

学习更好的配准以学习更好的少样本医学图像分割：真实性，多样性，鲁棒性

何宇霆, 葛荣骏, 戚晓明, 陈阳, 伍家松, Jean-Louis Coatrieux, 杨冠羽*, 李硕

摘要—在这项工作中, 我们致力于基于一种学习配准以学习分割 (learning registration to learn segmentation, LRLS) 范式, 提出新型的框架来解决少样本医学图像分割 (MIS) 的任务。为了解决现有 LRLS 框架中存在的真实性、多样性和鲁棒性不足的限制, 我们提出了优配强分 (Better Registration Better Segmentation, BRBS) 框架, 具有三个主要贡献, 经实验证明具有实际价值。1) 我们改善了基于配准的生成程序中的真实性, 提出了知识一致性约束策略, 该策略约束配准网络根据领域知识进行学习。它实现了语义对齐和拓扑保持的配准, 从而使生成程序以保持空间和样式真实性的情况下生成新数据。2) 我们深入研究了生成程序的多样性, 并提出了空间风格采样程序, 将样式和空间的转换路径建模到生成程序中, 引入了少量有标签图像和大量无标签图像之间的特征转换路径。进一步, 在转换路径上通过采样为生成的数据提供了更多样的空间和样式特征, 从而有效提高生成数据的多样性。3) 我们首次强调了 LRLS 范式中分割学习的稳健性, 并提出了混合不对齐正则化策略, 模拟了不对齐失真, 并约束网络减小对错误匹配区域的拟合程度。从而对这些区域的学习起到了正则化作用, 提高了分割学习的稳健性。我们的方法在两个具有挑战性的少样本医学图像分割任务中取得了新的最先进性能, 优于现有的基于 LRLS 的少样本分割方法。我们相信, 这种新颖而有效的框架将为医学图像领域提供一个强大的少样本分割基准, 并有效降低医学图像研究的成本。我们的所

这项工作部分得到了中国国家重点研发计划的支持, 项目编号为 2021ZD0113202; 部分得到了国家自然科学基金的支持, 项目编号为 61828101 和 62101249; 部分得到了江苏省自然科学基金的支持, 项目编号为 BK20210291; 部分得到了东南大学卓越项目资金的支持; 部分得到了 CAAI-Huawei MindSpore 开放基金的支持; 部分得到了东南大学研究生院科研基金的支持, 项目编号为 YBPY2139。(通讯作者: 杨冠羽)

何宇霆和戚晓明隶属于东南大学影像科学与技术实验室以及教育部计算机网络与信息集成重点实验室, 中国南京 210096。

葛荣骏隶属于南京航空航天大学计算机科学与技术学院, 中国南京 211106。

陈阳、伍家松和杨冠羽隶属于东南大学影像科学与技术实验室、教育部计算机网络与信息集成重点实验室、江苏省联合国际医学信息处理实验室以及中法生物医学信息研究中心 (CRIbs), 中国南京 210096 (电子邮件: yang.list@seu.edu.cn)。

Jean-Louis Coatrieux 隶属于江苏省国际联合医学信息处理实验室和中法生物医学信息研究中心 (CRIbs), 东南大学, 中国南京 210096。他还隶属于法国 Inserm、雷恩一大的 LTSI-UMR1099 实验室, 法国雷恩 35000。

李硕隶属于 Case Western Reserve University 的生物医学工程系和计算机与数据科学系, 美国俄亥俄州克利夫兰 44106 (电子邮件: shuo.li11@case.edu)。

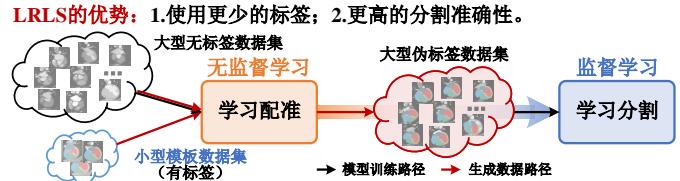


图 1: 学习配准以学习分割范式 (LRLS) 用于少标签医学图像分割。

有代码都将公开在线发布。

Index Terms—医学图像分割, 少样本学习, 图谱, 医学图像配准, 生成, 深度学习

I. 介绍

学习配准以学习分割范式 (Learning registration to learn segmentation, LRLS, 图1) [1]–[6] 是一种有效的少标签医学图像分割范式, 近来变得越来越受到研究领域的关注。该范式利用了人体结构的一致性, 利用配准的空间映射能力实现标签的映射, 因此, 任意满足解剖结构一致性的分割目标 (如心脏腔室、脑区域) 都能够通过配准来较为可靠地生成大量伪标签数据用于学习, 是一种通用的医学图像少标签学习范式, 已在大量具有挑战性的少标签医学图像分割场景 [2]–[4], [6] 中大大提升了模型的标签效率和分割准确性, 具有较大的应用和研究价值。如图1所示, 该范式有三个步骤: 1. 使用大量无标签数据学习一个配准网络。2. 使用配准网络将一个小型有标签数据集 (以下称为“模板集”) 与大型无标签数据集对齐, 从而生成大型伪标签数据集。3. 使用生成的伪标签数据集训练分割网络, 以获得具有强大的泛化能力的分割模型。这三个步骤合在一起构成了一个利用无监督的配准学习驱动需要监督的分割学习的工作流, 驱动了有效的少标签医学图像分割。然而, 如图2a-c) 所示, LRLS 范式的三个固有缺陷为基于 LRLS 的少样本医学图像分割模型带来了技术上的瓶颈: a) 生成数据缺乏真实性 [1], [2], [5]。在利用配准

S1: 学习配准

S2: 生成数据

S3: 学习分割

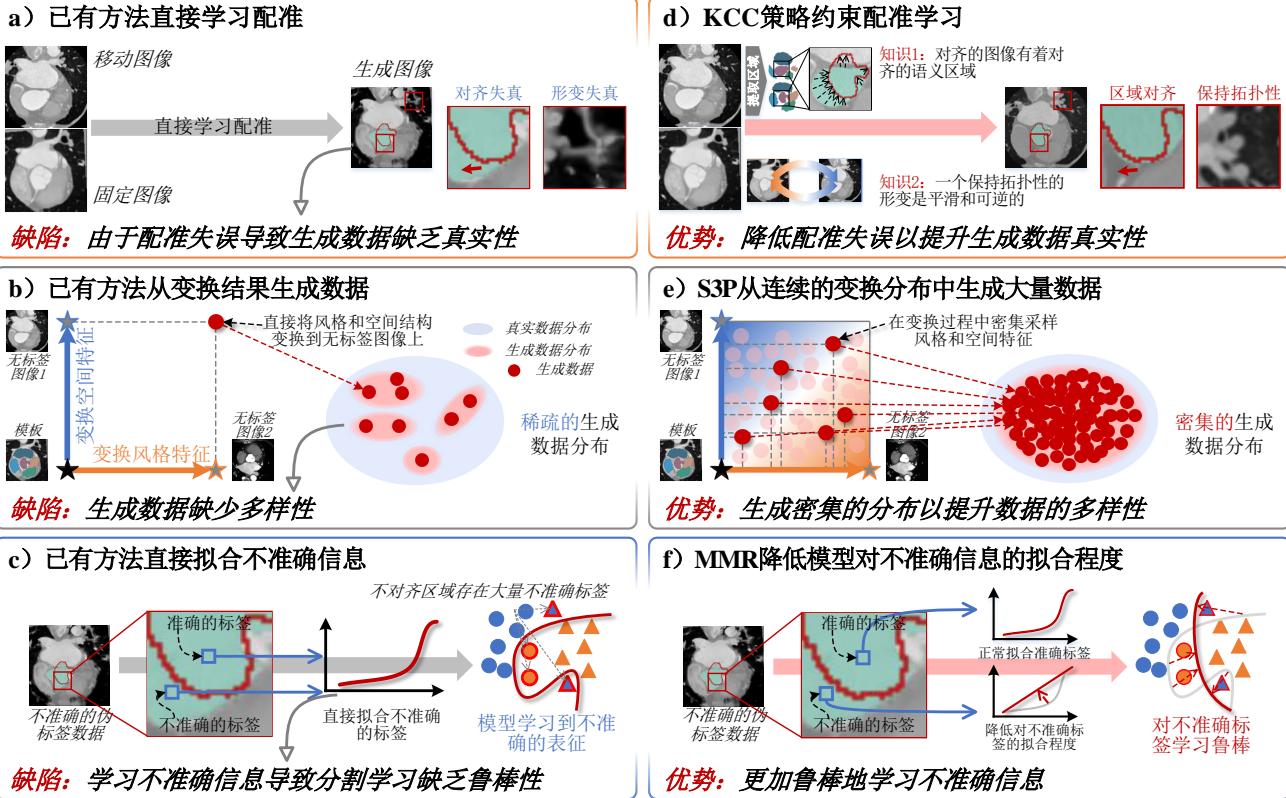


图 2: 现有的 LRLS 框架存在一定的局限性, 而本章所提出的 BRBS 框架突破了已有方法的瓶颈, 仅使用非常少标签即可取得强大的医学图像分割性能。a) 到 c) 是现有 LRLS 框架 (以 DataAug [2] 为例) 的局限性; d) 到 f) 是本章得到 BRBS 框架的设计动机和潜在优势。

生成数据的过程中, 有两种失真会使得伪标签数据缺乏真实性, 最终将误差传递给分割学习。对齐失真: [3] 伪标签和图像之间出现标签区域与图像上的真实区域间不对齐, 为分割学习带来错误的监督信息。形变失真: [7] 配准模型对区域发生的错误形变将破坏形变图像中真实的拓扑结构 [1], [5], 使得形变图像失真, 最终导致分割模型学习到不准确的图像特征, 限制分割模型性能。b) 生成数据缺乏多样性 [2], [4]。由于无标签图像数量的限制, 直接将模板集和无标签图像进行配准 [1] 或风格迁移 [2] 来生成数据, 只能将训练数据集扩大到与无标签数据集相同的数据规模。当无标签数据集非常小, 并且其数据特征分布较为稀疏时, 生成的伪标签数据也将较小, 限制了分割网络泛化潜力。c) 分割学习缺乏鲁棒性。现有的 LRLS 的相关工作 [1]–[4], [6] 专注于前两个步骤的研究, 无法在第三个步骤中抵抗来自配准误差的干扰。这些方法直接学习生成的缺乏真实性和多样性的伪标记数据集, 将使得分割模型学习到数据中不准确的特征和监督信息, 缺乏应对数据和标签噪声的鲁棒性。

为了突破 LRLS 范式的瓶颈, 使用非常少量的标签获得更加精确的医学图像分割模型, 我们提出了一种新的 LRLS 框架, 名为优配强分 (Better Registration Better Segmentation, BRBS) 框架。该框架融合了我们提出的三个关键创新点 (图 2 d-f), 解决了 LRLS 范式中缺乏真实性、多样性和鲁棒性的限制, 仅使用非常少的标签即可有效地生成高真实性和多样性的伪标签数据集, 并将对抗不准确数据中错误信息的能力融入分割模型, 提升分割学习的鲁棒性。

知识一致性约束策略 (Knowledge Consistency Constraint, KCC) 用于提高真实性: 如图 2d) 所示, KCC 根据两种领域知识构建一种更加有效的配准网络学习过程, 从而减少配准误差并提高生成数据的真实性。1. 应对对齐失真。对齐的图像具有对齐的语义区域 [4], [8], 因此我们构建了一种语义一致性 (SeC) 约束用于图像中语义区域间的对齐配准。它利用训练过程中的分割网络提取语义区域, 并约束对应的区域对齐以提高图像之间关键区域的对齐精度。2. 应对形变失真。保持拓扑结构不被破坏的形变是可逆且平滑的 [9], [10], 因

此我们构建了一种双向一致性 (BiC) 约束用于训练网络学习保持拓扑结构的配准。它训练网络同时学习正向和反向的平滑形变以获得形变可逆性和平滑性 [11]，从而保持形变图像中的拓扑结构不失真。

空间-风格采样过程 (Space-Style Sampling Process, S3P) 用于提高多样性：如图2e) 所示，S3P 构建了一种模板集与无标签数据集之间风格和空间结构信息密集采样的过程，在连续的流形上采样生成密集分布的数据，从而提高了生成数据的多样性。该过程通过建模模板和无标签图像之间的风格和空间相互转换的连续过程来构建图像间流形上的连续分布，并在该连续分布上密集采样生成数据，从而生成具有更好的多样性的数据。对于风格的多样性，该方法在模板和无标签图像之间建立一个条风格变换路径，并对此路径上对风格的偏移程度进行采样，从而调整图像原本的风格以生成具有新的风格特征的图像。对于空间的多样性，该方法在模板和无标签图像之间建立一条空间变换路径，并对此路径上的空间变形程度进行采样，从而调整图像原本的空间结构生成具有新的空间特征的图像。因此，同时对模板集中的数据进行密集的空间和风格的变换，可以生成大量的多样的生成数据，以提高多样性。

混合不对齐正则化 (Mix Misalignment Regularization, MMR) 用于提高鲁棒性：如图2f) 所示，MMR 通过模拟伪标签数据中标签与图像数据不对齐的情况，并训练网络减少这些数据中不对齐区域的拟合程度，从而降低不准确信息的影响，提高分割学习的鲁棒性。具体的，该方法将两个图像以一定比重混合以模拟风格采样时，模板图像与无标签图像中语义区域不对齐导致的图像失真，并以相同比重混合两个图像分别的分割结果以发现不对齐的语义区域。然后，训练分割网络对这些区域拟合一个线性函数（准确的区域拟合非线性函数），从而减少这些不准确信息的拟合程度。因此，分割网络将拟合一个更简单的函数，并更加鲁棒地应对生成图像中的不对齐问题 [12]，进一步提高分割模型性能。

总的来说，本章针对 LRLS 范式中的真实性、多样性和鲁棒性问题，设计了一种新型的 LRLS 框架，BRBS 框架，以激发少量标签信息对整体数据的监督能力，促进获得更好的少标签分割模型。具体如下：

- 基于 LRLS 范式，我们提出了一种高效的少标签医学图像分割框架，BRBS 框架，为医学图像领域提供了强有力的少标签学习基准，有效降低了获得有效医学图像分割模型的成本。

- 我们提出了一种基于领域知识驱动的配准学习策略，KCC，通过减少配准误差来提高生成图像的真实性，从而获得更符合真实世界图像生成形式的模拟生成过程。
- 我们提出了一种新颖的数据生成过程，S3P，在连续的图像风格和空间结构变换过程中密集采样生成数据，以提高生成数据的多样性。
- 我们提出了一种强大的数据驱动的正则化方法，MMR，通过模型对数据的表现来自我发现并降低不准确区域的拟合程度，提高分割网络学习过程中对不准确信息的鲁棒性。

总之，BRBS 框架有三个关键优势：a) 更小的标签需求：与全监督方法 [13]–[16] 相比，BRBS 框架几乎只需要一个标签就能够驱动整个训练过程，并且获得可靠的医学图像分割模型，有效地降低了该任务所需的标签成本。b) 更高的分割精度：与其他 LRLS 方法 [1]–[4], [6] 相比，BRBS 在真实性、多样性和鲁棒性方面的提升显著提高了分割模型的性能，在多个具有挑战性的医学图像分割任务上取得了最优的结果。c) 更大的简便性：在测试过程中，我们的框架仅使用单一分割网络就可以进行推理，没有任何额外的部分，因此相比于以往的工作 [4]，BRBS 具有更好的简便性。我们使用 PyTorch [17] 框架实现了所提出的方法。我们在一个伴随网站 <https://github.com/YutingHe-list/BRBS> 上共享了我们的代码和模型。

II. 相关工作

A. 少标签学习

少标签学习 (Few-shot learning) 是指从有限数量的带有监督信息的样本中学习的机器学习方法，能够有效提高模型学习的标签使用效率 [18]–[21]。近年来，在少标签学习已经在一些任务中取得了成功，如图像分类 [21]、视频分类 [19]、神经架构搜索 [20] 等。由于医学数据的标注成本高昂且数据稀缺 [22]，少标签学习在医学图像分析任务中有着较高的研究价值，并且在一些具体任务上已经取得了成功 [2]–[4], [6], [23]。然而，正如 Wang 等人 [18] 所指出的，少标签学习的核心问题是在学习中产生的不可靠的优化方向（我们称之为经验风险最小化方向）。当训练集中有标签数据非常少时，经验风险将与期望的风险相去甚远，无法为模型提供合理的优化方向，反而会使模型过拟合到错误的经验风险上，限制模型的性能。因此，在少标签的情况下进行可靠的

学习是一个具有挑战并十分重要的开放问题。尽管包括本文所关注的 LRLS 范式在内的现有工作已经提出了一些初步的方法以提升少标签情况下学习的可靠性 [1]–[6]，但该问题仍然是少标签学习的核心瓶颈。

B. 医学图像配准和分割

在过去的十年中，医学图像配准 [11] (medical image registration, MIR) 和医学图像分割 [24] 是医学图像分析中两个关键的密集预测任务。配准将医学图像中的解剖结构对齐到同一空间坐标系统中 [4], [11], [25], [26]，而分割则密集地预测每个像素位置所属的类别 [16], [24], [27]。随着深度学习 [28] 的发展，结合深度学习的配准和分割模型得到了广泛的关注并正被深入地研究，一些研究已经取得了显著的成果 [4], [11], [24], [29]。然而，随着研究的深入，一些重大的问题逐渐被暴露出来，导致了这类技术在应用中的巨大挑战。无监督的医学图像配准模型 [4], [11], [26] 缺乏对感兴趣区域的感知能力，导致其容易受到背景区域的影响，使得感兴趣区域的对齐精度受到限制。监督学习的医学图像分割模型 [16], [24] 需要大量的分割标签，而标注这些标签需要花费大量的成本。如果降低训练模型所使用的标签数量，深度学习强大的非线性能力又会引发严重的过拟合问题。

C. 基于模板的分割范式

基于模板的分割是一种被广泛研究的少样本医学图像分割范式 [5], [30]–[34]。这类方法的一个核心思想是利用图像-标签对中分割标签与图像之间的联系，通过有标签的图像与无标签的图像的配准，间接地构建标签到无标签图像的映射，从而实现无标签图像的分割 [33]。近年来，由于深度学习的强大特征表示能力和高运算速度，一些研究 [1]–[5], [30]–[32] 致力于将深度学习与基于模板的分割算法相结合，以较小的标注成本获得有效的分割结果。一种直观的方法 [5], [31] 是使用网络无监督地学习配准模型 [11]，以获得模板和无标签图像之间的快速配准。然而，如果模板与目标图像的空间差异较大，配准准确性的限制将使得分割效果较差，这些方法往往无法获有效的分割结果。虽然一些工作 [26] 使用分割标签来约束配准学习以获得更加精确的空间对齐，但这也额外为学习的过程引入了分割学习原有的标签成本问题，失去了该范式的少标签的优势。

D. 学习配准以学习分割范式

最近，基于深度学习的配准和分割模型之间的互补性研究 [3] 取得了一系列的成果 [1]–[4], [6]，该范式通过学习配准来生成大量数据以驱动分割学习 (LRLS)，取得了有效的少标签医学图像分割。该范式使用配准模型将少量模板图像和大量无标签图像对齐，从而生成大量的伪标签图像，利用这些伪标签图像来驱动分割模型的学习。DataAug [2] 基于该范式，并引入了一个风格转换操作，提高了生成数据的多样性。然而，不准确的配准导致生成数据失真，最终限制着分割模型的性能。Ding 等人 [6] 进一步增强了 DataAug 模型中生成数据的多样性，训练了一个自编码器从连续的潜在空间中采样生成伪标签图像。但是，该方法需要让自编码器学习模板图像固定的空间结构，因此只适用于一个模板图像的情况，限制了其可扩展性，同时，该方法中配准性能的缺陷仍然限制着生成数据的真实性。DeepRS [3], PC-Reg-RT [4] 和 DeepAtlas [1] 都利用了配准和分割任务之间的互补性，使用分割模型训练配准模型对语义区域的感知，从而使得目标区域获得更准确的配准性能，降低了配准失误导致的数据失真，从而获得更高质量的生成数据。然而，这些工作都没有考虑变形失真和生成数据的多样性问题，分割模型仍然在训练中受到部分失真问题的干扰，同时泛化能力受限。此外，上述的这些工作 [1]–[6] 都缺乏进一步关于鲁棒学习生成的伪标签数据的研究，分割模型的学习受到生成数据种种缺陷的干扰。

相关工作的讨论： LRLS 范式在少样本医学图像分割任务中的三个关键方面是真实性 [5], [7], [35]、多样性 [6], [20] 和鲁棒性 [36], [37]。更好的生成图像-标签对的真实性能将减少由于配准而引起的错误信息，使分割模型学习更可靠的表示。更好的生成数据的多样性将为分割网络提供更多知识，实现更好的泛化能力的有效表示。分割学习的更好的鲁棒性将提高对不准确数据的干扰能力，实现精确的特征表示和可靠的优化。因此，LRLS 范式需要提高其真实性、多样性和鲁棒性，学习更好的配准以实现更好的少样本医学图像分割。尽管一些工作在多样性 [2], [3], [6] 和真实性 [1], [3], [4] 方面进行了初步研究，但在 LRLS 范式的所有三个关键方面上还没有进行深入和全面的研究。在本文中，我们的 KCC、S3P 和 MMR 同时提高了真实性、多样性和鲁棒性，进一步提升了 LRLS 范式的性能。

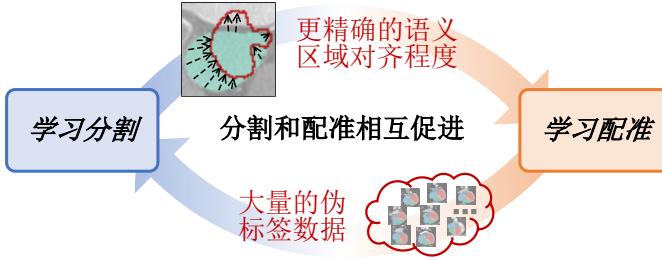


图 3: 分割 (S) 和配准 (R) 的互补学习过程。在整个训练过程中，配准和分割网络循环训练，互补优化，从而实现相互促进。

III. 方法

如图2所示，BRBS 框架学习语义对齐和保拓扑的配准，以获得更好的真实性 (KCC，第III-A节)，在变换路径上密集采样数据，以获得更好的多样性 (S3P，第III-B节)，并抑制不对齐区域的分割学习以提高对生成数据中不准确信息的鲁棒性 (MMR，第III-C节)，从而获得有效的少标签医学图像分割模型。BRBS 采用互补的分割和配准学习过程 [3] (图3)，通过交替迭代地训练配准和分割网络来实现相互促进的过程。

A. 知识一致性约束策略用于提升数据真实性

如图2d) 和图4所示，KCC 利用两种领域知识来约束配准学习，使模型生成符合领域知识的伪标签数据，大大提高了生成数据的真实性。我们将从训练数据集 \mathbb{D}_{train} 中采样的图像表示为 M 和 F。该策略有两个关键组件：

1) A) 双向一致性 (BiC) 策略：BiC 策略约束配准网络学习保持拓扑结构的配准映射，从而提高生成图像中空间结构的真实性。该方法约束配准过程的可逆性和平滑性，以获得保持拓扑结构不变的映射。对于可逆性，它不仅训练模型学习正向形变，即通过正向形变场 ϕ 将图像 M 形变到图像 F 的空间结构 [11]，还训练网络学习反向形变，即通过反向形变场 ϕ^{-1} 将图像 F 形变到图像 M 的空间结构：

$$\phi = \mathbb{R}(M, F), \phi^{-1} = \mathbb{R}(F, M). \quad (1)$$

正向和反向形变场还通过反向损失 \mathcal{L}^I 进行相互约束：

$$\begin{aligned} \mathcal{L}^I(\phi, \phi^{-1}) &= \sum_{p \in \phi} \|\phi - \mathbb{I}(\phi^{-1})\|^2, \\ \mathbb{I}(\phi^{-1}) &= -1 * \phi^{-1}(\phi^{-1}), \end{aligned} \quad (2)$$

其中， p 是 ϕ 中的体素位置， $\phi^{-1}(\phi^{-1})$ 表示使用空间变换操作 [11] 将 ϕ^{-1} (括号内) 按照位移向量场 ϕ^{-1} (括号外) 形变到目标空间位置， \mathbb{I} 表示将位移向量场取

逆变换。因此，该损失有两个步骤：1) $\mathbb{I}(\phi^{-1})$ 首先将逆变形场 ϕ^{-1} 与正向变形场 ϕ 的空间对齐并取负，因此逆变形场中的向量与其对应的正向变形场中的向量具有相同的方向和空间位置。2) 使用 $\mathcal{L}^I(\phi, \phi^{-1})$ 来计算对齐的逆变形场和正向变形场之间的差异，以优化网络学习可逆的形变，从而提高保持图像拓扑能力。我们同时计算形变图像 M^ϕ 与固定图像 F 对之间的相似度 \mathcal{L}^{sim} ，以及形变图像 $F^{\phi^{-1}}$ 与固定图像 M 对之间的相似度。因此，配准模型将学习道更好的可逆性，获得更加真实的形变。我们采用广泛使用的归一化互相关损失 [11] 作为 \mathcal{L}^{sim} ：

$$\mathcal{L}^{sim}(F, M^\phi) = \frac{\left(\sum_{p_i} (F(p_i) - \hat{F}(p_i))(M^\phi(p_i) - \hat{M}^\phi(p_i))^2 \right)}{\left(\sum_{p_i} (F(p_i) - \hat{F}(p_i))^2 \right) \left(\sum_{p_i} (M^\phi(p_i) - \hat{M}^\phi(p_i))^2 \right)}, \quad (3)$$

其中， Ω 是图像空间， $\hat{F}(p)$ ($\hat{M}^\phi(p)$) 是局部均值强度图像： $\hat{F}(p) = \frac{1}{n^3} \sum_{p_i} F(p_i)$ 。 p_i ，即在 p 周围的一个 n^3 的体积范围中求平均。为了学习平滑性，该组件还采用平滑性损失 [11] \mathcal{L}^{smooth} 同时约束位移向量场的相邻向量梯度较小，削弱局部夸张的形变。

$$\mathcal{L}^{smooth}(\phi) = \sum_{p \in \phi} \|\nabla \phi(p)\|^2, \quad (4)$$

因此，BiC 的整体损失 \mathcal{L}^{BiC} 是：

$$\begin{aligned} \mathcal{L}^{BiC}(F, M, \phi, \phi^{-1}) &= \\ \mathcal{L}^{sim}(F, \phi(M)) + \mathcal{L}^{sim}(M, \phi^{-1}(F)) + \\ \lambda^{smooth} (\mathcal{L}^{smooth}(\phi) + \mathcal{L}^{smooth}(\phi^{-1})) + \lambda^I \mathcal{L}_I(\phi, \phi^{-1}), \end{aligned} \quad (5)$$

其中， λ^{smooth} 和 λ^I 是 \mathcal{L}^{BiC} 中损失的权重。遵从论文 [11] 和论文 [38]，我们设置 $\lambda^{smooth} = 1$ 和 $\lambda^I = 0.1$ 以获得更好的拓扑保持能力。结合语义一致性策略 (SeC，下文)，BiC 将获得比论文 [11] 具有更好的拓扑保持能力，且降低了论文 [38] 中由于过度平滑所导致的配准效果差的风险。

2) B) 语义一致性 (SeC) 策略：SeC 策略约束配准网络学习感知两幅图像中语义区域并将其对齐的能力，提高了生成图像过程中进行样式偏移时的语义区域的真实性。它通过分割网络 (\mathcal{S}) 对两幅图像 (M、F) 进行分割来提取语义区域，并通过 SeC 损失 \mathcal{L}^{SeC} 约束配准网络将对应的语义区域对齐。在这个过程中，我们采用 DSC 损失 [1] 作为度量语义区域对齐程度的损

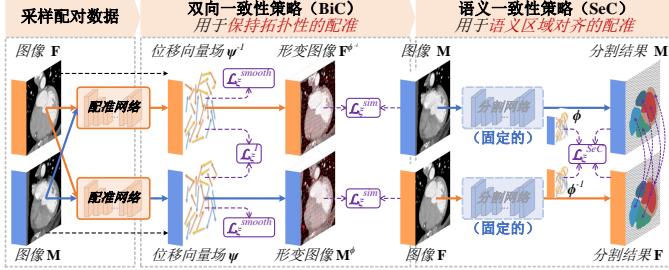


图 4: KCC 通过 BiC 和 SeC 策略来训练拥有保持拓扑性和语义区域对齐能力的配准模型, 从而生成符合领域知识伪标签数据生成过程, 提高生成数据的真实性。

失函数:

$$\begin{aligned} \mathcal{L}^{SeC}(F, M, \phi, \phi^{-1}) = \\ \mathcal{L}^{DSC}(\mathbb{S}(F), \phi(\mathbb{S}(M))) + \mathcal{L}^{DSC}(\mathbb{S}(M), \phi^{-1}(\mathbb{S}(F))). \end{aligned} \quad (6)$$

因此, 该损失函数将优化配准网络输出具有良好语义区域对齐能力的形变, 使得在生成图像时的风格偏移获得更好的真实性。

总的来说, 学习配准的过程可以被定义为最小化一个损失函数 $\mathcal{L}^{\mathbb{R}}$:

$$\hat{\mathcal{R}} = \arg \min_{\mathbb{R}} [\mathbb{E}_{(F, M) \sim \mathbb{D}_{train}} [\mathcal{L}^{\mathbb{R}}(F, M, \mathbb{R}(F, M), \mathbb{R}(M, F))]] \quad (7)$$

其中 F 和 M 是从我们的训练数据集 \mathbb{D}_{train} 中采样的图像。完整的损失函数为:

$$\begin{aligned} \mathcal{L}^{\mathbb{R}}(F, M, \mathbb{R}(F, M), \mathbb{R}(M, F)) = \\ \mathcal{L}^{BiC}(F, M, \phi, \phi^{-1}) + \lambda^{SeC} \mathcal{L}^{SeC}(F, M, \phi, \phi^{-1}), \end{aligned} \quad (8)$$

其中, λ^{SeC} 是 \mathcal{L}^{SeC} 的权重。

B. 空间-风格采样程序用于提升多样性

如图5, S3P 构建了一种在基于配准的生成程序中密集采样的策略, 从而理由少量有标签数据生成具有良好多样性的伪标签数据集。S3P 遵循论文 [2], 同时考虑医学图像空间结构和风格特征两个方面来生成数据。我们构建了一个拥有 M 个不同空间结构和风格特征无标签图像的无标签数据集 U , 以及一个拥有 N ($N, N \ll M$) 个有标签图像的模板数据集 A 。因此, 训练数据集 \mathbb{D}_{train} 可以描述为:

$$\mathbb{D}_{train} = \{\{(A, y)\}^N, \{U\}^M\} (N \ll M), \quad (9)$$

S3P 从无标签数据集 U 中随机采样两个拥有新的空间和风格特征的无标签图像 U_j, U_k , 然后从模板数

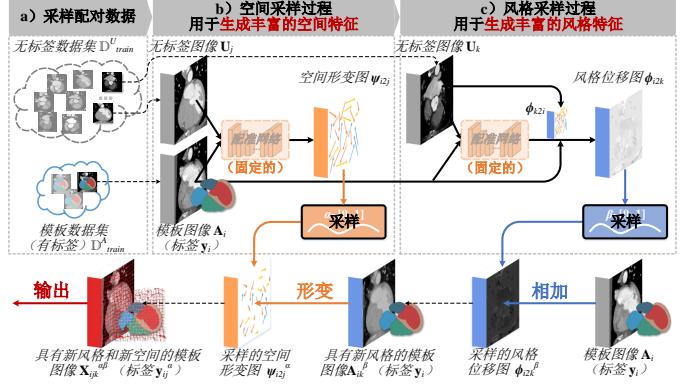


图 5: S3P 构建了一个基于配准概率生成程序, 通过对流形上图像间空间和风格路径进行采样, 生成更加丰富的空间和风格特征, 提高了生成数据的多样性。

据集中采样一个拥有基本的空间和风格特征的有标签图像 (A_i, y_i) : $\{(A_i, y_i), U_j, U_k\} \sim \mathbb{D}_{train}$ 。

对于生成更加多样的空间特征, S3P 将一个无标签图像 U_j 和一个模板图像 A_i 输入到配准网络 \mathbb{R} 中, 得到一个形变场 ϕ_{ij} (公式10), 该形变场能够将模板图像的空间特征转化为无标签图像 U_j 的空间特征。接着, S3P 生成一个随机值 $\alpha \in [0, 1]$ 对形变程度进行采样, 获得空间形变图, 从而实现从模板图像空间到无标签图像空间之间流型上的采样, 获得更加多样的空间信息:

$$\phi_{ij} = \mathbb{R}(A_i, U_j), \phi_{ij}^\alpha = \alpha \phi_{ij}. \quad (10)$$

对于生成更加多样的风格特征, S3P 将另一个无标签图像 U_k 通过配准网络对齐到模板图像 A_i 上, 得到形变后的无标签图像 $U_{ki} = \mathbb{R}(U_k, A_i)(U_k)$ 。然后, 通过对齐的两个图像 (U_{ki}, A_i) 之间进行像素值相减 (公式11), 获得模板图像的风格位移图 ψ_{ki} , 从而实现了从模板图像到无标签图像 U_k 的风格特征转换。然后, S3P 生成一个随机值 $\beta \in [0, 1]$ 对风格变换程度进行采样, 获得风格位移图, 从而实现从模板图像风格到无标签图像分风格之间流形上的采样, 获得更加多样的风格信息:

$$\psi_{ki} = U_{ki} - A_i, \psi_{ki}^\beta = \beta \psi_{ki}. \quad (11)$$

最后, 采样获得的风格位移图 ψ_{ki}^β 被添加到模板图像中生成新的风格特征, 采样获得的空间形变图 ϕ_{ij}^α 对具有新风格的模板图像和模板标签进行形变, 生成同时具有新风格和空间特征的伪标签图像 $(X_{ijk}^{\alpha\beta}, y_{ij}^\alpha)$:

$$X_{ijk}^{\alpha\beta} = \phi_{ij}^\alpha(A_i + \psi_{ki}^\beta), y_{ij}^\alpha = \phi_{ij}^\alpha(y_i). \quad (12)$$

如果 $\beta = 0$, 则生成的图像是仅空间形变后的模板图像 A_{ij}^α ; 如果 $\alpha = 0$, 则生成的图像是仅风格变换后的模板

表 I: 不同典型 LRLS 方法的概率生成程序的概率分布表明, BRBS 的生成过程具有更高的随机性, 将产生更加多样化的生成图像。

方法	概率生成程序
DeepAtlas [1]	$p((A_i, y_i))p(U_j)$
PC-Reg-RT [4]	$p(\phi_{i2j} \mathbb{R}, A_i, U_j)p((A_i, y_i))p(U_j)$
DataAug [2]	$p((A_i, y_i))p(U_j)p(U_k)$
DeepRS [3]	$p(\alpha)p(\phi_{i2j} \mathbb{R}, A_i, U_j)p((A_i, y_i))p(U_j)$
Our BRBS	$p(\beta)p(\psi_{i2k} \mathbb{R}, A_i, U_k)p(\alpha)p(\phi_{i2j} \mathbb{R}, A_i, U_j)p((A_i, y_i))p(U_j)p(U_k)$

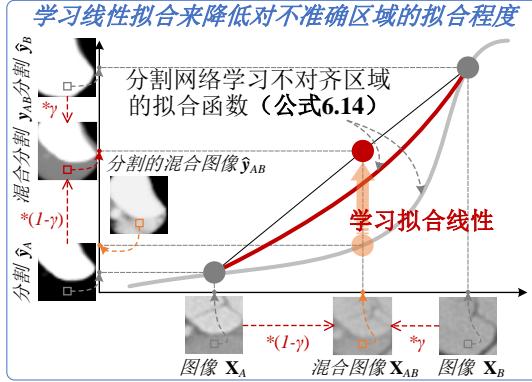


图 6: MMR 提高了分割网络训练的鲁棒性。从优化的角度来看, 它将两个图像混合以模拟不对齐失真, 并训练深度学习模型在不对齐区域学习线性函数, 从而减少对这些不准确信息的拟合。

图像 A_{ik}^β 。因此, 通过在风格和空间结构变换的流形上对数据进行采样, S3C 获得了更密集的生成数据分布, 大大提高了生成数据的多样性。

总的来说, 如表I所示, 与其他典型的 LRLS 方法相比 [1]–[4], BRBS 具有包括图像级、空间级和风格级的三个级别采样, 拥有更多样的数据生成能力。它使用三种随机性, 包括图像间空间和风格配对的随机性 (DataAug, $p((A_i, y_i))p(U_j)p(U_k)$), 配准网络训练过程的随机性 (DeepRS, PC-Reg-RT, $p(\phi_{i2j}|\mathbb{R}, A_i, U_j)$), 以及风格 ($p(\beta)$) 和空间 (DeepRS, $p(\alpha)$) 变换过程中采样的随机性, 这三重的随机性进一步提高了生成的多样性, 为分割模型的学习提供了多样化的生成数据。

C. 混合不对齐正则化提升鲁棒性

如图7所示, MMR 策略在使用生成的伪标签数据 $(X_{ijk}^{\alpha\beta}, y_{ij}^\alpha)$ 训练分割网络的同时限制不对齐区域的拟合程度, 从而削弱数据中不准确信息的干扰, 提高了分割学习的鲁棒性。

具体的, MMR 策略在不需要标签的情况下, 通过加权混合两个图像 X_A 和 X_B , 生成一种混合图像 X_{AB} 来模拟图像不对齐的风格失真情况, 并训练分割网络在不对齐区域内拟合线性函数, 对齐的区域内拟合非线性

函数, 从而降低模型对这些不准确信息的拟合度 (图7)。该策略首先从训练数据集 \mathbb{D}_{train} 中采样两个图像 X_A 和 X_B , 然后使用随机值 $\gamma \in [0, 1]$ 加权求和这两个图像从而生成混合图像 X_{AB} 以模拟生成数据中由于配准不对齐导致的风格失真。接着, 分割模型 S 对两个原始图像 X_A, X_B 和混合图像 X_{AB} 分别进行推理预测, 获得它们的分割结果 $\hat{y}_A = S(X_A)$, $\hat{y}_B = S(X_B)$ 和分割混合图 $\hat{y}_{AB} = S(X_{AB})$ 。然后, 分割结果图 \hat{y}_A, \hat{y}_B 也通过随机值 γ 加权求和, 从而获得对齐区域响应高、不对齐区域响应低的混合分割结果图 y_{AB} 。最后, 通过混合损失 \mathcal{L}_{Mix} , 将混合图像的分割结果图 \hat{y}_{AB} 与混合分割结果图 y_{AB} 约束抑制, 从而提高分割网络应对生成数据失真的鲁棒性。

$$\mathcal{L}_{Mix}(\hat{y}_A, \hat{y}_B, \hat{y}_{AB}) = \parallel \gamma \hat{y}_A + (1 - \gamma) \hat{y}_B - \hat{y}_{AB} \parallel^2. \quad (13)$$

如图6所示, 从优化过程的角度看, MMR 约束分割网络对不对齐的区域 (即失真的区域) 拟合更具线性的函数, 从而减少拟合不准确信息的风险。两张图像和它们的分割结果图 $(X_A, \hat{y}_A), (X_B, \hat{y}_B)$ 中不对齐区域是图中分割网络函数 S 上的两个位置点 (灰色)。通过加权求和混合这两张图像和它们的分割结果图得到的混合图像和混合分割结果图是两个灰点连接线上采样的点 (红点 (X_{AB}, y_{AB}))。由于两张图像的不对齐区域属于不同的类别, 混合后, 这些区域的响应会比对齐区域更低。因此, MMR 约束混合图像的分割结果图 \hat{y}_{AB} 和混合分割结果图 y_{AB} 保持一致, 能够使得分割网络函数在这些不对齐的区域拟合到更线性的位置上, 从而降低过拟合到错误信息的风险。该过程为:

$$S(\gamma X_A + (1 - \gamma) X_B) \Leftrightarrow \gamma S(X_A) + (1 - \gamma) S(X_B). \quad (14)$$

在实验中, 随机值 γ 从一个 Beta 分布 $Beta(a, b)$ 中采样 (a 和 b 均为 0.3), 如果混合的区域是对齐的, 它们的分割结果图将具有相同的值, 因此混合分割结果图中对齐区域的值将接近于它们的原始值, 不会影响这些对齐区域的拟合。因此, 分割网络最终将针对性地降低失真区域的拟合程度, 从而更加鲁棒地训练分割网络。

除了 MMR 的学习外, 分割网络还使用分割损失 \mathcal{L}^{seg} 学习生成的伪标签数据 $(X_{ijk}^{\alpha\beta}, y_{ij}^\alpha)$ 。如图7所示, 分割损失 \mathcal{L}^{seg} 计算生成图像的分割结果 \hat{y}_{ij}^α 与生成标签 y_{ij}^α 之间的距离以训练分割模型。在本文中, 我们采用 DSC 损失 [1] 作为分割损失。

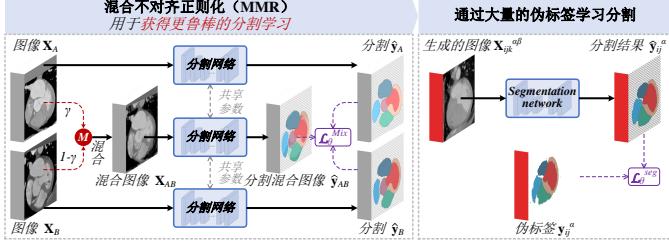


图 7: 分割网络的整体学习过程。它通过 MMR 的 Mix 损失 \mathcal{L}^{Mix} 和分割损失 \mathcal{L}^{seg} 来学习医学图像分割。

总的来说，分割学习过程可以被定义为最小化一个损失函数 \mathcal{L}^S :

$$\hat{\mathcal{S}} = \arg \min_{\mathcal{S}} [\mathbb{E}_{((A_i, y_i), U_j, U_k, X_A, X_B) \sim \mathbb{D}_{train}} [\mathcal{L}^S(\hat{y}_{ij}^\alpha, y_{ij}^\alpha, \hat{y}_{AB}, \hat{y}_A, \hat{y}_B)]]. \quad (15)$$

分割网络的完整损失为:

$$\begin{aligned} \mathcal{L}^S(\hat{y}_{ij}, y_{ij}, \hat{y}_{AB}, \hat{y}_A, \hat{y}_B) &= \omega^{Mix} \mathcal{L}^{Mix}(\hat{y}_A, \hat{y}_B, \hat{y}_{AB}) \\ &\quad + \mathcal{L}^{seg}(\hat{y}_{ij}^\alpha, y_{ij}^\alpha) \end{aligned} \quad (16)$$

其中， ω^{Mix} 是完整的分割网络损失中混合损失的权重。

IV. 实验设置

1) 数据集: 我们在两个不同成像模态和成像特点的任务上评估 BRBS 框架的少标签医学图像分割性能:
a) 少标签心脏结构分割 [39]: 该任务在 CT 图像上评估了 BRBS 框架对七个大的心脏结构的分割性能，总共包含三个公共的心脏 CTA 数据集，包括 MM-WHS 挑战赛 [39] 的 CT 图像数据（20 张带有心脏结构标签的图像和 40 张无标签图像），ASOCA 挑战赛 [40] 数据集（60 张无心脏标签图像），CAT08 挑战赛 [41] 数据集（32 张带有心脏结构标签的图像，标签来自该连接¹）。因此，该任务总共拥有 52 个有标签图像和 100 个无标签的图像。预处理过程中，我们裁剪心脏区域并将图像重采样为 $144 \times 144 \times 128$ 。我们同时进行了 5 个标签情况和 1 个标签情况的验证，对于 5 个标签（5-shot）的评估，我们随机选择五个有标签图像作为模板集，并将其余 47 个有标签图像作为测试集。对于 1 个标签（1-shot）的评估，我们随机选择一个有标签图像作为模板集，其余 51 个有标签图像作为测试集。剩下的 100 个无标签图像作为训练数据中的无标签数据集。
b) 少标签脑组织分割 [42]: 该任务在 T1 MR 图

像上评估 BRBS 框架对 28 个小的脑组织的分割性能，使用了 CANDI 数据集 [43]（包含 103 个带有脑组织标签的 T1 脑 MR 图像）。遵从论文 [6] 的实验设置，我们从原始图像中心裁剪出一个 $160 \times 160 \times 128$ 的体积，并将 20、82、1 个图像作为 1-shot 的评估中的测试、无标签和模板数据集。对于 5-shot 的评估，我们将 20、78 和 5 个图像作为测试、无标签和模板数据集。我们对所有训练过程执行 $[-20^\circ, 20^\circ]$ 范围内的旋转和 $[0.75, 1.25]$ 倍的缩放作为数据增强。

2) 对比设置: 我们将 BRBS 框架与 16 种广泛使用的框架在 1-shot 和 5-shot 情况下进行比较以展示 BRBS 框架的卓越性，共分为四种不同的类型对比: a) 对比直接学习分割 (learning segmentation, LS) 的框架 (3D U-Net (2016) [13], SegNet (2017) [14], U-Net++ (2019) [15], DBN (2020) [16]) 以展示少标签情况下监督学习的医学图像分割模型的基本性能。b) 对比半监督学习分割 (semi-supervised learning segmentation, SLS) 框架 (UA-MT (2019) [44], MASSL (2019) [45], DPA-DBN (2020) [16], CPS (2021) [46]) 以显示在少标签情况下由于缺乏领域知识而导致的半监督方法的局限性。c) 对比基于模板分割 (Atlas-based segmentation, ABS) 框架 (VoxelMorph (VM, 2018) [11], 标签约束的 VM (LC-VM, 2019) [26], LT-Net (2020) [5]) 以显示由于模板和目标图像之间的空间差异而导致的分割性能限制。d) 对比基于 LRLS 范式的方法 (DeepAtlas (2019) [1], DataAug (2019) [2], DeepRS (2020) [3], PC-Reg-RT (2021) [4], VAEAug (2021) [6]) 进以展示 BRBS 卓越的性能。除了 LS 方法外，我们对所有方法使用相同的骨干网络，即 3D U-Net [13]，以避免网络架构的不同对性能造成干扰而破坏公平性。

3) 实施细节和评价指标: 我们基于 PyTorch [17] 实施 BRBS 框架，在 24GB 显存的 NVIDIA GeForce RTX 3090 GPU 上训练和测试。我们将 α , β 和 γ 的值设为从 Beta 分布 $Beta(a, b)$ 中采样 (a 和 b 均为 0.3)，从而使得概率分布更接近于 0 和 1，降低失真风险。遵从论文 [26] 和论文 [38]，我们设置 $\lambda^{smooth} = 1$, $\lambda^I = 0.1$, $\lambda_{SeC} = 100$ (参照 V 节) 以获得语义区域对齐和保留拓扑的有效配准。我们进一步设置 $\omega_{Mix} = 0.1$ (参照 V 节) 以平衡分割网络的训练。训练过程采用 Adam 优化器 [47] 进行优化，学习率设置为 1×10^{-4} ，批大小为 1，以实现快速收敛 [4]。我们通过 AntsPy²对所有图像执

¹<http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/>

²<https://github.com/ANTsX/ANTsPy>

行仿射变换以统一图像的空间位置。

4) 评价指标: 本章实验从两个方面评估模型的分割性能 [48]: a) 基于区域的评估指标: 使用 Dice 系数 (DSC) 来评估分割结果与标签之间的重叠程度。DSC 越高, 区域的重合度就越好。b) 基于距离的评估指标: 使用平均 Hausdorff 距离 (AVD) 来评估表面的重合度。AVD 越低说明分割结果与标签之间表面重合度越好。对于实验分析部分对配准性能评估, 我们除了 DSC 和 AVD, 还计算每个变形体素位置上位移向量场的雅可比矩阵 $J_\phi < 0$ 的位置站全部位置的百分比 $J_\phi < 0$, 从而评估形变的失真程度 (越低越好)。

V. 实验结果和分析

A. 定量对比分析

如表II所示, 基于 LRLS 的方法在少样本医学图像分割任务中表现出极大的优越性。我们可以发现三个有趣的观察结果: 1) 尽管 SLS 方法使用了更多的无标签数据, 但是这类方法的性能竟然比只是用少量有标签数据的 LS 的方法表现更差。这是因为极少的标签只能够提供了很少的监督信息, 因此大量无标签数据在半监督学习中为模型提供了极不可靠的监督信息, 导致了 UA-MT 在 5-shot 的 (a) 任务中与 3D U-Net 相比 DSC 下降了 17.9%。虽然 CPS 在 5-shot 的 (a) 任务中取得了 3.1% 的 DSC 提高, 但在结构非常小的脑组织分割任务 (b) 中几乎没有分割能力 (25.3% 和 37.1% 的 DSC)。2) ABS 方法由于医学图像间基础解剖结构的一致性, 对于小结构具有较为稳定的性能, 因此 VM 在 5-shot (b) 任务中获得了第三高的 83.1% 的 DSC。然而, 这类方法受限于模板图像与测试图像之间的相似性, 有着较大的瓶颈, 因此 VM 在 5-shot (a) 任务中仅有 81.0% 的 DSC。3) 基于 LRLS 范式的方法由于其深度学习的强大表征能力和生成的大量训练数据的能力, 获得了全面高于 LS 和 SLS 的性能。其在两个任务上的 5-shot 设置中都取得了 80% 以上和 70% 以上的 DSC。

与其他 LRLS 方法相比, BRBS 在两个方面展示出强大的少标签分割优势: 1) BRBS 同时在大结构 (任务 (a)) 和小结构 (任务 (b)) 分割上都有着强大的能力。在 5-shot 情况下, 由于 BRBS 在生成图像的多样性和真实性方面的优势, 它在两个任务上都取得了最高的 DSC (91.1%, 87.2%) 和最低的 AVD (0.93mm, 0.43mm)。虽然 PC-Reg-RT 在大结构分割 (任务 (a))

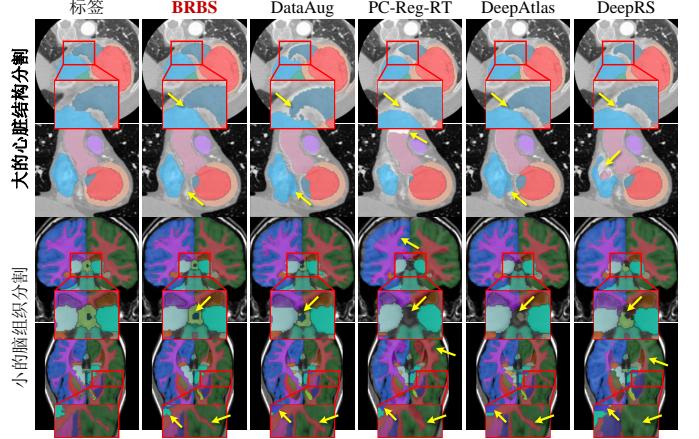


图 8: 定性评估显示, 与四种典型的 LRLS 方法相比, BRBS 在 5-shot 设置下的两个少标签医学图像分割任务上都具有更好的视觉效果。

上也取得了类似的性能, 但在小结构分割 (任务 (b)) 上表现得非常差, 比 BRBS 低了超过 10% 的 DSC。这是因为它在小结构上的严重误分割极大限制了配准性能, 从而导致生成数据中出现严重的对齐失真, 反过来限制了分割的性能。DataAug 在小结构分割 (任务 (b)) 上表现出有竞争力的性能, 其 DSC 只比我们低 4.4% 和 3.3%, 但该方法由于缺乏对语义区域对齐的约束, 在大结构分割 (任务 (a)) 上极为受限。2) BRBS 在 1-shot 和 5-shot 设置下都具有最好的分割能力, 表明其对标签数量的鲁棒性。尽管 DeepRS 在 5-shot 任务 (a) 中只比我们低了 4.1% 的 DSC, 但由于该方面的对抗训练扩大了其对数据数据数量的敏感性, 因此在 1-shot 任务 (a) 中, 该方法与本章的 BRBS 框架的差距扩大到了 15.8%。PC-Reg-RT 在 5-shot 情况下的 DSC 为具有 88.5%, 比 BRBS 低了 2.6%, 但在 1-shot 设置下, 该差距扩大到了 3.7%。这是因为 PC-Reg-RT 中的配准网络的输入基于其分割网络的输出, 当有标签图像数量极少时, 分割模型的退化使得配准模型性能进一步下降, 使得生成数据出现更大的失真, 进一步干扰并限制分割学习。

B. 定性对比分析

图8展示了在 5-shot 心脏结构和脑组织分割任务上展示了五种基于 LRLS 的方法的结果, 我们发现 BRBS 在大结构的边界上具有高准确性, 并且在小结构上也有着出色的表现。1) 对于大的心脏结构, BRBS 在边界的分割准确性上表现更好。如第一行的放大部分所示, BRBS 和 DeepAtlas 都能够精细地分割两个结构之间

表 II: 与 16 种常用方法相比, BRBS 在 1-shot 和 5-shot 情况下的两个少标签医学图像分割任务中, 都取得了最佳性能。表中“unable”表示分割结果非常差导致 AVD 无法计算。

方法	类型	a) CT 图像上的心脏结构分割				b) MR 图像上的脑组织分割			
		1-shot $\pm std$		5-shot $\pm std$		1-shot $\pm std$		5-shot $\pm std$	
		DSC% \uparrow	AVD $mm \downarrow$						
3D U-Net [13]	LS	63.8 ± 16.3	6.13 ± 3.46	84.3 ± 9.6	2.43 ± 2.14	54.4 ± 10.8	2.94 ± 1.23	69.5 ± 8.8	1.59 ± 0.84
SegNet [14]		57.5 ± 17.4	7.01 ± 4.53	78.8 ± 10.5	2.68 ± 1.72	52.3 ± 4.9	3.18 ± 0.37	62.7 ± 7.0	1.98 ± 0.72
U-Net++ [15]		42.9 ± 20.5	9.18 ± 3.78	84.0 ± 8.6	2.51 ± 2.26	51.2 ± 10.6	2.33 ± 1.06	66.4 ± 12.7	2.02 ± 1.62
DBN [16]		48.8 ± 16.5	10.70 ± 4.10	78.9 ± 12.0	3.90 ± 3.12	23.5 ± 15.9	13.83 ± 7.26	80.2 ± 5.6	0.92 ± 0.30
UA-MT [44]	SLS	54.8 ± 17.0	9.44 ± 4.77	66.4 ± 16.2	4.69 ± 2.27	36.7 ± 8.4	8.69 ± 2.29	75.5 ± 3.4	1.31 ± 0.95
CPS [46]		70.7 ± 9.4	4.01 ± 1.73	87.4 ± 5.4	1.40 ± 0.76	25.3 ± 1.2	unable	37.1 ± 1.8	unable
MASSL [45]		57.2 ± 12.5	13.86 ± 3.16	77.4 ± 8.7	9.07 ± 3.11	74.0 ± 3.1	1.32 ± 0.35	80.5 ± 3.1	0.92 ± 0.43
DPA-DBN [16]		49.0 ± 14.4	10.47 ± 3.81	68.0 ± 14.5	5.75 ± 3.89	28.1 ± 7.6	7.75 ± 1.78	68.7 ± 8.2	3.90 ± 2.39
VM [11]	ABS	77.6 ± 6.0	2.49 ± 0.73	81.0 ± 6.1	2.13 ± 0.78	78.7 ± 1.8	0.73 ± 0.07	83.1 ± 1.8	0.56 ± 0.08
LC-VM [26]		-	-	81.7 ± 6.0	2.04 ± 0.77	-	-	83.0 ± 1.8	0.56 ± 0.07
LT-Net [5]		67.2 ± 6.5	3.55 ± 0.90	77.8 ± 7.8	2.25 ± 0.95	76.9 ± 1.5	0.75 ± 0.51	82.6 ± 1.2	0.57 ± 0.05
DeepAtlas [1]	LRLS	85.4 ± 4.5	1.59 ± 0.56	87.9 ± 4.3	1.30 ± 0.57	73.0 ± 2.4	1.02 ± 0.10	79.3 ± 2.6	0.74 ± 0.12
DataAug [2]		81.4 ± 5.2	2.23 ± 0.67	82.2 ± 5.2	2.04 ± 0.73	81.3 ± 1.4	0.69 ± 0.06	83.9 ± 1.2	0.55 ± 0.06
DeepRS [3]		73.4 ± 12.3	3.40 ± 1.92	87.0 ± 5.0	1.60 ± 0.90	55.9 ± 12.0	1.81 ± 0.91	73.0 ± 5.9	0.93 ± 0.25
PC-Reg-RT [4]		85.5 ± 4.7	1.55 ± 0.63	88.5 ± 4.9	1.23 ± 0.72	66.9 ± 3.6	1.38 ± 0.19	73.1 ± 3.1	1.09 ± 0.17
VAEAug [6]		75.5 ± 11.0	4.29 ± 2.12	-	-	74.8 ± 12.2	1.71 ± 2.71	-	-
BRBS	LRLS	89.2 ± 3.4	1.24 ± 0.50	91.1 ± 3.9	0.93 ± 0.57	85.7 ± 1.0	0.49 ± 0.04	87.2 ± 1.0	0.43 ± 0.05

的边界(黄色箭头), 但 DataAug 在这个区域存在误分割。这是因为 DataAug 在配准时缺乏对语义区域对齐的约束, 因此在通过配准生成伪标签图像时, 存在较为严重的对齐失真。虽然 DeepRS 和 PC-Reg-RT 在训练配准模型时具有语义区域对齐的约束, 但 DeepRS 受到对抗学习的干扰(DeepRS)而 RC-Reg-RT 过于依赖分割模型, 这使得它们的配准容易在细节区域发生失真, 导致随之训练的分割模型在边界处准确性差。2)对于小的脑组织, BRBS 表现出更好的精细结构的分割性能。最后一行的放大部分有多个相邻的小结构。与其他基于 LRLS 的方法相比, BRBS 在这些区域均具有出色的完整性和一致性, 这归功于 BRBS 生成的伪标签图像具有较好的真实性。PC-Reg-RT、DeepAtlas 和 DeepRS 严重丢失了一些小结构, 如黄色箭头所指示的区域, 因为它们的生成过程没有考虑真实性, 而恰恰真实对于小结构来说非常重要。一旦发生失真, 这些小结构的特诊很容易被错误信息所干扰, 导致模型无法有效地学习这些信息。DataAug 在小结构上表现比其他三种方法更好, 这是因为它具有三个独立的训练阶段, 减弱了误分割对配准的干扰。

C. 消融实验

如表III所示, 在 5-shot 心脏结构分割任务上的消融实验展示出本章创新点带来的性能提升。3D U-Net(第一行, 我们的基准线)仅有 84.3% 和 2.43mm 的 DSC 和 AVD。当添加 S3P 但并不进行风格和空间的采样时, 由于缺少 KCC 中分割的约束, 使得配准生成的图像中存在严重的失真问题, 从而导致与 3D U-Net 相比降低了 3.6% 的 DSC。S3P 中的空间采样为分割模型提供了更多不同的空间特征, 使性能提高了 3.0%DSC, 但失真问题仍然干扰着分割学习, 因此 DSC 仍然低于 3D U-Net。加入风格采样后极大地提升了分割模型性能, 获得了 88.1% 和 1.25mm 的 DSC 与 AVD。这是因为不同的 CT 扫描协议导致了 CT 图像风格特征的变化, 导致了仅使用少量有标签数据训练的 3D U-Net 存在的巨大瓶颈, 而我们的风格采样为分割学习提供了更多的风格特征, 极大地提升了模型的泛化能力。当仅添加 MMR(没有风格采样)时, 其正则化能力带来了 0.7% 的 DSC 提升。但与风格级采样相比, 该方法仍然缺乏来自具有更多风格信息的生成数据的监督, 因此仍然比加入风格采样后的模型低 3.7% 的 DSC。KCC 为生成数据带来了更好的真实性, 因此当加入 SeC 时, 配准网络在相同语义区域上获得了更好的对齐, 从而带来

表 III: 在 5-shot 心脏结构分割任务上的消融实验显示了本章创新点的切实贡献。

KCC		S3P		MMR	分割性能 $\pm std$		配准性能 $\pm std$		
SeC	BiC	图像配准	空间采样		DSC% \uparrow	AVD _{mm} \downarrow	DSC% \uparrow	AVD _{mm} \downarrow	$ J_\phi \leq 0\% \downarrow$
					84.3 ± 9.6	2.43 ± 2.14	-	-	-
		✓			80.7 ± 9.6	2.52 ± 1.52	72.6 ± 13.8	2.89 ± 1.18	3.3 ± 0.7
		✓	✓		83.7 ± 8.0	2.33 ± 2.03	73.2 ± 13.8	2.84 ± 1.17	3.2 ± 0.7
		✓	✓	✓	88.1 ± 4.7	1.25 ± 0.63	73.0 ± 13.9	2.87 ± 1.20	3.5 ± 0.8
		✓	✓	✓	84.4 ± 6.5	1.85 ± 0.89	73.5 ± 13.8	2.84 ± 1.18	3.7 ± 0.8
✓		✓	✓	✓	90.0 ± 3.8	1.04 ± 0.49	85.9 ± 13.5	1.33 ± 0.67	6.2 ± 1.2
✓	✓	✓	✓	✓	90.4 ± 3.4	1.00 ± 0.44	86.0 ± 13.5	1.31 ± 0.64	2.5 ± 1.1
✓	✓	✓	✓	✓	91.1 ± 3.9	0.93 ± 0.57	86.7 ± 13.6	1.22 ± 0.62	1.7 ± 0.8

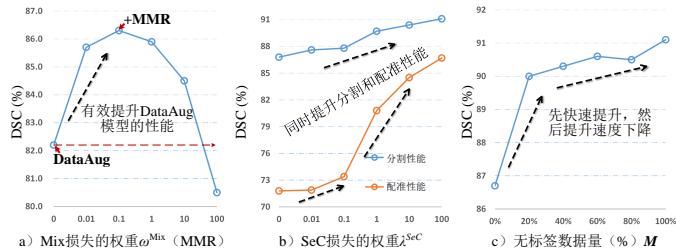


图 9: 该实验在 5-shot 心脏结构分割任务上进行了本章中新引入的超参数的分析。分析了三个超参数: a) MMR 中 Mix 损失的权重 ω^{Mix} , b) SeC 损失的权重 λ^{SeC} , c) 无标签数据集的大小 M 。

了 13.3% 的配准 DSC 提升，并进一步带来了 1.9% 的分割 DSC 提升。但是由于标签内部缺乏纹理信息导致的大形变破坏了图像的拓扑结构，使得 $|J_\phi| \leq 0$ 扩大到了 6.2%。BiC 进一步约束配准网络产生更平滑和可逆的形变，从而降低扭曲程度，获得了 2.5% 的 $|J_\phi| \leq 0$ 。由于 KCC 和 S3P 对生成数据真实性和多样性的提升，BRBS 框架比 3D U-Net 提高了 6.1% 的 DSC。进一步的，当加入 MMR 后，更鲁棒地学习生成的伪标签数据让最终的 BRBS 取得了 91.1% 和 0.93mm 的 DSC 和 AVD。

D. 框架分析

1) 超参数分析: 如图9所示, 该实验对三个新超参数的分析展示了我们创新点对模型的内在影响。a) MMR 显著提高了 DataAug [2] 学习生成数据的鲁棒性，取得了更高的性能，但当其权重大于 0.1 时，性能开始下降。本实验去除其他创新点对性能的干扰，仅在 DataAug 这种极易受到配准对齐失真干扰的模型上评估 MMR 的影响。由于 MMR 对不准确信息的抑制作用，当权重 ω^{Mix} 为 0.1 时，DataAug 模型取得了最高和显著的提升。但当将权重 ω^{Mix} 增加到 100 时，

表 IV: 归功于 KCC 的知识嵌入，5-shot 心脏结构分割任务的配准性能上的分析显示，BRBS 具有更高的配准精度 (DSC, AVD) 和更低的扭曲失真 ($|J_\phi| \leq 0$)。“Trad” 表示传统方法，“DL” 表示深度学习方法。

方法	类型	DSC% $\pm std$ \uparrow	AVD _{mm} $\pm std$ \downarrow	$ J_\phi \leq 0\% \pm std \downarrow$
Initial		62.1 ± 8.7	4.37 ± 1.34	-
B 样条法 [49]	Trad	77.0 ± 11.2	2.61 ± 1.57	8.4 ± 9.7
SyN [7]	Trad	79.2 ± 8.9	2.33 ± 0.92	0
VM [11]	DL	72.7 ± 13.9	2.89 ± 1.20	3.5 ± 0.7
LC-VM [26]	DL	73.0 ± 13.9	2.85 ± 1.19	3.1 ± 0.6
DeepAtlas [1]	DL	79.7 ± 13.6	2.10 ± 0.96	2.5 ± 0.6
PC-Reg-RT [4]	DL	80.8 ± 13.6	1.94 ± 0.91	0.5 ± 0.3
DeepRS [3]	DL	76.9 ± 13.3	2.39 ± 0.99	5.3 ± 1.3
LT-Net [5]	DL	68.2 ± 13.9	3.26 ± 1.29	2.5 ± 0.3
BRBS	DL	86.7 ± 13.6	1.22 ± 0.62	1.7 ± 0.8

MMR 过于强大的正则化作用导致模型欠拟合，极大地削弱了性能（甚至低于原始的 DataAug）。b) 随着其权重 λ^{SeC} 的增大，SeC 损失同时提高了分割和配准性能（我们设置 $\omega^{Mix} = 0.1$ ）。这是因为在配准中，语义区域对齐性能的提升可以促进生成数据的真实性。因此，通过更真实的数据训练的分割网络获得了更好的性能。c) 随着无标签数据集的增大，模型的分割性能迅速提高，当仅使用 20% 的无标签数据集时就取得了 90% 的 DSC。但是进一步增加更多的无标签数据，模型性能的增涨变缓。这是因为 S3P 极大地提高了生成数据的多样性，即是少量的无标签图像就能生成具有大量特征的数据。

2) 配准性能分析: 如表IV所示，与其他配准框架和 LRLS 框架中的配准模型相比，BRBS 在 5-shot 心脏结构分割任务上具有更高的配准精度和更低的形变失真风险。本实验比较了两种传统的 (Trad) 配准模型 (B 样条法 [49] 和 SyN [7])、两个基于深度学习的配准模型 (VM [11] 和 LC-VM [26]) 以及三个 LRLS 模型的配准部分 (DeepAtlas [1], PC-Reg-RT [4] 和 DeepRS

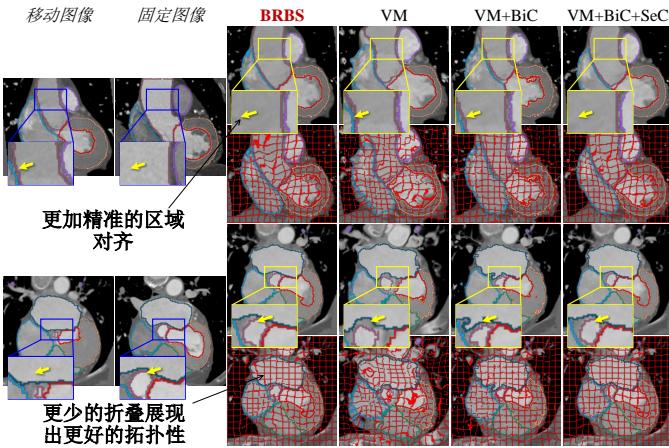


图 10: KCC 通过 BiC 和 SeC 策略优化配准网络的学习, 使其具有更好的保持拓扑性和语义区域对齐的能力。

[3])。B 样条法在 DSC、AVD 和 $|J_\phi| \leq 0$ 上分别为 77.0%、2.61mm 和 8.4%, 展示了该任务的基线性能。SyN 是一种针对平滑性优化设计的配准模型, 因此它取得了 0% 的 $|J_\phi| \leq 0$, 而配准精度也提升到了 79.2% 的 DSC。VM 通过深度学习, 获得了数据驱动的配准能力, 但是该方法缺乏语义区域感知能力, 仅拥有 72.7% 的 DSC。LC-VM 仅通过额外的 5 个标签训练 VM 以获得更好的语义感知和对齐能力, 但由于标签过少, 仅带来了微乎其微的提升 (0.3%DSC)。DeepAtlas 采用分割网络驱动配准模型学习语义区域的对齐, 取得了 79.7% 的 DSC。PC-Reg-RT 仅在感兴趣区域上进行变形, 避免了背景上的潜在失真, 显示出了强大的拓扑保持能力, 具有 0.5% 的 $|J_\phi| \leq 0$ 。但它没有对背景区域进行变形, 缺乏对一些无标签但潜在地与任务相关的区域的对齐。BRBS 取得了最高的 DSC (86.7%), 最低的 AVD (1.22mm) 和相对较低的 $|J_\phi| \leq 0$ (1.7%), 展示出来自本章所提出的 KCC 的强大的拓扑保持和语义对齐能力。BRBS 对对齐精度和扭曲的优化将有效提升生成数据的真实性, 从而进一步提升分割模型的准确性。

3) KCC 分析: 如图10所示, 本章所提出的 KCC 利用 BiC 和 SeC 策略优化配准网络性能, 使其同时拥有拓扑保留和语义对齐的配准能力。因此, 生成的图像将具有更好的真实性, 进而显着提高医学图像分割任务的性能。a) 因为缺乏对拓扑性和语义区域对齐的针对性约束, VM 在变形时产生了许多扭曲、折叠, 同时对于语义区域的对齐性能也较差。因此, 如表IV所示, 它的 $|J_\phi| \leq 0$ 较高 (3.5%) 而准确性较低 (72.7%)。b) BiC 利用了保持拓扑性的形变一定是可逆和平滑的这一领域知识, 构建了一种双向的可逆配准和平滑性约

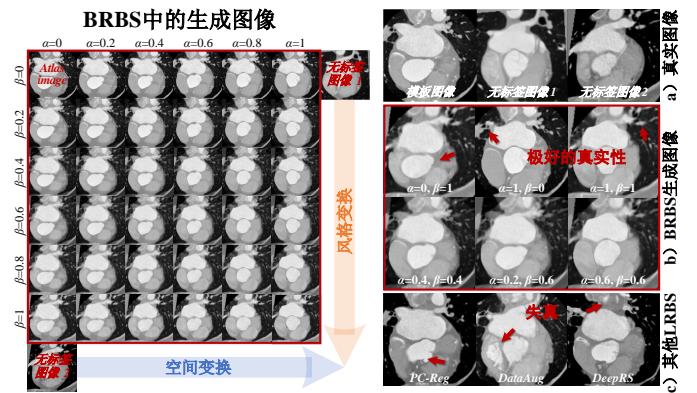
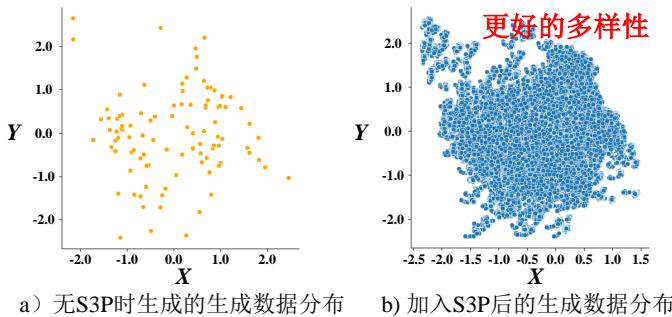


图 11: BRBS 中生成的图像具有很高的多样性和真实性。(左) S3P 通过空间和风格密集采样, 使得生成的图像特征覆盖图像间的风格流形和空间结构流形上的位移路径, 从而获得更加多样的空间和风格特征的数据。(右) 从 a) 到 c) 分别是来自真实世界的图像、来自 S3P 生成图像和来自其他 LRLS 方法的图像。

束。因此, 在加入 BiC 后, 图中的配准结果拥有更少的折叠和扭曲。c) SeC 利用了两个对齐的图像具有对齐的语义区域这一领域知识, 利用分割网络训练配准网络学习对齐两个图像中对应的语义区域。因此, 模型取得了更好的语义区域对齐。

4) 生成数据的真实性分析: 如图11中右侧图所示, 由于 KCC 中的 BiC 和 SeC 策略对数据真实性的促进作用, BRBS 生成的图像具有极好的真实性。a) 与其他基于 LRLS 的方法 ([2]–[4]) 生成的图像相比, BiC 和 SeC 有效地提高了 BRBS 模型生成图像的真实性。DataAug 没有语义对齐和保持拓扑性的约束, 因此生成的图像在边界上具有大量的畸变和伪影。虽然 PC-Reg-RT 和 DeepRS 设计了针对语义区域对齐的约束, 但是由于缺乏保持拓扑性的约束, 它们生成的图像形态结构发生了较大的畸变。我们的 KCC 通过两种领域知识 (BiC 和 SeC) 约束配准学习, 从而生成了具有高真实性的图像。b) 与真实世界的图像 (图11 a)) 相比, 我们可以发现 BRBS 生成的图像同样具有很高的相似性, 因此使用本章方法生成的数据训练的分割网络将学习到与真实数据相匹配的表征能力, 分割网络在真实世界数据上测试时将取得很好的泛化能力。

5) 生成数据的多样性分析: 如图12所示, S3P 通过对空间和风格的密集采样, 为分割网络的学习提供了更加多样化的生成数据。本实验通过从训练好的分割网络的瓶颈层提取各图像的特征, 并通过主成分分析 (PCA) 将这些特征压缩为两个值绘制在散点图上来展示所生成的图像的特征分布。a) 在没有 S3P 的情况下



a) 无S3P时生成的生成数据分布 b) 加入S3P后的生成数据分布

图 12: 从风格网络瓶颈层提取的特征降维 (PCA) 后的散点图展示了 S3P 对生成数据多样性的促进作用。a) 当没有 S3P 时 [1], 生成数据分布稀疏离散, 多样性较低。b) 当加入 S3P 时, BRBS 生成了具有极佳多样性的数据分布。

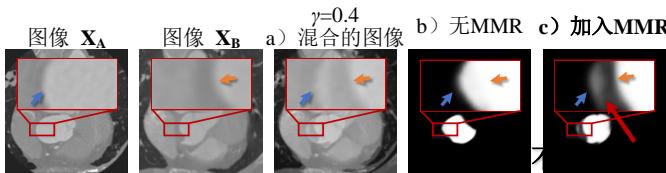


图 13: MMR 约束分割网络感知不对齐的区域, 降低这些不准确信息的干扰。a) 它通过加权融合两个图像来模拟生成具有不对齐区域的图像。b) 当没有 MMR 的时候, 3D U-Net 无法感知不对齐的区域, 使其学习到不对齐区域中不准确的信息。c) 加入 MMR 后, BRBS 拥有感知这些不对齐区域的能力, 并且学习降低对这些区域的拟合程度, 获得了更好的鲁棒性。

[1], 所生成的图像仅具有稀疏的特征分布, 多样性较差。这是因为数据生成程序直接从无标签数据集的稀疏离散分布中采样数据, 因此生成图像构建的特征分布也是稀疏离散的。b) S3P 构建了一种空间和风格连续的分布并该分布上密集采样, 因此生成了大量具有不同空间和风格特征的图像。因此, BRBS 中生成图像的特征分布是密集和多样化的, 能够有效地提高了分割网络的泛化能力。c) 从视觉上看, 图11的左图展示了所生成的图像的空间和样式特征的变换过程。随着 α 和 β 的增大, 模板图像的空间和样式逐渐转换到两个无标签图像上。因此, 所生成的图像的特征将涵盖从模板图像到无标签图像的变换流形路径上的所有特征, 得到更多样化的生成数据。

6) MMR 的正则化能力分析: 如图13所示, MMR 通过模拟生成数据中的对齐失真, 并约束网络降低对这些不准确信息的拟合程度来对网络进行正则化, 提高分割模型学习的鲁棒性。图像 X_A 和 X_B 通过加权求和混合在一起, 使得边界附近的区域 (蓝色和黄色箭头) 在混合图像中不对齐, 模拟了生成数据中可能产生的不对齐失真。MMR 训练模型对不对齐区域拟合一个线性函

数, 从而对这些不对齐区域产生更低的响应 (c), 因此提升了分割模型对对齐失真的鲁棒性。而 3D U-Net 无法感知这些不对齐的区域 (b), 容易过拟合到这些不准确信息, 因此在分割结果中对这些区域表现出过高的置信度。

VI. 总结和讨论

我们提出的 BRBS 框架为 MIS 领域提供了一个强大的少样本分割基准, 并有效降低了医学图像研究的成本。它在多样性、真实性和鲁棒性方面有效推进了 LRLS 范式, 突破了该范式的重要瓶颈。我们提出的 KCC 进一步约束配准网络学习语义对齐和拓扑保持的配准, 使生成程序能够输出具有空间和风格真实性的新数据。我们的 S3P 将风格和空间变化的转换路径建模到生成程序中, 从而在更多样的空间和样式特征上采样生成数据。最后, 我们的 MMR 模拟不对齐失真, 并约束网络减小失配区域的拟合度, 从而提高分割学习的稳健性。在两个具有挑战性的任务上进行的广泛实验证明了我们强大的少样本 MIS 性能和广泛的应用价值, 而且没有任何花哨的技巧。我们还希望在 MindSpore 上使用 BRBS, 这是一个新的深度学习计算框架。这些问题留待未来的研究。我们的所有代码将公开发布在 <https://github.com/YutingHe-list/BRBS>。

A. 失败案例和本方法限制讨论

在我们的实验中, 失败案例展示了本方法潜在的限制之一。失败案例: 如图14所示, 在 1-shot 心脏结构分割任务中, 测试图像上的心肌较厚, 但在分割结果中较薄。这是因为有标签图像上的心肌较厚, 导致大多数生成的图像具有较薄的心肌, 因此训练的分割网络更容易预测较薄的心肌。限制讨论: 有标签图像中特征的偏好是基于 LRLS 的方法的一种限制。这些偏好将被转化到生成的图像中, 从而使分割网络学习到一个偏向于有标签图像中特征的表示。因此, 这将限制模型对一些特征与训练集标签图像不同的图像的泛化能力。幸运的是, 使用更多的有标签图像能够减小这种限制。一种策略是精心选择具有不同特征的多个标签图像, 避免偏好到某种特征的问题, 从而实现具有更好泛化能力的分割网络。

B. 未来工作讨论

尽管我们的 BRBS 在医学图像中展现了巨大的优越性, 但它仍有很大的探索潜力。因此, 我们未来的工

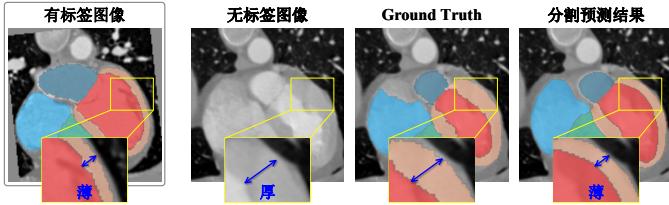


图 14: 由于标签图像中特征的偏差, 在 1-shot 心脏结构分割任务的情况下, 测试图像上的心肌较厚, 但在分割结果中却变得较薄。

作主要集中在两个方面: 1) 在医学图像中探索更多基于配准的高效学习任务。配准利用解剖一致性先验, 在许多常规结构化的医学图像(如 CT、MR)中广泛存在, 为这些图像的高效学习提供了一种有效的方式。因此, 我们将进一步研究学习配准以学习其他任务的范式。2) 扩大 LRLS 范式的应用范围。许多医学图像不具有解剖一致性先验, 例如组织病理学全切片图像 [50]。这使得 LRLS 范式中的配准无法在这些图像上应用, 阻碍了 LRLS 框架在更多任务上取得优异性能。因此, 我们将进一步研究在不规则结构图像上的少样本分割, 从而进一步扩展 LRLS 范式的应用范围。

参考文献

- [1] Z. Xu and M. Niethammer, “Deepatlas: Joint semi-supervised learning of image registration and segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019, pp. 420–429.
- [2] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, “Data augmentation using learned transformations for one-shot medical image segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2019, pp. 8543–8553.
- [3] Y. He, T. Li, G. Yang, Y. Kong, Y. Chen, H. Shu, J.-L. Coatrieux, J.-L. Dillenseger, and S. Li, “Deep complementary joint model for complex scene registration and few-shot segmentation on medical images,” in 16th european conference on computer vision, vol. 1, 2020, pp. 770–786.
- [4] Y. He, T. Li, R. Ge, G. Yang, J. Yang, Y. Kong, J. Zhu, H. Shu, and S. Li, “Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching,” IEEE Journal of Biomedical and Health Informatics, pp. 1–1, 2021.
- [5] S. Wang, S. Cao, D. Wei, R. Wang, K. Ma, L. Wang, D. Meng, and Y. Zheng, “Lt-net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9162–9171.
- [6] Y. Ding, X. Yu, and Y. Yang, “Modeling the probabilistic distribution of unlabeled data for one-shot medical image segmentation,” in AAAI, 2021, pp. 1246–1254.
- [7] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain,” Medical Image Analysis, vol. 12, no. 1, pp. 26–41, 2008.
- [8] F. Liu, J. Cai, Y. Huo, C.-T. Cheng, A. Raju, D. Jin, J. Xiao, A. Yuille, L. Lu, C. Liao et al., “Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. Springer, 2020, pp. 257–274.
- [9] M. Holden, “A review of geometric transformations for nonrigid body registration,” IEEE transactions on medical imaging, vol. 27, no. 1, pp. 111–128, 2007.
- [10] J. Ashburner, “A fast diffeomorphic image registration algorithm,” NeuroImage, vol. 38, no. 1, pp. 95–113, 2007.
- [11] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “An unsupervised learning model for deformable medical image registration,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9252–9260.
- [12] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, “Learning from noisy labels with deep neural networks: A survey,” IEEE Transactions on Neural Networks and Learning Systems, 2022.
- [13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 424–432.
- [14] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.
- [15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” IEEE transactions on medical imaging, vol. 39, no. 6, pp. 1856–1867, 2019.
- [16] Y. He, G. Yang, J. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao, S. Zhang, H. Shu, J.-L. Coatrieux, and S. Li, “Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation,” Medical Image Analysis, vol. 63, p. 101722, 2020.
- [17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in Advances in Neural Information Processing Systems, vol. 32, 2019, pp. 8026–8037.
- [18] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” ACM Computing Surveys (CSUR), 2019.
- [19] L. Zhu and Y. Yang, “Label independent memory for semi-supervised few-shot video classification,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 273–285, 2020.
- [20] M. Zhang, H. Li, S. Pan, X. Chang, C. Zhou, Z. Ge, and S. Su, “One-shot neural architecture search: Maximising diversity to overcome catastrophic forgetting,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 9, pp. 2921–2935, 2020.

- [21] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4077–4087.
- [22] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [23] S. Puch, I. Sánchez, and M. Rowe, "Few-shot learning with deep triplet networks for brain imaging modality recognition," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*. Springer, 2019, pp. 181–189.
- [24] Y. He, G. Yang, J. Yang, R. Ge, Y. Kong, X. Zhu, S. Zhang, P. Shao, H. Shu, J.-L. Dillenseger, J.-L. Coatrieux, and S. Li, "Meta grayscale adaptive network for 3d integrated renal structures segmentation," *Medical Image Analysis*, vol. 71, pp. 102055–102055, 2021.
- [25] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," in *Machine Vision and Applications*, vol. 31, no. 1, 2020, pp. 1–18.
- [26] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 1788–1800, 2019.
- [27] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [29] Y. He, G. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao, S. Zhang, H. Shu, J.-L. Coatrieux, and S. Li, "Dpa-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep priori anatomy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11769, 2019, pp. 139–147.
- [30] M. S. Elmahdy, J. M. Wolterink, H. Sokooti, I. Isgum, and M. Staring, "Adversarial optimization for joint registration and segmentation in prostate ct radiotherapy," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 - 22nd International Conference, Proceedings*, 2019, pp. 366–374.
- [31] N. K. Dinsdale, M. Jenkinson, and A. I. L. Namburete, "Spatial warping network for 3d segmentation of the hippocampus in mr images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 284–291.
- [32] H. Yang, J. Sun, H. Li, L. Wang, and Z. Xu, "Neural multi-atlas label fusion: Application to cardiac mr images." *Medical Image Analysis*, vol. 49, pp. 60–75, 2018.
- [33] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, 2015.
- [34] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in ct scans," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1000–1010, 2009.
- [35] X. Zhu, Y. Liu, J. Li, T. Wan, and Z. Qin, "Emotion classification with data augmentation using generative adversarial networks," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2018, pp. 349–360.
- [36] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Medical Image Analysis*, vol. 65, p. 101759, 2020.
- [37] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2017.
- [38] J. Zhang, "Inverse-consistent deep networks for unsupervised deformable image registration," *arXiv preprint arXiv:1809.03443*, 2018.
- [39] X. Zhuang, L. Li, C. Payer, D. Stern, M. Urschler, M. P. Heinrich, J. Oster, C. Wang, Örjan Smedby, C. Bian, X. Yang, P.-A. Heng, A. Mortazi, U. Bagci, G. Yang, C. Sun, G. Galisot, J.-Y. Ramel, T. Brouard, Q. Tong, W. Si, X. Liao, G. Zeng, Z. Shi, G. Zheng, C. Wang, T. J. MacGillivray, D. E. Newby, K. S. Rhode, S. Ourselin, R. Mohiaddin, J. Keegan, D. N. Firmin, and G. Yang, "Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge," *Medical Image Analysis*, vol. 58, no. 101537, pp. 101537–101537, 2019.
- [40] R. Gharleghi, D. G. Samarasinghe, P. A. Sowmya, and D. S. Beier, "Automated segmentation of coronary arteries," Mar. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3819799>
- [41] M. Schaap, C. T. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mollet, C. Bauer, H. Bogunović, C. Castro, X. Deng, E. Dikici, T. O' Donnell, M. Frenay, O. Friman, M. H. Hoyos, P. H. Kitslaar, K. Krissian, C. Kühnel, M. A. Luengo-Oroz, M. Orkisz, Örjan Smedby, M. Styner, A. Szymczak, H. Tek, C. Wang, S. K. Warfield, S. Zambal, Y. Zhang, G. P. Krestin, and W. J. Niessen, "Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms," *Medical Image Analysis*, vol. 13, no. 5, pp. 701–714, 2009.
- [42] K. Payette, P. de Dumast, H. Kebiri, I. Ezhev, J. C. Paetzold, S. Shit, A. Iqbal, R. Khan, R. Kottke, P. Grethen, H. Ji, L. Lanczi, M. Nagy, M. Beresova, T. D. Nguyen, G. Natalucci, T. Karayannis, B. Menze, M. B. Cuadra, and A. Jakab, "An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset," *Scientific Data*, vol. 8, no. 1, pp. 167–167, 2021.
- [43] D. N. Kennedy, C. Haselgrave, S. M. Hodge, P. S. Rane, N. Makris, and J. A. Frazier, "Candishare: A resource for pediatric neuroimaging data," *Neuroinformatics*, vol. 10, no. 3, pp. 319–322, 2012.
- [44] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, pp. 605–613.
- [45] S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 457–465.
- [46] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2613–2622.
- [47] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *ICLR 2015 : International Conference on Learning Representations 2015*, 2015.

- [48] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, pp. 29–29, 2015.
- [49] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [50] N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep learning for whole slide image analysis: an overview," *Frontiers in medicine*, p. 264, 2019.