arXiv:2302.02125v1 [eess.IV] 4 Feb 2023

利用几何先验和对比相似度的弱监督三维医学图像分割

Hao Du，Qihua Dong，Yan Xu，Jing Liao

摘要—医学图像分割几乎是计算机辅助诊断中最重要的预处理过程，但由于分割目标形状复杂，以及医学成像造成的各种伪影（即低对比度组织和非均匀纹理），这也是一项极具挑战性的任务。本文提出了一种简单而有效的分割框架，以基于损失的方式将几何先验和对比相似度融入弱监督分割框架中。所提出的基于点云构建的几何先验为弱监督分割候选区域提供了细致的几何信息，相较于边界框标注固有的属性（即高度和宽度），能够提供更好的监督。此外，我们提出对比相似度，以促使器官像素在对比嵌入空间中聚集，从而更好地区分低对比度组织。所提出的对比嵌入空间能够弥补传统使用的灰度空间在表征能力上的不足。我们在多个数据集上开展了大量实验，以验证所提出弱监督分割框架的有效性和鲁棒性。所提出的框架在以下公开可获取的数据集上优于现有最先进的弱监督方法：LiTS 2017 挑战赛、KiTS 2021 挑战赛和LPBA40。我们还对方法进行了剖析，并评估了各组成部分的性能。

关键词—弱监督分割、医学图像分割、对比相似度、几何先验、点云

I. 引言

SEGMENTATION 对于理解和解读医学图像至关重要，也是各种疾病诊断、治疗和后续康复的必要环节。随着深度卷积神经网络（CNNs）的兴起，这项任务近年来得到了广泛研究 [1], [2]。然而，其主要局限在于，这些方法需要大量带有像素级标注的训练图像。收集并标注这些训练图像的成本极高，这在很大程度上阻碍了深度卷积神经网络在医学领域的性能，也限制了其可扩展性。为减少对像素级标注的需求，一种常用范式是基于边界框标注的弱监督分割 [3]–[7]。该方法利用边界框标注生成候选区域，并将其作为伪标签，从而模拟全监督。

然而，尽管这些工作在某些实际场景中取得了良好的表现，但其适用性可能受到两个因素的限制：1）复杂形状：一些器官具有精细的结构，即肾内变异，在没有像素级监督的情况下难以精确分割；2）成像伪影：正如先前工作 [8]–[10], 中所讨论的，由技术或物理问题引起的各种医学成像伪影，使低对比度组织和非均匀纹理难以区分，尤其是在传统上广泛使用的灰度空间中。复杂形状和成像伪影极大地限制了弱监督分割模型在许多场景中的适用性，尤其是在分割复杂结构时。

为应对复杂形状带来的挑战，我们提出利用标准器官模板学习器官的几何先验。我们不使用体积表示，而是首先利用反向网格化 [11] 将分割

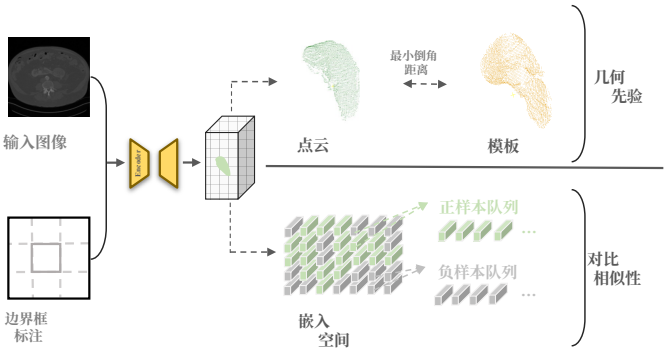


图1. 所提出的弱监督分割框架示意图。如图所示，我们为弱监督分割提出了几何先验和对比相似度。顶行展示了我们方法的几何先验。我们首先将传统使用的体积表示转换为点云表示，并将模板器官配准到预测器官。随后，我们最小化二者之间的倒角距离。底行解释了所提出的对比相似度的核心思想。通过将像素划分为正像素和负像素，我们促使器官像素在嵌入空间中聚集，从而更好地分割低对比度器官。

Hao Du、Qihua Dong 和 Jing Liao 任职于中国香港香港城市大学计算机科学系（电子邮箱：haodu8-c@my.cityu.edu.hk、qihuadong2-c@my.cityu.edu.hk 和 jingliao@cityu.edu.hk）。Yan Xu 就职于中国北京的北京航空航天大学生物科学与医学工程学院（电子邮箱：xuyan04@gmail.com）。Hao Du 和 Qihua Dong 对本工作贡献相同。通讯作者：Jing Liao（jingliao@cityu.edu.hk）和 Yan Xu（xuyan04@gmail.com）。本工作得到香港特别行政区创新科技署（ITC）ITF项目MHP/109/19 的资助，以及国家自然科学基金委员会项目62022010、中国111计划项目B13003和北京航空航天大学高性能计算（HPC）资源的支持。

结果从体积表示转换为点云表示，然后在点云空间中将其与模板进行比较。与体积表示相比，点云表示的基本单元粒度更细、灵活性更高（，即，灵活的点相较于均匀体素网格），有助于更好地描述精细的几何结构。另一方面，与传统使用的灰度空间 [8]，不同，我们利用对比学习 [12]将像素编码到高维嵌入空间中，并促使相同标签的像素聚集在一起。与灰度空间相比，这有助于缓解成像伪影，并在嵌入空间中获得更丰富的表达能力。

本文提出了一种新颖的弱监督分割框架，率先尝试将几何先验和对比相似度结合起来。该框架还具有通用性，可轻松用于改进多种采用边界框标注的弱监督分割模型，即 Ai+L [13], BoxInst [7]。通过从给定模板中学习几何先验，并利用对比相似度区分低对比度组织，我们的方法仅凭边界框监督即可生成高质量结果。

我们的方法由两个主要组件构成。在几何先验组件中，候选区域的形状受到由点云表示的给定模板约束。通过根据给定模板最小化距离，优化候选区域的外部边界和内部结构。第二个组件是对比相似度，用于解决医学成像伪影带来的问题。通过预训练对比头，我们成功学习了器官像素与非器官像素之间的差异。与传统上广泛使用的灰度空间相比，该组件能够更好地区分低对比度组织和非均质纹理。通过大量实验，我们证明了该方法能够生成高质量的分割结果，同时保留精细的内部细节和准确的边界。我们还表明，在类似设置下，我们的方法优于其他基于边界框监督的弱监督方法 [5], [7] 。此外，我们进行了大量实验，以验证该方法各组件的有效性。

总而言之，我们的主要贡献有三点。

- 我们提出了一种简单而有效的基于边界框标注的弱监督分割框架，可轻松应用于多种弱监督分割模型并提升其性能。
- 我们在点云表示中引入几何先验，以更好地引导形状学习，尤其适用于具有复杂结构的器官。
- 所提出的对比相似度弥补了传统灰度空间表示能力不足的问题，因此能够更好地区分存在医学成像伪影的组织。

我们的代码和数据将公开发布，以供后续研究使用。

II. 相关工作

在本节中，我们首先回顾自然图像和医学图像分割领域中使用边界框标注的现有弱监督医学分割方法，然后讨论近期利用几何先验的研究，最后介绍对比相似度的发展趋势。

A. 弱监督医学语义分割

通常，根据弱标注的类型，弱监督分割方法可分为四类：涂鸦标注 [14],点标注 [15], [16], 图像级标签 [17], [18] 和边界框标注 [5]。涂鸦标注和点标注监督至少为每个区域标注一个涂鸦或点，标注区域将直接纳入分割损失的计算。Wang等 [19] 提出利用随机游走算法 [20] 为未标注区域生成初始候选区域，然后使用初始分割区域监督分割模型的训练。Qu等 [16] 采用了类似的训练流程，但使用了不同的标签生成方法，该方法结合了K均值聚类和Voronoi划分图。Xu等 [18] 通过多实例学习（MIL）将图像级标签扩展为实例级标签，并仅使用体积级标签对图像进行分割。

由于简单且标注成本低，采用边界框标注的弱监督分割在医学图像分割领域日益受到关注。在实际场景中，我们可以使用两个角点坐标定义边界框，这些坐标易于存储。此外，边界框标注具有位置感知能力，能够提供目标对象的空间关系，因此成为近期研究的热门方向之一 [21]–[24]。在早期阶段，研究人员 [22], [25] 提出将边界框内的像素视为前景像素，并利用这些噪声标签训练分割框架。尽管这种方案取得了良好的性能，但在交替生成过程中可能会不断累积误差。最近，研究人员 [7] 尝试直接生成分割结果，而不是采用容易产生误差的替代方式。通常，他们构建一个掩码头来生成分割结果，并使用边界框标注训练该掩码头。在本文中，我们遵循这一分割方案，直接通过掩码头生成分割结果。此外，为了解决前述复杂形状和成像伪影问题，我们分别提出了几何先验和对比相似度。

B. 几何先验

与自然图像不同，医学图像中存在明显的解剖先验（，即图谱先验），具体体现在人体器官上（即形状和位置）。现有融合此类解剖先验的研究大多分为两类：基于损失的方法和基于图的方法。通常，基于图的方法 [26]–[30] 利用解剖结构出现概率图构建图模型，并根据输入图像灰度空间估计前景概率。为提高分割准确率，研究中采用了基本形式的外观模型 [29]。Gao 等 [30]提出使用外观卷积神经网络来表征前景。尽管这些方法取得了较高的准确率，但图模型给分割框架带来了沉重

且高昂的计算负担，使其在某些场景中难以应用。

将先验融入分割框架的另一种常见思路是基于损失的方法。研究者 [31]–[36]大多通过最小化分割网络输出与预定义解剖先验之间的距离来实现这一点。已有一些工作 [31], [32]在训练目标中加入正则项（即解剖邻接关系或边界条件）。此外，还会在潜在特征空间 [35], [36]中计算预测结果与图谱先验之间的距离。

与基于图的方法相比，基于损失的方法能以灵活的方式，在更广泛的尺度范围内融入解剖学先验，同时保持分割框架的计算效率。然而，先前的基于损失的方法在以下两个方面未能解决上述问题：1）以往的工作通常将体积切分成 2D 或 3D 图像块，再依次处理以节省内存。但这种划分方式会破坏全局几何关系，导致分割性能下降。与之不同，我们在三维嵌入空间中学习几何先验，以捕捉整体几何形状，并提出完整性头来确保候选区域的形状完整性。2）不同于采用体积表示的以往工作，我们利用反向网格化 [11] 将体积表示转换为点云表示。与受均匀体素网格约束的体积表示相比，不受网格限制的点云在表示精细结构时更加灵活。

C. 对比学习

对比学习旨在将特征空间划分为正负数据对，从而拉近正样本并推远负样本。对于语义分割，对比学习主要用于预训练 [37]–[39]。Van 等 [40] 将其用于区分来自各种显著掩码的特征，展现了其在无监督设置下的优势。Wang 等 [41]通过在像素和区域层面进行学习，展示了对比学习的优势。一些研究人员利用对比学习来解决医学图像分割中耗时的逐像素标注问题。Chaitanya 等 [12] 提出了一种两阶段自监督对比学习框架，在预训练阶段利用无标签数据，通过全局和局部机制学习特征匹配。Hu 等 [42]提出了一种半监督方案，用于学习自监督全局对比和有监督局部对比。在我们的工作中，我们发现，常用的灰度空间不足以区分正负像素（即器官像素和非器官像素），尤其是在磁共振成像中。因此，我们利用对比学习来计算像素之间的对比相似度。通过将像素编码为高维特征，并促使相同标签的像素在嵌入空间中聚集，从而增强判别能力，缓解灰度空间在处理医学成像伪影（即超声成像中的伪影）和相似周围组织时表现不佳的问题。

III. 方法

A. 总体框架

给定输入图像 $I \in \mathbb{R}^{S\times H\times W}$ （S 表示切片数量， H 表示高度， W 表示宽度）及其对应的边界框标注 $\mathbb{B}^{1\times 6}$（由左上角和右下角坐标限定），我们的弱监督框架$\mathbf{F}(\cdot)$得到像素级分割掩码 $\mathcal{M} = \mathbf{F}(I)$，训练目标是最小化损失函数 L_{frame}：

$$\min_{\mathbf{F}} L_{\text{frame}}(I, \mathbb{B}, \mathbf{F}) \tag{1}$$

流程 遵循 nnUNet [2], ，我们从原始输入图像中随机采样一个输入图像块 $p \in \mathbb{R}^{S'\times H'\times W'}$ ，并使用卷积网络编码器 E对这些图像块进行编码。与 [1], 类似，我们采用多层卷积网络作为解码器 G 来获得特征图 $P \in \mathbb{R}^{S'\times H'\times W'}$，其中解码器与编码器具有相同的层数。如图2所示，我们将所提出的几何先验和对比学习融入掩码头 的训练中，以解决前述复杂形状和成像伪影问题。训练损失 L_{frame} 由两部分组成： L_{ori}和 L_{mask}

$$L_{\text{frame}} = L_{\text{ori}} + L_{\text{mask}} \tag{2}$$

其中， L_{ori} 表示标准弱监督框架的原始训练损失（即，BoxInst [7]中的 L_{fcos} ），L_{mask} 表示掩码头的训练损失。在接下来的段落中，我们主要讨论掩码头的训练。掩码头的训练可以表示为式（3）。

$$L_{\text{mask}} = L_{\text{geo}} + L_{\text{cons}} \tag{3}$$

掩码头生成二值分割掩码，并通过我们提出的几何先验 L_{geo} 和对比相似度 L_{cons}进一步优化。具体而言，对于几何先验损失，我们构建了完整性头 ，用于预测每个候选区域的完整性分数，以表示输入中目标完整存在的条件概率。每个完整的候选区域都会转换为点云，并与模板器官的点云进行配准。我们采用倒角距离损失来最小化二者之间的距离。在对比相似度损失方面，我们构建了一个对比头 ，通过特征图获得对比相似度；对比头会为特征图中的每个位置分配正标签或负标签。接下来的段落中，我们首先介绍几何先验，然后详细阐述所提出的对比相似度的技术细节。

B. 几何先验

提出的几何先验应用于三维点云空间，原因有二：1）我们发现，二维切片无法很好地保留三维器官的几何连续性。因此，我们在三维嵌入空间中学习器官的几何形状。2）传统的体积表示无法处理精细结构的分割。体积表示的表达能力受到均匀体素网格的极大限制。为此，我们利用反向网格化 [11] 在点云嵌入空间中处理复杂形状的分割。

如图2所示，我们提出几何先验，以更好地对掩码头的训练进行弱监督。几何先验指模板器官的边界形状和内部空间分布。具体而言，我们引入反向网格化 [11] ，以构建体积表示与点云表示转换之间的桥梁。这有助于摆脱体积表示带来的表示限制。将模板器官$\mathbb{T}$ 和候选区域 $\mathbb{S}$ 都转换为点云表示后，我们使用广泛应用的 ICP 配准工具 [43]将模板器官配准到候选区域。最后，我们最小化点云嵌入空间中模板器官与候选区域之间的倒角距离所对应的几何先验损失。下面我们先介绍点云的转换与配准，然后说明几何先验损失的定义。

转换与配准为了利用点云表示丰富的表达能力，我们引入反向网格化 [11] ，以实现体积表示与点云表示之间的转换。对于每个体素网格，反向网格化会计算对应网格八个顶点的加权和，并将该加权和赋予新点的坐标。与均匀体素网格不同，点标注坐标具有很高的灵活性，使点云表示能够描述精细而复杂的结构。这有助于更好地学习弱监督分割中具有挑战性的器官内变异。此外，我们提出稀疏配准，并在计算 $\mathbb{S}$ 与 $\mathbb{T}$ 之间的倒角距离之前应用该方法。模板的微小旋转会显著影响倒角距离的计算，尤其是在对象结构更为复杂时。因此，我们对候选区域与模板的整体形状进行配准 [43]。具体而言，我们分别在模板和候选区域的区间内均匀采样 20% 的点，然后使用 ICP 配准工具 [43] 计算它们之间的变换矩阵。根据该变换矩阵对模板点云进行配准。

损失 随后，在掩码头 的训练损失函数中应用几何先验，通过优化候选区域 $\mathbb{S}$ 与配准模板器官 $\mathbb{T}$之间的倒角距离来实现。具体可表述如下:

$$L_{\text{geo}} = \frac{1}{|\mathbb{S}|}\sum_{x\in\mathbb{S}}\min_{y\in\mathbb{T}}||x-y||_2 + \frac{1}{|\mathbb{T}|}\sum_{y\in\mathbb{T}}\min_{x\in\mathbb{S}}||y-x||_2. \quad (4)$$

具体而言，掩码头为每个候选区域生成二值分割掩码。候选区域是由分割实例构成的概率分割掩码。随后，使用 Gumbel-Softmax [44] 进一步处理该分割掩码，以获得二值体素。概率较低的位置赋值为 0，反之亦然。然后，我们采用反向网格化 [11] ，从二值体素中获得点云候选区域$\mathbb{S}$ 。最后，计算 $\mathbb{S}$ 与 $\mathbb{T}$之间的倒角距离，作为几何先验损失。

C. 对比相似度

灰度空间对于医学成像中的伪影和相似的周围组织表现不佳。仅靠它不足以区分正像素和负像素（即，器官像素和非器官像素）。因此，我们利用对比学习来计算像素之间的对比相似度。将像素编码为高维特征，并鼓励相同标签的像素在嵌入空间中聚集，有助于提高可区分性。

为计算所提出的对比相似度，我们首先在解码器后接入一个卷积神经网络对比头 。遵循以往的对比学习研究 [42], [45], ，我们构建一个两层逐点卷积 $h(\cdot)$，从特征图 P中提取不同的表示。具体而言，在预训练阶段，我们首先以由粗到精的方式对所提出的对比头 进行预训练，此阶段仅使用边界框标注。我们将像素编码为嵌入特征 $\mathcal{C} = h(P)$，并促使具有相同标签的像素在嵌入空间中聚集。两个像素之间的对比相似度定义为它们在嵌入空间中的距离。为计算整幅图像的对比相似度损失，我们构建一个无向图，其中顶点对应像素，边则连接相邻像素。随后汇总与每条边相关的对比相似度，以计算整体对比相似度损失。下文将先介绍对比头的预训练，再定义对比相似度损失。

预训练 对比头的预训练采用弱监督方式进行，预训练阶段仅使用边界框标注。具体而言，预训练阶段分为两个子阶段：粗略 和细化。在粗略 阶段，我们首先将边界框内的像素标记为正标签，将边界框外的像素标记为负标签。然后利用这些标注训练对比头。然而，噪声标签在很大程度上限制了对比头的性能。因此，我们提出细化 阶段，以进一步提升对比头的性能。在细化 阶段，我们首先随机选取 K 个负像素作为参考像素。然后，对于边界框内的每个像素，我们计算其与所有 K 个参考像素之间的距离$\mathcal{D}$ ：

$$D_{u,v,z} = \sum_{i=1}^{K}\mathbb{1}\{\{\mathcal{C}_{u,v,z}\cdot\mathcal{C}_i\}\geq\tau\} \quad (5)$$

其中，$\mathcal{C}_{u,\ v,\ z}\in R^{S\times H\times W}$表示嵌入特征在（$u$，$v$，$z$）处的特征，$\mathbb{1}$ 表示当距离大于 τ 时取 1，否则取 0；τ 是用于判断像素为正或负的阈值。如果 $D_{u,\ v,\ z}$ 大于$K/2$，则位置（u，v，z）处的像素被视为正像素，反之亦然。然后，我们使用与粗略 阶段相同的训练损失来训练对比头，其公式如下:

$$loss = -\frac{1}{|\Omega|}\sum_{(u,v,z)\in\Omega}\frac{1}{|\mathcal{P}(u,v,z)|}\cdot log\frac{\sum_{(u_p,v_p,z_p)\in\mathcal{P}(u,v,z)}exp(\mathcal{C}_{u,v,z}\cdot\mathcal{C}_{u_p,v_p,z_p}/\tau)}{\sum_{(u_n,v_n,z_n)\in\mathcal{N}(u,v,z)}exp(\mathcal{C}_{u,v,z}\cdot\mathcal{C}_{u_n,p_n,z_n}/\tau)} \quad (6)$$

其中，$\mathcal{C}_{u,v,z}\in R^{S\times H\times W}$表示特征图中位于$(u,v,z)$处的特征，Ω 表示输入中的所有点。

$\mathcal{P}$（u, v, z）表示与位于（u, v, z）处的像素具有相同标签的点集，而 $\mathcal{N}$表示具有不同标签的点集。 τ 是温度常数。

损失 考虑在输入图像 I 上构建的无向图 $G = (V, E)$，其中 V对应像素， E 表示相邻像素之间的边。预测的分割掩膜可视为像素（u, v, z）属于前景的概率。那么，像素（u_1, v_1, z_1）和像素（u_2, v_2, z_2）具有相同标签的概率为：

$$\begin{aligned} Prob(y_e = 1) = \tilde{\mathcal{M}}_{u_1,v_1,z_1} \cdot \tilde{\mathcal{M}}_{u_2,v_2,z_2} \\ + (1 - \tilde{\mathcal{M}}_{u_1,v_1,z_1}) \cdot (1 - \tilde{\mathcal{M}}_{u_2,v_2,z_2}), \end{aligned} \tag{7}$$

其中 $\tilde{\mathcal{M}}$ 表示前景概率掩膜， y_e表示边的标签。

因此，我们可以为每条边定义一个指示量，用于指示其连接的像素是否属于同一标签。如果两个相邻像素之间的对比相似度高于预先定义的阈值 τ，则将连接它们的边上的指示量设为 1，反之设为 0。我们舍弃指示量为 0 的边，并进一步汇总正边的对比相似度损失，其形式如下：

$$L_{\text{cons}} = -\frac{1}{N} \sum_{e \in E_{in}} \mathbb{1}_{\{\mathcal{C}_{e_{start}} \cdot \mathcal{C}_{e_{end}} \geq \tau\}} \log Prob(y_e = 1). \tag{8}$$

这就是整幅图像的对比相似度损失。

IV. 实验

A. 数据集

- LiTS：公开的肝脏LiTS [49] 数据集包含使用各种CT扫描仪和设备采集的201例CT扫描。该数据集中图像的轴向分辨率为0.56毫米至1.0毫米， z 方向的分辨率为0.45毫米至6.0毫米。 z 中的切片数为42至1026。我们按4:1的比例将131个病例划分为训练集和评估集。
- KiTS21：公开可访问的KiTS21 [50]数据集包含2010年至2020年期间的300个病例。数据集中的每次CT扫描均由三位专家标注员标注以下语义类别：肾脏、肿瘤和囊肿。我们按4:1的比例将提供的CT扫描划分为训练集和验证集。
- LPBA40：该数据集包含从随机选取的病例中获得的40个T1加权图像体积（在40次扫描中：20名男性、20名女性，年龄为29.2 ± 6.3岁）。扫描的空间分辨率为0.86 × 1.5 ×0.86毫米^3。在此，我们针对LPBA40中的海马体子集开展实验。.

B. 实现细节与评估指标

所提出的结合几何先验和对比相似度的框架，可轻松应用于任何使用边界框标注的弱监督分割模型。为验证我们方法的鲁棒性，我们在两个模型上开展实验：Ai+L [13] 和 BoxInst [7]。此处，我们采用最新版本的 BoxInst [7]作为基线，除非另有说明，实验均基于该模型进行。

训练设置主要沿用 BoxInst 的训练配置：基础学习率为 0.01，权重衰减为 1e- 4，并采用带预热的 MultiStepLR 调度器。我们的框架在 GeForce RTX 3090 图形处理器上进行训练。在推理阶段，我们根据经验将完整性头的阈值设为 0.6，类别头的阈值设为 0.5。除非另有说明，我们为每个特定数据集随机选择一个训练样本作为模板。我们使用两种广泛采用的度量评估结果：以百分比表示的 Dice 系数（DSC）和豪斯多夫距离（95%）。1）DSC 系数的计算方式为两个掩码的重叠面积除以它们的面积之和。DSC 系数越高，表示与真实标注的重叠越好。2）豪斯多夫距离主要衡量分割结果与逐像素分割掩码之间的边界距离。与较差的结果相比，较好的分割结果具有更小的数值。

预处理与后处理 遵循 nnUNet [2]，预处理包括下采样、图像块划分和数据增强。在这里，我们对数据进行下采样，以减少内存使用，并确保单个图像块内包含完整实例。随后，在后处理阶段，流程包括重采样、图像块划分和图像块非极大值抑制（图像块之间的非极大值抑制）。我们首先调整间距，并以一定的步长将数据划分为与训练时大小相同的图像块。然后，我们为每个图像块预测一个分割掩膜，并将所有预测结果映射回原始空间。由于图像块之间存在重叠区域，我们在这些区域使用图像块非极大值抑制来消除重复项。

边界框标注 我们在 LiTS、KiTS 和 LPBA40 数据集上进行了实验。这三个数据集都具有逐像素分割标注。我们利用相应的逐像素标注来获取对应的边界框标注。在训练阶段，仅使用边界框标注。在推理阶段，我们使用逐像素标注评估性能。

C. 结果

定量结果。我们在表 I 中报告了三个数据集（LiTS17、KiTS21 和 LPBA40）的定量结果。我们的方法旨在结合几何先验和对比相似度，为弱监督分割框架的训练提供更好的监督。与作为我们方法基线的 BoxInst [7]，相比，我们的方法在这三个数据集上的分割性能均有显著提升。这是因为几何先验能够监督外部形状和内部结构的学习，而对比相似度则能在嵌入空间中更好地区分器官像素和非器官像素。同样，我们的方法优于最新的两种由边界框标注监督的弱监督方法 [42], [47]，后者基于多实例学习（即 MIL）。其原因在于，多实例学习主要关注候选区域的边界回归，而我们还利用几何先验学习器官的内部细节。我们还评估了我们方法的上限，即使用真实标注监督框架训练。

定性结果。 我们展示了与两种最先进弱监督分割方法的定性比较

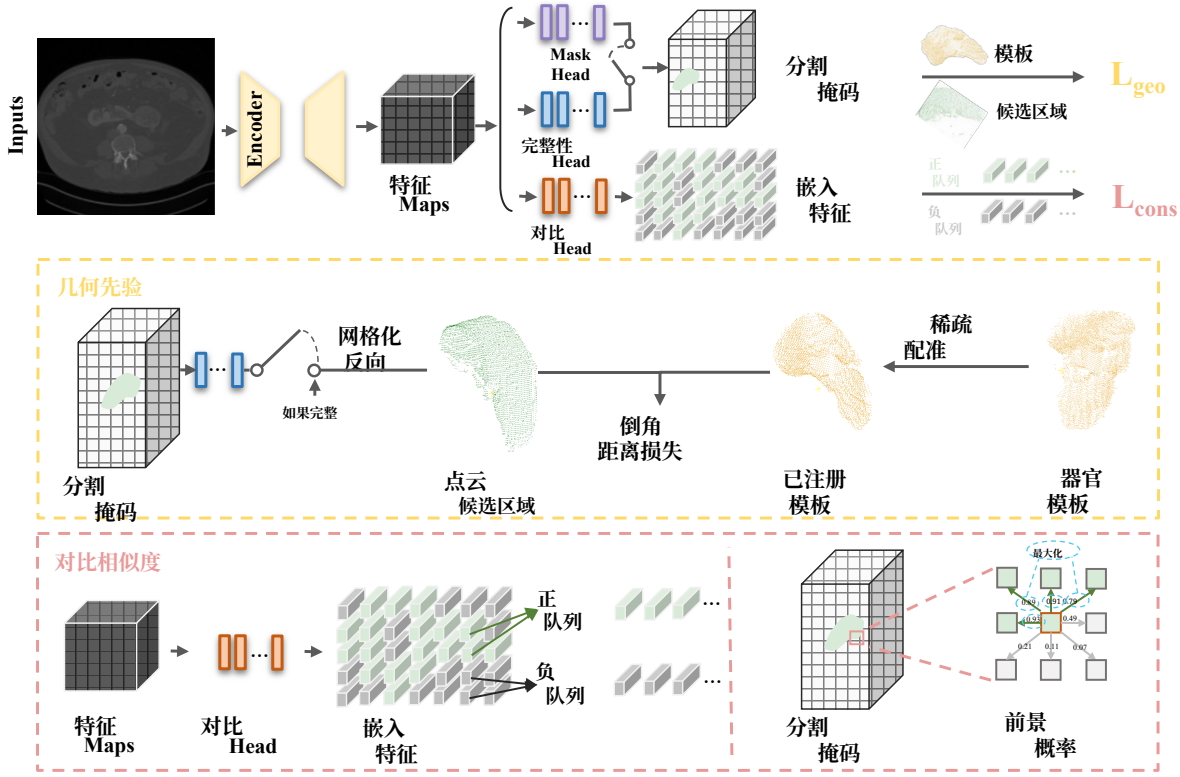


图2. 我们提出的结合几何先验与对比相似度的弱监督分割框架的总体框架。该框架的训练过程如下：首先对输入图像进行预处理，得到嵌入特征图。然后，通过提出的几何先验和对比相似度共同监督掩码头的训练。在几何先验方面，如果分割结果完整，我们首先将其转换为点云，然后计算该结果与配准模板之间的倒角距离。在对比相似度方面，我们首先仅使用边界框标注对比头进行预训练，然后通过预训练的对比头最小化相同标签像素之间的距离。

表 I 与其他方法的性能比较。我们使用两种度量评估性能：Dice 系数和 HD95。† 我们通过其开源代码获得它们的 Dice 系数。 ‡：我们基于其二维开源代码重建了三维框架。我们的方法（无几何先验） 表示移除了几何先验组件。我们的方法（无对比相似度） 表示移除了对比相似性组件。

方法	骨干网络	LiTS17		KITS21		LPBA40	
		↑DSC(%)	↓HD95	↑DSC(%)	↓HD95	↑DSC(%)	↓HD95
全监督	UNet [1]	95.5	5.3	96.0	3.2	83.7	2.1
DeepCut（2016年） [25]	-	37.1	15.2	36.2	14.7	-	-
SDI（2017年） [23]	VGG-16 [46]	49.2	11.7	-	-	38.7	9.0
多实例学习（2020年） [47]†	ENet [48]	69.4	9.4	72.3	8.7	-	-
GMIL（2021） [42]†	ENet [48]	71.1	8.8	71.7	5.9	58.2	4.7
BoxInst（2021） [7]‡	UNet [1]	47.1	11.3	48.4	11.6	37.9	8.9
我们的方法（不含几何信息）	UNet [1]	52.9	10.9	54.2	10.7	44.9	7.1
我们的方法（不含对比学习）	UNet [1]	69.7	9.4	68.5	9.1	57.3	4.9
Ours	UNet [1]	**79.8**	**8.7**	**80.2**	**5.3**	**65.4**	**4.2**

见图3。与作为我们框架基线的 BoxInst [7], 相比，我们很好地保留了候选区域的内部结构细节和外部形状。对比第4列和第7列可见，与图3中的基线相比，分割准确率有了显著提升。这是因为所提出的几何先验和对比相似度有助于更好地分割器官。此外，我们的结果质量远优于其他最先进方法，尤其是在内部细节方面。这些方法基于多实例学习（即，MIL、GMIL），主要关注候选区域的外部形状，而忽略了内部的空洞

器官。与他们的方法不同，我们从给定的模板器官中学习其内部结构和几何细节，同时利用外部形状来增强掩码头的训练。

D. 消融研究

在本小节中，我们首先对框架进行组件级分析，然后讨论所提出的几何先验和对比相似度的有效性。

1) 鲁棒性：在本小节中，我们主要讨论所提出的弱监督分割

表 II 鲁棒性分析。我们在 KiTS21 数据集上测量相应的 Dice 系数。我们基于其 2D 版本公开代码实现了 BoxInst 的 3D 版本 [7] 。

#	骨干网络	模型	不使用 我们的损失函数	With 我们的损失函数
1	ResNet [51]	BoxInst 3D [7]	47.8	**77.1**
2	UNet [1]	BoxInst 3D [7]	49.1	**80.2**
3	ResNet [51]	Ai+L [13]	67.2	**72.1**

框架的鲁棒性。我们的框架基于 BoxInst [7], 构建，由骨干网络和特征金字塔网络组成。我们分析骨干网络对弱监督分割性能的影响。如表 II 所示，我们的方法对其他骨干网络（即 ResNet [51] 和 U-Net [1]）具有鲁棒性。此外，我们通过将所提出的框架扩展到其他弱监督分割模型，验证了其通用性。遵循 Chu 等人 [13], 的做法，我们使用 Ai+L [13]模型，该模型采用 ResNet50 [51]骨干网络，广泛用于非医学 [52] 和医学 [53]分割任务。我们在解码器之后串联几何先验分支和对比相似性分支。如表 II 所示，基线模型在 KiTS 数据集上的得分为 67.2%。加入所提出的几何先验和对比相似性损失后，分割性能大幅提升了 4.9%。

2）几何先验：在所提出的几何先验中，我们最小化模板器官与点云嵌入空间中候选区域之间的倒角距离。点云表示比体积表示灵活得多，这是因为体积表示受到均匀体素网格的约束。每个网格的代表点固定在相应网格的中心。相比之下，点云表示不受此类约束。点的坐标更加精确，且具有更灵活的弹性（即，$\{0.17, 0.18, ...\}$ 与$\{1.00, 2.00, ...\}$）。与传统使用的体积表示相比，这有助于点云表示更好地描述精细且复杂的结构。因此，为了验证这一点，我们对采用体积表示的配准和优化进行了消融实验。具体而言，我们首先使用广泛应用的 SimpleITK [54] 对器官进行配准。配准设置如下：均方度量，随机采样策略的采样比例为 0.01，收缩因子为 $[4, 2, 1],$，平滑 sigma 为 $[2, 1, 0],$，迭代 100 次。配准后，我们使用 Dice损失来优化分割结果与几何先验之间的距离。如图 4 所示，点云表示能够处理更复杂的结构和更精确的几何形状，Dice 相似系数提升了 4.3%。

内部细节 我们分别从外部形状和内部结构的角度，进一步分析模板的影响。本小节的实验安排如下：我们首先补全模板的内部结构，仅保留外部形状。随后，使用变形后的模板训练掩码头。通常，器官内存在变异，即如图 5 所示，尤其是肾脏内部精细的结构。由于缺乏内部先验，分割性能显著下降（3.9%）。

完整性头分析 如第 III-A 节所述，遵循 nnUNet [2], ，对于输入图像，我们首先将其划分为图像块，并依次处理这些图像块。然而，在划分过程中，器官可能不完整，这可能导致无法从模板中学习。因此，为确保能够从模板中成功学习，所提出的完整性头会检测采样图像块中的候选区域是否完整。只有完整的候选区域才会被进一步处理。如表 III 所示，移除完整性头后，性能比我们的框架下降了 1.1%。

3）对比头的架构：本小节对所提出的对比头设计进行了实验。遵循以往的工作 [42], [45], ，对比头由一个两层逐点卷积组成，用于区分器官组织与非器官组织。为进一步分析对比头的结构，我们针对嵌入空间的维度进行了消融实验。如表 IV 所示，我们评估了不同嵌入维度下的性能和效率。综合考虑效率与性能之间的权衡，我们通过经验将嵌入维度设为 32。

对比相似度 我们提出了嵌入空间中的对比相似度。为了将嵌入空间与常用的灰度空间进行比较，我们选取了以下两种相似度指标：均方误差 和结构相似性 [55]。均方误差 和结构相似性 均基于灰度值计算，因为医学图像是灰度图像。如表 IV 所示，均方误差度量无法区分低对比度组织，而结构相似性

表 III 几何先验分析。†表示模板的内部结构。

#	消融设置	DSC 开启 KiTS
1	不含完整性头	79.1
2	不含内部细节†	76.3
3	基线	80.2

表 IV 对比头分析。 †: C 表示对比相似度，S 表示结构相似性相似度，G 表示灰度相似度。推理速度表示 1000 次运行的总时间。

#	组件设置	嵌入维度	嵌入空间†	推理速度	DSC（%）在 KiTS 上
1	对比头	8	C	3s	78.4
2		16	C	5s	79.1
3		32	C	8s	80.2
4		64	C	15s	80.4
5	相似度指标	32	C	8s	80.2
6		32	G	8s	73.4
7		32	S	12s	47.2

度量在分割性能和评估效率方面均不如对比相似度。

V. 结论

IN 在本研究中，我们提出了一种新颖的弱监督分割框架，使用边界框标注。我们引入几何先验和对比相似度，以应对复杂形状和成像伪影带来的挑战，从而增强弱监督分割框架的实用性和鲁棒性。几何先验能够促进精细复杂结构的学习，而对比相似度则有助于更好地区分器官像素与非器官像素。

我们的框架具有通用性，可轻松应用于多种弱监督分割模型，从而提升其性能。它有望成为未来弱监督分割研究的基础，尤其适用于具有特定形状器官的医学图像分割。我们通过大量实验验证了所提出的几何先验和对比相似度的有效性与优越性。

参考文献

[1] O. Ronneberger、P. Fischer 和 T. Brox，“U-net：用于生物医学图像分割的卷积网络”，载于医学图像计算与计算机辅助干预国际会议。Springer，2015 年，第 234–241 页。[2] 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”，*Nature Methods*，第 18 卷，第 2 期，第 203–211 页，2021 年。[3] D. Pathak、P. Krahenbuhl 和 T. Darrell，“用于弱监督分割的约束卷积神经网络”，载于论文集 *o f IEEE*国际计算机视觉会议，2015年，第1796–1804页。[4] Z. Jia、X. Huang、I. Eric、C. Chang和Y. Xu，《用于组织病理图像分割的约束深度弱监督》，*IEEE*医学影像汇刊，第36卷，第11期，第2376–2388页，2017年。[5] H. Kervadec、J. Dolz、M. Tang、E. Granger、Y. Boykov和I. B. Ayed，《用于弱监督分割的约束CNN损失》，*Medical Image Analysis*，第54卷，第88–99页，2019年。[6] M. Bateson、H. Kervadec、J. Dolz、H. Lombaert和I. B. Ayed，《用于分割的约束域适应》，载于医学图像计算与计算机辅助干预国际会议。Springer，2019年，第326–334页。[7] Z. Tian、C. Shen、X. Wang和H. Chen，《BoxInst：利用边界框标注实现高性能实例分割》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2021年，第5443–5452页。[8] T. Budrys、V. Veikutis、S. Lukosevicius、R. Gleizniene、E. Monastyreckiene和I. Kulakiene，《磁共振成像中的伪影：它们如何真正影响诊断图像质量并干扰临床诊断？》*Journal of Vibroengineering*，第20卷，2018年，第2期，第1202–1213页。[9] F. E. Boas、D. Fleischmann等，《CT伪影：成因与消减技术》，*Imaging Med*，第4卷，第2期，第229–240页，2012年。[10] S. N. Sarkar、D. B. Hackney、R. L. Greenman、B. A. Vachha、E. A. Johnson、S. Nagle和G. Moonis，《常规自旋回波与非对称自旋回波迭代分解在颈部软组织磁共振成像中的组织对比度和成像伪影：主观与客观比较》，*European Journal of Radiology*，第102卷，第202–207页，2018年。[11] H. Xie、H. Yao、S. Zhou、J. Mao、S. Zhang和W. Sun，《GrNet：用于稠密点云补全的网格残差网络》，载于欧洲计算机视觉会议。Springer，2020年，第365–381页。[12] K. Chaitanya、E. Erdil、N. Karani和E. Konukoglu，《在有限标注条件下为医学图像分割学习全局和局部特征的对比学习》，神经信息处理系统进展，第33卷，第12 546–12 558页，2020年。[13] T. Chu、X. Li、H. V. Vo、R. M. Summers和E. Sizikova，《利用多任务学习改进弱监督病灶分割》，载于医学影像与深度学习。PMLR，2021年，第60–73页。[14] D. Lin、J. Dai、J. Jia、K. He和J. Sun，《Scribblesup：用于语义分割的涂鸦监督卷积网络》，载于IEEE计算机视觉与模式识别会议论文集，2016年，第3159–3167页。[15] A. Bearman、O. Russakovsky、V. Ferrari和L. Fei-Fei，《点在哪里：点监督下的语义分割》，载于欧洲计算机视觉会议。Springer，2016年，第549–565页。[16] H. Qu等，《利用部分点标注对组织病理图像进行弱监督深度细胞核分割》，IEEE医学影像汇刊，第1–1页，2020年。[17] G. Patel和J. Dolz，《基于跨模态等变约束的弱监督分割》。Medical Image Analysis，第102374页，2022年。[18] G. Xu等，《CAMEL：用于组织病理图像分割的弱监督学习框架》，载于2019年IEEE/CVF国际计算机视觉会议（ICCV），2019年，第10 681–10 690页。[19] X. Wang等，《用于全切片肺癌图像分析的弱监督深度学习》，IEEE系统、人与控制论汇刊，2019年。[20] L. Grady，《图像分割的随机游走》，IEEE模式分析与机器智能汇刊，第28卷，2006年，第1768–1783页。[21] J. Dai、K. He和J. Sun，《Boxsup：利用边界框监督卷积网络进行语义分割》，载于IEEE国际计算机视觉会议论文集，2015年，第1635–1643页。[22] G. Papandreou、L.-C. Chen、K. P. Murphy和A. L. Yuille，《深度卷积网络的弱监督与半监督学习及其在语义图像分割中的应用》，载于IEEE国际计算机视觉会议论文集，2015年，第1742–1750页。[23] A. Khoreva、R. Benenson、J. Hosang、M. Hein和B. Schiele，《简单即有效：弱监督实例分割与语义分割》，载于IEEE计算机视觉与模式识别会议论文集，2017年，第876–885页。[24] M. Pu、Y. Huang、Q. Guan和Q. Zou，《Graphnet：为弱监督语义分割学习图像伪标注》，载于第26届ACM国际多媒体会议论文集，2018年，第483–491页。[25] M. Rajchl、M. C. Lee、O. Oktay、K. Kamnitsas、J. Passerat-Palmbach、W. Bai、M. Damodaram、M. A. Rutherford、J. V. Hajnal、B. Kainz等，《DeepCut：利用卷积神经网络从边界框标注中进行目标分割》，IEEE医学影像汇刊，第36卷，第2期，第674–683页，2016年。[26] B. Patenaude、S. M. Smith、D. N. Kennedy和M. Jenkinson，《用于皮质下脑分割的形状与外观贝叶斯模型》，神经影像，第56卷，第3期，第907–922页，2011年。[27] M. R. Sabuncu、B. T. Yeo、K. Van Leemput、B. Fischl和P. Golland，《基于标签融合的图像分割生成模型》，IEEE医学影像汇刊，第29卷，2010年，第10期，第1714–1729页。[28] B. Fischl、D. H. Salat、E. Busa、M. Albert、M. Dieterich、C. Haselgrove、A. Van Der Kouwe、R. Killiany、D. Kennedy、S. Klaveness等，《全脑分割：自动标注人脑神经解剖结构》，Neuron，第33卷，第3期，第341–355页，2002年。[29] J. E. Iglesias和M. R. Sabuncu，《生物医学图像的多图谱分割：综述》，Medical Image Analysis，第24卷，第1期，第205–219页，2015年。[30] M. Gao、Z. Xu、L. Lu、A. Wu、I. Nogues、R. M. Summers和D. J. Mollura，《利用深度卷积神经网络和稠密条件随机场进行分割标签传播》，载于2016年IEEE第13届生物医学成像国际研讨会（ISBI）。IEEE，2016年，第1265–1268页。[31] P.-A. Ganaye、M. Sdika和H. Benoit-Cattin，《语义约束下分割的半监督学习》，载于医学图像计算与计算机辅助干预国际会议。Springer，2018年，第595–602页。[32] A. BenTaieb和G. Hamarneh，《用于组织学腺体分割的拓扑感知全卷积网络》，载于医学图像计算与计算机辅助干预国际会议。Springer，2016年，第460–468页。[33] H. Chen、X. Qi、L. Yu、Q. Dou、J. Qin和P.-A. Heng，《DCAN：用于从组织学图像中进行目标实例分割的深度轮廓感知网络》

,”*Medical Image Analysis* ，第36卷，第135–146页，2017年。

[34] Y. Zhou, Z. Li, S. Bai, C. Wang, X. Chen, M. Han, E. Fishman, and A. L. Yuille, 《用于部分监督多器官分割的先验感知神经网络》，载于《IEEE/CVF国际计算机视觉会议论文集》，2019年，第10 672–10 681页。[35] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O'Regan等，《解剖约束神经网络（ACNN）：在心脏图像增强和分割中的应用》，《IEEE医学影像汇刊》，第37卷，第2期，第384–395页，2017年。[36] A. V. Dalca, J. Guttag, and M. R. Sabuncu, 《用于无监督生物医学分割的卷积网络中的解剖学先验》，载于《IEEE计算机视觉与模式识别会议论文集》，2018年，第9290–9299页。[37] S. Xie, J. Gu, D. Guo, C. R. Qi, L. Guibas, and O. Litany, 《Pointcontrast：面向三维点云理解的无监督预训练》，载于《欧洲计算机视觉会议》。Springer，2020年，第574–591页。[38] Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu, 《自我传播：探索无监督视觉表征学习中的像素级一致性》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2021年，第16 684–16 693页。[39] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, 《用于无监督视觉表征学习的动量对比》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2020年，第9729–9738页。[40] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, and L. Van Gool, 《通过对比对象掩码提议进行无监督语义分割》，载于《IEEE/CVF国际计算机视觉会议论文集》，2021年，第10 052–10 062页。[41] W. Wang, T. Zhou, F. Yu, J. Dai, E. Konukoglu, and L. Van Gool, 《探索用于语义分割的跨图像像素对比》，载于《IEEE/CVF国际计算机视觉会议论文集》，2021年，第7303–7313页。[42] X. Hu, D. Zeng, X. Xu, and Y. Shi, 《用于标签高效医学图像分割的半监督对比学习》，载于《医学图像计算与计算机辅助干预国际会议》。Springer，2021年，第481–490页。[43] Q.-Y. Zhou, J. Park, and V. Koltun, 《Open3D：用于三维数据处理的现代库》，arXiv:1801.09847，2018年。[44] E. Jang, S. Gu, and B. Poole, 《使用Gumbel-Softmax进行分类重参数化》，arXiv预印本 arXiv:1611.01144，2016年。[45] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, 《视觉表征对比学习的简单框架》，载于《国际机器学习会议》。PMLR，2020年，第1597–1607页。[46] K. Simonyan and A. Zisserman, 《用于大规模图像识别的超深卷积网络》，arXiv预印本 arXiv:1409.1556，2014年。[47] H. Kervadec, J. Dolz, S. Wang, E. Granger, and I. B. Ayed, 《用于弱监督分割的边界框：全局约束接近全监督》，arXiv预印本 ArXiv:2004.06816，2020年。[48] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, 《ENet：用于实时语义分割的深度神经网络架构》，arXiv预印本 arXiv:1606.02147，2016年。[49] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser等，《肝脏肿瘤分割基准（LiTS）》，arXiv预印本 arXiv:1901.04056，2019年。[50] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han等，《增强CT成像中肾脏及肾脏肿瘤分割的最新进展：KiTS19挑战赛结果》，Medical Image Analysis，第101821页，2020年。[51] K. He, X. Zhang, S. Ren, and J. Sun, 《用于图像识别的深度残差学习》，载于《IEEE计算机视觉与模式识别会议论文集》，2016年，第770–778页。[52] F. Lateef and Y. Ruichek, 《基于深度学习技术的语义分割综述》，《神经计算》，第338卷，第321–348页，2019年。[53] B. M. Anderson, E. Y. Lin, C. E. Cardenas, D. A. Gress, W. D. Erwin, B. C. Odisio, E. J. Koay, and K. K. Brock, 《使用全卷积网络自动勾画增强与非增强计算机断层扫描肝脏图像》，《放射肿瘤学进展》，第6卷，第1期，第100464页，2021年。[54] B. C. Lowekamp, D. T. Chen, L. Ib´a˜nez, and D. Blezek, 《SimpleITK的设计》，《神经信息学前沿》，第7卷，第45页，2013年。[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 《图像质量评价：从误差可见性到结构相似性》，《IEEE图像处理汇刊》，第13卷，第4期，第600–612页，2004年。

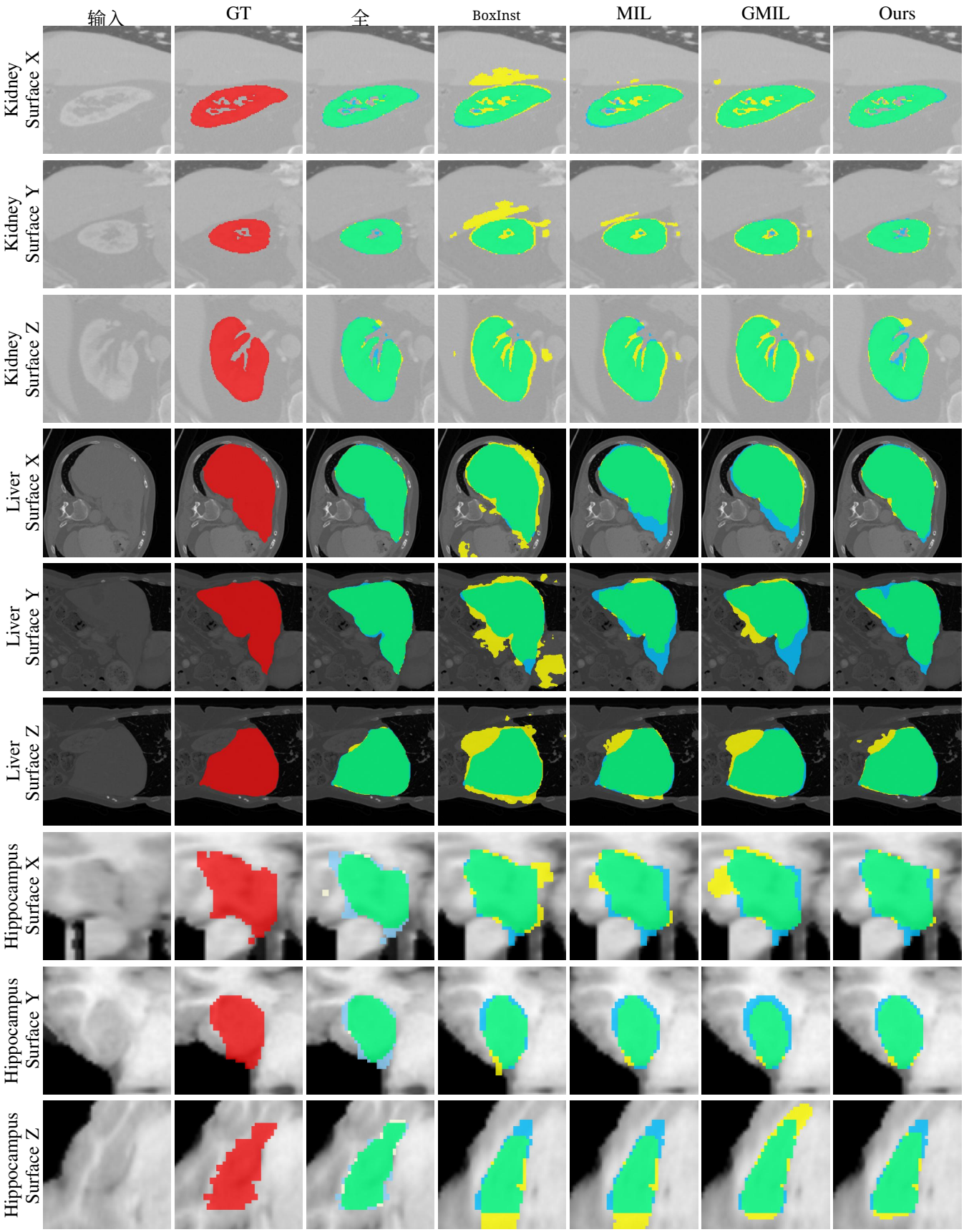


图3. 我们的框架在三个数据集上的定性结果。如图所示，分割结果中包含四种颜色。红色 表示真实标注。绿色 表示预测正确。蓝色 表示本应预测出来但未预测出的区域。黄色 表示预测错误的像素。我们的框架基于 BoxInst [7]，它是基线方法。随后，我们将多种方法进行比较，包括全监督结果和所提出的方法，并展示候选区域的高质量。

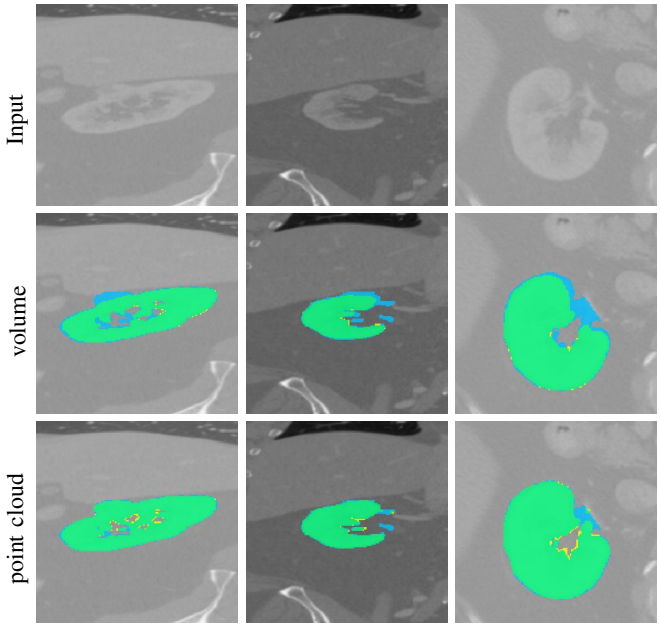


图 4. 体积与点云。我们通过 Dice损失，以体积表示的形式优化几何形状的学习。如图所示，所提出的点云表示显著改善了分割结果。

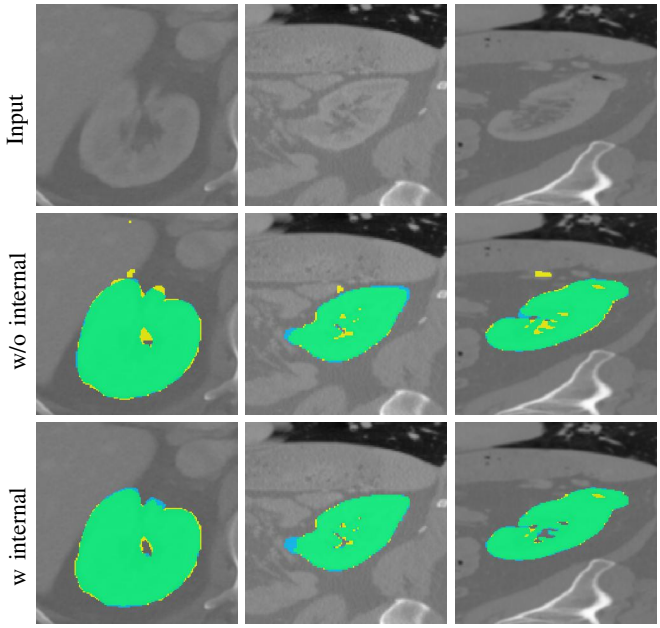


图 5. 内部细节。为验证模板内部细节的有效性，我们通过填充模板器官来消除器官内变异。如图所示，分割性能大幅下降，尤其是对于具有精细且复杂结构的器官。从左到右，图中器官的结构越来越精细、复杂。