arXiv:2611.28740v1 [eess.IV] 21 Nov 2026

US-Seg：通用超声分割的一种参考图像

Chen Ma，*IEEE* 研究生会员、Yunshu Li、Shuyu Liang、Qin Wang、Junhu Fu、Zeju Li、Yi Guo，*IEEE* 会员，以及 Yuanyuan Wang，*IEEE* 高级会员

摘要——**近期的超声基础模型已展现出跨器官、跨任务和跨成像条件的强大特征提取能力。然而，要将这些表征转化为分割结果，仍需在带标注的目标数据上进行微调。为减少对任务特定微调的依赖，现有通用超声分割方法通常采用基于提示的目标指定，主要通过 SAM式模型中的空间提示，或以 MOFO 为代表的任务提示来实现。然而，对于训练期间未见过的目标进行直接分割，仍需要对每幅图像进行重复交互或目标特定适应。我们提出 US-Seg，一种通用超声分割框架，仅使用单个已标注参考图像指定目标，并在无需微调的情况下直接预测查询图像中的目标掩膜。US-Seg 的核心思想是：通过掩膜标签重建，利用异构参考-查询对学习共享的结构恢复过程，将分割从针对预定义目标的直接图像到标签映射重新表述为结构恢复过程，从而使单个参考图像足以完成目标指定。以边界为中心的标签掩蔽强调了模糊边界，促使模型学习可迁移的结构线索，而非目标特定的外观模式。借助这些结构增强的表征，多级双向参考-查询交互将参考定义的目标与查询图像对齐。最后，预测感知的迭代解码器利用中间预测，在参考图像指导下逐步纠正残余结构误差。无需重复交互或繁琐标注，仅使用一个参考图像的单个 US-Seg 模型在涵盖跨站点和跨任务泛化的六项评估中取得了 81.85% 的平均 Dice系数，比最强的标签参考方法高出 24.30%，同时还实现了最佳域内性能。这些结果推动了通用超声分割向着在不断变化的临床目标和多样化成像环境中实现直接、可扩展应用的方向发展。**

索引词—**超声图像分割、通用分割、参考图像引导的重建**

I. 引言

RECENT 以 USFM [1], 为代表的超声基础模型已证明，在异质超声数据上进行大规模预训练，能够学习到适用于不同器官、成像条件和下游任务的可迁移表征。这些进展通过降低对任务特定特征学习的依赖，为通用超声图像分析奠定了良好基础。然而，表征层面的通用性并不能直接转化为分割级通用性，因为现有超声分割系统通常仍需要特定任务的解码器、带标注的目标域数据或额外的微调，才能将学习到的表征应用于每个新的解剖目标或成像域 [2]。因此，通用超声表征学习与直接通用分割之间仍存在一个关键差距，尤其是在训练期间未见过相应成像域或解剖目标的情况下。

为实现分割级通用性，近期的通用分割方法引入了灵活的目标指定机制。空间提示方法通过点和框等位置线索指定目标，以 SAM [3] 及其医学变体 MedSAM [4],为代表；SAMUS [5]和 UltraSAM [6] 则进一步将这一范式应用于超声，在不同目标上取得了出色的分割性能。这类空间提示主要传达目标位置，提供的结构和语义信息有限，而且通常需要针对每张查询图像反复提示。相比之下，MOFO [7]和 PerceptGuide [8], 等任务提示方法在任务层级指定分割目标，并支持多个器官或任务

在统一模型中。其目标空间在训练期间仍然是预先定义的，因此无法直接扩展到此前未见过的解剖结构。一个更灵活的替代方案是通过已标注参考图像来定义目标，正如 UniverSeg [9] 和 INSID3 [10]所探索的那样。与空间提示和任务提示相比，已标注参考图像能提供更丰富的结构与语义线索，支持超越固定类别的目标指定，并可在多个查询中重复使用。如果能够根据单个标注可靠地定义目标，那么可重复使用的参考图像便能在不同查询中提供灵活且一致的指定方式，无需重新定义固定的任务空间，也无需针对每张图像重复指定。

尽管具有这一潜力，但要使单个已标注参考足以实现通用超声分割仍然极具挑战性，并带来了三个层层关联的挑战。首先，如何在异构的超声分布之间**泛化结构**。与更加标准化的医学成像模态相比，超声尤其容易受到不同设备、操作人员和采集角度所引起的外观变化的影响。因此，同一解剖结构在不同成像条件下可能呈现出显著不同的外观，使依赖外观的对应关系变得不可靠，并阻碍稳定结构知识在不同域之间的泛化 [11], [12]。其次，在具备这种可迁移结构知识的情况下，面对参考图像与查询图像之间的多样差异，如何仅凭一个已标注参考**指定目标**。现有的基于参考图像的方法通常通过使用多个

本研究获国家重点研发计划（2024YFF0507300、2024YFF0507303）和国家自然科学基金（62531004）资助。（通讯作者：Yi Guo；Yuanyuan Wang）

Chen Ma、Yunshu Li、Shuyu Liang、Qin Wang、Junhu Fu、Zeju Li、Yi Guo 和 Yuanyuan Wang 均来自中国上海 200433 复旦大学生物医学工程学院（电子邮箱： {cma24、ysli24、syliang22、wangqin23、jhfu21}@m.fudan.edu.cn； {zejuli、guoyi、yywang}@fudan.edu.cn）。

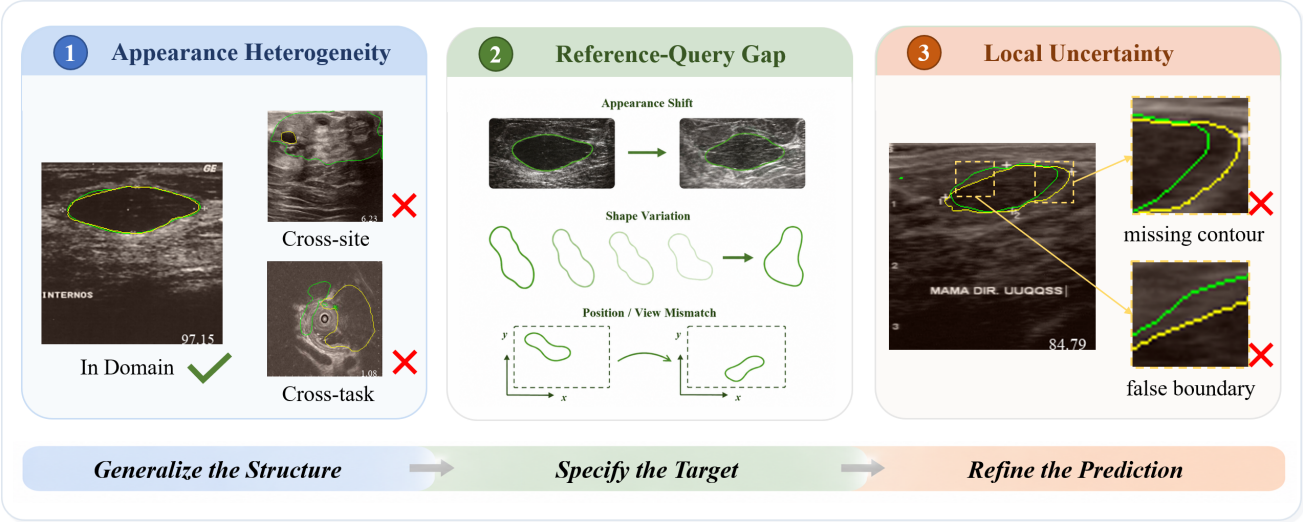


图1. 单参考通用超声分割的关键挑战。

已标注参考或支持示例来缓解较大的外观差异，有时需要数十个参考案例才能提供足够的覆盖范围 [9]–[13]。然而，依赖大量已标注参考会重新带来沉重的标注负担，并削弱基于参考的分割在实际应用中的动机。通用超声模型应当转而尽量降低标注负担，使单个可复用参考图像足以在不同查询中可靠地指定目标。第三，一旦目标得到指定，如何在局部结构存在歧义的情况下**细化预测**。现有方法可能能够近似定位目标结构，但弱边界、缺失轮廓、伪边缘、声影和低对比度区域往往会妨碍对其精细形状的准确分割 [14], [15]。值得注意的是，已标注参考不仅为目标识别提供初始线索，还为校正不确定区域提供可复用的结构引导。要充分利用这一信息，需要采用渐进式的参考引导细化，而不是一次性预测。

为应对这些挑战，我们提出了 US-Seg，这是一种通用超声分割框架，将传统的图像到标签预测重新定义为异构参考-查询对上的参考引导重建，从而仅凭一个可重复使用的参考图像即可实现可靠分割。在训练过程中，US-Seg 学习从异构参考-查询对中恢复被掩码的标签结构，促使模型在不同解剖目标和成像分布之间进行共享结构重建。在推理阶段，仅需提供一个已标注参考图像-掩膜对和一幅查询图像，US-Seg 即可直接分割数据集中的所有查询图像，无需进行任何目标特定适应。已标注参考通过提供目标身份及其结构和外观上下文，明确了分割目标。这种方法使单个模型能够在不同成像中心之间迁移分割能力，并仅凭一个已标注参考图像为整个数据集创建新的分割任务。本文的主要贡献总结如下：

- 我们提出了 US-Seg，这是一种参考引导重建用于通用超声分割的框架，仅凭一个可重复使用的已标注参考图像，即可让单个模型灵活地分割已观测和此前未见的解剖目标。
- 我们以异构参考-查询对作为学习单元，设计了一种统一的重建方案：通过边界中心标签掩蔽学习与目标无关的结构恢复，从而实现结构泛化；通过与已标注参考图像建立多级双向参考-查询交互来明确目标；并通过预测反馈和参考条件引导逐步纠正残余结构误差，以细化边界，从而仅凭单个可重复使用的参考图像即可实现可靠分割。
- 我们在多种超声基准上进行了广泛实验，结果表明，US-Seg 无需微调即可实现出色的跨站点和跨任务泛化，同时在域内分布上保持最先进的性能。

II. 相关工作

A. 通用表征学习

通用表征学习已成为医学图像分析中的重要范式，旨在从大规模异质数据中学习可迁移特征，以支持广泛的临床任务 [16], [17]。通用医学人工智能和医学基础模型进一步证明了可复用表征在不同器官、成像模态和任务中的价值 [18]–[20]。这一范式与超声尤其相关，因为图像外观会因解剖结构、探头位置、操作人员、成像协议和设备厂商的不同而显著变化。USFM [1]通过多器官、多中心和多设备预训练，构建了通用超声表征骨干网络，并展现出在标签高效分类、分割和增强任务中的强迁移能力。

URFM [21] 通过针对超声图像低信噪比特征设计的基于表征的掩码图像建模，进一步推进了超声表征学习。TinyUSFM [22] 则将这一方向拓展至效率层面，表明经过精心设计的蒸馏能够使紧凑型超声基础模型在不同器官和任务中保持通用性。综上，这些研究确立了通用表征学习作为实现可扩展超声图像分析的重要基础。

尽管取得了这些进展，现有的超声基础模型主要致力于学习可迁移表征，通常作为可复用编码器。对于下游分割，它们仍需要特定任务解码器、监督式适配或目标域微调。更重要的是，仅凭可迁移特征并不能确保在异质超声分布中始终如一地恢复解剖结构，因此分割层面的结构泛化问题仍未得到充分解决。

B. 可提示分割

可提示分割通过引入明确的机制来指定需要分割的对象，推动了通用分割的发展。空间提示方法通过查询图像中的位置线索来定义目标。SAM [3]采用点、框和掩码确立了这一范式，而 Med-SAM [4] 则将其扩展至医学图像分割。在此可提示范式的基础上，SAMUS [5]和 Ultra-SAM [6]进一步将 SAM 风格的分割适配于超声，并在各种超声目标上取得了出色表现。尽管空间提示能够实现超越固定语义类别的灵活分割，但每张查询图像都需要通过点、框或其他空间线索单独指定目标。因此，随着查询数量增加，需要在多张图像上反复提供提示，导致用户交互不断重复。

任务提示方法从另一种角度处理目标指定，即明确编码应执行分割的器官或分割任务。MOFO [7] 和 PerceptGuide [8] 等方法将任务级引导、解剖先验或感知信息纳入模型，以在统一模型中支持多个器官或任务。与空间提示相比，任务提示能更明确地表示目标身份，但仅适用于训练期间定义的分割目标，因此无法灵活指定此前未见过的解剖目标。

因此，现有的可提示分割在基于位置的灵活指定与明确但预先定义的目标身份之间面临权衡，这推动了目标指定机制的发展，使其既能提供灵活性，又能确保跨图像的目标定义一致。

C. 标签参照分割

标签参考分割通过带标注的示例来定义预期结构，而非依赖空间位置或预定义任务身份，从而灵活地指定目标。SegGPT [23]通过上下文图像-掩膜示例来构建分割任务，而 PerSAM [24] 则进一步展示了针对特定视觉目标的单参考个性化。然而，可靠的目标迁移仍依赖有效的参考-查询对应关系；当外观差异显著时，这一对应关系便难以建立。

在医学成像中，UniverSeg [9]和 Tyche [13] 通过多个带标注的支持样例改进目标指定，但代价是增加标注负担。INSID3 [10] 等较新的方法则转向单参考分割，通过跨图像对应关系迁移指定的目标。然而，超声中常见的较大外观差异和结构歧义会削弱跨图像对应关系，使可靠的目标迁移和精确的边界恢复变得困难。

总而言之，现有标签参考方法证明了灵活指定目标的可行性，但在超声中遇到的显著分布偏移和结构歧义，使得可靠的单参考分割仍然面临挑战。因此，通用超声分割需要一种参考引导机制，能够在不同分布间泛化结构、跨越参考与查询之间的差异保持目标身份，并持续利用参考信息来优化不确定的预测。

III. 方法

A. 方法概述

我们提出 US-Seg，这是一种通用超声分割框架，将异构分割任务建模为共享的参考引导重建问题，而非图像标签训练。给定参考图像 I_{ref} 及其对应标注 Y_{ref}（二者共同定义分割目标），以及查询图像 I_q，US-Seg 会预测 $\widehat{}$ 对应的查询掩膜 Y_{q}。在每个参考查询对中，参考图像和查询图像对应同一分割目标，但取自不同患者。参考标注指出应分割哪种解剖结构，而参考图像则提供该结构的视觉、结构及周围解剖背景。训练时，参考标签和查询标签被表示为补丁级标签标记，并部分或完全掩蔽以进行重建，从而使模型能够在异构参考图像-查询图像对之间学习共享的结构恢复。推理时，模型直接根据带标注的参考对 (I_{ref}, Y_{ref}) 和查询图像 I_q 预测查询标签，无需查询侧标注。

US-Seg由三个组件组成。边界中心标签掩蔽优先掩蔽结构模糊区域，以学习适用于异质解剖结构的、与目标无关的结构恢复。在这一可迁移的结构表征基础上，多级双向参考-查询交互通过双向跨分支交互和自适应融合，在参考图像与查询图像之间建立稳健的对应关系，从而能够在查询图像中识别由参考图像定义的目标。基于这一表征，预测感知迭代细化解码器将中间预测反馈至解码过程，并通过门控的、以参考图像为条件的潜在状态更新，逐步纠正残余结构误差。总而言之，

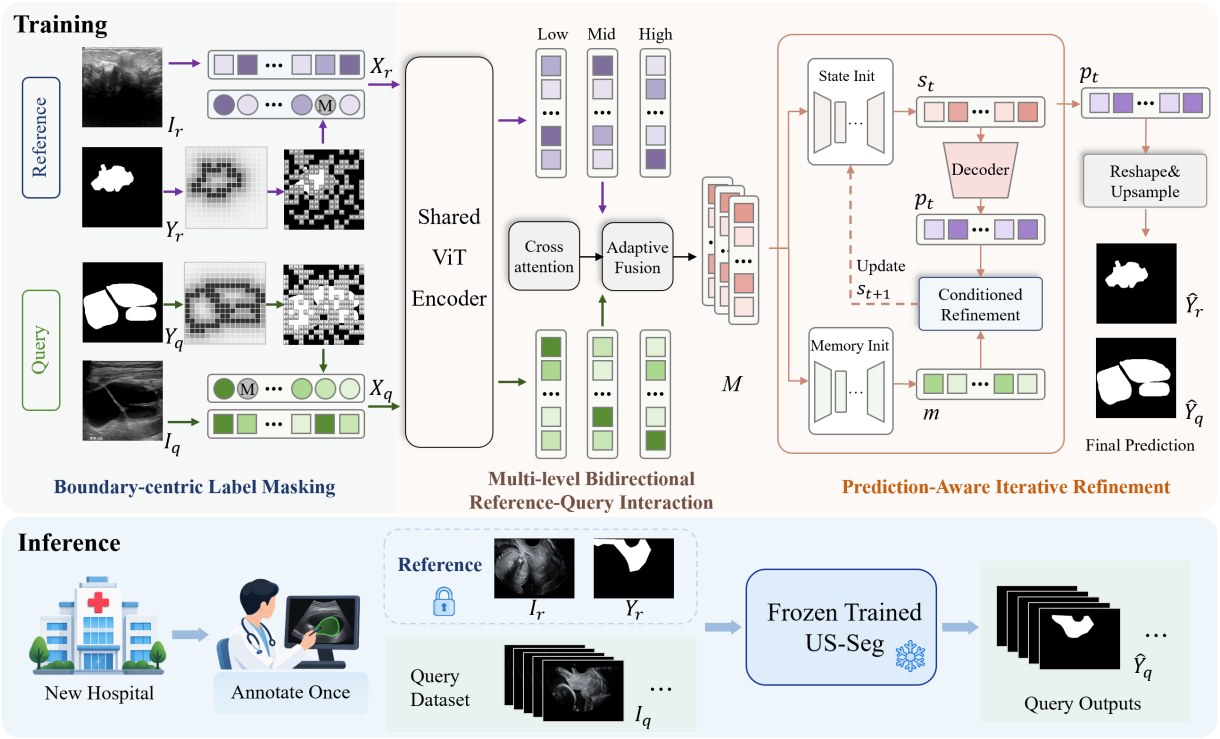


图2. US-Seg框架概览。

这些组件能够直接分割开放式超声目标，无需进行目标特定参数适配。

B. 边界中心标签掩蔽

我们引入边界中心标签掩蔽，以促进模型在异质超声分布之间的结构泛化。在此，我们为靠近目标边界的标签块赋予更高的掩码概率，鼓励模型依据稳定的上下文线索，而非依赖特定于目标的视觉模式，来恢复不确定的形状信息。

为两个分支构造标签令牌，令 Y 表示参考图像或查询图像的二值标注掩码。我们首先依据 ViT块网格 [25] 对 Y 进行划分。对于第 i 个块区域 P_i，相应的标签块定义如下：

$$L_i = Vec\left(Y|P_i\right), i = 1, \ldots, N. \tag{1}$$

每个可见标签块都通过可学习映射 $\phi(\cdot)$ 转换为标签令牌，而被掩蔽的块则由可学习的掩码嵌入 e_m表示：

$$z_i^{(0)} = \begin{cases} e_m, & m_i = 1, \\ \phi\left(L_i\right), & m_i = 0. \end{cases} \tag{2}$$

对于每个分支 $b \in \{ref, q\}$，图像令牌与对应的掩码标签令牌沿令牌维度拼接，形成联合输入序列 X_b。

为构建边界中心掩码分布，我们首先使用大小为 3×3 的结构元素，通过形态学梯度从 Y 中提取像素级边界图：

$$B = I\left(Dilate_{3\times3}(Y) - Erode_{3\times3}(Y)\right), \tag{3}$$

其中，$Dilate_{3\times3}(\cdot)$ 和 $Erode_{3\times3}(\cdot)$ 分别表示二值膨胀和腐蚀。令

$$\mathcal{B} = \{(u, v) \mid B\left(u, v\right) = 1\} \tag{4}$$

表示边界像素集合。随后计算每个像素（u，v）到最近边界像素的距离：

$$D\left(u, v\right) = \min_{(u', v') \in \mathcal{B}} \sqrt{\left(u - u'\right)^2 + \left(v - v'\right)^2}, \tag{5}$$

对于第 i 个图像块区域 P_i，通过对像素级距离取平均，得到块级边界距离：

$$d_i = \frac{1}{|P_i|} \sum_{(u,v) \in P_i} D\left(u, v\right). \tag{6}$$

基于 d_i，我们定义一种随距离衰减的掩码概率：

$$p_i = p_{min} + \left(p_{max} - p_{min}\right) exp\left(-\frac{d_i}{\tau}\right), \tag{7}$$

其中，p_{max} 和 p_{min} 分别是掩码概率的最大值和最小值，τ 控制衰减率。随后，根据由 p_i 参数化的伯努利分布，独立采样掩码指示变量 m_i：

$$m_i \sim Bernoulli\left(p_i\right). \tag{8}$$

其中，$m_i = 1$ 表示第 i 个标签块被掩码以进行重建，而 $m_i = 0$ 表示该标签块保持可见。

对于部分掩码重建，同样独立地将以边界为中心的采样机制应用于

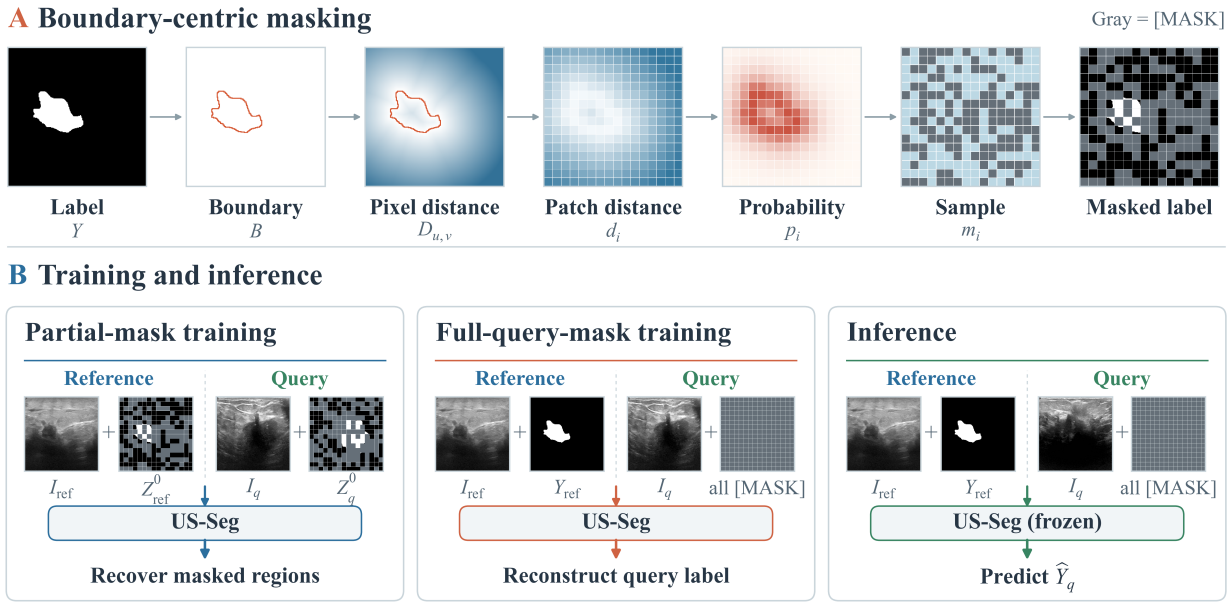


图3. 聚焦边界的标签掩蔽示意图。

参考图像和查询标签。给定一对训练对 $(I_{ref},\ Y_{ref})$ 和 $(I_q,\ Y_q)$，我们构建

$$Z^{(0)}_{ref} = E\left(Y_{ref}, m_{ref}\right), Z^{(0)}_{q} = E\left(Y_q, m_q\right). \tag{9}$$

其中，$E(Y,m)$ 表示标签掩膜嵌入算子，该算子根据 m 将上述逐块令牌规则应用于 Y 的每个图块。所得的 $Z^{(0)}_{ref}$ 和 $Z^{(0)}_{q}$ 分别是参考分支和查询分支的初始标签令牌序列。通过在异构训练目标中反复强调结构不确定区域，该掩蔽策略使共享重建过程倾向于与目标无关的结构恢复，而非特定外观模式。

C. 多级双向参考-查询交互

在聚焦边界的结构学习基础上，下一步是在查询图像中识别由参考图像定义的目标，尽管二者存在外观差异。为此，US-Seg 在参考分支与查询分支之间建立多级双向交互，使来自已标注参考图像的目标特定信息能够在多个语义深度上与查询图像对齐。给定联合输入序列，共享的 ViT-L 编码器从一组中间层 L 中提取参考特征和查询特征，分别记为 $\{x^l_{ref}\}_{l\in\mathcal{L}}$ 和 $\{x^l_q\}_{l\in\mathcal{L}}$。这些特征在不同语义深度上提供互补信息：较浅层保留局部边界和纹理线索，而较深层则编码更高层次的结构语义。

对于每个层级 l，都会在两个分支表征之间应用双向交叉注意力。查询分支关注参考图像，以获取目标特定结构线索；而参考分支则关注查询图像表征，以结合当前查询图像的外观对这些线索进行上下文建模：

$$\widetilde{x^l_q} = x^l_q + \mathrm{CA}\left(x^l_q, x^l_{ref}\right), \widetilde{x^l_{ref}} = x^l_{ref} + \mathrm{CA}\left(x^l_{ref}, x^l_q\right). \tag{10}$$

此处，$CA(\cdot)$表示多头交叉注意力。这种双向交互在特征融合之前建立了目标感知的参考-查询对应关系，而不是依赖单向的参考图像注入。

随后，利用可学习的空间门控对交互后的特征进行自适应融合。具体而言，将查询特征、参考特征及其差异进行拼接，以预测融合门控：

$$g^l = \sigma\left(MLP\left(\left[\widetilde{x^l_q}; \widetilde{x^l_{ref}}; \widetilde{x^l_q} - \widetilde{x^l_{ref}}\right]\right)\right). \tag{11}$$

第 l 层的融合特征通过以下方式获得：

$$f^l = g^l \odot \widetilde{x^l_q} + \left(1 - g^l\right) \odot \widetilde{x^l_{ref}}. \tag{12}$$

该门控在每个空间位置动态平衡查询特有的视觉证据与参考图像提供的结构引导。最后，将所有选定层级的融合表征沿通道维度进行拼接：

$$M = \mathrm{Concat}\left(f^{l_1}, f^{l_2}, \ldots, f^{l_K}\right). \tag{13}$$

所得表征融合了跨多个语义深度的参考-查询对应关系，并作为后续迭代细化的条件化特征场。

D. 预测感知迭代细化

给定参考条件多层表示，US-Seg 采用预测感知迭代细化解码器进行渐进式标签重建。解码器并非通过单次解码生成查询掩码，而是维护一个不断演化的潜在状态，并反复将其中间预测反馈到细化过程中。

具体而言，解码器从条件化表示 M 初始化潜在状态 s_0 和持久记忆 m：

$$s_0 = W_s M, \quad m = W_m M. \tag{14}$$

潜在状态表示不断演化的重建假设，而固定记忆在整个细化过程中保留经过参考图像条件化的目标证据，在预测逐步更新时提供稳定的结构引导。

在细化步骤 t中，当前状态 s_t 被解码为$\widehat{\ }$中间掩膜预测 Y_t。随后，该预测会被重新标记化并投影到潜在空间中，形成预测感知特征 p_t，使当前分割假设能够直接影响下一细化步骤。基于 s_t 和 p_t，门控模块估计空间细化门控 g_t，用于确定当前状态应在何处以及以何种程度进行更新。

状态校正由共享细化模块 R（·）生成，该模块结合了对当前状态的自注意力与对固定参考图像条件化记忆的交叉注意力，同时融入预测反馈特征p_t。状态转移可写为：

$$s_{t+1} = s_t + g_t \odot R\left(s_t, m, p_t\right). \tag{15}$$

其中，g_t 决定需要在何处以及以多大强度进行细化，而 R（·）则利用预测反馈和参考图像条件化记忆校正这些区域。重要的是，细化仅更新潜在状态，而不是模型参数，从而保留免微调推理。经过 T 次细化步骤后，最终查询掩码从最后一个$\widehat{\ }$潜在状态 $Y_q = \mathrm{Dec}\left(s_T\right)$。

通过这一闭环过程，US-Seg 将预测反馈转化为参考引导的状态校正，逐步消解局部结构歧义和残余分割误差，而无需进行目标特定参数适配。

E. 训练目标与推理

US-Seg 采用两个互补目标进行优化：一是迭代重建损失，在整个细化过程中监督标签恢复；二是门控正则化项，鼓励局部状态更新。总体训练目标为：

$$\mathcal{L}_{total} = \mathcal{L}_{iter} + \lambda_g \mathcal{L}_{gate}, \tag{16}$$

其中，$\mathcal{L}_{iter}$ 用于监督各细化步骤的重建准确性，$\mathcal{L}_{gate}$ 用于正则化空间细化门控。

为同时支持结构感知学习和完全掩蔽的推理设置，US-Seg采用混合式掩蔽标签重建策略。在部分掩码训练中，参考图像标签和查询标签会分别使用III-B节所述的边界中心掩码机制进行扰动，模型则重建两个分支中缺失的标签区域。在全查询掩码训练中，所有查询标签词元都会被掩蔽，因此必须完全依据已标注参考和查询图像重建查询掩膜。这样一来，模型便能直接接触到推理过程中遇到的目标缺失查询条件。

对于每个分支 $b \in \{ref, q\}$，细化步骤 t 的重建损失仅在掩码区域上计算：

$$\mathcal{L}_{rec,b}^{(t)} = \frac{\sum \ell\left(\widehat{Y_b^t}, Y_b\right) \odot M_b}{\sum M_b + \epsilon}, \tag{17}$$

其中，M_b 表示由相应图块级掩码指示符扩展得到的像素级掩码，ℓ(·)表示逐像素 Smooth L1 损失，其中 $\beta = 0.01$。两个分支的重建损失在各细化步骤中汇总如下：

$$\mathcal{L}_{iter} = \sum_{t=1}^{T} \lambda_t \left(\mathcal{L}_{rec,ref}^{(t)} + \mathcal{L}_{rec,q}^{(t)}\right), \tag{18}$$

其中，λ_t 控制每个细化步骤的贡献。为避免对潜在状态进行不必要的全局修改，我们还对空间细化门控进行正则化。设 g_t表示第 t 步的细化门控。门控正则化定义为

$$\mathcal{L}_{gate} = \sum_{t=1}^{T} \mathrm{mean}\left(g_t\right). \tag{19}$$

最小化该项可促使更新保持稀疏且局部化，使细化集中于需要结构校正的区域，而不是对整个预测结果进行统一修改。

在推理过程中，仅提供带注释的参考对（I_{ref}，Y_{ref}）和查询图像 I_q，而查询端的所有标签令牌均初始化为掩码令牌。带注释的参考图像指定了目标，解码器则通过预测反馈和参考条件引导，逐步重建并细化查询掩膜。

IV. 实验

A. 数据集

我们在涵盖多个解剖目标、成像视图、设备和临床来源的多样化超声数据集上评估 US-Seg，具体汇总于表 I。所有数据划分均在患者级别进行，以避免信息泄漏。因此，我们将评估组织为三个泛化难度逐步增加的设置：

（1）域内设置：分割任务和成像分布均在训练过程中有所体现。该设置在熟悉的成像条件下，评估 US-Seg 对训练所定义目标的处理能力。

(2) 跨站点设置：训练过程中包含该分割任务，但完全排除测试数据集。此设置用于评估在不进行目标数据集适配的情况下，已学习的分割能力能否适应临床来源、成像设备、采集协议和图像特征的变化并实现迁移。

(3) 跨任务设置：训练过程中既不包含该分割任务，也不包含其数据集。模型必须根据已标注参考图像推断目标定义，并将其学到的结构重建能力迁移到训练任务空间之外的新分割任务。

对于域内数据集，病例按患者级别划分为训练集、验证集和测试集，比例为

表I 实验中使用的超声数据集汇总。

设置	器官	目标	数据集	图像数量
域内	乳房	乳腺肿瘤	BrEaST [26], BUS-BRA [27], BUSI [28], BUSI-WHU [29]	3,838
	甲状腺	甲状腺结节和甲状腺	DDTI [30], TG3K [31], TN3K [31]	7,715
	心脏	左心室腔、心肌和左心房	CAMUS [32]	2,000
	腰椎	多裂肌	LUMINOUS [33]	341
	Neck	臂丛	颈部 [34]	5,635
	骨盆	耻骨联合和胎头	PSFHS [35]	1,358
	胃	胃肠道间质瘤	美元514 [36]	514
域外 跨站点	乳房	乳腺肿瘤	BUS-UCLM [37],UDIAT [38]	427
	甲状腺	甲状腺结节	TNSCUI [39]	3,644
域外 跨任务	肾脏	肾脏	KidneyUS [40]	487
	子宫颈	宫颈前唇和后唇	FUGC [41]	440
	前列腺	前列腺	MicroSeg [42]	2,621
	卵巢	卵巢肿瘤	MMOTU [43]	1,469

7:1:2，US-Seg 在所有训练集划分上进行联合训练。域内参考图像取自训练集划分，而域外参考图像则从相应数据集中选取，并从查询集中排除；其余所有域外（OOD）病例均用作查询。所有实验的查询集均固定不变。

训练期间，参考图像-查询图像对由来自不同患者、具有相同分割目标的两张图像组成。在标准评估期间，US-Seg、PerSAM 和 INSID3 使用一个固定参考图像，而 UniverSeg 和 Tyche 使用 16 个固定支持样本。对于 IV-F 中的鲁棒性分析，查询集保持固定，并将 100 个已标注病例中的每个病例分别作为参考图像进行评估。

这些设置共同构成了渐进式评估流程，从域内分割到跨站点迁移，最终到跨任务泛化。

B. 实现细节

US-Seg 使用 PyTorch 实现，并在单块 NVIDIA A100 GPU 上训练。所有图像和掩膜均调整为448×448。我们使用 Adam优化器，初始学习率为 1×10^{-3}，批量大小为 32，并训练模型 400 轮。训练期间，部分掩膜重建和完整查询掩膜重建的采样概率分别为 0.8 和 0.2。对于边界中心掩膜，我们设置$p_{\min} = 0.4$、 $p_{\max} = 0.8$和 $\tau = 24$ 像素。迭代解码器执行 T=3 个细化步骤，其 (λ_1, λ_2, λ_3) =(0.3，0.7，1.0)，门控正则化权重 λ_g 设为 0.01。对于每个分割目标，使用来自不同患者的一组已标注参考图像-掩膜对来指定目标，并且可在多个查询图像中重复使用。

C. 对比方法与评估指标

我们将 US-Seg 与来自四种分割范式的代表性方法进行比较，包括基于微调的方法（nnUNet [44] 和 USFM [1]）、空间提示方法（MedSAM [4], 、UltraSAM [6], 和 SAMUS [5]）、任务提示型超声分割方法（MOFO [7]和 PerceptGuide [8]），以及标签参考方法（UniverSeg [9], 、Tyche [13], 、PerSAM [24], 和 INSID3 [10]）。对于空间提示方法，我们为每张查询图像提供一个点提示。遵循各自的学习范式，nnUNet 和 USFM 分别在每个域内数据集上进行微调。对于 UltraSAM、SAMUS、MOFO 和 PerceptGuide，由于其原始训练数据与我们的部分评估数据集重叠，为避免信息泄漏，我们仅使用域内训练数据重新训练这些模型。对于 MedSAM、UniverSeg、Tyche、PerSAM 和 INSID3，我们直接使用其公开发布的预训练权重，不再进行额外的重新训练或适配，因为这些模型已在大规模医学图像数据集上完成训练，且与我们的评估集不存在重叠。

对于标签参考方法，我们遵循各自的参考范式。US-Seg、PerSAM 和 INSID3 使用单个已标注参考图像，而 UniverSeg 和 Tyche 则使用 16 个带标注的支持示例进行评估。使用 Dice 系数和交并比（IoU）评估分割性能。

D. 跨超声场景通用分割

1）统一域内学习的有效性：如表 II 所示，US-Seg 在全部七个域内器官组中均取得最佳性能，平均 Dice 系数和交并比分别为 88.23% 和 80.89%。与通用表征学习相比，USFM 提供了强大的可迁移特征，但仍需针对特定数据集进行分割适配。US-Seg 使用在异构数据集上联合训练的单一模型，Dice 系数和交并比分别比 USFM 高出 1.25% 和 1.68%，这表明统一分割在域内分布下不会牺牲性能。包括 MedSAM、UltraSAM 和 SAMUS 在内的空间提示方法进一步提升了目标灵活性，但需要对每张查询图像进行重复交互。相比之下，US-Seg 使用一个已标注参考图像，可将其重复用于多个查询图像，同时取得更强的分割性能。

MOFO 和 PerceptGuide 等任务提示方法通过在任务层级指定目标，避免了重复的空间提示，但其预测仍局限于训练期间定义的目标。US-Seg 的 Dice系数比

表 II 域内数据集上不同模型的比较。* 表示使用 16 个已标注参考进行评估。

方法	乳房		甲状腺		心脏		腰椎		Neck		骨盆		胃	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
基于微调的方法														
nnUNet	86.28	80.36	79.37	71.82	90.25	82.17	91.14	84.69	80.10	69.80	96.44	93.02	79.77	69.91
USFM	87.16	80.62	80.69	72.14	90.46	82.71	92.04	84.84	79.25	68.32	97.36	94.73	81.87	71.09
空间提示方法														
MedSAM	63.58	52.22	65.43	52.77	49.05	41.24	56.01	41.10	76.22	63.77	73.16	63.71	46.29	33.59
UltraSAM	75.77	67.82	76.63	68.29	90.06	81.58	86.40	81.93	69.84	57.36	91.61	84.86	74.39	64.02
SAMUS	87.24	80.07	81.13	74.71	90.60	83.26	90.45	83.99	80.95	69.32	94.98	91.35	81.69	72.48
任务提示方法														
MOFO	83.84	76.01	76.45	67.91	89.76	80.05	88.95	82.34	73.44	61.02	94.02	88.91	80.58	70.21
PerceptGuide	83.23	75.48	79.23	71.54	88.12	79.67	87.51	81.10	78.71	67.18	93.27	87.50	76.24	65.73
标签参考方法														
PerSAM	22.52	17.14	23.11	13.66	29.45	15.40	27.18	16.05	5.73	2.96	24.78	14.26	8.19	4.47
INSID3	28.52	21.41	24.77	16.10	25.89	14.75	24.28	13.94	7.76	4.09	23.31	14.52	13.66	7.85
UniverSeg*	56.65	44.37	36.60	25.59	52.17	37.93	82.87	72.51	30.25	20.39	48.90	35.26	27.59	19.50
Tyche*	57.91	45.15	36.90	25.92	52.95	38.85	77.14	64.21	25.12	16.82	42.87	30.47	18.41	13.08
US-Seg	**88.48**	**81.44**	**83.23**	**76.03**	**91.04**	**83.90**	**92.89**	**85.48**	**81.10**	**70.12**	**97.54**	**94.75**	**83.32**	**74.49**

表 III 不同模型在域外数据集上的比较。* 表示使用 16 个已标注参考进行评估。

Type	器官	基于微调				空间提示				任务提示				标签参考					
		nnUNet		USFM		MedSAM		SAMUS		MOFO		PerceptGuide		UniverSeg*		Tyche*		US-Seg	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
跨站点	乳房	61.68	53.17	70.30	62.06	62.11	51.20	83.93	75.20	78.59	71.72	76.47	67.89	69.84	58.56	57.45	46.52	**86.24**	**78.98**
	甲状腺	61.49	51.81	68.01	58.27	73.96	62.77	75.96	68.67	66.14	55.81	72.54	63.20	37.39	26.43	38.02	26.53	**80.81**	**72.40**
跨任务	肾脏	–	–	–	–	64.43	50.03	78.89	67.24	–	–	–	–	69.96	58.12	65.11	51.03	**82.95**	**72.81**
	子宫颈	–	–	–	–	46.39	36.83	75.58	67.82	–	–	–	–	53.29	40.54	51.25	38.37	**86.04**	**82.48**
	前列腺	–	–	–	–	61.83	47.22	73.51	65.05	–	–	–	–	60.33	47.42	57.71	44.15	**78.40**	**69.22**
	卵巢	–	–	–	–	66.94	54.51	70.29	63.82	–	–	–	–	54.48	41.91	57.29	44.78	**76.68**	**69.06**

MOFO 高 4.37%，交并比比 PerceptGuide 高 5.43%，这表明基于参考的指定无需依赖固定的目标空间即可实现准确的分割。标签参考方法提供了最接近的比较，因为它们同样通过带标注的示例定义目标。尽管使用了在大规模医学图像数据集上预训练的公开发布的模型，并为 UniverSeg 和 Tyche 提供了 16 个带标注的支持示例，但它们在异质超声数据上的表现仍显著较低。US-Seg 的表现优于最强的标签参考基线 UniverSeg：Dice系数高 40.37%，交并比高 44.38%。这些结果表明，当结构泛化和参考-查询对应关系仍然具有挑战性时，强大的预训练和额外的参考图像本身并不足够。

总体而言，US-Seg 将强大的域内性能与可靠的标签参考分割和可复用目标指定相结合，同时避免了重复的空间交互和预定义任务空间。

2）跨站点和跨任务的域外泛化：如表 III 所示，US-Seg 在全部六种域外评估场景中均取得最佳表现，平均 Dice系数和交并比分别为 81.85% 和 74.16%。跨任务设置还揭示了基于微调的方法和任务提示方法的一个根本局限：前者需要特定目标训练数据，而后者仍局限于训练期间定义的目标。相比之下，US-Seg 仅凭一张已标注参考图像即可指定新目标，并直接执行分割，无需进行参数适配，因此同一模型能够适用于跨站点和跨任务设置。

在跨站点评估中，分割任务保持不变但成像分布发生变化，US-Seg 达到 83.53% 的 Dice系数和 75.69% 的交并比。与最强的空间提示基线 SAMUS 相比，其 Dice系数和交并比分别进一步提升了 3.58% 和 3.76%；同时，它用一个可在不同查询之间复用的已标注参考，替代了重复的查询特定提示。US-Seg 的性能也显著优于现有的标签参考方法：尽管 UniverSeg 和 Tyche 获得了 16 个已标注支持样本，US-Seg 的 Dice系数仍比 UniverSeg 高出 29.92%，交并比高出 33.19%。这表明，仅仅增加参考图像的数量，并不足以克服由异质超声图像外观造成的参考-查询差异。

在跨任务评估中，这一优势更加显著，因为分割目标及其数据集均不属于训练任务空间。US-Seg 的 Dice系数为 81.02%，交并比为 73.39%，同时明显优于

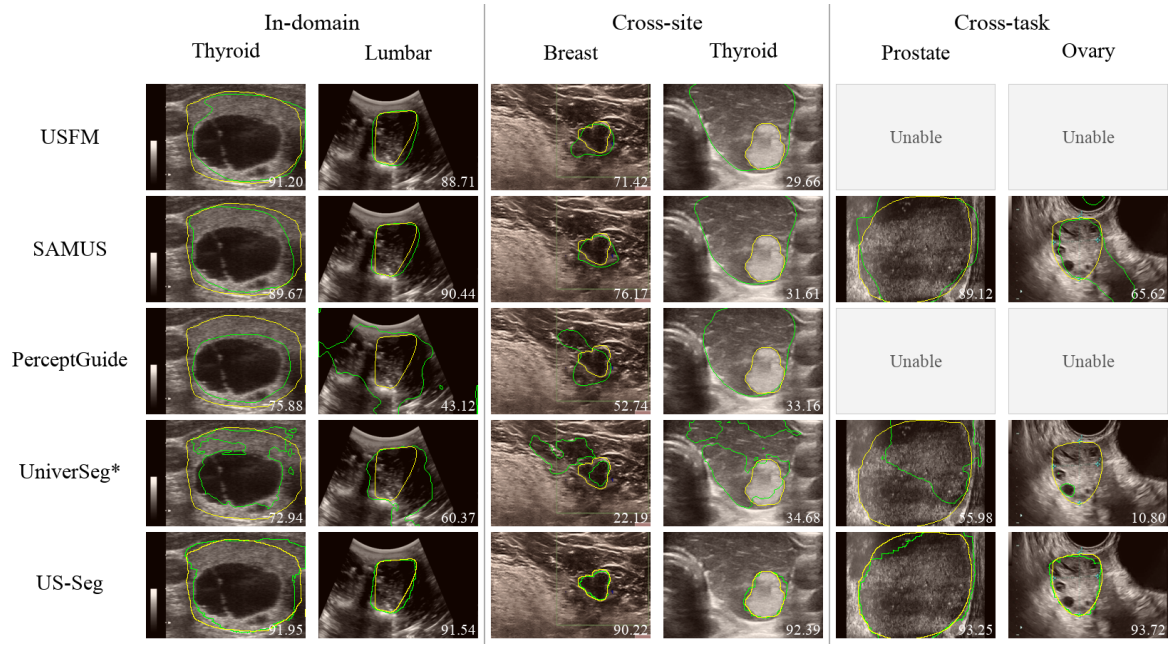


图4. 不同超声分割场景下的定性比较。黄色表示真实标注，绿色表示预测结果。*表示使用16幅已标注参考图像进行评估。

空间提示方法和标签参考方法。值得注意的是，尽管 US-Seg 仅使用一张参考图像，而对比方法使用了 16 个支持样本，但其 Dice系数和交并比分别比最强的标签参考基线 UniverSeg 高出 21.50% 和 26.39%。这些结果表明，US-Seg 只需一张可重复使用的参考图像，就能将参考条件化结构知识迁移到不同的成像分布和分割任务中。

3) 不同超声场景下的定性优势：图4展示了域内、跨站点和跨任务设置下的代表性结果，并选取各分割范式中的强方法进行比较。在域内条件下，US-Seg能够针对已观察到的目标生成准确且结构完整的预测。面对跨站点偏移时，其他方法会出现边界偏差、区域不完整或假阳性，而US-Seg的预测仍与目标结构高度吻合。在跨任务场景中，这一优势更加明显：基于微调和任务提示方法无法直接应用于未见目标，而US-Seg利用已标注参考图像指定目标结构，生成的预测比现有空间提示和标签参考方法更加完整。这些结果直观地证明了US-Seg在逐渐复杂的超声场景下具有鲁棒性。

E. 目标指定效率

除了分割准确率之外，我们还评估了六项分布外任务中目标指定所需的人工投入。为在统一尺度上比较不同交互范式，我们估算了分割整个数据集所需的总标注成本。完整标注一个参考掩膜平均需要 34.2 秒。

如图 5 所示，US-Seg 仅使用一个已标注参考图像，即可获得最高的平均分布外 Dice（81.85%）。在

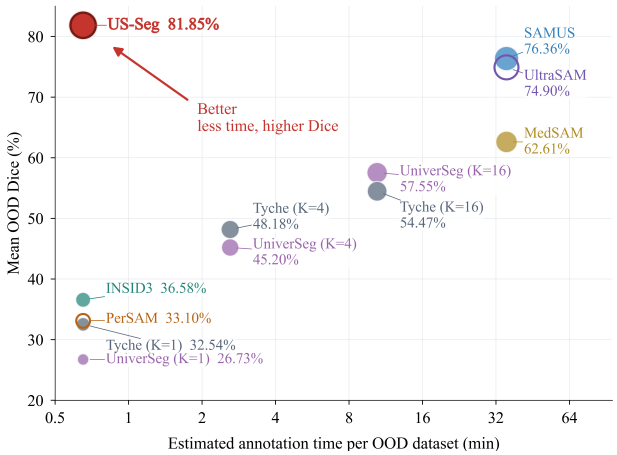
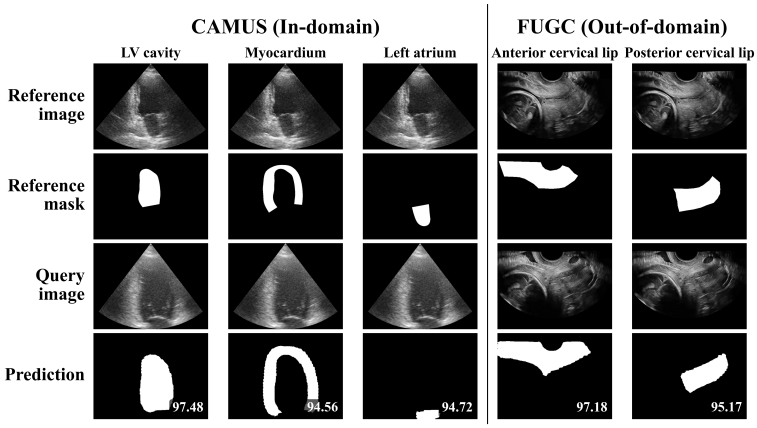


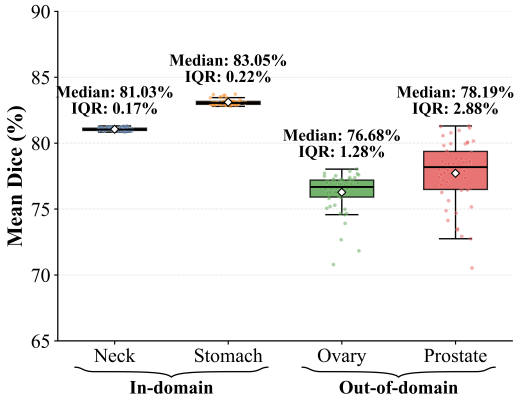
图 5. 分割准确率与目标指定成本的关系。K 表示使用 K 个已标注参考图像进行评估。

此对比之下，空间提示方法需要针对每张查询图像分别进行目标指定，导致其总交互成本随数据集规模增加。平均而言，为每个查询图像放置一个点提示需要 1.7 秒，因此估计每个分布外数据集的标注成本约为 35.2 分钟，约为 US-Seg 的 62 倍。尽管交互成本显著更高，SAMUS、UltraSAM 和 MedSAM 的 Dice分数却分别较低，仅为 76.36%、74.90% 和 62.61%。

增加参考预算可以提升 UniverSeg 和 Tyche 的表现，但使用 16 张参考图像需要约 9.12 分钟，标注成本是 US-Seg 的 16 倍，而它们的 Dice分数仍分别低 24.30% 和 27.38%。PerSAM 和 INSID3 使用与 US-Seg 相同的单张参考图像预算，但准确率显著较低，这表明



（a）参考图像控制的目标切换。



（b）跨参考实例的稳定性。

图6。US-Seg 的参考图像可控性与鲁棒性。

如果缺乏有效的结构泛化和参考图像-查询图像迁移，仅降低标注成本是不够的。

总体而言，US-Seg 通过将目标指定从重复标注或多示例标注转变为单个可重复使用的参考图像，实现了更有利的准确率–工作量权衡。这一特性确保其标注成本不会随后续查询数量的增加而上升，从而支持在临床队列和外部数据集中的可扩展部署。

F. 参考图像控制的目标指定与鲁棒性

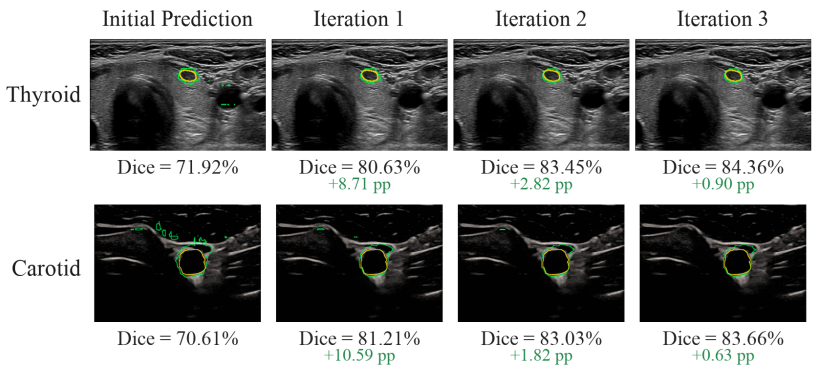
为进一步考察单个已标注参考图像能否在通用超声分割中有效指定预期目标，我们首先通过固定查询图像、仅更改参考图像-掩膜对，评估 US-Seg 的目标切换能力。如图6(a)所示，对于同一张 CAMUS 查询图像，US-Seg 能够在左心室腔、心肌和左心房之间准确切换，Dice分数分别达到 97.48%、94.56% 和 94.72%。在域外 FUGC 数据集上也观察到类似现象：模型能够成功区分宫颈前唇和宫颈后唇，Dice分数分别达到 97.18% 和 95.17%。由于查询图像和模型参数均保持不变，这些结果表明，分割目标由已标注参考图像决定，而不是由固定的输出类别或查询图像中视觉上最显著的结构决定。具体而言，参考掩膜指定了目标身份，而参考图像则提供了其结构、视觉和解剖背景。因此，无需重新训练或进行特定任务重新设计，US-Seg 仅通过参考图像即可切换预期目标。

实用的标签参考模型不仅应能正确响应目标语义的变化，还应在同一目标的不同参考实例之间保持可靠的性能。遵循第IV-A节所述的评估方案，我们分别将100个带标注病例中的每一个作为参考，并在固定查询集上进行评估，所得性能分布汇总于图6(b)中。在域内的颈部和胃数据集上，US-Seg表现出高度集中的性能分布，Dice中位数分别达到81.03%和83.05%，其四分位距仅为0.17%和0.22%。更重要的是，在域外评估中，这种鲁棒性仍得以良好保持。在未见过的卵巢和前列腺数据集上，Dice中位数仍分别高达76.68%和78.19%，而四分位距仅为1.28%和2.88%。持续较高的中位数和紧凑的分布表明，即使面对训练期间未遇到的解剖结构和数据分布，US-Seg也能从多样的参考实例中可靠地传递目标信息。因此，无需仔细选择或筛选有利的参考病例，也能获得出色的分割性能。

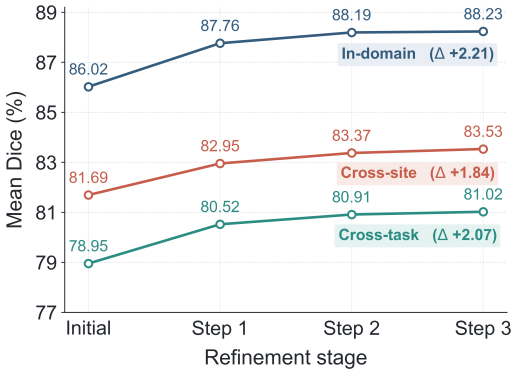
G. 通过迭代细化进行渐进式结构校正

为考察预测感知迭代细化如何校正残余结构误差，我们展示了经过三个细化步骤后中间预测的演变过程。如图7(a) 所示，初始预测已能定位目标，但存在明显的边界偏差，且结构恢复不完整。随着预测反馈不断融入潜在状态更新，这些误差得到了逐步校正。对于甲状腺示例，经过三个细化步骤，Dice系数从 71.92% 依次升至 80.63%、83.45% 和 84.36%，预测轮廓也逐渐与目标边界更加吻合。颈动脉示例也呈现出类似趋势，Dice系数从 70.61% 依次升至 81.21%、83.03% 和 83.66%。这些逐步变化表明，解码器并非只是重复生成掩码；相反，每个中间预测都会引导后续步骤校正残余结构误差。

我们进一步在图 7(b) 中对完整评估集上的这一表现进行了量化。对于域内评估，迭代细化将平均 Dice系数从 86.02% 提高至 88.23%；对于跨站点评估，从 81.69% 提高至 83.53%；对于跨任务评估，从 78.95% 提高至 81.02%，相应的总体增幅分别为 2.21%、1.84% 和 2.07%。大部分提升是在前两个细化步骤中实现的，此时三种设置下的 Dice系数已分别升至 88.19%、83.37% 和 80.91%，而第三步仅带来



（a）逐步结构细化。



（b）跨细化步骤的性能演变。

图7. 采用预测感知迭代细化的渐进式结构校正。黄色表示真实标注，绿色表示预测结果。

表IV US-Seg的消融研究。BLM、MBI和PIR分别表示边界中心标签掩蔽、多级双向参考-查询交互和预测感知迭代细化。

组件			域内		跨站点		跨任务	
BLM	MBI	PIR	Dice	IoU	Dice	IoU	Dice	IoU
✓			82.21	74.80	72.34	63.88	69.72	57.96
✓		✓	84.67	76.85	76.91	67.25	72.34	60.89
	✓	✓	85.48	76.39	76.35	66.41	73.92	62.31
✓	✓		86.02	77.31	81.69	73.05	78.95	69.90
✓	✓	✓	**88.23**	**80.89**	**83.53**	**75.69**	**81.02**	**73.39**

0.04%、0.16% 和 0.11% 的微小增益。增益逐渐减小表明结果正逐步收敛，同时也证实，迭代细化无需更新参数即可有效纠正残余结构误差。

H. US-Seg 设计的有效性

表 IV 评估了以边界为中心的标签掩膜（BLM）、多级双向参考-查询交互（MBI）和预测感知迭代细化（PIR）在域内、跨站点和跨任务设置下的贡献。完整的 US-Seg 在这三种设置下均取得最佳性能，Dice分数分别为 88.23%、83.53% 和 81.02%。

移除 BLM 后，在域内、跨站点和跨任务设置下，Dice系数分别下降 2.75%、7.18% 和 7.10%。在域外评估中，性能下降幅度显著更大，这凸显了以边界为中心的重建对于超越熟悉成像分布实现结构泛化的重要性。移除 MBI 后，Dice 系数分别下降 3.56% 、 6.62% 和 8.68%，其中跨任务评估下的降幅最大。相应的跨任务交并比也下降了 12.50%。这一结果证实，多层级双向交互对于跨越参考图像与查询图像之间的差异迁移参考定义的目标至关重要。移除 PIR 后，在三种设置下，Dice系数分别稳定下降 2.21%、1.84% 和 2.07%，表明迭代预测反馈可通过纠正残余结构误差带来额外收益。

总体而言，BLM、MBI 和 PIR 分别在结构泛化、参考图像引导的目标迁移以及残差纠正方面提供了互补收益。

V. 讨论

US-Seg 将通用超声分割从可迁移表征学习推进到直接的、基于参考图像的分割，无需进行目标特定适应。接下来，我们将讨论这一表述如何拓展分割级通用性、改变目标指定范式，并支持更具扩展性的临床交互，同时也会探讨其仍存在的局限性。

A. 从通用表征到通用分割

超声基础模型已表明，异质数据预训练能够生成可迁移至不同器官、成像条件和下游任务的表征。然而，这种表征层面的通用性并不能直接带来通用分割。现有框架仍依赖特定任务解码器或目标数据适配，将可迁移特征转化为分割输出；这意味着其通用性主要体现在特征复用上，而非直接生成适应不同目标的掩膜。

US-Seg 将通用性从可复用表征扩展到了直接、目标灵活的分割。通过在推理时使用已标注参考图像来定义目标，同一模型便可跨成像中心运行，并在不修改输出空间或使用目标数据进行微调的情况下，实现训练任务空间之外的分割任务。值得注意的是，联合训练的 US-Seg 在平均域内Dice系数上也比特定数据集的 USFM基线高出 1.25%，这表明，更广泛的泛化能力并不意味着必须牺牲在熟悉分布上的性能。这些结果说明，通用超声分割不仅需要可迁移特征，还需要一种能够将共享结构知识转化为可直接执行的分割任务的预测机制。

B. 从提示驱动交互到可复用的目标指定

现有通用超声分割方法主要通过任务提示或空间提示来实现目标灵活性。任务提示方法（如 MOFO 和 Percept-Guide）会显式编码目标身份，但仍局限于训练期间定义的分割任务。空间提示方法（包括 MedSAM、UltraSAM 和 SAMUS）能够适应更灵活的目标，但需要针对查询进行交互，因此在分布外评估中的标注成本约为 US-Seg 的 62 倍，同时 Dice 系数低 5.49%。这两种范式在预定义的目标身份与针对查询的灵活性之间进行权衡，但都未能提供一种可在不同图像间复用的目标指定机制。

标签参考方法通过已标注示例定义预期结构，从而提供了一种更具可复用性的目标指定方式。这类方法仍需在标注负担与迁移可靠性之间权衡。UniverSeg 和 Tyche 通过使用多个已标注参考来提升鲁棒性，而 PerSAM 和 INSID3 等单参考图像方法则大幅降低了标注需求，但仍容易受到较大参考-查询差异的影响。在我们的实验中，即使使用 16 个已标注参考，表现最强的多参考基线 UniverSeg 在分布外 Dice 上仍比 US-Seg 低 24.30%；而在相同参考图像预算下，表现最强的单参考图像基线 INSID3 仍低 45.27%。这些结果表明，有效的基于参考图像的分割并不只是依赖增加标注数量，还取决于学习可迁移的结构并建立鲁棒的参考-查询对应关系。

US-Seg 通过结合单参考图像目标指定、可迁移的结构学习以及鲁棒的参考-查询交互，解决了这一权衡，使单个已标注参考能够在不同查询中持续定义目标，而无需反复进行空间交互或目标特定适应。

C. US-Seg 的临床意义

US-Seg 提供了一种实用方法，可将超声分割扩展至目标、成像条件或数据集可能随时间变化的临床场景。遇到新的设置时，临床医生无需重新训练模型、进行额外微调，或针对每张图像反复提示；只需通过一张已标注参考图像定义一次预期目标，即可将其应用于后续病例。这对于队列级分析、纵向随访、回顾性数据库研究和新引入的测量尤为重要，因为反复交互或针对特定目标进行训练会显著限制可扩展性。

这一范式的主要临床价值在于，将目标指定从对每幅图像反复执行的操作，转变为一次性且可重复使用的指令。尽管生成完整的参考掩膜比单个点提示需要付出更多努力，但这一成本可以在后续查询中摊销。此外，观察到的对不同参考实例的鲁棒性表明，部署并不依赖于精心选择某个特殊的参考病例。结合无需适配的跨站点和跨任务泛化，这些特性使 US-Seg 特别适用于可扩展的临床工作流程，在此类流程中，分割需求不断演变，却无需相应增加标注和模型维护负担。

D. 局限性与未来方向

当前研究主要聚焦于二维 B 模式超声。未来工作将把 US-Seg 扩展至更广泛的超声设置，包括时间序列、三维成像和多模态采集，以利用更丰富的解剖与功能信息。此外，尽管目前的评估涵盖了多种器官、临床来源和具有挑战性的域外场景，仍有必要在规模更大的前瞻性多中心队列中进一步验证，以评估其在更广泛临床人群和工作流程中的适用性。

VI. 结论

本文提出了 US-Seg，这是一种通用超声分割框架，只需使用一个已标注参考图像指定目标，即可直接分割查询图像，无需进行目标特定适应。通过在统一的参考引导框架中对异构分割任务进行建模，US-Seg 使单一模型能够在保持强大域内性能的同时，有效泛化至不同成像中心和分割任务。大量实验进一步表明，仅凭一个可重复使用的参考图像即可准确、稳定地指定目标，且与反复进行空间提示或使用多个参考图像相比，标注和交互成本大幅降低。这些发现为实现通用超声分割提供了一条可扩展且具有良好泛化能力的路径，使其不再受限于固定的成像分布和预定义任务空间。

参考文献

[1] J. Jiao, J. Zhou, X. Li, M. Xia, Y. Huang, L. Huang, N. Wang, X. Zhang, S. Zhou, Y. Wang, 等，“Usfm：一种通用超声基础模型，可泛化至不同任务和器官，以实现标签高效的图像分析”，《*Medical Image Analysis*》，第96卷，文章编号103202，2024。[2] L. Chen, Y. Bian, J. Zeng, Q. Meng, W. Zhu, F. Shi, C. Shao, X. Chen, and D. Xiang，“风格一致性无监督域适应医学图像分割”，*IEEE* 图像处理汇刊，第33卷，第4882–4895页，2024。[3] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo，等，“Segment Anything”，载于*2023 IEEE/CVF* 国际计算机视觉大会（*ICCV*），第3992–4003页，IEEE，2023。[4] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang，“医学图像中的 Segment Anything”，《自然·通讯》，第15卷，第1期，文章编号654，2024。[5] X. Lin, Y. Xiang, L. Yu, and Z. Yan，“超越 SAM 适配：通过自动提示实现端到端超声图像分割”，载于国际医学图像计算与计算机辅助干预会议，第24–34页，Springer，2024。[6] A. Meyer, A. Murali, F. Zarin, D. Mutter, and N. Padoy，“UltraSAM：利用大型开放获取分割数据集构建的超声基础模型”，《国际计算机辅助放射学与外科学杂志》，第21卷，第1期，第93–102页，2026。[7] H. Chen, Y. Cai, C. Wang, L. Chen, B. Zhang, H. Han, Y. Guo, H. Ding, and Q. Zhang，“基于任务提示和解剖先验的通用超声图像分割多器官基础模型”，*IEEE* 医学影像汇刊，第44卷，1018，2024 2，1005 no. pp. –。

[8] Z. Lin, S. Li, S. Wang, Z. Gao, Y. Sun, C.-T. Lam, X. Hu, X. Yang, D. Ni, and T. Tan，“超声成像的编排学习框架：提示引导的超感知与注意力匹配下游同步”，《Medical Image Analysis》，第104卷，103639，2025 p. .[9] V. I. Butoi, J. J. G. Ortiz, T. Ma, M. R. Sabuncu, J. Guttag, and A. V. Dalca，“UniverSeg：通用医学图像分割”，载于2023 IEEE/CVF 国际计算机视觉大会（ICCV），第21381–21394页，IEEE，2023。[10] C. Cuttano, G. Trivigno, C. Reich, D. Cremers, C. Masone, and S. Roth，“INSID3：使用 DINOv3 进行免训练上下文分割”，载于 IEEE/CVF 计算机视觉与模式识别会议（CVPR）论文集，2026。[11] Y. Zhu, S. Wang, T. Zhou, Z. Li, H. Zhang, and L. Shao，“通过动态语义匹配实现跨域小样本医学图像分割”，IEEE 图像处理汇刊，2025。[12] Y. Zhang, S. Tian, M. Liao, G. Hua, W. Zou, and C. Xu，“学习形状不变表示以实现可泛化语义分割”，IEEE 图像处理汇刊，第32卷，第5031–5045页，2023。[13] M. Rakic, H. E. Wong, J. J. G. Ortiz, B. A. Cimini, J. V. Guttag, and A. V. Dalca，“Tyche：医学图像分割的随机上下文学习”，载于2024 IEEE/CVF 计算机视觉与模式识别会议（CVPR），第11159–11173页，IEEE，2024。[14] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li，“医学图像分割的边界感知上下文神经网络”，《Medical Image Analysis》，第78卷，文章编号102395，2022。[15] W. Qi, H.-C. Wu, and S.-C. Chan，“Mdf-net：用于超声图像乳腺肿瘤分割的多尺度动态融合网络”，IEEE 图像处理汇刊，第32卷，第4842–4855页，2023。[16] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar，“面向通用型医疗人工智能的基础模型”，《Nature》，第616卷，第7956期，第259–265页，2023。[17] J. Qiu, L. Li, J. Sun, J. Peng, P. Shi, R. Zhang, Y. Dong, K. Lam, F. P.-W. Lo, B. Xiao, et al.，“健康信息学中的大型人工智能模型：应用、挑战与未来”，IEEE生物医学与健康信息学期刊，第27卷，第12期，第6074–6087页，2023。[18] G. Wang, J. Wu, X. Luo, X. Liu, K. Li, and S. Zhang，“Mis-fm：使用在大型未标注数据集上预训练的基础模型进行三维医学图像分割”，arXiv 预印本 arXiv:2306 16925，2023 . .[19] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, et al.，“用于从视网膜图像中实现可泛化疾病检测的基础模型”，《Nature》，第622卷，第7981期，第156–163页，2023。[20] Z. Wang, C. Liu, S. Zhang, and Q. Dou，“通过大规模自监督预训练构建用于内窥镜视频分析的基础模型”，载于国际医学图像计算与计算机辅助干预会议，第101–111页，Springer，2023。[21] Q. Kang, Q. Lao, J. Gao, W. Bao, Z. He, C. Du, Q. Lu, and K. Li，“Urfm：用于推进超声图像诊断的通用超声表示基础模型”，《iScience》，第28卷，第8期，2025。[22] C. Ma, J. Jiao, S. Liang, J. Fu, Q. Wang, Z. Li, Y. Wang, and Y. Guo，“TinyUSFM：迈向紧凑高效的超声基础模型”，IEEE生物医学与健康信息学期刊，2026。[23] X. Wang, X. Zhang, Y. Cao, W. Wang, C. Shen, and T. Huang，“SegGPT：迈向上下文中分割一切”，载于2023 IEEE/CVF国际计算机视觉会议（ICCV），第1130–1140页，IEEE，2023。[24] R. Zhang, Z. Jiang, Z. Guo, S. Yan, J. Pan, H. Dong, Y. Qiao, G. Peng, and H. Li，“通过单次使用个性化 Segment Anything 模型”，载于国际学习表征会议，第2024卷，第18250–18279页，2024。[25] A. Dosovitskiy，“一幅图像胜过16×16个词：用于大规模图像识别的 Transformer”，arXiv 预印本 arXiv:2010.11929，2020。[26] A. Pawłowska, A. ´Cwierz-Pie´nkowska, A. Domalik, D. Jagu´s, P. Kasprzak, R. Matkowski, Ł. Fura, A. Nowicki, and N. ˙Zołek，“用于超声乳腺病变分析的精选基准数据集”，《Scientific Data》，第11卷，第1期，文章编号148，2024。[27] W. G´omez-Flores, M. J. Gregorio-Calas, and W. Coelho de Albu- querque Pereira，“BUS-BRA：用于评估计算机辅助诊断系统的乳腺超声数据集”，《Medical Physics》，第51卷，第4期，第3110–3123页，2024。[28] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy，“乳腺超声图像数据集”，《Data in Brief》，第28卷，文章编号104863，2020。[29] Z. Ye, J. Huang, Y. Zhang, J. Deng, J. Zhang, S. Liu, D. Wang, L. Mei, and C. Lei，“Dflnet：解耦特征学习网络乳腺癌超声图像分割”，《数字信号处理》，第165卷，文章编号105331，2025年。[30] L. Pedraza、C. Vargas、F. Narv´aez、O. Dur´an、E. Mu˜noz和E. Romero，“开放获取的甲状腺超声图像数据库”，载于第十届医学信息处理与分析国际研讨会，第9287卷，第188–193页，SPIE，2015年。[31] H. Gong、J. Chen、G. Chen、H. Li、G. Li和F. Chen，“用于甲状腺结节超声分割的甲状腺区域先验引导注意力”，《生物与医学中的计算机》，第155卷，文章编号106389，2023年。[32] S. Leclerc、E. Smistad、J. Pedrosa、A. Østvik、F. Cervenansky、F. Espinosa、T. Espeland、E. A. R. Berg、P.-M. Jodoin、T. Grenier、等，“利用开放的大规模二维超声心动图数据集进行分割的深度学习”，*IEEE*《医学影像汇刊》，第38卷，第9期，第2198–2210页，2019年。[33] C. J. Belasso、B. Behboodi、H. Benali、M. Boily、H. Rivaz和M. Fortin，“Luminous数据库：基于超声图像的腰椎多裂肌分割”，《*BMC*肌肉骨骼疾病》，第21卷，第1期，文章编号703，2020年。[34] A. Montoya、Hasnin、kaggle446、shirzad、W. Cukierski和yf-fud，“超声神经分割。”https://kaggle.com/competitions/ultrasound-nerve-segmentation，2016年。Kaggle。[35] G. Chen、J. Bai、Z. Ou、Y. Lu和H. Wang，“PSFHS：用于基于人工智能分割耻骨联合和胎头的产时超声图像数据集”，《*Scientific Data*》，第11卷，第1期，文章编号436，2024年。[36] Q. He、S. Bano、J. Liu、W. Liu、D. Stoyanov和S. Zuo，“Query2：通过查询查询改进内镜超声中的胃肠道间质瘤检测”，《生物与医学中的计算机》，第152卷，文章编号106424，2023年。[37] N. Vallez、G. Bueno、O. Deniz、M. A. Rienda和C. Pastor，“BUS-UCLM：乳腺超声病灶分割数据集”，《*Scientific Data*》，第12卷，第1期，文章编号242，2025年。[38] M. H. Yap、G. Pons、J. Marti、S. Ganau、M. Sentis、R. Zwiggelaar、A. K. Davison和R. Marti，“使用卷积神经网络自动检测乳腺超声病灶”，*IEEE*生物医学与健康信息学期刊，第22卷，第4期，第1218–1226页，2017年。[39] J. Zhou、X. Jia、D. Ni、A. Noble、R. Huang、T. Tan和M. T. Van，“超声图像中的甲状腺结节分割与分类”，2020年3月。[40] R. Singla、C. Ringstrom、G. Hu、V. Lessoway、J. Reid、C. Nguan和R. Rohling，“开放肾脏超声数据集”，载于医学超声简化技术进展国际研讨会，第155–164页，Springer，2023年。[41] J. Bai、Y. Tang、Z. Zhou、M. Islam、M. Tabassum、E. Almar-Munoz、H. Liu、H. Meng、N. Lv、B. Deng、等，“FUGC：宫颈分割半监督学习方法基准测试”，*IEEE*《医学影像汇刊》，2026年。[42] H. Jiang、M. Imran、P. Muralidharan、A. Patel、J. Pensa、M. Liang、T. Benidir、J. R. Grajo、J. P. Joseph、R. Terry、等，“Microseg-net：一种基于深度学习的微型超声图像前列腺分割方法”，《计算机化医学成像与图形学》，第112卷，文章编号102326，2024年。[43] Q. Zhao、S. Lyu、W. Bai、L. Cai、B. Liu、G. Cheng、M. Wu、X. Sang、M. Yang和L. Chen，“MMOTU：用于无监督跨域语义分割的多模态卵巢肿瘤超声图像数据集”，*arXiv*预印本 *arXiv:2207.06799*，2022年。[44] F. Isensee、P. F. Jaeger、S. A. Kohl、J. Petersen和K. H. Maier-Hein，“nnU-Net：一种用于基于深度学习的生物医学图像分割的自配置方法”，《自然·方法》，第18卷，第2期，第203–211页，2021年。