部分提出的高斯测地线距离编码方法的有效性, 本文将其与欧几里得距离变换 (Euclidean distance variation, EDV)、高斯热图 (Gaussian heatmap, GH) 以及测地线距离变换 (GEO) 进行比较. 对比实验在 MSD 数据集的 Spleen 和 Colon 子集上进行, 所有实验均采用相同的用户混合点击交互, 其他模块保持一致使用本文方法.

实验结果如表 5 所示,可以看出,在两个数据子集上,本文提出的高斯测地线距离编码方法在 *Dice* 系数和 *ASSD* 指标上均取得了优于其他编码方法的表现,

验证了其在交互编码中的有效性和优越性.特别是在边界精度要求较高的任务中,高斯测地线距离编码表现出显著的优势.

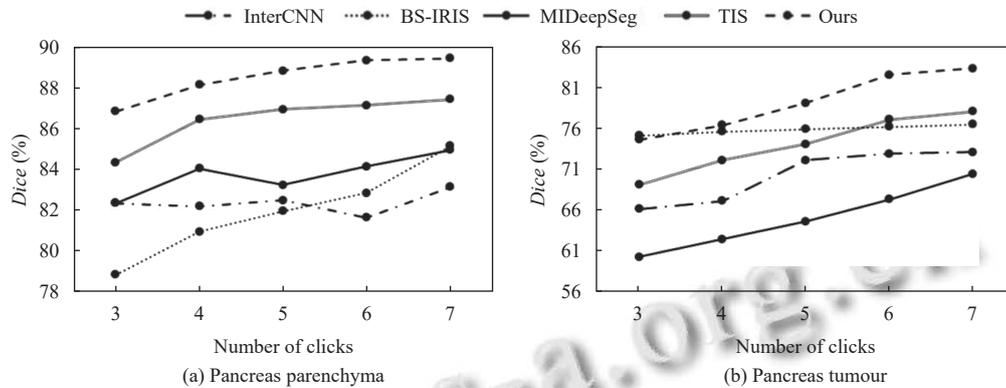


图 4 同一交互序列下不同分割方法的迭代性能对比

表 5 各编码方法在 Spleen 和 Colon 数据集上的定量比较

Method	Spleen		Colon	
	<i>Dice</i> (%)	<i>ASSD</i> (pix)	<i>Dice</i> (%)	<i>ASSD</i> (pix)
EDV	87.42±6.54	3.13±1.76	74.56±5.98	4.02±2.30
GH	88.47±3.27	2.34±1.42	75.91±6.18	3.86±2.43
GEO	94.17±6.23	2.21±0.91	80.17±6.38	3.37±2.13
Ours	<b>96.86±2.34</b>	<b>1.68±0.72</b>	<b>82.27±3.27</b>	<b>3.12±1.07</b>

### 3.6 消融实验

为了充分验证本文提出的方法中各个模块的必要性和有效性,本文进行了详细的消融实验,通过逐步移除或替换模型中的关键模块,评估各模块对整体性能能的贡献,具体设计如下.

基线模型(完整方法):包括融合点击交互、高斯测地线距离编码、特征提取模块和基于混合核函数高斯过程分类的像素分类模块.

消融实验 1:替换混合点击交互设计.该实验中,移除“近中心点+外边缘点”的交互形式,替换为传统的“前景点+背景点”交互.通过对比性能差异,评估融合交互设计的贡献.

消融实验 2:移除高斯测地线距离编码.在此实验中,不对用户交互信息进行任何编码操作,直接与待分割图像进行级联,输入特征提取网络,以验证高斯测地线距离编码对用户交互信息的捕捉效果.

消融实验 3:替换基于混合核函数高斯过程分类的像素分类模块.该实验将混合核函数高斯过程分类器替换为常见的 CNN 分类器,评估高斯过程分类器在像素分类任务中的表现优势.

表 6 展示了在 Spleen 和 Colon 两个子集上基线模

型及 3 个消融实验的 *Dice* 系数和 *ASSD* 值对比结果.对比基线模型,去除模块后的 3 个消融实验在分割准确性上都有明显下降.

表 6 各版本消融模型在 Spleen 和 Colon 数据集上的定量比较

模型版本	Spleen		Colon	
	<i>Dice</i> (%)	<i>ASSD</i> (pix)	<i>Dice</i> (%)	<i>ASSD</i> (pix)
消融实验1	94.17±4.27	2.41±1.39	77.45±6.22	3.78±2.47
消融实验2	90.12±5.47	3.15±1.43	74.03±5.47	4.06±2.13
消融实验3	91.58±3.98	2.54±1.45	76.78±5.84	3.63±2.34
基线模型	<b>96.86±2.34</b>	<b>1.68±0.72</b>	<b>82.27±3.27</b>	<b>3.12±1.07</b>

消融实验 1 中,替换“近中心点+外边缘点”的混合点击交互设计为传统“前景点+背景点”后,*Dice* 系数略有降低,但 *ASSD* 值增加显著,尤其在 Colon 数据集中,*Dice* 系数降低 4.82%,*ASSD* 值增加了 0.66,表明在复杂分割任务中,本文提出的混合交互设计更有助于提高分割精度和边界准确性,而传统交互形式在复杂分割任务中的表现则较为不足.其次,在消融实验 2 中移除高斯测地线距离编码导致了最为显著的性能下降.在 Spleen 数据集中,*Dice* 系数下降了 6.74%,*ASSD* 值增加了 1.47 pix.而在 Colon 数据集中,*Dice* 系数下降了 8.24%,*ASSD* 值增加了 0.94 pix.表明高斯测地线距离编码在有效捕捉用户交互中的聚焦指示信息和增强边界精度方面起到关键作用.在消融实验 3 中,Spleen 数据集的 *Dice* 下降 5.28%,*ASSD* 增加 0.86 pix,而 Colon 数据集的 *Dice* 下降 5.49%,*ASSD* 增加 0.51 pix.这充分说明像素分类模块通过高斯过程分类引入用户

交互信息,使模型拥有更强的稳定性和边界处理能力。

综合上述实验结果,本文提出的混合点击交互设计、高斯测地线距离编码方法以及基于混合核函数高斯过程分类的像素分类模块,在常规和复杂分割任务中,对于整体分割精度及边界精度的提升均发挥了重要作用。

#### 4 结论与展望

深度学习在医学图像交互式分割领域得到了广泛应用,但现有的交互设计与编码方法尚未完全满足医学图像分割的需求,且在像素分类过程中未能充分利用用户的交互信息,导致模型性能受到限制。为了解决这一问题,本文提出了一种基于改良测地线距离编码和混合高斯过程分类的医学图像交互式分割方法。通过高斯衰减函数对测地线距离进行加权,减弱图像噪声的干扰,增强交互编码的鲁棒性和准确性,并结合混合点击交互设计,精准定位用户关注的区域。在像素分类中,利用高斯过程分类器结合混合核函数捕捉医学图像中不同尺度的特征,进一步提升分类的准确性,并为模型提供一定的可解释性。为验证本文方法的有效性,我们在MSD数据集的4个子集的5项分割任务上进行了广泛实验,取得了优异结果。未来,我们将致力于减少医学图像分割模型对训练样本量的依赖,优化交互效率和模型的普适性,探索更多样化的交互设计,以提升模型在更具挑战性的医学图像分割任务中的表现。

#### 参考文献

- 1 Olabarriaga SD, Smeulders AWM. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 2001, 5(2): 127–142. [doi: [10.1016/S1361-8415\(00\)00041-4](https://doi.org/10.1016/S1361-8415(00)00041-4)]
- 2 Zhao F, Xie XH. An overview on interactive medical image segmentation. *Annals of the BMVA*, 2013, 2013(7): 1–22.
- 3 刘宇, 陈胜. 医学图像分割方法综述. *电子科技*, 2017, 30(8): 169–172.
- 4 Shelhamer E, Long J, Darrell T, *et al.* Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. [doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683)]
- 5 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 6 段杰, 崔志明, 沈艺, 等. 一种改进 FCN 的肝脏肿瘤 CT 图像分割方法. *图学学报*, 2020, 41(1): 100–107.
- 7 周涛, 董雅丽, 刘珊, 等. 用于肺部肿瘤图像分割的跨模态多编码混合注意力 U-Net. *光子学报*, 2022, 51(4): 0410006. [doi: [10.3788/gzxb20225104.0410006](https://doi.org/10.3788/gzxb20225104.0410006)]
- 8 Neelam B, Palakayala PK, Mbangweta K, *et al.* FCN based deep learning architecture for medical image segmentation. *Proceedings of the 2nd International Conference on Edge Computing and Applications (ICECAA)*. Namakkal: IEEE, 2023. 556–562. [doi: [10.1109/ICECAA58104.2023.10212108](https://doi.org/10.1109/ICECAA58104.2023.10212108)]
- 9 陈成, 张正, 肖迪. 基于改进 U-Net 网络的新冠病毒 CT 图像分割. *计算机仿真*, 2023, 40(7): 154–158, 452. [doi: [10.3969/j.issn.1006-9348.2023.07.028](https://doi.org/10.3969/j.issn.1006-9348.2023.07.028)]
- 10 Wang GT, Zuluaga MA, Li WQ, *et al.* DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(7): 1559–1572. [doi: [10.1109/TPAMI.2018.2840695](https://doi.org/10.1109/TPAMI.2018.2840695)]
- 11 Xu N, Price BL, Cohen S, *et al.* Deep grabcut for object selection. *arXiv:1707.00243*, 2017.
- 12 Bredell G, Tanner C, Konukoglu E. Iterative interaction training for segmentation editing networks. *Proceedings of the 9th International Workshop on Machine Learning in Medical Imaging*. Granada: Springer, 2018. 363–370. [doi: [10.1007/978-3-030-00919-9\\_42](https://doi.org/10.1007/978-3-030-00919-9_42)]
- 13 Ma CF, Xu QS, Wang XF, *et al.* Boundary-aware supervoxel-level iteratively refined interactive 3D image segmentation with multi-agent reinforcement learning. *IEEE Transactions on Medical Imaging*, 2021, 40(10): 2563–2574. [doi: [10.1109/TMI.2020.3048477](https://doi.org/10.1109/TMI.2020.3048477)]
- 14 Luo XD, Wang GT, Song T, *et al.* MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning. *Medical Image Analysis*, 2021, 72: 102102. [doi: [10.1016/j.media.2021.102102](https://doi.org/10.1016/j.media.2021.102102)]
- 15 Liu WT, Ma CF, Yang YH, *et al.* Transforming the interactive segmentation for medical imaging. *Proceedings of the 25th International Conference on Medical Image Computing and Computer-assisted Intervention*. Singapore: Springer, 2022. 704–713. [doi: [10.1007/978-3-031-16440-8\\_67](https://doi.org/10.1007/978-3-031-16440-8_67)]
- 16 Grady L. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(11): 1768–1783. [doi: [10.1109/TPAMI.2006.233](https://doi.org/10.1109/TPAMI.2006.233)]
- 17 Yushkevich PA, Piven J, Hazlett HC, *et al.* User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*,

- 2006, 31(3): 1116–1128. [doi: [10.1016/j.neuroimage.2006.01.015](https://doi.org/10.1016/j.neuroimage.2006.01.015)]
- 18 Criminisi A, Sharp T, Blake A. GeoS: Geodesic image segmentation. Proceedings of the 10th European Conference on Computer Vision. Marseille: Springer, 2008. 99–112. [doi: [10.1007/978-3-540-88682-2\\_9](https://doi.org/10.1007/978-3-540-88682-2_9)]
- 19 Wang GT, Zuluaga MA, Pratt R, *et al.* Dynamically balanced online random forests for interactive scribble-based segmentation. Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens: Springer, 2016. 352–360. [doi: [10.1007/978-3-319-46723-8\\_41](https://doi.org/10.1007/978-3-319-46723-8_41)]
- 20 Wang GT, Zuluaga MA, Pratt R, *et al.* Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. Medical Image Analysis, 2016, 34: 137–147. [doi: [10.1016/j.media.2016.04.009](https://doi.org/10.1016/j.media.2016.04.009)]
- 21 Rother C, Kolmogorov V, Blake A. “GrabCut”—Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics, 2004, 23(3): 309–314. [doi: [10.1145/1015706.1015720](https://doi.org/10.1145/1015706.1015720)]
- 22 Rajchl M, Lee MCH, Oktay O, *et al.* DeepCut: Object segmentation from bounding box annotations using convolutional neural networks. IEEE Transactions on Medical Imaging, 2017, 36(2): 674–683. [doi: [10.1109/TMI.2016.2621185](https://doi.org/10.1109/TMI.2016.2621185)]
- 23 Maninis KK, Caelles S, Pont-Tuset J, *et al.* Deep extreme cut: From extreme points to object segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 616–625.
- 24 Wang GT, Li WQ, Zuluaga MA, *et al.* Interactive medical image segmentation using deep learning with image-specific fine-tuning. IEEE Transactions on Medical Imaging, 2018, 37(7): 1562–1573. [doi: [10.1109/TMI.2018.2791721](https://doi.org/10.1109/TMI.2018.2791721)]
- 25 Hu Y, Soltoggio A, Lock R, *et al.* A fully convolutional two-stream fusion network for interactive image segmentation. Neural Networks, 2019, 109: 31–42. [doi: [10.1016/j.neunet.2018.10.009](https://doi.org/10.1016/j.neunet.2018.10.009)]
- 26 Koohbanani NA, Jahanifar M, Tajeddin NZ, *et al.* Leveraging transfer learning for segmenting lesions and their attributes in dermoscopy images. arXiv:1809.10243, 2018.
- 27 Sakinis T, Milletari F, Roth H, *et al.* Interactive segmentation of medical images through fully convolutional neural networks. arXiv:1903.08205, 2019.
- 28 Li ZW, Chen QF, Koltun V. Interactive image segmentation with latent diversity. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 577–585. [doi: [10.1109/CVPR.2018.00067](https://doi.org/10.1109/CVPR.2018.00067)]
- 29 Xu N, Price B, Cohen S, *et al.* Deep interactive object selection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas IEEE, 2016. 373–381. [doi: [10.1109/CVPR.2016.47](https://doi.org/10.1109/CVPR.2016.47)]
- 30 Wang Z, Acuna D, Ling H, *et al.* Object instance annotation with deep extreme level set evolution. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7492–7500. [doi: [10.1109/CVPR.2019.00768](https://doi.org/10.1109/CVPR.2019.00768)]
- 31 Khan S, Shahin AH, Villafraua J, *et al.* Extreme points derived confidence map as a cue for class-agnostic interactive segmentation using deep neural network. Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention. Shenzhen: Springer, 2019. 66–73. [doi: [10.1007/978-3-030-32245-8\\_8](https://doi.org/10.1007/978-3-030-32245-8_8)]
- 32 Chen LC, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv:1412.7062, 2014.
- 33 Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239. [doi: [10.1109/CVPR.2017.660](https://doi.org/10.1109/CVPR.2017.660)]
- 34 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615)]
- 35 Zhou MH, Wang H, Zhao Q, *et al.* Interactive segmentation as Gaussian process classification. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 19488–19497.
- 36 Paszke A, Gross S, Massa F, *et al.* PyTorch: An imperative style, high-performance deep learning library. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 721.
- 37 Isensee F, Jaeger PF, Kohl SAA, *et al.* nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18(2): 203–211. [doi: [10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z)]

(校对责编:王欣欣)