

多模式交互式图像分割

Zheng Lin

TMCC, College of Computer Science, Nankai
University
Tianjin, China
frazer.linzheng@mail.nankai.edu.cn

Zhao Zhang*

SenseTime Research
Shanghai, China
zzhang@mail.nankai.edu.cn

Ling-Hao Han

TMCC, College of Computer Science, Nankai
University
Tianjin, China
lhhan@mail.nankai.edu.cn

Shao-Ping Lu

TMCC, College of Computer Science, Nankai
University
Tianjin, China
slu@nankai.edu.cn

摘要

对于现在数据紧缺的医学图像分析模型来说,大规模的像素级图像标注是十分缺乏的。为了快速获取标注信息,迫切需要一种经济高效的交互式医学图像分割方法。然而,由于它们的交互方式无法处理医学图像中各种固有的歧义性,目前的技术在很多情况下都会失效,例如不规则的形状和模糊的边界。为了解决这一问题,我们提出了一种多模式交互的医学图像分割框架,其中可以选择不同的交互模式,并允许其相互协作来共同发挥作用。在我们的框架中用户可以根据目标结构复杂程度用各种初始交互方式包围目标区域。然后在初始分割的基础上,用户可以综合利用区域和边界的交互来修复不同歧义性导致的错误标注区域。

*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '22, October 10–14, 2022, Lisboa, Portugal

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9203-7/22/10...\$15.00

<https://doi.org/10.1145/3503161.3548096>

我们在广泛的医学图像上评估我们的框架,其中包括x光、CT图像、MRI图像、超声图像、内窥镜图像和病灶照片。充分的实验结果和用户调研表明该框架可用于各种真实场景下的医学图像标注。

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); Interaction design process and methods; • Computing methodologies → Artificial intelligence.

KEYWORDS

图像分割; 多模式; 人机交互

ACM Reference Format:

Zheng Lin, Zhao Zhang, Ling-Hao Han, and Shao-Ping Lu. 2022. 多模式交互式图像分割. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*, October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3503161.3548096>

1 介绍

交互式医学图像分割 (Interactive Medical Image Segmentation, IMIS) 旨在使用最少的用户交互来标注出

医疗图像中感兴趣的区域。IMIS 在目前的计算机辅助诊断 (Computer-aided Diagnosis, CAD) 系统中起着至关重要的作用, 它对于各种 CAD 算法 [7, 15, 57] 的训练数据标注也是必不可少的。它还可以帮助医护人员从头开始或基于粗略的自动分割的选区精确地突出特定区域来进行疾病进展的定量测量 [45]。考虑到医学影像的特殊性, 只有训练有素、经验丰富的工作人员才能胜任对应的标注工作。往往这样的专业人员非常稀缺, 尤其是在发生紧急医疗事故时, 比如 0COVID-19 的大流行等。在短期内, 寻找专业人员为 CAD 系统标注足够的数据既昂贵又具有挑战性。因此, IMIS 算法中以最小的交互成本快速、准确地完成像素级标注是非常重要的。

几十年来, 科学界一直致力于这项研究 [8]。到目前为止, 为了高效地获取用户的交互意图, 提出了多种方法来接收用户输入的包围框 (bounding box, Bbox) [42, 51], 边界关键点 [21], 点击/涂鸦 [3, 28, 53] 等信息。与自然图像不同, 医学图像的目标区域存在各种严重的歧义性 [62], 比如不规则的形状, 模糊的边界, 以及与相似的背景等等。不幸的是, 单一的交互模式不能有效地处理这些歧义性。例如, 很难使用包围框来精确定位不规则形状的区域, 也很难使用区域单击/涂鸦来定位模糊边界下的目标轮廓。理想的 IMIS 框架应该能高效地标注, 并灵活地选择多种交互方式, 并允许它们相互合作以克服多种歧义性。这种潜在的方法直到现在还没有得到很好的研究。

在本文中, 我们通过探索不同的交互模式, 提出了一个统一的交互框架, 以实现灵活、准确的医学图像分割。如图. 1所示, 初始分割模块接收包围框、包围多边形或包围涂鸦来生成初始预测。用户可以根据目标区域的结构复杂程度灵活选择初始输入。在这些模式中, 包围框初始定位目标的成本较低, 而包围多边形和涂鸦通过引入更多的交互, 可以有效降低歧义性, 使交互映射更接近目标区域轮廓。如果用户对初始分割结果不满意, 我们的细节修复模块 (detail refinement block, DRB) 将协同使用区域和边界点击/涂鸦修复错误标记的区域。将区域和边界相互作用相结合, 改进了现有基于区域 [28, 51, 53] 和边界 [3, 18, 24] 的局部

修复方法无法解决的多重模糊问题。除了修复初始分割, 细节修复模块还支持修复其他自动分割算法生成的粗糙分割图。当预先分割图给出时, 这将是一个减少交互负担的重要特性。简单地说, 与以往的工作相比, 以下因素使我们的模型更好地适用于 IMIS:

- (1) 我们提出了一种多模式交互的思想, 用户可以自由选择并联合使用合适的交互模式来处理分割中的不同歧义。
- (2) 在上述思想下, 我们的方法提供了不同的初始交互模式, 可以在初始定位目标实例时灵活地平衡精度和时间成本。
- (3) 在初始定位后, 我们提出区域-边界协同交互, 可以准确定位边界, 并快速修复错误标记的区域。

我们将我们的框架应用于多个医学图像标注任务, 如 COVID-19 感染和皮肤病变等。实验结果和用户调研表明, 我们的框架对不同的歧义性是稳健的, 是对现实世界的图像标注的一个有意义的探索。

2 相关背景

2.1 用户交互模式

2.1.1 初始定位. 包围框 (Bbox) 是定位对象的有效交互样式。用户可以通过确定矩形对角线上的两个端点来提供包含对象的区域。许多交互式图像分割 (IIS) 方法 [6, 11, 25, 42, 44, 56] 将包围框作为用户输入来分割所包含的实例。GrabCut [44] 分别对 box 内外的颜色分布进行建模。将 box 内部的像素重新分配给前景或背景, 通过多次迭代的颜色建模和像素重新分配得到修复的结果。DenseCut [11] 为了快速 IIS, 用密集连接的 CRF 替换了 GrabCut 中的全局颜色模型。Benenson *et al.* [6] 训练端到端 CNN, 它接收 Bbox 的图像, 并输出分割选区。考虑到 Bbox 可能会在场景中的重叠物体上冲突, IOG [60] 引入了一个额外的锚点来定位 box 内的目标对象。包围涂鸦在物体周围画出一个比较宽松的轮廓。和 Bbox 相比, 它离物体更近, 并提供形状线索。Pizenberg *et al.* [41] 提出了一个类似于 GrabCut 的 IIS 方法, 利用包围涂鸦代替 box, 在移动终端上实现

更高效的分割。边界关键点同样是广泛用于初始分割的一种交互方法 [18, 37]。智能剪刀 [39] 通过寻找加权图中边界关键点之间的最小代价路径来分割目标。点击雕刻 [18] 使用边界关键点选择 MCG [2] 生成的对象建议。DEXTR [37] 以目标实例的四个极值点作为实例分割的位置和边界线索。Le *et al.* [24] 引入边界监督, 提高边界感知。最近基于点击的 IIS 方法 [10, 19, 27, 30, 33, 36, 48, 58] 使用用户交互的前背景点分割实例。如 [32] 所示, 这些方法中的首次点击可以作为初始分割的潜在定位线索。

2.1.2 细节修复. IIS 方法在获得初始分割结果后, 通常进入第二阶段通过进一步的交互输入对局部细节区域进行修复, 其中可分为基于边界的交互和基于区域的交互。基于边界的交互在精确的边界拟合方面效果好。在这些方法中, 用户可以在错误标记的边界上提供额外的点击/涂鸦 [3, 24]。与它们不同的有, Lazy Snapping [26], Polygon-RNN [9], Polygon-RNN++ [1], 以及 Curve-GCN [34] 这些方法允许用户拖动初始阶段生成的多边形的顶点来匹配实例边界。基于区域的交互更适合错误标记区域的快速修复。他们要求用户在错误标记的区域上进一步提供点击/涂鸦 [6, 44, 58]。例如, GrabCut [44] 使用涂鸦或点作为下一轮迭代的硬约束。KnifeCut [31] 是专门为修复薄物体而提出的。

在细节修复过程中, 基于边界的交互能够很好地拟合复杂的边界, 但不适用于错误标记区域的快速修复。相反, 基于区域的交互方法可以快速处理标记错误的区域, 但当边界模糊或复杂时, 它们往往不可靠。遗憾的是结合不同交互风格的优势快速准确地修复初始分割并没有被社区提出和重视。

2.2 交互式医学图像分割

交互式医学图像分割 (Interactive Medical Image Segmentation, IMIS) 是 IIS 的一个特殊领域。这些方法旨在根据医学图像的特殊性, 以最少的交互来准确地标记物体。这里我们将这些方法按照其工作流程分为两类: 从头开始标记和通过粗略初始化的自动分割修复。

2.2.1 从头开始. 给定一幅图像, 根据用户简单的输入获取目标选区是 IMIS 方法的基本要求。DeepCut [42] 接收标有包围框的医学图像, 在类似于 GrabCut 的迭代过程中, 利用卷积神经网络 (CNN) 对前景/背景分布进行建模。Slic-Seg [55] 允许用户首先在 MRI 的单帧进行交互, 然后将 GraphCut 扩展到 4D 版本, 通过进一步的交互共同分割目标区域。DyBa-ORF [54] 以用户的涂鸦作为输入, 学习一个通用的动态平衡在线随机森林来解决不平衡的前背景涂鸦。BIFSeg [51] 针对训练集中不可见的医学图像域设计了一个特定于图像的微调。这种微调也可以与进一步输入的涂鸦一起工作, 以修复精确的细节。CM [21] 使用四个极点作为用户输入, 并对它们进行编码, 生成十字置信图, 与图像一起输入到 CNN 进行预测。IteR-MRL [28] 将三维 IMIS 表述为马尔可夫决策过程, 并使用强化学习求解。

2.2.2 从自动产生的结果改进. 自动分割算法可以为 IMIS 方法提供粗略的定位信息, 利用这些粗略的初始化是降低交互成本的有效方法。DeepIGeoS [53] 设计了一个两阶段的框架, 其中一个 CNN 用于自动分割, 另一个 CNN 是用用户输入的测地线距离来修复初始分割。遵循这种两阶段的想法, iW-Net [3] 接收两个点来计算由前景点电荷的产生的引力场激发的置信图。RefineNet [23] 在 U-Net 和全卷积网络等多层次的自动分割网络特征上附加交互式流。Zhou *et al.* [63] 在第二阶段的修复中引入膨胀/收缩操作。UGIR [50] 首先利用基于分组卷积的 CNN 获得初始分割。Then, this work effectively estimates the uncertainty score of each slide, and only relatively uncertain slides will require further user interactions. 然后, 这项工作有效地估计每一个切片的不确定性得分, 只有相对不确定的切片才需要进一步的用户交互。

3 提出的方法

3.1 框架结构

我们构建了一个统一的交互框架, 将多种交互模式有机地集成在一起, 以解决医学图像分割任务中的各种

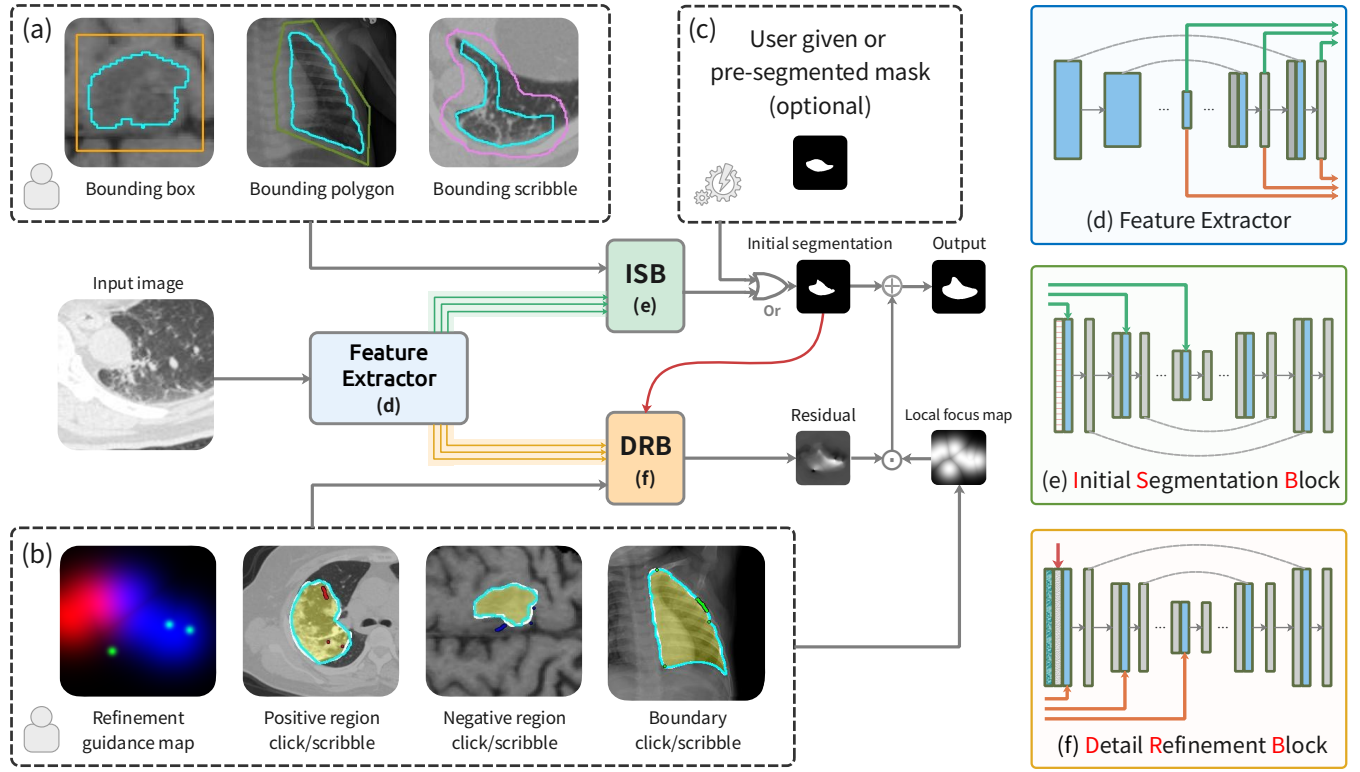


图 1: 我们框架的工作流程 (a) 说明了对于初始分割的交互。如 (c) 所示, 初始分割模块的初始分割可以被自动生成或手动标记的选区所取代。如图 (b) 所示, 在初始分割的基础上, 用户可以为区域和边界提供进一步的交互。然后将结果输送到细节修复模块进行迭代改进。对于可视化图像, 黄色选区是预测结果, 青色轮廓线是实际值。

歧性问题。如图. 1所示, 提出的框架可以分为三部分, 即特征提取器、初始分割模块 (ISB) 和细节修复模块 (DRB)。特征提取部分以裁剪后的图像作为输入, 提取医学图像特征。初始分割模块的目的是借助初始指导图获得更准确的分割结果。最后在修复阶段, 根据用户交互情况对粗糙分割结果进行修复。在我们的框架中, 三个部分中的相同分辨率的特征通过跨层连接直接通信。这些连接使初始分割模块和细节修复模块能够共享多级强大的表示同时保持轻量级结构。

3.1.1 特征提取器. 该特征提取器是一种类似 FPN 的编码器-解码器架构。它以裁剪后的医学图像为输入, 从解码器中提取多段 32 通道特征。具体来说, 我们采用了 VGG-16 骨干网络 [46] 的 5 层卷积层, 并额外增加了 32 个通道的层。我们将这些特征从低到高分别表示为 $\{F_i\}_{i=1}^5$ 和 G_6 。对应的解码器特征 $\{G_i\}_{i=1}^5$ 可以

从以下得到:

$$G_i = D_i((G_{i+1}) \uparrow \oplus R_i(F_i)), \quad i \in \{5, 4, 3, 2, 1\}. \quad (1)$$

在上式中, $(\cdot) \uparrow$ 是上采样操作。 $R_i(\cdot)$ 卷积 F_i 到 32 通道。在解码器中, $D_i(\cdot)$ 由两个卷积层和 32 个核组成。

3.1.2 初始分割模块. 初始分割模块接受用户输入的包围框、多边形或涂鸦来输出初始分割结果。如图. 1(d), 它是一个高效的沙漏形块, 这得益于通过具有相同分辨率的层之间的跨层连接共享表示。与特征提取器相连接的编码器可以形式化为

$$F_i^{ISB} = \begin{cases} E_i^{ISB}(G_i \oplus A^{ISB}), & i = 1 \\ E_i^{ISB}(G_i \oplus (F_{i-1}^{ISB}) \downarrow), & i \in \{2, 3, 4, 5\} \end{cases}, \quad (2)$$

其中 A^{ISB} 是初始指导图。 $(\cdot) \downarrow$ 是下采样操作, $E_i^{ISB}(\cdot)$ 是一个有 32 个核的卷积层。为了增加到原来的分辨率,

解码器可以表示为

$$\mathbf{G}_i^{\text{ISB}} = \begin{cases} D_i^{\text{ISB}}(\mathbf{G}_i \oplus \mathbf{F}_{i-1}^{\text{ISB}}), & i = 6 \\ D_i^{\text{ISB}}((\mathbf{G}_{i+1}^{\text{ISB}}) \uparrow \oplus (\mathbf{F}_i^{\text{ISB}})), & i \in \{5, 4, 3, 2, 1\} \end{cases}, \quad (3)$$

其中 $D_i^{\text{ISB}}(\cdot)$ 由两个卷积层和 32 个核组成。我们在 $\mathbf{G}_1^{\text{ISB}}$ 上使用 3×3 卷积; 然后对预测图进行阈值化, 得到二值的初始化的分割结果 \mathbf{M}^{ISB} 。

3.1.3 细节修复模块. 细节修复模块 (图. 1(f)) 的设计类似于初始分割模块, 区别在于输入输出部分细节修复模块连接初始分割 \mathbf{M}^{ISB} (来自初始分割模块或其他自动分割方法), 前景/背景区域交互图和边界交互图作为输入。细节修复模块生成残差来修复粗糙分割。它使用与式. (2) 和式. (3) 相同的方法获得 $\{\mathbf{F}_i^{\text{DRB}}\}_{i=1}^5$ 和 $\{\mathbf{G}_i^{\text{DRB}}\}_{i=1}^6$ 。残差 \mathbf{R}^{DRB} 可以从使用 1×1 卷积的 $\mathbf{G}_1^{\text{DRB}}$ 得到。迭代的最终结果 \mathbf{M} 可以从下式获得

$$\mathbf{M} = T(\mathbf{M}^{\text{ISB}} + \mathbf{U} \odot \mathbf{R}^{\text{DRB}}). \quad (4)$$

这里 $T(\cdot)$ 是二值化函数, \odot 是元素乘法运算符。如图. 1所示, \mathbf{U} 是局部焦点图。在我们的工作中, 这是一个标注的半径为 80 的高斯图, 它是为了减少局部修复交互的全局影响而设计的。

3.2 交互和用户界面

在本节中, 我们将介绍在我们的多模式交互框架中使用的交互方式 (图. 1(a)-(b)) 和用户界面 (UI) (图. 2)。

3.2.1 初始交互. 如图. 1(a) 我们为用户提供了便捷的方式来定位目标区域。它支持包围框、包围多边形和包围涂鸦。初始化指导图可以从这些交互中生成。它由三个通道组成, 会根据交互方式激活其中一个通道。具体地说, 当我们采用一种模式时, 根据绘制方式的内外区域, 将对应通道内的像素设为 1 或 -1, 而其他不相关的通道设为 0。

包围框. 如图. 1(a) 包围框是一个围绕目标区域的矩形。作为一种低成本的交互方式, 用户只需要确定矩形对角线上的两个端点。它可以处理大多数圆形和矩形形状的目标。

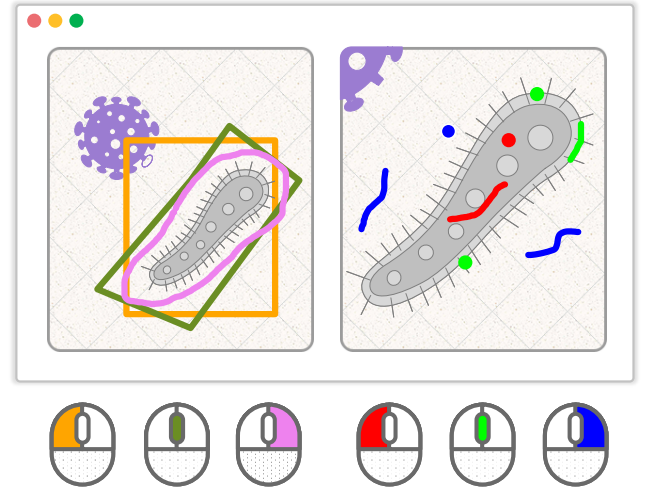


图 2: 用户交互及界面原型。用户首先在左侧窗口提供初始交互。框/多边形/涂鸦可以通过鼠标左键/中键/右键来点击或拖动获得。然后, 右侧窗口将提供缩放后的局部区域。前景/背景/边界的局部交互对应于左键/中键/右键的点击或拖动涂鸦。在修复过程中可以不断迭代进行, 直到得到满意的结果。

包围多边形. 包围多边形是一种对包围框的自然扩展方式, 将矩形替换为任意多边形。显然, 多边形比矩形更适合目标区域。它减少了无关背景的干扰, 比包围框更容易得到更好的引导。然而, 这些好处是以更大的交互性为代价的, 它需要用户确定多边形的所有顶点, 而不是像包围框那样只确定对角线顶点。

包围涂鸦. 除了包围框和多边形, 我们的框架还提供了包围涂鸦, 可以更灵活地包围目标。如图. 1(a) 所示, 当目标区域轮廓很复杂时候时, 包围涂鸦可以更好包裹区域, 并提供目标形状的先验信息。

图. 2 展示了我们的交互界面原型。在我们的 UI 中, 初始交互应该首先在左侧窗口的图像上提供。用户可以用鼠标左键绘制一个包围框。以同样的方式, 他们可以单击中键来提供包围多边形的顶点, 或长按右键来提供包围涂鸦。之后 UI 将在初始交互的基础上在右侧窗口显示局部放大的图像和初始分割结果。

3.2.2 修复交互. 在我们的框架中, 用户可以提供前景区域点击/涂鸦、背景区域点击/涂鸦和边界点击/涂鸦。这三个交互激活了对应的三个不同的修复指导图通道, 这些通道将随着初始分割被送入细节修复模块。

区域点击/涂鸦. 用户可以根据初始分割模块或给定的初始分割, 使用区域单击/涂鸦纠正错误的区域。前景区域的交互能修复标注错误的前景预测, 背景区域的交互能消除错误的背景预测。与涂鸦相比, 点击错误标记的前景和背景是一种低成本的交互方式; 而涂鸦比点击则包含更多的指导信息。区域交互引导图在标记像素周围具有一定的高斯半径 (80 像素)。对于前景或背景的交互, 两者的图是相互独立的。

边界点击/涂鸦. 医学图像经常面临严重的边界模糊。用户需要根据自己的专业经验来确定具体的边界位置。为了精确的边界定位, 上面区域的点击/涂鸦无法给出有效的约束。只有通过不断使用前景、背景区域交互来收紧边界, 才能实现精确定位。在实际的交互中是费时费力的。因此, 我们将引入边界点击/涂鸦与区域交互协同, 灵活地解决各种模糊问题。与区域指导图相似, 边界制导图在用户标记的像素周围具有较小的高斯半径 (10 像素)。

如图. 2中展示的修复细节的 UI, 用户可以在右侧窗口反复添加交互来修复分割结果。具体来说, 添加的前景、背景和边界点击/涂鸦对应于左键、右键和中键的点击/长按。

3.3 仿真机器人用户

为了对模型进行客观的训练和评估, 我们需要一个机器人用户来模拟真实用户的交互。我们在附录 A中介绍了针对不同交互模式的机器人用户的策略。

4 实验

4.1 数据集 & 度量标准

我们采用了包含不同类型的医学图像的六个数据集。他们分别是 **COVID19-CT**[57] (新冠肺炎/电子计算机

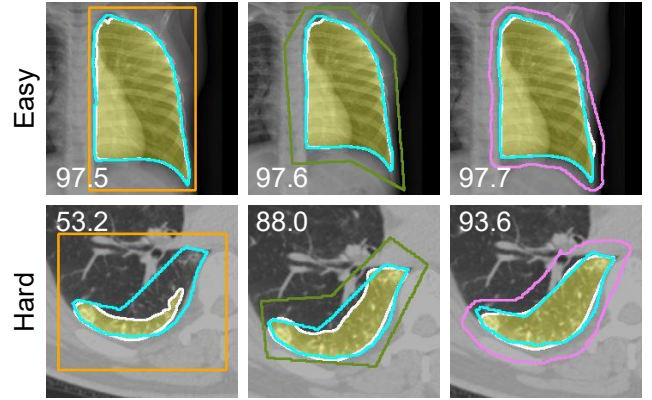


图 3: 在不同交互下的初始分割第一行是一个简单的例子, 目标对象具有规则的形状; 它们都能很好地定位对象 (角落里的白色数字是 Dice 的值). 如果目标区域是不规则的形状, 如第二行, 那么包围框就会面临很大的歧义, 导致不正确的分割。

断层扫描图像), **COVID19-Xray**[49] (COVID19/X 光图像), **BraTS-T1**[4, 5, 38] (脑肿瘤/核磁共振图像), **Nerve** (臂丛/超声图像), **Polyp** (息肉/内窥镜图像), **ISIC**[12] (皮肤病变/病灶照片图像). 我们采用常用的 Dice 评分、ASSD 和 mNoI 来评估我们的方法。详情请参考附录 B。

4.2 实现细节

我们以 VGG-16 [46] 为骨干并设置批大小为 16。我们将初始学习率设置为 1×10^{-4} 并使用权重衰减为 5×10^{-4} 的 Adam 方法 [22] 进行优化。我们在训练中使用二元交叉熵损失函数。我们采用多项式学习速率衰减策略。所有的实验都是用 PyTorch [40] 框架实现的, 并运行在单个 NVIDIA Titan XP GPU 上。每次交互的测试速度为 1.5×10^{-2} 秒, 对于真实的应用来说已经足够满足实时操作需求。

4.3 消融实验 & 讨论

在本节中, 我们验证了不同交互模式对这些广泛使用的医疗数据的影响。

4.3.1 初始交互性能. 我们首先在一个简单和一个困难的标注场景中研究初始交互。总的来说, 包围框/包围

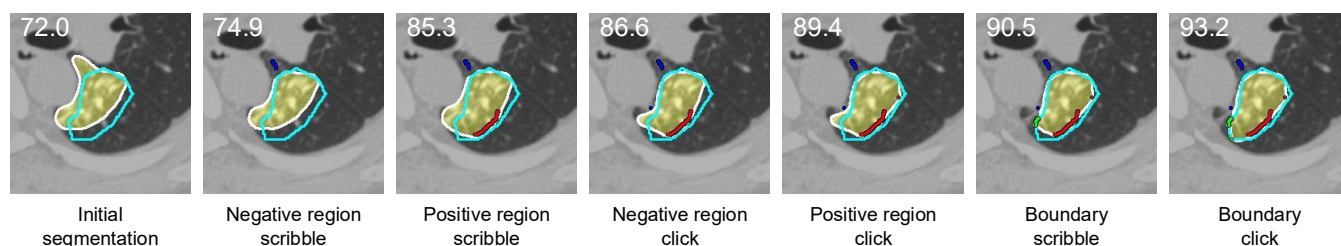


图 4: 基于不同修复交互模式协同工作可视样例。左上角的数字代表 Dice 得分。

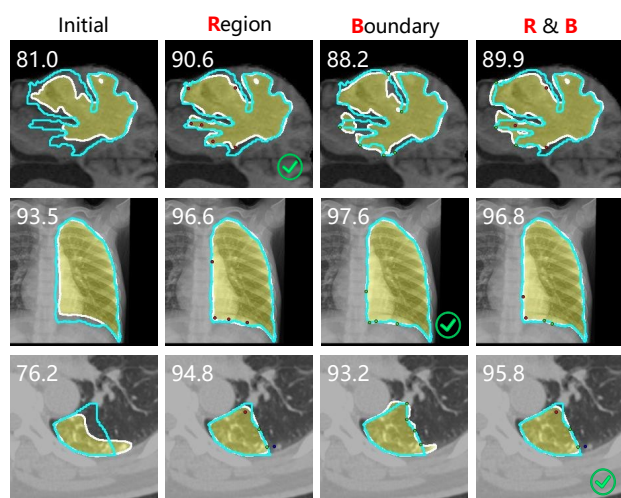


图 5: 区域与边界交互模式独自工作或相互合作的对比。白色为 Dice 指标, 打勾的为性能最佳选择。不同的医学图像场景面临不同的歧义性。在第一行, 这些初始分割有很大的区域被错误预测。第二行包含不确定边界。最后一行面临着更有挑战性的情况, 他们边界模糊, 区域难以区分。不同的场景适用不同的交互模式, 有时在更具挑战性的问题下, 区域和边界交互协作能取得更好的效果。

多边形/包围涂鸦的交互负担是逐渐增加的。同时, 他们带来的定位效果也会相应地更加准确。绘制包围框, 用户只需要确定两个关键点。这种轻量级交互方法可以处理规则形状的区域。如图 3 的第一行, 利用包围框可以获得与包围多边形/包围涂鸦相似的结果。当场景变得复杂时, 包围框就不能胜任了。例如, 第二行中目标对象是不规则形状的。包围框不可避免地包含了背景区域, 这反过来导致了不准确的预测。在这种

情况下, 包围多边形和包围涂鸦可以灵活地靠近目标边界, 同时带来更多的形状先验。因此, 他们可以得到更准确的结果。它也反映在表 1 的初始 Dice 得分。这就是为什么我们在最初的分割阶段允许多种交互。用户可以根据场景的复杂程度选择合适的模式, 避免了特定交互模式能力不足或交互成本过高的问题。

4.3.2 修复交互性能. 图 4 中, 我们通过联合使用六种交互模式 (包括前景区域点击/涂鸦、背景区域点击/涂鸦和边界点击/涂鸦) 来进行了一个实际的修复演示。从结果中, 我们可以看到涂鸦比点击更精确, 这是因为涂鸦以更高的交互成本为代价提供了更多的指导。通过比较不同类型的模式可以发现, 前景区域的交互可以很容易地修复错误标记的前景区域, 而背景区域的交互则可以很好地消除背景上的错误预测。用户可以利用这些交互来解决前景分布不均匀和背景相似区域干扰造成的歧义性问题。与区域交互不同, 边界交互带来的效果也不同。我们可以看到, 无论是边界点击还是涂鸦, 预测的边界都会固定到交互位置, 并延伸到两边, 寻找可能的轮廓。由于医学图像经常面临模糊不清的轮廓, 这一特性非常有利于医学图像的标注。解决这种歧义性需要有经验的医生使用边界点击或涂鸦来确定边界位置, 而在这种情况下使用区域交互将会大概率失败。在表 1, 我们定量评估细节修复模块在修复初始分割中的效果。我们可以看到修复交互在初始分割基础上带来了普遍的精度显著提高。

为什么细节修复需要多个交互模式? 如图 5, 我们列举了一些例子, 在这些例子中, 不同的交互模式共同作用或单独作用来修复最初的分割。每种情况下的迭代都在同一轮结束。我们可以看到, 区域交互、边

界交互和混合作用适用于不同的场景。此外,从表.1,我们可以看到区域交互在前几次迭代中表现良好。当预测逐渐修复时,区域相互作用不能准确定位边界,性能提升不如边界交互。

该框架可用于修复给定的粗糙选区,如自动算法预测等。在这一节中,我们展示了 COVID19-CT [57] 和 COVID19-Xray [49] 数据集上的修复性能。对 ASSD 的改进列在表.3中。可以看出,我们的方法可以对已有的预测进行稳定的修复。在现实世界的标注任务中,如新冠肺炎标注,基于自动预测进行交互 [14, 52, 59] 将减少手动标注的负担。在临床诊断中,我们的方法,配合自动算法,有助于产生更准确的分割掩膜,并提高了疾病感染的定量评估性能。

4.4 现有方法的比较

在本节中,我们将我们的框架的两个阶段,即初始分割和细节修复阶段,与现有相关方法进行比较。对于初始分割,我们比较了四种开源的交互式图像分割方法: GrabCut [44], Random Walks (RW) [16], FCA-Net [32], f-BRS [48] 以及三种自动分割算法: U-Net [43], FCN [35], and DSS [17]。从表.1可以看出,我们的方法远远超出了传统方法,如 GrabCut 和 Random Walks。此外,良好的初始化交互也使我们的方法性能优于 FCA-Net 和 f-BRS。在与表.2中的自动分割算法进行比较时,我们的三种初始交互模式带来了有用的指导信息,从而获得了更好的分割结果。对于细节修复,我们将我们的框架与上述包含修复阶段的方法进行比较。结果记录在表.1。实验结果表明,在相同的迭代次数下,该方法取得了较好的分割效果。

5 用户调研

我们进行了一项用户调研,以调查我们提出的方法在现实标注中的有效性。我们主要比较了每个样本的平均标注时间与当前最先进的方法的对比,并展示在表.4中。我们还探索了现实场景中多种交互模式的协作是否便利,并将相关数据展示在表.5中。和 [29] 一样,我们从 COVID19-CT、BraTS-T1 和 Nerve 数据集中随机抽

取 50 张图像进行该用户调研。我们为这项调研招募了五名参与者。要求每个参与者使用这三种初始交互和修复交互模式来标注这些样本。当 Dice 分数超过 90% 或达到 30 秒的最大限制时,停止对每个样本的时间记录。

在初始分割交互模式中,包围框是最受欢迎的选择,据统计占 81%。它的平均交互时间为 1.44 秒。包围多边形和包围涂鸦耗时分别为 3.64 秒和 2.25 秒,以用于定位复杂的物体。从表.5中,我们可以看到 17% 的样本仅使用初始交互实现了 90% 的 Dice 指标。

在细节修复中,点击时间约为 1.21 秒,涂鸦时间约为 1.82 秒(其中包括判断、交互和模型推断时间)。尽管涂鸦需要比点击更长的时间,但它给 Dice 带来的平均收益却更大。此外,25% 的样本被涂鸦和点击共同标注。参与者倾向于用涂鸦修复大的错误标记区域,用点击修复小的区域。当涉及到边界和区域交互的选择时,我们可以看到区域交互实现了更多的 Dice 指标的提升,因为它经常用于快速修复大的错误标记区域。边界交互也是精确定位边界的有效方式。在大多数情况下,初始分割足够好,33% 的样本仅通过边界交互进行修复。此外,22% 的样本同时使用区域和边界的协同作用模式。从结果可以看出,我们提供的交互模式在实际标注中是实用的。

我们还将我们的方法与 FCA-Net 和 f-BRS (它们的代码是公开的)进行比较,并在表.4中记录平均标注时间。我们的框架为每个图像节省了大约 5 秒的时间,快了近 40%,有效地减少了交互负担。

在分割修复的详细过程中,我们的模型提供了区域/边界和点击/涂鸦交互模式,它们可以共同协作来修复被错误标记的区域。在这里,我们进一步说明了用户在实际标注过程中的选择。统计结果表明,我们的方法符合用户的自然行为。

区域 vs. 边界. 我们认为区域交互的相互作用更有利于大错误区域的快速修复,而边界交互的相互作用则用来固定边界位置以获得更精确的结果。用户调研结果证实了我们的观点。如图.6(a)所示,我们计算了标注过程中区域和边界相互作用的比例。从图中可

表 1: 在 Dice 指标上交互模式的消融研究及与其他方法的比较 (%). 我们将这些交互模式缩写为: 包围框 (Bbox); 包围多边形 (Bpoly); 包围涂鸦 (Bscri); 区域点击 (RC); 区域涂鸦 (RS); 边界点击 (BC); 边界涂鸦 (BS). 在比较方法中, “-C” 和 “-S” 意味着分别基于 “点击” 和 “涂鸦” 进行交互。“0、5、10” 表示当每个实例上有 0、5 或 10 个交互时在一个数据集上 Dice 得分。mNoI (@95%) 是交互的平均数量。

| | COVID19-CT [57] | | | | COVID19-Xray [49] | | | | BraTS-T1 [5] | | | | Nerve [20] | | | | ISIC [12] | | | | Polyp [13] | | | |
|------------|-----------------|------|------|------|-------------------|------|------|------|--------------|------|------|------|------------|------|------|------|-----------|------|------|------|------------|------|------|------|
| | 0 | 5 | 10 | mNoI | 0 | 5 | 10 | mNoI | 0 | 5 | 10 | mNoI | 0 | 5 | 10 | mNoI | 0 | 5 | 10 | mNoI | 0 | 5 | 10 | mNoI |
| Bbox+RC | 84.8 | 93.4 | 95.2 | 8.4 | 96.0 | 98.0 | 98.3 | 0.4 | 81.0 | 89.7 | 92.4 | 14.8 | 81.9 | 93.6 | 96.0 | 7.2 | 88.5 | 92.9 | 93.8 | 9.9 | 85.9 | 91.1 | 92.1 | 9.7 |
| Bbox+RS | 84.8 | 95.2 | 96.3 | 5.6 | 96.0 | 98.4 | 98.6 | 0.2 | 81.0 | 93.0 | 94.5 | 11.0 | 81.9 | 96.0 | 97.2 | 4.1 | 88.5 | 95.5 | 95.7 | 5.7 | 85.9 | 94.4 | 94.6 | 7.8 |
| Bbox+BC | 84.8 | 93.9 | 96.2 | 6.2 | 96.0 | 98.2 | 98.6 | 0.3 | 81.0 | 88.9 | 91.8 | 13.5 | 81.9 | 94.4 | 96.9 | 5.1 | 88.5 | 94.9 | 96.0 | 5.1 | 85.9 | 93.7 | 95.3 | 4.0 |
| Bbox+BS | 84.8 | 95.7 | 97.8 | 4.3 | 96.0 | 98.6 | 98.9 | 0.3 | 81.0 | 92.0 | 95.0 | 8.8 | 81.9 | 96.5 | 98.3 | 3.8 | 88.5 | 96.1 | 97.0 | 4.5 | 85.9 | 96.6 | 98.1 | 2.8 |
| Bpoly+RC | 86.6 | 94.0 | 95.8 | 6.9 | 96.8 | 98.2 | 98.5 | 0.1 | 85.4 | 91.3 | 93.5 | 12.8 | 87.2 | 94.8 | 96.7 | 5.7 | 91.6 | 94.7 | 95.5 | 6.6 | 91.4 | 94.2 | 94.7 | 8.5 |
| Bpoly+RS | 86.6 | 95.8 | 97.1 | 4.0 | 96.8 | 98.5 | 98.7 | 0.1 | 85.4 | 93.9 | 95.4 | 8.5 | 87.2 | 96.9 | 97.9 | 2.8 | 91.6 | 96.5 | 96.6 | 3.3 | 91.4 | 95.7 | 95.9 | 5.9 |
| Bpoly+BC | 86.6 | 94.2 | 96.5 | 5.8 | 96.8 | 98.3 | 98.7 | 0.1 | 85.4 | 90.5 | 92.5 | 12.7 | 87.2 | 95.9 | 97.7 | 3.9 | 91.6 | 96.0 | 96.9 | 3.9 | 91.4 | 96.6 | 97.8 | 2.6 |
| Bpoly+BS | 86.6 | 96.2 | 98.3 | 3.7 | 96.8 | 98.7 | 99.0 | 0.1 | 85.4 | 93.2 | 96.0 | 7.4 | 87.2 | 97.2 | 98.6 | 3.1 | 91.6 | 96.6 | 97.4 | 3.5 | 91.4 | 97.4 | 98.7 | 2.0 |
| Bscri+RC | 91.6 | 95.8 | 96.8 | 3.2 | 98.7 | 98.7 | 98.7 | 0.0 | 91.8 | 93.9 | 94.9 | 8.4 | 92.5 | 95.9 | 97.2 | 3.6 | 92.5 | 95.1 | 95.8 | 6.3 | 93.5 | 95.2 | 95.6 | 6.5 |
| Bscri+RS | 91.6 | 97.0 | 97.5 | 1.7 | 98.7 | 98.9 | 98.9 | 0.0 | 91.8 | 95.6 | 96.2 | 5.3 | 92.5 | 97.8 | 98.1 | 1.2 | 92.5 | 96.8 | 96.9 | 2.8 | 93.5 | 96.4 | 96.4 | 4.3 |
| Bscri+BC | 91.6 | 96.1 | 97.0 | 2.7 | 98.7 | 98.9 | 98.9 | 0.0 | 91.8 | 93.8 | 94.6 | 8.5 | 92.5 | 97.3 | 98.1 | 1.8 | 92.5 | 96.3 | 97.1 | 3.7 | 93.5 | 97.0 | 98.0 | 1.9 |
| Bscri+BS | 91.6 | 97.4 | 98.6 | 1.7 | 98.7 | 99.0 | 99.0 | 0.0 | 91.8 | 95.6 | 97.2 | 4.1 | 92.5 | 98.0 | 98.7 | 1.5 | 92.5 | 96.8 | 97.5 | 3.4 | 93.5 | 97.8 | 98.7 | 1.4 |
| GrabCut-C | 14.6 | 20.3 | 35.0 | 19.9 | 12.9 | 13.3 | 13.8 | 20.0 | 43.6 | 44.2 | 45.1 | 20.0 | 0.1 | 0.2 | 0.4 | 20.0 | 72.2 | 73.0 | 73.9 | 18.2 | 39.9 | 40.3 | 41.7 | 19.7 |
| GrabCut-S | 14.6 | 83.1 | 92.0 | 13.8 | 12.9 | 88.0 | 95.2 | 9.2 | 43.6 | 73.9 | 87.5 | 18.3 | 0.1 | 83.9 | 93.2 | 13.1 | 72.2 | 91.5 | 95.1 | 8.5 | 39.9 | 82.7 | 93.8 | 10.5 |
| GraphCut-C | 1.0 | 5.2 | 11.7 | 20.0 | 15.1 | 15.8 | 12.9 | 20.0 | 3.5 | 2.6 | 6.6 | 20.0 | 2.6 | 3.2 | 2.6 | 20.0 | 13.8 | 20.0 | 14.0 | 19.9 | 5.1 | 6.8 | 5.4 | 20.0 |
| GraphCut-S | 35.7 | 84.8 | 92.3 | 13.9 | 30.9 | 84.1 | 94.8 | 9.5 | 30.7 | 79.3 | 89.2 | 17.9 | 29.0 | 83.9 | 93.3 | 13.6 | 19.2 | 83.3 | 93.3 | 12.4 | 16.2 | 87.2 | 95.1 | 9.5 |
| RW-C | 4.5 | 64.7 | 85.8 | 19.3 | 48.9 | 78.8 | 85.1 | 19.7 | 9.4 | 65.9 | 85.1 | 19.7 | 10.0 | 66.5 | 88.1 | 19.0 | 40.5 | 78.3 | 88.2 | 18.1 | 20.8 | 72.0 | 88.7 | 18.2 |
| RW-S | 44.6 | 88.3 | 94.3 | 10.9 | 57.7 | 87.7 | 94.7 | 10.8 | 25.6 | 85.7 | 92.0 | 15.2 | 44.0 | 88.2 | 95.6 | 9.0 | 52.6 | 91.3 | 95.6 | 8.9 | 61.9 | 91.2 | 96.2 | 7.8 |
| FCA-Net | 81.0 | 92.4 | 94.0 | 3.0 | 96.0 | 97.9 | 98.4 | 0.4 | 80.2 | 83.7 | 84.9 | 19.1 | 79.4 | 90.5 | 93.8 | 11.8 | 83.2 | 94.7 | 96.1 | 4.9 | 82.3 | 94.7 | 96.5 | 4.7 |
| f-BRS | 80.0 | 90.2 | 92.1 | 11.8 | 95.6 | 97.2 | 97.9 | 0.5 | 79.7 | 67.7 | 85.7 | 18.2 | 78.1 | 90.5 | 90.6 | 13.3 | 83.0 | 94.4 | 95.9 | 4.6 | 85.4 | 94.3 | 96.5 | 4.3 |

以看出, 在错误标注区域较大的前期, 用户更倾向于选择区域交互, 而在后期, 用户更倾向于使用边界交互来固定边界。曲线趋势与我们的观点是一致的。

涂鸦 vs. 点击. 与探索区域和边界交互的关系类似, 我们也探索用户的涂鸦和点击的选择。据统计, 用户涂鸦的时间比点击要长, 但收益会更大。在实际操

作中, 较大的错误区域更容易绘制涂鸦, 而较小的错误区域则更容易通过点击修复。如图. 6(b) 所示, 这些因素使用户在优化的早期阶段选择涂鸦, 并在后期阶段选择点击, 多种交互的联合使用, 让用户更自然地完成分割。

表 2: 比较了初始分割交互方法和自动分割方法的性能 (Dice (%)).

| | FCN [35] | U-Net [43] | DSS [17] | Bbox | Bpoly | Bscri |
|----------|----------|------------|----------|------|-------|-------|
| ISIC | 79.4 | 74.9 | 80.5 | 88.5 | 91.6 | 92.4 |
| Nerve | 47.1 | 50.7 | 60.3 | 81.9 | 87.2 | 92.5 |
| BraTS-T1 | 40.7 | 59.9 | 54.5 | 81.0 | 85.4 | 91.8 |

表 3: 修复 U-Net 生成的粗糙掩膜 (ASSD 指标).

| | COVID19-CT [57] | | | COVID19-Xray [49] | | |
|-------|-----------------|-----|-----|-------------------|-----|-----|
| U-Net | 0 | 1 | 5 | 0 | 1 | 5 |
| + RC | 26.7 | 9.5 | 7.9 | 5.7 | 5.4 | 4.3 |
| + RS | 26.7 | 9.3 | 7.9 | 5.7 | 5.0 | 3.6 |
| + BC | 26.7 | 8.7 | 6.7 | 5.7 | 5.5 | 4.3 |
| + BS | 26.7 | 8.5 | 6.5 | 5.7 | 5.3 | 3.7 |

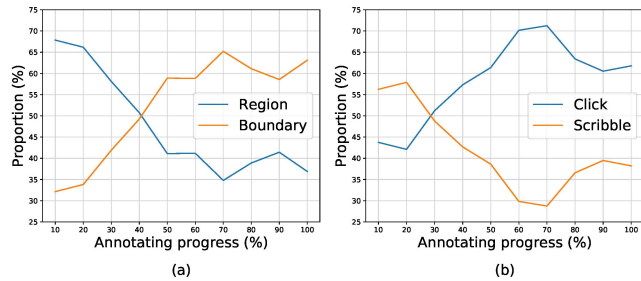


图 6: 真实用户交互分析。(a) 标注过程中区域与边界相互作用的比例 (b) 标注过程中点击和涂鸦交互的比例。

6 结论 & 未来工作

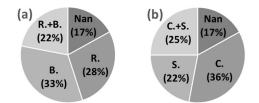
在本文中, 我们提出了一种多模式交互式图像分割框架, 以解决医学图像分割中各种图像模糊, 如不规则形状、模糊边界和相似背景干扰。与目前的 IMIS 方法相比, 这是第一次系统地将多种交互方式集成到一个统一的网络中。充分的实验结果和用户调研表明, 我们的方法是一个灵活有效地解决现实标注任务的方法。

表 4: 用户调研: 三个数据集中真实用户交互时间的比较。该值为包含模型推断时间的每幅图像的平均交互时间。

| | COVID19-CT [57] | BraTS-T1 [5] | Nerve [20] | All |
|-------------|-----------------|--------------|------------|-------|
| f-BRS[48] | 12.26 | 14.82 | 9.29 | 12.12 |
| FCA-Net[32] | 10.77 | 19.93 | 6.14 | 12.28 |
| Ours | 6.11 | 10.39 | 5.74 | 7.41 |

表 5: 用户调研: 真实场景中的交互分析. R. = 区域交互, B. = 边界交互, C. = 点击, S. = 涂鸦, Time: 平均交互时间; Ratio: 交互的比值; Δ Dice: 交互对 Dice 均值的提高; (a): 不使用任何东西, 只使用 R, 只使用 B, 并将 R 和 B 结合在一起的图片的比值; (b): 不使用任何东西, 只使用 C, 只使用 S, 并将 C 与 S 相结合的图片的比值。

| | R. | B. | C. | S. |
|------------------|------|------|------|------|
| Time(s) | 1.55 | 1.30 | 1.21 | 1.82 |
| Ratio(%) | 50.7 | 49.3 | 64.1 | 35.9 |
| Δ Dice(%) | 3.4 | 2.9 | 2.4 | 4.5 |



未来, 我们将继续改进所提模型的泛化, 并将我们的框架扩展到 3D 数据标注。

ACKNOWLEDGMENTS

This work is funded by the National Key Research and Development Program of China (NO. 2018AAA0100400), NSFC (NO. 61922046), The Fundamental Research Funds for the Central Universities, Nankai University, China Postdoctoral Science Foundation (NO.2021M701780).

REFERENCES

- [1] David Acuna, Huan Ling, Amlan Kar, and Sanja Fidler. 2018. Efficient interactive annotation of segmentation datasets with polygon-RNN++. In *CVPR*. 859–868.

- [2] Pablo Arbeláez, Jordi Pont-Tuset, Jonathan T Barron, Ferran Marques, and Jitendra Malik. 2014. Multiscale combinatorial grouping. In *CVPR*. 328–335.
- [3] Guilherme Aresta, Colin Jacobs, Teresa Araújo, António Cunha, Isabel Ramos, Bram van Ginneken, and Aurélio Campilho. 2019. iW-Net: an automatic and minimalistic interactive lung nodule segmentation deep network. *Scientific reports* 9, 1 (2019), 1–9.
- [4] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozyski, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data* 4 (2017), 170117.
- [5] Spyridon Bakas, Mauricio Reyes, and So on. 2018. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. arXiv:1811.02629 [cs.CV]
- [6] Rodrigo Benenson, Stefan Popov, and Vittorio Ferrari. 2019. Large-scale interactive object segmentation with human annotators. In *CVPR*. 11700–11709.
- [7] Jannis Born, Gabriel Brändle, Manuel Cossio, Marion Disdier, Julie Goulet, Jérémie Roulin, and Nina Wiedemann. 2020. POCOD-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). *arXiv preprint arXiv:2004.12084* (2020).
- [8] James F Brinkley. 1991. Spatial anatomic knowledge for 2-D interactive medical image segmentation and matching.. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, 460.
- [9] Lluís Castrejón, Kaustav Kundu, Raquel Urtasun, and Sanja Fidler. 2017. Annotating object instances with a polygon-RNN. In *CVPR*. 5230–5238.
- [10] Xi Chen, Zhiyan Zhao, Feiwei Yu, Yilei Zhang, and Manni Duan. 2021. Conditional Diffusion for Interactive Segmentation. In *ICCV*.
- [11] Ming-Ming Cheng, Victor Adrian Prisacariu, Shuai Zheng, Philip HS Torr, and Carsten Rother. 2015. Densecut: Densely connected crfs for realtime grabcut. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 193–201.
- [12] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kallou, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. 2017. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). arXiv:1710.05006 [cs.CV]
- [13] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. 2020. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. *MICCAI* (2020).
- [14] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. 2020. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. *IEEE Transactions on Medical Imaging* (2020).
- [15] Gusztáv Gaál, Balázs Maga, and András Lukács. 2020. Attention U-Net based adversarial architectures for chest x-ray lung segmentation. *arXiv preprint arXiv:2003.10304* (2020).
- [16] Leo Grady. 2006. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 11 (2006), 1768–1783.
- [17] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. 2019. Deeply Supervised Salient Object Detection with Short Connections. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 4 (2019), 815–828.
- [18] Suyog Dutt Jain and Kristen Grauman. 2019. Click Carving: Interactive Object Segmentation in Images and Videos with Point Clicks. *International Journal of Computer Vision* 127, 9 (2019), 1321–1344.
- [19] Won-Dong Jang and Chang-Su Kim. 2019. Interactive image segmentation via backpropagating refinement scheme. In *CVPR*. 5297–5306.
- [20] Kaggle. 2020. Ultrasound Nerve Segmentation | Kaggle. <https://www.kaggle.com/c/ultrasound-nerve-segmentation>
- [21] Shadab Khan, Ahmed H Shahin, Javier Villafruela, Jianbing Shen, and Ling Shao. 2019. Extreme Points Derived Confidence Map as a Cue for Class-Agnostic Interactive Segmentation Using Deep Neural Network. In *MICCAI*. Springer, 66–73.
- [22] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [23] Titinunt Kitrungratsakul, Iwamoto Yutaro, Lanfen Lin, Ruofeng Tong, Jingsong Li, and Yen-Wei Chen. 2020. Interactive Deep Refinement Network for Medical Image Segmentation. arXiv:2006.15320 [cs.CV]
- [24] Hoang Le, Long Mai, Brian Price, Scott Cohen, Hailin Jin, and Feng Liu. 2018. Interactive boundary prediction for object selection. In *ECCV*. 18–33.
- [25] Victor Lempitsky, Pushmeet Kohli, Carsten Rother, and Toby Sharp. 2009. Image segmentation with a bounding box prior. In *ICCV*. IEEE, 277–284.
- [26] Yin Li, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. 2004. Lazy snapping. *ACM Transactions on Graphics* 23, 3 (2004), 303–308.
- [27] Zhuwen Li, Qifeng Chen, and Vladlen Koltun. 2018. Interactive image segmentation with latent diversity. In *CVPR*. 577–585.
- [28] Xuan Liao, Wenhao Li, Qisen Xu, Xiangfeng Wang, Bo Jin, Xiaoyun Zhang, Yanfeng Wang, and Ya Zhang. 2020. Iteratively-Refined Interactive 3D Medical Image Segmentation with Multi-Agent Reinforcement Learning. In *CVPR*. 9394–9402.
- [29] Jun Hao Liew, Scott Cohen, Brian Price, Long Mai, Sim-Heng Ong, and Jiashi Feng. 2019. MultiSeg: Semantically Meaningful, Scale-Diverse Segmentations From Minimal User Input. In *ICCV*. 662–670.
- [30] Zheng Lin, Zheng-Peng Duan, Zhao Zhang, Chun-Le Guo, and Ming-Ming Cheng. 2022. FocusCut: Diving Into a Focus View in Interactive Segmentation. In *CVPR*.
- [31] Zheng Lin, Zheng-Peng Duan, Zhao Zhang, Chun-Le Guo, and Ming-Ming Cheng. 2022. KnifeCut: Refining Thin Part Segmentation with Cutting Lines. In *ACM MM*.
- [32] Zheng Lin, Zhao Zhang, Lin-Zhuo Chen, Ming-Ming Cheng, and Shao-Ping Lu. 2020. Interactive Image Segmentation with First Click Attention. In *CVPR*. 13339–13348.
- [33] Zheng Lin, Zhao Zhang, Zi-Yue Zhu, Deng-Ping Fan, and Xia-Lei Liu. 2022. Sequential Interactive Image Segmentation.
- [34] Huan Ling, Jun Gao, Amlan Kar, Wenzheng Chen, and Sanja Fidler. 2019. Fast interactive object annotation with curve-GCN. In *CVPR*. 5257–5266.
- [35] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*. 3431–3440.
- [36] Soumajit Majumder and Angela Yao. 2019. Content-aware multi-level guidance for interactive instance segmentation. In *CVPR*. 11602–11611.

- [37] Kevis-Kokitsi Maninis, Sergi Caelles, Jordi Pont-Tuset, and Luc Van Gool. 2018. Deep extreme cut: From extreme points to object segmentation. In *CVPR*. 616–625.
- [38] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. 2014. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34, 10 (2014), 1993–2024.
- [39] Eric N Mortensen and William A Barrett. 1995. Intelligent scissors for image composition. In *Annual conference on Computer Graphics and Interactive Techniques*. 191–198.
- [40] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*. 8024–8035.
- [41] Matthieu Pizenberg, Axel Carlier, Emmanuel Faure, and Vincent Charvillat. 2017. Outlining objects for interactive segmentation on touch devices. In *ACM MM*. 1734–1742.
- [42] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. 2016. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE Transactions on Medical Imaging* 36, 2 (2016), 674–683.
- [43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer, 234–241.
- [44] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "Grab-Cut" interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* 23, 3 (2004), 309–314.
- [45] Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaofei Han, Zhong Xue, and Yuxin Shi. 2020. Lung infection quantification of COVID-19 in ct images with deep learning. *arXiv preprint arXiv:2003.04655* (2020).
- [46] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [47] Jack Sklansky. 1982. Finding the convex hull of a simple polygon. *Pattern Recognition Letters* 1, 2 (1982), 79–83.
- [48] Konstantin Sofiiuk, Ilia Petrov, Olga Barinova, and Anton Konushin. 2020. f-BRS: Rethinking Backpropagating Refinement for Interactive Segmentation. In *CVPR*. 8623–8632.
- [49] V7Labs. 2020. COVID-19 xray dataset. <https://github.com/v7labs/covid-19-xray-dataset>
- [50] Guotai Wang, Michael Aertsen, Jan Deprest, Sebastien Ourselin, Tom Vercauteren, and Shaoting Zhang. 2020. Uncertainty-Guided Efficient Interactive Refinement of Fetal Brain Segmentation from Stacks of MRI Slices. In *MICCAI*. Springer.
- [51] Guotai Wang, Wenqi Li, Maria A Zuluaga, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Transactions on Medical Imaging* 37, 7 (2018), 1562–1573.
- [52] Guotai Wang, Xinglong Liu, Chaoping Li, Zhiyong Xu, Jiugen Ruan, Haifeng Zhu, Tao Meng, Kang Li, Ning Huang, and Shaoting Zhang. 2020. A Noise-robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images. *IEEE Transactions on Medical Imaging* (2020).
- [53] Guotai Wang, Maria A Zuluaga, Wenqi Li, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. 2018. DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 7 (2018), 1559–1572.
- [54] Guotai Wang, Maria A Zuluaga, Rosalind Pratt, Michael Aertsen, Tom Doel, Maria Klusmann, Anna L David, Jan Deprest, and Tom Vercauteren. 2016. Dynamically balanced online random forests for interactive scribble-based segmentation. In *MICCAI*. Springer, 352–360.
- [55] Guotai Wang, Maria A Zuluaga, Rosalind Pratt, Michael Aertsen, Tom Doel, Maria Klusmann, Anna L David, Jan Deprest, Tom Vercauteren, and Sébastien Ourselin. 2016. Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. *Medical Image Analysis* 34 (2016), 137–147.
- [56] Jiajun Wu, Yibiao Zhao, Jun-Yan Zhu, Siwei Luo, and Zhuowen Tu. 2014. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. In *CVPR*. 256–263.
- [57] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Rong-Guo Zhang, and Ming-Ming Cheng. 2021. JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing* 30 (2021), 3113–3126.
- [58] Ning Xu, Brian Price, Scott Cohen, Jimei Yang, and Thomas S Huang. 2016. Deep interactive object selection. In *CVPR*. 373–381.
- [59] Qingsen Yan, Bo Wang, Dong Gong, Chuan Luo, Wei Zhao, Jianhu Shen, Qinfeng Shi, Shuo Jin, Liang Zhang, and Zheng You. 2020. COVID-19 Chest CT Image Segmentation—A Deep Convolutional Neural Network Solution. *arXiv preprint arXiv:2004.10987* (2020).
- [60] Shiyin Zhang, Jun Hao Liew, Yunchao Wei, Shikui Wei, and Yao Zhao. 2020. Interactive Object Segmentation With Inside-Outside Guidance. In *CVPR*. 12234–12244.
- [61] T Y Zhang and C Y Suen. 1984. A fast parallel algorithm for thinning digital patterns. *Communications of The ACM* 27, 3 (1984), 236–239.
- [62] Feng Zhao and Xianghua Xie. 2013. An overview of interactive medical image segmentation. *Annals of the BMVA* 2013, 7 (2013), 1–22.
- [63] Bowei Zhou, Li Chen, and Zhao Wang. 2019. Interactive deep editing framework for medical image segmentation. In *MICCAI*. Springer, 329–337.

A 仿真机器人用户

为了对模型进行客观的训练和评估, 我们需要一个机器人用户来模拟真实用户的交互。在本节中, 我们将简要介绍不同交互模式下的机器人用户策略。对于初始交互, 我们用一定的偏移量扩展包围物体的交互, 以模拟真实的用户行为。偏移值在训练中是随机的, 在

测试中是固定的。包围框: 在目标区域周围生成带有偏移量的松弛矩形。偏移值在训练中是随机的, 在测试中是固定的。包围多边形: 我们使用凸包算法 [47] 来寻找最小凸包多边形。我们将这些凸包顶点和这些被随机值偏移的点作为一个新的点集。通过重新计算新生成的点集的凸包, 我们可以得到一个不规则包围多边形。偏移值在训练中是随机的, 在测试中是固定的。包围涂鸦: 采用扩张目标的边界轮廓进行模拟, 迭代次数在训练中是随机的, 在测试中是固定的。

对于精细化交互, 训练时采用随机抽样策略和迭代抽样策略, 测试时只采用迭代抽样策略。区域点击: 对于随机抽样策略: 从前景和背景区域或之前的错误预测区域中随机选择前背景点来模拟点击。对于迭代采样策略, 在之前标注的基础上, 新添加最大错误预测区域的中心作为交互点。区域涂鸦: 对于随机抽样策略, 我们模拟涂鸦为一个由 [61] 算法生成的目标骨架的两个端点之间的八连通路径。对于迭代采样策略, 则用最大错误区域的骨架作为涂鸦。边界点击: 对于随机抽样策略, 边界点从目标边界像素中随机选取。对于迭代采样策略, 边界点是距离上次预测图的轮廓边界最远的边界点。包围涂鸦: 对于随机抽样策略, 包围涂鸦是目标轮廓边界的一定比例部分。对于迭代采样策略, 我们将确保下一次涂鸦的中点与边界点击的选取的点相同。

其中 \mathcal{P}_b 表示预测的边界分割; \mathcal{G}_n 示边界分割的真实值; $\phi(p, \mathcal{X})$ 表示点 p 到点集 \mathcal{X} 之间的最短距离; $|\cdot|$ 表示像素数。

- **mNoI**: 我们使用交互的平均数量来衡量交互方法的交互性能。它被定义为修复阶段的平均交互次数, 直到每个实例达到指定的 Dice(@95%). 最大交互次数设置为 20。

B 度量指标

以下是三个指标的介绍:

- **Dice**: 我们利用 Dice 评分来评价医学图像的分割质量: $Dice = 2|\mathcal{P} \cap \mathcal{G}|/(|\mathcal{P}| + |\mathcal{G}|)$. 该指标的值越高, 分割质量越好。
- **ASSD**: 与 [53] 中的方法相同, 我们使用像素的平均对称表面距离来评价医学图像的边界分割质量。

$$ASSD = \frac{1}{|\mathcal{P}_b| + |\mathcal{G}_b|} \left(\sum_{p \in \mathcal{P}_b} \phi(p, \mathcal{G}_b) + \sum_{p \in \mathcal{G}_b} \phi(p, \mathcal{P}_b) \right),$$