感知对应解耦和反向教学法用于少样本弹性形变医学图像配准

何宇霆†, 李甜甜†, 葛荣骏, 杨健, 孔佑勇, 朱健, 舒华忠, 杨冠羽*, 李硕*

摘要—弹性形变医学图像配准旨在通过估计形变场将两幅图像的感兴趣区域（ROIs）对齐到相同的空间坐标系统。然而，最近的无监督配准模型仅具有对应能力而缺乏感知能力，导致在模糊的解剖结构上出现错位并在与任务无关的背景上出现失真。有标签约束（LC）配准模型通过标签嵌入感知能力，但由于标签中缺乏纹理约束以及昂贵的标注成本，导致 ROI 内部发生失真和对感知目标的过度拟合。为解决这些挑战，我们提出了第一个少样本弹性形变医学图像配准框架，感知对应配准（Perception-Correspondence Registration, PC-Reg），它通过仅使用少量标签将感知能力嵌入配准模型中，从而显著提高了配准准确性并减少失真。具体来说，1) 我们提出了感知对应解耦架构，将配准的感知和对应行为分离为两个卷积神经网络（CNNs）。因此，可以进行独立的优化和特征表征，避免由于缺乏纹理约束而对对应预测的干扰。2) 对于少样本学习，我们提出了 Reverse Teaching，将有标签和无标签图像彼此对齐，为无标签图像中的结构和样式知识提供监督信息，从而生成额外的训练数据。基于此，这些数据将反过来教导感知 CNN 更多的样式和结构知识，提高其泛化能力。

我们在仅有五个标签的三个数据集上进行的实验证明，PC-Reg 具有竞争力的配准准确性和有效的减少失真的能力。与 LC-VoxelMorph($\lambda = 1$) 相比，在三个数据集上，我们实现了分别为 12.5%、6.3% 和 1.0% 的 Reg-DSC 提升，显示了我们的框架在临床应用中具有巨大潜力。

Index Terms—**弹性形变医学图像配准，感知对应解耦，少样本学习，反向教学法，减少失真。**

(†共同一作：何宇霆，李甜甜；*通讯作者：杨冠羽，李硕。)

何宇霆†，李甜甜†，葛荣骏，孔侑勇，舒华忠，杨冠羽* 隶属于东南大学计算机网络与信息集成重点实验室（LIST），中国南京，210096。（邮箱: yang.list@seu.edu.cn）

杨冠羽*也隶属于东南大学网络空间安全学院，中国南京，210096。

杨健隶属于北京理工大学光电学院混合现实与先进显示北京市工程研究中心，中国北京

朱健隶属于山东大学齐鲁医院放射肿瘤科，中国济南 250117。

李硕* 隶属于加拿大西安大略省伦敦市的西安大略大学医学影像系。(邮箱: sli287@uwo.ca)

这项研究得到了中国国家重点研发计划（2017YFC0109202）、国家自然科学基金（31800825、31571001、61828101）、东南大学优秀项目基金以及东南大学研究生院科研基金（YBPY2139）的支持。我们感谢东南大学大数据计算中心在本文数值计算方面提供的设施支持。

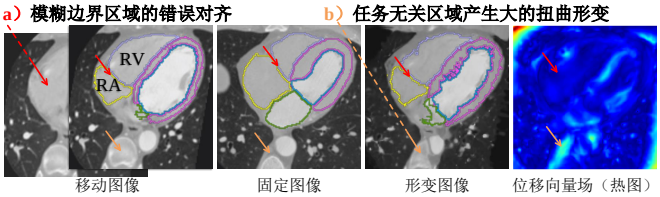

图 1: 弹性形变配准模型对感兴趣区域感知能力的缺失限制了模型对语义区域的精细对齐，导致：a）对比度低的解剖结构发生验重的错误配准 [1]；b）任务无关但显著的背景区域导致发生极大的形变扭曲，削弱配准的效果。

I. 介绍

弹性形变医学图像配准 [2], [3] 旨在将医学图像间的解剖结构对齐到同一空间中，从而帮助放射学医生在对齐的图像间观察并分析图象中的感兴趣区域，获得更加合理和准确的判断。因此该任务是临床诊断和研究中关键的图像处理任务之一 [2]。传统方法 [2] 采用迭代优化来估计弹性形变场，通常需要花费较多时间来获得配准结果。近年来，基于深度学习的弹性形变医学图像配准 [3], [4] 利用卷积神经网络学习对数据特征的表征能力，并利用图像特征直接预测配准形变，从而仅使用一次前向运算即可获得弹性形变场，结合图形计算单元（GPU）的快速发展，基于深度学习的配准展现出了兼具快速和准确的性能。

然而，如图2 a）所示，无监督弹性形变配准模型 [1], [3] 仅学习预测位移向量场（DVF）来获得图像上像素位置间的对应关系以对齐图像（对应能力），在该学习过程中，神经网络本身并没有获得表征“这些像素是什么”的语义感知能力，导致了两个巨大的挑战：1) **无法对齐模糊解剖结构**。如图图1 a）所示，由于缺乏对不同语义区域类型的监督，无监督的配准网络在学习的过程中无法学习对这些区域有鉴别力的特征表示，缺乏语义感知能力。一旦这些语义区域在外观上近似，如图中的右心房（RA）和右心室（RV）之间对比度低，边界模糊，网络将无法感知两者的差异，导致在配准过程

中无法将其对齐到准确的目标。2）**任务无关区域的干扰**。如图1 b）所示，一些类似于脊柱这种显著但与任务无关的区域将在配准学习过程中产生较大梯度，在缺乏感知能力的无监督配准模型中，这些显著无关区域将使其更倾向于将这些区域配准。又由于模型对于拓扑不变性的约束，任务感兴趣区域的对齐程度不得不为之妥协，从而限制了对这些重要区域的配准精度。

如图2 b）所示，尽管一些已有的分割标签驱动的配准模型 [4], [5] 采能够通过分割标签在学习配准的对应能力的同时学习到对任务感兴趣区域的感知能力，但是这种感知能力与对应能力被混合在神经网络内部，使地彼此之间的劣势相互影响后被放大，导致了两方面的限制：1）**标签内缺少纹理约束导致的边缘强扭曲**。分割标签缺少感兴趣区域内部的纹理信息，使得损失函数只存在对齐像素和不对齐像素这种 0/1 式迭代评估，导致那些不对齐区域产生巨大损失而对齐区域没有损失。因此，那些不容易对齐的边缘区域将产生非常大的梯度，促使模型对这些区域进行大的形变以使其对齐，最终破坏图像中原有的拓扑结构，使得配准后的图像失真。2）**分割标签数量限制感知学习性能**。由于标注分割标签成本高，使得在配准任务中难以获得大量的标签，而仅使用少量来训练配准模型学习语义感知将很容易使得模型的感知能力过拟合到这些少量的带标签图像上，限制了其感知的泛化能力 [6]。因此，这些少量分割标签驱动的配准模型无法适应测试数据的解剖结构特征的变化，在测试过程中，出现对语义区域的错误感知，最终导致错误对应，削弱了配准的性能。

在本文中，我们致力于仅使用极少量分割标签来训练基于深度学习的弹性形变医学图像配准模型，在几乎可以忽略不记的额外成本下，显著提升模型性能：1）为了减少感兴趣区域配准的失真问题，我们将整体模型的感知和对应行为解耦为两个网络（感知网络和对应网络），两者分别进行独立的优化和特征学习（图2 c））。因此，当模型通过学习分割标签获得了语义感知能力时时，模型的对应行为的学习不直接受到分割标签作为损失的影响，仍将结合语义区域内部纹理以学习对齐整个感兴趣区域，从而减少配准的失真问题。2）减少分割标签带来的额外成本问题，同时避免少量分割标签带来的感知过拟合问题，我们提出了一种反向教学方法（Reverse Teaching）。该方法利用对应网络从无标记的图像生成伪标签图像对，反向训练感知网络更多的解剖知识以提升感知网络的泛化能力。因此，随着感知网络性能的提升，对应网络也将在更准确的感兴趣区域上学习更精准的对齐。最终使得仅使用少量分割标签就能够获得出色的配准性能。具体如下：

创新点 1：*感知-对应解耦架构*（Perception-Correspondence Decoupling）本文设计了一种*感知-对应解耦架构*来帮助模型在通过分割标签学习感知能力的同时保留感兴趣区域内部的纹理结构，从而保持配准图像的真实性。该架构通过将神经网络在弹性形变配准中的感知和对应行为解耦成两个网络，分别独立地学习它们各自的特征表示 [7]。因此，当将分割标签中的解剖结构知识嵌入到配准模型 [4] 中时，感知网络学习分割的过程将不会干扰对应网络的学习，从而避免了由于分割标签中缺乏纹理信息而导致配准时对解剖结构的过度扭曲，获得同时考虑到感兴趣区域内部纹理的精细配准。感知网络还将使模型关注感兴趣区域并去除背景，因此对应网络只需要学习如何对齐感兴趣区域，而无需受到背景中与任务无关但外观显著区域的干扰，实现有效的配准。此外，在整个过程中，感知与对应将形成一种正向的相互促进的学习模式，感知结果越精确，对应网络对解剖结构的配准也将更加准确，从而间接地获得更好的配准性能。

创新点 2：*反向教学法*（Reverse Teaching）本文提出了一种*反向教学方法*，在少量分割标签的情况下提升感知网络的泛化能力，从而提高模型对感兴趣区域的感知能力，最终促进配准性能。该方法基于对应网络对图像配准的能力，利用少量带有分割标签的数据和大量无标签的数据，反过来训练感知网络学习更加丰富的解剖结构特征，从而提升感知能力。包括两个方面的增强：1）生成大量伪标签图像对。利用对应网络将大量无标签的图像配准到少量带标签图像上，使标签与无标签图像对齐，生成大量新的伪标签图像对。这些数据拥有来自无标签图像的丰富风格特征信息，将训练感知网络学习到更多的解剖结构风格的表征能力，获得更强大的泛化能力。2）生成大量增强标签图像对。利用对应网络将少量带标签图像配准到大量无标签图像上，使得被变换后的带标签图像拥有无标签图像的空间信息，从而利用这些拥有新风格信息的数据训练感知网络，使其学习到更多新的解剖结构空间特征的表征能力，拥有更强的泛化性能。

总的来说，在本文中，我们设计了“感知-对应配准（PC-Reg）”框架用于少标签情况下的弹性形变医学图像配准，仅使用几个标签就可以极大提升配准模型的性

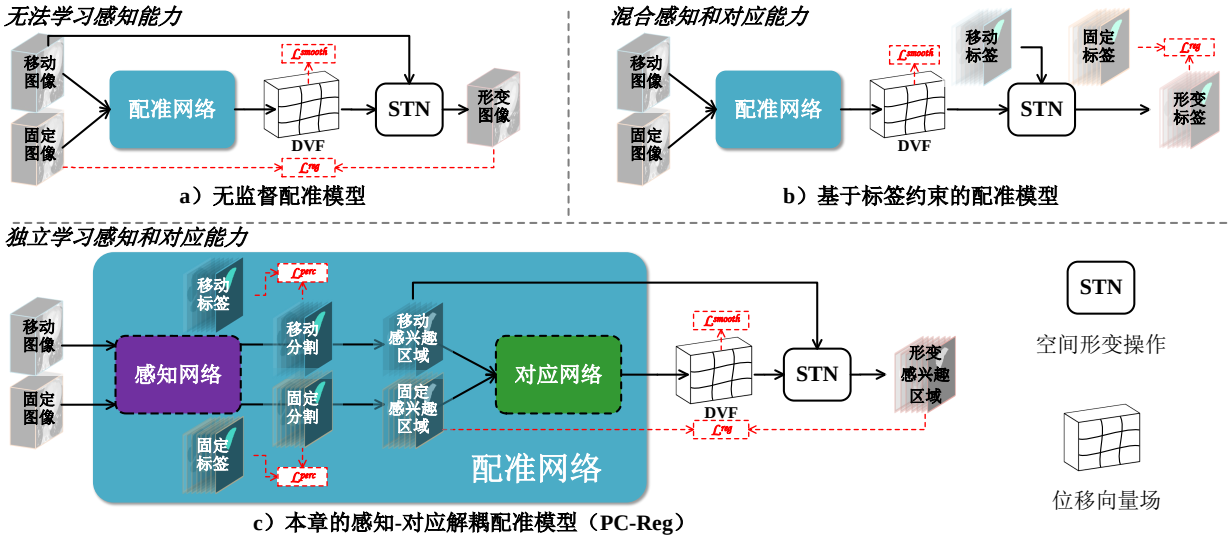

图 2: PC-Reg 与现有的基于深度学习的配准模型相比的优势。a）无监督配准模型只学习图像间的对应关系，神经网络没有学习感知语义区域的能力。b）基于标签约束的配准模型将感知和对应关系混合在一个网络中，使得对应学习受到缺乏纹理信息的标签的影响，限制最终的配准性能。c）我们的 PC-Reg 将配准模型的感知与对应行为解耦，利用感知网络和配准网络分别学习感知感兴趣区域并独立地对对应的感兴趣区域配准，获得极好的配准性能。

能。具体贡献如下：

- 我们提出了一种新的少标签医学图像配准学习框架，“感知-对应配准（PC-Reg）”框架。该框架只需要少量的分割标签，在极少的额外成本下就可以极大地提高了对感兴趣区域的配准精度，同时减少过度形变带来的失真问题。
- 我们提出了“感知-对应解耦”架构，通过分割标签提高模型的语义感知能力，同时获得图像内部纹理的有效配准。它将模型感知和对应行为分解为两个网络进行独立优化和特征表征。从而避免缺乏纹理的分割标签对配准模型对应学习的干扰，进而更好的利用模型对感兴趣区域的感知学习促进配准的性能。
- 我们提出了“反向教学法”，使用少量标签来提高模型中感知网络的泛化能力。它利用对应网络将大量无标签图像信息融入到感知学习中，反向引导感知网络学习更加丰富的解剖结构的特征表征，从而进一步促进配准性能。

本文的剩余部分组织如下。在第II节中，我们回顾了可变形医学图像配准、计算机视觉中的感知和对应、以及医学图像分析中的少样本学习的相关工作。然后，在第III节中，我们具体介绍了我们提出的 PC-Reg，包括我们的感知-对应解耦技术（第III-A节）、我们的反向教学技术（第III-B节）以及详细的网络结构（第III-C节）。然后，在第IV节中描述了数据集、比较方法、实现和评估指标。在第V节中展示和分析了结果。最后，在第VI节中得出了结论。

II. 相关工作

A. 弹性形变医学图像配准

弹性形变医学图像配准 [2], [3] 将医学图像中的解剖结构对准到同一空间中，是一项重要的医学图像处理任务。传统方法 [2]，如 B 样条配准 [8]、对称标准化（SyN）[9] 等，通过多次迭代的方式逐步优化图像间的对齐程度获得最终的配准，时间效率较差 [4]。随着 GPU 的快速发展和深度学习强大的可学习能力，基于深度学习的方法 [3], [4] 开始被广泛应用到弹性形变医学图像配准中，改方法通过端到端的方式直接学习预测位移向量场（DVF），避免了反复的迭代优化，大大提高了配准模型的时间效率和有竞争力的精度。

本文的研究重点是基于深度学习的弹性形变配准，包括三种不同的类型：1）基于监督学习的配准 [10], [11]。这类方法通过空间形变来模拟生成或采用传统配准方法迭代计算获得位移向量场，进而利用这些向量场来训练配准网络以获得预测配准的能力。然而，由于传统配准方法或模拟方法准确性和真实性的限制，生成 DVF 的误差进一步传递到模型中，使得网络预测的准确性受到限制。2）基于无监督学习的配准 [1], [12], [13]。受到可导空间变换操作 [14] 的启发，无监督配准将形变过程集成到网络中，使得配准可以通过度量形变

图像与固定图像间的对齐程度来端到端地训练 [1]。这些方法可以直接使用类似于均方误差距离 [1]、归一化互相关 [1] 等基于图像强度的相似性度量作为损失函数来训练网络，从而无需任何标签即可学习弹性形变配准。然而，这种无监督方法使得模型缺乏对感兴趣区域的感知能力，只能够根据强度相似性来寻求所有解剖结构的空间对齐，容易受到与任务无关的显著区域的干扰，从而限制了感兴趣区域的配准性能。3）基于弱监督学习的配准 [5], [15]。这类方法利用分割标签或关键点标签等辅助信息作为额外的约束来指导网络感知感兴趣区域，从而减弱任务无关区域对配准的影响。然而，这些辅助信息需要额外的精细标注，将带来额外的人工标注成本。一种常见的情况是只有少量标签参与到感知学习中，但是这样也将导致模型的感知能力出现过拟合问题。此外，标签中缺乏真实纹理约束也将导致边缘区域极大的形变，使得配准结果失真 [4]，降低了其临床实用性。

B. 计算机视觉中的感知和对应问题

感知和对应学习是计算机视觉中的两个基本问题。对于感知问题，目标检测任务 [16] 旨在感知图像中目标对象的位置和大小，实例分割 [17] 和语义分割 [18], [19] 任务旨在感知每个像素所属的实例对象和语义对象。对于对应问题，常见的光流估计 [20] 和弹性形变配准 [1], [21] 任务都旨在获得图像间像素的对应关系，而目标跟踪 [22]、仿射变换配准 [23] 和目标匹配 [24] 任务都旨在获得图像块之间的对应关系。

我们的少标签弹性形变医学图像配准任务将感知问题和对应问题结合，旨在使用少量的分割标签指导模型学习感知感兴趣区域的同时学习对这些感兴趣区域的像素级对应以实现配准。由于感兴趣区域标签的限制，我们构建了一种解耦感知和对应行为的网络架构，并且构建相互促进的学习模式，实现了感知和对应的互补拓扑，最终大大提升了弹性形变配准模型的性能。

C. 医学图像分析中的少标签学习

少标签学习是利用少量的监督信息从数据中学习有效表征的学习范式，该范式大大降低了构建大规模标签数据集的巨大成本 [6]，有着极大的应用潜力。在医学图像分析中，少标签学习以其标签高效性展现出极大的研究价值和迫切性，已经在许多场景中取得了成功，如，病理图像分类 [25]、脑影像诊断 [26]、分割 [18] 等。这些极好的研究大大减少医学图像分析模型所需的标签数量，降低了标注工作的专业性和成本，有着极大的大规模应用潜力。

然而，在本文研究之前，还没有针对少标签弹性形变医学图像配准的研究被报道，使得模型无法同时兼顾低标注成本和高配准精度。本文所提出的反向教学法利用对应网络的配准能力，利用大量无标签图像和少量带标签图像生成大量额外的监督信息，反向教授感知网络仿佛的解剖风格和空间结构知识，大大提高模型的泛化能力。

III. 方法

如图3所示，PC-Reg 使用少量标签就能实现有效的弹性形变医学图像配准。通过将配准学习中的感知和对应行为解耦到两个网络中（感知-对应解耦框架，第III-A节），并构建一种对应行为反向促进感知行为的学习模式（反向教学法，第III-B节），最终仅使用极少的数据即可获得精细的配准性能。

A. 感知-对应解耦框架

如图3所示，PC-Reg 将配准模型 $\mathbb{R}_{\theta,\xi}$ 解耦为一个感知网络 $\mathbb{P}_\theta$ 和一个对应网络 $\mathbb{C}_\xi$，通过独立的优化和特征表征学习，以避免感知学习对对应学习的负面影响。其中，θ 和 ξ 分别代表感知网络和对应网络内部的可训练参数。因此，整个配准框架将学习感知具有清晰边界的感兴趣区域，并进一步利用这些感兴趣区域进行准确的对应学习，避免了因分割标签内部缺乏真实纹理而导致的失真问题。

1) *感知具有清晰边界的感兴趣区域:* 感知网络为对应网络的学习提供了具有清晰边界的感兴趣区域，因此对应网络能够学习更加有效的像素间对应关系。如图3所示，将移动图像 x_m 和固定图像 x_f 被反别输入感知网络 $\mathbb{P}_\theta$ 中，获得具有 C 个语义类别的移动图像和固定图像的分割结果掩膜，$\hat{y}_m$ 和 $\hat{y}_f$。这些分割结果去除背景通道后，与它们的原始图像 x_m 和 x_f 相乘，从而获得具有 $C-1$ 个语义类型的带有真实纹理的移动和固定感兴趣区域组，r_m 和 r_f，从而消除在背景中与任务无关的显著区域并为对应网络提供清晰的感兴趣区域边界。在训练感知网络过程中，我们使用交叉熵损失函数 $\mathcal{L}^{ce}$ 计算预测的固定图像和移动图像分割结果与其对应标签距离（$\mathcal{L}_\theta^{fce} = \mathcal{L}_\theta^{ce}(\hat{y}_f, y_f)$，$\mathcal{L}_\theta^{mce} = \mathcal{L}_\theta^{ce}(\hat{y}_m, y_m)$），

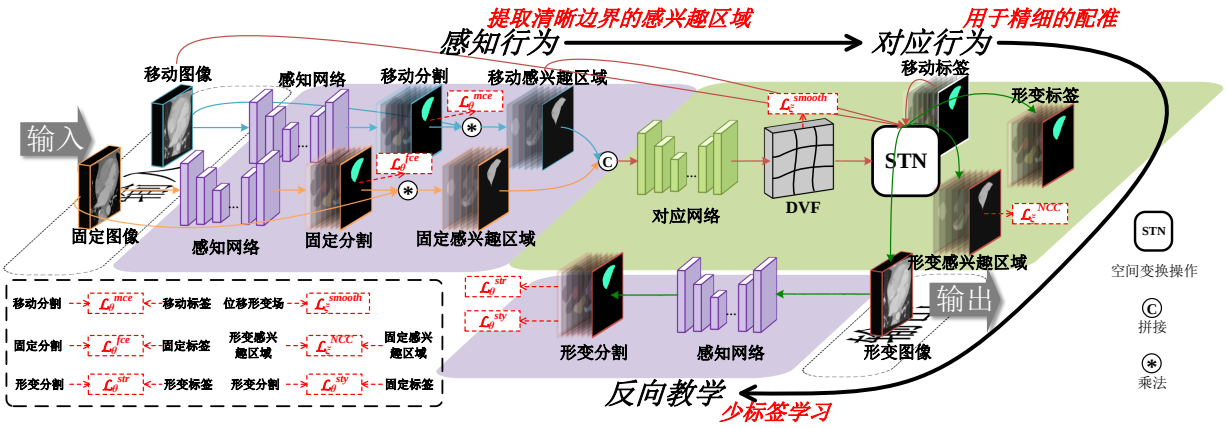

图 3: PC-Reg 框架的具体结构。配准模型的感知和对应行为被分到了两个神经网络中，独立地学习配准所需的感知和对应能力。反向教学法进一步从无标签图像中教授感知网络丰富的结构和风格特征，驱动其获得更强大的感知能力。

从而通过梯度反向传播以优化感知网络。交叉熵损失函数为：

$$\mathcal{L}_{\theta}^{ce}(\hat{y}, y) = -\frac{1}{CK}\sum_{n=0}^{C}\sum_{k=0}^{K} y_{k,n}\log \hat{y}_{k,n}, \qquad (1)$$

其中，C 是感知网络输出的语义类别数量（包括背景类），K 是每个类别分割掩膜中的体素数量，y 是像分割标签，$\hat{y}$ 是预测的分割结果。

2) *在纹理约束下的对应学习*：对应网络基于感知网络输出的感兴趣区域组，准确地学习对齐这些区域和区域内部的解剖结构。如图3所示，我们将利用感知网络获得的移动和固定感兴趣区域组在通道方向上连接起来，输入对应网络 $\mathbb{C}_{\xi}$，从而获得一个对感兴趣区域进行精细地配准但不会在背景上发生扭曲的位移向量场（DVF）ϕ。该 DVF 通过空间变换操作 [14] 将移动图像的感兴趣区域、移动图像的分割结果和移动图像本身同时进行变形，获得形变感兴趣区域 r_w，形变分割结果 y_w 和形变图像 x_w。我们将形变感兴趣区域与固定感兴趣区域一起计算归一化互相关损失函数 [1]$\mathcal{L}_{\xi}^{NCC}$，从而实现在感兴趣区域内部的纹理约束，防止直接使用分割掩膜带来的边缘过度扭曲。对于形变场，我们同时计算一个平滑正则化损失函数 [1] $\mathcal{L}^{smooth}$，以惩罚 DVF 中的过度扭曲，从而防止配准过程发生过度的畸变：

$$\mathcal{L}_{\xi}^{corr} = \sum_{c=0}^{C-1}\mathcal{L}_{\xi}^{NCC}(r_{w_n}, r_{f_n}) + \lambda^{smooth}\mathcal{L}_{\xi}^{smooth}(\phi) \qquad (2)$$

其中，λ^{smooth} 是平衡两个损失函数的权重。在我们的整体 PC-Reg 框架中，其整体的前向运算过程为 $x_w = \mathbb{R}(x_f, x_m)(x_m) \rightarrow \mathbb{C}(\mathbb{P}(x_f)_{\{1,\dots,C\}} * x_f, \mathbb{P}(x_m)_{\{1,\dots,C\}} * x_m)(x_m)$，其中 $\rightarrow$ 表示解耦。

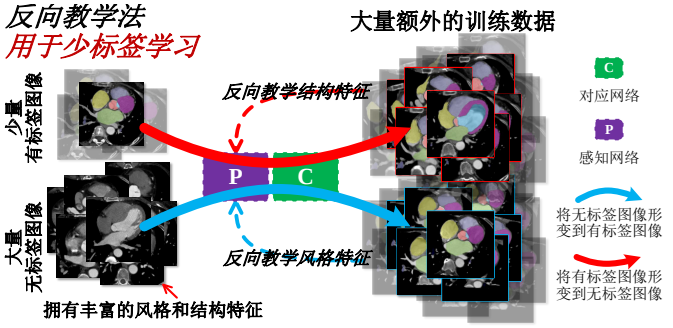

图 4: 反向教学法将少量有标签图像与大量的无标签图像相互对齐生成额外的训练数据，从而教授感知网络丰富的结构和风格特征以实现高效的少标签学习。

为避免一开始感知网络性能较差而干扰整体训练，我们构建了一种两阶段的训练策略。在第一阶段中，感知网络使用少量有标签数据独立训练，以获得基础的感知能力，而对应网络也仅在这些少量有标签数据上利用其来自分割标签的准确感兴趣区域训练，以获得基础的对应能力。接着，在第二阶段中我们加入大量的无标签图像，并结合我们的反向教学法（第III-B节），进一步提高感知网络性能的同时，利用感知网络输出结果训练对应网络，进一步获得更强大的泛化能力。

B. 反向教学用于少标签感知学习

如图4所示，反向教学法（Reverse Teaching）利用对应网络，将少量有标签图像和大量无标签图像进行对齐，从而将标签中的监督信息映射到无标签图像上，从而产生额外的训练数据，包括（形变图像和形变标签对、形变图像和固定标签对）。因此，如图3所示，这些生成的新数据被输入到感知网络中，以反向教授感知网络学

习对无标签图像中丰富的结构和风格特征表征，提高感知网络的泛化能力。该过程包括两个部分：

1）*形变有标签图像以学习丰富结构特征*：反向教学法将少量有标签图像形变到大量无标签图像的空间中，为感知网络提供学习来自无标签图像的丰富空间结构特征。大量的无标签图像拥有丰富的空间结构特征，利用本文的 PC-Reg 算法，反向教学法将少量的有标签图像形变并对齐到大量等等无标签图像上。因此，形变的有标签图像将拥有无标签图像的空间结构特征，从而产生丰富的形变图像和形变标签对。这些数据对为感知网络提供了大量的结构特征，促进感知网络获得对更广泛的空间结构的泛化能力。如图3所示，从对应网络中获得的形变图像 x_w 被输入感知网络以获得形变图像的分割结果 $\hat{y}_w$。该分割结果与形变标签 y_w 一起计算交叉熵（Equ. 1）作为结构损失函数 $\mathcal{L}_\theta^{str} = \mathcal{L}_\theta^{ce}(\hat{y}_w, y_w)$，训练感知网络泛化到更多的解剖结构形态上。

2）*形变无标签图像以学习丰富风格特征*：反向教学法进一步将大量的无标签图像形变到少量有标签图像的空间中与标签对齐，为感知网络提供来自无标签数据的丰富的风格特征。大量的无标签图像拥有丰富的风格特征，利用本文的 PC-Reg 算法，反向教学法将其形变到有标签的图像空间中，使标签与之对齐，构成形变图像和固定标签对，使大量无标签图像的丰富风格特征融入到生成的标签对中。因此，这些大量标签对将训练感知网络获得更加丰富的风格特征表征能力。如图3所示，来自对应网络的形变图像 x_w 被输入到感知网络中以获得形变图像的分割结果 $\hat{y}_w$。该分割结果与固定标签 y_f 一起计算计算交叉熵（Equ.1）作为风格损失函数 $\mathcal{L}_\theta^{sty} = \mathcal{L}_\theta^{ce}(\hat{y}_w, y_f)$，训练感知网络泛化到更多的图像风格上。

总的来说，在训练的过程中，随机采样图像获得的不同的有标签和无标签图像组合，感知网络将使用四种不同的损失函数组和 $\mathcal{L}_\theta^{perc}$ 来进行训练：1）固定图像和移动图像都是有标签图像，存在固定标签 y_f 和移动标签 y_m；2）仅固定图像为有标签图像，存在固定图像标签 y_f；3）仅移动图像为有标签图像，存在移动图像标签 y_m；4）两张图象均为无标签图像：

$$\mathcal{L}_\theta^{perc} = \begin{cases} \mathcal{L}_\theta^{fce} + \mathcal{L}_\theta^{mce} + \mathcal{L}_\theta^{str} + \lambda^{style}\mathcal{L}_\theta^{sty} & \text{if } y_f \ \& \ y_m \\ \mathcal{L}_\theta^{mce} + \mathcal{L}_\theta^{str} & \text{if } y_m \\ \mathcal{L}_\theta^{fce} + \lambda^{sty}\mathcal{L}_\theta^{sty} & \text{if } y_f \\ 0 & \text{if no label} \end{cases} \tag{3}$$

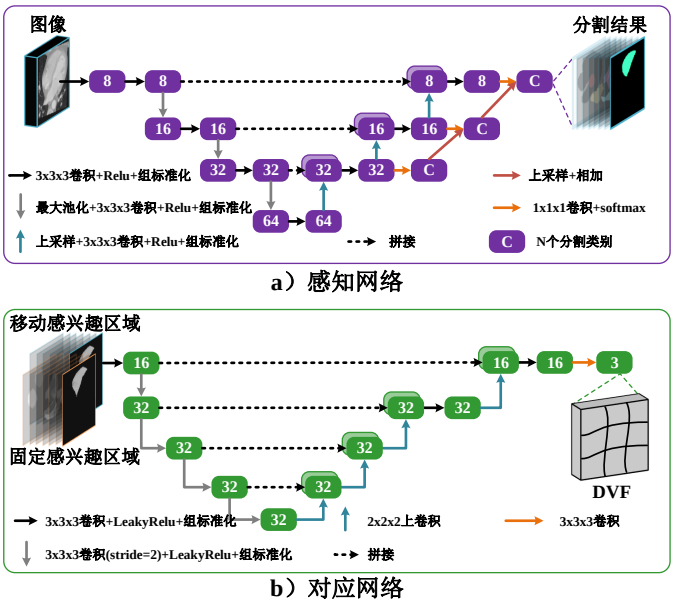

图 5: 感知网络和对应网络的详细网络结构。

其中，λ^{sty} 是形变图像和固定标签对的权重，用来削弱由于配准误差带来的形变的无标签图像与固定标签之间不对齐造成的影响。反向教学法在模型的整体训练过程的第二阶段使用，以避免由于刚开始较弱的配准性能使得形变图像-固定标签对对齐程度低而干扰感知网络学习。

C. 网络结构细节

感知网络：感知网络对固定图像和移动图像进行分割以获得感兴趣区域的掩模。如图5a）所示，它采用了 3D U-Net [27] 结构，包括编码器和解码器两个部分，并加入深度监督 [28] 以增强学习效率。编码器对输入图像逐层卷积的同时逐阶段进行下采样，从而节省 GPU 内存并扩展卷积感受野，解码器在逐层卷积的同时逐阶段进行上采样，以输出与原始分辨率相同的分割结果。整个网络具有四个分辨率阶段，每个阶段有两个 $3 \times 3 \times 3$ 卷积层，每层卷积后都跟随一个 LeakyReLU 激活函数层和组标准化层 [29]。在每个分辨率阶段分别使用最大池化层进行下采样和三线性插值进行上采样。编码器和解码器的相同分辨率阶段之间，使用跳跃连接将编码其中带有细节信息的特征图传输到解码器中。解码器中的每个分辨率阶段的特征图通过 $1 \times 1 \times 1$ 卷积输出越预测该分辨率的分割结果，所有分辨率的分割结果求平均获得最终结果以进行深度监督。

对应网络：对应网络利用移动和固定感兴趣区域预测对这些区域的位移形变场。如图5b）所示，它采用了

5 个分辨率阶段的 3D U-Net [27] 结构。与感知网络类似，在编码器中，每个分辨率阶段有一个 $3 \times 3 \times 3$ 卷积层，每层卷积后都跟随一个 LeakyReLU 激活函数层和组标准化层 [29]，然后使用最大池化在每个阶段之间进行特征图下采样。在解码器中，前两个阶段有一个 $3 \times 3 \times 3$ 卷积-LeakyReLU-组标准化层组，后两个阶段有两个这样的层组以更精细地预测形变。最后，使用 $1 \times 1 \times 1$ 卷积层预测具有三个通道的 DVF，以对每个体素在 x，y，z 方向上进行位移实现对整个图像的形变。

IV. 实验设置

1）**数据集：**我们在三个不同的器官不同模态的数据集上进行了实验以验证我们方法的有效性，包括心脏 CT、颈椎 CT 和脑 MR：

心脏 CT 图像配准：该数据集来自于 MM-WHS2017 挑战赛 [30] 中的 CT 图像数据集，其中包括 20 个有标签 CT 图像和 40 个无标签 CT 图像。该数据集分割了心脏 CT 图像上的七个心脏相关的解剖结构区域，包括升主动脉（AO）、左心房（LA）、左心室（LV）、左室心肌（Myo）、肺动脉（PA）、右心房（RA）和右心室（RV）。在我们的实验中，首先，将心脏区域裁剪到一个矩形区域内，以减少不匹配内容的情况。然后，我们将所有数据重采样为 $128 \times 128 \times 96$ 以获得统一的网络的输入大小。为了构建少标签设置，我们随机选择了 5 个有标签的图像和 40 个无标记的图像作为训练集，剩下的 15 个有标签的图像一一配对，形成 210 个图像对作为测试集。

颈椎 CT 图像配准：该数据集包括了来自 43 个患者的 43 张颈部 CT 图像，我们邀请了两位放射科医生对图像上的七个颈椎（C1、C2、C3、C4、C5、C6、C7）进行了精细的分割标注，并且对这两个医生分别标注的结果进行交叉检查以获得统一的一套分割标签。这些图像被重采样为 $128 \times 128 \times 128$ 以获得统一的网络输入大小。我们随机选择 27 个图像作为训练集，其中随机选择 5 个作为有标签图像和 22 个作为无标签图像以构建少标签设置，剩下的 16 个图像两两配对作为测试集。

脑部 MR 图像配准：该数据集源自于 LBPA40 [31]，包含 40 个脑部 MR 图像，每个图象拥有 56 个标签区域，我们选择了其中 4 个区域（B1、B2、B3、B4）作为我们任务的感兴趣区域。这些图像被裁剪并重新采样为 $128 \times 144 \times 112$，以获得统一达的网络输入尺寸。我们随机选择了 30 个图像作为训练集，其中随机选择 5 个作为有标签图像和 22 个作为无标签图像以构建少标签设置，剩余的 10 个图像两两配对作为测试集。

在输入网络之前，每个图像对都经过仿射变换进行初始化[1] [1]，使得所有数据统一在相同的全局空间中，以帮助实验专注于我们的弹性形变配准任务。

2）**对比设置：**为了展示 PC-Reg 的优越性，我们与五种流行的弹性形变配准方法进行了比较，包括两种传统方法（B 样条法 [8]，SyN [9]）和四种基于深度学习的方法（无监督（Unsup-）VoxelMorph [1]，标签约束（LC-）VoxelMorph [4]，CycleMorph [13] 和 DeepRS [5]）。B 样条法使用 Elastix[1] 实现，并采用其默认配准参数 [8]，SyN 采用 Advanced Normalization Tools（ANTs）[32] 实现，使用互信息相似度度量。Unsup-VoxelMorph 使用 NCC 损失函数以及平滑损失训练。在我们的少标签情况下，LC-VoxelMorph 使用 Dice 损失函数以及平滑损失训练，我们测试了具有不同的平滑损失权重的 LC-VoxelMorph（$\lambda^{smooth} = 1$ 和 $\lambda^{smooth} = 0.1$）来验证缺乏纹理的标签对形变失真的影响。

3）**实施细节：**PC-Reg 基于 Tensorflow 实现的，在单个 12GB 的 NVIDIA TitanX GPU 上运行训练与测试。在实验中，与 Unsup-VoxelMorph [1] 论文中的设置相同，我们设置 $\lambda^{smooth} = 1$ 以同时实现准确的配准和光滑的形变。我们将 λ^{sty} 设置为 0.5 以削弱由于形变图像-固定标签对不对齐对感知网络学习的影响。我们采用 Adam 优化器来训练所有基于深度学习的模型，学习率设置为 1×10^{-4}，批大小设置为 1 以节省内存。在整个训练中，数据集被输入网络迭代训练了 400 遍以获得良好的拟合，其中 200 遍用于 PC-Reg 的第一阶段训练，200 遍用于 PC-Reg 的第二阶段。为了提高泛化能力，我们使用了 x、y 和 z 轴方向随机镜像和 $[-20°, 20°]$ 范围内的旋转作为数据增强。我们的源代码发布在https://github.com/YutingHe-list/PC-Reg-RT。

4）**评价指标：**为了评估配准结果，我们计算了变形标签 $y_w = T(y_m, \phi)$ 与固定标签 y_f 之间的 Dice 相似系数（DSC）[%] 和平均表面距离（ASD）[19]。DSC 越高，ASD 越低，说明配准的准确性越好。为了评估我们的保纹理能力，我们计算了 Jacobian 矩阵 $J_\phi(p) = \nabla\phi(p) \in R^{3\times3}$，以捕捉围绕体素 p 的形变向量 ϕ 的局部性质 [4]。对于 $J_\phi(p) \leq 0$ 的那些体素，被记录为指示折叠的奇点。我们计算了每个 DL 方法的

[1]https://www.elastix.org/

$J_\phi(p) \leq 0$ 的比例 [%]，以定量衡量 DVF 的保纹理能力，该度量越小，变形的保纹理能力越好 [33]。还统计了这些模型的 GPU 时间和 CPU 时间，以分析它们的时间效率。计算了分割 y' 与其标签 y 之间的 DSC[%]，以分析感知 CNN 的感知准确性。还提供了这些指标的标准差（std），以评估这些模型的稳定性。

V. 结果和分析

我们的框架消除了背景的干扰，感知具有清晰边界的 ROIs，保留了真实的纹理，并在少样本情况下提高了泛化能力，从而在 ROIs 上实现了竞争性的配准。

A. 定量对比分析

如表I所示，使用了反向教学法的 PC-Reg（PC-Reg-RT）在三个配准任务中均取得了最高的配准精度，相对平滑的形变过程，较块的配准速度：

配准精度的优越性：如表I所示，PC-Reg 仅使用五个标签，即可在心脏 CT、颈椎 CT 和脑 MR 配准任务上取得最佳的配准精度。PC-Reg 采用感知网络去除了任务无关的显著背景区域，使对应行为能够关注到任务感兴趣的区域进行清晰的区域边界对齐，从而在心脏 CT 上获得了 79.0% 的配准 DSC 和 1.93 的 ASD，在颈椎 CT 上获得了 81.4% 配准 DSC 和 0.66 的 ASD。与 LC-VoxelMorph（$\lambda^{smooth} = 1$）相比，它在心脏和颈椎 CT 配准上分别取得了 5.8% 和 1.0% 的配准 DSC 提升，这是由于我们对感知和对应的解耦为这两个行为的学习提供了独立的特征表示，隔离了彼此之间的干扰。当进一步使用本文所提出的反向教学法时（PC-Reg-RT），整体模型在心脏 CT、颈椎 CT 和脑 MR 图像配准任务上分别取得了 6.7%、5.3% 和 1.0% 的配准 DSC 提升，以及 0.61、0.25 和 0.6 的 ASD 降低。这是由于无标签数据中丰富的结构信息和风格信息有效地促进了感知网络的性能，从而为对应网络提供更加精准的感兴趣区域，进一步促进了配准性能的提升。表中所示的分割 DSC 在三个任务中分别提升了 6.3%, 20.6% 和 2.9%。如图6所示的各具体结构的配准上，进一步可以发现，PC-Reg-RT 在心脏 CT、颈椎 CT 和脑 MR 上的所有结构均获得了具有竞争力的配准性能。尤其是在颈椎 CT 图像配准上，由于其背景的大量干扰，Unsup-VoxelMorph 和 SyN 都发生了变形严重，导致它们甚至降低了 ROI 的对齐度，并出现了大幅度的性能波动。

有效的避免失真能力：PC-Reg 消除了与任务无关的背景区域，同时避免了标签中缺乏真实纹理导致的非对齐边缘区域过大的形变失真。在心脏 CT、颈椎 CT 和脑 MR 配准任务中，PC-Reg 和 PC-Reg-RT 均获得了极小 $|J_\phi| \leq 0$ ($< 1\%$)，这说明在形变的过程中，模型可以有效的保持图像原本的拓扑结构，不发生失真。而 Unsup-VoxelMorph 受到了背景中的与任务无关的显著区域的影响，在三个任务上的配准都产生了较大的扭曲（4.48% a），10.96% b），1.37% c））。特别是在颈椎 CT 配准中，SyN 和 Unsup-VoxelMorph 寻求获得全局对齐，这使得感兴趣区域的配准精度将向那些显著的任务无关区域妥协，最终导致了在颈椎配准上极差的性能（39.4% 和 50.1% 的配准 DSC b））。而 SyN 的 ASD 和 $|J_\phi| \leq 0$ 甚至由于非常夸张的形变，根本无法计算。直接在配准训练中加入分割标签的损失使得 LC-VoxelMorph（$\lambda^{smooth} = 1$）在心脏 CT 上获得了更平滑的形变（0.38% a））， 这是由于标签帮助配准学习获得语义感知能力，这种能力将使得配准学习关注感兴趣区域，避免背景显著任务无关区域达到干扰，但是由于标签数据量的限制，无法进行大数据学习，因此该方法的配准 DSC 下降至 73.2% a)。当我们把平滑损失的权重降至 0.1 时，LC-VoxelMorph（$\lambda^{smooth} = 0.1$）的配准 DSC 在三个任务上分别提高至 77.0%、82.3% 和 79.9%。但是，由于标签中缺乏真实的纹理约束的同时平滑损失约束下降，导致了较大的扭曲失真，其 $|J_\phi| \leq 0$ 在心脏 CT 和颈椎 MR 配准任务上增加至 3.43% 和 1.85%。DeepRS 受到标签和背景的严重干扰，在三个任务上都发生了验重的扭曲，$|J_\phi| \leq 0$ 极高。PC-Reg-RT 去除了背景干扰，并且利用感兴趣区域原有的纹理特征约束内部的平滑配准，从而获得了极小的 $|J_\phi| \leq 0$（0.37%、0.11% 和 0.04%）和高的配准 DSC（85.7%、86.7% 和 80.0%)。与 Unsup-VoxelMorph 和 LC-VoxelMorph（$\lambda^{smooth} = 0.1$）相比，PC-Reg-RT 具有更好的保持形变前后拓扑一致的能力和优秀的配准精度。

出色的时间效率：基于深度学习的一次前向运算即可获得配准结果的能力以及 GPU 的快速运算能力，PC-Reg 的时间效率与其他 DL 模型相当，并且比传统模型快超过 10 倍。而传统的 B 样条法和 SyN 在每个任务上都需要超过 10 秒的时间，这导致这些方法在一些有实时性要求的场景下无法使用。

结构细节：我们的 PC-Reg-RT 在心脏 CT、颈椎

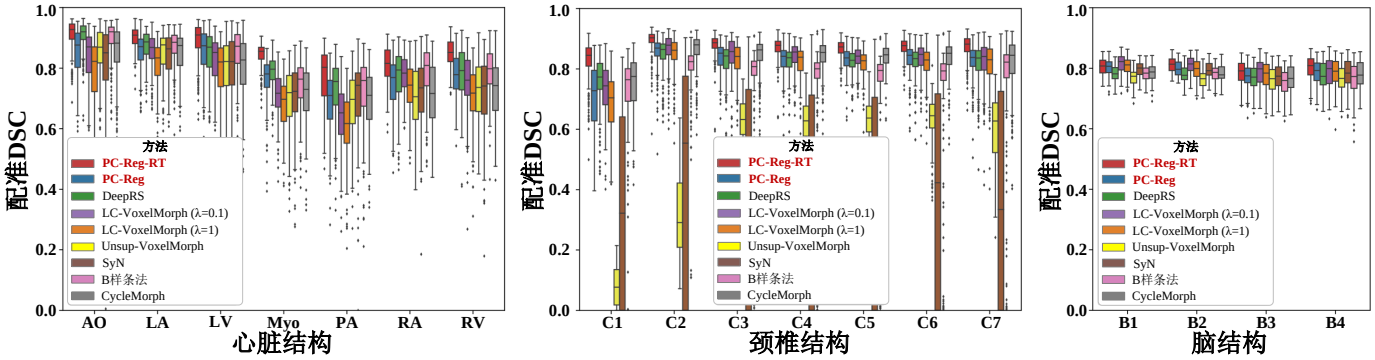

图 6: 箱线图展示了各模型在心脏、颈椎和脑各结构的具体性能，PC-Reg-RT 在每个结构上都取得了优秀的配准精度。

表 I: PC-Reg 框架在三个配准任务上拥有极大的优势。PC-Reg-RT 在配准 DSC 和平均表面距离（ASD）上取得了最优的性能。在时间效率上，也与其他基于深度学习的方法相当（$< 1s$）。在平滑性方面，PC-Reg 拥有极小的 $|J_\phi| \leq 0$，表明了我们的方法拥有保持配准后图像真实性的能力。

方法	配准			运算时间		分割
	$\text{DSC}_{\%\pm std}\uparrow$	$\text{ASD}_{voxel\pm std}\downarrow$	$\lvert J_\phi\rvert \leq 0_{\%\pm std}\downarrow$	$\text{CPU}_{s\pm std}\downarrow$	$\text{GPU}_{s\pm std}\downarrow$	$\text{DSC}_{\%\pm std}\uparrow$
a）**心脏** CT **图像配准**						
仅仿射变换初始化	$64.0_{\pm 12.5}$	$3.37_{\pm 0.86}$	-	$5.98_{\pm 0.55}$	-	-
B 样条法 [8]	$80.8_{\pm}10.4$	$1.69_{\pm 0.63}$	$0.34_{\pm 0.51}$	$40.19_{\pm 1.59}$	-	-
SyN [9]	$75.5_{\pm 12.7}$	$2.31_{\pm 0.90}$	$0.50_{\pm 0.16}$	$23.70_{\pm 4.33}$	-	-
Unsup-VoxelMorph [1]	$75.8_{\pm 11.8}$	$2.18_{\pm 0.74}$	$4.48_{\pm 1.61}$	-	$0.22_{\pm 0.16}$	-
LC-VoxelMorph（$\lambda = 1$）[4]	$73.2_{\pm 11.6}$	$2.43_{\pm 0.68}$	$0.38_{\pm 0.23}$	-	$0.23_{\pm 0.41}$	-
LC-VoxelMorph（$\lambda = 0.1$）[4]	$77.0_{\pm 11.6}$	$2.04_{\pm 0.58}$	$3.43_{\pm 0.79}$	-	$0.23_{\pm 0.31}$	-
CycleMorph [13]	$76.5_{\pm 9.4}$	$2.12_{\pm 1.01}$	$0.64_{\pm 0.18}$	-	$0.22_{\pm 0.25}$	-
DeepRS [5]	$81.5_{\pm 7.2}$	$1.71_{\pm 0.77}$	$7.03_{\pm 1.18}$	-	$0.65_{\pm 0.10}$	$87.4_{\pm 6.4}$
PC-Reg	$79.0_{\pm 9.9}$	$1.93_{\pm 0.56}$	$0.40_{\pm 0.18}$	-	$0.54_{\pm 0.51}$	$83.1_{\pm 12.9}$
PC-Reg-RT	$85.7_{\pm 7.3}$	$1.32_{\pm 0.38}$	$0.37_{\pm 0.28}$	-	$0.54_{\pm 0.19}$	$89.4_{\pm 6.1}$
b）**颈椎** CT **图像配准**						
仅仿射变换初始化	$64.8_{\pm 10.2}$	$1.37_{\pm 0.34}$	-	$6.35_{\pm 0.70}$	-	-
B 样条法 [8]	$74.2_{\pm 18.5}$	$1.15_{\pm 1.58}$	$0.45_{\pm 0.98}$	$38.62_{\pm 1.72}$	-	-
SyN [9]	$39.4_{\pm 34.7}$	-	-	$21.30_{\pm 9.70}$	-	-
Unsup-VoxelMorph [1]	$50.1_{\pm 22.1}$	$3.09_{\pm 0.58}$	$10.96_{\pm 0.41}$	-	$0.29_{\pm 0.17}$	-
LC-VoxelMorph（$\lambda = 1$）[4]	$80.4_{\pm 8.4}$	$0.72_{\pm 0.25}$	$0.25_{\pm 0.09}$	-	$0.29_{\pm 0.16}$	-
LC-VoxelMorph（$\lambda = 0.1$）[4]	$82.3_{\pm 7.6}$	$0.65_{\pm 0.22}$	$1.85_{\pm 0.42}$	-	$0.29_{\pm 0.17}$	-
CycleMorph [13]	$82.5_{\pm 6.8}$	$0.62_{\pm 0.35}$	$0.13_{\pm 0.06}$	-	$0.34_{\pm 0.38}$	-
DeepRS [5]	$81.7_{\pm 5.7}$	$0.65_{\pm 0.31}$	$2.06_{\pm 0.39}$	-	$0.86_{\pm 0.13}$	$86.3_{\pm 8.6}$
PC-Reg	$81.4_{\pm 8.1}$	$0.66_{\pm 0.23}$	$0.16_{\pm 0.08}$	-	$0.74_{\pm 0.60}$	$63.8_{\pm 20.4}$
PC-Reg-RT	$86.7_{\pm 5.0}$	$0.41_{\pm 0.15}$	$0.11_{\pm 0.06}$	-	$0.71_{\pm 0.23}$	$84.4_{\pm 12.6}$
c）**脑** MR **图像配准**						
仅仿射变换初始化	$75.5_{\pm 3.7}$	$1.25_{\pm 0.21}$	-	$7.14_{\pm 0.51}$	-	-
B 样条法 [8]	$77.0_{\pm 3.9}$	$1.15_{\pm 0.22}$	0	$40.32_{\pm 0.62}$	-	-
SyN [9]	$78.5_{\pm 3.8}$	$1.07_{\pm 0.21}$	0	$19.67_{\pm 1.46}$	-	-
Unsup-VoxelMorph [1]	$76.5_{\pm 3.7}$	$1.09_{\pm 0.19}$	$1.37_{\pm 0.19}$	-	$0.30_{\pm 0.30}$	-
LC-VoxelMorph（$\lambda = 1$）[4]	$79.0_{\pm 4.1}$	$1.07_{\pm 0.22}$	$0.14_{\pm 0.03}$	-	0.30±0.29	-
LC-VoxelMorph（$\lambda = 0.1$）[4]	$79.9_{\pm 3.9}$	$1.02_{\pm 0.20}$	$1.37_{\pm 0.15}$	-	$0.30_{\pm 0.28}$	-
CycleMorph [13]	$77.7_{\pm 3.7}$	$1.06_{\pm 0.20}$	0	-	$0.32_{\pm 0.42}$	-
DeepRS [5]	$77.6_{\pm 3.6}$	$1.05_{\pm 0.19}$	$1.34_{\pm 0.24}$	-	$1.04_{\pm 0.12}$	$81.7_{\pm 3.4}$
PC-Reg	$79.0_{\pm 3.4}$	$1.03_{\pm 0.18}$	$0.02_{\pm 0.01}$	-	$0.71_{\pm 0.39}$	$79.4_{\pm 3.5}$
PC-Reg-RT	$80.0_{\pm 3.4}$	$0.97_{\pm 0.18}$	$0.04_{\pm 0.02}$	-	$0.71_{\pm 0.40}$	$82.3_{\pm 3.3}$

CT 和脑 MR 上实现了竞争性的配准性能（图6）。在颈椎 CT 中，Unsup-VoxelMorph 和 SyN 的结果由于背景的大量干扰而出现严重变形和扭曲，以至于它们甚至减少了 ROIs 的对齐并出现了较大的性能波动。

B. 定性对比分析

如图7所示，PC-Reg 在感兴趣区域边界上具有良好的配准效果的同时拥有较少的扭曲，表现在三个方面：

边界上的精细配准：如图7中所展示的心脏案例 2 和颈椎案例 1 中，PC-Reg-RT 通过分离感知行为，为对应行为提供具有清晰边界的感兴趣区域，从而使得这些区域在边界上获得精细配准。由于完全无监督的配准方法缺乏对感兴趣区域的感知能力，Unsup-VoxelMorph，SyN 和 B 样条法的配准结果都较为粗糙，特别是在结构小且变化较大的颈椎 CT 中，甚至比最原始的仿射变换后的初始状态更差。添加标签后，LC-VoxelMorph（$\lambda^{smooth} = 0.1$），LC-VoxelMorph（$\lambda^{smooth} = 1$）和 DeepRS 获得了对感兴趣区域的感知能力，能够实现更好的对感兴趣区域的配准，但是这些方法把感知和对应混合在一个网络中进行学习，导致了这两个行为的学习互相干扰，使得这些方法对于和边界等细节区域的配准不够精确。

消除背景区域的大扭曲：PC-Reg-RT 通过感知网络感知感兴趣区域，从而消除区域外部背景区域中显著但与任务无关区域带来的巨大形变扭曲。如图7中的颈椎 CT 案例 2 所示，我们的 PC-Reg 和 PC-Reg-RT 专注于感兴趣区域，因此背景中任务无关的解剖结构不会干扰对应过程。最终，配准仅在感兴趣区域上进行独立的形变，从而消除了背景带来的任务无关的扭曲。SyN、B 样条法和 Unsup-VoxelMorph 没有感知能力，被背景中的结构，因此产生了夸张的形变。虽然标签将感知能力嵌入到 LC-VoxelMorph（$\lambda^{smooth} = 1$）、LC-VoxelMorph（$\lambda^{smooth} = 0.1$）和 DeepRS 中后使得这些方法同样更加关注感兴趣区域，削弱了背景中结构的影响，但由于它们的感知和对应能力被混合在一个网络中，无法去除背景的影响，因此他们依旧在背景上有着额外的扭曲。

平滑的感兴趣区域形变：如图7中的心脏案例 1 所示，PC-Reg 使用固定图像和移动图像的感兴趣区域来训练对应以获得位移形变场，在这样的训练过程中，感兴趣区域内部保持的纹理信息约束其内部不产生夸张的形变，使得 PC-Reg 的形变具有很强的纹理保持能力。而 LC-VoxelMorph（$\lambda^{smooth} = 1$）直接使用没有真实纹理信息的标签进行训练，导致配准结果在感兴趣区域内部存在严重的扭曲，降低了其配准结果的临床应用潜力。当其平滑损失的权重 λ^{smooth} 减少到 0.1 时，由于标签约束更加强烈，配准结果显示出更严重的扭曲问题。无监督的方法（Unsup-VoxelMorph，SyN 和 B 样条法）使用纹理和结构信息作为约束以避免在感兴趣区域内部产生夸张的扭曲，但背景上的大规模形变使得全局发生大的扭曲。

C. 模型感知能力分析

我们的反向教学方法只使用五个有标签图像就能从大量的无标签图像中生成额外的训练数据，从而教授感知网络丰富的结构和风格信息，并提高其感知更准确的感兴趣区域的能力，在实验中获得了 89.4%，84.4% 和 82.3% 的心脏，颈椎和大脑的分割 DSC。如图8所示，在视觉上也不难发现，加入反向教学法后分割结果得到了显着的提升，边界被分割得更加精细同时也减少了误分割的情况。在心脏 CT 图像的放大区域中，我们能够发现感知网络能够分割出完整的解剖结构区域和光滑的边界。同样地在颈椎分割中，PC-Reg-RT 也能够清晰地分割不同锥体，有效地避免了由于在少量图像上学习导致的过拟合问题（PC-Reg）。

D. 消融实验

1）**在更少标签情况下的有效性**：如图9所示，当标签数量减少时，PC-Reg-RT 仍能在心脏 CT 上取得良好的配准性能。我们可以发现两个有趣的现象：a）在非常少的标签下，本文所提出的反向教学法能够带来了显著的性能提升，哪怕只有一张有标签图像，PC-Reg-RT 仍然具有出色的配准性能。只有一个标签时，PC-Reg 仅拥有 64.1% 的配准 DSC 和 15.4% 的分割 DSC，而加入反向教学法之后，PC-Reg-RT 获得了 15.0% 的配准 DSC 和 65.8% 的分割 DSC 的显著提升。这归功于无标签图像丰富的结构和风格信息加入到感知网络的学习中，从而增强了感知网络的泛化能力，使得分割 DSC 提升的同时，配准性能也进一步提升。b）随着标签数量的增加，PC-Reg-RT 的性能将进一步提高。当标签数量增加到五个时，无反向教学法的 PC-Reg 就能够获得良好的性能，而 PC-Reg-RT 在反向教学法的促进下，依旧可以获得 6.6% 的配准 DSC 和 8.2%的分割 DSC 的提升。

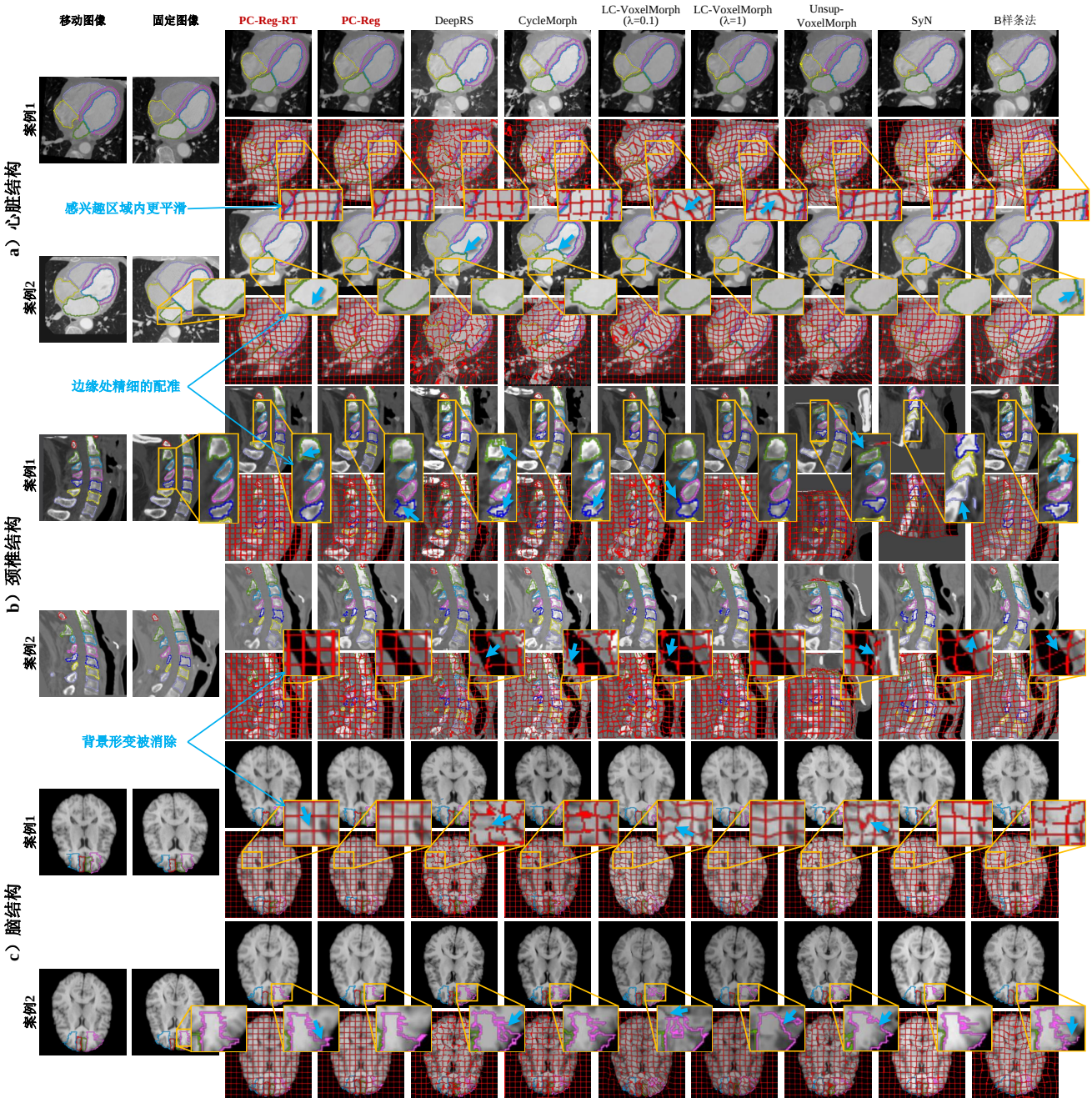

图 7: 定性评估展示了 PC-Reg 在三个配准任务上的视觉优越性。感知-对应解耦消除了背景的破坏性变形，减少 DVF 的扭曲，从而实现了对 ROI 的精细配准和出色的纹理保留能力。

表 II: 反向教学法中感知模型对风格特征和对结构特征的学习对整体分割和配准性能带来的贡献。

$\mathcal{L}_\theta^{str}$	$\mathcal{L}_\theta^{sty}$	配准 $\text{DSC}_{\%\pm std}\uparrow$	分割 $\text{DSC}_{\%\pm std}\uparrow$
		$79.0_{\pm 9.9}$	$83.1_{\pm 12.9}$
✓		$82.1_{\pm 9.2}$	$84.1_{\pm 14.0}$
	✓	$85.4_{\pm 7.3}$	$89.1_{\pm 6.1}$
✓	✓	$85.7_{\pm 7.3}$	$89.4_{\pm 6.1}$

2）**无标签图像的结构和风格特征的贡献**：如表II所示，我们对风格损失 $\mathcal{L}^{sty}$ 和结构损失 $\mathcal{L}^{str}$ 的消融实验表明了来自无标签图像的丰富的结构和风格特征对我们的少标签学习有着巨大的贡献。来自大量无标签图像的结构信息使得 PC-Reg-RT 中的感知网络学习到更多样的结构知识，从而在配准 DSC 和分割 DSC 上都取得了 3.1% 和 1.0% 的提升。无标签图像中的丰富的风格同样引导感知网络学习到对更多风格的泛化能力，从而使配准 DSC 和分割 DSC 分别提高了 6.4% 和 6.0%。

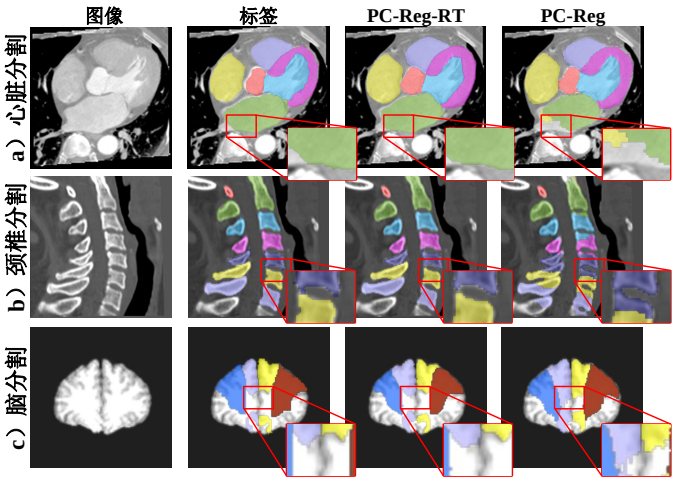

图 8: 反向教学法显著提高了感知网络的泛化能力，并且仅使用五个标签就能够感知极为准确的感兴趣区域。

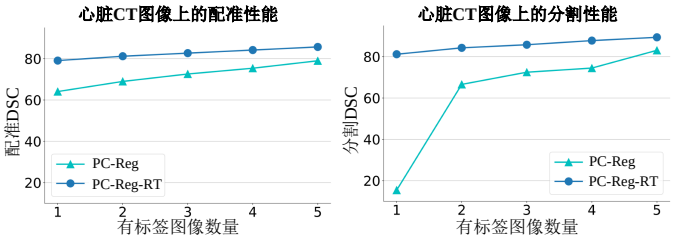

图 9: 我们的逆向教学使得我们的方法在少量标签情况下仍能保持优异的配准性能。线形图展示了我们的 PC-Reg（-RT）在不同标记数量下在心脏 CT 上的 Reg-DSC 和 Seg-DSC。

当同时使用风格损失和结构损失时，PC-Reg-RT 最终获得了惊人的 6.7% 配准 DSC 和 6.3% 分割 DSC 的提升。

VI. 总结和讨论

在本文中，我们提出了用于少样本弹性形变医学图像配准框架，PC-Reg，显著提高了对感兴趣区域的配准准确性。我们的感知对应结构框架将配准的感知和对应行为分离为两个卷积神经网络（CNNs），使它们拥有独立的优化和特征表示，隔离标签干扰的同时将 ROIs 的解剖知识嵌入到配准过程中，从而无需因标签中缺乏纹理信息而造成图像的纹理破坏。对于少样本学习，我们的反向教学法利用对应 CNN 的对齐能力，从多样的无标签图像中生成额外的训练数据，从而为感知 CNN 传授丰富的结构和样式知识。我们在心脏 CT、颈椎 CT 和脑 MR 图像上进行的实验证明，PC-Reg 具有竞争性的配准准确性、有效的保纹理能力和出色的时间效率。与 LC-VoxelMorph($\lambda = 1$) 相比，我们取得了分别

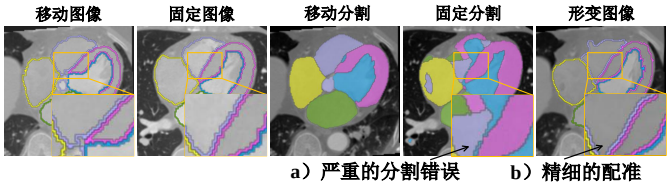

图 10: 由于训练过程中感知网络分割结果的误差，使得对应网络不得不学习到一种自我纠正的能力，从而实现在分割失误的情况下依旧可以获得较好的配准结果。

为 12.5%、6.3% 和 1.0% 的 Reg-DSC 提升，展现出我们框架在临床实践中的巨大潜力。

我们的 PC-Reg 显著促进了只有少量标签情况下的弹性形变医学图像配准性能，但框架中的解耦过程会导致更多的模块，进而导致这些模块之间的错误在传递过程中累积。幸运的是，即时在每个模块中进行检查和纠正将有效地避免错误的积累。感知和对应之间的后处理 [34] 将优化感知的 ROIs 和生成的额外训练数据，从而避免了错误结果的干扰。度量学习 [5] 还可以生成一个权重图，在训练中减弱不对齐区域，避免了低效的学习。

在训练过程中，我们的对应 CNN 将学习自我校正的能力，即使来自我们感知 CNN 的 ROIs 不准确，也能实现精细的配准。通过光滑损失和 LNCC 损失中的纹理约束，配准结果受到平滑变形和相同纹理对齐的限制。当不准确的分割结果出现时，对应 CNN 被约束输出平滑和纹理对齐的配准结果，因此这个不准确的训练过程将教导对应 CNN 具备自我校正的能力。如图 10所示，当将不准确的固定 ROIs 用作输入（a）时，配准仍将产生准确的配准结果（b）。

未来工作： 多模态配准仍然是医学图像可变形配准中的一个具有挑战性的任务，因此我们 PC-Reg 在多模态配准任务上的未来研究具有重要意义。

参考文献

[1] 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.

[2] B. Zitova and J. Flusser, “Image registration methods: a survey,” Image and vision computing, vol. 21, no. 11, pp. 977–1000, 2003.

[3] G. Haskins, U. Kruger, and P. Yan, “Deep learning in medical image registration: a survey,” Machine Vision and Applications, vol. 31, no. 1, p. 8, 2020.

[4] 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.

[5] 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.

[6] 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.

[7] A. Caldas, A. Micaelli, M. Grossard, M. Makarov, P. Rodriguez-Ayerbe, and D. Dumur, "On task-decoupling by robust eigenstructure assignment for dexterous manipulation," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015, pp. 5654–5661.

[8] 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.

[9] 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.

[10] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered cnn regression," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, pp. 300–308.

[11] J. Wang and M. Zhang, "Deepflash: An efficient network for learning-based medical image registration," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 4444–4452.

[12] T. C. Mok and A. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 4644–4653.

[13] B. Kim, D. H. Kim, S. H. Park, J. Kim, J.-G. Lee, and J. C. Ye, "Cyclemorph: Cycle consistent unsupervised deformable image registration," Medical Image Analysis, vol. 71, p. 102036, 2021.

[14] M. Jaderberg, K. Simonyan, A. Zisserman et al., "Spatial transformer networks," in Advances in neural information processing systems, 2015, pp. 2017–2025.

[15] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, "Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking," in Bildverarbeitung für die Medizin 2019. Springer, 2019, pp. 309–314.

[16] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," International journal of computer vision, vol. 128, no. 2, pp. 261–318, 2020.

[17] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," International Journal of Multimedia Information Retrieval, pp. 1–19, 2020.

[18] 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.

[19] Y. He, G. Yang, J. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao et al., "Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation," Medical Image Analysis, p. 101722, 2020.

[20] P. Liu, M. Lyu, I. King, and J. Xu, "Selflow: Self-supervised learning of optical flow," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4571–4580.

[21] A. V. Dalca, G. Balakrishnan, J. V. Guttag, and M. R. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces." Medical Image Analysis, vol. 57, pp. 226–236, 2019.

[22] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey," Neurocomputing, vol. 381, pp. 61–88, 2020.

[23] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1352–1363, 2016.

[24] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," International Journal of Computer Vision, pp. 1–57, 2020.

[25] A. Medela, A. Picon, C. L. Saratxaga, O. Belar, V. Cabezón, R. Cicchi, R. Bilbao, and B. Glover, "Few shot learning in histopathological images: reducing the need of labeled data on biological datasets," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 1860–1864.

[26] 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.

[27] 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.

[28] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in Artificial intelligence and statistics, 2015, pp. 562–570.

[29] Y. Wu and K. He, "Group normalization," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.

[30] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of mri," Medical Image Analysis, vol. 31, pp. 77 – 87, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841516000219

[31] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, "Construction of a 3d probabilistic atlas of human cortical structures," Neuroimage, vol. 39, no. 3, pp. 1064–1080, 2008.

[32] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration." NeuroImage, vol. 54, no. 3, pp. 2033–2044, 2011.

[33] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 729–738.

[34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.